

Глубинное обучение

Лекция 8: Свёрточные сети в задачах компьютерного зрения

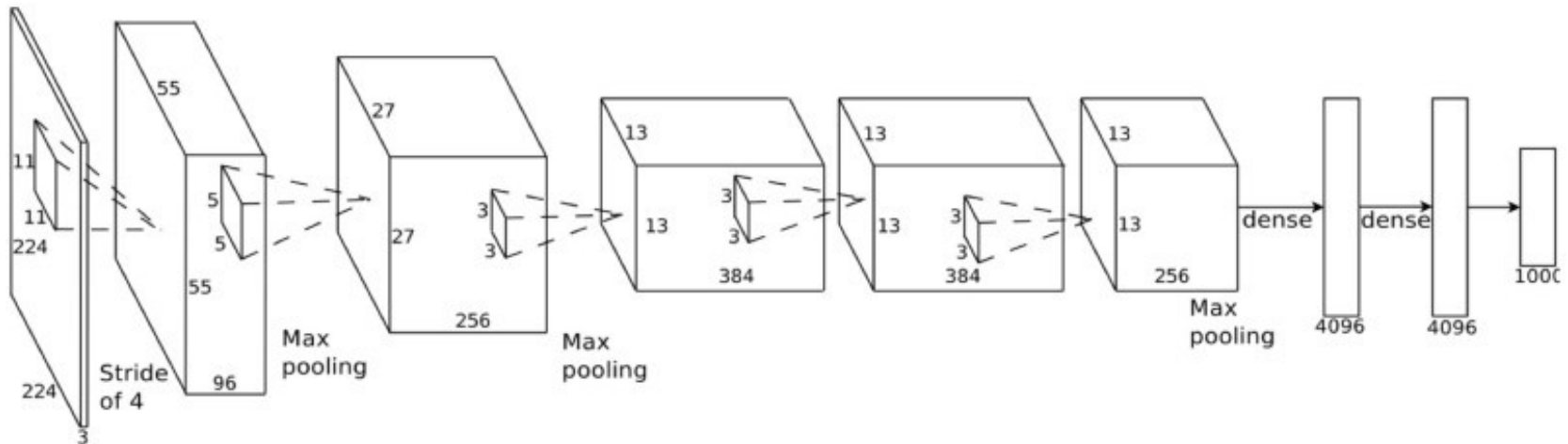
Лектор: Антон Осокин

ФКН ВШЭ, 2019



НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
УНИВЕРСИТЕТ

Ресар: классификация



- Задача классификации изображение решена! (почти)
- Вход сети – изображение
- Выходы сети соответствуют классам
- Функция потерь – кросс-энтропия (log loss)
- Много архитектур сетей (например, ResNet)
- Блок свёрточных слоев в начале сети
- Идея – переиспользовать выученные представления

План лекции

- Детекция объектов
 - R-CNN, Fast R-CNN, Region Proposal Networks
 - Быстрые детекторы: SSD and YOLO
- Сегментация изображений
 - Fully convolutional networks
 - Masked R-CNN
- Поиск похожих изображений
 - Siamese architecture
 - Отслеживание объектов на видео
- Распознавание действий на видео

Часть 1: детекция объектов

- Задача найти объекты на изображении
- Найти = поставить прямоугольник (bounding box)

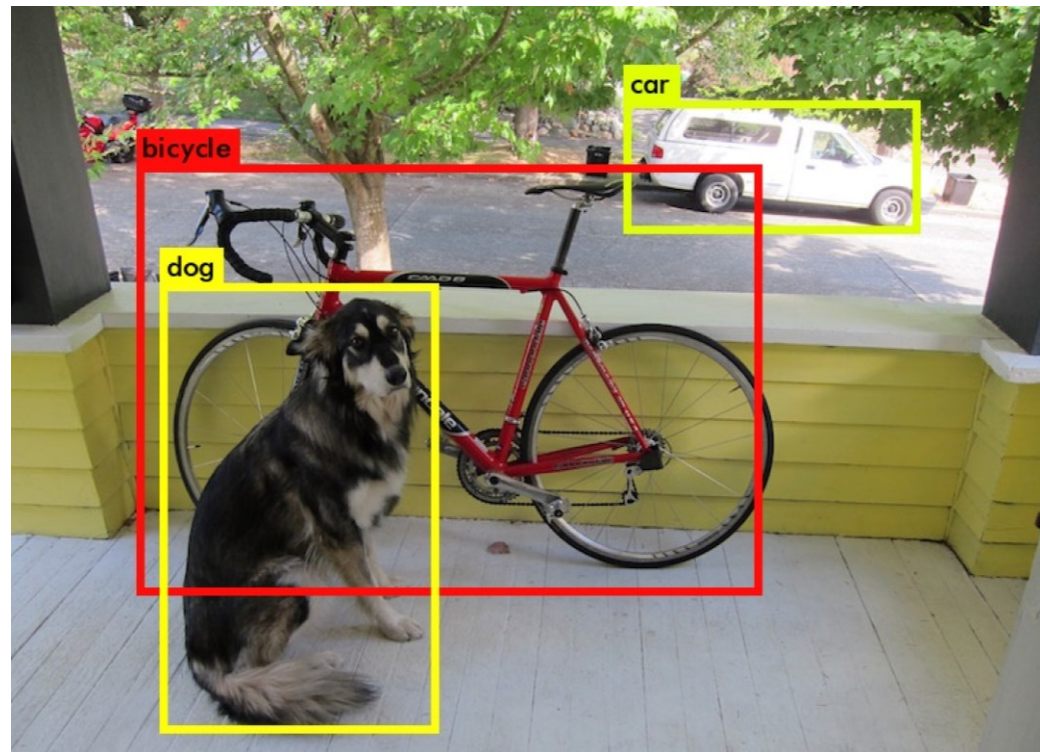
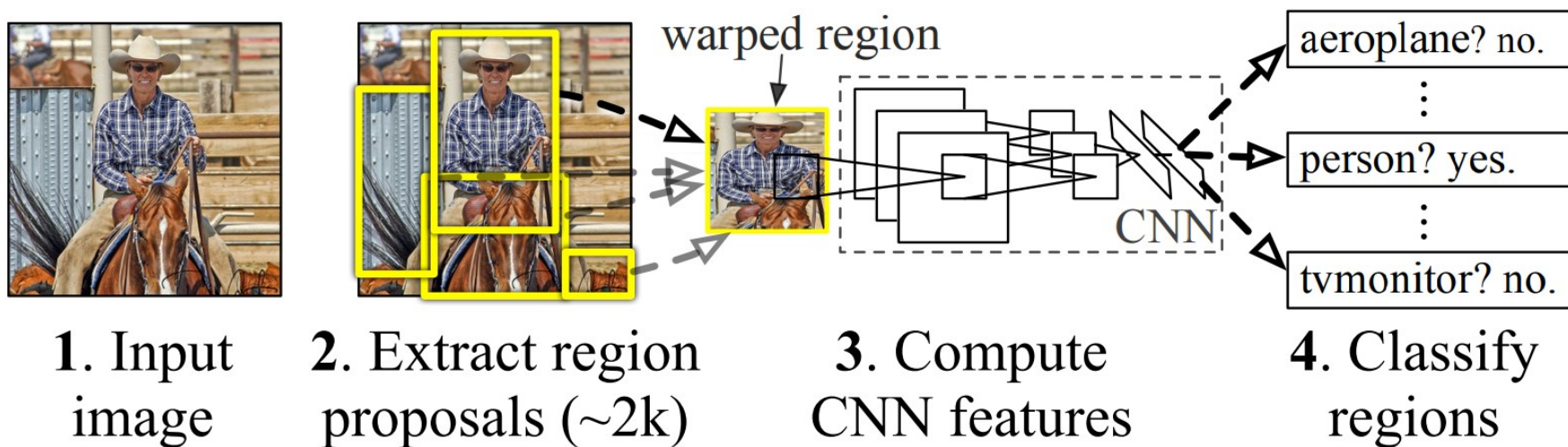


image credit: Joseph Redmon

Ранние методы: R-CNN

[Girshick et al., 2013]

R-CNN: *Regions with CNN features*

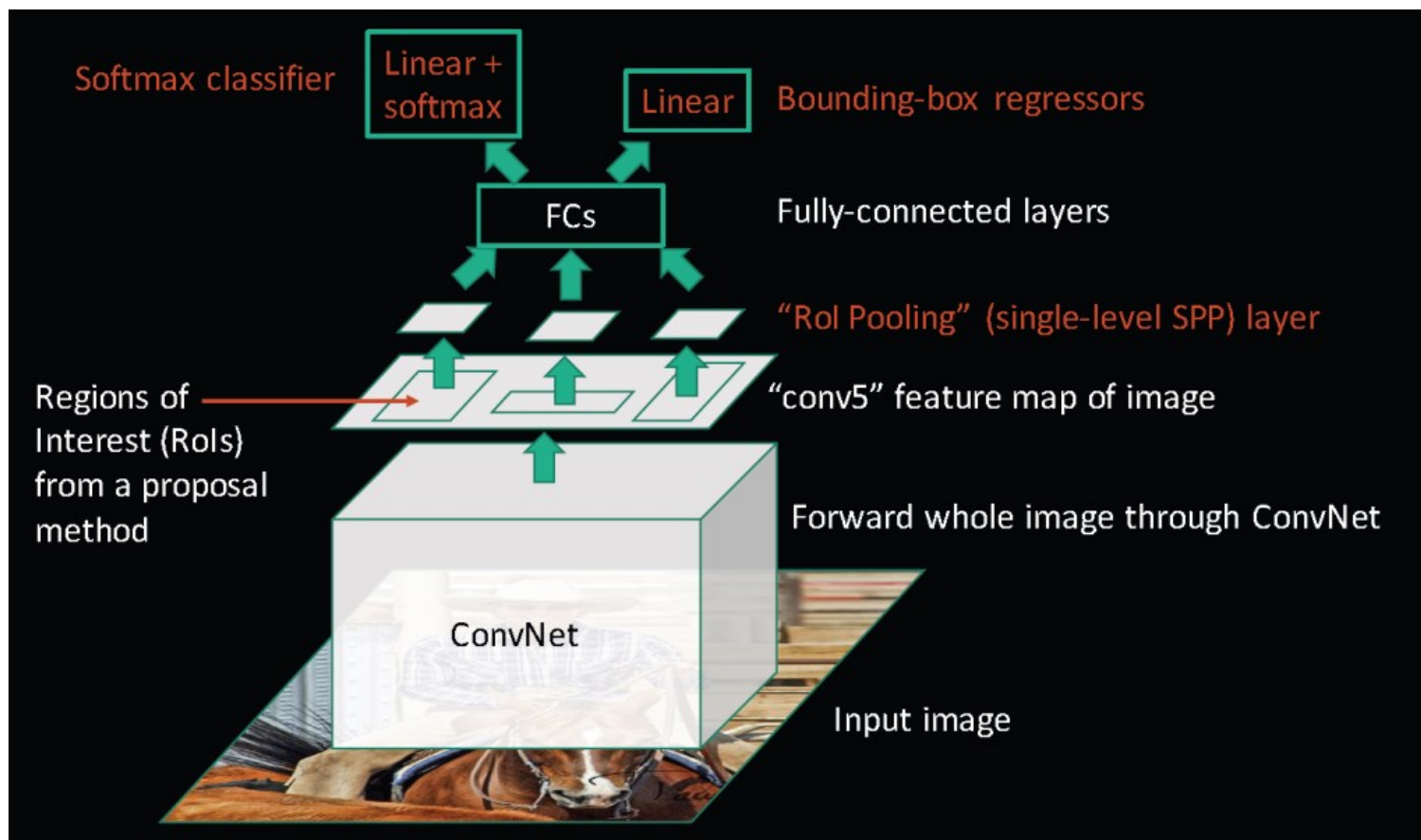


- Основная идея – классифицировать гипотезы (object proposals)
- Используем CNN для каждой гипотезы
- На выходе: метка класса и уточнение позиции объекта
- Проблема: сильный дисбаланс объектов и фона
 - Контроль баланса в батче, специальные функции потерь (focal loss)

Fast R-CNN

[Girshick 2015]

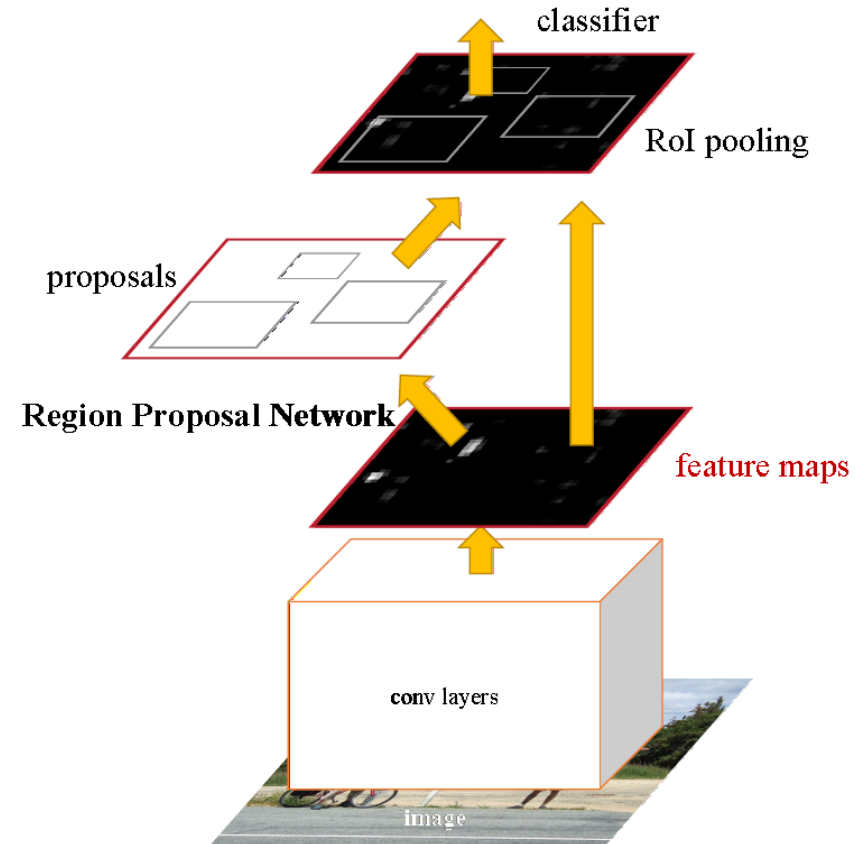
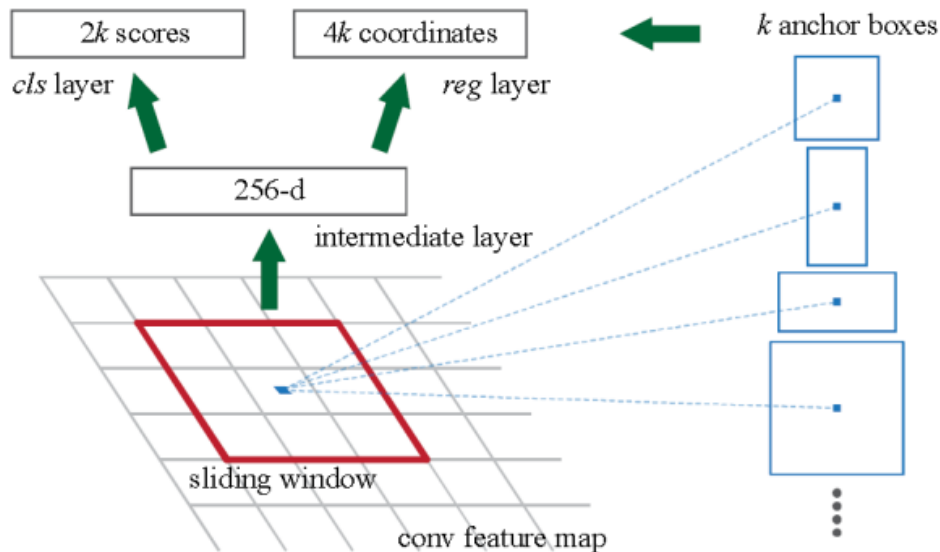
- Недостаток R-CNN – медленная скорость работы
- Много пересекающихся гипотез – неэффективно
- Идея: разделить вычисления свёрток между гипотезами



Region proposal network

[Ren et al., 2015]

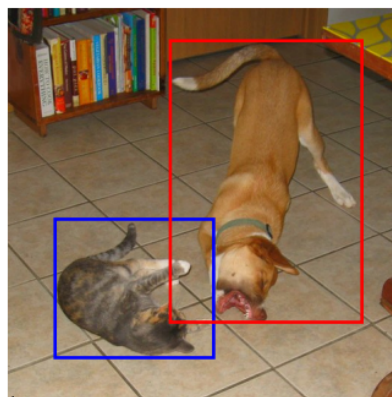
- Fast R-CNN нужны гипотезы
- Гипотезы считать медленно
- Идея: гипотезы из сети
- 5-17 FPS



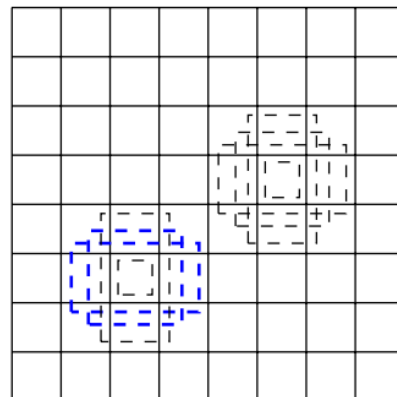
Fast detectors: YOLO (v3), SSD, RetinaNet

[Liu et al., 2016]

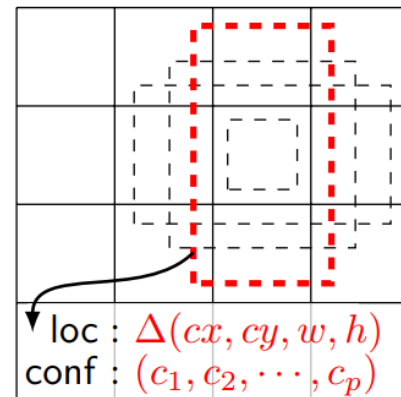
- Идея: отказ от двух стадий детекции, ответ за 1 проход
- Только RPN
- SSD: 59 FPS



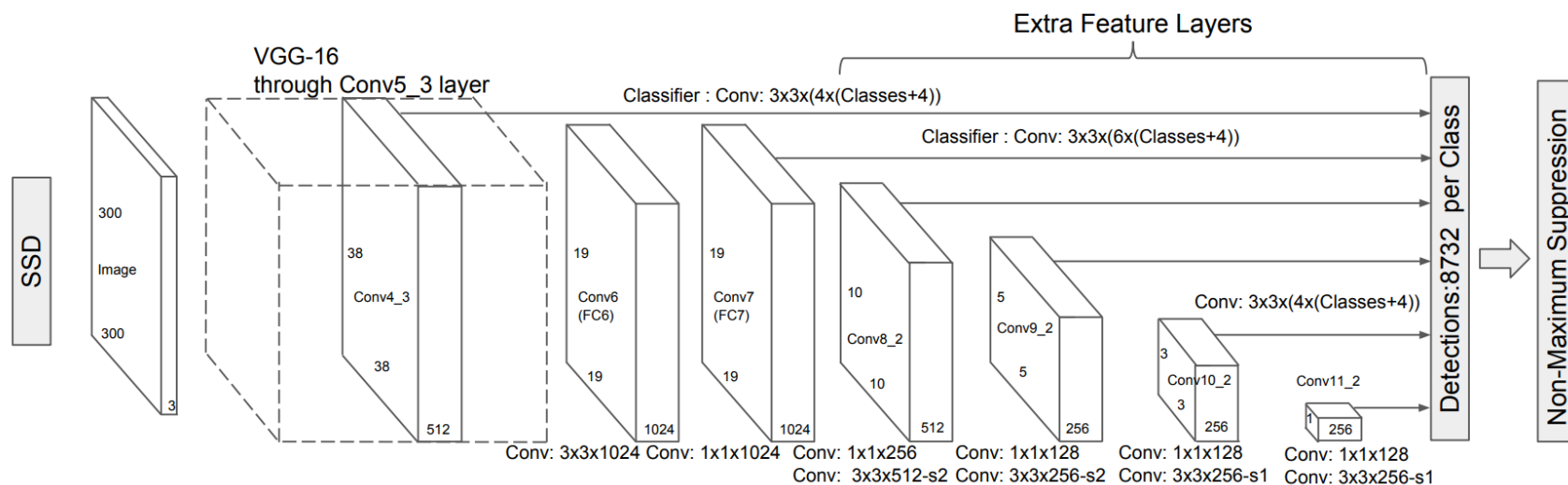
(a) Image with GT boxes



(b) 8×8 feature map



(c) 4×4 feature map



Часть 2: сегментация

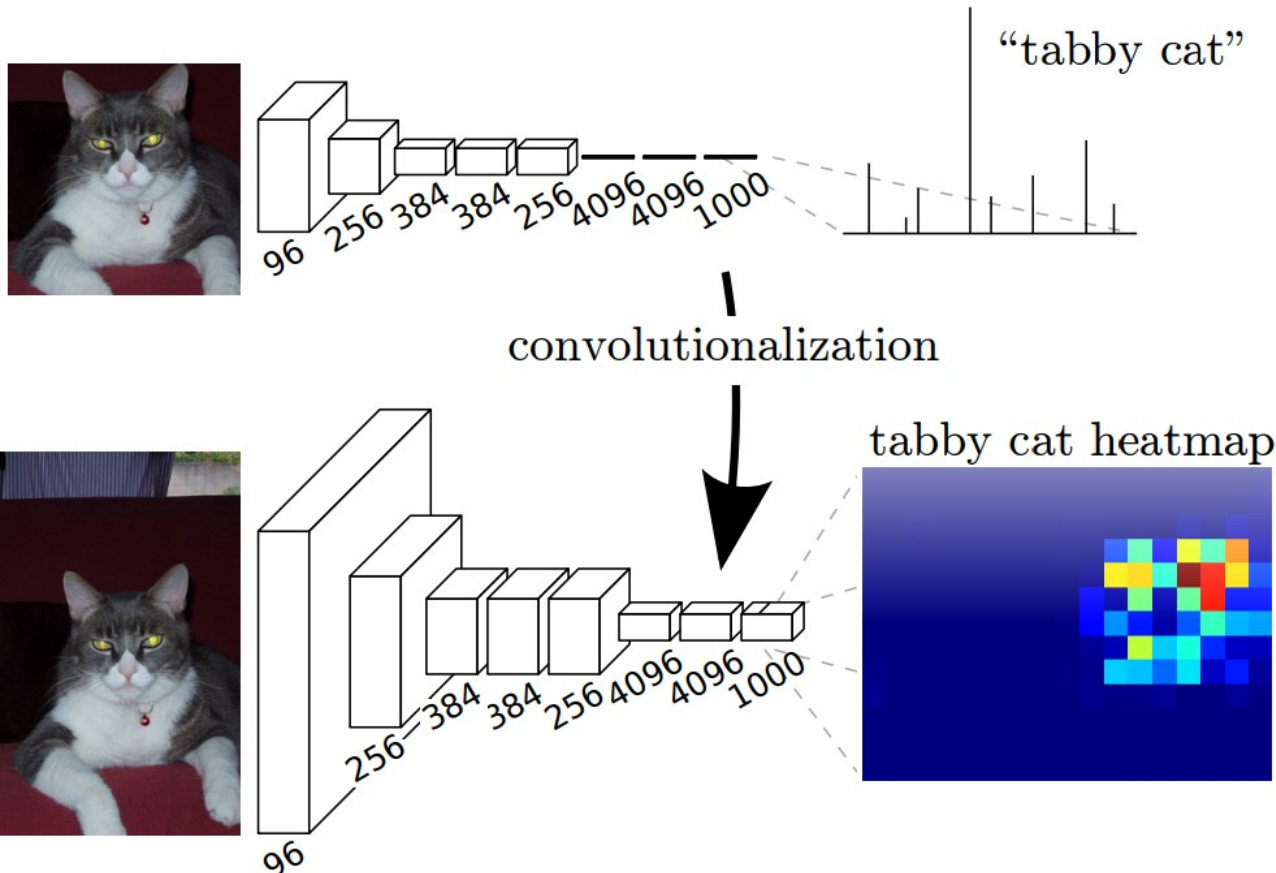
- Задача найти объекты на изображении
- Найти = метки класса для пикселей



Fully-convolutional CNN

Идея из 90-х, [Long et al., 2015]

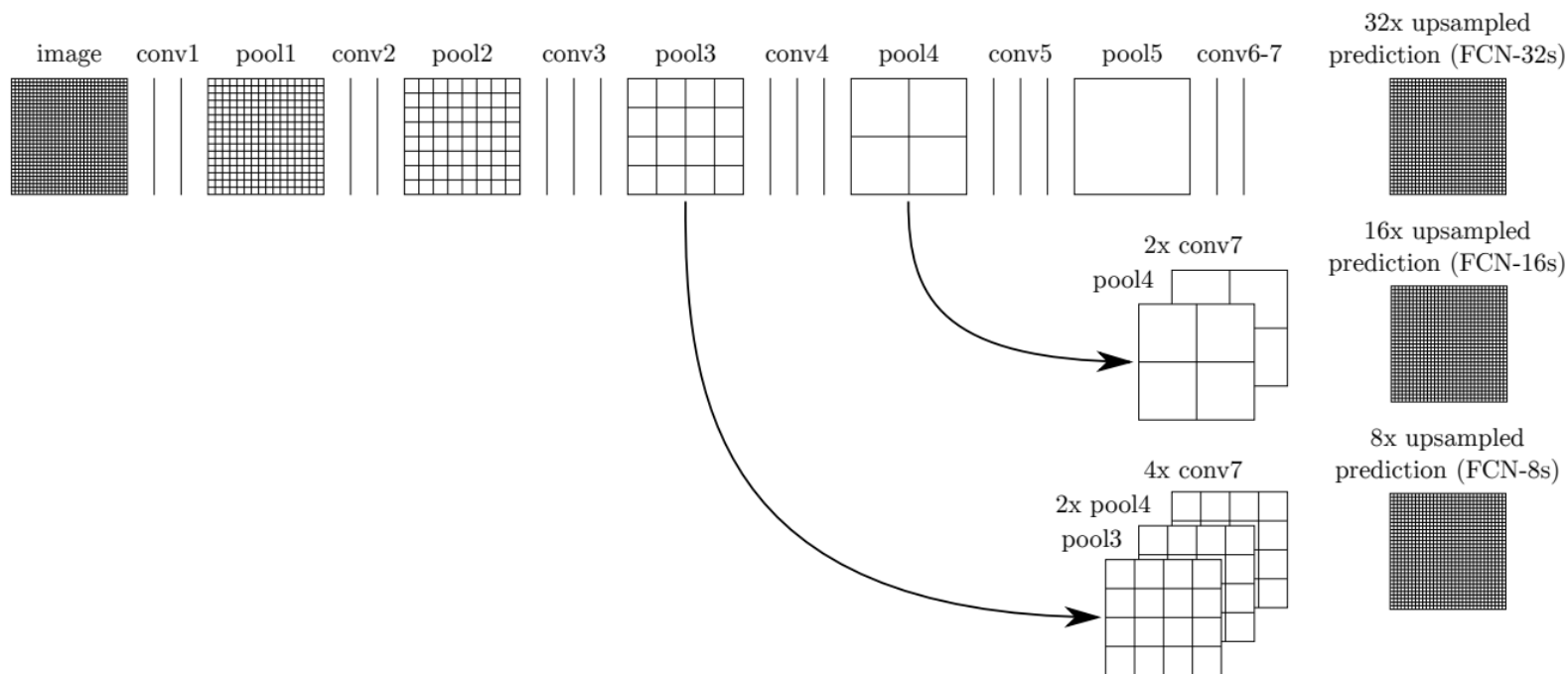
- Идея: применить CNN скользящим окном
- Недостаток – очень низкое разрешение выхода



Fully-convolutional CNN

[Long et al., 2015]

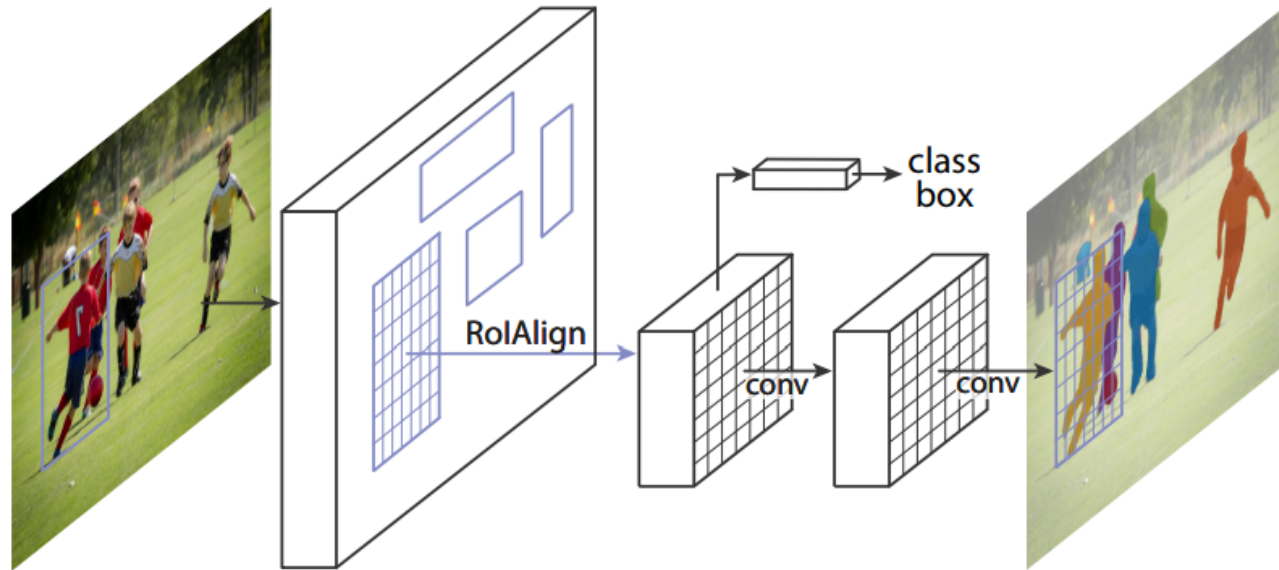
- Идея: применить CNN скользящим окном
- Недостаток – очень низкое разрешение выхода
- Идея: разрешение с помощью более глубоких слоев
- Используются upconv, dilated conv, etc.
- Модели такого типа известны как U-net (детали важны!)



Сегментация объектов: Mask R-CNN

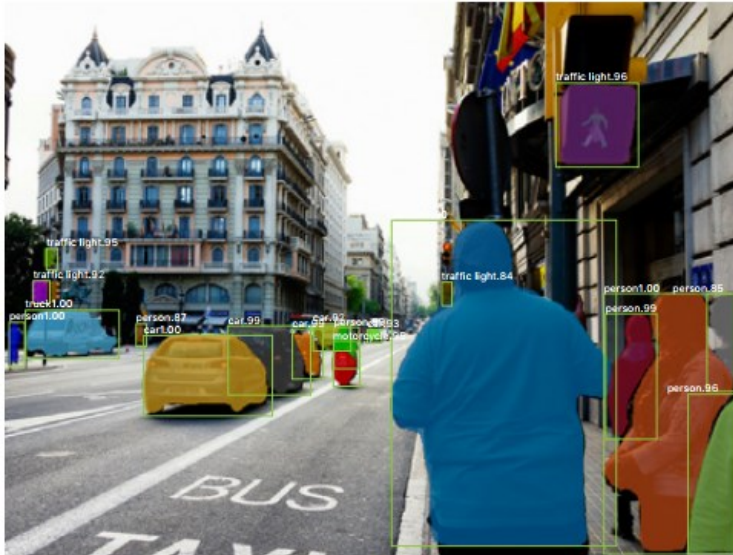
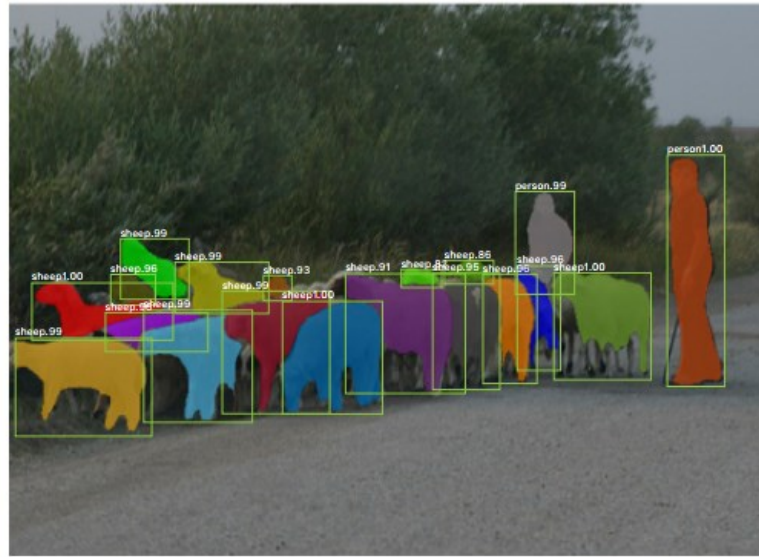
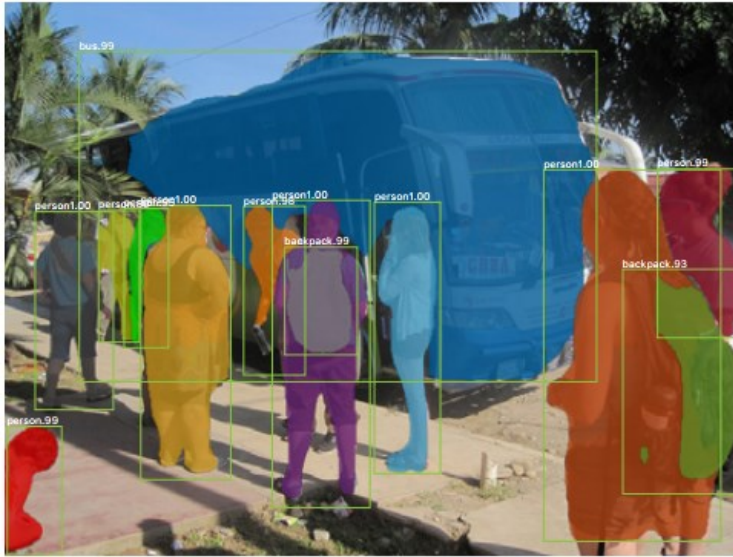
[He et al., 2017]

- Идея: использовать детекцию для сегментации
- Недостаток – из-за maxpool теряется точная позиция
- Идея: использовать «гладкий pooling»
- Билинейная интерполяция границ пикселей



Сегментация объектов: Mask R-CNN

[He et al., 2017]



Часть 3: поиск изображений (retrieval)

- Задача найти похожие изображения
- Задача идентификации (например, лица)
- Подход: описать изображение небольшим вектором (128, 256) и делать поиск ближайших соседей по L2 метрике
- Быстрые приближенный алгоритмы поиска
- Можно использовать предобученные сети
- Обучение специальных признаков!

Сиамские сети (siamese)

- Идея: использовать одну и ту же сеть на двух изображениях, и считать расстояние между признаками
- Вопрос – как обучать?

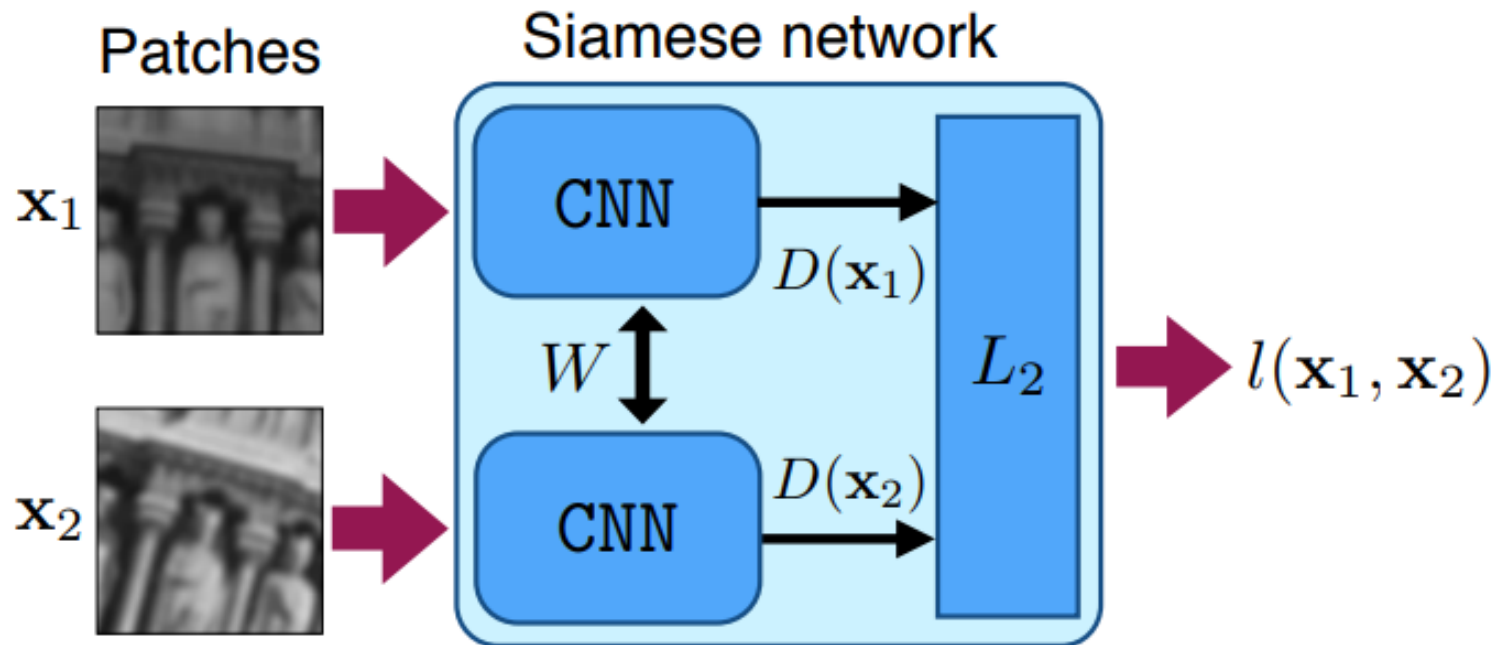


Image from [Simo-Serra et al., 2015]

Сиамские сети (siamese)

- Идея: использовать одну и ту же сеть на двух изображениях, и считать расстояние между признаками
- Вопрос – как обучать?

- Вариант 1 – Contrastive loss
 $y = 1$ - положительная пара
 $y = 0$ - отрицательная пара
 m – margin, чтобы не отталкивать непохожие

$$\ell(x_1, x_2) = y \|x_1 - x_2\|^2 + (1 - y) \max(0, m - \|x_1 - x_2\|)^2$$

Векторы x нормированные!

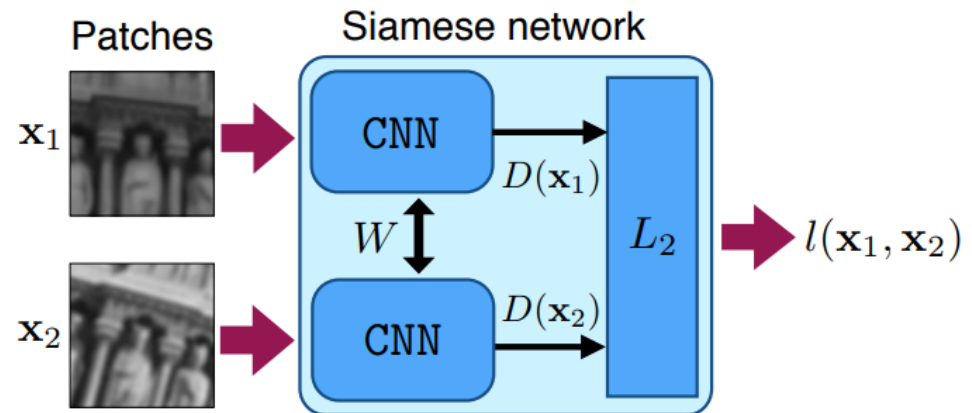
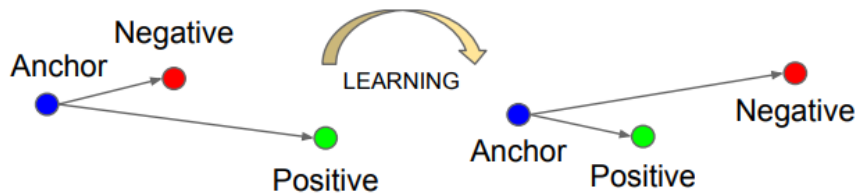


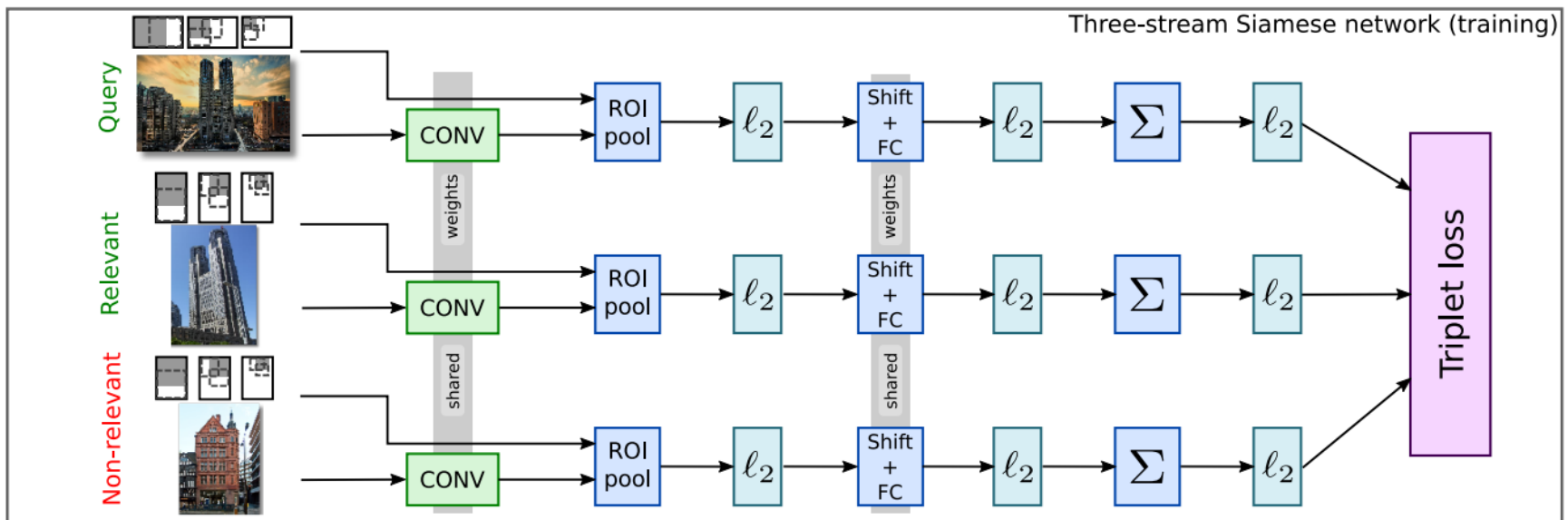
Image from [Simo-Serra et al., 2015]

Сиамские сети (siamese)

- Идея: использовать одну и ту же сеть на трёх изображениях, и считать расстояние между признаками
- Вариант 2 – Triplet loss



$$\max(0, m + \|q - d^+\|^2 - \|q - d^-\|^2)$$

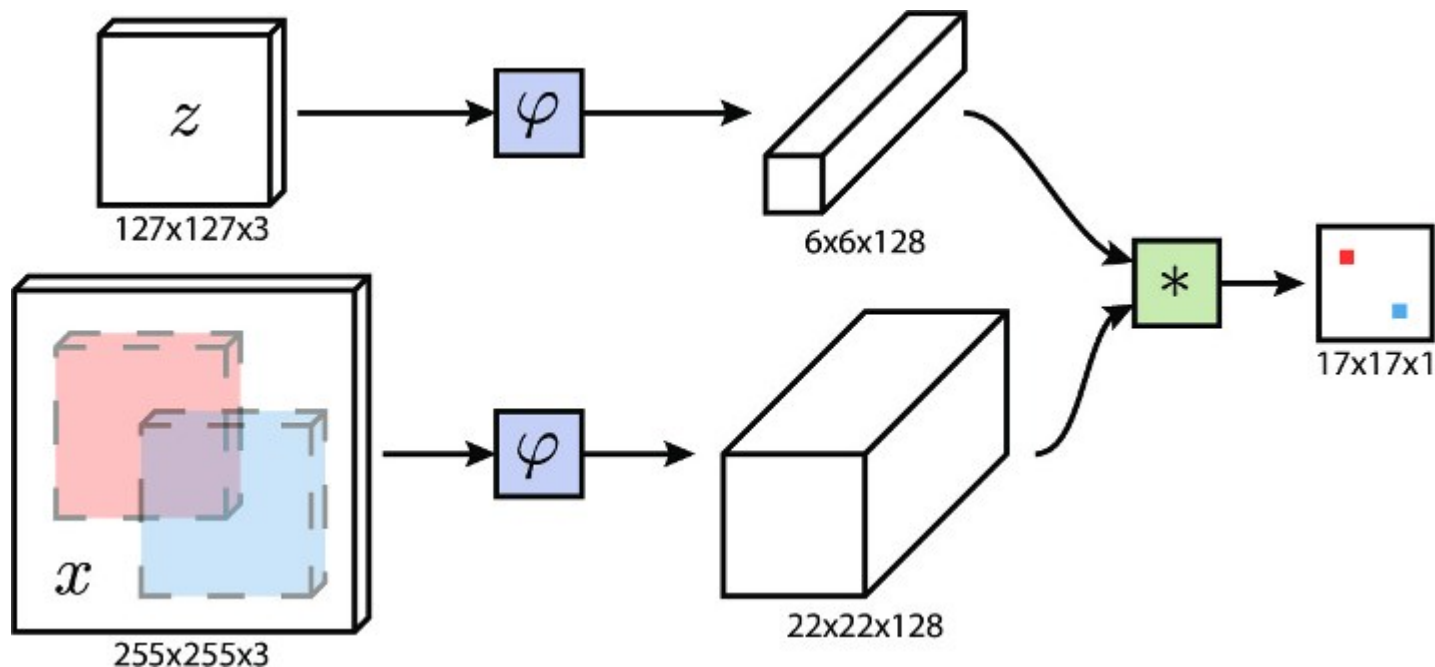


[Simo-Serra et al., 2015; Gordo et al., 2016]

Отслеживание объектов на видео

[Bertinetto et al., 2016]

- Идея: одну из веток сиамский сетей применять свёрточно



Отслеживание объектов на видео

[Bertinetto et al., 2016]

- Идея: одну из веток сиамский сетей применять свёрточно
- Real-time, online



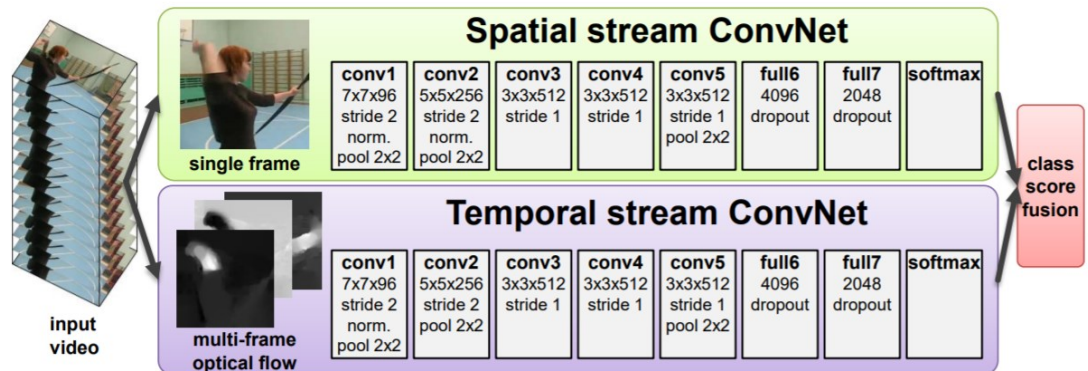
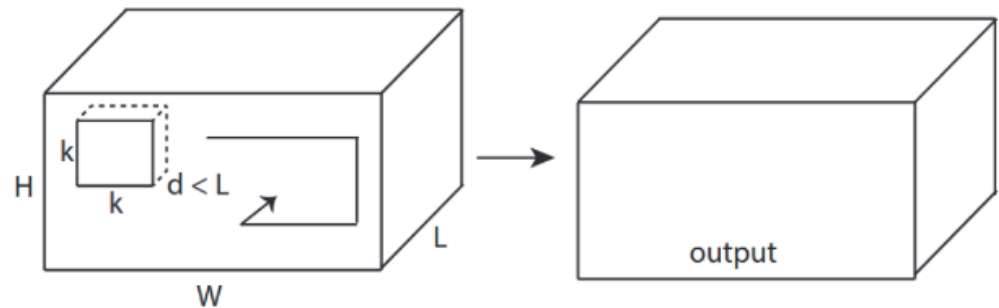
Часть 4: классификация видео

- Задача: распознавание действий на видео



Подходы к видео

- Задача: распознавание действий на видео
- Подходы:
 - Извлечь CNN признаки и каждого кадра и усреднить
 - Рекуррентная сеть над признаками с кадров [Karpathy et al., 2014] (часто работает плохо!)
 - 3D свёртки [Tran et al., 2015]
- Двупоточные сети [Simonyan&Zisserman, 2014]:



Заключение

- Компьютерное зрение активно использует нейросети
- Есть задачи зрения, где нейросети не работают
- Очень большая область
- Одна из самых вычислительно тяжелых областей
- Много специализированных курсов