Probabilistic Machine Learning and AI

Outline of the lecture

This lecture introduces you to the fascinating subject of classification and regression with convolutional neural networks.

Types of Learning

- **Supervised (inductive) learning**
  - Training data includes desired outputs
- **Unsupervised learning**
  - Training data does not include desired outputs
- **Semi-supervised learning**
  - Training data includes a few desired outputs
- **Reinforcement learning**
  - Rewards from sequence of actions

The machine learning framework

- Apply a prediction function to a **feature representation** of the image to get the desired output:

h() = "apple"

h() = "cow"

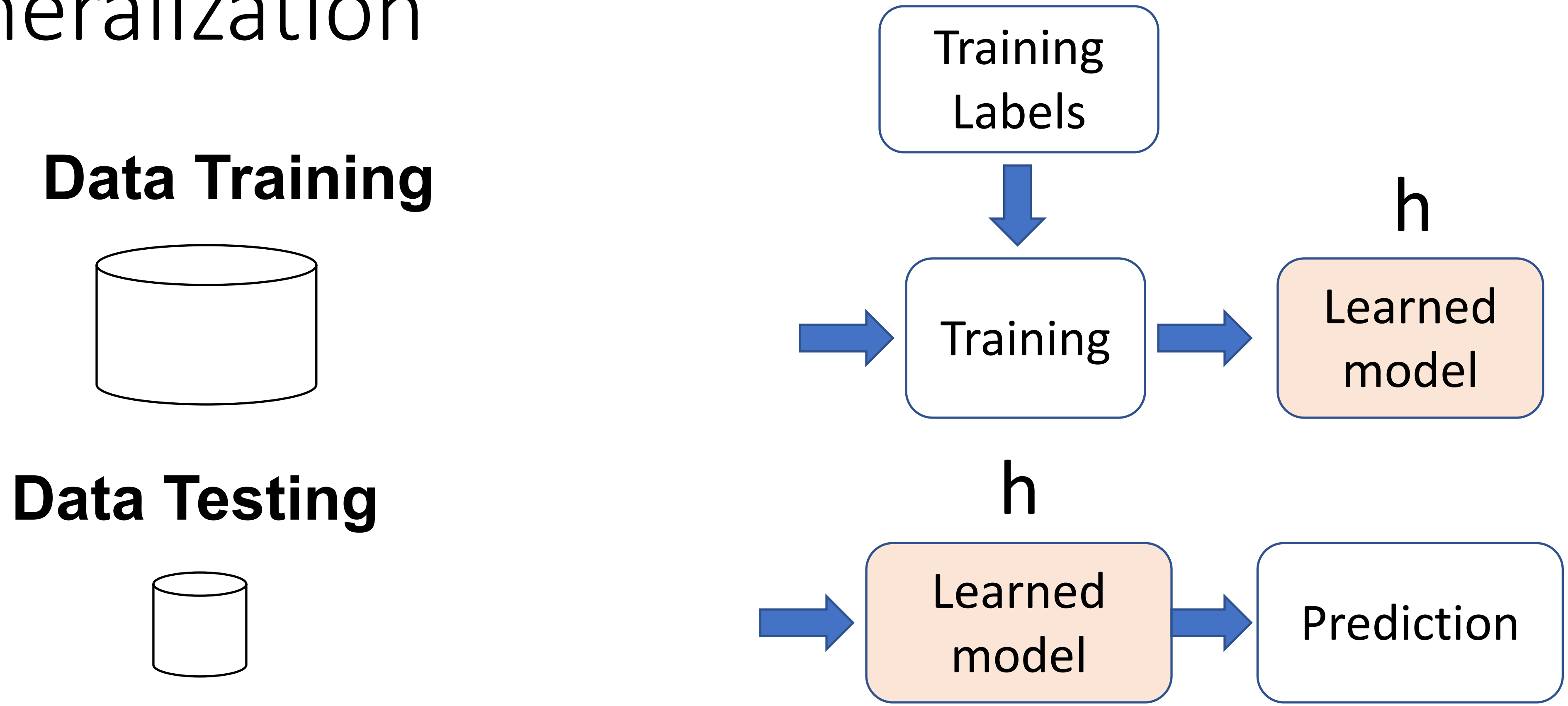
The machine learning framework

$$y = h(\mathbf{x})$$

output   prediction function   Image feature

- **Training:** given a *training set* of labeled examples $\{(\mathbf{x}_1,y_1), \ldots, (\mathbf{x}_N,y_N)\}$, estimate the prediction function h by minimizing the prediction error on the training set
- **Testing:** apply h to a never before seen *test example* **x** and output the predicted value $y = h(\mathbf{x})$

Generalization



How well does a learned model generalize from the data it was trained on to a new test set?

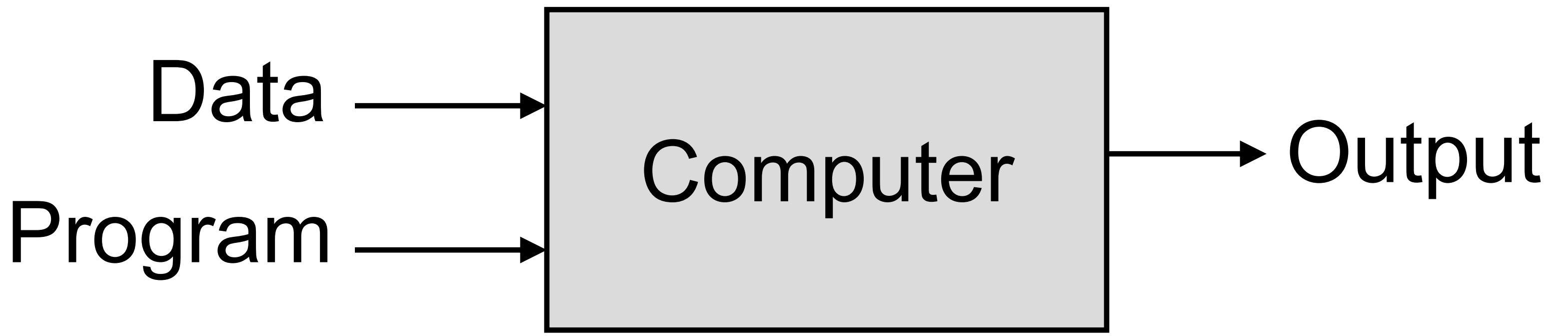
Generalization

- Components of generalization error
  - **Bias:** how much the average model over all training sets differ from the true model?
    - Error due to inaccurate assumptions/simplifications made by the model
  - **Variance:** how much models estimated from different training sets differ from each other

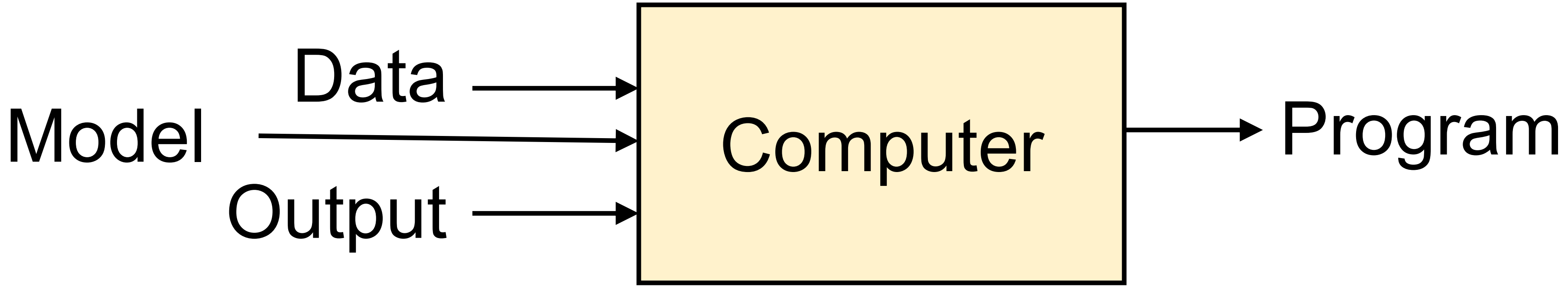
- **Underfitting:** model is too "simple" to represent all the relevant class characteristics
  - High bias and low variance
  - High training error and high test error

- **Overfitting:** model is too "complex" and fits irrelevant characteristics (noise) in the data
  - Low bias and high variance
  - Low training error and high test error

Traditional Programming

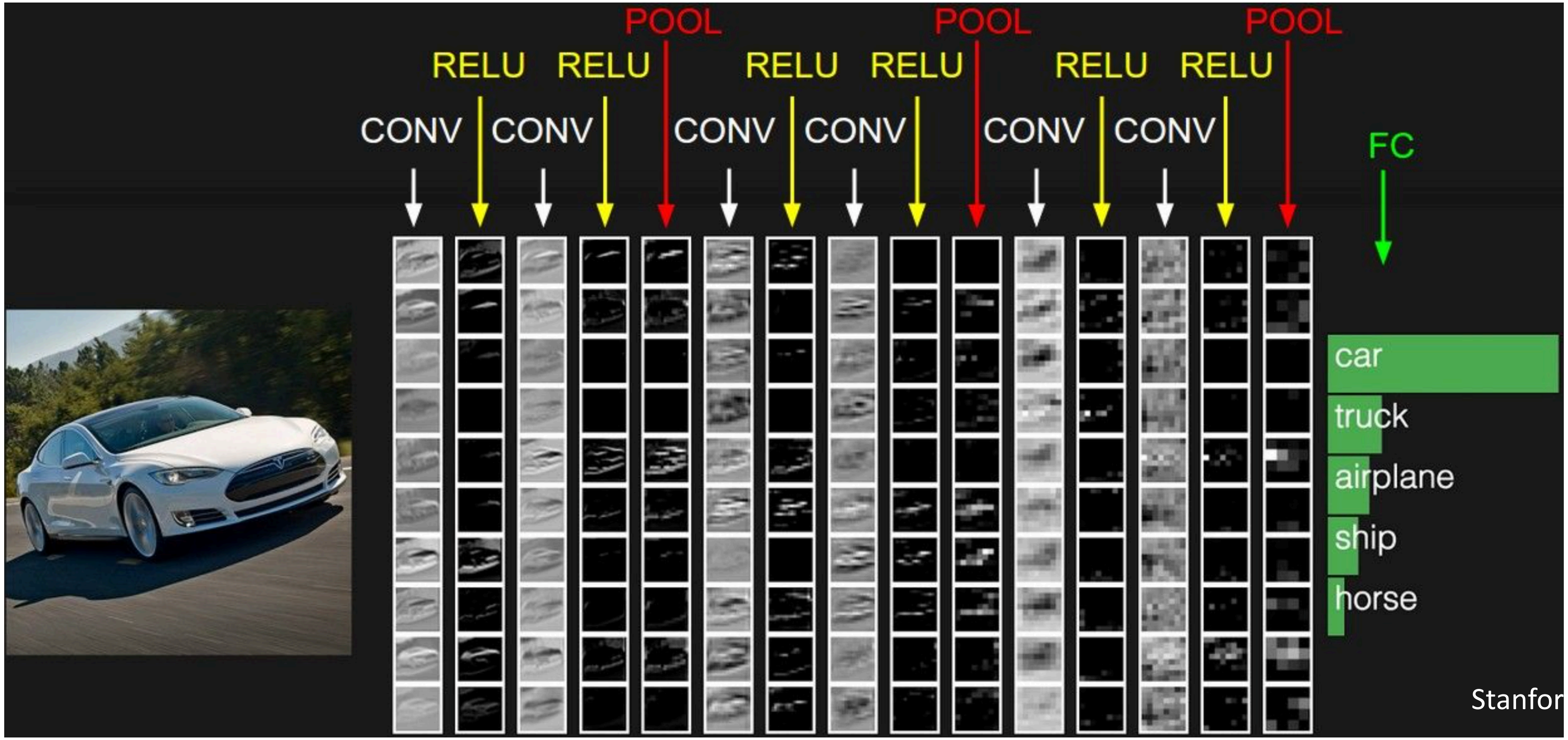


Machine Learning



Convolution Neural Network (CNN)

Neural Networks that share their parameters across space.



What is convolution?

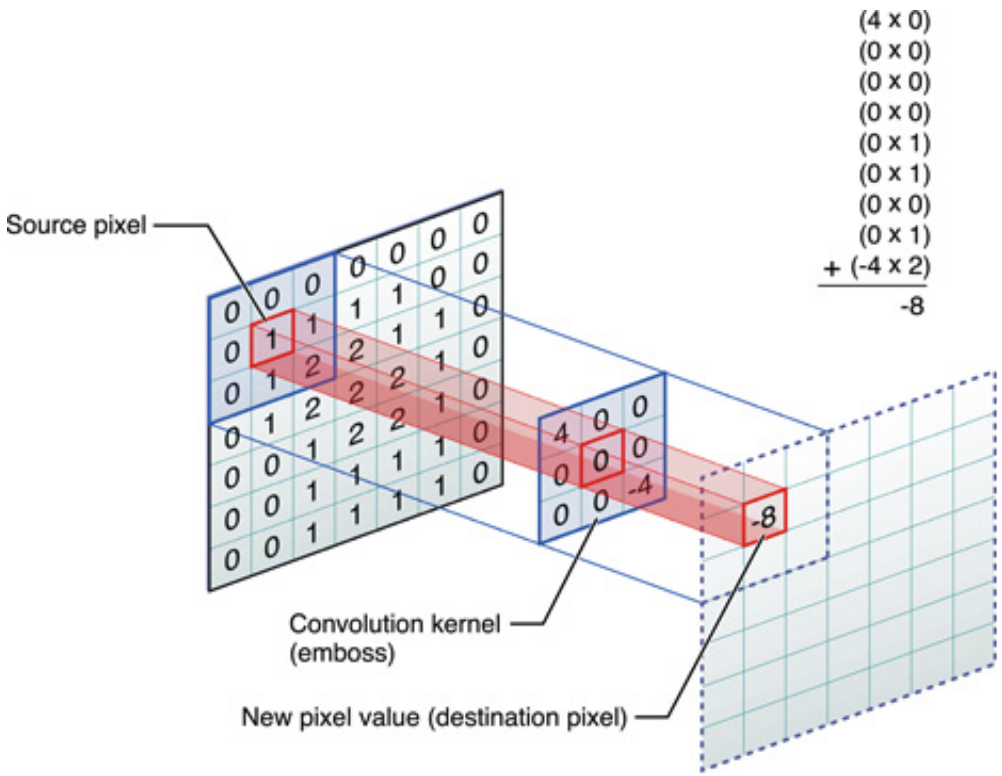
- Convolution is an important operation from signal processing
- A convolution is an integral that expresses the amount of overlap of one function as it is shifted over another function

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

- 2-Dimensional Discrete Function (Image)

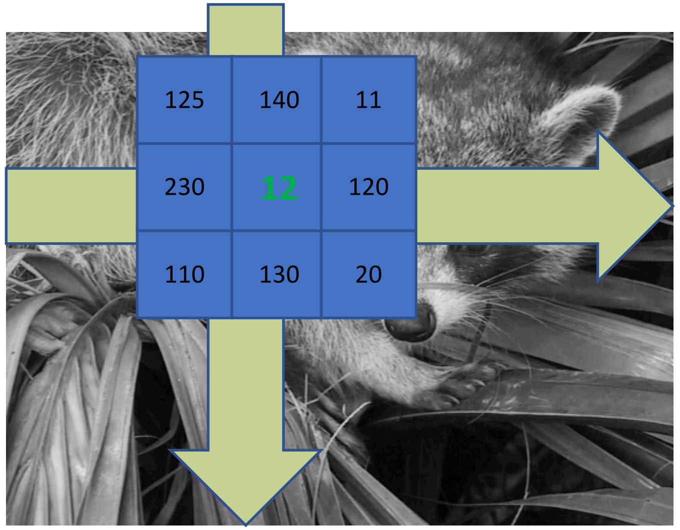
$$f[x,y] * g[x,y] = \sum_{n_1=-\infty}^{\infty} \sum_{n_2=-\infty}^{\infty} f[n_1,n_2] \cdot g[x-n_1, y-n_2]$$

2-Dimensional Convolution

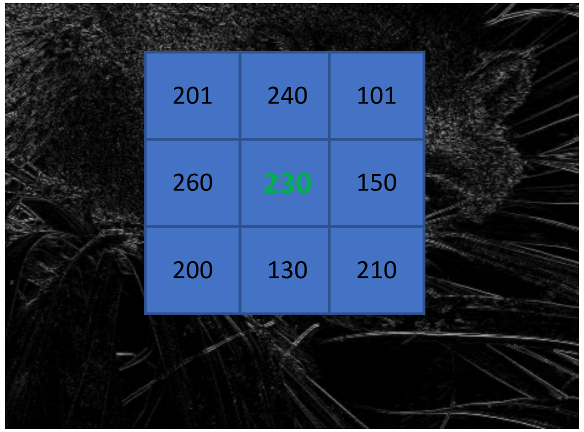


Example: 2-Dimensional Convolution

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

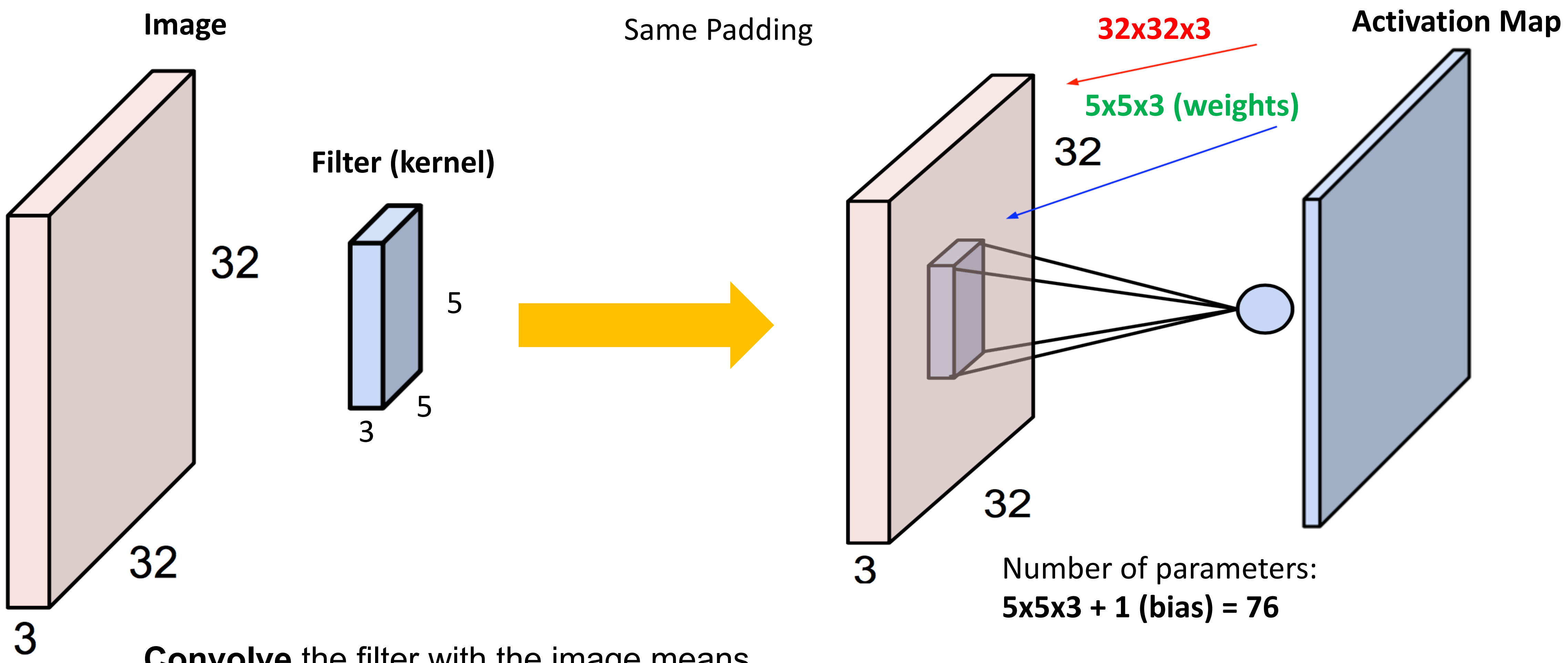


-1	-2	-1
0	0	0
1	2	1

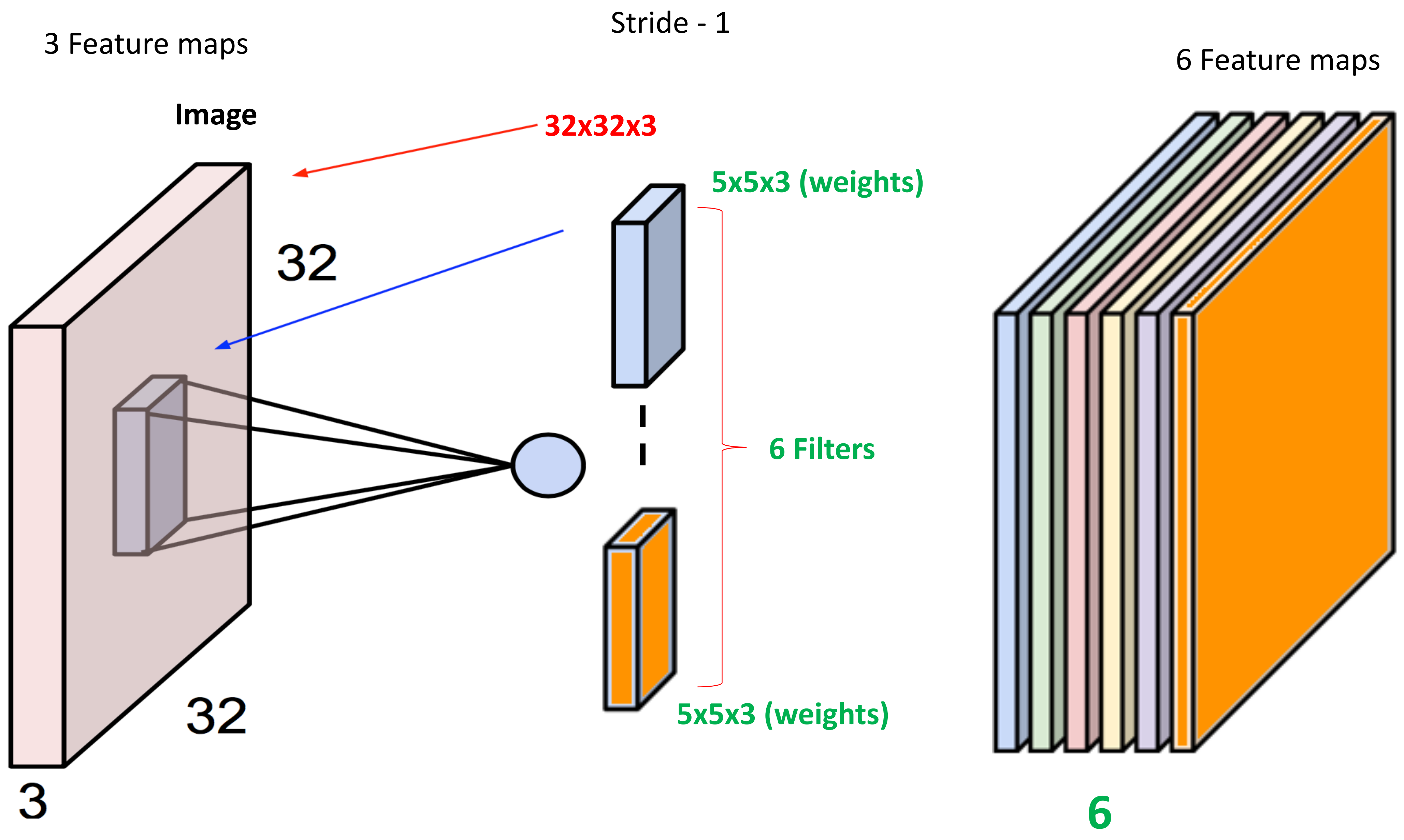


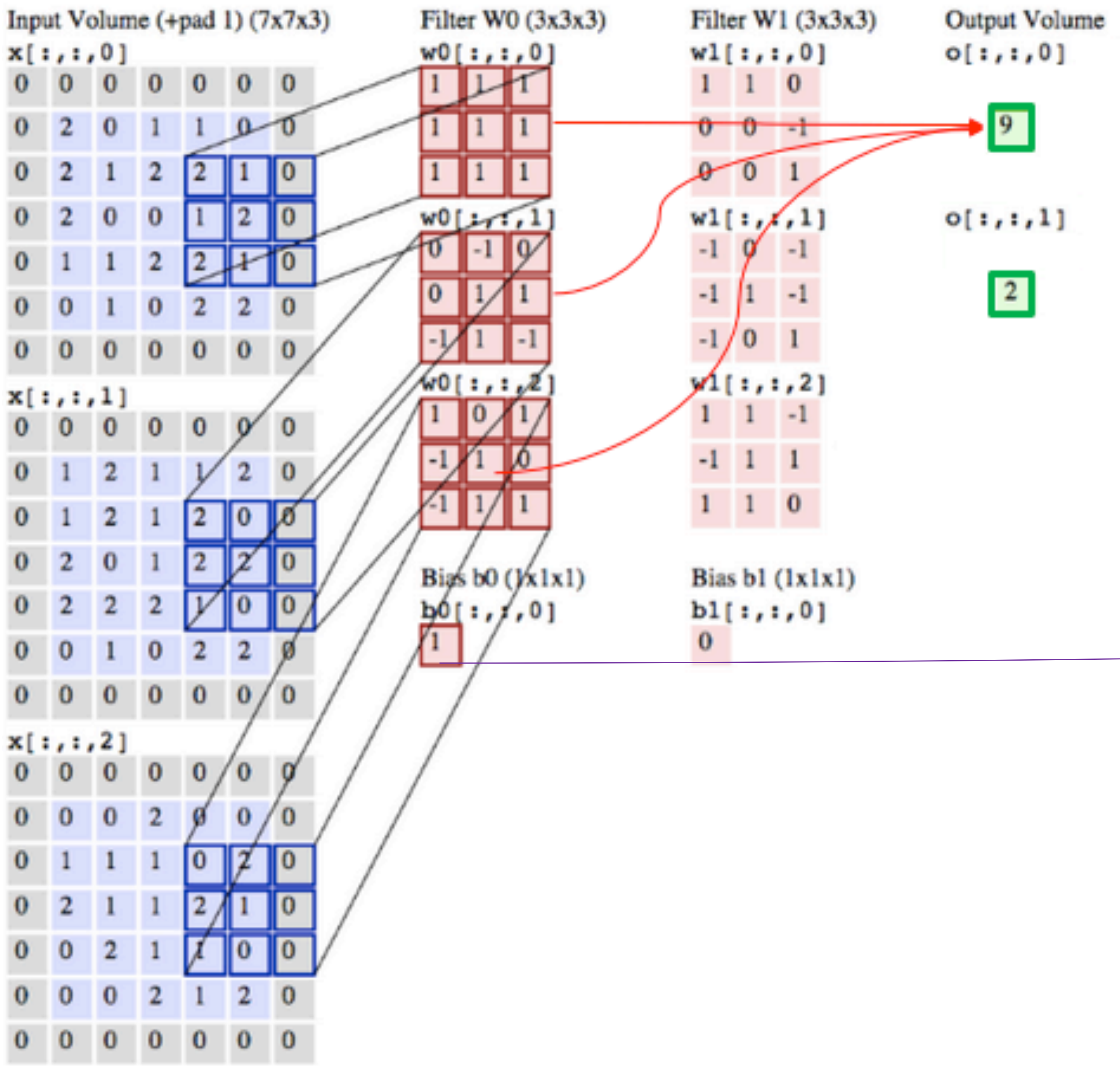
-1	0	1
-2	0	2
-1	0	1

Convolution Layer



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


3 Feature maps
Stride - 1
6 Feature maps
Image
32x32x3
5x5x3 (weights)
32
6 Filters
32
5x5x3 (weights)
3
6



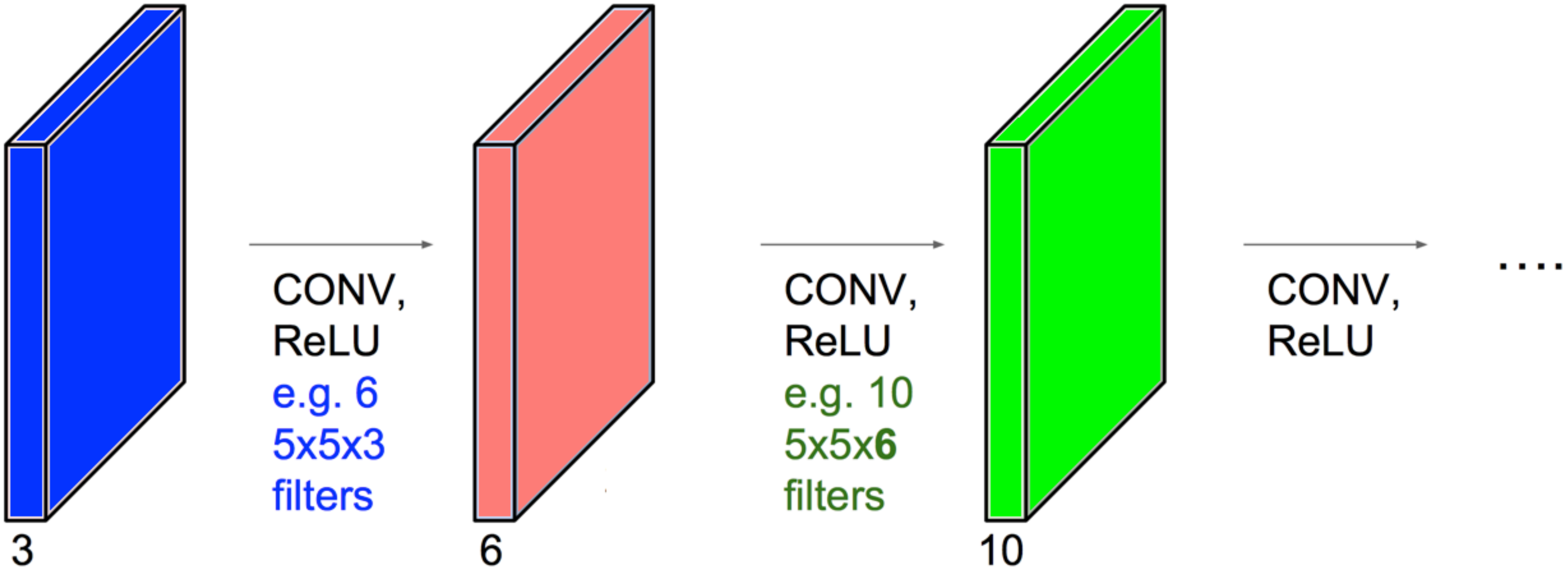
(2x1)+(1x1)+0+(1x1)+(2x1)+0+(2x1)+(1x1)+0 = 9

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

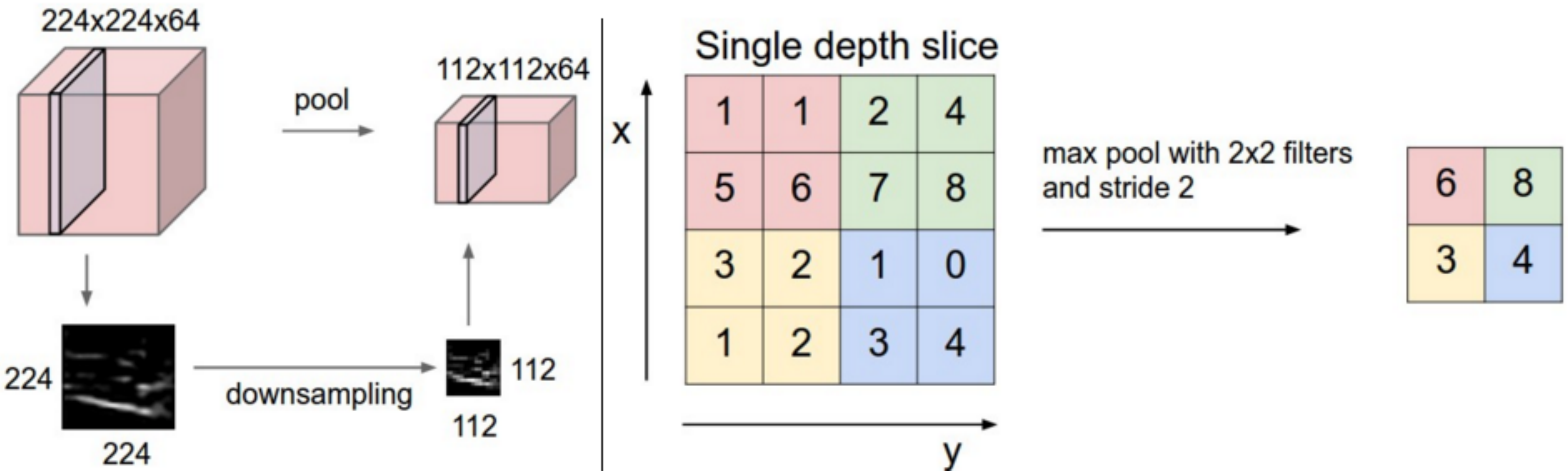
0+0+0+(2x-1)+(1x1)+0+(1x-1)+0+0 = -2

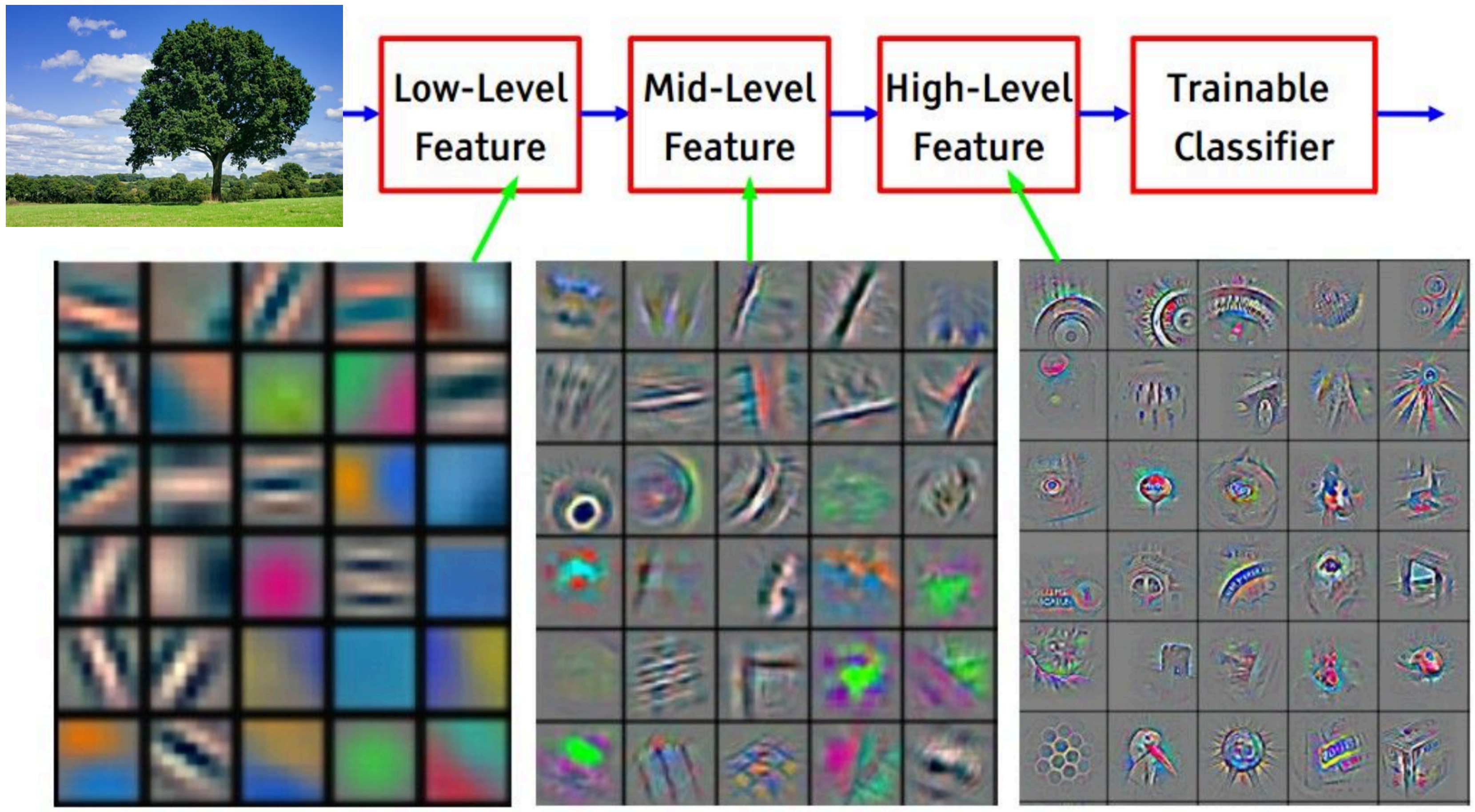
1 + 1 + (-2) + 9 = $\boxed{9}$

Convolution Network

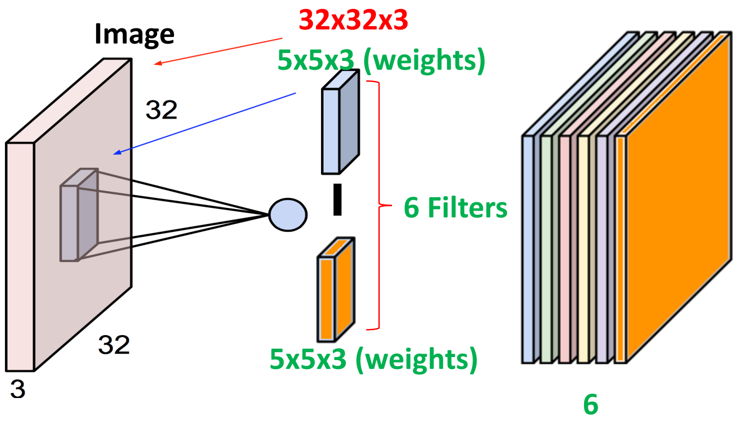


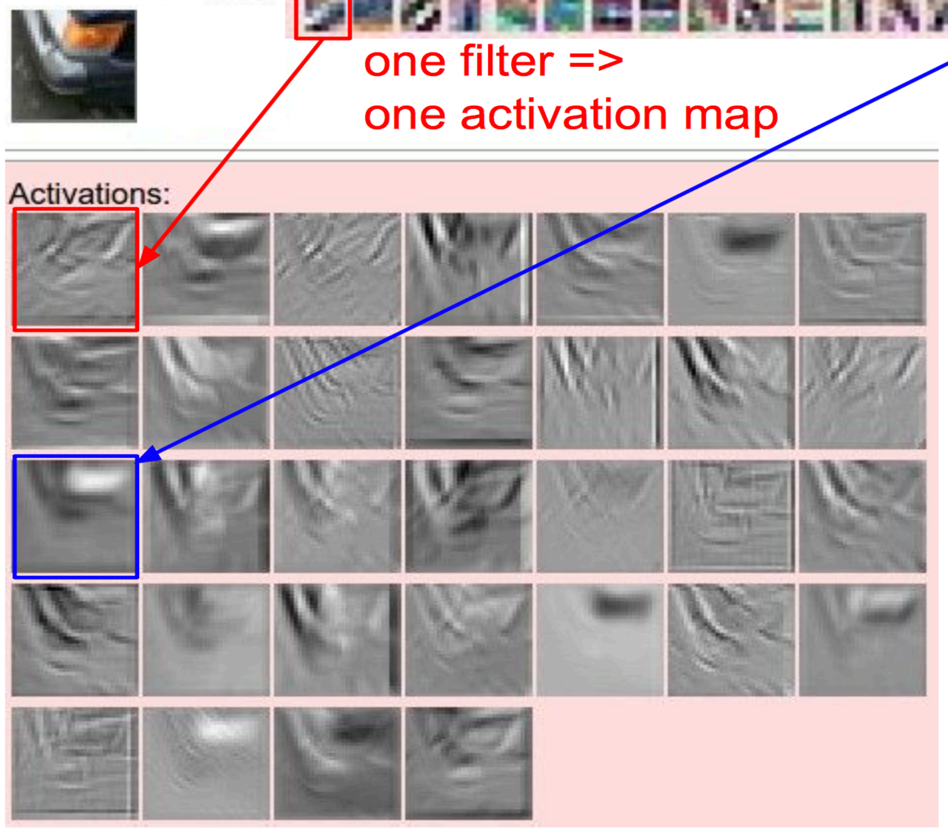
Pooling


224x224x64
pool
112x112x64
224
224
downsampling
112
112
Single depth slice
x
y
1 1 2 4
5 6 7 8
3 2 1 0
1 2 3 4
max pool with 2x2 filters
and stride 2
6 8
3 4



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





example 5x5 filters
(32 total)

We call the layer convolutional because it is related to convolution of two signals:

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

elementwise multiplication and sum of a filter and the signal (image)

Architecture of LeNet-5

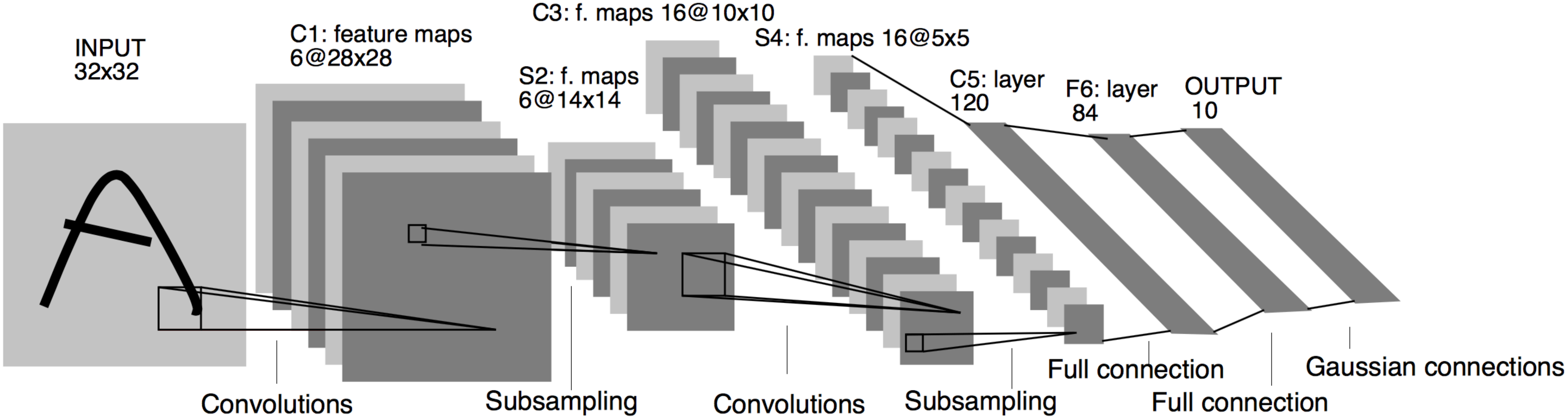


Fig. 2. Architecture of LeNet-5, a Convolutional Neural Network, here for digits recognition. Each plane is a feature map, i.e. a set of units whose weights are constrained to be identical.

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