



WEEK 5 PRESENTATION

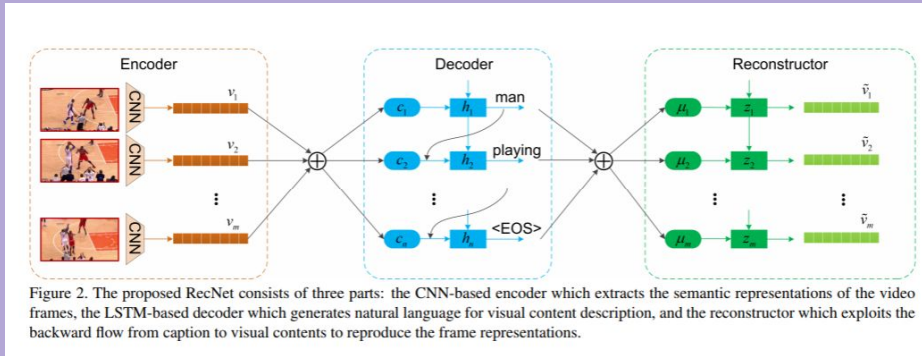
NGOC TA
AIDEAN SHARGHI



RECONSTRUCTION NETWORK FOR VIDEO CAPTIONING

- RecNet
- Dataset: MSR-VTT, MSVD
- Goal: to minimize the differences between original and reproduced video features
- Forward flow (video to sentence)
- Backward flow (sentence to video)

RECONSTRUCTION NETWORK FOR VIDEO CAPTIONING



- Encoder (CNN): extracts semantic representations of video frames
- Decoder (LSTM): generates natural language for visual content description
- Reconstructor: exploits the backward flow to reproduce frame representations



C3D Feature Extraction

- Pytorch
- Keras



THANK YOU