Convolutional Neural Network (CNN)

Outline

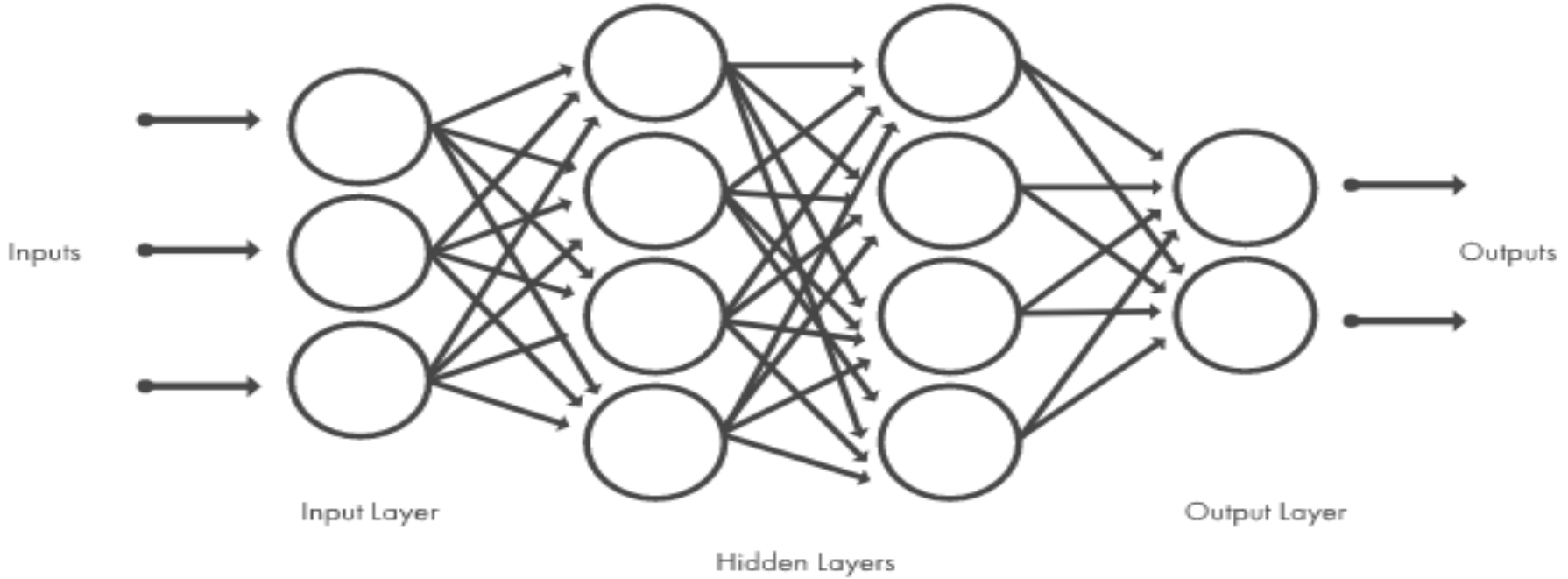
- Introduction to Deep Learning
- Introduction to CNN
- How Deep Learning Works
- What are neural network?
- What is a Neuron?
- Introduction to CNN
- How to represent the images to computers?
- CNN Layers.
- CNN Applications
- Different Architectures of CNN
- Benefits of Using CNN
- Challenges of CNN
- References

Introduction to Deep Learning

- Deep learning is a machine learning technique that teaches computers to do what comes naturally to humans: learn by example. Deep learning is a key technology behind driverless cars, enabling them to recognize a stop sign, or to distinguish a pedestrian from a lamppost. It is the key to voice control in consumer devices like phones, tablets, TVs, and hands-free speakers. Deep learning is getting lots of attention lately and for good reason.
- In deep learning, a computer model learns to perform classification tasks directly from images, text, or sound.

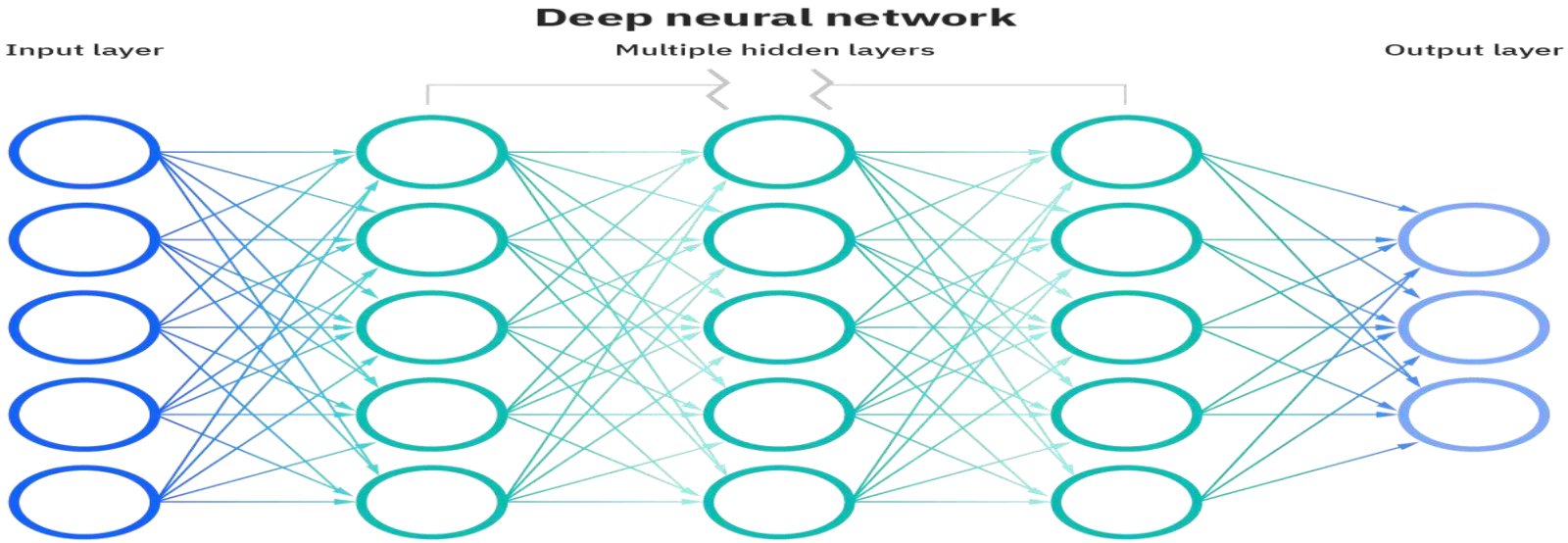
How Deep Learning Works…

- Most deep learning methods use neural network architectures, that is why deep learning models are often referred to as deep neural networks.
- One of the most popular types of deep neural networks is known as convolutional neural network **(CNN** or **ConvNet)**.



What are neural networks?

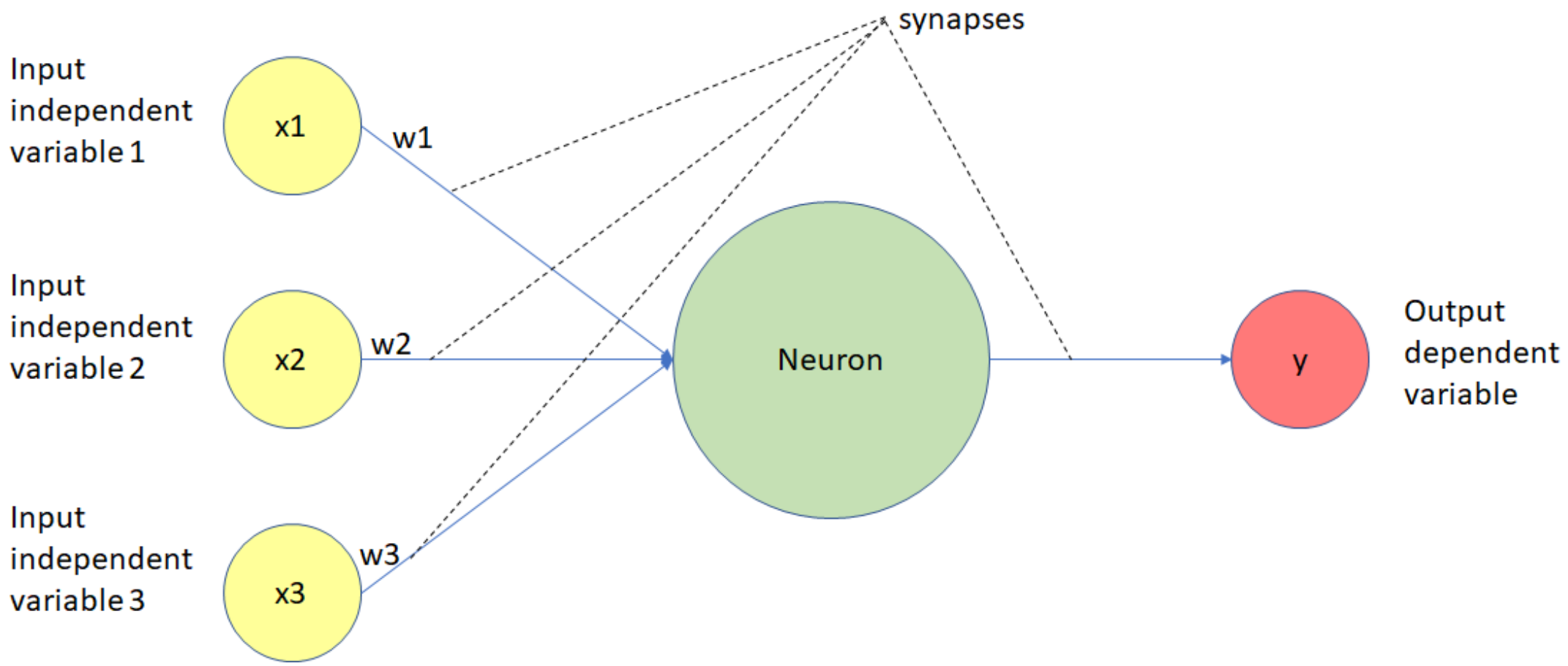
Neural networks, also known as artificial neural networks (ANNs) or simulated neural networks (SNNs), are a subset of machine learning and are at the heart of deep learning algorithms. Their name and structure are inspired by the human brain, mimicking the way that biological neurons signal to one another.



What is a Neuron?

- Neurons in deep learning were inspired by neurons in the human brain.
- Neurons in deep learning models are nodes through which data and computations flow.
- Neurons work like this:
- They receive one or more input signals. These input signals can come from either the raw data set or from neurons positioned at a previous layer of the neural net.
- They perform some calculations.
- They send some output signals to neurons deeper in the neural net through a synapse.

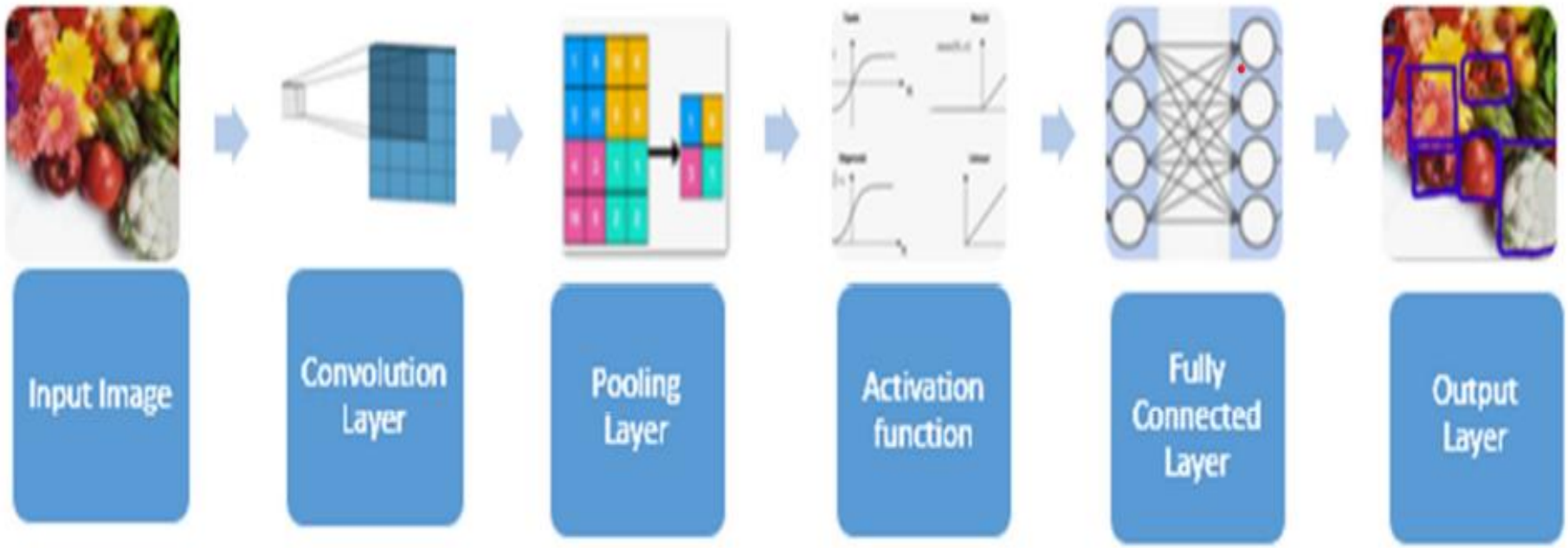
Functionality of a neuron in a deep learning neural network



Introduction to CNN

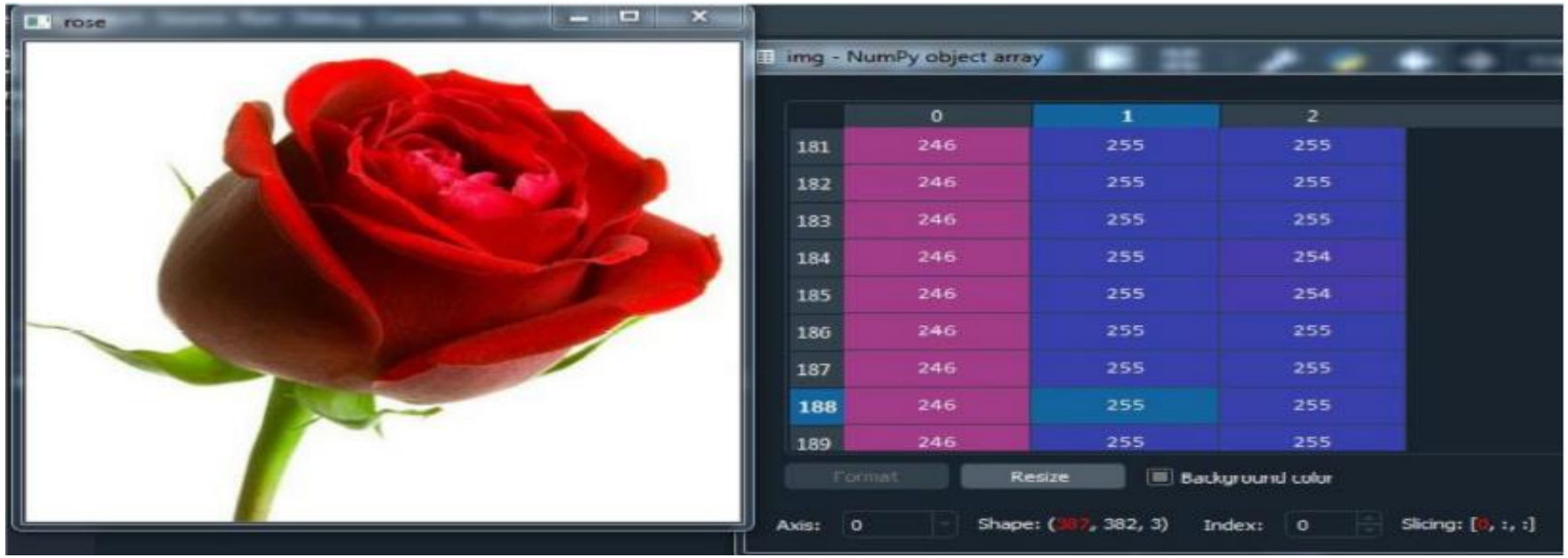
- A Convolutional Neural Network (ConvNet/CNN) is a Deep Learning algorithm.
- CNN is a multilayer neural network.
- Convolutional Neural network is comprised of one or more convolutional Layers(often with a pooling layer) and then followed by a fully connected layer.
- It can take an input image, assign importance (learnable weights and biases) to various aspects/objects in the image and be able to differentiate one from the other.
- The architecture of a ConvNet is analogous to that of the connectivity pattern of Neurons in the Human Brain and was inspired by the organization of the Visual Cortex.

CNN Layers


Input Image
Convolution Layer
Pooling Layer
Activation function
Fully Connected Layer
Output Layer

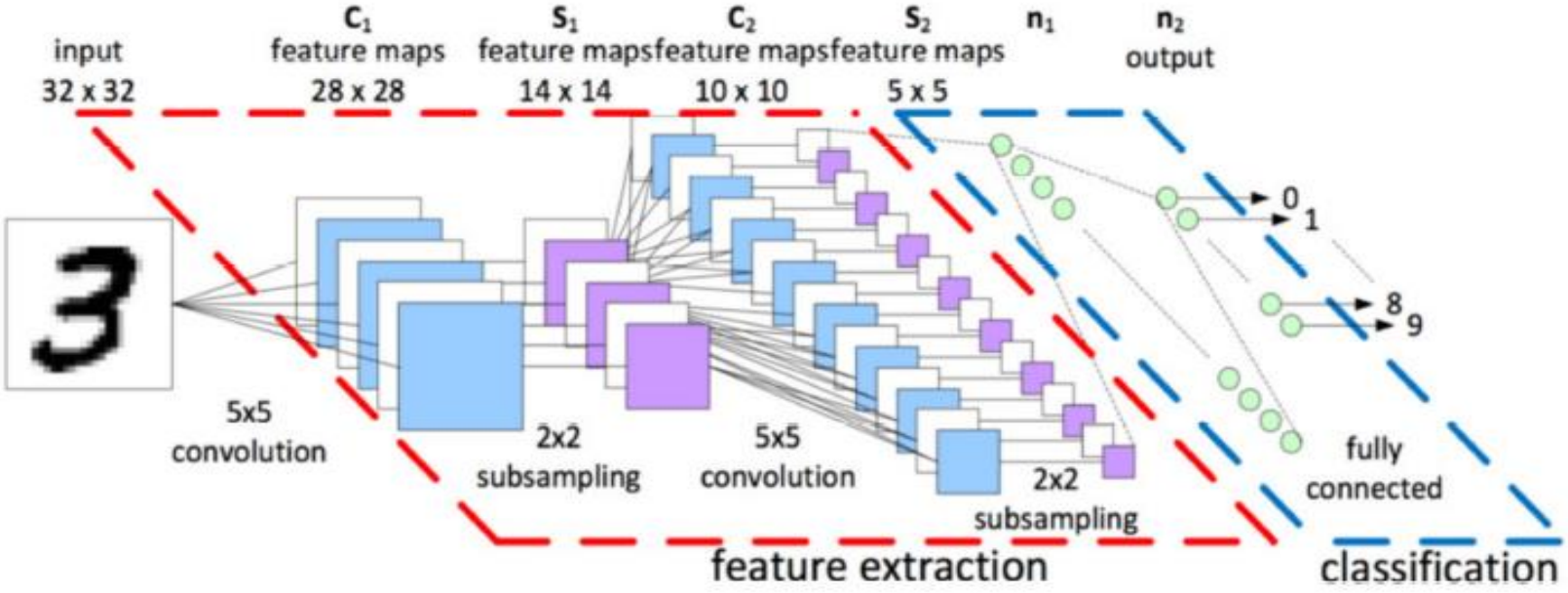
How to represent the images to computers?

- Humans can identify the image by seeing it, but computers can only understands numerical data. That's why to examine and identify the images by a computer program is an intricate task.
- Computer image is made up of pixels. Pixels are the binary representation of visual data
- For digital image pixels are arranged in a matrix format where pixel values range from 0-255 .The pixel value of an image specifies brightness and hue of each pixel.



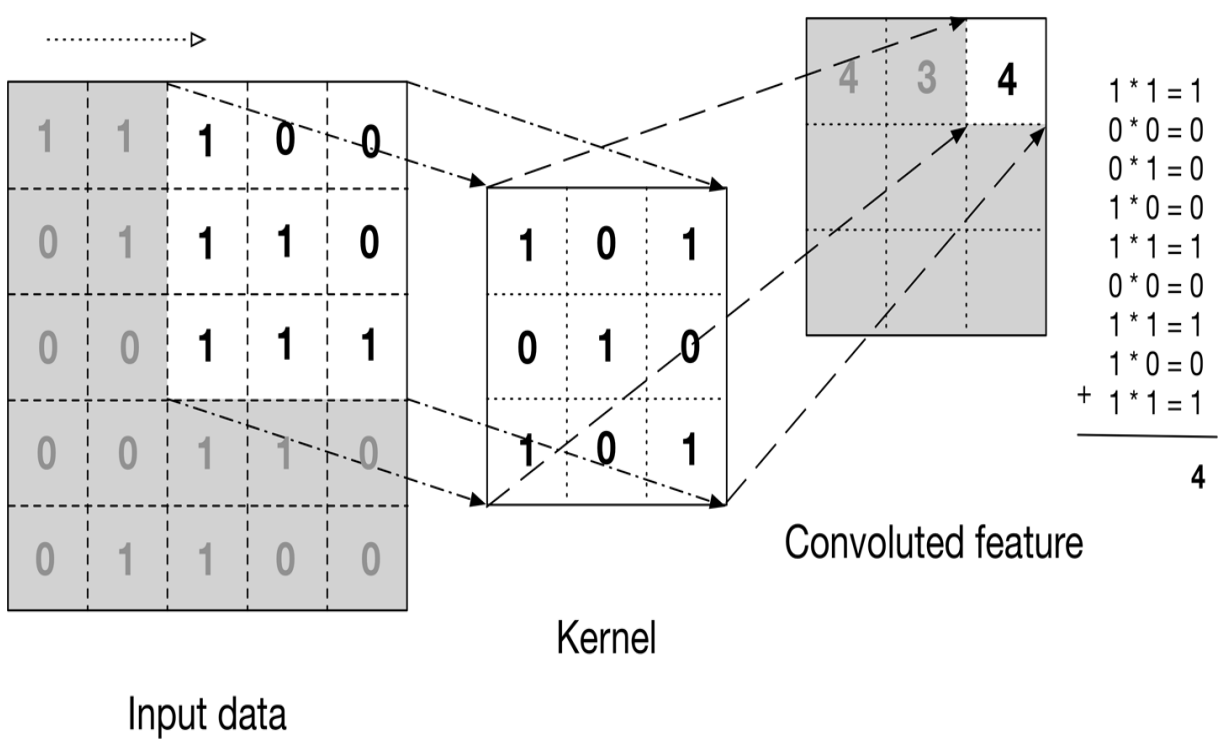
Convolution Layer

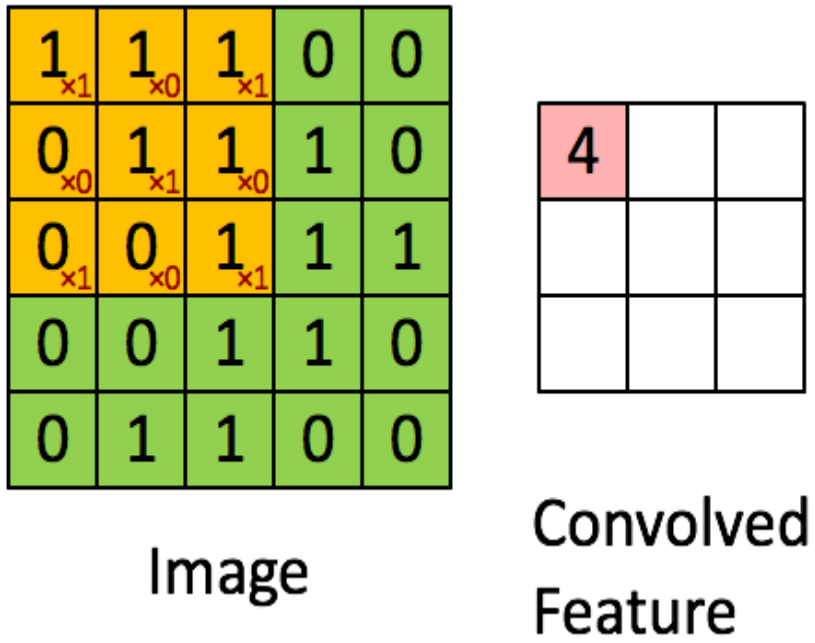
- The most important layer in CNN architecture is convolution layer .
- This layer acts as a feature extractor that extracts the features of the input image.
- It receives an input in the form of image and then apply 5*5 or 3*3 filter .



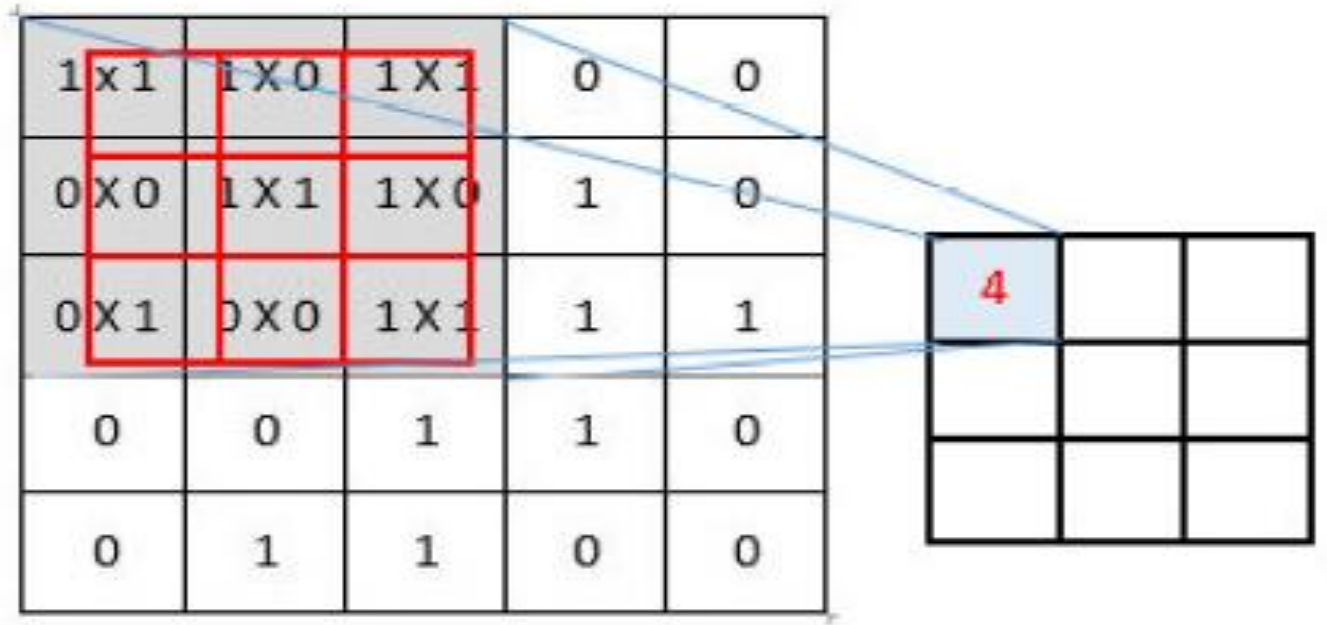
Convolution layer Cntd---

- The below image shows what a convolution is. We take a filter/kernel(3×3 matrix) and apply it to the input image to get the convolved feature.



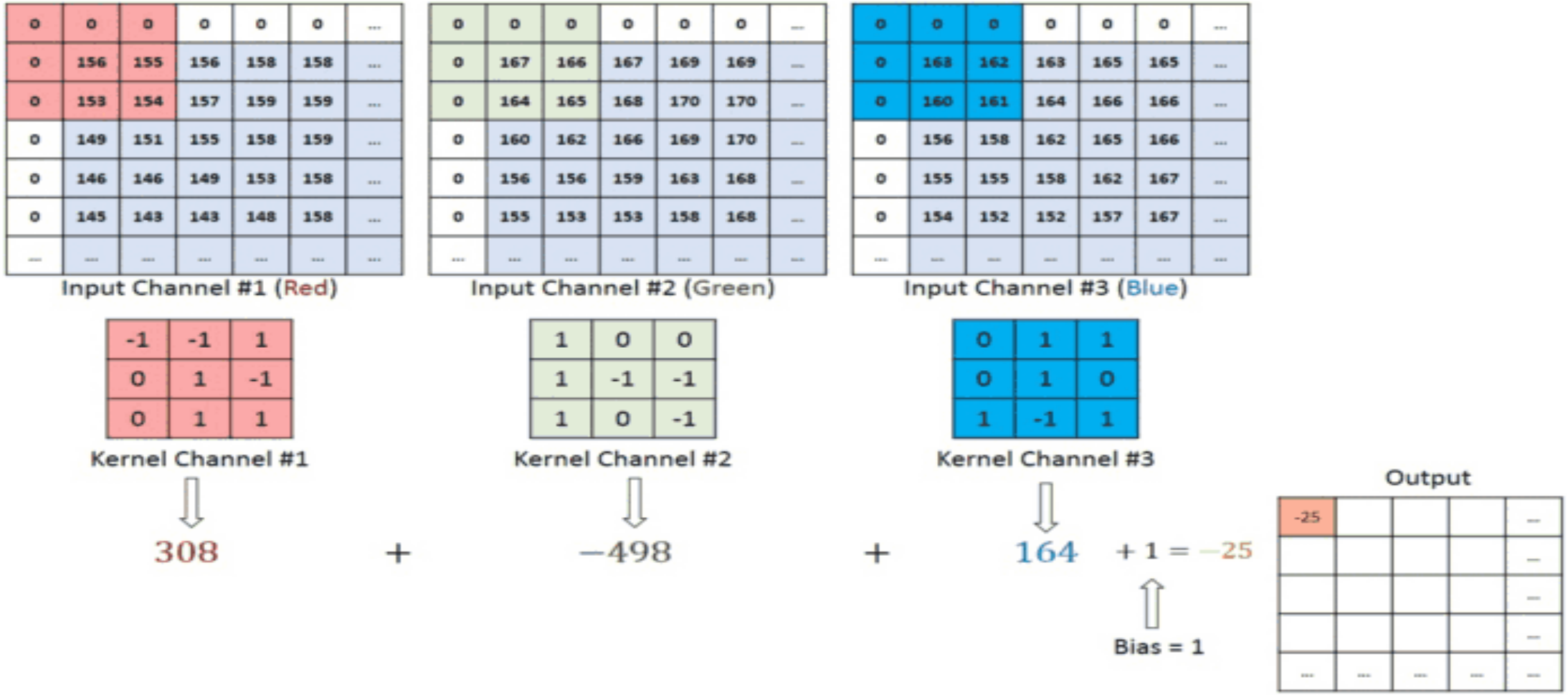


First Step of Convolution



$1 \times 1 + 1 \times 0 + 1 \times 1 + 0 \times 0 + 1 \times 1 + 1 \times 0 + 0 \times 1 + 0 \times 0 + 1 \times 1$

Convolution operation on a MxNx3 image matrix with a 3x3x3 Kernel



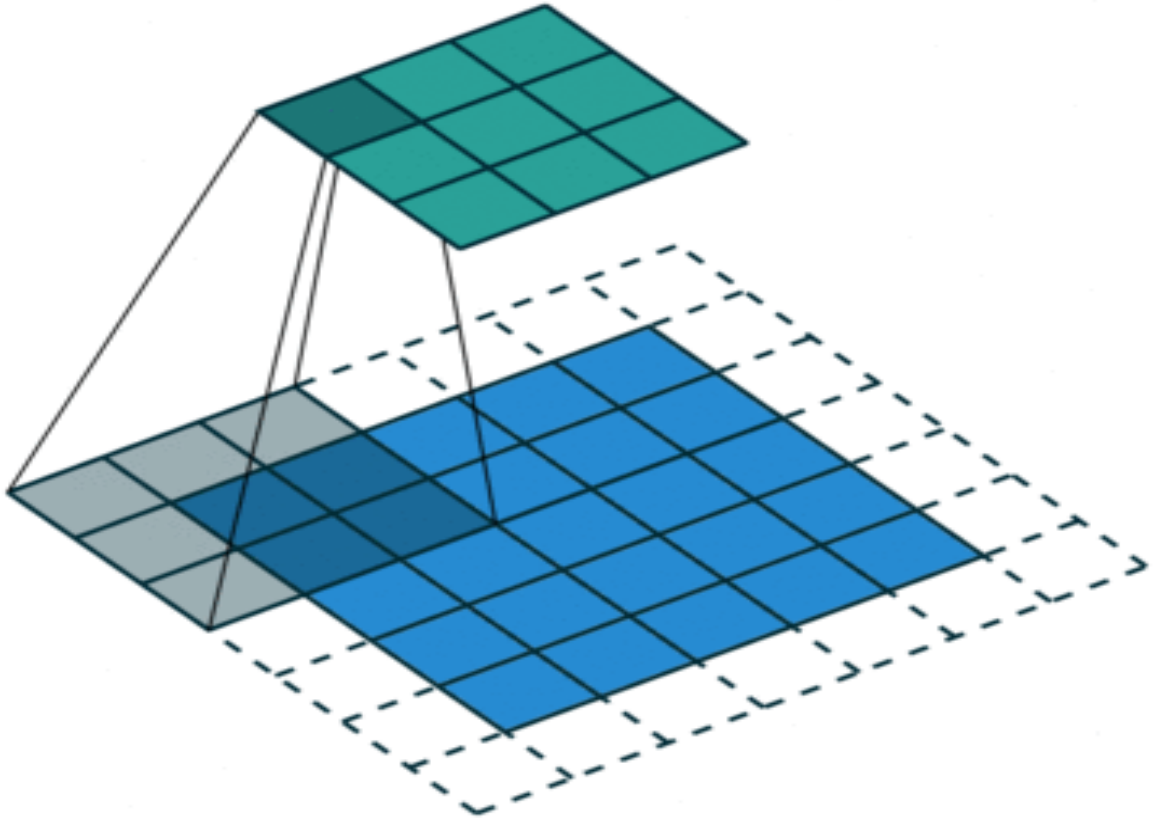
Feature Extraction

- CNN is acknowledged for its capability of automatically extracting features.
- Padding is commonly done in CNN for preventing the undesirable shrinking of size of feature map at each layer.
- The result of this operation is of two types:

✓ "A type in which the dimensionality of the convoluted feature is reduced in comparison to the input"[5]

✓ "A type in which the dimensionality is not reduced but is either enhanced or maintained and padding is used to satisfy this task"[5].

Convolution Operation with Stride Length = 2



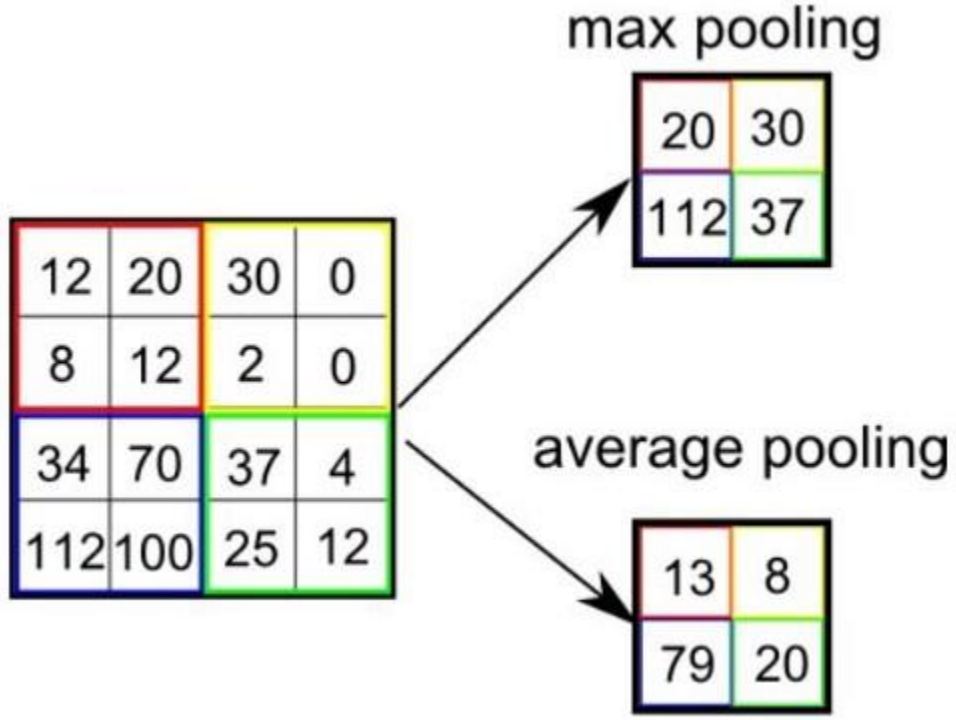
Pooling Layer

- After obtaining the feature maps, it is necessary to add a pooling (sub-sampling)layer in CNN, next to a convolution layer.
- The job of the pooling layer is to shrink the convolved feature's spatial size. As a result of the dimensionality reduction, the computer power required to process the data is reduced.
- This also aids in the extraction of leading characteristics that are positional and rotational invariant, which preserves the model's practical training.
- Pooling reduces the training time while also preventing over-fitting.

Types of Pooling

There are two forms of pooling:

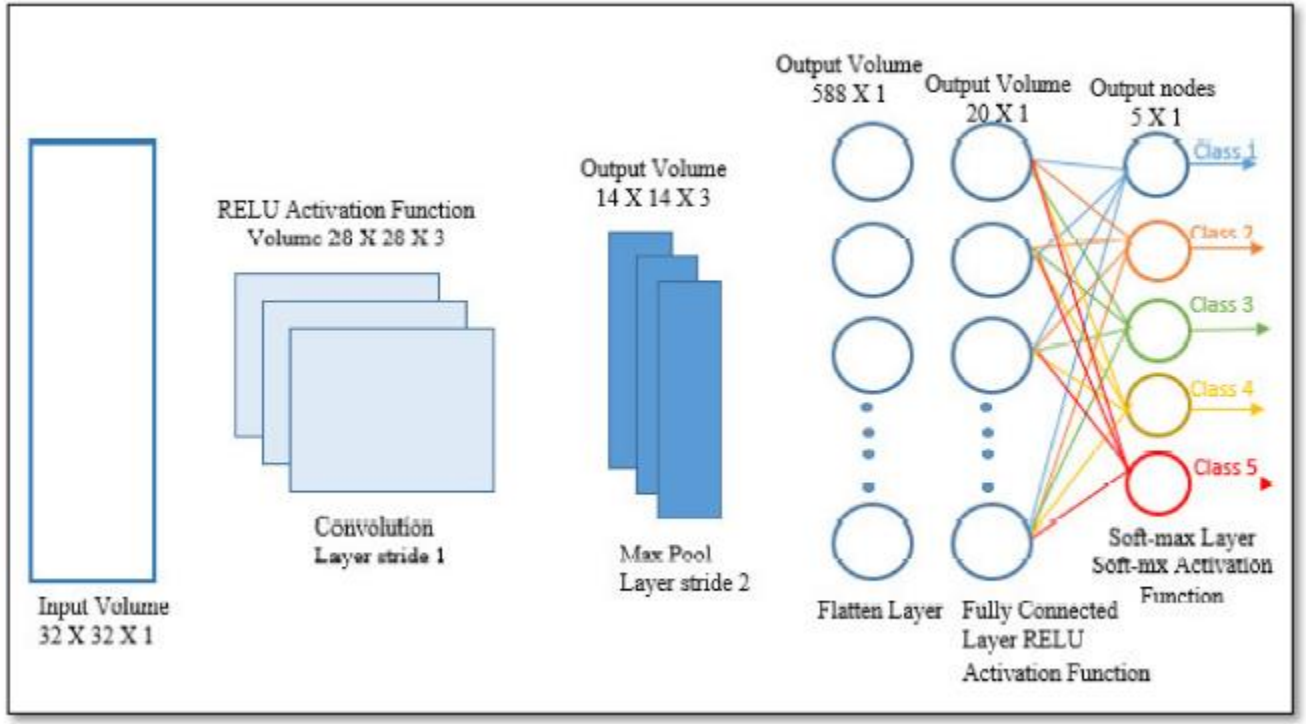
- Maximum pooling
- Average pooling



Nonlinearity Layer (Activation Function)

- A fully connected layer is nothing more than a feed-forward neural network .
- Fully connected layers are found at the network's very bottom layers.
- A fully connected layer receives input from the final pooling or convolutional layer's output layer, which is flattened before being delivered as input.
- Flattening the output entails unrolling all values from the output that were obtained after the last pooling or convolutional layer into a vector (3D matrix).

Fully Connected Layer



CNN Applications

- Object Detection and Segmentation
- Object Counting
- Computer Vision
- Image Classification
- Recognition of Speech
- Video Processing
- Images with Low Resolution
- Natural Language Processing

Different Architectures of CNN

- AlexNet
- High-resolution network (HRNet)
- CapsuleNet
- ResNet
- DenseNet
- GoogLeNet
- Visual geometry group (VGG)
- ZefNet

Benefits of Using CNNs

In computer Vision environment the advantage of employing CNN rather than other neural network is :

- The main purpose to use CNN is the weight distribution property. This decreases the amount of trainable network parameters also improves network generalization and avoids overfitting in turn benefits the network to improve
- The output is both extremely reliant and structured for extracted features because of simultaneously learning of classification and feature extraction layer.
- CNN is easy to use for the implementation of large-scale network rather than other neural networks.[6]

Challenges of CNN

- **Training data**: To attain good performance by the model it requires huge amount of data [8].
- **Under specification**: Some models display unexpectedly lowly performance when tested on some real-world applications like medical genomics, computer vision, natural language processing, and medical imaging. Under specification is the main reason of low performance. By making few changes to a model can lead to entirely different output and entirely dissimilar predictions [6].
- **Imbalanced data**: Usually, living data incline to be imbalanced, as negative samples are more in numbers than positive ones [8]. If the model is trained with the imbalanced data, then there are high chances of getting the undesirable results.