Convolutional Neural Network

Computer Vision

Contents

Introduction to Computer Vision

- Computer vision is concerned with the automatic extraction, analysis and understanding of useful information from a single image or a sequence of images.

  - The British Machine Vision Association and Society for Pattern Recognition (BMVA)

(or)

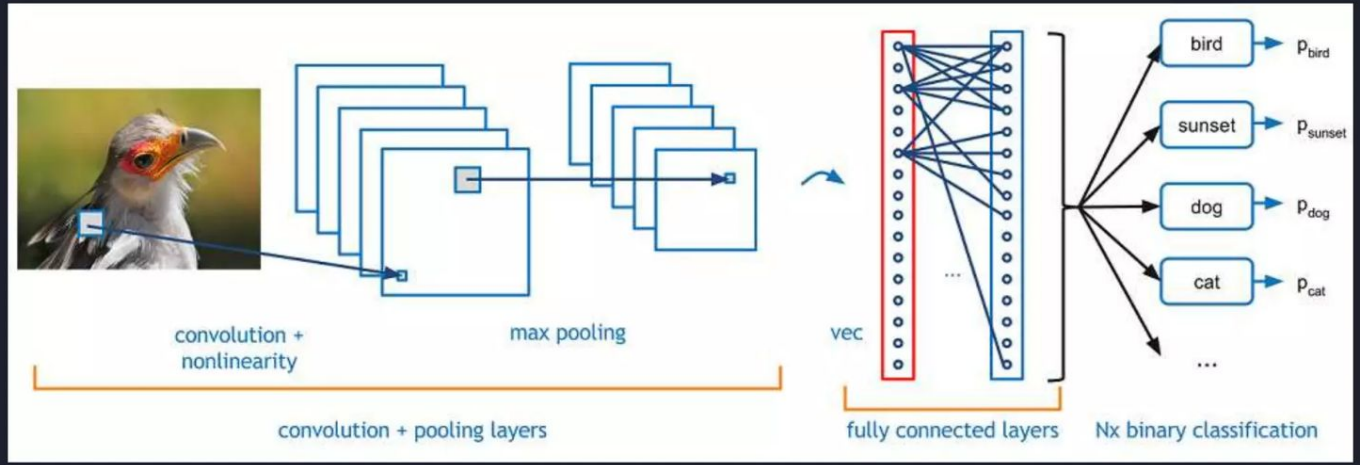
- It is an interdisciplinary field that deals with how computers can be made to gain high-level understanding from digital images or videos.

  - Wikipedia

What is CNN(Convolution Neural Network)

- It is a class of deep learning.
- Convolutional neural network (ConvNet's or CNNs) is one of the main categories to do images recognition, images classifications, objects detections, recognition faces etc.,
- It is similar to the basic neural network. CNN also have learnable parameter like neural network i.e., weights, biases etc.
- CNN is heavily used in computer vision
- There 3 basic components to define CNN
  - The Convolution Layer
  - The Pooling Layer
  - The Output Layer (or) Fully Connected Layer

Architecture of CNN



Convolution Layer

- Computers read images as pixels and it is expressed as matrix (NxNx3)— (height by width by depth).
- The Convolutional 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 (MxMx3), with a smaller dimension but the same depth as the input file.
- This filter is convolved (slided) across the width and height of the input file, and a dot product is computed to give an activation map.

Convolution Layer

INPUT IMAGE

18	54	51	239	244	188
55	121	75	78	95	88
35	24	204	113	109	221
3	154	104	235	25	130
15	253	225	159	78	233
68	85	180	214	245	0

WEIGHT

1	0	1
0	1	0
1	0	1

429

INPUT IMAGE

18	54	51	239	244	188
55	121	75	78	95	88
35	24	204	113	109	221
3	154	104	235	25	130
15	253	225	159	78	233
68	85	180	214	245	0

WEIGHT

1	0	1
0	1	0
1	0	1

429

Images source: Analytics Vidhya

Convolution Layer

Original Image

Convoluted image

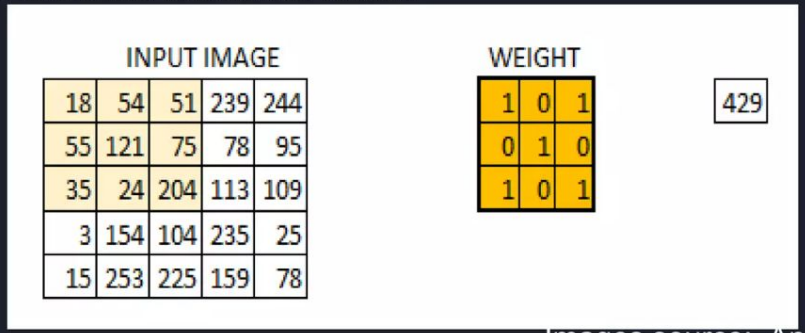
Images source: Analytics Vidhya

Convolution Layer

The concept of stride and padding:

- The weight of a matrix moves 1 pixel at a time is called as stride 1 (as we did in above case).

What if we increase the stride value?



Images source: Analytics Vidhya

Convolution Layer

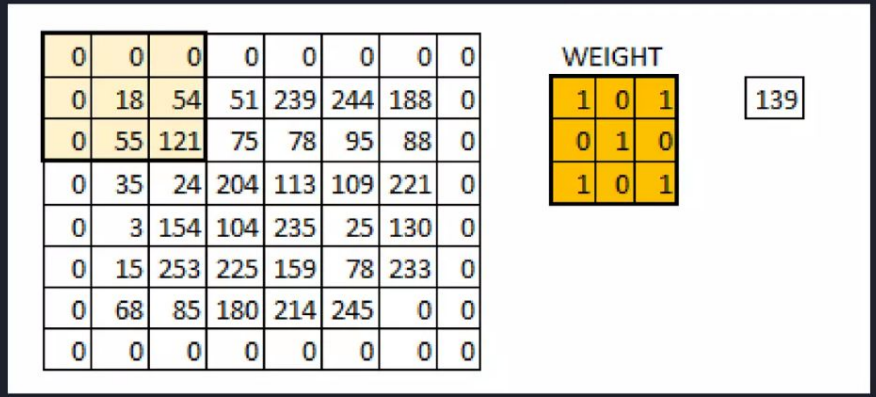
- As we can see in above image the increase in the stride value decreases the size of the image (which may cause in losing the features of the image).
- Padding the input image across it solves our problem, we add more than one layer of zeros around the image in case of higher stride values.

0	0	0	0	0	0	0	0
0	18	54	51	239	244	188	0
0	55	121	75	78	95	88	0
0	35	24	204	113	109	221	0
0	3	154	104	235	25	130	0
0	15	253	225	159	78	233	0
0	68	85	180	214	245	0	0
0	0	0	0	0	0	0	0

Images source: Analytics Vidhya

Convolution Layer

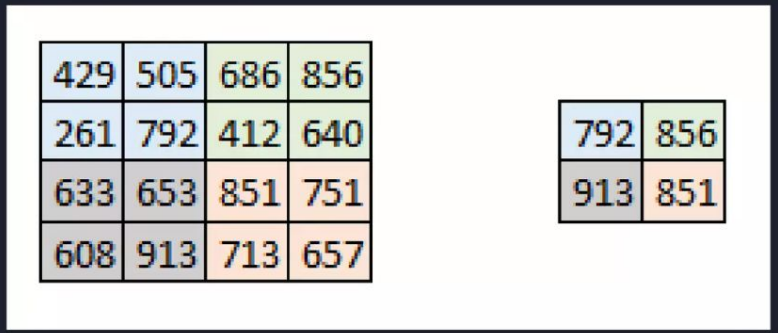
- when the input of 6x6 is padded around with zeros we get the output with same dimensions of 6x6 this is known as '**Same Padding'.**
- The middle 4x4 pixel remains the same, here we have retained the more information from borders and also preserved the size of image.



Images source: Analytics Vidhya

The Pooling Layer

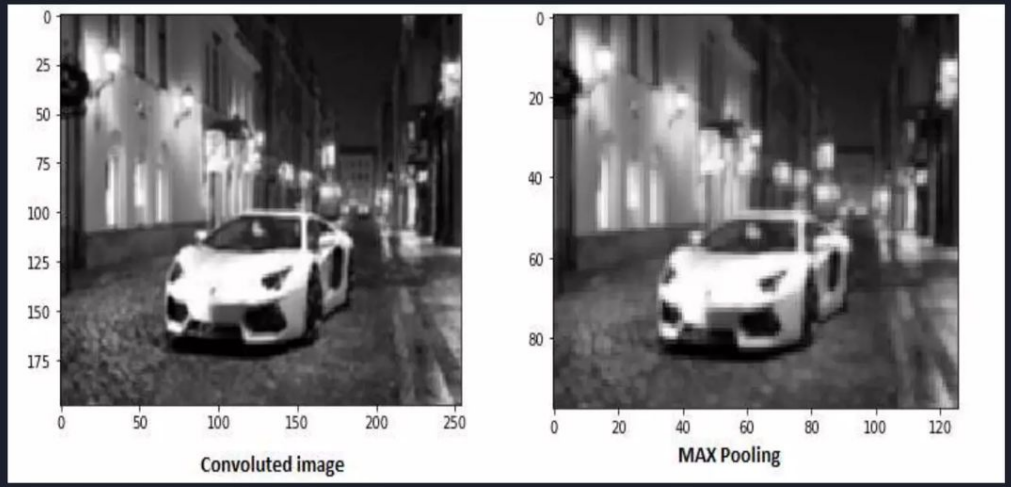
- It can be seen in between the convolution layers in CNN architecture.
- This layer basically reduces the amount of parameters and computation in the network.
- Pooling is done for the sole purpose of reducing the spatial size of the image.
- Pooling is done independently on each depth dimension, therefore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



Images source: Analytics Vidhya

The Pooling Layer

- Let's try to understand from below image how max pooling has affected our image below.



Images source: Analytics Vidhya

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.

CNN

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

- Deeper 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.

Types 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

Applications of Computer Vision

- Computer vision, an AI technology that allows computers to understand and label images, is now used in convenience stores, driverless car testing, daily medical diagnostics, and in monitoring the health of crops and livestock.
- Different use cases found in the computer vision as follows
    - Retail and Retail Security
    - Automotive
    - Healthcare
    - Banking
    - Agriculture
    - Industrial

END