Машинное обучение

лекция 13
Segmentation models

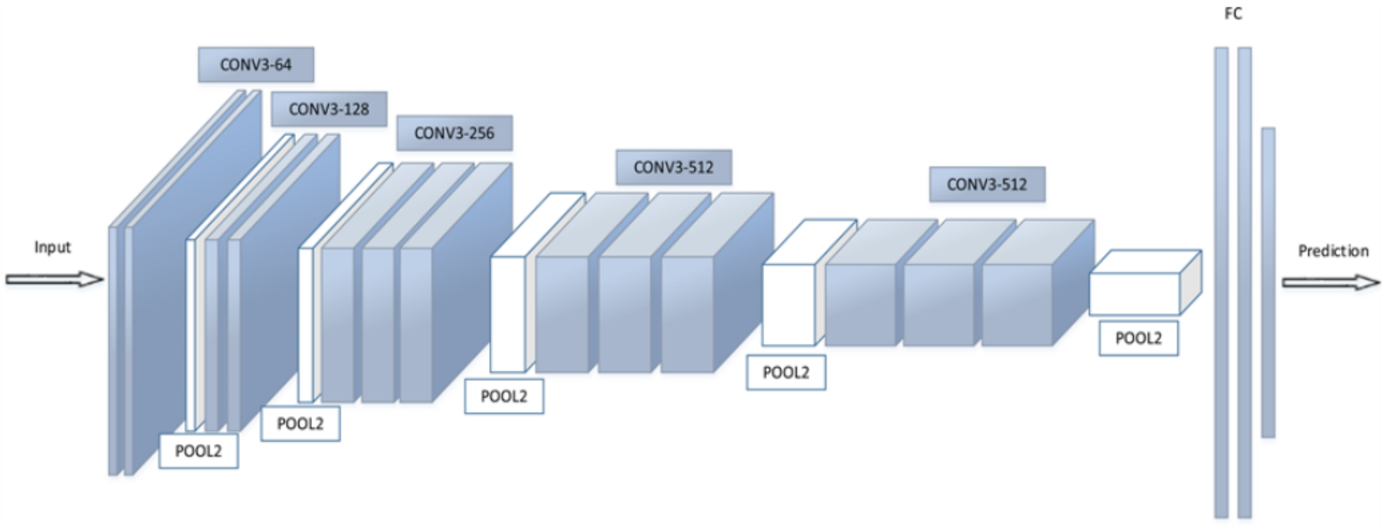
Гришин Никита Александрович

MIPT
MOSCOW INSTITUTE OF PHYSICS AND TECHNOLOGY

ФАЛТ

2019

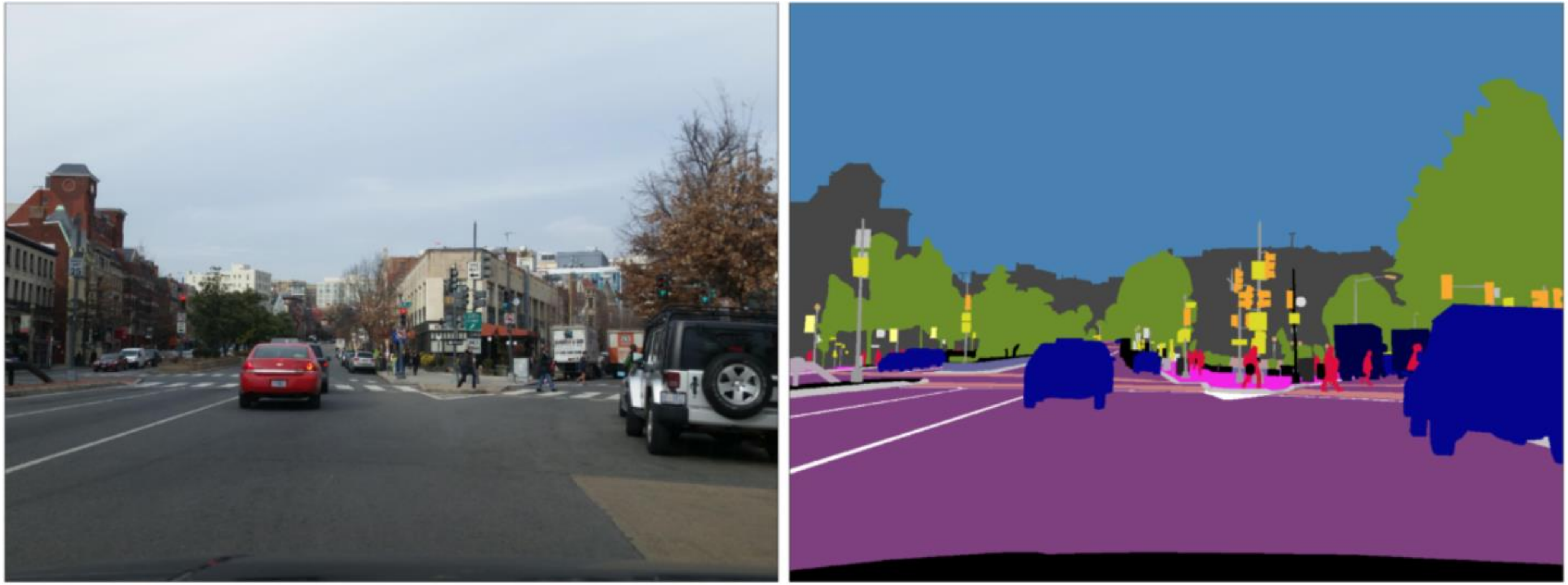
Classification



Cat (0.99)

Dog(0.01)

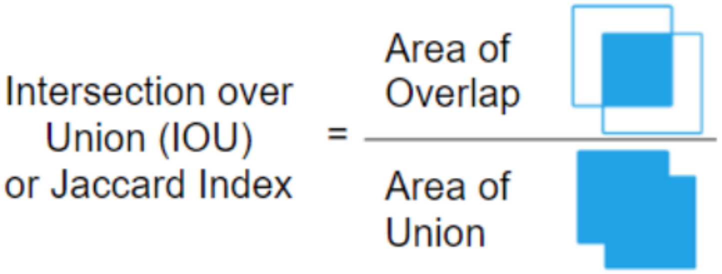
Segmentation



- Сегментация — попиксельная классификация
- Не требует большого количества данных
- Все сегментационные модели — это архитектуры вида FCN

slides by @Vladimir Iglovikov

Segmentation: Metric

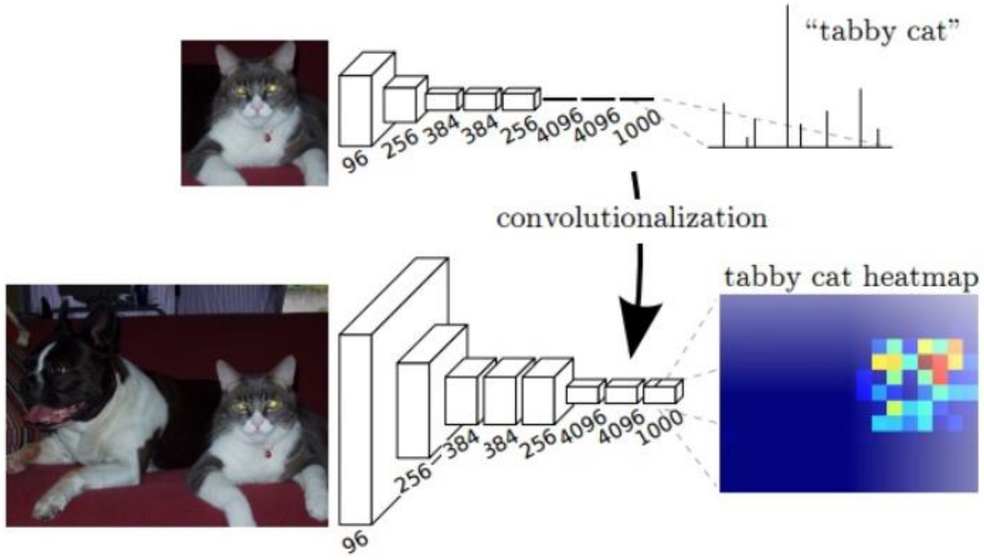


Чаще всего используют Dice - особенно в медицинских снимках и Jaccard (IoU)

Table 1. The three similarity coefficients

Similarity Coefficient (X,Y)	**Actual Formula**
Dice Coefficient	$2\frac{\lvert X \cap Y\rvert}{\lvert X\rvert + \lvert Y\rvert}$
Cosine Coefficient	$\frac{\lvert X \cap Y\rvert}{\lvert X\rvert^{1/2} . \lvert Y\rvert^{1/2}}$
Jaccard Coefficient	$\frac{\lvert X \cap Y\rvert}{\lvert X\rvert + \lvert Y\rvert - \lvert X \cap Y\rvert}$

Fully Convolutional Network: FCN

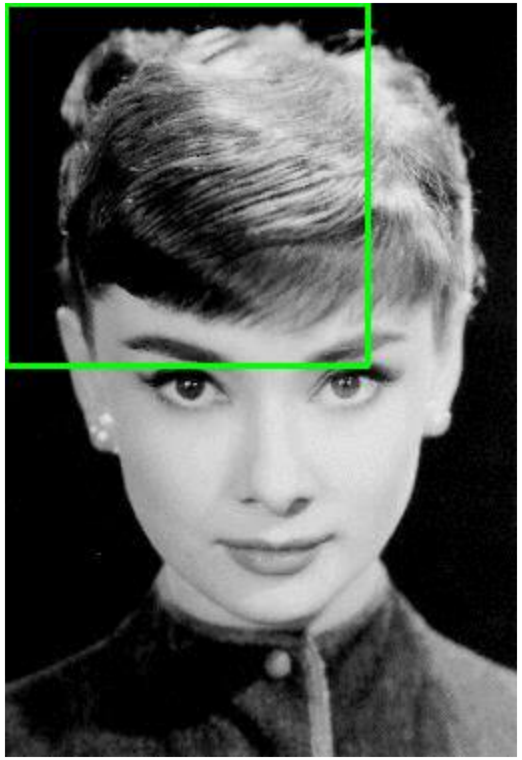


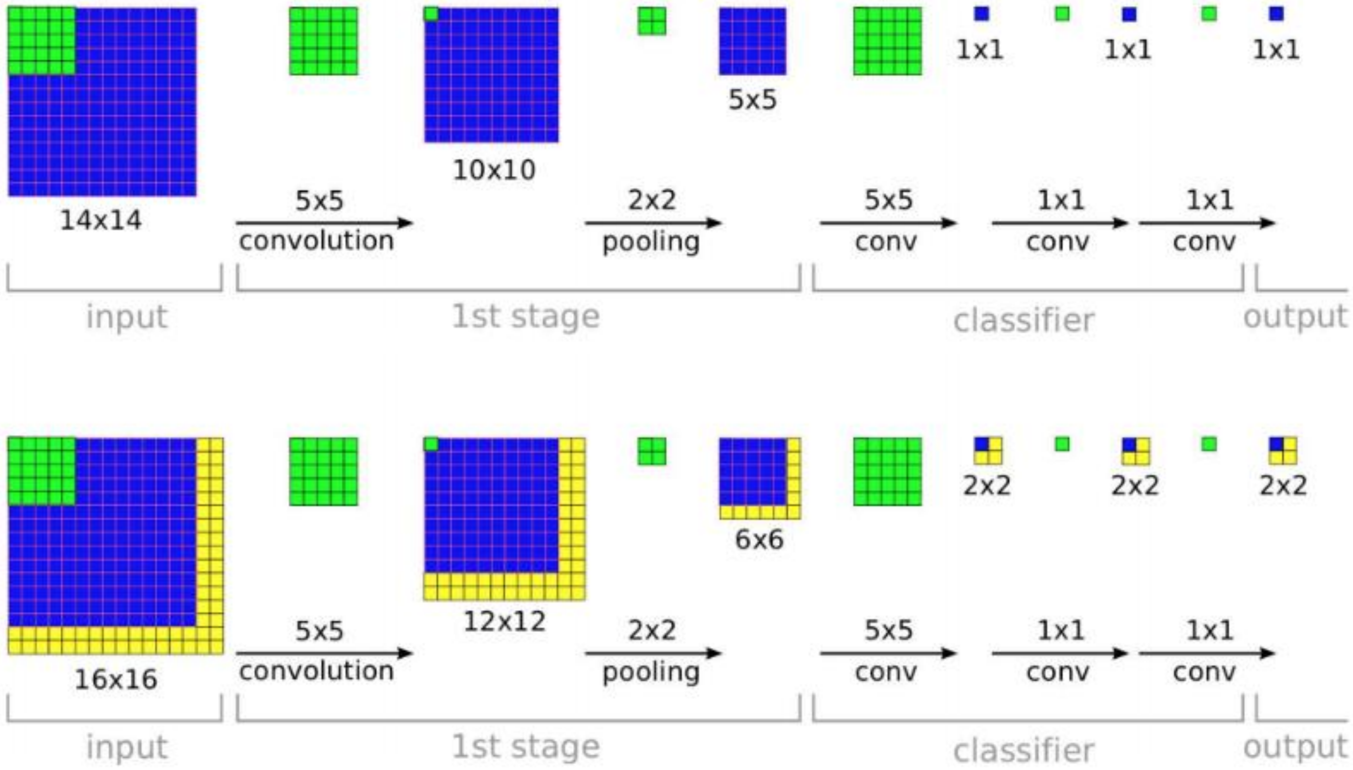
1. Оторвать Dense слой, или Dense => Conv.
2. В сети мало параметров.
3. Берет на вход картинки любого размера.

Jonathan Long, Evan Shelhamer, Trevor Darrell; The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015, pp. 3431-3440

slides by @Vladimir Iglovikov

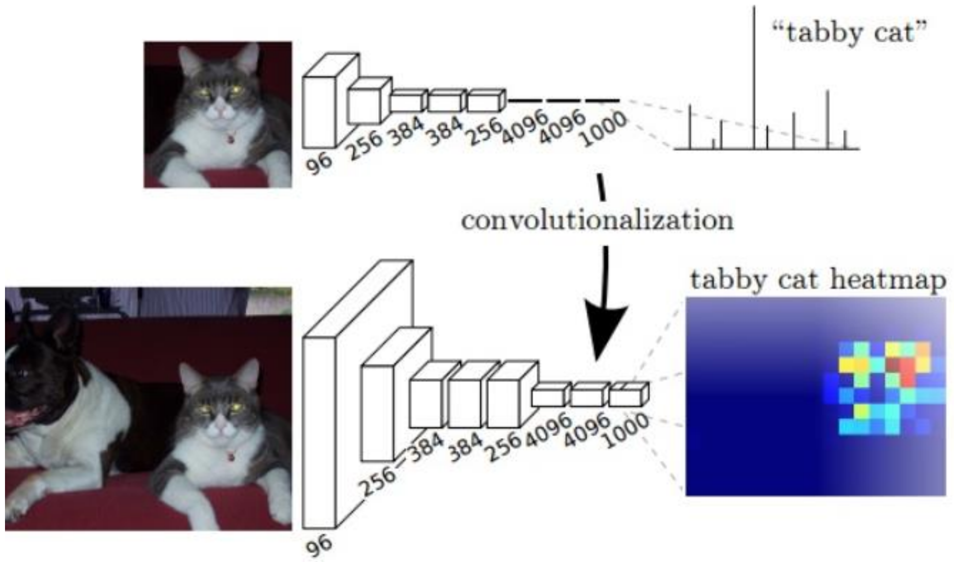
FCN = Efficient Sliding Window

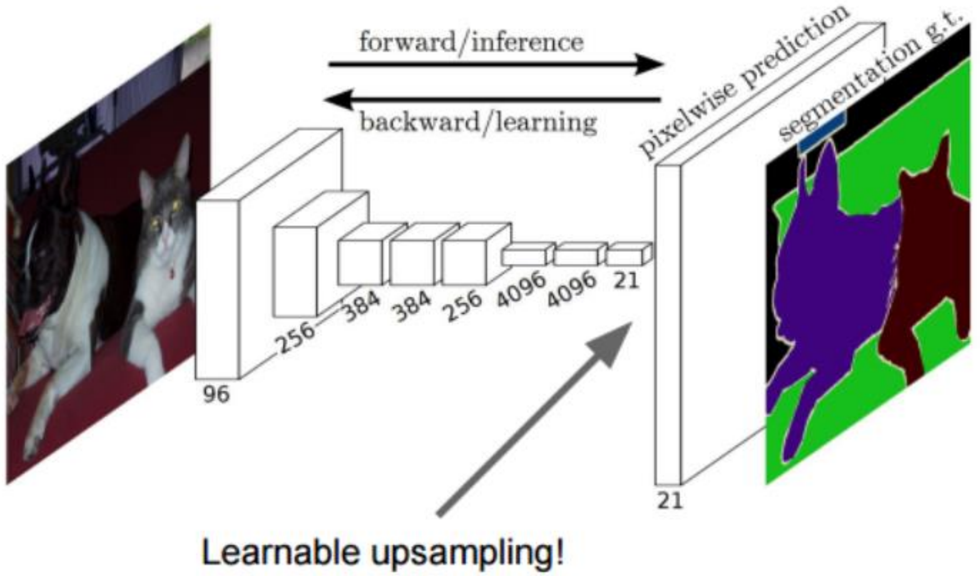




slides by @Vladimir Iglovikov

Classification to Segmentation



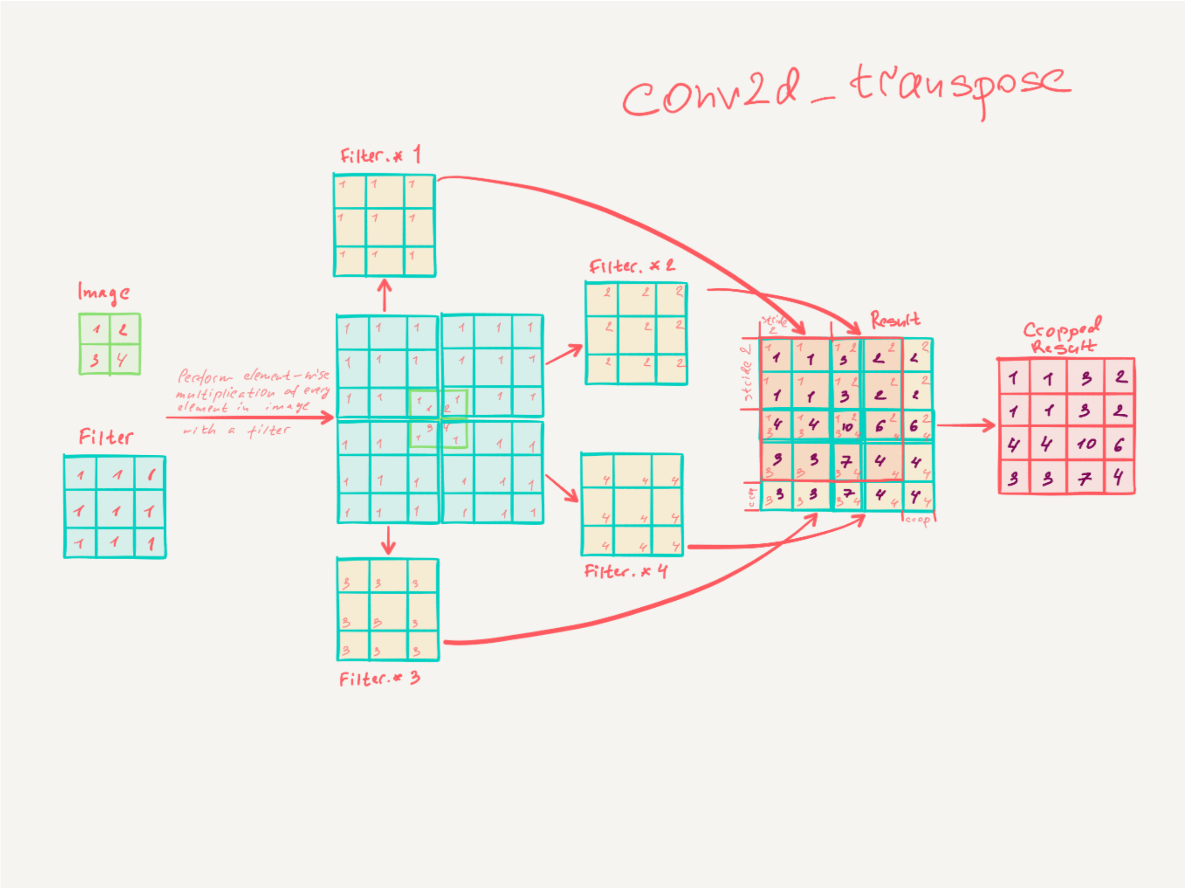


1. Оторвать Dense или Dense => Conv
2. Добавить декодер

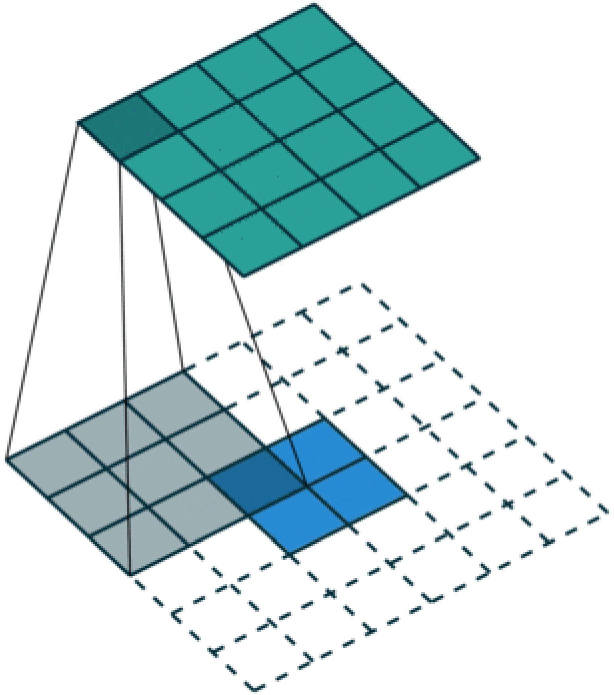
Jonathan Long, Evan Shelhamer, Trevor Darrell; The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015, pp. 3431-3440

slides by @Vladimir Iglovikov

Upsampling Understanding

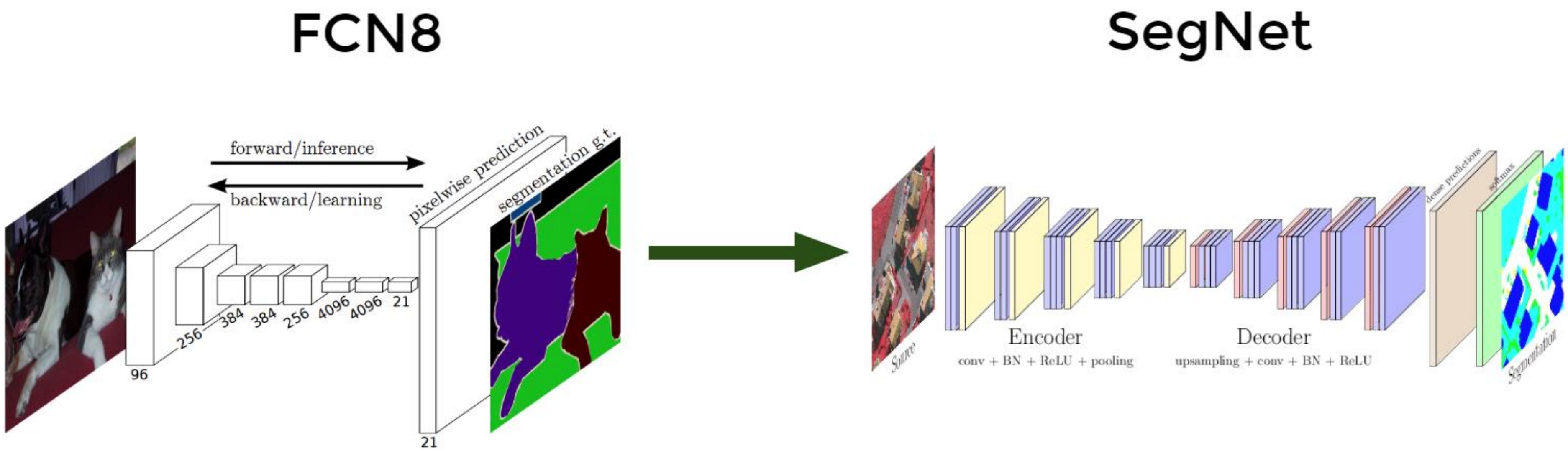


Stride = 2
Pad = 0



Stride = 1
Pad = 0

FCN8 to SegNet



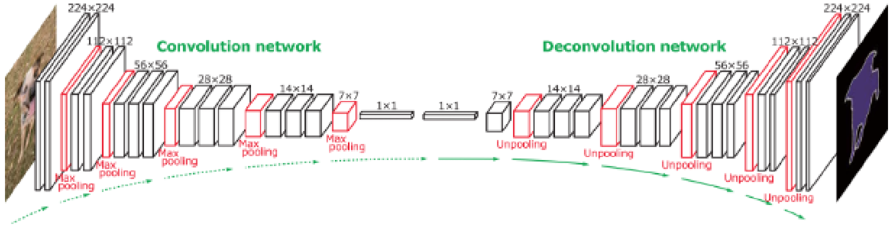
Заменить Upsampling на иерархический Upsampling

V. Badrinarayanan, A. Kendall, and R. Cipolla, “Segnet: A deep convolutional encoder-decoder architecture for image segmentation,” arXiv:1511.00561, 2015

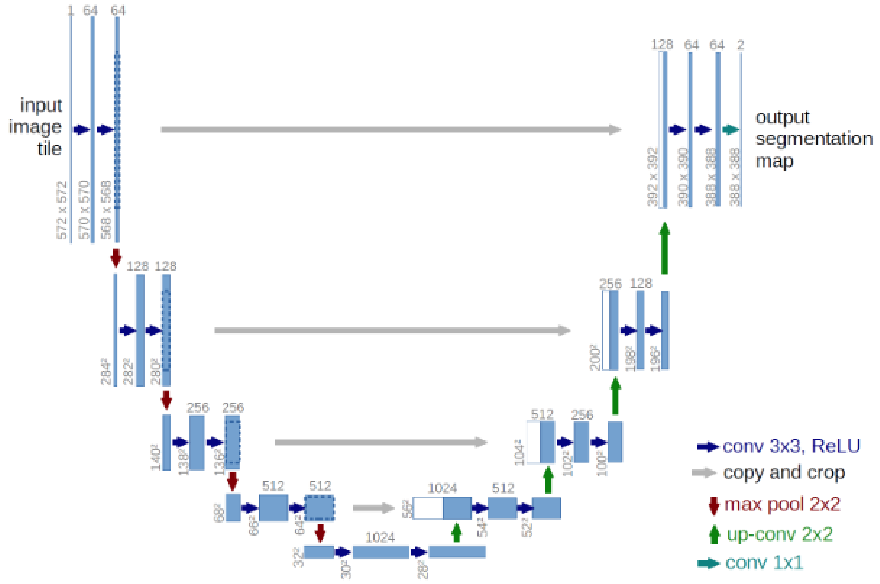
slides by @Vladimir Iglovikov

SegNet to UNet

SegNet



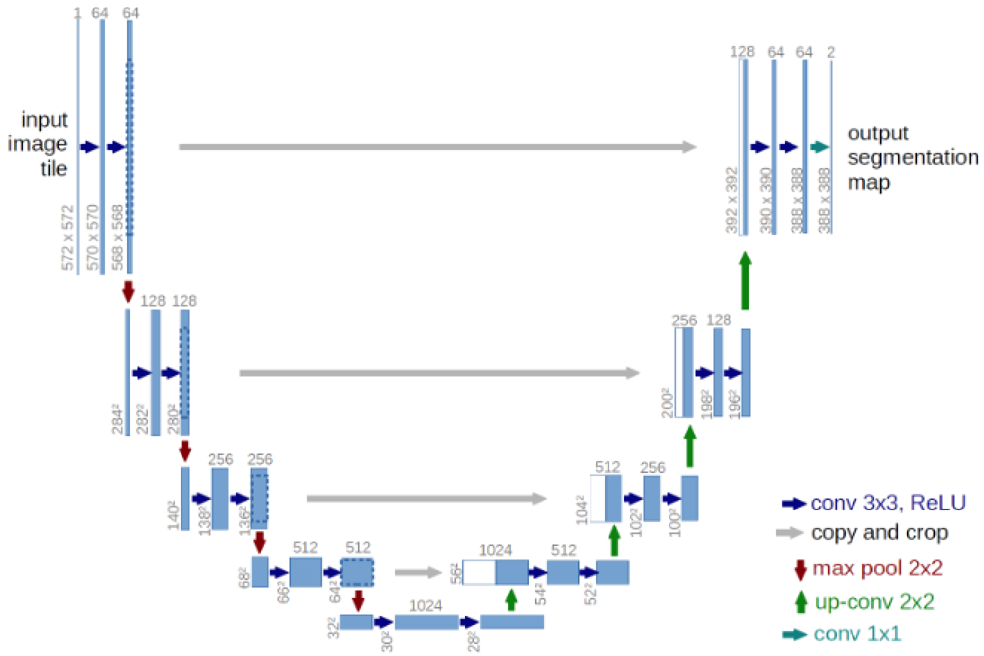
UNet



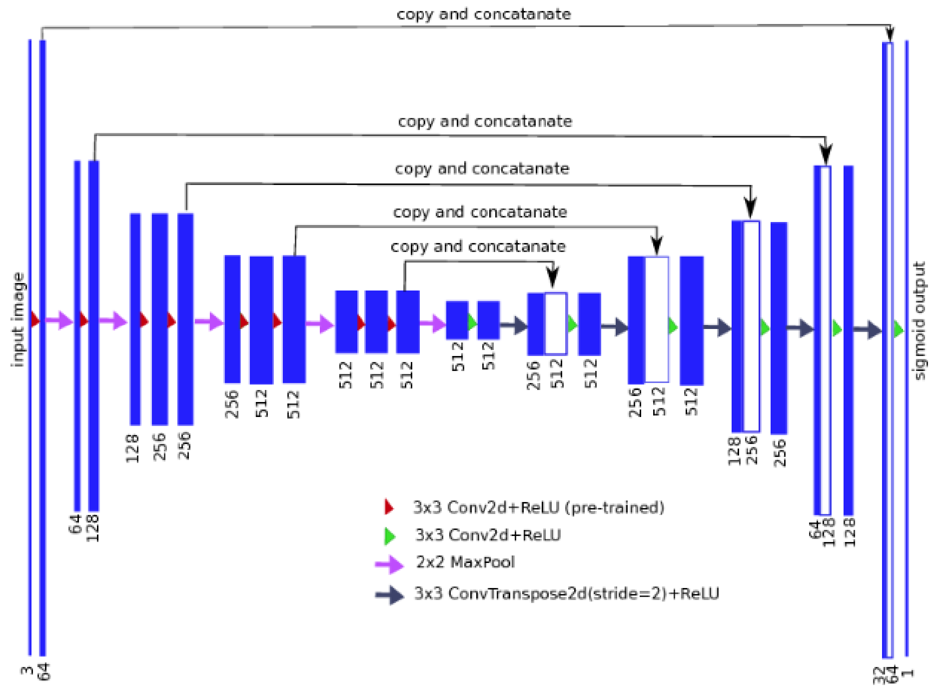
Added skip connections

O. Ronneberger P. Fischer T. Brox "U-net: Convolutional networks for biomedical image segmentation" Proc. Med. Image Comput. Comput.-Assisted Intervention pp. 234-241 2015.

Unet => TernausNet



Text



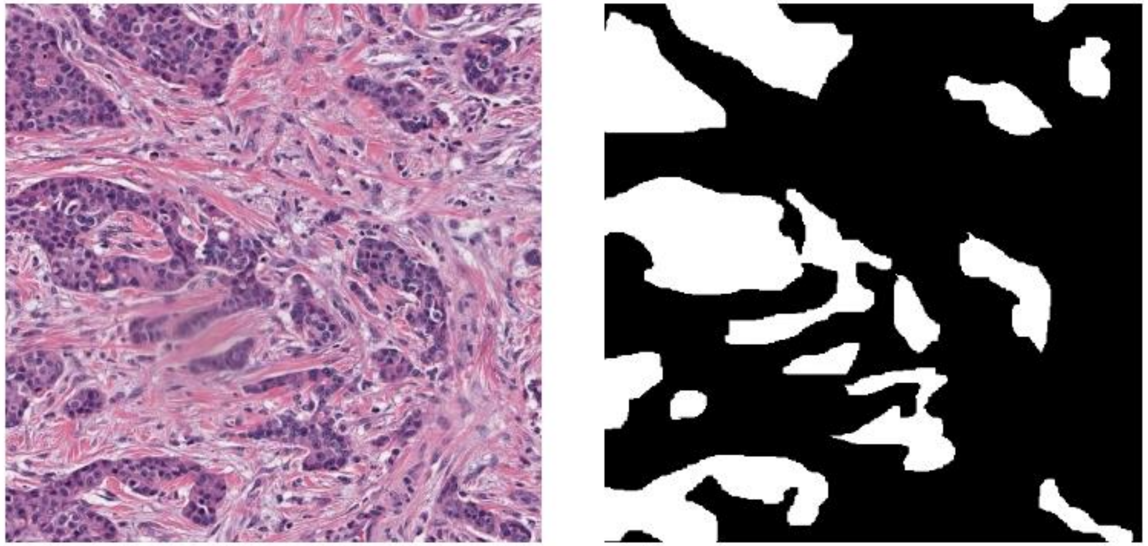
Энкодер инициализируем весами с ImageNet

Iglovikov, V., Shvets, A.: Ternausnet: U-net with vgg11 encoder pre-trained on imagenet for image segmentation. arXiv preprint arXiv:1801.05746 (2018)

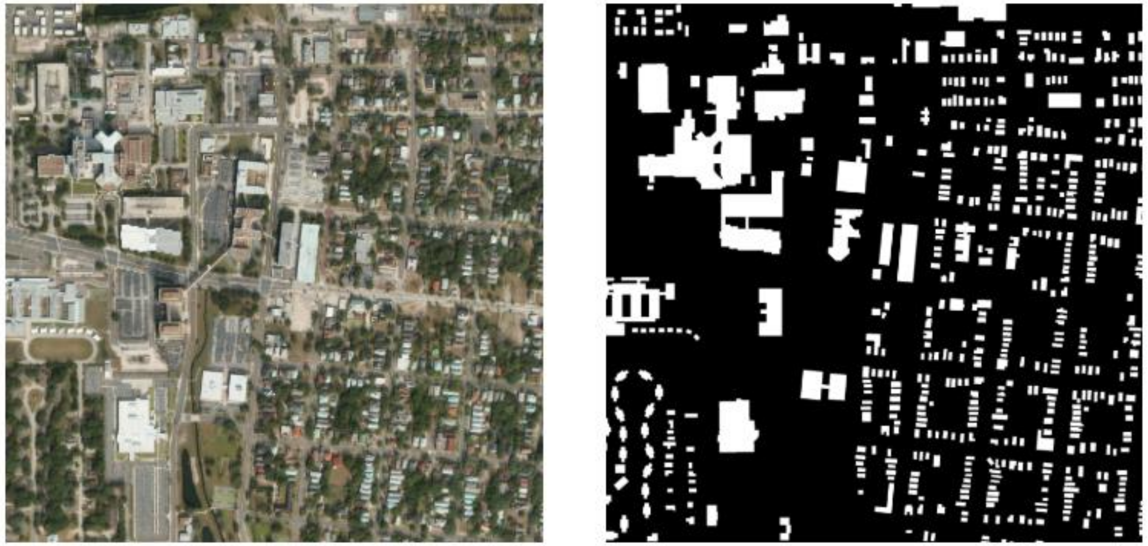
slides by @Vladimir Iglovikov

Segmentation => UNet

Medical Imaging

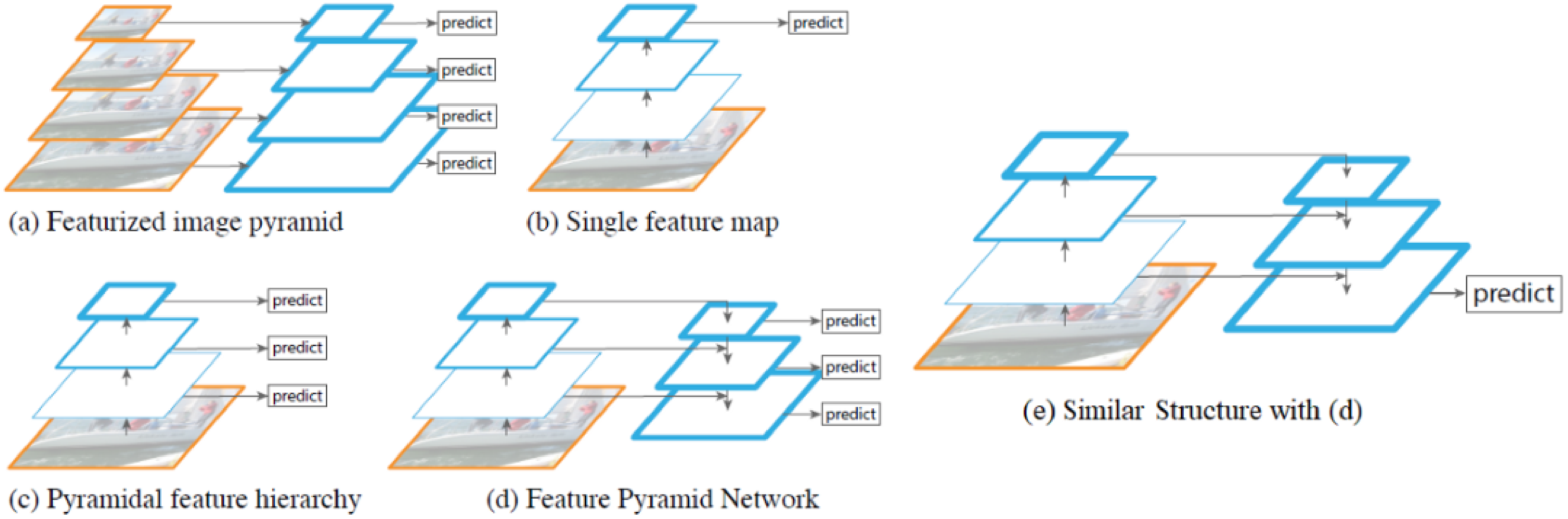


Satellite Imaging



slides by @Vladimir Iglovikov

Feature Pyramid Networks (FPN)

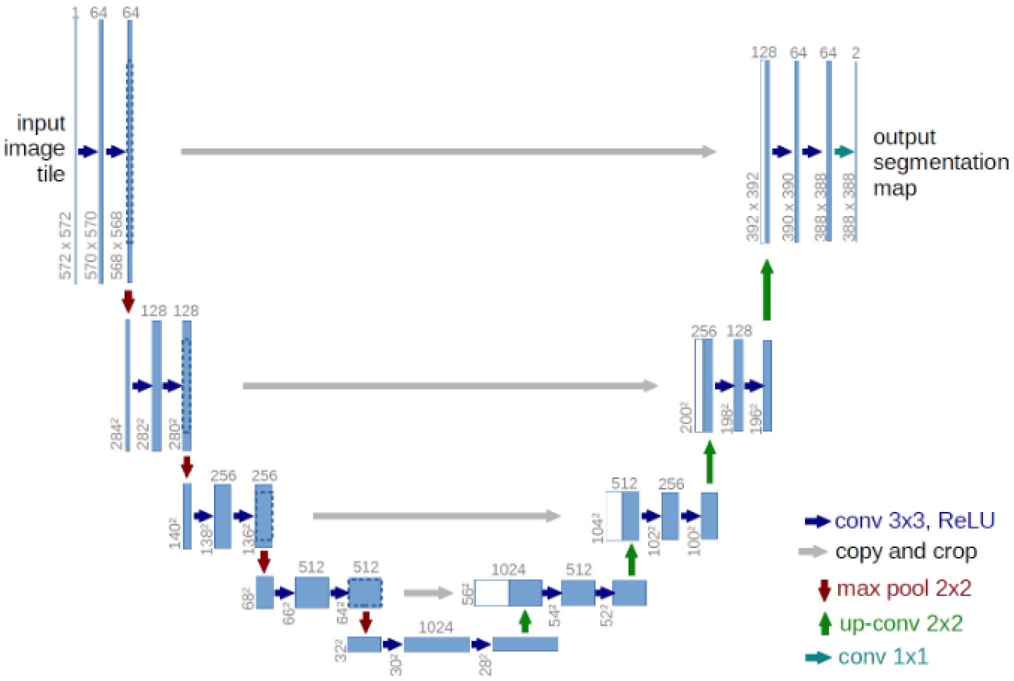


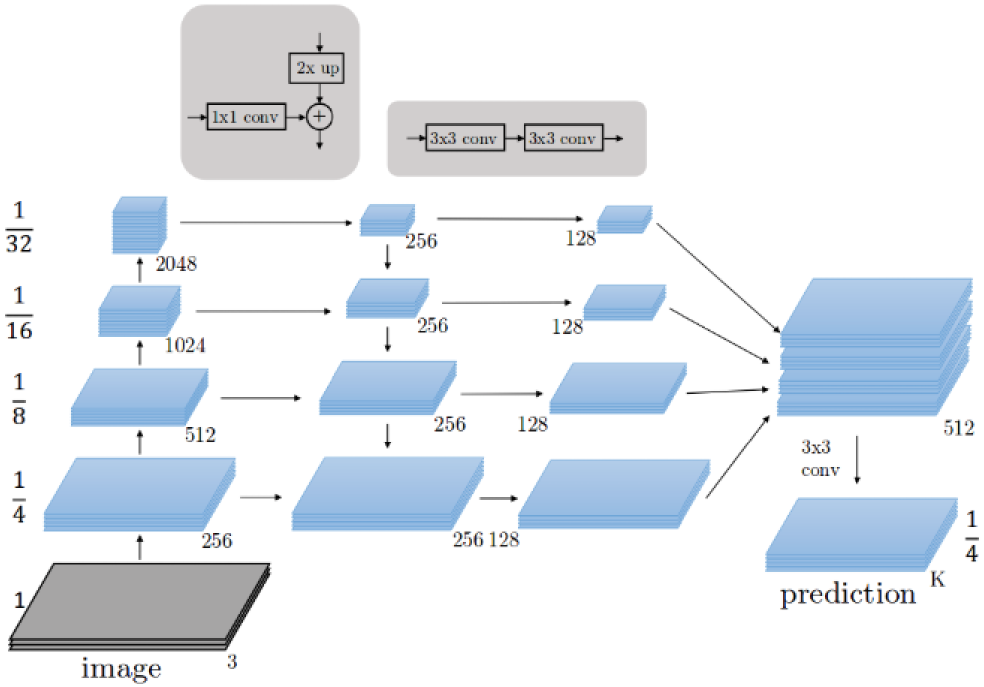
1. Легко добавить во многие архитектуры.
2. Помогает с multiscale

Tsung-Yi Lin, Piotr Dollar, Ross Girshick, Kaiming He, Bharath Hariharan, Serge Belongie; The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, pp. 2117-2125

slides by @Vladimir Iglovikov

Unet + FPN





slides by @Vladimir Iglovikov

Segmentation Loss Function

Каждый пиксель классификатор =>
Categorical / Binary Cross Entropy(CCE, BCE)

Но! Метрика Dice / Jaccard

Dice / Jaccard недифференцируемы =>

Soft Dice / Soft Jaccard

и добавляем в loss

Lovasz-Softmax loss

Использовать для FineTune

$$CCE = \sum_c p(x) \log q(x)$$

$$LOSS = BCE - \ln(DICE)$$

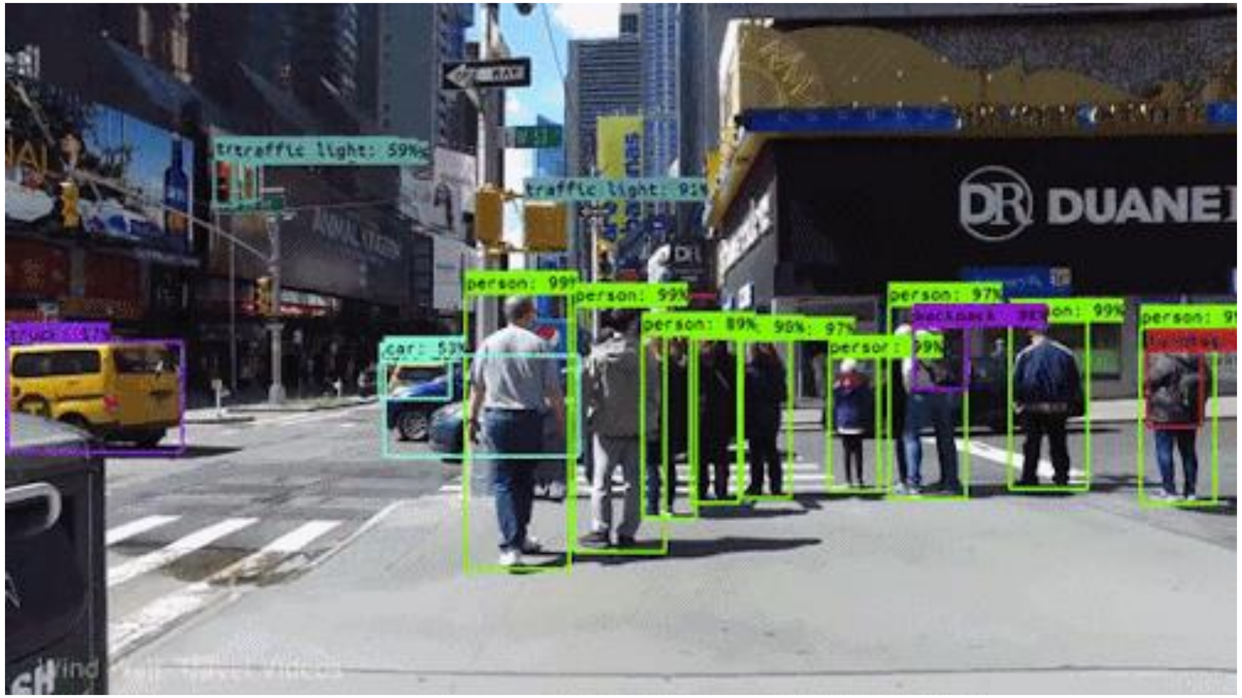
$$BCE = -\sum_i (y_i \ln(p_i) + (1 - y_i) \ln(1 - p_i))$$

$$DICE = 2 \frac{\sum_i y_i p_i}{\sum_u y_i + \sum p_i}$$

Berman, M., Rannen Triki, A., Blaschko, M.B.: The lovász-softmax loss: a tractable surrogate for the optimization of the intersection-over-union measure in neural networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 4413–4421 (2018)

slides by @Vladimir Iglovikov

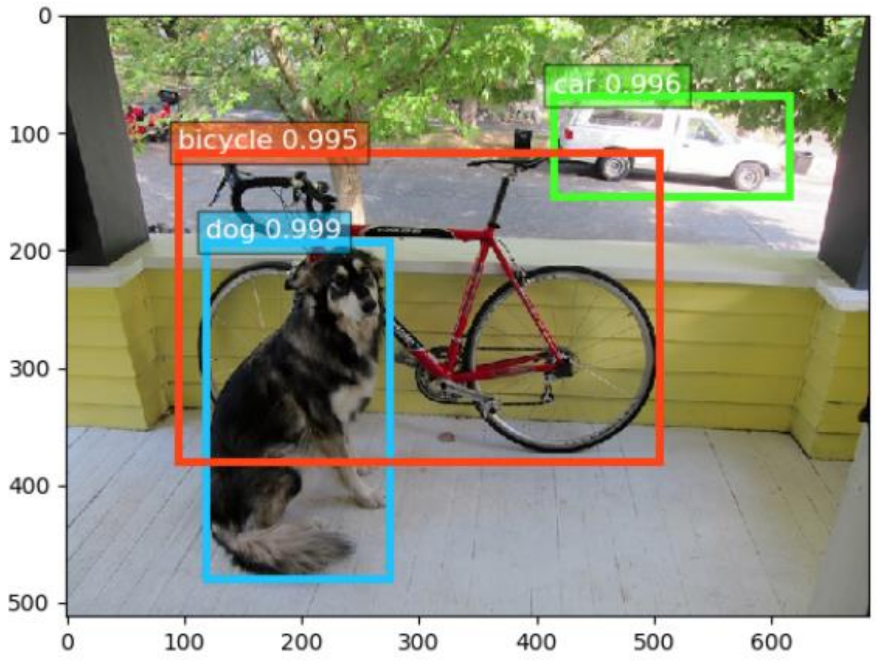
Detection

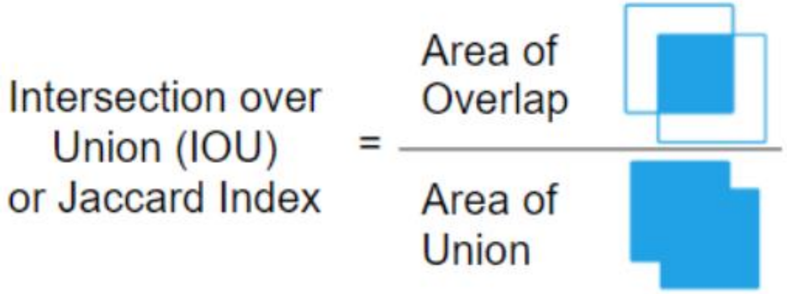


Предсказываем:

1. Координаты боксов
2. Класс
3. Атрибуты

Detection Metric: mAP



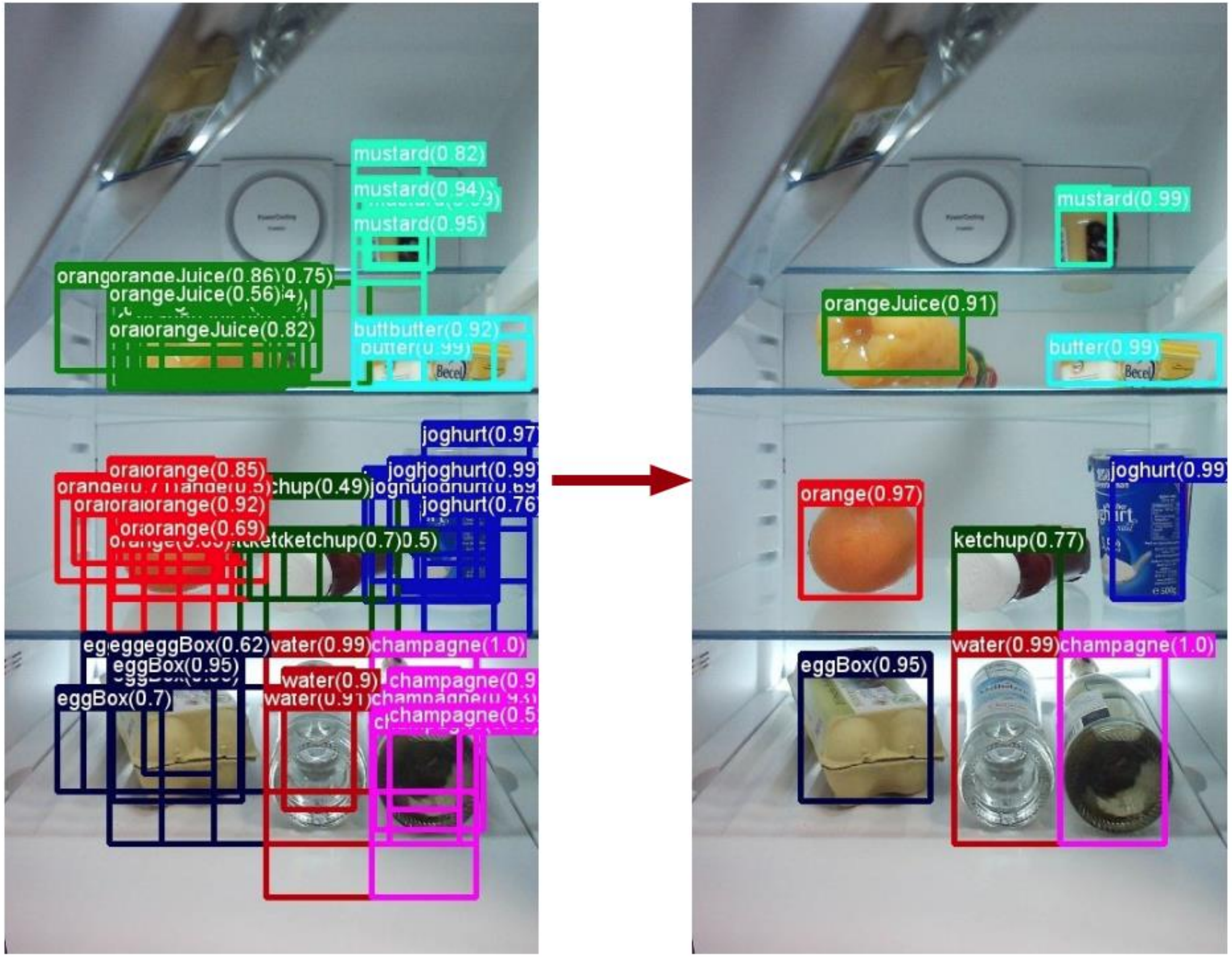


Для каждого класса:

1. Хотим получить Precision Recall Curve
2. Цикл по трешхолдам confidence
3. Трешхолд по IoU => TP, FP, FN
4. mAP = area under PR Curve

Non Maximum Suppression (NMS)

Detection = Предсказываем много боксов, а потом фильтруем



slides by @Vladimir Iglovikov

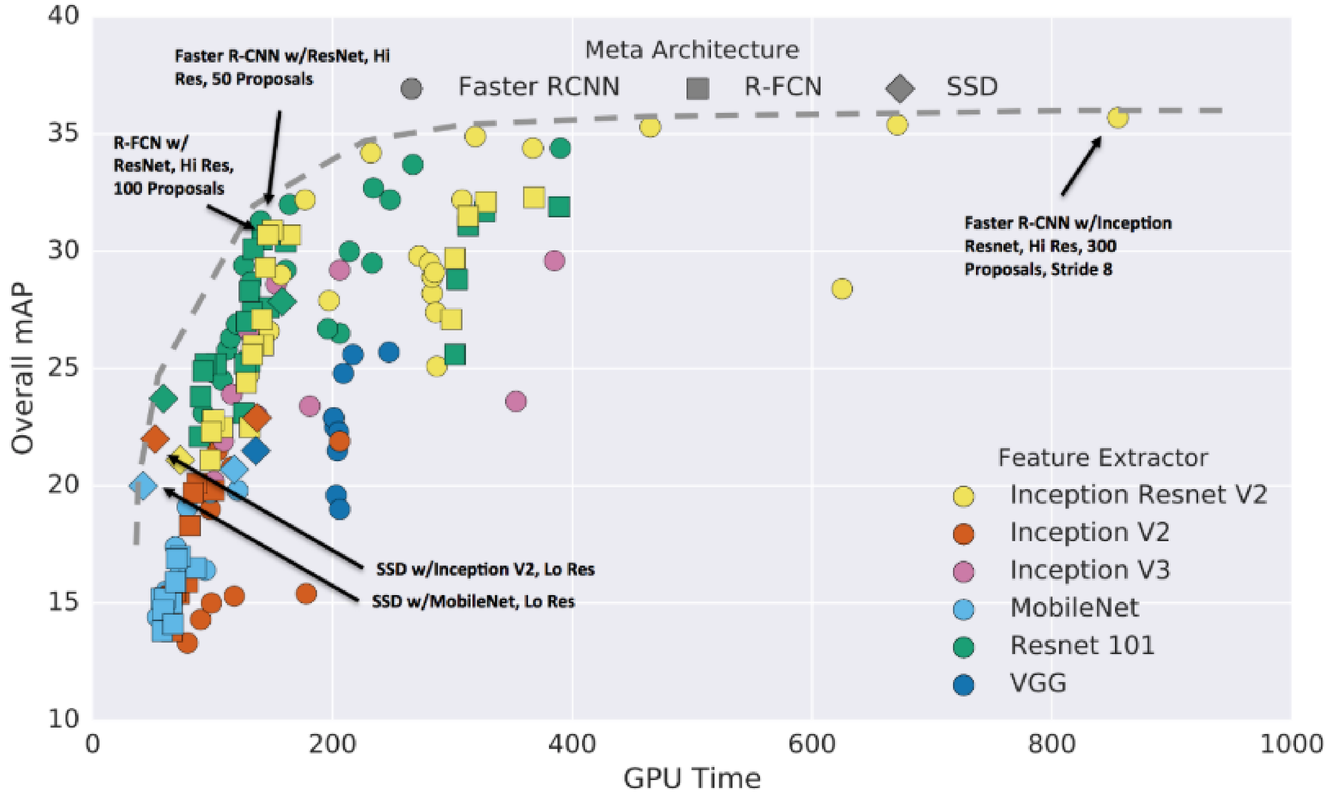
Detection

One-shot (быстрые)

YOLO, SSD, RetinaNet, SqueezeNet, DetectNet

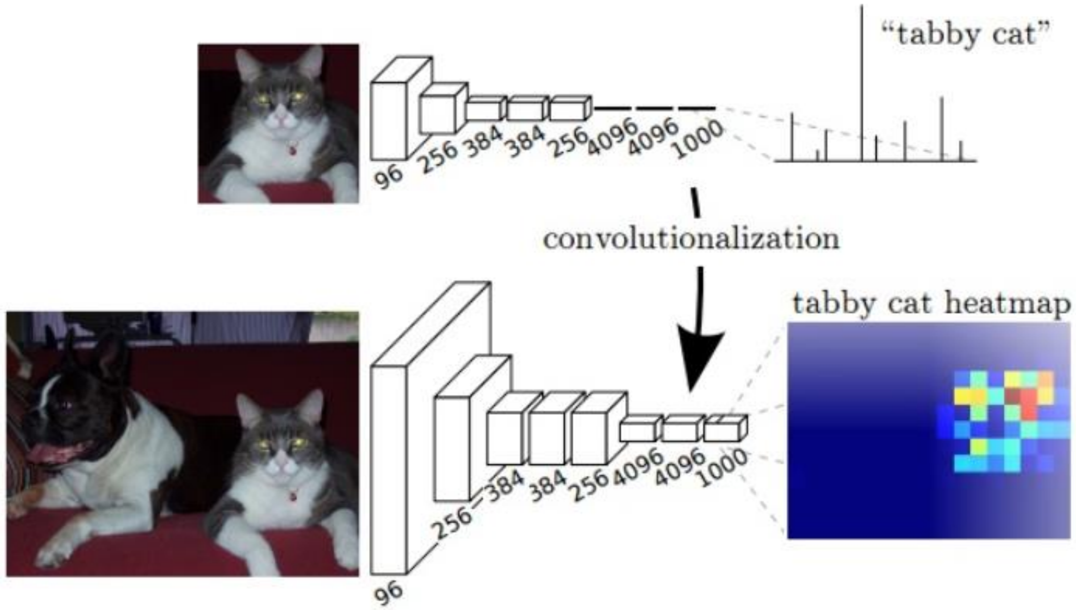
Two-shot (точные)

R-FCN, Fast RCNN, Faster-RCNN



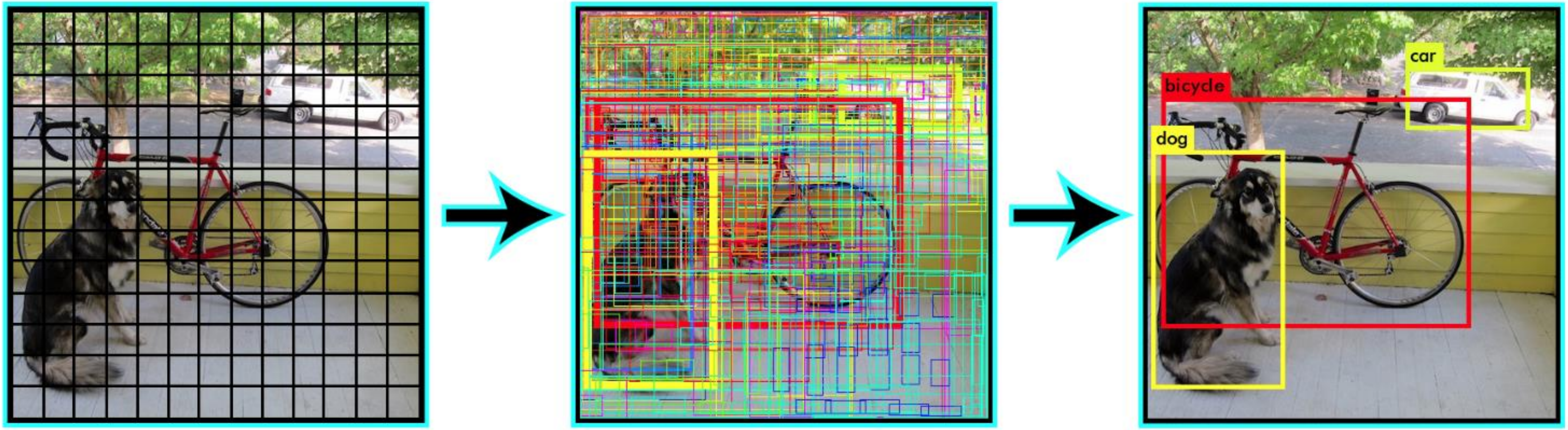
slides by @Vladimir Iglovikov

One Shot Detection



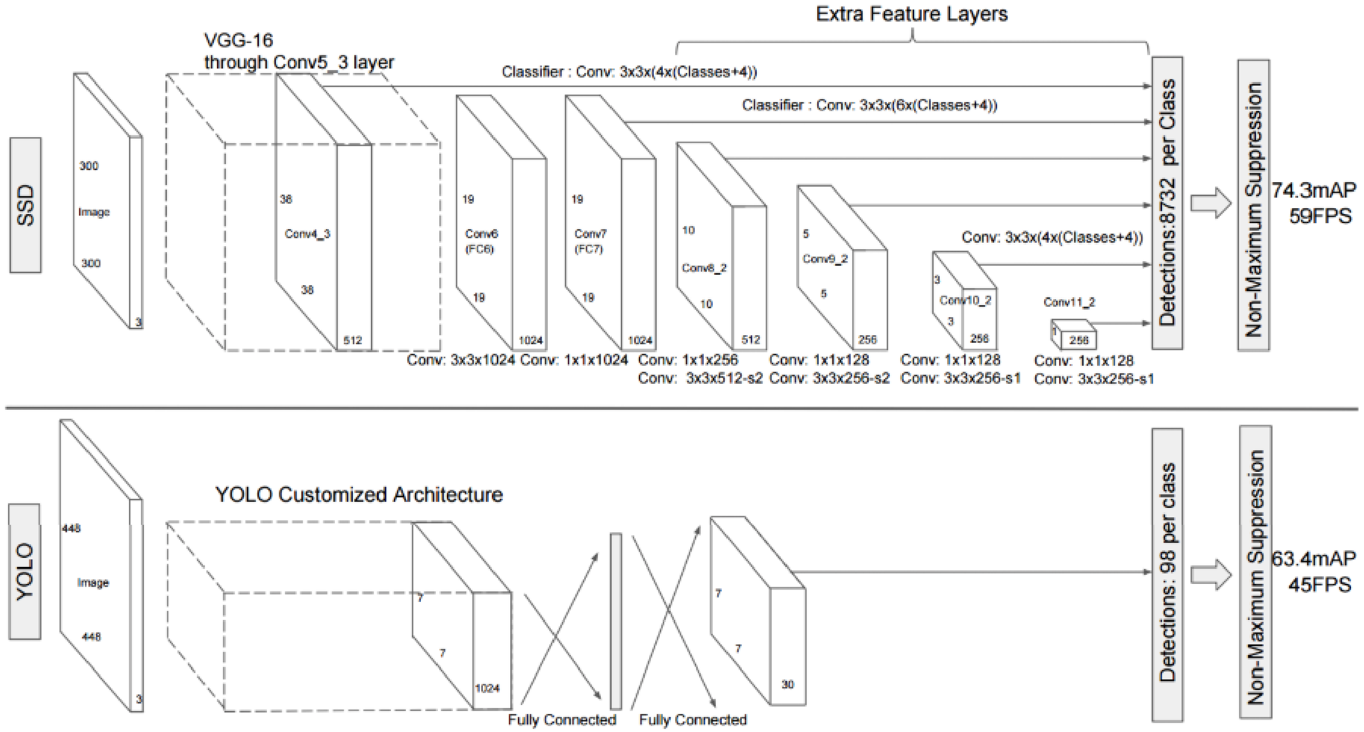
Для каждой ячейки в последнем conv слое предказываем координаты боксa и класс объекта с центром в ячейке.

One Shot Detector: YOLO



Для каждой ячейки в последнем conv слое предказываем координаты бокса и класс объекта с центром в ячейке.

One Shot Detector with FPN = SSD

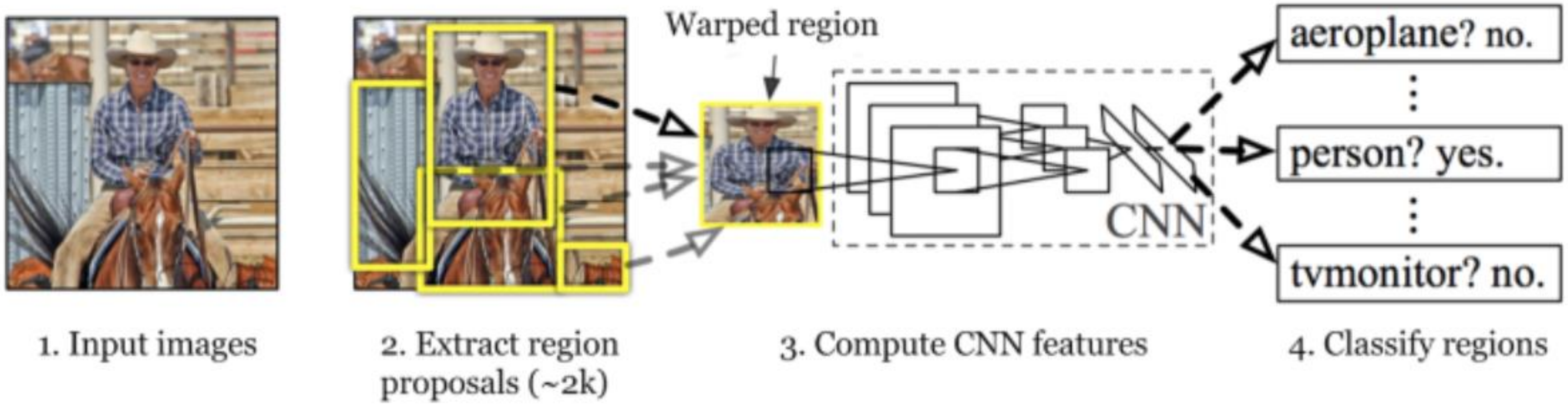


	mAP	FPS
YOLO v2	21.6	91
SSD	28.0	59

slides by @Vladimir Iglovikov

Two Shot Detector: R-CNN

R-CNN = Selective Search
+ Classification



Ross Girshick, Jeff Donahue, Trevor Darrell, Jitendra Malik; The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2014, pp. 580-587

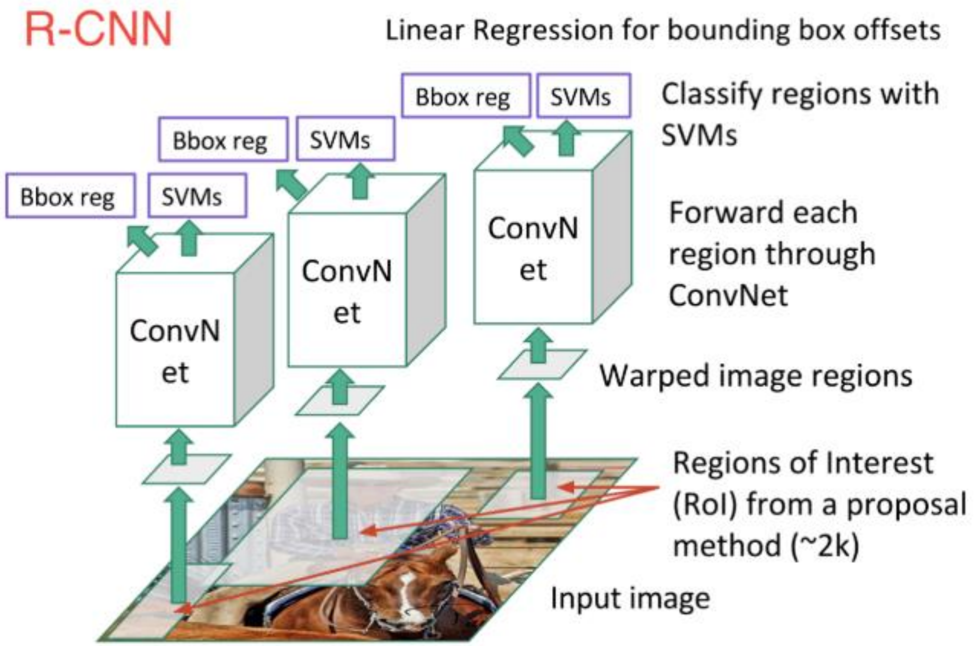
slides by @Vladimir Iglovikov

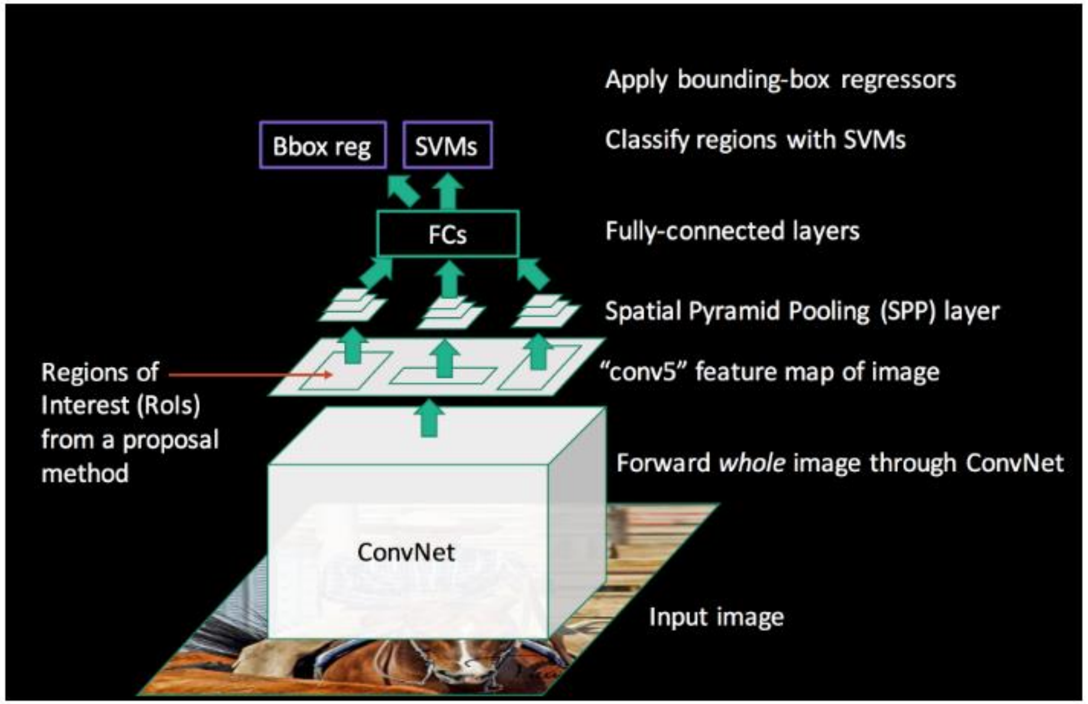
Two Shot Detector: Fast-RCNN

R-CNN => Fast R-CNN

Меняем порядок Crop и ConvNet

50 секунд => 2 секунды (25 раз быстрее)



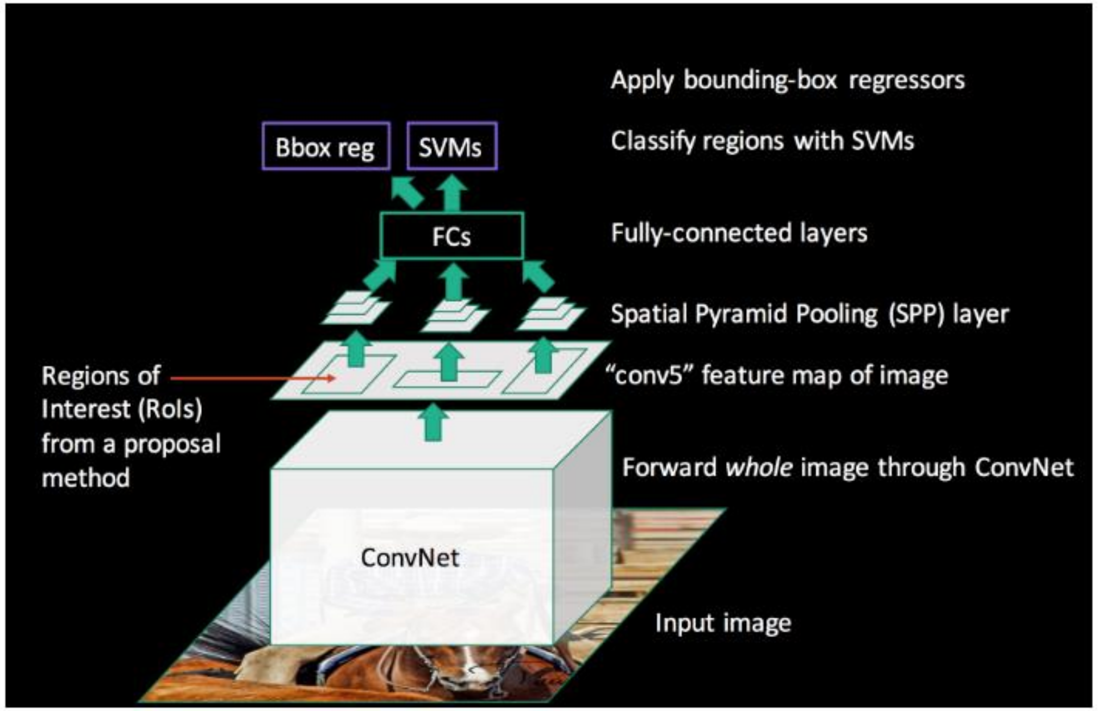


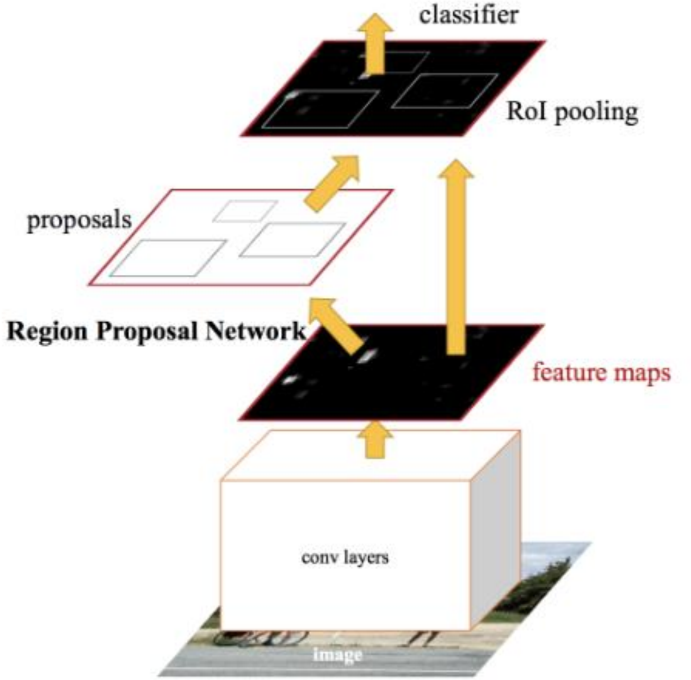
Fast => Faster

Fast R-CNN => Faster R-CNN

Вычисляем proposals самой сетью.

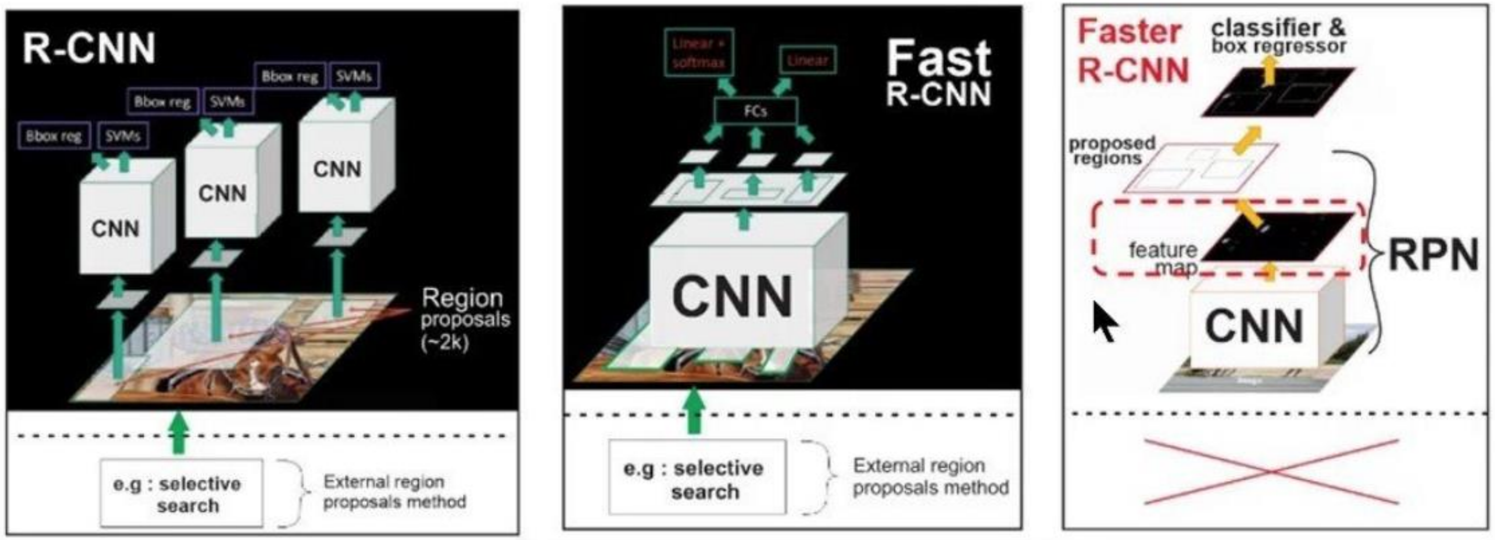
2 секунды => 0.2 секунды (10 раз быстрее)





slides by @Vladimir Iglovikov

Two Shot Detector: performance



	R-CNN	Fast R-CNN	Faster R-CNN
Test time per image	50 seconds	2 seconds	0.2 seconds
Speed-up	1x	25x	250x
mAP (VOC 2007)	66.0%	66.9%	66.9%