

К.Н. Гладкова, ассист.,
Я.М. Демяненко, доц., канд. физ.-мат. наук
(Южный федеральный университет, г. Ростов-на-Дону, Россия)

АНАЛИЗ МОДЕЛЕЙ ДЛЯ РЕШЕНИЯ ЗАДАЧИ РАСПОЗНАВАНИЯ ПОЗЫ ЧЕЛОВЕКА В СФЕРЕ ТРЕНИРОВОК ПО ЙОГЕ

Сфера применения моделей human pose estimation (HPE) в основном представлена спортом и медициной. Например, существует ряд приложений, выполняющих роль личного тренера и контролирующих технику выполнения различных упражнений. Приложение HomeCourt [1] анализирует движения баскетболистов и помогает им оценить их уровень игры и вырабатывает рекомендации для улучшения их навыков. Приложение Zenia Yoga [2] определяет правильность выполнения асан (поз) и таким образом делает опыт занятий дома безопаснее и продуктивнее. Однако с сентября 2023 года это приложение изъято из всех маркетплейсов.

Задача распознавания позы человека состоит в нахождении на изображении определенных ключевых точек (КТ) тела, таких как голова, плечи, локти, руки и так далее. По этим КТ и связям между ними воссоздается положение «скелета» позы. Главными проблемами при такой постановке задачи является широкое разнообразие поз и трудности точного определения КТ, когда они являются невидимыми с данного ракурса. Применительно к йоге это происходит почти во всех асанах сидя или «скручивающих» асанах.

Для решения этой задачи существует несколько различных подходов и моделей. Одной из первых успешных моделей определения позы одного человека стала модель «обучения с учителем» DeepPose [3] на основе глубоких нейронных сетей (DNN, Deep Neural Network). Она предсказывает конкретные координаты (x , y) КТ на входном изображении, при том, что КТ на изображении не ограничивается одним пикселем, а часто занимает некоторую область. Архитектура модели довольно проста: она состоит из каскада регрессионных блоков (рис. 1). Каждый блок состоит из семи слоев, где пять из них – сверточные и два – полносвязные. Цель блока – найти, а затем уточнить координаты интересующей точки тела, уменьшая зону поиска на каждой стадии.

Следующий подход использует сверточные нейросети (CNN, Convolutional Neural Network) [4] для регрессии не координат КТ, а пространственной карты распределения вероятности (heatmap) (рис.

2). Такой вариант предсказаний уменьшает ошибку при обучении, так как лучше описывает поведение КТ.



Рисунок 1 – Слева: Схематический вид DeepPose. Слои сети обозначены с собственными измерениями: сверточные слои – синие, полносвязные – зеленые. Справа: на стадии s уточняющий регрессор применен к части изображения для уточнения предсказания предыдущей стадии

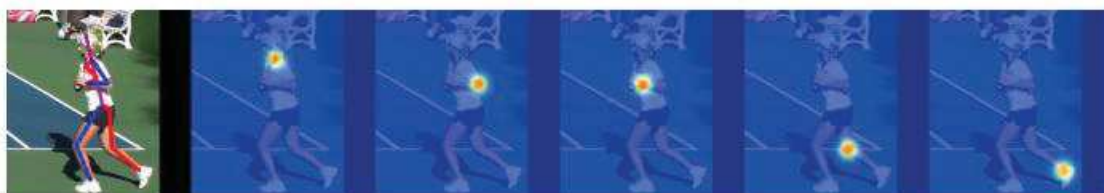


Рисунок 2 – Пример выхода в виде пространственной карты распределения вероятности, где каждая относится к определенной точке тела человека

Основной принцип этой модели – использование изображения в семи различных разрешениях (рис. 3), что помогает избавиться от необходимости использования метода «скользящего окна», что значительно ускоряет вычисления. Что позволяет заметить зависимости разного масштаба. В то время как локальные данные необходимы для идентификации таких особенностей, как лицо и руки, окончательная оценка позы требует глобального контекста.

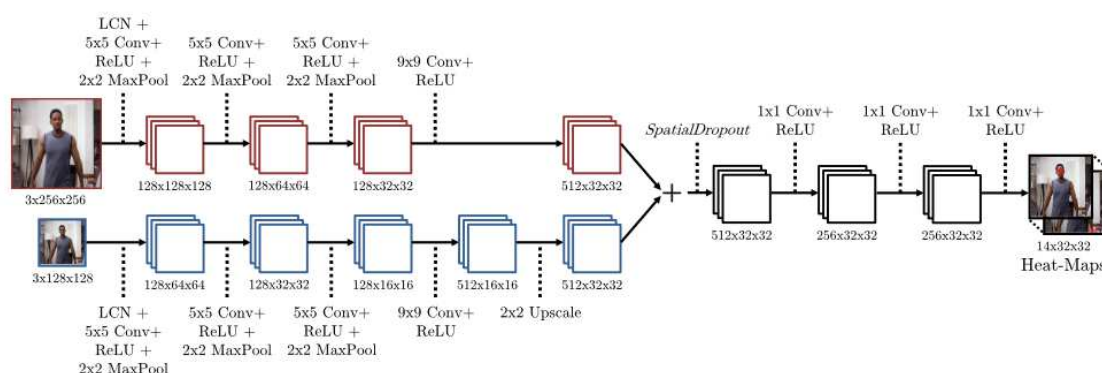


Рисунок 3 – Детектор со скользящим окном с перекрывающимися контекстами, применяющийся к изображению в различных разрешениях

Более продвинутым подходом является так называемая «последовательность песочных часов» (stacked hourglasses) [5]. Сходство с песочными часами блокам обучения придает последовательность

сверточных (pooling) и развертывающих (upsample) слоев, последовательно уменьшающих, а затем увеличивающих размерность изображения (рис. 4). Это повторяет предыдущую идею, но теперь модель может проследить взаимосвязь между масштабами и делать более точные целостные выводы о положении КТ.

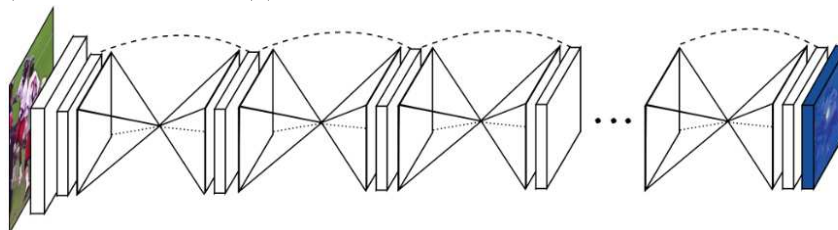


Рисунок 4 – Общий вид модели stacked hourglasses, состоящей из нескольких блоков «песочных часов»

Между слоями одинаковой размерности пропущены связи (рис. 5) для того, чтобы данные, потерянные при уменьшении размерности внесли свой вклад. Ориентация человека, расположение его конечностей и взаимоотношения соседних суставов входят в число многих признаков, которые лучше всего распознаются на изображении в разных масштабах. Применяя блоки песочных часов последовательно, модель вычисляет потери каждого блока и корректирует предсказания.

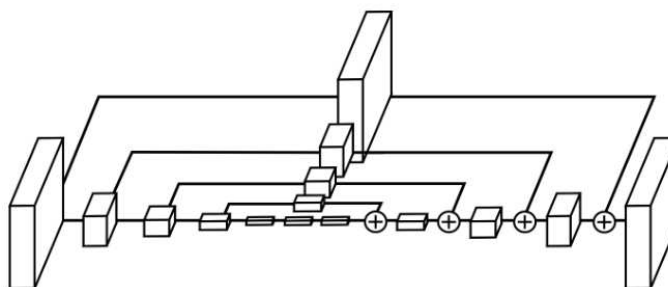


Рисунок 5 – Блок «песочных часов»

Одной из самых быстрых моделей на сегодняшний день является OpenPose [6], основная задача которой определять позы нескольких людей на изображении. Так же используя карту распределения вероятностей, как и в примерах выше, модель расширяет её картами аффинных полей. Это поля двумерных векторов, кодирующих расположение и ориентацию конечностей на изображении (рис. 6), что помогает ассоциировать части тела с конкретными людьми и корректно определить несколько поз на изображении. Карты распределения вероятностей и карты аффинных полей строятся одновременно независимо друг от друга, что позволяет ускорить обучение с помощью параллельных вычислений.

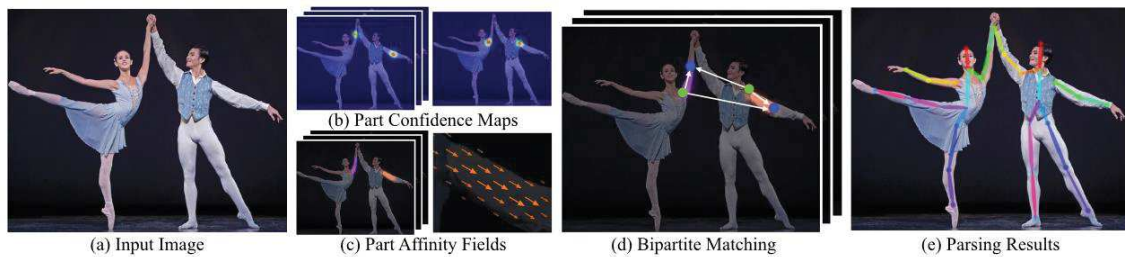


Рисунок 6 – Оценка изображений: (а) – входное изображение, (b) – карты распределения вероятностей для нахождения частей тела, (с) – карты аффинных полей для ассоциации частей тел. (d) – сведение соответствующих карт вероятностей и полей для ассоциации частей тел с людьми. (е) – объединение всех результатов

Все описанные выше модели являются представителями моделей «обучения с учителем», а значит – требуют большой набор качественно размеченных данных. К сожалению, пока нет специализированных размеченных датасетов, содержащих всё широкое разнообразие асан с разных ракурсов. Тем не менее, существуют модели «обучения без учителя», которые самостоятельно могут отыскивать зависимости на изображениях и не требуют дорогостоящей подготовки данных.

В статье [7] предложен подход определения позы через изменение базового шаблона «скелета» путем аффинных преобразований. Для обучения модель использует следующий алгоритм: из случайного изображения получает карту силуэта, потом берет другое изображение и преобразовывает его под вычисленный силуэт и в конце сравнивает выходное и первое изображения на совпадения (рис. 7). Такой подход отлично подходит для оценки поз младенцев и может использоваться в медицине для ранней диагностики различных неврологических заболеваний, но он сильно проигрывает в точности предсказаний КТ взрослых людей моделям «с учителем» ввиду сложности их поз и частой невидимостью КТ.

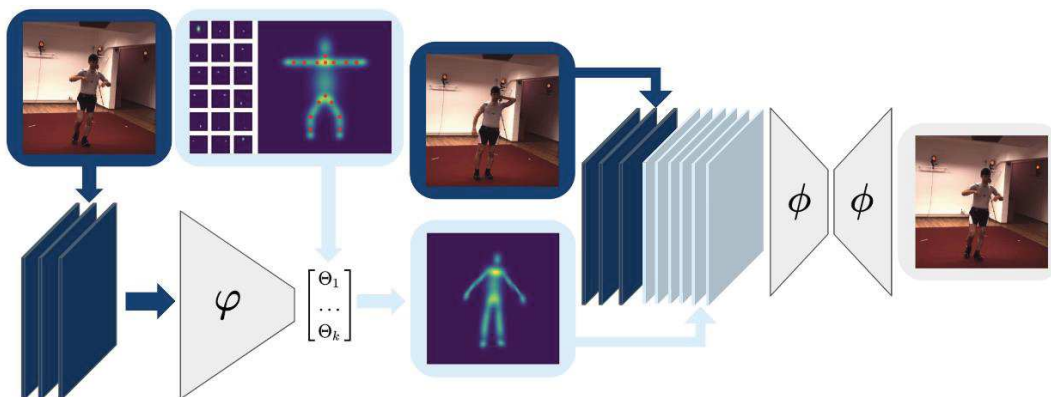


Рисунок 7 – Пример обучения модели

В рамках задачи распознавания позы человека в сфере тренировок по йоге нами используется модель на основе архитектуры «последовательности песочных часов», как быстрая и точная модель, строящая предсказания в виде карты распределения вероятностей и на основании системного анализа изображения в разных масштабах. При этом она была обучена на общем размеченном датасете поз людей MPII [8]. В дальнейшем предполагается дообучить её на самостоятельно собранном из открытых источников датасете, размеченном обученными и проверенными моделями в открытом доступе.

ЛИТЕРАТУРА

1. HomeCourt [Электронный ресурс] – Режим доступа: <https://www.homecourt.ai/> – Дата доступа 05.01.2024.
2. Йога глазами дата-сайентиста: как мы строили computer vision в мобильном приложении [Электронный ресурс] – Режим доступа: <https://habr.com/ru/articles/555162/> – Дата доступа: 05.01.2024.
3. Toshev A., Szegedy C. DeepPose: Human Pose Estimation via Deep Neural Networks [Электронный ресурс] – Режим доступа: <https://arxiv.org/pdf/1312.4659.pdf> – Дата доступа: 03.01.2024.
4. Tompson J., Goroshin R., Jain A., LeCun Y., Bregler C. Efficient Object Localization Using Convolutional Networks [Электронный ресурс] – Режим доступа: <https://arxiv.org/pdf/1411.4280.pdf> – Дата Доступа: 03.01.2024.
5. Newell A., Yang K., Deng J. Stacked Hourglass Networks for Human Pose Estimation [Электронный ресурс] – Режим доступа: <https://arxiv.org/pdf/1603.06937.pdf> – Дата доступа: 03.01.2024.
6. Cao Z., Hidalgo G., Simon T., Wei S., Sheikh Y. OpenPose: Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields [Электронный ресурс] – Режим доступа: <https://arxiv.org/pdf/1812.08008.pdf> – Дата доступа: 04.01.2024.
7. Schmidtke L., Vlontzos A., Ellershaw S., Lukens A., Arichi T., Kainz B. Unsupervised Human Pose Estimation through Transforming Shape Templates [Электронный ресурс] – Режим доступа: <https://arxiv.org/pdf/2105.04154.pdf> – Дата доступа: 04.01.2024.
8. MPII Human Pose Dataset [Электронный ресурс] – Режим доступа: <http://human-pose.mpi-inf.mpg.de/> – Дата доступа: 06.01.2024.