

НАБОРЫ ДАННЫХ ДЛЯ ОЦЕНКИ ЭФФЕКТИВНОСТИ МОДЕЛЕЙ ОБНАРУЖЕНИЯ ОБЪЕКТОВ НА ИЗОБРАЖЕНИИ

Верхов К.А.¹, ассистент, k.verkhov@bsuir.by

Ларькин А.Д.², ассистент, a.larkin@bsuir.by

2025

1. Белорусский государственный университет информатики и радиоэлектроники
2. Белорусский государственный университет информатики и радиоэлектроники

Ключевые слова: искусственный интеллект, компьютерное зрение, обнаружение объектов, датасеты, метрики эффективности.

Аннотация: Обнаружение объектов в компьютерном зрении представляет собой важную задачу, заключающуюся в нахождении и локализации объектов на изображениях и видео. Этот процесс включает определение присутствия объектов и точное указание их местоположения с использованием ограничивающих рамок. В статье рассматриваются различные датасеты, широко используемые для оценки алгоритмов обнаружения объектов, а также метрики для измерения их эффективности.

Обнаружение объектов в компьютерном зрении – это задача, направленная на нахождение и локализацию объектов на изображении или в видео. В процессе этой задачи алгоритмы не только определяют, присутствуют ли объекты, но и определяют их точное местоположение, обозначая их ограничивающими рамками [1]. Далее будут рассмотрены датасеты, которые чаще всего используются для оценки эффективности алгоритмов обнаружения объектов, а также метрики, применяемые для измерения производительности таких систем.

Датасет PASCAL-VOC является одним из старейших и наиболее известных наборов данных для задач обнаружения объектов. PASCAL-VOC включает более 11 000 изображений, в которых представлены 20 категорий объектов, таких как люди, машины, животные, бытовые предметы и другие [2]. Это делает датасет универсальным для тестирования алгоритмов, ориентированных на различные типы объектов в реальных условиях. Оценка качества детекции в PASCAL-VOC осуществляется с помощью средней точности. Средняя точность измеряется для каждой категории объекта, а затем вычисляется среднее значение по всем категориям. Учитываются как правильные, так и ошибочные ответы, минимизируя проблемы с ложными положительными результатами и пропущенными объектами. Также, эта метрика учитывает точность как в плане классификации, так и локализации объектов.

Датасет MS-COCO стал одним из самых популярных и широко используемых для задач обнаружения объектов и сегментации. MS-COCO

включает в себя более 450 000 изображений с 91 категорией объектов, таких как люди, транспортные средства, животные и другие [3]. Он также содержит разметку для объектов, расположенных в сложных и переполненных сценах. Для оценки качества применяется средняя точность.

Датасет VIS Drone был выпущен в 2018 году и содержит изображения и видео, снятые с дронов в различных городах Китая. Датасет включает в себя 263 видеосегмента и 10 209 аннотированных изображений. Общее количество аннотированных примеров составляет почти 2,5 миллиона. Для оценки производительности применяется стандартная метрика MSCOCO.

Датасет Open-Images включает в себя 1,9 миллиона изображений, распределенных по 600 категориям, и более 15 миллионов объектов, которые были аннотированы с использованием различных типов меток и объектов. Каждое изображение сопровождается метками классов, сегментацией объектов, а также аннотациями для визуальных отношений между объектами [4]. Одной из особенностей является то, что аннотированные ограничивающие рамки часто создаются вручную, что повышает точность и качество данных. Для предотвращения ошибок в виде ложных отрицательных результатов, из датасета исключаются группы классов, для которых отсутствуют аннотации.

Также, существуют специализированные датасеты, ориентированные на более узкие задачи. Например, датасеты для обнаружения лиц.

PASCAL FACE – включает в себя 851 изображение с 1335 лицами. В основном используется для проверки работы моделей для обнаружения лиц.

Fddb – включает в себя 2845 изображений с 5171 лицом. В основном используется для проверки эффективности работы алгоритмов обнаружения лиц.

Wider-Face – включает в себя 3203 изображения с 400 000 лиц. Разделен на три набора данных: 40 процентов изображений используется для обучения, 10 процентов для валидации, 50 процентов для проверки.

Список использованных источников

1. Edozie, E. Comprehensive review of recent developments in visual object detection based on deep learning / Edozie, E., Shuaibu, A.N., John, U.K. // Artificial Intelligence Review – Vol. 58, N. 227 – 2025.
2. Everingham, M. The pascal visual object classes (voc) challenge / Everingham, M., Van Gool, L., Williams, C. K., Winn, J., Zisserman, A. // International journal of computer vision – Vol. 88, N. 2 – 2010 – Pp. 303-338.
3. Dai, J. R-fcn: Object detection via region-based fully convolutional networks / Dai, J., Li, Y., He, K., Sun, J. // Advances in neural information processing systems – Vol. 29, N. 1 – 2016.
4. Kuznetsova, A. The open images dataset v4: Unified image classification, object detection, and visual relationship detection at scale / Kuznetsova, A., Rom, H., Alldrin, N., Uijlings, J., Krasin, I., Pont-Tuset, J., Kamali, S., Popov, S., Mallocci, M., Kolesnikov, A. and Duerig, T. // International journal of computer vision – Vol.128, N. 7 – 2020 – Pp. 1956-1981.