

本周完成了网络测试部分的代码,遇到的主要问题是由于我把每次输入的图片数量设的与训练时相同,但因为验证时不包含 **dropout** 等操作,所以显存爆掉了,把每次放入图片数量减少即 L 维长度减少解决了此问题。验证程序的逻辑是每次放入长度为 L 的视频,下一次放入时间序列上 L/4 以后的长度为 L 的视频,取得分>0.05 预测视频段经过 nms 后得到预测结果与 ground truth 比较。

其中做了一个小样本测试,取了三十个训练集视频在训练网络上继续训练,之后再相同的视频跑验证集,由于这些视频并不包括所有 20 类视频,但我在继续训练时没有考虑到删除多余的类,因此还是以 num_classes=21 进行的训练,在计算 map 时我将十多个预测到的不属于这 9 类的预测段删除,得到了如下结果,其中 CliffDiving 在预测时全部检测成 Diving,下图中二维矩阵中五行分别表示 iou 取 0.1-.0.5 对应的 ap,每行中每个精确率分别对应上面的 9 个类。

```
BaseballPitch 0
JavelinThrow 1
CleanandJerk 2
CliffDiving 3
ThrowDiscus 4
BilliardsShot 5
CricketBowling 6
CricketShot 7
LongJump 8
[INIT] Loaded annotations from test subset.
      Number of ground truth instances: 324
      Number of predictions: 1222
      Fixed threshold for tiou score: [0.1 0.2 0.3 0.4 0.5]
[[0.70677338 0.73584375 0.99450549 0.39237737 0.16666667 0.11688948
  0.56754464 0.7859192 0.98299808]
 [0.64673928 0.72692709 0.99450549 0.30419301 0.11111111 0.08592316
  0.52428156 0.76146618 0.98299808]
 [0.64673928 0.66596037 0.94527042 0.24623907 0.04761905 0.05639274
  0.50746112 0.61562047 0.98299808]
 [0.56851435 0.63950386 0.82671175 0.19298712 0.04761905 0.05188146
  0.47288422 0.61562047 0.98299808]
 [0.43731322 0.62128294 0.81682164 0.14711733 0.04761905 0.04154723
  0.43100144 0.61562047 0.98299808]]
[RESULTS] Performance on ActivityNet detection task.
mean AP at different thresholds: [0.60550201 0.570905 0.52381118 0.48874671 0.46014682]
Average-mAP: 52.98223421732713
```

之后又对整个测试集样本进行了测试,得到了如下结果,我查看自己的预测结果中,没有产生 cliffdiving 的类,这个和 ground truth 里面所有 cliffdiving 类的 annotation 都出现在 diving 类中应该有关系。整体 map 结果都偏低,有个原因可能是因为在训练时每个 batch 放入 1304 张图片,但我在验证时每个 batch 只取 512 张图片,另一个方面过拟合可能也有所影响。接下来想重新用那三十个样本训练一个 num_classes = 10 的网络,看看 mAP 和这次相比能提高多少,如果还有问题就把有问题的类的输入输出的数据提出来单独分析下。

```
BaseballPitch 0
CricketBowling 1
CricketShot 2
FrisbeeCatch 3
JavelinThrow 4
SoccerPenalty 5
BilliardsShot 6
VolleyballSpike 7
Diving 8
LongJump 9
HighJump 10
ThrowDiscus 11
CleanandJerk 12
ShotPut 13
HammerThrow 14
GolfSwing 15
PoleVault 16
CliffDiving 17
TennisSwing 18
BasketballDunk 19
[INIT] Loaded annotations from test subset.
      Number of ground truth instances: 3141
      Number of predictions: 13068
      Fixed threshold for tiou score: [0.1 0.2 0.3 0.4 0.5]
[[0.14536601 0.25102553 0.15092003 0.20195065 0.30819033 0.16334731
  0.15730543 0.08457138 0.20452745 0.57389798 0.20514586 0.25167671
  0.50918814 0.17866519 0.45489527 0.32331449 0.54206741 0.
  0.43889903 0.64086147]
[0.13933557 0.23095749 0.14417073 0.18015466 0.30316217 0.11889197
  0.12715814 0.07316709 0.19246578 0.57389798 0.18789123 0.23364863
  0.48710682 0.17356011 0.45022286 0.2848316 0.53177138 0.
  0.4067961 0.62290603]
[0.13395509 0.1950517 0.12345542 0.17219253 0.29271658 0.10216894
  0.10154075 0.06023107 0.18357267 0.55798666 0.17134631 0.20647976
  0.4513947 0.15585305 0.43792499 0.22462045 0.49870507 0.
  0.33345465 0.58102921]
[0.1198395 0.16913596 0.0986435 0.14525967 0.2644498 0.06529794
  0.06987811 0.04923192 0.16947706 0.51129826 0.14983683 0.18215484
  0.40205602 0.1169697 0.3975689 0.1960248 0.44058345 0.
  0.2018953 0.50585915]
[0.10671936 0.15622599 0.07355686 0.07390203 0.2273685 0.04805825
  0.05448385 0.0353764 0.13470856 0.37423477 0.09761343 0.10704425
  0.29462621 0.09638685 0.32589651 0.131343 0.33952388 0.
  0.14319771 0.39777862]]
[RESULTS] Performance on ActivityNet detection task.
mean AP at different thresholds: [0.28929078 0.27310482 0.24918398 0.21277304 0.16090225]
Average-mAP: 23.70509736009531
```