

Содержание

	Страница
1 Введение	4
2 Аналитический раздел	6
2.1 Обзор предметной области	6
Классические подходы к определению позы человека	6
2.2 Глубокое обучение в сверточных нейронных сетях в машине позирования	9
2.3 Точность и метрики	10
Определение точности и понятие метрики	10
Различные метрики, используемые в определении позы чело- века	11
Список литературы	12

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

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.

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.

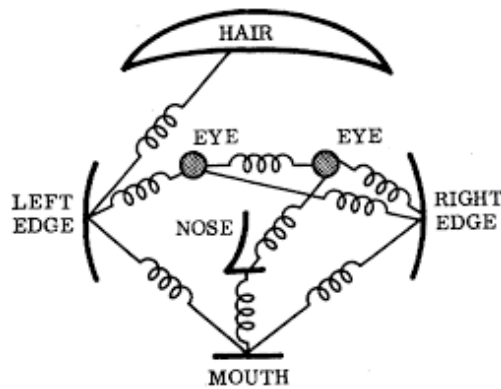


Рисунок 2.1 – Пример пиктографической структуры.

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.

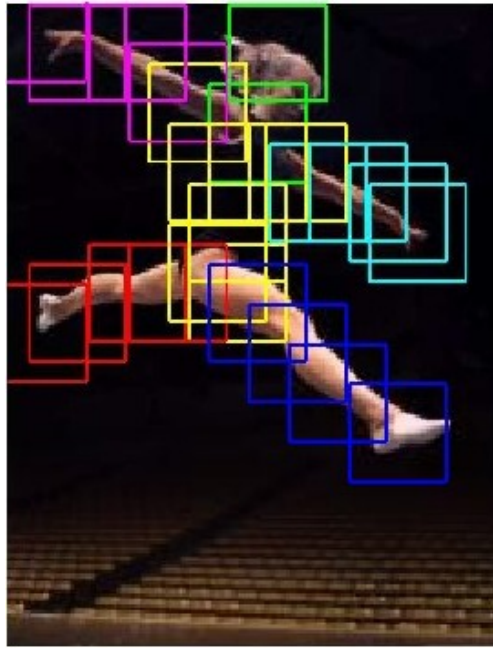


Рисунок 2.2 – Пример гибкого смещения частей.

- 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 другими 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 CPM 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 архитекту-

ре 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 CPM.

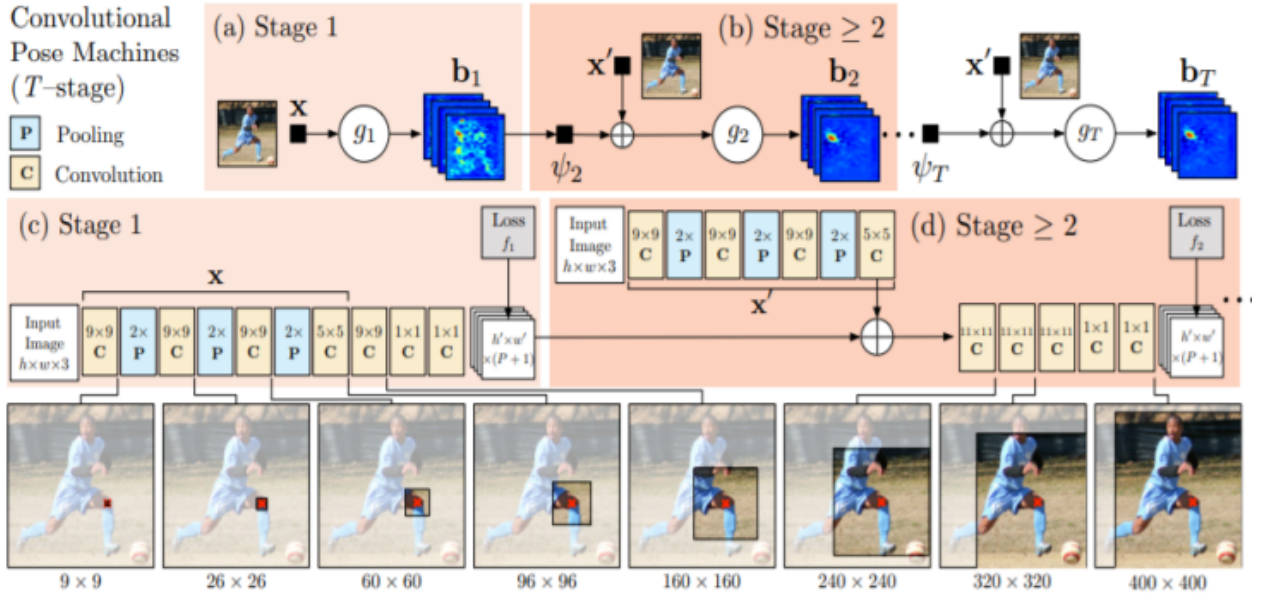


Рисунок 2.3 – Пример работы CPM.

2.3 Точность и метрики

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.

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 (СТ) [1] — это среднее значение средней точности по всем классам при различных пороговых значениях ПНО по всей модели.

Список литературы

- [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).