2021.11.8-11.15

上周工作

论文阅读：

- SlowFast Networks for Video Recognition - ICCV 2019
- Deep Layer Aggregation - CVPR 2018
- Actionness Estimation Using Hybrid Fully Convolutional Networks - CVPR 2016

论文实现：

YOWO

- Dataset：JHMDB21
- Evaluation Metric：frame-mAP

Setting	batch size	epoch 5	epoch 6	epoch 7	epoch 8	epoch 9
16 frame	6	56.27%	58.18%	58.86%	58.34%	59.29%

Setting	batch size	epoch 10	epoch 11	epoch 12	epoch 19	epoch 20
16 frame	6	59.25%	58.89%	58.93%	60.24%	59.41%

实验分析：

1. 上一周工作中模型在jhmdb上，仅对默认配置文件的batch size做了修改，以满足显卡的显存限制，但性能相比论文中的74.4%相差较多，batch size对模型的性能不应该这么大。通过这次的实验，将batch size改回预设的6，性能没有提升，证明了该结论。
2. 考虑到模型可能是欠拟合，增大epoch数至20 (预设为10)。通过表格中的实验结果发现，增大epoch对模型的性能提升很小。
3. 从paper中的结果可知该模型的性能潜力，而我复现时性能差别较多，可能还是欠拟合的问题。由于jhmdb的规模较小，为了防止过拟合，在训练过程中会freeze 2D，3D分支网络的参数。为了解决这里遇到的欠拟合问题，对两个分支的参数进行解冻，查看实验结果，目前仍在训练。

目前想法

1. YOWO的两个分支分别提取空间上的Appearance信息及时序上的Motion信息，在相关论文的阅读中有了以下认识：
    - 空间上的2D卷积有个条件，即空间上的二维，即x和y是等价的，所以可以通过卷积来提取两个维度的信息。而时序维度与空间维度是不同的，不应该一起处理，所以YOWO的3D卷积应该更注重时序的Motion提取。

        > 对应的想法是在*3D*分支上仅考虑时序上的卷积，即卷积核大小改成*1*1*，或者使用时序上的注意力。待实验。

2. frame-mAP作为评价指标仅能判断单帧上的性能，而video-mAP会考虑连接算法的性能。

    > 可以折中使用*clip-mAP*或者判别单帧性能的时候，考虑当前帧前后几帧的结果。等待多篇论文复现后再进行实验。