Yolo

第一阶段：Yolo V1

思想：把目标检测看成回归问题，直接从图片中预测边界框和类别

特点：端到端，实时

优点：快，预测会利用全局图像，学习目标的概括性表现

缺点：与最新的检测系统相比，YOLO会产生更多的定位错误，但预测背景的假阳性的可能性较小。精度落后，对小目标定位不准

算法概括：把输入图像分成S\*S个方格，如果一个目标的中心掉进了一个方格内，那这个方格就代表那个检测到的目标

误区：每个网格单元的视野有限而且很可能只有局部特征，这样就很难理解yolo为何能检测比grid cell大很多的物体。其实，yolo的做法并不是把每个单独的网格作为输入feed到模型，在inference的过程中，网格只是物体中心点位置的划分之用，并不是对图片进行切片，不会让网格脱离整体的关系。

1、每个方格预测B个边界框以及边界框的置信度，

置信度：Pr(object)\*IOU(truth vs pred)，

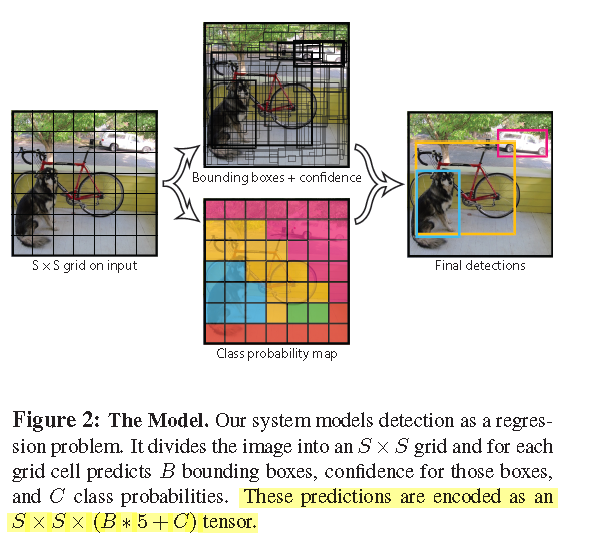
边界框有五个预测值：x,y,w,h和置信度

2、每个方格预测C的条件类别概率：Pr(Classi|Object)

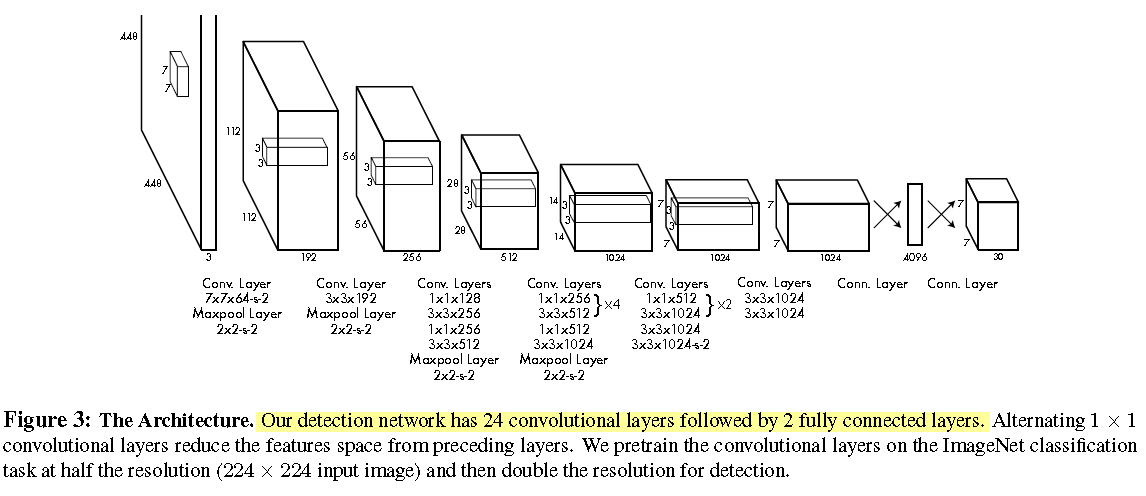
类别置信度：

Pr(Classi|Object)\*Pr(Object)\*IOU(truth vs pred)=Pr(Classi)\*IOU(truth vs pred)

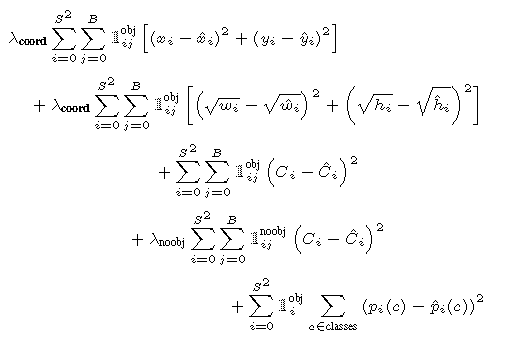
预测数量：S\*S\*(B\*(4+1)+C)



3、网络结构：



4、Loss function：



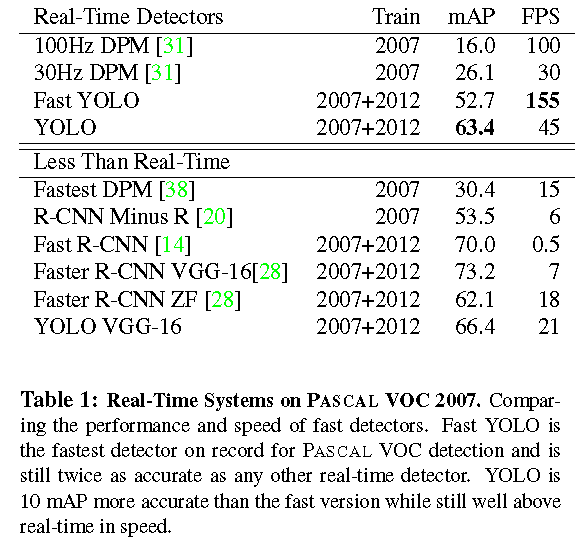
5、Yolo的主要问题：

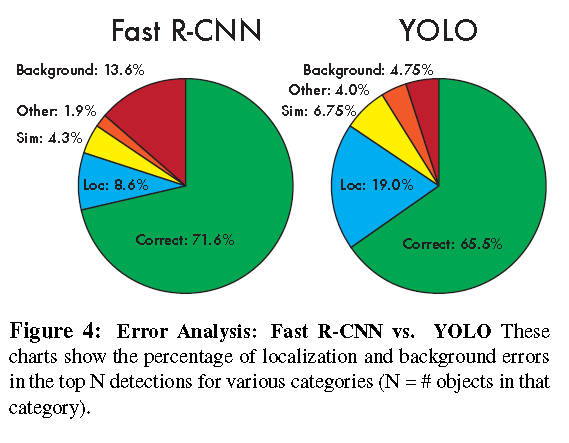
1、每个grid只能负责预测一个物体，当物体比较密集时会出现漏检

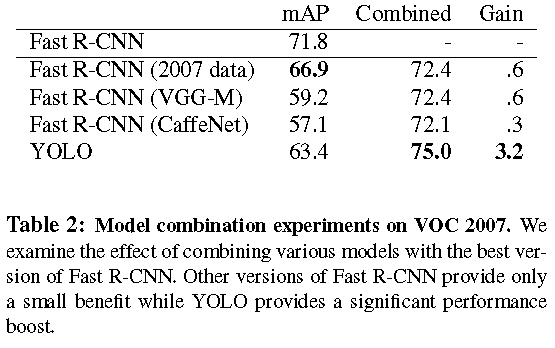
2、目标的bb是基于训练集训出来的，当测试集中出现一个新的或者不常见的纵横比的物体或者不常见构造的物体时，容易误检

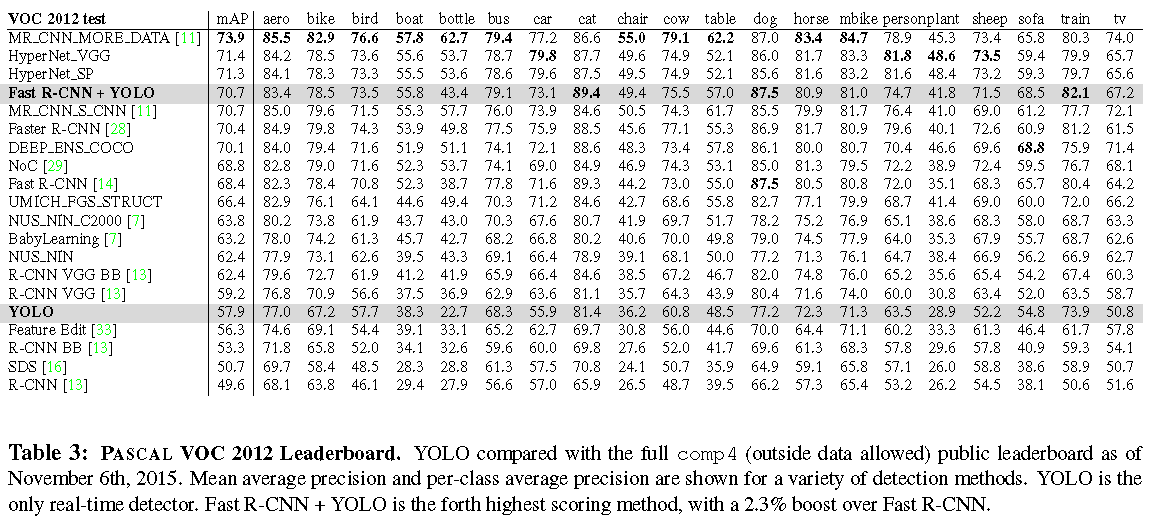
3、对待大目标和小目标的loss一致，错误主要源于不准确的定位

6、实验结果：









第二阶段：Yolo V2

思想：

几处改进

1、使用一个新的多尺度训练方法，v2能变化尺度运行，在速度和准确度上很好的平衡

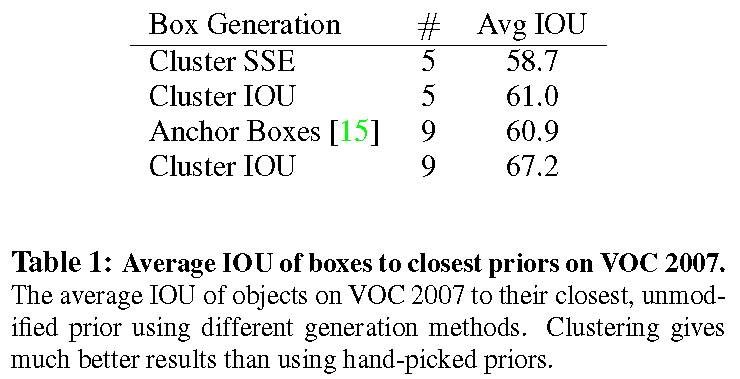
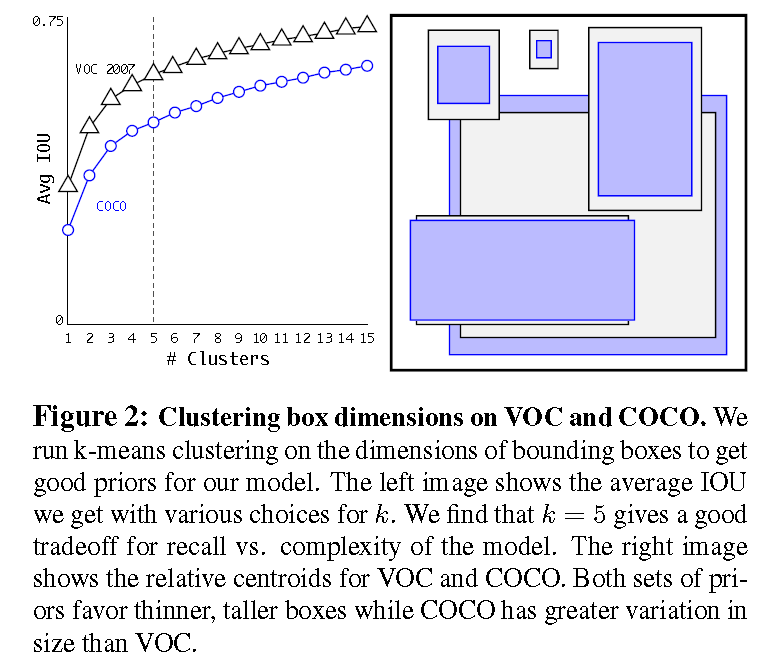
2、提出一种在检测数据集和分类数据集上联合训练

3、仍然能实时运行

一、Better

1、批归一化

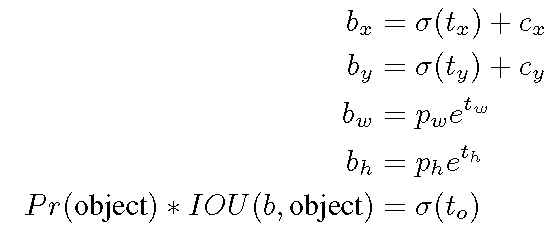
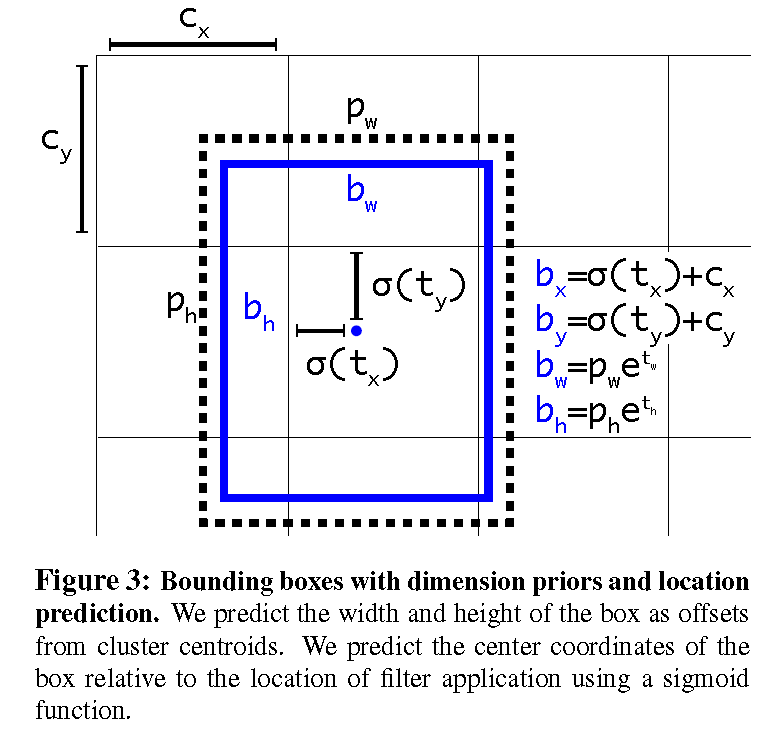
1. 高分辨率分类器：224\*224 ----> 448\*448
2. 用anchor box做卷积
3. 维度聚类



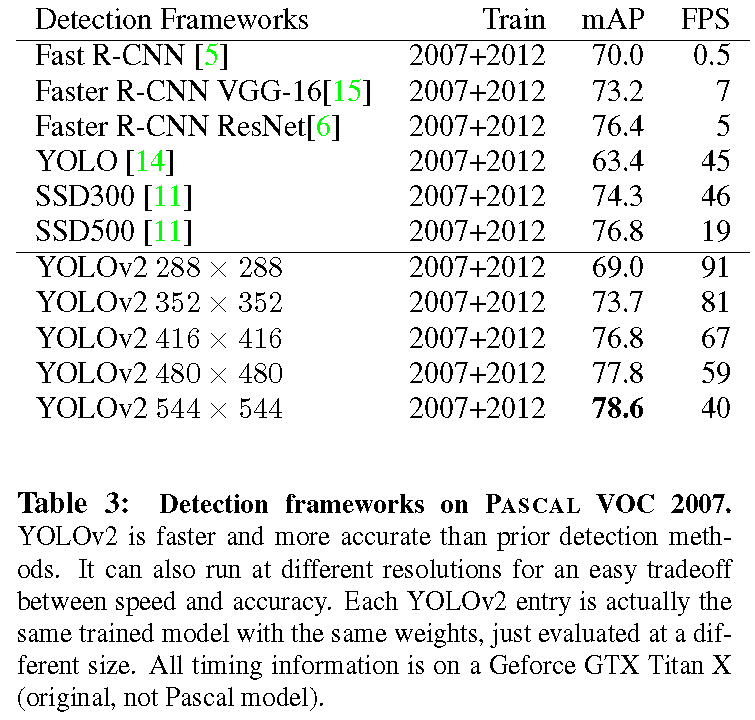
1. 直接位置预测

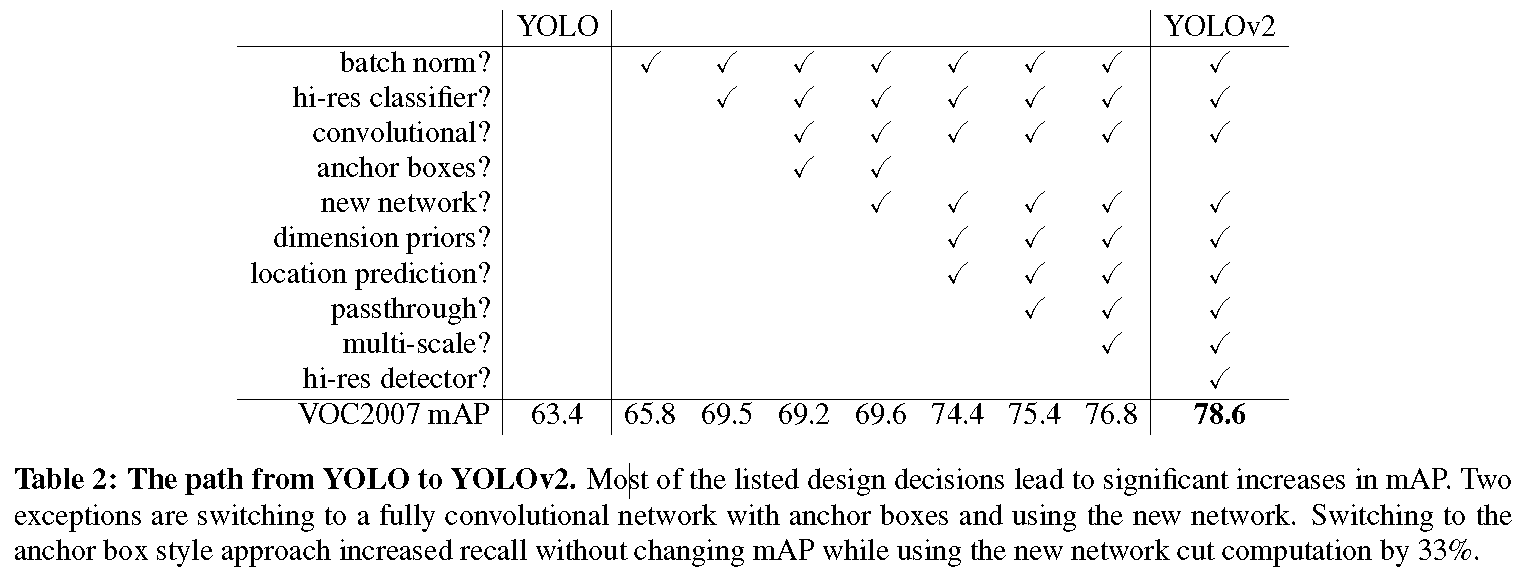
数量：5个边界框

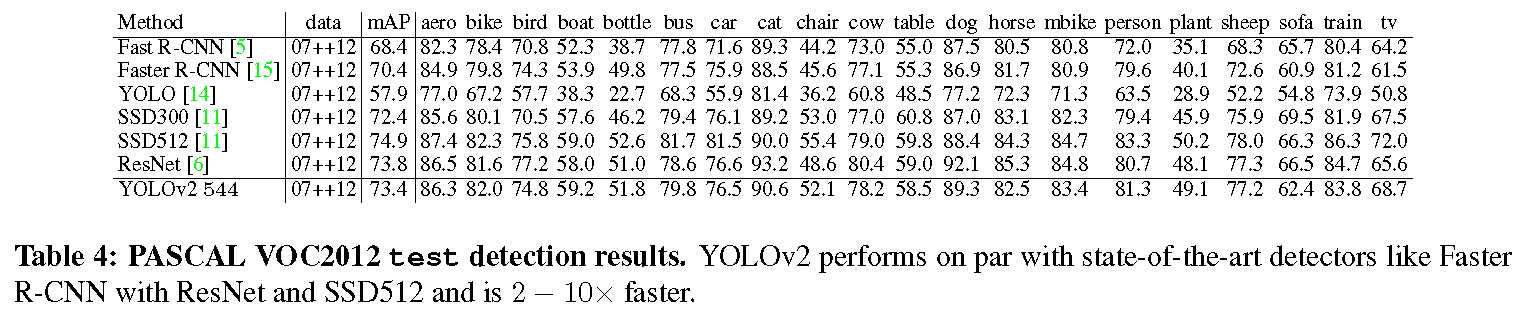
坐标计算：

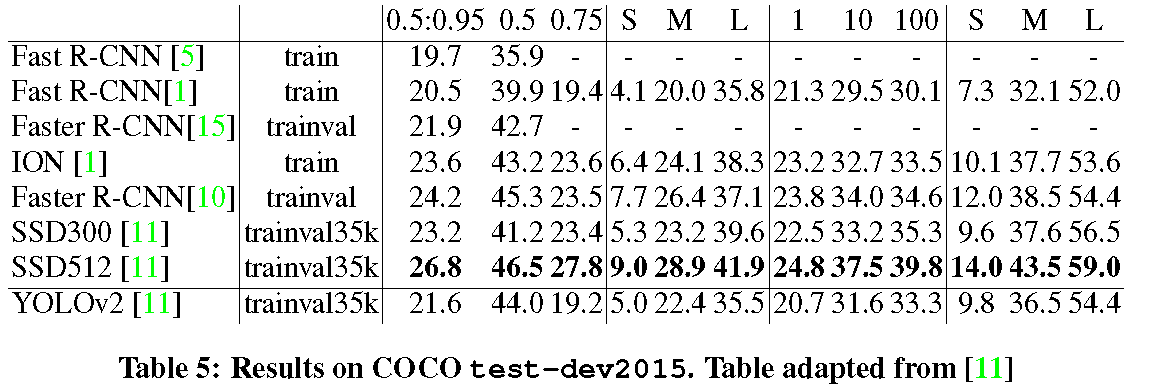


1. 有细密纹理的特征
2. 多比例训练





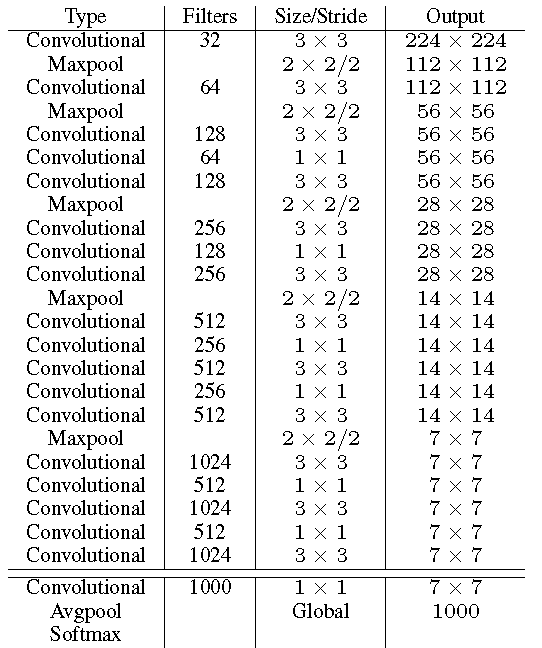




二、Faster

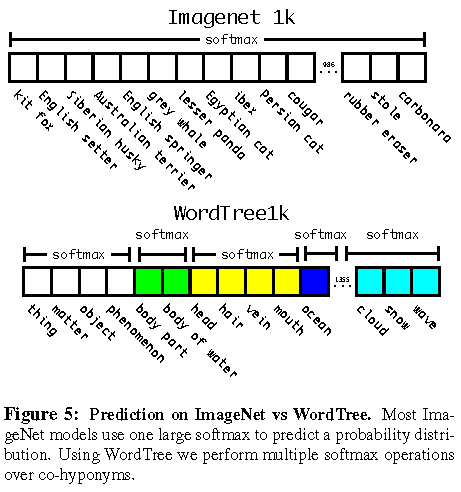
模型：

Darknet-19

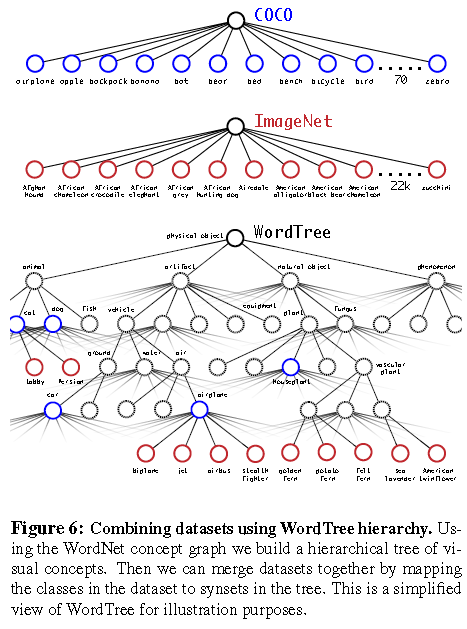


三、Stronger

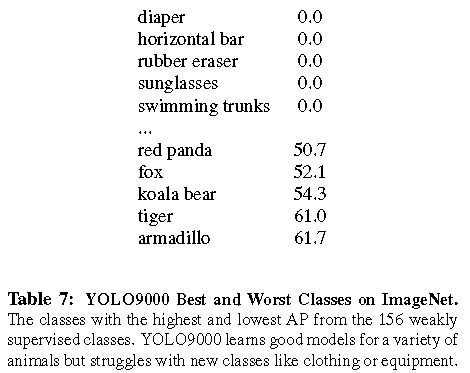
1. 分级分类



1. 使用世界树对数据集合并

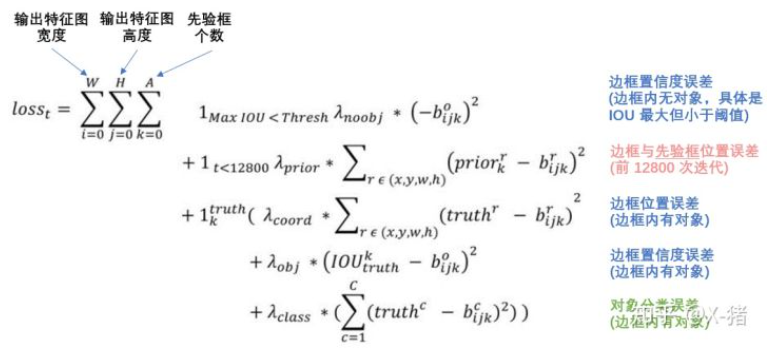


1. 联合分类和检测数据集



补充：

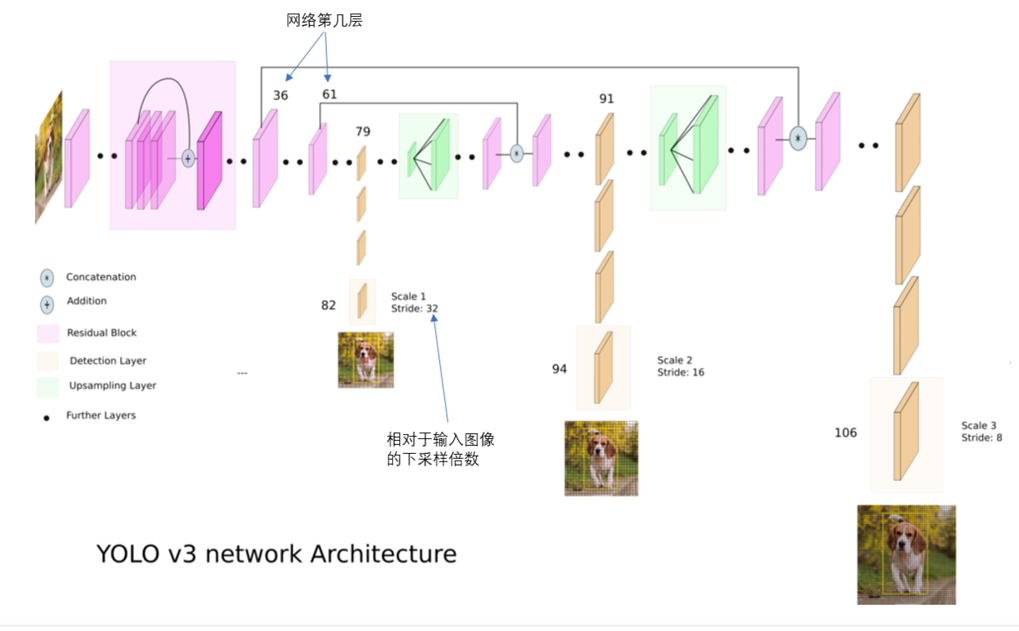
损失函数：



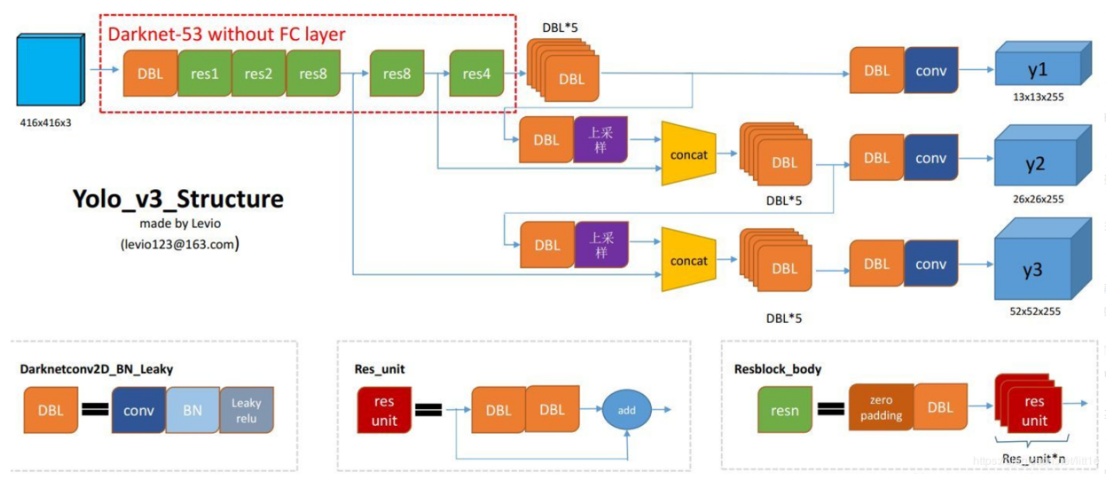
第三阶段：Yolo V3

网络结构：

第一种可视化：



第二种可视化：



y1,y2,y3是如何而来的？

网络中作者进行了三次检测，分别是在32倍降采样，16倍降采样，8倍降采样时进行检测,这样在多尺度的feature map上检测跟SSD有点像。在网络中使用up-sample（上采样）的原因:网络越深的特征表达效果越好，比如在进行16倍降采样检测，如果直接使用第四次下采样的特征来检测，这样就使用了浅层特征，这样效果一般并不好。如果想使用32倍降采样后的特征，但深层特征的大小太小，因此yolo\_v3使用了步长为2的up-sample（上采样），把32倍降采样得到的feature map的大小提升一倍，也就成了16倍降采样后的维度。同理8倍采样也是对16倍降采样的特征进行步长为2的上采样，这样就可以使用深层特征进行detection。

作者通过上采样将深层特征提取，其维度是与将要融合的特征层维度相同的（channel不同）。如下图所示，85层将13×13×256的特征上采样得到26×26×256，再将其与61层的特征拼接起来得到26×26×768。为了得到channel255，还需要进行一系列的3×3，1×1卷积操作，这样既可以提高非线性程度增加泛化性能提高网络精度，又能减少参数提高实时性。52×52×255的特征也是类似的过程。

v3的输出是至关重要的。

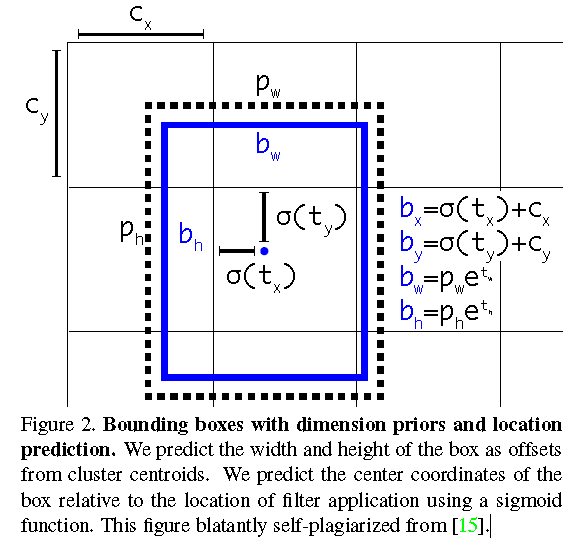
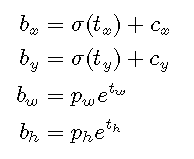
第一点， 9个anchor会被三个输出张量平分的。根据大中小三种size各自取自己的anchor。

第二点，每个输出y在每个自己的网格都会输出3个预测框，这3个框是9除以3得到的，这是作者设置的，我们可以从输出张量的维度来看，13x13x255。255是怎么来的呢，3\*(5+80)。80表示80个种类，5表示位置信息和置信度，3表示要输出3个prediction。在代码上来看，3\*(5+80)中的3是直接由num\_anchors//3得到的。

第三点，作者使用了logistic回归来对每个anchor包围的内容进行了一个目标性评分(objectness score)。根据目标性评分来选择anchor prior进行predict，而不是所有anchor prior都会有输出。

解决的问题：

2.1边界框预测



2.2类别预测

使用独立的逻辑分类器

为什么不适用softmax？

Softmax要求数据标签最好是互斥的，而实际数据集中包含许多重叠的标签，例如女人和人

2.3跨尺度的预测

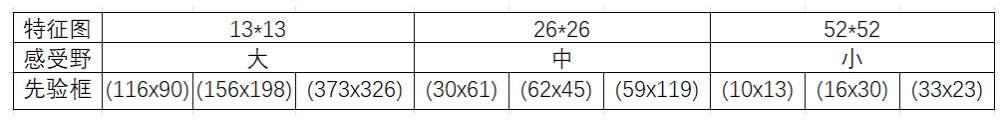
提出三个不同尺度的box

预测：N\*N\*[3\*(4+1+80)]

9类3个尺度：(10\*13),(16\*30),(33\*23),(30\*61),(62\*45),(59\*119)

(116\*90),(156\*198),(373\*326)

proposal box：（52×52+26×26+13×13）×3=10647个



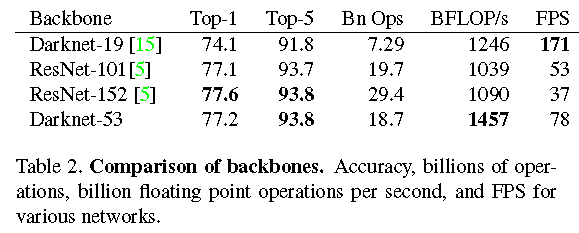
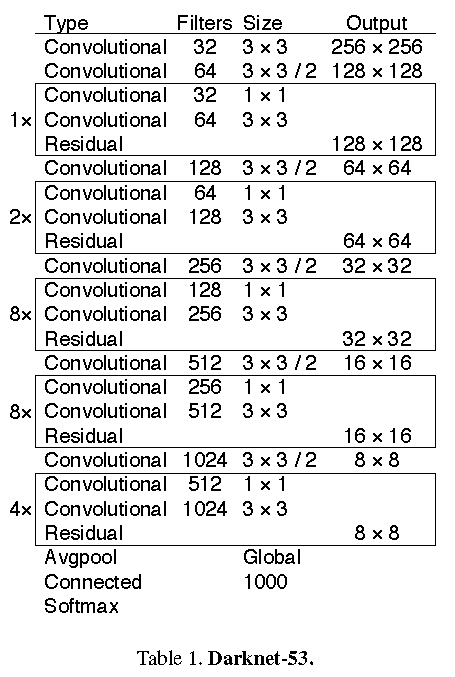
bounding box 与anchor box的区别：

Bounding box它输出的是框的位置（中心坐标与宽高），confidence以及N个类别。

anchor box只是一个尺度即只有宽高。

2.4特征提取

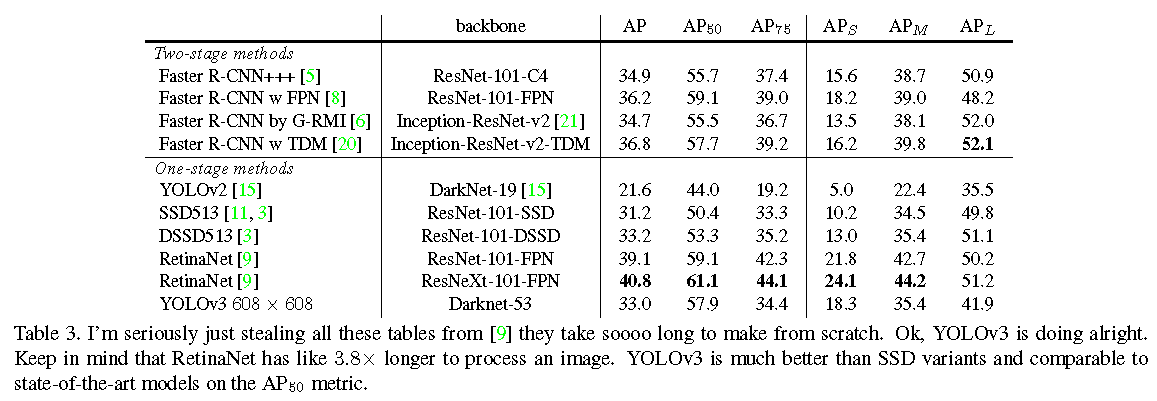
网络结构：



2.5训练

We still train on full images with no hard negative mining or any of that stuff

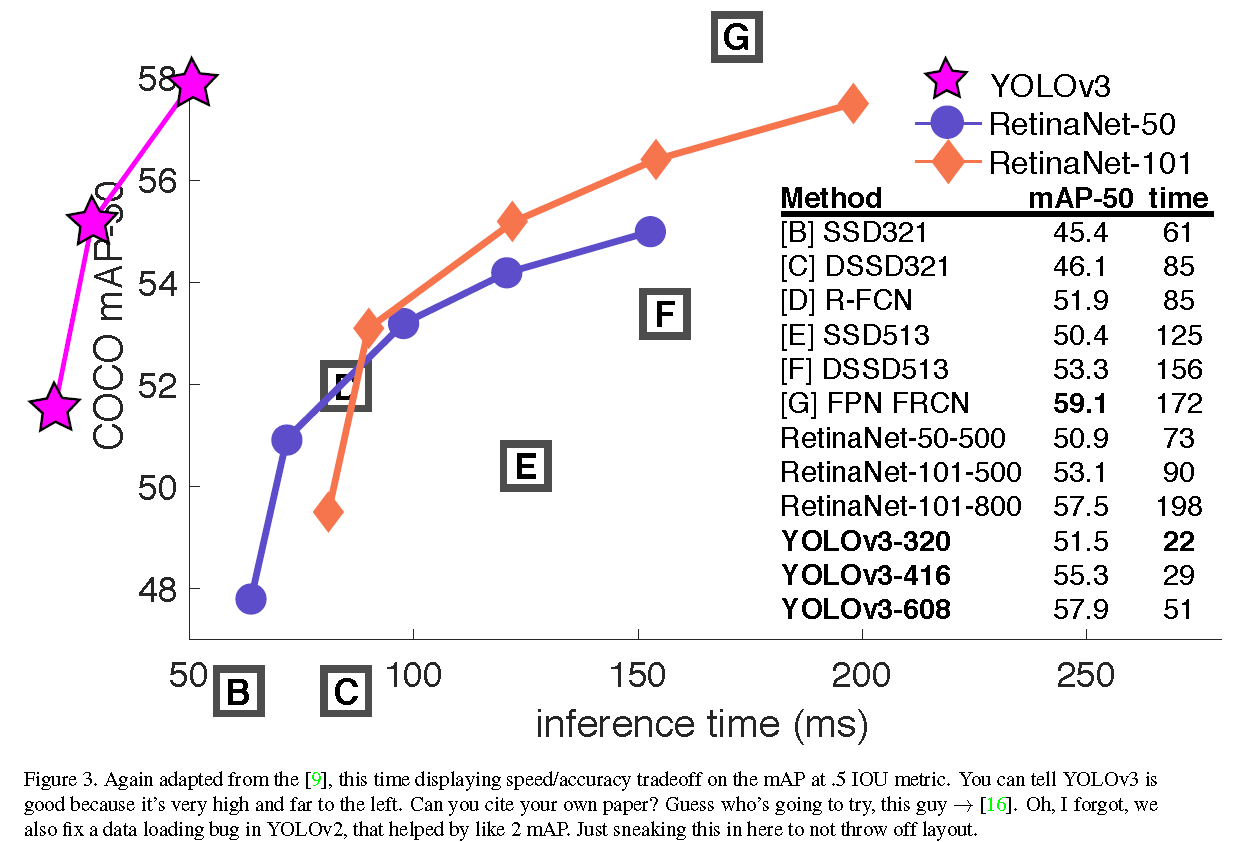
3、成果



缺点：

1、当IOU阈值增加时，性能会显着下降，这表明YOLOv3难以使框与对象完美对齐。

1. 在加入多尺度预测后，yolo在小目标上表现较好，而在中等和大目标上表现相对较差



4、尝试

1.锚框x,y偏移量预测

2.线性x,y预测而不是logistic

3.focus loss

YOLOv3 may already be robust to the problem focal loss is trying to solve because it has separate objectness predictions and conditional class predictions.

1. 双重IOU阈值和真相分配

思考？

1、Yolo V3在使用多尺度特征检测目标时使用率深层特征上采样来得到浅层特征，按照以往经验，利用深层特征检测大目标，利用浅层特征检测小目标，但是yolo在检测中大型目标上提升并不明显，反而在检测小目标上得到了提升

2、对自有数据集聚类确定b-box priors的数量和大小

3、根据特殊情景下某些物体的特性，设置focal loss来提高对特殊物体的检测，提高模型精度

目前工作计划：

长期目标：根据实际场景要求，提高通用检测模型的实际能力

三个月安排 ：

1. 对现有数据集进行分析，根据数据集中物体的尺寸数据重新设计先验anchor重新训练，根据实验结果制定具体改进措施

原因：原本模型中的先验anchor的设置是根据coco数据集设置的，与自有数据集的情况可能不一致

1. 根据特殊情景下某些物体的特性，设置focal loss来提高对特殊物体的检测，提高模型精度

原因：对于一些特殊尺寸特殊宽高比的物体，现有模型检测精度有限，造成误报，通过调整训练时loss可能会对特殊物体产生检测精度的提高有所帮助