

УДК 004.89

АЛГОРИТМ РАСПОЗНАВАНИЯ ДЕЙСТВИЙ ЧЕЛОВЕКА В ВИДЕОПОТОКЕ С ПОМОЩЬЮ СКЕЛЕТНЫХ МОДЕЛЕЙ КОМПЬЮТЕРНОГО ЗРЕНИЯ

Шпилева А.А. (ИТМО)

Научный руководитель – кандидат технических наук, ассистент Ефимова В.А.
(ИТМО)

Введение. Задача распознавания действий человека в видеопотоке заключается в анализе последовательных кадров видео для идентификации и классификации различных действий людей с помощью их пространственных и временных характеристик. На данный момент данная задача востребована в первую очередь в промышленной среде, так как решает проблему автоматизации мониторинга производственных процессов и повышения безопасности труда.

В условиях существующих систем видеонаблюдения и анализа данных существует потребность в эффективных инструментах, способных точно идентифицировать некоторые действия работников и предупреждать о потенциально опасных ситуациях. На данный момент как на отечественном, так и на зарубежном рынке решения в этой области ограничены коммерческими продуктами, доступ к которым затруднен для широкого круга пользователей.

Основная часть. В рамках данной задачи действия человека тесно связаны с взаимодействием с окружающими предметами, что требует эффективного обнаружения и анализа этих объектов. Одним из современных и эффективных подходов к решению этой проблемы является отслеживание объектов через их обнаружение и последующую классификацию.

В данной работе предлагается использовать новейшую версию модели YOLO [1] для определения скелетных точек человека и модели DepthAnything для того, чтобы получить глубину скелетных точек, а также применять комбинацию моделей SAM [2] и DINO [3] для обнаружения целевых объектов независимо от условий окружающей среды. Завершающим этапом является обработка полученных данных с использованием метрических методов или глубоких нейронных сетей для точной классификации действий.

Выводы. Реализовано программное обеспечение, которое классифицирует действия человека в производственной среде.

Список использованных источников:

1. Rahima Khanam, Muhammad Hussain, “YOLOv11: An Overview of the Key Architectural Enhancements, arXiv preprint arXiv: 2410.17725v1 2024.
2. Alexander Kirillov, Eric Mintun, Nikhila Ravi, Hanzi Mao, Chloe Rolland, Laura Gustafson, Tete Xiao, Spencer Whitehead Alexander C. Berg Wan-Yen Lo Piotr Dollar, Ross Girshick, “Segment Anything”, arXiv preprint arXiv: arXiv:2304.02643v1 2023.
3. Tianhe Ren, Yihao Chen, Qing Jiang, Zhaoyang Zeng, Yuda Xiong, Wenlong Liu, Zhengyu Ma, Junyi Shen, Yuan Gao, Xiaoke Jiang, Xingyu Chen, Zhuheng Song, Yuhong Zhang, Hongjie Huang, Han Gao, Shilong Liu, Hao Zhang, Feng Li, Kent Yu, Lei Zhang, “DINO-X: A Unified Vision Model for Open-World Object Detection and Understanding”, arXiv: 2411.14347v2 2024