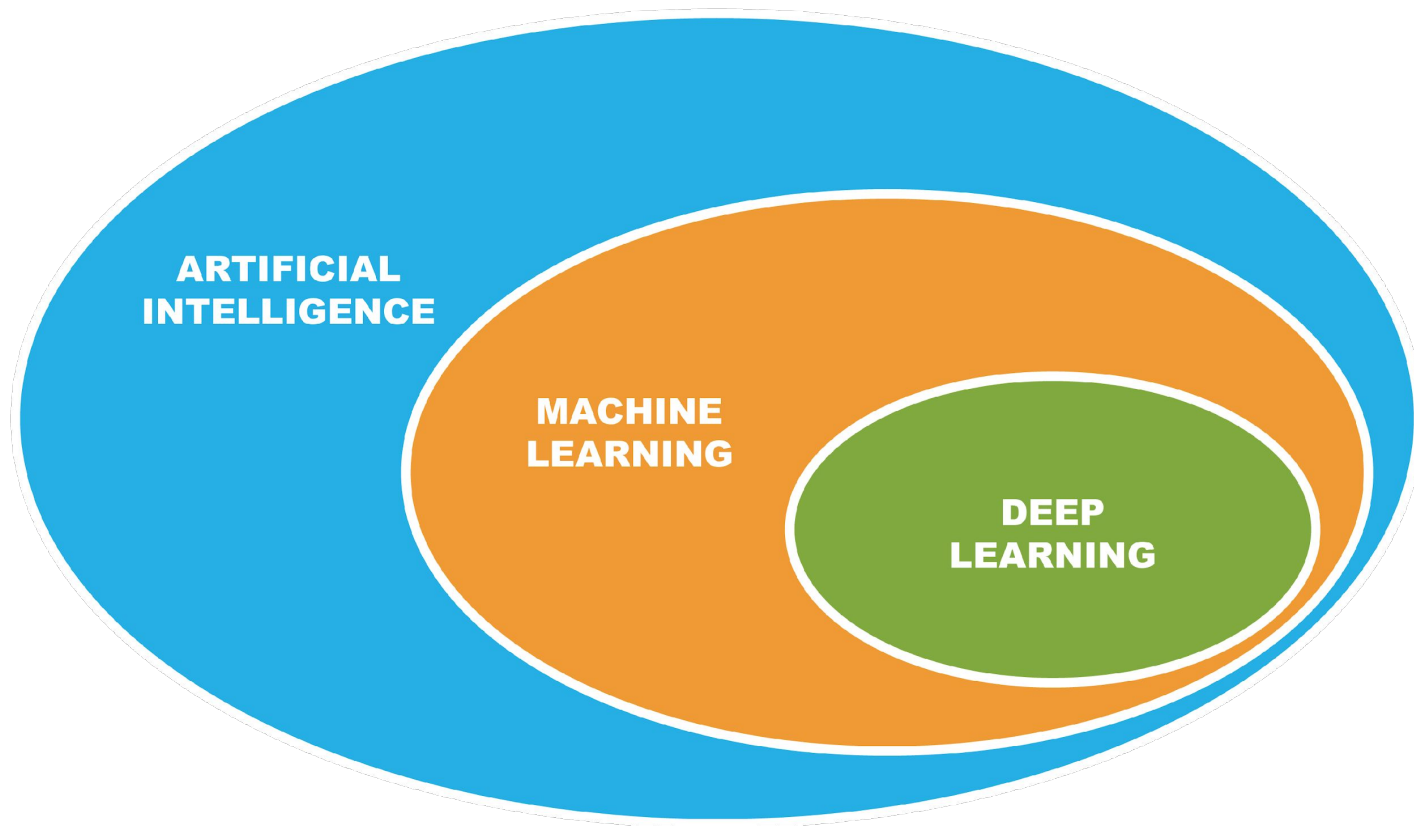


Deep learning CV

Свёрточные нейронные сети

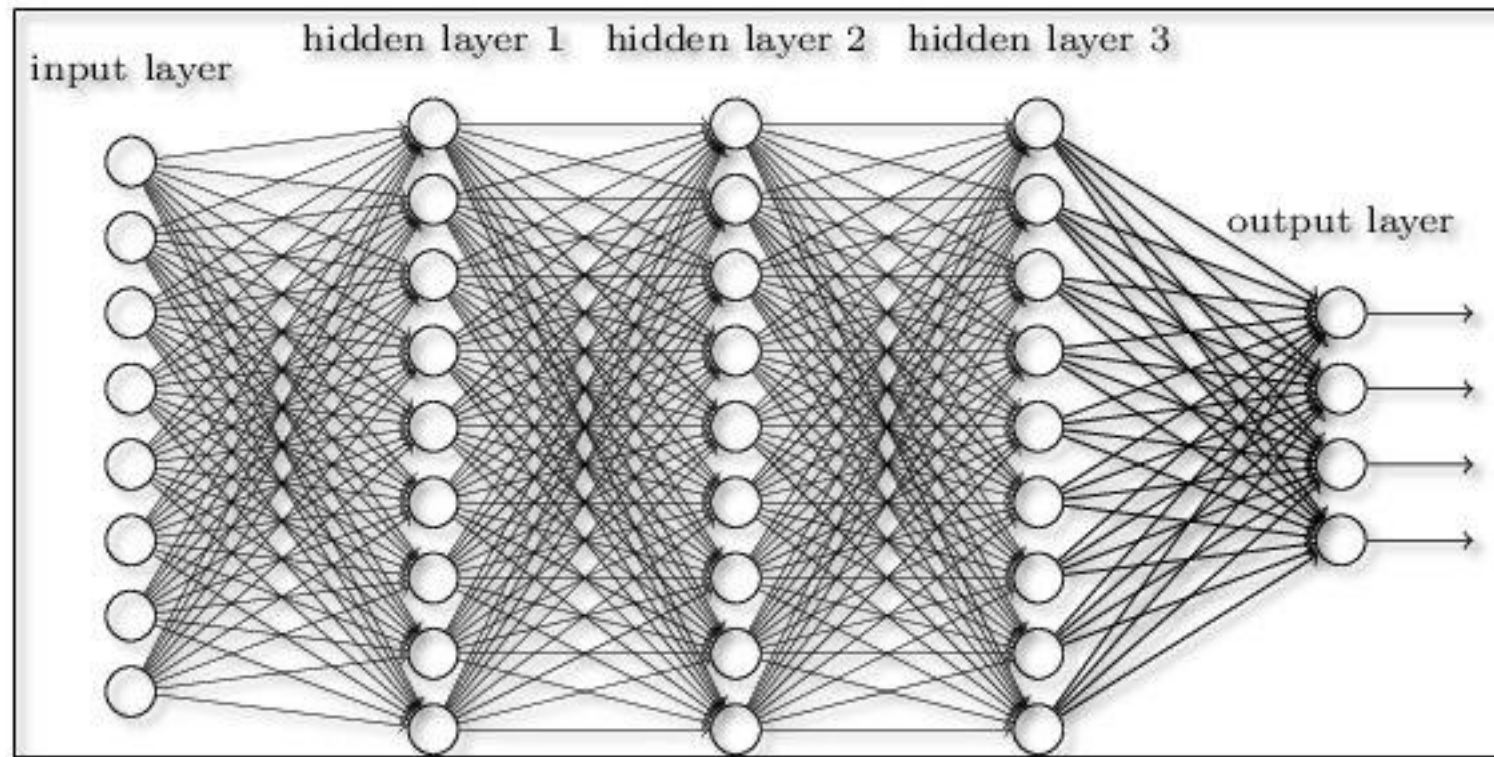
Что такое deep learning (глубокое обучение)?



- Только нейронные сети
- Много слоёв
- Долгое обучение (дни, недели)
- Специальные фреймворки
- Не нужно подготавливать признаки самим

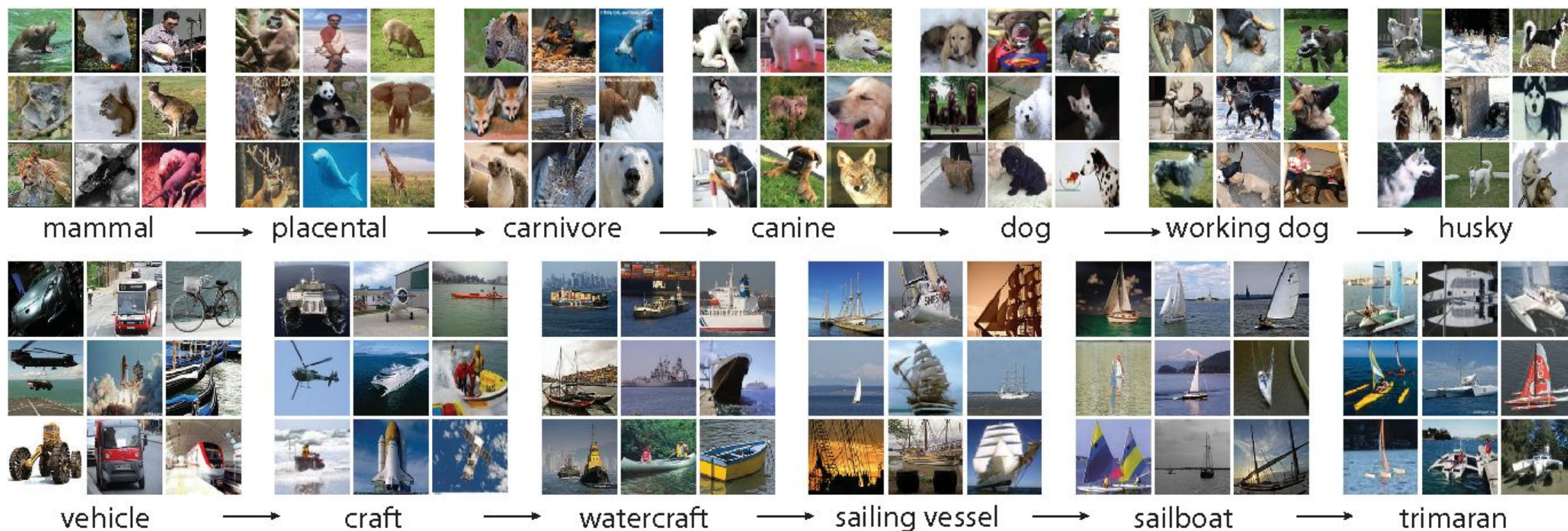
Нейронные сети – общий подход

$$a^m(x, w) = \sigma_m \left(\sum_{h=0}^H w_{hm} \sigma_h \left(\sum_{j=0}^n w_{jh} f_j(x_i) \right) \right)$$

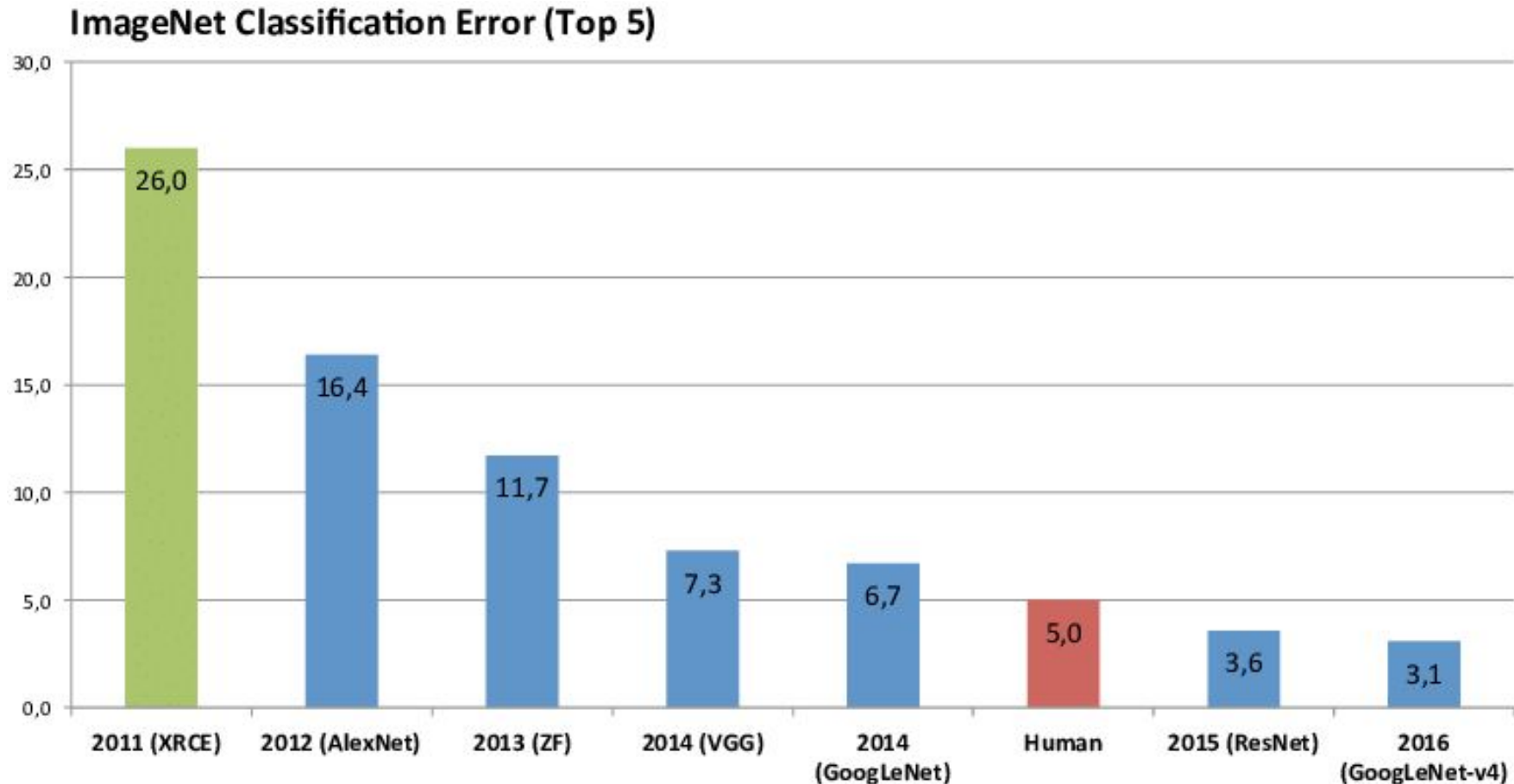


ImageNet датасет

- 14 197 122 картинок
- 1000 классов



Результаты соревнования ImageNet



Почему только топ5?

Какой класс у этой
картинки?



Почему результат всё ещё не идеальный?

Какой класс у этой картинки?

horse

women's clothing



Почему результат всё ещё не идеальный?

Какой класс у этой картинки?

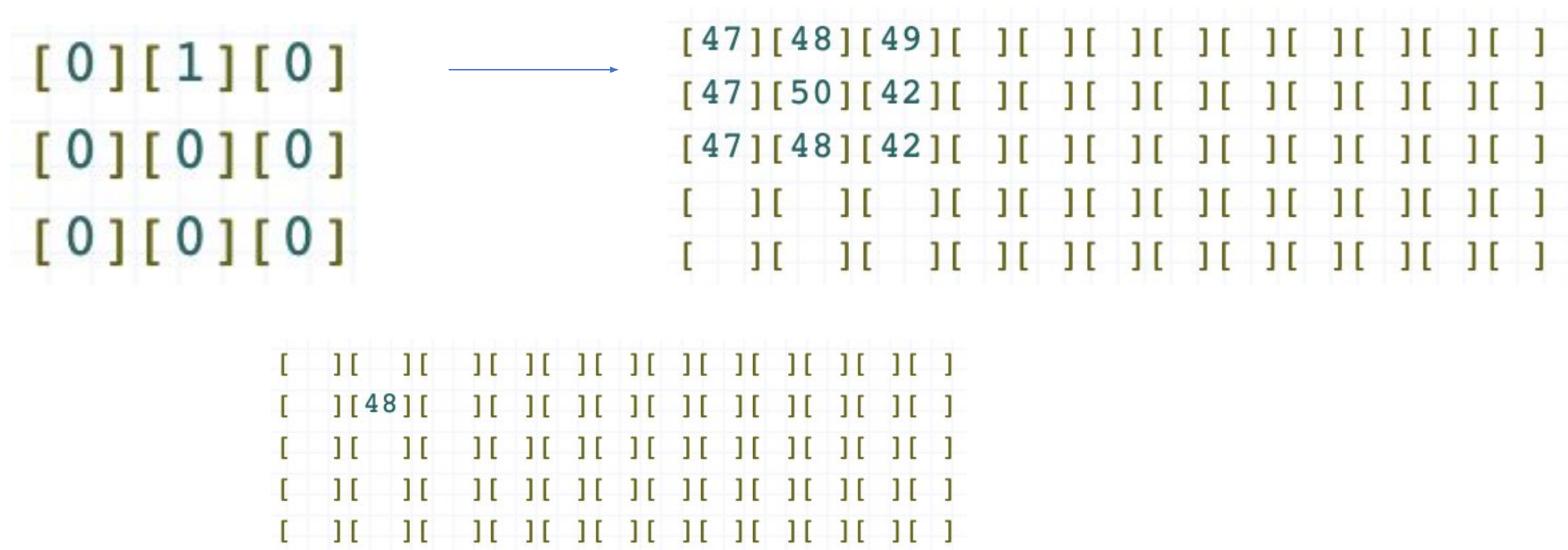
horse

women's clothing

hay (сено)



Свёрточный слой



Применение свёртки

$$\begin{pmatrix} -1 & 0 & 1 \\ -1 & 0 & 1 \\ -1 & 0 & 1 \end{pmatrix}$$



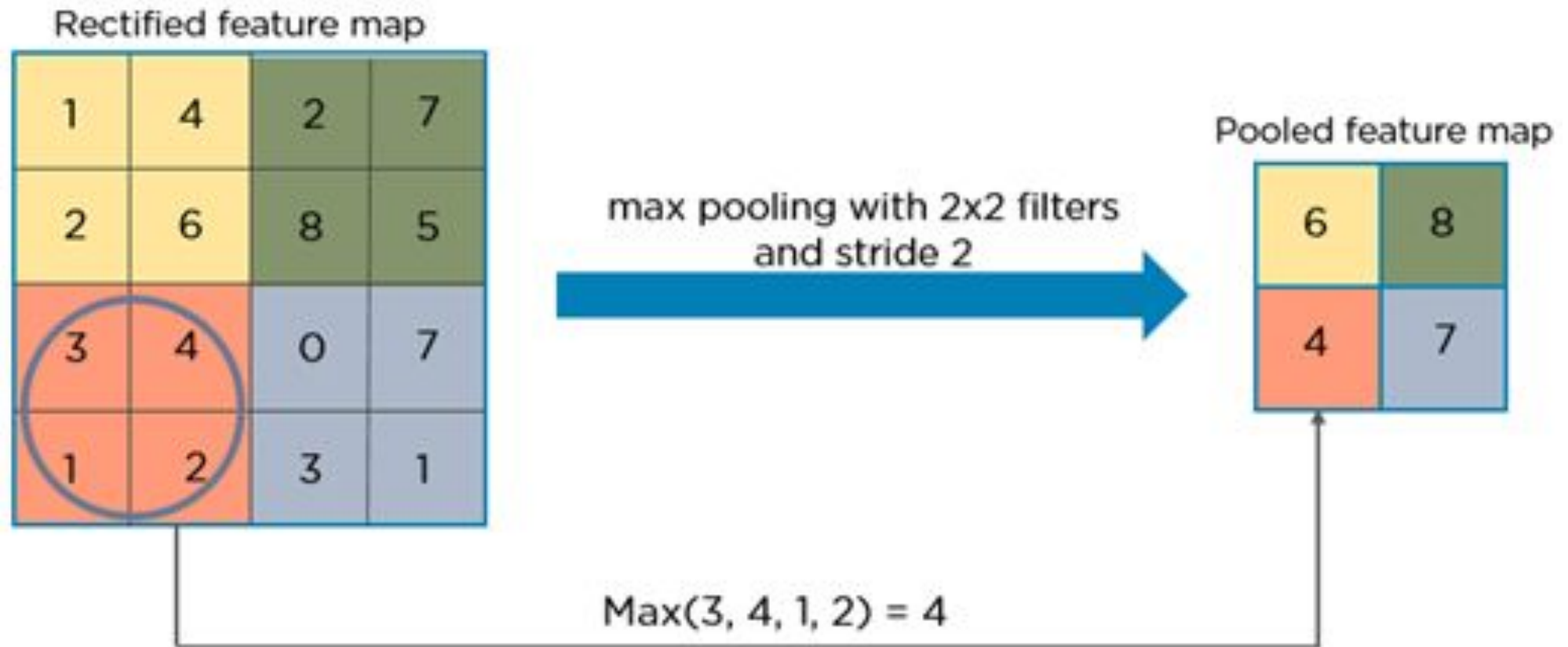
Или так

$$\begin{pmatrix} 0.0625 & 0.125 & 0.0625 \\ 0.125 & 0.25 & 0.125 \\ 0.0625 & 0.125 & 0.0625 \end{pmatrix}$$

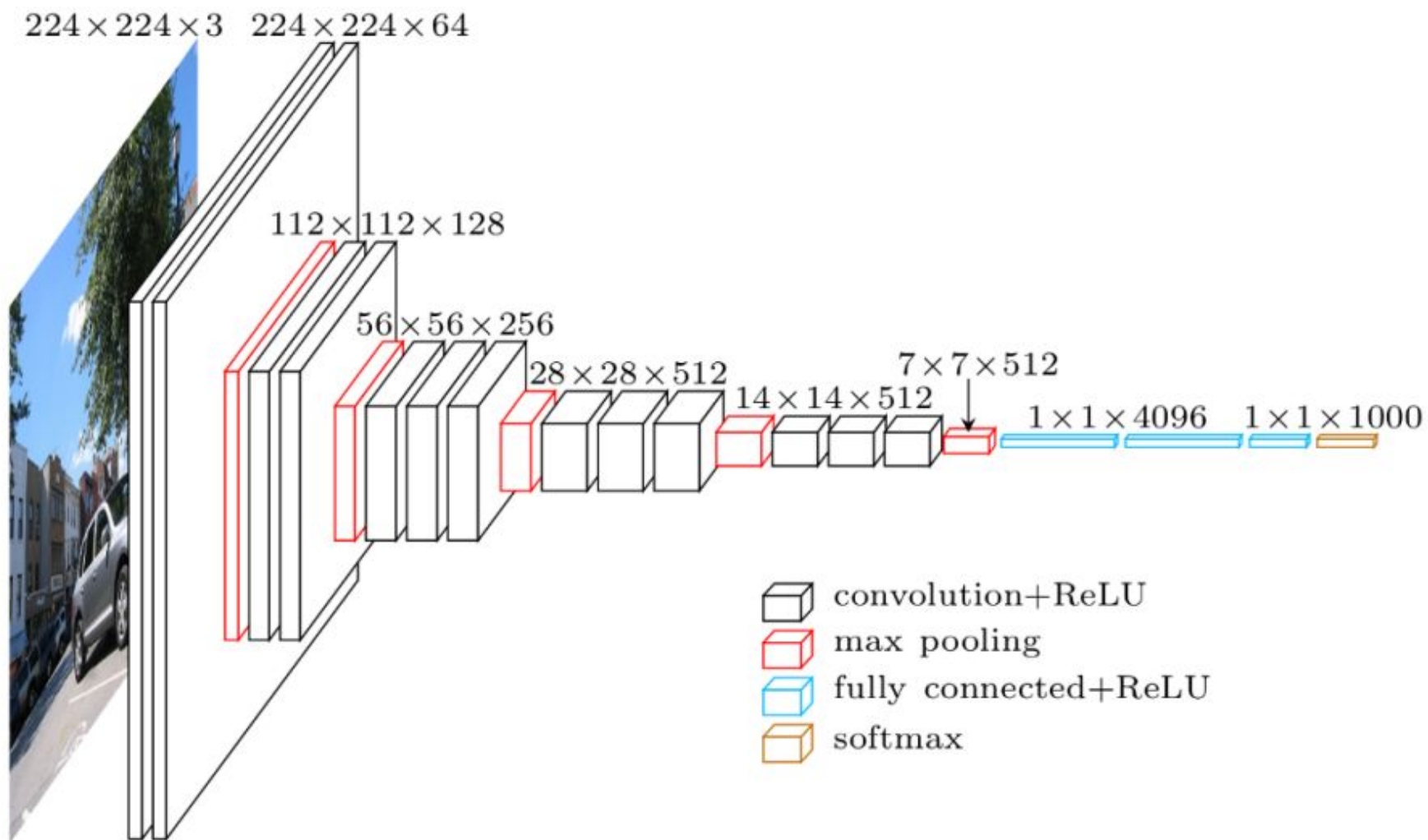


Pooling

- Существуют max-pooling и avg-pooling

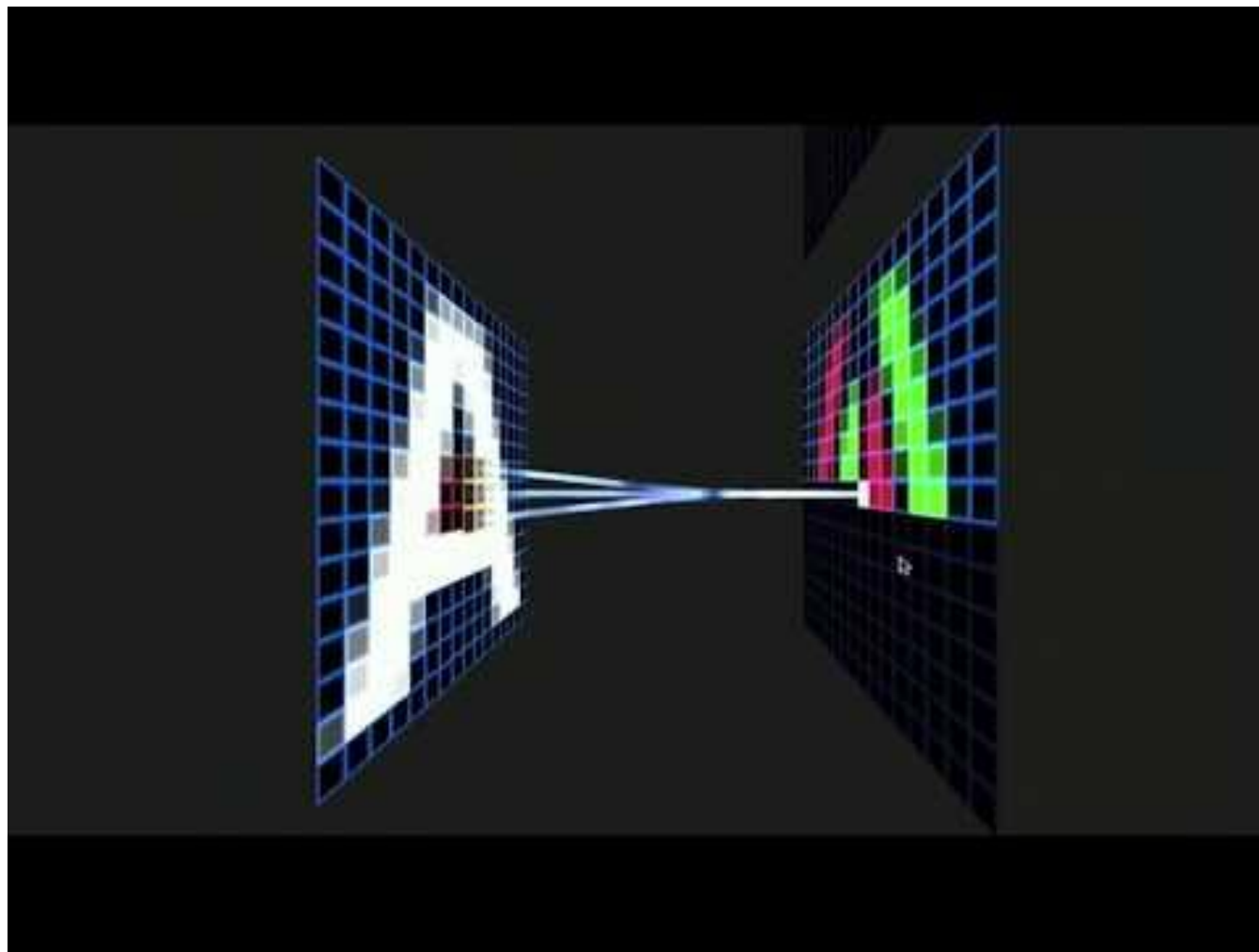


Что мы уже знаем отсюда?

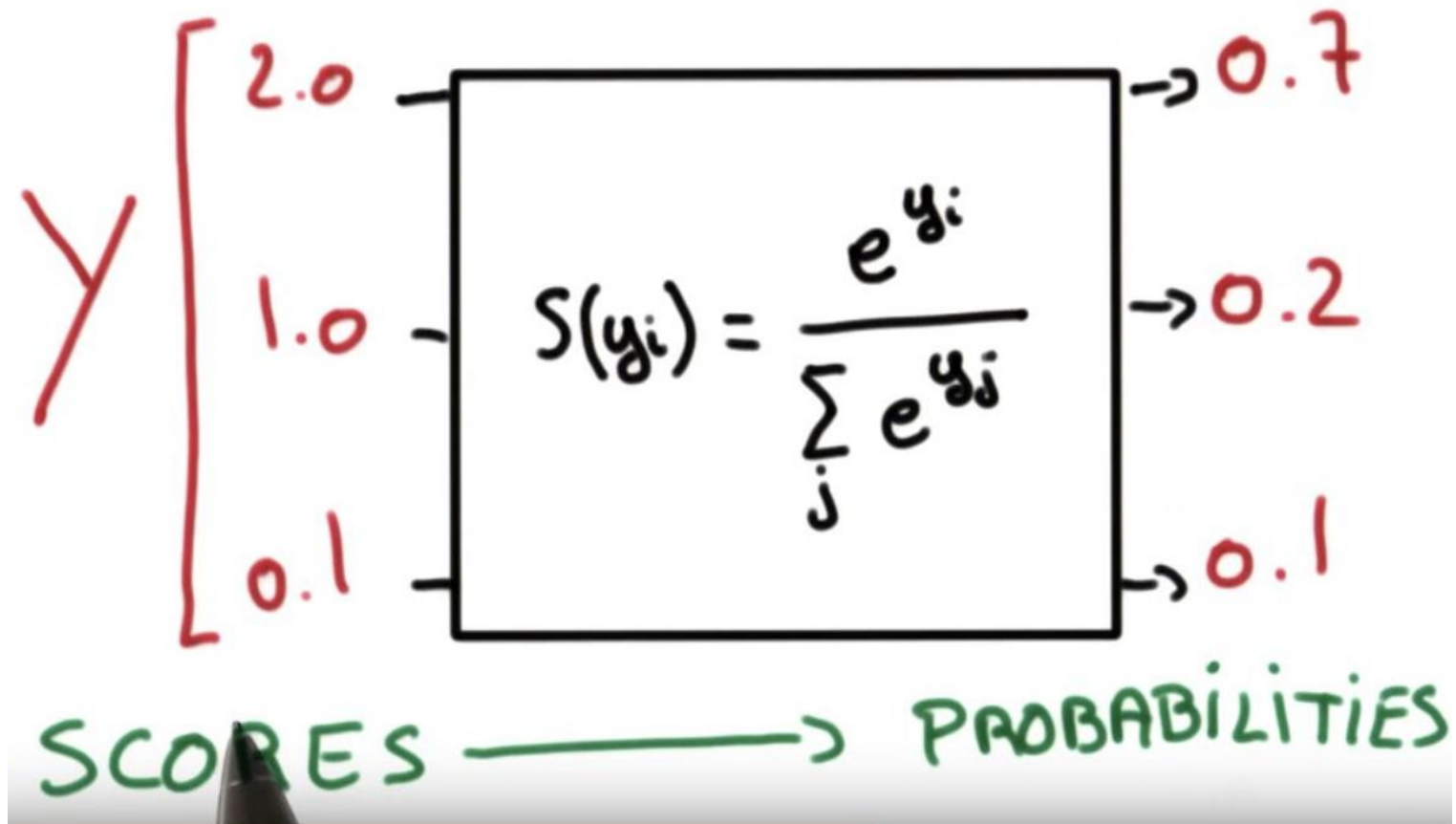


Визуализация работы CNN от Otavio Good

Ссылка на видео: <https://youtu.be/f0t-OCG79-U>

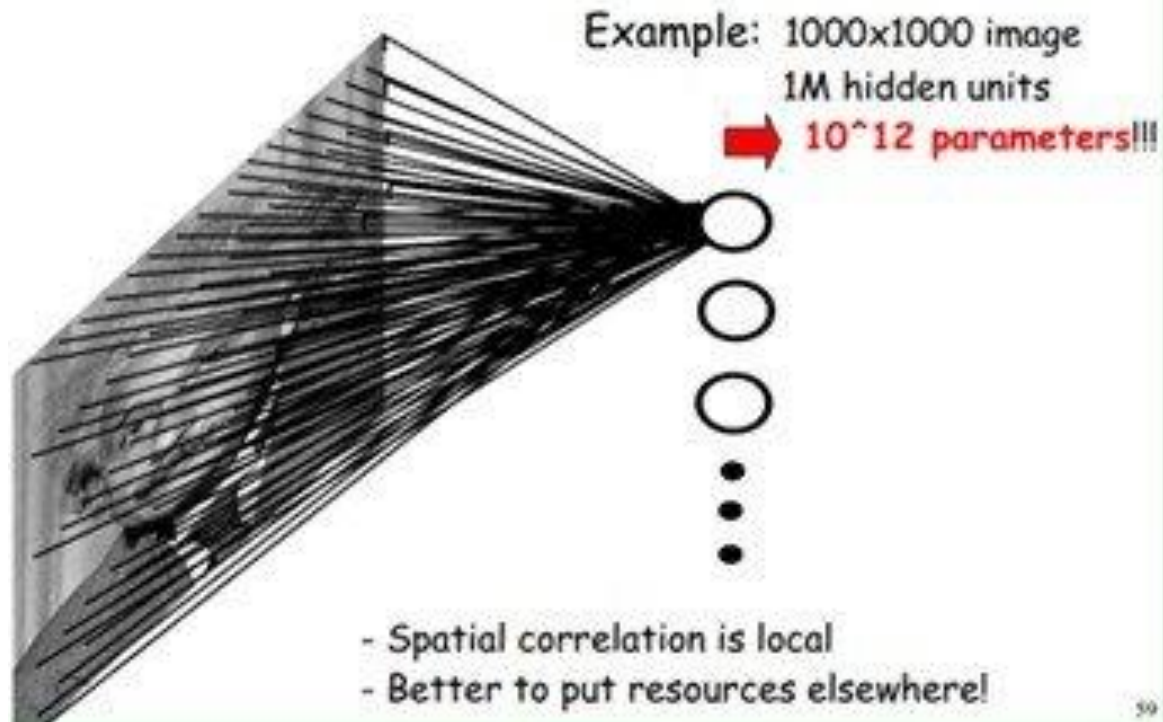


Softmax



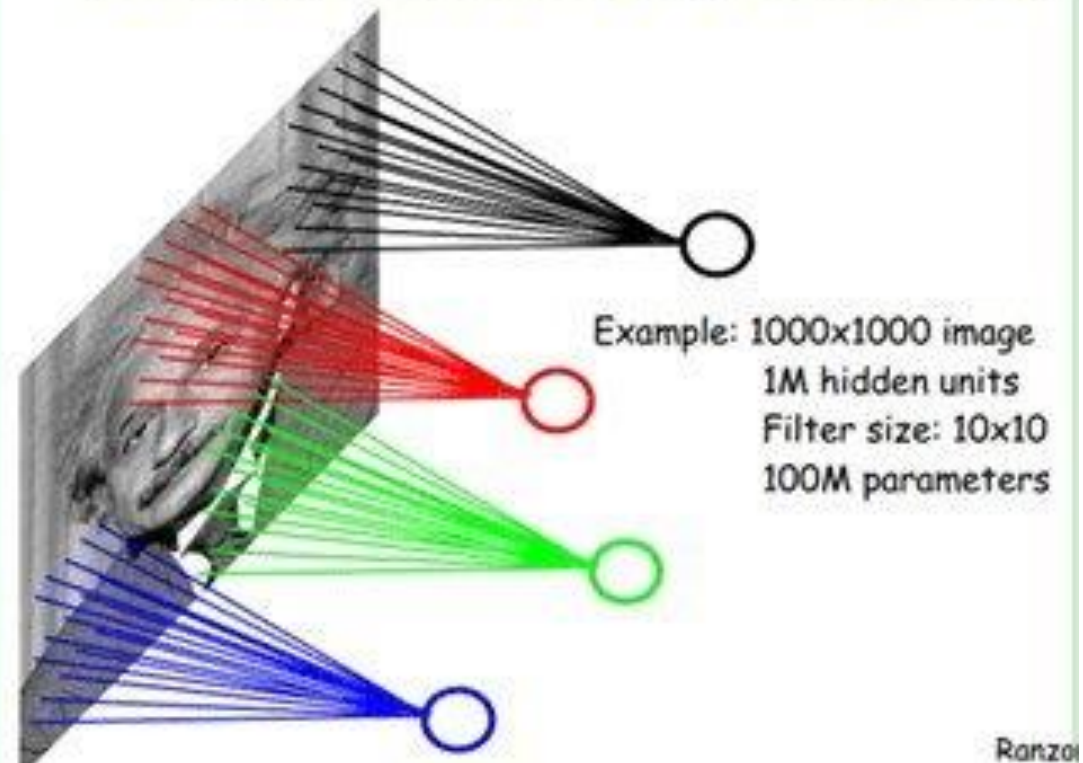
Почему свертки лучше?

FULLY CONNECTED NEURAL NET



59

LOCALLY CONNECTED NEURAL NET



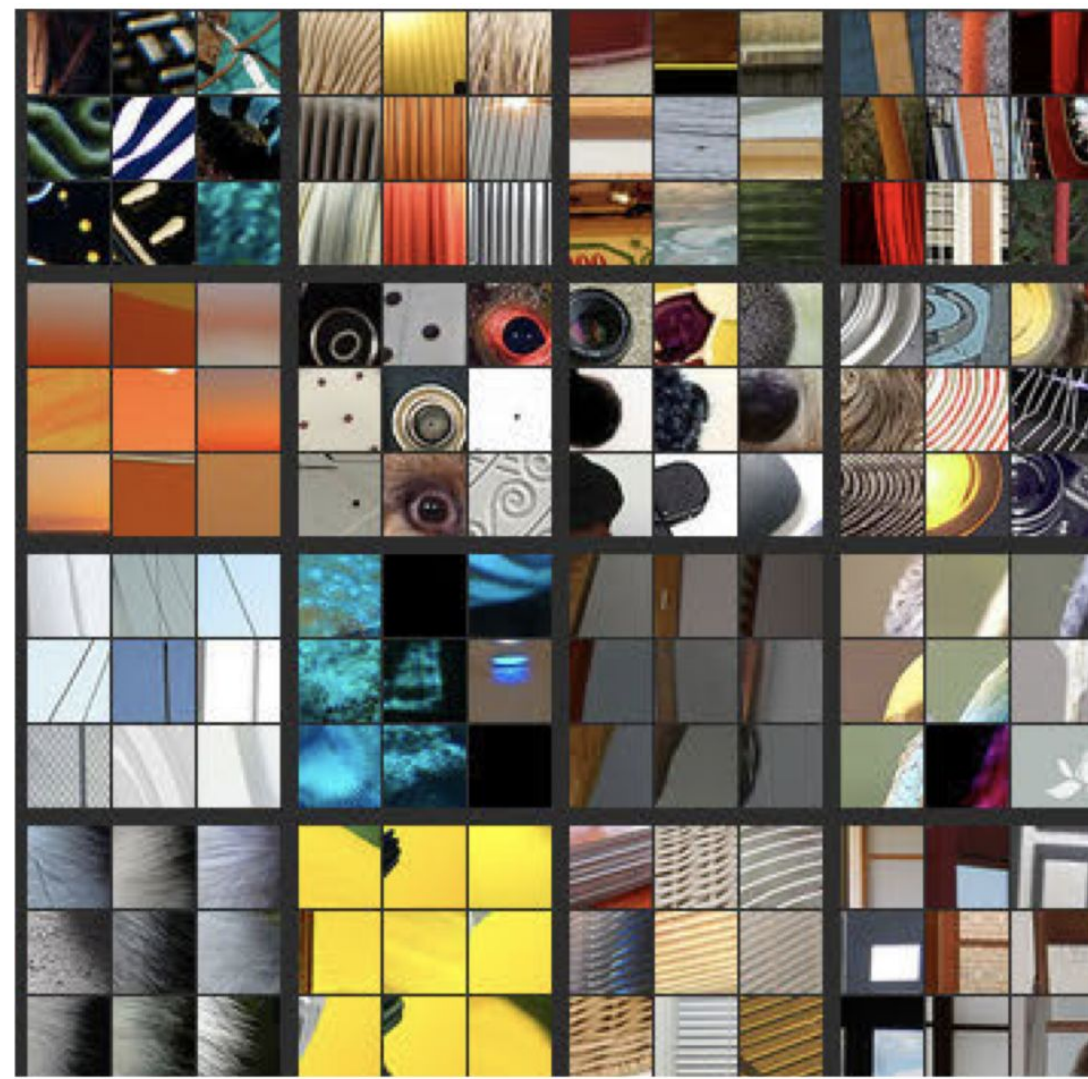
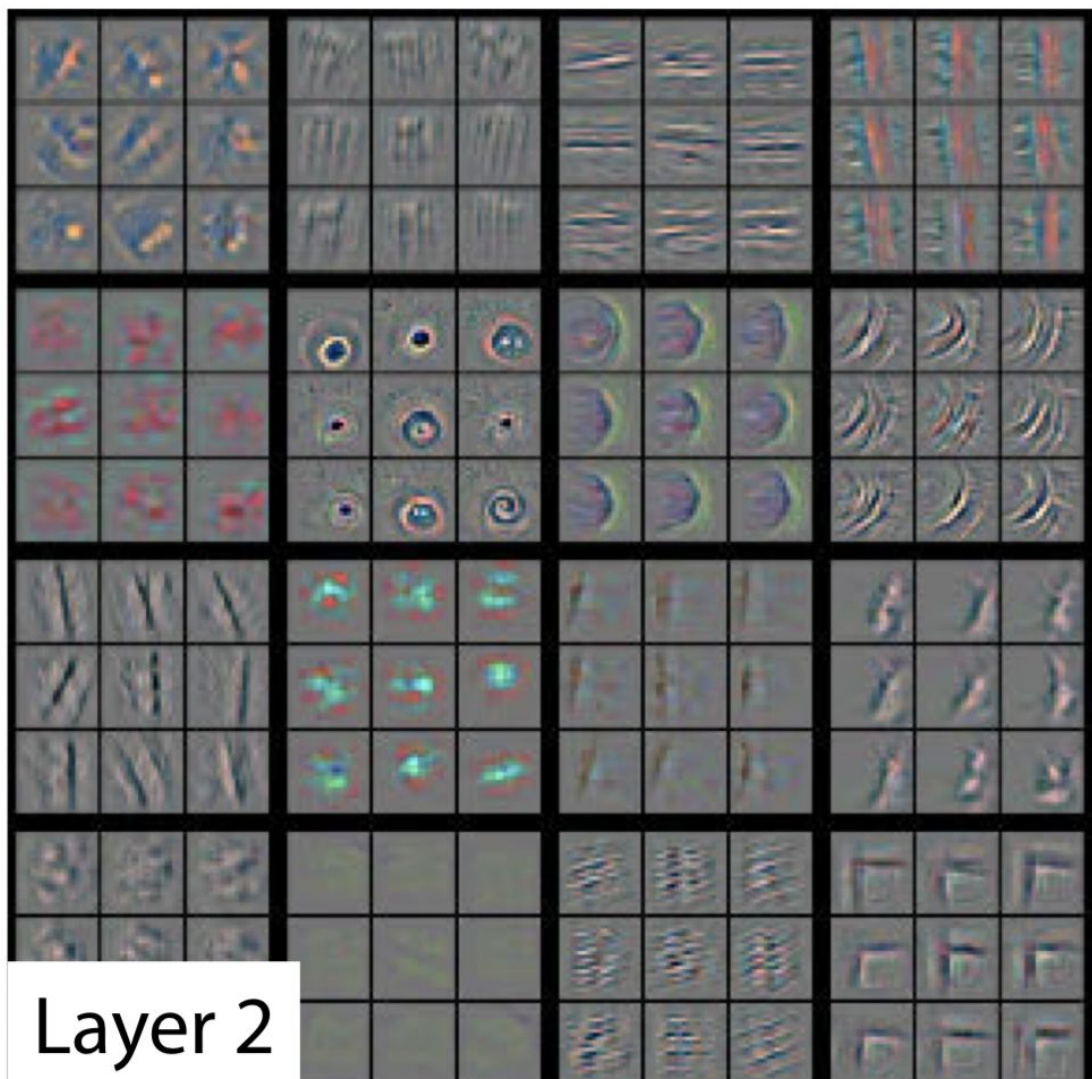
Визуализация первого слоя



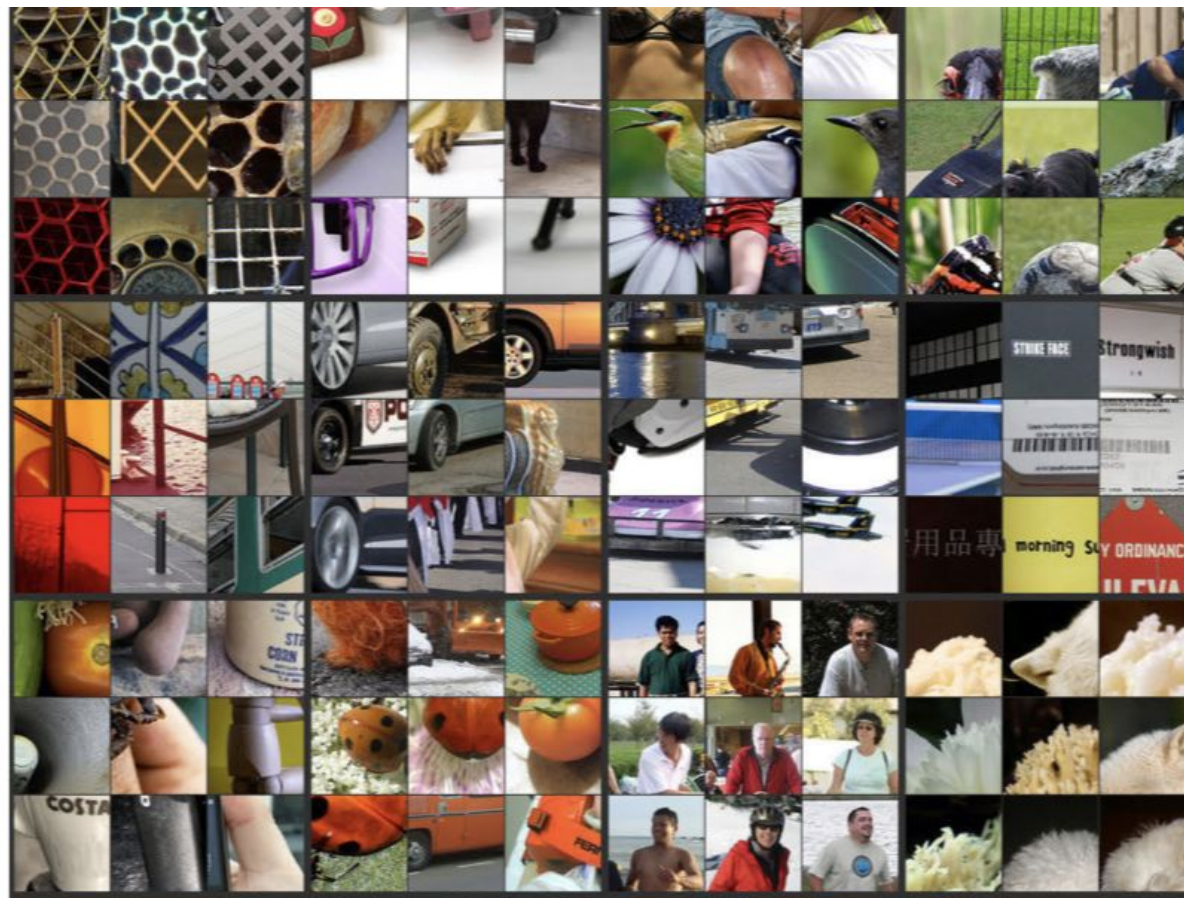
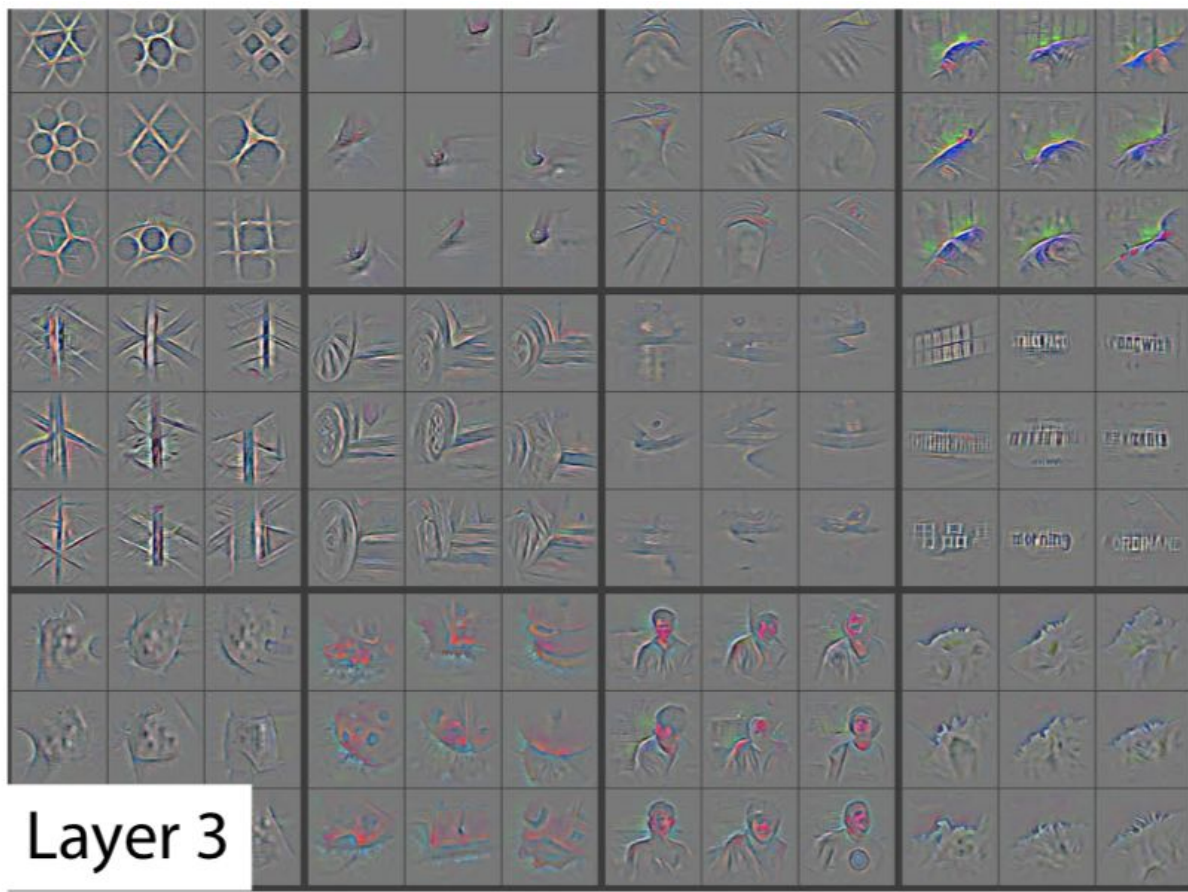
Layer 1



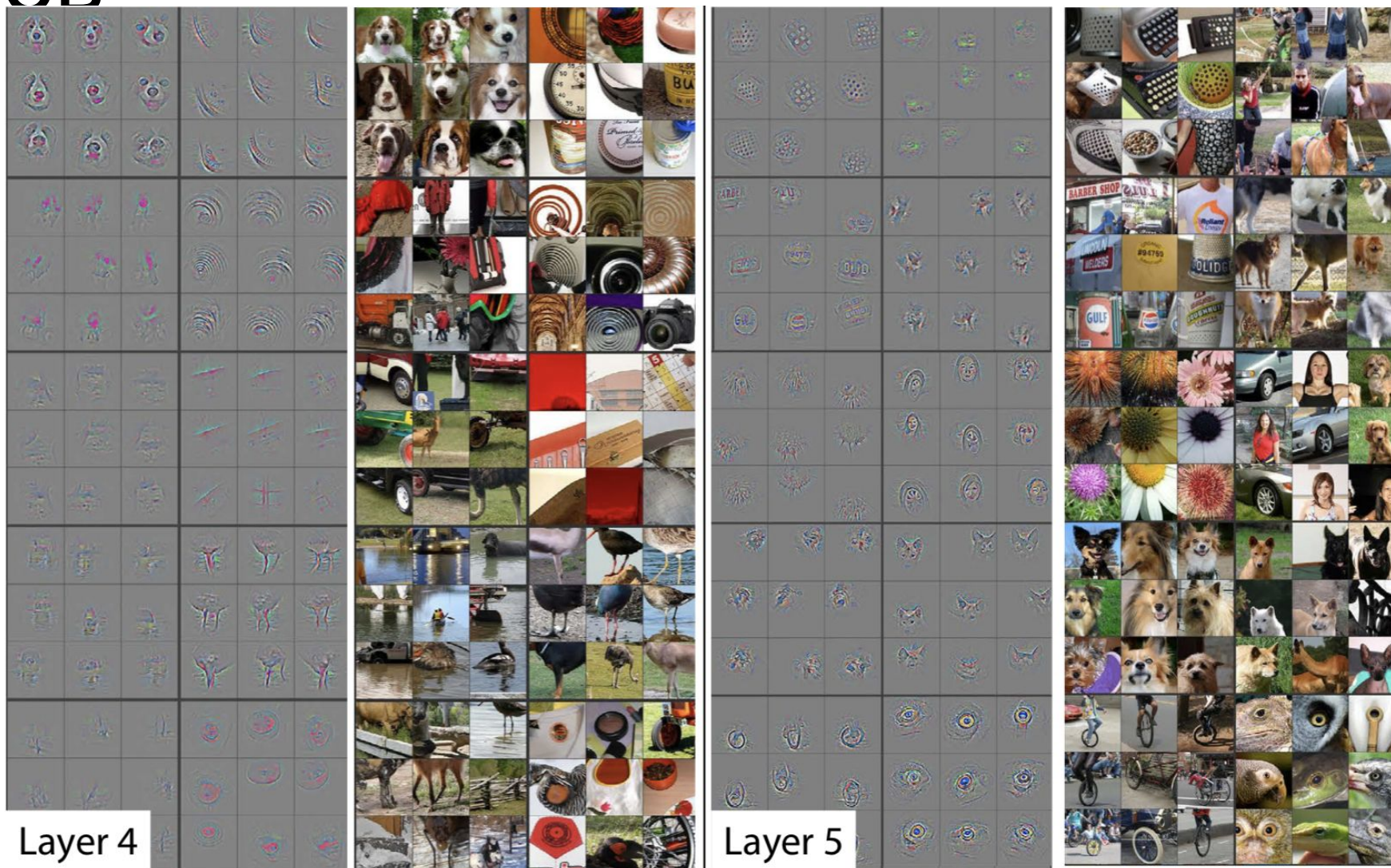
Визуализация второго слоя



Визуализация третьего слоя



Визуализация четвертого и пятого слоев



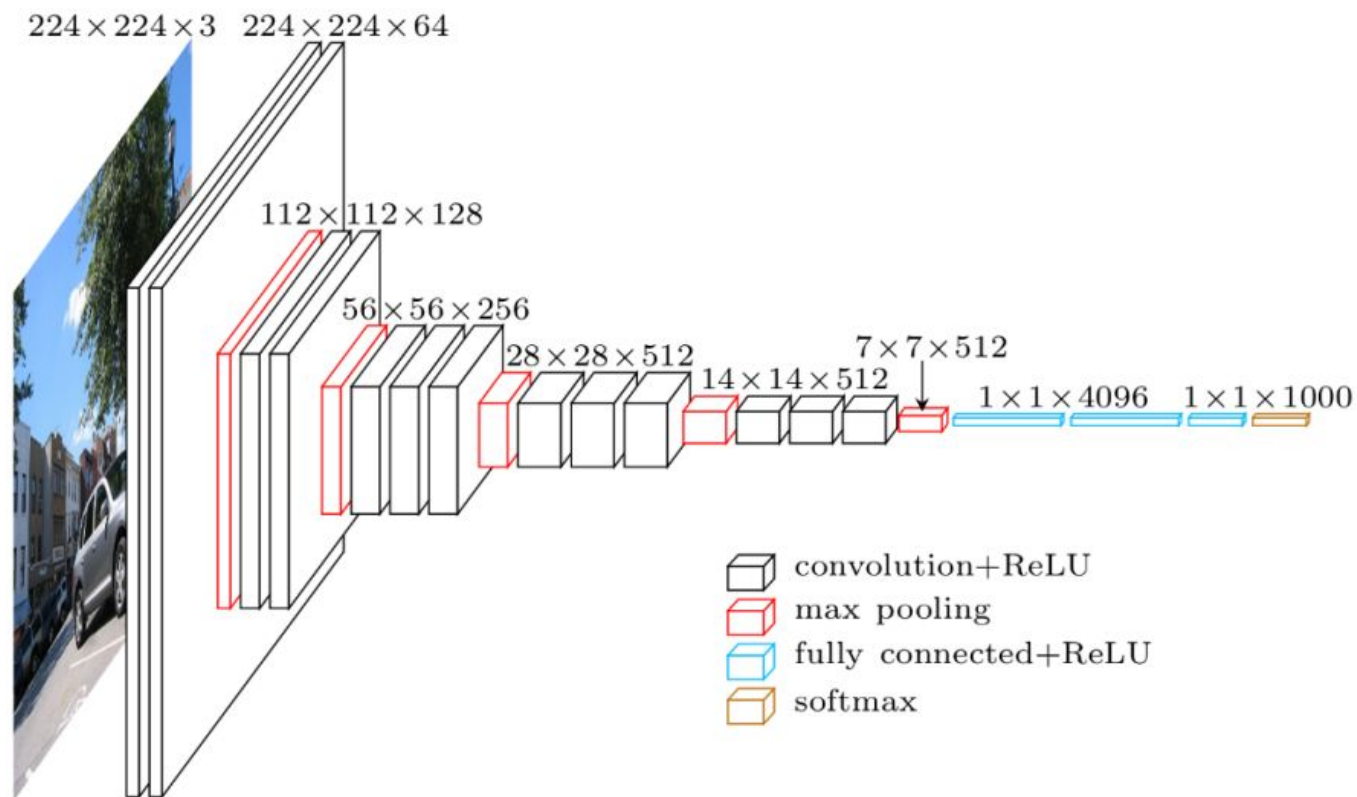
Развитие свёрточных нейронных сетей

С 2012 по 2015

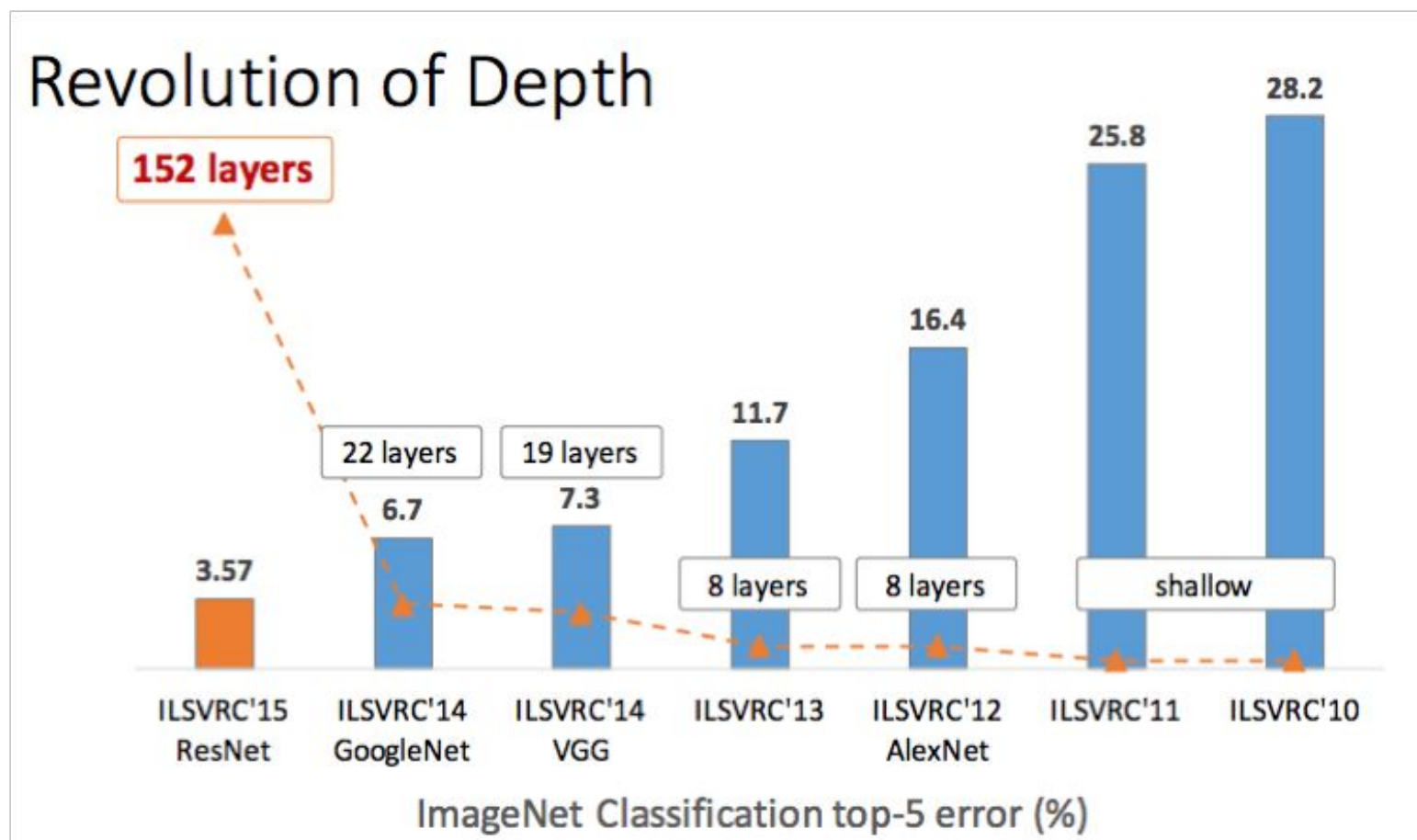
- Lenet-5 (Le Kun, 1998)
- AlexNet (2012)
- VGG19 (2014)
- GoogLeNet Inception v1 (2014)
- ResNet (2015)

VGG19

- Победил предыдущую сеть на ImageNet 7.3% vs 11.2%
- 3x3 CONV сдвиг 1, паддинг 1
- 2x2 MAX POOL сдвиг 2
- 138M параметров



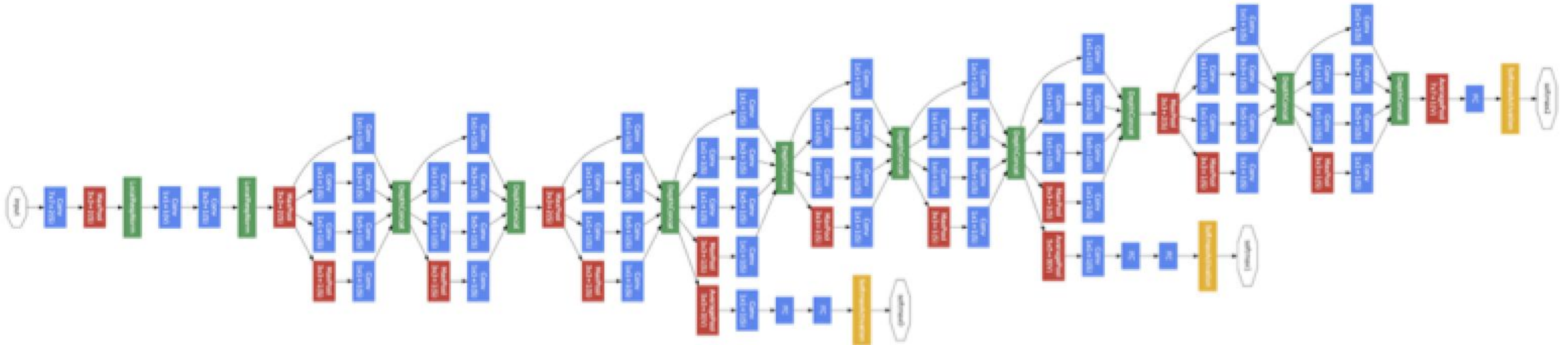
Как победить? Нужно больше слоёв



Inception (GoogLeNet)

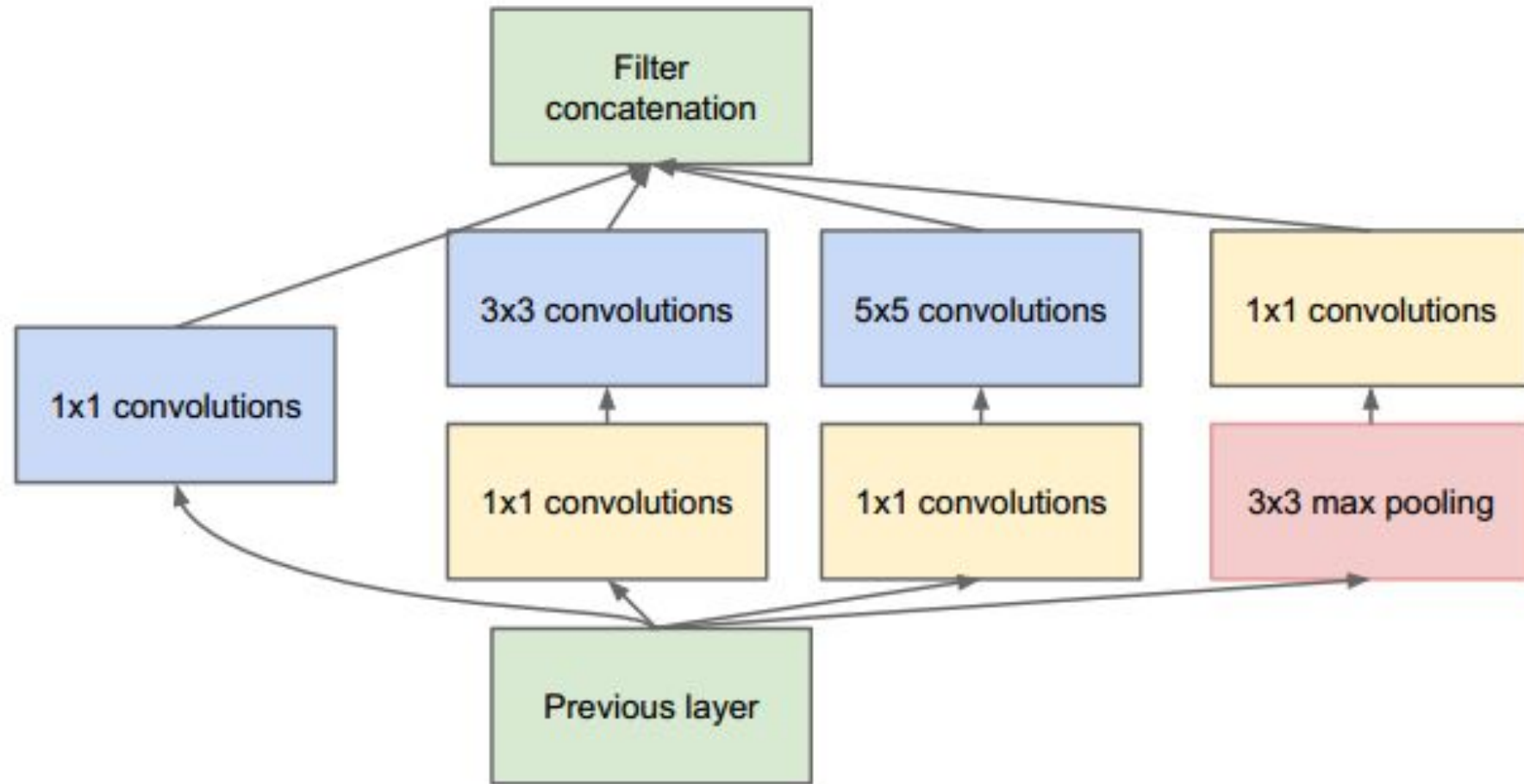


Inception (GoogLeNet)



Convolution
Pooling
Softmax
Other

Inception (GoogLeNet)



Inception: результаты

- Победитель 2014
- 6.7% против 7.3% у VGG19
- Нет полносвязных слоёв
- Всего лишь 5M параметров (у VGG - 128M)

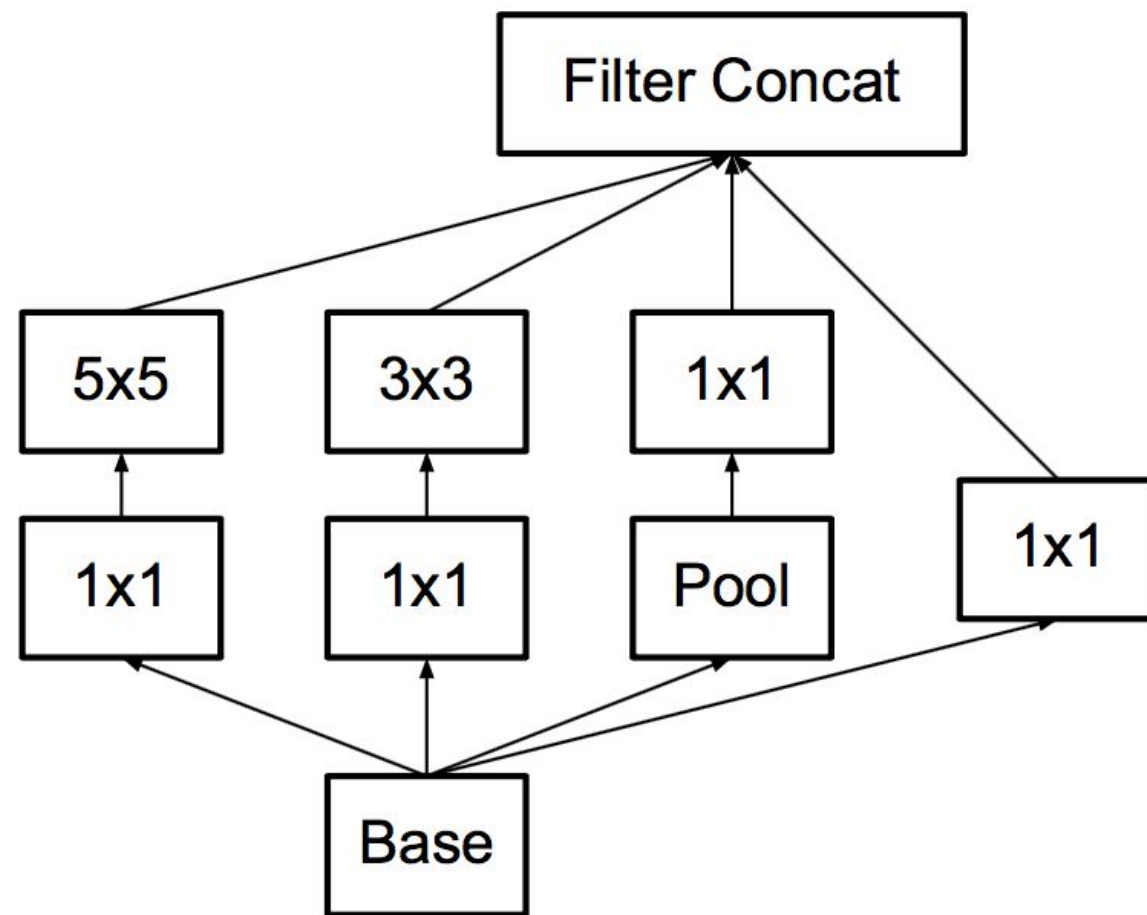
Inception v3

Основные принципы:

- 1) Очень много похожих сигналов расположены близко к друг другу в пространстве => можно уменьшать размерность перед применением свёрток
- 2) Чтобы использовать ресурсы эффективно нужно увеличивать и глубину и ширину сети, делать только что-то одно неэффективно
- 3) Нельзя иметь слои, которые слишком резко уменьшают число параметров
- 4) “Широкие” слои обучаются быстрее

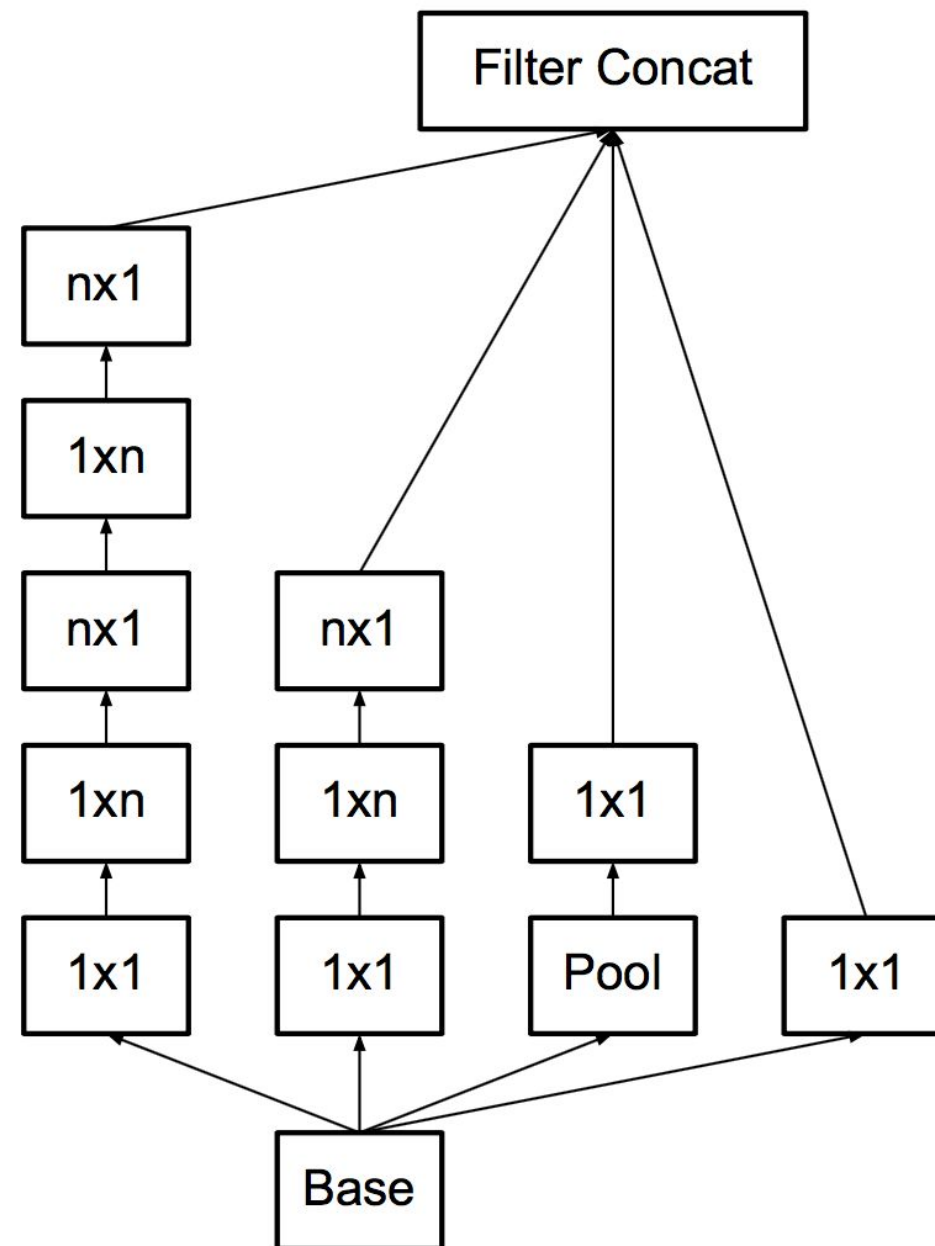
Inception v3

- Заменим свёртку 5x5 на две последовательные 3x3
- Заменим свёртку 3x3 на последовательные 1x3 и 3x1
- Раз теперь свёртки делать так “дёшево” давайте делать большие свёртки вроде 7x7

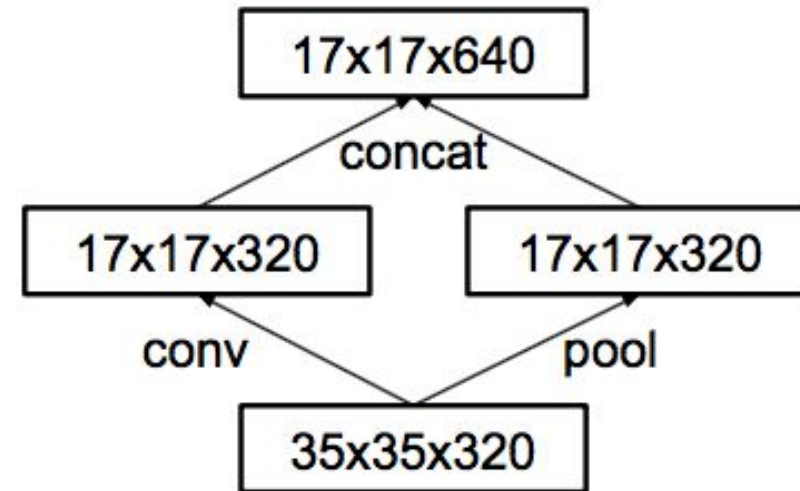
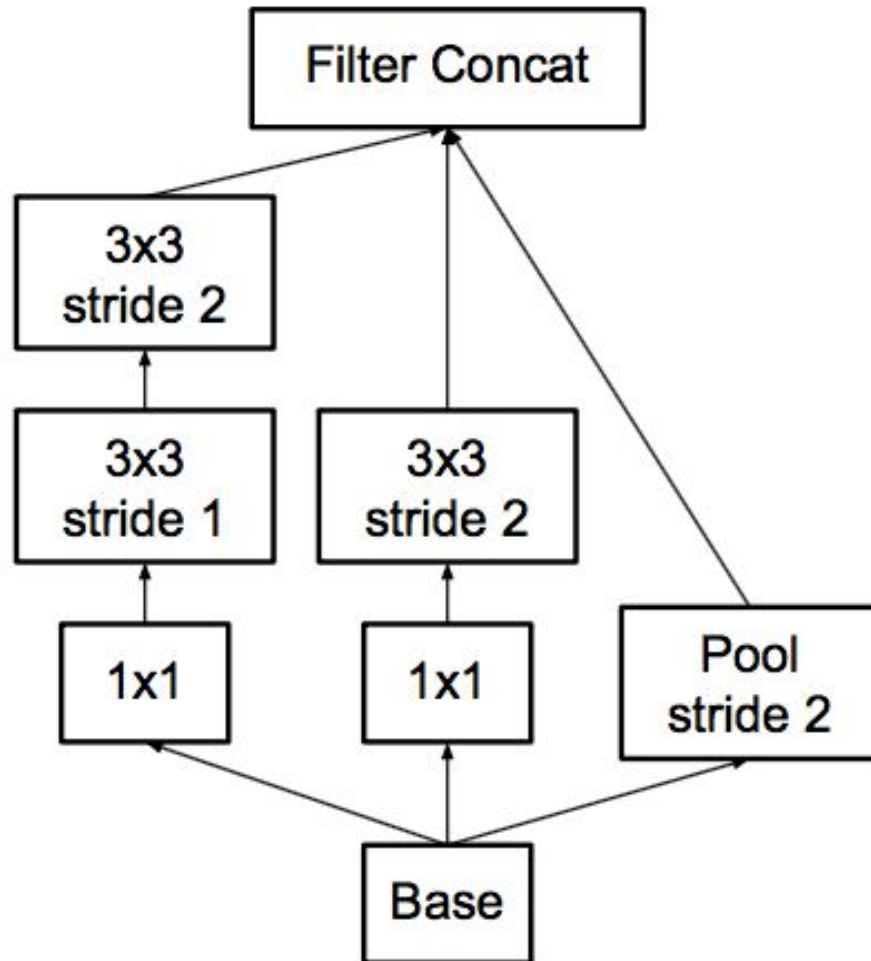


Inception v3

- Что делать с pulling? Ведь он уменьшает число параметров в 2 раза, а это плохо



Inception v3



Inception v3

- Результат 3.58% (против 6.7% у Inception v1 в 2014!)
- Столько инноваций, наверняка сеть ждёт первое место в 2015 году?
-

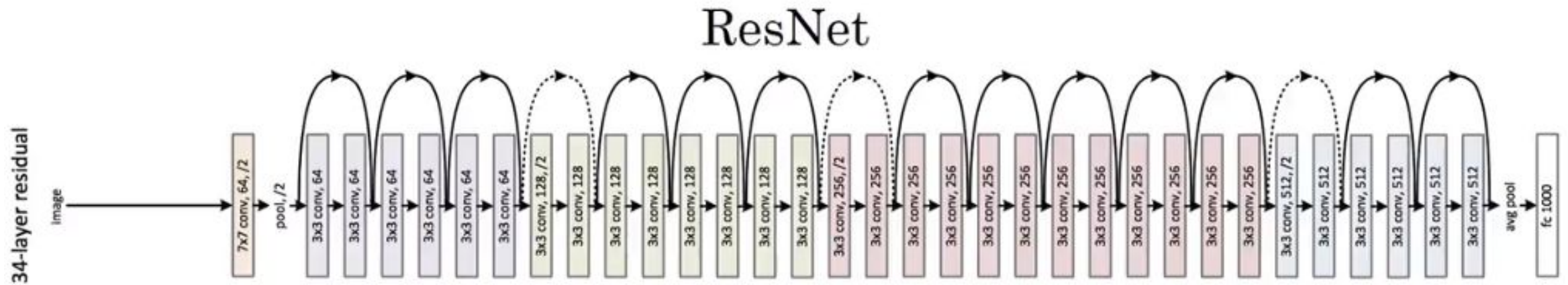
Inception v3

- Результат 3.58% (против 6.7% у Inception v1 в 2014!)
- Столько инноваций, наверняка сеть ждёт первое место в 2015 году?
-
- Выходит ResNet и показывает результат 3.57% (разница 0.01% !)

ResNet. Основная идея

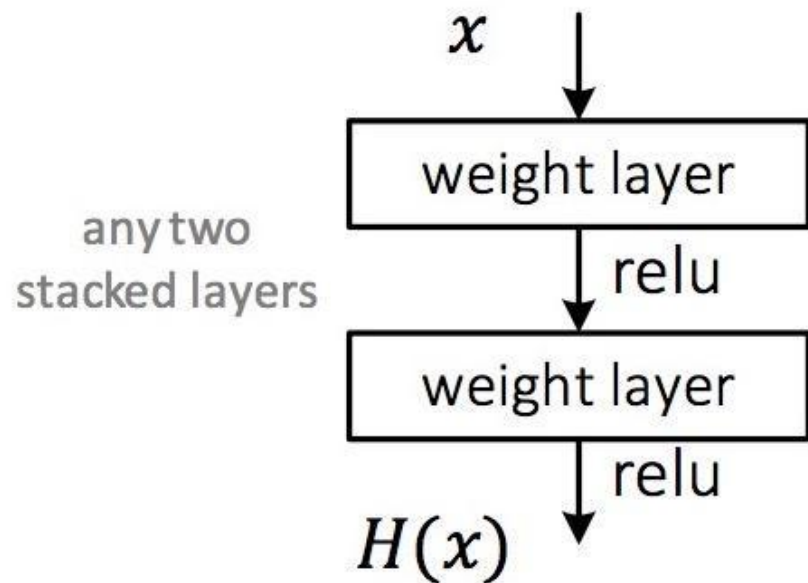
- **Теория:** если мы получаем некоторые результат с 5 слоями, то с 10 слоями мы должны иметь результат не хуже (новые 5 слоёв просто могут стать тождественными)
- **Практика:** с увеличением числа слоёв результат становится хуже с некоторого момента
- **Основная причина:** “затухание” градиента

ResNet (OT microsoft research Asia)

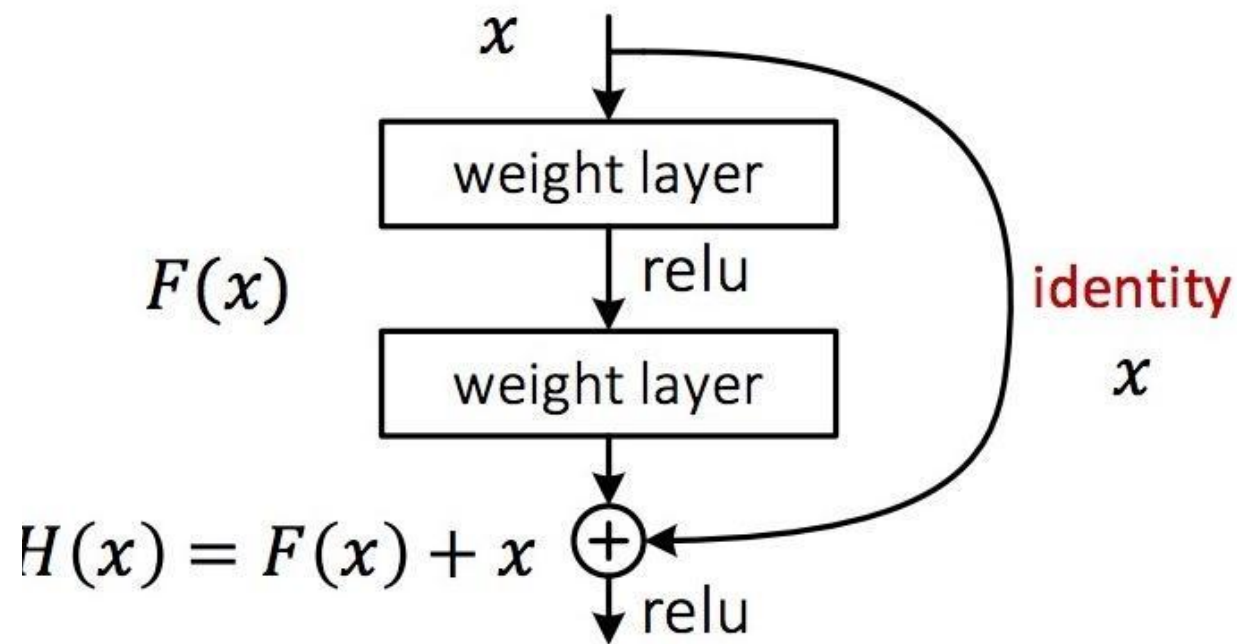


Главное нововведение: residual connections

- Plain net



- Residual net



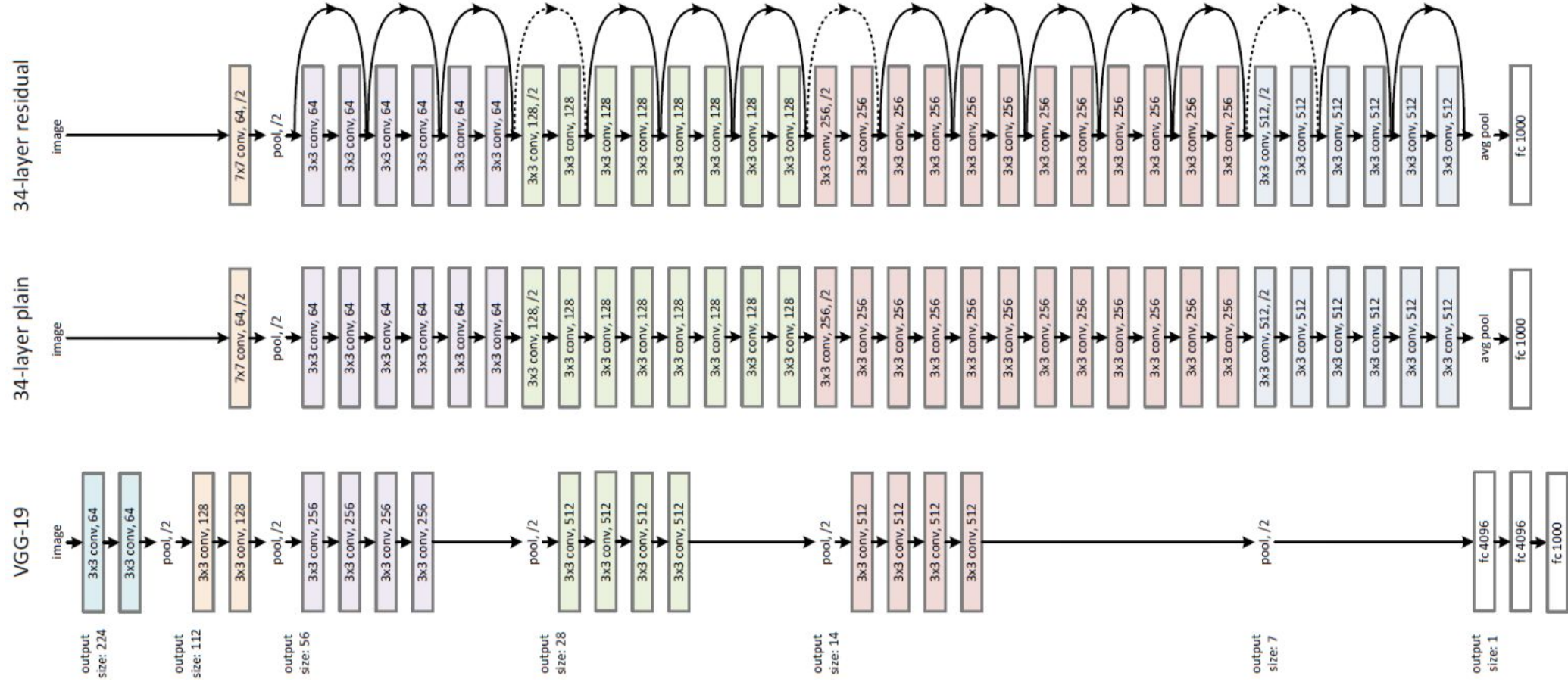
Что это даёт?

- Можно делать сверх-глубокие сети:
 - 101 слой
 - 152 слоя
 - и даже 1001 слой!
- Причём версии со 101 и 152 слоями будут иметь меньше параметров, чем VGG19

Resnet: results in 2015

- Первое место на всех основных подзадачах:
 - ImageNet classification (3.6% vs 6.7% by Inception)
 - ImageNet detection: 16% better than 2nd
 - ImageNet localization: 27% better than 2nd
 - COCO detection: 11% better than 2nd
 - COCO segmentation: 12% better than 2nd
- Работает быстрее VGGNet (но имеет в 8 раз больше слоёв)

ResNet vs VGG19



Что же дальше?

Inception-ResNet или Inception v4

- Комбинация ResNet и Inception V3
- Новые виде inception блоков
- Результат ансамбля моделей 3.08%

Transfer learning

