

Занятие № 11

# Свёрточные нейронные сети. Часть 3

# План занятия



1. Задачи детектирования и сегментации объектов
2. Модель RCNN
3. Модель Fast RCNN
4. Модель Faster RCNN
5. SegNet
6. U-net
7. Модель Mask RCNN

# Задачи обработки изображений

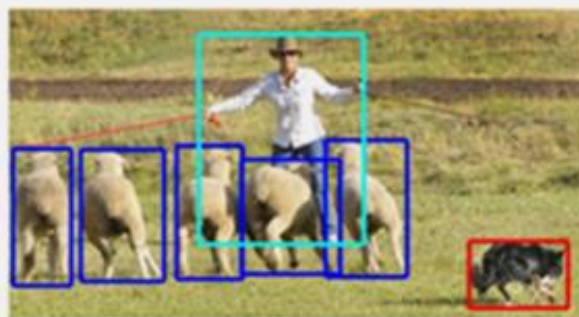


Основные задачи обработки изображений:

- а. Классификация объектов
- б. Детектирование объектов**
- с. Сегментация объектов



(a) classification



(b) detection



(c) segmentation

# Соревнование PASCAL Visual Object Classes



Цель соревнования:

Для 20 классов объектов надо:

1. Предсказать наличие объектов на изображении
2. Для каждого объекта предсказать рамку вокруг и метку класса

Объем данных:

~27k объектов на ~11k изображений

# Соревнование PASCAL Visual Object Classes



Классы объектов:

- Person: person
- Animal: bird, cat, cow, dog, horse, sheep
- Vehicle: airplane, bicycle, boat, bus, car, motorbike, train
- Indoor: bottle, chair, dining table, potted plant, sofa, TV/monitor

Метрика соревнования: Mean Average Precision (mAP)

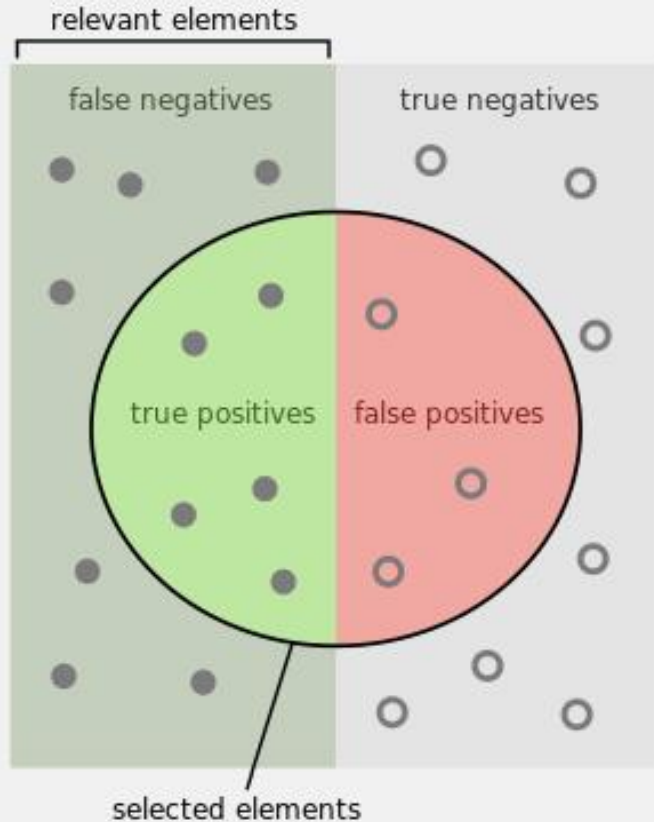
# Intersection over Union (IoU)



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



# Precision & Recall



How many selected items are relevant?

$$\text{Precision} = \frac{\text{true positives}}{\text{true positives} + \text{false positives}}$$

How many relevant items are selected?

$$\text{Recall} = \frac{\text{true positives}}{\text{true positives} + \text{false negatives}}$$

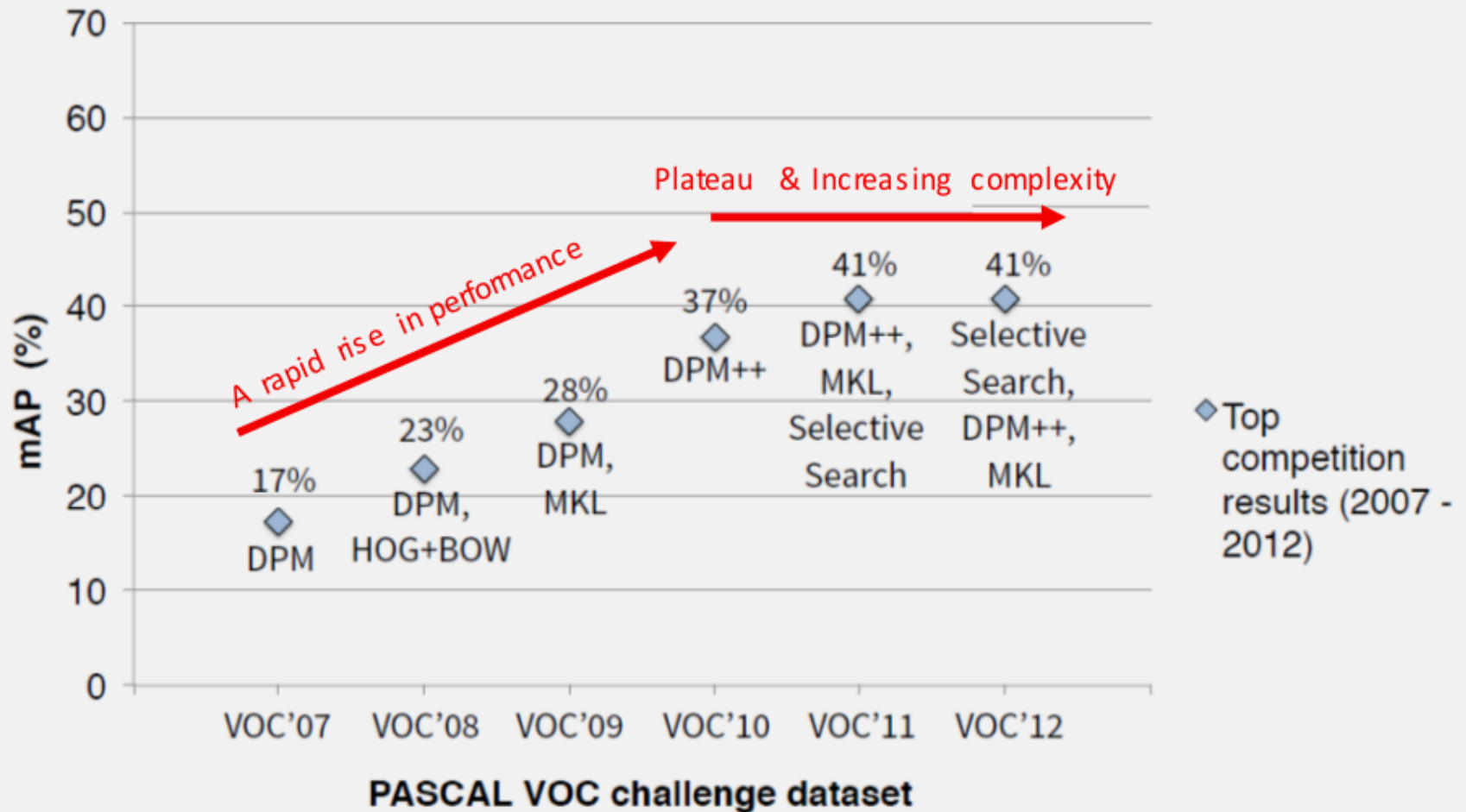
# Mean Average Precision (mAP)



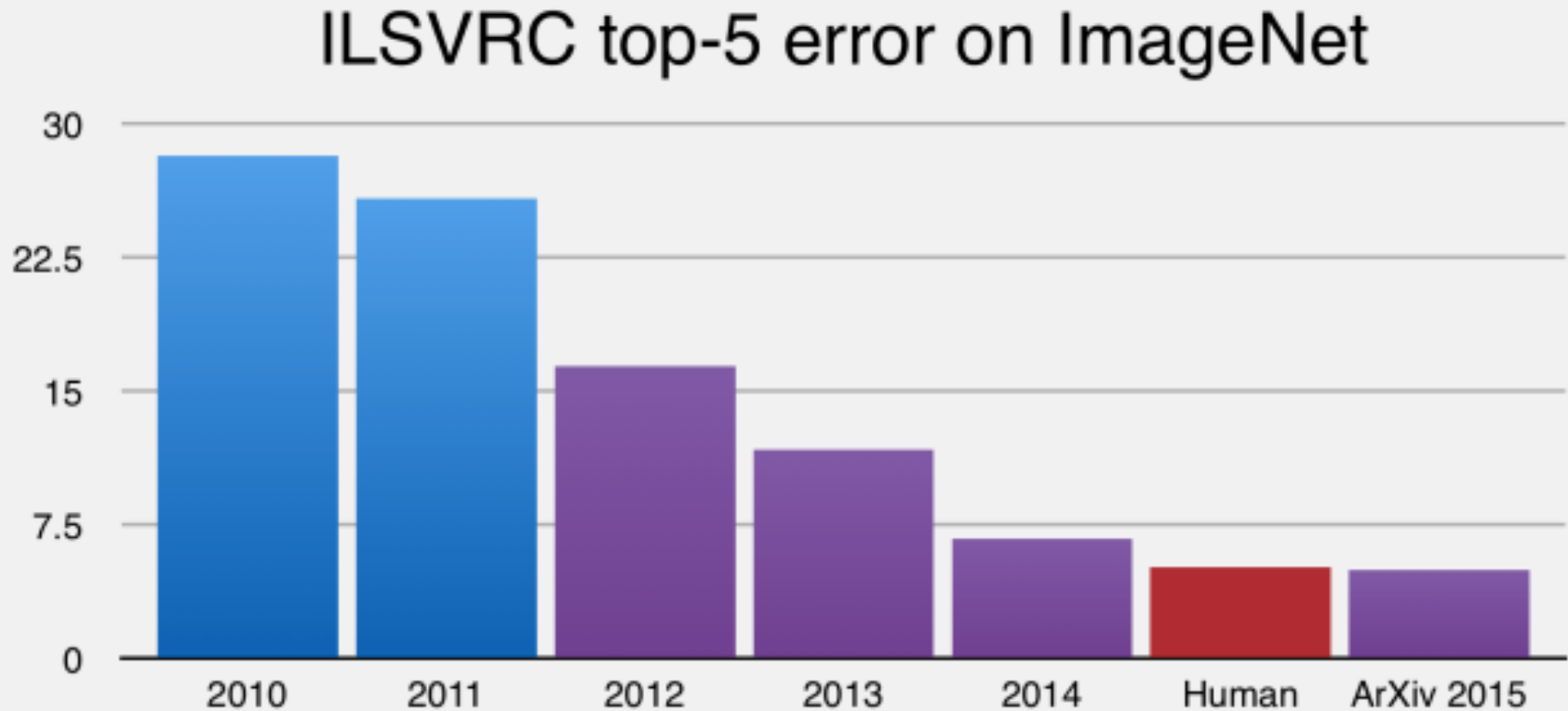
- $mAP = \text{MEAN}(AP \text{ для каждого класса объектов})$
- $AP = \text{MEAN}(\text{Precision для каждого из 11 значений Recall } [0.0, 0.1, \dots, 1.0])$



# Прогресс в соревновании PASCAL VOC

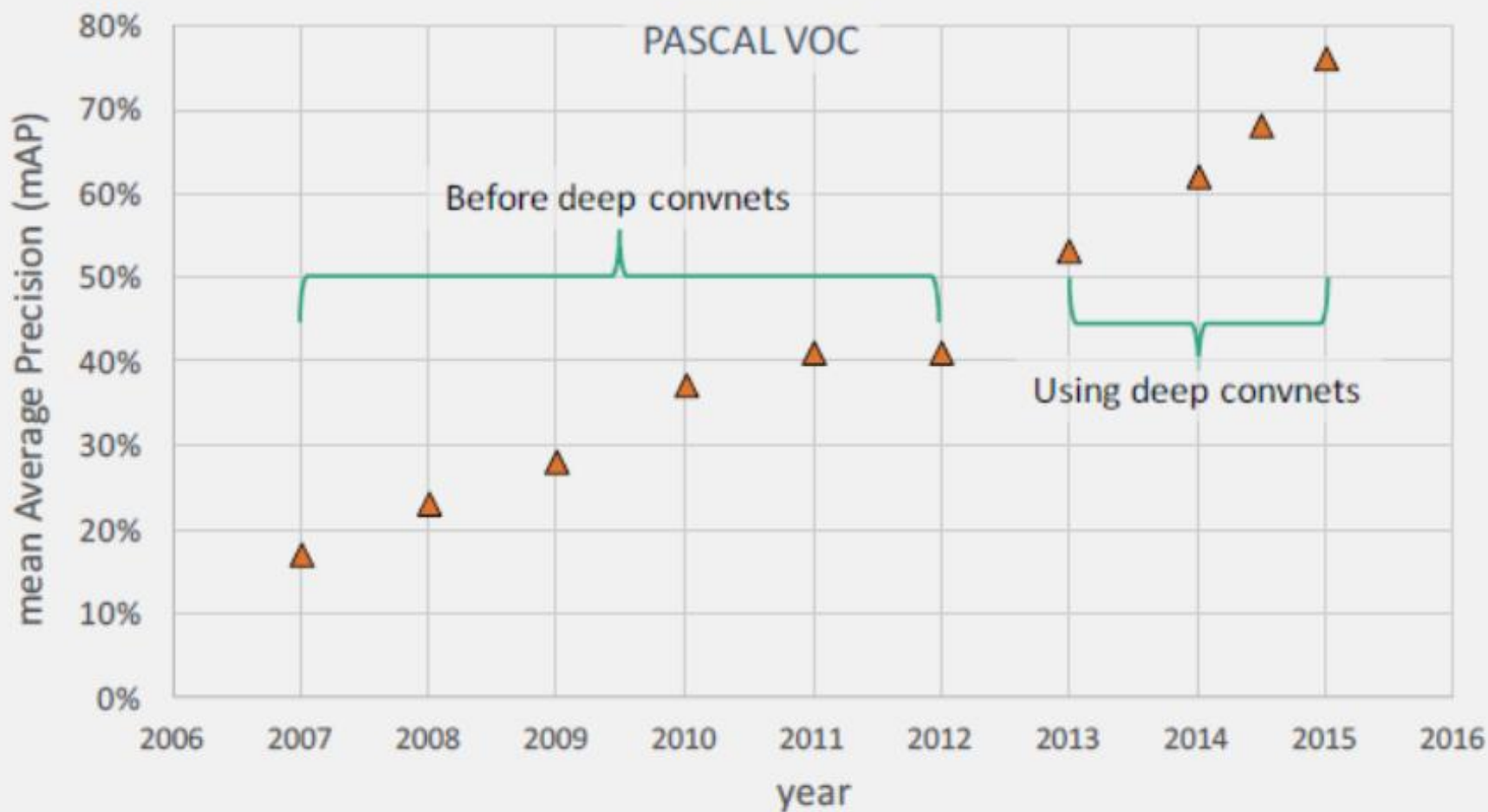


# Прогресс в соревновании ILSVRC



Фиолетовым отмечены модели на основе сверточных нейронных сетей

# Прогресс в соревновании PASCAL VOC после революции CNN

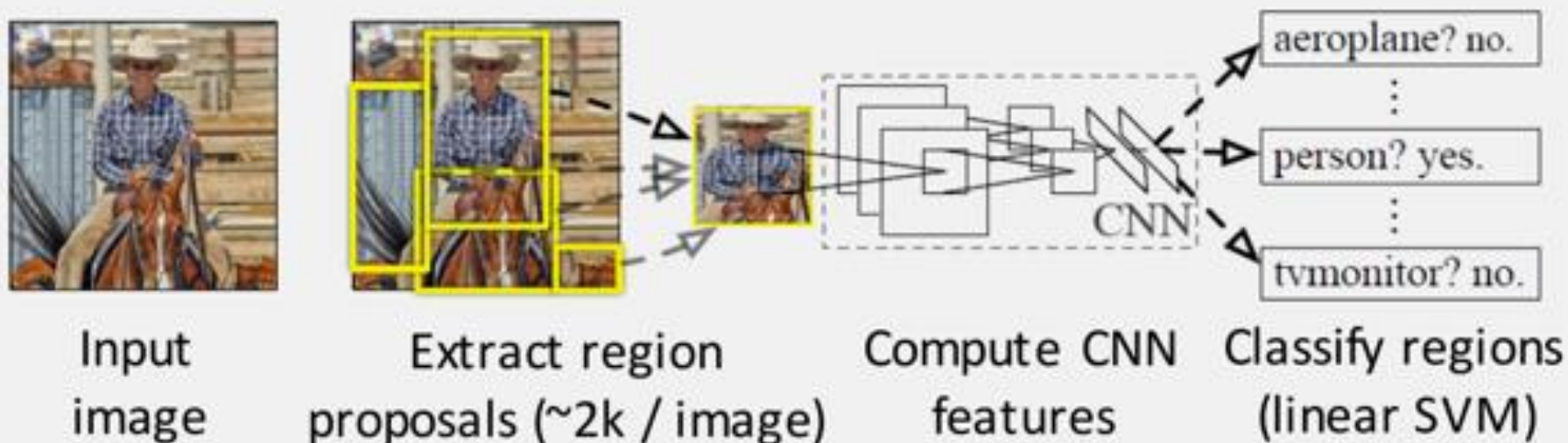


# RCNN (2014), идея



RCNN - Regions with CNN features

Идея: использование сверточной нейронной сети для детектирования объектов



# Выбор рамок потенциальных объектов на изображении



Проблема:

Проход скользящим окном на различных масштабах порождает очень много потенциальных рамок объектов

Решение:

Выбираем только те рамки, в которых потенциально могут быть объекты используя данные о цвете изображения

# Selective Search (2012)

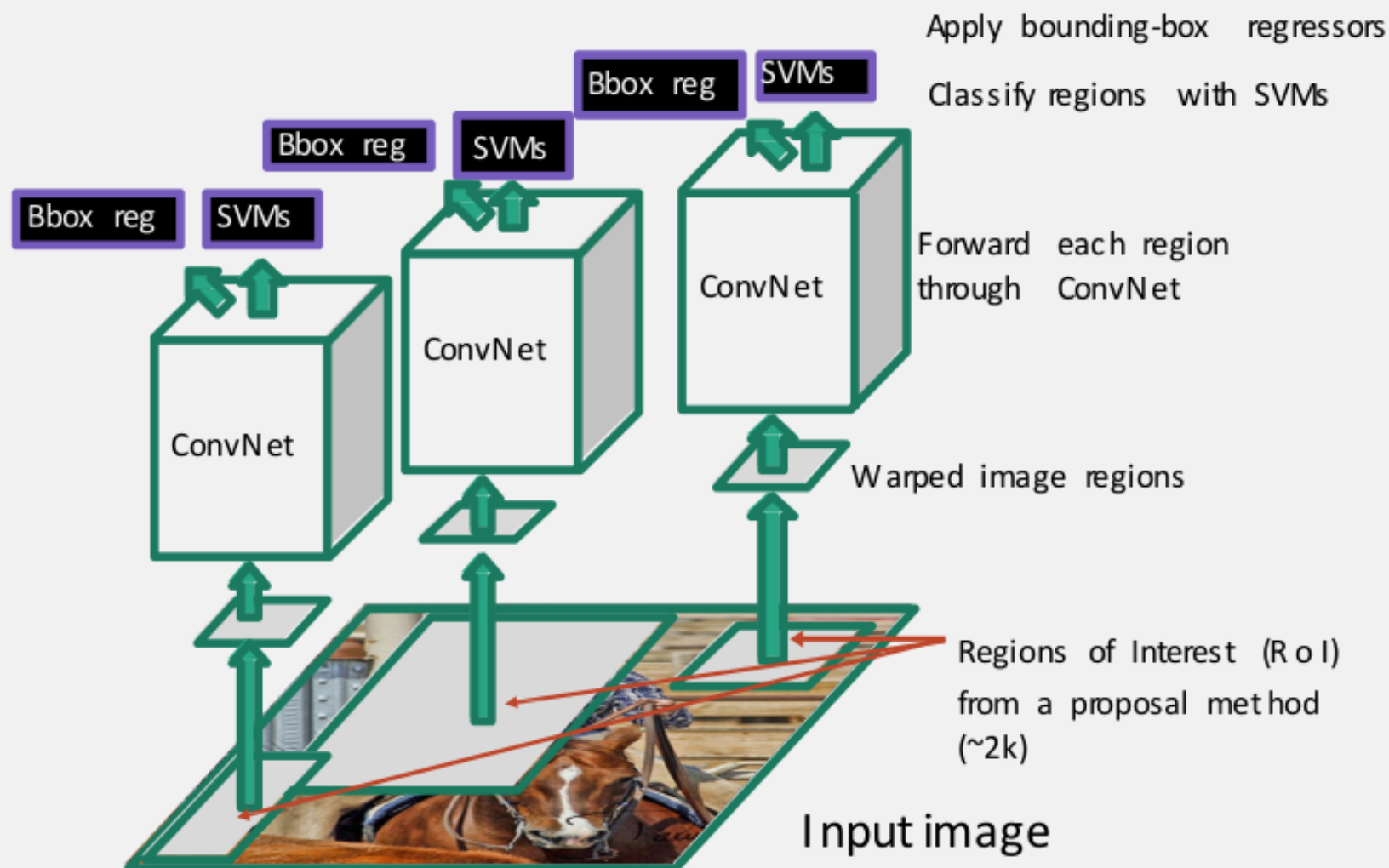


## Алгоритм Selective Search



Снижаем количество рамок с  $\sim 100k$  до  $\sim 2k$

# RCNN, архитектура



# Проблемы RCNN



- Используется предтренированная сверточная нейронная сеть
- Обучение экстрактора признаков, классификатора объектов и регрессора bbox-ов происходит раздельно
- Система работает медленно



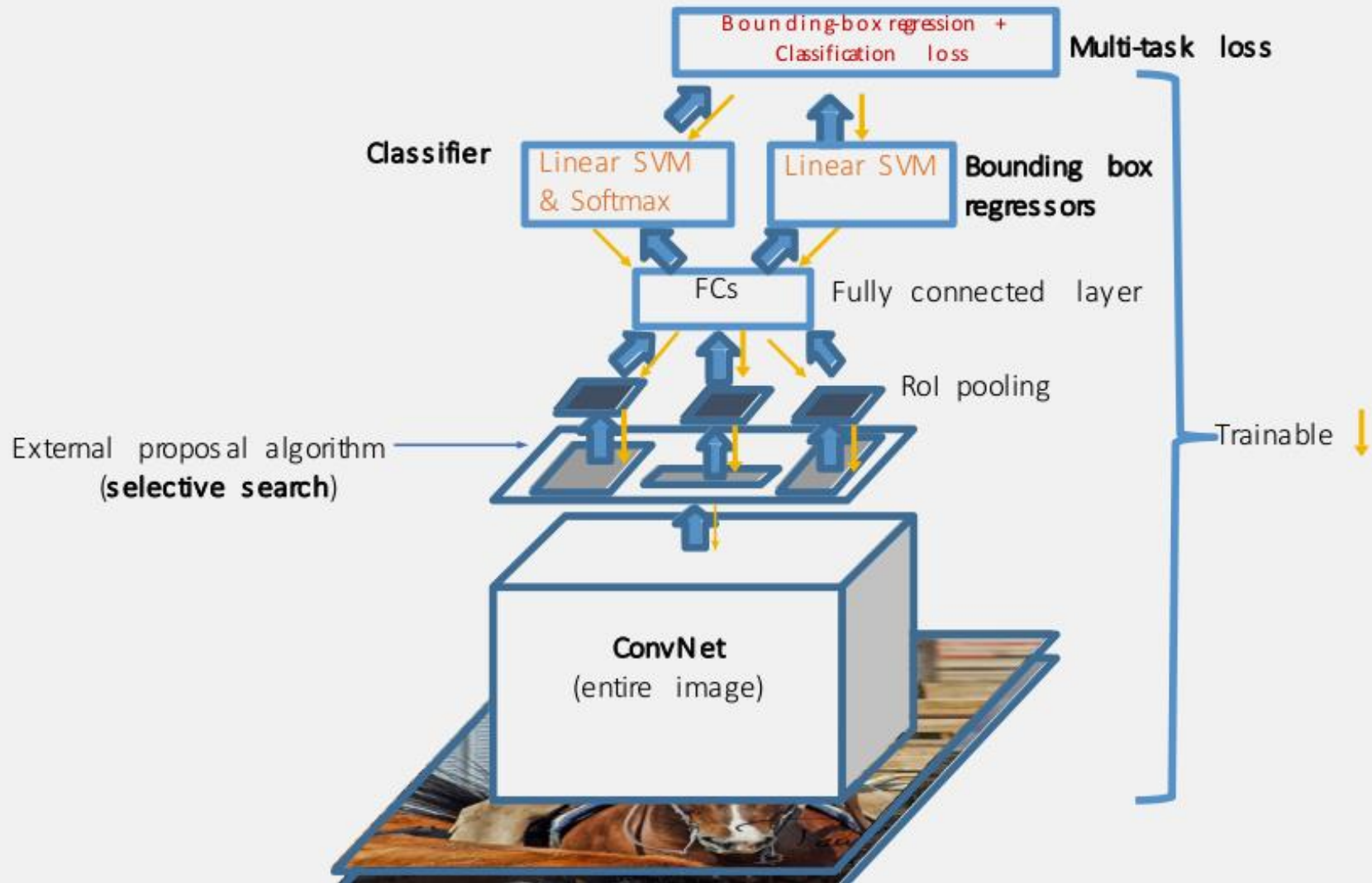
# Fast RCNN (2015), идея



Идея:

- Сделаем end-to-end систему
- Будем извлекать признаки из изображения только 1 раз

# Fast RCNN, архитектура



# RCNN и Fast RCNN



Training time:	84 hours / 8.75 hours
VOC07 test mAP:	66.0% / 68.1%
Testing time per image:	49s / 2.32s

# Проблемы Fast RCNN



- Selective Search не тренируется
- Selective Search работает медленно

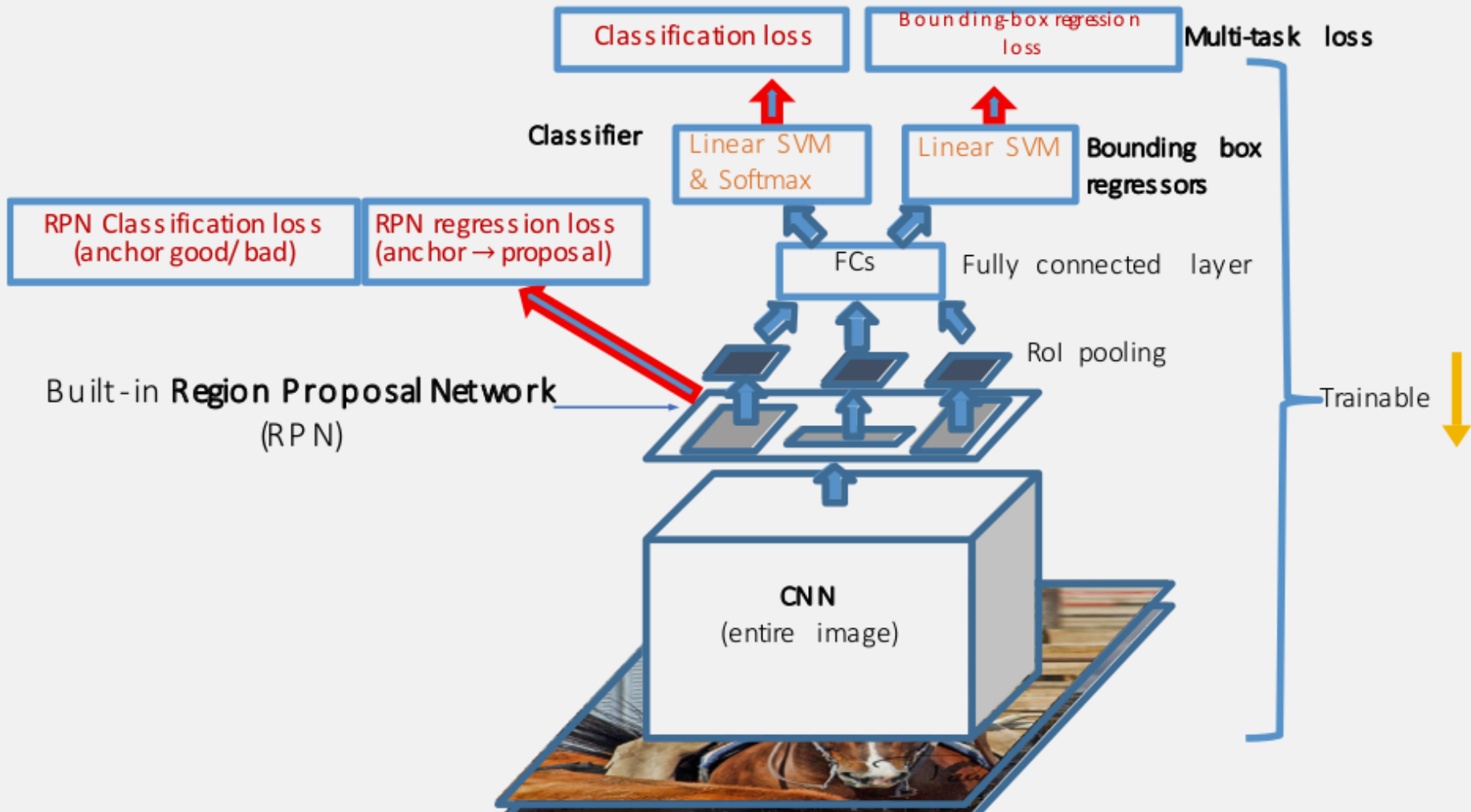
# Faster RCNN (2015), идея



Идея:

Заменим Selective Search на тренируемую Region Proposal Network

# Faster RCNN, архитектура



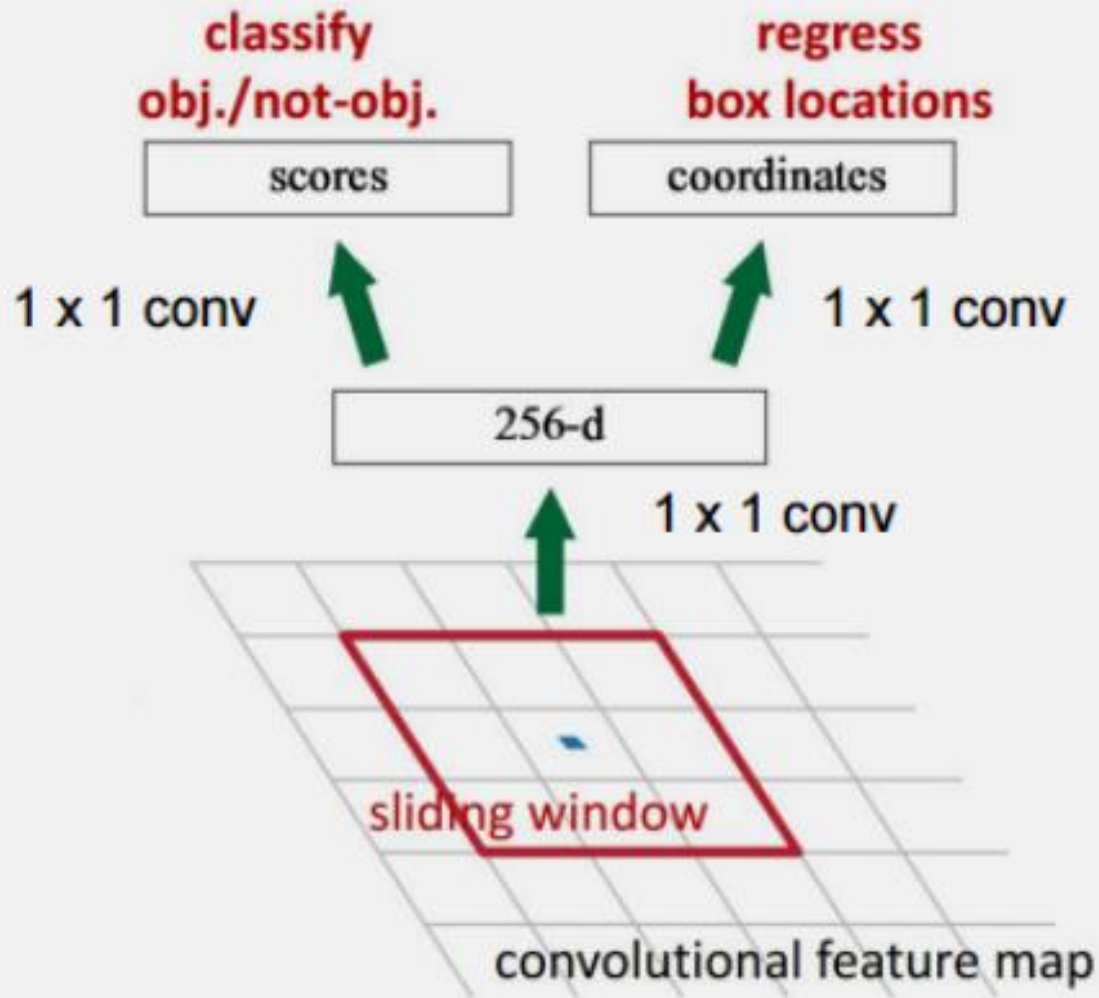
# Region Proposal Network



RPN - это небольшая нейронная сеть, которая на вход принимает карты признаков и предсказывает:

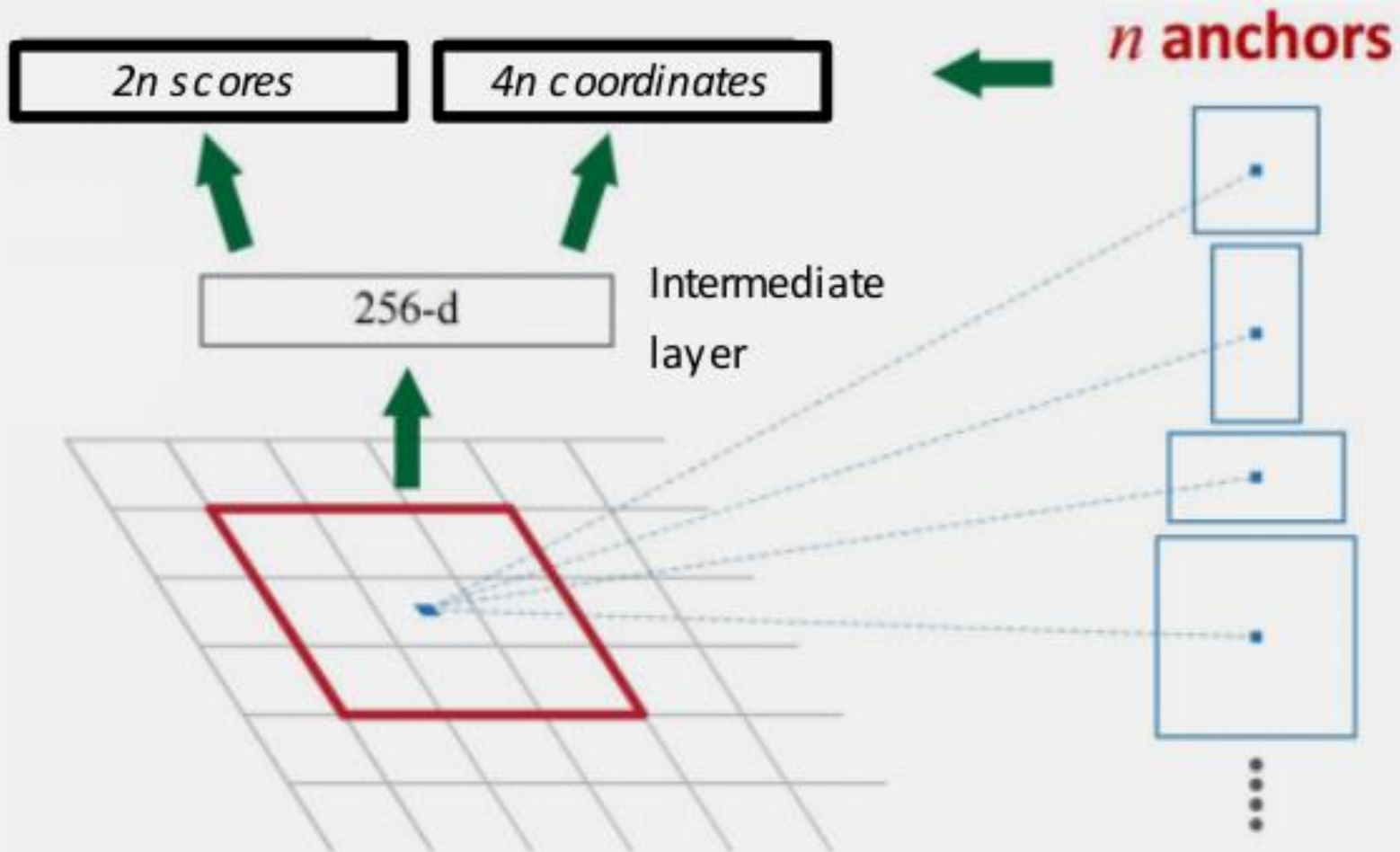
1. Есть ли в данной области объект
2. Уточняет локализацию Bbox-ов

# Region Proposal Network





# Region Proposal Network



# Сравниваем RCNN, Fast RCNN, Faster RCNN

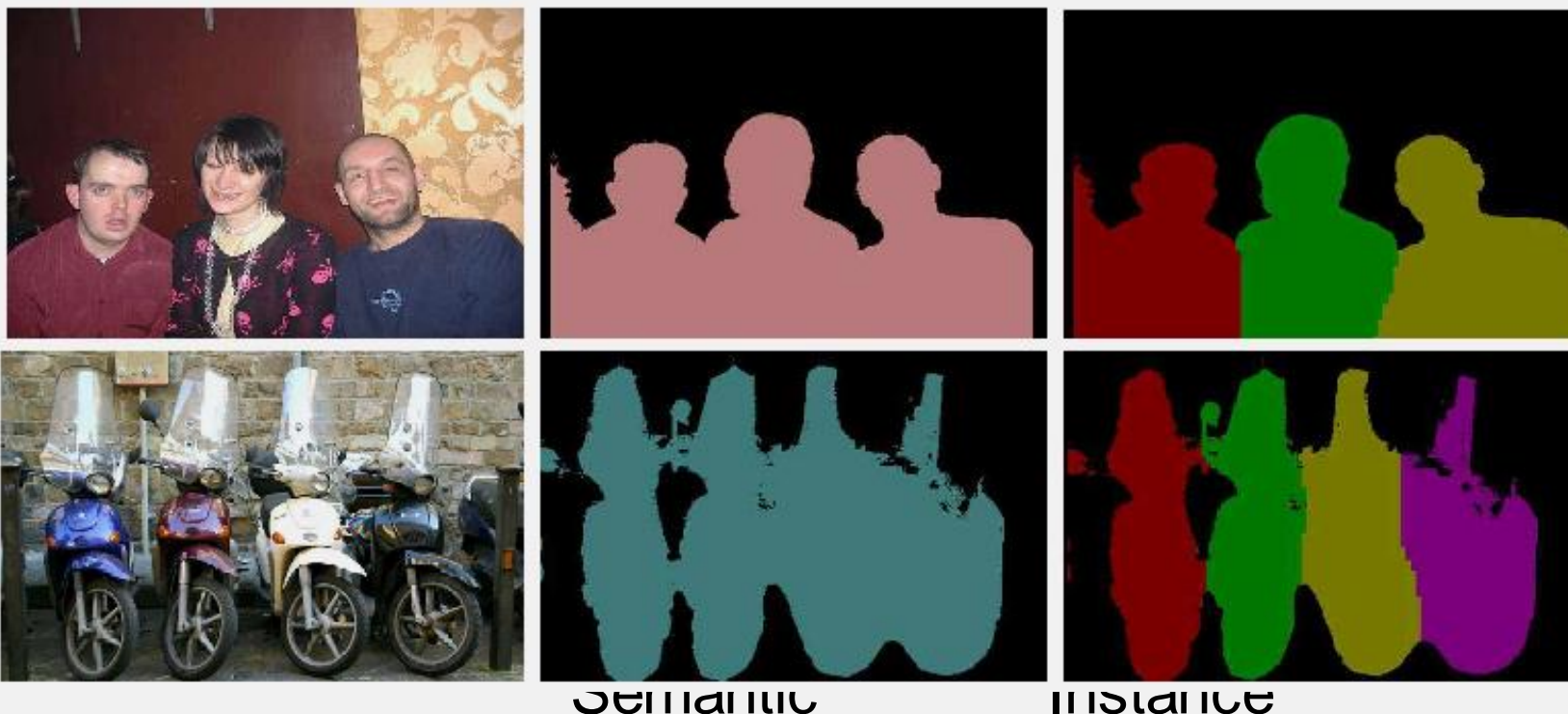


	RCNN	Fast RCNN	Faster RCNN
PASCAL VOC 2007 mAP	66.0	66.9	66.9
Время на предсказание	50 сек.	2 сек.	0.2 сек.
Ускорение	1x	25x	250x

# Виды сегментации



Задачи сегментации обычно разделяют на две группы:



# COCO dataset для задач сегментации



Common Objects in Context (COCO) - это большой, качественно размеченный датасет обычных объектов в их естественной среде от Microsoft.

- более 200к изображений
- 80 категорий объектов (Instance)
- 91 классов (semantic)



# Метрики соревнований в рамках COCO



Semantic segmentation:

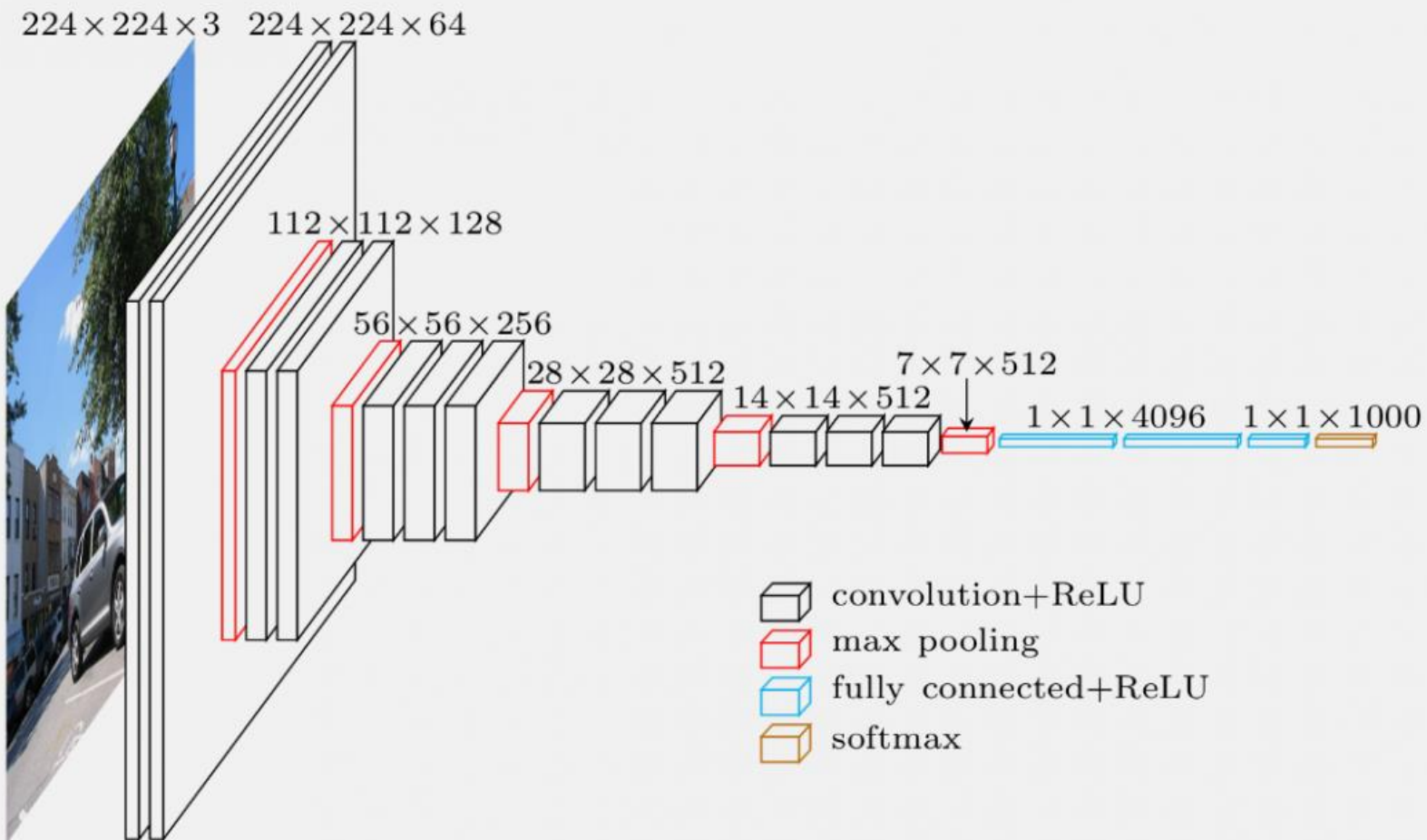
- усреднение по классам IoU

Instance segmentation:

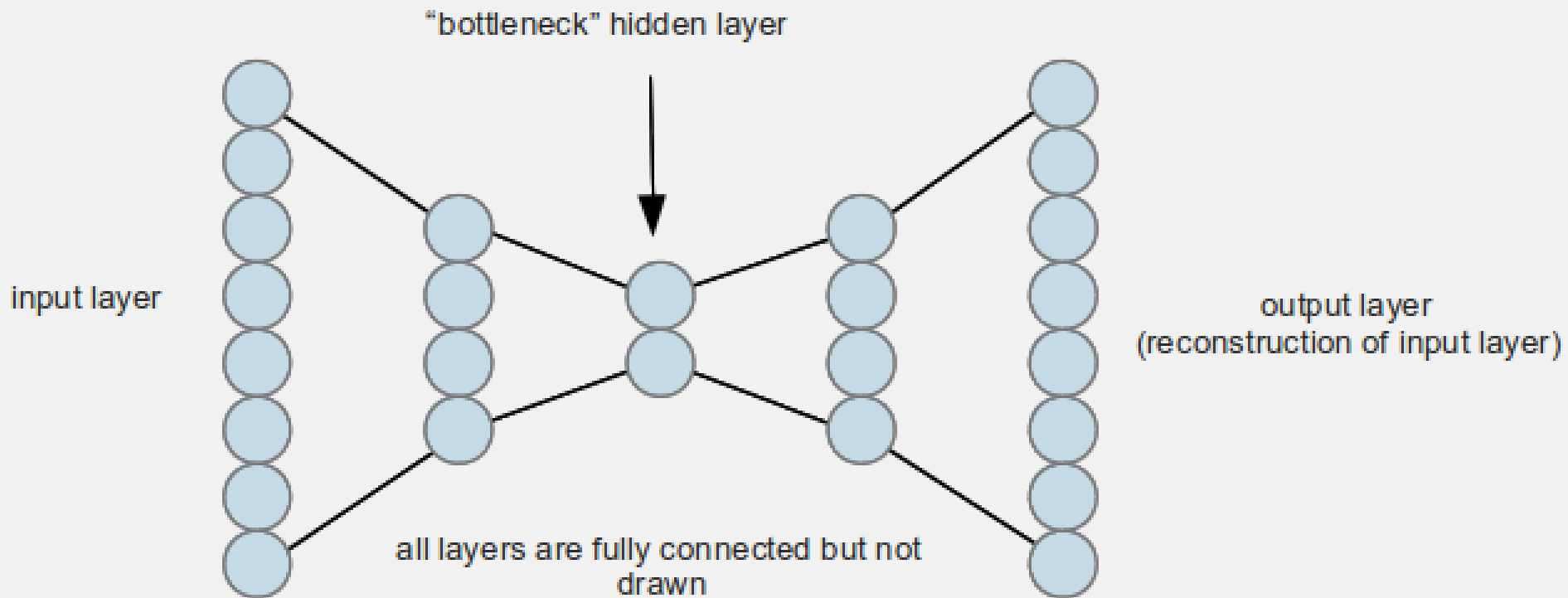
- mAP, усредненный по порогам принятия детекции по IoU [0.5, 0.55 ... 0.95]



## Архитектура сети VGG-16 (победитель ILSVRC 2014)



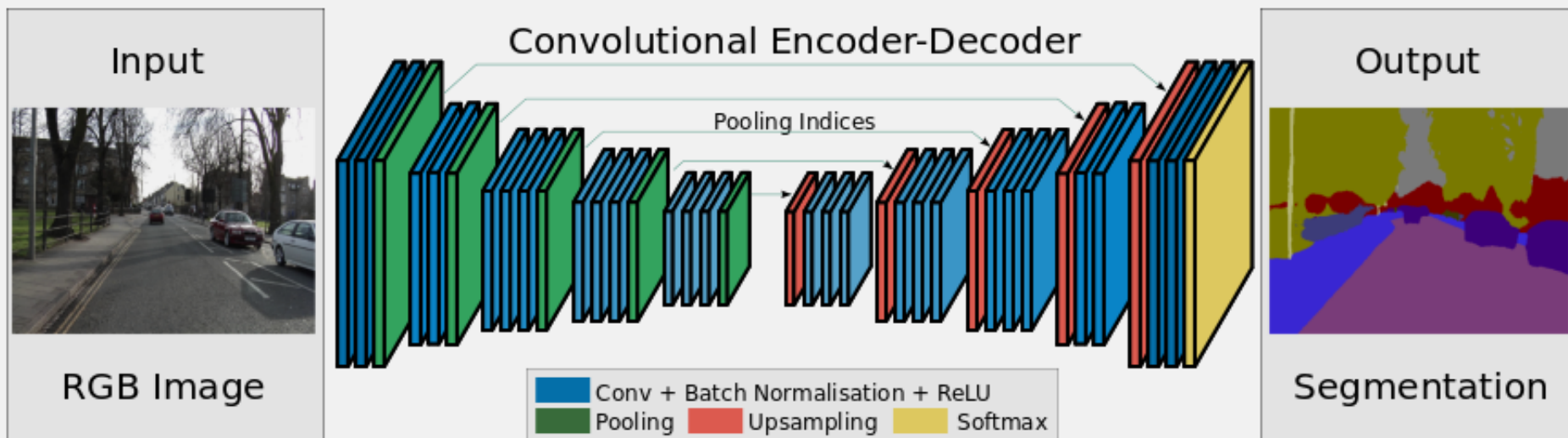
# Автоэнкодеры



# Архитектура SegNet (2015)



SegNet по сути сверточный автоэнкодер с хитрой реализацией операции unpooling





# Архитектура SegNet, unpooling



0.1	0.5	<b>1.2</b>	-0.7
<b>0.8</b>	-0.2	-0.5	0.3
0.4	<b>0.9</b>	-0.1	-0.2
-0.6	0.1	<b>0.5</b>	0.3

max-pooling

0.8	1.2
0.9	0.5

		x	
x			
	x		
		x	

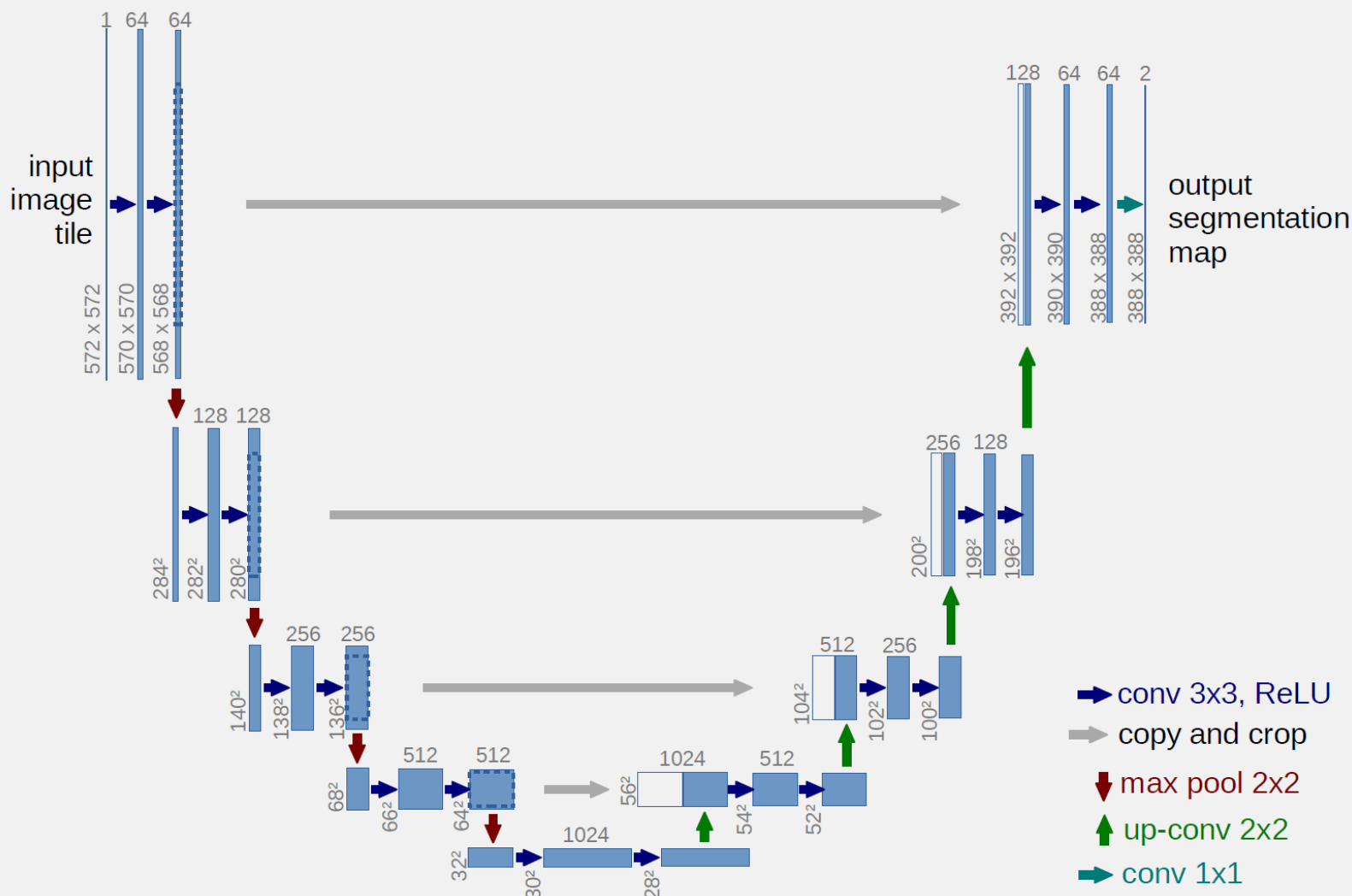
max locations

0	0	<b>0.5</b>	0
<b>1.3</b>	0	0	0
0	<b>0.4</b>	0	0
0	0	<b>0.1</b>	0

unpooling

1.3	0.5
0.4	0.1

# Архитектура U-net (2015)

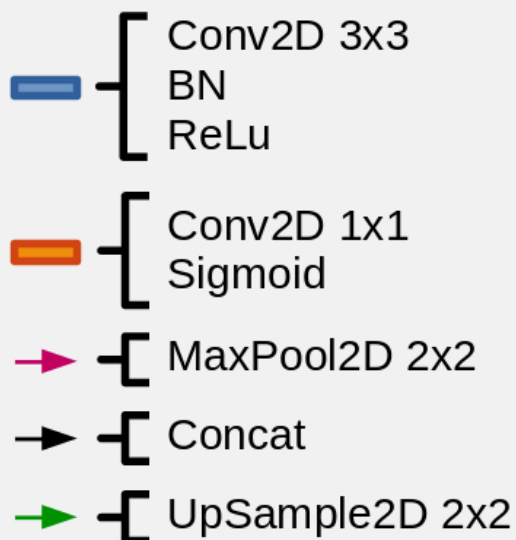




## Задача:

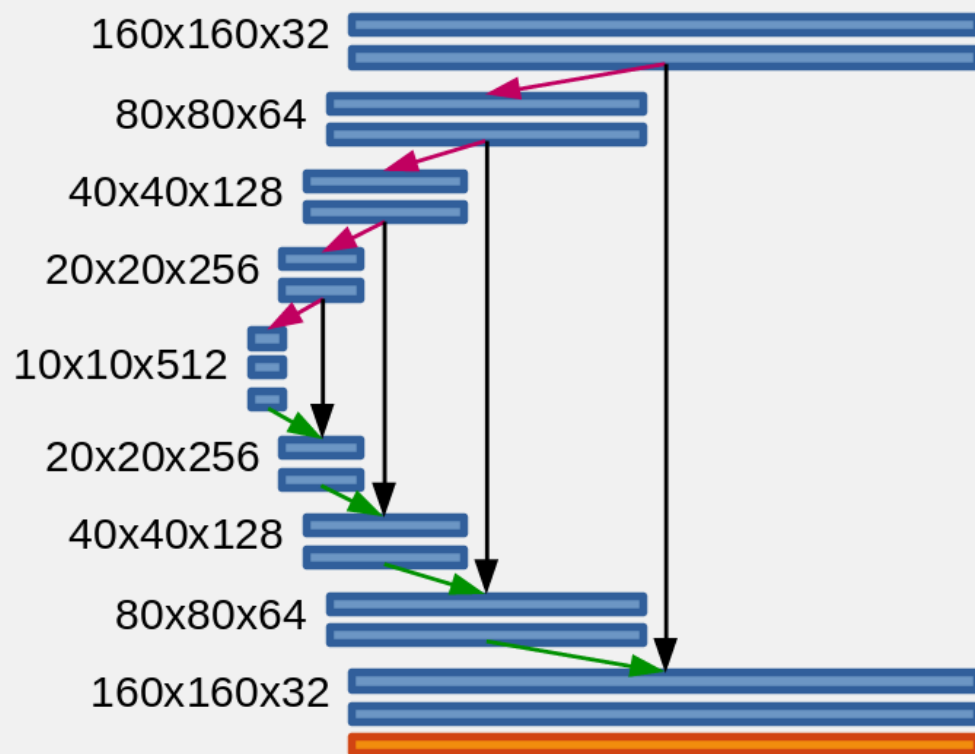
- 25 спутниковых изображений в Train
- 425 спутниковых изображений в Test
- 10 классов объектов
- Изображения представляют участки поверхности земли 1x1 км
- Для каждого участка 1x1 км даны 4 файла tiff с разных приборов

# U-net, пример



Input: 160x160x20

Output: 160x160x7 or 160x160x2

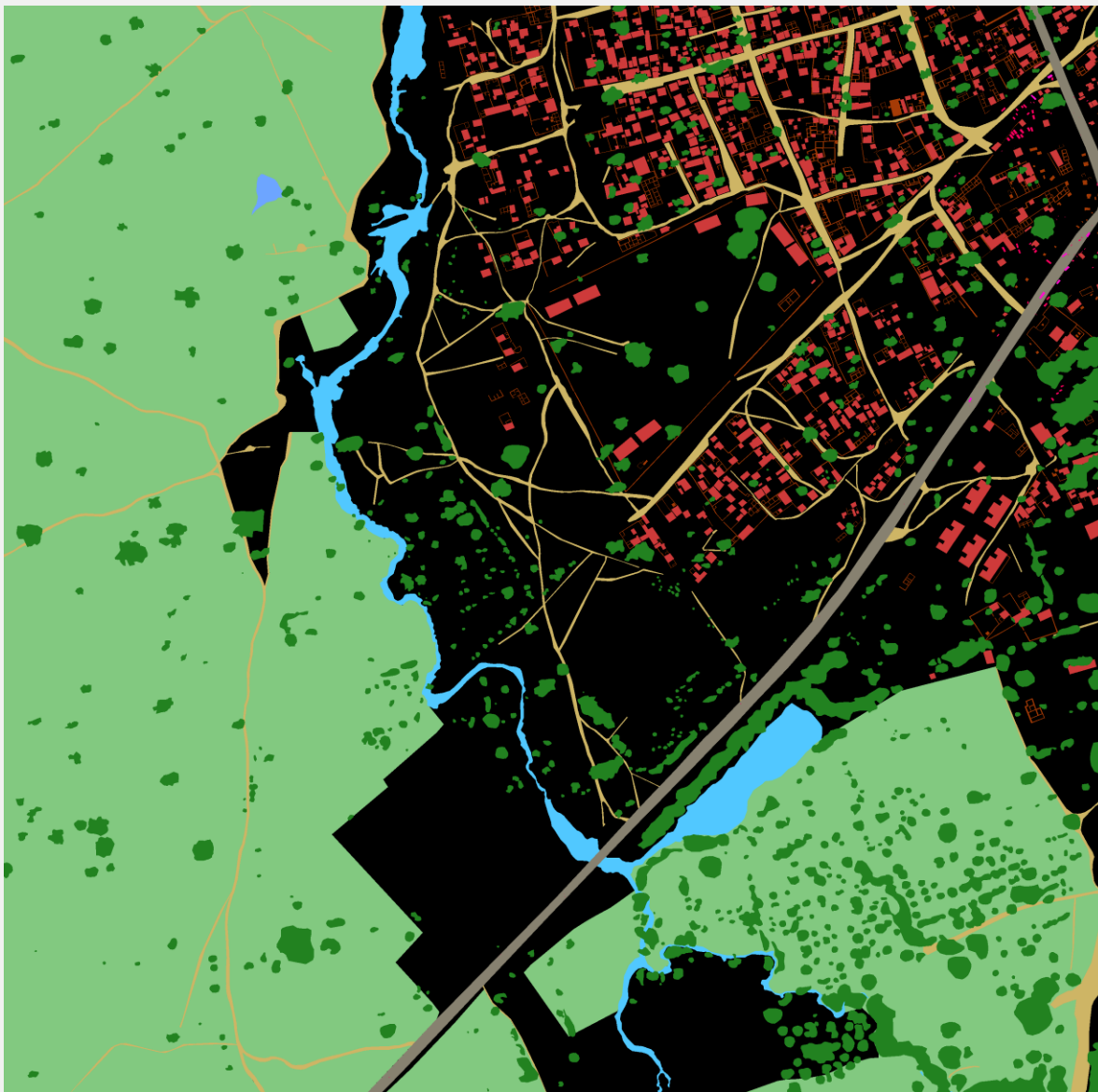


# Сегментация Train





# Сегментация Train

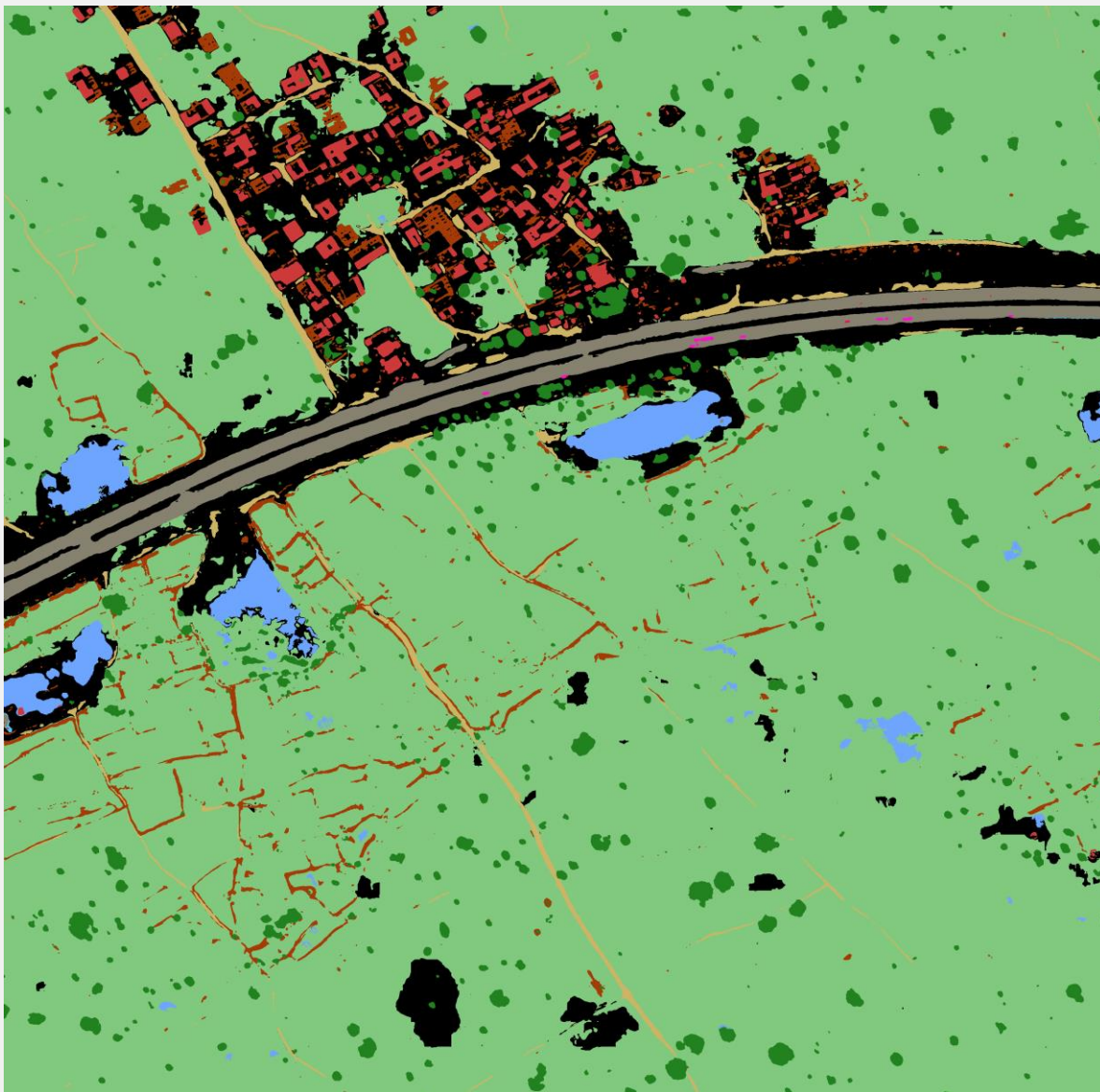


# Сегментация Test



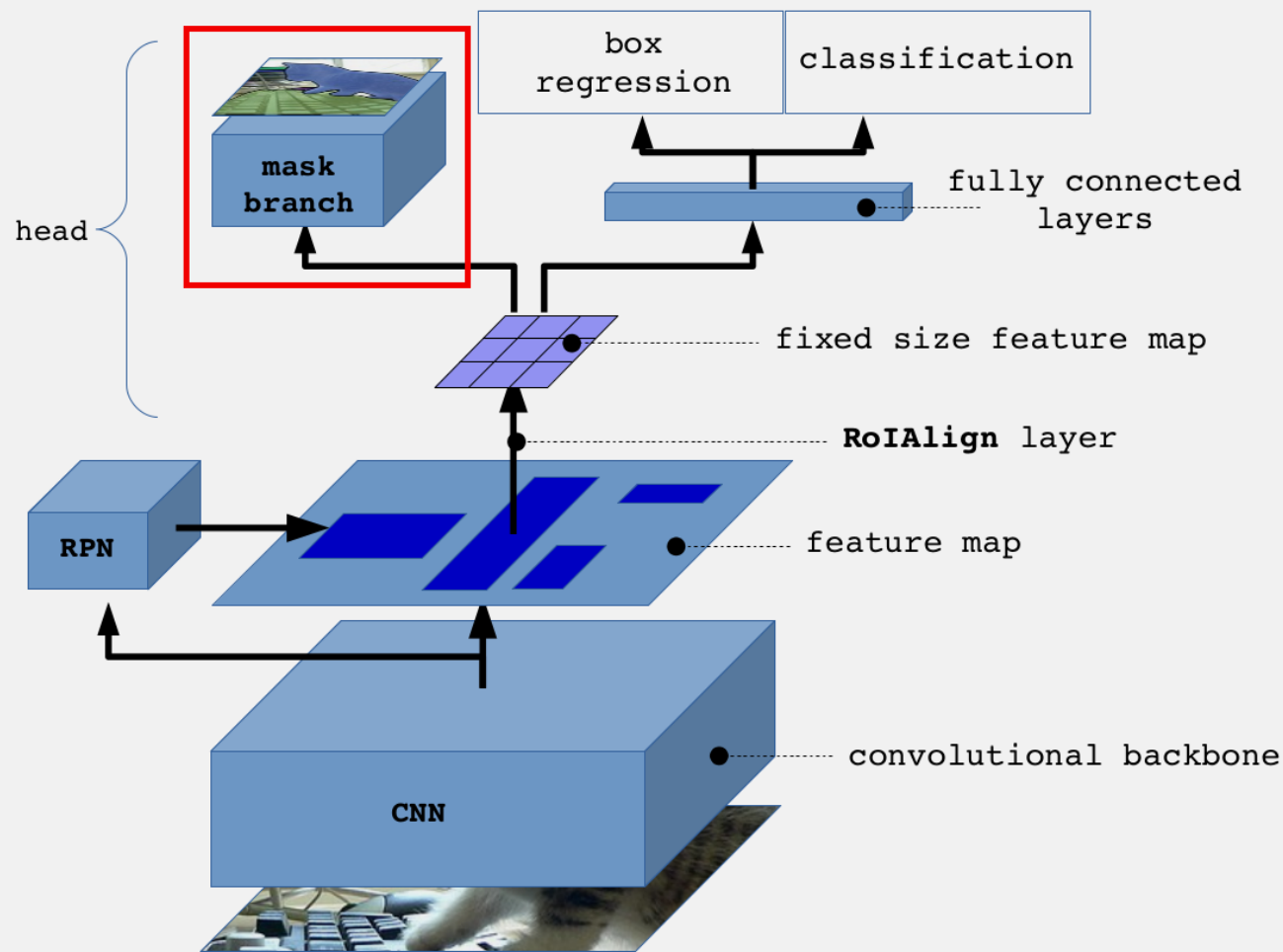


# Сегментация Test



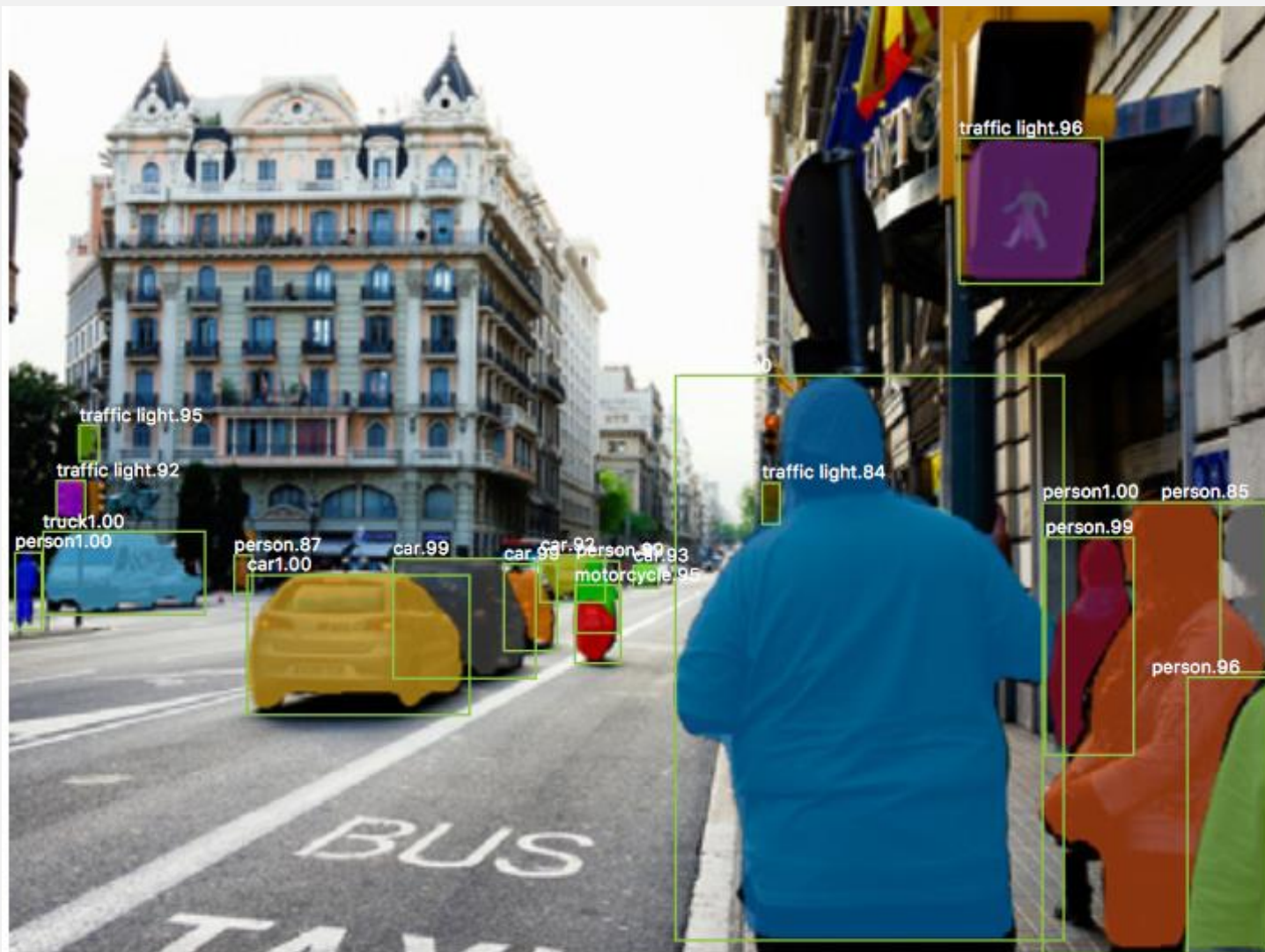


# Mask RCNN (2017)



Добавим еще одну ветвь, которая на основе карты признаков будет предсказывать маску объекта

# Mask RCNN, сегментация



# U-net (Instance segmentation)



Задачу instance segmentation можно свести к задаче semantic segmentation следующими способами:

- Предсказание границ объектов
- Разделение масок одного класса зазором в несколько пикселей
- Предсказание центров объектов и последующая постобработка



**Спасибо за  
внимание!**

Евгений Некрасов

[e.nekrasov@corp.mail.ru](mailto:e.nekrasov@corp.mail.ru)