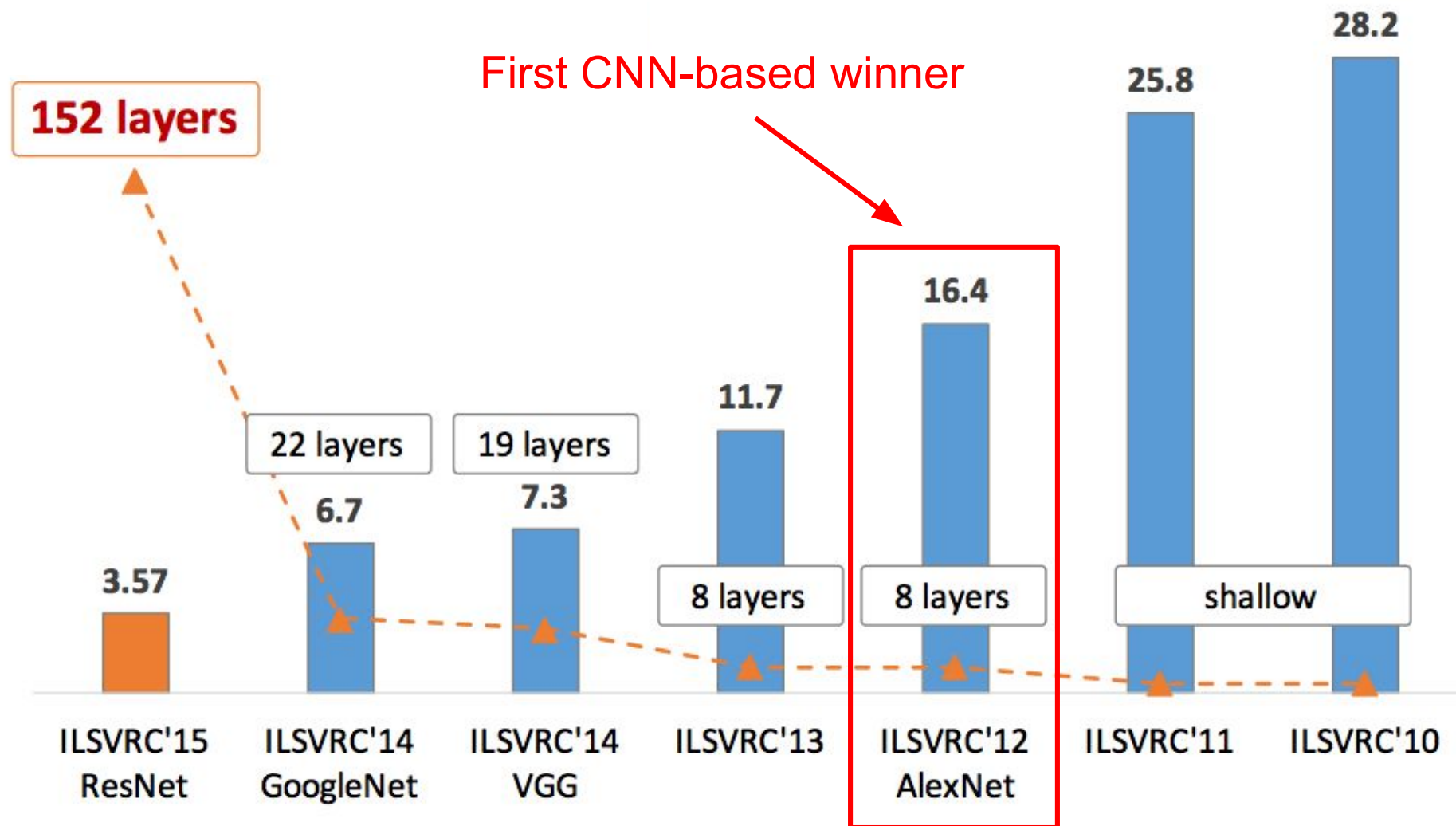
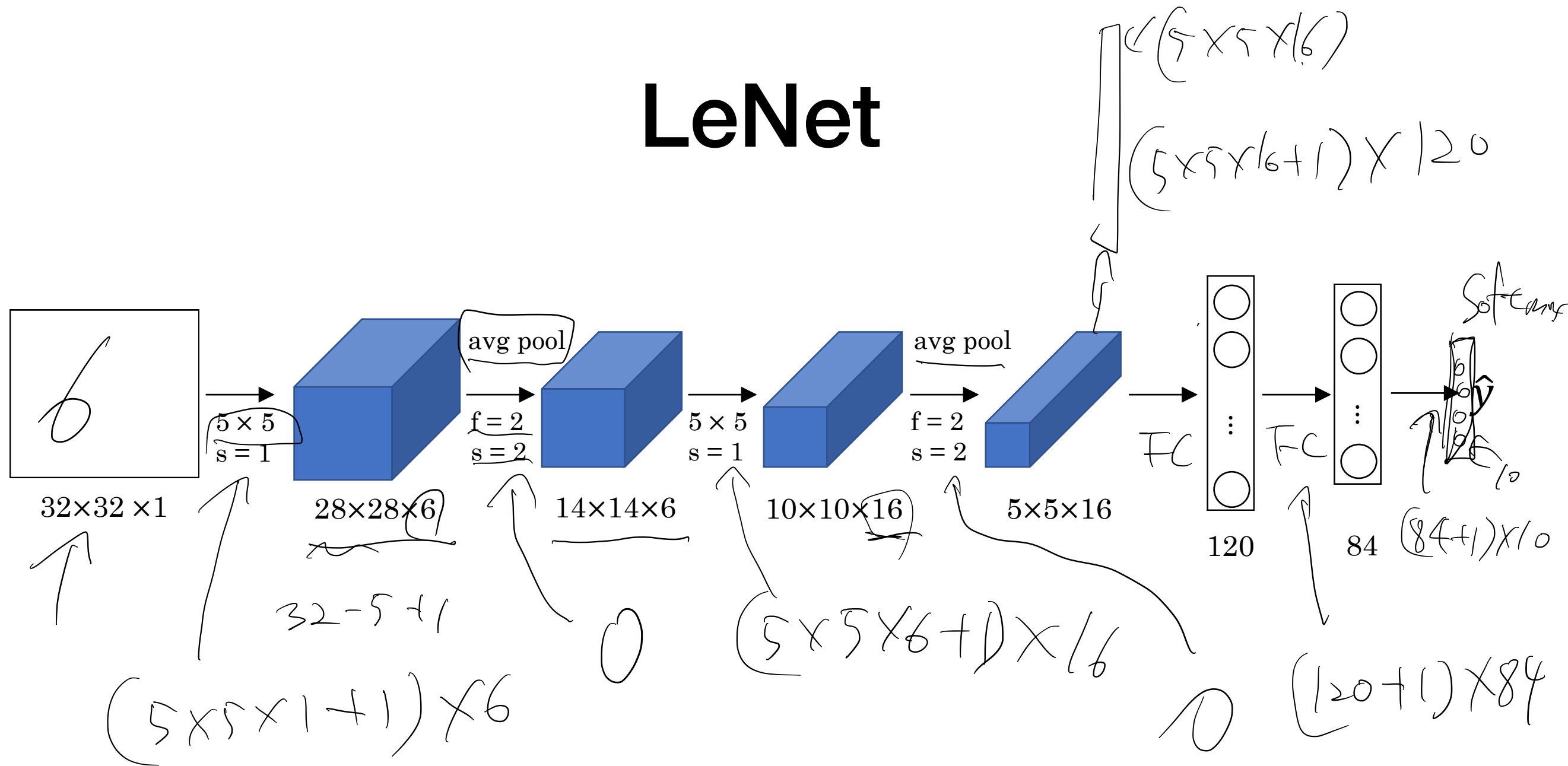


CNN Architectures

ImageNet Large Scale Visual Recognition Challenge (ILSVRC) winners



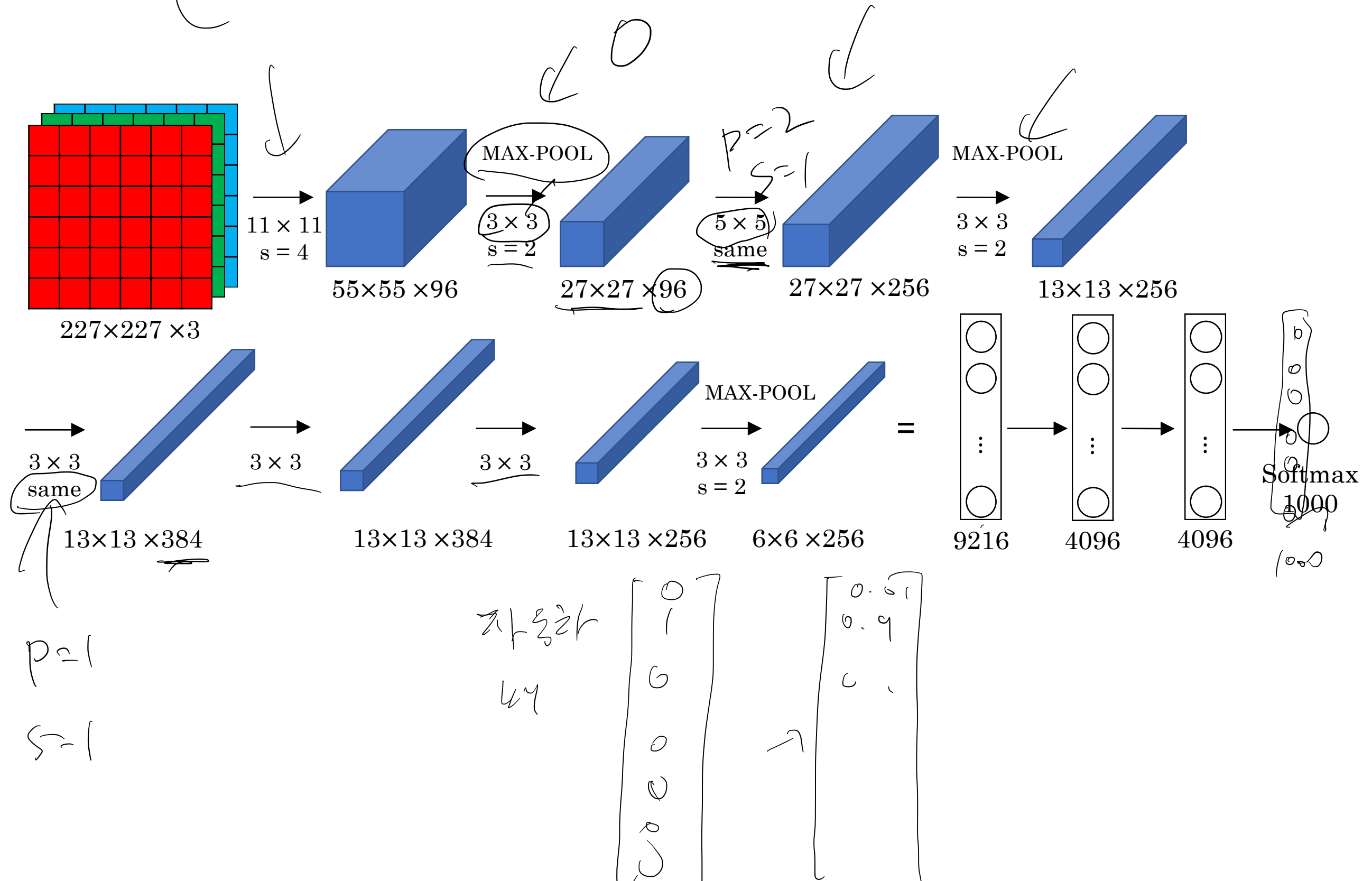
LeNet



AlexNet

$$(11 \times 11 \times 3 + 1) \times 96$$

$$(5 \times 5 \times 96 + 1) \times 256$$



VGG

filter size

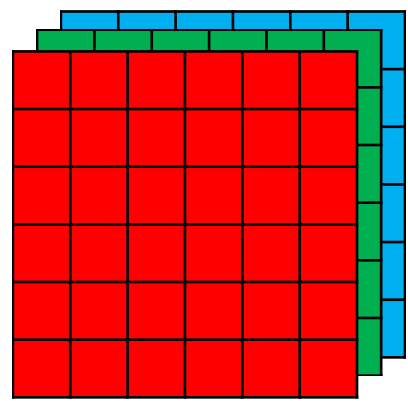
3x3

5x5

7x7

CONV = 3x3 filter, s = 1, same

MAX-POOL = 2x2, s = 2



224x224 x 3

$(3 \times 3 \times 3) \times 64$

x2

$(3 \times 3 \times 64 + 1) \times 64$

max pooling

224x224x64

POOL

112x112 x 64

[CONV 128]

x2

$(3 \times 3 \times 64 + 1) \times 128$

112x112 x 128

POOL

56x56 x 128

[CONV 256]

x3

56x56 x 256

POOL

28x28 x 256

[CONV 512]

x3

28x28 x 512

POOL

14x14 x 512

[CONV 512]

x3

14x14 x 512

POOL

7x7x512

FC
4096

FC
4096

Softmax
1000

Inception Net

- o vanishing gradient.
- o Too many parameters

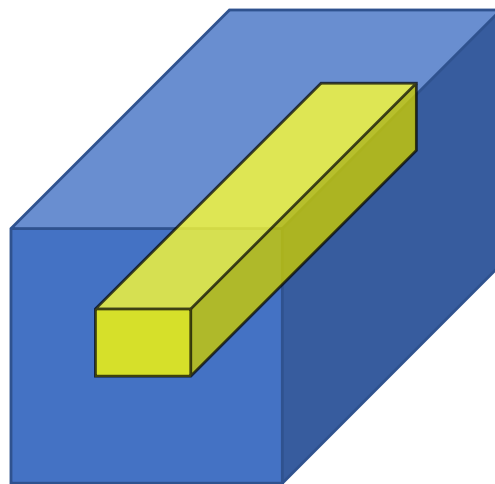


<http://knowyourmeme.com/memes/we-need-to-go-deeper>

Why does a 1×1 convolution do?

1	2	3	6	5	8
3	5	5	1	3	4
2	1	3	4	9	3
4	7	8	5	7	9
1	5	3	7	4	8
5	4	9	8	3	5

6×6



$6 \times 6 \times 32$

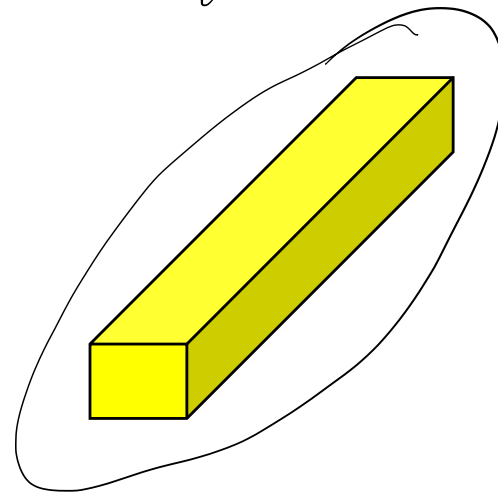
*

2

=

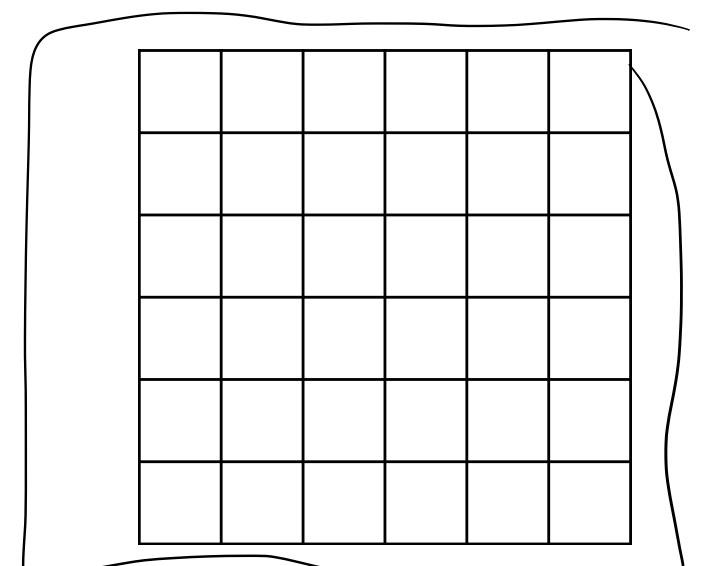
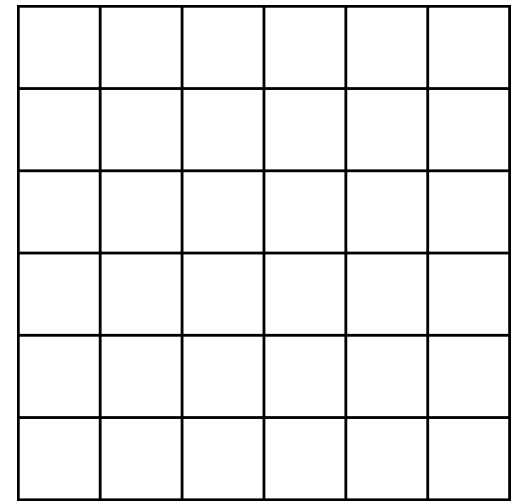
1/6, 1x1 filters

*



$1 \times 1 \times 32$

=

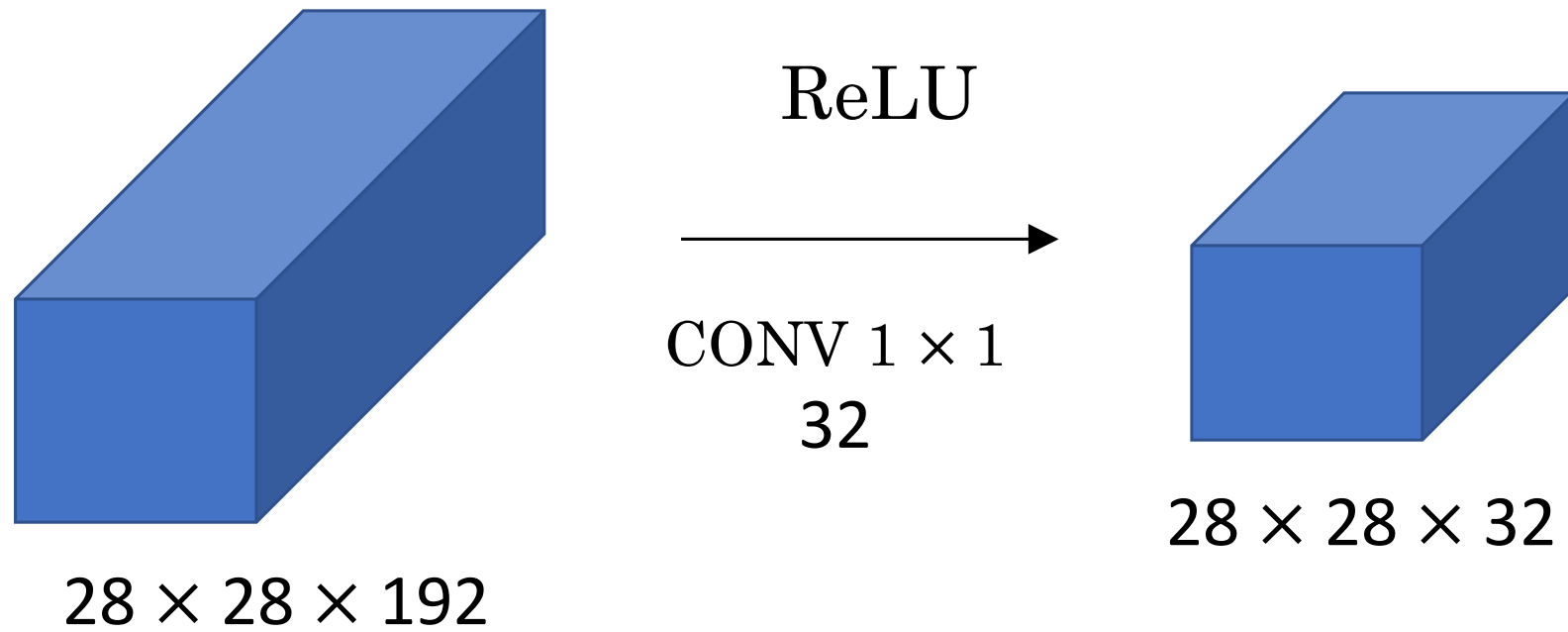


$6 \times 6 \times \# \text{ filters}$

[Lin et al., 2013. Network in network]

1/6

Using 1x1 convolutions

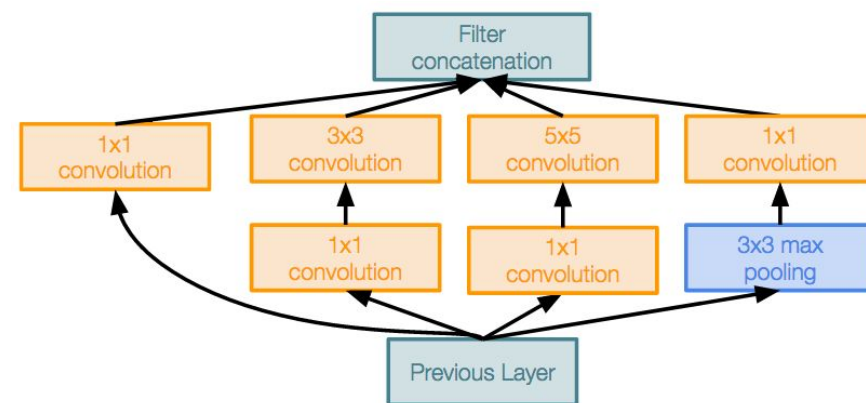


Inception Module

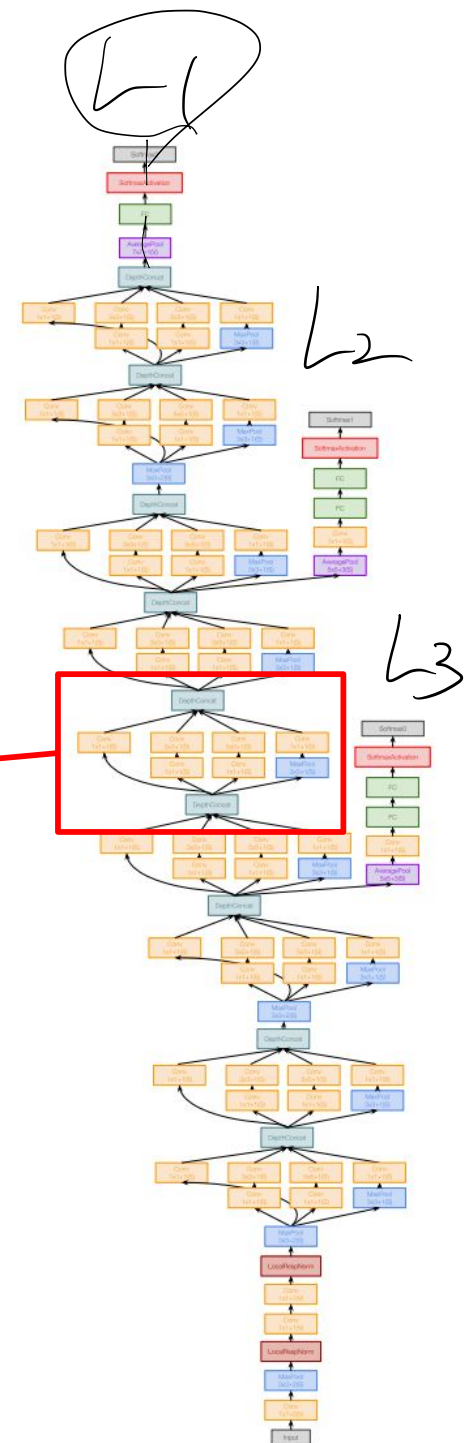
Case Study: GoogLeNet

[Szegedy et al., 2014]

“Inception module”: design a good local network topology (network within a network) and then stack these modules on top of each other



Inception module

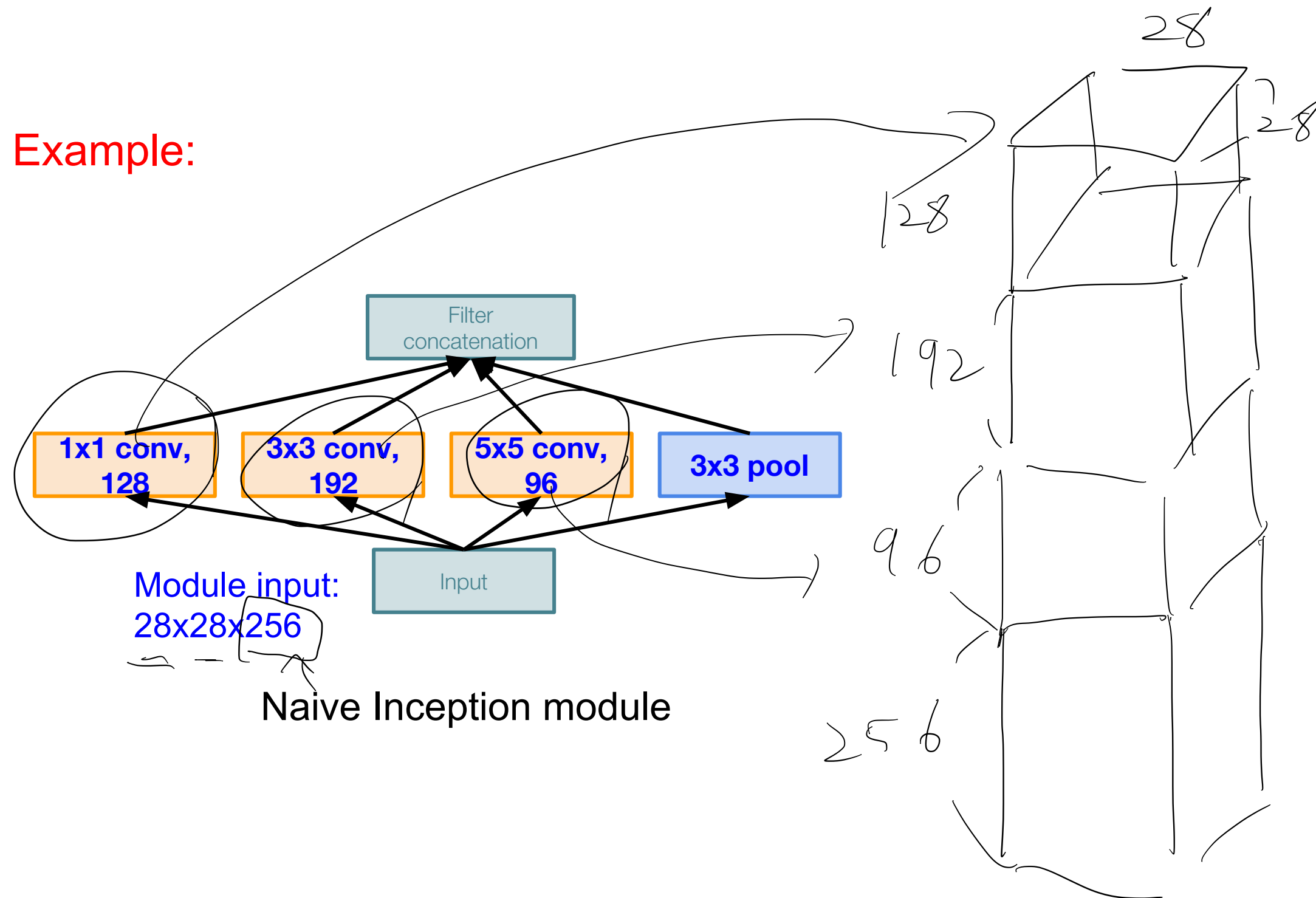


$$L = \lambda_1 L_1 + \lambda_2 L_2 + \lambda_3 L_3$$

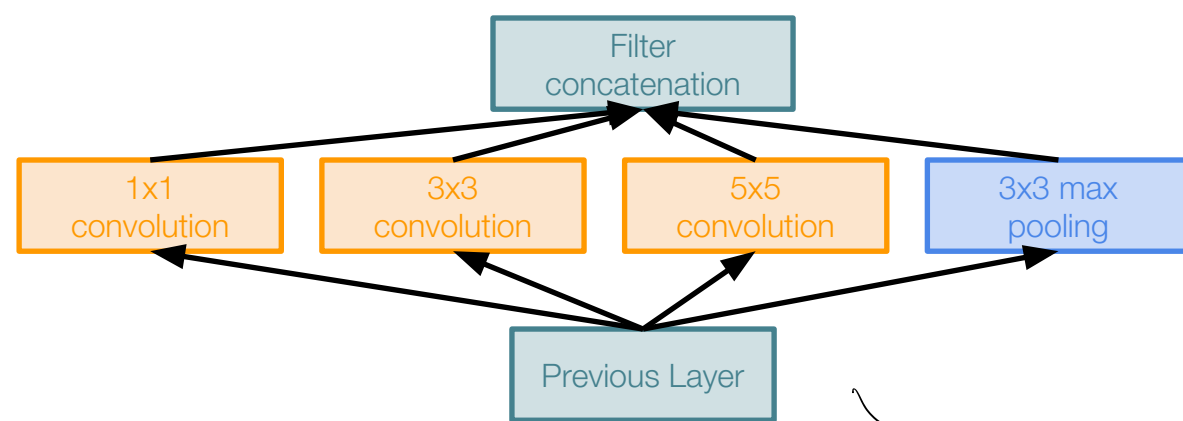
$$W \leftarrow W - \eta \nabla_W L = \eta (\lambda_1 \nabla L_1 + \lambda_2 \nabla L_2 + \lambda_3 \nabla L_3)$$

Inception Module

Example:

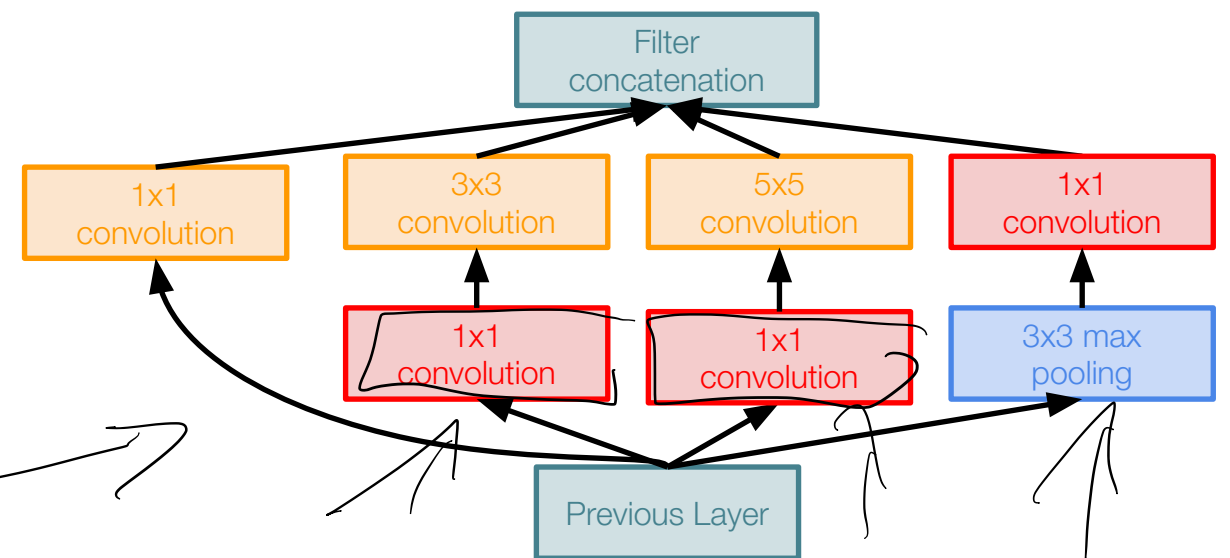


With Bottleneck



Naive Inception module

1x1 conv “bottleneck”
layers

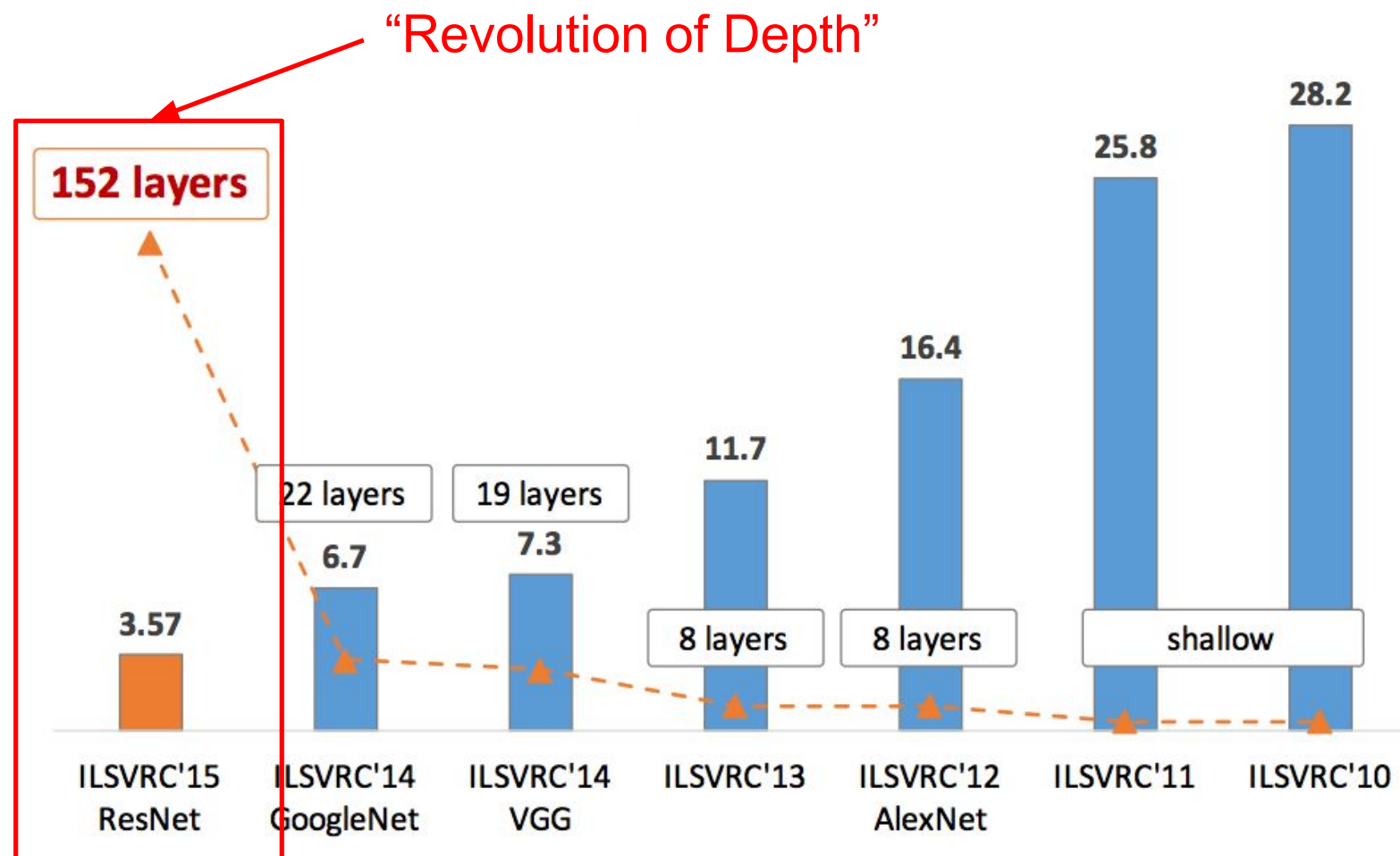


Inception module with dimension reduction

More Deeper

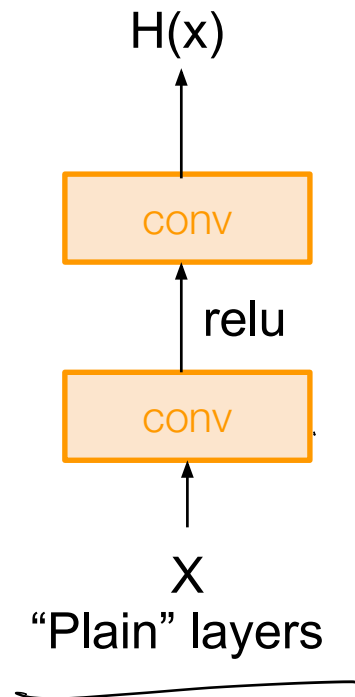


ImageNet Large Scale Visual Recognition Challenge (ILSVRC) winners



Residual Block

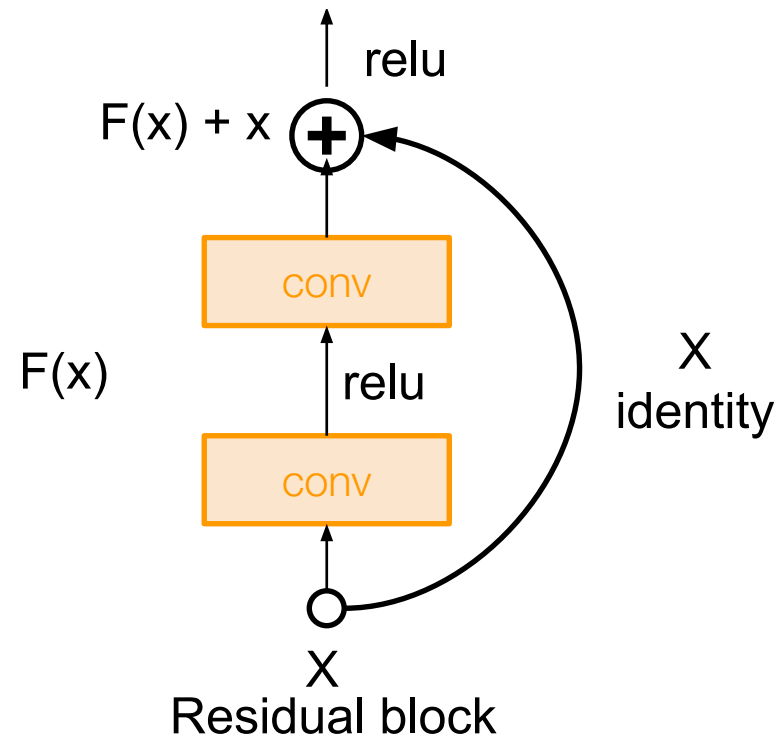
$H(x)$



\Downarrow

$H(x)$

2-layer NN



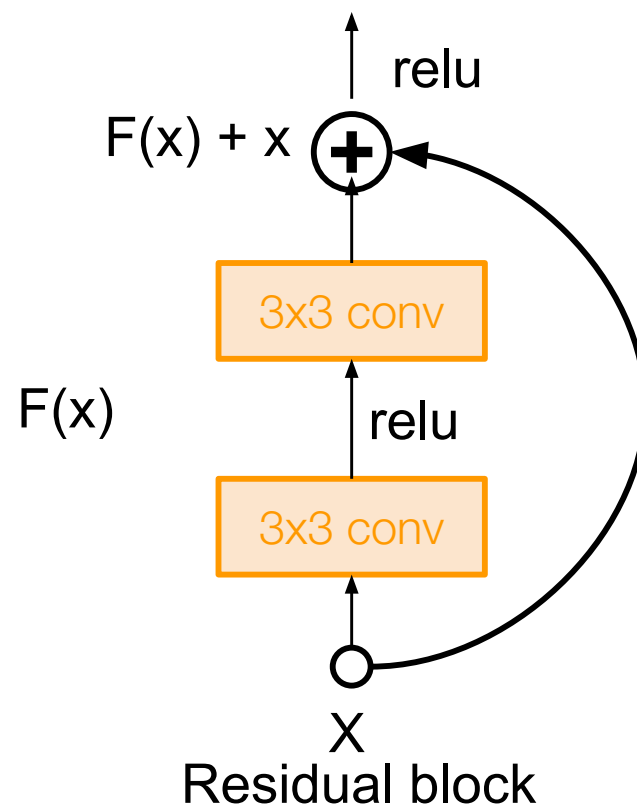
Residual block

2-layer NN

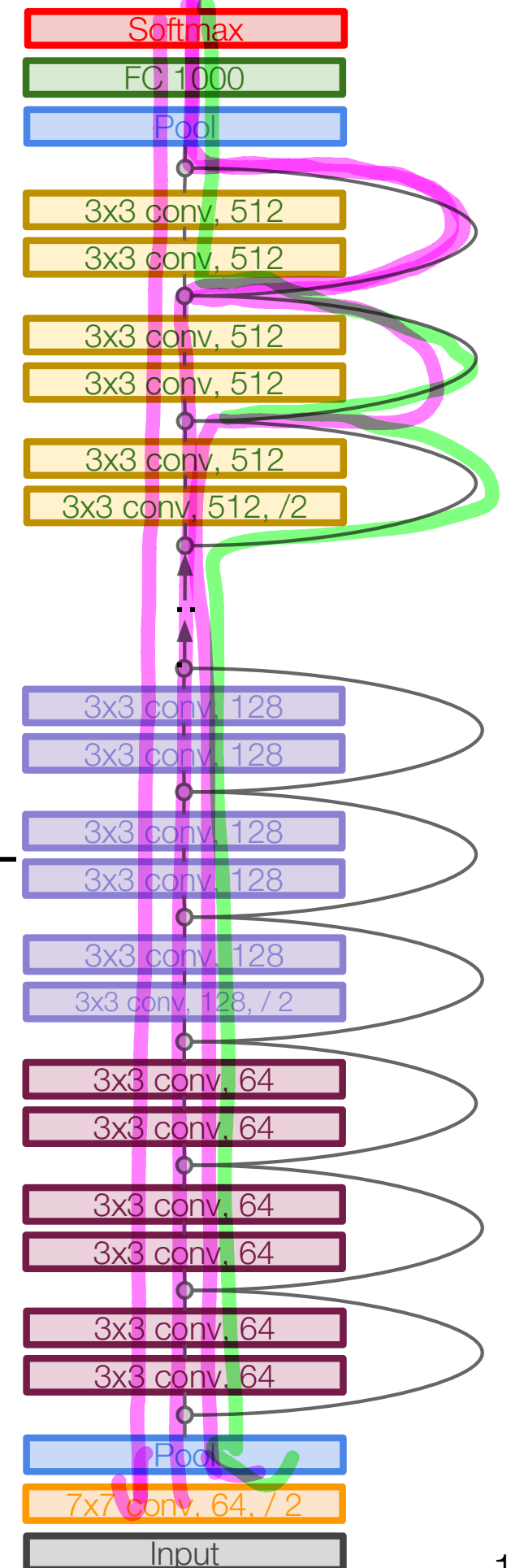
$$H(x) = x + F(x)$$

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

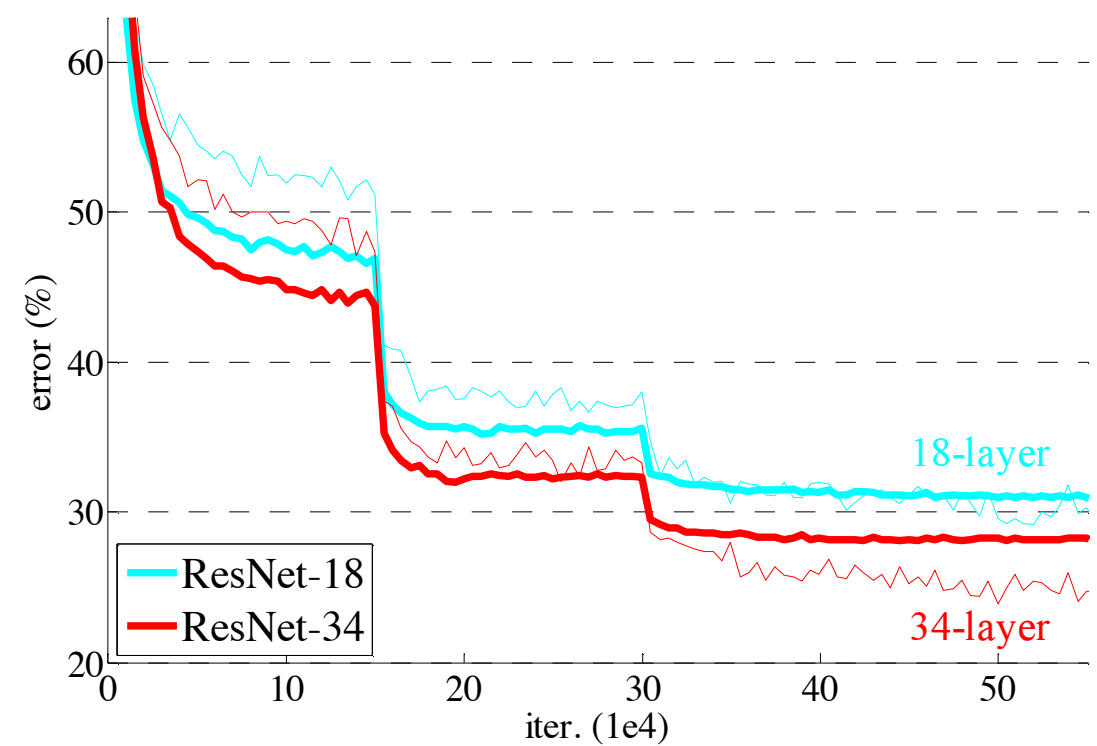
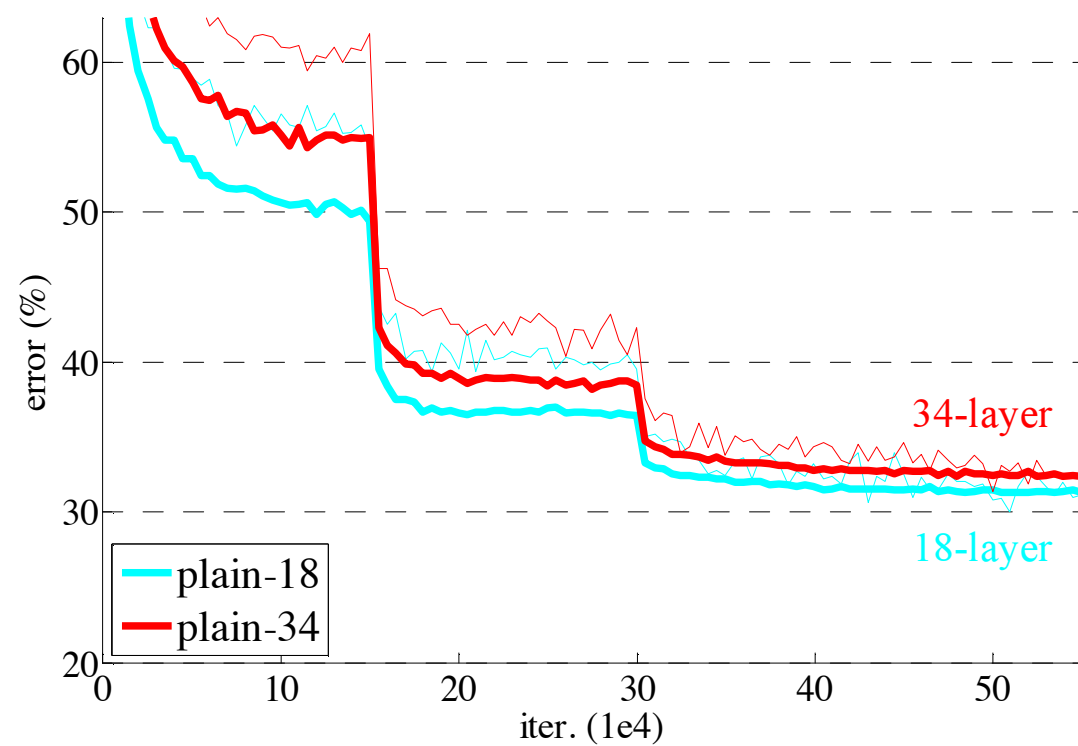
ResNet



X
identity

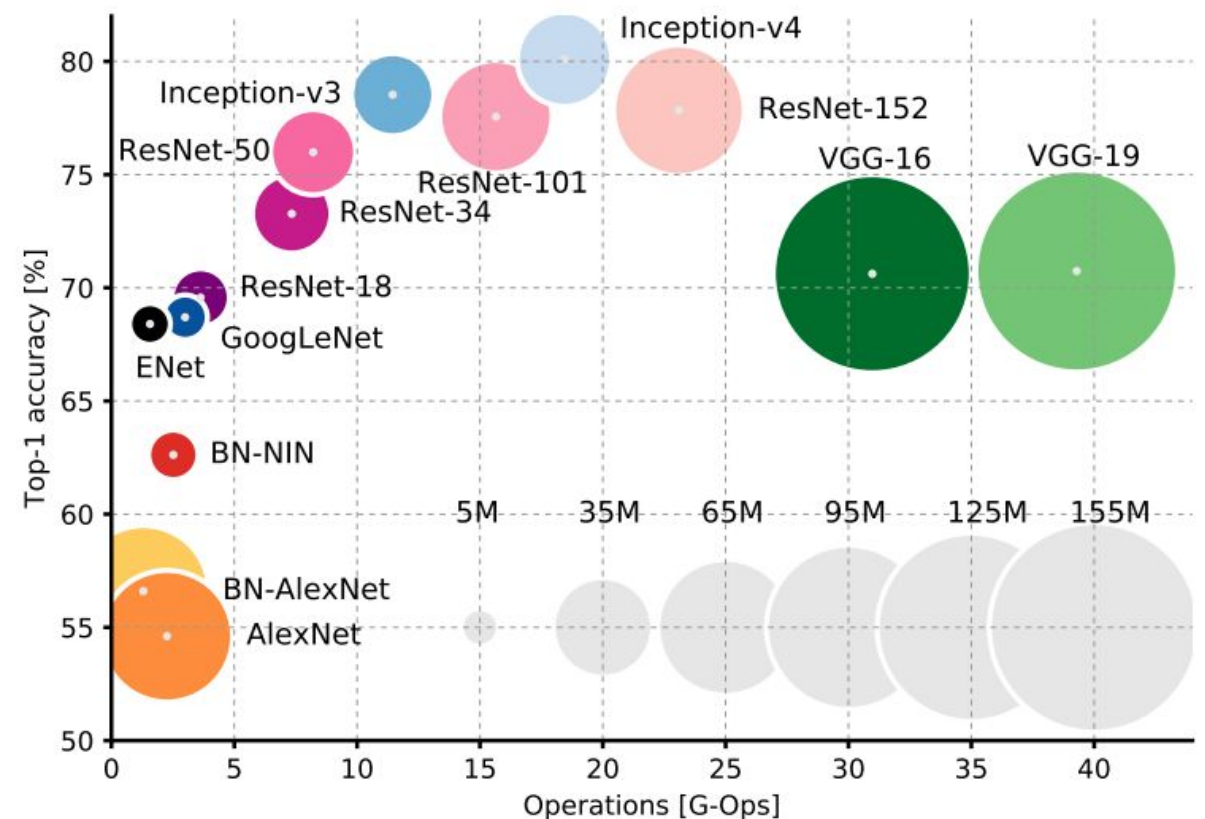
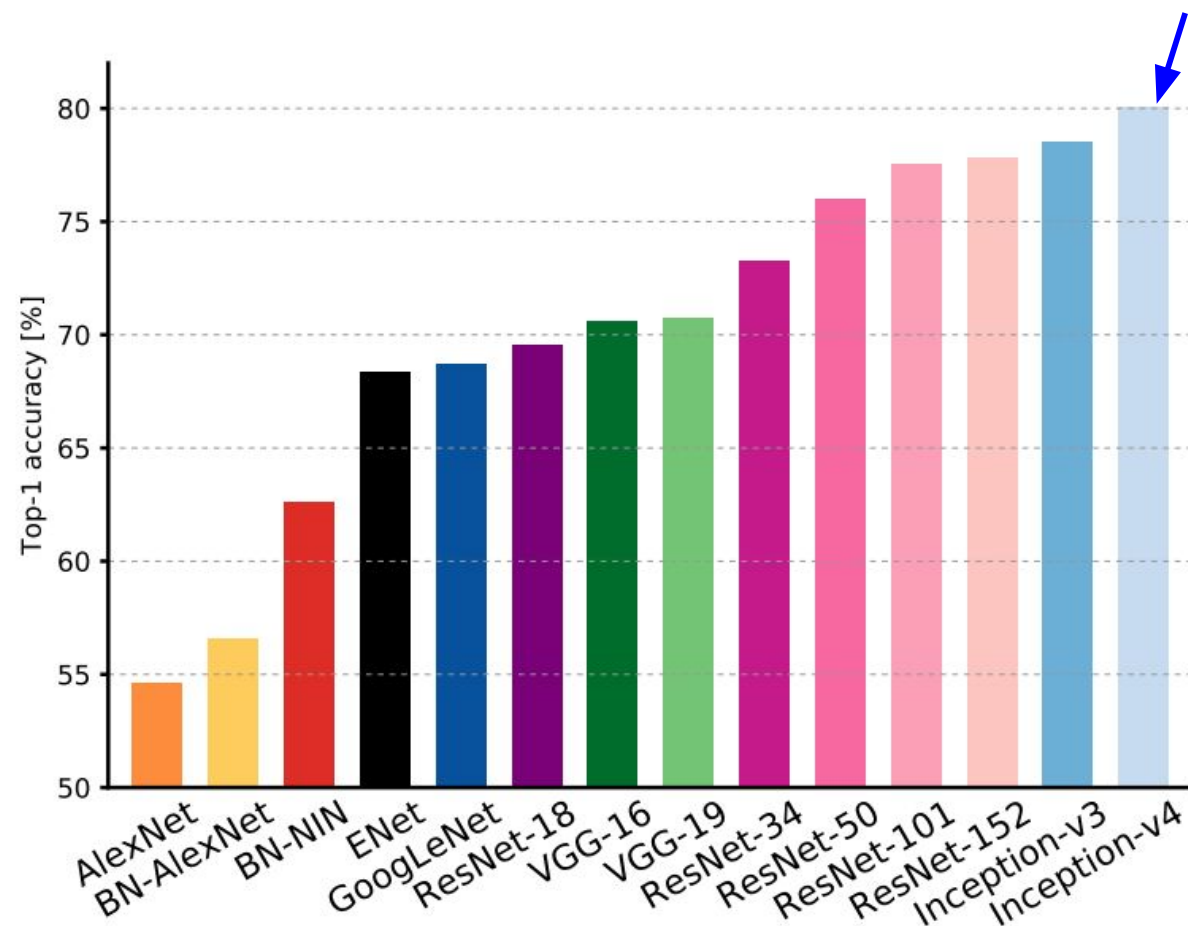


ResNet results



Comparing complexity...

Inception-v4: Resnet + Inception!



An Analysis of Deep Neural Network Models for Practical Applications, 2017.