DOI: 10.13382/j.jemi.B2407178

基于稀疏可学习 proposal 的车间工具目标检测*

刘珍兵 孙巧榆 王述文 夏嘉伟
(江苏海洋大学电子工程学院 连云港 222005)

摘 要:针对车间工具不同型号之间尺寸存在较大差异、形状种类繁多等问题,提出了一种基于稀疏可学习 proposal 的车间工具检测算法。首先,融入稀疏表示和可学习的 proposal 机制来提升模型的鲁棒性,并减少检测过程中所需的参数量;其次,引入 Swin-Transformer 结构,旨在增强模型的全局以及细节学习能力,有效地解决传统卷积神经网络在高层语义信息融合方面存在的不足;然后,使用一种改进的多尺度特征融合网络架构,通过有效融合不同尺度的特征,提高了模型对于各种尺度目标的检测能力;最后,将多头注意力和动态卷积结合,在不同特征层之间建立更精确且细致的联系,从而进一步提升了目标检测的准确性;采用了 CIoU 损失函数,通过综合考虑位置、尺度和形状信息,使得模型对边界框的回归预测更加全面与准确。实验结果显示,本文算法在车间工具目标检测任务上的平均检测精度达到了 91%,较当前主流算法至少提升了 2.3%以上。同时,单张图片的检测速度大约为 53 ms,满足了实时检测的需求,体现了综合性能优越。

关键词: 工具检测;稀疏可学习;多尺度特征;Swin-Transformer;多头注意力
中图分类号: TP391;TN9 **文献标识码**: A **国家标准学科分类代码**: 520.6040

Target detection of workshop tools based on sparse learnable proposal

Liu Zhenbing Sun Qiaoyu Wang Shuwen Xia Jiawei
(School of Electronic Engineering, Jiangsu Ocean University, Lianyungang 222005, China)

Abstract: Aiming at the significant size discrepancies and various shapes among different models of workshop tools, a workshop tool detection method based on sparse learnable proposal is proposed. Firstly, sparse representation and learnable proposal mechanism are integrated to improve the robustness of the model and reduce the required parameters in the detection process. Secondly, Swin-Transformer structure is introduced to enhance the global and detail learning ability of the model, which can effectively overcome the shortcomings of traditional convolution neural network in high-level semantic information fusion. Thirdly, an improved multi-scale feature fusion network architecture is used to improve the detection ability of the model for various scale targets according to effective fusion of different scale features. Finally, multi-head attention and dynamic convolution are combined to establish a more precise and detailed connection between different feature layers, thereby furtherly improving the accuracy of target detection. The CIoU loss function is applied to make the regression prediction of the boundary box more comprehensive and accurate by considering the location, scale and shape information. The experimental results show that the average detection accuracy of the proposed method for workshop tool detection reaches 91%, which is at least 2.3% higher than the current mainstream methods. At the same time, the detection speed of a single picture is about 53 ms, which meets the needs of real-time detection and reflects the excellent comprehensive performance.

Keywords: tool detection; sparse and learnable; multi-scale features; Swin-Transformer; multi-head attention

收稿日期: 2024-01-04 Received Date: 2024-01-04
*基金项目:国家自然科学基金(62271236)项目资助

0 引 言

在现代工业生产中,实施车间工具目标检测任务具有重要意义。运用人工智能技术,及时检测并记录工人携带的各类工具的进出状态,实现对工具使用情况的精确管理和监督至关重要[1]。这样不仅能够有效防止工具的遗漏,还能够提高工具的利用率,有助于减少工人在管理和准备工具上所花费的时间,进而提升整体工作效能;另一方面,对车间工具进行精确检测管理可以减少因为工具被遗忘在车间内而导致的潜在安全风险,如果工具被错误地遗留在运行中的机器上,可能导致机械故障或更加严重的事故。

随着深度学习技术的快速发展,基于深度学习的目标检测方法在不同领域取得了显著的进步[2]。文献[3]设计了一个空间感知模块,旨在实现双通道输入的空间特征融合,采用 ESNet(Enhanced ShuffleNet)重构主干网络,并且基于深度可分离卷积实现多尺度特征提取,从而有效提高了检测速度。文献[4]基于 RetinaNet,构建了一个特征融合增强模块,该模块能够实现对不同尺度铁路工机具的高精度检测。在文献[5]中,根据多尺度目标参数,采用了最佳锚框比来提高检测的召回率和精度。文献[6]以 Faster R-CNN 为基础,通过引入双特征融合算子,对提取的特征进行标记,然后再次输入模型进行特征提取,并融合了具有高重合度的特征,以显著提升检测效果。在文献[7]中,将 YOLOv5 中的加权非极大值抑制方式改进为 DIoU_NMS,相较于原算法取得了明显的性能提升。

本文针对车间工具目标检测问题,提出了一种基于稀疏可学习 proposal 的方法。避免了传统方法中密集 anchor 和大量候选框(proposal)的生成,通过学习目标的稀疏特性,从固定数量的 proposal 中筛选最可能包含目标的框。以深度学习目标检测网络 Sparse R-CNN 为基础,但在网络结构上进行了关键性的改进。首先,将原始的 ResNet 主干网络替换为 Swin-Transformer,这一变化有效地缓解了传统卷积操作的局限性,使得网络具备了更强的特征表达能力和更好的可扩展性。针对多尺度特征融合问题,通过不断优化网络结构,引入了双线性插值上采样策略和独特的特征融合机制,以增强模型对多尺度语义信息的捕获和整合能力,使得模型能够更全面地理解和处理各种尺度的目标,从而提高检测的准确性和鲁棒性。在检测头部分,通过在动态交互之前添加多头自注意力计算模块,将多头注意力和动态卷积结合,有助于模型推理不同目标之间的关系,进一步提升了模型的性能。

1 基于稀疏可学习 proposal 的车间工具目标检测

1.1 整体网络结构

稀疏可学习 proposal 算法的关键思想是将来自提取候选框网络[8](region proposal network,RPN)的数十万个 proposal 替换为一小组 proposal。该算法是一个简单、统一的网络,由 1 个主干网络、1 个特征融合层、1 个检测头和 2 个特定任务的预测层组成。总共 3 个输入,分别是图像、1 组 proposal boxes 和 1 组 proposal features,其中 proposal boxes 和 proposal features 是可学习的,可以与网络中的其他参数一起进行优化。整体框架如图 1 所示。

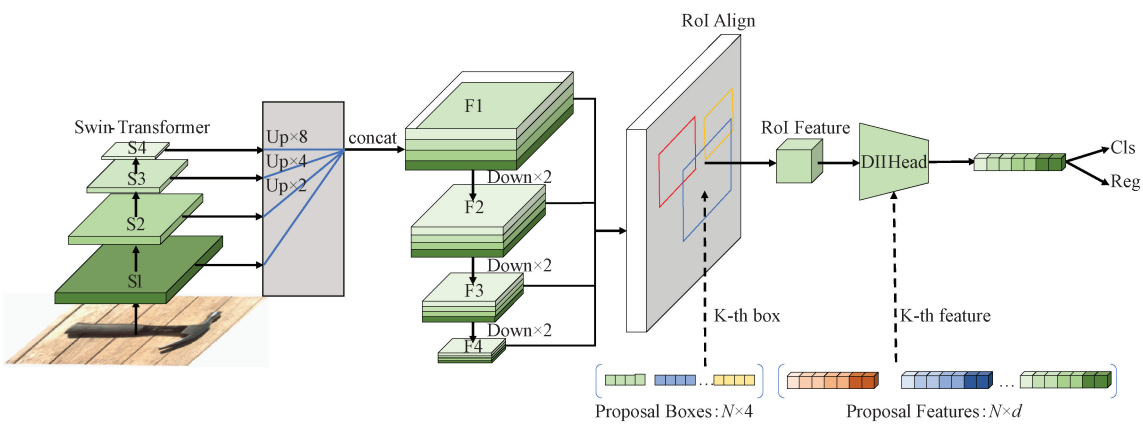


图 1 整体框架图

Fig. 1 Overall framework diagram

1.2　Swin-Transformer 结构

本文采用 Swin-Transformer[9] 作为主干网络，用于图像特征提取。Swin-Transformer 的整体结构层次分明，自底向上的下采样[10]倍数呈 2 倍增加，能够有效提取层次化的特征。相较于纯卷积结构，Swin Transformer 更关注图像中的感受野[11]，在多目标动态特征的目标检测任务中表现卓越。首先，输入图像经过 Patch Partition 层[12]处理，被划分为相同大小的块，每个块由 4×4 的 patch 组成。在嵌入向量后，进行通道方向上的展开平铺。通常采用 RGB 三通道图像，每个块包含 16 个通道，因此展开平铺后通道数变为 48。通过 Patch Partition 后，图像的宽高减少为原来的 1/4，而深度增加为原来的 16 倍，具体结构如图 2 所示。

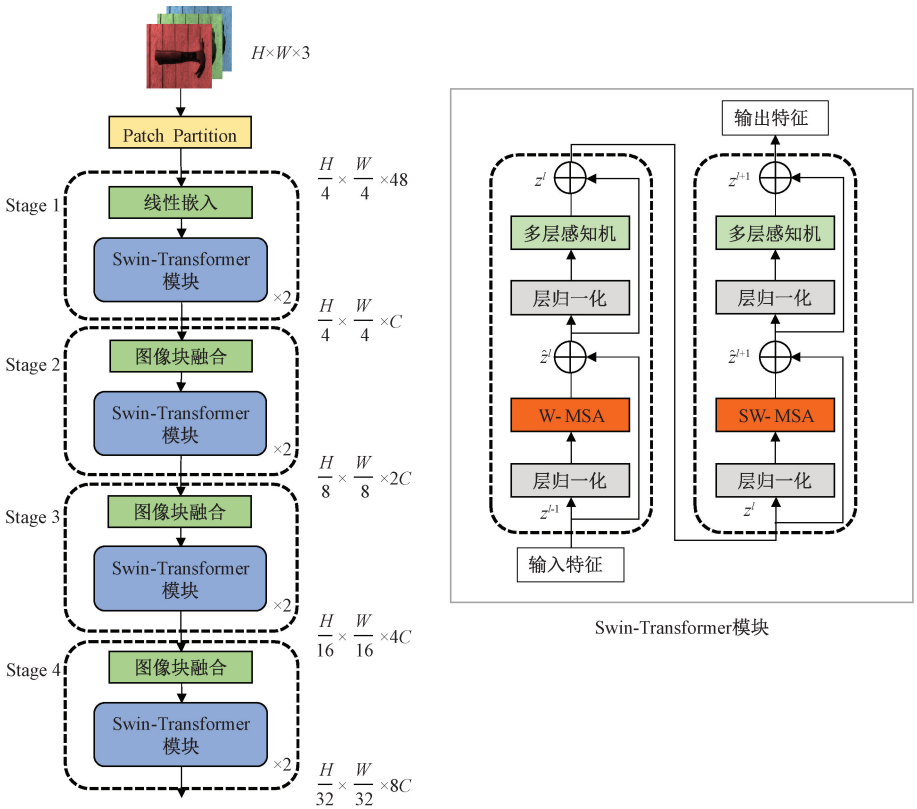


图 2　Swin-Transformer 结构

Fig. 2　Structure of Swin-Transformer

接着，经过 4 个阶段的融合，在第 1 个阶段，线性嵌入层对输入图像的每个像素通道做深度变换，将深度从 48 增加到 C，然后将其传递到 Swin-Transformer 模块。剩下的 3 个阶段都是重复经过图像块融合下采样和 Swin-Transformer 模块计算，这一操作导致特征图的高度和宽度减半，深度增加为原来的 4 倍。最后，经过一个全连接层[13]，特征图的高度和宽度保持不变，而深度翻倍。每个阶段中的 Swin-Transformer 模块都被设计为偶数次，其中内部首先经过一个窗口多头自注意力[14]结构（windows multi-head self-attention，W-MSA），然后经过一个滑动窗口多头自注意力结构（shifted windows multi-head self-attention，SW-MSA）。最终，通过多层感知机[15]层，实际相当于一个全连接层和激活函数。

1.3　改进的多尺度特征融合网络

鉴于车间工具的尺寸和比例可能存在较大差异，本文提出了一种改进的多尺度特征融合网络（multi-scale feature fusion network，MFFN）。该网络结合了特征提取技术 Swin-Transformer 和特征金字塔[16]（feature pyramid networks，FPN）。Swin-Transformer 生成 4 个不同尺度的特征图，分别命名为 S1～S4，而 FPN 由 F1～F4 这 4 个尺度递减的特征层构成，整体网络结构如图 3 所示。

MFFN 首先对 Swin-Transformer 输出的不同尺度的特征图进行处理，将它们通过上采样操作至统一的大小，这一上采样过程首先通过应用一个 1×1 的卷积，然后进行批次归一化，最后再进行双线性上采样[17]操作。接下来这些经过上采样得到的特征图被连接起来，并经过 ReLU 激活函数和 1×1 的卷积处理，随后被送入特征金字塔网

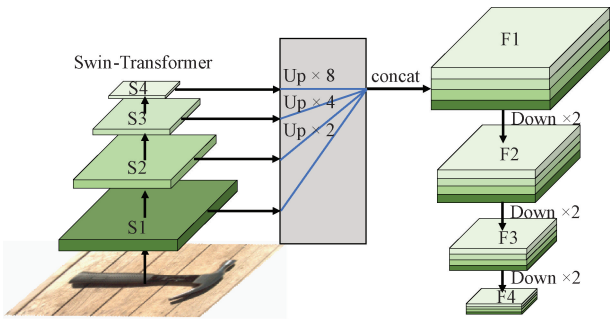


图 3 改进的多尺度特征融合网络

Fig. 3 Improved multi-scale feature fusion structure

络。在特征金字塔网络中,这些特征图经过连续的下采样操作,从而提取出更高层次的特征信息。这样,MFFN 能够有效地融合不同尺度的特征,提升模型对目标的检测能力,并在复杂场景下取得更好的性能,具体的特征融合过程如图 4 所示。

这项改进的特征融合策略旨在显著提升模型对多尺度信息的整合能力。通过将不同尺度的特征进行连接,能够更加全面地捕获和利用来自各个尺度的信息[18]。这样模型将能够更准确地理解车间工具目标中不同尺度的特征和结构,从而提升网络的整体性能,使其在复杂场景下更具鲁棒性和泛化能力。

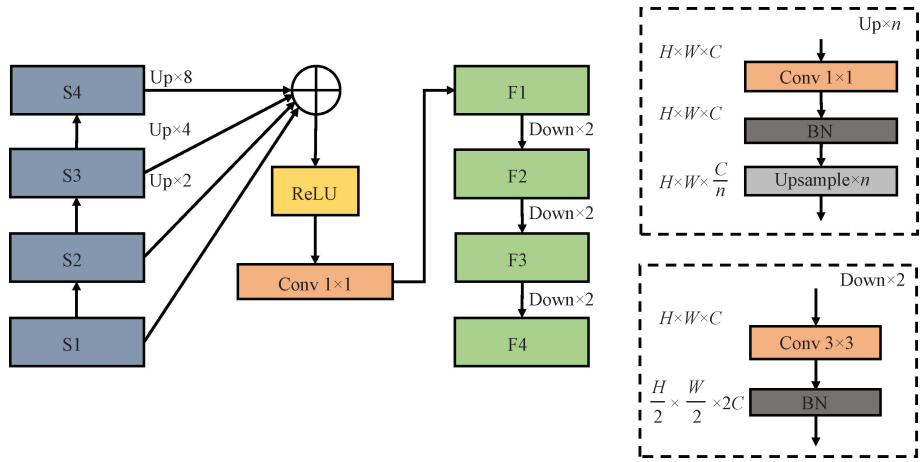


图 4 多尺度特征融合过程

Fig. 4 Multi-scale feature fusion process

1.4 稀疏可学习 proposal

与通常使用的提取候选框网络的生成不同,本文采用了一组固定的可学习 proposal 的方法,通过学习一组具有四维参数的 proposal boxes 来实现。这四维参数包括标准化的中心坐标、高度和宽度,取值范围在 0~1。这些可学习 proposal boxes 的参数在训练过程中通过反向传播算法进行更新,通过不断更新迭代适应目标检测任务的要求。

这些学习到的 proposal box 可以被视为训练集中潜在目标位置的统计数据,也就是对图像中目标区域最有可能的初始猜测。这种方法不仅提供了更灵活的 proposal 生成机制,同时也允许模型通过训练过程中的参数更新逐渐适应不同数据集的特征和目标分布,从而提高模型的泛化性能。

为了弥补可学习 proposal box 在描述物体的姿态和形状方面的不足,本文引入了一种特征层面的 proposal,即 proposal feature。proposal feature 也是一组可学习的参数,总共有 $N \times d$ 个参数,其中 N 代表 proposal feature 的个数,与 proposal boxes 对应,d 代表特征的维度。proposal feature 通过一对一的交互与从提取出的感兴趣区域(region of interest,RoI)特征进行交互,这种交互机制能够使得 RoI 特征更有利于准确地定位和分类物体。

1.5 动态实例交互头

首先,初始化 N 个 proposal boxes,通过 RoI Align[19] 操作为每个 box 提取其对应的 RoI feature。随后,将这些 RoI feature 输入到动态实例交互头[20](dynamic instance interactive head,DII head)中,DII head 的主要任务是实现 RoI features 和 proposal features 之间的有效信息交互。在交互之前,本文添加了一个多头自注意力(multi-head self-attention)模块,该模块首先对 proposal feature 进行多头自注意力计算,以推理不同目标之间的关系。通过这一步骤,模型能够在不同的特征之间建立更为准确和细致的联系,使得模型能够更有效地理解不同目标之间的关联和相互作用,从而提升检测的性能。

接着,经过多头自注意力计算后的 proposal feature 被送入全连接层并分为了两个部分。这两个部分随后与 RoI feature 进行卷积操作,实现了特征层面信息的交互,并得到了更为丰富的 proposal feature。然后,将得到的

proposal feature 进行展平和全连接操作,最后进行分类和回归的预测。回归预测阶段通过 1 个包含 3 层的感知计算完成,生成新的 proposal box,而分类预测由 1 个线性投影层实现。图 5 给出了动态实例交互的过程,其中每个 RoI feature 和 proposal feature 都被送入其专用的交互头中。

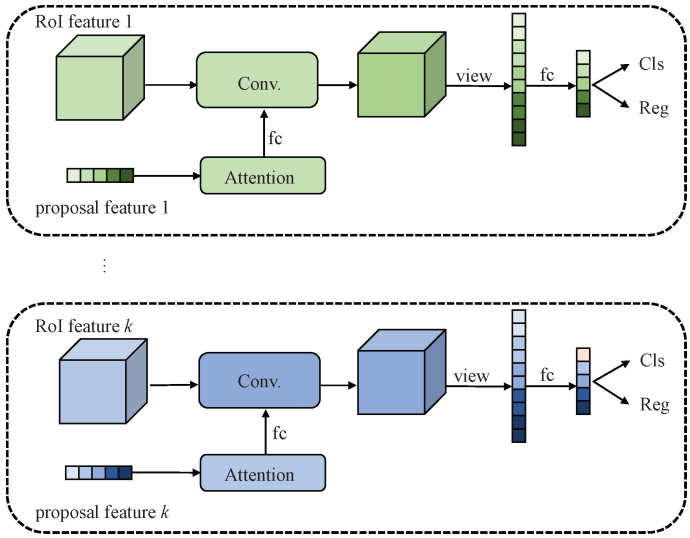


图 5　动态实例交互的过程

Fig. 5　Dynamic instance interaction process

DII head 是一个迭代的过程,它代表了一种连续的信息传递与优化的机制。在这个过程中,上一个 head 所预测的 proposal boxes 和 proposal features 被传递到下一个 head。在下一个 head 中,这些 proposal boxes 不仅被重新利用,并且结合了上一层的 proposal features,通过一系列的细致处理,进一步调整并筛选出更加精准的 object proposals。这种迭代的优化过程,可以有效地提升 object proposals 的质量和准确性,从而提高整体检测性能。

1.6　损失函数

本文采用基于集合的损失函数进行训练。基于集合的损失函数包含 3 部分损失,在预测和真实目标之间实现最佳的二分匹配。匹配的定义如下:

$$\mathcal{L} = \lambda_{Cls} \cdot \mathcal{L}_{Cls} + \lambda_{L1} \cdot \mathcal{L}_{L1} + \lambda_{CIoU} \cdot \mathcal{L}_{CIoU} \tag{1}$$

其中, $\mathcal{L}_{Cls}$ 为预测分类和真实值类别标签的焦点损失,定义如下:

$$L_{Cls} = -\alpha_t (1 - p_t)^{\gamma} \log(p_t) \tag{2}$$

其中, p_t 是模型对样本属于正类别的预测概率, α_t 是平衡参数,用于调整正类别和负类别的权重, γ 是焦点因子,用于调整易分类样本和难分类样本的权重。

$\mathcal{L}_{L1}$ 为归一化中心坐标与预测框和真实框的高度和宽度之间的 $L1$ 损失,其定义如下:

$$L_{L1} = \frac{1}{n} \sum_{i=1}^{n} | \boldsymbol{y}_i - \hat{\boldsymbol{y}}_i | \tag{3}$$

其中, n 为真实框数量, $\boldsymbol{y}_i$ 是第 i 个目标真实框的值, $\hat{\boldsymbol{y}}_i$ 是第 i 个目标预测框的值。

本文采用的 CIoU Loss 定义为式(4):

$$\mathcal{L}_{CIoU} = 1 - IoU + \frac{\rho^2(\boldsymbol{b}, \boldsymbol{b}^{gt})}{c^2} + \alpha v \tag{4}$$

其中, $\boldsymbol{b}$ 和 $\boldsymbol{b}^{gt}$ 表示两个矩形框的中心点, ρ 表示两个矩形框之间的欧氏距离, c 表示两个矩形框的闭包区域的对角线的距离。v 用来衡量两个矩形框相对比例的一致性, α 是权重系数:

$$v = \frac{4}{\pi^2} \left(\arctan \frac{w^{gt}}{h^{gt}} - \arctan \frac{w}{h} \right)^2 \tag{5}$$

$$\alpha = \frac{v}{(1 - IoU) + v} \tag{6}$$

λ_{Cls}、λ_{L1} 和 λ_{CIoU} 为各分量的系数。除了仅在匹配对上执行之外,训练损失和匹配成本相同,最终损失是由训练批次内的目标数量归一化损失的总和。

2　实　验

2.1　实验环境与评价指标

实验在独立服务器上进行,实验的软硬件配置详如表 1 所示。

表 1　实验环境

Table 1　Experimental environment

类别	环境条件	类别	环境条件
CPU	i7-13700K	CUDA 版本	CUDA 11.7
显卡	NVIDIA RTX A4000	深度学习框架	Pytorch
内存	16 GB	运行环境	Linux
系统	Ubuntu 20.04	脚本语言	Python 3.8

本文使用平均精确率(average precision,AP)和各类别 AP 的平均值(mean average precision, mAP)来衡量本文算法的有效性。同时,考虑到模型的计算量和速度,采用了每秒执行的浮点运算次数(giga floating point operations,GFlops)来衡量模型的计算复杂度,采用 Params 来衡量模型的参数量;通过每秒钟检测图像的帧数(frames per second,FPS)作为额外的评估指标来评估模型的检测速度。

2.2 数据集

本文所采用的车间工具数据集为自建数据集,使用数据标注工具 LabelImg,将目标的位置和类别信息进行了标注。所有标注的数据以 json 文件的格式保存,其中包含了目标在图像中的准确定位以及目标的具体类别。标注示意在图 6 中展示。

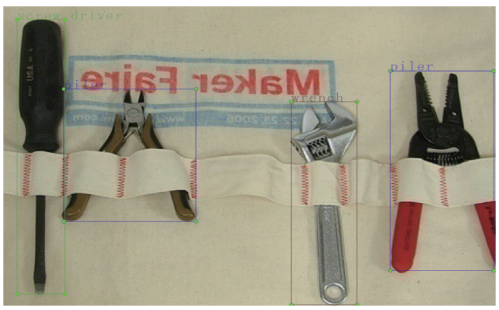


图 6 数据集标注示意图

Fig. 6 Annotated dataset illustration

该数据集涵盖了 5 种不同类别的车间工具,总计包含 2 717 张图像,各类别的标签数量如表 2 所示。按照 8∶1∶1 的比例将数据集中的图片划分成训练集、评估集和测试集。这样的划分有助于在训练模型、评估性能和验证泛化能力等方面进行全面的实验分析。

表 2 各类别标签数量

Table 2 Number of labels for each class

名称	hammer	piler	rope	screw driver	wrench
数量	1 086	440	202	815	995

为了增加图像样本数量,在模型训练之前,对图像应用了多种数据增强操作,其中包括平移变换、翻转变换、随机裁剪等技术。通过这些操作,原始图像被修改并扩充,生成了一系列具有不同变化的新图像样本。平移变换使得图像在空间位置上发生微小的移动,翻转变换则改变了图像的方向,而随机裁剪则在保持图像主题的同时剪裁了周围背景。这些增强操作旨在保持原始图像基本结构和颜色分布的基础上,微调图像的空间几何和像素值,从而增加样本的多样性。这样的处理方式不仅有助于模型更好地学习到对象的不同变化和视角,还可以提高模型的泛化能力和对新样本的适应性。

2.3 消融实验

为验证改进的各模块对模型的影响,本文进行了 3 种关键模块的消融实验,涵盖了改进的特征融合模块(Fusion)、多头自注意力模块(Attention)和 CIoU Loss 函数(Loss)。通过在主干网络之后添加改进的特征融合模块;在动态交互之前添加多头自注意力机制;将 IoU 损失替换为 CIoU Loss 完成消融实验。

为确保消融实验对比的公正性,各个实验间保持相同的软硬件配置。统一使用相同的训练硬件设备和软件环境,并且保持了一致的训练参数设置。具体而言,消融实验将迭代次数设置为 36 次,学习率为 0.000 025,batch_size 大小为 16,并采用 AdamW 优化器进行网络优化。消融实验的实验结果如表 3 所示,详细展示了引入各个模块的情况下网络性能的变化。

表 3 网络主要模块消融实验对比

Table 3 Comparative analysis of ablation experiments on key network modules

Fusion	Attention	Loss	hammer	piler	rope	screw driver	wrench	mAP	Params
			0.867	0.788	0.931	0.801	0.853	0.848	107.99
√			**0.911**	0.769	0.942	0.833	0.897	0.87	107.99
	√		0.882	0.884	0.93	0.835	0.903	0.887	109.57
		√	0.884	0.875	0.939	0.827	0.916	0.889	107.99
√	√		0.884	0.864	0.984	0.81	0.915	0.892	109.57
√		√	0.887	0.875	0.939	**0.846**	0.917	0.893	107.99
	√	√	0.883	0.87	0.937	0.864	0.922	0.895	109.57
√	√	√	0.905	**0.891**	**0.988**	0.841	**0.923**	**0.91**	**109.57**

由表 3 可知,引入改进后的特征融合模块后,除了 piler 类别的检测精确率下降了 0.019 外,其他类别均有所上升,平均精确率提高了 0.22,这表明仅通过添加特征融合模块,对于物体形态不会改变的情况能够有效检测,但对于像 piler 类别有张开和闭合多种形态的情况则有一定不足;引入特征融合模块后,参数量并无明显变化,说明特征融合层主要是对已有的特征图进行组合操作,并不会产生大量额外的参数。

通过在动态交互之前引入多头自注意力模块,所有类别的检测精确率都得到了提升,特别是 piler 类别提升最为显著,其精确率由 0.788 上升到 0.884,平均精确率增加了 0.039,参数量从 107.99 提升到 109.57,增加了 1.59,这表明多头自注意力机制能够建立不同目标之间更细致的联系,使模型能够更好地理解不同目标之间的关联,从而提高检测精度,但会增加额外的计算参数。

在引入 CIoU Loss 函数后,所有类别的检测精确率也有一定的提升,其中 wrench 类别的提升最为明显,从 0.853 提升到 0.916,平均精确率增加了 0.041,值得注意的是参数量并无明显变化,这表明引入 CIoU Loss 能在不增加模型计算复杂度的同时,有效提升检测精度。

2.4　目标检测算法性能对比实验

为验证所提出算法的有效性,本文算法分别与 Faster R-CNN、RetinaNet、Cascade、Foveabox、Dynamic R-CNN、Sparse R-CNN 和 YOLOv5x 在检测精度、计算量和推理速度 3 个指标上进行对比实验,实验结果如表 4 所示。

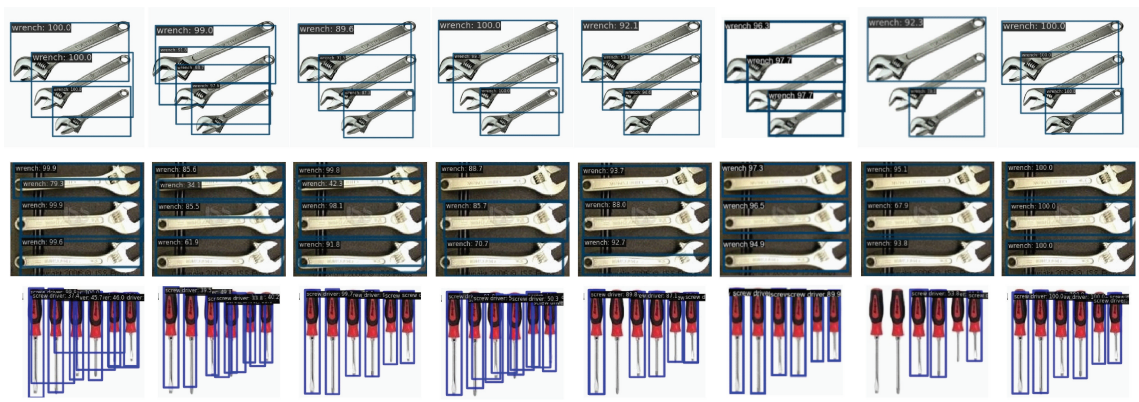
表 4　不同算法对比

Table 4　Comparison of different algorithms

算法	hammer	piler	rope	screw driver	wrench	mAP	GFlops	FPS
Faster R-CNN	0.799	0.82	0.948	**0.851**	0.921	0.868	210.33	16.2
RetinaNet	0.866	0.841	0.941	0.816	0.919	0.877	211.63	17.2
Cascade	0.854	0.844	0.961	0.823	0.916	0.88	238.12	14.4
Foveabox	0.862	0.866	0.937	0.846	0.915	0.885	208.09	17.5
Dynamic R-CNN	0.835	0.816	0.94	0.845	0.923	0.872	203.13	16.3
YOLOv5x	0.899	0.835	0.903	0.842	**0.956**	0.887	203.8	**50.25**
Sparse R-CNN	0.866	0.779	0.941	0.793	0.883	0.852	**149.9**	21.5
Ours	**0.905**	**0.891**	**0.988**	0.841	0.923	**0.91**	174.18	18.5

从表 4 可以看出,本文提出的模型相比较 Faster R-CNN 模型、RetinaNet 模型、Cascade 模型、Foveabox 模型、Dynamic R-CNN 模型、YOLOv5x 模型、Sparse R-CNN 模型的平均精确率分别提高了 0.042、0.033、0.03、0.025、0.038、0.023、0.058。对比 Faster R-CNN,算法的浮点运算次数从 210.33 M 降低至 174.18 M,推理速度 FPS 由 16.2 fps 提升至 18.5 fps,相较于 RetinaNet、Cascade 等方法,提升效果同样较为显著。YOLOv5x 的检测速度是绝对的优势,但本文所提算法的平均精确率比 YOLOv5x 高出了 0.023,计算复杂度略低于 YOLOv5x 模型。本文算法与 Sparse R-CNN 相比,平均精确率有了很大的提升,从 0.852 提升到 0.91,但 Swin-Transformer 结构计算复杂度高,故计算量和检测速度略低于 Sparse R-CNN 算法。

为了更清晰地展示检测结果,图 7 展示了本文算法与主流目标检测算法的对比结果。可以看出,Faster R-CNN 和 RetinaNet 算法在实际检测精度方面欠佳,对于 wrench 和 screw driver 两类目标出现误检;Cascade 算法在处理 wrench 时出现误检,而 Foveabox 则在 screw driver 类别的检测中存在误检;相比之下,Dynamic R-CNN 虽然能够避免误检,但在 screw driver 类别的检测中却可能出现漏检;Sparse R-CNN 出现多处漏检,检测效果不佳;本文算法与 YOLOv5x 相比,虽然速度稍慢,但不存在漏检,概率分数和定位精度高,整体检测效果较好,为其在实际工程中的应用提供了坚实的基础。



(a) Faster R-CNN 模型 (a) Faster R-CNN model　(b) RetinaNet 模型 (b) RetinaNet model　(c) Cascade 模型 (c) Cascade model　(d) Foveabox 模型 (d) Foveabox model　(e) Dynamic R-CNN 模型 (e) Dynamic R-CNN model　(f) YOLOv5x 模型 (f) YOLOv5x model　(g) Sparse R-CNN 模型 (g) Sparse R-CNN model　(h) 本文模型 (h) Our model

图 7　不同算法检测结果

Fig. 7　Detection results of different algorithms

3 结 论

本文提出了一种基于稀疏可学习 proposal 的车间工具目标检测方法,通过引入稀疏表示和可学习 proposal 的概念,有效地提升了检测的准确性和鲁棒性。稀疏表示的引入有效降低检测过程中的参数量,而可学习 proposal 方法则使得模型能够自适应生成适用于车间工具目标的候选框,从而解决了传统方法中固定尺度和比例的限制。在本文方法中,将 Swin-Transformer 结构引入,增加特征提取的全局精确度,减少模型量化误差;引入了改进的多尺度特征融合网络,增加了模型对于不同尺度目标的检测性能。同时,结合了多头注意力和动态卷积计算,在不同特征之间建立更为准确和细致的联系。采用 CIoU Loss 策略,通过综合考虑位置、尺度和形状信息,并引入尺度惩罚项,使得模型对边界框的回归更为全面和准确。实验结果表明,本文提出的方法在检测精度上表现优异,能够准确识别车间工具的类别,基本满足实时检测的需求。这些改进和策略的综合运用使得本文的方法在车间工具目标检测任务中取得了显著的性能提升。

参考文献

[1] 文笑雨, 王康红, 孙海强, 等. 集成式工艺规划与车间调度问题研究现状及发展[J]. 重庆大学学报, 2021, 44(2): 120-128.
WEN X Y, WANG K H, SUN H Q, et al. Research status and development of integrated process planning and workshop scheduling problems[J]. Journal of Chongqing University, 2021, 44(2): 120-128.

[2] 许德刚,王露,李凡. 深度学习的典型目标检测算法研究综述[J]. 计算机工程与应用, 2021, 57 (8): 10-25.
XU D G, WANG L, LI F. A review of typical target detection algorithms based on deeplearning [J]. Computer Engineering and Applications, 2021, 57(8): 10-25.

[3] 王呈,黄义超,杨桂锋. 基于空间特征融合的车间作业工具检测算法 [J]. 电子测量与仪器学报, 2023, 37 (3): 39-49.
WANG CH, HUANG Y CH, YANG G F. Workshop tool detection algorithm based on spatial feature fusion [J]. Journal of Electronic Measurement and Instrumentation, 2023, 37(3): 39-49.

[4] 杨瑾,陈灯,张彦铎,等. 基于多尺度特征融合网络的铁路工机具目标检测 [J]. 电子测量技术, 2022, 45 (17): 94-100.
YANG J, CHEN D, ZHANG Y D, et al. Railway engineering tool detection based on multi-scale feature fusion network [J]. Electronic Measurement Technology, 2022, 45(17): 94-100.

[5] 李玮,高林. 改进 RetinaNet 的工艺流程检测算法 [J]. 电子测量与仪器学报, 2023, 37 (7): 104-112.
LI W, GAO L. Process detection algorithm improvement for RetinaNet [J]. Journal of Electronic Measurement and Instrumentation, 2023, 37(7): 104-112.

[6] 叶飞,骆星智,宋永春,等. 基于双特征融合的改进 R-CNN 电力小金具缺陷检测方法研究 [J]. 电子测量与仪器学报, 2023, 37 (7): 213-220.
YE F, LUO X ZH, SONG Y CH, et al. Research on defect detection method of improved R-CNN electric small hardware based on dual feature fusion [J]. Journal of Electronic Measurement and Instrumentation, 2023, 37(7): 213-220.

[7] 赵梓杉,桑海峰. 基于改进的 YOLOv5 的交通锥标检测系统 [J]. 电子测量与仪器学报, 2023, 37 (2): 56-64.
ZHAO Z SH, SANG H F. Traffic cone detection system based on improved YOLOv5 [J]. Journal of Electronic Measurement and Instrumentation, 2023, 37 (2): 56-64.

[8] 蒋光峰,胡鹏程,叶桦,等. 基于旋转中心点估计的遥感目标精确检测算法[J]. 计算机应用研究, 2021, 38 (9): 2866-2870.
JIANG G F, HU P CH, YE H, et al. Accurate detection algorithm of remote sensing targets based on rotation center estimation [J]. Computer Application Research, 2021, 38(9): 2866-2870.

[9] 刘晓雯,郭继昌,郑司达. 采用渐进式网络的弱监督显著性目标检测算法[J]. 西安电子科技大学学报, 2023, 50 (1): 48-57.
LIU X W, GUO J CH, ZHENG S D. Weakly supervised saliency target detection algorithm based on progressive network [J]. Journal of Xidian University, 2023, 50(1): 48-57.

[10] 吴湘宁,贺鹏,邓中港,等. 一种基于注意力机制的小目标检测深度学习模型[J]. 计算机工程与科学, 2021, 43 (1): 95-104.
WU X N, HE P, DENG ZH G, et al. A deep learning model for small target detection based on attention mechanism [J]. Computer Engineering and Science, 2021, 43(1): 95-104.

[11] 王伟锋,金杰,陈景明. 基于感受野的快速小目标检测算法[J]. 激光与光电子学进展, 2020, 57 (2): 250-255.
WANG W F, JIN J, CHEN J M. Fast small target

detection algorithm based on receptive field [J]. Progress in Laser and Optoelectronics, 2020, 57(2): 250-255.

[12] 黄媛媛, 熊文博, 张宏伟, 等. 基于 U 型 Swin Transformer 自编码器的色织物缺陷检测[J]. 激光与光电子学进展, 2023, 60(12): 303-310.
HUANG Y Y, XIONG W B, ZHANG H W, et al. Fabric defect detection based on U-shaped Swin Transformer autoencoder [J]. Progress in Laser and Optoelectronics, 2023, 60(12): 303-310.

[13] AHMED B, GULLIVER A T, ALZAHIR S. Image splicing detection using mask-RCNN [J]. Signal, Image and Video Processing, 2020, 14(5): 1-8.

[14] ZHANG Y, XU B, ZHAO T. Convolutional multi-head self-attention on memory for aspect sentiment classification [J]. IEEE/CAA Journal of Automatica Sinica, 2020, 7(4): 1038-1044.

[15] 贾小云, 翁佳顺, 刘颜荤. 多网络和多头注意力融合的场景文本识别算法 [J]. 计算机时代, 2023(8): 46-51.
JIA X Y, WENG J SH, LIU Y L. Scene text recognition algorithm based on multi-networks and multi-head attention fusion [J]. Computer Era, 2023(8): 46-51.

[16] 陈景明, 金杰, 王伟锋. 基于特征金字塔网络的改进算法[J]. 激光与光电子学进展, 2019, 56(21): 165-170.
CHEN J M, JIN J, WANG W F. Improved algorithm based on feature pyramid network [J]. Progress in Laser and Optoelectronics, 2019, 56(21): 165-170.

[17] 董绍江, 刘伟, 蔡巍巍, 等. 基于分层精简双线性注意力网络的鱼类识别[J]. 计算机工程与应用, 2022, 58(5): 186-192.
DONG SH J, LIU W, CAI W W, et al. Fish recognition based on hierarchical and simplified bilinear attention network [J]. Journal of Computer Engineering and Applications, 2022, 58(5): 186-192.

[18] 戴媛, 易本顺, 肖进胜, 等. 基于改进旋转区域生成网络的遥感图像目标检测[J]. 光学学报, 2020, 40(1): 270-280.
DAI Y, YI B SH, XIAO J SH, et al. Remote sensing image target detection based on improved rotated region generating network [J]. Acta Optica Sinica, 2020, 40(1): 270-280.

[19] SEKAR A, VARALAKSHMI P. Automatic road crack detection and classification using multi-tasking faster RCNN [J]. Journal of Intelligent & Fuzzy Systems, 2021, 41(6): 6615-6628.

[20] 王爱丽, 刘美红, 薛冬, 等. 结合动态卷积和三重注意力机制的高光谱图像分类[J]. 激光与光电子学进展, 2022, 59(10): 341-351.
WANG AI L, LIU M H, XUE D, et al. Hyperspectral image classification combining dynamic convolution and triple attention mechanism [J]. Progress in Laser and Optoelectronics, 2022, 59(10): 341-351.

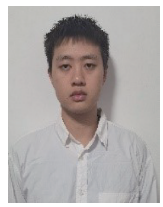
作者简介

刘珍兵, 2022 年于南通理工学院获得学士学位, 现为江苏海洋大学电子工程学院硕士研究生, 主要研究方向为图像分析和智能系统。
E-mail: lzb2312824955@163.com

Liu Zhenbing received his B. Sc. degree from Nantong Institute of Technology in 2022. Now he is a M. Sc. candidate at the School of Electronic Engineering of Jiangsu Ocean University. His main research interests include image analysis and intelligent system.

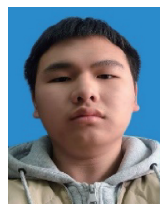
孙巧榆(通信作者), 2012 年于华东师范大学获得博士学位, 现为江苏海洋大学电子工程学院副教授, 主要研究方向为图像分析和智能系统。
E-mail: sunqy@jou.edu.cn

Sun Qiaoyu (Corresponding author) received the Ph. D. degree in 2012 from East China Normal University. Now she is an associate professor at the School of Electronic Engineering, Jiangsu Ocean University. Her main research interests include image analysis and intelligent system.



王述文, 2021 年于集美大学获得学士学位, 现为江苏海洋大学电子工程学院硕士研究生, 主要研究方向为图像分析和智能系统。
E-mail: 2250065581@qq.com

Wang Shuwen received his B. Sc. degree from Jimei University in 2021. Now he is a M. Sc. candidate at the School of Electronic Engineering of Jiangsu Ocean University. His main research interests include image analysis and intelligent system.



夏嘉伟, 2022 年就读于江苏海洋大学, 现为江苏海洋大学电子工程学院本科生。
E-mail: 3242342115@qq.com

Xia Jiawei studied at Jiangsu Ocean University in 2022. Now he is a B. Sc. candidate in the School of Electronic Engineering of Jiangsu Ocean University.