Deep Learning. Сверточные нейронные сети. Обзор ключевых архитектур. Задача сегментации. Задача локализации и детектирования объектов

Урок 7

Егор Конягин

23 июля 2019 г.

МФТИ & АО ”ЦОСиВТ”

Содержание

CNN. Повторение

CNN. Повторение

Мы обсудили, что полносвязные нейронные сети в задаче анализа изображений будут иметь два существенных недостатка

- огромное кол-во параметров (порядка 40-100 млн);
- неспособность к локальному анализу изображения.

Мы дали определение двумерной дискретной свёртке:

$$g(x, y) = (\mathcal{K} * f)(x, y) = \sum_{s=0}^{n_k} \sum_{t=0}^{n_k} \mathcal{K}_{st} \cdot f_{x-s, y-t}. \tag{1}$$

CNN. Принцип действия

В основе CNN лежит обучаемая свёртка!

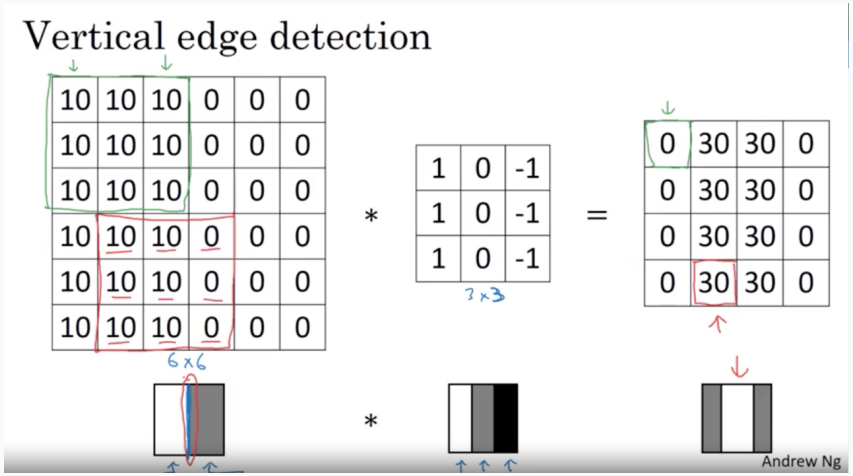

Рис. 1: Вычисление свертки изображения. Источник: Andrew Ng's classes

Если написать уравнения для backward propagation, то есть для вычисления $\frac{\partial J}{\partial w}$, то мы переходим к понятию обучаемой свёртки:

$$W = \begin{pmatrix} w_{11} & w_{12} & w_{13} \\ w_{21} & w_{22} & w_{23} \\ w_{31} & w_{32} & w_{33} \end{pmatrix}. \tag{2}$$

CNN. Принцип действия - II

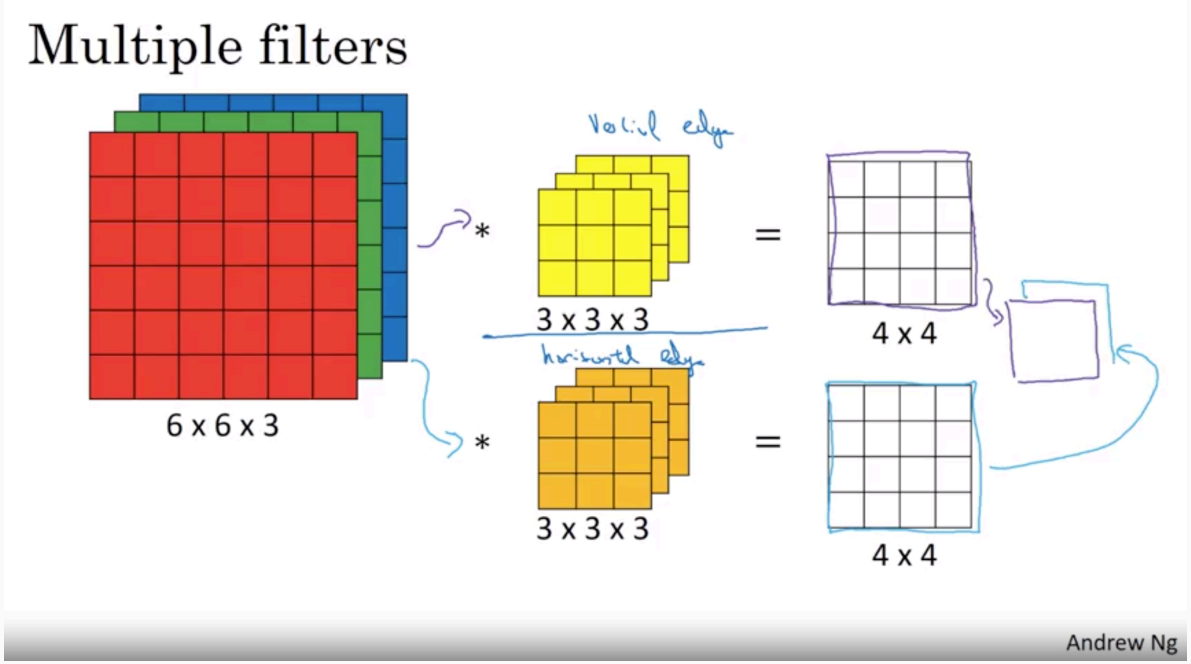

Рис. 2: Свертка над многоканальным изображением. Источник: Andrew Ng's classes

LeNet

LeNet (1998)

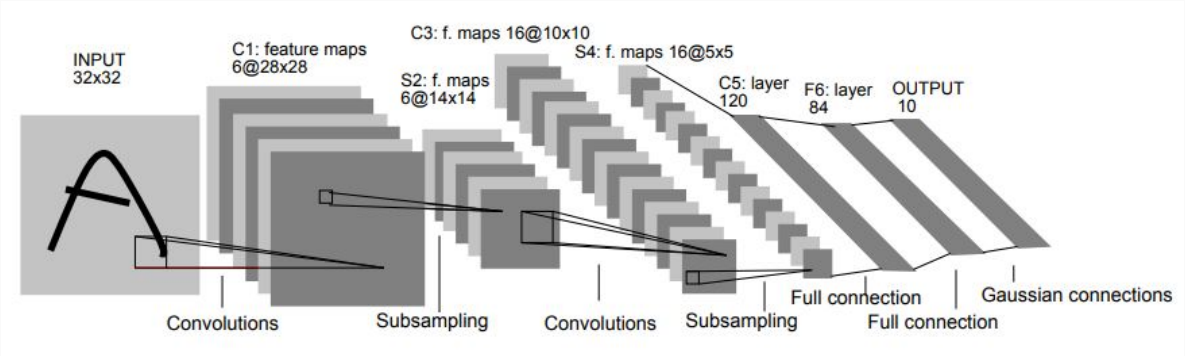

Рис. 3: Архитектура LeNet. Источник: original paper

AlexNet (2012)

AlexNet

Статья по AlexNet была опубликована в 2012 году Алексом Крижевским. Её задача - тоже многоклассовая классификация.

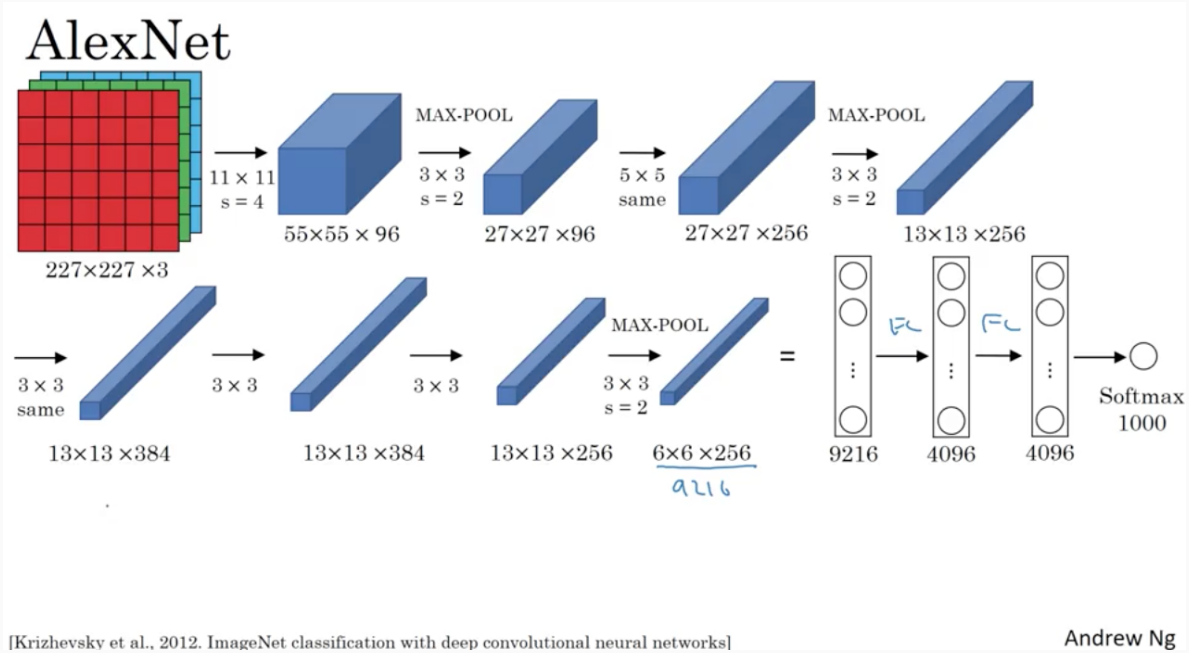

Рис. 4: Архитектура AlexNet. Источник: Andrew Ng's classes

Нейросеть содержит 62 000 000 параметров, из которых 58 000 000 приходится на три последние полносвязных слоя.

VGG-16

VGG-16 (2014)

Данная работа была разработана в 2014 году группой компьютерного зрения Visual geometry group, а именно Кареном Симоньяном и Эндрю Зиссерманом (Very Deep Convolutional Networks for Large-Scale Image Recognition).

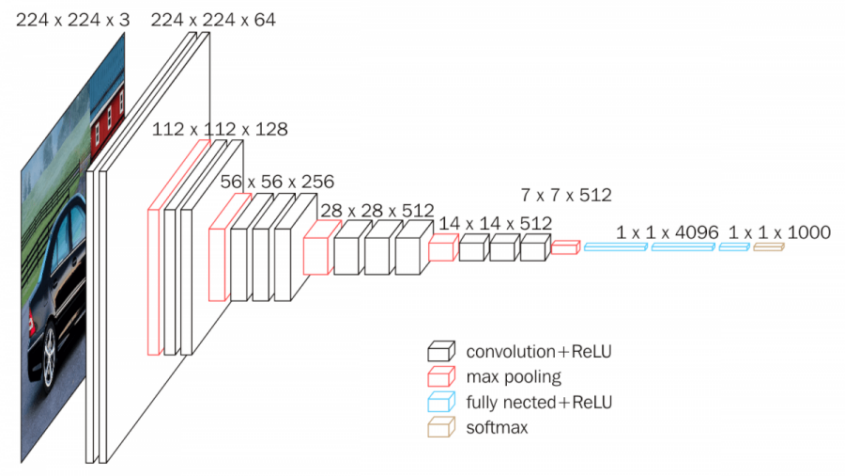

Рис. 5: Архитектура VGG-16

В отличие от AlexNet, все фильтры во всех слоях имеют размер 3x3.

VGG-16 - II. Недостатки

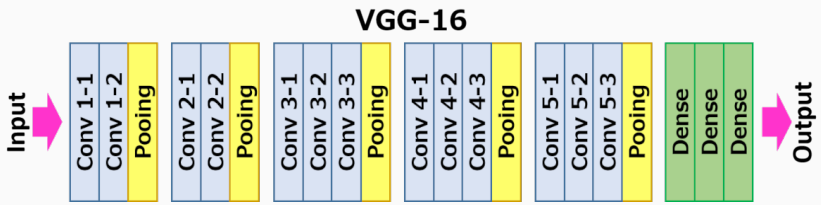

Рис. 6: Архитектура VGG-16. Слои

К сожалению, архитектура VGG-16 обладает двумя недостатками:

1. данная нейросеть обучается слишком медленно;
2. нейросеть имеет много параметров, которые занимают порядка 500 МБ при разворачивании (138 млн параметров).

GoogLeNet

GoogLeNet (2014)

Данная нейронная сеть существенно отличается от всех предыдущих нейросетей. Нам придется познакомиться со следующими концепциями перед тем, как непосредственно расмотреть эту архитектуру:

- свёртка с фильтром 1x1 - нейросеть в нейросети;
- модуль inception;
- global average pooling;
-

GoogLeNet. 1x1 convolution

Рассмотрим две модели сверточной нейросети и посчитаем кол-во совершаемых операций:

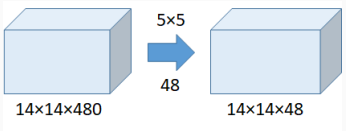

Рис. 7: Сверточный слой

Кол-во операций = (14×14×48)×(5×5×480) = 112.9M

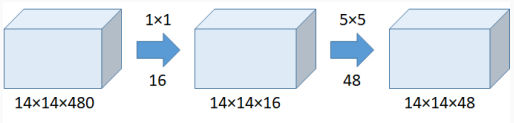

Рис. 8: Сверточный слой с использованием 1x1 convolution

Кол-во операций для свертки 1×1 = (14×14×16)×(1×1×480) = 1.5M
Кол-во операций для свертки 5×5 = (14×14×48)×(5×5×16) = 3.8M
Всего операций = 1.5M + 3.8M = 5.3M « 112.M.

GoogLeNet. Inception module

Рассмотрим следующую архитектуру, которая называется блоком inception module:

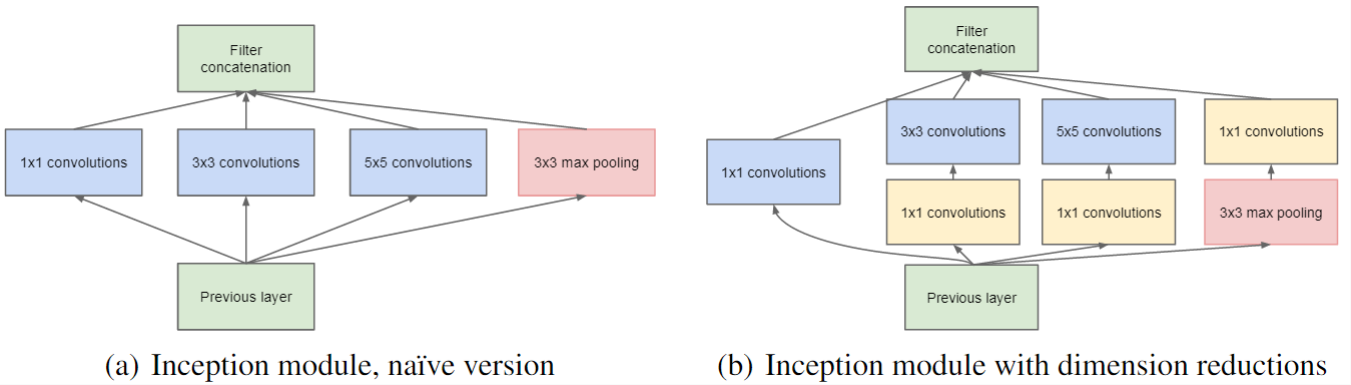

Рис. 9: Архитектура блока inception module. Свертка 1x1 нужна для снижения кол-ва операций

GoogLeNet. Global average pooling

Данный метод применяется для снижения кол-ва весов при переходе от сверточных к полносвязным слоям.

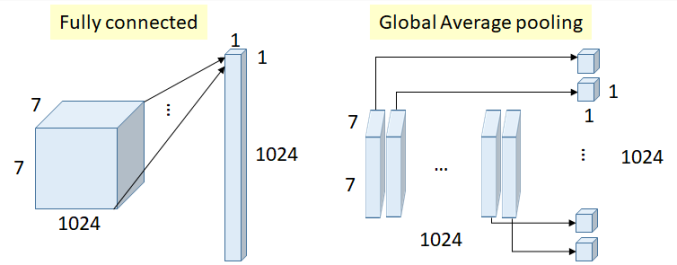

Рис. 10: Суть global avg pooling

Как было ранее замечено, большинство параметров в свёрточных нейросетях расположены в последних полносвязных слоях.

Кол-во весов слева: 7×7×1024×1024 = 51.3M.

Кол-во весов справа: 0. В данном случае считается среднее по каждой из матриц 7x7, это число записывается в 1024-мерный вектор.

GoogLeNet. Архитектура

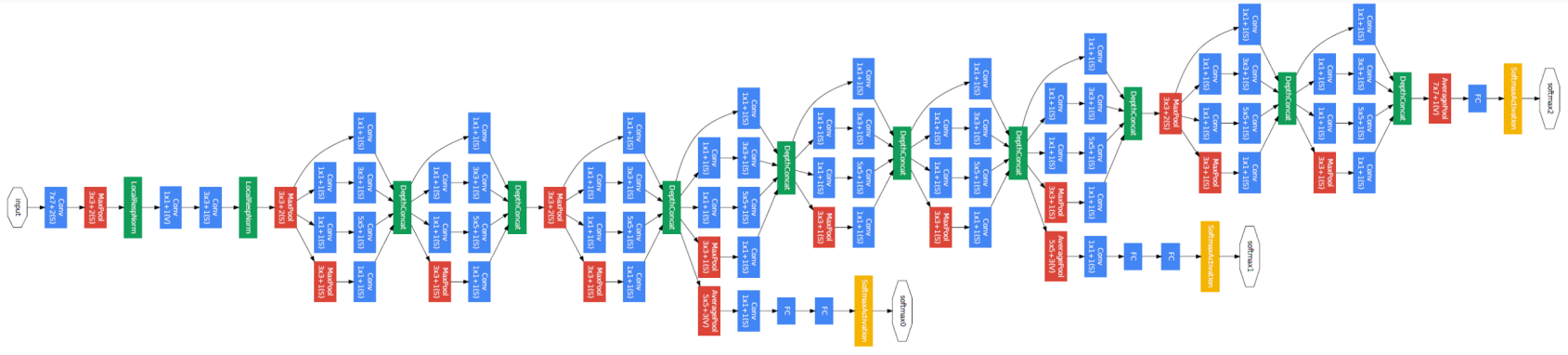

Рис. 11: Архитектура GoogLeNet

Несмотря на кажущуюся сложность, данная нейросеть имеет меньше 7 миллионов параметров (сравните с VGG-16). Будучи в 15 раз ”легче чем VGG-16, она не сильно уступила ей по качеству в классификации изображений.

ResNet

ResNet. Skip connection

Основу нейросети ResNet составляет т.н. блок skip (shortcut) connection:

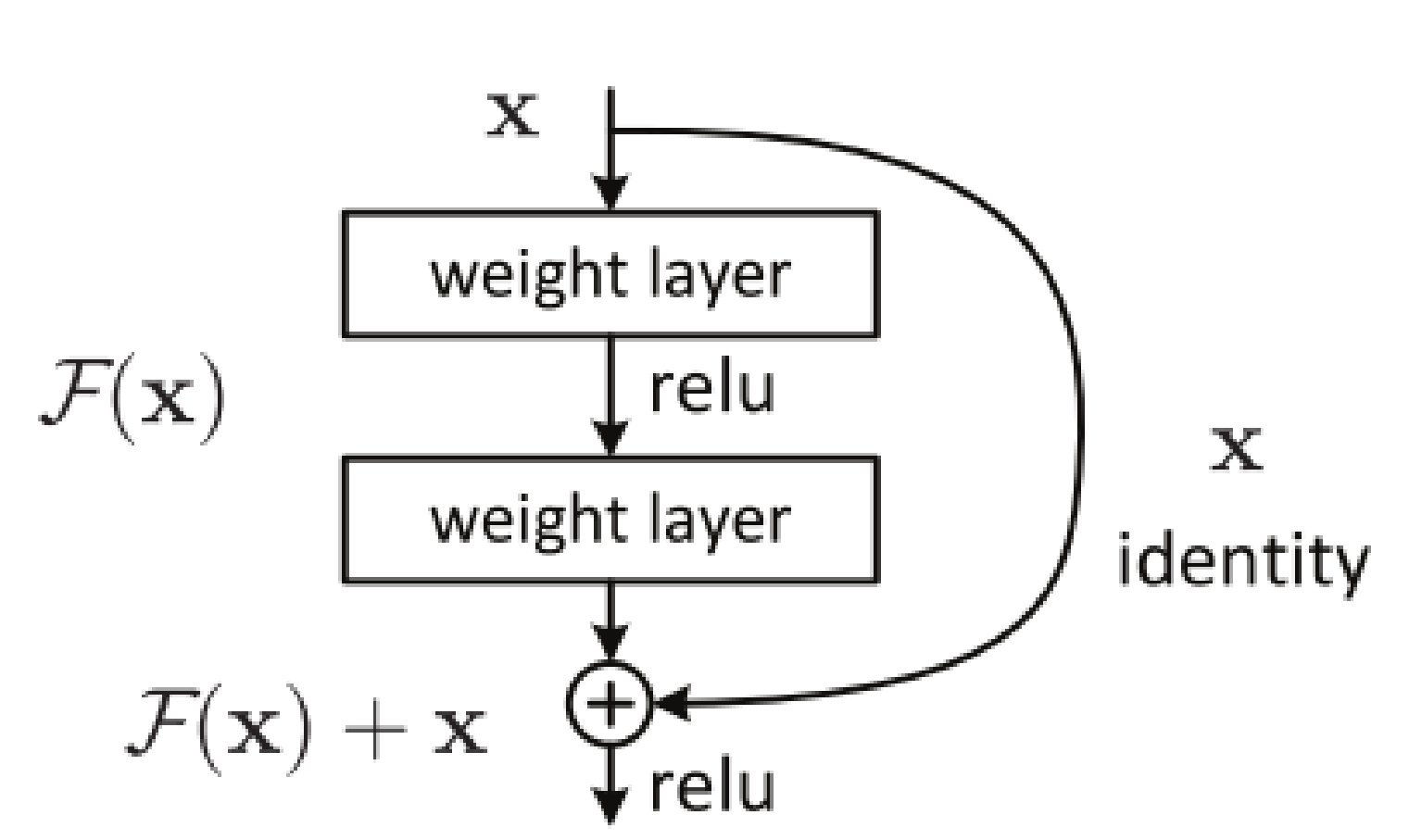

Рис. 12: Архитектура skip connection

ResNet. Архитектура

К сожалению, нейросеть ResNet слишком большая, чтобы уместиться на одной картинке :(отметим, что она состоит из сверточных skip connection-слоёв, которые завершаются одним fc-слоем (т. е. полносвязным). Кол-во слоев в сети варьируется в различных модификациях от 34 до 1002.

Сложность нейросети: примерно 20 миллионов параметров.

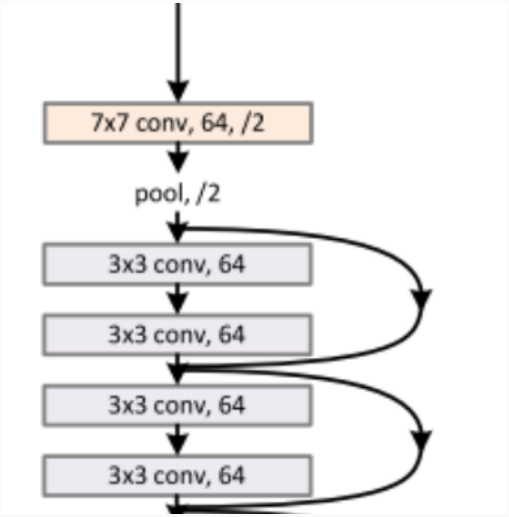

Рис. 13: Архитектура ResNet. Фрагмент

UNet. Краткий обзор задачи сегментации

Постановка задачи сегментации

Сегментация (sematic segmentation) - это задача попиксельной классификации изображения. Таким образом, изображению сопоставляется изображение, называемое маской, где каждый пиксель исходного отнесен к тому или иному классу:

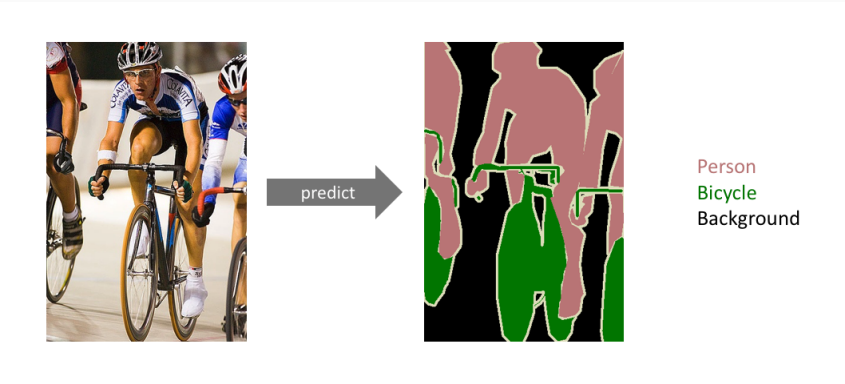

Рис. 14: Пример семантической сегментации

Convolution transposed

Пусть есть фильтр свертки размера 3x3 с данной матрицей:

$$W = \begin{pmatrix} w_{00} & w_{01} & w_{02} \\ w_{10} & w_{11} & w_{12} \\ w_{20} & w_{21} & w_{22} \end{pmatrix} \tag{3}$$

Тогда операцию свертки над изображением 4x4 можно представить в матричном виде (4,16):

$$C = \begin{pmatrix} w_{00} & w_{01} & w_{02} & 0 & w_{10} & w_{11} & w_{12} & \cdots \\ 0 & w_{00} & w_{01} & w_{02} & 0 & w_{10} & w_{11} & \cdots \\ 0 & 0 & 0 & 0 & w_{00} & w_{01} & w_{02} & \cdots \\ 0 & 0 & 0 & 0 & 0 & w_{00} & w_{01} & \cdots \end{pmatrix} \tag{4}$$

Convolution transposed - II

Таким образом, умножаться эта матрица будет на 16-мерный вектор, полученный ”вытягиванием”матрицы 4x4 в столбец. Запишем все размерности:

$$n_{(4,16)} \cdot n_{(16,1)} = n_{(4,1)} \tag{5}$$

4-мерный вектор преобразуем в матрицу 2x2, получим результат, который мы бы получили при свертке изображения.

Теперь рассмотрим матрицу C^T:

$$n_{(16,4)} \cdot n_{(4,1)} = n_{(16,1)}. \tag{6}$$

Тогда:

$$n_{(16,4)} \cdot n_{(4,1)} = n_{(16,1)}. \tag{7}$$

Итак, мы научились создавать матрицу для повышения размерности изображения!

UNet (2015). Обзор архитектуры

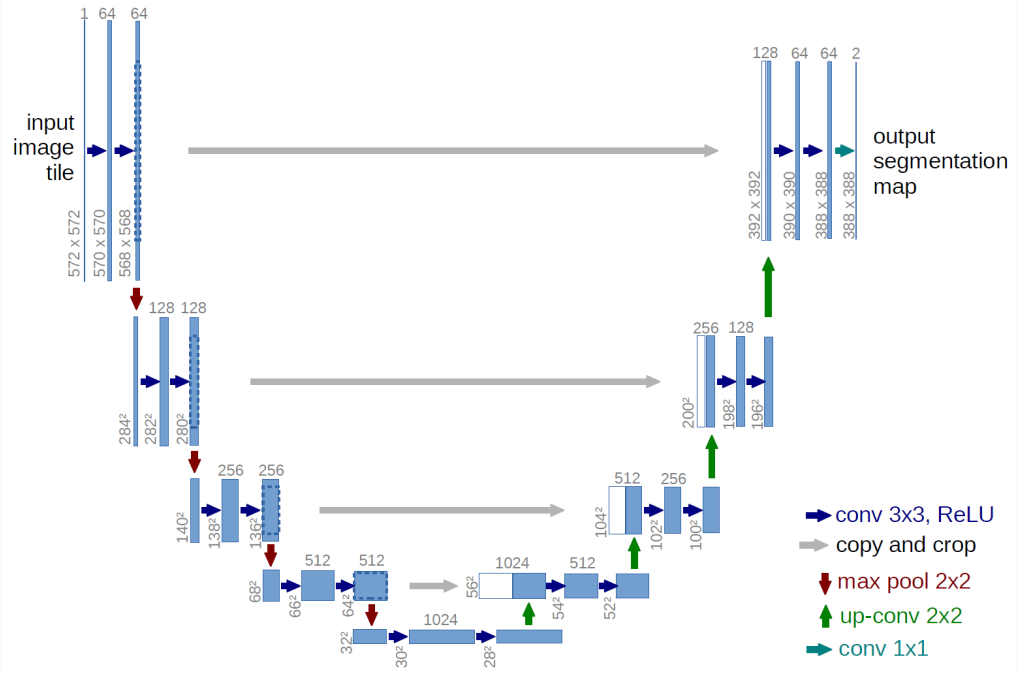

Рис. 15: Архитектура UNet

Исходя из поставленной задачи, выводом данной нейросети должно быть изображение, причем того же размера, что и входное изображение. Исходя из этого требования, эта архитектура построена только на свёрточных слоях (conv и conv_transpose).

Summary

Мы рассмотрели следующие примеры нейронных сетей:

1. LeNet;
2. AlexNet;
3. VGG-16;
4. GoogLeNet;
5. ResNet;
6. UNet.