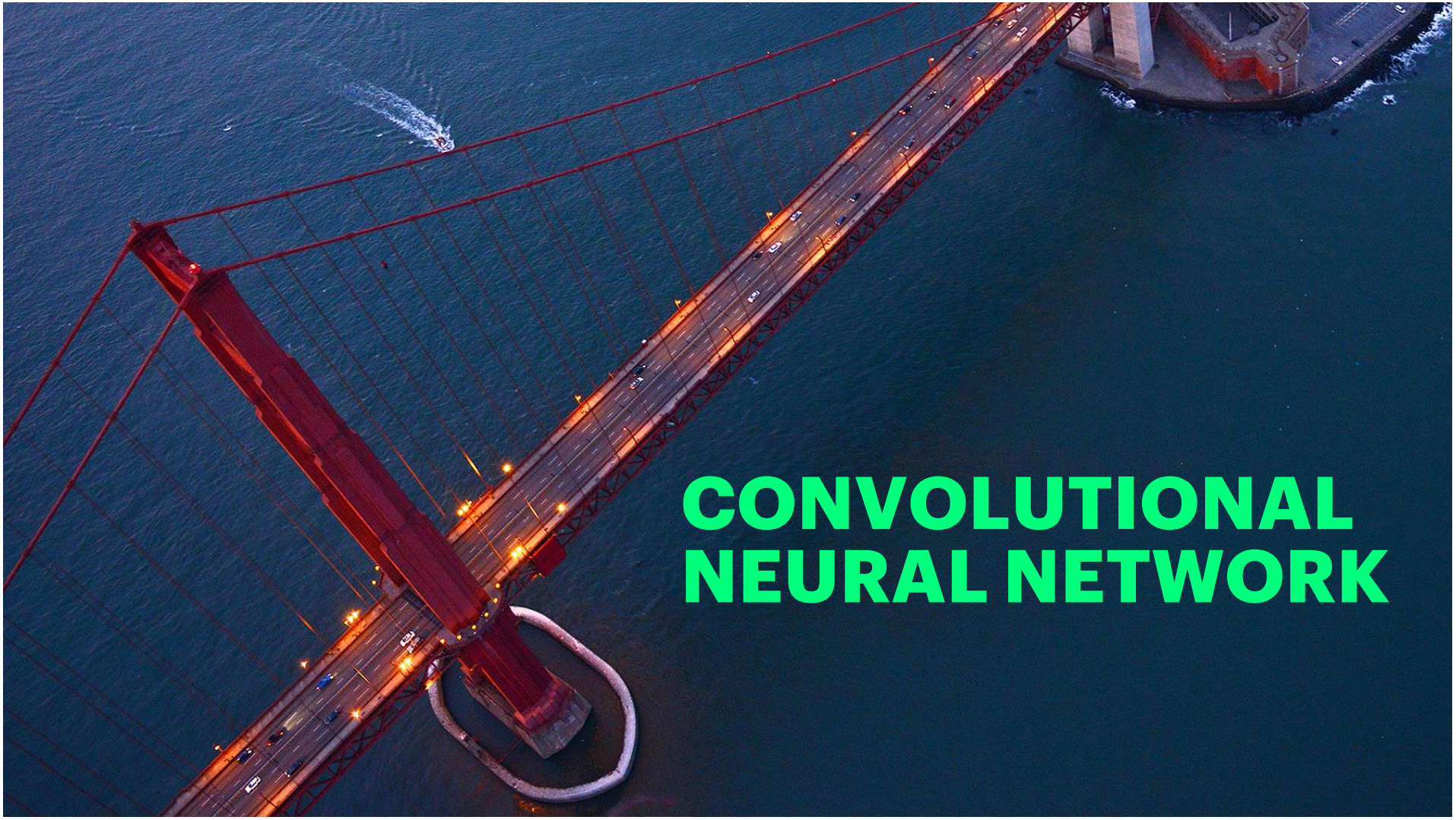
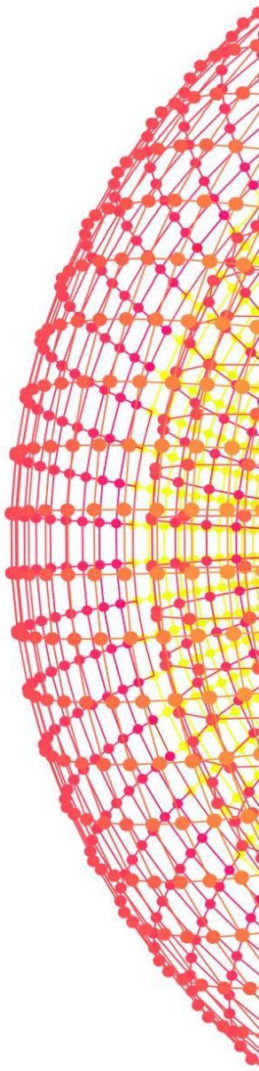
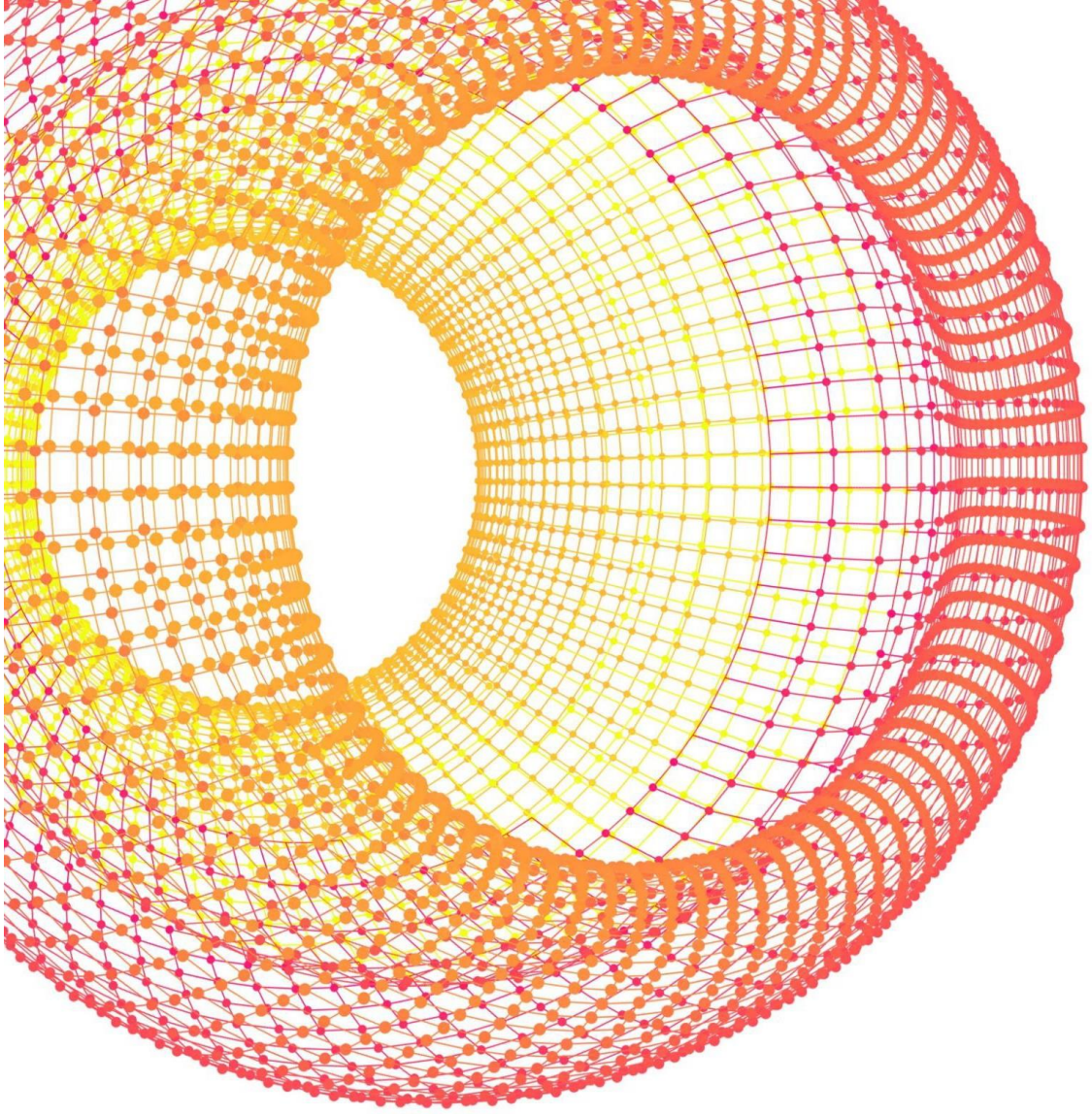


AGENDA

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





Professional School
ISO 9001 : 2000 Organisation

INTRODUCTION TO COMPUTER VISISON

Introduction to Computer Vision

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

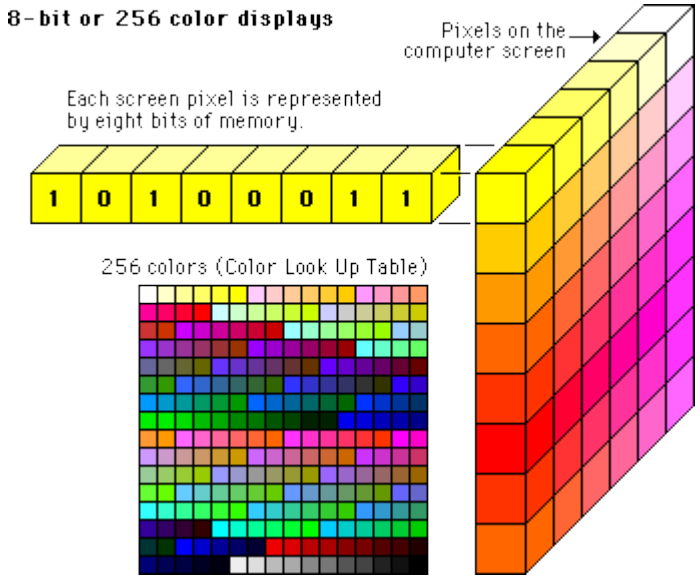
Computer Vision system is a neural network designed to solve a specific visual problem. It looks at a lot of examples of a problem and learns from those examples (training)..

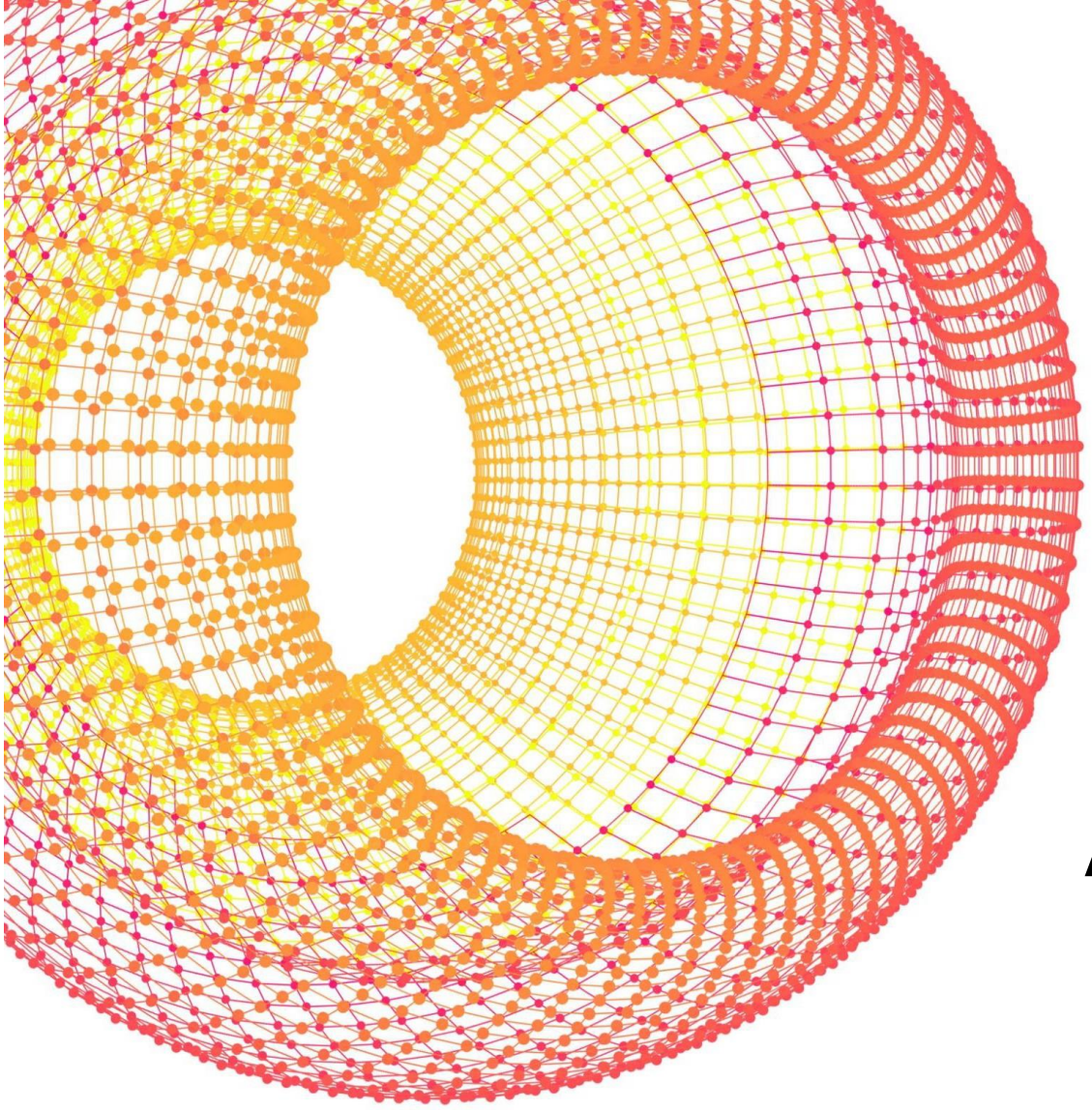
Introduction to Computer Vision

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



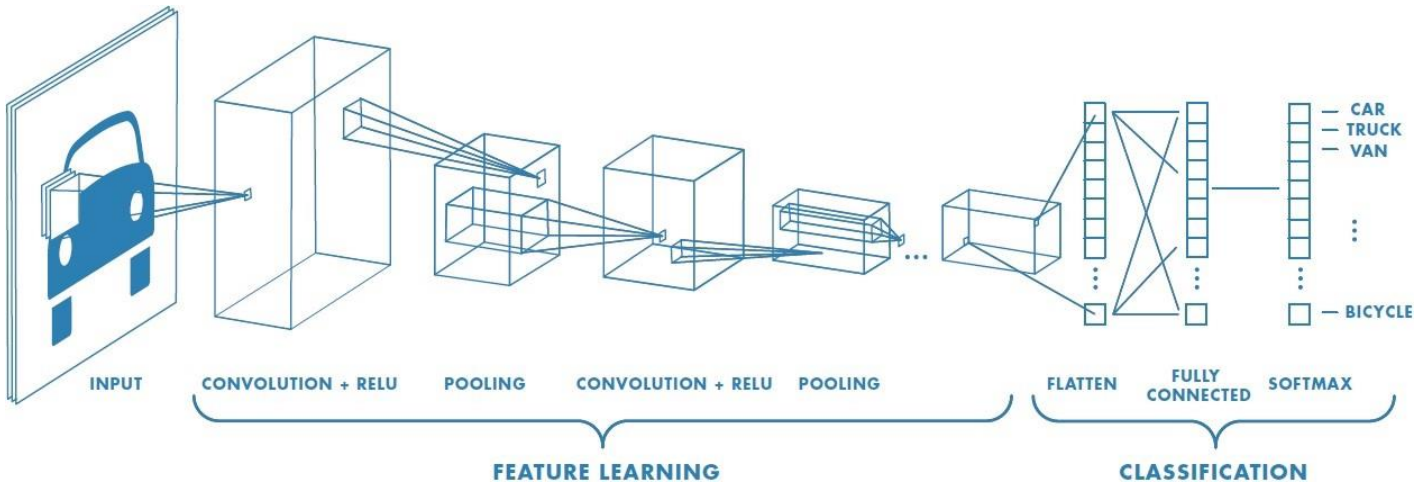


Professional School
ISO 9001 : 2000 Organisation

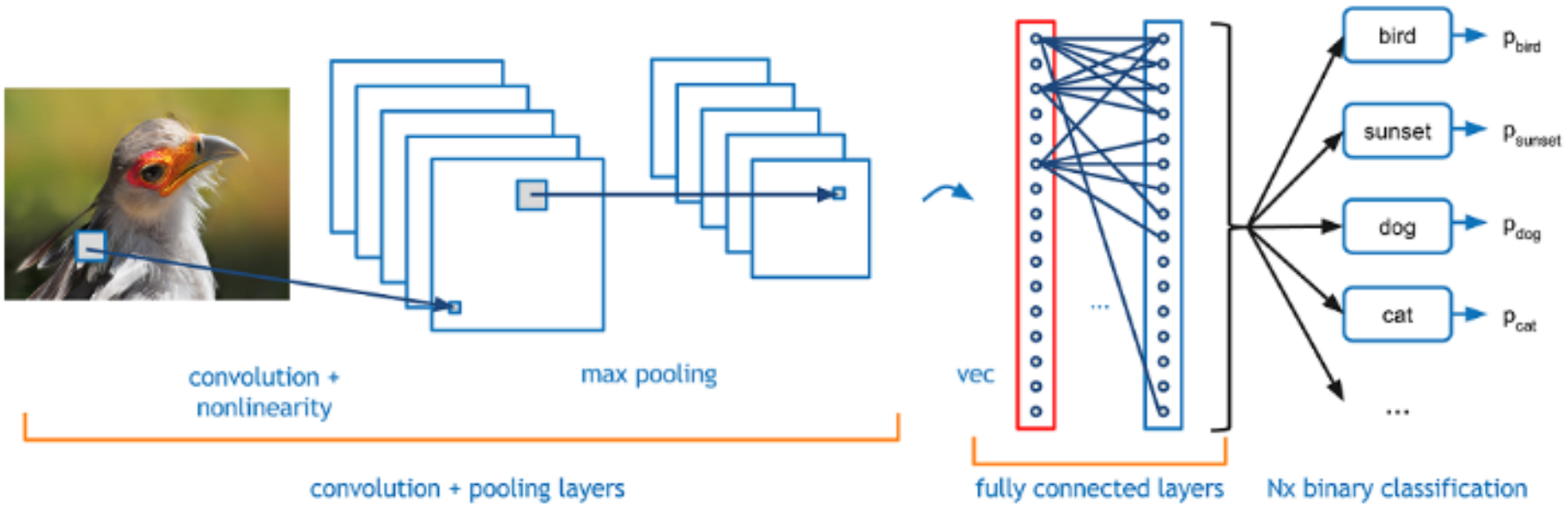
ABOUT CNN

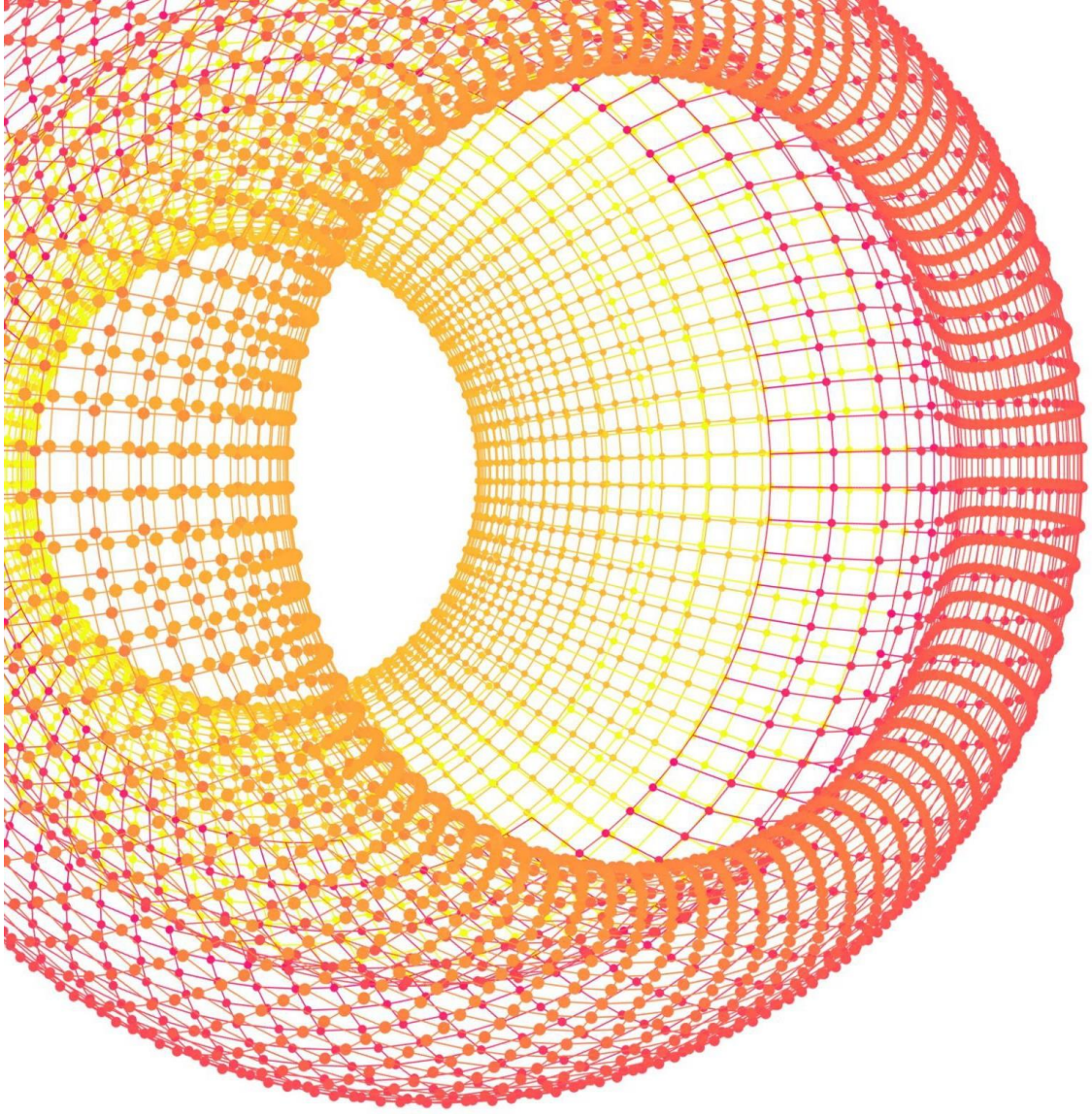
What is CNN

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





Professional School

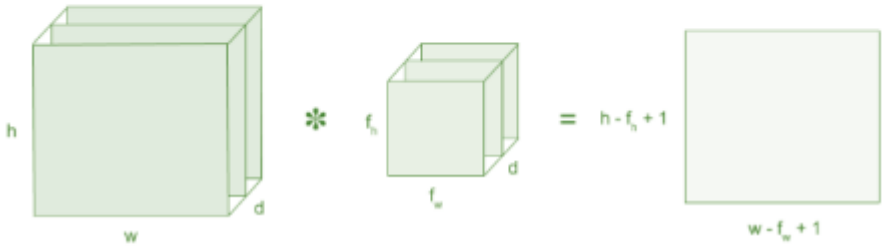
ISO 9001 : 2000 Organisation

UNDESTANDING CONVOLUTIONAL LAYERS IN CNN

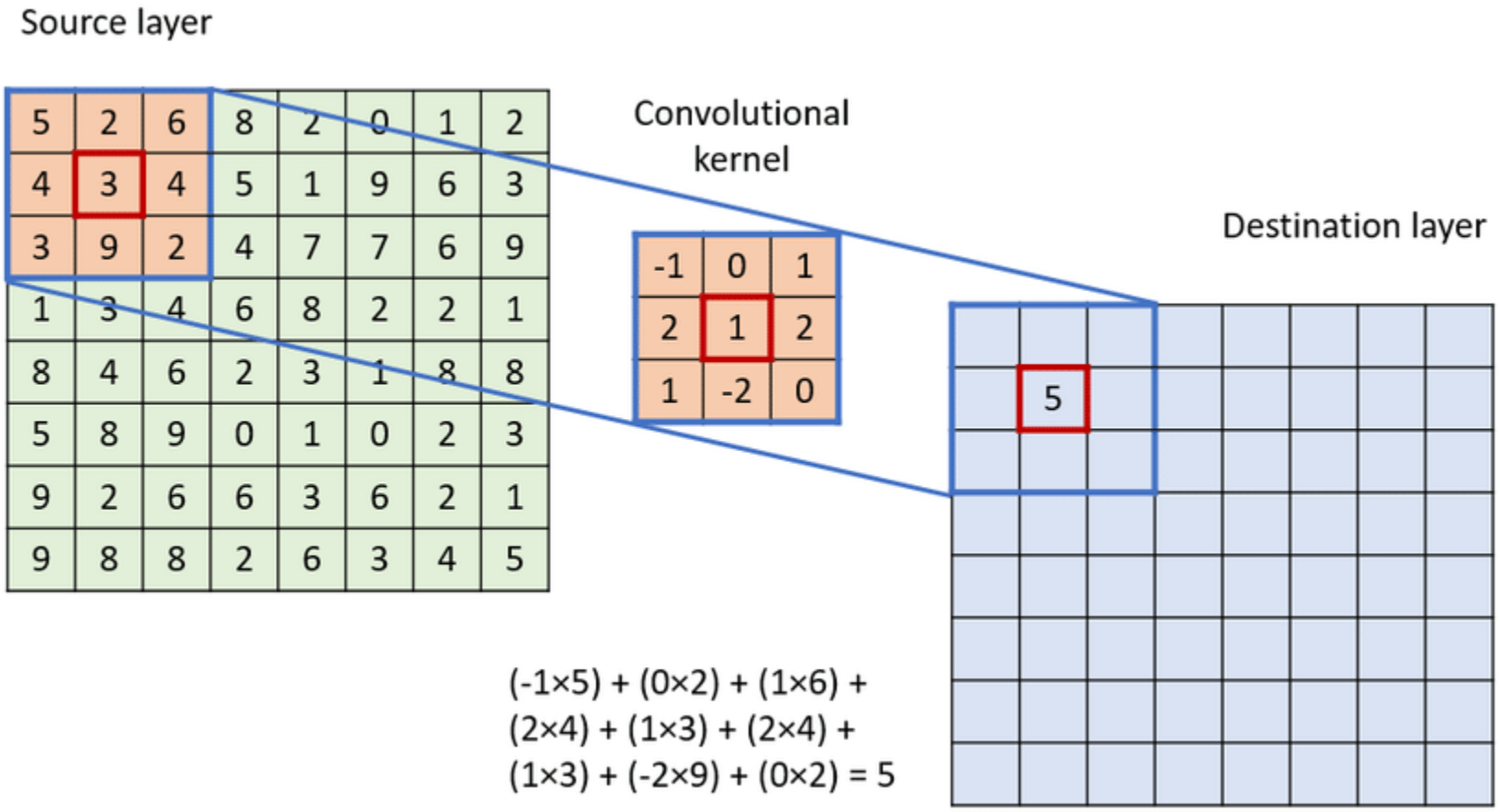
Convolution Layer

- Computer read images as pixels and it is expressed as matrix(NxNx3) – (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 x w x d)**
- A filter **(f_h x f_w x d)**
- Outputs a volume dimension **(h - f_h + 1) x (w - f_w + 1) x 1**



Convolution Layer



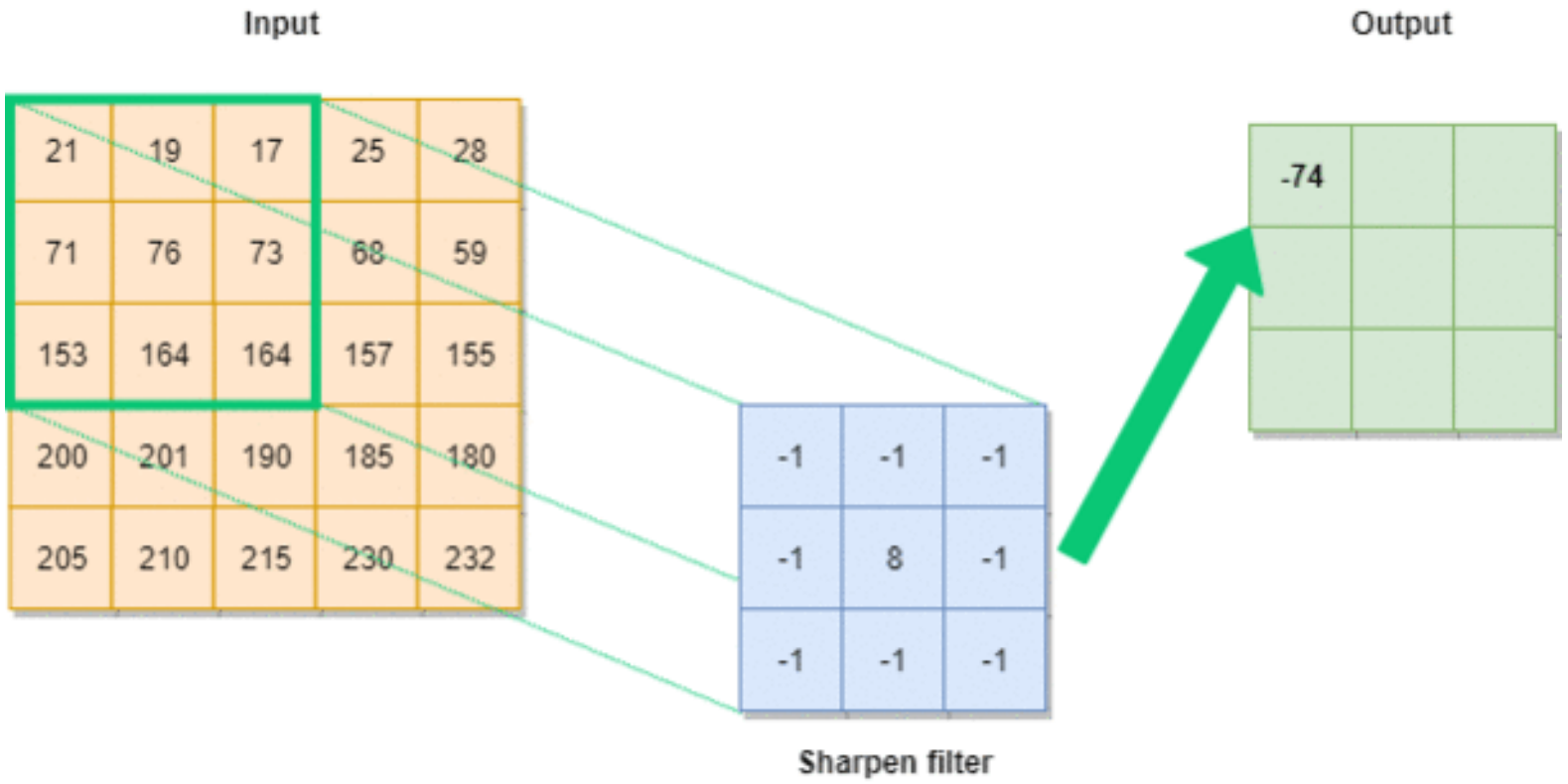
Convolution Layer

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

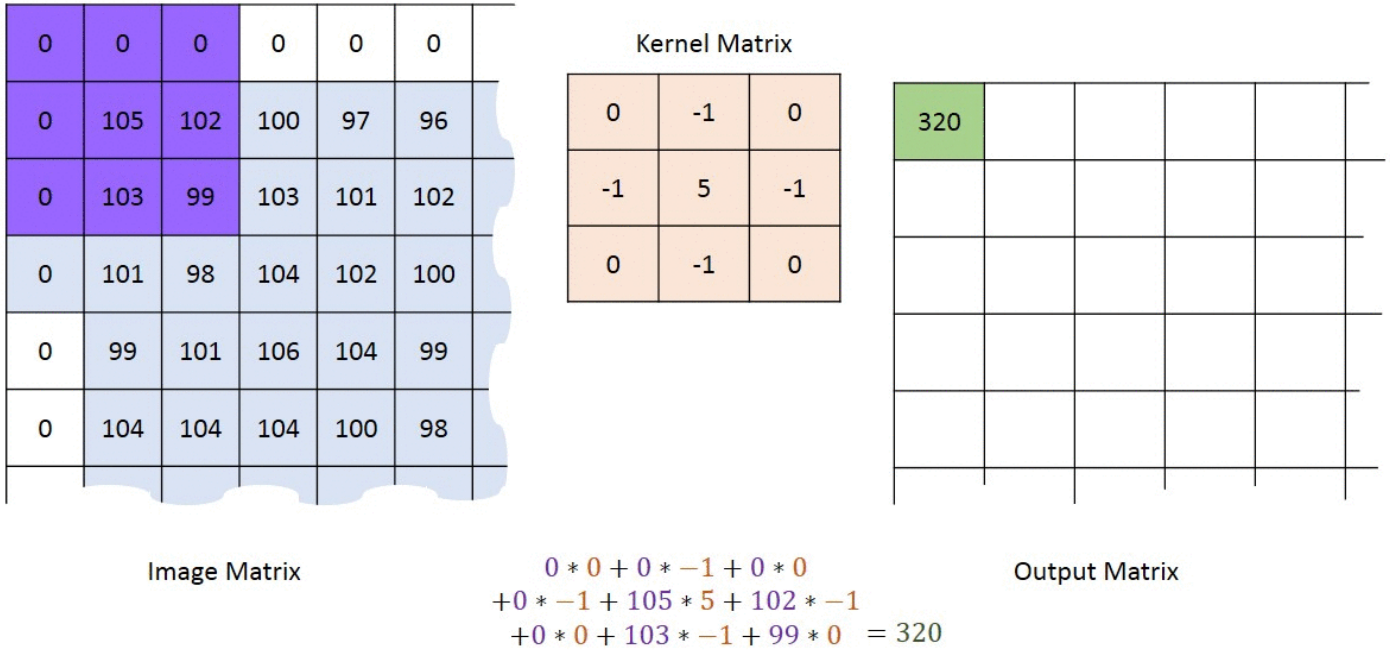
The concept of stride and padding:

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

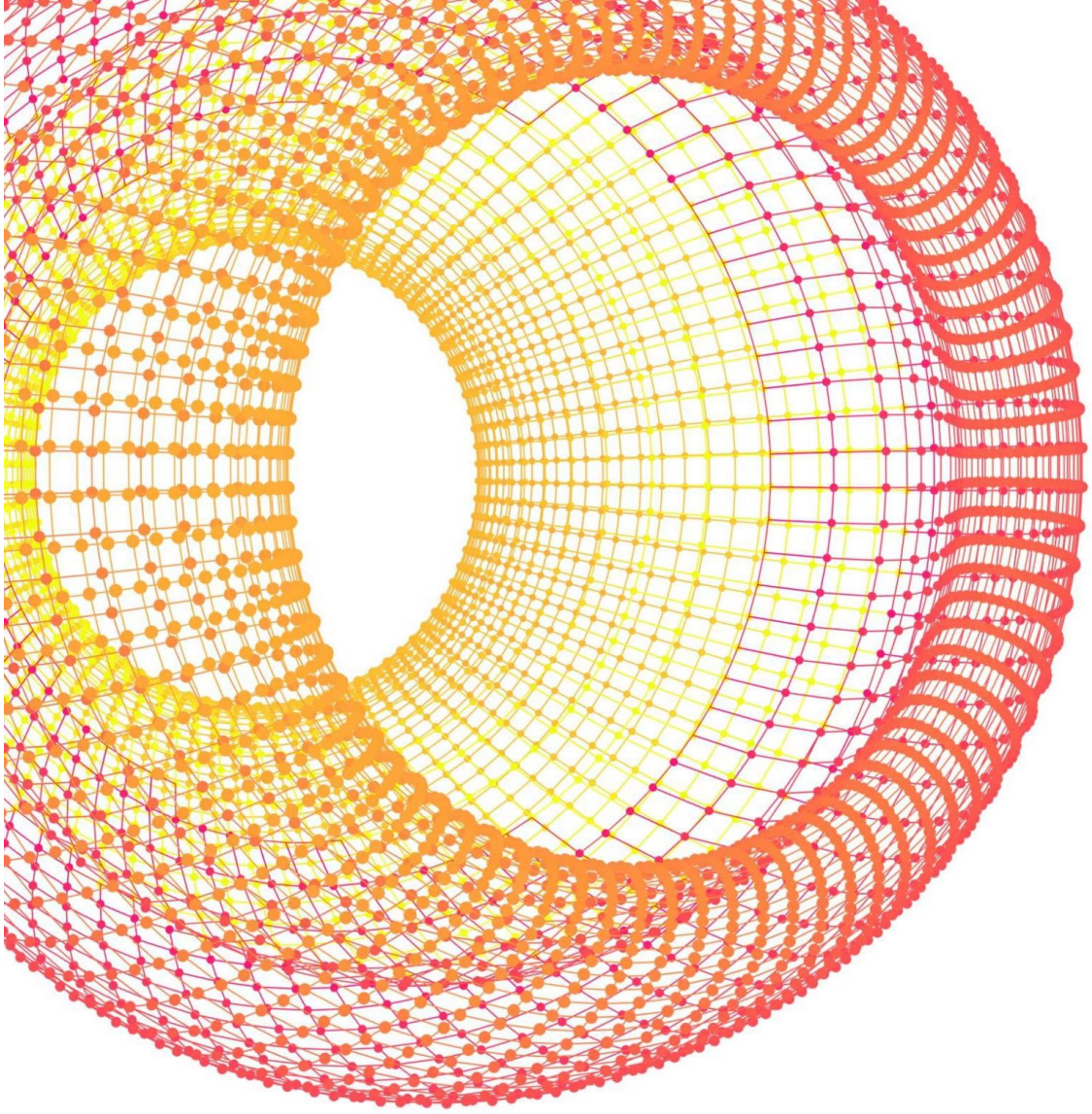


Convolution Layer

- 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



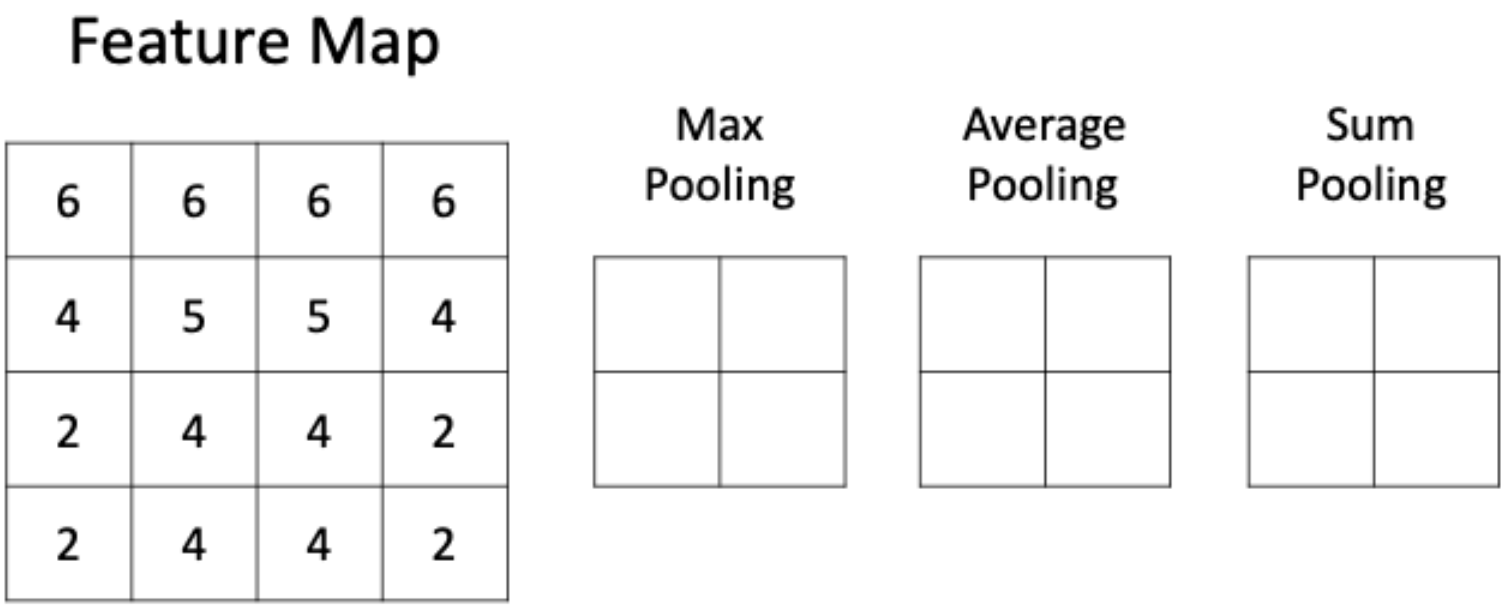
Professional School

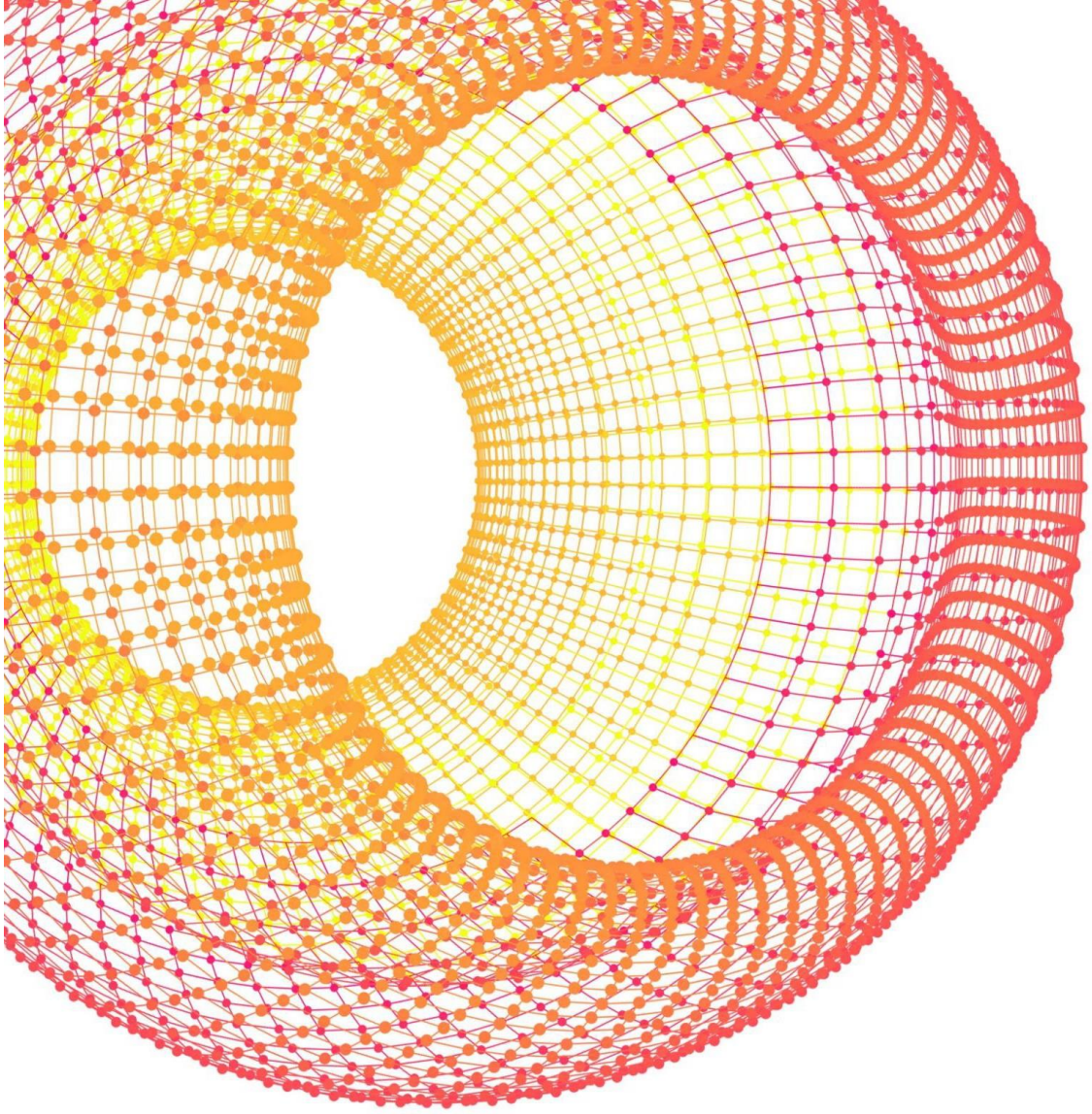
ISO 9001 : 2000 Organisation

UNDESTANDING 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.



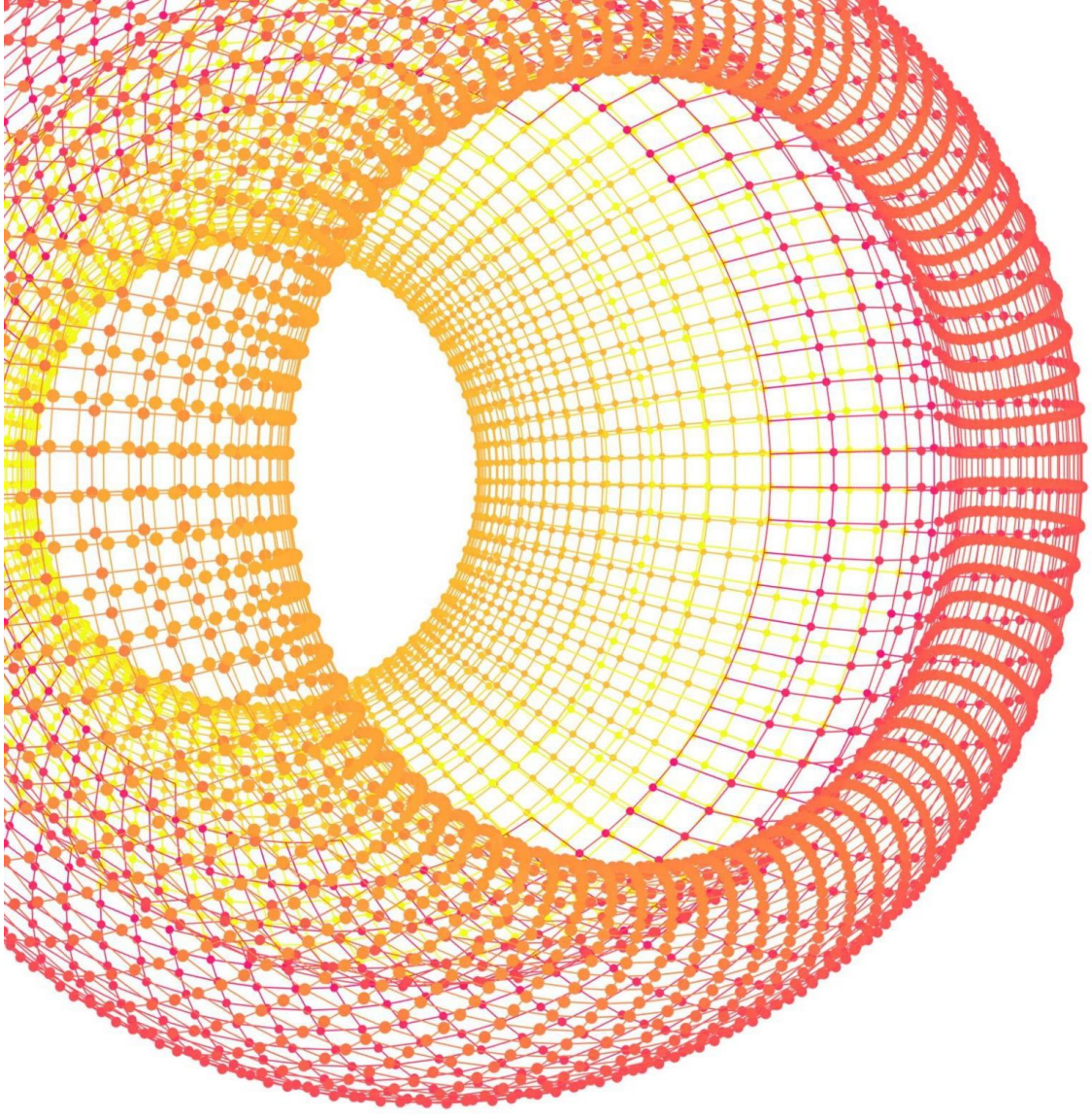


Professional School
ISO 9001 : 2000 Organisation

UNDESTANDING 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.



Professional School
ISO 9001 : 2000 Organisation

DECIDING NO OF LAYERS

DECIDING NO OF LAYERS

Professional School
ISO 9001 : 2000 Organisation

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

Professional School
ISO 9001 : 2000 Organisation

THANK YOU