прикладные науки

MDPI

Статья

# Новая симметричная мелко-грубая нейронная сеть для 3D-человека

## Распознавание действий на основе последовательностей облаков точек

Чанг Ли [1], Цянь Хуан [1,*], Инчи Мао [1], Вэйвэнь Цянь [1] и Син Ли [2]

[1] Колледж компьютерных наук и разработки программного обеспечения, Университет Хохай, Нанкин 211100, Китай; licang@hhu.edu.cn (CL); yingchimao@hhu.edu.cn (ЮМ); qianweiwen@hhu.edu.cn (WQ)

[2] Колледж информационных наук и технологий и Колледж искусственного интеллекта Нанкинского лесного хозяйства Университет, Нанкин 210037, Китай; lixing@njfu.edu.cn

* Переписка: huangqian@hhu.edu.cn

Аннотация: Распознавание действий человека способствовало разработке устройств искусственного интеллекта, ориентированных на человеческую деятельность и услуги. Эта технология получила развитие благодаря внедрению Трехмерные облака точек, полученные с помощью камер глубины или радаров. Однако человеческое поведение сложно и задействованные облака точек обширны, неупорядочены и сложны, что создает проблемы для трехмерных действий. признание. Для решения этих проблем мы предлагаем симметричную нейронную сеть мелкого и грубого анализа (SFCNet). который одновременно анализирует внешний вид и детали действий человека. Во-первых, последовательности облаков точек преобразуются и вокселизуются в структурированные наборы трехмерных вокселей. Эти наборы затем дополняются с дескриптором интервальной частоты для создания 6D-функций, отражающих пространственно-временную динамику информация. Оценивая занятость воксельного пространства с помощью пороговой обработки, мы можем эффективно идентифицировать существенные части. После этого все воксели с признаком 6D направляются в глобальный грубый поток, в то время как вокселы в ключевых частях направляются в локальный тонкий поток. Эти два потока извлекают глобальные функции внешнего вида и критические части тела с использованием симметричного PointNet++. Впоследствии Объединение функций внимания используется для адаптивного захвата более отличительных моделей движений. Эксперименты, проведенные на общедоступных эталонных наборах данных NTU RGB+D 60 и NTU RGB+D 120, подтверждают Эффективность и превосходство SFCNet в распознавании 3D-действий.

Ключевые слова: анализ облака точек; Распознавание 3D-действий; распознавание образов; глубокое обучение



Цитирование: Ли, К.; Хуан, К.; Мао, Ю.; Цянь, В.; Ли, Х. Роман симметричный Грубая нейронная сеть для 3D Распознавание действий человека на основе Последовательности облаков точек. Прил. наук. 2024, 14, 6335. https://doi.org/10.3390/app14146335

Академический редактор: Ацуши Масе

Поступила: 11 июня 2024 г.
Пересмотрено: 8 июля 2024 г.
Принято: 18 июля 2024 г.
Опубликовано: 20 июля 2024 г.



## 1. Введение

Распознавание действий человека направлено на то, чтобы помочь компьютерам понять семантику человеческого поведения на основе различных данных, записанных устройствами сбора данных. В частности, 3D-экшен Распознавание предназначено для извлечения моделей действий из трехмерных данных, включающих движения человека . Он привлекает все большее внимание из-за его широкого применения, такого как мониторинг общественной безопасности, аттестация, военная разведка и интеллектуальная транспорт [1].

Текущие основные методы распознавания трехмерных действий можно разделить на методы, основанные на глубине (включая карты глубины и последовательности облаков точек) [2–4] и методы, основанные на скелетах [5,6] в зависимости от используемого типа данных. Ограничено точностью алгоритм оценки позы — неизбежная первоочередная задача — скелетные методы сталкиваются с проблемами вычислительного потребления и надежности. Напротив, основанный на глубине методы более независимы от задач и привлекли широкое внимание. Существующий Подходы к распознаванию трехмерных действий на основе глубины в основном делятся на две основные категории. первый — закодировать 3D-движения в одно или несколько изображений [2,3,7,8] и использовать CNN [9] для признания действий. Однако плоскость 2D-изображения не может полностью охарактеризовать 3D-изображение. динамика, поскольку человеческие действия одновременно пространственно-временны и осуществляются в 3D пространство. Другой способ — преобразовать видео глубины в последовательность облаков точек [10], которая записывает трехмерные координаты точек в пространстве в несколько моментов времени. Таким образом, по сравнению с изображениями последовательности облаков точек имеют то преимущество, что сохраняют трехмерный вид и

динамика геометрии с течением времени, что позволяет проводить расширенный анализ и понимание действий человека. Кроме того, облака точек можно получать с помощью различных устройств, таких как лазерные сканеры, радары, датчики глубины и камеры RGB+D, которые можно устанавливать на дроны, уличные фонари, транспортные средства и самолеты наблюдения, что расширяет сферу применения распознавания действий. . Однако из-за сложной структуры и огромного объема облака точек существующие методы распознавания трехмерных действий на его основе сталкиваются со следующими проблема

Во-первых, последовательности облаков точек всегда имеют массивные точки, пропорциональные размерности времени, а схема обработки данных требует много времени. Поэтому разработка эффективной и легкой модели последовательности облаков точек имеет решающее значение для распознавания трехмерных действий. Во-вторых, точки в последовательностях нерегулярны, демонстрируют неупорядоченную внутрикадровую пространственную информацию и упорядоченные межкадровые временные детали, что затрудняет анализ основных моделей движения. Однако существующие методы обработки облаков точек обычно выполняют недифференцированную понижающую дискретизацию всего облака точек, что приводит к равномерной потере важной и малозаметной информации. Кроме того, существующие схемы анализа облака точек игнорируют критические части тела, участвующие в действиях, что приводит к отсутствию нюансов извлеченных признаков действий, что в конечном итоге ограничивает производительность распознавания действий

Чтобы решить эти проблемы, мы предлагаем структуру глубокого обучения под названием Symmetric Fine-Coarse Neural Network (SFCNet), которая симметрично сочетает в себе анализ особенностей движения с локальной и глобальной точек зрения, как показано на рисунке 1. Во-первых, для экономии вычислительных затрат. , мы уменьшаем количество точек путем выборки кадров и выборки самой дальней точки. Затем выбранные облака точек преобразуются в 3D-вокселы для создания компактного представления облака точек. Затем к исходным трехмерным положениям добавляется дескриптор интервальной частоты, чтобы отобразить общую пространственную конфигурацию и облегчить идентификацию основных частей тела, что позволяет нам разделить последовательности облаков точек на локальное мелкое пространство и глобальное грубое пространство. Мы рассматриваем вокселы, участвующие в этих двух пространствах, как точки и используем PointNet++ [11] для сквозного извлечения признаков. Наконец, наш модуль объединения функций объединяет общий внешний вид и локальные детали для получения отличительных признаков для распознавания трехмерных действий. Обширные эксперименты с крупномасштабными наборами данных NTU RGB+D 60 и NTU RGB+D 120 демонстрируют эффективность и превосходство SFCNet, с помощью которого можно судить о намерениях человека и помогать им в приложениях дистанционного зондирова

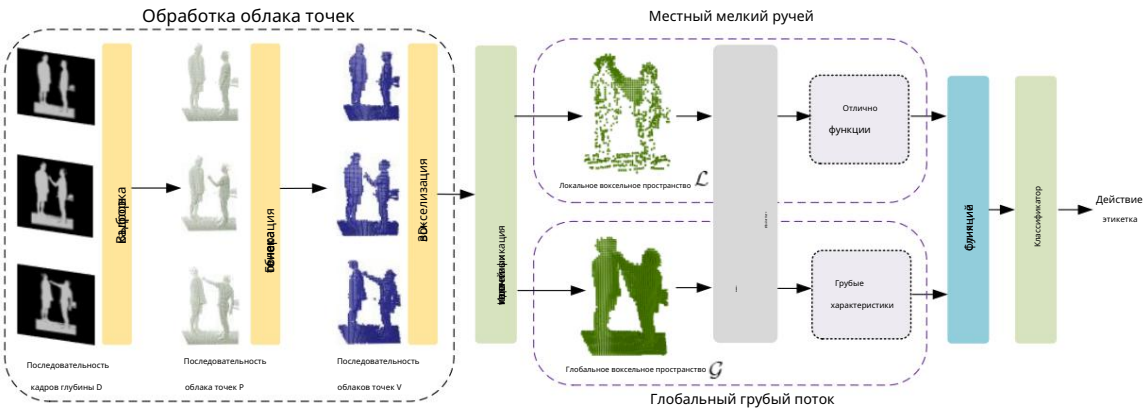



Рисунок 1. Конвейер SFCNet. Он преобразует кадры глубины в облако точек и применяет операции вокселизации. Симметричная структура кодирует 3D-вокселы, при этом важные части и глобальная динамическая информация обрабатываются отдельно. Прикрепленный дескриптор интервальной частоты первоначально характеризует информацию о движении, а затем обрабатывается PointNet++ [11] для получения более глубоких характеристик. Наконец, классификатор распознает трехмерные действия, используя агрегированный признак.

В целом, основные результаты нашей работы заключаются в следующем: •

Мы предлагаем дескриптор интервальной частоты для характеристики 3D-вокселов во время выполнения действий, который полностью сохраняет детали движения и дает важные подсказки для восприятия ключевых частей тела. Насколько нам известно, наша работа является первой, в которой подобным образом обрабатываются последовательности облаков точек.

• Мы создаем среду глубокого обучения под названием SFCNet, которая сначала использует симметричную структуру для обработки последовательностей облаков точек. Он кодирует динамику локальных важных частей тела с помощью тонкого потока, а затем дополняет эти сложные детали к общему виду, захваченному грубым потоком. SFCNet может выделять важные части тела и фиксировать более различимые модели движений, решая проблему эффективного представления действий на основе облаков точек. • Представленная сеть SFCNet продемонстрировала свою превосходную точность на двух общедоступных наборах данных, NTU RGB+D 60 и NTU RGB+D 120, что доказывает, что наш метод имеет значительный потенциал в распознавании различных типов действий, таких как повседневные действия, медицинские и связанные действия и действия взаимодействия двух человек.

## 2. Сопутствующие работы

### 2.1. Распознавание 3D-действий на основе скелетов

Существующие методы распознавания трехмерных действий можно разделить на скелетные методы [ 12–17] и глубинные методы [3,7,12,18,19]. Обычно существует четыре основных подхода к распознаванию действий на основе скелетов. Первый — использовать CNN [12] для изучения пространственно-временных закономерностей по псевдоизображениям [13,14]. Каэтано и др. [20] представили эталонное изображение соединений древовидной структуры (TSRJI) для представления последовательностей скелета. Второй предполагает рассмотрение скелетных последовательностей как временных рядов [15–17] и использование магистральных сетей, таких как RNN [21], для извлечения признаков. Третий — рассматривать скелетные данные в виде графов [6,22] с суставами в качестве вершин и костями в качестве ребер и обращаться к GCN [16] для представления действий.

Например, ST-GCN [23] эффективно представлял временную динамическую информацию скелетных последовательностей, используя пространственно-временную свертку графов и стратегии разделения. SkeleMotion [5] собирал временную динамическую информацию, вычисляя размер и ориентацию суставов скелета в разных временных масштабах. Четвертый — закодировать скелет в токены с помощью Transformer. Плиццари и др. [24] использовали пространственно-временное внимание в Transformer и фиксировали динамические межкадровые отношения суставов. Однако, поскольку по-прежнему существуют серьезные проблемы с точной трехмерной оценкой позы человека [25,26], методы распознавания действий на основе скелетов страдают от каскадного снижения производительности из-за этой неизбежной исходной задачи.

### 2.2. Распознавание 3D-действий на основе глубины

Для распознавания трехмерных действий на основе глубины ранние подходы в основном представляют видео глубины с помощью ручных дескрипторов [19]. Ян и др. [7] построили карты движения глубины (DMM) путем суммирования межкадровых различий проецируемых кадров глубины. Затем они рассчитали гистограмму ориентированных градиентов (HOG), чтобы представить действия. Такие методы имеют ограниченную выразительную силу и поэтому обычно нуждаются в помощи в улавливании пространственно-временной информации. В последние годы методы глубокого обучения стали мейнстримом с развитием нейронных сетей. Большинство исследователей пытались сжать глубокое видео в изображения и анализировать закономерности движения с помощью CNN [12]. Камель и др. [27] вводили изображения глубины движения (DMI) и перемещали дескрипторы суставов (MJD) в CNN для распознавания действий. Чтобы закодировать пространственно-временную информацию о последовательностях глубин, Адриан и др. [28] предложили 3D-CNN для извлечения признаков движения. Кроме того, они предложили ConvLSTM [29] для накопления отличительных паттернов движения от длительных краткосрочных единиц. Сяо и др. [3] вращали виртуальную камеру в трехмерном пространстве для плотного проецирования необработанного видео глубины с разных точек обзора виртуального изображения и, таким образом, создавали многоракурсные динамические изображения. Что касается инвариантов перспективного представления, Kumar et al. [30] предложили ActionNet на основе CNN и обучили набору данных с несколькими изображениями, собранному с использованием пяти камер глубины. Гош и др. [31] вычислили многовидовое изображение истории движения с обнаружением границ дескриптора глубины (ED-MHI) в качестве входных данных многоп
Ван и др. [2] использовали сегментированные видеопоследовательности глубины для создания трех типов изображений с динамической глубиной. Однако двумерная карта глубины по-прежнему с трудом может полностью использовать трехмерные модели движения из-за ее компактной пространственной структуры [10].

В последнее время преобразование карт глубины в облака точек для обработки позволило добиться лучших результатов как в области распознавания, так и в области сегментации. Многочисленные исследования показали

что облака точек имеют значительные преимущества при представлении трехмерной пространственной информации благодаря своим характеристикам, таким как беспорядок и инвариантность вращения. Глубокое обучение облаком точек не только широко использовалось в задачах классификации и сегментации, но также продемонстрировало мышечную силу при реконструкции сцены [32] и обнаружении целей [ Однако вышеуказанные методы ориентированы только на объекты в статических облаках точек. При использовании облаков точек для распознавания трехмерных действий необходимо выделять динамические признаки по временным интервалам и особенностям внешнего вида всего процесса действия. Ключом к эффективной обработке последовательностей облаков точек является выбор подходящего метода анализа облака точек. Томас и др. [34] разработали метод, основанный на свертке на основе изображений , и использовали набор точек ядра для распределения веса каждого ядра. PointNet++ [11] как эффективный инструмент анализа и обработки наборов точек широко применяется для распознавания трехмерных действий на основе последовательностей облаков точек. Первый метод — 3DV [10], который выполняет 3D-вокселизацию последовательностей облаков точек и описывает 3D-вид по пространственной занятости, а для извлечения 3DV используется пул временных рангов. Этот метод в первую очередь фокусируется на общем движении действия и изменениях внешнего вида. Однако он игнорирует детали действия, такие как едва заметное движение руки, что ограничивает его способность точно отображать поведение. Следовательно, мы стремимся уловить важнейшие части действий и их деликатную информацию, чтобы признать их более надежными человеческими

## 3. Методика 3.1.

### Конвейер

Конвейер предлагаемой сети SFCNet показан на рисунке 1. Сначала каждый кадр глубины преобразуется в облако точек, чтобы лучше сохранить динамические характеристики и характеристики внешнего вида в трехмерном пространстве. Чтобы облегчить анализ пространственного использования и очертить локальное пространство, мы выполняем операции вокселизации над облаками точек. Затем мы создаем симметричную структуру для кодирования 3D-вокселей, где ключевые части и глобальная динамическая информация обрабатываются отдельно в локальном тонком потоке и глобальном грубом потоке. Затем мы присоединяем дескриптор интервальной частоты, чтобы дополнить информацию о движении. Мы используем PointNet++ [11] для захвата моделей движения и отправки агрегированного признака в классификатор для распознавания трехмерных действий.

### 3.2. Трехмерная генерация вокселей

Глубинное видео имеет преимущество в устойчивости к внешним помехам, таким как фон и свет, по сравнению с модальным режимом RGB, поскольку оно содержит информацию о глубине объекта действия. По сути, видео глубины — это своего рода временные ряды данных, состоящие из карт глубины, расположенных в хронологическом порядке. Математически видео глубины с t кадрами можно определить как D = {d1, d2, . . . , dt}, где di — карта глубины кадра t , в которой каждый пиксель представляет собой трехмерную координату (x, y, z) , а z — расстояние от камеры глубины. Поскольку невозможно классифицировать важность действия во временном измерении по одному критерию, равномерная выборка может помочь нам лучше понять общий процесс движения по сравнению со случайной выборкой [10]. Поэтому мы сначала равномерно сэмплируем видео глубины, чтобы облегчить вычислительную нагрузку, сохраняя при этом целостность действия. Последовательность глубин после выборки обозначается как Dˆ = {d1, d2, . . . , dT}, где T — количество кадров, а

по умолчанию 64.

Некоторые современные методы распознавания действий [35,36] предпочитают отображать кадры глубины в 2D- пространства для прямой обработки. Хотя эти подходы иногда могут обеспечить хорошую производительность , они не могут решить проблему неадекватного представления трехмерной информации. Поэтому, чтобы лучше отобразить движение человека в трехмерном пространстве, мы преобразуем каждый кадр di в облако точек P = {p1, p2, . . . , pn}, где n — количество точек, образуя таким образом последовательность облаков точек S = {P1, P2, . . . , ПТ}. При создании облаков точек необходимы внутренние параметры камеры, поскольку они определяют модель изображения камеры, включая

фокусное расстояние и координаты главной точки (cx, cy). Для каждого пикселя (x, y, z) изображения глубины соответствующее ему облако точек p(x ) можно получить по следующей формуле:

$$p(x', y', z') = \left(\frac{(x - cx) \times z}{fx}, \frac{(y - cy) \times z}{fy}, \frac{z}{fz}\right) \quad (1)$$

где fx и fy представляют собой фокусное расстояние камеры глубины в горизонтальном и вертикальном направлениях, которое можно получить из параметров устройства. fz по умолчанию имеет значение 1 .

В отличие от традиционных изображений (обычных структурированных данных), точки в облаке точек неупорядочены, поэтому их сложно обрабатывать. Многие существующие алгоритмы разработаны для данных регулярной сетки. Однако неупорядоченное облако точек представляет собой группу случайно распределенных точек в трехмерном пространстве, поэтому их структуру сложно обрабатывать и анализировать напрямую. Чтобы решить эту проблему, мы преобразуем облако точек в обычную трехмерную сетку (пространство вокселов) путем вокселизации, чтобы упорядочить представление облака точек. Сначала мы определяем размер сетки вокселов Vgrid = (Vx, Vy, Vz) в трехмерных координатах, что определяет разрешение процесса вокселизации. Каждая ячейка в этой сетке является потенциальным вокселем, а размер каждой ячейки обозначается как Vvoxel(dx, dy, dz). Учитывая точку p(x, y, z) в облаке точек, она отображается в сетку путем нахождения соответствующего индекса воксела Vindex(x, y, z) согласно следующему уравнению:

$$\text{Виндекс}(x, y, z) = \left(\frac{x - x\min}{dx}, \frac{y - y\min}{dy}, \frac{z - z\min}{dz}\right) \quad (2)$$

где xmin, ymin и zmin — минимальные координаты всех облаков точек. dx, dy и dz рассчитываются как общий размер, разделенный на количество ячеек в каждом измерении (Vx, Vy, Vz). Функция пола . округляет до ближайшей точки. Мы определяем, что воксель занят, если он содержит облако точек. Затем информацию о трехмерном внешнем виде можно описать путем наблюдения за тем, были ли вокселы заняты или нет, без учета исключенной точки, как показано в уравнении (3):

$$B^{T}_{\text{воксель}}(x, y, z) = \begin{cases} 1, & \text{если } V^{T}_{\text{воксель}}(x, y, z) \text{ занят} \\ 0, & \text{в противном случае} \end{cases} \quad (3)$$

где $B^{T}_{voxel}(x, y, z)$ указывает определенный воксель в t-м кадре. (x, y, z) представляет собой обычный индекс трехмерного положения, т.е. Vindex в уравнении (2). Эта стратегия приносит две основные прибыли. Во-первых, полученные наборы двоичных 3D-вокселей являются регулярными, как показано на рисунке 2. Таким образом, сложность обработки облака точек снижается. Кроме того, вокселизация может эффективно сжимать облака точек, поскольку соседние вокселы могут иметь схожие характеристики. Такое сжатие не только уменьшает количество точек, но также помогает уменьшить накладные расходы на хранение и вычисления.

### 3.3. Идентификация и представление ключевых частей

Важнейшим вопросом в задачах распознавания трехмерных действий является эффективный захват и представление динамических элементов в последовательностях облаков точек. На данный момент методы оценки, основанные на потоке сцен [37,38], могут помочь понять трехмерное движение, но это требует очень много времени. В некоторых исследованиях используется пул временных рангов [3,39] для сохранения процессов движения в трехмерном пространстве путем разделения временных сегментов. Эти методы могут собирать больше временной информации, но часто разделяют лишь небольшое количество интервалов, что приводит к получению более крупнозернистых динамических характеристик. Мы предлагаем модуль идентификации и кодирования важнейших деталей, чтобы лучше сосредоточиться на критической динамике во время движения. Он может извлекать основные части из глобального трехмерного воксельного пространства в соответствии со временем занятости пространства и кодировать детали действий посред
В частности, мы сначала анализируем занятость пространства, создавая трехмерное пространство U с точными границами в виде последовательностей облаков точек для каждого пространственного местоположения и начальными значениями.

из U равны 0. Мы обрабатываем m равномерных групп последовательности Dˆ по порядку. Затем 3D
Использование пространства можно рассчитать по уравнению (4):

$$\text{Увоксель}(x, y, z) = \begin{matrix} \text{Uvoxel}(x, y, z) + 1, \text{vi} / = 0 \\ \text{Uvoxel}(x, y, z), \text{иначе} \end{matrix} \quad (4)$$

Общую занятость пространства u для каждой позиции можно получить, посчитав все m
наборы точек. Кроме того, поскольку множества точек естественным образом упорядочены во времени, мы можем легко
запишите первый и последний раз, f и l, соответственно, для каждого пространственного местоположения. Затем,
мы определяем порог θ для разделения заметного локального пространства. Места, занимаемые
менее θ рассматриваются как случайный шум, а те, которые записывают больше и меньше m
составляют критические части движения S согласно уравнению (5):

$$\text{Своксель}(x, y, z) = \begin{matrix} 0, & \text{Uvoxel}(x, y, z) < \theta \\ 1, \theta & \text{Uvoxel}(x, y, z) \quad m \end{matrix} \quad (5)$$

Если значение Svoxel(x, y, z) равно 0, это означает, что воксель принадлежит глобальному
пространство G; в противном случае оно принадлежит локальному пространству L. По сравнению с облаком точек
методы обработки [10,40], которые обычно используют унифицированные операции понижения дискретизации,
предлагаемый метод более эффективен, особенно для действий, включающих лишь небольшое количество
частей конечностей, ведь разделение локального пространства может не только преодолеть фон
влияет в определенной степени, но также эффективно увеличивает содержание золота в точке отбора проб
данные. Как показано на рисунке 2, локальное пространство L полностью сохраняет подробную информацию о
основные части тела, что дает важные подсказки для распознавания трехмерных действий, в то же время существенно
сокращение избыточности.

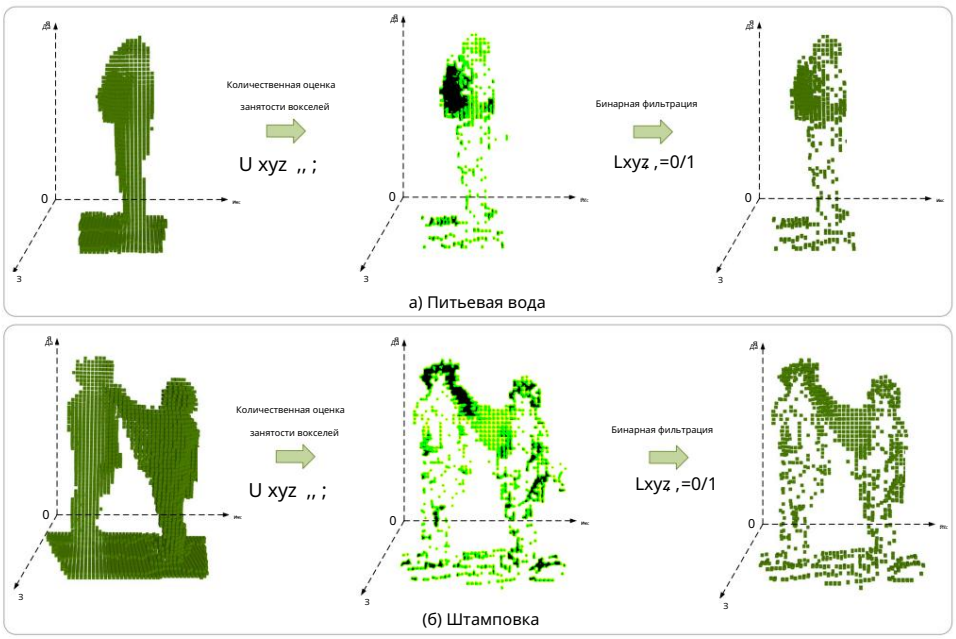



Рисунок 2. Процесс разделения локального пространства. Мы количественно оцениваем вклад вокселя в действие с помощью
его показатель занятости пространства. Установив порог, критические части можно разделить как сжатый
локальное пространство, которое удаляет избыточную информацию и снижает вычислительную нагрузку.

### 3.4. Симметричное извлечение признаков

Для обработанных 3D-вокселей наиболее интуитивным способом является использование 3DCNN [41,42], но
он ограничен размером воксела и требует много времени. Мы решили использовать PointNet++ [11] в качестве
экстрактор признаков в этой работе, как показано на рисунке 3. Он предназначен специально для иерархического
функция обучения на неупорядоченных наборах точек в метрическом пространстве, что позволяет захватывать локальные
мелкозернистые узоры в облаках точек. Для этого PointNet++ разделяет точку

облако на перекрывающиеся локальные регионы на основе метрики расстояния в базовом пространстве. Чтобы получить подробные трехмерные визуальные подсказки, PointNet++ рекурсивно использует PointNet [43] для извлечения локальных объектов, которые затем объединяются для глобального анализа внешнего вида. PointNet++ — отличная альтернатива 3DCNN, поскольку она хорошо справляется с захватом локальных 3D-шаблонов, необходимых для распознавания действий. Более того, его применение относительно просто и требует лишь преобразования 3D-вокселей в наборы точек.

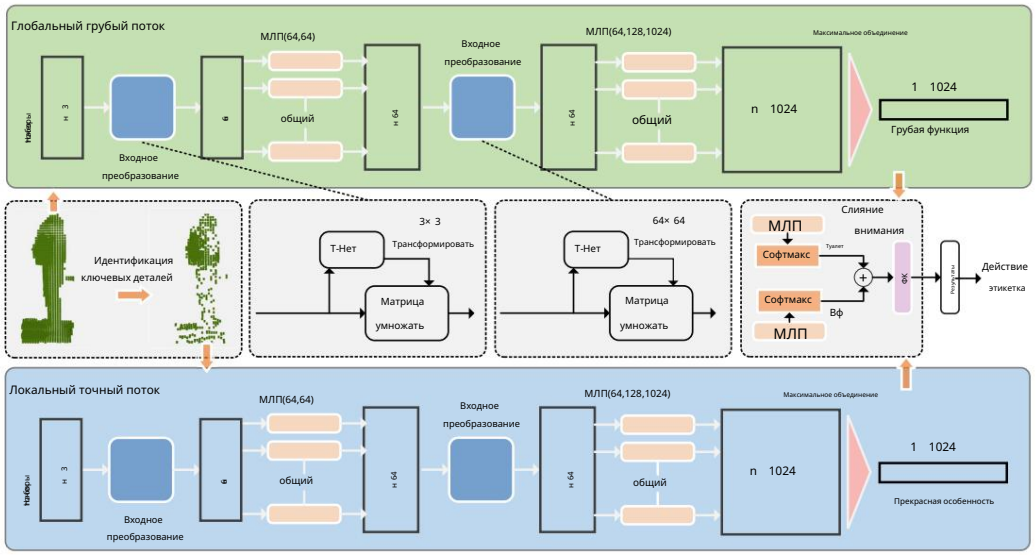



Рисунок 3. Структура сети SFCNet. Это симметричная сетевая структура, включающая глобальный грубый поток и локальный тонкий поток. Глобальный грубый поток принимает глобальную информацию о внешнем виде в качестве входных данных, тогда как локальный точный поток принимает только сжатую информацию о ключевой части. Характеристики каждого потока извлекаются с помощью PointNet++ и объединяются с обучаемыми весами в качестве окончательных характеристик действий, которые отправляются в классификатор для распознавания трехмерных действий.

Чтобы соответствовать PointNet++, каждая точка p(x, y, z) затем абстрагируется как воксель Vvoxel(x, y, z) с описательным признаком (x, y, z, I), где I — интервал-частота дескриптор, включающий три переменные (o, f , l) , которые обозначают начальные временные метки, конечные временные метки и общую частоту занятости вокселей соответственно. Мы присоединяем I к исходным 3D-положениям вокселей, чтобы получить 6D-точки p'(x, y, z, o, f , l), как показано на рисунке 4. Математически для вокселя Vvoxel мы можем определить время начала и окончания (o и f) занятия им пространства воксела(x, y, z) = 1 с использованием уравнения (3), которое указывает индекс времени, когда условия

$$V_t^{u} \; и \; V$$

voxel(x, y, z) = 0 сначала выполняются. Кроме того, общая занятость пространства u может быть рассчитана по уравнению (4). Следовательно, частота занятости l может быть вычислена как l = u/(f o). Дескриптор интервал-частота охватывает временной интервал, и общая частота занятости пространственных местоположений может не только помочь нам отличить обратные действия, охватывающие одно и то же пространство, но также помочь выделить различия между действиями, сохраняя подробную информацию в процессе действия. Наконец, полученные наборы точек используются в качестве входных данных PointNet++ для извлечения признаков действия.

Кроме того, мы проектируем двухпотоковую сеть для обработки облака точек в глобальном и локальном пространстве соответственно (как показано на рисунке 3). Глобальное пространство содержит все вокселизированные точки, а глобальный грубый поток фиксирует общие закономерности движения. Учитывая , что жизненно важные части тела могут предоставлять более целенаправленную и различительную динамическую информацию для распознавания действий, мы разделяем вокселизированные точки основных частей тела на локальные пространства и вводим тонкий поток для извлечения особенностей. После этого модуль объединения функций настроен для точного и грубого представления действий. Учитывая особенности различных действий, их зависимость от глобальных и локальных особенностей различна. Для действий, которые включают только движения конечностей, таких как взмахи руками и ногами, модель должна подчеркивать локальные детализированные особенности потока. Напротив, для крупных движений всего тела, таких как падение и прыжки, модель должна сосредоточиться на характ

Для этого восприятия, специфичного для действия, мы используем модуль слияния признаков с механизмом тонкого подхода. Сначала мы проецируем признаки внимания n× 1024 и Xc R n× 1024, извлеченным из Xf R и грубые потоки соответственно в нижнее пространство признаков, чтобы уменьшить вычислительную нагрузку и получить X и X'c . Затем промежуточные признаки извлекаются с помощью многослойного f-персептрона (MLP). После этого обучаемые веса Wf и Wc глобального грубого потока и локального тонкого потока получаются с помощью функции активации SoftMax соответственно.

Наконец, глобальные и локальные характеристики объединяются, как показано в уравнении (6), для получения характеристики движения Xˆ :

$$\begin{aligned} Wf &= \text{SoftMax}\left(\text{MLP}\left(X_f'\right)\right) \\ Wc &= \text{SoftMax}\left(\text{MLP}\left(X_c'\right)\right) \\ \hat{X} &= \varphi\left(\sum_{i=1}^{K} Wi_{ff}\left(X_f^{(i)}\right) + Wi\text{-}CX_c'^{(i)}\right) \end{aligned} \tag{6}$$

где φ — линейный слой, а K — длина признаков, выдаваемых MLP. Наконец, полностью связный слой размерно восстановил объединенный признак Xˆ , и классификатор SoftMax получил окончательные оценки прогноза. В отличие от существующих методов, которые напрямую анализируют состояние движения всего облака точек, наша работа фокусируется на критической части тела при выполнении действий, что помогает преодолеть влияние избыточных данных, таких как фон, на распознавание трехмерных действий. Кроме того, детали важнейших частей и внешний вид человеческого тела дополняют друг друга, что стимулирует извлечение отличительных признаков для распознавания трехмерных действий.

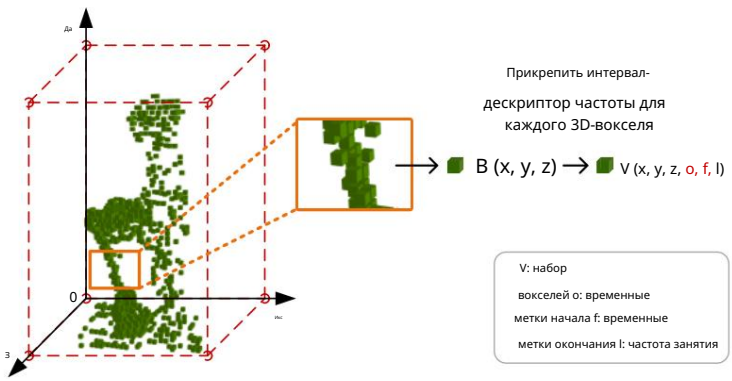



Рисунок 4. Иллюстрация дополнительного интервально-частотного дескриптора. Он содержит начальные временные метки, конечные временные метки и общую частоту занятости вокселей и обозначается как (o, f , l) в разделе 3.4; таким образом, трехмерная пространственная информация о глубине преобразуется в характеристики шести каналов.

## 4. Эксперименты 4.1. Наборы данных

Набор данных NTU RGB+D 60. NTU RGB+D 60 [44] представляет собой крупномасштабный набор данных для распознавания 3D-действий, который содержит около 56 880 образцов действий RGB+D. Он использует Microsoft Kinect V2 для захвата 60 категорий действий, выполняемых 40 субъектами. Набор данных соответствует двум принципам оценки. В случае перекрестного обзора в качестве тестового набора использовались образцы, снятые камерой 1, а в качестве обучающего набора рассматривались камеры 2 и 3. То есть количество тестовых выборок составляет 18 960, а для обучения используется 37 920 выборок. В случае перекрестного анализа данные делятся на тестовый набор из 16 560 образцов и обучающий набор из 40 320 образцов на основе идентификатора субъекта.

Набор данных NTU RGB+D 120. NTU RGB+D 120 [45] представляет собой обширный набор данных для распознавания трехмерных действий, состоящий из 114 480 образцов и 120 категорий действий, выполненных 106 субъектами. Этот набор данных содержит ежедневные действия, действия, связанные с медициной, а также действия взаимодействия двух человек . Образцы собраны в различных местах и фонах, которые обозначены как 32 установки. Помимо общих кросс-предметных настроек, кросс-настройка

введена оценка, где обучающий набор состоит из выборок с нечетными идентификаторами настройки, и набор для тестирования состоит из остальных.

### 4.2. Подробности обучения

По умолчанию SFCNet и ее варианты обучаются с использованием Adaptive Moment. Оптимизатор оценки (Адам) на 60 эпох в рамках платформы глубокого обучения PyTorch, если не указано иное. Мы используем стандартную перекрестную энтропийную потерю и применяем методы увеличения данных, такие как случайное вращение, сглаживание и исключение, к обучающим данным. Скорость обучения начинается с 0,001 и снижается на 0,5 каждые десять эпох. Чтобы обеспечить справедливость , мы строго следуем схеме выборочной сегментации двух наборов данных в соответствии с ориентиры.

### 4.3. Анализ параметров

Размер 3D-вокселей. Облака точек обычно состоят из большого количества неупорядоченных объектов. наборы точек. Из-за сложности этих данных их хранение может занять очень много времени. и процесс. Чтобы решить эту проблему, мы встраиваем неупорядоченные точки в 3D-пространстве в регулярная структура сетки посредством растеризации. Это преобразует облако точек в 3D-вокселы. в зависимости от занятости пространства, дискретизируя непрерывное трехмерное пространство в регулярную сетку. Этот процесс обеспечивает регулярную и хорошо понятную структуру, что сокращает вычислительные затраты. сложность и сжимает данные облака точек, значительно сокращая вычислительные затраты. груз. Очень важно правильно вокселировать облако точек, поскольку размер 3D-воксель определяет силу сжатия облака точек и степень детализации. представление облака точек. Чтобы изучить влияние вокселизации на результаты, мы оценили производительность SFCNet на наборе данных NTU RGB+D 60 для различных размеров воксели. Результаты представлены в таблице 1, что указывает на то, что модель работает лучше всего для Размер кубика 35 мм. Установка слишком большого или слишком маленького размера может привести к снижению точности.

Таблица 1. Производительность набора данных NTU RGB+D 60 с вокселизацией разного размера.

| Размер воксела (мм) | Межтематический | Перекрестный вид |
|---|---|---|
| 25 × 25 × 25 35 | 87,1% | 94,9% |
| × 35 × 35 45 × | 89,9% | 96,7% |
| 45 × 45 55 × 55 | 88,1% | 95,5% |
| × 55 | 86,5% | 93,6% |

Установка порога θ. Поведение человека обычно включает в себя только движение определенные части тела, такие как размахивание руками, ходьба ногами, поворот головы и т. д. Эта локальность означает что поведенческий анализ должен быть больше сосредоточен на основных частях тела, а не на все тело. С помощью переменной частоты присутствия l в интервале-частоте дескриптор, мы можем описать взаимодействие каждого воксела, что положительно связано с вклад части тела в процесс выполнения действия. Поскольку мы пробуем последовательность действий по глубине на группы равной продолжительности, частота занятости l положительная коррелирует с количеством занятий u в разделе 3.3. Затем порог θ используется для оценить внимание, уделяемое части тела. Чтобы исследовать влияние порога, мы сравните производительность SFCNet в наборе данных NTU RGB+D 60 с различными значениями. Результаты представлены в таблице 2. Оптимальный результат может быть получен, когда θ равен 30. Наше исследование показало, что небольшие изменения θ, не более чем на 5, приводили к в колебаниях точности, подчеркивая важность исследования порогов и поломка значительных частей тела.

Таблица 2. Производительность на наборе данных NTU RGB+D 60 с различными значениями θ.

| Значение порога θ | Межтематический | Перекрестный вид |
|---|---|---|
| 15 | 79,5% | 85,1% |
| 20 | 83,5% | 93,2% |
| 25 | 86,5 % | 94,4% |
| 30 | 89,9% | 96,7% |
| 35 | 87,3% | 94,9% |
| 40 | 86,9% | 93,7% |

4.4. Исследование абляции

Эффективность интервально-частотного дескриптора. Только исходные данные 3D-облака точек. содержат информацию о местоположении точек в трехмерном пространстве. Даже если измерение времени вводится в последовательности облаков точек, все еще сложно описать общую картину. пространственно-временную динамику точки, полагаясь только на эти подсказки. Мы разработали дескриптор интервальной частоты, который фиксирует время начала и количество вокселей заполняемость. Эта дополнительная информация помогает нам всесторонне описать человека. поведение путем захвата дополнительных функций движения. Мы провели исследования абляции на Набор данных NTU RGB+D 60, из которого мы удалили информацию о признаках движения в два потока, и входные данные наборов точек в SFCNet имели только трехмерные координаты (x, y, z). Результаты сравнения представлены в таблице 3. Мы заметили, что без дополнительных трехмерных функций произошло значительное снижение производительности SFCNet более чем на 10%.

Таблица 3. Эффективность интервально-частотного дескриптора на NTU RGB+D 60.

| Точечная функция | Межтематический | Перекрестный вид |
|---|---|---|
| (x, y, z) | 78,0% | 82,3% |
| (x, y, z, o, f, l) | 89,9% | 96,7% |

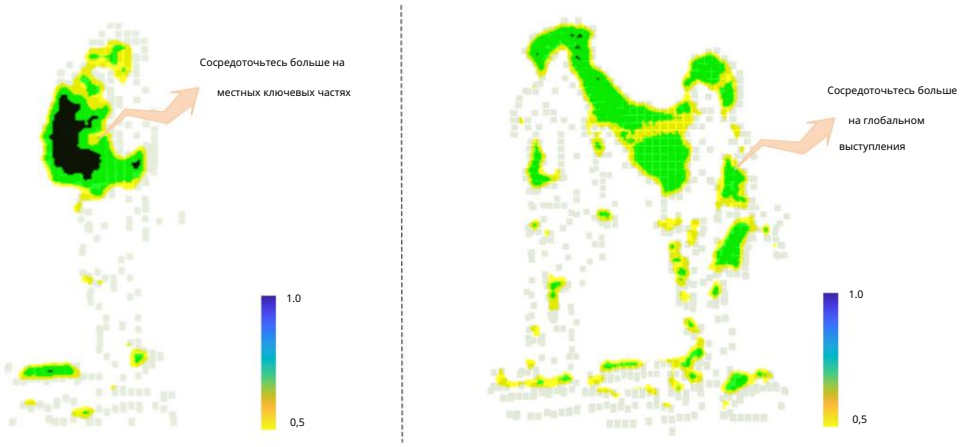
Это указывает на то, что дескриптор интервальной частоты эффективно представляет динамическую динамику. особенности всего процесса действия, что играет жизненно важную роль в распознавании трехмерных действий.

Эффективность объединения двух потоков функций. В разных действиях человека содержится разное глобальную и локальную динамику, мы описываем общую картину движения всего человеческого организма. тело во время действия через глобальный грубый поток. Напротив, местный мелкий ручей описывает динамику важнейших частей тела, уделяя больше внимания деталям и локальные особенности действий и помогает уловить тонкие изменения и сложные модели действий. Результаты в Таблице 4 показывают, что предлагаемая нами сеть SFCNet объединяет мелко-грубые характеристики два потока могут более полно понимать и распознавать действия человека. Сначала мы обсудим производительность распознавания в однопоточном состоянии. Видно, что местные поток имеет более высокую способность представлять действие, чем глобальный поток, что указывает на то, что он Важно обратить внимание на основные части, чтобы устранить избыточность. Кроме того, мы сравниваем три различные стратегии объединения функций, чтобы продемонстрировать превосходство объединение функций на основе внимания, предложенное в SFCNet (см. уравнение (6)). Как показано в таблице 4, SFCNet (слияние) имеет очевидные преимущества перед собственным каскадом или аддитивным слиянием. стратегии. Основная причина заключается в том, что объединение функций, основанное на внимании, может адаптивно распределять внимание модели к особенностям глобального грубого потока и локального мелкого потока.

Как показано на рисунке 5, питьевая вода предполагает только взаимодействие рук и головы, а амплитуда движения небольшая, поэтому модель подчеркивает особенности местного поток для захвата детальных моделей движения. С другой стороны, перфорирование предполагает взаимодействие двух людей, а ударное движение большое и мощное, поэтому больше внимания уделяется общему внешнему виду кузова, подчеркивая при этом некоторые детали руки и голова. Этот механизм извлечения функций, специфичных для конкретных действий, улучшает способность SFCNet к обобщению и точность распознавания трехмерных действий.

Таблица 4. Эффективность объединения двух потоков в наборе данных NTU RGB+D 60.

| Входной поток | Межтематический | Перекрестный вид |
|---|---|---|
| 1s-SFCNet (L) 1s-SFCNet (G) | 85,0% 81,8% | 94,6% 86,6% |
| SFCNet (конкат) | 88,9% | 94,8% |
| SFCNet (добавить) | 86,7% | 93,9% |
| SFCNet (слияние) | 89,9% | 96,7% |




а) Питьевая вода (б) Штамповка

Рисунок 5. Визуализация внимания к функциям. Визуализируем тепловую карту питьевой воды и пробивания.

4.5. Сравнение с существующими методами

Чтобы оценить производительность предлагаемой SFCNet, мы сравним ее с существующими методами. на двух больших наборах эталонных данных, как показано в таблицах 5 и 6. Мы разделяем существующие 3D-действия методы распознавания на скелетные и глубинные методы. В скелетных методах мы сравниваем различные базовые методы, включая CNN [5], LSTM [15,46] и ГЦН [6,23,47]. Для методов, основанных на глубине, мы сравниваем методы, основанные на 2D-изображениях [19,35,48], Методы на основе 3D CNN [28] и методы на основе 3D вокселей [10]. Результаты сравнения представлено в таблицах 5 и 6. Для набора данных NTU RGB+D 60 предлагаемая сеть SFCNet достигает Точность 89,9% и 96,7% в случае настроек перекрестного просмотра и объекта. Кроме того, мы сравните SFCNet с двумя методами, основанными на мультимодальных данных. По сравнению с ED-MHI [31], который объединяет данные глубины и скелета, наш метод повышает точность на 4,3% в межпредметная настройка. TS-CNN-LSTM [49] объединил данные трех модальностей, а именно RGB, глубины, и скелет, но он на 2,6% и 4,9% ниже, чем у SFCNet в настройках перекрестного просмотра и настройки перекрестного просмотра соответственно. Для набора данных NTU RGB+D 120 SFCNet также достигает конкурентоспособных результаты, достигая точности 83,6% и 93,8% при настройках перекрестного просмотра и перекрестного просмотра, соответственно. В целом, SFCNet эффективен и превосходен, превосходя традиционные методы, использующие ручное извлечение признаков [19,35,50] и методы глубокого обучения, которые сжимают глубину видео в изображения для обработки [2,3,36] или последовательности облаков точек [10]. Результаты эксперимента доказать, что SFCNet превосходно фиксирует дискриминационные модели человеческого поведения и таким образом, это полезно для распознавания трехмерных действий.

Таблица 5. Сравнение различных методов по точности распознавания действий (%) на НТУ RGB+D 60 наборов данных.

| Метод | Межтематический | Перекрестный вид | Год |
|---|---|---|---|
| | ввод: 3D-скелет | | |
| ГКА-ЛСТМ [15] | 74,4 | 82,8 | 2017 год |
| Двухпотоковое внимание LSTM [46] | 77,1 | 85,1 | 2018 год |
| СТ-ГКН [23] | 81,5 | 88,3 | 2018 год |
| Скеледвижение [5] | 69,6 | 80,1 | 2019 год |
| AS-GCN [6] 2s- | 86,8 | 94,2 | 2019 год |
| AGCN [47] | 88,5 | 95,1 | 2019 год |
| СТ-ТР (новый) [24] | 89,9 | 96,1 | 2021 год |
| DSwarm-Net (новый) [51] | 85,5 | 90,0 | 2022 год |
| ЭкшнНет [30] | 73,2 | 76,1 | 2023 год |
| СГМСН (новый) [52] | 90,1 | 95,8 | 2023 год |
| | Входные данные: карты глубины. | | |
| ХОН4Д[19] | 30,6 | 7.3 | 2013 |
| ХОГ2 [35] | 32,2 | 22,3 | 2013 |
| СНВ [50] | 31,8 | 13,6 | 2014 год |
| Ли. [36] | 68,1 | 83,4 | 2018 год |
| Ван. [2] | 87,1 | 84,2 | 2018 год |
| МВДИ [3] | 84,6 | 87,3 | 2019 год |
| 3DV-PointNet++ [10] | 88,8 | 96,3 | 2020 год |
| ДОГВ (новый) [53] | 90,6 | 94,7 | 2021 год |
| 3DFCNN [28] | 78,1 | 80,4 | 2022 год |
| 3D-обрезка [54] | 83,6 | 92,4 | 2022 год |
| ConvLSTM (новый) [29] | 80,4 | 79,9 | 2022 год |
| СВВМС (новый) [48] | 83,3 | 87,7 | 2023 год |
| PointMapNet (новый) [55] | 89,4 | 96,7 | 2023 год |
| SFCNet (наш) | 89,9 | 96,7 | - |
| | Ввод: мультимодальности | | |
| ЭД-МХИ [31] | 85,6 | - | 2022 год |
| ТС-CNN-LSTM [49] | 87,3 | 91,8 | 2023 год |

Таблица 6. Сравнение различных методов по точности распознавания действий (%) на НТУ RGB+D 120 наборов данных.

| Метод | Межтематический | Крест-набор | Год |
|---|---|---|---|
| | Сырьё: 3D-скелет. | | |
| ГКА-ЛСТМ [15] | 58,3 | 59,3 | 2017 год |
| Карта эволюции позы тела [56] | 64,6 | 66,9 | 2018 год |
| Двухпотоковое внимание LSTM [46] | 61,2 | 63,3 | 2018 год |
| СТ-ГКН [23] | 70,7 | 73,2 | 2018 год |
| Базовый уровень NTU RGB+D 120 [45] | 55,7 | 57,9 | 2019 год |
| ФССеть [57] | 59,9 | 62,4 | 2019 год |
| Скеледвижение [5] | 67,7 | 66,9 | 2019 год |
| ЦРДЖИ [20] | 67,9 | 62,8 | 2019 год |
| AS-GCN [6] 2s- | 77,9 | 78,5 | 2019 год |
| AGCN [47] | 82,9 | 84,9 | 2019 год |
| СТ-ТР (новый) [24] | 82,7 | 84,7 | 2021 год |
| СГМСН (новый) [52] | 84,8 | 85,9 | 2023 год |
| | Входные данные: карты глубины. | | |
| АПСР [45] | 48,7 | 40,1 | 2019 год |
| 3DV-PointNet++ [10] | 82,4 | 93,5 | 2020 год |
| ДОГВ (новый) [53] | 82,2 | 85,0 | 2021 год |
| 3D-обрезка [54] | 76,6 | 88,8 | 2022 год |
| SFCNet (наш) | 83,6 | 93,8 | - |

## 5. Обсуждение

Чтобы проанализировать преимущества и недостатки предлагаемого метода, мы представили точность распознавания SFCNet на наборе данных NTU RGB+60 при межпредметных настройках для каждой категории. Результаты показаны в виде матрицы путаницы на рисунке 6 (слева). Мы выбрали некоторые сбивающие с толку действия для более четкого отображения и увеличили их локально, как показано на рисунке 6 (справа). Результаты показывают, что SFCNet обладает надежной способностью анализа действий человека с точностью распознавания более 90% в большинстве категорий. Например, он достиг 100% точности прыжков и 99% точности прыжков вверх. Однако SFCNet смущен признанием некоторых подобных действий. Например , чтение и письмо, ношение обуви и снятие обуви — самые запутанные пары примеров. Кроме того, 25% тех, кто играл на своих телефонах, были ошибочно отнесены к чтению (9%), письму (8%) и набору текста на клавиатуре (8%). Точность набора текста на клавиатуре составляет всего 66%, а 12% образцов ошибочно классифицируются как письмо. В результате анализа мы обнаружили, что эти действия имеют лишь незначительные различия, а амплитуда движений невелика. Это основная причина, по которой такие действия трудно различить.

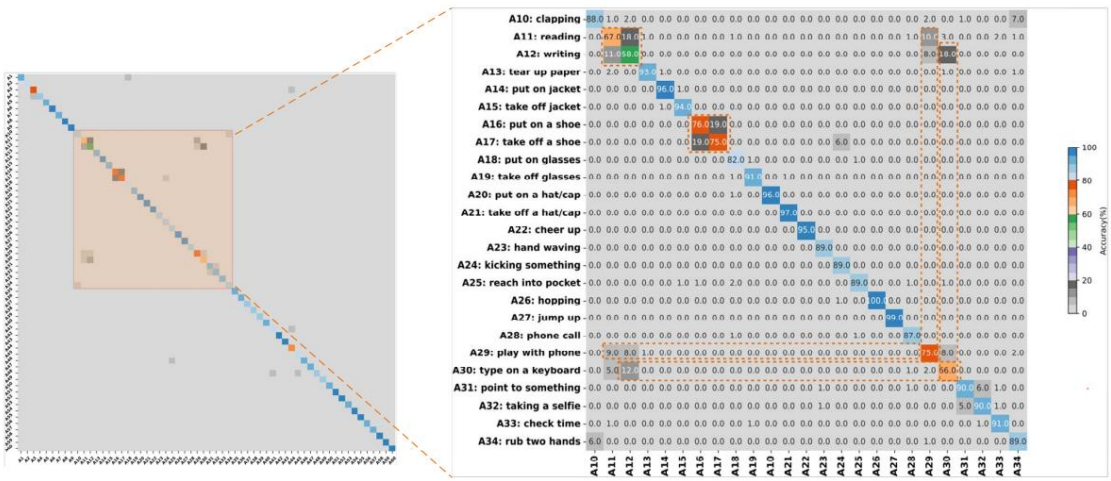



Рисунок 6. Матрица ошибок для точности распознавания в зависимости от класса. Изображение слева содержит все категории действий. Чтобы подчеркнуть некоторые действия по запутыванию, справа показано локальное увеличение .

## 6. Выводы

В этой статье мы предлагаем симметричную нейронную сеть SFCNet для распознавания трехмерных действий по последовательностям облаков точек. Он содержит глобальный грубый поток и локальный точный поток, который использует PointNet++ в качестве средства извлечения признаков. Последовательности облаков точек регуляризуются как структурированные наборы вокселов, к которым добавляется предложенный интервально-частотный дескриптор для создания 6D-объектов, которые фиксируют пространственно-временную динамическую информацию. Глобальный грубый поток фиксирует грубые модели действий по внешнему виду человеческого тела, а локальный деликатный поток извлекает детализированные характеристики конкретных действий из критических частей. После объединения функций SFCNet может анализировать различительные модели движения, которые включают общие пространственные изменения и подчеркивают важные детали от начала до конца. Согласно результатам экспериментов на двух больших наборах эталонных данных, NTU RGB+D 60 и NTU RGB+D 120, SFCNet эффективен для распознавания трехмерных действий и имеет потенциал для приложений диста

Однако предлагаемая сеть SFCNet по-прежнему имеет ограничения в различении аналогичных действий. Наша будущая работа будет сосредоточена на распознавании подобных действий и выявлении тонких закономерностей для повышения точности.

Вклад автора: концептуализация, CL и QH; методология, CL, WQ и XL; программное обеспечение, QH и YM; валидация, CL, WQ и XL; формальный анализ, QH и YM; расследование, CL и WQ; ресурсы, QH и YM; курирование данных, WQ; письмо — подготовка первоначального проекта, CL и WQ; написание — просмотр и редактирование, CL, YM, WQ, QH и XL; визуализация, WQ; надзор,

QH и YM; администрирование проекта, QH и YM; приобретение финансирования, QH, YM и CL. Все авторы прочитали и согласились с опубликованной версией рукописи.

Финансирование: данное исследование финансировалось Программой инновационных исследований и практики последипломного образования провинции Цзянсу (номер гранта KYCX23_0753), Фондами фундаментальных исследований для центральных университетов (номер гранта B230205027), Программой ключевых исследований и разработок Китая (номер гранта 2022YFC3005401). , Ключевая программа исследований и разработок Китая, провинция Юньнань (номер гранта 202203AA080009), 14-й пятилетний план педагогической науки провинции Цзянсу (номер гранта D/2021/01/39) и Исследовательский проект реформы высшего образования Цзянсу. (грант № 2021JSJG143); АПК финансировался Фондами фундаментальных исследований центральных университетов.

Заявление Институционального наблюдательного совета: Неприменимо.

Заявление об информированном согласии: Не применимо.

Заявление о доступности данных. Наборы данных NTU RGB+D 60 и NTU RGB+D 120, используемые в этом документе, являются общедоступными, бесплатными и доступны по адресу: https://rose1.ntu.edu.sg/dataset/actionRecognition/ . (по состоянию на 21 декабря 2020 г.).

Конфликты интересов: Авторы заявляют об отсутствии конфликта интересов.

Рекомендации

1. Риаз, В.; Гао, К.; Азим, А.; Сайфулла; Букс, Дж.А.; Улла, А. Система прогнозирования аномалий дорожного движения с использованием сети прогнозирования. Дистанционная сенсорика 2022, 14, 1–19.
2. Ван П.; Ли, В.; Гао, З.; Тан, К.; Огунбона, П.О. Крупномасштабное распознавание трехмерных действий на основе объединения по глубине с помощью сверточных нейронных сетей. IEEE Транс. Мультимед. 2018, 20, 1051–1061.
3. Сяо, Ю.; Чен, Дж.; Ван, Ю.; Цао, З.; Чжоу, Джей Ти; Бай, Х. Распознавание действий для видео глубины с использованием многоракурсных динамических изображений. Инф. наук. 2019, 480, 287–304.
4. Ли, К.; Хуан, К.; Ли, Х.; Ву, К. Распознавание действий человека на основе многомасштабных карт объектов из видеопоследовательностей глубины. Мультимед. Инструменты Прил. 2021, 80, 32111–32130.
5. Каэтано, К.; Сена, Дж.; Бремон, Ф.; Дос Сантос, JA; Шварц, В.Р. Skelemotion: новое представление последовательностей суставов скелета, основанное на информации о движении для распознавания трехмерных действий. В материалах Международной конференции IEEE по передовому видеонаблюдению и сигнальному наблюдению, Тайбэй, Тайвань, 18–21 сентября 2019 г.; IEEE: Пискатауэй, Нью-Джерси, США, 2019 г.; стр. 1–8.
6. Ли, М.; Чен, С.; Чен, Х.; Чжан, Ю.; Ван, Ю.; Тиан, К. Сверточные сети на графах с действием и структурой для распознавания действий на основе скелетов . В материалах конференции IEEE/CVF по компьютерному зрению и распознаванию образов, Лонг-Бич, Калифорния, США, 15–20 июня 2019 г.; стр. 3595–3603.
7. Ян, Х.; Чжан, К.; Тиан, Ю. Распознавание действий с использованием гистограмм ориентированных градиентов на основе карт движения глубины. В материалах Международной конференции ACM по мультимедиа, Нара, Япония, 29 октября – 2 ноября 2012 г.; стр. 1057–1060.
8. Эльмадани, Нидерланды; Привет.; Гуан, Л. Объединение информации для распознавания действий человека через бисетную/мультимножественную глобальную локальность сохранение канонического корреляционного анализа. IEEE Транс. Процесс изображения. 2018, 27, 5275–5287.
9. Он, К.; Чжан, Х.; Рен, С.; Сан, Дж. Глубокое остаточное обучение для распознавания изображений. В материалах конференции IEEE по компьютерному зрению и распознаванию образов, Лас-Вегас, Невада, США, 27–30 июня 2016 г.; стр. 770–778.
10. Ван Ю.; Сяо, Ю.; Сюн, Ф.; Цзян, В.; Цао, З.; Чжоу, Джей Ти; Юань, Дж. 3DV: 3D-динамический воксель для глубокого распознавания действий в видео. В материалах конференции IEEE/CVF по компьютерному зрению и распознаванию образов, Сиэтл, Вашингтон, США, 14–19 июня 2020 г.; стр. 508–517.
11. Ци, ЧР; Йи, Л.; Су, Х.; Гибас, Л. Дж. Pointnet++: Глубокое иерархическое изучение функций наборов точек в метрическом пространстве. Адв. Нейронная инф. Процесс. Сист. 2017, 30, 5105–5114.
12. Ван П.; Ли, В.; Гао, З.; Чжан, Дж.; Тан, К.; Огунбона, П.О. Распознавание действий по картам глубины с использованием глубокой свертки нейронные сети. IEEE Транс. Хм. -Max. Сист. 2015, 46, 498–509.
13. Ке, К.; Беннамун, М.; Ан, С.; Сохель, Ф.; Буссаид, Ф. Новое представление скелетных последовательностей для распознавания трехмерных действий. В материалах конференции IEEE/CVF по компьютерному зрению и распознаванию образов, Гонолулу, Гавайи, США, 21–26 июля 2017 г.; стр. 3288–3297.
14. Ли, К.; Чжун, К.; Се, Д.; Пу, С. Обучение признакам совместного возникновения на основе скелетных данных для распознавания и обнаружения действий с помощью иерархической агрегации. arXiv 2018, arXiv:1804.06055v1.
15. Лю, Дж.; Ганг, В.; Пинг, Х.; Дуань, Ливия; Кот, АС Глобальные контекстно-зависимые сети LSTM для распознавания трехмерных действий. В материалах конференции IEEE/CVF по компьютерному зрению и распознаванию образов, Гонолулу, Гавайи, США, 21–26 июля 2017 г.; стр. 3671–3680.
16. Ши, Л.; Чжан, Ю.; Ченг, Дж.; Лу, Х. Распознавание действий на основе скелета с помощью нейронных сетей с направленным графом. В материалах конференции IEEE/CVF по компьютерному зрению и распознаванию образов, Лонг-Бич, Калифорния, США, 15–20 июня 2019 г.; стр. 7912–7921.

17. Си, К.; Чен, В.; Ван, В.; Ван, Л.; Тан, Т. Сверточная сеть LSTM на графах с улучшенным вниманием для распознавания действий на основе скелетов . В материалах конференции IEEE/CVF по компьютерному зрению и распознаванию образов, Лонг-Бич, Калифорния, США, 15–20 июня 2019 г.; стр. 1227–1236.

18. Ю, З.; Вэньбинь, К.; Годун, Г. Оценка пространственно-временных особенностей точек интереса для распознавания действий на основе глубины. Изображение Виз. Вычислить. 2014, 32, 453–464.
19. Орейфей, О.; Лю, З. Hon4d: Гистограмма ориентированных 4d нормалей для распознавания активности по глубинным последовательностям. В материалах конференции IEEE/CVF по компьютерному зрению и распознаванию образов, Портленд, Орегон, США, 23–28 июня 2013 г.; стр. 716–723.
20. Каэтано, К.; Бремон, Ф.; Шварц, В.Р. Представление изображения скелета для распознавания трехмерных действий на основе древовидной структуры и эталонных соединений. В материалах тридцать второй конференции SIBGRAPI по графике, узорам и изображениям, Рио-де-Жанейро, Бразилия, 28–31 октября 2019 г.; стр. 16–23.

21. Лю, Дж.; Шахруди, А.; Сюй, Д.; Ван, Г. Пространственно-временной метод с воротами доверия для трехмерного распознавания действий человека. В материалах Европейской конференции по компьютерному зрению, Амстердам, Нидерланды, 11–14 октября 2022 г.; Springer: Берлин/Гейдельберг, Германия, 2016 г.; стр. 816–833.

22. Чжан П.; Лан, К.; Син, Дж.; Цзэн, В.; Сюэ, Дж.; Чжэн, Н. Взгляд на адаптивные нейронные сети для высокопроизводительных скелетных сетей. признание действий человека. IEEE Транс. Паттерн Анал. Мах. Интел. 2019, 41, 1963–1978.
23. Ян, С.; Сюн, Ю.; Лин, Д. Сверточные сети пространственно-временных графов для распознавания действий на основе скелетов. В материалах тридцать второй конференции AAAI по искусственному интеллекту, Новый Орлеан, Луизиана, США, 2–7 февраля 2018 г.; стр. 7444–7452.
24. Плиццари, К.; Канничи, М.; Маттеуччи, М. Распознавание действий на основе скелетов с помощью сетей пространственных и временных преобразователей. Вычислить. Вис. Изображение Понимание. 2021, 208-209, 103219.
25. Шоттон Дж.; Фитцгиббон, А.; Кук, М.; Шарп, Т.; Финоккио, М.; Мур, Р.; Кипман, А.; Блейк, А. Распознавание позы человека в реальном времени по частям изображений с одной глубиной. В материалах конференции IEEE/CVF по компьютерному зрению и распознаванию образов, Колорадо-Спрингс, Колорадо, США, 20–25 июня 2011 г.; IEEE: Пискатауэй, Нью-Джерси, США, 2011 г.; стр. 1297–1304.
26. Сюн, Ф.; Чжан, Б.; Сяо, Ю.; Цао, З.; Ю, Т.; Чжоу, Джей Ти; Юань, Дж. A2j: Сеть регрессии от привязки к суставу для оценки трехмерной шарнирной позы по одному изображению глубины. В материалах Международной конференции IEEE по компьютерному зрению, Сеул, Республика Корея, 27 октября – 2 ноября 2019 г.; стр. 793–802.

27. Камель А.; Шэн, Б.; Ян, П.; Ли, П.; Шен, Р.; Фэн, Д.Д. Глубокие сверточные нейронные сети для распознавания действий человека Использование карт глубины и поз. IEEE Транс. Сист. Человек Киберн. Сист. 2019, 49, 1806–1819.
28. Санчес-Кабальеро, А.; де Лопес-Дис, С.; Фуэнтес-Хименес, Д.; Лосада-Гутьеррес, К.; Маррон-Ромера, М.; Касильяс-Перес, Д.; Саркер, Мичиган 3DFCNN: Распознавание действий в реальном времени с использованием глубоких трехмерных нейронных сетей с необработанной информацией о глубине. Мультимед. Инструменты Прил. 2022, 81, 24119–24143.
29. Санчес-Кабальеро, А.; Фуэнтес-Хименес, Д.; Лосада-Гутьеррес, К. Распознавание действий человека в реальном времени с использованием необработанного видео глубины. на основе рекуррентных нейронных сетей. Мультимед. Инструменты Прил. 2022, 82, 16213–16235.
30. Кумар, Д.А.; Кишор, ПВВ; Мурти, Г.; Чайтанья, ТР; Субхани, С. Просмотр инвариантного распознавания действий человека с использованием карт поверхности через сверточные сети. В материалах Международной конференции по методологиям исследований в области управления знаниями, искусственного интеллекта и телекоммуникационной техники, Ченнаи, Индия, 1–2 ноября 2023 г.; стр. 1–5.
31. Гош, СК; МИСТЕР.; Мохан, БР; Гуддети, Р.М.Р. Многоракурсное распознавание трехмерных действий человека на основе глубокого обучения с использованием данных о скелете и глубине. Мультимед. Инструменты Прил. 2022, 82, 19829–19851.
32. Ли, Р.; Ли, Х.; Фу, CW; Коэн-Ор, Д.; Хенг, Пенсильвания Пу-ган: состязательная сеть с повышением дискретизации облака точек. В материалах Международной конференции IEEE/CVF по компьютерному зрению, Сеул, Республика Корея, 27 октября – 2 ноября 2019 г.; стр. 7203–7212.
33. Ци, ЧР; Литания, О.; Он, К.; Гибас, Л. Дж. Дип, голосуя за обнаружение трехмерных объектов в облаках точек. В материалах Международной конференции IEEE/CVF по компьютерному зрению, Сеул, Республика Корея, 27 октября – 2 ноября 2019 г.; стр. 9277–9286.
34. Томас, Х.; Ци, ЧР; Дешо, JE; Маркотеги, Б.; Гулетт, Ф.; Гибас, Л. КПКонв: Гибкая и деформируемая свертка для облаков точек. В материалах Международной конференции IEEE/CVF по компьютерному зрению, Сеул, Республика Корея, 27 октября – 2 ноября 2019 г.; стр. 6410–6419.

35. Он-Бар, Э.; Триведи, М. Сходство углов суставов и HOG2 для распознавания действий. В материалах конференции IEEE/CVF по компьютерному зрению и распознаванию образов, Портленд, Орегон, США, 23–28 июня 2013 г.; стр. 465–470.
36. Ли, Дж.; Вонг, Ю.; Чжао, К.; Канканхалли, М.С. Обучение без учителя представлений действий, инвариантных к представлению. Адв. Нейронная инф. Процесс. Сист. 2018, 31, 1262–1272.
37. Лю, Х.; Ци, ЧР; Гибас, Л. Дж. Flownet3d: Изучение потока сцены в трехмерных облаках точек. В материалах конференции IEEE/CVF по Компьютерное зрение и распознавание образов, Лонг-Бич, Калифорния, США, 15–20 июня 2019 г.; стр. 529–537.
38. Чжай, М.; Сян, Х.; Льв, Н.; Конг, Х. Оптический поток и оценка потока сцены: обзор. Распознавание образов. 2021, 114, 107861.
39. Фернандо, Б.; Гаввес, Э.; Орамас, Дж.; Годрати, А.; Туйтелаарс, Т. Объединение рангов для распознавания действий. IEEE Транс. Паттерн Анал. Мах. Интел. 2016, 39, 773–787.
40. Лю, Дж.; Сюй, Д. GeometryMotion-Net: надежная двухпотоковая базовая линия для распознавания трехмерных действий. IEEE Транс. Сист. цепей. Видео Технол. 2021, 31, 4711–4721.
41. Доу, В.; Чин, WH; Кубота, Н. Растущая сеть памяти со случайным весом 3DCNN для непрерывного распознавания действий человека . В материалах Международной конференции IEEE по нечетким системам, Инчхон, Республика Корея, 13–17 августа 2023 г.; стр. 1–6.

42. Фан, Х.; Ю, Х.; Дин, Ю.; Ян, Ю.; Канканхалли, М. PSTNet: Точечная пространственно-временная свертка в последовательностях облаков точек. В материалах Международной конференции по изучению представлений, Аддис-Абеба, Эфиопия, 26–30 апреля 2020 г.; стр. 1–6.
43. Ци, CR; Су, Х.; Мо, К.; Гибас, Л. Дж. Pointnet: Глубокое изучение наборов точек для 3D-классификации и сегментации. В материалах конференции IEEE по компьютерному зрению и распознаванию образов, Гонолулу, Гавайи, США, 21–26 июля 2017 г.; стр. 652–660.
44. Шахруди А.; Лю, Дж.; Нг, ТТ; Ван, Г. NTU RGB+D: крупномасштабный набор данных для трехмерного анализа человеческой деятельности. В материалах конференции IEEE/CVF по компьютерному зрению и распознаванию образов, Лас-Вегас, Невада, США, 27–30 июня 2016 г.; стр. 1010–1019.
45. Лю, Дж.; Шахруди, А.; Перес, М.; Ван, Г.; Дуань, Линия; Кот, АС Ntu rgb+ d 120: Крупномасштабный тест для трехмерной человеческой деятельности понимание. IEEE Транс. Паттерн Анал. Мах. Интел. 2019, 42, 2684–2701.
46. Лю, Дж.; Ван, Г.; Дуань, Ливия; Абдиева К.; Кот, АС Распознавание действий человека на основе скелетов с глобальным контекстом Внимание, сети LSTM. IEEE Транс. Процесс изображения. 2018, 27, 1586–1599.
47. Ши, Л.; Чжан, Ю.; Ченг, Дж.; Лу, Х. Двухпоточные адаптивные сверточные сети на графах для распознавания действий на основе скелетов. В материалах конференции IEEE/CVF по компьютерному зрению и распознаванию образов, Лонг-Бич, Калифорния, США, 15–20 июня 2019 г.; стр. 12026–12035.
48. Ли, Х.; Хуан, К.; Ван, З. Объединение пространственной и временной информации для распознавания действий человека посредством балансировки границ центра Мультимодальный классификатор. Дж. Вис. Коммун. Изображение представляет. 2023, 90, 103716.
49. Зан, Х.; Чжао, Г. Исследование распознавания человеческих действий на основе сетей Fusion TS-CNN и LSTM. Араб. Дж. Наук. англ. 2023, 48, 2331–2345.
50. Ян, Х.; Тиан, Ю. Супернормальный вектор для распознавания активности с использованием последовательностей глубины. В материалах конференции IEEE/CVF по компьютерному зрению и распознаванию образов, Колумбус, Огайо, США, 23–28 июня 2014 г.; стр. 804–811.
51. Басак, Х.; Кунду, Р.; Сингх, ПК; Иджаз, МФ; Возняк, М.; Саркар, Р. Объединение глубокого обучения и роевой оптимизации для трехмерного распознавания действий человека. наук. Отчет 2022 г., 12, 1–17.
52. Ци, Ю.; Ху, Дж.; Чжуан, Л.; Пей, Х. Многомасштабное распознавание действий человеческого скелета на основе семантики. Прил. Интел. Межд. Дж. Артиф. Интел. Нейронная сеть. Комплексная пробл. -Решение Технол. 2023, 53, 9763–9778.
53. Джи, Х.; Чжао, К.; Ченг, Дж.; Ма, К. Использование пространственно-временного представления для трехмерного распознавания действий человека по карте глубины последовательности. Знать. -Базовая система. 2021, 227, 107040.
54. Го, Дж.; Лю, Дж.; Сюй, Д. 3D-обрезка: платформа сжатия моделей для эффективного распознавания 3D-действий. IEEE Транс. Сист. цепей . Видео Технол. 2022, 32, 8717–8729.
55. Ли, Х.; Хуан, К.; Чжан, Ю.; Ян, Т.; Ван, З. PointMapNet: Сеть карт объектов облаков точек для трехмерного распознавания действий человека. Симметрия 2023, 15, 1–17.
56. Лю, М.; Юань, Дж. Признание человеческих действий как эволюция карт оценки позы. В материалах конференции IEEE/CVF по компьютерному зрению и распознаванию образов, Солт-Лейк-Сити, Юта, США, 18–23 июня 2018 г.; стр. 1159–1168.
57. Лю, Дж.; Шахруди, А.; Ван, Г.; Дуань, Линия; Кот, А.С. Онлайн-прогнозирование действий на основе скелетов с использованием сети выбора масштаба. IEEE Транс. Паттерн Анал. Мах. Интел. 2019, 42, 1453–1467.