Deep Learning Assignment Part 2

Pragya Khanna 2022122003

Link to the model: model *(since it exceeded the github's 25Mb quota)*. Continuing from where we left off the discussion in previous part

Approach 2: Non-Pipelined Model

In this approach, the task was divided into three distinct steps:

Segmentation of the Image into Constituent Digits

- The input image was processed to identify and isolate individual digits. Using image processing techniques like thresholding and connected component analysis, each digit in the image was segmented into separate regions.
  - Thresholding: Converted the image to a binary format where pixel intensities above a certain value were set to 1, and others to 0.
  - Connected Component Analysis: Labeled connected regions in the binary image, treating each labeled region as a potential digit.
  - Region Filtering: Eliminated noise by discarding regions that were too small to be valid digits.
- **Why?** Proper segmentation ensures that each digit can be independently processed for classification. Without segmentation, the classifier might struggle to differentiate overlapping or clustered digits.

Individual Digit Prediction

- After segmentation, each isolated digit was resized to a fixed shape (28x28) to match the input requirements of a pre-trained classifier. A digit recognition model, trained on the MNIST digit dataset, was used to classify each digit and output its numerical value.
- A digit recognition model simplifies the task of identifying individual digits. By breaking down the problem into smaller, localized predictions, the model could focus on recognizing one digit at a time, reducing the complexity of the task.

Summing the Predictions

- Once all the digits in the image were classified, their numerical values were summed to obtain the final prediction.

Issues encountered here:

- The accuracy of the entire pipeline depended heavily on the segmentation step. Even small errors in segmenting digits (e.g., missing or merging digits) led to incorrect predictions downstream.
- A deep learning pipeline (Approach 3) could integrate segmentation, recognition, and summation into a single model, leading to better performance by learning features and relationships directly from data.

Approach 3: Pipelined Model

Digit Region Extraction

The first task was to identify and extract individual digit regions from the input image. This step involved image preprocessing and connected component analysis, implemented in the extract_digit_regions function.

- The image was binarized using a threshold value (e.g., 0.5). Pixels with intensities above the threshold were set to 1 (foreground), and others to 0 (background).
- Connected components were labeled, and their properties (e.g., bounding box, center, area) were calculated.
- Small or noisy regions (e.g., fewer than 10 pixels) were discarded. The extracted regions were sorted by their x-coordinates to maintain the reading order of digits.
- If the number of regions exceeded the maximum allowed (i.e., 4), the closest regions were merged. Conversely, if fewer regions were detected, empty placeholders were added.

Dataset Creation

- Input images were processed to extract individual digit regions and their corresponding digit crops.
- The digits were normalized and repeated across three channels to match the input requirements of ResNet18.

Model Architecture

- Digit Classifier: A ResNet18 model, modified for digit recognition.
- Sum Predictor: A fully connected layer that aggregated predictions for all digits and predicted their sum.

Advantages and Roadblocks

- The pipeline integrated all steps, eliminating the need for separate manual segmentation and prediction stages.
- The pipeline demonstrated higher accuracy and generalization compared to the non-pipelined approach.
- The pipeline required more computational resources and time for training.