Meeting on August 18th

1.Report Contents:

Experiment: build an input-head for Transformer model(VIOLET)

Paper read: Two network structures for better video question&answering

A. Graph Neural Network

B. Neural Symbolic Network

2.Future plans:

· Do more experiment on the situation of constructional sites

· Try to build a GNN or NS module for VideoQA. Build model_head for fine-tuning

· Advanced experiment designing

3.Teacher's suggestions:

· For experiment part, there are many useful operation code in TBSI wiki. Read them carefully for more details. For example, if you have problems on running on GPU, there is relevant code to show how to activate GPU for training.

our wiki main-page: lab2c wiki

Cluster-usage guide: Yang Li / clusterHowTo · GitLab

· Also, transformer is often a type of large model, if you want to train&valid it on a single GPU, you should learn how to meet light-weight demands. Or you can try running with multi-gpu.

· For paper reading and learning, there are some tips to pay attention to:

1. Sometimes it is not necessary to grab detailed-network in a paper, but the most important thing is to realize what is the core of this paper. Taking GNN(DualVGR) as example, you need to figure out how the author transfer a video input to a "graph"! The truth is, the author treat video-frame as graph node, so GNN is applied to learning the "relationship" between nodes, which contains appearance and motion features that we need.

2. For end-to-end network or non end-to-end network, they both have their advantages and disadvantages. End-to-end network tend to suit more complicated environment as it doesn't constrained by human-design. However, its performance might not be state-of-art. Non end-to-end network performs well on particular practice(eg. Neural Symbolic Network), however the human-designed feature might not fit most problem. Consider these potential issues while comparing network structures.