1.Download Dataset

This project is built on the MSVD dataset. It contains 1450 training videos and 100 testing videos.

Usage

2.To use the models that have already been trained

3.Add a video to data/testing\_data/video folder and run the predict real time file as python predict\_realtime.py

4.For faster results extract the features of the video and save it in the feat folder of the testing\_data.

5.To convert into features run the extract\_features.py file as python extract\_features.py

6.Run train.py for local training.

Scripts

#train.py contains the model architecture

#predict\_test.py is to check for predicted results and store them in a txt file along with the time taken for each prediction

#predict\_realtime.py checks the results in real time.

#model\_final folder contains the trained encoder model along with the tokenizer and decoder model weights.

#features.py extracts 80 frames evenly spread from the video and then those video frames are processed by a pre-trained VGG16 so each frame has 4096 dimensions. So for a video we create a numpy array of shape(80, 4096) config.py contains all the configurations i am using