

新型 DRNet 结合 EIoU 的遮挡目标分割模型 *

陈丹¹ 令陈佩¹ 刘瑞瑜²

(1. 福州大学电气工程与自动化学院 福州 350100; 2. 福州大学梅努斯国际工程学院 福州 350100)

摘要: 实例分割是计算机视觉领域的重要研究方向,但由于遮挡问题的存在,使得该任务仍然没有得到充分探索。针对目前算法对遮挡物体的分割检测效果不佳,容易出现误检漏检问题,在 Mask R-CNN 框架基础上,提出一种新型双向残差网络(DRNet)结合 EIoU 的遮挡目标分割模型。首先,提出一种 DRNet 代替原有 ResNet 网络,使用更少的 BN 层和 ReLU 层取代传统 Conv-BN-ReLU 结构,利用传统卷积和深度可分离卷积串行连接增强图像感受野特征,通过跳跃连接减轻网络随深度增加出现退化问题,提升网络表征能力;其次,使用 CEIoU NMS 算法代替原有 NMS 算法,通过聚类思想有效处理重叠边界框抑制问题,引入 EIoU 评估指标增加边界框几何信息,更加精准地描述边界框之间的相似程度,减少网络对遮挡物体边界框的错误抑制;最后,使用 EIoU 损失替换原有 Smooth L1 损失,加速网络收敛速度,提升边界框检测精度。在公共 COCO 2017 数据集上进行预训练,再在不同程度的遮挡数据集上进行实验。实验结果表明,相比较于原网络,所提分割算法在 COCO 2017 数据集上 Box AP 和 Mask AP 分别提升了 1.7% 和 1.3%;在遮挡数据集上对遮挡物体边界框检测精度和掩码分割精度均有明显提升,证实该方法对遮挡物体分割的有效性。

关键词: 遮挡物体; 实例分割; DRNet; Cluster EIoU NMS; EIoU 损失

中图分类号: TP391.41; TN911.73 文献标识码: A 国家标准学科分类代码: 520.2060

Novel DRNet occlusion target segmentation model combined with EIoU

Chen Dan¹ Ling Chenpei¹ Liu Ruiyu²

(1. School of Electrical Engineering and Automation, Fuzhou University, Fuzhou 350100, China;
2. Maynooth International College of Engineering, Fuzhou University, Fuzhou 350100, China)

Abstract: Instance segmentation is an important research direction in the field of computer vision, but the existence of the occlusion problem still prevents this task from being fully explored. To address the poor segmentation detection of occluded objects by current algorithms, which are prone to the problems of misdetection and omission, a novel duplex residual network (DRNet) is proposed, combining the EIoU occluded target segmentation model with the Mask R-CNN framework. First, DRNet is proposed to replace the original ResNet network, using fewer BN and ReLU layers to replace the traditional Conv-BN-ReLU structure, utilizing the conventional convolution and depth-separable convolution serial connection to enhance the image sensory field features, and mitigating the degradation problem of the network with the increase of the depth by the hopping connection. Second, the CEIoU NMS algorithm is used instead of the original NMS algorithm to effectively deal with the overlapping bounding box suppression problem through the clustering idea, and the introduction of the EIoU evaluation index increases the bounding box geometric information, which more accurately describes the degree of similarity between the bounding boxes, and reduces the network's erroneous suppression of the bounding boxes of the occluded objects. Finally, the EIoU loss is used to replace the original Smooth L1 loss to accelerate the network convergence speed and improve the bounding box detection accuracy. In this paper, we first conduct pre-training on the public COCO 2017 dataset and experiments on different degrees of occlusion datasets, and the results show that compared with the original network, the proposed segmentation algorithm improves the Box AP and Mask AP by 1.7% and 1.3% on the COCO 2017 dataset, respectively; and both the bounding-box detection

accuracy of the occluded object and the mask segmentation accuracy on the occlusion dataset are significantly improved on the occlusion dataset, confirming the effectiveness of the method for occluded object segmentation.

Keywords: occluded objects; instance segmentation; DRNet; CEIoU NMS; EIoU loss

0 引言

目标检测和实例分割是计算机视觉领域中两大重要且关系紧密的任务,随着神经网络和深度学习技术的崛起,该领域研究取得突破性进展,进而衍生出大批具有实际意义的应用场景,例如自动驾驶,医学成像,无人机监测和机器人抓取等。

当前实例分割算法主要包括单阶段方法和双阶段方法。单阶段方法主要有如下 3 种:YOLACT 网络^[1],利用生成一组原型掩码和预测每个实例掩码系数并行得到最终的分割掩码,并利用 Fast NMS 优化模型,实现实时的实例分割;RDSNet 网络^[2],通过引入物体级别和像素级别学习特征,在目标检测和实例分割之间实现双向性,以获取更高精度的分割结果和更准确的边界框;SOLO 网络^[3],通过引入“实例类别”的概念,将实例掩码分割转化成分类问题,利用简单的框架实现较强的性能。双阶段网络主要是基于 Mask R-CNN 网络^[4]:在 Faster R-CNN 网络^[5]基础上增加全卷积网络分支,采用先检测后分割的策略,同时实现目标检测和掩码分割;Cascade Mask R-CNN 网络^[6],在 Mask R-CNN 网络的基础上,使用不同的 IoU 阈值训练多个级联的检测器,提高网络预测精度;HTC 网络^[7],在 Mask R-CNN 网络和 Cascade Mask R-CNN 网络基础之上,设计多任务多阶段的混合级连结构,实现多任务间的消息共享和联合优化。

然而,当目标物体被遮挡时,其部分或全部特征信息会丢失,导致模型难以准确地识别和分割出被遮挡的目标^[8]。此外,由于遮挡目标的存在,不同目标的检测框之间可能会严重重叠,容易被非极大抑制算法看作同一个预测目标而错误抑制,造成漏检的发生。为应对上述问题,目前已经有研究者提出了遮挡检测优化算法。Zhang 等^[9]在 Faster R-CNN 基础上设计部分遮挡感知区域池化层代替 RoI 池化层,将行人划分为 5 个部分,通过对各个部分的检测估计行人的遮挡状态,但行人各个部分均具有丰富的特征信息,当被检测物体特征信息不充分时,难以通过分部检测实现遮挡状态估计。Liu 等^[10]在 R-FCN^[11]基础上增加了多层特征提取网络,并将每一层的特征信息进行融合,该方法充分利用特征提取网络采集到的特征信息,增强输出特征图的表征能力,提高网络对遮挡目标的检测精度,但当特征提取网络本身效果不佳时,难以确保网络对遮挡问题的有效性。Yang 等^[12]在 RetinaNet 基础之上,提出 Rep-GIoU Loss,通过改进损失

函数来改善检测器在遮挡环境下的性能,是代价最小且最具解释性的方式之一,但是大多数损失函数的设计无法充分考虑遮挡条件的复杂性,容易造成漏检的情况;Luo 等^[13]将非极大值抑制算法(NMS)加入网络模型训练,提出 NMS Loss, NMS Loss 考虑了假阴性和假阳性对目标检测结果的影响,对假阳性没有被抑制而假阴性被 NMS 错误地删除进行惩罚,减少网络对遮挡目标检测框的过度抑制。但 NMS Loss 只适用于单类目标的检测^[14],其在一般检测中的使用仍需要进一步的研究。

针对上述问题,本文在 Mask R-CNN 框架基础之上,提出一种针对遮挡物体的检测与分割算法。设计了一种双向残差特征提取网络 (duplex residual network, DRNet),将传统卷积和深度可分离卷积^[15]串行连接,充分提取细节特征,利用跳跃连接^[16]减轻网络退化情况,显著增强特征提取能力;改进 NMS 算法,使用 CEIoU NMS 算法替换原有 NMS 算法,在 Cluster NMS 算法^[17]基础上,使用 EIoU 作为边界框评价指标,有效提升 NMS 算法推理速度,减少网络对遮挡物体边界框的过度抑制;使用 EIoU 损失^[18]计算边界框回归损失,促使模型更准确地预测目标框的位置和大小,提高目标检测任务的定位精度,加速网络收敛速度。

1 改进 Mask R-CNN 算法

1.1 算法整体框架

本文算法以 Mask R-CNN 网络为基础,整体框架如图 1 所示,训练图像经预处理以消除图像特征单位和尺度差异的影响,提高数据的多样性;将预处理后的图像输入至主干网络,提取图像特征信息;接着通过特征金字塔网络^[19]进行多尺度图像特征融合,生成预测特征金字塔;区域生成网络从特征金字塔网络输出的预测特征金字塔中寻找可能包含目标的兴趣区域,提高网络预测定位能力;感兴趣区域经不同大小的尺寸对齐后输入到快速卷积预测器^[20]和全卷积网络;快速卷积预测器对区域生成网络的输出特征图和感兴趣区域进行特征提取和分类,获取目标的类别信息和边界框信息;全卷积网络在对齐后的感兴趣区域内部对每个像素进行标签预测,实现实例级别的语义分割;后处理部分将网络预测的类别信息,边界框信息和掩码信息经过过滤、合并等处理,将最终预测结果绘制到原图像,实现目标检测与分割任务。

1.2 双向残差网络

在遮挡条件下,目标特征信息部分丢失,遮挡目标之

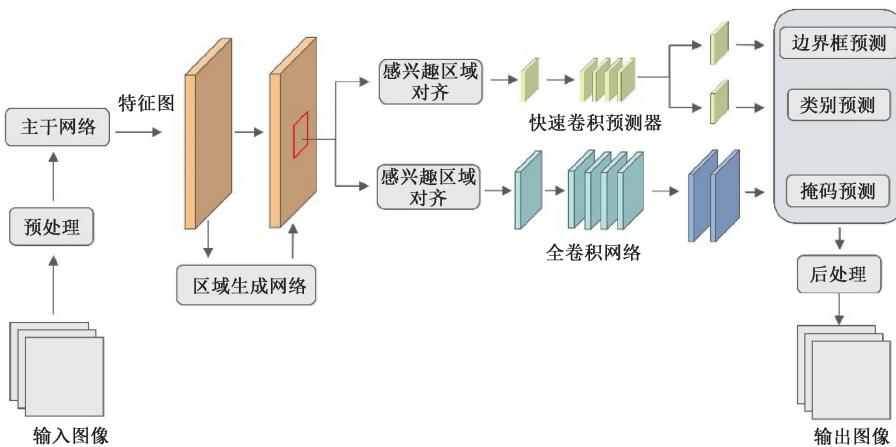


图 1 算法整体框架

Fig. 1 Overall framework of algorithm

间特征信息互相干扰,导致网络对遮挡目标的特征信息提取困难,造成检测和分割效果不佳。针对上述问题,本文设计了一种双向残差网络 DRNet,通过串联传统卷积和深度可分离卷积提取更加丰富的特征信息,利用跳跃连接减少网络退化^[21],增加网络深度,提升网络特征提取能力。

DRNet 网络组成结构如图 2 所示,当输入通道数和输出通道数相同时,网络采用双向残差基础结构。输入特征图 $F \in \mathbf{R}^{c \times w \times h}$,通过 DRB 结构正向传播,并利用跳跃连接将原始输入和网络输出叠加,得到最终得到输入输出尺寸相同的特征图 $\hat{F} \in \mathbf{R}^{c \times w \times h}$ 。第 1 个 1×1 卷积层对特征图降维处理,减少通道个数,降低计算量;再通过 3×3 卷积层捕捉输入特征图的细节特征信息,提高特征图表征能力。第 2 个 1×1 卷积层对特征图升维处理,以提高非线性能力;再通过 5×5 深度卷积层^[22],在更高的维度通过逐点卷积实现通道间的信息融合,使用更大的卷积核可以拥有更广阔的感受野,进一步增加特征信息。第 3 个 1×1 卷积层对特征图再做升维处理,保证输入输出特征通道数相同;最后通过跳跃连接减少梯度消失,防止网络退化。不同于传统的 Conv-BN-ReLU 结构,DRB 使用更少的 BN 层和 ReLU 层^[23],虽然 BN 层能够加速网络收敛,减少梯度消失或爆炸问题,但 BN 层引入了额外的计算量,并导致网络结构复杂化,增加调参难度,适当减少 BN 层可以降低计算开销,减少过度正则化,保留更多特征信息;ReLU 层通过截断负数值引入非线性特征,增加网络表征能力,过多的 ReLU 使得许多神经元输出为零,损坏特征信息,减少 ReLU 层以增强信息流动,缓解梯度消失问题。

当输入通道数和输出通道数不同时,网络采用双向残差跳跃结构,在双向残差基础结构上,将 3×3 卷积层的步距设置为 2,对输入特征图进行下采样;利用第 3 个 $1 \times$

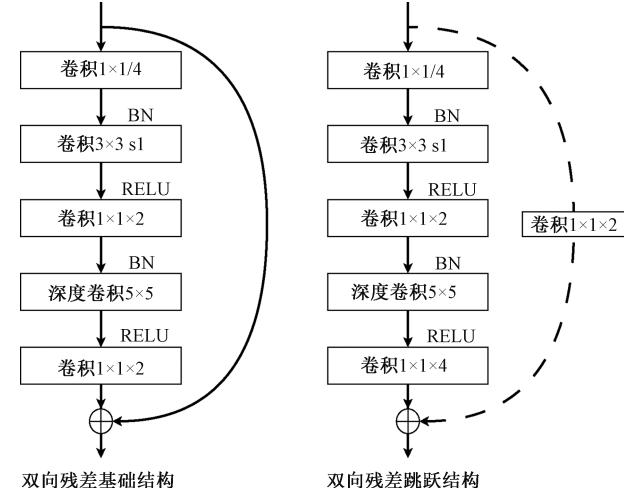


图 2 DRNet 网络组成结构

Fig. 2 DRNet network infrastructure

1 卷积层对特征图升维处理,并在跳跃连接时使用一个 1×1 卷积层对原始输入特征层升维,确保梯度传播,最终得到输出特征图 $\hat{F} \in \mathbf{R}^{2c \times w/2 \times h/2}$ 。

DRNet 网络完整结构如图 3 所示,结构参数如表 1 所示。

表 1 DRNet 网络结构参数

Table 1 DRNet Network structure parameter

层级	输出大小	网络参数
层 0	$48 \times \frac{w}{4} \times \frac{h}{4}$	卷积 7×7 , 步距 2 最大池化层, 步距 2
层 1	$192 \times \frac{w}{4} \times \frac{h}{4}$	双向残差基础结构 $\times 1$ 双向残差基础结构 $\times 2$
层 2	$384 \times \frac{w}{8} \times \frac{h}{8}$	双向残差跳跃结构 $\times 1$ 双向残差基础结构 $\times 2$
层 3	$768 \times \frac{w}{16} \times \frac{h}{16}$	双向残差跳跃结构 $\times 1$ 双向残差基础结构 $\times 8$
层 4	$1536 \times \frac{w}{32} \times \frac{h}{32}$	双向残差跳跃结构 $\times 1$ 双向残差基础结构 $\times 2$

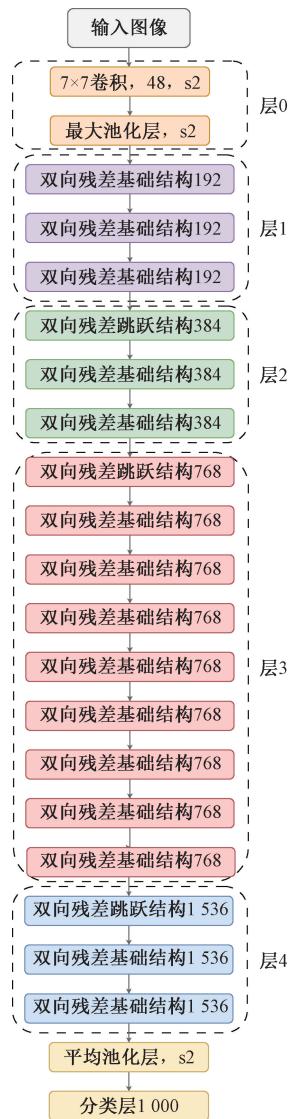


图 3 DRNet 网络结构

Fig. 3 DRNet network structure

1.3 CEIoU NMS 算法

在检测遮挡目标时,传统 NMS 算法存在如下两点问题:1)使用交并比(intersection over union, IoU)作为算法抑制评价指标,无法很好地衡量边界框的重叠程度,可能导致某些有效目标被错误移除,造成漏检问题;2)对于大量候选检测框,计算每个检测框之间的 IoU 并进行比较,计算复杂度较高,处理速度较慢。因此,本文使用 CEIoU NMS 算法,在 Cluster NMS 算法基础上,利用 EIoU 抑制评价指标,加速 NMS 算法推理速度,减少网络对遮挡物体边界框的过度抑制。

Cluster NMS 算法通过引入聚类步骤,在每个聚类内部执行 NMS 操作,通过分组形式减少需要比较的边界框数量,提高效率和效果。由于 \mathbf{b} 矩阵为 0 的元素不会对抑制结果产生影响,使得 \mathbf{A} 和矩阵 \mathbf{X} 的乘积操作将没有

重叠的预测框被隐式地分组到不同的聚类中,并在聚类间执行抑制,有效改善网络过度抑制问题,并且利用 GPU 加速矩阵运算,提升 NMS 推理速度。

算法: Cluster-NMS

```

输入:  $N$  个预测边界框,  $\mathbf{B} = [B_1, B_2, \dots, B_N]^T$ , 抑制框阈值  $\varepsilon$ 
输出:  $\mathbf{b} = \{b_i\}_{1 \times N}, \mathbf{b} \in [0, 1]$ 
1:  $T = N, t = 1, t^* = T, \mathbf{b}^0 = 1$ 
2: 计算 IoU 矩阵  $\mathbf{X} = \{x_{ij}\}_{N \times N}, x_{ij} = \text{IoU}(B_i, B_j)$ 
3:  $\mathbf{X} = \text{triu}(\mathbf{X}) \&& x_{ii} = 0$ 
4: while  $t \leq T$  do
5:    $\mathbf{A}^t = \text{diag}(\mathbf{b}^{t-1})$ 
6:    $\mathbf{C}^t = \mathbf{A}^t \times \mathbf{X}$ 
7:    $\mathbf{g} = \max_i \mathbf{C}^t$ 
8:    $\mathbf{b}^t = \begin{cases} b_i = 1, & g_i > \varepsilon \\ b_i = 0, & g_i < \varepsilon \end{cases}$ 
9:   if  $\mathbf{b}^t = \mathbf{b}^{t-1}$ 
10:     $t^* = t$ , break
11:  end if
12: end while
13: return  $\mathbf{b}^{t^*}$ 

```

传统 NMS 使用 IoU 作为抑制评价指标,用于衡量两个边界框之间的重叠程度, IoU 定义公式如下:

$$IoU = \frac{\text{area}_A \cap \text{area}_B}{\text{area}_A \cup \text{area}_B} \quad (1)$$

式中: area_A 和 area_B 分别表示 A 、 B 两个边界框的面积,但 IoU 仅计算两个边界框重叠区域与并集的比值,不考虑边界框的其他几何特征,当两个框高度重叠但形状和位置不同时, IoU 无法进行有效区分。本文在 Cluster NMS 基础上,引入 EIoU 作为抑制评价指标,增加边界框中心点距离和长宽比差异惩罚项,提供更多几何信息, EIoU 定义公式如下:

$$EIoU = IoU - \frac{\rho^2(b_A, b_B)}{c^2} - \frac{\rho^2(w_A, w_B)}{C_w^2} - \frac{\rho^2(h_A, h_B)}{C_h^2} \quad (2)$$

式中: c, C_w 和 C_h 分别表示 A, B 边界框最小外接矩形对角线长度,宽度和高度; $\rho^2(b_A, b_B)$ 、 $\rho^2(w_A, w_B)$ 和 $\rho^2(h_A, h_B)$ 分别表示 A, B 边界框几何中心点之间的欧氏距离,宽度差平方和高度差平方。 EIoU 能够抑制 IoU 得分,提供更高的定位精度,减少 NMS 错误抑制。

1.4 损失函数

损失函数提供衡量标准,用于衡量模型预测值和真实值之间的差异,通过最小化损失函数更新模型参数,提升网络效果, Mask R-CNN 网络损失定义如下:

$$\mathcal{L}_{loss} = \mathcal{L}_{act} + \mathcal{L}_{arl} + \mathcal{L}_{cla} + \mathcal{L}_{reg} + \mathcal{L}_{seg} \quad (3)$$

式中: \mathcal{L}_{act} 、 \mathcal{L}_{cla} 和 \mathcal{L}_{seg} 分别表示 RPN 分类损失、目标检测分类损失以及掩码分类损失。使用二元交叉熵损

失(binary cross-entropy loss)统计计算,损失函数定义如下:

$$\mathcal{L}_{ce} = \frac{1}{N_{ce}} \sum_i CE(\rho_i, \rho_i^*) \quad (4)$$

式中: ρ_i 和 ρ_i^* 分别表示第 i 个目标的预测概率和真实标签; CE 表示交叉熵损失。

\mathcal{L}_{arl} 和 \mathcal{L}_{reg} 分别表示 RPN 边界框回归损失和目标检测边界框回归损失, 使用 Smooth L1 损失统计计算, 损失函数定义如下:

$$\mathcal{L}_{sl} = \frac{1}{N_{sl}} \sum_i y_i^* \cdot \text{Smooth L1}(t_i - t_i^*) \quad (5)$$

式中: y_i^* 为指示变量, 当第 i 个边界框为正样本时, $y_i^* = 1$, 否则 $y_i^* = 0$; t_i 和 t_i^* 分别代表第 i 个边界框预测回归参数和真实回归参数。

Smooth L1 损失结合了 L1 和 L2 损失的优点, 对异常值具有一定的鲁棒性, 在误差较小时有较好的优化效果。然而, Smooth L1 损失函数在目标边界框回归损失的应用中存在一些局限性: 当误差较大时, Smooth L1 损失函数变为线性增长, 对大误差的惩罚力度不足; 在边界框定位问题上, 仅仅考虑边界框回归参数的差值, 无法很好衡量边界框的位置和几何信息, 影响网络边界框定位能力; 针对上述问题, 本文提出使用 EIoU 损失函数统计边界框回归损失, EIoU 引入了更多的几何信息, 提供了更丰富的梯度, 有助于加快模型训练的收敛速度, EIoU 损失函数定义如下:

$$\mathcal{L}_{EIoU} = 1 - \text{IoU} + \frac{\rho^2(b_A, b_B)}{c^2} + \alpha \frac{\rho^2(w_A, w_B)}{C_w^2} + \beta \frac{\rho^2(h_A, h_B)}{C_h^2} \quad (6)$$

式中: α 和 β 表示权重参数, 用于平衡长宽比差异对损失的影响。

2 实验与分析

2.1 数据集与评价指标

本文使用 COCO (common objects in context) 数据集和自制的遮挡数据集进行网络训练和测试实验。COCO 数据集是一个大型的、丰富的物体检测和分割数据集, 数据集中包含 80 种常见目标类别, 超过 330 000 张图像, 包含 150 万个目标。遮挡数据集共收集 3 580 张图像作为训练数据, 该数据集共包含 5 种检测目标, 钳子、扳手、剪刀、螺丝刀和美工刀。根据遮挡程度划分为无遮挡, 轻微遮挡 (0% ~ 20%), 一般遮挡 (20% ~ 40%) 和严重遮挡 (40% ~ 60%), 无遮挡数据集包括 930 张训练集, 310 张验证集; 轻微遮挡, 一般遮挡和严重遮挡数据集均包含

520 张训练集和 260 张验证集。模型在 COCO 数据集上进行预训练, 并在不同遮挡程度的数据集进行训练和测试。

本文实验采用 COCO 数据集中设定的平均精度 (average precision, AP) 以及帧率作为评价指标, 平均精度包括 IoU 阈值从 0.50 ~ 0.95, 步长为 0.05 的 10 个值的平均 AP, IoU 阈值为 0.50 和 0.75 下的 AP@0.5 和 AP@0.75。

2.2 实验环境与训练参数

本文实验得到福建省超算中心算力支持, 使用 Liunx 超算集群, 系统内核版本 CentOS Linux release 7.9.2009 (Core), 4 块 Nvidia Tesla P100 显卡, 在 PyTorch 框架下进行实验, 使用 SGD 优化器完成梯度下降, 开启混合精度训练, 学习率设置为 0.02, 每 100 轮训练学习率衰减 0.1, 共迭代 300 轮, EIoU 损失函数 α 和 β 权重参数设置为 1。

2.3 实验结果与分析

本文在 COCO 数据集上进行预训练, 结果如表 2 所示。相比较与原网络, 本文提出的网络在 COCO 数据集上 Box AP 提升了 1.7%, Mask AP@0.5 略有下降, 但平均 Mask AP 和 Mask AP@0.75 值分别提升了 1.3% 和 2.1%。

表 2 COCO 数据集网络预训练结果对比

Table 2 Comparison of network pre-training

results for the COCO dataset (%)

方法	Box AP	Mask AP	Mask AP@0.5	Mask AP@0.75
Mask R-CNN	37.5	33.6	55.2	35.3
本文	39.2	34.9	54.7	37.4

本文在遮挡数据集上与 Mask R-CNN, Yolact, Cascade Mask R-CNN, MS R-CNN 网络进行对比, 结果如表 3 所示。以 Mask R-CNN 网络为基准线, 本文在轻度遮挡, 一般遮挡和严重遮挡情况下的 Box AP 分别提升 5.8%、5.2% 和 7.8%, Mask AP 分别提升了 4.2%、4.0% 和 2.8%, 相比较于其他经典的二阶段分割算法, 本文算法的分割精度更高; Cascade Mask R-CNN 通过级联结构显著提高了检测和分割的精度, 但代价是更高的计算复杂度和训练难度; 与单阶段网络 YOLACT 相比, 检测精度与分割精度明显提升。

网络实例分割推理结果如图 4 所示, YOLACT 网络对钳子造成漏检, Mask R-CNN 和 MS R-CNN 网络对钳子分割结果不够精细; Mask R-CNN 和 YOLACT 网络对美工刀多余分割, Cascade Mask R-CNN 和 MS R-CNN 对钳子尖嘴部分分割不够完整; 本文的分割结果相对较好, 分割结果更加完整, 证明了本文所提方法的有效性, 提升了网络对遮挡物体的分割能力。

表 3 实例分割模型网络精度对比

Table 3 Example segmentation network model precision comparison

方法	遮挡程度	帧率/FPS	Box AP/%	Mask AP/%	Mask AP@0.5/%	Mask AP@0.75/%
Mask R-CNN	轻微遮挡	12.48	86.3	82.2	98.9	90.4
	一般遮挡	11.35	69.0	68.8	93.3	85.8
	严重遮挡	11.87	62.1	67.9	90.5	83.2
YOLACT	轻微遮挡	23.67	71.5	76.5	92.4	81.8
	一般遮挡	23.42	62.1	60.8	89.7	74.9
	严重遮挡	22.84	54.8	51.9	86.6	58.2
Cascade Mask R-CNN	轻微遮挡	10.54	90.5	83.0	99.3	91.8
	一般遮挡	11.79	75.6	70.6	95.4	89.0
	严重遮挡	11.20	71.4	68.5	93.1	82.8
MS R-CNN	轻微遮挡	13.11	87.2	82.3	98.7	90.8
	一般遮挡	12.97	73.8	69.4	93.6	80.7
	严重遮挡	12.91	68.0	66.9	94.3	83.1
本文	轻微遮挡	14.32	92.1	86.4	99.7	92.0
	一般遮挡	13.62	74.2	72.8	92.7	88.6
	严重遮挡	14.01	69.9	70.7	93.6	85.2

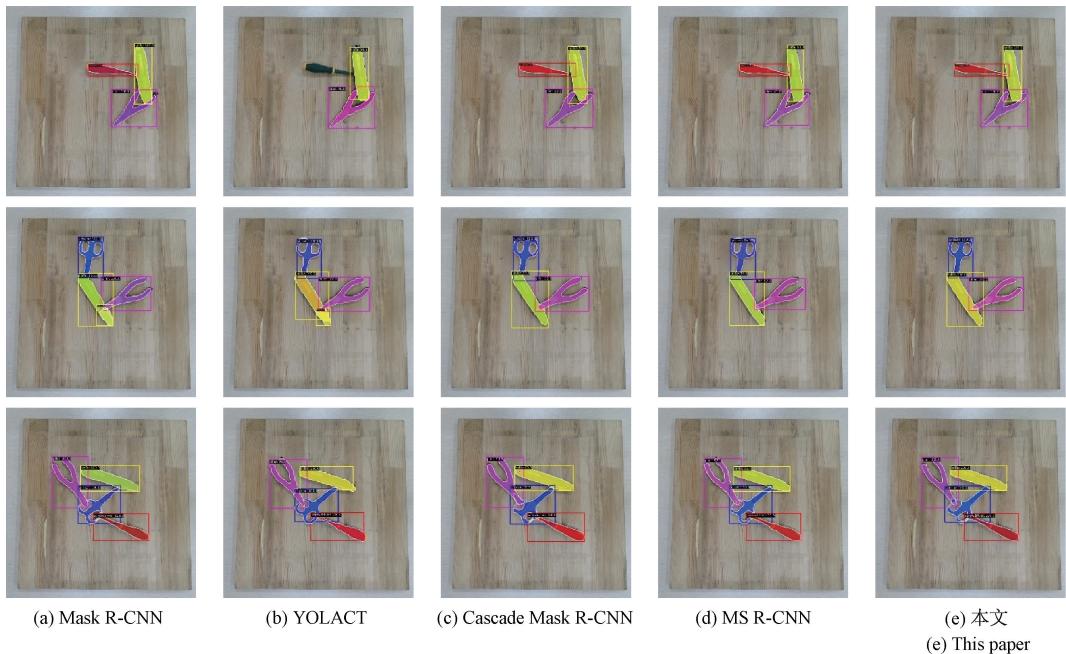


图 4 实例分割网络推理结果

Fig. 4 Example segmentation network inference results

2.4 消融实验

为验证本文所提出的 DRNet、CEIoU NMS 和 EloU 损失函数的有效性, 在遮挡数据集上设计消融实验, 实验结果如表 4 所示。

实验表明, 加入 CEIoU NMS 算法和 EloU 损失后网络推理速度和网络精度均有所提升, 证明了各个模块的有效性; 相比较于 ResNet-50 网络, 本文所提 DRNet 对于网络提升明显, 平均 Box AP 和 Mask AP 分别提升了 3.2% 和 1.7%, 证明本文所提主干网络具有更加出色的特征提取能力, 有利于增强网络定位和分割精度, 提升网

络推理速度, 结合可视化结果进一步说明本文所提方法的有效性。

表 4 消融实验结果

Table 4 Results of the ablation experiment

主干网络	CEIoU NMS	EIoU 损失	帧率/FPS	Box AP/%	Mask AP/%
ResNet-50	×	×	12.48	86.3	82.2
	√	×	13.67	86.7	83.3
	√	√	13.94	88.9	84.7
DRNet	√	√	14.32	39.2	34.9

首先,在原网络基础上引入 CEIoU NMS 模块,待网络训练收敛后,使用模型进行推理,可视化边界框预测结果,如图 5 所示。

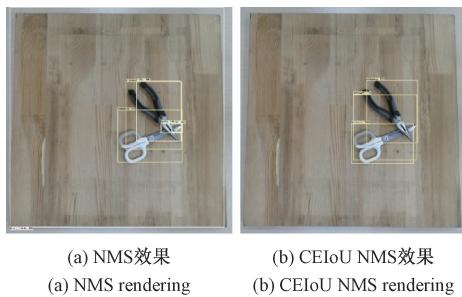


图 5 NMS 与 c EIoU NMS 效果对比

Fig. 5 Comparison of NMS and Cluster EIoU NMS effects

由推理结果可知,Cluster EIoU NMS 对边界框预测准确性更高,对多余边界框抑制效果更好;其次,引入 EIoU 损失函数进行训练,可视化网络训练过程中的边界框损失,如图 6 所示。

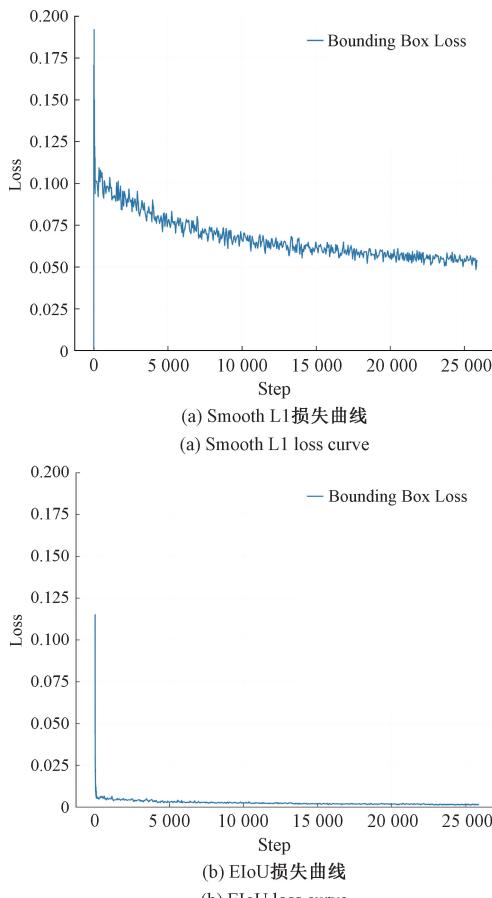


图 6 Smooth L1 与 EIoU 边界框回归损失曲线对比

Fig. 6 Comparison of Smooth L1 and EIoU boundary regression loss curves

加入 EIoU 损失后网络训练损失曲线更加平滑,边界框回归损失值明显降低,表明训练过程中 EIoU 损失函数能够更好地反映预测框和真实框之间的差异,提供更加丰富的梯度信息,有效指导网络模型优化方向,提升模型边界框检测精度;为验证不同 α 和 β 的值对网络模型的影响,采用多组不同值进行测试,实验结果如表 5 所示。

表 5 对比实验结果

Table 5 Comparison of experimental results

α	β	Box AP/%	Mask AP/%
0.5	2.0	85.7	82.9
0.7	1.4	87.4	83.5
1.0	1.0	88.9	84.7
1.4	0.7	88.1	84.2
2.0	0.5	86.1	83.2

通常图像中矩形检测框边宽跨度略微高于边高跨度,使得 α 权重对模型影响略高于 β 权重;综合场景下,由于目标预测边界框宽高比随机性较强,当 α 和 β 权重相当时,网络模型效果最优。

最后,将原有 ResNet-50 替换为 DRNet,结合网络推理结果对比模型性能,如图 7 所示。引入 DRNet 前,网络对物体部分检测效果不佳;引入 DRNet 后网络能够有效识别目标,减少检测错误率,表明 DRNet 能够有效提升特征信息,提升模型检测效果。

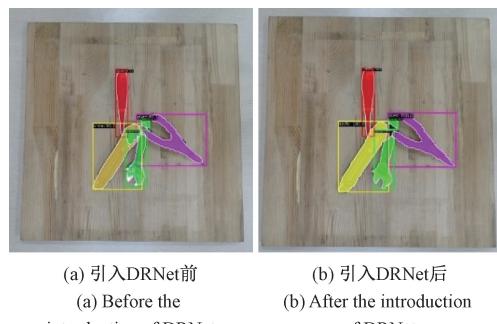


图 7 引入 DRNet 前后网络推理效果对比

Fig. 7 Comparison of network inference effects before and after the introduction of DRNet

3 结 论

为解决实例分割任务场景中的遮挡问题,本文提出了一种改进 Mask R-CNN 算法,采用双向残差网络提取特征,增强了网络的表征能力;使用 Cluster EIoU NMS 算法减少预测边界框过度抑制,在遮挡条件下能够有效减少网络漏检发生;同时使用 EIoU 损失统计边界框回归损失,有效增强网络定位能力,提升网络收敛速度。实验表明,以 Mask R-CNN 网络为基准线,本文所提算法在轻微

遮挡,一般遮挡,严重遮挡数据集下的平均 Box AP 分别提升了 5.8%、5.2% 和 7.8%,平均 Mask AP 分别提升了 4.2%、4.0% 和 2.8%;相比较于其他经典分割网络,本文模型能够有效提升遮挡物体的检测与分割精度。

然而,本文方法仍然存在一些局限性,仅考虑到当前遮挡物体的边缘进行检测与分割,并未考虑遮挡物体间的遮挡状态,因此,未来的研究方向应该进一步探讨遮挡物体间的遮挡关系,以提升模型的检测性能。

参考文献

- [1] BOLYA D, ZHOU C, XIAO F, et al. Yolact: Real-time instance segmentation [C]. Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019: 9157-9166.
- [2] WANG S, GONG Y, XING J, et al. RdsNet: A new deep architecture for reciprocal object detection and instance segmentation [C]. Proceedings of the AAAI Conference on Artificial Intelligence, 2020, 34 (7): 12208-12215.
- [3] WANG X, ZHANG R, SHEN C, et al. Solo: A simple framework for instance segmentation [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021, 44(11): 8587-8601.
- [4] HE K, GKIOXARI G, DOLLAR P, et al. Mask R-CNN [C]. Proceedings of the IEEE International Conference on Computer Vision, 2017: 2961-2969.
- [5] REN S, HE K, 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, 2016, 39(6): 1137-1149.
- [6] CAI Z, VASCONCELOS N. Cascade R-CNN: High quality object detection and instance segmentation [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2019, 43(5): 1483-1498.
- [7] CHEN K, PANG J, WANG J, et al. Hybrid task cascade for instance segmentation [C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019: 4974-4983.
- [8] 林哲,潘慧琳,陈丹.融合改进 YOLO 和语义分割的遮挡目标抓取方法 [J]. 电子测量与仪器学报, 2024, 38(12): 190-201.
- LING ZH, PAN H L, CHEN D. Grasp method for occlusion method by fusing improved YOLO with semantic segmentation [J]. Journal of Electronic Measurement and Instrumentation, 2024, 38 (12): 190-201.
- [9] ZHANG S, WEN L, BIAN X, et al. Occlusion-aware R-CNN: Detecting pedestrians in a crowd [C]. Proceedings of the European Conference on Computer Vision (ECCV), 2018: 637-653.
- [10] LIU T, LUO W, MA L, et al. Coupled network for robust pedestrian detection with gated multi-layer feature extraction and deformable occlusion handling [J]. IEEE Transactions on Image Processing, 2020, 30: 754-766.
- [11] DAI J, LI Y, HE K, et al. R-FCN: Object detection via region-based fully convolutional networks [J]. ArXiv preprint arXiv:1605.06409, 2016.
- [12] YANG S, WANG J, HU L, et al. Research on occluded object detection by improved RetinaNet [J]. Computer Engineering and Applications, 2022, 58(11): 209-214.
- [13] LUO Z K, FANG Z, ZHENG S X, et al. NMS-Loss: learning with non-maximum suppression for crowded pedestrian detection [C]. Proceedings of the 2021 International Conference on Multimedia Retrieval. Newyork: ACM, 2021: 481-485.
- [14] SUN F W, LI C Y, XIE Y Q, et al. Review of deep learning applied to occluded object detection [J]. Journal of Frontiers of Computer Science & Technology, 2022, 16(6): 1243-1259.
- [15] CHOLLET F. Xception: Deep learning with depthwise separable convolutions [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 1251-1258.
- [16] SZEGEDY C, IOFFE S, VANHOUCKE V, et al. Inception-v4, inception-resnet and the impact of residual connections on learning [C]. Proceedings of the AAAI Conference on Artificial Intelligence, 2017.
- [17] ZHENG Z, WANG P, REN D, et al. Enhancing geometric factors in model learning and inference for object detection and instance segmentation [J]. IEEE Transactions on Cybernetics, 2021, 52(8): 8574-8586.
- [18] ZHANG Y F, REN W, ZHANG Z, et al. Focal and efficient IOU loss for accurate bounding box regression [J]. Neurocomputing, 2022, 506: 146-157.
- [19] LIN T Y, DOLLAR P, GIRSHICK R, et al. Feature pyramid networks for object detection [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 2117-2125.
- [20] LONG J, SHELHAMER E, DARRELL T. Fully

- convolutional networks for semantic segmentation [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015: 3431-3440.
- [21] 赵恩玄, 何云勇, 沈宽, 等. 基于深度学习的铸件 CT 图像分割算法 [J]. 仪器仪表学报, 2023, 44(11): 176-184.
- ZHAO EN X, HE Y Y, SHEN K, et al. Casting CT image segmentation algorithm based on deep learning[J]. Chinese Journal of Scientific Instrument, 2023,44(11):176-184.
- [22] SANDLER M, HOWARD A, ZHU M, et al. Mobilenetv2: Inverted residuals and linear bottlenecks [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 4510-4520.
- [23] LIU Z, MAO H, WU C Y, et al. A convnet for the 2020s [C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022: 11976-11986.

作者简介



陈丹(通信作者), 2011 年于中国科学院沈阳自动化研究所获得博士学位, 现为福州大学副教授, 主要研究方向为机器视觉、机械臂路径规划等。

E-mail: 632151807@ qq. com

Chen Dan (Corresponding author)

received her Ph. D. degree from Shenyang Institute of Automation Chinese Academy of Sciences in 2011. Now she is an associate professor at the Fuzhou University. Her main research interests include robot vision, robotic arm path planning, etc.



令陈佩, 2017 年于西安文理学院获得学士学位, 现为福州大学硕士研究生, 主要研究方向为计算机视觉。

E-mail: 17612946906@ 163. com

Ling Chenpei received his B. Sc. degree

from Xi'an University in 2017. Now he is a M. Sc. candidate at Fuzhou University. His main research interest includes computer vision.