électronique

MDPI

Article

Prédire les caractéristiques des liaisons série à haut débit Basé sur un réseau neuronal profond (DNN) – Transformer Modèle en cascade

Liyin Wu [1], Jingyang Zhou [1], Haining Jiang [1], Xi Yang [1], Yong Zheng Zhan [2] et Yinhang Zhang [1,*]

[1] École de communication et d'ingénierie électronique, Université Jishou, Jishou 416000, Chine ; 2022700814@stu.jsu.edu.cn (LW); 2022700812@stu.jsu.edu.cn (JZ); 2023700812@stu.jsu.edu.cn (HJ); ynkej@163.com (XY)

[2] Shandong Yunhai Guochuang Cloud Computing Equipment Industry Innovation Co., Ltd., Jinan 250101, Chine ; sdzyz1989@163.com *

Correspondance : zyh149032@jsu.edu.cn

Résumé : Le niveau de conception des caractéristiques physiques du canal a une influence cruciale sur la qualité de transmission des liaisons série à haut débit. Cependant, la conception des canaux nécessite un processus complexe de simulation et de vérification. Dans cet article, un modèle de réseau neuronal en cascade construit à partir d'un réseau neuronal profond (DNN) et d'un transformateur est proposé. Ce modèle prend des caractéristiques physiques comme entrées et importe une réponse à un seul bit (SBR) en tant que connexion, qui est améliorée grâce à la prédiction des caractéristiques de fréquence et des paramètres de l'égaliseur. Dans le même temps, l'analyse de l'intégrité du signal (SI) et l'optimisation des liaisons sont réalisées en prédisant les diagrammes oculaires et les marges d'exploitation des canaux (COM). De plus, l'optimisation bayésienne basée sur le processus gaussien (GP) est utilisée pour l'optimisation des hyperparamètres (HPO). Les résultats montrent que le modèle en cascade DNN-Transformer permet d'obtenir des prédictions de haute précision de plusieurs métriques de prédiction et d'optimisation des performances, et que l'erreur relative maximale des résultats de l'ensemble de test est inférieure à 2 % dans le cadre de l'architecture d'égalisation d'un TX à 3 prises. FFE, un RX CTLE avec double gain DC et un RX DFE à 12 taps, qui est plus puissant que les autres modèles d'apprentissage profond en termes de capacité de prédiction.

Mots-clés : liaison à grande vitesse ; l'intégrité du signal; diagramme de l'œil ; marge opérationnelle du canal ; modèle en cascade

Référence : Wu, L. ; Zhou, J. ; Jiang, H. ; Yang, X. ; Zhan, Y. ; Zhang, Y. Prédire les caractéristiques des liaisons série à haut débit basées sur un réseau neuronal profond (DNN) : Modèle en cascade de transformateur. Électronique 2024, 13, 3064. https:// doi.org/ 10.3390/electronics13153064

Rédacteur académique : Shing-Hong Liu

Reçu : 1er juillet 2024
Révisé : 24 juillet 2024
Accepté : 31 juillet 2024
Publié : 2 août 2024

1. Introduction

À mesure que la bande passante de transmission de la technologie de liaison série filaire atteint le niveau GHz, il n'est plus possible d'assurer une transmission efficace du signal en optimisant simplement la structure diélectrique et la configuration. Les systèmes de liaison série à haut débit souffrent de graves problèmes d'intégrité du signal (SI) dus à l'effet pelliculaire, à la perte diélectrique, à la diaphonie, aux réflexions et à la gigue ; par conséquent, l'analyse SI devient de plus en plus stricte lors de la phase de conception des liaisons série à haut débit. L'analyse de simulation du SI comprend généralement deux étapes : les solveurs de champ électromagnétique (EMFS) et la simulation du système de circuits [1]. Premièrement, un EMFS est utilisé pour obtenir des paramètres S afin de caractériser la réponse en fréquence du circuit. Ensuite, ces paramètres S sont importés dans le système de circuit modèle pour une simulation dans le domaine temporel afin d'obtenir les principales métriques SI, notamment un diagramme oculaire, la réponse impulsionnelle et les formes d'onde transitoire

Bien que l'analyse SI traditionnelle basée sur des modèles physiques de liaisons à haut débit puisse offrir une grande précision, elle consomme beaucoup de temps et de ressources informatiques. L'interface de modèle algorithmique de spécification d'informations de tampon d'entrée/sortie (IBIS-AMI) est un modèle comportemental qui simule le comportement d'entrée/sortie et les algorithmes des liaisons série haute vitesse de bout en bout , simplifiant les détails physiques internes [2]. Par rapport aux modèles physiques, il présente les avantages de la rapidité, de la simplicité et d'une faible consommation de ressources, mais il est peu précis et manque de flexibilité. La comparaison des modèles de normes de conformité est une autre méthode couramment utilisée [3]. Cela permet d'estimer rapidement et intuitivement les performances du canal, mais cela entraîne inévitablement

rejette les marges des canaux en raison de la nécessité de satisfaire strictement des métriques discrètes [4]. Au contraire, la technique de marge opérationnelle du canal (COM) peut efficacement surmonter cet inconvénient en recherchant l'espace de conception optimal de l'ensemble de la liaison sous la forme d'un rapport signal sur bruit (SNR). Par rapport aux mesures SI traditionnelles telles que les diagrammes oculaires et le taux d'erreur binaire (BER), l'approche COM présente des avantages tels qu'un fonctionnement plus simple, une vitesse plus rapide et des tests plus efficaces. Évidemment, par rapport à l'évaluation du canal Annex69B pour Ethernet 10 Gb/s (10 GbE), l'utilisation de COM est plus précise [5]. Bien que l' approche COM puisse constituer un moyen efficace d'évaluer et d'optimiser les liaisons à haut débit, elle nécessite de nombreuses itérations pour la recherche spatiale et n'est pas suffisamment flexible pour analyser les canaux à haute capacité.

L'apprentissage automatique (ML) a été largement utilisé dans la simulation et la conception de liaisons à haut débit ces dernières années et montre la capacité d'améliorer l'efficacité de l'analyse SI. Les auteurs de [4,6–10] ont recherché des caractéristiques à partir de données de simulation et ont formé des modèles de réseau neuronal artificiel (ANN), de réseau neuronal profond (DNN) et de machine à vecteurs de support des moindres carrés (LS-SVM) pour remplacer les simulations de systèmes de circuits. pour la prédiction précise des mesures du domaine temporel (TD) et du domaine fréquentiel telles que la hauteur des yeux (EH)/ largeur des yeux (EW) et la perte de retour (RL)/perte d'insertion (IL) ; cependant, l'acquisition de données de simulation consomme beaucoup de ressources informatiques et de temps. Les réseaux de neurones Feedforward (FNN), la régression forestière aléatoire (RFR) et les machines à vecteurs de support (SVM) sont utilisés pour réaliser des prédictions simples de données de simulation [11-13], par exemple pour prédire les paramètres S et les réponses impulsionnelles. Lorsqu'elles traitent des données de simulation plus complexes, les méthodes traditionnelles de ML peuvent perdre de nombreuses fonctionnalités. Le réseau neuronal récurrent (RNN) est un modèle ML qui peut capturer efficacement les caractéristiques de données complexes, en particulier pour les données de séquence telles que les paramètres S et les réponses impulsionnelles. Les auteurs de [14–20] utilisent l'architecture RNN et Long Short-Term Memory (LSTM) pour créer des modèles de substitution permettant de prédire les formes d'onde de réponse transitoire de liaisons complexes à haut débit. Ces modèles de substitution indépendants peuvent effectuer une analyse SI basée sur des données de simulation, mais ne peuvent pas traiter les paramètres physiques des liens. L'analyse SI basée sur des paramètres physiques nécessite des méthodes de ML plus compliquées, et les auteurs de [21–24] ont obtenu des prédictions efficaces des paramètres physiques utilisés pour évaluer les performances des liaisons à haut débit en combinant plusieurs algorithmes d'apprentissage profond. De plus, une optimisation équilibrée de l'architecture pour les liaisons série à haut débit peut être obtenue sur la base du ML [4,25-27]. En conclusion, la prédiction des performances et l'optimisation de l'architecture [28] sont deux parties importantes des applica

Dans cet article, un modèle en cascade DNN-Transformer est proposé pour l'analyse SI et l'optimisation des liaisons série à haut débit. Ce modèle peut ignorer l'utilisation des EMFS et de la simulation de systèmes de circuits, et prédit directement les métriques SI, notamment les valeurs EH/EW, IL/RL, la réponse impulsionnelle et COM en fonction des paramètres physiques. Parallèlement, l'optimisation des liens peut être obtenue en utilisant ce modèle pour prédire les valeurs COM et les paramètres d'égalisation correspondants pour les liens avec différentes configurations d'égaliseur. De plus , l'optimisation bayésienne basée sur le processus gaussien (GP) est utilisée pour optimiser les hyperparamètres dans les mêmes conditions pour différentes combinaisons de modèles. Comparé à la prédiction des performances d'une liaison à haut débit à l'aide du ML traditionnel [6-10], DNN-Transformer peut utiliser directement les paramètres physiques de la liaison pour l'analyse, et sa précision de prédiction est nettement meilleure. Pour la prédiction de données de simulation telles que les réponses impulsionnelles, DNN-Transformer peut capturer plus de fonctionnalités et sa capacité de prédiction est plus précise que celle obtenue en utilisant un modèle RNN [14] ou LSTM [20] seul pour créer un modèle de substitution. Les résultats montrent que le modèle DNN-Transformer peut réaliser une prédiction plus efficace. Les performances et la faisabilité de la méthode proposée sont illustrées par les données de prédiction et les résultats graphiques présentés dans cet article.

Les principales contributions de cet article peuvent être résumées comme suit : (1) Sur

la base des paramètres physiques clés des canaux, des modèles de réseaux neuronaux sont utilisés pour analyser directement les performances des liaisons, ce qui évite les processus fastidieux de résolution des champs électromagnétiques et de simulation du système de circuits. .

Les principales contributions de cet article peuvent être résumées comme suit : (1) Sur la base des paramètres physiques clés des canaux, des modèles de réseaux neuronaux sont utilisés pour analyser directement les performances des liaisons, ce qui évite les processus fastidieux de résolution des champs électromagnétiques et de simulation du système de circuits. .

(2) Grâce à des modèles de réseaux neuronaux, la prédiction précise de plusieurs indicateurs SI et les paramètres de l'égaliseur sont atteints. De plus, nous montrons que l'analyse SI et les paramètres de l'égaliseur de liaison sont obtenus. De plus, nous montrons que l'analyse SI et le lien l'optimisation peut également être réalisée rapidement.

(3) Un modèle en cascade DNN – Transformateur est proposé, l'optimisation bayésienne est utilisée pour régler les hyperparamètres de ce modèle, et ses performances supérieures sont démontrées stratégique en le comparant avec d'autres modèles. en le comparant avec d'autres modèles.

Le reste de ce document est organisé comme suit : La section 2 discute des principes de base de ces travaux, incluant les principes des méthodes SI et COM. La section 3 décrit le fonctionnement de ce travail, y compris les principes des méthodes SI et COM. La section 3 décrit comme l'ensemble de données de simulation a été créé. Dans la section 4, l'idée de conception du DNN – Transformer, l'ensemble de données de simulation, a été créée. Dans la section 4, l'idée de conception du DNN-Transformer le modèle et les métriques de test sont présentés. La section 5 donne les résultats numériques pour démontrer le modèle et les métriques de test sont présentées. La section 5 donne les résultats numériques pour démontrer les performances de prédiction de notre modèle. Enfin, la section 6 conclut l'article. les performances de prédiction de notre modèle. Enfin, la section 6 conclut l'article.

2. Méthodes fondamentales 2. Méthodes fondamentales

2.1. Intégrité du signal et paramètres S 2.1. Intégrité du signal et paramètres S

À mesure que les taux de transmission augmentent, les inefficacités de la transmission de signaux parallèles traditionnelle. À mesure que les taux de transmission augmentent, les inefficacités des méthodes traditionnelles de transmission de signaux parallèles deviennent apparentes. Les méthodes de transmission numérique série à grande vitesse deviennent progressivement la transmission numérique en série devie à vitesse augmentent utilisé. La figure 1 montre la structure SerDes augmentent utilisée dans une liaison série haut débit système de liaison série. Il comprend une égaliseur (FFE), un capteur et une puce de réception avec un égaliseur à retour de décision (DFE) et un désérialiseur. un canal et une puce réceptrice avec un égaliseur à retour de décision (DFE) et un désérialiseur. à l'extrémité émettrice effectue principalement le signal pré-accentuation sur le signal. Le FFE à l'extrémité émettrice effectue nal pour améliorer sa composante haute fréquence. La composition d'un canal filaire comprend des traces, des vias et des composante haute fréquence. La composition d'un canal filaire est complexe, généralement connecteurs, etc. Ces structures physiques ex- hibit les caractéristiques du filtre passe-bas et peut augmenter la perte haute fréquence d'un signal, la qualité de transmission du signal et entraîne des problèmes SI. Le conseiller (CTLE) à l'extrémité de réception est principalement utilisé pour compenser la perte haute fréquence de L'égaliseur linéaire à temps continu (CTLE) à l'extrémité de réception est principalement utilisé pour compenser le canal, éliminer les précurseurs et supprimer la traînée de la réponse impulsionnelle. Le pour la perte haute fréquence du canal, éliminer les précurseurs et supprimer le DFE est principalement utilisé pour éliminer suivi de la réponse impulsionnelle. Le DFE est principalement utilisé pour éliminer les post-curseurs du et les réglages des réponse impulsionnelle. La combinaison et les réglages des paramètres des FFE, CTLE et DFE disposent de liaisons de paramètres du FFE, du CTLE et du DFE ont un impact significatif sur le SI de communication série filaires ultra-rapides un impact significatif sur le SI des liaisons de communication série filaires à très haut débit.

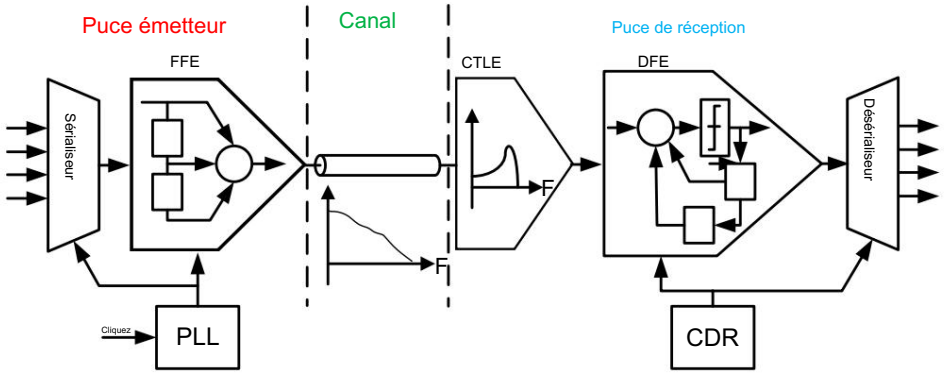
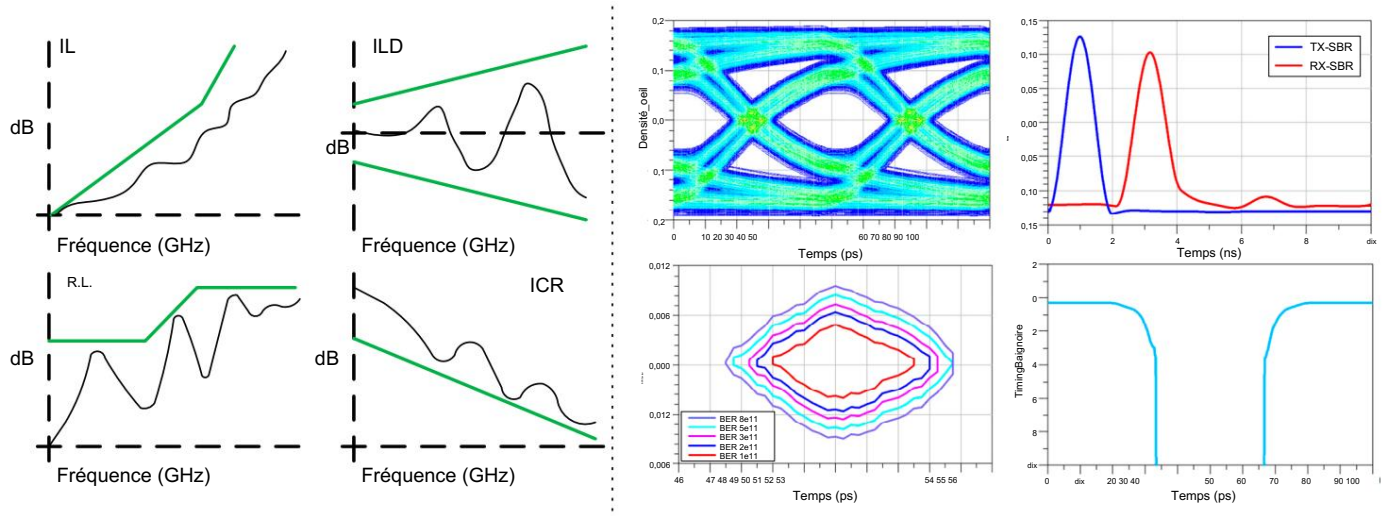


Figure 1. Structure commune d'un circuit SerDes.
Figure 1. Structure commune d'un circuit SerDes.

L'objectif principal de l'analyse SI est de détecter et de réduire les facteurs à l'origine des pertes, tels que la réflexion et la diaphonie. D'une part, le SI peut être analysé à partir du L'objectif principal de l'analyse SI est de détecter et de réduire les facteurs à l'origine des pertes, tels que le tramage, comme le tramage, la réflexion et la diaphonie. D'une part, le SI peut être analysé à partir du domaine fréquentiel, comme le montre la figure 2a. Cela comprend principalement IL, RL, perte d'insertion domaine fréquentiel, comme le montre la figure 2a. Cela inclut principalement IL, RL, la déviation de perte d'insertion (ILD) et le rapport perte d'insertion/diaphonie (ICR), qui sont ensuite viation (ILD) et le rapport perte d'insertion/diaphonie (ICR), qui sont ensuite placés dans un modèle pour évaluation. En revanche, le SI peut être évalué dans le placé dans un modèle pour évaluation. D'autre part, le SI peut être évalué dans le domaine temporel, par exemple au moyen de diagrammes oculaires, de simulations transitoires, de courbes de baignoire et de bits. taux d'erreur, comme le montre la figure 2b.

taux d'erreur sur les bits, comme le montre la figure 2b.

La méthode critique de l'analyse SI traditionnelle implique l'acquisition de paramètres S.
Ces paramètres contiennent des caractéristiques complètes du domaine de fréquence (FD) du canal de transmission et offrent une grande quantité d'informations sur des aspects tels que la réflexion.
tion, diaphonie et perte. De plus, les paramètres S peuvent être utilisés dans des simulations dans le domaine temporel pour générer des données telles que des diagrammes oculaires et des courbes de baignoire.



(un) (b)

Figure 2. (a) FD. (b) TD. Illustration de deux méthodes de simulation différentes pour détecter le SI.

Figure 2. (a) FD. (b) TD. Illustration de deux méthodes de simulation différentes pour détecter le SI.

2.2. Marge opérationnelle du canal

La méthode critique de l'analyse SI traditionnelle implique l'acquisition de paramètres S.
Ces paramètres contiennent des caractéristiques complètes du domaine de fréquence (FD) du canal de transmission et offrent une grande quantité d'informations sur des aspects tels que la réflexion.
perte. De plus, les paramètres S peuvent être utilisés dans les simulations dans le domaine temporel
données telles que des diagrammes oculaires et des courbes de baignoire.

La méthode COM est une méthode de caractérisation de liaison série à haut débit recommandée comme canal de par le groupe de travail IEEE 802.3 pour les tests de conformité des canaux. La définition officielle de la diaphonie et de la déclare qu'un « COM est une figure de mérite (FOM) pour les canaux déterminés à partir d'un minimum pour générer des référencez l'architecture PHY et les paramètres s du canal »[29]. Cette approche fournit un environnement relativement précis et équitable pour la conception physique des canaux en considérant 2.2. Marge opérationnelle du canal divers facteurs tels que la perte, la réflexion, l'interférence inter-symboles (ISI), la dispersion ISI, La méthode COM est une la diaphonie et les spécifications de l'appareil, permettant une évaluation relativement précise du canal par le groupe de performances et l'impact des canaux agresseurs sur les canaux victimes. L'impact du fait qu'un « COM est un facteur de mérite (FOM) pour les chaînes déterminé à partir d'une référence minimale de l'architecture PHY et des paramètres s du canal » [29]. Cette approche fournit un relativement d'un COM basé sur le FOM peut être amélioré en sélectionnant les paramètres de l'égaliseur. environnement précis et équitable pour la conception physique des canaux en tenant compte de divers Par conséquent, le calcul du COM peut également déterminer si la qualité du canal répond à des facteurs tels que la perte, la réflexion, l'interférence entre symboles (ISI), la dispersion ISI, la diaphonie et les exigences SI de l'émetteur-récepteur [30]. Comme le montre l'équation (1), le COM peut être une spécification provenant d'un appareil, permettant une évaluation relativement précise des performances du canal et pressé par le rapport entre l'amplitude du signal disponible As et l'amplitude du bruit statistique An . l'impact des canaux agresseurs sur les canaux victimes. L'impact des égalisations a été

pris en compte dans les versions ultérieures de la méthode COM, et la valeur d'une méthode basée sur COM sur le FOM peut être amélioré en sélectionnant les paramètres de l'égaliseur. Par conséquent,

$$COM = 20 \log 10 \left(\frac{UN_s}{UNE} \right) \quad (1)$$

Le calcul du COM peut également déterminer si la qualité du canal répond aux exigences de l'émetteur-récepteur. Exigences SI [30]. Comme le montre l'équation (1), le COM peut être exprimé par le rapport de

Le processus de dérivation du COM implique plusieurs étapes essentielles, notamment la détermination

de l'amplitude du signal disponible par rapport à l'amplitude statistique du bruit An.
extraire la fonction de transfert, convertir la fonction de transfert en réponse impulsionnelle, appliquer des algorithmes d'égalisation de l'émetteur et du récepteur et effectuer des calculs statistiques de bruit. La figure 3 présente un organigramme de dérivation du modèle COM et représente deux

$$\text{Comme} \quad COM = 20 \times \log 10 \quad (1)$$

Un chemins pour les canaux de la victime et de l'agresseur. De plus, le modèle prend également en compte les interférences S (tp/rp) du boîtier émetteur et récepteur, la résistance de terminaison Rd, les égaliseurs. Le processus de dérivation du

COM implique plusieurs étapes essentielles, notamment la dissuasion et les filtres. Pour mieux simuler les canaux

réels, le modèle COM intègre l' exploration gaussienne de la fonction de transfert, convertissant la fonction de transfert en réponse impulsionnelle,
bruit blanc et gigue à l'extrémité du récepteur. La figure 4 montre le processus détaillé de COM appliquant des algorithmes d'égalisation de l'émetteur et du récepteur et effectuant un bruit statistique.
calcul de la valeur.
calculs. La figure 3 présente un organigramme de dérivation du modèle COM et décrit deux
chemins pour les canaux de la victime et de l'agresseur. De plus, le modèle prend également en compte
les interférences S (tp/rp) des boîtiers émetteur et récepteur, la résistance de terminaison Rd, les égaliseurs,
et des filtres. Pour mieux simuler les canaux réels, le modèle COM intègre
bruit blanc et gigue à l'extrémité du récepteur. La figure 4 montre le processus détaillé de COM
calcul de la valeur.

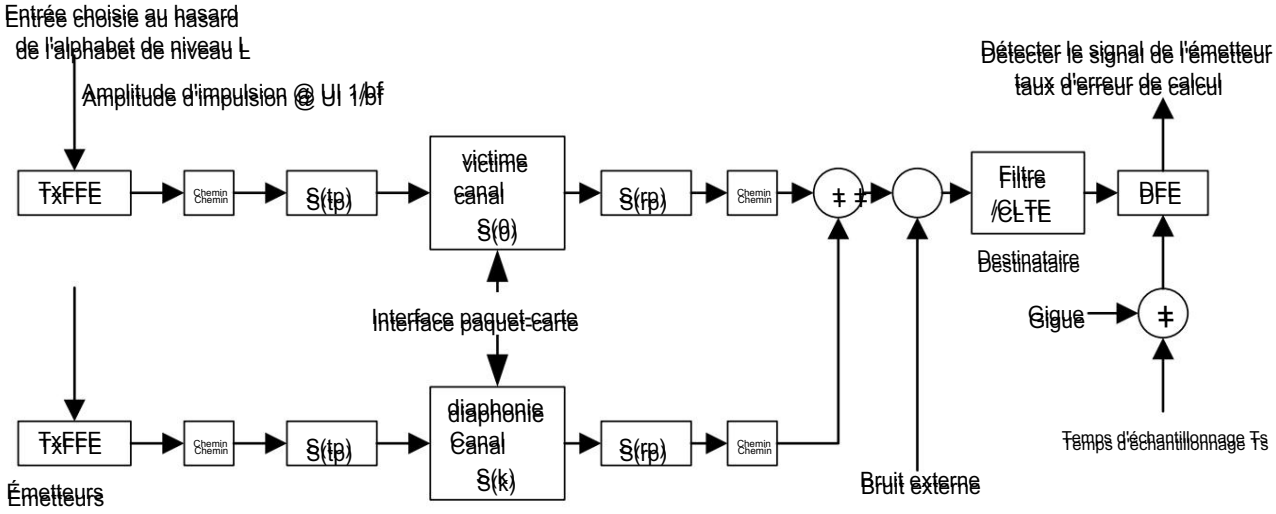


Figure 3. Un modèle COM typique.

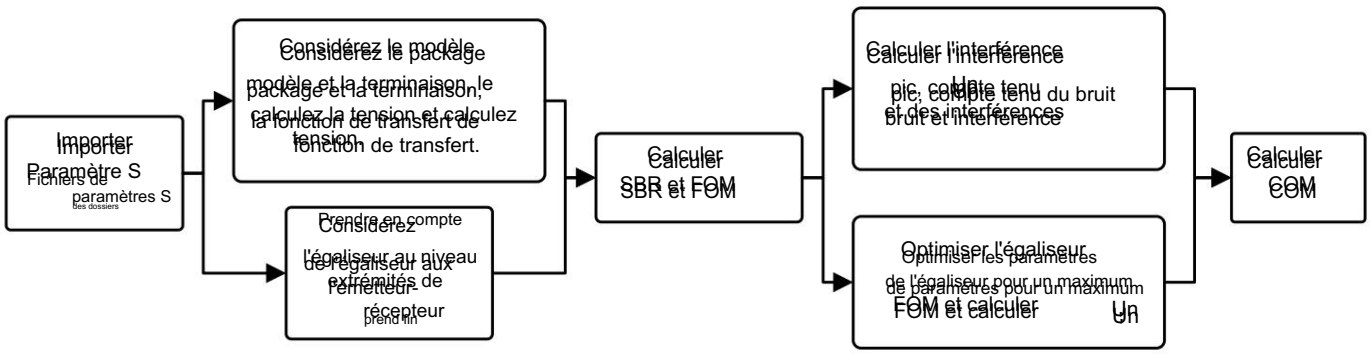


Figure 4. Processus de calcul COM.

L'approche COM implique d'explorer toutes les combinaisons de paramètres possibles du TX, l'approche égaliseur sont principalement définis en fonction de la perte du canal. Le FOM sert et les égaliseurs RX dans une plage définie pour trouver la configuration qui maximise le FOM, et les quantitative pour évaluer la qualité du canal et les performances de l'égaliseur. Ça prend Ce processus détermine les valeurs de As et An qui aboutissent au FOM optimal, en compte plusieurs facteurs affectant le SI, notamment l'ISI, la gigue, la diaphonie et le bruit. Le FOM

Les méthodes traditionnelles reposent sur divers indicateurs tels que la gigue, la hauteur des yeux et la des yeux, mais le COM sert de mesure complète pour évaluer une liaison série. Il peut s'étendre, mais le COM sert de métrique complète pour évaluer une liaison série. Ça peut

$$FOM = 10 \times \log 10 \frac{UN_s^2}{\sigma^2_{Émission} + \sigma^2_{ISI} + \sigma^2_{J.} + \sigma^2_{XTK} + \sigma^2_{N}} \quad (2)$$

Le numérateur As est dérivé de l'amplitude de la réponse impulsionnelle à ts , qui correspond au curseur principal de la réponse impulsionnelle h(t). Le dénominateur comprend la somme des écarts de tous les composants de bruit, de gigue et d'interférence. σ^2_{TX} représente le variance du bruit au niveau de l'émetteur, σ^2_{ISI} désigne la variance de l'amplitude ISI résiduelle, et $\sigma^2_{J.}$ indique la variance de l'amplitude de la gigue. Dans l'analyse COM, la gigue est prise en compte en convertissant la gigue horizontale en bruit vertical à l'instant d'échantillonnage ts . En plus, σ^2_{XTK} représente la variance totale de diaphonie de tous les chemins d'interférence, tandis que σ^2_{N} désigne le bruit blanc gaussien au point d'échantillonnage du récepteur.

L'approche COM consiste à explorer toutes les combinaisons possibles de paramètres du TX et les égaliseurs RX dans une plage définie pour trouver la configuration qui maximise le FOM.
Ce processus détermine les valeurs de As et An qui aboutissent au FOM optimal.

Les méthodes traditionnelles s'appuient sur divers indicateurs tels que la gigue, la hauteur des yeux et la largeur, mais le COM sert de métrique complète pour évaluer une liaison série. Il peut réduire considérablement le temps de calcul et le nombre d'itérations, et fournit un évaluation précise des performances du canal avant et après l'égalisation.

réduit considérablement le temps de calcul et le nombre d'itérations et fournit une évaluation précise des performances du canal avant et après l'égalisation.

3. Construction d'un ensemble de données

3. Construction d'un ensemble de données

3.1. Conception des canaux et fractionnement des ensembles de données

3.1. Conception des canaux et fractionnement des ensembles de données

Dans ce travail, les ensembles de données de canal ont été générés à l'aide d'un logiciel professionnel de conception de PCB. Comme le montre la figure 5, logiciels Altium Designer 20.0.13 et Allegro 17.4 (logiciel EDA) professionnel. ont été principalement utilisés pour créer les traces des lignes différentielles à grande vitesse. Les sous-marins ont été principalement utilisés pour créer les traces des lignes différentielles à grande vitesse. Ensuite, Ensuite, les fichiers de configuration de PCB ont été importés dans Advanced Design System 2021 (ADS) ou les fichiers de configuration de PCB ont été importés dans Advanced Design System 2021 (ADS) ou HFSS 2023 R1 pour la simulation électromagnétique (EM). Notre conception est conforme aux spécifications R1 pour la simulation électromagnétique (EM). Notre conception est conforme aux spécifications USB tion pour les lignes de transmission différentielles de bout en bout USB, y compris des paramètres tels que les lignes de transmission différentielles de bout en bout de ligne, y compris des paramètres tels que la largeur de ligne et la largeur et la longueur, l'impédance caractéristique des canaux de transmission et la longueur différentielle, l'impédance caractéristique des canaux de transmission et l'espacement différentiel des lignes. intérligne.

articles. Comme le montre la figure 5, logiciels Altium Designer 20.0.13 et Allegro 17.4 (logiciel EDA). Comme le montre la figure 5, Altium Designer 20.0.13 et Allegro 17.4 (logiciel EDA)

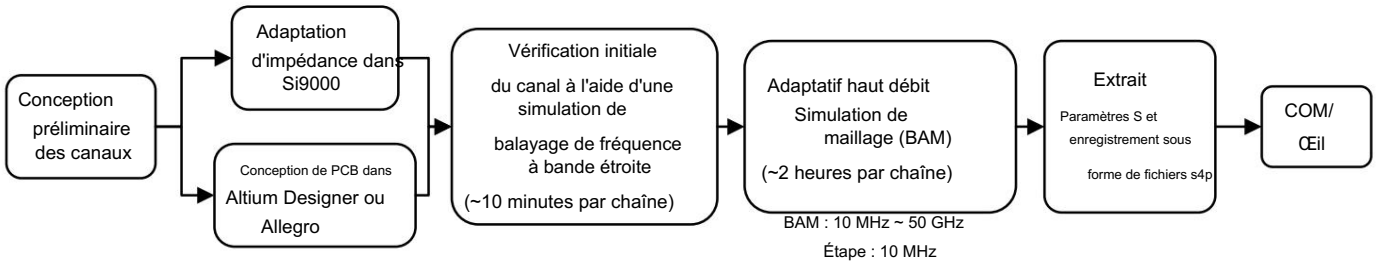


Figure 5. Processus de fabrication et de simulation des canaux.

Figure 5. Processus de fabrication et de simulation des canaux.

Les caractéristiques différenciées des données de canal, telles que présentées dans le tableau 1, principalement incluent des paramètres tels que le matériau, les dimensions des traces et le type de PCB. Les caractéristiques de la carte les deux caractéristiques numériques, comme la permittivité (Er), le facteur de dissipation. englobent et en l'épaisseur de la carte et en caractéristiques catégorielles telles que le type de PCB. Er consiste en (Df) la permittivité de la feuille et l'angle tangent de la perte diélectrique (TanD). Df est dérivé de la partie réelle de la permittivité de la feuille et de l'angle tangent de perte diélectrique (TanD). Df est la partie réelle de paramètres incluant le bas et le haut dérivé du modèle Djordjevic, qui est défini par des hautes fréquences de TanD et permittivité relative à haute fréquence. Dans cette étude, Df est quantifié TanD et la permittivité relative à haute fréquence. Dans cette étude, Df est quantifié selon les valeurs numériques de TanD et ses hautes et basses correspondantes selon les valeurs numériques de TanD et ses hautes et basses fréquences correspondantes. fréquences. Le type de PCB est classé soit en type stripline , représenté par une valeur numérique. valeur numérique de 1, ou le type microruban, représenté par 0. de 1, ou le type microruban, représenté par 0.

Ces caractéristiques de trace sont quantifiées en fonction de paramètres spécifiques tels que la longueur, la largeur, l'épaisseur, la conductivité, le profil et l'espacement différentiel des lignes. Le profil est le tableau 1. Paramètres des caractéristiques du canal et méthodes quantitatives. caractérisé par la forme géométrique de la trace, comme le nombre de coins. L'autre entrée sortie

Variable de fonctionnalité

les caractéristiques sont directement représentées par leurs valeurs numériques correspondantes.

Variable de fonctionnalité		Sortir
Euh	Permittivité de la feuille PCB pièce réelle et TanD	
Df	Facteur de dissipation de la feuille PCB	
Tracer	Longueur de trace du PCB d'entrée Largeur de trace de PCB Espace de ligne différentiel Longueur de trace PCB Épaisseur de trace PCB Largeur de trace de PCB Conductivité de trace de PCB Structure de trace PCB à espace de lignes différentielles (profil)	Diagrammes oculaires Valeurs COM Paramètres de l'égaliseur Diagrammes oculaires Valeurs COM Paramètres de l'égaliseur
Zd	Impédance différentielle des traces de PCB Structure des traces de PCB (profil)	
Type de carte PCB	Stripline est 1 et microruban est 0	

Tableau 1. Paramètres des caractéristiques du canal et méthodes quantitatives.

Variable de fonctionnalité Euh Permittivité de la feuille PCB pièce réelle et TanD Df Facteur de dissipation de la feuille PCB Variable de fonctionnalité Sortir Euh Df Tracer Zd Type de carte PCB

Impédance différentielle Type de PCB Stripline est 1 et microruban est 0 Ces caractéristiques de trace sont quantifiées en fonction de paramètres spécifiques tels que la trace longueur, largeur, épaisseur, conductivité, profil et espacement différentiel des lignes. Le profil est caractérisé par la forme géométrique de la trace, comme le nombre de coins. L'autre les caractéristiques sont directement représentées par leurs valeurs numériques correspondantes.

3.2. Configuration COM et réglage de l'égaliseur

Ce travail utilise le programme version COM 4.0 fourni par IEEE 802.3. La configuration fait référence à la configuration de simulation COM (eCOM) améliorée avec l'USB4 Gen4 standard et intègre les éléments de la norme 50 GBASE-KR. Les paramètres de configuration du PAM4, comme indiqué dans le tableau 2, utilisent une combinaison d'un émetteur FFE et CTLE et DFE côté récepteur. Cette configuration inclut les coefficients de prise du TX FFE et RX DFE, la plage adaptative de gain DC et les positions du pôle zéro du RX CTLE. Les paramètres supplémentaires incluent le nombre de niveaux de signal, noté L. Pour les signaux PAM4, L est réglé sur 4. Selon la configuration PAM4 recommandée basée sur USB4, le symbole Le débit est réglé sur 20 GBd et la sortie de tension de crête différentielle de l'émetteur est réglée sur 0,4 V, noté Av. La norme USB4 Gen4 suggère que PAM3 soit utilisé comme méthode de modulation. Les paramètres COM pour PAM3 sont à peu près les mêmes que ceux de PAM4, sauf que le débit de symboles est fixé à 25,6 God et que L est fixé à 3.

Tableau 2. Configuration des paramètres COM.

Paramètre	Symbole	Paramètre	Unité
Taux de symboles	fb	20	GBd
Nombre de niveaux de signal	L	4	
Échantillons par interface utilisateur	M.	32	
Taux d'erreur du détecteur cible	DER0	10−8	
Tension de sortie de l'émetteur, victime	Un V	0,4	V
Gain CC CTLE	gDC	[−12:1:0]	dB
Gain CC CTLE2	gDC_HP	[−6:1:0]	dB
Pôle CTLE HP	fHP_PZ	0,25	GHz
CTLE zéro	fZ	fb/2,5	GHz
Pôle CTLE1	fP1	fb/2,5 fb	GHz
Pôle CTLE2	fP2		
Curseur principal FFE	c (0)	0,62	
Précurseur FFE	c (−1)	[−0,18:0,02:0]	
Post-curseur FFE	c (1)	[−0,38:0,02:0]	
Longueur DFE	Nb	12	
Limite d'ampleur du DFE	bmax(1)	0,75	
	bmax(2 Nb)	0,2	
Seuil de réussite COM	ème	3	dB

Basé sur la configuration adoptée de TX FFE + RX CTLE + RX DFE, la prédiction les paramètres incluent les coefficients de prise du FFE, le gain DC du CTLE et les coefficients de prise du DFE. Les données TX FFE et RX CTLE sont obtenues en recherchant les valeurs FOM dans tout l'espace de conception, tandis que les coefficients DFE sont déterminés en fonction du réponse impulsionnelle h(t) correspondant au FOM optimal. Le SI est fortement impacté par les post-curseurs plus proches du curseur principal, tandis que les post-curseurs les plus éloignés exercent un minimum d'activité. influence, donc seuls les quatre premiers post-curseurs du DFE sont prédits.

3.3. Fractionnement de l'ensemble de données

Pour ce travail, un total de 325 chaînes ont été collectées. Dans cet ensemble de données, 280 canaux ont été utilisé pour créer un ensemble de données de diagramme de l'œil, appelé ensemble de données A, et 215 canaux ont été utilisés pour créer un ensemble de données pour les COM et prédire les paramètres des égaliseurs, référencés comme ensemble de données B. En raison de l'impact du FOM, l'approche COM utilise différents égaliseurs méthodes d'optimisation des paramètres pour les systèmes stripline et microruban, ce qui aboutit à

ensembles de données	:	: ROM
Ensemble d'entraînement	184	140
Ensemble de validation	40	40
Ensemble de test	56	40 (5 de l'ensemble de validation) 325
Total		

4. Construction du modèle en cascade et formation d'une
capacité d'ajustement affaiblie du modèle ML. Par conséquent, nous avons standardisé l'ensemble de données B
4.1. DNN
utilisant des canaux multilignes. Parmi les 215 canaux de l'ensemble de données B, 170 sont des canaux stripline.
l'ensemble de données A, tandis que 45 canaux sont des canaux stripline nouvellement créés. Selon
Le modèle MP, proposé par la psychologue McCulloch et le logicien Pitts, est un canal mathématique dans
la norme, les deux ensembles de données sont divisés en formation, validation et test.
Modèle mathématique qui a été développé grâce à une analyse et une synthèse de la méthodologie de base propre à
partie de l'ensemble de test de l'ensemble de données B est extraite de l'ensemble de validation. Les paramètres de partage
lignes de neurones [31]. Ce modèle est crucial pour la mise en œuvre de réseaux de neurones et de
constitue l'unité fondamentale d'un DNN. La structure de ce modèle est décrite en détail dans le tableau
Figure 6. Le =(, représente les n caractéristiques d'entrée du neurone, et
=(, ω ω ω ω) désigne le vecteur de poids responsable de la connexion pondérée linéaire.

Tableau 3. Paramètres de fractionnement de l'ensemble de données.

	Nombre de canaux	
	R : 280	B : 215 (170 de A)
Ensemble d'entraînement	184	140
Ensemble de validation	40	40
Ensemble de test	56	40 (5 de l'ensemble de validation)
Total	325	

vecteur de poids ω est optimisé au sein des réseaux de neurones pour améliorer la précision des prédictions. La sortie linéaire
pondérée Z peut être exprimée comme suit :

$$Z = \sum_{j=1}^{n} \omega_j x_j \quad (3)$$

La figure 6 représente la structure du modèle de neurone MP. La différence entre Z
et un décalage θ est utilisé comme entrée de la fonction f( ) . La fonction f( ) est appelée
fonction d'activation et utilisée pour dériver la sortie du neurone. θ peut être considéré
comme poids de l'entrée 0 x avec une valeur fixe de −1 et constitue également un objectif d'optimisation de 4,1. DNN
le réseau neuronal. Dans cet article, la fonction ReLU est adoptée comme fonction d'activation
car elle peut atténuer plus efficacement l'apparition d'un surapprentissage. La formule pour cela est présentée dans
Le modèle MP, proposé par la psychologue McCulloch et le logicien Pitts, est un modèle
mathématique qui a été développé à travers une analyse et une synthèse des propriétés de base
des neurones [31]. Ce modèle est crucial pour la mise en œuvre de réseaux de neurones et
DNN. La structure de ce modèle est représentée dans

$$fReLU(x) = \max(0, x) \quad (4)$$

constitue l'unité fondamentale d'un
Figure 6. Le x = (x1, x2, · · · représente les n caractéristiques d'entrée du neurone, et
ω i2 ω in) désigne le vecteur de poids responsable de la connexion pondérée linéaire. Les DNN peuvent effectuer des tâches
La sortie de la fonction d'activation est généralement utilisée comme entrée pour les couches adjacentes. ω i = (ω i1 ,
telles que la régression, la classification et la reconnaissance. Les DNN sont des solutions. Le vecteur de poids ω est optimisé
largement utilisé pour les tâches de régression en production et en évaluation des performances [4,6–
au sein des réseaux de neurones pour améliorer la précision de
9,32]. Les prédictions. La sortie linéaire pondérée Z peut être exprimée comme suit :
Le DNN utilisé dans cet article, comme le montre la figure 7, adopte une structure standard avec
plusieurs couches cachées, ainsi qu'une couche d'entrée et une couche de sortie de régression linéaire
utilise une architecture de réseau entièrement connectée entre ses neurones. [33], et (3)

$$z_i = \sum_{j=1}^{n} \omega_{ij} x_j$$

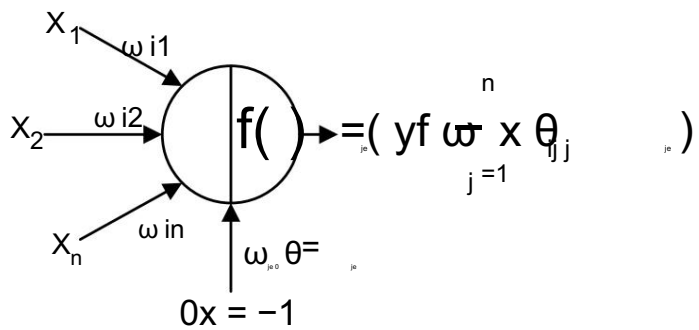


Figure 6. La structure du modèle de neurone MP.

La figure 6 représente la structure du modèle de neurone MP. La différence entre Z et
un décalage θ est utilisé comme entrée de la fonction f(·). La fonction f(·) est appelée la
fonction d'activation et utilisée pour dériver la sortie du neurone. θ peut être considéré comme le
poids de l'entrée x0 avec une valeur fixe de −1 et est également une cible d'optimisation du système neuronal
réseau. Dans cet article, la fonction ReLU est adoptée comme fonction d'activation car elle
peut atténuer plus efficacement l'apparition d'un surapprentissage. La formule pour cela est présentée dans
Équation (4).

$$fReLU(x) = \max(0, x) \quad (4)$$

La sortie de la fonction d'activation est généralement utilisée comme entrée pour les couches adjacentes.
Les DNN peuvent effectuer des tâches telles que la régression, la classification et la reconnaissance. Les DNN sont
largement utilisé pour les tâches de régression en production et en évaluation des performances [4,6–9,32]. Le

Le DNN utilisé dans cet article, comme le montre la figure 7, adopte une structure standard avec plusieurs des couches cachées, ainsi qu'une couche d'entrée et une couche de sortie de régression linéaire [33], et utilise une architecture de réseau entièrement connectée entre ses neurones.

Figure 6. La structure du modèle de neurone MP.

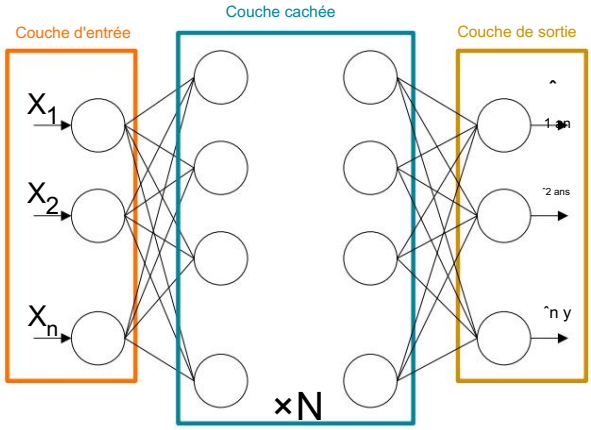


Figure 7. La structure du DNN de cette étude.

Afin de mieux faire converger le modèle, le modèle de formation nécessite généralement une fonction de coût pour évaluer l'écart entre les valeurs prédites par le modèle et les valeurs réelles. Dans ce panier, Smooth L1 [34] est adopté comme fonction de coût :

$$\text{lisse}_{L1} = \begin{cases} 0{,}5\,(x_i - y_i)^2, & \text{si } |x_i - y_i| < 1 \\ |x_i - y_i| - 0{,}5, & \text{sinon} \end{cases}, \quad (5)$$

où y_n et $\hat{y}_n$ représentent respectivement la valeur réelle et la valeur prédite.

La performance du modèle est évaluée à l'aide de la fonction de coût, ensuite, un optimiseur est utilisé pour rétropropager et minimiser continuellement la fonction de coût. Les paramètres et les pondérations du modèle sont continuellement mis à jour jusqu'à ce que les performances attendues soient atteintes. Après avoir comparé plusieurs optimiseurs différents, l'optimiseur Adam a été ultimately sélectionné pour être utilisé dans cet article [35].

4.2. Transformateur pour la régression

La précision des paramètres d'égalisation et des valeurs COM directement prédites par les paramètres de caractéristiques du canal est relativement faible. Dans ce travail, la réponse à un bit (SBR) avant et après l'égalisation est d'abord prédite en fonction des paramètres caractéristiques du canal, et puis les paramètres de l'égaliseur et les valeurs COM sont déterminés en fonction de la réponse pré-réponse pulsée dictée. La réponse impulsionnelle doit être utilisée comme série chronologique pour la prédiction. Traditionnellement, les modèles RNN ou LSTM sont utilisés pour la prévision de séries chronologiques, mais ils présentent certaines limites. Le modèle Transformer peut améliorer efficacement le problème de gradi-explosion qui se produit dans LSTM en utilisant des mécanismes d'attention, améliorant ainsi la précision de prédiction [36].

La formule d'un mécanisme d'attention est présentée dans l'équation (6). Il se compose principalement de trois entrées : Q (Requête), K (Clé) et V (Valeur), ainsi qu'un module softmax. Lorsque Q et V sont identiques, on parle de mécanisme d'auto-attention.

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (6)$$

L'architecture complète de Transformer comprend principalement une couche d'intégration, une couche de codage positionnel, une couche de codeur, une couche de décodeur et une couche de sortie. Dans ce travail, puisque notre tâche est une tâche de prédiction de séries chronologiques plutôt qu'une tâche de PNL, la couche de codage positionnel a été supprimée. L'essence de la prédiction de séries chronologiques peut être considérée comme une tâche de régression, le modèle n'a besoin que de générer une seule valeur de sortie correspondant à chaque instant et n'a pas besoin d'un décodeur pour générer des sorties de séquence. Utiliser uniquement

à chaque instant et n'a pas besoin d'un décodeur pour générer des sorties de séquence. En utilisant
la couche d'encodeur est suffisante, ce qui simplifie la structure du modèle et réduit la com- la couche d'encodeur
ce qui simplifie la structure du modèle et réduit la com- seule la couche d'encodeur est suffisante,
charge putationnelle. Comme le montre la figure 8, notre conception finale conserve la couche d'intégration, la
charge de calcul. Comme le montre la figure 8, notre conception finale conserve l'intégration
couche d'encodeur et la couche de sortie finale. De plus, afin d'obtenir une conversion encodeur plus rapide et
finale. De plus, afin d'obtenir une conversion plus rapide, la couche d'encodeur et la couche de sortie
plus rapidement.
C'est pourquoi l'unité linéaire d'erreur gaussienne (GELU) a été choisie comme fonction d'activation. C'est pourquoi
l'unité linéaire d'erreur gaussienne (GELU) a été choisie comme fonction d'activation. convergence, l'unité linéaire

$$f(x) = x\,P(X \le x) = x\,\Phi(x) = x \cdot \frac{1}{2}\left[1 + \operatorname{erf}(x/\sqrt{2})\right] \quad (7)$$

$$f(x) = x\,P(X \le x) = x\,\Phi(x) = x \cdot \frac{1}{2}\left[1 + \operatorname{erf}(x/\sqrt{2})\right] \quad (7)$$

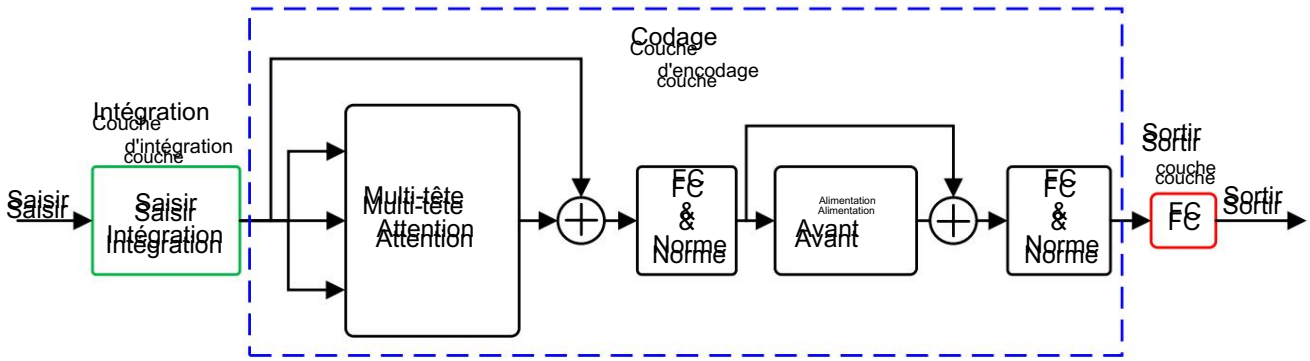


Figure 8. La structure de notre Transformer pour les tâches de régression.
Figure 8. La structure de notre Transformer pour les tâches de régression.
Figure 8. La structure de notre Transformer pour les tâches de régression.

4.3. Modèle en cascade et formation
4.3. Modèle en cascade et formation
4.3. Modèle en cascade et formation

Comme l'illustre la figure 9, un modèle de transformateur et de DNN en cascade est utilisé dans ce cas.
Comme l'illustre la figure 9, un modèle de transformateur et de DNN en cascade est utilisé dans ce cas.
travailler pour prédire avec précision la valeur COM et ses coefficients d'égalisation correspondants. Un travail
avec précision la valeur COM et ses coefficients d'égalisation correspondants. Un travail pour prédire
un total de 325 entités de canal ont été collectées sous forme d'ensemble de données et un modèle DNN est
est utilisé. Un total de 325 entités de canal ont été collectées sous forme d'ensemble de données et un modèle DNN est
pour prédire la caractéristique spectrale de chaque canal, en particulier l'IL et le RL au Nyquist utilisé pour prédire la
Spectres de fréquence du canal, en particulier l'IL et le RL au Nyquist utilisé pour prédire la caractéristique
points de fréquence. Ces données spectrales prédites, ainsi que les caractéristiques du canal, sont alors des points de fréquence
Nyquist. Ces données spectrales prédites, ainsi que les caractéristiques du canal, sont alors des points de fréquence de
introduit dans un modèle de transformateur pour dériver le SBR de pré-égalisation. Utilisation de la pré-égalisation dans
un modèle de transformateur pour dériver le SBR de pré-égalisation. En égalisation SBR de présentation
configuration COM. Sous la configuration COM sont SBR d'égalisation, les paramètres d'égaliseur correspondants sous la
paramètres d'égalisation correspondants sous la configuration COM sont SBR de isation, les paramètres d'égalisation
calculé par le modèle DNN. Afin de calculer le SBR post-égalisation, égaliseur calculé par le modèle DNN. Afin de calculer
le SBR post-égalisation, l'égaliseur est calculé par le modèle DNN. Afin de calculer le SBR post-égalisation, l'égaliseur
paramètres combinés avec les caractéristiques du canal et les caractéristiques du domaine fréquentiel sont ainsi
les envoyé dans un autre modèle de Transformer. Enfin, le canal SBR post-égalisation échantillonné, envoyé dans un autre
modèle de SBR post-égalisation, et canal SBR, post-égalisation échantillonné, envoyé dans un autre modèle de Transformer.
les caractéristiques et les caractéristiques spectrales sont utilisées pour calculer la valeur COM correspondante finale dans
Les caractéristiques et les caractéristiques spectrales sont utilisées pour calculer la valeur COM finale
correspondante dans
le modèle DNN. le modèle DNN.

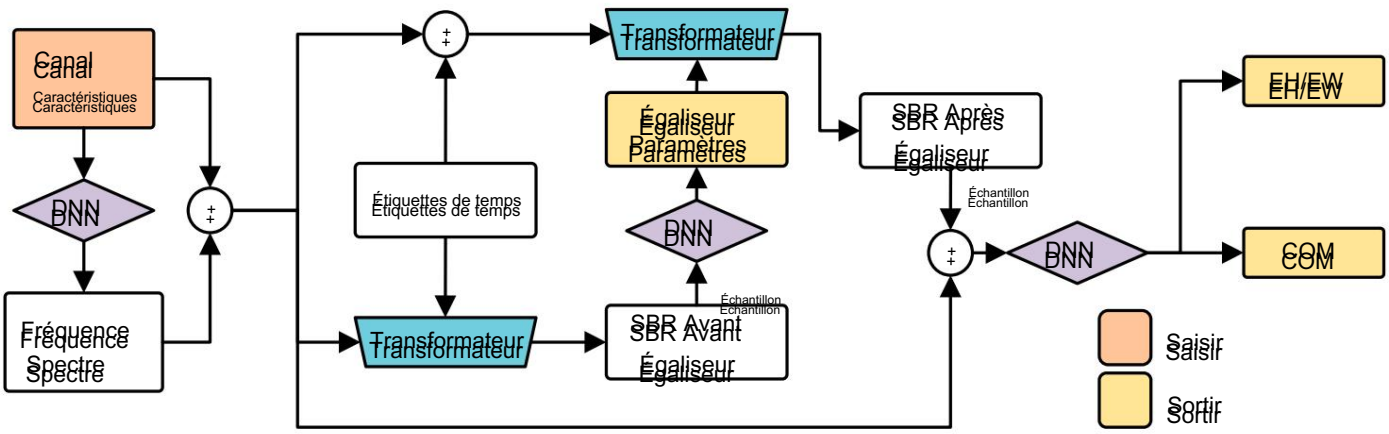


Figure 9. La structure du modèle en cascade.
Figure 9. La structure du modèle en cascade.
Figure 9. La structure du modèle en cascade.

Le vecteur d'entrée des caractéristiques physiques de chaque canal peut être défini comme suit :
Le vecteur d'entrée des caractéristiques physiques de chaque canal peut être défini comme suit : bas :

$$X = x_{ij}\ j=1:NC, j=1:NP, \quad (8)$$

où NC représente le nombre de canaux dans l'ensemble de données, NP est le nombre total d' entités de canal et X est une matrice NC × NP . L'élément xij désigne la valeur quantifiée d'une caractéristique physique spécifique d'un canal : par exemple, x11 correspond à la valeur de la caractéristique de longueur du premier canal.

Le DNN utilisé dans le modèle en cascade peut être représenté par l'équation (9) :

$$\begin{aligned} Y &= FDNN(Z, \theta) \\ &= W(l)\, fReLU(\cdots(W(2)\, fReLU(W(1)X + \theta^{(1)}) + \theta^{(2)})\cdots) + \theta^{(l)} \end{aligned}, \quad (9)$$

où Z représente la sortie pondérée linéairement de la matrice d'entrée X et les poids du modèle W, comme décrit dans l'équation (3), et θ représente le décalage. L'index l désigne la couche du modèle DNN, qui comprend la couche d'entrée, les couches cachées et la couche de sortie. Notre réseau utilise un total de quatre modèles DNN différents avec des hyperparamètres variables : Y_f pour prédire les caractéristiques spectrales, Y_e pour prédire les paramètres de l'égaliseur, Y_F pour prédire le diagramme de l'œil et Y_C pour prédire le COM.

La matrice augmentée Xf est obtenue en concaténant horizontalement la matrice d'entrée X du premier modèle DNN avec sa sortie Y_f :

$$Xf = [X Y_f] = \begin{bmatrix} x11 & x12 & \cdots & x1NP & y1I & y1R \\ x21 & x22 & \cdots & x2NP & y2I & y2R \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ xNC1 & xNC2 & \cdots & xNCNP & yNC\,I & yNCR \end{bmatrix}, \quad (dix)$$

où le vecteur colonne $yI = (y1I, y2I, \cdots, yNc\,I)^T$ représente le canal prédit IL et $yR = (y1R, y2R, \cdots, yNcR)^T$ représente le canal prédit RL.

Pour améliorer la précision de la valeur COM prévue, la réponse impulsionnelle est extraite du SBR pré- et post-égalisation dans l'outil COM. Le point temporel correspondant au pic est pris comme curseur principal, et les amplitudes de quatre précurseurs et de huit post- curseurs déterminées par l'intervalle d'échantillonnage COM sont sélectionnées comme SBR requis. Ensuite, l'impulsion de chaque canal est aplatie et les étiquettes temporelles correspondantes sont insérées, formant un vecteur de caractéristiques séquentielles. L'équation (11) représente la séquence de valeurs d'amplitude. Après ce traitement, le modèle Transformer peut être utilisé pour prédire avec précision le SBR.

$$\begin{aligned} SBR1(t) &= FTrans[x11(t), x12(t), \cdots, x1NP(t), y1I(t), y1R(t)] \\ SBR1 &= [SBR1(t - t0), \cdots, SBR1(t), \cdots, SBR1(t + t1)] \end{aligned}, \quad (11)$$

où FTrans désigne le modèle de transformateur utilisé, SBR1(t) représente le SBR correspondant au vecteur de caractéristiques d'entrée du canal à un instant spécifique, et SBR1 représente la séquence SBR du canal. t0 et t1 sont respectivement les limites gauche et droite de la plage horaire.

En introduisant le SBR dans le modèle DNN suivant avec des paramètres distincts, la prédiction des paramètres de l'égaliseur peut être réalisée via l'équation (12) :

$$Sortie = FDNN(SBR1, SBR2, \cdots, SBRNC) \quad (12)$$

En intégrant les paramètres de l'égaliseur avec l'entrée présentant les caractéristiques tics, une matrice augmentée $SBReq_{NC}$ spectrales, (t) est construit pour la prédiction de la post-égalisation

SBR. Cette matrice, via l'équation (13), est ensuite traitée en une séquence dépendante du temps suivant la méthodologie présentée dans l'équation (11).

$$\begin{aligned} SBReq_{NC}(t) &= FTrans[xNC1(t), \cdots, xNCNP(t), \\ &\quad yNCI(t), yNCR(t), Y^{e}_{NC}(t)] \\ SBReq_{NC} &= [SBReq_{NC}(t-t0), \cdots, \\ &\quad SBReq_{NC}(t), \cdots (t+tf)]_{NC} \\ {}^{Oui}{}^{e}_{NC} Robinets(t) &= [y_{Robinets}(t), y(t), y_{Gain\,DC}(t)] \\ FFE\text{-}NC &\quad CTLE\text{-}NC \quad DFE\text{-}NC \end{aligned} \tag{13}$$

Enfin, le SBR post-égalisation obtenu est concaténé avec X f pour former le résultat final. fonctionnalités d'entrée pour prédire la valeur COM.

Avant de former formellement le modèle en cascade, nous avons standardisé les données d'entrée et f req a utilisé une stratégie de pré-formation. Initialement, les modèles DNN F DNN(X) et F DNN(X) conçus respectivement pour prédire les caractéristiques spectrales et les paramètres de l'égaliseur, ont été formés indépendamment et leurs paramètres correspondants ont été enregistrés. Ces modèles DNN pré-entraînés ont ensuite été intégrés au modèle Transformer nécessaire pour construire le modèle en cascade complet pour prédire les paramètres de l'égaliseur et COM, comme illustré dans la figure 9. Pour la formation du modèle en cascade, Smooth L1 a été utilisé comme fonction de coût. et l'algorithme d'Adam a été appliqué pour mettre à jour les paramètres du modèle. À l'aide de la boîte à outils optuna, la structuration du modèle et l'optimisation des hyperparamètres (HPO) ont été réalisées via l'optimisation bayésienne à l'aide du GP [37,38]. La fonction d'acquisition choisie était la fonction d'amélioration attendue (EI), avec un ensemble initial de 10 points d'observation, permettant une HPO efficace dans une portée informatiquement réalisable.

Par rapport aux méthodes de recherche de grille globale, de recherche de grille aléatoire et de recherche de moitié, l'optimisation bayésienne est plus efficace pour HPO et nécessite moins de temps d'optimisation. De plus , la méthode de validation croisée K-Fold est utilisée dans cet article pour améliorer la fiabilité de l'évaluation du modèle.

Une fois la formation du modèle terminée, ses capacités prédictives peuvent être évaluées à l'aide de plusieurs métriques différentes. L'une de ces mesures, présentée dans l'équation (14), est le RMSE. L'opération de mise au carré du RMSE augmente sa sensibilité aux erreurs, le rendant particulièrement réactif aux valeurs aberrantes. Cette sensibilité accrue peut également le rendre trop réactif aux valeurs aberrantes, mais l'opération de racine carrée ne peut pas refléter intuitivement l'ampleur réelle de l'erreur.

$$RMSE = \sqrt{\frac{1}{n}\sum_{je=1}^{n}(yn - \hat{y}n)^2} \tag{14}$$

Le MAE est calculé en fonction de la valeur moyenne des différences absolues entre les valeurs prédites et réelles. Comparé au RMSE, le MAE est moins sensible aux valeurs aberrantes et fournit un reflet plus intuitif de l'ampleur réelle des erreurs. Cependant, cela peut conduire à une sous-estimation de l'impact des erreurs.

$$MAE = \frac{1}{n}\sum_{je=1}^{n}|yn - \hat{y}n| \tag{15}$$

MAPE est la valeur moyenne des différences en pourcentage absolu entre les valeurs prédites et réelles. Comparé au RMSE et au MAE, MAPE est plus facile à comprendre et permet de comparer les résultats de prédiction à différentes échelles. Il est plus adapté aux situations présentant des changements importants dans les valeurs réelles. Cependant, lorsque les valeurs réelles sont proches de zéro, l'erreur peut devenir importante. MAPE ne convient donc pas aux ensembles de données contenant des valeurs nulles ou proches de zéro.

$$MAPE = \frac{1}{n}\sum_{je=1}^{n}\frac{yn - \hat{y}n}{yn} \tag{16}$$

valeurs nulles ou proches de zéro.

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{a_i - \hat{a}_i}{a_i}\right| \quad (16)$$

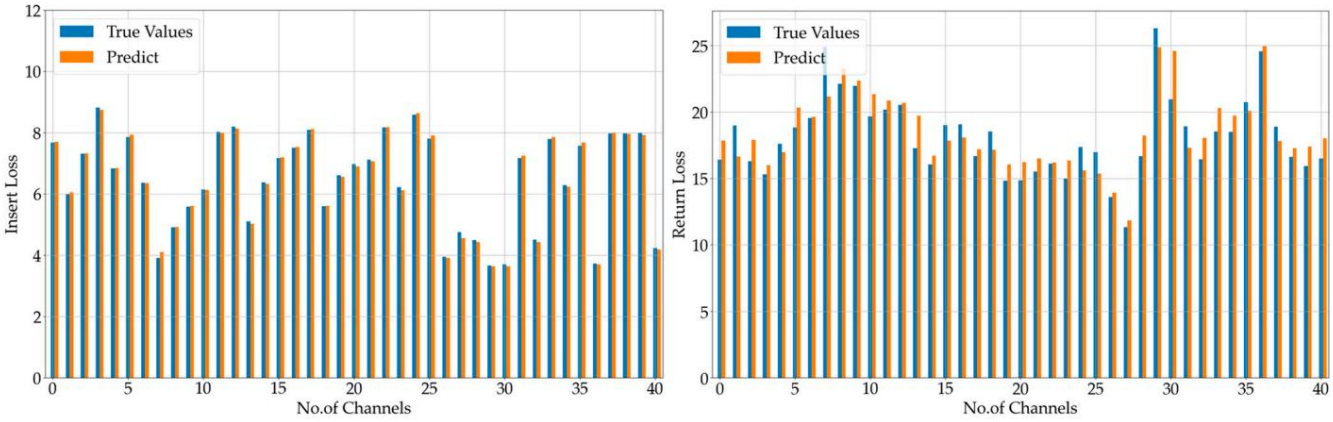
5. Résultats numériques

Dans ce travail, le modèle de réseau neuronal a été construit à l'aide d'une bibliothèque Python basé sur Pytorch 2.0.1. L'ordinateur utilisé pour le calcul du modèle comprenait une GeForce RTX 4070 et un Intel Core i5-12600 KF 3,7 GHz.

5.1. IL et RL

Comme le montre la figure 10, l'IL et le RL sont des pertes à la fréquence de Nyquist et celles du modèle les prédictions de IL sont plus précises que ses prédictions de RL. En effet, IL est primaire. Comme le montre la figure 10, IL et RL sont des pertes à la fréquence de Nyquist, et le modèle fortement influencé par la perte diélectrique et les caractéristiques d'entrée, y compris les prédictions de IL liées à la perte diélectrique, sont plus précises que ses prédictions de RL. C'est parce que l'IL est avant tout paramètres. À l'inverse, RL est affecté par des facteurs supplémentaires tels que l'impédance mal influencée par la perte diélectrique et les caractéristiques d'entrée, notamment les paramètres liés à la perte diélectrique et les formes de traces complexes, compliquant sa prédiction. Par conséquent, la précision des eters. À l'inverse, RL est affecté par des facteurs supplémentaires tels que les inadéquations d'impédance et La prédiction IL est meilleure que celle de la prédiction RL. L'erreur relative de la prédiction RL réside dans les formes de traces complexes, ce qui complique sa prédiction. Par conséquent, la précision de la prédiction de l'IL 17,4 %, mais cela est suffisant pour prendre en charge la prédiction précise des diagrammes oculaires et COM est meilleure que celle de la prédiction RL. L'erreur relative de prédiction RL est de 17,4 %, mais c'est valeurs.

suffisant pour prendre en charge la prédiction précise des diagrammes oculaires et des valeurs COM.



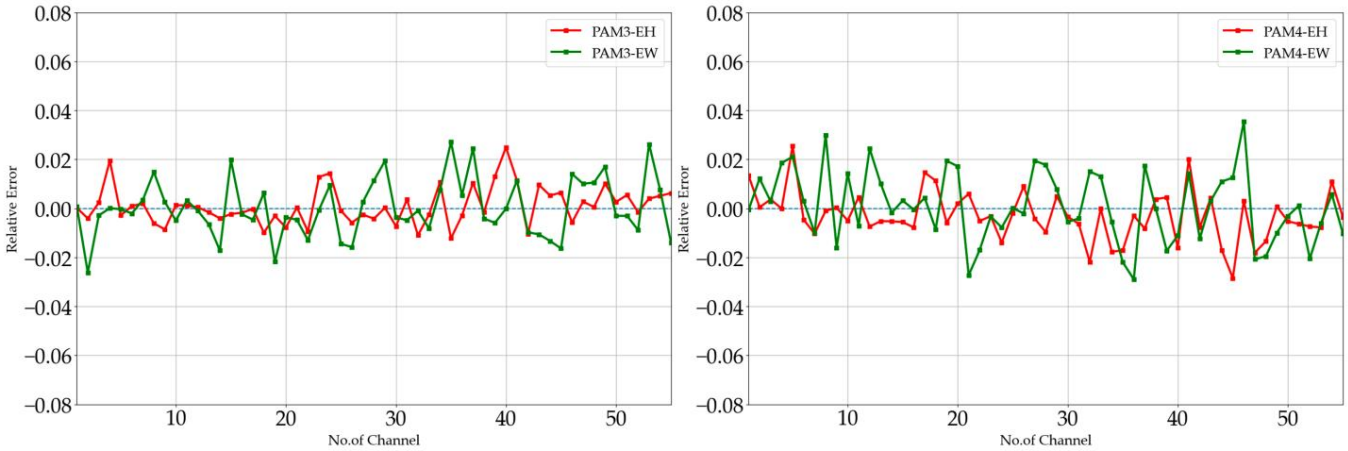
(un) (b)

Figure 10. Comparaison des valeurs prévues et réelles pour (a) IL et (b) RL.

5.2. EH/EW des diagrammes oculaires

La figure 11 montre les erreurs relatives de hauteur et de largeur des yeux pour 40 canaux de test avec les méthodes de modulation PAM3 et PAM4. Les canaux n° 1 à n° 32 sont des canaux à bande tandis que les canaux n° 33 à n° 56 sont des canaux microruban. Canaux n°1 à n°10 présentent des chemins de trace plus complexes, ce qui entraîne des erreurs plus importantes en termes de hauteur et de largeur des yeux. En revanche, les canaux n° 11 à n° 23, qui ont un routage plus simple avec des canaux plus droits ou moins nombreux, présentent une meilleure précision d'ajustement. La plupart des canaux de l'ensemble de données utilisent FR-4 comme couche diélectrique, tandis que les canaux n° 24 à n° 32 utilisent des matériaux alternatifs, ce qui entraîne des erreurs relativement plus importantes. Par rapport aux chaînes stripline, les chaînes n°33 à n°56 qui utilisent des lignes microruban, présentent des erreurs relatives significativement plus élevées, démontrant que les lignes microruban ont des capacités anti-interférences plus faibles que les lignes ruban. Le tableau 4 montre le RMSE, le MAE, le MAPE et le MRE pour la prédiction du diagramme oculaire avec les méthodes de modulation PAM3. Il est évident que les résultats de prédiction présentent une bonne convergence et précision. Étant donné que l'unité EH est choisie en mV, les métriques RMSE et MAE augmentent d'un ordre de grandeur par rapport à EW. Les résultats de ce tableau, combinés à ceux dans la figure 11, démontrent que les modèles PAM3 et PAM4 sont efficaces et présentent haute précision.

Œil	RMSE de modulation		MAE	MAPE	MRE (%)
EH (mV)	PAM3	4,8 × 10−1	3,3 × 10−1	5,7 × 10−3	2.5
	PAM4	4,0 × 10−1	2,9 × 10−1	8,1 × 10−3	2.8
EW (UI)	PAM3	4,4 × 10−3	3,4 × 10−3	9,3 × 10−3	2,7
	PAM4	4,4 × 10−3	3,6 × 10−3	1,2 × 10−2	3,5



(a) PAM3 (b) PAM4

Figure 11. Erreurs relatives des diagrammes oculaires pour les canaux de l'ensemble de test dans les deux formats de modulation.
Figure 11. Erreurs relatives des diagrammes oculaires pour les canaux de l'ensemble de test dans les deux formats de modulation.

5.3. COM

Tableau 4. Prédiction des diagrammes oculaires avec le modèle DNN.

Œil	RMSE de modulation		MAE	MAPE	MRE (%)
EH (mV)	PAM3	4,8 × 10−1	3,3 × 10−1	5,7 × 10−3	2.5
	PAM4	4,0 × 10−1	2,9 × 10−1	8,1 × 10−3	2.8
EW (UI)	PAM3	4,4 × 10−3	3,4 × 10−3	9,3 × 10−3	2.7
	PAM4	4,4 × 10−3	3,6 × 10−3	1,2 × 10−2	3.5

Étant donné que de nombreuses valeurs SBR avant et après le curseur sont proches ou égales à zéro, Modulation Eye MAPE et MRE ne conviennent pas pour évaluer la précision des prédictions SBR, et seuls RMSE et MAE sont utilisés pour évaluer la précision des prédictions du SBR. Pour le SBR 7,3 × 10−3 avant égalisation des 40 canaux de test, le RMSE est de 4,8 × 10−4 et le MAE est de 3,3 × 10−4. Pour le SBR après égalisation, le RMSE est de 8,1 × 10−3 et le MAE est de 1,6 × 10−4. Ci-dessous, nous utilisons un canal sélectionné au hasard (ID : 3) dans l'ensemble de test pour illustrer la prédiction SBR performances PAM4 . Comme le montre la figure 12, le SBR avant et après égalisation peut être prédit par le modèle Transformer avec une grande précision.

5.3. COM

Étant donné que de nombreuses valeurs SBR avant et après le curseur sont proches ou égales à zéro, MAPE et MRE ne conviennent pas pour évaluer l'exactitude des prédictions SBR, et seul RMSE et MAE sont utilisés pour évaluer la précision de la prédiction du SBR. Pour le SBR avant égalisation des 40 voies de test, le RMSE est de 7,3 × 10−4 et le MAE est de 6,3 × 10−4 . Pour le SBR après égalisation, le RMSE est de 1,1 × 10−3 et le MAE est de 1,6 × 10−4 . Ci-dessous, nous utiliser un canal sélectionné au hasard (ID : 3) dans l'ensemble de test pour illustrer la prédiction SBR performance. Comme le montre la figure 12, les phases de pré-égalisation et de post-égalisation

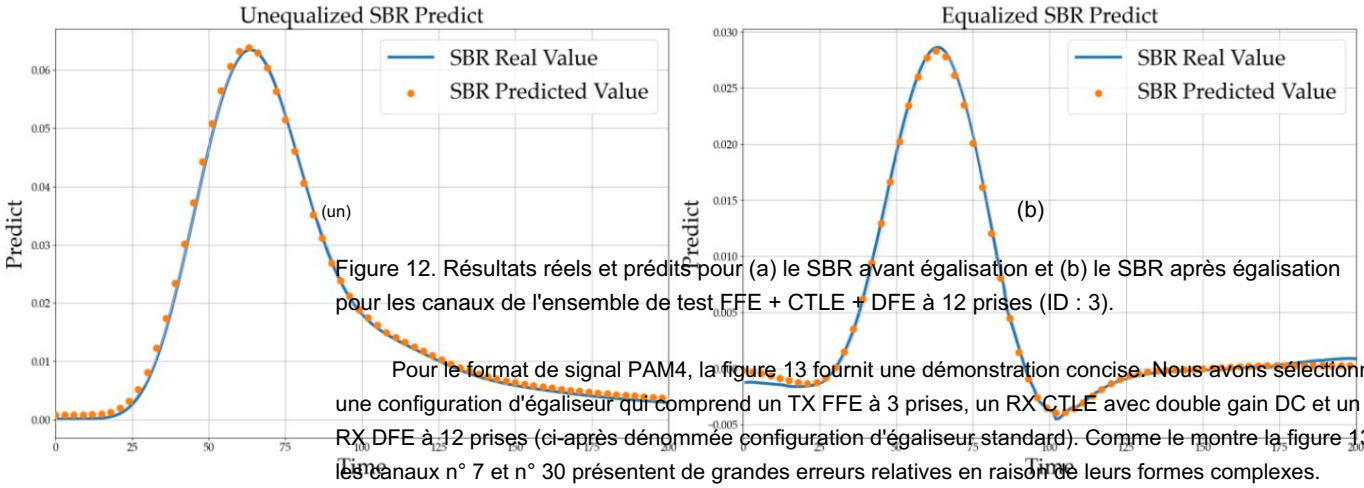
Le SBR peut être prédit par le modèle Transformer avec une grande précision.

Pour le format de signal PAM4, la figure 13 fournit une démonstration concise. Nous avons sélectionné une configuration d'égaliseur qui comprend un TX FFE à 3 prises, un RX CTLE avec double gain DC, et un RX DFE à 12 prises (ci-après dénommé la configuration d'égalisation standard). Comme Comme le montre la figure 13, les canaux n°7 et n°30 présentent des erreurs relatives importantes en raison de leur formes complexes. Certaines chaînes avec un RL élevé, telles que les chaînes n° 7 à n° 7. 9, expose également erreur relative importante. L'erreur relative des valeurs COM prédites pour chaque test

Le canal est inférieur à 2 %, démontrant la capacité du modèle DNN-Transformer à atteindre la précision de prédiction requise.

Ici, les influences des différentes combinaisons d'égaliseurs sur la précision de la prédiction des valeurs COM sont analysées en profondeur. Comme le montre le tableau 5, nous avons configuré cinq combinaisons d'égalisation. Les résultats indiquent que notre modèle en cascade proposé atteint les performances de prédiction attendues sur diverses combinaisons d'égaliseurs. Cependant, il est Il est évident qu'une réduction de la variété des égaliseurs affaiblit la généralisation du modèle capacité et diminue sa précision de prédiction. Lorsque seul le DFE était activé, le RMSE augmenté à 1,6 × 10−1 , et le MRE a atteint 17,5%. Nous attribuons cela à l'optimisation méthode de cet égaliseur dans le cadre COM et le fait que le modèle en cascade prend également en compte les prévisions pour le CTLE et le FFE. Nous avons donc ajusté la structure du réseau en supprimant les égaliseurs inutilisés dans le modèle en cascade pour différents

combinaisons d'égalisation. Cette modification a entraîné une amélioration de la fonctionnalité du modèle. précision de prédiction pour les valeurs COM, le MRE diminuant de 17,5 % à 11,5 %.



(un) (b)

Figure 12. Résultats réels et prédits pour (a) le SBR avant égalisation et (b) le SBR après égalisation pour les canaux de l'ensemble de test FFE + CTLE + DFE à 12 prises (ID : 3).

Pour le format de signal PAM4, la figure 13 fournit une démonstration concise. Nous avons sélectionné une configuration d'égaliseur qui comprend un TX FFE à 3 prises, un RX CTLE avec double gain DC et un RX DFE à 12 prises (ci-après dénommée configuration d'égaliseur standard). Comme le montre la figure 13, les canaux n° 7 et n° 30 présentent de grandes erreurs relatives en raison de leurs formes complexes. Certaines chaînes avec un RL élevé, telles que les chaînes n° 7 à n° 7. 9, présentent également une erreur relative significative. L'erreur relative des valeurs COM prévues pour chaque canal de test est inférieure à 2 %, démontrant la capacité du modèle DNN-Transformer à atteindre la précision de prédiction requise.

(b)

Figure 12. Résultats réels et prédits pour (a) le SBR avant égalisation et (b) le SBR après égalisation pour les canaux de l'ensemble de test FFE + CTLE + DFE à 12 prises (ID : 3).

Figure 12. Résultats réels et prédits pour (a) le SBR avant égalisation et (b) le SBR après égalisation pour les canaux de l'ensemble de test FFE + CTLE + DFE à 3 prises (ID : 3).

Pour le format de signal PAM4, la figure 13 fournit une démonstration concise. Nous avons sélectionné une configuration d'égaliseur qui comprend un TX FFE à 3 prises, un RX CTLE avec double gain DC et un RX DFE à 12 prises (ci-après dénommée configuration d'égaliseur standard). Comme le montre la figure 13, les canaux n° 7 et n° 30 présentent de grandes erreurs relatives en raison de leurs formes complexes. Certaines chaînes avec un RL élevé, telles que les chaînes n° 7 à n° 7. 9, présentent également une erreur relative significative. L'erreur relative des valeurs COM prévues pour chaque canal de test est inférieure à 2 %, démontrant la capacité du modèle DNN-Transformer à atteindre la précision de prédiction requise.

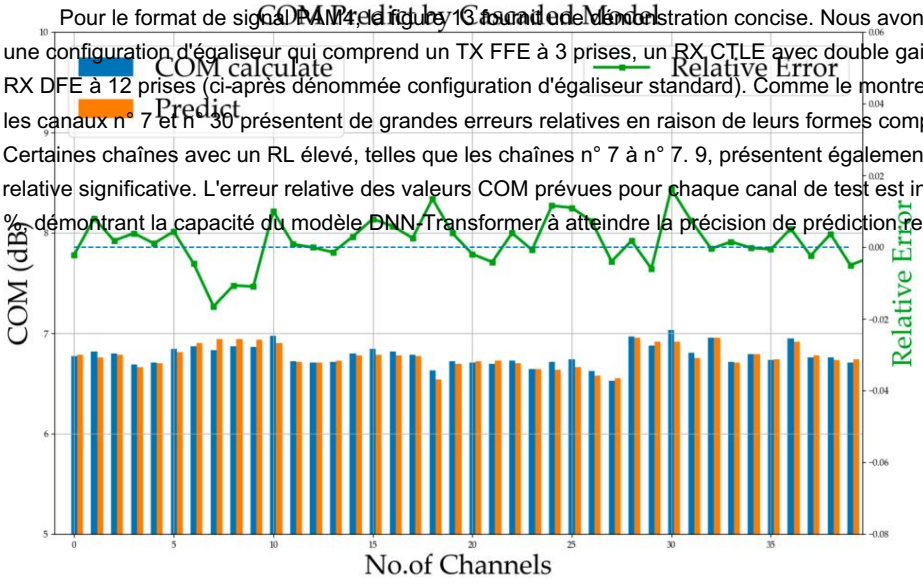


Figure 13. Erreur relative et valeurs COM pour le codage PAM4 (ID : 1–40) dans l'égaliseur standard configuration.

Figure 13. Erreur relative et valeurs COM pour le codage PAM4 (ID : 1–40) dans la configuration d'égaliseur standard.

Ici, les influences de différentes combinaisons d'égaliseurs sur la prédiction précise. Tableau 5. Valeurs d'erreur COM prévues par les modèles de différentes combinaisons d'égaliseurs sur la prédiction précise. Tableau 5. Valeurs d'erreur COM sion des valeurs COM sont analysées en profondeur. Comme le montre le tableau 5, nous avons configuré cinq combinaisons d'égalisation. Les résultats indiquent que notre modèle en cascade proposé atteint 4,5 × 10−2 3,5 × 10−2 5,1 × sur diverses combinaisons d'égaliseurs. Cependant, il apparaît qu'une réduction de la variété des égaliseurs affaiblit le modèle 10−3 les performances de prédiction attendues 1,6%

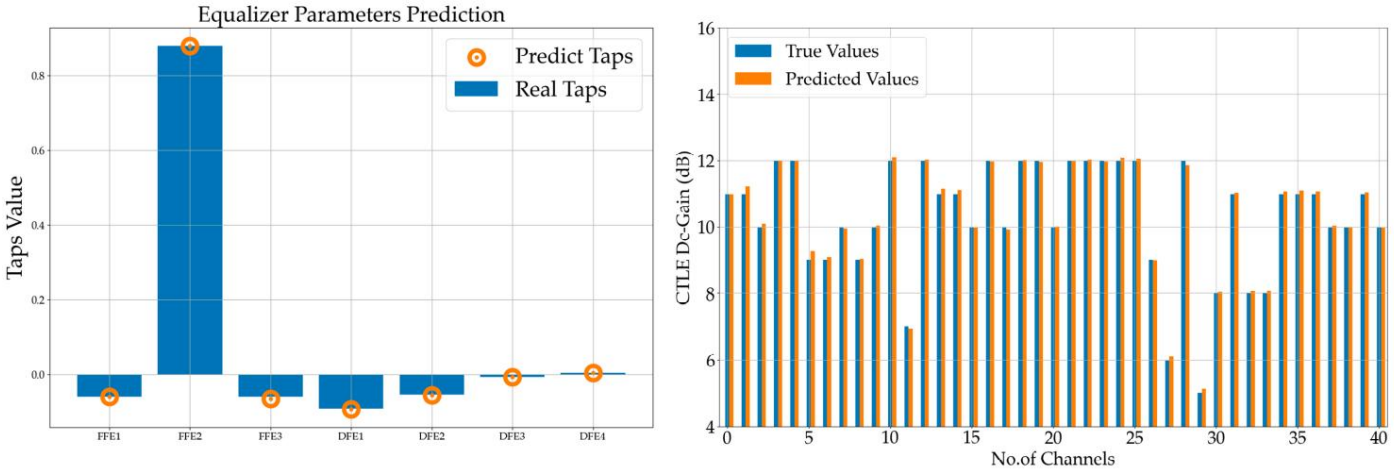
5,0 × 10−2 3,7 × 10−2 5,6 × 10−3

COM	RMSE	MAE	MAPE MRE (%) différents
FFE à 3 prises + CTLE + DFE à 12 prises			
FFE à 3 prises + CTLE + DFE à 8 prises			2,0%
+ DFE à 4 prises	4,2 × 10−2 3,2 × 10−2 4,1 × 10−3 FFE à 3 prises + CTLE		1,7%
CTLE + DFE à 12 prises .	9,6 × 10−2 7,1 × 10−2 1,4 × 10−2 Configuration		5,7%
DFE à 12 prises	1,6 × 10−1 1,1 × 10−1 3,0 × 10−2		17,5%

Ici, les influences de différentes combinaisons d'égaliseurs sur la précision de prédiction des valeurs COM sont analysées en profondeur. Comme le montre le tableau 5, nous avons configuré cinq combinaisons d'égaliseur différentes. Les résultats indiquent que notre modèle en cascade proposé atteint les performances de prédiction attendues sur diverses combinaisons d'égaliseurs. Cependant, il apparaît qu'une réduction de la variété des égaliseurs affaiblit la performance du modèle.

3 prises FFE + CTLE + 12 prises DFE	4,5 × 10−2	3,5 × 10−2	5,1 × 10−3	1,6%
FFE + CTLE + 8 prises DFE	5,0 × 10−2	3,7 × 10−2	5,6 × 10−3	2,0%
3 prises FFE + CTLE + 4 prises DFE	4,2 × 10−2	3,2 × 10−2	4,1 × 10−3	1,7%
CTLE + DFE à 12 prises	9,6 × 10−2	7,1 × 10−2	1,4 × 10−2	5,7%
DFE à 12 prises	1,6 × 10−1	1,1 × 10−1	3,0 × 10−2	17,5%

Ci-dessous, un canal de test (ID : 3) est utilisé pour illustrer la capacité de prédiction du modèle
canal de test (ID : 3) est utilisé pour illustrer la capacité de prédiction du modèle
avec les quatre premiers taps du DFE et les trois premiers taps du FFE. Comme le montre la figure 14a, le
avec les quatre premiers taps du DFE et les trois premiers taps du FFE. Comme le montre la figure 14a, le
modèle en cascade réalise des prédictions très précises. De plus, les résultats de la prédiction
le modèle en cascade réalise des prédictions très précises. De plus, les résultats de la prédiction
pour le gain DC principal du CTLE pour les 40 canaux de test, comme présenté sur la figure 14b, également
pour le gain DC principal du CTLE pour les 40 canaux de test, comme présenté sur la figure 14b, également
indiquent un haut niveau de précision.
indiquent un haut niveau de précision.



(a) Prises FFE et DFE (b) Gain CTLE DC

Figure 14. Comparaison des résultats des paramètres d'égalisation prévus et réels.
Figure 14. Comparaison des résultats des paramètres d'égalisation prévus et réels.

Les capacités prédictives de différentes combinaisons de modèles en cascade ont été étudiées. Les capacités
prédictives de différentes combinaisons de modèles en cascade ont été étudiées.
formé comme suit. La configuration de l'égaliseur a été définie comme configuration standard, et comme suit. La
configuration de l'égaliseur a été définie comme configuration standard et le
les paramètres HPO ont été configurés uniformément avec la plage de recherche du taux d'apprentissage définie
paramètres HPO ont été configurés uniformément, avec la plage de recherche du taux d'apprentissage définie à partir de
de 0,0001 à 0,01, en effectuant 20 essais d'optimisation. Compte tenu des caractéristiques du SBR,
de 0,0001 à 0,01, en effectuant 20 essais d'optimisation. Compte tenu des caractéristiques du SBR,
nous avons testé les performances prédictives de modèles autonomes, dont un DNN, un LSTM,
performances prédictives de modèles autonomes, dont un DNN, un LSTM
Le DNN combine deux modèles de prédiction de séquence,
Le DNN combine à deux modèles de prédiction de séquence,
un transformateur. Ils ont été sélectionnés comme système d'un réseau de prédiction de séquence et à un transformateur.
un transformateur. Ont été sélectionnés comme système d'un réseau de prédiction de séquence et à un transformateur.
Le système DNN-LSTM et un transformateur. Le DNN combiné à deux réseaux de prédiction de séquence et à un transformateur.
Le système DNN-LSTM et un transformateur. Le DNN combiné à deux réseaux de prédiction de séquence et à un transformateur.
Le transformateur de la figure 9 par un réseau LSTM typique. Le remplace le module
Le transformateur de la figure 9 par un réseau LSTM typique. Le remplace le module
DNN, LSTM et Transformer utilisent des modèles indépendants pour traiter l'étiquette de temps.
DNN, LSTM et Transformer utilisent des modèles indépendants pour traiter l'étiquette de temps.
modules, de sorte que ces modèles ne peuvent réaliser que des diagrammes oculaires et des prédictions de valeurs
modules, de sorte que ces modèles ne peuvent réaliser que des diagrammes oculaires et des prédictions de valeurs
meilleurs paramètres structurels de ces modèles, tels que le nombre de couches cachées et la taille de la couche
meilleurs paramètres structurels de ces modèles, tels que le nombre de couches cachées et la
cachée du LSTM, la mémoire de tête, la dimension du réseau du mécanisme d'attention du Transformateur et la
taille de la couche cachée du LSTM, la mémoire de tête, la dimension du réseau du mécanisme d'attention du Transformateur et la taille de la couche cachée à rétroaction du
dimension réseau du mécanisme d'attention du Transformateur - ont tous été obtenus grâce à
Optimisation bayésienne. Comme présenté dans le tableau 6, les modèles en cascade démontrent une qualité supérieure
capacités prédictives par rapport aux modèles autonomes. Parmi les modèles en cascade,
ceux intégrant un Transformer affichent les meilleures performances prédictives, indiquant que
le modèle en cascade DNN – Transformer est le modèle de prédiction COM optimal.

Tableau 6. Valeurs d'erreur prévues de différents modèles.

Modèle	RMSE	MAE	MAPE	MRE (%)
DNN	6,9 × 10−2	4,3 × 10−2	7,8 × 10−3	6.5
LSTM	5,3 × 10−2	4,4 × 10−2	6,4 × 10−3	2.0
DNN-LSTM	5,0 × 10−2	4,1 × 10−2	6,0 × 10−3	1.8
Transformateur	4,7 × 10−2	3,7 × 10−2	5,5 × 10−3	1.7
DNN–Transformateur	4,6 × 10−2	3,5 × 10−2	5,1 × 10−3	1.6

Comme le montre le tableau 7, nous avons comparé le modèle DNN-Transformer avec les modèles
présenté dans la section Introduction en termes de fonctionnalités. Le LSTM traditionnel [20]

la structure n'a pas pu réaliser de prédiction sous des variables de paramètres physiques. Traditionnel
les modèles d'apprentissage automatique tels que les modèles SVM [13], RFR [12] et DNN [9] ont des limites
capacités de traitement des séquences et ne peut donc pas prédire les formes d'onde transitoires. Le
Le modèle GNN-RNN [21] simplifie les composants et circuits interconnectés, ce qui signifie
que certains des paramètres réels des liens sont perdus et que les performances sont diminuées. Le
les modèles ci-dessus n'ont pas pu réaliser l'optimisation des liens, et RFR [12] n'a pu compléter que FD
prédiction. En revanche, le DNN-Transformer proposé est plus pratique. Ce modèle peut
réaliser efficacement ces tâches comportementales ci-dessus et fonctionne mieux que les autres modèles
en RMSE et MAE pour la prédiction de la valeur COM. En raison de la complexité accrue de cette question
modèle, son temps d'inférence est augmenté, mais il présente toujours un avantage temporel significatif par rapport
à la simulation EM traditionnelle dont le temps de solution unique est de plusieurs heures.

Tableau 7. Comparaison fonctionnelle des différents modèles de liaisons haut débit.

Modèle	Physique Paramètres Variable	Transitoire Forme d'onde Prédiction	FD Prédiction	Lien Optimisation	GPU/CPU Temps (ms)	RMSE	MAE
SVM [13]	Oui	Non	Non	Non	<5 *	3,8 × 10−1	3,1 × 10−1
RFR [12]	Oui	Non	Oui	Non	<5 *	4,8 × 10−1	3,8 × 10−1
DNN [9]	Oui	Non	Non	Non	<5	6,9 × 10−2	4,3 × 10−2
LSTM [20]	Non	Oui	Non	Non	<5	5,3 × 10−2	4,4 × 10−2
GNN-RNN [21]	Oui	Oui	Non	Non	567.1		\
DNN–Transformateur	Oui	Oui	Oui	Oui	792,8 *	\ 4,6 × 10−2	3,5 × 10−2

* Temps d'inférence GPU.

6. Conclusions

Dans cet article, un réseau neuronal en cascade DNN-Transformer est proposé pour efficacement
analyser le SI dans les liaisons série à haut débit. Au cours du processus de création de l'ensemble de données, nous
avons fait référence aux normes USB4 Gen4 et 50GBASE-KR pour la conception de PCB et les systèmes électromagnétiques.
simulation et utilisé les paramètres de conception physique de chaque canal comme entrées pour le
modèle. Ce modèle DNN – Transformer est utilisé dans cet article pour extraire les fonctionnalités de
paramètres de conception physique des canaux et prédire avec succès les données du diagramme oculaire
et les valeurs COM des liens de test. De plus, ce modèle d'apprentissage profond peut réussir
Prédire le SBR avant et après l'égalisation et les principaux paramètres de l'égaliseur pour différents
les combinaisons sont également prédites avec précision. Pour la formation du modèle, nous avons employé un
Méthode d'optimisation bayésienne basée sur le GP pour HPO. Enfin, cet article compare
DNN-Transformer avec divers autres modèles tels que les modèles DNN, LSTM, Transformer et DNN-
LSTM. Les résultats montrent que notre modèle en cascade DNN-Transformateur est précis.
réalise la prédiction des performances et l'optimisation de l'architecture d'égalisation pour le haut débit
liaisons série, et le MRE dans ses résultats de prédiction COM pour l'ensemble de test, avec un égaliseur
configuration comprenant un TX FFE à 3 prises, un RX CTLE avec double gain DC et un 12 prises
RX DFE, est de 1,6%.

Contributions des auteurs : Conceptualisation, LW, JZ et YZ (Yinhang Zhang) ; méthodologie, LW,
JZ et HJ ; logiciels, LW; validation, LW, JZ et YZ (Yinhang Zhang) ; analyse formelle, LW,
HJ et YZ (Yinhang Zhang) ; enquête, LW; ressources, LW, YZ (Yinhang Zhang) et
YZ (Yong Zheng Zhan); conservation des données, LW et YZ (Yinhang Zhang) ; rédaction – version originale
préparation, LW; rédaction : révision et édition, XY, YZ (Yinhang Zhang) et YZ (Yongzheng
Zhan); visualisation, LW ; supervision, XY, YZ (Yinhang Zhang) et YZ (Yongzheng Zhan) ;
administration de projet, LW; acquisition de financement, LW, YZ (Yinhang Zhang) et XY Tous les auteurs
avoir lu et accepté la version publiée du manuscrit.

Financement : Ce travail a été soutenu par la Fondation nationale des sciences naturelles de Chine (Grant No.
61861019, 62161012), le ministère provincial de l'Éducation du Hunan (subvention n° 22B0525, 21A0335),

la China Postdoctoral Science Foundation (subvention n° 2024M751268) et le Postgraduate Research Programme de l'Université Jishou (subvention n° Jdy23035).

Déclaration de disponibilité des données : Les données utilisées pour étayer les conclusions de l'étude sont disponibles dans l'article.

Conflits d'intérêts : l'auteur Yongzheng Zhan était employé par la société Shandong Yunhai Guochuang Cloud Computing Equipment Industry Innovation Company Limited. Le reste les auteurs déclarent que la recherche a été menée en l'absence de toute relation commerciale ou financière relations qui pourraient être interprétées comme un conflit d'intérêts potentiel.

Abréviations

Les abréviations suivantes sont utilisées dans ce manuscrit :

DNN	Réseau neuronal profond ;
LSTM	Réseau neuronal de mémoire à long terme et à court terme ;
SBR	Réponse sur un seul bit ;
SI	Intégrité du signal
COM	Marge opérationnelle du canal ;
eCOM	Marge opérationnelle améliorée du canal ;
GP	Processus gaussien ;
HPO	Optimisation des hyperparamètres ;
EM	électromagnétique;
CEM	solveur de champ électromagnétique ;
IBIS-AMI	Spécification des informations sur le tampon d'entrée/sortie Interface de modèle algorithmique ;
EH	hauteur des yeux ;
CE	largeur des yeux;
R.L.	perte de retour ;
IL	perte d'insertion;
TD	dans le domaine temporel;
FD	domaine fréquentiel ;
ML	apprentissage automatique ;
SVM	Machine à vecteurs de support ;
LS-SVM	Machine à vecteurs de support des moindres carrés ;
DNN	Réseau neuronal à action directe ;
RFR	Régression de forêt aléatoire FNN ;
RNN	Réseau neuronal récurrent ;
FFE	égaliseur à action directe ;
DFE	égaliseur de retour de décision ;
CTLE	égaliseur linéaire à temps continu ;
ILD	écart de perte d'insertion ;
ICR	rapport perte d'insertion/diaphonie ;
ISI	interférence entre symboles ;
FOM	symbole de mérite.

Les références

1. Salle, SH ; Heck, HL Intégrité avancée du signal pour les conceptions numériques à grande vitesse ; John Wiley & Sons : Hoboken, New Jersey, États-Unis, 2009 ; pp. 201-206.
2. Yan, J. ; Zargaran-Yazd, A. Modélisation IBIS-AMI des interfaces mémoire haute vitesse. Dans les actes du 24e congrès électrique de l'IEEE 2015 Performance of Electronic Packaging and Systems (EPEPS), San Jose, Californie, États-Unis, 25-28 octobre 2015 ; pp. 73-76.
3. Comité des normes LAN/MAN de la IEEE Computer Society. Norme IEEE pour Ethernet. 2012. Disponible en ligne : https ://ieeexplore.ieee.org/document/6419735 (consulté le 26 juin 2024).
4. Wang, Y. ; Hu, QS Une optimisation de liaison série haute vitesse basée sur COM à l'aide de l'apprentissage automatique. IEICE Trans. Électron. 2022, 105, 684-691. [Référence croisée]
5. Gore, B. ; Richard, M. Un exercice sur l'application de la marge opérationnelle de canal (COM) pour la conception de canal 10GBASE-KR. Dans les actes du Symposium international IEEE 2014 sur la compatibilité électromagnétique (EMC), Raleigh, Caroline du Nord, États-Unis, du 4 au 8 août 2014 ; pp. 653-658.

6. Ambasana, N. ; Gopé, B. ; Mutnury, B. ; Anand, G. Application de réseaux de neurones artificiels pour la prédiction de la hauteur/largeur des yeux à partir des paramètres S. Dans les actes de la 23e conférence IEEE sur les performances électriques des emballages et systèmes électroniques, Portland, Oregon, États-Unis, 26-29 octobre 2014 ; pp. 99-102.
7. Ambasana, N. ; Anand, G. ; Mutnury, B. ; Gope, D. Prédiction de la hauteur/largeur des yeux à partir des paramètres S à l'aide de modèles basés sur l'apprentissage. IEEETrans. Composant. Emballage. Fab. Technologie. 2016, 6, 873-885. [Référence croisée]
8. Ambasana, N. ; Anand, G. ; Gopé, D. ; Mutnury, B. Méthode d'identification des paramètres S et de la fréquence pour les yeux basés sur ANN Prédiction hauteur/largeur. IEEETrans. Composant. Emballage. Fab. Technologie. 2017, 7, 698-709. [Référence croisée]
9. Lu, T. ; Soleil, J. ; Wu, K. ; Yang, Z. Modélisation de canaux à grande vitesse avec des méthodes d'apprentissage automatique pour l'analyse de l'intégrité du signal. IEEETrans. Électromagn. Compat. 2018, 60, 1957-1964. [Référence croisée]
10. Lho, D. ; Parc, J. ; Parc, H. ; Kang, H. ; Parc, S. ; Kim, J. Méthode d'estimation de la largeur et de la hauteur des yeux basée sur le réseau neuronal artificiel (ANN) pour USB 3.0. Dans les actes de la 27e conférence IEEE 2018 sur les performances électriques des emballages et systèmes électroniques (EPEPS), San Jose, Californie, États-Unis, 14-17 octobre 2018 ; pp. 209-211.
11. Sánchez-Masís, A. ; Rimolo-Donadio, R. ; Roy, K. ; Sulaiman, M. ; Schuster, C. Modèles FNN pour la régression des paramètres S dans les interconnexions multicouches avec différentes longueurs électriques. Dans les actes de la conférence IEEE MTT-S Latin America Microwave (LAMC) 2023, San Jose, Californie, États-Unis, 6-8 décembre 2023 ; pp. 82-85.
12. Li, X. ; Hu, Q. Une modélisation de canal basée sur l'apprentissage automatique pour une liaison série à haut débit. Dans Actes de la 6e Conférence internationale de l'IEEE 2020 sur l'informatique et les communications (ICCC), Chengdu, Chine, 11-14 décembre 2020 ; pages 1511 à 1515.
13. Trinchero, R. ; Canavero, FG Modélisation de la hauteur du diagramme de l'œil dans les liaisons à grande vitesse via une machine à vecteurs de support. Dans Actes du 22e atelier IEEE 2018 sur l'intégrité du signal et de l'alimentation (SPI), Brest, France, 22-25 mai 2018 ; p. 1 à 4.
14. Cao, Y. ; Zhang, QJ Une nouvelle approche de formation pour une modélisation robuste de réseaux neuronaux récurrents de circuits non linéaires. IEEETrans . Microw. Technologie théorique. 2009, 57, 1539-1553. [Référence croisée]
15. Mutnury, B. ; Swaminathan, M. ; Libous, JP Macromodélisation de pilotes d'E/S numériques non linéaires. IEEETrans. Av. Emballage. 2006, 29, 102-113. [Référence croisée]
16. Cao, Y. ; Erdin, moi.; Zhang, QJ Modélisation comportementale transitoire de pilotes d'E/S non linéaires combinant des réseaux de neurones et des circuits équivalents. IEEE Microw. Fil. Composant. Lett. 2010, 20, 645-647. [Référence croisée]
17. Yu, H. ; Michalka, T. ; Larbi, M. ; Swaminathan, M. Modélisation comportementale des pilotes d'E/S réglables avec préaccentuation, y compris Bruit de l'alimentation. IEEETrans. Intégration à très grande échelle. (VLSI) Syst. 2020, 28, 233-242. [Référence croisée]
18. Yu, H. ; Chalamalasetty, H. ; Swaminathan, M. Modélisation d'oscillateurs contrôlés en tension, y compris le comportement des E/S à l'aide Réseaux de neurones augmentés. Accès IEEE 2019, 7, 38973-38982. [Référence croisée]
19. Faraji, A. ; Noohi, M. ; Sadrossadat, SA; Mirvakili, A. ; Na, WC; Feng, F. Réseau neuronal récurrent profond normalisé par lots pour la macromodélisation de circuits non linéaires à grande vitesse. IEEETrans. Microw. Technologie théorique. 2022, 70, 4857-4868. [Référence croisée]
20. Moradi, M. ; Sadrossadat, A. ; Derhami, V. Réseaux de neurones à mémoire longue et à court terme pour la modélisation de l'électronique non linéaire Composants. IEEETrans. Composant. Emballage. Fab. Technologie. 2021, 1, 840-847. [Référence croisée]
21. Li, Z. ; Li, CX; Wu, ZM; Zhu, Y. ; Mao, JF Modélisation de substitution de liaisons à haut débit basée sur GNN et RNN pour l'intégrité du signal Applications. IEEETrans. Microw. Technologie théorique. 2023, 71, 3784-7796. [Référence croisée]
22. Lho, D. ; Parc, H. ; Parc, S. ; Kim, S. ; Kang, H. ; Sim, B. ; Kim, S. ; Parc, J. ; Cho, K. ; Chanson, J. ; et coll. Modèles de réseaux neuronaux profonds basés sur les caractéristiques des canaux pour une estimation précise du diagramme oculaire dans un interposeur de silicium à mémoire à haute bande passante (HBM). IEEETrans. Électromagn. Compat. 2021, 64, 196-208. [Référence croisée]
23. Li, GS ; Mao, CS ; Zhao, WS Modèle de régression semi-supervisé pour l'estimation du diagramme oculaire de l'interposeur de silicium à mémoire à large bande passante (HBM). Dans les actes du symposium 2023 de la Société internationale d'électromagnétique computationnelle appliquée, Hangzhou, Chine, 15-18 août 2023 ; p. 1–3.
24. Goay, CH ; Ahmad, Nouvelle-Écosse; Goh, P. Réseaux convolutifs temporels pour la simulation transitoire des canaux à grande vitesse. Alex. Ing. J. 2023, 74, 643-663. [Référence croisée]
25. Li, PC ; Jiao, B. ; Chou, CH; Mayder, R. ; Franzon, P. Modèle d'apprentissage profond en cascade à auto-évolution pour récepteur haute vitesse Adaptation. IEEETrans. Composant. Emballage. Fab. Technologie. 2020, 10, 1043-1053. [Référence croisée]
26. Li, PC ; Jiao, B. ; Chou, CH; Mayder, R. ; Franzon, P. Adaptation CTLE utilisant l'apprentissage profond dans SerDes Link haut débit. Dans les actes de la 70e conférence IEEE sur les composants électroniques et la technologie (ECTC), Orlando, FL, États-Unis, 3-30 juin 2020 ; pp. 952-955.
27. Zhang, HH ; Xue, ZS ; Liu, XY; Lèvre.; Jiang, LJ ; Shi, GM Optimisation du canal haut débit pour l'intégrité du signal avec Deep Algorithme génétique. IEEETrans. Électromagn. Compat. 2022, 64, 1270-1274. [Référence croisée]
28. Shan, G. ; Li, G. ; Wang, Y. ; Xing, C. ; Zheng, Y. ; Yang, Y. Application et perspectives des méthodes d'intelligence artificielle dans la prévision de l'intégrité du signal et l'optimisation des microsystèmes. Micromachines 2023, 14, 344. [CrossRef]
29. Mellitz, R. ; Ran, A. ; Mous; Ragavassamy, V. Channel Operating Margin (COM) : évolution des spécifications des canaux pour 25 Gbps et au-delà. Dans Actes de la DesignCon 2013, Santa Clara, Californie, États-Unis, 28-31 janvier 2013.
30. Pu, B. ; Lui, J. ; Harmon, A. ; Guo, Y. ; Liu, Y. ; Cai, Q. Méthodologie de conception de l'intégrité du signal pour les boîtiers optiques co-packagés basée sur le facteur de mérite en tant que marge d'exploitation du canal. Dans les actes du symposium international conjoint EMC/SI/PI et EMC Europe 2021 de l'IEEE, Raleigh, Caroline du Nord, États-Unis, 26 juillet-13 août 2021 ; pp. 492-497.
31. McCulloch, WS; Pitts, W. Un calcul logique des idées immanentes à l'activité nerveuse. Taureau. Mathématiques. Biophysique. 1943, 5, 115-133. [Référence croisée]

32. Bono, FM; Radicioni, L. ; Cinquemani, S. Une nouvelle approche pour le contrôle qualité des lignes de production automatisées travaillant sous Conditions très incohérentes. Ing. Appl. Artif. Intell. 2023, 122, 106149. [Réf. croisée]
33. Acier, RGD ; Torrie, JH Principes et procédures statistiques ; McGraw Hill : New York, New Jersey, États-Unis, 1960.
34. Girshick, R. Fast R-CNN. Dans Actes de la Conférence internationale de l'IEEE sur la vision par ordinateur (ICCV), Santiago, Chili, 7-13 décembre 2015 ; pages 1440 à 1448.
35. Kingma, DP; Ba, J. Adam : Une méthode d'optimisation stochastique. arXiv 2014, arXiv : 1412.6980.
36. Vaswani, A. ; Shazeer, N. ; Parmar, N. ; Uszkoreit, J. ; Jones, L. ; Gomez, AN; Kaiser, L. ; Polosukhin, I. L'attention est tout ce dont vous avez besoin. Dans Proceedings of the Advances in Neural Information Processing Systems, Long Beach, Californie, États-Unis, 4-9 décembre 2017 ; pages 6000 à 6010.
37. Akiba, T. ; Sano, S. ; Yanase, T. ; Ohta, T. ; Koyama, M. Optuna : Un cadre d'optimisation d'hyperparamètres de nouvelle génération. Dans Actes de la 25e Conférence internationale ACM SIGKDD sur la découverte des connaissances et l'exploration de données, Anchorage, AK, États-Unis, 4-8 août 2019.
38. Frazier, PI Un didacticiel sur l'optimisation bayésienne. arXiv 2018, arXiv : 1807.02811.

électronique

MDPI

Article

Améliorer la perception des véhicules autonomes par mauvais temps :

Un modèle multi-objectifs pour la classification météorologique intégrée et la détection d'objets

Nasser Aloufi*, Abdulaziz Alnori et Abdullah Basuhail

Département d'informatique, Faculté d'informatique et de technologie de l'information, Université King Abdulaziz, Djeddah 21589, Arabie Saoudite ; asalnori@kau.edu.sa (AA); abasuhail@kau.edu.sa (AB)
* Correspondance : nhassanaloufi@stu.kau.edu.sa

Résumé : Une détection d'objets robuste et une classification météorologique sont essentielles au fonctionnement sûr des véhicules autonomes (VA) dans des conditions météorologiques défavorables. Alors que les recherches existantes traitent souvent ces tâches séparément, cet article propose un nouveau modèle multi-objectifs qui traite la classification météorologique et la détection d'objets comme un problème unique en utilisant uniquement le système de détection de caméra AV. Notre modèle offre une efficacité améliorée et des gains de performances potentiels en intégrant l'évaluation de la qualité de l'image, le réseau contradictoire génératif à super résolution (SRGAN) et une version modifiée de You Only Look Once (YOLO) version 5. De plus, en tirant parti de la difficile détection dans des conditions météorologiques défavorables. Nature (DAWN), qui comprend quatre types de conditions météorologiques extrêmes, y compris le temps sablonneux souvent négligé, nous avons appliqué plusieurs techniques d'augmentation, ce qui a entraîné une expansion significative de l'ensemble de données de 1 027 images à 2 046 images. De plus, nous optimisons l' architecture YOLO pour une détection robuste de six classes d'objets (voiture, cycliste, piéton, moto, bus, camion) dans des scénarios météorologiques défavorables. Des expériences approfondies démontrent l'efficacité de notre approche, atteignant une précision moyenne (mAP) de 74,6 %, soulignant le potentiel de ce modèle multi-objectifs pour améliorer considérablement les capacités de perception des caméras des véhicules autonomes dans des environnements difficiles.

Citation : Aloufi, N. ; Alnori, A. ; Basuhail, A. Améliorer la perception des véhicules autonomes dans des conditions météorologiques défavorables : un modèle multi-objectifs pour la classification météorologique intégrée et la détection d'objets. Électronique 2024, 13, 3063. https://doi.org/10.3390/électronique13153063

Rédacteurs académiques : Shiping Wen et Jian Ying Xiao

Reçu : 14 mai 2024
Révisé : 22 juillet 2024
Accepté : 29 juillet 2024
Publié : 2 août 2024

Mots-clés : véhicules autonomes ; réseau neuronal conventionnel; détection d'objets ; l'apprentissage en profondeur; capteurs de caméras; des conditions météorologiques défavorables; classification météorologique

1. Introduction

Les progrès rapides de la technologie des véhicules autonomes (AV) ont attiré l'attention des chercheurs, des ingénieurs, des décideurs politiques et du public. Au cœur du développement audiovisuel se trouvent les capteurs qui permettent la perception et la prise de décision dans des environnements de conduite dynamiques. Parmi ceux-ci, les capteurs de caméra jouent un rôle essentiel en tant que principale source de perception visuelle dans les systèmes audiovisuels. Les caméras capturent des images haute résolution en temps réel de l'environnement du véhicule, fournissant des données visuelles cruciales pour la détection et la classification précises de divers objets. En tirant parti d'algorithmes avancés de détection d'objets, les caméras contribuent à diverses fonctionnalités audiovisuelles telles que le maintien de la voie et la planification des trajectoires en surveillant en permanence les marquages au sol et les changements dans le tracé des routes. Cela permet au véhicule de maintenir sa position dans les voies et de prendre des décisions éclairées concernant la trajectoire et les manœuvres, améliorant ainsi la sécurité routière globale et la fluidité du trafic. De plus, les capteurs de caméra contribuent à la planification du trajet en identifiant les obstacles, les panneaux de signalisation et d'autres entités, permettant au véhicule d'adapter sa trajectoire en conséquence et de naviguer dans des scénarios de circulation complexes.

L'estimation de la profondeur est une autre capacité clé des capteurs de caméra, permettant aux AV de percevoir avec précision les distances des objets environnants. Grâce à des techniques avancées de traitement d'images , les caméras peuvent fournir une perception de la profondeur, améliorant ainsi la conscience spatiale du véhicule et ses capacités d'évitement des obstacles. La segmentation d'images est une tâche supplémentaire effectuée

capacités de sensibilisation et d'évitement des obstacles. La segmentation d'images est une tâche supplémentaire effectuée par les capteurs de caméra, dans laquelle la scène visuelle est segmentée en régions sémantiquement significatives . Cette segmentation permet à l'AV de distinguer différents capteurs de caméra, dans lesquels la scène visuelle est segmentée en éléments sémantiquement significatifs. éléments dans son champ de vision, facilitant la détection et la classification robustes des objets, des régions. Cette segmentation permet à l'AV de distinguer les différents éléments au sein de indispensable pour une navigation en toute sécurité. De plus, les capteurs de caméra jouent un rôle crucial dans la fusion de son champ de vision, facilitant une détection et une classification robustes des objets, essentielles pour une navigation sûre , intégrant les données de plusieurs caméras positionnées autour du véhicule jusqu'à la consignation . De plus, les capteurs des caméras jouent un rôle crucial dans la fusion de la perception, intégrant construire une carte complète de connaissance de la situation. Cette fusion améliore les données du véhicule provenant de plusieurs caméras positionnées autour du véhicule pour construire une image complète. compréhension de son environnement, lui permettant de prendre des décisions éclairées en temps réel. carte de connaissance de la situation. Cette fusion améliore la compréhension du véhicule de son environnement, lui permettant de prendre des décisions éclairées en temps réel. En plus de la perception externe, les capteurs de caméra contribuent également à la surveillance et à la perception de la cabine, les capteurs de caméra contribuent également à la surveillance de la cabine et fournissent des données essentielles fournir des données essentielles sur le comportement et le statut des passagers. La figure 1 répertorie les principaux rôles des comportements et du statut des passagers. La figure 1 répertorie les principaux rôles des caméras et leurs caméras et leurs systèmes dans les AV. systèmes dans les AV.

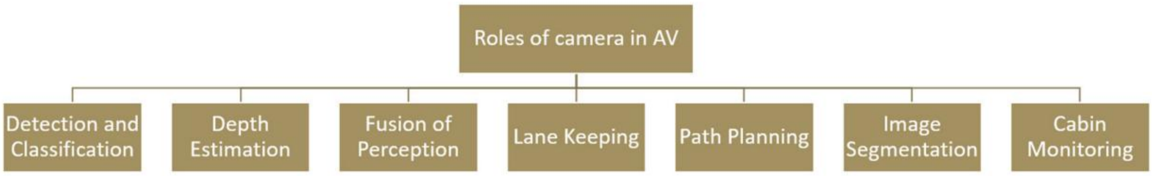


Figure 1. Principaux rôles des caméras et de leurs systèmes dans les AV.

Figure 1. Principaux rôles des caméras et de leurs systèmes dans les AV.

En plus de leurs fonctions principales, les caméras offrent un support économique et léger. En plus de leurs fonctions principales, les caméras offrent une solution rentable et légère par rapport aux technologies de capteurs alternatives telles que le LiDAR et le radar. Ce solution de poids par rapport aux technologies de capteurs alternatives telles que le LiDAR et le radar. l'abordabilité facilite l'adoption et le déploiement généralisés de la technologie audiovisuelle, ouvrant la voie à Ce prix abordable facilite l'adoption et le déploiement généralisés de la technologie audiovisuelle, ouvrant la voie à un avenir où les véhicules autonomes seront omniprésents sur nos routes. ouvrir la voie à un avenir où les véhicules autonomes seront omniprésents sur nos routes.

Cela dit, à propos des fonctions des caméras, atteindre le plus haut niveau d'AV développement, « Automatisation complète » [1], exige que les véhicules détectent chaque objet dans leur environnement d'arrondi AV dans toutes les conditions et scénarios avec rien de moins qu'un environnement semblable à celui d'un humain Le développement, « full automation » [1], exige que les véhicules détectent chaque objet dans leur comportement . Atteindre ce niveau reste très difficile en cas de conditions météorologiques défavorables, ce qui l'environnement arrondi dans toutes les conditions et scénarios avec au moins un environnement semblable à celui d'un humain présente des défis importants pour les capteurs de caméra, ayant un impact sur leur capacité à capturer des images claires. comportement. Atteindre ce niveau reste très difficile en cas de conditions météorologiques défavorables et d'images fiables de l'environnement. Les conditions météorologiques telles que la pluie, la neige, le brouillard et présente des défis importants pour les capteurs de caméra, ayant un impact sur leur capacité à capturer des tempêtes de sable claires, pose des défis importants pour les capteurs de caméra, et les principaux défis comprennent : et des images fiables de l'environnement. Les conditions météorologiques telles que la pluie, la neige, le brouillard et • Visibilité réduite : des conditions météorologiques défavorables entraînent souvent une visibilité réduite, les tempêtes de sable posent des défis importants aux capteurs de caméra, et les principaux défis comprennent : l'efficacité des capteurs de caméra à capturer des images claires de l'environnement.

- Visibilité réduite : les conditions météorologiques défavorables entraînent souvent une visibilité réduite, altérée. La pluie, la neige et le brouillard peuvent obscurcir le champ de vision, ce qui rend difficile la surveillance des caméras. l'efficacité des capteurs de caméra pour capturer des images claires de l'environnement. • Gouttelettes d'eau et accumulation de neige : la pluie et la neige peuvent produire des gouttelettes d'eau ment. La pluie, la neige et le brouillard peuvent obscurcir le champ de vision, ce qui rend difficile l'accumulation de neige sur les objectifs de l'appareil photo, entraînant une distorsion, un flou ou une occlusion. caméras pour discerner les objets et les obstacles avec précision.
d'images capturées. Cette accumulation peut dégrader la qualité de l'image et gêner l'objet
- Gouttelettes d'eau et accumulation de neige : la pluie et la neige peuvent entraîner des capacités de détection et de reconnaissance des gouttelettes d'eau. ou une accumulation de neige sur les objectifs de l'appareil photo, entraînant une distorsion, un flou ou une occultation. • Brouillard et brume : les conditions de brouillard créent une atmosphère brumeuse qui réduit le contraste et la brume. sion des images capturées. Cette accumulation peut dégrader la qualité de l'image et nuire à la clarté des images de la caméra, entravant ainsi la détection et la localisation des objets. La présence de capacités de détection et de reconnaissance d'objets. le brouillard et la brume empêchent les caméras de distinguer les objets de leur arrière-plan, systèmes de perception audiovisuelle.
- Brouillard et brume : les conditions de brouillard créent une atmosphère brumeuse qui réduit le contraste et compromet la fiabilité des clartés des images de la caméra, ce qui entrave la détection et la localisation des objets. La présence • Éblouissement et reflets : reflets dans les images de l'appareil photo, entraînant des images surexposées ou délavées. Éblouissement le brouillard et la brume rendent difficile pour les caméras de distinguer les objets de leur dos- et les reflets peuvent obscurcir des informations visuelles importantes, ce qui rend la tâche difficile pour les caméras. sol, compromettant la fiabilité des systèmes de perception audiovisuelle.
Des véhicules utilitaires pour naviguer en toute sécurité dans des conditions météorologiques défavorables.
- Éblouissement et reflets : l'éblouissement provenant des routes mouillées ou des surfaces réfléchissantes peut provoquer • Particules de sable et de poussière : les tempêtes de sable et les conditions poussiéreuses peuvent entraîner une accumulation reflets dans les images de l'appareil photo, entraînant des images surexposées ou délavées. de particules de sable et de poussière sur les objectifs des caméras, obstruant le champ de vision et se dégradant. L'éblouissement et les reflets peuvent obscurcir des informations visuelles importantes, ce qui rend la qualité de l'image problématique. Cette accumulation de particules peut compromettre les performances de permettre aux VA de naviguer en toute sécurité dans des conditions météorologiques défavorables. capteurs de caméra, affectant la fiabilité des systèmes de perception AV.
- Particules de sable et de poussière : les tempêtes de sable et les conditions poussiéreuses peuvent entraîner une accumulation • Conditions d'éclairage dynamiques : des conditions météorologiques défavorables peuvent provoquer des changements rapides de lation de particules de sable et de poussière sur les objectifs des appareils photo, obstruant le champ de vision et les conditions d'éclairage, y compris des variations de luminosité, de contraste et de température de couleur. Les systèmes de caméras doivent s'adapter à ces conditions d'éclairage dynamiques pour maintenir des dégradant la qualité de l'image. Cette accumulation de particules peut compromettre la perception de l'environnement et garantir une détection et une reconnaissance fiables des objets. performances des capteurs de caméra, affectant la fiabilité des systèmes de perception AV.
- Conditions d'éclairage dynamiques : des conditions météorologiques défavorables peuvent provoquer des changements rapides des conditions d'éclairage, notamment des variations de luminosité, de contraste et de température de couleur.

Les systèmes de caméras doivent s'adapter à ces conditions d'éclairage dynamiques pour maintenir une perception précise de l'environnement et garantir une détection et une reconnaissance fiables des objets.

• Étalonnage du capteur : des conditions météorologiques défavorables peuvent nécessiter des ajustements de la caméra paramètres d'étalonnage pour compenser les changements d'éclairage, de visibilité et de capteur performances. Garantir un étalonnage précis du capteur est essentiel pour maintenir la fonctionnalité et l'efficacité des capteurs de caméra dans des conditions météorologiques défavorables.

• Fiabilité et robustesse : des conditions météorologiques défavorables exigent fiabilité et robustesse. défis pour les capteurs de caméra, les obligeant à continuer à fonctionner efficacement dans des conditions environnementales difficiles. Assurer la durabilité et la résilience des Les capteurs sont essentiels au fonctionnement sûr et fiable des AV par mauvais temps.

La figure 2 montre certains problèmes de détection d'objets par mauvais temps. Chaque type de la météo a ses propres obstacles. Par exemple, lors de fortes chutes de neige et de tempêtes de sable, les limites de la route peuvent être obscurcies par la neige et le sable. Pendant la pluie, des gouttelettes d'eau sur les lentilles de les objectifs entraînent une distorsion et un flou des images capturées. Tous les défis mentionnés ci-dessus font qu'il est difficile pour la caméra de percevoir avec précision les objets.

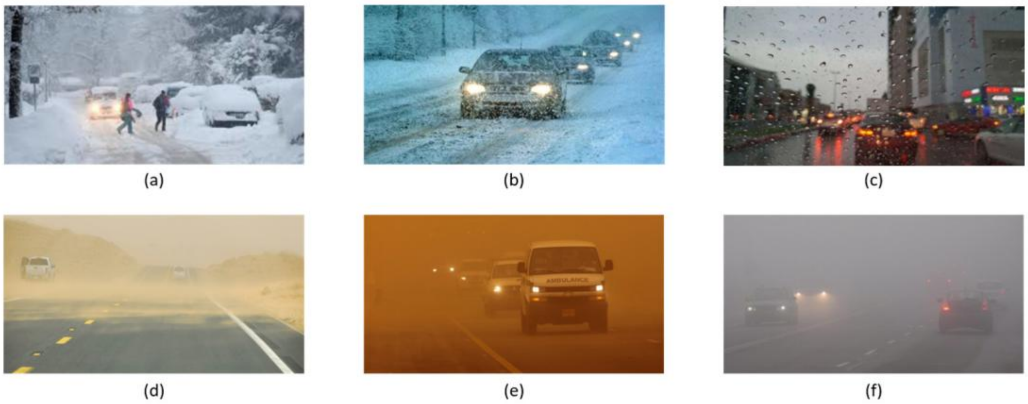


Figure 2. Quelques défis météorologiques défavorables. (a,b) De fortes chutes de neige peuvent obscurcir les voies routières. (c) La pluie mène à la distorsion et au flou des images capturées. (d,e) Une tempête de sable peut masquer les limites des routes et l'éclairage de la scène. (f) Le brouillard rend difficile pour les caméras de distinguer les objets en raison de l'éclairage modifié de la scène. arrière-plans.

Lors de l'exploration de la détection d'objets à l'aide de réseaux de neurones convolutifs (CNN), nous rencontrons deux approches principales : en une étape et en deux étapes. L'approche en deux étapes, lancé par l'introduction du modèle Region-Based CNN (R-CNN) en 2014, implique une étape de proposition de région pour identifier les régions contenant potentiellement des objets, suivie de l'extraction de caractéristiques et de la classification des objets [2]. Cependant, cette méthode était lente en raison du traitement de chaque région proposée séparément. Fast R-CNN, introduit l'année suivante, a amélioré la vitesse en faisant passer l'image entière via le CNN, générant une carte de caractéristiques pour la détection d'objets [3]. Un R-CNN plus rapide [4] a ensuite été introduit pour améliorer encore les performances. Une avancée significative a eu lieu en 2017 avec l'introduction de Mask R-CNN [5]. Mask R-CNN adopte le Feature Pyramid Network (FPN) comme épine dorsale [6] et introduit une nouvelle phase dans le processus de détection en générant un masque de segmentation pour chaque objet.

D'autre part, l'approche en une étape, démontrée pour la première fois par Redmon et al. [7] avec le modèle YOLO, encapsule l'ensemble du processus de détection en un seul passage via CNN. YOLOv2 et les itérations ultérieures comme YOLOv3 [8] ont introduit un nouveau squelette nommé Darknet-53. Une nouvelle version de YOLO appelée YOLOv4 a ensuite proposé de viser à améliorer à la fois la précision et la vitesse et a réalisé une amélioration notable de l'architecture avec un rapport à ses prédécesseurs. YOLOv5, YOLOv7 [9] et YOLOv8 étaient des itérations ultérieures de YOLO. Un autre modèle appelé Single-Shot multibox Detector (SSD), proposé dans [10], a obtenu des résultats compétitifs sur l'ensemble de données VOC2007, avec des améliorations du mAP.

Dans cet article, notre objectif est de proposer une solution pour les AV basée sur des capteurs de caméra qui permettant non seulement de détecter des objets, mais également de classer la météo en fonction de l'état de la scène. La portée de notre article est présentée dans la figure 3.

Dans cet article, notre objectif est de proposer une solution pour les AV basée sur des capteurs de caméra capables non seulement de détecter des objets, mais également de classer la météo en fonction de l'état de la scène. La portée de notre article est présentée dans la figure 3.

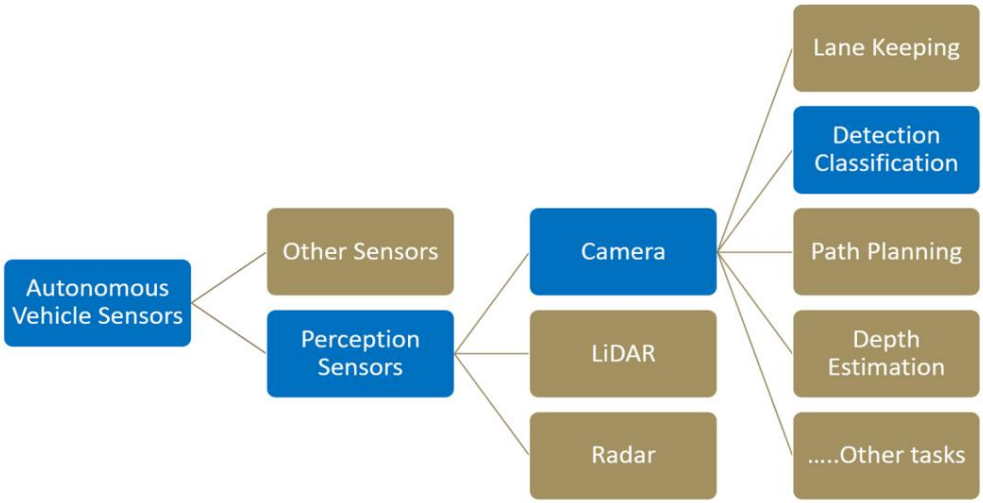


Figure 3. La portée de cet article est mise en évidence en bleu.

Figure 3. La portée de ce document est mise en évidence en bleu.

Les principales contributions de nos travaux sont les suivantes :

Les principales contributions de nos travaux sont les suivantes :

• Nous proposons un modèle multi-objectifs pour classifier la météo et détecter des objets. Comme nous le démontrerons dans la section 2, et au meilleur de nos connaissances, les articles AV existants traitent la classification météorologique et la détection d'objets comme des problèmes distincts. Notre proposition traiter la classification des conditions météorologiques et la détection d'objets comme des problèmes distincts. Notre modèle proposé traite la classification météorologique routière et la détection d'objets routiers comme un problème unifié. Le modèle traite la classification météorologique routière et la détection d'objets routiers comme un problème unifié pour les systèmes de détection par caméra.

pour les systèmes de détection de caméra.

• Nous avons élargi l'ensemble de données de détection dans des conditions météorologiques défavorables (DAWN) en

• Nous avons élargi l'ensemble de données DAWN (Detection in Adverse Weather Nature) en ajoutant des images

augmentées qui couvrent les quatre types de conditions météorologiques (sable, ajout d'images augmentées qui couvrent les quatre types de conditions météorologiques (sable, pluie, brouillard et neige). La taille totale de l'ensemble de données a presque doublé, passant de pluvieux, brumeux et neigeux). La taille totale de l'ensemble de données a presque doublé, augmentant par rapport à sa taille d'origine. sa taille originale.

• En plus de fournir un modèle unique pour classer la météo et détecter des objets,

En plus de fournir un modèle unique pour classer la météo et détecter des objets, nous comblons une lacune critique dans la recherche sur les véhicules autonomes en considérant le temps sablonneux.

nous comblons une lacune critique dans la recherche sur les véhicules autonomes en prenant en compte les conditions météorologiques sablonneuses, qui ont été largement négligées par les études existantes. conditions, qui ont été largement négligées par les études existantes. • L'architecture de base du modèle de détection d'objets You Only Look Once (YOLO)

• L'architecture de base du modèle de détection d'objets You Only Look Once (YOLO) version 5 a été adaptée et

modifiée pour s'adapter à notre domaine. En conséquence, nous avons suc- version 5 a été adaptée et modifiée

pour s'adapter à notre domaine. En conséquence, nous avons progressivement augmenté la précision moyenne (mAP) à 74,6 %, ce qui est prometteur. a réussi à augmenter la précision moyenne moyenne (mAP) à 74,6 %, ce qui est un résultat par rapport à d'autres articles utilisant le même ensemble de données. résultat prometteur par rapport à d'autres articles utilisant le même ensemble de données.

2. Travaux connexes

2. Travaux connexes

La détection d'objets dans des conditions météorologiques difficiles présente des difficultés car le

La détection d'objets dans des conditions météorologiques difficiles présente des difficultés car la qualité des images

se dégrade et les caractéristiques visuelles sont compromises en raison de phénomènes météorologiques. La qualité des images se dégrade et les caractéristiques visuelles sont compromises en raison de phénomènes météorologiques tels que la pluie, le brouillard, la neige et les tempêtes de sable. Ces conditions ont un impact sur les performances de détection en comme la pluie, le brouillard, la neige et les tempêtes de sable. Ces conditions ont un impact sur les performances de détection d'objets, entravant l'éclairage de la scène, en réduisant la visibilité des objets et en compliquant la différenciation des par rapport aux éléments environnants. Plusieurs articles ont été publiés visant à proposer objets en compliquant la différenciation des objets par rapport aux éléments environnants. Plusieurs articles ont été publiés visant à proposer des éléments environnants. Plusieurs articles ont été publiés visant à proposer des solutions adaptées. solutions adaptées.

Dans [11], les auteurs ont étudié la classification des conditions météorologiques défavorables ainsi que le niveau de lumière dans le ciel.

Dans [11], les auteurs ont étudié la classification des conditions météorologiques défavorables ainsi que le niveau de

Environnement nuageux individuel. Aborder la question de la perception en cas de mauvais temps et de faible luminosité

Environnement nuageux individuel. Aborder le problème de la perception dans des conditions météorologiques défavorables et en cas de faible luminosité,

Dans des conditions où la dégradation de la précision constitue un problème important, les auteurs ont introduit la neige) et

Dans des conditions où la dégradation de la précision constitue un problème important, les auteurs ont introduit la neige) et

leur propre ensemble de données. L'ensemble de données a été conçu pour couvrir trois types de temps (brouillard, pluie et neige) et trois niveaux d'éclairage (fort, modéré et faible), ainsi que trois types de rues liés au type de météo, au niveau (asphalte, herbe et pavés). Les images traitées de l'ensemble de données contiennent trois étiquettes d'éclairage et au type de rue. Les auteurs ont utilisé ResNet18 comme

(asphalte, herbe et pavés). Les images traitées de l'ensemble de données contiennent trois étiquettes sur leur base et ont

conclu que le système fonctionnait avec une faible précision sur l'ensemble de données. liés au type de temps, au niveau d'éclairage et au type de rue. Les auteurs ont utilisé ResNet18 au fur et à mesure et nécessitaient des améliorations supplémentaires.

Dans [12], les auteurs abordent les défis des AV lors de conditions météorologiques défavorables, où les modèles perceptuels typiques luttent. Les recherches existantes se concentrent principalement sur la classification conditions météorologiques; cependant, les auteurs ont étudié les transitions entre ces types de

météo. Ils ont proposé une méthode pour définir et comprendre six états de transition météorologiques intermédiaires (nuageux à pluvieux, pluvieux à nuageux, ensoleillé à pluvieux, pluvieux à ensoleillé, ensoleillé à brumeux et brumeux à ensoleillé). L'approche consiste à interpoler des données de transition météorologique intermédiaire à l'aide d'un auto-encodeur de variation, à extraire des caractéristiques spatiales avec des réseaux convolutifs très profonds VGG (Visual Geometry Group) et à modéliser la distribution temporelle avec une unité récurrente fermée pour la classification. Les auteurs ont proposé un nouvel ensemble de données à grande échelle appelé AIWD6 (Adverse Intermediate Weather Driving), et les résultats ont montré un modèle de transition météorologique efficace.

Dans [13], les auteurs introduisent un nouveau framework appelé WeatherNet, qui utilise quatre modèles CNN profonds basés sur l'architecture ResNet50. WeatherNet extrait de manière autonome les informations météorologiques de l'image d'entrée et classe la sortie dans la bonne catégorie. Cependant, l'inconvénient du framework présenté est l'incapacité de partager des fonctionnalités, puisque les quatre modèles fonctionnent séparément.

Réf. [14] se concentre sur l'impact significatif des conditions météorologiques défavorables sur le trafic urbain et souligne l'importance de la reconnaissance des conditions météorologiques pour des applications telles que l'assistance audiovisuelle et les systèmes de transport intelligents. Tirant parti des progrès de l'apprentissage profond, l'article présente un nouveau modèle simplifié appelé ResNet15, une version proposée du célèbre ResNet50 [15]. Le modèle proposé comporte une couche entièrement connectée qui utilise le classificateur Softmax. Le document présente également un nouvel ensemble de données appelé « WeatherDataset-4 » contenant environ 5 000 images couvrant le temps brumeux, pluvieux, enneigé et ensoleillé. Bien que le réseau proposé ait surpassé le ResNet50 traditionnel, le document ne couvre pas les environnements nocturnes et sablonneux.

Dans [16], les auteurs ont proposé l'algorithme MCS-YOLO pour améliorer la détection d'objets en intégrant un mécanisme d'attention coordonnée, une structure multi-échelle pour les petits objets et en appliquant la structure Swin Transformer [17]. Grâce à des expériences sur l' ensemble de données BDD100K, ils ont démontré une précision moyenne (mAP) de 53,6 %.

L'article [18] est l'un des premiers articles à avoir appliqué CNN pour la classification météorologique AV. Les auteurs ont ajouté deux couches entièrement connectées pour extraire les caractéristiques des images des conditions de service routier (RSC). L'article s'est concentré sur les conditions routières hivernales, où le problème des routes enneigées a été divisé en trois expériences : (a) une classification en deux classes, (b) une classification en trois classes et (c) une classification en cinq classes. Le modèle a surpassé les techniques de classification traditionnelles et a enregistré une précision de 78,5 % lors de l'application d'une classification en cinq classes. Dans [19], YOLOv4 a été amélioré pour détecter des objets en proposant une tête sans ancre et découplée. Le document a utilisé BDD100k comme ensemble de données original et a créé une nouvelle version qui se concentre sur trois types de temps (pluie, neige, brouillard). Les résultats expérimentaux ont montré un mAP de 60,3 %.

Dans [20], les auteurs ont extrait des données de mouvement de haute précision et ont proposé un nouveau mécanisme de suivi des véhicules appelé SORT++. Image-Adaptive YOLO (IA-YOLO) a été présenté dans [21] et a montré une amélioration dans la détection d'objets dans des environnements de faible luminosité et de brouillard.

Réf. [22] ont proposé un réseau à double sous-réseau (DSNet) pour détecter des objets et ont obtenu un mAP de 50,8 % par temps brumeux. Dans [23] YOLOv5 a été étudié pour détecter des objets de plusieurs classes, et le mAP de toutes les classes a obtenu un score de 25,8 %. Dans [24], des images de drones ont été créées et appliquées à une version modifiée de YOLOv5, qui a obtenu un mAP d'environ 50 %.
L'article [25] a comparé les performances de YOLOv3, YOLOv4 et Faster R-CNN dans différents types de temps (pluie, brouillard, neige). Le document concluait que YOLOv4 surpassait YOLOv3 et Faster R-CNN.

Le tableau 1 présente un résumé des publications récentes sur la classification météorologique et la détection d'objets dans l'environnement AV. Alors que les modèles standard de détection d'objets se concentrent principalement uniquement sur le processus de détection, nos travaux et le modèle proposé introduisent plusieurs différences clés par rapport aux études récentes connexes. Premièrement, nous avons incorporé une nouvelle phase dans notre modèle appelée « bloc de qualité », conçue pour évaluer et améliorer la scène Deuxièmement, nous avons ajouté un score seuil réglable pour réduire le nombre d'images entrant dans la phase d'amélioration. Troisièmement, notre étude porte uniquement sur les conditions météorologiques sablonneuses, qui n'ont pas été prises en compte dans les publications récentes.

Tableau 1. Des publications récentes dans le domaine AV éliminent le temps sablonneux de leurs études.
De plus, il existe une lacune dans la combinaison de la classification météorologique et de la détection d'objets.

Papier	Météo Classification	Objet Détection	Entrez en voiture Neige	Entrez en voiture Pluie	Entrez en voiture Brouillard	Entrez en voiture Sable
[11]	√	×	√	√	√	×
[13]	√	×	√	√	×	×
[14]	√	×	√	√	√	×
[16]	×	√	√	√	√	×
[18]	√	×	√	×	×	×
[19]	×	√	√	√	√	×
[20]	×	√	√	√	×	×
[21]	×	√	×	×	√	×
[22]	×	√	×	×	√	×
[23]	×	√	×	√	×	×
[24]	×	√	√	√	×	×
[25]	×	√	√	√	√	×
Les notres	√	√	√	√	√	√

3. Méthodologie

Notre méthodologie pour développer un modèle capable à la fois de classification météorologique et la détection d'objets par temps violent a commencé par l'application de la détection dans des conditions météorologiques défavorables. (DAWN) [26]. Nous nous sommes concentrés sur quatre types de temps clés (sable, pluvieux, brumeux, et neigeux) avec six classes (piéton, vélo, voiture, moto, bus et camion). Étendre l'ensemble de données et introduisons une nouvelle variation des images existantes, nous avons inclus les données augmentation de notre travail. Cet ensemble de données augmenté a été combiné avec le DAWN original ensemble de données pour augmenter le nombre d'échantillons d'entraînement. Une description complète de l'augmentation sera fourni dans la section 6. Nous avons ensuite divisé l'ensemble de données combiné en formations et ensembles de validation. Notre pourcentage de répartition est de 80 % des images destinées à la formation, tandis que (20 %) étaient utilisé pour la validation et les tests (10 % pour la validation et 10 % pour les tests). L'ensemble de formation a été utilisé pour entraîner à la fois les modèles de classification météorologique et de détection d'objets, tandis que le L'ensemble de validation a joué le rôle essentiel de prévention du surapprentissage. Après cela, les étapes d'optimisation sont impliqués pour trouver les meilleures performances du modèle en modifiant les hyperparamètres. Enfin, nous avons évalué les modèles optimisés en utilisant la précision moyenne moyenne standard (mAP), de précision et de mesures de rappel. La figure 4 montre la séquence de notre méthodologie.

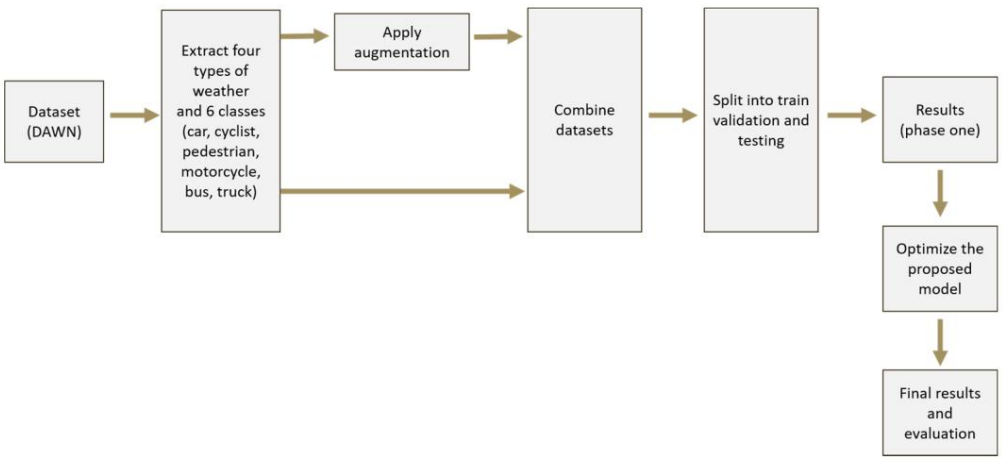


Figure 4. Séquence de notre méthodologie dans cet article.
Figure 4. Séquence de notre méthodologie dans cet article.

Pour optimiser l'efficacité de calcul étant donné les ressources GPU limitées , nous avons utilisé
La plate-forme Colab basée sur le cloud de Google comme environnement expérimental. Colab fournit un framework d'apprentissage automatique PyTorch et des GPU hautes performances (tels que le Tesla T4). Colab, nous avons pu exécuter efficacement nos expériences, en particulier avec le
Un framework d'apprentissage automatique PyTorch et des GPU hautes performances (tels que le Tesla T4). Grâce à Colab, nous avons pu exécuter efficacement nos expériences, en particulier avec le
avec l'intégration avec CUDA (Compute Unified Device Architecture), qui facilite l'intégration avec CUDA (Compute Unified Device Architecture), qui aide à accélérer
en accélérant le processus de calcul de notre pipeline et de la partie CNN tout en détectant les objets (en particulier dans les tâches telles que les couches de convolution, de pooling, de normalisation et d'activation).

Diverses mesures sont disponibles pour quantifier l'efficacité des modèles de détection d'objets. Dans notre article, nous avons donné la priorité à trois mesures principales : (a) la précision moyenne moyenne (mAP), (b) la précision et (c) le rappel. mAP est une mesure d'évaluation répandue dans le domaine.

le processus de calcul de notre pipeline et de la partie CNN lors de la détection d'objets
(en particulier dans les tâches telles que les couches de convolution, de pooling, de normalisation et d'activation).

Diverses mesures sont disponibles pour quantifier l'efficacité des modèles de détection d'objets. Dans Dans notre article, nous avons priorisé trois mesures principales : (a) la précision moyenne moyenne (mAP), (b) la précision et (c) le rappel. mAP constitue une métrique d'évaluation répandue dans le domaine de l'objet détection, offrant une évaluation holistique de la compétence du modèle en matière d'identification d'objets et la localisation. mAP combine précision et rappel en calculant la précision moyenne (AP) pour chaque classe ou catégorie d'objet, dérivant ensuite la moyenne pour toutes les classes. PA sert de mesure de la qualité de la détection, encapsulant à la fois la précision de objets identifiés et l'exhaustivité de la détection sur la scène. Grâce au calcul de mAP, les performances de notre pipeline peuvent être comparées et évaluées numériquement sur divers domaines et scénarios.

Nous avons également considéré la précision et le rappel comme des mesures indispensables dans le contexte de détection d'objets. La précision est la proportion ou le pourcentage d'éléments récupérés qui sont pertinent pour la classe correcte, tandis que le rappel mesure le pourcentage d'objets pertinents qui sont récupéré avec succès. La précision est exprimée comme le rapport des vrais positifs (TP) à la somme de vrais positifs et de faux positifs (FP), représentés par :

$$\text{Précision} = TP/(TP + FP)$$

Le rappel est le rapport du TP à la somme des vrais positifs et des faux négatifs (FN), représenté par :

$$\text{Rappel} = TP/(TP + FP)$$

4. Ensemble de données

Pour l'ensemble de données, comme nous l'avons mentionné précédemment, nous avons utilisé DAWN dans notre développement et expérimentation. L'ensemble de données DAWN couvre quatre types de conditions météorologiques défavorables : tempête de sable, pluie, neige et brouillard. La figure 5 montre un échantillon des différents types de temps couverts par AUBE. L'ensemble de données contient 1027 images couvrant les quatre types de conditions météorologiques et différents contextes environnementaux tels que les autoroutes et les paysages urbains, garantissant un une large représentation de scénarios du monde réel.

Figure 5. L'ensemble de données DAWN fournit diverses conditions météorologiques difficiles telles que le brouillard, la pluie, la neige et sable [26].

Bien que de nombreux autres ensembles de données couvrent des conditions météorologiques défavorables, l'ensemble de données DAWN l'avantage d'inclure des images de tempête de sable ou de temps sableux, qui sont souvent absentes. a à partir d'autres ensembles de données. Cette fonctionnalité unique de DAWN a permis à notre modèle d'aborder la classification météorologique et la détection d'objets dans plusieurs types d'environnements géographiques.

L'ensemble de données DAWN utilisé se compose à l'origine de 1027 images d'une taille de 640 × 640.

L'annotation d'image contient la classe de l'objet et les limites correspondantes de : x, y, largeur et hauteur du cadre de délimitation (x_center, y_center, width, height). Figue-

Figure 5. L'ensemble de données DAWN fournit diverses conditions météorologiques difficiles telles que le brouillard, la pluie, la neige et le sable [26].

Bien que de nombreux autres ensembles de données couvrent des conditions météorologiques défavorables, l'ensemble de données DAWN présente l'avantage d'inclure des images de tempêtes de sable ou de temps sablonneux, qui sont souvent absentes des autres ensembles de données. Cette fonctionnalité unique de DAWN a permis à notre modèle de prendre en compte d'autres ensembles de données. Cette fonctionnalité unique de DAWN a permis à notre modèle de répond classification météorologique et détection d'objets dans plusieurs types d'environnements géographiques. météorologique et détection d'objets dans plusieurs types d'environnements géographiques.

L'ensemble de données DAWN utilisé se compose à l'origine de 1027 images d'une taille de 640 × 640. L'ensemble de données DAWN utilise se compose à l'origine de 1027 images d'une taille de 640 × 640.

L'annotation d'image contient la classe de l'objet et les limites correspondantes. L'annotation d'image contient la classe de l'objet et les limites correspondantes de : x, y, largeur et hauteur du cadre de délimitation (x_center, y_center, width, height). Fig- x, y, largeur et hauteur du cadre de délimitation (x_center, y_center, width, height). Figure 6

La figure 6 représente un échantillon de nos images étiquetées, considérées comme une référence de vérité terrain. représente un échantillon de nos images étiquetées, considérées comme une référence de vérité terrain.

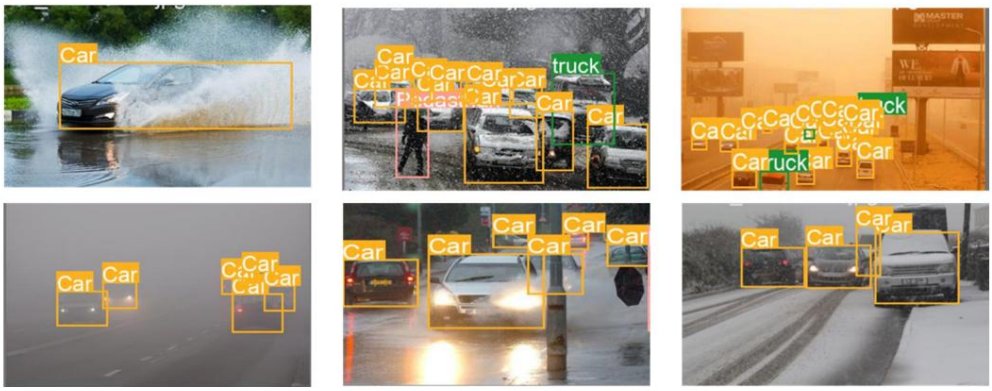


Figure 6. Échantillon d'images étiquetées. Figure 6. Échantillon d'images étiquetées.

5. Modèle proposé 5. Modèle proposé Notre

pipeline proposé établit une classification météorologique complète et un objet Notre pipeline proposé établit une classification météorologique complète et une détection d'objets en intégrant tâches principales : (1) évaluation de la qualité de l'image, (2) amélioration de l'image. ment, (3) la classification quatre tâches principales : (1) évaluation de la qualité de l'image, (2) détection d'amélioration de l'image en intégrant quatre météorologique et (4) la détection d'objets. Nous avons combiné les tâches 1 et 2 ment, (3) la classification météorologique et (4) la détection d'objets. Nous avons combiné la tâche 1 et en un seul bloc appelé « Bloc Qualité », tandis que les tâches 3 et 4 sont combinées dans un autre bloc 2 en un seul bloc appelé « Bloc qualité », tandis que les tâches 3 et 4 sont combinées dans un autre appelé « Bloc Classifier et détecter ». Une fois l'image entrée dans le modèle, une évaluation doit être effectuée bloc appelé « Classifier et détecter le bloc ». Une fois l'image entrée dans le modèle, une évaluation sera fait pour évaluer la qualité de l'image. Le but de cette étape est de décider si l'image capturée a besoin d'être pour évaluer la qualité de l'image. Le but de cette étape est de décider si l'image capturée a besoin d'être améliorées ou non. L'aveugle sans référence de savoir si l'image capturée a besoin d'être améliorée ou non. La méthode Blind Refer- Image Spatial saisie a un score supérieur à un point seuil Quality Evaluator (BRISQUE) [27] a été utilisée pour vérifier la qualité de l'image sans quance . Si l'image pour quantifier la qualité de l'image. Si l'image saisie a un score supérieur à un seuil (faible qualité), l'image sera refusée et transférée vers une étape d'amélioration ;

sinon, il sera approuvé et transféré directement au service de classification et de détection bloc. Le point seuil peut être changé et modifié en fonction de la situation de la scène ; pour Par exemple, dans notre expérience, nous avons utilisé un seuil de 42,7, comme nous l'expliquerons dans la section 7. «Expériences et résultats». Il est à noter que, dans la méthode BRISQUE, généralement un un score inférieur indique une meilleure qualité de perception, tandis qu'un score BRISQUE plus élevé indique moins bonne qualité de perception. L'algorithme BRISQUE présente plusieurs avantages qui en font un solution adaptée à notre modèle et pour évaluer les scènes météorologiques défavorables. Premièrement, c'est un métrique de qualité d'image sans référence, dans la mesure où elle ne nécessite pas une image de référence parfaite pour comparaison. Ceci est très avantageux dans des conditions météorologiques défavorables, où l'obtention des images idéales et non déformées peuvent être difficiles, voire impossibles. BRISQUE fonctionne par analyser les statistiques de scène naturelle (NSS) d'une image et les comparer aux statistiques attendues d'images naturelles (non déformées). Tout écart par rapport à ce naturel est signalés comme indicateurs de dégradation de la qualité, ce qui en fait un bon choix pour détecter les types de distorsions introduites par les phénomènes météorologiques. De plus, BRISQUE propose des outils informatiques efficacité par rapport à plusieurs autres options, ce qui peut être important lorsque vous travaillez avec de grands ensembles de données d'images ou dans des scénarios où une évaluation de la qualité en temps réel est souhaitée. Pour la phase d'amélioration de l'image, nous avons utilisé le Super-Resolution Generative Adversarial Technique de réseau (SRGAN) [28], qui consiste en des réseaux générateurs et discriminateurs.

Le réseau générateur vise à valoriser les images basse résolution, tandis que le discriminateur Le réseau vise à affiner la sortie du générateur, ce qui entraîne une meilleure clarté de l'image. Suite au Quality Block, l'image est traitée par deux réseaux YOLOv5. Un YOLO Le réseau, largement formé sur un ensemble de données d'images météorologiques, classe avec précision conditions météorologiques telles que sable, pluie, neige ou brouillard. Simultanément, un distinct

technique [28], qui consiste en des réseaux générateurs et discriminateurs. Le réseau de générateurs vise à améliorer les images à basse résolution, tandis que le réseau de discriminateurs vise à affiner la sortie du générateur, ce qui améliore la clarté de l'image. Suite au Quality Block, l'image est traitée par deux réseaux YOLOv5. Un réseau YOLO,

formé de manière approfondie sur un ensemble de données d'images météorologiques, classe avec précision les conditions météorologiques telles que le sable, la pluie, la neige ou le brouillard. Simultanément, un réseau YOLO distinct, formé pour identifier et localiser les objets avec des cadres englobants, détectera le réseau YOLO ciblé, formé pour identifier et localiser les objets avec des cadres englobants, détectera

objets, tels que les voitures, les cyclistes, les piétons, les motos, les bus et les camions. Généralement, nos objets ciblés, tels que les voitures, les cyclistes, les piétons, les motos, les bus et les camions.

Le modèle proposé propose une approche à deux volets, donnant la priorité à la qualité de l'image. Généralement , notre modèle proposé propose une approche à deux volets, donnant la priorité à la qualité de l'image.

transition transparente vers une classification météorologique et une détection d'objets robustes basées sur YOLO avant de passer de manière transparente à une classification météorologique et des objets robustes basés sur YOLO

tion pour une analyse d'image fiable. La figure 7 montre une illustration de notre proposition.

détection pour une analyse d'image fiable. La figure 7 montre une illustration de notre proposition.

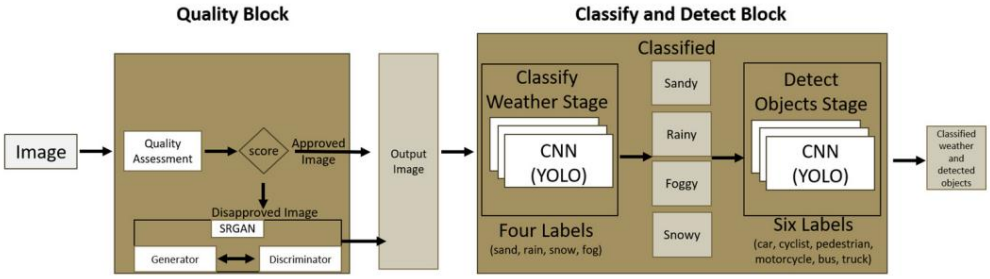


Figure 7. Notre modèle proposé.

Figure 7. Notre modèle proposé.

6. AUBE augmentée

6. DAWN augmenté Étant

donné le nombre limité d'images météorologiques défavorables dans l'ensemble de données DAWN, nous avons construit créé une nouvelle version de DAWN utilisant l'augmentation pour étendre notre ensemble de données expérimentales. Données

Compte tenu du nombre limité d'images météorologiques défavorables dans l'ensemble de données DAWN, nous avons construit une nouvelle version de DAWN en utilisant l'augmentation pour étendre notre ensemble de données expérimentales.

L'augmentation est une méthode largement utilisée pour agrandir artificiellement des ensembles de données en créant de nouveaux

L'augmentation des données est une méthode largement utilisée pour agrandir artificiellement des ensembles de données en créant

créent des images d'entraînement à partir d'un ensemble de données actuellement disponible. Divers articles, tels que [29,30], et le

de nouvelles images d'entraînement à partir d'un ensemble de données actuellement disponible. Divers articles, tels que [29,30], et le

tels que [29,30], amélioration incrémentielle de YOLOv3 [8] ont utilisé l'augmentation des données soit pour leur

d'augmentation des données ou de détection d'objets. Notre version augmentée a augmenté

augmenté le nombre d'ensembles de données DAWN de 1027 images à 2046 images, soit presque le double de la taille actuelle. La

figure 8 montre un aperçu général de l'ensemble de données DAWN avant et après notre augmentation appliquée.

avant et après notre augmentation appliquée.

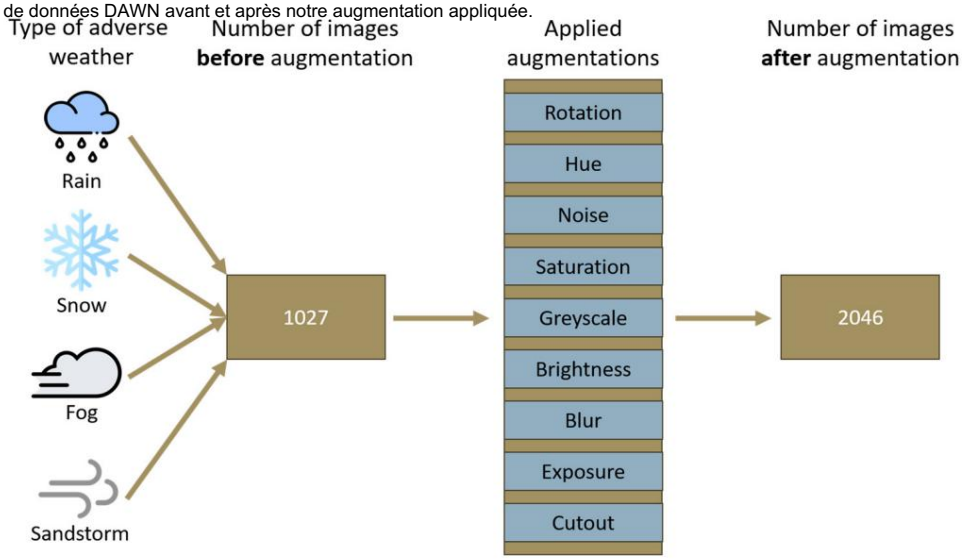


Figure 8. L'ensemble de données DAWN est passé de 1 027 à 2 046 images.

Figure 8. L'ensemble de données DAWN est passé de 1 027 à 2 046 images.

Les techniques d'augmentation englobent une gamme de réglages d'image, tels que l'image

Les techniques d'augmentation englobent une gamme de réglages d'image, tels que la rotation, le recadrage, le réglage des couleurs, le réglage de la netteté, ainsi que des modifications plus complexes telles que la

recadrage, réglage des couleurs, réglage de la netteté, ainsi que des modifications plus complexes telles que la

rotation, le recadrage, le réglage des couleurs, le réglage de la netteté, ainsi que des modifications plus complexes telles que la

opérations. L'application d'une augmentation pour la classification météorologique et la détection d'objets peut être

opérations. L'application d'une augmentation pour la classification météorologique et la détection d'objets peut être très avantageuse

très avantageux pour les raisons suivantes :

pour les raisons suivantes :

la variabilité des données de formation, facilitant la généralisation du modèle. • Améliorer la diversité et

• Améliorer la diversité et la variabilité des données de formation, facilitant la généralisation du modèle à des scénarios non représentés.

adaptation à des scénarios non représentés.

• Renforcer la résilience du modèle face aux facteurs affectant l'apparence des objets, tels que

• Renforcer la résilience du modèle face aux facteurs affectant l'apparence des objets, tels que les conditions d'éclairage

conditions d'éclairage variables, les occultations ou les changements de point de vue.

variables, occlusions ou changements de point de vue.

• Remédier au déséquilibre des classes en suréchantillonnant ou en sous-échantillonnant les classes minoritaires.

majoritaires.

• Atténuer le surapprentissage en introduisant la régularisation et le bruit dans les données d'entraînement.

Les sections suivantes décrivent les augmentations que nous avons effectuées dans cet article.

• Remédier au déséquilibre des classes en suréchantillonnant ou en sous-échantillonnant les classes minoritaires.
majoritaires.
• Atténuer le surapprentissage en introduisant la régularisation et le bruit dans les données d'entraînement.
Les sections suivantes décrivent les augmentations que nous avons effectuées dans cet article.

6.1. Rotation

Ceci est utilisé pour présenter l'objet sous différents angles de vue. Dans des scénarios réels, les objets peuvent apparaître sous différents angles ou rotations, et l'ajouter à notre augmentation peut aider le modèle à mieux gérer ces variations de vue.

6.2. Teinte

La teinte est une technique d'augmentation d'image basée sur la couleur qui modifie la teinte ou le ton de couleur de l'image.
une image tout en préservant sa luminosité et sa saturation.

6.3. Bruit

Nous avons également incorporé du bruit synthétique dans notre processus d'augmentation pour élargir notre ensemble de données. Ce type d'augmentation améliore la résilience de notre modèle au bruit et améliore sa capacité à s'adapter à de nouvelles données ou scénarios.

6.4. Saturation

La saturation ajuste l'intensité des couleurs dans une image. En saturant une image, nous mettons efficacement à l'échelle les valeurs des pixels selon un facteur aléatoire dans une plage spécifiée. Augmenter la valeur de saturation d'une image peut rendre les couleurs plus vibrantes et plus vives, tandis que la diminuer peut rendre les couleurs plus atténuées et atténuées. Nous avons augmenté la saturation de notre ensemble de données d'environ 25 %.

6.5. Niveaux de gris

Nous avons incorporé l'augmentation des niveaux de gris, qui convertit une image en niveaux de gris. Cette technique est couramment utilisée pour augmenter le contraste d'une image et améliorer ses détails.

6.6. Luminosité

En augmentant de manière aléatoire la luminosité des images, nous avons soumis notre modèle à une gamme plus large de conditions d'éclairage, améliorant ainsi sa résilience aux changements d'éclairage.
Nous avons augmenté les images, les rendant environ 15 % plus lumineuses.

6.7. Se brouiller

Le flou est utilisé pour introduire des effets de flou dans les images. Pour nos données augmentées, nous
utilisé le flou gaussien jusqu'à 1,25 px.

6.8. Exposition

De plus, nous avons modifié artificiellement le niveau d'exposition des images, en le réglant entre 10 % et −10 %.

6.9. Découper

Nous avons également découpé de petites parties d'objets de la scène. Le but est d'ajouter une occlusion à notre expérience, qui consiste à bloquer, couvrir ou masquer un objet de la vue de la caméra.
Le tableau 2 montre nos valeurs de paramètres d'augmentation et leurs impacts sur les images.

Tableau 2. Résumé des augmentations appliquées et de leur impact sur l'image.

Augmentation	Valeur	Impact
Rotation	90 degrés	Aide le modèle à être insensible à l'orientation de la caméra
Teinte	15%	Ajustement aléatoire des couleurs
Bruit	Bruit aléatoire	Plus d'obstacles ajoutés à l'image

Tableau 2. Suite

Augmentation	Valeur	Impact
Saturation	25%	Change l'intensité des pixels
Niveaux de gris	15%	Convertit l'image en canal unique
Luminosité	15%	L'image apparaît plus claire
Se brouiller	1,25px	Fait la moyenne des valeurs de pixels avec celles voisines
Exposition	10%	Résilient aux changements d'éclairage et de réglage de la caméra
Découper	Couper des parties aléatoires de l'image	Plus résistant pour détecter la moitié des objets

7. Expériences et résultats

Pour tester notre modèle, nous avons mené plusieurs expérimentations, en commençant par définir notre Score seuil BRISQUE à 42,75. Ce score est le score de qualité moyen pour le DAWN ensemble de données, et toute image supérieure à ce score moyen passera par l'étape d'amélioration. Le tableau 3 explique la raison pour laquelle nous avons choisi 42,75 comme seuil et illustre le impact de la qualité de l'image en mesurant les scores BRISQUE avant et après augmentation. Le Le tableau compare les images de l'ensemble de données DAWN (tempêtes de sable, pluie, neige et brouillard) avec notre images augmentées étendues du même ensemble de données visant à simuler les conditions météorologiques défavorables conditions. Dans toutes les conditions météorologiques, les images augmentées présentent généralement des Scores BRISQUE, indiquant une baisse de la qualité de l'image par rapport au DAWN original images. Comme le montre le tableau, les images augmentées sont pires d'environ 9 % en ce qui concerne la qualité moyenne de la scène. Cette faible qualité des images augmentées peut être attribuée à les augmentations effectuées (flou, teinte, saturation, bruit, coupure, luminosité et exposition), qui sont des effets habituels en cas de mauvais temps. Les différences observées soulignent importance de concevoir un modèle d'évaluation de la qualité pour préserver la qualité des images, en particulier dans des conditions météorologiques défavorables où la clarté visuelle est cruciale pour une observation précise de la scène et détection d'objets.

Tableau 3. Comparaison de la qualité de l'image avec et sans augmentation. Les résultats montrent un qualité d'image moyenne de 46,59 pour l'ensemble de données DAWN augmenté, contre 42,75 pour l'original Ensemble de données DAWN.

	Sablonneux	Brumeux	Neigeux	Pluvieux	Moyenne
Images de DAWN Niveau de qualité	44.05	45.21	40.18	41.57	42,75
Images DAWN augmentées Niveau de qualité	48.71	49,83	43.19	44.64	46.59

Le scénario expérimental pour l'ensemble de données DAWN augmenté a été exécuté dans le cadre du Environnement Google Colab, exploitant la puissance de calcul d'un GPU Tesla T4. Nous ont apporté plusieurs modifications à l'architecture YOLOv5, dans le but de créer un modèle pour notre domaine. Cette modification inclut le changement des fonctions d'activation et testez le modèle avec les fonctions SiLU et LeakyRelu. Nous avons également modifié le backbone et partez tester les performances des architectures BottleneckCSP et C3. En outre à cela, les hyperparamètres tels que les époques et la taille des lots ont été modifiés tout au long nos expériences.

Après avoir conçu notre modèle, nous avons initié notre phase expérimentale en mettant en œuvre BottleneckCSP comme architecture de base et de tête. Notre modèle a démontré des résultats prometteurs, atteignant une précision moyenne moyenne (mAP) de 55,6 % et 45,6 % lorsque formé pendant 128 époques avec une taille de lot de 32, en utilisant l'activation SiLU et LeakyReLU fonctions, respectivement. Le tableau 4 montre clairement que lorsque nous avons mis en œuvre BottleneckCSP dans notre modèle, le mAP augmentait pour les fonctions SiLU et LeakyRelu chaque fois que nous augmentions le nombre d'époques. On peut également voir que LeakyRelu a une valeur inférieure performances que SiLU avec le squelette et la tête BottleneckCSP. Le tableau 4 montre le résultats complets de notre modèle en utilisant BottleneckCSP.

Tableau 4. Performances de notre modèle utilisant BottleneckCSP comme colonne vertébrale et tête.

Fonction d'activation de la colonne vertébrale et de la tête		Époque	Lot	carte
Goulot d'étranglementCSP	SiLU	32	16	33,7%
Goulot d'étranglementCSP	SiLU	32	32	34,5%
Goulot d'étranglementCSP	SiLU	64	16	40,2%
Goulot d'étranglementCSP	SiLU	64	32	43,9%
Goulot d'étranglementCSP	SiLU	128	16	55,2%
Goulot d'étranglementCSP	SiLU	128	32	55,6%
Goulot d'étranglementCSP	LeakyRelu	32	16	24,0%
Goulot d'étranglementCSP	LeakyRelu	32	32	25,1%
Goulot d'étranglementCSP	LeakyRelu	64	16	34,7%
Goulot d'étranglementCSP	LeakyRelu	64	32	34,9%
Goulot d'étranglementCSP	LeakyRelu	128	16	38,7%
Goulot d'étranglementCSP	LeakyRelu	128	32	45,6%

Nous avons poursuivi nos expérimentations en incluant désormais les solutions concentrées-complètes. Convolution (C3) [31] comme épine dorsale et tête dans notre modèle proposé. Le modèle réalisé un meilleur résultat, atteignant 71,8 % de mAP en utilisant SiLU avec seulement 32 époques et 16 lots, comme Le tableau 5 le montre. Ce score est supérieur à celui de LeakyRelu de 7,4 points de pourcentage, avec le même métrique. Ce résultat est également supérieur au score le plus élevé obtenu grâce au BottleneckCSP. de base (tableau 4), qui était de 55,6 %. Nous avons continué à augmenter le nombre d'époques et des lots jusqu'à ce que nous atteignions 74,6 % après 64 époques avec 16 lots, ce qui est le plus élevé mAP dans cet article. Comme nous le verrons plus loin, ce score est le plus élevé par rapport à d'autres publications récentes sur la détection d'objets qui utilisent DAWN comme base de données. Les mAP, la précision et le rappel de notre modèle sont visibles dans la figure 9. Le haut à gauche graphique représente la précision, avec un score de 85 %, le graphique en haut à droite montre le rappel, qui a atteint 68 %, et le graphique du bas représente la mAP, qui a atteint 74,6 %. La figure 10a montre notre courbe F1, où nous pouvons voir que les pics pour la plupart des classes se produisent à des seuils de confiance compris entre 0,4 et 0,6. Cela suggère que la confiance du modèle dans cette plage offre un meilleur équilibre entre précision et rappel. Cela montre également que le modèle excelle dans la détection des voitures, tout en étant confronté à des difficultés avec les cyclistes, suivant un schéma similaire à celui de la moyenne dans toutes les classes. La figure 10b représente notre mAP, avec une intersection sur union de 0,5 à 0,95.

Électronique 2024, 13, x POUR EXAMEN PAR LES PAIRS 13 sur 21

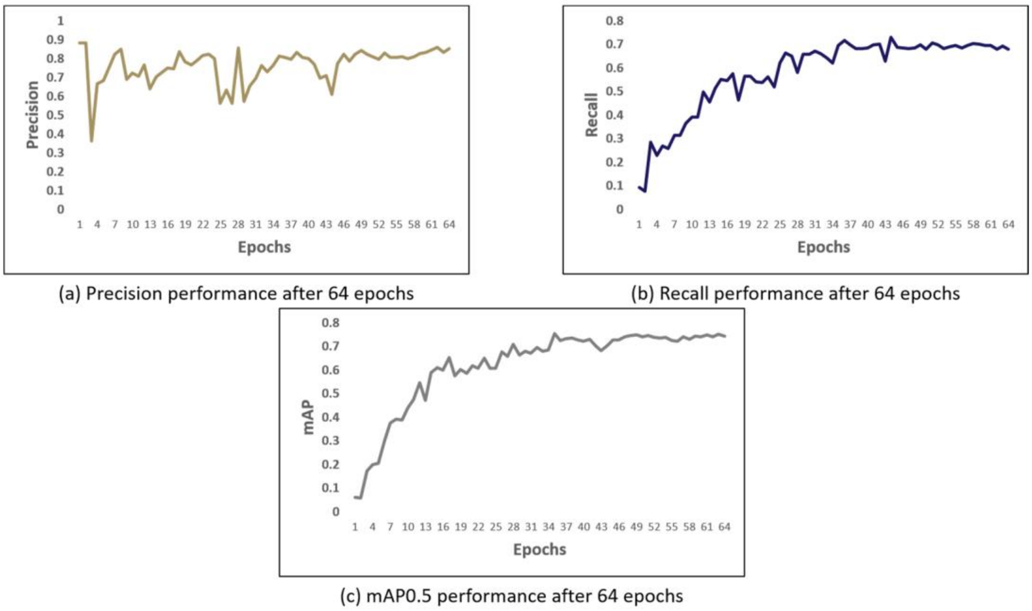


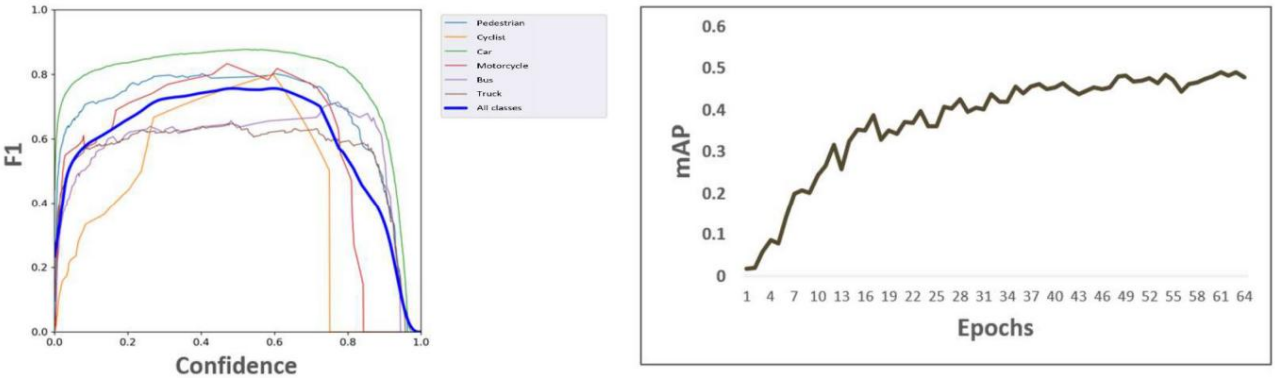
Figure 9. Précision, rappel et mAP à 0,5 de notre modèle après 64 époques.

Tableau 5. Performances de notre modèle utilisant C3 comme colonne vertébrale et tête.

Colonne vertébrale et tête	Fonction d'activation	Époque	Lot	carte
C3	SiLU	32	16	71,8%
C3	SiLU	32	32	68,6%

Tableau 5. Performances de notre modèle utilisant C3 comme colonne vertébrale et tête.

Fonction d'activation de la colonne vertébrale et de la tête		Époque	Lot	carte
C3	SiLU	32	16	71,8%
C3	SiLU	32	32	68,6%
C3	SiLU	64	16	74,6%
C3	SiLU	64	32	74,1%
C3	SiLU	128	16	74,0%
C3	SiLU	128	32	73,1%
C3	LeakyRelu	32	16	64,4%
C3	LeakyRelu	32	32	67,1%
C3	LeakyRelu	64	16	62,9%
C3	LeakyRelu	64	32	63,2%
C3	LeakyRelu	128	16	21 72,9%
C3	LeakyRelu	128	32	72,4%



(a) F1 score (b) mAP 0.5:0.95 performance after 64 epochs

Figure 10. Score F1 et mAP à 0,5 : 0,95 de notre modèle après 64 époques.

Les tableaux mentionnés précédemment, tels que le tableau 4 et 5, mettent en évidence une observation notable : l'augmentation du nombre d'époques n'est pas systématiquement corrélée à une précision moyenne (mAP) plus élevée. Étonnamment, notre modèle a atteint son score mAP le plus élevé lorsqu'il a été entraîné pendant 64 époques au lieu de 128 époques, contrairement aux attentes initiales. Cette découverte met en évidence l'importance d'un réglage minutieux des hyperparamètres et d'une expérimentation de validation pour identifier le programme de formation le plus efficace pour une architecture de modèle et un ensemble de données donnés. Parmi les publications récentes utilisant l'ensemble de données DAWN, la précision moyenne (mAP) de notre méthode de 74,6 % est la plus élevée atteinte, comme détaillé dans le tableau 6.

Tableau 6. Comparaison de nos résultats avec certaines publications récentes utilisant l'ensemble de données DAWN.

Réf. CARTE	Réf. carte	Objectif de l'ensemble de données Base de données	But
[32]	32,75% AUBE 32,75%		Approche d'ensemble pour améliorer la détection d'objets dans les AV dans des conditions météorologiques défavorables DAWN . Approche d'ensemble pour améliorer la détection d'objets dans les AV par mauvais temps conditions.
[33]	Brouillard 29,66% Pluie 41,21% Neige 43,01% Sable 24,13% 24,13%	Brouillard 29,66% Pluie 41,21% AUBE Neige 43,01% Sable	Modification de YOLO et utilisation de plusieurs ensembles de données pour détecter des objets dans l'environnement AV . Modification de YOLO et utilisation de plusieurs ensembles de données pour détecter des objets dans l'environnement AV.
[34]	39,19%	AUBE	Architecture pour améliorer l'apprentissage de l'ensemble de données DAWN à l'aide de GAN et CycleGAN.
[35]	55,85%	AUBE	Transformateur de détection de faible luminosité DAWN (LDETR) 55,85% de faible luminosité (LDETR) pour améliorer les performances de détection.
[36]	72,8% 72,8%	AUBE	DAWN Améliorer YOLO à l'aide d'algorithmes métaheuristiques. Algorithmes métaheuristiques.
74,6%	Les nôtres 74,6% Les nôtres	AUBE	Modification de YOLO utilisant l'ensemble de données DAWN pour classer la météo et détecter des objets dans l'environnement AV. Modification utilisant l'ensemble de données DAWN pour classer la météo et détecter dans l'environnement audiovisuel.

Pour l'évaluation de la classification météorologique, le modèle proposé a obtenu une précision de 74,3 % après 64 époques, comme le montre la figure 11. Le modèle a réussi à classer la plupart des scènes ; cependant, dans certains cas, le modèle n'a pas réussi à classer la météo réelle. Par exemple, si l'on regarde le tableau 7, qui montre le résultat expérimental de la classification météorologique, dans l'image numéro 5, le temps réel était une forte tempête de sable, alors que le modèle classifiait

Pour l'évaluation de la classification météorologique, le modèle proposé a obtenu une précision de 74,3 % après 64 époques, comme le montre la figure 11 . Le modèle a réussi à classer la plupart des scènes ; cependant, dans certains cas, le modèle n'a pas réussi à classer la météo réelle. Pour Par exemple, si l'on regarde le tableau 7, qui montre le résultat expérimental de la classification météorologique, dans l'image numéro 5, le temps réel était une forte tempête de sable, alors que le modèle classifiait c'est comme un temps brumeux. Ce cas est un exemple où la luminosité et l'éclairage du La scène pourrait être difficile pour les modèles de classification météorologique par mauvais temps.

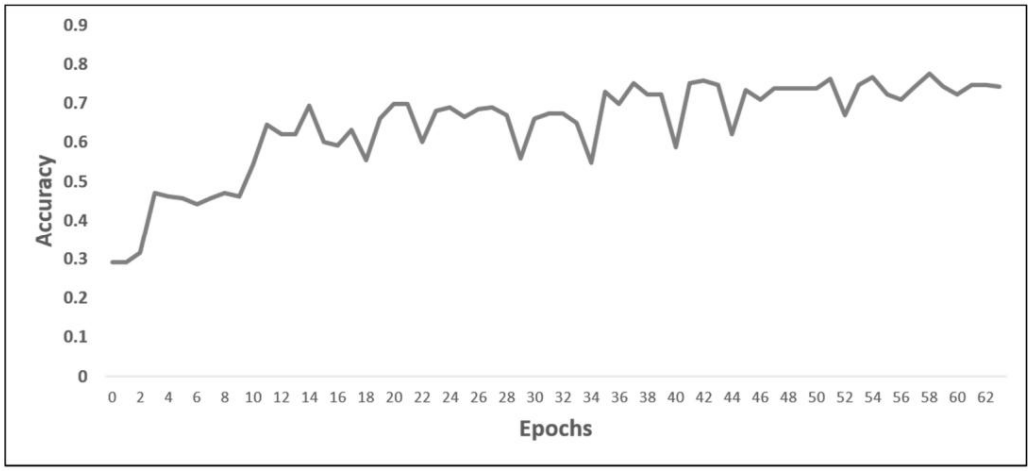


Figure 11. Pour la classification météorologique, notre modèle a obtenu un score de 74,3 % après 64 époques.
Figure 11. Pour la classification météorologique, notre modèle a obtenu un score de 74,3 % après 64 époques.
Figure 11. Pour la classification météorologique, notre modèle a obtenu un score de 74,3 % après 64 époques.
Figure 11. Pour la classification météorologique, notre modèle a obtenu un score de 74,3 % après 64 époques.

Tableau 7. Résultats de classification.
Tableau 7. Résultats de classification.
Tableau 7. Résultats de classification.
Tableau 7. Résultats de classification.

Numéro	Image	Sol	Classifié
1		Temps sablonneux	Temps sablonneux
2		Temps brumeux	Temps brumeux
3		Climat pluvieux	Climat pluvieux

LES PAIRS Tableau 7. Suite.

Nombre	Image	Sol	Classifié
4		Temps enneigé	Temps enneigé
5		Temps sablonneux	Temps brumeux

Les figures 12 à 14 sont des exemples de nos expériences où le modèle a réussi à se classer. Les figures 12 à 14 sont des exemples de nos expériences où le modèle a réussi à classer le type de temps dans la scène et à détecter les objets. Le coin supérieur gauche montre le type de météo dans la scène et les objets détectés. Le coin supérieur gauche affiche les scores de probabilité pour la météo actuelle, en les classant de la probabilité la plus élevée à la faible, les scores de probabilité pour la météo actuelle, en les classant de la probabilité la plus élevée à la plus faible. Par exemple, dans la figure 12, le modèle a classé la scène comme un temps sablonneux à 87 %, ce qui est correct et correspond à la vérité sur le terrain. Il est également détecté avec succès que les deux véhicules sont corrects et correspondent à la vérité terrain. Les deux véhicules apparaissant sur les lieux ont également été détectés avec succès, avec des pourcentages de détection de 96% et 94%, respectivement, apparaissant sur la scène, avec des pourcentages de détection de 96% et 94%, respectivement:

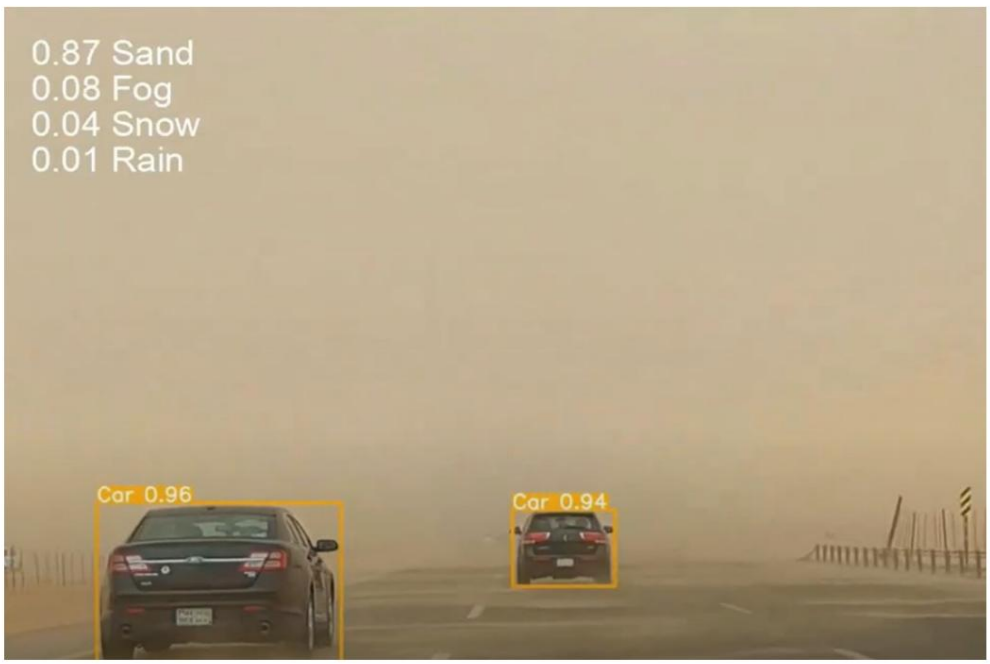


Figure 12. Figure 12. Le modèle a réussi à classer le temps sablonneux et à détecter des objets dans la scène.
Figure 12. Le modèle a réussi à classer le temps sablonneux et à détecter des objets dans la scène.

Électronique 2024, 13, x POUR EXAMEN PAR LES PAIRS 16 sur 20
Électronique 2024, 13, x POUR EXAMEN PAR LES PAIRS 17 sur 21

Figure 13. Le modèle a réussi à classer le temps brumeux et à détecter des objets dans la scène.
Figure 13. Le modèle a réussi à classer le temps brumeux et à détecter des objets dans la scène.

Figure 14. Le modèle a réussi à classer le temps pluvieux et à détecter des objets dans la scène.
Figure 14. Le modèle a réussi à classer le temps pluvieux et à détecter des objets dans la scène.
Figure 14. Le modèle a réussi à classer le temps pluvieux et à détecter des objets dans la scène.

8. Discussion

Bien que la section précédente démontre le potentiel de notre méthode pour détecter diverses véhicules par mauvais temps, dans les points suivants, nous discuterons des informations clés et observations qui ont émergé au cours de l'élaboration de ce travail :

• Si nous regardons notre score F1 (Figure 10), la classe « voiture » obtient systématiquement les scores F1 les plus élevés à différents niveaux de confiance, ce qui indique que le modèle est particulièrement apte à détecter les voitures avec précision. À l'inverse, la classe « camions » présente généralement les scores F1 les plus bas, ce qui suggère que le modèle pourrait avoir plus de difficulté à distinguer les camions ou être confronté à davantage de faux positifs/négatifs dans cette catégorie. La courbe « toutes les classes » représente la performance moyenne de toutes les classes d'objets, démontrant une tendance similaire à celle des classes individuelles, avec un score F1 maximal autour du seuil de confiance de 0,6.

• Comme le montre le tableau 6 , notre travail proposé a atteint un mAP de 74,6 %. Ce résultat surpasse les performances d'autres publications sur l'ensemble de données DAWN, notamment les méthodes d'ensemble [32], les modifications YOLO [33], les architectures basées sur GAN [34], le transformateur LDETR [35] et YOLO amélioré avec des algorithmes métaheuristiques [36] . • Notamment, DAWN est un ensemble de données très complexe, comme le corrobore notre propre expérience et le soulignent les observations des auteurs dans [33], qui ont fait remarquer : « [nous] trouvons l'ensemble de données DAWN un peu plus difficile que les autres. » Ce défi est dû au fait que certaines images et objets se caractérisent par une visibilité extrêmement faible, ce qui est un facteur qui peut avoir un impact sur le score obtenu de tout modèle développé. • Le domaine de la détection d'objets par mauvais temps nécessite encore des ensembles de données plus fiables offrant une variabilité suffisante pour couvrir diverses apparences d'objets, conditions d'éclairage et occlusions. La création de tels ensembles de données prend du temps et coûte cher. Un article récemment publié par Liu et al. [37] ont démontré une approche basée sur un simulateur qui permet une manipulation facile des conditions environnementales, du placement des objets et des perspectives de la caméra. L'utilisation de la collecte de données sur simulateur ouvre la porte à des ensembles de données diversifiés et complets sans collecte approfondie de données réelles. Cette approche peut accélérer la collecte de données en définissant et en exécutant divers scénarios météorologiques défavorables sans, par exemple, attendre les changements saisonniers de la météo. De plus , il offre une évolutivité des données, surmontant les contraintes géographiques de la collecte de données réelles .

• Bien que les publications récentes existantes et les ensembles de données publics offrent des ressources précieuses pour la détection d'objets dans diverses conditions météorologiques, il existe un besoin évident de travaux supplémentaires incluant des scénarios météorologiques sablonneux. • La combinaison d'images avec LiDAR en utilisant la fusion peut être une approche prometteuse pour améliorer la détection d'objets dans les environnements de véhicules autonomes. Des études récentes, comme celles de Dai et al. [38] et Liu et al. [39], ont démontré que cette technique améliore considérablement la détection d'objets dans des environnements difficiles en tirant parti des fonctionnalités complémentaires du LiDAR et des caméras. Les caméras constituent une solution économique et légère qui capture des détails riches en couleurs et en textures, facilitant ainsi la classification et l'identification des objets. D'autre part, le LiDAR offre des mesures de distance précises et des informations spatiales 3D, particulièrement utiles dans des conditions de faible visibilité où les caméras peuvent avoir des difficultés. En fusionnant les données des deux capteurs, la précision et la robustesse des systèmes de détection d'objets peuvent être considérablement améliorées. • Nous avons étendu nos expériences pour tester notre modèle en utilisant l'ensemble de données UAVDT [40]. L' ensemble de données UAVDT original comprend plus de 77 000 images capturées de jour, de nuit et dans des conditions météorologiques brumeuses. Après avoir mené l'expérience pendant 64 époques, nous avons obtenu les résultats suivants : mAP de 94,1 %, rappel de 90,8 % et précision de 97,0 %. Nous pensons que l'ensemble de données UAVDT nécessite un prétraitement supplémentaire avant de pouvoir être pleinement utilisé. Par exemple, ajuster le délai de capture des images pourrait aider à diversifier les images obtenues.

• Les données synthétiques peuvent être utilisées pour relever les défis et les limites des ensembles de données réelles. Dans une publication récente [41], les auteurs ont proposé CrowdSim2, un ensemble de données synthétiques, pour les tâches de détection d'objets, en particulier la détection de personnes et de véhicules. Une telle technique peut être très bénéfique pour le domaine audiovisuel en offrant un environnement contrôlé dans lequel des facteurs tels que les conditions météorologiques, la densité des objets et l'éclairage peuvent être pris en compte.

manipulé avec précision, permettant de tester des modèles de détection d'objets dans divers scénarios. De plus, il peut être utilisé pour simuler des événements rares mais critiques, tels que des accidents ou des obstacles inhabituels, qui peuvent être sous-représentés dans les ensembles de données du monde réel.

9. Conclusions

Classer les conditions météorologiques et détecter des objets dans des environnements météorologiques extrêmes est une tâche critique et difficile. Dans cet article, nous avons présenté un modèle multi-objectifs qui intègre la classification météorologique et la détection d'objets et les traite comme un problème unifié dans le domaine des systèmes de perception des véhicules autonomes. Notre modèle se compose de deux blocs principaux. Tout d'abord, le bloc qualité vérifie la qualité de l'image en fonction du score BRISQUE, et si l'image a un score supérieur au seuil, elle est ensuite améliorée par une méthode SRGAN. Deuxièmement, le bloc Classifier et détecter classe quatre types de conditions météorologiques défavorables (neige, pluie, brouillard et sable) et détecte six classes (voiture, cycliste, piéton, moto, bus et camion). Au cours de notre développement, nous avons utilisé l'ensemble de données complexe DAWN comme source d'images et utilisé YOLO comme structure de base pour la classification et la détection. Les résultats expérimentaux montrent que notre modèle a atteint une précision moyenne (mAP) de 74,6 % pour la détection d'objets en utilisant l' architecture YOLO avec l'architecture C3 comme épine dorsale et SiLU comme fonction d'activation.

De plus, pour classer la météo de la scène, notre modèle a atteint une précision de 74,3 %, ce qui correspond étroitement au mAP. Cela dit, certains défis dans ce domaine doivent encore être pris en compte lors du développement de modèles de détection et de classification. Les modifications des caractéristiques de la scène telles que l'éclairage et la nébulosité conduisent à une mauvaise classification du temps correct.

10. Travaux futurs

Les intempéries restent un domaine très difficile dans les environnements audiovisuels. Pour atteindre le plus haut niveau d'automatisation, les capteurs de caméra ont besoin d'un système robuste, capable de naviguer en toute sécurité dans divers scénarios météorologiques et d'observer avec précision l' environnement. À l'avenir, nous étendrons notre domaine pour inclure des ensembles de données supplémentaires qui pourraient être fusionnés avec l'ensemble de données DAWN actuel. Cela pourrait nous amener à élargir nos classes de détection pour inclure de nouvelles classes plus détaillées que nous observons dans un environnement de conduite réel, comme les feux de circulation, les enfants, les animaux domestiques (comme les chiens) et les forces de l'ordre (comme les policiers). Chacune de ces classes représente des composants à part entière de la scène routière, et il est essentiel de détecter et de réagir avec précision à leur présence pour garantir la sécurité et l'efficacité des systèmes de conduite autonome. En incorporant ces classes supplémentaires dans notre cadre de détection, nous visons à améliorer le mAP global. De plus, nous visons à améliorer les capacités de perception des systèmes autonomes grâce à la fusion perceptuelle, qui implique de combiner les informations provenant de plusieurs capteurs, tels que des caméras, des LiDAR, des radars et des capteurs à ultrasons, pour créer une représentation complète et précise de l'environnement. En développant un système aussi robuste , nous pensons pouvoir atténuer l'impact des conditions météorologiques défavorables sur les performances des capteurs et améliorer la fiabilité et la robustesse des systèmes généraux de perception audiovisuelle.

Contributions de l'auteur : Logiciel NA ; analyse de données, NA, AA et AB ; conservation des données, NA ; méthodologie, NA; rédaction – préparation du projet original, NA ; rédaction – révision et édition, AA et AB ; supervision, AA et AB; conceptualisation, NA, AA et AB Tous les auteurs ont lu et accepté la version publiée du manuscrit.

Financement : Cette recherche n’a reçu aucun financement externe.

Déclaration de disponibilité des données : les données sont contenues dans l'article.

Conflits d'intérêts : Les auteurs ne déclarent aucun conflit d'intérêts.

Les références

1. J3016_202104 ; Taxonomie et définitions des termes liés aux systèmes d'automatisation de la conduite pour les véhicules automobiles routiers. Society of Automotive Engineers (SAE) : Warrendale, PA, États-Unis, 2021.
2. Girshick, R. ; Donahue, J. ; Darrell, T. ; Malik, J. Riches hiérarchies de fonctionnalités pour une détection précise des objets et une segmentation sémantique. Dans Actes de la conférence IEEE sur la vision par ordinateur et la reconnaissance de formes, Columbus, OH, États-Unis, 23-28 juin 2014 ; pp. 580-587.
3. Girshick, R. Fast R-CNN. Dans les actes de la conférence internationale de l'IEEE sur la vision par ordinateur, Séville, Espagne, 17-19 mars 2015 ; pages 1440 à 1448.
4. Ren, S. ; Lui, K. ; Girshick, R. ; Sun, J. Faster R-CNN : Vers une détection d'objets en temps réel avec des réseaux de proposition de région. Av. Informations neuronales . Processus. Système. 2015, 28, 91-99. [Référence croisée] [Pub Med]
5. Lui, K. ; Gkioxari, G. ; Dollarar, P. ; Girshick, R. Masque R-CNN. Dans les actes de la conférence internationale de l'IEEE sur la vision par ordinateur, Jaipur, Inde, 18-21 décembre 2017 ; pages 2961 à 2969.
6. Lin, TY ; Dollarar, P. ; Girshick, R. ; Lui, K. ; Hariharan, B. ; Belongie, S. Présentez des réseaux pyramidaux pour la détection d'objets. Dans Actes de la conférence IEEE sur la vision par ordinateur et la reconnaissance de formes, Penang, Malaisie, 5-8 novembre 2017 ; pp. 2117-2125.
7. Redmon, J. ; Divvala, S. ; Girshick, R. ; Farhadi, A. Vous ne regardez qu'une seule fois : détection d'objets unifiée et en temps réel. Dans Actes des actes de la conférence IEEE sur la vision par ordinateur et la reconnaissance de formes, Bangalore, Inde, 20-21 mai 2016 ; pp. 779-788.
8. Redmon, J. ; Farhadi, A. Yolov3 : Une amélioration progressive. arXiv 2018, arXiv : 1804.02767.
9. Wang, CY ; Bochkovskiy, A. ; Liao, HYM YOLOv7 : un sac de cadeaux entraînables établit un nouvel état de l'art pour les objets en temps réel détecteurs. arXiv 2022, arXiv :2207.02696.
10. Liu, W. ; Anguélov, D. ; Erhan, D. ; Szegedy, C. ; Reed, S. ; Fu, CY ; Berg, AC Ssd : Détecteur multibox monocoup. Dans Actes de la Conférence européenne sur la vision par ordinateur, Amsterdam, Pays-Bas, 11-14 octobre 2016 ; Springer : Berlin/Heidelberg, Allemagne, 2016 ; pp. 21-37.
11. Dhananjaya, MM ; Kumar, VR ; Yogamani, S. Classification des conditions météorologiques et des niveaux de luminosité pour la conduite autonome : ensemble de données, référence et apprentissage actif. Dans les actes de la conférence internationale sur les systèmes de transport intelligents (ITSC) de l'IEEE 2021, Indianapolis, Indiana, États-Unis, du 19 au 22 septembre 2021 ; IEEE : Piscataway, New Jersey, États-Unis, 2021 ; pages 2816 à 2821.
12. Kondapally, M. ; Kumar, KN ; Vishnu, C. ; Mohan, CK Vers une approche de reconnaissance de scènes météorologiques de transition pour Véhicules autonomes. IEEETrans. Intell. Transp. Système. 2023, 25, 5201-5210.
13. Ibrahim, MR ; Haworth, J. ; Cheng, T. WeatherNet : Reconnaître les conditions météorologiques et visuelles à partir d'images au niveau de la rue à l'aide apprentissage résiduel profond. ISPRS Int. J. Géo-Inf. 2019, 8, 549. [Réf. croisée]
14. Xia, J. ; Xuan, D. ; Tan, L. ; Xing, L. ResNet15 : Reconnaissance météo sur route avec réseau neuronal convolutif profond. Av. Météorol. 2020, 2020, 6972826. [Réf. croisée]
15. Lui, K. ; Zhang, X. ; Ren, S. ; Sun, J. Apprentissage résiduel profond pour la reconnaissance d'images. Dans les actes de la conférence IEEE sur Vision par ordinateur et reconnaissance de formes, Bangalore, Inde, 20 et 21 mai 2016 ; pp. 770-778.
16. Cao, Y. ; Li, C. ; Peng, Y. ; Ru, H. MCS-YOLO : Une méthode de détection d'objets multi-échelles pour l'environnement routier de conduite autonome reconnaissance. Accès IEEE 2023, 11, 22342-22354.
17. Liu, Z. ; Lin, Y. ; Cao, Y. ; Hein.; Wei, Y. ; Zhang, Z. ; Lin, S. ; Guo, B. Transformateur Swin : Transformateur de vision hiérarchique utilisant des fenêtres décalées. Dans Actes de la conférence internationale IEEE/CVF sur la vision par ordinateur, Montréal, Colombie-Britannique, Canada, 11-17 octobre 2021 ; pages 10012 à 10022.
18. Poêle, G. ; Fu, L. ; Yu, R. ; Muresan, MI Reconnaissance de l'état de la surface des routes d'hiver à l'aide d'un réseau neuronal à convolution profonde pré-entraîné. Dans les actes de la 97e réunion annuelle du Transportation Research Board, Washington, DC, États-Unis, 7-11 janvier 2018 ; pp. 838-855.
19. Wang, R. ; Zhao, H. ; Xu, Z. ; Ding, Y. ; Li, G. ; Zhang, Y. ; Li, H. Détection de cibles de véhicules en temps réel dans des conditions météorologiques défavorables basé sur YOLOv4. Devant. Neurorobotique 2023, 17, 34.
20. Li, X. ; Wu, J. Extraction de données de mouvement de véhicule de haute précision à partir d'une vidéo de véhicule aérien sans pilote capturée dans diverses conditions météorologiques. Remote Sens.2022 , 14, 5513. [CrossRef]
21. Liu, W. ; Ren, G. ; Yu, R. ; Guo, S. ; Zhu, J. ; Zhang, L. YOLO adaptatif à l'image pour la détection d'objets dans des conditions météorologiques défavorables. Dans Actes de la conférence AAAI sur l'intelligence artificielle, Philadelphie, PA, États-Unis, 27 février-2 mars 2022 ; Volume 36, pages 1792-1800.
22. Huang, Caroline du Sud ; Le, TH; Jaw, DW DSNet : Apprentissage sémantique conjoint pour la détection d'objets dans des conditions météorologiques défavorables. IEEETrans. Modèle Anal. Mach. Intell. 2020, 43, 2623-2633.
23. Sharma, T. ; Débaque, B. ; Duclos, N. ; Chehri, A. ; Kinder, B. ; Fortier, P. Détection d'objets et perception de scènes basées sur l'apprentissage profond dans de mauvaises conditions météorologiques. Électronique 2022, 11, 563. [CrossRef]
24. Jung, Hong Kong ; Choi, GS Yolov5 amélioré : détection efficace d'objets à l'aide d'images de drones dans diverses conditions. Appl. Sci. 2022, 12, 7255. [Réf. croisée]
25. ABDULGHANI, AMA; DALVEREN, Détection d'objets en mouvement GGM en vidéo avec les algorithmes YOLO et R-CNN plus rapide dans différentes conditions. Avrupa Bilim Ve Teknoloji Dergisi 2022, 33, 40-54. [Référence croisée]
26. Kenk, MA; Hassaballah, M. DAWN : Détection de véhicules dans un ensemble de données sur la nature des conditions météorologiques défavorables. arXiv 2020, arXiv :2008.05402.
27. Mittal , A. ; Moorthy, AK ; Bovik, AC Évaluation de la qualité des images sans référence dans le domaine spatial. IEEETrans. Processus d'images. 2012, 21, 4695-4708. [Référence croisée] [Pub Med]

28. Ledig, C. ; Théis, L. ; Huszár, F. ; Caballero, J. ; Cunningham, A. ; Acosta, A. ; Aitken, A. ; Tejani, A. ; Totz, J. ; Wang, Z. ; et coll.
Super-résolution d'image unique photo-réaliste utilisant un réseau contradictoire génératif. Dans Actes de la conférence IEEE sur la vision par ordinateur et la reconnaissance de formes, Penang, Malaisie, 5-8 novembre 2017 ; pages 4681 à 4690.
29. Zoph, B. ; Cubuk, ED; Ghiasi, G. ; Lin, TY ; Shlens, J. ; Le, QV Apprentissage des stratégies d'augmentation des données pour la détection d'objets.
Dans Actes de Computer Vision–ECCV 2020 : 16e Conférence européenne, Glasgow, Royaume-Uni, 23-28 août 2020 ; Springer : Berlin/Heidelberg, Allemagne, 2020 ; pp. 566-583, partie XXVII 16.
30. Volk, G. ; Müller, S. ; Von Bernuth, A. ; Hospach, D. ; Bringmann, O. Vers une détection d'objets robuste basée sur CNN grâce à une augmentation avec des variations de pluie synthétiques. Dans les actes de la conférence IEEE sur les systèmes de transport intelligents (ITSC) 2019 ; IEEE : Piscataway, New Jersey, États-Unis, 2019 ; pp. 285-292.
31. Parc, H. ; Yoo, Y. ; SEO, G. ; Main.; Yun, S. ; Kwak, N. C3 : Convolution concentrée-complète et son application à la sémantique
segmentation. arXiv 2018, arXiv :1812.04920.
32. Walambé, R. ; Marathe, A. ; Kotecha, K. ; Ghinea, G. Cadre d'ensemble de détection d'objets légers pour les véhicules autonomes en
conditions météorologiques difficiles. Calculer. Intell. Neurosci. 2021, 2021, 5278820. [CrossRef] [Pub Med]
33. Farid, A. ; Hussein, F. ; Khan, K. ; Shahzad, M. ; Khan, U. ; Mahmood, Z. Une méthode de détection de véhicules en temps réel rapide et précise
utiliser l'apprentissage profond pour des environnements sans contraintes. Appl. Sci. 2023, 13, 3059. [Réf. croisée]
34. Musat, V. ; Fursa, moi; Newman, P. ; Cuzzolin, F. ; Bradley, A. Ville multi-météo : empilement de conditions météorologiques défavorables pour la conduite autonome.
Dans Actes des actes de la conférence internationale IEEE/CVF sur la vision par ordinateur, Montréal, Colombie-Britannique, Canada, 11-17 octobre 2021 ; pages 2906 à 2915.
35. Tiwari, AK ; Pattanaik, M. ; Sharma, G. Low-light DEtection TRansformer (LDETR) : Détection d'objets dans des conditions de faible luminosité et défavorables
conditions météorologiques. Multimed. Outils Appl. 2024, 1-18. [Référence croisée]
36. Özcan, I. ; Altun, Y. ; Parlak, C. Améliorer les performances de détection YOLO des véhicules autonomes dans des conditions météorologiques défavorables
Utilisation d'algorithmes métaheuristiques. Appl. Sci. 2024, 14, 5841. [Réf. croisée]
37. Liu, S. ; Zhang, H. ; Qi, Y. ; Wang, P. ; Zhang, Y. ; Wu, Q. Aerialvln : Navigation visuelle et linguistique pour les drones. Dans Actes des actes de la conférence internationale IEEE/CVF sur la vision par ordinateur, Paris, France, 2-6 octobre 2023 ; pages 15384 à 15394.
38. Dai, Z. ; Guan, Z. ; Chen, Q. ; Xu, Y. ; Sun, F. Détection améliorée d'objets dans les véhicules autonomes grâce au LiDAR—Capteur de caméra
La fusion. Monde Electr. Véh. J. 2024, 15, 297. [CrossRef]
39. Liu, H. ; Wu, C. ; Wang, H. Détection d'objets en temps réel à l'aide du LiDAR et de la fusion de caméras pour la conduite autonome. Sci. Rep.2023 , 13, 8056. [CrossRef] [Pub Med]
40. Du, D. ; Qi, Y. ; Yu, H. ; Yang, Y. ; Duan, K. ; Li, G. ; Zhang, W. ; Huang, Q. ; Tian, Q. La référence des véhicules aériens sans pilote : détection et suivi d'objets. Dans Actes de la conférence européenne sur la vision par ordinateur (ECCV), Munich, Allemagne, 8-14 septembre 2018 ; pp. 370-386.
41. Foszner, P. ; Szczesna, A. ; Ciampi, L. ; Messine, N. ; Cygan, A. ; Bizon', B. ; Cogiel, M. ; Golba, D. ; Macioszek, E. ; Staniszewski, M.
CrowdSim2 : Un benchmark synthétique ouvert pour les détecteurs d'objets. arXiv 2023, arXiv :2304.05090.

électronique

MDPI

Article

Impact de la surcharge du lien de communication sur le flux de puissance et Transmission de données dans les systèmes électriques cyber-physiques

Xinyu Liu 1,2,3, Yan Li 1,2,3* et Tianqi Xu 1,2,3

1 École d'électricité et de technologie de l'information, Université Yunnan Minzu, Kunming 650504, Chine ; xyliu7918@gmail.com (XL); xu.tianqi@ymu.edu.cn (TX)
2 Laboratoire clé du système autonome sans pilote du Yunnan, Kunming 650504, Chine
3 Laboratoire clé du système d'alimentation cyber-physique des collèges et universités du Yunnan, Kunming 650504, Chine *
Correspondance : yan.li@ymu.edu.cn

Résumé : Le volume de la demande de flux dans les systèmes électriques cyber-physiques (CPPS) fluctue de manière inégale à travers les réseaux couplés et est susceptible de connaître une congestion ou une surcharge en raison de la demande énergétique des consommateurs ou de catastrophes extrêmes. Par conséquent, compte tenu de l'élasticité des réseaux réels, les liens de communication avec un flux d'informations excessif ne se déconnectent pas immédiatement mais présentent un certain degré de redondance. Cet article propose un modèle itératif de défaillance en cascade dynamique basé sur la distribution de la surcharge du flux d'informations dans un réseau de communication et l'interdépendance du flux d'énergie dans le réseau électrique physique. Tout d'abord, un modèle de capacité de charge non linéaire d'un réseau de communication avec surcharge et bords pondérés est introduit, prenant pleinement en compte les trois états de liaison : normal, échec et surcharge. Ensuite, les flux intermédiaires remplacent les flux de branchement dans le réseau électrique physique, et le flux d'énergie sur les lignes défaillantes est redistribué à l'aide du modèle de capacité de charge, simplifiant ainsi les calculs. Troisièmement, sous l'influence des relations de couplage, un modèle complet basé sur une théorie améliorée de la percolation est construit, avec des stratégies d'attaque formulées pour évaluer plus précisément les réseaux couplés. Les simulations sur le système de bus IEEE-39 démontrent que la prise en compte de la capacité de surcharge des liens de communication à petite échelle améliore la robustesse des réseaux couplés. De plus, les attaques de liens délibérées provoquent des dégâts plus rapides et plus étendus que les attaques aléatoires.

Citation : Liu, X. ; Li, Y. ; Xu, T.Impact de la surcharge du lien de communication sur Flux de puissance et transmission de données dans Systèmes d'alimentation cyber-physiques. Électronique 2024, 13, 3065. https://doi.org/10.3390/electronics13153065

Rédacteur académique : Christos J. Bouras

Reçu : 8 juillet 2024
Révisé : 30 juillet 2024
Accepté : 31 juillet 2024
Publié : 2 août 2024

Mots-clés : bords surchargés ; flux d'information; flux de puissance ; échecs en cascade ; théorie améliorée de la percolation ; systèmes d'alimentation cyber-physiques

1. Introduction

1.1. Contexte

Avec le développement du réseau intelligent et de l'Internet énergétique, le système électrique est devenu profondément couplé au système d'information. Le système d'alimentation physique et le système de communication du côté information ont progressivement évolué vers le système d'alimentation cyber-physique (CPPS) [1]. Si le système couplé a apporté de nombreux avantages, il a également accru le risque de pannes en cascade dans l'espace. Les vulnérabilités dans les deux systèmes via un réseau qui se chevauchent augmenteront le risque de propagation de pannes, de sorte que même une seule défaillance d'un bord ou d'un nœud peut avoir un impact sur l'ensemble du réseau, conduisant souvent à un effondrement global [2,3]. Par exemple, la panne massive dans l'ouest des États-Unis en 2003, la panne en Ukraine en 2015 et la panne de 815 heures au Brésil en 2023 [4,5] ont toutes été causées par la défaillance de certaines périphéries du réseau d'information. Ces pannes se sont propagées au réseau électrique par couplage fonctionnel, aboutissant finalement à la paralysie simultanée des deux systèmes.

1.2. Travaux connexes

L'analyse des cas passés montre que lorsque le système de communication tombe en panne ou est attaqué, des paquets de données peuvent être perdus ou manipulés, empêchant ainsi un contrôle en boucle fermée. En raison de la connexion de couplage cyber-physique, la panne se propagera et affectera les nœuds de puissance du réseau physique avec une certaine probabilité. Ensuite, la panne continue de se propager dans le réseau physique, causant finalement de graves dommages au système [6-8]. Par conséquent, la modélisation du réseau électrique physique, du réseau de communication et de sa connexion de couplage est essentielle pour comprendre le processus de propagation des défaillances en cascade dans l'espace.

Dans [9], un modèle de couplage un-à-un entre les nœuds de puissance et de communication a été proposé, analysant la robustesse du système de défaillance en cascade après la suppression d'une petite fraction de nœuds sur la base de son modèle topologique. Dans [10], les auteurs ont étudié la robustesse d'un réseau de communication à double couche sans échelle basé sur la théorie de la percolation. Basé sur [10], Réf. [11] ont considéré les interactions entre les nœuds de différentes couches comme hétérogènes, étudiant un type de dynamique en cascade dans les réseaux à double couche qui présentent à la fois interdépendance et connectivité. Sur la base de [9,10], Chen et al. [12] ont différencié les nœuds du réseau électrique physique en nœuds de générateur et de charge, proposant un nouveau mécanisme interactif pour les pannes en cascade. Dans [13], les caractéristiques opérationnelles et la structure topologique du réseau de transmission ont été intégrées pour établir un modèle de défaillance en cascade pour les défauts aléatoires dans les lignes de transmission sous différentes stratégies de couplage, visant à obtenir un réseau couplé robuste et optimal. Cependant, les modèles de couplage établis dans ces études se concentrent uniquement sur les structures topologiques, négligeant les caractéristiques opérationnelles des deux côtés des réseaux couplés. Dans [14,15], l'optimisation du flux de puissance dans le réseau électrique physique a été prise en compte et les résultats de vulnérabilité sous différentes stratégies et topologies de réseau d'information ont été comparés, mais les caractéristiques opérationnelles du réseau d'information n'ont pas été prises en compte. Dans [16], la propagation dynamique des défaillances en cascade entre le réseau électrique et le réseau de communication a été étudiée, en considérant les caractéristiques du flux d'énergie et du flux de données dans deux systèmes différents, mais l'impact de la surcharge de données dans le réseau de communication sur le réseau couplé n'a pas été prise en compte.

Dans [17], les caractéristiques de récupération de différentes forces de couplage et topologies de réseau basées sur un modèle en cascade lié à la charge ont été étudiées. Bien que l'état de surcharge des nœuds ait été pris en compte dans ce modèle, la charge supplémentaire n'a pas été redistribuée. Sur la base de [16,17], Ding et al. [18] ont proposé un modèle amélioré de défaillances en cascade. Ce modèle prend en compte l'état de surcharge et le processus de récupération des cyber-nœuds, ainsi que l'optimisation du flux d'énergie dans la couche physique et la redistribution du flux d'informations lors de la propagation des pannes. Sur la base de cela, Wang et al. [19] ont utilisé un modèle de flux de puissance CA pour caractériser les caractéristiques opérationnelles du réseau électrique, améliorant ainsi la précision du modèle de réseau électrique. Simultanément, il a construit un réseau de communication pondéré avec des centres de contrôle et appliqué un modèle de redistribution des flux. Dans [20], les auteurs ont proposé deux types de modèles de dépendance forte et faible et ont analysé les changements de robustesse du réseau de couplage en utilisant un schéma d'équilibrage de charge tenant compte de la congestion sous des fautes aléatoires initiales dans la couche d'alimentation. Cependant, les flux de données dans la couche de communication n'ont pas été pris en compte. Les auteurs de [21,22] ont considéré les défaillances des nœuds de communication et ont établi un modèle de défaillance en cascade amélioré basé sur la répartition de la charge de la couche physique. Dans [23], le modèle proposé par les auteurs considère les différences pratiques entre un réseau de communication et un réseau électrique en termes de structure du réseau, de fonctionnement physique et de comportement dynamique, en se concentrant sur l'analyse des défauts survenant du côté du réseau électrique. Il ressort clairement de ces études que la plupart des chercheurs ont accordé moins d'attention aux retards de transmission causés par les surcharges de trafic dans les réseaux de communication et à l'établissement de modèles couplés intégrant les caractéristiques opérationnelles des deux réseaux.

1.3. Motivation

En fait, de nombreux dispositifs Edge connectés possèdent souvent une capacité redondante. Par exemple, la figure 1a illustre un réseau de communication à cinq nœuds. La matrice F représente la matrice de demande de transmission du flux d'information, où les éléments Fij indiquent le flux d'information

demande qui doit être transmise de la source à la destination. Chaque lien eij est associé à son attribut de qualité qij [24,25], où le niveau opérationnel d'un lien eij est défini comme le rapport entre la capacité du lien et la charge du lien. En supposant que le seuil de défaillance du lien est ρ, si les défauts initiaux dans le réseau de communication sont des liens avec ρ < 0,5, ces liens seront supprimés du réseau d'origine (par exemple, supprimer 1 → 2, 2 → 3). A ce stade, la transmission du flux d'informations sur le lien de communication n'est pas affectée (les liens 1→4→2 et 2→1→3 existent toujours), mais les liens 1→4 et 2→1 sont dans un état de surcharge. La demande de flux d'informations sur le lien d'origine 1→2 sera redistribuée vers le lien 1→4→2. Même si le réseau n'est pas immédiatement affecté et que les liaisons surchargées ne tombent pas en panne, la qualité de la transmission va continuer à diminuer. Lorsque le seuil est augmenté à ρ = 0,7, l' efficacité de fonctionnement du lien 1 → 4 tombe en dessous du seuil critique, ce qui fait passer l'état du dispositif Edge de surchargé à défaillant, et il est supprimé du réseau d'origine. À ce stade, 25 demandes de transmission de flux d'informations sont concernées (surlignées en rouge sur la figure).

Lorsque le seuil est encore augmenté jusqu'à ρ = 0,9, les liens avec qij < 0,9 sont supprimés. Comme le montre la figure 1d, seules onze unités de demande de trafic peuvent être efficacement transmises au centre de contrôle. On peut constater que lors du processus de modification du seuil et de suppression des liaisons, si l'état de surcharge et la demande de flux de transmission des liaisons ne sont pas pris en compte, le réseau s'effondrera prématurément, entraînant des pertes importantes dans le réseau électrique.

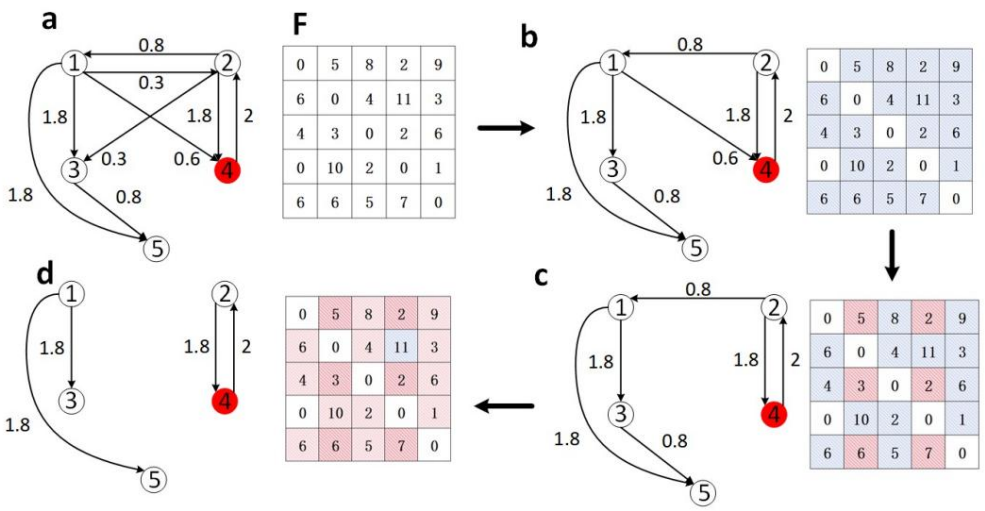


Figure 1. Schéma du réseau de communication électrique d'une défaillance de liaison basé sur la demande de flux d'informations. (a) Réseau de communication G de taille n = 5, où le nœud quatre est le centre de contrôle et les autres sont des nœuds de transmission réguliers. La qualité qij de chaque ligne eij est indiquée à côté du lien. La matrice F représente la demande de transmission du flux d'informations. (b) En supposant que les défauts initiaux dans le réseau de communication sont des liaisons avec ρ < 0,5. (c) Augmenter le seuil à ρ = 0,7, tout en supprimant les liens avec qij < 0,7. (d) Lorsque le seuil est encore augmenté jusqu'à ρ = 0,9, les liaisons avec qij < 0,9 sont supprimées et seules onze unités de demande de trafic peuvent être effectivement transmises au réseau. centre de contrôle.

Généralement, lorsqu'une panne N-1 survient dans le réseau électrique, la convergence des flux doit être recalculée pour chaque scénario. Si le flux converge, il faut déterminer si le flux sur chaque ligne dépasse ses limites ; si tel est le cas, la ligne concernée doit être coupée. S'il ne converge pas, un délestage de charge ou des ajustements de sortie du générateur sont généralement mis en œuvre pour équilibrer les flux d'énergie. Énumérer tous les scénarios peut être à la fois complexe et prendre beaucoup de temps Par conséquent, cet article propose d'appliquer l'interdépendance du flux de ligne [26] au modèle charge-capacité du système électrique, en l'utilisant comme charge électrique sur la ligne. Cette approche reflète et quantifie efficacement le rôle des lignes dans la transmission de la puissance des générateurs aux charges, en tenant également compte de l'impact de la puissance de transport maximale disponible entre les générateurs et les charges sur les lignes critiques. Ce contexte physique s'aligne plus étroitement avec

le fonctionnement réel des systèmes électriques et reflète mieux leurs caractéristiques opérationnelles. Lors de la modélisation du processus de transmission du flux d'informations, il est crucial de le distinguer du réseau électrique. Les liaisons de la couche d'informations ne se déconnecteront pas en raison d'une surcharge du flux d'informations mais provoqueront une congestion si le flux est excessif. Il est nécessaire de prendre en compte l'état de surcharge lors de la redistribution du flux d'informations des liaisons non opérationnelles vers les liaisons voisines afin de développer un modèle de flux de réseau puissance-communication plus réaliste . Par conséquent, cet article, considérant les interruptions de liaison dans le système de communication, propose un modèle de défaillance en cascade charge-capacité basé sur une théorie améliorée de la percolation, intégrant la surcharge du flux d'informations et les caractéristiques opérationnelles du système électrique. Ce modèle analyse l'impact des liens surchargés sur la robustesse du système

1.4. Contributions

Les contributions de cet article sont résumées comme suit :

- Compte tenu de l'état de congestion des liaisons du réseau de communication, un modèle d'allocation de transmission dynamique pour le flux d'informations sous trois états de liaison (normal, surcharge et défaillance) est établi. Des mesures d'intégrité topologique et de caractéristiques opérationnelles du système de communication sont proposées pour évaluer la vulnérabilité du système en cas de pannes.

- Considérant les caractéristiques de génération de topologie des réseaux de communication d'énergie réels , une théorie améliorée de la percolation est proposée. Les nœuds de communication qui se trouvent initialement à l'extérieur du plus grand composant connecté mais qui disposent de liens de communication avec le centre de contrôle sont considérés comme des nœuds efficaces. Les nœuds de puissance qui perdent leur couplage avec le réseau de communication mais restent cohérents sont également considérés comme des nœuds efficaces.

- Compte tenu des caractéristiques physiques des réseaux couplés, l'indicateur d'interconnexion des flux de ligne est utilisé pour mesurer les caractéristiques électriques des flux d'énergie de ligne. Un modèle de distribution charge-capacité basé sur l'interdépendance des flux est proposé pour transférer la charge des lignes défaillantes.

Le reste du document est organisé comme suit. Dans la section 2, nous construisons le modèle de dépendance unidirectionnel du CPPS et fournissons une description détaillée du processus de propagation des fautes à travers l'espace en tenant compte des conditions de surcharge des liaisons de communication dans le modèle couplé. Dans la section 3, nous proposons deux mesures d' évaluation de la robustesse du système pour mesurer les résultats des défaillances en cascade. Dans la section 4, la simulation numérique est effectuée pour analyser la défaillance en cascade et présente les discussions connexes. La section 5 conclut le document.

2. Méthodes

2.1. Modélisation des réseaux couplés puissance-communication basée sur une dépendance unidirectionnelle

CPPS comprend la modélisation du réseau d'information, du réseau électrique et de la couche de couplage. Le réseau d'informations comprend la couche d'accès, la couche principale et la couche centrale. Les bords interdépendants entre les nœuds de puissance et les nœuds d'information forment le réseau de couplage [27]. Dans le réseau de couplage, le réseau d'alimentation fournit un support d'alimentation au réseau de communication, tandis que le réseau de communication offre un support 3C au réseau d'alimentation. Cependant, étant donné que les nœuds de communication sont largement équipés de sources d'alimentation de secours, la défaillance des nœuds d'alimentation couplés n'entraîne pas la défaillance des nœuds de communication due à une panne de courant. Le fonctionnement normal des nœuds de puissance repose sur la surveillance et le contrôle des nœuds de communication. La défaillance des liaisons de communication empêchera le centre de contrôle de recevoir les informations sur les pannes du système électrique en temps opportun, ce qui entraînera l'incapacité de traiter rapidement les pannes de courant et élargira la portée de l'impact des pannes. Par conséquent, cet article étudie principalement le modèle de dépendance du réseau électrique vis-à-vis du réseau d'information. En référence à l'organigramme de la figure 2, le processus de modélisation est détaillé comme suit :

(1) Le réseau électrique physique est résumé sous la forme d'un graphe Gp(Vp, Ep) composé de nœuds électriques Vp et de lignes de transmission Ep. L'ensemble Vp comprend les équipements physiques du réseau électrique tels que les centrales électriques, les sous-stations et les charges.

(2) Divisez le réseau électrique en régions [28].

un. Le cloisonnement des communautés s'applique à la division des régions du réseau électrique. Dans le réseau électrique, plus la distance entre deux bus est proche, plus la réactance de ligne est petite et plus sa réciproque est grande (c'est-à-dire un poids plus élevé), indiquant une plus grande intimité entre la paire de nœuds. Ainsi, ces nœuds sont plus susceptibles d'être partitionnés dans la même région. L'inverse de la réactance de ligne est utilisée comme poids pour les éléments non nuls de la matrice de contiguïté du réseau électrique E. b. En utilisant la méthode Fast Newman, la matrice E modifiée à partir des étapes ci-dessus est substituée à la matrice E d'origine pour calculer la modularité Q pour le partitionnement initial du réseau électrique. Le processus de partitionnement doit satisfaire aux conditions suivantes : chaque région doit contenir au moins un générateur et une charge ; le nombre de régions doit être inférieur au minimum du nombre de générateurs et de charges ; et les sous-régions doivent parvenir à un équilibre des pouvoirs. c. Pour parvenir à un équilibre de puissance au sein des régions, les flux de puissance du système sur les lignes sont utilisés comme valeurs de pondération basées sur l'algorithme Prim. Ces valeurs de poids sont utilisées pour déterminer les chemins d'écoulement des générateurs aux charges. Les régions sont ensuite fusionnées selon ces chemins et combinées avec les résultats de partitionnement initiaux pour former les zones finales.

(3) Le nombre de nœuds pour chaque couche est déterminé : la couche d'accès, la couche de base et la couche centrale. Les nœuds de couche d'accès sont des nœuds de collecte d'informations, avec le nombre de nœuds de communication égal au nombre de nœuds de puissance ; les nœuds de la couche dorsale sont des nœuds d'équipement de routage, également considérés comme des nœuds de contrôle, avec un nombre de nœuds égal au nombre de partitions du réseau électrique ; et les nœuds de couche centrale sont deux nœuds de contrôle représentant la répartition principale et de secours.

(4) La modélisation du réseau d'information : a. La connexion entre le réseau électrique et la couche d'accès : selon le schéma abstrait du réseau électrique Gp, les nœuds de la couche d'accès sont connectés de manière correspondante un à un pour rendre le diagramme de topologie de la couche d'accès cohérent avec le diagramme de topologie. du réseau électrique. b. La connexion entre la couche d'accès et la couche de base : les nœuds du plus haut degré dans chaque région de la couche d'accès et les nœuds générateurs sont connectés aux nœuds correspondants de la couche de base. c. Les nœuds de la couche dorsale sont connectés en interne en fonction de la connexion relation entre les cloisons du réseau électrique.

d. La connexion entre la couche d'accès et la couche centrale : calculez la proportion qi de la sortie du générateur Gi par rapport à la sortie totale du système, et connectez-la aux appels principaux et de secours avec une probabilité de qi .

e. Tous les nœuds de routage sont connectés aux nœuds du centre de planification principal et de secours. F. Les centres de planification principal et de secours sont connectés.

(5) Extraire le réseau d'information pour former Gc.

(6) Connectez les nœuds de la couche d'accès aux nœuds de puissance pour former des bords dépendants unidirectionnels.

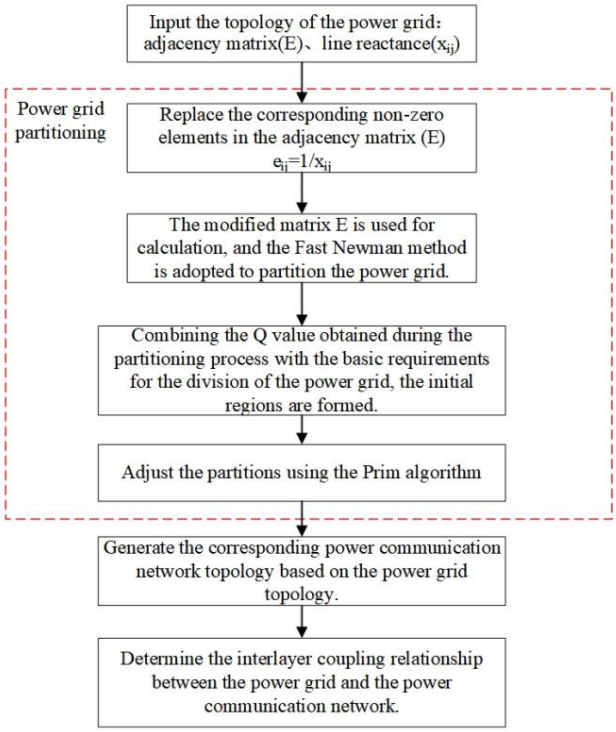


Figure 2. Organigramme de la modélisation du système électrique cyber-physique.

Ainsi, les réseaux couplés sont représentés sur la figure 3.

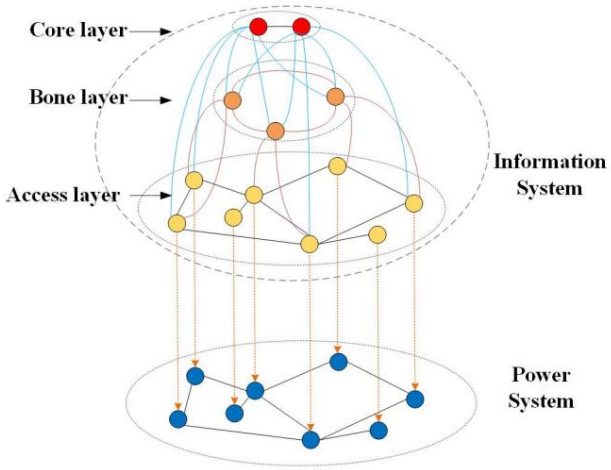


Figure 3. Schéma de topologie du réseau électrique dépendant du réseau d'information.

2.2. Modèle de défaillance en cascade basé sur la théorie de la percolation prenant en compte les surcharges des liaisons de communication

2.2.1. Modèle de flux non linéaire pour les liens

Le modèle de flux d'informations de liaison du réseau de communication utilise un modèle capacité-charge non linéaire qui prend en compte l'état de surcharge [29]. Dans les réseaux complexes, un degré de nœud plus élevé indique davantage de connexions avec d'autres nœuds. Par conséquent, les valeurs de degré des points d'extrémité sont utilisées pour calculer le flux d'informations de chaque liaison de couche d'accès. Plus la valeur du degré d'un lien est élevée, plus l'information qui le traverse est importante. Le taux de flux d'informations d'un lien est utilisé comme poids du lien, et le coefficient de surcharge δ décrit la capacité du bord à gérer un flux d'informations supplémentaire, comme suit :

$$C_{c,ij} = L_{c,ij} + \beta L_{c,ij}^{\alpha} \quad (1)$$

$$C_{c,ij}^{maximum} = \delta C_{c,ij} \quad (2)$$

$$W_{c,ij} = \begin{cases} \dfrac{L_{c,ij}}{C_{c,ij}} & e_{ij} \in EC \\ 0 & e_{ij} \notin EC \end{cases} \quad (3)$$

où

$$L_{c,ij} = w_{ij} = k_i k_j^{\theta} \quad (4)$$

où Lc,ij représente la quantité d'informations transmises par le lien et ki et kj désignent respectivement les degrés des nœuds i et j. θ est un paramètre qui ajuste le flux d'informations.

Cc,ij est la capacité du lien, C $^{maximum}_{c,ij}$ représente le débit maximum que peut supporter la ligne, et Wc,ij est le poids de l'arête eij. α et β sont des coefficients de capacité. Lorsque α = 1, le modèle est linéaire.

2.2.2. Théorie de la percolation améliorée La

théorie traditionnelle de la percolation [30] est largement utilisée pour décrire la structure, la fonction et la résilience des systèmes de réseau. Les modèles de percolation simulent des scénarios de défaillance de liaison en supprimant progressivement les liaisons du réseau. Au fur et à mesure que les liens sont progressivement supprimés, la réduction de la taille du plus grand sous-graphe connecté peut être utilisée pour mesurer les conséquences des défaillances de liens. Ainsi, la théorie de la percolation est applicable à la modélisation des défaillances en cascade dans les systèmes électriques cyber-physiques. Cependant, la théorie traditionnelle de la percolation ne prend en compte que le plus grand sous-ensemble connecté dans un réseau monocouche lors de la détermination du sous-ensemble de nœuds de travail, sans envisager d'autres scénarios. Son application directe pour modéliser les défaillances en cascade dans les systèmes électriques cyber-physiques rend difficile la simulation précise du processus de défaillance. Compte tenu des caractéristiques de transmission du flux d'informations du réseau d'information, où la couche centrale et la couche principale ne correspondent pas directement à la couche d'accès, et où l'état de la liaison prend en compte la situation de surcharge, il est nécessaire d'améliorer le modèle classique de la théorie de la percolation. Le modèle de défaillance du système d'alimentation cyber-physique établi dans cet article est le suivant :

- Système de communication. Pour les liens de la couche d'accès, les arêtes dont le poids est supérieur au coefficient de surcharge sont considérées comme ayant échoué. Pour les nœuds d'information de la couche d'accès, les nœuds qui ne peuvent pas établir un chemin vers les nœuds de contrôle sont considérés comme ayant échoué.
- Système du pouvoir. Un nœud de charge de puissance doit être connecté à au moins un nœud de générateur ; sinon, il est considéré comme un échec. De même, un nœud générateur doit être connecté à au moins un nœud de charge électrique ; sinon, il est considéré comme un échec. Les nœuds fonctionnant à la fois comme générateurs et comme charges sont considérés comme des nœuds auto-cohére
- Interaction. Les nœuds de puissance couplés à des nœuds de communication défaillants échoueront également. De plus, les nœuds de puissance couplés aux nœuds de communication sortant des délais de transmission tomberont en panne avec une certaine probabilité Pdi.

2.2.3. Modèle charge-capacité du réseau électrique basé sur l'interdépendance des flux

Dans des études antérieures, le principe de propagation du chemin le plus court a été couramment utilisé pour étudier la transmission de puissance entre les bus d'un réseau électrique. Cependant, en réalité, le flux d'énergie dans un réseau électrique ne suit pas seulement le chemin présentant l'impédance la plus faible, mais se propage plutôt le long de tous les chemins possibles, conformément à la loi de Kirchhoff.

Pour refléter avec précision le rôle de chaque ligne de transport dans la propagation de l'énergie et l' impact variable des différentes paires générateur-charge sur chaque ligne, considérons que dans un réseau électrique donné, chaque ligne de transport transporte une proportion variable de puissance de transport P(m, n) du générateur m à la charge n. Par conséquent, chaque ligne joue un rôle distinct et a différents niveaux d'importance dans la puissance de transmission P(m, n). Étant donné que les chemins de transmission de puissance pour les paires générateur-charge qui traversent chaque ligne diffèrent, l'importance de la ligne dans l'ensemb

Le réseau est quantifié à l'aide de l'indice d'intermédiarité des flux [26]. Cet indice est calculé en considérant toutes les paires générateur-charge utilisant la ligne. La formule de calcul est la suivante :

$$FBIj = \sum_{m \ G} \sum_{n \ L} \min(Sm, Sn) \ \frac{Pij,mPij,nPn}{PijPLnA \ unmPGm^{-1}} \quad (5)$$

où G est l'assemblage des nœuds de génération et L est l'assemblage des nœuds consommateurs. min(Sm, Sn) est le poids de l'interconnexion du flux d'une seule ligne, qui dépend de la valeur minimale entre la sortie réelle du générateur m et la charge réelle n, reflétant la puissance de transmission maximale disponible entre m et n. Pij,m est la partie du flux d'énergie sur la ligne eij provenant du générateur m. Pij,n est la partie du flux de puissance sur la ligne eij dirigée vers la charge n. Pn est le flux de puissance du nœud de n. Pij est la puissance active passant par la ligne eij. PLn est la charge $_{hum}^{-1}$ contient l'ordre inverse active au nœud de charge n. Une matrice de répartition des éléments. PGm est la sortie active du nœud

Par conséquent, la charge de puissance Lp(ij) sur le bord eij peut être calculée comme suit :

$$Lp(ij) = FBij \quad (6)$$

Compte tenu de la capacité des liaisons à gérer des charges de puissance, nous utilisons le modèle charge-capacité proposé par Motter et Lai [31]. Selon ce modèle, la capacité de la liaison est directement proportionnelle à sa charge électrique initiale, comme suit :

$$Cp(ij) = (1 + \gamma)Lp(ij) \quad (7)$$

où γ est le paramètre de tolérance.

2.2.4. Processus de défaillance en cascade

Le réseau d'information a la capacité de contrôler les nœuds de puissance. Par conséquent, les défauts survenant au sein du réseau d'information peuvent se propager au réseau électrique interconnecté. Cette section détaille le processus dynamique de pannes en cascade déclenchées par certains liens de communication défectueux.

En fonction de l'ampleur de la transmission du flux d'informations dans le réseau de communication , les états opérationnels des liens sont classés en trois types : normal, surchargé et défaillant. Pour analyser les pannes en cascade entre réseaux couplés, certains liens du réseau de communication sont supprimés de manière aléatoire en tant que pannes initiales. En référence à l'organigramme de la figure 4, le processus dynamique détaillé des défaillances en cascade déclenchées par des liens endommagés spécifiques est décrit comme suit :

- Étape 1 : L'ensemble des liaisons défaillantes dans le réseau de communication de la couche d'accès en raison de pannes accidentelles ou d'attaques est noté eij. Le flux d'informations sur ces liens défaillants sera supporté par les fronts connectés aux nœuds des liens défaillants. • Étape 2 : Processus de distribution du flux d'informations sur les branches défaillantes. Cet article utilise le principe de redistribution locale du flux d'informations, avec la formule de calcul fournie comme suit.

$$\Delta Lc,ia = Lc,ij \ \frac{Lc,ia}{\sum m \quad \Omega 1 \ Lc,im + \sum n \quad \Omega 2 \ Lc,jn} \quad (8)$$

où Lc,ij est le flux d'information initial sur une branche eij, Lc,ia est le flux d'information initial sur une branche eia, Ω1 est l'ensemble des nœuds voisins du nœud i, et Ω2 est l'ensemble des nœuds voisins du nœud j.

• Étape 3 : Le processus permettant de déterminer les états surchargés et défaillants est le suivant.

$$\begin{cases} Wc,im > \delta & \text{échouer} \\ 1 < Wc,im < \delta \text{ et rand} > \text{surcharge pim} & \\ 1 < Wc,im < \delta \text{ et rand} \le \text{pim échouent} & \\ Wc,im \le 1 & \text{normale} \end{cases} \tag{9}$$

où rand (0, 1).

Puisque chaque branche a des capacités différentes pour gérer un flux d'informations supplémentaire, un coefficient de distribution ω est introduit pour caractériser cette propriété.

$$pia = \left(\frac{Wc,ia - 1}{\delta - 1}\right)^{\omega} \tag{dix}$$

• Étape 4 : Processus de diffusion des flux d'informations sur les branches surchargées :

$$\Delta_{ak} = (Lc,ia - Cc,ia)Tak \tag{11}$$

$$Tak = \frac{Cc,ak - Lc,ak}{\sum e \quad \Omega i\,(Cc,ie - Lc,ie) + \sum f \quad \Omega a\,(Cc,af - Lc,a\,f)} \tag{12}$$

où Δ_{ak} est la stratégie de distribution de la branche surchargée, Ωi est l'ensemble des nœuds voisins du nœud i avec des branches dans un état normal, et Ωa est l'ensemble des nœuds voisins du nœud a avec des branches dans un état normal. • Étape 5 : Déterminez s'il existe de nouvelles branches défaillantes. Si de nouvelles branches défaillantes sont

détecté, passez à l'étape 2 ; sinon, passez à l'étape 6.

• Étape 6 : compter les ensembles de liens défaillants et surchargés au sein du réseau de communication . Mettez à jour l'ensemble efficace de liaisons de communication et calculez les incréments de délai de transmission pour chaque nœud de communication. Évaluez les conditions de défaillance des nœuds de communication, supprimez les nœuds qui ne peuvent pas former un chemin vers le centre de contrôle et mettez à jour l'ensemble des nœuds de communication efficaces.

• Étape 7 : Analyse des dépendances de contrôle : En raison du couplage un à un entre les nœuds de communication de la couche d'accès et les nœuds de puissance, les nœuds de communication défaillants identifiés à l'étape 6 entraîneront la défaillance des nœuds de puissance correspondants via les bords dépendants. Les nœuds sont sélectionnés sur la base de la probabilité de retard de chaque nœud de communicatio Si les informations de ces nœuds ne sont pas transmises au centre de contrôle en temps opportun, leurs nœuds d'alimentation correspondants sont considérés comme défaillants. Marquez les nœuds d'alimentation qui ont échoué au cours de

cette étape. • Étape 8 : Défaillance physique : Tout d'abord, supprimez les nœuds défaillants du réseau électrique. Ensuite, en fonction des conditions de défaillance des nœuds du système électrique, vérifiez s'il existe des nœuds défaillants supplémentaires parmi les nœuds restants et supprimez-les si nécessaire. Recalculer la charge électrique sur les lignes ; si la charge redistribuée dépasse la capacité d'une ligne, marquez cette ligne comme défectueuse. Répétez ce processus jusqu'à ce qu'il ne reste plus de nœuds défectueux dans le

réseau électrique. • Étape 9 : Résultat : la simulation de défaillance en cascade unique se termine. Sortez les données pou le réseau électrique et le réseau de communication.

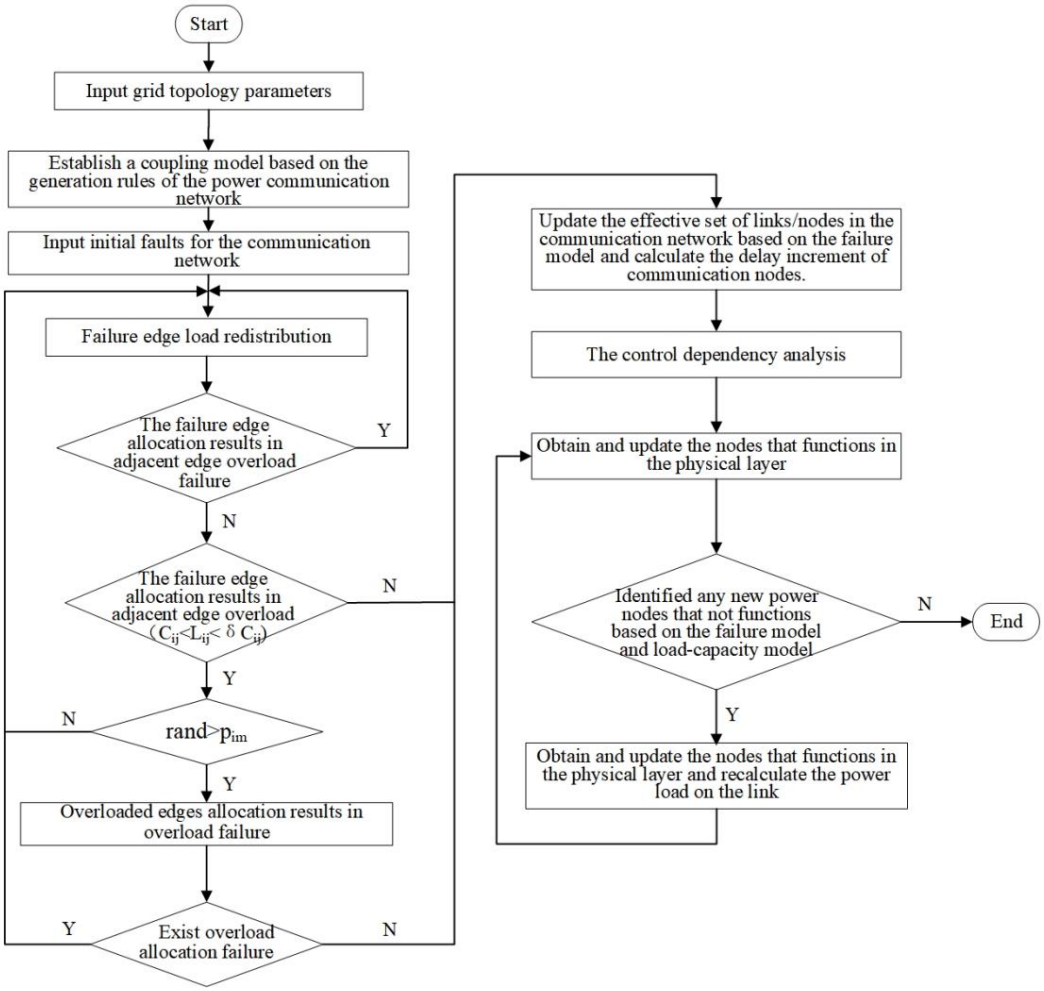


Figure 4. Organigramme du processus de défaillance en cascade dans les systèmes électriques cyber-physiques.

3. Mesures d'évaluation de la robustesse du système face aux pannes en cascade

Dans CPPS, tout lien défectueux a le potentiel de se propager à travers des relations de couplage et d'évoluer vers des pannes en cascade. Pour analyser l'impact des pannes initiales sur le système, nous définissons des métriques d'évaluation pour les deux réseaux couplés en fonction de l'intégrité de la topologie du réseau et des caractéristiques opérationnelles.

3.1. Métriques du réseau de communication

Dans le réseau de communication, nous utilisons des taux de survie de nœud/lien ajustés pour évaluer l'intégrité topologique du réseau de communication. Étant donné que les liens de communication dans un état surchargé ont un flux d'informations dépassant leur capacité – contrairement aux conditions normales – ces liens fonctionnent de manière inefficace et ont une certaine probabilité de passer à un état défaillant. Par conséquent, pour distinguer les liens normaux et surchargés et représenter plus précisément l'impact de l'état de surcharge sur le système, la taille relative du plus grand composant connecté Gc est utilisée pour évaluer les liens de communication après ajustement [29]. Le taux de survie des nœuds/liens après une défaillance est utilisé pour évaluer l'intégrité topologique, comme détaillé ci-dessous :

$$Gc = \frac{\sum h \quad \Psi sh}{NC} \tag{13}$$

où Nc est le nombre de nœuds dans la couche d'accès du réseau de communication. Ψ désigne l'ensemble des nœuds non défaillants. Lorsque les liens d'information sont dans un état normal, sh = 1.

Machine Translated by Google

Cependant, lorsque les liaisons d'information sont dans des conditions de surcharge, sh est calculé comme suit :

$$= \delta Ch \frac{\delta Ch - Lh\ sh}{- Ch} \quad (14)$$

Ensuite, le taux de survie ajusté des nœuds/liens est

$$Fc = \frac{1}{2} \left(\frac{V_c'}{Vc} + Gc \right) \quad (15)$$

En raison de la déconnexion des liaisons d'information, le chemin depuis les nœuds de couche d'accès du réseau de communication jusqu'au centre de contrôle peut changer, entraînant des retards de communication. Dans cet article, le délai de communication des données est simplifié et calculé comme suit [32] : La transmission des données dans le réseau de communication suit le principe du chemin le plus court, et chaque fois qu'elles passent par un nœud de données, le délai augmente d'une unité de temps τ. L'unité de temps de retard reflète le retard provoqué par la transmission et le traitement des données du nœud source au nœud de destination dans le réseau de communication, y compris le passage par chaque nœud d'informations et le chemin de communication vers le nœud d'informations suivant.

Par conséquent, l'incrément de retard de transmission T provoqué par le réseau de communication est calculé comme suit :

$$T = \sum_{k \quad Lc} T'_{\pi c} + \sum_{k \quad Lc} T\pi c \quad (16)$$

où $T'_{\pi c}$ et Tπc sont le retard du chemin de transmission des mêmes paires source-destination après et avant la panne en cascade. Lc désigne l'ensemble des chemins de transmission les plus courts pour toutes les paires source-destination du réseau de communication.

3.2. Métriques du réseau électrique

Dans le réseau électrique physique, nous utilisons l'impact de défaillance Fp pour évaluer l'influence de la défaillance en cascade [12], exprimée sous la forme

$$Fp = 2 \frac{1}{} \left(\frac{Vp'}{\text{Vice-président}} \right) + \frac{E_p'}{Ep.} \quad (17)$$

où V_p' et E_p', respectivement, sont le nombre de nœuds et de liaisons défaillants dans le réseau électrique. Vp et Ep représentent respectivement le nombre total de nœuds et de liaisons dans le réseau électrique physique.

4. Étude de cas et discussion

Dans cette section, en prenant le système de bus IEEE 39 comme exemple, la topologie du réseau électrique est illustrée à la figure 5 et la topologie d'origine est divisée en quatre régions basées sur des principes de zonage. En conséquence, le réseau de communication de puissance, généré selon les règles susmentionnées, est représenté sur la figure 6. Dans ce réseau, la couche d'accès se compose de 39 nœuds de communication, chacun correspondant à l'un des 39 nœuds de puissance, qui sont utilisés pour télécharger des informations sur les défauts. et émettre des instructions d'expédition.

La couche de base comprend quatre nœuds de contrôle, chacun correspondant à sa région de couche d'accès respective. Le centre de répartition est équipé de deux nœuds de contrôle.

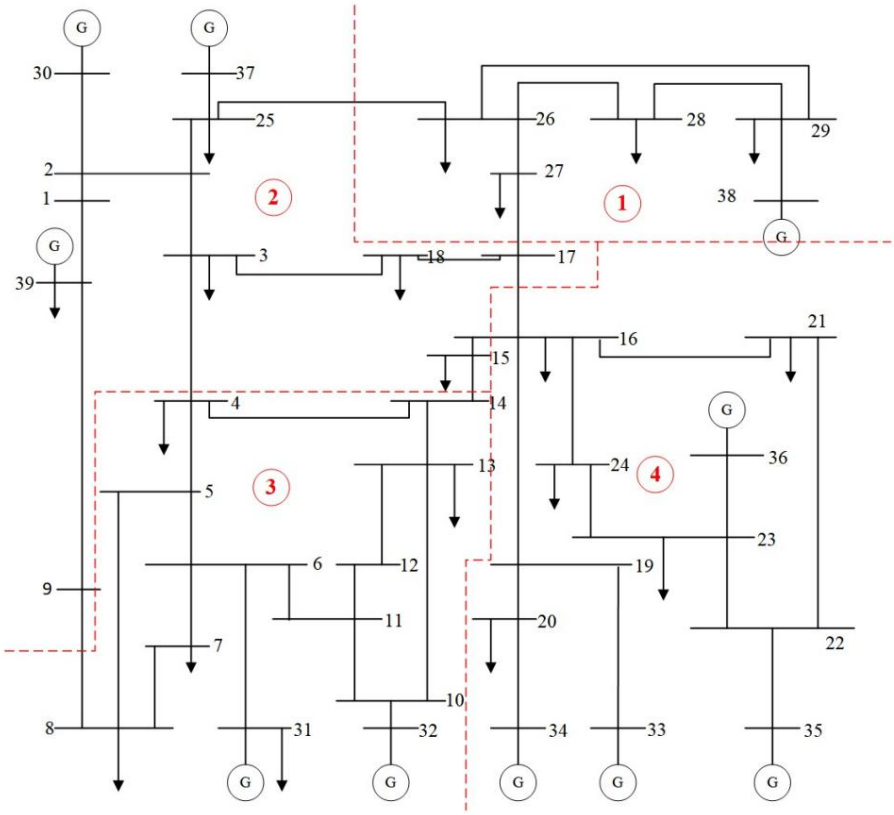


Figure 5. Schéma de partition du système de bus IEEE 39.

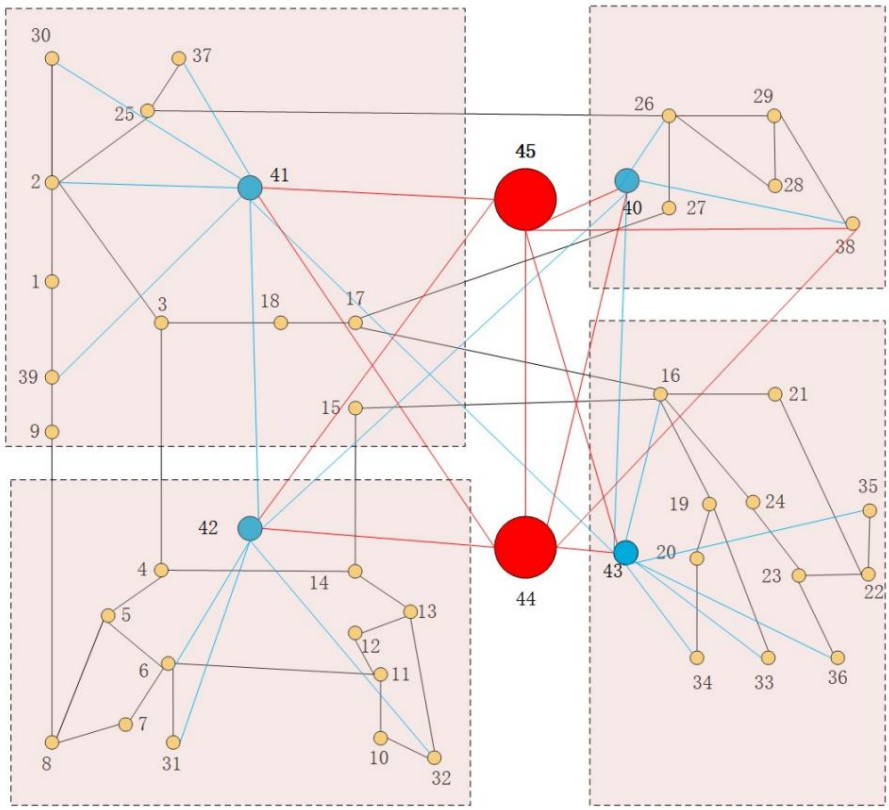


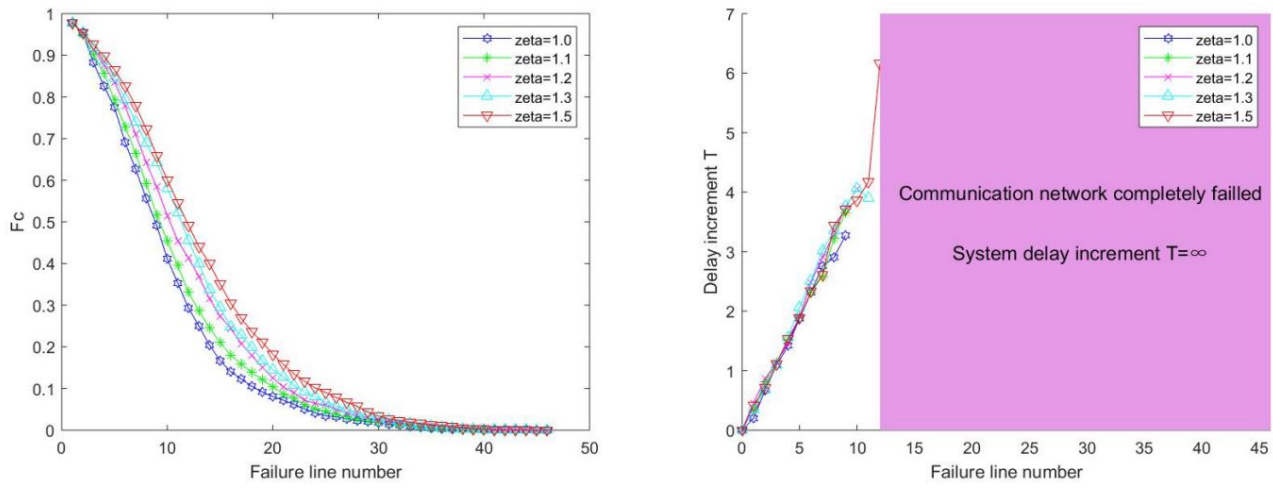
Figure 6. Diagramme de topologie du réseau d'informations généré par le système de bus IEEE 39.

4.1. Impact du coefficient de surcharge

Pour explorer l'impact du coefficient de surcharge sur la robustesse du réseau, le nombre de défauts de liaison initiaux a été varié, chaque ensemble de liens défectueux étant généré de manière aléatoire. Les simulations ont été exécutées indépendamment 50 fois et la valeur moyenne a été prise comme mesure du réseau de communication. Pour plus de clarté dans la présentation dans les graphiques, des descriptions textuelles ont été utilisées lorsque le système de communication a complètement échoué, et l'ensemble de données comportant le plus grand nombre de liens défaillants ayant provoqué la défaillance complète du système de communication est sélectionné pour une explication textuelle. Selon la figure 7a, à mesure que le nombre de liens défectueux initiaux augmente, le taux de survie des nœuds/liens du réseau de communication diminue progressivement. Lorsque la capacité de surcharge des liens n'est pas prise en compte, le taux de survie des nœuds/liens du réseau est le plus bas. Lorsque le coefficient de surcharge est de 1,2, le taux de survie des nœuds/liens du réseau de communication s'améliore considérablement. Cependant, on peut observer que lorsque les coefficients de surcharge sont de 1,3 et 1,5, le taux de survie des nœuds/liens ne s'améliore pas significativement.

La figure 7b montre que le CPPS sous différents coefficients de surcharge présente une transition de percolation de premier ordre, avec une tendance globale similaire. À mesure que le nombre de lignes défaillantes augmente, la topologie de la couche d'informations est perturbée et certains nœuds perdent leurs chemins de transmission de données. La modification du chemin de transmission augmente le délai du système. Lorsque le coefficient de surcharge est de 1,5, le nombre de liaisons défaillantes que le réseau de communication peut supporter avant son effondrement complet est le plus élevé. Comme le montrent la figure 7b et le tableau 1, il existe un seuil pour le nombre initial de liaisons défaillantes, à proximité duquel l'effet de défaillance en cascade s'étend à toutes les liaisons de communication, ne laissant aucun chemin de transmission entre les nœuds de communication et le centre de contrôle. À ce stade, le système de communication tombe complètement en panne, rendant impossible le contrôle du réseau électrique. Le tableau 1 montre les seuils spécifiques pour différents coefficients de surcharge.

Ainsi, compte tenu des coûts de construction du réseau, le coefficient de surcharge est fixé à 1,3 selon la figure 7 et le tableau 1.



(un) (b)

Figure 7. La robustesse du réseau de communication et l'augmentation du délai de communication sous différents coefficients de surcharge. (a, b) montrent respectivement la relation entre la robustesse du réseau de communication et le retard du système avec l'augmentation du nombre de pannes de liaison de communication sous différents coefficients de surcharge, tandis que les autres paramètres restent constants.

Tableau 1. Les seuils du système sous différents coefficients de surcharge.

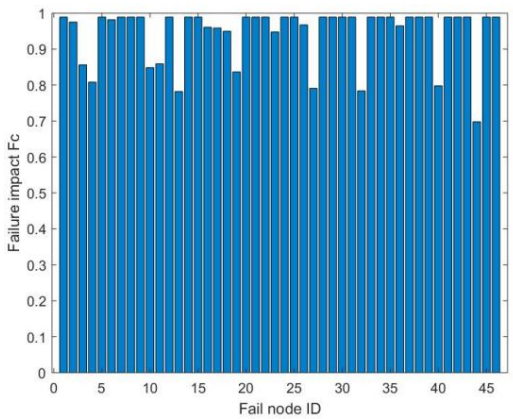
Coefficient de surcharge	Le seuil des lignes initiales défaillantes	Seuil Pourcentage %
1,0	9	19h56
1,1	9	19h56
1,2	10	21.74
1,3	11	23.91
1,5	12	26.09

4.2. Impact des liens défaillants

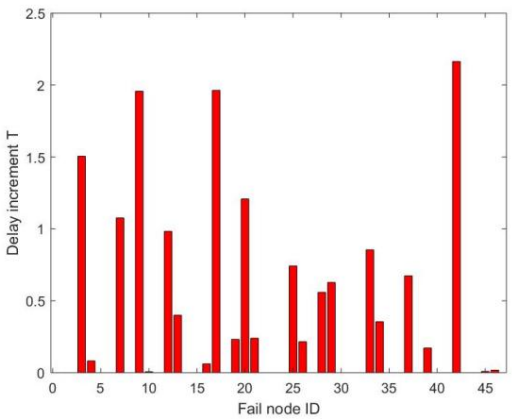
Dans cette section, nous analyserons l'impact sur la transmission des données au sein du réseau électrique et réseaux de communication en supprimant séquentiellement chaque lien du réseau de communication.

Comme le montre la figure 8, la plupart des liens de communication défectueux ont des impacts variables sur l'intégrité topologique et les caractéristiques opérationnelles du CPPS. Des liens qui provoquent les perturbations importantes du système sont identifiées comme des liens critiques. Dans des conditions de défaut N-1, tous les scénarios possibles d'attaques du système sont énumérés.

Comme le montre la figure 8, on peut observer que la défaillance d'un seul système de communication Le lien peut réduire le taux de survie des nœuds et des liens dans le réseau de communication à un niveau aussi bas à 69,69 %, induisent un incrément de retard du système de 2,1635 et provoquent une paralysie de 35,07 % dans les nœuds et les maillons du réseau électrique. Il est également à noter que certaines défaillances de liaison ne affecter le reste des réseaux de communication et d'alimentation au-delà de la liaison défaillante. C'est car, lors de pannes en cascade, le réseau de communication considère l'état encombré du les liens, redistribuant le trafic sur les liens défaillants et surchargés. De plus, surchargé les liaisons peuvent fonctionner pendant une courte période avant de tomber en panne, améliorant ainsi la performance du système. une certaine robustesse. De plus, le lien de communication qui provoque le maximum L'incrément de retard n'entraîne pas nécessairement le taux de survie de nœud/liaison le plus bas dans le réseau de communication ou le taux de défaillance le plus élevé dans les nœuds/liaison du réseau électrique. Cependant, les taux de défaillance des nœuds/liens dans les réseaux de communication et d'énergie sont notamment similaire. Par exemple, lorsque la liaison 42 échoue, l'incrément de délai le plus élevé de 2,1635 se produit. À À ce stade, le taux de survie des nœuds/liens de communication est de 98,91 %, et le taux d'échec de les nœuds/liens électriques sont de 3,46 %. En effet, l'échec du lien 42 modifie le chemin de nœud de communication 27 vers le centre de contrôle, du chemin d'origine 27 → 26 → 40 → 44 vers 27 → 17 → 16 → 43 → 44, ce qui entraîne un retard de communication. Le taux d'échec de la communication le nœud 27 est de 36,06 %. Cependant, la suppression du lien 42 ne provoque pas de panne ou de surcharge dans d'autres liens, il y a donc 45 liens opérationnels, et le nombre de nœuds de communication efficaces reste inchangé par rapport à avant le défaut. En raison de la relation de couplage, le le taux de défaillance du nœud de puissance 27 est également de 36,06 %, conduisant finalement à une panne de courant. nœud 27.

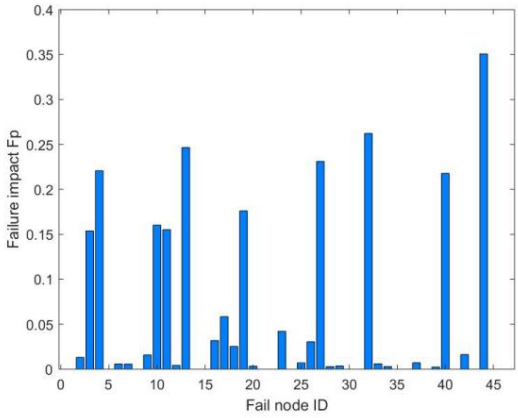


(un)



(b)

Figure 8. Suite.



(c)

Figure 8. L'impact des pannes initiales sur le système cyber-physique. (a, b) montrent l'impact de chaque défaillance de liaison de communication sur la robustesse et l'incrément de retard du système de communication, tandis que (c) montre l'impact de chaque défaillance de liaison de communication sur le réseau électrique physique.

4.3. Impact des stratégies d'attaque

La section précédente a analysé l'impact N-1 provoqué par la défaillance d'un seul lien de communication. À mesure que le nombre de lignes défectueuses augmente, différents ensembles initiaux de lignes défectueuses auront des effets variables sur le système. Par conséquent, sur la base des caractéristiques structurelles et électriques , nous utilisons à la fois une attaque de ligne aléatoire et une attaque de ligne critique pour supprimer les liaisons du réseau de communication une par une, en analysant l'impact des différentes stratégies d'attaque sur la transmission des données du réseau d'alimentation et de communication.

La stratégie d'attaque de ligne critique est dérivée du classement initial de l'importance de la centralité du pouvoir du réseau électrique. En raison de l'incertitude de la stratégie d'attaque par ligne aléatoire, chaque ligne défectueuse dans la stratégie d'attaque aléatoire est générée aléatoirement et les résultats sont moyennés après 50 exécutions indépendantes.

Comme le montre la figure 9, par rapport aux attaques aléatoires, le CPPS présente une grande vulnérabilité face aux attaques sur les lignes critiques. En cas d'attaques de lignes critiques, même en tenant compte au préalable de la redistribution des flux d'informations en raison d'une surcharge de liaison, le réseau subit toujours des dommages importants. En effet, la défaillance de lignes critiques perturbe les chemins de transmission des informations vers le centre de contrôle, empêchant ainsi le téléchargement des informations sur les pannes d'alimentation . Par conséquent, le centre de contrôle ne peut pas répondre et les nœuds sources transmettant des informations dans le réseau de communication provoqueront la défaillance des nœuds de puissance couplés, élargissant ainsi l'ampleur des pannes et accélérant la propagation des pannes en cascade et l'effondrement du système.

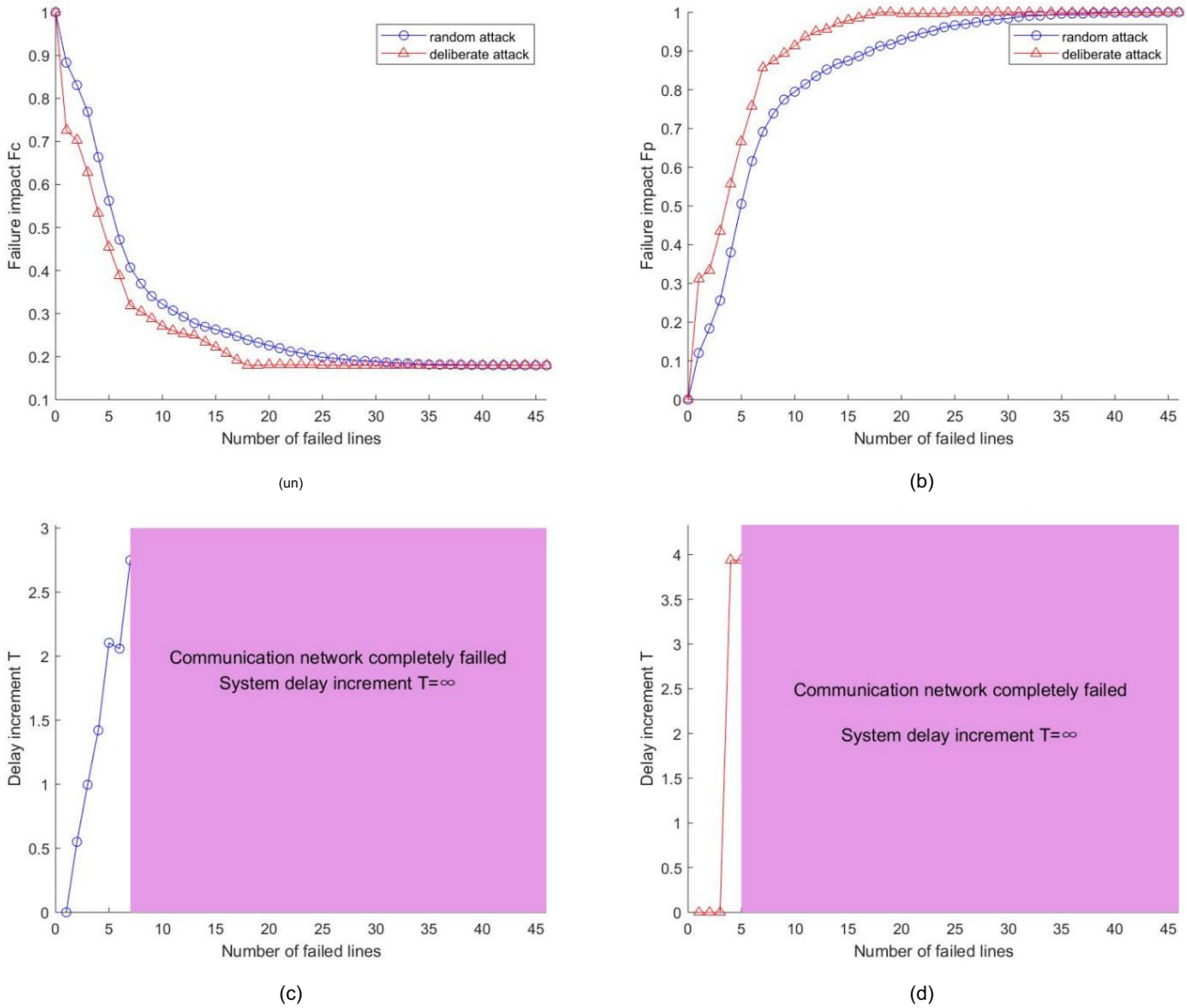


Figure 9. Impact des stratégies d'attaque sur le système cyber-physique. (a – d) illustrent la relation entre les métriques d'évaluation du réseau couplé et le nombre de liens de communication défaillants sous deux stratégies d'attaque différentes (attaque aléatoire et attaque délibérée). Plus précisément, (a) montre la relation entre le taux de survie des nœuds/liens dans le réseau de communication électrique et le nombre de liens de communication défaillants, (b) montre la relation entre le taux de défaillance des nœuds/liens dans le réseau de communication électrique et le nombre de liens de communication défaillants. (c, d) décrivent respectivement la tendance de l'augmentation du retard dans le système sous une attaque aléatoire et une attaque délibérée.

5. Conclusions

Cet article étudie la propagation dynamique des défaillances en cascade dans CPPS. Tout d'abord, la relation de couplage entre le réseau de communication et le réseau électrique physique est analysée, et la structure topologique du système de communication correspondante est générée sur la base du réseau électrique. Ensuite, un modèle de distribution de surcharge pour le flux d'informations est établi au sein du réseau de communication pour gérer les surcharges résultant de pannes. Dans le réseau électrique, un modèle de distribution charge-capacité basé sur l'interdépendance du flux de puissance est établi pour gérer la complexité des calculs de flux de puissance dus aux pannes. Pour décrire plus précisément les défaillances de nœuds/lignes causées par les transferts de flux de puissance dus à des défauts, la théorie traditionnelle de la percolation est améliorée pour établir un modèle de propagation des défaillances pour les réseaux couplés.

Dans l'ensemble, cet article considère de manière exhaustive à la fois les caractéristiques électriques de le réseau électrique physique et les caractéristiques de transmission des informations de l'électricité réseau de communication lors de l'établissement du modèle de couplage. Il améliore le couplage modèle de réseau à partir de la structure topologique et des caractéristiques fonctionnelles et construit un modèle de défaillance en cascade qui reflète plus précisément les scénarios du monde réel. La plupart des études existantes sont basées sur la théorie de la percolation et analysent uniquement les défaillances en cascade. processus de CPPS du point de vue de l'intégrité topologique du réseau. Cette approche échoue pour capturer de manière réaliste les phénomènes d'îlotage dans les opérations du système électrique, nécessitant ainsi une amélioration de la théorie de la percolation.

Les simulations analysent l'impact des lignes défectueuses sur la capacité de surcharge des liaisons du réseau de communication, la transmission de données, la capacité de survie du réseau et la puissance physique. taux de pannes du réseau. Les résultats indiquent que même une légère augmentation de la capacité des liaisons peut améliorer considérablement la robustesse du réseau. De plus, les performances du système se dégradent lorsque des liens situés à des positions topologiques critiques du réseau de communication sont attaqués, conduisant à des incréments de retard système accrus par rapport aux attaques aléatoires et provoquant le le réseau de la couche d'information s'effondrerait plus tôt.

En résumé, ces résultats offrent des informations précieuses pour la planification future des réseaux électriques. Cependant, ces études sont basées sur des conditions post-faute, en considérant uniquement les interactions entre les nœuds comme un échec normal ou complet dans la répartition du flux de puissance par le centre de contrôle. En réalité, le processus d'interaction entre les nœuds de pouvoir et l'information Les nœuds sont extrêmement complexes et souvent analysés uniquement qualitativement. De plus, comparé Pour la répartition du flux d'énergie, la planification du flux d'informations est relativement simple. Réduire le l'apparition de défauts provenant de la source du côté cyber est une solution plus efficace et plus fiable au lieu d'établir un mécanisme de résistance aux pannes en cascade du côté du réseau électrique. Par conséquent, nos futures recherches se concentreront sur l'allocation des flux d'informations critiques à des sources fiables. chemins dans des conditions de surcharge des liens de communication pour garantir l'accessibilité et réduire la probabilité de pannes en cascade causées par des interruptions de communication.

Contributions des auteurs : Conceptualisation, XL et YL ; Méthodologie, XL et YL ; Écriture—originale brouillon, XL; Rédaction – révision et édition, YL et TX Tous les auteurs ont lu et accepté le version publiée du manuscrit.

Financement : Ce travail a été soutenu par la Fondation nationale des sciences naturelles de Chine dans le cadre Subvention 62062068 et programme de jeunes leaders académiques et techniques de la province du Yunnan dans le cadre Subvention 202305AC160077.

Déclaration de disponibilité des données : les contributions originales présentées dans l'étude sont incluses dans le article, des demandes complémentaires peuvent être adressées à l'auteur correspondant.

Remerciements : Merci pour l'aide du Kunming AI Computing Center.

Conflits d'intérêts : Les auteurs ne déclarent aucun conflit d'intérêts.

Abréviations et nomenclature

CPPS	Système d'alimentation cyber-physique ;
Fij	Demande de flux d'informations sur le lien eij ;
qij	Le rapport entre la capacité du lien et la charge du lien ;
ρ	Le seuil de défaillance du lien ;
[illegible]	Les ensembles de nœuds de puissance ;
Ep.	Les ensembles de lignes de transport d'énergie ;
Q	La valeur de modularité de l'union combinée de deux communautés ;
qi	La proportion de la production du générateur i par rapport à la production totale du système ;
δ	La capacité du lien à gérer un flux d'informations supplémentaire ;
ki	Les degrés du nœud i ;
θ	Le paramètre qui ajuste le flux d'informations ;
Lc,ij	La quantité d'informations transmises par le lien eij ;
Cc,ij	La capacité du lien eij ;

$C_{c,ij}^{maximum}$	Le débit maximum que peut supporter le lien eij ;
$WC_{,ij}$	Le poids du lien eij ;
α	Le coefficient de capacité ;
β	Le coefficient de capacité ;
min(Sm, Sn)	Le poids de l'intermédiarité du flux d'une seule ligne ;
$P_{ij,m}$	La partie du flux d'énergie sur la ligne eij provenant du générateur m ;
$P_{ij,n}$	La partie du flux d'énergie sur la ligne eij dirigée vers la charge n ;
Pij	La puissance active passant par la ligne eij;
PLn	La charge active au nœud de charge n ;
UN_{hum}^{-1}	Les éléments de la matrice de distribution d'ordre inverse ;
PGm	La sortie active du nœud générateur m ;
FBIj	L'interdépendance du flux eij dans le réseau électrique ;
Lp(ij)	La charge électrique en pointe eij dans le réseau électrique ;
Cp(ij)	La capacité du bord eij dans le réseau électrique ;
γ	Le paramètre de tolérance ;
Δ_{ak}	La stratégie de distribution pour la succursale surchargée ;
FC	Le taux de survie ajusté des nœuds/liens suite à une défaillance du réseau de communication ;
T	L'incrément du délai de transmission ;
Fp	Le taux de nœuds/liens défaillants suite à une panne du réseau électrique.

Les références

1. Abdelmalak, M. ; Venkataramanan, V. ; Macwan, R. Une enquête sur les méthodes de modélisation des systèmes électriques cyber-physiques pour l'énergie future systèmes. Accès IEEE 2022, 10, 99875-99896. [Référence croisée]
2. Li, Y. ; Wang, B. ; Wang, H. ; Maman, F. ; Maman, H. ; Zhang, J. ; Zhang, Y. ; Mohamed, MA Un interdépendant efficace de nœud à bord Analyse de réseau et de vulnérabilité pour les réseaux électriques couplés numériques. Int. Trans. Electr. Système énergétique. 2022, 2022, 5820126. [Référence croisée]
3. Gao, X. ; Peng, M. ; Chi, KT ; Zhang, H. Un modèle stochastique de dynamique de défaillance en cascade dans les systèmes électriques cyber-physiques. IEEE Système. J. 2020, 14, 4626-4637. [Référence croisée]
4. Liang, G. ; Weller, SR ; Zhao, J. ; Luo, F. ; Dong, ZY La panne d'électricité en Ukraine en 2015 : implications pour les attaques par fausse injection de données. IEEE Trans. Système d'alimentation. 2016, 32, 3317-3318. [Référence croisée]
5. Zhao, Q. ; Qi, X. ; Hua, M. ; Liu, J. ; Tian, H. Bilan des récentes pannes d'électricité et des Lumières. Dans les Actes du CIRED 2020 Atelier de Berlin (CIRED 2020), conférence en ligne, 22-23 septembre 2020 ; IET : Stevenage, Royaume-Uni, 2020 ; Tome 2020 ; pp. 312-314.
6. Wu, J. ; Chen, Z. ; Zhang, Y. ; Xia, Y. ; Chen, X. Récupération séquentielle de réseaux complexes souffrant de pannes de courant en cascade. IEEETrans. Réseau. Sci. Ing. 2020, 7, 2997-3007. [Référence croisée]
7. Peng, C. ; Wu, J. ; Tian, E. Contrôle H∞ déclenché par un événement stochastique pour les systèmes en réseau soumis à des attaques par déni de service. IEEE Trans. Système. Homme Cybern. Système. 2021, 52, 4200-4210. [Référence croisée]
8. Hu, S. ; Yue, D. ; Han, QL; Xie, X. ; Chen, X. ; Dou, C. Contrôle déclenché par un événement basé sur un observateur pour les systèmes linéaires en réseau Soumis aux attaques par déni de service. IEEETrans. Cybern. 2020, 50, 1952-1964. [Référence croisée]
9. Bouldyrev, SV; Parshani, R. ; Paul, G. ; Stanley, S.E. ; Havlin, S. Cascade catastrophique de pannes dans les réseaux interdépendants. Nature 2010, 464, 1025-1028. [Référence croisée]
10. Huang, X. ; Gao, J. ; Bouldyrev, SV; Havlin, S. ; Stanley, HE Robustesse des réseaux interdépendants face à des attaques ciblées. Phys. Rév. E—Stat. Physique de la matière molle non linéaire. 2011, 83, 065101. [Réf. croisée] [Pub Med]
11. Cao, YY; Liu, RR ; Jia, CX; Wang, BH Percolation dans les réseaux complexes multicouches avec connectivité et interdépendance structures topologiques. Commun. Sci non linéaire. Numéro. Simul. 2021, 92, 105492. [Réf. croisée]
12. Chen, L. ; Yue, D. ; Dou, C. ; Cheng, Z. ; Chen, J. Robustesse des systèmes d'alimentation cyber-physiques en cas de panne en cascade : survie de clusters interdépendants. Int. J. Electr. Système d'énergie électrique. 2020, 114, 105374. [Réf. croisée]
13. Poêle, H. ; Li, X. ; Na, C. ; Cao, R. Modélisation et analyse des défaillances en cascade dans les systèmes électriques cyber-physiques sous différents stratégies de couplage. Accès IEEE 2022, 10, 108684-108696. [Référence croisée]
14. Jianfeng, D. ; Jian, Q. ; Jing, W. ; Xuesong, W. Une méthode d'évaluation de la vulnérabilité du système d'alimentation cyber-physique prenant en compte défaillance des infrastructures du réseau électrique. Dans les actes de la conférence IEEE sur l'énergie et l'énergie durables (iSPEC) 2019, Pékin, Chine, 21-23 novembre 2019 ; IEEE : Piscataway, New Jersey, États-Unis, 2019 ; pages 1492 à 1496.
15. Tian, M. ; Dong, Z. ; Gong, L. ; Wang, X. Stratégies de renforcement des lignes pour les systèmes électriques résilients en tenant compte de la cyber-topologie interdépendance. Fiable. Ing. Système. Saf. 2024, 241, 109644. [Réf. croisée]
16. Chen, L. ; Yue, D. ; Dou, C. ; Xie, X. ; Li, S. ; Zhao, N. ; Zhang, T. Impact des pannes en cascade sur la distribution électrique et les données transmission dans les systèmes électriques cyber-physiques. IEEETrans. Réseau. Sci. Ing. 2023, 11, 1580-1590. [Référence croisée]
17. Zhong, J. ; Zhang, F. ; Yang, S. ; Li, D. Restauration d'un réseau interdépendant contre les surcharges en cascade. Phys. Une statistique. Mécanique. Son application. 2019, 514, 884-891. [Référence croisée]

18. Ding, D. ; Wu, H. ; Yu, X. ; Wang, H. ; Yang, L. ; Wang, H. ; Kong, X. ; Liu, Q. ; Lu, Z. Évaluation de la vulnérabilité du système d'alimentation cyber-physique basée sur un modèle de défaillance en cascade amélioré. J. Electr. Ing. Technologie. 2024, 1-12. [Référence croisée]
19. Wang, S. ; Gu, X. ; Chen, J. ; Chen, C. ; Huang, X. Stratégie d'amélioration de la robustesse des systèmes cyber-physiques à faible interdépendance. Fiable. Ing. Système. Saf. 2023, 229, 108837. [Réf. croisée]
20. Ghasemi, A. ; de Meer, H. Robustesse du réseau électrique interdépendant et des réseaux de communication face aux pannes en cascade. IEEETrans . Réseau. Sci. Ing. 2023, 10, 1919-1930. [Référence croisée]
21. Cao, R. ; Dong, X. ; Wang, B. ; Liu, K. Discussion sur la protection et les pannes en cascade du point de vue de la communication. Dans Actes de la Conférence internationale 2011 sur l'automatisation et la protection avancées des systèmes électriques, Pékin, Chine, 16-20 octobre 2011 ; IEEE : Piscataway, New Jersey, États-Unis, 2011 ; Volume 3, pages 2430 à 2437.
22. Zhang, G. ; Shi, J. ; Huang, S. ; Wang, J. ; Jiang, H. Un modèle de défaillance en cascade prenant en compte les caractéristiques de fonctionnement du couche de communication. Accès IEEE 2021, 9, 9493-9504. [Référence croisée]
23. Gao, X. ; Peng, M. ; Chi, KT Analyse de robustesse des systèmes électriques cyber-couplés avec des considérations d'interdépendance des structures, des opérations et des comportements dynamiques. Phys. Une statistique. Mécanique. Son application. 2022, 596, 127215. [Réf. croisée]
24. Wang, F. ; Couvercle.; Xu, X. ; Wu, R. ; Havlin, S. Propriétés de percolation dans un modèle de trafic. Europhys. Lett. 2015, 112, 38001. [Référence croisée]
25. Sabéri, M. ; Hamedmoghadam, H. ; Ashfaq, M. ; Hosseini, SA; Gu, Z. ; Shafiei, S. ; Nair, DJ ; Dixit, V. ; Gardner, L. ; Waller, ST ; et coll. Un processus de contagion simple décrit la propagation des embouteillages dans les réseaux urbains. Nat. Commun. 2020, 11, 1616. [Réf. croisée]
26. Liu, W. ; Liang, C. ; Xu, P. ; Dan, Y. ; Wang, J. ; Wang, W. Identification de la ligne critique dans les systèmes électriques basée sur l'intermédiarité des flux. Proc. CSEE 2013, 33, 90-98.
27. Wang, T. ; Soleil, C. ; Gu, X. ; Qin, X. Modélisation et analyse de vulnérabilité du réseau couplé de communication électrique. Proc. CSEE 2018, 38, 3556-3567.
28. Newman, ME Trouver la structure communautaire dans les réseaux en utilisant les vecteurs propres des matrices. Phys. Rév. E—Stat. Non linéaire doux Matière Phys. 2006, 74, 036104. [Réf. croisée]
29. Chen, CY; Zhao, Y. ; Gao, J. ; Stanley, HE Modèle non linéaire de défaillance en cascade dans des réseaux complexes pondérés prenant en compte les bords surchargés. Sci. Rep.2020 , 10, 13428. [CrossRef]
30. Stauffer, D. ; Aharony, A. Introduction à la théorie de la percolation ; Taylor et Francis : Abingdon, Royaume-Uni, 2018.
31. Motter, AE; Lai, YC Attaques basées sur Cascade sur des réseaux complexes. Phys. Rév. E 2002, 66, 065102. [CrossRef]
32. Han, Y. ; Guo, C. ; Zhu, B. ; Xu, L. Modéliser les défaillances en cascade dans le système d'alimentation cyber-physique sur la base d'une théorie améliorée de la percolation. Automatique. Electr. Système d'alimentation. 2016, 40, 30-37.

électronique

MDPI

Article

Une méthodologie pratique d'évaluation de la sécurité du système électrique Opérations tenant compte de l'incertitude

Nhi Thi Ai Nguyen [1], Dinh Duong Le 1,*, Van Duong Ngo [1], Van Kien Pham [1] et Van Ky Huynh [2]

[1] Faculté de génie électrique, Université de Danang — Université des sciences et technologies, Danang 550000, Vietnam ; ntanhi@dut.udn.vn (NTAN); nvduong@dut.udn.vn (VDN); pvkien@dut.udn.vn (VKP)

[2] L'Université de Danang, Danang 550000, Vietnam ; hvky@ac.udn.vn *

Correspondance : lddduong@dut.udn.vn

Résumé : Aujourd'hui, les sources d'énergie renouvelables (SER) sont de plus en plus intégrées dans les systèmes électriques. Cela signifie ajouter davantage de sources d'incertitude au système électrique. Pour faire face à l'incertitude des variables aléatoires d'entrée (RV) dans les problèmes de calcul et d'analyse du système électrique, des techniques de flux de puissance probabiliste (PPF) ont été introduites et se sont avérées efficaces. Actuellement , bien qu'il existe de nombreuses techniques proposées pour résoudre le problème PPF, la méthode de simulation de Monte Carlo (MCS) est toujours considérée comme la méthode la plus précise et ses résultats sont utilisés comme référence pour l'évaluation d'autres méthodes. Cependant, le MCS nécessite souvent une intensité de calcul très élevée, ce qui rend son application pratique difficile, en particulier avec les systèmes électriques à grande échelle. Dans le présent article, une technique avancée de regroupement de données est proposée pour traiter les données RV d'entrée afin de réduire la charge de calcul liée à la résolution du problème PPF tout en maintenant un niveau de précision acceptable. La méthode proposée peut être appliquée efficacement pour résoudre des problèmes pratiques liés à l'horizon temporel de fonctionnement des systèmes électriques. L'approche développée est testée sur le système de bus IEEE-300 modifié, indiquant de bonnes performances en matière de

Mots-clés : flux de puissance probabiliste ; système du pouvoir; énergie renouvelable

Citation : Nguyen, NTA ; Le, DD; Ngo, VD; Pham, VK ; Huynh, VK Une méthodologie pratique d'évaluation de la sécurité pour les opérations du système électrique en tenant compte de l'incertitude. Électronique 2024, 13, 3068. https://doi.org/10.3390/electronics13153068

Rédacteur académique : Ahmed Abu-Siada

Reçu : 26 juin 2024
Révisé : 31 juillet 2024
Accepté : 31 juillet 2024
Publié : 2 août 2024

1. Introduction

Lors du fonctionnement d'un système électrique, les paramètres du mode de fonctionnement tels que la tension aux bus, la puissance transmise à travers les branches, etc., doivent être régulièrement calculés pour évaluer la sécurité du système en comparant les paramètres avec leurs limites admissibles. En cas de risque pour la sécurité, des solutions raisonnables doivent être proposées pour le résoudre. Le flux d'énergie déterministe (DPF) est l'un des outils essentiels pour l'exploitation et la planification du système électrique. Néanmoins, pendant le processus de calcul, l' approche traditionnelle utilise des valeurs fixes d'injections de puissance aux nœuds (provenant de la production d'électricité, de la charge, etc.) et de la structure de réseau connue afin que les sources d'incertitude liées à ces facteurs ne soient pas prises en compte. C'est la principale limitation de la méthode traditionnelle du flux de puissance (PF) [1].

Pour pallier l'inconvénient mentionné ci-dessus, le PPF a été proposé et est devenu un outil de calcul très efficace. La charge, la production d'énergie d'une centrale et le fonctionnement d'un élément tel qu'une ligne, un transformateur, etc. peuvent suivre certaines règles de probabilité . En particulier, pour les systèmes électriques actuels, lorsque des énergies renouvelables supplémentaires telles que l'énergie solaire et éolienne, etc., sont connectées au système, la modélisation de l'intermittence des énergies renouvelables est très difficile. L'intermittence change souvent très rapidement et de manière stochastique, augmentant le niveau d'incertitude du système. Une méthode de calcul est donc nécessaire pour pouvoir intégrer les incertitudes dans le processus de calcul. En utilisant les méthodes PPF, les sorties, c'est-à-dire la tension aux bus, le PF transmis sur les branches, etc., changent également de manière aléatoire selon une loi de distribution de probabilité [1]. L'analyse PPF permet d'évaluer la probabilité de surcharge de ligne, la probabilité de sur/sous-tension, etc. A partir de là,

caractéristiques du système et la gravité de la violation, l'opérateur pourrait envisager et suggérer des solutions appropriées pour améliorer la sécurité du système.

L'approche PPF a été introduite pour la première fois par Borkowska en 1974 [2] et, depuis lors, plusieurs travaux de recherche sur la PPF ont été proposés à travers le monde. Généralement, les méthodes de calcul du PF utilisant la technique PPF peuvent être classées en trois catégories principales, à savoir les approches numériques, analytiques et d'approximation.

L'approche analytique [3–6] utilise des algorithmes et des techniques analytiques telles que les techniques de convolution et de cumulant. L'application de ces techniques analytiques combinées à la relation entre l'entrée et la sortie d'un problème PPF permet de déterminer la fonction de distribution des RV de sortie telles que la transmission de puissance sur la ligne, la tension nodale, l'angle de phase, etc., en fonction du système. paramètres, par exemple l' impédance totale de la ligne, l'impédance totale du transformateur, etc., et les distributions de probabilité des RV d'entrée de la charge et de la production d'électricité à partir des générateurs et RES traditionnels, ainsi que l'état de fonctionnement des appareils. La relation entre l'entrée et la sortie du problème de calcul PF est non linéaire. Néanmoins, la méthode analytique fonctionne bien avec une relation linéaire entre l'entrée et la sortie du problème. Par conséquent, la relation doit d'abord être linéarisée à l'aide d'une technique d'expansion, par exemple l'expansion de Taylor. L'un des avantages majeurs de l'approche analytique est qu'elle peut donner des résultats très rapides. Parmi les approches cumulant et convolution, l'approche convolution est plus gourmande en calcul que l'approche cumulant. Ainsi, actuellement, l'approche cumulante est plus populaire que l'approche convolution. Pour réaliser les fonctions de distribution pour les RV de sortie, l'approche cumulante est souvent utilisée simultanément avec des techniques d'expansion telles que l'expansion de Gram-Charlier ou de Cornish-Fisher [4]. En raison de l'avantage d' un calcul rapide, l'approche analytique pourrait être appliquée dans la pratique à un système électrique à grande échelle . Néanmoins, l'approche analytique présente certains inconvénients. Premièrement, la précision de l'approche analytique est considérablement affectée par l'utilisation de techniques qui linéarisent la relation entrée-sortie, en particulier lorsque la VR d'entrée change sur une large plage, par exemple dans le cas des SER. Deuxièmement, l'approche analytique utilise des techniques d'expansion qui peuvent bien fonctionner dans le cas où les fonctions de distribution des RV d'entrée sont soit une distribution gausienne, soit une distribution proche de la distribution gausienne. En fait, les fonctions de distribution des RV d'entrée du problème PPF pour un système électrique suivent souvent, en pratique, une distribution non gausienne, de sorte que les résultats obtenus seront limités. Pour pouvoir intégrer des fonctions de distribution discrètes des RV d'entrée dans le processus de calcul, la méthode de Von Mises est proposée [1].

L'approche typique pour le groupe des approximations dans le calcul du PPF est l' approche de l'estimation ponctuelle [7,8]. Dans cette approche, le RV d'entrée est décomposé en une séquence de paires de valeurs et de poids. Ensuite, le moment du RV de sortie est calculé en fonction du RV d'entrée, puis la fonction de distribution RV de sortie est obtenue. L'approche d'estimation ponctuelle peut fournir des résultats relativement rapides. De plus, contrairement à l' approche analytique, cette approche utilise la relation non linéaire entre l'entrée et la sortie du problème de calcul PF. Cependant, la principale limite de l'approche d'estimation ponctuelle est que sa précision diminue à mesure que l'ordre du moment augmente. Un autre inconvénient est que le temps de calcul requis augmente considérablement à mesure que le nombre de RV d'entrée augme

Une approche numérique typique est la MCS [9–14]. Dans MCS, les RV d'entrée sont échantillonnées, puis le calcul DPF est effectué pour tous les échantillons. Il répète la simulation avec un très grand nombre d'échantillons pour obtenir un résultat très précis. MCS utilise la relation non linéaire entre l'entrée et la sortie du problème PF, comme l'approche traditionnelle Le principal avantage de l'approche MCS est qu'elle donne des résultats très précis et fiables. De plus, les distributions de probabilité des RV d'entrée dans MCS sont faciles à représenter. Cependant, le plus gros inconvénient est que le volume de calcul est lourd et le temps de calcul est relativement long, ce qui rend difficile son application pour un réseau électrique pratique à grande échelle. Pour réduire la charge de calcul de la méthode MCS, plusieurs algorithmes de clustering sont proposés pour réduire le nombre d'échantillons, puis DPF est exécuté pour chaque cluster au lieu de l'exécuter pour tous les échantillons. Dans [15], un algorithme PSO est proposé à utiliser p

la tâche de clustering, tandis que K-means est utilisé dans [16,17]. Chaque algorithme de clustering a ses propres faiblesses. Pour PSO, son taux de convergence itérative est modeste et il reste souvent bloqué aux optimaux locaux lors du traitement d'ensembles de données de grande dimension. Pour l'algorithme K-means, il est sensible au choix de k et il est difficile de trouver le k optimal pour un ensemble de données donné. De plus, K-means ne s'adapte pas bien aux grands problèmes, c'est pourquoi, dans la présente étude, ce problème se concentre.

De l'analyse ci-dessus, il ressort que chaque approche PPF a ses propres caractéristiques. caractéristiques, avantages et inconvénients.

En plus de l'aperçu ci-dessus du PPF, pour résoudre divers problèmes liés à l'incertitude , des progrès récents ont également été constatés. Dans [18], une méthode de répartition coordonnée robuste en deux étapes pour les micro-réseaux multi-énergies est développée pour atténuer tous les effets négatifs des diverses incertitudes liées à l'énergie éolienne et aux charges. La méthode des intervalles est utilisée dans [18] pour caractériser les incertitudes. Une région d'exploitation d'émissions de carbone engagées (CCEOR) de systèmes énergétiques intégrés (IES) est proposée dans [19]. La méthode développée convertit le modèle CCEOR non linéaire incertain proposé en un modèle CCEOR convexe déterministe à nombres entiers mixtes. Dans [20], pour prendre en compte différents types d'incertitudes liées aux résultats des catastrophes, des événements extrêmes, des charges et de la production renouvelable, les mesures de pré-restauration et en temps réel sont coordonnées via une méthode de programmation stochastique en deux étapes .

Les principales contributions de cet article sont résumées comme suit : (1) L'objectif principal de cette étude est de développer une approche de calcul du PPF qui assure un certain niveau de précision par rapport au MCS mais qui doit donner des résultats très rapides proches du « temps réel ». » fonctionnement du système électrique. Afin d'exploiter les avantages de la précision de la méthode MCS tout en réduisant le temps et le volume du calcul, une technique de clustering en temps réel combinée au MCS dans le calcul PPF est proposée. La technique de clustering appliquée est simple mais efficace et adaptée à une application pratique. Le grand nombre d'échantillons de RV d'entrée du problème de calcul PPF utilisant MCS est réduit de manière significative et efficace, de sorte que le calcul PPF donne des résultats rapides. Grâce à cette caractéristique exceptionnelle, l'approche PPF peut être appliquée aux grands systèmes électriques dans la pratique et aux délais d'exploitation. (2) En outre, la discussion sur l'application des méthodes PPF dans le calcul et l'analyse du système électrique est également présentée en détail dans cet article. Cela fournit une image plus claire et plus intuitive de l'application de l'analyse PPF à la fois dans la planification et dans l'exploitation des systèmes électriques. Les limites actuelles des méthodes PPF en général sont également soulignées afin de fournir des sujets de recherches futures.

Le reste de cet article est structuré comme suit. La section 2 présente la méthodologie développée, tandis que les résultats obtenus par l'approche développée sont discutés dans la section 3. Dans la section 4, une discussion plus approfondie sur l'applicabilité de diverses méthodes PPF aux problèmes d'analyse du système électrique est donnée. Les remarques finales sont fournies dans la sectior

2. Méthodologie 2.1.

Technique de clustering en temps réel

Le clustering est la division des données en un certain nombre de groupes afin que les points de données d'un même groupe aient des caractéristiques similaires les uns aux autres et soient différents des points de données d' autres groupes. Il s'agit essentiellement de diviser les données d'un ensemble de données en fonction des similitudes et des différences entre elles. Parmi les techniques de clustering, K-means est connue comme la plus populaire et est appliquée dans tous les domaines en raison de sa simplicité, de son efficacité, de son évolutivité et de sa facilité de mise en œuvre. Il peut gérer efficacement de grands ensembles de données, ce qui en fait un choix pratique pour de nombreuses applications. Un examen complet de l'application du clustering K-means dans les systèmes électriques modernes est présenté dans [21]. L'algorithme de clustering K-means est un type d'apprentissage automatique non supervisé qui divise l'ensemble de données non étiqueté en k clusters différents par un algorithme itératif.

L'algorithme K-means est implémenté comme suit : Étape 1 :

Choisir aléatoirement k points ou centroïdes à partir des données considérées pour initialiser les groupes ou clusters ;

Étape 2 : Pour chaque point de l'ensemble de données, calculez la distance entre le point et chaque des k centroïdes ; attribuez chaque point à son centroïde le plus proche pour former k clusters ;

Étape 3 : Remplacez le centroïde de chaque cluster par la moyenne de tous les points de données attribués au cluster ;

Étape 4 : Répétez les étapes 2 et 3 jusqu'à ce que les centroïdes ne changent plus de manière significative ou après un nombre maximum d'itérations présélectionné. Les sorties obtenues sont les derniers centroïdes du cluster et les points de données attribués aux clusters.

Bien que la méthode K-means présente de nombreux avantages, elle présente certains inconvénients qui peuvent affecter son applicabilité et ses performances. L'algorithme K-means n'est pas considéré comme ayant une bonne évolutivité pour les gros problèmes. Elle est sensible au choix de k et il est difficile de trouver le k optimal pour un ensemble de données donné. K-means converge vers un minimum local, donc différentes initialisations entraîneront des résultats différents.

Les K-means peuvent prendre beaucoup de temps avec de grands ensembles de données. Pour un ensemble de données comprenant n points de données, il doit être exécuté O(nkT) fois pour calculer les distances entre les n points de données et chacun des k centroïdes (T est le nombre d'itérations) [22]. Dans l' algorithme K-means, chaque itération prend un temps proportionnel à k et n. Cela explique pourquoi l' algorithme K-means a une faible évolutivité. Le temps d'exécution augmentera avec l'augmentation de n ou de k, ou les deux. Par conséquent, son efficacité peut être considérablement améliorée en diminuant le temps d'exécution lié à n. Pour résoudre le problème de la mauvaise évolutivité, dans [22], les auteurs développent une approche K-means-lite qui peut obtenir les centroïdes visés en un temps O(1) par rapport à n et présente un facteur d'accélération amélioré. à mesure que k et n augmentent. Il a également été démontré que la précision augmente. La technique d'inférence statistique est utilisée, dans laquelle les k centroïdes sont calculés à l'aide de quelques petits échantillons, au lieu d'une comparaison exhaustive et répétée entre les centroïdes et les points de données. Cette idée vient d'une extension intuitive du théorème central limite classique. En particulier, son utilisation ne nécessite pas de structures de données particulières, ne nécessite pas de conserver les distances calculées en mémoire et ne nécessite pas d'affectations exhaustives répétées. Il est démontré que l'utilisation de K-means-lite permet d'obtenir un gain d'efficacité drastique et peut résoudre de grands ensembles de données en temps réel ; c'est ce qu'on appelle le clustering de données avancé (ADC) dans cet article [22].

2.2. Représentation des incertitudes d'entrée

Dans cet article, pour tenir compte des incertitudes concernant la puissance de sortie des générateurs , des charges, etc., ainsi que les paramètres des composants, elles sont représentées par des distributions probabilistes. Sur la base de leurs données historiques, les distributions peuvent être estimées. Ils peuvent également être apportés par une technique de prévision, notamment dans la résolution de problèmes opérationnels.

• Génération éolienne Pour

la modélisation de la vitesse du vent, la distribution de Weibull [23] est couramment utilisée. Sa fonction de densité de probabilité (PDF) est représentée par :

$$f(v) = \frac{h}{c}\cdot\frac{v}{c}^{h-1}\cdot \exp -\frac{v}{c}^{h} \quad (1)$$

où h : le paramètre de forme ; c : le paramètre d'échelle ; v : vitesse du vent.

La courbe caractéristique de l'éolienne peut être estimée par puissance éolienne – vitesse du vent paires de données de mesure [24]. Il peut également être modélisé par une fonction par morceaux comme suit :

$$Pwo(v) = \begin{cases} 0 & v \le vci \text{ ou } v > vco \\ Mot\ \dfrac{v - vci}{\text{de passe } vr - vci} & vci < v \le vr \\ Pwr & vr < v \le vco \end{cases} \quad (2)$$

où vci, vco et vr sont respectivement la vitesse d'enclenchement, de coupure et la vitesse nominale du vent ; Pwr et Pwo sont respectivement la puissance nominale et la production de la production éolienne.

• Production solaire

La distribution du rayonnement solaire peut être estimée par ses données observées. C'est aussi généralement représenté par une distribution Beta [25], comme :

$$f(r) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \cdot \left(\frac{r}{max}\right)^{\alpha-1} \cdot 1 - \left(\frac{r}{max}\right)^{\beta-1} \quad (3)$$

où r et rmax sont respectivement les rayonnements solaires réel et maximum ; α et β sont deux paramètres principaux de la distribution ; Γ(·) est la fonction Gamma bien connue.

$$Pvo(r) = \begin{cases} Pvr \frac{2r}{rcrstd\ r} & r < rc \\ Pv_{initial} & rc \le r \le rstd \\ PVR & r > rstd \end{cases} \quad (4)$$

où rc est le rayonnement en un certain point ; rstd est le rayonnement standard (correspondant à l'environnement standard) ; Pvr et Pvo sont respectivement les puissances nominale et de sortie de l' unité photovoltaïque. La production solaire doit généralement fonctionner en mode facteur de puissance unitaire, c'est-à-dire que sa puissance réactive est égale à zéro.

• Charger

L'incertitude de chaque charge est généralement représentée par une distribution gaussienne ou normale [1]. La fonction de distribution normale est une fonction continue et l'une des fonctions les plus couramment utilisées dans la plupart des domaines.

Le PDF d'une distribution normale est le suivant :

$$f(x) = \sqrt{\frac{1}{2\pi\sigma}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (5)$$

dans laquelle μ et σ sont respectivement l'espérance (valeur moyenne) et l'écart type.

La fonction de distribution cumulative (CDF) de la fonction de distribution normale est calculée comme suit :

$$F(x) = \sqrt{\frac{1}{2\pi\sigma}} \int_{-\infty}^{x} e^{-\frac{(t-\mu)^2}{2\sigma^2}} dt \quad (6)$$

Pour modéliser une charge, la valeur attendue (moyenne) est sa puissance de base tandis que la valeur standard l'écart est supposé être égal à un certain pourcentage, par exemple 10 %, de la moyenne.

En plus des distributions de probabilité populaires présentées ci-dessus, qui sont très appropriées pour représenter les incertitudes liées aux SER et aux charges dans les systèmes électriques, il existe actuellement, dans les domaines de la probabilité et des statistiques, plusieurs autres distributions de probabilité et fonctions de densité utilisées pour incorporer les incertitudes dans le problème du PPF. En d'autres termes, le fait de supposer les fonctions de distribution ci-dessus pour les SER et les charges ne fait pas perdre la généralité de l'utilisation de la méthode PPF développée dans cette étude.

2.3. Flux de puissance probabiliste basé sur le clustering de données avancé

L'organigramme de l'approche proposée, c'est-à-dire le flux de puissance probabiliste basé sur le regroupement de données avancé (ADCPPF), utilisé pour l'évaluation probabiliste de la sécurité, est présenté dans la figure 1. Dans l'organigramme de la figure 1, la partie principale qui contribue à améliorer considérablement le temps de calcul est dans le bloc ADC.

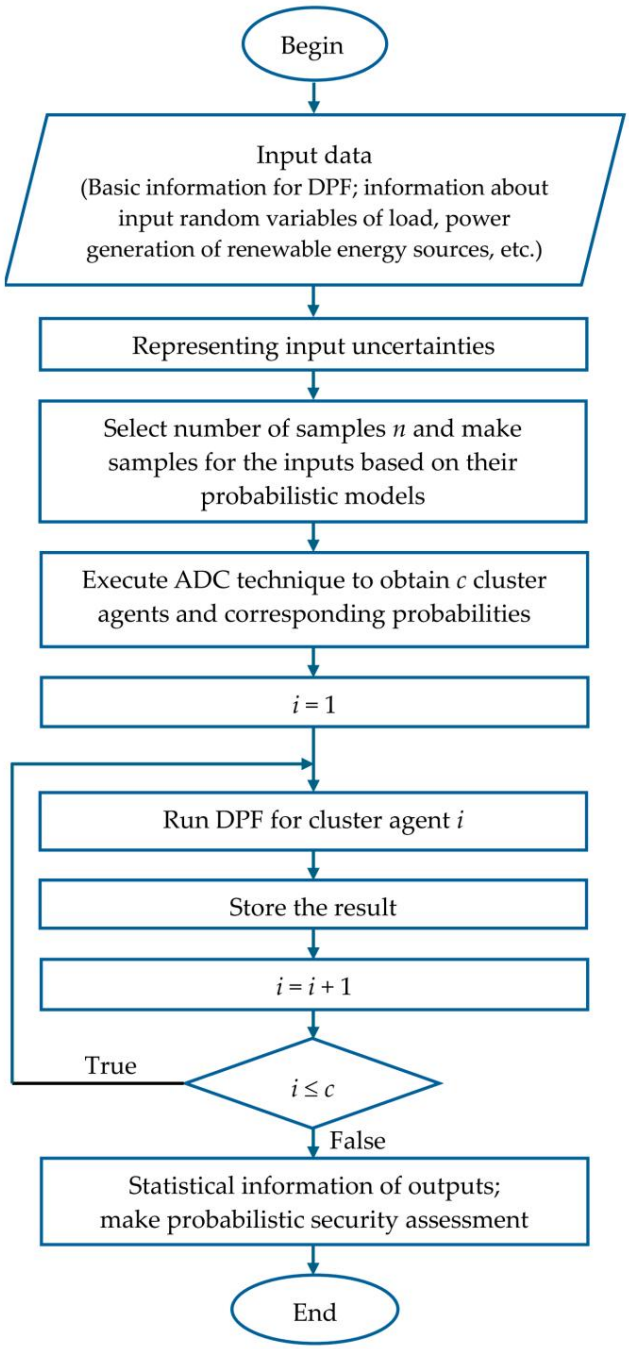


Figure 1. Organigramme de l'évaluation probabiliste de sécurité proposée.

Figure 1. Organigramme de l'évaluation probabiliste de sécurité proposée.

3. Tests et résultats

3. Tests et résultats

L'application de l'approche développée est illustrée sur un bus IEEE 300 modifié.

L'application de l'approche développée est illustrée sur un système de bus IEEE 300 modifié . Les informations nécessaires à l'analyse DPF du système, c'est-à-dire le réseau électrique, les données du bus, de la dérivation et du générateur, sont données dans [26]. Le système comprend 300 bus, 409 branches, 195 charges et 69 générateurs. Dans ce test, les incertitudes liées aux charges et aux RES sont prises en compte. Le système est modifié en ajoutant 10 centrales solaires photovoltaïques et 8 centrales éoliennes aux bus, comme indiqué respectivement dans les tableaux 1 et 2.

système. Les informations nécessaires à l'analyse DPF du système, c'est-à-dire le schéma du réseau électrique, le diagramme, le bus, la branche, les données du générateur, sont données dans [26]. Le système comprend 300 bus, branches, 195 charges et 69 générateurs. Dans ce test, les incertitudes liées aux charges et aux RES sont prises en ; sont considérés. Le système est modifié en ajoutant 10 centrales solaires photovoltaïques et 8 centrales des usines aux bus, comme le montrent respectivement les tableaux 1 et 2.

Tableau 1. Informations sur les distributions bêta de l'énergie solaire photovoltaïque.

Bus	Puissance nominale (MW)	Paramètre α	β Paramètre 8
196	50	2.5	
198	80	1.6	9
203	40	1.2	6

Tableau 1. Informations sur les distributions bêta de l'énergie solaire photovoltaïque.

Bus	Puissance nominale (MW)	Paramètre α	Paramètre β
196	50	2.5	8
198	80	1.6	9
203	40	1.2	6
204	60	2.2	8
215	70	3.2	7
217	50	3.5	dix
221	35	4.2	11
229	90	2.8	8
245	65	1.9	7
246	95	3.1	8

Tableau 2. Informations sur les distributions Weibull de l'énergie éolienne.

Bus	Puissance nominale (MW)	Paramètre d'échelle	Paramètre de forme
118	90	dix	2.4
121	80	15	1.6
126	100	11	2.4
142	50	14	1,5
154	40	20	2.2
156	60	28	1.8
159	70	16	1.7
161	95	12	2.3

Par souci de simplicité, mais sans perte de généralité, les incertitudes des charges et les RES sont supposées connues par les techniques de prévision. L'incertitude de chaque charge est modélisé par une distribution normale avec une valeur attendue égale à la valeur de base et écart type égal à 10% de la valeur attendue. L'énergie solaire photovoltaïque l'incertitude au niveau de chaque centrale est supposée suivre la distribution bêta, ses paramètres étant donnés dans le tableau 1. Les distributions sont également supposées être corrélées à un coefficient de corrélation de 0,7. Dans cette étude, on suppose que le scénario de simulation se déroule pendant les heures de clarté, avec des conditions potentielles pour une certaine production d'énergie solaire photovoltaïque. De plus, nous supposer que les charges et les centrales solaires et éoliennes ne sont pas soumises à des conditions météorologiques anormales conditions (par exemple, vague de chaleur extrême, tempête hivernale à grande échelle et ouragans), extrêmement les phénomènes rares (par exemple, une éclipse solaire) et les catastrophes (par exemple, la guerre, les tremblements de terre et les tsunamis). Certains de ces événements anormaux sont extrêmement difficiles à prévoir avec des normes de faible qualité. déviation. Ces événements anormaux et rares peuvent également être décrits par une distribution appropriée fonctions et inclus dans le problème PPF. Toutefois, cette question sort du cadre du étude actuelle et est destiné à être pris en compte dans les études futures. De même, l'incertitude de la puissance produite par chaque parc éolien est supposée avoir des distributions de Weibull, avec la paramètres présentés dans le tableau 2. Les distributions sont corrélées à un coefficient de corrélation de 0,8.

Dans le test actuel, la puissance de base de 100 MVA est utilisée. Les résultats MCS sont utilisés comme la référence pour évaluer les résultats obtenus par d'autres méthodes. Tous les tests sont exécutés dans Matlab (R2015b) sur un processeur Intel Core i5 2,53 GHz et 4,00 Go de RAM.

Le calcul PPF est effectué pour obtenir tous les résultats intéressants des RV de sortie dans termes de PDF et/ou CDF. À des fins d'illustration, les distributions d'un certain nombre des RV de sortie sélectionnés sont affichés. La figure 2 représente les CDF du PF réel à travers les branches 2 à 8.

Électronique 2024, 13, x POUR EXAMEN PAR LES PAIRS 8 sur 13
Electronique 2024, 13, x POUR EXAMEN PAR LES PAIRS 8 sur 13

(c'est-à-dire noté P2–8), tandis que celle de la tension sur le bus 89 (c'est-à-dire noté V89) est donnée dans la figure 3. Les figures montrent que l'approche ADCPPF développée peut correspondre bien avec les courbes de MCS, indiquant la bonne performance de l'approche ADCPPF.

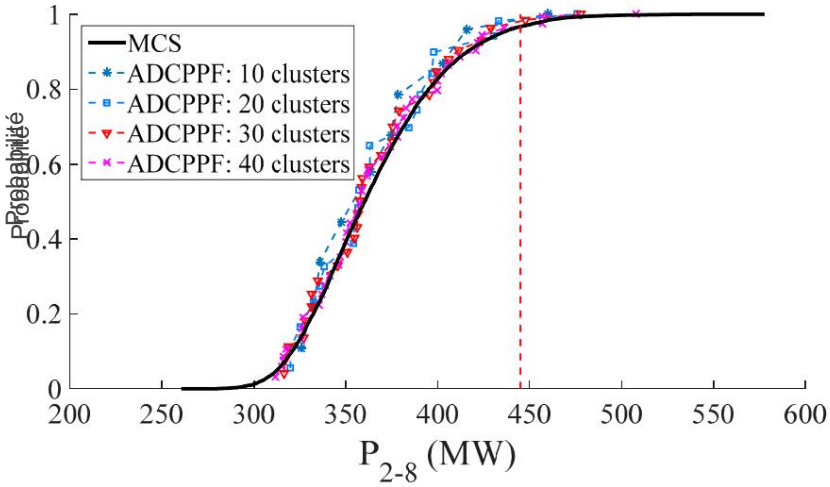


Figure 2. CDF du PF actif via les branches 2 à 8.

Figure 2. CDF du PF actif via les branches 2 à 8.

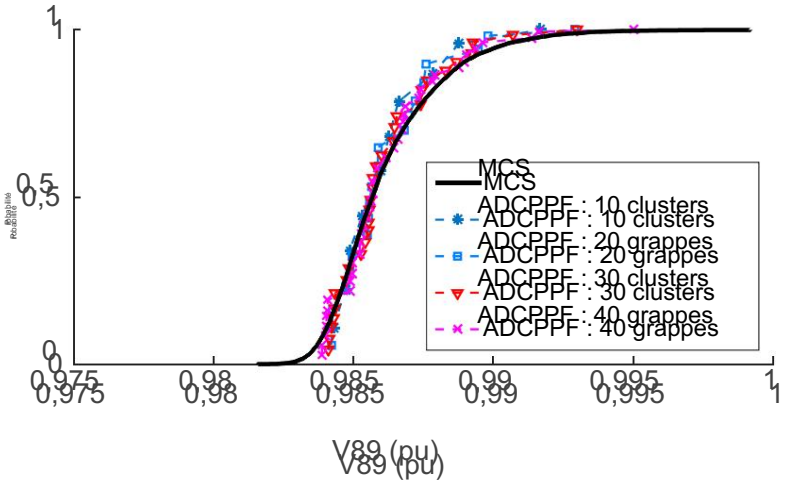


Figure 3. CDF de tension sur le bus 89.

Figure 3. CDF de tension sur le bus 89.

Parce qu'il est difficile d'observer clairement en traçant tous les résultats sur la même figure. Parce qu'il est difficile d'observer clairement en traçant tous les résultats sur la même figure, seuls les résultats correspondant à la méthode ADCPPF sont représentés. Toutefois, seuls les résultats correspondant à la méthode ADCPPF sont représentés. Cependant, pour figurer, seuls les résultats démontrons son efficacité, nous comparons également les résultats du PPF utilisant l'ADC avec K- pour démontrant son efficacité, nous comparons également les résultats du PPF utilisant l'ADC avec le clustering K-means. K signifie regroupement.

Comme indiqué dans la section 2.1, le regroupement de K-moyennes peut prendre beaucoup de temps avec de grands ensembles de données en raison du faible problème d'évolutivité. Différent des K-means, les K-means-lite l'approche K-means-lite a été développée dans [16] et peut donner des résultats très rapidement. Son précision a également que sa précision augmente par rapport à la technique des K-means. Par conséquent, il a été démontré que cela Il a également été démontré qu'elle augmente par rapport à la technique K-means. Par conséquent, dans ce test, nous ne nous concentrons pas sur la preuve de l'exactitude de la méthode ADC par rapport à la méthode K-means. Au lieu de cela, l'accent est mis sur les performances du temps de traitement, indiquant ainsi la méthode. Au signifie méthode. Au lieu de cela, la performance du temps de traitement est axée sur la performance, indiquant ainsi que l'approche proposée a une bonne applicabilité pour résoudre des problèmes dans le délai d'exploitation que l'approche proposée a une bonne applicabilité pour résoudre des problèmes dans le délai d'exploitation du système électrique. Il est clairement démontré dans le tableau 3 que l'approche ADCPPF peut alimenter du fonctionnement du système électrique. Il est clairement montré dans le tableau 3 que l'approche ADCPPF peut donner le résultat en quelques secondes, en comparaison aux centaines de secondes nécessaires pour donne le résultat en quelques secondes, en comparaison aux centaines de secondes nécessaires au MCS. En particulier, comme mentionné précédemment, pour un ensemble de données comprenant n points de MCS. En particulier, comme mentionné précédemment, pour un ensemble de données comprenant n points de données, le clustering K-means a besoin d'un temps égal à O (nk1) pour s'exécuter (dans lequel T est le nombre - le signifie que le clustering a besoin d'un temps égal à O (nk1) pour s'exécuter (dans lequel T est le nombre d'itérations). Le temps d'exécution des K-moyennes augmentera fortement avec l'augmentation de n et/ou bre d'itérations). Le temps d'exécution des K-means augmentera fortement avec l'augmentation de n et/ou k. Le système de bus IEEE 300 modifié est un système à grande échelle, donc basé sur K-means et/ou k. Le système de bus IEEE 300 modifié est un système à grande échelle, donc le PPF basé sur K-means fonctionne très dur et prend beaucoup de temps.
PPF fonctionne très dur et prend beaucoup de temps.

k. Le système de bus IEEE 300 modifié est un système à grande échelle, donc PPF basé sur K-means fonctionne très dur et prend beaucoup de temps.

Tableau 3. Comparaison des temps d'exécution.

Méthode	Temps (s)
MCS	725
ADCPPF avec 5 clusters	2,64
ADCPPF avec 10 clusters	2,79
ADCPPF avec 20 clusters	2,98
ADCPPF avec 30 clusters	3.26
ADCPPF avec 40 clusters	3.47
ADCPPF avec 50 clusters	3,72
ADCPPF avec 70 clusters	4.21
PPF basé sur K-means avec 5 clusters	140
PPF basé sur K-means avec 10 clusters	420

D'après les figures 2 et 3 et le tableau 3, la précision de l'ADCPPF augmente (c'est-à-dire intuitivement, la courbe correspondante des figures 2 et 3 suit de plus près la courbe MCS) et le temps requis pour exécuter la méthode augmente également avec un nombre croissant de groupes. En comparant MCS utilisant K-means et ADCPPF, K-means est un défi dans ce domaine cas et entraîne une augmentation du temps de calcul bien plus importante que lors de l'utilisation d'ADCPPF. A travers l'analyse ci-dessus, il est montré que la méthode ADCPPF présente l'avantage de à la fois une précision relativement élevée et un temps d'exécution considérablement réduit.

PPF peut fournir des distributions pour les RV de sortie qui sont bonnes pour la sécurité du système électrique évaluation. La probabilité de sous/surtension, de surcharge de ligne, etc. peut être jugé. Par exemple, la limite supérieure du PF réel de la branche 2 à 8 est supposée être égale à 445 MW, soit correspondant à la ligne verticale de la figure 2, et la probabilité que la puissance transmis via la branche dépasse sa limite peut être calculé comme suit :

$$P\{P2-8 > 445\} = 1,4\ \% \quad (7)$$

De même, la probabilité que la tension sur un bus considéré soit hors de la plage de fonctionnement peuvent être évalués. Cependant, dans ce test, les tensions sur tous les bus du système (par exemple, V89 sur la figure 3) sont dans la plage, c'est-à-dire [0,9, 1,1] pu

Les résultats obtenus par la méthode ADCPPF peuvent aider l'opérateur du système à évaluer les états de fonctionnement afin de prendre des décisions adaptées et de proposer des solutions pour le système.

4. Discussion plus approfondie sur l'applicabilité des diverses méthodes PPF

Les méthodes PPF peuvent être sélectionnées pour être appliquées à la fois aux problèmes de planification et d'exploitation. des systèmes électriques.

- Application de PPF aux problèmes de planification : la méthode MCS convient à la résolution de problèmes de planification. problèmes avec des délais longs (tels que des années, des saisons, des mois, des semaines) ou opérationnels problèmes de planification en quelques jours. Dans de tels cas, le temps nécessaire pour atteindre les résultats n'ont pas besoin d'être très rapides. En plus des données de configuration du réseau, les données sur les sources (notamment SER) et les charges collectées sur de longues périodes, c'est-à-dire des mois, un année, ou plusieurs années, sont utilisées pour estimer les PDF. Ces données peuvent également être utilisées par un technique de prévision pour fournir des résultats pour le fonctionnement du système. Si la prévision technique suit l'approche de prévision ponctuelle, les résultats de prévision sont fournis sous forme une valeur définie à chaque instant de prévision et une erreur correspondante. Ces valeurs sont pris en compte l'espérance et l'écart type de la fonction de distribution normale.

Ces fonctions sont les informations d'entrée du problème MCS. Si la prévision la technique suit une approche de prévision de l'incertitude telle que la prévision probabiliste ou des approches de prévision de scénarios [24], alors la fonction de distribution de probabilité sera plus utile. En fait, lors de l'application pratique du MCS, si le système est trop grand, avec beaucoup de variables d'entrée, le traitement du MCS est très difficile, prenant un temps extrêmement une quantité de mémoire extrêmement importante, ce qui rend le traitement très difficile, et, une grande quantité de mémoire, le rend très difficile à traiter et, dans de nombreux cas, cela peut dans de nombreux cas, cela peut ne pas être possible. Dans de tels cas, le regroupement (comme dans le document actuel) n'est pas possible. Dans de tels cas, le clustering (comme dans l'article actuel) et la dimensionnalité et des techniques de réduction de dimensionnalité doivent être utilisées. des techniques de réduction doivent être utilisées.

• Application aux problèmes opérationnels : Pour les problèmes opérationnels avec des temps extrêmement courts. • Application aux problèmes opérationnels : Pour les problèmes opérationnels avec des temps extrêmement courts. des délais allant de quelques minutes (c'est-à-dire un délai à très court terme) à quelques heures des délais allant de quelques minutes (c'est-à-dire un délai à très court terme) à quelques heures (c'est-à-dire un délai à court terme) et dans les 24 heures (pour à venir), méthodes analytiques et approximation (c'est-à-dire cadre à court terme) et dans les 24 h (jour à venir), méthodes analytiques et approximation (cadre des méthodes d'imagerie avec un temps de traitement rapide peuvent être appliquées. La méthode MCS peut être appliquée à des méthodes avec un temps de traitement rapide. La méthode MCS peut également être utilisée également être utilisé dans un laps de temps de plusieurs heures ou plus lorsque le système électrique est dans un laps de temps de plusieurs heures ou plus lorsque le système électrique est de petite taille à petite échelle et le nombre de RV d'entrée est également faible. La méthode MCS combinée au clustering RV d'entrée sont également faibles. La méthode MCS combinée et le nombre de avec des techniques de clustering, telles que celle proposée dans cet article, peuvent être mises en œuvre des techniques, telles que celle proposée dans cet article, peuvent être mises en œuvre avec de courtes avec des délais de fonctionnement courts, voire extrêmement courts. Basé sur les données collectées auprès des RES très chargés (ou des charges (en plus de configurations de réseau), distributions de probabilités ou les moments de l'entrée du problème opérationnel sont construits. Ces fonctions peuvent aux moments du problème opérationnel sont construites. les fonctions également être assurée par une technique de prévision. Ces fonctions peuvent également être assurées par une technique de prévision.

La partie suivante est un exemple illustratif. Supposons une technique de prévision de la charge. La partie suivante est un exemple illustratif. Supposons une technique de prévision de la charge ou la production d'électricité des SER fournit des informations prévisionnelles à différents moments de la production d'électricité des SER fournit des informations prévisionnelles à différents moments de la ou la délai opérationnel. À cette époque, les résultats obtenus grâce aux méthodes PPF seront dans un délai très opérationnel. À cette époque, les résultats obtenus par les méthodes PPF seront très utile pour l'opérateur du système. utile pour l'opérateur du système.

La figure 4 illustre la PDF du courant circulant sur une ligne d'intérêt à des vitesses consécutives. La figure 4 illustre la PDF du courant circulant sur une ligne d'intérêt à des instants consécutifs (résolution de 1 h) sur l'horizon temporel de 24 h de fonctionnement journalier du système. (résolution de 1 h) pour l'horizon de fonctionnement journalier du système. L'opérateur du système aura une image très claire de l'insécurité du système, de sorte que l'opérateur L'opérateur du système aura une très claire de l'insécurité du système afin qu' ils peuvent proposer des solutions appropriées. Sur la base de la comparaison de la courbe PDF avec la qu'ils peuvent proposer des solutions appropriées. Sur la base de la comparaison de la courbe PDF limite maximale avec la limite maximale Imax (correspondant à la limite de la puissance) de la ligne d'intérêt, il est Imax (correspondant à la limite de la puissance) de la ligne d'in- possible de déterminer le point où il existe un risque que le courant la traverse. terêt, il est possible de déterminer le point où il existe un risque que la ligne de passage actuelle dépasse la valeur admissible. à travers cette ligne dépassant la valeur autorisée.

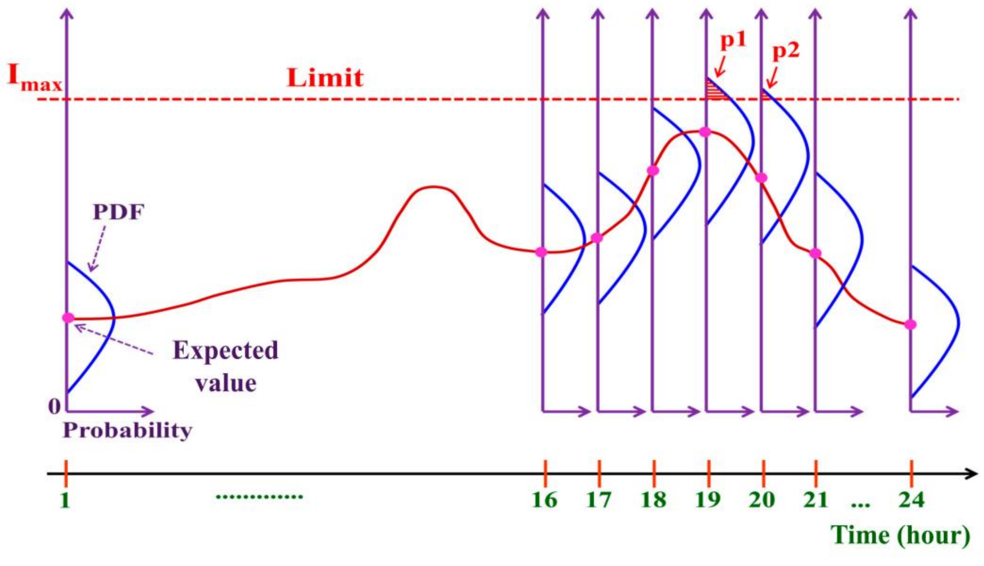


Figure 4. L'exemple illustre la PDF du courant circulant sur une ligne intéressée à des moments consécutifs. Figure 4. L'exemple illustre la PDF du courant circulant sur une ligne intéressée à des moments consécutifs. heures au cours de l'horizon temporel de 24 heures de fonctionnement journalier du système. heures au cours de l'horizon temporel de 24 heures de fonctionnement journalier du système.

Par exemple, sur la figure 4, de 16h00 à 18h00, le courant augmente progressivement, et, par Exemple, sur la figure 4, de 16h00 à 18h00, le courant augmente progressivement et, près de vers 19h00, la limite supérieure de la ligne PDF commence à toucher et dépasser la limite Imax et 19h00, la limite supérieure de la ligne PDF commence à toucher et dépasser la limite Imax. et la probabilité calculée qui la dépasse est p1. Il convient de noter que la probabilité calculée qui et la probabilité à 19h00 est p1. Il convient de noter que, comme le soir approche, la source d'énergie solaire diminue progressivement, et dans cet exemple, vers 19h00, cette source ne produit plus d'électricité. Cependant, il s'agit de la période de pointe du système où la charge augmente rapidement jusqu'à atteindre le pic. À l'heure actuelle, l'incertitude dans le système provient des charges et d'autres sources, le cas échéant, et non de la source d'énergie solaire. Puis, à 20h00, le courant tend à diminuer et le niveau d'intrusion a

smaer an em max et saey rss ue oe overoang ne sont pas vus. C'est un point très nouveau dans la perspective de l'évaluation de la sécurité des systèmes.

La figure 5 est un exemple illustratif d'évaluation du risque de surtension/sous-tension sur un bus d'intérêt. Dans cet exemple, le risque de sous-tension aux heures 19h00 et 20h00 est calculé respectivement comme p1 et p2.

La méthode de calcul et d'analyse de la sécurité du système est basée sur des informations et des données obtenues à partir de quantités d'entrée aléatoires du problème telles que les charges et les puissances de sortie des RES. Le résultat est qu'en termes de distributions de probabilité de production en soirée, la source d'énergie solaire diminue progressivement, et ne produit plus d'électricité. Cependant, c'est la période de pointe du système dans cet exemple, de des grandeurs telles que la tension des nœuds, ou la puissance ou le courant transmis sur les branches. À 19h00, cette source Le processus est effectué avant l'opération réelle pour trouver les distributions de probabilité où la charge augmente rapidement jusqu'à atteindre le pic. À l'heure actuelle, l'incertitude dans systèmes intégrant des charges et d'autres sources, le cas échéant, et non de l'énergie solaire 20h00, Pour les systèmes électriques réels dotés d'un système SCADA EMS, ce système fournira une source interne. Puis, à informations sur les paramètres de mode et être mis à jour régulièrement en temps quasi réel, donc mon- de p2. Ainsi, dans le la surveillance du fonctionnement des paramètres de mode degrés est effectué en continu. En plus, dure environ 1 h. Le temps de De plus, lorsqu'il existe un système SCADA EMS, les données collectées pour des facteurs aléatoires seront plus petites (30 min, 15 min, etc.), et le système plus pratique et continuellement mis à jour. Lorsque l'ensemble de données de facteurs aléatoires est plus complète, les informations obtenues sont plus claires. C'est également un autre avantage lors de l'utilisation d'un système Système SCADA EMS combiné à la méthode PPF. Lorsqu'il existe un risque de surcharge pendant une longue période et que le niveau de surcharge est une certaine précision acceptable mais qui fournit également des résultats très rapides pour répondre à la valeur application dans un laps de temps quasi « temps réel » d'exploitation du système électrique. Comme les résultats obtenus avec le DPF traditionnel), toutes les valeurs actuelles calculées sont plus petites de sécurité dus à la surcharge ne sont « pas visibles ». C'est un très nouveau point du point de vue de l'évaluation de la sécurité du système. exemple illustratif d'évaluation du risque de surtension/sous-tension sur un bus. exemple, le risque de sous-tension aux heures 19h00 et 20h00 est calculé. telles que p1 et p2, respectivement.

L'objectif principal de cet article est de développer une approche de calcul du PPF qui garantit une sécurité sévère, souhaitée. Il convient de noter que, sur la figure 4, si l'on se soucie uniquement de la valeur attendue (similaire à mentionné ci-dessus, la méthode MCS fait face à de nombreux défis et est même impossible lorsque la limite Imax et les risques appliqué aux systèmes électriques réels, en particulier aux systèmes à grande échelle avec un fonctionnement très court, délais. L'algorithme de clustering proposé pour être appliqué pour résoudre le problème PPF. La figure 5 est un dans ce document est simple, facile à mettre en œuvre et aide à faire face efficacement à la faible échelle d'intérêt. Dans cet capacité de l'algorithme K-means. Par conséquent, ADCPPF peut donner des résultats rapides avec des données volumineuses qui peut être appliqué pour résoudre les problèmes de PPF pour les systèmes électriques à grande échelle.

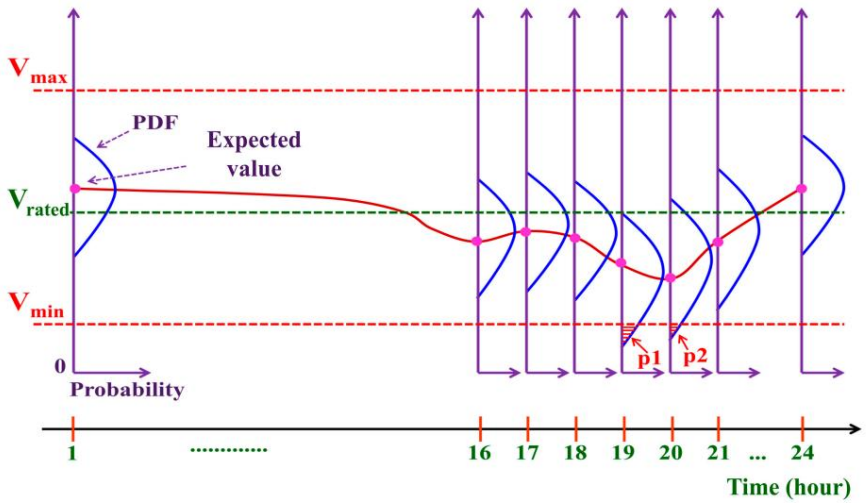


Figure 5. L'exemple illustre la PDF de la tension à un nœud d'intérêt à des instants consécutifs dans l'horizon temporel de 24 heures pour l'exploitation journalière du système électrique.

La méthode de calcul et d'analyse de la sécurité du système est basée sur informations et données obtenues à partir de quantités d'entrée aléatoires du problème telles que le charges et puissances de sortie des RES. Le résultat est en termes de distributions de probabilité de grandeurs de sortie telles que la tension du nœud, ou la puissance ou le courant transmis sur les branches. Ce processus est effectué avant l'opération réelle pour trouver les distributions de probabilité de les quantités qui nous intéressent.

Pour les systèmes électriques réels dotés d'un système SCADA EMS, ce système fournira informations sur les paramètres de mode et être mis à jour régulièrement en temps quasi réel, donc la surveillance du fonctionnement des paramètres de mode degrés est effectuée en permanence. Dans De plus, lorsqu'il existe un système SCADA EMS, les données collectées pour des facteurs aléatoires seront plus pratique et continuellement mis à jour. Lorsque l'ensemble de données de facteurs aléatoires est plus complète, les informations obtenues sont plus claires. C'est également un autre avantage lors de l'utilisation d'un Système SCADA EMS combiné à la méthode PPF.

L'objectif principal de cet article est de développer une approche de calcul du PPF qui garantit une certaine précision acceptable mais qui fournit également des résultats très rapides pour répondre aux exigences souhaitées

application dans un laps de temps quasi « temps réel » d'exploitation du système électrique. Comme mentionné ci-dessus, la méthode MCS est confrontée à de nombreux défis et est même impossible lorsqu'elle est appliquée à des systèmes électriques réels, en particulier des systèmes à grande échelle avec des délais d'exploitation très courts . L'algorithme de clustering proposé dans cet article pour résoudre le problème PPF est simple, facile à mettre en œuvre et permet de gérer efficacement la faible évolutivité de l'algorithme K-means. Par conséquent, ADCPPF peut donner des résultats rapides avec des données volumineuses qui peuvent être appliquées pour résoudre les problèmes de PPF pour les systèmes électriques à grande échelle.

Cependant, en plus des avantages et des contributions de la méthode proposée dans des applications pratiques, elle présente également une limitation qui doit être résolue : la prédétermination de la valeur de k comme dans l'algorithme des K-moyennes. De plus, les limites inhérentes au PPF, qui ne couvre pas les transitoires électromagnétiques, et les conditions de fonctionnement anormales extrêmes restent des problèmes non résolus. La planification et l'exploitation du système électrique doivent également se concentrer sur les pires scénarios futurs et anormaux (par exemple, vagues de chaleur, tempêtes hivernales, etc.). Les événements météorologiques extrêmes exacerbés par le changement climatique et le réchauffement planétaire suscitent un niveau élevé d'incertitude. Par conséquent, une plus grande diversité de types d'incertitude devrait être envisagée pour être intégrée au problème PPF. Les systèmes de stockage d'énergie (par exemple, les batteries, l'hydroélectricité par pompage et les systèmes de transport de véhicules électriques vers le réseau) peuvent atténuer l'incertitude liée aux SER qui n'est pas non plus prise en compte dans cette étude. Le matériel est l'un des facteurs importants affectant l'analyse PPF, notamment en ce qui concerne le temps de calcul. Les recherches futures pourraient également se concentrer sur le traitement de l'algorithme dans des ordinateurs hautes performances, en surmontant les problèmes de traitement du temps et en permettant de se concentrer davantage sur la précision du modèle. Ces limites ouvrent des sujets à considérer dans de futures recherches.

5. Conclusions

PPF est un outil efficace pour calculer et analyser les systèmes électriques et prendre en compte les incertitudes existant dans le système. Cela peut aider l'opérateur du système à évaluer la sécurité. Parmi les différentes techniques de PPF, la MCS donne des résultats très précis mais nécessite souvent beaucoup de calculs. Cet article se concentre sur la résolution du problème de la réduction du temps de calcul pour la simulation Monte Carlo afin d'obtenir un outil pratique avec une grande précision qui donne des résultats de calcul rapides à utiliser pour résoudre des problèmes dans l'horizon temporel des opérations du système électrique. Pour atteindre cet objectif, nous utilisons une technique avancée de regroupement de données appelée K-means-lite. L'approche développée, ADCPPF, est testée de manière approfondie sur un système de bus IEEE-300 modifié, montrant de bonnes performances en matière de réduction du temps de calcul.

Contributions des auteurs : Méthodologie, NTAN, DDL, VDN, VKP et VKH ; logiciels, NTAN et DDL ; Tous les auteurs ont écrit et édité le manuscrit. Tous les auteurs ont lu et accepté la version publiée du manuscrit.

Financement : Cette recherche a été financée par le ministère de l'Éducation et de la Formation du Vietnam sous le numéro de projet CT2022.07.DNA.03.

Déclaration de disponibilité des données : les données sont contenues dans l'article.

Conflits d'intérêts : Les auteurs ne déclarent aucun conflit d'intérêts.

Les références

1. Le, DD; Berizzi, A. ; Bovo, C. Une approche probabiliste d'évaluation de la sécurité des systèmes électriques avec des ressources éoliennes intégrées. Renouveler. Énergie 2016, 85, 114-123. [Référence croisée]

2. Borkowska, B. Flux de charge probabiliste. IEEETrans. Application de puissance. Système. 1974, PAS-93, 752-759. [Référence croisée]

3. Allan, infirmier autorisé ; Al-Shakarchi, MRG Techniques probabilistes dans l'analyse de flux de charge AC. Proc. Inst. Élire. Ing. 1977, 124, 154-160. [Référence croisée]

4. Zhang, P. ; Lee, ST Calcul probabiliste du flux de charge en utilisant la méthode des cumulants combinés et du développement de Gram-Charlier. IEEETrans. Système d'alimentation. 2004, 19, 676-682. [Référence croisée]

5. Fan, M. ; Vittal, V. ; Heydt, GT; Ayyanar, R. Études probabilistes de flux de puissance pour les systèmes de transmission photovoltaïques Génération à l'aide de cumulants. IEEETrans. Système d'alimentation. 2012, 27, 2251-2261. [Référence croisée]

6. Le, DD; Berizzi, A. ; Bovo, C. ; Ciapessoni, E. ; Cirio, D. ; Pitto, A. ; Gross, G. Une approche probabiliste de l'évaluation de la sécurité du système électrique dans des conditions d'incertitude. Dans Actes de la dynamique et du contrôle du système électrique en vrac — IX Optimisation, sécurité et contrôle du réseau électrique émergent (Symposium IREP 2013), Crète, Grèce, 25-30 août 2013.
7. Su, CL Calcul probabiliste de flux de charge à l'aide de la méthode d'estimation ponctuelle. IEEETrans. Système d'alimentation. 2005, 20, 1843-1851. [Référence croisée]
8. Mohammadi, M. ; Shayegani, A. ; Adaminejad, H. Une nouvelle approche de la méthode d'estimation ponctuelle pour le flux de charge probabiliste. Int. J. Electr. Pouvoir 2013, 51, 54-60. [Référence croisée]
9. Liu, Y. ; Gao, S. ; Cui, H. ; Yu, L. Flux de charge probabiliste considérant les corrélations de variables d'entrée suivant des distributions arbitraires. Electr. Système d'alimentation. Rés. 2016, 140, 354-362. [Référence croisée]
10. Yu, H. ; Chung, CY; Wong, KP; Lee, HW; Zhang, JH Évaluation probabiliste du flux de charge avec échantillonnage d'hypercube latin hybride et décomposition de Cholesky. IEEETrans. Système d'alimentation. 2009, 24, 661-667. [Référence croisée]
11. Zou, B. ; Xiao, Q. Résolution du problème de flux de puissance optimal probabiliste à l'aide de la méthode quasi Monte Carlo et du neuvième ordre Transformation normale polynomiale. IEEETrans. Système d'alimentation. 2014, 29, 300-306. [Référence croisée]
12. Hajian, M. ; Rosehart, DEO ; Zareipour, H. Flux de puissance probabiliste par simulation de Monte Carlo avec échantillonnage Latin Supercube. IEEETrans. Système d'alimentation. 2013, 28, 1550-1559. [Référence croisée]
13. Leite da Silva, AM; Milhorance de Castro, A. Évaluation des risques dans l'écoulement de charge probabiliste via la simulation de Monte Carlo et Méthode d'entropie croisée. IEEETrans. Système d'alimentation. 2018, 14, 1193-1202. [Référence croisée]
14. Carpinelli, G. ; Caramia, P. ; Varilone, P. Méthode de simulation Monte Carlo multilinéaire pour le flux de charge probabiliste de distribution systèmes avec des systèmes de production éolienne et photovoltaïque. Renouveler. Énergie 2015, 76, 283-295. [Référence croisée]
15. Hagh, MT; Amiyan, P. ; Galvani, S. ; Valizadeh, N. Flux de charge probabiliste utilisant le clustering d'optimisation d'essaim de particules méthode. IET Gén. Transm. Distribuer. 2018, 12, 780-789. [Référence croisée]
16. Khalghani, M. ; Ramezani, M. ; Mashhadi, MR Flux d'énergie probabiliste basé sur la simulation de Monte-Carlo et le regroupement de données pour analyser le système électrique à grande échelle, y compris le parc éolien. Dans Actes de la conférence IEEE Kansas Power and Energy (KPEC) 2020, Manhattan, KS, États-Unis, 13 et 14 juillet 2020.
17. Haschisch, MS ; Hasanien, HM; Ji, H. ; Alkuhayli, A. ; Alharbi, M. ; Akmaral, T. ; Turquie, RA ; Jurado, F. ; Badr, AO Simulation de Monte Carlo et technique de clustering pour résoudre le problème de flux de puissance optimal probabiliste pour les systèmes hybrides d'énergie renouvelable. Durabilité 2023, 15, 783. [CrossRef]
18. Zhang, R. ; Chen, Z. ; Li, Z. ; Jiang, T. ; Li, X. Fonctionnement robuste en deux étapes de micro-réseaux multi-énergies intégrés électricité-gaz-chaleur en considérant des incertitudes hétérogènes. Appl. Énergie 2024, 371, 123690. [CrossRef]
19. Jiang, Y. ; Ren, Z. ; Li, W. Région d'exploitation des émissions de carbone engagées pour les systèmes énergétiques intégrés : concepts et analyses. IEEETrans. Soutenir. Énergie 2024, 15, 1194-1209. [Référence croisée]
20. Li, Z. ; Xu, Z. ; Wang, P. ; Xiao, G. Restauration d'un système de distribution multi-énergies avec reconfiguration du réseau de district commun via une programmation stochastique distribuée. IEEETrans. Réseau intelligent 2024, 15, 2667-2680. [Référence croisée]
21. Miraftabzadeh, SM; Colombo, CG ; Longo, M. ; Foiadelli, F. K-Means et méthodes de clustering alternatives dans les systèmes électriques modernes. Accès IEEE 2023, 11, 119596-119633. [Référence croisée]
22. Olukanmi, PO ; Nelwamondo, F. ; Marwala, T. k-Means-Lite : Clustering en temps réel pour les grands ensembles de données. Dans Actes de la 5e Conférence internationale sur le soft computing et l'intelligence artificielle, Nairobi, Kenya, 21 et 22 novembre 2018 ; p. 54-59.
23. Wang, X. ; Chiang, HD ; Wang, J. ; Liu, H. ; Wang, T. Analyse de stabilité à long terme des systèmes électriques utilisant l'énergie éolienne basée sur des équations différentielles stochastiques : développement et fondations de modèles. IEEETrans. Soutenir. Énergie 2015, 6, 1534-1542. [Référence croisée]
24. Le, DD; Gross, G. ; Berizzi, A. Modélisation probabiliste de la production de parcs éoliens multisites pour des applications basées sur des scénarios. IEEE Trans. Soutenir. Énergie 2015, 6, 748-758. [Référence croisée]
25. Salamé, ZM ; Borowy, BS; Amin, ARA Adaptation module photovoltaïque-site basée sur les facteurs de capacité. IEEETrans. Convertisseurs d'énergie . 1995, 10, 326-332. [Référence croisée]
26. Archives des scénarios de test du système électrique. Disponible en ligne : http://labs.ece.uw.edu/pstca/pf300/pg_tca300bus.htm (consulté le 6 juin 2024).

électronique

MDPI

Article

Un système d'imagerie radar à synthèse d'ouverture inverse rapide Combinant la prise de vue accélérée par GPU et le rayon rebondissant et Algorithme de rétroprojection sous larges bandes passantes et angles

Jiongming Chen [1], Pengju Yang [1,2,*], Rong Zhang [1] et Rui Wu [1,3]

[1] École de physique et d'information électronique, Université de Yan'an, Yan'an 716000, Chine ; jmchen@yau.edu.cn (JC); zr2108700733@yau.edu.cn (RZ); wurui@yau.edu.cn (RW)
[2] Laboratoire clé pour la science de l'information sur les ondes électromagnétiques (MoE), Université Fudan, Shanghai 200433,
[3] Institut chinois de technologie et de logiciels de radiofréquence, Institut de technologie de Pékin, Pékin 100081, Chine
* Correspondance : pjyang@fudan.edu.cn

Résumé : Les techniques d'imagerie radar à synthèse inverse (ISAR) sont fréquemment utilisées dans les applications de classification et de reconnaissance de cibles, en raison de leur capacité à produire des images haute résolution pour des cibles en mouvement. Afin de répondre à la demande de l'imagerie ISAR pour le calcul électromagnétique avec une efficacité et une précision élevées, une nouvelle méthode de tir et de rebond accélérés (SBR) est présentée en combinant une unité de traitement graphique (GPU) et une structure arborescente de hiérarchies de volumes limites (BVH) . Pour surmonter le problème des images floues par une procédure ISAR basée sur Fourier dans des conditions de grand angle et de large bande passante, un algorithme d'imagerie par rétroprojection (BP) parallèle efficace est développé en utilisant la technique d'accélération GPU. Le SBR accéléré par GPU présenté est validé par comparaison avec la méthode RL-GO dans le logiciel commercial FEKO v2020. Pour les images ISAR, il est clairement indiqué que de forts centres de diffusion ainsi que des profils de cibles peuvent être observés sous de grands angles, et de larges bandes passantes, 3 GHz. C'est d'azimut d'observation. Δφ= 90◦ indique également que l'imagerie ISAR est très sensible aux angles d'observation. De plus, des lobes secondaires évidents peuvent être observés, en raison de la distorsion de l'histoire des phases de l'onde électromagnétique résultant de la diffusion multipolaire. Les résultats de simulation confirment la faisabilité et l'efficacité de notre système en combinant le SBR accéléré par GPU avec l'algorithme BP pour une simulation d'imagerie ISAR rapide dans des conditions de grand angle et de large bande passante.

Mots-clés : imagerie ISAR ; méthode de tir et de rebond des rayons ; Accélération GPU ; arbre BVH ; algorithme de rétroprojection

Citation : Chen, J. ; Yang, P. ; Zhang, R. ; Wu, R. Un système d'imagerie radar à ouverture synthétique inverse rapide combinant un algorithme de prise de vue accéléré par GPU et de rayons rebondissants et de rétroprojection sous des largeurs de bande et des angles larges. Électronique 2024, 13, 3062. https://doi.org/10.3390/électronique13153062

Rédacteur académique : Chiman Kwan

Reçu : 4 juin 2024
Révisé : 29 juillet 2024
Accepté : 30 juillet 2024
Publié : 2 août 2024

1. Introduction

Le radar à synthèse d'ouverture inverse (ISAR) est un puissant système radar d'imagerie à micro-ondes actif largement utilisé dans les applications militaires et civiles en raison de sa capacité à produire des images haute résolution pour des cibles mobiles dans presque tous les temps et dans des conditions quotidiennes [1-3] . Les images ISAR peuvent être obtenues en focalisant les données du champ de diffusion sous plusieurs angles et fréquences, qui sont une représentation bidimensionnelle du centre de diffusion cible [4–6]. La simulation de l'imagerie ISAR pour de grandes cibles électriques prend énormément de temps, en raison du calcul du champ de diffusion de plusieurs angles et fréquences. Plusieurs méthodes et leurs versions améliorées ainsi que des techniques d'accélération ont été développées pour calculer efficacement la diffusion à partir de grandes cibles électriques, y compris à la fois la méthode numérique basse fréquence et les méthodes d'approximation haute fréquence [7,8]. En raison du temps de calcul et des besoins en mémoire énormes, les méthodes numériques pures telles que la méthode des moments (MoM) et la méthode des éléments finis (FEM) sont confrontées à d'énormes défis [9]. En raison du bon compromis entre précision et efficacité, les méthodes d'approximation haute fréquence sont largement utilisées dans la simulation d'imagerie ISAR pou

grandes cibles électriques. Parmi elles, la méthode de tir et de rebondissement des rayons (SBR) est la plus populaire. Elle est une combinaison d'optique physique (PO) et d'optique géométrique (GO) et convient pour prendre en compte la diffusion multiple.

Suite à la proposition de la méthode SBR par Ling en 1989, les chercheurs ont mis en œuvre de nombreuses améliorations [10], telles que la méthode de tir dans le domaine temporel et de rayons rebondissants (TDSBR), le traçage de rayons analytique bidirectionnel [11], etc. contribué à la popularité croissante de la méthode des rayons rebondissants. Ces dernières années, une méthode de rayons rebondissants accélérée par GPU a été proposée, basée sur un algorithme de traversée d'arbre sans pile à dimension k (Kd) , permettant au processus de traçage de rayons d'être effectué efficacement dans le GPU [12]. Dans [13], une méthode améliorée de rayons rebondissants utilisant une technique de propulsion de rayons est proposée pour accélérer le processus de traçage de rayons et améliorer l'efficacité de l'intersection des rayons, la rendant ainsi capable de calculer efficacement les caractéristiques de diffusion de cibles électriquement grandes . Dans [14], l'inclusion des trajets de rayons inversés dans la méthode SBR est proposée pour améliorer la précision des prévisions de section efficace radar à cavité (RCS), qui peuvent être implémentées de manière presque triviale dans le code SBR existant, tout en produisant des améliorations potentiellement substantielles de la précision des prévisions. . Une technique de traçage de rayons inverse est proposée, basée sur la structure de données Kd-tree de cordes, qui s'est avérée donner des résultats satisfaisants dans le calcul des caractéristiques de diffusion haute fréquence [15]. En plus de l'amélioration du SBR grâce à l'utilisation de GPU et de structures de données, la méthode SBR a également été intégrée à d'autres méthodes de calcul électromagnétique, rendant ainsi cette méthode plus complète dans sa prise en compte des caractéristiques de diffusion électromagnétique de structures cibles complexes [16– 19]. Par exemple, la méthode SBR basée sur l'octree en combinaison avec la théorie physique de la diffraction (PTD) est présentée pour l' analyse de la diffusion électromagnétique (EM) à partir d'une cible en mouvement [20]. Dans [21], une méthode hybride de moment dipolaire équivalent (EDM), MOM et SBR est proposée pour améliorer l' efficacité de calcul du RCS d'objets complexes dans le cadre EDM. Dans cette méthode hybride, une approche itérative est introduite pour améliorer les performances de l'algorithme, offrant une grande précision et réduisant le temps de calcul.

Sur la base de la modélisation de la diffusion électromagnétique, des images ISAR focalisées peuvent être obtenues en appliquant un algorithme de traitement du signal, notamment l' algorithme Range Doppler (RD), l'algorithme de format polaire (PFA), l'algorithme de rétroprojection (BP), etc.]. L'algorithme d'imagerie ISAR le plus couramment utilisé interpole les données polaires sur une grille cartésienne, puis applique une FFT 2D pour réaliser la reconstruction ISAR. Dans un cas particulier, dans des conditions de petit angle et de petite bande passante, les images ISAR peuvent être obtenues approximativement en effectuant une transformée de Fourier inverse de données de champ rétrodiffusées en 2D, et les images ISAR résultantes sont composées des centres de diffusion de la cible avec leur coefficient de réflexion électromagnétique. . En raison de son adéquation au traitement parallèle GPU et de sa capacité à réaliser l'imagerie ISAR dans n'importe quel mode, l'algorithme BP et ses versions modifiées sont largement utilisés dans les applications d'imagerie SAR/ISAR [25]. Dès 1989, un algorithme BP simplifié et son architecture de traitement parallèle ont été proposés en utilisant la forme d'onde radar comme réponse impulsionnelle du filtre pour obtenir la projection filtrée [26]. Jusqu'à présent, l' algorithme BP basé sur GPU est toujours en cours de développement pour optimiser les performances maximales de l' algorithme BP sur les serveurs et les appareils GPU miniaturisés, qui peuvent gérer les différences de plates-formes matérielles ainsi que les différences d'échelles de données [27] . Dans [28], un algorithme d'imagerie ISAR pour les scènes composites cible-océan basé sur les rayons de tir et de rebondissements dans le domaine temporel (TDSBR) est développé. Dans [29], l'optique physique itérative dans le domaine temporel (TD-LIPO) est proposée pour analyser la diffusion à partir de cibles électriquement grandes et complexes. Dans [30], une méthode d'optique physique itérative accélérée dans le domaine temporel est développée pour analyser la diffusion de cibles électriquement grandes et complexes, et une IFFT est réalisée pour obtenir l'image ISAR dans des conditior

Visant l'imagerie ISAR pour de grandes cibles électriques dans des conditions de large bande passante et de grand angle, cet article est consacré à un schéma d'imagerie ISAR combinant le SBR accéléré par GPU basé sur l'accélération d'arbre GPU et BVH avec le BP accéléré par GPU.

algorithme. Pour améliorer l'efficacité de l'intersection des rayons, une structure arborescente BVH est construite

SBR accéléré basé sur l'accélération du GPU et de l'arborescence BVH avec l'algorithme B accéléré par GPU. Pour
améliorer l'efficacité de l'intersection des rayons, une structure arborescente BVH est construite en fonction de la structure
cible, qui est implémentée en C++AMP pour réaliser un calcul d'accélération parallèle GPU. La méthode SAH est incorporée
dans la division bo de délimitation de la scène, atténuant efficacement l'impact du chevauchement du cadre de délimitation
sur l'efficacité du parcours des rayons de la structure arborescente BVH. Pour effectuer efficacement des simulations
d'imagerie ISAR, en fonction de la structure cible, qui est implémentée en C++AMP pour obtenir un parallèle GPU
tion, un algorithme d'imagerie BP accéléré basé sur GPU est développé grâce à un calcul d'accélération informatique. La
méthode SAH est incorporée dans le cadre de sélection de la scène
architecture de périphérique unifiée (CUDA).
division, atténuant efficacement l'impact du chevauchement du cadre de délimitation sur la traversée des rayons
Cet article est organisé comme suit : La section 2 présente un SBR accéléré par GPU utilisant l'efficacité de la
structure arborescente BVH. Pour effectuer efficacement une simulation d'imagerie ISAR, un
Arborescence BVH. Dans la section 3, l'algorithme d'imagerie BP accéléré par GPU est présenté.
Dans la section 4, les résultats et la discussion sont présentés, et plusieurs simulations sont basées sur une architecture
par appareil (CUDA).
L'algorithme d'imagerie BP accéléré basé sur GPU est développé grâce à un calcul unifié
formé pour confirmer la faisabilité et l'efficacité de notre schéma SBR combiné par GPU utilisant
Arborescence BVH. Dans la section 3, l'algorithme d'imagerie BP accéléré par GPU est présenté. Dans
SBR associé à l'algorithme BP pour la simulation rapide d'imagerie ISAR sous grand angle.
Section 4, les résultats et la discussion sont présentés et plusieurs simulations sont effectuées.
conditions de large bande passante. La section 5 conclut cet article.
confirmer la faisabilité et l'efficacité de notre schéma en combinant l'accélération GPU
SBR avec l'algorithme BP pour une simulation d'imagerie ISAR rapide sous grand angle et grand angle. 2. Un
SBR accéléré par GPU utilisant la structure arborescente BVH
conditions de bande passante. La section 5 conclut cet article.

2. Un SBR accéléré par GPU utilisant la structure arborescente BVH

2.1. Calcul de diffusion multiple à l'aide de PO et GO

En tant que méthode d'approximation à haute fréquence, la méthode de tir et de rebond des rayons
est une combinaison de PO et GO, qui utilise GO pour tracer la réflexion des ondes électromagnétiques. En tant que
méthode d'approximation à haute fréquence, la méthode de tir et de rayon rebondissant
tion chemin et PO pour calculer le champ de diffusion, ce qui entraîne un grand avantage dans la résolution est une
combinaison de PO et GO, qui utilise GO pour tracer les problèmes de diffusion re- électromagnétique des
ondes électromagnétiques pour des cibles complexes [31]. Selon le chemin d'approximation PO et PO pour calculer le
champ de diffusion, ce qui entraîne un grand avantage dans
Selon la théorie des partenaires, les surfaces des cibles sont divisées en zones claires et sombres, selon que
l'on résout des problèmes de diffusion électromagnétique pour des cibles complexes [31]. Selon PO
les surfaces sont éclairées ou non par des ondes électromagnétiques. Le champ total dispersé selon la théorie
approximative, les surfaces cibles sont divisées en zones claires et sombres, en fonction
la zone sombre est supposée être nulle. Cependant, cette hypothèse n'est valable que si les surfaces sont éclairées ou
non par des ondes électromagnétiques. Le champ total
La longueur d'onde électromagnétique est beaucoup plus petite que la géométrie cible. La diffusion depuis la
zone sombre est supposée être nulle. Cependant, cette hypothèse n'est
valable lorsque la longueur d'onde de l'onde électromagnétique est beaucoup plus petite que la cible
Il est difficile de déterminer le champ total de la zone sur la cible qui n'est pas une géométrie directe. Il devient difficile de
déterminer le champ total de la zone sur la cible qui
éclairé par l'onde incidente lorsque les dimensions de la cible sont très grandes et n'est pas directement éclairé par l'onde
incidente lorsque les dimensions de la cible sont très grandes.
direction perpendiculaire de l'onde incidente. En supposant un champ total nul sur l'ombre grande dans la direction
perpendiculaire de l'onde incidente. En supposant un champ total de zéro
la surface impliquerait une discontinuité dans le champ sur la limite ombrée. Par conséquent, t sur la surface ombrée
impliquerait une discontinuité dans le champ sur la frontière ombrée.
Pour résoudre la discontinuité dans le champ limite, une intégrale de ligne doit être ajoutée. Par conséquent, pour résoudre
la discontinuité dans le champ limite, une intégrale de ligne doit être ajoutée.
limite [32,33]. Sur la figure 1, un diagramme schématique de la diffusion multiple du SBR est illustré sur
la frontière [32,33]. Sur la figure 1, un diagramme schématique de diffusion multiple de SBR est
traité, dans lequel seule la 1ère réflexion (en flèches jaunes) et la 2ème réflexion (en vert illustrée, dans laquelle seule la
1ère réflexion (en flèches jaunes) et la 2ème réflexion (en vert)
flèches) sont représentées.
flèches) sont représentées.

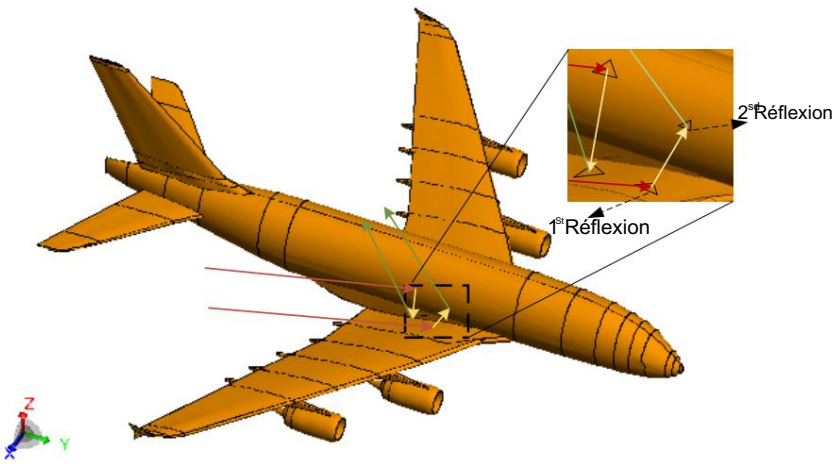


Figure 1. Diagramme schématique de la diffusion multiple.
Figure 1. Diagramme schématique de la diffusion multiple.

Le champ PO de la cible du conducteur électrique parfait (PEC) en position rs peut être analysé.
Le champ PO de la cible du conducteur électrique parfait (PEC) à la position sr peut être exprimé par [34]

enfoncé comme [34]

$$E^{po}_{s}(rs) = -jk\eta \int_{S} k\hat{}_{s} \times k\hat{}_{s} \times J\, G(rs, r')\, dS'$$

$$= -jk\eta \int_{S} \frac{Salut\ -jk(rs-r')}{2\pi rs} k\hat{}_{s} \times k\hat{}_{s} \times n\hat{} \times h\hat{}_{c'est\ à\ dire} -jk(k\hat{}i - k\hat{}s)\cdot r' dS' \quad (1)$$

$$E^{po}(r_s) = -jk\eta \int_S \hat{k}_s \times \left(\hat{k}_s \times \hat{n}\right) G(r_s, r')\, dS' \quad (1)$$

$$H^{je} = -\frac{jk\eta}{2\pi r_s} \left\{ \int_S \hat{k}_s \times \hat{n} \times \left(\hat{n} \times \hat{i}\right) \right\} e^{jk(\hat{k}_s \cdot r')}\, dS'$$

où $\hat{n}$ est le vecteur direction du champ magnétique incident et $\hat{i}$ est électromagnétique

où salut $\hat{n}$ est le vecteur direction du champ magnétique incident et ki $\hat{i}$ est électromagnétique

vecteur de direction de propagation des ondes. ks est le vecteur unitaire de la direction de l'onde de diffusion.

Sorepresente la zone éclairée par l'onde incidente, et r′ est la position du point source . S représente la zone éclairée par

cette. Dans la zone de champ lointain, la fonction de Green dans l'espace libre peut être approchée comme suit : [35]

$$G(r,r') = \frac{e^{-jk|r-r'|}}{4\pi|r-r'|} \approx \frac{e^{-jk(r-\hat{r}\cdot r')}}{4\pi r} \quad (2)$$

Lorsque le rayon de courbure en un point de la cible est beaucoup plus grand que la longueur d'onde

Lorsque le rayon de courbure en un point de la cible est beaucoup plus grand que la longueur d'onde de l'onde incidente, une approximation du plan tangentiel peut être appliquée, et le

longueur de l'onde incidente, une approximation du plan tangentiel peut être appliquée et le courant induit

peut être exprimé comme suit : [36]

le courant induit peut être exprimé par [36]

$$J = \hat{n} \times H = \begin{cases} 2\hat{n} \times H^i & \text{Région éclairée} \\ 0 & \text{Région d'ombre} \end{cases} \quad (3)$$

Sur la figure 2, Dm est l'aire du tube à rayons au niveau du rayon m , et Dm+1 est l'aire du rayon

Sur la figure 2, Dm est l'aire du tube à rayons au niveau du rayon m et Dm+1 est l'aire du tube au niveau du

rayon réfléchi du rayon m . Les rayons provoquent des changements dans l'amplitude et la phase du

le tube à rayons au rayon réfléchi du rayon m . Les rayons provoquent des changements dans l'amplitude et le champ électrique

à mesure qu'ils se propagent et se réfléchissent, ainsi que des informations sur la force et la phase de ces rayons.

phase du champ électrique au fur et à mesure de leur propagation et de leur réflexion, et des informations sur l'intensité du

champ électrique de chaque tube à rayons sont suivies à chaque point de réflexion. Dans GO, le champ électrique

Dans GO, au moment du champ électrique de chaque tube à rayons réfléchi suivi au point de réflexion.

le champ électrique en mr est lié au champ électrique en son r réfléchi selon (rm+1) = Γm . E

$$E^s(r_m) \cdot (DF)_m \cdot e^{-j\beta} \quad (4)$$

équation suivante.

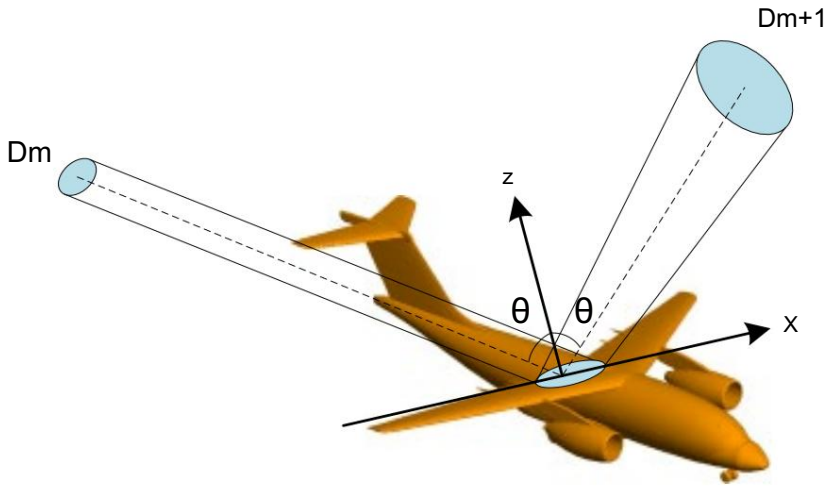


Figure 2. Diagramme schématique de la réflectance du faisceau. La forme du tube à rayons change après chaque re-

Figure 2. Diagramme schématique de la réflectance du faisceau. La forme du tube à rayons change après chaque re-

flexion, qui est déterminée par le facteur de divergence (DF)m .

Dans l'équation (4), Γm désigne le coefficient de réflexion à rm. Pour un chef d'orchestre parfait,

Γm = −1 pour la polarisation horizontale et Γm = 1 pour la polarisation verticale. La divergence

$$E_{ref}(r_m) = \Gamma_m \left(\frac{DF}{DD}\right)_m e^{-j\beta} \quad (4)$$

facteur e− (DF)m à rm est généralement noté (DF)m ≈ Dm/Dm+1.β = k0|rm+1 − rm|, ce qui

représente la différence de phase entre deux points de réflexion voisins. Remplacement

Dans l'équation (4), Γm désigne le coefficient de réflexion à mr . Pour un conducteur parfait, pour une

L'équation (1) dans l'équation (4), le champ diffusé après réflexion du m-ième rayon peut être

= exprimé comme [37–40] pour la polarisation horizontale et 1 Γ =− 1 Γ polarisation verticale. La divergence

$$\left(\frac{DF}{DD}\right)_m \approx \sqrt{\frac{\Delta x_m}{\Delta x_{m+1}}} \cdot \beta = k_0 \left| r_{m+1} - r_m \right|,$$

le facteur ( )m DF à rm est généralement noté (représente lequel

la différence de phase entre deux points de réflexion voisins. La substitution

L'équation (1) dans l'équation (4), le champ diffusé après réflexion du m-ième rayon peut être

$$E^{ref}(r_m) = \Gamma_m (DF)_m e^{-jk|rm+1-rm|} \sum_{x\Delta x} \hat{k}_s \times \hat{k}_s \times \hat{n} \times \hat{h} \; e^{-jk(\hat{k}^i - \hat{k}^s)\cdot r'} dS' \quad (5)$$

exprimé comme [37–40]

Dans l'équation (5), $\sum_{x} E_{\Delta x}^{po}(rm)$ est le champ PO généré par l'élément facette frappé par le m-ième rayon, et x est le nombre d'éléments de facettes frappés. Le champ diffusé résultant pour chaque rayon est ensuite superposé pour obtenir le champ diffusé total.

$$E_{s}^{SBRtotal}(rm) = \sum_{1}^{''} \sum_{x} E_{\Delta x}^{po}(rm) \quad (6)$$

Dans l'équation (6), I est le nombre total de rayons et Δx est l'aire d'un élément de face triangulaire qui a été touché.

2.2. Ray Tracing accéléré par GPU à l'aide de la structure arborescente
BVH 2.2.1. Processus d'accélération GPU

Dans cet article, le SBR accéléré par GPU est parallélisé à l'aide du parallélisme massif accéléré C++ (C++AMP). Par rapport aux processeurs, les GPU possèdent un plus grand nombre de cœurs, ce qui les rend plus adaptés au traitement parallèle massif. C++AMP est une plate-forme informatique parallèle hétérogène basée sur C++ publiée par Microsoft, qui est un modèle de programmation natif avec l'avantage de fonctionner sur tous les appareils sur la plate- forme Windows [41]. La plupart des méthodes de programmation pour GPU, telles que Direct Compute et OpenCL, nécessitent différents langages de programmation et compilateurs. C++AMP unifie le langage de programmation et le compilateur, ce qui le distingue des autres approches. La bibliothèque C++ AMP permet le calcul parallèle via un ensemble d'abstractions et une API de haut niveau, le matériel GPU sous-jacent étant accessible directement via Direct Compute [42,43]. Il est important d'allouer un tableau pour appliquer C++AMP afin d'implémenter le calcul parallèle. Le modèle de tableau se trouve dans l'espace de noms de concurrence. Il prend deux paramètres : un pour le type d'élément de collection et l'autre pour la dimension. La dimension du tableau est définie en fonction du type d'éléments de collection dans cette méthode papier.

Par exemple, lors de la collecte de données de fréquence balayées, en définissant un <tableau a (comptes de fréquences)>, cet exemple définit un tableau unidimensionnel dont la taille est le nombre de fréquences. Les tableaux jouent un rôle extrêmement important dans C++AMP en représentant une vue qui peut accéder aux données sur le GPU et en encapsulant des tableaux ou des vecteurs C++, qui sont des tableaux sur <accelerator_view>. En C++AMP, le GPU n'est pas le seul accélérateur, et chaque accélérateur possède sa propre vue par défaut. Une fois la matrice construite, les données seront transférées vers la mémoire du GPU, où elles seront directement accessibles par le GPU. La fonction <parallel_for_each> est utilisée pour exécuter des tâches de calcul parallèles. La fonction <parallel_for_each> est une fonction d'exécution parallèle en C++AMP qui accepte une plage d'index et une fonction lambda, et elle exécute cette fonction lambda en parallèle sur le GPU pour chaque index. La fonction <parallel_for_each> délègue des tâches de calcul parallèles aux fonctions du noyau du GPU , qui peuvent être directement affectées au matériel du GPU via l'API Direct Compute.
Le mot clé <restrict(amp)> est utilisé pour spécifier que la fonction doit être exécutée uniquement sur le GPU. Il est utilisé pour identifier des blocs spécifiques de code et des fonctions lambda à exécuter sur le GPU, et il permet au compilateur d'optimiser la fonction pour une instruction unique, plusieurs threads (SIMT) [44]. Les opérations au sein de la fonction utiliseront des instructions SIMT pour obtenir l'effet d'un calcul parallèle multithread à instruction unique. C++AMP synchronise les données et les copie de la mémoire GPU vers la mémoire hôte après que la fonction <parallel_for_each> exécute la tâche de calcul parallèle.

La figure 3 présente un processus de calcul parallèle pour C++AMP utilisant l' API Direct Computing pour envoyer des instructions parallèles au périphérique GPU. La structure arborescente BVH ainsi que les données de rayons, etc., sont construites dans le CPU et stockées dans la mémoire globale du GPU. Les données seront automatiquement copiées de l'hôte CPU vers la mémoire GPU en créant le tableau <array_view>, permettant au GPU d'accéder directement à ces données. La mémoire constante du GPU stockera les données qui resteront inchangées lors du calcul parallèle.
La mémoire de texture est utilisée pour stocker les données pendant le rendu du modèle. Il existe de nombreux multiprocesseurs de streaming (SM) dans l'architecture matérielle des GPU, et les SM dans les GPU utilisent l' architecture SIMT. Chaque SM contient un certain nombre de processeurs de streaming (SP), et chaque SP

ing le tableau <array_view> , permettant au GPU d'accéder directement à ces données. La mémoire constante du GPU stockera les données qui resteront inchangées lors du calcul parallèle.

La mémoire de texture est utilisée pour stocker les données pendant le rendu du modèle. Il existe de nombreux multiprocesseurs de streaming (SM) dans l'architecture matérielle des GPU, et les SM dans les GPU utilisent l'architecture SIMT. Chaque SM contient un certain nombre de processeurs de streaming (SP) et chaque SP correspond à un thread. Une seule chaîne est composée de 32 threads, le nombre de threads par chaîne étant déterminé par l'architecture GPU. Warp est la planification et l'exécution correspond à un thread. Une seule chaîne est composée de 32 fils, avec le nombre de unité de tion dans SM, où les threads de chaque chaîne exécutent des calculs en parallèle. Comme précédemment, les threads par chaîne sont déterminés par l'architecture GPU. Warp est la planification et jacente dans SM, ou les threads de chaque chaîne exécutent des calculs en parallèle. Comme Comme indiqué, le fonctionnement parallel_for_each> attribue des tâches directement à l'unité d'exécution GPU sous- matériel via DirectCompute, puis le GPU parallel_for_each> attribue des tâches directement à l'unité d'exécution GPU sous-jacente matériel via DirectCompute. Le GPU attribue des instructions aux warps. SM avec l'architecture SIMT. L'architecture SIMT autorise les threads au sein de chacun du SM avec l'architecture Compute. Le GPU attribue des instructions aux warps. SIMT. L'architecture SIMT permet aux threads à l'intérieur warp pour exécuter une séquence d'instructions, y compris la détection de collision, le lancer de rayons, PO chaque chaîne pour exécuter une séquence d'instructions, y compris la détection de collision, le tracage de rayons, calcul de champ et autres opérations, en modes parallèles. Dans l'architecture SIMD, le calcul du champ PO et d'autres opérations, en modes parallèles. Dans l'architecture SIMD, le Le GPU attribue un fil à chaque rayon. La structure arborescente BVH est obtenue en accédant au thread GPU qui attribue un thread à chaque rayon. L'arborescence BVH est obtenue par accès aux threads la mémoire globale du GPU et est responsable du calcul du chemin de lancer de rayons et de la mémoire globale du GPU et est responsable du calcul du chemin de lancer de rayons et du PO Champ du bon de commande. Dans la fonction <parallel_for_each>, détection d'intersection et champ d'accélération parallèle . Dans la fonction <parallel_for_each>, la détection d'intersection et l'accélération parallèle La ation des rayons et des boîtes englobantes est obtenue en parcourant la structure arborescente BVH. Les rayons et les boîtes englobantes sont obtenus en parcourant la structure arborescente BVH. Après le l'exécution de la fonction <parallel_for_each>, C++AMP copie le disperse parallel_for_each est terminée, C++AMP copie le scat- l' exécution de la fonction données de champ près de la mémoire GPU vers la mémoire hôte. Le processeur additionnera les champs diffusés de champ de la mémoire GPU vers la mémoire hôte. Le GPU additionnera les champs dispersés de tout par les rayons pour obtenir le champ de diffusion cible total. de tous les rayons pour obtenir le champ de diffusion total de la cible.

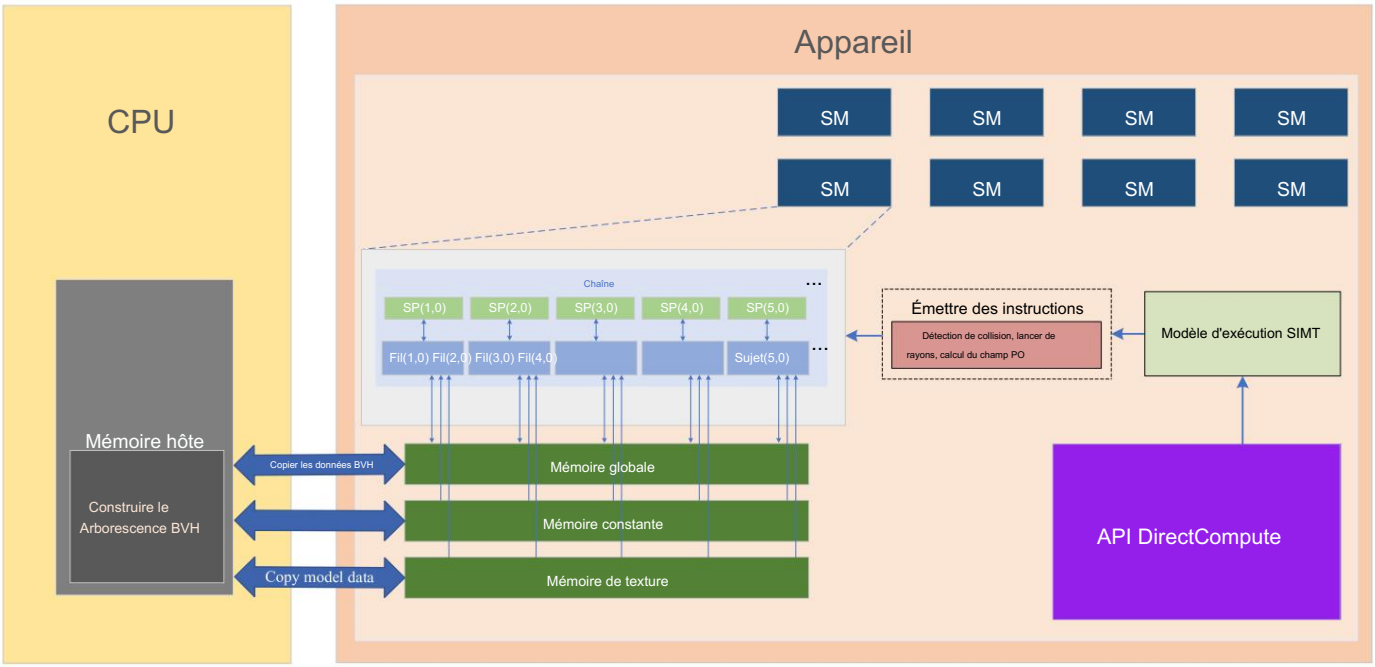


Figure 3. Processus de calcul parallèle de C++AMP utilisant l'API Direct Compute pour envoyer du parallèle instructions à l'appareil (GPU).

Figure 3. Processus de calcul parallèle de C++AMP utilisant l'API Direct Compute pour envoyer du parallèle instructions à l'appareil (GPU).

Les informations sur les rayons sont initialisées et les nœuds ainsi que la structure de l'arborescence L'arborescence BVH est générée dans le CPU comme illustré sur la figure 4. Le CPU partagera les données de l' BVH sont générés dans le CPU comme illustré sur la figure 4. Le CPU partagera les données de Arbre BVH à la mémoire globale du GPU et il partagera ces données avec chaque thread. Dans l'architecture SIMT , arborescence BVH avec la mémoire globale du GPU et partagera ces données avec chaque thread. Architecture SIMT, les instructions sont émises vers le warp par l'API Direct Compute. les instructions sont émises vers la chaîne par l'API Direct Compute. Les fils d'une chaîne exécuteront les instructions reçues séquentiellement et en parallèle, et chaque fil d'une chaîne Les threads d'une chaîne exécuteront les instructions reçues séquentiellement et en parallèle, et sera responsable du calcul d'un rayon. A ce stade, tout chaque fil d'une chaîne sera responsable du calcul d'un rayon. A ce stade, tous les rayons traversant l'arbre BVH coupant la cible sont enregistrés, et leurs rayons réfléchis les rayons traversant l'arbre BVH couvrant la cible sont enregistrés et leurs rayons réfléchis sont tracés jusqu'à ce que le rayon quitte la surface cible. Le GPU calcule le champ dispersé de sont tracés jusqu'à ce que le rayon quitte la surface cible. Le GPU calcule le champ diffusé de ce rayon et transfère les données de tous les rayons pour déterminer le champ de diffusion total de la cible. ce rayon et transfère les données de tous les rayons pour déterminer le champ de diffusion total de la cible. Enfin, le CPU combinera le champ diffusé de tous les rayons pour déterminer le champ de diffusion total de la cible.

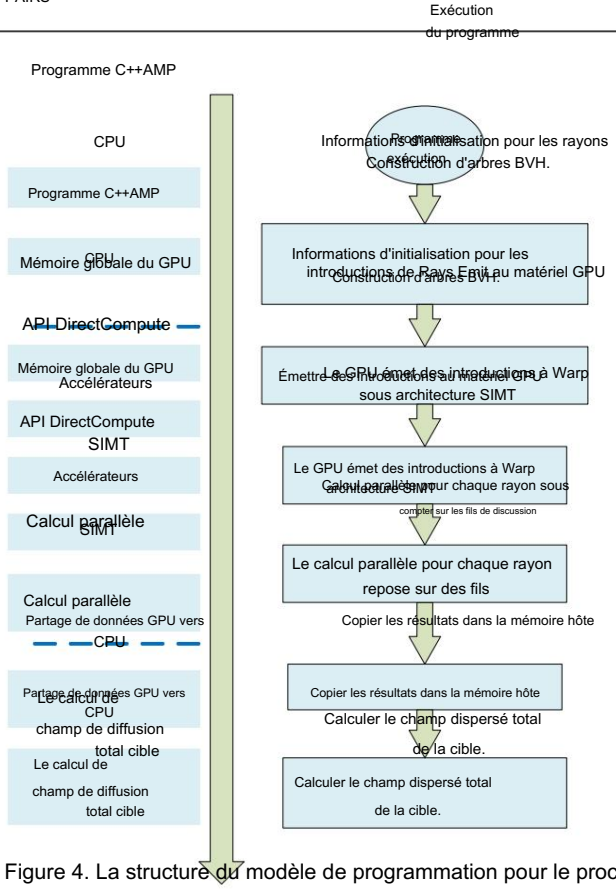


Figure 4. La structure du modèle de programmation pour le processus d'accélération GPU.

Figure 4. La structure du modèle de programmation pour le processus d'accélération GPU.

2.2.2. Algorithme de traçage de rayons utilisant l'arbre BVH

2.2.2. Algorithme de traçage de rayons utilisant l'arbre BVH

La figure 5 illustre le processus de lancer de rayons utilisant la structure arborescente BVH avec le multip. La figure 5 illustre le processus de lancer de rayons utilisant la structure arborescente BVH avec le multiple processus de diffusion d'une onde électromagnétique entre les structures arborescentes. GO est utilisé pour suivre le chemin de diffusion de l'onde électromagnétique entre la surface triangulaire et suivre le chemin de diffusion de l'onde électromagnétique entre la surface triangulaire et utilisé pour l'onde électromagnétique entre les tuples de la surface triangulaire, et PO est utilisé pour calculer le champ diffusé de l'onde électromagnétique lorsqu'elle frappe les tuples de surface triangulaire.

processus de diffusion multiple d'une onde électromagnétique entre les structures arborescentes. GO est utilisé pour suivre le chemin de diffusion de l'onde électromagnétique entre la surface triangulaire et utilisé pour l'onde électromagnétique entre les tuples de la surface triangulaire, et PO est utilisé pour calculer le champ diffusé de l'onde électromagnétique lorsqu'elle frappe les tuples de surface triangulaire.

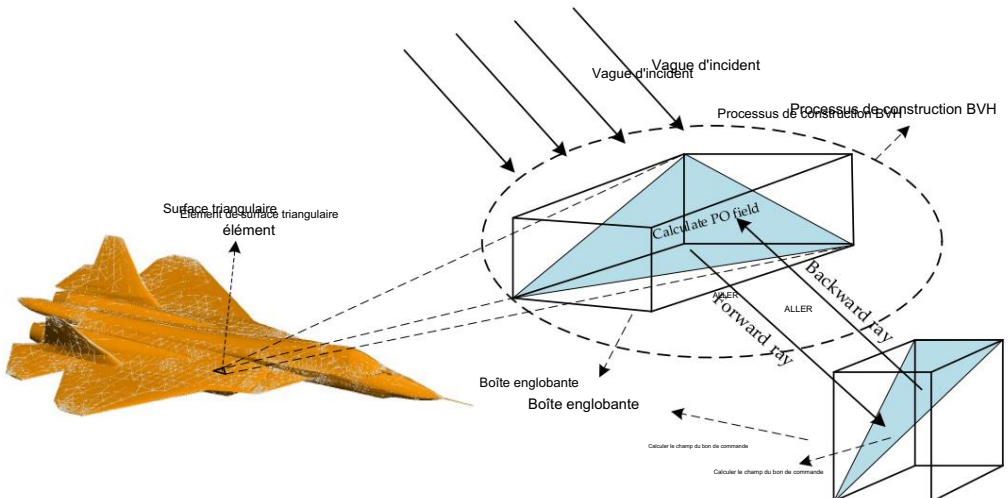


Figure 5. Structure arborescente BVH prenant en compte la diffusion multiple entre les parcelles de surface triangulaires.

Figure 5. Structure arborescente BVH prenant en compte la diffusion multiple entre les parcelles de surface triangulaires.

Afin de résoudre le processus fastidieux d'intersection de rayons en lancer de rayons, nous avons recours.

Afin de résoudre le processus fastidieux d'intersection de rayons en lancer de rayons, nous recourons

à l'arborescence BVH sur la base de l'accélération GPU. Un arbre BVH est une infographie

structure arborescente BVH sur la base de l'accélération GPU. Un arbre BVH est une infographie

informatique qui est une technique de détection d'intersection de rayons basée sur des tuples. Un arbre BVH se divise

BVH se divise en arborescence BVH sur la base de l'accélération GPU. Un arbre BVH est une structure graphique
les tuples en une structure hiérarchique d'ensembles disjoints, largement utilisée dans la structure de traçage
informatique, qui est une technique de détection d'intersection de rayons basée sur des tuples. Un arbre BVH se divise
de rayons, qui est une technique de détection d'intersection de rayons basée sur des tuples. Un arbre BVH divise les
tuples en une structure hiérarchique d'ensembles disjoints, largement utilisée en lancer de rayons

et détection de collision. Nous construirons le cadre de délimitation qui entoure la cible selon les tuples dans une structure hiérarchique d'ensembles disjoints, largement utilisée dans le traçage de rayons et la détection de collisions

Nous allons construire le cadre de délimitation qui entoure la cible en fonction

aux caractéristiques géométriques cibles, et sa structure de boîte de clôture est une boîte de rebond axisymétrique, une boîte de délimitation AABB. Dans la structure arborescente BVH, tous les tuples sont stockés dans la détection de t et de collision. Nous allons construire la boîte englobante qui entoure la cible selon les nœuds feuilles de l'arborescence BVH et les nœuds du milieu stockent les informations sur la boîte. Enfin, les caractéristiques géométriques cibles et sa structure de boîte d'enceinte sont une délimitation axisymétrique les informations de la scène entière sont stockées dans la structure arborescente BVH [45]. Lors de la traversée de la boîte t , une boîte englobante AABB. Dans l'arborescence BVH, tous les tuples sont stockés dans la feuille Arbre BVH, le rayon jugera d'abord s'il croise la boîte ou non. S'il ne s'agit pas d'entre- nœuds de l'arborescence BVH, les nœuds du milieu stockent les informations sur la boîte. Enfin, l'ensemble sectionné, alors il sautera tous les tuples de la boîte, obtenant une amélioration car les informations de la scène ne sont stockées dans la structure arborescente BVH [45]. Lors de la traversée de l'arborescence BVH, le rayon jugera d'abord s'il croise la boîte ou non. S'il n'est pas coupé, alors efficacité du traçage.

les tuples de la boîte seront ignorés, ce qui permettra d'améliorer l'efficacité du traçage des rayons.

Dans la figure 6, la scène contient huit tuples. À l'étape 1, les tuples sont divisés en différents éléments. Tous deux parties dans la scène. Si les rayons lumineux ne croisent pas le cadre de délimitation, alors les parties de la scène. Si les rayons lumineux ne coupent pas la boîte englobante, ils le seront ne croisera pas les tuples, ce qui peut exclure la moitié des tuples à la fois et ne croisera pas les tuples, ce qui peut exclure la moitié des tuples à la fois et donc réduire le nombre de carrefours. À l'étape , nous continuons à diviser les tuples en deux pour réduire le nombre d'intersections. À l'étape 2, nous continuons à diviser les tuples en deux, et et ce faisant, la complexité d'intersection des rayons peut être réduite de O n( ) t. Ce faisant, la complexité d'intersection des rayons peut être réduite de O(n) à O(log(n)) [46].

Après une division récursive par paire, la boîte englobante qui entoure un seul tuple peut être trouvé. Si un rayon coupe cette boîte englobante, le rayon continue de croiser le tuple. L'équation du paramètre de rayon peut être exprimée comme suit

un seul tuple peut être trouvé. Si un rayon coupe cette boîte englobante, le rayon continue à former un tuple. croiser le tuple. L'équation du paramètre de rayon peut être exprimée comme suit

$$r(t) = o + t \cdot d \quad (7)$$

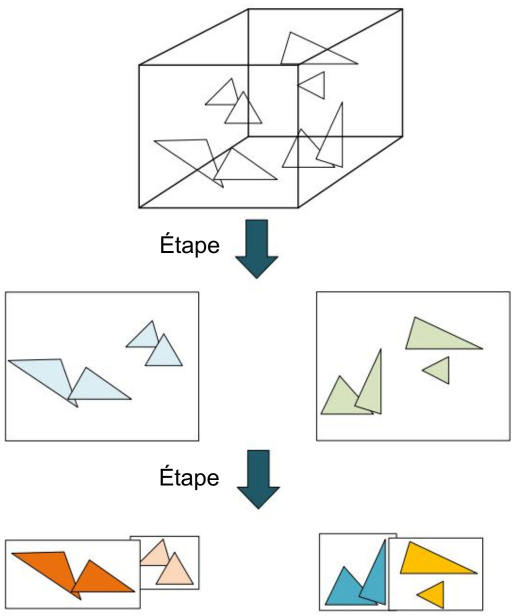


Figure 6. Processus de segmentation des tuples de scénarios.

Dans l'équation (7), o et d sont le point de départ du rayon et le vecteur de direction du rayon normalisé. En remplaçant l'équation des paramètres du rayon dans l'équation plane implicite de où se trouve le tuple, l'équation du plan implicite peut s'écrire sous la forme

Dans l'équation (7), o et d sont respectivement le point de départ du rayon et la direction normalisée du rayon . vecteur, respectivement. En substituant l'équation du paramètre de rayon dans l'équation du plan implicite - le plan tion du plan où se trouve le tuple, l'équation implicite du plan peut s'écrire sous la forme

$$N^T(o + t \cdot d) = c \quad (8)$$

D'après l'équation (8), le paramètre t correspondant au point d'intersection du rayon comme

À partir de l'équation (8), le paramètre t correspondant au point d'intersection de t avec le plan peut être résolu le rayon avec l'avion peut être résolu comme

$$t = \frac{c - N^T \cdot o}{N^T \cdot d} \quad (9)$$

Dans cet article, l'élément de surface est un élément de surface triangulaire et le plan l'équation paramétrique de l'élément de surface triangulaire est

Dans cet article, l'élément de surface est un élément de surface triangulaire, et le plan l'équation métrique de l'élément de surface triangulaire est

$$f(u, v) = (1 - u - v) \cdot p_0 + u \cdot p_1 + v \cdot p_2 \quad \text{(dix)}$$

Électronique 2024, 13, x POUR EXAMEN PAR LES PAIRS 9 sur 24

où u et v sont les coordonnées du centre de masse du triangle, satisfaisant u ≥ 0 et u+v ≤ 1 . L'élément de face triangulaire peut être considéré comme la cartographie de l'unité où u et v sont les coordonnées du centre de masse du triangle, satisfaisant u ≥ 0 élément de face de triangle sur ses trois arêtes. et u+v ≤ 1 . L'élément de face triangulaire peut être considéré comme la cartographie de l'unité où u et v sont les coordonnées du centre de masse du triangle, satisfaisant u ≥ 0 et u+v ≤ 1. L'élément de face triangulaire peut être considéré comme la cartographie du triangle unitaire élément de face sur ses trois arêtes

En combinant l'équation (11) avec l'équation (7), on peut obtenir

$$\begin{bmatrix} u \\ v \\ t \end{bmatrix} = M^{-1} \cdot \begin{bmatrix} p_{0x} - o_x \\ p_{0y} - o_y \\ p_{0z} - o_z \end{bmatrix} \quad (11)$$

En combinant l'équation (11) avec l'équation (7), on peut obtenir

$$\begin{bmatrix} u \\ v \\ t \end{bmatrix} = M^{-1} \cdot \begin{bmatrix} p_{0x} - o_x \\ p_{0y} - o_y \\ p_{0z} - o_z \end{bmatrix} \quad (12)$$

Dans l'équation (12), M −1 est la matrice transformant un élément de face triangulaire en un élément de face triangulaire unitaire dans le plan u,v , dans lequel le rayon cartographié est orthogonal à l'élément de face triangulaire unitaire. La cartographie des éléments de surface triangulaires dans le u,v plan, est illustré à la figure 7. La figure 7a est le cas avant la cartographie, et la figure 7b est le cas correspondant à l'équation (11). La figure 7c représente l'intersection du rayon avec le élément de surface triangulaire unitaire après mappage.

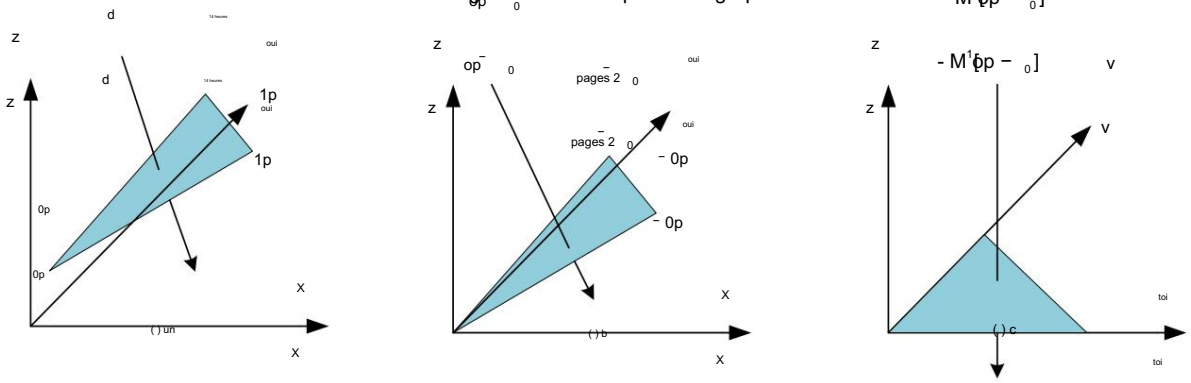


Figure 7. Cartographie d'éléments de surface triangulaires dans le plan u,v . (a) Le cas avant la cartographie ; (b) Le cas correspondant à l'équation (11) ; (c) L'intersection du rayon avec l'unité triangulaire Figure 7. Cartographie des éléments de surface triangulaires dans le plan u, v . Figure 7. Cartographie des éléments de surface triangulaires dans le plan u,v . (a) Le cas avant la cartographie ; (b) Le cas correspondant à l'équation (11) ; (c) L'intersection du rayon avec l'élément triangulaire unitaire après cartographie. mappage. (a) Le cas avant la cartographie ; (b) Le cas correspondant à l'équation (11) ; (c) L'intersection du rayon avec l'élément triangulaire unitaire après cartographie. élément de surface après mappage.

Selon la méthode de demi-division, les cadres de délimitation de la scène peuvent se chevaucher ou se croiser, et le chevauchement des cadres de délimitation est illustré à la figure 8. Sur la figure 8, lorsque la lumière traverse l'arbre BVH, le chevauchement des boîtes englobantes amène la lumière à pour croiser les deux boîtes englobantes, ce qui entraîne une diminution de l'efficacité de l'intersection.

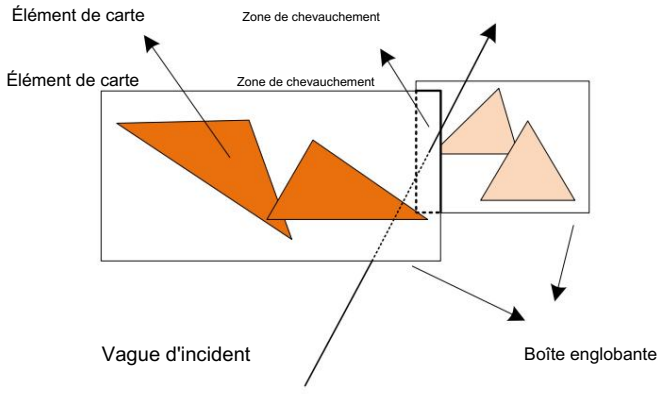


Figure 8. Diagramme de chevauchement des boîtes englobantes des nœuds enfants. Figure 8. Diagramme de chevauchement des cadres de délimitation des nœuds enfants.

Figure 8. Diagramme de chevauchement des cadres de délimitation des nœuds enfants.

Afin d'éliminer le phénomène de chevauchement des boîtes d'enveloppe, la méthode de division SAH au lieu de la méthode de demi-division est adoptée dans le processus de construction de l'arbre BVH. Afin d'éliminer le phénomène de chevauchement des boîtes d'enveloppe, la structure de division SAH. Le SAH est basé sur la méthode heuristique de division de la surface, et la méthode ultérieure au lieu de la méthode de demi-division est adoptée dans le processus de construction de l'arbre BVH. Le SAH est basé sur la méthode heuristique de division de la surface, et après

Afin d'éliminer le phénomène de chevauchement des boîtes enveloppes, la méthode de division SAH au lieu de la méthode de demi-division est adoptée dans le processus de construction de l'arborescence BVH . Le SAH est basé sur la méthode de division heuristique de la surface, et après avoir ajouté le SAH, nous pouvons estimer la probabilité que le rayon lumineux frappe les boîtes enveloppantes en termes de taille de la surface de la boîte enveloppante parent dans laquelle il y a deux ou davantage de boîtes d'enveloppement pour enfants qui se chevauchent.

Dans l'arborescence BVH, sous l'hypothèse que le nœud actuel a trois boîtes englobantes A, B et C, le coût de l'intersection du rayon avec le nœud actuel est

$$c(UNE, B, C)= p(UNE)\sum_{je\ \ A} t(i)+p(B)\sum_{je\ \ B} t(j)+p(C)\sum_{je\ \ C} t(k) + ttrav \quad (13)$$

Dans l'équation (13), ∑ t(i) est le coût d'intersection de chaque boîte de sous-boîtier, et t(i) est le i-ème tuple de la boîte de clôture enfant. p(A), p(B) et p(C) sont les probabilités que la lumière frappe les objets dans les cadres englobants A, B et C, respectivement. ttrav est le coût de la lumière traversant l'arbre BVH.

Dans SAH, nous utilisons la surface du cadre de délimitation enfant dans le nœud parent au lieu de la probabilité que le rayon frappe le cadre de délimitation. En supposant que les surfaces des cadres de délimitation enfants A, B et C sont respectivement S(A), S(B) et S(C), et que la surface du cadre de délimitation du nœud parent D est S(D) , L'équation (13) peut être réécrite comme

$$S(UNE)\ c(UNE, B, C) = \frac{}{DAKOTA\ DU\ SUD)}\sum_{je\ \ A} t\ (i) + \frac{S(B)}{DAKOTA\ DU\ SUD)}\sum_{je\ \ B} t(j) + \frac{S(C)}{DAKOTA\ DU\ SUD)}\sum_{je\ \ C} t(k) + ttrav \quad (14)$$

Après avoir résolu la méthode de division optimale en calculant la valeur minimale de l'équation (14), le parcours des rayons de l'arbre BVH est le plus efficace [47].

3. Algorithme d'imagerie BP accéléré par GPU 3.1.

Algorithme BP pour l'imagerie ISAR

Dans des conditions de petit angle et de petite bande passante, les images ISAR peuvent être obtenues approximativement en effectuant la transformée de Fourier inverse des données de champ rétrodiffusées en 2D, et les images ISAR résultantes sont composées des centres de diffusion de la cible avec leur coefficient de réflexion électromagnétique. En raison de son adéquation au traitement parallèle GPU et de sa capacité à réaliser l'imagerie ISAR dans n'importe quel mode, l'algorithme BP et ses versions modifiées sont largement utilisés dans les applications d'imagerie SAR/ISAR. L'algorithme d'imagerie BP est une méthode avec une précision d'imagerie élevée. Cependant, en raison de sa grande complexité, l'algorithme BP n'est pas aussi performant que les autres algorithmes d'imagerie en termes de vitesse d'imagerie. Visant l'imagerie ISAR pour de grandes cibles électriques dans des conditions de large bande passante et de grand angle, un algorithme d'imagerie BP accéléré basé sur GPU est développé dans cet article grâce à CUDA, tout en conservant la précision d'imagerie de l'algorithme BP.

L'idée fondamentale de l'algorithme BP consiste à superposer de manière cohérente les échos calculés de chaque impulsion en transmettant des impulsions électromagnétiques et en calculant le délai bidirectionnel entre les points de pixel dans la zone d'imagerie et le radar au moment de chaque impulsion. La superposition dépend de la relation de phase entre les points de pixels. Si les échos sont en phase, les échos des points pixels superposés deviennent de plus en plus forts. Lorsque des points de pixels avec des phases différentes sont superposés, l'effet est plus faible. En tant qu'algorithme précis dans le domaine temporel, le profil de distance dans l'algorithme BP est obtenu à l'aide de la technique de compression d'impulsions, similaire à l'algorithme Range-Doppler.

Le traitement de la direction azimutale est réalisé en calculant les échos des points pixels pour une superposition cohérente, liée à l'angle de rotation de la cible par rapport au radar. On observe que la résolution azimutale augmente à mesure que l'angle entre la cible et le radar augmente, sans limite apparente. Pour toute trajectoire de mouvement de la cible, si la trajectoire de mouvement peut être prédite à l'avance, l'algorithme BP peut alors obtenir une imagerie précise.

est la grille de l'image. Ra et Rb désignent la distance radar de la cible à la figure
instants a et b. xyz est un système de coordonnées global. La cible spatiale XY est le système de coordonnées local. S est le
(X a , Y a) et (X b , Y b) sont les positions de la cible dans la grille d'imagerie aux instants a et a et b,
b , respectivement. v est la vitesse et la direction de la cible en mouvement. Le signe de transmission correspond aux
nal est [48]
positions de la cible dans la grille d'imagerie aux instants a et b, respectivement. v est le
vitesse et direction de la cible en mouvement. Le signal de transmission est [48]

$$s(t, m) = \text{rectangle}\left(\frac{t}{T_p}\right) \cdot \exp\left(j2\pi f_0 t + jK t^2\right) \quad (15)$$

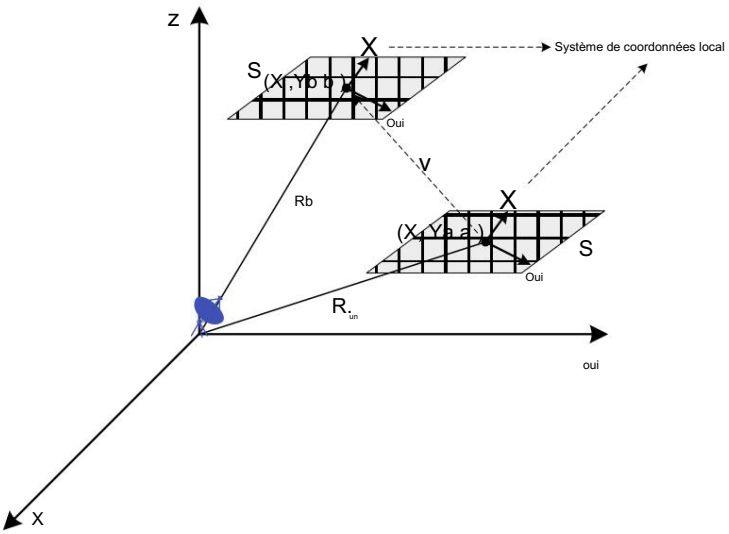


Figure 9. Diagramme schématique de rétroprojection pour l'imagerie ISAR.
Figure 9. Diagramme schématique de rétroprojection pour l'imagerie ISAR.

Dans l'équation (15), t est le temps rapide, T_p est la largeur d'impulsion du signal, rectangle est une
Dans l'équation (15), t est le temps rapide, T_p est la largeur d'impulsion du signal, le temps rapide est un
fonction de fenêtre rectangulaire. f0 est la fréquence porteuse du signal. $K = \frac{B}{T_p}$, K est le signal
Tp , est une fonction de fenêtre rectangulaire. f est la fréquence porteuse du signal. $K = \frac{B}{T_p}$, K est la
fréquence de modulation. B est la bande passante du signal. L'écho reçu est le suivant
fréquence de modulation du signal. B est la bande passante du signal. L'écho reçu est le suivant

$$sr(t, m) = \sum_{j=1}^{\text{rectifier}} \text{rectangle}\left(\frac{t - \tau_{i,j(m)}}{T_p}\right) \exp\left(j2\pi f_0 (t - \tau_{i,j(m)}) + jK(t - \tau_{i,j(m)})^2\right) \quad (16)$$

Dans l'équation (16), $\tau_{i,j(m)} = \frac{2\sqrt{(x(m)+X_i)^2+(y(m)+Y_i)^2+z(m)^2}}{c}$ est le retard bidirectionnel de
pixel (Xi ,Yi) dans le plan imageur au radar au temps lent m. x(m), y(m) et z(m)
Dans l'équation (16), $\tau_{i,j(m)} = \frac{2\sqrt{(x(m)+X_i)^2+(y(m)+Y_i)^2+z(m)^2}}{c}$ est le dé-
figurent les positions des pixels (Xi , Yi) dans la grille d'imagerie dans le système de coordonnées spatiales cibles.
Le filtre correspondant est le suivant
s'étend du pixel (X Yi i ,) dans le plan d'imagerie jusqu'au radar au temps lent m . x m( ) , y m( ) et z

$$h(t, m) = \exp\left(j2\pi f_0 t - jK t^2\right) \quad (17)$$

m( ) sont les positions du pixel (X Yi i ,) dans la grille d'imagerie dans l'espace
système de coordonnées cible. Le filtre correspondant est le suivant

Dans l'équation (17), t0 est le délai bidirectionnel à la distance la plus proche entre la cible et
Conversion de la convolution du domaine temporel en multiplication du domaine fréquentiel
traitement, la sortie du filtre adapté peut être exprimée comme suit

Dans l'équation (17), t_0 est le retard bidirectionnel à la distance la plus proche entre la cible et le
radar. En convertissant la convolution dans le domaine temporel en multiplication dans le domaine

$$s_{ac}(t_m) = \text{IFFT}(\text{FFT}(s(t_m)) \cdot \text{FFT}(h(t_m))) \quad (18)$$

fréquentiel, la sortie du filtre adapté peut être
exprimée comme suit :

La haute résolution en azimut est obtenue par l'accumulation cohérente d'impulsions.
La formule intégrale pour l'accumulation cohérente est la suivante [49]

La haute résolution en azimut est obtenue par l'accumulation cohérente d'impulsions. La
formule intégrale pour l'accumulation cohérente est la suivante [49]

$$I(x, y) = \int_m s(t, m) \cdot \exp\left(j2\pi f_0 \tau_{i,j(m)}\right) dm \quad (19)$$

L'énergie de chaque point de pixel dans la grille est superposée de manière cohérente sur le temps de mouvement cible. La valeur de pixel est accumulée par chaque point de pixel pendant le temps de mouvement pour synthétiser l'image finale.

3.2. Accélération GPU de l'algorithme BP pour l'imagerie

ISAR Lancé par NVIDIA en 2006, CUDA est une plate-forme de calcul parallèle à usage général et un modèle de programmation construit sur des GPU. Les calculs pour des tâches complexes peuvent être effectués plus efficacement avec la programmation CUDA. Ces dernières années, les techniques de programmation CUDA ont été développées tant au niveau matériel que logiciel [50]. Au moment de sa sortie initiale, CUDA était capable d'utiliser des GPU avec un nombre limité de cœurs, généralement de l'ordre de quelques dizaines ou quelques centaines. Par conséquent, il n'a pas été possible de faire des comparaisons significatives en termes de puissance de calcul entre ces premiers GPU et les GPU disponibles aujourd'hui [51]. Par exemple, le NVIDIA RTX 2080 de NVIDIA, sorti le 20 septembre 2018, est un GPU doté de 2 944 cœurs CUDA et d'une puissance de calcul FP32 de 10 070 milliards de fois par seconde. Cependant, le NVIDIA RTX 4090 dispose désormais de 16 384 cœurs CUDA, avec une puissance de calcul FP32 atteignant 82 580 milliards de fois par seconde. Ces dernières années, CUDA a largement utilisé ses puissantes capacités de calcul parallèle dans le domaine du calcul scientifique [52].

Dans cet article, nous avons réalisé l'algorithme d'imagerie BP hautement parallèle de CUDA. Les données de champ dispersé obtenues par le calcul SBR sont d'abord chargées dans la mémoire globale du GPU à partir de la mémoire hôte sous la fonction <cudaMemcpy>. Ces données contiennent des informations telles que l'azimut, l'angle d'incidence, la fréquence, les points d'échantillonnage de fréquence, les points d'échantillonnage d'angle, le mode de polarisation, etc. L'espace mémoire spécifié est alloué pour ces paramètres à partir du GPU avec une taille de mémoire Nf × Nphi × Nf ft par le Fonction <cudaMalloc> . Les variables globales sur l'appareil sont définies via le symbole _device_, y compris le signal de compression de distance et les variables utilisées pour stocker les données brutes du champ diffusé, les variables utilisées pour stocker les coordonnées et la position du radar, et les variables utilisées pour stocker les résultats d'imagerie finaux. . La compression d'impulsions de plage des données d'écho brutes est effectuée à l'aide de la bibliothèque de fonctions <cufft> dans CUDA, grâce à laquelle des FFT et des IFFT hautement parallèles peuvent être réalisées. Le signal de compression de plage est copié dans la mémoire GPU à l'aide de la fonction <cudaMemcpy> avec une taille de mémoire allouée Nf × Nphi × Nf f t.

Le calcul accéléré parallèle est principalement implémenté sur l'appareil par la fonction du noyau CUDA. Le noyau est un concept important dans CUDA, et c'est une fonction qui est exécutée en parallèle dans un thread sur l'appareil. La fonction noyau est déclarée avec le symbole <_global_>, et le nombre de threads requis lors de l'appel de cette fonction doit être spécifié, notamment par <<<grid, block>>>. La fonction noyau est exécutée par chaque thread. Le processus de focalisation azimutale de l'algorithme BP est écrit sous forme de fonction noyau, et le nombre correspondant de threads est alloué à la fonction noyau pour permettre un calcul parallèle sur le GPU.

Sur la figure 10, la fonction noyau est chargée de calculer les distances de tous les pixels du radar à chaque moment azimutal, ainsi que la compensation de phase. La position du signal azimutal correspondant est déterminée par la distance du point pixel au radar. L'imagerie ISAR utilisant l'algorithme BP est obtenue en superposant de manière cohérente les échos de tous les points de pixels à chaque moment azimutal. Les blocs et les threads sont définis comme unidimensionnels lors de l'exécution de la fonction CUDA de l'algorithme BP. La répartition des threads dans chaque bloc est Ny, 1, 1 avec Nx blocs au total et (Nx, 1, 1) répartition des blocs. Nx et Ny sont respectivement le nombre de pixels dans la direction x et le nombre de pixels dans la direction y . Afin d'optimiser l'efficacité de la fonction noyau dans le traitement des données à grande échelle, les tâches de calcul pour l'imagerie de pixels de taille supérieure à 500 × 500 sont calculées par lots. Le nombre maximum de points de pixels calculés dans chaque lot est de 500 × 500 afin de s'adapter aux limitations de ressources du GPU. Dans la fonction noyau, le GPU alloue 250 000 threads pour chaque lot de tâches de calcul, tous les 500 threads constituant un bloc, pour un total de 500 blocs. Chaque thread est responsable du calcul de la valeur d'un point de pixel, et plusieurs lots de données seront exécutés en parallèle sur le co

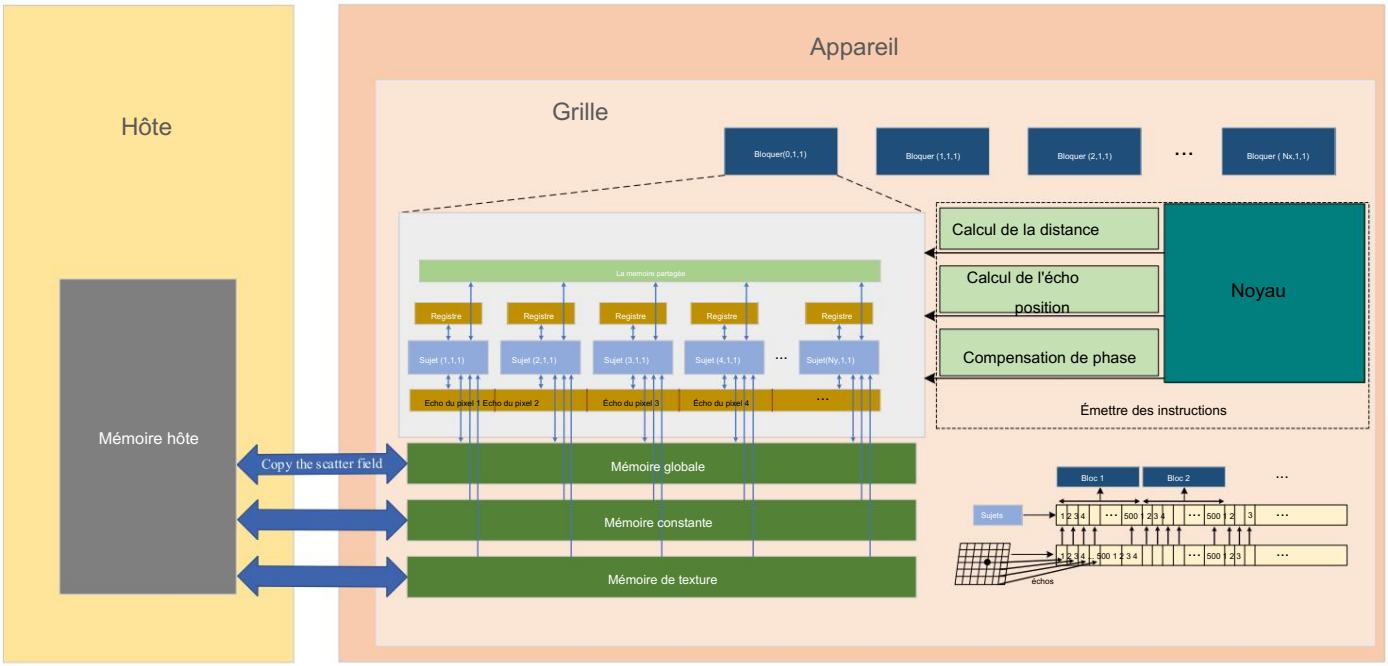


Figure 10. Processus de calcul parallèle de l'algorithme BP pour l'imagerie ISAR utilisant l'accélération CUDA.

Dans la figure 10, les données de champ dispersées sont d'abord copiées de la mémoire hôte vers le GPU. Dans le GPU, les données de champ dispersées sont stockées dans la mémoire globale. La mémoire constante est utilisée pour stocker les données de texture. Le nombre de blocs et de threads est déterminé par le nombre de pixels Nx dans la direction x et le nombre de pixels Ny dans la direction y, c'est-à-dire la fonction du noyau s'appuie sur des threads pour réaliser le fonctionnement hautement parallèle des calculs de localisation azimutale.

La figure 11 présente l'organigramme de l'algorithme BP accéléré par GPU. Après la plage, un certain nombre de threads sont alloués pour être responsables du calcul des données d'écho de ces pixels. Si les grilles et les accumulant aux positions correspondantes. La combinaison de ceux-ci correspondante est divisée en plusieurs sous-régions pour un traitement parallèle si les pixels sont dépassés 500 × 500. Chaque thread du GPU est responsable du traitement d'un point d'un pixel, y compris l'écho vers la grille de pixels correspondante. Une fois que chaque bloc a terminé le calcul de ses données du sous-région, les données sont transférées à leur position respective dans l'image finale. Enfin, l'image entière est sortie.

Figure 10. Processus de calcul parallèle de l'algorithme BP pour l'imagerie ISAR utilisant l'accélération CUDA.

Dans la figure 10, les données de champ dispersées sont d'abord copiées de la mémoire hôte vers le GPU. Dans le GPU, les données de champ dispersées sont stockées dans la mémoire globale. La mémoire constante est utilisée pour stocker les données de texture. Le nombre de blocs et de threads est déterminé par le nombre de pixels Nx dans la direction x et le nombre de pixels Ny dans la direction y, c'est-à-dire la fonction du noyau s'appuie sur des threads pour réaliser le fonctionnement hautement parallèle des calculs de localisation azimutale.

La figure 11 présente l'organigramme de l'algorithme BP accéléré par GPU. Après la plage, un certain nombre de threads sont alloués pour être responsables du calcul des données d'écho de ces pixels. Si les grilles et les accumulant aux positions correspondantes. La combinaison de ceux-ci correspondante est divisée en plusieurs sous-régions pour un traitement parallèle si les pixels sont dépassés 500 × 500. Chaque thread du GPU est responsable du traitement d'un point d'un pixel, y compris l'écho vers la grille de pixels correspondante. Une fois que chaque bloc a terminé le calcul de ses données du sous-région, les données sont transférées à leur position respective dans l'image finale. Enfin, l'image entière est sortie.

La mémoire est une mémoire morte utilisée pour diffuser les instructions envoyées par l'hôte. Texturer le nombre de pixels Nx dans la direction x et le nombre de pixels Ny dans la direction y, respectivement. Dans un bloc, tous les threads exécutent les mêmes instructions, et chaque thread exécute la fonction du noyau s'appuie sur des threads pour réaliser le fonctionnement hautement parallèle les calculs de localisation azimutale.

La figure 11 présente l'organigramme de l'algorithme BP accéléré par GPU. Après la plage, un certain nombre de threads sont alloués pour être responsables du calcul des données d'écho de ces pixels. Si les grilles et les accumulant aux positions correspondantes. La combinaison de ceux-ci correspondante est divisée en plusieurs sous-régions pour un traitement parallèle si les pixels sont dépassés 500 × 500. Chaque thread du GPU est responsable du traitement d'un point d'un pixel, y compris l'écho vers la grille de pixels correspondante. Une fois que chaque bloc a terminé le calcul de ses données du sous-région, les données sont transférées à leur position respective dans l'image finale. Enfin, l'image entière est sortie.

Électronique 2024, 13, 3082 14 sur 24

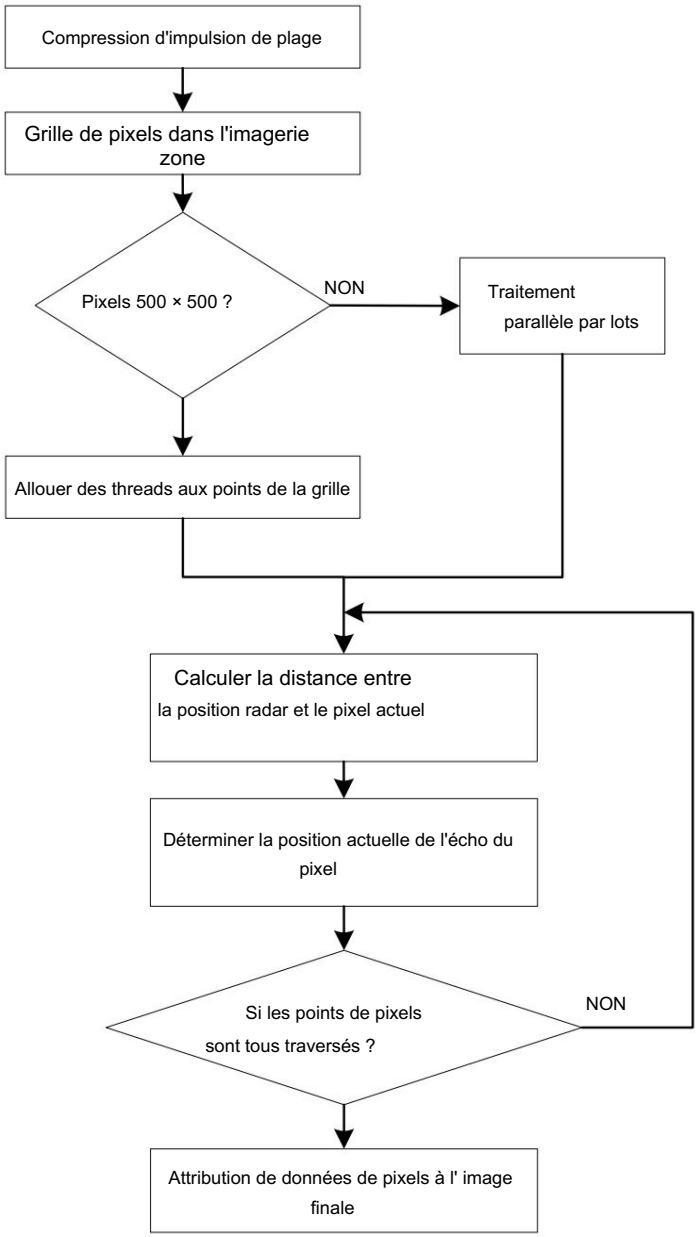


Figure 11. Organigramme de l'algorithme BP accéléré par GPU.
Figure 11. Organigramme de l'algorithme BP accéléré par GPU.

4. Résultats et discussion
4. Résultats et discussion

4.1. Validation de la méthode SBR accélérée par GPU
4.1. Validation de la méthode SBR accélérée par GPU

La mise en œuvre de la méthode de calcul électromagnétique accélérée par GPU et de la méthode d'imagerie BP accélérée par GPU a été présentée précédemment, comme décrit dans cet article. Dans cette section, la validité du SBR accéléré par GPU combinant les approximations GO avec PO est vérifiée par comparaison avec RL-GO dans le logiciel FEKO v2020. En prenant comme exemple un chasseur F-22 à grande échelle, RL-GO est configuré avec deux types de densités de rayons, l'une est de λ/10 et l'autre de λ/100. Une comparaison du RCS en champ lointain est faite avec l'élec- /10 et l'autre est /100 fréquence d'onde tromagnétique 3 GHz . La dissection du modèle a produit 2444 f triangulaires = maillages, et l'arbre BVH a été construit pour produire 4887 nœuds feuilles. Le modèle CAO est comme dans la figure 12. Le temps de calcul et le coût de la mémoire sont répertoriés dans le tableau 1. Le processeur de notre ordinateur est un Intel(R) Core (TM) i5-12490F de 12e génération avec une vitesse de référence de 3,00 GHz, fabriqué par Intel Corporation pour les produits en édition spéciale en Chine. Le GPU est un NVIDIA GeForce RTX 2080 avec 8,0 Go de mémoire GPU dédiée, fabriquée par NVIDIA Corporation en Chine continentale.

La mise en œuvre de la méthode de calcul électromagnétique accélérée par GPU et de la méthode d'imagerie BP accélérée par GPU a été présentée précédemment, comme décrit dans cet article. Dans cette section, la validité du SBR accéléré par GPU combinant les approximations GO avec PO est vérifiée par comparaison avec RL-GO dans le logiciel FEKO v2020. Prise À titre d'exemple de chasseur F-22 à grande échelle, le RL-GO est configuré avec deux types de densités de rayons, l' une est de λ/10 et l'autre de λ/100. Une comparaison du RCS en champ lointain est faite avec l'élec- /10 et l'autre est /100 . Une comparaison du RCS en champ lointain est effectuée avec la fréquence des ondes électromagnétiques f0 = 3 GHz. La dissection du modèle a produit 2444 triangles maillages, et l'arbre BVH a été construit pour produire 4887 nœuds feuilles. Le modèle CAO est Le temps de calcul et le coût de la mémoire sont répertoriés dans le tableau 1. Les performances de notre ordinateur Le processeur est un Intel(R) Core (TM) i5-12490F de 12e génération avec une vitesse de référence de 3,00 GHz, fabriqué par Intel Corporation pour les produits en édition spéciale en Chine. Le GPU est un NVIDIA GeForce RTX 2080 avec 8,0 Go de mémoire GPU dédiée, fabriquée par NVIDIA Corporation en Chine continentale.

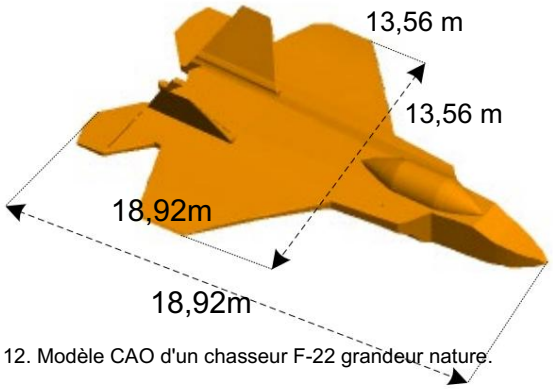


Figure 12. Modèle CAO d'un chasseur F-22 grandeur nature.

Figure 12. Modèle CAO d'un chasseur F-22 grandeur nature.

La figure 13 montre la comparaison de la distribution angulaire RCS d'un F-22 à grande échelle.
chasseur obtenu par notre méthode SBR accélérée par GPU et la méthode parallèle RL-GO dans le tableau 1. Une
comparaison du temps de calcul et du coût de la mémoire pour un chasseur F-22 à grande échelle
La figure 13 montre la comparaison de la distribution angulaire RCS d'un F-22 à grande échelle.

Polariser-tion	Temps de calcul (s)			Mémoire (Mo)		
	Cet article	RL-GO	RL-GO	Cet article	RL-GO	RL-GO
		λ /100	λ /dix		λ /100	λ /dix
Papier	207.96	186,4	2303.3	557.6	130.9	181.4
Papier	433.195	103,1	1427.43	552,1	136,4	185.1

Méthode GO avec une densité de rayons de λ /100 dans FEKO en raison de l'effet de la densité des rayons, et /
la méthode RL-GO avec une densité de rayons de λ 100 prend également en compte le champ de diffraction [53]
Méthode GO avec une densité de rayons de la λ /100 dans FEKO en raison de l'effet de la densité des rayons, et /
méthode RL-GO avec une densité de rayons de compte. λ 100 prend également en compte le champ de diffraction [53

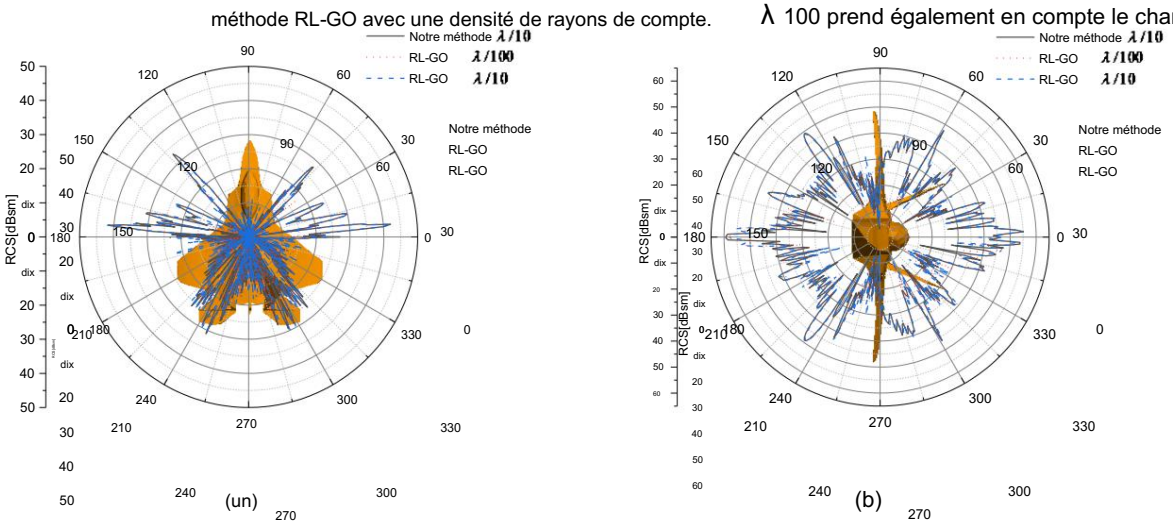


Figure 13. Distribution angulaire RCS d'un chasseur F-22 à grande échelle : (a) angle d'incidence θ = 90°, azimut
angle φ=0 °, polarisation VV ; (b) angle d'incidence θ = 90°, angle d'azimut φ=0 °,
polarisation 0 ~ 360 HH.
Figure 13. Distribution angulaire RCS d'un chasseur F-22 à grande échelle : (a) angle d'incidence θ = 90°, azimut
= 0 ~ 360 polarisation VV ; (b) angle d'incidence θ = 90°, angle d'azimut φ= 0 °,
Polarisation HH.

Tableau 1. Comparaison du temps de calcul et du coût de la mémoire pour un chasseur F-22 à grande échelle.
comparant obtenu par notre méthode SBR accélérée par GPU et la méthode parallèle RL-GO dans
FEKO. À partir des figures 13a, b, on observe que le RCS calculé par notre processeur accéléré par GPU
Tableau 1. Comparaison du temps de calcul et du coût de la mémoire pour un chasseur F-22 à grande échelle.

Polariser-tion	Temps de calcul (s)			Mémoire (Mo)		
	Cet article	RL-GO	RL-GO	Cet article	RL-GO	RL-GO
		λ /100	λ /dix		λ /100	λ /dix
Papier	207.96	186,4	2303.3	557.6	130.9	181.4
Papier	433.195	103,1	1427.43	552,1	136,4	185.1

Figure 13a VV Polarisation
Figure 13b HHtion
Figure 13a VV 207,96 186,4 557,6 181,4
Graphique 13b Ménages 103,1 552,1 185,1

Il existe une légère différence entre notre méthode SBR accélérée par GPU et la méthode parallèle
La méthode RL-GO avec une densité de rayons de λ /100 dans FEKO en raison de l'effet de la densité des rayons, et le
La méthode RL-GO avec une densité de rayons de λ 100 prend également en compte le champ de diffraction [53].

4.2. Simulations d'imagerie ISAR

Dans cette section, les résultats d'imagerie ISAR d'une cible d'avion représentative sont présentés, calculés
et discuté. La matrice du champ rétrodiffusé des cibles aériennes représentatives est
par notre méthode SBR accélérée par GPU, puis les échos sont focalisés pour obtenir
ISAR
Images ISAR en utilisant l'algorithme BP accéléré par GPU développé dans cet article. Pour
calculé par notre méthode SBR accélérée par GPU, puis les échos sont focalisés pour obtenir des images ISAR en
Dans cette section, les résultats d'imagerie ISAR d'une cible d'avion représentative sont présentés et discutés.
utilisant l'algorithme BP accéléré par GPU développé dans cet article. À titre de comparaison, notre algorithme BP
La matrice du champ rétrodiffusé de cibles d'avions représentatives est calculée par notre méthode SBR accélérée
accéléré par GPU a également été appliqué pour concentrer le
par GPU, puis les échos sont focalisés pour obtenir des images ISAR en utilisant l'algorithme BP accéléré par GPU
développé dans cet article. À titre de comparaison, notre algorithme BP accéléré par GPU a également été utilisé
pour concentrer le

La méthode RL-GO de FEKO dans les conditions de la même densité de rayons, et elle surpasse même la méthode RL-GO sous des angles spécifiques. Par exemple, sur les figures 16a, d, la différence de diffusion dans la partie moteur de l'avion peut être considérée comme plus évidente, et les différences de diffusion dans la zone moteur de l'avion peuvent être considérées comme plus évidentes. évident sur la figure 16c, f. La figure 16f présente un fort fouillis qui submerge les informations telles que les caractéristiques structurelles de l'avion, et les résultats de la figure 16f ne sont pas aussi bons que la comparaison. Notre algorithme BP accéléré par GPU a également été appliqué pour concentrer les signaux rétrodiffusés, utilisant la méthode RL-GO dans le logiciel FEKO v2020 pour obtenir un ISAR focalisé. Les caractéristiques structurelles géométriques de l'avion sont détaillées, donc la figure 16g–j'ai très bien. La figure 14 montre le modèle CAO d'un modèle d'avion A380 à l'échelle avec ses dimensions. Ce comporte 3 770 triangles. Les figures 15a à c illustrent trois configurations d'observation typiques résultats d'imagerie comparés aux figures 16a – c et 16d – f, ce qui confirme que le modèle effectif balayage azimutal. L'angle d'incidence est fixé à θ = 60◦ . Le ness des algorithmes d'imagerie BP accélérés par GPU dans cet article de côté. Table avec différentes plages de montre les comparaisons du temps de calcul et de la mémoire maximale pour la figure 16.

Les paramètres d'imagerie ISAR pour les figures 16 et 18 sont définis comme dans le tableau 2.

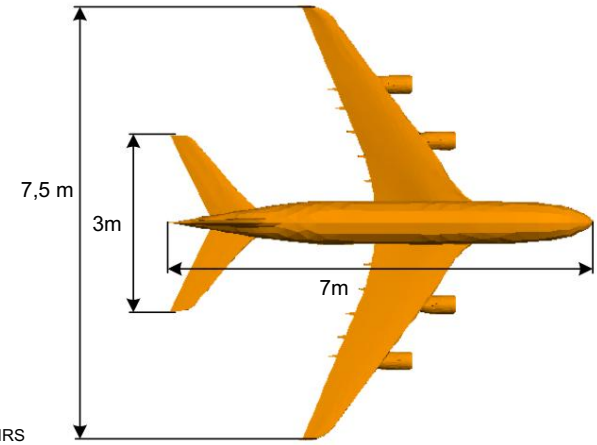
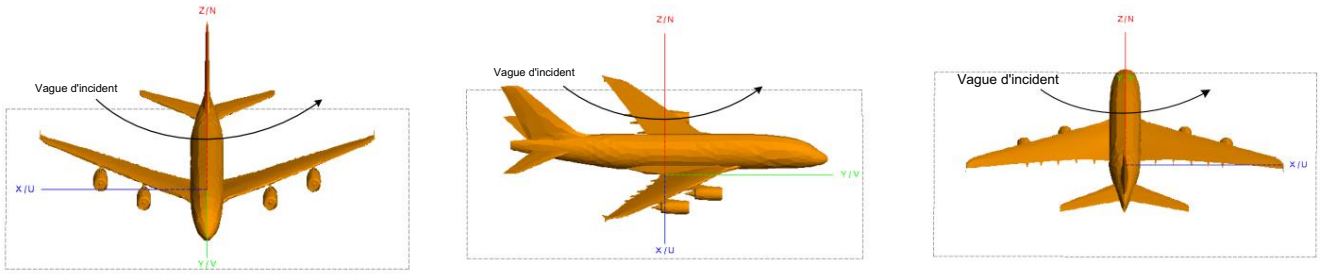


Figure 14. Modèle de conception assistée par ordinateur (CAO) et dimensions d'un modèle d'avion A380 à l'échelle.

Figure 14. Modèle de conception assistée par ordinateur (CAO) et dimensions d'un modèle d'avion A380 à l'échelle.



(un) (b) (c)

Figure 15. Trois configurations d'observation typiques avec différents angles d'azimut sous angle fixe. Figure 15. Trois configurations d'observation typiques avec différents angles d'azimut sous angle fixe angle d'incidence θ = 60◦ . (a) φ = 45◦ ~ 135◦ ; (b) φ = −45◦ ~ 45◦ ; (c) φ = 135◦ ~ −45◦ . angle de descente θ = 60◦ (une) φ = 45 ~ 135◦ ; (b) φ = −45 ~ 45◦ ; (c) φ = 135◦ ~ −45◦ .

Tableau 2. Paramètres de simulation ISAR pour les figures 16 et 18.

Paramètre	Valeur
f0	1,75 GHz
B	3 GHz
Δφ	90
	60 ou 120
θ Δx	0,05 m
Δy	0,05 m
Points de prélèvement	200
Polarisation	VV

La figure 16 présente les résultats d'imagerie ISAR pour trois configurations d'observation typiques avec différentes plages de balayage azimutal. L'angle d'incidence est fixé à θ = 60◦ . Notre Un algorithme d'imagerie BP accéléré par GPU est appliqué pour concentrer les champs de rétrodiffusion sur obtenir des images ISAR focalisées. Sur la figure 16, les figures 16d à f sont RL-GO à une densité de rayons de λ/10 sans diffraction et la figure 16g – i sont RL-GO à une densité de rayons de λ/100 avec diffraction.

En comparant les figures 16a à c et 16d à f, on peut constater que les résultats de la méthode de cet article et la méthode RL-GO sont en meilleur accord avec celles obtenues par la méthode RL-GO de FEKO

(d) (e) (F)

méthode dans les conditions de la même densité de rayons, et elle surpasse même le RL-GO à angles spécifiques. Par exemple, sur les figures 16a et d, la différence de diffusion dans le moteur partie de l'avion peut être considérée comme plus évidente, et les différences dans la diffusion dans la zone du moteur de l'avion est plus évidente sur la figure 16c, f ; Figure 16f a un fort encombrement qui submerge les informations telles que les caractéristiques structurelles du avion, et les résultats de la figure 16f ne sont pas aussi bons que ceux de la figure 16c. Quand le la densité des rayons est λ/100, les échos sont capables d'enregistrer les caractéristiques structurelles géométriques de l'avion en détail, donc figure 16g – j'ai de très bons résultats d'imagerie par rapport aux figures 16a – c et 16d – f, qui confirment l'efficacité de l'imagerie BP accélérée par GPU dans cet article vu de côté. Le tableau 3 montre le temps de calcul et le pic comparaisons de mémoire pour la figure 16.

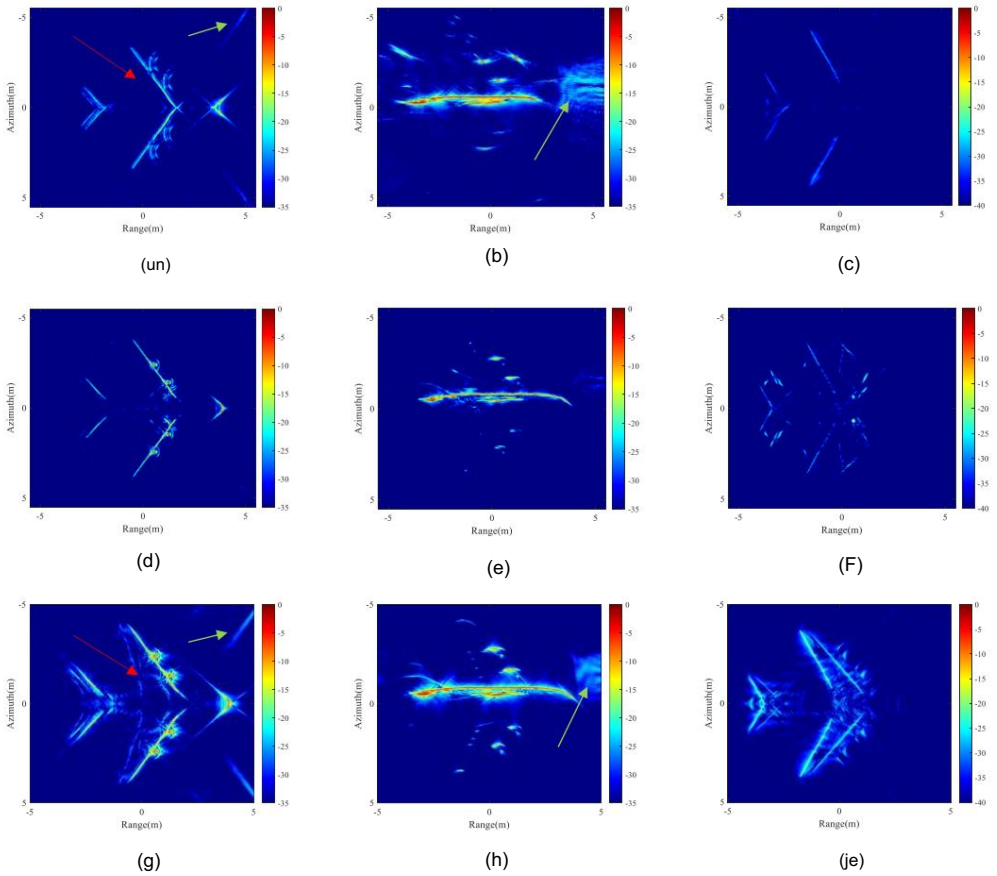


Figure 16. Résultats d'imagerie SAR utilisant l'algorithme d'imagerie BP accéléré par GPU sous incidence fixe angle θ = 60°. Dans (a – c), les champs de rétrodiffusion sont calculés par notre méthode SBR accélérée par GPU avec une densité de rayons de λ/10 ; en (d – f), les champs de rétrodiffusion sont obtenus par le RL-GO de FEKO avec la méthode avec une densité de rayons de λ/10 ; dans (g – i), les champs de rétrodiffusion sont obtenus par RL-GO avec un rayon de rayons de λ/100. (a,d,g) sont les résultats pour l'angle d'azimut φ = 45° ~ 135° ; (b,e,h) sont les résultats pour l'angle d'azimut φ = −45° ~ 45° ; (c,f,i) sont les résultats pour l'angle d'azimut φ = −135° ~ −45° ; (a,d) sont les résultats pour l'angle d'azimut φ = 45° ~ 135° ; (b,e) sont les résultats pour l'angle d'azimut φ = −45° ~ 45° ; (c,f) sont les résultats pour l'angle d'azimut φ = −135° ~ −45°.

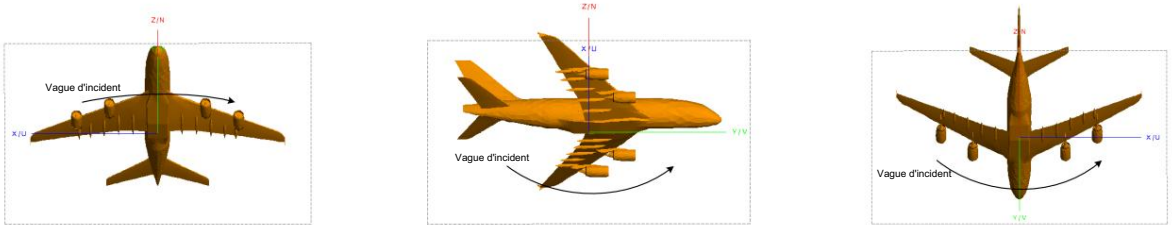
Tableau 3. Temps de calcul et mémoire maximale pour la figure 16.

	SBR		RL-GO	
	Temps (s)	Mémoire (Mo)	Temps (s)	Mémoire (Mo)
Figures 16a,d	423,2 668,8 72,5	79,2 78,4 2 623,3	1 584,2	177,8
Figure 16b,e	423,2 668,8 72,5	79,2 78,4 2 623,3	1 614,3	179,6
Graphique 16c,f	423,2 668,8 72,5	79,2 78,4 2 623,3	1 474,1	179,5

On peut conclure que lorsque la densité des rayons est la même, les résultats de la méthode présentée dans cet article sont en bon accord avec ceux de RL-GO. Ensuite, nous analysons la différence entre les résultats des deux méthodes lorsque les densités de rayons ne sont pas les mêmes. Sur la figure 16a, il y a de forts échos provenant à la fois du moteur et de la partie d'aile fixée au moteur. La zone indiquée par la flèche rouge en (a) représente la position de l'aile. Aux angles d'observation, le rayon sera réfléchi une fois après avoir heurté l'aile. On peut conclure que lorsque la densité des rayons est la même, les résultats de la méthode dans

partie en position 1. En raison de la planéité de cette partie de la structure, les rayons de cet article sont en bon

accord avec ceux de RL-GO. Ensuite, nous analysons la différence entre les résultats des deux méthodes lorsque les densités de rayons ne sont pas les mêmes. Dans la diffusion multiple est prise en compte par le lancer de rayons, dans lequel la réflexion maximale vers le moteur. La zone indiquée par la flèche rouge 1 en (a) représente la position de l'aile en position 1. En raison de la planéité de cette partie de la structure, le rayon va quittez ensuite la cible après avoir été réfléchi. Dans notre méthode SBR accélérée par GPU, figure 16a, il y a de forts échos provenant à la fois du moteur et de la partie de l'aile attachée, le nombre est 10, tandis que le champ de diffraction du aux bords de la cible n'est pas pris en compte, ce qui conduit à une différence prononcée entre la position 1 dans (a) et la position 2 dans (g). Des points de diffusion apparaissent aux positions indiquées par les flèches vertes et sur la figure 16a,b,g,h. Les figures 16a, b sont les résultats calculés par la méthode de cet article, la diffusion multiple est prise en compte par lancer de rayons, dans laquelle la réflexion maximale et les figures 16g, h sont les résultats calculés par le RL-GO avec une diffraction à une densité de rayons de égal à 10, tandis que le champ de diffraction du aux bords de la cible n'est pas pris en compte, /100 en FEKO. Nos SBR et RL-GO accélérés par GPU dans FEKO sont tous deux basés sur les rayons, ce qui entraîne une différence prononcée entre la position 1 dans (a) et la position 2 dans (g). Des points de diffusion sont liés au mécanisme de diffusion de l'électromagnétique. Ces points de diffusion sont liés au mécanisme de diffusion des ondes électromagnétiques et le mécanisme de lancer de rayons. En raison de la négligence du champ de diffraction, En raison des bords cibles, l'écho cible sur la figure 16c est plus faible que celui sur la figure 16i. λ/100 en FEKO. Une comparaison des résultats d'imagerie ISAR de la figure 16a – c avec la figure 16d – f et la figure. Ces points de 16g – i démontre la faisabilité et l'efficacité de notre système en combinant l'accélération GPU et le mécanisme de diffusion sont liés au mécanisme de diffusion des ondes électromagnétiques. traçage de rayons. En raison de la négligence du champ de diffraction résultant de SBR géré avec l'algorithme BP pour une simulation d'imagerie ISAR rapide. bords cibles, l'écho cible sur la figure 16c est plus faible que celui sur la figure 16i. Une comparaison résultats d'imagerie ISAR de la figure 16a à c avec les figures 16d à f et 16g à i démontrent le Les figures 17a à c illustrent trois configurations d'observation typiques avec différents azi- des efficacité de notre système en combinant GPU accéléré SBR avec le BP plages de balayage muthal. Résultats d'imagerie ISAR pour trois configurations d'observation typiques, faisabilité et avec différentes plages de balayage azimutal sont illustrées respectivement sur les figures 18a, b et c. algorithme pour une simulation rapide d'imagerie ISAR. Les figures 17a à c illustrent trois configurations d'observation typiques avec différents az- θ = 120 Résultats d'imagerie ISAR pour trois configurations d'observation typiques différentes plages de balayage azimutal sont illustrées sur les figures 18a, b et c, respectivement. l'angle d'incidence est défini comme θ = 120 , et les autres paramètres sont les identiques à ceux de la figure 16. Dans les figures 18a à c, les champs de rétrodiffusion sont calculés par notre correspondante dans le logiciel FEKO v2020, où (d – f) sont RL-GO à une densité de rayons de λ/10 et (g – i) sont RL-GO à une densité de rayons de λ/100. Puisque les densités de rayons sont les mêmes et aucun d'entre eux n'inclut le champ enveloppant, les résultats pour les plages d'angles correspondantes en bon accord. Une comparaison des résultats de (c) et (f) montre d'écho du corps de l'avion est très faible en (f). Il y a aussi du bruit. Seulement le sous les mêmes densités de rayons et il peut mieux enregistrer la structure géométrique de la cible. La zone du moteur présente de forts échos, ce qui rend la méthode de cet article plus efficace sous la de rayons et peuvent mieux enregistrer les informations sur la structure géométrique de la cible. Dans formation. En général, les images ISAR focalisées des champs rétrodiffusés sont calculées par nos mêmes densités images ISAR focalisées des champs rétrodiffusés calculées par notre méthode SBR accélérée par GPU La méthode SBR accélérée par GPU est en bon accord avec celles obtenues par le RL de FEKO. En général, les Méthode GO. Certaines taches de diffusion peuvent être observées, comme l'indiquent les flèches rouges dans la ure 18a,b,g,h comme sur la figure 16. Cela est dû à la diffusion multiple des ondes électromagnétiques, des points de diffusion peuvent être observés comme indiqué par les flèches rouges sur la figure 16a,b,g,h, comme les surfaces diffusion multiples des rayons réfléchis plusieurs fois sur la cible. dans la figure 16. Cela surface. Pour les rayons réfléchis plusieurs fois, l'histoire de phase de l'onde électromagnétique est l'onde magnétique est déformée. Par conséquent, les points de diffusion indiqués par les flèches rouges sont déformés. Par conséquent, les points de diffusion indiqués par les flèches rouges sur les figures 18a, b, g, h Les figures 18a, b, g, h peuvent être éliminées en réduisant les nombres de réflexion dans le traçage de rayons peuvent être éliminées en réduisant les nombres de réflexion dans l'algorithme de traçage de rayons. Tableau 4 algorithme. Le tableau 4 montre les comparaisons du temps de calcul et de la mémoire maximale pour la figure 18. 18.



(un) (b) (c)

Figure 17. Trois configurations d'observation typiques avec différents angles d'azimut sous angle fixe d'incidence θ = 120 . (a) φ = 45 ~ 135 ; (b) φ = −45 ~ 45 ; (c) angle φ = −135 ~ −45 .

Figure 17. Trois configurations d'observation typiques avec différents angles d'azimut sous angle fixe

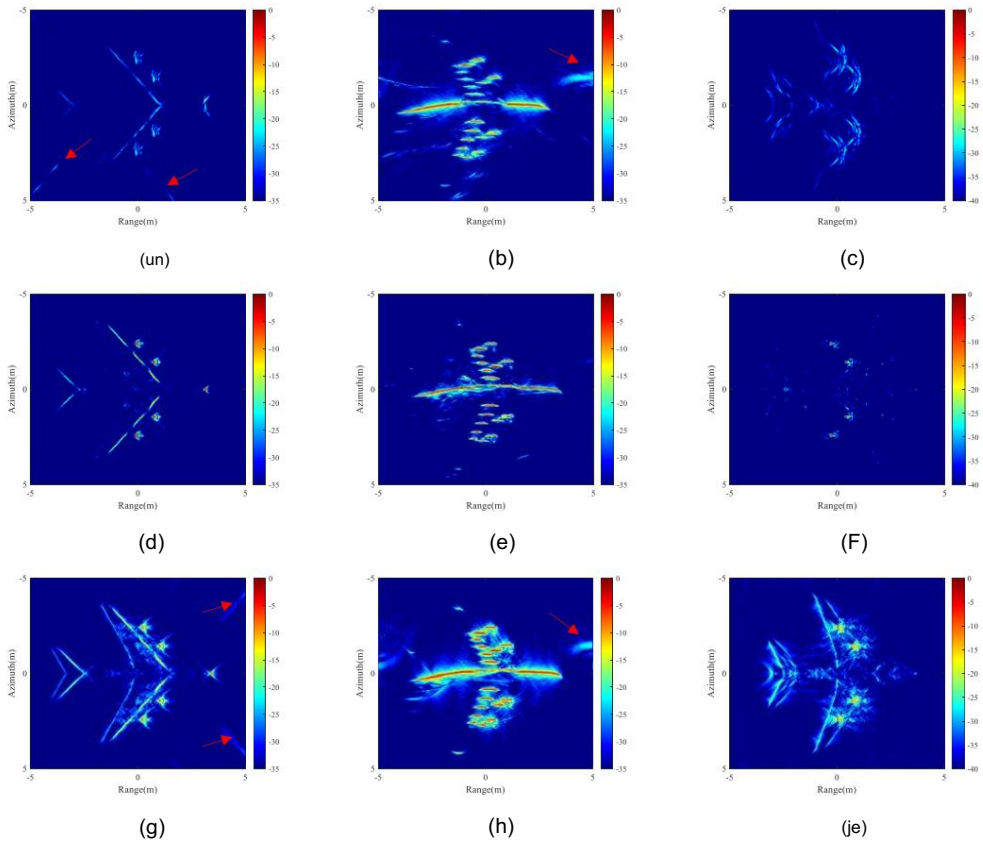


Figure 18. Similaire à la figure 16 mais avec un angle d'incidence $\theta = 120^{\circ}$.

Tableau 4. Temps de calcul et mémoire maximale pour la figure 18.

Tableau 4. Temps de calcul et mémoire maximale pour la figure 18.

	SBR		RL-GO	
	Temps (s)	Mémoire (Mo)	Temps (s)	Mémoire (Mo)
Figure 18a,d	2815,2	72,4	1655,6	194,6
Figure 18b,e	2647,05	74,2	1702,4	202
Figure 18c,f	2872,9	77,1	1764,2	196,9

La figure 19 montre un modèle CAO d'une cible d'avion électriquement de grande taille, qui appartient à la catégorie électriquement grande taille avec une taille électrique de 171 × 104 longueurs d'onde. Ce modèle CAO comporte 712 éléments de visage. Les images ISAR pour ce modèle ont été calculées en utilisant la méthode de cet article et la méthode RL-GO dans FEKO. Les densités de rayons pour les deux modèles CAO La méthode dans cet article et la méthode RL-GO sont définies sur $\lambda/10$. Les centres azimutaux de la 90° -90° dans cet article et la méthode RL-GO sont définis sur 0°. Les centres azimutaux de 0° -90° sont définis sur -90°. La figure 20 montre trois plages d'azimut différentes. Dans les figures 20a à c, $\varphi = -90^{\circ}$, 0°, 90°. La figure 20 montre trois plages d'azimut différentes. Dans les simulations suivantes de la figure 21, les paramètres d'imagerie ISAR sont définis comme dans le tableau 5.

La figure 19 montre un modèle CAO d'une cible d'avion électriquement de grande taille, qui est utilisant la catégorie électriquement de grande taille avec une taille électrique de 171 × 104 longueurs d'onde. comportent 712 éléments de visage. Les images ISAR pour ce modèle ont été calculées en utilisant la méthode dans cet article et de la méthode RL-GO dans FEKO. Les densités de rayons pour les deux 90° -90° dans cet article et la méthode RL-GO sont définies sur $\lambda/10$. Les centres azimutaux de 0° -90° sont définis sur -90°. La figure 20 montre trois plages d'azimut différentes. Dans les simulations suivantes de la figure 21, les paramètres d'imagerie ISAR sont définis comme dans le tableau 5.

La figure 21 montre les résultats d'imagerie ISAR des deux méthodes pour des images électriquement de grande taille. cibles d'avions, avec des densités de rayons de λ/10 pour cette méthode papier et RL-GO dans FEKO. comparant les figures 21b et 21e, 1 et 2 sur la figure 21b sont les informations structurelles de l'avion, et 1 est l'aile et 2 est la queue, alors que la figure 21e ne peut pas le montrer informations structurelles, la méthode de cet article est donc meilleure que le résultat de l'imagerie RL-GO quand $\varphi = -45^{\circ}$ 45°. En comparant les figures 21a et 21d, le résultat de la figure 21a est évidemment meilleure que celle de la figure 21d, où la distorsion de l'historique de phase se produit à l'endroit indiqué par les flèches blanches sur la figure 21d. Informations détaillées sur les ailes, le nez et la queue de l'avion peut être clairement affichée sur la figure 21a. Cela se produit également sur la figure 21c,f, où pointent les flèches blanches de la figure 21c,f. On voit clairement que les taches éparses d'échantillonnage qui apparaissent à ces endroits indiqués par les flèches ne sont pas des informations structurelles de l'avion. est un fait bien établi que les ondes électromagnétiques présentent un effet de trajets multiples pendant

Tableau 5. Paramètres de simulation ISAR pour la figure 21.

Paramètre	Valeur
f_0	1,75 GHz
B	6 GHz
θ_0	90
θ	60
x	0,05 m
y	0,05 m
Les points	600
Polarisation II	VV

propagation. Ce phénomène implique qu'une onde incidente traverse une multitude de chemins pour atteindre un point de réception désigné. Modifications des données de l'historique des phases causées par ces différents chemins peuvent être superposés, ce qui entraîne une distorsion de l'historique des phases. Le tableau 6 montre comparaisons du temps de calcul et de la mémoire maximale pour la figure 21.

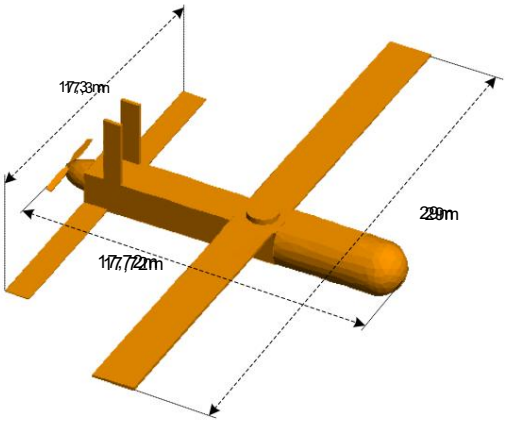


Figure 19. Modèle CAO d'un gros avion électrique.

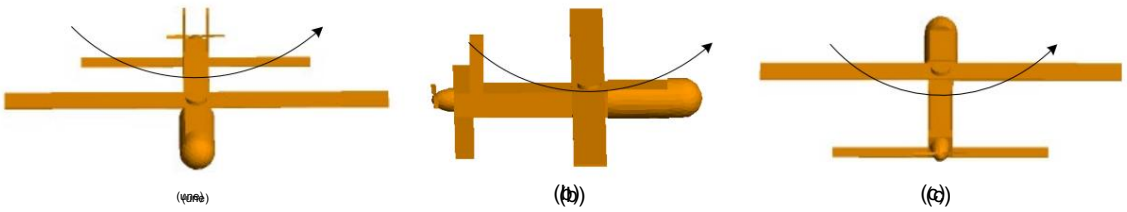


Figure 20. Trois configurations d'observation typiques avec différents angles d'azimut sous incidence fixe angle d'incidence θ = 60°. (a) φ= 45° ~ 135°; (b) φ= −45° ~ 45°; (c) φ= −135° ~ −45°.

Tableau 5. Paramètres de simulation ISAR pour la figure 21.

Paramètre	Valeur
f0	1,75 GHz
B	3 GHz
Δφθ	90
	60
Δx	0,05 m
Δy	0,05 m
Points de prélèvement	600
Polarisation	VV

Tableau 6. Temps de calcul et mémoire maximale pour la figure 21.

	SBR		RL-GO	
	Temps (heures)	Mémoire (Mo)	Temps (heures)	Mémoire (Mo)
Figures 21a,d	20.1	141,2	16,3	269,6
Figure 21b,e	19,6	140,7	15,5	258,5
Graphique 21c,f	20,7	145,6	16,9	273,1

Les simulations numériques de l'imagerie ISAR montrent clairement que de forts centres de diffusion ainsi que des profils de cibles peuvent être observés sous un grand azimut d'observation. Figure 21. Résultats d'images ISAR utilisant l'algorithme RD accéléré par GPU sous angles d'incidence fixes et large bande passante. Il est également indiqué que les images ISAR sont très sensibles à θ = 60 (a,c) les champs de rétrodiffusion sont calculés par la SBR accélérée par GPU par rapport aux angles d'observation. En raison de la diffusion multiple, plusieurs patchs triangulaire méthode, dans (d–f) les champs de rétrodiffusion sont obtenus par la méthode RL-GO de FEKO avec une densité de rayons de 10 ; (a,d) sont les résultats pour l'azimut φ = 45 ~ 135°; (b,e) sont les résultats pour l'azimut φ = −45 ~ 45°; (c,f) sont les résultats pour l'azimut φ = −135 ~ −45°. Polarisation VV.

Figure 19. Modèle CAO d'un gros avion électrique.

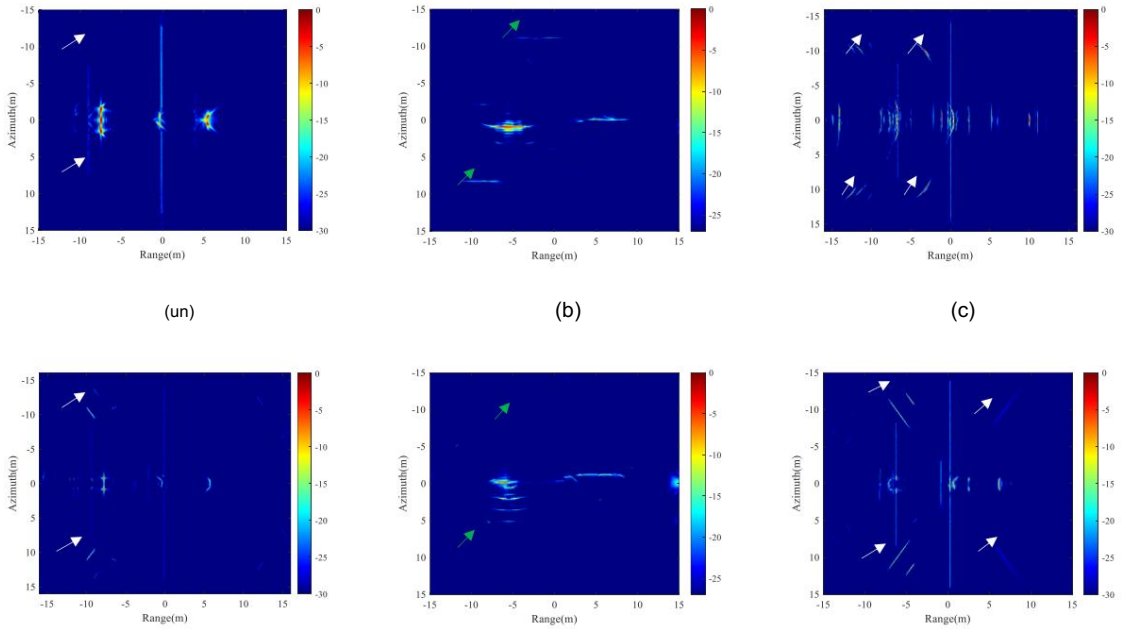
frappés par des rayons identiques, ce qui entraîne une distorsion de l'histoire de phase des ondes électromagnétiques.

(un) La distorsion de l'historique de phase est un problème courant avec les méthodes par rayons. Ainsi, des lobes secondaires évidents (b) (c)

peuvent être observés sur des images ISAR focalisées. En comparaison avec RL-GO dans FEKO v2020

Figure 20. Trois configurations d'observation typiques avec différents angles d'azimut sous un logiciel inci- fixe , la faisabilité et l'efficacité de notre schéma sont démontrées en combinant l'angle de densité GPU. (une) θ = 60 = 45 ~ 135 ; (b) = −45 ~ 45 ; (c) φ= − .

SBR accéléré avec algorithme BP pour des simulations d'imagerie ISAR rapides sous grand angle et des conditions de large bande passante.



(un) (b) (c)

(d) (e) (F)

Figure 21. Résultats d'imagerie ISAR utilisant un algorithme d'imagerie de la pression artérielle accéléré par GPU sous incidence fixe

GPU. θ = 60 ◦ dans (a–c), les champs de rétrodiffusion sont calculés par notre méthode SBR accélérée par angle θ = 60 ◦ dans (a–c), les champs de rétrodiffusion sont calculés par notre méthode SBR accélérée par GPU ; dans méthode; dans (d – f), les champs de rétrodiffusion sont obtenus par la méthode RL-GO de FEKO à une densité de rayons de (d – f), les champs de rétrodiffusion sont obtenus par la méthode RL-GO de FEKO à une densité de rayons de λ/10 ; (un défi /dix ; (a,d) sont les résultats pour l'azimut φ = 45 ◦ ~ 135 ◦ ; (b,e) sont les résultats pour l'azimut φ = −45 ◦ ~ 45 ◦ ; résultats pour l'azimut φ= 45 ◦ 135 ◦ ; (b,e) sont les résultats pour l'azimut φ= −45 ◦ 45 ◦ ; (c,f) sont des résultats (c,f) sont des résultats pour l'azimut φ = 135 ◦ − 135 ◦ ~ 45 ◦ , Polarisation VV. pour l'azimut φ= 135 ◦ − 45 ◦ , Polarisation VV.

5. Conclusions

Dans cet article, une nouvelle méthode de rayons rebondissants basée sur l'accélération des arbres GPU et BVH. est présenté pour les simulations d'imagerie ISAR. Il utilise C++AMP pour réaliser le parallèle GPU calcul d'accélération, dans lequel une structure arborescente BVH est construite selon le structure cible, améliorant ainsi l'efficacité de l'intersection des rayons. La méthode SAH est incorporée dans la division du cadre de délimitation de la scène, atténuant ainsi efficacement l'impact de la délimitation de la scène. boîte se chevauchant sur l'efficacité de traversée des rayons de la structure arborescente BVH. Pour efficacement effectuer des simulations d'imagerie ISAR, un algorithme d'imagerie BP accéléré basé sur GPU a été développé grâce à CUDA. La précision du SBR accéléré par GPU est validée en comparant le RCS calculé par notre méthode SBR avec celui obtenu par RL-GO en FEKO. Il est démontré que le SBR accéléré par GPU présenté montre une bonne validité et la fiabilité. Pour les simulations d'imagerie ISAR, en prenant un A380 et un avion simplifié modèle à titre d'exemple, les champs de rétrodiffusion ont été calculés à l'aide du modèle accéléré par GPU. Algorithme SBR sous grands angles d'azimut, Δφ= 90◦, et de larges bandes passantes, 3 GHz. Le les échos rétrodiffusés sont focalisés à l'aide de l'algorithme d'imagerie BP accéléré par GPU, et le les images ISAR focalisées de notre méthode SBR accélérée par GPU sont en bon accord avec celles de la méthode RL-GO de FEKO, indiquant la faisabilité et l'efficacité de notre accélération GPU Algorithme d'imagerie BP ISAR. Les simulations indiquent que de forts centres de diffusion car les profils de cibles peuvent être observés clairement à partir des images ISAR sous de grands angles d'observation et de larges bandes passantes. Des lobes secondaires évidents dans les images ISAR focalisées peuvent être observés, en raison

à l'histoire de phase des ondes électromagnétiques déformées résultant de l'effet multipolaire diffusion. Les simulations numériques indiquent également que les résultats de l'imagerie ISAR sont très sensible aux angles d'observation. Dans le futur, les travaux actuels seront étendus pour essaimer l'imagerie ISAR des cibles ainsi que l'imagerie ISAR 3D en développant un système efficace algorithme de simulation électromagnétique en combinant SBR et PTD accélérés par GPU pour en tenant compte de la diffraction de bord.

Contributions des auteurs : Méthodologie, JC, RZ, PY et RW ; logiciel, JC; validation, JC ; officiel analyse, JC ; simulation de lancer de rayons, JC ; enquête, JC et PY ; ressources, PY ; écriture – originale préparation du projet, JC ; rédaction – révision et édition, PY ; visualisation, JC ; supervision, PY; projet administration, JC et PY Tous les auteurs ont lu et accepté la version publiée du manuscrit.

Financement : Ce travail a été soutenu en partie par la Fondation nationale des sciences naturelles de Chine. [Grant Nos. 62061048, 62261054 et 62361054], en partie par le Shaanxi Key Research and Development Programme [Grant Nos. 2023-YBGY-254 et 2024GXYBXM-108], en partie par le Natural Science Basic Plan de recherche dans la province chinoise du Shaanxi (subvention n° 2023-JC-YB-539), et en partie par le diplômé Programme d'innovation éducative de l'Université de Yan'an (subvention n° YCX2024086).

Déclaration de disponibilité des données : Les données présentées dans cette étude sont disponibles sur demande auprès du auteur correspondant.

Remerciements : Les auteurs remercient les éditeurs et les relecteurs pour leurs suggestions constructives.

Conflits d'intérêts : Les auteurs ne déclarent aucun conflit d'intérêts.

Abréviations

Les abréviations suivantes sont utilisées dans ce manuscrit :

PA	Rétroprojection
BVH	Hiérarchies de volumes englobants
C++AMP	C++ Parallélisme massif accéléré
CUDA	Architecture de périphérique unifiée de calcul
GOUJAT	Conception assistée par ordinateur
CPU	Unité centrale de traitement
GED	Moment dipolaire équivalent
FFT	Transformée de Fourier Rapide
GPU	Processeur graphique
ALLER	Optique Géométrique
ISAR	Radar à synthèse d'ouverture inversée
IFFT	Transformée de Fourier rapide inverse
Arbre Kd	Arbre à K dimensions
MOM	Méthode des Moments
PTD	Théorie physique de la diffraction
PO	Optique physique
PEC	Conducteur électrique parfait
RCS	Section efficace du radar
RL-GO	Optique géométrique de lancement de rayons
RMSE	Erreur quadratique moyenne
SBR	Rayon de tir et rebondissant
SAH	Heuristique de surface
SMS	Multiprocesseurs de streaming
SP	Processeurs de streaming
SIMT	Instruction unique, plusieurs threads
TDSBR	Tir dans le domaine temporel et rayon rebondissant
TDPO	Optique physique dans le domaine temporel

Les références

1. Bhalla, R. ; Ling, H. Formation d'images ISAR à l'aide de données bistatiques calculées à partir de la technique de prise de vue et de rayons rebondissants. J. Électromagn. Application Vagues. 1993, 7, 1271-1287. [Référence croisée]
2. Lui, XY ; Zhou, XY; Cui, TJ Simulation rapide d'images 3D-ISAR de cibles à des angles d'aspect arbitraires grâce à une transformation de Fourier rapide non uniforme (NUFFT). IEEETrans. Propag. d'antennes. 2012, 60, 2597-2602. [Référence croisée]
3. Zhang, K. ; Wang, CF; Jin, JM Broadband Monostatic RCS et calcul ISAR de cavités ouvertes grandes et profondes. IEEETrans. Propag. d'antennes. 2018, 66, 4180-4193. [Référence croisée]
4. Prickett, M. ; Chen, C. Principes du radar/ISAR/imagerie à synthèse inverse. Dans Actes de l'EASCON 1980, Conférence sur l'électronique et les systèmes aérospatiaux, Arlington, VA, États-Unis, 29 septembre-1er octobre 1980 ; pp. 340-345.
5. García-Fernández, AF ; Yeste-Ojeda, OA; Grajal, J. Modèle à facettes de cibles mobiles pour l'imagerie ISAR et la rétrodiffusion radar Simulation. IEEETrans. Aérosp. Électron. Système. 2010, 46, 1455-1467. [Référence croisée]
6. Lee, JI; Yun, DJ ; Kim, HJ ; Yang, Wyoming ; Myung, NH Formations d'images ISAR rapides sur des angles multiaspects à l'aide de la prise de vue et Rayons rebondissants. Fil d'antennes IEEE. Propagande. Lett. 2018, 17, 1020-1023. [Référence croisée]
7. Guo, G. ; Guo, L. ; Wang, R. ; Li, L. Un cadre d'imagerie ISAR pour des cibles grandes et complexes utilisant TDSBR. Antennes IEEE Fil. Propagande. Lett. 2021, 20, 1928-1932. [Référence croisée]
8. Meng, W. ; Li, J. ; Xi, YJ; Guo, LX ; Li, ZH; Wen, SK Une méthode améliorée de tir et de rayon rebondissant basée sur Blend-Tree pour la diffusion EM de plusieurs cibles mobiles et l'analyse de l'écho. IEEETrans. Propag. d'antennes. 2024, 72, 2723-2737. [Référence croisée]
9. Yang, PJ ; Wu, R. ; Ren, XC ; Zhang, YQ; Zhao, Y. Spectres Doppler d'ondes électromagnétiques diffusées par un objet volant au-dessus de surfaces marines non linéaires variant dans le temps. J. Electromagn. Application Vagues. 2019, 33, 2175-2198. [Référence croisée]
10. Ling, H. ; Chou, RC; Lee, SW Tir et rayons rebondissants : Calcul du RCS d'une cavité de forme arbitraire. IEEETrans. Propag. d'antennes. 1989, 37, 194-205. [Référence croisée]
11. Xu, F. ; Jin, YQ Traçage de rayons analytique bidirectionnel pour le calcul rapide de la diffusion composite à partir d'une grande cible électrique sur un surface aléatoirement rugueuse. IEEETrans. Propag. d'antennes. 2009, 57, 1495-1505. [Référence croisée]
12. Tao, Y. ; Lin, H. ; Bao, H. Méthode de prise de vue et de rayon rebondissant basée sur GPU pour une prévision RCS rapide. IEEETrans. Propag. d'antennes. 2010, 58, 494-502. [Référence croisée]
13. Meng, W. ; Li, J. ; Guo, LX ; Yang, QJ Une méthode SBR accélérée pour la prédiction RCS d'une cible électriquement grande. Antennes IEEE Fil. Propagande. Lett. 2022, 21, 1930-1934. [Référence croisée]
14. Baden, JM; Tripp, inversion de VK Ray dans les calculs SBR RCS. Dans les actes de la 31e Revue internationale des progrès en matière de Electromagnétique computationnelle appliquée (ACES), Williamsburg, VA, États-Unis, 22-26 mars 2015 ; p. 1–2.
15. Feng, TT; Guo, LX Un algorithme de lancer de rayons amélioré pour le calcul de diffusion EM basé sur SBR de cibles électriquement grandes. Fil d'antennes IEEE. Propagande. Lett. 2021, 20, 818-822. [Référence croisée]
16. Yun, KC ; Fu, WC Implémentation GPU efficace de la méthode SBR-PO haute fréquence. Fil d'antennes IEEE. Propagande. Lett. 2013, 12, 941-944. [Référence croisée]
17. Wu, R. ; Wu, PAR ; Lui, PX ; Guo, Kentucky ; Sheng, XQ Un algorithme d'expansion rapide des ondes planes pour une analyse de diffusion rigoureuse de Cibles d'essaim. IEEETrans. Propag. d'antennes. 2023, 71, 7426-7437. [Référence croisée]
18. Dong, C.-L.; Guo, L.-X.; Meng, X. Un algorithme accéléré basé sur GO-PO/PTD et CWMFSM pour la diffusion EM depuis le navire sur une surface de la mer et la formation d'images SAR. IEEETrans. Propag. d'antennes. 2020, 68, 3934-3944. [Référence croisée]
19. Dong, C.-L. ; Guo, L.-X.; Meng, X. ; Wang, Y. Un SBR accéléré pour la diffusion EM à partir d'objets complexes électriquement grands. Fil d'antennes IEEE. Propagande. Lett. 2018, 17, 2294-2298. [Référence croisée]
20. Meng, W. ; Li, J. ; Chai, SR ; Xi, YJ; Wen, Saskatchewan ; Liu, RF Une méthode SBR-PTD améliorée pour la diffusion EM à partir d'une cible mobile. Dans les actes du symposium 2023 de la Société internationale d'électromagnétique computationnelle appliquée (ACES-Chine), Hangzhou, Chine, 15-18 août 2023 ; p. 1–3.
21. Li, Hz ; Dong, CL; Meng, X. ; Guo, LX ; Wei, QH Une nouvelle méthode hybride MoM-SBR basée sur le moment dipolaire équivalent pour le calcul de diffusion EM de cibles complexes électriquement grandes. Dans les actes de la Conférence internationale 2023 sur la technologie des micro-ondes et des ondes millimétriques (ICMMT), Qingdao, Chine, 14-17 mai 2023 ; p. 1–3.
22. Wang, Z. ; Wei, F. ; Huang, Y. ; Zhang, X. ; Zhang, Z. Méthode d'imagerie RD améliorée basée sur le principe du calcul étape par étape. Dans les actes du 2e Symposium international SAR de Chine (CISS) 2021, Shanghai, Chine, 3-5 novembre 2021 ; p. 1 à 5.
23. Dong, L. ; Han, S. ; Zhu, D. ; Mao, X. Un algorithme de format polaire modifié pour le SAR à missiles hautement louches. Géosci IEEE. Télédétection Lett. 2023, 20, 1–5. [Référence croisée]
24. Boag, AGA Imagerie par rétroprojection d'objets en mouvement. IEEETrans. Propag. d'antennes. 2021, 69, 4944-4954. [Référence croisée]
25. Zhang, M. ; Ren, Z. ; Zhang, G. ; Zhang, C. Imagerie THz ISAR utilisant un algorithme de rétroprojection à compensation de phase accélérée par GPU. J. Infrarouge Millim. Vagues 2022, 41, 448-456. [Référence croisée]
26. Arikan, O. ; Munson, DC, Jr. Un nouvel algorithme de rétroprojection pour le SAR et l'ISAR en mode projecteur. Dans les Actes du Haut Speed Computing II, Los Angeles, Californie, États-Unis, 17-18 janvier 1989 ; pp. 107-117.
27. Gong, H. ; Liu, Y. ; Chen, X. ; Wang, C. Optimisation de scène de l'algorithme de rétroprojection basé sur GPU. J. Supercalculateur. 2023, 79, 4192-4214. [Référence croisée]
28. Guo, G. ; Guo, L. ; Wang, R. ; Liu, W. ; Li, L. Simulation d'écho de diffusion transitoire et imagerie ISAR pour un océan-cible composite Scène basée sur la méthode TDSBR. Remote Sens.2022 , 14, 1183. [CrossRef]

29. Sun, T.-P. ; Cong, Z. ; Lui, Z. ; Ding, D. Une méthode d'optique physique itérative accélérée dans le domaine temporel pour l'analyse électrique Cibles vastes et complexes. Électronique 2022, 12, 59. [CrossRef]
30. Guo, G. ; Guo, L. ; Wang, R. Algorithme d'image ISAR utilisant l'écho de diffusion dans le domaine temporel simulé par la méthode TDPO. Fil d'antennes IEEE . Propagande. Lett. 2020, 19, 1331-1335. [Référence croisée]
31. Zhou, J. ; Han, Y. Analyse des caractéristiques de diffusion électromagnétique pour les cibles plasma basées sur le tir et le rebond des rayons méthode. AIP Adv. 2019, 9, 065106. [Réf. croisée]
32. Gordon, WB approximations à haute fréquence de l'intégrale de diffusion optique physique. IEEETrans. Propag. d'antennes. 1994, 42, 427-432. [Référence croisée]
33. Ventilateur, TT ; Zhou, X. ; Yu, WM ; Zhou, XY; Cui, TJ Représentations intégrales de ligne dans le domaine temporel des champs dispersés d'optique physique. IEEETrans. Propag. d'antennes. 2017, 65, 309-318. [Référence croisée]
34. Liao, C. Recherche sur la modélisation de diffusion en champ proche d'un navire à la surface de la mer basée sur la méthode à haute fréquence (en chinois). Thèse de maîtrise, Université des sciences et technologies électroniques de Chine, Chengdu, Chine, 2021. Disponible en ligne : https://kns.cnki.net/kcms2/article/abstract?v=n6BwBobH4uvU7PG733EVJdhS9-f9LApXUEAHzK60Kgv6ciotFWwf1 1njOZZqPyjAOLTnfvU-beMgAqMR8blHbC9mFOJ0F5tkD-vf1xHSqT6eY_XonBN7ouPAQRKMghtFziV-7qPBjXZhfcEiaH_ takq8r-_JoHJuM61BQbTbKyScQeIPY8_hM3AAmLBaJTfo9XIwNvOUOSI=&uni (consulté le 29 juillet 2024).
35. Stratton, JA; Chu, L. Théorie de la diffraction des ondes électromagnétiques. Phys. Rév.1939 , 56, 99. [CrossRef]
36. Chu, LJ; Stratton, JA Fonctions d'onde elliptique et sphéroïdale. J. Math. Phys. 1941, 20, 259-309. [Référence croisée]
37. Lui, X.-Y. ; Wang, X.-B. ; Zhou, X. ; Zhao, B. ; Cui, T.-J. Simulation rapide d'images ISAR de cibles à des angles d'aspect arbitraires à l'aide d'un nouveau Méthode SBR. Programme. Électromagn. Rés. B 2011, 28, 129-142. [Référence croisée]
38. Guo, G. ; Guo, L. ; Wang, R. L'étude sur la diffusion en champ proche d'une cible sous irradiation d'antenne par la méthode TDSBR. IEEE Accès 2019, 7, 113476-113487. [Référence croisée]
39. Tang, X. ; Feng, Y. ; Gong, X. Mo Algorithme M-PO/SBR basé sur une plateforme collaborative et un modèle mixte. Trans. Université de Nankin. Aéronaute. Astronaute. 2019, 36, 589-598. [Référence croisée]
40. Li, J. ; Meng, W. ; Chai, S. ; Guo, L. ; Xi, Y. ; Wen, S. ; Li, K. Une méthode hybride accélérée pour la diffusion électromagnétique d'un Modèle composite cible-sol et son image SAR Spotlight. Remote Sens.2022 , 14, 6632. [CrossRef]
41. Zhu, R. Accélération des simulations micromagnétiques avec C++ AMP sur les unités de traitement graphique. Calculer. Sci. Ing. 2016, 18, 53-59. [Référence croisée]
42. Sihai, W. ; Hu, Z. ; Haotian, P. ; Lu, C. Parallélisme accéléré dans la simulation numérique avec C++ AMP. Dans les actes de l'édition 2016 Atelier : Atelier sur le calcul haute performance, Pékin, Chine, 14-16 novembre 2016 ; pp. 53-55.
43. Wynters, E. Traitement parallèle rapide et facile sur GPU utilisant C++ AMP. J. Informatique. Sci. Coll. 2016, 31, 27-33.
44. Shyamala, K. ; Kiran, KR ; Rajeshwari, D. Conception et implémentation d'une multiplication de chaînes matricielles basée sur GPU à l'aide de C++AMP. Dans Actes de la deuxième Conférence internationale 2017 sur les technologies électriques, informatiques et de communication (ICECCT), Coimbatore, Inde, 22-24 février 2017 ; p. 1 à 6.
45. Damkjær, J. Détection de collision BVH sans pile pour la simulation physique ; Université de Copenhague Universitetsparken : København, Danemark, 2007 ; Disponible en ligne : http://image.diku.dk/projects/media/jesper.damkjaer.07.pdf (consulté le 29 juillet 2024).
46. Chung, S. ; Choi, M. ; Vous, D. ; Kim, S. Comparaison de BVH et KD-tree pour l'accélération GPGPU sur de vrais appareils mobiles. Dans Proceedings of the Frontier Computing : Theory, Technologies and Applications (FC 2018), Kyushu, Japon, 9-12 juillet 2019 ; pp. 535-540.
47. Sopin, D. ; Bogolepov, D. ; Ulyanov, D. Construction SAH BVH en temps réel pour les scènes dynamiques de lancer de rayons. Dans les actes du Графикон'2011, Moscou, Russie, 26-30 septembre 2011 ; pp. 74-77.
48. Huipeng, Z. ; Junling, W. ; Di, X. ; Xiaoyang, Q. L'algorithme de rétroprojection modifié pour l'imagerie ISAR bistatique des objets spatiaux. Dans les actes de la conférence internationale IEEE 2016 sur le traitement du signal, les communications et l'informatique (ICSPCC), Hong Kong, Chine, 5-8 août 2016 ; p. 1 à 5.
49. Xiao, D. Étude sur l'imagerie ISAR de cibles spatiales utilisant la technologie BP ; Université de Nanjing : Nanjing, Chine, 2017.
50. Pu, L. ; Zhang, X. ; Ouais.; Wei, S. Un algorithme d'imagerie par rétroprojection tridimensionnel rapide dans le domaine fréquentiel basé sur GPU. Dans les actes de la conférence IEEE Radar 2018 (RadarConf18), Oklahoma City, OK, États-Unis, 23-27 avril 2018 ; pages 1173 à 1177.
51. Li, Z. ; Qiu, X. ; Yang, J. ; Meng, D. ; Huang, L. ; Song, S. Un algorithme BP efficace basé sur TSU-ICSI combiné avec GPU Parallel L'informatique. Remote Sens.2023 , 15, 5529. [CrossRef]
52. Afif, M. ; Dit, Y. ; Atri, M. Accélération des algorithmes de vision par ordinateur à l'aide des processeurs graphiques NVIDIA CUDA. Groupe. Calculer. 2020, 23, 3335-3347. [Référence croisée]
53. Ufimtsev, PY Ondes de bord élémentaires et théorie physique de la diffraction. Électromagnétique 1991, 11, 125-160. [Référence croisée]

électronique

MDPI

Article

Libérer la valeur commerciale : intégrer l'IA Prise de décision dans les systèmes d'information financière

Alin Emanuel Artène 1,*, Aura Emanuela Domil 2 et Larissa Ivascu 1

1 Département de gestion, Faculté de gestion de la production et des transports, Université Politehnica de Timisoara, 14 rue Remus, 300009 Timisoara, Roumanie ; larisa.ivascu@upt.ro
2 Département de comptabilité et d'audit, Faculté d'économie et d'administration des affaires, Université Ouest de Timisoara, 300223 Timisoara, Roumanie ; aura.domil@e-uvt.ro
* Correspondance : alin.artene@upt.ro

Résumé : Cet article de recherche étudie les synergies entre l'intelligence artificielle (IA), la transformation numérique (DT) et les systèmes d'information financière dans le contexte commercial. Le thème central explore la manière dont les organisations améliorent leurs processus décisionnels en intégrant les technologies d'IA dans les initiatives de transformation numérique, en particulier dans le reporting financier. L'objectif principal est de comprendre comment la synergie de ces systèmes intégrés peut générer une valeur commerciale substantielle, stimuler l'innovation stratégique et élever l'analyse financière globale grâce à l'adoption de méthodologies de prise de décision intelligentes et basées sur les données. En exploitant les capacités avancées d'analyse, d'automatisation et d'aide à la décision adaptative, les organisations naviguent dans les complexités d'un environnement commercial en évolution rapide, dans lequel les réseaux de neurones apparaissent comme un outil précieux pour calibrer les résultats dans l' environnement de comptabilité financière, démontrant leur efficacité dans le traitement de données financières complexes. identifier des modèles et faire des prédictions, ouvrant la voie à une nouvelle ère de possibilités de trar L'introduction d'une matrice de gains de théorie des jeux dans cet outil d'aide à la décision d'IA ajoute un cadre stratégique pour analyser les interactions entre les décideurs, en considérant les choix stratégiques et les résultats dans un contexte dynamique et compétitif.

Mots-clés : transformation numérique ; systèmes de prise de décision; systèmes intégrés; les réseaux de neurones; rapports financiers d'entreprise; matrice de gains de la théorie des jeux

Citation : Artène, AE ; Domil, AE; Ivascu, L. Libérer la valeur commerciale : intégrer la prise de décision basée sur l'IA dans les systèmes d'information financière. Électronique 2024, 13, 3069. https://doi.org/10.3390/electronique13153069

Rédacteur académique : Domenico Ursino

Reçu : 12 juillet 2024
Révisé : 24 juillet 2024
Accepté : 30 juillet 2024
Publié : 2 août 2024

1. Introduction

L'automatisation, l'intelligence artificielle et l'analyse des données entraînent des changements importants dans le fonctionnement des institutions financières. Dans le paysage en constante évolution de l'information financière, les organisations sont confrontées à un impératif urgent : exploiter le potentiel de transformation des technologies numériques. L'impact de la numérisation sur la finance est indéniable. Le changement de paradigme vers la transformation numérique représente un point critique où les entreprises doivent non seulement s'adapter mais aussi innover pour rester compétitives sur le marché mondial. L'intégration des technologies d'intelligence artificielle (IA), qui offrent des opportunités sans précédent pour révolutionner la gestion stratégique au sein des systèmes d'information financière économique, est au cœur des objectifs de cette évolution. Le processus de contrôle, dans ce paradigme, évolue d' une surveillance rigide à une orchestration dynamique.

Les systèmes d'IA embarqués équipés d'algorithmes sophistiqués démontrent leur capacité à apprendre en permanence à partir des flux de données, à s'adapter à l'évolution des modèles et à prendre des décisions en temps réel. Cela représente un passage d'un modèle de contrôle déterministe à un cadre plus adaptatif et probabiliste où le contrôle est exercé par le biais d'une gouvernance algorithmique et de boucles de rétroaction continues.

À l'ère de la transformation numérique post-pandémique, l'intégration des technologies d'intelligence artificielle est une pierre angulaire pour les entités économiques publiques et privées qui aspirent à naviguer dans le paysage complexe de l'information financière et de la productivité économique avec

précision et agilité. La convergence de l'intelligence artificielle et de la numérisation des systèmes d'information financière [1] ne représente pas seulement un changement de paradigme technologique mais introduit également une profonde redéfinition des mécanismes de contrôle et des processus de gestion. Les systèmes d'IA intégrés remodèlent les cadres de contrôle et responsabilisent les décideurs dans le contexte dynamique de l'information financière dans un contexte de transformation numérique.

Alors que les entreprises se lancent dans leur transformation numérique et leur numérisation [2], l' injection stratégique et subventionnée des technologies d'IA promet de libérer une valeur commerciale sans précédent. Les systèmes d'information financière traditionnels, ancrés dans des méthodologies déterministes, sont désormais à la croisée des chemins alors que les organisations cherchent à intégrer des algorithmes avancés d'intelligence artificielle. L'intégration de l'apprentissage automatique, de l'analyse prédictive et de l'informatique cognitive dans ces systèmes introduit une nouvelle dimension d'autonomie et d'adaptabilité. À mesure que l'IA devient un participant actif dans les processus décisionnels, les contours des mécanismes de contrôle doivent être réévalués.

Les modèles hiérarchiques traditionnels sont remis en question par la nature distribuée de la prise de décision dans les systèmes d'IA embarqués [3], où les algorithmes apprennent, s'adaptent et contribuent de manière autonome . La nécessité de comprendre les complexités et les implications de cette intégration est devenue primordiale, notamment en termes de son impact sur la prise de décision en matière d'information financière. Les systèmes de reporting traditionnels, bien qu'ils fassent partie intégrante du fonctionnement organisationnel, sont désormais confrontés à la nécessité d'évoluer. La question suivante se pose : comment une intégration judicieuse des technologies d'IA peut-elle améliorer les processus décisionnels dans le contexte du reporting financier lors du processus de transformation numérique ?

La figure 1 montre la prise de décision basée sur l'IA dans le cadre de la transformation numérique. Cela inclut les initiatives de transformation numérique, l'intégration de l'IA, l'amélioration de la prise de décision, les avantages, les cas d'utilisation, la prise en compte, les tendances futures, les stratégies de mise en œuvre et les indicateurs de mesure . Chacune de ces directions est importante dans l'évolution de la numérisation. En évaluant chaque direction,

on peut affirmer ce qui suit : • Initiatives de transformation numérique : celles-ci représentent des étapes qui contribuent à la DT. Les composants qui contribuent au succès du processus DT, les éléments stratégiques ainsi que les opportunités et les défis peuvent être inclus.

• Intégration de l'IA : cela implique de définir le rôle important de l'IA dans la DT, les différents types de technologies associées et la mise en œuvre d'opportunités et de solutions. Tout cela contribue à un soutien important dans le processus DT. • Amélioration de la prise de décision : cela

fait référence à l'utilisation des données pour obtenir des avantages concurrentiels, des analyses prédictives et l'automatisation des processus organisationnels, et rationaliser les processus grâce à une prise de décision efficace.

• Avantages : les avantages identifiés par les bénéficiaires sont multiples et parmi eux figurent l'amélioration de la précision, de la rentabilité et du temps, l'amélioration de l' expérience client et d'autres avantages connexes.

• Exemple : en évaluant les éléments spécifiques dans lesquels l'IA peut être utilisée, nous pouvons mentionner le reporting financier, la gestion des opérations, la relation client avec les applications de l'IA, et bien

d'autres. • Considération : les éléments importants qui doivent être évalués dans cette approche sont les implications éthiques, sécurité et collaboration homme-IA.

• Tendances futures : ce domaine est en pleine croissance et les approches futures doivent être anticipées au niveau organisationnel à travers des stratégies et des approches réfléchies (tendances émergentes, évolution du domaine et approches technologiques). • Stratégies de mise en œuvre : les éléments

de gestion stratégique sont opportuns au niveau organisationnel pour un bon alignement avec l'avancée technologique. • Paramètres de mesure : toute cette approche doit être mesurée, et les

indicateurs de performance et de succès pour la mise en œuvre de l'IA incluent des approches organisationnelles correctes.

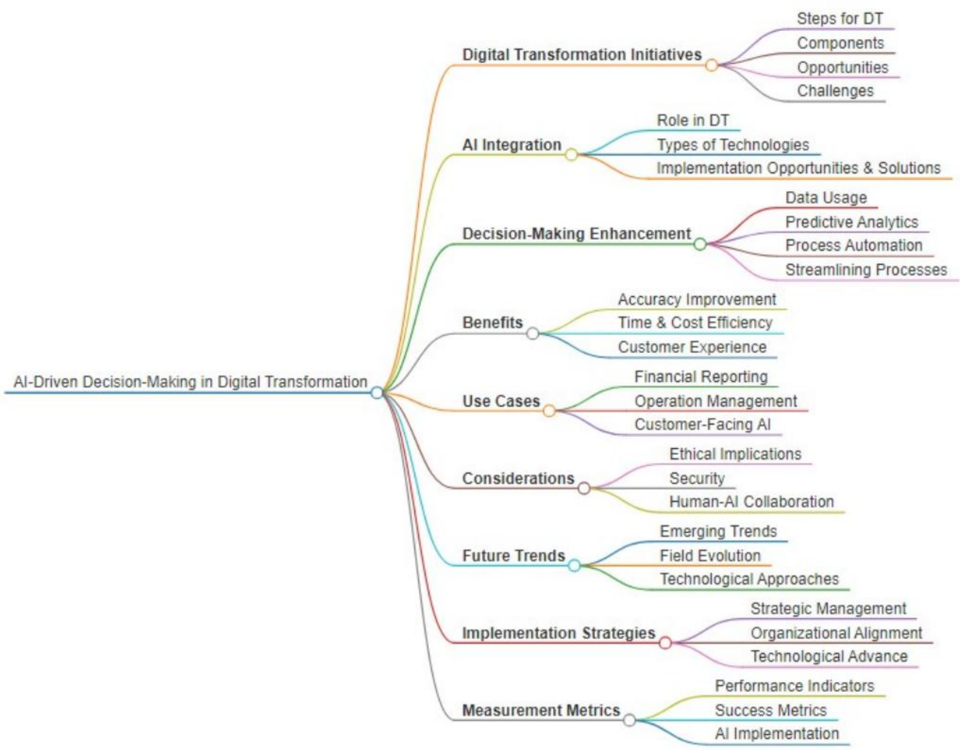


Figure 1. Carte mentale de prise de décision basée sur l'IA.

Figure 1. Carte mentale de prise de décision basée sur l'IA.

Dans un paysage où les informations basées sur les données sont le moteur du succès organisationnel, le potentiel les avantages potentiels de l'IA (de la rationalisation des processus à la découverte d'informations exploitables) invitent les organisations à explorer de nouvelles frontières en matière de capacités de prise de décision. Face à cela, les contexte , cet article entreprend une exploration du rôle transformateur que jouent les technologies de l'IA. la refonte du paysage décisionnel en matière d'information financière à l'ère de transformation numérique. Plonger dans l'interaction complexe entre l'intégration de l'IA et évolutive des systèmes d'information financière, nous cherchons à fournir des informations nuancées qui nuancées et liées avec des implications pratiques. Notre objectif est de guider les organisations compréhension globale de la façon dont l'IA, lorsqu'elle est parfaitement intégrée, peut non seulement mais dépasse les besoins de prise de décision dans le domaine dynamique de l'information financière dans le le portage à l'ère du numérique.

Dans un paysage où les informations basées sur les données sont le moteur de la réussite organisationnelle, Les avantages de l'IA (de la rationalisation des processus à la découverte de renseignements exploitables) invitent organisations doivent explorer de nouvelles frontières en matière de capacités de prise de décision. Dans ce Sur le terrain, cet article entreprend une exploration du rôle transformateur que jouent les technologies de l'IA dans Les stratégies contribuent à remodeler le paysage décisionnel en matière d'information financière à l'ère de la de la transformation numérique. En nous plongeant dans l'interaction complexe entre l'intégration de l'IA et la nature et la nature évolutive des systèmes d'information financière, nous cherchons à fournir des avancées théoriques des vues qui relient les avancées théoriques aux implications pratiques. Notre objectif est de guider vers une les organisations à une compréhension globale de la façon dont l'IA, lorsqu'elle est parfaitement intégrée, répond, peut non seulement répondre, mais aussi dépasser les besoins de prise de décision dans le domaine dynamique de l'ère de la re-numérisation financière.

2. Revue de la littérature

2. Revue de la littérature

L'intégration de l'intelligence artificielle (IA) dans les systèmes d'information financière a gagné attention particulière dans la littérature récente [4] en raison de son potentiel à transformer les systèmes traditionnels. processus financiers et à libérer de la valeur commerciale. Les auteurs de [5] étudient l'impact l'intelligence artificielle dans les systèmes d'information comptable pour améliorer non financière des entreprises manufacturières, révélant des résultats positifs significatifs efficacité des systèmes d'information comptable et sur l'ensemble des performances non financières. nombreuses études ont souligné le rôle de l'IA dans l'amélioration des finances décisionnels. Réf. [1] a démontré l'efficacité de l'apprentissage automatique les tendances financières, optimiser l'allocation des ressources et fournir informations précieuses pour la planification stratégique. L'intégration des réseaux de neurones, comme discuté dans [6], s'est révélée prometteuse dans l'automatisation des tâches, l'amélioration de la précision des prévisions et l'assistance aux décideurs dans le domaine financier.

L'intégration de l'intelligence artificielle (IA) dans les systèmes d'information financière fait l'objet d'une a attiré une attention considérable dans la littérature récente [4] en raison de son potentiel à transformer les processus financiers traditionnels et libérer de la valeur commerciale. Les auteurs de [5] étudient l'intégration de l'impact de l'intégration de l'intelligence artificielle dans les systèmes d'information comptables sur la performance améliorant la performance non financière des entreprises manufacturières, révélant des effets significatifs à la fois sur effets positifs tant sur l'efficacité des systèmes d'information comptable que sur la non- performance globale. De performance financière. De nombreuses études ont souligné le rôle de l'IA dans l'amélioration des processus processus de prise de décision financière. Réf. [1] ont démontré l'efficacité des algorithmes automatiques pour prédire des algorithmes d'apprentissage pour prédire les tendances financières, optimiser l'allocation des ressources et des fournir des informations précieuses pour la planification stratégique. L'intégration des réseaux de neurones, comme discuté par [6], s'est montré prometteur dans l'automatisation des tâches, l'amélioration de la précision des prévisions et accompagner les décideurs dans le domaine financier.

Le rôle du Big Data et l'exploitation de la prise de décision basée sur l'IA et de la transformation numérique Le les systèmes d'information financière libèrent de la valeur commerciale, en s'alignant sur les préférences des parties prenantes dans le système d'information financière en pleine évolution [5]. s'alignent sur les parties préférences, et relever les défis et les opportunités dans le paysage financier en évolution. Dans le contexte de paysage [5]. Dans le contexte de l'économie moderne, l'impact transformateur du Big Data vers la compréhension

comment l'utilisation des données par d'autres influence les résultats du marché [5]. Cette exploration implique d'intégrer le big data dans les théories économiques et financières contemporaines. Une application consiste notamment à exploiter le Big Data pour améliorer la prise de décision des acteurs des marchés financiers, en influençant les prix des entreprises, le coût du capital et la dynamique des investissements.

Le concept de transformation numérique dans l'information financière a été largement exploré dans la littérature. Des chercheurs tels que les auteurs de [7,8] ont discuté de l'évolution des systèmes d'information financière depuis des processus manuels vers des plateformes numériques avancées. L'évolution vers des solutions basées sur le cloud, l'automatisation des tâches de saisie de données et l'utilisation d' outils d'analyse avancés ont été identifiées comme des éléments clés de la transformation numérique dans le reporting financier [9].

Les recherches de Sorensen [10] étudient l'importance d'aligner les décisions financières sur les préférences des parties prenantes. Réf. [10] souligne la nécessité d'outils qui prennent en compte les points de vue de diverses parties prenantes, telles que la direction de l'entreprise et les actionnaires.
L'intégration d'outils de prise de décision basés sur l'IA, informés par les réseaux de neurones, offre un mécanisme permettant d'aligner les choix stratégiques sur les préférences des principales parties prenantes.

La littérature reconnaît également les défis associés à l'intégration de l'IA et de la transformation numérique dans l'information financière. Les préoccupations liées à la confidentialité des données, à la conformité réglementaire et au besoin de professionnels qualifiés capables de naviguer à l'intersection de la finance et de l'IA ont été discutées [3,11]. Cependant, les études mettent également en évidence les opportunités d'économies de coûts, d'amélioration de l'efficacité et d'avantages stratégiques qui découlent d' une mise en œuvre réussie [12]. De nombreuses études soulignent l'importance des KPI financiers dans la mesure de la performance et la prise de décision stratégique au sein des organisations [13]
Des indicateurs tels que le retour sur investissement (ROI), le bénéfice par action (BPA) et la marge bénéficiaire sont reconnus comme des indicateurs essentiels de la santé financière. Les chercheurs soutiennent que l'intégration de ces KPI dans les réseaux neuronaux peut améliorer la précision des prévisions financières et des systèmes d'aide à la décision [14,15].

L'examen de l'intersection de la théorie des jeux et de la prise de décision financière [16] explore comment les interactions stratégiques entre les acteurs du marché, influencées par l'asymétrie de l'information et la concurrence, ont un impact sur les résultats financiers. L'étude ne se concentre peut-être pas directement sur les KPI, mais fournit un aperçu de la dynamique décisionnelle. Même s'il n'existe peut-être pas une abondante littérature combinant explicitement les KPI financiers, la prise de décision et la théorie des jeux, certaines études fournissent une base pour comprendre l'interdépendance de ces concepts. Aborder la transparence et l'interprétabilité des modèles de décision d'IA, Réf. [17] discute de l'importance de l'explicabilité pour gagner la confiance et l'acceptation des outils d'IA dans la prise de décision, en particulier dans des domaines sensibles comme la santé et la finance.

L'analyse documentaire indique un nombre croissant de recherches mettant l'accent sur les diverses applications, défis et considérations éthiques associés à l'intégration des outils d'IA dans la prise de décision. Comprendre l'impact de l'IA dans divers domaines constitue une base pour les développements futurs visant à créer un système d'aide à la décision plus efficace, transparent et éthique .

Ces dernières années, les sociétés de comptabilité et d'audit ont exploité le potentiel de l'IA pour révolutionner les processus traditionnels au sein des institutions financières et ont développé des applications d'apprentissage automatique dans le domaine financier. JP Morgan Chase, un leader mondial des services financiers, a développé la plateforme Contract Intelligence, ou COiN, un outil basé sur l'IA conçu pour examiner les documents juridiques et extraire des points de données et des clauses critiques, réduisant ainsi jusqu'à 360 000 heures par an. consommant d'importantes ressources humaines et améliorant la précision et l'évolutivité des opérations, établissant ainsi une nouvelle norme d'efficacité en matière d'information financière. Un autre exemple convaincant vient de Deloitte, une référence en matière de services d'audit et d'assurance qui a utilisé ACL Analytics, un logiciel sophistiqué d'analyse de données, pour améliorer ses processus d'audit. Cet outil exploite la puissance de l'IA et de l'apprentissage automatique pour passer au crible de vastes volumes de données financières, identifiant les anomalies, les tendances et les modèles qui justifient un examen plus approfondi. En intégrant des réseaux de neurones, les auditeurs peuvent concentrer leurs efforts sur les domaines à haut risque, améliorant ainsi la qualité et

fournir aux clients de précieuses recommandations basées sur des données, favorisant une prise de décision et une planification stratégique plus éclairées.

3. Matériels et méthodes

Ce document de recherche utilise une méthodologie de recherche fondamentale complète pour libérer de la valeur commerciale en intégrant la prise de décision basée sur l'IA à la transformation numérique dans les systèmes d'information financière. Le processus logique décrit dans cette méthodologie sert de feuille de route étape par étape, guidant l'enquête sur l'intersection entre l'intelligence artificielle (IA) et la transformation numérique dans l'économie dans le contexte de la numérisation de l'information financière. L'objectif principal est de développer un cadre théorique qui améliore les processus décisionnels en assimilant les technologies avancées d'intelligence artificielle dans les systèmes d'information financière existants. Au cours de cette recherche, nous sommes passés par des étapes systématiques, en commençant par l'identification du problème de recherche, suivie d'une revue approfondie de la littérature pour recueillir des informations pertinentes sur l'intelligence artificielle , la transformation numérique et les systèmes d'information financière actuels. Les fondements théoriques ont ensuite été cartographiés en synthétisant les connaissances existantes et en identifiant les lacunes dans la compréhension actuelle. Par la suite, les auteurs ont formulé un plan de recherche pour guider l'enquête empirique.

L'intégration de la prise de décision basée sur l'IA et de la transformation numérique dans les systèmes d'information financière nécessite un examen attentif des dimensions technologiques, organisationnelles et éthiques . La méthodologie aborde ces questions en intégrant une approche multidisciplinaire, garantissant une compréhension globale des défis et des opportunités associés à la mise en œuvre de l'intelligence artificielle dans l'information financière.

En parcourant ce processus méthodique, la recherche vise à fournir des informations précieuses aux universités, aux entreprises et à l'industrie. Le résultat attendu est un cadre théorique solide qui non seulement clarifie les synergies entre l'intelligence artificielle et la transformation numérique, mais fournit également des conseils pratiques aux organisations cherchant à améliorer leurs capacités décisionnelles en matière d'information financière.

Cette recherche vise à combler le fossé entre les avancées théoriques et les implications pratiques, libérant ainsi la valeur commerciale inexploitée inhérente à l'intégration de la prise de décision basée sur l'IA avec la transformation numérique dans les systèmes d'information financière.

4. Outils de transformation numérique pour le processus décisionnel

Les outils de transformation numérique pour l'information financière peuvent soutenir de manière significative le processus de prise de décision au sein des organisations, en développant des stratégies résilientes aux différentes conditions économiques. Ils jouent un rôle crucial dans le soutien aux processus décisionnels en matière de reporting financier en fournissant des données en temps réel, des analyses avancées, des fonctionnalités d'automatisation et de collaboration. Ces outils contribuent à une prise de décision plus éclairée, stratégique et basée sur les données au sein des organisations. L'intégration de technologies avancées et d'outils numériques dans les systèmes d'information financière offre plusieurs avantages qui améliorent les capacités de prise de décision. Les décideurs peuvent accéder à des informations à jour sur les mesures financières clés, les indicateurs de performance et les tendances du marché, permettant une prise de décision plus éclairée et plus rapide [17,18].

Les outils avancés d'analyse et de visualisation des données aident à transformer des données financières complexes en représentations visuelles facilement compréhensibles. Cela aide les décideurs à identifier les modèles, les tendances et les anomalies, facilitant ainsi la prise de décision basée sur les données. Les outils numériques intègrent souvent des capacités de prévision et d'analyse prédictive. En analysant les données historiques et en identifiant des modèles, ces outils peuvent aider les décideurs à faire des prévisions plus précises sur les tendances et les résultats financiers futurs. L'intégration d'algorithmes d'intelligence artificielle et d'apprentissage automatique (ML) améliore les capacités des systèmes d'information financière . Ces technologies peuvent fournir des informations intelligentes, identifier les opportunités et soutenir la prise de décision en analysant de grands ensembles de données et en facilitant la collaboration et la communication entre les différentes parties prenantes impliquées dans le processus déci

Les décideurs peuvent évaluer les implications financières de diverses décisions, en garantissant l'alignement avec les exigences réglementaires et les stratégies d'atténuation des risques et en permettant aux décideurs d'évaluer les implications financières de diverses décisions, en garantissant

nario planification en permettant aux décideurs de modéliser différents scénarios financiers et leur alignement avec les exigences réglementaires et les stratégies d'atténuation des risques et en permettant d'évaluer leur impact potentiel. Analyse des interactions stratégiques entre plusieurs planifications décisionnelles en permettant aux décideurs de modéliser différents scénarios financiers

Les décideurs, souvent dans des situations de concurrence, peuvent être employés pour recruter les managers et évaluer leur impact potentiel. Analyse des interactions stratégiques entre plusieurs décideurs dans le processus de prise de décision. Dans des contextes commerciaux, il peut être utilisé pour modéliser et analyser des scénarios de prise de décision impliquant plusieurs parties prenantes, utilisé pour accompagner les managers et analyser des scénarios de prise de décision impliquant plusieurs parties prenantes,

telles que des concurrents, des détaillants et dans le processus de prise de décision. Dans des contextes commerciaux, il peut être

les fournisseurs et les clients [19,20]. risque de décision impliquant plusieurs parties prenantes, telles que des concurrents,

La théorie des jeux fournisseurs et les clients [19,20].

analysant divers ensembles de données et identifiant des tendances et des informations stratégiques pour l'évaluation des risques en

aider à la prise de décision en tenant compte des réponses probables des concurrents et collaborer à la prise de décision sur

laboratoires, conduisant à des stratégies plus éclairées. Les réseaux de neurones, un sous-ensemble de machines menant à des

l'apprentissage et l'IA, sont adeptes de la reconnaissance et de la prédiction des formes. Dans la prise de décision, la neu- IA est

Les réseaux réels peuvent analyser de grands ensembles de données, identifier des modèles et faire des prédictions basées sur

informations historiques et en temps réel. Cette fonctionnalité est précieuse pour les gestionnaires et partage des informations en

titulaires lors de l'évaluation des différentes options de décision et des risques associés. évaluer diverses options

Cette carte conceptuelle (Figure 2), fournit une représentation visuelle de l'interconnexion. Cette carte

concepts liés, illustrant comment la théorie des jeux et les réseaux de neurones, lorsqu'ils sont intégrés dans des concepts, illustrant

systèmes d'aide à la décision, contribuer aux processus de prise de décision stratégique et améliorer les systèmes de soutien,

qualité globale des décisions pour les parties prenantes [21]. La combinaison de la théorie des jeux et de la qualité des

les réseaux peuvent être particulièrement puissants. Les réseaux de neurones peuvent analyser les données pour éclairer les

les décideurs sur les résultats probables de différentes stratégies, tandis que la théorie des jeux peut aider à déterminer les résultats

modéliser les interactions stratégiques et la dynamique concurrentielle entre les parties prenantes. Par levier- interactions

réseaux de neurones vieillissants pour analyser les données historiques et prédire les résultats, réseaux de décideurs pour analyser

peuvent utiliser la théorie des jeux pour optimiser leurs stratégies sur la base d'une compréhension plus approfondie de la théorie

paysage concurrentiel. L'IA, y compris la théorie des jeux et les réseaux neuronaux, peut être intégrée au paysage. L'IA, y compris

exploitables aux dirigeants et aux systèmes d'aide aux actions qui fournissent des informations exploitables aux dirigeants et aux

titulaires.

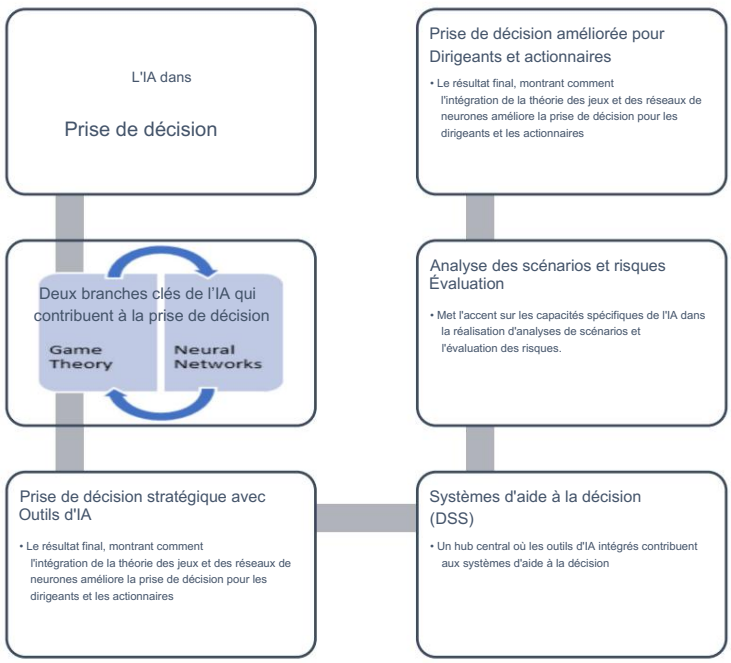


Figure 2. Carte conceptuelle de l'IA intégrée aux systèmes d'aide à la décision.

Figure 2. Carte conceptuelle de l'IA intégrée aux systèmes d'aide à la décision.

Ces systèmes exploitent des analyses avancées pour améliorer les processus de prise de décision.

Les gestionnaires peuvent utiliser des outils d'IA dans le processus de prise de décision pour effectuer une analyse de scénarios,

évaluer les résultats potentiels de différentes décisions dans diverses conditions. Cela facilite la prise de décisions plus solides et adaptables à différentes circonstances.

4.1. Développer un réseau neuronal pour intégrer l'IA dans les processus décisionnels

Développer un réseau neuronal pour intégrer l'IA dans l'information financière implique la conception d' un modèle capable d'analyser et d'interpréter les données financières des bilans. Le choix d' un réseau neuronal à trois couches pour les tâches de comptabilité et de reporting financier n'est pas une règle inhérente, mais plutôt une architecture couramment utilisée qui a fait ses preuves dans diverses applications. Une architecture à trois couches est relativement simple et plus facile à interpréter lorsque l'on la compare à des homologues plus profondes. Cette architecture comporte moins de paramètres et de couches, ce qui la rend apparemment plus gérable. Cependant, cela ne doit pas occulter le fait que même un réseau neuronal à trois couches peut fonctionner comme une boîte noire complexe. Les transformations non linéaires inhérentes et les interactions entre les couches rendent difficile le déchiffrement de la façon dont les entrées spécifiques sont traduites en sorties. Dans le domaine de l'information financière, où la transparence et la responsabilité sont primordiales, la nature opaque des réseaux de neurones, quelle que soit leur profondeur, peut entraîner d'importants problèmes d'interprétabilité. Parce que l'interprétabilité est souvent cruciale et que les parties prenantes doivent comprendre et faire confiance aux résultats du modèle, cette complexité n'est pas seulement une préoccupation théorique mais aussi une préoccupation pratique qui a un impact sur la confiance et la fiabilité dans les applications du monde réel.

Les réseaux neuronaux à trois couches sont particulièrement puissants en raison de leur capacité à se rapprocher d'un large éventail de fonctions. Cette capacité trouve son origine dans le théorème de superposition de Kolmogorov, qui affirme qu'un réseau neuronal à trois couches peut représenter n'importe quelle fonction multivariée, qu'elle soit continue ou discontinue. Cela rend ces réseaux polyvalents pour diverses applications, y compris les systèmes d'information financière où la capture de relations complexes et non linéaires au sein des données est cruciale [22,23]. L'un des principaux avantages d'un réseau neuronal à trois couches est sa relative simplicité et interprétabilité par rapport aux réseaux plus profonds. Bien qu'il puisse être difficile d'interpréter le processus décisionnel des réseaux neuronaux très profonds, les modèles à trois couches trouvent un équilibre en étant suffisamment complexes pour capturer des modèles complexes tout en restant plus simples à analyser et à déboguer. Cette interprétabilité est particulièrement importante dans les contextes financiers où la transparence et la responsabilité sont primordiales [24,25].

Les modèles plus simples sont également moins sujets au surajustement et peuvent être plus robustes, en particulier lorsqu'ils traitent des données limitées. Pour de nombreuses tâches de comptabilité et de reporting financier, un nombre modéré de couches cachées peuvent capturer efficacement les modèles et les relations sous-jacentes dans les données.

Nous avons conçu le modèle mathématique d'un réseau neuronal avec 31 unités d'entrée, 16 unités cachées et 2 unités de sortie comme suit :

x1, x2, ..., x31 comme caractéristiques d'entrée

h1, h2, ..., h16 pour être les unités de couche cachées

y1, y2 comme unités de couche de sortie

Le modèle mathématique du réseau de neurones peut être exprimé comme suit :

Calcul des couches cachées :

$$hj = \text{ReLU} \sum i = 131wij(1)xi + bj(1) \text{ , pour } j = 1, 2, \ldots, 16$$

où ReLU(a) = max(0, a) est la fonction d'activation de l'unité linéaire rectifiée.

Calcul de la couche de sortie :

$$yk = \text{Softmax} \sum j = 116wjk(2)hj + bk(2) \text{ , pour } k = 1, 2$$

étaient

$$\text{Softmax}(a)i = \frac{eai}{\sum j = \text{est la fonction d'activation SoftMax.} \sum 1eai}$$

Softmax ( ) = ∑ = ∑ est la fonction d'activation SoftMax.

Les paramètres du modèle (poids et biais) sont appris pendant la formation processus pour minimiser une fonction de perte spécifique, généralement associée à la tâche à accomplir (par exemple, classification ou régression). La formation consiste à ajuster les poids et les biais en utilisant des algorithmes d'optimisation tels que la descente de gradient stochastique (SGD).

Le modèle est entraîné en effectuant de manière itérative une propagation vers l'avant, en calculant la perte, puis en utilisant la rétropropagation pour mettre à jour les paramètres du modèle. La formation le processus se poursuit jusqu'à ce que le modèle converge ou jusqu'à un nombre prédéterminé d'époques atteint. Une fois formé, le réseau neuronal peut être utilisé pour faire des prédictions sur de nouvelles données en effectuant une propagation vers l'avant avec les paramètres appris. Cette application mathématique donne des L'approche constitue la base de la formation utilisant un réseau neuronal à trois couches pour les rapports financiers. tâches de reporting, connexion de la couche d'entrée avec la structure du bilan à la couche cachée, la couche de sortie, et finalement prédire les KPI financiers dans la couche de sortie. Des ajustements peuvent être effectués au niveau de en fonction des caractéristiques spécifiques de vos données et des exigences de votre tâche.

Dans des contextes financiers où les données peuvent être limitées et les ressources informatiques peuvent être limitées. contrainte, une architecture à trois couches établit un équilibre entre complexité et efficacité . Les réseaux extrêmement profonds peuvent souffrir de problèmes de gradient qui disparaissent ou explosent. lors de la formation, ce qui rend difficile pour le modèle d'apprendre efficacement. Une formation de trois couches est moins sujette à ces problèmes, en particulier lors de l'utilisation de fonctions d'activation telles que ReLU (Unité Linéaire Rectifiée) qui atténuent le problème du gradient de disparition [26].

La couche d'entrée doit être conçue pour capturer les informations pertinentes du bilan feuille. Comme le montre la figure 3. Et expliqué plus en détail dans l'annexe A. Chaque nœud de la couche d'entrée correspond à une caractéristique ou une variable spécifique du bilan, telle que l'actif, le passif, les capitaux propres, les impôts, les amortissements, le revenu net, etc. Y compris les caractéristiques liées à la structure du bilan permet au réseau neuronal d'apprendre des modèles et des relations au sein des données financières. Ces informations sont cruciales pour comprendre la santé financière et la position de l'entité économique. Bien qu'il s'agisse d'un modèle de proposition pour une entité économique, l'efficacité de notre réseau de neurones dépend de la qualité et de la pertinence de la comptabilité fournie. Il convient de s'assurer que les entités sélectionnées dans la couche d'entrée sont significatives et qu'il existe suffisamment de données pour entraîner efficacement le modèle.

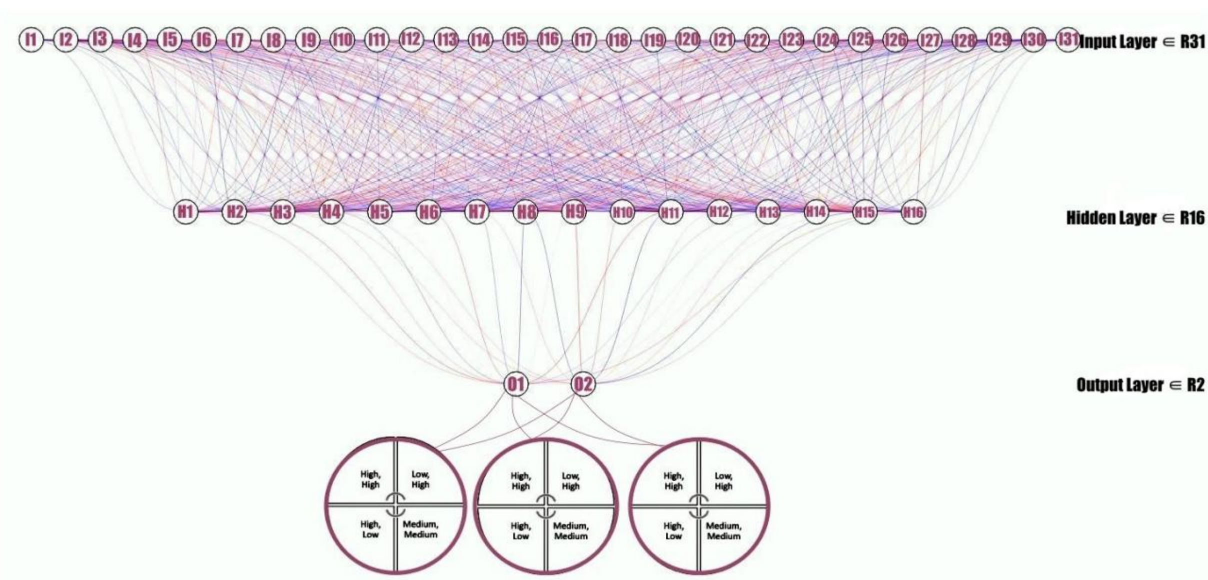


Figure 3. Réseau neuronal à trois couches pour les tâches de comptabilité et de reporting financier.
Figure 3. Réseau neuronal à trois couches pour les tâches de comptabilité et de reporting financier.

La couche cachée est l'endroit où le réseau neuronal apprend des représentations complexes et modèles à partir des entités en entrée. En connectant la couche d'entrée à la couche cachée, le réseau peut extraire des informations et des relations significatives. Si l'objectif de l'entité économique est d'augmenter la valeur de l'entreprise, il est logique de corréler la couche cachée avec les aspects financiers KPI. Les KPI financiers servent souvent d'indicateurs de performance qui ont un impact direct sur l'ensemble de l'activité.

la valeur et le succès d'une entreprise. La couche cachée peut apprendre à faire abstraction et à combiner les fonctionnalités de la couche d'entrée pour identifier les modèles associés à des résultats financiers réussis ou à une valeur commerciale accrue. Le choix des deux niveaux de résultats dépend des objectifs spécifiques et de la manière dont l'entreprise définit et se rapporte à la valeur commerciale, que ce soit par la rentabilité ou la croissance des revenus. Les KPI exacts peuvent varier en fonction du secteur, des objectifs de l'entreprise et de la nature de l'entreprise qui intègre l'outil d'IA de réseau neuronal.
L'un des indicateurs clés de la réussite d'une entreprise est sa capacité à générer des revenus croissants au fil du temps. La première couche de résultats pourrait représenter la croissance prévue des revenus. Le réseau neuronal serait ensuite formé pour apprendre des modèles dans les données d'entrée qui sont en corrélation avec des revenus plus élevés.

Le modèle mathématique aurait deux nœuds de sortie, chacun correspondant à l'un de ces résultats. Pour la classification binaire, l'entreprise peut utiliser une fonction d'activation sigmoïde dans la couche de sortie, et pour les tâches de régression, l'entreprise peut utiliser une activation linéaire [17].

$$(1)\ Y1 = \sigma\ (H \cdot W1\ (2) + b1\ (2))$$

$$(2)\ Y2 = \sigma\ (H \cdot W2\ (2) + b2\ (2))$$

où σ est la fonction d'activation sigmoïde, et Y1 et Y2 représentent respectivement les valeurs prévues pour la croissance des revenus et la rentabilité.

Le troisième nœud de sortie peut être inclus si l'entité économique peut être en mesure de prendre en compte les KPI financiers négatifs qu'elle souhaite minimiser, comme les coûts d'exploitation ou le ratio des dépenses.

$$(3)\ Y3 = \sigma\ (H \cdot W3\ (2) + b3\ (2))$$

4.2. Utiliser les outils d'IA de la théorie des jeux dans le processus de prise de décision

La théorie des jeux est puissante à l'ère de la numérisation [20] et de l'IA car elle propose une approche structurée de la prise de décision dans des environnements complexes, dynamiques et incertains. Il aide les décideurs à gérer les interactions stratégiques, à allouer efficacement les ressources et à s'adapter à l'évolution rapide du paysage technologique et commercial. Comme nous l'avons mentionné précédemment , nous encourageons les entités économiques à appliquer la théorie des jeux à leurs processus commerciaux pour libérer de la valeur commerciale grâce à une prise de décision basée sur l'IA et à la transformation numérique des systèmes d'information financière et à créer une matrice de gains qui représente les interactions entre les différentes décisions. décideurs ou parties prenantes. Dans le contexte de nos recherches, ces décideurs pourraient inclure la direction de l'entreprise, les actionnaires, le système d'IA et d'autres entités pertinentes.

Nous supposons qu'il y a deux décideurs clés : la direction de l'entreprise (CM) et les actionnaires (SH), et qu'il existe deux choix stratégiques pour chacun : mettre en œuvre un reporting financier (AI) basé sur l'IA ou s'en tenir au reporting traditionnel (TR). Les gains sont exprimés en termes de valeur commerciale ou d'utilité pour chaque combinaison de choix.

Explication des gains • Élevé, élevé :

si la direction de l'entreprise et les actionnaires choisissent de mettre en œuvre des rapports financiers basés sur l'IA, ils reçoivent tous deux une valeur commerciale ou une utilité élevée. • Faible, élevé : si la direction de l'entreprise opte pour l'IA alors que les actionnaires s'en tiennent au reporting traditionnel, la direction de l'entreprise pourrait connaître une faible valeur (en raison des coûts de mise en œuvre, par exemple), tandis que les actionnaires recevront une valeur élevée (car ils préféreront peut-être le reporting traditionnel familier). reporting).

• Élevé, faible : si la direction de l'entreprise s'en tient au reporting traditionnel alors que les actionnaires préfèrent les rapports basés sur l'IA, la direction de l'entreprise pourrait atteindre une valeur élevée (car elle évite les coûts de mise en œuvre), mais les actionnaires recevront une faible valeur (car ils souhaitent bénéficier des avantages de l'IA). reporting piloté). • Moyen, Moyen : Si la direction de l'entreprise et les actionnaires s'en tiennent au reporting traditionnel, ils atteignent tous deux un niveau moyen de valeur commerciale ou d'utilité.

Nous pouvons développer les bénéfices en termes d'augmentation de la valeur pour la direction de l'entreprise et les actionnaires sur la base de deux choix stratégiques : mettre en œuvre un reporting financier basé sur l'IA ou s'en tenir au reporting traditionnel. Les valeurs représentent la valeur commerciale ou l'utilité perçue, les valeurs plus élevées indiquant une valeur perçue plus élevée. Élevé, moyen et faible sont des représentations qualitatives de la valeur perçue, et les valeurs numériques réelles dépendront du contexte spécifique et des préférences des parties prenantes.

Les bénéfices sont décrits dans le tableau 1, et les valeurs réelles dépendraient de facteurs tels que le secteur, les objectifs et préférences spécifiques des parties prenantes, ainsi que les avantages et les inconvénients perçus du reporting basé sur l'IA par rapport au reporting traditionnel dans le contexte donné. Cette matrice fournit une manière structurée de comprendre comment les décisions de chaque partie influencent la valeur perçue à la fois par la direction de l'entreprise et par les actionnaires :

Tableau 1. Matrice de gains.

		Actionnaires	
		IA	TR
Entreprise Gestion	IA	Haut, haut	Faible, élevé
	TR	Haut, bas	Moyen, Moyen

• Élevé, Élevé (AI) :

Gestion d'entreprise (CM) : valeur élevée, car les rapports financiers basés sur l'IA devraient améliorer l'efficacité, la précision et la prise de décision stratégique, conduisant à une augmentation des performances globales de l'entreprise.
Actionnaires (SH) : valeur élevée, car ils bénéficient d'une transparence améliorée, d'une prise de décision mieux informée de la part de la direction et d' une augmentation potentielle des bénéfices. • Faible,

Élevé (TR) : Gestion

d'entreprise (CM) : faible valeur, car le fait de s'en tenir aux rapports traditionnels peut entraîner des opportunités manquées de gains d'efficacité, d'informations stratégiques et d'économies de coûts offertes par les rapports basés sur l'IA.
Actionnaires (SH) : valeur élevée, car ils peuvent préférer la familiarité et la stabilité des rapports traditionnels, percevant potentiellement moins de risques ou de perturbations dans leurs investissements.

• Élevé, faible (IA) :

Gestion d'entreprise (CM) : valeur élevée, car la mise en œuvre de rapports basés sur l'IA satisfait l'objectif de la direction d'adopter des technologies innovantes et de rester compétitif.

Actionnaires (SH) : faible valeur, car ils pourraient être déçus par la décision de ne pas s'en tenir au reporting traditionnel, percevant potentiellement des risques ou des incertitudes plus élevés.

• Moyen, Moyen (TR) :

Gestion d'entreprise (CM) : valeur moyenne, car s'en tenir aux rapports traditionnels peut assurer la stabilité, mais peut ne pas tirer parti des avantages potentiels offerts par les rapports basés sur l'IA.

Actionnaires (SH) : valeur moyenne, car ils maintiennent un sentiment de stabilité mais peuvent passer à côté d'améliorations potentielles en matière de prise de décision et d'efficacité.

Les directeurs financiers sont confrontés à la fois à des défis et à des opportunités dans le processus décisionnel en matière d'IA dans un environnement industriel et réglementaire donné. Ceux-ci peuvent varier en fonction des circonstances spécifiques de chaque organisation. Ils doivent se familiariser avec des réglementations complexes sur la confidentialité des données pour garantir que l'utilisation de l'IA est conforme aux lois sur la confidentialité, en particulier dans les secteurs traitant d'informations sensibles sur les clients. Un autre défi auquel la direction sera bientôt confrontée consiste à suivre l'évolution des réglementations liées à l'IA et à garantir que les systèmes d'IA sont conformes aux lois et normes spécifiques au secteur.

Les matrices de la théorie des jeux peuvent être adaptées à divers scénarios de prise de décision, et lorsqu'elles sont appliquées au contexte des décisions des directeurs financiers (CFO) dans le domaine de l'IA, nous proposons quatre matrices qui peuvent être prises en compte dans le processus de prise de décision. : Les directeurs financiers qui prennent des décisions liées à l'IA peuvent utiliser des matrices de théorie des jeux pour naviguer dans quatre scénarios clés : adoption précoce ou tardive de l'IA, fournisseur externe ou développement interne de l'IA, collaboration pour le partage de données ou protection des données, et conformité réglementaire proactive ou réactive. , en équilibrant les avantages, les coûts et les risques pour l'avantage concurrentiel et l'innovation.

4.3. Problèmes d'éthique et de confidentialité dans les rapports financiers basés sur l'IA

Dans cet environnement de reporting financier relativement nouveau basé sur l'IA, les préoccupations en matière de confidentialité des données sont primordiales. Les exigences étendues en matière de données des systèmes d'IA englobent souvent des informations financières et personnelles sensibles, soulevant le spectre d'accès non autorisés et de violations de données [27]. Le fondement même de notre travail repose sur l'intégrité et la confidentialité des données que nous traitons. À mesure que nous intégrons l'IA, nous devons être vigilants dans la protection de ces données grâce à des méthodes de cryptage robustes, garantissant que les informations sensibles restent sécurisées tant en transit qu'au repos. Les parties prenantes doivent être pleinement informées de la manière dont leurs données sont collectées, stockées et utilisées. Cette transparence s'étend à l'obtention du consentement explicite des personnes dont les données sont utilisées. En mettant en œuvre des politiques strictes de gouvernance des données, nous pouvons définir des contrôles d'accès clairs et garantir que les données ne sont accessibles qu'au personnel autorisé dans des circonstances appropriées. La responsabilité et la transparence dans les systèmes d'IA sont essentielles et il est impératif que les développeurs d'applications établissent des cadres de responsabilité clairs qui définissent les responsabilités des utilisateurs, des opérateurs et des déci
Lorsque des erreurs se produisent, il doit y avoir un processus transparent pour identifier et résoudre la source du problème. La mise en œuvre de techniques d'IA explicables peut grandement faciliter ce processus. Ces techniques nous permettent de comprendre et d'articuler comment les systèmes d'IA prennent des décisions, favorisant ainsi la confiance et la responsabilité entre les parties prenantes [28].

Avec l'adoption généralisée des normes comptables internationales, la conformité réglementaire est une préoccupation constante dans notre profession. Les applications d'IA dans le reporting financier doivent respecter une multitude d'exigences réglementaires, depuis les réglementations financières jusqu'aux lois sur la protection des données. S'engager dans un dialogue continu avec les organismes de réglementation peut fournir des conseils précieux et contribuer à garantir que nos systèmes d'IA sont conformes aux lois en vigueur. Des audits de conformité réguliers sont essentiels pour vérifier le respect et combler toute lacune dans nos processus.

Les risques de sécurité sont un aspect inhérent à l'intégration de l'IA dans l'information financière. Les systèmes d'IA peuvent être la cible de cyberattaques, pouvant conduire à des violations de données ou à la manipulation d'informations financières [29]. Pour atténuer ces risques, nous devons mettre en œuvre des mesures avancées de cybersécurité et élaborer des plans complets de réponse aux incidents. Des évaluations de sécurité régulières aideront à identifier et à corriger les vulnérabilités, garantissant ainsi l'intégrité et la fiabilité de nos systèmes d'IA. Dans des situations réelles, plusieurs mesures avancées de cybersécurité se sont révélées efficaces pour protéger les systèmes de reporting financier basés sur l'IA, comme l'authentification multifacteur qui améliore considérablement la sécurité en exigeant que les utilisateurs fournissent au moins deux facteurs de vérification pour accéder à un système. , un chiffrement de bout en bout qui garantit que les données sont chiffrées depuis le moment où elles sont créées jusqu'à ce qu'elles soient reçues et déchiffrées par le destinataire prévu, ou la réalisation régulière d'audits de sécurité et de tests d'intrusion qui permettent aux organisations d'identifier et de traiter de manière proactive les vulnérabilités de leurs systèmes [29].

5. Résultats et discussion

La combinaison d'un modèle de réseau neuronal avec la théorie des jeux, telle que représentée dans une matrice, peut constituer une approche puissante. Elle est souvent appelée apprentissage automatique de la théorie des jeux et est utilisée pour l'aide à la décision dans les entreprises. Le réseau neuronal peut être utilisé pour prédire les KPI financiers ou d'autres résultats pertinents, et la matrice de la théorie des jeux peut aider à analyser les interactions stratégiques entre différentes entités ou parties prenantes. Il vise à comprendre comment différents agents ou acteurs, chacun avec ses propres objectifs, prennent des décisions dans un contexte compétitif ou

environnement coopératif. Nous utilisons la théorie des jeux comme branche des mathématiques et de l'économie pour surveiller les interactions entre décideurs rationnels, souvent évoquées dans la littérature en tant que joueurs, dans des situations où le résultat de la décision d'un joueur dépend du décisions des autres, et nous proposons de combiner des techniques d'apprentissage automatique et d'intégrer avec la théorie des jeux pour modéliser et prédire le comportement des joueurs qui peuvent bénéficier de création de valeur commerciale positive. Dans notre théorie des jeux, les joueurs représentent les Actionnaires et l'équipe de direction, et les stratégies comprennent une augmentation de la capitalisation boursière de augmenter les revenus ou la rentabilité et les gains représentés dans la matrice. Stratégies proposés par la recherche sont les choix qui s'offrent à chaque joueur comme le choix du numérisation et intégration de l'IA ou respect des méthodes comptables traditionnelles et les rapports, et les gains représentent les résultats associés à des combinaisons spécifiques de stratégies choisies par tous les joueurs. Pour appliquer les stratégies, des modèles prédictifs, tels que des réseaux de neurones avec des couches d'arbres peuvent être utilisés pour estimer les choix ou les actions probables des acteurs sur la base de données comptables historiques extraites du bilan et du résultat et les données de perte ou d'autres fonctionnalités pertinentes.

Compte tenu des spécifications de notre réseau neuronal et de la structure de la matrice de gains de la théorie des jeux, nous suggérons de nommer le modèle théorique « Decision Harbor AI » qui reflète l'intégration de l'intelligence artificielle, de la prise de décision financière et de la stratégie analyse et symbolise un port sûr et informé pour la prise de décision assistée par l'intelligence artificielle. intelligence. L'apprentissage automatique de la théorie des jeux fournit un cadre pour comprendre et prédire les interactions stratégiques [30]. En intégrant des modèles d'apprentissage automatique avec concepts de la théorie des jeux, nous prévoyons qu'il deviendra possible de mieux comprendre les processus de prise de décision dans des environnements complexes et dynamiques. Cette approche peut être précieuse pour prendre des décisions éclairées dans des scénarios compétitifs ou coopératifs où plusieurs les parties prenantes sont impliquées [31].

Dans notre scénario, nous avons une matrice de théorie des jeux avec deux acteurs : la direction de l'entreprise et les actionnaires et la matrice représente les résultats possibles en fonction des décisions prises par ces joueurs. Nous pouvons entraîner le réseau neuronal à prédire les résultats financiers KPI ou résultats d'intérêt, tels que la croissance des revenus, la rentabilité et la rentabilité, ainsi que comme le montre la figure 4. Après avoir entraîné le réseau neuronal à prédire les indicateurs de performance positifs , nous avons utilisé le réseau neuronal entraîné pour faire des prédictions pour chaque joueur (Société Management, AI et TR) en fonction des fonctionnalités d'entrée et du KPI dans la couche cachée.

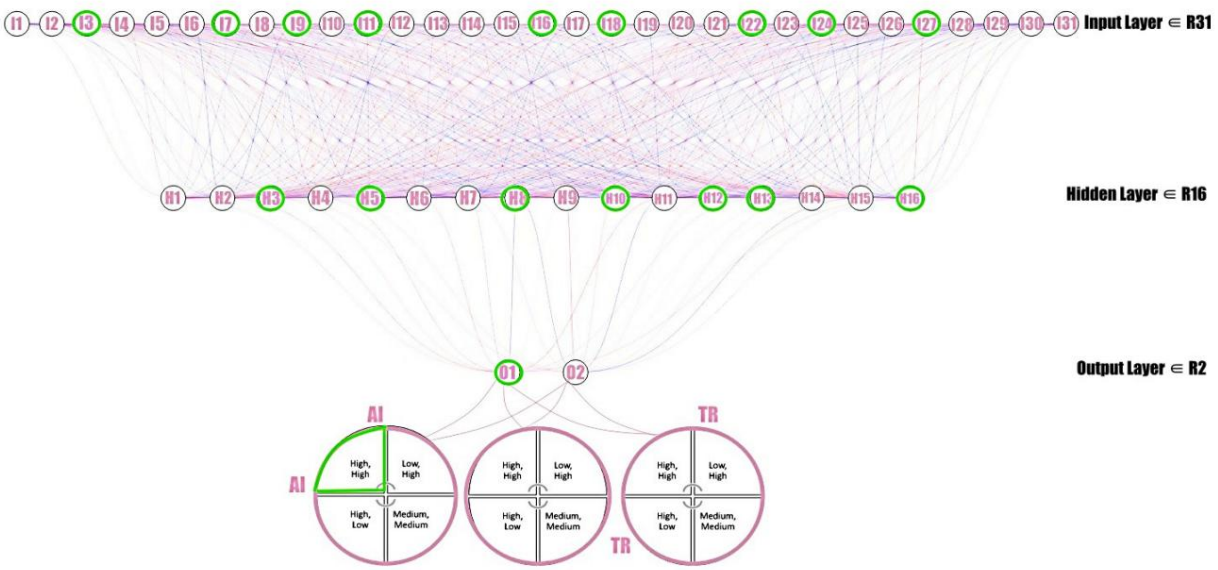


Figure 4. Modèle théorique d'IA de Decision Harbor – scénario 1.

Figure 4. Modèle théorique d'IA de Decision Harbor – scénario 1.

Lors des mauvaises années financières, la répartition des données peut changer. Les modèles et les relations entre les caractéristiques d'entrée et les résultats financiers peuvent changer, entraînant une inadéquation entre les données de formation et de test. Le réseau neuronal, formé sur des données historiques, pourrait avoir du mal à se généraliser à ces nouvelles conditions. Les performances prédictives du modèle peuvent se dégrader lors de mauvais exercices financiers si les schémas observés lors de la formation ne se vérifient plus. Le modèle peut faire des prédictions inexactes, surtout s'il n'a pas rencontré de scénarios similaires dans les données d'entraînement (Figure 5).

Figure 4. Modèle théorique d'IA de Decision Harbor – scénario 1.

Lors des mauvaises années financières, la répartition des données peut changer. Les modèles et les relations
des intrants et les résultats financiers peuvent changer, conduisant à une inadéquation entre les caractéristiques des intrants et les résultats financiers.
Les relations entre les fonctionnalités d'entrée et les résultats financiers peuvent changer, entraînant une inadéquation
entre les données de formation et de test. Le réseau neuronal, formé sur des données historiques, pourrait avoir du mal
entre les données de formation et de test. Le réseau neuronal, formé sur des données historiques, pourrait se
généraliser à ces nouvelles conditions. Les performances prédictives du modèle peuvent se dégrader
du mal à généraliser à ces nouvelles conditions. Les performances prédictives du modèle peuvent être affectées lors
se dégrader lors des mauvais exercices financiers si les schémas observés lors de la formation ne permettent plus au
de mauvais exercices financiers si les tendances observées lors de la formation ne sont plus valables. Le
prise. Le modèle peut faire des prédictions inexactes, surtout s'il n'a pas rencontré de phénomènes similaires
modèle de faire des prédictions inexactes, surtout s'il n'a pas rencontré de scénarios dans les données
d'entraînement (Figure 5).
scénarios similaires dans les données de formation (Figure 5).

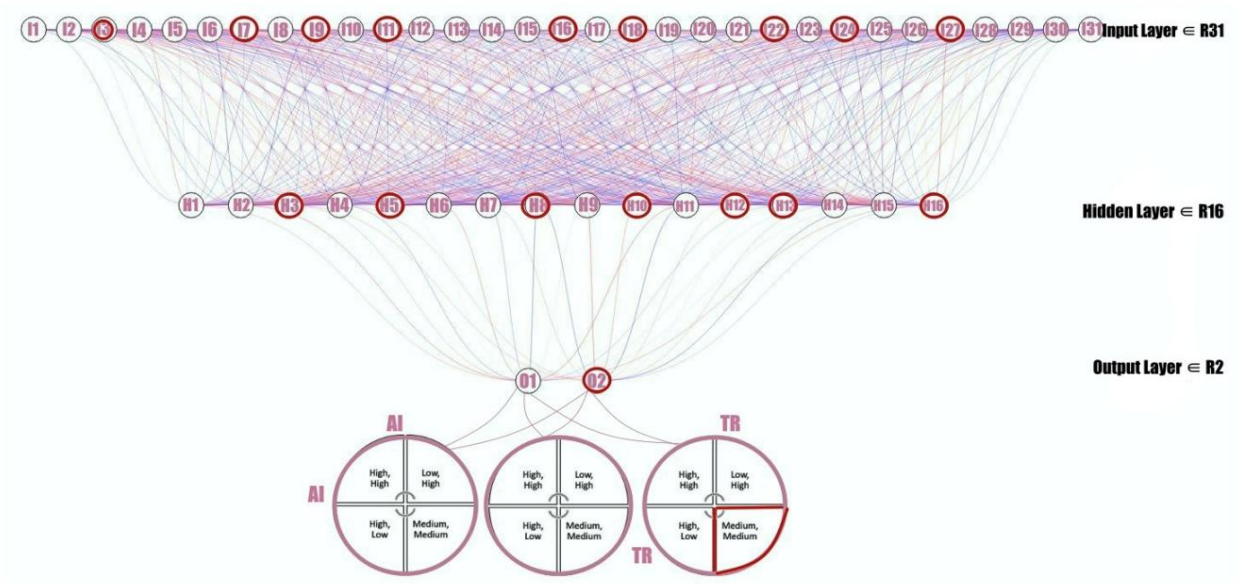


Figure 5. Modèle théorique d'IA de Decision Harbor – scénario 2.
Figure 5. Modèle théorique d'IA de Decision Harbor – scénario 2.

Nous suggérons à l'entreprise de mettre en œuvre une stratégie de formation continue des modèles. Nous
pour s'adapter à l'évolution des conditions financières. Le modèle doit être régulièrement mis à jour pour s'adapter à
l'évolution des conditions financières. Le modèle doit être régulièrement mis à jour avec de nouveaux
de nouvelles données, en particulier pendant les périodes de turbulences financières, pour garantir leur pertinence.
données, en particulier pendant les périodes de turbulences financières, pour garantir qu'elles restent pertinentes. Pour
Pour que le modèle conduise à de bonnes décisions financières, le management doit conduire un scénario. Pour que
le modèle conduise à de bonnes décisions financières, le management doit conduire un scénario.
Analyse nario pour évaluer les performances du modèle dans diverses conditions financières. Cette analyse permet
d'évaluer la performance du modèle dans diverses conditions financières. Ceci peut
peut impliquer de tester le modèle par rapport à des scénarios simulés représentant différents scénarios économiques.
États économiques, y compris les bonnes et les mauvaises années financières. Si le réseau de neurones fonctionne
bien , il y a de bons et de mauvais exercices financiers. Si le réseau neuronal fonctionne bien pendant
bons exercices, il valide la robustesse du modèle. Cela suggère que
le modèle peut s'adapter à différentes conditions financières et continue de fournir des informations précises
prévisions lorsque les conditions sont favorables. De bons exercices financiers peuvent renforcer positivement
l'utilisation du modèle pour la prise de décision. Des prévisions fiables dans notre entreprise pendant
ces périodes peuvent contribuer à une planification stratégique efficace, à l'allocation des ressources et à d'autres
décisions qui augmentent la valeur commerciale. Les décideurs pourraient gagner en confiance
dans les prévisions du modèle lors des bons exercices financiers. Cette confiance peut conduire à davantage
confiance dans le modèle pour guider les décisions.

Le cycle de Deming (PDCA) peut être utilisé comme méthode de gestion itérative en quatre étapes
pour une amélioration continue. Appliquer le cycle PDCA à la formation continue et
analyse de scénarios de notre modèle de réseau neuronal dans le contexte de conditions financières évolutives
est une approche sensée et efficace. Nous devons surveiller le processus de formation et évaluer
si les performances du modèle s'améliorent ou restent stables, exécutez l'analyse du scénario
en utilisant les scripts ou outils de simulation préparés, et évaluer la précision, la sensibilité,
et spécificité dans différents scénarios pour identifier tout modèle de sous-performance
ou un surapprentissage.

Si l'analyse des scénarios de notre réseau de neurones révèle des lacunes, il faut considérer
ajuster le modèle ou le processus de génération de scénarios. Cela peut impliquer de redéfinir les
fonctionnalités , d'introduire de nouvelles fonctionnalités ou de modifier l'approche de simulation. Le théorique

Le modèle peut intégrer toutes les améliorations nécessaires pour améliorer sa capacité à gérer diverses conditions financières. Nous pouvons appliquer des tests de résistance à notre modèle [23], comme la façon dont les banques effectuent des tests de résistance pour évaluer la résilience de leurs systèmes financiers dans des conditions défavorables. Dans le contexte de notre modèle de réseau neuronal et de notre matrice de théorie des jeux, les tests de résistance peuvent nous aider à comprendre les performances du modèle et la robustesse de la prise de décision face à des scénarios financiers extrêmes ou inattendus. L'apprentissage profond peut être appliqué à la couche d'entrée de notre réseau dans le contexte de tests de résistance dynamiques du bilan [31,32]. Si le bilan qui génère des informations financières pour les données de la couche d'entrée a une dimension temporelle telle que trimestrielle ou annuelle, nous pouvons envisager des réseaux de neurones récurrents ou des réseaux de mémoire à long terme et à court terme (LSTM) pour capturer les modèles temporels et les dépendances dans les données. Si nos données d'entrée incluent diverses sources telles que des états financiers, des données de marché et des indicateurs économiques, nous suggérons d'appliquer des architectures prenant en charge l'intégration de données multimodales. Cela permet au modèle d'apprendre simultanément de différents types d'informations. Dans l'ensemble, il est important d'adapter l'approche d'apprentissage en profondeur aux exigences et caractéristiques spécifiques de nos données financières et aux objectifs des tests de résistance. Notre modèle théorique DecisionHarborAI peut également être mis en œuvre dans les organisations publiques. L'application d'outils d'IA, comme ce modèle théorique à deux niveaux, dans le secteur public a le potentiel d'apporter des avantages significatifs, comme aider à analyser l'impact de différentes politiques, aider les décideurs politiques à comprendre les résultats potentiels et à prendre des décisions éclairées, fournir des informations sur l'optimisation. l'allocation des ressources pour atteindre les résultats souhaités, l'évaluation de l'efficacité des programmes et initiatives publics et le renforcement de l'engagement des citoyens en fournissant des informations basées sur des données sur les processus décisionnels , en favorisant la transparence et la confiance entre le public et le gouvernement [33].

Pour mettre en œuvre efficacement un tel modèle, plusieurs recommandations pratiques émergent. Des mises à jour régulières et un recyclage du réseau neuronal avec de nouvelles données sont cruciaux pour maintenir l'exactitude et la pertinence.

La première étape consiste à fournir une image ou un fichier PDF d'un bilan à notre API. Cela se déroulera dans une application mobile ou une application Web. Dès qu'une image ou un PDF est reçu, il est converti en fichier TXT. Un analyseur prend le TXT obtenu grâce à l'OCR et le convertit en JSON structuré à l'aide de l'apprentissage automatique. Le JSON est ensuite renvoyé en tant que sortie de l'API. À partir de là, les données du bilan peuvent être traitées davantage. L'ensemble de données complet est ensuite utilisé pour former un réseau neuronal, capable de prédire des indicateurs de performance financière clés tels que la croissance des revenus, la rentabilité et la rentabilité. Le modèle DecisionHarborAI va encore plus loin en intégrant un cadre de théorie des jeux. Cela permet une analyse nuancée des interactions stratégiques entre les parties prenantes, telles que la direction et les actionnaires. En comprenant ces dynamiques, l'institution peut aligner ses décisions stratégiques sur les intérêts de toutes les parties impliquées, maximisant ainsi la valeur commerciale.

Pour garantir la robustesse, une analyse continue de scénarios et des tests de résistance sont effectués. Cette pratique valide non seulement les prédictions du modèle dans diverses conditions, mais prépare également l'institution à d'éventuelles fluctuations économiques. Les résultats sont profonds : des capacités de prise de décision améliorées, des gains d'efficacité significatifs grâce à l'automatisation et un alignement stratégique qui stimule la performance globale de l'entreprise.

Des limites doivent également être prises en compte lors de la mise en œuvre de modèles de reporting financier basés sur l'IA tels que DecisionHarborAI. L'une des limites les plus importantes réside dans la qualité des données introduites dans les modèles d'IA. Les systèmes d'IA nécessitent de grandes quantités de données de haute qualité pour fonctionner de manière optimale. Cependant, les données financières peuvent souvent être incomplètes, incohérentes ou bruitées. Des données inexactes ou de mauvaise qualité peuvent conduire à des prévisions erronées et à des processus décisionnels erronés. Les organisations doivent investir des efforts considérables dans le nettoyage et la validation des données pour garantir que les données utilisées à des fins de formation et opérationnelles sont exactes et fiables. Cela peut nécessiter beaucoup de ressources et n'est pas toujours réalisable, en particulier pour les petites institutions disposant de budgets limités. Comme tous les modèles d'IA, DecisionHarborAI sera formé sur des données historiques et pourra fonctionner correctement dans des conditions similaires à celles présentes dans l'ensemble de données de formation. La capacité de notre modèle théorique à s'adapter aux changements économique

les changements réglementaires et les événements mondiaux imprévus tels que les crises financières ou les pandémies constituent une préoccupation majeure [34]. Le modèle DecisionHarborAI peut ne pas parvenir à se généraliser efficacement à de nouveaux scénarios invisibles, entraînant une dégradation des performances. Un recyclage et une validation continus du modèle seront nécessaires pour garantir que DecisionHarborAI reste pertinent et précis, mais cela peut être difficile à gérer et nécessite un investissement continu.

6. Conclusions

Lors de la mise en œuvre d'un réseau neuronal pour la comptabilité financière, il est important de prendre en compte des facteurs tels que la qualité des données, l'interprétabilité du modèle et l'environnement réglementaire spécifique. Le modèle proposé par cette recherche offre une approche intégrée de l'aide à la décision en combinant les capacités prédictives des réseaux de neurones avec les informations stratégiques fournies par la théorie des jeux. Cette intégration vise à améliorer les processus décisionnels dans des environnements complexes et dynamiques et à recentrer le modèle sur la prévision des indicateurs financiers et le soutien à la prise de décision financière. Le composant réseau neuronal du nouvel outil d'IA est conçu pour fournir des prévisions et des informations financières basées sur des données historiques. Cela inclut la prévision d'indicateurs financiers clés tels que la croissance des revenus, la rentabilité et la rentabilité. Les réseaux de neurones peuvent être un outil précieux pour calibrer les résultats dans l'environnement de comptabilité financière et ont démontré leur efficacité dans le traitement de données financières complexes, l'identification de modèles et la réalisation de prévisions . L'inclusion d'une matrice de gains de théorie des jeux dans ce nouvel outil d'aide à la décision en IA introduit un cadre stratégique pour analyser les interactions entre les décideurs.
Cela permet de considérer les choix et les résultats stratégiques dans un contexte plus dynamique et compétitif. Les outils de transformation numérique de l'IA ont le potentiel d' améliorer considérablement les processus de prise de décision financière et la gestion stratégique globale. Ces outils peuvent aider les organisations à tirer parti des informations basées sur les données et à automatiser les tâches de routine, améliorant ainsi potentiellement l'efficacité et la précision des décisions. Cependant, il est important de reconnaître que les technologies d'IA comportent également des défis et des inconvénients, tels que des problèmes de confidentialité des données, le risque de biais algorithmiques et la nécessité de ressources informatiques importantes. De plus, la complexité des modèles d'IA peut les rendre difficiles à interpréter et à faire confiance, ce qui peut limiter leur application pratique dans certains contextes. Par conséquent, même si l'IA offre des avancées prometteuses, il est essentiel d'aborder son intégration dans une perspective équilibrée, en reconnaissant à la fois ses avantages et ses limites. Tous les futurs outils d'aide à la décision en développement doivent continuellement apprendre et mettre à jour leurs prédictions, garantissant ainsi leur pertinence et leur exactitude au fil du temps. Le modèle proposé dans ce document doit être doté de la capacité de subir des tests de résistance et une analyse de scénarios, lui permettant d' évaluer sa performance dans diverses conditions financières, y compris des scénarios à la fois favorables et défavorables, contribuant ainsi à sa robustesse.

La fusion des matrices de gains des réseaux neuronaux représente un outil d'aide à la décision complet et adaptatif qui exploite les atouts des réseaux neuronaux et de la théorie des jeux pour fournir des informations précieuses pour la prise de décision financière dans les secteurs privé et public et apporter une valeur tangible aux entreprises. La formation continue, l'analyse de scénarios et le cadre stratégique contribuent à son efficacité potentielle dans des environnements dynamiques et incertains. Le concept d'intégration de la prise de décision basée sur l'IA se reflète directement dans les conclusions de cet article. Les outils numériques combinés combinent les capacités prédictives des réseaux de neurones avec un cadre de théorie des jeux, mettant l'accent sur l'intégration de techniques avancées d'IA pour une prise de décision plus éclairée. Le succès du modèle dépendra de tests approfondis, de la collaboration avec des experts du domaine et d'améliorations continues basées sur des retours concrets.

Les outils de transformation numérique peuvent faciliter considérablement le processus de prise de décision financière et la gestion stratégique globale d'une entreprise en analysant les données financières historiques pour faire des prédictions précises sur les tendances et les résultats futurs. Ces outils peuvent évaluer et prédire les risques potentiels en analysant diverses sources de données. Cela permet aux organisations d'identifier et d'atténuer de manière proactive les risques liés à la prise de décision financière et d'automatiser les tâches répétitives et routinières, telles que la saisie des données, le rapprochement et le reporting.

Les outils de transformation numérique de l'IA jouent un rôle crucial dans l'amélioration des processus de prise de décision financière et de la gestion stratégique globale. Ces outils permettent aux organisations de tirer parti des informations basées sur les données, d'automatiser les tâches de routine et de naviguer plus efficacement dans les complexités du paysage commercial moderne.

Contributions des auteurs : Conceptualisation, AEA et AED ; AED et LI : analyse formelle, AEA : enquête, LI : ressources, AED : rédaction – préparation du projet original, AEA : rédaction – révision et édition, AED : visualisation, LI : supervision, AEA : administration du projet. Tous les auteurs ont lu et accepté la version publiée du manuscrit.

Financement : Cette recherche n'a reçu aucun financement externe.

Déclaration de disponibilité des données : toutes les données sous-jacentes aux résultats sont disponibles dans le cadre de l'article et aucune donnée source supplémentaire n'est requise.

Conflits d'intérêts : Les auteurs ne déclarent aucun conflit d'intérêts.

Annexe A

Tableau A1. Description des couches dans le modèle neuronal proposé : Decision Harbor AI.

Couche d'entrée R^{31} je	Couche cachée R 16 H	Couche de sortie R^{2} Ô
Revenu net	Revenu brut	Hausse des revenus
Revenu net		
Impôts	Revenu brut	
Frais d'intérêt	BAIIA	
Dépréciation		
Amortissement	Marges bénéficiaires	
Coût des marchandises vendues		
Dépenses de fonctionnement	Flux de trésorerie opérationnels	
Résultat opérationnel		
Variation du fonds de roulement	Libre circulation des capitaux	
Flux de trésorerie opérationnel		
Dépenses en capital	CHEVREUIL	
Capitaux propres		
Actif total moyen	ROA	
Responsabilités totales		
Actifs courants	Ratio dette/fonds propres	
Passifs courants		
Espèces	Ratio actuel	Rentabilité
Equivalents de trésorerie		
Créances nettes	Rapport rapide	
Dépenses payées d'avance		
Cours de l'action	Ratio cours/bénéfice	
Bénéfice par action		
Actions privilégiées	La valeur comptable par action	
Diviser par partage		
Valeur par action	Rendement en dividendes	
Total des actions		
Hausse des revenus	Capitalisation boursière	
Gains		
Volonté divine	Indicateurs de croissance	
Immobilisations		

Les références

1. Imoniana, JO ; Réginato, L. ; Junior, EBC; Benetti, C. Perceptions des parties prenantes sur l'adoption des Normes internationales d'information financière par les petites et moyennes entreprises : une analyse multidimensionnelle. Int. J.Autobus. Émerger. Marque. 2025, 1. [Référence croisée]
2. Radicic, D. ; Petkovi´c, S. Impact de la numérisation sur les innovations technologiques dans les petites et moyennes entreprises (PME). Technologie. Prévision. Soc. Chang. 2023, 191, 122474. [Réf. croisée]
3. Saura, JR ; Ribeiro-Soriano, D. ; Palacios-Marqués, D. Évaluation des problèmes de confidentialité de la science des données comportementales dans le secteur artificiel gouvernemental déploiement du renseignement. Gov. Inf. Q. 2022, 39, 101679. [CrossRef]
4. Faccia, A. ; Al Naqbi, MYK ; Lootah, SA Cycle de comptabilité financière intégré dans le cloud. Dans Actes de la 3e Conférence internationale 2019 sur le cloud et le Big Data Computing, Oxford, Royaume-Uni, 28-30 août 2019 ; ACM : New York, NY, États-Unis, 2019 ; p. 31-37. [Référence croisée]
5. Allam, K. ; Rodwal, A. Analyse Big Data basée sur l'IA : dévoiler des informations pour l'avancement des entreprises. EPH-Int. J. Sci. Ing. 2023, 9, 53-58. [Référence croisée]
6. Mengu, D. ; Luo, Y. ; Rivenson, Y. ; Ozcan, A. Analyse des réseaux de neurones optiques diffractifs et de leur intégration avec l'électronique Les réseaux de neurones. IEEE J. Sel. Haut. Électron quantique. 2020, 26, 1–14. [Référence croisée]
7. Shengélia, NSN ; Tsiklauri, ZTZ ; Rzepka, ARA; Shengelia, RSR L'impact des technologies financières sur la transformation numérique formation de la comptabilité, de l'audit et de l'information financière. Économie 2022, 105, 385-399. [Référence croisée]
8. Pacurari, D. ; Nechita, E. Quelques considérations sur la comptabilité cloud. En Etudes et Recherches Scientifiques. Édition économique ; Faculté des Sciences Economiques, Université « Vasile Alecsandri » de Bacau : Bacau, Roumanie, 2013. [CrossRef]
9. Li, M. ; Yu, Y. ; Li, X. ; Zhao, JL ; Zhao, D. Déterminants de la transformation des PME vers les services cloud : perspectives de rationalités économiques et sociales. Pac. Asie J. Assoc. Inf. Système. 2019, 11, 65-87. [Référence croisée]
10. Sorensen, M. ; Yasuda, A. Impact des investissements en capital-investissement sur les parties prenantes. Dans Manuel d'économie de la finance d'entreprise : capital-investissement et finance entrepreneuriale ; Elsevier : Amsterdam, Pays-Bas, 2023 ; pp. 299-341. [Référence croisée]
11. Naik, N. ; Hameed, BZ ; Shetty, DK ; Swain, D. ; Shah, M. ; Paul, R. ; Aggarwal, K. ; Ibrahim, S. ; Patil, V. ; Smriti, K. ; et coll. Considérations juridiques et éthiques liées à l'intelligence artificielle dans les soins de santé : qui en assume la responsabilité ? Devant. Surg. 2022, 9, 862322. [CrossRef]
12. Vo-Thanh, T. ; Zaman, M. ; Hasan, R. ; Akter, S. ; Dang-Van, T. La digitalisation des services dans les restaurants gastronomiques : un rapport coût-bénéfice perspective. Int. J.Contemp. Hôpital. Gérer. 2022, 34, 3502-3524. [Référence croisée]
13. Agustí-Juan, I. ; Verre, J. ; Pawar, V. Un tableau de bord équilibré pour évaluer l'automatisation dans la construction. Dans Actes de la Creative Construction Conference 2019, Budapest, Hongrie, 29 juin-2 juillet 2019 ; Université de technologie et d'économie de Budapest : Budapest, Hongrie, 2019 ; 155-163. [Référence croisée]
14. Wang, Z. ; Wang, Q. ; Nie, Z. ; Li, B. Prédiction de la détresse financière des entreprises basée sur l'engagement de capitaux propres de l'actionnaire majoritaire. Appl. Écon. Lett. 2022, 29, 1365-1368. [Référence croisée]
15. Du, G. ; Elston, F. ARTICLE RÉTRAIT : Évaluation des risques financiers pour améliorer l'exactitude des prévisions financières dans le secteur financier sur Internet à l'aide de modèles d'analyse de données. Opéra. Gérer. Rés. 2022, 15, 925-940. [Référence croisée]
16. Chanson, Z. ; Wu, S. Théorie des jeux de prévisions financières et prise de décision. Finance. Rés. Lett. 2023, 58, 104288. [Réf. croisée]
17. Murali, M. ; Kanmani, S. Modélisation d'applications de théorie des jeux utilisant des éléments de réseau pour assurer l'équilibre de Nash via le prix de l'anarchie et le prix de la stabilité. Int. J. Syst. Système. Ing. 2024, 14. [Réf. croisée]
18. Hamido˘glu, A. Une approche théorique des jeux sur la construction d'un nouveau modèle de chaîne d'approvisionnement agroalimentaire soutenu par le gouvernement. Expert. Système. Appl. 2024, 237, 121353. [Réf. croisée]
19. Zhuang, Yukon ; Wu, F. ; Chen, C. ; Pan, YH Défis et opportunités : du big data à la connaissance de l'IA 2.0. Devant. Inf. Technologie. Électron. Ing. 2017, 18, 3-14. [Référence croisée]
20. Mahmood, A. ; Al Marzooqi, A. ; El Khatib, M. ; AlAmeemi, H. Comment l'intelligence artificielle peut tirer parti du système d'information de gestion de projet (PMIS) et de la prise de décision basée sur les données dans la gestion de projet. Int. J.Autobus. Anal. Sécurisé. (IJBAS) 2023, 3, 184-195. [Référence croisée]
21. Chongcs, J. ; Kathiarayan, V. ; Chong, J. ; Sin, C. Le rôle de l'intelligence artificielle dans les opportunités, les défis et les implications de prise de décision stratégique pour les managers à l'ère numérique. Int. J. Manag. Commer. Innover. 2023, 11, 73-79.
22. Laczkovich, M. Un théorème de superposition de type Kolmogorov pour les fonctions continues bornées. J. Env. Théorie 2021, 269, 105609. [Référence croisée]
23. Quraisy, A. Normalité des données à l'aide des tests de Kolmogorov-Smirnov et Shapiro-Wilk. J-HEST J. Éduc. Santé. Écon. Sci. Technologie. 2020, 3, 7-11. [Référence croisée]
24. Schmidt-Hieber, J. Le théorème de représentation de Kolmogorov-Arnold revisité. Réseau neuronal. 2021, 137, 119-126. [Référence croisée]
25. Ismailov, VE Un réseau neuronal à trois couches peut représenter n'importe quelle fonction multivariée. J. Math. Anal. Appl. 2023, 523, 127096. [Référence croisée]
26. Elahi, M. ; Afolaranmi, SO ; Lastra, JLM; Garcia, JAP Une revue complète de la littérature sur les applications des techniques d'IA tout au long du cycle de vie des équipements industriels. Découverte. Artif. Intell. 2023, 3, 43. [Réf. croisée]
27. Habbal, A. ; Ali, MK ; Abuzaraida, MA Confiance en intelligence artificielle, gestion des risques et de la sécurité (AI TRiSM) : cadres, applications, défis et orientations de recherche futures. Expert. Système. Appl. 2024, 240, 122442. [Réf. croisée]
28. Cheng, L. ; Liu, X. Des principes aux pratiques : l'interaction intertextuelle entre les discours éthiques et juridiques de l'IA. Int. J. Leg. Discours 2023, 8, 31-52. [Référence croisée]

29. Humphreys, D. ; Koay, A. ; Desmond, D. ; Mealy, E. Le battage médiatique sur l'IA en tant que risque de cybersécurité : la responsabilité morale de la mise en œuvre IA générative en entreprise. Éthique de l'IA 2024. [CrossRef]
30. Cho, S. ; Vasarhelyi, MA; Soleil, T. ; Zhang, C. Apprendre de l'apprentissage automatique en comptabilité et en assurance. J. Emerg. Technologie. Compte. 2020, 17, 1–10. [Référence croisée]
31. Ganaie, MA; Hum.; Malik, AK ; Tanveer, M. ; Suganthan, PN Ensemble deep learning : une revue. Ing. Appl. Artif. Intell. 2022, 115, 105151. [Réf. croisée]
32. Zhang, Z. ; Cui, P. ; Zhu, W. Deep Learning sur les graphiques : une enquête. IEEETrans. Connaître. Ingénierie des données. 2022, 34, 249-270. [Référence croisée]
33. Chandana Charitha, P. ; Hemaraju, B. Impact de l'intelligence artificielle sur la prise de décision dans les organisations. Int. J. Multidisciplinaire. Rés. 2023, 5. [Référence croisée]
34. Aldoseri, A. ; Al-Khalifa, KN ; Hamouda, AM Repenser la stratégie et l'intégration des données pour l'intelligence artificielle : concepts, Opportunités et défis. Appl. Sci. 2023, 13, 7082. [Réf. croisée]

électronique

MDPI

Article

Apprendre à améliorer l'efficacité opérationnelle à partir d'une erreur de pose Estimation en pollinisation robotique

Jinlong Chen [1], Jun Xiao 1,*, Minghao Yang [2] et suspendre la poêle [3]

1 École d'informatique et de sécurité de l'information, Université de technologie électronique de Guilin, Guilin 541000, Chine ; 7259@guet.edu.cn Centre
2 de recherche sur l'intelligence inspirée du cerveau (BII), Institut d'automatisation, Académie chinoise des sciences (CASIA), Pékin 100190, Chine ; mhyang@nlpr.ia.ac.cn
3 Département d'informatique, Université de Changzhi, Changzhi 046011, Chine ; panhang@czc.edu.cn
* Correspondance : 22032303110@mails.guet.edu.cn

Résumé : Les robots de pollinisation autonomes ont été largement discutés ces dernières années. Cependant, l' estimation précise de la pose des fleurs dans des environnements agricoles complexes reste un défi. À cette fin, ce travail propose la mise en œuvre d'une architecture basée sur un transformateur pour apprendre les erreurs de translation et de rotation entre l'effecteur final du robot de pollinisation et l'objet cible dans le but d'améliorer l'efficacité de la pollinisation robotique dans les tâches de croisement. Les contributions sont les suivantes : (1) Nous avons développé un modèle d'architecture de transformateur, équipé de deux réseaux de neurones feedforward qui régressent directement les erreurs de translation et de rotation entre l' effecteur terminal du robot et la cible de pollinisation. (2) De plus, nous avons conçu une fonction de perte de régression guidée par les erreurs de translation et de rotation entre l'effecteur final du robot et les cibles de pollinisation. Cela permet au bras du robot d'identifier rapidement et précisément la cible de pollinisation à partir de la position actuelle. (3) De plus, nous avons conçu une stratégie permettant d'acquérir facilement un nombre important d'échantillons d'entraînement à partir de l'observation œil dans la main, qui peuvent être utilisés comme entrées pour le modèle. Pendant ce temps, les erreurs de translation et de rotation identifiées dans le système de coordonnées cartésiennes du manipulateur final sont désignées simultanément comme cibles de perte. Cela permet d'optimiser la formation du modèle. Nous avons mené des expériences sur un système de pollinisation robotique réaliste. Les résultats démontrent que la méthode proposée surpasse la méthode de l'état de l'art, en termes de précision et d'efficacité.

Mots-clés : robot de pollinisation ; transformateur; erreurs de décalage

Citation : Chen, J. ; Xiao, J. ; Yang, M. ; Pan, H. Apprendre à s'améliorer Efficacité opérationnelle de Pose Estimation des erreurs en robotique Pollinisation. Électronique 2024, 13, 3070. https://doi.org/10.3390/électronique13153070

Rédacteur académique : Dah-Jye Lee

Reçu : 24 juin 2024
Révisé : 26 juillet 2024
Accepté : 1er août 2024
Publié : 2 août 2024

1. Introduction

La pollinisation assistée par robot est un sujet depuis plus d'une décennie [1]. Il devient un contributeur de plus en plus important à la production végétale dans l'industrie agricole et apporte des solutions à des défis urgents, tels que le déclin des pollinisateurs naturels.
Les robots de pollinisation sont classés en plusieurs types, pour répondre aux divers besoins agricoles. Des robots autonomes, dotés d'une vision avancée et d'une intelligence artificielle, naviguent et pollinisent de manière indépendante. Les robots télécommandés ou semi-autonomes nécessitent une surveillance humaine, adaptés aux environnements complexes. Les véhicules aériens sans pilote (UAV) pollinisent efficacement de grands champs depuis le haut. Ces systèmes robotiques sont conçus pour imiter le processus de pollinisation traditionnellement effectué par les abeilles et autres insectes. Utilisant des capteurs avancés, des systèmes de vision et des manipulateurs de précision, les pollinisateurs robotisés identifient et interagissent avec les fleurs, déposant du pollen avec une grande précision et cohérence.

Diverses études ont été discutées dans le domaine de la pollinisation automatisée avec des applications dans la pollinisation des kiwis [2,3], des fleurs de ronce [4], de vanille [5] et des fleurs de tomates [6]. La technique de pollinisation nécessite le transfert physique du pollen du mâle vers les pistils (les organes reproducteurs des fleurs femelles) [7]. Pour minimiser la perte de pollen et assurer un transfert précis vers les stigmates, il est nécessaire de remédier aux difficultés

de la collecte du pollen et les défis du maintien de son activité [8]. Malheureusement, cela nécessite beaucoup de temps et d'efforts, qui n'ont pas été soigneusement pris en compte dans les approches de pollinisation existantes. Les difficultés proviennent de la très petite taille des pistils, ainsi que du défi d'identifier avec précision leur rotation et leur orientation.

À cette fin, ce travail propose une méthodologie basée sur une architecture de transformateur qui vise à améliorer l'efficacité opérationnelle de la pollinisation robotisée dans les tâches de croisement. La couche de sortie du modèle intègre deux réseaux neuronaux à action directe distincts, chacun conçu pour régresser les écarts de translation et de rotation entre le robot et la cible de pollinisation, respectivement. Contrairement aux fonctions de perte de transformateur traditionnelles, nous avons conçu une nouvelle fonction de perte qui s'appuie sur les écarts de translation et de rotation entre l'effecteur terminal du robot et les cibles de pollinisation. Cela permet au bras du robot d'identifier rapidement et précisément la cible de pollinisation à partir de la vue d'observation actuelle. Parallèlement, nous avons également conçu une stratégie permettant de collecter des échantillons d'apprentissage à grande échelle à partir d'observations œil dans la main et de désigner les erreurs de translation et de rotation identifiées dans le système de coordonnées cartésiennes du manipulateur final comme cibles de perte. Les images observées dans les positions actuelle et cible, combinées aux erreurs de translation et de rotation entre elles, fournissent une grande échelle pour les données de formation, améliorant ainsi la formation du modèle. La figure 1 montre le robot de pollinisation utilisé dans ce travail.

La suite de ce travail est organisée comme suit : les travaux associés sont présentés dans la section 2 ; les descriptions détaillées de la méthode sont introduites dans la section 3 ; Les sections 4 à 6 présentent respectivement les expériences, la discussion et la conclusion.

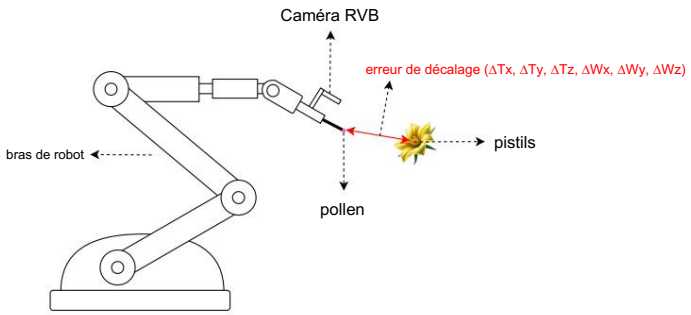


Figure 1. Le robot de pollinisation se compose principalement d'un bras robotique UR5 et d'une caméra RVB monoculaire, configurés dans une structure œil dans la main. Au bout du bras robotique se trouve une brosse de pollinisation utilisée pour polliniser les pistils. L'illustration représente un bras robotique engagé dans la pollinisation, où « l'erreur de décalage » marquée en rouge représente le contenu que le modèle doit apprendre. Ici, ΔTx, ΔTy et ΔTz désignent les composantes de l'erreur de translation le long des axes X, Y et Z , respectivement, tandis que ΔWx, ΔWy et ΔWz représentent les composantes du vecteur rotation, indiquant la erreurs de rotation.

2. Travaux connexes

2.1. Estimation de pose

Atteindre une haute précision dans la pose de l'effecteur final d'un bras robotique pose des défis importants. Pour améliorer la précision de la pose de l'effecteur terminal, des capteurs supplémentaires, tels que des capteurs LiDAR ou de pression, sont souvent utilisés, ce qui peut augmenter les coûts. Certaines méthodes sont basées sur l'estimation de la pose d'objets 6D, construisant des modèles pour numériser différentes positions dans l'image d'entrée, calculant les scores de similarité à chaque position et obtenant la meilleure correspondance en comparant les scores de similarité [9,10]. D'autres utilisent des méthodes basées sur les caractéristiques, où les modèles CNN régressent directement les coordonnées 3D de chaque pixel [11], ou décrivent la densité postérieure d'une pose d'objet spécifique via des CNN, en comparant l'image observée avec l' image rendue [12]. De plus, certaines méthodes combinent les avantages des approches basées sur des modèles et des fonctionnalités dans un cadre d'apprentissage en profondeur. Le réseau intègre un étiquetage ascendant au niveau des pixels avec une régression descendante de la pose d'objet, prédisant

la position 3D et la rotation 3D de l'objet de manière découplée [13]. Toutes ces méthodes utilisent invariablement des capteurs de profondeur pour capturer les informations de profondeur D de l'objet cible. Cependant , pour les scénarios dans lesquels les informations de profondeur ne peuvent pas être capturées, comme dans le cas du robot de pollinisation agricole évoqué dans cet article, les régions creuses des fleurs de pollinisation entraînent une mauvaise acquisition des informations de profondeur. De plus, les modèles entraînés intégrant des informations de profondeur [11-13] ne régressent pas et ne prédisent pas directement les informations de position 3D de l' objet cible. Les modèles susmentionnés basés sur des modèles et des fonctionnalités sont formés sur des nuages de points, mais ils sont toujours confrontés à des défis lors de la gestion de polyèdres non convexes et de données non multiples. Les méthodes existantes de formation par nuages de points conviennent principalement aux formes régulières, telles que les polyèdres convexes et les variétés. Ces méthodes sont limitées lorsqu'il s'agit de formes complexes et irrégulières de fleurs en milieu agricole, qui présentent souvent des structures non polyédriques et non multiples. Cette limitation rend difficile l' application directe des méthodes traditionnelles de formation par nuages de points. De plus, il existe un écart entre la formation aux nuages de points dans les environnements virtuels et les applications du monde réel, ce qui peut affecter les performances du modèle dans des scénarios réels.

Contrairement aux techniques d'estimation de pose introduites précédemment, qui s'appuient sur des capteurs supplémentaires pour capturer des données d'image améliorées en profondeur afin de construire une formation en nuages de points - une pratique difficile à mettre en œuvre dans des environnements agricoles non convexes et non variés - notre méthode utilise uniquement caméras conventionnelles, pour acquérir des images RVB à partir de l'effecteur final du robot de pollinisation. Cette approche permet une formation transparente de bout en bout dans des environnements réels, éliminant ainsi les écarts couramment constatés entre la formation par nuages de points dans des environnements virtuels et leur application pratique sur le terrain.

2.2. Méthodes traditionnelles et d'apprentissage profond pour améliorer la

précision Pour améliorer la précision de l'effecteur final de pollinisation, des méthodes traditionnelles peuvent être mises en œuvre grâce à la conception du bras robotique. Par exemple, LI K a conçu un bras robotique pour la pollinisation des kiwis [14], atteignant une haute précision en exprimant la vitesse de l'effecteur final et les vitesses angulaires des articulations à travers la matrice jacobienne. Lors de la planification de la trajectoire, l'interpolation polynomiale quintique assure la continuité et la douceur des courbes de vitesse et d'accélération de chaque articulation. La cinématique directe et les méthodes de Monte Carlo sont utilisées pour l'analyse de simulation, afin de couvrir toute la zone de pollinisation avec l'effecteur final. Dans le système de vision, une caméra binoculaire [15] est utilisée pour obtenir les coordonnées 3D des fleurs en temps réel, combinée au système de contrôle de planification de trajectoire et de calcul de point d'interpolation, pour contrôler avec précision chaque moteur commun et réaliser un point à point. pollinisation précise . Cependant, ce bras robotique s'appuie sur des caméras binoculaires pour obtenir les coordonnées 3D des fleurs, et une forte lumière ou des occlusions peuvent affecter la précision du sys

Les algorithmes traditionnels peuvent encore produire certains effets dans des domaines spécifiques de pollinisation agricole. Une étude récente de N Duc Tai [16] impliquait l'utilisation de méthodes de segmentation pour localiser et segmenter les fleurs de cantaloup, l'utilisation de modèles mathématiques basés sur les caractéristiques biologiques des fleurs de cantaloup pour déterminer les points clés de la direction de croissance de chaque fleur, et l'utilisation d'une projection inverse. méthode pour convertir la position de la fleur d'une image 2D en un espace 3D, réalisant ainsi la localisation de la pose de la fleur.

Ces dernières années, avec le développement et l'application de l'apprentissage profond, le robot de pollinisation BrambleBee [4] de STRADER J a combiné des algorithmes traditionnels et des algorithmes d'apprentissage profond pour estimer la pose des fleurs. Le système utilise un cadre de traitement d'image en deux étapes pour reconnaître les fleurs et estimer leurs poses. Il utilise d'abord un algorithme de segmentation Naive Bayes au niveau des pixels, puis utilise un réseau neuronal convolutionnel pour la classification et l'affinement des poses, garantissant ainsi l'exactitude des résultats de reconnaissance. Enfin, une carte d'obstacles basée sur un octree et un graphique de facteurs représentent la carte des fleurs, cartographiant les fleurs et leurs positions en 3D pour optimiser l'estimation de la pose. Le robot de pollinisation automatique des fleurs de pastèque de Khubaib Ahmad [17] utilise l'apprentissage profond pour détecter les informations 2D des fleurs et les combine avec des algorithmes traditionnels, pour calculer la profondeur de la position des fleurs, réalisant ainsi la localisation des fleurs. Le robot ajuste ensuite le bras mécanique, à l'aide d'une recherche servo, jusqu'à ce qu'il atteigne la pos

Le robot de pollinisation asservi automatisé basé sur la vision de YANG [18] utilise des algorithmes de détection de cible plus avancés, tels que YOLOv5, YOLACT++ [19] et DETR [20], pour détecter les fleurs, identifiant la position et l'orientation du pistil par rotation. technologie de détection d'objets . Le système utilise ensuite une stratégie de télémétrie pseudo-binoculaire, calculant les coordonnées 3D du pistil en déplaçant la position de la caméra et en utilisant l'étalonnage œil-main pour transformer ces coordonnées dans le système de coordonnées opérationnel du bras du robot . Pendant la tâche de pollinisation, le système combine des stratégies d'asservissement visuel, effectuant un positionnement grossier en déplaçant l'effecteur terminal près de la fleur, suivi d' un positionnement précis à l'aide d'une trajectoire de recherche circulaire pour assurer un contact précis entre la brosse à pollen et le pistil. Cependant, le système présente toujours une erreur de précision de l'effecteur final de 15 mm. Bien que la trajectoire de recherche circulaire puisse éventuellement permettre d'obtenir un positionnement précis, le processus prend beaucoup de temps, puisqu'il faut près de 19 secondes pour terminer la pollinisation d'une seule fleur.
Si la précision de l'effecteur final peut être améliorée, l'efficacité du positionnement précis sur la trajectoire de recherche circulaire en bénéficiera considérablement.

Les méthodes traditionnelles de N Duc Tai et les robots de pollinisation basés sur l'apprentissage profond de STRADER J, Khubaib Ahmad et YANG ont un impact significatif sur le taux de réussite et l'efficacité des opérations de pollinisation en raison d'erreurs dans l'estimation de la pose des cibles de pollinisation. Par exemple, les robots de pollinisation de N Duc Tai et YANG ne peuvent assurer une pollinisation réussie qu'en maintenant une vaste zone de recherche d'asservissement, ce qui entraîne une efficacité réduite de la pollinisation. Par conséquent, notre travail propose une méthode d'apprentissage profond pour apprendre les erreurs de translation et de rotation entre la fleur et l'effecteur final de pollinisation.
En compensant ces erreurs, la précision de positionnement de l'effecteur final du robot est améliorée. Cette réduction de la portée et de la trajectoire de recherche du mécanisme d'asservissement améliore finalement l'efficacité du processus de pollinisation.

2.3. Architecture du transformateur et codage de position

L'architecture du transformateur, introduite par Vaswani et al. [21], a eu un impact significatif sur le traitement du langage naturel (NLP) et la vision par ordinateur. Les transformateurs utilisent le mécanisme d'auto-attention pour capturer les dépendances complexes au sein des données d'entrée, ce qui les rend efficaces pour comprendre les informations visuelles. En vision par ordinateur, les transformateurs traitent les images en les segmentant en patchs et en intégrant linéairement ces segments, permettant ainsi de puissants mécanismes d'attention pour des tâches telles que la classification d'images, la détection d'objets et la génération de descriptions d'images. En robotique, les transformateurs peuvent prédire les décalages d'erreur de l'extrémité d'un bras robotique en analysant des séquences d'images ou des données de capteurs, améliorant ainsi la précision et l'exactitude.

L'introduction du codage de position [22] dans la prédiction de l'erreur de décalage de translation et de l'erreur de décalage de pose de rotation à l'extrémité du bras robotique améliore encore la conscience spatiale et la précision. En intégrant les informations de position spatiale directement dans le modèle de transformateur, il améliore la capacité du modèle à discerner les positions relatives et absolues des objets dans une scène, ce qui est essentiel pour une estimation précise des erreurs.

Le mécanisme d'auto-attention des transformateurs peut capturer des dépendances à longue portée, permettant au modèle de calculer directement les relations entre deux éléments quelconques au sein d' une séquence de caractéristiques d'image d'entrée, en considérant toutes les données d'entrée plutôt que simplement les informations locales . Cette caractéristique permet de réduire le problème des optima locaux. De plus, grâce à une allocation d'attention fine, les modèles basés sur des transformateurs peuvent identifier et mettre en valeur les caractéristiques et les relations les plus pertinentes pour la tâche actuelle. Par conséquent, l'utilisation d'un modèle de transformateur équipé d'un codage de position pour prédire les erreurs de translation et de rotation entre l'effecteur final du robot de pollinisation et l'objet cible constitue une approche viable.

3. Méthode

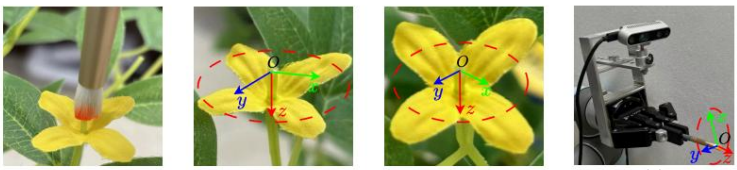
3.1. Erreur de

décalage Pour décrire l'erreur de décalage de translation et l'erreur de décalage de pose en rotation, deux systèmes de coordonnées cartésiennes, CA et CB, doivent être construits au niveau de la fleur cible et à

l'effecteur final du robot de pollinisation, respectivement. Le repère cartésien CA, construit sur la fleur, a son origine O à l'extrémité du pistil. Le plan formé par les axes x et y est parallèle au plan des pétales de fleur et l' axe z est parallèle au pistil, pointant vers l'intérieur, comme le montrent les figures 2b, c. Le système de coordonnées cartésiennes CB, construit à l'extrémité du robot de pollinisation, a son origine O à l'extrémité de la brosse, obtenue en traduisant le système de coordonnées TCP du bras robotique UR5, comme le montre la figure 2d. CA peut être dérivé de CB via une matrice de translation-rotation 4 × 4 TR :

$$TR = \begin{matrix} R11 & R12 & R13 & t1 \\ R21 & R22 & R23 & t2 \\ R31 & R32 & R33 & t3 \\ 0 & 0 & 0 & 1 \end{matrix} \quad (1)$$

$$CA = TR \cdot CB \quad (2)$$



(un) (b) (c) (d)

Figure 2. (a) La position idéale de pollinisation sans erreur. (b – d) Positions des systèmes de coordonnées cartésiennes sur la fleur et à l'extrémité du bras robotique, respectivement.

L'erreur de décalage de translation et l'erreur de décalage de pose de rotation sont respectivement le vecteur de translation t et le vecteur de rotation dérivé de la matrice de rotation R de la matrice de translation-rotation TR :

$$t = t1\ t2\ t3^{T} \quad (3)$$

$$R = \begin{matrix} R11 & R12 & R13 \\ R21 & R22 & R23 \\ R31 & R32 & R33 \end{matrix} \quad (4)$$

Lorsque l'effecteur terminal du robot de pollinisation se trouve dans la posture de pollinisation idéale, les valeurs de l'erreur de décalage de translation et de l'erreur de décalage de pose de rotation se rapprochent de zéro. A ce moment, la brosse à l'extrémité du robot de pollinisation est perpendiculaire au plan des pétales de la fleur cible et juste en contact avec le pistil, c'est-à-dire que les systèmes de coordonnées CA et CB coïncident, comme le montre la figure 2a.

$$\theta = \cos^{-1}\left(\frac{\text{trace}(R) - 1}{2}\right)$$

$$\vec{V} = \frac{1}{2\,\text{péché}(\theta)} \begin{matrix} R32 - R23 \\ R13 - R13 \\ R21 - R12 \end{matrix} \quad (5)$$

$$\vec{W} = \theta\,\vec{V}$$

L'erreur de décalage de translation et l'erreur de décalage de pose de rotation prédites par le modèle sont notées respectivement $\hat{t}$ et $\hat{R}$. Les différences entre l'erreur de décalage de translation réelle et prévue et l'erreur de décalage de pose en rotation depuis l'extrémité du bras robotique jusqu'au pistil de la fleur cible sont calculées respectivement en tant qu'erreur de translation (TE) et erreur de rotation (RE) . D'après l'équation (5), les matrices de rotation R et $\hat{R}$ peuvent être converties en vecteurs de rotation $\vec{W}$ et $\hat{\vec{W}}$:

$$TE = \hat{t} - t \quad (6)$$

$$RE = 1 - \frac{\vec{W}}{\vec{W}\ \vec{W}\ \vec{W}} \quad (7)$$

L'unité de TE est le centimètre (cm), représentant la distance spatiale entre l'erreur de translation prédite et l'erreur de translation réelle. RE est sans dimension, représentant la distance cosinusoïdale entre l'erreur vectorielle de rotation prédite et l' erreur vectorielle de rotation réelle.

3.2. Module Attention

Le module d'attention est implémenté sur la base du mécanisme d'auto-attention, qui permet au modèle de peser l'importance des différentes caractéristiques dans les données d'entrée.
En appliquant ce mécanisme, le modèle peut concentrer de manière adaptative les ressources de calcul sur les régions de l'image qui contiennent des informations clés pour prédire les décalages de pose (ΔTx, ΔTy, ΔTz, ΔWx, ΔWy, ΔWz). De plus, avec l'introduction du codage de position dans le mécanisme d'auto-attention, le modèle parvient à une compréhension globale de l'image. Cela fournit des informations supplémentaires pour résoudre les problèmes de symétrie dans la prédiction de la pose des objets, améliorant ainsi la précision du modèle et augmentant sa capacité générale dans divers scénarios rencontrés par les robots de pollinisation. Cette méthode améliore non seulement la précision du modèle mais renforce également sa polyvalence dans les diverses conditions rencontrées par les robots de pollinisation.

3.3. Extraction de caractéristiques

Compte tenu de la capacité unique des modèles de transformateur à traiter les données d'image, nous utilisons un réseau neuronal convolutionnel pré-entraîné, ResNet-50 [23], comme réseau d'extraction de caractéristiques, pour convertir les images originales en vecteurs de caractéristiques de haute dimension. Ces caractéristiques sont ensuite introduites dans le modèle de transformateur pour un traitement ultérieur. Nous comparons des réseaux d'extraction de caractéristiques de différentes profondeurs, pour déterminer la représentation optimale des caractéristiques. Cette approche exploite la forte capacité de ResNet-50 à capturer des hiérarchies spatiales détaillées dans les images, fournissant ainsi un riche ensemble de fonctionnalités à analyser par le modèle de transformateur. En incorporant cette architecture hybride, combinant les atouts des CNN en matière d'extraction de caractéristiques avec le mécanisme d'attention avancé des transformateurs, le modèle permet une compréhension nuancée du contenu de l'image pertinent pour prédire les erreurs de décalage de pose. Cette méthode facilite l'identification et la concentration sur les aspects cruciaux des données d'entrée, améliorant ainsi la précision et l'efficacité du processus d'estimation de pose.

3.4. Approche proposée

Ce travail présente un nouveau modèle de réseau intégrant un module d'attention, conçu pour améliorer la précision des prédictions en se concentrant sur les parties les plus critiques des données d'entrée. Ce modèle est particulièrement adapté à l'analyse des données d'image capturées par l'effecteur final d'un robot de pollinisation, visant à prédire avec précision le décalage de position de l'effecteur terminal par rapport à l'objet cible dans un système de coordonnées cartésiennes. Pendant l'entraînement, le modèle prend en entrée les données d'image de l'extrémité du bras robotique et génère un vecteur à six dimensions (ΔTx, ΔTy, ΔTz, ΔWx, ΔWy, ΔWz) représentant le décalage de translation $\vec{W}$. = (ΔWx, ΔWy, ΔWz) dans les erreurs ΔT = (ΔTx, ΔTy, ΔTz) et les erreurs de décalage de rotation trois directions du système de coordonnées cartésiennes.

3.5. Prédiction des erreurs

de décalage L'entrée du modèle est une image RVB capturée par une caméra RVB « œil dans la main » sur le bras robotique, illustrant l'état de position de l'effecteur terminal du bras robotique pendant la pollinisation et la fleur de pollinisation. La sortie est un vecteur représentant le vecteur de décalage de pose (ΔTx, ΔTy, ΔTz, ΔWx, ΔWy, ΔWz), qui inclut les erreurs de translation et de rotation de l'effecteur final du robot par rapport à l'objet cible. Pour y parvenir, le modèle utilise deux réseaux de neurones à action directe pour cartographier directement les caractéristiques centrées sur l'attention.

dans l'espace de décalage tridimensionnel et l'espace de rotation tridimensionnel. La figure 3 illustre l'ensemble du processus algorithmique.

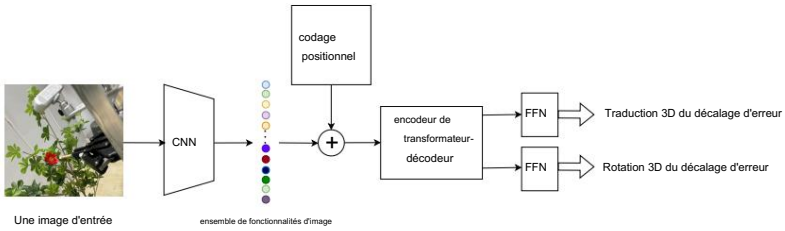


Figure 3. Le diagramme illustre le flux de travail de l'algorithme proposé. Le processus commence par l'image d'entrée, qui subit une extraction de caractéristiques via un réseau neuronal convolutif (CNN), pour obtenir des caractéristiques d'image de grande dimension. Ces caractéristiques sont ensuite complétées par un codage de position avant d'être entrées dans le module transformateur. La sortie du transformateur est ensuite traitée par deux réseaux neuronaux à action directe distincts, chargés respectivement de prédire les erreurs de translation et de rotation.

3.6. Architecture de réseau

Dans ce travail, nous avons adopté un modèle de transformateur personnalisé conçu pour traiter les caractéristiques de l'image sérialisée et prédire l'erreur de décalage de translation et l'erreur de décalage de pose en rotation à l'extrémité d'un bras robotique. Nous avons modifié le modèle de transformateur original en introduisant un codage de position bidimensionnel, afin de préserver les informations spatiales des images d'entrée. La couche de sortie est personnalisée pour générer un décalage de translation 3D et un vecteur d'erreur de décalage de pose en rotation. Comme le montre la figure 4, un ResNet50 sert de base pour extraire un riche ensemble de fonctionnalités à partir des données d'image d'entrée. Pour conserver les informations spatiales parmi les éléments des images, nous appliquons un codage de position sinusoïdale aux caractéristiques extraites, qui sont ensuite additionnées aux caractéristiques avant d'être introduites dans l'encode

Pour capturer des informations provenant de différents sous-espaces, un mécanisme d'attention multi-têtes à cinq têtes est utilisé dans l'encodeur. La sortie du codeur est ensuite introduite dans le décodeur. Suivant l'architecture standard du transformateur, le décodeur utilise un mécanisme d'attention multi-têtes pour transformer deux intégrations de taille d. Ces objets de requête sont transformés par le décodeur en intégrations de sortie, qui sont ensuite décodées par deux réseaux de rétroaction distincts en erreur de décalage de translation et en erreur de décalage de pose de rotation, respectivement.

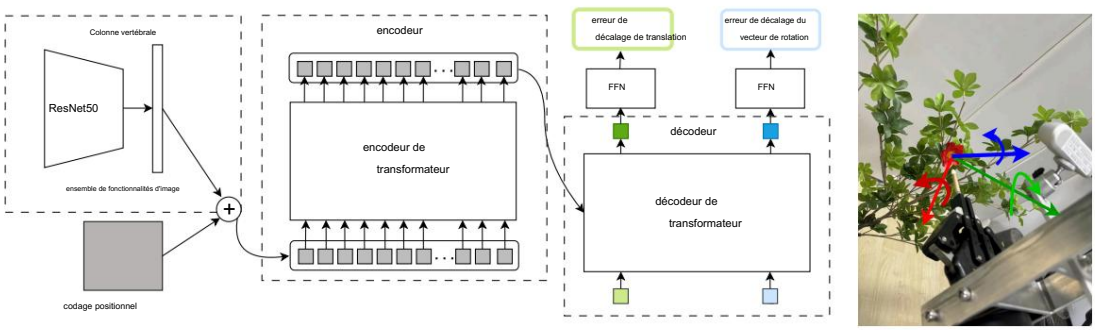


Figure 4. Le modèle final utilise ResNet50 comme réseau fédérateur pour apprendre les fonctionnalités de haut niveau à partir de l'image d'entrée. Ces fonctionnalités sont ensuite complétées par un codage positionnel avant d'être transmises au codeur. Par la suite, le décodeur génère d'abord un vecteur caractéristique pour l'erreur de décalage de translation, qui est utilisé comme entrée pour un réseau neuronal à action directe conçu pour la prédiction. Ce vecteur de caractéristiques est ensuite utilisé comme entrée de requête pour le décodeur, ce qui donne lieu à un autre vecteur de caractéristiques qui est introduit dans un réseau neuronal à action directe distinct pour prédire l'erreur de décalage de pose en rotatio

4. Expériences

4.1. Prétraitement des données

Les données d'image utilisées dans cette expérience proviennent toutes d'un ensemble de données propriétaire, qui se compose d'images capturées par une caméra montée sur le bras robotique pendant la pollinisation processus, où chaque image représente de manière unique la pose d'une fleur à polliniser. Comme le montre la figure 5, l'orientation des fleurs par rapport à l'extrémité du bras robotique a été classé en cinq directions : gauche, droite, haut, bas et avant. Détaillé des informations statistiques sont disponibles dans le tableau 1.

Figure 5. Fleurs avec différentes orientations par rapport à l'effecteur final du bras robotique.

Tableau 1. L'ensemble de données comprend le nombre d'images de fleurs orientées dans différentes directions par rapport à l'effecteur final du bras robotique.

Orientations	Nombre de fleurs	Proportion (%)
L	261	20.6
R.	274	21.7
F	230	18.2
U	256	20.3
D	243	19.2

« F », « L », « R », « U » et « D » représentent les orientations avant, gauche, droite, vers le haut et vers le bas.

Pour garantir la cohérence des données d'entrée, toutes les images capturées par le bras robotique la caméra finale a d'abord été redimensionnée à une résolution uniforme de (224 × 224). Par la suite, les images ont été normalisées, en mettant les valeurs de pixels à l' échelle [0, 1] , pour améliorer la stabilité de la formation du modèle. De plus, une série de techniques d'augmentation des données, notamment la mise à l'échelle, le recadrage et la transformation des couleurs, ont été appliquées pour augmenter la diversité des données et éviter le surajustement. Concernant les données de l'étiquette, nous avons supposé la position hors-l'erreur de réglage $\Delta T = (\Delta Tx, \Delta Ty, \Delta Tz)$ et l'erreur de décalage de rotation $\Delta -\vec{W} = (\Delta Wx, \Delta Wy, \Delta Wz)$, avec $||\Delta T|| < D$, où D était une constante. L'équation d'erreur de décalage de position (8) et Erreur de décalage de rotation L'équation (9) a été normalisée et mise à l'échelle à [−1, 1], comme suit :

$$\Delta Tn = \frac{\Delta T}{D} \tag{8}$$

$$-\vec{W} = \theta\ -\vec{V} \tag{9}$$

$$\theta = \Delta W2_x + \Delta W2_{oui} + \Delta W2_z \tag{dix}$$

$$-\vec{V} = (\frac{\Delta Wx}{\theta}, \frac{\Delta Wy}{\theta}, \frac{\Delta Wz}{)\theta} \tag{11}$$

Le symbole θ dans l'équation (10) représente l'angle de rotation et $-\vec{V}$ désigne le vecteur unitaire le long de l'axe de rotation. Les données d'étiquette normalisées étaient $(\Delta Tn, -\vec{V}, \frac{\theta}{2\pi})$.

4.2. Paramètres d'évaluation

Dans le modèle présenté dans cet article, nous prédisons séparément le décalage translationnel l'erreur et l'erreur de décalage de pose de rotation, concevant ainsi deux fonctions de perte. La première défaite

La fonction, nommée LossT, mesure la distance quadratique moyenne entre l' erreur de décalage de position spatiale prédite par le modèle pour l'effecteur final du bras robotique de pollinisation et le pistil de la fleur cible et l'erreur de décalage spatial réelle. LossT est défini comme suit :

$$\text{PerteT} = \frac{1}{2m_x} \sum_{M} (\hat{T}_n - T_n)^2 \quad (12)$$

où M est l'ensemble de l'ensemble de données de test, m est le nombre d'éléments dans l'ensemble et $\hat{T}_n$ T_n et sont respectivement l'erreur de décalage de translation prédite par le modèle et la véritable erreur de décalage de translation obtenue via l'équation (8).

La deuxième fonction de perte, nommée LossR, mesure l'écart entre l' erreur de rotation spatiale prédite par le modèle pour l'effecteur final du bras robotique de pollinisation par rapport au pistil de la fleur cible et l'erreur de rotation spatiale réelle. LossR est défini comme suit :

$$\text{PerteR} = \frac{1}{2m_x} \sum_{M} \left(\frac{1}{\sigma_1^2} \left(\hat{\theta} - \theta \right)^2 + \frac{1}{\sigma_2^2} \left(1 - \frac{\vec{V} \cdot \vec{\hat{V}}}{\vec{V}\,\vec{\hat{V}}}\right) + \log(\sigma_1 \sigma_2)\right) \quad (13)$$

où M est l'ensemble de l'ensemble de données de test et m est le nombre d'éléments dans l'ensemble. Variables $\hat{\theta}$, θ, $\vec{V}$, et $\vec{\hat{V}}$ représentent les angles et axes de rotation prédits et réels obtenus via équations (9) – (11), σ_1 et σ_2 étant les paramètres qui doivent être appris.

La fonction de perte combinée utilisée pour la formation du modèle est définie comme

$$\text{Perte} = \alpha \text{LossT} + \beta \text{LossR} \quad (14)$$

Cette fonction de perte dans l'équation (14) mesure de manière exhaustive la perte d'erreur de décalage de translation et la perte d'erreur de décalage de rotation pendant l'entraînement du modèle. Les paramètres α et β sont des hyperparamètres représentant des poids qui doivent être systématiquement ajustés lors de la formation du modèle.

4.3. Détails de la

formation La formation du modèle a été réalisée dans un environnement informatique équipé de GPU NVIDIA V100. Nous avons utilisé l'optimiseur Adam [24], avec le taux d'apprentissage initial fixé à 0,01, et nous avons utilisé une stratégie de décroissance du taux d'apprentissage qui réduisait progressivement le taux d'apprentissage à 0,00001 au fur et à mesure de la progression de la formation. Une fonction de perte personnalisée, l'équation (14), a été utilisée pendant le processus de formation. Lors de la formation du modèle, le réglage des hyperparamètres dans l'équation (14) a affecté la capacité du modèle à converger. Après des expériences approfondies, nous avons finalement fixé α = 0,0025 et β = 1, ce qui a permis au modèle de converger plus facilement pendant l'entraînement.

Dans l'ensemble de données, des fleurs d'orientations différentes ont été mélangées au hasard, puis les données ont été divisées en 10 sous-ensembles, en utilisant la méthode de validation croisée K-fold. À chaque fois, un sous-ensemble a été utilisé comme ensemble de test et les neuf sous-ensembles restants ont été utilisés comme ensemble de formation. Ce processus a été répété 10 fois. Lors de la formation du modèle, nous avons observé que le modèle convergeait initialement rapidement puis se stabilisait progressivement. La figure 6 montre les changements dans les valeurs de perte, d'erreur de translation et d'erreur de rotation au cours du processus de formation du modèle :

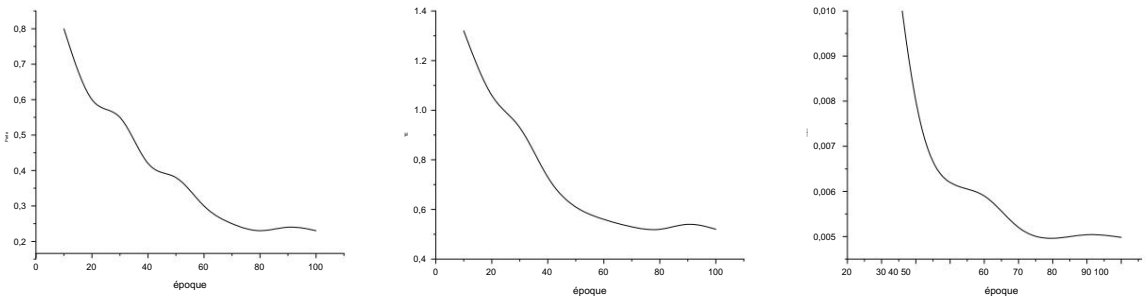


Figure 6. Modifications de la perte, du TE et du RE au cours du processus de formation du modèle.

4.4. Résultats

Au meilleur de nos connaissances, nous sommes les premiers à prédire l'erreur de décalage de translation et une erreur de décalage de pose en rotation d'un effecteur terminal d'un bras robotique par rapport à un objet cible à l'aide uniquement les informations sur l'image RVB. Par conséquent, cet article se concentre sur les améliorations de la précision de l'erreur de décalage de translation et de l'erreur de décalage de pose en rotation de l'effecteur terminal du robot lors de l'utilisation de notre méthode proposée par rapport à la méthode YANG de l'état de l'art [18] comme la ligne de base. De plus, nous soulignons l'amélioration de l'efficacité de la pollinisation d'un fleur unique.

Nous avons mené des expériences sur des fleurs d'orientations différentes en groupes, pour analyser erreurs de décalage de translation et de rotation ainsi que la vitesse de détection. Les résultats expérimentaux a montré de légères variations pour les fleurs avec des orientations différentes. Comme le montrent le tableau 2 et Figure 7, les fleurs tournées vers l'avant ont obtenu les meilleurs résultats, en termes de précision expérimentale, avec la plus petite erreur de décalage de translation et d'erreur de décalage de pose en rotation par rapport à fleurs dans d'autres orientations. Viennent ensuite les fleurs tournées vers le haut. En raison de l'environnement symétrie, il n'y avait presque aucune différence dans les résultats pour les fleurs orientées vers la gauche et vers la droite. Fleurs face vers le bas a eu les pires résultats, à la fois en termes d'erreur de décalage de translation et erreur de décalage de pose de rotation. Cependant, la vitesse de détection du modèle pour les fleurs avec les différentes orientations sont restées presque constantes.

Tableau 2. Résultats expérimentaux du modèle sur des fleurs avec différentes orientations.

Orientations	TE	CONCERNANT	FPS
L	0,82	0,0049	41
R.	0,82	0,0049	42
F	0,78	0,0047	43
U	0,80	0,0048	42
D	0,87	0,0051	41

« F », « L », « R », « U » et « D » représentent les orientations avant, gauche, droite, vers le haut et vers le bas.

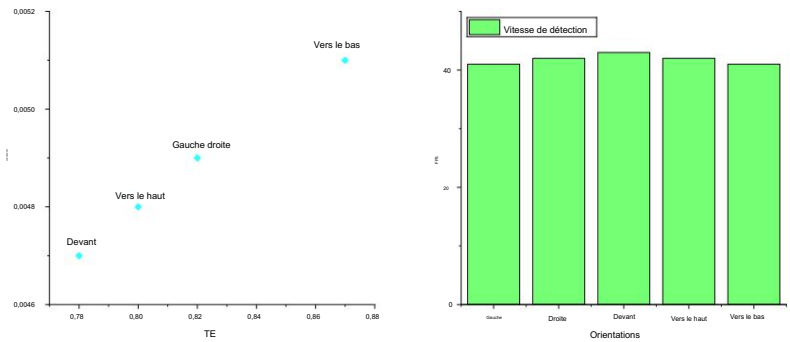


Figure 7. Distribution de la détection des erreurs de rotation et de translation du modèle pour les fleurs avec différentes orientations, ainsi que la vitesse de détection.

Nous avons mené nos expériences à l'aide du robot de pollinisation YANG. Le processus de pollinisation d'une seule fleur du robot de pollinisation YANG peut être divisé en cinq étapes.

ordre chronologique. Après la quatrième étape, la précision de positionnement de l' effecteur terminal du robot de pollinisation YANG par rapport au pistil cible était de 1,5 cm. À ce stade, nous avons introduit une nouvelle étape appelée « Position de réglage fin ». Nous saisissons l'image de la pose de la fleur par rapport à l'effecteur final du robot dans notre modèle entraîné, pour prédire la translation et erreurs de décalage de rotation. Sur la base des valeurs prédites, la pose du robot de pollinisation l'effecteur final a été ajusté.

Comme le montre le tableau 3, grâce à la nouvelle étape « Ajustement fin de la position », la distance de translation et l'écart de rotation entre l'effecteur terminal du robot de pollinisation et les objectifs de pollinisation ont été encore réduits, réduisant ainsi la plage pour la prochaine recherche d'asservissement étape de pollinisation. Nos expériences ont montré que le temps moyen de pollinisation par recherche asservie n'était que de 3,1 s, ce qui permet d'obtenir un taux de réussite de pollinisation comparable à celui de YANG. robot à 86,19%. Comme le montre le tableau 4, après avoir appliqué notre méthode au robot de pollinisation, la précision de distance moyenne entre l'effecteur terminal du robot et le pistil cible atteinte 0,81 cm, soit une amélioration de 46,67%. L'erreur de décalage de pose de rotation calculée selon à l'équation (14) a également atteint 0,0049. Le temps total pour terminer la pollinisation d'un seul la fleur a été réduite de près de moitié, avec une amélioration moyenne de l'efficacité de 50,9 %.

Tableau 3. Comparaison du coût moyen en temps pour chaque étape du système de pollinisation après intégration de notre méthode.

Étape	YANG (référence) Coût en temps (S)	Notre coût en temps (S)
Détection des fleurs	0,0928	0,0927
Identification des pistils	0,025	0,024
Calcul de position (mouvement du robot inclus)	1.8	1.8
Fleur atteignant (mouvement du robot inclus)	4.2	4.2
Position de réglage fin	/	0,024
Asservissement (mouvement du robot inclus)	12.7	3.1

L'étape « Position de réglage fin » est une étape supplémentaire. Avec cette étape incluse, le temps passé dans le « Serving » Le pas est considérablement réduit, atteignant le même taux de réussite de pollinisation en seulement 3,1 s. Tous les numéros sont enregistrés en secondes.

Tableau 4. Comparaison de l'erreur de décalage de translation et de l'erreur de décalage de pose en rotation lors de l'application de notre algorithme aux fleurs orientées dans des directions différentes par rapport à la ligne de base.

Orientations	Méthode	TE	CONCERNANT	Temps Coût (s)
L	YANG	1,53	/	18h78
	Notre	0,82	0,0050	9.22
R.	YANG	1,52	/	18h80
	Notre	0,82	0,0050	9.23
F	YANG	1,48	/	18h85
	Notre	0,80	0,0048	9.22
U	YANG	1,53	/	18.79
	Notre	0,81	0,0049	9.24
D	YANG	1,55	/	18.83
	Notre	0,83	0,0052	9.26

« F », « L », « R », « U » et « D » représentent les orientations avant, gauche, droite, vers le haut et vers le bas. La méthode de YANG a servi de base de référence.

Cet article a mené plusieurs expériences d'ablation, pour vérifier l'impact de différents et l'ajout d'un codage de position sur la prédiction du modèle de l' erreur de décalage de translation et de l'erreur de décalage de pose en rotation. ResNet50, ResNet18, ResNet101, VGG16, VGG19, DenseNet-121 et DenseNet-201 ont été sélectionnés comme fonctionnalité de base réseaux d'extraction. Pour chaque backbone, des expériences ont été menées avec et sans codage positionnel. Les résultats expérimentaux, présentés dans le tableau 5 et la figure 8, indiquent que différents squelettes et l'ajout d'un codage de position affectent de manière significative la

la précision de la prédiction finale du modèle. La version du modèle avec ResNet50 comme épine dorsale
Le réseau d'extraction de fonctionnalités et l'encodage de position ajoutés ont permis d'obtenir les meilleures performances dans la prédiction de l'erreur de décalage de translation et de l'erreur de décalage de pose en rotation, atteignant 8,1 mm et 0,0049, respectivement.

Tableau 5. L'impact des différents piliers du modèle sur l'exactitude de la prévision du l'erreur de décalage de translation et l'erreur de décalage de pose de rotation.

Colonnes vertébrales	Encodage positionnel	TE	
ResNet50		0,81	0,0049
		1.19	0,0051
ResNet18		0,92	0,0053
		1,33	0,0054
ResNet101		0,86	0,0050
		1.21	0,0050
VGG16		0,96	0,0053
		1.32	0,0055
VGG19		0,99	0,0054
		1,33	0,0055
DenseNet-121		0,92	0,0054
		1.23	0,0055
DenseNet-201		1.10	0,0055
		1,38	0,0057

Une coche () indique que le modèle incluait un codage de position, tandis qu'une croix () indique que le modèle le modèle n'incluait pas le codage de position.

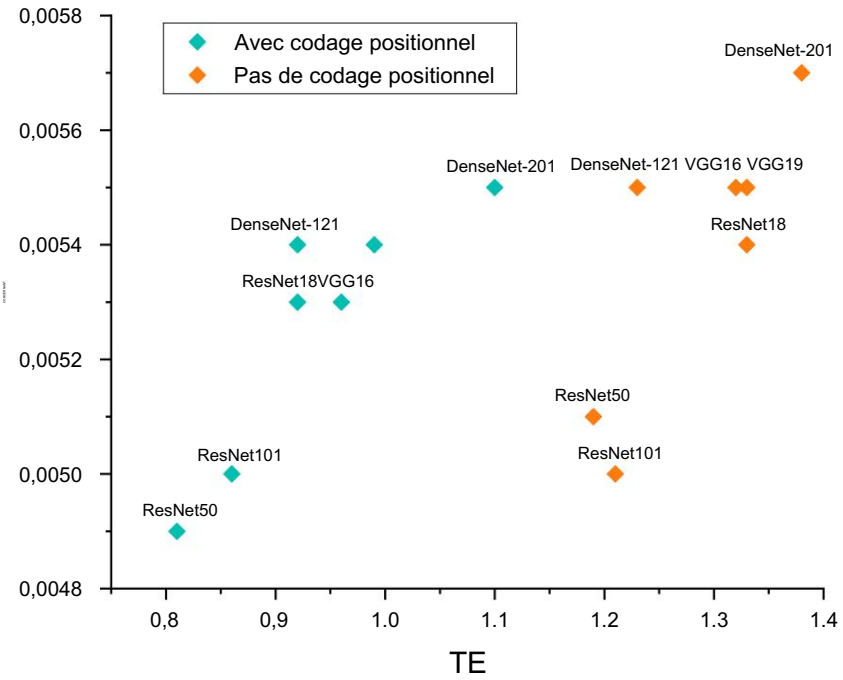


Figure 8. Répartition des erreurs de translation et de rotation pour différents modèles de base avec et sans l'ajout d'un codage de position.

5. Discussion

Ce travail propose une approche basée sur un transformateur qui permet d'obtenir une prédiction de bout en bout d'erreurs de translation et de rotation entre l'effecteur terminal du bras robotique de pollinisation

et la position de pollinisation cible, en utilisant uniquement les informations d'image RVB. Nos résultats expérimentaux démontrent que cette méthode réduit efficacement la plage d'erreur de l' effecteur final robotique dans une marge d'erreur connue, améliorant ainsi l'efficacité globale de la pollinisation robotique.

Dans ce travail, notre méthode a montré de légères variations dans les résultats lorsqu'il s'agissait de fleurs d'orientations différentes, en particulier des erreurs de translation et de rotation plus importantes avec des fleurs orientées vers le bas. Cela peut être attribué au positionnement relatif de l'angle de la caméra du robot par rapport à l'orientation des fleurs, ce qui complique la reconnaissance et la localisation précises des fleurs dans des orientations spécifiques. Grâce à l'expérimentation, nous avons également observé que les versions du modèle utilisant le codage de position étaient plus performantes pour prédire avec précision les erreurs de translation et de rotation. De plus, les différents réseaux fédérateurs utilisés pour extraire les caractéristiques des images d'entrée ont eu un impact significatif sur les performances du modèle.

Ce travail présente des limites potentielles malgré l'efficacité démontrée de la méthode proposée pour réduire les erreurs de translation et de rotation entre l'effecteur final du robot de pollinisation et la cible : (1) L'ensemble de données utilisé dans cette étude a été spécifiquement collecté à des fins expérimentales, ce qui peut limiter les capacités de généralisation du modèle à travers des environnements et des types de fleurs variés ; (2) Les performances du modèle peuvent être compromises dans différentes conditions d'éclairage et dans des environnements obstrués, ce qui a un impact sur son efficacité globale. Ces problèmes soulignent la nécessité d'améliorations supplémentaires avant le déploiement pratique, ce qui nécessite des recherches futures pour explorer des méthodes supplémentaires améliorant l'adaptabilité et la fiabilité du modèle dans divers environnements agr

6. Conclusions

Ce travail présente une méthode innovante qui utilise les puissantes capacités d'apprentissage et de compréhension spatiales d'un modèle d'apprentissage profond basé sur un transformateur pour réaliser une prédiction de bout en bout des erreurs de translation et de rotation entre l' effecteur final du robot de pollinisation et la position de pollinisation cible en utilisant uniquement RVB. images. Nos résultats expérimentaux démontrent que cette méthode est efficace pour réduire davantage la plage d'erreur de l'effecteur final du robot de pollinisation dans une plage d'erreur connue, améliorant ainsi l'efficacité globale du robot de pollinisation. Les travaux futurs pourraient se concentrer sur l'étude de la prédiction des erreurs de translation et de rotation entre l'effecteur final robotique et les positions cibles au sein d'ensembles de données contenant une gamme plus diversifiée de types de fleurs, dans diverses conditions d'éclairage et d'occlusions. De tels travaux viseraient à améliorer la robustesse et les capacités de généralisation du modèle, fournissant ainsi une approche réalisable pour améliorer la précision des effecteurs terminaux robotiques génériques.

Contributions des auteurs : Conceptualisation, JX ; méthodologie, JX; validation, JX ; analyse formelle, JX, JC, HP et MY ; enquête, JX; ressources, JX ; conservation des données, JX ; rédaction : préparation du brouillon original, JX, HP et MY ; rédaction – révision et édition, JX, HP et MY ; visualisation, JX ; supervision, HP, MY et JC ; administration du projet, JC Tous les auteurs ont lu et accepté la version publiée du manuscrit.

Financement : Ce travail a été financé par le projet de plan de R&D clé du Guangxi (AB24010164 ; AB21220038).

Déclaration de disponibilité des données : Les données qui étayent les conclusions de ce travail sont disponibles auprès de l'auteur correspondant sur demande raisonnable.

Conflits d'intérêts : Les auteurs ne déclarent aucun conflit d'intérêts.

Abréviations

Les abréviations suivantes sont utilisées dans ce manuscrit :

Erreur de traduction TE
Erreur de rotation RE

Les références

1. Binns, C. Les insectes robotiques pourraient polliniser les fleurs et trouver des victimes de catastrophes. Science populaire, 17 décembre 2009.
2. Williams, H. ; Néjati, M. ; Hussein, S. ; Penhall, N. ; Lim, JY ; Jones, MH; Bell, J. ; Ahn, HS; Bradley, S. ; Schaare, P. ; et coll. Pollinisation autonome des fleurs individuelles de kiwi : vers un pollinisateur robotisé de kiwi. Robot J. Field. 2020, 37, 246-262. [Référence croisée]
3. Gao, C. ; Lui, L. ; Croc, W. ; Wu, Z. ; Jiang, H. ; Li, R. ; Fu, L. Un nouveau robot de pollinisation des fleurs de kiwi basé sur la préférence sélection des fleurs et cible précise. Calculer. Électron. Agricole. 2023, 207, 107762. [Réf. croisée]
4. Strader, J. ; Nguyen, J. ; Tatsch, C. ; Du, Y. ; Lassak, K. ; Buzzo, B. ; Watson, R. ; Cerbone, H. ; Ohi, N. ; Yang, C. ; et coll. Sous-système d'interaction avec les fleurs pour un robot de pollinisation de précision. Dans les actes de la conférence internationale IEEE/RSJ 2019 sur les robots et systèmes intelligents (IROS), Macao, Chine, 3-8 novembre 2019 ; pages 5534 à 5541.
5. Shaneyfelt, T. ; Jamshidi, MM; Agaian, S. Un système de grue d'amarrage robotique à retour de vision avec application à la pollinisation de la vanille. Int. J.Autom. Contrôle 2013, 7, 62-82. [Référence croisée]
6. Yuan, T. ; Zhang, S. ; Sheng, X. ; Wang, D. ; Gong, Y. ; Li, W. Un robot de pollinisation autonome pour le traitement hormonal des fleurs de tomates en serre. Dans Actes de la 3e Conférence internationale sur les systèmes et l'informatique (ICSAI) 2016, Shanghai, Chine, 19-21 novembre 2016 ; pp. 108-113.
7. Abrol, DP Pollinisation Biologie : Conservation de la biodiversité et production agricole : Volume 792 ; Springer : New York, NY, États-Unis, 2012.
8. Yang, X. ; Miyako, E. Pollinisation par bulles de savon. Iscience 2020, 23, 101188. [CrossRef] [Pub Med] 9. Hinterstoisser, S. ; Cagniart, C. ; Ilic, S. ; Sturm, P. ; Navab, N. ; Fua, P. ; Lepetit, V. Cartes de réponse dégradées pour la détection en temps réel d' objets sans texture. IEEETrans. Modèle Anal. Mach. Intell. 2011, 34, 876-888. [Référence croisée] [Pub Med]
10. Cao, Z. ; Cheikh, Y. ; Banerjee, NK Estimation de pose 6dof évolutive en temps réel pour les objets sans texture. Dans les actes de la conférence internationale IEEE 2016 sur la robotique et l'automatisation (ICRA), Stockholm, Suède, 16-21 mai 2016 ; pages 2441 à 2448.
11. Brachmann, E. ; Krull, A. ; Michel, F. ; Gumhold, S. ; Shotton, J. ; Rother, C. Apprentissage de l'estimation de la pose d'objets 6D à l'aide des coordonnées d'objets 3D. In Proceedings, Part II 13, Actes de Computer Vision–ECCV 2014 : 13e Conférence européenne, Zurich, Suisse, 6-12 septembre 2014 ; Springer : Berlin/Heidelberg, Allemagne, 2014 ; pp. 536-551.
12. Krull, A. ; Brachmann, E. ; Michel, F. ; Yang, MON ; Gumhold, S. ; Rother, C. Apprentissage de l'analyse par synthèse pour l'estimation de pose 6D dans des images RVB-D. Dans les actes de la conférence internationale de l'IEEE sur la vision par ordinateur, Santiago, Chili, 7-13 décembre 2015 ; pp. 954-962.
13. Xiang, Y. ; Schmidt, T. ; Narayanan, V. ; Fox, D. Posecnn : Un réseau neuronal convolutif pour l'estimation de la pose d'objets 6D dans des scènes encombrées. arXiv 2017, arXiv : 1711.00199.
14. Li, K. ; Huo, Y. ; Liu, Y. ; Shi, Y. ; Lui, Z. ; Cui, Y. Conception d'un bras robotique léger pour la pollinisation des kiwis. Calculer. Électron. Agricole. 2022, 198, 107114. [Réf. croisée]
15. Ma, WP ; Li, WX; Cao, PX Méthode de positionnement d'objets par vision binoculaire pour robots basée sur une correspondance stéréo grossière-fine. Int. J. Automatique. Calculer. 2020, 17, 562-571. [Référence croisée]
16. Tai, ND; Trieu, Nouveau-Mexique ; Thinh, NT Modélisation des positions et orientations des fleurs de cantaloup pour une pollinisation automatique. Agriculture 2024, 14, 746. [Réf. croisée]
17. Ahmad, K. ; Park, JE-E.; Ilyas, T. ; Lee, J.-H. ; Kim, S. ; Kim, H. Pollinisations précises et robustes pour les pastèques grâce à l'intelligence asservissement visuel guidé. Calculer. Électron. Agricole. 2024, 219, 108753. [Réf. croisée]
18. Yang, M. ; Lyu, H. ; Zhao, Y. ; Soleil, Y. ; Poêle, H. ; Soleil, Q. ; Chen, J. ; Qiang, B. ; Yang, H. Livraison de pollen aux pistils de fleurs de forsythia de manière autonome et précise à l'aide d'un bras robot. Calculer. Électron. Agricole. 2023, 214, 108274. [Réf. croisée]
19. Zhou, C. Yolact++ Meilleure segmentation des instances en temps réel ; Université de Californie : Davis, Californie, États-Unis, 2020.
20. Carion, N. ; Massa, F. ; Synnaève, G. ; Usunier, N. ; Kirillov, A. ; Zagoruyko, S. Détection d'objets de bout en bout avec transformateurs. Dans Conférence européenne sur la vision par ordinateur ; Springer : Cham, Suisse, 2020 ; pp. 213-229.
21. Vaswani, A. ; Shazeer, N. ; Parmar, N. ; Uszkoreit, J. ; Jones, L. ; Gomez, AN; Kaiser, Ł.; Polosukhin, I. L'attention est tout ce dont vous avez besoin. arXiv 2017, arXiv:1706.03762v7.
22. Jiang, K. ; Peng, P. ; Lian, Y. ; Xu, W. La méthode de codage des intégrations de position dans le transformateur de vision. J. Vis. Commun. L'image représente. 2022, 89, 103664. [Réf. croisée]
23. Lui, K. ; Zhang, X. ; Ren, S. ; Sun, J. Apprentissage résiduel profond pour la reconnaissance d'images. Dans Actes de la conférence IEEE sur la vision par ordinateur et la reconnaissance de formes, Las Vegas, NV, États-Unis, 27-30 juin 2016 ; pp. 770-778.
24. Kingma, DP; Ba, J. Adam : Une méthode d'optimisation stochastique. arXiv 2014, arXiv : 1412.6980.

électronique

MDPI

Article

Filtres de corrélation adaptatifs à aberrance réprimée avec Optimisation coopérative dans les véhicules sans pilote de grande dimension Répartition des tâches des véhicules aériens et planification de la trajectoire

Zijie Zheng [1,*], Zhijun Zhang [1], Zhenzhang Li [2], Qiuda Yu [1] et Ya Jiang [1]

[1] École des sciences et de l'ingénierie de l'automatisation, Université de technologie de Chine du Sud, Guangzhou 510006,
[2] École chinoise de mathématiques et de sciences des systèmes, Université normale polytechnique du Guangdong, Guangzhou 510665, Chine ; zhenzhangli@gpnu.edu.cn
* Correspondance : 201910102802@mail.scut.edu.cn

Résumé : Dans le domaine en évolution rapide des applications des véhicules aériens sans pilote (UAV), la complexité de la planification des tâches et de l'optimisation de la trajectoire, en particulier dans les environnements opérationnels de grande dimension, est de plus en plus difficile. Cette étude répond à ces défis en développant l' algorithme d'optimisation coopérative du filtre de corrélation de suppression de distorsion adaptative (ARCF-ICO), conçu pour l'allocation de tâches et la planification de trajectoire de drones de grande dimension. L'algorithme ARCF-ICO combine des technologies avancées de filtrage de corrélation avec des techniques d'optimisation multi-objectifs, améliorant ainsi la précision de la planification des trajectoires et l'efficacité de l'allocation des tâches. En intégrant les conditions météorologiques et d'autres facteurs environnementaux, l'algorithme garantit des performances robustes à basse altitude. L'algorithme ARCF-ICO améliore la stabilité et la précision du suivi des drones en supprimant les distorsions, facilitant ainsi la sélection optimale de la trajectoire et l'exécution des tâches. La validation expérimentale à l'aide des ensembles de données UAV123@10fps et OTB-100 démontre que l'algorithme ARCF-ICO surpasse les méthodes existantes dans les métriques Area Under the Curve (AUC) et Precision. De plus, la prise en compte par l'algorithme de la consommation et de l'endurance de la batterie valide en outre son applicabilité aux technologies actuelles d'UAV. Cette recherche fait progresser la planification des missions des drones et établit de nouvelles normes pour le déploiement des drones dans les applications civiles et militaires, où l'adapta

Mots-clés : optimisation multi-objectifs ; Planification de trajectoire de drone ; filtres de corrélation ; algorithmes adaptatifs ; répartition des tâches

Citation : Zheng, Z. ; Zhang, Z. ; Li, Z. ; Yu, Q. ; Jiang, Y. Filtres de corrélation adaptatifs à aberrance réprimée avec optimisation coopérative dans l'attribution des tâches et la planification des trajectoires des véhicules aériens sans pilote de grande dimension . Électronique 2024, 13, 3071. https://doi.org/10.3390/électronique13153071

Rédacteur académique : Zhiquan Liu

Reçu : 29 mai 2024
Révisé : 20 juillet 2024
Accepté : 30 juillet 2024
Publié : 2 août 2024

1. Introduction

À l'ère de l'Internet des objets, les véhicules aériens sans pilote (UAV) sont devenus des outils essentiels dans divers domaines en raison de leur flexibilité et de leur efficacité. À mesure que la complexité et la diversité des missions des drones augmentent, le besoin d' une allocation avancée de tâches multi-objectifs et d'une planification de trajectoire de grande dimension devient critique. Des études récentes de Chen et al. [1–3] ont introduit des algorithmes innovants qui améliorent la planification de trajectoire et le contrôle coopératif du comportement des drones hétérogènes. Ces méthodes améliorent l'efficacité opérationnelle et l'adaptabilité des drones dans des environnements dynamiques et complexes, répondant ainsi aux défis contemporains du déploiement des drones.

Le développement rapide de la technologie des drones a conduit à une demande croissante d' applications de drones dans des environnements multitâches. Cependant, dans des environnements multi-objectifs complexes , un seul drone a souvent du mal à gérer plusieurs tâches simultanément, ce qui nécessite la collaboration de plusieurs drones. La recherche sur l'allocation de tâches multi-objectifs et de grande dimension des drones et la planification de trajectoire basée sur l'apprentissage profond vise à relever les défis d'optimisation dans l'allocation de tâches collaboratives multi-UAV, en réalisant une gestion et une planification intelligentes des drones pour améliorer l'efficacité et la précision de l'exécution des tâches [4].

Dans le domaine du développement stratégique national, la progression et l'utilisation de la technologie des véhicules aériens sans pilote (UAV) apparaissent comme des catalyseurs technologiques essentiels pour les initiatives stratégiques. Grâce au raffinement de la planification des tâches des drones, des secteurs clés au sein d'un pays connaîtront des niveaux de productivité et d'efficacité accrus, favorisant ainsi la culture d'une productivité nouvelle et de haute qualité, un concept défendu par Yao et Liu [5]. Leurs travaux soulignent l'importance de définir des cadres théoriques et des voies pratiques pour stimuler le progrès. De plus, l'évolution de la technologie des drones augmente non seulement les capacités technologiques d'un pays, mais renforce également sa compétitivité mondiale, comme l'ont souligné Ma et Chen [6]. Leur exploration des transitions nationales souligne le rôle incontestable de l'innovation technologique dans l'élaboration des trajectoires économiques et stratégiques. La technologie des drones constituant une innovation fondamentale, ses applications multiformes dans des domaines tels que la défense nationale, les transports et l'agriculture contribuent de manière significative à améliorer la compétitivité globale d'un pays. Ainsi, en alignant la technologie des drones sur des impératifs stratégiques plus larges, les aspirations à faire progresser le progrès et à renforcer la compétitivité nationale, telles qu'épouses par le nouveau paradigme de développement, sont sur le point de se réaliser.

Les modèles d'apprentissage profond ont joué un rôle important dans l'attribution des tâches et la planification de la trajectoire des drones. Les modèles d'apprentissage profond courants comprennent les réseaux de neurones convolutifs (CNN) [7], les réseaux de neurones récurrents (RNN) [8], l'apprentissage par renforcement profond (DRL) [9], les réseaux contradictoires génératifs (GAN) [10] et l'apprentissage par transfert, chacun avec ses avantages et ses limites. Par exemple, les CNN conviennent aux tâches de traitement d'images mais manquent d'efficacité dans la gestion des données séquentielles [11], tandis que les RNN peuvent traiter des données séquentielles mais souffrent de problèmes tels que la disparition et l'explosion des gradients. DRL peut gérer des tâches avec des récompenses différées mais implique un processus de formation complexe, tandis que les GAN peuvent générer des données mais présentent une instabilité pendant la formation. L'apprentissage par transfert peut exploiter les connaissances acquises précédemment pour accélérer l'apprentissage de nouvelles tâches, mais nécessite de prendre en compte les différences entre les

Cette étude vise à résoudre les problèmes d'attribution de tâches et de planification de trajectoire d'UAV multi-objectifs de grande dimension à l'aide d'un filtre de corrélation de suppression de distorsion adaptative (ARCF). Nous développerons un réseau ARCF qui intègre les états, les actions et les fonctions de récompense des missions de drones pour permettre une prise de décision intelligente en matière d'attribution de tâches et de planification de trajectoire. Plus précisément, nous utiliserons des techniques d'apprentissage par renforcement profond pour entraîner le réseau, lui permettant ainsi d'apprendre des stratégies comportementales optimales pour les drones dans des environnements complexes. Au cours du processus de formation, nous utiliserons des unités de mémoire de relecture pour stocker les expériences historiques et combinerons des réseaux de cibles et d'estimation pour améliorer la stabilité et l'efficacité du processus d'apprentissage.

Les principales contributions de cet article sont les suivantes :

- Proposer un nouvel algorithme d'optimisation coopérative de filtre de corrélation de suppression de distorsion adaptative (ARCF-ICO), améliorant la précision et la stabilité de la planification des missions des drones.
- Intégrez des techniques d'optimisation multi-objectifs, permettant une prise de décision intelligente et efficace pour l'attribution des tâches des drones et la planification des trajectoires dans des environnements complexes. • Présenter des résultats expérimentaux qui montrent que l'algorithme ARCF-ICO surpasse les méthodes existantes en termes de métriques d'AUC et de précision sur les ensembles de données UAV123@10fps et OTB-100.

2. Travaux connexes

2.1. Optimisation à objectif unique pour la planification de la trajectoire des véhicules aériens sans pilote

La planification de trajectoire pour les véhicules aériens sans pilote (UAV) est fondamentalement un problème d'optimisation ayant des implications pratiques. Li et Duan [12] ont intégré le coût de la menace et le coût du carburant dans un objectif d'optimisation pondéré et ont utilisé un algorithme de recherche gravitationnelle universelle amélioré pour améliorer la convergence globale de la recherche, améliorant ainsi la qualité des solutions optimales pour les trajectoires des drones. Qu et al. [13] ont combiné un optimiseur simplifié de loup gris avec une recherche améliorée d'organismes symbiotiques pour proposer un nouvel algorithme hybride permettant d'obtenir des itinéraires réalisables et efficaces. Dasdemir et coll. [14] ont conçu un géné

algorithme évolutif à objectif unique basé sur les préférences pour optimiser à la fois la distance totale des itinéraires planifiés et les menaces de détection radar. Yao et coll. [15] ont introduit un algorithme hybride basé sur un modèle de contrôle prédictif et un optimiseur de loup gris amélioré pour planifier des trajectoires optimales pour le suivi de cibles de drones en environnement urbain. Papaioannou et coll. [16] ont abordé les défis de la surveillance passive de plusieurs cibles mobiles dans des environnements obstrués avec des drones en concevant un modèle de contrôleur de guidage prédictif combiné à une stratégie conjointe d'estimation et de contrôle. Ren et coll. [17] ont proposé une approche de planification de chemin multi-objectifs (MOPP) utilisant l'algorithme génétique de tri non dominé II (NSGA-II), optimisé à la fois pour la distance et la sécurité, démontrant son efficacité en milieu urbain en utilisant des subdivisions spatiales basées sur des octrees. et des cartes d'index de sécurité.

Cependant, les recherches actuelles sur la planification de trajectoire pour les drones simples et multiples se concentrent souvent sur l'optimisation d'un seul objectif, soit en considérant un seul objectif, soit en intégrant plusieurs objectifs d'optimisation en un seul via une pondération linéaire. De telles approches d'optimisation s'appuient fortement sur des coefficients de pondération subjectifs fixés par les décideurs , ayant un impact direct sur les résultats d'optimisation, et peuvent négliger les trajectoires présentant des performances exceptionnelles dans des objectifs relativement mineurs. Ces dernières années, malgré l' attention croissante portée aux problèmes de planification de trajectoire basés sur l'optimisation multi-objectifs, ne considérant généralement que deux ou trois objectifs optimisés, les exigences pratiques d'optimisation pour la planification de trajectoire des drones s'étendent au-delà d'un nombre limité d'objectifs. Pour résoudre ce problème, l'établissement d'un modèle de planification de trajectoire basé sur une optimisation multi-objectifs de grande dimension devient particulièrement crucial pour optimiser simultanément divers aspects de performance des trajectoires.

2.2. Optimisation multi-objectifs de grande dimension pour la planification de trajectoire de drones

Les problèmes d'optimisation multi-objectifs de grande dimension sont répandus dans les pratiques de vie et d'ingénierie, où plusieurs objectifs nécessitent une optimisation simultanée, souvent avec des corrélations inter-objectifs conduisant à des situations conflictuelles. Dans de tels cas, des schémas d'optimisation alternatifs doivent être envisagés pour garantir la génération de solutions équivalentes. en l'absence d'informations corrélées provenant d'autres systèmes.

Storn et Price [18] ont proposé l'algorithme d'évolution différentielle (DE), une méthode basée sur la population similaire aux mécanismes de remplacement en régime permanent, pour résoudre des problèmes d'optimisation de paramètres réels . Les nouveaux descendants ne rivalisent qu'avec leurs parents correspondants, et si les descendants présentent une meilleure forme physique, ils les remplacent. Avec l'émergence de nouvelles heuristiques bio-inspirées telles que l'optimisation des essaims de particules [19], l'algorithme du loup gris [20], l'algorithme des baleines [21] et l'algorithme de recherche Sparrow [22], comprendre comment elles s'appliquent à différents types de problèmes d'optimisation d'objectifs. devient crucial. La littérature a optimisé la distance de trajectoire des drones et le coût de la menace de trajectoire à l'aide d'algorithmes génétiques et les a lissés [23].

La démarcation des algorithmes d'optimisation multi-objectifs de grande dimension réside dans le fait que le nombre d'objectifs optimisés dépasse quatre [24]. Avec un nombre croissant d'objectifs optimisés, le nombre de solutions non dominées générées lors du processus de résolution d'algorithmes d'optimisation multi-objectifs augmente de façon exponentielle, affectant considérablement les performances et l'efficacité de l'algorithme [25,26]. De plus, la pression de sélection générée par les algorithmes d'optimisation multi-objectifs dans la résolution de problèmes multi-objectifs de grande dimension est souvent insuffisante pour guider les individus de la population vers des points idéaux.

Les stratégies visant à améliorer ces deux indicateurs se répartissent principalement en trois catégories : (1) Renforcer la pression de sélection des algorithmes en modifiant les méthodes de dominance de Pareto pour accélérer le taux de convergence des populations. GrEA [27] utilise des métriques d'évaluation basées sur une grille pour améliorer la pression de sélection des algorithmes. NSGA-III [28] utilise une stratégie de point de référence au lieu de la stratégie de distance de foule de NSGA-II [29] pour sélectionner d'excellents individus à partir de solutions non dominées, améliorant ainsi la convergence des algorithmes. 1by1EA [30] sélectionne les descendants un par un sur la base de la convergence individuelle lorsque la sélection environnementale se produit, puis améliore la diversité des populations grâce à des techniques de niche. RPEA [31] en continu

génère une série de points de référence bien convergents et distribués basés sur la population actuelle pour guider l'évolution. (2) Décomposer un problème d'optimisation multi-objectif complexe de haute dimension en un groupe de sous-problèmes et co-optimiser ces sous-problèmes.

MOEA [32] décompose le problème en une série de sous-problèmes d'optimisation à objectif unique décomposés de manière adaptative, puis agrège les informations des problèmes voisins. MaOEA/Ds [33] utilise un ensemble de vecteurs de référence autoguidés uniformément distribués dans l'espace pour diviser l'espace de décision en plusieurs petits sous-espaces et juger les mérites des individus dans les sous-espaces. Yi et coll. [34] ont proposé une méthode d'optimisation évolutive multi-objectif basée sur la décomposition objective, décomposant le problème en plusieurs sous-problèmes et résolvant chaque sous-problème en parallèle, en utilisant pleinement les informations d'autres sous-populations pour améliorer la pression de sélection de solutions non dominées. (3) Stratégie basée sur des paramètres d'évaluation. Évaluer la supériorité et l'infériorité des individus de manière globale grâce à des mesures d'évaluation, puis sélectionner d'excellents individus pour les opérations génétiques. Cependant, la complexité informatique de telles stratégies est généralement élevée et de telles métriques d'évaluation sont couramment utilisées pour évaluer la qualité des résultats d'optimisation des algorithmes [35].

Dans le processus de planification collaborative de trajectoire pour plusieurs drones, il est nécessaire de prendre en compte non seulement les attributs de trajectoire individuels, mais également la coordination spatiale entre plusieurs drones. Pour éviter l'impact de la combinaison de plusieurs objectifs d'optimisation en un seul via une pondération sur la planification de trajectoire, un modèle basé sur une optimisation multi-objectifs de grande dimension est proposé pour la planification collaborative de trajectoire pour plusieurs drones. Ce modèle optimise le coût de la distance de trajectoire d'un drone, le coût de sécurité de la trajectoire, le coût énergétique de la trajectoire et la coordination spatiale entre plusieurs drones en tant qu'objectifs d'optimisation. Contrairement aux approches existantes qui traitent la trajectoire d'un seul drone en tant qu'individu dans la population et optimisent les trajectoires de plusieurs drones séparément, dans ce modèle, plusieurs trajectoires de drones sont traitées comme une entité globale dans la population et l'optimisation est effectuée simultanément . sur plusieurs trajectoires de drones. De plus, des évaluations complètes de la convergence et de la diversité des individus dans la population sont menées, et les stratégies d'accouplement de l'algorithme pour les problèmes de planification de trajectoire sont améliorées pour améliorer les performances de convergence (35, 36). Grâce à l'algorithme, un ensemble de trajectoires Pareto-optimales pour plusieurs drones est obtenu que les décideurs peuvent utiliser. Les décideurs peuvent sélectionner les trajectoires les plus adaptées aux attributs de leur mission à partir de cet ensemble de trajectoires Pa

3. Modélisation de la planification de la trajectoire de vol multi-UAV 3.1. Description

du problème Dans le contexte de la planification collaborative de trajectoire pour plusieurs véhicules aériens sans pilote (UAV), un groupe d'UAV est chargé de naviguer à partir de plusieurs points de départ vers une série de points cibles spécifiés pour exécuter des missions complexes. Le scénario de mission se déroule dans une zone protégée par divers systèmes de défense, où les drones doivent échapper intelligemment aux menaces présentes dans la couverture radar ennemie et dans les zones de tirs anti-aériens, tout en tenant compte des limitations de performances et des exigences de coopération de chaque drone. Sur la base de ce scénario, nous effectuons des analyses de simulation des chemins de navigation et d'exécution de missions d'un groupe de drones dans un espace tridimensionnel. Les hypothèses de calcul sont les suivantes :

(1) Tous les drones maintiennent une vitesse de vol constante pendant l'exécution de la mission. (2) Chaque segment de trajectoire est divisé en trajectoires de vol rectilignes. (3) Chaque drone possède des caractéristiques de performances identiques.

Le cœur de la planification collaborative de trajectoire pour plusieurs drones implique la conception de modèles de coûts de trajectoire et de modèles spatiaux collaboratifs. Le coût de la trajectoire prend principalement en compte la distance, la menace et la consommation d'énergie. L'objectif du modèle est de minimiser ces coûts tout en améliorant la coordination spatiale entre les drones.

En supposant qu'il y ait Nz points cibles, chacun nécessitant des tâches de reconnaissance, de frappe et de confirmation représentéespar Smj , où j [1, Nz] désigne le mième type de tâche du jème point cible , correspondant à la reconnaissance, à la frappe et à la confirmation (m = 1, 2, 3). Par conséquent, le nombre total $\sum_{n=1}^{\sim\sim\sim} 3$ m=1 Smn . Chaque tâche a des fenêtres horaires spécifiques [cmn , dmn], ∑ de tâches est ∑ et les besoins en munitions désignés zj pour les tâches de frappe à chaque point cible.

Le groupe d'UAV se compose d' UAV de reconnaissance Np , d'UAV d'attaque Nq et d'UAV intégrés de reconnaissance et d'attaque Npq , totalisant Nv = Np + Nq + Npq UAV, indexés comme u [1, Nv]. Chaque drone u possède un vecteur de capacité Capabilityu correspondant à son type de tâche, où Capabilityu(m) représente la capacité du drone u à effectuer une tâche de type m, en prenant des valeurs de 1 ou 0. La capacité de charge utile de chaque drone u est zu, et si Le drone u n'a pas de capacité de frappe, alors zu * = 0. On suppose que la relation entre la consommation de carburant f pu et la vitesse de vol Vu de chaque drone u est f pu = αu × Vu, où α représente la proportionnalité entre la consommation de carburant et la vitesse de vol. , et Vu est dans la plage [ku, ju], désignant les vitesses minimale et maximale.

Dans les environnements urbains, le problème de plusieurs drones poursuivant plusieurs cibles au sol est décrit dans la figure 1, où N drones sont prêts à exécuter M tâches. L'ensemble UAV est noté U = {U1, U2, . . . , UN}, et l'ensemble cible est noté H = {H1, H2, . . . , HM}, avec des types de tâches représentés par Mi = {1, 2} (Mi = 1 pour la reconnaissance, Mi = 2 pour la frappe).

Pour démontrer l'hétérogénéité des drones et les exigences de performances spécifiques des scénarios de mission, Zum est utilisé pour représenter la matrice de performances des drones, où les éléments de la matrice représentent la capacité du drone à exécuter une certaine tâche. Par exemple, s'il y a trois drones et que leur matrice de performances Zum est présentée dans le tableau 1, Zum (1, 1) = 0,9 indique que la capacité de tâche de reconnaissance de l'UAV 1 est de 0,9. Si H1 est une tâche de reconnaissance (M1 = 1) avec une capacité minimale requise de 0,6, alors parmi les trois drones U1, U2, U3, seul U1 (0,9) peut répondre aux exigences de la tâche H1.

Tableau 1. Valeur de capacité du drone.

Véhicule aérien sans pilote	Valeur de capacité	
	Scout	Piste
U1	0,9	0,4
U2	0,3	0,9
U3	0,5	0,5

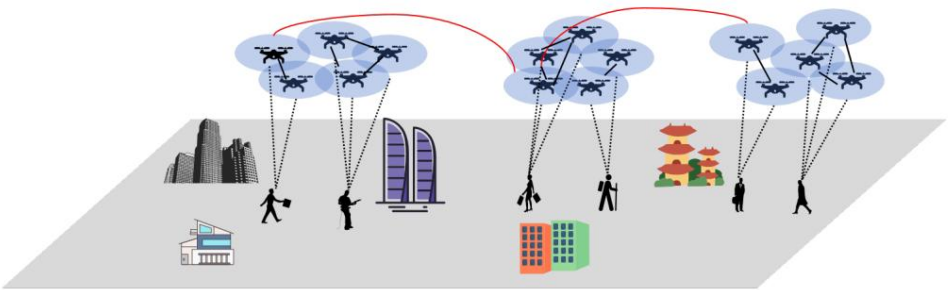


Figure 1. Allocation de tâches multi-UAV dans des scénarios urbains.

3.2. Modèle objectif de conception d'optimisation

Notre objectif est de minimiser la consommation totale de carburant de l'ensemble de la flotte de drones pendant l' exécution des tâches tout en maximisant l'efficacité de l'exécution des tâches. Par conséquent, nous définissons les fonctions de coût suivantes.

Coût total de consommation de carburant : cela représente la consommation totale de carburant des drones depuis le décollage, l'exécution des tâches jusqu'au retour à l'aéroport. La consommation de carburant de chaque drone dépend de la distance de vol et de la vitesse. Si nous définissons la consommation de carburant du drone u effectuant la tâche k du point Pi à Pj comme fu,k,i,j , alors la consommation totale de carburant Ff uel peut être exprimée comme

$$F_{f\ carburant} = \sum_{u=1}^{Nu} \sum_{k=1}^{3} \sum_{je=0}^{Np} \sum_{j=1}^{Np} X_{u,k,i,j} \cdot f_{u,k,i,j} \quad (1)$$

Coût de remise en forme du temps de tâche : cela mesure la différence entre le temps d'achèvement des tâches et le point médian de leurs fenêtres temporelles. L'objectif est de rendre la tâche des drones

des temps d'exécution aussi proches que possible du milieu des fenêtres temporelles des tâches. Si nous désignons l'aptitude de la fenêtre temporelle du drone u par Tadapt,u, alors l'aptitude totale de la fenêtre temporelle Ftime est

$$\text{Ftemps} = \sum_{u=1}^{Nu} \text{Tadapter,u} = \sum_{u=1}^{Nu} \sum_{je=0}^{Np} \sum_{j=1}^{Np} \sum_{k=1}^{3} Xu,k,i,j \cdot (tu,k,i,j -) \quad \frac{sk,j + ek,j}{2} \tag{2}$$

En combinant les deux fonctions de coût ci-dessus, nous formons un problème d'optimisation bi-objectif :

$$\text{Minimiser } F = \alpha \cdot \text{Ff carburant} + \beta \cdot \text{Ftime} \tag{3}$$

où α et β sont des paramètres qui équilibrent l'importance des deux objectifs. Dans l'équation (3), les fonctions objectifs sont normalisées pour garantir que chacune contribue de manière égale à l'optimisation globale. Le processus de normalisation implique la mise à l'échelle de chaque fonction objectif dans une plage [0, 1] en fonction de leurs valeurs maximales et minimales respectives observées lors des exécutions initiales.

3.3. Contraintes

Dans le problème de planification de trajectoire d'UAV, nous devons prendre en compte les contraintes suivantes, y compris les conditions météorologiques pour tenir compte des opérations à basse altitude.

Contrainte d'exécution des tâches : s'assurer que chaque tâche est exécutée par au moins un drone avec la capacité correspondante. La formule de contrainte pour l'exécution des tâches est la suivante :

$$\sum_{u=1}^{Nu} Xu,k,i,j \geq 1 \quad k, \quad i, \quad j \tag{4}$$

Ici, Xu,k,i,j représente la variable de décision binaire indiquant si le drone u exécute la tâche k du point Pi au point Pj . Nu est le nombre total de drones. k représente le type de tâche. i représente l'indice du point de départ. j représente l'indice du point final.

Contrainte de correspondance de capacité : les tâches exécutées par les drones doivent être conformes à leurs limitations de capacité. La formule de contrainte pour la correspondance des capacités est la suivante :

$$\sum_{u=1}^{Nu} Xu,k,i,j \geq 1 \quad k, \quad i, \quad jXu,k,i,j \leq \text{Capacitéu,k} \quad u, \quad k, \quad i, \quad j \tag{5}$$

où Capabilityu,k représente la capacité du drone u à effectuer la tâche k.

Contrainte des conditions météorologiques : Les opérations à basse altitude sont influencées par les conditions météorologiques telles que la vitesse du vent, les précipitations et la visibilité. Ces facteurs sont intégrés à l'estimation de la trajectoire pour garantir des opérations sûres et fiables. La formule de contrainte pour les conditions météorologiques est la suivante :

$$Wu,k,i,j \leq Wmax \quad u, \quad k, \quad i, \quad j \tag{6}$$

où Wu,k,i,j représente l'impact météorologique sur le drone u lors de l'exécution de la tâche k du point Pi au point Pj . Wmax est l'impact météorologique maximum autorisé.

Contrainte de temps de vol : assurez-vous que le temps nécessaire à un drone pour voler d'un point de tâche à un autre ne dépasse pas la valeur maximale spécifiée. La formule de contrainte pour le temps de vol est la suivante :

$$tu,k,i,j - tu,k,i,j-1 \leq Tmax \quad u, \quad k, \quad i, \quad j > 1 \text{ où } tu,k,i,j \tag{7}$$

représente l'heure à laquelle le drone u arrive à la tâche point j en effectuant la tâche k, à partir du point Pi . Tmax est le temps de vol maximum autorisé entre des points de tâche consécutifs.

3.4. Mesures de performance

L'évaluation des performances est cruciale pour valider l'efficacité de notre modèle d'optimisation. Nous utilisons les mesures de performances suivantes :

Couverture des tâches : rapport entre les tâches assignées et terminées avec succès et le nombre total de tâches. La couverture des tâches peut être exprimée comme

$$\text{Couverture} = \frac{\text{Nombre de tâches assignées et terminées avec succès}}{\text{Nombre total de tâches}} \quad (8)$$

Consommation moyenne de la batterie : la consommation moyenne de la batterie de la formation du drone pour effectuer toutes les tâches s'exprime comme suit :

$$\text{Consommation moyenne de la batterie} = \frac{\sum_{u=1}^{Nu} \text{Batterie utilisée par vous}}{Nu} \quad (9)$$

Endurance : L'endurance mesure le temps de fonctionnement des drones, garantissant qu'ils peuvent effectuer des tâches dans les limites de leur batterie. Il s'exprime comme

$$\text{Endurance} = \frac{\sum_{u=1}^{Nu} \text{Temps de fonctionnement de vous}}{Nu} \quad (dix)$$

Efficacité temporelle : la différence entre le temps d'exécution de toutes les tâches et le heure de début la plus rapprochée. Cela peut être exprimé comme

$$\text{Efficacité du temps} = \max_{u \in U} \{\text{tachèvement},u\} - \min_{u \in U} \{\text{tstart},u\} \quad (11)$$

Grâce à ces mesures de performances, nous pouvons évaluer de manière globale l'efficacité de l'algorithme d'optimisation, permettant des ajustements supplémentaires des paramètres du modèle ou des améliorations algorithmiques.

4. Algorithme d'estimation multi-objectif génétique adaptatif L'algorithme ARCF

(Adaptive Response Map Correction Filter) vise à intégrer la distorsion de la carte de réponse se produisant pendant le processus de suivi avec le processus de formation du filtre, améliorant ainsi les performances de l'algorithme (comme illustré dans la figure 2).). Pour supprimer la distorsion de la carte de réponse, la première étape consiste à identifier la distorsion (c'est-à-dire déterminer quand la distorsion de la carte de réponse se produit). Il introduit la norme euclidienne pour définir la différence entre les cartes de réponse de la trame précédente M1 et la trame actuelle M2 .

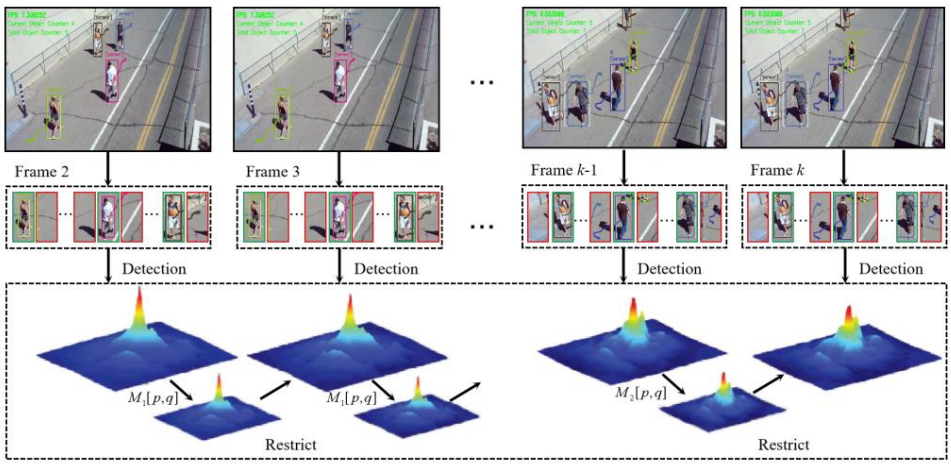


Figure 2. Organigramme de l'algorithme ARCF.

L'objectif principal de l'algorithme ARCF est de combiner la distorsion de la carte de réponse générée lors du suivi de cible avec le processus d'entraînement du filtre, améliorant ainsi les performances de suivi en mettant à jour dynamiquement le filtre. La distorsion de la carte de réponse est principalement causée par le mouvement rapide de la cible ou par des facteurs environnementaux externes tels que l'occlusion et les changements d'éclairage. L'algorithme ARCF identifie et supprime ces distorsions en analysant les changements dans les cartes de réponse entre trames consécutives, notamment en calculant la distance euclidienne entre les cartes de réponse de deux trames.

Dans l'algorithme, les cartes de réponse de deux trames F1 et F2 consécutives sont respectivement notées M1 et M2 et alignées à l'aide de l'opération de décalage ψ pour calculer leurs différence La formule pour calculer la différence est

$$\Delta M = \psi(M1, M2) = \sum_{p,q} \sqrt{(M1[p, q] - M2[p, q])2} \quad (12)$$

Ici, p et q représentent les coordonnées spatiales de la carte de réponse.

Sur la base de la mesure de différence susmentionnée, la fonction objective de l' algorithme ARCF peut être décrite comme un problème d'optimisation visant à minimiser la distorsion de la carte de réponse tout en maximisant la précision du suivi de la cible. La fonction objectif est constituée d'un terme de distorsion et d'un terme de régularisation, exprimés par

$$\min h \quad \lambda \quad h \quad ^2 + y \quad \sum_{d=1}^{D} h \quad Md - Yd \quad ^2 + \quad \Delta M \quad (13)$$

où h est le filtre, désigne l'opération de convolution, Md est la carte de réponse d'entrée pour le d-ième canal, Yd est la carte de réponse de sortie souhaitée et λ, γ et sont des coefficients ajustant l'importance de chaque terme.

Pour plus d'efficacité de calcul, la fonction objectif est ensuite transformée en domaine fréquentiel. Dans le domaine fréquentiel, la formule devient

$$\min_{H} \quad \lambda \quad H \quad ^2 + \gamma \quad F(H) \quad X - Y \quad ^2 + \quad \Delta M \quad (14)$$

où F représente la transformée de Fourier, désigne la multiplication par éléments, et X et Y sont respectivement les représentations du domaine fréquentiel d'entrée et de sortie souhaitée.

En utilisant l'algorithme ADMM (Alternating Direction Method of Multipliers) pour résoudre le problème d'optimisation, le filtre peut être mis à jour efficacement. Le processus de résolution implique deux sous-problèmes principaux : l'optimisation du filtre et la mise à jour de la carte de réponse. Cette approche supprime efficacement la distorsion de la carte de réponse provoquée par un mouvement rapide ou des changements environnementaux externes, améliorant ainsi la stabilité et la précision de l' algorithme de suivi. L'algorithme ARCF peut être résumé comme l'algorithme 1.

Algorithme 1 : La procédure de l'ARCF

```
Initialiser le filtre h0, taux d'apprentissage λ, γ,
Définir les itérations maximales
T pour t = 1 à T do
    Capturer l'image actuelle Ft
    Calculer la carte de réponse Mt en utilisant
    ht−1 si t > 1 alors
        Calculer la distorsion ΔM = ∑p,q (Mt [p, q] − Mt−1[p, q])2
    fin si
    Mettre à jour le filtre dans le domaine fréquentiel :
    Ht = argminH λ  H  2 + γ  F(H)  Xt − Yt  2 +  ΔM
    Mettre à jour la carte de réponse Mt :
    Mt = Ht  Ft
    Vérifier la convergence :
    si  F(Ht) − F(Ht−1)  < seuil alors
        pause
    fin si
    Mettez à jour ht avec Ht à la fin du domaine
spatial pour
Afficher la meilleure solution globale Gbest = hT
```

4.1. Algorithme ARCF-ICO

L'algorithme ARCF-ICO intègre des améliorations significatives à la stratégie ARCF originale, en se concentrant à la fois sur la convergence et la diversité des solutions au sein d'un espace d'optimisation de grande dimension . Cela est particulièrement pertinent dans le contexte de la planification de missions d'UAV , où des environnements opérationnels diversifiés et des exigences de mission en évolution rapide nécessitent une approche d'optimisation robuste et adaptative.

Dans la phase initiale de l'algorithme ARCF-ICO, la priorité est donnée à l'atteinte de taux de convergence élevés. Cela garantit un alignement rapide des drones vers des trajectoires ou des ensembles de solutions optimaux, répondant ainsi efficacement aux besoins opérationnels immédiats tels que la surveillance ou la détection des menaces. Au fur et à mesure que l'algorithme progresse, l'accent est mis sur la préservation de la diversité des solutions. Ceci est crucial dans les opérations de drones pour explorer une gamme de trajectoires de vol potentielles ou de stratégies tactiques, évitant ainsi les optima locaux et améliorant la robustesse des résultats de la mission. L'algorithme ARCF-ICO est conçu pour être implémenté par diverses catégories de drones, notamment les drones à voilure fixe, à voilure tournante et hybrides. Ces drones peuvent être utilisés dans des applications commerciales et militaires, en fonction de leurs capacités et des exigences de leur mission.

L'indicateur d'évaluation complet de la convergence et de la diversité (CAD) [37] est une mesure nouvellement proposée conçue pour évaluer à la fois la convergence et la diversité des solutions dans notre cadre ARCF-ICO. Ce nouvel indicateur est défini comme suit :

$$CAD(ui , U) = 1 + rand(0,8, 1) \times M \times \left(\frac{t}{tmax}\right)^{\theta} \times D(ui , U) \times (1 - C(ui , U)) \quad (15)$$

où D(ui , U) représente la mesure de diversité et C(ui , U) désigne la mesure de convergence des drones ui au sein de la flotte U. Le paramètre M désigne le nombre d'objectifs, et θ régit l'équilibre entre convergence et diversité comme le l'algorithme itère de t à tmax – le nombre maximum de générations.

La diversité D(ui , U) est calculée comme

$$SDE(ui , U) = \min_{ui \quad U,j \quad =i} \sqrt{\sum_{k=1}^{m} sde(f_k'(ui , U), f_k'(uj , U))2} \quad (16)$$

où $sde(f_k'(ui , U), f_k'(uj , U))$ est défini comme :

$$sde(f_k'(ui , U), f_k'(uj , U)) = \begin{cases} f_k'(uj , U) - f_k'(ui , U) & f_k'(uj , U) > f \text{ sinon } f_k'(ui , U) \\ 0 & \end{cases} \quad (17)$$

La convergence C(ui , U) d'un drone par rapport à la flotte est quantifiée comme

$$C(ui , U) = \frac{Disque (ui , U)}{\sqrt{m}} \quad (18)$$

où Disc(ui , U) représente la distance euclidienne entre le drone ui et la solution idéale dans l'espace objectif normalisé.

En intégrant ces stratégies, l'ARCF-ICO adapte dynamiquement la planification des missions et les tactiques de réponse des UAV en fonction de l'évolution des conditions environnementales et des demandes opérationnelles, optimisant à la fois l'efficience et l'efficacité des UAV déployés.

4.2. Planification multi-objectifs ARCF-ICO

Dans le modèle d'optimisation multi-objectif ARCF-ICO, nous avons conçu une stratégie d'accouplement efficace adaptée aux environnements de tâches complexes rencontrés par les drones. Cette stratégie combine les traits favorables des individus au sein de la population et introduit des éléments stochastiques pour augmenter la diversité de la population, améliorant ainsi l'adaptabilité et la flexibilité de l'algorithme. L'algorithme ARCF-ICO est applicable aux missions menées en ligne de visée visuelle (VLOS) [38], au-delà de la ligne de visée visuelle (BVLOS) [38], et entière

opérations autonomes. La flexibilité de l'algorithme lui permet de s'adapter aux différentes contraintes et exigences opérationnelles.

Dans le cadre de la planification de trajectoire des drones, les points de trajectoire générés par chaque Les drones pour chaque segment de trajectoire sont représentés par la matrice suivante :

$$X = [X1, X2, \ldots, Xm]^{T} \quad (19)$$

où Xi est une matrice N × 3L représentant le ième individu de la population, N est le nombre de drones et L est le nombre de segments de trajectoire.

Au cours du processus d'accouplement dans la population parentale, deux parents P1 et P2 sont sélectionnés sur la base de leur indicateur d'évaluation complet CAD. La formule de calcul CAD est la suivante :

$$CAD(P) = \frac{1}{N}\sum_{je=1}^{N} 1 + rand(0,8, 1) \times M \times \left(\frac{t}{tmax}\right)^{\theta} \times D(Pi , P) \times (1 - C(Pi , P)) \quad (20)$$

où D(Pi , P) et C(Pi , P) représentent les indicateurs de diversité et de convergence des individus Pi , M est le nombre de fonctions objectives, θ est le paramètre d'équilibre, t est la génération actuelle et tmax est le paramètre d'équilibre. génération maximale.

L'opération d'accouplement est la suivante, combinant les informations de trajectoire des deux parents pour générer de nouvelles trajectoires de progéniture :

$$Pnew,j = \alpha P1,j + (1 - \alpha)P2,j \text{ où } P1,j \quad (21)$$

et P2,j désignent respectivement les coordonnées de position des parents P1 et P2 au jème point de la trajectoire, et α est un nombre aléatoire compris entre 0 et 1 est utilisé pour contrôler le ratio de contribution des deux parents dans la progéniture nouvellement générée. L'opération d'accouplement individuelle est illustrée à la figure 3.

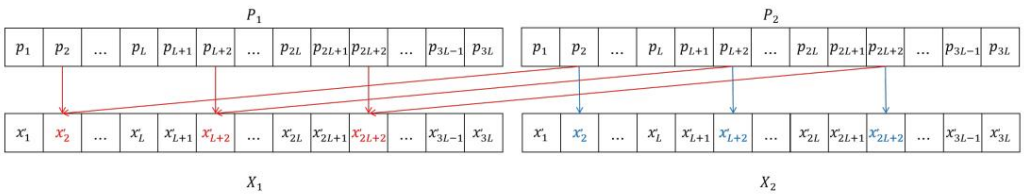


Figure 3. Opération d'accouplement individuel.

Cette conception garantit que les drones peuvent ajuster leurs stratégies de vol en fonction des exigences réelles des tâches et des changements environnementaux lors de l'exécution de tâches telles que la surveillance, la reconnaissance ou d'autres missions complexes. En outre, l'approche d'optimisation multi-objectifs permet aux drones d'optimiser d'autres paramètres de tâche importants, tels que le temps de vol et le rendement énergétique, tout en garantissant l'efficacité des tâches, garantissant ainsi une exécution complète et performante de la mission.

5. Simulation et analyse des résultats 5.1. Ensembles de données

Pour évaluer de manière exhaustive les performances de l'algorithme, nous utilisons deux ensembles de données largement utilisés : l'ensemble de données UAV123@10fps [39] et l'ensemble de données OTB-100 [40]. Vous trouverez ci-dessous les détails spécifiques de chaque ensemble de données :

Ensemble de données UAV123@10fps : cet ensemble de données comprend 123 scénarios de suivi capturés à l'aide de drones dans des environnements aériens. Il comprend un mélange de scènes réelles et synthétiques générées à l'aide de simulateurs. L'ensemble de données couvre 12 environnements de défi de suivi différents, offrant un ensemble diversifié de scénarios d'évaluation.

Ensemble de données OTB-100 : L'ensemble de données OTB-100 se compose de 100 scénarios de suivi réels capturés manuellement. Il englobe 11 environnements de défi de suivi distincts, offrant un large éventail de scénarios pour évaluer les performances des algorithmes.

L'utilisation de ces ensembles de données permet une évaluation complète de l'algorithme à travers divers défis de suivi et conditions environnementales. Ses informations spécifiques est présenté dans le tableau 2 ci-dessous.

Tableau 2. UAV123@10fps et détails de l'ensemble de données OTB-100.

UAV123 @ 10fps			OTB-100	
Numéro de série	Nom du défi	Nombre	Nom du défi	Nombre
1	Variation d'échelle (SV)	109	Variation d'échelle (SV)	65
2	Changement de rapport d'aspect (ARC)	68	Occlusion (OCC)	49
3	Fouillis d'arrière-plan (C.-B.)	21	Variation d'éclairage (IV)	38
4	Mouvement de la caméra (CM)	70	Flou de mouvement (Mo)	31
5	Mouvement rapide (FM)	28	Déformation (DEF)	43
6	Occlusion complète (FOC)	33	Mouvement rapide (FM)	43
7	Variation d'éclairage (IV)	31	Rotation hors plan (OPR)	64
8	Basse résolution (LR)	48	Rotation dans le plan (IPR)	52
9	Hors de vue (OV)	30	Clutters d'arrière-plan (BC)	33
dix	Occlusion partielle (POC)	73	Hors de vue (OV)	14
11	Objet similaire (SOB)	39	Basse résolution (LR)	dix
12	Changement de point de vue (VC)	60	-	-

5.2. Paramétrage expérimental

Dans la configuration des expériences, la configuration matérielle utilisée comprend un processeur Intel Core i9-13700 et 32 Go de mémoire. La configuration du logiciel est basé sur la plateforme MATLAB R2019a. Concernant le paramétrage, la régularisation Le paramètre est défini sur 1,2, suivant les paramètres de l'algorithme ARCF d'origine. Après Ajustements expérimentaux, un autre paramètre de régularisation est fixé à 0,001. L'apprentissage le taux pour le modèle cible, noté η, est fixé à 0,0192, toujours en suivant les directives à partir de l'algorithme ARCF original.

5.3. Indices d'évaluation de corrélation

Pour l'algorithme d'optimisation multi-objectifs ARCF-ICO dans la planification de missions de drones, nous utilisons trois mesures d'évaluation améliorées pour évaluer de manière exhaustive les performances de l'algorithme : l'aire sous la courbe (AUC), l'erreur de localisation centrale (CLE) et la précision. Ce les mesures évaluent efficacement les performances et la précision des drones pendant l'exécution de la mission.

Aire sous la courbe (AUC) : cette mesure évalue le taux de réussite global des drones dans de multiples tâches de vol, en particulier dans le maintien des objectifs cibles dans des environnements complexes . L'ASC est calculée comme

$$\text{ASC} = \frac{\text{zone (Bpred} \cap \text{Btrue)}}{\text{zone (Bpred} \quad \text{Btrue)}}, \tag{22}$$

où Bpred représente le rectangle de localisation de la cible prédit par l'algorithme UAV, et Btrue représente le rectangle du véritable emplacement cible.

Erreur de localisation centrale (CLE) : cette métrique mesure la distance euclidienne moyenne entre la position centrale prévue du drone et la véritable position centrale de la cible. Une grande précision en CLE est essentielle pour garantir une exécution précise de tâches telles que la surveillance et des reconnaissances. CLE est calculé comme

$$\text{CLE} = \sqrt{(\text{xpred} - \text{xtrue})^2 + (\text{ypre} - \text{yvrai})^2}, \tag{23}$$

où (xpred, ypred) est la position centrale prédite par le drone, et (xtrue, ytrue) est la vraie position centrale de la cible.

Précision : La précision mesure la précision de la localisation du drone dans un seuil spécifique t, c'est-à-dire la proportion d'images où l'erreur de prédiction de la position centrale de la cible se situe dans le seuil. Il s'agit d'une mesure essentielle pour évaluer les performances de suivi en temps réel des drones. La précision est calculée comme

$$\text{Précision} = \frac{Nt \leq seuil}{Ntotal}, \quad (24)$$

où Nt≤threshold est le nombre de trames où CLE est inférieur ou égal au seuil t, et Ntotal est le nombre total de trames. Le seuil t est typiquement fixé à 20 pixels.

Grâce à ces trois mesures d'évaluation, les performances des drones dans l'exécution de tâches dans des environnements complexes, telles que la précision et la stabilité de la planification de trajectoire, peuvent être évaluées de manière exhaustive. Ces mesures aident non seulement à optimiser les stratégies opérationnelles des drones, mais fournissent également des informations cruciales pour une amélioration ultérieure des algorithmes et des ajustements des paramètres de vol.

5.4. Étude comparative

Dans ce chapitre, l'algorithme ARCF-ICO est évalué à l'aide de deux ensembles de données : UAV123@10fps et OTB-100. Pour mieux comprendre les performances de l'algorithme ARCF-ICO proposé, il sera comparé à huit algorithmes populaires dans le domaine du suivi d'objets vidéo, notamment KCF [41], LDES [42], MCCT-H [43], Staple [44].], fDSST [45] et AutoTrack [46]. Les algorithmes LDES et AutoTrack ont été publiés respectivement dans AAAI2019 et CVPR2020, en se concentrant sur les ensembles de données UAV, tandis que les cinq algorithmes restants sont des algorithmes de suivi classiques basés sur des filtres de corrélation ces dernières années.

La figure 4 illustre la comparaison complète de l'algorithme ARCF-ICO avec les six autres algorithmes traditionnels de l'ensemble de données UAV123@10fps en termes d'AUC et de précision. D'après la figure, on peut observer que l'algorithme ARCF-ICO proposé atteint les performances les plus élevées, avec une AUC globale de 0,516 et une précision globale de 0,712, surpassant les six autres algorithmes. Les algorithmes AutoTrack et LDES se classent deuxième et troisième, avec des valeurs globales d'AUC de 0,504 et 0,492 et des valeurs globales de précision de 0,682 et 0,655, respectivement. Par rapport à l' algorithme ARCF-ICO, les algorithmes AutoTrack et ARCF ont une AUC inférieure de 1,2 % et 2,4 %, et une précision inférieure de 3,0 % et 5,7 %, respectivement. Les cinq algorithmes restants, MCCT-H, Staple, fDSST et KCF, ont des valeurs d'AUC inférieures, allant de 0,285 à 0,456, et des valeurs de précision inférieures, allant de 0,384 à 0,581, par rapport à l'algorithme ARCF-ICO.

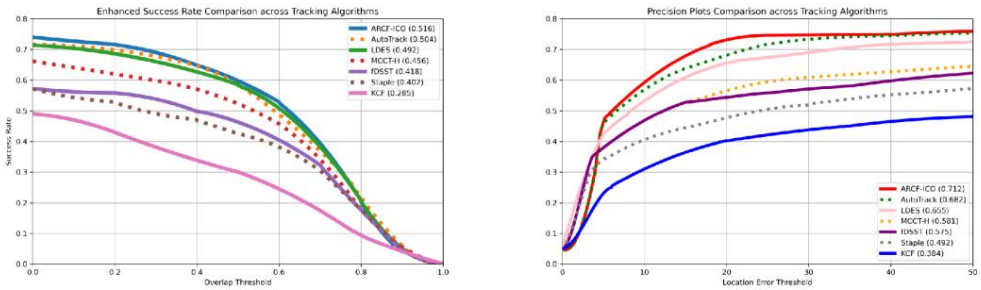


Figure 4. Comparaison entre l'AUC et la précision dans l'ensemble de données UAV123@10fps.

Le tableau 3 présente la comparaison complète de précision de divers algorithmes dans les 12 environnements difficiles de l'ensemble de données UAV123@10fps. Le tableau 3 montre que l'algorithme ARCF-ICO se classe systématiquement premier en termes de précision dans les 12 environnements difficiles. Les algorithmes AutoTrack et LDES se classent respectivement deuxième dans dix et deux environnements difficiles. Ainsi, les résultats de comparaison du tableau 3 démontrent que l' algorithme ARCF-ICO s'adapte bien aux défis de suivi les plus complexes et présente des performances de suivi robustes.

Tableau 3. Tableau de comparaison de précision UAV123@10fps pour divers scénarios de défi de l'ensemble de données.

Nom du défi	Notre algorithme	Algorithme de contraste					
	ARCF-ICO	AutoTrack	LDES	MCCT-H	fDSST	Agrafe	KCF
SV	0,724	0,672	0,642	0,545	0,521	0,496	0,372
ARC	0,692	0,686	0,631	0,591	0,552	0,503	0,336
avant JC	0,702	0,621	0,603	0,503	0,521	0,467	0,415
CM	0,684	0,691	0,633	0,512	0,488	0,472	0,375
FM	0,696	0,664	0,628	0,498	0,508	0,475	0,311
FOC	0,693	0,677	0,689	0,582	0,571	0,436	0,388
IV	0,669	0,673	0,642	0,603	0,598	0,479	0,406
G / D	0,688	0,645	0,667	0,654	0,635	0,527	0,416
VO	0,741	0,714	0,656	0,672	0,626	0,426	0,403
POC	0,744	0,725	0,702	0,625	0,645	0,539	0,369
SANGLOT	0,752	0,714	0,693	0,564	0,586	0,545	0,388
Capital-risque	0,763	0,702	0,671	0,625	0,645	0,535	0,426

Le tableau 4 présente les résultats comparatifs des analyses de précision complètes de divers algorithmes. dans 11 environnements difficiles au sein de l'ensemble de données OTB-100. Il ressort du tableau 4 que l'algorithme ARCF-ICO se classe premier en précision dans les 11 environnements, avec le Les algorithmes AutoTrack et LDES se classent respectivement deuxième dans cinq et trois environnements. De plus, les algorithmes MCCT-H et Staple obtiennent la deuxième place dans le classement Fast Motion. et les défis de rotation hors plan, respectivement. La comparaison résulte donc de Le tableau 4 valide en outre l'efficacité de l'algorithme ARCF-ICO.

Tableau 4. Tableau de comparaison de précision OTB-100 pour divers scénarios de défi de l'ensemble de données.

Nom du défi	Notre algorithme	Algorithme de contraste					
	ARCF-ICO	AutoTrack	LDES	MCCT-H	fDSST	Agrafe	KCF
FM	0,517	0,485	0,497	0,505	0,422	0,388	0,263
avant JC	0,483	0,476	0,496	0,476	0,412	0,433	0,314
Mo	0,514	0,502	0,467	0,443	0,402	0,367	0,278
DEF	0,482	0,477	0,472	0,425	0,378	0,336	0,266
IV	0,541	0,538	0,533	0,501	0,445	0,388	0,325
DPI	0,477	0,432	0,375	0,468	0,305	0,325	0,213
G / D	0,533	0,532	0,596	0,492	0,462	0,439	0,306
OCC	0,488	0,482	0,457	0,426	0,430	0,422	0,239
OPR	0,563	0,552	0,512	0,423	0,458	0,563	0,325
VO	0,534	0,529	0,505	0,412	0,433	0,368	0,336
SV	0,545	0,537	0,546	0,447	0,448	0,392	0,268

5.5. Visualisation de simulation

La figure 5 illustre les résultats d'allocation de tâches obtenus par l'algorithme ARCF-ICO, démontrant que ce problème d'allocation dynamique est essentiellement une programmation non linéaire problème avec des solutions optimales. Le panneau de gauche de la figure 5 montre les solutions Pareto pour l'attribution des tâches des drones et la planification de trajectoires dérivées de l'ensemble de données UAV123@10fps, tandis que le panneau de droite présente les solutions de l'ensemble de données OTB-100. Le front de Pareto les deux panneaux indiquent les compromis entre différents objectifs, mettant en valeur l'efficacité et l'efficacité de l'algorithme ARCF-ICO dans la gestion de l'optimisation multi-objectifs. Dans De plus, nous montrons également les résultats d'affectation des tâches de simulation MATLAB de 10 drones dans

Figure 6. Comme le montre la figure 6d, 10 drones sont sur le point de trouver l' objet cible correspondant.

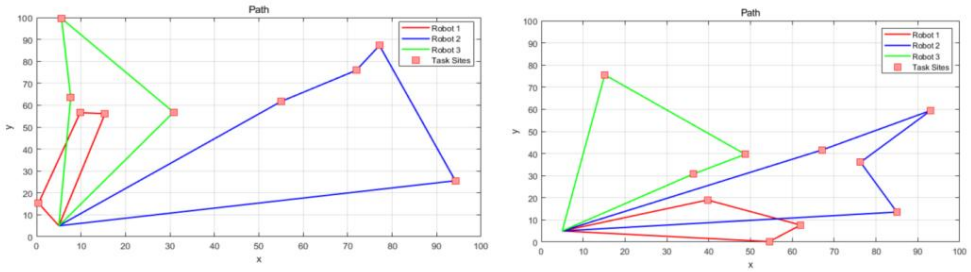


Figure 5. Affectation des tâches de solution de l'algorithme ARCF-ICO. Solutions Pareto pour l'attribution des tâches des drones et la planification de trajectoires dérivées de l'ensemble de données UAV123@10fps (à gauche) et de l'ensemble de données OTB-100 (à droite).

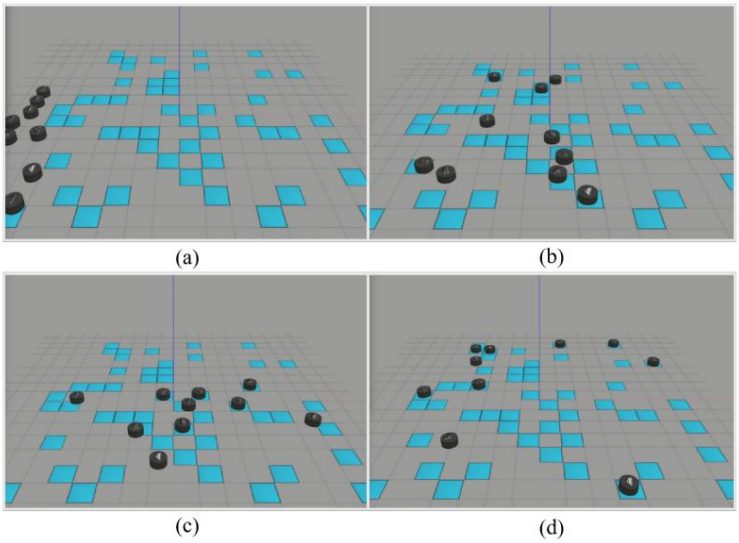


Figure 6. Démonstration d'affectation de tâches de simulation MATLAB de 10 drones. (a) L'état initial de l' UAV ; (b, c) Statut intermédiaire de l'affectation des tâches du drone ; (d) Statut final de l'affectation des drones.

De plus, des tests de simulation sont effectués sur des scènes de l'ensemble de données OTB-100. La carte de l'environnement est divisée en cinq sous-cartes, mais l'utilisation de l'algorithme ARCF-ICO seul pour la planification de la trajectoire des drones risque de ne pas atteindre la vitesse de formation du modèle la plus rapide. Cela pourrait être dû au processus d'apprentissage continu par essais et erreurs de l'algorithme original ARCF dans l'environnement, qui nécessite plus de temps. La figure 7 illustre la planification de trajectoire des tâches d'UAV dans les cartes d'environnement recadrées de l'ensemble de données OTB-100 à l'aide de l'algorithme ARCF-

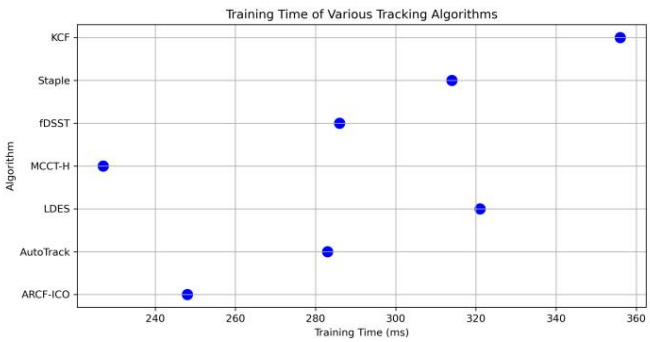


Figure 7. L'algorithme ARCF-ICO est démontré dans la carte de l'environnement de culture de l' ensemble de données OTB-100 pour la planification de mission de drone. Le point bleu représente le temps de formation pour chaque algorithme.

6. Conclusions et discussion

Cette étude présente l'algorithme ARCF-ICO pour aborder l'optimisation multi-objectifs dans la planification des missions des drones, améliorant ainsi les performances de suivi des drones dans des environnements complexes. En se concentrant à la fois sur la convergence et la diversité, l'algorithme ARCF-ICO s'adapte aux changements environnementaux rapides et aux demandes dynamiques des missions grâce à des mises à jour de filtres en temps réel. La validation à l'aide des ensembles de données UAV123@10fps et OTB-100 démontre que l'algorithme ARCF-ICO surpasse les méthodes existantes en matière de métriques AUC et Precision, indiquant une précision et une robustesse de suivi supérieures. Cependant, deux principales limites sont identifiées : les performances sous-optimales de l'algorithme avec des cibles dynamiques extrêmement rapides et la nécessité d'une efficacité de calcul améliorée. Les recherches futures exploreront des structures algorithmiques plus efficaces pour réduire la consommation de ressources informatiques tout en optimisant la réponse aux cibles en mouvement à grande vitesse. Dans l'ensemble, l'algorithme ARCF-ICO offre une avancée technologique significative pour l'optimisation multi-objectifs des drones, apportant une valeur théorique et pratique substantielle pour les applications civiles

Contributions de l'auteur : conceptualisation, méthodologie, rédaction – préparation de l'ébauche originale, visualisation et supervision : ZZ (Zijie Zheng). Validation, analyse formelle et enquête : ZZ (Zhijun Zhang). Logiciel, validation, rédaction — révision et édition : ZL Ressources, conservation des données, rédaction — révision et édition : QY Validation, analyse formelle et méthodologie : YJ Tous les auteurs ont lu et accepté la version publiée du manuscrit.

Financement : Cette recherche n'a reçu aucun financement externe.

Déclaration de disponibilité des données : adresse de téléchargement de l'ensemble de données UAV123 : https://cemse.kaust.edu.sa/ivul/ drone123 (consulté le 10 mars 2024). Adresse de téléchargement de l'ensemble de données OTB : http://cvlab.hanyang.ac. kr/tracker_benchmark/datasets.html (consulté le 10 mars 2024). Le code soutenant les résultats de cette étude n'est pas accessible au public mais peut être obtenu auprès de l'auteur correspondant sur demande raisonnable.

Conflits d'intérêts : Les auteurs ne déclarent aucun conflit d'intérêts.

Les références

1. Chen, J. ; Zhang, Y. ; Wu, L. ; Vous, T. ; Ning, X. Un algorithme basé sur le clustering adaptatif pour la planification automatique des chemins de données hétérogènes. Les drones. IEEETrans. Intell. Transp. Système. 2021, 23, 16842-16853. [Référence croisée]
2. Chen, J. ; Ling, F. ; Zhang, Y. ; Vous, T. ; Liu, Y. ; Du, X. Planification du chemin de couverture de véhicules aériens sans pilote hétérogènes basée sur système de colonies de fourmis. Essaim Evol. Calculer. 2022, 69, 101005. [Réf. croisée]
3. Chen, J. ; Li, T. ; Zhang, Y. ; Vous, T. ; Lu, Y. ; Tiwari, P. ; Kumar, N. Apprentissage par renforcement basé sur l'attention globale et locale pour le contrôle coopératif du comportement de plusieurs drones. IEEETrans. Véh. Technologie. 2024, 73, 4194-4206. [Référence croisée]
4. Ning, X. ; Tian, W. ; Lui, F. ; Bai, X. ; Soleil, L. ; Li, W. Modèle de neurone à fonction de couverture hyper-saucisse et algorithme d'apprentissage pour classement des images. Reconnaissance de modèles. 2023, 136, 109216. [Réf. croisée]
5. Yao, Y. ; Liu, Z. Le nouveau concept de développement contribue à accélérer la formation d'une nouvelle productivité de qualité : logique théorique et les voies de mise en œuvre. J.Xian Univ. Finance. Écon. 2024, 37, 3-14.
6. Ma, G. ; Chen, X. De la puissance financière à la puissance financière : comparaison internationale et approche chinoise. J.Xian Univ. Finance. Écon. 2024, 37, 46-59.
7. Wang, J. ; Li, F. ; N'importe lequel.; Zhang, X. ; Sun, H. Vers une fusion robuste de caméras LiDAR dans l'espace BEV via une attention mutuelle déformable et agrégation temporelle. IEEETrans. Système de circuits. Technologie vidéo. 2024, 34, 5753-5764. [Référence croisée]
8. Chen, CC; Wei, CC; Chen, SH; Soleil, LM ; Lin, HH AI a prédit un modèle de compétences pour maximiser les performances professionnelles. Cybern. Système. 2022, 53, 298-317. [Référence croisée]
9. Kaufmann, E. ; Bauersfeld, L. ; Loquercio, A. ; Müller, M. ; Koltun, V. ; Scaramuzza, D. Courses de drones de niveau champion utilisant l'apprentissage par renforcement profond. Nature 2023, 620, 982-987. [Référence croisée]
10. Uthamacumaran, A. Détection de modèles sur le paysage de Waddington du glioblastome via des réseaux adverses génératifs. Cybern. Système. 2022, 53, 223-237. [Référence croisée]
11. Ning, E. ; Wang, C. ; Zhang, H. ; Ning, X. ; Tiwari, P. Ré-identification des personnes occultées avec l'apprentissage profond : une enquête et des perspectives. Système expert. Appl. 2024, 239, 122419. [Réf. croisée]
12. Li, P. ; Duan, H. Planification du chemin d'un véhicule aérien sans pilote basée sur un algorithme de recherche gravitationnelle amélioré. Sci. Technologie chinoise. Sci. 2012, 55, 2712-2719. [Référence croisée]
13. Qu, C. ; Gai, W. ; Zhang, J. ; Zhong, M. Un nouvel algorithme d'optimisation hybride de loup gris pour la planification de trajectoire de véhicule aérien sans pilote (UAV). Système basé sur les connaissances. 2020, 194, 105530. [Réf. croisée]

14. Dasdemir, E. ; Köksalan, M. ; Öztürk, DT Un algorithme évolutif multi-objectif flexible basé sur des points de référence : une application au problème de planification d'itinéraire des drones. Calculer. Opéra. Rés. 2020, 114, 104811. [Réf. croisée]
15. Yao, P. ; Wang, H. ; Ji, H. Suivi de cibles multi-UAV en environnement urbain par contrôle prédictif de modèle et optimiseur de loup gris amélioré. Aérosp. Sci. Technologie. 2016, 55, 131-143. [Référence croisée]
16. Papaioannou, S. ; Laoudias, C. ; Kolios, P. ; Théocharides, T. ; Panayiotou, CG Estimation et contrôle conjoints pour la surveillance passive multi-cibles avec un agent UAV autonome. Dans Actes de la 31e Conférence méditerranéenne sur le contrôle et l'automatisation (MED) 2023, Limassol, Chypre, 26-29 juin 2023 ; pp. 176-181.
17. Ren, Q. ; Yao, Y. ; Yang, G. ; Zhou, X. Planification de trajectoire multi-objectifs pour les drones en environnement urbain basée sur CDNSGA-II. Dans Actes de la conférence internationale IEEE 2019 sur l'ingénierie des systèmes orientés services (SOSE), San Francisco, Californie, États-Unis, 4-9 avril 2019 ; pp. 350-3505.
18. Price, KV Une introduction à l'évolution différentielle. Dans Nouvelles idées en optimisation ; McGraw : Berkshire, Royaume-Uni, 1999 ; pp. 79-108.
19. Gad, AG Algorithme d'optimisation des essaims de particules et ses applications : une revue systématique. Cambre. Calculer. Méthodes Ing. 2022, 29, 2531-2561. [Référence croisée]
20. Yu, X. ; Jiang, N. ; Wang, X. ; Li, M. Un algorithme hybride basé sur l'optimiseur de loup gris et l'évolution différentielle pour la planification de trajectoire de drone. Système expert. Appl. 2023, 215, 119327. [Réf. croisée]
21. Liu, J. ; Wei, X. ; Huang, H. Un algorithme amélioré d'optimisation du loup gris et son application dans la planification de chemins. Accès IEEE 2021, 9, 121944-121956. [Référence croisée]
22. Xue, J. ; Shen, B. ; Pan, A. Un algorithme de recherche intensifié de moineaux pour résoudre des problèmes d'optimisation. J. Ambiante. Intell. Humaniser. Calculer. 2023, 14, 9173-9189. [Référence croisée]
23. Sahingoz, OK Planification de trajectoire pilotable pour un système multi-UAV avec algorithmes génétiques et courbes de Bézier. Dans Actes de la Conférence internationale 2013 sur les systèmes d'aéronefs sans pilote (ICUAS), Atlanta, GA, États-Unis, 28-31 mai 2013 ; p. 41-48.
24. Cui, Z. ; Zhang, J. ; Wu, D. ; Cai, X. ; Wang, H. ; Zhang, W. ; Chen, J. Algorithme hybride d'optimisation d'essaim de particules à plusieurs objectifs pour le problème de production de charbon vert. Inf. Sci. 2020, 518, 256-271. [Référence croisée]
25. Zhang, J. ; Xue, F. ; Cai, X. ; Cui, Z. ; Chang, Y. ; Zhang, W. ; Li, W. Protection de la vie privée basée sur un algorithme d'optimisation à plusieurs objectifs. D'accord. Calculer. Entraînez-vous. Exp. 2019, 31, e5342. [Référence croisée]
26. Gong, D. ; Soleil, J. ; Miao, Z. Un algorithme génétique basé sur des ensembles pour les problèmes d'optimisation d'intervalles à plusieurs objectifs. IEEETrans. Évol. Calculer. 2016, 22, 47-60. [Référence croisée]
27. Yang, S. ; Li, M. ; Liu, X. ; Zheng, J. Un algorithme évolutif basé sur une grille pour l'optimisation à plusieurs objectifs. IEEETrans. Évol. Calculer. 2013, 17, 721-736. [Référence croisée]
28. Cui, Z. ; Chang, Y. ; Zhang, J. ; Cai, X. ; Zhang, W. NSGA-III amélioré avec opérateur de sélection et d'élimination. Essaim Evol. Calculer. 2019, 49, 23-33. [Référence croisée]
29. Hobbie, JG; Gandomi, AH; Rahimi, I. Une comparaison des techniques de gestion des contraintes sur NSGA-II. Cambre. Calculer. Méthodes Ing. 2021, 28, 3475-3490. [Référence croisée]
30. Liu, Y. ; Gong, D. ; Soleil, J. ; Jin, Y. Un algorithme évolutif à objectifs multiples utilisant une stratégie de sélection un par un. IEEETrans. Cybern. 2017, 47, 2689-2702. [Référence croisée] [Pub Med]
31. Gao, X. ; Zhang, H. ; Song, S. Un algorithme évolutif basé sur les propriétés de régularité pour l'optimisation multiobjectif. Essaim Evol. Calculer. 2023, 78, 101258. [Réf. croisée]
32. Bao, C. ; Gao, D. ; Gu, W. ; Xu, L. ; Goodman, ED Un nouvel algorithme évolutif basé sur la décomposition adaptative pour les optimisation à plusieurs objectifs. Système expert. Appl. 2023, 213, 119080. [Réf. croisée]
33. Liu, S. ; Lin, Q. ; Wong, KC ; Coello, CAC ; Li, J. ; Ming, Z. ; Zhang, J. Une stratégie de vecteur de référence autoguidée pour plusieurs objectifs optimisation. IEEETrans. Cybern. 2020, 52, 1164-1178. [Référence croisée]
34. Yi, J. ; Zhang, W. ; Bai, J. ; Zhou, W. ; Yao, L. Algorithme évolutif multifactoriel basé sur une décomposition dynamique améliorée pour des problèmes d'optimisation à plusieurs objectifs. IEEETrans. Évol. Calculer. 2021, 26, 334-348. [Référence croisée]
35. Nondy, J. ; Gogoi, TK Comparaison des performances d'algorithmes évolutifs multi-objectifs pour l'exergétique et l'exergoenvironnementomique optimisation d'un système de cogénération de référence. Énergie 2021, 233, 121135. [CrossRef]
36. Cai, X. ; Hu, Z. ; Chen, J. Un algorithme de recommandation d'optimisation à plusieurs objectifs basé sur l'exploration de connaissances. Inf. Sci. 2020, 537, 148-161. [Référence croisée]
37. Cai, T. ; Wang, H. Une méthode générale d'analyse de convergence pour un algorithme d'optimisation multi-objectif évolutif. Inf. Sci. 2024, 663, 120267. [Référence croisée]
38. Davies, L. ; Bolam, RC; Vagapov, Y. ; Anuchin, A. Examen des technologies des systèmes d'avions sans pilote pour permettre des opérations au-delà de la ligne de vue visuelle (BVLOS). Dans les actes de la Xe Conférence internationale 2018 sur les systèmes d'entraînement électrique (ICEPDS), Novotcherkassk, Russie, 3-6 octobre 2018 ; p. 1 à 6.
39. Mueller, M. ; Forgeron, N. ; Ghanem, B. Une référence et un simulateur pour le suivi des drones. Dans Actes de Computer Vision – ECCV 2016 : 14e Conférence européenne, Amsterdam, Pays-Bas, 11-14 octobre 2016 ; Actes, partie I 14 ; Springer : Berlin/Heidelberg, Allemagne, 2016 ; pp. 445-461.
40. Wu, Y. ; Lim, J. ; Yang, MH Suivi d'objets en ligne : une référence. Dans les actes de la conférence IEEE sur la vision par ordinateur et Pattern Recognition, Portland, OR, États-Unis, 23-28 juin 2013 ; pages 2411 à 2418.
41. Henriques, JF; Caseiro, R. ; Martins, P. ; Batista, J. Suivi à grande vitesse avec filtres de corrélation Kernelisés. IEEETrans. Modèle Anal. Mach. Intell. 2015, 37, 583-596. [Référence croisée]

42. Li, Y. ; Zhu, J. ; Hoi, Caroline du Sud ; Chanson, W. ; Wang, Z. ; Liu, H. Estimation robuste de la transformation de similarité pour le suivi d'objets visuels. Dans Actes de la conférence AAAI sur l'intelligence artificielle, Honolulu, HI, États-Unis, 27 janvier-1er février 2019 ; Volume 33, pages 8666 à 8673.

43. Wang, N. ; Zhou, W. ; Tian, Q. ; Hong, R. ; Wang, M. ; Li, H. Filtres de corrélation multi-repères pour un suivi visuel robuste. Dans Actes de la conférence IEEE sur la vision par ordinateur et la reconnaissance de formes, Salt Lake City, UT, États-Unis, 18-23 juin 2018 ; pages 4844 à 4853.
44. Bertinetto, L. ; Valmadre, J. ; Golodetz, S. ; Miksik, O. ; Torr, PH Staple : Apprenants complémentaires pour un suivi en temps réel. Dans Actes de la conférence IEEE sur la vision par ordinateur et la reconnaissance de formes, Las Vegas, NV, États-Unis, 27-30 juin 2016 ; pages 1401 à 1409.

45. Danelljan, M. ; Hager, G. ; Khan, FS ; Felsberg, M. Suivi spatial à l'échelle discriminante. IEEETrans. Modèle Anal. Mach. Intell. 2016, 39, 1561-1575. [Référence croisée]
46. Li, Y. ; Fu, C. ; Ding, F. ; Huang, Z. ; Lu, G. AutoTrack : Vers un suivi visuel performant pour drones avec régularisation spatio-temporelle automatique. Dans Actes de la conférence IEEE/CVF sur la vision par ordinateur et la reconnaissance de formes, Seattle, WA, États-Unis, 17-21 juin 2020 ; pp. 11923-11932.

électronique

MDPI

Article

Optimisation du placement des capteurs et des techniques d'apprentissage automatique pour une classification précise des gestes de la main

Lakshya Chaplot [1], Sara Houshmand, [1] Karla Beltrán Martinez [1], John Andersen [2,3] et Hossein Rouhani [1,*]

[1] Département de génie mécanique, Université de l'Alberta, Edmonton, AB T6G 2R3, Canada ; lakshyachaplot@gmail.com (LC) ; shoushma@ualberta.ca (SH); Beltranm@ualberta.ca (KBM)
[2] Département de pédiatrie, Université de l'Alberta, Edmonton, AB T6G 2R3, Canada ; john.andersen@albertahealthservices.ca 3
Glenrose Rehabilitation Hospital, Edmonton, AB T5G 0B7, Canada *
Correspondance : hrouhani@ualberta.ca

Résumé : Des millions de personnes vivent avec des amputations des membres supérieurs, ce qui en fait des bénéficiaires potentiels de prothèses de main et de bras. Si les prothèses myoélectriques ont évolué pour répondre aux besoins des amputés, des défis restent liés à leur contrôle. Cette recherche exploite des capteurs d'électromyographie de surface et des techniques d'apprentissage automatique pour classer cinq gestes fondamentaux de la En utilisant des fonctionnalités extraites des données d'électromyographie, nous avons utilisé un classificateur de machine à vecteurs de support non linéaire, basé sur l'apprentissage à plusieurs noyaux, pour la reconnaissance des gestes. Notre ensemble de données comprenait huit jeunes participants non handicapés. De plus, notre étude a mené une analyse comparative de cinq configurations distinctes de placement de capteurs. Ces configurations capturent les données d'électromyographie associées aux mouvements de l'index et du pouce, ainsi qu'aux mouvements de l'index et de l'annulaire . Nous avons également comparé quatre classificateurs différents pour déterminer celui qui est le plus capable de classer les gestes de la main. La configuration à double capteur stratégiquement placée pour capturer les mouvements du pouce et de l'index était la plus efficace : cette configuration à double capteur a atteint une précision de 90 % pour classer les cinq gestes à l'aide du classificateur de machine à vecteurs de support. De plus, l'application de l'apprentissage à noyaux multiples au sein du classificateur de machines à vecteurs de support démontre son efficacité, atteignant la précision de classification la plus élevée parmi tous les classificateurs. Cette étude a montré le potentiel des capteurs d'électromyographie de surface et de l'apprentissage automatique pour améliorer le contrôle et la fonctionnalité des prothèses myoélectriques pour les personnes amputées des membres supérieurs.

Mots clés : capteur myoélectrique ; geste de la main; machine à vecteurs de support ; main prothétique; classification; apprentissage automatique

Référence : Chaplot, L. ; Houshmand, S. ; Martinez, KB ; Andersen, J. ; Rouhani, H. Optimisation du placement des capteurs et des techniques d'apprentissage automatique pour Classification précise des gestes de la main. Électronique 2024, 13, 3072. https://doi.org/10.3390/electronics13153072

Rédacteur académique : Janos Botzheim

Reçu : 7 juin 2024
Révisé : 24 juillet 2024
Accepté : 1er août 2024
Publié : 3 août 2024

1. Introduction

En 2017, le nombre mondial d'amputés unilatéraux des membres supérieurs dépassait 11,3 millions, avec 11,0 millions de personnes supplémentaires ayant subi une amputation bilatérale des membres supérieurs [1]. Au Canada, environ 6 800 personnes vivent avec une amputation proximale au poignet [2]. Une étude récente [2] a comparé les résultats d'utilité et les coûts associés à deux interventions pour traiter les amputations de la main : l'allotransplantation composite vascularisée de la main et les prothèses myoélectriques de la main. La conclusion était que le traitement des amputations unilatérales avec des prothèses myoélectriques était plus rentable.

Les prothèses de mains myoélectriques sont apparues comme une voie essentielle pour restaurer les capacités gestuelles et préhensiles chez les amputés des membres supérieurs, offrant une alternative non invasive aux interventions chirurgicales permanentes [2]. Les systèmes de contrôle répandus dans les prothèses utilisent souvent un mécanisme basé sur un déclencheur reposant sur un ou deux canaux d'électromyographie de surface (EMG) [3,4]. Cette configuration mappe des événements de contraction musculaire uniques à des séquences de mouvements prédéfinies, nécessitant des commandes utilisateur explicites pour le changement de mode [3,4]. Cette commutation séquentielle introduit une latence dans les temps d'exécution, nécessitant plusieurs commandes distinctes pour passer d'un mode de préhension à l'autre [3,4]. Le

ce processus de changement de préhension, associé à une dynamique de contrôle maladroite et à un manque de rétroaction suffisante, a été identifié comme l'un des principaux contributeurs aux faibles taux d'acceptation observés pour les prothèses myoélectriques [5,6]. Relever ces défis est essentiel pour améliorer la convivialité et l'acceptation des dispositifs prothétiques myoélectriques au sein de la communauté des utilisateurs [5,6].

Plusieurs tentatives visant à classer les signaux sEMG (électromyographie de surface) provenant des muscles de l'avant-bras humain ont été documentées dans des travaux antérieurs. Pour atténuer les problèmes d'intuitivité, les solutions prototypes de la littérature existante se concentrent sur le déchiffrement des intentions gestuelles de l'utilisateur en ciblant les muscles fléchisseurs et extenseurs distincts de l'avant-bras [7]. Les contractions volontaires des muscles restants de l'avant-bras après l'amputation peuvent être identifiées grâce à des classificateurs d'apprentissage automatique, tels que les classificateurs de réseaux neuronaux artificiels (ANN), d'analyse discriminante linéaire (LDA) et de machines à vecteurs de support (SVM). SVM est souvent choisi en raison de son interprétabilité mathématique et de son optimisation globale. Il fonctionne bien même avec un petit ensemble d'entraînement [8]. Le paramètre d'élasticité de SVM, également connu sous le nom d' hyperparamètre de contrainte de boîte C, contrôle la pénalité maximale imposée aux observations violant les marges et aide à prévenir le surajustement [8]. Palkowski et Redlarski [9] ont utilisé deux capteurs EMG sur l'avant-bras, échantillonnant des données à 16 Hz, pour discerner six gestes entiers de la main et du poignet à l'aide d'un classificateur SVM. Lee et coll. [10] ont réussi à classifier dix gestes de la main en utilisant les caractéristiques obtenues par trois capteurs EMG et ont atteint une précision supérieure à 90 % pour chaque participant. Cependant, leur modèle d'apprentissage automatique a subi une formation et des tests sur des ensembles de données de participants sans fusion de données entre participants pour évaluer la capacité de généralisation du modèle. Cette approche de test restrictive introduit un biais dans l'exactitude de la classification, soulevant des questions sur l'applicabilité de cette technologie aux mains prothétiques. D'autres travaux antérieurs [8,11,12] ont atteint une précision élevée de plus de 90 % pour classer plusieurs gestes de la main et du poignet entiers, comme l'ouverture/fermeture du poignet, la déviation ulnaire et radiale et la flexion-extension. Cependant, les gestes de la main entière et du poignet sont moins difficiles à classer et offrent une application fonctionnelle limitée pour les amputés des membres supérieurs cherchant à restaurer leur dextérité manuelle.

L'efficacité du ciblage de muscles spécifiques nécessite le déploiement stratégique des électrodes et leur configuration – une considération essentielle pour les approches basées sur les caractéristiques qui sont relativement inexplorées dans la littérature.

Ce projet visait à améliorer la précision de la classification des gestes de la main, le placement stratégique des capteurs et la sélection de gestes pratiques pour une intégration potentielle avec les mains prothétiques myoélectriques. Par conséquent, cette étude a examiné le développement d'un classificateur SVM basé sur l'apprentissage à noyaux multiples (MKL) pour classifier cinq gestes de la main complexes cruciaux pour les amputés : la prise puissante (serrement du poignet), la prise en crochet (saisie à quatre chiffres), le pincement fin (en utilisant l'index et le pouce), un pincement grossier (en utilisant les cinq chiffres) et un geste de pointage (flexion des chiffres 3, 4 et 5). La méthodologie de recherche impliquait la construction du classificateur à l'aide de données collectées dans cinq configurations de capteurs distinctes, chacune utilisant un ou deux capteurs EMG sur l'avant-bras vers une approche minimaliste de collecte de données, tout en s'efforçant d'identifier le placement optimal du capteur pour obtenir la plus grande précision de classif

2. Matériels et méthodes

2.1. Les données sEMG

de la procédure expérimentale ont été collectées sur la main droite de huit participants sans déficience neurologique/musculo-squelettique ni diagnostic (âge : 21 ± 2 ans (moyenne ± écart-type), taille : 169,5 ± 2,8 cm (moyenne ± écart- type), poids corporel : 57,9 ± 9,5 kg (moyenne ± ET), 7 mâles et 1 femelle). Tous les participants connaissaient les procédures expérimentales. Le consentement éclairé a été obtenu de tous les sujets impliqués dans l'étude. L'étude a été menée conformément à la Déclaration d'Helsinki et approuvée par le comité d'éthique institutionnel de l'Université de l'Alberta (AB T6G 2N2, approuvé le 25 avril 2022).

Les signaux sEMG ont été acquis à l'aide de capteurs musculaires bipolaires MyoWare 2.0 [13] (SparkFun Electronics, Niwot, CO, USA) choisis pour leur potentiel d'intégration dans des prothèses de main à faible coût. La version précédente de ce capteur a été fréquemment utilisée dans

Les signaux sEMG ont été acquis à l'aide de capteurs musculaires bipolaires MyoWare 2.0 [13] (SparkFun Electronics, Niwot, CO, USA) choisis pour leur potentiel d'intégration dans des prothèses de main à faible coût. La version précédente de ce capteur a été fréquemment utilisée dans la littérature en raison de son faible coût, de ses fonctionnalités faciles à personnaliser et de ses rapports de performances favorables dans les études de validation, ce qui montre qu'elle est comparable à des produits plus coûteux . son faible coût, ses fonctionnalités faciles à personnaliser et ses performances favorables.

systèmes EMG commerciaux [14,15]. MyoWare 2.0 est doté de trois électrodes : mi-musculaire et fin

musculaire – rapportent les études de validation, montrant qu'il est comparable aux produits commerciaux plus chers.

clé et référence [13]. Avant d'être acquis par le microcontrôleur, le signal différentiel

Systèmes EMG [14,15]. MyoWare 2.0 comporte trois électrodes : celle du muscle médian, celle de l'extrémité du muscle et

passe à travers un amplificateur d'instrumentation avec un CMRR élevé (référence de rejection de mode commun [13]. Avant d'être acquis par le microcontrôleur, le signal différentiel passait

ratio, qui est le rapport entre le gain différentiel et le gain en mode commun de l'étage amplificateur) via un

amplificateur d'instrumentation avec un CMRR élevé (taux de réjection en mode commun.

(140 dB) et gain unitaire pour éliminer les sources de bruit courantes, telles que le bruit de ligne 50 Hz. qui

est le rapport entre le gain différentiel et le gain en mode commun de l'étage amplificateur) (140 dB)

Par la suite, le signal a été filtré par un filtre passe-bande de premier ordre avec fréquence de coupure et

gain unitaire pour éliminer les sources de bruit courantes, telles que le bruit de ligne à 50 Hz. Sous-séries à

20 Hz et 498 Hz [13]. Suite à cela, le signal a subi une rectification et peu à peu, le signal a été filtré par un

filtre passe-bande du premier ordre avec des fréquences de coupure à

lissage, obtenu par un circuit de détection d'enveloppe passe-bas (3,6 Hz) intégré dans les 20 Hz et 498 Hz [13]. Suite à cela, le signal a subi une rectification et un lissage,

matériel du capteur, ce qui entraîne une enveloppe linéaire du signal EMG. L'EMG enveloppé obtenu par un

circuit de détection d'enveloppe passe-bas (3,6 Hz) intégré dans la sortie matérielle du capteur a ensuite été

acquis par le microcontrôleur (Arduino UNO) avec un logiciel d'échantillonnage , ce qui a donné une

enveloppe linéaire du signal EMG. La sortie EMG enveloppée était

fréquence de 780 Hz (SD = 5 Hz) pour répondre aux problèmes de sous-échantillonnage précédemment

acquis par le microcontrôleur (Arduino UNO) avec une fréquence d'échantillonnage de 780 Hz

études [16].

(SD = 5 Hz) pour répondre aux préoccupations concernant le sous-échantillonnage dans les études précédentes [16].

Les configurations de placement du capteur EMG étaient basées sur les gestes destinés à être Les

configurations de placement du capteur EMG étaient basées sur les gestes destinés à être

classifié. Les capteurs ont été placés uniquement le long des muscles de l'avant-bras les plus responsables .

Les capteurs ont été placés uniquement le long des muscles de l'avant-bras principalement responsables

pour la flexion de l'index, de l'annulaire et du pouce. Les cinq gestes utilisés étaient (1) un pour la flexion de

l'index, de l'annulaire et du pouce. Les cinq gestes utilisés étaient (1) un

pincement grossier en utilisant les cinq chiffres, (2) un pincement fin avec l'index et le pouce, (3) un crochet

pincement grossier en utilisant les cinq chiffres, (2) un pincement fin avec l'index et le pouce, (3) un crochet

(4) geste de point (flexion des chiffres 3, 4 et 5), et (5) une prise puissante (prise à quatre chiffres), (4) geste

de point (flexion des chiffres 3, 4 et 5), et (5) une prise de pouvoir

saisir (serrement du poignet), étiqueté respectivement de 0 à 4 pour le classificateur. Ces cinq (serrement du

poignet), étiquetés respectivement de 0 à 4, pour le classificateur. Ces cinq gestes

les gestes (Figure 1) sont considérés comme ayant une grande utilité pour les personnes amputées dans leur vie quotidienne et

(Figure 1) sont considérés comme ayant une grande utilité pour les personnes amputées dans leur vie quotidienne et sont également

sont également observés dans les prothèses commerciales

[3,4]. observé dans les prothèses commerciales [3,4].

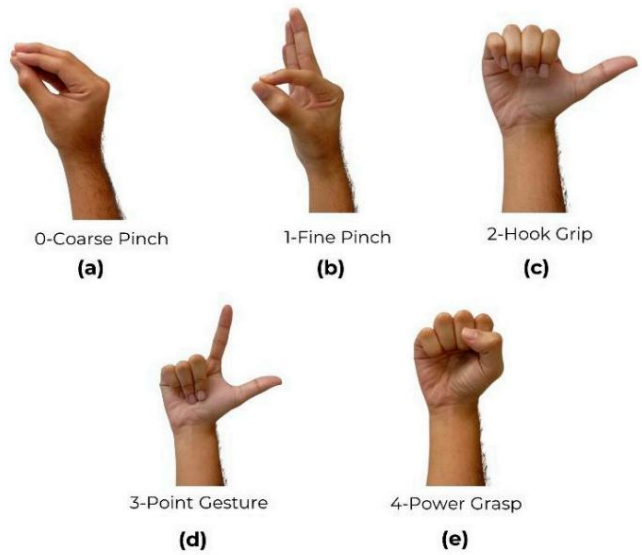


Figure 1. Les cinq gestes différents avec leurs étiquettes de prédiction correspondantes : (a) pincement grossier,

Figure 1. Les cinq gestes différents avec leurs étiquettes de prédiction correspondantes : (a) pincement grossier, (b) pincement fin, (c) prise en crochet, (d) geste de point et (e) prise puissante.

(b) pincement fin, (c) prise en crochet, (d) geste de point et (e) prise puissante.

Les capteurs EMG ont été placés à trois endroits sur la face ventrale de l'avant-bras.

Les capteurs EMG ont été placés à trois endroits sur la face ventrale de l'avant-bras (Figure 2) pour cinq

configurations différentes afin d'acquérir les signaux sEMG des muscles principalement responsables de la

configurations différentes pour acquérir les signaux sEMG des muscles principalement responsables de la

flexion de l'index, de l'annulaire et du pouce. Le capteur

Il est responsable de la flexion de l'index, de l'annulaire et du pouce. Les configurations des capteurs sont basées sur [17–19]

et sont décrites comme suit :

les figurations étaient basées sur [17–19] et sont décrites comme suit : C1 : un

- C1 : un capteur placé à proximité du poignet le long du long fléchisseur du pouce pour acquérir

capteur placé à proximité du poignet le long du long fléchisseur du pouce pour acquérir

de flexion du pouce.

les données de flexion du

- pouce. • C2 : un capteur placé à proximité du poignet le long du fléchisseur superficiel des orteils pour

C2 : un capteur placé à proximité du poignet le long du fléchisseur superficiel des doigts pour acquérir

les données de flexion de l'index.

acquérir les données de flexion de l'index. •

C3 : un capteur placé le long du fléchisseur superficiel des orteils et du fléchisseur des orteils

profundus pour acquérir des données de flexion de l'annulaire.

- C4 : deux capteurs placés à proximité du poignet le long du fléchisseur superficiel des orteils pour

acquérir des données de flexion de l'index et proximal du coude le long du fléchisseur des orteils

superficiel et fléchisseur profond des doigts pour acquérir des données de flexion de l'annulaire.

Machine Translated by Google

- C3 : Un capteur placé le long du fléchisseur superficiel des orteils et du fléchisseur des orteils. fond d'œil pour acquérir des données de flexion de l'annulaire.
- C4 : Deux capteurs placés à proximité du poignet le long du fléchisseur superficiel des doigts pour acquérir les données de flexion de l'index et à proximité du coude le long du fléchisseur superficiel des doigts et du fléchisseur profond des doigts pour acquérir les données de flexion de l'annulaire. • C5 : deux capteurs placés à proximité du poignet le long du fléchisseur superficiel des orteils. C5 : deux capteurs placés à proximité du poignet le long du fléchisseur superficiel des orteils.
- pour acquérir les données de flexion de l'index et le long du long fléchisseur du pouce pour acquérir les données de flexion de l'index et le long du long fléchisseur du pouce pour acquérir le données de flexion du pouce.

Il a été demandé aux participants d'effectuer chaque geste de manière répétée 20 fois, avec un temps. Il a été demandé aux intervalle de deux secondes entre chaque répétition. Après avoir effectué un cycle de données, intervalle de deux secondes entre chaque répétition. Après avoir effectué une série de données pré-collecte pour un seul geste, le participant a pris 4 minutes de repos pour détendre ses muscles avant de pré- collecter pour un seul geste, le participant a pris 4 minutes de repos pour détendre ses muscles évacuer la fatigue avant de commencer le tour suivant pour un nouveau geste. Les signaux mesurés évitent la fatigue avant d'entamer le tour suivant pour un nouveau geste. Les signaux mesurés des deux capteurs EMG ont été enregistrés simultanément. des deux capteurs EMG ont été enregistrés simultanément.

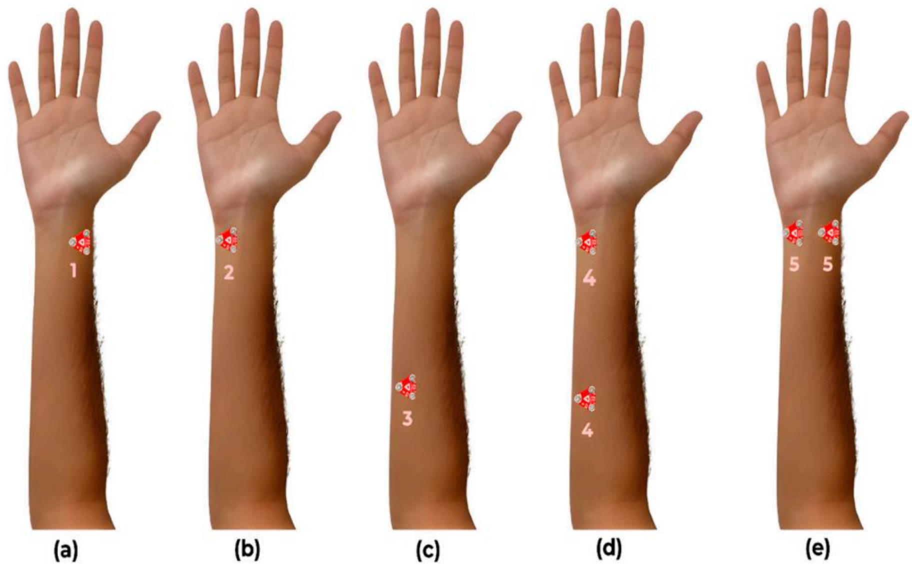


Figure 2. Les cinq configurations différentes de placement du capteur utilisées pour acquérir des données EMG : (a) pouce Figure 2. Les cinq configurations différentes de placement du capteur utilisées pour acquérir des données EMG : (a) pouce données de flexion, (b) données de flexion de l'index, (c) données de flexion de l'annulaire, (d) données de flexion de l'index et de l'anneau, données de flexion, (b) données de flexion de l'index, (c) données de flexion de l'annulaire, (d) données de flexion d'index et d'anneau, et (e) les données de flexion de l'index et du pouce. et (e) les données de flexion de l'index et du pouce.

2.2. Traitement des données

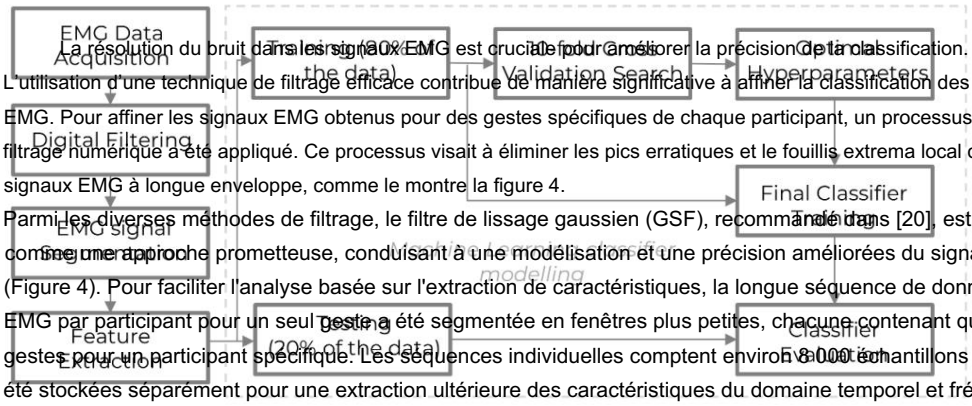
2.2. Traitement de l'information

La figure 3 montre un organigramme des étapes de traitement des données impliquées dans l'ensemble de l'expérience et du pipeline de modélisation. Tout d'abord, les données EMG ont été enregistrées par des capteurs et filtrées par un pipeline de modélisation. Premièrement, les données EMG ont été enregistrées par des capteurs et filtrées pour éliminer le bruit. Deuxièmement, les signaux ont été segmentés en séquences plus courtes et le temps a éliminé le bruit. Deuxièmement, les signaux ont été segmentés en séquences plus courtes, et le temps et les caractéristiques du domaine fréquentiel ont été extraites pour capturer les caractéristiques distinctes des caractéristiques du domaine fréquentiel ont été extraites pour capturer les caractéristiques distinctes de l'EMG Données EMG. Troisièmement, les données traitées ont été utilisées pour la classification des gestes de la main à l'aide de ma- data. Troisièmement, les données traitées ont été utilisées pour la classification des gestes de la main à l'aide d'une machine. Chine Learning, où une répartition de 80 à 20 trains-tests a été effectuée pour entraîner et évaluer l' apprentissage, où une répartition de 80 à 20 trains-tests a été effectuée pour entraîner et évaluer les modèles. des modèles. Quatrièmement, les hyperparamètres des modèles d'apprentissage automatique pour l'ensemble de données donné. Quatrièmement, les hyperparamètres des modèles d'apprentissage automatique pour l'ensemble de données donné étaient ont été réglés à l'aide d'une méthode de validation croisée de recherche sur grille 10 fois pour améliorer les performances. optimisé à l'aide d'une méthode de validation croisée de recherche sur grille 10 fois pour améliorer les performances. Cinquièmement, Cinquièmement, les classificateurs ont été formés à l'aide de l'ensemble de données de formation complet et de ces hyperparamètres optimisés. Sixièmement, les performances du modèle ont été évaluées à l'aide de données de test invisibles et de la métrique paramètres. Sixièmement, les performances du modèle ont été évaluées à l'aide de données de test invisibles et de la métrique d'évaluation (le score d'exactitude) pour refléter l'efficacité du modèle en classe. La métrique d'évaluation (le score d'exactitude) reflète l'efficacité du modèle dans la classification des données EMG. les données EMG.

La résolution du bruit dans les signaux EMG est cruciale pour améliorer la précision de la classification. L'utilisation d'une technique de filtrage efficace contribue de manière significative à affiner le signal EMG classification. Pour affiner les signaux EMG obtenus pour des gestes spécifiques de chaque participant , un processus de filtrage numérique a été appliqué. Ce processus visait à éliminer les pics erratiques et le fouillis extrema local provenant des signaux EMG à longue enveloppe, comme le montre la figure 4. Parmi les différentes méthodes de filtrage, le filtre de lissage gaussien (GSF), recommandé dans [20], est apparue comme une approche prometteuse, conduisant à une modélisation améliorée du signal EMG et la précision (Figure 4). Pour faciliter l'analyse basée sur l'extraction de caractéristiques, les données EMG longues La séquence par participant pour un seul geste a été segmentée en fenêtres plus petites, chacune contenant quatre gestes pour un participant spécifique. Les séquences individuelles sont d'environ

8 000 échantillons de long et ont été stockés séparément pour les domaines temporel et fréquentiel ultérieurs extraction de caractéristiques.

Figure 3. Organigramme de traitement des données EMG.



La résolution du bruit dans les signaux EMG est cruciale pour améliorer la précision de la classification.
L'utilisation d'une technique de filtrage efficace contribue de manière significative à affiner la classification des signaux EMG. Pour affiner les signaux EMG obtenus pour des gestes spécifiques de chaque participant, un processus de filtrage numérique a été appliqué. Ce processus visait à éliminer les pics erratiques et le fouillis extrema local des signaux EMG à longue enveloppe, comme le montre la figure 4.

Parmi les diverses méthodes de filtrage, le filtre de lissage gaussien (GSF), recommandé dans [20], est apparu comme une approche prometteuse, conduisant à une modélisation et une précision améliorées du signal EMG (Figure 4). Pour faciliter l'analyse basée sur l'extraction de caractéristiques, la longue séquence de données EMG par participant pour un seul geste a été segmentée en fenêtres plus petites, chacune contenant quatre gestes pour un participant spécifique. Les séquences individuelles comptent environ 8 000 échantillons et ont été stockées séparément pour une extraction ultérieure des caractéristiques du domaine temporel et fréquentiel.

Figure 3. Organigramme de traitement des données EMG.

La résolution du bruit dans les signaux EMG est cruciale pour améliorer la précision de la classification.
L'utilisation d'une technique de filtrage efficace contribue de manière significative à affiner la classification des signaux EMG. Pour affiner les signaux EMG obtenus pour des gestes spécifiques de chaque participant, un processus de filtrage numérique a été appliqué. Ce processus visait à éliminer les pics erratiques et le fouillis extrema local des signaux EMG à longue enveloppe, comme le montre la figure 4.

Parmi les diverses méthodes de filtrage, le filtre de lissage gaussien (GSF), recommandé dans [20], est apparu comme une approche prometteuse, conduisant à une modélisation et une précision améliorées du signal EMG (Figure 4). Pour faciliter l'analyse basée sur l'extraction de caractéristiques, la longue séquence de données EMG par participant pour un seul geste a été segmentée en fenêtres plus petites, chacune contenant quatre gestes pour un participant spécifique. Les séquences individuelles comptent environ 8 000 échantillons et ont été stockées séparément pour une extraction ultérieure des caractéristiques du domaine temporel et fréquentiel.

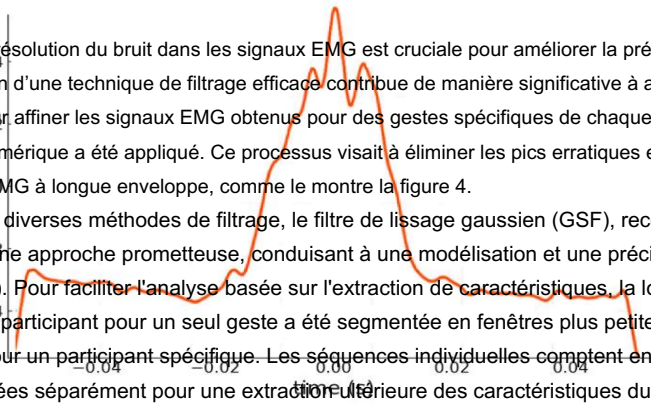


Figure 4. Signal EMG filtré pour le geste de saisie de puissance (exemple illustré pour la configuration C3 et le geste de saisie du pouvoir).

Figure 4. Signal EMG filtré pour le geste de saisie de puissance (exemple présenté pour la configuration C3 et les transformations étaient le geste de saisie du pouvoir).

2.3. Extraction et sélection de fonctionnalités

2.3. Extraction et sélection de fonctionnalités

L'extraction de fonctionnalités est cruciale pour réduire la dimensionnalité des données représentant un geste, permettant ainsi la classification des gestes. Au total, 22 caractéristiques du domaine temporel et fréquentiel ont été extraites des 8 000 échantillons de signaux EMG segmentés en long pour chaque geste. Les fonctionnalités comprenaient des mesures statistiques telles que la moyenne quadratique (RMS), variance (VAR), valeur absolue moyenne (MAV), changement de signe de pente (SSC), passages par zéro (ZC), longueur de forme d'onde (WL), écart absolu médian (MAD), asymétrie, aplatissement et énergie. Ces caractéristiques sont bien établies dans la littérature pour leur utilité dans la reconnaissance gestuelle [21,22]. Les fonctionnalités du domaine temps-fréquence incluent un modèle autorégressif (5ème ordre), l'entropie (basée sur des coefficients approximatifs d'ondelettes discrètes 1D transformée du niveau 4), estimations de la variance (basées sur les ondelettes discrètes à chevauchement maximum transformée de niveau 3), entropie spectrale, fréquence moyenne et puissance de bande. L'ondelette basées sur l'ondelette Daubechies-2. Les fonctionnalités ci-dessus ont été utilisées

L'extraction de fonctionnalités est cruciale pour réduire la dimensionnalité des données représentant un geste, permettant ainsi la classification des gestes. Un total de 22 heures et fréquences
les caractéristiques du domaine ont été extraites des 8 000 échantillons de signaux EMG segmentés longuement pour chaque geste. Les fonctionnalités comprenaient des mesures statistiques telles que la moyenne quadratique (RMS), variance (VAR); valeur absolue moyenne (MAV), changement de signe de pente (SSC), passages par zéro (ZC), (ZC), longueur de la forme d'onde (WL), déviation absolue médiane (MAD), asymétrie, aplatissement et énergie . Ces caractéristiques sont bien établies dans la littérature pour leur utilité dans la reconnaissance des gestes
nition [21,22]. Les caractéristiques du domaine temps-fréquence incluent les coefficients du modèle autorégressif

Figure 4. Signal EMG filtré pour le geste de saisie de puissance (exemple présenté pour la configuration C3 et les transformations étaient

le geste de saisie du
pouvoir). plusieurs fois dans la littérature [10,23,24]. Pour faciliter une classification efficace, les extraits

les caractéristiques ont été normalisées à une moyenne de zéro en soustrayant la moyenne de chaque échantillon
2.3. Extraction et sélection de fonctionnalités
et normalisation à l'aide de l'écart type.

L'extraction de fonctionnalités est cruciale pour réduire la dimensionnalité des données représentant l'analyse en composantes principales (ACP) a été utilisée dans la phase de sélection des fonctionnalités comme élément clé
un geste, permettant ainsi la classification des gestes. Un total de 22 pas de temps et de fréquence pour réduire davantage la dimensionnalité de l'espace des fonctionnalités, dérivé du
les caractéristiques du domaine ont été extraites des 8 000 échantillons de signaux EMG segmentés longs pour les données EMG segmentées obtenues par les deux capteurs. Avec un ensemble de fonctionnalités initial comprenant
chaque geste. Les fonctionnalités comprenaient des mesures statistiques telles que la moyenne quadratique (RMS), 44 fonctionnalités distinctes (22 × 2 capteurs), la PCA a été utilisée pour transformer ces fonctionnalités en un
variance (VAR); valeur absolue moyenne (MAV), changement de signe de pente (SSC), passages par zéro (ZC), nouvel ensemble de composantes principales. L'objectif premier était de retenir le contenu le plus informatif
longueur de la forme d'onde (WL), déviation absolue médiane (MAD), asymétrie, aplatissement et composants tout en réduisant la dimensionnalité. Les principales fonctionnalités capturant plus de 95 % de variabilité
ergie. Ces caractéristiques sont bien établies dans la littérature pour leur utilité dans la reconnaissance gestuelle [21,22]. Les fonctionnalités du domaine temps-fréquence incluent un modèle autorégressif

dans l'ensemble de données ont été méticuleusement sélectionnés. Ce critère de sélection garantissait la conservation uniquement des composants qui contribuaient de manière significative à la variance des données, compressant ainsi l'espace des fonctionnalités tout en préservant ses informations essentielles. Ces composants sélectionnés, qui capturaient le plus de variabilité de l'ensemble de données, ont ensuite été exclusivement introduits dans le classificateur. Cette utilisation stratégique de fonctionnalités dérivées de la PCA, réduites mais informatives, visait à améliorer les performances de classification en se concentrant sur les aspects les plus critiques des données EMG.

2.4. Classificateur d'apprentissage automatique

Dans cet article, SVM a été utilisé pour classer cinq gestes de la main. Comme SVM est une méthode basée sur le noyau, la sélection des fonctions appropriées du noyau et des hyper-paramètres associés est une tâche importante. Ce problème est généralement résolu par une approche par essais et erreurs. De plus, une application SVM typique à noyau unique adopte fréquemment les mêmes hyper-paramètres pour chaque classe, et elle peut ne pas convenir lorsque la distribution des modèles de fonctionnalités est significativement différente entre les différentes classes. Bien qu'il existe différents noyaux, tels que le noyau gaussien, le noyau polynomial et le noyau sigmoïde, il est souvent difficile de savoir quel est le noyau le plus approprié pour un ensemble de données donné, et il est donc souhaitable que les méthodes du noyau utilisent une fonction de noyau optimisée qui s'adapte bien à l'ensemble de données disponible et au type de frontières entre les classes. Un moyen efficace de concevoir un noyau optimal pour un ensemble de données donné consiste à considérer le noyau comme une combinaison convexe de noyaux de base, comme illustré dans l'équation (1).
Un tel SVM basé sur MKL est inspiré de [25].

$$K(x, y) = a \cdot K1(x, y) + b \cdot K2(x, y) + c \cdot K3(x, y) + d \cdot KRBF(x, y) \quad (1)$$

Ici, K1 est le noyau linéaire, K2 est le noyau quadratique, K3 est le noyau cubique et KRBF est le noyau de fonction de base gaussienne ou radiale avec un écart type unitaire.
Les coefficients {a, b, c, d} sont des hyperparamètres à ajuster à l'aide d'une validation croisée de recherche sur grille 10 fois. Le paramètre d'élasticité ou contrainte de boîte C est également un hyperparamètre exprimant le degré de perte de contrainte. Un grand C peut classer les échantillons d'apprentissage plus correctement, mais finit également par un surajustement et une réduction de la précision des tests. Par conséquent, la contrainte de boîte est également ajustée à l'aide d'une validation croisée de recherche de grille 10 fois. La formation et le test des données sont effectués à l'aide d'une répartition typique de 80 à 20 trains-tests. La PCA a été utilisée pour la sélection des caractéristiques afin de réduire davantage la dimensionnalité du vecteur d'entrée donné au classificateur et de choisir uniquement les caractéristiques statistiquement significatives.

Une analyse plus approfondie a consisté à utiliser trois classificateurs d'apprentissage automatique : Bayes naïf, arbre de décision et KNN. Cela a été entrepris pour évaluer les performances de SVM dans un ensemble de données spécifique et des conditions expérimentales, en se concentrant sur la robustesse avec un nombre croissant de gestes et des configurations de capteurs variables. Bien que les SVM soient reconnus comme étant efficaces pour la reconnaissance gestuelle basée sur l'EMG [26,27], la recherche vise à étudier ces résultats dans le cadre de notre configuration unique qui utilise des capteurs sEMG à faible coût disponibles dans le commerce . En confirmant la robustesse et l'efficacité des SVM dans cette application, des preuves supplémentaires seront ajoutées à la théorie établie, en tenant compte des nuances de notre ensemble de données spécifique.

3. Résultats

3.1. Évaluation de la configuration du capteur

Le meilleur ensemble d'hyperparamètres résultant en la plus grande précision sur les données de test pour les classificateurs SVM, Bayes naïfs, KNN et d'arbre de décision, construit à l'aide d'un ensemble de données fusionné de tous les participants dans différentes configurations de capteurs, est présenté dans le tableau 1, tableau 2, tableau 3 et tableau 4, respectivement. Les hyperparamètres associés sont ajustés à l'aide d'une validation croisée de recherche sur grille 10 fois, et les étiquettes utilisées pour chaque geste sont données dans la figure 1.

Tableau 1. Configuration optimale des hyperparamètres pour la classification des gestes basée sur SVM en utilisant 10 fois validation croisée. L'ordre du noyau SVM est l'ordre du noyau polynomial pur. {a, b, c, d} sont les MKL coefficients. C est la contrainte de boîte. C1 , C2 , C3 , C4 , C5 sont les cinq configurations de capteurs différentes.

	Hyperparamètres		Configurations de capteurs				
			C1	C2	C3	C4	C5
Classer cinq gestes dont pincement grossier, pincement fin, prise en crochet, geste de pointage et prise de pouvoir	Ordre du noyau SVM		3	-	-	-	-
	Coefficients MKL	un	-	dix	0,1	0,1	1
		b	-	0,5	1	1	0,001
		c	-	0,1	dix	dix	0
		d	-	0	0	0	0
	C (contrainte de boîte)		1	0,1	0,5	0,5	1
Classer deux gestes dont pincement fin et prise puissante	Ordre du noyau SVM		1	3	1	2	3
	Coefficients MKL	un	-	-	-	-	-
		b	-	-	-	-	-
		c	-	-	-	-	-
		d	-	-	-	-	-
	C (contrainte de boîte)		dix	1	0,5	0,02	1

Tableau 2. Configuration optimale des hyperparamètres pour la classification gestuelle naïve basée sur Bayes à l'aide Validation croisée 10 fois. Le noyau de distribution « N » est le noyau normal et « B » est le noyau boîte (uniforme).

	Hyperparamètres	Configurations de capteurs				
		C1	C2	C3	C4	C5
Classer cinq gestes dont pincement grossier, pincement fin, prise en crochet, geste de pointage et prise de pouvoir	Noyau de distribution (N/B)	N	B	N	N	N
	Bande passante	0,05	0,8	0,08	0,15	0,42
Classer deux gestes dont pincement fin et prise puissante	Noyau de distribution (N/B)	N	N	N	N	N
	Bande passante	0,1	0,05	0,1	0,05	0,05

Tableau 3. Configuration optimale des hyperparamètres pour la classification gestuelle basée sur KNN en utilisant 10 fois validation croisée. D1 à D4 désignent les fonctions de distance euclidienne, cosinus, pâté de maisons et Minkowski.

	Hyperparamètres	Configurations de capteurs				
		C1	C2	C3	C4	C5
Classer cinq gestes dont pincement grossier, pincement fin, prise en crochet, geste de pointage et prise de pouvoir	Fonction distance	D1	D2	D1	D1	D1
	Valeur K	1	1	1	1	2
Classer deux gestes dont pincement fin et prise puissante	Fonction distance	D1	D3	D2	D1	D4
	Valeur K	1	1	1	1	1

Tableau 4. Configuration optimale des hyperparamètres pour la classification des gestes basée sur un arbre de décision en utilisant une validation croisée 10 fois.

	Hyperparamètres	Configurations de capteurs				
		C1	C2	C3	C4	C5
Classer cinq gestes dont pincement grossier, pincement fin, prise en crochet, geste de pointage et prise de pouvoir	Taille minimale des feuilles	1	9	2	2	2
Classer deux gestes dont pincement fin et prise puissante	Taille minimale des feuilles	2	9	9	2	2

Le tableau 5 présente les résultats de performance de différents classificateurs sur deux gestes distincts tâches de classification avec différentes configurations (C1, C2, C3, C4, C5) et fournit un aperçu comparatif de l'exactitude des différents classificateurs dans cinq catégories distinctes et non liées. configurations de capteurs pour les tâches de classification des gestes. En classant cinq gestes, le SVM les performances du classificateur varient de 75 % à 90 %, avec la plus grande précision observée dans configuration C5. La précision du KNN fluctue, atteignant son maximum à 82 % avec C4 et légèrement inférieur à 80% avec C5. Naïve Bayes affiche son meilleur résultat à 70% avec C5, tandis que le le classificateur d'arbre de décision plafonne à 75 % avec la même configuration.

Tableau 5. Précision des tests pour les classificateurs SVM, KNN, Bayes naïfs et d'arbre de décision dans différentes configurations de capteurs. C1 à C5 sont les cinq configurations différentes de capteurs.

	Classificateur	Configuration Ci				
		C1	C2	C3	C4	C5
Classer cinq gestes dont pincement grossier, pincement fin, prise en crochet, geste de pointage et prise de pouvoir	SVM	75%	75%	84,6%	87,2%	90%
	KNN	67,9%	73,9%	69,2%	82%	80%
	Bayes naïfs	53,6%	51,3%	61,5%	61,5%	70%
	Arbre de décision	57,2%	59%	56,4%	59%	75%
Classer deux gestes dont pincement fin et prise puissante	SVM	100%	100%	100%	100%	100%
	KNN	100%	86,7%	100%	93,3%	100%
	Bayes naïfs	100%	73,3%	100%	93,3%	93,3%
	Arbre de décision	100%	86,7%	100%	80%	100%

En revanche, la tâche de classification de deux gestes, le pincement fin et la saisie puissante, voit une performance nettement supérieure et parfaite de la part du classificateur SVM, maintenant une précision de 100 %. dans toutes les configurations. Les classificateurs KNN et d'arbre de décision fonctionnent également exceptionnellement eh bien, KNN atteignant une précision de 100 % dans tous les cas sauf C2 et C4, où il obtient un léger score inférieur à 86,7% et 93,3%, respectivement. Les classificateurs Naïve Bayes et arbres de décision présentent une précision parfaite de 100% pour la configuration C3. Cette différence nette de performances entre les tâches suggère que certains classificateurs, en particulier SVM, peuvent être plus robustes à changements de configurations ou sont mieux adaptés aux tâches de classification binaire dans le cadre de reconnaissance gestuelle.

3.2. Évaluation du classificateur

La matrice de confusion pour le classificateur SVM dans différentes configurations a été méticuleusement tracé pour fournir une visualisation complète des performances du modèle (Figure 5).

La matrice de confusion pour le classificateur SVM dans différentes configurations a été méticuleusement tracé pour fournir une visualisation complète des performances du modèle (Figure 5).

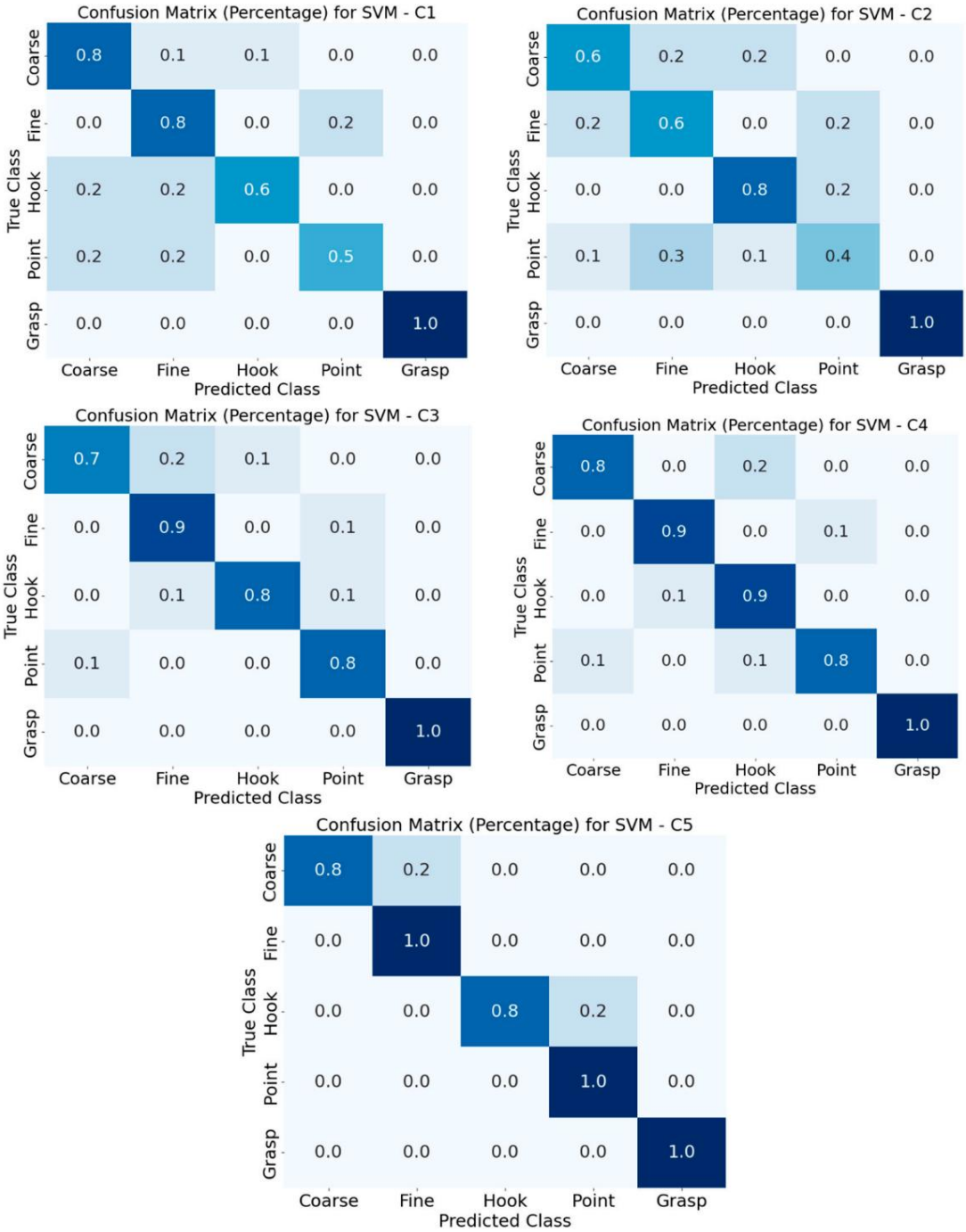


Figure 5. Matrice de confusion pour le classificateur SVM dans différentes configurations.

Figure 5. Matrice de confusion pour le classificateur SVM dans différentes configurations.

4. Discussion

L'étude s'étend au-delà de l'évaluation de l'exactitude de la classification, en approfondissant la complexité. comment classer un éventail croissant de gestes, comme le corroborent des recherches [28]. Cette enquête révèle en outre les limites de l'utilisation d'un seul capteur pour l'acquisition EMG, qui limite le nombre de classes distinctement identifiables. Nous avons évalué l'efficacité de quatre classificateurs d'apprentissage automatique (SVM, Bayes naïf, KNN et décision arbre de configuration : plus de cinq configurations de capteurs distinctes pour collecter des données détaillées sur les gestes de la main des avant-bras de huit participants. Un ensemble complet de 22 fonctionnalités, englobant les domaines temporel et fréquentiel par capteur ont été extraits des signaux EMG. Avant des études suggèrent qu'un ensemble de fonctionnalités à domaines mixtes peut renforcer les performances du classificateur [29].

L'étude s'étend au-delà de l'évaluation de l'exactitude de la classification, en approfondissant les complexités liées à la classification d'un éventail croissant de gestes, comme le corroborent les recherches antérieures. Cette enquête révèle en outre les limites de l'utilisation d'un seul capteur pour l'acquisition EMG, qui limite le nombre de classes distinctement identifiables. Nous avons évalué l'efficacité de quatre classificateurs d'apprentissage automatique (SVM, Bayes naïf, KNN et décision arbre de configuration : plus de cinq configurations de capteurs distinctes pour collecter des données détaillées sur les gestes de la main des avant-bras de huit participants. Un ensemble complet de 22 fonctionnalités, englobant les domaines temporel et fréquentiel par capteur ont été extraits des signaux EMG. Avant des études suggèrent qu'un ensemble de fonctionnalités à domaines mixtes peut renforcer les performances du classificateur [29].

L'application de la PCA pour sélectionner des fonctionnalités clés a considérablement réduit la dimensionnalité de l'espace des fonctionnalités, améliorant ainsi la précision de la classification. Les résultats, tels que résumés dans les tableaux 1 à 5, ont identifié la configuration de capteur C5 (deux capteurs placés à proximité du poignet le long du fléchisseur superficiel des doigts et du long fléchisseur du pouce) comme la plus efficace, donnant la plus grande précision avec SVM, Bayes naïf et arbre de décision. classificateurs. Le classificateur SVM, optimisé par MKL, a systématiquement atteint une précision de plus de 90 % dans la classification des gestes via C5, comme le montrent le tableau 2 et la figure 5. La configuration à double capteur de C5, qui capture les données des muscles contrôlant les mouvements du pouce et de l'index, suggère une stratégie avantageuse pour le placement des capteurs dans la conception de prothèses de main.

Le geste de pincement fin, représenté sur la figure 1b, s'est révélé particulièrement difficile en raison de ses niveaux de contraction musculaire plus faibles et du recours à seulement deux doigts, ce qui le rend plus sujet au bruit. Le positionnement précis du capteur C5 sur les muscles actifs lors de ce geste a permis la capture de données plus nuancées. Une comparaison de C1 et C5 (Figure 2) illustre l'influence significative d'un capteur EMG supplémentaire sur la précision de la classification, confirmant la supériorité de l'EMG double canal sur l'EMG monocanal en matière de reconnaissance gestuelle.

Lorsque les contraintes d'espace limitent le placement du capteur près du poignet sur un bras prothétique, C4 (deux capteurs placés à proximité du poignet le long du fléchisseur superficiel des orteils et le long du fléchisseur superficiel des orteils et du fléchisseur profond des orteils) apparaît comme une alternative réalisable , atteignant une précision de 87,2 % avec SVM (Tableau 5). Ces résultats mettent en évidence l'importance à la fois de l'emplacement du capteur sur le muscle et de la proximité du ventre musculaire pour une collecte de données optimale.

L'amélioration des performances SVM par MKL est évidente dans la configuration du capteur C3 (un capteur placé le long du fléchisseur superficiel des doigts et du fléchisseur profond des doigts), où un seul capteur rivalise étroitement avec le double capteur C4 en termes de précision (précision des tests SVM, tableau 5). La sélection adaptative des fonctions du noyau par MKL pour les données de C3 constitue une avancée significative par rapport aux méthodes à noyau unique. L'hyperparamètre C du SVM joue un rôle central dans l'équilibre entre la complexité du modèle et le surajustement, avec une plage de valeurs allant de 0,01 à 10, évaluées via une validation croisée de recherche de grille 10 fois (Tableau 1). Ce réglage fin était crucial pour développer un modèle SVM optimisé avec de fortes capacités de généralisation pour une classification précise des gestes. Ainsi, pour les bras prothétiques ne pouvant intégrer qu'un seul capteur EMG, C3 est la configuration recommandée. L'utilisation de MKL avec SVM améliore considérablement les performances par rapport aux SVM de base, en particulier pour la classification de plusieurs gestes et nous a permis d'obtenir des coefficients MKL non nuls pour classer les cinq gestes, comme détaillé dans le tableau 1. Alors que les SVM de base excellent dans la classification binaire, MKL gère le meilleure complexité des signaux EMG en combinant plusieurs noyaux. Il en résulte une classification robuste de cinq gestes distincts, justifiant la complexité supplémentaire de MKL.

Les matrices de confusion démontrent les hautes performances de SVM sur différents gestes , autour de 80 % à 100 % avec la configuration C5 . De plus, SVM a excellé en tant que classificateur binaire pour les gestes de pincement fin et de saisie puissante (Figure 1), atteignant une précision de 100 %. Une tendance notée est la réduction de la précision de la classification avec un nombre croissant de gestes , un phénomène qui fait écho aux découvertes précédentes [26]. Ce déclin est attribué à la dispersion plus large des signaux EMG dans l'avant-bras et au chevauchement des signaux résultant de contractions musculaires simultanées lors de l'exécution de gestes complexes.

Une limite potentielle de la présente étude, lorsqu'elle est étendue à la classification EMG en ligne en temps réel, est la légère variation dans la longueur des segments EMG dans le domaine temporel utilisé pour l'extraction de caractéristiques dans différentes configurations gestuelles. Cette limitation peut être facilement résolue en sélectionnant des longueurs de segment précisément égales lors de l'application en ligne.

Une autre limite du travail actuel est la taille de l'échantillon de huit participants. Cette taille d'échantillon est similaire aux autres tailles d'échantillon dans la littérature [25,27]. Cependant, la taille limitée de l'échantillon peut affecter la robustesse des précisions obtenues jusqu'à un certain niveau. L'étude actuelle sert à valider la méthodologie proposée à petite échelle, et il est prévu de mener des expériences futures avec un échantillon plus grand et d'inclure des individus présentant des différences de membres pour évaluer directement l'application de nos résultats dans le contrôle de la main prothétique. Cela améliorera la généralisabilité et la pertinence de la recherche actuelle.

Des recherches futures appliqueront les résultats de cet article pour reproduire en temps réel les gestes de la main dans une prothèse de main myoélectrique. Il explorera également différentes configurations de capteurs et techniques d'apprentissage automatique pour améliorer encore les performances des prothèses myoélectriques. De plus, le développement de méthodes plus sophistiquées d'extraction de caractéristiques et d'optimisation des classificateurs pourrait être bénéfique pour gérer l'analyse des signaux EMG à multiples facettes. Cette étude sert de tremplin vers la réalisation de prothèses myoélectriques plus efficientes et plus efficaces, et on espère que les connaissances acquises inspireront une exploration plus approfondie dans ce domaine prometteur.

5. Conclusions

Cette étude a fait des progrès significatifs dans le domaine de la technologie des prothèses myoélectriques, avec des résultats clés qui pourraient potentiellement façonner les recherches et les applications futures. Les performances supérieures de la configuration à double canal index-anneau et index-pouce (C4 et C5) soulignent l'importance d'un placement optimal du capteur pour améliorer la fonctionnalité des prothèses myoélectriques. L'application de MKL dans le classificateur SVM, notamment en plaçant un capteur en anneau (configuration C3), a démontré son efficacité pour atteindre une précision de classification élevée, même avec un seul capteur. La diminution observée de la précision avec l'augmentation du nombre de gestes met en évidence la nécessité d' une extraction complète des caractéristiques et d'une optimisation du classificateur pour l'analyse complexe des signaux EMG. Ces découvertes ouvrent collectivement la voie aux progrès des prothèses myoélectriques, soulignant la nécessité de configurations de capteurs raffinées et de méthodologies d'apprentissage automatique pour améliorer la précision de la reconnaissance des gestes.

Contributions des auteurs : Conceptualisation, LC, SH, KBM, JA et HR ; Méthodologie, LC, SH, KBM, JA et HR ; Logiciels, LC et SH ; Validation, LC, SH, KBM et RH ; Analyse formelle , LC, SH, JA et HR ; Enquête, LC, SH, KBM, JA et HR ; Ressources, KBM, JA et RH ; Curation des données, LC, SH, KBM et HR ; Rédaction – ébauche originale, LC, SH, KBM, JA et HR ; Rédaction – révision et édition, LC, SH, KBM, JA et HR Tous les auteurs ont lu et accepté la version publiée du manuscrit.

Financement : Cette recherche a été financée par la Glenrose Hospital Foundation et Mitacs, numéro de subvention : IT36690.

Déclaration du comité d'examen institutionnel : l'étude a été menée conformément à la Déclaration d'Helsinki et approuvée par le comité d'éthique institutionnel de l'Université de l'Alberta (AB T6G 2N2, approuvé le 25 avril 2022).

Déclaration de consentement éclairé : tous les participants ont signé un formulaire de consentement, la procédure a été approuvée par le numéro de demande du comité d'éthique de la recherche de l'Université de l'Alberta : Pro00117863.

Déclaration de disponibilité des données : Données disponibles sur demande.

Conflits d'intérêts : Les auteurs ne déclarent aucun conflit d'intérêts.

Les références

1. McDonald, CL ; Westcott-McCoy, S. ; Tisserand, M. ; Haagsma, J. ; Kartin, D. Prévalence mondiale des membres traumatiques non mortels amputation. Prothèse. Orthot. Int. 2021, 45, 105-114. [Référence croisée] [Pub Med]
2. Efanov, J. ; Tchiloemba, B. ; Izadpanah, A. ; Harris, P. ; Danino, M. Un examen des utilités et des coûts du traitement des amputations des membres supérieurs par allotransplantation composite vascularisée par rapport aux prothèses myoélectriques au Canada. JPRAS Ouvert 2022, 32, 150-160. [Référence croisée] [Pub Med]
3. Guide de l'utilisateur de Hero Arm : ouvrez Bionics. Disponible en ligne : https://openbionics.com/hero-arm-user-guide/ (consulté le 9 décembre 2023).
4. bebionic|Ottobock États-Unis. Disponible en ligne : https://www.ottobockus.com/prosthetics/upper-limb-prosthetics/solution-overview/main-bebionic/ (consulté le 9 décembre 2023).
5. Peerdeman, B. ; Boère, D. ; Witteveen, H. ; Hermens, H. ; Stramigioli, S. ; Rietman, H. ; Veltink, P. ; Misra, S. Prothèses myoélectriques de l'avant-bras : état de l'art d'un point de vue centré sur l'utilisateur. J. Réadaptation. Rés. Dév. 2011, 48, 719-738. [Référence croisée] [Pub Med]
6. Cordella, F. ; Ciancio, AL; Sacchetti, R. ; Davalli, A. ; Cutti, AG; Guglielmelli, E. ; Zollo, L. Revue de la littérature sur les besoins des utilisateurs de prothèses de membre. Devant. Neurosci. 2016, 10, 209. [Réf. croisée] [Pub Med]
7. Geethanjali, P. Contrôle myoélectrique des mains prothétiques : état de l'art. Méd. Appareils 2016, 9, 247-255. [Référence croisée] [Pub Med]

8. Tavakoli, M. ; Benussi, C. ; Lopes, Pennsylvanie ; Osorio, LB; de Almeida, AT Reconnaissance robuste des gestes de la main avec un brassard portable EMG de surface à double canal et un classificateur SVM. Bioméde. Processus de signal. Contrôle. 2018, 46, 121-130. [Référence croisée]
9. Palkowski, A. ; Redlarski, G. Classification de base des gestes de la main basée sur l'électromyographie de surface. Calculer. Mathématiques. Méthodes Med. 2016, 2016, 6481282. [Réf. croisée] [Pub Med]
10. Lee, KH ; Min, JY ; Byun, S. Classification basée sur l'électromyogramme des gestes de la main et des doigts à l'aide de réseaux de neurones artificiels. Capteurs 2021, 22, 225. [CrossRef] [Pub Med]
11. MAhsan, R. ; Ibrahimy, Michigan ; Khalifa, OO Reconnaissance des gestes de la main basée sur le signal d'électromygraphie (EMG) à l'aide d' un réseau neuronal artificiel (ANN). Dans Actes de la 4e Conférence internationale 2011 sur la mécatronique : ingénierie intégrée pour le développement industriel et sociétal, ICOM'11 — Actes de la conférence, Kuala Lumpur, Malaisie, 17-19 mai 2011. [CrossRef]
12. Kisa, DH ; Özdemir, MA ; Guren, O. ; Akan, A. Classification des gestes de la main basée sur EMG à l'aide de séries chronologiques de décomposition en mode empirique et d'apprentissage profond. Dans Actes du Congrès TIPTEKNO 2020—Tip Teknolojileri Kongresi—2020 sur les technologies médicales , TIPTEKNO 2020, Antalya, Turquie, 19-20 novembre 2020. [CrossRef]
13. Capteur musculaire MYOWARE® 2.0. Disponible en ligne : https://myoware.com/products/muscle-sensor/ (consulté le 9 décembre 2023). 14. del Toro, SF ; Wei, Y. ; Olmeda, E. ; Ren, L. ; Guowu, W. ; Díaz, V. Validation d'un système d'électromyographie (EMG) à faible coût via un Appareil EMG commercial et précis : étude pilote. Capteurs 2019, 19, 5214. [CrossRef] [Pub Med]
15. Heywood, S. ; Pua, YH; McClelland, J. ; Geigle, P. ; Rahmann, A. ; Bower, K. ; Clark, R. Électromyographie à faible coût – Validation par rapport à un système commercial utilisant des seuils de synchronisation d'activation manuels et automatisés. J. Electromyogr. Kinésiol. 2018, 42, 74-80. [Référence croisée] [Pub Med]
16. Kurniawan, SR; Pamungkas, D. Capteurs de brassard MYO et algorithme de réseau neuronal pour contrôler le robot manuel. Dans Actes de la Conférence internationale 2018 sur l'ingénierie appliquée, ICAE 2018, Batam, Indonésie, 3 et 4 octobre 2018. [Référence croisée]
17. Mao, Z.-H. ; Lee, H.-N.; Sclabassi, RJ; Sun, M. Capacité informationnelle du pouce et de l'index en communication. IEEETrans . Bioméde. Ing. 2009, 56, 1535-1545. [Référence croisée] [Pub Med]
18. Hioki, M. ; Kawasaki, H. Estimation des angles d'articulation des doigts à partir de sEMG à l'aide d'un réseau neuronal incluant le facteur de retard et Structure récurrente. ISRN Réhabilitation. 2012, 2012, 604314. [Réf. croisée]
19. Expérience : Classification des signaux. Disponible en ligne : https://backyardbrains.com/experiments/RobotHand (consulté le 9 décembre 2023).
20. Ghalyan, FIJ ; Abouélénine, ZM; Annamalai, G. ; Kapila, V. Filtre de lissage gaussien pour une modélisation améliorée du signal EMG. Dans Traitement du signal en médecine et en biologie : tendances émergentes en matière de recherche et d'applications ; Springer : Cham, Suisse, 2020 ; pp. 161-204. [Référence croisée]
21. Jaramillo-Yánez, A. ; Benalcazar, MOI ; Mena-Maldonado, E. Reconnaissance des gestes de la main en temps réel à l'aide de l'électromyographie de surface et de l'apprentissage automatique : une revue systématique de la littérature. Capteurs 2020, 20, 2467. [CrossRef] [Pub Med]
22. Tkach, D. ; Huang, H. ; Kuiken, TA Étude de la stabilité des caractéristiques du domaine temporel pour la reconnaissance de formes électromyographiques. J. Neuroeng. Rééducation. 2010, 7, 21. [Réf. croisée] [Pub Med]
23. Khairuddin, IM ; Sidek, SN ; Majeed, APPA; Razman, MAM ; Puzi, AA; Yusof, HM La classification de l'intention de mouvement grâce à des modèles d'apprentissage automatique : l'identification de caractéristiques EMG significatives dans le domaine temporel. Calcul PeerJ. Sci. 2021, 7, e379. [Référence croisée] [Pub Med]
24. Tomaszewski, J. ; Amaral, T. ; Dias, O. ; Wołczowski, A. ; Kurzy´nski, M. Classification du signal EMG utilisant un réseau neuronal avec AR coefficients du modèle. Procédure IFAC. Vol. 2009, 42, 318-325. [Référence croisée]
25. Elle, Q. ; Luo, Z. ; Meng, M. ; Xu, P. Classification des modèles EMG basée sur l'apprentissage à plusieurs noyaux pour le contrôle des membres inférieurs. Dans Actes de la 11e Conférence internationale sur le contrôle, l'automatisation, la robotique et la vision, ICARCV 2010, Singapour, 7-10 décembre 2010 ; pp. 2109-2113. [Référence croisée]
26. Odeyemi, J. ; Ogbeyemi, A. ; Wong, K. ; Zhang, W. Sur la saisie automatisée d'objets pour des mains prothétiques intelligentes utilisant l'apprentissage automatique. Bio-ingénierie 2024, 11, 108. [CrossRef] [Pub Med]
27. Avilés, M. ; Sánchez-Reyes, L.-M.; Fuentes-Aguilar, RQ; Toledo-Pérez, DC ; Rodríguez-Resendiz, J. Une nouvelle méthodologie pour classifier les mouvements EMG basée sur le SVM et les algorithmes génétiques. Micromachines 2022, 13, 2108. [CrossRef] [Pub Med]
28. Fajardo, JM; Gomez, O. ; Prieto, F. Classification des gestes de la main EMG utilisant des fonctionnalités artisanales et profondes. Bioméde. Processus de signal. Contrôle. 2021, 63, 102210. [Réf. croisée]
29. Abbaspour, S. ; Lindén, M. ; Gholamhosseini, H. ; Naber, A. ; Ortiz-Catalan, M. Évaluation des algorithmes de reconnaissance basés sur l'EMG de surface pour décoder les mouvements de la main. Méd. Biol. Ing. Calculer. 2020, 58, 83-100. [Référence croisée] [Pub Med]

électronique

MDPI

Article

Amélioration de l'utilité des données dans les données de localisation préservant la confidentialité Collecte via le partitionnement adaptatif de la grille

Jongwook Kim

Département d'informatique, Université Sangmyung, Séoul 03016, République de Corée ; jkim@smu.ac.kr

Résumé : La disponibilité généralisée des appareils compatibles GPS et les progrès des technologies de positionnement ont considérablement facilité la collecte de données de localisation des utilisateurs, ce qui en fait un atout inestimable dans diverses industries. En conséquence, il existe une demande croissante pour la collecte et le partage de ces données. Compte tenu de la nature sensible des informations de localisation des utilisateurs, des efforts considérables ont été déployés pour garantir la confidentialité, les systèmes basés sur la confidentialité différentielle (DP) apparaissant comme l' approche la plus privilégiée. Cependant, ces méthodes représentent généralement les emplacements des utilisateurs sur des grilles uniformément divisées, qui ne reflètent souvent pas avec précision la véritable répartition des utilisateurs dans un espace. Par conséquent, dans cet article, nous introduisons une nouvelle méthode qui ajuste la grille de manière adaptative en temps réel pendant la collecte de données, représentant ainsi les utilisateurs sur ces grilles partitionnées dynamiquement pour améliorer l'utilité des données collectées. Plus précisément, notre méthode capture directement la répartition des utilisateurs pendant le processus de collecte de données, éliminant ainsi le besoin de s'appuyer sur des données de répartition des utilisateurs préexistantes. Les résultats expérimentaux avec des ensembles de données réels montrent que le schéma proposé améliore considérablement l'utilité des données de localisation collectées par rapport à la méthode existante.

Mots-clés : confidentialité de la localisation ; répartition de la densité ; confidentialité différentielle; géo-indiscernabilité

check for updates

Citation : Kim, J. Améliorer les données Utilitaire de collecte de données de localisation préservant la confidentialité via le partitionnement adaptatif de la grille. Électronique 2024, 13, 3073. https://doi.org/10.3390/électronique13153073

Rédacteurs académiques : Swapnoneel Roy et Guosheng Xu

Reçu : 27 juin 2024
Révisé : 27 juillet 2024
Accepté : 29 juillet 2024
Publié : 3 août 2024

1. Introduction

La prolifération des appareils compatibles GPS et les progrès récents des technologies de positionnement ont facilité la collecte de données de localisation des utilisateurs, ce qui en fait un atout précieux pour divers secteurs. Ces données jouent un rôle important dans des domaines tels que le marketing personnalisé, l'analyse du trafic en temps réel, les recommandations, etc. Par exemple, l'analyse du trafic en temps réel utilise les données de localisation pour optimiser la circulation, réduire les embouteillages et améliorer les systèmes de navigation pour des déplacements plus efficaces. [1,2]. De plus, les recommandations basées sur la localisation pour les services, les restaurants et les événements fournissent aux utilisateurs des suggestions pertinentes et opportunes, améliorant ainsi leur expérience globale [3,4]. En conséquence, la demande de collecte et de partage de données de localisation des utilisateurs continue d'augmenter.

Les données de localisation des utilisateurs sont sensibles car elles contiennent des informations personnelles, telles que les adresses du domicile ou de l'entreprise, les dossiers de visites à l'hôpital et même les affiliations politiques [5-7]. Par exemple, en collectant et en analysant les informations de positionnement des visiteurs dans un grand centre commercial couvert, il est possible de déduire des détails sensibles, tels que leurs habitudes d'achat. De plus, les données de localisation peuvent être croisées avec d'autres ensembles de données pour tirer des conclusions encore plus précises sur le mode de vie et les choix d'un individu [8]. Par exemple, des visites fréquentes dans certains types d'entreprises ou de lieux peuvent indiquer des problèmes de santé, des passe-temps ou même des pratiques religieuses spécifiques. En conséquence, la collecte aveugle de données de localisation soulève d'importantes préoccupations en matière de confidentialité. Par conséquent, des efforts considérables ont été déployés pour protéger la confidentialité des données de localisation des utilisateurs lors du traitement de ces données.

Alors que la confidentialité différentielle (DP) [9,10] est devenue la norme de facto pour le traitement des données personnelles sensibles, des efforts importants ont été déployés pour l'appliquer aux données de localisation. En conséquence, de nombreuses méthodes basées sur DP ont été proposées pour collecter, traiter et analyser les données de localisation tout en préservant la confidentialité. Beaucoup de ces approches représentent les données de localisation des utilisateurs à l'aide de grilles, où l'ensemble du domaine est uniformément divisé en éléments disjoints.

grilles, et l'emplacement d'un utilisateur est représenté par la grille dans laquelle se trouve sa position réelle [11-14]. Bien que représenter l'emplacement des utilisateurs à l'aide de grilles uniformément partitionnées soit simple, cela ne tient pas compte de la répartition réelle des utilisateurs dans l'espace.
Cette approche se traduit souvent par une moindre utilité des données de localisation collectées, car elle suppose que les utilisateurs sont répartis de manière égale et uniforme dans toute la zone. Cependant, cela n'est pas vrai dans la plupart des scénarios réels, où certaines zones sont plus denses que d'autres. Par exemple, en milieu urbain, le centre-ville peut avoir une forte concentration d'usagers, tandis que les banlieues ont une densité plus faible. Cet écart peut avoir un impact négatif sur l'exactitude et l'efficacité des analyses ultérieures. Par conséquent, des représentations de grille plus sophistiquées , alignées sur les distributions réelles des utilisateurs, sont nécessaires pour améliorer l'utilisation des données et améliorer les performances de ces applications.

Les solutions existantes supposent l'existence d'informations préalables sur la répartition des utilisateurs, généralement obtenues à partir de données historiques. Cependant, ces données historiques ne sont pas toujours disponibles pour de nombreuses applications. Plus important encore, les informations antérieures sur la répartition des utilisateurs dérivées de données historiques peuvent ne pas correspondre à la répartition actuelle, car elles peuvent changer au fil du temps ou en réponse à des événements sociaux particuliers. Par exemple, des événements majeurs tels que des festivals peuvent modifier radicalement les schémas de déplacement et les densités des utilisateurs, rendant les données historiques obsolètes ou trompeuses [15]. Par conséquent, il est préférable d' extraire instantanément des informations sur la répartition des utilisateurs lors de la collecte de données et d'ajuster la grille de manière adaptative en conséquence.

Dans cet article, nous proposons une nouvelle méthode qui extrait simultanément la répartition des utilisateurs et ajuste la grille de manière adaptative en temps réel lors de la collecte de données de localisation. Les apports de ce travail peuvent être résumés comme suit :

- Tout d'abord, nous introduisons une méthode pour calculer efficacement la répartition des utilisateurs lors de la collecte de données de localisation basée sur DP. Cette approche est capable de capturer efficacement la répartition des utilisateurs en temps réel et de s'adapter aux changements dynamiques du comportement des utilisateurs. • Ensuite, nous proposons une méthode pour ajuster la grille de manière adaptative afin de maximiser l'utilité des données de localisation collectées sous DP. Cet ajustement adaptatif de la grille est conçu pour améliorer la granularité et la pertinence des données, en garantissant que les zones les plus importantes et les plus densément peuplées soient prioritaires, améliorant ainsi la qualité globale et l'applicabilité des données. • Nous avons évalué les performances des algorithmes proposés à l'aide d'ensembles de données du monde réel. Les résultats de l'évaluation ont démontré que le système proposé améliore considérablement l'utilité des données de localisation collectées par rapport aux méthodes existantes. Le reste de cet

article est organisé comme suit : La section 2 passe en revue les travaux connexes. Dans la section 3, nous fournissons des informations générales. Dans la section 4, nous introduisons une nouvelle méthode qui extrait simultanément la distribution des utilisateurs et ajuste la grille de manière adaptative en temps réel lors de la collecte de données de localisation. Dans la section 5, nous évaluons expérimentalement l' approche proposée avec des ensembles de données réels. Enfin, la section 6 présente nos conclusions

2. Travaux connexes

De nombreuses méthodes basées sur DP ont été développées pour collecter, traiter et analyser les données de localisation tout en préservant la confidentialité. Dans cette section, nous fournissons un bref aperçu de ces méthodes.

La confidentialité différentielle locale (LDP) est une variante de DP dans laquelle chaque utilisateur perturbe individuellement ses propres données sensibles avant de les signaler au serveur. Kim et Jang [12] proposent une approche d'agrégation de données basée sur LDP conçue pour la collecte de données de positionnement intérieur en fonction de la charge de travail, tout en garantissant la confidentialité des utilisateurs. Leur méthode identifie une stratégie optimale de codage des données et de perturbation dans le cadre LDP afin de minimiser l' erreur d'estimation globale pour la charge de travail donnée. LDPTrace [14] est conçu pour générer synthétiquement des données de trajectoire localement différentiellement privées. Dans cette méthode, les informations de localisation des utilisateurs sont collectées à l'aide de LDP pour garantir la confidentialité, et ces données perturbées sont ensuite utilisées pour générer des trajectoires synthétiques. Kim et coll. [3] présentent une méthode pour recommander le prochain point d' intérêt, en utilisant les données de localisation collectées dans le cadre du LDP.

La confidentialité différentielle métrique (MDP) étend le cadre de confidentialité différentielle standard pour gérer les données avec des mesures métriques ou de distance inhérentes [16]. Cette extension est particulièrement utile pour les données basées sur la localisation. La géo-indiscernabilité (Geo-Ind) est une application spécifique du MDP conçue pour les services basés sur la localisation [11,17,18]. Les cadres Mobile Crowdsens-ing (MCS) utilisent souvent Geo-Ind pour collecter des informations de localisation auprès des travailleurs et attribuer des tâches de manière à préserver la confidentialité. Wang et coll. [19] a été le premier à utiliser Geo-Ind pour protéger la confidentialité de la localisation des travailleurs dans le processus MCS. Le cadre proposé comprend trois étapes : Premièrement, le serveur MCS génère une fonction qui satisfait Geo-Ind. Ensuite, chaque travailleur télécharge cette fonction, masque son véritable emplacement et télécharge l'emplacement obscurci sur le serveur. Enfin, le serveur MCS attribue des tâches aux travailleurs en fonction des informations de localisation obscurcies. Dans [20], la protection de la confidentialité de la localisation dans les MCS basés sur les véhicules est étudiée, où la feuille de route est modélisée comme un graphe orienté pondéré avec les emplacements des tâches et des travailleurs comme points sur le graphe. Les auteurs proposent un schéma d'obscurcissement basé sur un mécanisme d'optimisation qui permet d'obfusquer la localisation grâce à une distribution probabiliste sur le graphique qui satisfait Geo-Ind. Jin et coll. [21] propose un cadre d'échange de confidentialité de localisation centré sur l'utilisateur pour MCS. Suivant la notion de Geo-Ind, ils conçoivent un mécanisme d'obscurcissement de la localisation qui permet à chaque travailleur de masquer de manière probabiliste sa véritable localisation en utilisant son propre budget de confidentialité. Zhang et coll. [13] introduit une méthode d'obscurcissement qui satisfait Geo-Ind pour collecter des informations de localisation auprès des travailleurs dans MCS. Huang et coll. [22] proposent un schéma respectueux de la confidentialité pour la surveillance du bruit basée sur MCS, dans lequel le serveur publie les tâches et les travailleurs signalent les emplacements perturbés et les niveaux de bruit sous DP. Chaque travailleur collabore avec un maître, soigneusement sélectionné parmi les travailleurs du même groupe, pour réaliser une Geo-Ind au niveau du groupe. Zhao et coll. [23] ont exploré la protection de la vie privée des emplacements des individus dans le contexte de l'analyse de la répartition géographique directionnelle de la communauté. Ils ont défini les informations sur la communauté à l'aide d'une matrice de covariance et les ont intégrées dans l'indiscernabilité de la géo-ellipse proposée basée sur Geo-Ind. Cette indiscernabilité géo-ellipsique offre des garanties quantifiables de confidentialité pour les emplacements dans l'espace de Mahalanobis. Yu et coll. [24] ont souligné les faiblesses des mécanismes actuels d'obfuscation de localisation basés sur Geo-Ind, en particulier lorsque les utilisateurs partagent systématiquement leurs emplacements avec plusieurs fournisseurs LBS sur une longue période de temps. Pour résoudre ce problème, ils ont introduit PrivLocAd, un système qui utilise le profilage de localisation pour générer des emplacements obscurcis, protégeant ainsi la vie privée des utilisateurs contre les attaques contradictoires multiplateformes . Zhao et coll. [25] ont introduit un nouveau concept de confidentialité appelé indiscernabilité vectorielle, qui s'appuie sur Geo-Ind pour fournir une garantie de confidentialité pour les relations dépendantes de l'emplacement.
Ils ont développé quatre mécanismes pour obtenir l'indiscernabilité des vecteurs, en utilisant à la fois des distributions de Laplace et des distributions uniformes. Mendès et coll. [26] ont utilisé la vitesse de l'utilisateur et la fréquence des rapports pour mesurer la corrélation entre les emplacements. Ils ont étendu Geo-Ind pour améliorer la préservation de la confidentialité dans les scénarios de reporting en ligne continu. Plus précisément, ils ont introduit un Geo-Ind sensible à la vitesse qui équilibre automatiquement la confidentialité et l'utilité en fonction de la vitesse de l'utilisateur et de la fréquence des rapports de localisation.

EGeoIndis [27] est un cadre de protection de la confidentialité de la localisation des véhicules conçu pour l'estimation de la densité du trafic. Il exploite Geo-Ind pour protéger la confidentialité de l'emplacement des véhicules pendant le processus d'estimation de la densité du trafic. Dans [28], les auteurs ont proposé une méthode basée sur l'apprentissage profond pour estimer la distribution de densité à l'aide de données de localisation collectées sous Geo-Ind. Chen et coll. [29] développent une méthode pour créer une carte de vulnérabilité au COVID-19 en utilisant la répartition de la densité des participants volontaires présentant des symptômes du COVID-19. Ils exploitent Geo-Ind pour collecter la localisation des participants dans le respect de la vie privée afin de garantir la confidentialité des informations sensibles sur la santé. Fathalizadeh et al. [30] présentent un cadre pour la mise en œuvre de Geo-Ind pour les environnements intérieurs. Le cadre proposé considère deux scénarios d'application de Geo-Ind, signalant un point obscurci au fournisseur de services de localisation qui satisfait DP.

L'approche proposée dans cet article s'appuie sur Geo-Ind, qui est un modèle représentatif dans le domaine de la collecte de données de localisation préservant la confidentialité. Cependant, l' approche proposée diffère des autres méthodes basées sur Geo-Ind à plusieurs égards. Premièrement, existant

les méthodes reposent généralement sur la disponibilité de données historiques pour déduire la répartition des utilisateurs, ce qui présente des défis importants. Les données historiques peuvent ne pas toujours être accessibles ou à jour, ce qui entraîne des déductions inexactes sur la répartition des utilisateurs. Deuxièmement, la plupart des méthodes existantes utilisent des structures de grille statiques, ce qui donne lieu à une représentation fixe qui ne tient pas compte du mouvement dynamique des utilisateurs. En revanche, la méthode proposée répond à ces limitations en ajustant de manière adaptative les grilles en temps réel lors de la collecte de données. Cet ajustement dynamique capture la répartition actuelle des utilisateurs sans s'appuyer sur des données historiques, améliorant ainsi l'utilité des données collectées.

3. Contexte et énoncé du problème

Dans cette section, nous fournissons le contexte nécessaire à cet article et exposons le problème abordé dans cet article.

3.1. Contexte

Récemment, DP est devenu la norme de facto pour le traitement des données préservant la confidentialité . DP est basé sur une définition mathématique formelle qui fournit une garantie probabiliste de confidentialité contre les attaquants ayant des connaissances de base arbitraires [9]. Cela garantit qu'un attaquant ne peut pas déterminer avec un degré de confiance élevé si un individu donné est inclus dans les données diffusées. DP est formellement défini comme [9,10] :

Définition 1. (-DP) Un algorithme randomisé A satisfait -DP, si et seulement si pour (1) deux ensembles de données voisins, D1 et D2, et (2) toute sortie O de A, ce qui suit est satisfait :

$$\Pr[A(D1) = O] \leq e \times \Pr[UNE(D2) = O]. \quad (1)$$

Deux ensembles de données, D1 et D2, sont considérés comme voisins s'ils ne diffèrent que par un seul enregistrement. La définition ci-dessus indique que, pour n'importe quelle sortie de A, un adversaire possédant une quelconque connaissance de base ne peut pas déterminer de manière fiable si D1 ou D2 était l'entrée de A. Le paramètre , connu sous le nom de budget de confidentialité, régule le niveau de confidentialité : plus petit les valeurs offrent une protection de la vie privée plus forte mais ajoutent plus de bruit au résultat, tandis que des valeurs plus élevées offrent une protection de la vie privée plus faible avec moins de bruit.

Il y a eu plusieurs propositions visant à appliquer le concept de DP à la protection des données de localisation. Dans cet article, nous utilisons Geo-Ind, un concept basé sur le cadre DP bien établi et reconnu comme la définition standard de la confidentialité pour la protection des données de localisation dans les services basés sur la localisation [11,17,18]. En plus des données de localisation, Geo-Ind est également utilisé pour collecter d'autres types de données, telles que des microdonnées textuelles, d'une manière conforme à DP [31,32]. Geo-Ind est formellement défini comme suit :

Définition 2. (-Geo-Ind) Considérez X comme l'ensemble des emplacements d'utilisateurs possibles et Y comme l'ensemble des emplacements signalés, qui sont généralement supposés égaux. Soit K un mécanisme aléatoire qui génère une localisation perturbée à partir de la véritable localisation d'un utilisateur. Un mécanisme aléatoire K satisfait -Geo-Ind si et seulement si la condition suivante est vraie pour (1) tous x1, x2 X et (2) tout emplacement de sortie y Y :

$$K(x1)(y) \leq e^{d(x1, x2)} \times K(x2)(y), \text{ où } d(x1, x2) \quad (2)$$

correspond à la distance entre x1 et x2.

Il existe deux méthodes principales pour implémenter Geo-Ind : le mécanisme de Laplace et le mécanisme matriciel. Il est bien connu que le mécanisme matriciel est plus efficace que la méthode de Laplace, étant donné les informations préalables sur la répartition des utilisateurs qui peuvent être obtenues à partir des données historiques disponibles [11]. Cette efficacité accrue est due au fait que le mécanisme matriciel intègre des informations de distribution préalable lorsqu'il perturbe la localisation réelle des utilisateurs. En conséquence, la distribution des emplacements perturbés collectés à l'aide du mécanisme matriciel se rapproche davantage de la distribution réelle que la distribution collectée à l'aide du mécanisme de Laplace.

Dans le mécanisme basé sur une matrice, l'espace est d'abord divisé en un ensemble de grilles, puis le serveur de collecte de données calcule une matrice d'obscurcissement, M, sur ces grilles qui satisfont -Geo-Ind. Cette matrice est ensuite distribuée aux utilisateurs. Par la suite, les utilisateurs perturbent leurs données de localisation en fonction des probabilités intégrées dans M et signalent la localisation perturbée au serveur au lieu de leurs véritables données. Plusieurs approches pour calculer la matrice d'obscurcissement qui satisfait -Geo-Ind ont été proposées dans la littérature [11,13,14,32,33]. Nous notons cependant que la méthode proposée dans cet article est suffisamment générale pour être appliquée à tout mécanisme matriciel.

3.2. Énoncé du problème

Soit U = {u1, u2, · · · , uk} un ensemble d'utilisateurs qui acceptent de fournir leurs informations de localisation au serveur. Cependant, les utilisateurs ne font pas entièrement confiance au serveur et ainsi, au lieu de fournir de véritables informations de localisation, chaque utilisateur fournit des informations de localisation perturbées (et donc préservées de la confidentialité) qui satisfont -Geo-Ind. Supposons que l'aire entière soit divisée en grilles disjointes, et soit G l'ensemble de ces grilles. La localisation de chaque utilisateur est alors représentée par la grille en G à laquelle appartient sa véritable localisation.

Le problème abordé dans cet article est de collecter des données de localisation très utiles tout en protégeant la confidentialité de la localisation des utilisateurs avec -Geo-Ind. Les méthodes existantes utilisent soit un partitionnement en grille statique qui ne s'adapte pas aux changements en temps réel dans la répartition des utilisateurs, soit s'appuient sur des données de répartition des utilisateurs préexistantes, qui peuvent ne pas être disponibles ou précises dans des scénarios en temps réel. Afin de combler ces lacunes, nous proposons une nouvelle méthode de partitionnement adaptatif de la grille qui ajuste dynamiquement la grille pendant le processus de collecte de données de localisation. En particulier, la méthode proposée capture directement la distribution des utilisateurs lors de la collecte de données, éliminant ainsi le besoin d'informations de distribution préexistantes.

4. Méthode proposée La

figure 1 donne un aperçu du schéma de collecte de données de localisation proposé à l'aide de -Geo-Ind. • Collecte de

données de localisation perturbées auprès d'utilisateurs échantillonnés : le serveur calcule d'abord la matrice d'obscurcissement, M, sur des grilles uniformément partitionnées et la distribue aux utilisateurs échantillonnés, qui renvoient ensuite leurs emplacements perturbés au serveur.

• Estimation de la répartition des utilisateurs : Le serveur estime la répartition des utilisateurs sur la base des données de localisation perturbées collectées auprès des utilisateurs échantillonnés.

• Calcul de grilles partitionnées de manière adaptative : Le serveur utilise la répartition estimée des utilisateurs pour calculer des grilles partitionnées de manière adaptative. • Collecte de données de

localisation perturbées auprès des utilisateurs restants à l'aide de grilles partitionnées de manière adaptative : une nouvelle matrice d'obscurcissement est calculée à l'aide des grilles partitionnées de manière adaptative. Cette nouvelle matrice d'obscurcissement est ensuite utilisée pour collecter les données de localisation des utilisateurs restants.

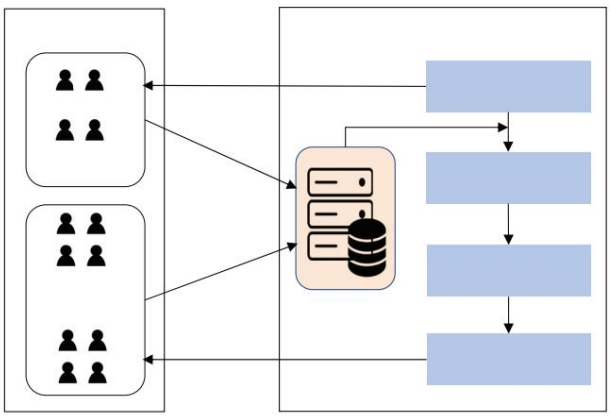


Figure 1. Un aperçu du système proposé de collecte de données de localisation préservant la confidentialité.

Dans les sous-sections suivantes, nous fournissons une explication détaillée de chaque étape.

4.1. Collecte de données de localisation perturbées auprès d'utilisateurs échantillonnés

Supposons que la zone entière soit uniformément divisée en m grilles, G = {g1, g2, · · · , gm}. Le serveur de collecte de données calcule une matrice d'obscurcissement, M, sur G. Il existe diverses approches pour calculer M qui satisfont -Geo-Ind. Dans cet article, nous utilisons la méthode proposée dans [13], où la matrice d'obscurcissement est définie comme une matrice m × m . Chaque élément M[i, j], qui représente la probabilité qu'un emplacement perturbé gj soit généré aléatoirement à partir du véritable emplacement gi , est défini comme suit :

$$M[i,j] = \frac{e^{-\frac{}{2}\cdot d(gi,gj)}}{\sum gk \quad G\, e^{-\frac{}{2}\cdot d(gi,gk)}} \tag{3}$$

Une fois la matrice d'obscurcissement M calculée, elle est distribuée aux utilisateurs échantillonnés, qui perturbent ensuite leur véritable localisation selon les probabilités codées dans M et signalent la localisation perturbée au serveur.

4.2. Estimation de la distribution de probabilité

Après avoir collecté les données de localisation perturbées auprès des utilisateurs échantillonnés, l'étape suivante consiste à estimer la répartition des utilisateurs sur la base de ces données. Pour chaque grille gi G, soit P(gi) la probabilité qu'un utilisateur se trouve en gi . Ensuite, dans cette sous-section, nous estimons P(gi) pour tout gi G à partir des données de localisation

Soit g ' perturbées échantillonnées. G soit les données perturbées que le serveur reçoit d'un utilisateur j à titre d'explication, nous utiliserons g j ' échantillonné. Pour que le désigne l'emplacement perturbé et que gj désigne l' emplacement réel. La probabilité que cet emplacement perturbé soit généré aléatoirement à partir de l' emplacement réel gi G peut être calculée comme suit :

$$P(gi|g\ j') \quad \frac{P(gi)P(g\ )=j'|gi)}{p(\ gj')} = \frac{P(gi)P(g\ \ j'|gi)}{\sum gk \quad G\ P(gk\ )P(g\ j'|gk\ )} = \frac{P(gi)M[i,j]}{\sum gk \quad G\ P(gk\ )M[k,j]} \tag{4}$$

Notez que P(g j '|gi) = M[i, j] par la définition de la matrice d'obscurcissement. Puisqu'il n'est pas possible de calculer la probabilité a priori P(gi) directement à partir de l'équation ci-dessus, nous devons l'approcher. Il existe plusieurs méthodes pour approximer la probabilité a priori, notamment l'inférence variationnelle [34], la chaîne de Markov Monte Carlo [35] et la propagation des espérances [36].

Dans cet article, nous utilisons l'algorithme d'espérance-maximisation (EM) [37] pour estimer P(gi). L'algorithme EM est particulièrement efficace lorsque la vraisemblance est bien définie, ce qui correspond dans notre cas à la matrice d'obscurcissement, M.

Supposons que DB soit un sac de données de localisation perturbées provenant d'utilisateurs échantillonnés. Le processus EM pour estimer P(gi) pour tout gi G à partir de DB est le suivant.

- Initialisation : le paramètre (c'est-à-dire la probabilité a priori) est initialisé comme suit :

$$P\ (0)\ (g1) = P\ (0)\ (g2) = \cdots = P\ (0)\ (gm) = \frac{1}{m} \tag{5}$$

- E-step : la probabilité a posteriori est calculée sur la base des paramètres actuels comme suit :

$$P(gi \mid gj' \quad \frac{P\ (t)\ (gi)M[i,j]\ ) =}{\sum gk \quad G\ P\ (t)\ (gk\ )M[k,j]} \tag{6}$$

- Étape M : le paramètre est mis à jour en utilisant les probabilités a posteriori actuelles calculées lors de l'étape E précédente :

$$P\ (t+1)\ (gi) = \frac{\sum g_{j}' \quad DB\ P(gi\ |g\ j')}{|DB|} \tag{7}$$

Ici, |DB| représente le nombre de données dans la base de données. Après avoir mis à jour les probabilités a priori, nous effectuons une étape de normalisation pour garantir que la somme de toutes les probabilités a priori est égale à 1 comme suit :

$$P^{(t+1)}(gi) = \frac{P^{(t+1)}(gi)}{\sum_{gk \in G} P^{(t+1)}(gk)} \quad (8)$$

Les étapes E et M ci-dessus sont itérées jusqu'à ce que le paramètre converge vers une valeur stable ou que le nombre d'itérations atteigne un seuil prédéfini.

4.3. Calcul de grilles partitionnées de manière adaptative

Dans cette sous-section, nous introduisons une méthode qui partitionne de manière adaptative les grilles en fonction de la distribution de probabilité (c'est-à-dire P(gi) pour tout gi G) calculée dans la phase précédente. Initialement, la méthode proposée traite toutes les grilles de G comme un seul cluster de grilles, puis partitionne itérativement ce cluster de manière descendante en utilisant un algorithme glouton.

Soit $GCv = \{C1, C2, \cdots, C_{|GCv|}\}$ représente un ensemble de clusters de grille après la v-ème partition. Supposons que pour chaque Ck GCv, grid(Ck) G désigne l'ensemble des grilles qui appartiennent au cluster Ck . Soit n le nombre total d'utilisateurs auprès desquels le serveur collecte des données de localisation. Ensuite, le nombre attendu d'utilisateurs situés dans la grille gi est calculé comme $Cnt(gi) = n \cdot P(gi)$.

De plus, soit MGCv un |GCv| × |GCv| matrice d'obscurcissement, satisfaisant -Geo-Ind, construite sur les éléments de GCv à l'aide de l'équation (3). La distance entre deux clusters nécessaires au calcul de MGCv est déterminée à l'aide des centroïdes des grilles appartenant à chaque cluster. Ensuite, en supposant que les utilisateurs perturbent leur localisation selon les probabilités codées dans MGCv , le nombre attendu de données de localisation perturbées correspondant aux grilles appartenant à CK que le serveur reçoit de n utilisateurs est calculé comme suit :

$$Cntpert(Ck) = \sum_{Cj \in GCv} \sum_{gi \in grid(Cj)} Cnt(gi) \times M[j, k] \quad (9)$$

Soit Clus() une fonction qui prend une grille en entrée et affiche le cluster auquel appartient cette grille. En supposant que les utilisateurs sont répartis uniformément sur les grilles de chaque cluster, l'erreur attendue due à Geo-Ind avec GCv est calculée comme suit :

$$ErrGCv = \sum_{gi \in G} \left| Cnt(gi) - \frac{Cntpert(Clus(gi))}{taille(Clus(gi))} \right| \quad (dix)$$

Ici, size(Ck) désigne le nombre de grilles appartenant au cluster Ck .

Supposons que dans la (v + 1)-ième partition suivante, Ch GCv soit sélectionné pour être divisé en sous-groupes. Dans cet article, nous partitionnons Ch en quatre sous-groupes de taille égale en divisant la région associée horizontalement et verticalement. Soit GCh v+1 représenter l'ensemble des clusters de grille nouvellement obtenus en subdivisant Ch . En utilisant la méthode décrite ci-dessus, nous pouvons de la même manière estimer l'erreur attendue, ErrGCh v+1 . L'ensemble des clusters de grille pour la (v + 1)-ème itération est ensuite déterminé comme suit :

$$GCv+1 = \operatorname{argmax}_{1 \le h \le |GCv|} ErrGCh_{v+1} - ErrGCv \quad (11)$$

En d'autres termes, un cluster de grille Ch qui fournit le gain maximum de réduction d'erreur est sélectionné pour le partitionnement lors de la prochaine itération.

L'algorithme 1 présente un pseudocode pour partitionner les grilles de manière adaptative en utilisant la distribution de probabilité des utilisateurs. L'algorithme prend en entrée un ensemble de grilles, G, et de distributions de probabilité, P(g1), · · · , P(gm), et génère un ensemble de clusters de grilles, GC. À la ligne 1, GCcur est initialisé pour contenir un seul cluster incluant toutes les grilles de G. Ensuite, à la ligne 3, ErrGCcur est calculé à l'aide de GCcur. Entre les lignes 4 et 12, l'algorithme identifie le cluster Ch GCcur qui produit le gain maximum de réduction d'erreur. Ce processus est répété jusqu'à ce que le gain de réduction d'erreur soit supérieur à 0. Enfin, l'algorithme renvoie GCcur.

Algorithme 1 : Pseudo-code pour l'entrée de partition de grille

adaptative : G = {g1, · · · , gm} et P(gi) pour toute sortie gi
G : un ensemble de clusters de grille GC
1 Initialisez GCcur ; 2
alors que c'est vrai
3 ErrGCcur = EstimateErr(GCcur, MGCcur) ;
4 Idx = 0, Gainmeilleur = −∞ ;
5 pour h = 1 à |GCcur| faire
6 GCh_{cur} = PartitionGrid(GCcur, h) ;
7 $ErrGCh_{cur}$ = EstimateErr(GCh cur, $MGCh_{cur}$) ;
8 si ($ErrGCh_{cur}$ − ErrGCcur) > Gainbest puis
9 Idx = h;
dix $Gainmeilleur = ErrGCh_{cur}$ − ErrGCcur ;
11 fin
12 fin
13 si Gainbest > 0 alors
14 GCcur = $GCIdx_{cur}$;
15 sinon
16 pause
17 fin
18 fin
19 retour GCcur ;

La méthode de partitionnement de grille adaptative proposée dans cette sous-section repose sur la distribution de probabilité des utilisateurs estimée à partir de données échantillonnées. Ainsi, comme pour d'autres méthodes basées sur l'échantillonnage, il est possible que les données échantillonnées soient biaisées. De tels biais peuvent conduire à l'utilisation d'une répartition des utilisateurs non représentative pour le partitionnement de la grille, ce qui peut conduire à un partitionnement inefficace. Ceci, à son tour, peut nuire à l'utilité globale des données de localisation collectées, car les grilles adaptatives peuvent ne pas refléter avec précision la véritable densité d'utilisateurs. Afin d'atténuer les biais et les inexactitudes potentiels lors de la capture de la répartition des utilisateurs en temps réel, des variations spatiales peuvent être prises en compte dans le processus d'échantillonnage. Une méthode efficace consiste à utiliser un échantillonnage stratifié [38], qui consiste à diviser la région entière en sous-régions disjointes. En garantissant que chaque sous-région est représentée proportionnellement dans l' échantillon, l'échantillonnage stratifié contribue à réduire les biais d'échantillonnage et fournit une estimation plus précise de la répartition des utilisateurs.

4.4. Collecte de données de localisation perturbées auprès des utilisateurs restants à l'aide de grilles partitionnées

de manière adaptative Une nouvelle matrice d'obscurcissement, M, est calculée à l'aide de l'ensemble de grilles partitionnées de manière adaptative, GC, calculé lors de la phase précédente. Les grilles partitionnées de manière adaptative permettent un processus d'obscurcissement plus précis et plus pertinent en capturant la nature dynamique de la répartition des utilisateurs plus efficacement que les grilles statiques. Une fois la nouvelle matrice d'obscurcissement calculée, elle est distribuée aux utilisateurs restants. Ces utilisateurs utilisent ensuite la matrice mise à jour pour perturber leurs véritables données de localisation, garantissant ainsi que leur confidentialité est préservée selon les principes de Geo-Ind. Les données de localisation obscurcies sont ensuite renvoyées au serveur, où elles sont intégrées aux données précédemment collectées auprès des utilisateurs échantillonnés.

5. Expériences

Dans cette section, nous décrivons d'abord la configuration expérimentale. Ensuite, nous discutons des résultats expérimentaux.

5.1. Configuration

expérimentale Dans cette section, nous décrivons les expériences que nous avons réalisées pour évaluer l' approche proposée. Pour nos expériences, nous avons utilisé le jeu de données des trajectoires de taxi de Porto [39], qui

se compose de trajectoires de taxi composées d'une série de coordonnées GPS enregistrées à partir de 442 taxis opérant dans la ville de Porto, au Portugal. Nous avons extrait de manière aléatoire 50 000 données de localisation de ces trajectoires, dont 10 000 ont été considérées comme des données de localisation des utilisateurs échantillonnés. Dans l'expérience, nous avons fait varier le nombre de grilles de 400 (c'est-à-dire des grilles de 20 x 20) à 10 000 (c'est-à-dire des grilles de 100 x 100). Dans les expériences, les résultats sont rapportés pour les alternatives suivantes : la méthode de grille non adaptative (NG) existante dans [13] et la méthode de grille adaptative (AG) introduite dans cet article. Nous utilisons les métriques suivantes pour l'évaluation :

- La métrique au niveau des données mesure la similarité entre l'ensemble de données de localisation réelle et l' ensemble de données de localisation perturbée collecté sous -Geo-Ind. Pour l'évaluation au niveau des données, nous utilisons à la fois l'erreur de comptage moyenne (ACE) et l'erreur de densité. L'erreur de comptage moyenne quantifie la différence entre le nombre réel d'utilisateurs, numtrue(gi), et le nombre dérivé de l'ensemble de données perturbé, numpert(gi), pour chaque grille. C'est calculé comme

$$\text{Erreur de comptage moyenne} = \sum_{1 \le i \le m} \frac{|\text{numtrue}(gi) - \text{numpert}(gi)|}{\max(\text{numvrai}(gi), 1)} \quad (12)$$

  L'erreur de densité mesure la différence entre la distribution réelle de la densité des utilisateurs et la version perturbée calculée à partir des ensembles de données collectés dans le cadre de -Geo- Ind. Cette erreur est mesurée comme

$$\text{Erreur de densité} = \text{JSD}(D(OD), D(PD)) \quad (13)$$

  Ici, JSD() représente la divergence Jenson-Shannon entre deux distributions de l'ensemble de données de localisation d'origine, D(OD) et de l'ensemble de données de localisation perturbé, D(PD).

- Les mesures au niveau de l'application évaluent l'utilité des données collectées du point de vue des applications qui les utilisent. Nous utilisons l'erreur de requête de plage pour cette métrique, une mesure largement reconnue pour évaluer l'efficacité des données de localisation [14]. Dans l'expérience, nous générons une requête de plage, QR, avec une région aléatoire R, et comparons le nombre de résultats de l'ensemble de données de localisation d'origine, QR(OD), avec ceux des ensembles de données de localisation perturbés, QR(PD). Il est calculé comme

$$\text{Erreur de requête de plage} = \frac{|QR(OD) - QR(PD)|}{\max(QR(OD), 1)} \quad (14)$$

  Dans les expériences, nous avons généré 200 requêtes de plage et signalé l' erreur moyenne des requêtes de plage.

Dans l'expérience, divers budgets de confidentialité ( ) allant de 0,2 à 2,0 ont été utilisés. Un budget de confidentialité inférieur à 2 est généralement considéré comme acceptable dans les applications pratiques [14]. Nous avons implémenté NG et AG à l'aide de Python 3.8, et toutes les expériences ont été menées dans un environnement équipé de processeurs Intel Xeon 5220R et de 64 Mo de mémoire.

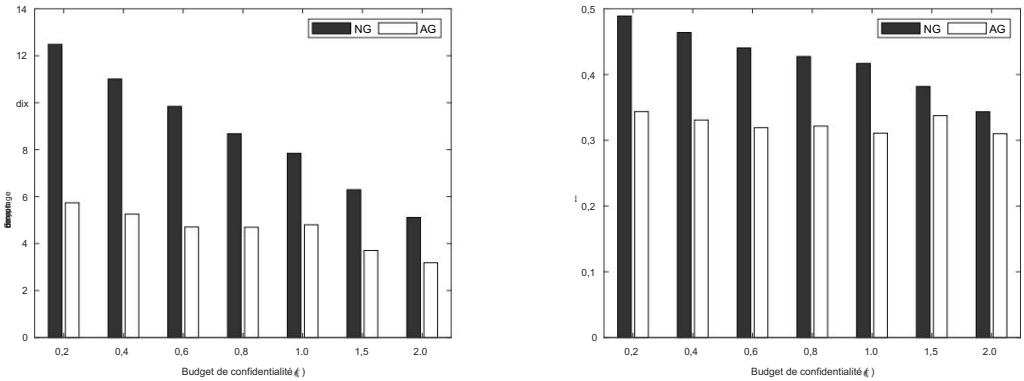
5.2. Résultats

Dans cette sous-section, nous présentons d'abord les résultats de l'évaluation du niveau des données, puis présentons les résultats de l'évaluation du niveau de l'application.

5.2.1. Évaluation au niveau des données

La figure 2 montre l'effet du budget de confidentialité sur l'erreur de comptage moyenne et l' erreur de densité. Dans cette expérience, le budget de confidentialité varie de 0,2 à 2,0, tandis que la taille de la grille est fixée à 400. À mesure que diminue, les deux erreurs augmentent. En effet, à mesure que diminue, le degré de perturbation provoqué par Geo-Ind augmente, entraînant une erreur accrue, ce qui est couramment observé avec les méthodes basées sur DP. Comme le montrent les figures, la méthode proposée (AG) surpasse systématiquement la méthode existante (NG) à tous les niveaux de budget de confidentialité. De plus, l'écart de performance entre les deux méthodes augmente à mesure que la confidentialité

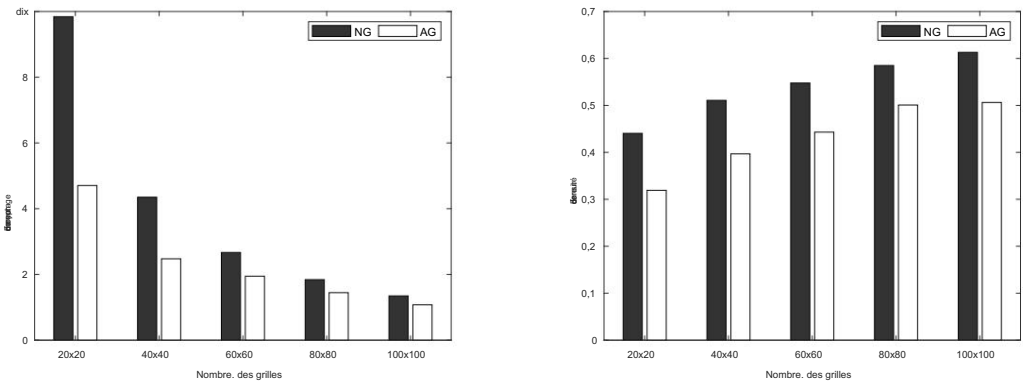
le budget diminue, et donc le niveau de confidentialité augmente. Cela montre que la proposition
Cette méthode est plus avantageuse pour les applications qui nécessitent un niveau élevé de confidentialité, ce qui est
typique de la plupart des applications qui gèrent des données de localisation.



(un) (b)

Figure 2. Effet du budget de confidentialité sur (a) l'erreur de comptage moyenne et (b) l'erreur de densité pour le la méthode de grille non adaptative (NG) existante et la méthode de grille adaptative (AG) proposée.

La figure 3 montre l'effet du nombre de grilles sur l'erreur de comptage moyenne et le
erreur de densité. Dans cette expérience, le nombre de grilles varie de 400 à 10 000, tandis que le
le budget de confidentialité est fixé à 0,6. La figure montre que la méthode proposée systématiquement
surpasse la méthode existante sur toutes les tailles de grille. Plus précisément, comme on peut le voir dans
Sur la figure, l'erreur de comptage moyenne diminue à mesure que le nombre de grilles augmente. Noter que
à mesure que le nombre de grilles augmente, le nombre d'utilisateurs par grille diminue car le total
le nombre d'utilisateurs est fixe. Ceci, à son tour, réduit l'erreur de comptage moyenne, qui est basée sur
la différence absolue entre les utilisateurs obtenus à partir des données de localisation réelles et les
données de localisation perturbées. En revanche, à mesure que le nombre de grilles augmente, la densité
erreur, qui mesure la divergence de Jenson – Shannon entre deux distributions du
l'ensemble de données de localisation d'origine et les ensembles de données de localisation perturbés augmentent. C'est parce que le
La divergence de Jenson – Shannon mesure la différence relative entre les distributions, et
par conséquent, n'est pas affecté par le nombre d'utilisateurs par grille. À mesure que la taille de la grille devient plus fine,
les perturbations dans les données ont un effet plus prononcé sur la distribution, ce qui entraîne
dans une divergence plus élevée entre les ensembles de données d'origine et perturbés.



(un) (b)

Figure 3. Effet du nombre de grilles sur (a) l'erreur de comptage moyenne et (b) l'erreur de densité pour le la méthode de grille non adaptative (NG) existante et la méthode de grille adaptative (AG) proposée.

Les résultats présentés dans les figures 2 et 3 confirment que la méthode proposée permet
collecte de données de localisation qui sont plus similaires aux données originales sous Geo-Ind que
la méthode existante. Ces résultats mettent en évidence les avantages significatifs de notre approche

dans la collecte de données de localisation préservant la confidentialité. En ajustant dynamiquement les grilles en fonction de la répartition des utilisateurs en temps réel sous Geo-Ind, nous obtenons une précision et une utilité des données plus élevées.

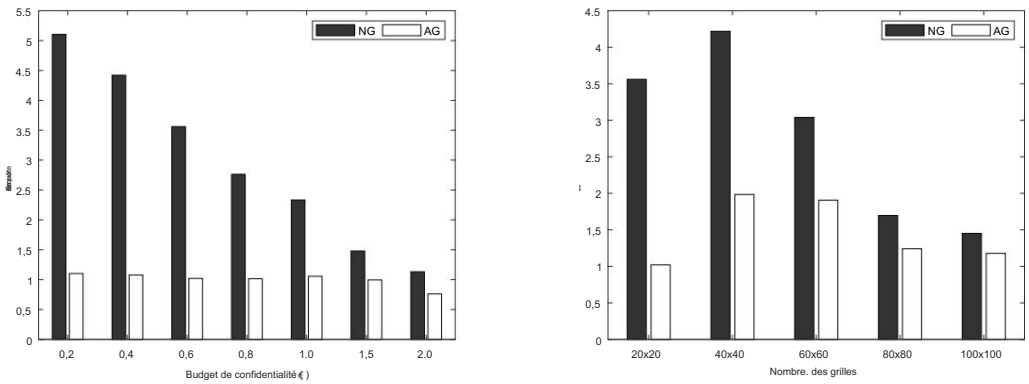
5.2.2. Évaluation au niveau de l'application La

figure 4 illustre l'effet du budget de confidentialité et du nombre de grilles sur l' erreur de requête de plage. Dans la figure 4a, la taille de la grille est fixée à 400, tandis que dans la figure 4b, le budget de confidentialité est fixé à 0,6. Comme le montre la figure, à mesure que diminue, l'erreur associée à la méthode existante augmente considérablement, tandis que l'erreur de la méthode proposée n'augmente que marginalement. Cela se produit parce que la méthode proposée est capable de collecter des ensembles de données de localisation plus proches des ensembles de données d'origine, comme l'a vérifié l'évaluation au niveau des données.

La robustesse de notre approche malgré différents budgets de confidentialité met en évidence son efficacité à équilibrer confidentialité et exactitude. À mesure que devient plus petit, indiquant des garanties de confidentialité plus fortes, la méthode proposée parvient toujours à préserver l'utilité des données, les rendant plus fiables pour les applications nécessitant des informations de localisation précises.

De plus, la méthode proposée surpasse systématiquement la méthode existante sur toutes les tailles de grille. Cela vérifie que notre méthode proposée est robuste quel que soit le nombre de grilles. La capacité à maintenir de faibles taux d'erreur de requête dans différentes configurations de grille démontre l'adaptabilité et l'efficacité de notre approche. Cette robustesse est cruciale pour les applications pratiques qui nécessitent une granularité différente dans la représentation de l'emplacement de l'utilisateur (c'est-à-dire le nombre de grilles) en fonction des exigences de l'application.

Ces résultats expérimentaux indiquent que la méthode proposée peut être utilisée pour un large éventail de services et d'applications géolocalisés nécessitant différents niveaux de confidentialité et granularité dans la représentation de l'emplacement de l'utilisateur.



(un) (b)

Figure 4. Effet de (a) le budget de confidentialité et (b) du nombre de grilles sur l'erreur de requête de plage pour la méthode de grille non adaptative (NG) existante et la méthode de grille adaptative (AG) proposée.

5.2.3. Évaluation de la variabilité du réseau et des effets d'adaptation au réseau Tous

les résultats expérimentaux des sous-sections précédentes ont été obtenus sous l'hypothèse que les conditions du réseau sont stables. Cependant, dans les scénarios réels, les conditions du réseau peuvent être très variables et imprévisibles. La méthode proposée divise la grille de manière adaptative en fonction des données de localisation échantillonnées collectées au cours du processus de collecte de données de localisation. Cependant, des conditions de réseau instables, telles qu'une latence élevée du réseau, une perte de paquets et une faible bande passante, peuvent retarder la collecte en temps opportun de ces données échantillonnées. Par conséquent, certaines données échantillonnées peuvent ne pas être disponibles pour le calcul du partitionnement de grille adaptatif, ce qui peut conduire à un partitionnement de grille moins précis.

Dans cette sous-section, pour relever les défis des réseaux instables, nous évaluons l' efficacité de la méthode proposée dans des scénarios de réseau réels. L'expérience présentée dans le tableau 1 considère un scénario dans lequel certaines des données de localisation échantillonnées nécessaires pour estimer la répartition des utilisateurs dans la section 4.2 (qui sont ensuite utilisées pour calculer de manière adaptative les partitions de grille dans la section 4.3) sont perdues ou ne sont pas reçues à temps en raison de réseau instable

conditions telles qu'une latence élevée, une perte de paquets et une faible bande passante. Dans cette expérience, le Le taux de perte des données de localisation échantillonnées varie de 1 % à 20 %, couvrant une plage allant de typique à des conditions de réseau sévères. Dans l'expérience, le nombre de grilles est fixé à 400 et le budget de confidentialité est fixé à 0,6. Comme le montre le tableau 1, même si le taux de perte des échantillons les données de localisation augmentent de 1 % à 20 %, les deux erreurs restent stables avec seulement un très petit augmenter. En particulier, même dans des conditions de réseau sévères avec un taux de perte de 20 %, le La méthode de grille adaptative (AG) proposée surpasse considérablement la grille non adaptative (NG) méthode. Ces résultats vérifient la robustesse et l'efficacité de la méthode proposée dans maintenir l'utilité des données dans des scénarios de réseau réels avec différents niveaux de perte de données dus à des conditions de réseau instables.

Tableau 1. Effet du taux de perte des données de localisation échantillonnées sur l'erreur de comptage moyenne et l'erreur de densité.

	AG				NG
Taux de perte des données de localisation échantillonnées 0 %	1%	5%	dix%	20%	
Erreur de comptage moyen	4.705	4.867	4.896	4.936	5.018	9.842
Erreur de densité	0,319	0,321	0,328	0,336	0,341	0,440

La méthode proposée ajuste la grille de manière adaptative pendant le processus de collecte de données de localisation, ce qui introduit une latence supplémentaire en raison de la surcharge de calcul de calculer la grille adaptative. Par conséquent, nous évaluons expérimentalement la latence introduite par la méthode proposée. Le tableau 2 montre les résultats de latence provoqués par la grille adaptative calcul de la méthode proposée. Dans cette expérience, le nombre de grilles varie de 400 à 10 000, alors que le budget vie privée est fixé à 0,6. Comme le montre le tableau, la latence augmente à mesure que la taille de la grille augmente. En effet, à mesure que le nombre de grilles augmente, le nombre d'itérations nécessaires pour partitionner les grilles de manière adaptative augmente également. Par conséquent, les grilles plus grandes nécessitent plus de ressources de calcul, ce qui entraîne une latence plus élevée.

Tableau 2. Latence provoquée par le calcul de la grille adaptative.

Taille de la grille	20 × 20	40 × 40	60 × 60	80 × 80	100 × 100
Temps d'exécution (s)	4.10	17h41	43.12	139,56	439.56

Nous notons que bien que le calcul de la grille adaptative introduise une latence supplémentaire comme le montre le tableau 2, il s'agit d'un processus unique dans le cadre de la collecte globale de données de localisation. procédure. Par conséquent, l'impact de cette surcharge sur le temps de traitement global du la collecte de données de localisation est limitée. De plus, la latence supplémentaire provoquée par la surcharge de calcul liée au partitionnement adaptatif de la grille peut être atténuée par diverses méthodes parallèles. techniques de traitement. En particulier, des techniques telles que les frameworks de calcul distribués [40,41] et l'accélération GPU peuvent réduire considérablement le temps de calcul, atténuant cette latence. En répartissant la charge de travail sur plusieurs processeurs, ces Les approches peuvent améliorer l'efficacité du processus de partitionnement adaptatif de la grille, garantissant analyse de données dans des applications en temps réel.

6. Conclusions et travaux futurs

Récemment, il y a eu une demande croissante pour la collecte et le partage de données de localisation. Compte tenu de la nature sensible des informations de localisation des utilisateurs, des efforts considérables ont été mis en place pour garantir la confidentialité, avec des systèmes différentiels basés sur la confidentialité émergeant comme l'approche privilégiée. Cependant, ces schémas représentent généralement les emplacements des utilisateurs sur des grilles uniformément divisées, qui souvent ne reflètent pas avec précision la véritable répartition des utilisateurs au sein d'un espace. Dans cet article, nous avons présenté une nouvelle approche qui ajuste dynamiquement la grille en temps réel pendant la collecte de données de localisation à l'aide de Geo-Ind pour améliorer l'utilité de les données collectées. La méthode proposée capture la distribution des utilisateurs directement pendant la collecte de données collecte, éliminant ainsi le recours aux informations de distribution préexistantes. Expérimental

les résultats sur des données réelles ont confirmé que le système proposé améliore considérablement l'utilité des données de localisation collectées au niveau des données et des applications. Plus précisément, les résultats ont montré que par rapport à la solution existante, la méthode proposée peut réduire le taux d'erreur jusqu'à 52 % dans les expériences au niveau des données et jusqu'à 75 % dans les expériences au niveau des applications.

Malgré les résultats prometteurs, la méthode proposée présente les limites suivantes. Étant donné que la méthode proposée ajuste la grille de manière adaptative en temps réel pendant la collecte de données, il existe une surcharge de calcul supplémentaire associée au calcul de la grille adaptative. Cette surcharge est particulièrement importante lors de l'utilisation de grilles de grande taille. Ainsi, les travaux futurs se concentreront sur l'amélioration de l'efficacité du calcul de grille adaptative, en particulier pour les grilles de grande taille avec un grand nombre d'utilisateurs. Ceci peut être réalisé en parallélisant le processus de partitionnement adaptatif de la grille afin de réduire la surcharge de calcul. Nous explorerons diverses techniques de traitement parallèle, telles que la mise en œuvre de cadres informatiques distribués tels qu'Apache Hadoop [40] ou Apache Spark [41], qui distribuent les données et les calculs sur un cluster de machines. De plus, le multithreading au sein d'une seule machine et l' utilisation de l'accélération GPU peuvent être envisagés pour augmenter l'efficacité. De plus, l'intégration des services de cloud computing améliorera l'évolutivité en fournissant une infrastructure dynamique et évolutive pour effectuer un partitionnement de grille adaptatif sur de grands ensembles de données.

Une autre direction de recherche future consiste à analyser théoriquement l'impact du partitionnement adaptatif de la grille sur l'utilité des données de localisation collectées. Cette analyse élucidera les principes sous-jacents du partitionnement adaptatif de la grille et son impact sur l'utilité des données, fournissant ainsi un support théorique plus robuste à la méthode proposée. En outre, le compromis confidentialité-utilité peut être étudié plus en détail afin d'optimiser l'équilibre entre confidentialité et utilité. Cela comprendra le développement de modèles et de mesures pour évaluer quantitativement ce compromis dans diverses conditions.

Financement : Cette recherche a été financée par une subvention de recherche 2023 de l'Université Sangmyung (2023-A000-0119).

Déclaration de disponibilité des données : les données originales présentées dans l'étude sont librement disponibles dans Kaggle à l'adresse https://www.kaggle.com/c/pkdd-15-predict-taxi-service-trajectory-i (consulté le 20 juin 2024).

Conflits d'intérêts : L'auteur ne déclare aucun conflit d'intérêts.

Abréviations

Les abréviations suivantes sont utilisées dans ce manuscrit :

DP	Confidentialité différentielle
PDL	Confidentialité différentielle locale
MDP	Confidentialité différentielle des métriques
Géo-Ind	Géo-indiscernabilité
MCS	Détection de foule mobile
EM	Attente-Maximisation
JSD	Divergence Jenson-Shannon

Les références

1. Wang, X. ; Peut.; Wang, Y. ; Jin, W. ; Wang, X. ; Tang, J. ; Jia, C. ; Yu, J. Prédiction des flux de trafic via un graphique neuronal spatio-temporel réseau. Dans Actes de la conférence Web, Taipei, Taiwan, 20-24 avril 2020 ; pages 1082 à 1092.
2. Poêle, Z. ; Liang, Y. ; Wang, W. ; Yu, Y. ; Zheng, Y. ; Zhang, J. Prédiction du trafic urbain à partir de données spatio-temporelles utilisant le méta- apprentissage profond. Dans les actes de la conférence internationale ACM SIGKDD sur la découverte des connaissances et l'exploration de données, Anchorage, AK, États-Unis, 4-8 avril 2019 ; pages 1720-1730.
3. Kim, JS ; Kim, JW ; Chung, YD Recommandation de points d'intérêt successifs avec confidentialité différentielle locale. Accès IEEE 2021, 9, 66371-66386. [Référence croisée]
4. An, J. ; Li, G. ; Jiang, W. NRDL : Apprentissage décentralisé des préférences de l'utilisateur pour la prochaine recommandation de POI préservant la confidentialité. Expert Système. Appl. 2024, 239, 122421. [Réf. croisée]
5. Primault, V. ; Boutet, A. ; Mokhtar, SB ; Brunie, L. Le long chemin vers la confidentialité de la localisation informatique : une enquête. Communique IEEE. Survivre. Tuteur. 2019, 21, 2772-2793. [Référence croisée]

6. Liu, B. ; Zhou, W. ; Zhu, T. ; Gao, L. ; Xiang, Y. Confidentialité de la localisation et ses applications : une étude systématique. Accès IEEE 2019, 6, 17606-17624. [Référence croisée]
7. Alharthi, R. ; Banihani, A. ; Alzahrani, A. ; Alshehri, A. ; Alshahrani, H. ; Fu, H. ; Liu, A. ; Zhu, Y. Défis de confidentialité de localisation dans le crowdsourcing spatial. Dans Actes de la conférence internationale de l'IEEE sur l'électro/technologie de l'information, Rochester, MI, États-Unis, 3-5 mai 2018.
8. Henriksen-Bulmer, J. ; Jeary, S. Attaques de réidentification – Une revue systématique de la littérature. Int. J. Inf. Gérer. 2016, 36, 1184-1192. [Référence croisée]
9. Dwork, C. Confidentialité différentielle. Dans Actes du Colloque international sur les automates, les langages et la programmation, Venise, Italie, 10-14 juillet 2006 ; p. 1 à 12.
10. Dwork, C. ; McSherry, F. ; Nissim, K. ; Smith, A. Calibrage du bruit en fonction de la sensibilité dans l'analyse de données privées. Dans les actes du Troisième conférence sur la théorie de la cryptographie, New York, NY, États-Unis, 4-7 mars 2006.
11. Bordenabé, NE ; Chatzikokolakis, K. ; Palamidess, C. Mécanismes géo-indiscernables optimaux pour la confidentialité de la localisation. Dans Actes de la conférence ACM SIGSAC sur la sécurité informatique et des communications, New York, NY, États-Unis, 3-7 novembre 2014 ; pp. 251-262.
12. Kim, J. ; Jang, B. Collecte de données de positionnement en intérieur tenant compte de la charge de travail via la confidentialité différentielle locale. Communique IEEE. Lett. 2019, 23, 1352-1356. [Référence croisée]
13. Zhang, P. ; Cheng, X. ; Su, S. ; Wang, N. Recrutement de travailleurs basé sur la couverture géographique sous géo-indiscernabilité. Calculer. Réseau. 2022, 217, 109340. [Réf. croisée]
14. Du, Y. ; Hu, Y. ; Zhang, Z. ; Croc, Z. ; Chen, L. ; Zheng, B. ; Gao, Y. LDPTrace : Synthèse de trajectoires localement différentiellement privées. Dans Actes du VLDB Endowment, Vancouver, BC, Canada, 28 août-1er septembre 2023 ; pages 1897-1909.
15. Ghaemi, Z. ; Farnaghi, M. Une approche de clustering basée sur une densité variée pour la détection d'événements à partir de données Twitter hétérogènes. ISPRS Int. J. Géo-Inf. 2019, 8, 82. [Réf. croisée]
16. Alvim, M. ; Chatzikokolakis, K. ; Palamidessi, C. ; Pazii, A. Confidentialité différentielle locale sur les espaces métriques : optimisation du compromis avec l'utilité. Dans les actes du symposium IEEE sur les fondations de sécurité informatique, Oxford, Royaume-Uni, 9-12 juillet 2018.
17. Andrés, MOI ; Bordenabé, NE ; Chatzikokolakis, K. ; Palamidessi, C. Géo-indiscernabilité : confidentialité différentielle pour les systèmes basés sur la localisation. Dans Actes de la conférence ACM SIGSAC sur la sécurité informatique et des communications, Berlin, Allemagne, 4-8 novembre 2013 ; pp. 901-914.
18. Chatzikokolakis, K. ; Palamidessi, C. ; Stronati, M. Géo-indiscernabilité : une approche fondée sur des principes en matière de confidentialité de localisation. Dans Actes de la Conférence internationale sur l'informatique distribuée et la technologie Internet, Bhubaneswar, Inde, 5-8 février 2015 ; pp. 49-72.
19. Wang, L. ; Yang, D. ; Han, X. ; Wang, T. ; Zhang, D. ; Ma, X. Allocation de tâches de localisation préservant la confidentialité pour la détection de foule mobile avec obscurcissement géographique différentiel. Dans Actes de la Conférence internationale sur le World Wide Web, Perth, Australie, 3-7 avril 2017 ; pp. 627-636.
20. Qiu, C. ; Squicciarini, AC Protection de la confidentialité de la localisation dans le crowdsourcing spatial basé sur les véhicules via la géo-indiscernabilité. Dans les actes de la conférence internationale de l'IEEE sur les systèmes informatiques distribués, Dallas, Texas, États-Unis, 7-10 juillet 2019 ; pages 1061 à 1071.
21. Jin, W. ; Xiao, M. ; Guo, L. ; Yang, L. ; Li, M. ULPT : Un cadre d'échange de confidentialité de localisation centré sur l'utilisateur pour la détection de foule mobile. IEEETrans. Foule. Calculer. 2022, 21, 3789-3806. [Référence croisée]
22. Huang, P. ; Zhang, X. ; Guo, L. ; Li, M. Incitation à la surveillance du bruit basée sur la détection de foule avec des emplacements différentiellement privés. IEEETrans. Foule. Calculer. 2021, 20, 519-532. [Référence croisée]
23. Zhao, Y. ; Yuan, D. ; Du, JT; Chen, J. Geo-Ellipse-Indistinguishability : Protection de la confidentialité de l'emplacement sensible à la communauté pour les directions distribution. IEEETrans. Connaître. Ingénierie des données. 2023, 35, 6957-6967. [Référence croisée]
24. Yu, L. ; Zhang, S. ; Meng, Y. ; Du, S. ; Chen, Y. ; Ren, Y. ; Zhu, H. Publicité géolocalisée préservant la confidentialité via la géo-indiscernabilité longitudinale. IEEETrans. Foule. Calculer. 2024, 23, 8256-8273. [Référence croisée]
25. Zhao, Y. ; Chen, J. Vector-indiscernability : protection de la confidentialité basée sur la dépendance à l'emplacement pour les données de localisation successives. IEEE Trans. Calculer. 2024, 73, 970-979. [Référence croisée]
26. Mendès, R. ; Cunha, M. ; Vilela, JP Géo-indiscernabilité sensible à la vitesse. Dans Actes de la conférence ACM sur les données et Sécurité des applications et confidentialité, Charlotte, Caroline du Nord, États-Unis, 24-26 avril 2023 ; pp. 141-152.
27. Ren, W. ; Tang, S. EGeoIndis : Un cadre de protection de la confidentialité de localisation efficace et efficient dans la détection de la densité du trafic. Véh. Commun. 2020, 21, 100187. [Réf. croisée]
28. Kim, JW ; Lim, B. Estimation efficace et respectueuse de la vie privée de la distribution de densité des utilisateurs LBS sous géo- indiscernabilité. Électronique 2023, 12, 917. [CrossRef]
29. Chen, R. ; P'tit.; Peut.; Gong, Y. ; Guo, Y. ; Ohtsuki, T. ; Pan, M. Construction d'une carte mobile de vulnérabilité au COVID-19 en mode participatif Avec géo-indiscernabilité. IEEE Internet Things J. 2022, 9, 17403-17416. [Référence croisée]
30. Fathalizadeh, A. ; Moghtadaiee, V. ; Alishahi, M. Géo-indiscernabilité intérieure : adopter une confidentialité différentielle pour l'intérieur Protection des données de localisation. IEEETrans. Émerger. Haut. Calculer. 2023, 12, 293-306. [Référence croisée]
31. Feyisetan, O. ; Balle, B. ; Drake, T. ; Diethe, T. Analyse textuelle préservant la confidentialité et l'utilité via des perturbations multivariées calibrées . Dans Actes de la Conférence internationale sur la recherche sur le Web et l'exploration de données, Houston, Texas, États-Unis, 3-7 février 2020 ; pp. 178-186.

32. Chanson, S. ; Kim, JW Adaptation de la géo-indiscernabilité pour la collecte de microdonnées médicales préservant la confidentialité. Electronique 2023, 12, 2793. [Réf. croisée]
33. Ahuja, R. ; Ghinita, G. ; Shahabi, C. Une technique de préservation de l'utilité et évolutive pour protéger les données de localisation avec géo- indiscernabilité. Dans Actes de la Conférence internationale sur l'extension de la technologie des bases de données, Lisbonne, Portugal, 26-29 mars 2019 ; pp. 210-231.
34. Blei, DM; Kucukelbir, A. ; McAuliffe, JD Inférence variationnelle : une revue pour les statisticiens. Confiture. Stat. Assoc. 2017, 112, 859-877. [Référence croisée]
35. Chib, S. Markov Méthodes de Monte Carlo en chaîne : calcul et inférence. Handb. Économ. 2001, 5, 3569-3649.
36. Li, Y. ; Hernández-Lobato, JM ; Turner, RE Propagation des attentes stochastiques. arXiv 2018, arXiv : 1506.04132.
37. Sammaknejad, N. ; Zhao, Y. ; Huang, B. Un examen de l'algorithme de maximisation des attentes dans l'identification des processus basés sur les données. J. Contrôle des processus 2019, 73, 123-136. [Référence croisée]
38. Howell, CR; Su, W. ; Nassel, AF; Agnès, AA ; Cherrington, AL Échantillonnage aléatoire stratifié basé sur une zone utilisant des données géospatiales technologie dans une enquête communautaire. BMC Public Health 2020, 20, 1678. [CrossRef] [Pub Med]
39. Moreira-Matias, L. ; Gama, J. ; Ferreira, M. ; Mendes-Moreira, J. ; Damas, L. Prédire la demande de taxi et de passagers à l'aide du streaming données. IEEETrans. Intell. Transp. Système. 2013, 14, 1393-1402. [Référence croisée]
40. Apache Hadoop. Disponible en ligne : https://hadoop.apache.org/ (consulté le 22 juillet 2024).
41. Apache Spark : moteur unifié pour l'analyse de données à grande échelle. Disponible en ligne : https://spark.apache.org/ (consulté sur 22 juillet 2024).