**Mask R-CNN**

**1 引言**

Faster RCNN通过将selective search区域提议算法改进为RPN、进一步实现了整体结构的端到端训练以及加入“注意力”机制，提高了目标检测的效率以及性能。但它仍然无法用于实例检测（instance segmentation）。实例检测需要再精确分割所有实例的同时进行目标检测，所以需要目标检测以及语义分割的共同实现。

所以，本文在Faster RCNN的基础上进行了一系列改进。

首先，在Faster RCNN的基础上增加了一个用于对每个ROI进行分割掩码预测的分支，这在结构上和原来的分类分支以及边界框回归分支并行执行。

此外，将Faster RCNN中的ROI Pooling部分替换为了ROIAlign，实现了无特征错位的尺寸变换，将掩码准确度提高了10%到50%。

经过以上改进以及一些细节的优化，Mask RCNN在没有复杂技巧的情况下，超过了所有先前在COCO实例分割任务上的最先进单模型结果。同时在COCO目标检测任务上的表现也非常优益。此外，作者认为Mask RCNN优秀的训练和测试速度以及框架的灵活性和准确性，将有助于并简化未来实例分割的研究。

**2 Mask RCNN框架**

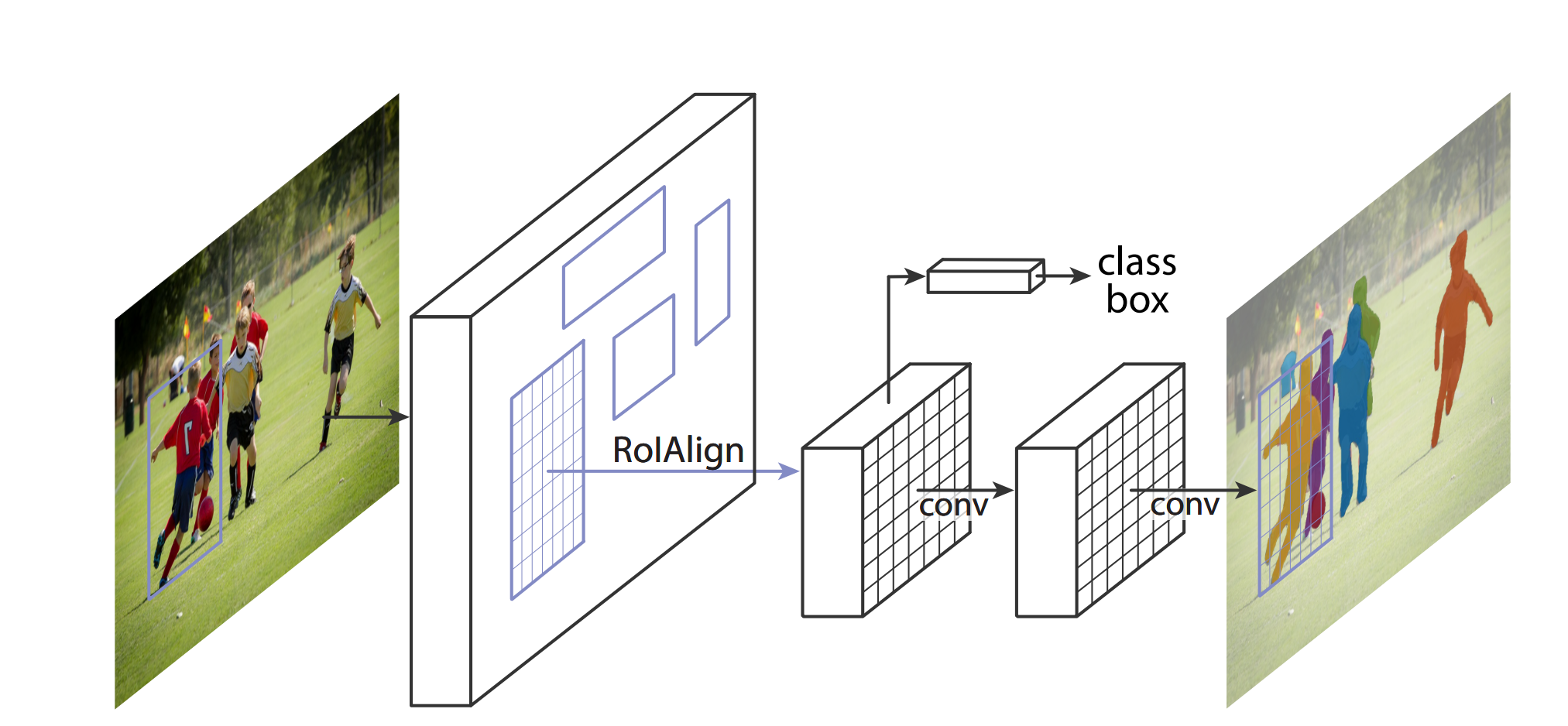


图2-1 Mask RCNN架构

**2.1 Mask RCNN算法步骤**

首先，输入一幅你想处理的图片，然后进行对应的预处理操作，或者预处理后的图片；

然后，将其输入到一个预训练好的神经网络中（ResNeXt等）获得对应的feature map；

接着，对这个feature map中的每一点设定预定个的ROI，从而获得多个候选ROI；

接着，将这些候选的ROI送入RPN网络进行二值分类（前景或背景）和BB回归，过滤掉一部分候选的ROI；

接着，对这些剩下的ROI进行ROIAlign操作（即先将原图和feature map的pixel对应起来，然后将feature map和固定的feature对应起来）；

最后，对这些ROI进行分类（N类别分类）、BB回归和MASK生成（在每一个ROI里面进行FCN操作）。

可以看到，前四步和Faster RCNN几乎相同，主要的改动也就是出现在后面两部，在接下来的内容中作详细介绍。

**2.2ROIAlign**

ROI Pooling存在一系列问题。ROI Pooling首先将浮点数表示的ROI量化到特征图的离散粒度，然后将该量化后的ROI细分成空间区域(bins)，并对这些区域进行进一步的量化处理，最后通过聚合（在Faster RCNN中使用最大池化）每个区域内的特征值来提取ROI的特征。在这个过程中，会涉及到由于原始图像和卷积网络需要的输入之间尺寸转换不能整除的问题，ROI Pooling的方法是直接进行四舍五入，这就会导致ROI和特征之间对不起，也就是产生misalignment，导致信息丢失。

这对于分类任务并没有太大的影响，因为分类任务并不特别关注像素级别的损失，具有鲁棒性。但这对于掩码预测这类像素级别的任务却又很大的影响。为了解决这一问题，作者提出了ROIAlign层，它消除了ROIPooling中严格的量化过程，将提取的特征与输入进行对齐。如何实现呢？作者使用了双线性差值对每个ROI区域进行采样，并对结果进行聚合（使用最大值池化或者均值池化），如图2-2所示。这样就解决了由于边界问题产生的空间量化和信息丢失，更准确地提取了ROI内的特征。

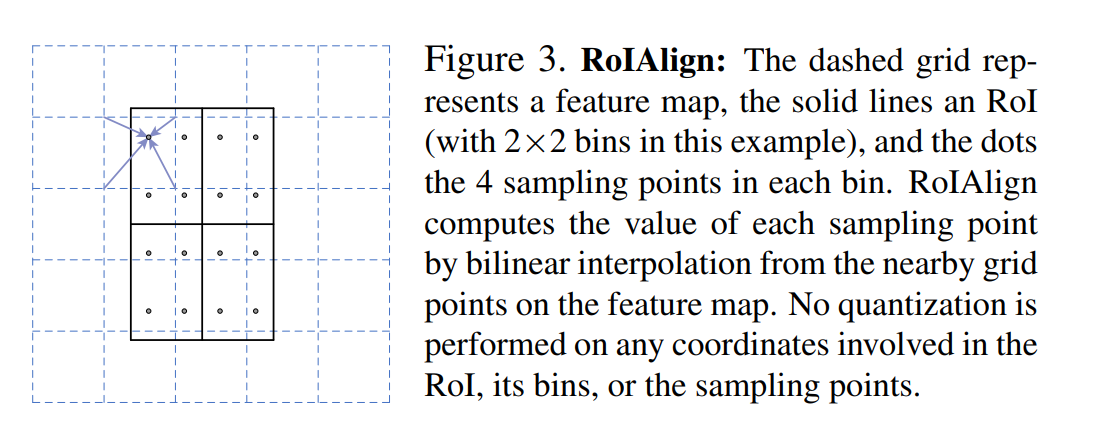


图2-2 ROIAlign示例

**2.3 Head Architecture**

为了实现掩码预测，文中提出了如图2-3所示的两种架构，左边的Faster RCNN/ResNet和右边的Faster RCNN/FPN。在原来的卷积层之后会添加一个逐像素预测层，也称为全卷积层。这一层使用逐像素的方式，对ROI内的每个像素进行分类，判断该像素是够属于目标类别。当然这需要对每个像素进行sigmoid激活并通过损失函数进行训练。在逐像素预测之后，得到的输出是一个概率图，用于表示每个像素属于目标类别的概率。并设置阈值用于得到目标的最终二值掩码。

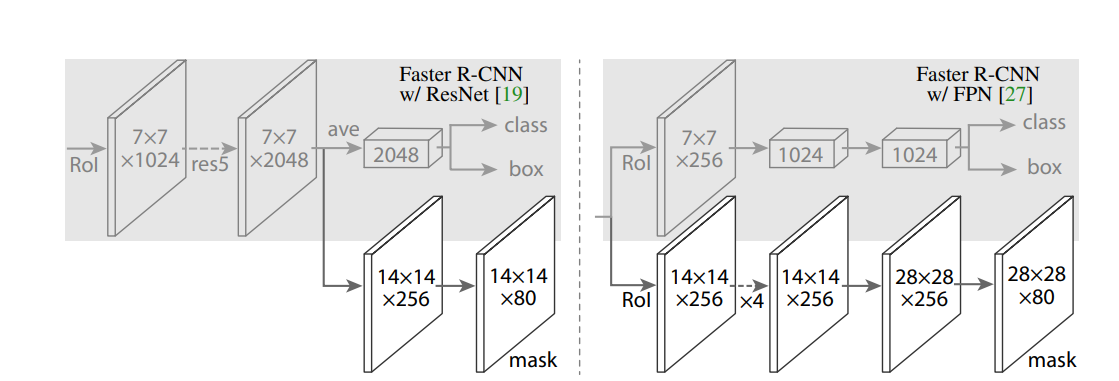


图2-3两种掩码预测分支实现

总的来说，Mask R-CNN的掩码预测分支通过RoIAlign将每个RoI对齐为固定大小的特征图，并在此基础上利用全卷积网络生成目标的精确掩码，实现了目标实例的像素级分割。这使得Mask R-CNN成为一个强大的目标检测和实例分割模型。

**2.4 损失函数**

Mask RCNN在Faster RCNN的损失函数基础上，增加了项，用于计算新的掩码预测分支的损失。新的损失函数为其中和和Faster RCNN中相同，用于计算分类分支和边界框回归分支的损失。每个类别中有一个m\*m的二进制掩码，k表示类别索引。为此，我们为每个像素应用sigmoid，并定义项为平均二进制交叉熵损失。特别要注意的是，对于真实类别为k的ROI，仅计算隶属于类别k的掩码的损失，其他掩码不计入当前损失计算。这与语义分割中的对每个像素应用softmax激活以及多项交叉熵损失函数不同，这种计算方式会导致掩码检产生跨类别竞争。本文对于损失函数的设计师改善实例分割结果的关键（通过实验得出这一结论）。

**3 实验**

实验部分不再赘述，几乎相同。需要提到的一点是Mask RCNN使用了多种网络架构，且使用了8块GPU共同训练。Mask RCNN在COCO的目标检测与实例分割任务中均取得了很好的成绩，且泛化到人体关节检测任务中，同样取得了非常好的成绩。