<https://arxiv.org/pdf/1804.02767.pdf>

В статье рассматриваются обновления функционала для нейронной сети YOLO.

Поскольку автор потратил много времени на Twitter, GAN, а также помогал в исследованиях других людей, YOLOv3 имеет небольшое количество улучшений относительно предыдущей версии YOLOv2 и основные идеи позаимствованы у других людей.

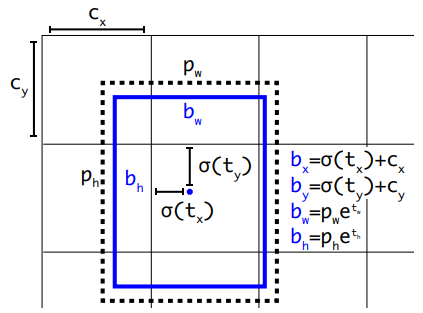
## Предсказание Bounding Box

Работает аналогично предыдущей версии YOLOv2.

Сеть предсказывает 4 координаты для каждого bounding box: tx, ty, tw, th.

В качестве функции потерь используется сумма квадратов ошибок.

И оценка наличия объектов для одного bounding box предсказывается с помощью логистической регрессии.



## Предсказание класса

Для предсказания класса объекта в YOLOv3 не используется Softmax.

Вместо этого, используются независимые логистические классификаторы и binary cross entropy в качестве функции потерь.

Этот метод лучше для сложных наборов данных, таких, как Open Images, в котором много перекрывающихся классификаций (например, Женщина и Человек). Использование softmax налагает ограничение, что у каждого bounding box есть только один класс, который часто не соответствует действительности.

## Предсказание в масштабе

Объекты предсказываются в 3 разных масштабах.

YOLOv3 извлекает признаки для разных масштабов, подобно тому, как это делается в функциональных пирамидальных сетях (FPN)..

В базовый экстрактор Darknet-53 добавлены несколько сверточных слоев.

Последний из этих слоев предсказывает bounding box, наличие объекта и класс объекта.

Из 2 предыдущих слоев берется карта признаков. Затем ее разрешение увеличивается в 2 раза. Карта признаков также мигрировала из более ранней версии YOLO.

Затем добавляется еще несколько сверточных слоев для обработки этой комбинированной карты признаков и, в конечном итоге, прогнозирования тензора.

Для определения наиболее подходящего bounding box используется кластеризация по методу k-mean.

## Выделение признаков: Darknet-53

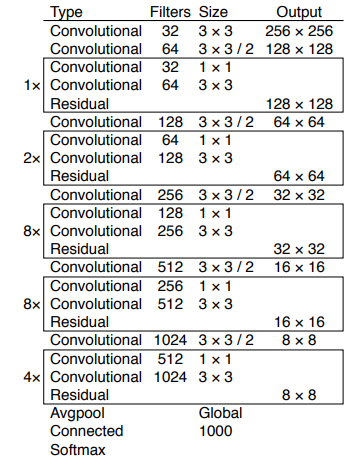


Рис. 2 Архитектура Darknet-53.

В YOLOv2 для извлечения признаков используется сеть классификации Darknet-19.

В YOLOv3 используется гораздо более глубокая сеть Darknet-53, т.е. имеющая 53 сверточных слоя.

И YOLOv2, и YOLOv3 также используют Batch Normalization.

Сеть использует соединения быстрого доступа (shortcut connections) см. рис выше.

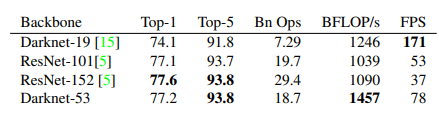


Рис. 3 сравнение сетей на 1000-классовом наборе данных ImageNet

В сравнении используются изображения размером 256 × 256 на графическом процессоре Titan X.

По сравнению с ResNet-101 Darknet-53 имеет лучшую производительность и Darknet-53 в 1,5 раза быстрее.

По сравнению с ResNet-152 Darknet-53 имеет аналогичную производительность и Darknet-53 в 2 раза быстрее.

## Результаты

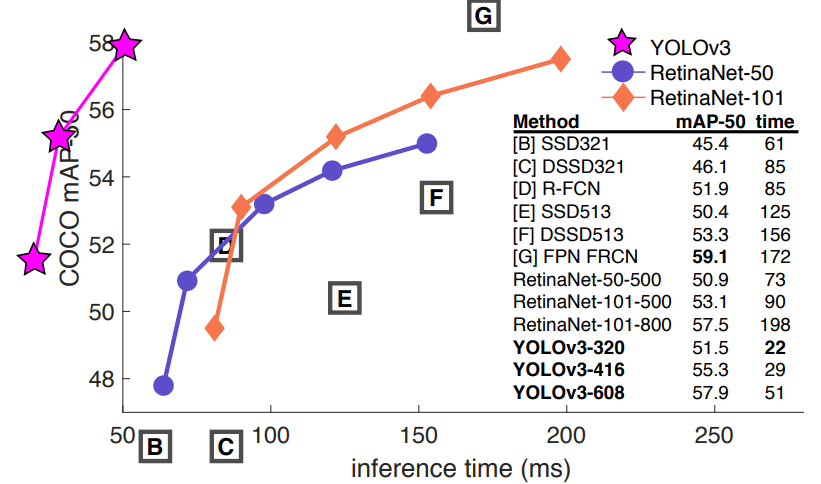
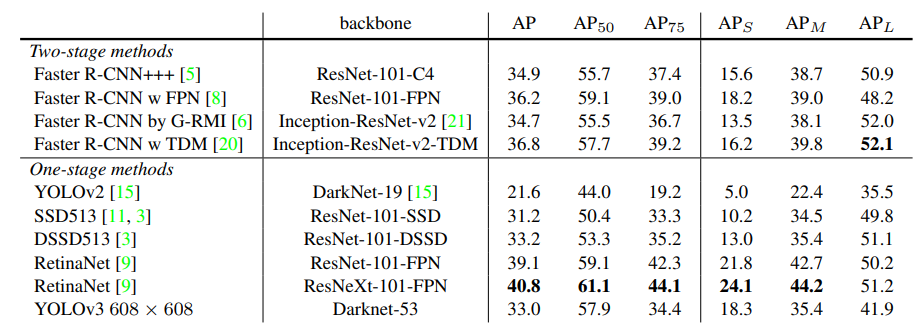


Рис. 4 сравнение работы сетей различной архитектуры

На рисунке выше видно, что по сравнению с RetinaNet, YOLOv3 получил сопоставимый показатель mean Average Precision(mAP) при гораздо меньшем времени работы.

Например, YOLOv3–608 получил 57,9% mAP за 51 мс, а RetinaNet-101–800 только 57,5% mAP за 198 мс, что почти в 4 раза медленнее.



YOLOv3 намного лучше, чем SSD, и имеет производительность, аналогичную DSSD.