

Идеи современных архитектур нейросетей для классификации изображений

Практикум на ЭВМ 2017/2018

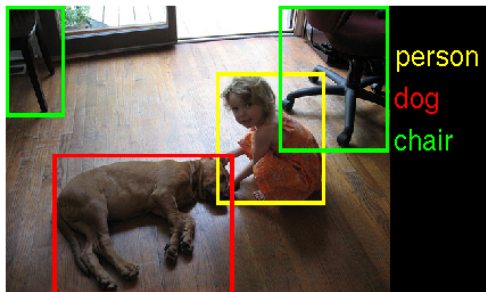
Сагайдак Олег Игоревич

МГУ имени М. В. Ломоносова, факультет ВМК, кафедра ММП

22 марта 2018 г.

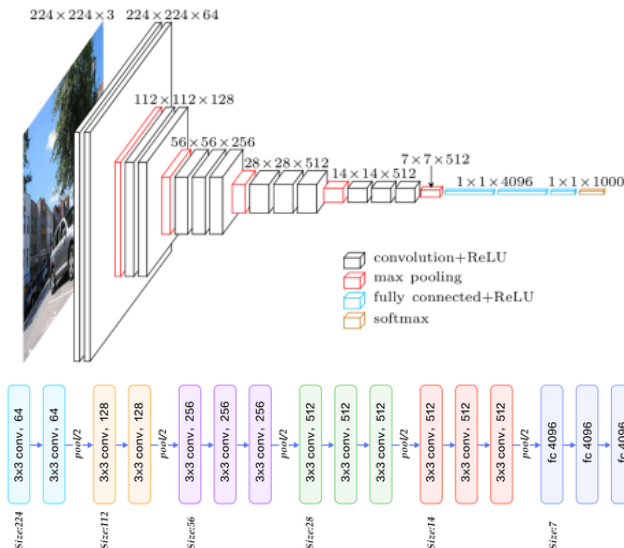
- ImageNet — крупная база данных изображений, на данный момент насчитывает более 14 миллионов вручную размеченных фотографий с 20 тысячами категорий;
- ImageNet Large Scale Visual Recognition Challenge (ILSVRC) — соревнование, которое проводится каждый год и в котором соревнуются самые современные методы распознавания образов в изображениях;
- В соревновании используется урезанное множество категорий (только 1000).

Примеры картинок



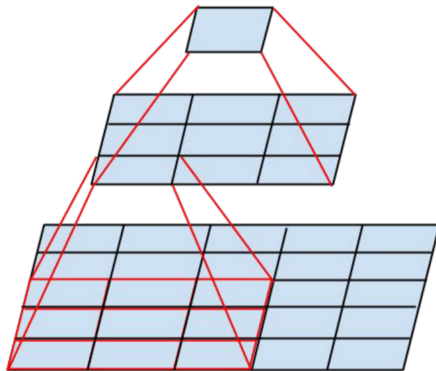
- Есть проблема, что на одном изображении находится сразу несколько объектов;
- Чтобы не наказывать алгоритм за то, что он нашёл не то, что нужно было, используется метрика top-5 error rate;
- Алгоритм выдаёт 5 категорий, в которых больше всего уверен и ошибкой является отсутствие правильного ответа среди этих 5 категорий.

Архитектура VGG

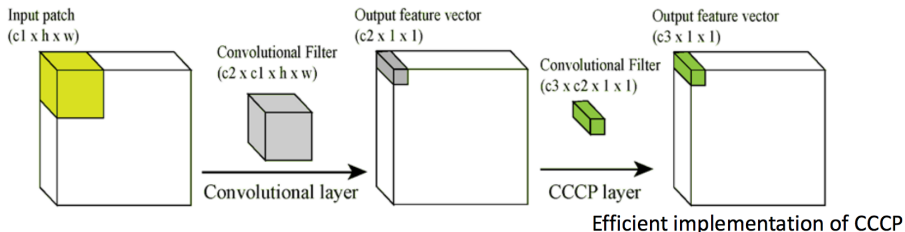


Сокращаем количество параметров

Идея — заменить свёртку большого размера на последовательные свёртки меньшего размера, например, заменим 5×5 на две последовательные 3×3 .



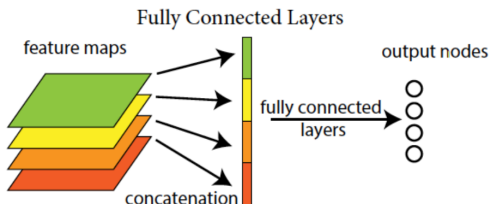
Cascaded Cross-Channel Pooling — использование 1×1 свёрток для изменения глубины представления изображения. Значения между каналами часто скоррелированы, поэтому мы не теряем много информации.



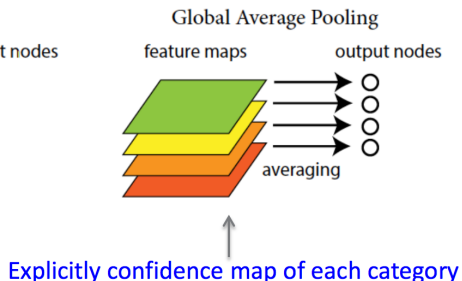
Отказ от полносвязных слоёв

Полносвязные слои очень дорогостоящие, поэтому их можно заменить с помощью CCCP и global average pooling.

CNN



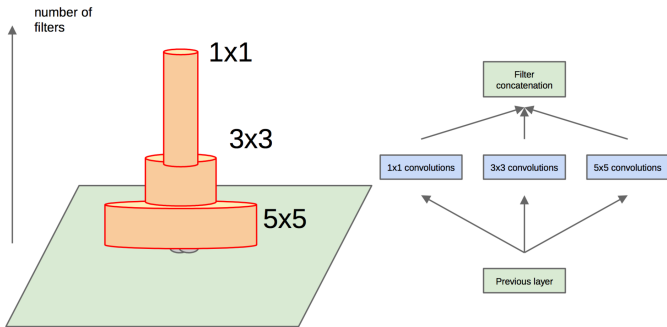
NIN



Свёртки разного размера

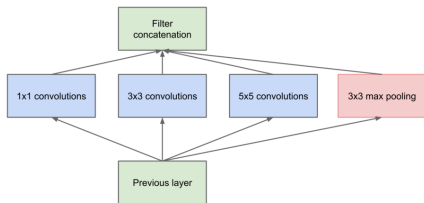
Соседние нейроны часто хранят коррелирующие значения, поэтому на следующем слое хотелось бы объединить их в один признак. Но масштаб таких признаков может быть разный, больше всего признаков находится в области 1×1 , чуть меньше в 3×3 и ещё меньше в 5×5 .

Schematic view (naive version)

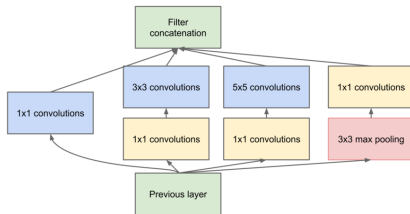


Inception-блок

Введём понятие Inception-блока. 1×1 свёртки используются для сокращения размерности, увеличения размерности и выделения признаков.



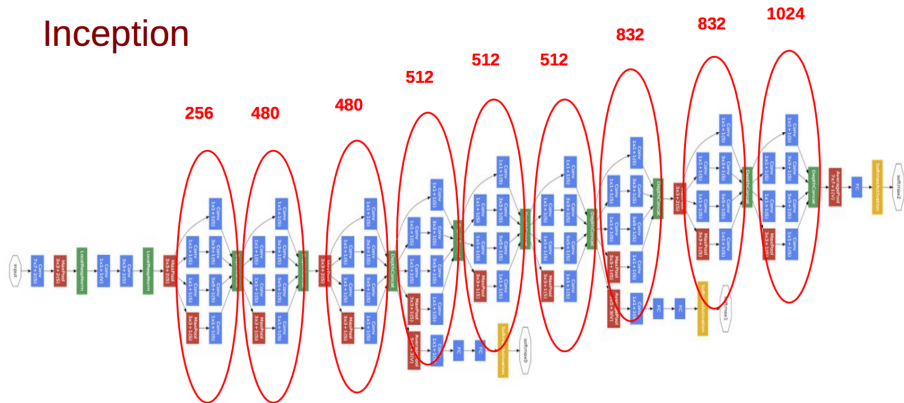
(a) Inception module, naïve version



(b) Inception module with dimension reductions

Figure 2: Inception module

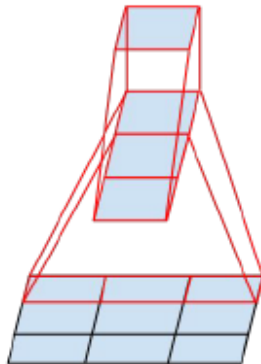
Inception



Width of **inception modules** ranges from 256 filters (in early modules) to 1024 in top inception modules.

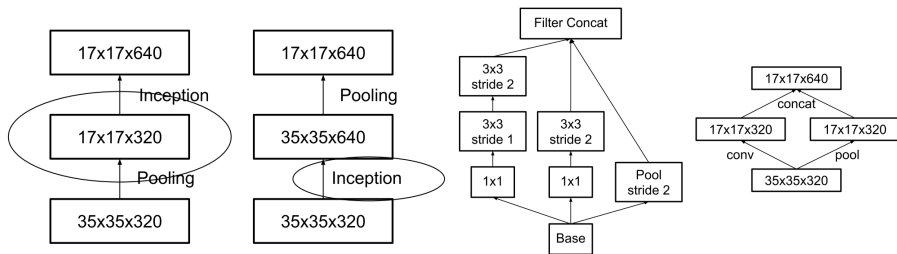
Сокращаем количество параметров ($n \times 1$)

Чтобы ещё сильнее сократить количество параметров, свёртку 3×3 можно представить как две свёртки 3×1 и 1×3 . То же самое можно сделать и для больших свёрток.

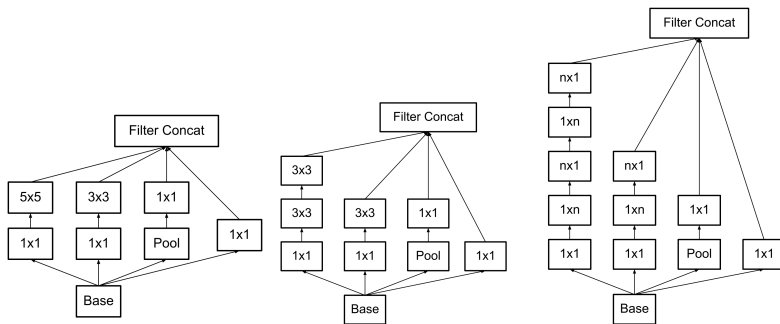


Бутылочное горлышко (bottleneck)

Резкое сокращение количества параметров ведёт к ухудшению качества. Поэтому предлагается проводить параллельно свёртку со $\text{stride} > 1$ и max pooling.



Посмотрим, как менялся Inception-блок после применения тех трюков, о которых мы говорили.

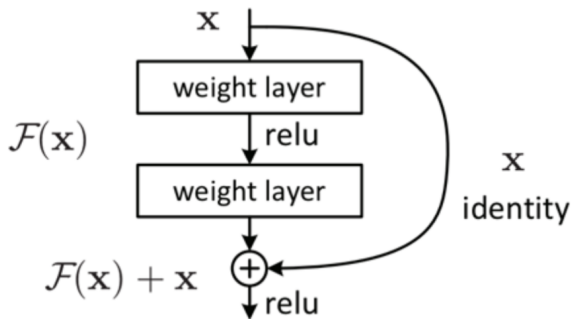


Идея ResNet

В 2015 году соревнование ImageNet выиграла новая архитектура ResNet, в которой строительным блоком является остаточный слой (residual layer).

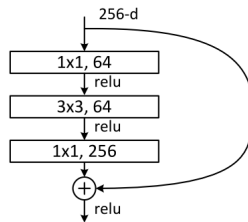
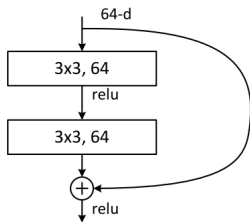
Идея заключается в том, что результат слоя не сразу подаётся на следующий, а предварительно складывается со своим входом.

Слой должен получить нужный результат не с нуля, а лишь найти разницу между входом и выходом, отсюда и название.



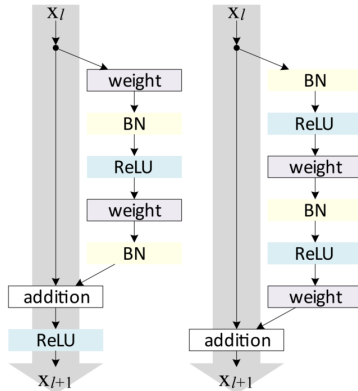
Скрытое сокращение размерности

Поскольку размеры входа и выхода должны совпадать, предлагается перед свёрткой сжать размерность представления, а потом расширить её обратно.



Перенос активации

Строго говоря, применять функцию активации после сложения некорректно. Чтобы информация перешла полностью через всю сеть, предлагается производить активацию перед применением свёрточного слоя.



(a) original

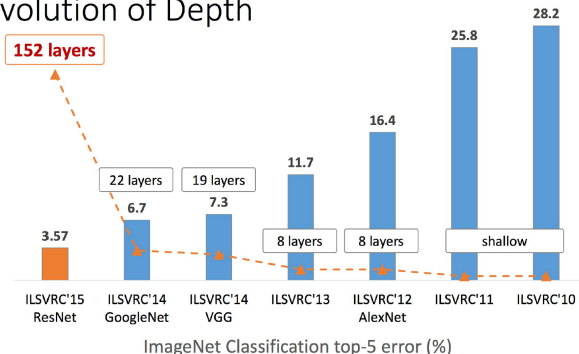
(b) proposed

Скачок глубины сети

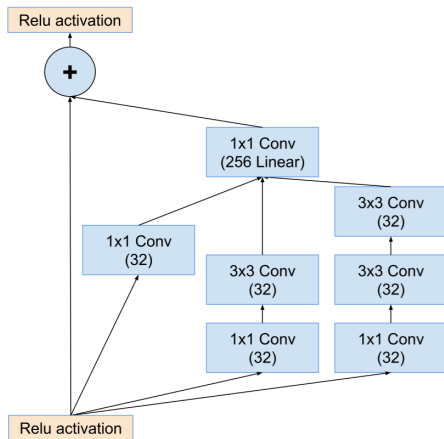
Благодаря тождественным связям сеть сама может решать, насколько глубокой ей нужно быть, что позволило резко увеличить количество слоёв без потери качества.

Какие-то признаки выделяются на первых слоях и передадутся к выходу, а какие-то удастся вычислить только ближе к концу сети.

Revolution of Depth



В дальнейшем Google позаимствует концепцию ResNet и модифицирует вид своего inception-блока.



Обзор топологий глубоких сверточных нейронных сетей, 2016 г.

ImageNet: VGGNet, ResNet, Inception, and Xception with Keras, 2017 г.

Neural Network Architectures, 2017 г.

Эволюция нейросетей для распознавания изображений в Google: GoogLeNet, 2016 г.

Эволюция нейросетей для распознавания изображений в Google: Inception-v3, 2016 г.

Эволюция нейросетей для распознавания изображений в Google: Inception-ResNet, 2016 г.