

Qubitrics Deep Learning Assignment

INTRODUCTION:

*Nowadays Artificial Intelligence is the most advanced technology across the globe. **Artificial intelligence** refers to the simulation of human **intelligence** in machines that are programmed to think like humans and mimic their actions. It works on the principle of learning and training itself from past records. This technology is helpful when we have some hand-written document and we want to store them in our computers without manually typing them in computer. It is also used in google translate and google lens.*

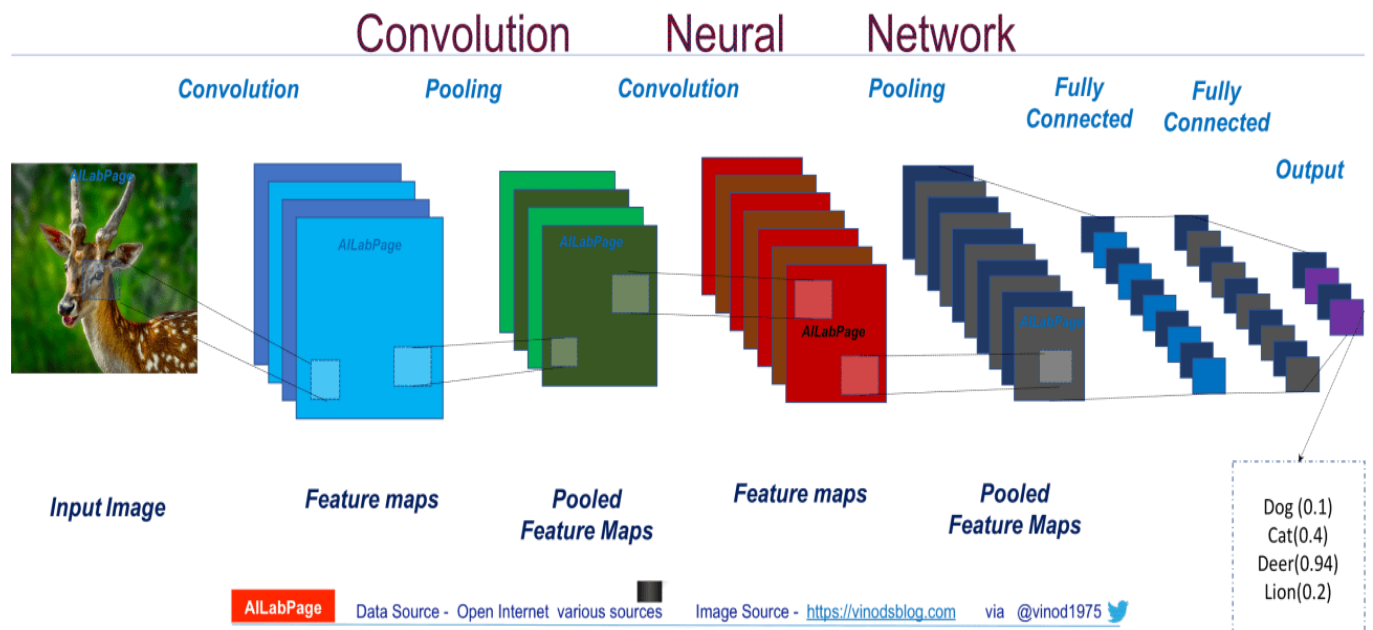
Technology Used:

Machine/Deep learning (Convolutional Neural Network)

Basically, this system works on the concept of Image Recognition and for image recognition we use the concept of Neural Networks which works on the principal of pattern recognition.

The handwritten text is converted in image format and it is fetched in the system. The image is then converted into a matrix of size equal to the pixel size of the image. The matrix contains the RGB values of the Image. This matrix is then interpreted by the trained system and the corresponding correct text is predicted. To create a text recognizer,

A machine learning model is created using Convolutional Neural Network algorithm, which is first trained manually with a large dataset with correct output text. During training the parameters of the algorithm are set according to the pattern of the text and are saved for future prediction. A particular unique pattern is created for each letter, digits and symbols. Then the model is tested with some correct and incorrect datasets and based on the prediction accuracy, further modifications are done in the model.



Description of Project

This project is based on handwritten text recognition. The model created should be able to detect text from any given image and predict that text correctly.

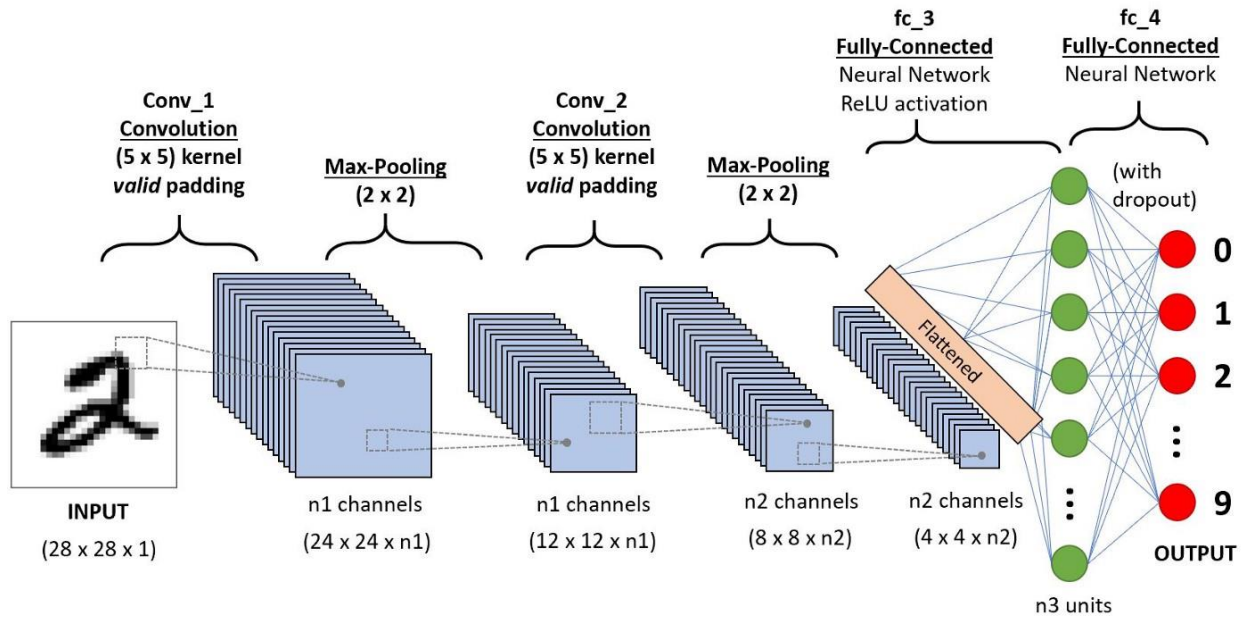
As it is a problem of computer vision, the neural networks will be able to do the prediction. Here the model is trained over an image dataset. A colored image has 3 color channels Red, Green and Blue. For processing the image, it is converted into a numpy array of its pixel values. That pixel value is given as input to the Deep learning model, to be trained. The model processes those inputs and calculates the values of its parameters (called weights) according to the correct outputs provided during the training. Here for processing an image, a large amount of inputs and parameters are present due to which our simple neural network will not be able to handle those large amount of inputs.

*So for creating this Deep Learning Model, we will use **Convolutional Neural Network**.*

Convolutional Neural Network

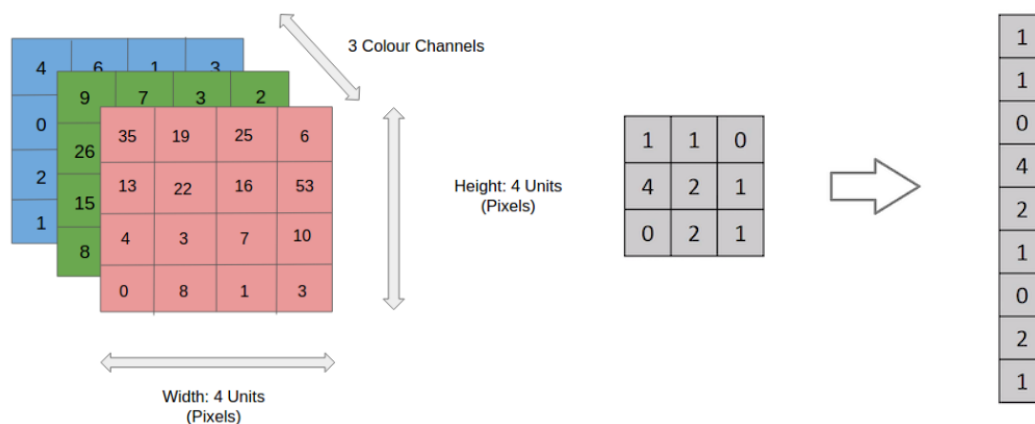
*A **Convolutional Neural Network (ConvNet/CNN)** is a Deep Learning algorithm which can take in an input image, assign importance (learnable weights and biases) to various aspects/objects in the image and be able to differentiate one from the other.*

The pre-processing required in a ConvNet is much lower as compared to other classification algorithms. While in primitive methods filters are hand-engineered, with enough training, ConvNets have the ability to learn these filters/characteristics.



A CNN sequence to classify handwritten digits

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**. Individual neurons respond to stimuli only in a restricted region of the visual field known as the **Receptive Field**. A collection of such fields overlap to cover the entire visual area.



Input For the model:

The pixel value array is flattened as a vector as shown in picture.

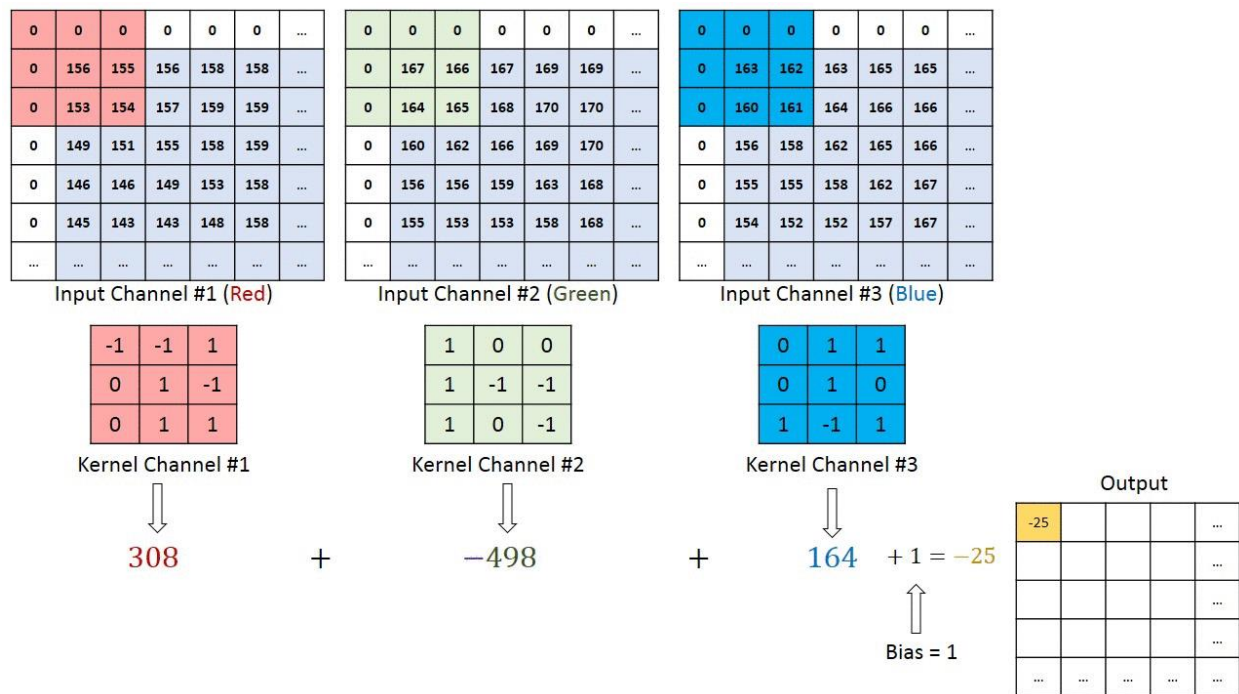
Edge Detection:

Convolutional Layer is used

For detecting anything present in an image ,we have to first analyse the edge of the object in the image,which gives a particular pattern which helps further to detect the complete object.

For edge detection we use kernel which convolves with the image pixel array and detects the edges.

The element involved in carrying out the convolution operation in the first part of a Convolutional Layer is called the **Kernel/Filter**.



Pooling Layer

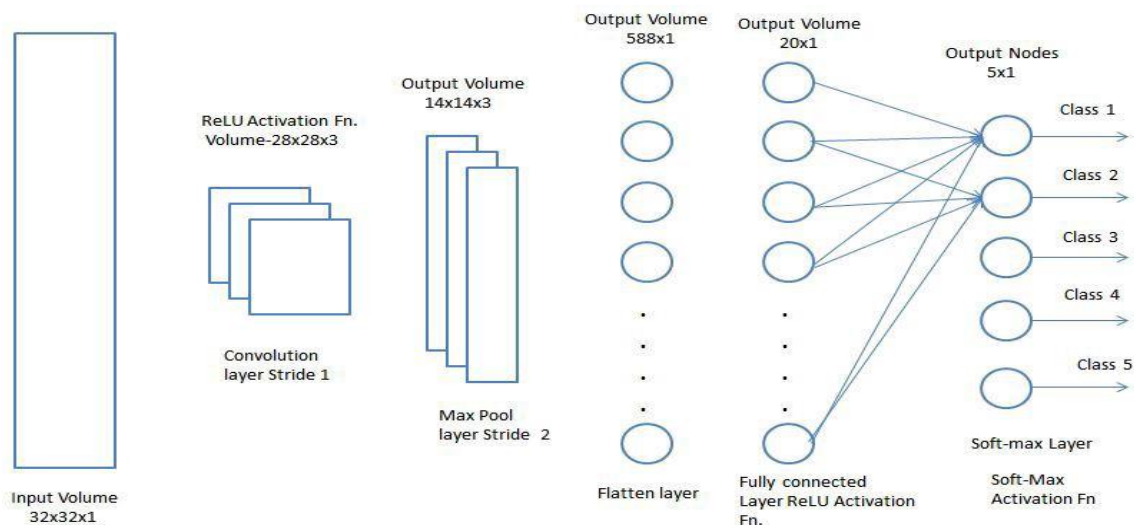
Pooling layer is responsible for reducing the spatial size of the Convolved Feature. This is to **decrease the computational power required to process the data** through dimensionality reduction.

3.0	3.0	3.0
3.0	3.0	3.0
3.0	2.0	3.0

3	3	2	1	0
0	0	1	3	1
3	1	2	2	3
2	0	0	2	2
2	0	0	0	1

Fully Connected Layer

Connected layer is a (usually) cheap way of learning non-linear combinations of the high-level features as represented by the output of the convolutional layer.



Softmax Classification:

It converts the output of the last **layer** in your neural network into what is essentially a probability distribution.

Coding of Project

For the implementation of our project, I have used the following python libraries.

1. Keras (TensorFlow backend)
2. NumPy
3. OpenCV
4. copy

3 Python files are created:

1. model.py
2. detect_text.py
3. predict.py

