

ДЕТЕКТИРОВАНИЕ ТРУДНОРАЗЛИЧИМЫХ ОБЪЕКТОВ С ПРИМЕНЕНИЕМ СВЁРТОЧНОЙ НЕЙРОННОЙ СЕТИ YOLO NANO

С.И. Беляев, студент гр. 8К92
Е.А. Мыцко, к.т.н., доцент ОИТ,
Томский политехнический университет
E-mail: sib5@tpu.ru

Введение

На сегодняшний день технологии компьютерного зрения находят применение в различных областях человеческой деятельности. Одной из главных задач компьютерного зрения является детектирование объектов. Нейронные сети являются классическим инструментом решения задач компьютерного зрения, однако одними из наиболее эффективных являются свёрточные нейронные сети [1], позволяющие выделять признаки у изображений с достаточно высокой точностью.

При разработке моделей нейронных сетей для мобильных систем (например, беспилотные летательные аппараты - БПЛА) необходимо учитывать ограниченные ресурсы (объем памяти, энергопотребление) устройств на основе таких систем. Иными словами, такие сети должны быть «компактными». Один из наиболее известных примеров реализации «компактных» моделей нейронных сетей – YOLO Nano [2]. Таким образом, целью данной работы является подготовка датасета и последующее обучение модели нейронной сети архитектуры YOLO Nano.

Работа нейронной сети

Принадлежность модели нейронной сети к семейству YOLO означает, что все прогнозы сети выполняются за один ее проход, что значительно ускоряет работу и дает преимущество перед архитектурами, основанными на построении сети регионов. Также YOLO Nano является свёрточной нейронной сетью, что подразумевает, что каждый фрагмент изображения умножается на матрицу (ядро свертки) поэлементно, а результат суммируется и записывается в аналогичную позицию выходного изображения. Оригинальная статья [2] предлагает реализацию с расширением нейронной сети на входе (416x416x12).

В качестве основы для подготовки разметки обучающей, валидационной и тестовой выборок выбран открытый датасет Stanford Drone Dataset [3], содержащий изображения людей, сделанные с БПЛА. Создание и разметка датасета были осуществлены с применением сервиса Roboflow, позволяющим упростить лейбеллинг изображений. Для обучения модели датасет был разбит на 3 части: тренировочная часть, валидационная (для настройки гиперпараметров) и тестовая части в отношении соответственно 60%, 20%, 20%.

Перед непосредственным обучением был создан config файл, который и будет реализовывать архитектуру YOLO Nano. Для обучения модели выбран фреймворк DarkNet, обычно использующийся для обучения моделей семейства YOLO. Его реализация на языке C и использование технологии CUDA позволяет добиться высокой производительности работы фреймворка. Пример результатов работы модели на тестовой выборке представлен на рисунке 1.

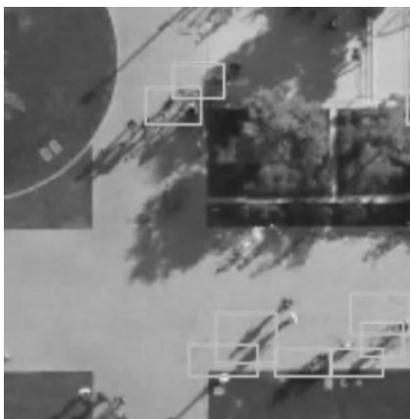


Рис. 1. Поиск фрагмента сцены в видеофайле

Тестирование алгоритма поиска трудноразличимых объектов

Источником для создания тестовой выборки послужил тот же датасет, что и для обучающей выборки. Для тестирования нейронной сети необходимо подать на вход модели новое изображение, в результате чего объекты, имеющие необходимые признаки будут помещены в прямоугольники.

На примере верхних bounding boxes (рис. 1) можно наблюдать правильную работу классификатора. Однако в некоторых случаях, как видно на нижних прямоугольниках, классификатор вместо пешеходов срабатывает на тени, что объясняется тем, что это гораздо более простой объект для распознавания. Данный недостаток в работе классификатора можно устранить обучением модели на выборке большего размера и на датасете, на котором отсутствуют тени. В таком случае у классификатора не будет иного выбора, как выделять признаки из изображений самих пешеходов, а не их теней даже в случае тестирования нейронной сети на датасетах, подобных использованным в данной работе.

Графики метрик в зависимости от размера обучающей выборки, полученные по результатам тестирования, представлены на рисунке 2. По оси абсцисс откладывается количество batch с изображениями, на которых обучалась сеть, по оси ординат — определенная метрика точности работы сети.

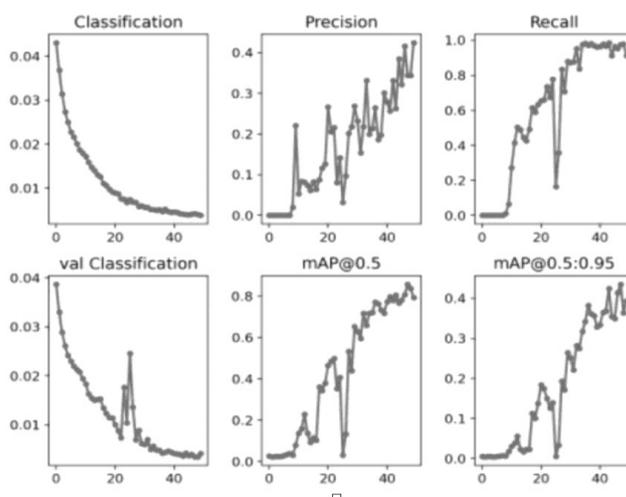


Рис. 2. Графики метрик точности обучения модели

Заключение

По результатам проверки сети на тестовой выборке были проанализированы графики метрик. В течение обучения значение метрики precision увеличивается, так как уменьшается количество срабатываний классификатора false positive. Аналогично ведет себя метрика recall, только в ее случае уменьшается количество срабатываний false negative. Метрика classification и val classification (процент ошибок и процент ошибок на валидационной выборке соответственно) наоборот снижается. Метрика mAP, являющаяся средним значением средней точности, показывает рост на протяжении обучения и достигает 80%.

Примерно в середине процесса обучения можно наблюдать увеличение процента ошибок, что может объясняться ошибками при лейбеллинге некоторых изображений обучающей выборки.

Список использованных источников

1. LeCun Y., Haffner P., Bottou L., Bengio Y. Object Recognition with Gradient-Based Learning. In: Shape, Contour and Grouping in Computer Vision. Lecture Notes in Computer Science. – 1999. – V.1681. – Springer, Berlin, Heidelberg.
2. Wong A., Famouri M., Shafiee M., Li F., Chwyl B., Chung J. YOLO Nano: a Highly Compact You Only Look Once Convolutional Neural Network for Object Detection. 2019 Fifth Workshop on Energy Efficient Machine Learning and Cognitive Computing. – 2019. – 22-25 p.
3. Stanford Drone Dataset [Электронный ресурс]. – URL: https://cvgl.stanford.edu/projects/uav_data/, свободный. (дата обращения: 24.02.2022).