

МЕТОДЫ ОБНАРУЖЕНИЯ ОБЪЕКТОВ НА ИЗОБРАЖЕНИИ

Верхов К.А.

Белорусский государственный университет информатики и радиоэлектроники,
г. Минск, Республика Беларусь

Научный руководитель: Тонкович И.Н. – канд.хим.наук, доцент

Аннотация. Представлен аналитический обзор известных методов обнаружения объектов на изображении. Описаны их принципы работы. Выявлены преимущества и недостатки рассмотренных методов, а также предложены актуальные варианты использования в повседневной жизни.

Ключевые слова: компьютерное зрение, обработка изображения, обнаружение объектов, машинное обучение, глубокое обучение, нейронные сети

Введение. Обнаружение объектов – это технология, которая относится к области компьютерного зрения и цифровой обработке изображений. Ее задача – обнаружение на цифровом изображении или видео объектов определенного вида (живые существа, машины, здания).

Для каждого вида объектов существует набор специфических особенностей, с помощью которых можно классифицировать объект. Например, для идентификации лица такими особенностями будут глаза, губы, нос, цвет кожи, расстояние между глазами. Эти же специфические особенности используются и для обнаружения объектов. Или, еще один пример, с развитием технологий умных городов требуется наличие быстрых и эффективных систем для распознавания объектов, чтобы минимизировать требования к аппаратному обеспечению данных технологий, а также повысить точность результатов их работы. Области применения алгоритмов обнаружения объектов на изображении разнообразны: медицинская помощь, розничная торговля, охранные системы, идентификация личности, виртуальные помощники и многое другое [1].

Обнаружение объектов на изображении является актуальной задачей в компьютерном зрении. Решение данных задач требует всестороннего анализа существующих методов обнаружения объектов на изображении.

Основная часть. Методы обнаружения объектов, как правило, основаны либо на машинном обучении, либо на глубоком обучении.

Для методов, основанных на глубоком обучении, свойственно использование сверточных нейронных сетей, которые позволяют осуществлять обнаружение объекта без использования списка специфических особенностей данного объекта.

Примеры методов с использованием глубокого обучения:

- Region Proposals (с использованием различных региональных сверточных нейронных сетей: R-CNN, Fast R-CNN, Faster R-CNN, cascade R-CNN);
- Single-Shot MultiBox Detector;
- You Only Look Once;
- RetinaNet;
- Single-Shot Refinement Neural Network for Object Detection (RefineDet);
- Deformable convolutional networks.

Данные методы базируются на сверточных нейронных сетях, основной идеей которых является чередование сверточных и субдискретизирующих слоев, а также наличие операции свертки, в процессе которой на матрицу свертки поэлементно умножается каждый фрагмент изображения, суммируется и записывается в соответствующую позицию в выходном изображении.

В методе *Region Proposals* используются региональные сверточные нейронные сети – разновидность сверточных нейронных сетей, основной целью которых является выделение на изображении областей интереса, и последующее определение и классификация объектов в каждой из областей.

Метод *You Only Look Once (YOLO)* менее точен, чем региональные сверточные нейронные сети, но при этом значительно быстрее, что позволяет обнаруживать объекты в реальном времени. Суть данного метода заключается в первоначальном разделении изображения на сетку ячеек. Каждая ячейка отвечает за расположение области объекта на изображении, если центр данной области находится в пределах ячейки. Для каждой области определяются координаты x и y , ширина и высота области, а также коэффициент уверенности, который показывает вероятность наличия в данной области какого-либо объекта. Кроме того, каждая клетка определяет класс объекта в области, которая относится к этой клетке [2].

Метод *Single-Shot MultiBox Detector* представлен двумя компонентами: нейронной сетью для классификации изображений и сверточного слоя для обнаружения и классификации объектов на изображении. Данный метод, как и YOLO, делит изображение на сетку, но также оперирует понятием якорной области. Для каждой ячейки сетки можно назначить несколько якорных областей. Каждая из них предопределена и отвечает за размер и форму объекта внутри ячейки [3]. Преимуществом данного метода, как и метода YOLO, является скорость выполнения и возможность обнаружения объекта за один прогон изображения, в отличие от региональных сверточных нейронных сетей, которым необходимо пройти два этапа (выделение возможного региона с объектом и собственно обнаружение объекта в данном регионе изображения).

Метод *RetinaNet* так же является одноэтапным методом обнаружения объекта. Он включает в себя пирамиду признаков Feature Pyramid Network и функцию ошибки для обучения нейронных сетей Focal Loss.

Feature Pyramid Network состоит из трех частей: восходящего пути, нисходящего пути и боковых соединений. Восходящий путь похож на пирамиду и выглядит как последовательность слоев сверточной нейронной сети с уменьшающейся размерностью. Таким образом, нижние слои имеют большее разрешение, но меньшее семантическое значение; верхние же слои – наоборот. Данная часть уязвима к шуму на изображении, так как из-за зашумленности может потеряться информация об объекте.

В нисходящем пути размерности слоев соответствуют размерности слоев в восходящем пути, но из-за движения вниз по слоям происходит увеличение карты признаков с помощью метода ближайшего соседа.

Боковые соединения нужны для того, чтобы убрать затухание сигналов в процессе движения по слоям, совмещая семантически важную информацию, полученную к концу первой пирамиды и более детальную информацию, полученную в ней ранее.

Focal Loss, как правило, используется для решения проблемы дисбаланса классов на основе одноступенчатых моделей обнаружения объектов, т.е. данная функция необходима для исправления проблемы перекрытия объекта с меньшей якорной областью объектом с большей якорной областью. Из-за данной проблемы метод может не обнаружить маленький объект на большом фоне. Focal Loss на каждом уровне пирамиды оставляет только несколько якорных областей, как искомый объект. Все остальные области будут считаться объектами на заднем фоне [4].

Метод RetinaNet хорошо зарекомендовал себя в обнаружении объектов на изображениях со спутников и из воздуха.

Для методов, основанных на машинном обучении, сначала требуется определить особенности объекта, перед тем как классифицировать его. Далее, с помощью техник схожих с методом опорных векторов, уже можно осуществлять классификацию объекта.

Примеры методов с использованием машинного обучения:

- алгоритм Виолы-Джонса, основанный на признаках Хаара;
- масштабно-инвариантная трансформация признаков;
- гистограмма направленных градиентов.

Алгоритм Виолы-Джонса включает четыре стадии:

- выделение признаков Хаара на изображении;
- создание интегрального изображения (для каждого пикселя записывается сумма пикселей выше и левее его);
- обучение с использованием алгоритма Adaptive Boost (алгоритм машинного обучения для улучшения алгоритмов классификации);

- каскадная классификация.

Признаки, которые алгоритм ищет, включают в себя сумму пикселей внутри определенной прямоугольной области. Но в отличие от признаков Хаара алгоритм использует больше одной области, тем самым делая признаки более комплексными. Преимущественно данный алгоритм используется для обнаружения лиц на изображениях и видео.

Алгоритм масштабно-инвариантной трансформации признаков основан на выделении ключевых признаков (точек) объекта из набора контрольных изображений и последующем хранении данных признаков в базе данных. Распознавание объекта на новом изображении происходит за счет сравнения признаков на новом изображении с признаками, хранящимися в базе данных. Совпадение признаков основано на евклидовом расстоянии между векторами признаков. Данный метод может использоваться для трекинга, распознавания жестов, сшивания изображений и трехмерного моделирования.

Алгоритм гистограммы направленных градиентов основан на подсчете случаев определенного направления градиентов на выделенном участке изображения. Данный алгоритм схож с алгоритмом масштабно-инвариантной трансформации признаков, но вычисляется на плотной сетке равномерно распределенных клеток и для повышения точности использует нормализацию перекрывающегося локального контраста. Алгоритм основан на возможности описания внешнего вида и формы объекта с помощью распределения градиентов интенсивности. Преимуществом данного алгоритма является стойкость к геометрическим и фотометрическим изменениям изображения. Благодаря этому он хорошо подходит для обнаружения людей на изображениях.

Заключение. Основное преимущество методов, основанных на машинном обучении, – отсутствие необходимости предварительного выделения признаков искомого объектов. Такие методы используют преимущественно сверточные нейронные сети и обучаются в ходе своей работы. Для методов, основанных на глубоком обучении, требуется предварительно выделить признаки искомого объекта, с которыми алгоритм в последующем сравнивает признаки, выделенные у нового объекта. Таким образом, методы с использованием глубокого обучения лучше использовать в случаях, когда искомым предметом небольшое количество или их признаки можно конкретно определить. Методы с использованием машинного обучения подойдут для ситуаций, когда искомые объекты очень разнообразны, либо они ещё не определены.

Список литературы

1. Верхов, К.А. Обнаружение объектов на изображении с использованием машинного обучения/ К.А. Верхов // Новые информационные технологии в научных исследованиях: материалы XXV Юбилейной Всероссийской научно-технической конференции студентов, молодых ученых и специалистов; Рязань: ИП Коняхин А.В. (Book Jet), 2020 – С. 226-227.
2. A Gentle Introduction to Object Recognition with Deep Learning [Электронный ресурс]. – Режим доступа: <https://machinelearningmastery.com/object-recognition-with-deep-learning/>. – Дата доступа: 25.03.2021.
3. How single-shot detector (SSD) works? [Электронный ресурс]. – Режим доступа: <https://developers.arcgis.com/python/guide/how-ssd-works/>. – Дата доступа: 25.03.2021.
4. How RetinaNet works? [Электронный ресурс]. – Режим доступа: <https://developers.arcgis.com/python/guide/how-retinanet-works/>. – Дата доступа: 25.03.2021.

UDC 621.3.049.77–048.24:537.2

METHODS FOR OBJECTS DETECTION IN THE IMAGE

Verkhov K.A.

Belarusian State University of Informatics and Radioelectronics, Minsk, Republic of Belarus

Tonkovich I.N. – PhD in Chemistry, associate professor

Annotation. An analytical review of the known methods for detecting objects in the image is presented. Their operating principles are described. The advantages and disadvantages of the considered methods are revealed, as well as actual options for use in everyday life are proposed.

Keywords: computer vision, image processing, objects detection, machine learning, deep learning, neural networks