

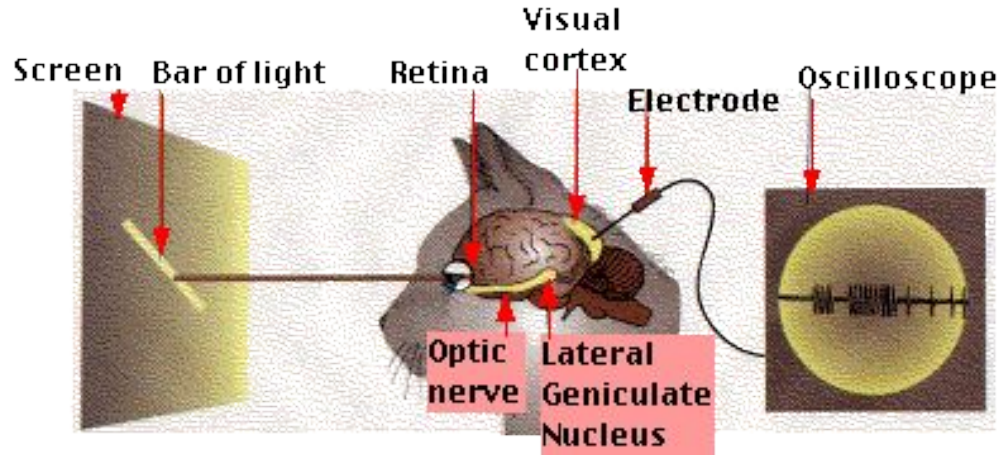
Новые архитектуры для зрения

Антон Кудинов, Даниил Петроченко

ИСТОКИ

эксперимент с котиками

Дэвид Хьюбел и Торстен Визель: “Принципы переработки информации в нейронных структурах”



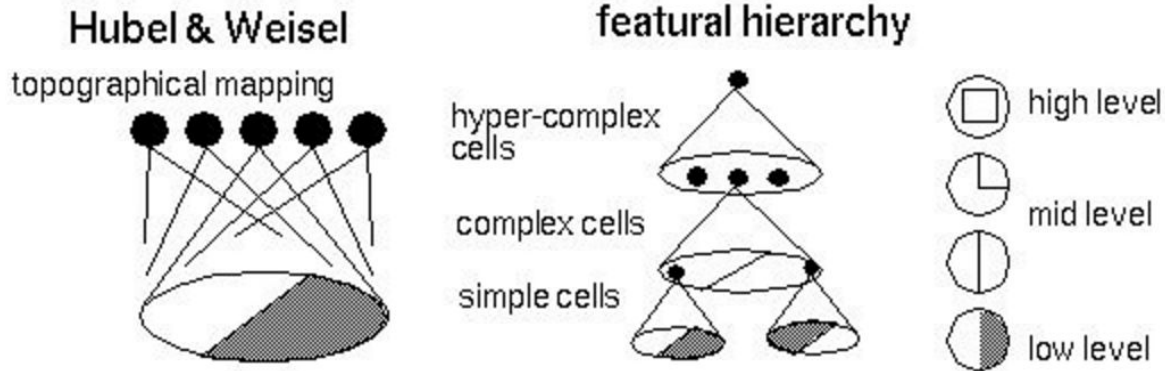
- определенные области зрительной коры активируются только тогда, когда линия проецируется на определенную часть сетчатки;
- уровень активности нейронов области изменяется при изменении угла наклона прямоугольника;
- некоторые области активируются только тогда, когда объект движется в определенном направлении.

Истоки

эксперимент с котиками

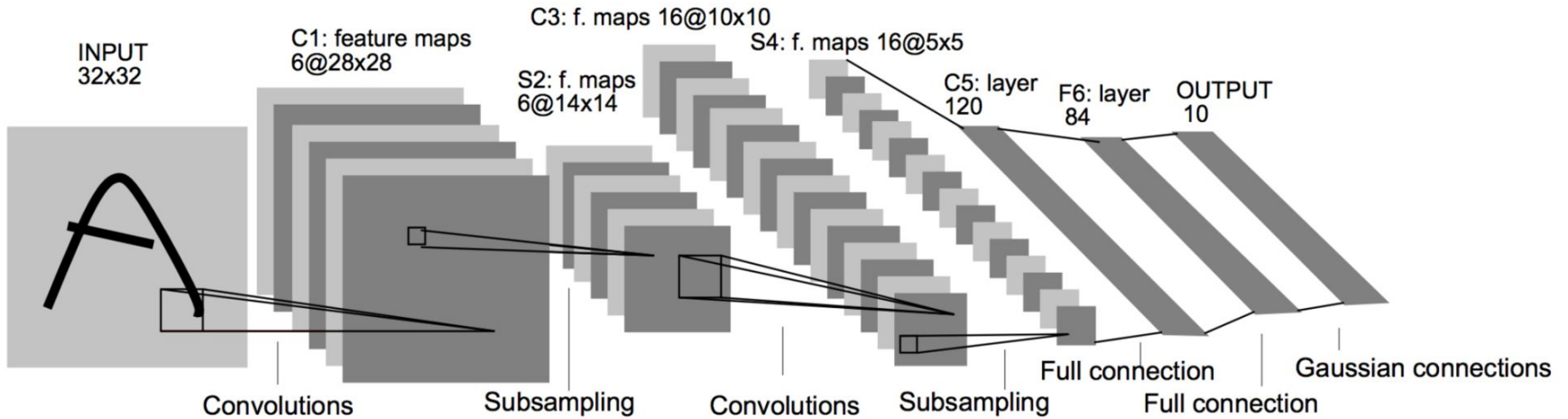
модель зрительной системы со свойствами:

- соседние нейроны обрабатывают сигналы с соседних областей сетчатки;
- нейроны образуют иерархическую структуру (изображение ниже), где каждый следующий уровень выделяет все более и более высокоуровневые признаки.
- нейроны организованы в так называемые колонки — вычислительные блоки, которые трансформируют и передают информацию от уровня к уровню.



LeNet

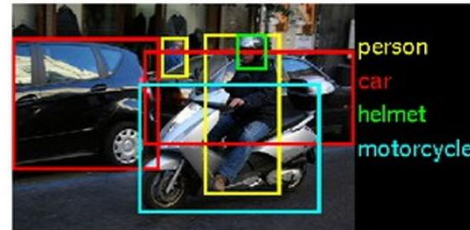
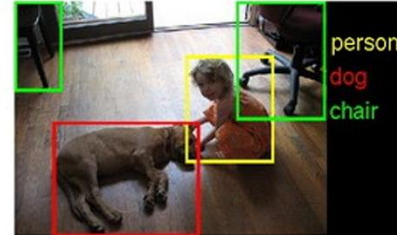
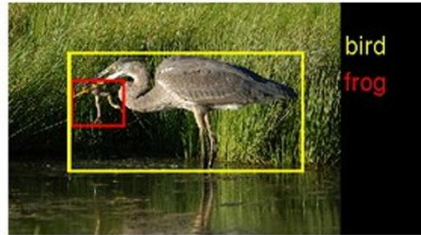
by Yann LeCun



ImageNet classification

ImageNet Large Scale Visual Recognition Challenge (ILSVCR) - соревнование по масштабному распознаванию графических образов, спонсируемое проектом ImageNet, который представляет собой огромную базу изображений. В настоящее время в базе имеется более 14 миллионов изображений.

Смысл соревнования в том, чтобы выявить на произвольно взятом графическом изображении определенные объекты, которые можно выразить осмысленным текстом (словами).



ImageNet classification

Один из методов оценки алгоритмов в соревновании - это top-5 error метрика. Метрика, в которой алгоритм выдает 5 лучших вариантов класса картинки и ошибка засчитывается, если среди всех этих вариантов нет правильного.

cat



dog

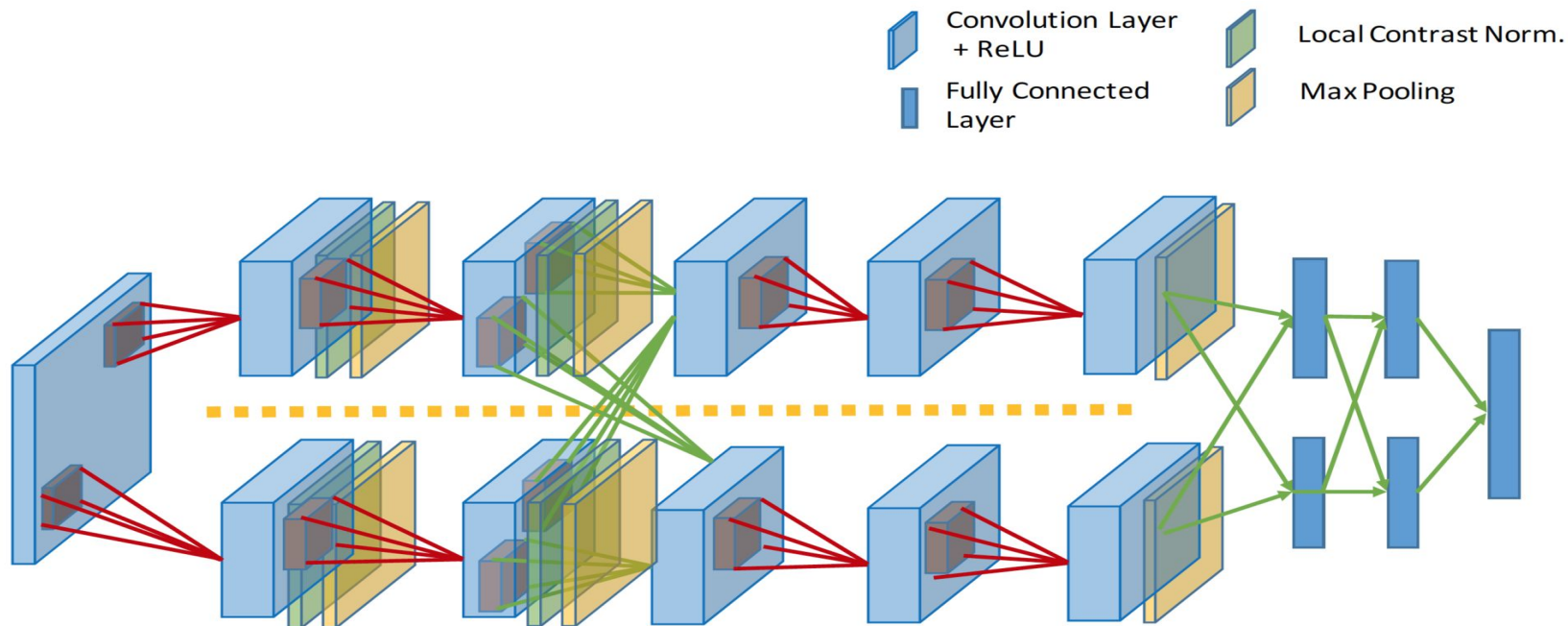


?



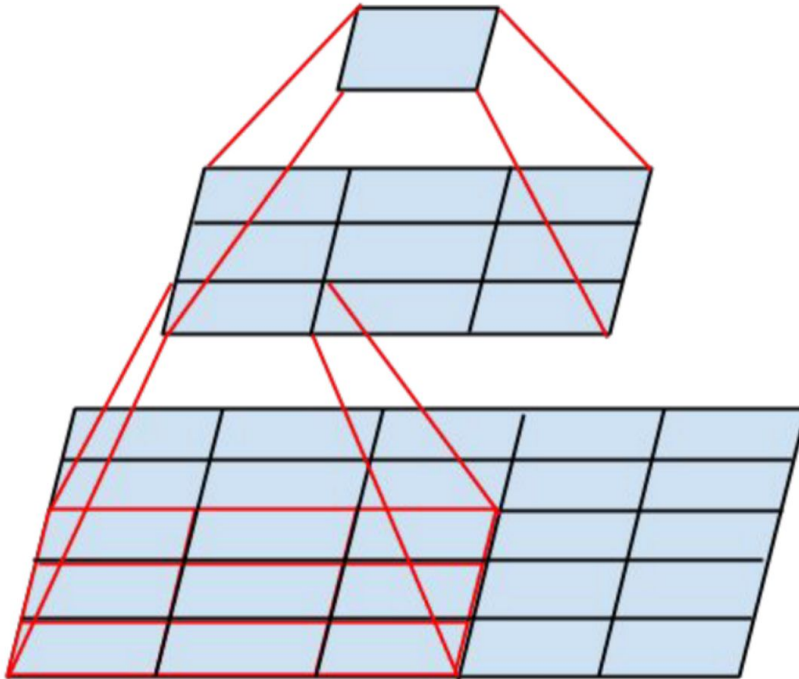
AlexNet

Глубокое обучение = модель + теория обучения + большие данные + железо.
На три порядка больше параметров, два графических ускорителя. 8 уровней (5 convolutional и 3 fully-connected). Размер ядер свертки уменьшается от входа сети к выходу



VGG

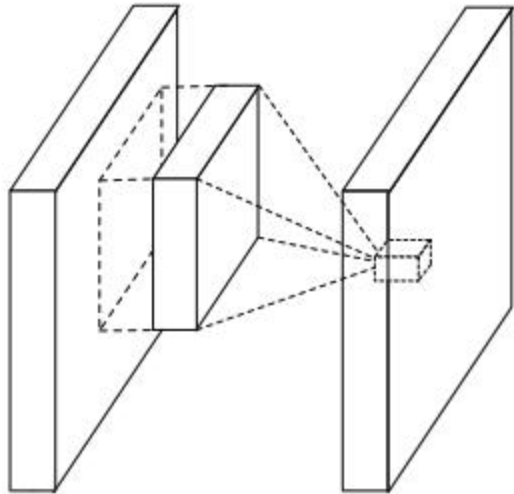
VGG-19 состоит из 144 миллионов параметров.



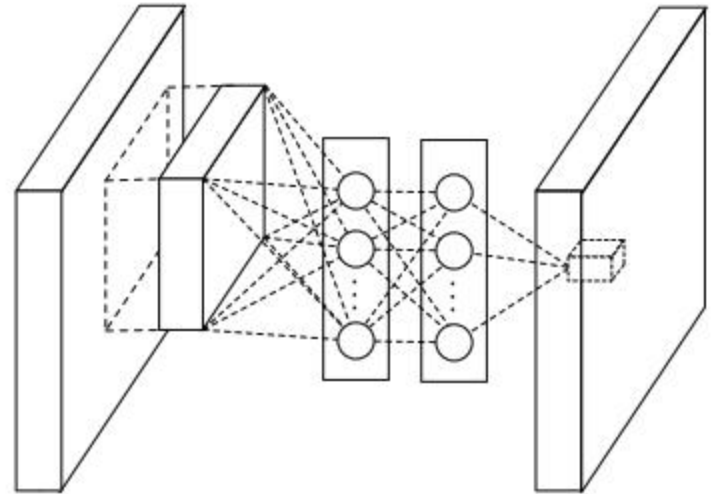
Network in network

Проблема: в сверточной сети чрезвычайно много признаков - ядер свертки.

Решение проблемы: заменить свертки на маленькие нейросети.



(a) Linear convolution layer



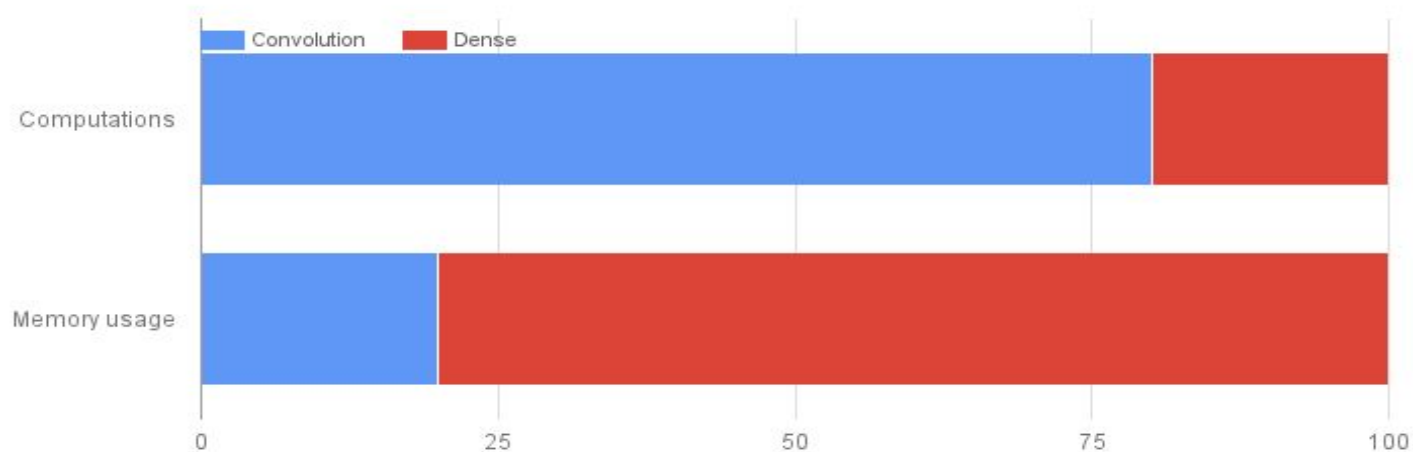
(b) Mlpconv layer

Network in network: CSCP pooling

Новая проблема: огромное количество параметров во внутренних нейросетях

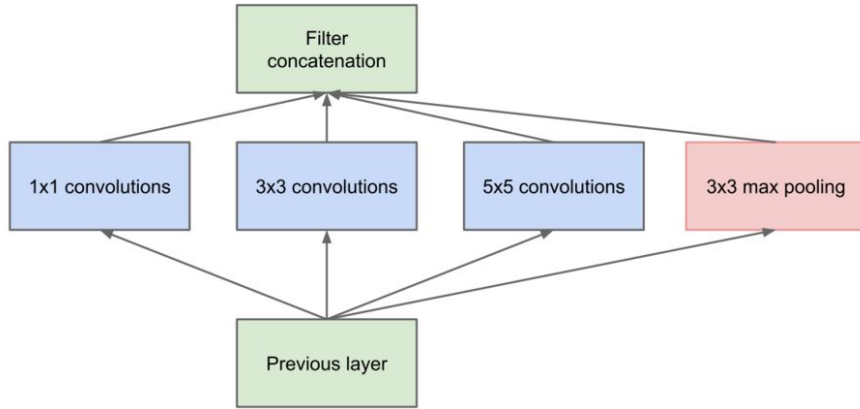
Решение новой проблемы: сделать внутренние сети сверточными - Cascaded Cross Channel
Parameter pooling

Network in network

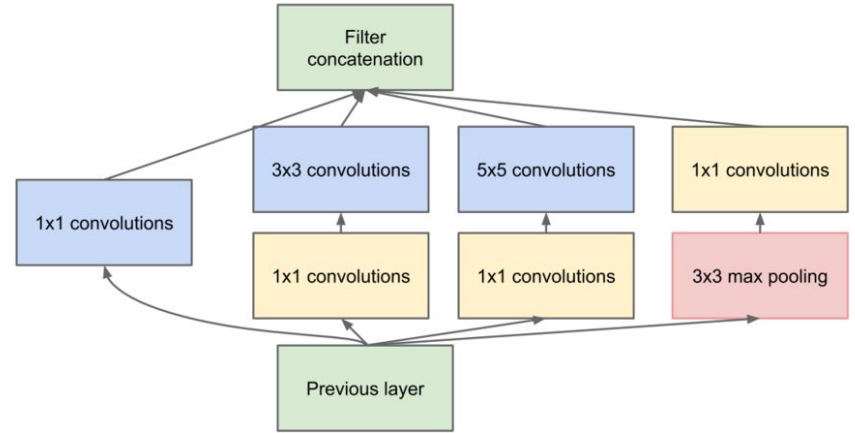


Еще одна идея - полный отказ от полносвязных слоев.

Inception



(a) Inception module, naïve version

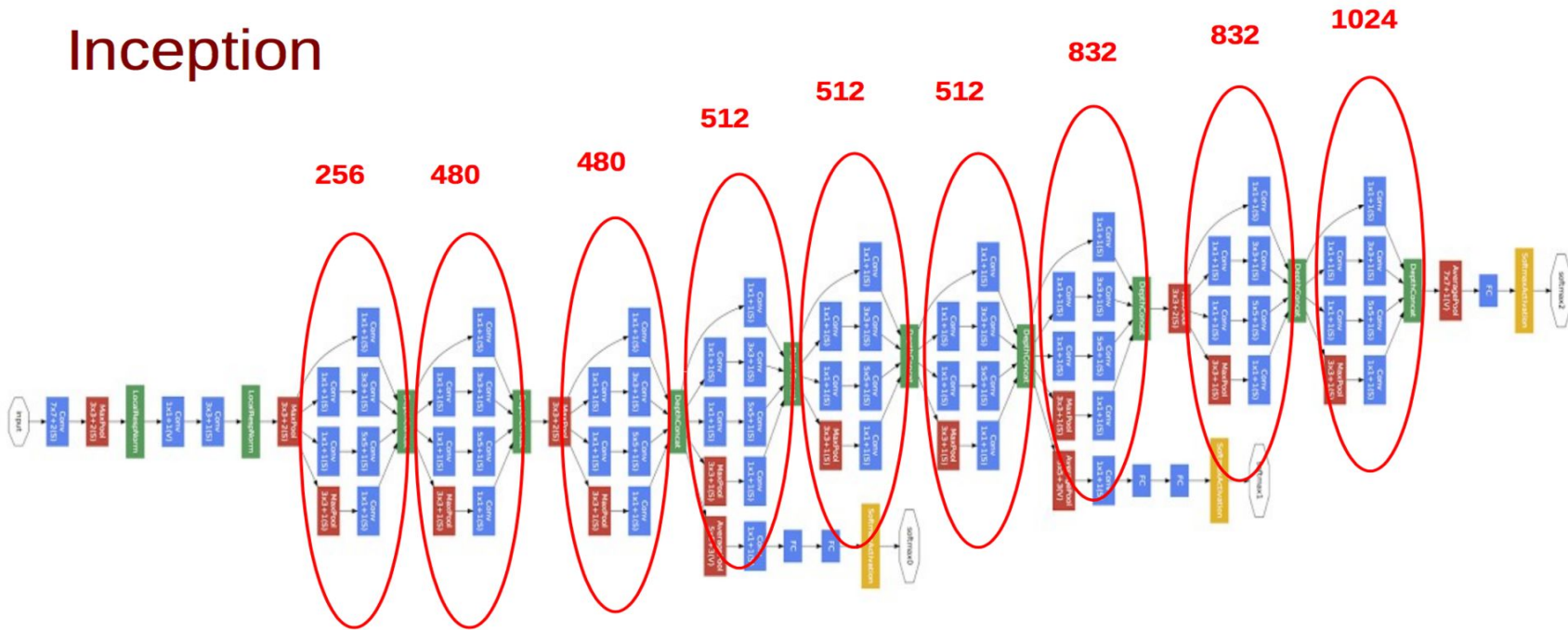


(b) Inception module with dimension reductions

Figure 2: Inception module

Inception

Inception

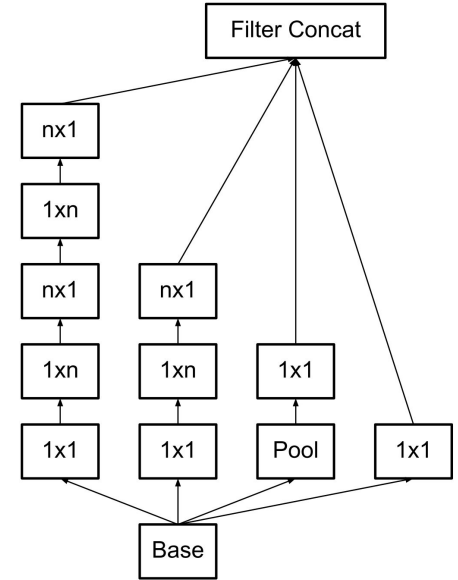
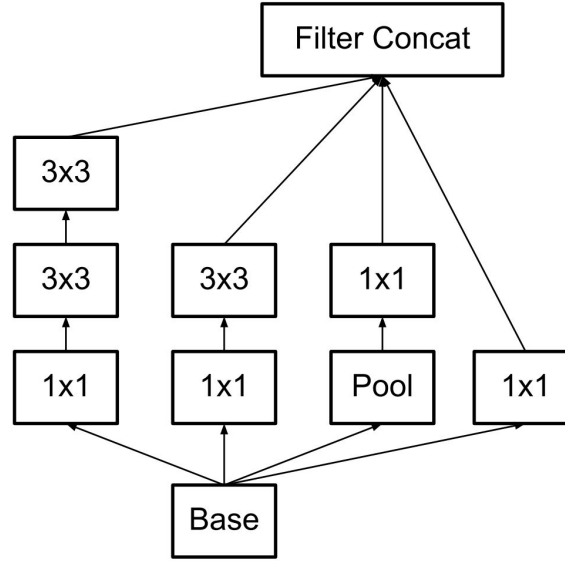
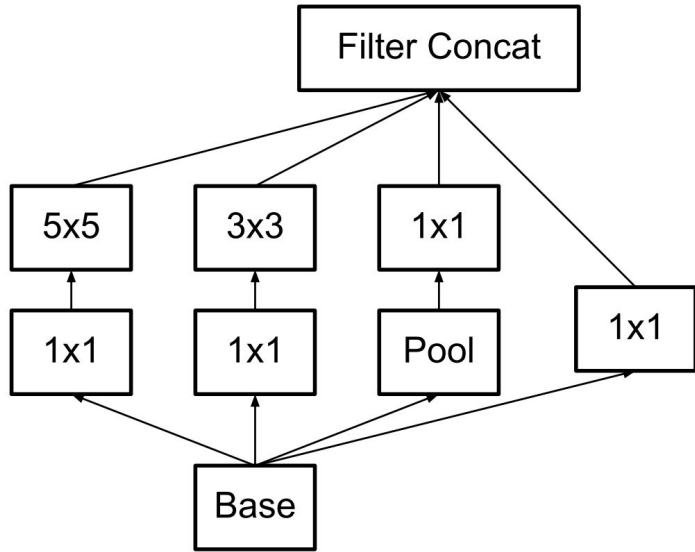


Width of **inception modules** ranges from 256 filters (in early modules) to 1024 in top inception modules.

Принципы построения глубоких нейросетей

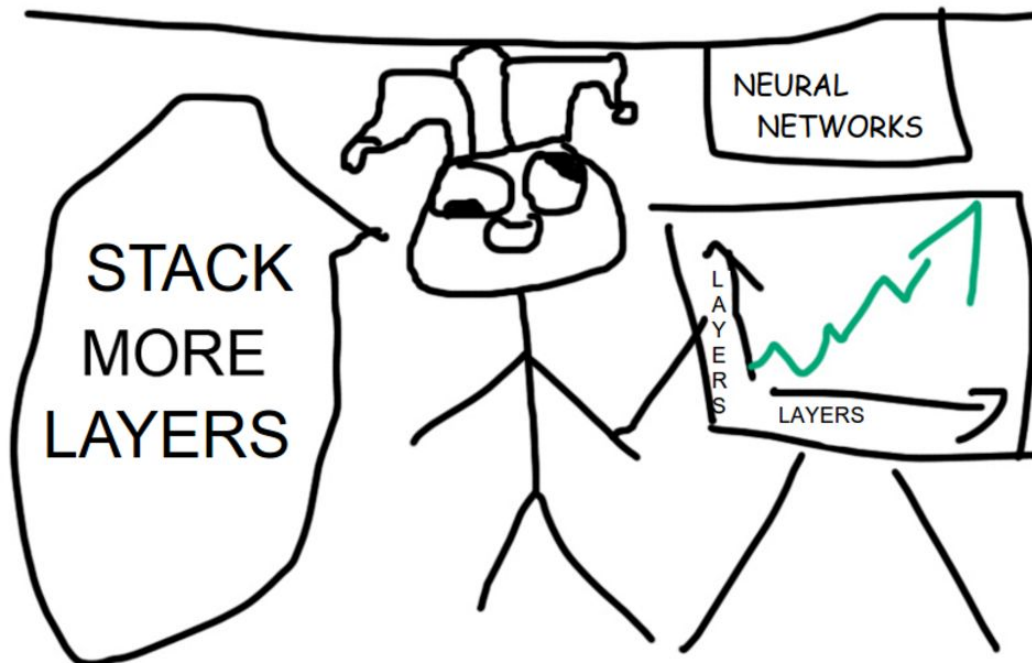
- Избегать representational bottlenecks: не стоит резко снижать размерность представления данных, это нужно делать плавно от начала сети и до классификатора на выходе.
- Высокоразмерные представления следует обрабатывать локально, увеличивая размерность: недостаточно плавно снижать размерность, стоит анализировать и группировать коррелированные участки.
- Пространственные сверки можно и нужно факторизовывать на еще более мелкие: это позволит сэкономить ресурсы и пустить их на увеличение размера сети.
- Необходимо соблюдать баланс между глубиной и шириной сети: не стоит резко увеличивать глубину сети отдельно от ширины, и наоборот; следует равномерно увеличивать или уменьшать обе размерности.

Inception v2



Что мешает просто увеличить сеть?

- Переобучение сетей с большим количеством параметров
- Затухание градиента в глубоких сетях
- Количество вычислений в сверточном слое пропорционально квадрату количества сверток



ResNet

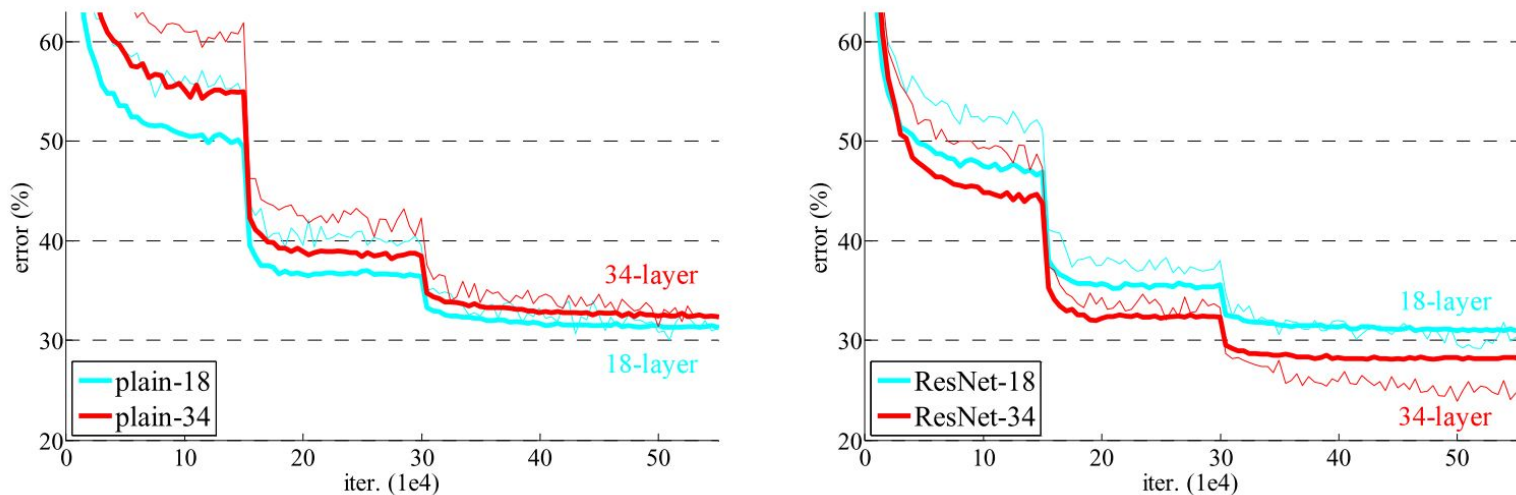
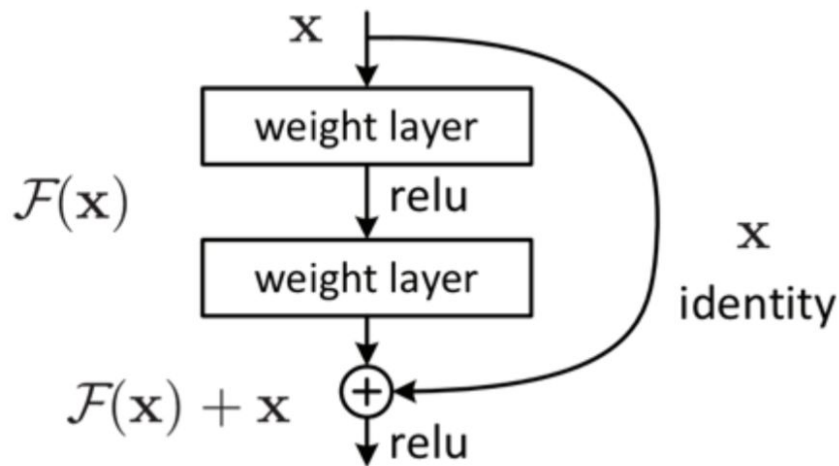


Figure 4. Training on **ImageNet**. Thin curves denote training error, and bold curves denote validation error of the center crops. Left: plain networks of 18 and 34 layers. Right: ResNets of 18 and 34 layers. In this plot, the residual networks have no extra parameter compared to their plain counterparts.

ResNet



$\mathcal{H}(\mathbf{x})$ is the true function we want to learn

Let's pretend we want to learn

$$\mathcal{F}(\mathbf{x}) := \mathcal{H}(\mathbf{x}) - \mathbf{x}$$

Figure 2. Residual learning: a building block.

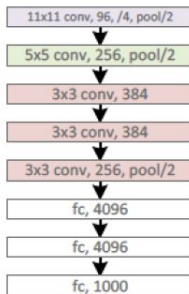
instead.

The original function is then

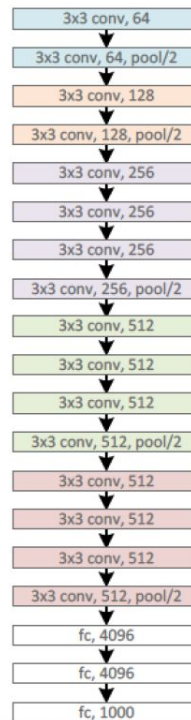
$$\mathcal{F}(\mathbf{x}) + \mathbf{x}$$

Сравнение архитектур

AlexNet, 8 layers
(ILSVRC 2012)



VGG, 19 layers
(ILSVRC 2014)



GoogleNet, 22 layers
(ILSVRC 2014)



Сравнение архитектур

AlexNet, 8 layers
(ILSVRC 2012)



VGG, 19 layers
(ILSVRC 2014)

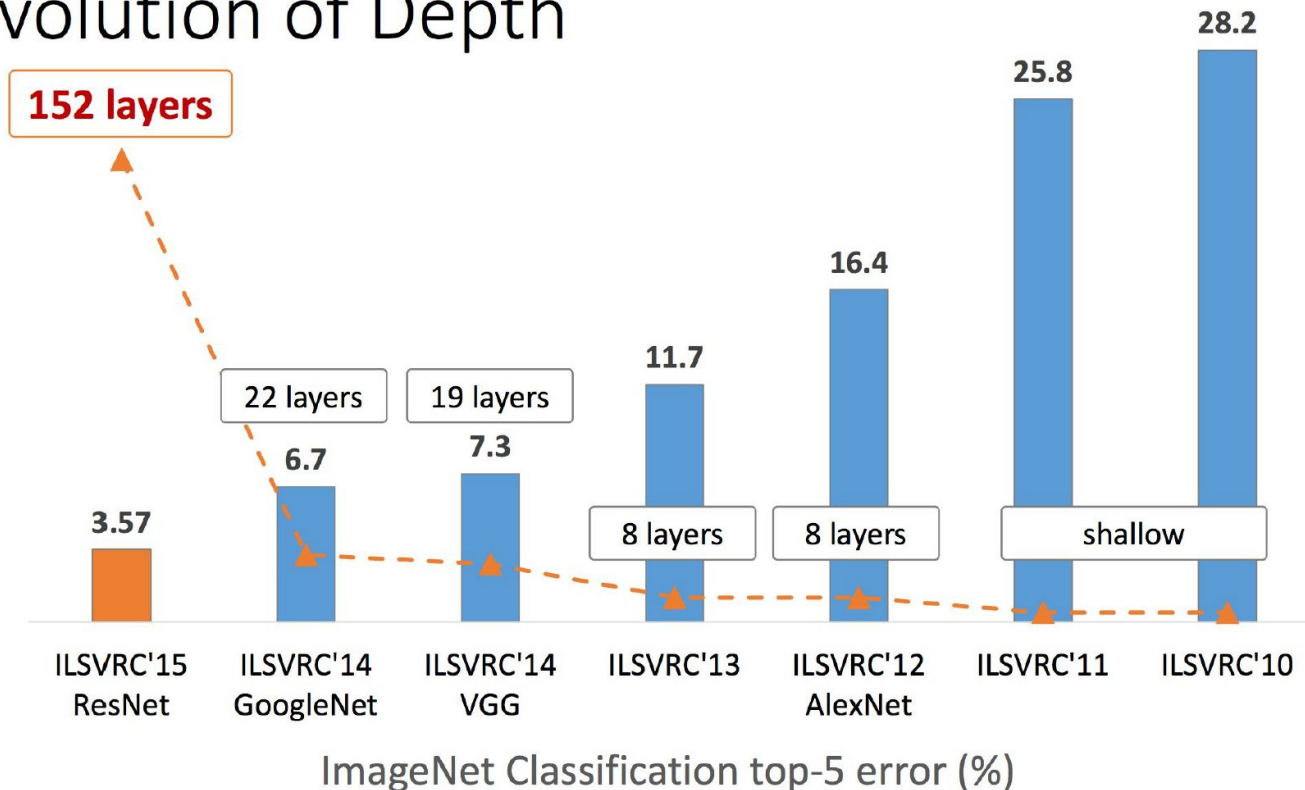


ResNet, 152 layers
(ILSVRC 2015)



Сравнение архитектур

Revolution of Depth



Inception v4

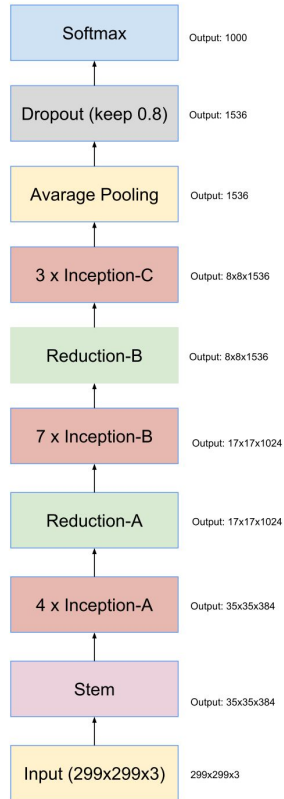


Figure 9. The overall schema of the Inception-v4 network. For the detailed modules, please refer to Figures 3, 4, 5, 6, 7 and 8 for the detailed structure of the various components.

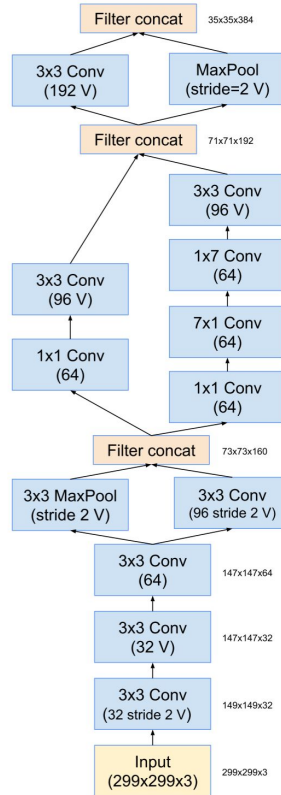


Figure 3. The schema for stem of the pure Inception-v4 and Inception-ResNet-v2 networks. This is the input part of those networks. Cf. Figures 9 and 15

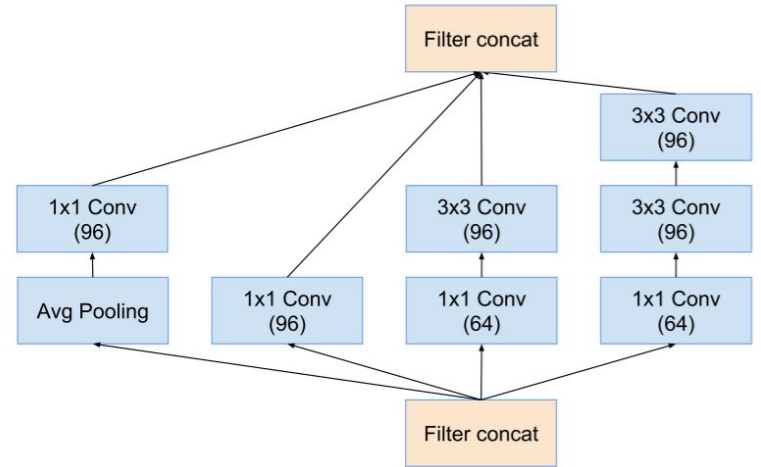
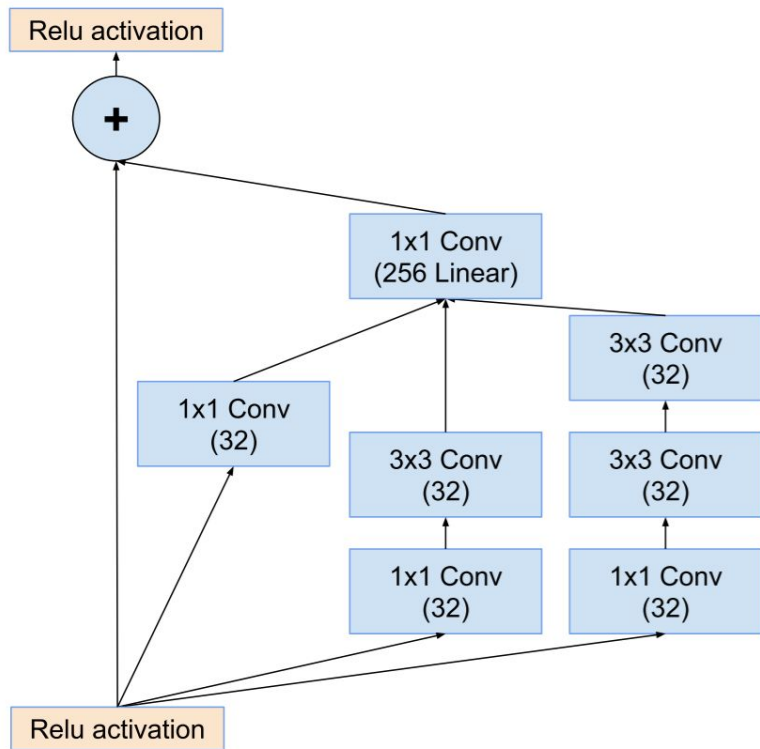


Figure 4. The schema for 35×35 grid modules of the pure Inception-v4 network. This is the Inception-A block of Figure 9.

- Рис.1 - общая схема сети
- Рис.2 - Stem-блок
- Рис.3 - Inception-A блок

Inception ResNet



Inception

ResNet



Figure 10. The schema for 35×35 grid (Inception-ResNet-A) module of Inception-ResNet-v1 network.

Спасибо за внимание!

