本周在应付期末考试，所以没有花多少时间在这方面。主要是在解决上周留下的几个问题，通过 with torch.no_grad()，解决了跑验证集时显存不够的问题。另一方面，将生成 ground truth 的 json 文件的代码修改了，通过 annotation 的 txt 文件以及视频的 FPS 来生成 json 文件，而不是通过 xgtf 文件。小样本的 ground truth 文件可以通过提取完整 ground truth 文件中相应的字典得到。使用新的 ground truth 对 30 个视频的小样本的预测结果求 mAP，得到如下结果：

```
Diving 0
CliffDiving 1
ThrowDiscus 2
CricketBowling 3
CricketShot 4
JavelinThrow 5
BaseballPitch 6
CleanAndJerk 7
LongJump 8
Billiards 9
[INIT] Loaded annotations from validation subset.
        Number of ground truth instances: 381
        Number of predictions: 609
        Fixed threshold for tiou score: [0.1 0.2 0.3 0.4 0.5]
[[0.99013685 0.         0.33333333 0.9950542  0.99134199 0.98148148
  0.97959711 1.         0.99892473 0.9848247 ]
 [0.99013685 0.         0.33333333 0.97278271 0.99134199 0.98148148
  0.97959711 1.         0.99892473 0.9848247 ]
 [0.98033293 0.         0.33333333 0.97278271 0.99134199 0.98148148
  0.97537413 1.         0.99892473 0.9848247 ]
 [0.98033293 0.         0.33333333 0.97278271 0.99134199 0.98148148
  0.94834711 1.         0.99892473 0.9848247 ]
 [0.98033293 0.         0.33333333 0.97278271 0.99134199 0.98148148
  0.92144169 1.         0.99892473 0.9848247 ]]
[RESULTS] Performance on ActivityNet detection task.
mean AP at different thresholds: [0.82546944 0.82324229 0.8218396  0.8191369  0.81644636]
        Average-mAP: 82.12269170191408
```

其中 AP 为 0.333333 的那类一共只有三段 ground truth。关于 diving 和 cliffdiving 两类由于 ground truth 的重复而出现有一个类概率为 0 的问题，我没想到好的解释和解决方法，我最早以为是 NMS 的逻辑出错了把不同类别的数据抵消了，但并没有，我查看了相应视频对应的网络输出值，经过 softmax 以后 diving 类的分数都是 0.9 这种，也就是说他把这段视频判断为其中一类了。在训练时输入的 ground truth 是包含两个标签的，想通过一个只含这两类标签的小样本进行一下测试。

之后对整个验证集样本进行了测试，得到的 mAP@0.5 为 0.21 多，取了多个 epoch 结束生成的模型（epoch：30-35）进行了验证，得到的 mAP@0.5 均在 0.21 到 0.22 之间，其中最好的一次结果如下

```
SoccerPenalty 0
VolleyballSpiking 1
HighJump 2
CricketBowling 3
HammerThrow 4
JavelinThrow 5
LongJump 6
Diving 7
BasketballDunk 8
CliffDiving 9
Billiards 10
BaseballPitch 11
CleanAndJerk 12
Shotput 13
ThrowDiscus 14
GolfSwing 15
PoleVault 16
CricketShot 17
FrisbeeCatch 18
TennisSwing 19
[INIT] Loaded annotations from test subset.
        Number of ground truth instances: 3358
        Number of predictions: 11664
        Fixed threshold for tiou score: [0.1 0.2 0.3 0.4 0.5]
[[0.14126179 0.0909502  0.27008698 0.33940109 0.48569624 0.46934591
  0.60208686 0.63166549 0.66208889 0.         0.21519126 0.10575802
  0.50937633 0.21885359 0.36154424 0.40911241 0.60235927 0.19128831
  0.21404075 0.39422464]
 [0.13153683 0.0878691  0.26729422 0.31651769 0.47849652 0.46656526
  0.59933233 0.61769901 0.64201854 0.         0.17507022 0.10239482
  0.49214519 0.21182422 0.35915247 0.34950906 0.59015652 0.18446093
  0.18817858 0.36339451]
 [0.12202038 0.07640791 0.2561499  0.30944082 0.44706632 0.44299876
  0.59910582 0.59058985 0.6091736  0.         0.14066528 0.08209235
  0.4373949  0.18707164 0.3308914  0.3298481  0.55235292 0.18221903
  0.18089981 0.28365533]
 [0.09895799 0.04861469 0.23359649 0.28844953 0.41386816 0.41588871
  0.56963673 0.54777072 0.55223542 0.         0.10055526 0.07379434
  0.39593508 0.16399511 0.3157431  0.25582491 0.50088606 0.15724521
  0.14397717 0.19104849]
 [0.0517728  0.03736745 0.14218738 0.25144129 0.38324008 0.35790184
  0.43464043 0.47274223 0.46834742 0.         0.06045907 0.06340013
  0.31323748 0.11280368 0.2522218  0.21587922 0.3859838  0.10654344
  0.10472849 0.13212736]]
[RESULTS] Performance on ActivityNet detection task.
mean AP at different thresholds: [0.34571661 0.3311808  0.30800221 0.27340116 0.21735127]
        Average-mAP: 29.513040952347307
```

还未到达论文给出的 mAP 值，之后我把整个训练集用 40 个 epoch 训练后得到的模型拿来跑了一下测试，得到的结果如下：

```
Diving 0
CliffDiving 1
ThrowDiscus 2
Shotput 3
CricketBowling 4
CricketShot 5
PoleVault 6
LongJump 7
BasketballDunk 8
JavelinThrow 9
BaseballPitch 10
FrisbeeCatch 11
CleanAndJerk 12
SoccerPenalty 13
HighJump 14
TennisSwing 15
GolfSwing 16
HammerThrow 17
VolleyballSpiking 18
Billiards 19
[INIT] Loaded annotations from validation subset.
        Number of ground truth instances: 3007
        Number of predictions: 7002
        Fixed threshold for tiou score: [0.1 0.2 0.3 0.4 0.5]
[[0.98052087 0.         0.82357307 0.93387785 0.96664473 0.94113407
  0.90547951 0.97050455 0.98654992 0.9853786  0.98707483 0.86156987
  0.93696425 0.91774511 0.97164245 0.88473871 0.7717376  0.97139519
  0.86900762 0.91960512]
 [0.97922882 0.         0.82309614 0.93126471 0.96301137 0.93729811
  0.90547951 0.97023811 0.98654992 0.9853786  0.98707483 0.85623411
  0.93696425 0.91774511 0.97164245 0.88473871 0.730613   0.97130636
  0.86687826 0.91395696]
 [0.97869669 0.         0.80776466 0.92622269 0.96301137 0.93729811
  0.90547951 0.96818369 0.98440399 0.9853786  0.98707483 0.85358709
  0.93696425 0.91774511 0.97164245 0.88426364 0.730613   0.96654623
  0.85558774 0.90958681]
 [0.97714269 0.         0.77224869 0.9185038  0.9516744  0.9325044
  0.90420563 0.96243563 0.98440399 0.9853786  0.96666667 0.83890743
  0.93696425 0.90458043 0.97164245 0.88426364 0.730613   0.9631522
  0.83013516 0.90867564]
 [0.97013788 0.         0.75244888 0.90766636 0.94691002 0.9325044
  0.89477499 0.95810025 0.96842331 0.98489724 0.96666667 0.80972603
  0.91531723 0.88522723 0.96935631 0.86273238 0.70549918 0.96048861
  0.7792432  0.89079928]]
[RESULTS] Performance on ActivityNet detection task.
mean AP at different thresholds: [0.8792572  0.87593497 0.87350252 0.86620494 0.85304597]
        Average-mAP: 86.958911885275
```

可能我得到的模型与训练集过拟合了？暂时没想到好的解决办法。