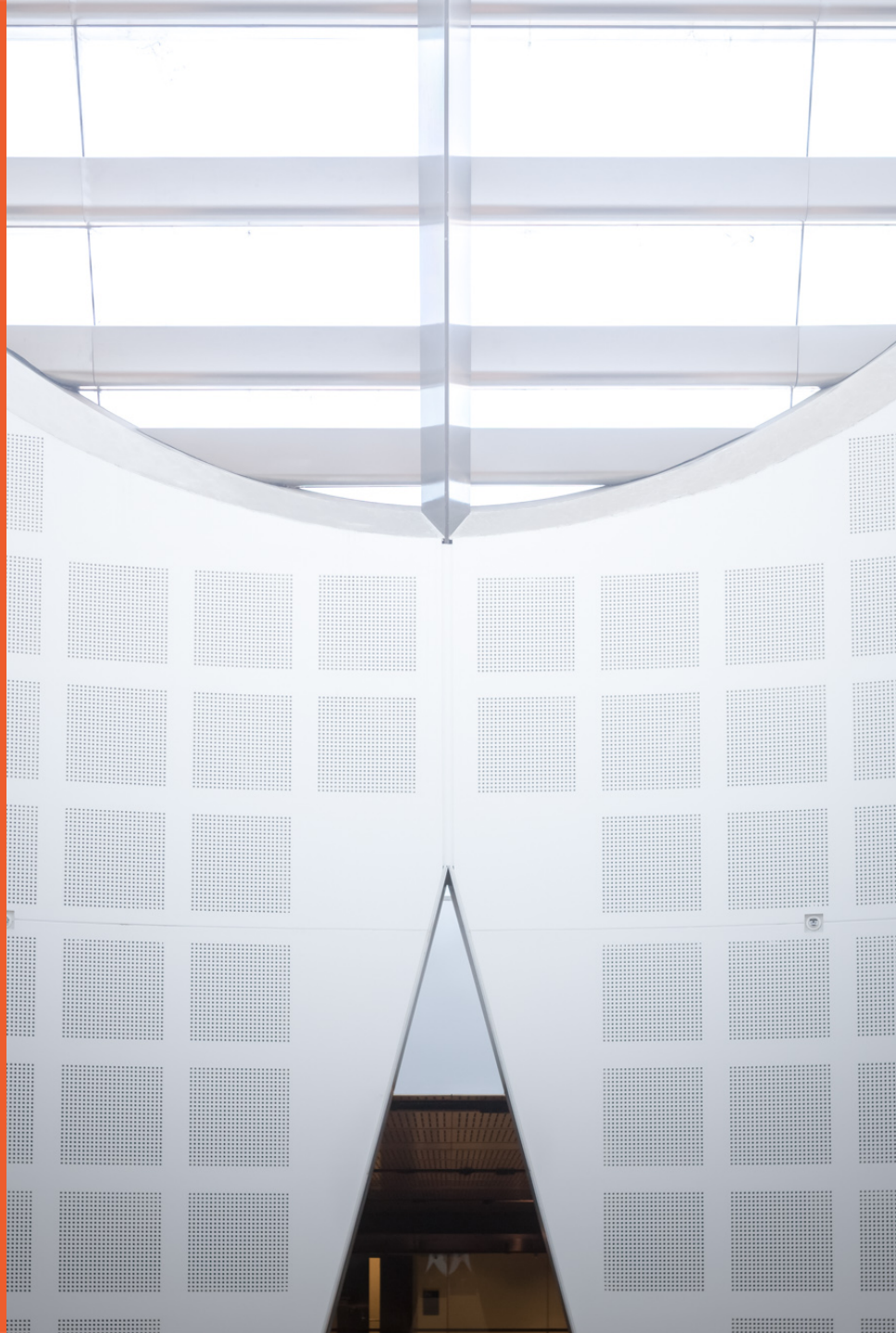
Convolutional Neural Networks

Dr Chang Xu
UBTECH Sydney AI Centre

Acknowledgements:
Chaoyue Wang

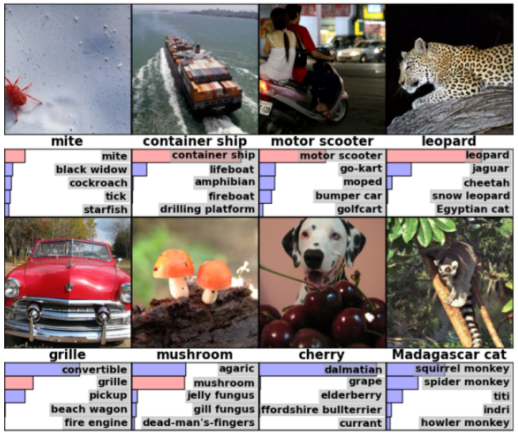
THE UNIVERSITY OF SYDNEY



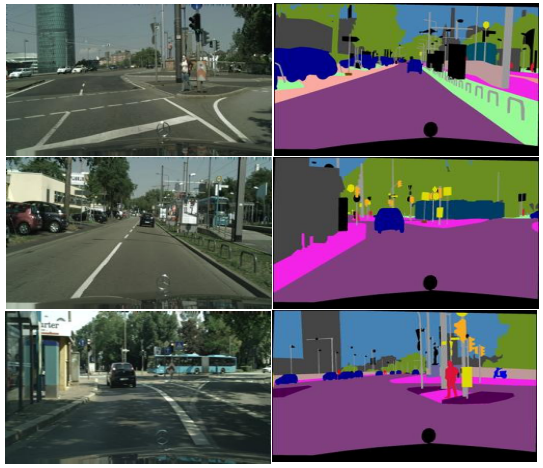
Quick Review

Today, CNNs are everywhere

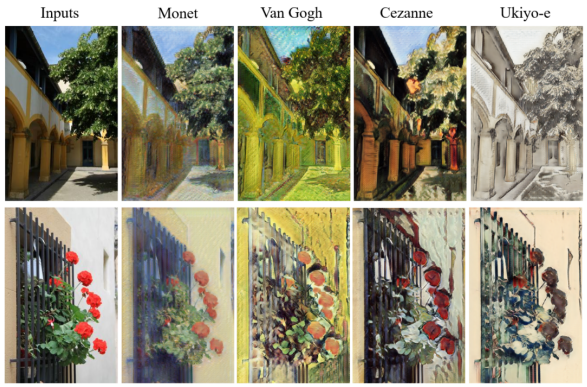
- Image classification, Image segmentation, Pose estimation, Style transfer, Image detection, Image caption …



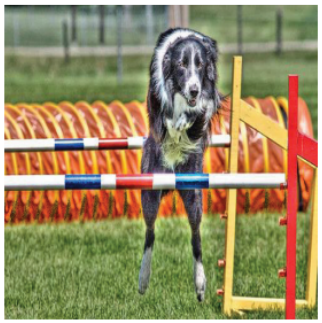
(Krizhevsky et al, 2012)



(Shaoli et al, 2017)



(Xinyuan et al, 2018)



Large black and white dog jumping hurdle

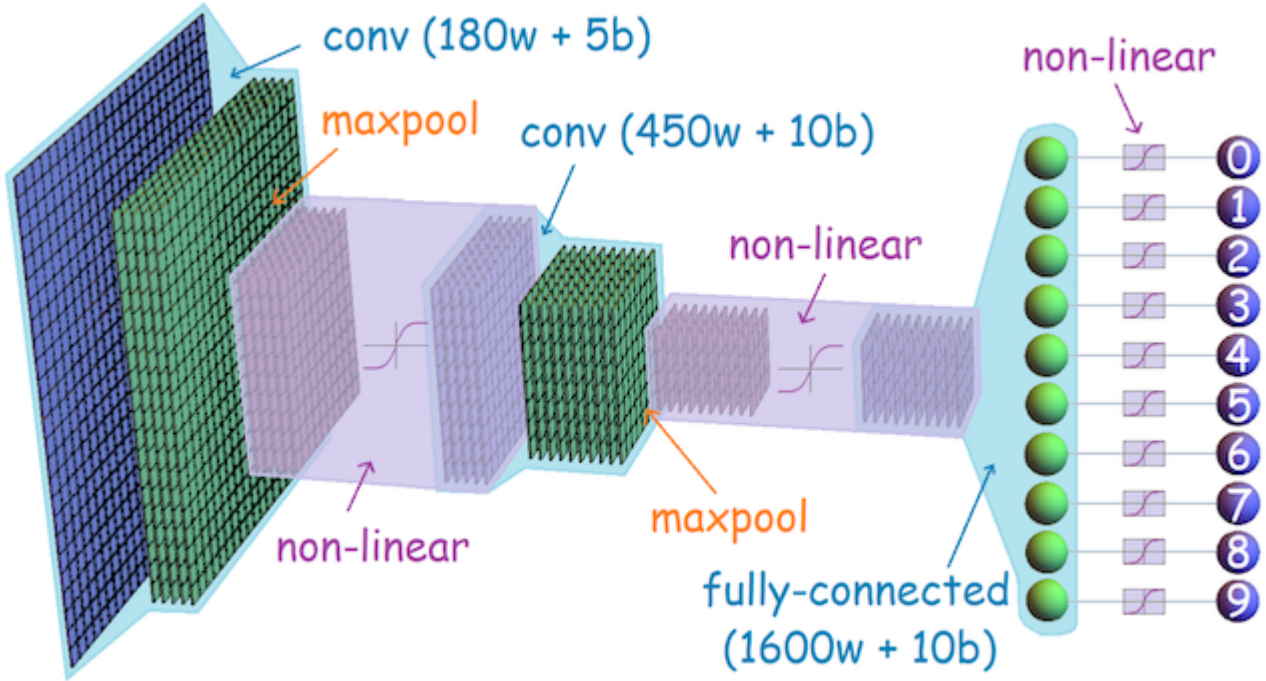
(Jianfeng et al, 2017)

A man in a red shirt is playing a saxophone outside a business

Basic CNNs Components

Basic CNNs Components

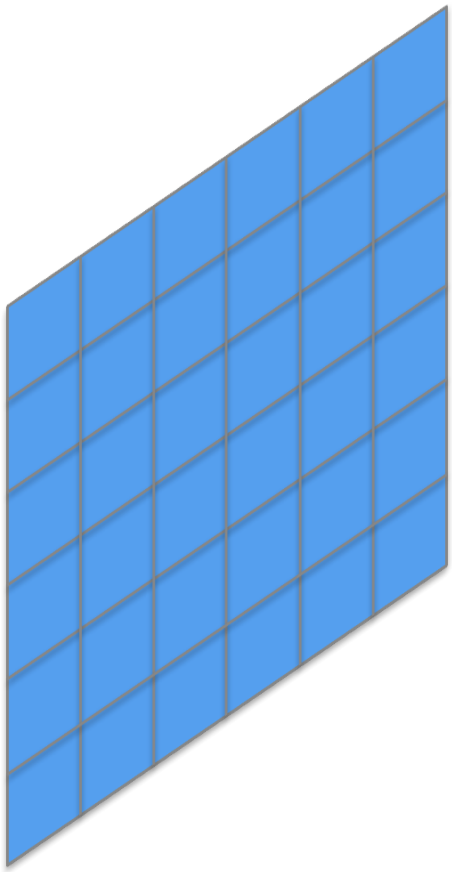
- **Convolutional Layer**
- **Pooling**
- Fully-connected Layer



(https://leonardoaraujosantos.gitbooks.io)

Convolutional Layer

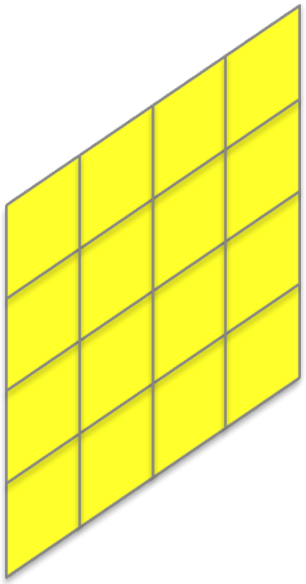
- Give a simple example: take a grayscale image as input



Grayscale Image: $\boldsymbol{X}$

$w_{0,1}$	$w_{0,2}$	$w_{0,3}$
$w_{1,1}$	$w_{1,2}$	$w_{1,3}$
$w_{2,1}$	$w_{2,2}$	$w_{2,3}$

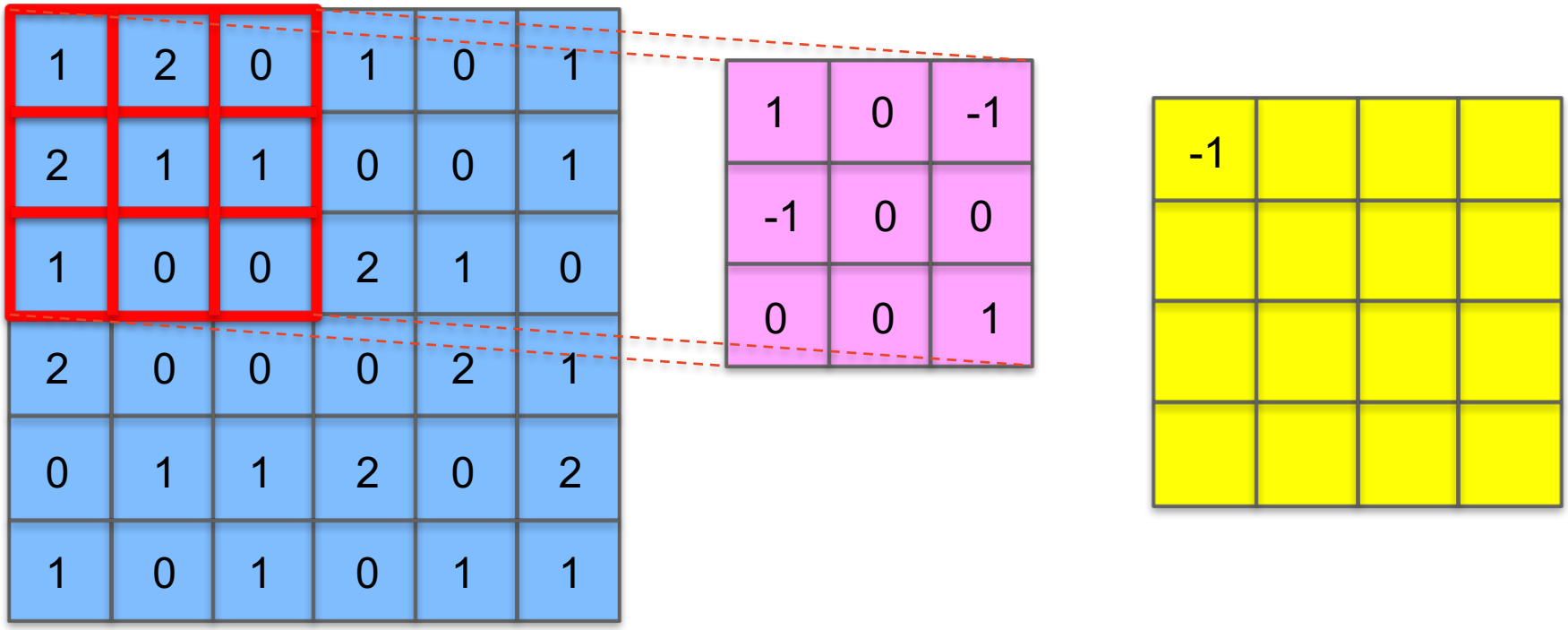
Filter: $\boldsymbol{w}$



Feature

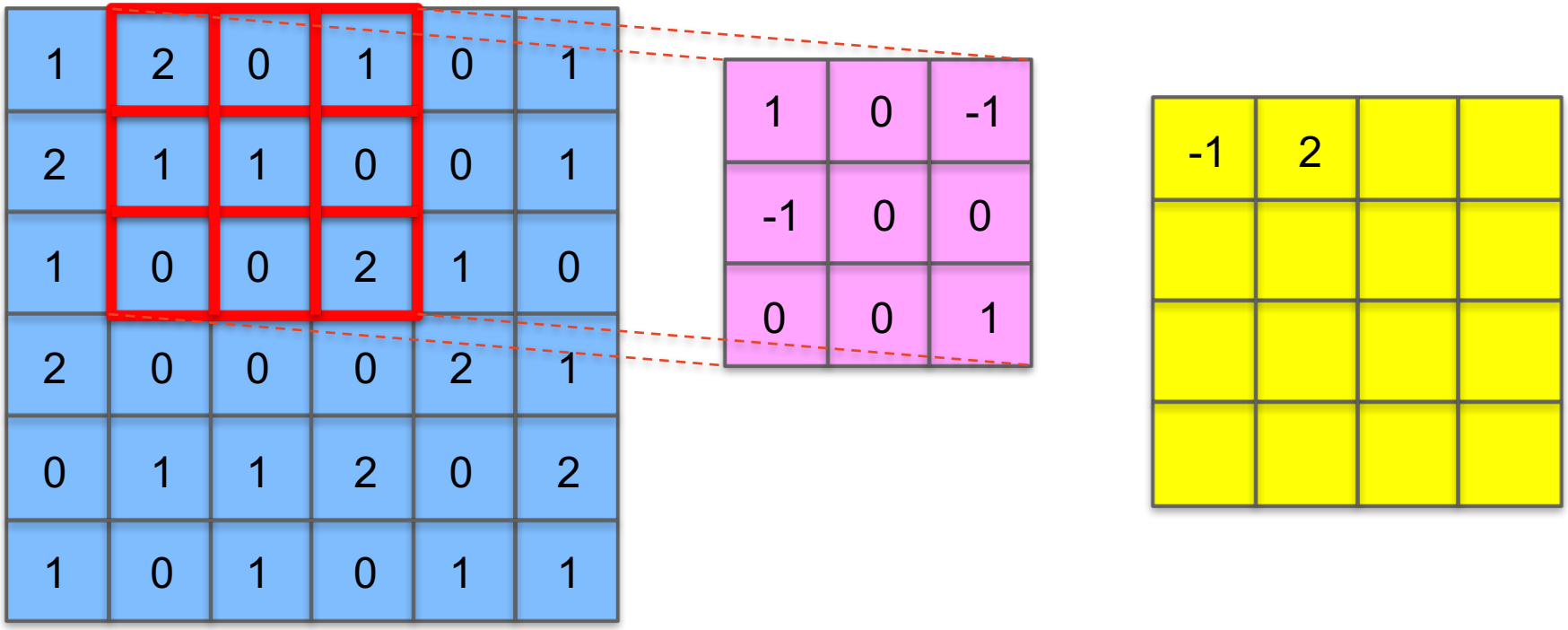
Convolutional Layer

- Convolution



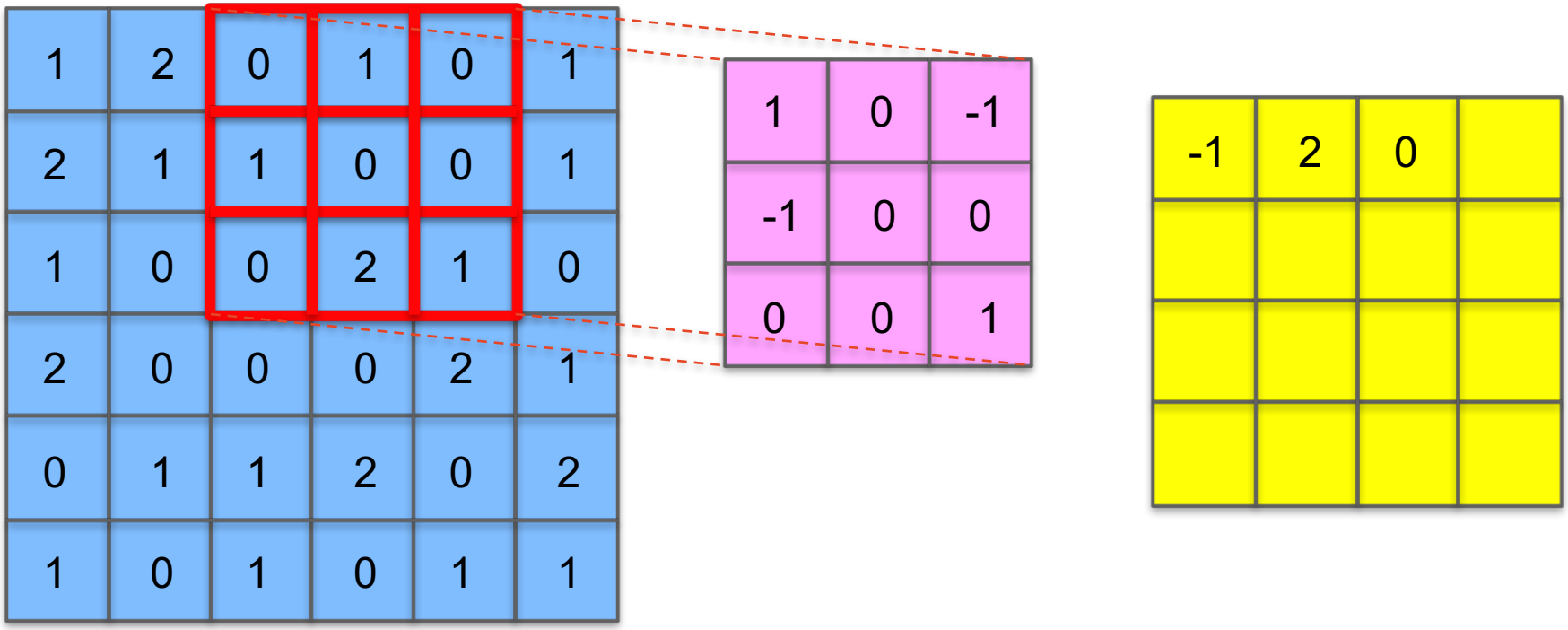
Convolutional Layer

- Convolution



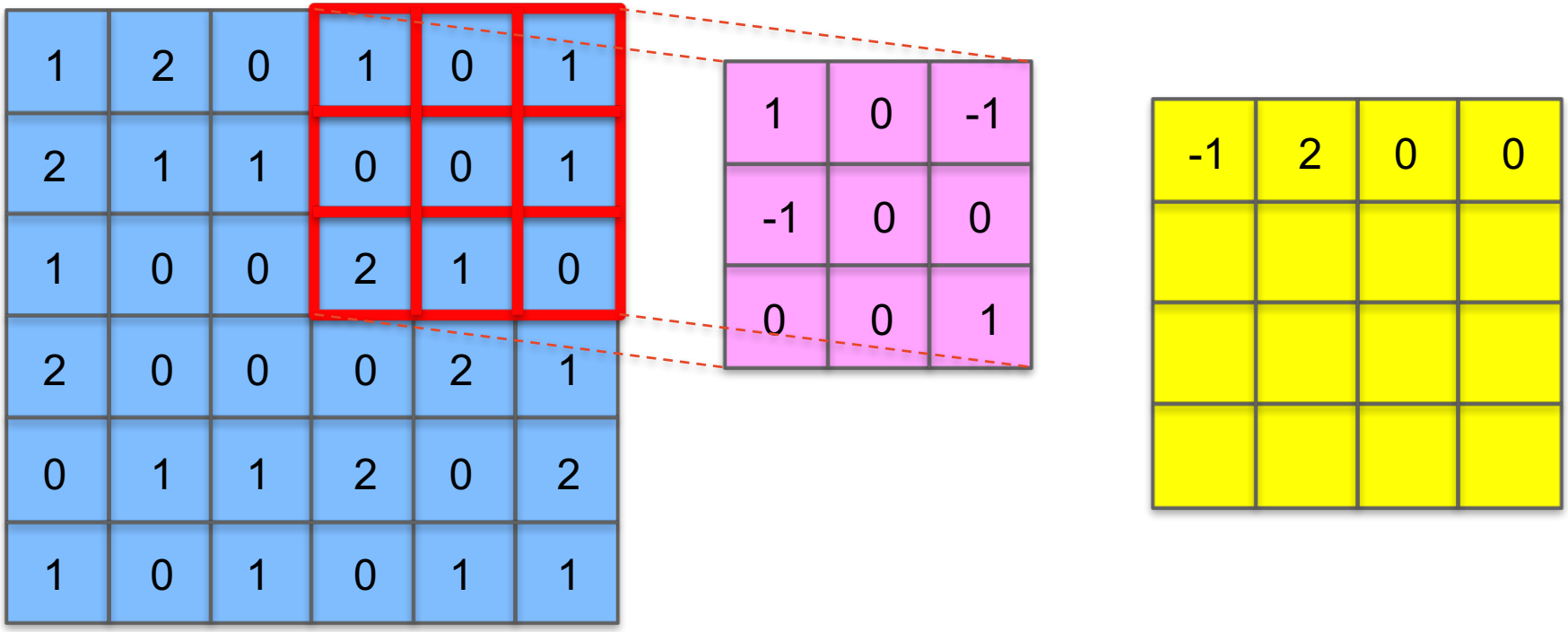
Convolutional Layer

- Convolution



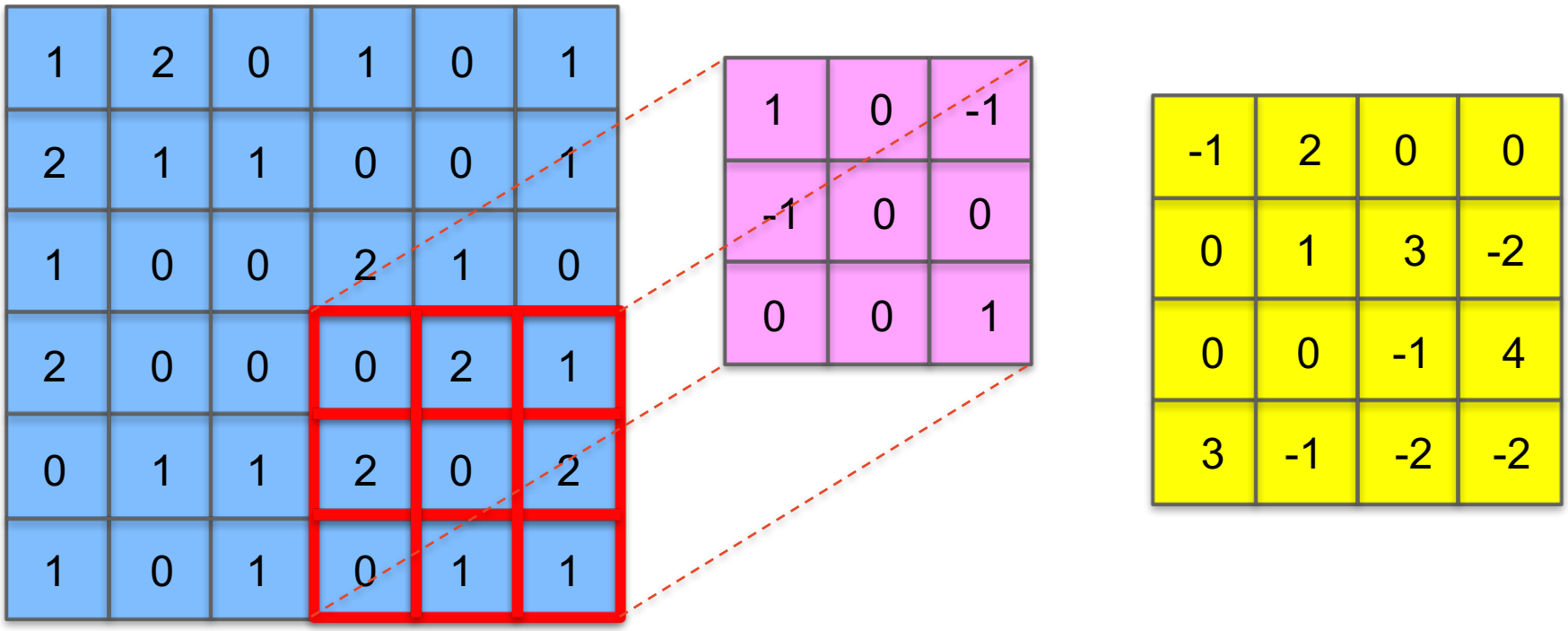
Convolutional Layer

- Convolution



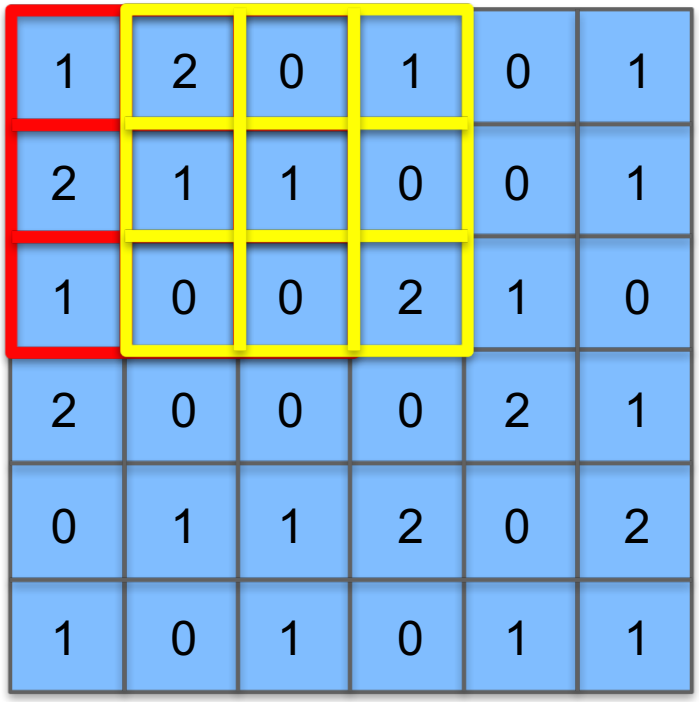
Convolutional Layer

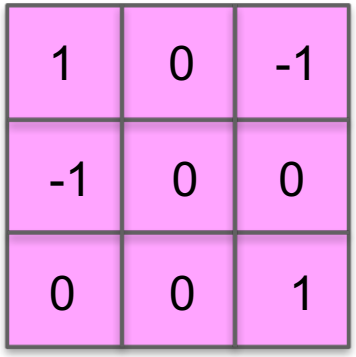
- Convolution

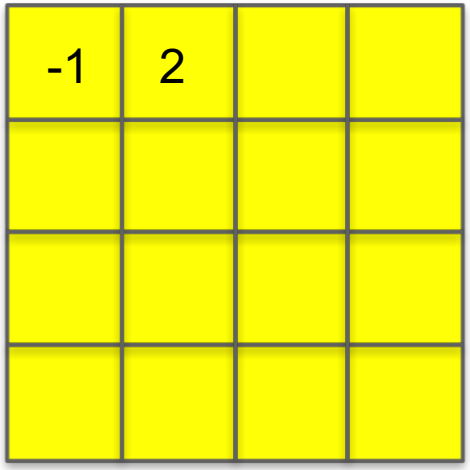


Convolutional Layer

- Stride



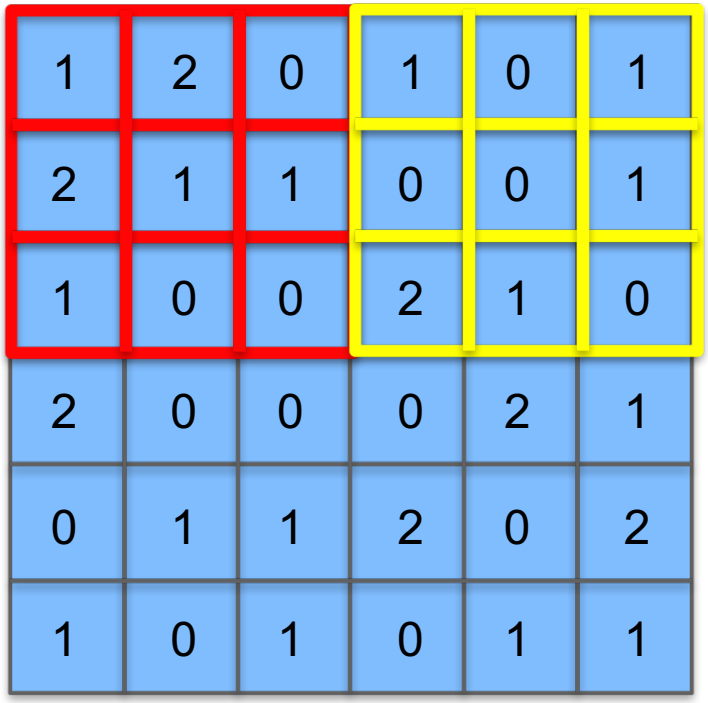


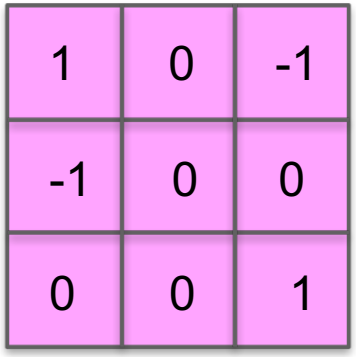


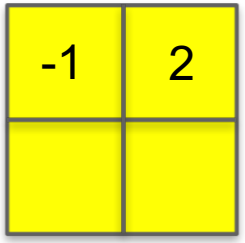
Stride = 1

Convolutional Layer

- Stride



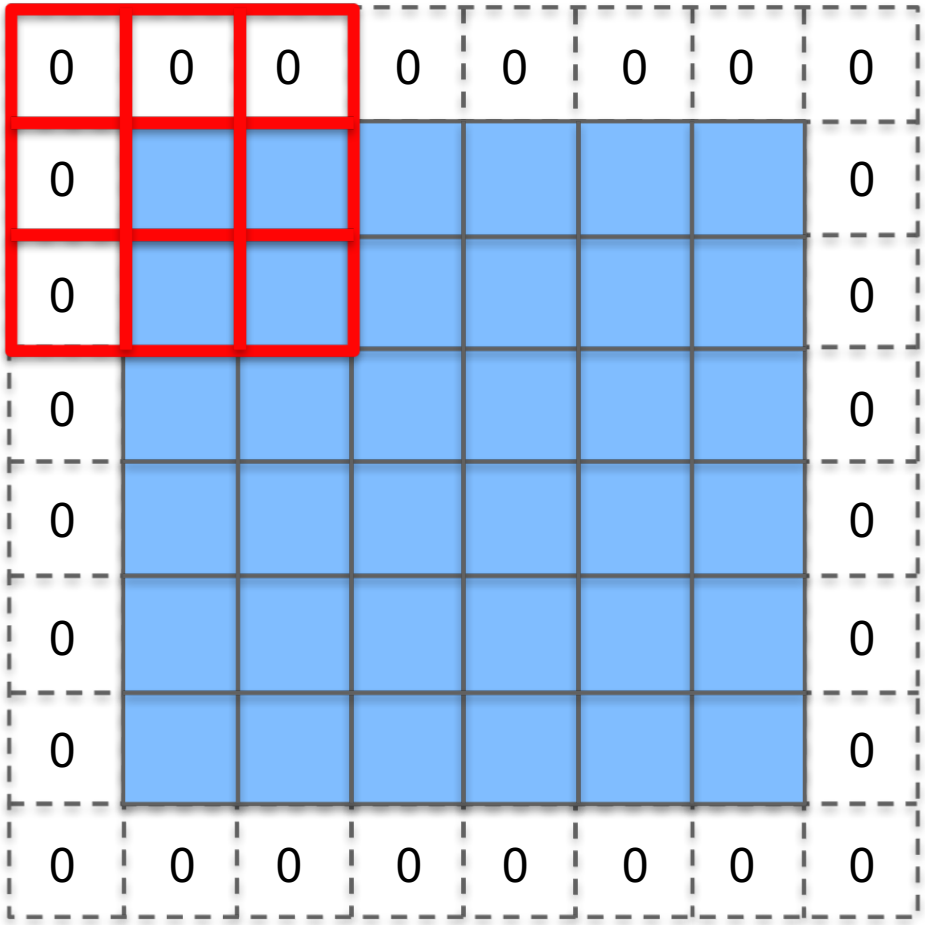




Stride = 3

Convolutional Layer

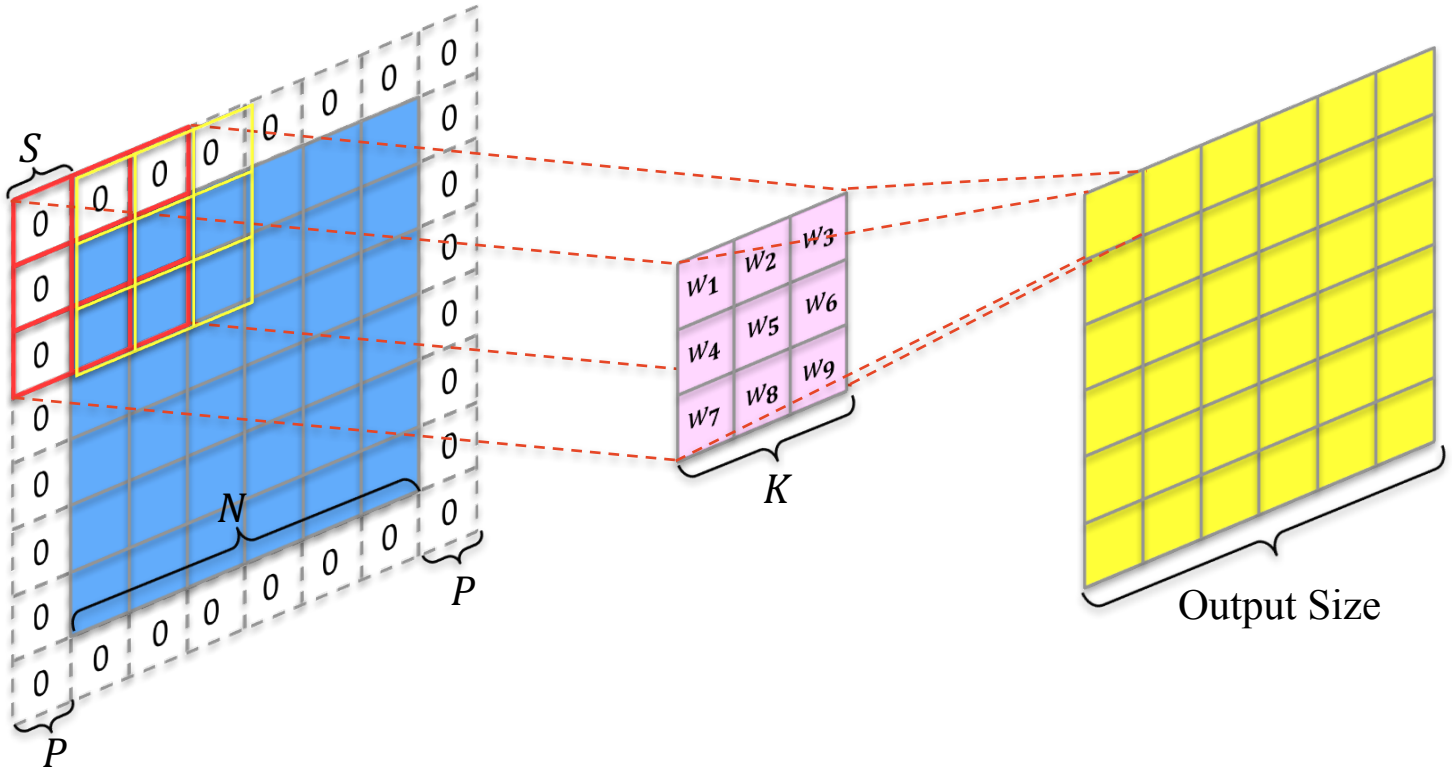
- Zero padding (pad = 1)



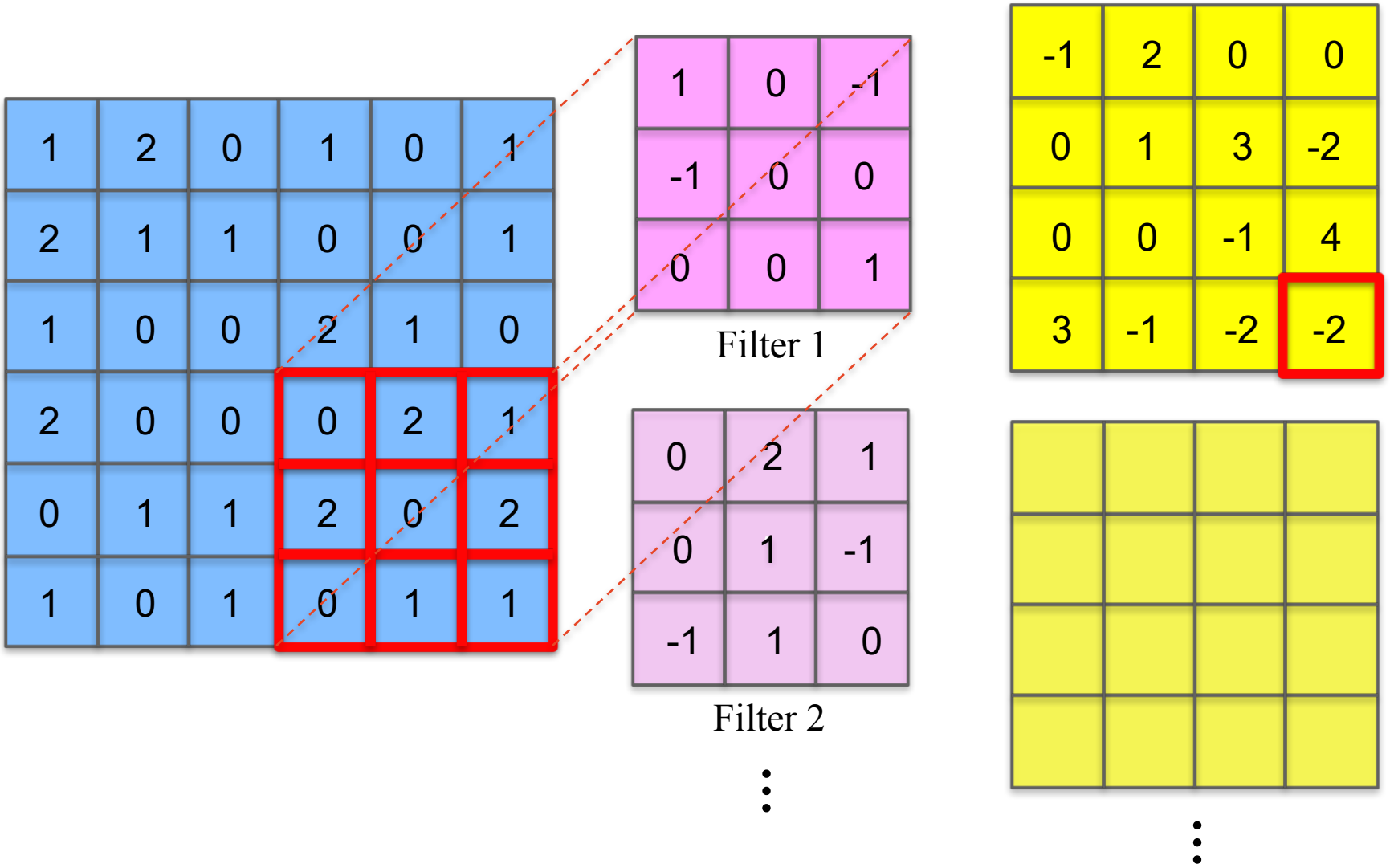
Convolutional Layer

- Output Size

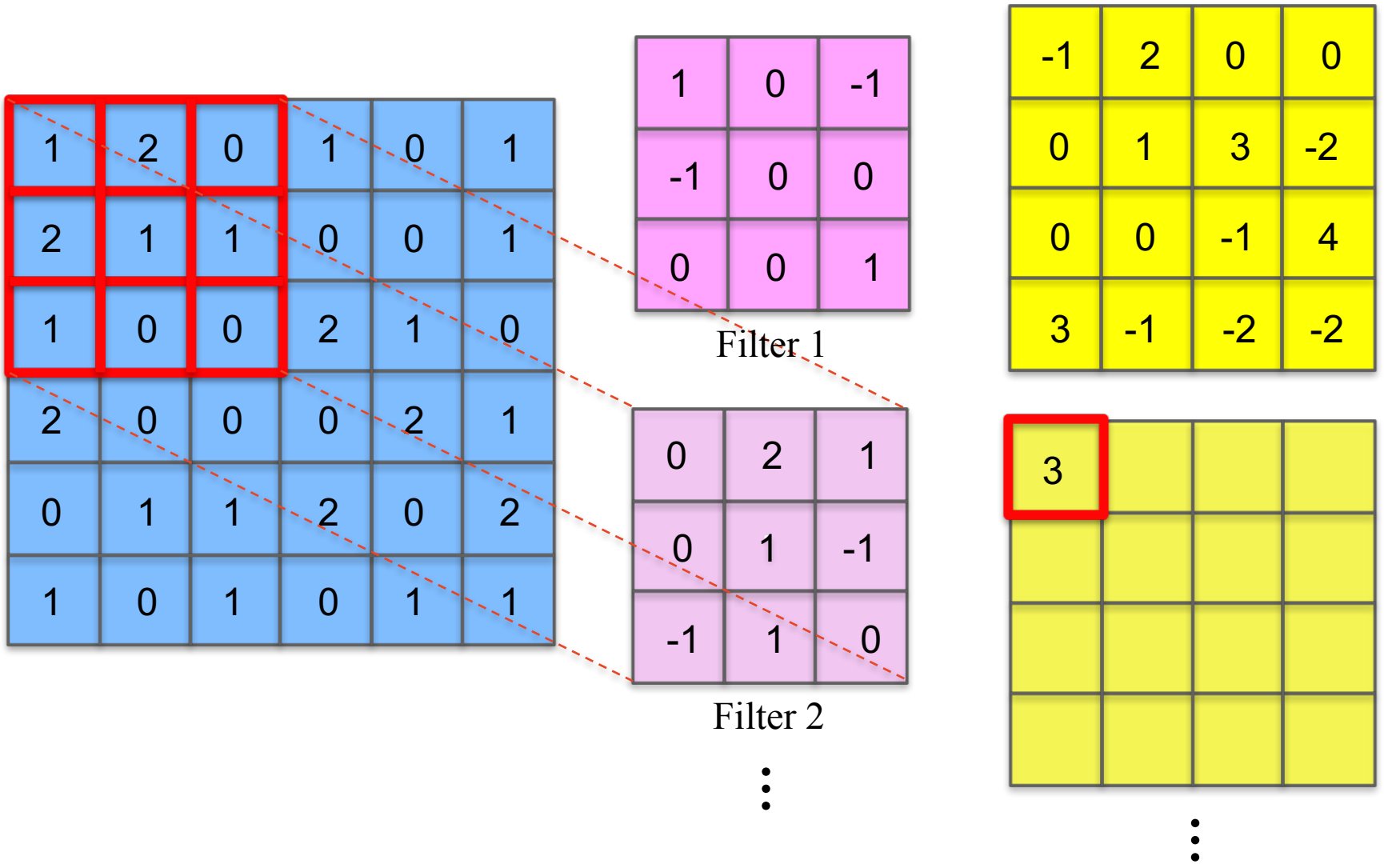
$$\text{Output Size} = \frac{N+2P-K}{S} + 1$$



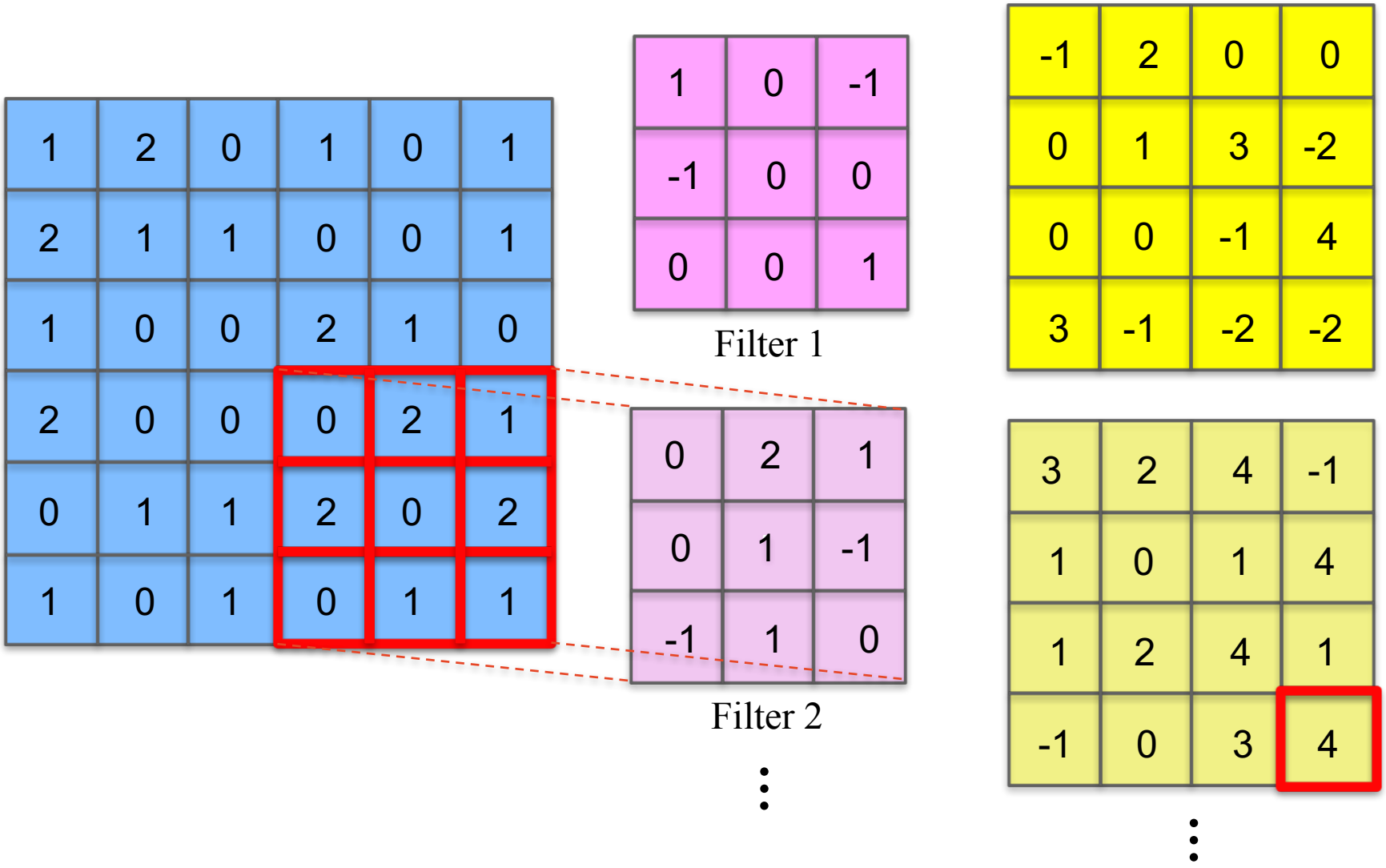
Learn multiple filters



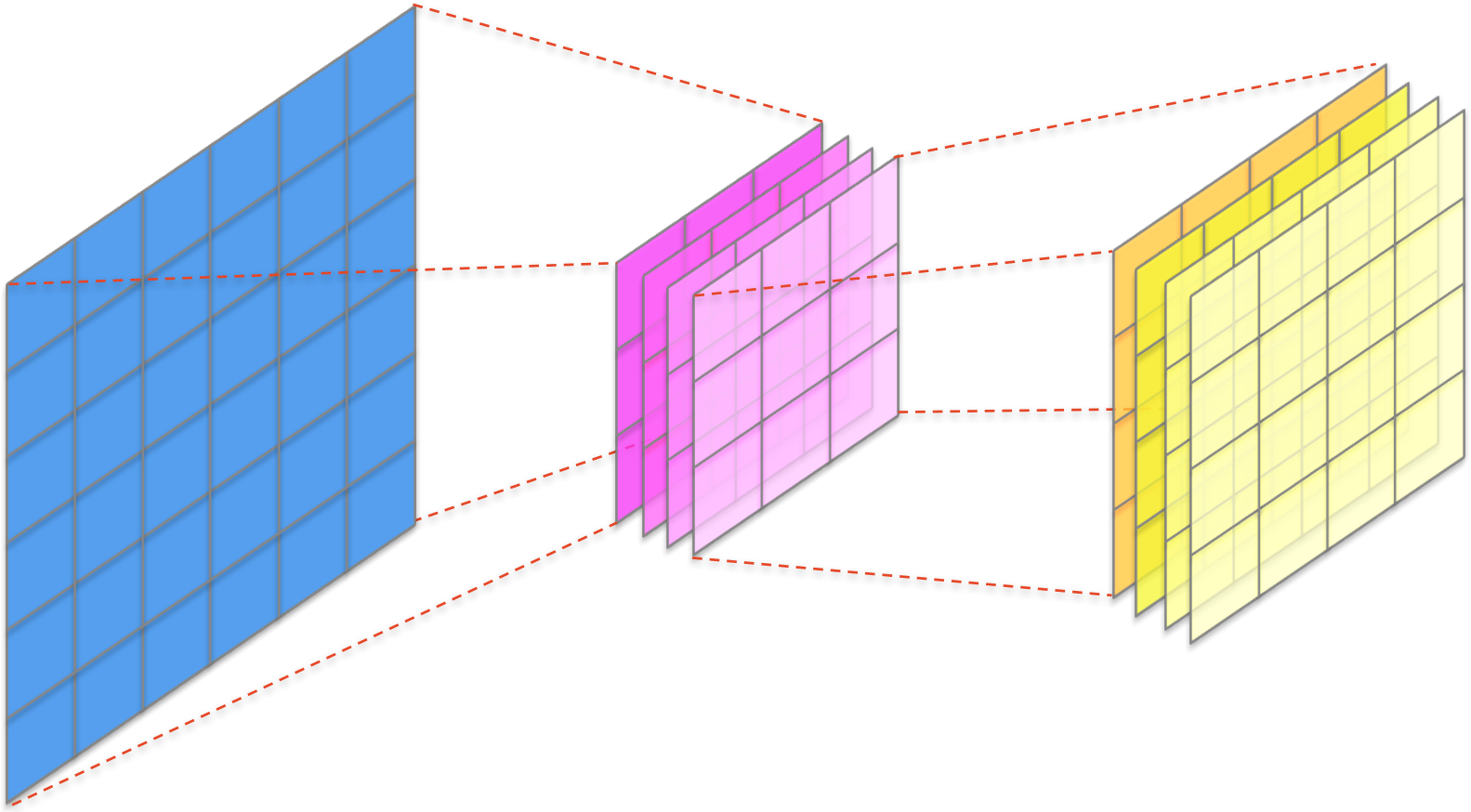
Learn multiple filters



Learn multiple filters

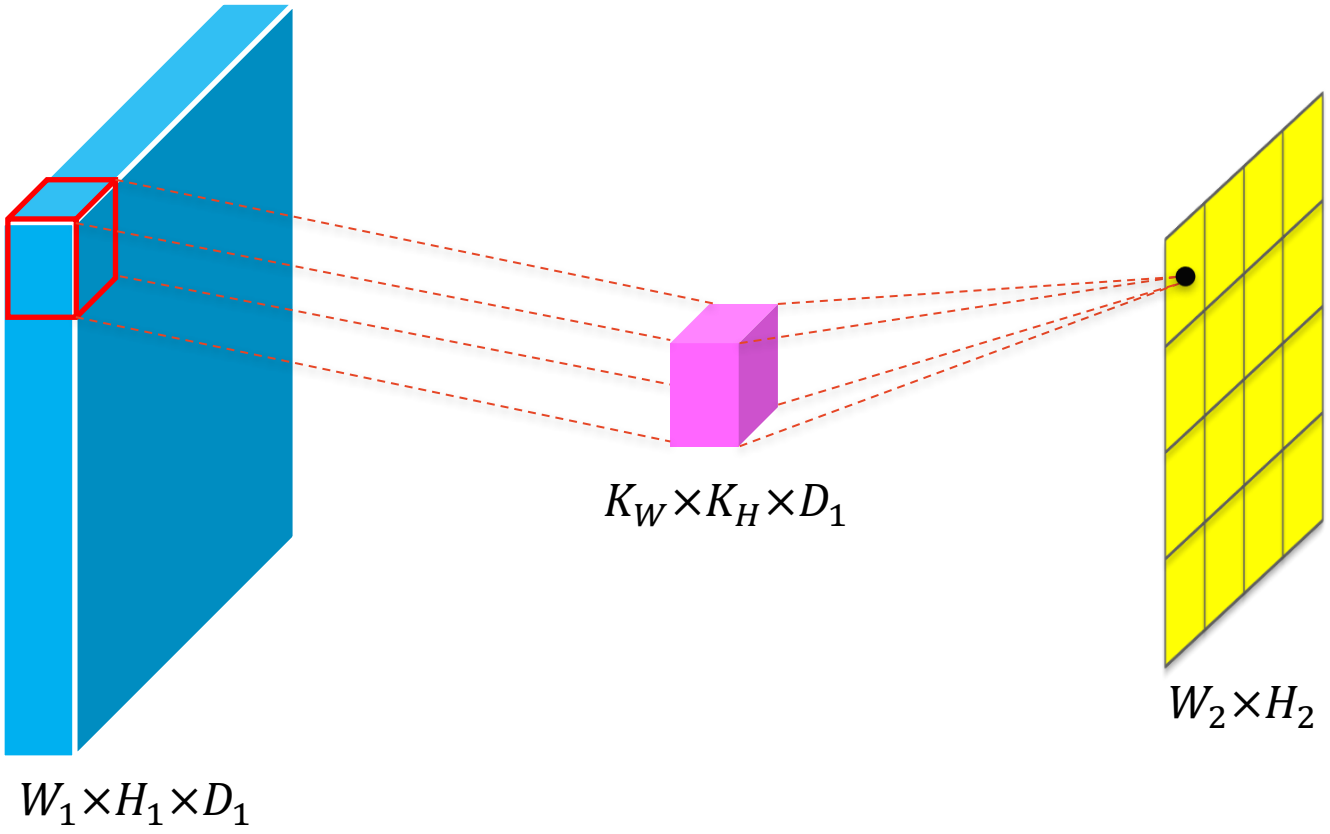


Learn multiple filters

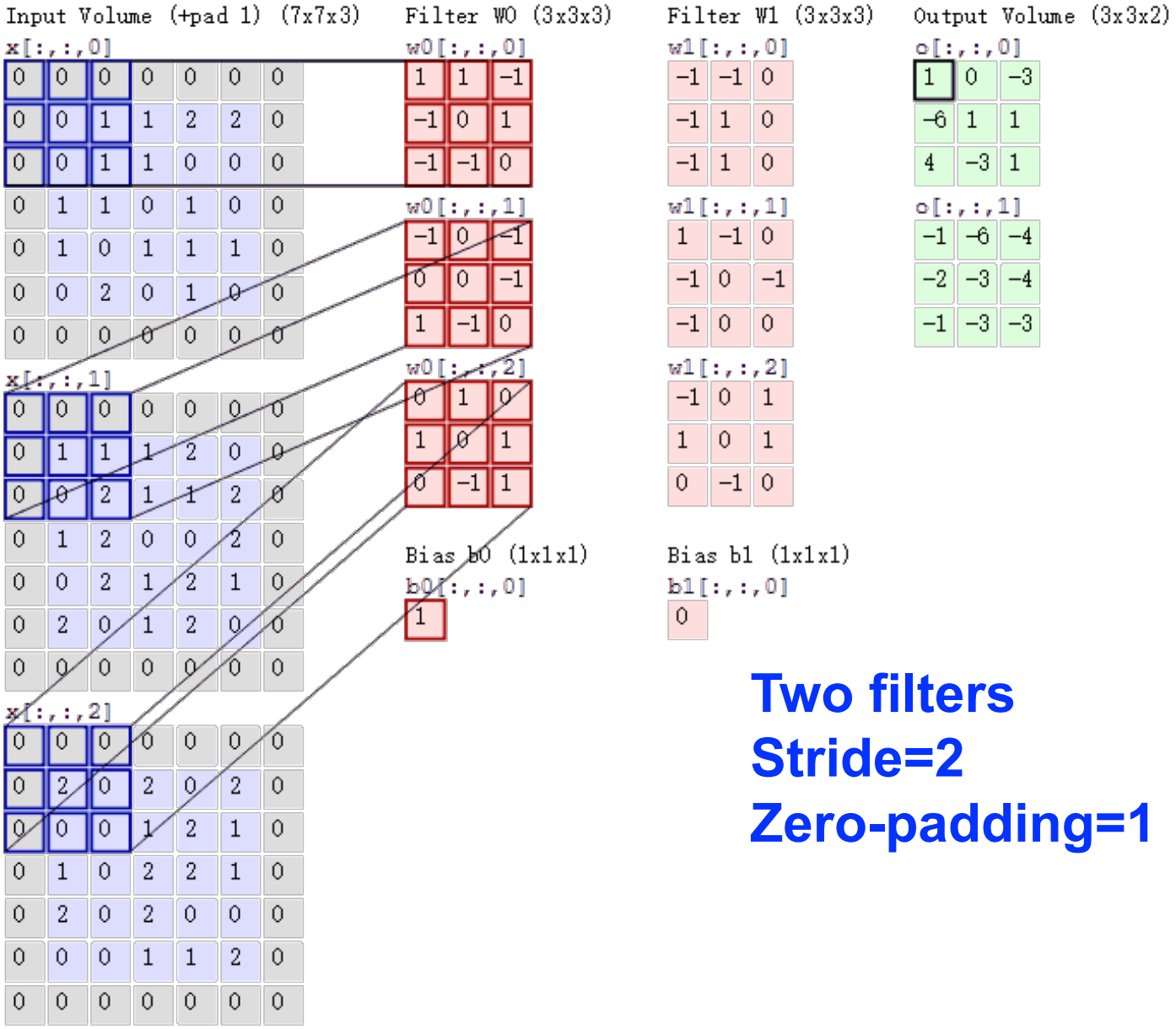


Convolutional Layer

- Above, we have only considered a 2-D image as input
- When the input has depth (e.g. RGB images), the convolution ops should be…

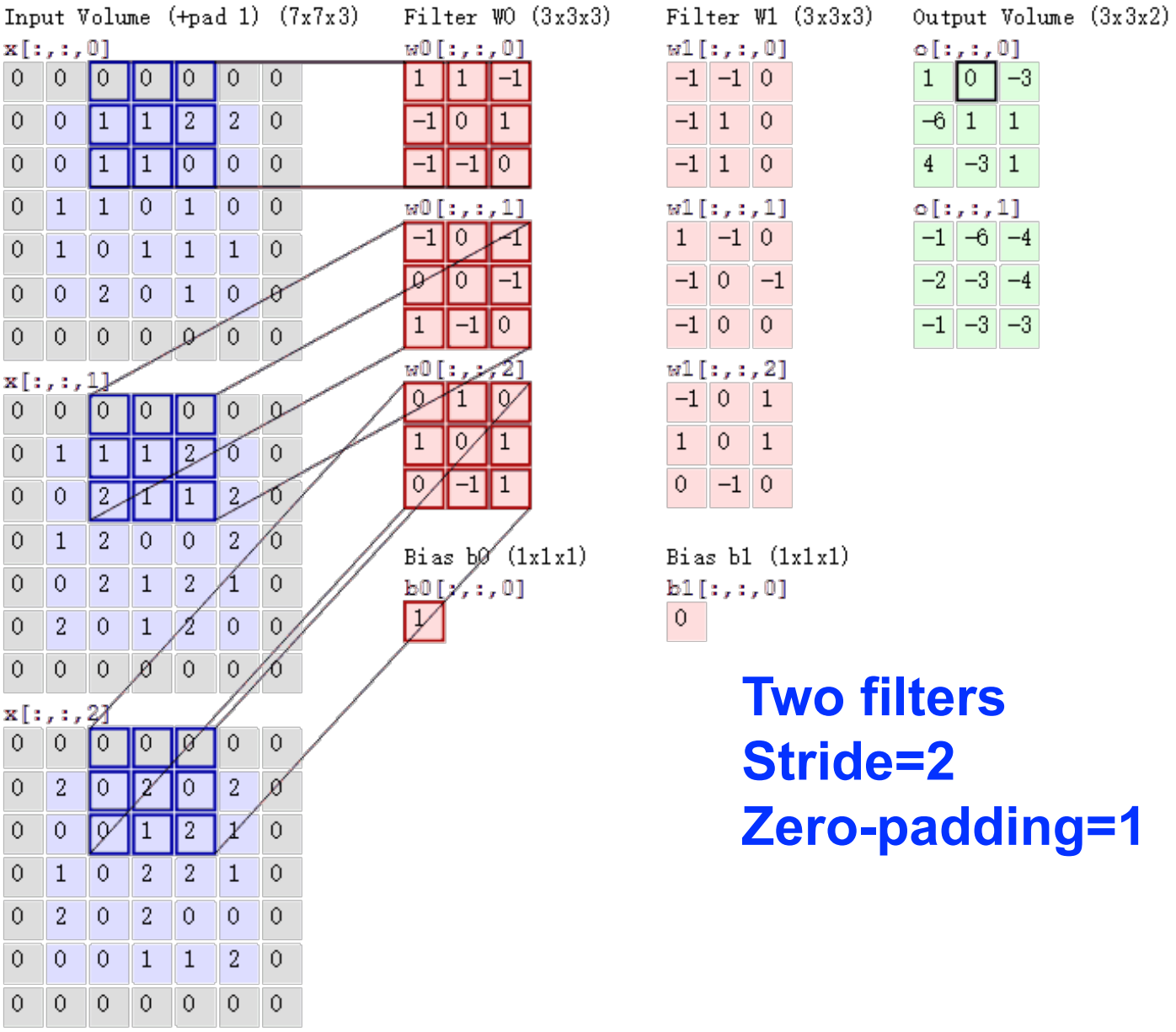


Convolutional Layer



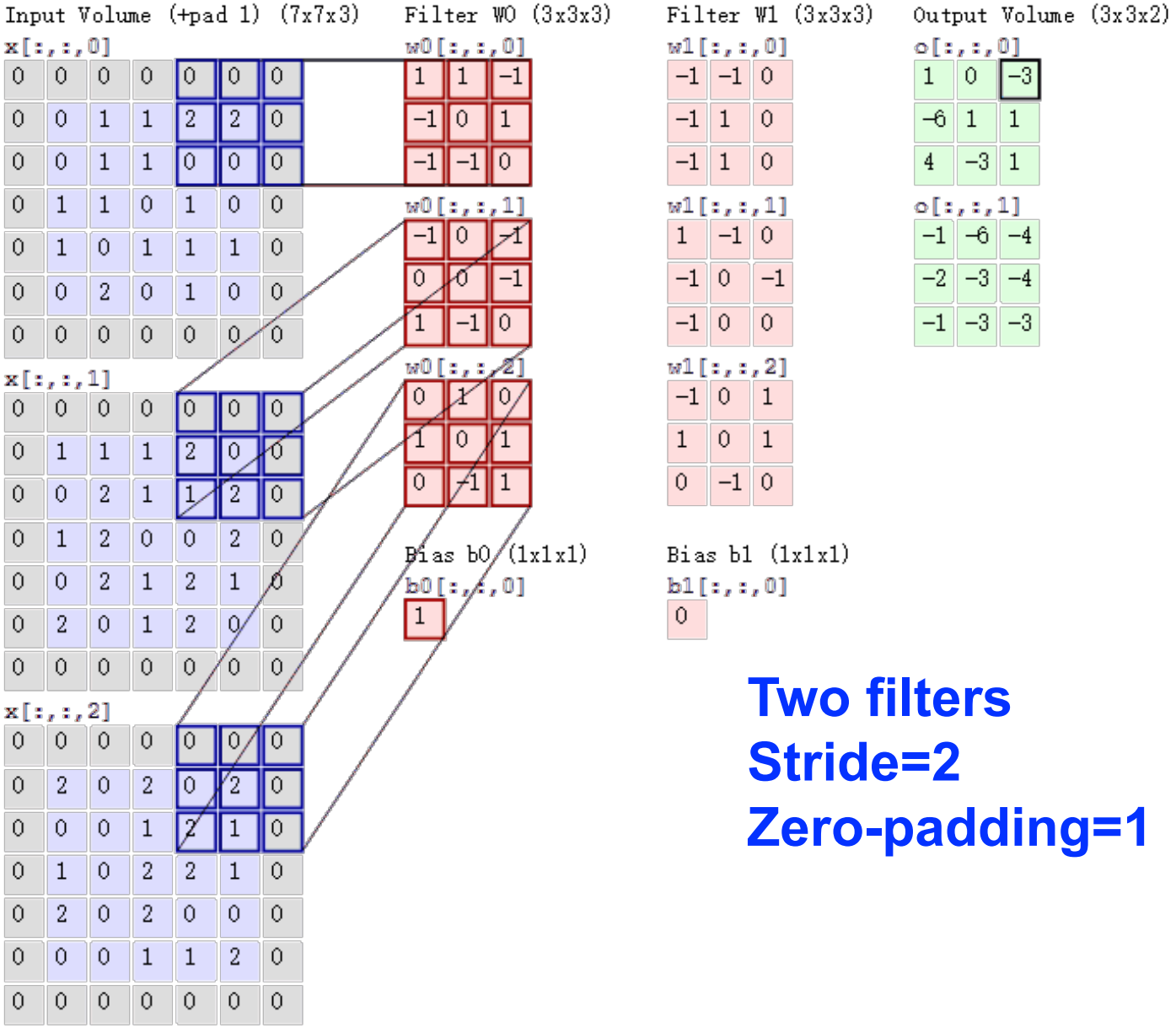
Two filters
Stride=2
Zero-padding=1

Convolutional Layer



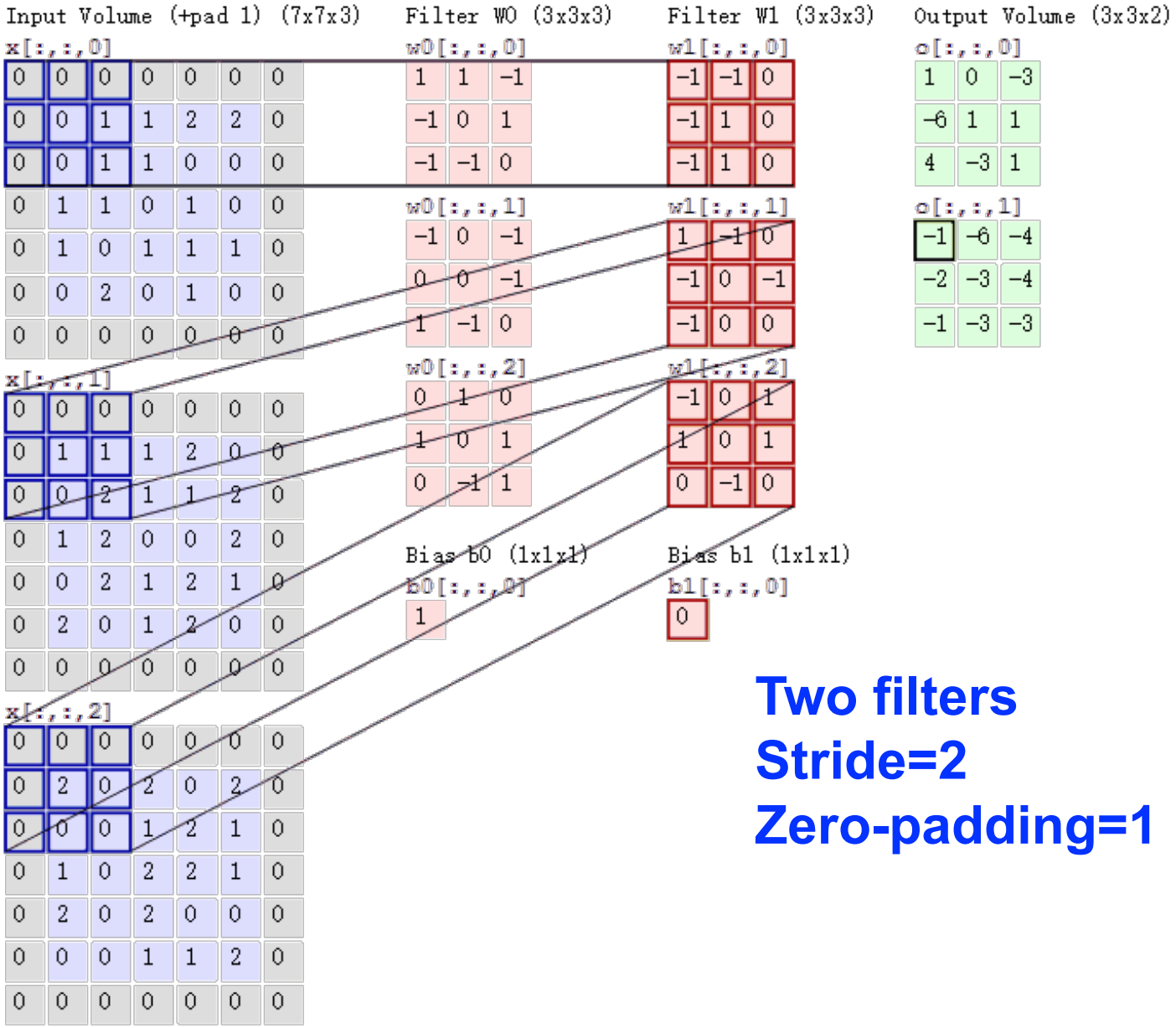
Two filters
Stride=2
Zero-padding=1

Convolutional Layer



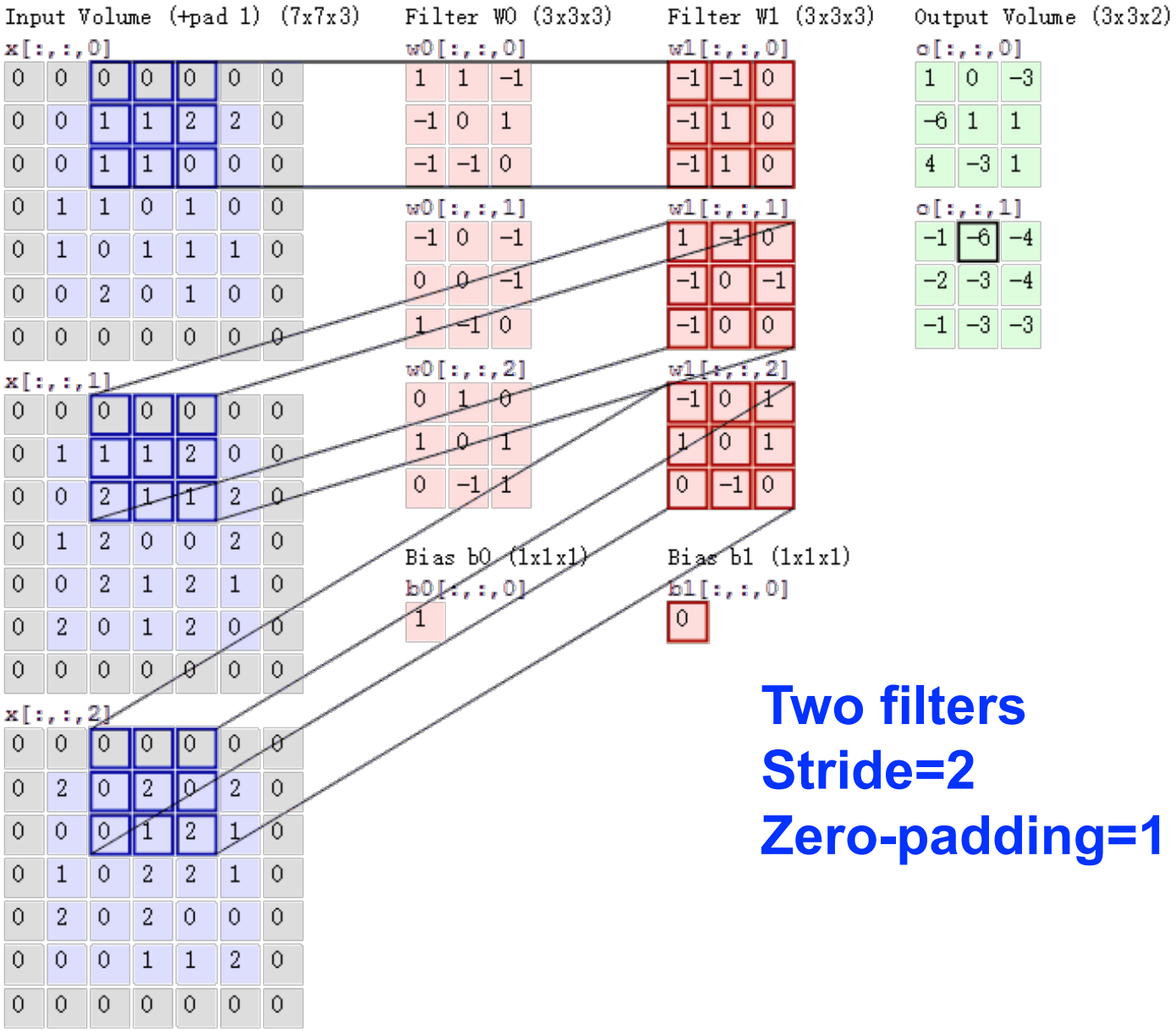
Two filters
Stride=2
Zero-padding=1

Convolutional Layer



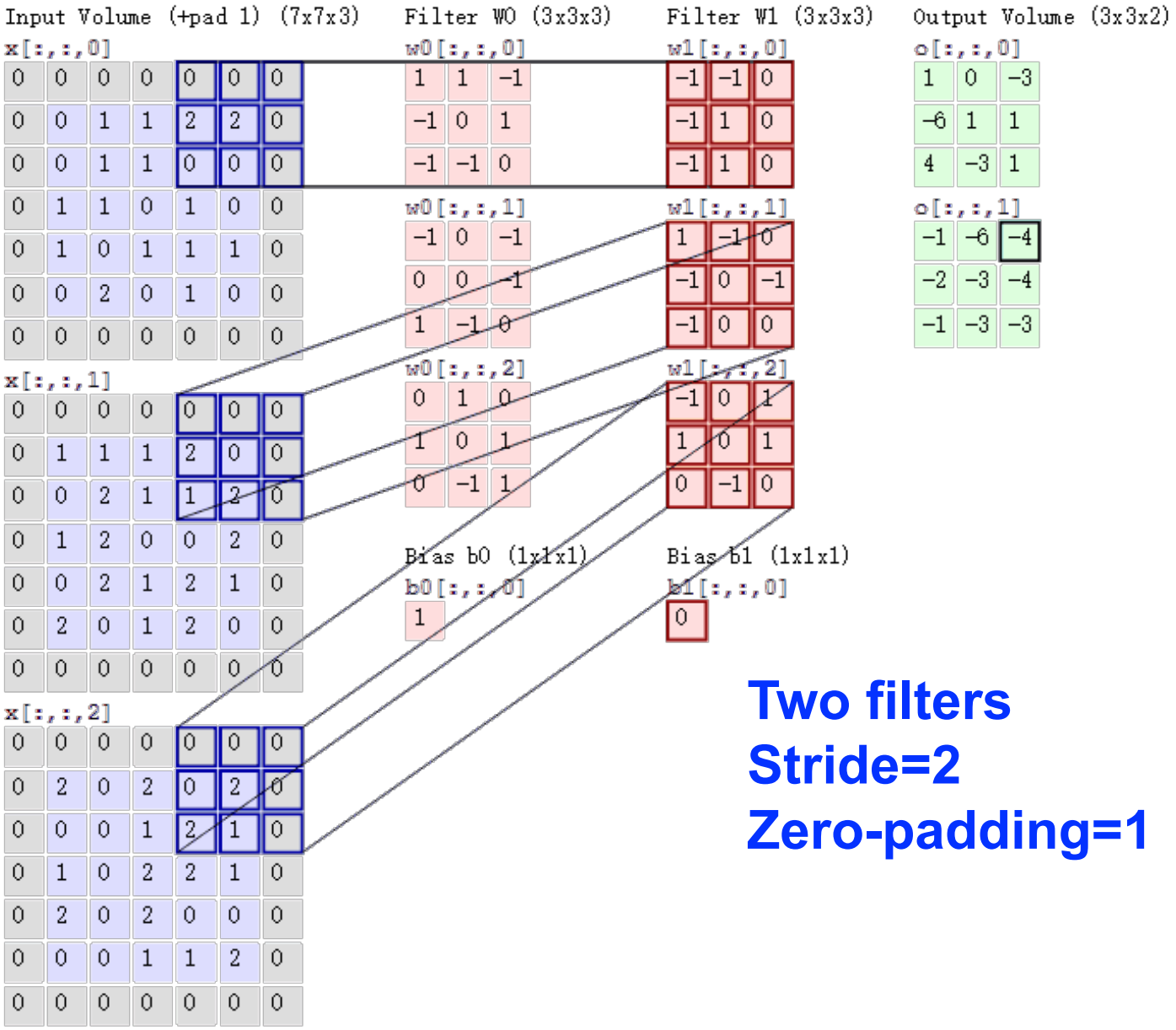
Two filters
Stride=2
Zero-padding=1

Convolutional Layer



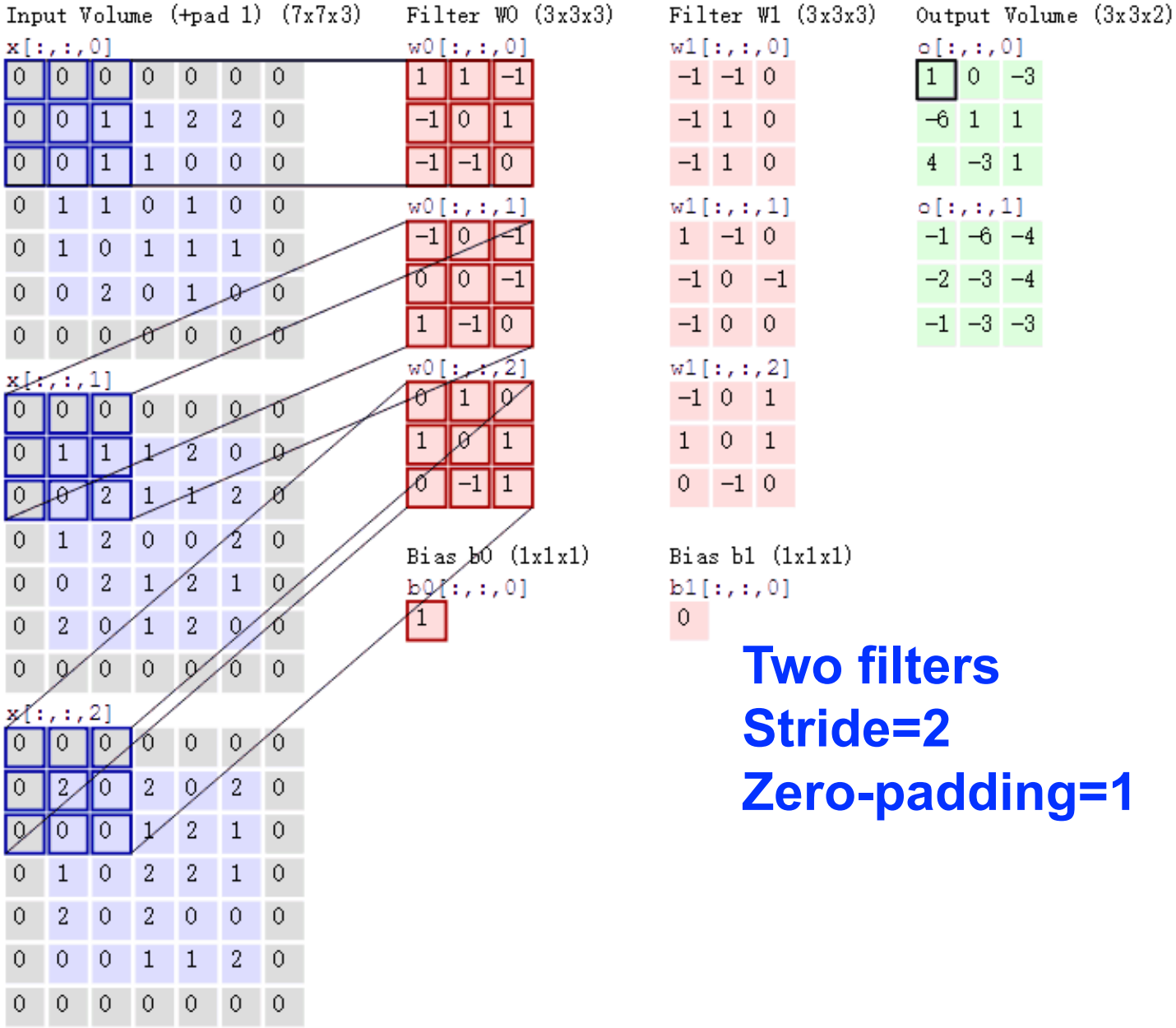
Two filters
Stride=2
Zero-padding=1

Convolutional Layer



Two filters
Stride=2
Zero-padding=1

Convolutional Layer

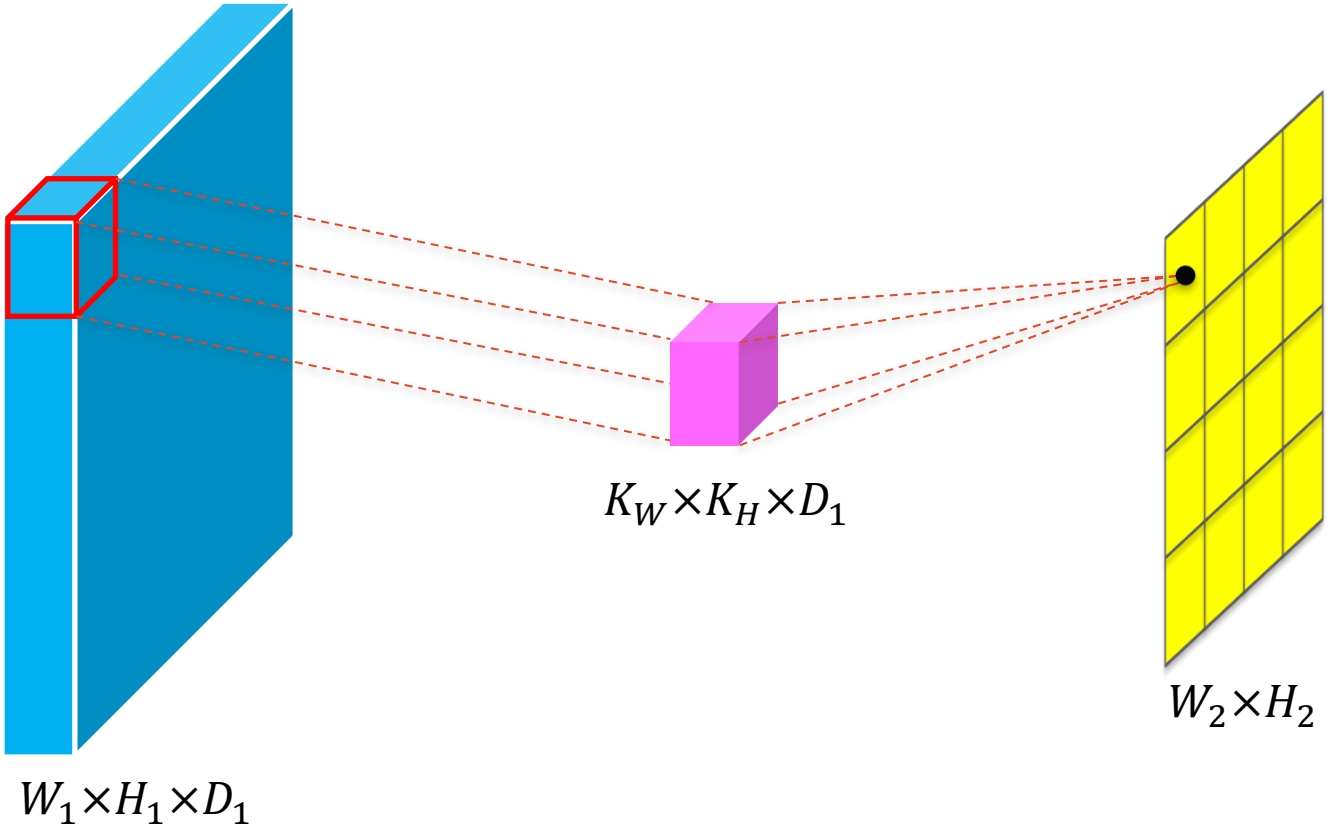


Two filters
Stride=2
Zero-padding=1

Convolutional Layer

- Suppose stride is (S_W, S_H) and pad is (P_W, P_H)

$$W_2 = \frac{W_1 + 2P_W - K_W}{S_W} + 1 \quad \text{and} \quad H_2 = \frac{H_1 + 2P_W - K_H}{S_H} + 1$$

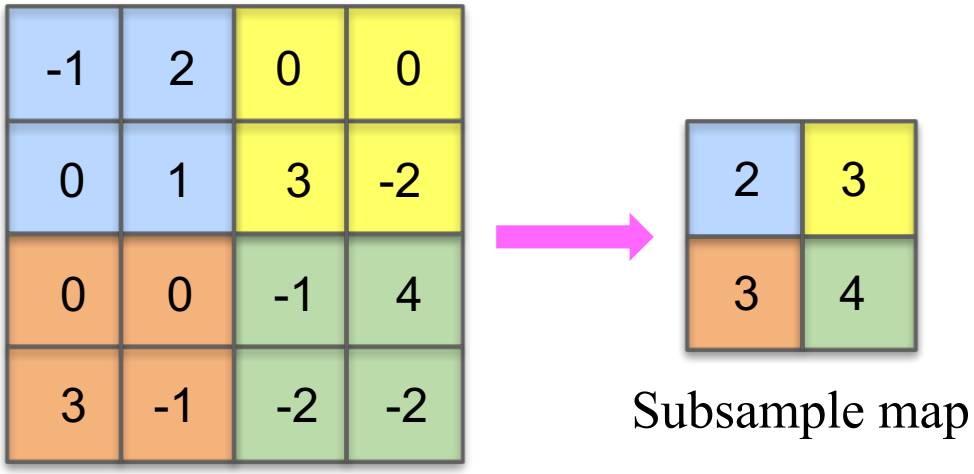


Pooling

Pooling

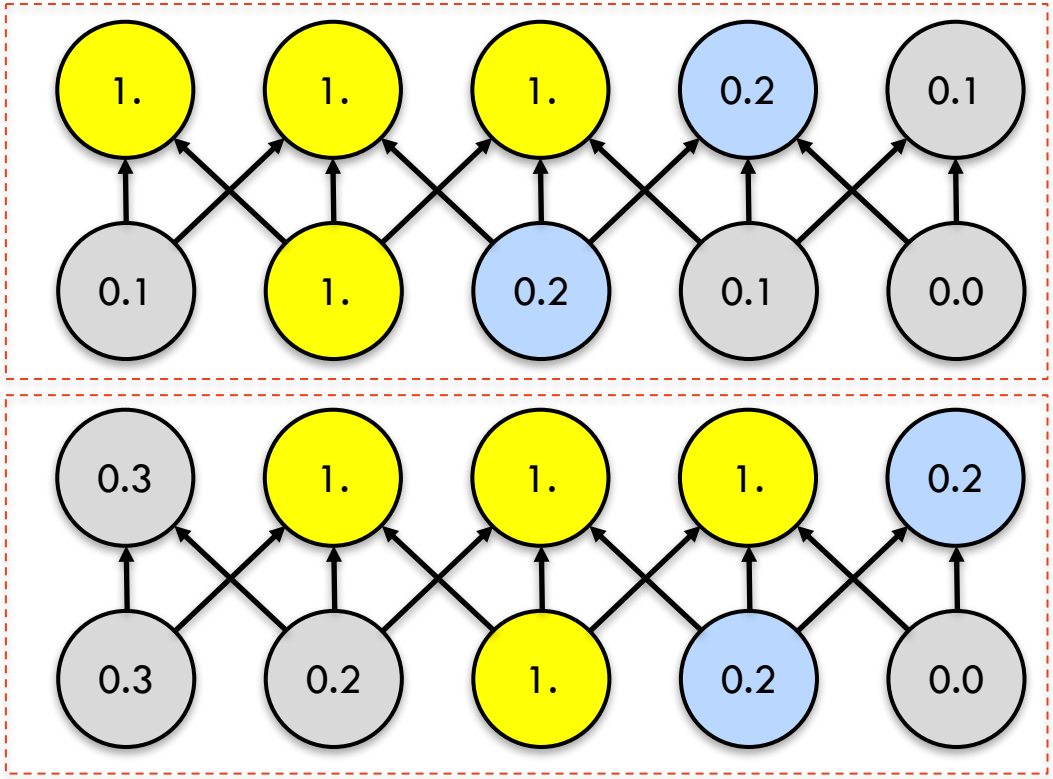
Max pooling

- Filter size: (2,2)
- Stride: (2,2)
- Pooling ops: max(·)

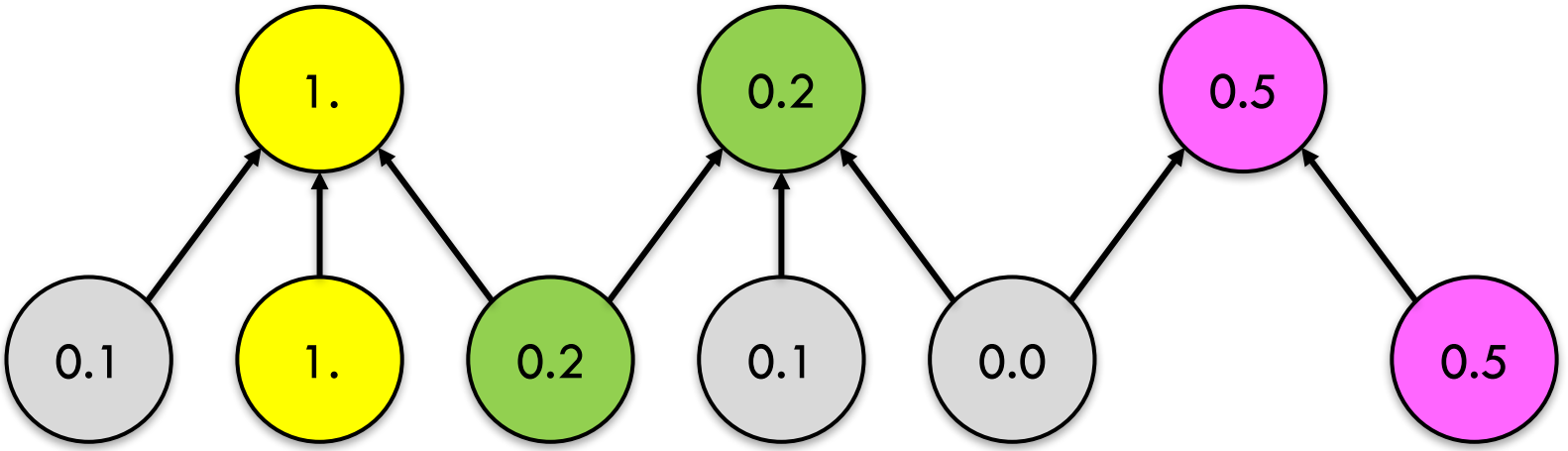


Motivation: Pooling

- Taking max pooling as an example:



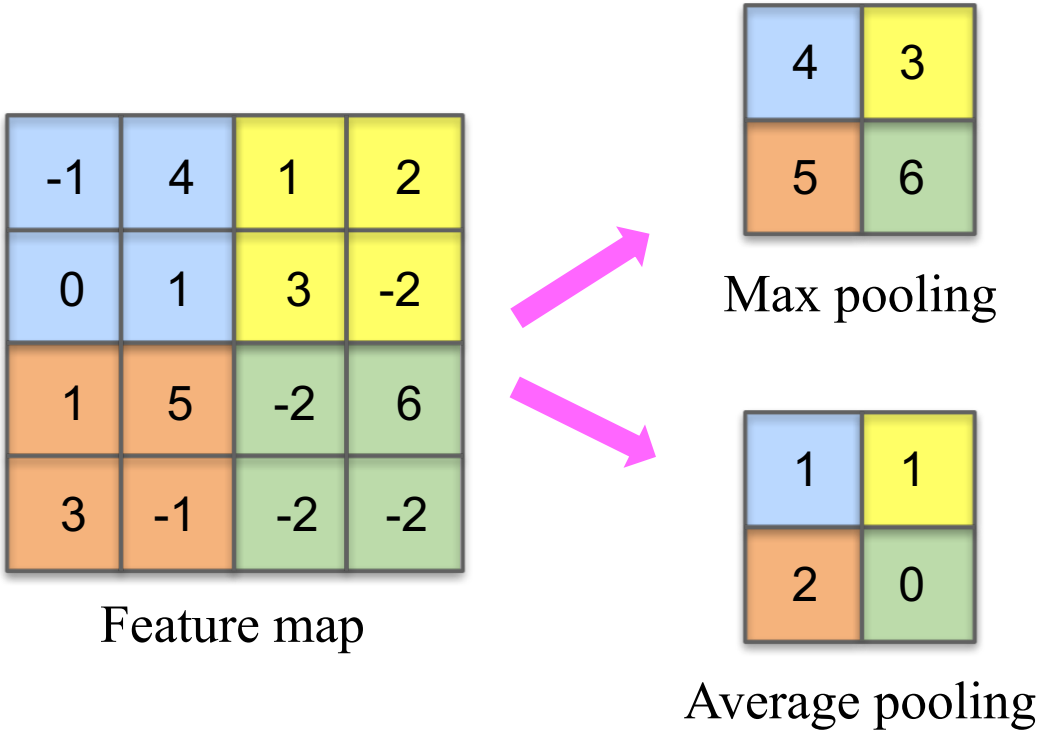
Motivation: Pooling



Pooling

Average pooling

- Filter size: (2,2)
- Stride: (2,2)
- Pooling ops: mean(·)

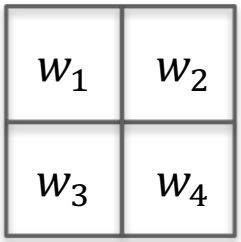


Pooling

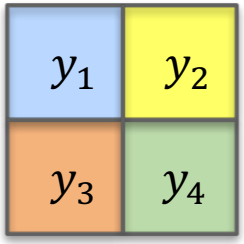
L_2 norm pooling

- Filter size (Gaussian kernel size): (2,2)
- Stride: (2,2)
- Pooling ops: $y_i = \sqrt{\sum_j w_j x_{i,j}{}^2}$

Feature map



Gaussian window



Output

Pooling

- Other pooling
  - L_p **pooling** (preserves the class-specific spatial/geometric information in the pooled features)

$$y_i = \left(\sum_j w_j {x_{i,j}}^p \right)^{1/p}$$

  - **Mixed pooling** (addresses the over-fitting problem)

$$y_i = a \max(x_{i,1}, \dots, x_{i,n}) + (1 - a)\text{mean}(x_{i,1}, \dots, x_{i,n})$$

  - **Stochastic pooling** (hyper-parameter free, regularizes large CNNs)

$$y_i = x_l, \text{ where } l \sim P(p_1, \cdots, p_n) \text{ and } p_j = \frac{x_{i,j}}{\sum_j x_{i,j}}$$

  - **Spectral pooling** (preserves considerably more information per parameter than other pooling strategies)

$$\mathbf{y} = \mathcal{F}(\mathbf{x}) \in \mathbb{C}^{M \times N}, \hat{\mathbf{y}} = \mathcal{F}^{-1} (\mathbf{y} \in \mathbb{C}^{H \times W})$$

  - …

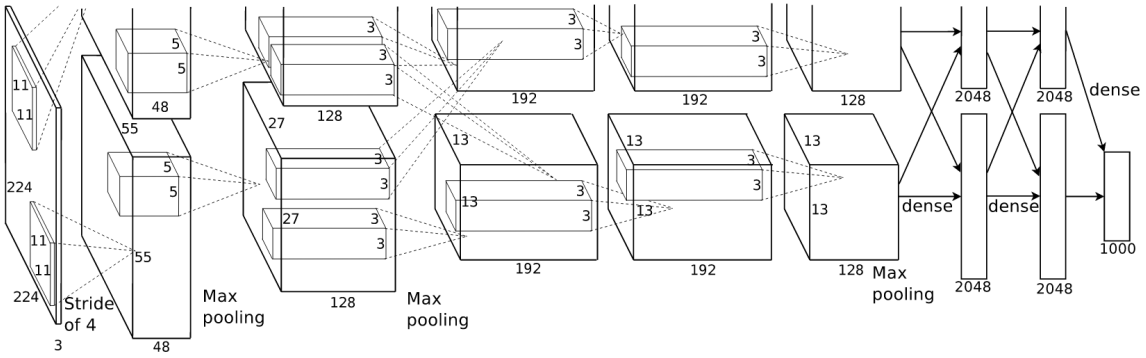
Visualize features

Visualize features

- Why CNNs work so well?

Hierarchical Convolution,
Nonlinear operations (ReLU, max pooling…)

Input image

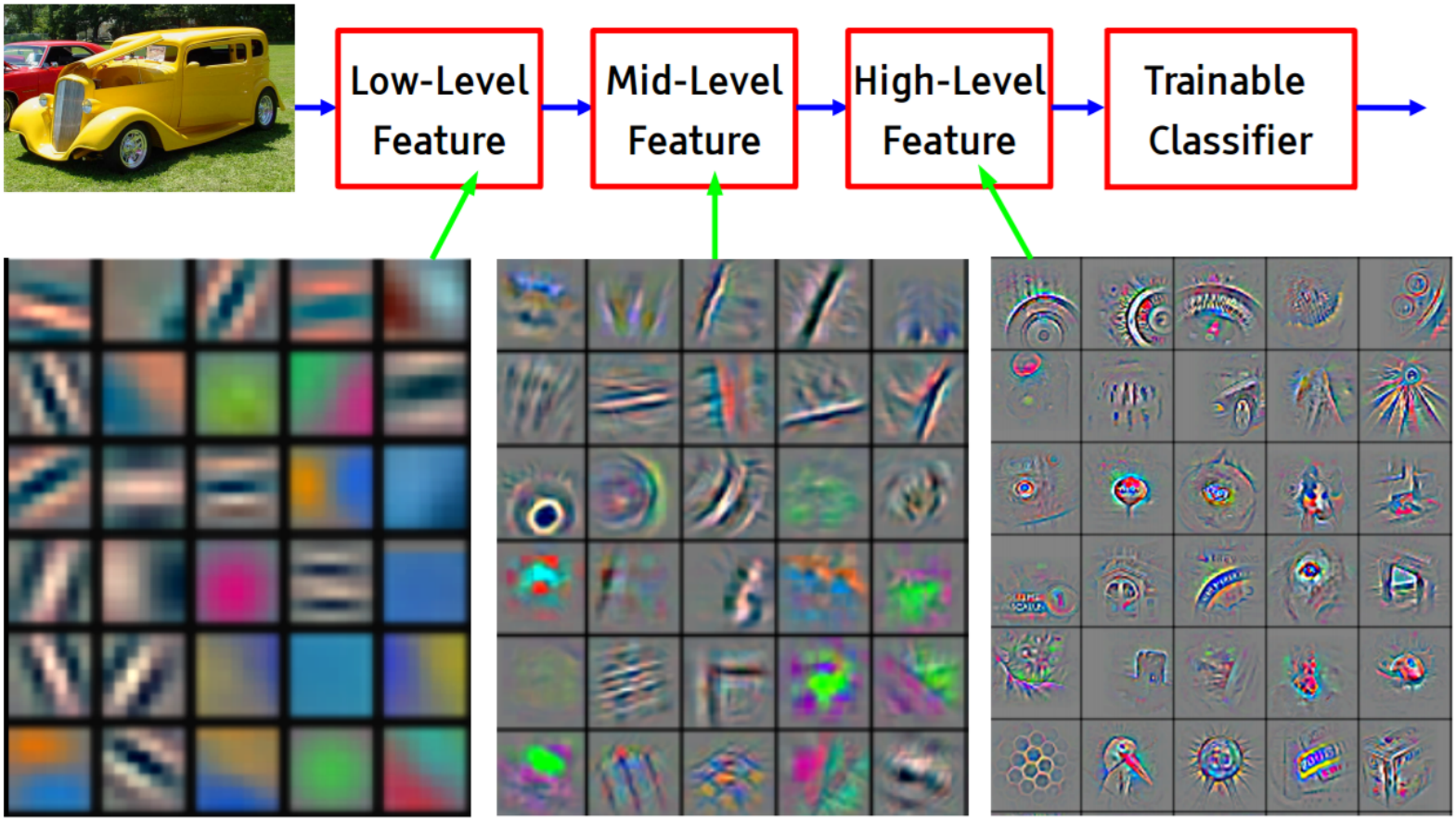


Class Scores:
1000 numbers

What happens inside hidden layers?

Visualize features

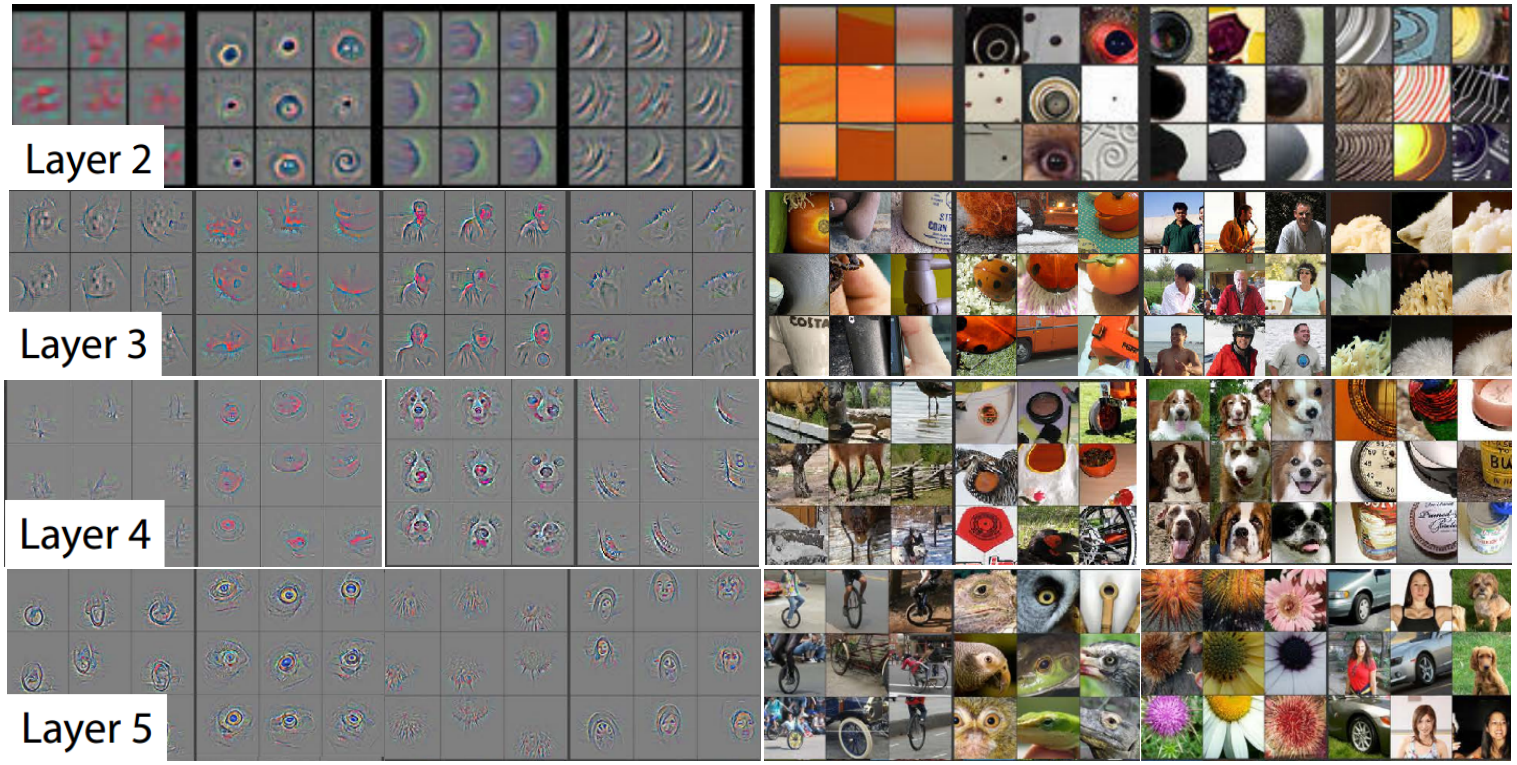
- Gives insight into the function of intermediate feature layers and the operation of the classifier



(Zeiler and Fergus, 2014)

Visualize features

- Layer 2 responds to corners and other edge/color conjunctions.
- Layer 3 has more complex invariances, capturing similar textures (e.g. mesh patterns).
- Layer 4 shows significant variation, but is more class-specific.
- Layer 5 shows entire objects with significant pose variation.



(Zeiler and Fergus, 2014)