YOLO9000:Better, Faster, Stronger
http://blog.csdn.net/jesse_mx/article/details/53925356
https://zhuanlan.zhihu.com/p/25167153

摘要

YOLO9000可以检测9000个不同类别的物体
YOLOv2 可以在不同的尺度上运行
在VOC 2007 数据集，67FPS的速度mAP是76.8，40FPS的速度mAP是78.6

1.介绍

大多数物体检测方法受约束于一个很小的物体集
分类的数据集比检测的数据集更广，想让检测可以达到分类的数据水平，但标记数据用于检测很难，拿到跟分类数据一样多的检测数据在近期是不可能的
利用大量的分类数据：物体分类的分层视角，结合不同的数据集
联合训练方法：在检测和分类数据上训练物体检测器，检测数据用来学习精确的定位物体，分类数据用来增加词汇量和鲁棒性
对YOLO的提升是YOLOv2,然后使用结合数据集和联合训练的方法，训练一个可以检测9000多种类别的模型叫做YOLO9000,用的数据集是分类的9000类ImageNet和检测的COCO
代码位置：http://pjreddie.com/yolo9000/

2.Better

主要改进YOLO定位错误和底的召回率，同时维持分类精度
批规范化：在收敛上有很大的提升，不需要其他的正则化方法，给所有的卷积层都加上批规范化，提高2%mAP,移除了dropout也不会导致正则化
BN的具体情况：http://blog.csdn.net/hjimce/article/details/50866313
高分辨率分类：原来的YOLO是在Image net用224×224的分辨率上训练分类网络，然后增加分辨率到448×448用于检测
YOLOv2首先使用10次448×448的Image net图片微调分类网络，然后再微调用于检测
提高mAP4%
Convolutional With Anchor Boxes：移除YOLO的全连接层，使用Anchor Boxes预测Bounding box ,首先消除一个池化层，让卷积层有更高的分辨率，把448×448（下采样因子是32，feature map是14×14）的图片缩小成416×416（下采样因子是32，feature map是13×13）,使得feature map的位置数为奇数，这样就有一个唯一的中心
通过在卷积层使用anchor boxes，网络可以预测超过1000个窗口，使用Anchor Boxes 精确率是降低，但是recall会从81%上涨到88%
维度聚类： 之前Anchor Box的尺寸是手动选择的，所以尺寸还有优化的余地。 为了优化，在训练集（training set）Bounding Boxes上跑了一下k-means聚类，来找到一个比较好的值。
如果我们用标准的欧式距离的k-means，尺寸大的框比小框产生更多的错误。因为我们的目的是提高IOU分数，这依赖于Box的大小，所以距离度量的使用：

$$d(\text{box}, \text{centroid}) = 1 - \text{IOU}(\text{box}, \text{centroid})$$

直接位置预测

模型不稳定

最终，网络在特征图（13 *13 ）的每个cell上预测5个bounding boxes，每一个bounding box预测5个坐标值：tx，ty，tw，th，to。如果这个cell距离图像左上角的边距为（cx，cy）以及该cell对应的box维度（bounding box prior）的长和宽分别为（pw，ph），那么对应的box为：

$$
\begin{aligned}
b_x &= \sigma(t_x) + c_x \\
b_y &= \sigma(t_y) + c_y \\
b_w &= p_w e^{t_w} \\
b_h &= p_h e^{t_h} \\
Pr(\text{object}) * IOU(b, \text{object}) &= \sigma(t_o)
\end{aligned}
$$

使用Dimension Clusters和Direct location prediction这两项anchor boxes改进方法，mAP获得了5%的提升

细粒度特征