

Оглавление

1	Введение	4
2	Аналитический раздел	6
2.1	Обзор предметной области	6
2.1.1	Классические подходы к определению позы человека	6
2.2	Глубокое обучение в сверточных нейронных сетях в машине позирования	8
2.3	Точность и метрики	9
2.3.1	Определение точности и понятие метрики	9
2.3.2	Различные метрики, используемые в определении по- зы человека	10
2.4	Стандарты и классификация для оценки позы человека . . .	12
2.4.1	Таблица стандартов	12
2.4.2	Категоризация для определения позы человека	13
	Список использованных источников	14

1 Введение

Феномен оценки позы человека — это проблема, которая изучалась в течение 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 специфическим образом. 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 руки (см. рисунок 1.1), а также множество аспектов.



Рисунок 1.1 – 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.

Целью данной работы является изучение методов определения позы человека [1]. 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 позы человека.

2 Аналитический раздел

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

2.1 Обзор предметной области

2.1.1 Классические подходы к определению позы человека

Глубокое обучение — тип машинного обучения [2], в котором используются искусственные нейронные сети с несколькими слоями для обработки сложных данных, которые активно используются в настоящее время в определении позы человека. Однако до его внедрения использовались другие методы, а именно:

- 1) Модель пиктографических структур [3]:

Эта структура моделирует пространственные взаимосвязи частей твердого тела, выражая их в виде древовидной графической модели, чтобы предсказать местоположение суставов тела. Эти пространственные связи показаны с помощью 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 2.1 продемонстрировано наглядное представление этой модели.

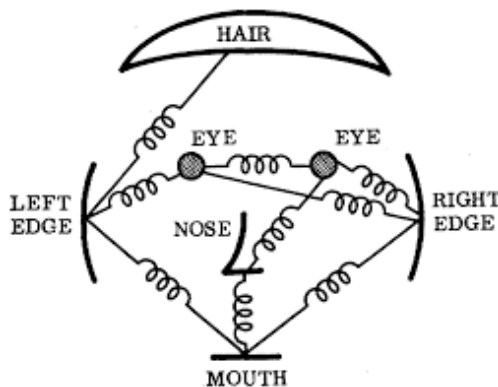


Рисунок 2.1 – Пример пиктографической структуры.

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 этого заключается в том, что с различиями в том, как выглядят конечности, и изменениями в точке зрения.
- 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 2.2 демонстрируется визуальное представление этой модели.

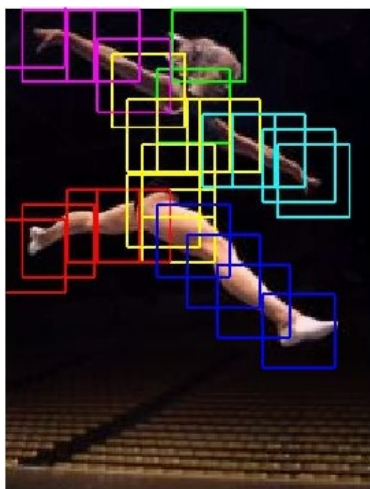


Рисунок 2.2 – Пример гибкого смещения частей.

- Гибкое смещение частей способна хорошо выражать сложные отношения между суставами, поэтому она также может моделировать артикуляцию. Однако у нее есть свои проблемы, которые включают ограниченную выразительность и отсутствие учета глобального контекста.

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 [3].

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

2.2 Глубокое обучение в сверточных нейронных сетях в машине позирования

Машина позирования состоит из последовательных мультиклассовых предсказаний, которых 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 [5], что позволяет изучать как изображения, так и контекстуальные представления признаков из данных. Именно эта идея привела к созданию Convolutional Pose Machine (CPM), которая является первой моделью оценки позы человека на основе глубокого обучения [6].

CPM полностью дифференцируема, что позволяет обучать ее многоступенчатую архитектуру по принципу обратного распространения, алгоритму, используемому для обучения нейронных сетей с прямой передачей. Кроме того, ее последовательная структура предсказания, состоящая из сверточных сетей и обучается неявным пространственным моделям, использует большие рецепторные поля на картах убеждений, полученных на предыдущих этапах, что помогает в изучении пространственных связей между деталями на большом расстоянии и приводит к повышению точности за счет все более точных оценок местоположения деталей на последующих шагах.

Проблема исчезающих градиентов, когда при обратном распространении ошибки, градиенты уменьшаются по мере прохождения через многие слои решается с помощью промежуточного контроля после каждого этапа [6].

На первом этапе СРМ предсказывает предположения о деталях, используя только локальные данные изображения, с помощью глубокой сверточной сети, состоящей из 7 общих сверточных слоев. Карты убеждений, 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.

На рисунке 2.3 показан пример работы СРМ.

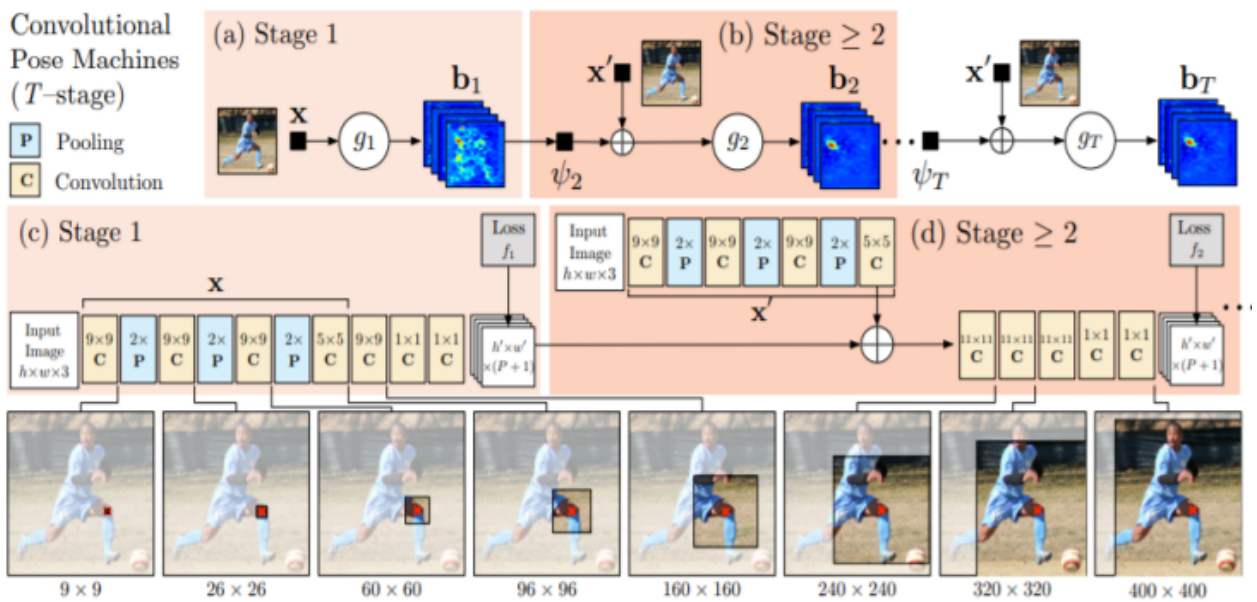


Рисунок 2.3 – Пример работы СРМ.

2.3 Точность и метрики

2.3.1 Определение точности и понятие метрики

Определение точности — это оценка машинного обучения путем вычисления показателей их алгоритмов [2]. 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.

2.3.2 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) [7]: это метрика, которая находит разницу между истинными и предсказанными ограничительными рамками. Удаляет все ненужные на основе установленного порогового значения, которое обычно составляет 0,5.
- 2) Процент правильных частей (ППЧ) и Процент обнаруженных соединений (ПОС) [8]: это метрика, которая сейчас не так часто используется, но ее цель заключалась в том, чтобы сообщить о точности локализации конечностей. Это 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 длиной. 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 определенной доли диаметра туловища, сустав считается правильно обнаруженным. Использование этой метрики подразумевает, что точность определения всех суставов основывается на этом пороге.
- 3) Процент правильных ключевых точек (ППКТ) [1]: эта метрика используется для измерения точности локализации различных ключевых точек в пределах определенного 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.
- 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.

- 5) Средневзвешенная точность (СвТ) [1] — это среднее значение средней точности по всех классов при различных пороговых значениях ПНО по всей модели.
- 6) Среднее значение отзыва (СО) [1]: СО измеряет точность обнаружения ключевых точек в соответствии с показателем отзыва, который представляет собой отношение истинных положительных результатов к общему количеству положительных результатов. Другими словами, сколько из всех истинных положительных результатов было найдено моделью. Таким образом метрика СО представляет собой среднее значение отзыва по всем значениям отзыва от 0 и 1 при различных ПНО.
- 7) Сходство ключевых точек объектов (СКТО) [9]. 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 схождения от 0 до 1, а СКТО — это среднее значение всех этих значений по всем ключевым точкам. Эта метрика помогает в определении СТ и СО.
- 8) Средняя погрешность взаимного расположения (СПВР) [9]: это наиболее широко используемая метрика для трехмерного определения позы человека. Рассчитывается с помощью евклидова расстояния между оценочным трехмерным суставами и истинным положением следующим образом:

$$\text{СПВР} = \frac{1}{N} \sum_{i=1}^N \|J_i - J_i^*\|_2, \quad (2.1)$$

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

2.4 Стандарты и классификация для оценки позы человека

2.4.1 Таблица стандартов

Стандарты в определении позы человека относятся к количеству ориентиров, используемых для проведения оценки, то есть количество суставов человека или предопределенных ориентиров, которые необходимо локализовать. Это число зависит от типа выполняемой оценки позы и от самого метода. В таблице 2.1 перечислены несколько методов оценки позы тела, лица и руки, а также соответствующее число ориентиров, которые они используют.

Таблица 2.1 – Стандарты для разных типов оценки позы человека

Тип позы человека	Методы/Имя	Количество ориентиров
Тело	HPE OpenCV Github [10]	18
	Lightweight OpenPose Multi-Person HPE [11]	18
	Whole-Body HPE in the Wild [12]	23
	BlazePose [13]	33
Лицо	Whole-Body HPE in the Wild [12]	68
	CNN Facial Landmark Github [14]	68
	Facial Keypoint Detection Github [15]	68
	MediaPipe Face Mesh [16]	468
Руки	Whole-Body HPE in the Wild [12]	21
	MediaPipe Hands [16]	21
	CNN for 3D Hand Pose Estimation[17]	21

Как показано в таблице 2.1, большинство методов оценки позы лица использовали 68 ориентиров, поскольку поиск алгоритмов, кроме MediaPipe Face Mesh, которые использовали число ориентиров лица кроме 68 оказались безуспешным.

Аналогично, все методы оценки положения рук использован 21 ориентир на руку, потому что поиск алгоритмов, использующих число ориентиров на руку кроме 21 на руку, тоже оказались безуспешными.

На рисунке 2.4 приведены примеры одного и того же типа оценки позы, различающиеся по количеству ориентиров.

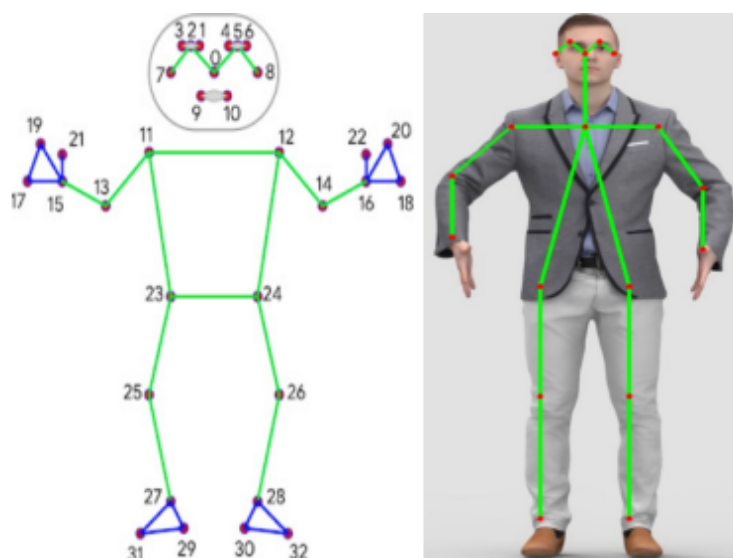


Рисунок 2.4 – Пример оценки позы.

2.4.2 Категоризация для определения позы человека

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 2.1. 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 2.1. Аналогично, при оценке позы руки требуется всего 42 ориентира 2.1. Для того чтобы учесть большое количество и размер ориентиров, необходимо высокое разрешение.

2.5 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 убедиться, 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.

2.5.1 Наборы данных по определению тела

- 1) COCO [9]:
- 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 2.1. Аналогично, при оценке позы руки требуется всего 42 ориентира 2.1. Для того чтобы учесть большое количество и размер ориентиров, необходимо высокое разрешение.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

- [1] A 2019 guide to Human Pose Estimation with Deep Learning [Электронный ресурс]. Режим доступа: <https://nanonets.com/blog/human-pose-estimation-2d-guide/> (дата обращения: 05.11.2022).
- [2] В.В.Вьюгин. Математические основы теории машинного обучения и прогнозирования. – М., 2013. – 387 с.
- [3] 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).
- [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] О.С. Сикороский. Обзор сверточных нейронных сетей для задачи классификации изображений // Новые информационные технологии в автоматизированных системах. 2017. С. 421–432.
- [6] Convolutional pose machines / Shih-En Wei, Varun Ramakrishna, Takeo Kanade [и др.] // CVPR. 2016.
- [7] 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).
- [8] Pose Estimation.Metrics. [Электронный ресурс]. Режим доступа: <https://stasiuk.medium.com/pose-estimation-metrics-844c07ba0a78> (дата обращения: 05.11.2022).
- [9] Keypoint Evaluation [Электронный ресурс]. Режим доступа: <https://cocodataset.org/#keypoints-eval> (дата обращения: 05.11.2022).
- [10] HPE OpenCV Github [Электронный ресурс]. Режим доступа: <https://github.com/quanhua92/human-pose-estimation-opencv> (дата обращения: 05.11.2022).
- [11] Osokin Daniil. Real-time 2D Multi-Person Pose Estimation on CPU: Lightweight OpenPose // arXiv preprint arXiv:1811.12004. 2018.

- [12] Whole-Body Human Pose Estimation in the Wild / Sheng Jin, Lumin Xu, Jin Xu [и др.] // Proceedings of the European Conference on Computer Vision (ECCV). 2020.
- [13] Pose Detection by Extending BlazeFace [Электронный ресурс]. Режим доступа: <https://ai.googleblog.com/2020/08/on-device-real-time-body-pose-tracking.html> (дата обращения: 05.11.2022).
- [14] cnn-facial-landmark [Электронный ресурс]. Режим доступа: <https://github.com/yinguobing/cnn-facial-landmark1> (дата обращения: 05.11.2022).
- [15] Facial-Keypoint-Detection-Udacity-PPB [Электронный ресурс]. Режим доступа: <https://github.com/ParthaPratimBanik/Facial-Keypoint-Detection-Udacity-PPB> (дата обращения: 05.11.2022).
- [16] mediapipe [Электронный ресурс]. Режим доступа: <https://github.com/google/mediapipe> (дата обращения: 05.11.2022).
- [17] CNN for 3D Hand Pose Estimatio [Электронный ресурс]. Режим доступа: <https://www.hindawi.com/journals/mpe/2020/8432840/> (дата обращения: 05.11.2022).