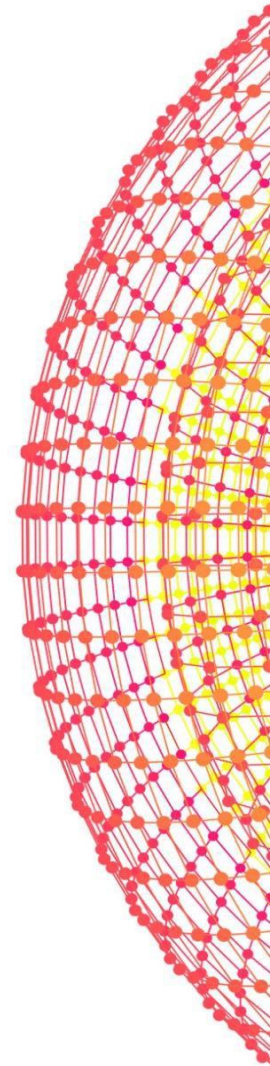


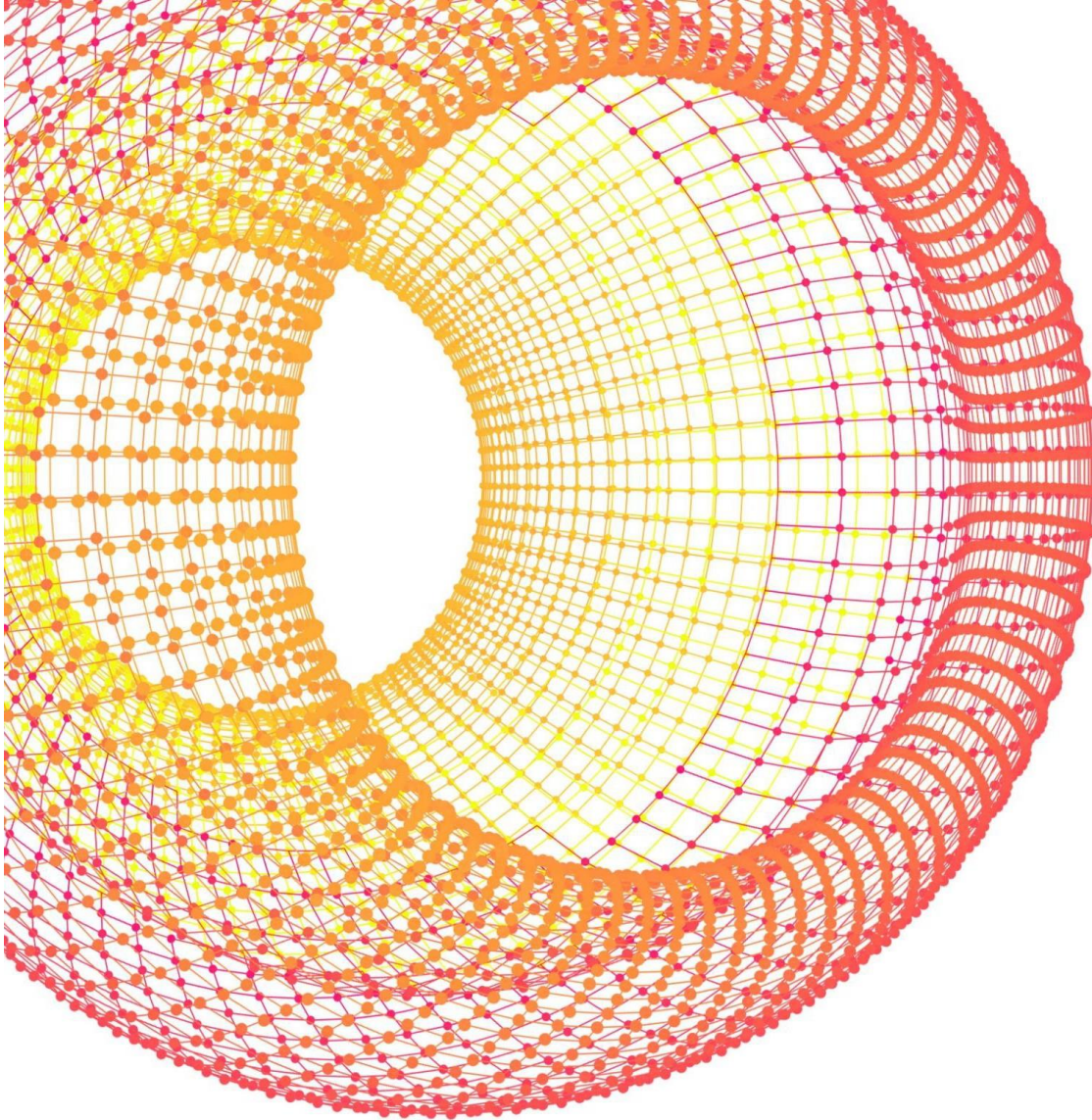


CONVOLUTIONAL NEURAL NETWORK

AGENDA

- Introduction to Computer Vision
- What is CNN
- Understanding Convolutional Layer
- Understanding Pooling Layer
- Applications of CNN





INTRODUCTION TO COMPUTER VISION

Introduction to Computer Vision

Computer vision is a branch of AI that focuses on extracting useful information from pictures and videos



Computer Vision system is a neural network designed to solve a specific visual problem. It looks at a lot of examples of a problem and learns from those examples (training)..

Introduction to Computer Vision

Represent colors by numbers: In computer science, each color is represented by a specified HEX value. That is how machines are programmed to understand what colors the image pixels are made up. Whereas as humans we have an inherited knowledge to differ between the shades

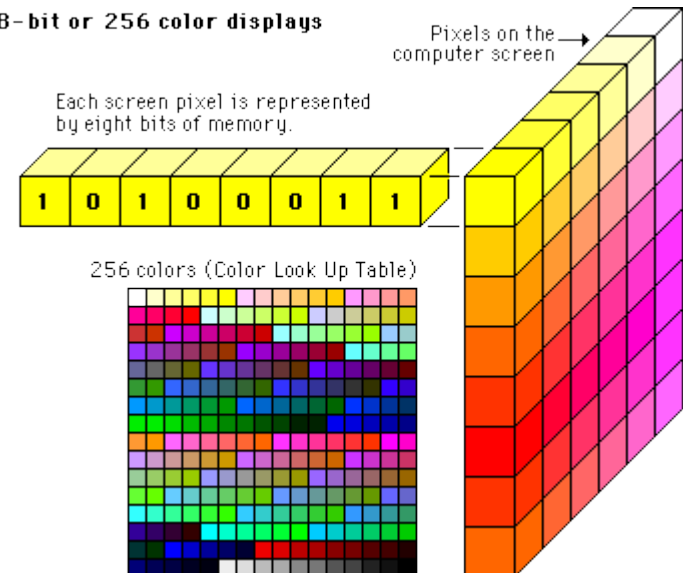
What I see

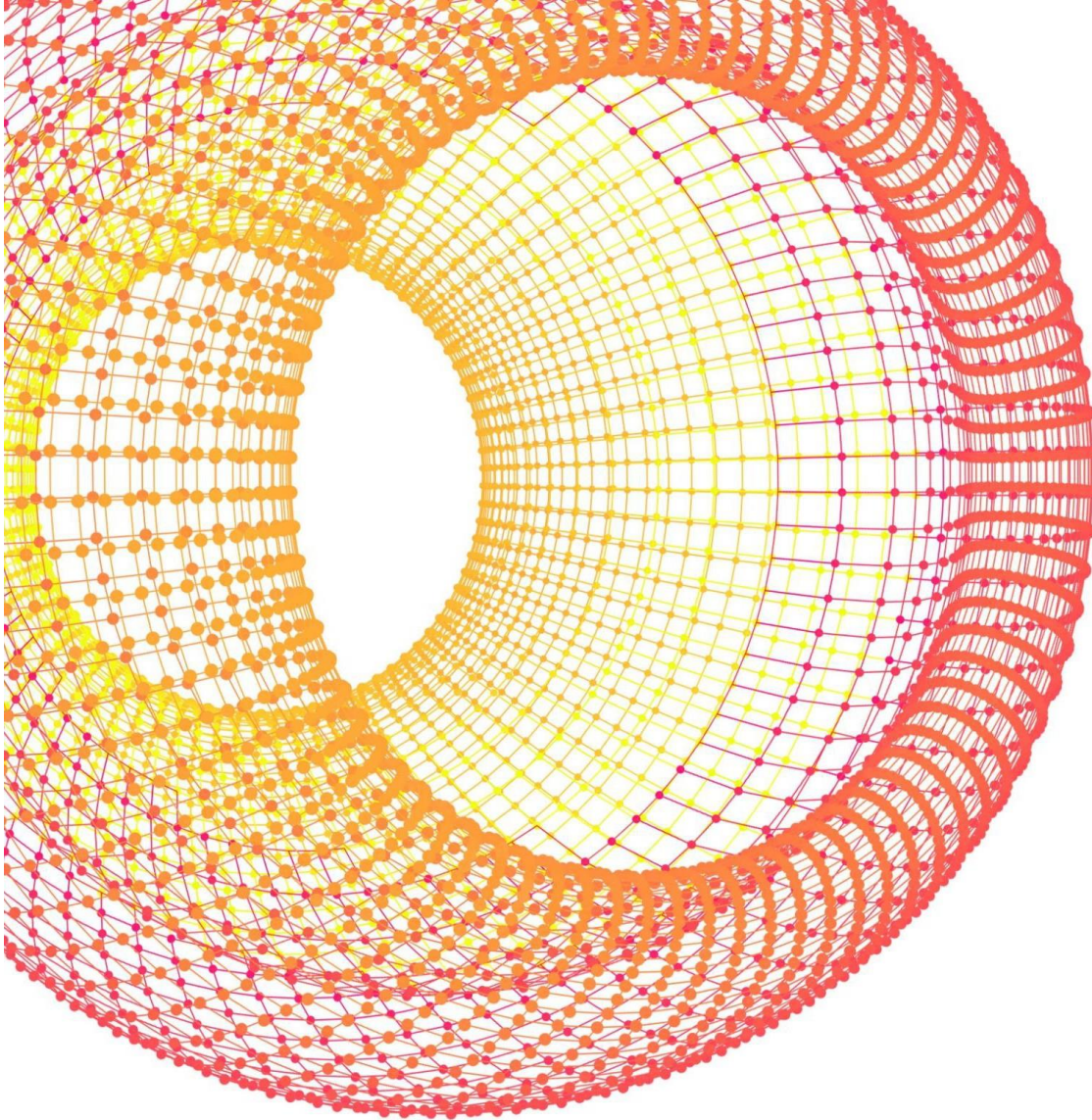


What a computer sees

08	02	22	97	38	15	00	40	00	75	04	05	07	78	52	12	50	77	91	08
49	49	99	40	17	81	18	57	60	87	17	40	98	43	69	48	04	56	42	00
81	49	31	73	55	79	14	29	93	71	40	67	53	88	30	03	49	13	36	45
52	70	95	23	04	40	11	42	49	24	48	54	01	32	56	71	37	02	36	91
22	31	16	71	51	67	63	89	41	92	36	54	22	40	40	28	66	33	13	80
24	47	32	40	99	03	45	02	44	75	33	53	78	36	84	20	35	17	12	50
32	98	81	28	64	23	67	10	24	38	40	67	59	54	70	66	18	38	64	70
47	24	20	48	02	62	12	20	95	63	94	39	43	08	40	91	66	49	94	21
24	55	58	05	64	73	99	24	97	17	78	78	94	63	14	88	34	59	43	72
21	36	23	09	75	00	74	44	20	45	35	14	80	61	33	97	34	31	33	95
78	17	53	28	22	75	31	67	15	94	03	80	04	62	16	14	09	53	56	92
16	39	05	42	96	35	31	67	55	58	88	24	00	17	54	24	36	29	85	57
86	54	00	48	35	71	89	07	05	44	44	37	44	40	21	58	51	54	17	58
19	80	81	48	05	94	47	49	28	73	92	13	94	52	17	77	04	89	55	40
04	52	08	83	97	35	98	14	07	97	57	32	14	24	79	33	27	99	46	
88	36	48	87	57	62	20	72	03	46	33	47	46	55	12	32	63	93	53	69
04	42	16	73	38	25	39	11	24	94	72	18	08	46	29	32	40	62	76	36
20	49	36	41	72	30	23	88	34	62	99	69	82	47	59	85	74	04	36	16
20	73	35	29	78	31	90	01	74	31	49	71	48	84	81	14	23	57	05	54
01	70	54	71	83	51	54	49	16	92	33	48	41	43	52	01	89	19	47	48

8-bit or 256 color displays

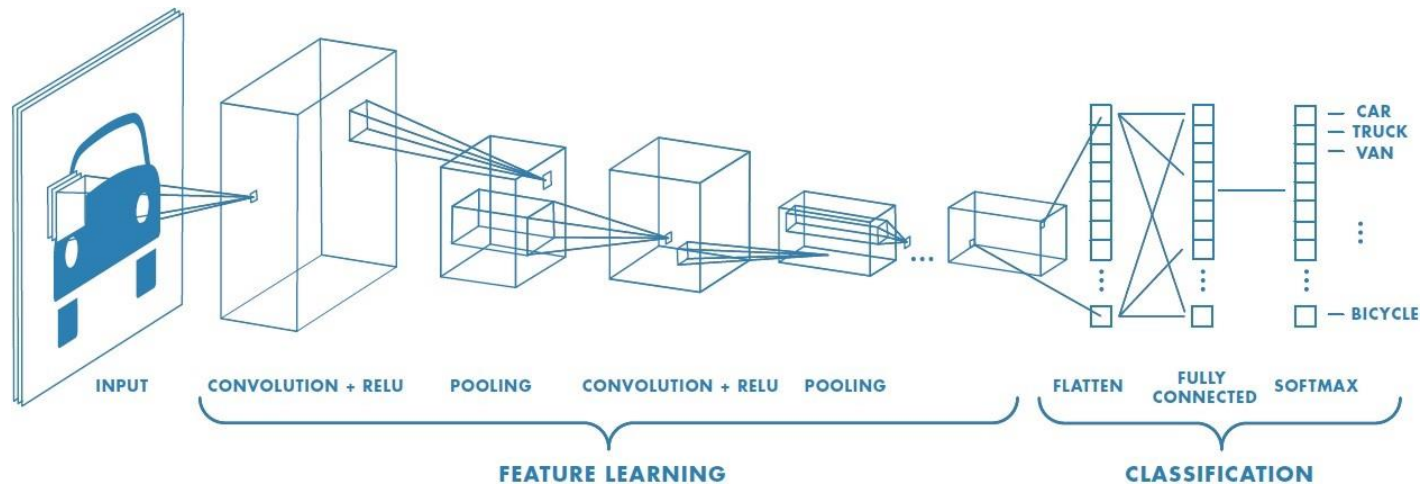




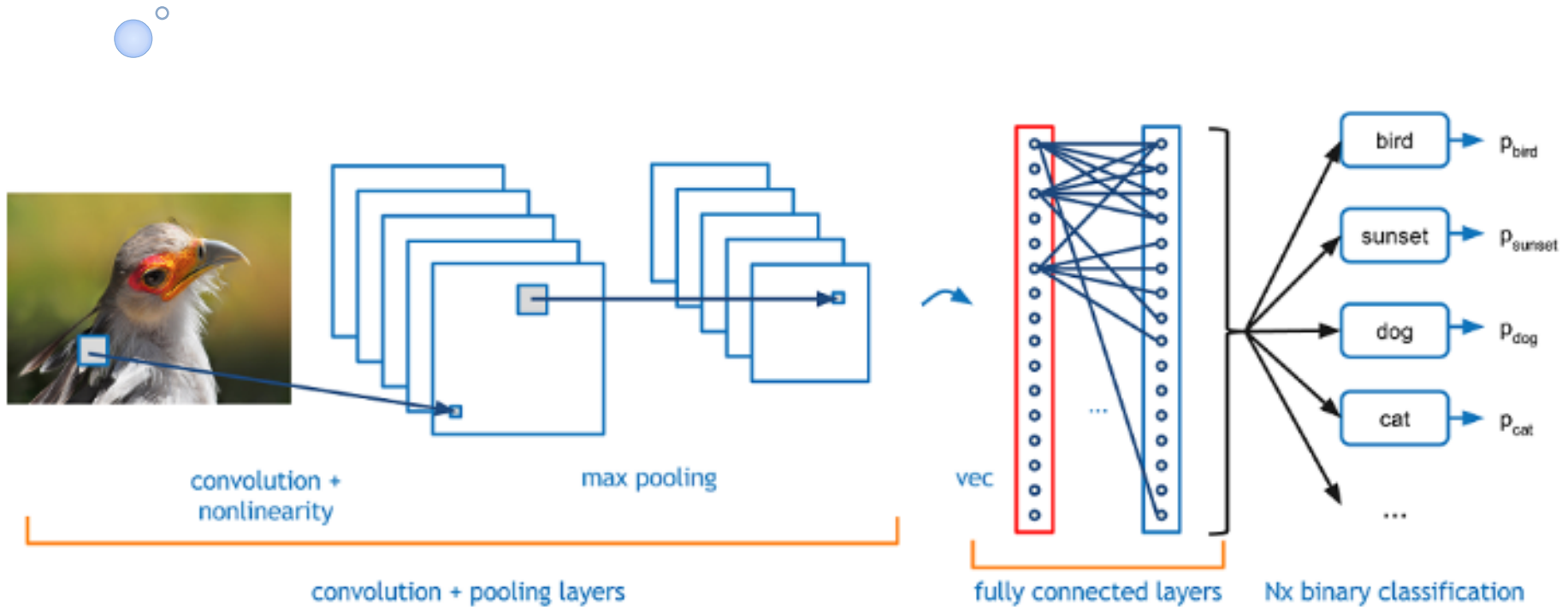
ABOUT CNN

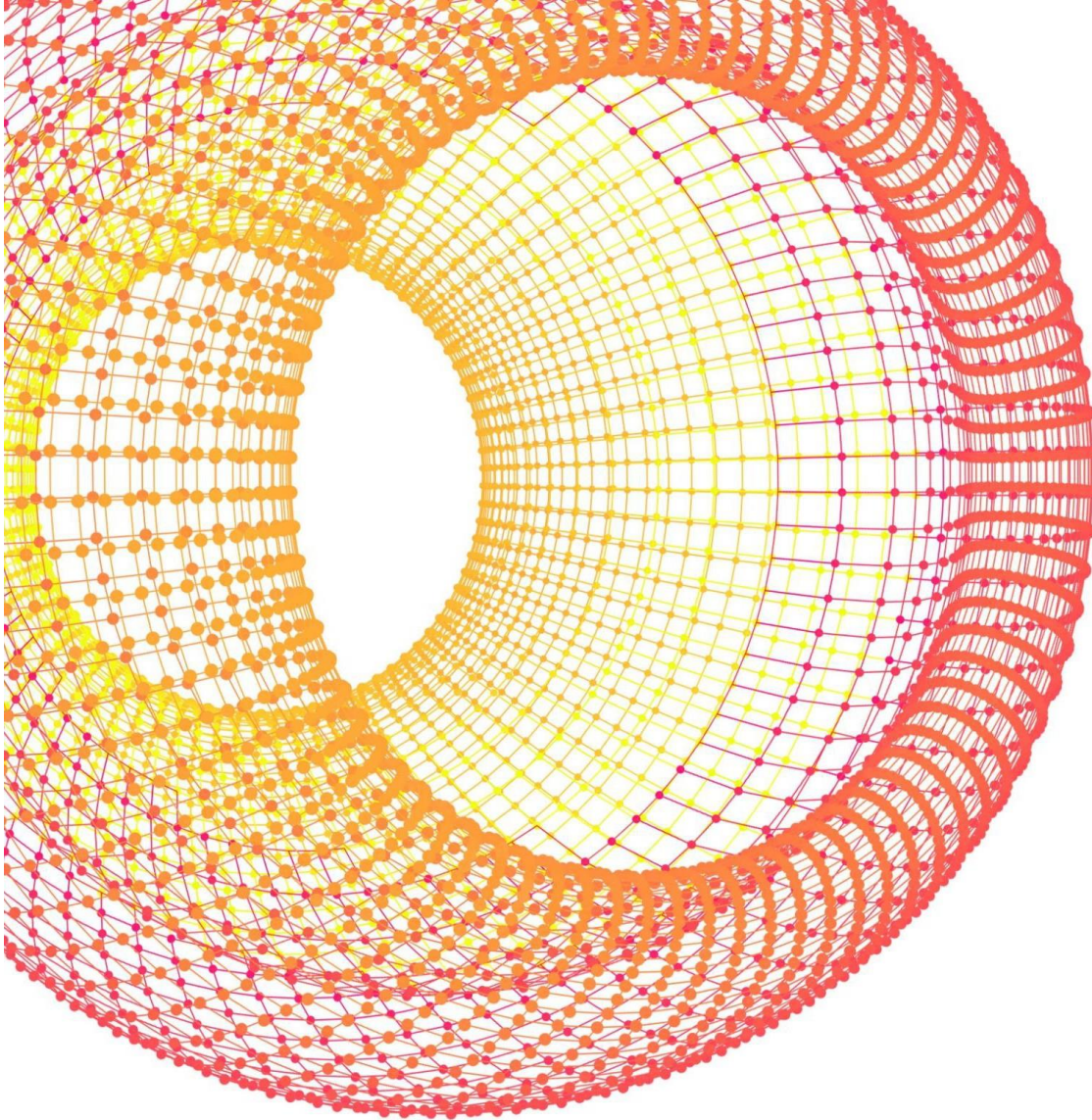
What is CNN

Convolutional neural network (ConvNets or CNNs) is one of the main categories to do images recognition, images classifications. Objects detections, recognition faces etc., are some of the areas where CNNs are widely used



Architecture of CNN



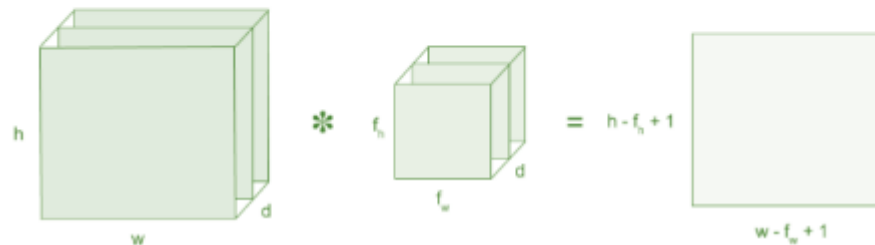


UNDERSTANDING CONVOLUTIONAL LAYERS IN CNN

Convolution Layer

- Computer read images as pixels and it is expressed as matrix($N \times N \times 3$) – (height by width by depth)
- The convolution layer makes use of a set of learnable filters. A filter is used to detect the presence of specific features or patterns present in the original image(input).
- It is usually expressed as a matrix, with a smaller dimension but the same depth as the input file.
- This filter is convolved across the width and height of the input file and a dot product is computed to give an activation map

- An image matrix (volume) of dimension **($h \times w \times d$)**
- A filter **($f_h \times f_w \times d$)**
- Outputs a volume dimension **($h - f_h + 1$) \times ($w - f_w + 1$) \times 1**



Convolution Layer

Source layer

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

Convolutional
kernel








-1	0	1
2	1	2
1	-2	0

Destination layer

	5						

$$\begin{aligned} &(-1 \times 5) + (0 \times 2) + (1 \times 6) + \\ &(2 \times 4) + (1 \times 3) + (2 \times 4) + \\ &(1 \times 3) + (-2 \times 9) + (0 \times 2) = 5 \end{aligned}$$

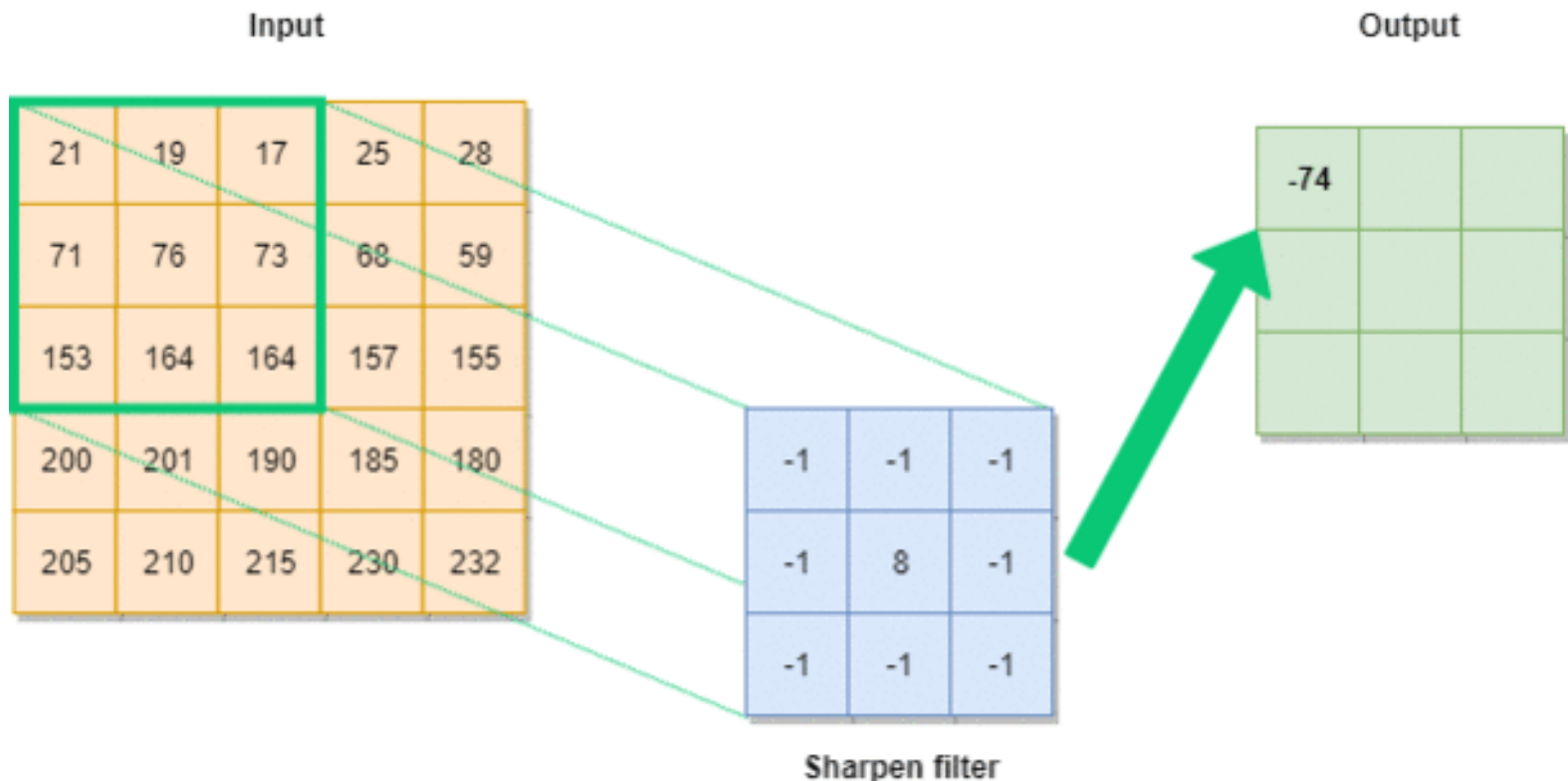
Convolution Layer

Operation	Filter	Convolved Image
Identity	$\begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$	
Edge detection	$\begin{bmatrix} 1 & 0 & -1 \\ 0 & 0 & 0 \\ -1 & 0 & 1 \end{bmatrix}$	
	$\begin{bmatrix} 0 & 1 & 0 \\ 1 & -4 & 1 \\ 0 & 1 & 0 \end{bmatrix}$	
	$\begin{bmatrix} -1 & -1 & -1 \\ -1 & 8 & -1 \\ -1 & -1 & -1 \end{bmatrix}$	
Sharpen	$\begin{bmatrix} 0 & -1 & 0 \\ -1 & 5 & -1 \\ 0 & -1 & 0 \end{bmatrix}$	
Box blur (normalized)	$\frac{1}{9} \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$	
Gaussian blur (approximation)	$\frac{1}{16} \begin{bmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{bmatrix}$	

Convolution Layer

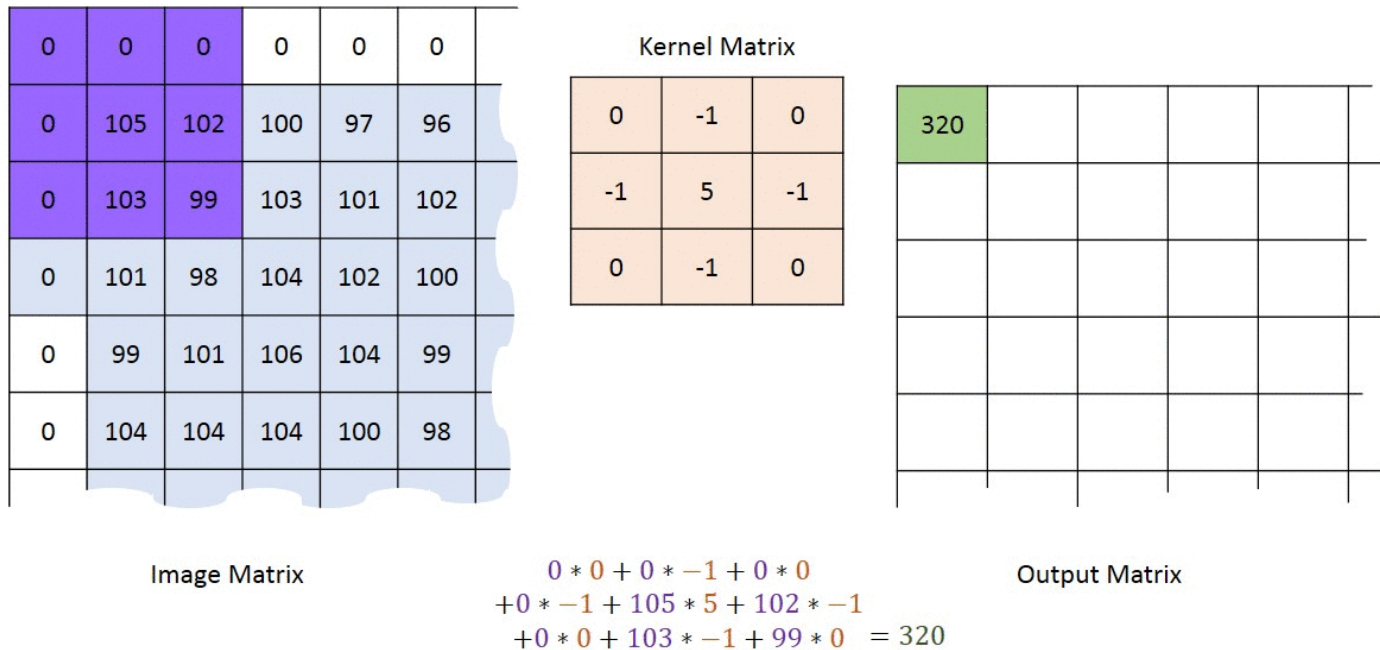
The concept of stride and padding:

- The weights of a matrix moves 1 pixel at a time is called as stride 1
- This is how stride works

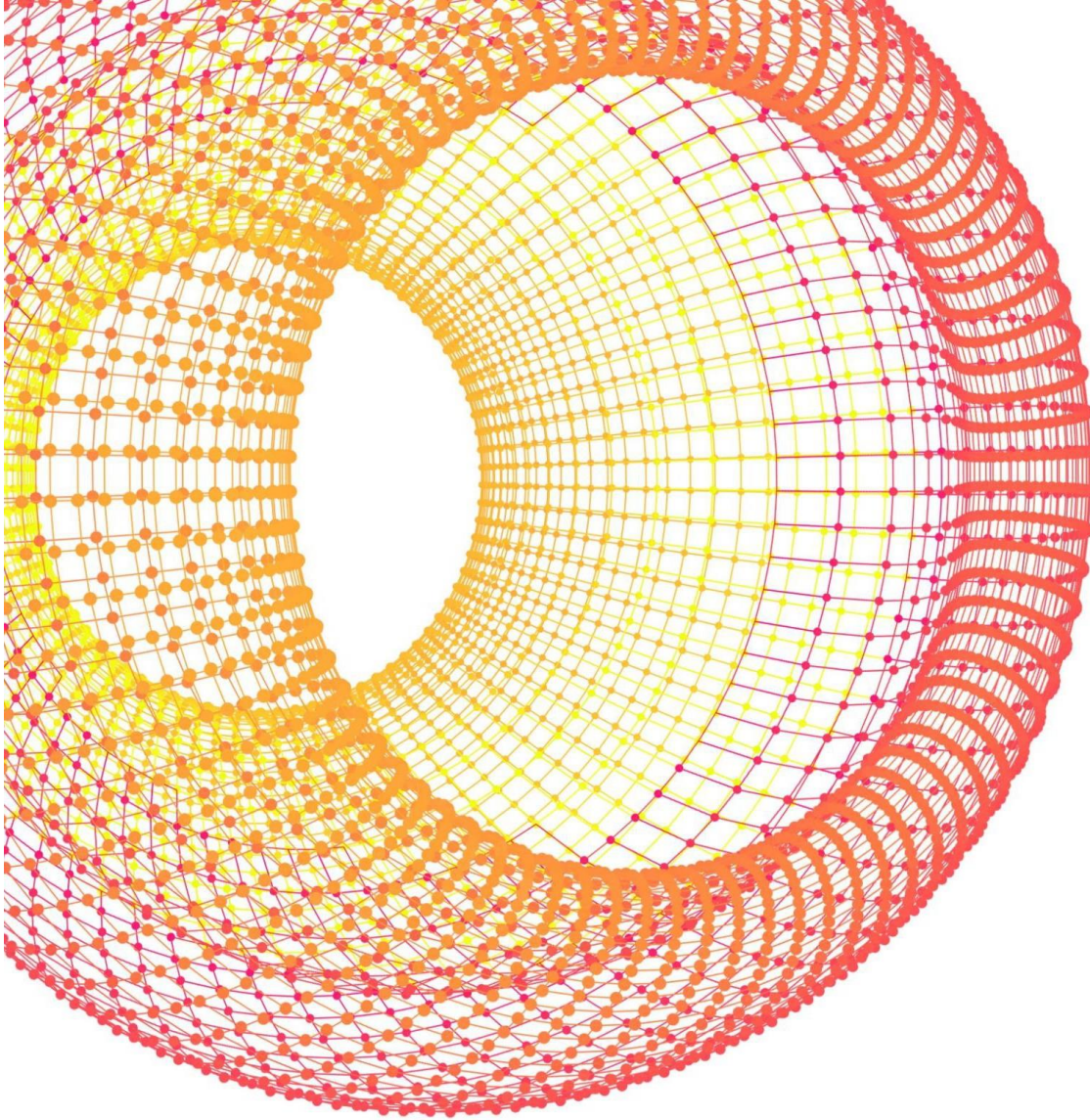


Convolution Layer

- As we can see in above slide the increase in the stride value decreases the size of the image (which may cause in losing the features of the image).
- Padding the input image across it solves our problem, we add more than one layer of zeros around the image in case of higher stride values.



**Convolution with horizontal and
vertical strides = 1**



UNDERSTANDING POOLING LAYERS IN CNN

The Pooling Layer

- It can be seen in between the convolution layers in CNN Architecture.
- The layer basically reduces the amount of parameters and computation in the network.
- Pooling is done for the sole purpose of reducing the spatial size of the image.
- Pooling is done independently on each depth dimension, therefore the depth of the image remains unchanged. The most common form of pooling layer generally applied is the max pooling.

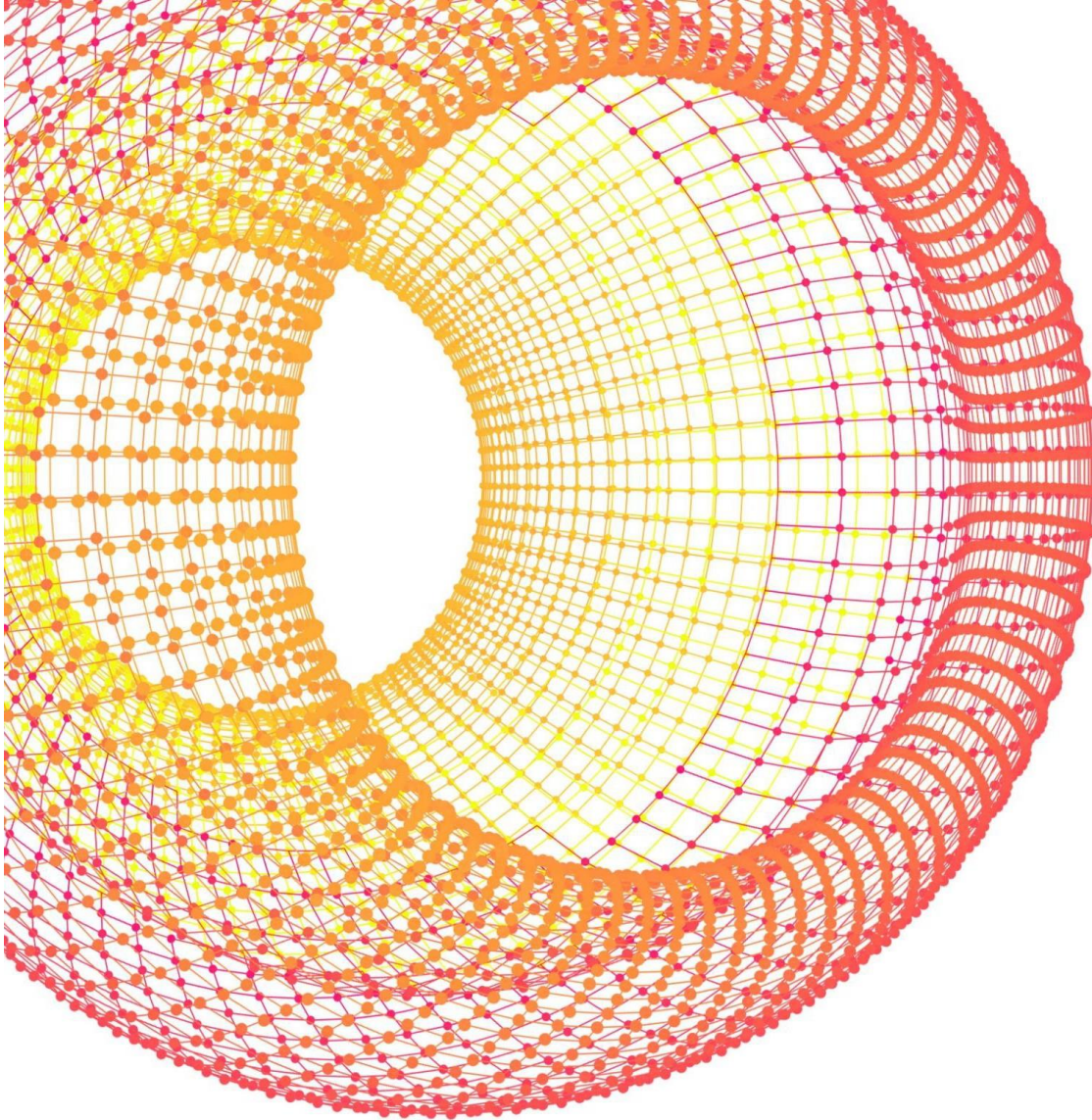
Feature Map

6	6	6	6
4	5	5	4
2	4	4	2
2	4	4	2

**Max
Pooling**

**Average
Pooling**

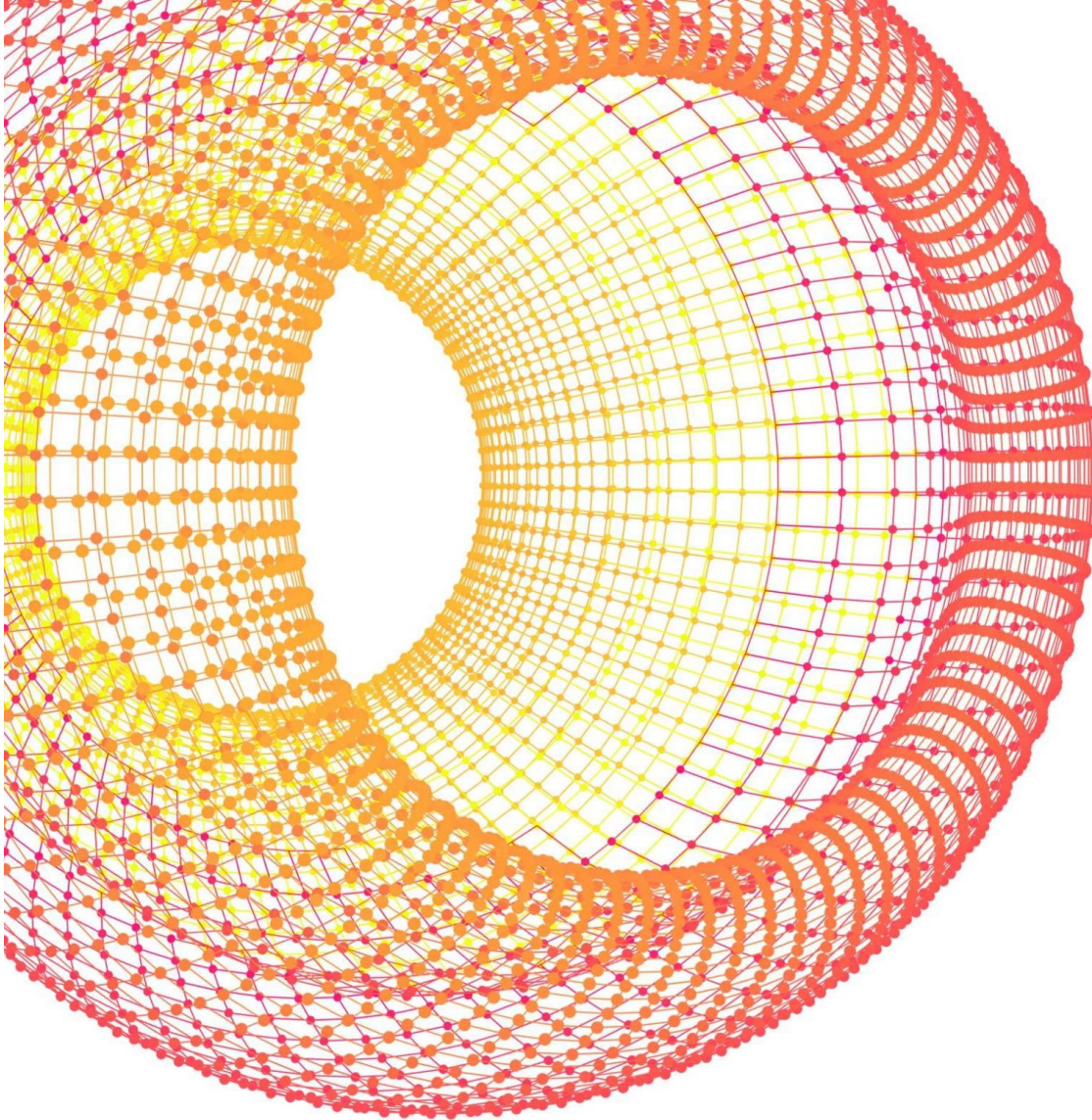
**Sum
Pooling**



UNDERSTANDING OUTPUT LAYERS

The Output Layer

- After multiple layers of convolution and padding, we would need the output in the form of a class.
- The convolution and pooling layers would only be able to extract features and reduce the number of parameters from the original images.
- However, to generate the final output we need to apply a fully connected layer to generate an output equal to the number of classes we need.
- Convolution layers generate 3D activation maps while we just need the output as whether or not an image belongs to a particular class.
- The output layer will have the loss function like categorical cross-entropy, to compute the error in prediction.
- Once the forward pass is complete the backpropagation begins to update the weight and biases for error and loss reduction.



DECIDING NO OF LAYERS

DECIDING NO OF LAYERS

How to decide the number of convolution layers and number of filters in CNN ?

- More layers networks is always better, at the cost of more data and increased complexity of learning.
- You should initially use fewer filters and gradually increase and monitor the error rate to see how it is varying.
- Very small filter sizes will capture very fine details of the image. On the other hand having a bigger filter size will leave out minute details in the image

Applications of CNN

Based on the problems, we have the different CNN's which are used in computer vision. The five major computer vision techniques which can be addressed using CNN.

- ❖ Image Classification
- ❖ Object Detection
- ❖ Object Tracking
- ❖ Semantic Segmentation
- ❖ Instance Segmentation

THANK YOU