

AI51701/CSE71001 Assignment 3

Submission Instructions: You shall submit this assignment on BlackBoard as two submission files – your code as assignment3coding.zip and write up for assignment3.pdf; Run the collect_submission.sh script to produce your assignment3.zip file. This time, we use Jupyter Notebook and Google Colab. It is your responsibility to make sure your code is runnable at Google Colab. If your code is not runnable, you will get no point. Also do not write down your name/student ID in your submission.

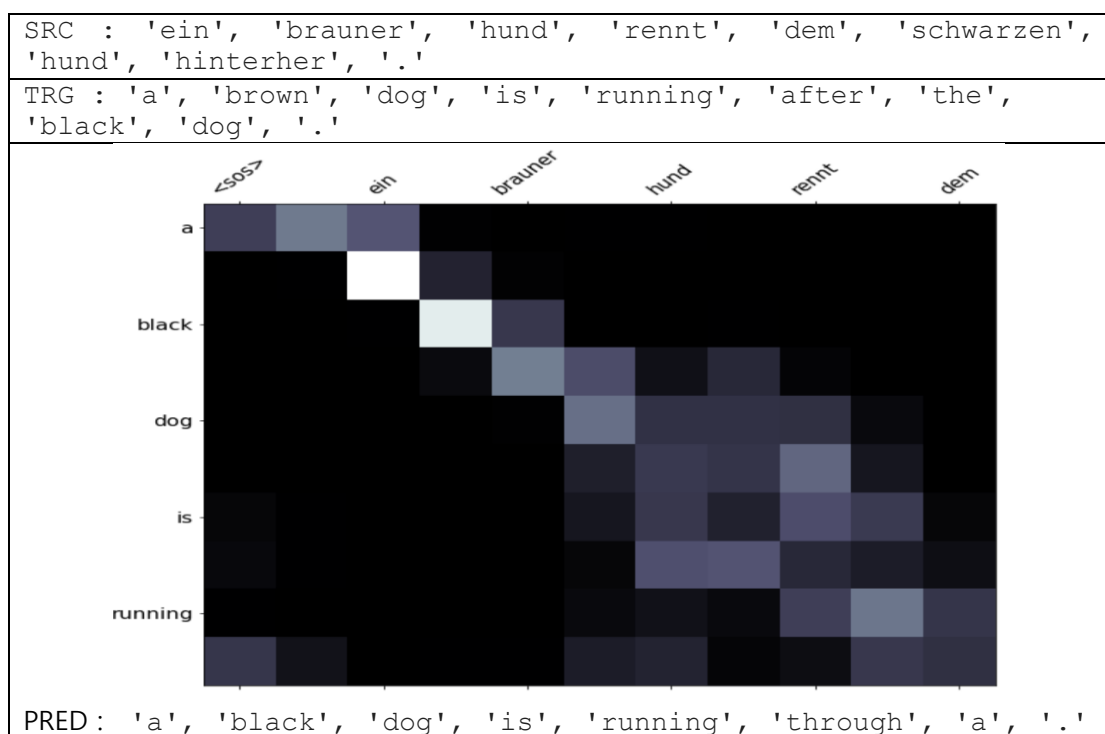
1 Implementing sequence-to-sequence model with attention (100 points)

In this problem, we implement sequence-to-sequence model with attention as learned in our class to build a Neural Machine Translation (NMT) system. We provide jupyter notebook file (asn3_1_seq2-seq_attention.ipynb) for implementing the model. For decode and encoder models, we use unidirectional LSTMs. For computing attention weights, we use the basic dot-product attention that we also learned in our class.

- (a) Implement encoder, decoder, and training and inference code as instructed in comments. Report your translation results for a few examples in the validation set and attention visualization of them. Report BLUE score for the test set (80 points)

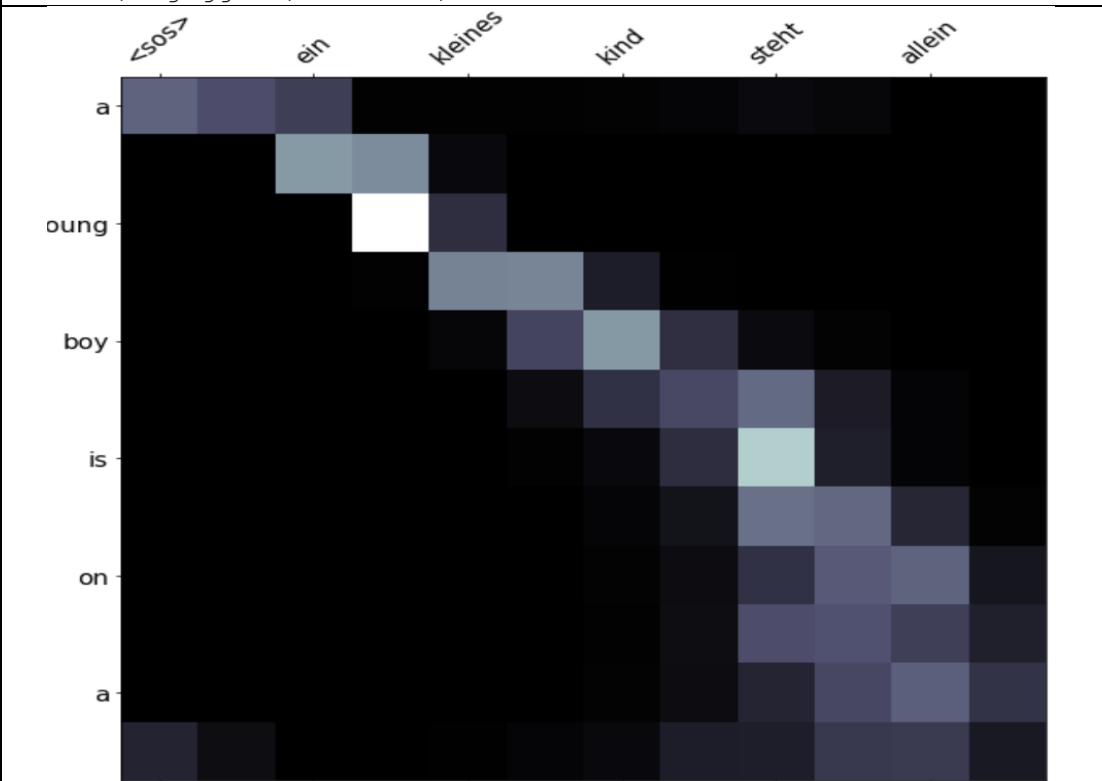
Ans:

Here is my model's visualization for output.



SRC : 'ein', 'kleines', 'kind', 'steht', 'allein', 'auf',
'einem', 'zerklüfteten', 'felsen', '.'

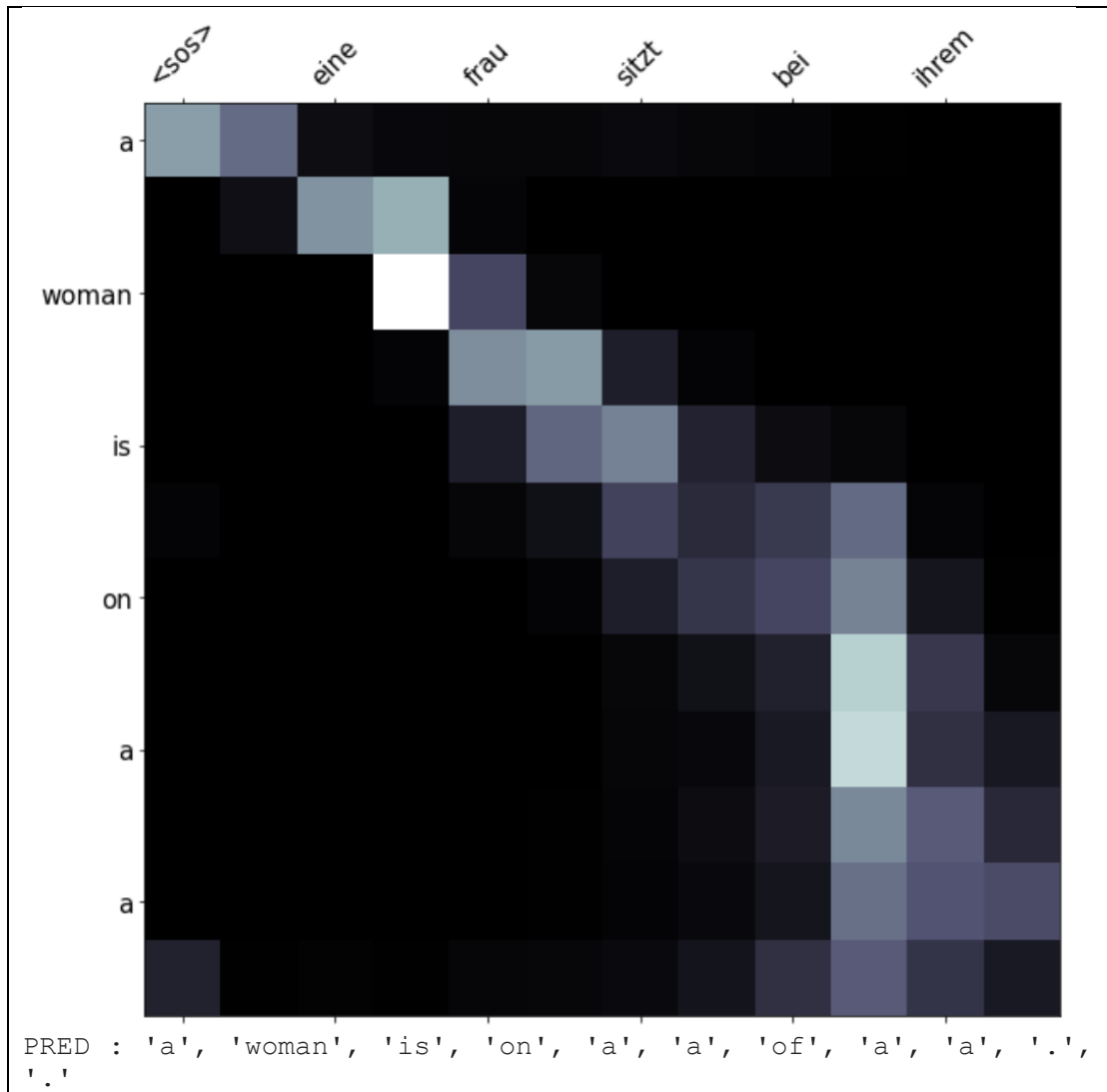
TRG : 'a', 'young', 'child', 'is', 'standing', 'alone', 'on',
'some', 'jagged', 'rocks', '.'



PRED: 'a', 'young', 'boy', 'is', 'on', 'a', 'a', 'of', 'a', '.',
'.'

SRC : 'eine', 'frau', 'sitzt', 'bei', 'ihrem',
'trockenblumensortiment', 'auf', 'einem', 'freiluftmarkt', '.'

TRG : 'a', 'woman', 'is', 'sitting', 'by', 'her', 'dried',
'flower', 'display', 'at', 'an', 'outside', 'market', '.'



BLEU score is 9.01. Using unidirectional LSTM, Because of using unidirectional lstm, it results low performance.

- (b) Tune your model and code to get the best performance you can get, and report BLUE score for the test set. Explain why your setting achieves this (10 points). Note that training of your model should be done in reasonable time at Google Colab (10 points).

Ans:

I replace LSTM unit to GRU. GRU and LSTM are similar, but GRU is more light unit than LSTM. So I want to see how they show different performance. But BLUE score is 4.08 because, LSTM has large memory than GRU.

- (c) Implement bidirectional LSTMs for the encoder and report BLEU score for the test set. What is difference in BLUE score compared to the encoder with unidirectional LSTMs and explain

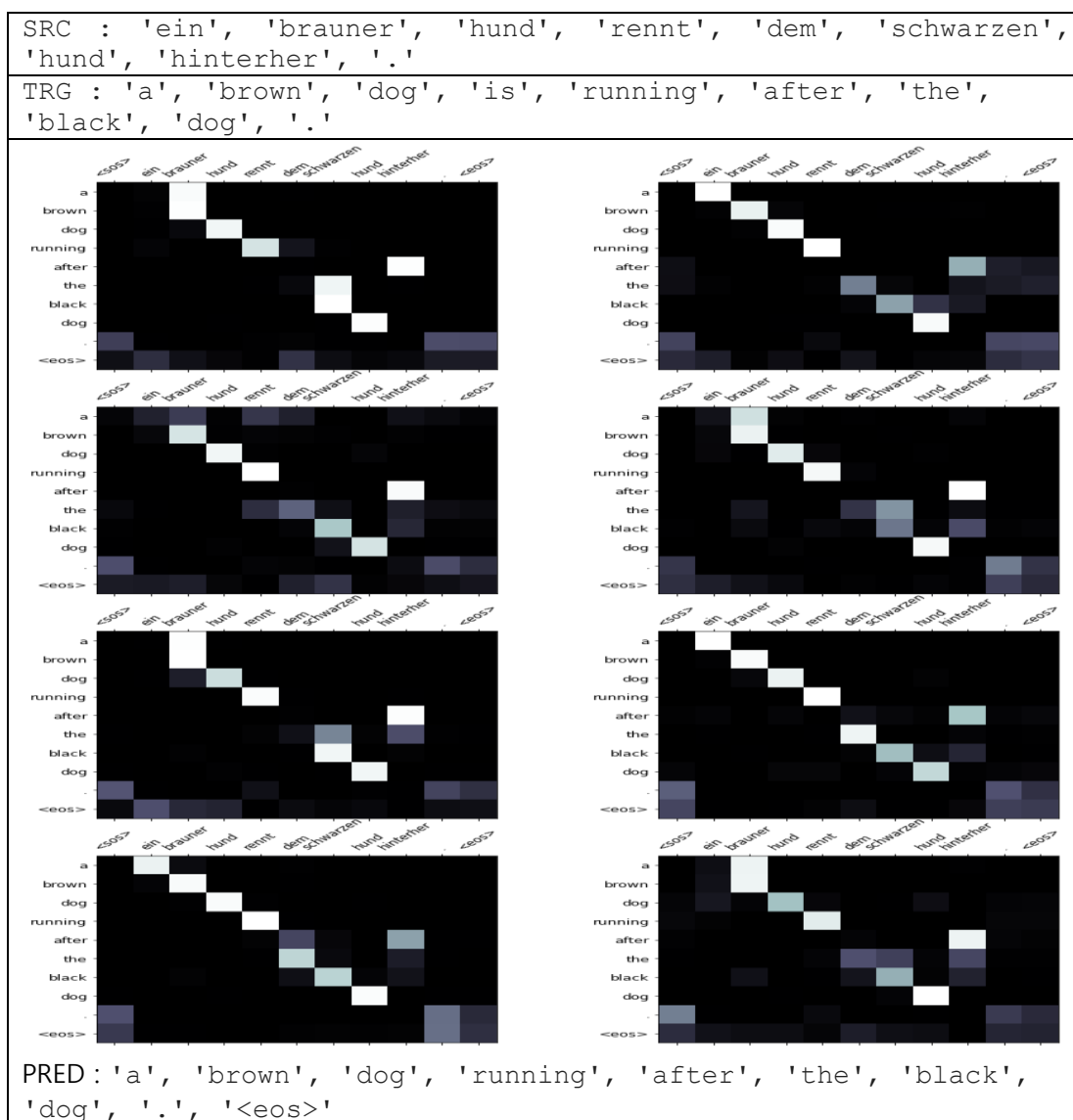
why (10points).

Ans: The Bleu score is 12.77. The model which use bidirectional LSTM is more better than (a)'s model. Using Bidirectional LSTM refers bi-directional side unless normal LSTM. So this is helpful to improve model's performance.

2 Implementing Transformer model (70 points)

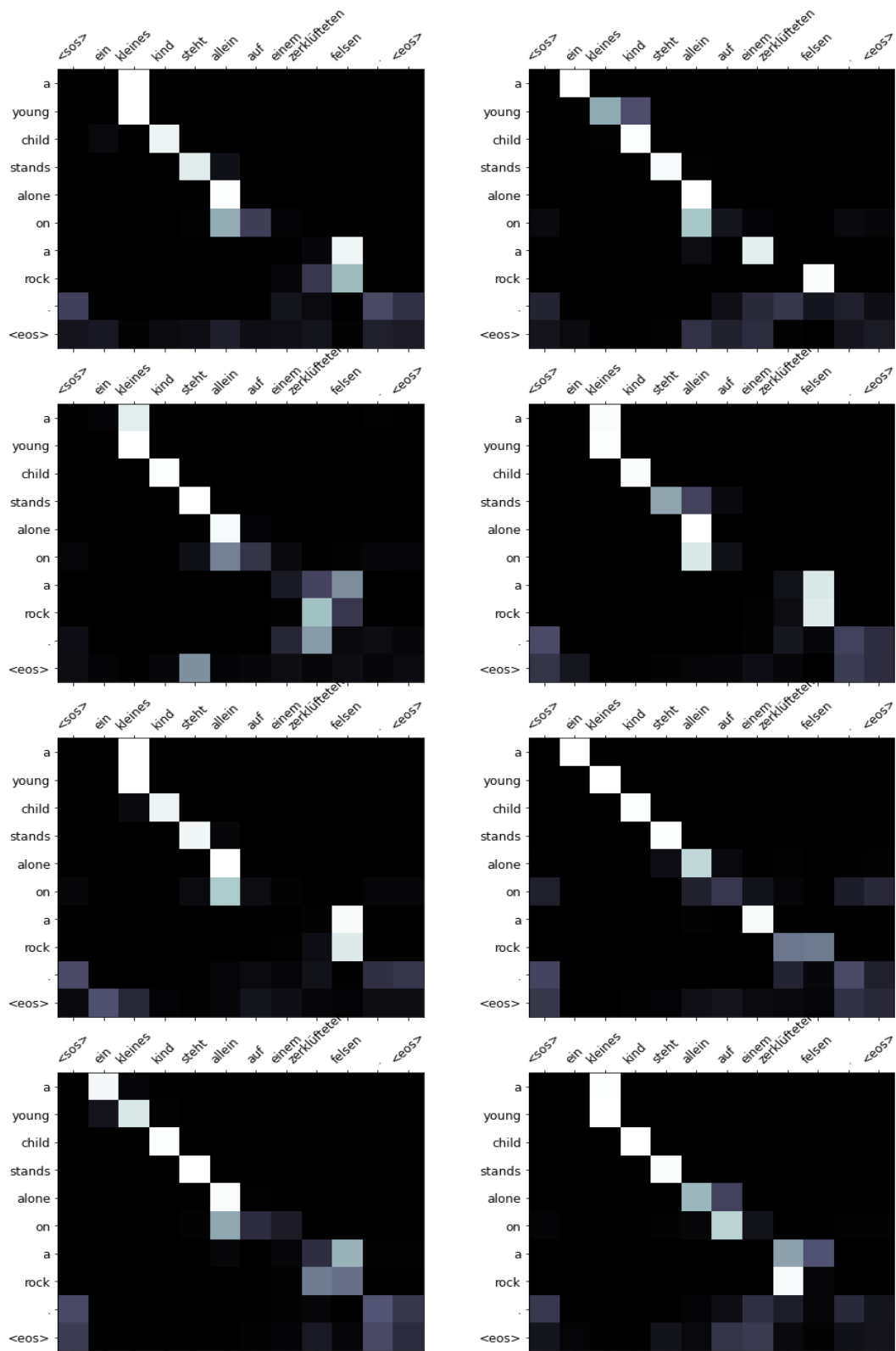
(a) We provide jupyter notebook file (assn3_2_transformer.ipynb) for implementing the model. Implement encoder, decoder, and training and inference code as instructed in comments. Note that, for positional encoding, we use positional embedding as in BERT unlike the positional encoding in the original Transformer paper. Report your translation results for a few examples in the validation set and attention visualization of them. Report BLUE score for the test set (60 points)

Ans: Here is visualization of transformer result.

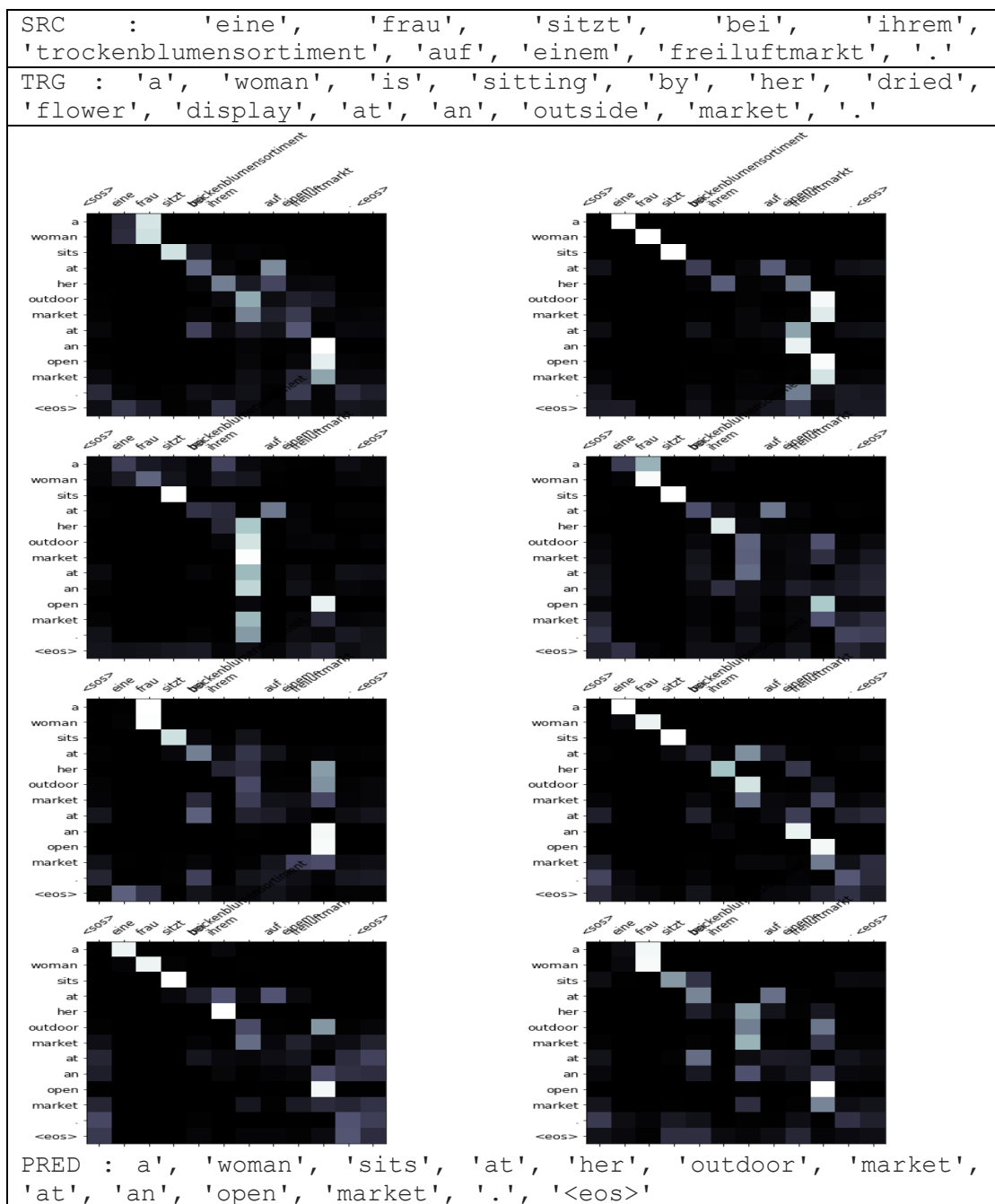


SRC : 'ein', 'kleines', 'kind', 'steht', 'allein', 'auf',
'einem', 'zerklüfteten', 'felsen', '.'

TRG : 'a', 'young', 'child', 'is', 'standing', 'alone', 'on',
'some', 'jagged', 'rocks', '.'



PRED : 'a', 'young', 'child', 'stands', 'alone', 'on', 'a',
'rock', '.', '<eos>'



BLEU Score is 35.91

(b) Tune your model and code to get the best performance you can get, and report BLUE score for the test set. Explain why your setting achieves this. Note that training of your model should be done in reasonable time at Google Colab (10 points).

Ans: To reinforce attention energy, I apply sigmoid and tanh. I combined them and add previous energy. Bleu score is 32.27.