

本周把代码的 base model 改成了 resnet101，用了在 Kinectics 上训练过的与训练模型，RPN 的输出在 resnet101 的 layer3 和 layer4 之间，暂时测试了 epoch 从 40-60 的结果。测试集的 mAP 结果如下所示

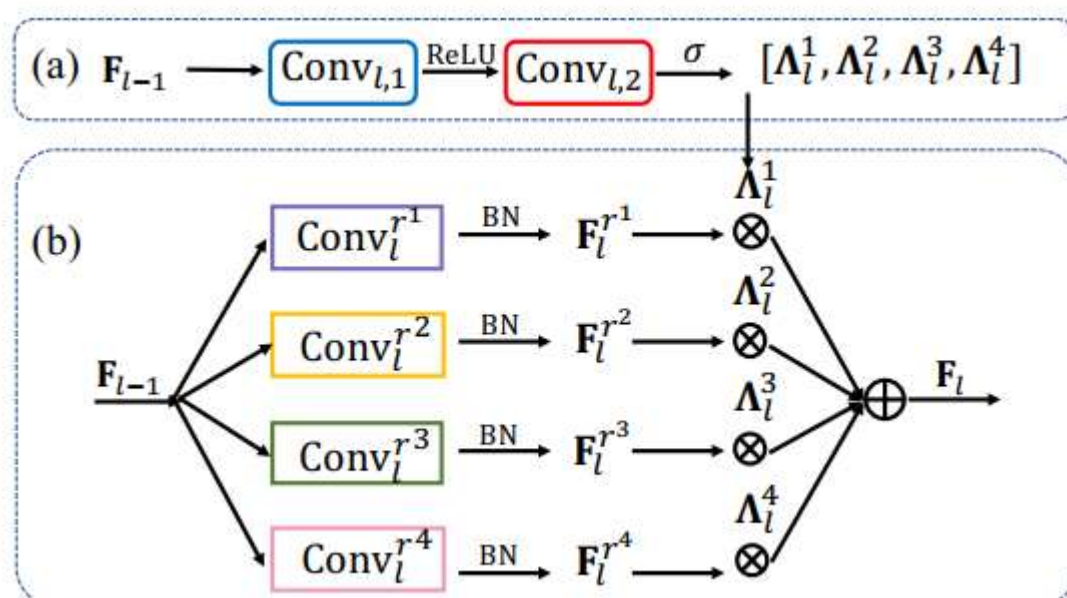
Epoch	$\alpha = 0.1$	$\alpha = 0.2$	$\alpha = 0.3$	$\alpha = 0.4$	$\alpha = 0.5$
40	49.09	47.47	45.29	41.92	35.60
45	49.11	47.45	45.20	40.95	34.01
50	50.97	49.38	46.59	42.20	36.32
55	51.55	49.55	46.97	43.30	36.01
60	50.88	49.22	46.47	41.78	35.40

mAP 还在波动，之后做更多的测试来做进一步分析。

本周还看了两篇论文，总结如下

Autofocus Layer for Semantic Segmentation

本文阐述了一种基于 soft attention 的平行网络模块自适应地改变感受野的尺寸来取得更好的特征。



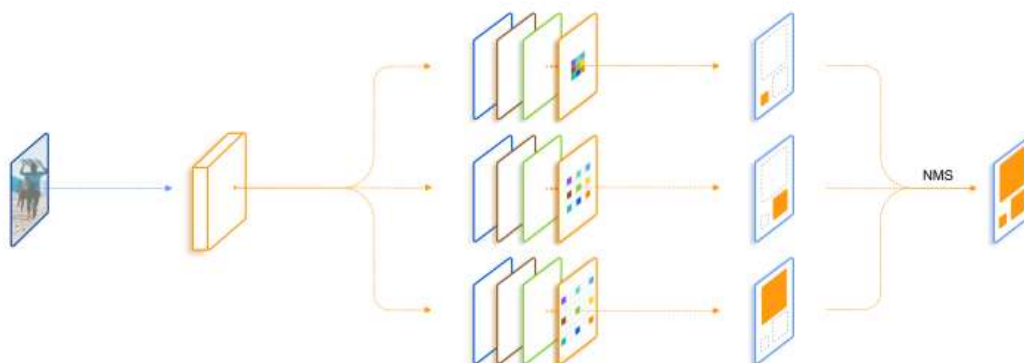
通过有不同 dilation rate 为 r^k 的 K 个平行卷积层得到不同感受野的 feature map 再通过注意力机制学习不同平行卷积层间的权重，最后融合为输出。图 (a) 为注意力模型，结构可写成下述表达式：

$$\Lambda_l = \sigma(\text{Conv}_{l,2}(f(\text{Conv}_{l,1}(\mathbf{F}_{l-1}))))$$

$\text{Conv}_{l,1}$ 是一个 $3*3*3$ 的卷积，作用是将 channel 层个数减半，经过 relu 激活函数到达 $\text{Conv}_{l,2}$ ，一个 channel 个数为 K 的 $1*1*1$ 卷积，之后接一个 softmax 函数进行归一化 (softmax 在 $\text{dim}=1$ 即 channel 维上进行)， K 个 feature map 分别对应图 (b) 中的 K 个不同权重，图 (b) 表示对上一层的特征进行不同 r^k 的空洞卷积，得到不同的 size 相同但感受野不同的 feature map，再通过权重的对每个元素对应相乘求和得到最终的输出。

Scale-Aware Trident Networks for Object Detection

本文提供了一种只改变感受野来对多尺度目标进行检测的思路，主要提出了 **weight sharing** 和 **scale-aware training** 两个方法，作者做了大量的对比实验来测试 **trident** 模块，思路很清晰，值得学习。本文主要讨论感受野问题对不同尺寸目标检测的影响，通过将网络中卷积层改成不同 **dilation rate** 的空洞卷积处理，得到了不错的效果。网络的结构图如下



baseline 网络为 resnet101 为基础的 faster rcnn，思路是将 resnet 的 backbone 中 conv4 部分的 3×3 卷积层改成 dilation rate 分别为 1,2,3 的并行空洞卷积（1 即为普通卷积），这样得到多路预测区域的 bounding box 通过 NMS 处理得到最后总的预测结果。由于每个 brunch 的结构除了空洞卷积层其它层相同，所以可以通过 **weight sharing** 操作只改变感受野范围而不对其它参数进行改进，同时也减少了计算量。在训练的时候针对不同感受野的 brunch 用不同面积（gt 和 proposal 的面积开方作为阈值）的样本进行训练，若不在尺寸范围内则把其视作负样本（scale-aware）。在 inference 的时候用单独的一个在训练时增加了 scale 范围 $(0, \infty)$ 的 brunch 也可以取得比 baseline 更好的结果，作者分析是因为 **weight-sharing** 的原因，**weight-sharing** 也能防止使用 **scale-aware** 时导致过拟合的情况。

文章在 resnet 不同位置的 block 都使用过 **trident** 结构进行测试，最终得到放在最后效果会比较好，若有多个 **trident** 连续，则会是相同 scale 的 brunch 直接连接在最后进行 NMS。所以我没想通他是怎么在前面的层做实验的，以及 **weight-sharing** 的具体结构和反向传播时的逻辑，想等下周代码公布以后再细看一下。