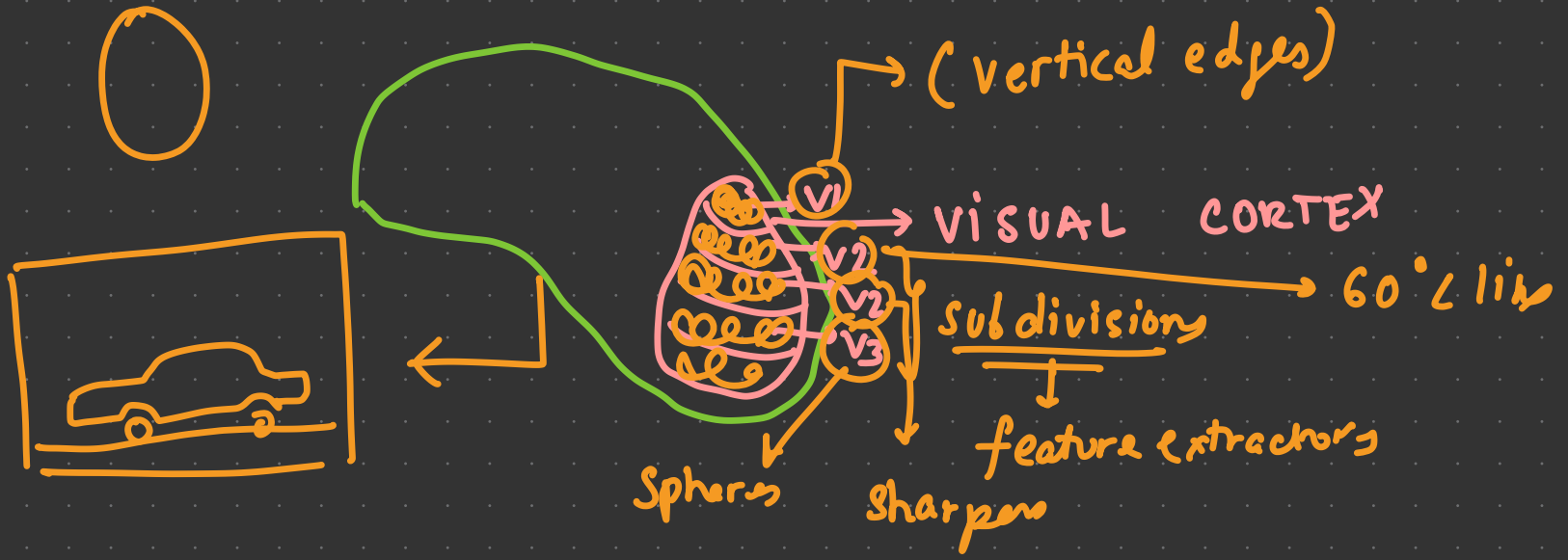
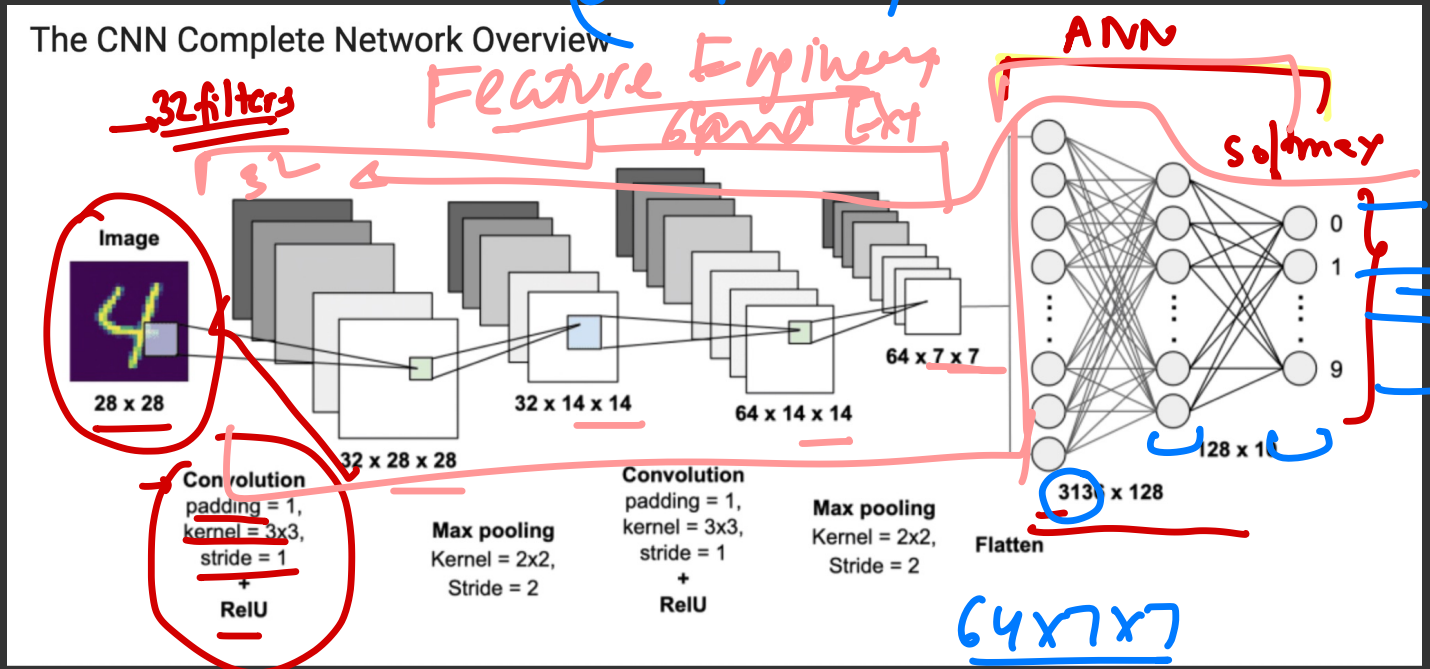
CONVOLUTION NEURAL NETWORKS

1 →

USE → these are used to work on visual data (videos & images)

How does a person recognize different objects




The CNN Complete Network Overview
32 filters
Image
28 x 28
32 x 28 x 28
Convolution
padding = 1,
kernel = 3x3,
stride = 1
+
ReLU
Max pooling
Kernel = 2x2,
Stride = 2
32 x 14 x 14
Convolution
padding = 1,
kernel = 3x3,
stride = 1
+
ReLU
64 x 14 x 14
Max pooling
Kernel = 2x2,
Stride = 2
64 x 7 x 7
Flatten
0
1
9

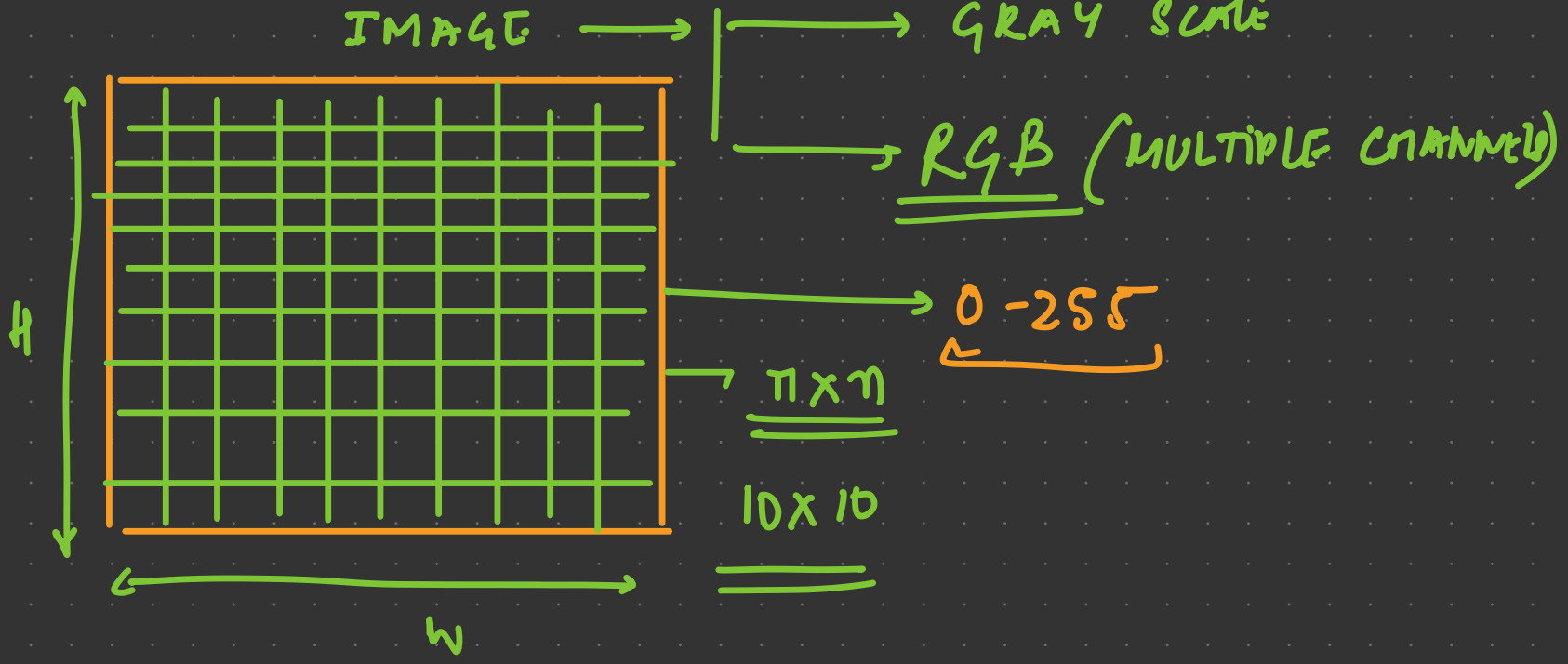
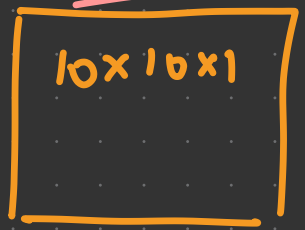


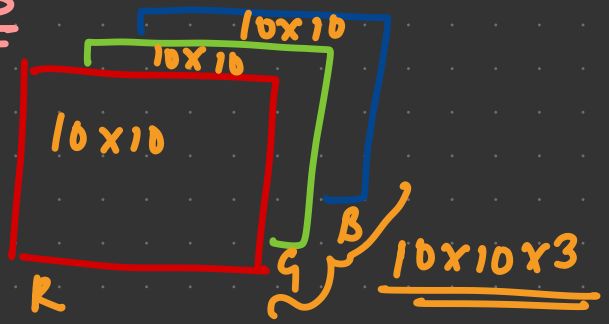
IMAGE
GRAY SCALE
RGB (MULTIPLE CHANNELS)
0 - 255
n x n
10 x 10
H
W

GRAY

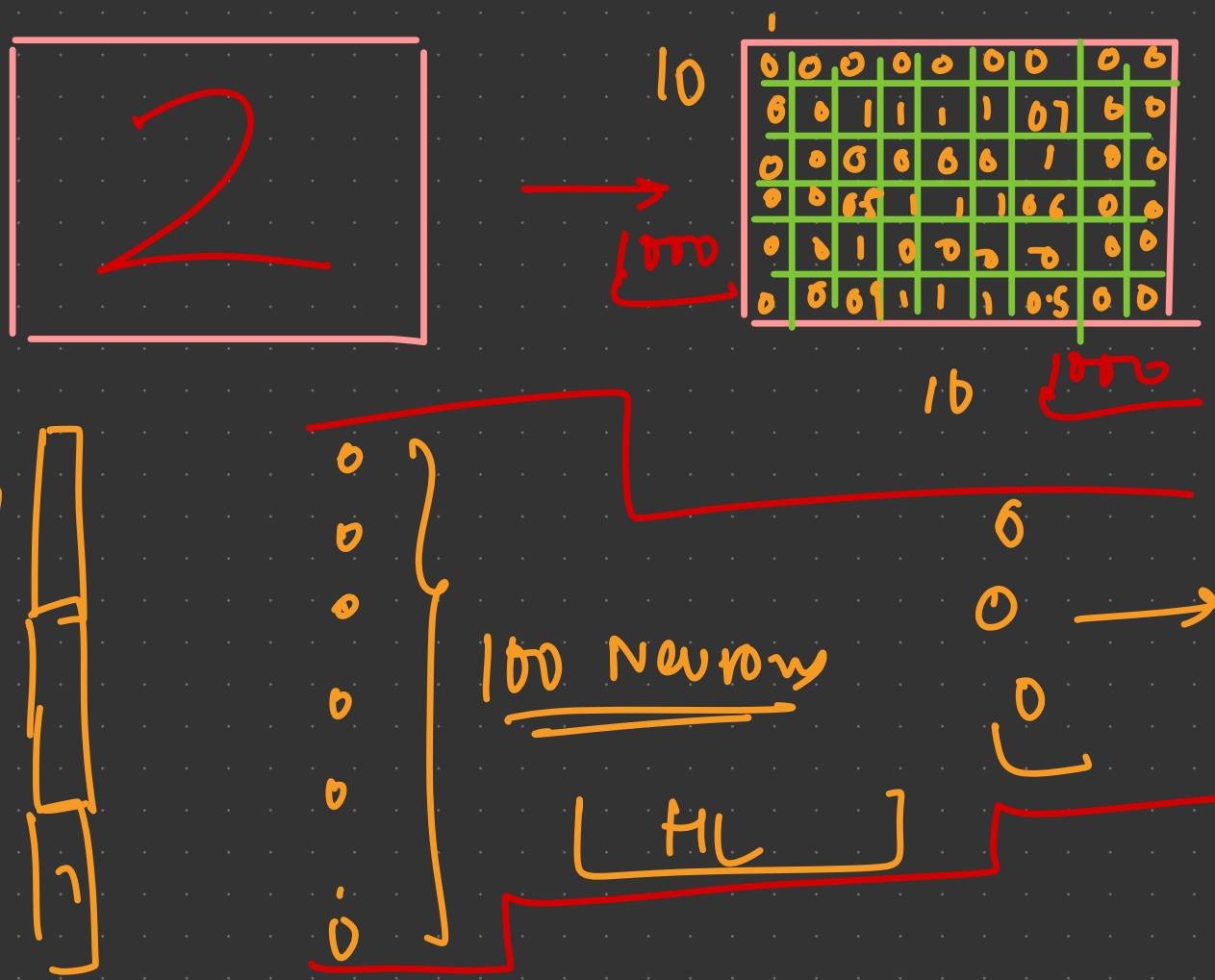


GRAY CHANNEL

RGB

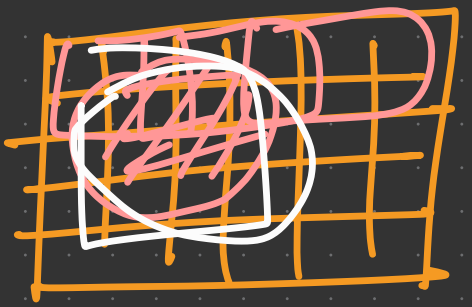


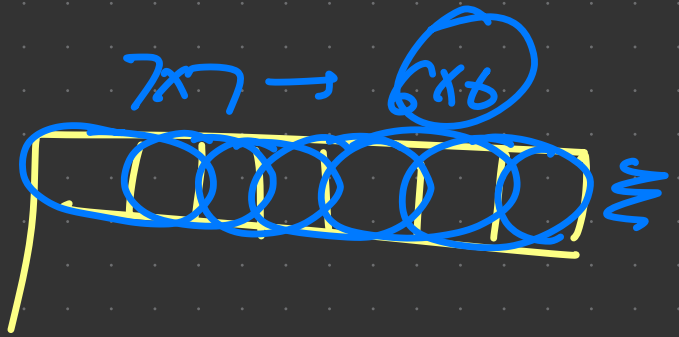
☆ CONVOLUTION


2
10
100
16
7
7
100 Neurons
HL

1000 * 1000]

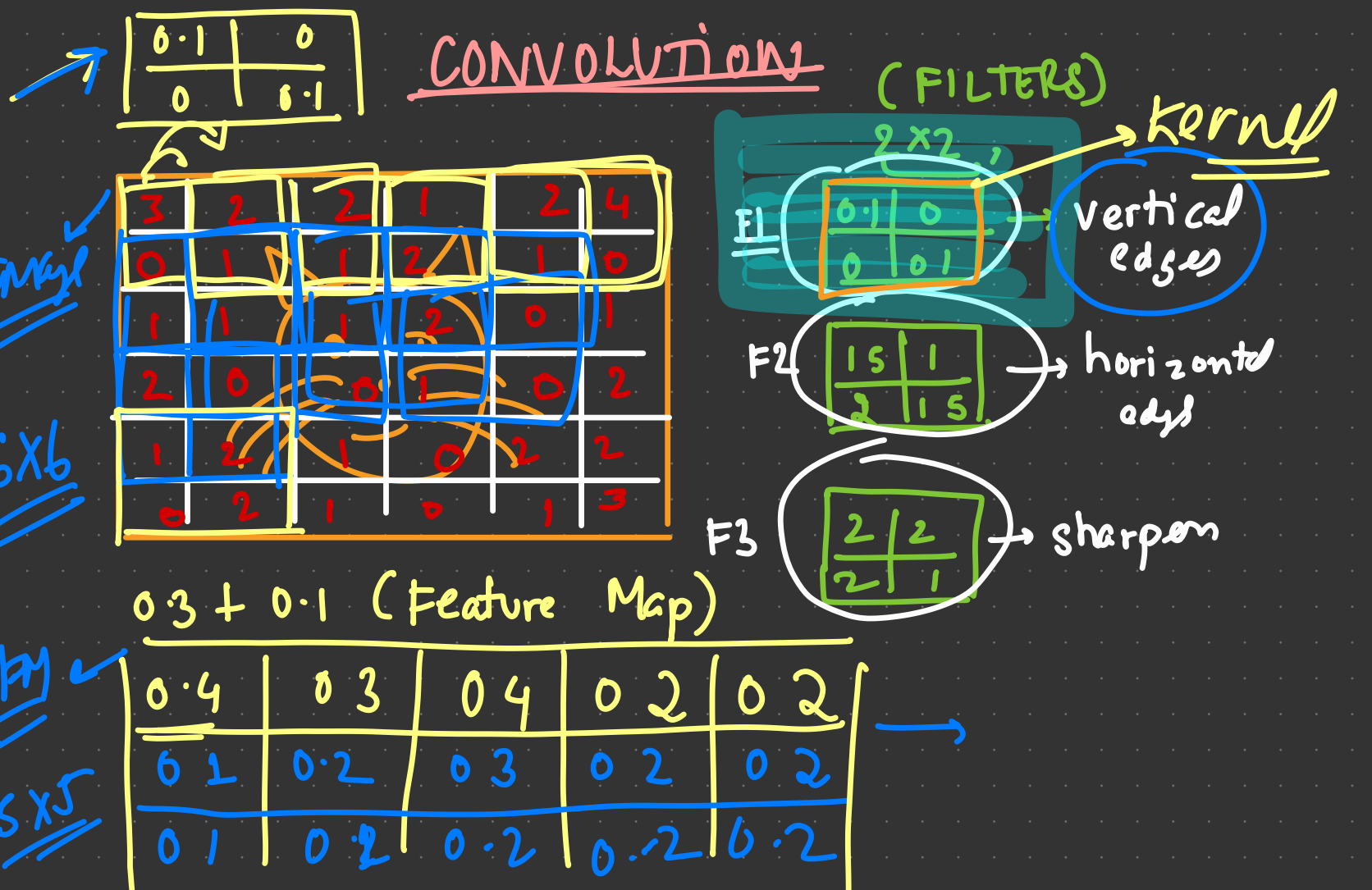
(weights & biases)





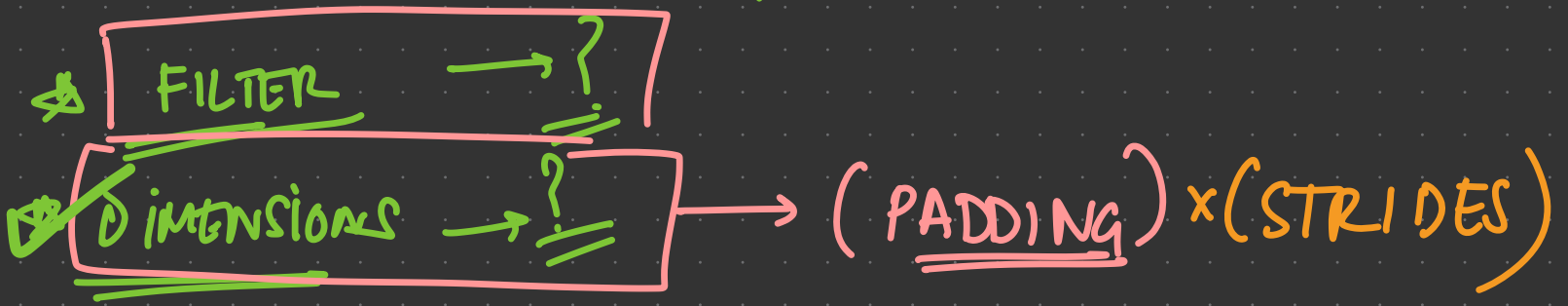
10 Lakh → 1 Billion

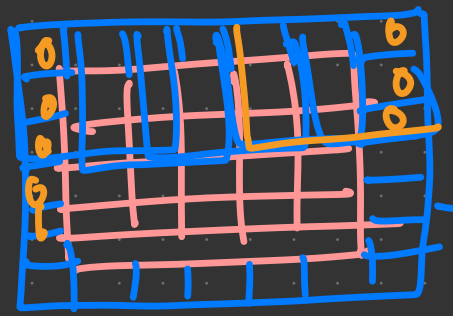
CONVOLUTION



→ We were able to extract information from your Image using a filter

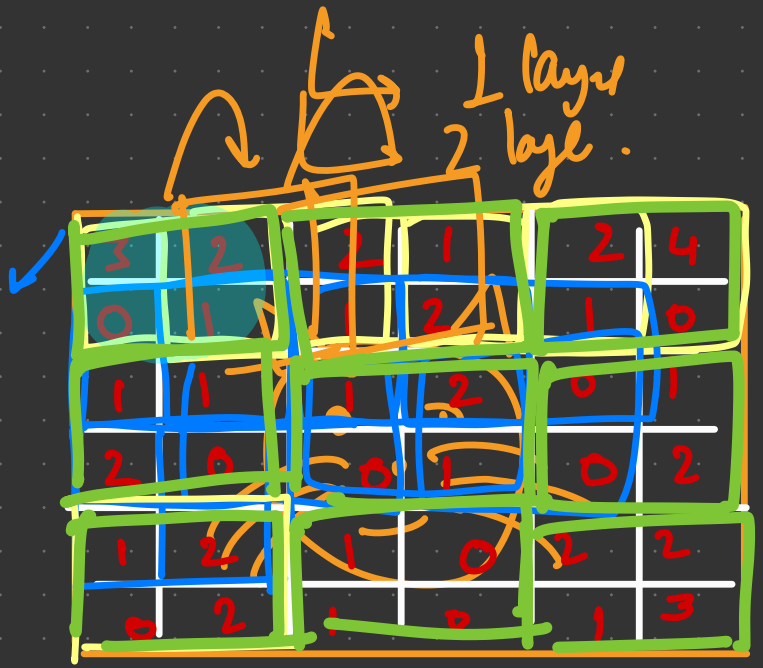
100 Feature Maps

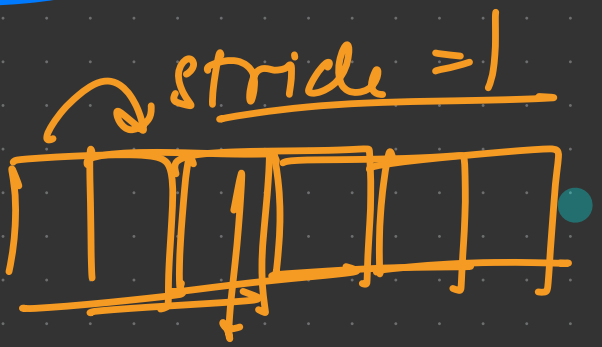




→ 7x7 (3x3)

5x5

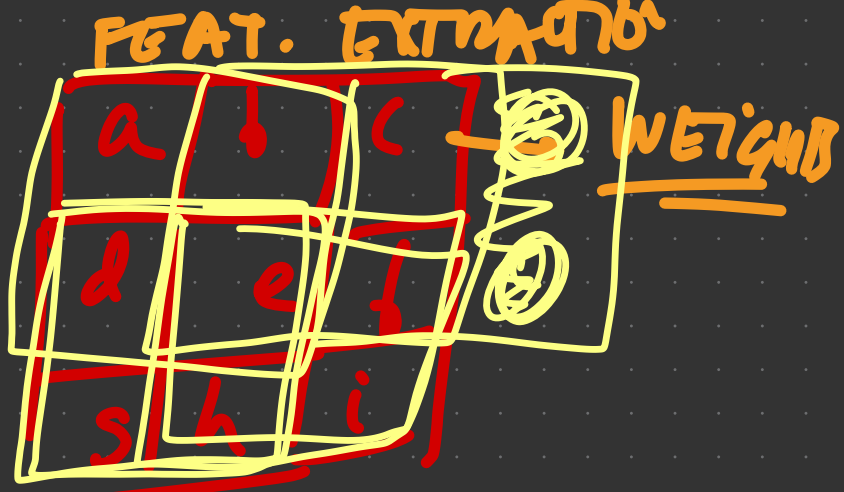


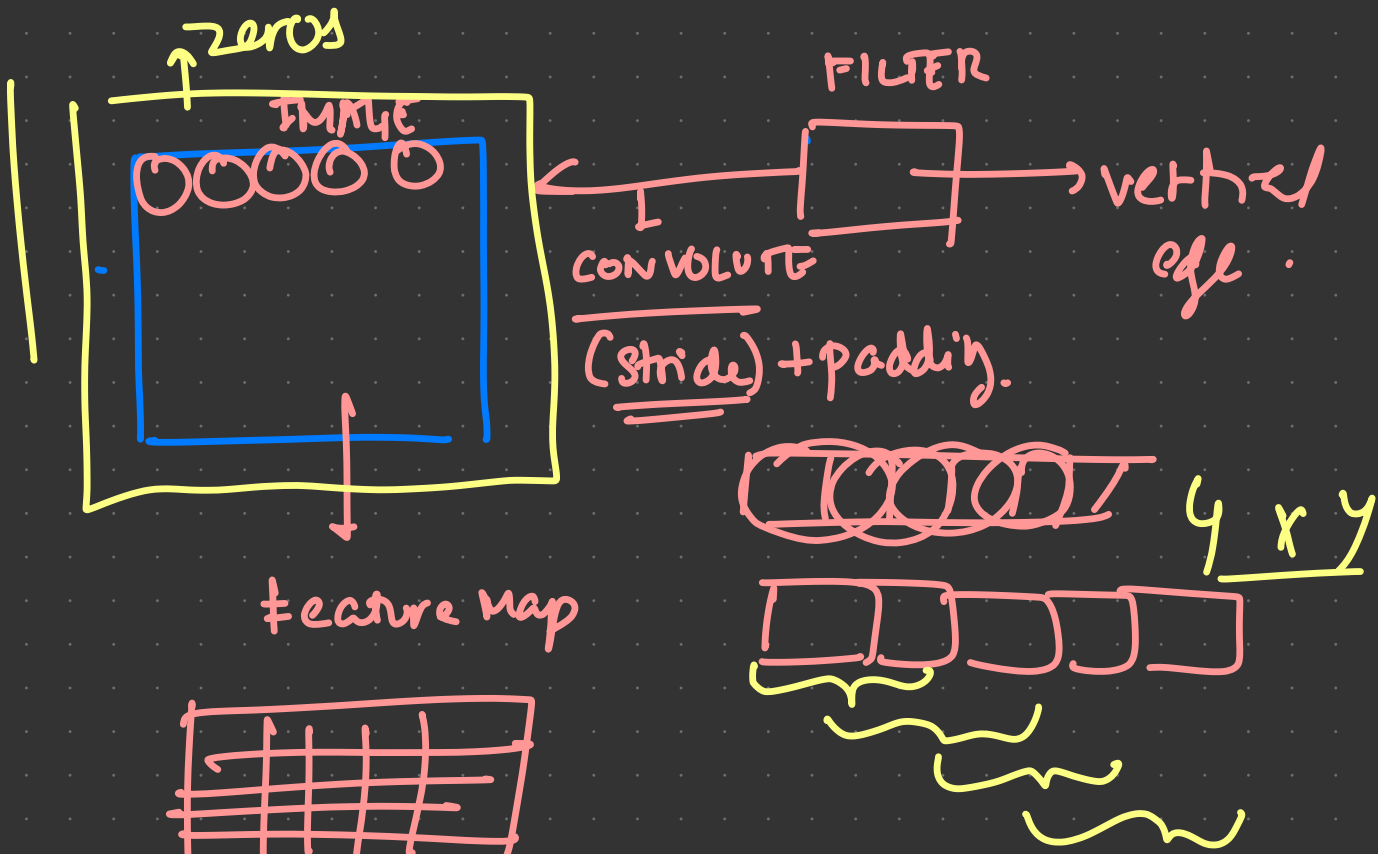


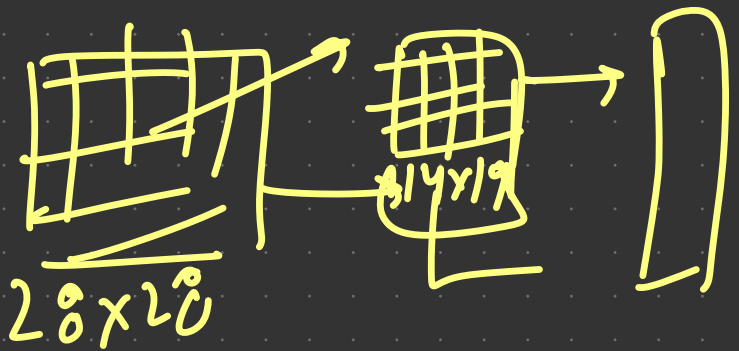
→ STRIDE

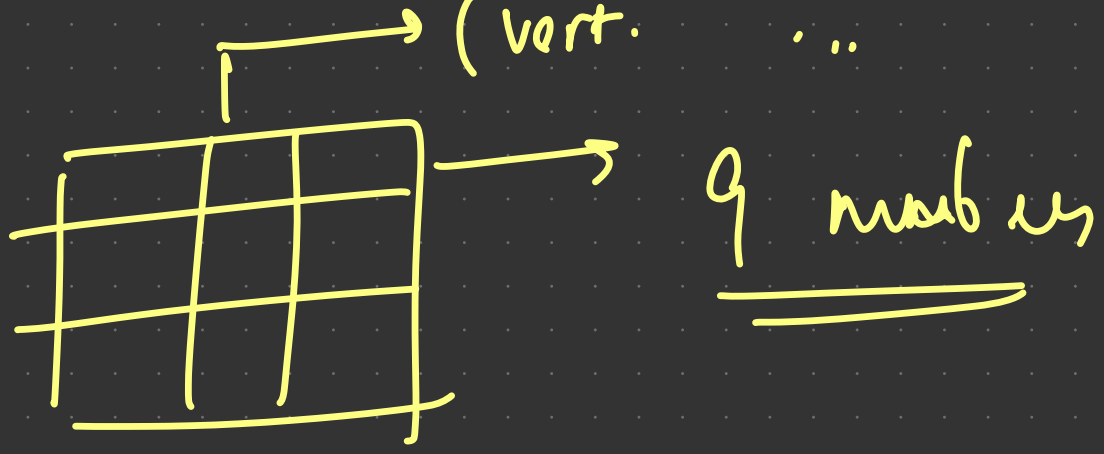
0.5	0.7
0.3	0.4

- USE CASES CNN
- HUMAN BRAIN → DETECT IMAGES and EXTRACT FEAT
- CONVOLUTION
- PADDING
- STRIDES
- FILTERS
- POOLING } DIMENSIONS REDUCED




zeros
IMAGE
FILTER
CONVOLUTE
vertical edge.
(stride) + padding.
feature map
4 x 4
28x28
NxN

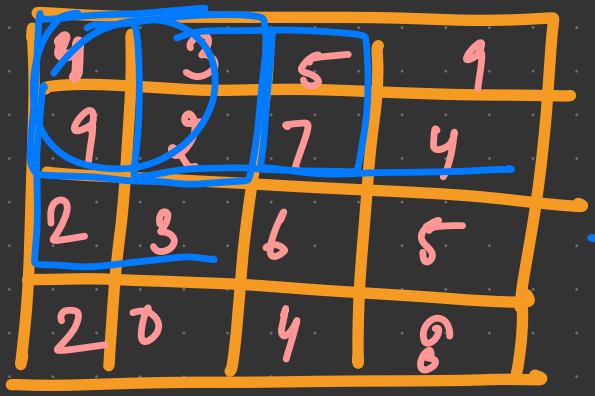

20x20
3x3


(vert.
...
9 numbers

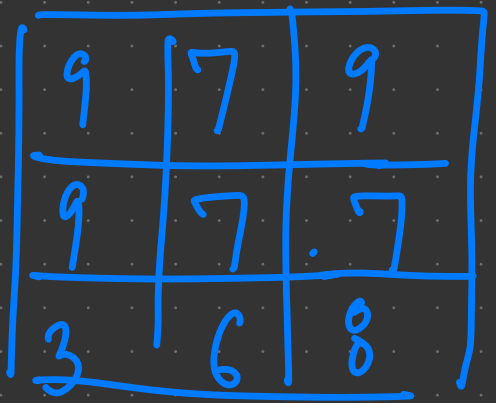
POOLING →

max pooling
min pooling
stride avg pool.

→ (2×2)



→ Minimum value in Min pool.



→ Avg pooling → avg.

IMPORTANT FORMULAS

Input Dimensions:

Width (W): The width of the input image.
Height (H): The height of the input image.
Channels (C): The number of channels or color planes in the input image (e.g., 1 for grayscale, 3 for RGB).

Convolutional Layer:

Kernel Size (K): The width and height of the convolutional kernel or filter.
Padding (P): The number of zeros added to the border of the input image to preserve spatial dimensions.
Stride (S): The step size at which the kernel moves across the input image.
Number of Filters (N): The number of convolutional filters applied in the layer.

Output Dimensions:

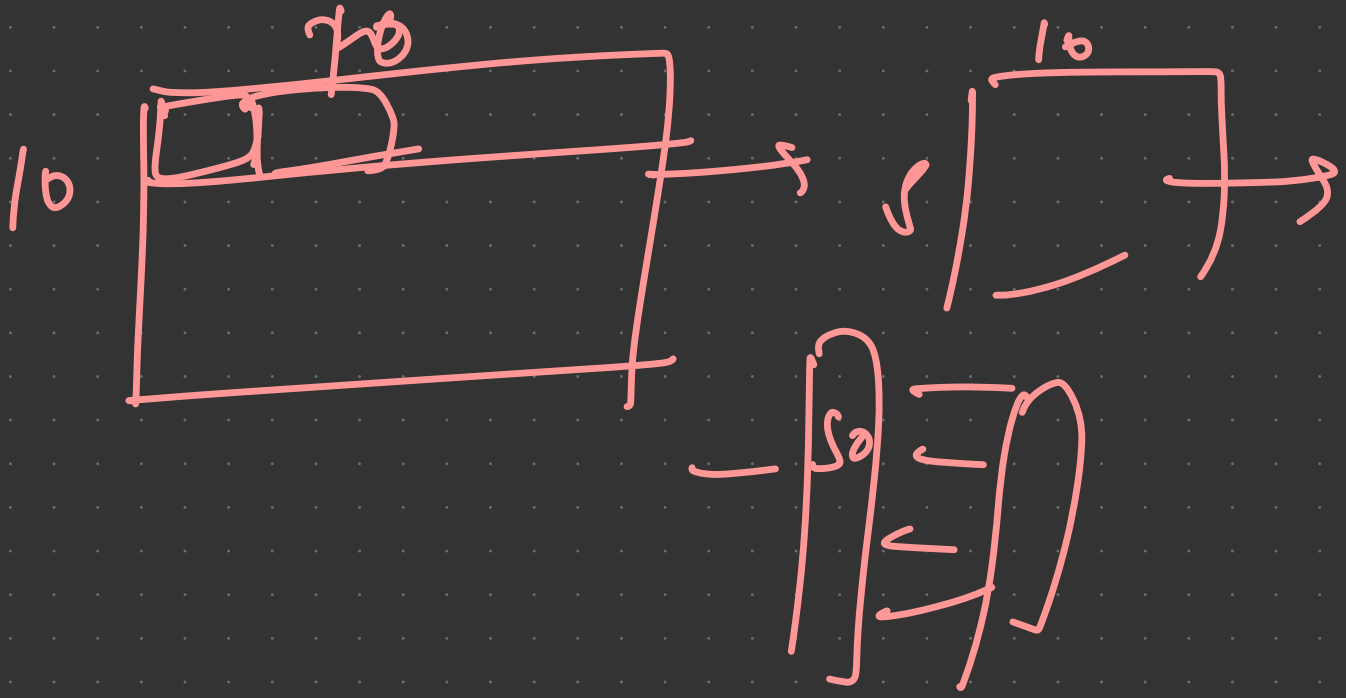
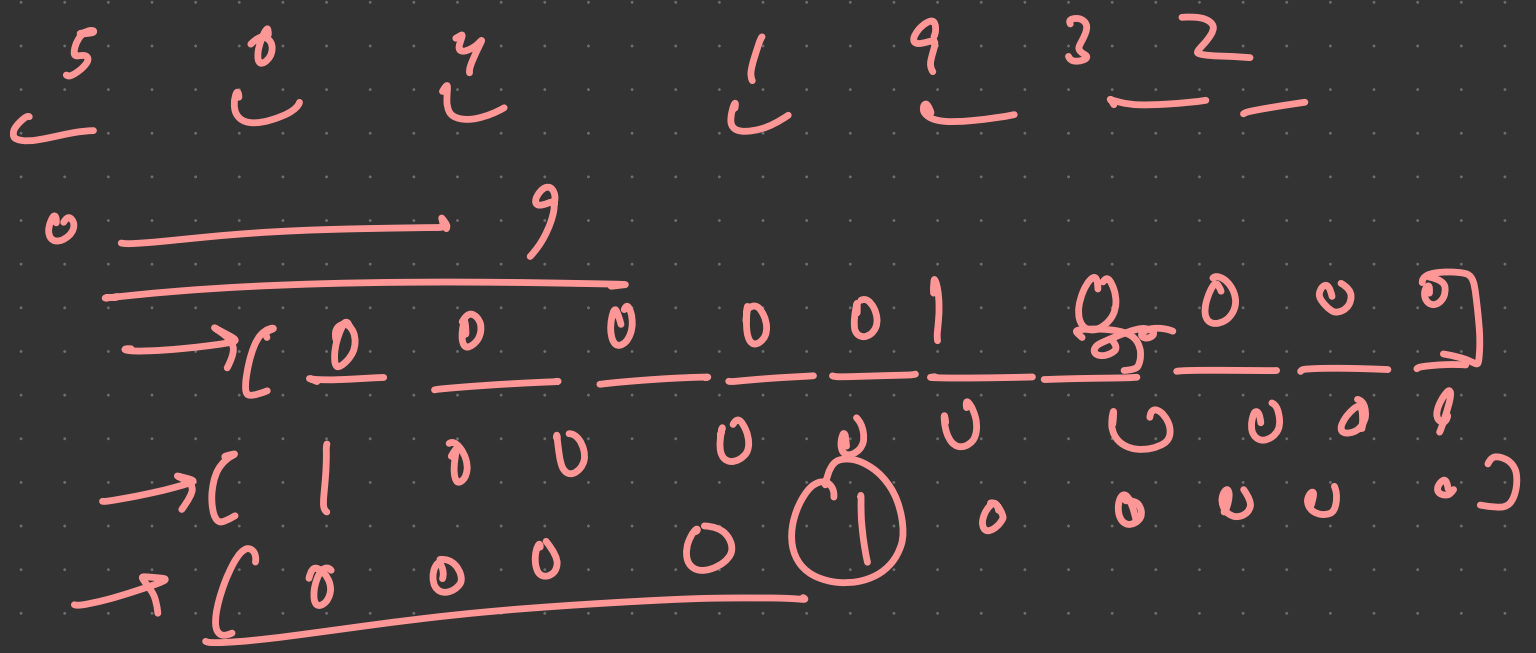
Width (W_out): ((W - K + 2P) / S) + 1
Height (H_out): ((H - K + 2P) / S) + 1
Channels (C_out): N

Pooling Layer:

Pool Size (P): The width and height of the pooling window.
Stride (S): The step size at which the pooling window moves across the input.

Output Dimensions:

Width (W_out): ((W - P) / S) + 1
Height (H_out): ((H - P) / S) + 1
Channels (C_out): C (remains the same)


20
10
10
50