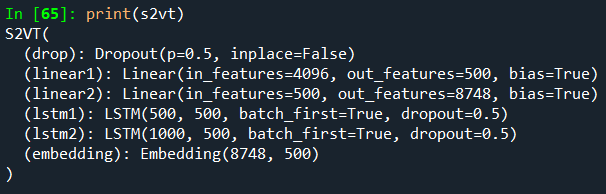
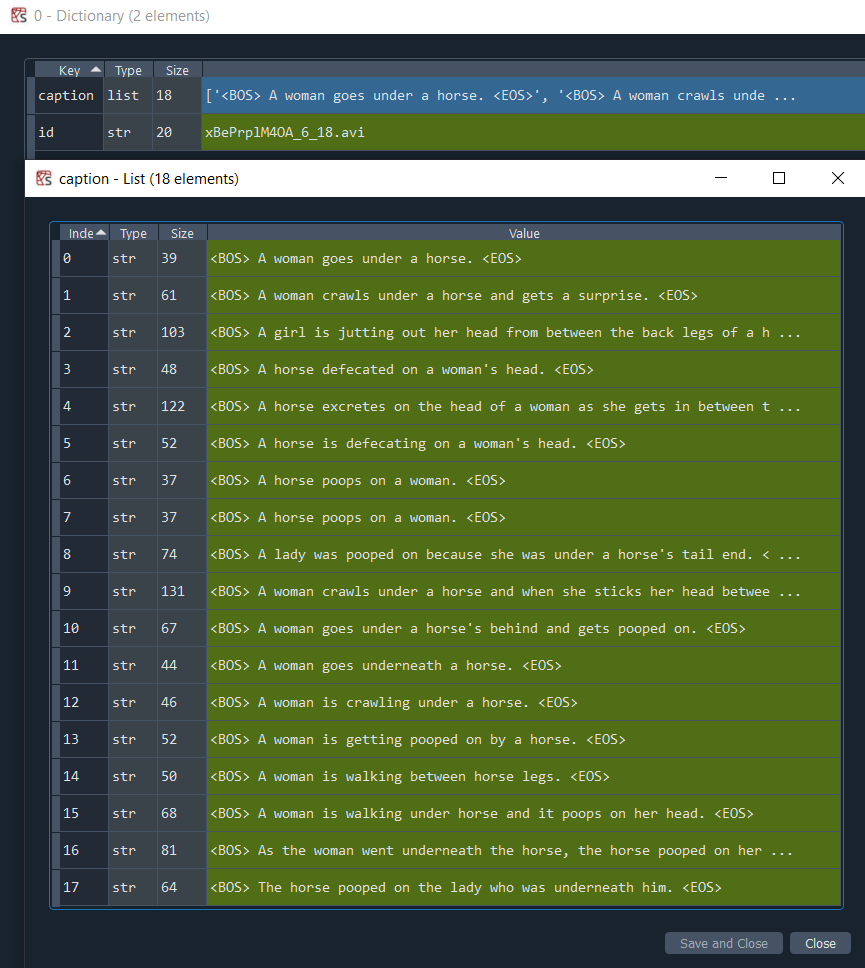
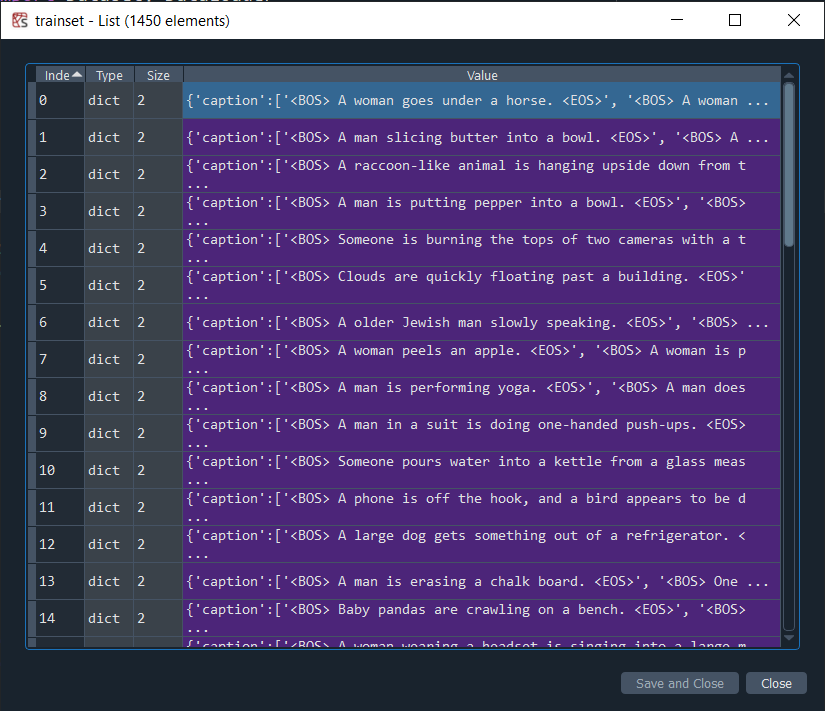
**HW2 Video caption generation:**

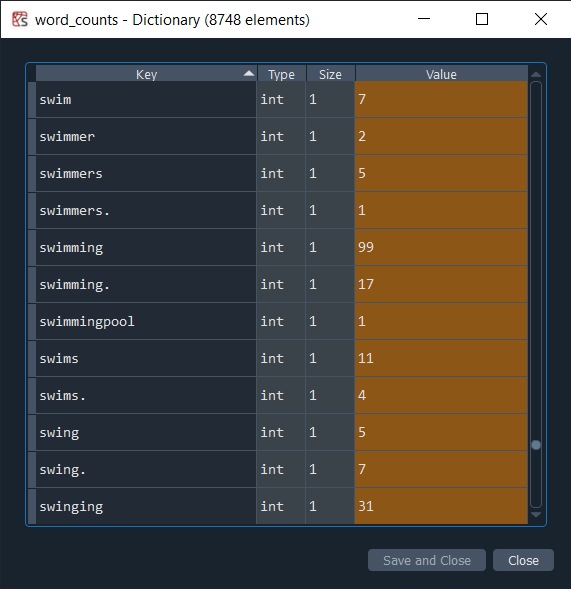
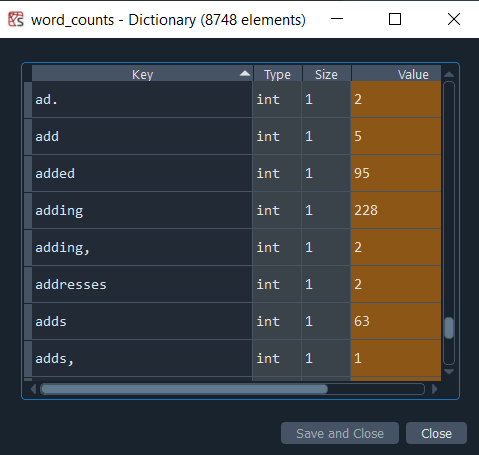
1. Model summary:



1. Data prepare:



1. Build vocabulary:

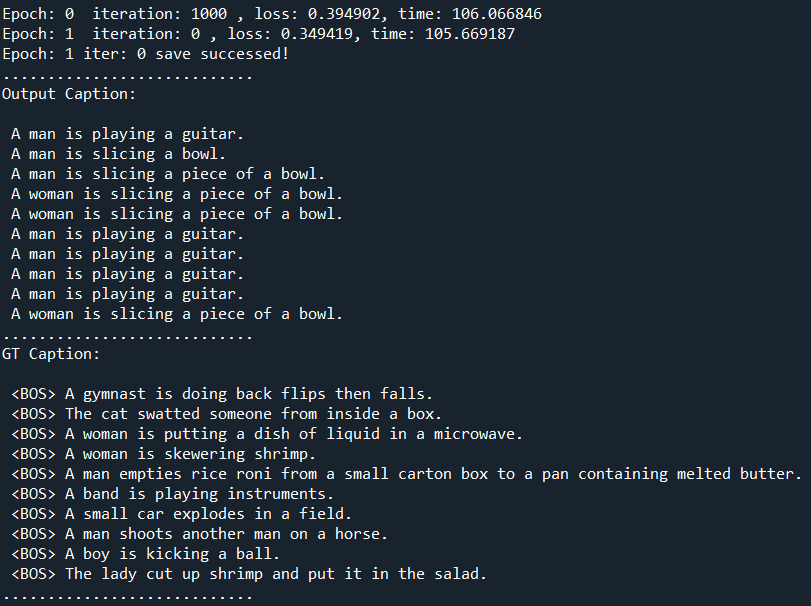


1. Training setup:

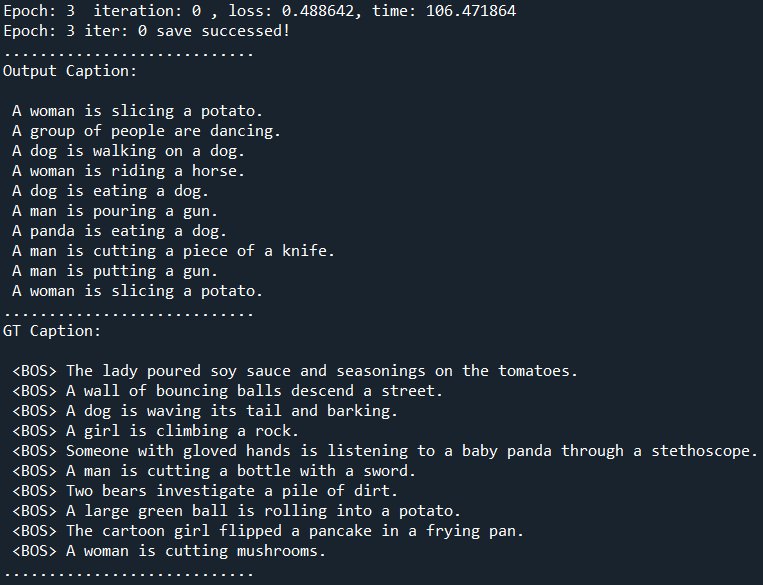
* Dataset = MSVD (1450/100, training/testing)
* Epochs = 30
* iteration = 2000
* Batch size = 10
* Learing rate= 0.0001

1. Training process tracking:

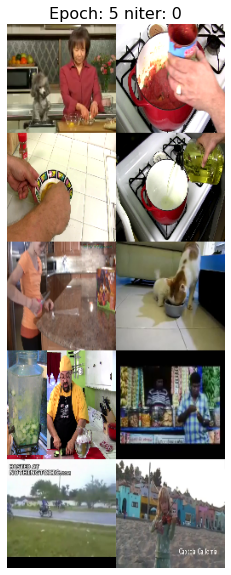
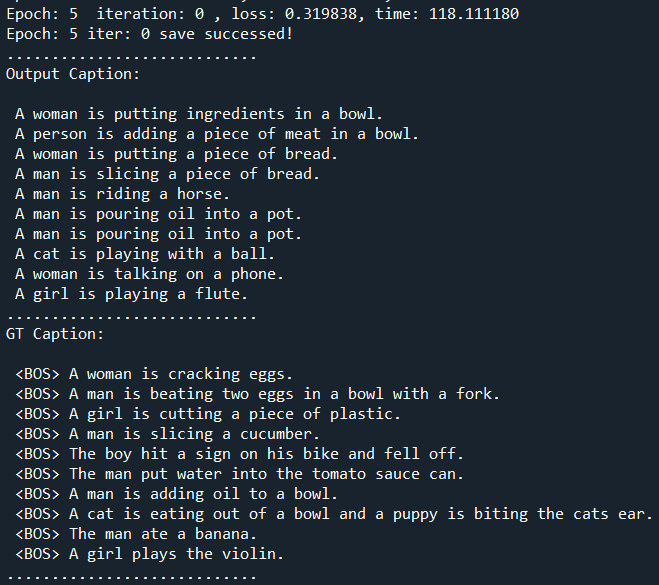
* Epoch 1: Training just started, weights are initializing, results tend to randomness.



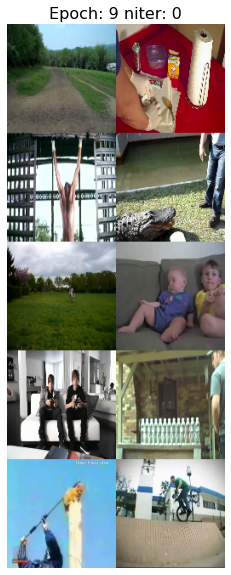
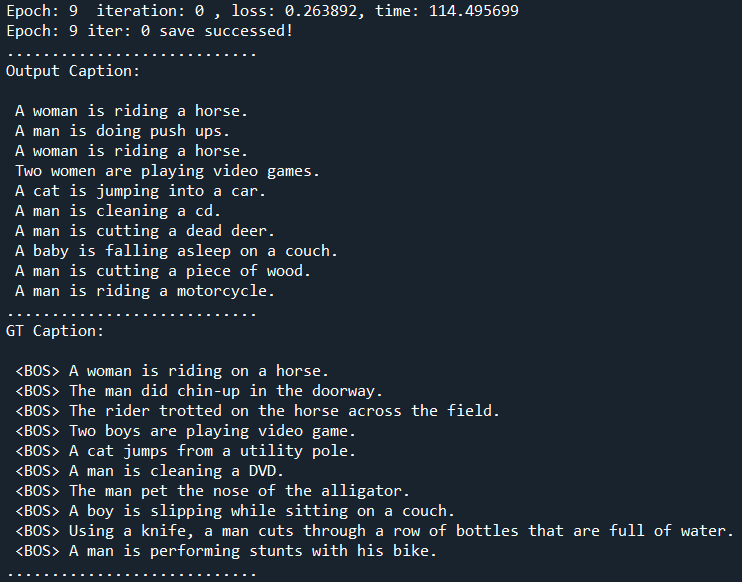
* Epoch 2: The model started to generate some right sentences in the correct order.



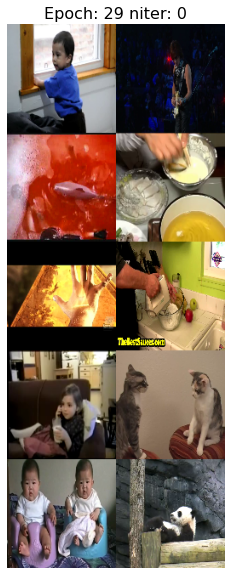
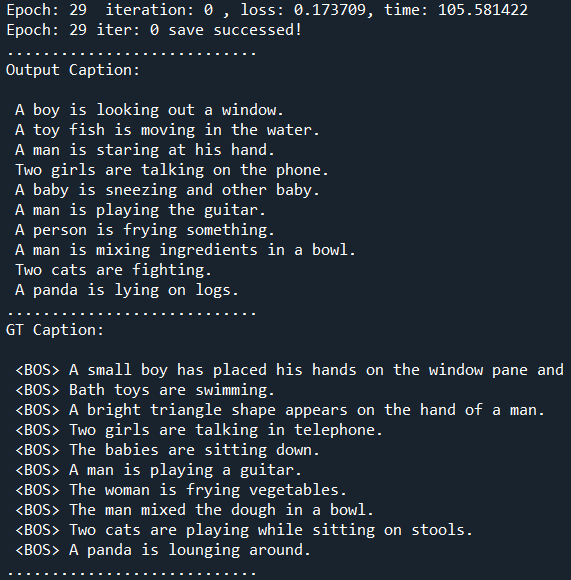
* Epoch 5: Output accuracy improves.



* Epoch 9: Loss decreased, the result is getting better.

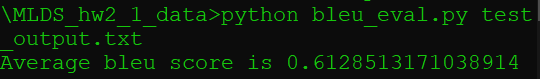


* Epoch 29: Training done, the result is very close to the ground truth.



1. Bleu evaluation:

* Score: 0.61 > baseline = 0.6



1. Summary:

* Total time cost: 6340 seconds ~ 1.8 hours
* Loss plot:

