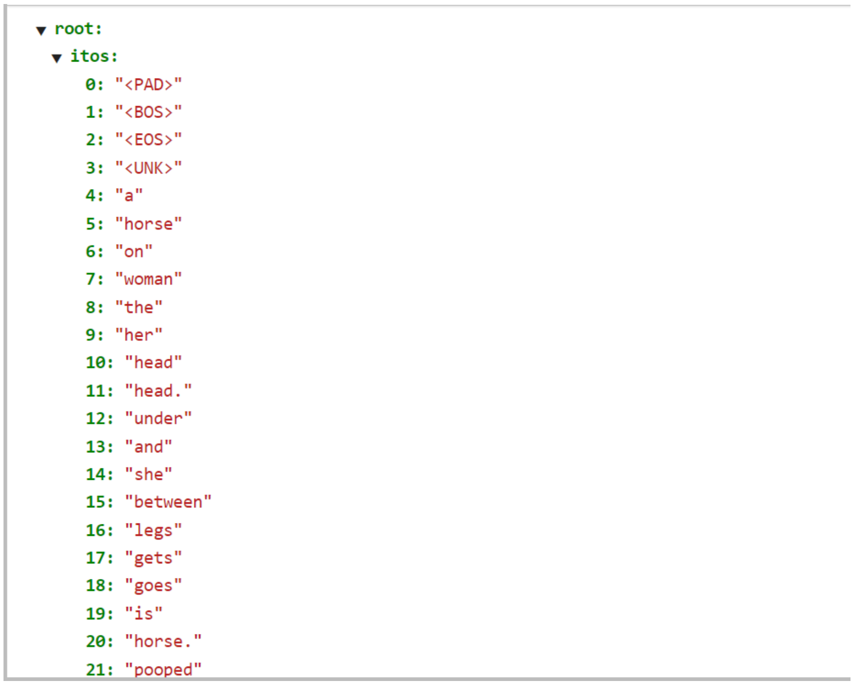
Homework 2 – Updated!

Joshua Lumpkin

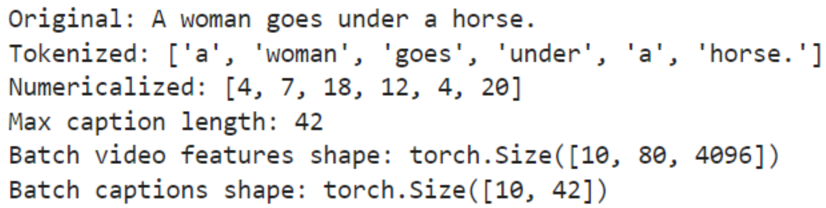
Vocabulary:

Saved as a JSON, vocab size of 3528



```
▼ root:
  ▼ itos:
      0: "<PAD>"
      1: "<BOS>"
      2: "<EOS>"
      3: "<UNK>"
      4: "a"
      5: "horse"
      6: "on"
      7: "woman"
      8: "the"
      9: "her"
      10: "head"
      11: "head."
      12: "under"
      13: "and"
      14: "she"
      15: "between"
      16: "legs"
      17: "gets"
      18: "goes"
      19: "is"
      20: "horse."
      21: "pooped"
```

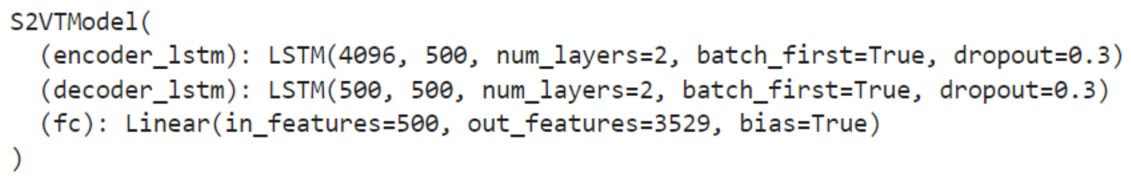
Example of preprocessing, tokenizing. I do append BOS and EOS to the caption as well, filling with PAD to the max caption length



```
Original: A woman goes under a horse.
Tokenized: ['a', 'woman', 'goes', 'under', 'a', 'horse.']
Numericalized: [4, 7, 18, 12, 4, 20]
Max caption length: 42
Batch video features shape: torch.Size([10, 80, 4096])
Batch captions shape: torch.Size([10, 42])
```

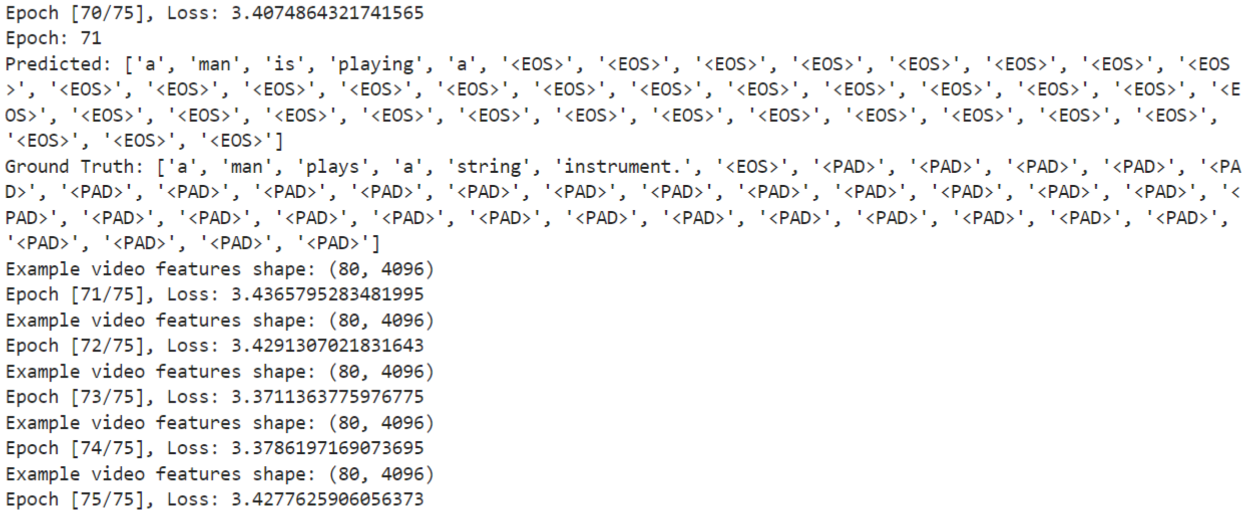
Training setup:

- Epochs = 75
- Learning rate = .0001
- Batch size = 10
- Dropout = 0.3 on both encoder and decoder



```
S2VTModel(
  (encoder_lstm): LSTM(4096, 500, num_layers=2, batch_first=True, dropout=0.3)
  (decoder_lstm): LSTM(500, 500, num_layers=2, batch_first=True, dropout=0.3)
  (fc): Linear(in_features=500, out_features=3529, bias=True)
)
```

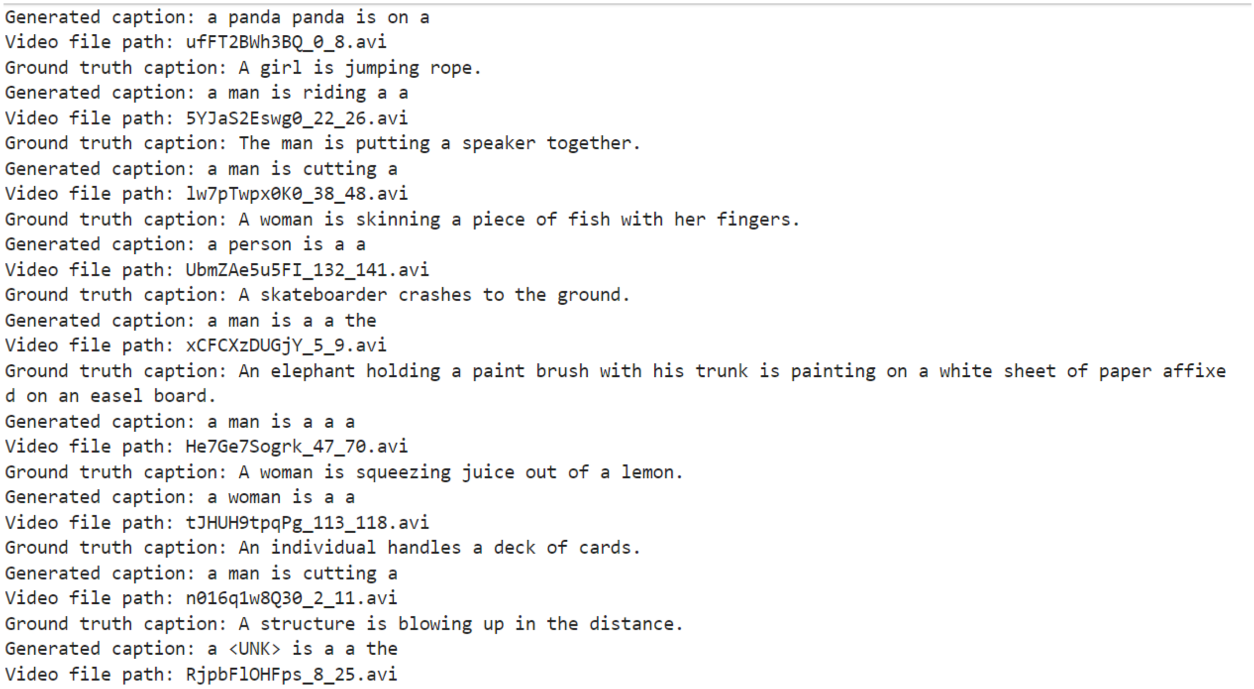
Training predictions:



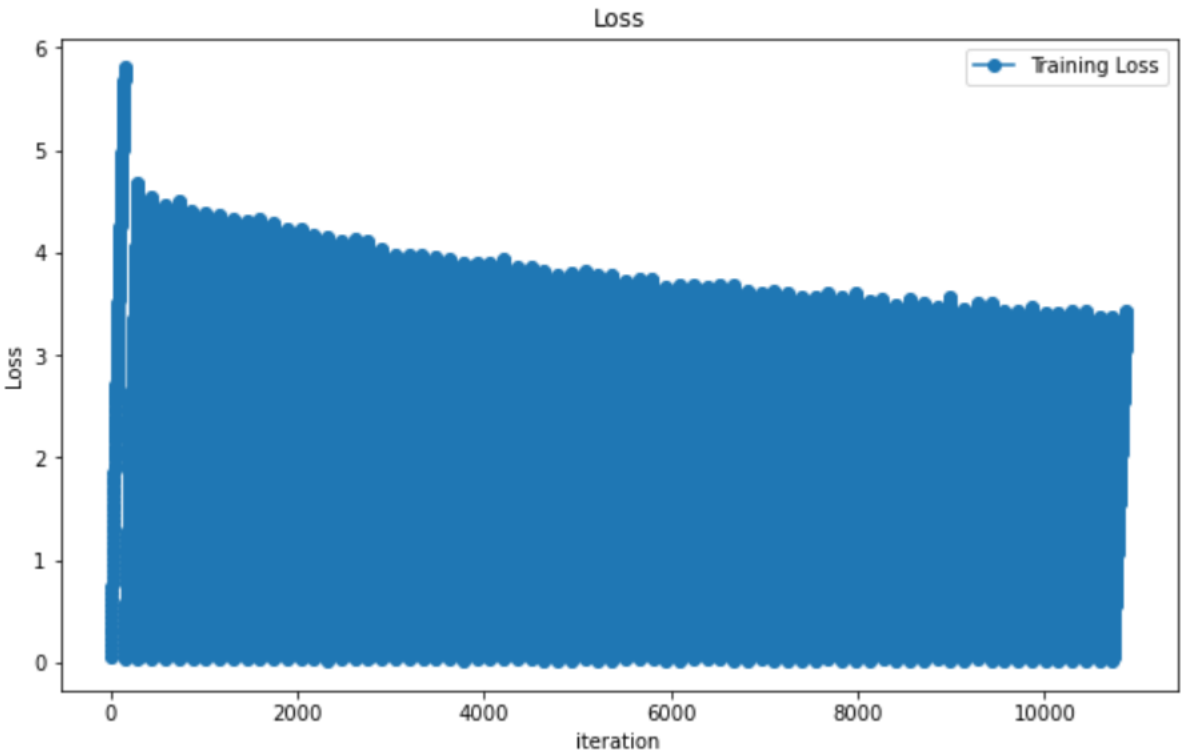
```
Epoch [70/75], Loss: 3.4074864321741565
Epoch: 71
Predicted: ['a', 'man', 'is', 'playing', 'a', '<EOS>', '<EOS>', '<EOS>', '<EOS>', '<EOS>', '<EOS>', '<EOS>', '<EOS
>', '<EOS>', '<EOS>', '<EOS>', '<EOS>', '<EOS>', '<EOS>', '<EOS>', '<EOS>', '<EOS>', '<EOS>', '<EOS>', '<EOS>', '<E
OS>', '<EOS>', '<EOS>', '<EOS>', '<EOS>', '<EOS>', '<EOS>', '<EOS>', '<EOS>', '<EOS>', '<EOS>', '<EOS>', '<EOS>',
'<EOS>', '<EOS>', '<EOS>']
Ground Truth: ['a', 'man', 'plays', 'a', 'string', 'instrument.', '<EOS>', '<PAD>', '<PAD>', '<PAD>', '<PAD>', '<PA
D>', '<PAD>', '<PAD>', '<PAD>', '<PAD>', '<PAD>', '<PAD>', '<PAD>', '<PAD>', '<PAD>', '<PAD>', '<PAD>', '<PAD>', '<
PAD>', '<PAD>', '<PAD>', '<PAD>', '<PAD>', '<PAD>', '<PAD>', '<PAD>', '<PAD>', '<PAD>', '<PAD>', '<PAD>', '<PAD>',
'<PAD>', '<PAD>', '<PAD>', '<PAD>']
Example video features shape: (80, 4096)
Epoch [71/75], Loss: 3.4365795283481995
Example video features shape: (80, 4096)
Epoch [72/75], Loss: 3.4291307021831643
Example video features shape: (80, 4096)
Epoch [73/75], Loss: 3.3711363775976775
Example video features shape: (80, 4096)
Epoch [74/75], Loss: 3.3786197169073695
Example video features shape: (80, 4096)
Epoch [75/75], Loss: 3.4277625906056373
```

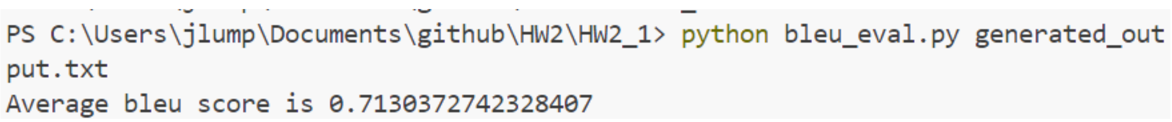
Results:

Was able to start getting more coherent output, still needs refinement, but can recognize a panda in one caption, man, vs women in some.



```
Generated caption: a panda panda is on a
Video file path: ufFT2BWh3BQ_0_8.avi
Ground truth caption: A girl is jumping rope.
Generated caption: a man is riding a a
Video file path: 5YJaS2Eswg0_22_26.avi
Ground truth caption: The man is putting a speaker together.
Generated caption: a man is cutting a
Video file path: lw7pTwpx0K0_38_48.avi
Ground truth caption: A woman is skinning a piece of fish with her fingers.
Generated caption: a person is a a
Video file path: UbmZAe5u5FI_132_141.avi
Ground truth caption: A skateboarder crashes to the ground.
Generated caption: a man is a a the
Video file path: xCFCXzDUGjY_5_9.avi
Ground truth caption: An elephant holding a paint brush with his trunk is painting on a white sheet of paper affixe
d on an easel board.
Generated caption: a man is a a a
Video file path: He7Ge7Sogrk_47_70.avi
Ground truth caption: A woman is squeezing juice out of a lemon.
Generated caption: a woman is a a
Video file path: tJHUH9tpqPg_113_118.avi
Ground truth caption: An individual handles a deck of cards.
Generated caption: a man is cutting a
Video file path: n016q1w8Q30_2_11.avi
Ground truth caption: A structure is blowing up in the distance.
Generated caption: a <UNK> is a a the
Video file path: RjpbFlOHFps_8_25.avi
```





```
PS C:\Users\jlump\Documents\github\HW2\HW2_1> python bleu_eval.py generated_out
put.txt
Average bleu score is 0.7130372742328407
```