ТЕХНИЧЕСКИЕ НАУКИ

УДК 004.932.2

Обзор методов автоматического определения позы человека по изображениям

Д.В. Атоян
Донской государственный технический университет, г. Ростов-на-Дону, Российская Федерация

Аннотация
В настоящее время технологии определения позы человека на основе анализа изображений находят широкое применение во многих сферах деятельности. Задача детектирования положения тела заключается в обработке визуальных данных с использованием различных методов для последующего построения модели человеческой фигуры. В современной практике принято различать двумерные и трехмерные модели, а также классифицировать их по структуре на скелетные, плоские и объемные типы представления. Данная статья посвящена подробному обзору существующих аналитических методов и архитектур нейронных сетей, предназначенных для оценки позы человека по цифровому изображению, а также рассмотрению специализированных наборов данных и метрик, применяемых в процессе обучения моделей.

Ключевые слова: определение позы человека, искусственный интеллект, глубокое обучение, компьютерное зрение

Для цитирования. Атоян Д.В. Обзор методов автоматического определения позы человека по изображениям. *Молодой исследователь Дона*. 2026;11(4):54–58.

An Overview of the Image-Based Human Pose Estimation Technologies

David V. Atoyan
Don State Technical University, Rostov-on-Don, Russian Federation

Abstract
Currently, image-based human pose estimation technologies are widely used in many fields of life. The problem of body pose detection implies processing visual data using various methods, which eventually results in building a model of a human body. At present, two-dimensional and three-dimensional human body models are commonly distinguished, and body modeling is classified into skeleton-based, planar (contour-based) and volumetric types. The article provides a detailed overview of existing analytical methods and neural network architectures designated for human pose estimation based on the digital images, as well as investigates the specialised datasets and metrics used in model training.

Keywords: human pose estimation, artificial intelligence, deep learning, computer vision

For Citation. Atoyan DV. An Overview of the Image-Based Human Pose Estimation Technologies. *Young Researcher of Don*. 2026;11(4):54–58.

Введение. Автоматизированное определение позы человека на основе видеопотока и статических изображений представляет собой технологию, позволяющую реконструировать модель человеческого тела по заданному визуальному ряду. Исследование и последующее внедрение подобных решений в различных отраслях характеризуются высокой степенью актуальности.

Цель исследования — проведение аналитического обзора методов оценки положения тела по видеоданным и изображениям, рассмотрение существующих наборов данных, а также изучение метрик, применяемых для обучения моделей.

Основная часть. В современной практике обработки и анализа визуальной информации задача детекции позы человека занимает приоритетное место наряду с обнаружением, классификацией, сегментацией и трекингом объектов. Процесс оценки состояния фигуры заключается в исследовании входного кадра различными аналитическими методами для последующей генерации цифровой модели обнаруженных субъектов. Данные технологии находят применение в следующих областях:

https://mid-journal.ru

© Атоян Д.В., 2026

– обеспечение общественной безопасности — в этой сфере методы используются для идентификации противоправных, несанкционированных или аномальных действий. Во многих государствах активно развиваются интеллектуальные системы поддержки правопорядка, функционирующие на базе анализа данных с камер наружного наблюдения;

– медицина — технологии востребованы в процессах реабилитации и терапии для прецизионного анализа двигательной активности пациентов;

– кинематограф и мультипликация — методики применяются для цифрового моделирования актеров и захвата движений (motion capture).

– робототехника и человеко-машинное взаимодействие — в данной отрасли алгоритмы обеспечивают коммуникацию с роботизированными комплексами, оснащенными системами компьютерного зрения.

Доступные методы классифицируются на алгоритмы построения двумерных и трехмерных моделей, а также на решения для детекции одного или нескольких субъектов. При этом структурно модели подразделяются на: скелетные типы, основанные на ключевых точках; контурные модели, базирующиеся на простых геометрических примитивах (прямоугольники, эллипсоиды); объемные модели, представленные в виде трехмерных полигональных сеток [1]. Визуализация данных подходов приведена на рис. 1.

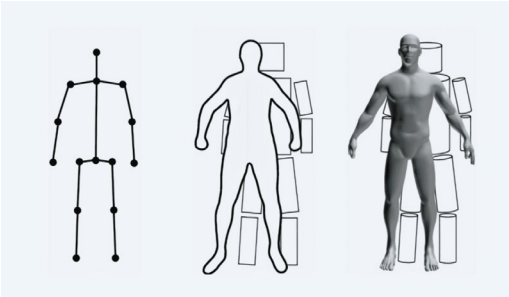

Рис. 1. Существующие модели тела человека [1]

Одним из основополагающих методов идентификации пространственного положения тела является концепция Pictorial Structure Framework (PSF), использующая алгоритм «случайный лес» для аппроксимации узловых точек [2]. Структура метода включает два компонента:

Discriminator — модуль, предназначенный для вероятностного моделирования нахождения конкретной части тела в определенной позиции на кадре.

Prior — блок, служащий для формирования распределения вероятностей пространственного расположения фигуры на основе данных, полученных на начальном этапе.

Главная задача PSF — описание тела как совокупности координат каждого элемента на входном изображении. В рамках алгоритма применяются нелинейные регрессоры, в частности — двухслойный регрессор. Однако традиционные методы демонстрируют недостаточно высокую точность. Подобные системы эффективно функционируют при условии четкой видимости всех элементов, но показывают низкую результативность при частичном или полном перекрытии (окклюзии) частей тела.

Методы на основе глубокого обучения. Прогресс в области глубоких нейронных сетей способствовал появлению более прецизионных способов оценки положения человека.

Модель DeepPose. Первым решением на базе сверточных нейросетей стала архитектура DeepPose, представленная в 2014 г. Система интегрирует кодировщик и регрессионный декодировщик. Первый компонент базируется на сети AlexNet, включающей пять сверточных и три полносвязных слоя. Входные данные анализируются сетью с последующим вычислением координат суставов, соответствующих ключевым точкам.

К преимуществам данного метода следует отнести: возможность извлечения контекста каждой точки за счет анализа полноразмерного изображения и относительную простоту реализации по сравнению с детекторами отдельных фрагментов. Схематически структура нейросети изображена на рис. 2 [3].

Рис. 2. Архитектура метода DeepPose [3]

При верификации на наборе данных LSP (Leeds Sports Pose) показатель корректно определенных элементов составил 61 %. К недостаткам алгоритма относятся: затруднения при обработке кадров с множеством людей, рост вычислительной сложности при увеличении числа субъектов и замедление работы в режиме реального времени.

Для минимизации указанных ограничений при анализе группы людей были разработаны два подхода: нисходящий и восходящий (рис. 3) [4]. Нисходящий метод сначала локализует каждого человека, а затем ищет части тела. Восходящий подход, напротив, детектирует отдельные фрагменты и объединяет их в модели. Нисходящие методы часто менее эффективны в условиях взаимного перекрытия тел.

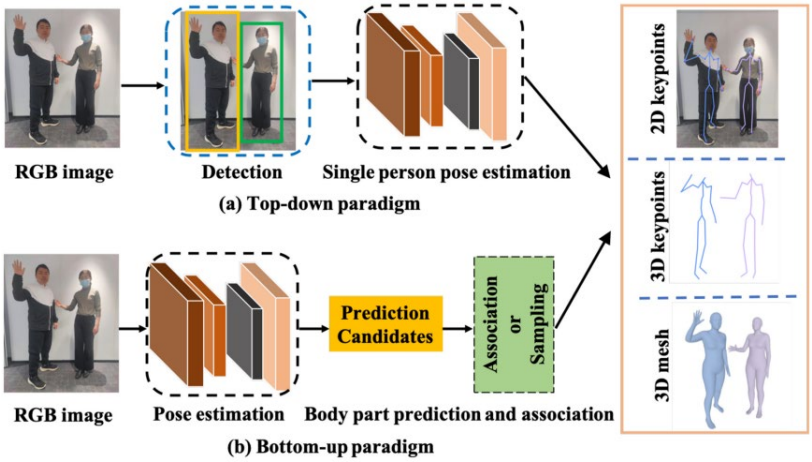

Рис. 3. Схематичное представление нисходящих и восходящих методов [4]

Модель RMPE. Система Regional Multi-Person Estimation (RMPE), разработанная в 2016 г., реализует нисходящий принцип. Она стала одной из первых успешных реализаций оценки позы для нескольких субъектов. Архитектура объединяет пространственную трансформирующую сеть для локализации областей с людьми и сеть для поиска ключевых точек. Процесс основан на последовательном извлечении ограничивающих прямоугольников (рис. 4) [5].

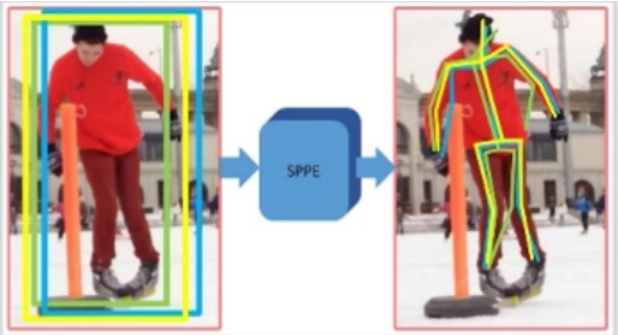

Рис. 4. Схема работы алгоритма RMPE [5]

Модель определения позы человека OpenPose. Модель определения позы человека OpenPose разработанная в 2019 году, реализует восходящий подход. Является первой моделью, способной решать задачу оценки позы нескольких человек в режиме реального времени. Архитектура данного метода состоит из глубокой нейронной сети VGG-19, используемой для извлечения признаков входного изображения, результаты работы которой передаются в две параллельные ветви сверточных нейронных сетей. Схематично архитектура данного метода показана на рисунке 5 [6].

Первая из двух параллельных ветвей создает карты достоверности (Confidence Maps) — набор матриц для каждой ключевой точки. Каждая матрица отражает достоверность появления определенной ключевой точки в определенном месте изображения. Большие значения означают более высокую вероятность появление опорной точки в данной части изображения.

Вторая ветвь алгоритма производит поиск связей между ключевыми точками на изображении с использованием метода PAS (Part Affinity Fields) — поля соответствия деталей. Поле соответствия представляет собой набор матриц, отражающих степень близости для найденных ранее опорных точек, и отвечает за кодирование двухмерного вектора положения и направления конечностей тела человека.

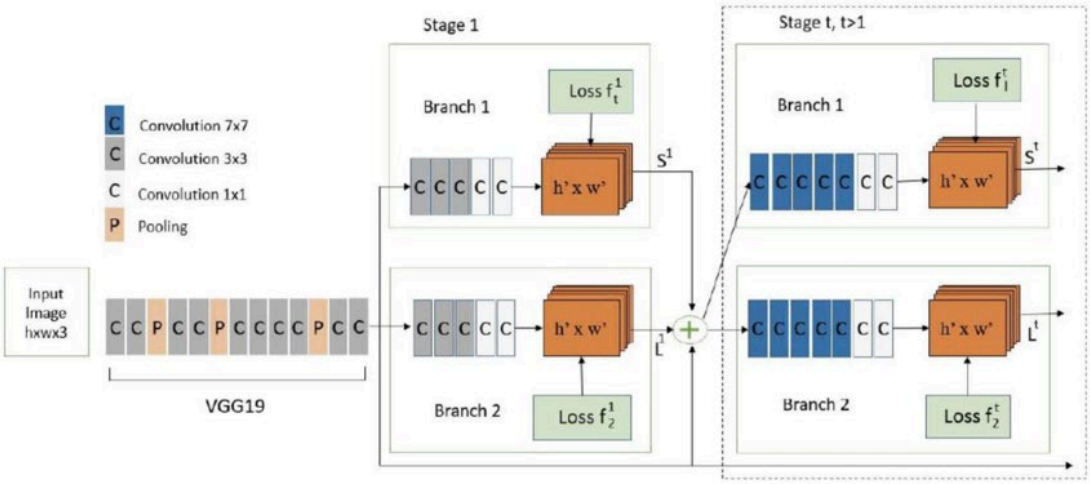

Рис. 5. Архитектура метода OpenPose [6]

Современные нейронные сети для определения позы человека. Одной из актуальных предварительно обученных нейронных сетей для определения ключевых точек тела человека является YOLO26-Pose, разработанная Ultralytics. Данная модель обучена на наборе данных COCO-Pose и поддерживает определение 17 ключевых точек тела человека. Модель представлена в пяти версиях (Nano, Small, Medium, Large, Extra Large), различающихся по точности, скорости работы, количеству параметров в нейронной сети и производительности [7]. Как и другие модели YOLO, YOLO26-Pose анализирует входное изображение всего за один проход по нейронной сети, в отличие от других моделей, реализующих двухэтапный подход. Отличительной особенностью данной модели является поддержка определения позы нескольких человек в режиме реального времени. Вместе с предварительно обученными моделями Ultralytics предоставляет удобный программный интерфейс для запуска, обучения и работы с моделями. Модель YOLO26-Pose является хорошим выбором для решения задачи определения позы нескольких человек при разработке различных прикладных программ. Другой современной моделью определения ключевых точек тела человека является MediaPipe Pose. Нейронная сеть MediaPipe Pose для определения ключевых точек тела человека разработана компанией Google в 2023 году [8]. Представлена в трех вариантах: Lite, оптимизированном для обеспечения максимальной скорости работы; Full — модель по умолчанию, сбалансированная для большинства приложений, и Heavy, показывающей лучшую производительность и требующей больших вычислительных ресурсов. Данная модель основана на анализе изображения с помощью сверточных нейронных сетей. Особенностью MediaPipe Pose является возможность построения различных моделей человеческого тела, например, построение маски лица, захват движения, идентификация эмоций, обнаружение движений кисти руки. В процессе работы MediaPipe Pose анализирует входное изображение, обнаруживает присутствующего человека и строит модель его тела. Выходными данными после обработки является массив из 33 ключевых точек, каждая из которых соответствует определенной части человеческого тела.

Наборы данных и метрики оценки играют важную роль при создании глубоких нейронных сетей. Наиболее популярными наборами для обучения являются следующие:

– Human3.6M — большой набор данных, содержащий аннотации в двумерном и трехмерном форматах, ориентирован на оценку позы одного человека, записан в лаборатории с использованием системы захвата движений профессиональными актерами.

– LSP (Leeds Sports Pose) — набор данных с записями спортивных тренировок, содержит аннотации 14 двумерных ключевых точек для более чем 2000 изображений.

– PoseTrack — набор для отслеживания позы нескольких человек на видео.

– MPII Human Pose — набор данных, содержащий более 25 тыс. изображений людей, выполняющих 410 различных действий.

– COCO-Pose — набор данных для двумерной оценки позы нескольких человек, содержит более 330 тыс. изображений, аннотированных 17 ключевыми точками.

– HumanEva — набор данных, содержащий двумерные и трехмерные аннотации для видеозаписей людей, выполняющих различные действий (бег, ходьба, жестикуляция и др.). Особенностью является запись движений человека с семи камер одновременно.

– В качестве метрик оценки нейронных сетей используются следующие:

– PCK (Percentage of Correct Keypoints) — процент правильно определенных ключевых точек тела.

– PCP (Percentage of Correct Parts) — процент правильно определенных частей тела.

– MPJPE (Mean Per Join Position Error) — среднее евклидово расстояние между исходными данными и прогнозируемыми, используется при трехмерной оценке позы человека.

Заключение. Оценка положения человека по изображениям является актуальной задачей науки и техники. Технологии оценки позы человека применяются во многих сферах от общественной безопасности до медицины. В данной статье рассмотрены методы обнаружения ключевых точек тела человека: DeepPose, RMPE, OpenPose. Описаны глубокие нейронные сети YOLO26 и MediaPipe Pose для определения ключевых точек тела человека, а также наборы данных и метрики для обучения нейронных сетей. Данная информация может быть использована при построении системы оценки общественной безопасности для анализа действий человека.

Список литературы

1. Barla N. *A Comprehensive Guide to Human Pose Estimation*. URL: https://v7labs.com/blog/human-pose-estimation-guide (дата обращения: 17.02.2026).

2. Felzenszwalb PF, Huttenlocker D. Pictorial Structures for Object Recognition. *International Journal of Computer Vision.* 2005;(61):55–79.

3. Ребрик Ю (Rebryk Y). *Детектирование частей тела с помощью глубоких нейронных сетей*. Habr.com. URL: https://habr.com/ru/companies/ods/articles/354850 (дата обращения: 19.02.2026).

4. Knap P. Human Modelling and Pose Estimation Overview. *arXiv*. 2024;arXiv:2406.19290v1. https://doi.org/10.48550/arXiv.2406.19290

5. Куцев Р. *Руководство по Human Pose Estimation.* Habr.com. URL: https://habr.com/ru/articles/687728 (Дата обращения: 26.02.2026).

6. Abdulghani MM, Ghazal MT, Salih ABM. Discover Human Poses Similarity and Action Recognition Based on Machine Learning. *Bulletin of Electrical Engineering and Informatics.* 2023;12(3):1570–1577. https://doi.org/10.11591/eei.v12i3.4930

7. *Документация Ultralytics YOLO26*. Ultralytics.com. URL: https://docs.ultralytics.com/ru/models/yolo26 (Дата обращения: 26.02.2026).

8. *YOLOv7 pose vs MediaPipe при оценке позы человека*. URL: https://habr.com/ru/articles/701268 (дата обращения: 27.02.2026).

Об авторах:

Давид Валерьевич Атоян, аспирант кафедры «Программное обеспечение вычислительной техники и автоматизированных систем» Донского государственного технического университета (344003, Российская Федерация, г. Ростов-на-Дону, пл. Гагарина, 1), atoyan.dawid@yandex.ru

Конфликт интересов: **авторы заявляют об отсутствии конфликта интересов.**

Все авторы прочитали и одобрили окончательный вариант рукописи.

About the Authors:

David V. Atoyan, Postgraduate Degree Student of the Computer Engineering and Automated Systems Software Department, Don State Technical University (1, Gagarin Sq., Rostov-on-Don, 344003, Russian Federation), atoyan.dawid@yandex.ru

Conflict of Interest Statement: **the authors declare no conflict of interest.**

All authors have read and approved the final manuscript.