

ROI Pooling 层解析

ROI Pooling 的意义

ROIs Pooling [顾名思义](#)，是 Pooling 层的一种，而且是针对 RoIs 的 Pooling，他的特点是**输入特征图尺寸不固定，但是输出特征图尺寸固定**；

1. 什么是 ROI 呢？

ROI 是 Region of Interest 的简写，指的是在“特征图上的框”；

2. 在 Fast RCNN 中，RoI 是指 Selective Search 完成后得到的“候选框”在特征图上的映射，如下图所示；

3. 在 Faster RCNN 中，候选框是经过 RPN 产生的，然后再把各个“候选框”映射到特征图上，得到 RoIs。

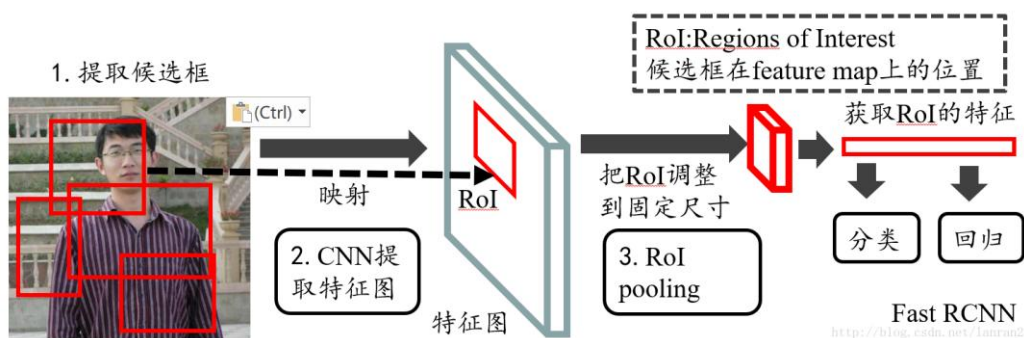


图 1 Fast RCNN 整体结构

往往经过 rpn 后输出的不止一个矩形框，所以这里我们是对多个 ROI 进行 Pooling。

ROI Pooling 的输入

输入有两部分组成：

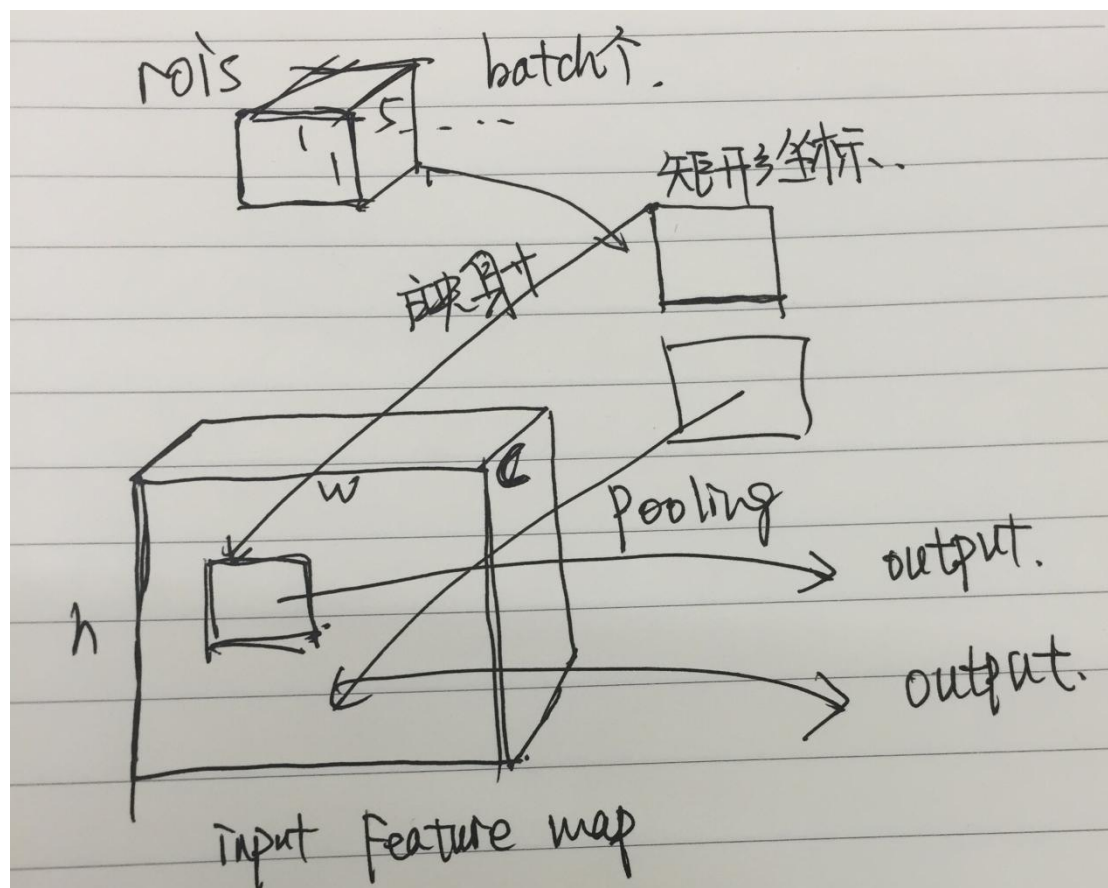
1. 特征图：指的是图 1 中所示的特征图，在 Fast RCNN 中，它位于 RoI Pooling 之前，在 Faster RCNN 中，它是与 RPN 共享那个特征图，通常我们常常称之为“share_conv”；

2. rois：在 Fast RCNN 中，指的是 Selective Search 的输出；在 Faster RCNN 中指的是 RPN 的输出，一堆矩形候选框框，形状为 $1 \times 5 \times 1 \times 1$ （4 个坐标+索引 index），其中值得注意的是：坐标的参考系不是针对 feature map 这张图的，而是针对原图的（神经网络最开始的输入）

ROI Pooling 的输出

输出是 batch 个 vector，其中 batch 的值等于 RoI 的个数，vector 的大小为 $\text{channel} * w * h$ ；RoI Pooling 的过程就是将一个个大小不同的 box 矩形框，都映射成大小固定（ $w * h$ ）的矩形框；

ROI Pooling 的过程



如图所示，我们先把 roi 中的坐标映射到 feature map 上，映射规则比较简单，就是把各个坐标除以“输入图片与 feature map 的大小的比值”，得到了 feature map 上的 box 坐标后，我们使用 Pooling 得到输出；由于输入的图片大小不一，所以这里我们使用的类似 Spp Pooling，在 Pooling 的过程中需要计算 Pooling 后的结果对应到 feature map 上所占的范围，然后在那个范围中进行取 max 或者取 average。

原文：<https://blog.csdn.net/lanran2/article/details/60143861>