RCNN
Rich feature hierarchies for accurate object detection and semantic segmentation
用CNN从底向上的产生候选域，用于定位和分割物体
当训练数据比较少时，监督的预训练作为一个辅助任务，之后采用特定领域的微调提升性能

1.介绍

用深度网络定位物体
只用少量的有标签的检测数据去训练一个高容量的模型

把定位当作回归问题来做：在实际中效果不好 ，需要提前知道图像中物体的个数
滑动窗口：CNN中一般采用这种方法定位，如果采用这种方法，精确定位有一个大的挑战，因为，网络的高层单元的5个卷积层有一个大的感受野和跨度195×195，32×32
文章中采用的方法：recognition using regions

训练数据太少，不足以训练一个大的CNN网络： 以前的解决方法：非监督的预训练，有监督的微调
文中采用的方法：使用一个大的辅助数据库（ILSVRC）有监督的预训练，在小数据库(PASCAL)中做特定领域中微调

2.RCNN做物体检测

目标检测系统有三个模块组成：
1. 产生类别无关的候选域
2. 大的CNN网络，从候选域中提取出固定长度的特征
3. 一系列的线性SVM，用于分类

模块设计

候选域：采用的方法是：selective search
特征提取：使用CNN从每个候选域中提取一个4096维的特征
特征的计算是通过向前传播一个 减掉均值的227×227的RGB图像，用于计算特征的CNN网络有5个卷积层2个全连接层
CNN网络只能接受227×227的RGB图像，所以需要先把候选域图像转化成这个尺寸

检测的测试过程

1. 在测试图片上运行selective search,得到2000个候选域 ，使用的是selective fast版本
2. 扭曲每个候选域，向前传送到CNN网络，从特定的层读出特征
3. 对每个类，用SVM给提取到的特征向量打分
4. 对于打过分的候选域，使用一个NMS，拒绝那些跟比他分高的候选域的IoU大于学到的阈值的候选域
   运行时间分析：
   所有类别之间共享CNN参数

通过CNN计算得到的特征向量维度比较低
计算候选域和特征向量的时间：13s/img 在GPU或者53s/img 在CPU

训练

有监督预训练：使用辅助数据库ILSVRC 2012，只有图像级的标签，没有bounding box 标签
特定领域的微调：使用VOC中经过扭曲的候选窗口继续用随机梯度下降训练CNN参数，把原来1000-way分类器，替换成21-way分类器
把所有跟真实样本的IoU大于等于0.5的作为正样本，类就是真实样本的类，其余的作为负样本
SVG的min batch是128，32个正样本窗口，96个背景窗口
类别分类器：决定是不是背景的阈值设置为0.3，跟真实物体的IoU低于0.3的标记为背景，正样本就是真实物体框
为每个类优化一个SVM分类器，训练数据太多，不能放在内存中，采用一个standard hard negative mining method，可以比较快的收敛 微调阶段和SVM训练阶段正负样本的定义不同

在PASCAL VOC 2010-12上的结果

用VOC 2012训练集微调CNN,用VOC 2012训练验证集优化SVM
有bounding box 回归和没有的各提交了一次
没有bounding box回归的在2010测试集上是 50.2%
有bounding box回归的在2010测试集上是 53.7% VOC 2011/2012测试集上是53.3%

3.可视化，消除，错误模型

可视化学到的特征

第一层滤波器抓取有向边和颜色变化
用一种简单的非参数化的方法展示网络学到的东西
挑选出一个单独的单元，在后面接一个物体检测器

消除研究

逐层性能没有微调：为了理解那个层对性能是重要的，在VOC2007数据集上研究CNN的最后三层
pool5的可视化文章中的图3详细的介绍了
fc6:连着pool5的一个全连接层，为了计算特征，要用一个4096×9216的权值矩阵乘以一个9216维的特征向量，然后加上一个偏置向量，最后得到的向量经过一个半波整流处理$x \rightarrow max(0,x)$
fc7:用fc6得到的特征向量乘以一个4096×4096的权值矩阵，加一个偏置，经过半波整流处理

逐层性能有微调：VOC2007训练验证集上微调，mAP增加了8，fc6,fc7的提升比pool5的大