Task 3 - Report

André Bastos, nº 56969
Rafaela Cruz, nº 56926

May 31, 2020

The loss function that we used was `tf.keras.losses.CategoricalCrossentropy` with logits from Keras, as we send the one hot encoding version of the predictions and the ground truth. The accuracy function was one from Keras as well: `tf.keras.metrics.Accuracy`.

We used 6 epochs to train our model, obtaining, for the training set, a loss of $\simeq 0.127$ and an accuracy of $\simeq 96\%$. For the validation set, we achieved a a loss of $\simeq 1.984$ and an accuracy of 56%, using teacher forcing. Lastly, for the test set, we achieved a loss of $\simeq 2.325$ and an accuracy of $\simeq 22\%$ without teacher forcing.

Taking into account the results, it is clear that the model is overfitting, as the validation and test loss and accuracy are much different from the training. Therefore, adding more epochs would be irrelevant. This could probably be solved by performing a grid search for hyperparameter optimization. These parameters could be the number of units in each layer (in the attention and in the GRU), the number of RNNs used in the decoder, changing the GRUs to LSTMs networks or even changing the whole image shape of the encoder outputs.