Part 1: Training a CNN for Digit Sum Prediction

Gaurav Bhole - 2021113015

Introduction:

This report describes the development and training of a Convolutional Neural Network (CNN) model designed to predict the sum of digits in handwritten number images. The model was implemented using PyTorch and trained on the given custom dataset. The model's performance was assessed using a comprehensive set of metrics and visualizations.

Dataset and Preprocessing:

- The dataset consists of 30,000 images (40x168 pixels) of handwritten numbers.
- Images were normalized and reshaped to (1, 40, 168) for input to the CNN.
- Data was split into training (75%), validation (10%), and test (15%) sets.

Model Architecture: A custom CNN was implemented with the following key features:

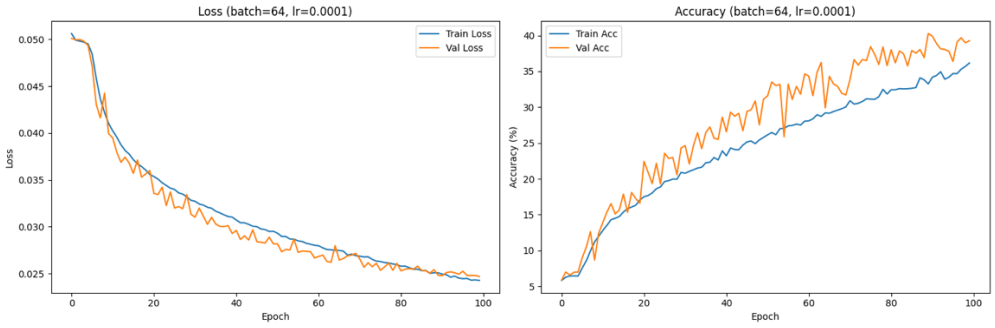
- Three convolutional blocks, each containing:
  - Two Conv2D layers with BatchNorm and ReLU activation
  - MaxPooling and Dropout for regularization
- Fully connected layers for classification
- Output layer with 37 classes (for sums 0-36)

Training Process:

- Optimizer: AdamW with weight decay for regularization
- Loss function: Cross-Entropy Loss
- Batch size: 64
- Learning rate: 0.0001
- Number of epochs: 100

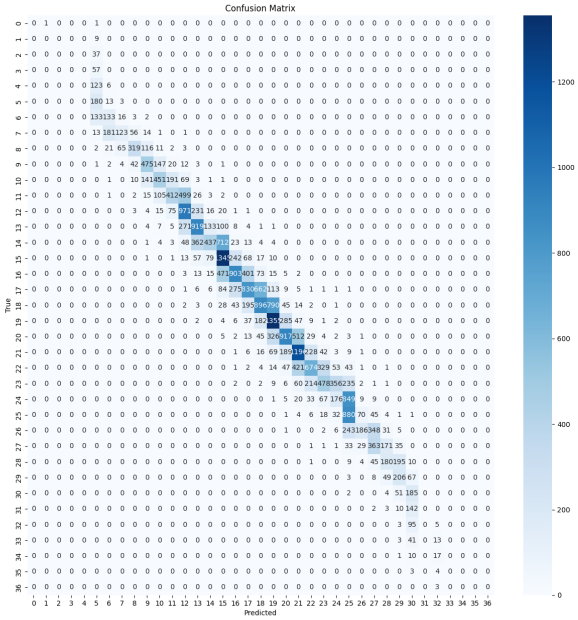
Results and Analysis:

1) Training Curves:

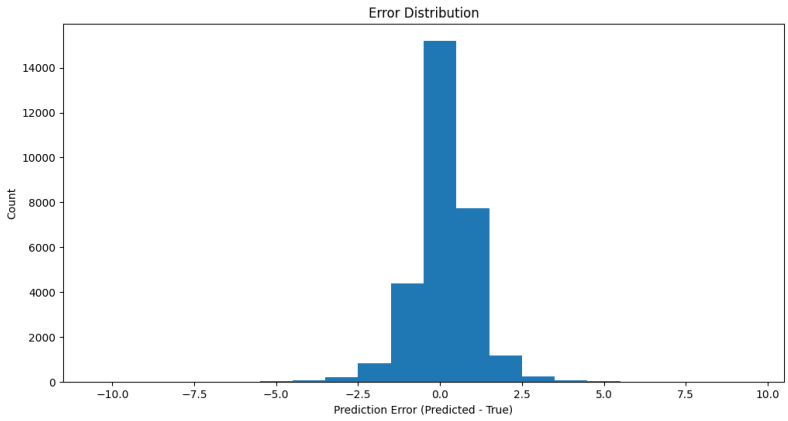


- Loss steadily decreased for both training and validation sets.
- Accuracy improved consistently, reaching ~36% for training and ~40% for validation.
- No significant overfitting observed.

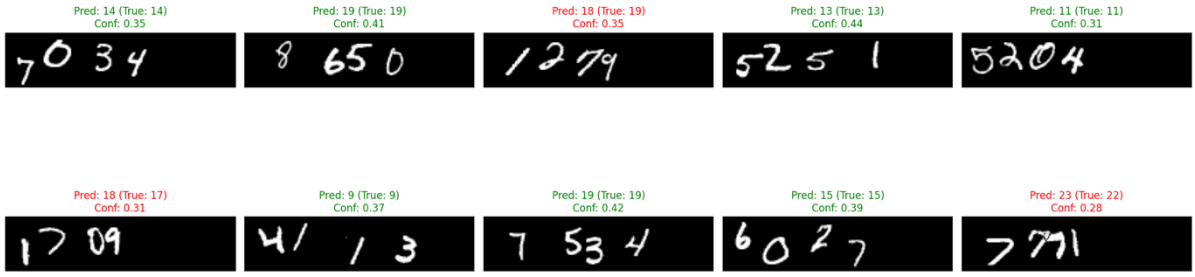
2) Confusion Matrix: Strong diagonal indicates good performance across multiple classes. Some off-diagonal elements suggest confusion between nearby sum values. Higher accuracy observed for lower sum values, possibly due to more training examples.



3) Error Distribution: Roughly normal distribution centred around 0. Most errors within ±2 range, indicating reasonable accuracy.



4) Performance Metrics: The model achieves moderate accuracy, with over 90% of predictions within ±1 of the true sum. This suggests good overall performance for the complex task of summing multiple handwritten digits.
    - Accuracy: 50.65%
    - Mean Absolute Error (MAE): 0.6142
    - Root Mean Square Error (RMSE): 0.9729
    - Within-One Accuracy: 91.12%
5) Sample Predictions: Misclassifications often involve visually similar digits (e.g., 7 and 1).



Conclusion:

The CNN model demonstrates promising performance in the challenging task of predicting digit sums from handwritten number images. With an accuracy of 50.65% and 91.12% of predictions within one of the true value, the model shows good generalization. However, there is room for improvement, particularly in handling more complex cases and in calibrating prediction confidence.