

УДК 004.946.2+004.932.72+004.272

СРАВНИТЕЛЬНЫЙ АНАЛИЗ БИБЛИОТЕК ДЛЯ ДЕТЕКТИРОВАНИЯ ПОЗЫ ЧЕЛОВЕКА В УСЛОВИЯХ РАБОТЫ НА МОБИЛЬНЫХ УСТРОЙСТВАХ

Е. И. Ярко^[0009-0009-0502-1066]

Казанский (Приволжский) федеральный университет, г. Казань, 420008, Россия

yarkoeg@gmail.com

Аннотация

Оценка положения тела человека (Human Pose Estimation, HPE) стала одной из наиболее актуальных тем в исследованиях в области компьютерного зрения. Эта технология может применяться в различных сферах, таких как видеонаблюдение, медицинская помощь и анализ спортивных движений.

В связи с растущим спросом на HPE за последние 20 лет было разработано большое количество библиотек для этой технологии. С 2017 года опубликовано множество алгоритмов HPE, основанных на скелетной модели, которые были упакованы в библиотеки для удобства использования исследователями. Эти библиотеки важны для исследователей, которые хотят интегрировать их в реальные приложения для видеонаблюдения, медицинской помощи и анализа спортивных движений.

В работе рассмотрены преимущества и недостатки четырёх популярных передовых библиотек HPE для распознавания поз человека, которые могут работать на мобильных устройствах: Lightweight OpenPose, PoseNet, MoveNet и Blaze Pose.

Ключевые слова: *поза человека, Human Pose Estimation, детектирование позы, компьютерное зрение, мобильные устройства, дополненная реальность, Lightweight OpenPose, PoseNet, MoveNet, BlazePose, скелетная модель.*

ВВЕДЕНИЕ

Распознавание поз человека (Human Pose Estimation, HPE) – это процесс определения положения всех частей тела человека на входных изображениях или видео. В настоящее время технология HPE стала востребованной в сфере компьютерного зрения и дополненной реальности (Augmented Reality, AR). Она широко применяется в системах видеонаблюдения [1–6], медицинской помощи [7–15], анализе спортивных движений [16–28] и во многих других сферах. В частности, технологии определения позы человека успешно применяются в реабилитационных играх для постинсультных пациентов [47]. Такие игры позволяют сделать процесс реабилитации более эффективным и увлекательным для пациентов.

Для классификации поз и оценки их корректности в областях использования, перечисленных выше, используются ключевые точки на теле человека (рис. 1). С помощью интеллектуальной системы компьютерного зрения можно определить ключевые точки на скелете, что может быть применимо для решения различного спектра задач.

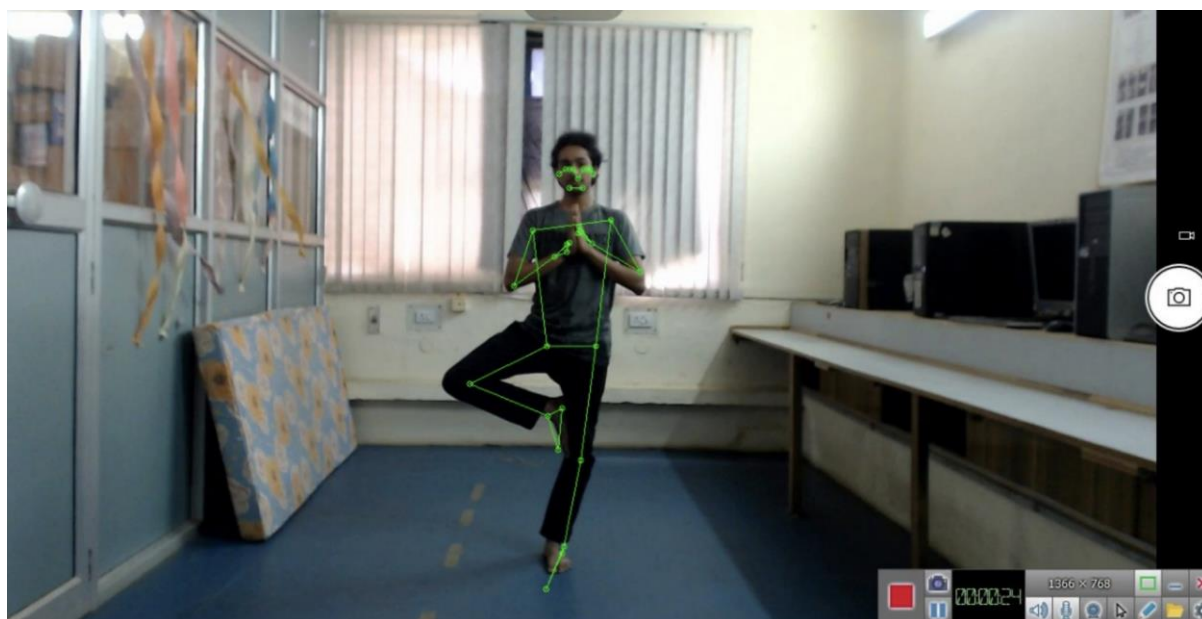


Рис. 1. Построение ключевых точек с помощью BlazePose

В сфере медицинской помощи детектирование ключевых точек на теле может быть использовано для оценки правильности выполнения упражнений

физиотерапии, обнаружения падений и реабилитации на дому. Кроме того, анализ ключевых точек на теле спортсмена позволяет оценить его производительность и корректность движений по сравнению с эталонными позами (ground truth).

HPE можно разделить на два типа: двумерную (2D) и трёхмерную (3D). Ключевое отличие между 2D- и 3D-оценками позы заключается в наличии дополнительной z-координаты, которая отвечает за глубину. Данные о глубине могут включать несколько камер (стерео- или мультивидовые системы), данные от глубинных датчиков (например, RGB-D-камеры) или использовать методы, которые восстанавливают глубину на основе одного RGB-изображения (моно-сетап), опираясь на статистическую регрессию.

Кроме того, HPE можно классифицировать по количеству определяемых в кадре людей: одного человека или нескольких. В обоих случаях, HPE можно дополнительно разделить на методы сверху вниз (top-down) и снизу вверх (bottom-up) в зависимости от способов детектирования ключевых точек скелета [29].

Основная задача настоящей работы – изучить практики применения библиотек детектирования позы человека с целью выбора оптимального решения в условиях работы программного обеспечения на мобильных устройствах, не обладающих глубинными датчиками. Мы представим сравнительное исследование двумерной оценки поз одного человека.

С ростом спроса на HPE было разработано большое количество алгоритмов, основанных на скелетной модели, которые объединены в библиотеки для удобства использования. Производительность этих библиотек HPE важна для обеспечения надёжности различных практических приложений, в которые они интегрируются. Например, если библиотека HPE используется в системе реабилитации на дому, она должна точно определять позы пациентов, выполняющих реабилитационные упражнения в различных домашних условиях, чтобы гарантировать надёжность приложения. Эта задача становится ещё более сложной, когда выявление ключевых точек скелета затруднено такими распространён-

ными проблемами, как неправильное положение камеры и самозакрытие (явление, при котором одна часть тела закрывается другой, что затрудняет детекцию, например, когда человек закрывает рукой лицо) [11, 20, 22, 23].

В последние годы четыре библиотеки HPE нашли применение в различных областях: PoseNet [30], MoveNet [31], Lightweight OpenPose [32] и Blaze Pose [33]. Во всех вышеприведенных библиотеках HPE возникают распространенные проблемы в оценке поз, такие как неправильное положение камеры и эффект самозакрывания.

Хотя каждая библиотека HPE использует свои подходы для преодоления отмеченных проблем, плюсы и минусы этих четырёх библиотек HPE пока не были исследованы. Поэтому необходимо провести сравнительный анализ производительности этих библиотек, чтобы оценить их надёжность в детектировании различных поз человека. Итак, цель этой работы – сравнить четыре передовые библиотеки HPE для распознавания поз человека и проанализировать их преимущества и недостатки.

2. СРАВНИТЕЛЬНЫЙ АНАЛИЗ БИБЛИОТЕК HPE

2.1. Существующий сравнительный анализ

В работах [33, 34], посвящённых сравнительному анализу, сопоставлена эффективность нескольких библиотек HPE, работающих с наборами изображений.

В одном из исследований [34] для сравнения OpenPose и BlazePose [33] был использован набор данных AR. В качестве критерия оценки была применена метрика PCK@0.2. Результаты показали, что OpenPose продемонстрировала несколько более высокую производительность по сравнению с BlazePose, набрав 87,8 и 84,1 балла соответственно.

В другом исследовании [31] были использованы OpenPose, PoseNet, MoveNet Lightning и MoveNet Thunder. Выбраны два набора данных изображений: COCO [35] и MPII [36]. Эффективность библиотек HPE была оценена с помощью собственной метрики. Наилучшие результаты показала PoseNet, в то время как MoveNet Lightning продемонстрировала худшие результаты.

Еще одно исследование [38] было посвящено сравнению четырех HPE библиотек: OpenPose, PoseNet, MoveNet и MediaPipe Pose. Авторы сравнили библиотеки на наборах данных из видеороликов и статичных изображений, а в качестве оценки производительности использовали показатель процентного охвата суставов (Percentage of Detected Joints, PDJ). MoveNet показала наилучшую производительность в выявлении различных поз человека на статических изображениях и видео. В [38] описаны проведенные эксперименты с использованием наборов данных изображений и видео.

2.2. Детальное сравнение библиотек HPE

В этом разделе рассмотрены библиотеки HPE, которые могут функционировать на мобильных устройствах, и их характеристики сведены в табл. 1.

Таблица 1. Характеристики библиотек оценки позы человека

Название библиотеки	Год выпуска	Максимальное число ключевых точек	Расположение ключевых точек в частях тела	Метод определения	Нейросетевая архитектура
PoseNet 2.0	2022	17	Глаза, уши, плечи, локти, запястья, бедра, колени, лодыжки	Bottom-up	ResNet
MoveNet	2021	17	Глаза, уши, плечи, локти, запястья, бедра, колени, лодыжки	Top-down	MobileNetV2
Lightweight OpenPose	2018	18	Включает шею, нос и дополнительные суставы рук	Bottom-up	MobileNet
BlazePose	2020	33	Полное тело с пальцами рук и ног	Top-down	BlazePose CNN

В библиотеках HPE было выявлено 17 ключевых точек, которые встречаются наиболее часто. Среди них пять точек расположены на голове: это уши,

глаза и нос. Шесть ключевых точек находятся на плечах, локтях и запястьях и относятся к верхней части тела, а шесть точек на бёдрах, коленях и лодыжках – к нижней части тела. Однако в Lightweight OpenPose добавлена одна дополнительная ключевая точка на шее, а BlazePose – более детализированные точки на лице и руках (например, мочки ушей, кончики пальцев).

В библиотеках НРЕ можно выделить два подхода к обнаружению ключевых точек: «сверху вниз» (Top-down) и «снизу вверх» (Bottom-up).

В подходе «сверху вниз» сначала определяют количество людей на изображении, после чего каждому человеку присваивают отдельную рамку. Далее в каждой рамке происходит обнаружение ключевых точек. В отличие от подхода «сверху вниз», в подходе «снизу вверх» сначала устанавливаются ключевые точки, а затем они группируются по человеческим фигурам.

Среди рассматриваемых библиотек подход «сверху вниз» используют MoveNet и BlazePose, тогда как PoseNet 2.0 и Lightweight OpenPose – подход «снизу вверх».

PoseNet 2.0 использует свёрточные нейронные сети MobileNetV1 и MobileNetV2 для эффективной оценки позы на мобильных устройствах и в браузерах. MobileNet обладает легкой архитектурой, что позволяет сети работать быстро и с низкими требованиями к вычислительным ресурсам, сохраняя точность распознавания позы на базовом уровне.

Библиотека MoveNet представлена в двух вариантах: MoveNet Lightning и MoveNet Thunder. MoveNet Lightning работает на основе MobileNetV3, обеспечивая высокую производительность на устройствах с ограниченными ресурсами, MoveNet Thunder использует более крупную архитектуру, разработанную для высокой точности на более мощных устройствах. Эти сети оптимизированы для точной и быстрой работы в режиме реального времени.

Сеть Lightweight OpenPose была разработана с использованием облегченных версий сетей MobileNet и VGG. В этой версии OpenPose применены упрощенные и сжимаемые сети для обеспечения более эффективной работы на устройствах с низкими вычислительными возможностями, с сохранением при этом основной структуры и подхода к оценке позы оригинального OpenPose.

Сеть BlazePose, разработанная Google Research и MediaPipe, использует модифицированную архитектуру BlazeFace, адаптированную для работы с полным телом, а также более легкие мобильные свёрточные нейронные сети. Эти сети оптимизированы для точной оценки позы с минимальной задержкой, что делает BlazePose популярной для приложений в фитнесе и отслеживании движений в реальном времени.

Первая версия PoseNet была выпущена в 2017 году [30]. Она построена на платформе машинного обучения TensorFlow и предоставляет реализацию HPE в реальном времени в браузере. В PoseNet встроены два алгоритма: один – для оценки позы одного человека, а другой – для оценки поз нескольких человек на входном изображении или видео. Оба алгоритма способны детектировать 17 ключевых точек на одном человеке. Время вычисления в алгоритме HPE для нескольких человек немного больше, чем для одного, но оно не зависит от количества обнаруженных людей. При использовании алгоритма для одного человека ключевые точки могут смешиваться, если на входном изображении или видео имеется более одного человека. Кроме этого, в PoseNet существуют две архитектуры – ResNet [37] и MobileNet [40]. MobileNet предназначена для мобильных устройств, она легче, чем ResNet, но имеет более низкую точность. Хотя ResNet достигает более высокой точности по сравнению с MobileNet, её большее количество слоёв требует более длительного времени загрузки и выполнения.

В 2022 году была выпущена вторая версия библиотеки PoseNet. PoseNet 2.0 была оптимизирована для меньшего потребления ресурсов и повышения производительности за счёт использования архитектуры MobileNetV2.

MoveNet (год выпуска – 2021) представляет собой модель для распознавания поз, которая в режиме реального времени определяет 17 ключевых точек на одном человеке. MoveNet существует в двух версиях: Lightning и Thunder. Точность версии Lightning ниже, чем у Thunder, но время выполнения у неё меньше.

MoveNet использует тепловые карты для точного определения местоположения ключевых точек человека. Архитектура модели состоит из двух частей: извлекателя признаков и набора предсказательных модулей.

Предсказательные модули MoveNet частично основаны на технологии CenterNet, которая применяется для повышения точности и скорости работы. CenterNet – это детектор объектов, который использует сети оценки ключевых точек для нахождения центральных точек и определения размера, местоположения и ориентации объектов.

Извлекатель признаков – это MobileNetV2 с прикреплённой сетью пирамиды признаков, которая генерирует высокоразрешённый и семантически насыщенный выходной признак.

MobileNetV2 – это нейронная сеть, оптимизированная для мобильных устройств, которая используется для извлечения признаков для обнаружения объектов, их классификации и семантической сегментации.

В предсказательных модулях MoveNet выделены четыре части: тепловая карта центра человека, поле регрессии ключевых точек, тепловая карта ключевых точек и поле 2D-смещения для каждой ключевой точки. Эти компоненты отвечают за предсказание ключевых точек человека с использованием тепловых карт.

Lightweight OpenPose была создана как упрощённая версия оригинальной OpenPose, разработанной исследовательской группой Carnegie Mellon University в 2017 году. OpenPose является одной из первых библиотек для оценки позы человека, способной отслеживать несколько ключевых точек на теле, лице и конечностях, что позволило применить её в задачах компьютерного зрения, таких как отслеживание движений, жестов и анализа поведения. Однако оригинальная OpenPose требовала значительных вычислительных ресурсов, поскольку использовала глубокую архитектуру на базе VGG-19, что делало её неподходящей для работы на мобильных устройствах и устройствах с ограниченными ресурсами.

Основное улучшение в Lightweight OpenPose достигалось за счёт использования облегченной архитектуры сети, которая сочетает компоненты MobileNet и модифицированную версию VGG. Вместо глубокой сети VGG-19, как в оригинальной версии, Lightweight OpenPose использует MobileNet в качестве базовой архитектуры для выделения признаков, что значительно снижает по-

требление ресурсов, но сохраняет достаточную точность для большинства приложений. В процессе модификации сеть была также дополнительно сжата и оптимизирована для работы в реальном времени.

MobileNet служит основным компонентом для извлечения признаков, поскольку её свёрточные слои сконструированы так, чтобы работать быстрее и эффективнее на маломощных устройствах за счёт использования глубинно-разделимых свёрток.

Для повышения точности при минимальных затратах производительности была также применена сжимаемая архитектура с уменьшенным количеством слоев и параметров, что делает модель компактной и подходящей для встраиваемых приложений.

BlazePose была разработана в 2020 году командой Google Research и MediaPipe как специализированное решение для точной оценки человеческой позы. В отличие от предыдущих моделей для оценки позы, таких как OpenPose, основная задача BlazePose состояла в создании модели, способной обеспечивать высокие точность и производительность при минимальном потреблении ресурсов, что стало необходимым условием для мобильных приложений, особенно в фитнесе, спорте и других задачах реального времени.

BlazePose опирается на уникальную архитектуру, в основе которой лежит комбинация BlazeFace и специализированных свёрточных нейронных сетей, адаптированных для оценки позы всего тела. Перечислим ключевые архитектурные особенности.

- **Двухэтапная архитектура:** BlazePose использует два этапа для оценки позы. На первом этапе сеть обнаруживает грубую позу тела, на втором – уточняет позиции ключевых точек. Это позволяет снизить общую вычислительную нагрузку и повысить точность модели.
- **BlazeFace** как основа для BlazePose: изначально BlazeFace разрабатывалась для анализа лицевой позы в реальном времени, и её архитектура была улучшена для работы с позой тела. Основная сеть – легковесная свёрточная сеть, ориентированная на скорость и производительность на мобильных устройствах.

- **Использование MediaPipe:** BlazePose интегрирована с фреймворком MediaPipe, что позволяет использовать GPU и CPU для параллельной обработки данных и обеспечивает высокую производительность даже на устройствах со средними и низкими ресурсами. MediaPipe включает готовые инструменты для предобработки данных, оптимизированные для работы на мобильных и веб-платформах.
- **33 ключевые точки:** BlazePose выделяется в сравнении с другими сетями за счёт отслеживания 33 ключевых точек, включая детализированные точки на лице и руках, что делает её подходящей для более точного анализа движений в приложениях, таких как спортивная аналитика и отслеживание позы в фитнесе.

3. МЕТОДОЛОГИЯ ОЦЕНКИ

В этом разделе рассмотрена методология оценки работоспособности библиотек HPE, самые распространенные наборы данных и аннотаций, на которых проверяют работоспособность библиотек и их подготовку к оценке, а также ключевую метрику.

3.1. Наборы данных

Одними из самых распространенных наборов данных (датасетов), используемых в научной среде, являются Microsoft Common Object in Context (COCO), который содержит изображения и аннотации к ним, а также Penn Action, который содержит отобранные видеоролики. Характеристики этих наборов приведены в табл. 2.

Таблица 2. Характеристики библиотек оценки позы человека

Набор данных	Тип данных	Ключевых точек на человека	Предоставленная аннотация частей тела
COCO	Изображения	17	Нос, Левый глаз, Правый глаз, Левое ухо, Правое ухо, Левое Плечо, Правое плечо, Левый локоть, Правый локоть, Левое запястье, Правое запястье, Левое бедро, Правое бедро, Левое Колено, Правое Колено, Левая Лодыжка, Правая Лодыжка
Penn Action	Видеоролики	13	Голова, Левое плечо, Правое плечо, Левый локоть, Правый локоть, Левое запястье, Правое запястье, Левое бедро, Правое Бедро, Левое колено, Правое колено, Левая лодыжка, Правая лодыжка

Оба набора содержат по шесть ключевых точек для верхней и нижней частей тела. Однако есть различие в количестве аннотированных ключевых точек для головы. В COCO их пять, включая нос, глаза и уши, а в Penn Action для головы указана только одна ключевая точка.

Примеры изображений из наборов данных COCO и Penn Action можно увидеть на рис. 2 и 3 соответственно.



Рис. 2. Пример изображений из набора данных COCO

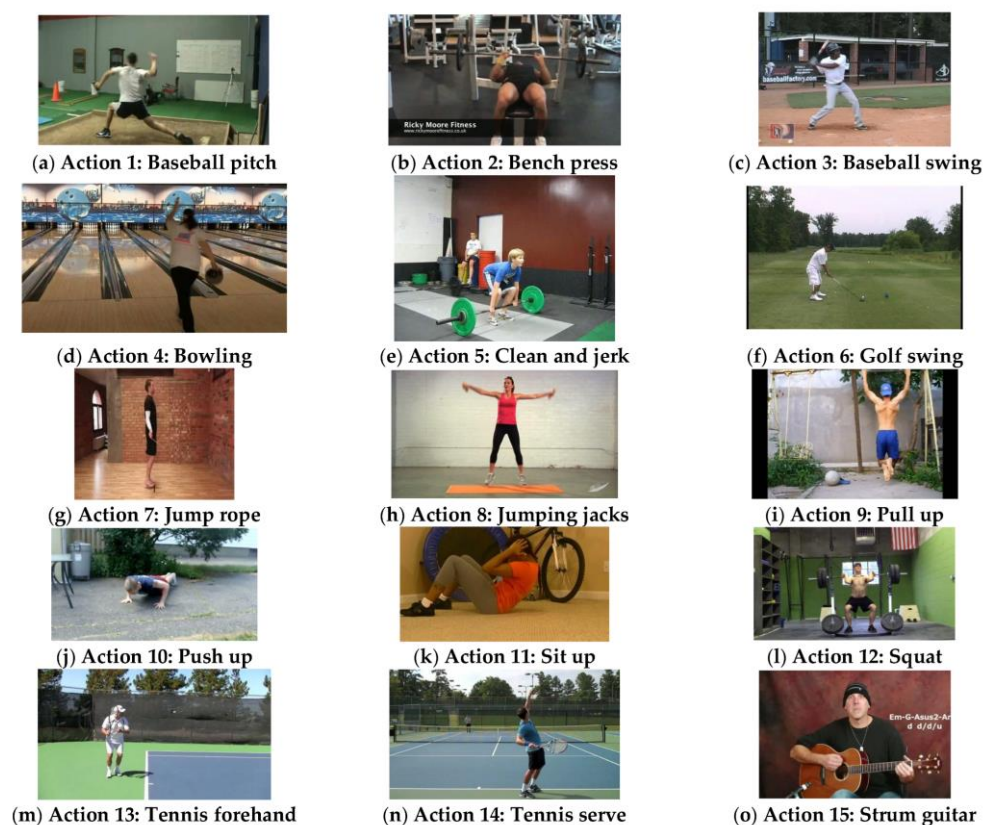


Рис. 3. Примеры кадров видео из датасета Penn Action

3.2. Подготовка наборов данных

Перед оценкой производительности библиотек НРЕ нужно провести предобработку данных для отсеивания нерелевантных данных в перечисленных выше наборах. В наборе данных COCO имеются три типа изображений: изображения с одним человеком, с несколькими людьми и без людей.

Мы рассматриваем задачу НРЕ для одного человека, поэтому изображения с несколькими людьми и без людей были удалены. Нужно также исключить изображения, на которых присутствует только часть тела. Для сравнения производительности четырех библиотек НРЕ 17 общих ключевых точек тела нужно сопоставить с 17 аннотациями, предоставленными набором данных (истинными значениями).

Что касается видео из набора Penn Action, в эксперименте использовались первые 14 действий (см. рис. 2). Поскольку Penn Action предоставил только одну

аннотацию ключевой точки головы, отличную от аннотаций в четырех библиотеках HPE (см. табл. 3), аннотация для головы была удалена из экспериментов для обеспечения объективного сравнения между библиотеками.

3.3. Метрика оценки

Метрики оценки играют важную роль в определении качества библиотек HPE. В качестве основной используется метрика «процент обнаруженных суставов» (PDJ), которая позволяет оценить производительность библиотеки HPE [37, 39]. PDJ использует евклидово расстояние между истинными и предсказанными ключевыми точками в пикселях для измерения точности обнаружения библиотек HPE. Чем выше значение PDJ, тем выше точность. Евклидово расстояние $d(x, y)$ между истинными (x_1, y_1) и предсказанными ключевыми точками (x_2, y_2) находится по формуле (1), представленной ниже.

Порог PDJ установлен на уровне 0.05 от диаметра туловища. Диаметр туловища рассчитывался как евклидово расстояние от левого плеча до правого бедра с координатами $[x_{ls}, y_{ls}]$ и $[x_{rh}, y_{rh}]$, как показано в формуле (2).

Если величина $d(x, y)$ между предсказанными и истинными ключевыми точками меньше порогового значения, предсказанная ключевая точка считалась обнаруженной корректно. Таким образом, PDJ может быть рассчитан, как показано в формуле (3), где n обозначает общее количество предсказанных суставов.

$$d(x, y) = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2} \text{ (pixel)} \quad (1)$$

$$\text{torso diameter} = \sqrt{(x_{ls} - x_{rh})^2 + (y_{ls} - y_{rh})^2} \quad (2)$$

$$PDJ = \frac{\sum_{i=1}^n \text{bool}(d_i < 0.05 \times \text{torso diameter})}{n} \quad (3)$$

4. Преимущества и недостатки рассматриваемых библиотек

В этом разделе рассмотрены четыре популярных библиотеки для HPE: PoseNet, OpenPose, MoveNet и BlazePose. Представлены ключевые преимущества и недостатки каждой из них, установленные на основе анализа научных работ и открытых исходных кодов.

4.1. PoseNet

Библиотека PoseNet и её версии на основе TensorFlow.js отличаются простотой интеграции и использования, что делает её идеальным инструментом для разработки практических приложений в браузере без необходимости установки сложных зависимостей [40].

Модель PoseNet оптимизирована для работы с мобильными и веб-приложениями. Её цель – обеспечить работу в режиме реального времени на устройствах с ограниченными ресурсами, что упрощает интеграцию в пользовательские сервисы.

Кроме того, PoseNet не требует специальных знаний и навыков, что делает её доступной широкому кругу пользователей. Благодаря подробной документации и готовым примерам модель часто используется в образовательных и любительских проектах. Но при этом библиотека PoseNet создана для быстрого и лёгкого определения ключевых точек, поэтому точность может быть ниже, чем у более сложных моделей [40].

PoseNet ориентирована на работу с одним объектом в кадре. Для распознавания нескольких объектов одновременно требуется дополнительная настройка или использование других методов.

4.2. OpenPose

Это одна из первых библиотек, которая позволяет определять позы нескольких человек в реальном времени с высоким качеством.

OpenPose использует карту связей (Part Affinity Fields), что позволяет эффективно устанавливать взаимосвязи между конечностями и сегментировать несколько человек даже при наличии перекрытий [41].

OpenPose широко известна и активно используется в научных исследованиях и промышленных прототипах. Но при этом OpenPose требует значительных

ресурсов. Модель занимает много памяти и нагружает процессор и графический процессор, что затрудняет её использование на мобильных устройствах и в приложениях с ограниченными вычислительными ресурсами [41].

OpenPose требует более глубокого понимания параметров и окружения для корректного развёртывания, особенно при необходимости ускорения вычислений. В отличие от PoseNet и MoveNet, OpenPose не так проста в настройке.

4.3. MoveNet

Библиотека MoveNet разработана с учётом ограниченных вычислительных ресурсов и предназначена для работы в режиме реального времени, включая мобильные устройства [42]. Это позволяет использовать ее в различных сферах, в том числе в медицине, спорте и робототехнике.

Существует несколько версий MoveNet, которые позволяют выбирать между точностью и скоростью. Например, версии Thunder и Lightning отличаются по этим параметрам.

MoveNet легко интегрируется с экосистемой TensorFlow. Модель доступна через TensorFlow Hub и может быть легко встроена в различные приложения, поддерживающие TF Lite, что упрощает разработку и развёртывание этой библиотеки. MoveNet хорошо подходит для отслеживания общих положений тела, но ее архитектура не всегда обеспечивает такую же точность локализации мелких деталей, как у более тяжёлых моделей. MoveNet использует упрощённый набор ключевых точек, аналогичный COCO, что может быть недостаточно для некоторых специализированных задач [42], например для анализа движений в танце или в медицинских целях.

4.4. BlazePose

Библиотека BlazePose, разработанная в рамках MediaPipe, способна отслеживать до 33 ключевых точек, включая дополнительные ориентиры кистей и стоп, что обеспечивает более детальное представление о позе.

Перечислим основные преимущества этой библиотеки.

1) Оптимизация для мобильных устройств: BlazePose, как и другие решения MediaPipe, предназначены для работы в режиме реального времени

на смартфонах и планшетах, поддерживая ускорение вычислений на графическом процессоре [44].

2) Гибкость в сочетании с другими моделями: благодаря интеграции в MediaPipe, BlazePose часто используется совместно с другими библиотеками для распознавания жестов и выражений лица, формируя комплексные решения в области компьютерного зрения [45, 46].

3) Зависимость от инфраструктуры MediaPipe: для разработки с использованием BlazePose необходимо понимание экосистемы MediaPipe, что может усложнить интеграцию в сторонние проекты.

Таким образом, каждая из рассмотренных библиотек обладает собственным набором преимуществ и ограничений, что позволяет выбирать конкретное решение под определённые задачи – от мобильных приложений и веб-сервисов (PoseNet, MoveNet, BlazePose) до комплексных систем отслеживания позы множества людей в реальном времени (OpenPose).

ЗАКЛЮЧЕНИЕ

Рассмотрев самые распространённые библиотеки определения позы человека (HPE), а также метрику оценки их качества, можно сделать вывод, что для решения задачи определения позы человека оптимальным выбором будет библиотека BlazePose, так как она способна отслеживать 33 ключевые точки и при этом показывает лучший результат в условиях работы на мобильных устройствах. В будущем планируется представить результаты эксперимента детекции позы человека, подтверждающие выводы, сформулированные выше, а в качестве ключевой метрики использовать предложенную выше метрику, равную проценту правильно определенных суставов.

СПИСОК ЛИТЕРАТУРЫ

1. *Su C., Li J.; Zhang S., Xing J., Gao W., Tian Q.* Pose-driven deep convolutional model for person re-identification // *Proceedings of the IEEE International Conference on Computer Vision*. 2017. P. 3960–3969.
2. *Xu J., Zhao R., Zhu F., Wang H., Ouyang W.* Attention-aware compositional network for person re-identification // *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2018. P. 2119–2128.

3. *Thyagarajmurthy A., Ninad M.G., Rakesh B.G., Niranjan S., Manvi B.* Anomaly detection in surveillance video using pose estimation // *Proceedings of the Emerging Research in Electronics, Computer Science and Technology*; Springer. 2019. P. 753–766.
4. *Lamas A., Tabik S., Montes A.C., Pérez-Hernández F., García J., Olmos R., Herrera F.* Human pose estimation for mitigating false negatives in weapon detection in video-surveillance // *Neurocomputing*. 2022. Vol. 489. P. 488–503.
5. *Yoo H.R., Lee B.H.* An OpenPose-based child abuse decision system using surveillance video // *J. Korea Inst. Inf. Commun. Eng.* 2019. Vol. 23. P. 282–290.
6. *Park J.H., Song K., Kim Y.-S.* A Kidnapping Detection Using Human Pose Estimation in Intelligent Video Surveillance Systems // *J. Korea Soc. Comput. Inf.* 2018. Vol. 23. P. 9–16.
7. *Chang Y.J., Chen S.F., Huang J.D.* A Kinect-based system for physical rehabilitation: A pilot study for young adults with motor disabilities // *Res. Dev. Disabil.* 2011. Vol. 32. P. 2566–2570.
8. *Hassan H.A., Abdallah B.H., Abdallah A.A., Abdel-Aal R.O., Numan R.R., Darwish A.K., El-Behaidy W.H.* Automatic Feedback for Physiotherapy Exercises Based On PoseNet // *FCAI-Inform. Bull.* 2020. Vol. 2. P. 10–14.
9. *Shapoval S., García Zapirain B., Mendez Zorrilla A., Mugueta-Aguinaga I.* Biofeedback applied to interactive serious games to monitor frailty in an elderly population // *Appl. Sci.* 2021. Vol. 11. P. 3502.
10. *Chua J., Ong L.Y., Leow M.C.* Telehealth using PoseNet-based system for in-home rehabilitation // *Future Internet*. 2021. Vol. 13. P. 173.
11. *Kim W., Sung J., Saakes D., Huang C., Xiong S.* Ergonomic postural assessment using a new open-source human pose estimation technology (OpenPose) // *Int. J. Ind. Ergon.* 2021. Vol. 84. P. 103164.
12. *Jawale C.D., Joshi K.A., Gogate S.K., Badgujar C.* Elcare: Elderly Care with Fall Detection // *J. Phys. Conf. Ser.* 2022. Vol. 2273. P. 012019.
13. *Kapoor R., Jaiswal A., Makedon F.* Light-weight seated posture guidance

system with machine learning and computer vision // Proceedings of the 15th International Conference on Pervasive Technologies Related to Assistive Environments. 2022. P. 595–600.

14. Landi H. Google, ProMedica Team up with Include Health to Tap into Growing Virtual MSK Market // Fierce Healthcare. URL: <https://www.fiercehealthcare.com/tech/google-promedica-team-up-include-health-to-tap-into-virtual-msk-market>

15. Chen W., Jiang Z., Guo H., Ni X. Fall detection based on key points of human skeleton using OpenPose // Symmetry. 2020. Vol. 12. P. 744.

16. Zou J., Li B., Wang L., Li Y., Li X., Lei R., Sun S. Intelligent fitness trainer system based on human pose estimation // Proceedings of the International Conference On Signal and Information Processing, Networking and Computers. 2018. P. 593–599.

17. Suda S., Makino Y., Shinoda H. Prediction of volleyball trajectory using skeletal motions of setter player // Proceedings of the 10th Augmented Human International Conference. 2019. P. 1–8.

18. Wang J., Qiu K., Peng H., Fu J., Zhu J. AI coach: Deep human pose estimation and analysis for personalized athletic training assistance // Proceedings of the 27th ACM International Conference on Multimedia. 2019. P. 374–382.

19. Jeon H., Yoon Y., Kim D. Lightweight 2D human pose estimation for fitness coaching system // Proceedings of the 2021 36th International Technical Conference on Circuits/Systems, Computers and Communications (ITC-CSCC). 2021. P. 1–4.

20. Li Y.C., Chang C.T., Cheng C.C., Huang Y.L. Baseball Swing Pose Estimation Using OpenPose // Proceedings of the 2021 IEEE International Conference on Robotics, Automation and Artificial Intelligence (RAAI). 2021. P. 6–9.

21. Park H.J., Baek J.W., Kim J.H. Imagery based Parametric Classification of Correct and Incorrect Motion for Push-up Counter Using OpenPose // Proceedings of the 2020 IEEE 16th International Conference on Automation Science and Engineering (CASE). 2020. P. 1389–1394.

22. Nguyen H.T.P., Woo Y., Huynh N.N., Jeong H. Scoring of Human Body-Balance Ability on Wobble Board Based on the Geometric Solution // Appl. Sci. 2022. Vol. 12. P. 5967.

23. *Patil A., Rao D., Utturwar K., Shelke T., Sarda E.* Body Posture Detection and Motion Tracking using AI for Medical Exercises and Recommendation System // ITM Web Conf. 2022. Vol. 44. P. 03043.

24. *Devanandan M., Rasaratnam V., Anbalagan M.K., Asokan N., Panchendrarajan R., Tharmaseelan J.* Cricket Shot Image Classification Using Random Forest // Proceedings of the 2021 3rd International Conference on Advancements in Computing (ICAC). 2021. P. 425–430.

25. *Joseph R., Ayyappan M., Shetty T., Gaonkar G., Nagpal A.* BeFit—A Real-Time Workout Analyzer // Proceedings of the Sentimental Analysis and Deep Learning. 2022. P. 303–318.

26. *Mahendran N.* Deep Learning for Fitness // arXiv. 2021. arXiv:2109.01376.

27. *Agarwal S., Gupta M., Khandelwal S., Jain P., Aggarwal A., Singh D., Mishra V.K.* FitMe: A Fitness Application for Accurate Pose Estimation Using Deep Learning // Proceedings of the 2021 2nd International Conference on Secure Cyber Computing and Communications (ICSCCC). 2021. P. 232–237.

28. *Nakai M., Tsunoda Y., Hayashi H., Murakoshi H.* Prediction of basketball free throw shooting by OpenPose // Proceedings of the JSAI International Symposium on Artificial Intelligence. 2018. P. 435–446.

29. *Zheng C., Wu W., Chen C., Yang T., Zhu S., Shen J., Kehtarnavaz N., Shah M.* Deep learning-based human pose estimation: A survey // arXiv. 2020, arXiv:2012.13392.

30. *Papandreou G., Zhu T., Kanazawa N., Toshev A., Thompson J., Bregler C., Murphy K.* Towards accurate multi-person pose estimation in the wild // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu. 2017. P. 4903–4911.

31. *Jo B., Kim S.* Comparative Analysis of OpenPose, PoseNet, and MoveNet Models for Pose Estimation in Mobile Devices // Trait. du Signal. 2022. Vol. 39. P. 119–124.

32. *Cao Z., Simon T., Wei S.E., Sheikh Y.* Realtime multi-person 2d pose estima-

tion using part affinity fields // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017. P. 7291–7299.

33. *Erramchetty S.K. et al.* Human pose estimation using blaze pose // AIP Conference Proceedings. 2024. Vol. 2971. No. 1.

34. *Gadhiya R., Kalani N.* Analysis of Deep Learning Based Pose Estimation Techniques for Locating Landmarks on Human Body Parts // Proceedings of the 2021 International Conference on Circuits, Controls and Communications (CCUBE). 2021. P. 1–4.

35. *Lin T.Y., Maire M., Belongie S., Hays J., Perona P., Ramanan D., Dollár P., Zitnick C.L.* Microsoft coco: Common objects in context // Proceedings of the European Conference on Computer Vision. 2014. P. 740–755.

36. *Andriluka M., Pishchulin L., Gehler P., Schiele B.* 2D human pose estimation: New benchmark and state of the art analysis // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2014. P. 3686–3693.

37. *Toshev A., Szegedy C.* DeepPose: Human pose estimation via deep neural networks // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2014. P. 1653–1660.

38. *Chung J.-L., Ong L.-Y., Leow M.-C.* Comparative Analysis of Skeleton-Based Human Pose Estimation // Future Internet. 2022. Vol. 14, No. 12. P. 380.
<https://doi.org/10.3390/fi14120380>

39. *Ahmedt-Aristizabal D., Nguyen K., Denman S., Sridharan S., Dionisio S., Fookes C.* Deep motion analysis for epileptic seizure classification // Proceedings of the 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. 2018. P. 3578–3581.

40. *Ishaq A., Haq I.U.* Real-time single-person pose estimation on low-power embedded devices // 2019 International Conference on Frontiers of Information Technology (FIT).

41. *Cao Z., Simon T., Wei S.-E., Sheikh Y.* Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2017. P. 7291–7299.

42. Papandreou G., Zhu T., Chen L.-C., Gidaris S., Thompson J., Murphy K. PersonLab: Person Pose Estimation and Instance Segmentation with a Bottom-Up, Part-Based, Geometric Embedding Model // Proceedings of the European Conference on Computer Vision (ECCV). 2018. P. 269–286.

43. TensorFlow Blog. MoveNet: Ultra fast and accurate pose detection model. 2021.

URL: <https://blog.tensorflow.org/2021/05/movenet-and-posture-analysis.html>

44. Bazarevsky V., Zhang F. BlazePose: On-Device Real-Time Body Pose Tracking. Google AI Blog. 2020.

URL: <https://ai.googleblog.com/2020/08/on-device-real-time-body-pose-tracking.html>

45. Gabdullina D.R., Zykov E.Yu., Kugurakova V.V. Using OpenCV for real-time image recognition through augmented reality devices // The 2023 International Conference on Artificial Life and Robotics (ICAROB2023). 2023. P. 868–874.

46. Gabdullina D.R., Zykov E.Yu., Kugurakova V.V. Using OpenCV for augmented reality applications // Journal of Advances in Artificial Life Robotics. 2023. Vol. 3. Iss. 3(3). P. 179–184.

47. Бакиров А.Р. и др. Использование контроллера Microsoft Kinect в разработке реабилитационных игр // Электронные библиотеки. 2016. Т. 19. №. 6. С. 521–537.

48. You Y., Zhao Y. A human pose estimation algorithm based on the integration of improved convolutional neural networks and multi-level graph structure constrained model // Pers. Ubiquitous Comput. 2019. Vol. 23. P. 607–616.

COMPARATIVE ANALYSIS OF LIBRARIES FOR HUMAN POSE DETECTION IN MOBILE DEVICE ENVIRONMENTS

E. I. Yarko^[0009-0009-0502-1066]

Kazan Federal university, Kazan, 420008, Russia

yarkoeg@gmail.com

Abstract

Human Pose Estimation (HPE) has become one of the most relevant topics in computer vision research. This technology can be applied in various fields such as video surveillance, medical care, and sports motion analysis. Due to the increasing demand for HPE, many libraries for this technology have been developed in the last 20 years. Since 2017, many HPE algorithms based on skeletal model have been published and packaged into libraries for easy use by researchers.

These libraries are important for researchers who want to integrate them into real-world applications for video surveillance, medical care, and sports motion analysis.

This paper investigates the strengths and weaknesses of four popular HPE advanced human pose recognition libraries that can run on mobile devices: Lightweight OpenPose, PoseNet, MoveNet, and Blase Pose.

Keywords: *Human pose, Human Pose Estimation, HPE, pose detection, computer vision, mobile devices, augmented reality, Lightweight OpenPose, PoseNet, MoveNet, BlazePose, skeletal model.*

REFERENCES

1. Su C., Li J.; Zhang S., Xing J., Gao W., Tian Q. Pose-driven deep convolutional model for person re-identification // Proceedings of the IEEE International Conference on Computer Vision. 2017. P. 3960–3969.
2. Xu J., Zhao R., Zhu F., Wang H., Ouyang W. Attention-aware compositional network for person re-identification // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018. P. 2119–2128.
3. Thyagarajmurthy A., Ninad M.G., Rakesh B.G., Niranjana S., Manvi B.

Anomaly detection in surveillance video using pose estimation // Proceedings of the Emerging Research in Electronics, Computer Science and Technology; Springer. 2019. P. 753–766.

4. *Lamas A., Tabik S., Montes A.C., Pérez-Hernández F., García J., Olmos R., Herrera F.* Human pose estimation for mitigating false negatives in weapon detection in video-surveillance // Neurocomputing. 2022. Vol. 489. P. 488–503.

5. *Yoo H.R., Lee B.H.* An OpenPose-based child abuse decision system using surveillance video // J. Korea Inst. Inf. Commun. Eng. 2019. Vol. 23. P. 282–290.

6. *Park J.H., Song K., Kim Y.-S.* A Kidnapping Detection Using Human Pose Estimation in Intelligent Video Surveillance Systems // J. Korea Soc. Comput. Inf. 2018. Vol. 23. P. 9–16.

7. *Chang Y.J., Chen S.F., Huang J.D.* A Kinect-based system for physical rehabilitation: A pilot study for young adults with motor disabilities // Res. Dev. Disabil. 2011. Vol. 32. P. 2566–2570.

8. *Hassan H.A., Abdallah B.H., Abdallah A.A., Abdel-Aal R.O., Numan R.R., Darwish A.K., El-Behaidy W.H.* Automatic Feedback for Physiotherapy Exercises Based On PoseNet // FCAI-Inform. Bull. 2020. Vol. 2. P. 10–14.

9. *Shapoval S., García Zapirain B., Mendez Zorrilla A., Mugueta-Aguinaga I.* Biofeedback applied to interactive serious games to monitor frailty in an elderly population // Appl. Sci. 2021. Vol. 11. P. 3502.

10. *Chua J., Ong L.Y., Leow M.C.* Telehealth using PoseNet-based system for in-home rehabilitation // Future Internet. 2021. Vol. 13. P. 173.

11. *Kim W., Sung J., Saakes D., Huang C., Xiong S.* Ergonomic postural assessment using a new open-source human pose estimation technology (OpenPose) // Int. J. Ind. Ergon. 2021. Vol. 84. P. 103164.

12. *Jawale C.D., Joshi K.A., Gogate S.K., Badgujar C.* Elcare: Elderly Care with Fall Detection // J. Phys. Conf. Ser. 2022. Vol. 2273. P. 012019.

13. *Kapoor R., Jaiswal A., Makedon F.* Light-weight seated posture guidance system with machine learning and computer vision // Proceedings of the 15th International Conference on Pervasive Technologies Related to Assistive Environments.

2022. P. 595–600.

14. *Landi H.* Google, ProMedica Team up with Include Health to Tap into Growing Virtual MSK Market // *Fierce Healthcare*.

URL: <https://www.fiercehealthcare.com/tech/google-promedica-team-up-include-health-to-tap-into-virtual-msk-market>

15. *Chen W., Jiang Z., Guo H., Ni X.* Fall detection based on key points of human-skeleton using OpenPose // *Symmetry*. 2020. Vol. 12. P. 744.

16. *Zou J., Li B., Wang L., Li Y., Li X., Lei R., Sun S.* Intelligent fitness trainer system based on human pose estimation // *Proceedings of the International Conference On Signal and Information Processing, Networking and Computers*. 2018. P. 593–599.

17. *Suda S., Makino Y., Shinoda H.* Prediction of volleyball trajectory using skeletal motions of setter player // *Proceedings of the 10th Augmented Human International Conference*. 2019. P. 1–8.

18. *Wang J., Qiu K., Peng H., Fu J., Zhu J.* AI coach: Deep human pose estimation and analysis for personalized athletic training assistance // *Proceedings of the 27th ACM International Conference on Multimedia*. 2019. P. 374–382.

19. *Jeon H., Yoon Y., Kim D.* Lightweight 2D human pose estimation for fitness coaching system // *Proceedings of the 2021 36th International Technical Conference on Circuits/Systems, Computers and Communications (ITC-CSCC)*. 2021. P. 1–4.

20. *Li Y.C., Chang C.T., Cheng C.C., Huang Y.L.* Baseball Swing Pose Estimation Using OpenPose // *Proceedings of the 2021 IEEE International Conference on Robotics, Automation and Artificial Intelligence (RAAI)*. 2021. P. 6–9.

21. *Park H.J., Baek J.W., Kim J.H.* Imagery based Parametric Classification of Correct and Incorrect Motion for Push-up Counter Using OpenPose // *Proceedings of the 2020 IEEE 16th International Conference on Automation Science and Engineering (CASE)*. 2020. P. 1389–1394.

22. *Nguyen H.T.P., Woo Y., Huynh N.N., Jeong H.* Scoring of Human Body-Balance Ability on Wobble Board Based on the Geometric Solution // *Appl. Sci*. 2022. Vol. 12. P. 5967.

23. *Patil A., Rao D., Utturwar K., Shelke T., Sarda E.* Body Posture Detection and Motion Tracking using AI for Medical Exercises and Recommendation System // ITM Web Conf. 2022. Vol. 44. P. 03043.
24. *Devanandan M., Rasaratnam V., Anbalagan M.K., Asokan N., Panchendrarajan R., Tharmaseelan J.* Cricket Shot Image Classification Using Random Forest // Proceedings of the 2021 3rd International Conference on Advancements in Computing (ICAC). 2021. P. 425–430.
25. *Joseph R., Ayyappan M., Shetty T., Gaonkar G., Nagpal A.* BeFit—A Real-Time Workout Analyzer // Proceedings of the Sentimental Analysis and Deep Learning. 2022. P. 303–318.
26. *Mahendran N.* Deep Learning for Fitness // arXiv. 2021. arXiv:2109.01376.
27. *Agarwal S., Gupta M., Khandelwal S., Jain P., Aggarwal A., Singh D., Mishra V.K.* FitMe: A Fitness Application for Accurate Pose Estimation Using Deep Learning // Proceedings of the 2021 2nd International Conference on Secure Cyber Computing and Communications (ICSCCC). 2021. P. 232–237.
28. *Nakai M., Tsunoda Y., Hayashi H., Murakoshi H.* Prediction of basketball free throw shooting by OpenPose // Proceedings of the JSAI International Symposium on Artificial Intelligence. 2018. P. 435–446.
29. *Zheng C., Wu W., Chen C., Yang T., Zhu S., Shen J., Kehtarnavaz N., Shah M.* Deep learning-based human pose estimation: A survey // arXiv. 2020, arXiv:2012.13392.
30. *Papandreou G., Zhu T., Kanazawa N., Toshev A., Thompson J., Bregler C., Murphy K.* Towards accurate multi-person pose estimation in the wild // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu. 2017. P. 4903–4911.
31. *Jo B., Kim S.* Comparative Analysis of OpenPose, PoseNet, and MoveNet Models for Pose Estimation in Mobile Devices // Trait. du Signal. 2022. Vol. 39. P. 119–124.

32. Cao Z., Simon T., Wei S.E., Sheikh Y. Realtime multi-person 2d pose estimation using part affinity fields // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017. P. 7291–7299.
33. Erramchetty S.K. et al. Human pose estimation using blaze pose // AIP Conference Proceedings. 2024. Vol. 2971. No. 1.
34. Gadhiya R., Kalani N. Analysis of Deep Learning Based Pose Estimation Techniques for Locating Landmarks on Human Body Parts // Proceedings of the 2021 International Conference on Circuits, Controls and Communications (CCUBE). 2021. P. 1–4.
35. Lin T.Y., Maire M., Belongie S., Hays J., Perona P., Ramanan D., Dollár P., Zitnick C.L. Microsoft coco: Common objects in context // Proceedings of the European Conference on Computer Vision. 2014. P. 740–755.
36. Andriluka M., Pishchulin L., Gehler P., Schiele B. 2D human pose estimation: New benchmark and state of the art analysis // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2014. P. 3686–3693.
37. Toshev A., Szegedy C. DeepPose: Human pose estimation via deep neural networks // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2014. P. 1653–1660.
38. Chung J.-L., Ong L.-Y., Leow M.-C. Comparative Analysis of Skeleton-Based Human Pose Estimation // Future Internet. 2022. Vol. 14, No. 12. P. 380.
<https://doi.org/10.3390/fi14120380>
39. Ahmedt-Aristizabal D., Nguyen K., Denman S., Sridharan S., Dionisio S., Fookes C. Deep motion analysis for epileptic seizure classification // Proceedings of the 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. 2018. P. 3578–3581.
40. Ishaq A., Haq I.U. Real-time single-person pose estimation on low-power embedded devices // 2019 International Conference on Frontiers of Information Technology (FIT).
41. Cao Z., Simon T., Wei S.-E., Sheikh Y. Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields // Proceedings of the IEEE Conference on Computer

Vision and Pattern Recognition (CVPR). 2017. P. 7291–7299.

42. Papandreou G., Zhu T., Chen L.-C., Gidaris S., Tompson J., Murphy K. PersonLab: Person Pose Estimation and Instance Segmentation with a Bottom-Up, Part-Based, Geometric Embedding Model // Proceedings of the European Conference on Computer Vision (ECCV). 2018. P. 269–286.

43. TensorFlow Blog. MoveNet: Ultra fast and accurate pose detection model. 2021.

URL: <https://blog.tensorflow.org/2021/05/movenet-and-posture-analysis.html>

44. Bazarevsky V., Zhang F. BlazePose: On-Device Real-Time Body Pose Tracking. Google AI Blog. 2020.

URL: <https://ai.googleblog.com/2020/08/on-device-real-time-body-pose-tracking.html>

45. Gabdullina D.R., Zykov E.Yu., Kugurakova V.V. Using OpenCV for real-time image recognition through augmented reality devices // The 2023 International Conference on Artificial Life and Robotics (ICAROB2023). 2023. P. 868–874.

46. Gabdullina D.R., Zykov E.Yu., Kugurakova V.V. Using OpenCV for augmented reality applications // Journal of Advances in Artificial Life Robotics. 2023. Vol. 3. Iss. 3(3). P. 179–184.

47. Bakirov A.R. et al. Using Microsoft Kinect in the rehabilitation games development // Russian Digital Libraries Journal. 2016. V. 19. No. P. 521–537.

48. You Y., Zhao Y. A human pose estimation algorithm based on the integration of improved convolutional neural networks and multi-level graph structure constrained model // Pers. Ubiquitous Comput. 2019. Vol. 23. P. 607–616.

СВЕДЕНИЯ ОБ АВТОРЕ



ЯРКО Егор Игоревич – аспирант кафедры индустрии разработки игр института ИТИС. Сфера научных интересов – приложения дополненной реальности с детектированием позы человека в реальном времени.

Egor Igorevich YARKO – postgraduate student of the Chair of Game Development Industry at ITIS Institute. His research interests include augmented reality applications with real-time human pose detection.

email: yarkoeg@gmail.com

ORCID: 0009-0009-0502-1066

Материал поступил в редакцию 28 января 2025 года