Intro Deep Learning Homework 4

Jaskin Kabir
Student Id: 801186717
GitHub:
https://github.com/jaskinkabir/Intro_Deep_Learning/tree/master/HM4

March 2025

1 Model Architectures

The problem is to develop a sequence to sequence machine translation solution for translating English sentences to French and vice versa. To achieve this, GRU-based encoder-decoder models were tested with and without Bahdanau attention. As shown in Table 1 below, attention added 39% more parameters to each model.

	No Attention	**With Attention**	**% Difference**
English → French	13418773	18664726	39
French → English	13391098	18637051	39

Table 1: Model Parameter Counts

2 Training

The models were trained on the English-French dataset provided by the assignment, and the qualitative validation dataset was generated by ChatGPT using the same dataset. The models were trained over 100 epochs with a batch size of 8. Teacher forcing was used with an initial ratio of 0.6 which gradually decreased to 0 over the training process. The plots of the training and validation losses can be seen in Figure 1 below. Additionally, the training time for each model is reported in Table 2.
The models with attention in both cases took more than 50% longer to train, and this difference was more pronounced in the French to English translation tasks. In all cases, the model was able to fully memorize the training data, with the training loss plummeting to almost zero by the end of training. However, the validation loss began to diverge after around 25 epochs in all cases. Attention made these curves smoother and decreased the rate of divergence of the validation loss, but did not change the overall behavior of the loss curves.

	No Attention	**With Attention**	**% Difference**
English → French	15.9	24.1	52
French → English	15.1	26.4	75

Table 2: Model Training Times

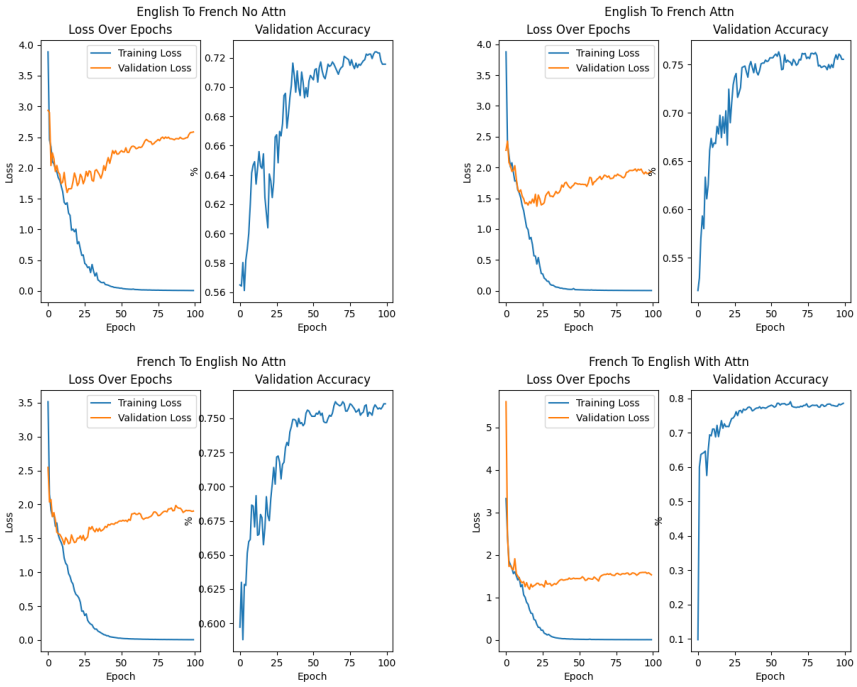

Figure 1: Training Curves

3 Results

The models were evaluated on the test dataset in terms of accuracy on the sentence level and the word level. A sentence was considered correct if all the words in the sentence were translated correctly. A word was considered correct if the model output the exact word in the target language. A comparison in accuracy between the models with and without attention can be seen in Table 3 below.

The models with attention outperformed the models without attention in both sentence and word level accuracy. Without attention, both translation tasks performed identically in sentence accuracy, whereas translating from French to English was slightly more accurate at the word level. This is mirrored when attention is added to the models.

However, the English to French translation task benefitted much more from attention, and this is reflected by this task showing twice the sentence level accuracy as the French to English translation task.

While the attention models performed well, all models still suffered from overfiting in the form of complete memorization of the training data. This is reflected in the sentences that the best performing model, the English to

French Translator with attention, produced. For example, the model correctly translated the sentences: "He cooks dinner" and "He sings a song" because these sentences are in the training data but with 'He' replaced with 'She'. The model can only correctly translate sentences that are very similar to those found in the training set.
This is a limitation of the encoder-decoder architecture. Even with Badhanau attention, the model only uses one head of attention, and has no mechanism to allow the output tokens to update each other with context information. Additionally, there is no positional encoding present in the model. These issues compound the model's inability to understand the contextual meaning of each word in the input sequence and the output sequence. Finally, the training dataset only contains around 100 samples, which is not enough to train a model that can generalize well to unseen data.

English → French			
	No Attention	**With Attention**	**% Difference**
Word Accuracy (%)	71.56	76.83	7
Sentence Accuracy (%)	1.83	12.84	602
French → English			
	No Attention	**With Attention**	**% Difference**
Word Accuracy (%)	76.07	78.59	3
Sentence Accuracy (%)	1.83	6.42	251

Table 3: Accuracy Results