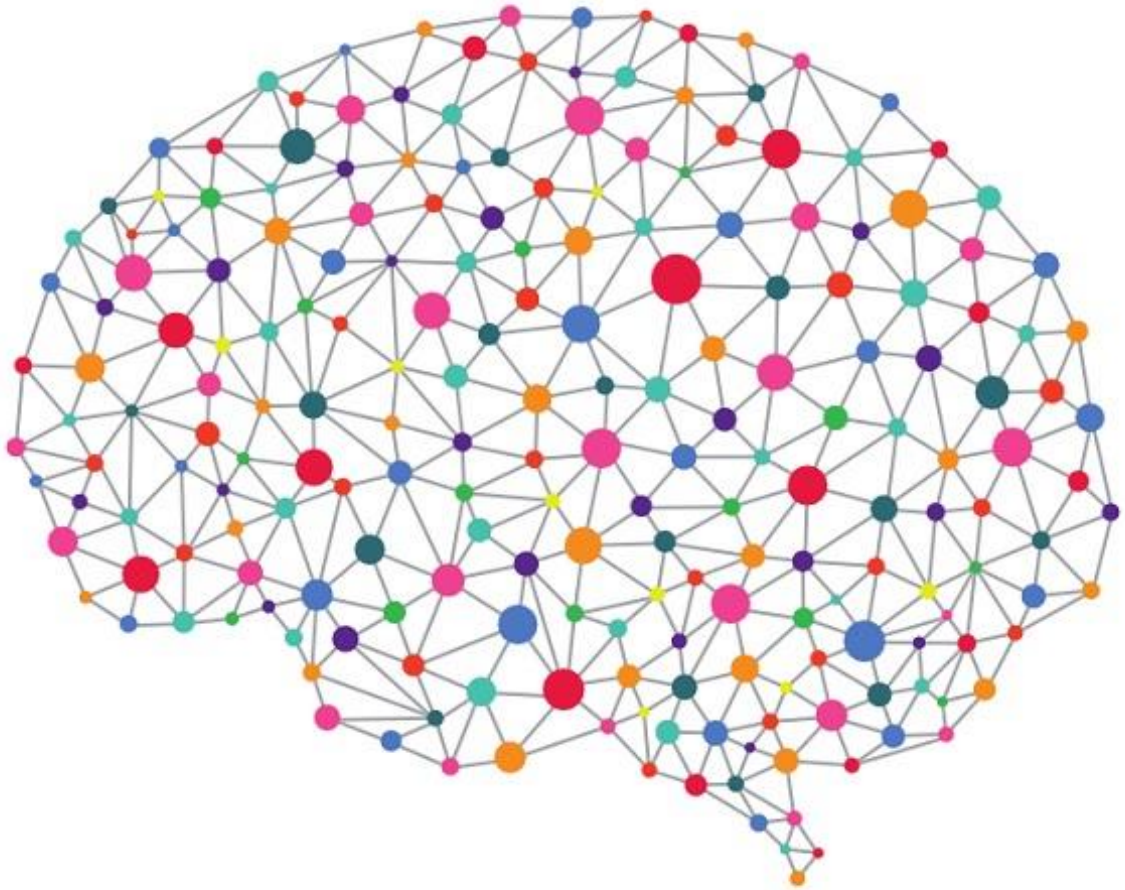
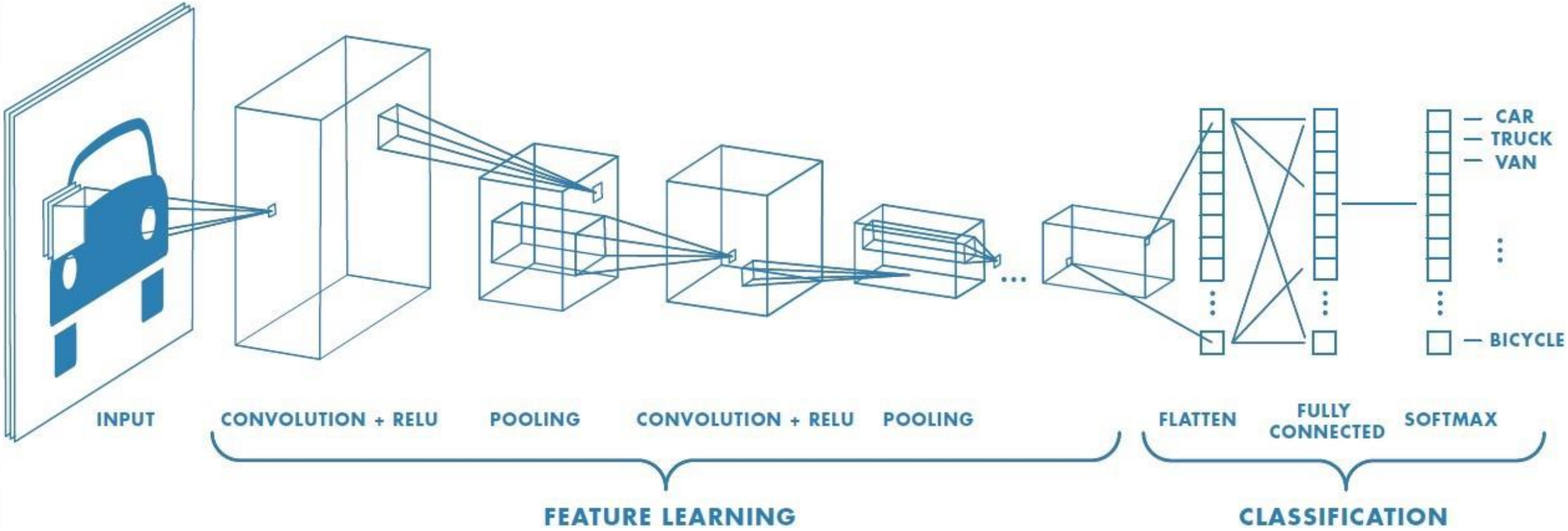
Deep Learning For beginners!

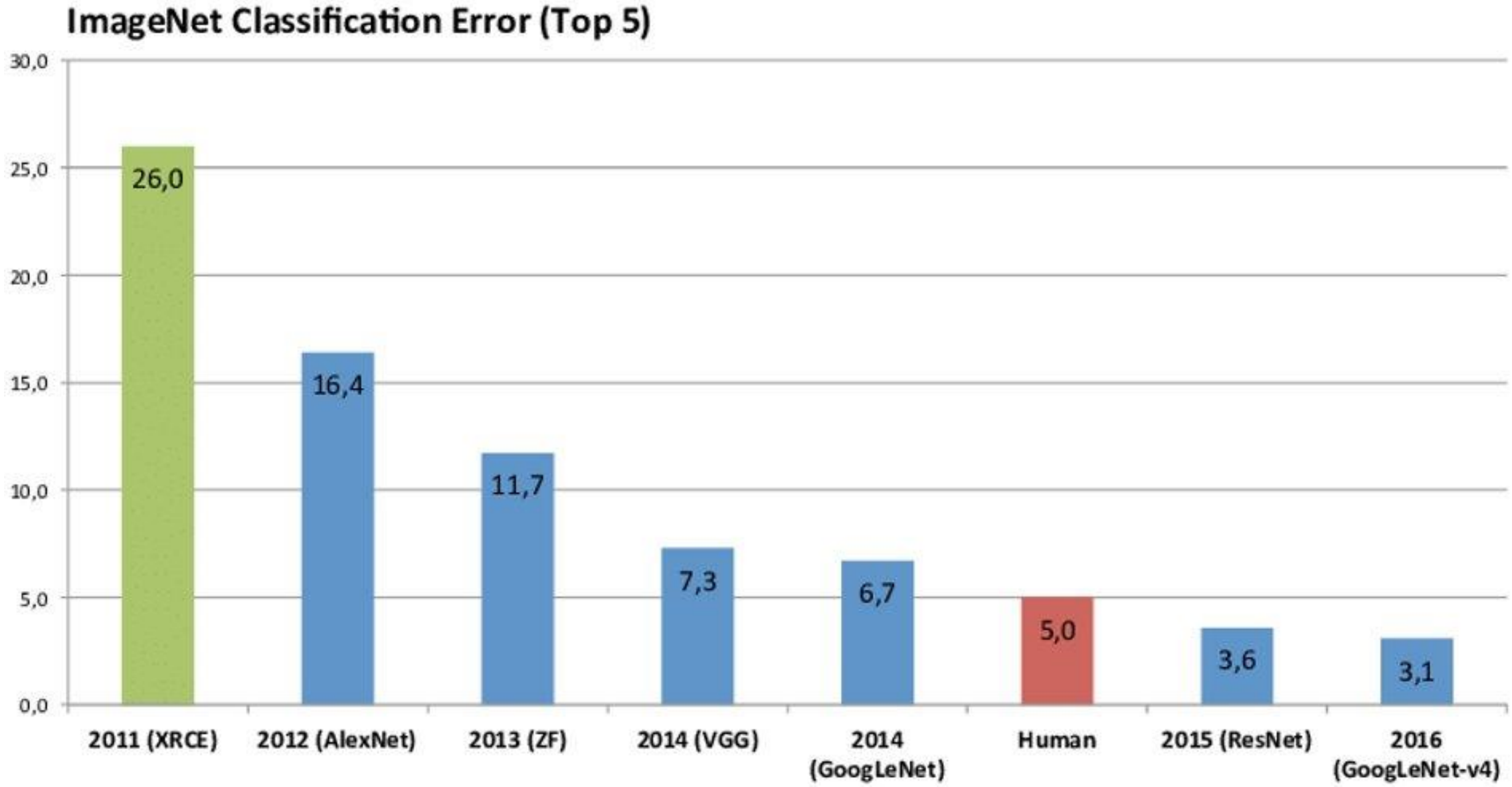


Session 2

- **CNN**
- **Convolution , Conv layers , Padding , MaxPool**
- **FashionMnist CNN Project (keras)**

Convolutional Neural Networks

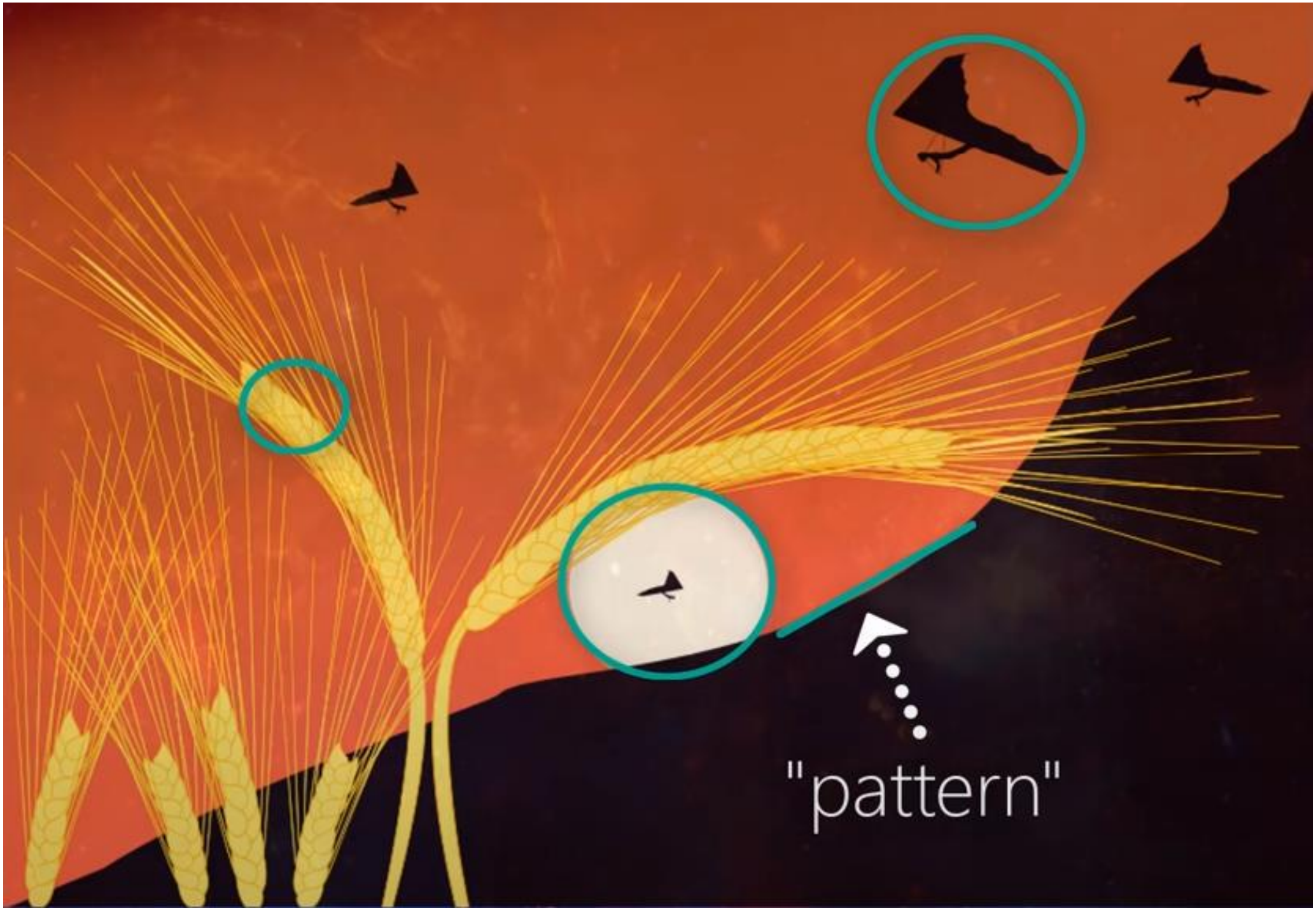

CAR
TRUCK
VAN
BICYCLE
INPUT
CONVOLUTION + RELU
POOLING
CONVOLUTION + RELU
POOLING
FLATTEN
FULLY CONNECTED
SOFTMAX
FEATURE LEARNING
CLASSIFICATION

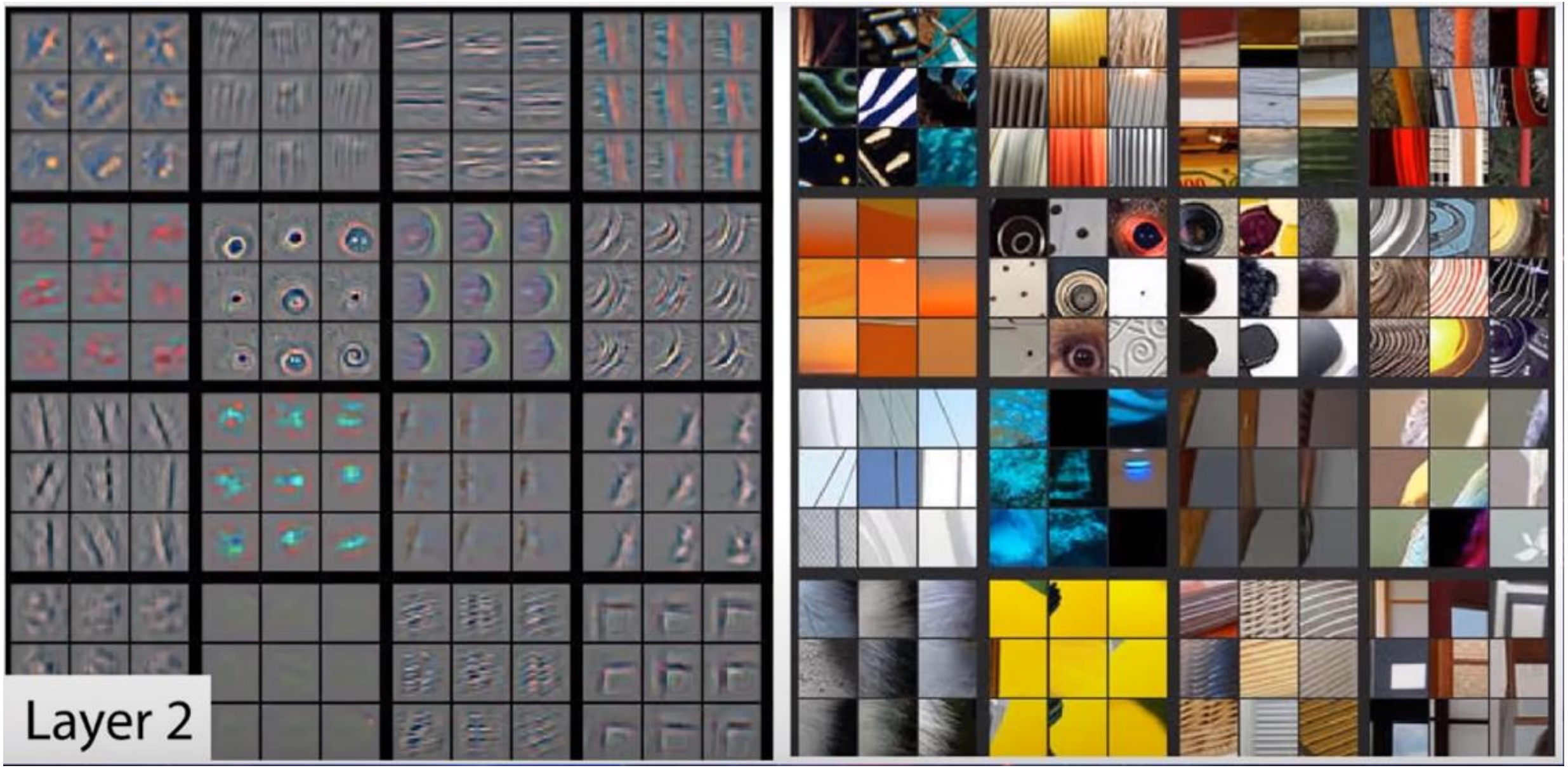

ImageNet Classification Error (Top 5)
30,0
25,0
20,0
15,0
10,0
5,0
0,0
26,0
16,4
11,7
7,3
6,7
5,0
3,6
3,1
2011 (XRCE)
2012 (AlexNet)
2013 (ZF)
2014 (VGG)
2014 (GoogLeNet)
Human
2015 (ResNet)
2016 (GoogLeNet-v4)

Convolutional layers consists of "filters" that can detect a repeating pattern in the images.

In the first layers it only detect basic patterns like lines and circles.

Deeper in the layers it detects the specific image properties.


"pattern"

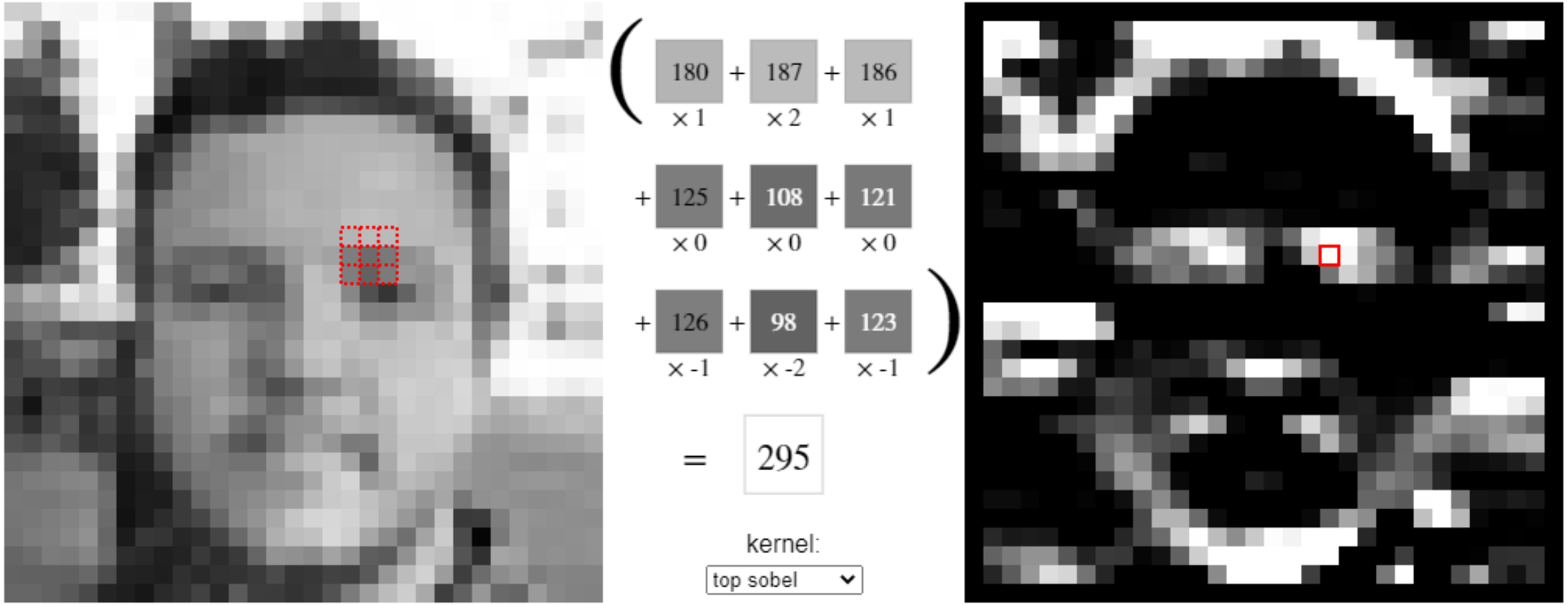

Layer 2

Convolution

- https://setosa.io/ev/image-kernels/

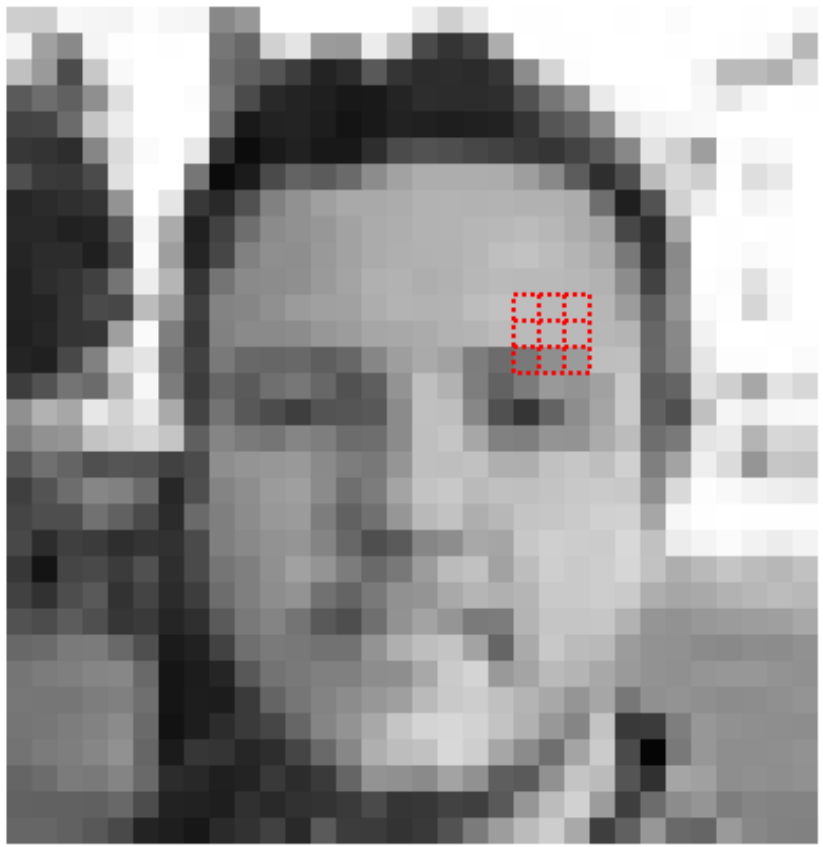
top sobel

$$\begin{pmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{pmatrix}$$



sharpen

$$\begin{pmatrix} 0 & -1 & 0 \\ -1 & 5 & -1 \\ 0 & -1 & 0 \end{pmatrix}$$



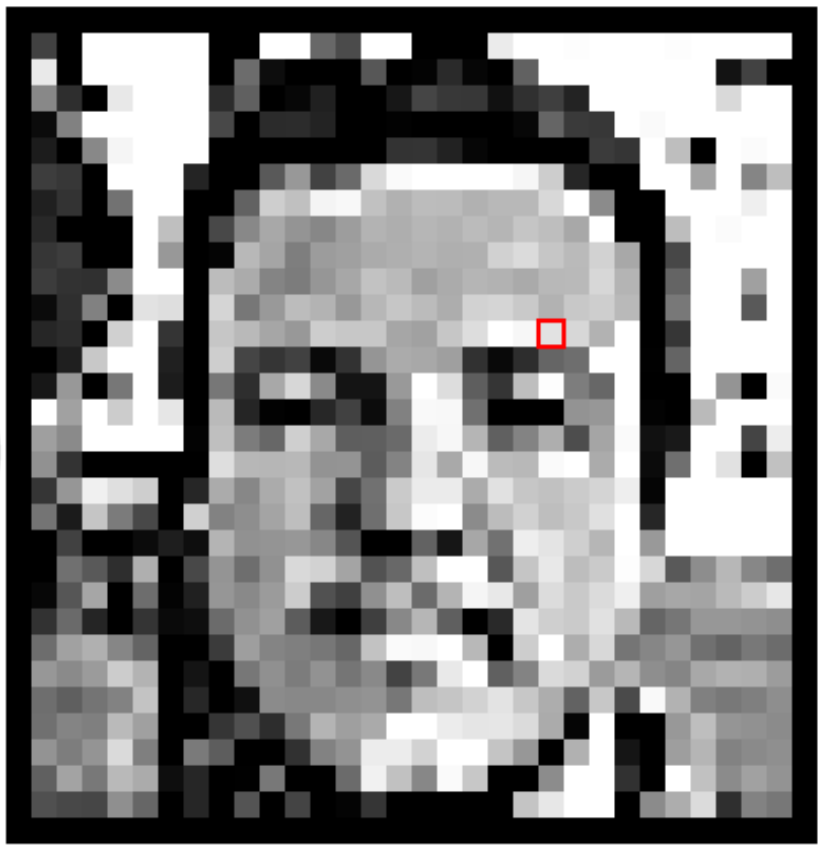
input image

$$\left(\begin{array}{ccc} 190 & + 185 & + 187 \\ \times 0 & \times -1 & \times 0 \\ + 186 & + 185 & + 185 \\ \times -1 & \times 5 & \times -1 \\ + 121 & + 143 & + 155 \\ \times 0 & \times -1 & \times 0 \end{array} \right)$$

$$= 226$$

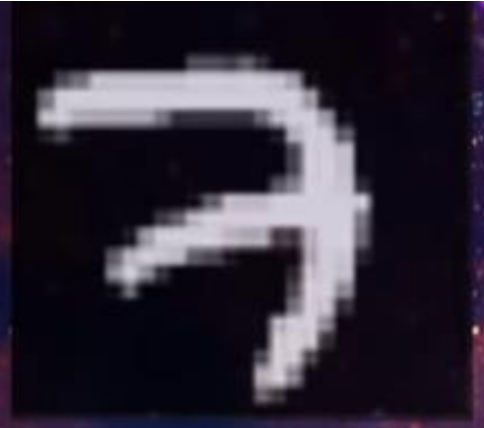
kernel:

sharpen



output image

Line detection filters as example :



filter 1		
-1	-1	-1
1	1	1
0	0	0

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

filter 3		
0	0	0
1	1	1
-1	-1	-1

filter 4		
0	1	-1
0	1	-1
0	1	-1

"bright" corresponds to white in this example

Now you know what the convolution process is .

n x n image

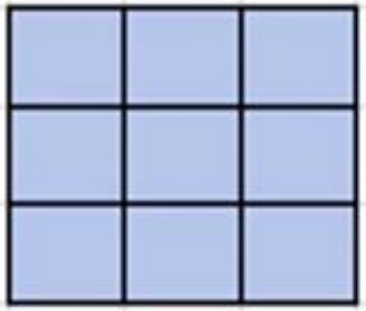
f x f filter

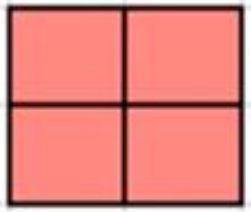
output size = $(n - f + 1) \times (n - f + 1)$

$= (4 - 3 + 1) \times (4 - 3 + 1)$

$= 2 \times 2$

0.3	0.5	0.9	1.0
1.0	1.0	1.0	1.0
0.9	0.9	0.5	0.3
0.2	0.0	0.0	0.0





Input
4 x 4

Filter
3 x 3

Output
2 x 2

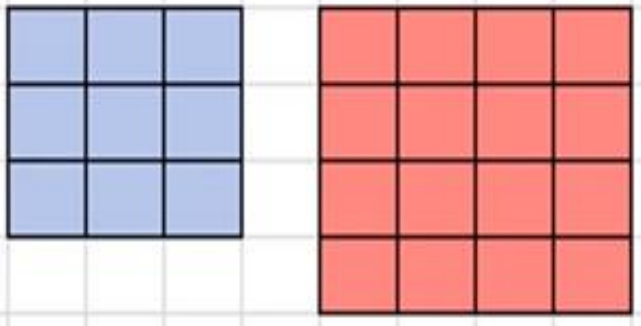
Padding : along the network your network will keep getting lower in size.
Padding adds a frame around the image

Padding:

"valid" - no padding

"same" - padding to make output size same as input size

0	0	0	0	0	0
0	0.3	0.5	0.9	1.0	0
0	1.0	1.0	1.0	1.0	0
0	0.9	0.9	0.5	0.3	0
0	0.2	0.0	0.0	0.0	0
0	0	0	0	0	0



Input
4 x 4

Filter
3 x 3

Output
4 x 4

Strided Convolution :

- The hop length of the convolution process
- Notice how the shape changes

Strided convolution

2	3	7	4	6	2	9
6	6	9	8	7	4	3
3	4	8	3	8	9	7
7	8	3	6	6	3	4
4	2	1	8	3^{3}	4^{4}	6^{4}
3	2	4	1	9^{1}	8^{0}	3^{2}
0	1	3	9	2^{-1}	1^{0}	4^{3}

7×7

$*$

3	4	4
1	0	2
-1	0	3

3×3

$=$

91	100	83
69	91	127
44	72	74

Stride = 2

The formula for the output size : if no padding then p = 0.

$n \times n$ image $\quad f \times f$ filter

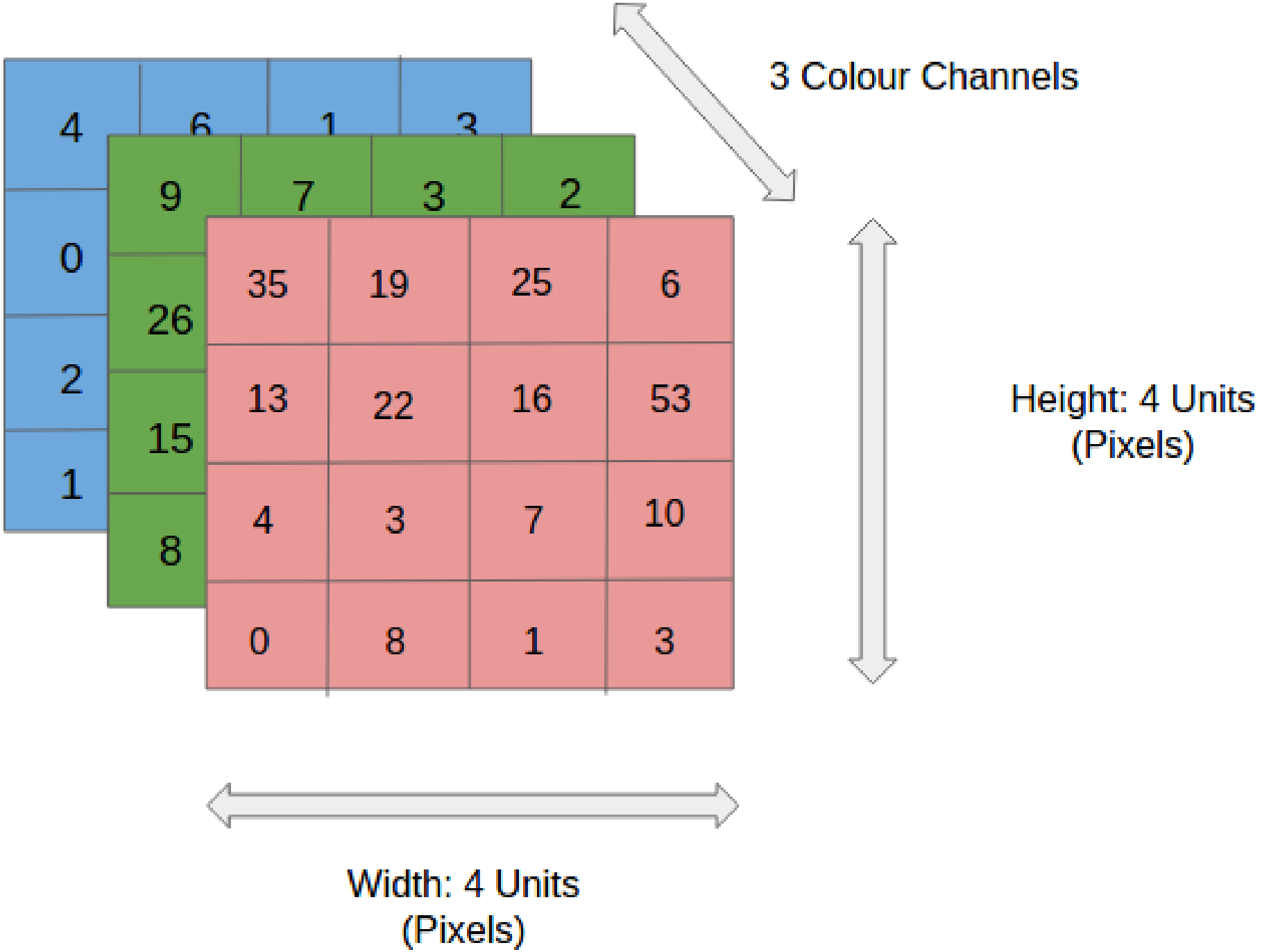
padding p $\quad$ stride s

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

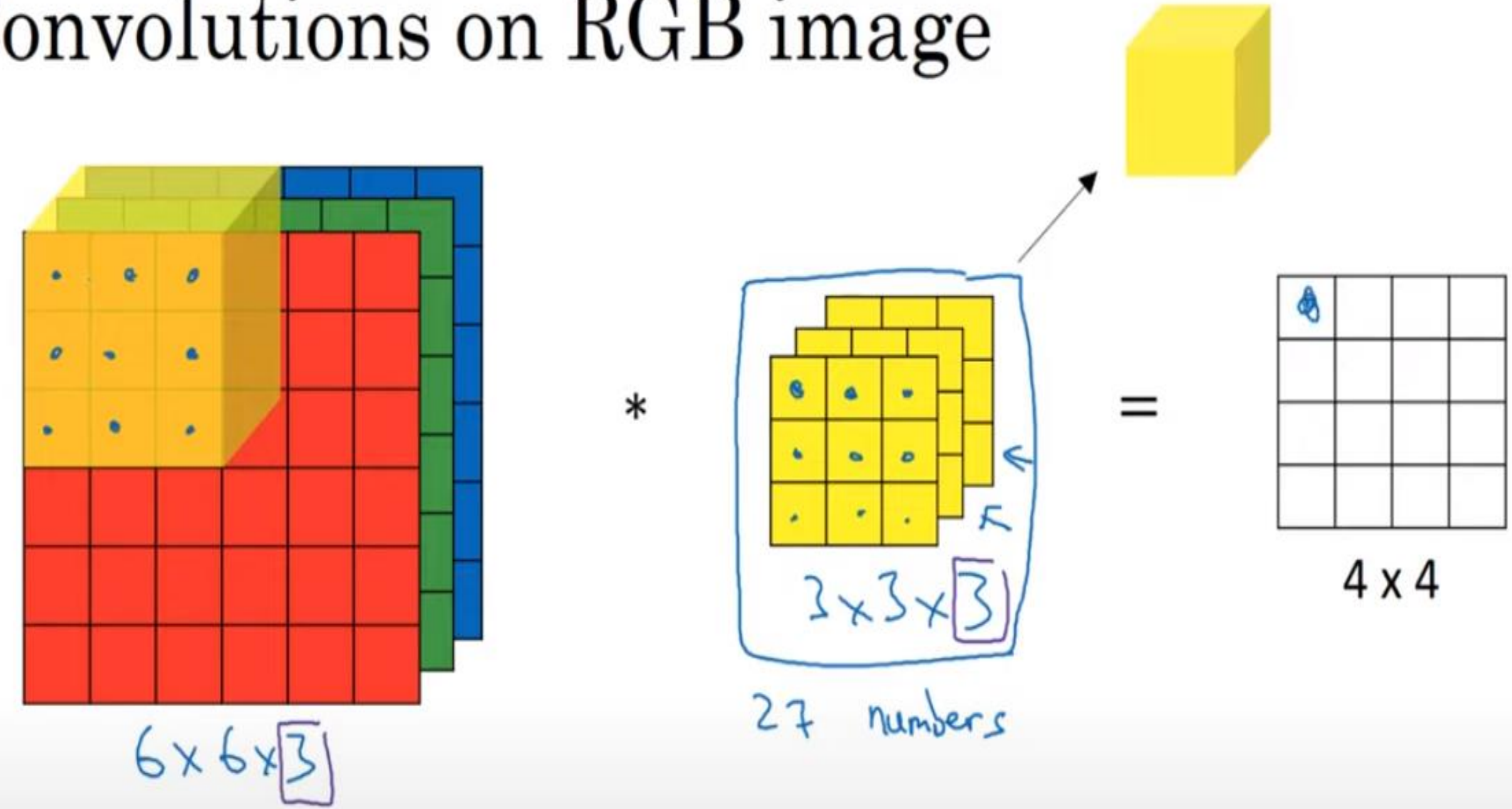
How is convolution applied to RGB colored "3Dimentional" Images?

We consider colored images as 3 dimensional matrices :

- (height , width , depth)
- What is the depth ? Depth is the number of "channels in an image and in RGB its 3 channels (red , green , blue)
- Every image is a matrix of pixel values.
- So for each channel if you want the output you do a convolution on the 3 dimensional image by a 3 Dimensional filter
- In your model you will have more channels in the filter , these channels will contain the features of the input images

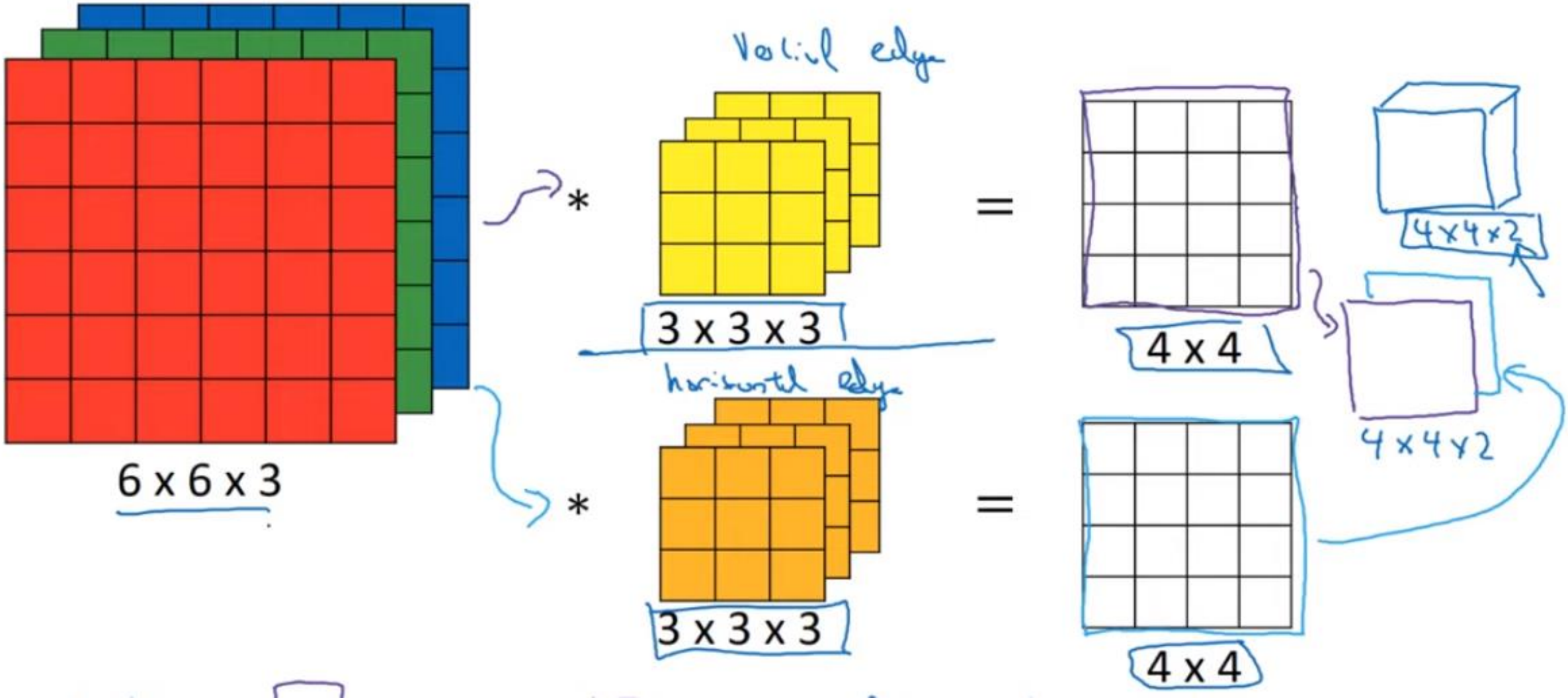

3 Colour Channels
4
6
1
3
0
2
1
9
7
3
2
26
15
8
35
19
25
6
13
22
16
53
4
3
7
10
0
8
1
3
Height: 4 Units
(Pixels)
Width: 4 Units
(Pixels)

Convolutions on RGB image



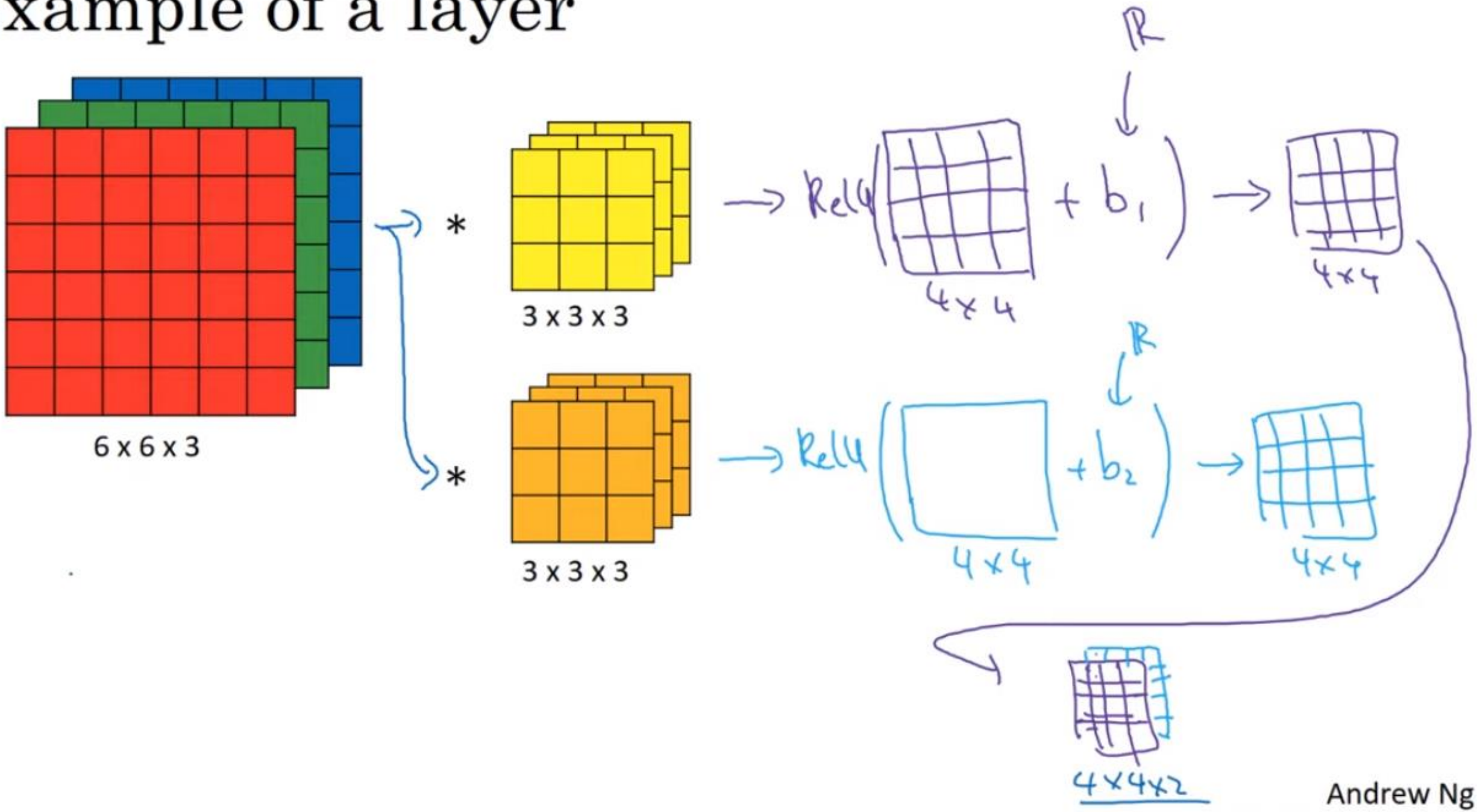
How do you get multiple channels in the output? By doing multiple convolutions on multiple filters

Multiple filters



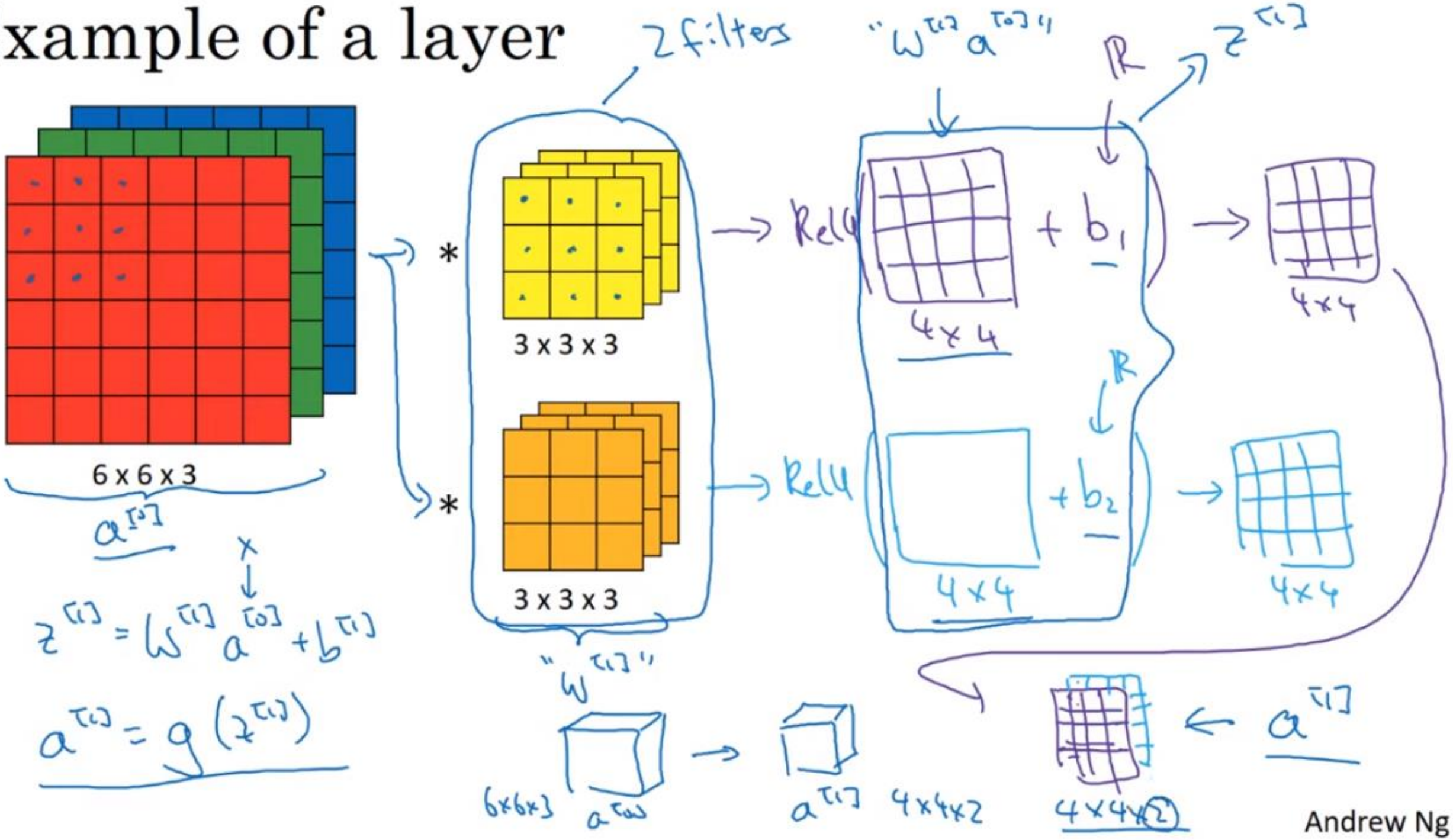
The output channels are stacked and given as input to the next convolution layer

Example of a layer



Andrew Ng

Example of a layer



$$z^{[1]} = W^{[1]} a^{[0]} + b^{[1]}$$

$$a^{[1]} = g(z^{[1]})$$

Types of layers in convolutional neural networks :

Convolution (done)

Fully Connected [linear layer] (done)

Pooling layers

What is max pooling (aslo known as down sampling) and why do we need it?

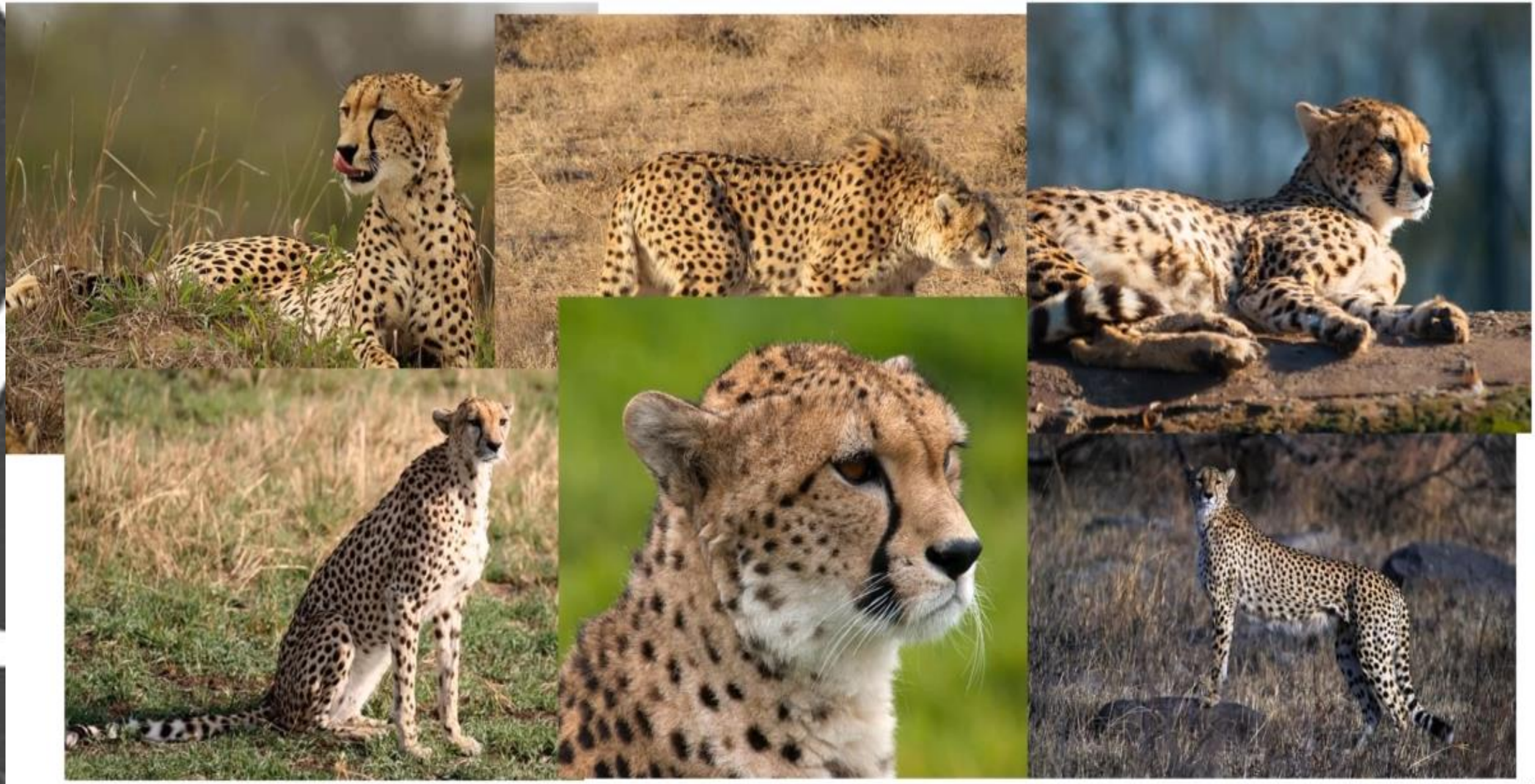


Image Source: Wikipedia

0	1	0	0	0
0	1	1	1	0
1	0	1	2	1
1	4	2	1	0
0	0	1	2	1

Feature Map

Max Pooling

1	1	0
4	2	1
0	2	1

Pooled Feature Map

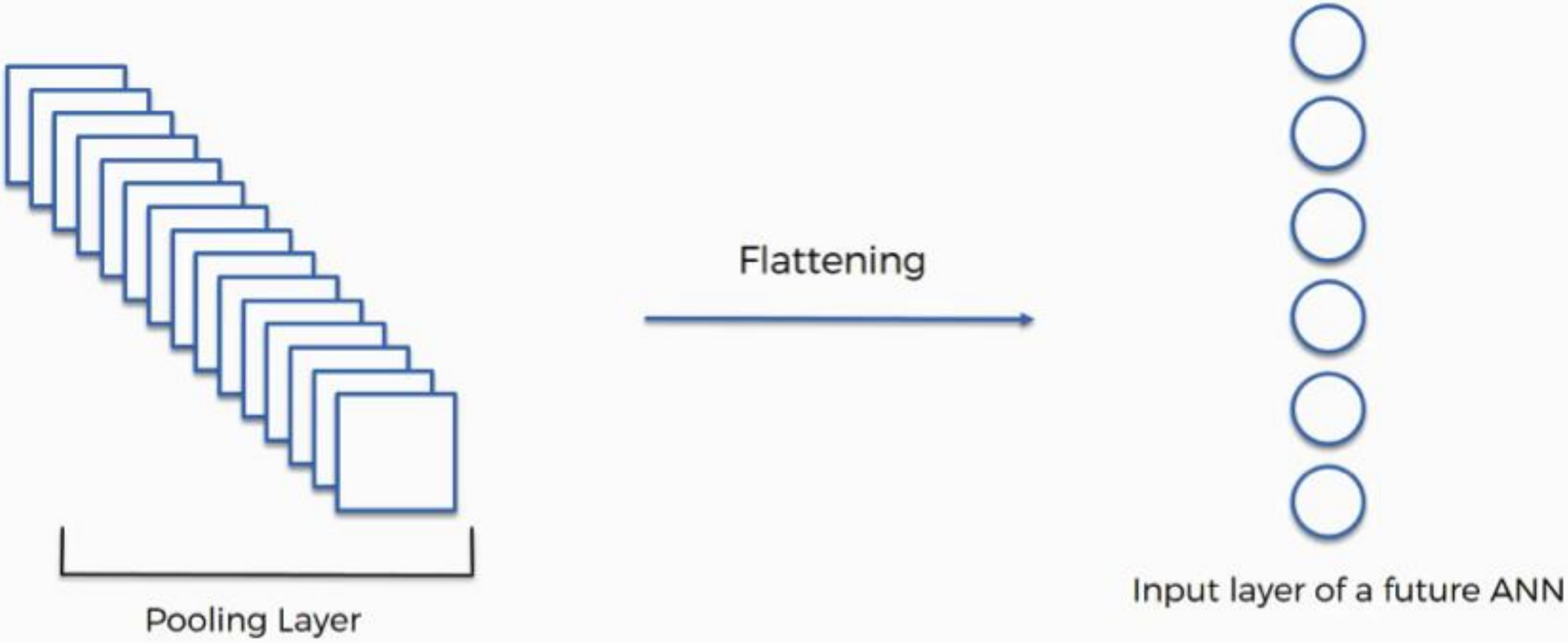
Why Pooling?

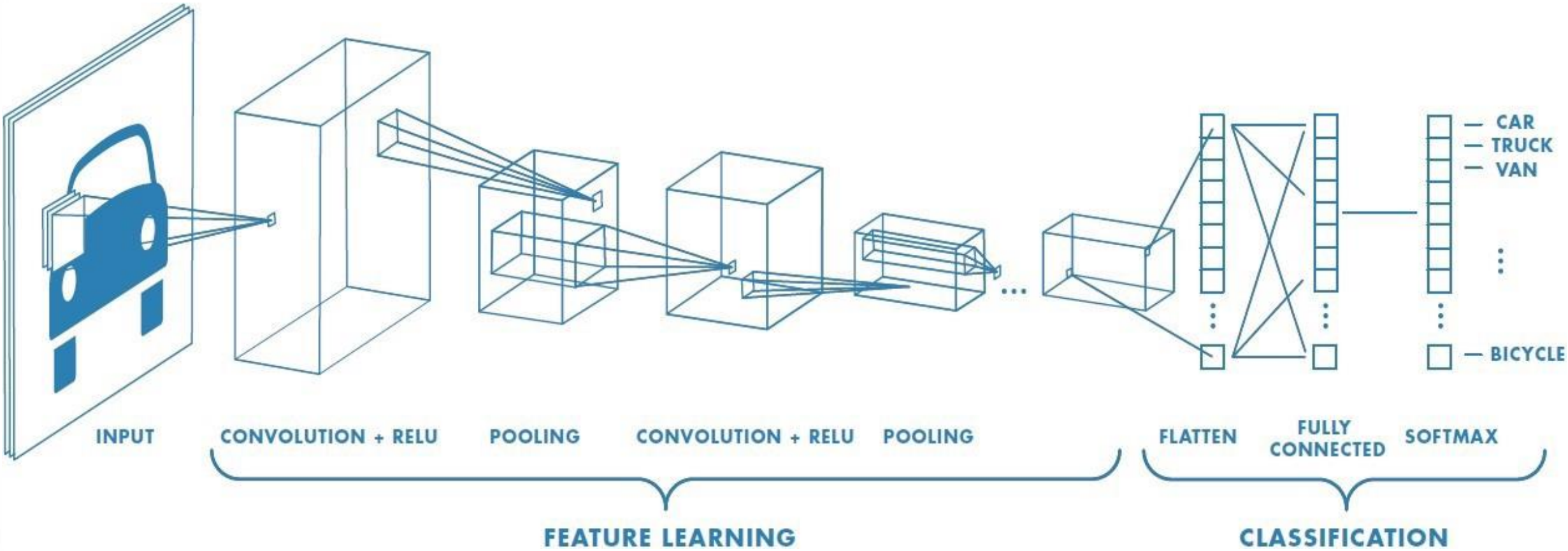
1 . Generalize better (prevents overfitting)

2 .Lowers the Computational Cost

What do we do after the pooling or the convolution layer ?

- **We Flatten the images as seen in the ANN Mnist project and we apply the linear dense layers that we learned about in session 1 .**


Flattening
Pooling Layer
Input layer of a future ANN


CAR
TRUCK
VAN
BICYCLE
INPUT
CONVOLUTION + RELU
POOLING
CONVOLUTION + RELU
POOLING
FLATTEN
FULLY CONNECTED
SOFTMAX
FEATURE LEARNING
CLASSIFICATION

Let's Code Fashion MNIST!


Thank you