Министерство образования и науки Российской Федерации
Московский физико-технический институт (национальный исследовательский университет)

Физтех-школа радиотехники и компьютерных технологий
Кафедра интеллектуальных информационных систем и технологий

Выпускная квалификационная работа бакалавра

Исследование методов классификации движений человека

Автор:
Студент Б01-815а группы
Токарев Андрей Сергеевич

Научный руководитель:
Ст. Преподаватель
Воронков Илья Михайлович

Москва 2022

Аннотация

Исследование методов классификации движений человека

Токарев Андрей Сергеевич

Краткое описание задачи и основных результатов, мотивирующее прочитать весь текст

Содержание

1 Введение

Понимание движений человека является необходимой частью нашей жизни. При общении людьми часто используется жестикуляция, так как это помогает выражать чувства, эмоции и доносить свои мысли до окружающих . Из анализа позы человека можно сделать вывод о его состоянии. К примеру, хромота или нахождение в неестественном положении говорят о необходимости не только медицинской, но, возможно, и вашей помощи. Ещё можно обратится к психоанализу, а точнее к разделу о языке телодвижений. В нем по позе можно сделать вывод о характере человека или о текущем состоянии, его заинтересованности в беседе.Также работает распознавание движений. Если мы видим бегущих в панике людей, то наш мозг получает сигнал об опасности и спасает нас. Из приведенных ситуацию становится понятно, почему определение позы и классификация движений являются важными аспектами нашей жизни. В связи с развитием информационных технологий, человечество задумалось над выполнением данной задачи с помощью компьютера. Тогда можно будет добавить дополнительный источник информации для взаимодействия искусственного интеллекта с человеком.

При рассмотрении данной задачи через призму машинного обучения, получим, что нам нужно классифицировать положение человека, данные о котором необходимо каким-то образом получать. Первый способ - надеть на добровольца датчики и, считывая координаты каждого из них, построить на компьютере его позу и, таким образом, восстановить скелет для последующего анализа. Второй способ - искать особые точки на фотографии с помощью компьютерного зрения. Установим камеру и начнем анализировать положение и скелет человека, исходя из картинки. Тогда не придется закупать большое количество датчиков для снятия данных, а нужна будет только камера и вычислительные мощно-

сти для работы алгоритмов глубокого обучения. Несмотря на сложность реализации первого варианта, его удобно использовать для подготовки тренировочных датасетов [1].

Движение - это растянутый во времени процесс. Он анализируется по видеозаписям, каждая из которых представляет собой последовательность кадров. Поэтому первостепенно научиться работать с изображением. Как же собирать данные для модели классификации?

Восстановление скелета (Skeletal Representation), детекция (Pose Detection) и оценка позы (Pose Estimation), распознавание движения (Action Recognition) являются расширением одной задачи: распознавание ключевых точек на теле человека (Key-points Detection). Задачи, которая имеет прикладной смысл не только в связке с классификацией движений. В работе мы будем рассматривать распознавание только на картинке, то есть в 2-х мерном пространстве. Но ведь можно восстанавливать положение человека (скелет человека) в 3-х мерном пространстве [2, 3]. Используя генеративные нейронные сети можно воссоздавать не только скелет человека, но и тело человека [4]. Объединяя две предыдущие задачи можно получить набор данных из 3-х мерных людей в различных позах. Некоторые исследователи уже пробуют реализовать этот симбиоз на практике [5].

Если перейти в тематику биологических и медицинских наук, то можно развить данную тему на примере восстановления структуры тканей человека. Получается, мы сможем по фотографии моделировать распределение мышечных, жировых и других тканей в теле человека. Это поможет более детально изучать проблемы персонально, каждого человека и подбирать индивидуальные курсы лечения или диеты.

Восстановление скелета человека поможет спасателям анализировать положение человека под завалами и строить планы по его спасению, имея более детальную информацию. Правда в данном случае необходимо

быстродействие алгоритма и очень важно получить изображение человека.

В современном мире, где повсюду слышны разговоры о технологиях дополненной реальности и мета вселенной, найдем ещё одно применение для алгоритмов детектирования позы. Для нахождения в виртуальной вселенной необходимо транслировать человека туда, а значит можно с помощью видеокамер определять положение, восстанавливать скелет и получать итоговое изображение или 3-х мерную модель. Чем-то напоминает фильм "Первому Игроку Приготовиться" Стивена Спилберга. Добавим алгоритм генерации аватара вместо реального человека и получим рабочий алгоритм трансляции живого человека в мета вселенную.

Второй частью работы является задача классификации (Pose Classification), которая использует данные, полученные в первой части. Таким образом, мы построили алгоритм анализа движений человека на статическом изображении. По изображению мы не можем давать оценку поведению человека, но своеобразный "помощник" из полученного алгоритма будет хороший. Рассмотрим некоторые идеи применения.

Начать можно с медицины. Восстановление больных после операций, травм и несчастных случаев - это длительный и трудоемкий процесс, требующий постоянного присмотра врача. Если человек учится двигаться, то нужен тренер, который укажет на ошибки и исправит вас. Решение нашей задачи помогает таким пациентам. Анализ движений может сравнивать человека с эталоном и указывать на ошибки. Также при наблюдении за больным алгоритм может идентифицировать отклонения от нормального поведения и вызвать врача (к примеру увидеть приступы эпилепсии у человека). Это может спасти множество жизней по всему миру, просто вызывая врача в необходимый момент, а также помочь в востановлении.

Также можно выявлять у здорового человека заболевания или дефекты скелета. Можно анализировать сколиоз или сутулость и подсказывать людям, что надо стараться держать спину прямо. Хотя лучше направлять к врачу на консультацию и лечить дефекты позвоночника сразу. Проведя исследование населения, мы получим статистику тех или иных отклонений. Так уже сделали производители кроссовок и с помощью gait-анализа [6] помогут выбрать подходящую обувь.

Посмотрим теперь на спорт. Из классификатора можно сделать хорошего судью соревнований в тех видах, где надо различать, отслеживать положения тела. К примеру, GOOGLE придумали использовать классификатор как счетчик подтягиваний, приседаний или отжиманий [7] и это можно поместить в современный смартфон. Если углубиться дальше, то решение можно обернуть интерфейсом и создать хорошего робота-фитнес-тренера. Ведь настроив камеру смартфона на наблюдение за вами во время тренировки, приложение будет подсказывать вам правильную позу для упражнения и укажет на ошибки, если таковые имеются.

При развитии моделей в будущем, можно будет найти другие варианты применения технологии классификации движений человека. Можно анализировать поведение группы людей, но для этого надо хорошо восстанавливать скелет нескольких человек на одном изображении [8, 9, 10]. Также есть возможность предсказывать будущие действия человека при изучении уже имеющихся [11]. Если опять затронуть идею генеративных нейронных сетей, то можно генерировать движения человека по заданному начальному условию. Следовательно, можно создавать искусственные видеозаписи или добавлять неигровых персонажей (npc - non-player character) в виртуальную реальность. В медицине можно моделировать восстановление двигательной активности человека или моделирование

протезов индивидуально под каждого пациента.

Как можно заметить, применений можно придумать множество - необходимо реализовать проект и получить модель. В текущий момент в мире уже существует какое-то количество решений описанных выше задач. В предложенной вашему чтению работе я рассмотрю некоторые из них, приведу качественную оценку результатам эксперимента и сделаю вывод с определением дальнейшего моего развития в данной теме. (МОЖЕТ СТОИТ УБРАТЬ ПРО МОЕ РАЗВИТИЕ В ДАННОЙ ТЕМЕ?)

2 Постановка задачи

Здесь надо максимально формально описать суть задачи, которую потребуется решить, так, чтобы можно было потом понять, в какой степени полученное в результате работы решение ей соответствует. Текст главы должен быть написан в стиле технического задания, т.е. содержать как описание задачи, так и некоторый набор требований к решению

Как уже было сказано в главе 1, будет произведена классификация движений человека на изображении. Из изображения надо получить данные о принимаемой субъектом позе и классифицировать её на род деятельности человека. Получается мы решаем две задачи: предобработки данных, то есть извлечение расположения ключевых точек на теле человека, и их последующая категоризация. Рассмотрим их по отдельности.

2.1 Задача распознавания ключевых точек на теле человека

Первоначально необходимо понять каким образом можно распознать позу человека, чтобы в дальнейшем взять оттуда информацию для классификации. Человек смотрит на другого человека и анализирует его позицию исходя из данных о его расположении частей тела анализируемого. Получается нам необходимо найти части тела человека, каждая из которых ограничена какими-либо суставами. Последние можно и взять за ключевые точки, которые будут распознаваться моделью. Если соединить выходные данные, то получим рисунок, который большинство из нас рисовало в детстве. (ДОБАВИТЬ РИСУНОК И ВОССТАНОВЛЕННЫЙ СКЕЛЕТ?)

(Можно сказать, что на сегодня шний момент существует три типа моделей оценки позы человека

Необходимо определиться сколько точек на теле человека необходимо различать. На текущий момент стандартом является топология COCO (см. рис. 1a), которая включает в себя 17 ориентиров на теле человека [12, 13]. Данная топология не учитывает расположение ступней и кистей рук, а также рассматривает всего 5 точек на лице человека: нос, два глаза и два уха. Но стандартом многие исследователи не ограничиваются и добавляют дополнительные точки. Приведу два примера:

1. Топология от BlazePose (см. рис. 1b)

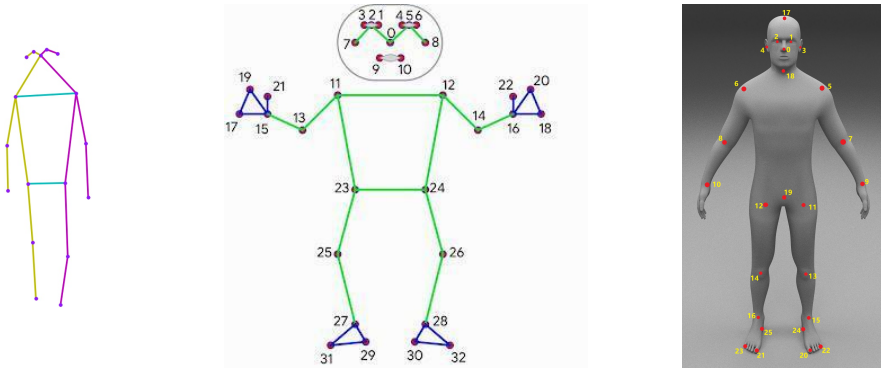
   Включает в себя 33 точки расположенные на теле человека. Данная топология представляет собой объединение COCO, BlazeFace [14] и BlazePalm [15]. В итоге мы получаем дополнительную информацию о направлении стоп и кистей, а также больше понимаем насчет точек на лице. Данная модель расположения точек используется в одноименной модели (BlazePose [16]) и ориентирована на использование в фитнес приложениях. Также у данной компании есть более развитая модель, которая определяет положение всех пальцев кисти и распознает мимику на лице [17].

2. Halpe (см. рис. 1c)

   Данная топология - это совместный проект AlphaPose [10] и HAKE [18]. Представлено две модели: на 26 и на 136 точек. Здесь добавлено рассмотрение ориентации стоп, распознавание шеи, паха и макушки головы. В расширенной модели присутствует ещё 68 точек на лице, а также по 21 на ладонях.

В итоге мы разобрались с тем что нам необходимо искать в нашей работе и сейчас необходимо понять как это делать. Данная работа проводится в два этапа:

1. Локализация человека и его частей тела



(a) Топология COCO (b) Топология BlazePose (c) Топология Halpe

Рис. 1: Примеры расположения точек на теле человека.

2. Упорядочивание и распределение суставов в правильном порядке.

В реальном мире мы имеем два подхода к поиску ключевых точек и восстановлению скелета на изображении:

- Bottom-up

  Когда сначала распознаем точки, потом собираем их в скелет

- Top-down

  Сначала происходит локализация людей или объектов, а потом происходит распознование ключевых точек

ДУМАЮ ТУТ СТОИТ ПРОДОЛЖИТЬ ИЗ РАЗДЕЛА МНОГО ТЕОРИИ ДЛЯ РАЗБАВКИ РАБОТЫ. О ТОМ КАК ДЕТЕКТИРОВАТЬ И ИСКАТЬ ЭТИ ТОЧКИ. ТАКЖЕ МОЖНО ДОБАВИТЬ ПРО РАЗЛИЧНЫЕ ФУНКЦИИ И ФОРМУЛКИ. ДОЛЖНО ПОЛУЧИТСЯ КРАСИВО И ИНТЕРЕСНО.

2.2 Задача классификации движений/позы человека

Полученные координаты ключевых точек можно использовать как признаки для различных классификаторов. Думаю тут не стоит сильно за-

морачиваться с описанием задачи машинного обучения классификации. Может просто забить, а может расписать и добавить ещё несколько страниц. СПРОСИТЬ У НАУЧРУКА ПРО ДАННЫЙ РАЗДЕЛ!!!

3 Обзор существующих моделей

3.1 Модели для распознавания ключевых точек на теле человека

В данном разделе мы рассмотрим 5-6 различных моделей. В главе 5 мы выберем 4 наиболее удобные в использовании и в обучении и проведем эксперимент по оценке данных моделей.

Так же хочется сказать, что, помимо приведенных, есть множество моделей от одиночных авторов, не объединенных в лаборатории (ССЫЛКИ). Они в основном брали какую-то из представленных моделей и проводили небольшое улучшение.

А теперь перейдем к моделям.

3.1.1 BlazePose

MediaPipe является одним из проектов компании GOOGLE и в своей работе решает задачи компьютерного зрения. В нем уже были представлены модели для распознавания лица (Face Detection) и его поверхности (Face Mesh), ладоней (Hands), объектов (Object Detection и Objectron) и другие [19]. Для нас же интересна задача поиска ключевых точек, которую и решает модель BlazePose [16]. На момент исследования модель умеет отслеживать движения человека на видеофрагменте и строить покадровую маску человека.

Для предложенной модели была создана топология, которая представляет собой суперпозицию топологии COCO и двух других топологий, уже использовавшихся в других подпроектах MediaPipe. Об этом более подробно написано в разд. 2.1.

В BlazePose используется top-down подход оценки позы человека. Сна-

чала запускается Pose Detector (см рис. 2), который возвращает координаты интересующей нас области (region-of-interest или ROI). Алгоритм используем расширение модели BlazeFace для определения наличия человека в кадре. Поэтому данная модель чувствительна к видимости головы, лица в частности, на фотографии. Взяв идею витрувианского человека Леонардо Да Винчи, исследователям понадобилось ещё две точки для точной локализации человека на изображении.

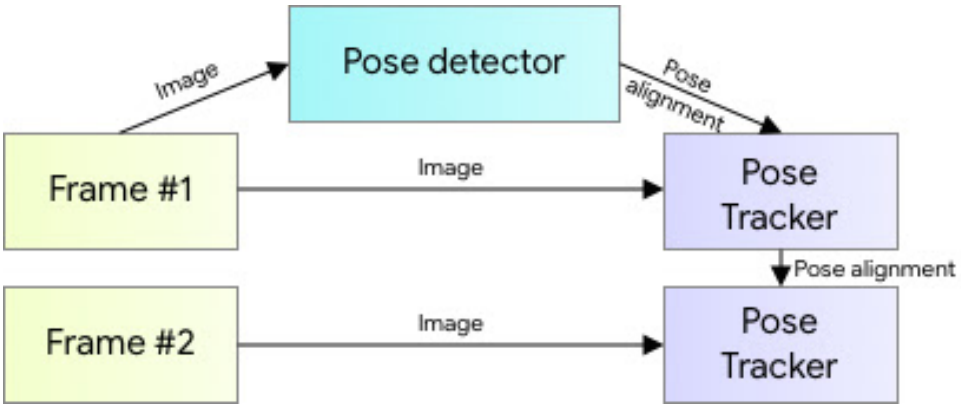


Рис. 2: Структура модели BlazePose для работы в реальном времени.

Оригинальное изображение

Следующим шагом Pose Tracker производит локализацию каждой точки в заданной ROI. Данное действие производится путем комбинированной обработки тепловой карты и данных о смещении с использованием регрессионной модели (см рис. 3). ВОЗМОЖНО СТОИТ ПОЧИТАТЬ ПРО ТАКУЮ ОБРАБОТКУ И ПОПОДРОБНЕЕ ЕЕ ОПИСАТЬ.

Как можно заметить из рис. 2, при аналитике видеофрагмента Pose Detector используется только на первом кадре, ведь позже данные об интересующей нас области передаются от кадра к кадру. Это упрощает вычисления и позволяет ускорить работу модели в реальном времени.

Развитием данной моели есть ее полное объединение с моделями BlazeFace и BlazeHand в модель Holistic [17]. Она рассматривает намного большее количество точек на лице и ладонях. НАПИСАТЬ ПРО ЕЕ ПРИМЕНЕНИЕ...

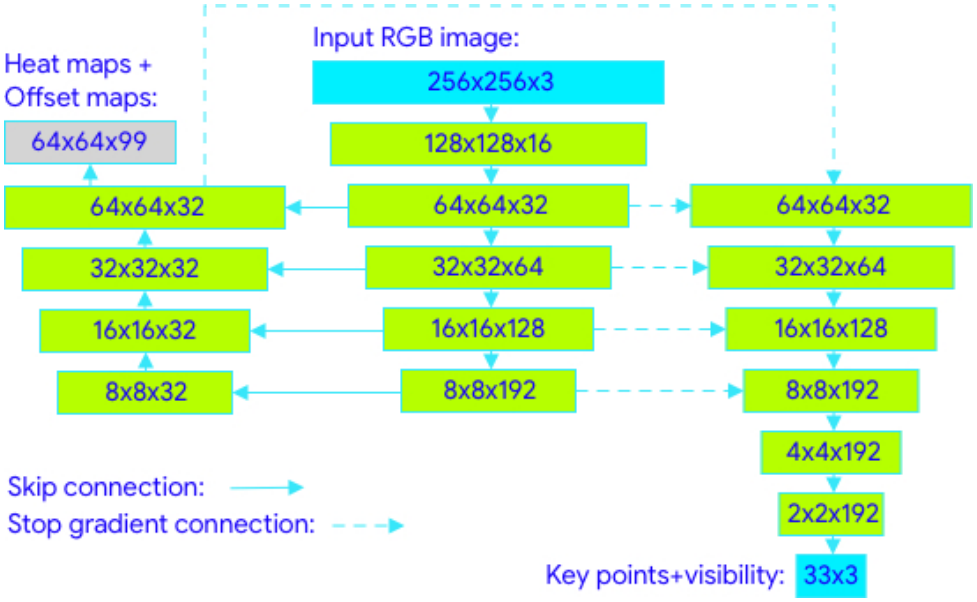


Рис. 3: Архитектура модели Pose Tracker.

Оригинальное изображение

3.1.2 MoveNet.SinglePose

SinglePose создана для работы в веб-интерфейсах или на мобильных устройствах в режиме реального времени. Модель представлена в двух спецификациях: lightning и thunder. Первая является менее требовательной в плане мощностей и вычислений и способна обрабатывать до 50 кадров в секунду. В то же время, по заверениям создателей, вторая модель имеет большие запросы по ресурсам, но дает лучшую точность распознавания, правда со скоростью до 30 кадров в секунду.

За расположение ключевых точек выбрана классическая топология COCO. Поэтому возвращает модель координаты 17 точек, которые нормированы на размер изображения (лежат в отрезке [0, 1]).

Представленная модель реализована на архитектуре MobileNetV2 [20]. СЛОЖНАЯ АРХИТЕКТУРА. НЕОБХОДИМО ПРО НЕЕ ПРОЧИТАТЬ, А ПОТОМ УЖЕ ДОПИСАТЬ ДАННЫЙ РАЗДЕЛ.

Также существует развитие проекта MoveNet в MoveNet.MultiPose

для распознавания сразу нескольких людей на изображении. Усовершенствованная модель также представлена в двух вариациях. В данной работе много персональное распознавание позы не исследовалось.

3.1.3 OpenPose

Проект CMU-Perceptual-Computing-Lab. В лаборатории есть множество моделей для работы отдельно с лицом, руками или что-то подобное модели Holistic от MediaPipe.

В данной модели используется топология, похожая на топологию Halpe, рассмотренную в разд. 2.1. Рассматриваются 25 точек на теле человека. Особенностью топологии является определение положения стоп за счет детекции 3 точек на каждой.

АРХИТЕКТУУУУУУУУУУУУУРА

3.1.4 MMPose

Азиаты

3.1.5 AlphaPose

Азиаты?

3.1.6 Detectron2

Работает не очень приятно, но как бы она есть и то ладно. Может обойдемся без данной модели?

3.1.7 DeepPose

Является очень старым решением, но про него нельзя было не рассказать.

3.2 Модели для классификации позы человека

В данном разделе мы рассмотрим 4 различных моделей. Позже выберем 3 наиболее удобные в использовании и в обучении и проведем эксперимент по оценке данных моделей.

3.2.1 MMaction2 by OpenMMlab

Что-то сложное...

3.2.2 BlazePose by MediaPipe

Просто накинули сверху предыдущей модели kNN.

3.2.3 mmakos

Чувак взял OpenPose и добавил классификатор.

3.2.4 HPC

Уже и не помню про что тут...

4 Поиск данных

4.1 Data for Pose Estimation

4.1.1 LSP

4.1.2 MPII HRD

4.1.3 COCO keipoints

4.2 Data for Pose Classification

4.2.1 HPC/mmakos

4.2.2 MPII

4.2.3 Yoga-82

4.3 Not Available data

4.3.1 Human3.6M

4.3.2 Surreal

4.3.3 BUFF

5 Исследование моделей

В данной главе я опишу поставленный эксперимент по исследованию моделей и приведу полученные рузльтаты.

5.1 Описание эксперимента

Эксперимент, как и вся работа разделен на две части: исследование моделей распознавания ключевых точек на теле человека и исследование моделей классификации поз человека.

В первой части мы рассмотрели 4 модели на работоспособность. Все они показали хороший результат классификации. Для определения наиболее хорошей можели использовались метрики: PCK и PDJ (возможно OKS, но с ней пока что я не разобрался). Метрика высчитывалась на точках туловища, так как в датасете нет размеченных точек лица (только верхушка головы, а она в исследуемых моделях не присутствует).

Во второй части ... ее ещё надо написать и создать. Надеюсь успеть это сделать на выходных...

5.2 Полученные результаты эксперимента

Выбранный мной датасет имеет фотографии низкого качества и маленького размера, поэтому приведу показ работоспособности моделей на фотографиях собственной работы.

(ФОТОЧКИ)

Также хочу привести результаты высчитывания метрик для нескольких моделей и различных пороговых значений в метриках.

В дополнение приведем средние значения по обработке одного изображения моделью.

Вторую чать работы пишем...

6 Заключение

Здесь надо перечислить все результаты, полученные в ходе работы. Из текста должно быть понятно, в какой мере решена поставленная задача.

Список литературы

[1] Human3.6M: Large Scale Datasets and Predictive Methods for 3D Human Sensing in Natural Environments / Catalin Ionescu, Dragos Papava, Vlad Olaru, Cristian Sminchisescu // *IEEE Transactions on Pattern Analysis and Machine Intelligence*. — 2014. — jul. — Vol. 36, no. 7. — Pp. 1325–1339.

[2] Deep 3D human pose estimation: A review / Jinbao Wang, Shujie Tan, Xiantong Zhen et al. // *Computer Vision and Image Understanding*. — 2021. — Vol. 210. — P. 103225. https://www.sciencedirect.com/science/article/pii/S1077314221000692.

[3] *Tome, Denis*. Lifting from the Deep: Convolutional 3D Pose Estimation from a Single Image / Denis Tome, Chris Russell, Lourdes Agapito // 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). — 2017. — Pp. 5689–5698.

[4] Detailed, Accurate, Human Shape Estimation From Clothed 3D Scan Sequences / Chao Zhang, Sergi Pujades, Michael J. Black, Gerard Pons-Moll // The IEEE Conference on Computer Vision and Pattern Recognition (CVPR). — 2017. — July.

[5] Learning from Synthetic Humans / Gül Varol, Javier Romero, Xavier Martin et al. // CVPR. — 2017.

[6] *Whittle, Michael W.* Clinical gait analysis: A review / Michael W. Whittle // *Human Movement Science*. — 1996. — Vol. 15, no. 3. — Pp. 369–387. https://www.sciencedirect.com/science/article/pii/0167945796000061.

[7] *google.github.io*. Pose Classification. — https://google.github.io/mediapipe/solutions/pose_classification.html.

[8] OpenPose: Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields / Z. Cao, G. Hidalgo Martinez, T. Simon et al. // *IEEE Transactions on Pattern Analysis and Machine Intelligence*. — 2019.

[9] *Kocabas, Muhammed*. MultiPoseNet: Fast Multi-Person Pose Estimation using Pose Residual Network. — 2018. https://arxiv.org/abs/1807.04067.

[10] RMPE: Regional Multi-person Pose Estimation / Hao-Shu Fang, Shuqin Xie, Yu-Wing Tai, Cewu Lu // ICCV. — 2017.

[11] Prediction of Human Activities Based on a New Structure of Skeleton Features and Deep Learning Model / Neziha Jaouedi, Francisco J. Perales, José Maria Buades et al. // *Sensors*. — 2020. — Vol. 20, no. 17. https://www.mdpi.com/1424-8220/20/17/4944.

[12] *Tsung-Yi Lin Matteo Ruggero Ronchi, Alexander Kirillov Yin Cui*. COCO 2020 Keypoint Detection Task. — https://cocodataset.org/#keypoints-2020.

[13] *Lin, Tsung-Yi*. Microsoft COCO: Common Objects in Context. — 2014. https://arxiv.org/abs/1405.0312.

[14] *Bazarevsky, Valentin*. BlazeFace: Sub-millisecond Neural Face Detection on Mobile GPUs. — 2019. https://arxiv.org/abs/1907.05047.

[15] *Zhang, Fan*. MediaPipe Hands: On-device Real-time Hand Tracking. — 2020. https://arxiv.org/abs/2006.10214.

[16] *Bazarevsky, Valentin*. BlazePose: On-device Real-time Body Pose tracking. — 2020. https://arxiv.org/abs/2006.10204.

[17] *Ivan Grishchenko, Valentin Bazarevsky*. MediaPipe Holistic — Simultaneous Face, Hand and Pose Prediction, on

Device. — https://ai.googleblog.com/2020/12/mediapipe-holistic-simultaneous-face.html.

[18] PaStaNet: Toward Human Activity Knowledge Engine / Yong-Lu Li, Liang Xu, Xinpeng Liu et al. // CVPR. — 2020.

[19] *google.github.io*. MediaPipe.Home. — https://google.github.io/mediapipe/.

[20] *Sandler, Mark*. MobileNetV2: Inverted Residuals and Linear Bottlenecks. — 2019.

[21] *Contributors, MMPose*. OpenMMLab Pose Estimation Toolbox and Benchmark. — https://github.com/open-mmlab/mmpose. — 2020.

[22] Hand Keypoint Detection in Single Images using Multiview Bootstrapping / Tomas Simon, Hanbyul Joo, Iain Matthews, Yaser Sheikh // CVPR. — 2017.

[23] Convolutional pose machines / Shih-En Wei, Varun Ramakrishna, Takeo Kanade, Yaser Sheikh // CVPR. — 2016.