

Image Caption Models

Mandatory assignment 2

Bendik Nyheim, 15500

IN5400 - Machine learning for Image Analysis

University of Oslo

Norway

March 2022

Contents

1	How to run	2
2	Hyperparameters	3
3	Results	4

1 How to run

The models for task 1-4 will be trained by running the scripts

"Exercise_Train_an_image_captioning_network_task*.py". It will train the model corresponding to the task. The path to the root of the dataset need to be set at line 40 in these scripts.

The folder "storedModels__test" contains a pretrained model for task4, that is a two layer LSTM with attention, which is trained for 80 epochs. Running the "validation_only.py" script will test this model on the validation set.

A small note: I made the mistake of training the models without setting the seed, which means that for the pretrained model from task 4, the reported best measures is not necessarily the best measure for the seed I set later. I hope you let this slide as the only thing needed to fix this is retrain the models using a seed, which I feel is a waste of resources as everything else is done properly. The scores reported for task 4 in table 2 shows the scores from epoch 70, and you should get the same results.

2 Hyperparameters

Table 1: Hyperparameters used in training and validation

Parameter	Value
Optimizer	<i>AdamW</i>
Learning rate η	0.001
Training batch size	128
Validation batch size	64
Weight decay	0.00001
Number of epochs task 1-3	50
Number of epochs task 4	80
Number of CNN features	2048
Hidden state size	512
Embedding size	300
Vocabulary size	10000
Scheduler factor	0.2
Truncated backpropagation length	25
Momentum γ	0.9
Torch seed	321
Numpy seed	321

3 Results

Figure 1 shows the train/validation loss while training the model, with the corresponding meteor score on the validation set.

Table 2 shows the Meteor and BLEU-4 score for all the models at the epoch with the highest Meteor score on validation set.

Figure 2 shows five images with the corresponding predicted caption as title. The predictions were made using the two layer LSTM with attention, which was trained for 80 epochs.

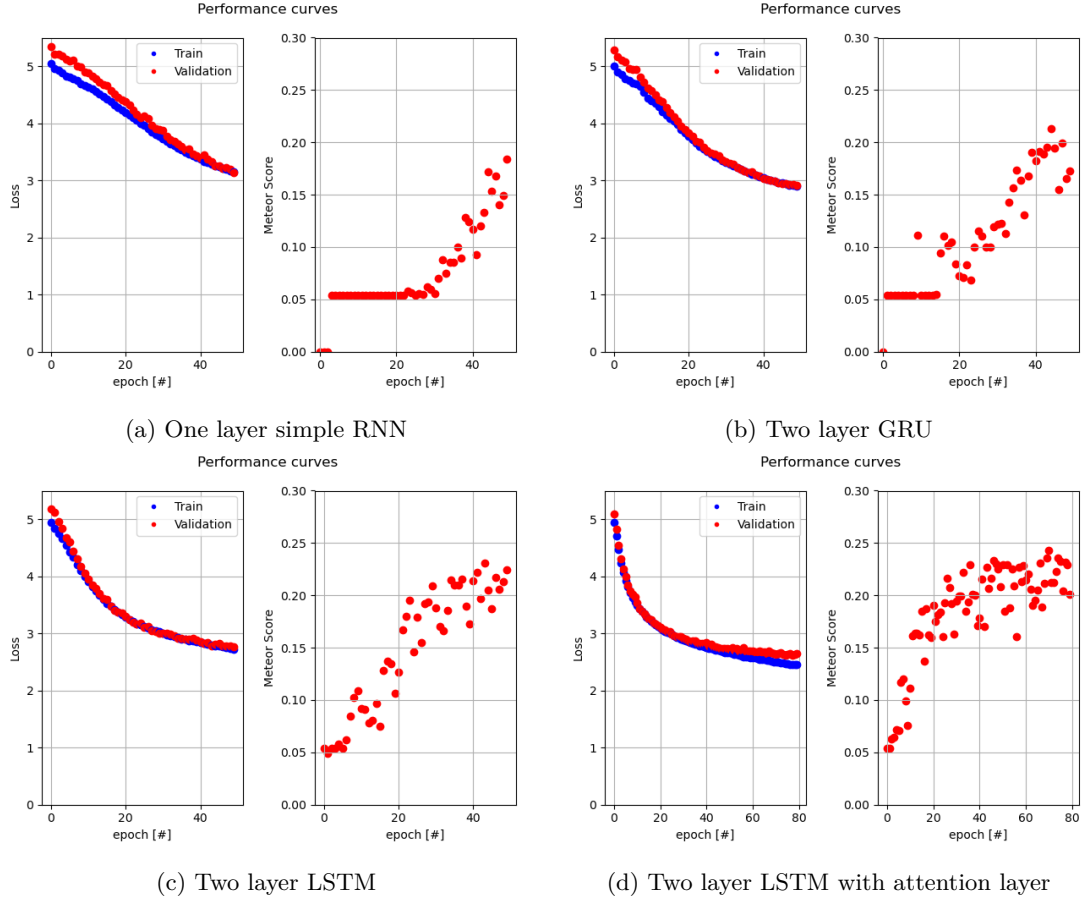


Figure 1: Loss and performance metrics for every epoch while training.

Table 2: Meteor and BLEU-4 score on validation set at epoch with highest Meteor score. All the models trained for 50 epochs, except for the two layer LSTM with attention, which trained for 80 epochs.

Model	Epoch	Meteor	BLEU-4
One layer RNN	49	0.1841	0.1740
Two layer GRU	44	0.2133	0.2237
Two layer LSTM	43	0.2307	0.2598
Two layer LSTM with attention	70	0.2394	0.2743



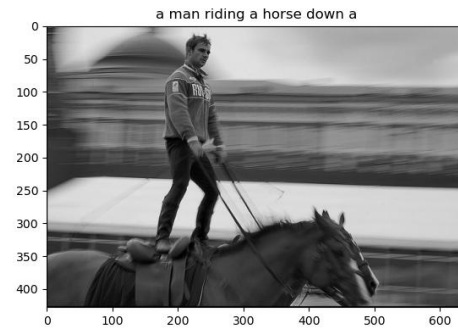
(a)



(b)



(c)



(d)



(e)

Figure 2: Predicted captions from the two layer LSTM with attention model.