Yolo：

与FasterRcnn不同：

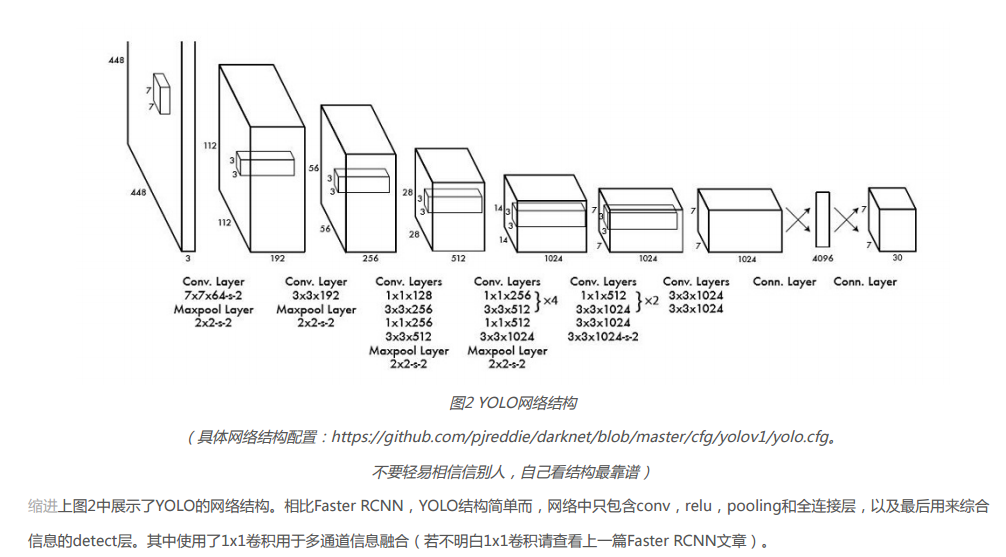
1.FasterRcnn将目标检测分解为分类问题与回归问题：首先采用独立的rpn网络求取region proposal，然后利用bounding box regression对提取的位置进行修正，最后采用softmax进行分类。

2.YOLO将物体预测作为一个回归问题进行求解，将图像输入一次网络，便能够得到图像中所有物体的位置和其所属类别对应的置信概率

优点：yolo将检测问题整合为一个回归问题，是网络简单，训练速度加快

缺点：1.输入图像尺寸固定，因为最后一层是全联接层

2.占比较小的目标检测效果不好.虽然每个格子可以预测B个bounding box，但是最终只选择只选择IOU最高的bounding box作为物体检测输出，即每个格子最多只预测出一个物体。当物体占画面比例较小，如图像中包含畜群或鸟群时，每个格子包含多个物体，但却只能检测出其中一个。



YOLO思想：

1.将原图划分为SxS的网格，如果一个目标的中心落入某个格子，这个格子就负责检测该目标

2.每个网格要预测B个bounding boxes以及c个类别的概率，在yolo中每个格子只有一个c类别，即相当于忽略了B个bounding boxes，每个格子只判断一次类别

3.每个bounding box除了要回归自身的位置之外，还要预测一个confidence值，这个值代表了所预测的box中所含有目标的置信度和这个bounding box预测有多准



有目标落在中心格子里Pr(obiect)=1,否则Pr（object）=0,第二个是预测的框与GT之间的IOU。所以每个bounding box都包含了五个预测量：（x,y,w,h,confidence）,其中（x,y）代表预测box相对于格子的中心，（w,h）为预测的宽度相对于图片width和height的比例。

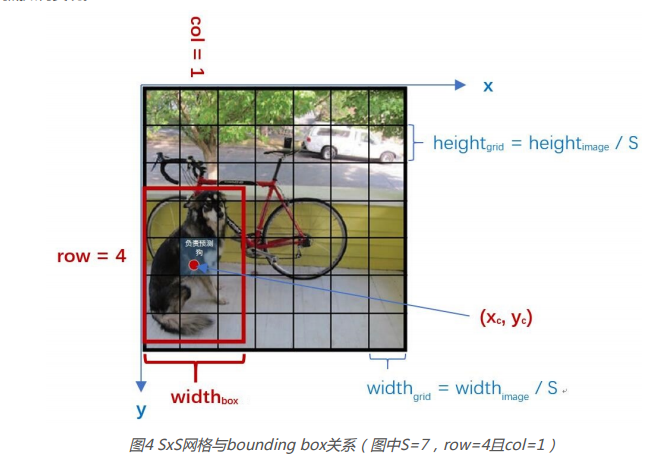
4.每个网格预测的类别条件概率和bounding box预测的confidence信息相乘，就得到，每个bounding box的class-specific-confidence score:



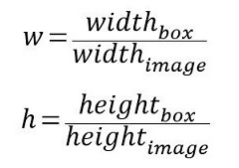
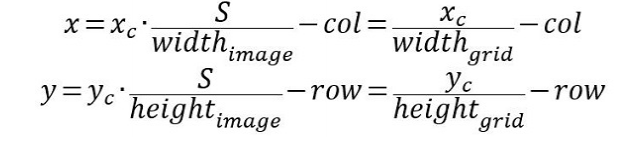
最后设置一个阈值，去掉小于阈值的值，然后对留下的进行非极大值抑制，最终得到检测框。

YOLO中的bounding box normalization:

对bound box的坐标（x,y,w,h）进行了normalization,以便进行回归。



中心落到了（row,clo）中，这个网格复制检测狗这个目标，那么归一化后的坐标为：



YOLO的损失函数：

1.坐标预测误差：大的框预测稍差一点还能接受，但是小的框预测差一点就不能接受了，所以这里使用开根号进行回归

2.confidence预测误差：由于大多数网格中不含目标，导致大多数confidence=0，所以同等对待函目标的box与不含目标的box是不对的，所以在不含目标的confidence预测误差中乘以权重0.5，还有同等对待4个坐标与一个confidence误差也不合理，所以作者在坐标预测误差前乘以权重5

