

融合多尺度特征的竹筷表面缺陷检测^{*}莫太平^{1,2} 陈良伟¹ 张向文^{1,2} 孙 鹏^{1,2,3}(1. 桂林电子科技大学电子工程与自动化学院 桂林 541004; 2. 智能综合自动化广西高校重点实验室 桂林 541004;
3. 广西自动检测技术与仪器重点实验室(桂林电子科技大学) 桂林 541004)

摘 要:为解决竹筷质检过程中小目标检测精度低、多目标聚集重叠时出现易漏检等问题,提出了一种基于改进实时检测 Transformer (RT-DETR) 的高效竹筷切片检测技术 EBCS-DETR。首先,设计高效紧凑的空间到深度挤压激励模块 SDSEM 增强 ResNet18 浅层特征表达能力;其次,构建快速高效注意力模块 FasterEMA,减少深层网络参数量,提升骨干网络的特征提取率;然后,采用内容感知上采样模块 CARAFE 与双向特征融合模块 Bi-FPN 修改颈部网络,改善小目标缺陷边界的定位及区分能力,减少背景噪声干扰。实验结果显示,在包含 5 400 张图像的自建数据集中,模型在 3 个测试集的平均检测精度(mAP_{50})相较于基线模型分别提升了 3.3%、3.4% 和 3.1%,均优于其他相似参数量的先进模型。鲁棒性实验进一步证实了该模型检测竹筷切片表面缺陷的可靠性和应用潜力,这对于复杂生产环境下的竹筷质检至关重要,可以为精准竹筷质检提供新的参考和思路。

关键词: 竹筷缺陷检测;多尺度目标检测;RT-DETR;注意力机制;特征融合

中图分类号: TP391.4; TN919.8 **文献标识码:** A **国家标准学科分类代码:** 520.6040

Surface defect detection of bamboo chopsticks with fusion of multiscale features

Mo Taiping^{1,2} Chen Liangwei¹ Zhang Xiangwen^{1,2} Sun Peng^{1,2,3}

(1. School of Electronic Engineering and Automation, Guilin University of Electronic Technology, Guilin 541004, China;

2. Key Laboratory of Intelligence Integrated Automation in Guangxi Universities, Guilin 541004, China; 3. Guangxi Key Laboratory of Automatic Detecting Technology and Instruments (Guilin University of Electronic Technology), Guilin 541004, China)

Abstract: Aiming at the problems of low detection accuracy of small targets in the quality inspection of bamboo chopsticks and leakage of detection when multiple targets are gathered and overlapped, this paper proposes an efficient bamboo chopstick slice detection method based on an improved real-time detection Transformer (RT-DETR), named EBCS-DETR. SDSEM module is designed to enhance the expressive capability of shallow features in ResNet18, improving the representation of fine-grained details. Second, FasterEMA module is introduced to reduce the number of parameters in deep layers and accelerate feature extraction in the backbone network. Third, the neck network is enhanced by integrating CARAFE module with Bi-FPN, which improves localization accuracy for small defective regions, strengthens discrimination at object boundaries and suppresses background noise. Experimental results on a self-constructed dataset of 5 400 images demonstrate that the proposed EBCS-DETR achieves mean average precision (mAP_{50}) improvements of 3.3%, 3.4%, and 3.1% over the baseline model across three test sets, consistently outperforming other state-of-the-art models with comparable parameter counts. In addition, robust experiments further confirmed the reliability and application potential of EBCS-DETR for detecting defects in bamboo chopsticks slices, which is crucial for complex production environments. This work provides a novel and effective solution for high-precision automated inspection of bamboo chopsticks, offering valuable insights for intelligent quality assurance in similar manufacturing scenarios.

Keywords: bamboo chopsticks defect detection; multi-scale object detection; RT-DETR; attentional mechanisms; feature fusion

0 引 言

具有自然表面的圆柱形物体,如竹签、医用棉签、木筷、

木制器皿和生物可降解吸管等,给自动光学检测带来了巨大挑战^[1-3]。竹筷表面表现出复杂纤维纹理,是具有代表性的样本,本文围绕竹筷的表面缺陷检测展开研究。筷子作

收稿日期:2025-10-23

^{*} 基金项目:国家自然科学基金(62541310,62263006)、广西自动检测技术与仪器重点实验室基金(YQ26102)、2025 年度广西高校中青年教师科研基础能力提升项目(2025KY0258)、广西自动检测技术与仪器重点实验室基金(YQ21107)、桂林电子科技大学研究生教育创新计划(2024YCXS125)项目资助

为东亚最主要的餐饮用具,广泛用于中国、日本、韩国和越南,在全球也越来越受欢迎,其中竹筷因为方便、卫生和成本低廉,在餐饮行业中的使用占比很高^[4]。尽管竹子本身具有抗菌特性,但部分竹筷的裂缝、毛刺和脏污等缺陷使竹筷存在使用风险。因此,在生产过程中准确甄别劣质品对保护消费者权益、提高产品质量和促进行业健康发展至关重要。

近年来,基于深度学习的工业缺陷检测方法,因为高效率、高质量和强鲁棒性,能够适应复杂的生产环境,已成为工厂批量竹筷缺陷检测的首选方法。实现缺陷检测具体有目标检测和异常检测两大途径^[5]。其中,基于异常检测的方法属于无监督的范畴,仅需使用正常样本训练模型,能够检测出已知和未知的缺陷,具有较强的泛化能力。然而,这类方法通常对小目标支持不好,推理过程复杂,实际应用中一般用于需要图像分割的工业检测^[6-7]。相比之下,基于目标检测的方法虽然通过监督学习,需要大量标注数据训练模型,但因模型输出简单,不需要复杂推理就能够同时实现缺陷的精确定位与准确分类,特别是以 YOLO^[8-9]、Faster 区域卷积神经网络(region-convolutional neural network, R-CNN)^[10]、PPYOLOE^[11-12] 和实时检测 Transformer (real-time detection Transformer, RT-DETR)^[13] 为代表的现代检测框架,已在多种工业场景中展现出优异性能。

目前国内外的研究大都使用 YOLO 作为缺陷目标检测的框架。朱黎颖等^[14]将空间到深度卷积(space-to-depth convolution, SPD-Conv)模块^[15]直接应用于 YOLOX 的骨干网络中,加强了印刷电路板(Printed Circuit Board, PCB)表面缺陷全局信息与细节信息提取,然而这种直接的应用并未能充分发挥 SPD 模块的优势。杨瑞峰等^[16]基于 YOLOv10 在颈部网络中插入高效多尺度注意力机制(efficient multi-scale attention, EMA)模块^[17],提升了液晶显示屏表面缺陷检测的准确性。Liang 等^[18]基于 YOLOv5 使用双向特征融合模块(bidirectional feature pyramid network, Bi-FPN)^[19]改进颈部网络来更好地解决无人机遥感检测中的多尺度问题,进而提高小尺寸物体的识别效果。但是 YOLO 模型属于单阶段检测,因为卷积操作接受域存在不可避免的局限性和卷积结构中的归纳偏差使得建立长距离依赖关系和全局上下文关系具有挑战性。

与基于卷积神经网络的单阶段检测和双阶段检测进行对比。基于 Transformer^[20]的 DETR^[21]提供了一种完全端到端的可训练架构,这种架构无需先验知识和非极大值抑制(non-maximum suppression, NMS)后处理步骤,也不存在提议和锚点等约束。尽管 DETR 高效且直接,但其较大的参数量和计算成本成为了挑战,这促使研究界提出多种改进方案。其中,百度飞桨团队提出的 RT-DETR 模型,在实时性和准确性方面超越了 YOLOv8。Li 等^[22]基于 RT-DETR 提出一种组合 SPD-Conv 的方案 SDTM,并将其应用于 HGNet 骨干网络的浅层中,提高了模型捕获局部细

节特征的能力,并且这对模型的参数量影响不高。张路等^[23]基于 RT-DETR 提出了 Faster-EMA 模块用于替换骨干网络 ResNet18 的全部 BasicBlock,能够在有效降低模型的参数量和模型深度的同时,显著提升对水下目标的特征提取能力。

本文基于现有的研究,在 RT-DETR 框架上进行改进,提出一种适用于竹筷缺陷检测的算法,高效竹筷切片检测(efficient bamboo chopstick slices-detection transformer, EBCS-DETR),以克服竹筷缺陷检测中存在的难题,提升检测精度并保证实时性,主要贡献如下。

1) 将基于图像切片的检测方法应用于竹筷表面缺陷检测,有效解决了整图检测在高密度、小目标场景下的漏检与误检问题。

2) 设计空间到深度挤压激励模块(space-to-depth squeeze-and-excitation module, SDSEM)应用于 ResNet18 骨干网络的浅层,增强局部特征的捕获能力,并弥补全局特征;将快速高效多尺度注意力机制模块(faster efficient multiscale attention, FasterEMA)仅用于 ResNet18 骨干网络的深层,以降低网络参数量和计算复杂度,并提高特征提取效率。

3) 重构特征融合颈部网络,使用内容感知上采样模块(content-aware reassembly of features, CARAFE)基于内容感知进行插值,有效恢复高分辨率特征图细节;同时,基于 Bi-FPN 模块,实现高、低分辨率特征的自适应双向聚合。

1 EBCS-DETR 模型

Zhao 等^[13]提出的 RT-DETR 系列模型中,RT-DETR-R18 使用轻量级的 ResNet18 作为主干网络,推理迅速能保证实时性,因此本研究选择该模型作为基线进行改进。改进的模型结构如图 1 所示,其中,主干网络 ResNet18 由初始卷积层和 4 个残差层构成,每个残差层又由 BasicBlock 实现的卷积块(convolutional block, Conv_Block)和恒等块(identity block, ID_Block)组成,二者区别在于 stride 参数的不同。本研究将第 2、3 层的 Conv_Block 替换为 SDSEM 模块,以增强浅层特征表达能力;同时,将第 4、5 层的 ID_Block 替换为 FasterEMA 模块,在减少参数量的同时提升深层特征提取效率。此外,在混合编码器中引入 CARAFE 与 Bi-FPN,通过内容感知的上采样机制与双向跨尺度连接,增强小目标缺陷的边界定位能力,抑制背景噪声干扰,提升了整体检测性能。

1.1 SDSEM 模块

在竹筷切片图像检测中存在大量小目标和边缘目标,这极大影响了模型的检测精度,为克服这一难题,本研究将 SPD-Conv 模块^[15]和 SE^[24]两种互补的技术相互集成。

SPD-Conv 可减轻向下采样过程中的信息丢失,这对小目标检测很重要。如图 2 所示,包括一个 SPD 层和一个

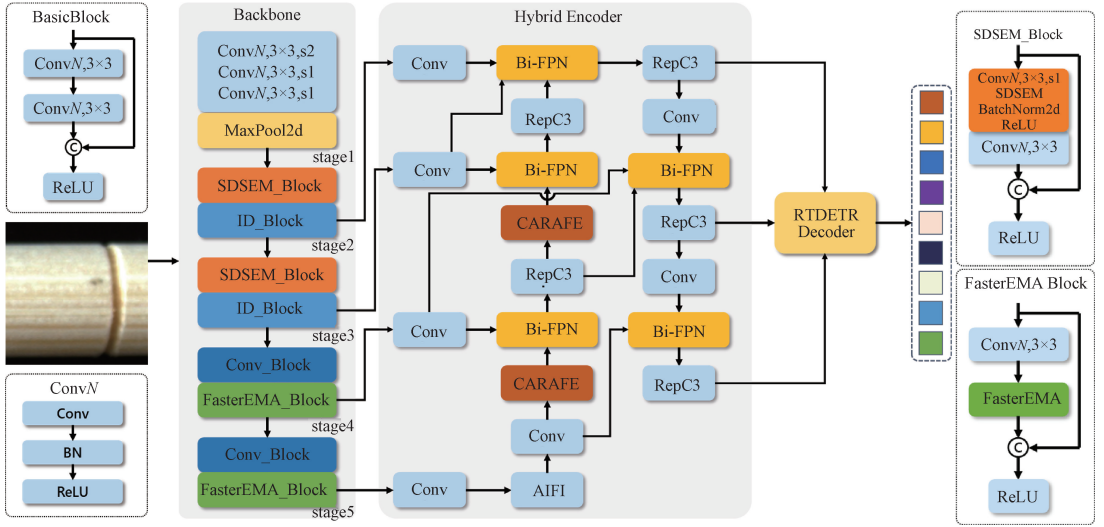


图1 EBCS-DETR 模型网络结构

Fig. 1 Architecture of the proposed EBCS-DETR model

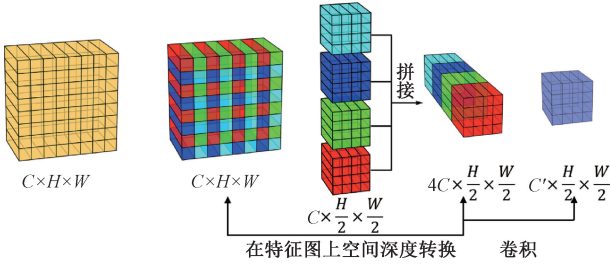


图2 SPD-Conv 处理流程

Fig. 2 The SPD-Conv processing flowchart

非分步卷积层,SPD层重新连接信道,保留了所有信息,然后非分步卷积调整信道维度。SPD-Conv 在减少下采样信息损失方面的功效显著,在小目标检测和分割任务中表现出强劲的性能^[14,22]。

如图3所示,作为补充,SE注意力机制模块通过自适应重校信道特征权重来减少小目标漏检。利用全局上下文,该模块会强调有用的信道,而抑制无用的信道,提高了特征灵敏度和表征能力,同时尽可能地降低计算开销。在小目标检测中,SE能有效降低漏检率,提高整体鲁棒性。

为了解决简单串联 SPD-Conv 和 SE 带来的通道膨胀和计算负担问题,并实现性能和效率的平衡,本研究提出了一种高效而紧凑的整合 SPD-Conv 和 SE 的模块——SDSEM。此模块能结合 CNN 和 Transformer 的优势,加强全局特征提取,同时保留捕捉局部细节和边界特征的能力。如图4所示,该模块包括4个深度卷积层 DWConv、2个 SPD-Conv、挤压激励层 SELayer 和转置卷积层 DeConv,然后是融合模块,用于整合多尺度特征以改进表示。

SDSEM 模块由两个协同增强局部和全局特征表示的并行分支组成。首先,负责局部特征的分支通过深度卷积 DWConv 处理得出 X_1 。接着。负责全局特征的分支结合

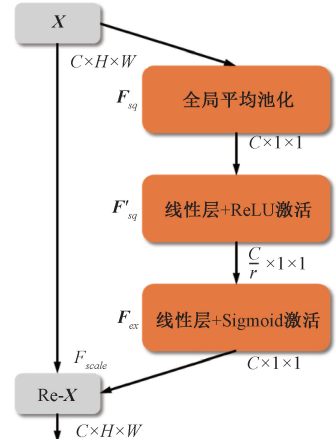


图3 SELayer 网络结构

Fig. 3 The illustration of SELayer

了 DWConv 和 SPD 操作进行下采样,这在保留关键缺陷和边缘信息的同时压缩了特征图的维度。然后,SELayer 识别长距离区域的突出特征并结合全局背景完善特征表示,再通过上采样卷积得出 X_3 。最后,将特征图 X_1 、下采样过程中的 X_2 与 X_3 沿着通道轴进行串联并通过 DWConv 层进一步处理,生成最终特征图。这种设计有效整合了多尺度信息,确保输出结果既能保持详细的局部结构,又能捕获全面的上下文关系语义。SDSEM 模块的数学实现为:

$$X_1 = \text{DWConv}_1(X) \quad (1)$$

$$U = \text{SPD}(\text{DWConv}_1(X)) \quad (2)$$

$$X_2 = \text{DWConv}_2(U) \quad (3)$$

$$V = \text{SE}(\text{SPD}(U)) \quad (4)$$

$$X_3 = \text{DeConv}(V) \quad (5)$$

$$Y = \text{DWConv}_3(\text{Concat}[X_1, X_2, X_3]) \quad (6)$$

在 RT-DETR 的骨干网络 ResNet 中,浅层网络主要

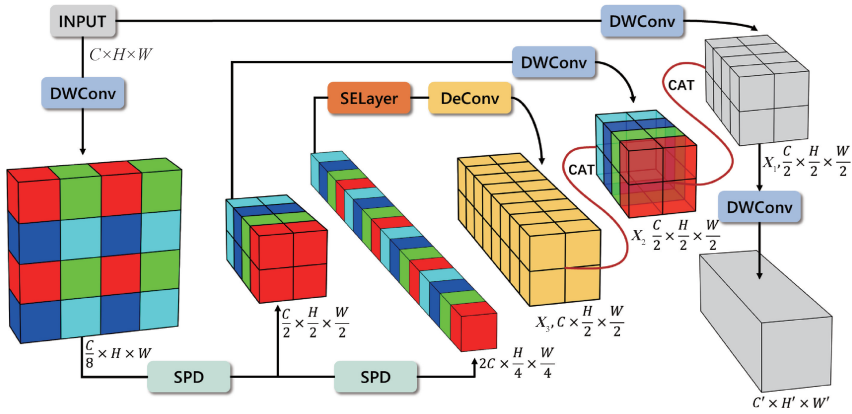


图 4 SDSEM 网络结构

Fig. 4 The structure network of SDSEM

捕捉局部细节特征,而深层网络则捕捉全局语义信息。本研究设计 SDSEM 改进了 ResNet 中较浅层(第 2、3 层)的 BasicBlock。这种改进保留更增强了浅层捕捉局部细节信息的能力,同时弥补了浅层缺乏全局特征的缺陷,便于模型建立长距离依赖关系和全局上下文关系。

1.2 FasterEMA 模块

Chen 等^[25]提出部分卷积(partial convolution, PConv)通过优化空间特征提取中的冗余计算和内存访问来提高浮点运算效率(floating-point operations per second, FLOPS)。Ouyang 等^[17]在 CA 注意力机制基础上提出 EMA 注意力机制,能高效保留通道信息并降低计算开销。为了进一步缩小理论效率与硬件部署之间的差距,张路等^[23]将这两个互补的模块进一步整合在一起,提出了 FasterEMA 模块,其结构如图 5 所示。这种整合方法既保留了关键的通道信息,又减少了内存访问周期。

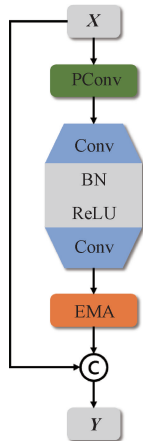


图 5 FasterEMA 网络结构

Fig. 5 The structure network of FasterEMA

本研究引入 FasterEMA 模块,改进 ResNet 负责高层次特征整合的深层(第 4、5 层)的 BasicBlock,该模块可在保持轻量化的同时提升模型效率与精度,实现高效性与准确性的平衡。

1.3 CARAFE 和 Bi-FPN 模块

RT-DETR 中的跨尺度特征融合(cross-scale feature fusion, CCFF)模块在检测小目标缺陷时存在局限性,尤其是当目标边界模糊时,检测精度受到较大影响。原因在于 ResNet 骨干网络的浅层网络只负责感受边界和位置等细粒度信息,而深层网络负责捕捉全局语义依赖关系,但容易丢失局部的细节。这些多尺度特征的融合处理不好会造成信息冲突或丢失,削弱小目标边界与背景之间的辨别能力,从而影响检测性能。本研究采用的解决方案是将 CARAFE^[26]与 Bi-FPN^[19]协同起来,克服这些局限性。

CARAFE 增强上采样:CARAFE 通过感知动态生成内核进行上采样,相比固定内核,这能根据局部内容自适应地重新组合特征,既保留缺陷边界,又抑制背景纹理干扰。

BiFPN 拓扑:通过移除单输入节点并在 P2~P5 层之间建立输入/输出节点的直接跨尺度连接,增强了特征聚合能力,并减少了特征冲突,重新连接的结构如图 6 所示。通过整合浅层局部细节和深层全局语义信息,该结构增强了小目标和模糊边界的表征能力,显著提高了缺陷区域检测的准确性,有效抑制了假阳性和假阴性^[18,27]。

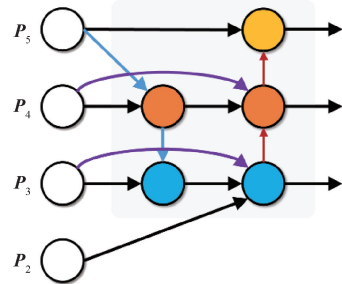


图 6 EBCS-DETR 中的 Bi-FPN 结构

Fig. 6 The structure of Bi-FPN in EBCS-DETR

2 实 验

2.1 数据集制作

如图 7 所示,图像采集系统采用了海康机器人工业面

阵相机(MV-CS200-10UC)和海康机器人工业自动化镜头(MVL-KF1624M-25MP),相机与镜头坚直固定于固特加视觉实验支架上。图像采集系统的详细参数如表 1 所示。采集方案为在阳光充足的条件下,将竹筷水平放置于实验台上,手动对焦,设置合适曝光时间和白平衡度以后,将每根竹筷手动旋转并完成 2~3 张图像的采集。

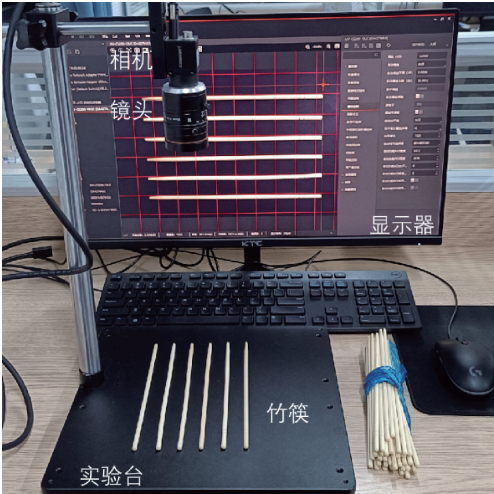


图 7 图像采集系统

Fig. 7 Image acquisition system

表 1 图像采集系统参数

Table 1 Image acquisition system parameter

属性	参数
分辨率	5 472×3 648 pixels
焦距	16 mm
镜头畸变	0.07%
物距	280 mm
视野	254 mm×219 mm
光源类型	白日自然光
单位像素表示的实际距离	60.3 μm

本实验总共拍摄竹筷不少于 900 根,共 3 320 张图像。

表 2 数据集详细信息

Table 2 Details of datasets

数据集	图像数量	框数量	小目标框/%	中目标框/%	大目标框/%
训练集	3 600	7 693	64.59	31.99	3.42
验证集	450	703	63.16	34.14	2.70
测试集 1	450	1 174	65.25	32.54	2.21
测试集 2	450	1 016	68.41	27.95	3.64
测试集 3	450	1 052	66.92	29.85	3.23
总计	5 400	11 638	65.11	31.63	3.26

表 2 归纳总结了每个数据集的目标框数量和不同大小目标框比例,附,小目标($\text{Area}<32\times32$)、中目标($32\leq\text{Area}<96\times96$)和大目标($\text{Area}\geq96\times96$)。所有数据集共包含 11 638 个目标,其中小目标占比最大,为

鉴于竹筷的实际尺寸为 200 mm×50 mm,直接检测整张图像会丢失很多细节,因此本研究采用了基于图像切片的检测方法,即对每张图像使用边缘检测的方法识别竹筷区域,该区域大小为 4 352×128 pixels,再将该区域通过滑动窗口裁剪成 128×128 pixels 的竹筷切片图像。该处理策略有效降低模型输入的尺寸,显著减小模型训练与推理阶段的计算复杂度,提升处理效率,同时也为后续精细化标注提供了便利。

采用 Label Studio 标注工具,依据预定义的 7 类常见竹筷表面缺陷进行标注,包括裂缝(crack)、脏污(dirt)、粗糙(rough)、黑斑(spot)、缺材(lack)、凹槽(groove)和毛刺(burr),典型示例如图 8 所示。其中,“凹槽”指竹筷尾端设计应有的刻槽结构,部分样本存在凹槽缺失、未闭合或不完整的缺陷形态,故将此类别纳入检测范围。

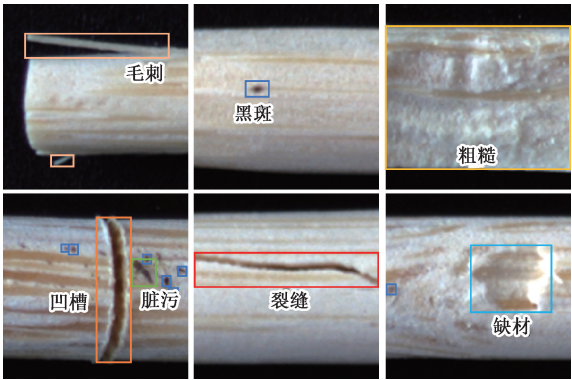


图 8 竹筷表面缺陷示意图

Fig. 8 Schematic diagram of surface defects in bamboo chopsticks

2.2 数据集结构和特征

本研究将 4 500 张有竹筷表面缺陷的切片图像作为数据集,按照 8:1:1 的比例划分为训练集、验证集和测试集。为了进一步评估模型的泛化能力,还额外补充了两个测试集。因此,训练集共 3 600 张图像,验证集共 450 张图像,3 个测试集均包含图像 450 张,数据集详细信息如表 2 所示。

65.11%,且测试集 2 的小目标占比最高。从图 9 提供的 7 种缺陷类型占比分布来看,每个数据集中裂缝和毛刺的占比较高,且 3 个测试集中裂缝的占比呈上升趋势。由于凹槽特征明显且为中目标,易于学习和检测,因此调整了

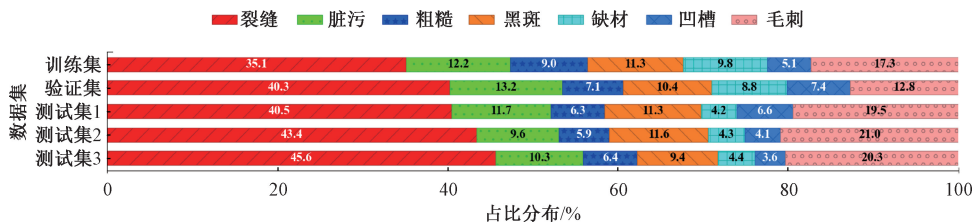


图 9 缺陷类型占比分布

Fig. 9 Proportion distribution of defect types

数据集中凹槽的数量,控制在较少的比例。

2.3 实验环境与参数

本实验的实现是通过搭载 AMD EPYC 9354 16 核的 CPU,64 GB 运行内存,NVIDIA RTX 4090 24 GB 显卡的设备。该设备的系统为 Ubuntu 22.04,驱动为 CUDA 12.6,软件是 Python 3.11.8 和 PyTorch 2.2.2。

为保证实验的公平性,所有模型均设置为不使用预训练权重,训练次数为 100 轮,Batch Size 为 16,输入图像分辨率大小为 224×224 pixels,且图像均为 RGB 三通道彩色,优化器均使用 AdamW,初始学习率为 0.001,动量系数为 0.9,权值衰减系数为 5×10^{-4} 。同时为保持统一性和公平性,所有模型在以上配置的条件下使用的训练、测试和验证数据集均相同。

2.4 评价指标

本研究采用的评估指标包括参数 (parameters, Params),千兆浮点运算 (giga floating point operations, GFLOPs),每秒帧数 (the number of frames per second, FPS) 和平均精度均值 (mean average precision, mAP)。Params 表示模型的参数量;GFLOPs 表示模型总浮点运算量;FPS 表示模型每秒处理的图像数量;AP 表示精确率-召回率曲线下面积;mAP 是多类别 AP 的平均值。其

中,mAP 是衡量目标检测任务性能的关键指标^[18]。在本实验中,检测速度 FPS 是在输入 Batch Size 为 64 的 $3\times 224\times 224$ 张量条件下,对 300 次推理的平均耗时计算得出。部分指标计算公式为:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{7}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{8}$$

$$\text{AP} = \int_0^1 (\text{Precision} \times \text{Recall}) d(\text{Recall}) \tag{9}$$

$$\text{mAP} = \frac{1}{c} \sum_1^c \text{AP}_i \tag{10}$$

其中,TP、FP 和 FN 分别表示真阳性、假阳性和假阴性的个数。平均精度 AP_{50} 和 $\text{AP}_{50,95}$ 的数值部分代表交并比(intersection over union,IoU)阈值。

2.5 SDSEM 模块实验

通过在自建数据集上实验,本文研究不同数量与位置的 SDSEM 模块改进方案对模型检测精度和运行效率的影响。在表 3 中,本文对比了 8 种不同的改进方案,这些方案是使用 SDSEM 模块替换 ResNet 残差层的 Conv_Block 产生的。通过 mAP_{50} 以及 3 个测试集上的平均增量,可以观察不同方案对模型整体检测性能的影响。

表 3 SDSEM 模块实验

Table 3 Experiments of SDSEM module

SDSEM 数量	替换层	平均 增量/%	测试集 1 mAP ₅₀ /%	测试集 2 mAP ₅₀ /%	测试集 3 mAP ₅₀ /%	Params/ ($\times 10^6$)	GFLOPs	FPS/fps
0	—	—	75.3	76.0	73.3	20.2	58.6	27
1	3	1.6	77.3	77.7	74.3	20.7	68.3	26
1	4	1.6	76.7	78.3	74.4	22.3	68.2	26
2	2, 3	1.7	77.2	78.0	74.4	20.7	68.3	26
2	3, 4	0.7	76.4	77.2	73.2	22.8	77.9	26
2	4, 5	0.0	76.4	75.7	72.5	30.9	77.8	23
3	2, 3, 4	0.5	76.9	77.7	71.6	22.9	77.9	25
3	3, 4, 5	1.1	76.6	78.3	73.0	31.4	87.5	22
4	2, 3, 4, 5	1.1	76.1	77.5	74.3	31.4	87.5	22

实验结果表明,当使用一个 SDSEM 模块时,模型的检测精度已有明显提升, mAP_{50} 平均提高了 1.6%。使用两

个 SDSEM 模块时,位于第 2、3 层的组合取得了最高的平均增长, mAP_{50} 平均增加了 1.7%,并在测试集 2 中表现最

优,有效地平衡了精度和泛化性,且相比原模型帧率仅降低了 1 fps,这说明 SDSEM 模块具有明显的紧凑性和高效性。相反,位于第 3、4 层以及位于第 4、5 层的 SDSEM 模块在测试集 2、测试集 3 中表现出较差的泛化能力,表明使用 SDSEM 改进在网络浅层 BasicBlock 更为合适。当使用位于 3、4、5 层的 3 个 SDSEM 模块时,模型实现了次优的平均增长,但模型参数显著增加。

总之,在浅层放置 SDSEM_Block 提高效果显著,这是因为在 ResNet 浅层 SDSEM 不仅捕获了细节特征,而且补充了整个网络的全局信息。另外,在网络较深层放置

SDSEM_Block 可能导致梯度爆炸问题,严重影响模型的泛化能力。基于各种改进方法在 3 个测试集上的表现和平均增量统计,本研究选择将 SDSEM_Block 替换位于原始模型第 2、3 层 BasicBlock 的方案,并在此基础上继续改进。

2.6 颈部网络实验

为验证 CARAFE 与 Bi-FPN 对模型性能的影响,本文在自制数据集上对颈部网络进行消融对比实验,其中 w/o 为英文 without 的缩写,表示基于 EBCS-DETR 颈部网络修改的不同子模型,具体实验结果如表 4 所示。

表 4 颈部网络消融实验

Table 4 Ablation experiments results of neck network

模型	测试集 1		测试集 2		测试集 3		Params/ ($\times 10^6$)	GFLOPs	FPS/ fps
	mAP ₅₀ /	mAP _{50,95} /	mAP ₅₀ /	mAP _{50,95} /	mAP ₅₀ /	mAP _{50,95} /			
	%	%	%	%	%	%			
EBCS-DETR	78.6	45.4	79.4	47.5	76.4	44.2	18.4	73.1	19
w/o CARAFE	77.3	44.8	76.7	46.0	71.4	42.0	18.2	72.8	24
w/o Bi-FPN	74.9	44.0	78.4	46.5	76.6	43.0	18.0	65.1	20
w/o CARAFE & Bi-FPN	75.4	43.6	79.2	46.9	74.4	43.1	17.8	64.8	26

由表 4 可知,使用 CARAFE 与 Bi-FPN 构建的颈部网络比原模型使用的 CCFF 颈部网络泛化性能更优,具体来说,对比第 1 行与最后 1 行,发现在 3 个测试集中 mAP₅₀ 分别提升了 3.2%、0.2% 和 2%,mAP_{50,95} 分别提升了 1.8%、0.6% 和 1.1%,即在测试集 1 与 3 中提升明显。而测试集 2 提升不明显的原因是其包含更多的毛刺(小目标),相对来说更为复杂。另外,可以看到,测试集 2 中仅使用 CARAFE 或 Bi-FPN 的模型相比颈部 CCFF 的模型表现降低了,这是模块耦合不好造成的。仅使用 Bi-FPN 修改颈部网络,帧率降低了 2 fps,但在测试集 1 上相对 CCFF 颈部网络的模型 mAP₅₀ 和 mAP_{50,95} 分别提升了 1.9% 和 1.2%;仅使用 CARAFE 修改颈部网络使帧率降低了 6 fps,但在测试集 3 上相对 CCFF 颈部网络的模型 mAP₅₀

提升了 2.2%。可见使用 CARAFE 进行上采样虽然能提升模型的检测能力,但是使用动态内核进行特征的上采样需要消耗更多的计算资源,这明显影响了模型的运行效率。将 CARAFE 与 Bi-FPN 相结合后,模型检测性能获得显著提升;虽然帧率有所下降,但该损耗处于可接受范围,因此综合考量后选用该方案。

2.7 消融实验

为验证 3 种模块对 RT-DETR-R18 改进的有效性,在骨干网络 ResNet 中使用 SDSEM、FasterEMA 改进 BasicBlock,在颈部网络中使用 CARAFE 和 Bi-FPN 模块改进。本文在自建数据集上围绕基线模型 RT-DETR-R18 进行消融实验,实验结果如表 5 所示,其中 A 代表 SDSEM,B 代表 FasterEMA,C 代表 CARAFE+BiFPN 使用情况。

表 5 消融实验

Table 5 Ablation experiments

模型	A	B	C	测试集 1		测试集 2		测试集 3		Params/ ($\times 10^6$)	GFLOPs	FPS/ fps
				mAP ₅₀ /	mAP _{50,95} /	mAP ₅₀ /	mAP _{50,95} /	mAP ₅₀ /	mAP _{50,95} /			
				%	%	%	%	%	%			
Base				75.3	43.4	76.0	44.7	73.3	42.4	20.2	58.6	27
Base-A	✓			77.4	45.1	78.1	46.6	74.8	43.1	20.7	68.3	26
Base-B		✓		76.8	44.3	77.9	45.9	73.2	42.4	17.3	55.1	26
Base-C			✓	77.2	44.9	79.2	47.0	74.0	43.1	20.7	66.9	20
Base-AC	✓		✓	76.9	44.9	78.7	46.6	75.2	43.2	21.3	76.6	19
Base-AB	✓	✓		75.4	43.6	79.2	46.9	74.4	43.1	17.8	64.8	26
Base-BC		✓	✓	77.1	43.7	77.7	46.4	74.5	42.9	17.8	63.5	20
EBCS-DETR	✓	✓	✓	78.6	45.4	79.4	47.5	76.4	44.2	18.4	73.1	19

由表 5 可知,仅使用 SDSEM 模块改进的 Base-A,虽然 Params 和 GFLOPs 与基线模型相比分别上升了 2.5% 和 16.6%,但是帧率只下降了 1 fps,并且在 3 个测试集上的检测精度 mAP_{50} 分别提升了 2.1%、2.1% 和 1.5%,表明 SDSEM 模块的改进是显著有效的。Base-B 利用 FasterEMA 改进深层的 BasicBlock,Params 和 GFLOPs 分别下降了 14.4% 和 6.0%,但是帧率没有上升,这是由于 EMA 注意力机制提升了内存的访问量,尽管该优化对模型整体复杂度的绝对数值影响相对有限,但在前两组的测试集上 mAP_{50} 分别提升了 1.5% 和 1.9%,有效提升了检测精度,同时降低了模型的资源消耗,为进一步的轻量化部署奠定了基础。Base-C 仅改进颈部网络,虽然相比基线模型帧率下降了 6 fps,但 mAP_{50} 显著提升,尤其测试集 2 提升了 3.2%,这表明 CARAFE 与 Bi-FPN 能增强模型对小目标缺陷的定位与区分能力。Base-AC 将 SDSEM 和颈部网络的修改相配合,参数量和计算量进一步上升,达到了这 8 个模型的最高,帧率下降到 19 fps,但是相对 Base-C 在测试集 1 和 2 中并没有提升,仅在测试集 3 中 mAP_{50} 提升了 1.2%,在仔细对比中发现 Base-AC 对黑斑的检测能

力相对较弱,说明 SDSEM 直接与颈部网络的改进耦合,性能不是很稳定。当使用 Base-AB 时帧率和 Base-A、Base-B 是一样的,其帧率不变的原因是 SDSEM 和 FasterEMA 的运算量基本上与 BasicBlock 一致,此外检测精度在测试集 2 和 3 均有明显提升。Base-BC 相较于 Base-B 多了颈部网络的修改,检测精度有所提升,尤其测试集 3 提升了 1.3%,表明 CARAFE 和 Bi-FPN 能与 FasterEMA 有效耦合。本文的改进算法在 3 个测试集中的表现是 8 个模型中最好的,分别在 mAP_{50} 提升了 3.3%、3.4% 和 3.1%,相比原型 Param 下降了 8.9%,GFLOPs 上升了 24.9%,帧率下降到 19 fps,但这在实时检测的允许范围内。

2.8 对比实验

为了进一步验证所提出的 EBCS-DETR 模型在检测精度和泛化性能方面的优势,本研究将另外 17 个先进检测模型的总体性能与 EBCS-DETR 进行比较。对比模型包括基于 CNNs 的模型和基于 Transformer 的模型。为保证公平性,比较模型基本选择了与 EBCS-DETR 相近的参数或 GFLOPs。所有对比模型在 3 个测试集中取得的检测结果如表 6 所示。

表 6 主流算法对比实验
Table 6 Comparison results of different models

类别	模型	测试集 1		测试集 2		测试集 3		Params/ ($\times 10^6$)	GFLOPs	FPS/ fps
		mAP_{50} /	$mAP_{50:95}$ /	mAP_{50} /	$mAP_{50:95}$ /	mAP_{50} /	$mAP_{50:95}$ /			
		%	%	%	%	%	%			
CNNs- based models	YOLOv5m	72.5	41.9	72.9	43.100	68.0	39.5	25.1	64.6	62
	YOLOv8m	72.1	41.7	72.7	42.900	68.7	39.1	25.9	79.3	53
	YOLOv9m	72.8	41.3	72.0	42.300	68.6	38.9	20.2	77.9	49
	YOLOv10m	68.4	40.3	69.7	42.200	66.0	38.4	16.6	64.5	53
	YOLO11m	72.4	41.9	72.0	43.600	67.9	40.0	20.1	68.5	51
	YOLO12m	72.3	42.2	74.3	44.300	67.8	39.9	20.2	68.1	45
	YOLOF	67.2	37.5	65.5	38.100	64.6	35.8	44.2	78.5	27
	YOLOX	76.2	43.0	78.8	45.700	75.0	42.1	25.3	73.8	21
Transformer- based models	PPYOLOE+ L	78.0	45.8	79.4	0.481	75.6	43.5	52.2	110.7	30
	EfficientNetV2	75.6	44.3	74.9	45.2	72.6	43.0	28.7	71.0	21
	RepViT	73.0	42.0	74.3	44.4	71.2	41.2	22.6	62.5	20
	SwinTransformer	72.5	41.8	72.1	42.9	69.7	41.5	36.6	98.1	15
	DINO	72.9	43.5	74.5	46.8	71.4	41.8	47.5	279.6	5
	Deformable DETR	72.5	40.8	72.8	41.6	70.1	38.3	40.0	173.0	9
	RT-DETR-R18	75.3	43.4	76.0	44.7	73.3	42.4	20.2	58.6	27
	RT-DETR-R34	75.4	43.9	76.7	44.9	72.1	42.1	30.3	88.9	23
	RT-DETR-R50	76.4	43.0	79.4	46.6	75.3	42.6	42.9	134.8	12
	EBCS-DETR	78.6	45.4	79.4	47.5	76.4	44.2	18.4	73.1	19

整体来看,CNN 架构模型普遍具备更高的检测速度,而 Transformer 架构模型则在检测精度方面表现更为优越。在裂缝占比最高的测试集 3 中,本文改进的 EBCS-

DETR 模型 mAP_{50} 和 $mAP_{50:95}$ 分别达到了 76.4% 和 44.2%,相比同样采用了 ResNet 作为主干网络的 PPYOLOE+ L,本文模型表现出优秀的泛化能力。RT-

DETR-R50 在众多的 Transformer 模型中性能仅次于本文提出的 EBCS-DETR,在 3 个测试集中本文模型相比 RT-DETR-R50 mAP_{50} 分别提高了 2.2%、0% 和 1.1%, $mAP_{50,95}$ 分别提高了 2.4%、0.9% 和 1.6%,但是 RT-DETR-R50 的 GFLOPs 约为 EBCS-DETR 的 2 倍,这也导致了 RT-DETR-R50 的帧率只有 12 fps。在 YOLO 系列的原型中,YOLO12m 的参数量级与计算复杂度都接近本文所提出的模型,但是从 3 个测试集的表现来看,本文模型相比 YOLO12m 在 mAP_{50} 提升了约 5%, $mAP_{50,95}$ 提升了约 2%。因此,在所有模型中,EBCS-DETR 表现出良好的检测性能和最佳的泛化性能,并且具有最低的参数量。

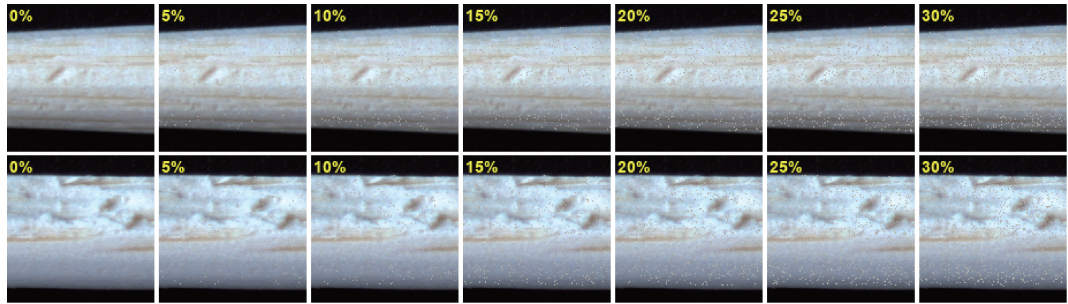


图 10 含不同比例盐噪声的稳健性检验样品

Fig. 10 Robustness test data with different proportion of salt noise

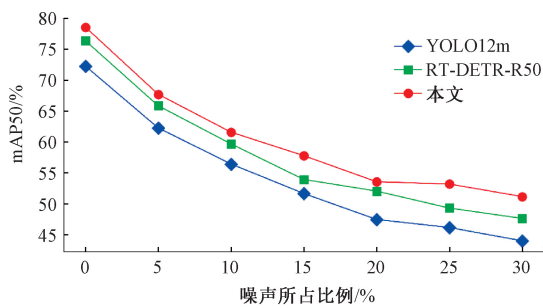
图 11 所示为 EBCS-DETR、RT-DETR-R50 和 YOLO12m 模型加噪后的测试结果。随着噪声比例的增加,所有模型都表现出检测性能的下降。在测试集 1 中,当比例噪声在 15% 以下时,EBCS-DETR 准确率的下降速率和 YOLO12m 的大致相同。但当比例噪声超过 15% 时,EBCS-DETR 的准确率下降速度慢于 YOLO12m。与 RT-DETR-R50 相比,EBCS-DETR

2.9 泛化分析实验

为评估模型 EBCS-DETR 在噪声干扰下的稳定性,选择模型表现优异的测试集 1(mAP_{50} 为 78.6%)和测试集 2(mAP_{50} 为 79.4%)进行实验,并引入模型 RT-DETR-R50 和 YOLO12m 作为对比。

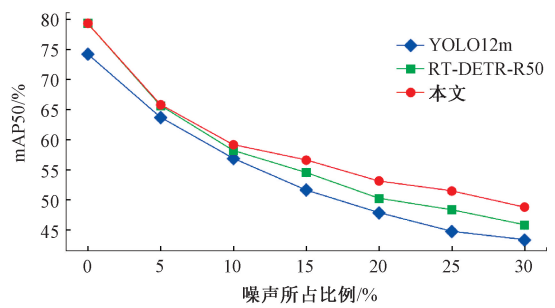
采用椒盐噪声模拟实际干扰,但为避免与竹篾缺陷(黑斑、毛刺)混淆,噪声设置为贴近竹篾材质的淡棕色,并限制噪声在竹篾区域内。对两个测试集,分别添加 5%~30% 的噪声(步长 5%),生成鲁棒性测试数据。图 10 所示为不同比例噪声下的图像示例,可见随着噪声比例增加,干扰显著增强。

在所有比例噪声水平下的精度更高。在测试集 2 中,当比例噪声低于 10% 时,EBCS-DETR 表现出比 YOLO12m 更快的退化速率,但与 RT-DETR-R50 基本相同,当比例噪声高于 10% 时,EBCS-DETR 的准确率下降速率比另外两个模型都慢。总体而言,EBCS-DETR 比 RT-DETR-R50 和 YOLO12m 表现出更好的稳定性。



(a) 测试集1

(a) Test set 1



(b) 测试集2

(b) Test set 2

图 11 不同模型在不同比例噪声的测试集 1 和测试集 2 中的表现

Fig. 11 Performance of different models in testset1 and testset2 with varied proportional noise

3 结 论

本文通过 EBCS-DETR 实现了对竹篾切片图像表面缺陷的实时准确识别。这种基于图像切片的检测方式,效率和准确率比整图检测更高。所设计的 SDSEM 模块,增强了 ResNet 浅层网络捕获细节特征的能力,并弥补浅

层网络对全局特征的理解能力。在 ResNet 深层引入的 FasterEMA 模块,有效降低了模型的参数量和计算复杂度并提升骨干网络特征提取效率。在颈部网络中的 CARAFE 上采样模块与 Bi-FPN 双向融合模块,优化了模型的小目标边界定位能力并有效抑制了背景噪声。在自建数据集上的实验表明,EBCS-DETR 具备优异的特征提

取与小目标边界区分能力,超越了现有规模相当的先进模型。泛化分析实验进一步表明该模型具备可靠性和应用潜力。

后续的研究将开发竹筷缺陷评估模块,对检测结果进行评分,在符合质检标准的条件下实现生产成本的节约。同时针对检测能力较弱的缺陷类型,将补充相应的数据集,以提高模型的检测水平。

参考文献

- [1] LIU C, CHEN K, WANG N C, et al. A lightweight multi-scale feature fusion method for detecting defects in water-based wood paint surfaces [J]. *Measurement*, 2025, 253: 117505.
- [2] WANG R J, LIANG F L, WANG B, et al. An efficient and accurate surface defect detection method for wood based on improved YOLOv8 [J]. *Forests*, 2024, 15(7): 1176.
- [3] SUN P, HUA C C, DING W L, et al. A real-time detection framework for surface defects in ceramic tableware based on deep learning [J]. *Expert Systems with Applications*, 2025, 286: 128101.
- [4] 王永香, 代亚男, 韦存茜, 等. 一次性筷子行业发展分析 [J]. *轻工科技*, 2019, 35(4): 135-136, 171.
WANG Y X, DAI Y N, WEI C Q, et al. Analysis of disposable chopsticks industry development [J]. *Light Industry Science and Technology*, 2019, 35(4): 135-136, 171.
- [5] 杨思念, 曹立佳, 杨洋, 等. 基于机器视觉的PCB缺陷检测算法研究综述 [J]. *计算机科学与探索*, 2025, 19(4): 901-915.
YANG S N, CAO L J, YANG Y, et al. Review of PCB defect detection algorithm based on machine vision [J]. *Journal of Frontiers of Computer Science and Technology*, 2025, 19(4): 901-915.
- [6] 陆畅, 李文举, 王旭彬, 等. 基于特征重构的工业图像无监督异常检测 [J/OL]. *电子测量技术*, 1-11 [2025-12-19]. <https://link.cnki.net/urlid/11.2175.tn.20250617.1107.013>.
LU CH, LI W J, WANG X B, et al. Unsupervised industrial anomaly detection based on feature reconstruction [J/OL]. *Electronic Measurement Technology*, 1-11 [2025-12-19]. <https://link.cnki.net/urlid/11.2175.tn.20250617.1107.013>.
- [7] 王俞霏, 杜弘志, 胡蕴博, 等. 基于多尺度特征融合网络的电阻点焊缺陷精确定位与分割方法 [J]. *仪器仪表学报*, 2025, 46(7): 202-213.
WANG Y F, DU H ZH, HU Y B, et al. Precise localization and segmentation method for resistance spot welding defects based on multi-scale feature fusion network [J]. *Chinese Journal of Scientific Instrument*, 2025, 46(7): 202-213.
- [8] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: Unified, real-time object detection [C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016: 779-788.
- [9] 冯哲, 王彬, 黄鹏程, 等. 小天体表面着陆区岩石目标检测算法 [J]. *仪器仪表学报*, 2024, 45(4): 194-205.
FENG ZH, WANG B, HUANG P CH, et al. Algorithm of detection rock object in landing zone of small celestial body surface [J]. *Chinese Journal of Scientific Instrument*, 2024, 45(4): 194-205.
- [10] REN S Q, HE K M, 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.
- [11] ZHANG W K, HONG Z Y, XIONG L Q, et al. Sinextnet: A new small object detection model for aerial images based on PP-Yoloe [J]. *Journal of Artificial Intelligence and Soft Computing Research*, 2024, 14(3): 251-265.
- [12] 赵云涛, 黄哲辉. 面向特钢车间内物料实时跟踪的钢管目标检测算法研究 [J]. *电子测量与仪器学报*, 2024, 38(11): 210-218.
ZHAO Y T, HUANG ZH H. Research on steel pipe target detection algorithm for real-time material tracking in special steel workshop [J]. *Journal of Electronic Measurement and Instrumentation*, 2024, 38(11): 210-218.
- [13] ZHAO Y, LYU W Y, XU S L, et al. DETRs beat YOLOs on real-time object detection [C]//IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2024: 16965-16974.
- [14] 朱黎颖, 王森, 沈爱萍, 等. 复杂背景下的电路板表面焊接缺陷视觉检测 [J]. *光学精密工程*, 2024, 32(14): 2256-2271.
ZHU L Y, WANG S, SHEN AI P, et al. Visual inspection of soldering defects on board surfaces against complex backgrounds [J]. *Optics and Precision Engineering*, 2024, 32(14): 2256-2271.
- [15] SUNKARA R, LUO T. No more strided convolutions or pooling: A new CNN building block for low-resolution images and small objects [C]//Machine Learning and Knowledge Discovery in Databases-European Conference, 2022: 443-459.
- [16] 杨瑞峰, 廖映华, 罗覃鹏, 等. 基于改进 YOLOv10 的

- 液晶显示屏表面缺陷检测算法[J]. 电子测量技术, 2025, 48(15): 70-79.
- YANG R F, LIAO Y H, LUO Q P, et al. Liquid crystal display surface defect detection algorithm based on improved YOLOv10[J]. Electronic Measurement Technology, 2025, 48(15): 70-79.
- [17] OUYANG D L, HE S, ZHANG G Z, et al. Efficient multi-scale attention module with cross-spatial learning[C]//ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing(ICASSP), 2023: 1-5.
- [18] LIANG J T, CHEN X, LIANG CH J, et al. A detection approach for late-autumn shoots of litchi based on unmanned aerial vehicle (UAV) remote sensing[J]. Computers and Electronics in Agriculture, 2023, 204: 107535.
- [19] TAN M X, PANG R M, LE Q V. EfficientDet: Scalable and efficient object detection [C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition(CVPR), 2020: 10778-10787.
- [20] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need [C]//31st International Conference on Neural Information Processing Systems, 2017: 6000-6010.
- [21] CARION N, MASSA F, SYNNAEVE G, et al. End-to-end object detection with transformers [C]//European Conference on Computer Vision, 2020: 213-229.
- [22] LI X X, CAI M R, TAN X J, et al. An efficient transformer network for detecting multi-scale chicken in complex free-range farming environments via improved RT-DETR[J]. Computers and Electronics in Agriculture, 2024, 224: 109160.
- [23] 张路, 魏本昌, 魏鸿奥, 等. 基于改进 RT-DETR 的水下目标检测[J]. 计算机系统应用, 2024, 33(12): 131-140.
- ZHANG L, WEI B CH, WEI H AO, et al. Underwater target detection based on improved RT-DETR[J]. Computer Systems & Applications, 2024, 33(12): 131-140.
- [24] HU J, SHEN L, SUN G. Squeeze-and-Excitation networks [C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018.
- [25] CHEN J R, KAO S H, HE H, et al. Run, don't walk: Chasing higher FLOPS for faster neural networks [C]//2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2023: 12021-12031.
- [26] WANG J Q, CHEN K, XU R, et al. CARAFE: Content-aware reassembly of features [C]//2019 IEEE/CVF International Conference on Computer Vision (ICCV), 2019: 3007-3016.
- [27] WANG S H, WANG Y D, CHANG Y J, et al. EBSE-YOLO: High precision recognition algorithm for small target foreign object detection[J]. IEEE Access, 2023, 11: 57951-57964.

作者简介

莫太平, 博士, 教授, 博士生导师, 主要研究方向为深度学习、图像处理等。

E-mail: mtp@guet.edu.cn

孙鹏(通信作者), 博士, 讲师, 硕士生导师, 主要研究方向为医学影像处理、目标检测、图像分割等。

E-mail: sunpeng@guet.edu.cn