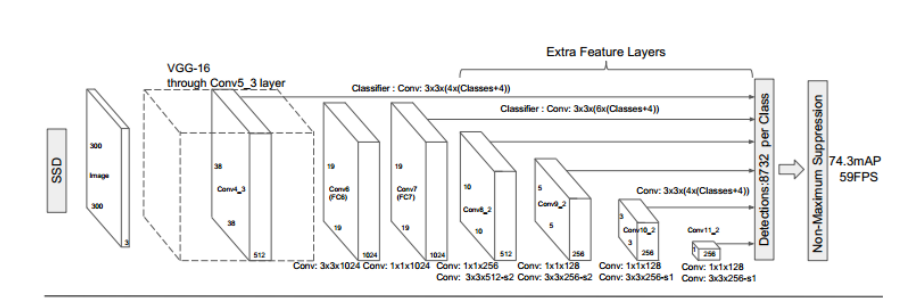
SSD:  
特点：1.继承了yolo中将detection转化为regression的思路，所以一次可以完成网络训练

2. 基于faster Rcnn提出的anchor，提出了相似的prior box

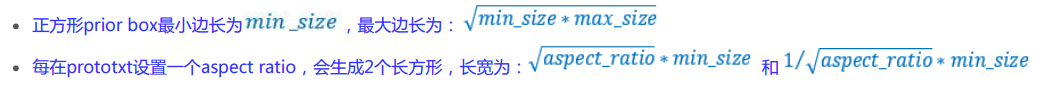
3.加入基于特征金字塔的检测方式（利用了前面的特征进行分类与回归）



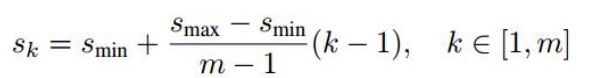
Prior Box:

就是一些目标预选框，后面通过softmax 与bounding box regression 获得真实的目标位置，SSD按下面规则生成prior box.

以feature map 上每个点中为中心，生成一系列的prior box（然后坐标乘以step，相当于从feature map映射回原图）



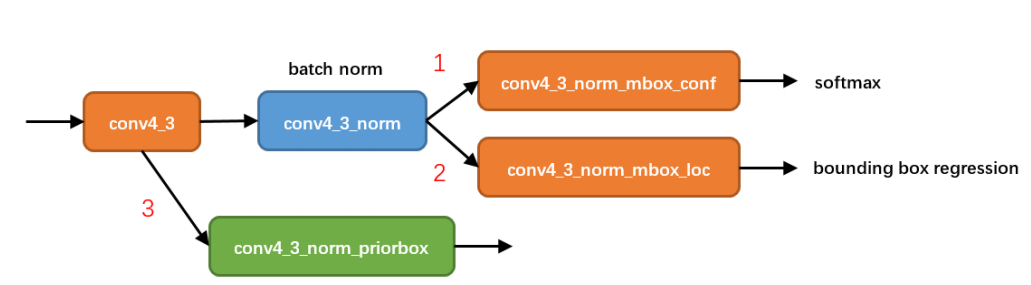
不同层上的特征图的检测框大小是不同的，对应一个计算公式



M是使用feature map 的数量，第一层的feature map 对应的min\_size = S1,max\_size = S2

第二层min\_size = S2,max\_size = S3,以此类推。这样较低层可以检测小目标，较高层检测较大目标。

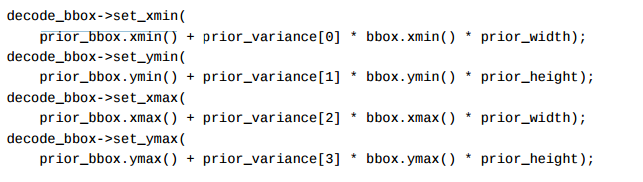
每个特征图是如何对prior box 进行分类与回归的：



第一条线：经过batch norm + 一次卷积，生成了[1,num\_class\*num\_priorbox,layer\_height,layer\_wdith]的特征图，用于softmax分类。

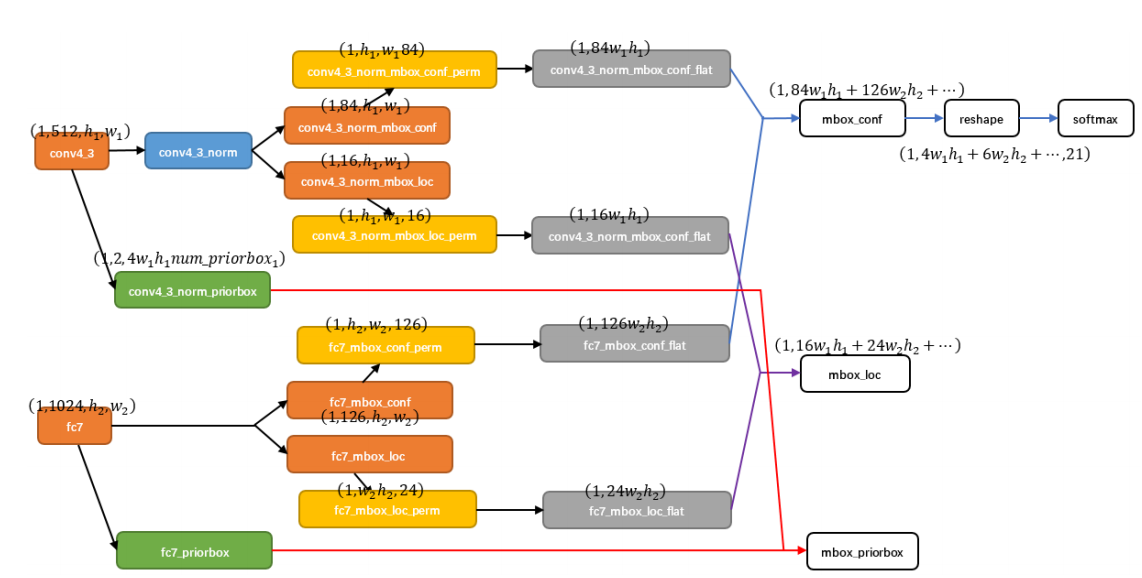
第二条线：经过batch norm +一次卷积后，生成了[1,4\*num\_priorbox,layer\_height,layer\_width]

大小的特征图用于bounding box regression

第三条线：生成了[1,2,4\*num\_priorbox]大小的prior box blob,其中两个channel分别存储prior box的4个坐标和对应的4个variance，这四个值是bounding regression中的权重，线路二中返回四个regerssion:[dxmin,dymin,dxmax,dymax],然后用下面的方法对prior box进行位置更新，

网络训练：

上面展示了一个特征图进行softmax+box regression，但是在SSD中一共有6个特征图进行分类与回归，所以不能一个一个单独计算。



展示了两个feature map合并进行softmax+box regession。

Conv4\_3每个点有4个prior box ,有21类，所以是 4\*21 = 84，fc7每个点有6个prior box 所以是21 \* 6 = 126，然后转化为（1,84\*w1\*h1）,(1,126\*h2\*w2),然后再进行拼接。