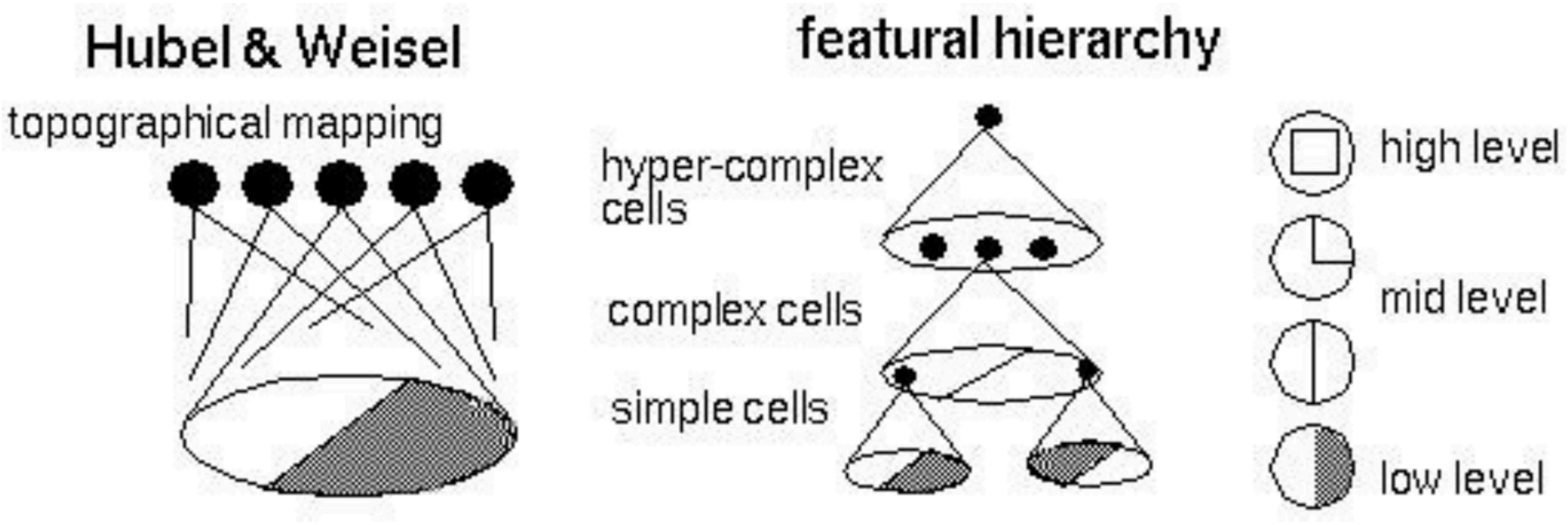
Convolution Neural Networks

Paul Rad, Ph.D.

Associate Professor
Cyber Analytics and AI
Information Systems and Cyber Security
College of Business School
210.872.7259

Vision: Hierarchical Organization (Year: 1962)



Why use hierarchical multi-layered models?

Biological vision is hierachically organised

object

↑

object parts

↑

primitive features

↑

input image

trees

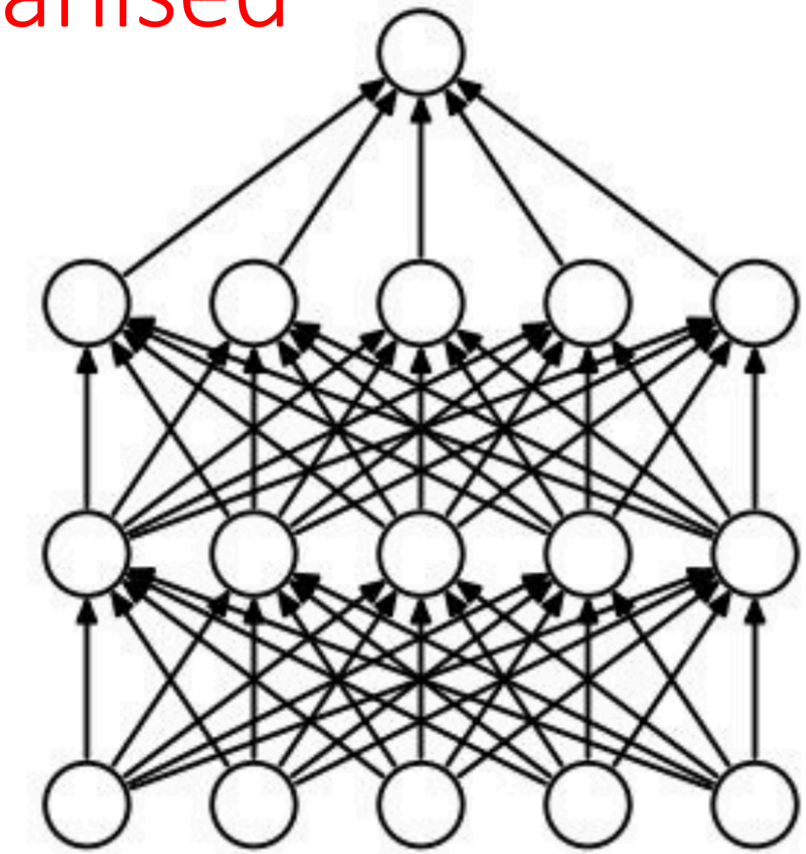
↑

bark, leaves, etc.

↑

oriented edges

↑



What's wrong with standard neural networks?

Hard to Train

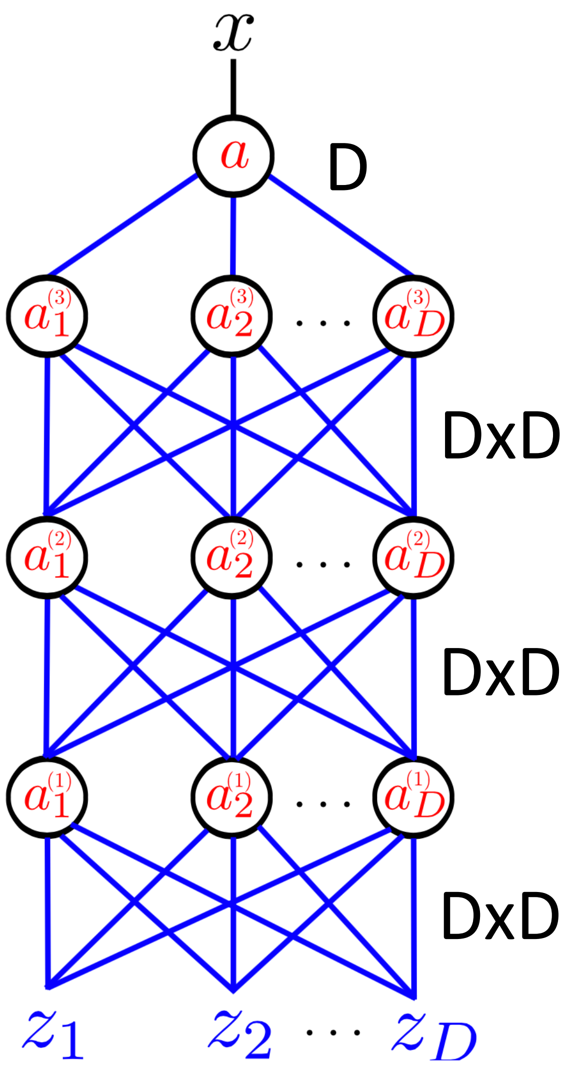
How many parameters does this network have?

Number of Parameters = 3 x (DxD) + D

For a small D = 32 x 32 = 1024 MNIST image:

Number of Parameters = 3 x (1024x1024) + 1024

~ $3x10^6$



Architecture of LeNet-5, Convolution Neural Network

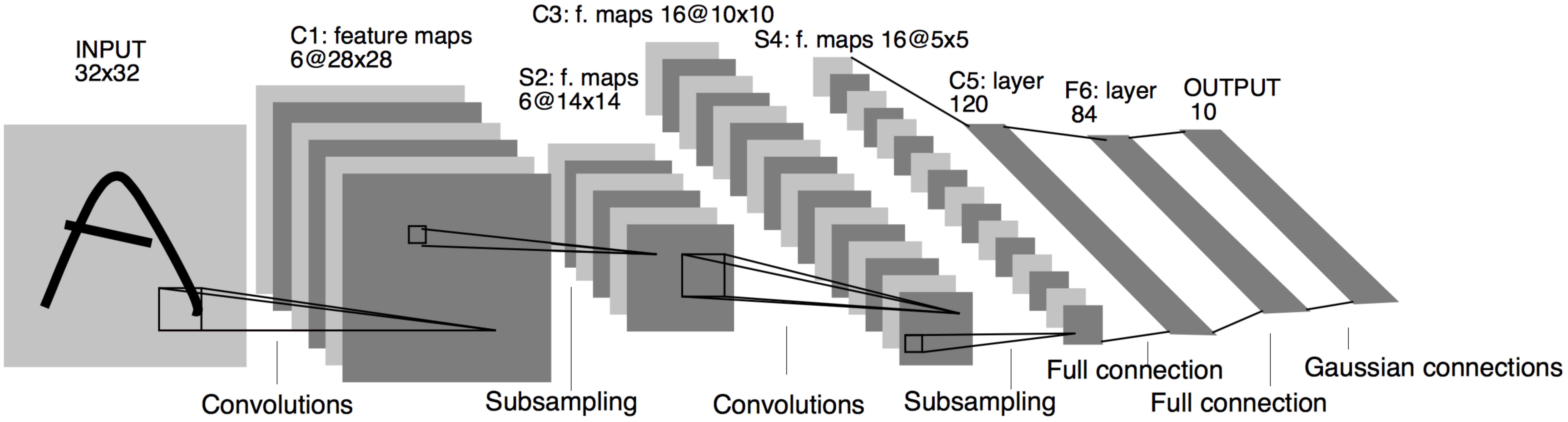
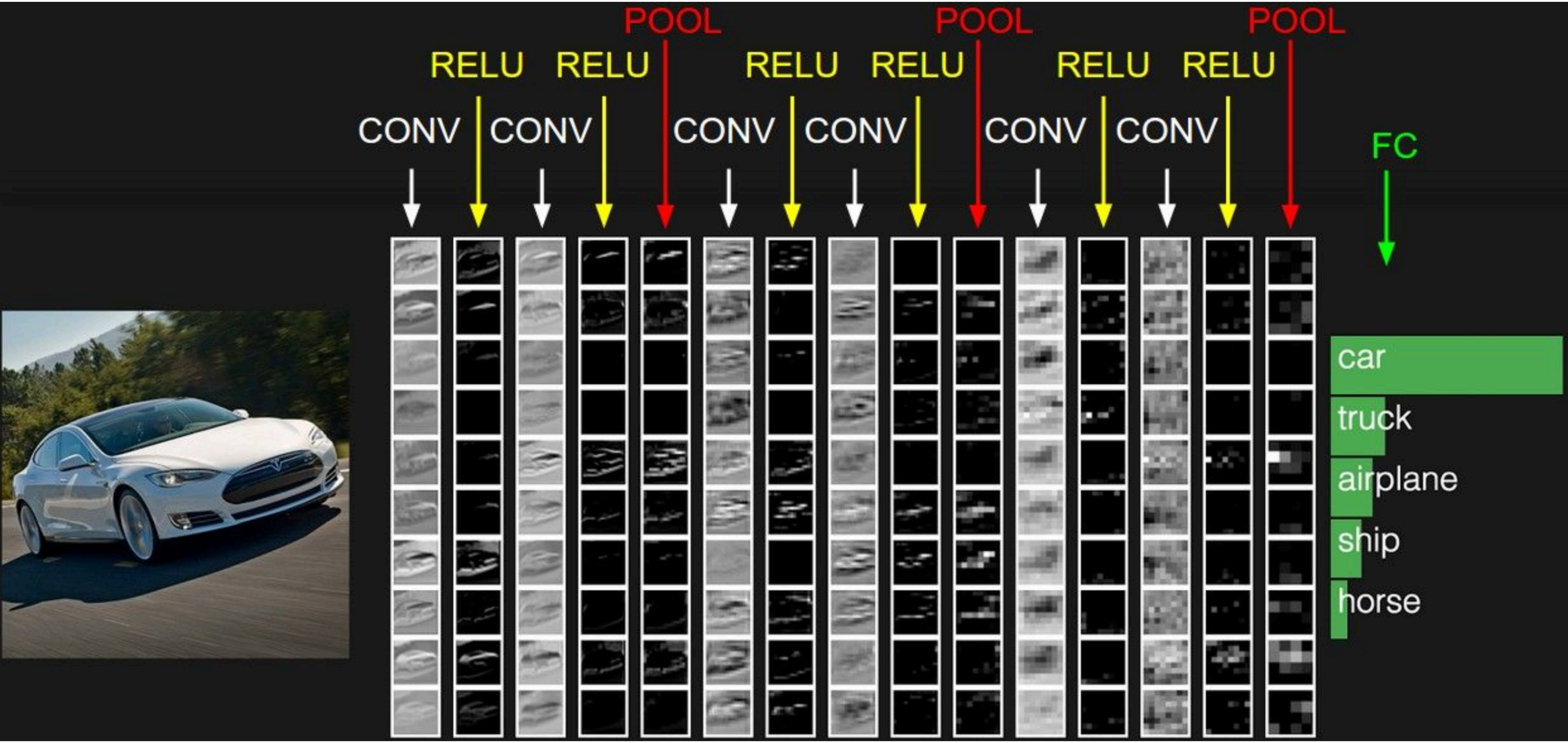


Fig. 2. Architecture of LeNet-5, a Convolutional Neural Network, here for digits recognition. Each plane is a feature map, i.e. a set of units whose weights are constrained to be identical.

Proc. Of the IEEE, November 1998, "Gradient-Based Learning Applied to Document Recognition"


POOL
POOL
POOL
RELU
RELU
RELU
RELU
RELU
RELU
CONV
CONV
CONV
CONV
CONV
CONV
FC
car
truck
airplane
ship
horse

Review: What is convolution?

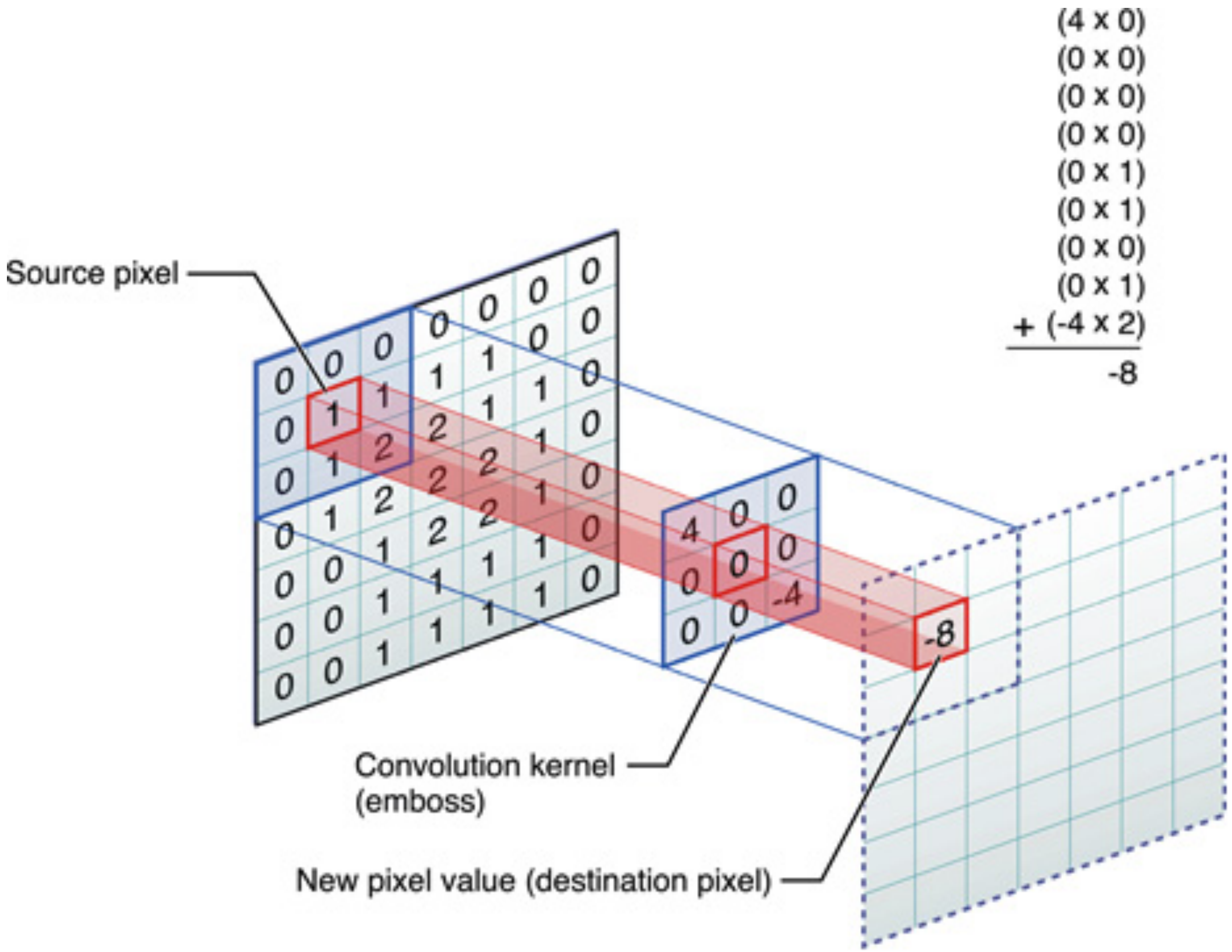
- Convolution is an important operation from signal processing
- A convolution is an integral that expresses the amount of overlap of one function as it is shifted over another function .

$$f * g = \int_{-\infty}^{\infty} f(\tau)\, g(t-\tau)\, d\tau = \int_{-\infty}^{\infty} g(\tau)\, f(t-\tau)\, d\tau$$

- 2 Dimensional Discrete Function (Image)

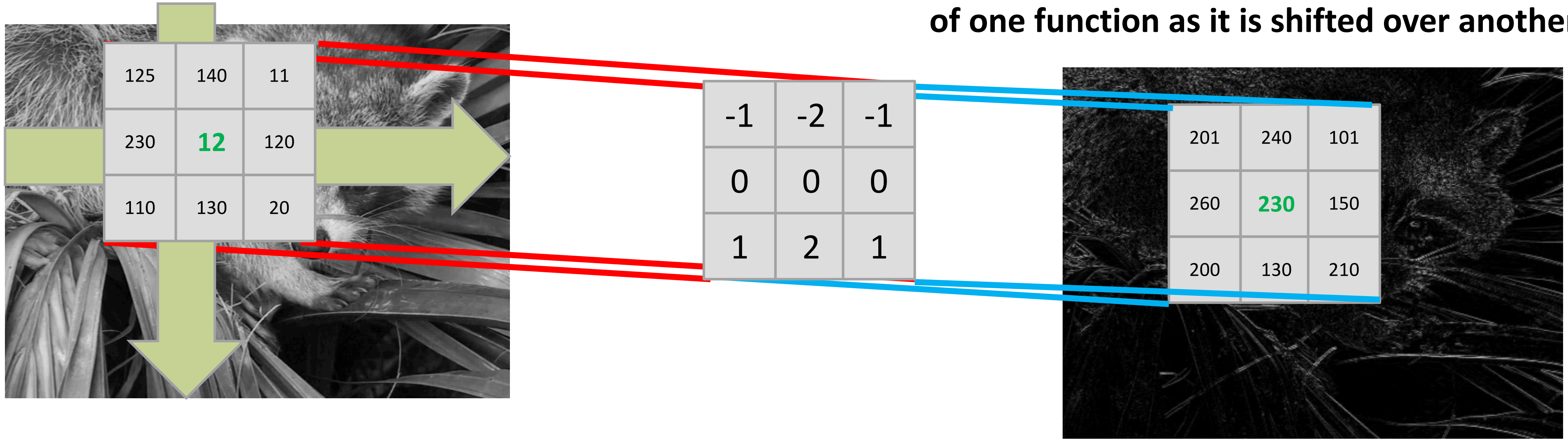
$$f[x,y] * g[x,y] = \sum_{n_1=-\infty}^{\infty} \sum_{n_2=-\infty}^{\infty} f[n_1,n_2] \cdot g[x-n_1, y-n_2]$$

2-Dimensional Convolution



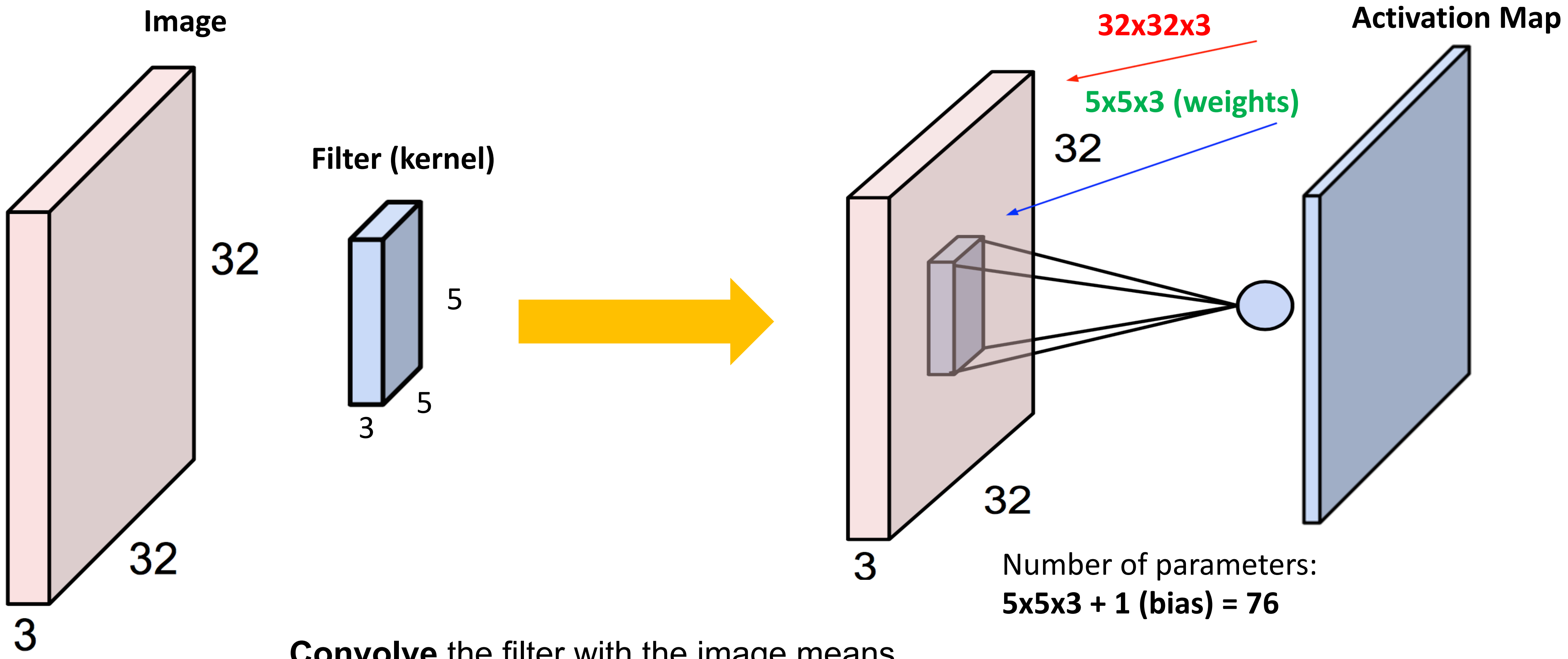
Example: 2-Dimensional Convolution

A convolution is an integral (discrete signals :Matrix Dot Product) that expresses the amount of overlap of one function as it is shifted over another function



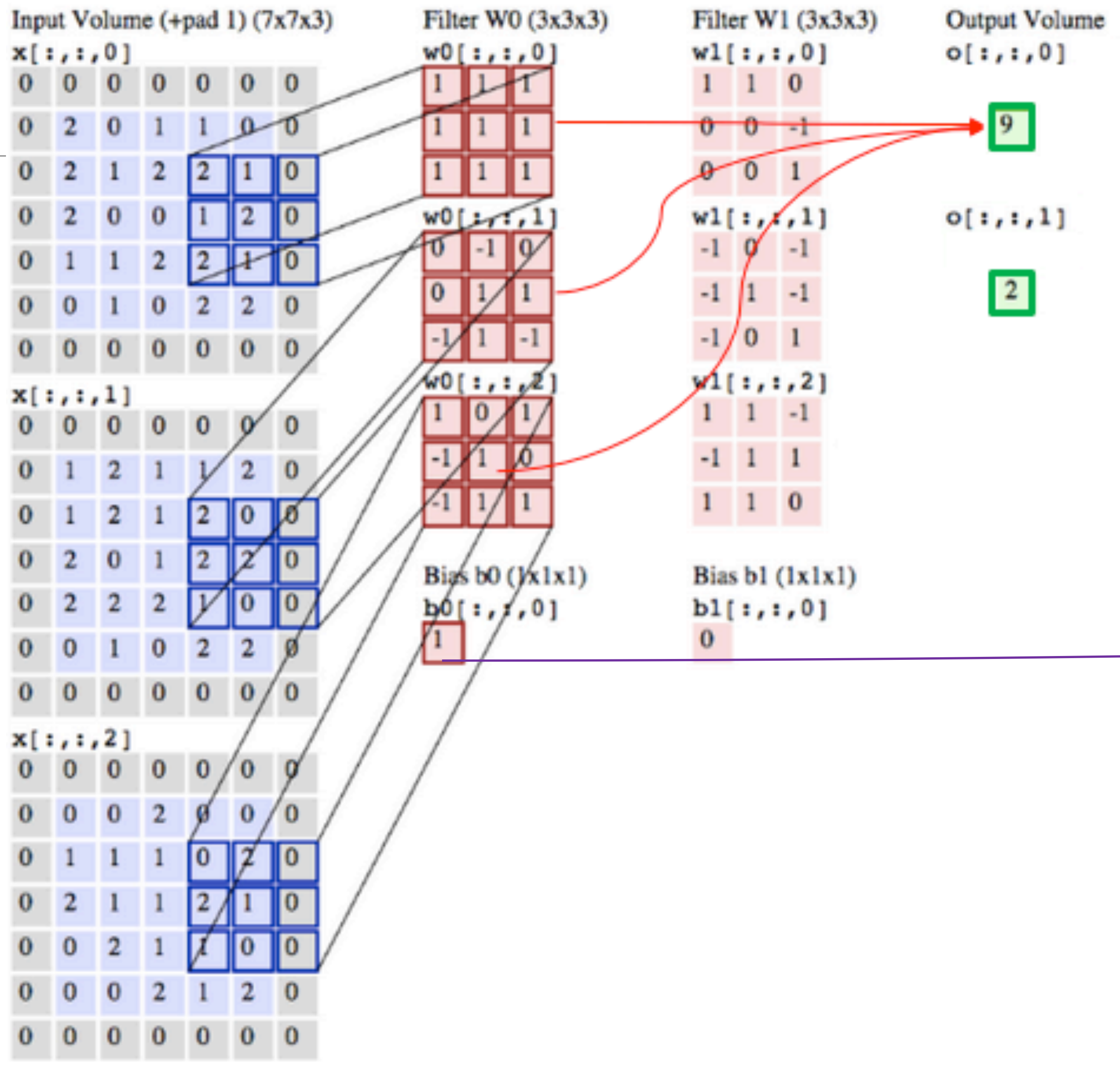
-1	0	1
-2	0	2
-1	0	1

Convolution Layer



Number of parameters:
5x5x3 + 1 (bias) = 76

Convolve the filter with the image means
"slide over the image spatially, compute dot products"



(2x1)+(1x1)+0+(1x1)+(2x1)+0+(2x1)+(1x1)+0 = 9

0+0+0+0+(2x1)+0+(1x-1)+0+0 = 1

0+0+0+(2x-1)+(1x1)+0+(1x-1)+0+0 = -2

1 + 1 + (-2) + 9 = $\boxed{9}$

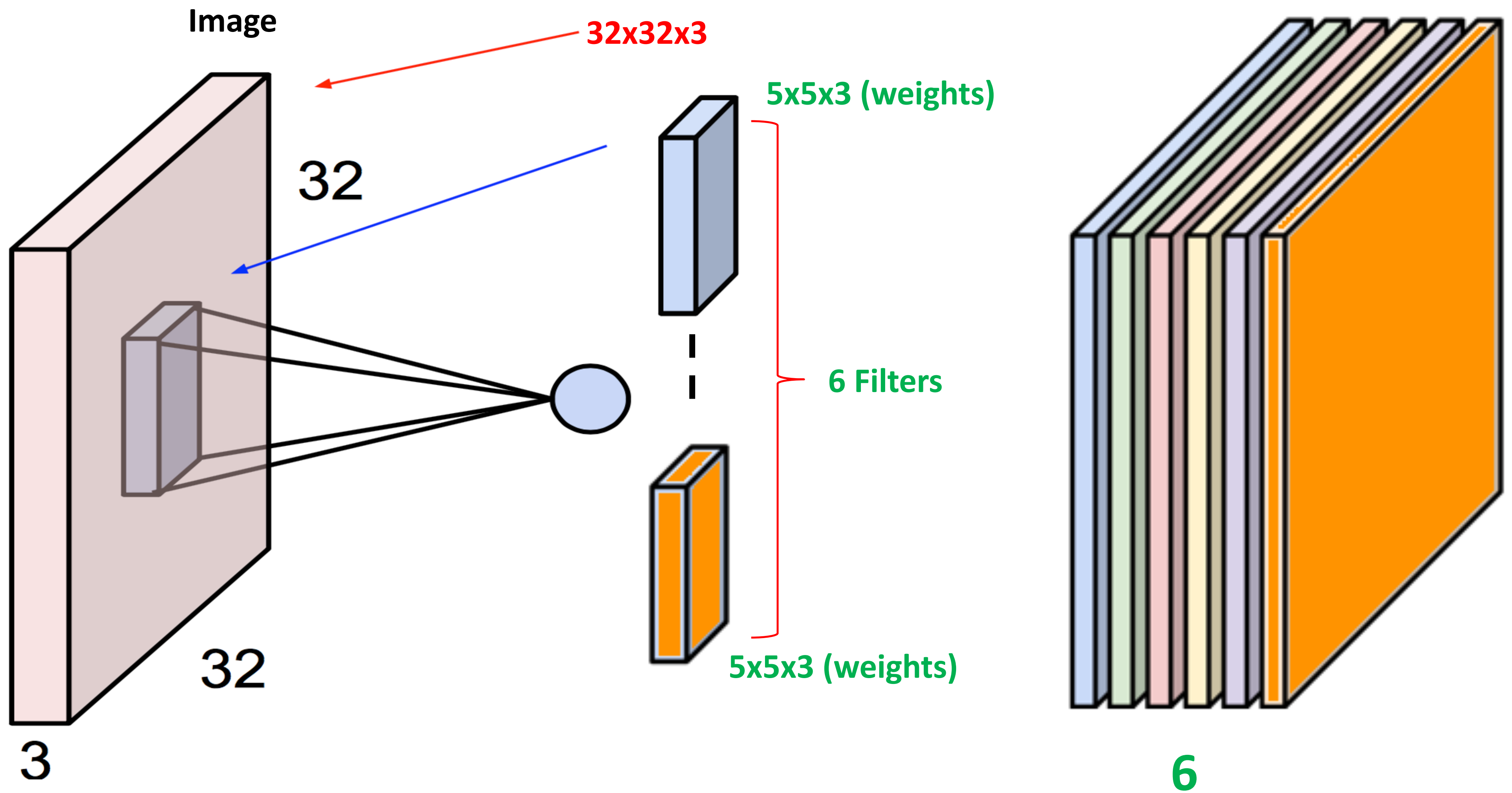
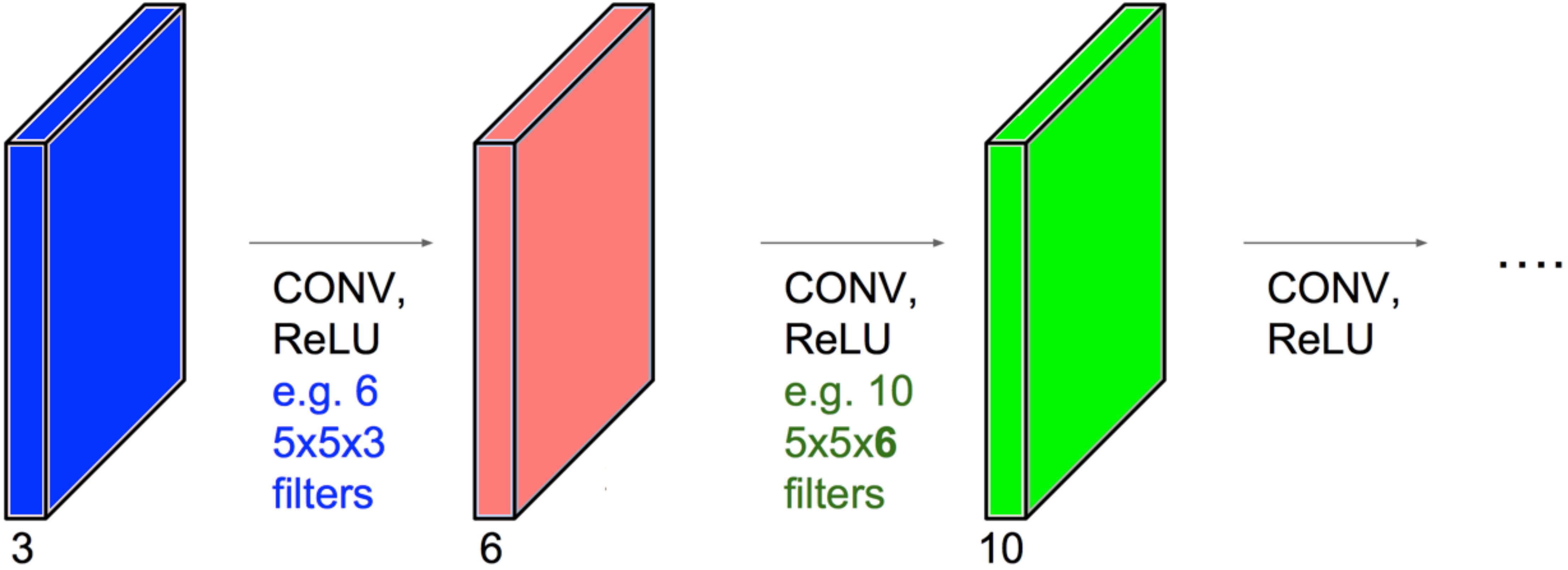
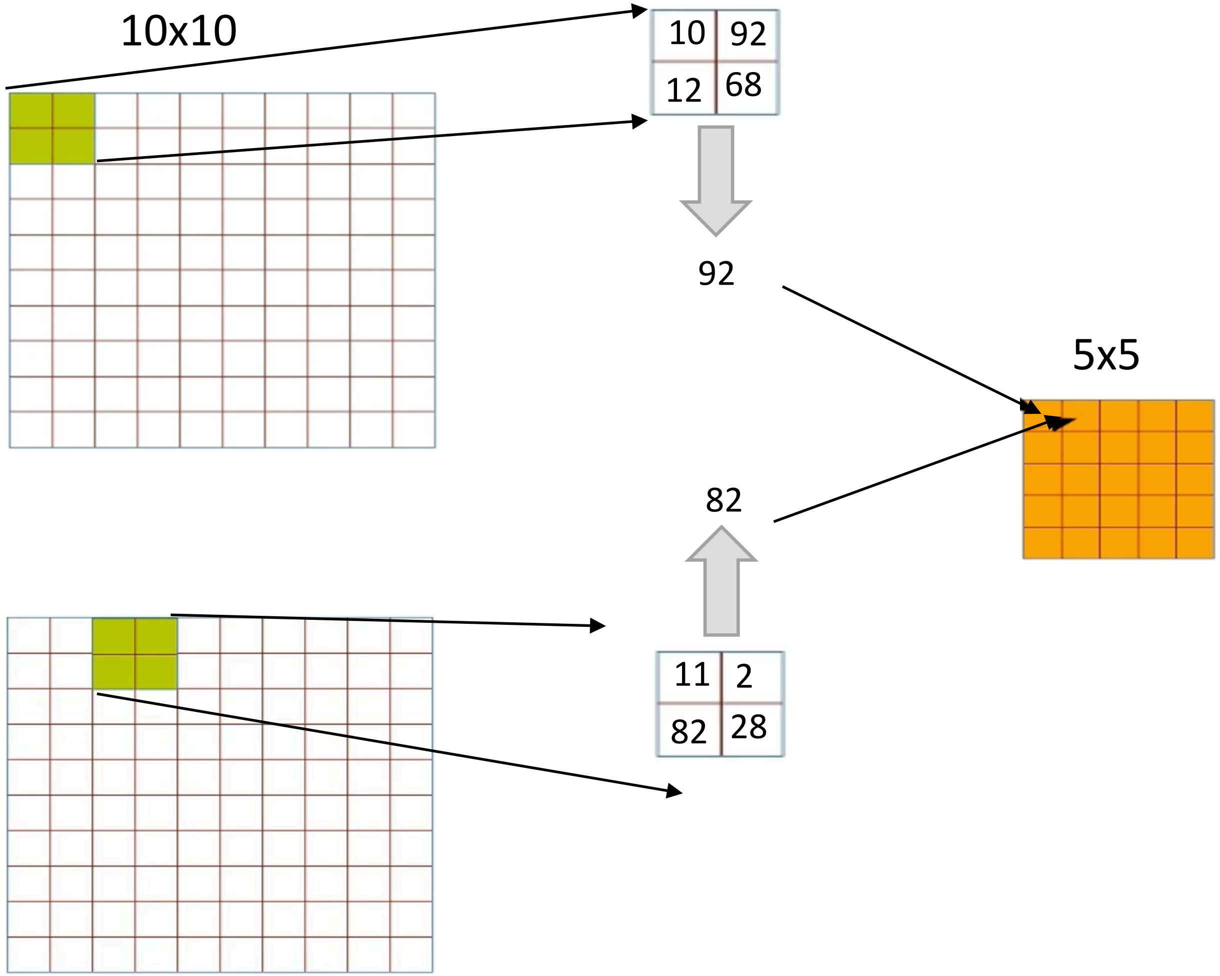


Image
32x32x3
5x5x3 (weights)
32
6 Filters
32
5x5x3 (weights)
3
6

Convolution Network



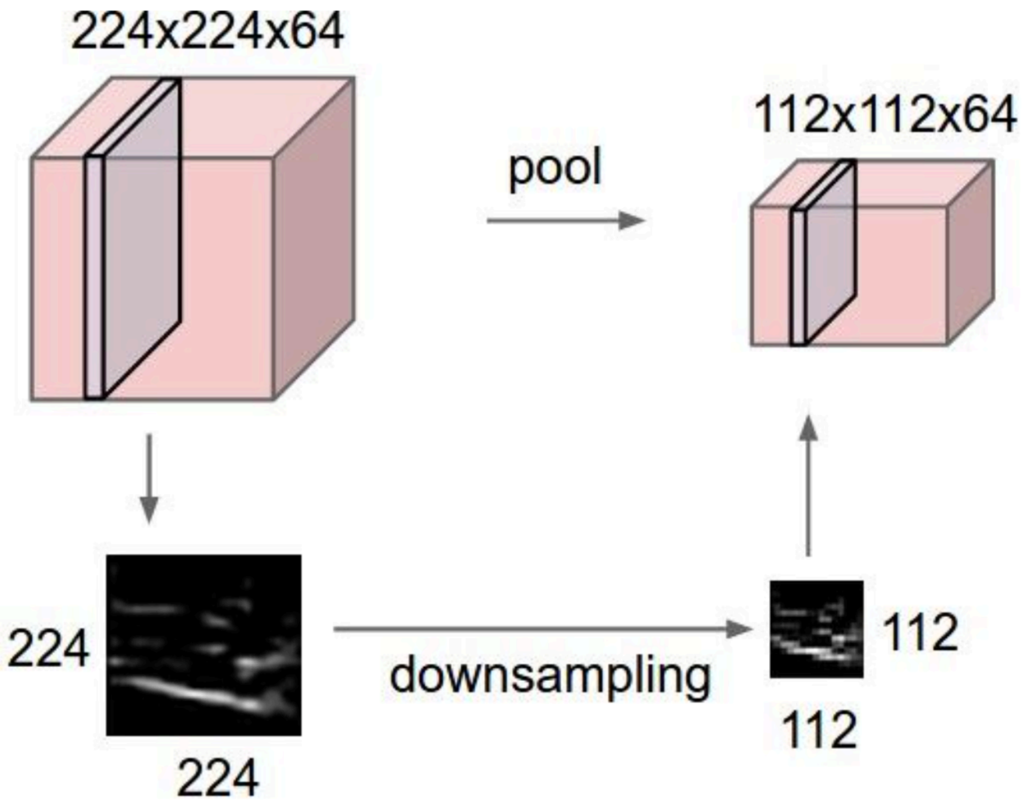
Downsampling = Max Pooling

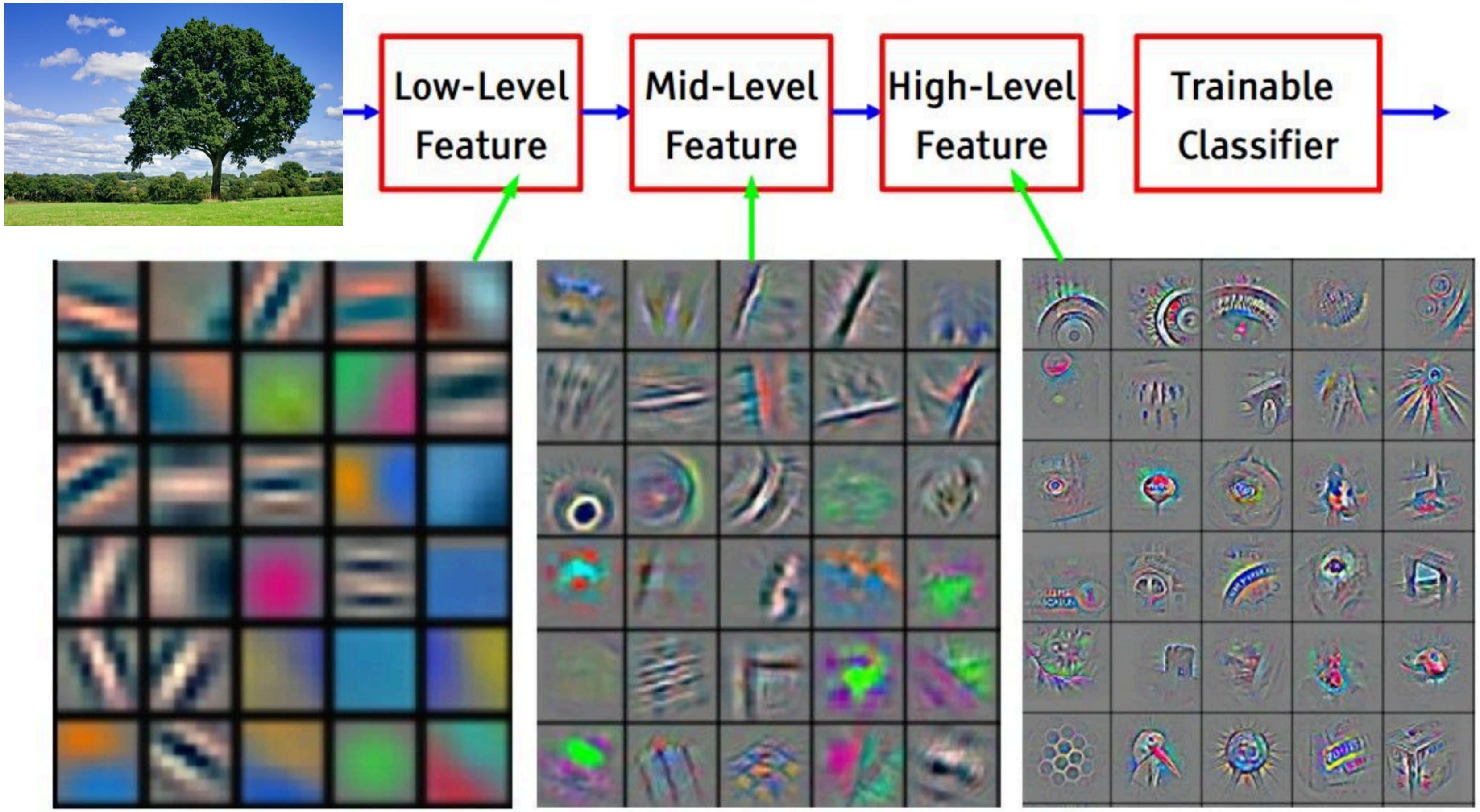


“Max Pooling” Layers to extract the “best” local feature

Pooling Layer

- Makes the representations smaller and more manageable
- Operates over each activation map independently





Feature visualization of convolution net trained on ImageNet from [Zeiler & Fergus 2013]