**System introduction**

**brief introduction**

This system use RecNet which leverages both the forward (video to sentence) and backward (sentence to video) flows for video captioning.The generation loss yielded by the encoder-decoder and the reconstruction loss introduced by the reconstructor are jointly drawn into training the proposed RecNet in an end-to-end fashion.

RecNet consists of three components, the encoder, the decoder,and the reconstructor. Reconstructor can collaborate with different classical encoder-decoder architectures for video captioning. Moreover,this system employ the attention-based video captioning and S2VT.

**network**

Encoder Inception-V4

Decoder LSTM

Reconstructor LSTM (Reconstructor uses the hidden state of decoder to generate approximate vision feature (video to sentence and then sentence to video), the generated video features are compared with the source domain video features, and constraints are added.)

**Pre-training Parameter**

epochs:30

batch\_size :240

Learning rate: 2e-4

**Training Parameter**

epochs:30

batch\_size :240

Learning rate: 5e-5