ActivityNet task 1 summary

2019年7月29日 星期一　　下午1:57

1. TSN(用来提取特征，这里每隔16帧，提取一帧的特征）

a. 准备数据：

i. 需要视频的RGB帧和光流，在caffe版的TSN的开源代码里有脚本，可以用来抽取视频的帧和光流。可以参考如下链接：
https://github.com/yjxiong/temporal-segment-networks#extract-frames-and-optical-flow-images
RGB帧总是比光流帧少一帧的，这里丢掉RGB的第一帧。

ii. 根据抽好的帧和光流，创建TSN训练、验证、测试数据文件列表（train_list.txt, val_list.txt, test_list.txt)：

1) 文件夹路径：是每个视频的路径，路径下包含RGB帧和Flow帧
2) 总帧数：指RGB帧的数目
3) 类别：ActivityNet 数据集200类到0～199的hash（这边就按照字典排序）
4) 开始：官方标注json里（activity_net.v1-3.min.json）对应视频的标注，action的起始帧（时间和帧率换算而来）
5) 结束：action的结束帧
若一个视频有多个action，则有多条记录
test_list.txt中没有标注，类别、开始、结束三个字段都为-1

文件夹路径	总帧数	类别	开始	结束
./datasets/v_sJFgo9H6zNo	4164	57	0	3697

./datasets/v_sJFgo9H6zNo	4164	57	3750	4160

b. 训练TSN：

TSN是two-stream网络，RGB和Flow组成，这两个stream分开训练，得到两个模型

i. RGB stream 训练，

ii. Flow stream 训练，

c. 提取特征：

i. 对于全部视频（train，val， test），每隔特定距离提取一帧的特征（我们比赛中step_size设置为16），RGB和Flow特征分别用各自的网络的提取，时间维度长度应该相等

ii. RGB特征：时间维度*200， Flow特征：时间维度*200，这里的200是logits（200类的scores，before softmax）

2. BSN（proposal预测）

a. 特征缩放与拼接：

i. 因为视频时间长度不一，训练BSN需要固定时间长度，这里我们把所有视频特征，时间维度都rescale到100，然后拼接起来每个视频得到100*400大小的特征（100时间维度，400：200RGB+200Flow），得到rescaled feature。相对应我们还可以得到没做缩放的的non-rescaled feature，这个特征长短不一，不能用来训练只能在BSN inference阶段使用，且BSN的TEM，PEM两个模块的inference batch size要设置为1（时间维度长度不等）

b. 训练BSN（即训练TEM， PEM两个网络）：

i. BSN介绍：

1) 组成及流程：rescale feature->TEM->PGM->PEM->soft-NMS->Top100 proposal (if total proposal>1000)

a) TEM：3层Conv layer，对于每个时间点预测

start，end，action三个概率值，output 3×100, loss function是三个概率的label balance binary loss之和；

b) PGM：根据start，end生成候选proposal，在action对应位置抽取32维特征做为下一个模块PEM的输入。32维特征：抽取start和end左右各4个时间点，中间部分抽取16个时间点；

c) PEM：2层FC layer，得到每个proposal的score；

d) soft-NMS：抑制冗余的proposal。

c. Inference + soft-NMS得到最后的100个proposal

d. Evaluation:

i. 评价指标 average AUC, 在10个不同IoU thresholds下AUC的平均。

IoU_thresholds=(start, end, step_size)=(0.5, 0.95, 0.05)；AUC是曲线 Average Recall@1~100下的面积