Our approach, named SSD, discretizes the output space of bounding boxes into a set of default boxes over different aspect ratios and scales per feature map location. At prediction time, the network generates scores for the presence of each object category in each default box and produces adjustments to the box to better match the object shape. Additionally, the network combines predictions from multiple feature maps with different resolutions to naturally handle objects of various sizes

把输出空间，离散成一组默认框（具有不同的横纵比例），预测时，生成每个默认框中每个对象类别存在的分数，并对框进行调整，以更好匹配对象。此外，该网络结合了多个具有不同分辨率的特征映射的预测，可以自然地处理各种大小的对象。

优点：简单，易于训练，完全消除了建议区域生成的问题，和后续像素或特征重采样阶段。

For 300× 300in-put, SSD achieves 74.3% mAP1 on VOC2007 testat 59 FPS on a Nvidia TitanX and for 512× 512 input, SSD achieves 76.9% mAP, outperforming a comparable state-of-the-art Faster R-CNN model

在不同数据集上表现良好，达到74.3%的准确率，优于相对较新的Faster R-CNN

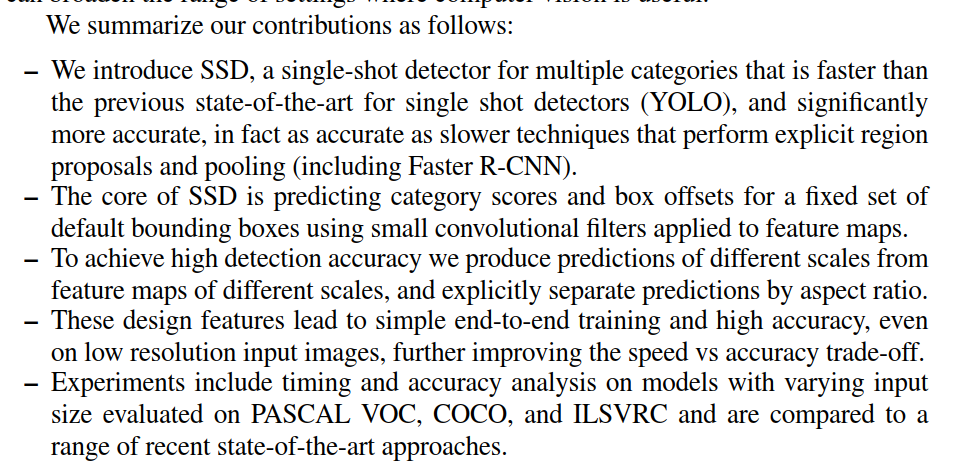
Often detection speed for these approaches is measured in seconds per frame (SPF), and even the fastest high-accuracy detector, Faster R-CNN, operates at only 7 frames per second (FPS). There have been many attempts to build faster detectors by attacking each stage of the detection pipeline (see related work in Sec. 4), but so far, signiﬁcantly increased speed comes only at the cost of signiﬁcantly decreased detection accuracy

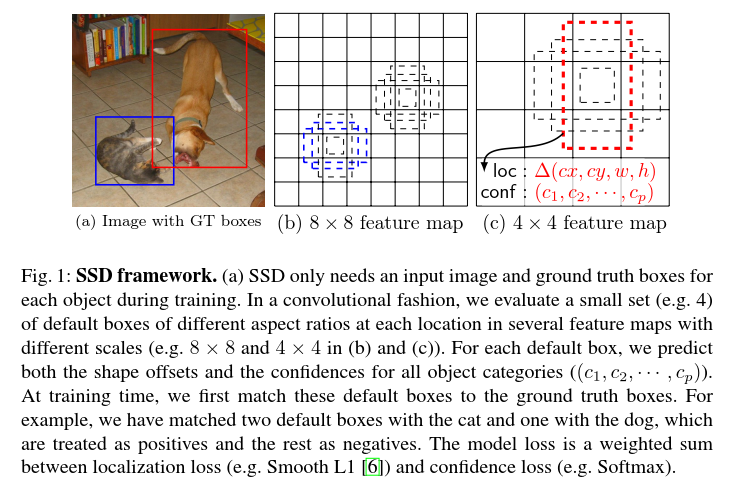
一般的检测速度以每秒的帧数来定义，即便是目前最快的Faster R-CNN，也只能以7FPS工作。有这方面提速的不少研究，但都以降低检测精度为代价。

Our improvements include using a small convolutional ﬁlter to predict object categories and offsets in bounding box locations, using separate predictors (ﬁlters) for different aspect ratio detections, and applying these ﬁlters to multiple feature maps from the later stages of a network in order to perform detection at multiple scales

改进：使用一个小卷积滤波器来预测边界框中对象类别和偏移量，为不同的衡纵比检测使用单独的预测器（滤波器），并将这些过滤器应用于网络后期的多个特征映射，以便在多个尺度上执行检测。

SSD模型的显著优点





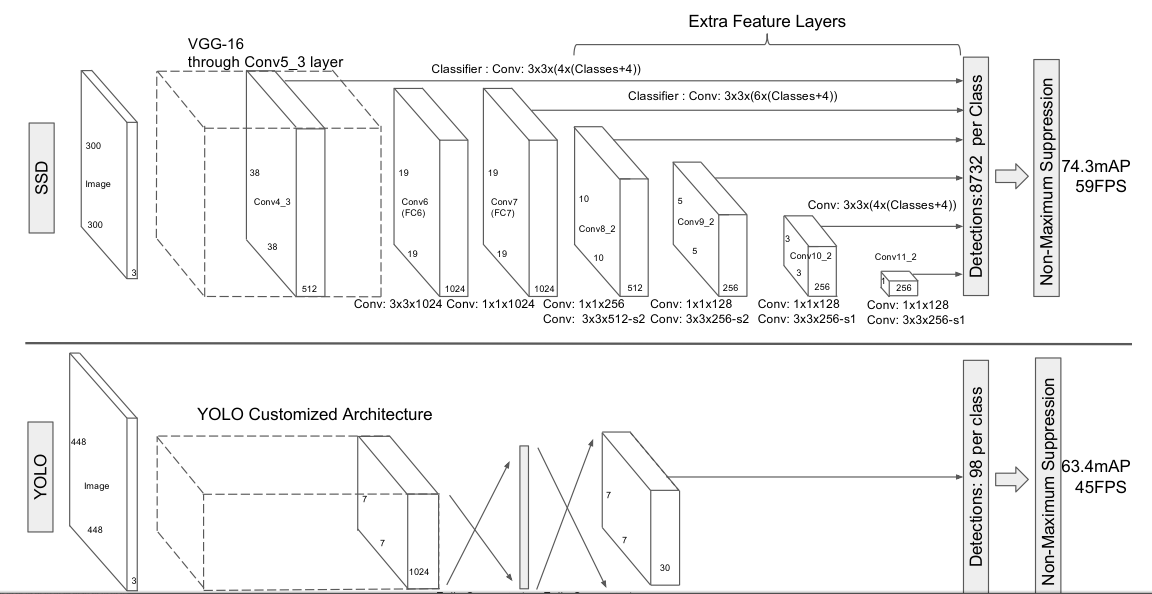
在卷积的方式中，我们在几个具有不同比例的feature map的每个位置计算一小组(例如4)不同纵横比的默认框

**算法模型框架的关键技术**

（1）Multi-scale feature maps for detection

（2）Convolutional predictors for detection

（3）Default boxes and aspect ratios



SSD网络的训练过程

1.**Matching strategy**:During training we need to determine which default boxes correspond to a ground truth detection and train the network accordingly. For each ground truth box we are selecting from default boxes that vary over location, aspect ratio, and scale. We begin by matching each ground truth box to the default box with the best jaccard overlap (as in MultiBox [7]). Unlike MultiBox, we then match default boxes to any ground truth with jaccard overlap higher than a threshold (0.5). This simpliﬁes the learning problem, allowing the network to predict high scores for multiple overlapping default boxes rather than requiring it to pick only the one with maximum overlap.

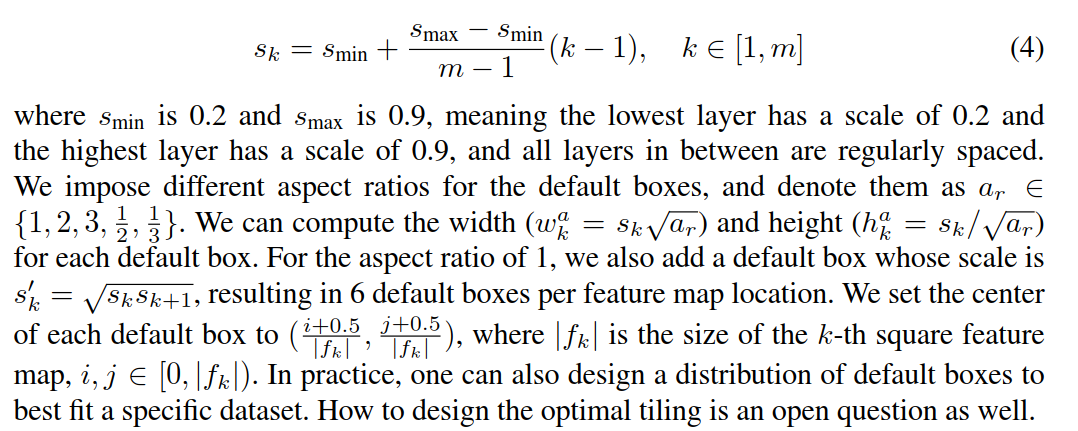
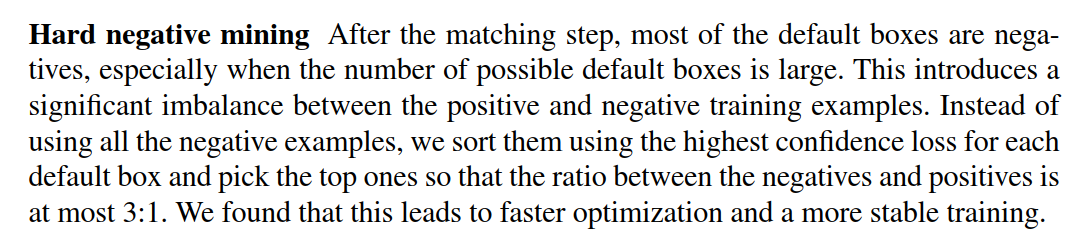
每个网格都有多个default boxes，给ground truth找到其最接近的default boxes。Jaccard overlap超过0.5门限。让网络预测多个default boxes。

2.借鉴语义分割中context相关结论

Previous works [10,11] have shown that using feature maps from the lower layers can improve semantic segmentation quality because the lower layers capture more ﬁne details of the input objects. Similarly, [12] showed that adding global context pooled from a feature map can help smooth the segmentation results

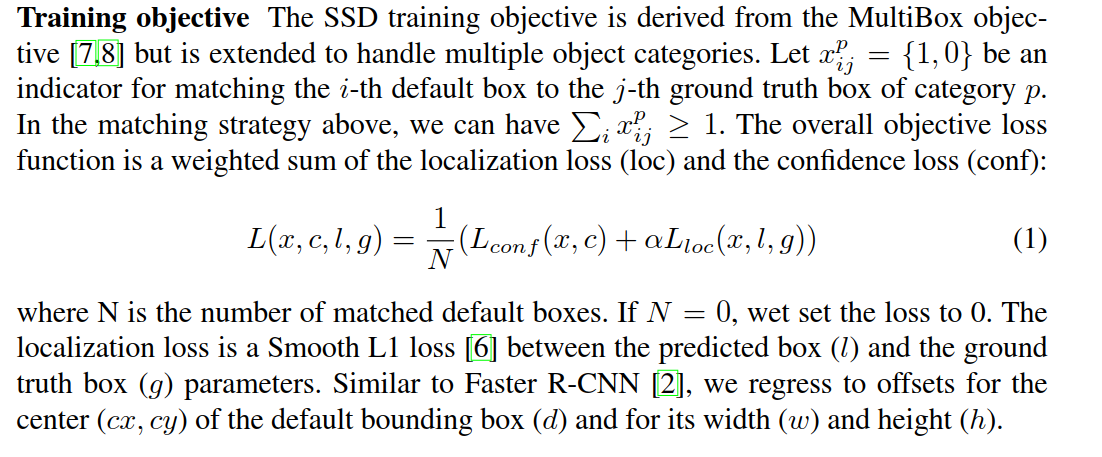
浅层的特征图可以改善语义分割的质量，因为底层次的特征图有更多的细节信息。

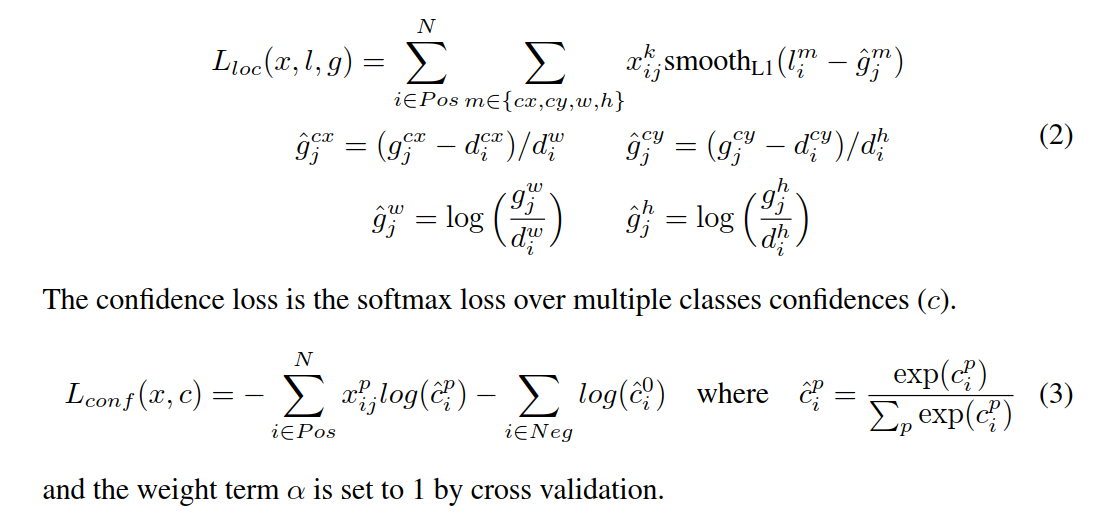
3.计算个feature map的格子需要的default boxes数目

****

同时，还应该去除negative的box，就是跟真实框，差距太大的default boxes。

4.损失函数





5.Compared to R-CNN [22], SSD has less localization error, indicating that SSD can localize objects better because it directly learns to regress the object shape and classify object categories instead of using two decoupled steps

相较于基于区域提名的目标检测算法而言，SSD算法有更低的定位误差。

6. SSD is very sensitive to the bounding box size. In other words, it has much worse performance on smaller objects than bigger objects.This is not surprising because those small objects may not even have any information at the very top layers. Increasing the input size (e.g. from 300 × 300 to 512 × 512) can help improve detecting small objects, but there is still a lot of room to improve

SSD对于包围盒的尺寸很敏感，不便于进行小目标检测。这并不奇怪，因为这些小目标即使在考前的卷积层，都很难提取到很多信息。可以通过放大图片来解决，但是效果不是那么明显。

