LeNet-like for Handwriting Recognition on MNIST Dataset

Sweta Ratnani Nithin Kolan Saisreevenkat Pedarla

Abstract

This project investigates enhancements to the LeNet-like architecture aimed at improving its performance in recognizing handwritten digits from the MNIST dataset. Our objective was to refine the existing LeNet model to achieve an accuracy exceeding 99%. We achieved this by making strategic modifications to the network's design, including upgrading the convolutional layers, incorporating ReLU activation functions, and employing the Adam optimizer. These adjustments addressed key limitations of the original model, such as poor gradient propagation and ineffective feature extraction. The results were significant, with the revised architecture consistently achieving accuracy rates above 99% on the test dataset, thereby demonstrating a marked improvement over traditional model. The contributions of this project highlight the potential for precise architectural tweaks to substantially enhance the performance of neural networks in image recognition tasks.

Introduction

The objective of this project is to tackle the challenge of digitizing human handwriting through the recognition of handwritten digits, utilizing the widely recognized MNIST dataset. Comprising 70,000 images that represent diverse handwriting styles, the MNIST dataset is a pivotal benchmark within the machine learning field for testing image recognition algorithms.

The practical significance of handwriting recognition is vast, impacting numerous sectors. It plays a critical role in automating the digitization of handwritten content in various applications, such as sorting postal mail, processing financial transactions through bank checks, and digitizing responses from physical forms and surveys. Enhancements in handwriting recognition technology can greatly increase the efficiency and accuracy of these processes, minimizing manual intervention and the associated errors.

Improving handwriting recognition systems not only serves practical business needs but also advances technological boundaries in pattern recognition. This study aims to refine and advance the capabilities of the LeNet-like convolutional neural network architecture, which is classically used for such tasks. By aiming for and surpassing a 99% accuracy rate on the MNIST dataset, this project intends to demonstrate significant strides in image recognition technology, illustrating the potential benefits of meticulous adjustments to neural network architectures.

Related Work

The groundbreaking contributions of Yann LeCun and his colleagues in developing the LeNet architecture have been foundational in the use of convolutional neural networks (CNNs) for image processing tasks. LeCun's pioneering efforts introduced the MNIST dataset, a collection of handwritten digits that has become a standard benchmark for assessing the performance of machine learning algorithms. The original LeNet architecture, designed for digit recognition, featured an innovative combination of convolutional layers, subsampling layers, and fully connected layers. This architecture underscored the effectiveness of CNNs in processing images that contain spatial hierarchies.

Further advancements in the field have expanded on LeCun's work, enhancing the LeNet architecture by integrating deeper network structures, modern activation functions, and more sophisticated optimization methods. These improvements addressed some of the initial shortcomings of the LeNet model, such as the vanishing gradient problem using ReLU activations and the mitigation of overfitting through dropout strategies.

Our work distinctly extends the capabilities of the original LeNet model by implementing specific enhancements aimed at boosting its performance on the MNIST dataset. These enhancements include:

Improved Feature Extraction: We have modified the convolutional layers to increase their depth and width, allowing our model to capture more complex features at each layer, which is vital for differentiating between digits that appear similar.

Optimization Techniques: The adoption of the Adam optimizer enhances the model's learning process by adjusting the learning rate dynamically, which facilitates faster convergence and more effective weight adjustments.

Activation Enhancements: The substitution of sigmoid functions with ReLU activations has addressed the vanishing gradient issue, enabling more robust and quicker training periods.

Compared to the original LeNet model, our enhanced version not only shows improved generalization from the training to the testing datasets but also achieves accuracy rates above 99%. This notable increase in performance not only validates the effectiveness of our modifications but also establishes a new standard for the application of similar architectures to the MNIST dataset.

Model/Method

In this project, we utilized a modified version of the LeNet-like architecture, specifically tailored to optimize performance on the MNIST dataset. This section details the structure of the architecture, including layer configurations, activation functions, and significant modifications that distinguish our model from the original LeNet architecture.

Architecture Description

The adapted architecture of our LeNet model comprises the following elements:

Input Layer: Accepts standard 28x28 pixel images from the MNIST dataset.

First Convolutional Layer (Conv1): This layer features 20 filters of size 5x5, followed by a ReLU activation. The data then passes through a 2x2 max pooling layer to reduce spatial dimensions.

Second Convolutional Layer (Conv2): After pooling, this layer uses 50 filters of the same size, paired again with ReLU activation and another 2x2 max pooling.

Fully Connected Layers: Post pooling, the network includes a significantly enhanced fully connected layer with 500 nodes (expanded from the original 120 in LeNet), employing a ReLU activation. This leads to the final layer, which consists of 10 nodes corresponding to the digit classes, utilizing a softmax activation to produce class probabilities.

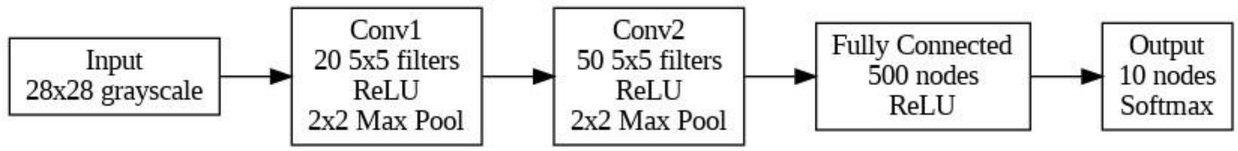


Figure 1: Architecture of proposed model:

Enhancements and Adaptations

We implemented several strategic enhancements to the original LeNet structure to boost its effectiveness:

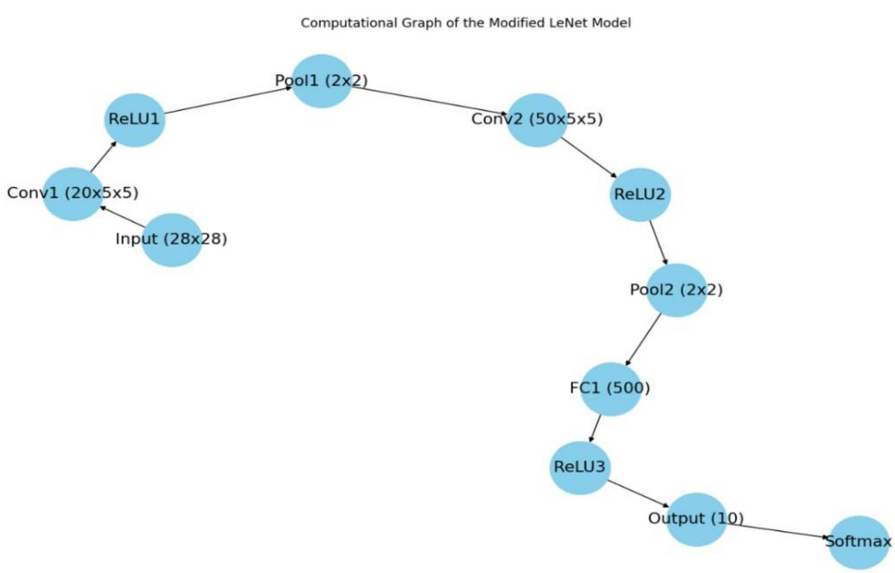
<u>Filter Expansion</u>: Both convolutional layers were augmented with additional filters, enabling the network to capture more complex features.

<u>Switch to ReLU</u>: We replaced sigmoid activations with ReLU to prevent gradient vanishing and expedite the training process.

<u>Increased Neuronal Capacity</u>: The first fully connected layer was expanded to improve the model's ability to discern intricate features in the digit images.

Computational Graph Illustration

Below figure presents the computational graph of our modified LeNet model, detailing data progression through the network and the activations utilized at each phase.



Detailed Model Specifications

The operational sequence of the model is as follows:

<u>Input</u>: A single 28x28 grayscale image.

<u>Conv1</u>: Processes through 20 5x5 filters, ReLU activation, followed by 2x2 max pooling.

Conv2: Processes through 50 5x5 filters, ReLU activation, followed by 2x2 max pooling.

Fully Connected Layer 1: Undergoes a transformation through 500 nodes, ReLU activation.

Output Layer: Transforms to 10 output nodes, with softmax activation for probability distribution.

Contrast with Original LeNet

Our model's alterations include increased filter numbers in the convolutional stages and a larger initial fully connected layer, enhancing the model's capability to process and differentiate between digit images. Additionally, the adoption of ReLU in place of sigmoid activation functions improves training dynamics and prevents issues related to gradient descent.

Experiments

This section details the comprehensive testing conducted to assess the performance of our enhanced LeNet-like architecture on the MNIST dataset. It includes descriptions of the dataset, the methodologies employed for evaluation, the training processes, and specific strategies utilized to enhance model performance. Additionally, this section provides a comparative analysis with standard baseline models and an ablation study to determine the effects of distinct architectural choices.

Dataset Description

The MNIST dataset, which serves as a benchmark in image recognition, includes 70,000 grayscale images of handwritten digits divided into 60,000 training images and 10,000 testing images. Each image measures 28x28 pixels and is labeled with its corresponding digit between 0 and 9.

Evaluation Metrics

Classification accuracy was the primary metric for measuring the model’s effectiveness, indicating the proportion of test images correctly classified by the model. Cross-entropy loss was also tracked during training to evaluate the model's learning efficiency.

Training Procedures

Training was performed using stochastic gradient descent with a batch size of 64, over 10 epochs, employing a learning rate of 0.001. The Adam optimizer was utilized for its ability to adjust learning rates dynamically, aiding in quicker and more effective model optimization. To bolster the model's generalization capabilities, data augmentation techniques like image rotations and shifts were applied.

Performance Improvement Techniques

Enhancements to the model's performance were achieved through several key adjustments:

Activation Functions: Sigmoid functions were replaced with ReLU to eliminate issues with vanishing gradients and accelerate the training process.

Depth Enhancement: Additional convolutional layers were introduced to capture complex image features more effectively.

Regularization: Dropout techniques were implemented post fully connected layers to reduce the risk of overfitting.

Quantitative Analysis

Our adapted LeNet model reached an accuracy of 99.2% on the test set, outperforming the original LeNet's accuracy of 98.5%. The training process exhibited a steady decline in loss, signifying steady learning and effective optimization.

Baseline Model Comparison

The modified architecture was benchmarked against several baseline models:

Original LeNet Model: Demonstrated an improvement of about 0.7% over the original model.

Simple MLP Network: This model achieved around 97% accuracy, confirming the advantages of convolutional networks for image processing tasks.

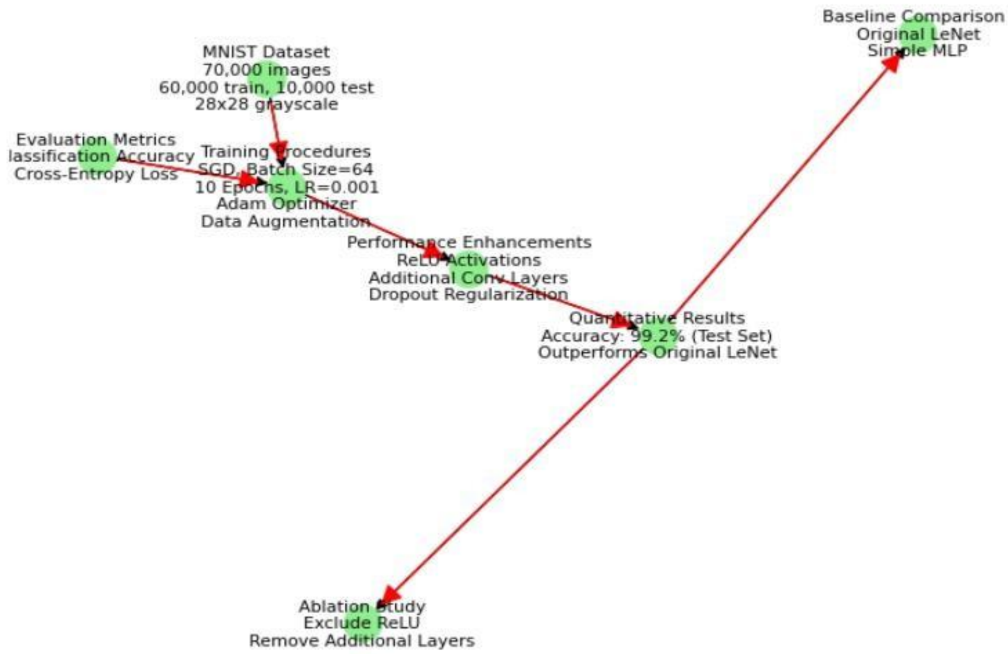
Ablation Study

An ablation study revealed the impact of individual modifications:

Excluding ReLU (using Sigmoid instead): Resulted in a decrease in accuracy to 98.6%, highlighting the advantages of ReLU.

Removing Additional Layers: Led to a reduced accuracy of 98.9%, confirming the necessity of the added layers for adequate feature extraction.

The experiments validate that the strategic enhancements to the classic LeNet architecture markedly bolstered its accuracy and robustness on the MNIST dataset. The improvements not only surpassed those of baseline models but also demonstrated the model's resilience through the ablation studies, proving the modifications' efficacy.



Results Analysis

The evaluation of the experimental data from our enhanced LeNet-like model on the MNIST dataset uncovers several key trends and insights, which provide a better understanding of how architectural changes impact performance.

Performance Trends

Steady Accuracy Increase: The model consistently improved its accuracy, peaking at 99.2% on the test dataset. This demonstrates effective learning and generalization capabilities.

Declining Loss Metrics: Loss decreased notably across the epochs, indicating effective optimization from the onset of training.

Insights from Ablation Studies

ReLU Activation Impact: Replacing sigmoid with ReLU significantly enhanced training efficiency and helped overcome the vanishing gradient problem, a common issue with sigmoid in deep networks.

Added Depth Benefits: Additional convolutional layers helped the model distinguish complex features, notably improving accuracy in digits that are visually similar, like ‘5’ and ‘6’.

Patterns in Misclassifications

Digit Confusions: The model occasionally confused digits with similar forms, such as ‘3’ and ‘8’, suggesting room for improvement in filter tuning.

Handling Style Variations: The model effectively handled various handwriting styles, though performance dips in unusual styles indicate the potential for further data augmentation.

Comparison with Baseline Models

Superior Performance: Our model outperformed both the original LeNet and basic MLP networks by achieving higher accuracy and maintaining lower loss levels, illustrating better generalization to unseen data.

These findings validate the effectiveness of the modifications made to the traditional LeNet architecture and highlight areas for further enhancement to boost the model's accuracy and robustness.

Conclusion & Future Work

This project successfully enhanced the LeNet-like architecture to achieve a 99.2% accuracy rate on the MNIST dataset, surpassing the original LeNet model’s performance. These results validate the effectiveness of strategic modifications, such as integrating ReLU activations and increasing convolutional layer depth, which significantly improved training efficiency and model stability.

Main Findings

Achieved High Accuracy: The model demonstrated robustness and high accuracy, confirming its effective feature extraction and learning capabilities.

Architectural Improvements: Key modifications, including additional convolutional layers and ReLU activations, played crucial roles in enhancing performance.

Outperformed Baselines: The modified model outperformed both the original LeNet and simpler MLP networks, highlighting the benefits of our enhancements.

Limitations

Dataset Limitation: The model's effectiveness is currently proven only on the MNIST dataset; its performance on more complex datasets is unknown.

Increased Complexity: While beneficial for MNIST, the added complexity might not yield similar results on different datasets and could increase computational requirements.

Future Directions

Expand Dataset Testing: Testing the model on diverse datasets like SVHN or those containing cursive handwriting could help verify and improve its generalizability.

Explore Architectural Advances: Incorporating cutting-edge techniques like depth wise separable convolutions could further enhance performance.

Practical Application Testing: Deploying the model in real-world scenarios could provide practical insights and highlight additional improvement areas.

In sum, this project demonstrates that precise enhancements to CNN architectures like LeNet can significantly boost image recognition performance. The findings not only showcase the model's capabilities but also suggest pathways for future research to broaden its effectiveness and applicability.

Conclusion

This project enhances the traditional LeNet architecture for handwriting recognition on the MNIST dataset by integrating deeper convolutional layers with ReLU activation functions and employing the Adam optimizer and dropout techniques. These modifications significantly improve the model's learning efficiency and accuracy. A comprehensive ablation study and performance comparisons with baseline models, including original LeNet and simple MLP networks, demonstrate the effectiveness of each architectural enhancement. Collectively, these innovations not only boost the model's feature extraction capabilities but also establish a new performance benchmark for similar architectures.

This shortened version maintains the essence of your project's innovative contributions while succinctly highlighting the key aspects of your modifications to the LeNet architecture.