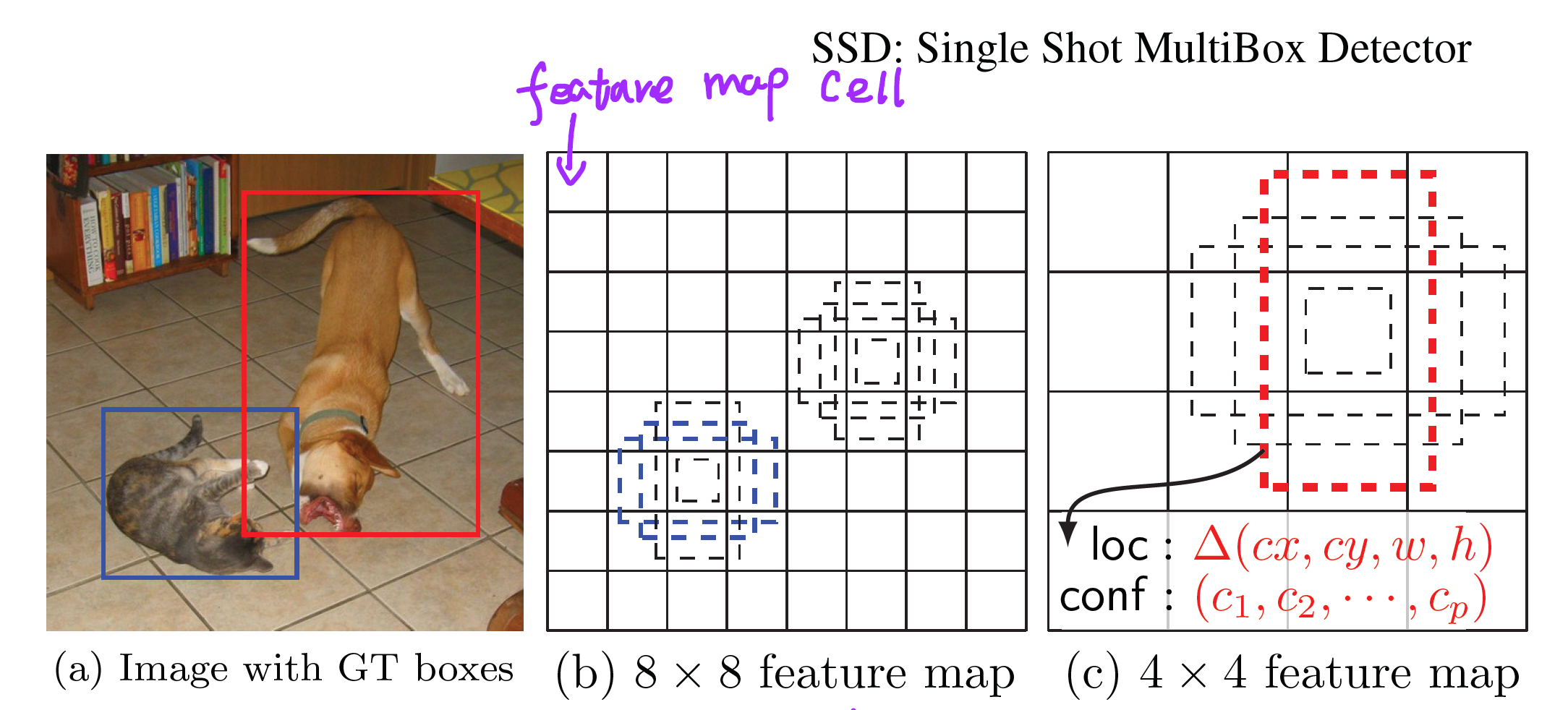
SSD Notes

1. **概述**

针对不同大小的目标检测，传统的做法是先将图像转换成不同大小（图像金字塔），然后分别检测，最后将结果综合起来（NMS）。而SSD算法则利用不同卷积层的feature map进行综合也能达到同样的效果。算法的基础网络结构是VGG16，并将最后两个全连接层改成卷积层，并随后增加了4个卷积层来构造网络结构。对其中5种不同的卷积层的输出（feature map）分别用两个不同的 3×3卷积核进行卷积，一个输出分类用的confidence，每个default box 生成21个类别confidence；一个输出回归用的 localization，每个default box生成4个坐标值（x, y, w, h）。上述5个feature map中每一层default box的数量是给定的，最后共生成8732个default box。

1. **一些概念**

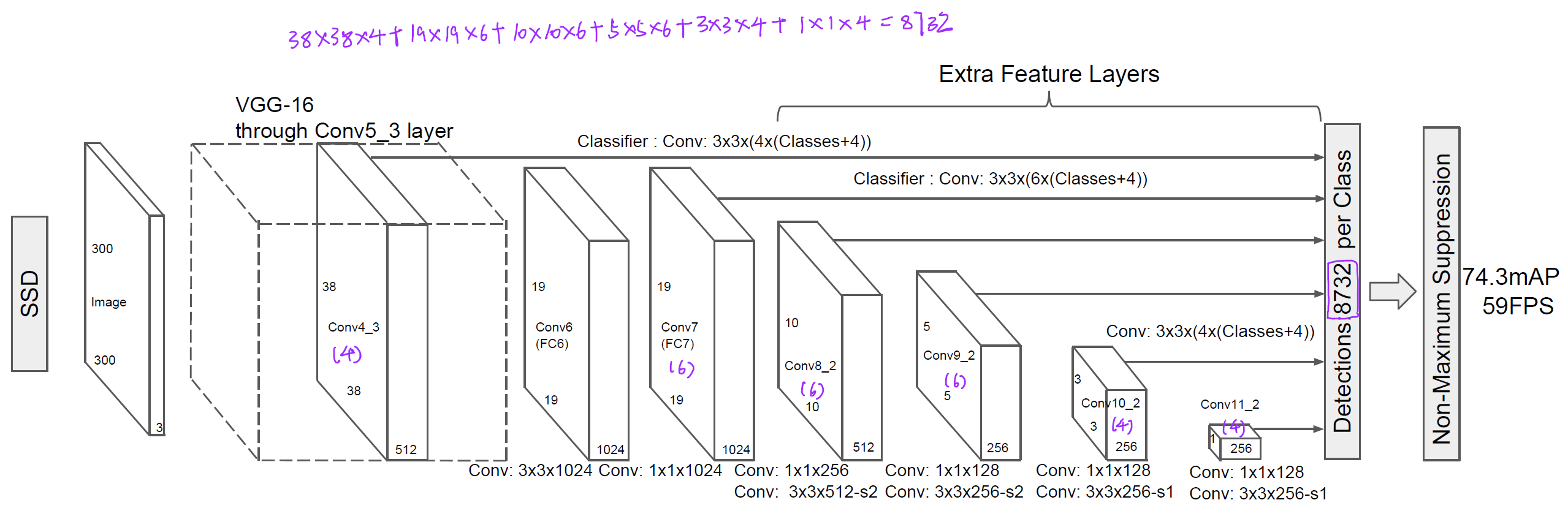


上图中（a）是画上GT boxes的原始图片，（b）是网络结构中的一个8\*8大小的特征图，（c）是网络结构中的一个4\*4大小的特征图。

第一个概念是**feature map cell**，是指feature map中每一个小格子，如图中分别有64和16个cell。另外有一个概念：**default box**，是指在feature map的每个小格(cell)上都有一系列固定大小的box，如上图中有4个（虚线框，仔细看格子的中间有比格子还小的一个box）。训练中还有一个东西：**prior box**，是指实际中选择的default box（每一个feature map cell 不是k个default box都取）。也就是说default box是一种概念，prior box则是实际的选取。

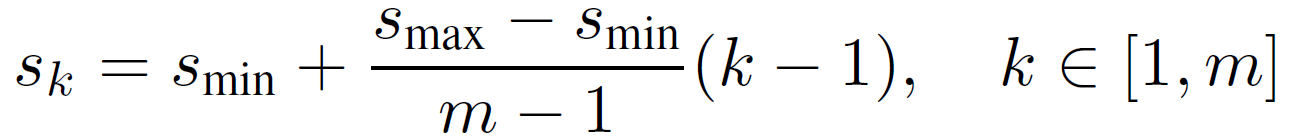
训练中一张完整的图片送进网络获得各个feature map，对于正样本训练来说，需要先将prior box与ground truth box做匹配，匹配成功说明这个prior box所包含的是个目标，但离完整目标的ground truth box还有段距离，训练的目的是保证default box的分类confidence的同时将prior box尽可能回归到ground truth box。作者的实验表明default box的shape数量越多，效果越好。

1. **网络结构**



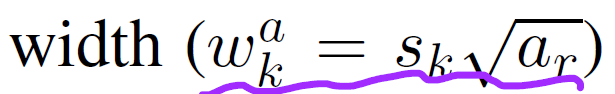
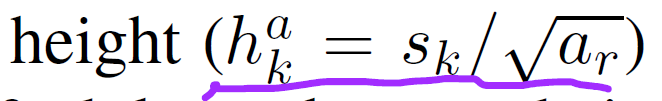
1. **Default Boxes的scale和aspect ratios**

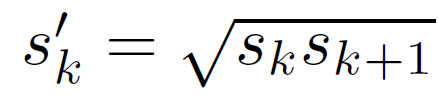
假设有m个特征图：



smin=0.2 smax=0.9 sk在0.2~0.9之间

ar∈{1，2，3，1/2，1/3}

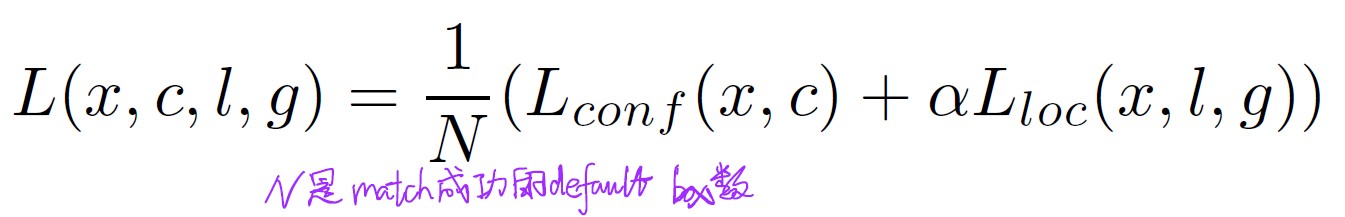
当aspect ratio=1时，增加一个default box，将scale设置成

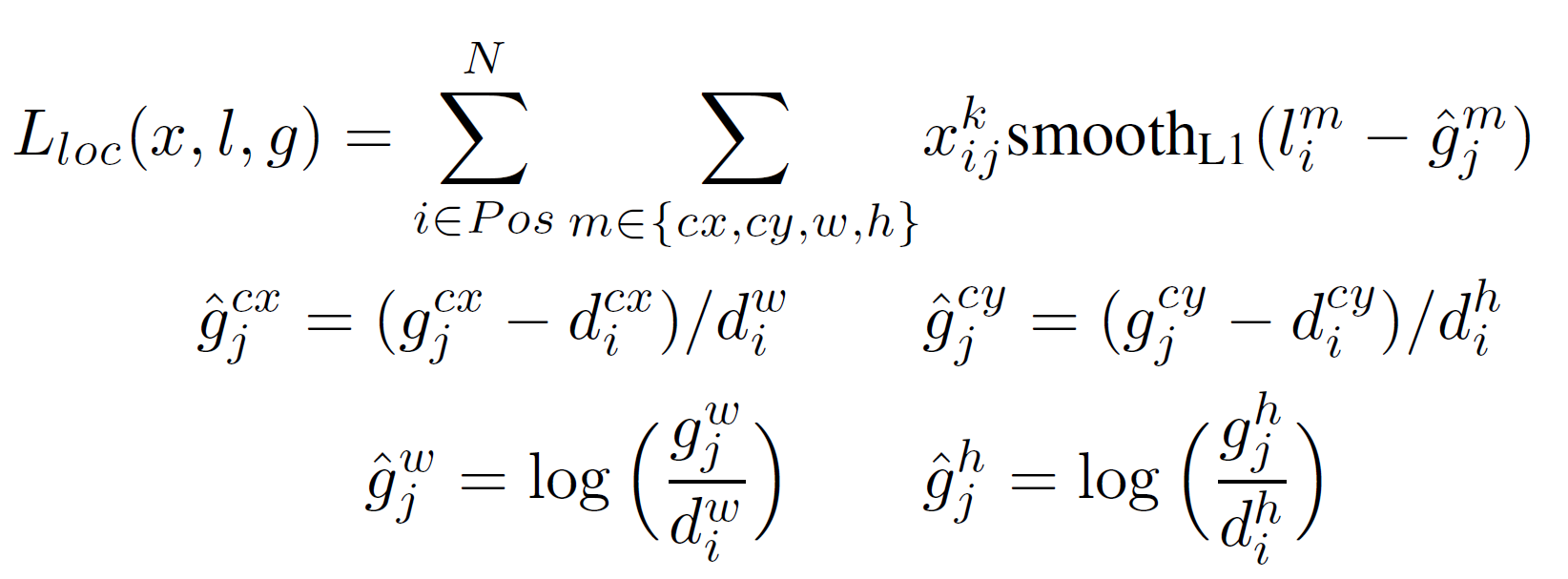
因此，对于每个feature map cell而言，一共有6种default box。可以看出这些default box在不同层的feature map上有不同的scale，在同一层的feature map上又有不同的aspect ratio，因此基本上可以覆盖输入图像中的各种形状和大小的object！

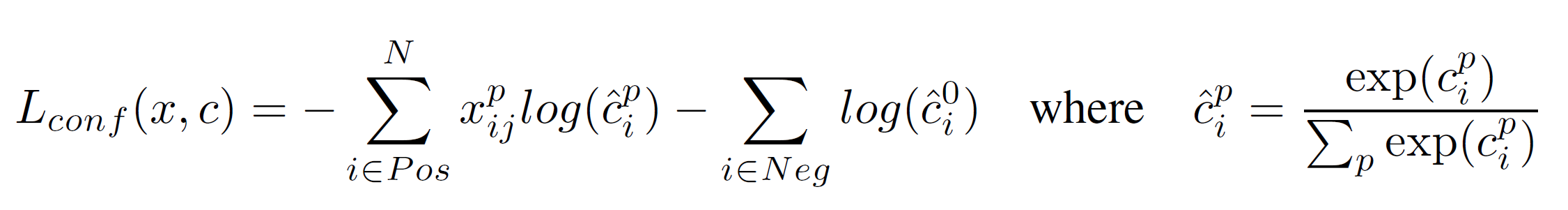
1. **难例挖掘**

将分类损失排序，从最高的开始选，使正负样本比例保持在1:3

1. **训练过程**







Lloc只有在正例上累加损失，Lconf在正例和负例上都累加损失

