Handwritten Digit Recognition (MNIST Database) using CNN

Madhukshara Sarkar (B18735)

M.Sc. Big Data Analytics

Ramakrishna Mission Vivekananda Educational and Research Institute

Objective:

In this study, we have developed a handwritten digit recognition system using CNN. The dataset used is the MNIST Handwritten Digit Database.

Samples provided from MNIST (Modified National Institute of Standards and Technology) database include handwritten digits total of 70,000 images consisting of 60,000 examples in training set and 10,000 examples in testing set, both with labelled images from 10 digits (0 to 9). All these black and white digits are size normalized, and centred in a fixed-size image where the centre of the intensity lies at the centre of the image with 28 × 28 pixels. The dimensionality of each image sample vector is 28 * 28 = 784, where each element is binary. We will implement two network models - the LeNet-5 and a base CNN model and compare their performances with respect to the MNIST dataset.

Methodology:

In this study, we use two CNN architectures, the widely known LeNet-5 architecture and a base CNN model to recognise handwritten digits from their images. The two models are illustrated and explained below:

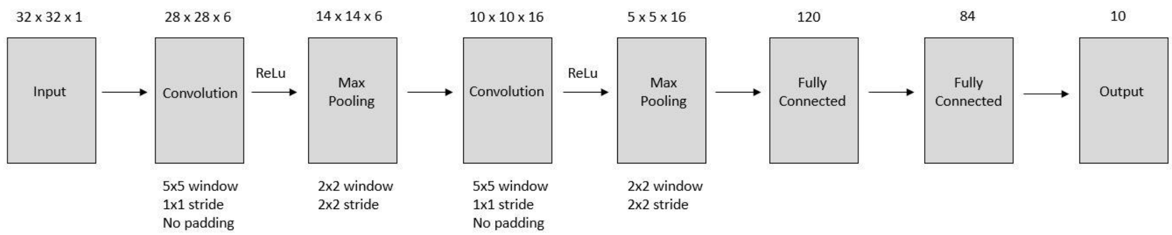

Figure 1: The architecture of the LeNet-5 Model

The LeNet-5 model has the following architecture:

1) Input: 32 x 32 x 1
2) CONV: 5 x 5 size, 6 filters, 1 stride, 0 padding
3) ReLU: $\max(0, h_\theta(x))$
4) POOL: 2 x 2 size, 1 stride
5) CONV: 5 x 5 size, 16 filters, 1 stride, 0 padding
6) ReLU: $\max(0, h_\theta(x))$
7) POOL: 2 x 2 size, 1 stride
8) FC: 120 Hidden Neurons
9) FC: 84 Hidden Neurons
10) FC: 10 Output Classes

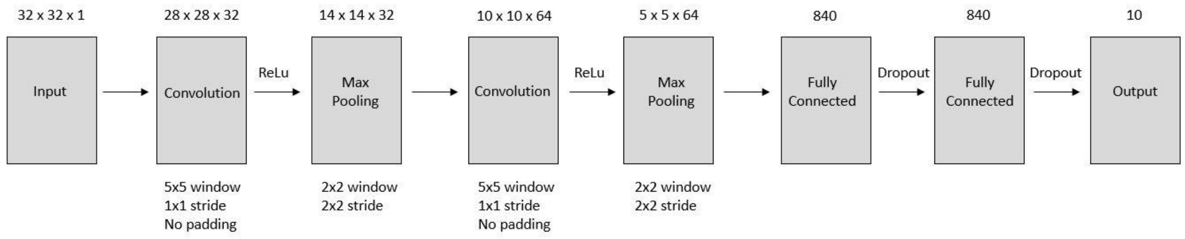

Figure 2: The architecture of the CNN Model

The base CNN model has the following architecture:

1) Input: 32 x 32 x 1
2) CONV: 5 x 5 size, 32 filters, 1 stride, 0 padding
3) ReLU: $\max(0, h_\theta(x))$
4) POOL: 2 x 2 size, 1 stride
5) CONV: 5 x 5 size, 64 filters, 1 stride, 0 padding
6) ReLU: $\max(0, h_\theta(x))$
7) POOL: 2 x 2 size, 1 stride
8) FC: 840 Hidden Neurons
9) DROPOUT: $p = 0.5$
10) FC: 840 Hidden Neurons
11) DROPOUT: $p = 0.5$
12) FC: 10 Output Classes

The following hyper-parameters were manually assigned and used for the experiments.

Hyper-parameters	**LeNet-5**	**CNN**
Training Batch Size	100	100
Testing Batch Size	100	100
Dropout Rate	–	0.5
Learning Rate	$1e^{-3}$	$1e^{-3}$
Steps	10000	10000

Table 1: Hyper-parameters used for the LeNet-5 and CNN Model

Results and Discussion:

Figure 3 shows the training loss of LeNet-5 and CNN on image classification using MNIST.

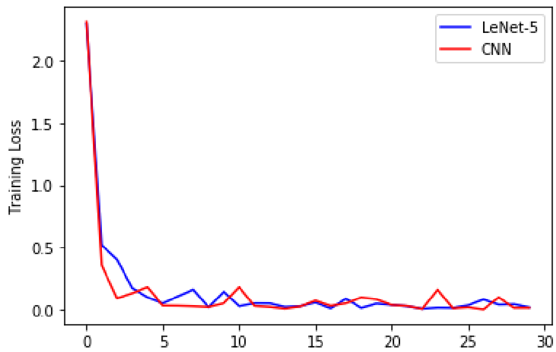

Figure 3: Training Loss of LeNet-5 and CNN Model

It can be seen from Figure 3 that the plot of the CNN model is smoother than that of the LeNet-5 model. At 10,000 steps, the LeNet-5 model was able to finish its training in 7 minutes 33 seconds and the CNN model was able to finish its training in 22 minutes 36 seconds. The LeNet-5 model had an average training loss of 0.155947887369742 and the CNN model had an average training loss of 0.14052866715161752.

After 10,000 training steps, both models were tested on the test cases of each dataset. As shown in Table 1, both datasets have 10,000 test cases each. Table 2 shows the test accuracies of the LeNet-5 and CNN models. Note that there are two test accuracies reported: the maximum accuracy achieved amongst all epochs, and the average accuracy over all epochs.

Model	**Number of Epochs**	**Maximum Test Accuracy (%)**	**Average Test Accuracy (%)**
LeNet-5	5	99.07	98.47
CNN	5	99.25	98.84

Table 2: Test Accuracy of LeNet-5 and CNN Model

Conclusion:

The purpose of this study was to implement a handwritten digit recognition system for the MNIST Handwritten Digit Database. We implemented two networks – the LeNet-5 architecture and a base CNN model and tested their performances.

On the basis of the training loss and testing accuracy of the two models we can conclude that the CNN model has performed better than the LeNet-5 model in classification of the MNIST dataset. The only limitation of the CNN model as compared to the LeNet-5 is that it has a much higher time complexity than the LeNet-5 model.