INF721

2024/2

UFV

Deep Learning

L11: Convolutional Neural Networks (Part II)

Logistics

Announcements

- PA3: Convolutional Neural Networks will be out today

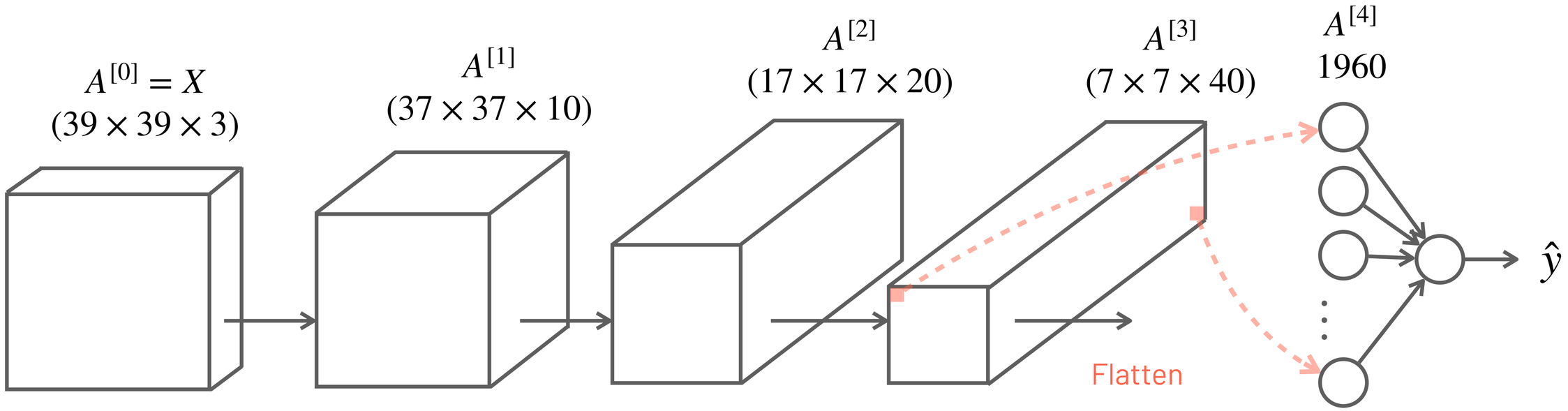
Last Lecture

- Filters
- Convolutions
  - Padding
  - Stride
- Convolutions over volumes
- A first CNN architecture

Lecture Outline

- Pooling Layers
  - Max pooling
  - Average pooling
- Classic CNNs
  - LeNet-5
  - AlexNet
  - VGG-16
- Residual Neural Networks

CNN for Image Classification



$f^{[1]} = 3$
$s^{[1]} = 1$
$p^{[1]} = 0$
$n^{[1]} = 10$

$f^{[2]} = 5$
$s^{[2]} = 2$
$p^{[2]} = 0$
$n^{[2]} = 20$

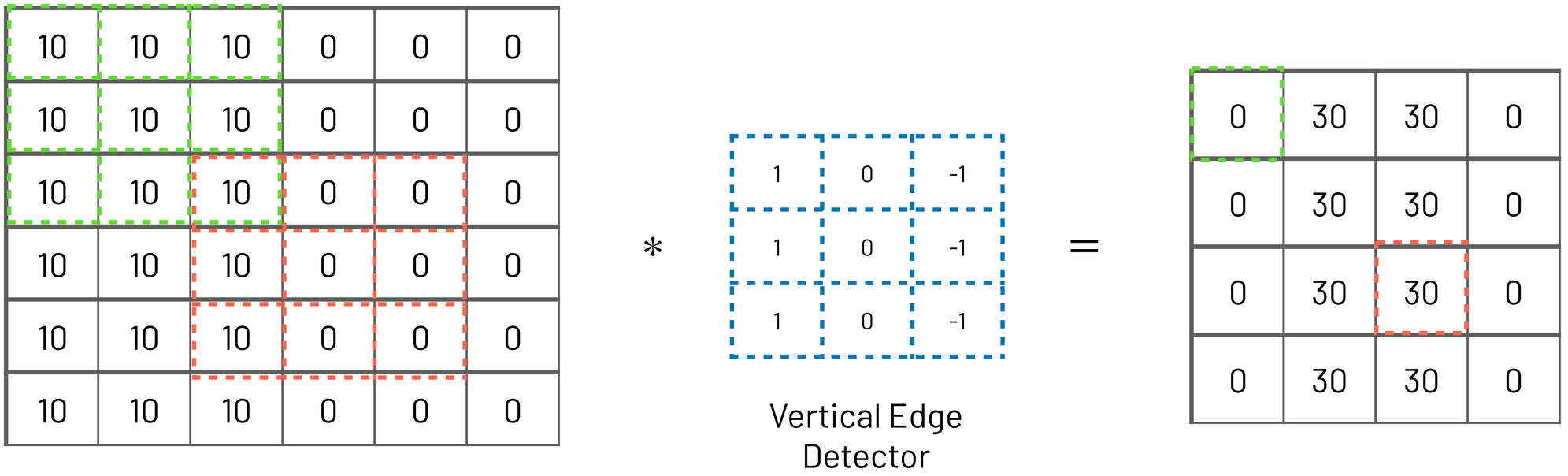
$f^{[3]} = 5$
$s^{[3]} = 2$
$p^{[3]} = 0$
$n^{[3]} = 40$

Notation:

- $f^{[l]}$ size of filters in layer l
- $s^{[l]}$ stride size in layer l
- $p^{[l]}$ padding size in layer l
- $n^{[l]}$ number of filters in layer l

Why Convolutions?

Reducing the number of parameters



Sharing parameters
A filter (e.g., edge detector) that works well in one part of the image will likely work well in another part of the image

Sparse Connections
An ouput depends only on a small number of inputs.

Pooling Layers

Besides convolution layers, CNNs typically also have **pooling layers** to:

1. Reduce dimensions of the feature maps
2. Summarize the features present in a region

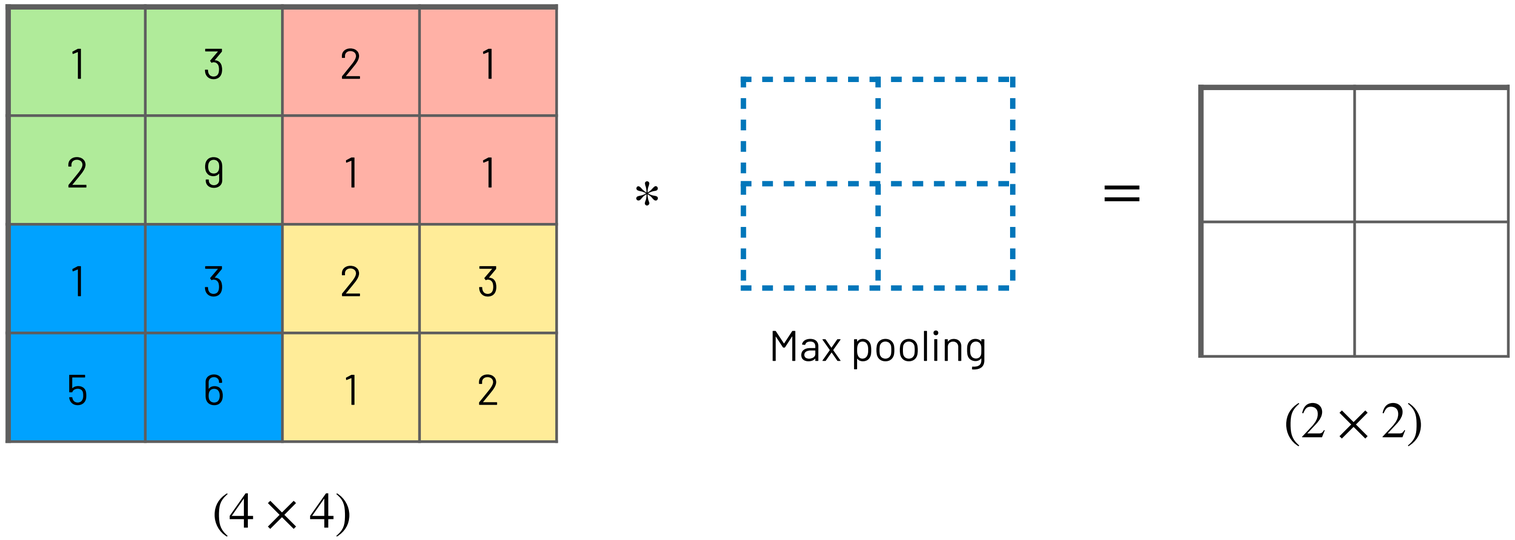
The most common pooling layers in CNNs are:

- Max Pooling
- Average Pooling

These layers perform pre-defined computations and do not have weights to learn!

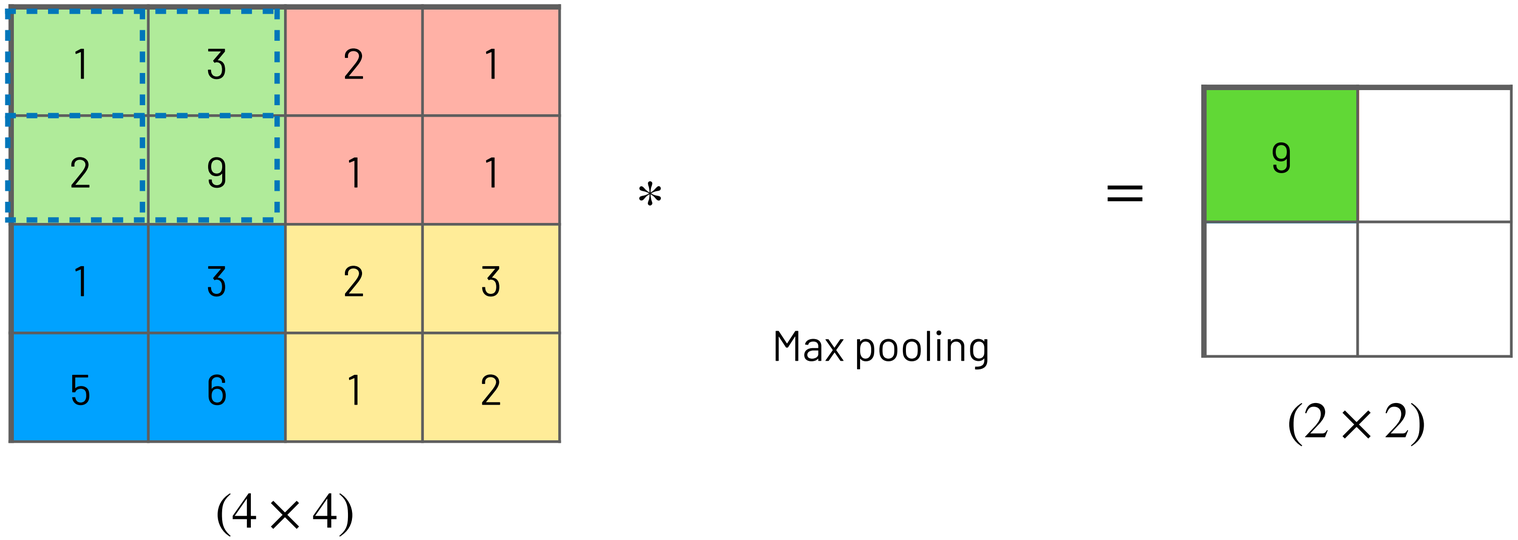
Max pooling

Max pooling is a filter to extract the maximum element of an image region:



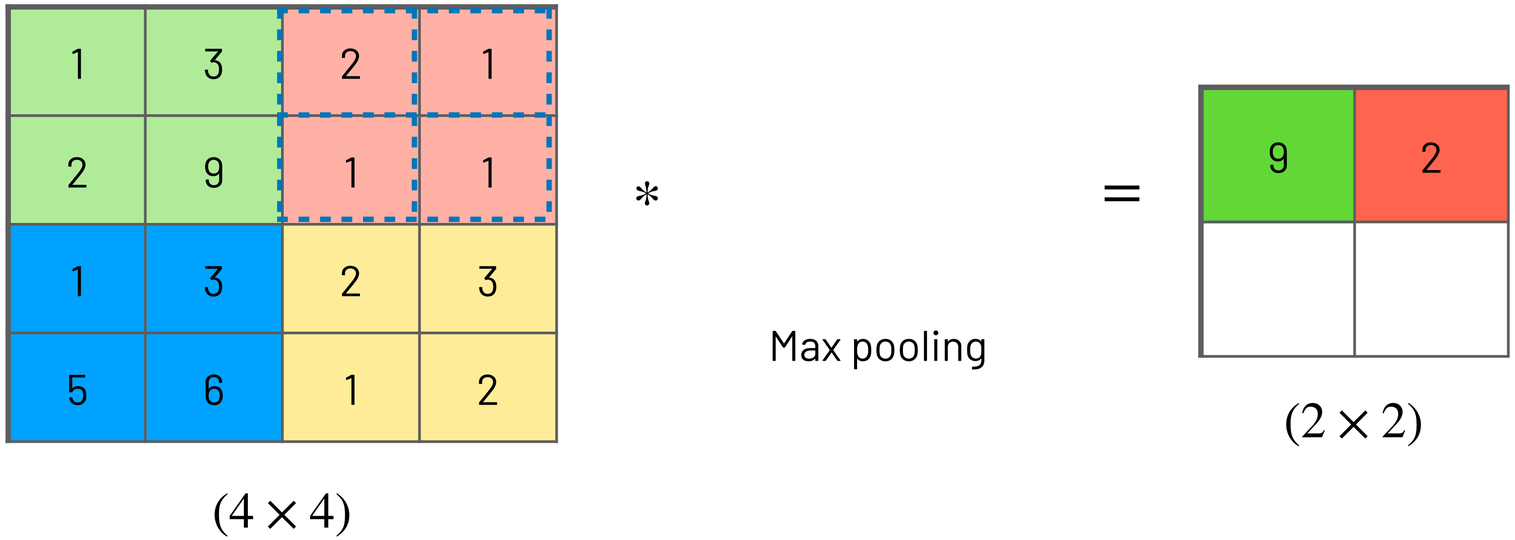
Max pooling

Max pooling is a filter to extract the maximum element of an image region:



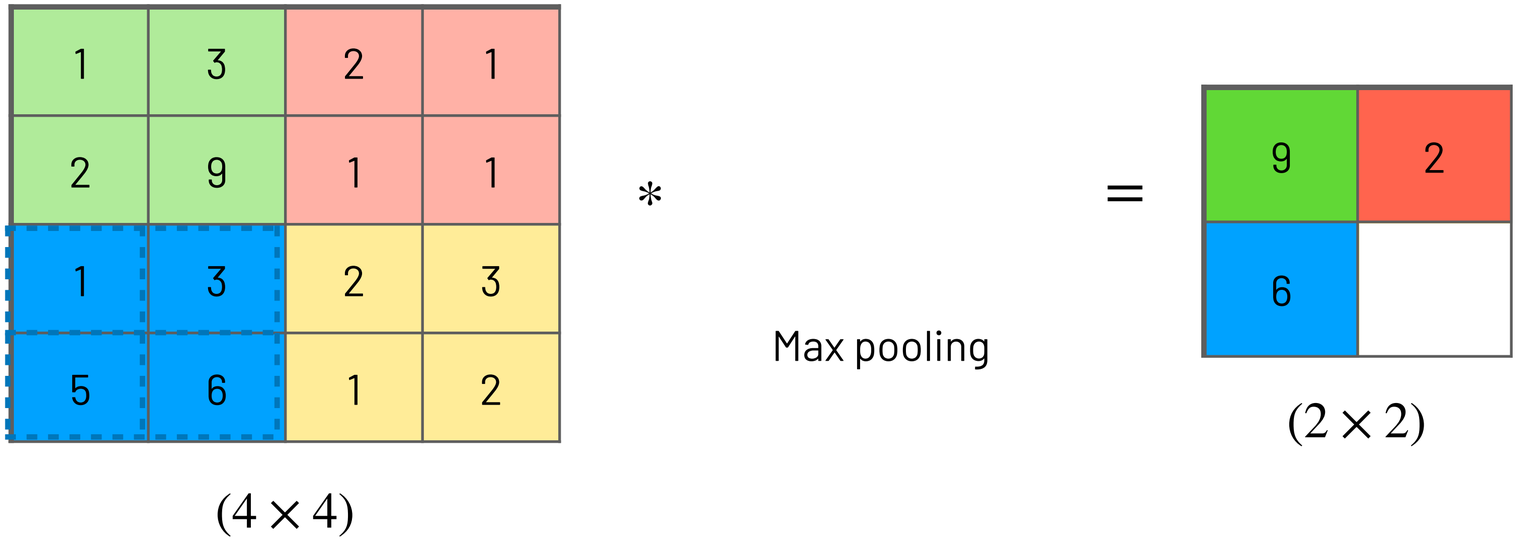
Max pooling

Max pooling is a filter to extract the maximum element of an image region:



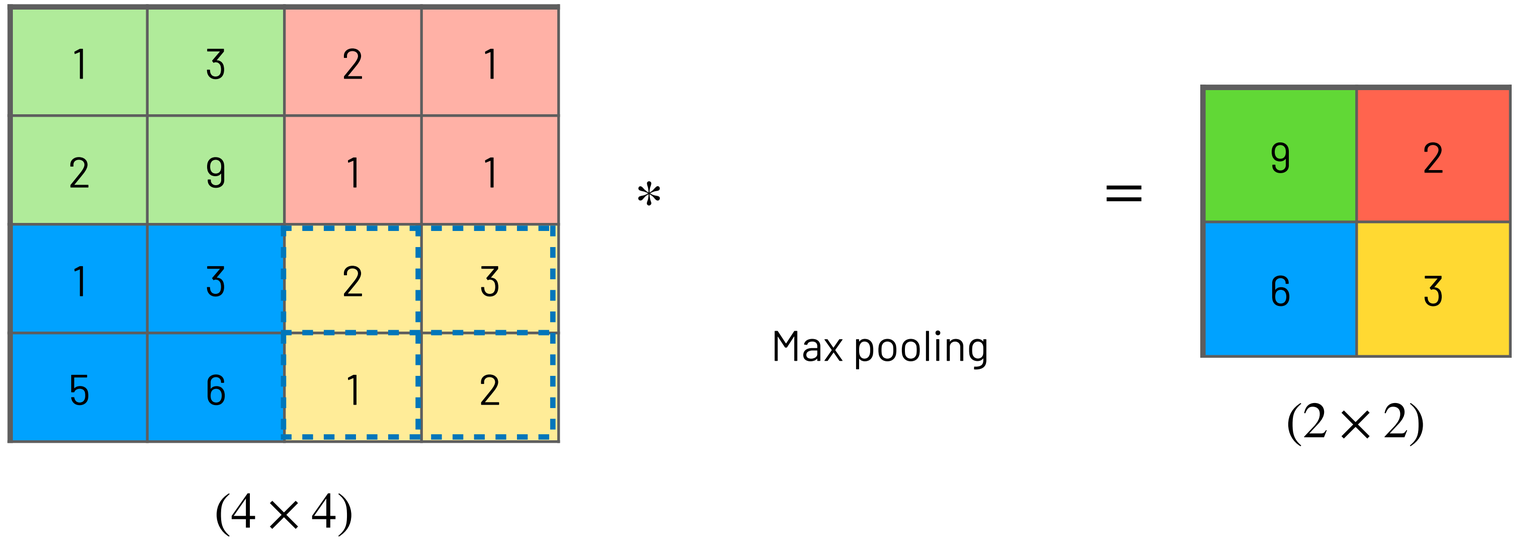
Max pooling

Max pooling is a filter to extract the maximum element of an image region:



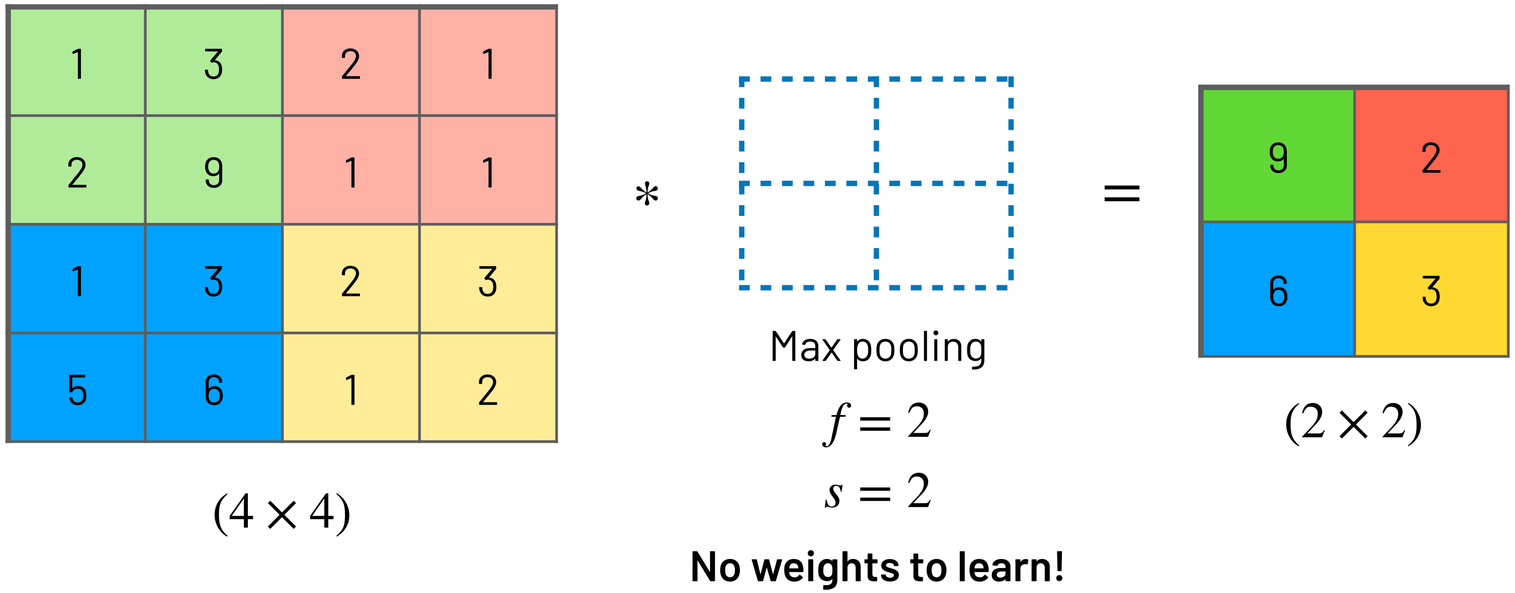
Max pooling

Max pooling is a filter to extract the maximum element of an image region:



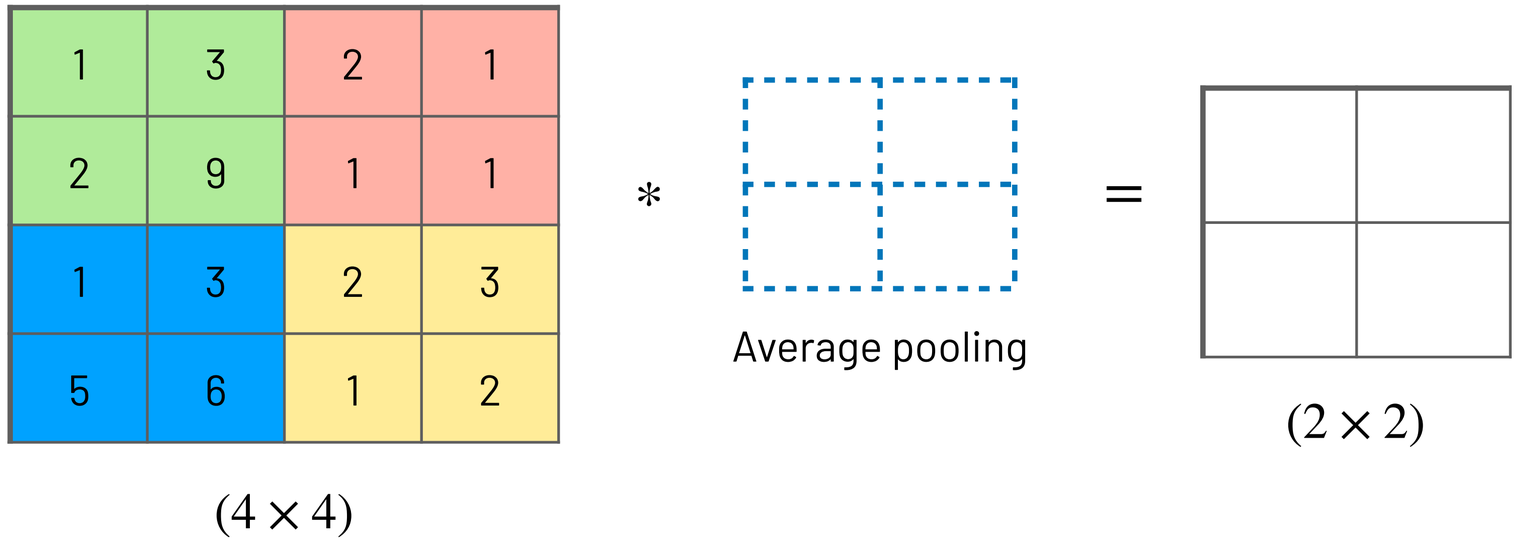
Max pooling

Max pooling is a filter to extract the maximum element of an image region:



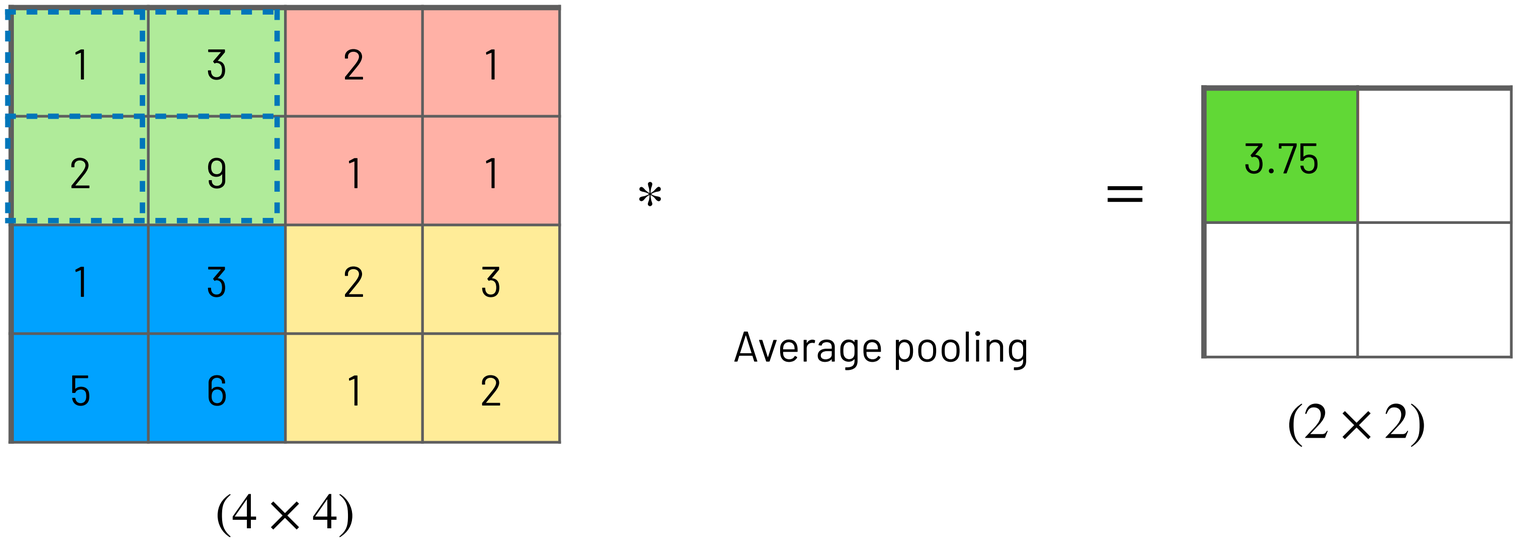
Average pooling

Average pooling is a filter to extract the mean of an image region:



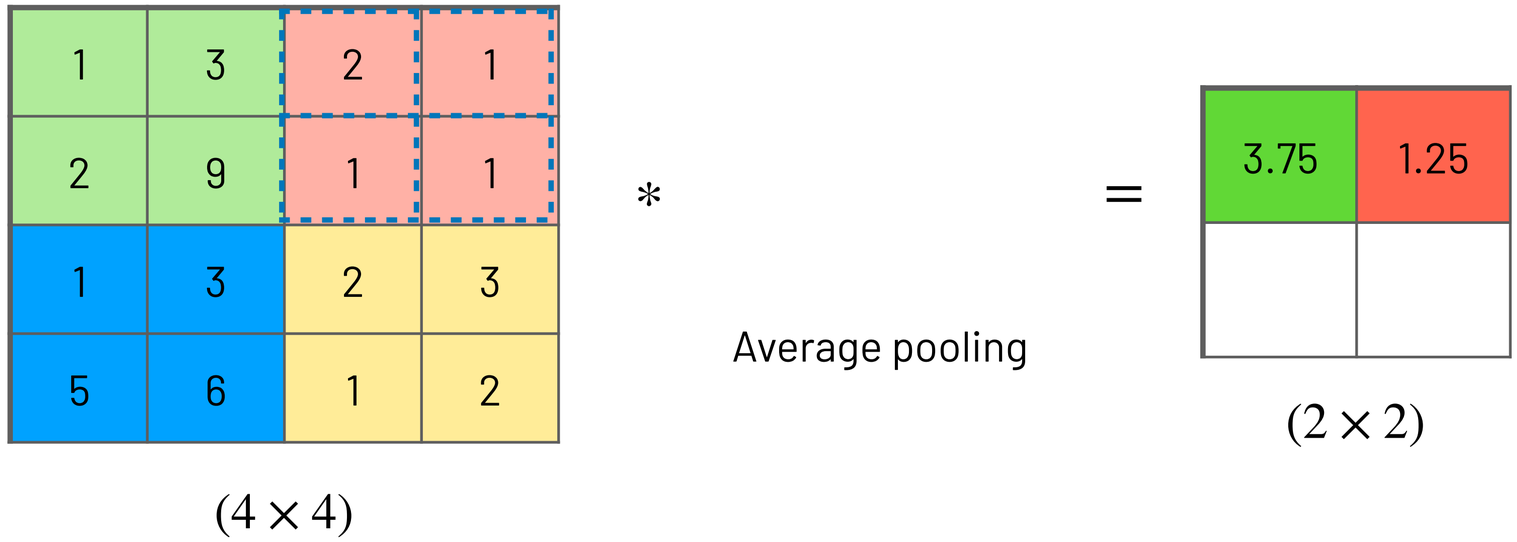
Average pooling

Average pooling is a filter to extract the mean of an image region:



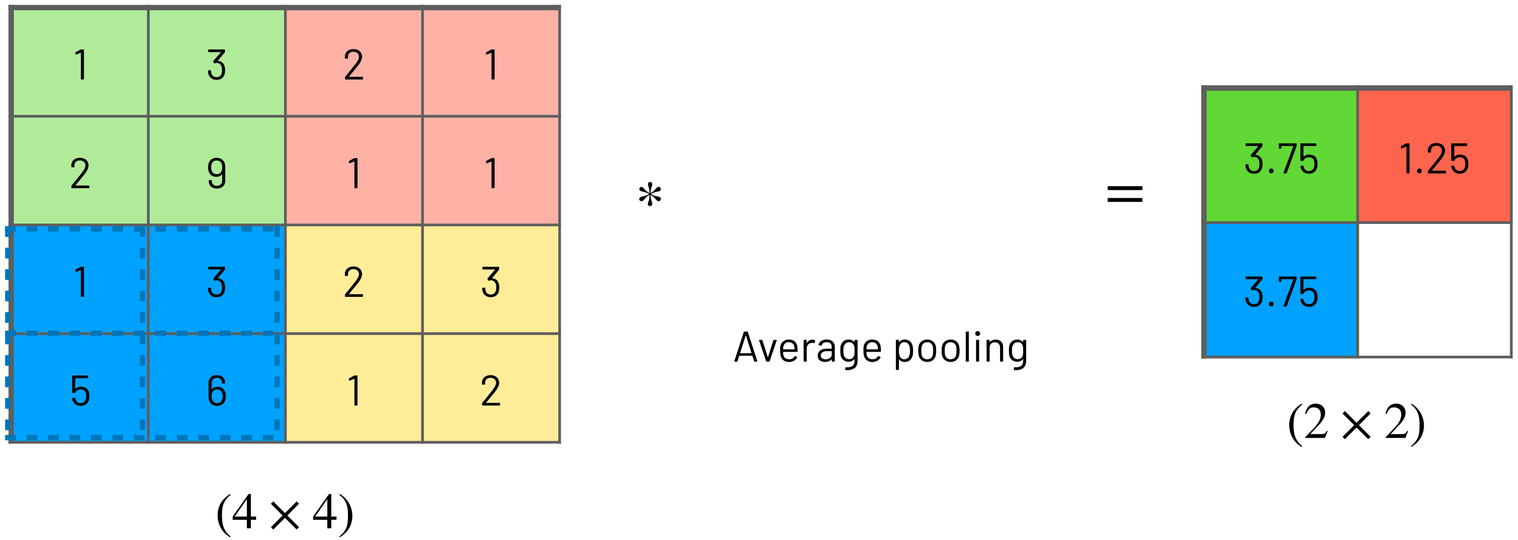
Average pooling

Average pooling is a filter to extract the mean of an image region:



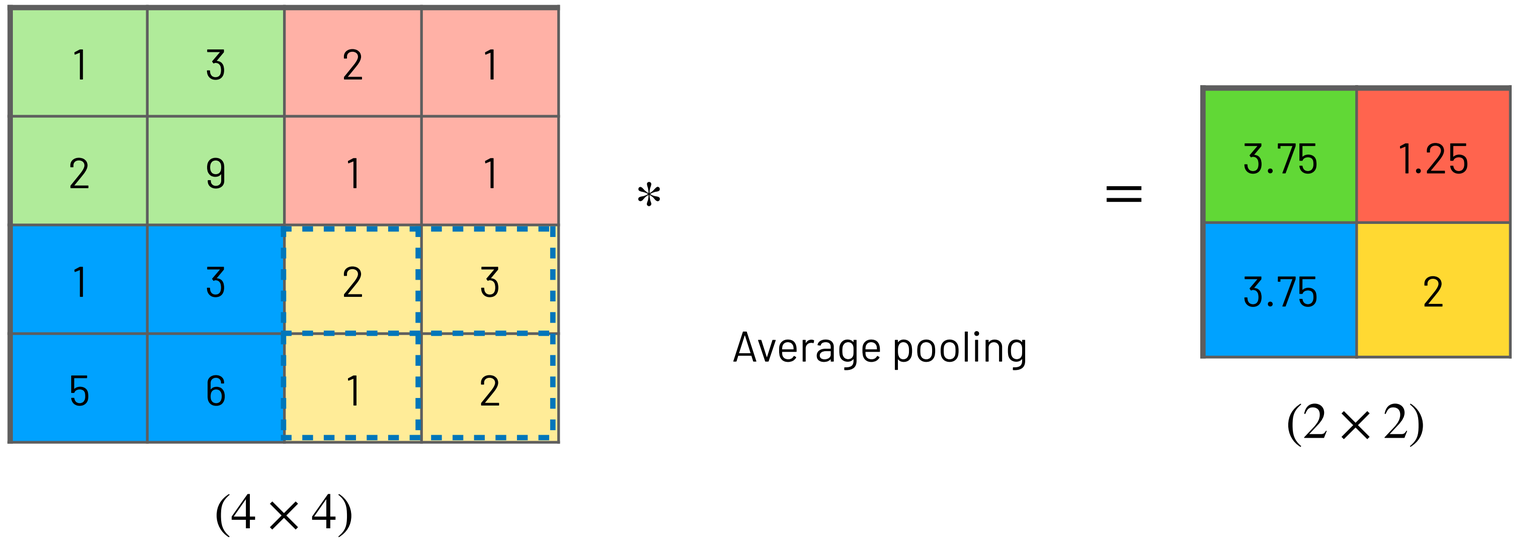
Average pooling

Average pooling is a filter to extract the mean of an image region:



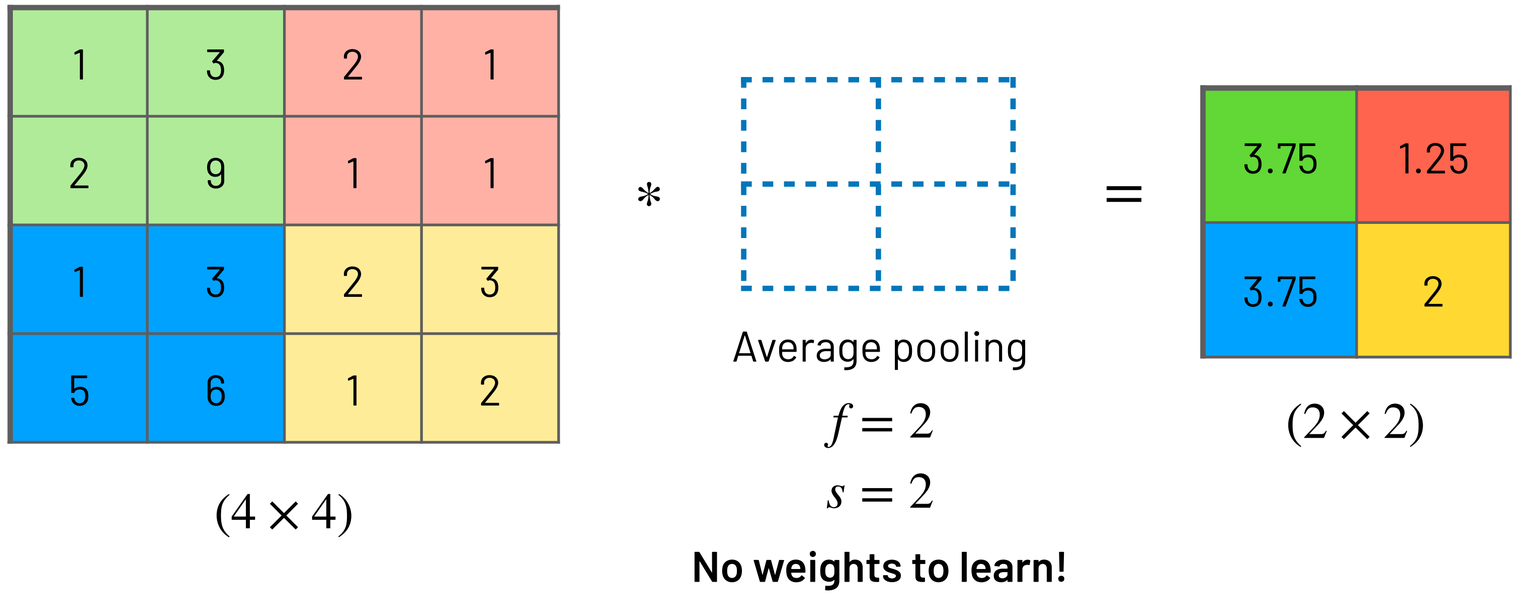
Average pooling

Average pooling is a filter to extract the mean of an image region:



Average pooling

Average pooling is a filter to extract the mean of an image region:



Exercise

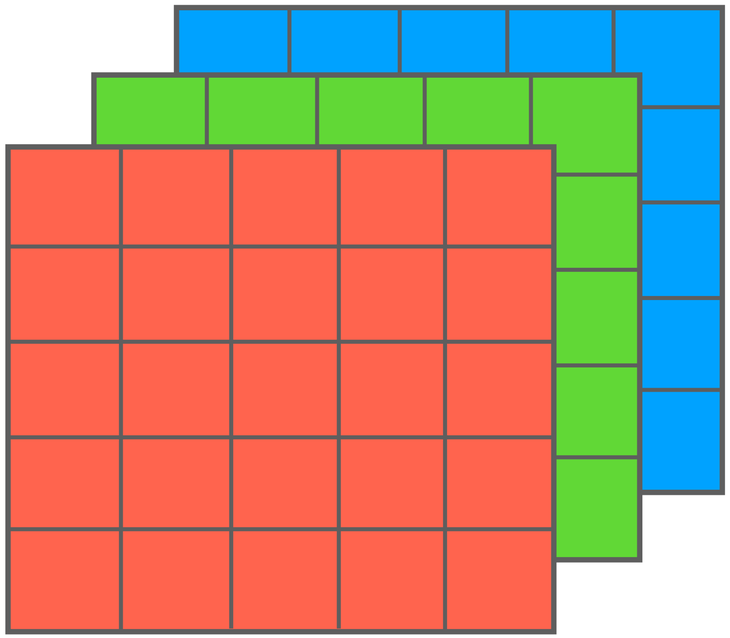
Apply the max pooling filter with $f = 3$ e $s = 1$ to the following image:

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

(5×5)

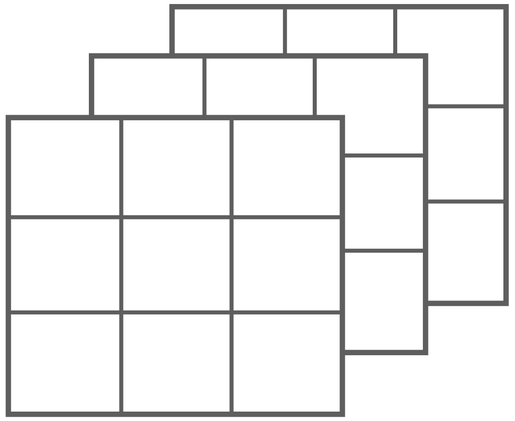
Pooling Layers Over Volumes

Pooling filters are applied to each channel independently:



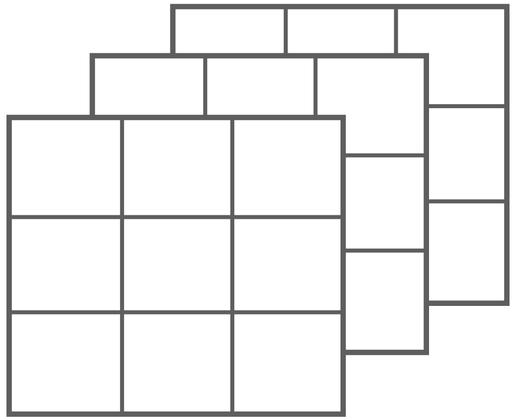
$(5 \times 5 \times 3)$

$*$



$(3 \times 3 \times 3)$

$=$

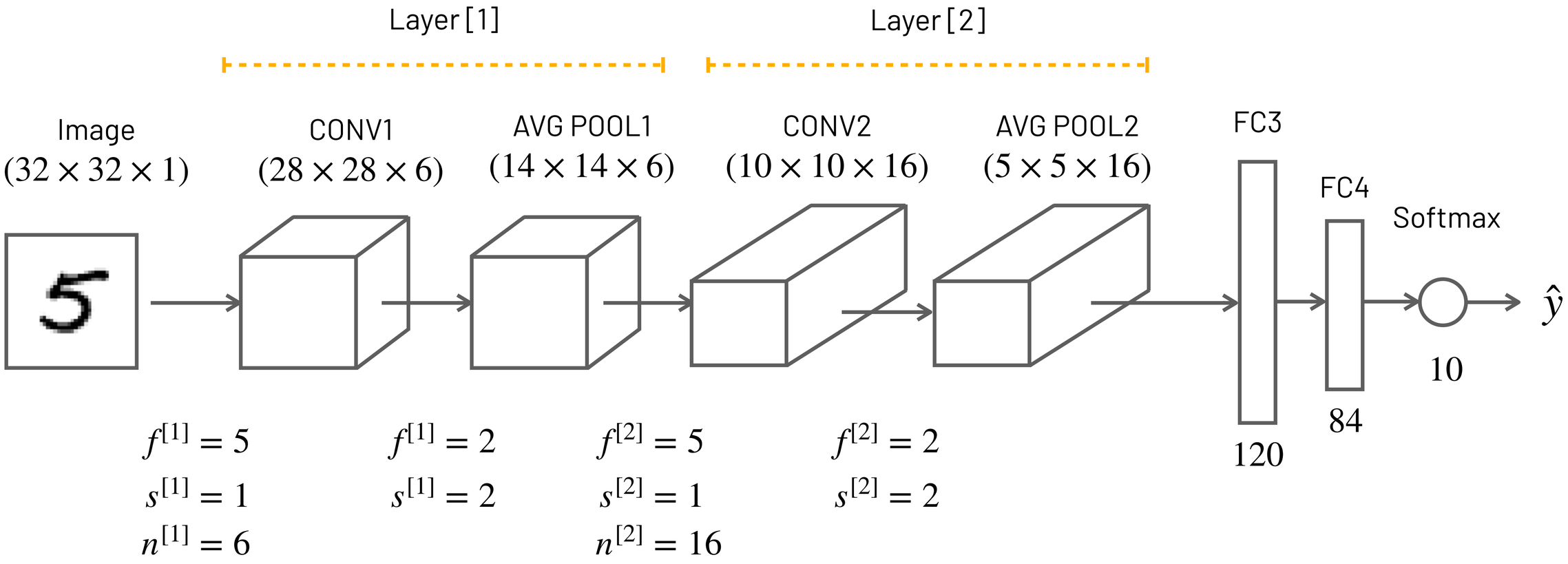


$(3 \times 3 \times 3)$

Pooling in volumnes does not change the number of channels!

LeNet-5

LeNet-5 is a classic CNNs introduced in 1998 by Yann Lecun et al. for handwritten digit classification:



LeCun, Yann, et al. "Gradient-based learning applied to document recognition." *Proceedings of the IEEE* 86.11 (1998): 2278-2324.

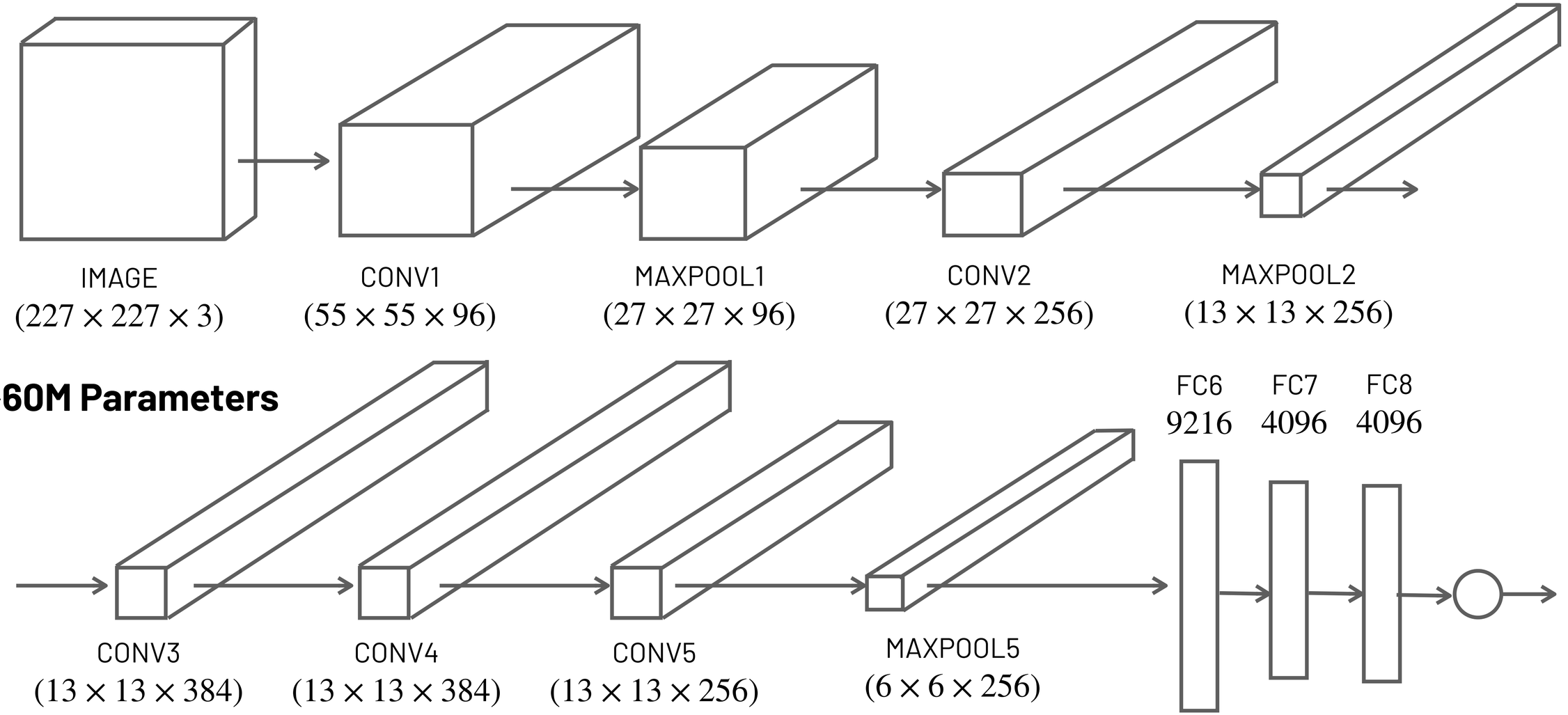
LeNet-5 Analysis

	Activation Shapes	Activation Size	Number of Parameters
Input	(32, 32, 1)	1024	0
CONV1 (f=5, s=1)	(28, 28, 6)	4704	156
AVG POOL1	(14, 14, 6)	1176	0
CONV2 (f=5, s=1)	(10, 10, 16)	1600	416
AVG POOL2	(5, 5, 16)	400	0
FC3	(120, 1)	120	48000
FC4	(84, 1)	84	10080
Softmax	(10,1)	10	840

~60K Parameters

AlexNet

AlexNet won the ImageNet 2012 competition and is considered the Big Ben of Deep Learning

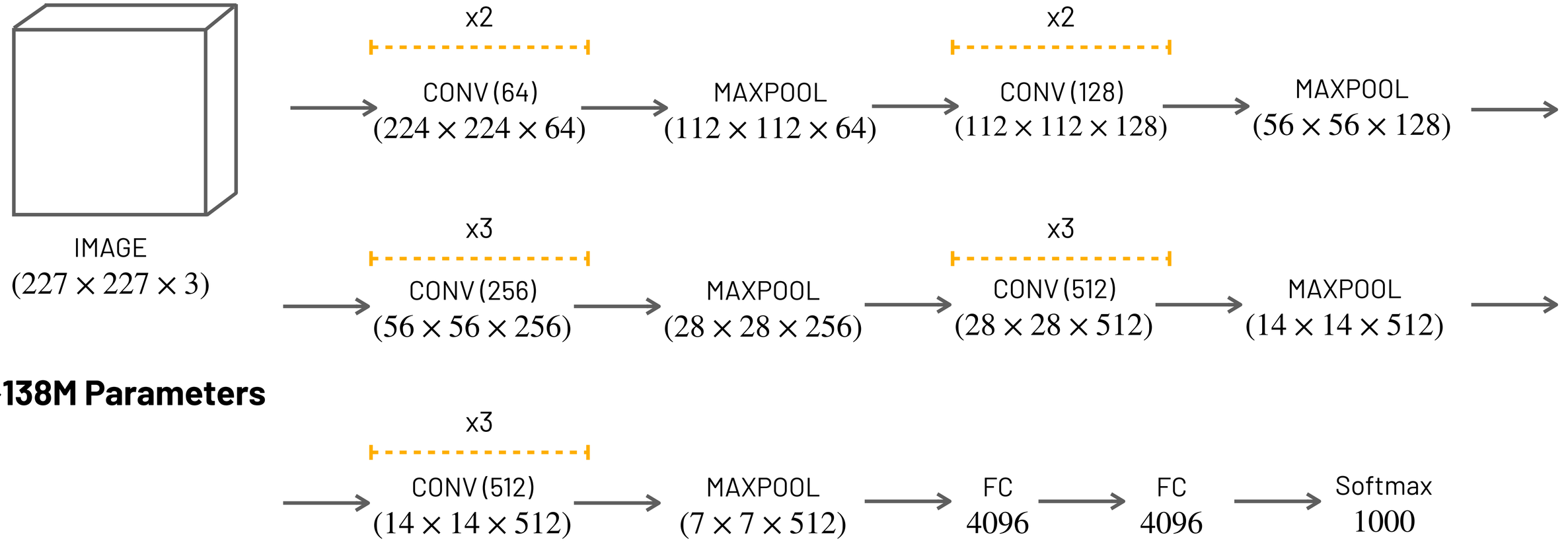


Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. "Imagenet classification with deep convolutional neural networks." Advances in neural information processing systems (2012).

VGG-16

VGG-16 simplified deep CNNs by applying a sequence of CONV + MAXPOOL layers:

CONV	MAXPOOL
$f = 3$	$f = 2$
$s = 1$	$s = 2$
$p = 1$	



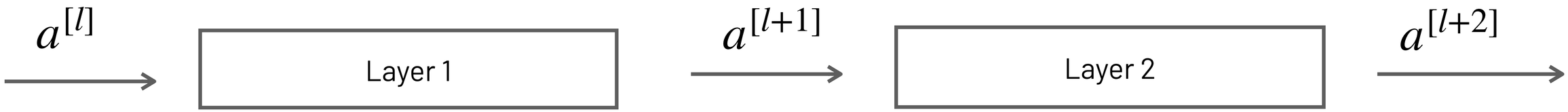
~138M Parameters

Simonyan, K., & Zisserman, A. "Very deep convolutional networks for large-scale image recognition". International Conference on Learning Representations (2015).

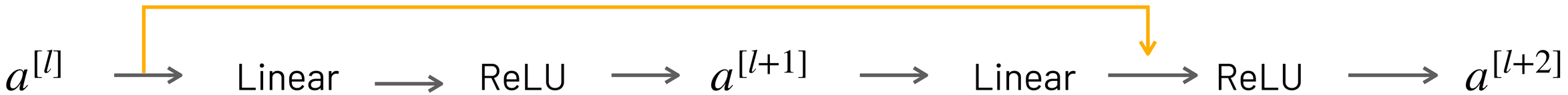
Resnet

ResNet introduced **residual connections** to train much larger CNNs

- **Residual connections** are additional links that connect some layers in a neural network to other layers that are not directly adjacent



skip connection



$$z^{[l+1]} = W^{[l+1]}a^{[l]} + b^{[l+1]} \qquad a^{[l+1]} = g(z^{[l+1]}) \qquad z^{[l+2]} = W^{[l+2]}a^{[l+1]} + b^{[l+2]}$$

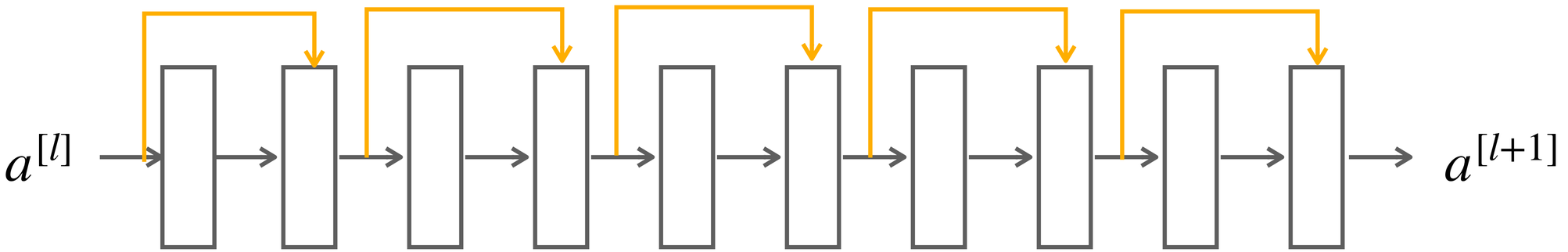
~~$a^{[l+2]} = g(z^{[l+2]})$~~

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

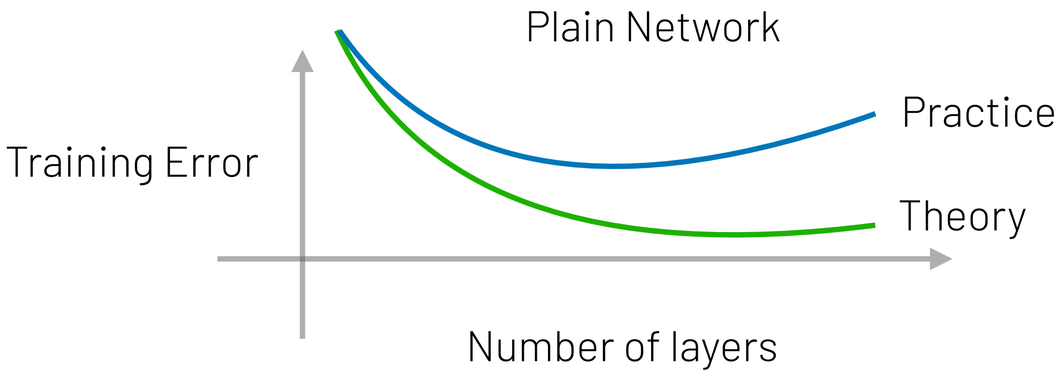
He, Kaiming, et al. "Deep residual learning for image recognition." Proceedings of the IEEE conference on computer vision and pattern recognition. 2016

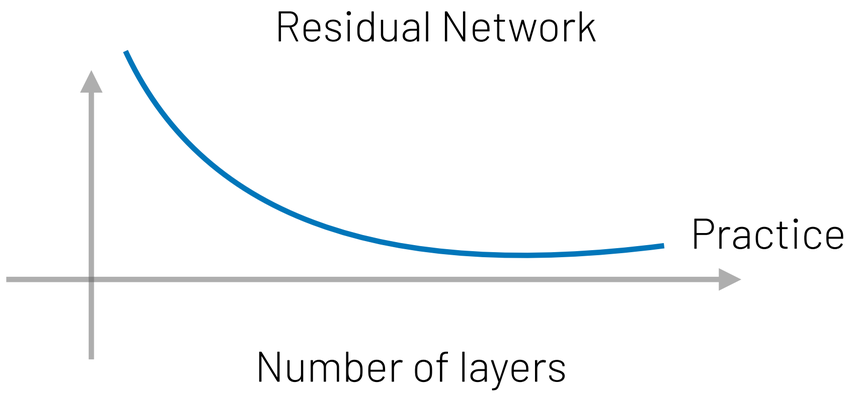
Resnet

A **ResNet** is a stack of residual blocks (layers with residual connections)



Residual blocks help dealing with the problem of exploding/vanishing gradients:





He, Kaiming, et al. "Deep residual learning for image recognition." Proceedings of the IEEE conference on computer vision and pattern recognition. 2016

Next Lecture

L12: Recurrent Neural Networks

Sequential problems, basic recurrent neural networks, backpropagation through time, one-hot encodging, language models