

Детектирование объектов с использованием методов глубокого обучения

Залесская Галина

Специалист по глубокому обучению Intel

Летняя школа по компьютерному зрению. ННГУ имени Н.И. Лобачевского

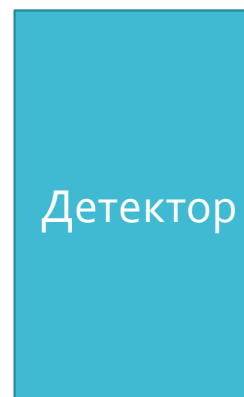
7 июля 2022 г

Agenda

- Постановка задачи
- Датасеты
- Метрики точности и производительности
- Общая архитектура детектора
- Классические архитектуры
 - 2-стадийные
 - 1-стадийные
- Сравнение точности и прозводительности

Постановка задачи детектирования

- Что и где?



Класс
«КОТ»

(x, y, height, width,
confidence)

Класс
«собака»

(x, y, height, width,
confidence)



Датасеты

- Маленькие разнообразные датасеты на Roboflow (~1000 картинок):
<https://public.roboflow.com/object-detection>
- MS COCO (320K картинок с разметкой 80-ти классов):
<http://cocodataset.org/>
- Open Images Dataset (1.9M картинок с разметкой 600 классов):
<https://storage.googleapis.com/openimages/web/index.html>

Метрика ТОЧНОСТИ

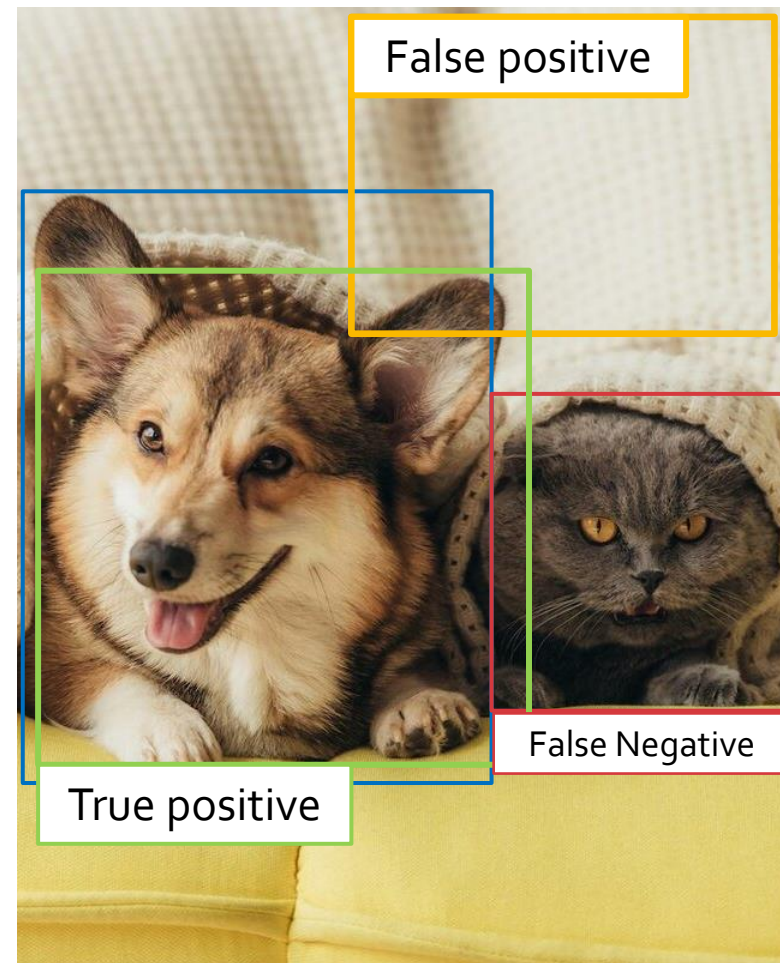
$$\text{IoU} = \frac{\text{Area of Overlap}}{\text{Area of Union}}$$


Фиксируем порог $\text{IoU}=0.5$

$$\text{Precision} = \frac{TP}{TP + FP}$$

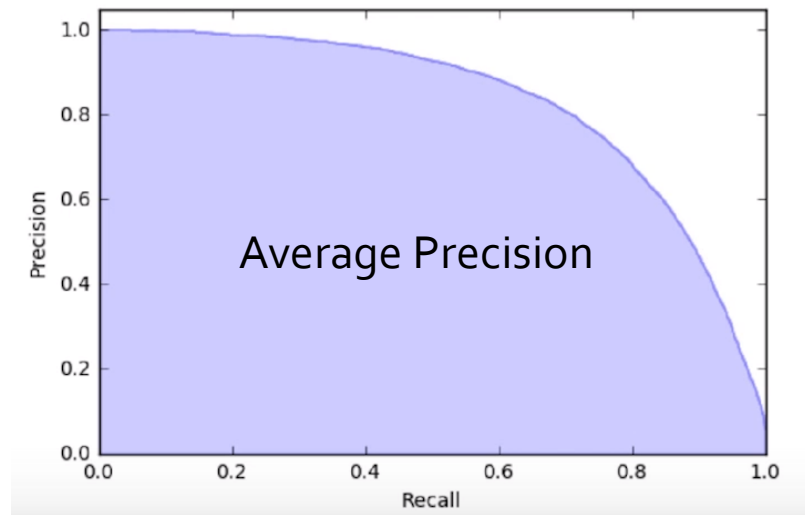
$$\text{Recall} = \frac{TP}{TP + FN}$$

$$F1 = \frac{2 * \text{precision} * \text{recall}}{\text{precision} + \text{recall}}$$



Метрика точности

- **Проблема:** необходимость подбора «хорошего» IoU для максимизации f_1
- **Решение:** Average Precision (AP) метрика
- Варьируем порог IoU от 0 до 1 и считаем Precision и Recall для каждого конкретного IoU.
- Значение AP – площадь под образовавшейся кривой



Метрики производительности

FPS (frames per second)

- сколько изображений обрабатывает модель за 1 секунду

FLOPs (Floating Point Operations)

- сколько операций с нецелыми числами нужно для обработки 1 изображения

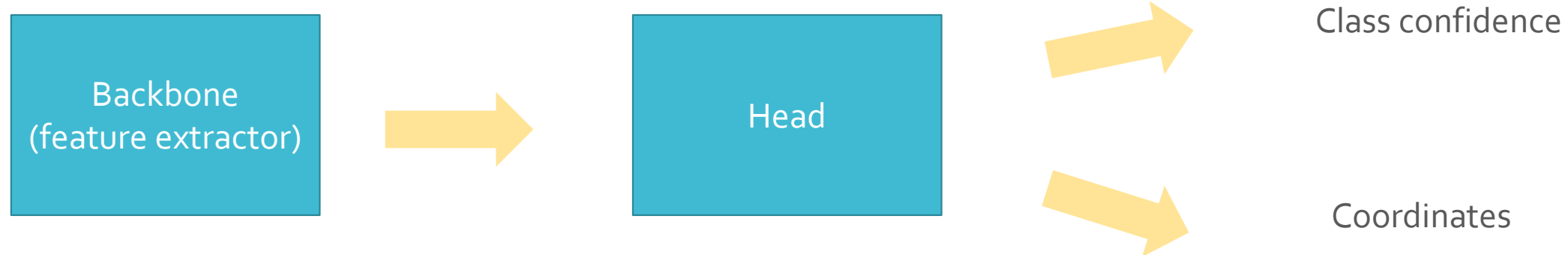


Архитектура

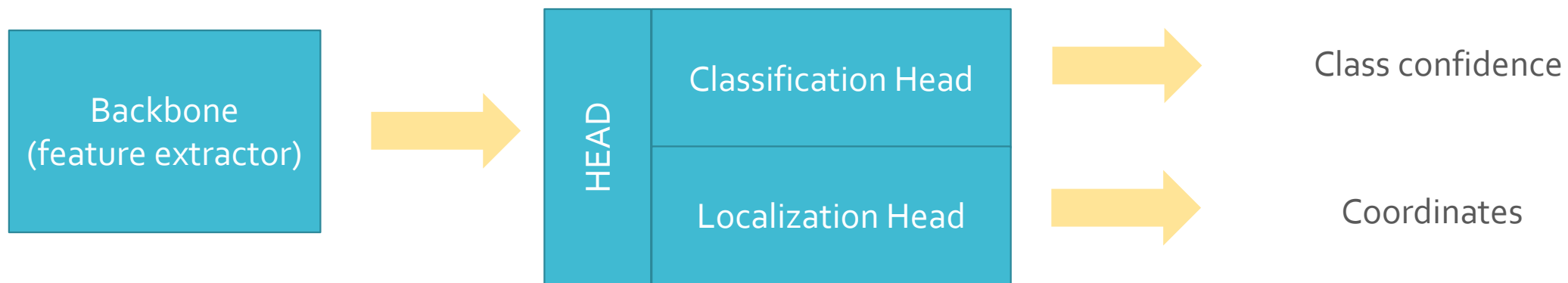
Архитектура классификатора



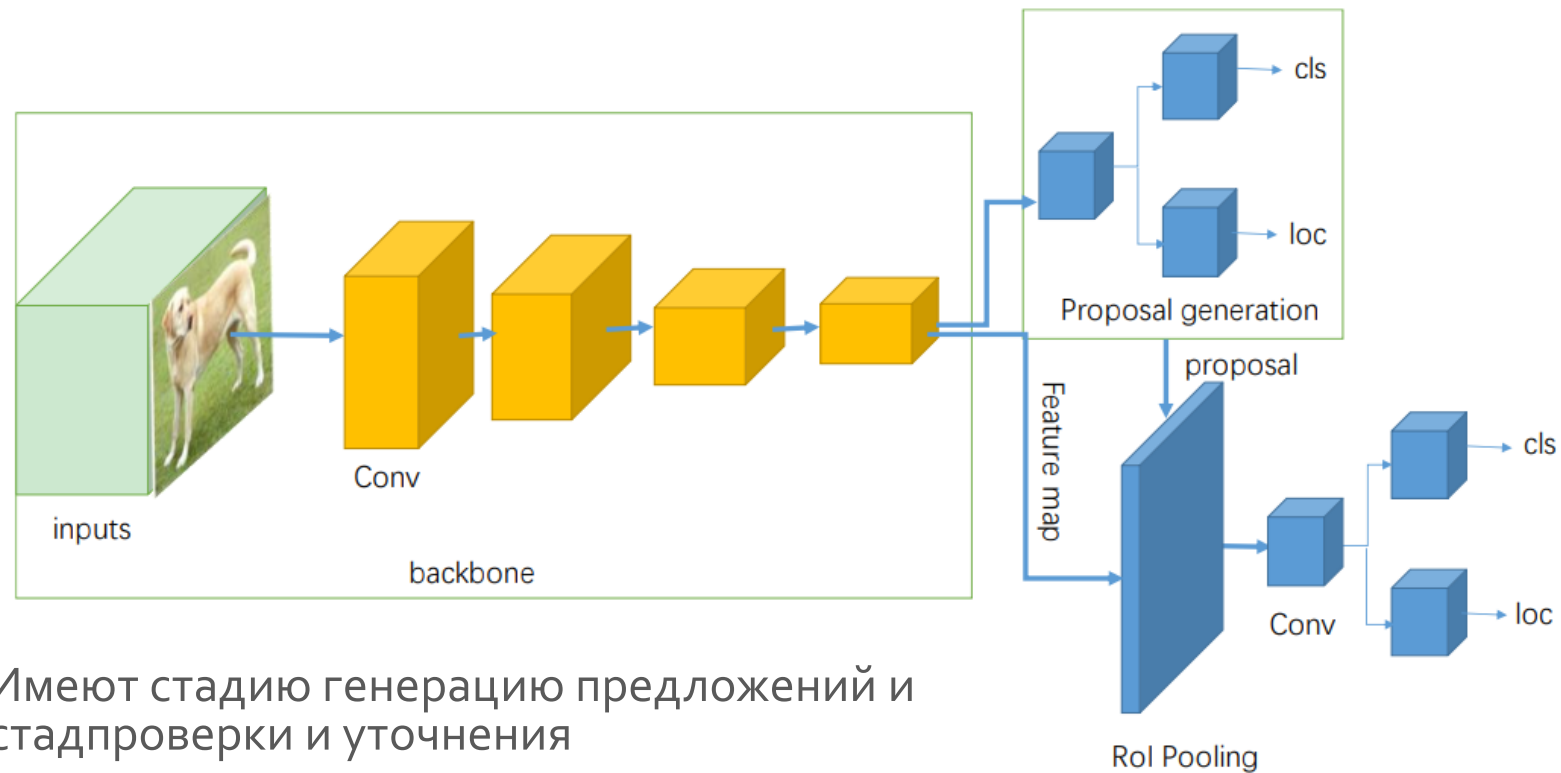
Архитектура детектора



Архитектура детектора

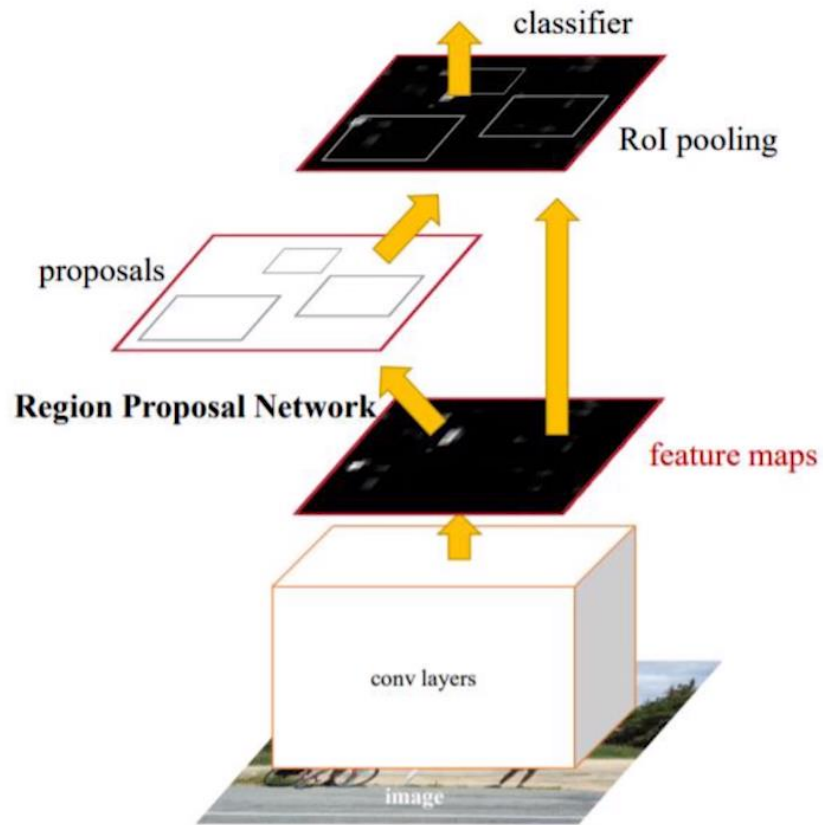


Двухстадийные детекторы

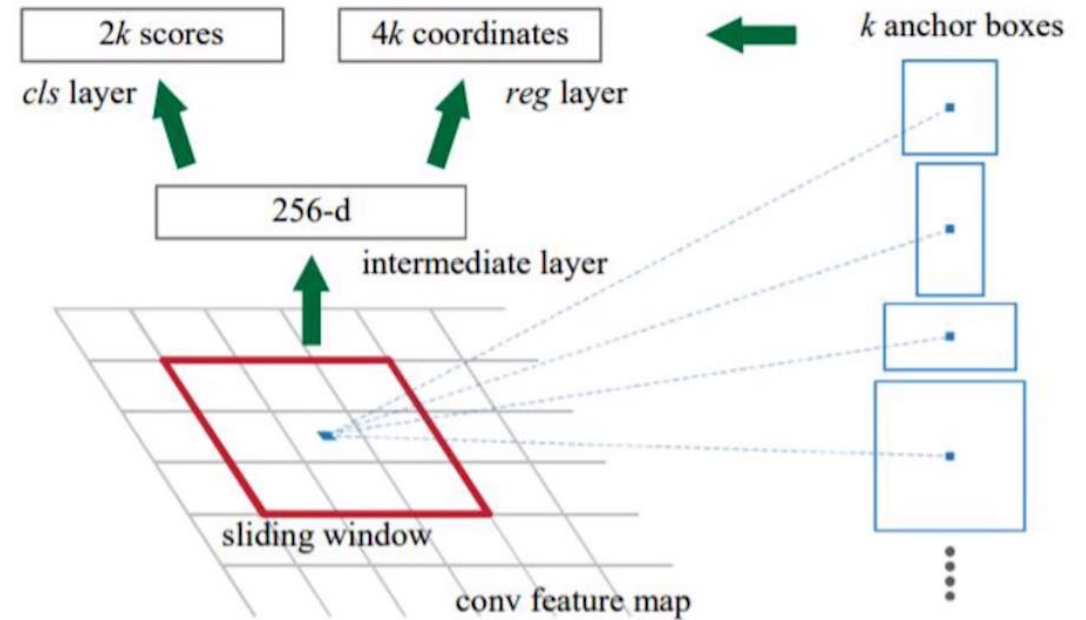


- Имеют стадию генерацию предложений и стад проверки и уточнения
- R-CNN (2014)
- Fast R-CNN (2015)
- Faster R-CNN (2015)
- + : Высокая точность
- -: Медленная производительность

Faster R-CNN (2015)

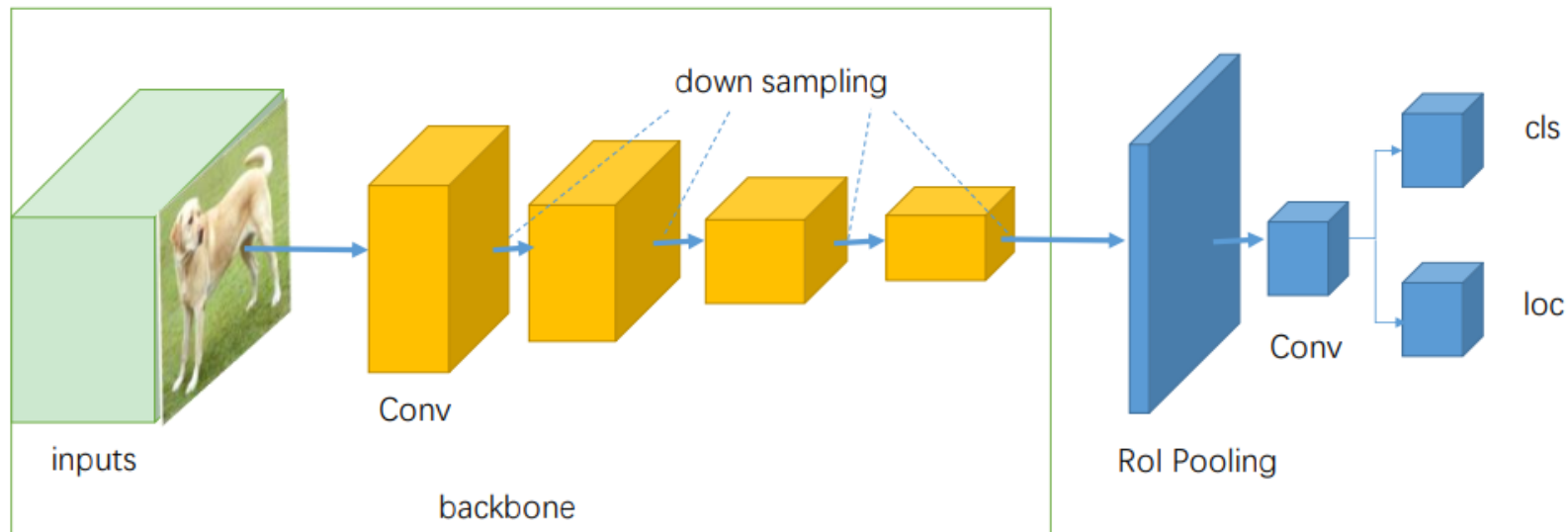


Faster R-CNN pipeline



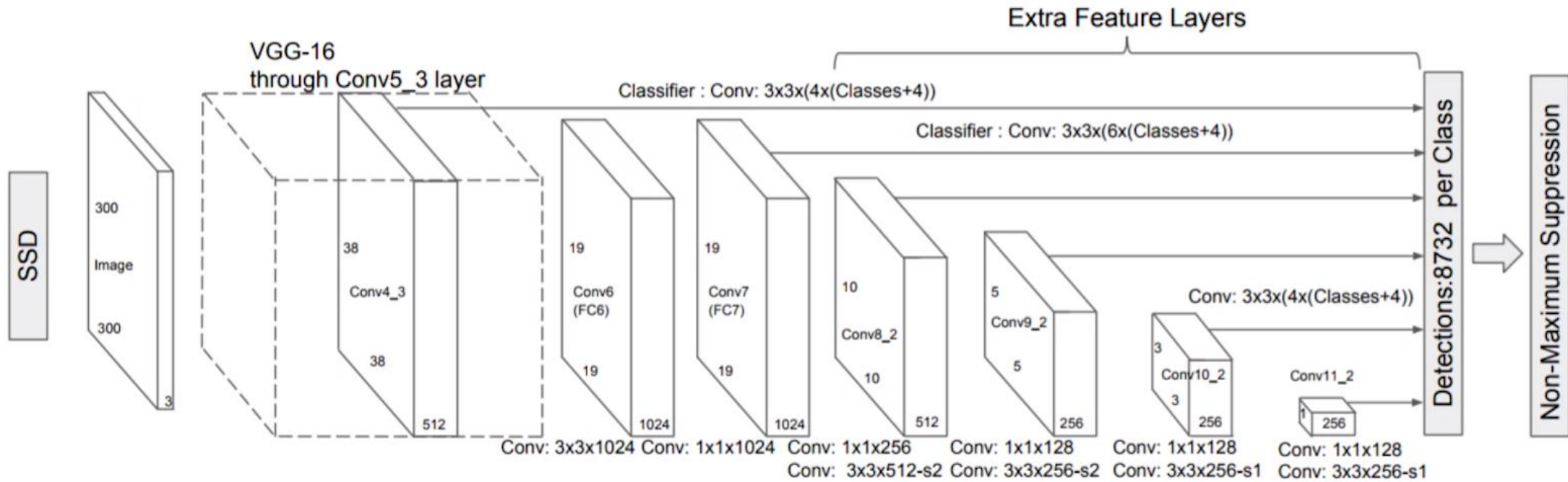
Region proposal network

Одностадийные детекторы



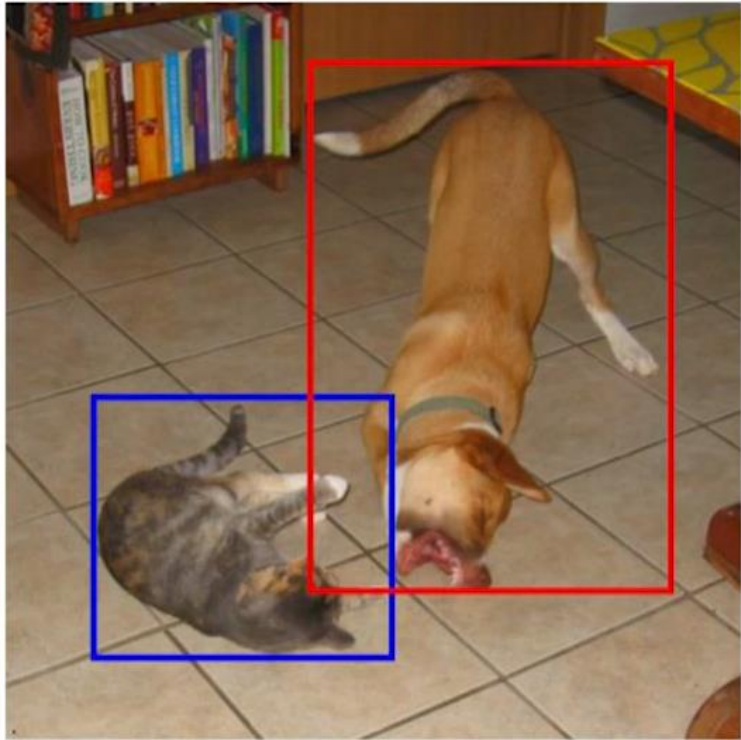
- Минуют стадию генерации предложений
- Благодаря высокой производительности, активно используются на практике
- Подразделяются на anchor-based и anchor-free

SSD (2016): Single Shot Multibox Detector

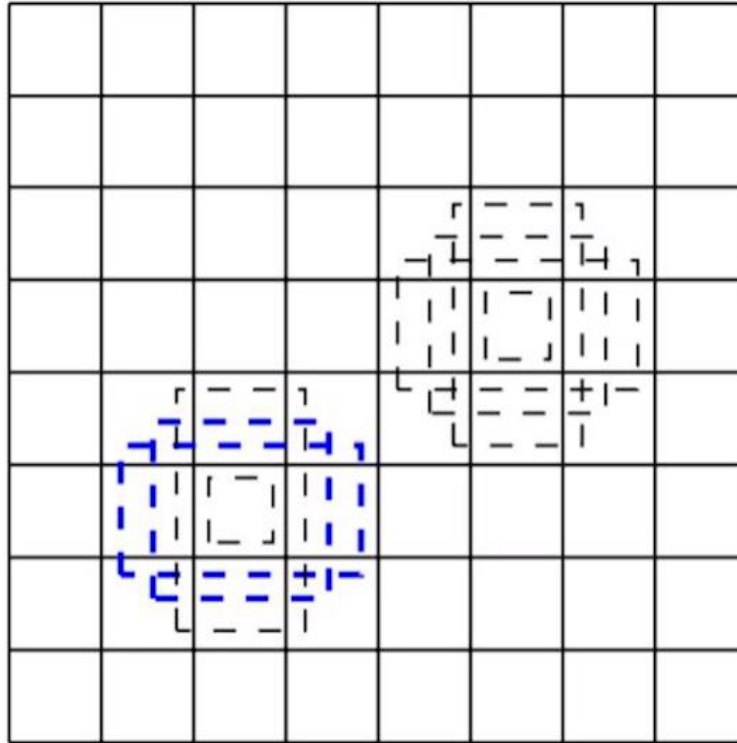


SSD: Single Shot MultiBox Detector. <https://arxiv.org/pdf/1512.02325.pdf>

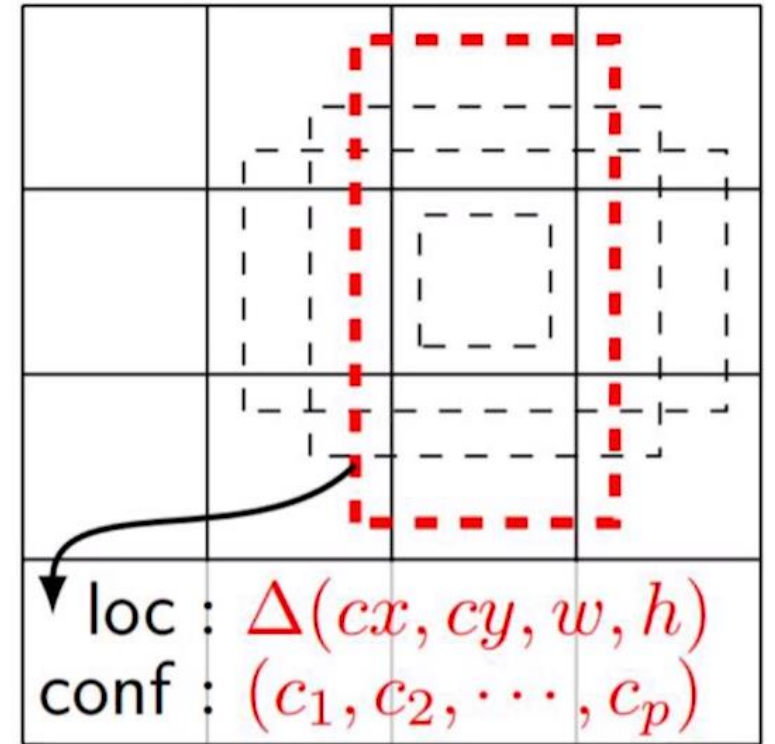
SSD (2016): Single Shot Multibox Detector



(a) Image with GT boxes



(b) 8×8 feature map

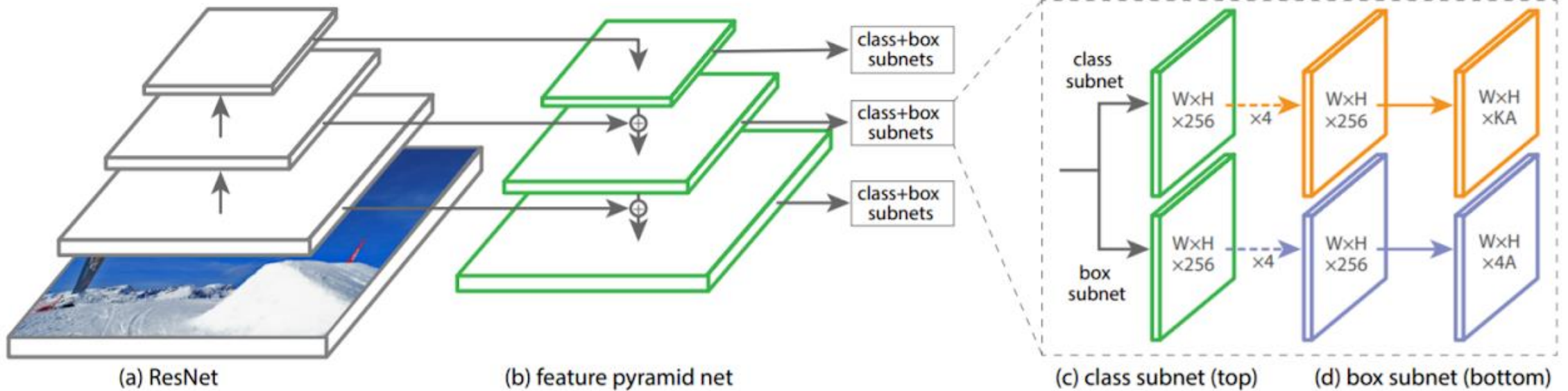


(c) 4×4 feature map

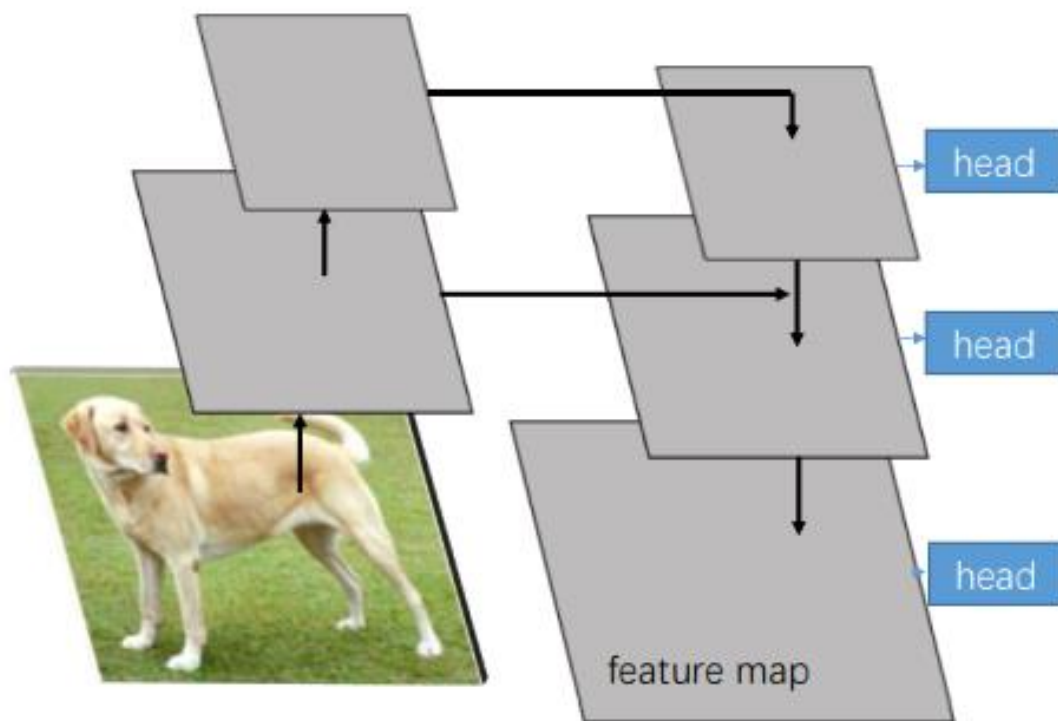
Non-Maximum Suppression (NMS)



RetinaNet (2018)



Feature Pyramid Network



(d) Feature pyramid network

Сравнение точности и производитель ности

- Тренировочные данные: PASCAL VOC 2007+2012
- Тестовые данные: PASCAL VOC 2007
- Показатель качества: средняя точность предсказания, усредненная по 20 классам (mean average precision)
- Инфраструктура: NVIDIA M40 или Titan X (сравнение качественное)

Модель	mAP, %	FPS
Fast R-CNN	70.0	0,5
Faster R-CNN VGG-16	73.2	7
Faster R-CNN ResNet	76.4	5
YOLO	63.4	45
SSD500	76.8	19
YOLOv2 544x544	78.6	40

Спасибо за внимание