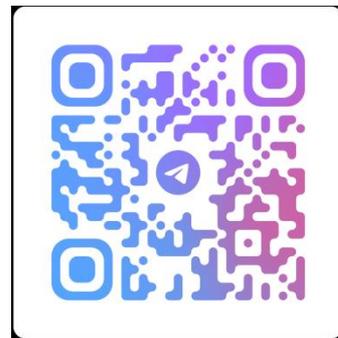


Основы практического использования нейронных сетей.

Лекция 9. Обзор современных глубоких сетей для
обнаружения объектов на изображениях.

Дмитрий Буряк.
к.ф.-м.н.
dyb04@yandex.ru



Задача обнаружения на изображениях

- ❑ Каждое изображение содержит несколько объектов. Классы объектов заданы.
- ❑ Необходимо определить классы объектов и их позиции.

❑ Метрики:
$$Precision = \frac{TP}{TP + FP}$$

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

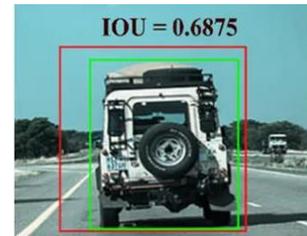
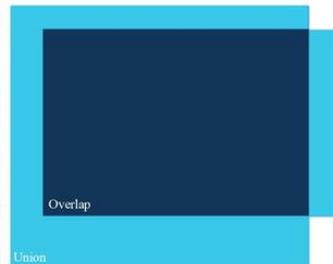
- ❑ IoU (Intersection over Union) – степень пересечения обнаруженного объекта с истинностью;
$$AP@α = \int_0^1 p(r) dr$$
- ❑ Average Precision (AP)

α – порог для IoU

- ❑ mAP (Mean Average Precision) - усреднение AP по всем классам для заданного порога для IoU.



$$IoU = \frac{\text{area of overlap}}{\text{area of union}}$$



База изображений PASCAL VOC

❑ PASCAL VOC (PASCAL Visual Object Classification)

- Создана в 2005г;
- Около 10К изображений;
- 20 классов.

❑ Соревнования PASCAL VOC

- Проводились в 2005-2012гг;
- Обнаружение (сегментация) объектов.



База изображений COCO

☐ COCO (Microsoft Common Objects in Context)

- Создана в 2015г;
- Более 123К изображений;
- Более 886К объектов;
- 91 класс.



☐ Конкурсы COCO

- Проводятся с 2015
- Обнаружение (сегментация) объектов, фона, ключевых точек объектов



☐ COCO. Обнаружение объектов на изображениях

- 200К изображений
- обучающая выборка: 500К объектов
- 80 классов



R-CNN

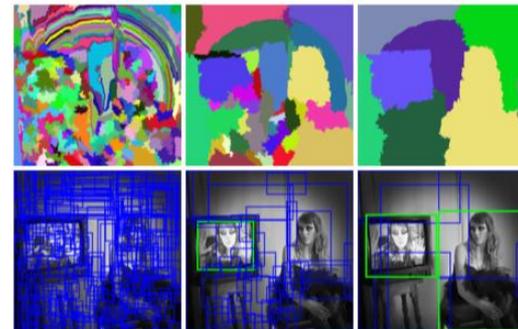
- ❑ R-CNN: Region-based Convolutional Network (R.Girshick et al., 2014);
- ❑ Предварительный отбор регионов интереса (Selective Search: J.R.R. Uijlings et al, 2012):

- Простые дескрипторы формы и цвета, различный масштаб;
- Низкие вычислительные затраты.

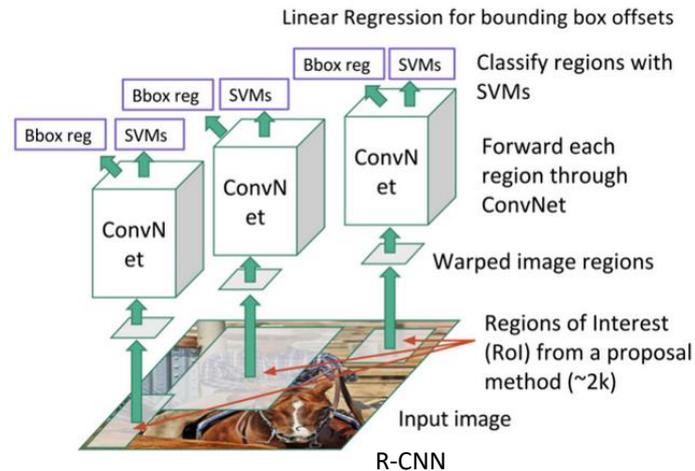
- ❑ Масштабирование региона → CNN → вектор признаков → классификатор для каждого класса (SVM) → регрессия (для позиции);

❑ **mAP = 62.4% (PASCAL VOC 2012)**

❑ **mAP = 31.4% (ImageNet 2013)**

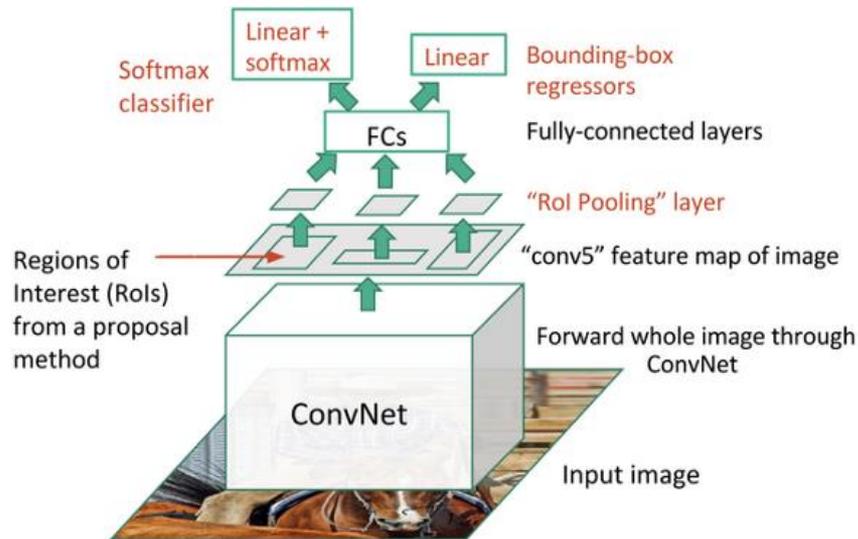


Selective Search: результаты сегментации и выбранные регионы



Fast R-CNN

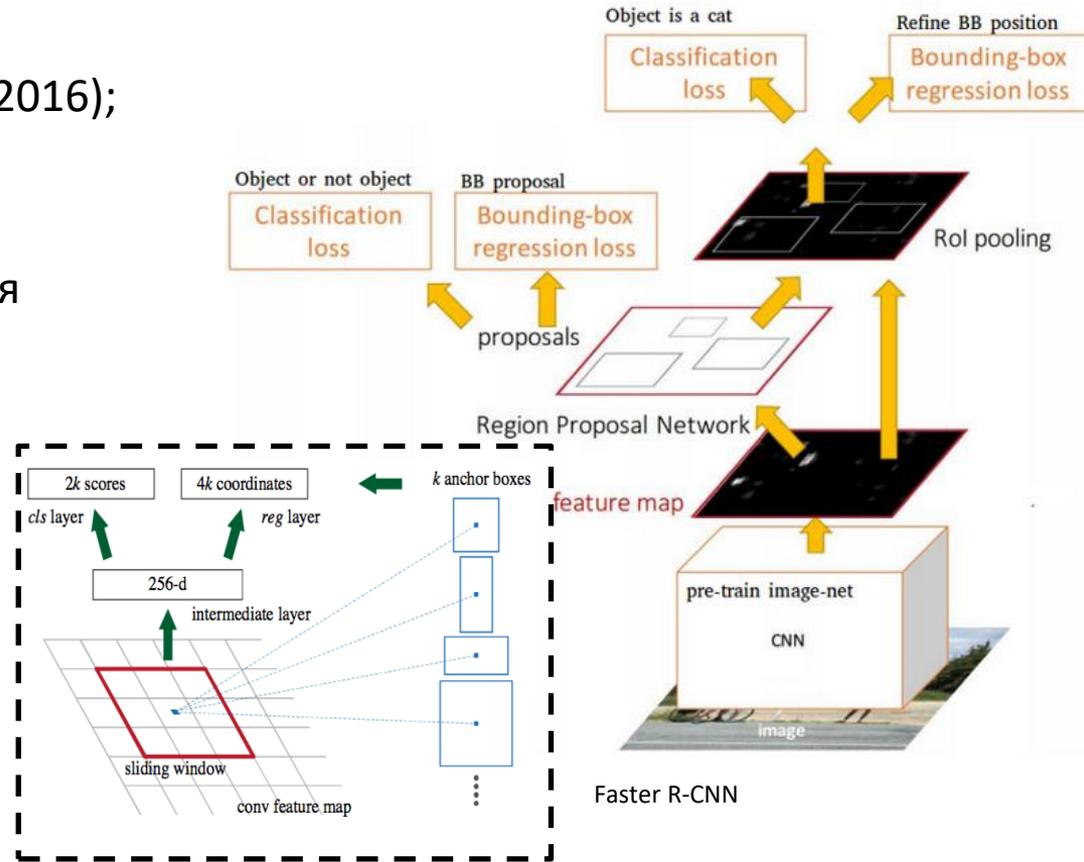
- ❑ Fast R-CNN: Fast Region-based Convolutional Network (R.Girshick et al., 2015);
- ❑ Однократное применение CNN для выделения признаков.
- ❑ CNN на всем изображении → Выбор регионов на картах → классификатор → регрессия (для позиции);
- ❑ **mAP = 70% (PASCAL VOC 2007)**
- ❑ **mAP = 68.8% (PASCAL VOC 2010)**
- ❑ **mAP = 68.4% (PASCAL VOC 2012)**



Fast R-CNN

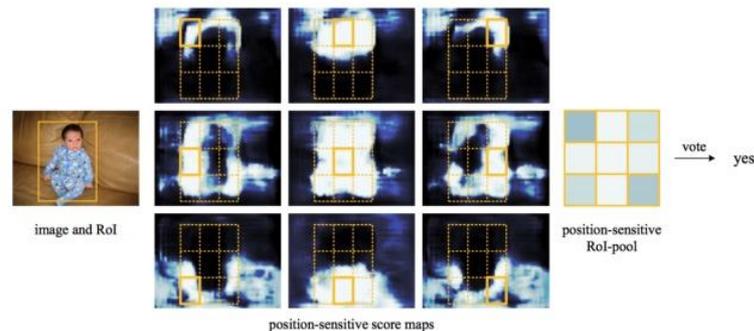
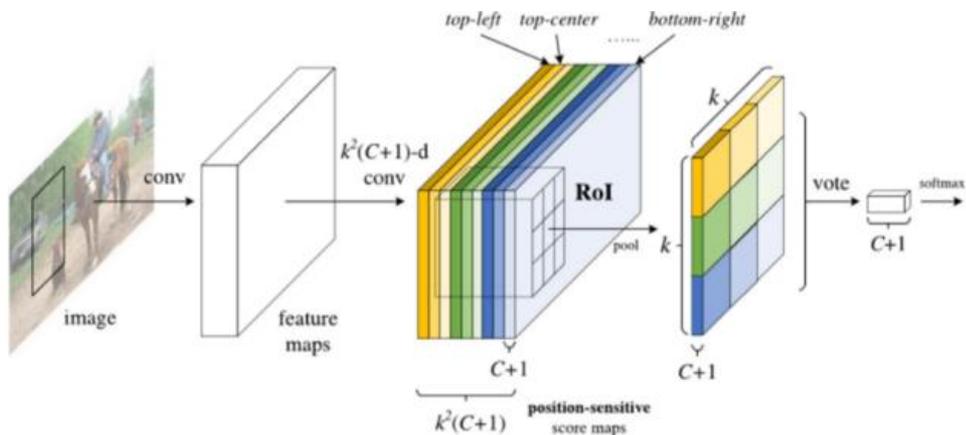
Faster R-CNN

- ❑ Faster R-CNN: Faster Region-based Convolutional Network (S. Ren et al., 2016);
- ❑ Selective Search – вычислительно «дорогой» .
- ❑ Region Proposal Network (RPN) для генерации регионов.
- ❑ Faster R-CNN = RPN+Fast R-CNN;
- ❑ Fast R-CNN обрабатывает карты признаков;
- ❑ 34x быстрее Fast R-CNN;
- ❑ **mAP = 78.8% (PASCAL VOC 2007)**
- ❑ **mAP = 75.9% (PASCAL VOC 2012)**

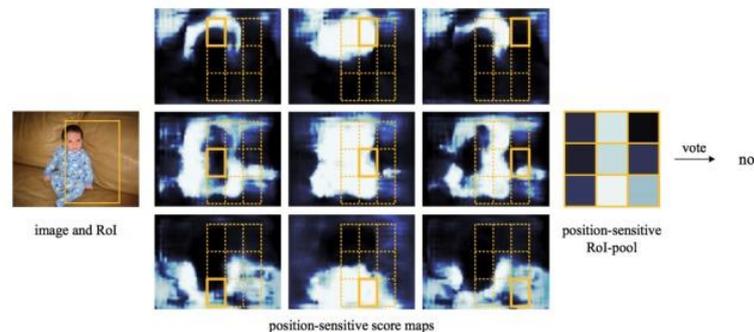


R-FCN

- ❑ R-FCN: Region-based Fully Convolutional Network (J. Dai et al., 2016);
- ❑ «Упрощение» этапа обработки регионов;
- ❑ Вычисление карт признаков-частей объектов;
- ❑ Применяется к результатам RPN
- ❑ 2.5x быстрее Faster R-CNN;
- ❑ **mAP = 83.6% (PASCAL VOC 2007)**
- ❑ **mAP = 53.2% (COCO 2015)**



Visualization of R-FCN ($k \times k = 3 \times 3$) for the *person* category.



Visualization when an RoI does not correctly overlap the object.

YOLO

❑ YOLO: You Only Look Once (J. Redmon et al., 2016);

❑ Нет выделения регионов;

❑ SxS сегментов изображения;

❑ Выходной тензор $S \times S \times (C+B \times 5)$:

- C – число классов,

- B – число гипотез (прямоугольников)

обнаруженных объектов (4 координаты +
степень уверенности)

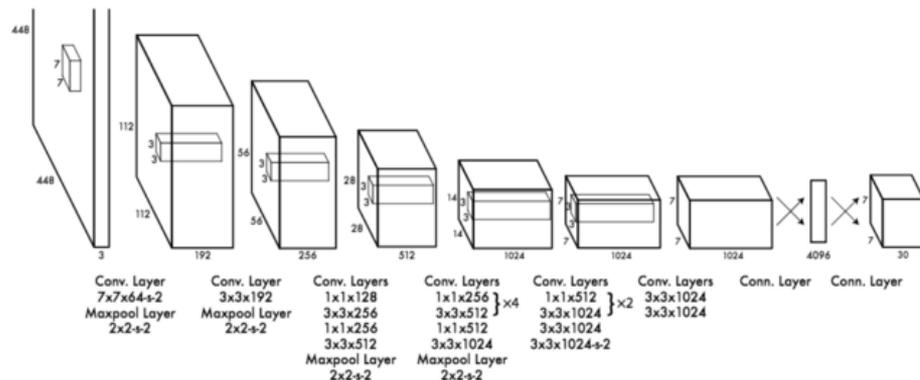
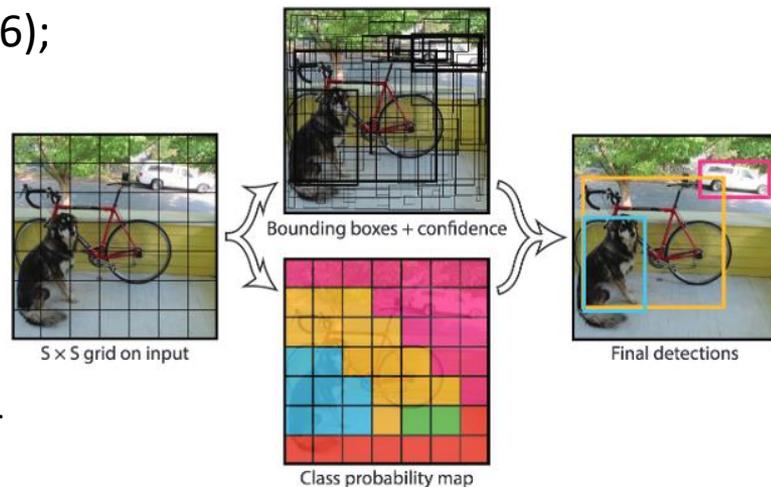
❑ Основана на GoogLeNet;

❑ Non-Maximum Suppression;

❑ Работа в «реальном времени»;

❑ mAP = 63.7% (PASCAL VOC 2007)

❑ mAP = 57.9% (PASCAL VOC 2012)



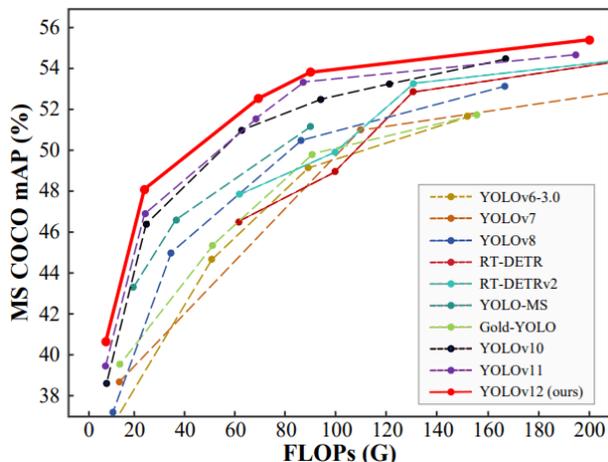
Non Maximum Suppression (NMS)

- ❑ Объединение нескольких перекрывающихся регионов.
- ❑ - Найти регион с максимальной уверенностью,
 - Удалить все регионы, пересекающиеся с ним,
 - Повторять пока список регионов не пуст.



Семейство YOLO

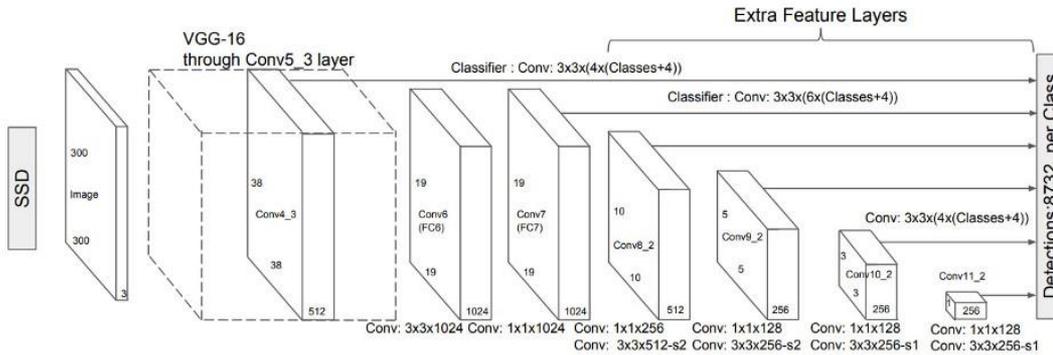
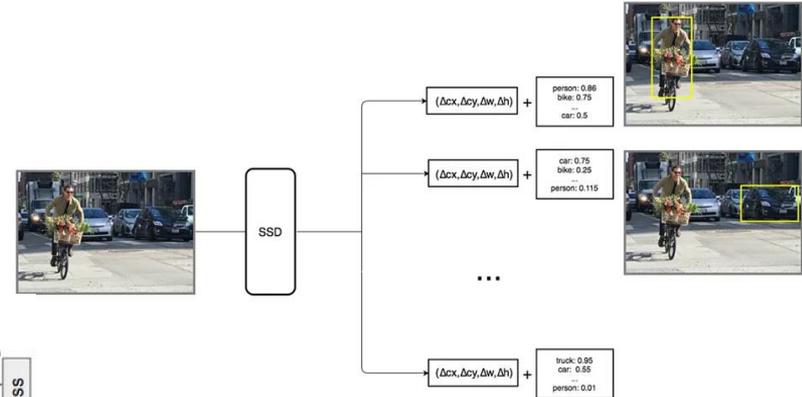
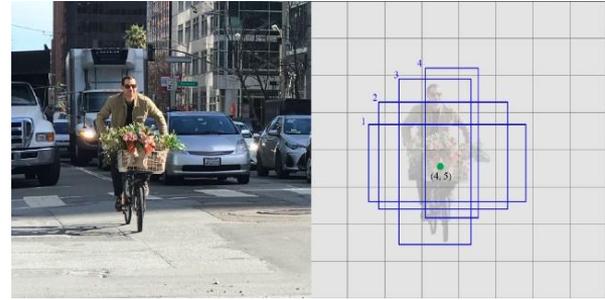
- ❑ YOLOv2, YOLOv3 (2018): введение «якорных» рамок, обнаружение на трех уровнях масштаба (FPNet)
- ❑ YOLOv4 (2020): изменение базовой сети (backbone), оптимизация обработки пирамиды изображений, подбор оптимальных параметров (генетические алгоритмы)
- ❑ YOLOv5 (2020): несколько вариантов размера сети, оптимизация скорости
- ❑
- ❑ YOLOv12 (2025)



Model Configuration	Precision	Recall	mAP@50
YOLOv12n	0.916	0.969	0.978
YOLOv12s	0.898	0.956	0.974
YOLOv12m	0.898	0.956	0.974
YOLOv12l	0.898	0.956	0.974
YOLO11n	0.84	0.76	0.862
YOLO11s	0.874	0.826	0.909
YOLO11m	0.809	0.821	0.879
YOLO11l	0.836	0.877	0.866
YOLO11x	0.857	0.85	0.91
YOLOv10n	0.84	0.8	0.89
YOLOv10s	0.82	0.83	0.88
YOLOv10m	0.83	0.8	0.87
YOLOv10b	0.85	0.82	0.88
YOLOv10l	0.85	0.75	0.83
YOLOv10x	0.77	0.81	0.85

SSD

- ❑ SSD: Single Shot Detector (W. Liu et al., 2016);
- ❑ Оценка «формы» объекта для каждой ячейки выходной карты;
- ❑ Оценка положения и вероятности класса;
- ❑ Обнаружение на разных масштабах;
- ❑ Non-maximum suppression;
- ❑ **mAP = 83.2% (PASCAL VOC 2007)**
- ❑ **mAP = 82.2% (PASCAL VOC 2012)**

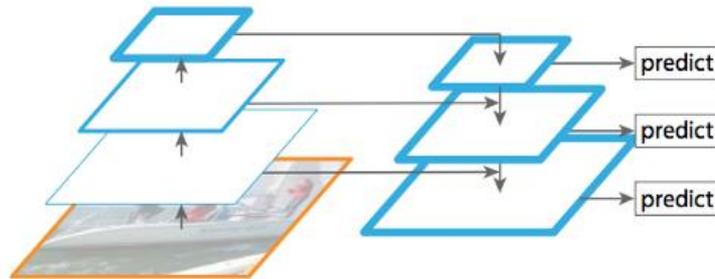


Сравнение

Модель	PASCAL VOC 2007	PASCAL VOC 2010	PASCAL VOC 2012	COCO 2015
R-CNN	-	62.4%	-	-
Fast R-CNN	70%	68.8%	68.4%	-
Faster R-CNN	78.8%	-	75.9%	-
R-FCN	82%	-	-	53.2%
YOLO	63.7%	-	57.9%	-
SSD	83.2%	-	82.2%	48.5%

Feature Pyramid Networks (FPN)

- ❑ Проблема надежного обнаружения объектов разного размера;
- ❑ Сверточная сеть → пирамида карт признаков (разные масштабы) → признаки разного уровня;
- ❑ FPN – объединение карт признаков.

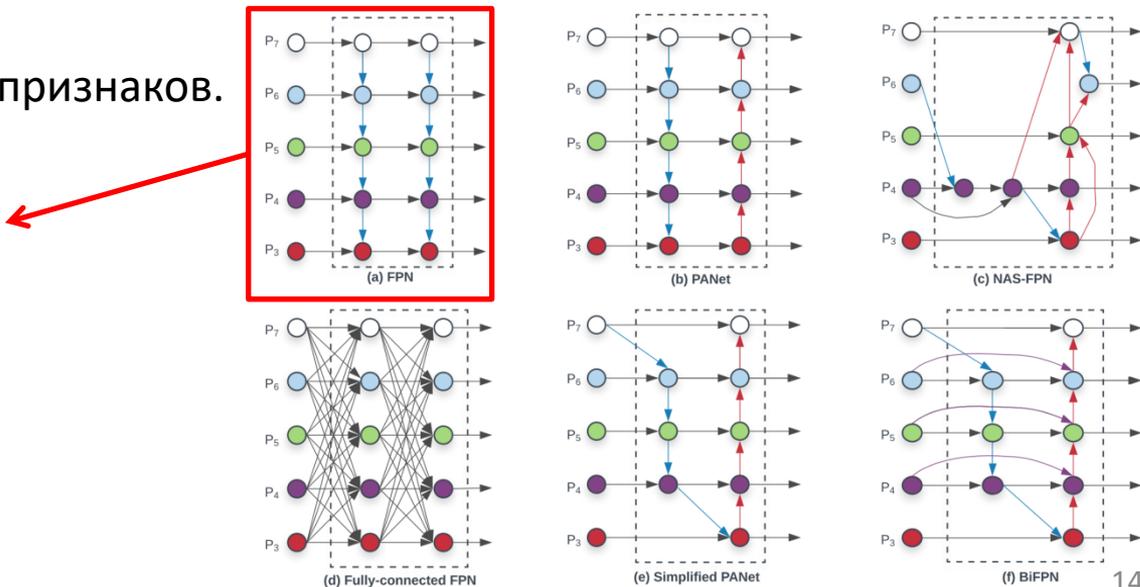


$$P_7^{out} = Conv(P_7^{in})$$

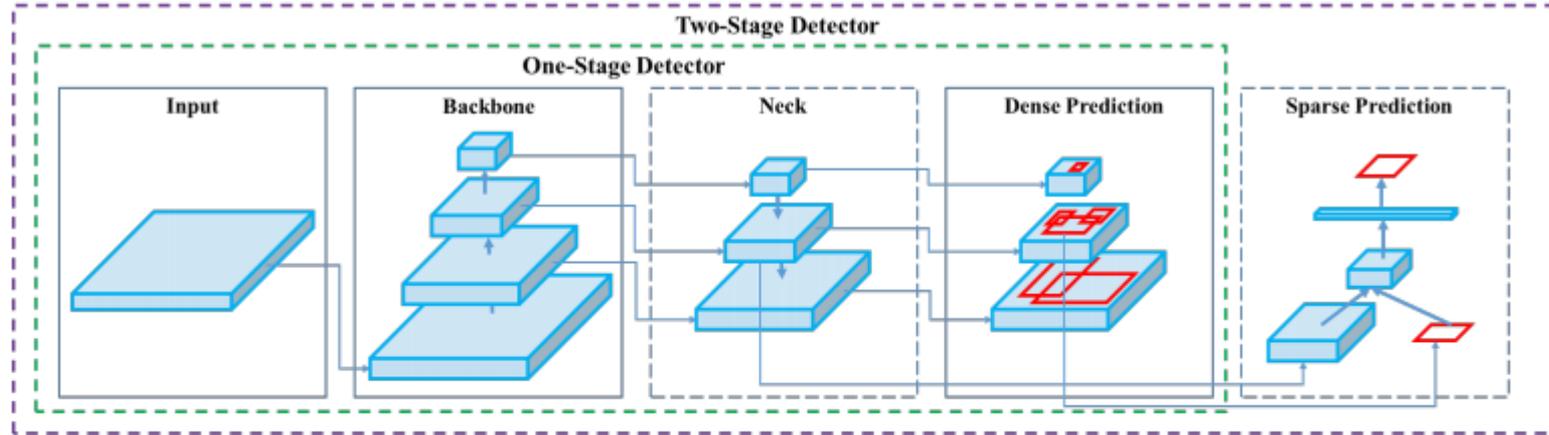
$$P_6^{out} = Conv(P_6^{in} + Resize(P_7^{out}))$$

...

$$P_3^{out} = Conv(P_3^{in} + Resize(P_4^{out}))$$



Структура детектора объектов



Input: { Image, Patches, Image Pyramid, ... }

Backbone: { VGG16 [68], ResNet-50 [26], ResNeXt-101 [86], Darknet53 [63], ... }

Neck: { FPN [44], PANet [49], Bi-FPN [77], ... }

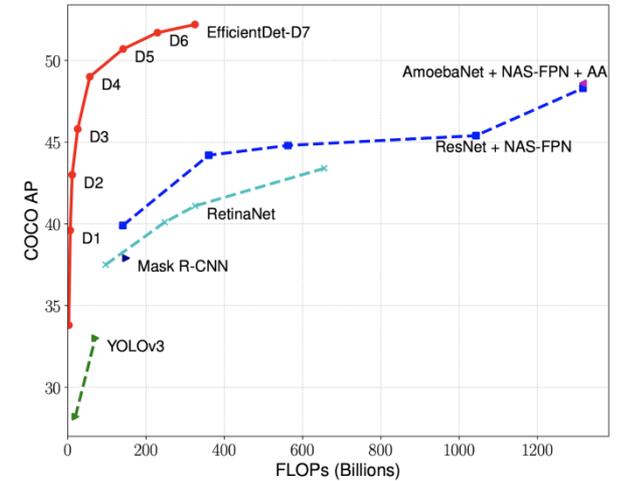
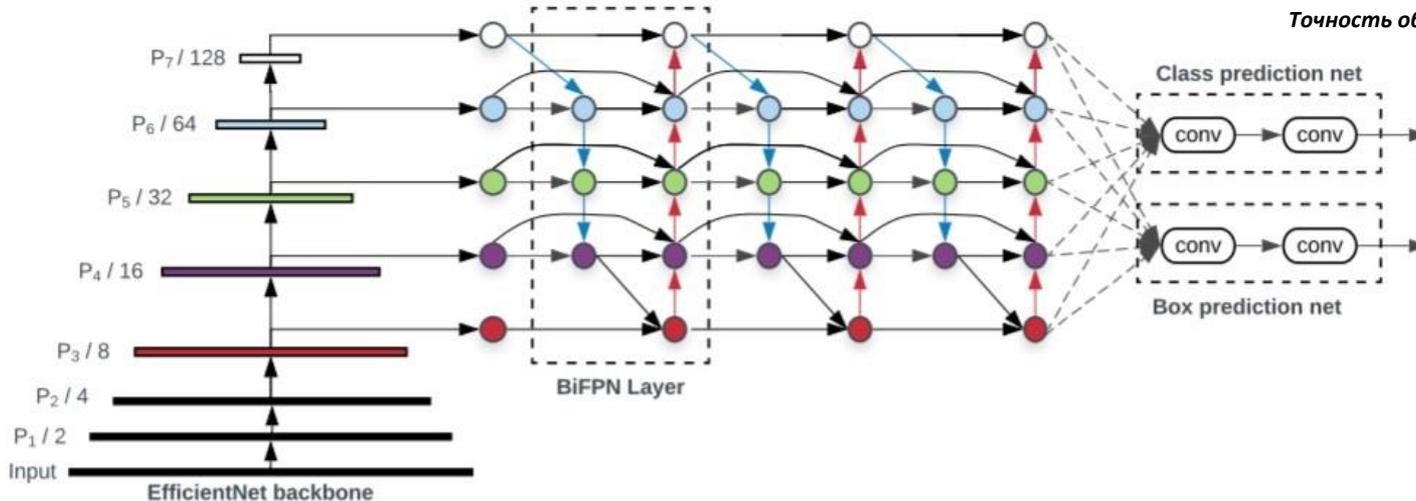
Head:

Dense Prediction: { RPN [64], YOLO [61, 62, 63], SSD [50], RetinaNet [45], FCOS [78], ... }

Sparse Prediction: { Faster R-CNN [64], R-FCN [9], ... }

EfficientDet

- ❑ M. Tan, et al., EfficientDet: Scalable and Efficient Object Detection, 2019.
- ❑ Базовая сеть EfficientNet;
- ❑ Взвешенная двунаправленная пирамида признаков (BiFPN)
- ❑ Смешанное масштабирование.



Точность обнаружения на базе COCO

ССЫЛКИ

- ❑ [A. Ouaknine, Review of Deep Learning Algorithms for Object Detection. 2018](#)
- ❑ R. Girshick et al., Rich feature hierarchies for accurate object detection and semantic segmentation. 2014.
- ❑ J.R.R. Uijlings et al., Selective Search for Object Recognition. 2012.
- ❑ R. Girshick, Fast R-CNN. 2015.
- ❑ S. Ren et al., Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. 2016.
- ❑ J. Dai et al., R-FCN: Object Detection via Region-based Fully Convolutional Networks. 2016.
- ❑ J. Redmon et al., You Only Look Once: Unified, Real-Time Object Detection. 2016.
- ❑ W. Liu et al., SSD: Single Shot MultiBox Detector. 2016.
- ❑ M. Tan, et al., EfficientDet: Scalable and Efficient Object Detection, 2019.

Вопросы

- ❑ Какие основные модули у сети Fast R-CNN?
- ❑ Какое назначение Feature Pyramid Networks?
- ❑ Общая структура детектора объектов на изображениях на основе НС?