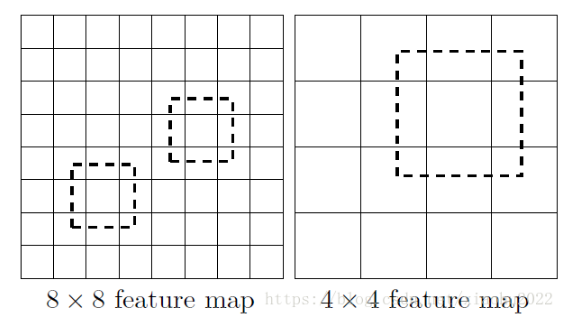
SSD算法

**设计理念：**

**1）采用多尺度特征图用于检测**

所谓多尺度采用大小不同的特征图，CNN网络一般前面的特征图比较大，后面会逐渐采用stride=2的卷积或者pool来降低特征图大小，多尺度就是用比较大的特征图来用来检测相对较小的目标，而小的特征图负责检测大目标，

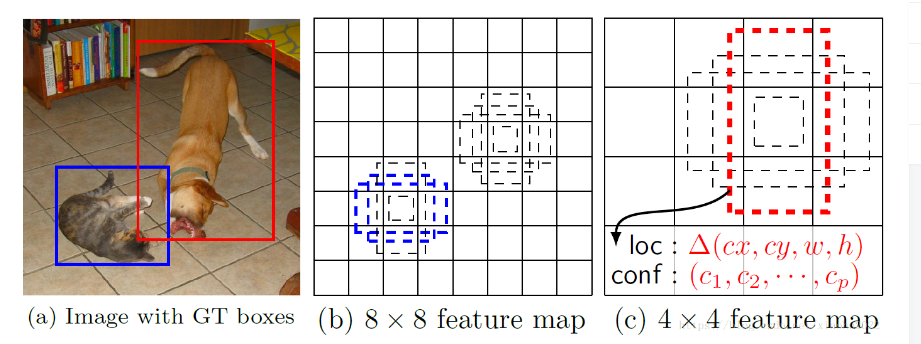


**（2）采用卷积进行检测**

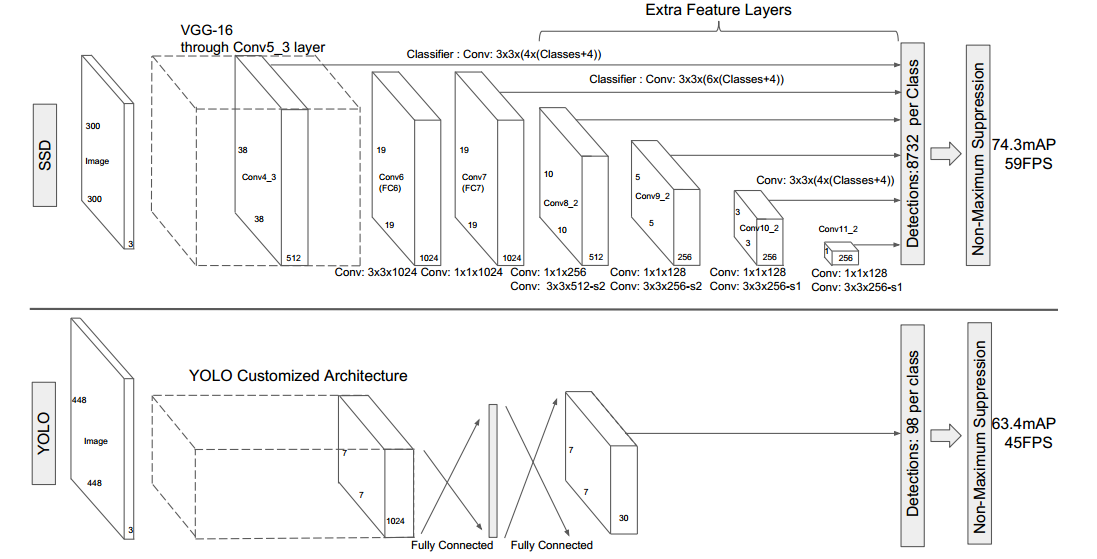
与Yolo最后采用全连接层不同，SSD直接采用卷积对不同的特征图来进行提取检测结果。对于形状为m×n×p的特征图，只需要采用3×3×p这样比较小的卷积核得到检测值。

**（3）设置先验框**

SSD借鉴了Faster R-CNN中anchor的理念，每个单元设置尺度或者长宽比不同的先验框，预测的边界框（bounding boxes）是以这些先验框为基准的，在一定程度上减少训练难度。一般情况下，每个单元会设置多个先验框，其尺度和长宽比存在差异，如下图所示，可以看到每个单元使用了4个不同的先验框。



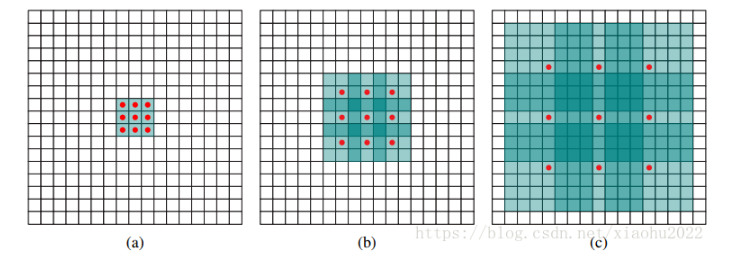
**网络结构**



SSD采用VGG16作为基础模型，然后在VGG16的基础上新增了卷积层来获得更多的特征图以用于检测。SSD的网络结构如图6所示。上面是SSD模型，下面是Yolo模型，可以明显看到SSD利用了多尺度的特征图做检测。

扩张卷积

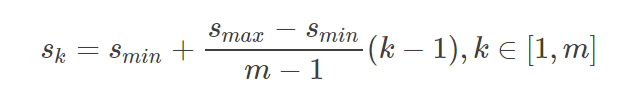
conv6采用扩展卷积或带孔卷积（Dilation Conv），其在不增加参数与模型复杂度的条件下指数级扩大卷积的视野，其使用扩张率(dilation rate)参数，来表示扩张的大小，如下图7所示，(a)是普通的3×3×3卷积，其视野就是3×3×3，(b)是扩张率为2，此时视野变成7×7×7，(c)扩张率为4时，视野扩大为15×15，但是视野的特征更稀疏了。Conv6采用3×3×3大小但dilation rate=6的扩展卷积。



从后面新增的卷积层中提取Conv7，Conv8\_2，Conv9\_2，Conv10\_2，Conv11\_2作为检测所用的特征图，加上Conv4\_3层，共提取了6个特征图，其大小分别是(38,38),(19,19),(10,10),(5,5),(3,3),(1,1)(38,38),(19,19),(10,10),(5,5),(3,3),(1,1)，但是不同特征图设置的先验框数目不同。先验框的设置，包括尺度（或者说大小）和长宽比两个方面。

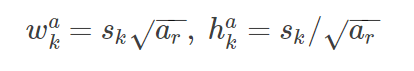
**先验框的计算**

对于先验框的尺度，其遵守一个线性递增规则：随着特征图大小降低，先验框尺度线性增加，假设我们用m个特征图去预测：



表示先验框大小相对于图片的比例，论文里取值，和表示最低层的比例为0.2与最高层的比例为0.9

对于特定的长宽比，按如下公式计算先验框的宽度与高度（后面的均指的是先验框实际尺度，而不是尺度比例）：



令为该特征图所采用的先验框数目，那么类别置信度需要的卷积核数量为，而边界框位置需要的卷积核数量为。由于每个先验框都会预测一个边界框，所以SSD300一共可以预测38×38×4+19×19×6+10×10×6+5×5×6+3×3×4+1×1×4=8732边界框。

**先验框匹配**

在训练过程中，首先要确定训练图片中的ground truth（真实目标）与哪个先验框来进行匹配，与之匹配的先验框所对应的边界框将负责预测它

原则一：对于图片中每个ground truth，找到与其IOU最大的先验框，该先验框与其匹配，这样，可以保证每个ground truth一定与某个先验框匹配。通常称与ground truth匹配的先验框为正样本，反之，若一个先验框没有与任何ground truth进行匹配，那么该先验框只能与背景匹配，就是负样本

第二个原则是：对于剩余的未匹配先验框，若某个ground truth的IOU大于某个阈值（一般是0.5），那么该先验框也与这个ground truth进行匹配。这意味着某个ground truth可能与多个先验框匹配

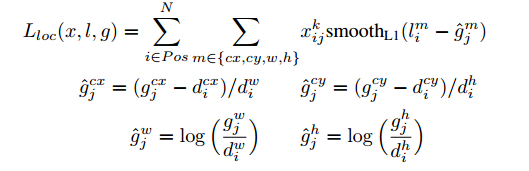
注：第二个原则一定在第一个原则之后进行，仔细考虑一下这种情况，如果某个ground truth所对应最大IOU小于阈值，并且所匹配的先验框却与另外一个ground truth的IOU大于阈值，那么该先验框应该匹配谁，答案应该是前者，首先要确保某个ground truth一定有一个先验框与之匹配

**损失函数：**

损失函数定义为位置误差（locatization loss）与置信度误差（confidence los）的加权和： 这里为一个指示参数，当时表示第i个先验框与第j个ground truth匹配，并且ground truth的类别为p：



位置误差仅针对正样本进行计算，对于位置误差，其采用Smooth L1 loss，定义如下：



对于置信度误差，其采用softmax loss:

