2024年2月
第43卷 第2期

DOI:10.19652/j.cnki.femt.2305474

基于级联网络的散乱堆叠物体分割算法*

罗雄炜　朱铮涛
(广东工业大学机电工程学院 广州 511400)

摘　要:提出了一种针对处理散乱堆叠物体的改进方法。在YOLOv5模型中采用了加权双向特征金字塔网络(BiFPN)替代路径聚合网络(PANet),结合Gfocal损失函数,使得漏检和误检问题得到有效改善,平均精度均值(mAP)mAP@0.5达到了90.1%。利用Mask R-CNN进行目标物体分割,使用轻量化的Mobilenetv3替代ResNet101主干网络以减少参数量,同时借用CFNet思想加强特征融合机制,使得分割精度提高至92.1%。通过级联改进后的YOLOv5和改进后的Mask R-CNN,算法在实时性和精确性上得到了平衡,在有效感兴趣区域(region of interest,ROI)中提取准确的物体形状信息。与单独使用实例分割算法相比,检测速度提升了1 s。实验证明所提出的算法不仅提高了推理速度,还提高了分割精度,解决了复杂堆叠场景下物体特征提取效果差且检测速度慢的问题。

关键词:目标检测;实例分割;深度学习;级联检测
中图分类号:TP391　　**文献标识码**:A　　**国家标准学科分类代码**:510.4030

Segmentation algorithm for scattered stacked objects based on cascade network

Luo Xiongwei　Zhu Zhengtao
(School of Mechanical and Electrical Engineering, Guangdong University of Technology, Guangzhou 511400, China)

Abstract: This paper proposes an improved method for handling scattered stacked objects. In the YOLOv5 model, the BiFPN feature pyramid is used to replace PANet, and combined with the Gfocal loss function, the problem of missed detection and false detection is effectively improved, and mAP@0.5 reaches 90.1%. Mask R-CNN is used for target object segmentation, the lightweight Mobilenetv3 is used to replace the ResNet101 backbone network to reduce the number of parameters, and the CFNet idea is used to strengthen the feature fusion mechanism, increasing the segmentation accuracy to 92.1%. By cascading the improved YOLOv5 and the improved Mask R-CNN, the algorithm achieves a balance between real-time performance and accuracy, and extracts accurate object shape information in the effective region of interest (ROI) area. Compared with using the instance segmentation algorithm alone, the detection speed is increased by 1 s. Experiments have shown that the algorithm proposed in this article not only improves the inference speed, but also improves the segmentation accuracy, and solves the problem of poor object feature extraction and slow detection speed in complex stacking scenes.

Keywords: object detection; instance segmentation; deep learning; cascade detection

0　引　言

近年来,智能制造技术的快速发展促进了机器人产业迅猛发展,特别是工业领域对智能化机器人需求日渐增加。采用视觉感知技术和机械臂抓取技术是实现工业生产自动化和柔性化的核心技术,尽管基于视觉引导的机械手[1]在一定程度上可以取代工人完成一些重复性和机械性工作,但在一些工业非结构化作业场景中,例如散乱堆叠场景或复杂随机场景的目标识别存在检测速度慢、精度低、鲁棒性等问题。

随着2012年深度学习浪潮的兴起,许多学者在2014年将深度学习应用于检测领域,开创了传统方法向深度学

收稿日期:2023-08-28
* **基金项目**:广东省自然科学基金(2019A1515011229)项目资助

习的转变。在2014年之前检测方法主要采用传统方法，特征提取和分类器是当时检测的重点，针对传统方法，在2021年，喻强等[2]提出对散乱堆叠的铆钉使用模板匹配方法，识别精度改善至86%，传统方法在检测特征明显、易于区别的场景中有较好的效果，但在复杂散乱场景中的识别效果不太理想，传统方法的特征提取和分类器之间的关联比较微弱，特征表达能力受限，是因为该特征是人工设计的，因此层次较浅而导致物体过分割和欠分割[3]等问题，从而无法达到真正分类的目的。而在2014年～2016年，检测技术得到了快速的发展，随着各种神经网络结构的出现，检测算法发生了巨大的变化。深度学习的目标检测方法相比于传统视觉方法，在速度和精度上都有很大的提升，能够有效的为后续机器人抓取识别提供更好的分割信息。从2013年～2014年，基于深度学习的物体检测有了飞速的发展，检测方式主要分为两类。第1种是one stage，包括Densebox、YOLO、SSD[4]和RetinaNet[5]等。第2种是two stage，包括区域卷积神经网络(region-based CNN，R-CNN)系列、基于区域的全卷积网络(region-based fully convolutional networks，RFCN)、特征金字塔网络(feature pyramid network，FPN)和MaskR-CNN等[6]。在2019年，周伟亮[7]使用SSD检测方法对随机分布工件进行识别，识别率达到90%以上。2022年，韩雪松[8]针对更快速的基于区域的卷积神经网络(faster region-based convolutional network，Faster R-CNN)提出改进以改善堆叠工件的识别，最终检测精度为86.6%。

本文针对堆叠物体识别存在的问题进行研究，提出级联YOLOv5和Mask R-CNN模型的分割算法，分别在YOLOv5模型的基础上改善有效感兴趣区域(region of interest，ROI)漏检误检的情况，同时在Mask R-CNN模型基础上改善有效ROI区域目标物体的分割，进一步通过级联网络的方法使得算法兼具YOLO的实时性和实例分割精度高的特点。

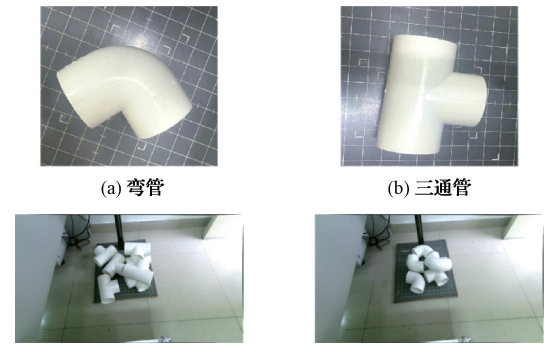
1 堆叠工件数据集制作

1.1 图像采集

本文根据6D位姿估计算法中具有代表性工件的堆叠场景，采集635张图像以构建初始数据集，其中包括三通管和弯管两种类型的工件。针对堆叠工件的散乱堆叠性，采集图像时不仅对单个工件样本采集，也同时对多目标散乱和堆叠场景进行采集，使得采集的样本更具有鲁棒性和泛化能力，如图1所示。

1.2 数据增强

图像增强技术是一种将原始图像通过一系列的处理方法增强其特征，提高图像质量和更好地展示图像细节的技术。对于目标检测任务来说，图像增强技术可以提高目标检测算法的准确性和稳定性，最终使得目标检测任务的效果更加优秀。通过图像增强技术，可以在原始图像中提取出目标对象更加明显的特征信息，如图像对比度、亮度、



(a) 弯管　(b) 三通管　(c) 堆叠三通管　(d) 堆叠弯管

图1　数据样本

锐利度等。这些信息能够帮助检测算法更加清晰地识别目标区域，并从复杂的背景中进行区分。同时，图像增强技术还有助于去除一些不必要的噪点和失真信息，从而提高目标检测算法的鲁棒性和稳定性。在实验中首先采用的是图片翻转的方式，对图片进行上下翻转和左右翻转。通过加入椒盐噪声的扰动，使得卷积神经网络(convolutional neural network，CNN)有着更加强大的噪声处理能力。加入高斯噪声，利用高斯分布中的随机矩阵，将噪声加入图片中的每一个像素RGB中，实现噪声随机扰动。最后也通过加入曝光对比度的操作，进行对图片明暗度的增强，增强后的部分效果如图2所示。将增强后的图片以8∶1∶1的比例分为训练集、验证集、测试集，选取其中8成图片作为训练集，其他为验证集和测试集，如表1所示。

(a) 亮度增强　(b) 椒盐噪声+旋转　(c) 高斯噪声　(d) 暗度增强

图2　各种数据增强效果

表1　训练集、测试集、验证集的组成

工件名称	样本总数	训练集样本数	验证集样本数	测试集样本数
三通管	2 222	1 777	222	223
弯管	2 223	1 778	223	222

2 级联网络结构设计

2.1 基于改进 YOLOv5 的 ROI 区域检测算法

随着目标检测网络结构的不断升级，YOLO 系列算法在 COCO 数据集上的性能逐渐提升，并展现出更为出色的表现，最新的模型比前几代的模型更具性能优势。然而，YOLOv5 并不是直接基于 YOLOv4 改进而来的，且不同数据集的复杂度会导致算法模型的表现出现差异。因此，最终选择在指定评价标准上表现更佳、更适用于工业应用部署的目标检测 YOLOv5 网络作为基准目标检测模型，YOLOv5 网络[9]结构如图 3 所示。

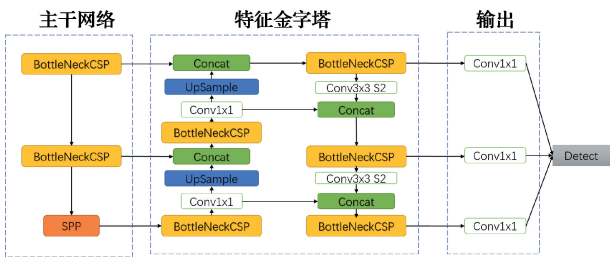


图 3 YOLOv5 网络结构

模型中的 Neck 是为了增加网络提取特征的能力，将增强特征图输入到 Head。自 2017 年 FPN 的提出，逐渐实现了多尺寸特征融合。近年来 FPN 的变体结构 PANet，NAS-FPN 等新型融合网络出现，使得多尺寸特征融合不再受制于简单的从上到下相邻尺度融合。

FPN 是一个从上到下进行融合的网络，PANet 结构[10]如图 4 所示。在 FPN 的基础上增加了从下到上的融合机制，相当于增多了一条由下到上的路线完成二次融合。研究证明简单的双向融合网络可以增强图像特征表达，随即出现图 5 所示的复杂双向融合网络 NAS-FPN，在简单双向融合的基础上中加入了跨尺度融合方式，通过搜索的方式对不同尺度的特征图进行重组，同样是增强了特征表达，分辨率不同的图片输入特征信息对于融合后的输出特征信息权重不同。

融合网络 BiFPN 如图 6 所示，是具有加权性质的双向特征融合网络，主要是在双向融合 PANet 的基础上消除了无融合效果的单一输入节点，并且添加了输入到输出的连接，并且在同尺度中采用跳跃连接的方法融合。

图 6 中蓝色虚线框内是用来重复多次的模块，BiFPN[11]能够对不同尺度的特征图更好地融合，通过重复叠加 FPN 中有效的 block 模块，增加不同尺度目标的敏感度以提升特征融合的性能，达到提高模型检测性能的目的。因此为了改善散乱堆叠物体特征信息之间关联性弱的问题，将采用 BiFPN 作为模型的 Neck 替代原来的 PANet。

对于 one-stage 检测器，通常将目标检测表述分为分类、检测框回归、检测框置信度，如图 7 所示。

现有的回归表示通常采用图 8 所示的 Dirac delta 的

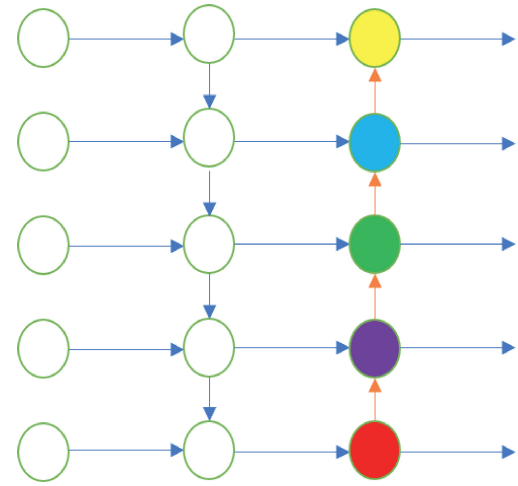


图 4 简单双向融合网络 PANet

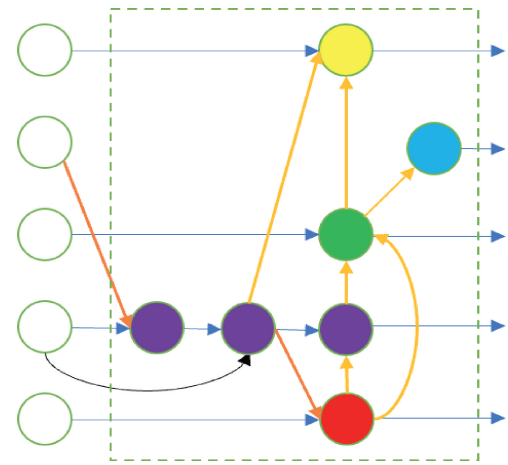


图 5 复杂双向融合网络 NAS-FPN

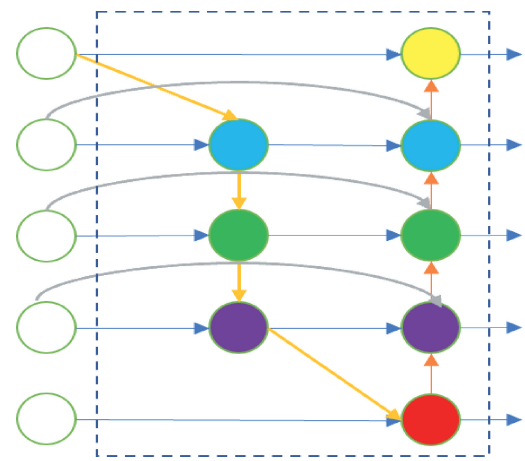


图 6 融合网络 BiFPN

分布方式，这种分布方式比较单一，当图像中出现模糊以及界定不清晰边界时表现不够灵活，而该问题在散乱堆叠场景中经常出现。Dirac delta 表示框分布形式如式(1)

所示。

$$y = \int_{-\infty}^{+\infty} \delta(x - y)x\,\mathrm{d}x \tag{1}$$

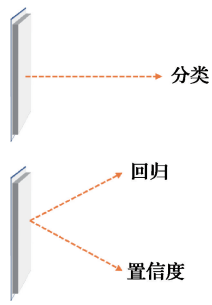


图 7　3 个常用表示

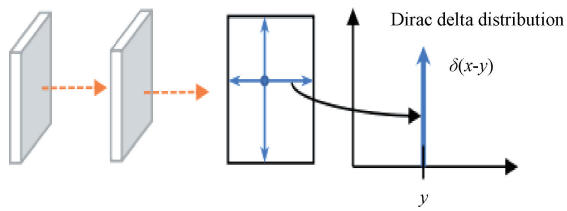


图 8　单一回归分支

图 9 所示的 Gfocal[12] 回归分支将坐标的预测从原来的预测单一数值，变成预测一个区间内的概率分布后求期望。这里为了和 CNN 一致，采用连续积分变离散表达如式(2)所示，使用 softmax 函数来实现，并且该任意分布方式建模回归框是可以通过积分形式引入到任意当前存在的回归框损失函数中。

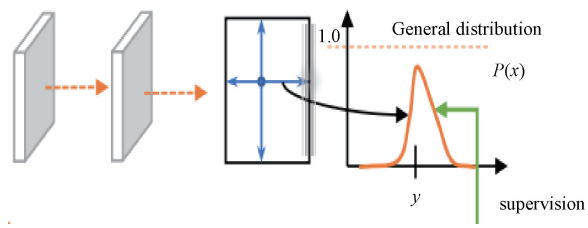


图 9　任意回归分支

$$\hat{y} = \sum_{i=0}^{n} P(y_i)x_i \tag{2}$$

由于真实分布的标注位置不会离得非常远，过于任意也会导致网络学习效率的下降，为了能够快速找到标注位置附近的数据，并使得概率尽可能地大，引入 Distribution Focal Loss 去优化标签附近一左一右两个位置的概率，使得网络能够高效聚焦到目标位置相邻区域的分布中去，公式如式(3)所示，S_i 相当于式(2)的 $P(y_i)$。综上所述，在做数据集标注的时候可能存在坐标框标注不够精确的问题，并且散乱堆叠场景中物体边界信息比较模糊问题很常见，为了改善散乱物体识别的误检或者漏检，将此损失函数加入检测任务。

$$D(S_i, S_{i+1}) = -((y_{i+1} - y)\log(S_i) + (y - y_i)\log(S_i + 1)) \tag{3}$$

由于数据标注存在误差且存在堆叠情况下局部特征识别效果不佳的问题，通过加入数据增强，且提出改进特征融合网络和改进坐标预测的策略。图 10 所示是将原始 YOLOv5 的 Neck 由 PANet 替换为特征融合网络 BiFPN，且加入改进坐标预测的 Gfocal 损失函数的改进 YOLOv5 后的模型示意图。

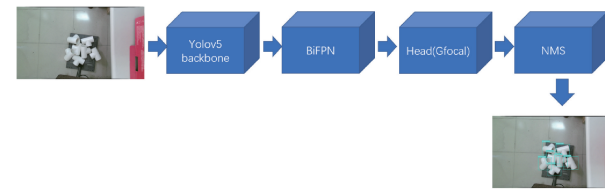


图 10　模型示意图

2.2　基于改进 Mask R-CNN 有效 ROI 区域分割

Mask R-CNN 是一种同时进行目标检测和实例分割的网络，其效果非常出色，算法流程如图 11 所示。在 Faster R-CNN 的基础上，增加了一个用于语义分割的分支，从而实现了实例分割。Mask R-CNN[13] 是一个两阶段的网络。第 1 阶段是生成候选区域(region proposal network，RPN)，即候选目标对象的边界框。第 2 阶段本质上是 Fast R-CNN，使用 ROIPool 从每个候选框中提取图像特征并进行分类、边界框回归等操作，同时为每个 ROI 输出一个并行的二值化掩码。Mask R-CNN 具有多种结构，包括不同的卷积主干结构和网络头。不同的主干结构主要有 ResNet50 和 ResNet101，通过第 4 阶段的最后一个卷积层进行特征提取，即 C4 层。除此之外，Mask R-CNN 还有一个独特的主干特征提取网络，即 FPN，使用自上而下的横向连接结构，从单一尺度输入图像并构建出特征金字塔。这种方法类似于 ResNet，可以从金字塔的不同层级上提取 ROI，从而提高网络的精度和速度。

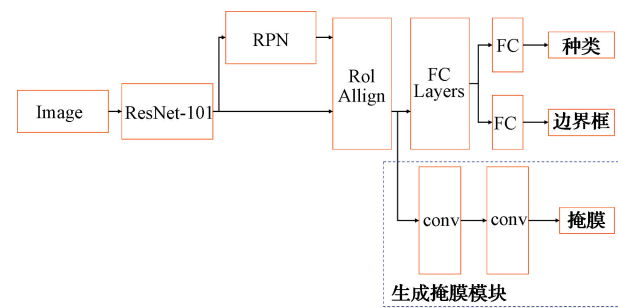


图 11　Mask R-CNN 示意图

Mask R-CNN 主干网络主要采用了 ResNet 残差网络，是被指定的构建特征图的 backbone，一般常用的有 ResNet-50，ResNet-101，以及加入了主流注意力模块的 ResNeXt，虽然这些主干网络在提取特征时有着不错的效

果，但参数量较大，使得网络检测速度较为缓慢，因此采用轻量化的主干网络替代原有的残差网络主干网络。

深度可分离卷积示意图如图 12 所示，针对所得到的特征矩阵每一个 channel，给定一个 channel 为 1 的卷积核，每一个卷积核只负责一个 channel，所得到的特征图也是通道数为 1 的，相对于传统的卷积，在准确率小幅度下降的前提下能够大幅度地减少模型的参数量和运算量。后续 1×1 的卷积层起到一个降维的作用，同样这里也有捷径分支 shortcut 连接，使输入特征矩阵和输出特征矩阵在相同纬度上进行直接相加。

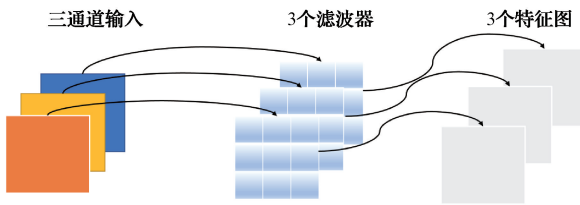


图 12　深度可分离卷积

Mobilenetv3 Block 结构[14] 如图 13 所示，加入了 SE 注意力机制的模块，针对特征矩阵的每一个 channel 进行池化处理，然后通过两个全连接层得到一个输出向量，并且更新了一个非线性激活函数 swish 等，对于耗时层的结构，采取了减少第 1 个卷积层个数的操作由 32 个减到 16 个，并且简化了 Last Stage。Mobilenetv3 不仅能提高对图片的处理速度，也能够更好的拟合网络，因此采用 Mobilenetv3 轻量化网络替代原始 Mask R-CNN 的主干网络。

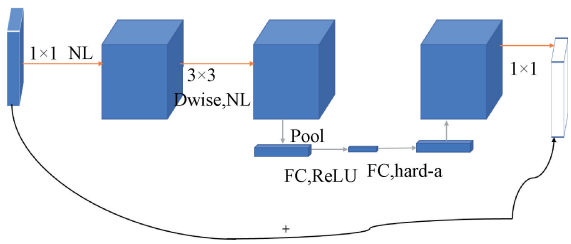


图 13　Mobilenetv3 Block 模块

主干网络轻量化后对 Mask R-CNN 的 FPN 进行对应的改进，FPN 是通过构建特征金字塔多级尺度来解决计算机视觉中的问题。金字塔由深度神经网络中的特征图层组成，这些特征图层在每个金字塔级别提供不同的图像尺度，在网络中逐层级联这些特征图，以促进多尺度[15] 的目标检测。FPN 示意图如图 14 所示。

对现今的 SOTA 方法在处理复杂场景下的散乱遮挡堆叠物体时，多尺度特征的密集预测是必不可少的。无论是实例分割还是目标检测，都需要多尺度特征。为了提取不同尺度的特征，常常使用主干网络结合轻量级模块进行特征融合。然而，这种方法可能无法充分融合多尺度特征，因为与主干网络相比，用于特征融合的参数非常有限。通常使用的 FPN 通过逐元素相加和 3×3 卷积的变换来

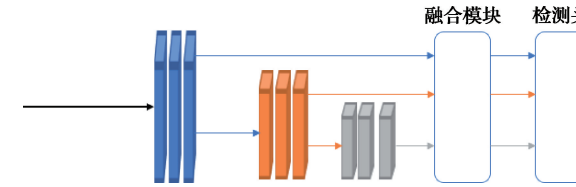


图 14　FPN 及其变体

实现多尺度特征的融合，但可能会引入更多的参数，从而减少主干网络的参数数量。为了解决这个问题，提出了一种称为级联融合网络 CFNet[16] 的架构，如图 15 所示，能够在相同算力情况下获得更多特征融合。CFNet 将特征集成操作嵌入到主干网络中，使更多参数参与特征融合，极大地增强了特征融合机制，从而提高了对遮挡堆叠物体特征的提取和物体之间特征的关联。

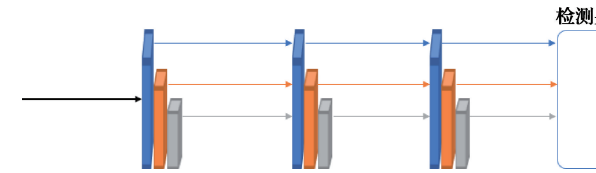


图 15　CFNet 模块

以上是针对传统视觉对于堆叠物体容易发生欠分割或者过分割以及检测速度慢的问题，提出对 Mask R-CNN 的主干网络进行轻量化，并且利用特征融合机制，使得更多的特征参数进行融合。

3　实验分析

3.1　实验环境与数据集

本实验所用计算机详细配置如表 2 所示，实验数据集采用自制的堆叠弯管和三通管检测识别对象。

表 2　机器参数及编程环境指标

	服务器参数
CPU	Inter i9-9900 K
内存	32 G C16 3 000 MHz
硬盘	Samsung pm981a 512 G
显卡	Nvidia RTX 2080TI 11 G
主板	ASUS B460-PLUS
系统	Ubuntu18.04(64 位)
编程语言	Python 3.7
编译界面	Pycharm community
深度学习框架	Pytorch 1.7.1

3.2　实验数据评价指标

对于检测模型的评价指标在本文使用精确度(precision,P)和召回率(recall,R)，评价一个目标有没有被正确检测出来，主要是预测的目标和真实物体间的交并比

(IOU)是否大于提前设定的阈值并且预测的物体是否与真实物体类别一样。精确度表示检测到的目标中正确检测的比例,召回率是表示所有实际目标中被正确检测到的比例,公式如式(4)、(5)所示。为了综合考虑精确度和召回率,需要计算平均精度(AP)。AP 通过在不同召回率水平上对精确率求平均得到,实验中,选取 $IOU=0.5$ 阈值下计算 AP。

$$P=\frac{TP}{TP+FP} \tag{4}$$

$$R=\frac{TP}{TP+FN} \tag{5}$$

式中:TP 是正确检测到的目标数量;FP 表示错误检测到的目标数量;FN 表示未被检测到的实际目标数量。

同样对于分割模型的评价指标采用交并比计算公式作为评价标准。交并比是通过将像素的真实值与预测值的交集与它们的并集进行比较,可以得到一个值,该值表示像素的置信度:

$$IOU=\frac{S_{预}\cap S_{真}}{S_{预}\cup S_{真}} \tag{6}$$

式中:$S_{预}$ 表示预测实例的面积;$S_{真}$ 该面积代表了真实的实例分割区域。

3.3 目标检测实验结果与分析

对原始 YOLOv5 模型进行消融实验,结果如表 3 所示,可以看出,通过加入 BiFPN 结构使得准确率和召回率以及 mAP 都有一定程度的提升,因此可以推出 BiFPN 这种基于双向流的结构可以将不同尺度的特征图结合在一起,可以提升该场景下的有效 ROI 区域检测的准确率;加入 Gfocal 的损失函数,同样对 mAP 等评价标准有提高;将这两个改进加入原始模型,最终得到 mAP@0.5 增长了 5%,达到 90.1%,准确率和回召同样也有一定幅度的提升,分别是 89.4%和 88.5%。

表 3 改进 YOLOv5 模型对比实验

模型方案	P	R	mAP@0.5
YOLOv5	0.843	0.854	0.851
YOLOv5+BiFPN	0.873	0.866	0.869
YOLOv5+Gfocal	0.881	0.893	0.891
本文	0.894	0.885	0.901

通过选取散乱堆叠情况下容易发生误检的图片,采用原始 YOLOv5 基础网络与改进后的 YOLOv5 网络进行检测,检测效果如图 16 和 17 所示。通过算法模型可视化结果可以更加直观的对比算法改进带来的识别效果,从图 16 可以清晰地看到,改进后的 YOLOv5 算法训练的网络能够大大改变了对于一些边缘物体信息误检的情况,同时,如图 17 所示,由箭头指向的物体可以看出,改进的 YOLOv5 算法有效地解决了在存在较严重堆叠的情况下,无法正确识别有效区域 ROI 的问题,使得可以识别到堆叠在下面的物体,在一定程度上改善了对于堆叠有效 ROI 区域漏检的情况。以上实验证明本文提出的算法在对散乱堆叠场景下有效区域 ROI 检测精度有一定的提升。

(a) 原始YOLOv5算法检测效果 (b) 改进后的YOLOv5算法检测效果

图 16 散乱堆叠目标物体的误检情况减少

(a) 原始YOLOv5算法检测效果 (b) 改进后的YOLOv5算法检测效果

图 17 散乱堆叠目标物体的漏检情况减少

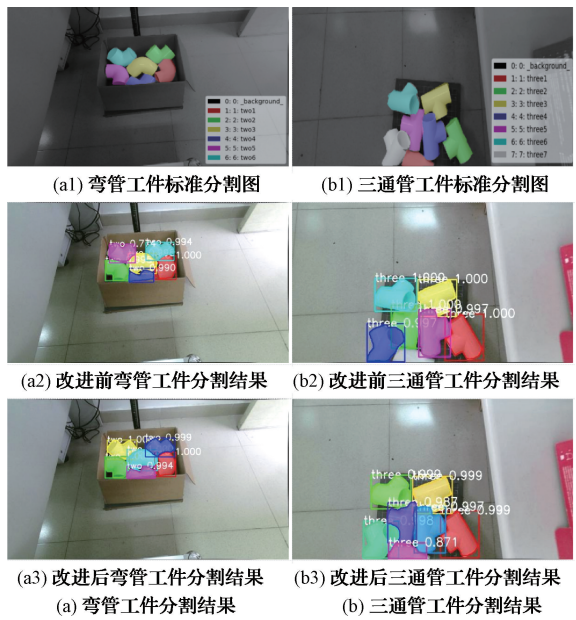
3.4 实例分割实验结果与分析

改进 Mask R-CNN 模型对比试验结果如表 4 所示,可以看出,通过对将主干网络用轻量级的 Mobilenetv3 进行替换,大大减少了模型的大小,在 mAP 有小幅度的提升情况下,显著加快了检测单张图片的速度,在保留 ResNet101 主干网络,借用 CFNet 思想重构特征金字塔结构,对于原始的 Mask R-CNN 的 mAP 提升了 5%。这说明改进特征融合机制使得 Mask R-CNN 可以更好地识别散乱堆叠的物体。最终是将两种模型进行消融实验,通过对比发现,最终 mAP 提升了 9%且 mIoU 提升了接近 5%。说明改进 Mask R-CNN 在对目标的上下文信息,获得了更多不同层次的目标特征,并且通过加入轻量级主干网络,在检测精度不下降的情况下,显著加快了检测速度,因此可以得出该改进在检测散乱堆叠场景下可以取得更好结果。

图 18(a)所示是对弯管工件的识别,图 18(b)所示是对三通管工件的识别,可以看出经过对模型特征金字塔和主干网络的改进,改善了对于严重遮挡堆叠物体漏检的情况,对于单个物体的识别准确率有小幅度的提升,并且没有出现欠分割和过分割的情况,不会对背景等进行误分割。综上所述,对于上述两种工件物体分割实验中,本文提出的改进方法提升了分割精度。

表 4　改进 Mask R-CNN 模型对比实验

模型	模型大小/MB	检测速度/fps	*P*	*R*	mAP	mIoU
MaskR-CNN(ResNet101+FPN)	255.9	2.5	0.854	0.841	0.846	0.654
Mask-R-CNN(MobileNetv3+FPN)	94.8	0.5	0.864	0.842	0.858	0.663
MaskR-CNN(ResNet101+CFNet)	230.55	2.7	0.895	0.882	0.889	0.687
本文改进算法	150.55	1.4	0.924	0.917	0.921	0.715



(a1) 弯管工件标准分割图　(b1) 三通管工件标准分割图
(a2) 改进前弯管工件分割结果　(b2) 改进前三通管工件分割结果
(a3) 改进后弯管工件分割结果　(b3) 改进后三通管工件分割结果
(a) 弯管工件分割结果　(b) 三通管工件分割结果

图 18　堆叠工件分割效果

3.5　级联网络实验结果

通过对比改进后的 YOLOv5 和改进后的 Mask R-CNN 对散乱堆叠物体识别的效果，对于 YOLOv5 来说它的模型更小，几乎可以实时检测[17]，经过改进后的 YOLOv5 可以很好识别目标物体的有效区域 ROI，但 YOLOv5 是通过坐标矩形框作为定位，在散乱堆叠场景下容易识别到其他物体的边缘信息，这对于后续点云信息的转化存在新的噪声。相对于检测模型，Mask R-CNN 模型参数量比较大，是一个更耗费算力和时间的一个模型，经过改进后的 Mask R-CNN 尽管减少了模型大小也提高了图片检测速度，速度提升的前提下也提高了分割精度，对堆叠物体的识别不会产生过分割和欠分割的情况，相较于检测模型来说，改进后的 Mask R-CNN 对于堆叠情况识别精度更高并且不会识别到粘连堆叠的边缘信息[18]，这对于后续完成位姿估计是更加鲁棒的算法。

改进的 YOLOv5 算法对 960×540 分辨率的图片预测能实现 40 fps 的检测效率，改进后的 Mask R-CNN 算法对单张图片检测速度需要 1.4 s。由于实例分割是获取每一个目标物体得像素位置也就是 Mask，Mask 可以提供精确的目标形状，而目标检测是获得矩形框 ROI 位置信息，目标物体的形状会受到堆叠情况下其他物体的噪声影响，所以在检测精度上 Mask R-CNN 更胜一筹。针对该问题，提出将 YOLOv5 算法和 Mask R-CNN 算法模型进行级联检测[19]，综合两个模型各自的优势，首先使用 YOLOv5 目标检测对全图进行有效 ROI 区域的检测，充分利用检测模型实时性的特点，得到有效 ROI 区域后使用 Mask R-CNN 对检测到的每一个区域 ROI 再进行实例分割获取目标形状的 Mask 信息[20]，级联模型结构如图 19 所示。

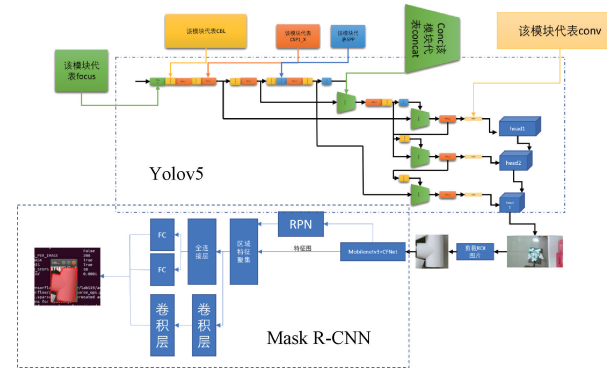


图 19　级联检测模型

级联检测效果如图 20 所示。该方法在不损失精度的情况下比单独使用 Mask R-CNN 算法检测速度提升了 1 s，并且在分割精度上也有着小幅度的提升，结果证明该算法很好的级联了 YOLO 算法的实时性和 Mask R-CNN 算法分割精度高的特点，有效改善了目标检测漏检误检、过分割以及欠分割的问题，显著提高了散乱堆叠物体的识别效果。

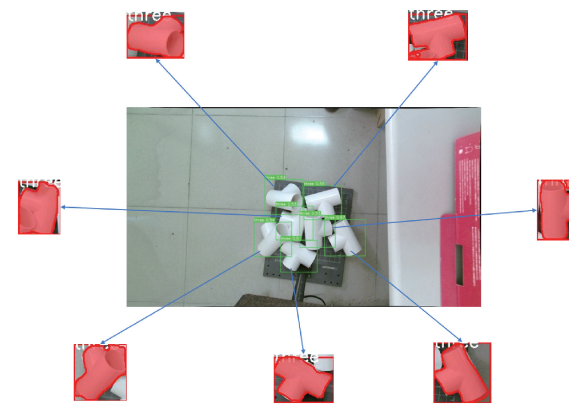


图 20　级联检测效果

4 结 论

针对机器人无序抓取场景中散乱堆叠物体的识别问题进行算法研究和实验分析，采用 KinectV2 深度相机采集常见的两种工件类型并制作数据集完成人工标注，同时采用离线数据增强，对该数据集进行算法研究。解决传统方法对于复杂场景识别精度效果不佳、效率低等问题，提出改进 YOLOv5 算法和改进 Mask R-CNN 算法，通过两阶段算法完成目标物体 ROI 的提取。该算法可配合物体定位方法实现机器人无序抓取场景，具有一定的实用价值。

参 考 文 献

[1] 葛俊彦，史金龙，周志强，等. 基于三维检测网络的机器人抓取方法[J]. 仪器仪表学报，2021，41(8)：146-153.

[2] 喻强，庄志炜，田威，等. 基于单目视觉的散堆铆钉识别与定位技术[J/OL]. 计算机集成制造系统：1-20. http://kns. cnki. net/kcms/detail/11. 5946. TP. 20211029. 1927. 016. html.

[3] 王瑞丰，朱铮涛，冯端奇. 基于改进 LCCP 的堆叠物体分割算法[J]. 电子测量技术，2022，45(3)：118-124.

[4] LIU W，ANGUELOV D，ERHAN D，et al. SSD：Single shot multibox detector [C]. European Conference on Computer Vision. Cham：Springer，2016：21-37.

[5] LIN T Y，GOYAL P，GIRSHICK R，et al. Focal loss for dense object detection[C]. Proceedings of the IEEE International Conference on Computer Vision，2017：2980-2988.

[6] ZOU Z，CHEN K，SHI Z，et al. Object detection in 20 years：A survey[C]. Proceedings of the IEEE，2023.

[7] 周伟亮. 基于机器视觉的随机分布工件自动排序研究[D]. 广州：华南农业大学，2019.

[8] 韩雪松. 基于深度学习的堆叠工件识别与定位系统的设计与实现[D]. 沈阳：中国科学院大学(中国科学院沈阳计算技术研究所)，2022.

[9] ZHU X，LYU S，WANG X，et al. TPH-YOLOv5：Improved YOLOv5 based on transformer prediction head for object detection on drone-captured scenarios[C]. Proceedings of the IEEE/CVF International Conference on Computer Vision，2021：2778-2788.

[10] LIU S，QI L，QIN H，et al. Path aggregation network for instance segmentation [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition，2018：8759-8768.

[11] TAN M，PANG R，LE Q V. Efficientdet：Scalable and efficient object detection [C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition，2020：10781-10790.

[12] LI X，WANG W，WU L，et al. Generalized focal loss：Learning qualified and distributed bounding boxes for dense object detection [J]. Advances in Neural Information Processing Systems，2020，33：21002-21012.

[13] 王陶然，王明泉，张俊生，等. 基于 Mask R-CNN 的轮毂缺陷分割技术[J]. 国外电子测量技术，2021，40(2)：1-5.

[14] HOWARD A，SANDLER M，CHU G，et al. Searching for mobilenetv3[C]. Proceedings of the IEEE/CVF International Conference on Computer Vision，2019：1314-1324.

[15] 徐晓光，李海. 多尺度特征在 YOLO 算法中的应用研究[J]. 电子测量与仪器学报，2021，35(6)：96-101.

[16] ZHANG G，LI Z，LI J，et al. CFNet：Cascade fusion network for dense prediction[J]. Computer Science，2023，DOI：10. 48550/arXiv. 2302. 06052.

[17] VASU P K A，GABRIEL J，ZHU J，et al. An improved one millisecond mobile backbone [J]. Computer Science，2022，DOI：10. 48550/arXiv. 2206. 04040.

[18] YANG J，WU S，GOU L，et al. SCD：A stacked carton dataset for detection and segmentation[J]. Sensors，2022，22(10)：3617.

[19] QI X，DONG J，LAN Y，et al. Method for identifying litchi picking position based on YOLOv5 and PSPNet[J]. Remote Sensing，2022，14(9)：2004.

[20] XI D，QIN Y，WANG S. YDRSNet：An integrated Yolov5-Deeplabv3＋real-time segmentation network for gear pitting measurement[J]. Journal of Intelligent Manufacturing，2021，34(12)：1-15.

作 者 简 介

罗雄炜，硕士研究生，主要研究方向为计算机视觉和深度学习。

E-mail：531581825@qq. com

朱铮涛(通信作者)，副教授，主要研究方向为机器视觉应用与自动化技术。

E-mail：gzzzt@gdut. edu. cn