

**РАЗРАБОТКА УЛУЧШЕННОГО АЛГОРИТМА
ДЕТЕКТИРОВАНИЯ ОБЪЕКТОВ ПО
ВИДЕОПОСЛЕДОВАТЕЛЬНОСТИ**

Попов Алексей Павлович

Студент

Факультет ВМК МГУ имени М. В. Ломоносова, Москва, Россия

E-mail: alexey.popov@graphics.cs.msu.ru

Научный руководитель — Шахуро Владислав Игоревич

Детектирование объектов на изображении является одной из важнейших задач компьютерного зрения. Алгоритмы детектирования применяются во множестве прикладных задач: регистрация нарушителей правил дорожного движения, контроль за соблюдением правил техники безопасности, алгоритмы автономного движения различных транспортных средств, определение направления взгляда водителя.

Для детектирования объектов на изображениях используются различные алгоритмы детектирования, но в современных приложениях абсолютное большинство алгоритмов реализуются на основе нейронных сетей. Во многих задачах необходимо осуществлять детектирование объектов по последовательности кадров, что позволяет сделать дополнительное предположение о том, что целевые объекты на соседних кадрах находятся достаточно близко друг к другу в пространстве кадра. Применение детектора объектов по отдельным кадрам, в таких задачах, не позволяет извлекать всю полезную информацию, поэтому для этих задач используют алгоритмы детектирования объектов по видеопоследовательности.

Алгоритмы детектирования объектов по видеопоследовательности делятся на две большие группы по предмету анализа: анализ низкоуровневых признаков изображений (внутреннее представление изображений в архитектуре нейронной сети), анализ высокоуровневых признаков изображений (результаты детектирования объектов на отдельных изображениях). В данной работе мы анализируем и предлагаем новый алгоритм из первой группы, а именно алгоритм, в котором анализ пространственно-временной информации производится за счет вычисления связей в нейросетевых представлениях для последовательных кадров.

В данной работе мы предлагаем несколько модификаций базового алгоритма YOLOV [1], применяющегося для детектирования объектов по видеопоследовательности, которые мы назвали YOLOV+v8 и

YOLOV+v8+TROI. В YOLOV+v8 базовая архитектура YOLOX [2] заменена на YOLOv8, а в YOLOV+v8+TROI кроме замены базовой архитектуры детектора применяется модифицированный механизм пространственно-временного внимания TROI [3]. Кроме того мы проводим сравнительный анализ предложенных нами детекторов с другими детекторами семейства YOLO [4] последних поколений. Мы сравниваем результаты работы предложенных алгоритмов с YOLOv7 [5], YOLOv8, YOLOX, а также с базовой реализацией алгоритма YOLOV, который является модификацией алгоритма YOLOX для эффективного применения в задаче детектирования по видео-последовательности. Для вычисления сравнительных метрик мы используем наборы данных OVIS [6] и YouTube-VOS [7].

В результате проведенных исследований мы показываем, что предложенные алгоритмы YOLOV+v8 и YOLOV+v8+TROI превосходят базовый алгоритм YOLOV по метрикам качества на выбранных наборах данных, при этом скорость работы алгоритмов возрастает, либо остается неизменной.

Литература

1. Shi Y., Wang N., Guo X. YOLOV: Making Still Image Object Detectors Great at Video Object Detection //arXiv preprint arXiv:2208.09686. – 2022.
2. Ge Z. et al. Yolox: Exceeding yolo series in 2021 //arXiv preprint arXiv:2107.08430. – 2021.
3. Gong T. et al. Temporal ROI align for video object recognition //Proceedings of the AAAI Conference on Artificial Intelligence. – 2021. – Т. 35. – №. 2. – С. 1442-1450.
4. Redmon J. et al. You only look once: Unified, real-time object detection //Proceedings of the IEEE conference on computer vision and pattern recognition. – 2016. – С. 779-788.
5. Wang C. Y., Bochkovskiy A., Liao H. Y. M. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors //arXiv preprint arXiv:2207.02696. – 2022.
6. Qi J. et al. Occluded video instance segmentation: Dataset and challenge. – 2021.
7. Xu N. et al. Youtube-vos: A large-scale video object segmentation benchmark //arXiv preprint arXiv:1809.03327. – 2018.