Part-2 Report

Approaches tried:

- Using a deeper CNN + ViT framework
    a. This approach provided better accuracy compared to the baseline but remained limited in complexity and depth. As a result, it struggled to adequately learn from the data and did not meet performance expectations.
- Training on normal MNIST-classification dataset, then segmenting images using:
    a. YOLO
    b. Cv2.contours

    After obtaining bounding boxes for digits, additional preprocessing steps were performed, including thinning the numbers using convolution operations and padding them with black pixels to standardize their size. Labels were predicted for each digit, summed up, and compared with the ground truth dataset labels.

    This approach worked sub-optimally due to the absence of direct training on the original dataset. Consequently, it was abandoned in favor of approaches that utilized the provided data directly.

- Using a deep pretrained model like ResNet (final approach)
    - This gave better results like 65% accuracy within 0.5 of true value and 98% within 5 of true value
    - Trained on a learning rate of 1e-3 and on the original data given
    - Loss -> MSE
    - Model was trained for 10 epochs

Conclusion:

The final approach of using a deep pretrained model, ResNet, yielded the best results among all the approaches explored. By training directly on the original dataset and fine-tuning the model, significant improvements in accuracy were observed. Better methods for this could include use of pretrained ViTs.

Code:

Can be found in Training.ipynb & Inference.ipynb