MILESTONE REPORT

- ADVANCEMENTS IN CNN ARCHITECTURES -

By: Tanvi Gandhi

➢ PROPOSAL

• MOTIVATION

In recent years, we have witnessed the explosion of Deep Neural Networks, especially Convolutional Neural Networks (CNNs) due to their ability in recognizing visual patterns directly from pixel images with minimal preprocessing. CNNs have been a part of the machine learning family since 1989 and were considered breakthrough due to their exemplary performance. This has inspired several ideas to bring advancements in CNNs have been explored, such as the use of different activation and loss functions, parameter optimization, regularization, and architectural innovations. Thus, we want to further explore the current state of the art practices in Deep Learning, relating to image recognition. These architectures include LeNet, Alexnet, ResNet, VGG-16 and many more. It would be interesting to tweak and play with the architectural components to try and come up with an improved model for the handwritten digits.

• WHO MIGHT CARE?

The domain of deep learning has still a lot to explore and it would be wise to understand the intrinsic details of the different dimensions of a Deep Neural Network so that we could come up with an improved version. Even the mere understanding of the basic principles would be fruitful and pave a way to dive into it. Further, it would be very beneficial for the organizations who are working in the image recognition field to have the comprehended architectures. The goal of this study is not only to have a deep understanding of the working of the NNs but also demonstrate myself and my knowledge and skills in this vast domain.

• DATA

The data comprises the MNIST dataset and Dog-Cat classification dataset which is kind of a benchmark dataset for trivial image processing systems. Due to the limited GPU availability, we would be using the mentioned datasets although better datasets are

available like ImageNet, CIFAR10 etc. The MNIST dataset contains 60,000 training images and 10,000 testing images. For the dog-cat classification, the training set comprises 25,000 images and test set contain 12,500 images.

- APPROACH

As an initial step, it seems apt to decide that we will implement the LeNet and AlexNet architecture. We will tweak and play with the parameters and compare their accuracy on the MNIST and dog-cat dataset.

- DELIVERABLES

  - **Jupyter Notebook:** A python notebook which will contain all the code involved in the process.
  - **Final Report:** A document to highlight the entire process followed and key takeaways.
  - **Presentation and/or Paper:** A presentation for the audience and/or a research paper to highlight the significant insights.

➢ DATASET

- MNIST-10

The MNIST (Modified National Institute of Standards and Technology) is a large database of handwritten digits and is commonly used for training various image processing systems. The images are black and white and fit into a 28 * 28 pixel bounding box. It consists of 60,000 training images and 10,000 testing images.

- DOG-CAT

The Dog-Cat dataset was provided by the Microsoft Research team and is taken from Kaggle website (link: https://www.kaggle.com/c/dogs-vs-cats/data). The training data comprises 25,000 colorful images of dogs and cats and the test data consists of 12,500 images to test upon.

➢ IMPLEMENTATION

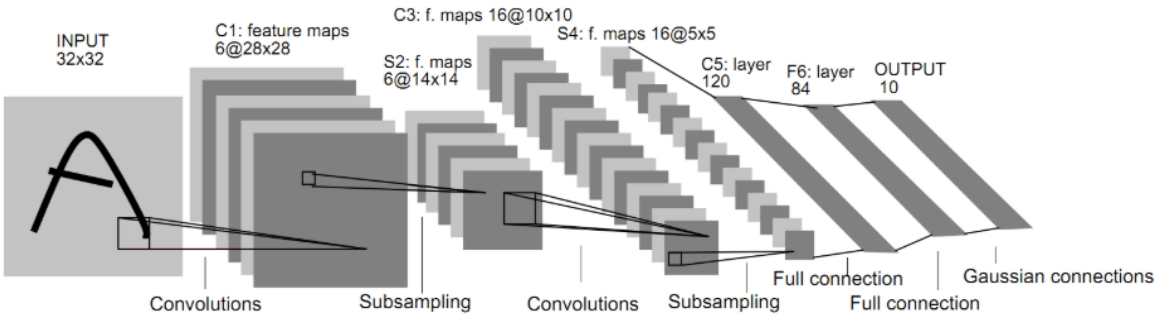
- ## HARDWARE

The implementation of LeNet and AlexNet was done on my local system. The details are:

- Processor : Intel(R) Core(TM) i3-6006U CPU @2.00GHz 1.99 GHz
- Installed Memory (RAM) : 4.00 GB
- System type : 64-bit Windows 10 Operating System, x64-based processor

- ## LENET-5 ON MNIST DATASET

The LeNet architecture comprises total 7 layers, excluding the input layer, with 3 convolutional layers, 2 pooling layers and 2 fully connected layers. I have used 'Relu' as an activation function for the convolutional layers and 'softmax' at the output layer. Below is the architecture used:

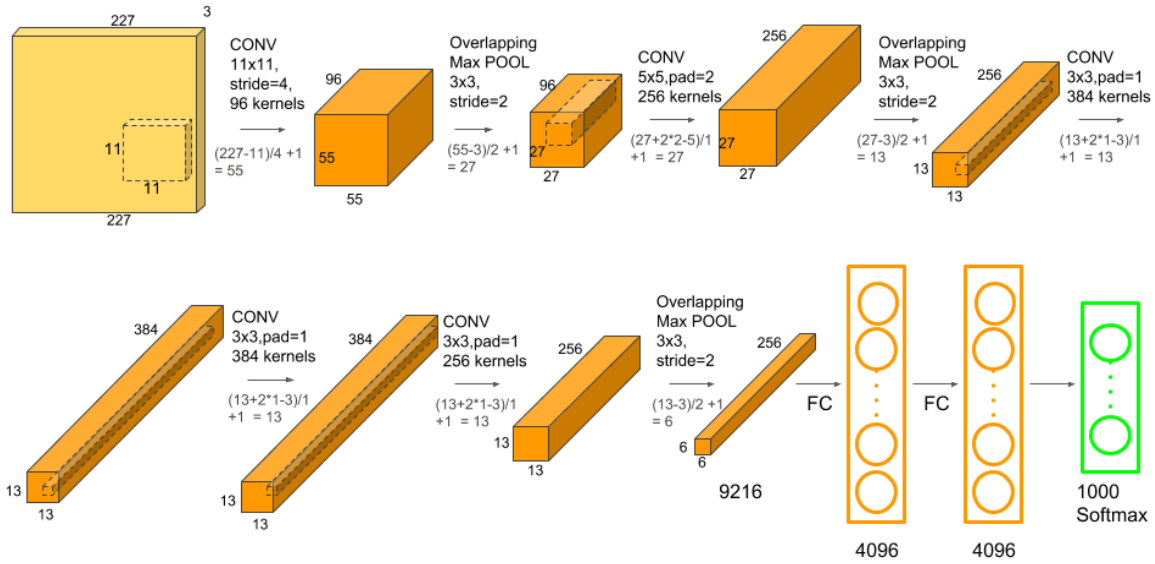


The architecture was trained for 60000 samples and tested on 10000 samples. The total time was almost half hour since the network was really simple for 20 epochs with 128 batch size. The Loss was 0.035 and Accuracy was 0.989 for the testing data.

- ## ALEXNET ON MNIST DATASET

Alexnet is quite a deeper network and came out in 2012 as the winner of the ImageNet competition. It consists of eight layers: five convolutional layers and three fully-connected layers. The features which were not standard at the time when the architecture came out were usage of Relu function which reduces the training time a lot, the overlapping pooling instead traditional "pool" outputs of neighboring groups of neurons with no overlapping and Dropout i.e. turning off neurons with a predetermined probability (e.g. 50%) which means that every iteration uses a different sample of the model's parameters, which

forces each neuron to have more robust features that can be used with other random neurons.



The network was trained on 60,000 samples and tested on 10,000 samples. The total time was almost 3 days for 20 epochs with batch size of 128. The loss was 0.054 and accuracy was 0.9893 for the testing data.