

R-FCN: Object Detection via Region-based Fully Convolutional Networks

笔记本: Detection

创建时间: 2019/4/15 9:13

更新时间: 2019/10/29 19:31

作者: ming71

1. Motivation

- Insight

- (1) 平移可变性和平移不变性

分类网络的位置不敏感性 (translation invariance) : 对于分类任务而言, 希望目标在图片中不断的移动, 网络不敏感这种变化, 仍然可以准确地进行类别判定。实验表明, 深度全卷积网络能够具备这个特性, 如ResNet-101等。

检测网络的位置敏感性 (translation variance) : 对于检测任务而言, 希望目标在图像上进行移动, 而网络能够敏感这种变化, 从而仍然能够准确的检测到它。但是, 深度全卷积网路由于最后输出的特征图很小, 不具备这样的一个特征。

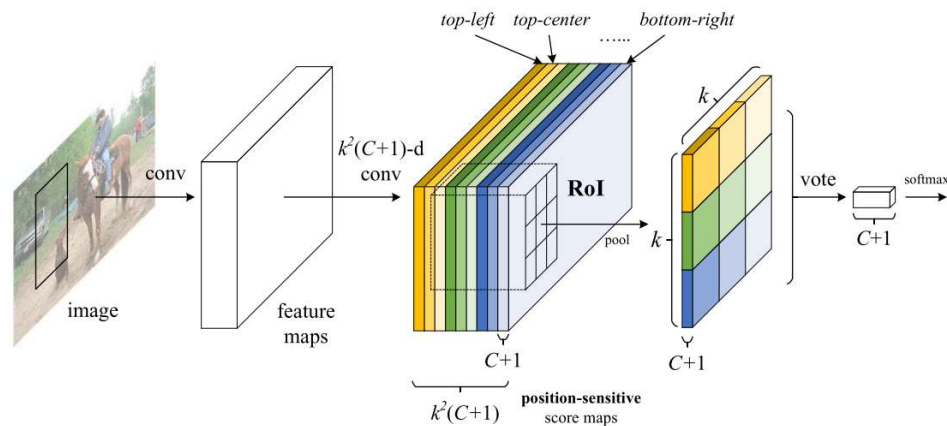
- (2) RoI带来的位置敏感性

在Faster RCNN中, 如果只采取卷积特征提取, 将导致定位回归精度极差; 卷积层之间加入RoIPooling后, 生成的proposals再经过后面的卷积层是敏感位置的, 因而给回归带来更高的精度; 显然如果RoIpooling加到FC分类回归之前门后面不加卷积层, 那么相当于只是提取分类特征然后分类回归, 会造成位置不敏感, 回归精度差

- Motivation

对于RoI得到的特征进行分别的分类回归, 这部分子网络的卷积以及后面的FC是不共享特征计算的, 导致计算量很大; 也不能单纯把卷积提前计算进行共享、RoI放到后面提取, 因为这样会导致上述的位置不敏感性(平移不变性)。为了更多地共享这部分计算, 进一步提速需要提出新的位置敏感方法。

2. Position-sensitive score map/Position-sensitive RoI Pooling



- 流程和解析

(1) 上图是分类的位置敏感得分图示意：

ResNet-101应用到R-FCN时会把最后的average pool和1000-d fc全连接层都去掉了，仅保留前100层，再新加一个 $1 \times 1 \times 1024$ 的卷积层用来降维，和一个很特殊的卷积来生成 $k^2 * (C+1)$ 维的Position-sensitive score map。比如论文中 $k=3$ 就是对ROI在长宽方向各三等分形成9个小区域。Position-sensitive score map的值对小区域相对于ROI中的位置很敏感，

最后一个特殊卷积输出Position-sensitive score map后，要做ROI Pooling了，和Faster R-CNN中的ROI Pooling一样要对9个小区域分别进行pooling，要注意的是R-FCN中9个小区域并不是在所有 $k^2 * (C+1)$ 维度上都做pooling，每个小区域只会在对应的 $(C+1)$ 个维度上作poolin，每个维度上的 $k \times k$ 个数据再加到一起(图2的vote过程)形成 $C+1$ 个单点数据，就代表了 $C+1$ 个类别的分类概率。

(2) 对于目标定位的输出和上面的分类输出过程类似，只是维度不再是 $k^2 * (C+1)$ ，而是 $k^2 * 4$ ，表示9个小区域的 $[dx, dy, dw, dh]$ 4个偏移坐标。上图的 k 的划分，相当于把proposal/ROI分成 $k \times k$ 个区域，每个区域负责检测物体对应的部分，只有九个区域都达到最大的响应了才认为检测到了物体，这一步通过对九个区域进行投票选出最佳可能的类别。例如如果gt是人，则中间上面的部分应该响应人头，依次类推。