layout: post
title: "RCNN系列文章之Mask RCNN详解"
date: 2020-07-14
description: "目标检测"

tag: 目标检测

RCNN系列的文章主要是RCNN，Fast RCNN， Faster RCNN， Mask RCNN, Cascade RCNN,这一系列的文章是目标检测two-stage算法的代表，这系列的算法精度高，效果好，是一类重要的方法。

论文地址：Mask R-CNN

简要介绍

Mask RCNN并不是一个目标检测的算法，而是一个语义分割的算法，但是作为一个RCNN系列的又一个神一般的扩展，必须要读一下，这可是ICCV2017的best paper，出自何凯明大神。

- Mask RCNN是一个通用的实例分割的框架
- 在Faster rcnn的基础上添加一个mask分支，multi-task来实现实例分割
- 这种multi-task的结构不仅能够用于实例分割，在人体姿态 检测上文中也给出了不错的结果
- 利用ROI_Align替代ROI_pooling，提升实例分割的准确率

网络结构图：

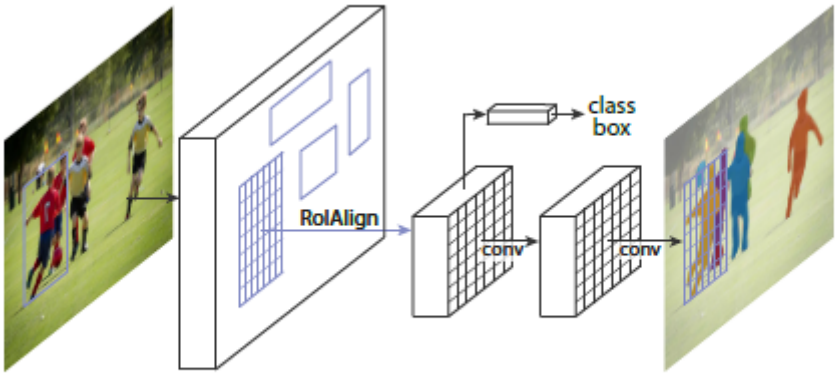

Figure 1. The Mask R-CNN framework for instance segmentation.

mask rcnn在faster rcnn预测框并行的位置添加一个预测掩码的分支，在每一个ROI上利用掩码分支预测一个二分类的mask。用来预测mask的掩码分支是一个在像素级别上对于每个ROI预测语义掩码的小全卷积网络。

由于Faster rcnn主要不是为了像素级别的实力分割所设计的算法，尤其是在ROI pooling中，对于特征仅仅获得了粗糙的表述，为了修正像素间可能出现的misalign现象，作者采用ROI_align来取代ROI_pooling从而能得到很大的提升（对于mask rcnn的ROIalign的细节介绍以及其对于ROIpooling的改进在后边介绍）。

另外，作者解耦了类别与掩码(mask)的预测，对于每一个类都独立的预测一个mask，依赖于网络的ROI分类分支，来进行类别的预测，这种方式相比较FCN（全卷积网络）来说，性能有很大的提升，FCN通常为每个像素预测多个类别，这种方式把语义信息与类别信息耦合在一起，这种方式进行实例分割效果较差。

Mask RCNN

mask rcnn 采用和faster rcnn相似的两步法结构， 第一阶段RPN网络，提取出候选的目标边界框，第二阶段mask rcnn对于来自RPN的候选区域，利用ROI align提取特征并进行类别分类、边界框回归与二进制掩码生成。

mask rcnn采用multi-task的损失函数的和作为最终的损失函数：

$$L = L_{cls} + L_{box} + L_{mask}$$

利用三个任务的损失函数之和作为最终的损失函数。

ROIAlign与ROIpooling的差别

ROIpooling实现的算法过程操作：

1. 假如原图为800x800，经过32倍下采样之后得到的最终feature map的大小为25*25，
2. 假设其中有一个665*665的region proposal，那么映射到特征图的大小变换为665/32=20.78,然后经过**第一次量化**，此region proposal在特征图上的尺寸大小为20x20，
3. 假定最终要变换为7x7的特征图，因此在feature map上为20x20的特征图需要划分为49个小区域，每个小区域的大小均为20/7=2.86， 然后经过**第二次量化**，此时每个小区域的大小变为2x2
4. 最终每个2x2的小区域中选择最大的像素值，构成7x7的固定大小的特征图

ROIAlign实现的算法过程：

同样与前边ROIpooling描述的算法过程类似，

1. 假如原图为800x800，经过32倍下采样之后得到的最终feature map的大小为25*25
2. 假设其中有一个665*665的region proposal，那么映射到特征图的大小变换为665/32=20.78，此时与ROIpooling不同之处在于此时不进行量化操作，保留浮点数
3. 假定最终要变换为7x7的特征图，因此在feature map上为20.78x20.78的特征图需要划分为49个小区域，那么每个小区域的大小为20.78/7=2.97，那么最终的小区域为2.97x2.97
4. 假定采样点数为4，即表示，对于每个2.97*2.97的小区域，平分四份，每一份取其中心点位置，而中心点位置的像素，采用双线性插值法进行计算，这样，就会得到四个点的像素值

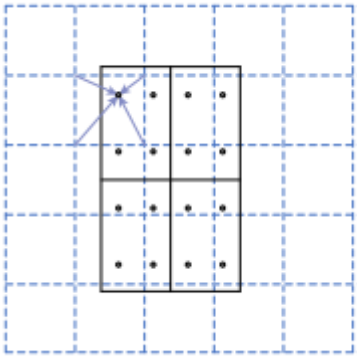

Figure 3. **RoIAlign:** The dashed grid represents a feature map, the solid lines an RoI (with 2×2 bins in this example), and the dots the 4 sampling points in each bin. RoIAlign computes the value of each sampling point by bilinear interpolation from the nearby grid points on the feature map. No quantization is performed on any coordinates involved in the RoI, its bins, or the sampling points.

ROIAlign操作相比于ROIpooling操作而言，免去了两次量化操作，从而减少了misalignment在实例分割中对于算法精度的影响，虽然对于目标检测的影响并不十分大，在实例分割中会有很大的提升。

总结

Mask RCNN创新点（精华之处）：

1. **multi-task结构**的提出，在不提升较大复杂度的情况下，提升实例分割的性能
2. **ROIAlign**解决ROIpooling算法的misalignment的问题，保证原图与feature map，feature map到ROI的像素对齐，提升目标检测精度，更符合实例分割的问题，极大程度的提升精度
3. **decouple mask and classification**，实现掩码预测与类别预测的解耦，对于每个ROI预测k个二分类的掩码，不想FCN那样对每个像素预测k个类别，在一定程度上提升算法的复杂度。