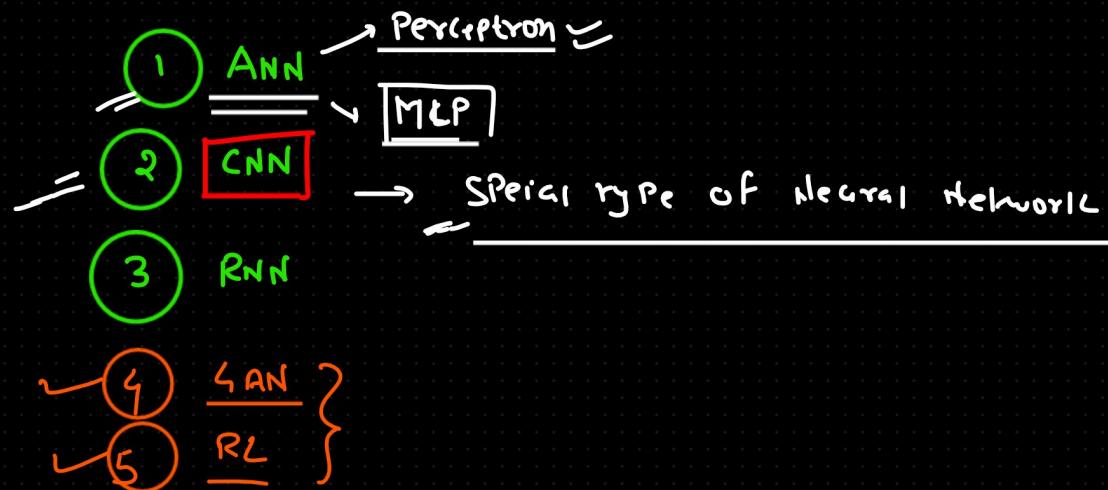


Convolution Neural Network



CNN is the building block of Computer Vision.
CNN is used for the image classification and other image based operations such as object detection, object segmentation etc.

Ques: Why ANN alone cannot be used for image based dataset and we need a special kind of neural network called CNN?

For Example: 2D image dataset like MNIST or Fashion MNIST dataset can be flatten in 1D and then passed to the input layer of ANN and then the resultant ANN can be used to make the prediction with good accuracy. Then why one needs a special type of neural network in the form of CNN?

Ans: This is because as the resolution of the images increases (which is true in the real world scenario where we have advanced cameras capturing much smaller pixels), no. of trainable parameters needed(which is propositional to the number of pixel in the image) to make the intended image based operations(classification) also increases with a much higher rate. So, to overcome that we avoid using ANN and use CNN in order to extract only important features or pixels(by applying filters and pooling), which in turn reduces the no. of neurons or nodes needed at the input layer thereby, decreasing the trainable parameters as well.

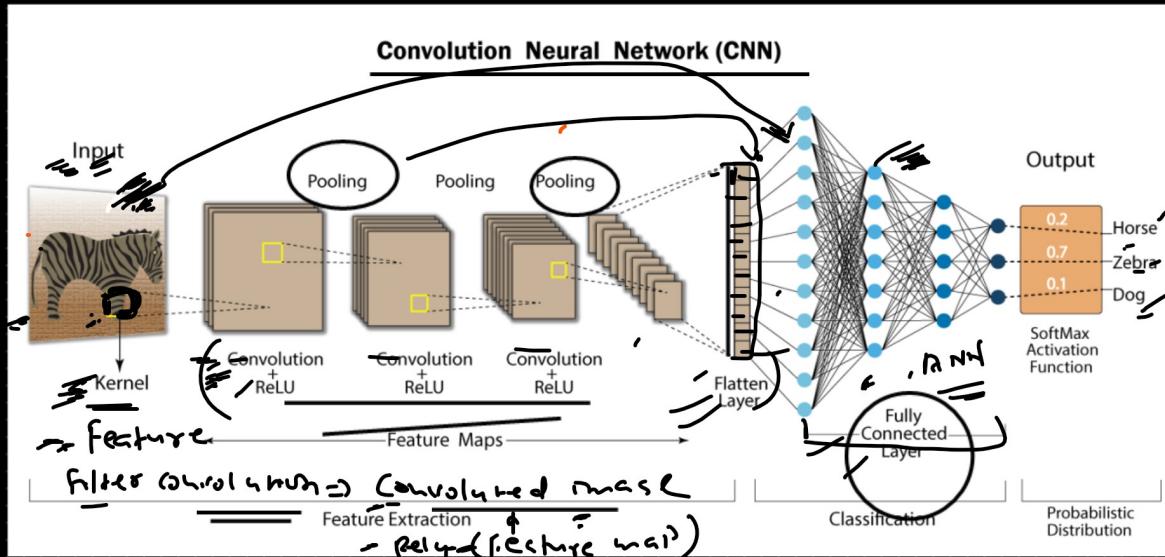
Which type of Data we use [CNN?]

= image, video \Rightarrow grid like structure
 $\Leftrightarrow (2-D, 3D) =$
 $\Leftrightarrow (\text{Matrix}) =$

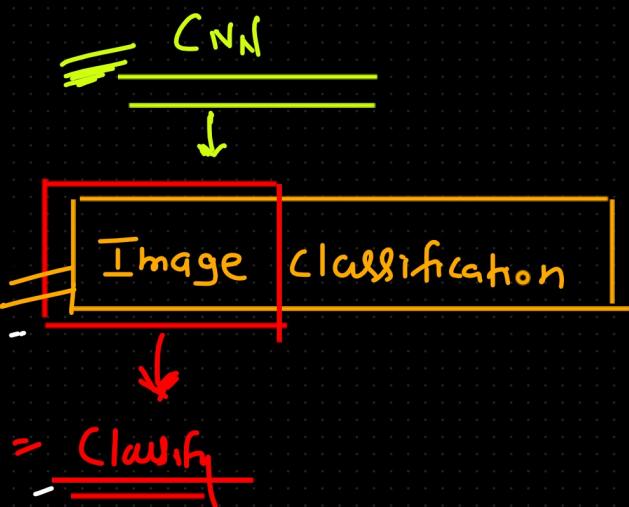
Dataset having the Grid Like structure (meaning in 2d, 3d, matrix, vector) is utilized by CNN.
Eg; image, video etc

Please note the FILTERS or KERNELS in CNN are similar to what we have as weights in ANN which is a learnable parameter.

LENET-5 was the first CNN architecture by Yann LeCun who is known as the father of CNN.



See all the terms carefully in this image



$\equiv \boxed{\text{boy}} \Rightarrow ?$

Sunny | Krish |
Sudhanshu |
Happy |

Image \Rightarrow Collection of Pixels \Rightarrow Pixel \rightarrow Numerical value

\hookrightarrow Gray image
(Black and White)
Single channel

$\hookrightarrow [0-255]$
Black, white

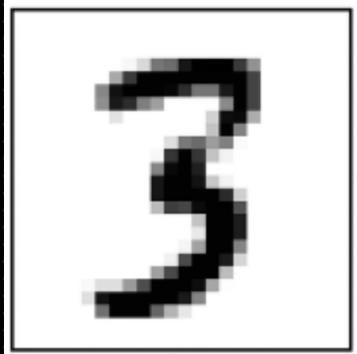


\hookrightarrow Color image (R, G, B)
 \hookrightarrow (Red) (Green) (Blue)
3 channel.

- Pixel is the numeric value having the range [0-255].
- In Colored Image we have 3 channels which is represented using RGB intensities whereas, in the case of Grey image we have only Single Channel(Black and white)



Practical implementation of CNN is used for converting old Black and White movies into colored movies. Eg; Colored version of Sholay movie which was originally shot in black and white version



This is how humans see a digit 3

This is how computers will see a digit 3. Where each pixel is internally mapped to a numeric value

trained your mind to identify this particular

(CNN)

Convolution Neural Network

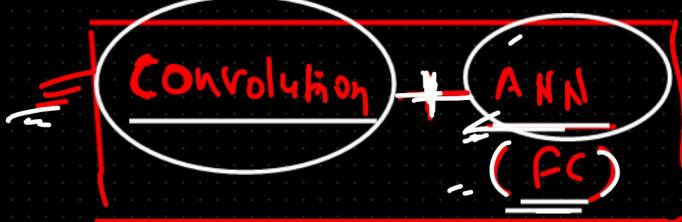


Image | Picture \Rightarrow Collection of Pixels

Numerical value

[0 - 255]

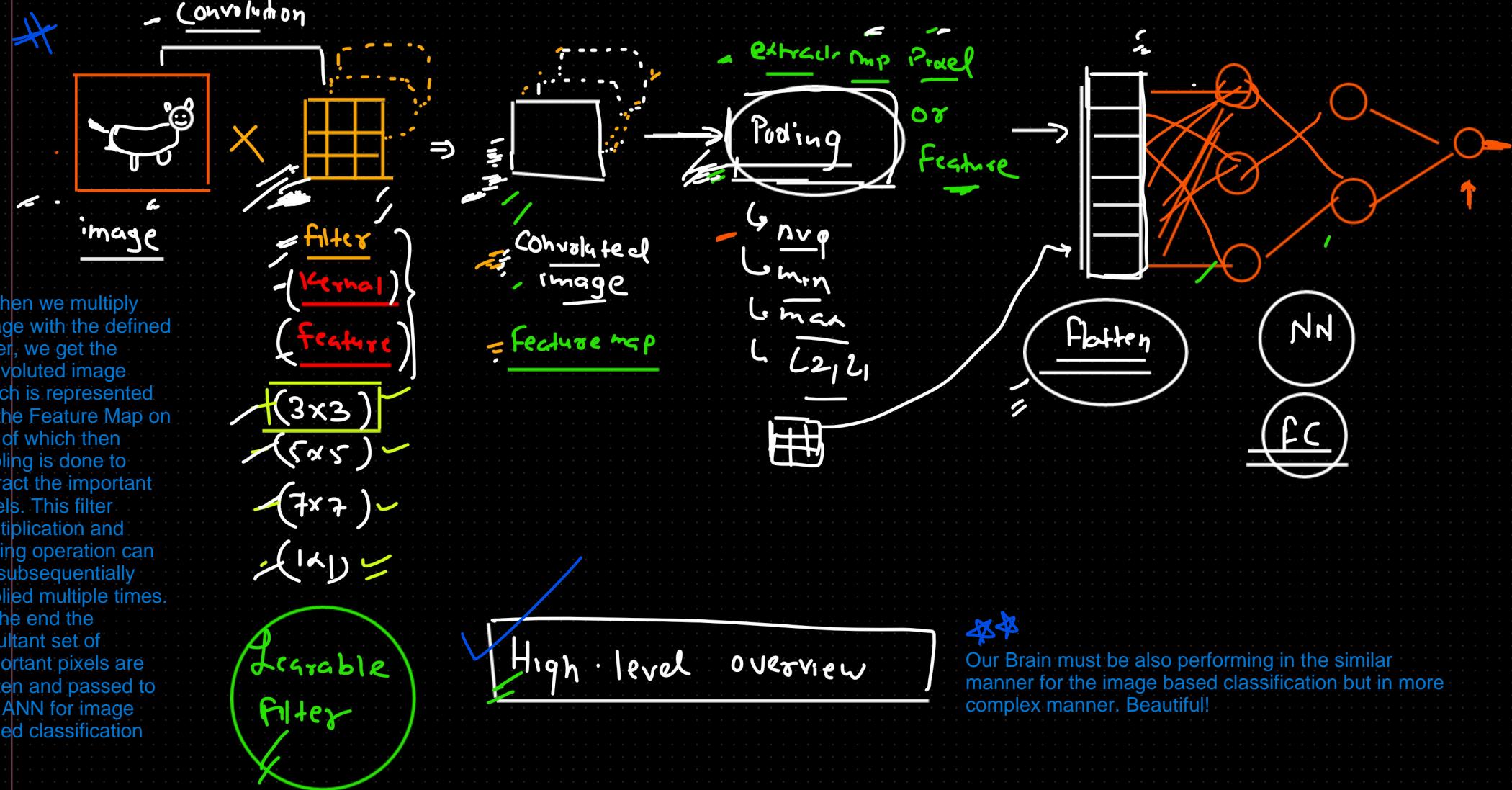
Block ↓ ← -
white



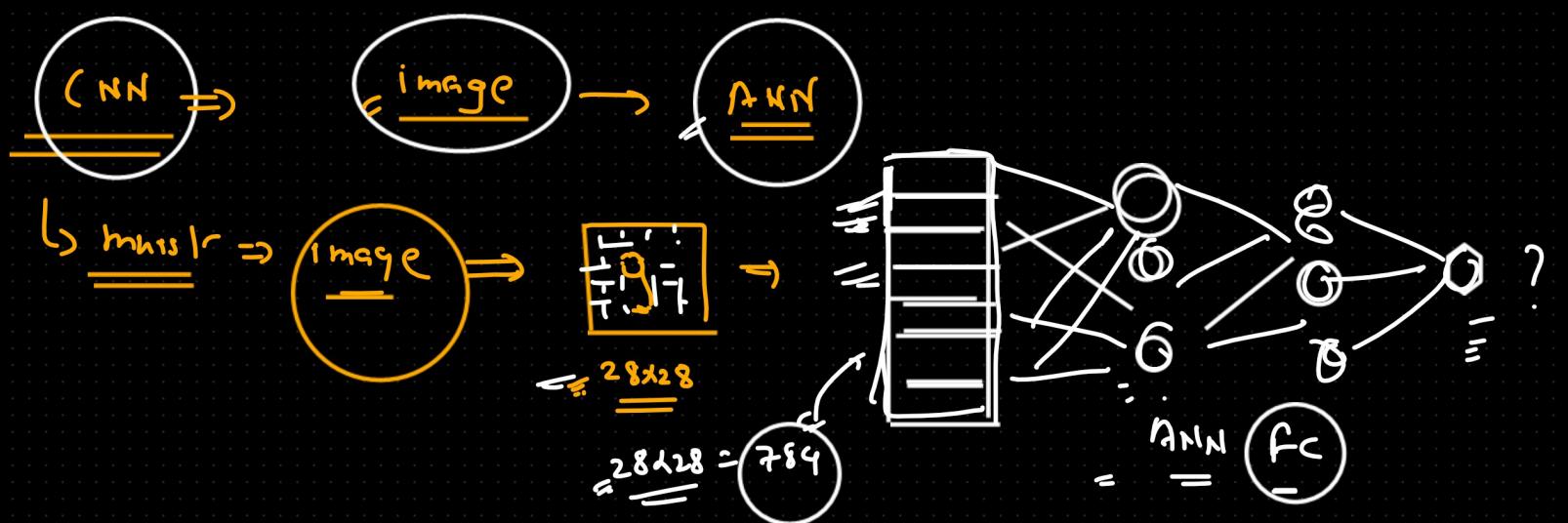
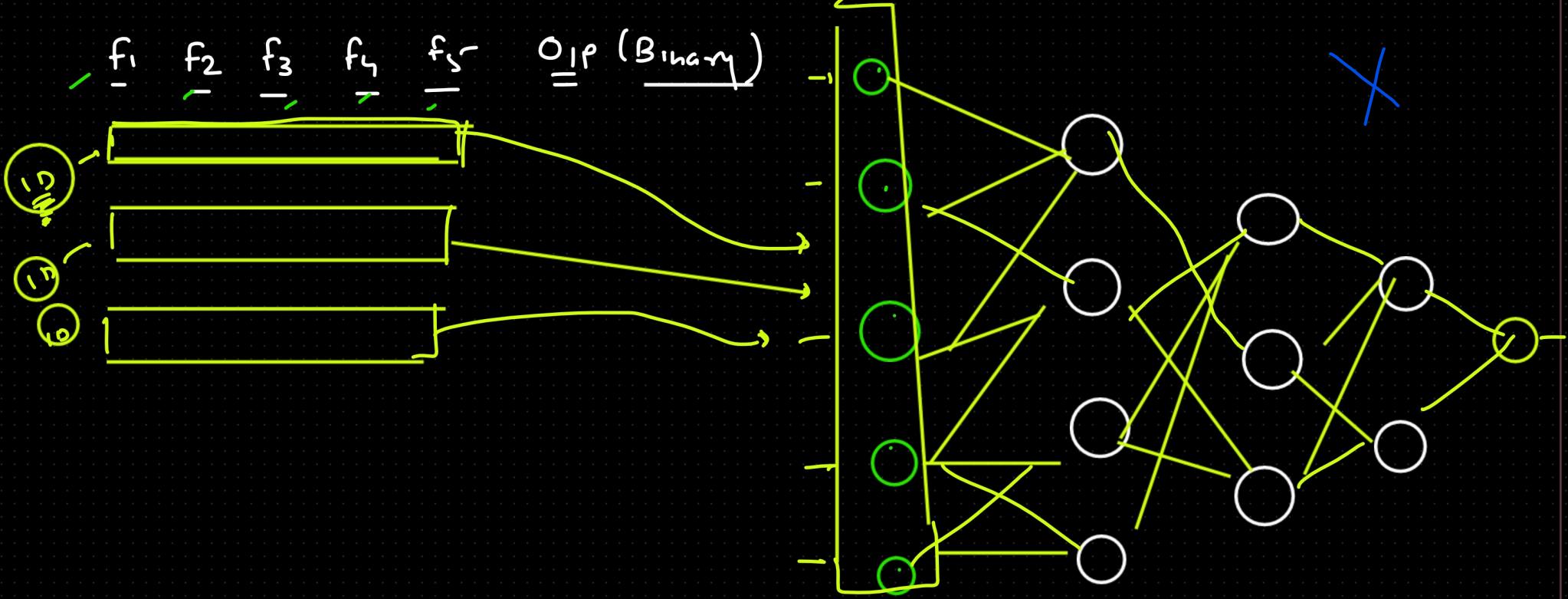
Ques: What is the Fundamental use case of the CNN?

Ans: For the Image Classification

To summarize CNN = Convolution + ANN.
Through Convolution we form convoluted image where we attempt to perform Feature Extraction where we extract only important features from the resultant feature map and then use them to train the fully connected ANN to perform the image based classification



Our Brain must be also performing in the similar manner for the image based classification but in more complex manner. Beautiful!



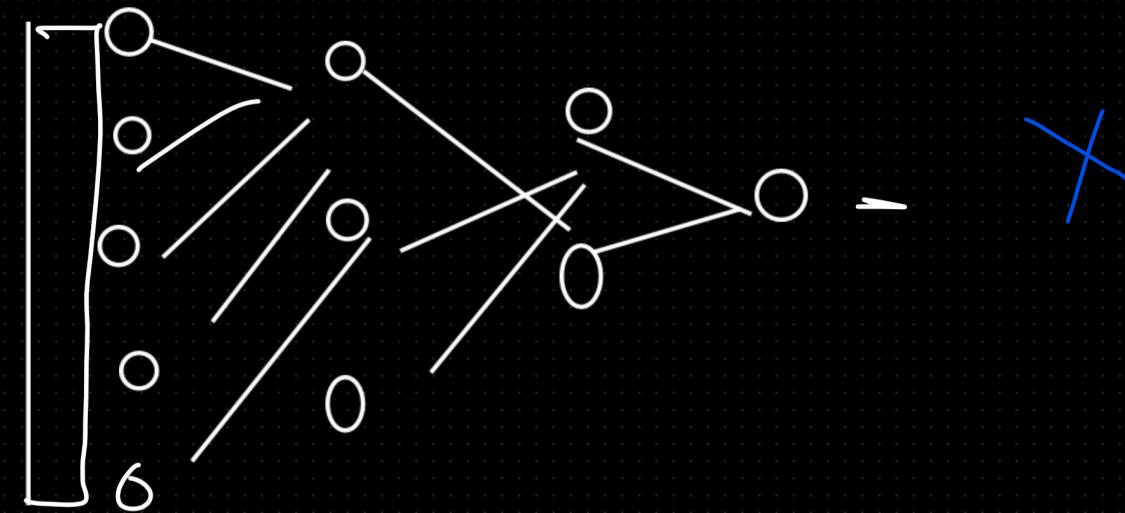
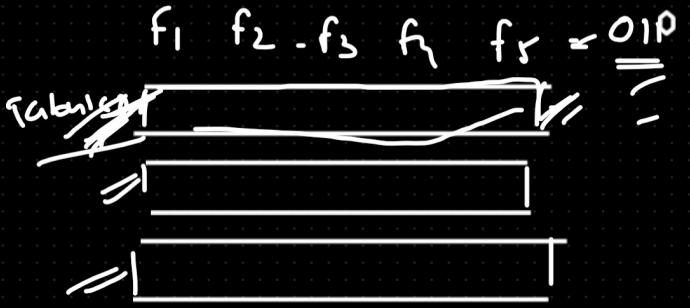
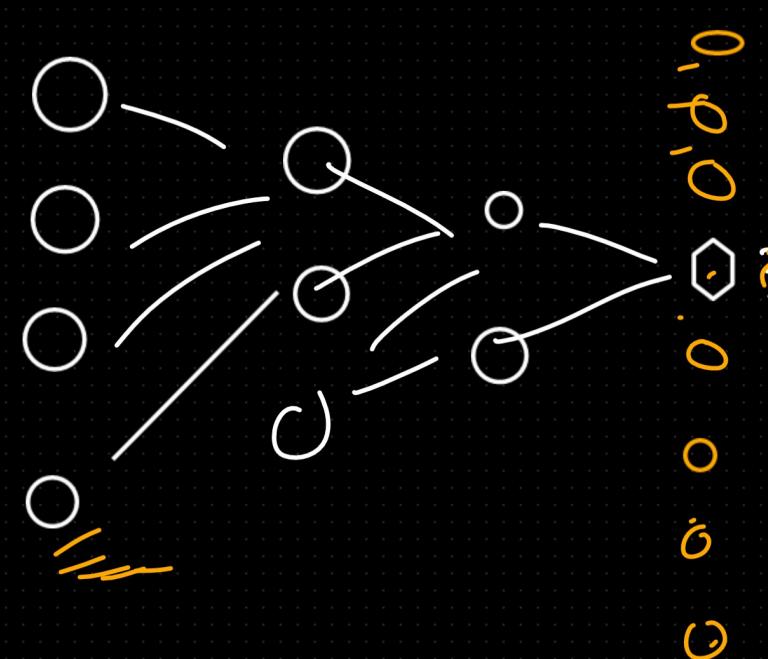
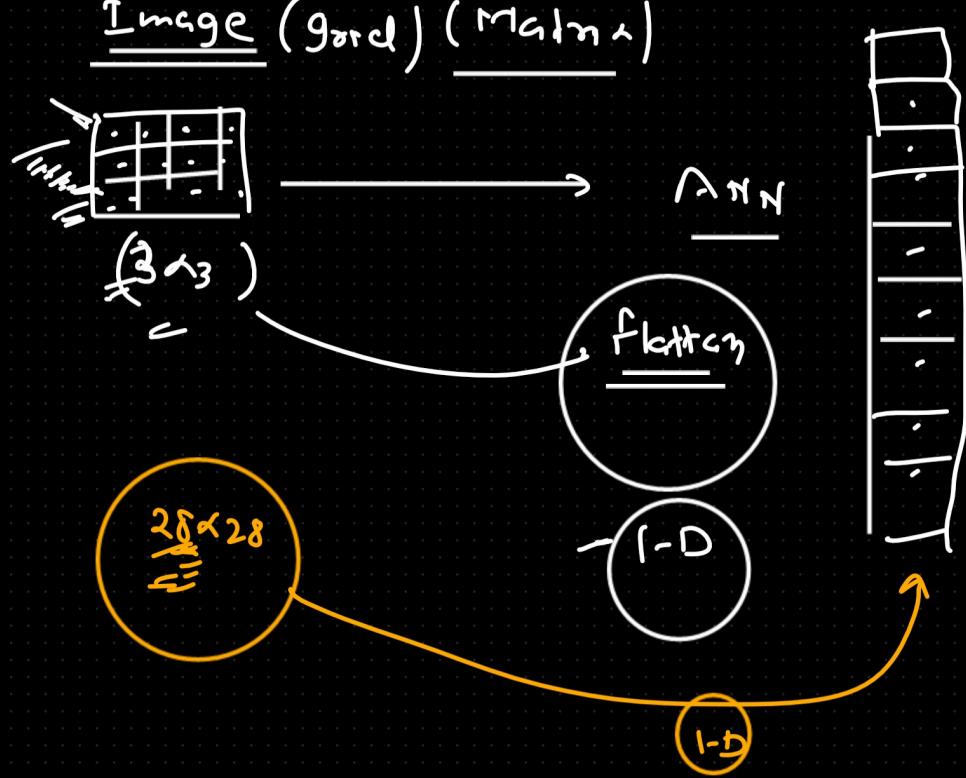


Image (Grid) (Matrix)



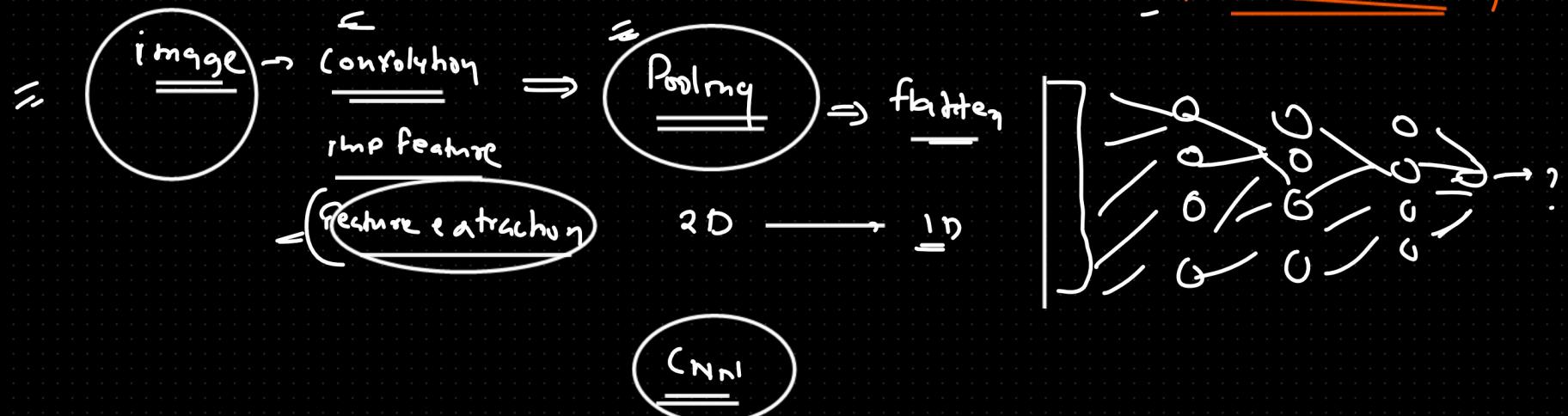


->NEED OF CNN:

This is the reason we are performing the convolution of the images so that dimension of image can be reduced so the computational cost can be reduced and also increasing the classification ability.

CNN is basically used for feature(pixel) extraction reducing the dimension of neural network (specifically at input layer) which leads to less trainable parameters which otherwise would have generated very huge trainable parameter.

On a whole we are trying to reduce the number of trainable parameters for image based dataset which other would have given a huge number of trainable parameters which if directly passed to ANN would have led to computational expensive process

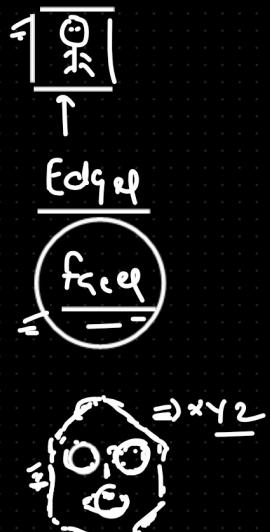
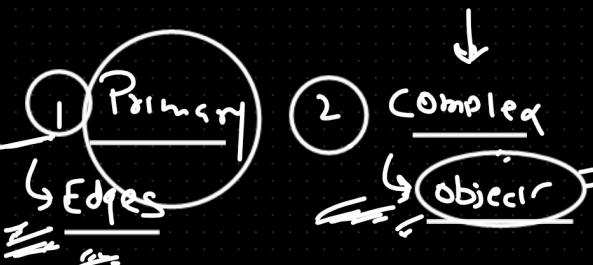
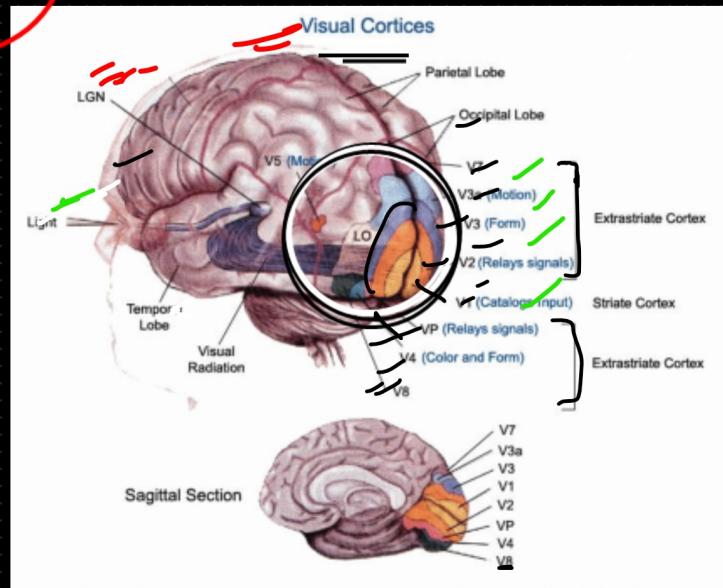


From where CNN is inspired and related history?

1960-1970 Human brain

From where this CNN inspired

Artificial Human vision (Visual cortex)



human

You may read these 2 articles to understand more about the biological inspiration of CNN (No required just read if interested):

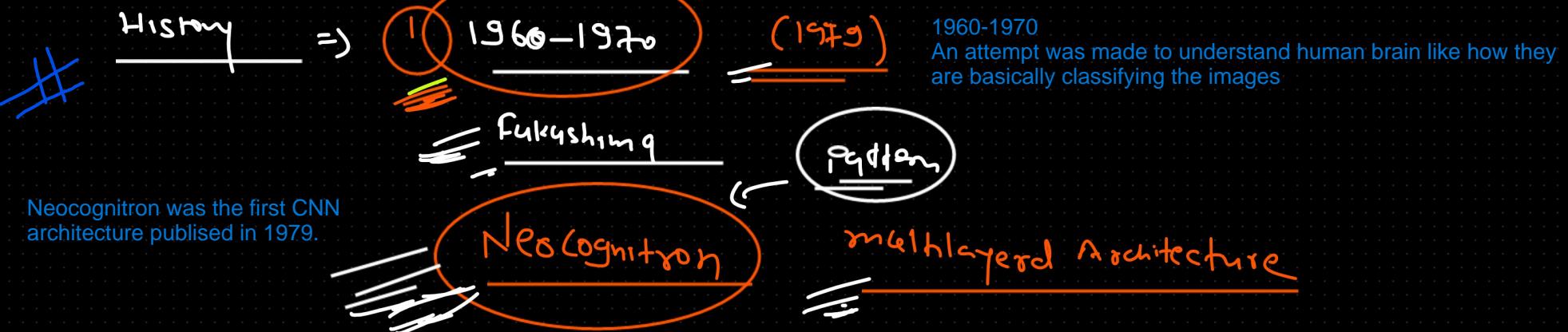
- <https://towardsdatascience.com/convolutional-neural-networks-the-biologically-inspired-model-f2d23a301f71>

- <https://gracewinday.com/2018/05/17/deep-convolutional-neural-networks-as-models-of-the-visual-system-qa/>

1960

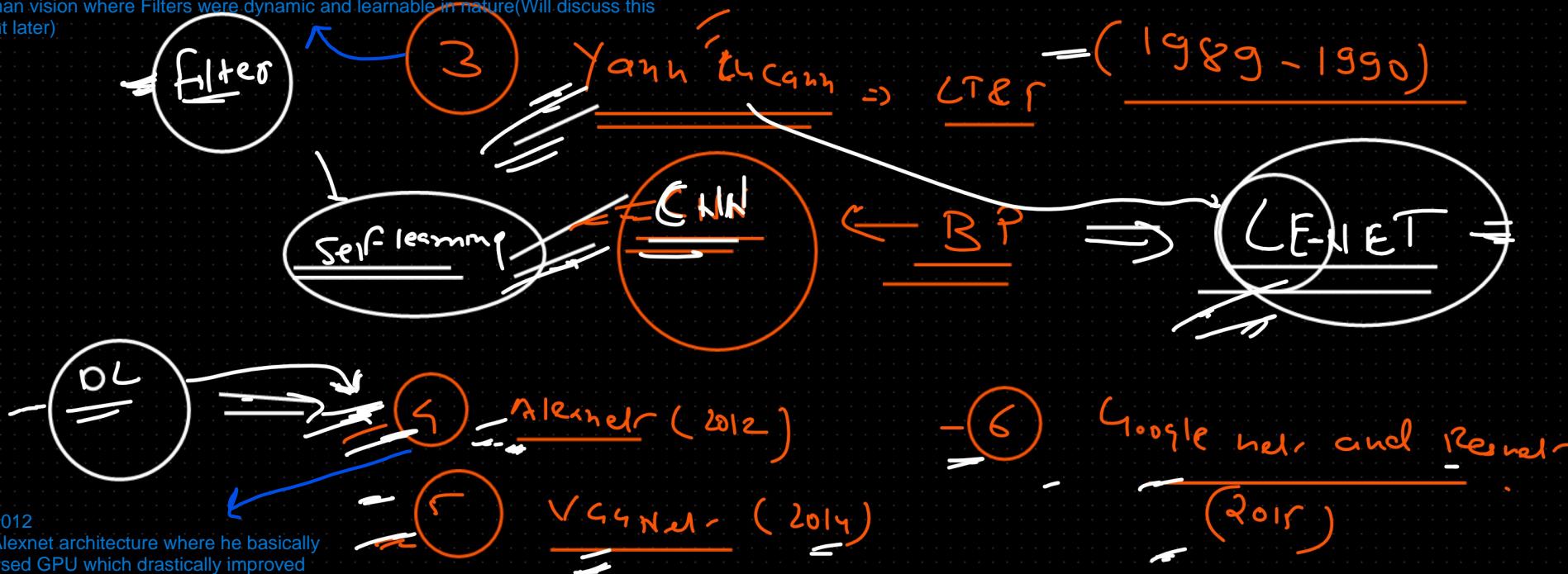
So CNN is inspired by the Human vision called Visual Cortex-->Biological stuff
We have different layers: V1, V2, V3, ..., V7
(Primary: Responsible to detect edges whereas, Complex: Responsible for detecting the objects present within the detected edges).

See every new technology starts with a real life inspiration which later is tried to make a reality by using technology where we are basically trying to mimic the similar architecture of real life inspiration. This real life architecture of inspiration is obtained through extensive research by scientist in that particular domain and then later engineers just try to make use of there research and create a product or technology out of it.



1989-1990

LENET by Yann LeCun who basically made use of Back propagation and earlier studies on human brain to give more reasonable CNN architecture that mimics human vision where Filters were dynamic and learnable in nature(Will discuss this point later)

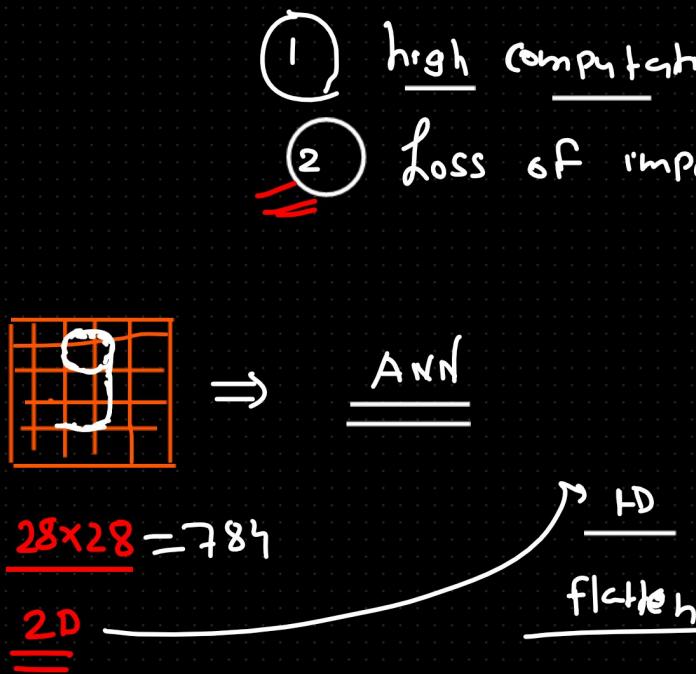


2012

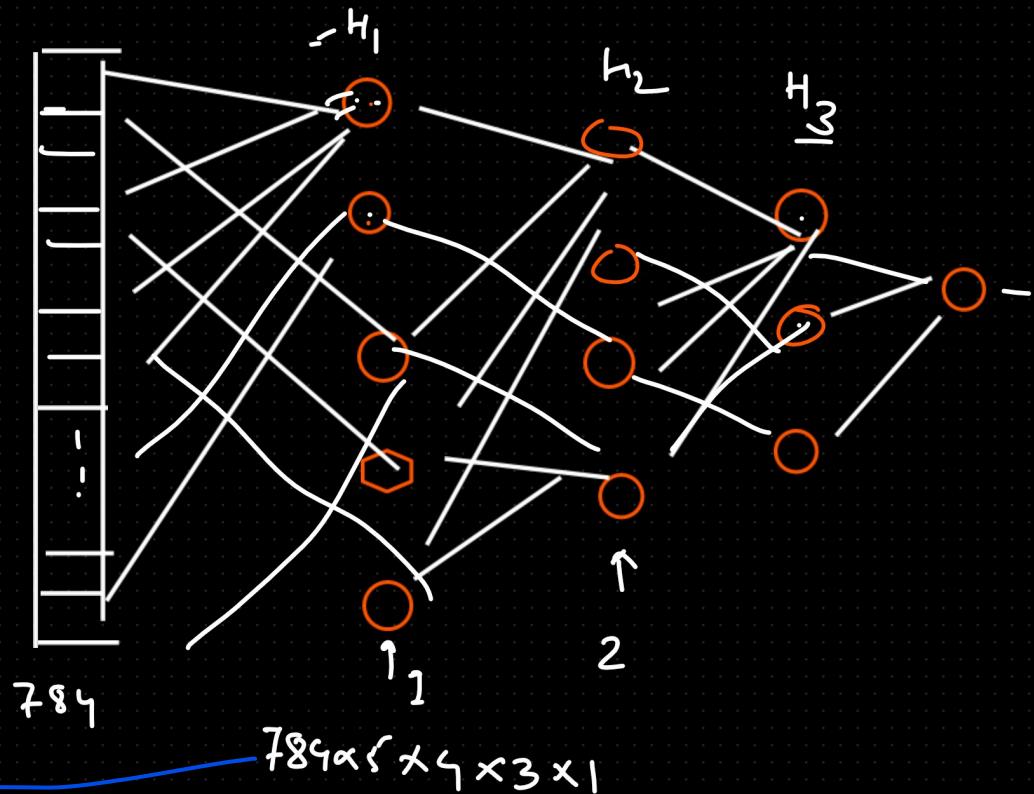
Alexnet architecture where he basically used GPU which drastically improved accuracy

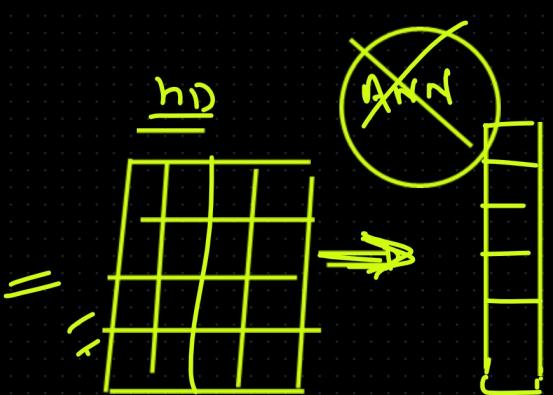
Transfer Learning

Why not ANN instead of CNN? --



Here we are basically calculating the trainable parameters when using ANN alone of image based classification of dimension 28X28. Please note the number of trainable parameters in the same fully connect ANN will be much lesser if images are convoluted before hand using CNN where only important features are passed to the network.





1000×1000

1000000

$$\leq \frac{1}{1000000} \times 100 = 0.0001$$

GDP \Rightarrow 176 Billion \equiv Billion, Trillion

$$\begin{array}{r}
 \overset{5}{\cancel{2}} \\
 787 \cancel{a} 60 \\
 \underline{-\quad 000} \\
 61 \, 0\cancel{3} \times \\
 \hline
 61,0\cancel{5} 0
 \end{array}$$

$$\frac{10 \times 100}{100}$$

franchise

(Precise)

256 \times 256



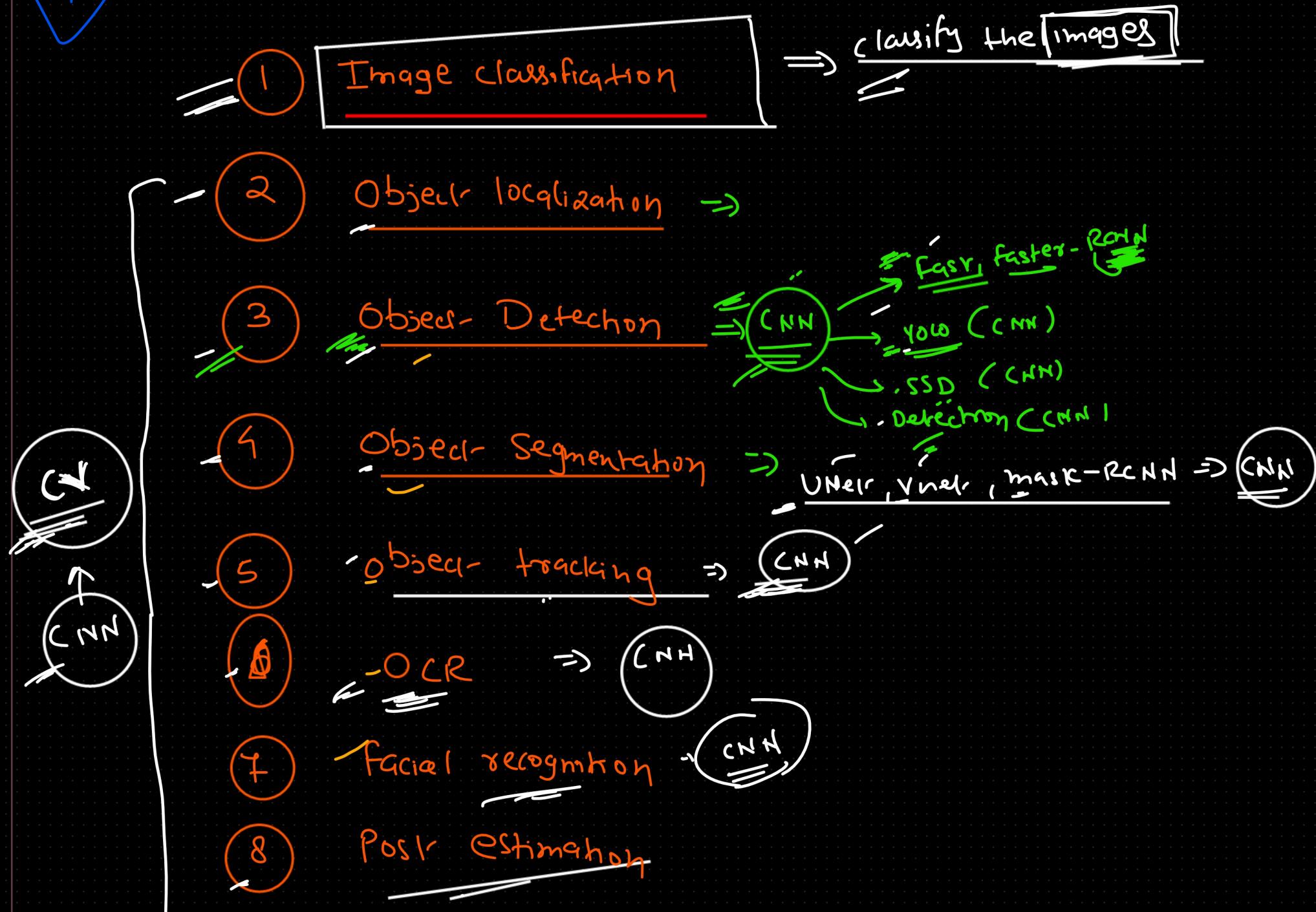
(Con \rightarrow Pooling)

128 \times 128

\downarrow
256 \times 256

↑
Padding

Application of CNN ~~(Image, Videos)~~



L 9

Gray to Color, Super resolution \rightarrow CNN

OpenCV \Rightarrow Python Lib \Rightarrow CV

OpenCV is the pythonic library used for implementing Computer Vision

(high level overview)

High level overview of what needs to be learned inside CNN

Convolution mathematically

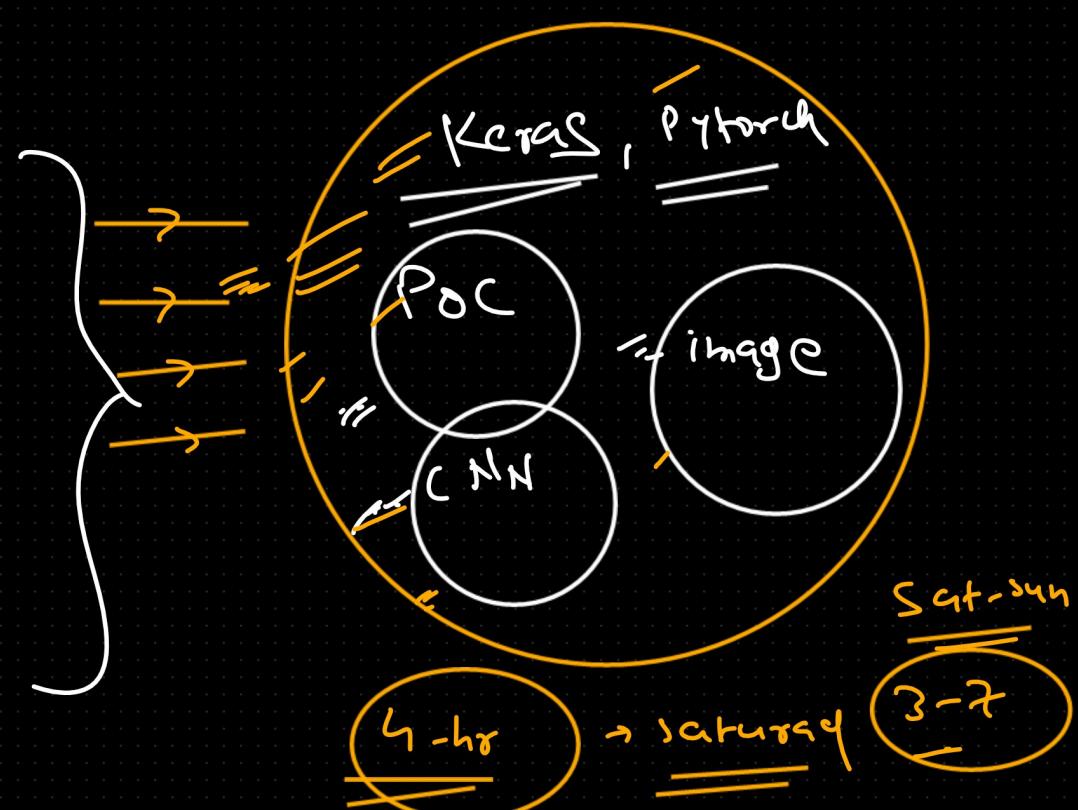
1 Filter, Kernel, feature

2 Convolution

3 Rely \rightarrow use

4 Stride, Padding

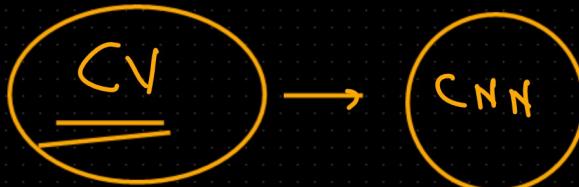
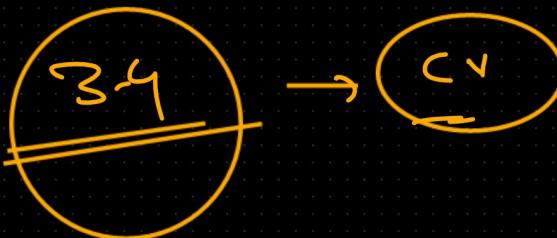
5 Pooling



Data augmentation

Popular CNN Architecture (LeNet, Alexnet, VGG, Resnet, Inception)

(transfer learning) \Rightarrow



This is really amazing. Visualizing how CNN works:

<https://deeplizard.com/resource/pavq7noze2>

<https://poloclub.github.io/cnn-explainer/>