

1.1 亮点

- (1) 经典的目标检测算法使用滑动窗法依次判断所有可能的区域。本文则预先提取一系列较可能是物体的**候选区域**，之后仅在这些候选区域上提取特征，进行判断。
- (2) 采用大样本下有监督预训练+小样本微调的方式解决小样本难以训练甚至过拟合等问题

1.2 原理

RCNN 的流程可以分为以下几步：

- (1) 一张图像生成 1K~2K 个候选区域
- (2) 对每个候选区域，使用深度网络提取特征
- (3) 特征送入每一类的 SVM 分类器，判别是否属于该类
- (4) 使用回归器精细修正候选框位置

而在候选区域生成的时候，论文采用了 **Selective Search** 方法，基本思路为：

- (1) 使用一种过分割手段，将图像分割成小区域
- (2) 查看现有小区域，合并可能性最高的两个区域。重复直到整张图像合并成一个区域位置，合并规则是优先合并颜色相近的，纹理相近的，合并后总面积小的，合并后总面积在其 **BBOX** 中所占比例打的。
- (3) 输出所有曾经存在过的区域，所谓候选区域

而在特征提取的时候，采取的网络基本借鉴 Hinton 2012 年在 Image Net 上的分类网络，而分类器是使用一个线性 SVM 二类分类器进行判别。输入为深度网络输出的 4096 维特征，输出是否属于此类。由于负样本很多，使用 **hard negative mining** 方法。正样本是本类的真值标定框。负样本是如果和本类所有标定框的重叠都小于 0.3，认定其为负样本。

1.3 结论

论文的前两个步骤（候选区域提取+特征提取）与待检测类别无关，可以在不同类之间共用。这两步在 GPU 上约需 13 秒。同时检测多类时，需要倍增的只有后两步骤（判别+精修），都是简单的线性运算，速度很快。这两步对于 100K 类别只需 10 秒。