

基于深度学习的物体检测 - 作业 4

peng00bo00

August 23, 2020

1. 本次作业是对 RetinaNet 代码进行调试，现将代码核心内容总结如下：

RetinaNet 主要由 3 部分组成，分别为 FPN、Anchor_Generator 以及 RetinaNet_Head

- (a) FPN 与 Faster RCNN 中的功能相同，负责生成特征金字塔。在 RetinaNet 中 FPN 接受一个包含 3 个不同尺度的特征金字塔 (res3, res4, res5)，然后利用卷积运算自上而下进行特征融合生成对应的特征图 p3-p5。最后利用卷积对 res5 进行降采样，生成特征图 p6-p7，从而得到一个包含 5 层的特征金字塔。
- (b) Anchor_Generator 与 Faster RCNN 中作用相同负责生成不同尺度的锚框。
- (c) RetinaNet_Head 负责在特征图上预测锚框的类别和偏移量。RetinaNet_Head 主要由 cls_subnet 和 bbox_subnet 两个子网络构成，每个网络利用卷积对输入的特征图计算出特征并通过 cls_score 和 bbox_pred 两个卷积模块计算得到最终类别得分以及偏移量。值得说明的是在不同尺度特征图上 cls_subnet 和 bbox_subnet 两个子网络是共享的，因此不同尺度的 anchor 都会对两个子网络进行训练。
得到锚框类别和偏移量后通过对 anchor 进行匹配来计算分类损失和偏移量损失。其中分类损失是在不同类别上利用 focal loss 进行多个二分类，不像 Faster RCNN 那样直接使用 softmax 计算多分类损失，此外在 RetinaNet 中分类损失是考虑全部锚框而 Faster RCNN 中则是通过抽样来控制锚框正负样本的数量。RetinaNet 的偏移量损失则与 Faster RCNN 类似，只考虑正样例 (非背景类) 的锚框而忽略其他锚框。得到损失后即可通过优化器来更新网络的参数。