北京大学
PEKING UNIVERSITY

Convolutional Neural Network

Kun Yuan

Center for Machine Learning Research @ Peking University

Oct. 24, 2023

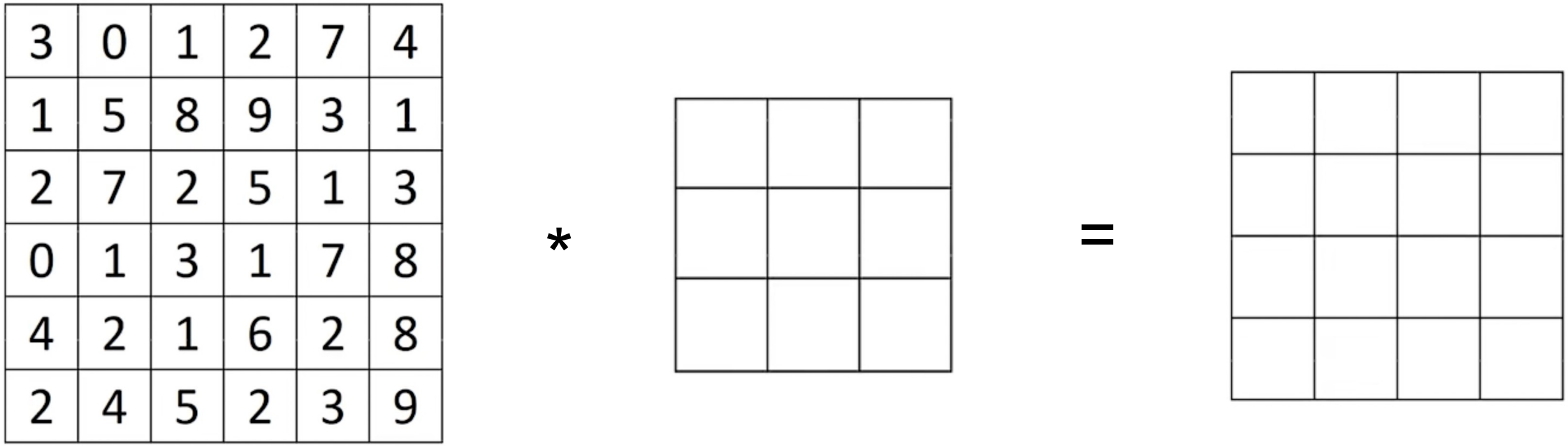
Most of the materials are from a great blog [1]

How to extract features in images?

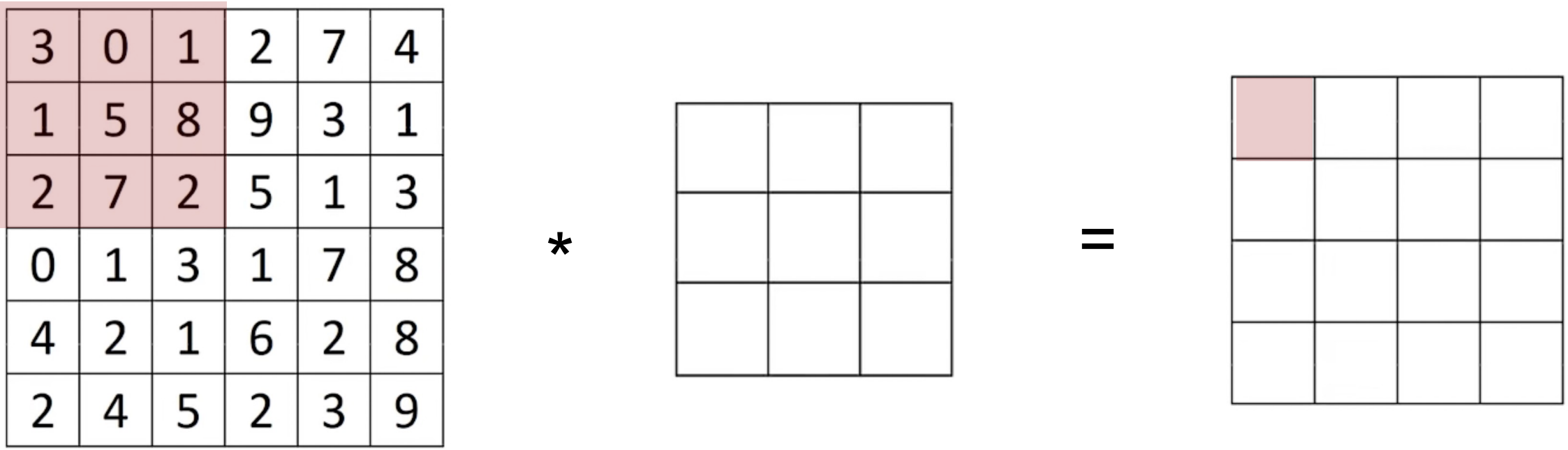
vertical edges

horizontal edges

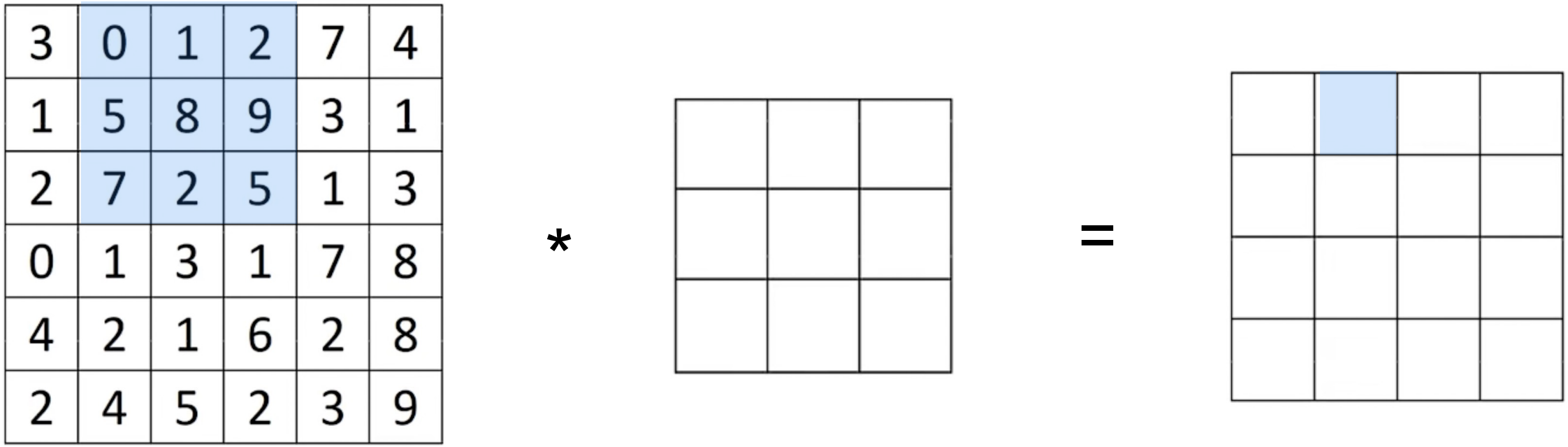
Convolution



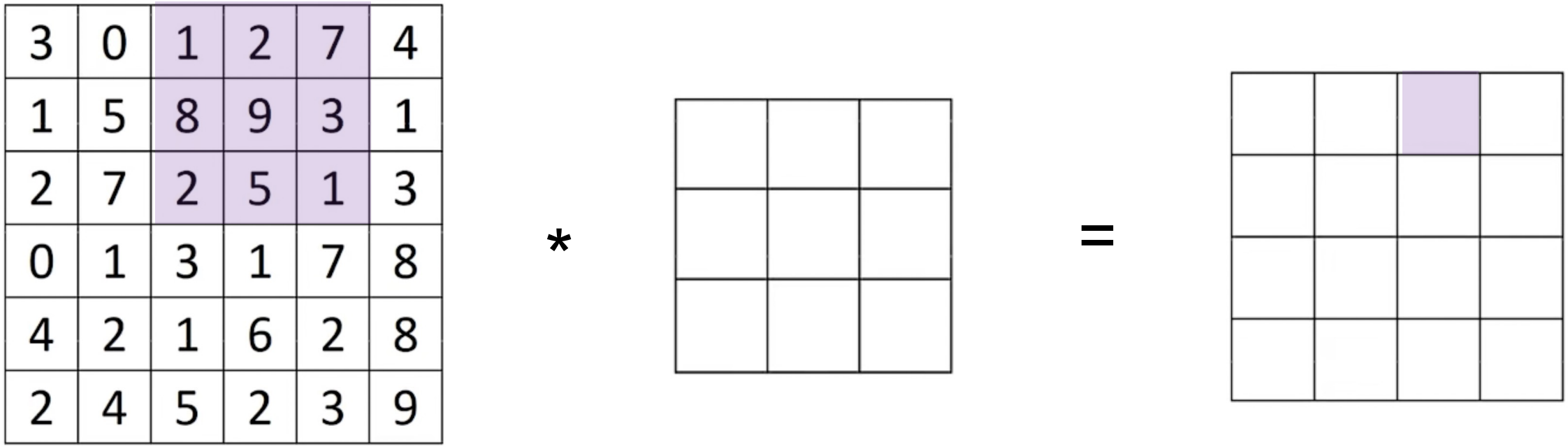
Convolution



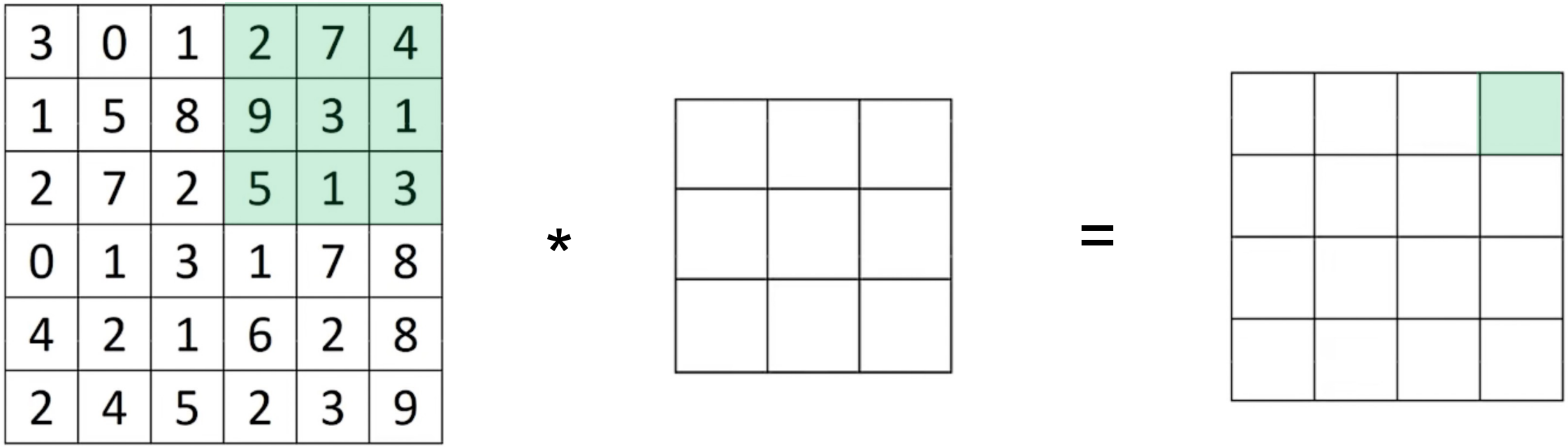
Convolution



Convolution



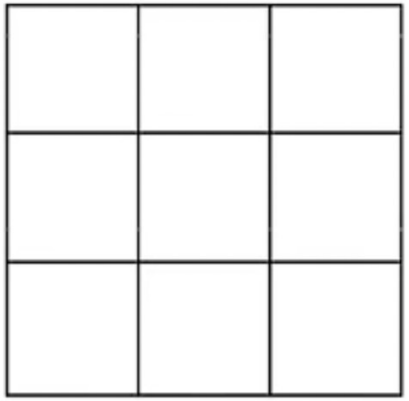
Convolution



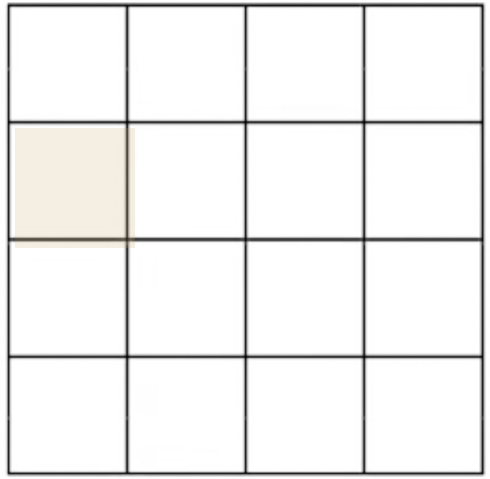
Convolution

3	0	1	2	7	4
1	5	8	9	3	1
2	7	2	5	1	3
0	1	3	1	7	8
4	2	1	6	2	8
2	4	5	2	3	9

*



=



Vertical edge detection

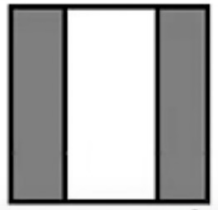
10	10	10	0	0	0
10	10	10	0	0	0
10	10	10	0	0	0
10	10	10	0	0	0
10	10	10	0	0	0
10	10	10	0	0	0

*

1	0	-1
1	0	-1
1	0	-1

=

0	30	30	0
0	30	30	0
0	30	30	0
0	30	30	0



Vertical edge detection

0	0	0	10	10	10
0	0	0	10	10	10
0	0	0	10	10	10
0	0	0	10	10	10
0	0	0	10	10	10
0	0	0	10	10	10

*

1	0	-1
1	0	-1
1	0	-1

=

0	-30	-30	0
0	-30	-30	0
0	-30	-30	0
0	-30	-30	0

Vertical and horizontal edge detection

1	0	-1
1	0	-1
1	0	-1

Vertical

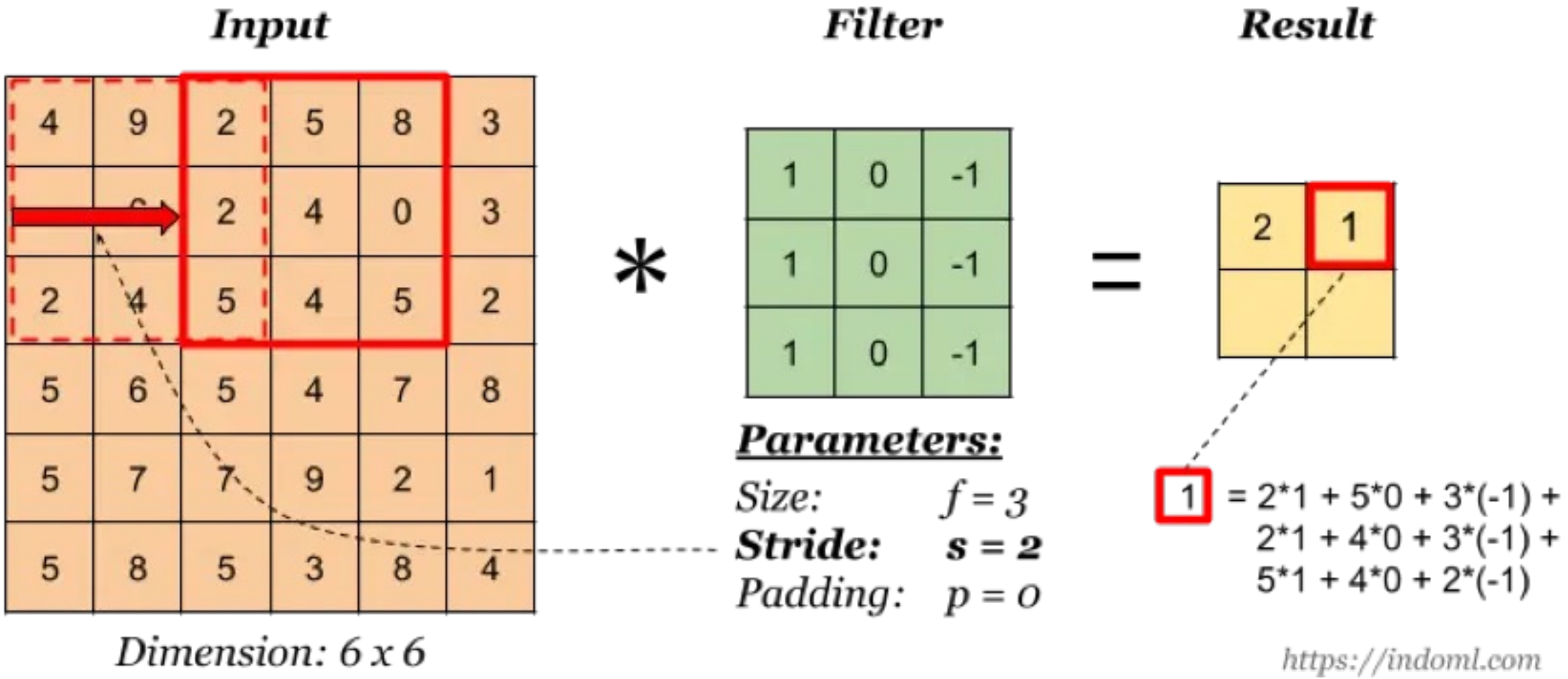
1	1	1
0	0	0
-1	-1	-1

Horizontal

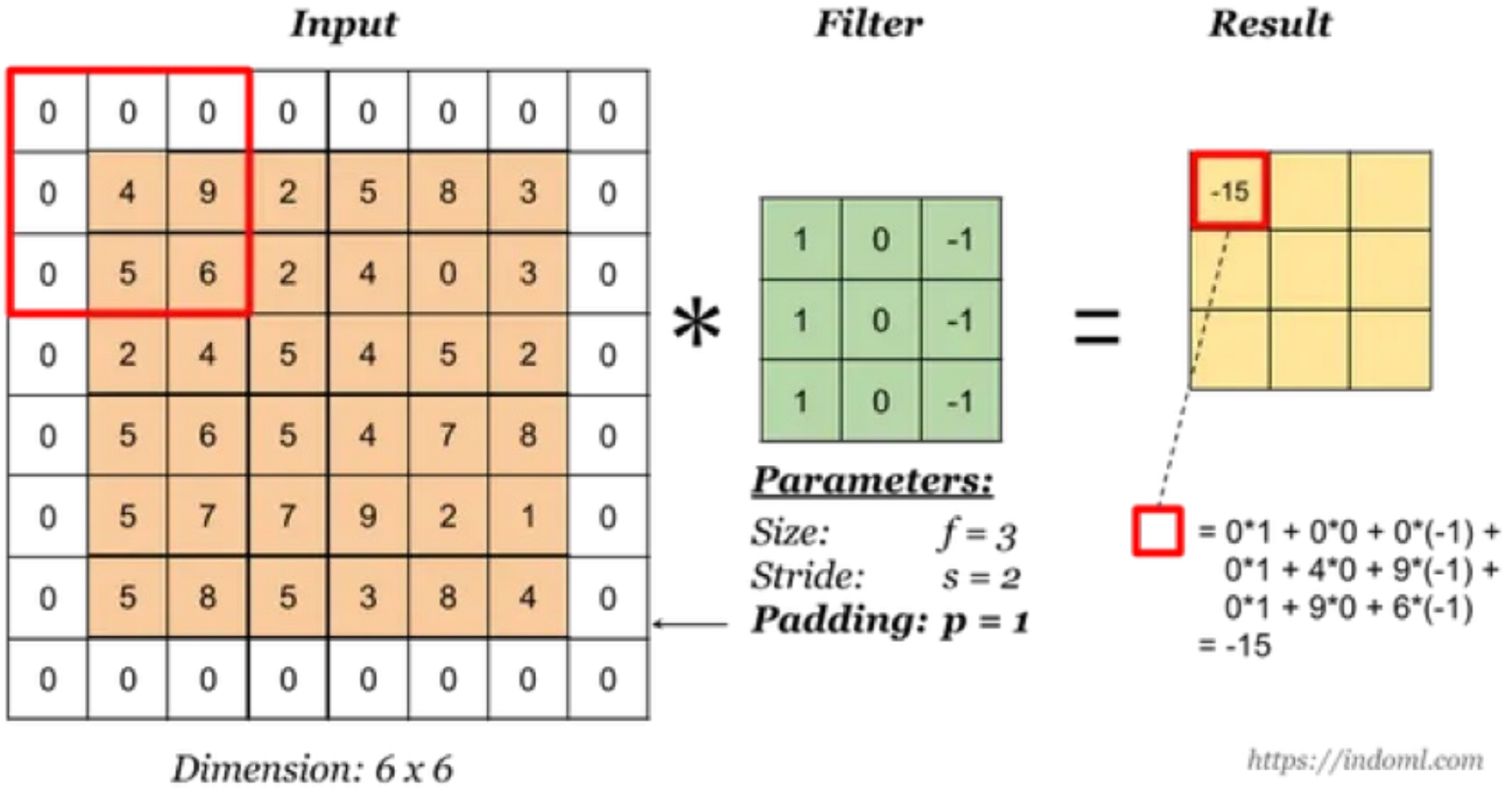
We can also develop other filters to extract curves, circles, etc.

In CNNs, we will let neural network to learn filters by itself; no need to design!

Stride(移动步长)


Input
Filter
Result
4 9 2 5 8 3
2 4 0 3
2 4 5 4 5 2
5 6 5 4 7 8
5 7 7 9 2 1
5 8 5 3 8 4
*
1 0 -1
1 0 -1
1 0 -1
=
2 1
Parameters:
Size: f = 3
Stride: s = 2
Padding: p = 0
Dimension: 6 x 6
1 = 2*1 + 5*0 + 3*(-1) +
2*1 + 4*0 + 3*(-1) +
5*1 + 4*0 + 2*(-1)
https://indoml.com

Padding (填充)


Input
Filter
Result
0 0 0 0 0 0 0 0
0 4 9 2 5 8 3 0
0 5 6 2 4 0 3 0
0 2 4 5 4 5 2 0
0 5 6 5 4 7 8 0
0 5 7 7 9 2 1 0
0 5 8 5 3 8 4 0
0 0 0 0 0 0 0 0
*
1 0 -1
1 0 -1
1 0 -1
=
-15
Parameters:
Size: f = 3
Stride: s = 2
Padding: p = 1
= 0*1 + 0*0 + 0*(-1) +
0*1 + 4*0 + 9*(-1) +
0*1 + 9*0 + 6*(-1)
= -15
Dimension: 6 x 6
https://indoml.com

Padding (填充)

Padding can maintain the influence of the corner elements in input matrix

Padding can help maintain the matrix size

Input size: n x n

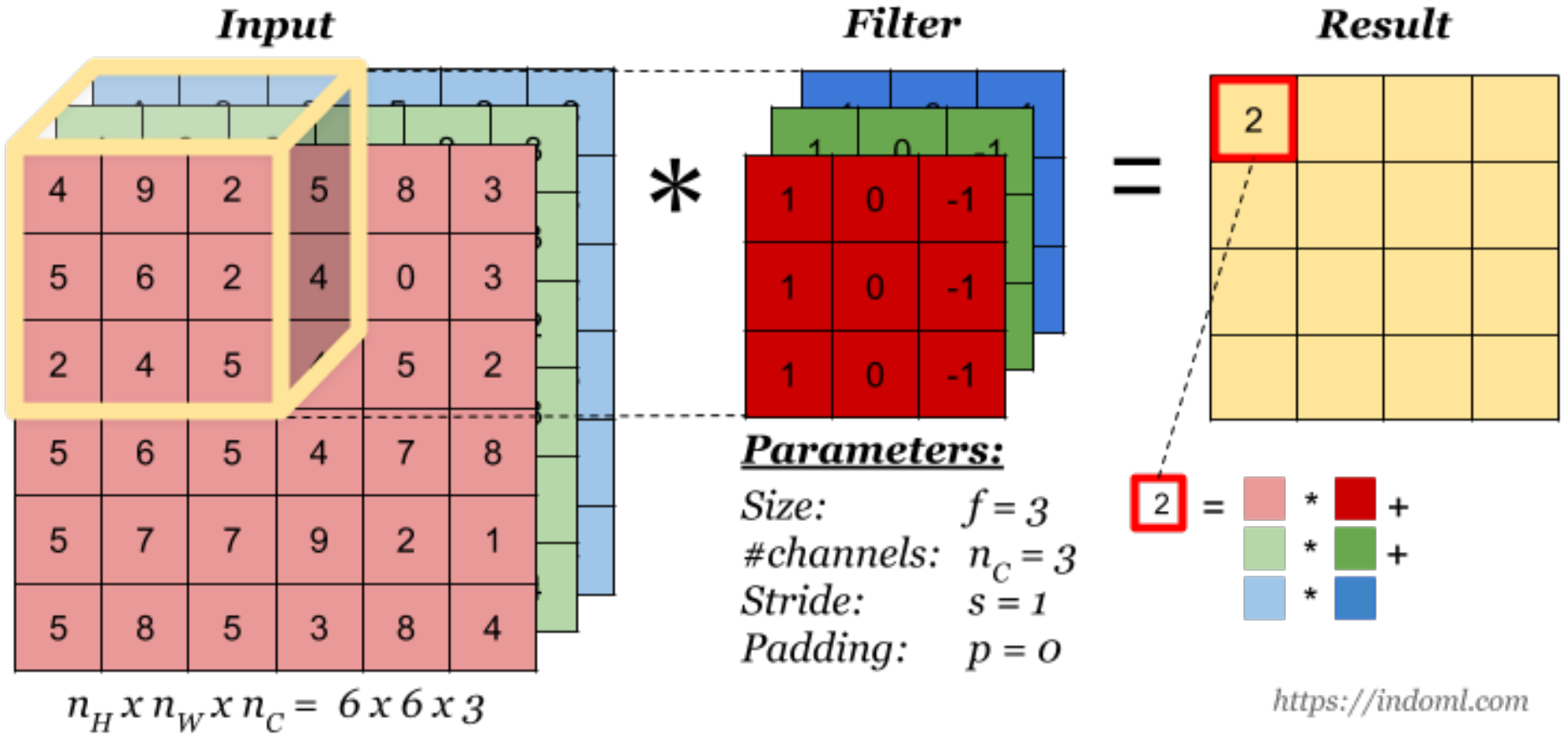
Filter size: f x f

Padding: p

Sride: s

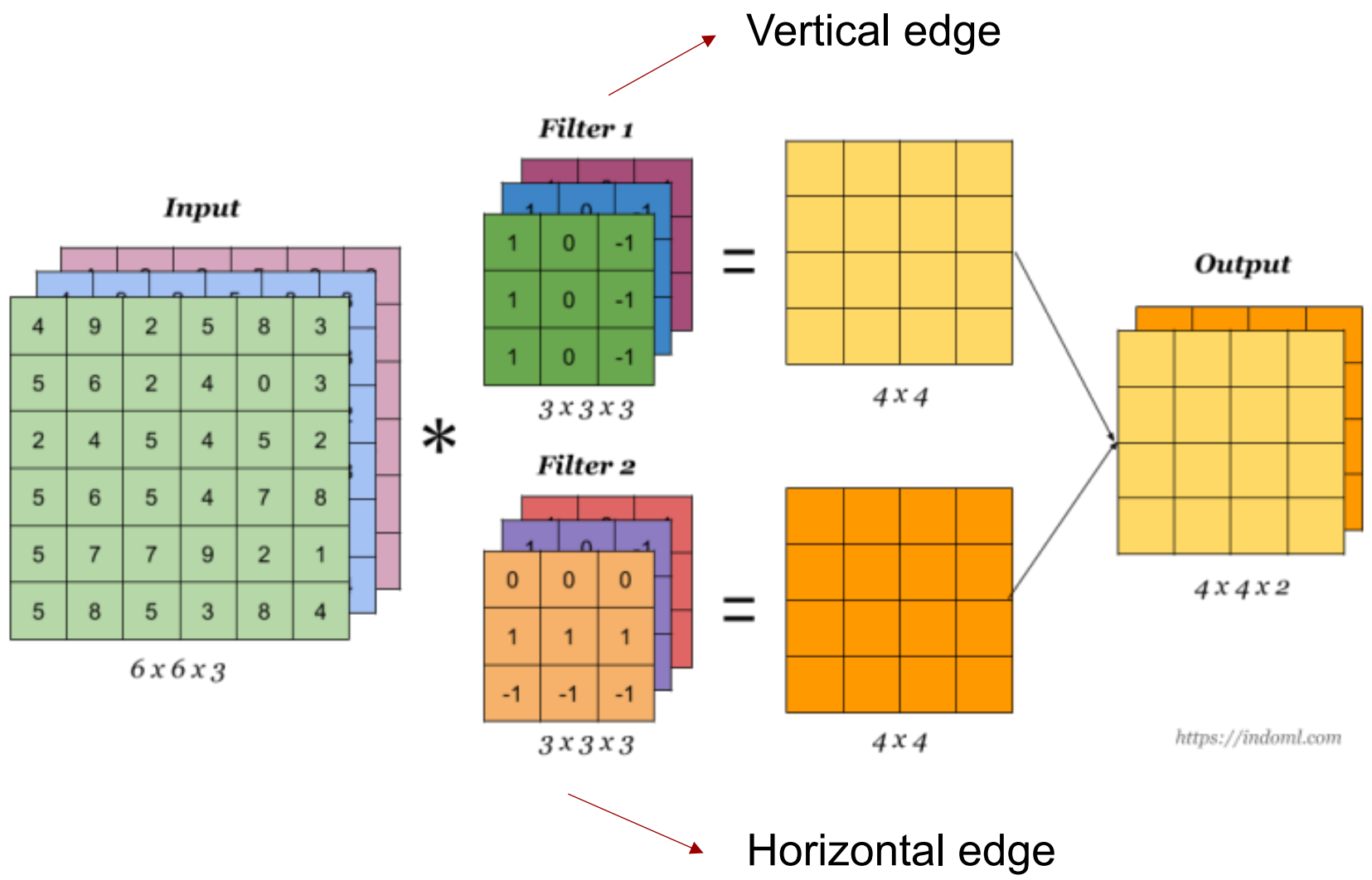
Output size: $\left\lfloor \frac{n+2p-f}{s} + 1 \right\rfloor$

3D convolution

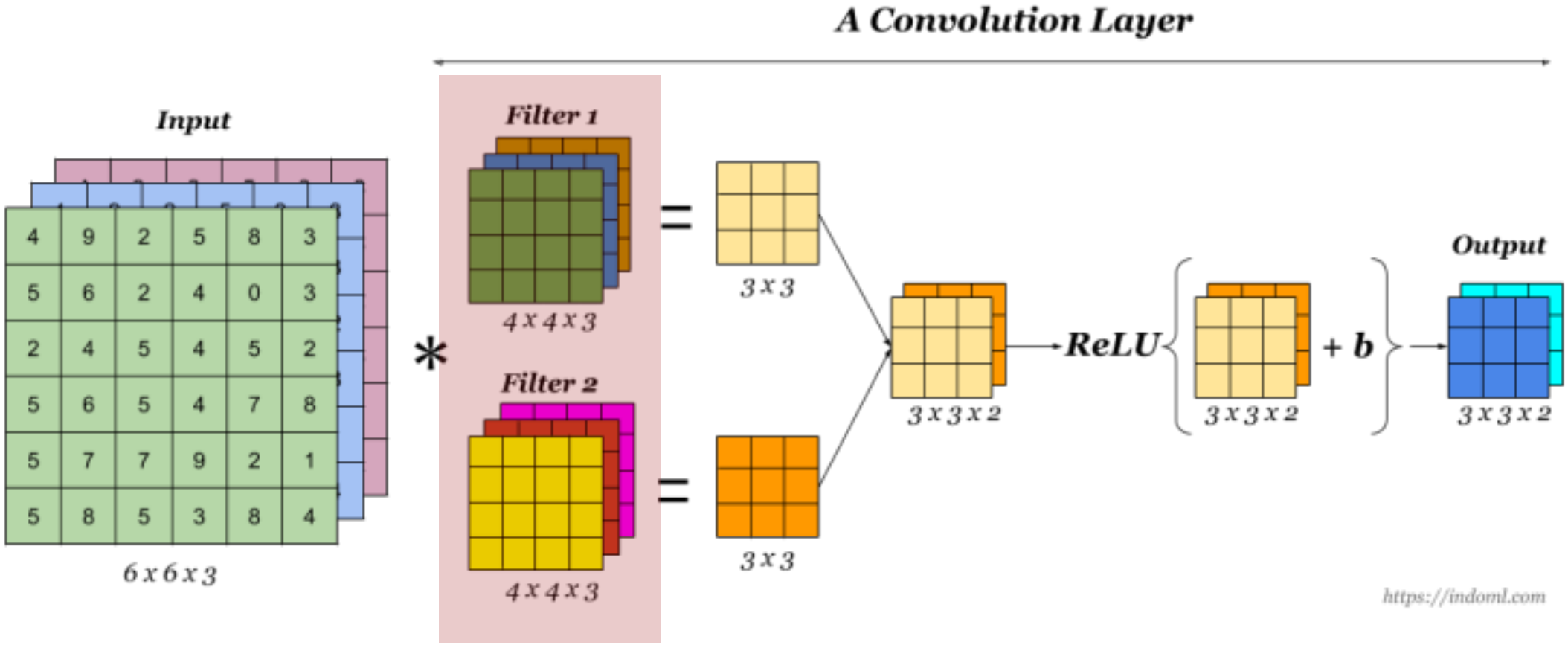


Filters in different channels can be different

Use multiple filters to extract different features

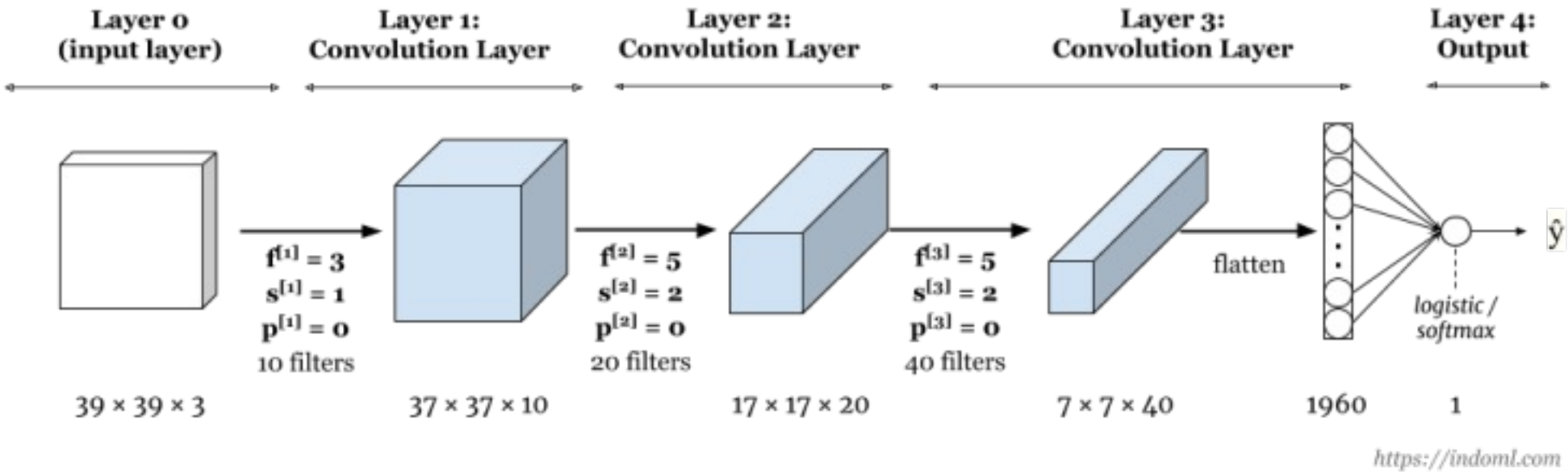


One convolutional layer



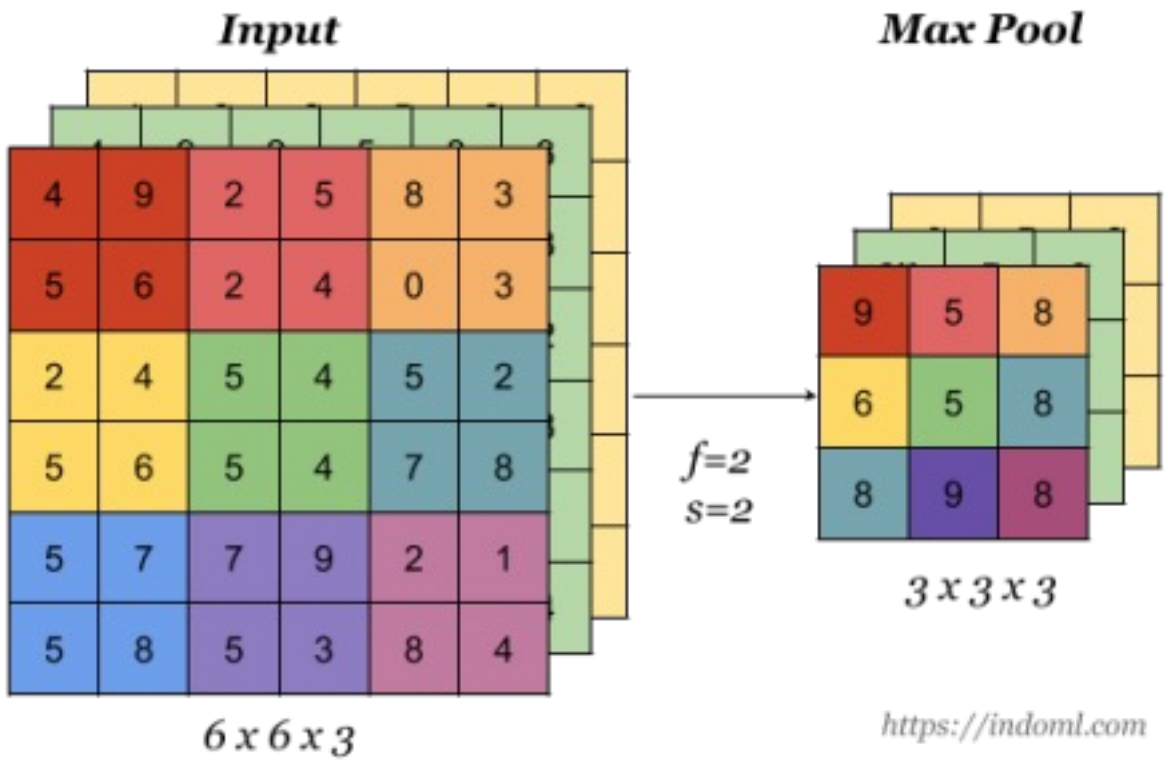
Parameters to learn

Sample complete network



Pooling

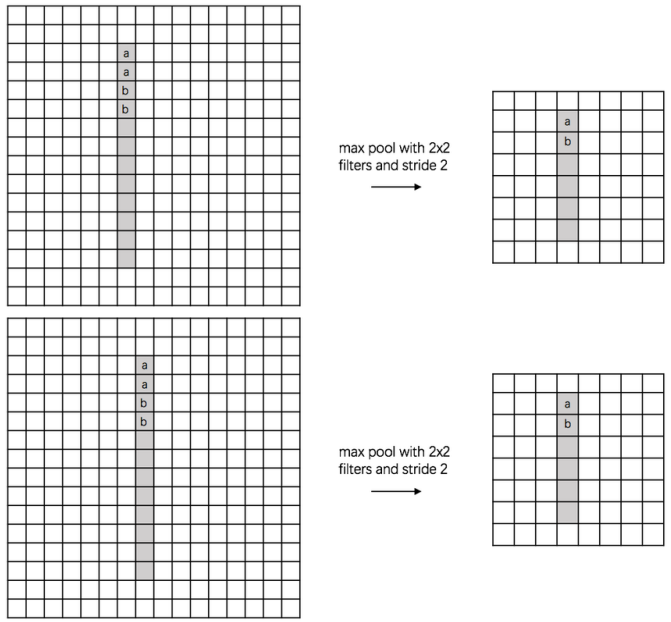
Pooling can reduce the size of representations, speedup calculations, and make feature extraction more robust



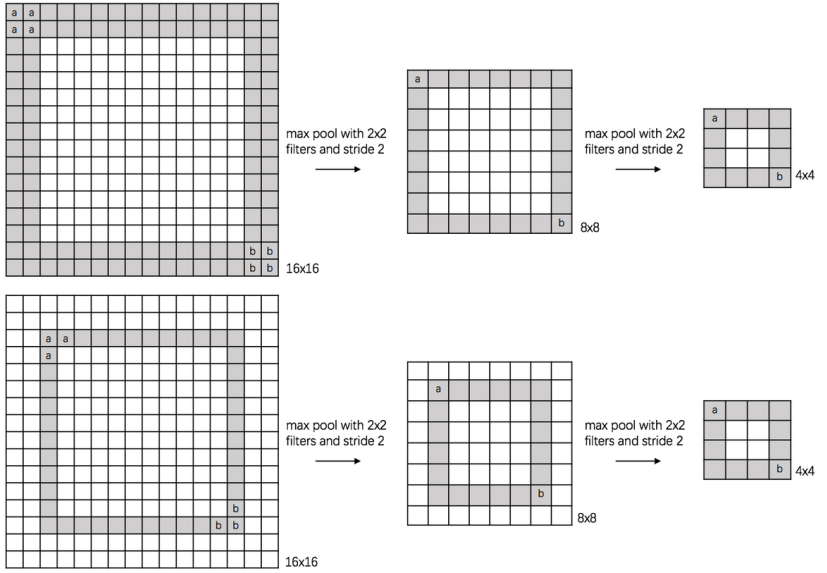
Pooling

Pooling can reduce variance (from 知乎-谢志宁)

https://www.zhihu.com/question/36686900/answer/130890492

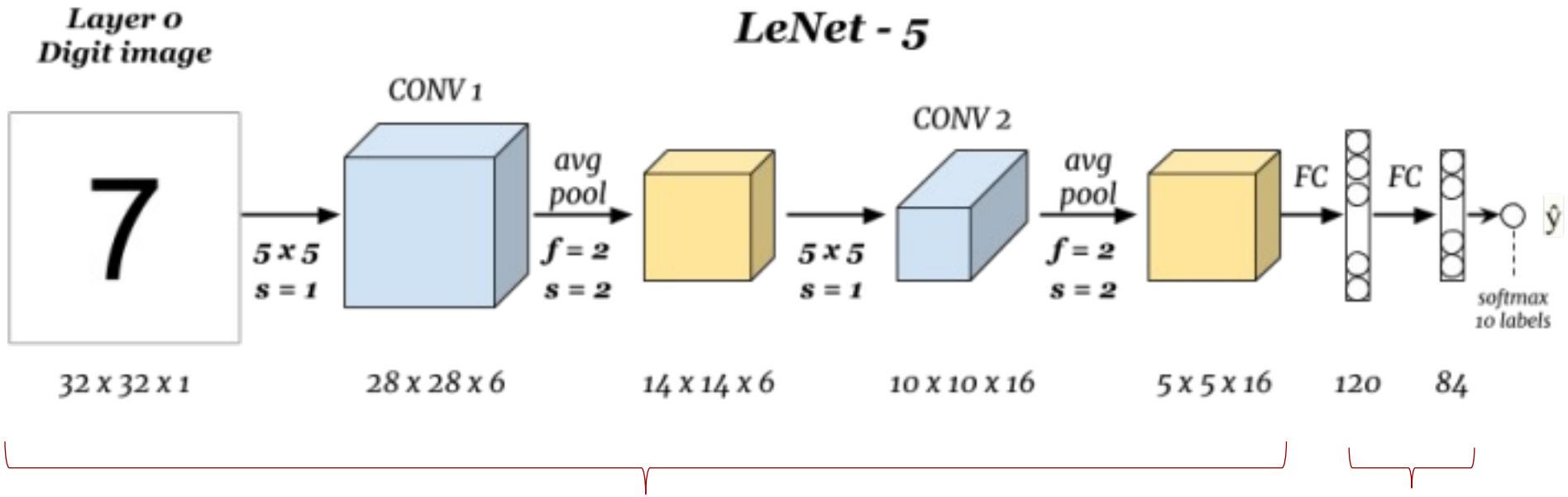


Shift invariance



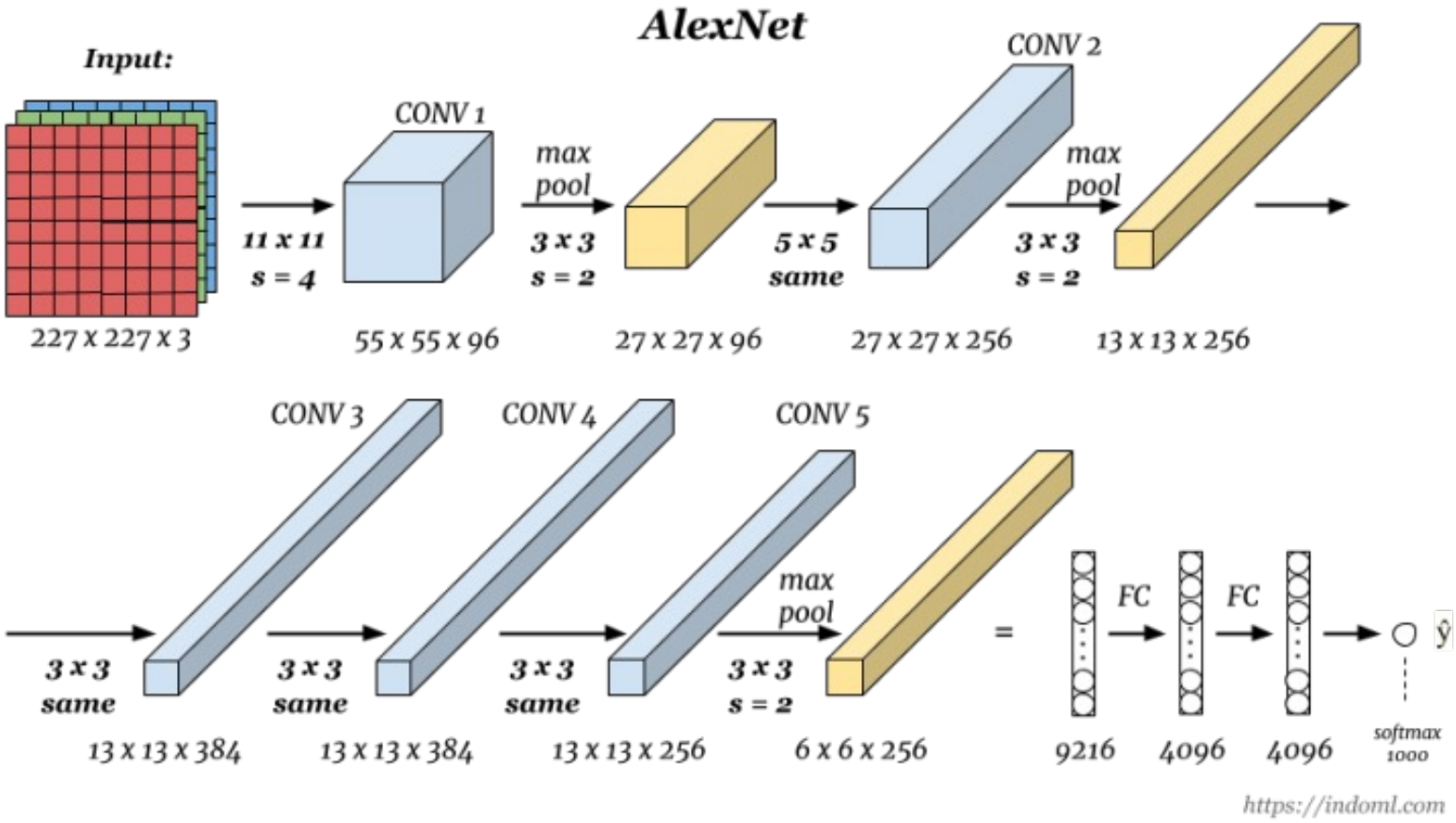
Scale invariance

Well-known architectures: LeNet

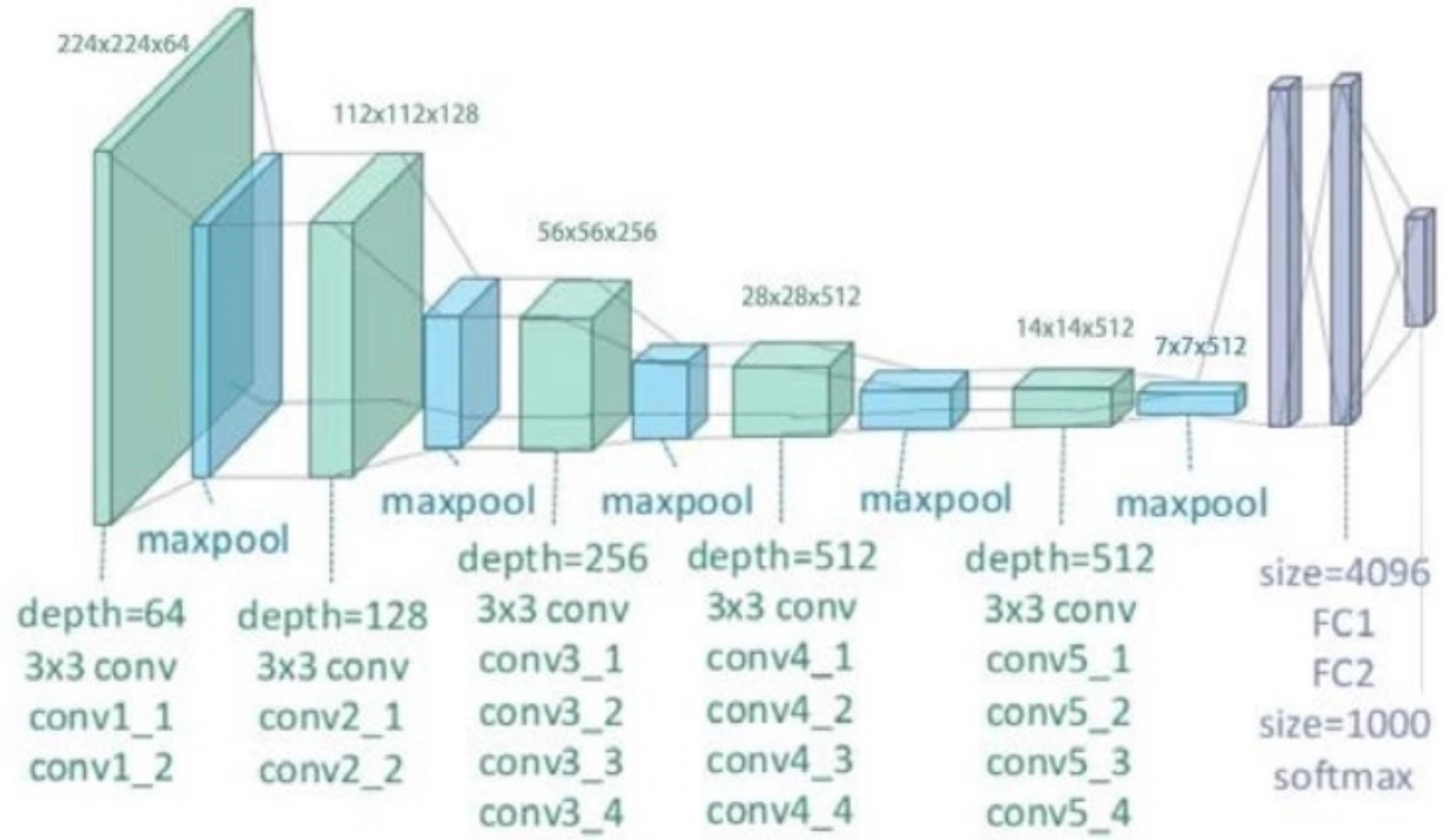


Conv layers only have a few parameters to learn; lightweight **99% weights**

Well-known architectures: AlexNet



Well-known architectures: VGG

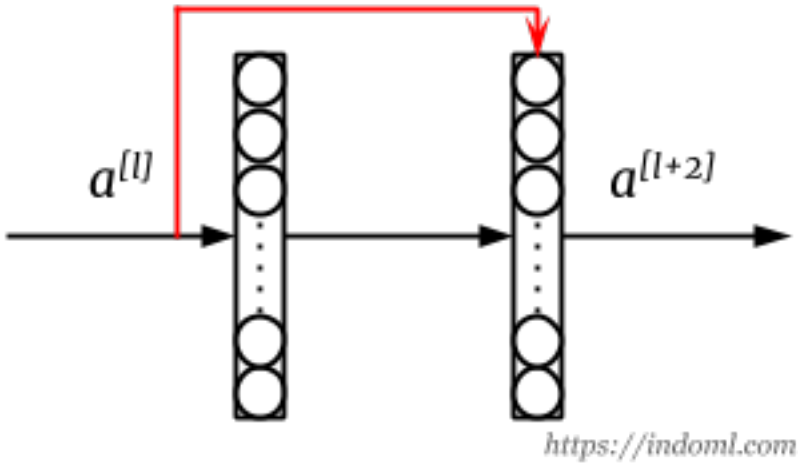


The strength is in the simplicity: the dimension is halved and the depth is increased on every step

ResNet

Deeper neural networks are harder to train; gradient vanishing or exploding

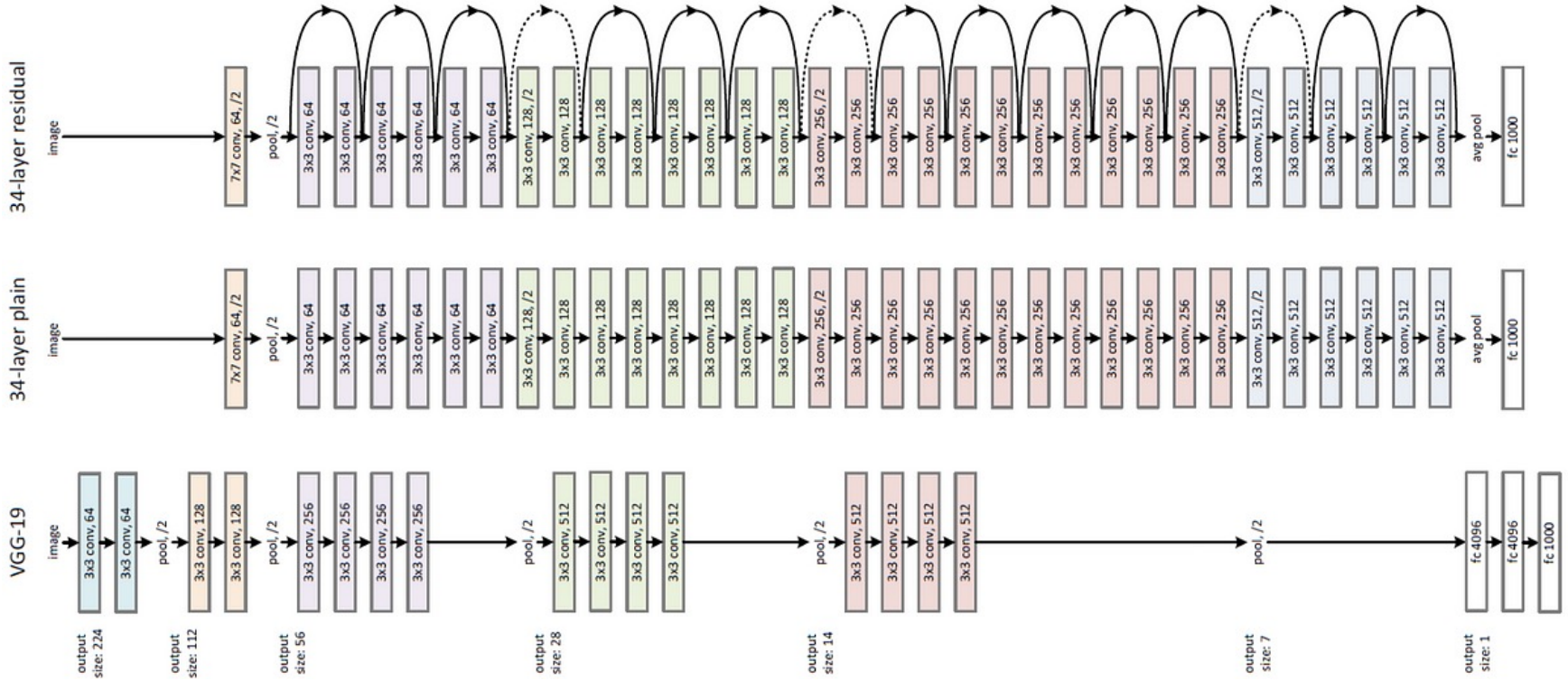
Skip connection helps the gradient to back-propagate



$$z^{[l+2]} = W^{[l+2]} a^{[l+1]} + b^{[l+2]}$$

$$a^{[l+2]} = g^{[l+2]}(z^{[l+2]} + a^{[l]})$$

ResNet



ResNet can train very deep neural networks

Reference

[1] Student Notes: Convolutional Neural Networks (CNN) Introduction

https://indoml.com/2018/03/07/student-notes-convolutional-neural-networks-cnn-introduction/

[2] Andrew Ng, Convolutional Neural Networks

https://www.bilibili.com/video/BV1BF411w7xQ/?spm_id_from=333.337.search-card.all.click