基于改进Mask R-CNN的模糊图像实例分割的研究

陈卫东[①②] 郭蔚然[①] 刘宏炜[①] 朱奇光*[①②]

[①](燕山大学信息科学与工程学院 秦皇岛 066004)

[②](河北省特种光纤与光纤传感重点实验室 秦皇岛 066004)

摘 要：Mask R-CNN是现阶段实例分割相对成熟的方法，针对Mask R-CNN算法当中还存在的分割边界精度以及对于模糊图片鲁棒性较差等问题，该文提出一种基于改进的Mask R-CNN实例分割方法。该方法首先提出在Mask分支上使用卷积化条件随机场(ConvCRF)来优化Mask分支对于候选区域进一步分割，并使用FCN-ConvCRF分支来代替原有分支；之后提出新锚点大小和IOU标准，使得RPN候选框能够涵盖所有实例区域；最后使用一种添加部分经过转换网络转换的数据进行训练的方法。总的mAP值与原算法相比提升了3%，并且分割边界精确度和鲁棒性都有一定提高。

关键词：图像实例分割；Mask R-CNN；条件随机场；RPN层

中图分类号：TN911.73 **文献标识码**：A **文章编号**：1009-5896(2020)11-2805-08

DOI: 10.11999/JEIT190604

Research on Fuzzy Image Instance Segmentation Based on Improved Mask R-CNN

CHEN Weidong[①②] GUO Weiran[①] LIU Hongwei[①] ZHU Qiguang[①②]

[①](*School of Information Science and Engineering, Yanshan University, Qinhuangdao* 066004, *China*)

[②](*Key Laboratory for Special Fiber and Fiber Sensor of Hebei Province, Yanshan University, Qinhuangdao* 066004, *China*)

Abstract: Mask R-CNN is a relatively mature method for image instance segmentation at this stage. For the problems of segmentation boundary accuracy and poor robustness of fuzzy pictures in Mask R-CNN algorithm, an improved Mask R-CNN method for image instance segmentation is proposed. This method first proposes that on the Mask branch, Convolution Condition Random Field(ConvCRF) is used to optimize the Mask branch, and the candidate area is further segmented, and uses FCN-ConvCRF branch to replace the original branch. Then, a new anchor size and IOU standard are proposed to enable the RPN candidate box cover all the instance areas. Finally, a training method is used to add a part of data transformed by the transformation network. Compared with the original algorithm, the total mAP value is improved by 3%, and the accuracy and robustness of segmentation boundary are improved to some extent.

Key words: Image instance segmentation; Mask R-CNN; Conditional Random Field(CRF); RPN level

1 引言

近年来，随着全卷积网络(FCN)[1]的发展，使得计算机对图像的特征提取和高级推理的精度、合理性都有了新提高。但FCN仅仅是预测像素到像素的类别标签，对于类别当中的实例却无法做出分割，既要区分类别又要区分实例成为了图像分割中新的挑战问题。

为了降低实例分割的复杂性，很多方法都是以预测物体框为重点，之后细化分类结果得到像素级别并且可以区分实例的预测，这类方法通常以最先进的物体检测模型如Faster R-CNN[2], YOLO[3], YOLOv2[4]为基础来获得感兴趣区域，然后对感兴趣区域内像素进行进一步分类以获得最终掩码。例如Dai等人[5]提出的MNC，Dai等人[6]的ISFCN以及FCIS[7]。另一种方法是利用深度学习来解决传统图像分割当中的问题来实现实例分割。如Bai等人[8]

收稿日期：2019-08-08；改回日期：2020-08-26；网络出版：2020-09-03

*通信作者： 朱奇光 zhu7880@ysu.edu.cn

基金项目：国家自然科学基金(61773333)，河北省教育厅高等学校科技计划重点项目(ZD2018234)

Foundation Items: The National Natural Science Foundation of China (61773333), The Key Project of Science and Technology Plan of Colleges and Universities of Hebei Provincial Department of Education (ZD2018234)

的深度分水岭算法，Liu等人[9]的SGN网络等等都有不错的效果。He等人[10]的Mask R-CNN是目前实例分割最好的方法之一，由DeepMask[11], SharpMask[12]和MultiPathNet[13]发展而来。深层卷积神经网络(DCNN)在处理图像分割、语义分割等图像问题当中取得了较为显著的成果，但卷积神经网络在处理图像问题时大都依靠局部特征和较小的感受野，容易产生粗糙边缘。罗会兰等人[14]使用多尺度相互重叠区域来优化分割边界。其他学者也提出了依靠深度神经网络和条件随机场结合的方法，比如DeepLabv2[15], CRF as RNN[16]算法，韩铮等人[17]使用弱监督数据集和条件随机场完成图像分割任务，所以神经网络与条件随机场结合的结构化模型也是现阶段图像分割的方法之一。在实际生活和操作中，图像可能由于大雨大雾天气，或者是本身像素问题导致图像变模糊，基于检测的实例分割算法在模糊图像中很容易因为分割的感兴趣区域出现检测偏差而产生像素信息丢失，或者出现分割边缘过度平滑问题，需要神经网络本身能够更加有效精准地处理模糊图像。

本文采用实例分割相对成熟的Mask R-CNN为基础，使用卷积化条件随机场(ConvCRF)等方法作为优化，重点改进网络在模糊图像当中容易出现的边界平滑问题、检测偏差问题，并重新利用带有模糊图像的数据作为训练集进行训练来改进Mask R-CNN算法，使Mask R-CNN的精确度、鲁棒性都有提高。

2 FCN-ConvCRF分支

2.1 Mask分支

Mask R-CNN当中的Mask分支是作用于RoIAlign上的小FCN，由4个3×3卷积层，1个2×2反卷积层和1个全连接层组成，其中进行4次批归一化，该网络的特征提取部分和RPN层共享特征图。Mask分支和Mask R-CNN的其他两个分支并行处理，为每一个检测类别独立预测出掩码从而消除了跨类别竞争，结构简单而且行之有效。

尽管Mask R-CNN和其他算法相比都体现出一定优越性，但在具体实验当中Mask R-CNN依然有改进的余地，在实例边界处出现了不准确预测。为了优化这一点，本文提出了结合卷积化条件随机场的新Mask分支来代替原有分支进而细化Mask R-CNN输出结果。

2.2 ConvCRF

传统条件随机场(Condition Random Field, CRF)[18]学习速度非常缓慢，本文使用卷积化条件随机场策略[19]而不采用全连接条件随机场。卷积化条件随机场把条件独立性假设添加到传统全连接条件随机场当中，进而把大部分推断表示为卷积化操作，从而可以在GPU当中高效实现。同时，卷积化条件随机场的能量函数依然依靠一元势函数和二元势函数共同确定，其中二元势函数可以使具有相近特征的像素更容易分配到一致结果，进而弥补卷积神经网络因每个像素独立输出结果而导致的不一致性，卷积化条件随机场优化每一张图片均可以在10 ms之内完成。

全连接条件随机场[20]一直是条件随机场优化当中最常见的模型。现有一张由n个像素组成的图片$\boldsymbol{I}$，对于图像分割任务，每一个像素对应着k个类别，对于图片里每一个像素i, $\boldsymbol{I}_j$是像素的颜色向量，$\boldsymbol{X}_j$是对于每一个像素的标签。现有一个条件随机场$(\boldsymbol{I},\boldsymbol{X})$，服从高斯分布，其对应的分布率和关于$\boldsymbol{X}$的能量函数如式(1)所示

$$P(\boldsymbol{X}|\boldsymbol{I}) = \frac{1}{Z(\boldsymbol{I})}\exp\left(-\sum_{c\in C_G}\phi_c(\boldsymbol{X}_c|\boldsymbol{I})\right) \tag{1}$$

$$E(x|\boldsymbol{I}) = \sum_{c\in C_G}\phi_c(x_c|\boldsymbol{I}) \tag{2}$$

式中$G=(\nu,\varepsilon)$是定义在$\boldsymbol{X}$上的一个图，c表示子图，$\boldsymbol{X}_c$和x_c为不同维度下子图对应变量，该图当中每一个子图c都伴随着一个势函数ϕ_c，在全连接条件随机场模型当中，$\boldsymbol{G}$是$\boldsymbol{X}$的完全图而C_G则是所有一元和二元子图的集合，相关吉布斯能量函数可简化成一元和二元部分相加的形式。

$$E(x) = \sum_i \varphi_u(x_i) + \sum_{i<j}\varphi_p(x_i,x_j) \tag{3}$$

对于该能量函数当中的一元势函数是基于该像素点的特征(位置颜色向量等)独立计算的，在本文算法中该部分由神经网络输出。对于二元势函数有式(4)的形式

$$\psi_p(x_i,x_j) = \mu(x_i,x_j)\sum_{m=1}^{K}\boldsymbol{w}^{(m)}\boldsymbol{k}^{(m)}(\boldsymbol{f}_i,\boldsymbol{f}_j) \tag{4}$$

$$k^{(m)}(\boldsymbol{f}_i,\boldsymbol{f}_j) = \exp\left(-\frac{1}{2}(\boldsymbol{f}_i-\boldsymbol{f}_j)^{\mathrm{T}}\boldsymbol{\Delta}^{(m)}(\boldsymbol{f}_i-\boldsymbol{f}_j)\right) \tag{5}$$

式中$\boldsymbol{k}$为高斯卷积核，$\boldsymbol{f}_i$和$\boldsymbol{f}_j$为对应像素在某特征空间中的特征向量，$\boldsymbol{w}$为线性伴随相量，μ为标签兼容性函数，$\boldsymbol{\Delta}$为相关图形形状的正定矩阵。

卷积化条件随机场就是用条件独立性假设来补充全连接条件随机场，当两个像素的曼哈顿距离大于k时，就认为它们之间是条件独立关系，而且不使用Permutohedral lattice近似，把信息传递步骤重置为带截断的高斯卷积以便在GPU上高效学习和计算，该过程与CNN卷积非常相似。考虑一个

大小为$[bs,c,h,w]$的输入本文为它定义一个核矩阵表达式式(6)

$$\boldsymbol{k}_g[b,dx,dy,x,y] = \exp\left(-\sum_{i=1}^{d}\frac{|f_i^{(d)}[b,x,y]-f_j^{(d)}[b,x-dx,y-dy]|^2}{2\theta_i^2}\right) \tag{6}$$

式中θ_i为可学习参数，对于一组核矩阵$\boldsymbol{g}_1$到$\boldsymbol{g}_s$，定义合并的矩阵$\boldsymbol{k}=\sum_{i=1}^{s}w_i\cdot\boldsymbol{g}_i$定义所有$s$个核的组合信息传递结果$\boldsymbol{Q}$为

$$\boldsymbol{Q}[b,c,x,y] = \sum_{dx,dy\le kk}\boldsymbol{K}[b,dx,dy,x,y]\cdot P[b,c,x+dx,y+dy] \tag{7}$$

式中d为曼哈顿距离，b表示批次维度，$\boldsymbol{K}$为核矩阵，总体计算过程相当于一个c^2维度上的卷积。

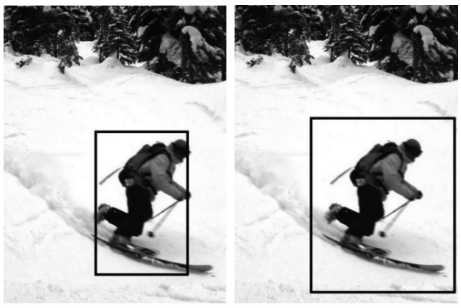
本文用分离式训练方法训练并使用CRF，相比与总体的端到端方法，分离式训练更具灵活性，它可以用于任何细化方法之中。本文将ConvCRF用在Mask R-CNN的Mask分支的类似FCN结构当中，以FCN-ConvCRF结构来代替原先Mask分支当中的FCN结构，再整合类别判断确定最终输出结果，虽然ConvCRF和Mask R-CNN在训练上没有实际交互，但增强了网络可解释性，也使网络训练更容易收敛，解决了梯度消失、梯度爆炸等难题。在正式设计实验之前，本文提取了coco数据图片集在Mask R-CNN上生成的掩码并对其用ConvCRF进行优化测试以观察ConvCRF的推理优化能力。

3　改进的RPN层

Mask R-CNN是以目标检测系统为主体来完成实例分割的工程，和Faster R-CNN边框选取机制一样，都是在RPN层和Head分支进行两次边框回归从而得出最终边框选取结果，最后的边框回归精确度要优于类似SSD[21], YOLO这样使用1次回归的算法。但Mask R-CNN是并行处理结构而不是级联结构，传到RoIAlign的数据是RPN层第1次粗略的回归结果，相对于分割任务来讲使用粗略的回归结果容易造成信息丢失。

图1为coco数据集图片在Mask R-CNN上运行过程可视化图像，具体内容为RPN层产生的两个离目标较近的候选框。

从图1可以看出，对于分割任务而言，偏大的候选框可以更全面地包含被分割对象，从而更适合分割任务。而Mask R-CNN的训练和计算过程当中，候选框的优劣是以候选框和边界框之间的IOU来判断，偏大的候选框在训练过程中容易因为



(a) 候选框1　　(b) 候选框2

图 1 RPN层运行当中两个可视化候选框

IOU值偏小而被舍弃。为了优化Mask R-CNN的RPN层因为候选框过小而导致后续分割出现虚假边缘，本文做了如下改进。

首先，本文重新设定了Mask R-CNN的锚点所对应的候选框的相对长宽，把锚点候选框相对大小增加了10%左右，使得新候选框能够多保留进一部分背景区域进行分类，从而减少丢失信息的可能性。此外，本文对Mask R-CNN数据发生器的IOU进行了改动，传统的IOU标准如式(8)所示

$$\mathrm{IOU}=\frac{S_1\cap S_2}{S_1\cup S_2} \tag{8}$$

式中S_1为候选框，S_2为真实值，对于区域$S_2-S_1\cap S_2$也就是没有囊括进去的区域惩罚力度和$S_1-S_1\cap S_2$的惩罚力度一样。对于分割任务来讲，一般更希望候选框能够涵盖所有待分割区域而不是贴紧目标，所以本文增加一个惩罚因子λ来额外惩罚目标区域没有涵盖到的情况，当$\lambda=0$时，新IOU($\mathrm{IOU}_{\mathrm{new}}$)和传统IOU一样，$\mathrm{IOU}_{\mathrm{new}}$计算公式如式(9)

$$\mathrm{IOU}_{\mathrm{new}}=\frac{S_1\cap S_2}{S_1\cup S_2+\lambda(S_2-S_1\cap S_2)} \tag{9}$$

改进后的Mask R-CNN使用新$\mathrm{IOU}_{\mathrm{new}}$产生RPN层需要的训练数据，使得囊括能力更强的候选框更有可能保留下来，从而在训练过程当中参数调整向着更多内容涵盖到候选框以内的方向发展，λ参数既保证了IOU的合理性，又给对于没有涵盖到的区域增加惩罚提供了一定弹性空间。改进的Mask R-CNN整体结构如图2所示。

4　训练数据

为了让Mask R-CNN对于模糊图像或者是受恶劣天气干扰的图像有更高的精确度和鲁棒性，本文使用一种特殊训练方法，把coco2014数据集当中的部分图像挑选出来，使用风格转换网络让图像转变成一定程度上受噪声污染的图像。与随机增加图像噪声相比，这种增加噪声的方式更符合实际情况，使得coco数据集图像能够更好地转换。

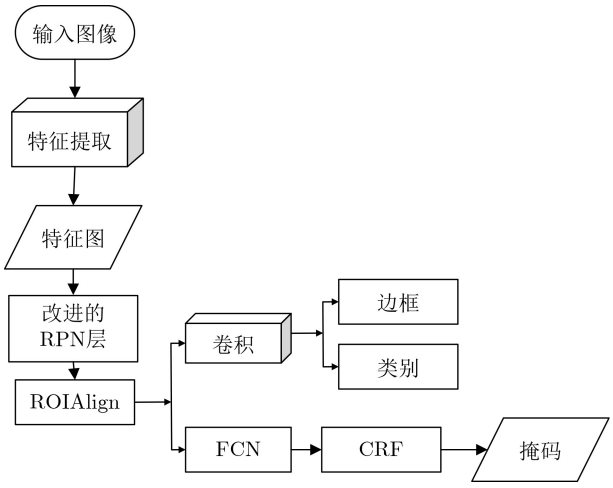


图 2 改进后Mask R-CNN流程图

风格转化网络基于VGG[22]网络改编而成，并借鉴了文献[23]的设计。VGG网络原本是根据图像识别而设计，对于图像原本信息和特征有着很强大的提取能力，本文使用了VGG网络的16个卷积层和5个池化层提供的特征空间，并通过放缩权重来规范化网络，使每个卷积滤波器的平均激活值等于1，由于只涉及激活函数的规范化，该修改可以在不改变输出的情况下对VGG进行，并采用平均池化来代替最大值池化以保留图像边缘特征。给定输入图像x在VGG给定的滤波器组上进行编码，网络当中的高层根据输入对象的特征形状和排列方式捕获高级别内容，对于图像每个位置上的其他细节特征(比如像素值)则不予保留，图像细节内容保留在网络底层当中，所以本文用高级别表示作为输入图像x的内容。对于图像的纹理特征，则建立一个由滤波器之间的相关性组成的特征空间，这些相关性由特征矩阵给出。对于输入的纹理特征图片a，仍然用VGG网络进行采样，通过对每个层次的特征提取得出关于纹理特征的多尺度表示，最终损失函数和每一层纹理特征相关。在获得图像的内容和纹理特征后初始化一个白噪声图像，从白噪声图像开始迭代使得图像内容最终和内容图像保持一致，纹理和纹理特征保持一致。图3经过风格转换网络转换的coco数据集图像与原图像的对比。

5 实验

5.1 实验数据

实验数据以coco2014数据集为基础，为了弥补coco数据集关于模糊数据的不足，本文用第4节描述的风格转换网络把coco的train2014和test2014的随机10%的图片转换成了几种不同形式的噪声图，包括模拟大雾天气的噪声图片，模拟大雨天气的噪声图片和低像素模糊图片。转换后的图片类别掩码标签和转换之前均保持不变。

(a) 原coco数据集图像 (b) 模糊化之后的coco数据集图

图 3 图像转换前后对比

5.2 网络训练

实验在4核酷睿i5CPU和4块1080ti GPU环境下进行，采用keras2.0.8和tensorflower1.12框架搭建的Mask R-CNN为基础，特征提取主干网络为ResNet101+FPN，大部分超参数设置参考了Faster R-CNN但做了一定改动。RPN层的k值从3调整到了5，训练时图片大小调整到了512×512为了更合理地分配显存，最小批次设置为8，对于本文提出的过度惩罚因子λ值设置为1。

训练时采用coco数据集预训练参数来初始化特征提取部分和Mask R-CNN的3个分支，对于改动比较大的RPN层采用随机初始化，经过两个训练阶段，每个阶段15个epoch，每个epoch有1000张图片，在第1个阶段当中锁住Mask R-CNN的特征提取和Head层只训练RPN层，经过15个epoch的迭代，实现初步候选框提取。在第2阶段当中只锁住Mask R-CNN的特征提取层训练RPN层和所有Head分支，实现对于候选框提取的进一步训练和对于Mask层的优化。

为了更好地比较改进之后Mask R-CNN的性能，排除训练数据、程序框架版本等一些无法控制的因素。本文重新训练了原Mask R-CNN以便与改进的Mask R-CNN进行对比。重新训练的Mask R-CNN和原Mask R-CNN参数设置完全相同，但使用了添加了模糊数据的coco数据集并且使用与本文一样的初始化方法。另外，为了更好地与原Mask R-CNN进行对比，本文使用未经改动的coco数据集重新训练了经过改动后的Mask R-CNN模型，初始化方法和超参数设置除了图片大小调整到512×512以外和Mask R-CNN完全相同。

5.3 实验评估

5.3.1 改进Mask分支对实验结果影响分析

为了单独观察使用ConvCRF改进后Mask分支对实验结果的影响，在使用4.2节训练出的本文模型和对比模型分割coco数据集图片过程当中，分别随机选取了100张图像对应检测框内的分割结果，计算了所有分割结果相对于coco数据集标签的平均IOU值，此外，为了更好地比较使用FCN-ConvCRF分支算法在时间上和其他CNN结合CRF结构的优越性，本文用FullCRF和ConvCRF同时优化候选框内分割结果，并单独计算程序在这一部分当中所需要的时间。

图4给出了改进Mask分支在某候选框的输出结果和原Mask分支的输出对比，由图可见，本文算法的输出结果明显修正了原Mask分支输出边界过度平滑问题，并且使得输出的分割结果对于色彩转变更为敏感。

表1对比了Mask分支和两种改进方法改进的Mask分支在时间和IOU上的差异，可以看出ConvCRF速度更快，对于处理每个候选框内分割结果，ConvCRF只使用了FullCRF时间的不到十分之一，100个随机候选框分割后的平均IOU相比于原Mask R-CNN高了0.004，可见，卷积化CRF使用GPU高效处理条件随机场运算可以在运算时间增加不大的情况下达到更好分割结果，分割边界也更为细致。

(a) coco 数据原图　(b) coco 数据集标签　(c) Mask 分支输出结果　(d) ConvCRF 分支输出结果

图 4 改进的Mask分支和原分支输出图像对比

表 1　原Mask分支与两种改进Mask分支的IOU时间(ms)对比

	Mask R-CNN	FullCRF	ConvCRF
时间	–	120	10
平均IOU	0.8831	–	0.8871

5.3.2 改进RPN层对实验结果影响分析

为了验证RPN层的对应改动是否能使候选框变得相对较大从而使信息能够完整落入候选框范围内，本文使用模型检查程序对于训练模型进行检查，该程序可以可视化展示Mask R-CNN和改进Mask R-CNN在RPN层非极大值抑制之后提取的候选框，并绘制AP曲线来查看两个模型在目标检测上的精确度和召回率。

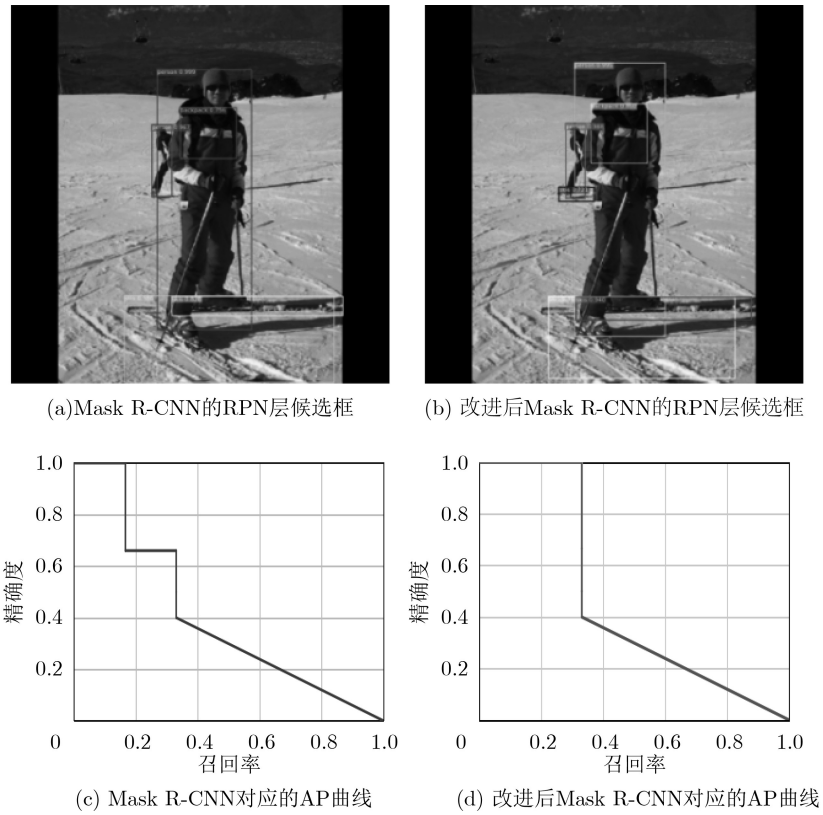
图5显示了coco数据集图片在改进Mask R-CNN在向前传播计算过程当中RPN层可视化图像和AP曲线，可以明显看出，改进的Mask R-CNN的RPN层能够产生更大候选框，从而使图像信息能够完整地保存在候选框内部，这使得分割边界进一步合理。对于coco数据集抽样图片的AP曲线分析也可以看出，在IOU等于50%的条件下，改进的Mask R-CNN综合表现要略高于原Mask R-CNN，在调整阈值使精度不变的情况下改进后的Mask R-CNN可以拥有更高召回率。

表2对比了关于RPN层改动之后算法和原算法的比较，mAP值(IOU=75)提升了0.01而mAP值(IOU=50)大致不变，对于候选框囊括能力的改进对于总体mAP值有一定贡献，这说明提升候选框囊括力不仅可以优化边界，还可以让正检率进一步提高。

5.3.3 总mAP值表现

本文采用voc风格的mAP计算方法来计算mAP值，使用框内掩码IOU为50%和75%以上的检测结果作为成功匹配项，分别计算召回率在一定给定范围内的精确度平均值作为mAP值。对比了复现Mask R-CNN和原Mask R-CNN以及两种不同训练方法的mAP值，同时也和PANet[23]同样使用Mask R-CNN为主要模型的算法以及FCIS算法等一些其他经典算法做比较以体现本文算法的优越性，最终mAP值在coco数据集和转换后的coco子数据集当中各计算1次，总mAP值如表3所示。

Mask R-CNN, mAP值(IOU=50)提升了0.05左右，mAP值(IOU=75)提升了0.03左右，一方面改进的算法候选框可以纳入更多像素进行分割，有效解决了原Mask R-CNN信息流失问题，为后面计算框内精确像素提供了更为有效的数据，另一方面，改进后Mask R-CNN的Mask分支由于使用了ConvCRF，使得分割不仅能够利用神经网络的输出也能够利用像素与像素之间的关系，提高了图像信息的利用率。使用模糊数据训练的Mask R-CNN在模糊数据集表现要优于原数据集，说明模糊训练可以提升模型鲁棒性，改进的Mask R-CNN相比较于原



(a)Mask R-CNN的RPN层候选框 (b) 改进后Mask R-CNN的RPN层候选框

(c) Mask R-CNN对应的AP曲线 (d) 改进后Mask R-CNN对应的AP曲线

图 5 RPN层可视化结果

Mask R-CNN和从新训练的Mask R-CNN在模糊处理的数据集上mAP值分别提升了0.02和0.03。但是mAP值不如原数据集，毕竟部分图像模糊化处理之后会丢失一部分图像数据导致像素分类更为困难。但使用改进之后的Mask分支进行模糊图片的分割效果依然好于原Mask R-CNN，这表明改进后的算法对于模糊数据的鲁棒性要优于原Mask R-CNN。在coco数据集上相比较于其他实例分割经典方法提升比较大。例如比MNC算法mAP值提升了0.16左右，原因是Mask R-CNN本身的优势，经典算法最终像素分类还存在比较大的跨类竞争，而且RPN层之后的ROIPool也容易发生数据丢失。Masklab[24]系列算法在Mask R-CNN基础之上融合了方向预测，可以产生更为精细的掩码结构，本文算法相比于Masklab也有一定提升，mAP值(IOU=50)和mAP值(IOU=75)上分别提升0.05和0.04左右。最近的PANet[25]是类似于Mask R-CNN的实例

表 2 mAP值对比

	mAP值(IOU=50)	mAP值(IOU=75)
原Mask R-CNN	0.60	0.39
改进的Mask R-CNN	0.60	0.40

表 3 总mAP值对比

	mAP值(IOU=50)	mAP值(IOU=75)	mAP值(模糊数据)
原Mask R-CNN	0.60	0.39	0.49
复现的Mask R-CNN(coco)	0.59	0.37	0.48
复现的Mask R-CNN(模糊数据)	0.58	0.37	0.50
改进的Mask R-CNN(模糊数据)	0.66	0.43	0.51
改进的Mask R-CNN(coco)	0.65	0.44	0.49
Mnc	0.44	0.24	–
Fcis	0.49	–	–
Masklab	0.57	0.37	
Masklab+	0.60	0.40	
PANet	0.65	0.43	–

分割算法，该算法有更为有效的特征提取网络和更复杂的Mask产生分支，改进后Mask R-CNN有着和PANet大致相当的表现。

6　结束语

本文针对实例分割的边界精确度和鲁棒性问题，提出了一种改进的Mask R-CNN算法。在网络结构上，一方面改进的Mask R-CNN采用FCN-ConvCRF分支来生成图片掩码，使得图像分割边界更为精确，另一方面改进的Mask R-CNN使用新RPN层，最终模型能够有效产生更适合分割的候选框，在一定程度上解决了图片像素丢失问题。在网络训练上，改进的Mask R-CNN使用添加了模糊图片的数据进行训练，进一步增加了鲁棒性。最终实验表明相比于原Mask R-CNN和其他经典算法，改进的Mask R-CNN在mAP值和边界精确度都有一定提高。由于R-CNN系列算法在速度上和其他算法相比耗时较多，网络较为复杂，如何提高运行速率将会是下一步研究的重点。

参 考 文 献

[1] SHELHAMER E, LONG J, and DARRELL T. Fully convolutional networks for semantic segmentation[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017, 39(4): 640–651. doi: 10.1109/TPAMI.2016.2572683.

[2] REN Shaoqing, HE Kaiming, GIRSHICK R, *et al*. Faster R-CNN: Towards real-time object detection with region proposal networks[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017, 39(6): 1137–1149. doi: 10.1109/TPAMI.2016.2577031.

[3] REDMON J, DIVVALA S, GIRSHICK R, *et al*. You only look once: Unified, real-time object detection[C]. The Conference on Computer Vision and Pattern Recognition, Las Vegas, USA, 2016: 779–788. doi: 10.1109/CVPR.2016.91.

[4] REDMON J and FARHADI A. YOLO9000: Better, faster, stronger[C]. The Conference on Computer Vision and Pattern Recognition, Honolulu, USA, 2017: 6517–6525. doi: 10.1109/CVPR.2017.690.

[5] DAI Jifeng, HE Kaiming, and SUN Jian. Instance-aware semantic segmentation via multi-task network cascades[C]. The Conference on Computer Vision and Pattern Recognition, Las Vegas, USA, 2016: 3150–3158. doi: 10.1109/CVPR.2016.343.

[6] DAI Jifeng, HE Kaiming, LI Yi, *et al*. Instance-sensitive fully convolutional networks[C]. The 14th European Conference on Computer Vision, Amsterdam, The Netherlands, 2016: 534–549.

[7] LI Yi, QI Haozhi, DAI Jifeng, *et al*. Fully convolutional instance-aware semantic segmentation[C]. The Conference on Computer Vision and Pattern Recognition, Honolulu, USA, 2017: 4438–4446. doi: 10.1109/CVPR.2017.472.

[8] BAI Min and URTASUN R. Deep watershed transform for instance segmentation[C]. The Conference on Computer Vision and Pattern Recognition, Honolulu, USA, 2017: 2858–2866. doi: 10.1109/CVPR.2017.305.

[9] LIU Shu, JIA Jiaya, FIDLER S, *et al*. SGN: Sequential grouping networks for instance segmentation[C]. 2017 IEEE International Conference on Computer Vision, Venice, Italy, 2017: 3516–3524. doi: 10.1109/ICCV.2017.378.

[10] HE Kaiming, GKIOXARI G, DOLLÁR P, *et al*. Mask R-CNN[C]. 2017 IEEE International Conference on Computer Vision, Venice, Italy, 2017: 2980–2988.

[11] PINHEIRO P O, COLLOBERT R, and DOLLÁR P. Learning to segment object candidates[C]. The 28th International Conference on Neural Information Processing Systems, Montreal, Canada, 2015: 1990–1998.

[12] PINHEIRO P O, LIN T Y, COLLOBERT R, *et al*. Learning to refine object segments[C]. The 14th European Conference on Computer Vision, Amsterdam, The Netherlands, 2016: 75–91. doi: 10.1007/978-3-319-46448-0_5.

[13] ZAGORUYKO S, LERER A, LIN T Y, *et al*. A multipath network for object detection[C]. The British Machine Vision Conference, Edinburgh, England, 2016. doi: 10.5244/C.30.15.

[14] 罗会兰, 卢飞, 孔繁胜. 基于区域与深度残差网络的图像语义分割[J]. 电子与信息学报, 2019, 41(11): 2777–2786. doi: 10.11999/JEIT190056.

LUO Huilan, LU Fei, and KONG Fansheng. Image semantic segmentation based on region and deep residual network[J]. *Journal of Electronics & Information Technology*, 2019, 41(11): 2777–2786. doi: 10.11999/JEIT190056.

[15] CHEN L C, PAPANDREOU G, KOKKINOS I, *et al*. DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2018, 40(4): 834–848. doi: 10.1109/TPAMI.2017.2699184.

[16] ZHENG Shuai, JAYASUMANA S, ROMERA-PAREDES B, *et al*. Conditional random fields as recurrent neural networks[C]. 2015 IEEE International Conference on Computer Vision, Santiago, Chile, 2015: 1529–1537.

[17] 韩铮, 肖志涛. 基于纹元森林和显著性先验的弱监督图像语义分割方法[J]. 电子与信息学报, 2018, 40(3): 610–617. doi: 10.11999/JEIT170472.

HAN Zheng and XIAO Zhitao. Weakly supervised semantic segmentation based on semantic texton forest and saliency prior[J]. *Journal of Electronics & Information Technology*,

2018, 40(3): 610–617. doi: 10.11999/JEIT170472.

[18] KRÄHENBÜHL P and KOLTUN V. Efficient inference in fully connected CRFs with Gaussian edge potentials[C]. The 24th International Conference on Neural Information Processing Systems, Granada, Spain, 2011: 109–117.

[19] TEICHMANN M T T and CIPOLLA R. Convolutional CRFs for semantic segmentation[EB/OL]. https://arxiv.org/abs/1805.04777, 2018.

[20] LAFFERTY J, MCCALLUM A, and PEREIRA F C N. Conditional random fields: Probabilistic models for segmenting and labeling sequence data[C]. The 18th International Conference on Machine Learning, San Francisco, CA, USA, 2001: 282–289.

[21] LIU Wei, ANGUELOV D, ERHAN D, *et al*. SSD: Single shot MultiBox detector[C]. The 14th European Conference on Computer Vision, Amsterdam, The Netherlands, 2016: 21–37. doi: 10.1007/978-3-319-46448-0_2.

[22] SIMONYAN K and ZISSERMAN A. Very deep convolutional networks for large-scale image recognition[EB/OL]. http://arxiv.org/abs/1409.1556v6, 2014.

[23] GATYS L A, ECKER A S, and BETHGE M. Image style transfer using convolutional neural networks[C]. 2016 IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, USA, 2016: 2414–2423. doi: 10.1109/CVPR.2016.265.

[24] CHEN L C, HERMANS A, PAPANDREOU G, *et al*. MaskLab: Instance segmentation by refining object detection with semantic and direction features[C]. The IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, USA, 2018: 4013–4022.

[25] LIU Shu, QI Lu, QIN Haifang, *et al*. Path aggregation network for instance segmentation[C]. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, USA, 2018: 8759–8768. doi: 10.1109/CVPR.2018.00913.

陈卫东：男，1971年生，教授，研究方向为智能算法及应用.

郭蔚然：男，1992年生，硕士生，研究方向为深度学习图像分割.

刘宏炜：男，1995年生，硕士生，研究方向为深度学习图像分割.

朱奇光：男，1978年生，副教授，研究方向为智能机器人检测与控制.

责任编辑：余 蓉