UNIT-III

Convolution Neural Networks (CNN)

Convolution

Finding good representations of images objects and features has been the main goal since the begining of Computer Vision.

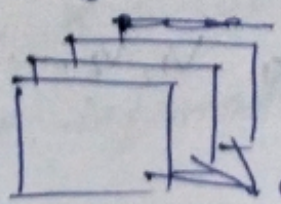
Therefore many tools have been invented to deal with images. Many of these are based on a mathematical operation, called convolution.

* Convolutional Neural Networks finally take the advantages of Neural Networks in general and goes even further to deal with two-dimensional data.
* Thus, The training parameters are elements of two dimensional filters. As a result of applying a filter to an image a feature map is created which contains information about how well the patch corresponds to the related position in the image.

Image Information

* Gray scale image (black & weight) channel-2
* Colored Images (RGB) - channel-3

Colored image



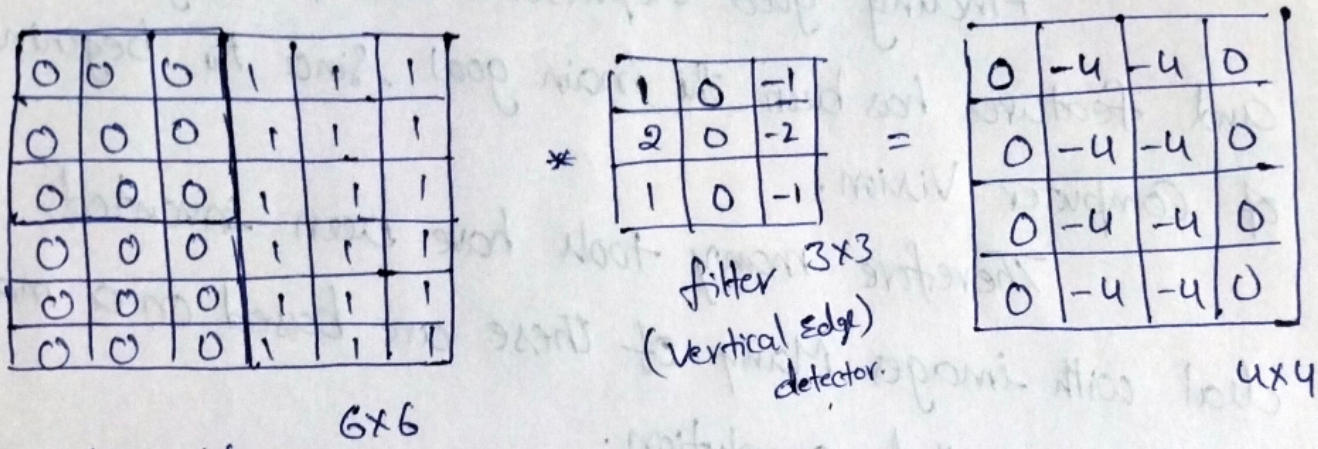
Pixel → Picture element

channel.

* Generall any image pixel range from 0 to 255.

Example

Convolution

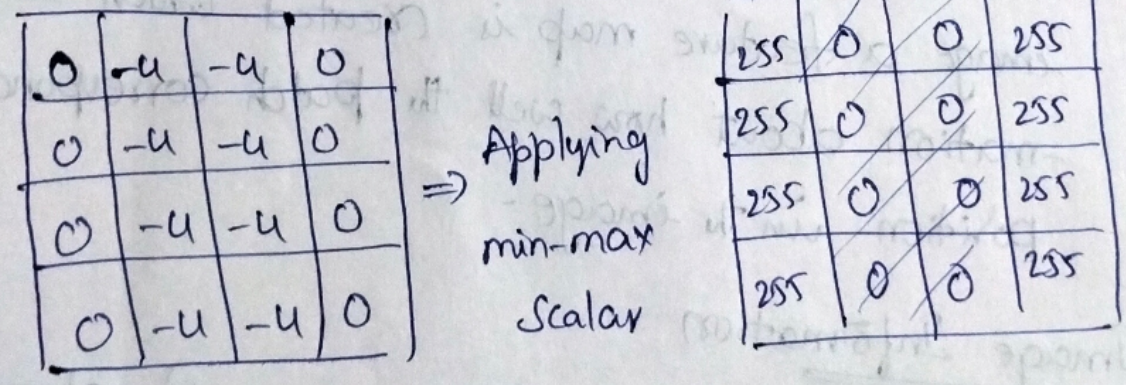


0 - indicate white
1 - indicate black.

Convolution.

* filter are mainly used for finding / representing the edges of image.

* After getting output



minimum = 0 maximum = 255

white represented Edge of image.

* filters / Kernal layers like $V_1, V_2, \ldots V_n$

* Image input = n×n

filter = 3×3

output = 4×4

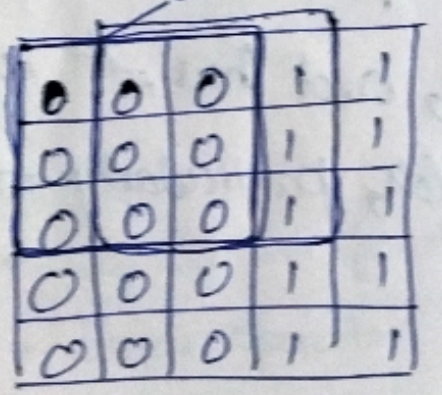
$$n-f+1$$

$6-3+1 \Rightarrow 3+1 = (4)$ → output matrix size.

* Stride

In the context of Convolutional neural networks (CNN), the term "Stride" refers to the number of pixels by which we move the filter across the input image.

Example



*

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

n - input size.

f - filter size.

$$n-f+1 = 6$$

$$n = 6+f-1$$

$$\boxed{n = 8}$$

Increase the input image size with '0' (zeros) (or) other nearest bit. * is called padding. Without padding we lose the image information.

PADDING

Padding in CNN refers to the addition of extra pixels around the borders of the input images or feature map This process removes aggregation bias from the convolution operation. In other words, it makes sure every pixel gets considered.

Types of padding
- Same – adding zeros.
- Valid – adding ones
- Casual – Sequence-to-Sequence

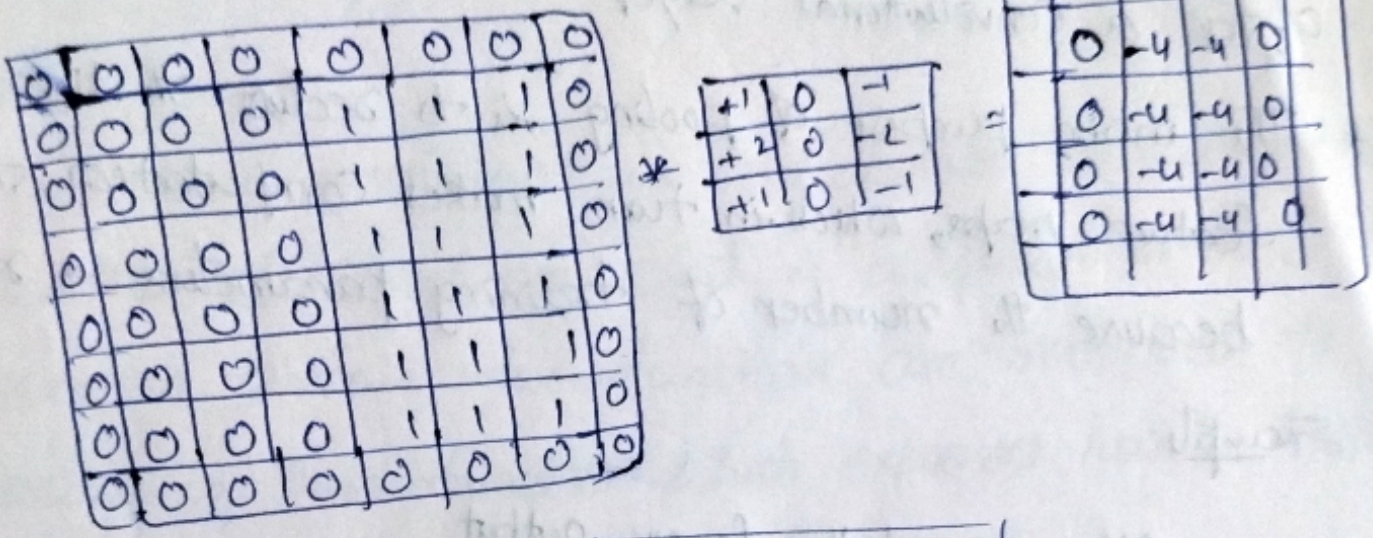
* It adds elements to the input matrix before any convolutional filter is applied, and thus, it aids in preventing any information loss, particularly from the edges of the images.

* In addition, it adds extra elements, and thus the computational cost is increased. Lastly, in some cases, padding has seemed to contribute to overfitting.

$$n - f + 1 = 6$$
$$n = 6 + f - 1$$
$$n = 6 + 3 - 1$$
$$= 9 - 1$$
$$= \boxed{8}$$

$n = 8$

Adding one row at top, one row at bottom, one colaum at right hand side & left hand side.



* $P = 1$

$n + 2P - f + 1$

$6 + 2(1) - 3 + 1$

$8 - 3 + 1 = 6$

Summary

* Convolution operation
* Stride operation
* Padding.
* filter / Kernal

POOLING

MAX - Pooling

It is performed on the convolutional layer of a CNN. It involves sliding a window (often called a filter/kernel) across the input data, similar to the convolution step, but instead of performing a matrix multiplication, max pooling takes the maximum value within the window

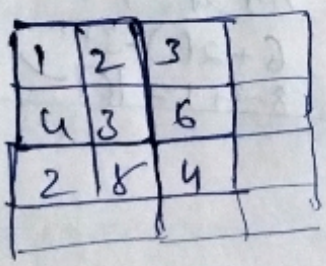
* It is a pooling operation that calculates the maximum value for patches of a feature map, and uses it to create a downsampled (pooled) feature map. It is usually used after a convolutional layer.

* The main purpose of pooling is to reduce the size of feature maps, which in turn makes computation faster because the number of training parameters is reduced

Example

After Convolution layer. output

Maxpooling



Stride = 2

output =

4	6
8	4

Max pooling is mainly used for "location Invariant"

* filter will be updated after back propagation.

CNN Architecture

⇒ CNNs are a class of Deep Neural Networks that can recognize and classify particular features from images and are widely used for analyzing visual images.

⇒ Their applications range from image and video recognition, image classification, medical image analysis, CV & NLP.

⇒ CNN has high accuracy, and because of the same, it is useful in image recognition. Image recognition has a wide range of uses in various industries such as medical image analysis, phone, security, recommendation systems, etc.

⇒ The term "convolution" in CNN denotes the mathematical function of convolution which is a special kind of linear operation wherein two functions are multiplied to produce a third function which expresses how the shape of one function is modified by the other.

⇒ In simple terms, two images which can represented as matrices are multiplied to give an output that is used to extract features from the image.

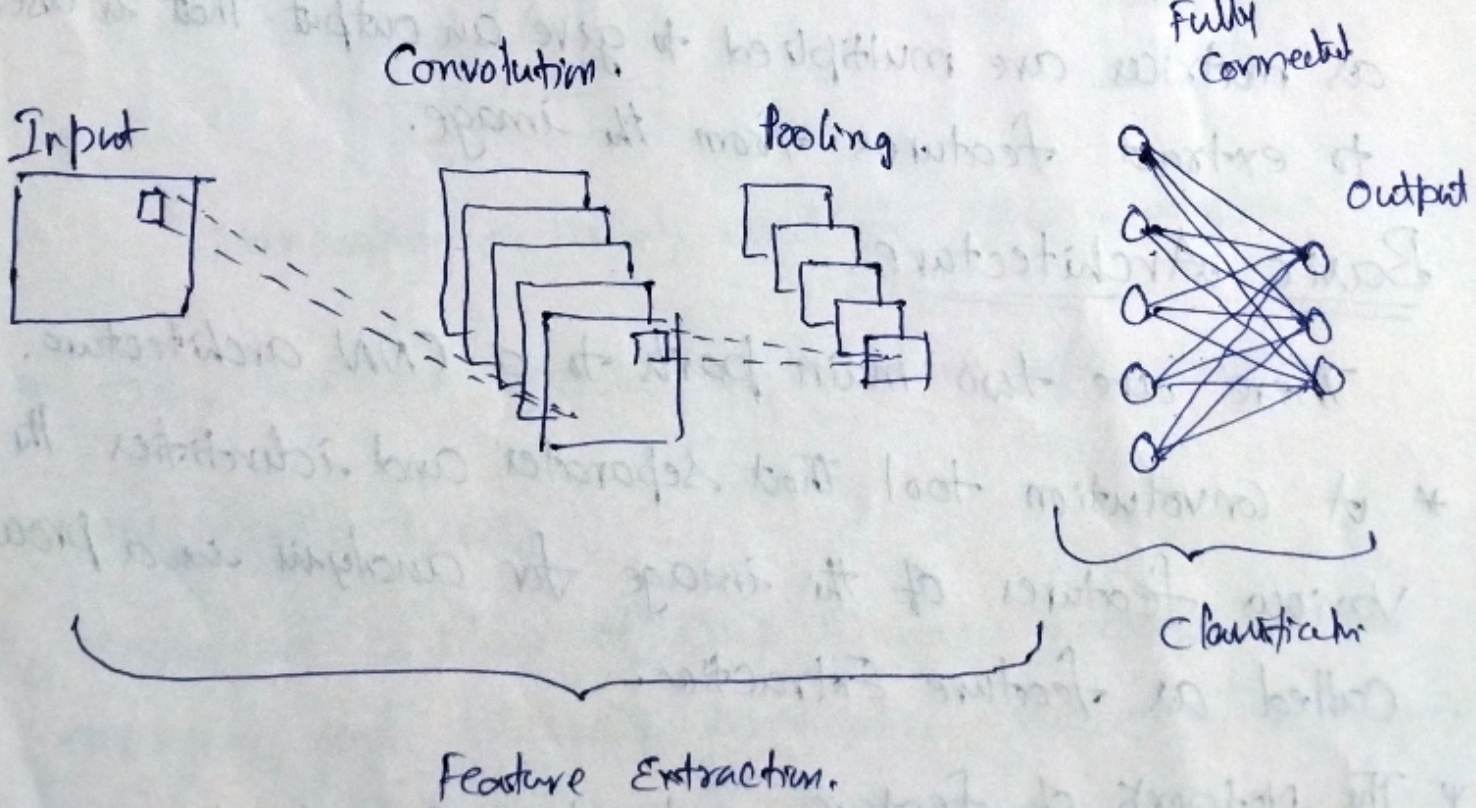
Basic Architecture

There are two main parts to a CNN architecture.

* A convolution tool that separates and identifies the various features of the image for analysis in a process called as feature Extraction.
* The Network of feature extraction consists of many pairs of convolutional or pooling layers.

* A fully connected layer That utilizes the output from the Convolution process and predicts the class of the image based on the feature extracted in previous stages.
* This CNN modul of feature extraction aims to reduce the number of features present in a dataset. It creates new features which summarises the existing features contained in an original set of features.

There are many CNN layers as shown in CNN architecture diagram.



Convolutional Neural Networks

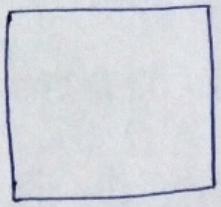
* Convnet — 1989
* LeNet — 1998
* AlexNet — 2012
* GoogleNet —
* Inception (V_2, V_3, V_4) — } 2014
* VGG
* ResNet — 2015
* DenseNet — 2016

LeNet Architecture

The architecture of LeNet-5 was relatively easy, with only five layers of convolution and sub-sampling. Later, two more layers were added to improve its performance.

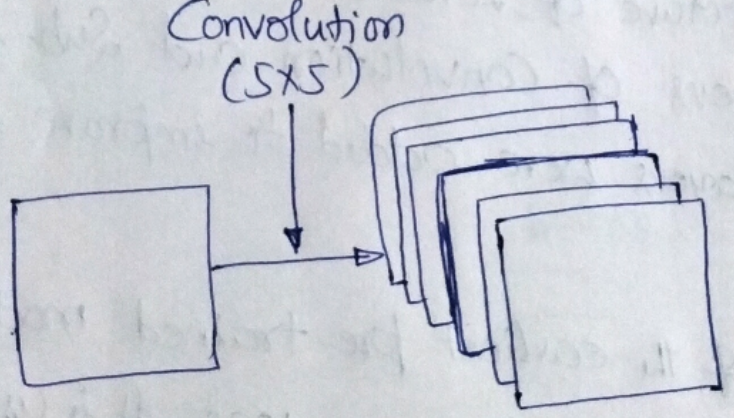
* Lenet-5 is one of the earliest pre-trained models proposed by Yann and others in the year 1998. It is used for recognizing the handwritten and machine-printed characters
* The main reason behind the popularity of this model was its simple and straightforward architecture. It is a multi layer convolution neural network for image classification
* The LeNet-5 Architecture has 5-layers with learnable parameters and hence named Lenet-5.

* It has Three set of convolution layers with a combination of average pooling.
* After the convolution and average pooling layers, we have two fully connected layers. At last, a Softmax classifier which classifies the images into respective class.



Input
32x32x1

The Input to this model is a 32x32 grayscale image hence the number of channels is one.

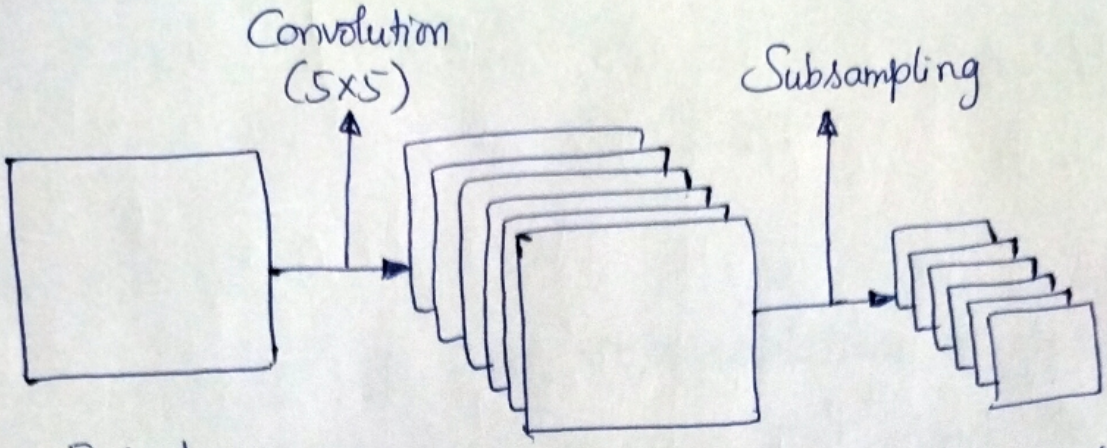


Output shape =
((32-5+1)x(32-5+1)x6)
= (28x28x6)

Input
32x32x1

Feature map
28x28x6

We Then apply the first Convolution operation with the filter size 5x5 and we have 6 such filters. As a result, we get a feature map of size 28x28x6. Here the number of channels is equal to the number of channels is equal to the number of filters applied.



like That final architecture of the Lenet-5 model.