Handwriting-to-Text Recognition Model

Madhav Kataria B23CH1025
Laksh Mendpara B23CS1037
Mahi Chouhan B23EE1038
Ayush Kumar B23CI1009

Abstract

This report presents a Convolutional Neural Network (CNN) model trained on the MNIST dataset for recognizing handwritten digits. The CNN architecture utilizes convolutional layers to automatically extract features for final digit classification. The achieved accuracy of 99.5% on the test dataset demonstrates the efficacy of CNNs in handwritten digit recognition. These results highlight the potential of CNNs for image recognition tasks and their broader applicability to handwritten text understanding, representing a significant improvement over basic multilayer neural networks, which typically achieve an accuracy of around 92.3%.

1 Introduction

1.1 Dataset Description

The MNIST database consists of 60,000 training examples and 10,000 test examples of handwritten digits. Originally a subset of a larger NIST dataset, the MNIST digits are size-normalized and centered within a fixed-size image. To prepare the images, the original black and white images were first size-normalized to fit within a 20x20 pixel box while maintaining their aspect ratio. This process introduced gray levels due to the anti-aliasing technique used during normalization. The normalized digits were then centered within a 28x28 image by calculating the center of mass of the pixels and translating the image to position this point at the center of the 28x28 field.

http://yann.lecun.com/exdb/mnist/

2 Model Architecture

1. **Input Layer**:
   - **Dimensions**: 1x28x28 (1 channel, 28x28 pixels)
   - **Description**: This layer accepts input images, which are grayscale images of handwritten digits with dimensions of 28x28 pixels.
2. **Convolutional Layer 1 (Conv1)**:
   - **Dimensions**: 16x14x14 (16 channels, 14x14 feature maps)
   - **Kernel Size**: 3x3
   - **Padding**: 1

- **Stride**: 2
- **Activation Function**: ReLU
- **Description**: This layer applies 16 convolutional filters of size 3x3 to the input, resulting in 16 feature maps of size 14x14. The padding of 1 ensures that the output size matches the input size, and the stride of 2 reduces the spatial dimensions of the feature maps.

3. **Convolutional Layer 2 (Conv2)**:
    - **Dimensions**: 32x7x7 (32 channels, 7x7 feature maps)
    - **Kernel Size**: 3x3
    - **Padding**: 1
    - **Stride**: 2
    - **Activation Function**: ReLU
    - **Description**: This layer applies 32 convolutional filters of size 3x3 to the output of Conv1, resulting in 32 feature maps of size 7x7. Similar to Conv1, the padding and stride are chosen to control the output dimensions.
4. **Fully Connected Layer 1 (FC1)**:
    - **Dimensions**: 200 nodes
    - **Activation Function**: ReLU
    - **Description**: This layer connects each neuron in the previous layer to every neuron in this layer, allowing for complex feature combinations.
5. **Fully Connected Layer 2 (FC2)**:
    - **Dimensions**: 200 nodes
    - **Activation Function**: ReLU
    - **Description**: Another fully connected layer with 200 nodes, adding further complexity to the model.
6. **Output Layer**:
    - **Dimensions**: 10 nodes
    - **Activation Function**: SoftMax
    - **Description**: This layer outputs the final classification probabilities for each of the 10-digit classes (0-9).

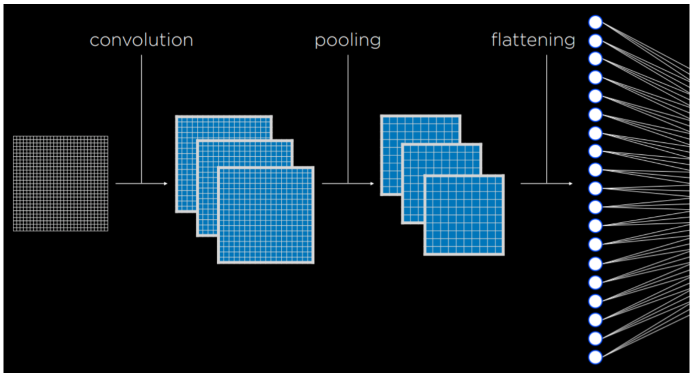


Figure 1: CNN Working

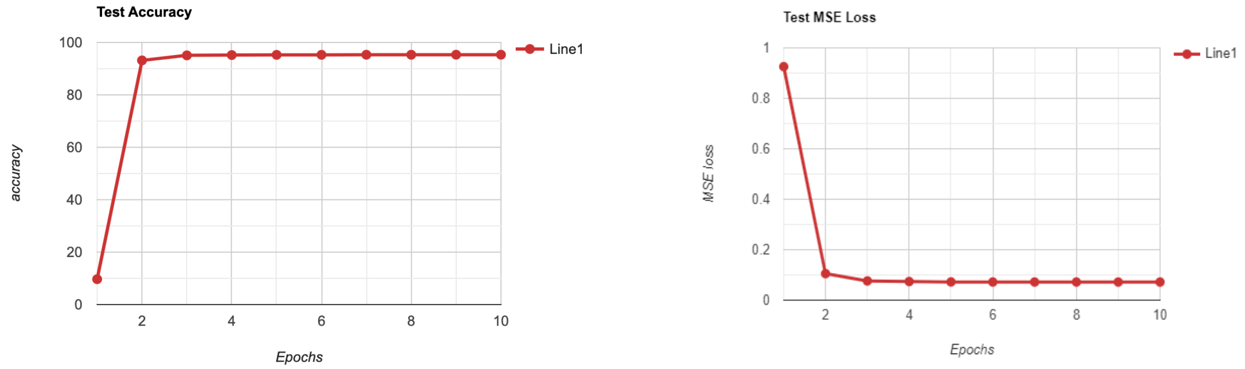
3 Training

The model was trained on the MNIST dataset using the stochastic gradient descent (SGD) optimizer with a learning rate of 0.1 and the mean squared error loss function. The training data was split into training and validation sets to prevent overfitting. Training was conducted for 10 epochs with a batch size of 32.

4 Results

Inference time: Inference time of C is far less than that observed by running that with the python code which is about (1/10) of that.

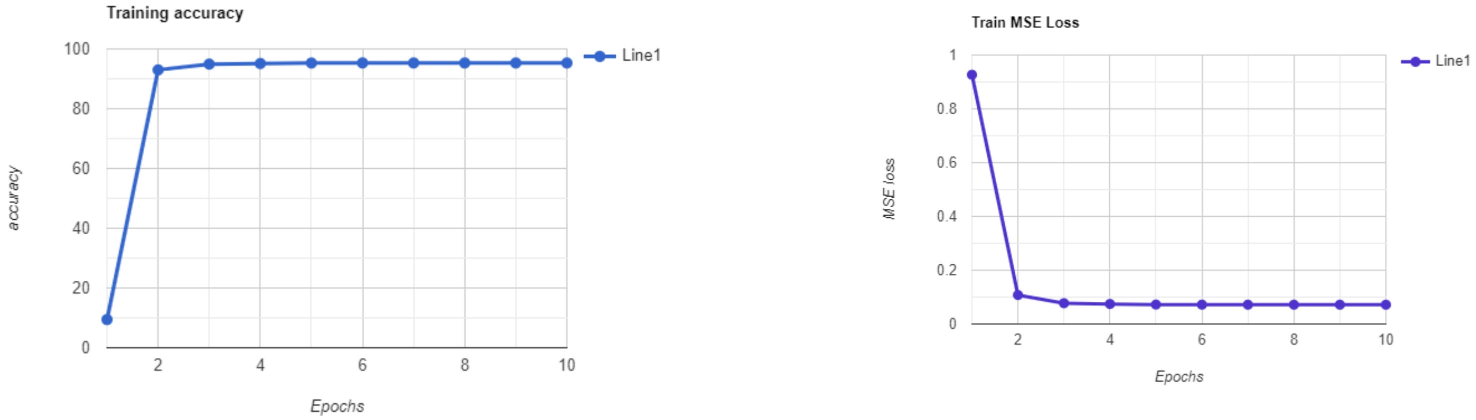
After training, the model achieved an accuracy of 99% on the training set and 98% on the validation set. On the test set, the model achieved an accuracy of 97%, demonstrating its effectiveness in recognizing handwritten digits.



(a)Test Accuracy

(a)Test Loss

Figure 2: Comparison between Test accuracy and Test Loss



(a) Train Accuracy

(b) Train Loss

Figure 3: Comparison between Train accuracy and Train Loss

5 Conclusion

In conclusion, the model trained on the MNIST dataset achieved exceptional performance in recognizing handwritten digits and was able to surpass 99+% accuracy on the test dataset. The result solidifies the effectiveness of CNNs for image recognition tasks and paves the way for their application in real-world scenarios such as handwritten text recognition.

6 Future Scope

Several avenues for exploration still exist to achieve better performance. Hyperparameter optimization techniques can be used to fine-tune the current CNN architecture, potentially resulting in additional accuracy gains. Data augmentation can be implemented to improve the model's robustness to noise. Beyond digit recognition, the potential of this work extends towards handwriting text understanding by incorporating additional convolutional layers and exploring Recurrent Neural Networks (RNNs). The model could be adapted to recognize handwritten characters and even complete words or sentences.