本周解决了在训练过程中和作者精度差很多的问题，并且将自己的数据增强做法和作者的进行比较，得出作者在数据增强方面最为关键的两个方法，便于日后应用于自己的设计中以提高精度。另外又重新巩固了一遍 resnet 和 inception 网络的特征，希望更加清晰其特点以便迁移应用，了解了论文 Rethinking the Faster R-CNN Architecture for Temporal Action Localization 中对于视频在时间上进行定位的思想。

这周再次仔细翻看作者的代码后，发现在数据的预处理和数据训练的处理上有比较大的区别。其中在光度变化的处理、随机裁剪、图像的扩展、以及 ground truth 的训练方法均有所不同。

首先对于光度变化只是简单的参数设置不同，区别不太大。

随机裁剪中，我在裁剪的过程中保证了每一张图是按照原图的比例裁剪的，但是作者的做法中会将图像的形状进行一定的改变，而且比例变化区间为[0.5,2]。

作者的图像扩展是指，将一张图按照 1-4 间的随机比例对长和宽进行扩张，然后将图片依然按照原来的大小放到扩张后的图像的右下角位置位置，而多余出来的部分全部用认为是黑色（全部置 0）。

作者对于 ground truth 的训练，是一次性将所有的 gt 全部放入网络中进行训练，也就是说网络每训练一个 epoch 都会将所有视频的所有 gt 全部过一遍。但是我刚开始的做法是以单个视频为单位，每一个 epoch 随机选取一段进行训练，同样的 loss 可以降到很低但是过拟合很严重，这应该是因为本来数据集就很小。

为了对比作者的做法与我的做法上的根本区别，为找出对 map 影响最大的数据增强方案，进行了以下实验：

Dataset	modality	方案	Best map
UCFSports	RGB	作者训练好的参数验证结果（对比）	0.8249
		上周训练结果（对比）	0.6559
		除了随机裁剪外，全部按照作者提供的思路来，即采用作者的光度变换参数，图像扩展参数和方案，随机等比例裁剪，随机镜像，一次性将所有 gt 全部放入训练	0.8259
		从自己的思路出发，保留自己的光度变换参数，随机等比例裁剪，随机镜像，不采用图像扩展，仅仅将所有的 gt 一次性放入训练	0.7753 (这一个操作效果提升 10%+)
		在上面思路的基础上，仅仅加入图像的随机扩展的增强方案，和自己最初的思路相比，仅仅增加了 gt 一次性放入训练和随机扩展。	0.8601 (加入图像随机扩展后精度提高也非常明显)

（训练策略：先 freeze 住提取 feature map 层，以 0.001 学习率训练最后随机生成的两层，然后以 0.0001 训练所有层，步长的缩减设置为每 2 个 epoch 减少 0.94 倍）

采用一次性将所有 gt 放入训练的方法在 UCF101 数据集上测试时数据量巨大，所以训练的速度非常慢，而且数据量过大时对超参数的改变时间代价往往比较大，一个 epoch 需要 10h+，虽然只要几个 epoch 应该能达到效果但是因为前面的步长设置不太合理所以现在还没有得到比较满意的结果，仍然在训练。

2018 年的论文 Rethinking the Faster R-CNN Architecture for Temporal Action Localization 中作者将 Faster R-CNN 的思想应用到视频的时序上，其主要的核心思想有两个:

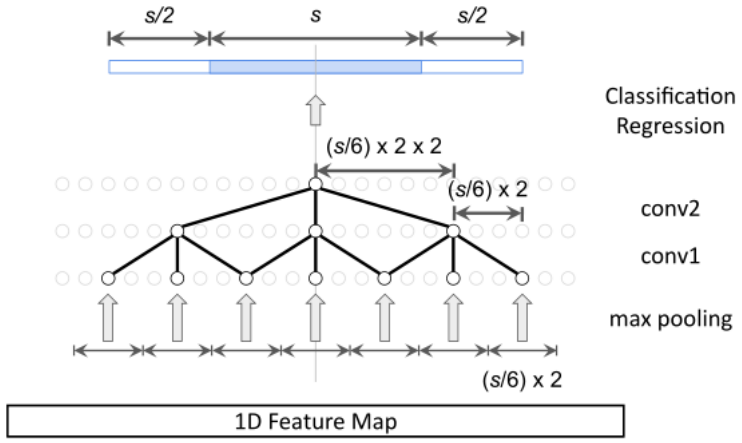

Figure 4: Incorporating context features in proposal generation.

（1）为了能够提高召回率，anchor 段就必须要有足够长的，然而如果感受野太小（在时间上看到的很小一段），在对比较大的 anchor 分类的时候，提取出的 feature 可能会没有足够的信息。如果感受野太大，在对比较小的 anchor 进行分类的时候容易被无关信息支配。

为了解决这个问题我们提议把每一个 anchor 的感受野和它的时间跨度对齐。这个有两个关键：a multi-tower network 和 dilated temporal convolutions（膨胀时间卷积）。在给定一维 feature map 的情况下，Segment Proposal Network 是由 K 个时间卷积网络组合而成，每一个网络只对一个特定长度的 anchor 负责进行分类，这 K 个时间卷积网络每一个都是设计为 anchor 长度和时间轴上的感受野是对齐的。在 K 个网络的每一个的最后，都会用 2 个大小为 1 的卷积层用来分类和边界回归。

（2）在 anchor 的前后需要加长 s/2 ,将上下文特征纳入 anchor 中，上下文特征提供了强大的语义线索，为识别边界内的动作类型提供信息。

下周要继续了解现有在动作识别和定位方面的文献，了解现有的基本思想和方法。待训练完成后要对现有结果进行错例分析。学习网易云机器学习课程巩固基础。