Deep Learning Assignment

1 Data Cleaning and Preprocessing

The following preprocessing steps were performed:

- All the data given was loaded and concatenated.
- Upon inspection, each image contains 4 digits which are non-overlapping, have 1 channel (grayscale) and have size of 40×168 pixels.
- Created a dataloader and split it into training validation and test sets in the ratio 80:10:10 and batch size 32.

2 Phase 1: CNN Baselines

- Since a maximum of 4 digits are contained in any image, this was treated as a classification problem with 37 classes corresponding to the possible sums.
- Cross Entropy loss was used for this.
- Used a basic CNN Architecture with two convolution and pooling layers. Implemented early stopping based on the validation loss.
- Got an accuracy of ∼10%
- Weights for this model are stored in `best_model.pth`
- With a similar approach, instead of defining the architecture, a ResNet model was also tested. However, since pretrained ResNets have 3 channels (RGB) and use very different data, training the architecture from scratch on the given data seemed to be a better option.
- This performs better with an accuracy of ∼
- Weights for this are stored in `best_resnet.pth`