

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

Reporter

冯浩哲

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

论文概要

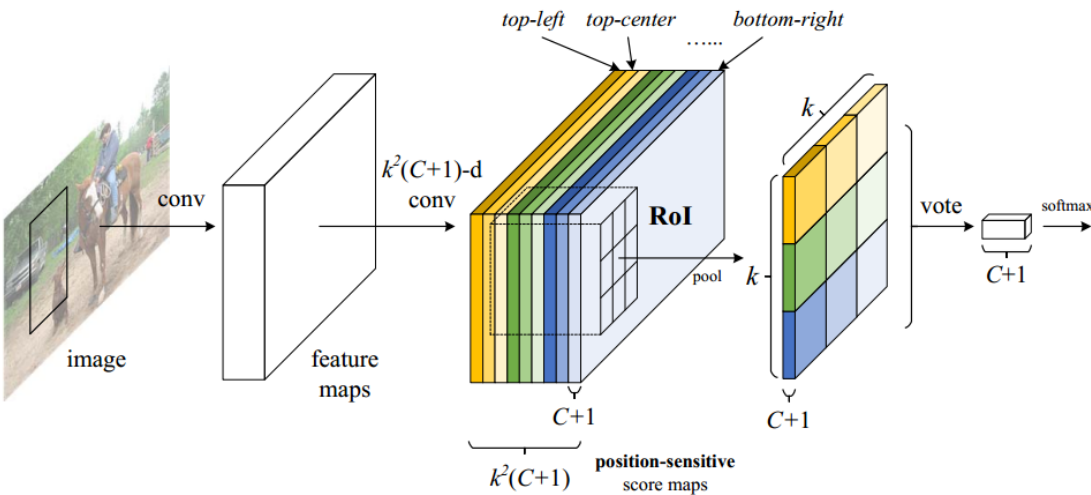
什么是全卷积，如何减少冗余计算

论文概要

论文主要是对Faster R-CNN的一个修改，即使用全部都是卷积，没有全连接的网络来进行Object Detection任务，而它针对的问题就是Faster R-CNN进行Feature projection与ROI Pooling后需要对很多Pooling后的重复特征进行全连接的问题，这里的计算量很大，而该文章就是要解决这个问题。

什么是全卷积，如何减少冗余计算

可以用一张图说清什么是全卷积



这是一个全卷积后分类的示意图。简单地来说，就是直接Feature Map到底部，然后map之后不加入参数。具体来说，针对classification的问题，对于Feature Extractor提出的Feature Map，我们卷积生成与Feature Map相同大小的score maps，但是channel规定为 $k^2 * (C + 1)$ 个。C代表分类的数目，那么C+1就多了一个背景，这个是好理解的。但是 k^2 如何理解呢？文中认为 $k^2 = k * k$ 代表的是不同方位下的特征，可以用一张图来解释：

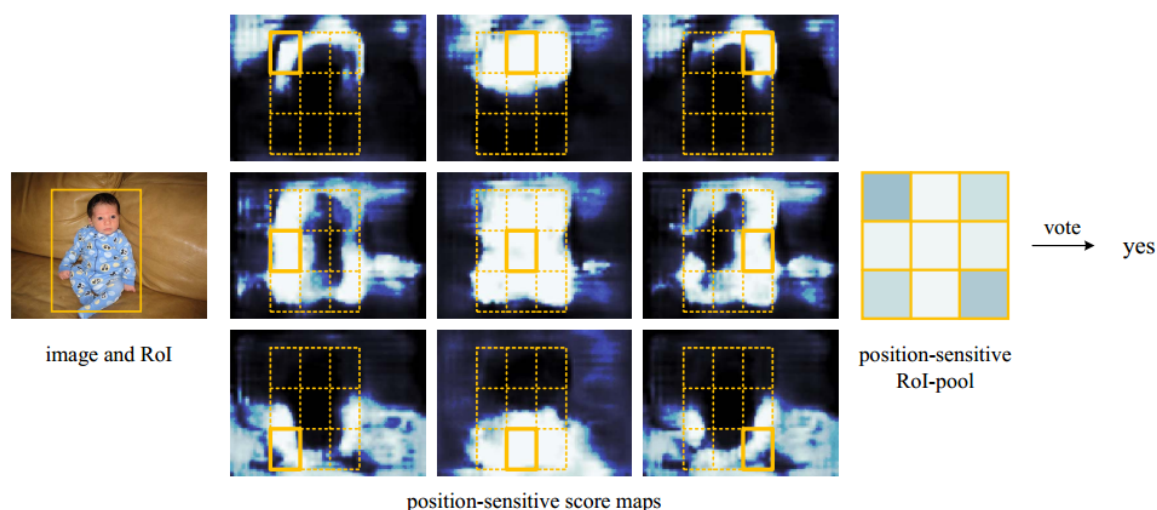


Figure 3: Visualization of R-FCN ($k \times k = 3 \times 3$) for the *person* category.

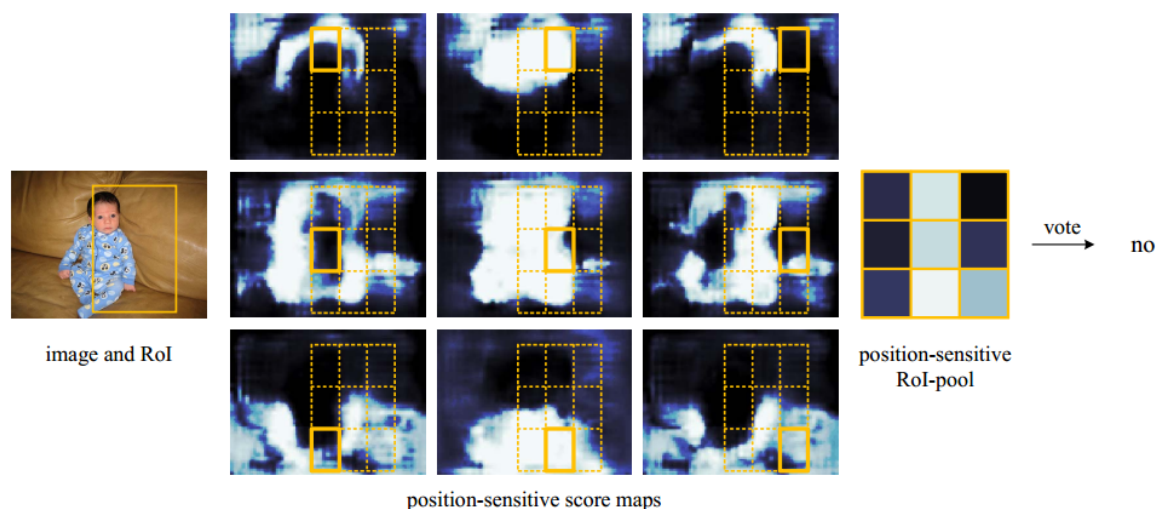


Figure 4: Visualization when an RoI does not correctly overlap the object.

对于每一个分类 C ，取 $k=3$ 意味着 k^2 表示9个区域，当然也可以取的更多。然后我们对 $k - th$ 特征图进行 $k * k$ 剖分为若干个**bins**，依次取对应的bins上的点的平均值得到一个 k^2 大小 $C + 1$ 个channels的带有方向性的Feature map组，然后对每个feature map求和，最后用softmax进行分类。这么做法的假设是特征图将自己想要表现的特征值很大。如果当前区域不是表现的特征，那么值就很小以至于无法激活。

对于bbox regression 也是如此，只是此时取的是 $k^2 * 4$ 的特征图罢了。

话说我一直没有弄清bbox regression的原理。我知道它要拟合的是什么，但是我仍然不知道它是依据什么来拟合的，所以作者提出这个一定是对bbox有着非常深入的了解了。