Оглавление

Введение

Феномен оценки позы человека — это проблема, которая изучалась в течение нескольких лет, в сфере компьютерного зрения. Что же это такое? Чтобы ответить на этот вопрос, необходимо понять концепцию позы. Позу можно определить как расположение суставов человека в определенной позиции. Таким образом, мы можем определить проблему оценки позы человека как локализацию суставов человека или заранее определенных ориентиров на изображениях и видео. Существует несколько типов оценки позы, включая оценки тела, лица и рук (см. рисунок 1).

Рисунок 1 – Эти изображения пример разных типов определения позы человека. Верхняя левая — это пример определения позы тела человека, верхняя правая — это пример позы руки. Нижнее изображение — это пример определение позы лица [1].

Целью данной работы является представить обзор и сравнение методов определения позы человека.

Для достижения поставленной цели необходимо выполнить следующие задачи:

- изучить методы по определению позы человека;
- выбрать критерии классификации и сравнить эти методы;
- определить области возможного применения методов определения поз человека.

1 Аналитический раздел

В данном разделе будут представлены обзор, существующих методов определения позы человека, а также проведен сравнительный анализ этих методов.

1.1 Обзор предметной области определения позы человека

Для определения позы человека существует множество методов. Одними из первых использовались данные методы:

1) Модель пиктографических структур [2]:

   Эта структура моделирует пространственные взаимосвязи частей твердого тела, выражая их в виде древовидной графической модели, чтобы предсказать местоположение суставов тела. Эти пространственные связи представлены в виде пружин, и части представляют собой шаблоны внешнего вида, основанные на изображении. Путем параметризации частей с помощью расположения и ориентации пикселей, полученная структура может моделировать артикуляции.

   На рисунке 1.1 продемонстрировано наглядное представление этой модели.

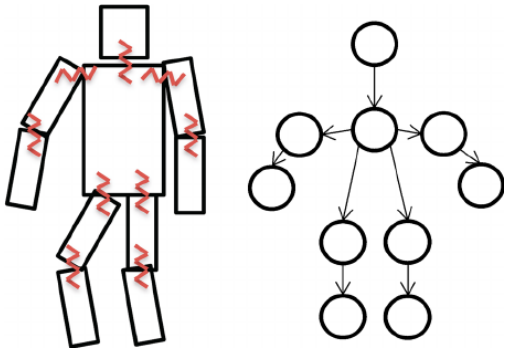

Рисунок 1.1 – Пример пиктографической структуры [3].

   Проблема этого подхода заключается в том, что он не может уловить корреляции между невидимыми и деформируемыми частями тела,

что означает, что модель подвержена ошибкам, если не все конечности человека видны. Она также не зависит от данных изображения.

2) Гибкое смещение частей [4]:

- Этот подход использует деформируемые модели частей, которые представляют собой коллекцию шаблонов, которые подбираются по изображению и располагаются в деформируемой конфигурации. Кроме того, каждая модель имеет глобальные шаблоны и шаблоны деталей. Основная идея заключается в том, чтобы использовать смесь мелких неориентированных деталей в отличие от использования семейства деформированных, то есть повернутых и ракурсных шаблонов. Причина такого подхода заключается в том, чтобы уловить различия в том, как выглядят конечности в деформированном виде и в спокойствие.
- Гибкое смещение частей одновременно фиксирует пространственные отношения между расположением деталей и отношения совпадения между смесями деталей, что приводит к моделям пиктографического структуры, которые кодируют исключительно пространственные отношения. Благодаря динамическому программированию, модели разделяют вычисления между аналогичными искривлениями, что делает этот подход не только значительно быстрым, но и высокоэффективным. Кроме того они моделируют экспоненциально большой набор глобальных смесей через композицию смесей локальных частей для того, чтобы изучить понятия локальной жесткости, а также уловить влияние глобальной геометрии на локальный внешний вид, то есть внешний вид деталей различается в разных местах. На рисунке 1.2 демонстрируется визуальное представление этой модели.
- Гибкое смещение частей способна хорошо выражать сложные отношения между суставами, поэтому она также может моделировать артикуляцию. Однако у нее есть свои проблемы, которые включат ограниченную выразительность и отсутствие учета глобального контекста.

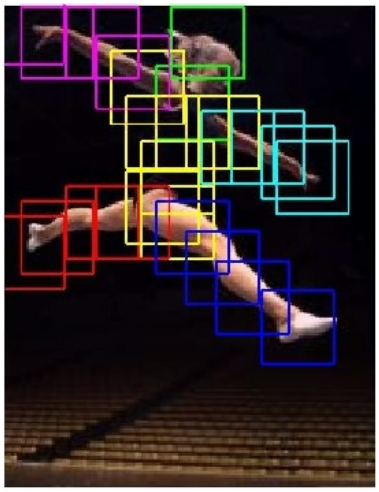

Рисунок 1.2 – Пример гибкого смещения частей [5].

3) Края, цветочные гистограммы, контуры и гистограмма ориентированных градиентов были альтернативными характеристиками, которые применялись в ранних работах определения позы человека и служили основными строительными блоками различных классических моделей для определения точного местоположения частей тела [2].

К общим проблемам классических подходов относятся плохое обобщение и неточное обнаружение частей тела. Поэтому для решения этих проблем было применено глубокое обучение.

1.2 Архитектуры сетей

1.2.1 ResNet

ResNet [6] — это сокращенное название для Residual Network («остаточная сеть»).

Глубокие сети извлекают признаки сквозным многослойным способом. Когда глубокая сеть начинает свертку, возникает проблема: с увеличением глубины сети точность сначала увеличивается, а затем быстро ухудшается.

Чтобы преодолеть эту проблему, Microsoft ввела глубокую «остаточную» структуру обучения. В данном методе используется соединение быстрого доступа — пропускается один или несколько слоев и выполняют составление идентификаторов.

На рисунке 1.3 представлена схема перехода между слоями свертки.

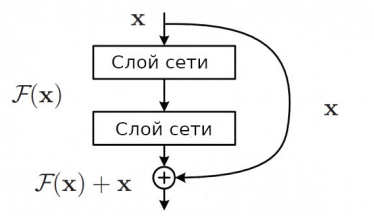

Рисунок 1.3 – Пример остаточного перехода

На рисунке 1.4 демонстрируется основа для архитектуры ResNet.

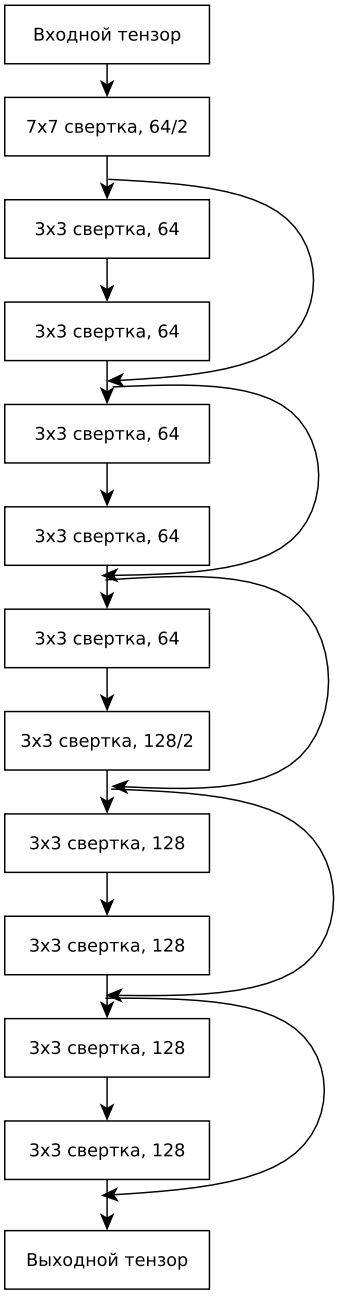

Рисунок 1.4 – Пример архитектуры ResNet.

1.2.2 HRNet

HRNet [7] — это нейронная сеть для распознавания позы человека. Модель решает проблему вариации масштаба людей на изображении.

Методы оценки позы человека имеют сложность с предсказанием позы для людей с низким ростом из-за вариации масштабов на изображении. В данной сети проблема решается за счет параллельного контроля разрешения.

На рисунке 1.5 демонстрируется работы архитектуры HRnet.

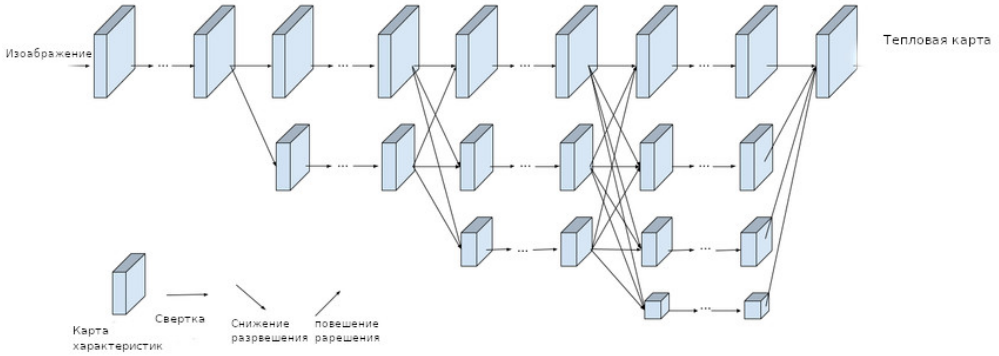

Рисунок 1.5 – Пример архитектуры HRnet [7].

1.2.3 MobileNet

MobileNet [8] — это нейронная сеть, которая создана для облегчения архитектуры сетей со сверточными слоями.

Особенность MobileNet является отсутствие слоя сжатия с выделением большего на изображении, вместо этого используется свертка 2 на 2.

На рисунке 1.6 основной слой MobileNet, который показывает особенность работы сети.

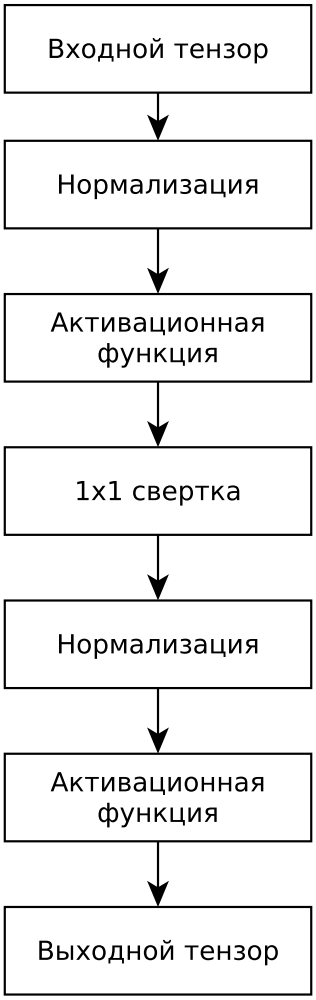

Рисунок 1.6 – Пример слоя MobileNet.

1.3 Методы глубокого обучения для определения позы человека

1.3.1 Сверточная машина поз

Машина позирования состоит из последовательных мультиклассовых слоев, которые обучены предсказывать местоположение каждой детали на каждом уровне иерархии. Она также имеет модуль вычисления характеристик изображения и модуль предсказания, оба из которых могут быть заменены сверточной архитектурой [9], что позволяет позволяет изучать как изображения, так и контекстуальные представления признаков из данных. Именно эта идея привела к созданию Свертоной машине поз (СМП),

которая является первой моделью оценки позы человека на основе глубокого обучения [10].

СМП полностью дифференцируема, что позволяет обучать ее многоступенчатую архитектуру по принципу обратного распространения ошибки, алгоритму, используемому для обучения нейронных сетей с прямой передачей ошибки.

Проблема исчезающих градиентов, когда при обратном распространении ошибки, градиенты уменьшаются по мере прохождения через многие слои решается с помощью промежуточного контроля после каждого этапа [10].

На первом этапе СМП предсказывает предположения о деталях, используя только локальные данные изображения, с помощью глубокой сверточной сети, состоящей из 7 общих сверточных слоев. Карты доверия, созданные на этом этапе, добавляются к вводимым данным перед обработкой несколькими сверточными слоями.

На более поздних этапах эффективное восприимчивое поле увеличивается для повышения точности. В целом, этот подход позволяет архитектуре изучать как особенности изображения, так и пространственные модели, зависящие от изображения, для задач прогнозирования без необходимости использовать графический стиль моделирования выводы.

1.3.2 Integral Pose

Integral Pose [11] решает проблемы недифференцируемой постобработки и ошибки квантования. Данный метод объединяет в себе представление тепловой карты и совместный подход с интегральной регрессией, таким, образом, разделяет достоинства обоих подходов, что позволяет избежать вышеуказанных проблем. Высокая производительность достигается при использовании архитектуры ResNet-101.

Уравнение тепловой карты H_k для k^{th} сустава, каждое место на карте представляет вероятность того, что место является суставом. Окончательная координата J_k получается как местоположение p с максимальным

правдоподобием как:

$$J_k = \arg\max_p H_k(p) \tag{1.1}$$

Данное уравнение является недифференцируемым и приводит к ошибке квантования.

В Integral Pose используется уравнение:

$$J_k = \int_{p\in\Omega} p \cdot \dot{H_k}(p), \tag{1.2}$$

H_k‘ — это нормализованная тепловая карта, а Ω — ее область. Данное уравнение решает проблемы недифференцируемости и ошбики квантования.

1.3.3 SimpleBaseline

Особенностью метода SimpleBaseline [12] является использование оптического потока, который является некоторым расширение кадра для оценки позы, и сходства опредленный позы на основе оптического потока предыдущего кадра. В этом методе используется архитектура ResNet-152.

В обычных методах сначала оцениваются позы нескольких людей на кадрах, затем отслеживаются позы с присвоенным уникальным идентификатором. Данный способ может приводит к ошибкам идентификации. SimpleBaseline решает эту проблему отслеживания позы нескольких человек, сначала оценивая позу человека по кадрам с помощью Mask RCNN [13], а затем выполняя отслеживание с помощью двухстороннего алгоритма сопоставления кадр за кадром.

1.3.4 HRNet-W32

HRNet-W32 [14] — это метод, который назван в честь архитектуры HigherHRNet, особенностью которой является использование пирамид масштаба, то есть отделение признаков высокого разрешения. Поза может содержать до 17 ориентиров.

Большинство существующих методов восстанавливают представления высокого разрешения из представлений низкого разрешения. Вместо этого, HRNet-W32 поддерживает представление высокого разрешения на протяжении всего процесса.

1.3.5 DARK

DARK [15]. Основой этого метода является представление координат для оценки позы человека. Обычно в методах применяется тепловая карта на входе и выходах из нейронных сетей для определения ориентиров, но в DARK тепловая карта используются во время всего обучения. В этом методе используется архитектура HRNet.

В Dark рассматривается проблема представления координат, включающие кодирование и декодирование, при оценки позы человека. Целью является предсказание координат суставов в заданном входном изображении. Для этого необходим выучить регрессионную модель от входного изображения до выходных координат. Для облегчения обучения модели, в представленном методе кодируются помеченные истинные координаты сустава в тепловую карту в качестве цели обучения. После этого происходит процесс декодирования.

1.3.6 UDP

UDP [16] — несмещенное преобразование систем координат (Unbised Coordinate System Transformation). Особенностью данного подхода состоит из несмещенного преобразования координат и ориентиров, которое повышает производительность, при том качество обучения не снижается.

Матрицы изображений и координаты ориентиров являются основными данными, входящее в проблему оценки позы человека. Изображения хранятся и обрабатываются в дискретном формате, но координаты ключевых точек определяются, обрабатываются и оцениваются в непрерывном пространстве. Чтобы избежать ухудшения точности в системе координат

преобразования, необходима единая парадигма для единообразного анализа.

Для этого в UDP предполагается, что существует непрерывная плоскость изображения и рассматривается каждая матрица изображения как дискретный результат, где каждый пиксель в матрице является определенной точкой дискретизации.

1.3.7 MobileNetV2

MobileNetV2 [17] — эта модель предназначена для повышения производительности определения позы человека на основе архитектуры MobileNet. Данный метод делает выводы с минимальными затратами памяти и использует стандартные операции, присутствующие во всех нейронных системах.

Как и в MobileNet, здесь есть сверточные блоки с шагом 1 и с шагом 2. Блоки с шагом 2 предназначены для снижения пространственной размерности тензора и, в отличие от блока с шагом 1, не имеют слоя остаточного соединения.

1.3.8 Lite-HRNet

Lite-HRNet [18] — это модель является оптимизацией архитектуры HRNet, для повышения производительности. Основной подход для ускорения сети является использования свертки 1 на 1, которая уменьшает размеры тензоров во время вычисления.

Сверточное ядро 1 на 1 является ключевым слоем для оптимизации нейронных сетей. Оно играет важную роль в обмене информацией между каналами, поскольку операция перестановки и свертка в глубину не влияют на обмен информацией между каналами. Она имеет квадратичную сложность по отношению к количеству каналов.

1.4 Точность и метрики

1.4.1 Определение точности и понятие метрики

Определение точности — это оценка машинного обучения путем вычисления показателей их алгоритмов [19]. Существует множество оценочных метрик. используемых для проведения таких вычислений. Причина этого заключается в том, что существует множество характеристик и требований, которые необходимо учитывать при оценке показателей модели оценки позы человека. Таким образом, другими словами, точность модели определяется с помощью метрик, то есть метрики — это способ количественной оценки точности модели.

1.4.2 Различные метрики, используемые в определении позы человека

Как было сказано ранее, существует несколько метрик, используемых для оценки эффективности моделей определения поз человека.

Ниже перечислены некоторые из них:

1) Пересечение над объединением (ПНО) [20]: это метрика, которая находит разницу между истинными и предсказанными ограничительными рамками. Удаляет все ненужные на основе установленного порогового значения, которое обычно составляет 0,5. Вычисляется по формуле:

$$\text{ПНО} = \frac{\text{Площадь пересечения двух прямоугольников}}{\text{Площадь объединения двух прямоугольников}}; \qquad (1.3)$$

2) Процент правильных частей (ППЧ) и Процент обнаруженных соединений (ПОС) [21]: это метрика, которая сейчас не так часто используется, но ее цель заключалась в том, чтобы сообщить о точности локализации конечностей. ППЧ определяется, когда расстояние между предсказанными и истинными суставами меньше, чем доля длины

конечности, которая составляет от 0,1 и 0,5. Если порог равен 0,5, то показатель ППЧ называется ППЧ@0,5. Более высокий показатель ППЧ означает лучшую производительность. Ограничение этой метрики, в тоже время заключается в том, что она является неточной для конечностей с небольшой длинной. В связи с этим был внедрен ПОС, который следует той же логике, что и ППЧ; если расстояние между предсказанным и истинным суставами находится в пределах опредленной доли диаметра туловища, сустав считается правильно обнаруженным. Использование этой метрики подразумевает, что точность определения всех суставов основывается на этом пороге. Вычисляется по формуле:

$$\text{ППЧ} = \frac{||s_n - s'_n|| + ||e_n - e'_n||}{2} \leq \alpha||s_n - e_n||, \tag{1.4}$$

где s_n и e_n — это истинные начало и конец суставов, а s'_n и e'_n — это предсказанное начало и конец суставов.

3) Процент правильных ключевых точек (ППКТ) [1]: эта метрика используется для измерения точности локализации различных ключевых точек в пределах опредленного порога. Он установлен на 50% от длины сегмента головы каждого тестового изображения. Связано с ПОС, когда расстояние между обнаруженными и истинными суставами меньше, чем 0,2 диаметра туловища, это называется ППКТ@0,2. Чем выше значение ППКТ, тем лучше показатели. Вычисляется по формуле:

$$\text{ППКТ} = \frac{\text{Предсказанное положение суставов}}{\text{Истинное положение суставов}}; \tag{1.5}$$

4) Средняя точность (СТ) [1]: СТ измеряет точность обнаружения ключевых точек в соответствии с точностью, которая представляет собой отношение истинно положительных результатов к общему количеству положительных результатов. Другими словами, насколько точным являются предсказания. Таким образом, метрика СТ представляет собой среднее значение точности по всем значениям отзыва от 0 до 1 при различных пороговых значений ПНО. Вычисляется по формуле:

$$\text{СТ} = \frac{\text{Истинное количество результатов}}{\text{Общее количество положительных результатов}}; \tag{1.6}$$

5) Сходство ключевых точек объектов (СКТО) [22]. Эта метрика представляет собой среднее сходство ключевых точек по всем ключевым точкам объекта. Она рассчитывается на основе масштаба объекта и расстояния между предсказанными и истинными точками. Масштаб и константа ключевых точек требуются, чтобы придать равную значимость каждой ключевой точке. Каждой ключевой точке присваивается значение сходства от 0 до 1, а СКТО — это среднее значение всех этих значений по всем ключевым точкам. Эта метрика помогает в определении СТ. Вычисляется по формуле:

$$\text{СКТО} = \frac{\sum_i (\frac{\exp(-d_i^2)}{2s^2k_i^2})\delta(v_i > 0))}{\sum_i (\delta(v_i > 0))}, \tag{1.7}$$

где d_i — это евклидово расстояние между каждой соответствующей опорной точкой, v_i — флаги истинной видимости опорной точки, s – масштаб объекта, k_i — константа для каждой ключевой точки в диапазоне от 0 до 1.

6) Средняя погрешность взаимного расположения (СПВР) [22]: это наиболее широко используемая метрика для трехмерного определения позы человека. Рассчитывается с помощью евклидового расстояния между оценочным трехмерным суставами и истинным положением следующим образом:

$$\text{СПВР} = \frac{1}{N}\sum_{i=1}^{N} ||J_i - J_i^*||_2 \,, \tag{1.8}$$

где N количество суставов, J_i и J_i^* это истинное и оценочное положение i-го сустава.

1.5 Наборы данных, используемых в определении позы человека

Наборы данных являются важным аспектом в машинном обучении. Для того чтобы модели машинного обучения выполняли задачу, их алгоритмы должны быть сначала обучены, а затем протестированы, чтобы убедится, что они правильно интерпретируют данные для выполнения задачи.

Это делается с помощью наборов данных, которые состоят из обучающих и тестовых данных. Для каждого типа определения позы человека существует свой набор данных.

1.5.1 Наборы данных по определению позы тела

1) COCO [22]: это наиболее широко используемый двумерный набор данных тела, в первую очередь для определения позы нескольких людей. Хотя он используется для обнаружения объектов и содержит изображения, помогающие в этом, он содержит более 330 тыс. изображений и 200 тыс. людей, помеченных ориентиров до 17 по всему телу. Первый набор был выпущен в 2014 году, но с тех пор был изменен. Существет 2 версии наборов данных COCO для определения позы человека: ориентиры COCO 2016 и 2017 года, отличаются в разделении на обучение, проверку и тестировании.

2) MPII [23]: этот набор данных двумерных тел используется в основном для оценки позы одного человека. Он содержит около 25 тысяч изображений, содержащих 40 тыс. человек с 16 вручную отмеченных суставов тела, это отличается от COCO, так как там отмечается 17 суставов тела. Изображения охватывают 410 различных видов человеческой деятельности, таких как танцы, бег, охота, и помечаются соответствующим видами деятельности. Каждое изображение было взято из видеоролика YouTube и были предоставлены предшествующие были предоставлены предшествующие и последующие кадры, которые не были отмечены, что является еще одним отличием от COCO.

Кроме того, были добавлены расширенные описания, такие как перекрытые части тела и трехмерная ориентация туловища и голова. Набор был сделан в 2014 году.

3) AI Challenger Human Keypoint Detection [24]: этот набор данных по двумерным телам является крупнейшим, когда речь идет о двухмерном описании положении человека. Он содержит более 300 тыс. отмеченных изображений высокого разрешения для обнаружения ориентиров (14 ориентиров на человека) и более 600 тыс. тестовых изображений. Все изображения были собраны из поисковых систем, и, как и в наборе данных MPII, сфокусированы на повседневной деятельности людей в различных позах. Разница в системе подбора изображений, и в дополнительных функциях: распознавание пустых снимков на основе атрибутов. Набор был сделан в 2017 году.

4) PoseTrack [25]: это двумерный набор данных по телу на основе видео, который содержит около 1356 видео, 46 тыс. отмеченных видеокадров и 276 тыс. помеченных поз тела. Он в основном используется для оценки позы тела несколько людей, где каждый человек имеет уникальный идентификатор трека с отметок, включающими до 15 ключевых точек тела. PoseTrack отличается от других набор данных тела тем, что в нем используется видеозаписи вместо изображений и другое количество ключевых точек. Набор был сделан в 2017 году.

1.5.2 Наборы данных по определению направления головы

1) 300W [26]: этот двумерный набор лиц является коллекцией нескольких других наборов данных, включая HELEN, AFW, LFPW и IBUG, которые были отмечены 68 ориентирами, что означает, что 300W тоже маркирует свои изображения с помощью 68 ориентиров. В общей сложности в нем общей сложности в нем имеется около 4000 учебных изображений и 600 тестовых изображений: 300 изображений лиц в помещении и 300 изображений лиц на улице. Все они имели различ-

ные условия съемки, такие как освещение, цвет, эмоции, размер лица, угол наклона и количество присутствующих лиц. Набор был сделан в 2013 году.

2) AFLW [27]: этот набор двумерных лиц содержит около 25 тыс. изображений лиц, различающихся по внешнему виду, такому как как поза, пол, возраст, этническая принадлежность и выражение лица. Отличие этого набора в том, что он маркирован только 21 ориентиром и имеет больший угол съемки лица по сравнение 300W. Набор был сделан в 2011 году.

3) COFW [28]: этот набор данных двумерных лиц в основном ориентирован на маркировку изображений лиц, которые частично закрыты другими объектами или самим лицом. Состоит из около 1,3 тыс. учебных и 507 тестовых изображений, которые помечены 29 ориентирами, что он отличается от других наборов данных лиц. Набор сделан в 2013 году.

4) WFLW [29]: этот двумерный набор лиц является одним из самых подробных наборов данных. Он состоит из 7,5 тыс. учебных и 2,5 тыс. тестовых изображения, в частности, они широкий диапазон эмоций, поз, размытости и условий освещения. Кроме того, они маркированы с помощью 98 плотно привязанных ориентиров, что также отличает их от других наборов данных лиц. Набор данных собран в 2018 году.

1.5.3 Наборы данных по определению позы руки

1) BigHand2.2M [30]: этот набор данных трехмерных рук является самым большим набором данных рук на сегодняшний день, поскольку содержит около 2,2 миллиона глубинных изображений отдельных рук с 21 ориентирами, полученных от 10 испытуемых. Варьирование положения испытуемых и ориентации рук помогло создать разнообразный набор точек зрения на руки. Набор данных разбит на 3 части: 1,5 миллиона по 13 кадров схематичными позами, которые охватывают все сочленения которые может свободно принять человеческая

рука; 375 тыс. кадров случайных поз, которые показывают, как испытуемые используют свои руки для ориентации в пространстве; 290 тыс. кадров эгоцентрических поз, в которых испытуемые выполняют 32 экстремальные позы, которые представляют собой позы рук, где каждый палец принимает максимально согнутое или вытянутое положение, в сочетании со случайными движениями. Набор сделан в 2017 году.

2) GANerated Hand Dataset [31]: это набор двумерных и трехмерных данных рук, содержащий 330 тыс. кадров, которые представляют собой синтезированне формы рук и имеют 21 ориентир. В отличие от BigHand2.2M, для захвата поз рук использовались искусственные объекты, удерживаемые руками. Набор создан в 2018 году.

3) NYU Hand [32]: этот набор данных трехмерных рук содержит более 72 тыс. кадров в обучающем наборе от одного испытуемого и более 8 тыс. кадров от двух разных испытуемых в тестовом наборе. По сравнению с другими наборами данных рук этот набор данных выделяется тем, что это снимки сделаны с помощью тепловизора. Набор собран в 2014 году.

4) HandNet [33]: этот набор данных трехмерных изображений рук является одним из самых больших набор данных по глубине. Он содержит 202 тыс. учебных и 10 тыс. тестовых кадров. В наборе присутствует 5 мужчин и 5 женщин для того, чтобы набор содержал разные размеры рук. Изображения были промаркированы 6 ориентирами. Набор собран в 2015 году.

1.6 Сравнение методов определения позы человека

1.6.1 Таблицы сравнения методов определения позы человека

В таблице 1.1 сравниваются методы определения позы человека по метрикам средней точности при разном ПНО на наборе данных COCO.

Таблица 1.1 – Сравнение методов определения позы человека.

Метод	Архитектура	Вход	Параметры	CT	CT^{50}	CT^{75}
G-RMI [34]	ResNet-101	353 × 257	42.6M	64.9	85.5	71.3
Integral Pose [11]	ResNet-101	256 × 256	45.0M	67.8	88.2	74.8
SimpleBaseline [12]	ResNet-152	384 × 288	68.6M	73.7	91.9	81.1
HRNet-W32 [14]	HRNet-W32	384 × 288	28.5M	74.9	92.5	82.8
HRNet-W48 [35]	HRNet-W48	384 × 288	63.6M	75.5	92.5	83.3
DARK [15]	HRNet-W48	384 × 288	63.6M	76.2	92.5	83.6
UDP [16]	HRNet-W48	384 × 288	63.6M	**76.5**	92.7	84.0

На основе данной таблицы можно сделать вывод, что одним из самых точных методов является UDP. Также, видно, что архитектура HRNet лучше справляется с задачей определения позы человека.

В таблице 1.2 сравниваются методы определения позы человека на легковесных архитектурах.

Таблица 1.2 – Сравнение методов определения позы человека.

Метод	Архитектура	Вход	Параметры	CT	CT^{50}	CT^{75}
Small HRNet [14]	HRNet-W16	384 × 288	1.3M	55.2	85.8	61.4
MobileNetV2 [17]	MobileNetV2	384 × 288	9.8M	66.2	90.0	74.0
Lite-HRNet [18]	Lite-HRNet-30	384 × 288	1.8M	**69.7**	90.7	77.5

Как видно из таблицы выше самым точным методом является Lite-HRNet. Наиболее подходящей архитектурой для определения позы человека — HRNet-W16. Проблема легковесных сетей — это сильное падения точности за счет уменьшения количества входных параметров.

1.6.2 Категоризация для определения позы человека

На рисунке 1.7 приведены примеры одного и того же типа оценки позы, различающиеся по количеству ориентиров.

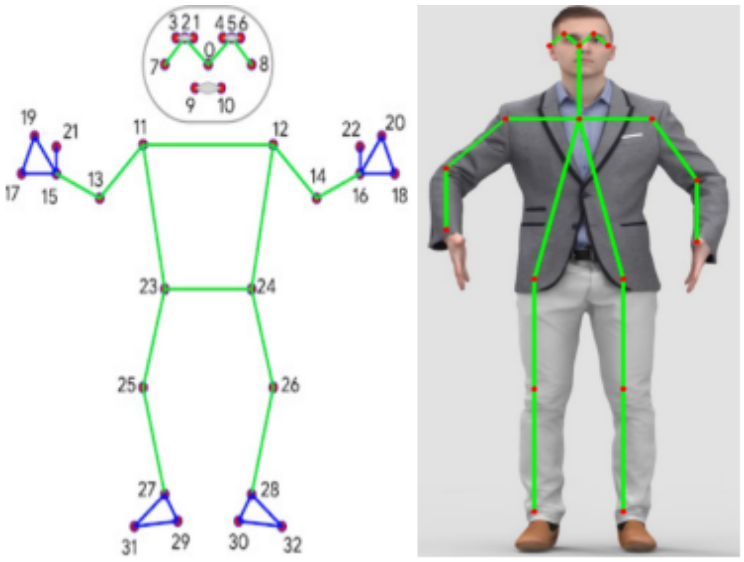

Рисунок 1.7 – Пример оценки позы [36].

Способ классификации различных типов оценки позы основан на разрешении и количестве ориентиров, поскольку это отражает тип выполняемой оценки позы, то есть руки, лицо или тело. Такой подход имеет смысл из-за свойств этих оценок, которые будут подчеркнуты при классификации ниже.

1) Низкое разрешение при наличии до 30 ориентиров. Если необходимо определить не так много ориентиров, то низкого разрешения будет достаточно для выполнения данной задачи. Это относится к оценке позы тела, поскольку в большинстве случаев имеется около 20 ориентиров, в то время как в некоторых случаях их немного больше 30, как показано в таблице 1.1. Кроме того, проблема наложения и сложных поз, которая широко распространена при оценке положения тела, решается с помощью большого «рецепторного поля» и не требует высокого разрешения, таким образом, для идентификации ориентиров достаточно низкого разрешения

2) Высокое разрешение для сложных поз. При наличии более 30 ориентиров, которые являются крупными и сложными, для точной локализации требуется более высокое разрешение для их точной локализации. Оценка позы лица обычно имеет 68 ориентиров 1.1. Аналогично, при оценке позы руки требуется всего 42 ориентира 1.1. Для того чтобы учесть большое количество и размер ориентиров, необходимо высокое разрешение.

1.7 Примеры использования определения позы человека

Определение положения человека можно использовать во множестве областей. В данном разделе будут приведены некоторые из них.

1.7.1 Личные тренеры на основе глубоко обучения

Для поддержания физического здоровья было создано приложение Zenia [37] — это приложение для занятия йогой на основе искусственного интеллекта, использующее определение позы человека. Оно помогает принимать правильные позы при занятиях йогой. Приложение использует камеру для распознавания позы и оценивает ее точность.

Также определение позы человека используется не только в йоге, но и в тяжелой атлетике, для предотвращения травм.

1.7.2 Роботехника

Еще одной из областей применения определения позы человека является роботехника.

Для обучения роботов используются глубокое обучение на основе моделей определения позы человека, так как это упрощает задачу программистам, которые пишут логику ходьбы для роботов.

1.7.3 Дополнительная реальность

За недавний период быстрыми темпами начало развиваться дополнительная реальность и виртуальная реальность. Для данных направлений необходимо отслеживать движений человека. Как раз для отслеживания используются нейронные сети на основе определения позы человека.

1.7.4 Распознавание поз спортсменов

В наши дни во всех видов спорта вводится анализ данных и одним из ключевых направлений является определение позы спортсменов.

Распознавание позы спортсмена необходимо для обучения спортсменов, чтобы совершенствовать их технику и улучшать результаты. Для этой области используются нейронные сети на основе определения позы человека.

Заключение

В ходе выполнения данной работы были выполнены следующие задачи:

- изучены методы по определению позы человека;
- выбраны критерии классификации и сравнены методы;
- определены области возможного применения.

Самым точным методом определения позы человека является UDP, средняя точность составляет 76.5, что на 0.3 лучше остальных методов при одинаковом количестве параметров. Самым точным оптимизированным методом является Lite-HRNet, средняя точность составляет 69.7, что на 3 лучше остальных оптимизированных методов.

Поставленная цель достигнута: методы определения позы человека были рассмотрены и определены.

Список использованных источников

[1] A 2019 guide to Human Pose Estimation with Depp Learning [Электронный ресурс]. Режим доступа: https://nanonets.com/blog/human-pose-estimation-2d-guide/ (дата обращения: 05.11.2022).

[2] Human Pose Estimation with Deep Learning - Part 1 [Электронный ресурс]. Режим доступа: https://nanonets.com/blog/human-pose-estimation-2d-guide/https://neuralet.com/article/human-pose-estimation-with-deep-learning-part-i/?id=3 (дата обращения: 05.11.2022).

[3] Pictorial structure with springs and tree based model [Электронный ресурс]. Режим доступа: https://www.researchgate.net/figure/Pictorial-structure-with-springs-and-tree-based-model_fig1_257627726 (дата обращения: 20.11.2022).

[4] Articulated pose estimation with flexible mixtures-of-parts [Электронный ресурс]. Режим доступа: https://www.researchgate.net/publication/224254989_Articulated_pose_estimation_with_flexible_mixtures-of-parts (дата обращения: 05.11.2022).

[5] Articaulated Human Detection with Flexible Mixtures of Parts [Электронный ресурс]. Режим доступа: https://escholarship.org/content/qt7sk1s10g/qt7sk1s10g_noSplash_a8d7d492292a22ca3c20c0c99cbd9d1f.pdf?t=oub9r6 (дата обращения: 20.11.2022).

[6] ResNet (34, 50, 101): «остаточные» CNN для классификации изображений [Электронный ресурс]. Режим доступа: https://neurohive.io/ru/vidy-nejrosetej/resnet-34-50-101/ (дата обращения: 20.11.2022).

[7] HRNet explained: Human Pose Estimation, Semantic Segmentation and Object Detection [Электронный ресурс]. Режим доступа: https://towardsdatascience.com/hrnet-explained-human-pose-estimation-sematic-segmentation-and-obj (дата обращения: 20.11.2022).

[8] MobileNet [Электронный ресурс]. Режим доступа: https://keras.io/api/applications/mobilenet/ (дата обращения: 20.11.2022).

[9] О.С. Сикороский. Обзор сверточных нейронных сетей для задачи классификации изображений // Новые информационные технологии в автоматизированных системах. 2017. С. 421–432.

[10] Convolutional pose machines / Shih-En Wei, Varun Ramakrishna, Takeo Kanade [и др.] // CVPR. 2016.

[11] Integral human pose regression / Xiao Sun, Bin Xiao, Shuang Liang [и др.] // arXiv preprint arXiv:1711.08229. 2017.

[12] Xiao Bin, Wu Haiping, Wei Yichen. Simple Baselines for Human Pose Estimation and Tracking // European Conference on Computer Vision (ECCV). 2018.

[13] Abdulla Waleed. Mask R-CNN for object detection and instance segmentation on Keras and TensorFlow. https://github.com/matterport/Mask_RCNN. 2017.

[14] Deep High-Resolution Representation Learning for Human Pose Estimation / Ke Sun, Bin Xiao, Dong Liu [и др.] // CVPR. 2019.

[15] Distribution-Aware Coordinate Representation for Human Pose Estimation / Feng Zhang, Xiatian Zhu, Hanbin Dai [и др.] // IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2020. Июнь.

[16] AID: Pushing the Performance Boundary of Human Pose Estimation with Information Dropping Augmentation, author=Huang, Junjie and Zhu, Zheng and Huang, Guan and Du, Dalong, journal=arXiv preprint arXiv:2008.07139, year=2020.

[17] MobileNetV2 [Электронный ресурс]. Режим доступа: https://github.com/PINTO0309/MobileNetV2-PoseEstimation (дата обращения: 20.11.2022).

[18] Deep High-Resolution Representation Learning for Visual Recognition / Jingdong Wang, Ke Sun, Tianheng Cheng [и др.] // TPAMI.

[19] В.В.Вьюгин. Математические основы теории машинного обучения и прогнозирования. – М.:, 2013. – 387 с.

[20] Object Detection Evaluation Metrics [Электронный ресурс]. Режим доступа: https://analyticsindiamag.com/5-object-detection_evaluation-metrics-that-data-scientists-should-know/ (дата обращения: 05.11.2022).

[21] Pose Estimation.Metrics. [Электронный ресурс]. Режим доступа: https://stasiuk.medium.com/pose-estimation-metrics-844c07ba0a78 (дата обращения: 05.11.2022).

[22] Keypoint Evaluation [Электронный ресурс]. Режим доступа: https://cocodataset.org/#keypoints-eval (дата обращения: 05.11.2022).

[23] MPII [Электронный ресурс]. Режим доступа: https://paperswithcode.com/dataset/mpii (дата обращения: 06.11.2022).

[24] AI challenger 2017 [Электронный ресурс]. Режим доступа: https://github.com/AIChallenger/AI_Challenger_2017 (дата обращения: 06.11.2022).

[25] PoseTrack [Электронный ресурс]. Режим доступа: https://paperswithcode.com/dataset/posetrack (дата обращения: 06.11.2022).

[26] 300 Faces In-the-Wild Challenge (300-W) [Электронный ресурс]. Режим доступа: https://ibug.doc.ic.ac.uk/resources/300-W/ (дата обращения: 06.11.2022).

[27] AFLW (Annotated Facial Landmarks in the Wild) [Электронный ресурс]. Режим доступа: https://paperswithcode.com/dataset/aflw (дата обращения: 06.11.2022).

[28] COFW (Caltech Occluded Faces in the Wild) [Электронный ресурс]. Режим доступа: https://paperswithcode.com/dataset/cofw (дата обращения: 06.11.2022).

[29] WFLW (Wider Facial Landmarks in the Wild) [Электронный ресурс]. Режим доступа: https://paperswithcode.com/dataset/wflw (дата обращения: 06.11.2022).

[30] BigHand2.2M [Электронный ресурс]. Режим доступа: https://paperswithcode.com/dataset/bighand2-2m-benchmark (дата обращения: 06.11.2022).

[31] GANerated Hands for Real-Time 3D Hand Tracking from Monocular RGB / Franziska Mueller, Florian Bernard, Oleksandr Sotnychenko [и др.] // Proceedings of Computer Vision and Pattern Recognition (CVPR). 2018. June. 11 с. URL: https://handtracker.mpi-inf.mpg.de/projects/GANeratedHands/.

[32] Jonathan Tompson, Murphy Stein, Yann Lecun [и др.].

[33] Wetzler Aaron, Slossberg Ron, Kimmel Ron. Rule Of Thumb: Deep derotation for improved fingertip detection // Proceedings of the British Machine Vision Conference (BMVC) / под ред. Mark W. Jones Xianghua Xie, Gary K. L. Tam. BMVA Press, 2015. September. С. 33.1–33.12.

[34] Fathi Alireza, Korattikara Anoop, Sun Chen [и др.]. G-RMI Object Detection. 2016.

[35] Deep High-Resolution Representation Learning for Human Pose Estimation / Ke Sun, Bin Xiao, Dong Liu [и др.] // CVPR. 2019.

[36] On-device, Real-time Body Pose Tracking with MediaPipe BlazePose [Электронный ресурс]. Режим доступа: https://ai.googleblog.com/2020/08/on-device-real-time-body-pose-tracking.html (дата обращения: 20.11.2022).

[37] Zenia [Электронный ресурс]. Режим доступа: https://www.zenia.tech/en (дата обращения: 06.11.2022).