# [7]2017 Mask R-CNN

参考<https://blog.csdn.net/wangdongwei0/article/details/83110305>

## Abstract

The method, called Mask R-CNN, extends Faster R-CNN by adding a branch for predicting an object mask in parallel with the existing branch for bounding box recognition. Mask R-CNN is simple to train and adds only a small overhead to Faster R-CNN, running at 5 fps. Moreover,Mask R-CNN is easy to generalize to other tasks.

Mask RCNN是Faster RCNN的扩展，对于Faster RCNN的每个Proposal Box都要使用FCN进行语义分割，分割任务与定位、分类任务是同时进行的。

引入了RoI Align代替Faster RCNN中的RoI Pooling。

引入语义分割分支，实现了mask和class预测的关系的解耦，mask分支只做语义分割，类型预测的任务交给另一个分支。

## Mask R-CNN

与Faster RCNN采用了相同的two-state步骤：首先是找出RPN，然后对RPN找到的每个RoI进行分类、定位、并找到binary mask。这与当时其他先找到mask然后在进行分类的网络是不同的。

### Lost function

IMG_256

这里主要介绍一下Lmask L\_{mask}L

mask

​

，Lmask L\_{mask}L

mask

​

是对每个像素进行分类，其含有K∗m∗m维度的输出，K代表类别的数量，m\*m是提取的ROI图像的大小。Lmask 被定义为 average binary cross-entropy loss（平均二值交叉熵损失函数）。这里解释一下是如何计算的，首先分割层会输出channel为K的Mask，每个Mask对应一个类别，利用sigmoid函数进行二分类，判断是否是这个类别，然后在计算loss的时候，假如ROI对应的ground-truth的类别是Ki，则计算第Ki个mask对应的loss，其他的mask对这个loss没有贡献计算二值交叉熵搞的公式如下图中的函数接口。

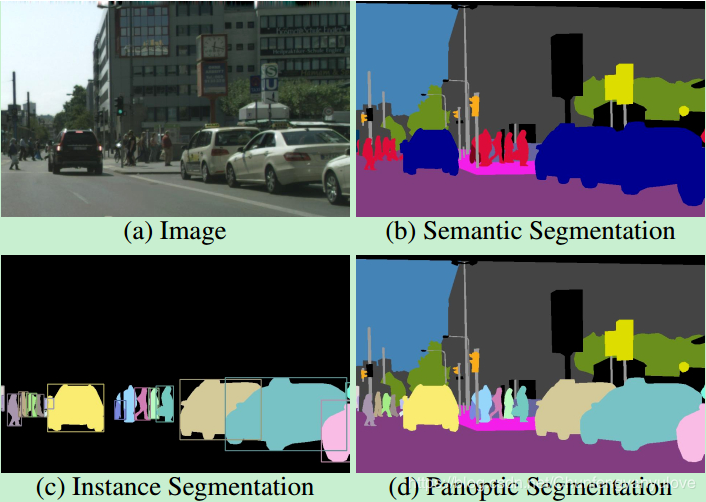
### RoIAlign

RoIPool的目的是为了从RPN网络确定的ROI中导出较小的特征图(a small feature map，eg 7x7)，ROI的大小各不相同，但是RoIPool后都变成了7x7大小。RPN网络会提出若干RoI的坐标以[x,y,w,h]表示，然后输入RoI Pooling，输出7x7大小的特征图供分类和定位使用。问题就出在RoI Pooling的输出大小是7x7上，如果RON网络输出的RoI大小是8\*8的，那么无法保证输入像素和输出像素是一一对应，首先他们包含的信息量不同（有的是1对1，有的是1对2），其次他们的坐标无法和输入对应起来（1对2的那个RoI输出像素该对应哪个输入像素的坐标？）。这对分类没什么影响，但是对分割却影响很大。RoIAlign的输出坐标使用插值算法得到，不再量化；每个grid中的值也不再使用max，同样使用差值算法。

## IMG_256

## 为什么Mask R-CNN能取得很好的效果？

原文链接：<https://blog.csdn.net/chunfengyanyulove/article/details/83545784>



语义分割（semantic segmentation）：对图像中逐像素进行分类。

实例分割（instance segmentation）：对图像中的object进行检测，并对检测到的object进行分割。

全景分割（panoptic segmentation）：对图像中的所有物体进行描述。

其中Mask RCNN做的是其中的instance segmentation.

实例分割（instance segmentation）的难点在于：**需要同时检测出目标的位置并且对目标进行分割**

**所以这就需要融合目标检测（框出目标的位置）以及语义分割（对像素进行分类，分割出目标）方法。在Mask R-CNN之前，Faster R-CNN在目标检测领域表现较好，同时FCN在语义分割领域表现较好。很自然的方法是将Faster R-CNN与FCN相结合。**

Mask R-CNN是建立在Faster R-CNN基础上的，那么我们首先回顾一下Faster R-CNN，Faster R-CNN是典型的two stage的目标检测方法，首先生成 RPN候选区域， 然后候选区域经过Roi Pooling进行目标检测（包括目标分类以及坐标回归），分类与回归共享前面的网络。

而Mask R-CNN在Faster R-CNN的基础上，**增加了第三个支路，输出每个ROI的Mask（这里是区别于传统方法的最大的不同，传统方法一般是先利用算法生成mask然后再进行分类，这里平行进行）**