MLNS Deep Learning Assignment - Part 2

Yashas SB

January 2025

Problem statement and methods explored

In this problem, we aim to classify images containing multiple digits, where the label is the sum of those digits. Before arriving at my final model, I tried to work with segmentation models and Multi-label Classifiers. While they are potential approaches, they have notable limitations for this task.

1. Segmentation Models

A segmentation model is typically used to identify and separate different regions in an image, such as detecting individual objects or digits. However, this approach assumes that bounding boxes or annotations are available to guide the model in distinguishing each object. Since the dataset in this case does not include any bounding box annotations or explicit digit boundaries, a segmentation model would struggle to effectively separate and classify the digits. Without proper annotations, the model would not know how to localize the digits within the image, making it difficult to perform meaningful segmentation. I had initially planned on performing segmentation before doing ordinary single label classification on the digits. I would have compared the sum of the predicted digits with the labels to find the accuracy of the model. But due to reasons mentioned above, I had to give up on the idea.

$$\text{Segmentation problem} \quad \longrightarrow \quad \text{Requires bounding boxes or pixel-wise annotations.}$$

2. Multi-label Classifier

A multi-label classifier assigns multiple labels to an image, which might seem suitable for classifying multiple digits. However, this approach fails in cases where the digits may repeat within an image. In our dataset, it's possible to have repeated digits (e.g., two instances of the digit "3"), which would result in the same label being assigned more than once in a multi-label classification. Since the task requires predicting the sum of digits, and repeated digits can affect the sum, a multi-label classifier would not be able to handle the combinatorial nature of repeated digits, making it unsuitable for accurately solving the problem. Mathematically, if the sum of digits $\{d_1, d_2, \ldots, d_n\}$ is required, repeated digits would create ambiguity:

$$\text{Sum} = \sum_{i=1}^{n} d_i \quad \text{where} \quad d_i = \{1, 1, 3, 5\}, \quad \text{but multi-labels can misrepresent repeated digits.}$$

ResNet

ResNet models are particularly effective for this task due to their ability to transfer learned features, handle complex patterns, and directly classify the sum of digits.

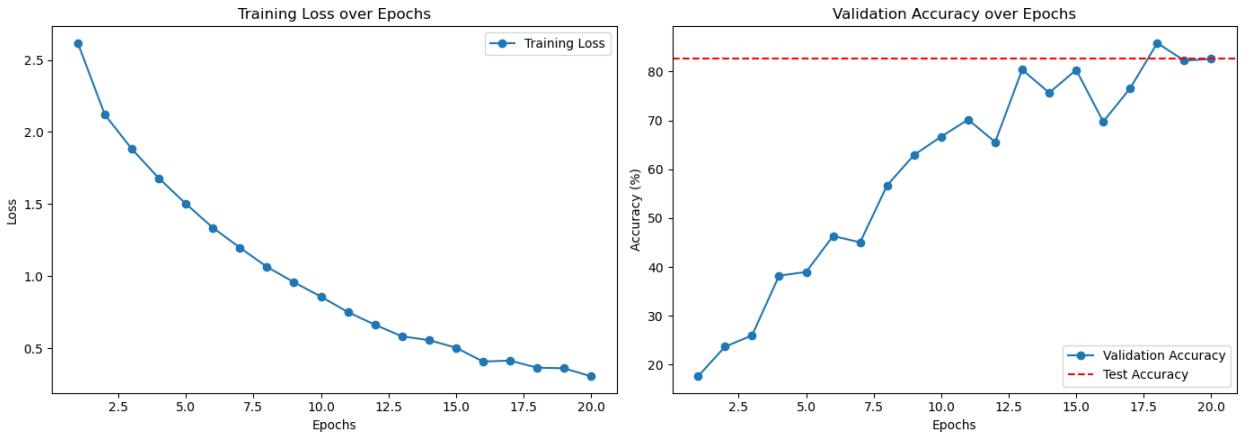

Figure 1: Perfoemance of the ResNet Model on available datasets

1. Transfer Learning

Pretrained ResNet models, trained on large datasets like ImageNet, can leverage learned features such as edges and textures. This allows them to recognize the structure of MNIST digits even without annotations or bounding boxes:

$$F = \mathcal{F}(X)$$

where F represents the feature map learned from the input image X.

2. Handling Complex Patterns

ResNet's deep architecture with residual connections captures complex patterns in images. The residual blocks ensure the model can learn intricate representations while avoiding training issues in deep networks:

$$\mathbf{y} = \mathcal{F}(x, \{W_i\}) + x$$

where $\mathbf{y}$ is the output of the residual block.

3. Direct Sum Classification

ResNet directly predicts the sum of digits S in an image, avoiding the complexity of segmentation or multi-label classification. This global classification task is simpler and more efficient for predicting the sum.