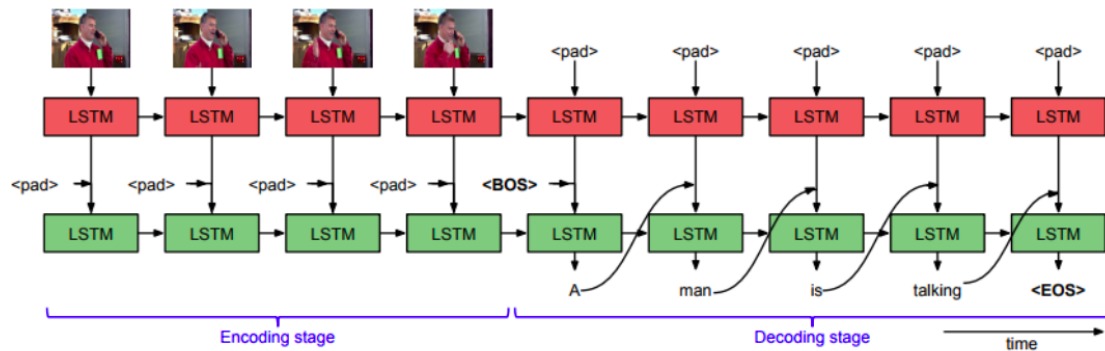


HW2 Report

Yining Hu

1 model:

I use the Sequence-to-Sequence Based Model: S2VT, which is a encoder-decoder model composed of 2 LSTM.



In the data preparation stage, I used the prepared VGG-16 model to process the data. First extract 80 frames from each video, and then use the VGG-16 model to obtain RGB features as the input of LSTM. Each video can be represented by an 80x4096 feature.

The hidden layer size of both LSTMs is 1000, and shared weights are used. And the attention mechanism is used to increase the efficiency of the model. The optimization model uses Adam optimizer. The learning rate is 0.0001, epoch is 100, and the batch size is 50.

2 performance:

I used BLUE@1score to calculate the performance of the model. Due to the lack of computing resources and time, I only used a training set of 150 videos and 100 epochs, and finally only got a BLUE@1 score of about 0.2.

3 code:

<https://github.com/Irene-233/CPSC8430/tree/master/hw2>