

ОСОБЕННОСТИ ПАКЕТОВ ДЛЯ РАСПОЗНАВАНИЯ ПОХОДКИ

Дин АОДИ¹, Виктор АНОСОВ², канд. техн. наук, доц. О. В. НЕДЗЬВЕДЬ¹

¹(Белорусский государственный университет, г. Минск)

²(ГУ «РНПЦ травматологии и ортопедии», г. Минск, Беларусь)

Аннотация. *С быстрым развитием науки и техники, использование характеристик походки человека для определения медицинского диагноза может стать ранним обнаружением скрытых заболеваний. В этом эксперименте сравниваются YOLOv8, Openpose и MediaPipe для выбора оптимальной модели распознавания костей, модель GCN используется для анализа данных о походке, и, наконец, модель классификации используется для постановки медицинского диагноза. Классификационная модель для движения выполняет диагностику по четырем частям: здоровье, проблемы с суставами, мышечная слабость, неврологические недостатки, чтобы получить лучшую предсказательную эффективность.*

Ключевые слова: *анализ медицинского диагноза, YOLOv8, Openpose, MediaPipe.*

Распознавание походки как разновидность биометрии – это технология, позволяющая идентифицировать человека путем анализа характеристик его походки. Эта технология применяется во многих областях, таких как безопасность, медицина и т. д. Данная статья посвящена применению этой технологии в медицине для прогнозирования и эффективного лечения заболеваний опорно-двигательного аппарата и неврологических расстройств. с помощью YOLOv8, Openpose и MediaPipe для идентификации ключевых узлов человеческого тела [1], на видео путем преобразования скелетных координат, сохранения данных о движении суставов человеческого скелета и использования классификационных моделей для выявления возможных заболеваний. Целью диагностики походки [2] является поиск и извлечение признаков различий между людьми при одинаковом поведении при ходьбе для автоматической идентификации [1].

Для обучения и тестирования использовался публичный эталонный набор данных, который был создан Rueangsiraraketal [5]. Определение ключевых точек в YOLOv8 – это модель "верх-низ", которая определяет целевой кадр, прежде чем предсказать ключевые точки в целевом кадре. Параметры модели YOLOv8 наилучшим образом соответствуют специфическим требованиям задачи оценки позы. Это включает в себя настройку различных гиперпараметров, скорости обучения, размера набора данных, количества периодов и настроек оптимизатора. Скорость обучения контролирует размер шага при оптимизации градиентного спуска, и нахождение оптимального значения является критическим для сходимости модели [4].

Точность весов снижается с помощью квантования модели YOLOv8, что обычно приводит к значительному ускорению при минимальной потере точности. Во-вторых, с помощью TensorRT или OpenVINO модель может быть дополнительно оптимизирована под конкретное оборудование, что приводит к снижению задержки [6]. Кроме того, можно использовать пакетную обработку для одновременной обработки нескольких кадров, что позволяет эффективно использовать ресурсы GPU и увеличить пропускную способность. Реализация этих стратегий позволяет значительно повысить производительность оценки позы YOLOv8, и на рисунке 1 показаны результаты работы YOLOv8 для распознавания скелета. В данном скелете для диагностики ходьбы используется только шесть точек. Не учитывается положение стопы, что является важным параметром при диагностике.

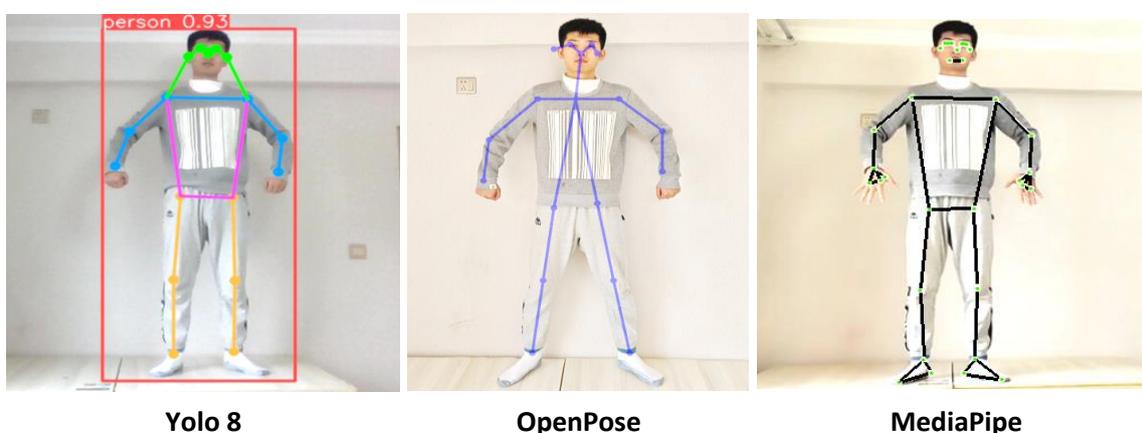


Рисунок 1. – Различия построения скелетов в пакетах Yolo 8, OpenPose, MediaPipe

Пакет **OpenPose** не обнаруживает цель. OpenPose способна детектировать позы сразу нескольких людей на изображении. Скелеты, построенные в этом пакете, не сопровождают одновременно несколько фигур человека. OpenPose получает положение ключевых точек скелета на основе построения тепловой карты и поля сродства части (PAFs) для частей тела во входном изображении. Каждая тепловая карта показывает вероятность, что конкретный тип части тела расположен на уровне каждого пикселя в изображении. PAFs являются векторными полями, которые указывают, соединятся ли две части тела. Для каждого заданного типа соединения части тела, такого как шея к левому плечу, существует два PAFs, которые показывают x-и y-компонент векторного поля между экземплярами частей тела. Таким образом определяются 18 карт признаков для 18 ключевых точек [9].

Пакет **MediaPipe Pose** способен извлечь 33 ключевые точки. Функция определения позы mediapipe берет части человеческого тела и разбивает их на 32 точки. Она обучена на двух наборах глубоких нейронных сетей: BlazePose на основе модели GHUM и AlphaPose на основе модели ResNet50. MediaPipe используется для полного определения скелета человека, получая 33 точки позиционирования

из модели GHUM[8]. Данные точки учитываю положения стоп, что является важным фактором. Однако, детектирование не всегда происходит на достаточно качественном уровне.

Для отрисовки ключевых точек был использован метод `mpDraw.draw_landmarks` при рисовании ключевых точек человеческого тела получает всю информацию о ключевых точках; если параметр не передан в метод `POSE_CONNECTIONS`, то соединительные линии между ключевыми точками не будут нарисованы.

Результаты производительности алгоритмов в пакетах Yolo 8, OpenPose, MediaPipe показаны в таблице. Следует отметить, что когда входные изображения одинаковы, при обработке последовательности из 640 изображений, YOLOv8 обрабатывает изображения быстрее, чтобы следовать модели анализа походки ближе к реальному времени, поэтому удобнее использовать именно эту модель [3].

Таблица. – Сравнение скорости модели

| Модель | Fps | Size |
|-----------|------|------|
| YOLOv8 | 0.82 | 640 |
| OpenPose | 1.6 | 640 |
| MediaPipe | 29.2 | 640 |

Из рисунка 1 видно, что модель YOLOv8 использует 17 ключевых узлов для описания скелета человека, модель OpenPose использует 18 ключевых узлов для описания скелета человека, а модель MediaPipe использует 33 ключевые точки для описания скелета человека. Модель YOLOv8 является самой быстрой – 0,82 кадра в секунду. Модель MediaPipe добавляет слишком много обнаружений ключевых точек на лице и ступнях, которые не всегда показывают адекватный угол. Поэтому YOLOv8 является наиболее подходящей моделью для обнаружения этой ключевой точки.

Видеораспознавание походки использует силуэтные изображения для представления человека в каждом кадре. Для оценки положения человека и распознавания походки выстраивается граф движения. Таким образом, объединение скелетных поз можно выполнить посредством графовых конволюционных сетей GCN[3].

Анализ походки выполняется на основе построения графиков функций углов ребер по изменению положения ключевых точек. Они позволяют определить, какой период пошагового цикла соответствует данному кадру, а также изменения функций в зависимости от периода пошагового цикла. На основе этой информации принимаются решения о поэтапной сегментации шага как на рисунке 2.

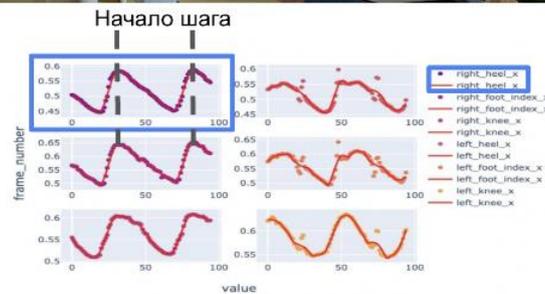


Рисунок 2. – Фрагмент программы анализа шага для начала шага/начало периода первой двойной опоры/окончание периода переноса

Использование модели Yolo v8 для создания скелетного каркаса а объединением ключевых точек посредством модели GCN для извлечения особенностей походки реализован анализ цикла шага, который используется для диагностики походки. Результаты в модель классификации ключевых точек диагностики основаны на анализе углов между точками что значительно снижает количество ошибок диагностики. Во всем процессе анализа, используется три этапа, позволяющие прийти на правильность решения, что позволяет упростить разработку и анализ видео для тестирования походки при заболевании, сократить использование медицинских ресурсов и улучшить качество диагностики заболевания.

ЛИТЕРАТУРА

1. I. Akhterand, M. J. Black Pose-Conditioned Joint Angle Limits for 3D Human Pose Reconstruction. In IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015, p. 1446–1455.
2. T. Alldieck, M. Magnor, W. Xu, C. Theobalt, and G. Pons-Moll. Detailed Human Avatars from Monocular Video. In 3D Vision(3DV), 2018. Sixth International Conference on, 2018.
3. T. Alldieck, M. Magnor, W. Xu, C. Theobalt and G. Pons-Moll. Video Based Reconstruction of 3D People Models, IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 2018, pp. 8387-8397, doi: 10.1109/CVPR.2018.00875.
4. M. Andriluka, L. Pishchulin, P. Gehler and B. Schiele, "2D Human Pose Estimation: New Benchmark and State of the Art Analysis," 2014 IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 2014, pp. 3686-3693, doi: 10.1109/CVPR.2014.471.
5. Rueangsirarak, W., Zhang, J., Aslam, N., Ho, E.S.L., Shum, H.P.H.: Automatic Musculoskeletal and Neurological Disorder Diagnosis With Relative Joint Displacement From Human Gait. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 2018, 26(12), p. 2387–2396, doi: 10.1109/tnsre.2018.2880871.

6. Sonata I, Heryadi Y, Lukas L, et al. Autonomous car using CNN deep learning algorithm[C] // Journal of Physics: Conference Series. IOP Publishing, 2021, 1869(1): 012071.
7. Bai H, Wu J, Liu C. Motion and haar-like features based vehicle detection[C] // 2006 12th International Multi-Media Modelling Conference. IEEE, 2006: 4 p.
8. Mo G, Zhang Y, Zhang S, et al. A method of vehicle detection based on SIFT features and boosting classifier[J]. Journal of Convergence Information Technology, 2012, 7(12): 328–334.
9. Tian S, Bhattacharya U, Lu S, et al. Multilingual scene character recognition with co-occurrence of histogram of oriented gradients[J]. Pattern Recognition, 2016, 51: 125–134.
10. Tsai L W, Hsieh J W, Fan K C. Vehicle detection using normalized color and edge map[J]. IEEE transactions on Image Processing, 2007, 16(3): 850–864.