Team Members: Guanxiang Diao, Hanlin Yang, Songyan Ma, Wangshuang Xu, Boliang Chen

Evaluating Deep Learning Models for Handwritten Number Recognition

This paper presents an analysis and evaluation of two deep learning architectures, Convolutional Neural Network (CNN) and Recurrent Neural Network (RNN), in the context of recognizing handwritten numbers. The study utilizes the MNIST dataset and delves into various aspects of model development, including pre-processing, architectural choices, training details, and performance evaluation. Additionally, the paper discusses challenges encountered during the implementation and proposes solutions to overcome them.

The primary task is to develop models capable of recognizing handwritten numbers from the MNIST dataset, a collection of grayscale images representing digits 0 to 9. Key preprocessing steps involve normalization and scaling pixel values to a 0-1 range, facilitating network convergence during training. Moreover, one-hot encoding of labels is employed to create a binary matrix format suitable for the models' output layers.

The paper describes the implementation of two distinct architectures: CNN and RNN. The CNN architecture is characterized by its convolutional layer for feature extraction, followed by a pooling layer to reduce spatial dimensions, and fully connected layers for classification. In contrast, the RNN architecture focuses on processing data sequences, interpreting image rows as sequences to capture dependencies along horizontal pixel alignments.

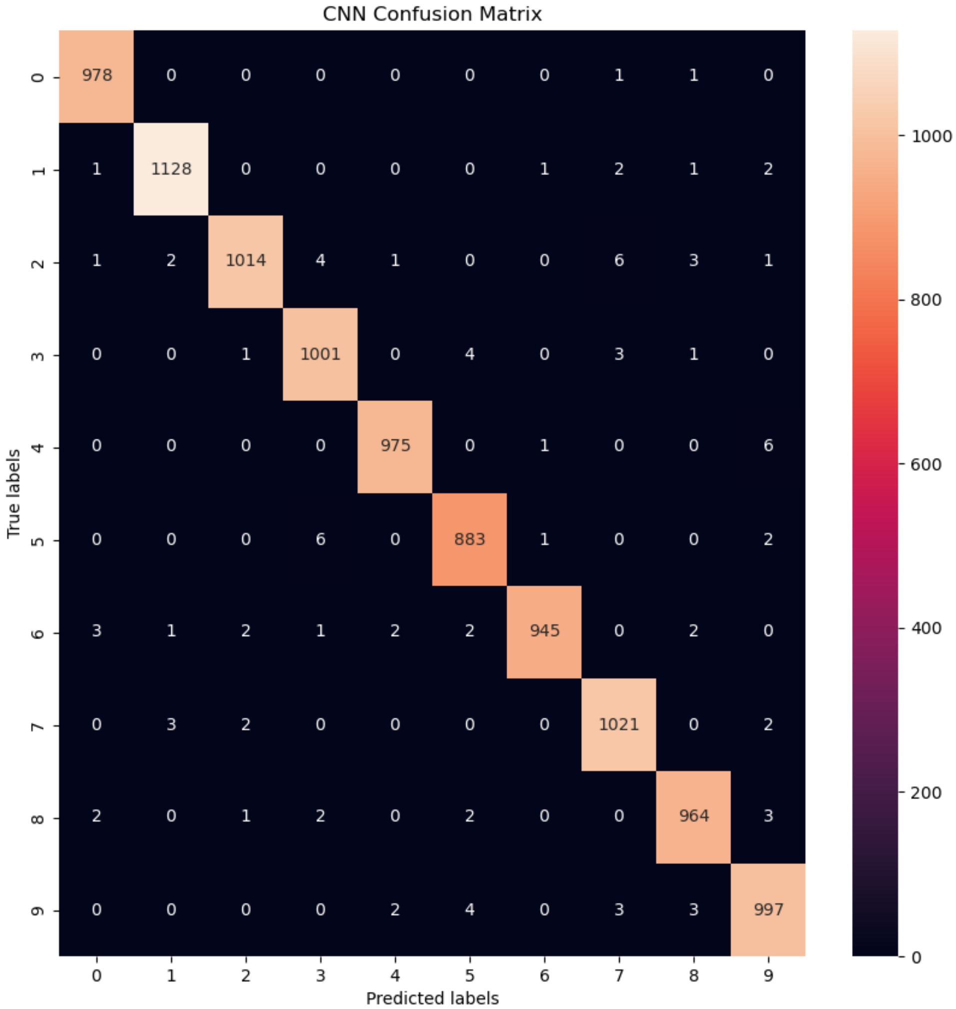
Both models underwent extensive training, iteratively adjusting learning rates and other hyperparameters to achieve convergence. The training process is documented through loss and accuracy graphs, which also help in monitoring overfitting. The CNN demonstrates consistent improvements in training and validation accuracy, while the RNN displays a more volatile training pattern with potential overfitting issues, as indicated by increasing validation losses.

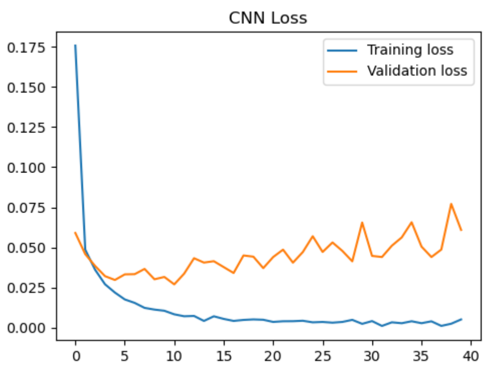
Evaluation metrics include accuracy, loss graphs, and a confusion matrix. The CNN outperforms the RNN, aligning with the general consensus on CNN's suitability for image-based tasks. The confusion matrix highlights high classification accuracy with some notable misclassifications, more prevalent in the RNN model.

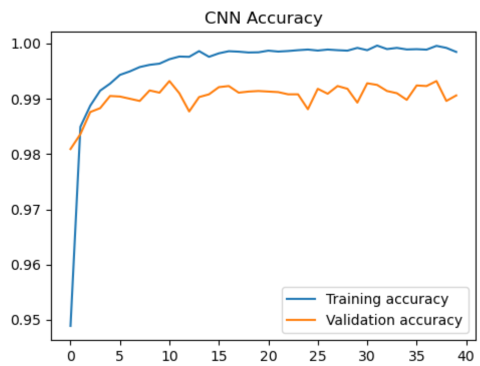
Significant challenges included model selection, hyperparameter tuning, and overfitting prevention. Solutions for the RNN's overfitting involved dropout, early stopping, or regularization techniques. For the CNN, the focus was on building an architecture that captures complex image patterns while ensuring generalization.

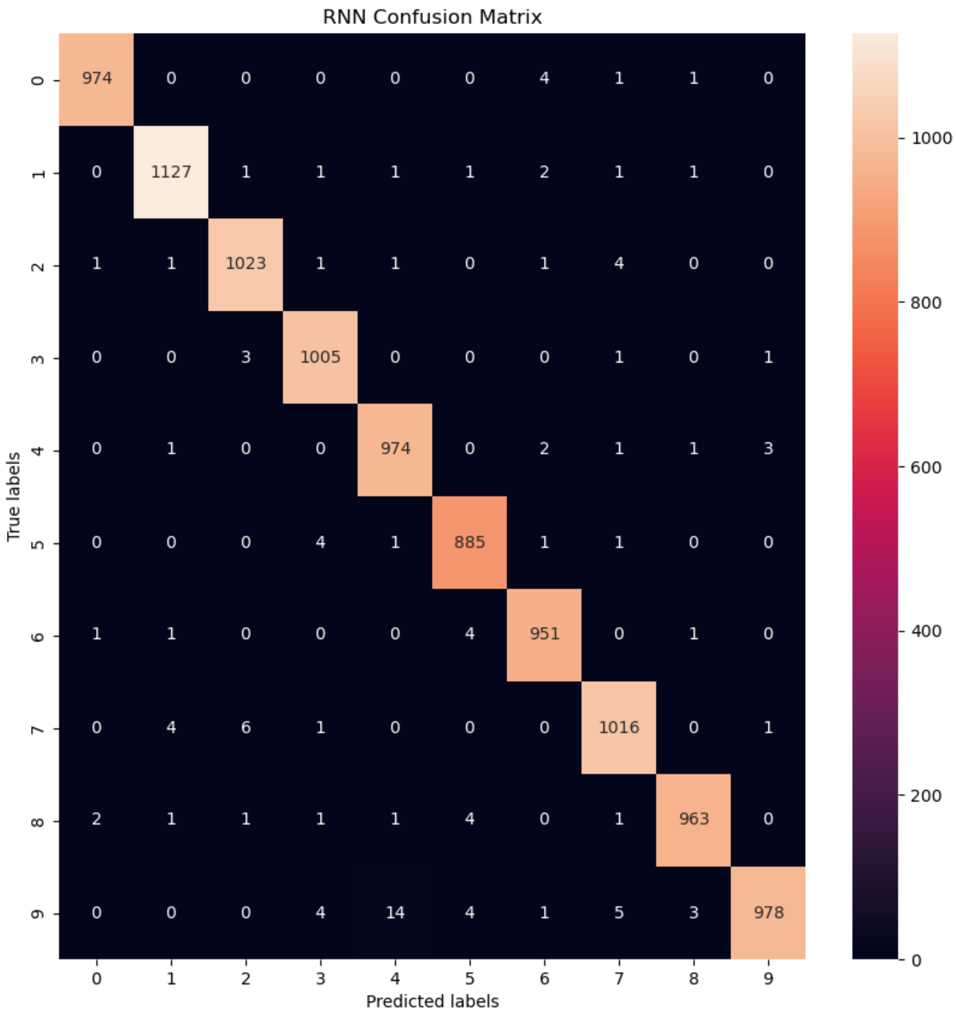
Monitoring and validation sets played a crucial role in preventing overfitting, supplemented by data augmentation techniques to enhance generalization.

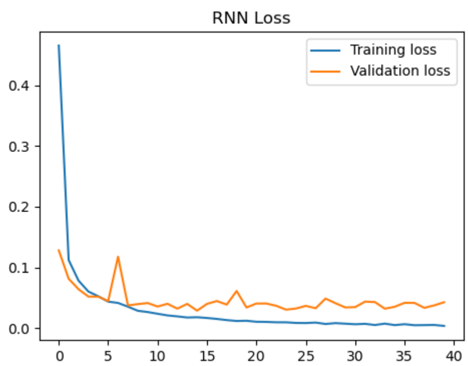
This study underscores the efficacy of CNN over RNN for handwritten number recognition, attributing this to CNN’s spatial feature extraction capabilities. The challenges encountered emphasize the importance of careful model and hyperparameter selection, as well as the need for vigilant monitoring to avoid overfitting. The findings contribute valuable insights into the application of deep learning models in image-based classification tasks.

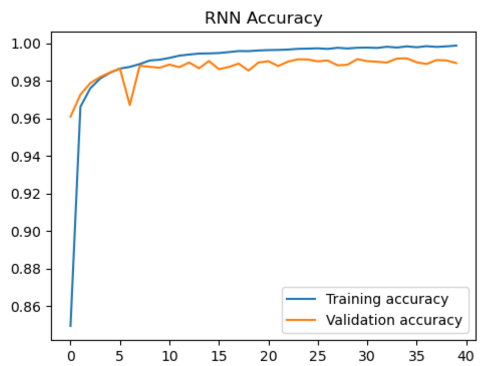

CNN Confusion Matrix
True labels
Predicted labels


CNN Loss
Training loss
Validation loss


CNN Accuracy
Training accuracy
Validation accuracy


RNN Confusion Matrix
True labels
Predicted labels
0 1 2 3 4 5 6 7 8 9
974 0 0 0 0 0 4 1 1 0
0 1127 1 1 1 1 2 1 1 0
1 1 1023 1 1 0 1 4 0 0
0 0 3 1005 0 0 0 1 0 1
0 1 0 0 974 0 2 1 1 3
0 0 0 4 1 885 1 1 0 0
1 1 0 0 0 4 951 0 1 0
0 4 6 1 0 0 0 1016 0 1
2 1 1 1 1 4 0 1 963 0
0 0 0 4 14 4 1 5 3 978
0 200 400 600 800 1000


RNN Loss
Training loss
Validation loss
0.0 0.1 0.2 0.3 0.4
0 5 10 15 20 25 30 35 40


RNN Accuracy
Training accuracy
Validation accuracy
0.86 0.88 0.90 0.92 0.94 0.96 0.98 1.00
0 5 10 15 20 25 30 35 40