

Курс “Анализ изображений”

Лекция#11.  
Свёрточные нейронные сети. Основные  
архитектуры.

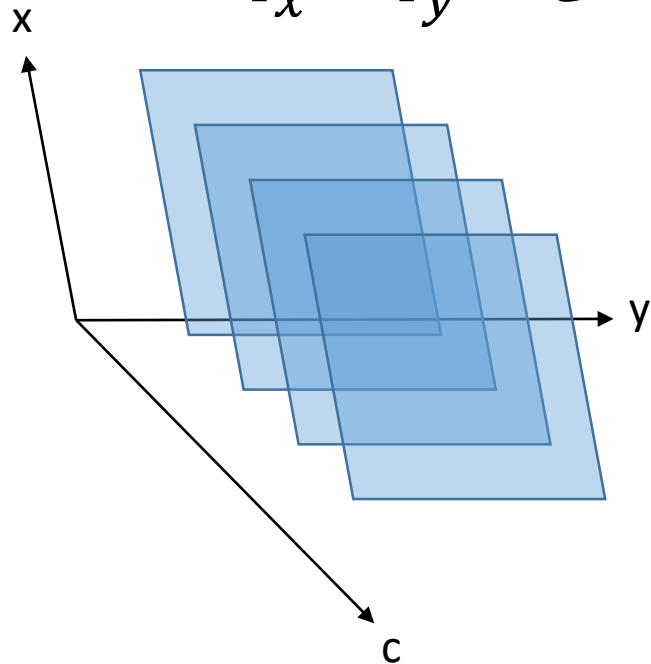
ФИВТ МФТИ

2017

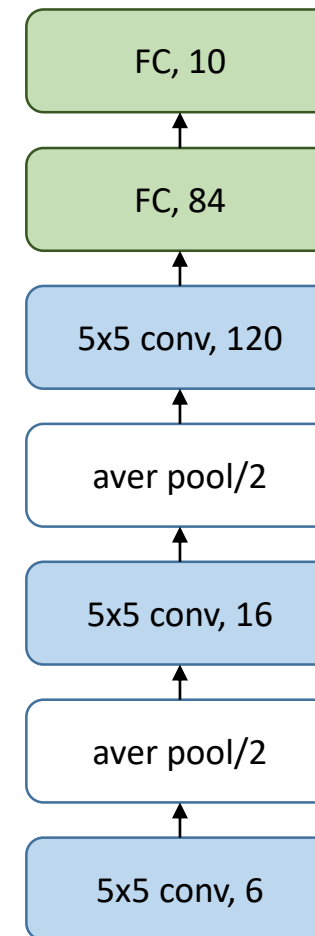
# Сверточные сети (CNN)

- Размерность слоя:  
 $X \times Y \times C \times N$

- Вход:  
 $I_x \times I_y \times C$



## LeNet, 1998



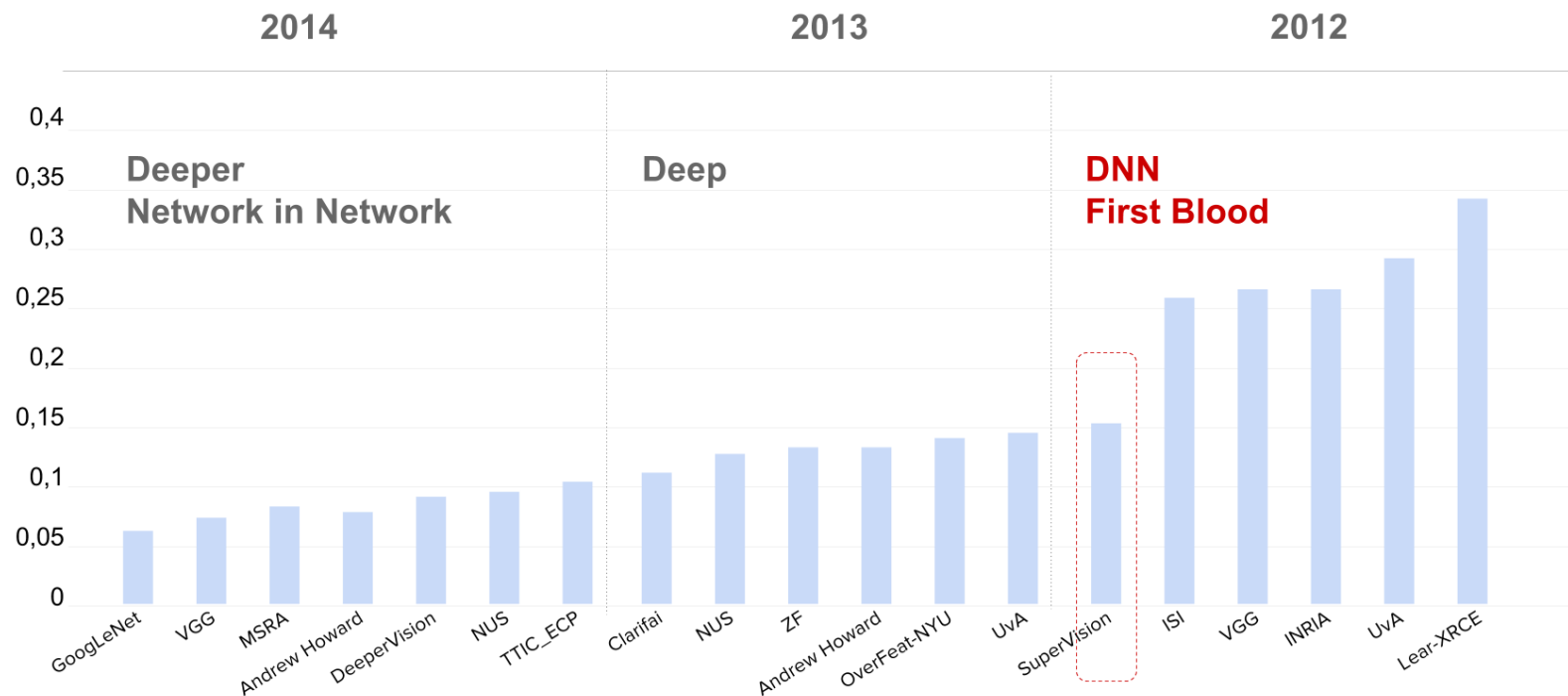
# ImageNet Large Scale Visual Recognition Challenge

- Более 15М размеченных изображений высокого разрешения
- Примерно 22К категорий
- Собрано и размечено с Amazon Mechanical Turk



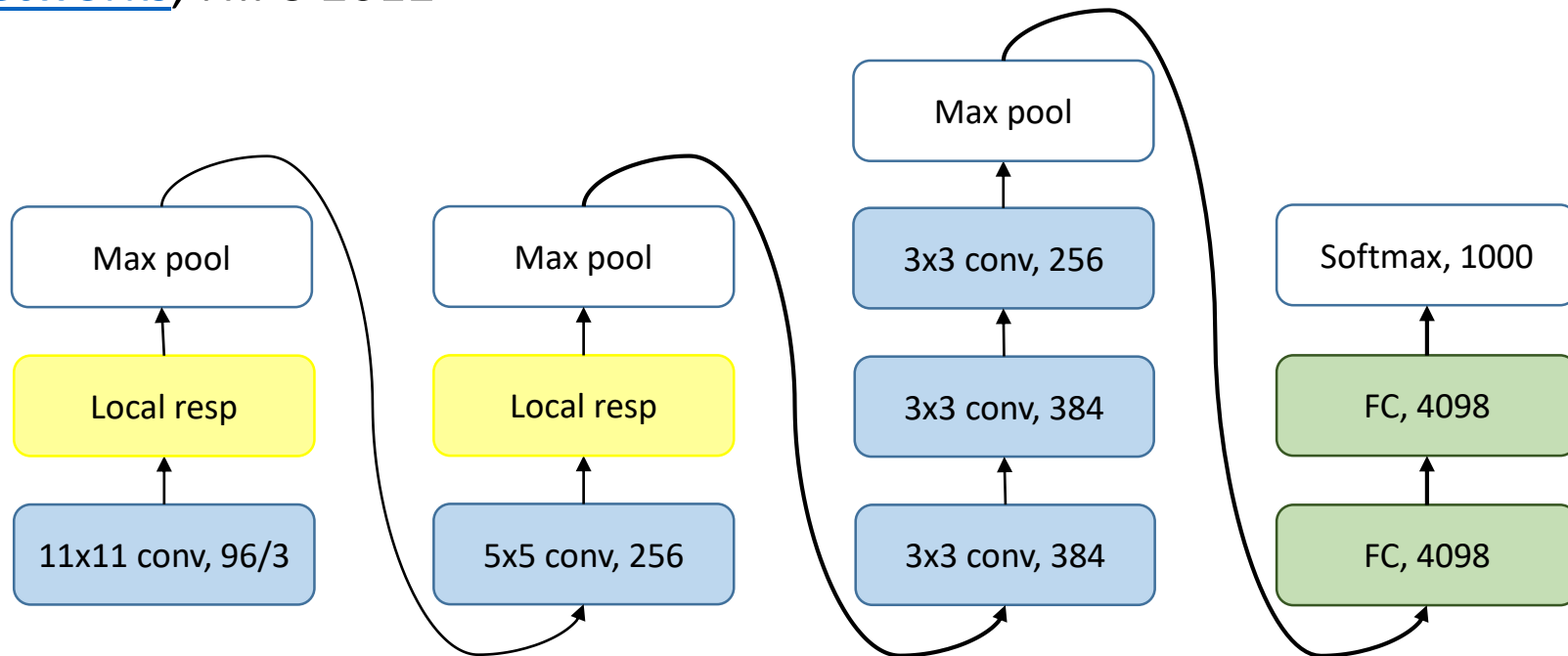
# ImageNet Large Scale Visual Recognition Challenge (ILSVRC)

- Ежегодное соревнование по классификации
- 1.2M обучающих изображений, 50K валидационных, 150K тестовых, 1K категорий
- Метрики: top-1, top-5



# AlexNet

- Победа в ILSVRC 2012:  $top5 = 16\%$
- Параллельное обучение на двух GPU
- 60M параметров
- A. Krizhevsky, I. Sutskever, G. Hinton, [ImageNet Classification with Deep Convolutional Neural Networks](#), NIPS 2012



# Local Response normalization

Нормализованный ответ слоя:

$$b_i = \frac{a_i}{\left(k + \alpha \sum_{j=\max(0, i-n/2)}^{\min(N-1, i+n/2)} (a_j)^2\right)^\beta}$$

$a_i$  - выход нейрона, посчитанный после применения ядра  $I$  в некоторой позиции  $(x, y)$  и после применения ReLU

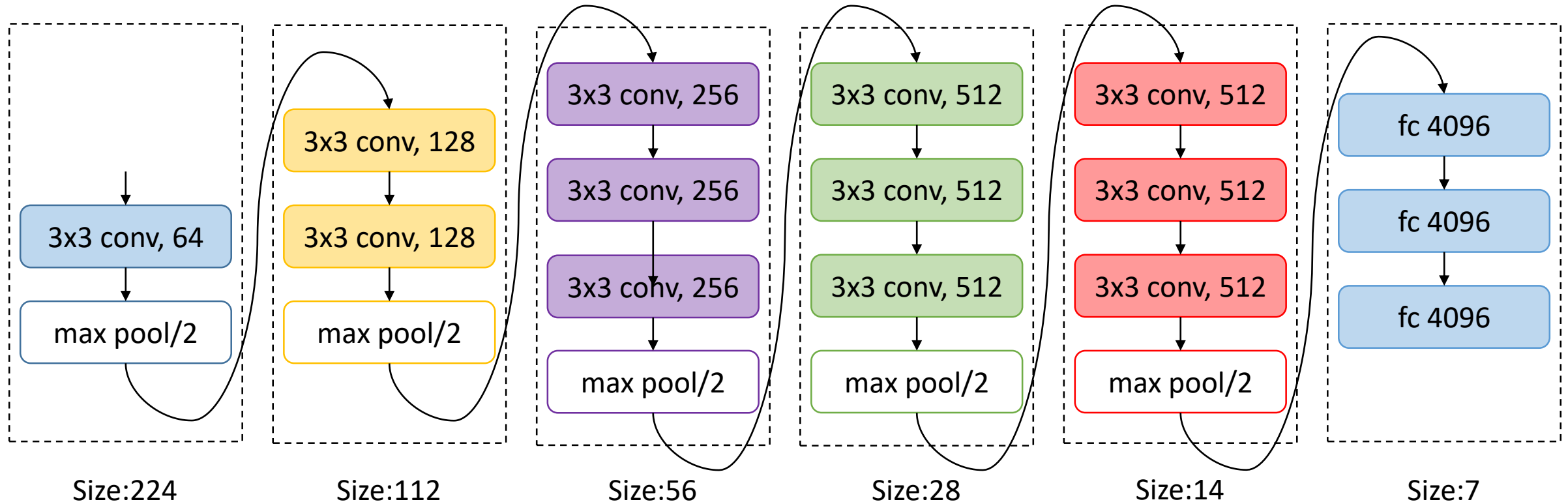
- Не требуется нормализация входа с ReLU
- Улучшает обобщение

# ZFNet – Zeiler & Fergus Net

- Победа в ILSVRC 2013
- Предложили вариант визуализации сети (deconvnet)
- Улучшение AlexNet, с лучшим подбором гиперпараметров (в частности – увеличением размера средних сверточных слоев)
- М. Zeiler and Fergus, [Visualizing and Understanding Convolutional Networks](#), 2013

# VGG-16

- Oxford Visual Geometry Group (VGG)
- Simonyan and Zisserman, [Very Deep Convolutional Networks for Large-Scale Image Recognition](#), ICLR, 2015



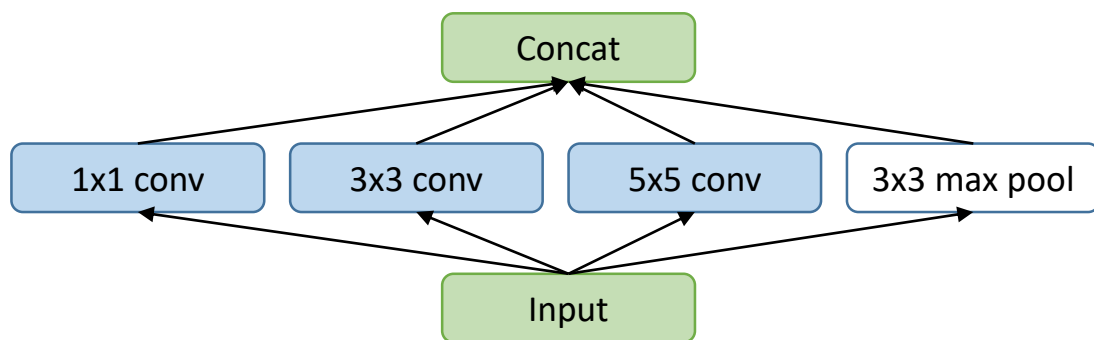


# VGG-16

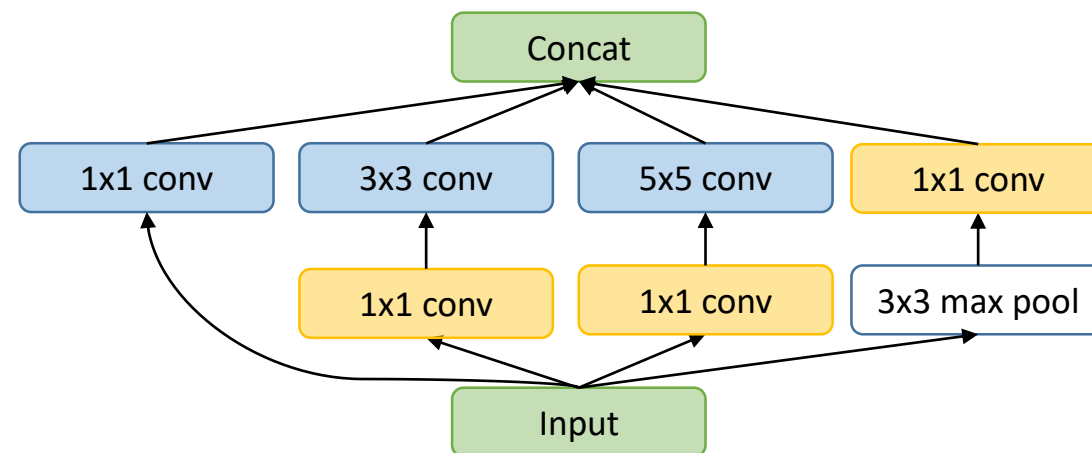
- 2 место в ILSVRC 2014:  $top5 = 7.3\%$
- Показали, что глубина сети является важной компонентой для качества
- Уменьшение количества параметров:
  - 1 фильтр  $7 \times 7$ :  $C \times (7 \times 7 \times C) = 49C^2$
  - 3 фильтра  $3 \times 3$ :  $3 \times (C \times (3 \times 3 \times C)) = 27C^2$
- 138M параметров (VGG-16) 144M параметров (VGG-19), 100M – в первом полносвязном слое

# Inception, 2014

Inception, naive



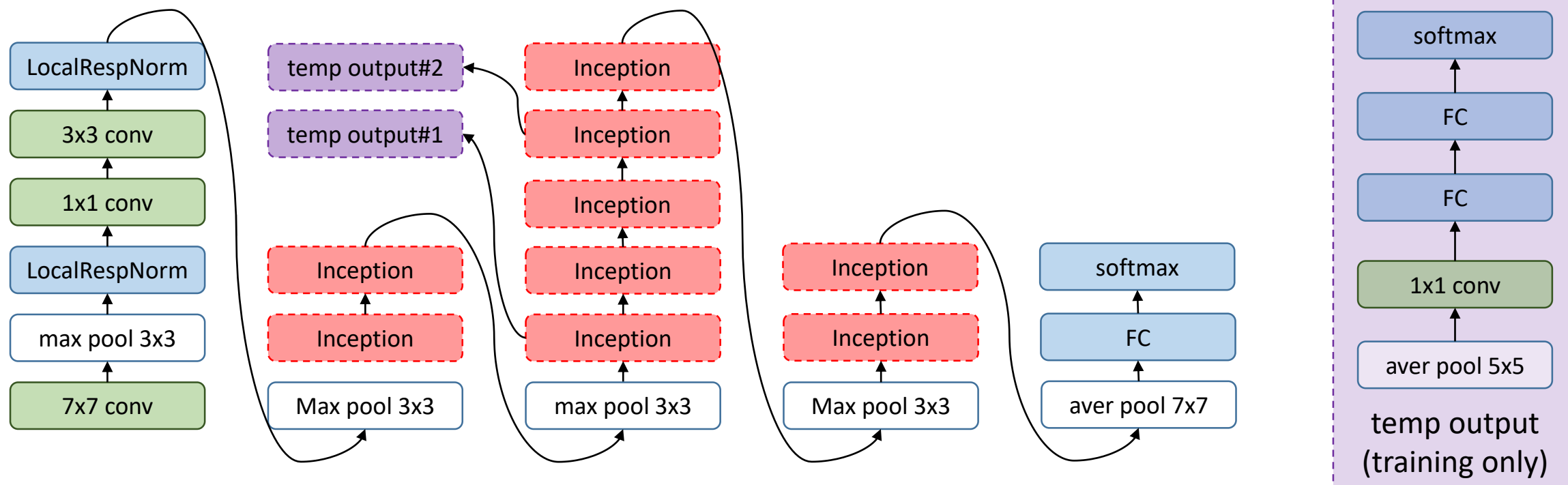
Inception, with dimensionality reduction



C. Szegedy et al, [Going deeper with convolutions](#), CVPR, 2015

# GoogLeNet-v1

- Победа в ILSVRC 2014:  $top5 = 6.7\%$
- 4M параметров
- C. Szegedy and etc, [Going deeper with convolutions](#), CVPR, 2015



# GoogLeNet-v2

Идеи:

- Разделение пространственных сверток на более мелкие:  $N \times N \rightarrow N \times 1$  и  $1 \times N$
- Параллельное исполнение max-pool слоя и сверточного слоя с более редким шагом для более плавного снижения размерности

# Batch normalization

$$\hat{x} = \frac{x - Mx}{\sqrt{Dx + \varepsilon}}$$
$$y = \gamma \hat{x} + \beta$$

- $\gamma, \beta$  — обучаемые параметры
- Обычно идет перед функцией активации
- Увеличивает скорость сходимости
- Используется в GoogLeNet v3
- S. Ioffe, C. Szegedy, [Batch normalization: Accelerating deep network training by reducing internal covariate shift](#), ICML 2015

# ResNet

Желаемый выход:  $H(x)$

Хотим научить блок остатку:

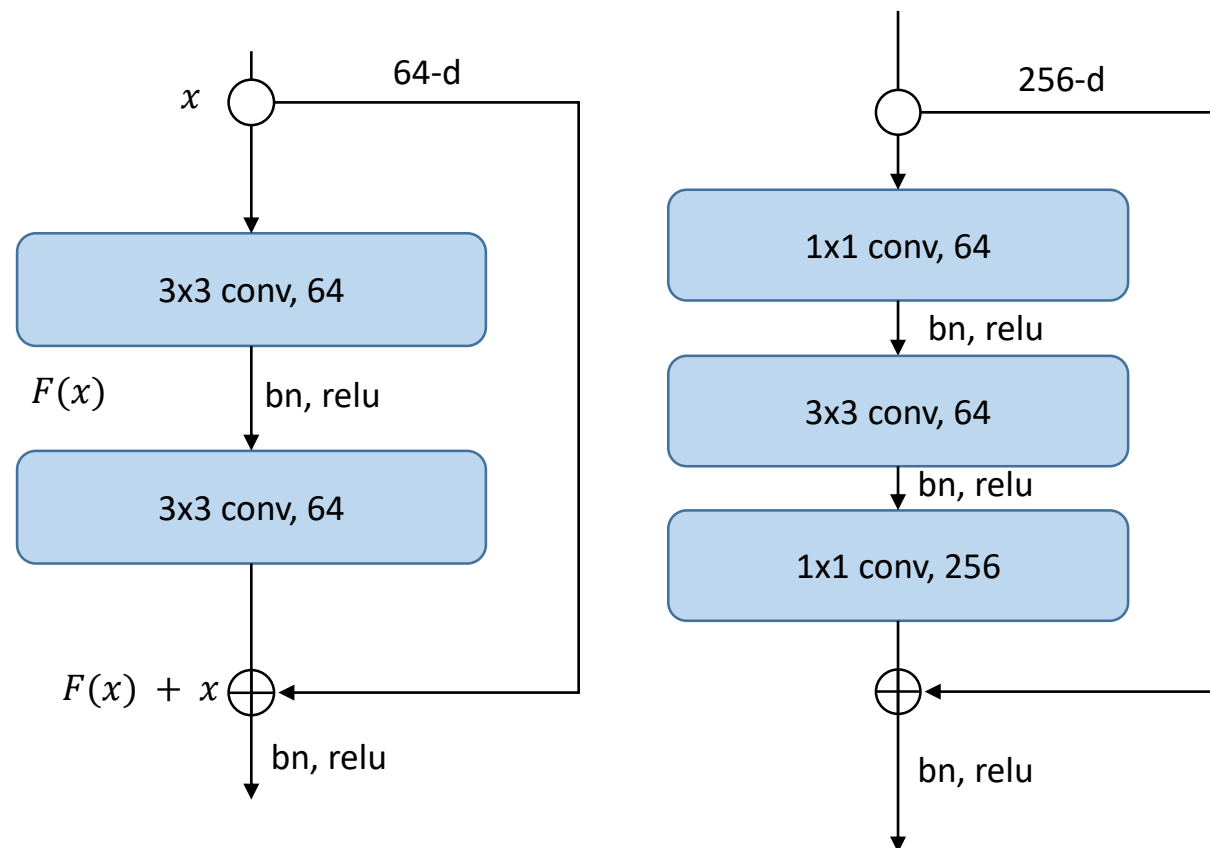
$$F(x) = H(x) - x$$

Для этого учим:

$$H(x) = F(x) + x$$

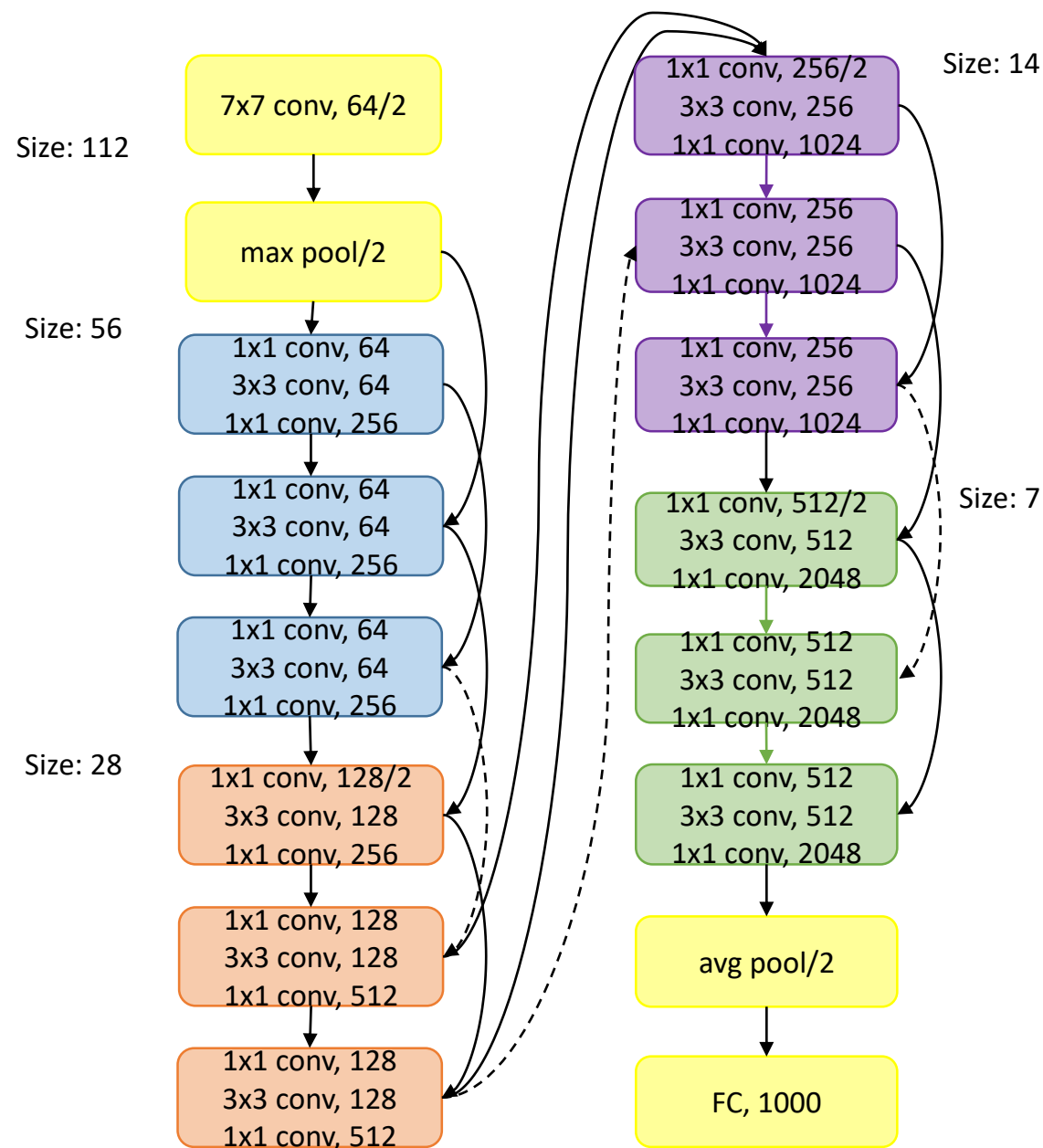
В случае оптимального входа:

$$F(x) = 0$$

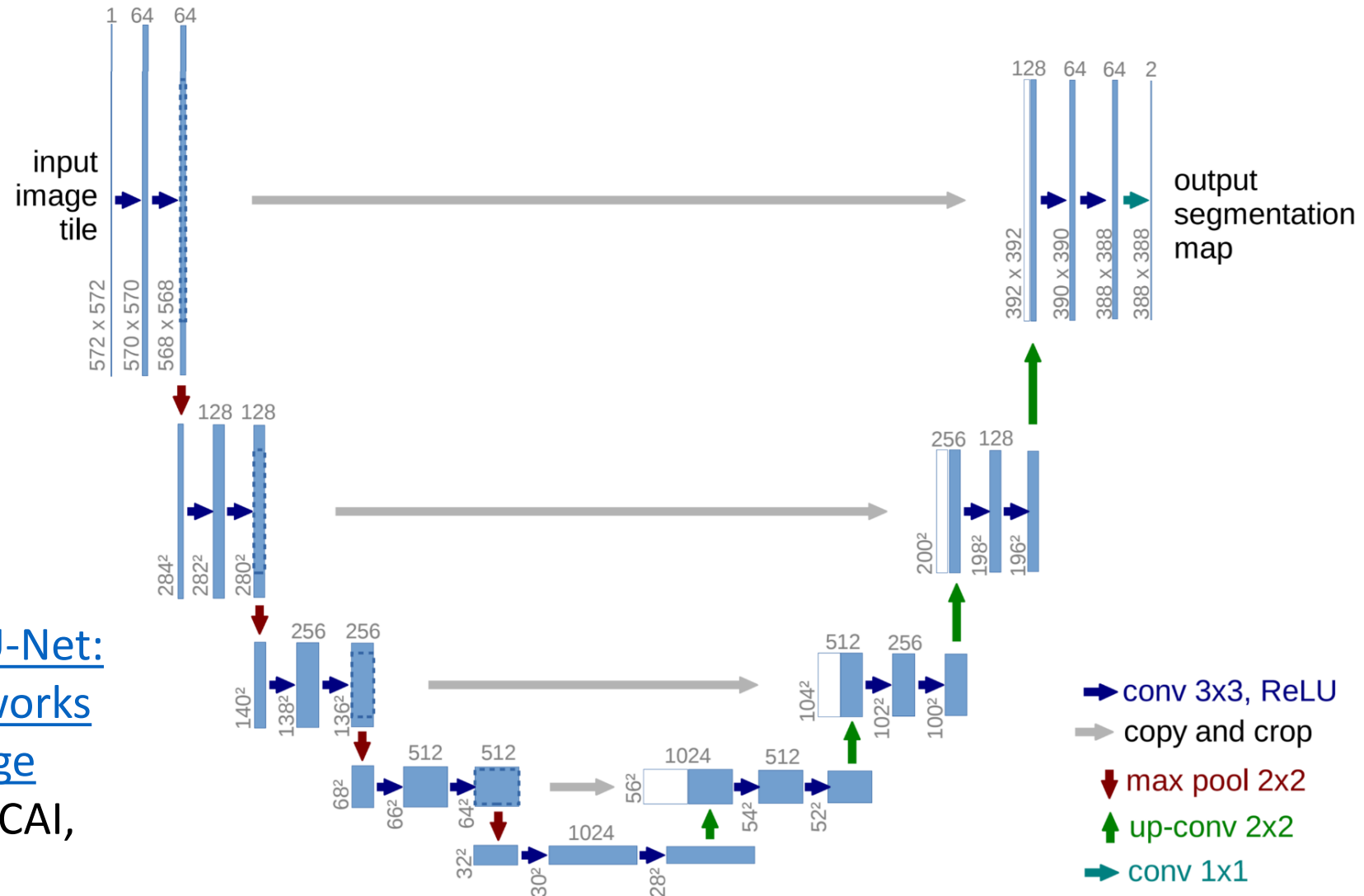


# ResNet

- Победа в ILSVRC 2015
- Позволяет обучать действительно глубокие сети (150 слоев)
- Можно удалять слои без значительной потери в качестве
- К. He, X. Zhang, S. Ren, J. Sun, [Deep Residual Learning for Image Recognition](#), CVPR 2016



# U-net



O. Ronneberger,  
P. Fischer, T. Brox, [U-Net: Convolutional Networks for Biomedical Image Segmentation](#), MICCAI, 2015