

Поиск объектов на изображениях с использованием TensorFlow Object Detection

Е. Ю. Митрофанова, email: mitrofanova@sc.vsu.ru

Воронежский государственный университет

***Аннотация.** Одной из самых сложных задач в информационных технологиях является обработка и распознавание объектов на изображениях. Исследования по распознаванию объектов, анализу изображений и речи включены в перечень приоритетных направлений развития науки и техники. В статье процесс поиска и классификации объектов на изображениях реализован с помощью TensorFlow Object Detection.*

***Ключевые слова:** поиск объектов, классификация объектов.*

Введение

Одна из востребованных задач компьютерного зрения, которая может быть решена с помощью глубоких нейронных сетей – это поиск на изображении объектов заданного типа.

В случае наличия нескольких объектов на изображении нужно решить две задачи:

- сегментация - выделение участков изображения, которые относятся к разным объектам;
- классификация - то есть определение типа объекта, для каждого выделенного сегмента отдельно.

Эти задачи составляют обнаружение объектов. Таким образом, обнаружение объектов - это процесс поиска и классификации объектов в изображении.

К решению задач детектирования и классификации применяются различные подходы: статистические, специально разработанные теории ключевых точек, генетические алгоритмы. Другим методом является машинное обучение, зачастую использующее в работе дополнительно один из предыдущих. Нейронные сети являются наиболее перспективным методом для детектирования и классификации объектов на изображении.

Для выполнения классификации, задач распознавания, обнаружении объектов на изображении используются технология сверточных нейронных сетей (CNNs, или ConvNets). Сверточные нейронные сети являются важными инструментами для глубокого

обучения и особенно полезны для классификации изображений, обнаружения объектов и задач распознавания. Сверточные нейронные сети реализованы в виде серии взаимосвязанных слоев. Слои состоят из повторяющихся блоков сверточных, ReLU (выпрямленных линейных блоков) и слоев объединения. Сверточные слои свертывают их вход с набором фильтров. Фильтры автоматически узнавались во время обучения в сети. Уровень ReLU добавляет нелинейность в сеть, что позволяет сети аппроксимировать нелинейное отображение между пикселями изображения и семантическим содержанием изображения.

1. Сверточные нейронные сети для поиска объектов на изображениях

Рассмотрим архитектуру полносверточных нейронных сетей [1,2]. При проектировании сверточных сетей в архитектуру могут добавляться разнообразные слои для повышения точности распознавания (drop out layer, local response normalization layer и т.д.). На рисунке 1 представлена базовая архитектура, которая определяет работу полносверточной нейронной сети.

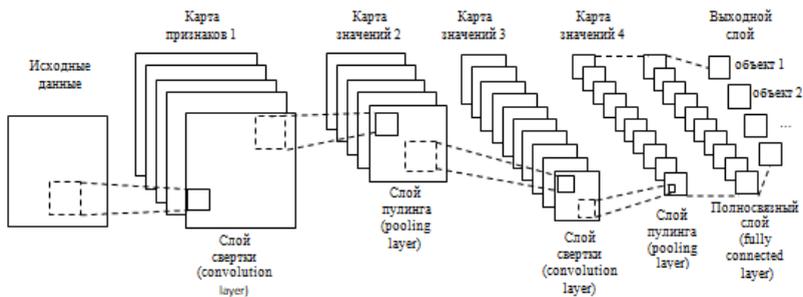


Рис. 1. Архитектура CNN

Основа работы полносверточных сетей – свертка изображения. Ключевыми слоями являются сверточные слои (convolution layer). В сверточном слое указывается количество выходов со слоя, ядро свертки, его шаг, размеры, отступ.

Операция свертки проходит ядром по всему изображению, в результате на выходе получаем величину отклика на ядро свертки в каждой точке изображения. Количество ядер каждого слоя свертки равно произведению количества выходов со слоя на количество входных картинок. Далее результаты проходят через следующий слой свертки,

получая значения уже для других ядер. К каждому сверточному слою можно добавлять слои регуляризации или нормализации (в зависимости от выбора разработчика). Прохождение изображений через многие сверточные слои позволяет получить богатое разнообразие возможных интерпретаций изображения.

Пройдя через требуемое количество слоёв свёртки, изображение затем попадает в слой пулинга (подвыборки, субдискретизации), pooling layer. Этот слой уменьшает размер входных изображений, не уменьшая их количества. Слой имеет ядро, которое движется подобно ядру свертки, вычисляя единственное значение для каждой области изображения. Уменьшение изображения способствует более быстрой обработке сетью большего количества данных. Это позволяет добавлять в следующих свёрточных слоях большее количество выходов, а также повышает точность результатов. Дело в том, что на уменьшенном изображении ядра свёртки того же размера способны захватывать большую область искомого объекта.

Последовательность: convolution/.../convolution/pooling (где количество convolutions layers определяется разработчиком) может повторяться несколько раз, вплоть до того, пока не будет достигнут минимальный размер изображения. Этот размер определяется экспериментально.

Для того чтобы выделенные объекты соответствовали оригинальному масштабу, уменьшенное изображение требуется вернуть к изначальному размеру. Слой upsample (upsample layer) выполняет увеличение изображения. На каждый выход имеется два входных изображения: первое – это обработанная картинка с предыдущего слоя (это может быть convolution или pooling), второе – это картинка из pooling layer, количество выходов которого равно количеству входов соответствующего upsample, а также размеры выходной картинки pooling равны размеру входной картинки upsample.

Таким образом, сверточная нейронная сеть имеет симметричную архитектуру относительно последнего pooling слоя и первого upsample слоя. Между слоями увеличения размеров изображения также помещаются сверточные слои, но количество выходов с них постепенно уменьшается.

Последовательность upsample/convolution/.../convolution необходима, чтобы привести изображение к исходным размерам, при этом сократив количество возможных интерпретаций изображения до количества искоемых групп объектов.

Процесс поиска объектов на изображении с использованием сверточных нейронных сетей состоит из следующих этапов [1]:

1. поиск возможных областей, содержащих объект;
2. построение карты признаков предполагаемых областей, содержащих объект;
3. классификация найденных объектов.

2. Поиск объектов на изображении

Обучение и использование нейросетевых моделей, способных локализовать и идентифицировать несколько объектов на одном изображении, остается основной задачей компьютерного зрения. TensorFlow Object Detection API — это платформа с открытым исходным кодом, входящая в состав TensorFlow, которая реализует создание, обучение, тестирование нейросетевых моделей обнаружения объектов.

Для поиска объектов на изображении использовались предварительно обученные нейронные сети на наборе данных COCO [3] (Common Object in Context — набор данных COCO состоит из 80 подкатегорий, 330000 изображений):

- faster_rcnn_nas;
- faster_rcnn_inception_resnet_v2;
- faster_rcnn_resnet101;
- ssd_resnet50.

Указанные модели на выходе выдают прямоугольник или прямоугольники (Boxes), внутри которого находится найденный объект (объекты).

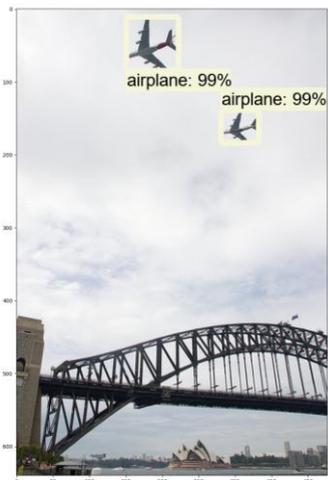
В табл. 1 представлены основные характеристики используемых нейронных сетей.

Таблица 1

Характеристики предварительно обученных нейронных сетей

№	Название	Скорость (мс)	Средняя точность (%)
1	faster_rcnn_nas	1833	43
2	faster_rcnn_inception_resnet_v2	620	37
3	faster_rcnn_resnet101	106	32
4	faster_rcnn_resnet50	89	30

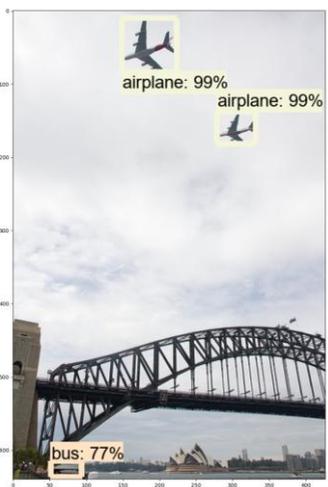
На рис. 2 представлены полученные результаты поиска объектов на изображениях из набора данных COCO (airplane_COCO.jpg).



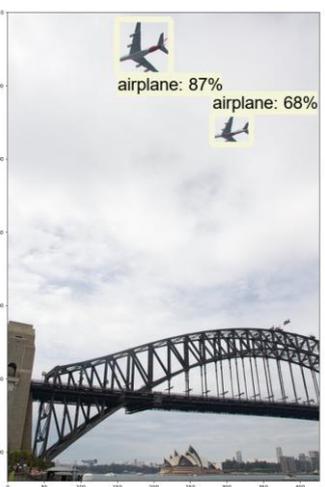
a



б



в

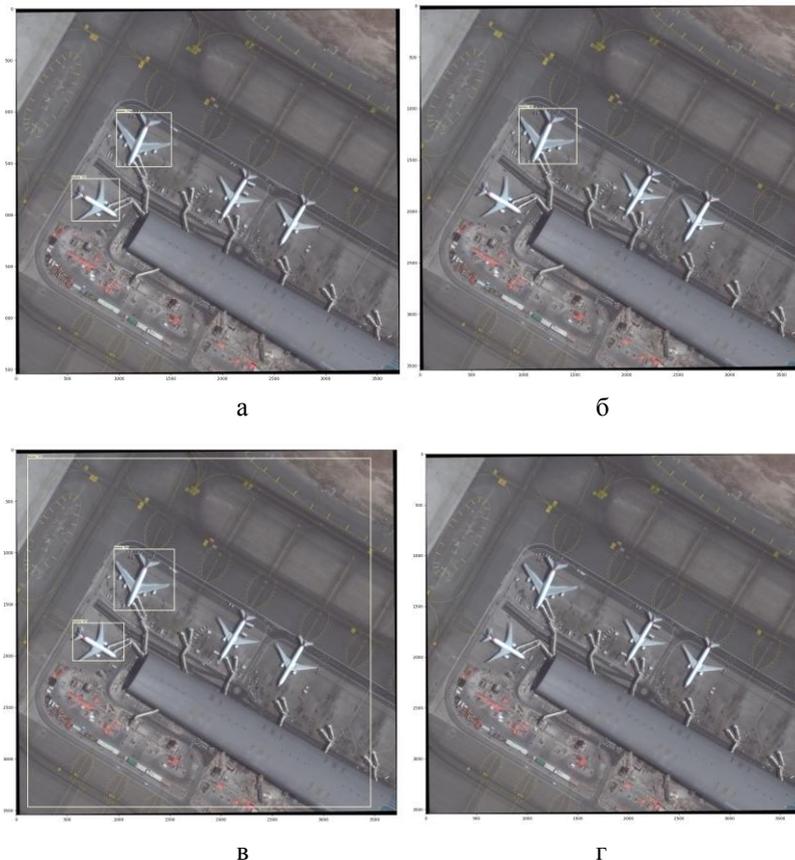


г

*a – faster_rcnn_nas, б – faster_rcnn_inception_resnet_v2,
в – faster_rcnn_resnet101, г – faster_rcnn_resnet50*

Рис. 2.

При визуализации полученных результатов поиска объектов на изображении выводится исходная картинка, на которой найденные объекты выделяются прямоугольниками. Рядом с каждым прямоугольником отображается название класса объекта и вероятность, с которой объект относится к этому классу.



*а – faster_rcnn_nas, б – faster_rcnn_inception_resnet_v2,
в – faster_rcnn_resnet101, г – faster_rcnn_resnet50*

Рис. 3.

Для тестирования использовались изображения из набора данных DOTA [4] (Dataset for Object DeTecton in Aerial Images), который состоит из изображений RGB и изображений в градациях серого. Изображения

RGB взяты из Google Earth и CycloMedia, а изображения в градациях серого взяты из панхроматического диапазона спутниковых изображений.

На рис. 7 представлены полученные результаты поиска объектов на изображениях из набора данных DOTA (airplane_DOTA.png).

Полученные результаты более наглядно представлены в табл. 1.

Таблица 2

Результаты поиска объектов на изображениях

	airplane_COCO.jpg	airplane_DOTA.png
faster_rcnn_nas	1 объект – 99% 2 объект – 99% 3 объект – не обнаружен	1 объект – 93% 2 объект – 74% 3 объект – не обнаружен 4 объект – не обнаружен
faster_rcnn_inception_resnet_v2	1 объект – 99% 2 объект – 99% 3 объект – не обнаружен	1 объект – не обнаружен 2 объект – 98% 3 объект – не обнаружен 4 объект – не обнаружен
faster_rcnn_resnet101	1 объект – 99% 2 объект – 99% 3 объект – 77%	1 объект – 64% 2 объект – 55% 3 объект – 94% 4 объект – не обнаружен
ssd_resnet50_v1	1 объект – 87% 2 объект – 68% 3 объект – не обнаружен	1 объект – не обнаружен 2 объект – не обнаружен 3 объект – не обнаружен 4 объект – не обнаружен

Представленные результаты показывают, что для COCO изображений выполняется поиск всех объектов, находящиеся на изображении (вероятность принадлежности к конкретному классу

достигает 99%). При использовании предварительно обученной нейронной сети `faster_rcnn_resnet101` выполняется более детальный поиск объектов меньшего размера. Нейронная сеть `ssd_resnet50_v1` работает достаточно быстро, но при этом для изображений из набора данных DOTA не удается выполнить поиск объектов на изображении.

Для улучшения полученных результатов, необходимо провести дообучение нейронной сети, с использованием характерных изображений, а также добавить дополнительную метку класса.

Список литературы

1. Митрофанова, Е. Ю. Нейросетевые технологии обработки информатики. Методы и технологии глубокого обучения : учебное пособие / Е. Ю. Митрофанова, А. А. Сирота, М. А. Дрюченко .— Воронеж : Издательский дом ВГУ, 2019 .— 197 с.
2. Сверточные нейронные сети: взгляд изнутри [Электронный ресурс]. – Режим доступа : <http://ru.datasides.com/code/cnn-convolutional-neural-networks/>
3. COCO - Common Objects in Context [Электронный ресурс]. – Режим доступа : <https://cocodataset.org/>
4. DOTA: A Large-Scale Benchmark and Challenges for Object Detection in Aerial Images [Электронный ресурс]. – Режим доступа : <https://arxiv.org/abs/2102.12219v1>