CPSC 8430 DEEP LEARNING
HW-2

HRUDAY CHARAN REDDY SANTIMALLA
C14698654

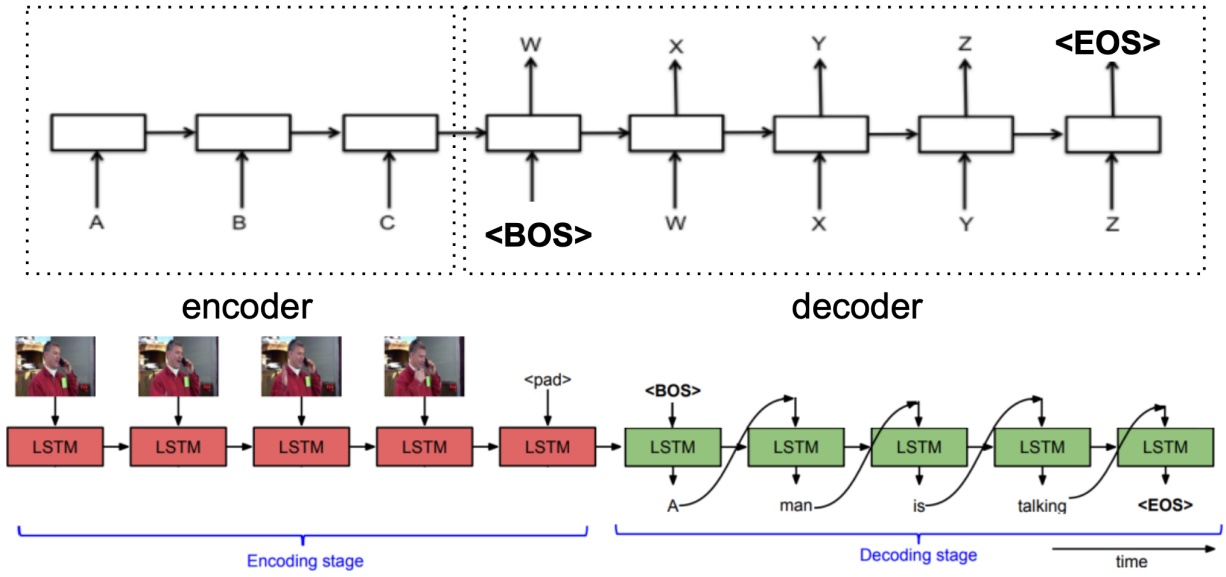
Introduction:

Input: A short video
Output: A corresponding caption that depicts the video

Example:

video properties
video ID: IhwPQL9dFYc_171_175.avi
Features: (80, 4096)
Caption: Smoke is rolling out of a jackolanterncaption
Total no of Unique tokens: 6443.

Sequence-to-Sequence: S2VT model is composed of 2LSTM
Two RNN's, Encoder and Decoder
I used 32 max encoder steps and 8 decoder steps



Requirements:
(I ran my code on palmetto cluster)

Python – 3.7.4
Pip – 22.0.4
Pandas – 0.25.1
Numpy - 1.19.5
Cuda – 11.0.3 (Provided by palmetto)
Tf-gpu – 1.15.0 (provided by palmetto)

Training and Model:

No of hidden layers used for LSTM's: 1024
Learning Rate: 0.001
Batch size: 40
Total size: 1550
Training size: 1550*0.8 = 1240 (I used 80% data to train and 20% to test)
Attention Mechanism as explained is used to increase efficiency.
Total no of epochs: 600
Optimizer: Adam
Each video has 80 frames and 4096 feature dimensions for each frame.
So, total dimensions are 80*4096.
In case of beam search beam size is given a limit of 6

Result:

The BLEU@1 score fluctuated in between 0.50 ~ 0.61 with Average length of caption ~7 and Max length of caption reaching 40.

Code: https://github.com/hruday48/CPSC-8430/tree/main/hw2/hw2_1