COL774 Assignment 4

Devang Garg & Aman Singh Dalawat

November 28, 2023

1 Image to LaTeX Converter

1.1 Approach

- In our approach, we employed an Encoder-Decoder architecture to address the given problem. The Encoder is implemented as a Convolutional Neural Network (CNN) with five layers and Rectified Linear Unit (ReLU) activation functions, following the suggested architecture in the assignment PDF. The output of the Encoder is a context vector, representing the flattened kernels obtained during the encoding process.

- On the other hand, the Decoder is designed as a Long Short-Term Memory (LSTM) network, consisting of an embedding layer, a single hidden layer, and an output layer with a size corresponding to the vocabulary. The Decoder generates predictions for each timestep, producing the predicted token at that specific point in the sequence.

- The vocabulary for the model is constructed from the Synthetic Dataset's train.csv file, with separate entries for the start and end tokens. During training, we employed Cross Entropy as the loss function for the decoder and the encoder.

- In training, the teacher forcing technique is utilized 50% of the time. This involves concatenating the context vector with the embedding of the ground truth label from the previous timestep. Alternatively, we use the learned embedding of the output from the previous timestep, concatenated with the context vector, as the input for the current timestep.

- During the prediction phase, the model generates LaTeX sequences using the decoder for a given context vector obtained from the forward propagation in the encoder. The generation process stops either when the maximum length is reached or when the *end* token is predicted. The output is a list of generated sequences.

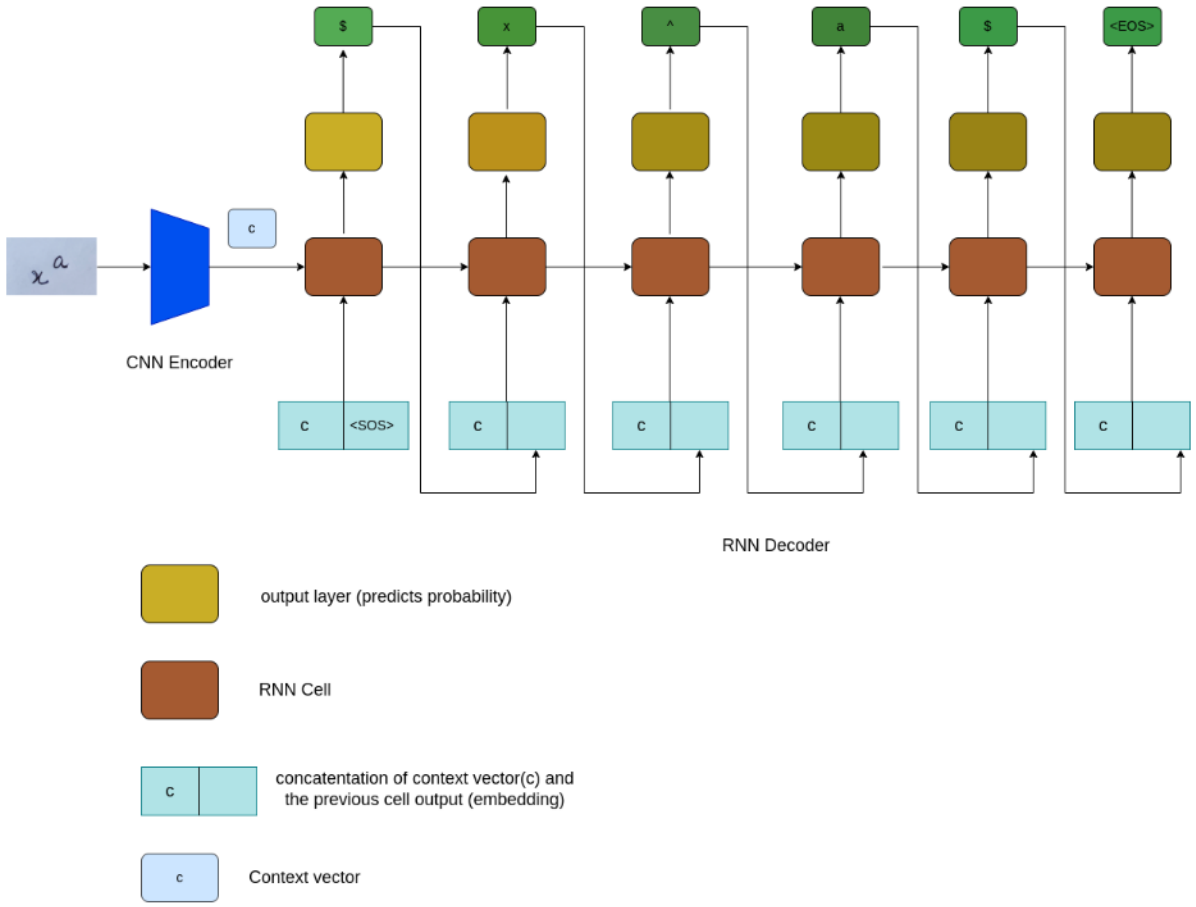

Figure 1: Model architecture

1.2 BLEU Scores

1.2.1 Part 1a

- Score on Test Data of Synthetic Data: 0.0312
- Score on Validation set of Handwritten Data: 0.0049

1.2.2 Part 1b

- Score on Test Data of Synthetic Data: 0.019
- Score on Validation set of Handwritten Data: 0.002

1.2.3 Competitive Part

- Score on Sample Sub csv: 0.0312

The Learning Rate used was 0.001 in all parts for 10 epochs.