

基于深度学习的物体检测 - 作业 5

peng00bo00

August 30, 2020

1. 多阶段法 Faster R-CNN 性能好于单阶段法 SSD 的原因如下：

- (a) 主要原因在于 Faster R-CNN 中使用了 RPN 来产生候选框而 SSD 则是直接使用锚框。RPN 相当与首先对锚框进行了一次筛选，去除了锚框中属于背景的部分使得后续的网络可以更容易工作；而 SSD 中由于没有使用 RPN 需要对大量的背景类锚框进行处理，这部分锚框容易造成正负样本不均衡的问题从而影响结果。
- (b) 在 Faster R-CNN 中得到 proposal 后使用了 ROI pooling 来整合锚框内的信息，而 SSD 则是直接使用锚框对应的特征向量，因此 Faster R-CNN 能够更好地把握候选区域的整体信息，效果更好。
- (c) Faster R-CNN 中在 RPN 首先进行了一次锚框位置的回归并且在 Fast R-CNN 网络中又对 proposal 的位置进行了修正；而 SSD 仅对锚框的位置进行了一次回归，因此 Faster R-CNN 得到的区域会更接近于实际的目标。

2. 基于锚框的单阶段法 RetinaNet 和无需锚框的中心域法 FCOS 的区别如下：

- (a) anchor 的使用方法不同：RetinaNet 把 anchor 作为锚框中心，而 FCOS 则把 anchor 作为目标内的一个点
- (b) anchor 的数量不同：RetinaNet 在每个 anchor 的位置要设置多个锚框以适应不同的高宽比，而 FCOS 每个 anchor 位置仅有 1 个
- (c) anchor 的回归目标不同：RetinaNet 对 anchor 回归得到锚框中心和长宽的偏移量，而 FCOS 对 anchor 回归得到 anchor 到目标框 4 个边界的距离
- (d) anchor 的匹配方式不同：RetinaNet 是把每个 anchor 匹配到具有最大 IoU 的目标上，容易产生正负样本不均衡的问题；而 FCOS 则是将落在目标框内的所有 anchor 都匹配上，所以正负样本的比例要好很多
- (e) 损失函数不同：RetinaNet 对每个 anchor 要预测类别和偏移量，而 FCOS 除此之外还要预测一个中心性 (centerness)
- (f) 多尺度机制不同：RetinaNet 在不同尺度的特征图上对不同大小的目标进行匹配，而 FCOS 则是按 anchor 到目标中心的距离进行多尺度匹配