

Focal Loss

Reporter

冯浩哲

Focal Loss

论文概要

什么是Focal Loss,它是怎么起作用的

作者用了什么样的结构,并在这个结构中使用了Focal Loss达到了state of art的成果

论文概要

论文面向one stage的网络,提出了focal loss以解决one stage网络比不上two stage的表现的问题。整个文章可以着眼于以下两个问题上:

1. 什么是Focal Loss,它是怎么起作用的
2. 作者用了什么样的结构,并在这个结构中使用了Focal Loss达到了state of art的成果

什么是Focal Loss,它是怎么起作用的

作者研究了 one stage detector和 two stage detector为什么精度不一样这个问题后发现,这是因为 one stage detector具有非常非常严重的正负例样本不均衡的问题。因为对于一阶段的detector,每一个location有若干个Anchors,但是只有少数anchor是有物体的,这就是正负例的不均衡了。在真正的训练过程中,检测器可以把背景anchor大部分都对(也就是没有物体的confidence达到0.5以上),但是在这之后训练器会仍然会努力让这些confidence接近于1,如果有200个anchor,只有10个有物体的话,那么显然训练器就会着重于让背景部分的confidence达到1,这部分梯度把更重要的10个anchor分类正确的梯度给在某种程度上稀释了。

一般可以使用的解决方法是做加权,但是效果并不是非常好。作者提出了所谓的Focal Loss就是一个根据预测的自加权问题,具体而言就是这个公式:

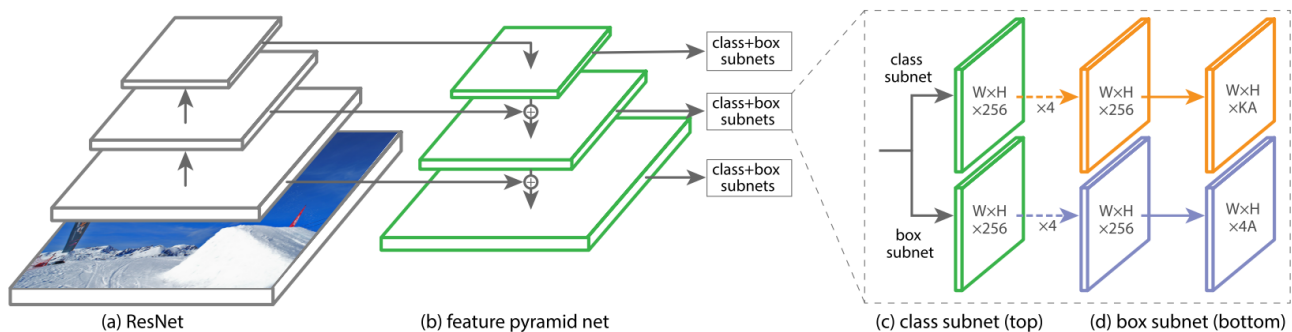
$$FL(p_t) = -(1 - p_t)^\gamma \log(p_t)$$

前面的加权项表示在当前的GT下,我们的预测越接近于GT,那么这个损失的权重就越小。而所谓的 γ 项就是对这个损失的权重衡量,一般取偶数,值越大加权就越厉害,一般取2。

从这个公式我们可以看出的是,这样的自加权方法就把训练重心转移到对未分类好的anchor的梯度上了,这样就很有重点了。

作者用了什么样的结构,并在这个结构中使用了Focal Loss达到了state of art的成果

作者用了一个一阶段的RetinaNet来达到这种效果。RetinaNet继承了FPN的一些特性,可以用一张图简述:



这张图中，对每一个层次的Feature都做了预测，预测输出的是特征图大小 $W * H$ 的预测，表示有 $W * H$ 个 location，同时对每个location有A个anchor boxes，每个anchor box有k个类别预测与4个bbox regression预测。

在类别预测的时候我们就用了focal loss， $\gamma = 2$ 的参数。anchor box与训练细节可以在论文中自行寻找。