Refined G-TAD: Sub-Graph Localization for Temporal Action Detection

Ziqun Ye

CS231n: Deep Learning for Computer Vision

Background/Introduction

More Videos published on the online-platform

- Task: Temporal Action Localization (TAL)

Problem Statement

TAL: find the starting time and end time of an action from an untrimmed video along with the category of the action.

Predict:
N annotations
- start and end time
- Action class

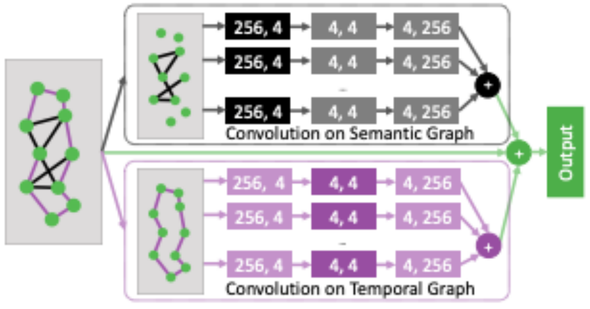
Dataset/Evaluation

THUMOS-14 dataset

This dataset contains untrimmed temporal localization video data on 20 human action classes.
- size of train data: 200 videos
- size of validation data: 213 videos

Evaluation:
- mAP

Method



This figure is from g-tad paper

Proposed Method
- Self Attention

$$v = Vx_i \quad i \in \{1, \ldots, \ell\} \tag{1}$$
$$k = Kx_i \quad i \in \{1, \ldots, \ell\} \tag{2}$$
$$q = Qx_i \quad i \in \{1, \ldots, \ell\} \tag{3}$$

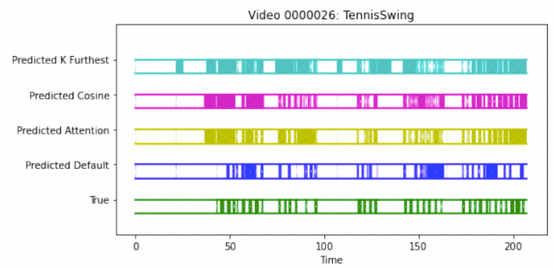
- Cosine Similarity

$$cos(x, y) = \frac{\vec{x} \cdot \vec{y}}{\|\vec{x}\|\|\vec{y}\|} \tag{4}$$

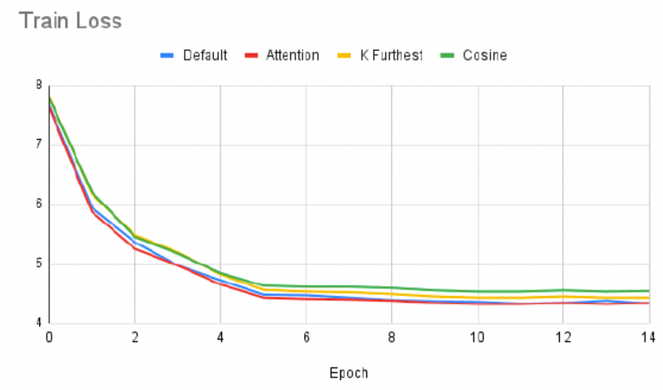
- K- Furthest
K furthest neighbor as semantic edges

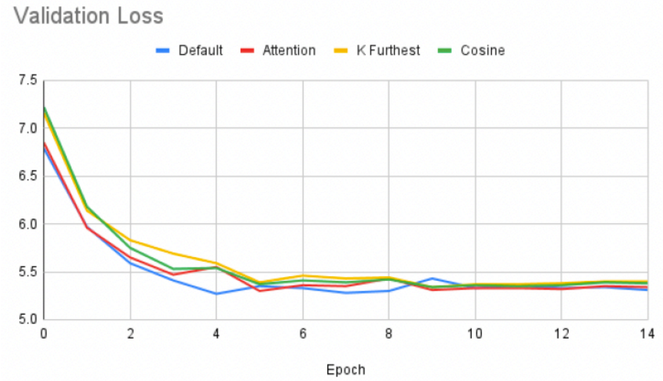
Results: mAP

Threshold	Cosine	Attention	K Furthest	Default
0.3	0.466	0.481	0.491	0.500
0.4	0.388	0.413	0.431	0.439
0.5	0.306	0.333	0.358	0.365
0.6	0.208	0.238	0.263	0.271
0.7	0.113	0.142	0.151	0.174



Losses





Conclusion

- Training losses and validation losses are very similar across different methods, suggesting that the structure of the graph proposed is irrelevant in terms of minimizing the loss

- mAP are slightly different across different methods indicating that the loss might not be a good proxy for the mAP metric