COL 774 Assignment - 3 Report

Aman Hassan, Purna Chandra Sekhar V
2021CS50607, 2021MT60270

October 2023

1 Non-Competitive Part

This section aims to use a CNN-based encoder and an LSTM-based decoder model to convert images into latex code. The Model is trained using 50% teacher forcing ratio for better learning. The BLEU scores on the data compute the performance of the model.

1.1 CNN-LSTM Model trained on Synthetic Data

Firstly, we aim to train the Model purely on the synthetic data and compute the BLEU scores as given in the problem statement. The model training statistics are as follows:

1. No of Epochs trained = 25
2. Time Taken = 6.5 hrs

The BLEU Scores are as follows:

- BLEU Scores on the Validation set of Handwritten data set = 0.0912
- BLEU Scores on the Validation set of Synthetic data set = 0.1731
- BLEU Scores on the Test set of Synthetic data set = 0.1682

1.2 CNN-LSTM Model trained on Handwritten Data

Firstly, we train the Model purely on the synthetic data and further fine-tune it by training it on the Handwritten data. The model training statistics are as follows:

1. No of Epochs trained = 100
2. Time Taken = 2.5 hrs

The BLEU Scores are as follows:

- BLEU Scores on the Validation set of Handwritten data set = 0.1392
- BLEU Scores on the Validation set of Synthetic data set = 0.1483
- BLEU Scores on the Test set of Synthetic data set = 0.1327

Note: The reported BLEU scores are tested on the given BLEU score script on Kaggle as dated November 27, 2023, at 11:32 PM. Any future updates to the BLEU script (if any) were not noticed due to the time crunch experienced while working on other parts of the assignment.
Observations: It can be observed that the score of the Handwritten data set increases as we train on the handwritten data set. However, the scores corresponding to the synthetic data set slightly experienced a drop.

2 Competitive Part

- We explored the Nougat Model by Facebook AI, an OCR model specifically trained for image-to-latex conversion.
- We also explored the Tr-OCR model by Microsoft, which is a transformer-based OCR Model. We have tried fine-tuning it from a pre-trained OCR model using the data set provided in the assignment. Due to a larger set of parameters in this model, it was difficult to train them on Google Colab or Kaggle.
- However, we decided to use a version of the pre-trained Donut model, a transformer-based model used for Document Understanding.
- This Model has a transformer-based image encoder and a decoder. The decoder takes the encoded image and the true labels as the input and learns the parameters accordingly.
- This Model does not use OCR engines, making it computationally efficient and significantly flexible in terms of various applications.

Training statistics:

1. No of epochs: 5
2. Time Taken: 7 hrs