

DOI:10.19651/j.cnki.emt.2518773

基于多尺度特征融合的钢缆小目标缺陷检测算法*

杨昌昌 张笃振 王斯豪

(江苏师范大学计算机科学与技术学院 徐州 221116)

摘要: 钢缆小目标缺陷检测任务中普遍存在检测精度低、漏检率高以及误检频发的问题,尤其在小尺寸较多的检测场景下表现尤为明显。造成该类问题的主要原因包括:传统检测算法特征提取能力不足,缺乏有效的多尺度信息融合机制,以及现有损失函数对小目标不敏感等。针对上述问题,提出一种基于改进型 RT-DETR 的钢缆缺陷检测方法。该方法在骨干网络中设计 BasicStar 特征提取模块,提升模型在高维空间中的语义表征能力;同时设计了新型多尺度特征融合策略小目标金字塔网络(SOPN),强化对小目标的关注和表达能力;在损失函数方面,提出焦点增强型 Focaler-SIoU 损失函数,以提升小目标定位精度与训练收敛稳定性。在钢缆缺陷数据集上的实验结果表明,该改进模型在平均检测精度 mAP50 上较原始 RT-DETR 提升了 2.1%。综合性能优于现有主流目标检测算法,验证了所提方法在工业场景下对小目标缺陷检测任务的有效性和实用性。

关键词: 目标检测;小目标;RT-DETR;特征融合;损失函数

中图分类号: TP391.4;TN911.73 **文献标识码:** A **国家标准学科分类代码:** 520.6040

Small target defect detection algorithm for steel cables based on multi-scale feature fusion

Yang Changchang Zhang Duzhen Wang Sihao

(School of Computer Science and Technology, Jiangsu Normal University, Xuzhou 221116, China)

Abstract: In the task of small target defect detection on steel cables, there are common problems such as low detection accuracy, high missed detection rate and frequent false detection, which are particularly obvious in the detection scenario with more small sizes. The main reasons for such problems include: insufficient feature extraction capability of traditional detection algorithms, lack of effective multi-scale information fusion mechanism, and insensitivity of existing loss functions to small targets. To address the above problems, a steel cable defect detection method based on improved RT-DETR is proposed. The BasicStar feature extraction module was designed in the backbone network to improve the semantic representation ability of the model in high-dimensional space; at the same time, a new multi-scale feature fusion strategy small object pyramid network(SOPN) is designed to strengthen the attention and expression ability of small targets; in terms of loss function, a focal enhancement Focaler-SIoU loss function is proposed to improve the positioning accuracy of small targets and the stability of training convergence. Experimental results on the steel cable defect dataset show that the improved model improves the average detection accuracy mAP50 by 2.1% compared with the original RT-DETR. The comprehensive performance is better than the existing mainstream target detection algorithms, which verifies the effectiveness and practicality of the proposed method for small target defect detection tasks in industrial scenarios.

Keywords: object detection; small target; RT-DETR; feature fusion; loss function

0 引言

钢缆作为一种关键承力构件,广泛应用于桥梁、索道等工程中,其运行安全性依赖于早期损伤的高效检测。然而,

钢缆损伤具有尺寸小、特征弱、位置分散等特点,传统检测手段在实时性、精度和鲁棒性方面存在明显不足。随着深度学习的发展,基于目标检测的自动化方法在钢缆缺陷检测中展现出广阔前景。

收稿日期:2025-05-09

* 基金项目:江苏省高等学校自然科学研究面上项目(19KJB520032)、江苏师范大学科研与实践创新项目(2024XKT2613)资助

当前目标检测算法主要分为基于 CNN^[1] 和基于 Transformer^[2] 的算法。其中, CNN 算法又包括双阶段与单阶段算法两类。双阶段算法如 R-CNN^[3]、Faster R-CNN^[4], 首先生成目标候选区域, 再利用卷积网络精确检测目标, 具有较高精度, 但检测速度较慢, 难以高效处理大规模数据集。单阶段算法如 SSD^[5]、YOLOv11^[6] 则将候选区域生成与目标检测合并到单一阶段, 网络结构简单, 检测速度快, 更适合大规模数据场景, 但在准确性、鲁棒性尤其是小目标检测能力上仍有不足。此外, 这些算法普遍采用非极大值抑制(non-maximum suppression, NMS) 进行后处理, 以抑制多余候选框, 但此方法存在运行效率低、易漏检等问题。基于 Transformer 的检测算法如 DETR^[7]、Deformable DETR^[8], 自提出以来备受关注, 具备全局感知优势, 然而计算复杂度高、拟合速度慢, 难以满足工业应用的实时性要求。

为解决上述问题, 本文以最新的 RT-DETR^[9] 模型为基础, 提出了一种基于改进模型的钢缆检测方法。

1) 设计特征提取模块 BasicStar, 引入深度可分离卷积, 利用其减少参数量和计算量, 强化局部空间特征提取的优势, 同时结合星型网络结构, 通过中心节点与周围节点的稀疏连接机制, 有效建模局部细节与全局关系, 提升特征在高维特征空间中的表达与区分能力。

2) 设计新型特征融合策略小目标金字塔网络(small object pyramid network, SOPN)。引入空间深度转换卷积模块, 利用其在空间特征变换与深度特征压缩方面的优势, 提升特征表达的跨尺度适应能力。同时设计 CrossOpt 子模块, 分别关注全局分支、大分支和局部分支特征。通过分支间协同优化, 有效学习从全局到局部的特征表征, 增强模型对小目标细粒度特征的感知能力。该策略能够提升特征融合的多尺度建模能力, 特别是小目标的感知能力, 增强模型对目标空间分布的敏感性和表征能力。

3) 提出融合改进的 Focaler-SIoU 损失函数。针对数据集中小目标数量较多的特点, 基于 Focaler-IoU 损失函数对困难样本的关注机制与 Siou 损失函数对定位偏差的优化特性, 融合两者优点, 构建了 Focaler-SIoU 损失函数。该方法在提高小目标检测精度的同时, 进一步增强了模型对小尺寸目标的回归能力, 显著提升整体检测性能。

改进后的模型在钢缆缺陷数据集上的检测精度相比原始 RT-DETR 算法提升 2.1%。特别适用于工业中包含大量小目标的钢缆缺陷检测任务。

1 相关工作

1.1 钢缆缺陷检测的进展

最初钢丝绳缺陷检测多依靠人工目测巡检, 虽然操作简单, 但受限于检测者经验, 效率低且精度不稳定。因此, 国内外研究者转向使用电磁、超声波、声发射等物理传感器检测方法, 通过捕获异常信号定位缺陷, 减少人为因素干

扰。但这些物理传感器方案在区分缺陷类型的精度、复杂噪声环境下的表现及设备成本方面仍存在不足, 难以广泛应用于大规模工业场景。

近年来, 随着深度学习在计算机视觉领域的迅猛发展, 钢丝绳缺陷检测迎来技术变革。卷积神经网络及其演进模型在工业缺陷检测中表现突出, 如刘航等^[10] 在 YOLOv8 基础上实现绝缘子自爆缺陷的轻量化高效检测; 吴葛等^[11] 利用 YOLOv10 检测钢材表面缺陷; 吴斌斌等^[12] 则将 RT-DETR 应用于 PCB 缺陷检测。上述研究表明, 深度学习模型具备较高的检测精度、稳定性及推广潜力, 为工业智能检测提供了新的技术路径和发展方向。

1.2 RT-DETR

RT-DETR 是一种基于 Transformer 的实时端到端目标检测模型, 通过解耦尺度内交互与跨尺度融合, 有效处理多尺度特征, 显著降低了 DETR 模型的计算成本, 速度与精度均优于同等规模的 YOLO 模型。

RT-DETR 的整体架构包括主干网络、混合编码器以及带辅助检测头的解码器。主干网络采用 ResNet^[13] 结构, 通过多层卷积神经网络提取图像特征, 并输出 3 个不同尺度的特征图, 为编码器提供丰富的多尺度输入信息。混合编码器由 AIFI 模块和 CCFM 组成。其中, AIFI 模块自适应调整不同特征间的交互方式, 以更好地捕捉特征间的依赖关系, 提升特征表达能力。CCFM 模块则通过自底向上和自顶向下双向路径进行跨尺度融合, 以提高特征表示的丰富性与准确性。

解码器部分, RT-DETR 首先利用 IoU 感知查询机制, 从编码器输出的序列中选取具有较高置信度的特征作为初始对象查询。随后通过迭代优化持续更新预测框的位置、尺寸以及置信度分数, 最终输出精确的检测结果。

2 本文方法

为了实现钢缆小目标缺陷检测的高精度与高效性, 基于 RT-DETR 模型, 提出一种改进的多尺度特征融合检测模型, 其架构如图 1 所示。该模型的核心改进由 3 部分构成: 主干网络部分设计 BasicStar 结构替换原始 BasicBlock 模块, 以增强高维特征映射能力并提升特征提取效果; 颈部特征融合网络设计新型 SOPN 策略, 有效缓解小目标语义信息丢失问题, 同时强化多尺度特征的上下文关联; 检测头部分提出改进后的 Focaler-SIoU^[14-15] 损失函数, 结合角度敏感性与困难样本聚焦机制, 在优化边界框回归精度的同时提升模型对小目标的定位鲁棒性。

2.1 特征提取模块 BasicStar

RT-DETR 模型主要由两个部分构成, 分别是作为 Backbone 的 ResNet 以及 RT-DETR 的检测头。一般来说, 深度的提升为模型带来了更强的表达能力和更丰富的信息利用途径, 从而有望取得更好的结果。但是在 VGG^[16] 上的实验发现单纯堆叠网络层数反而导致训练损

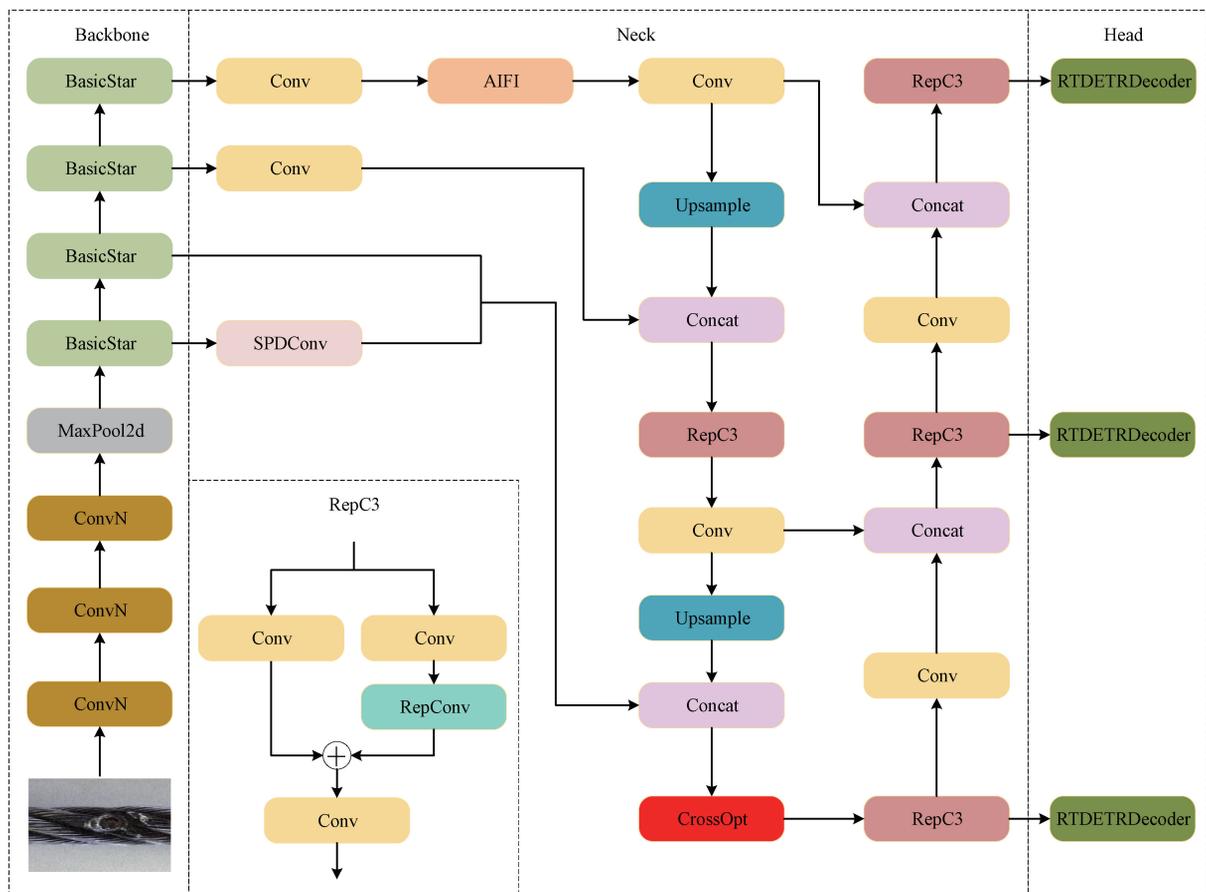


图 1 改进后的模型结构图

Fig. 1 Improved model structure diagram

失逐渐增高。何凯明认为深层的网络存在着梯度消失以及梯度爆炸的问题。即网络的层数越深,反向传播回来的梯度就越小,甚至接近于没有梯度,最后模型难以拟合。基于此现象 ResNet 增加了“短路”连接(shortcut connection),这种结构在残差神经网络中增加了一个从第 1 层连接至第 2 层激活函数前的跳跃连接,让更深的卷积层不会因为网络更深而产生退化的问题。

在 ResNet 架构中,BasicBlock 的第 2 个卷积操作通常使用 3×3 的卷积核,该设计在特征提取方面取得了显著的效果。然而,随着网络的加深和任务的复杂化,单纯的卷积操作面临着计算效率和表达能力的双重挑战。首先, 3×3 卷积虽然能够提取局部特征,但其计算量较大,尤其是在深层网络中,卷积层的参数数量和计算量呈指数级增长,导致训练和推理的速度受到影响。其次,传统卷积操作在处理多尺度信息时存在一定的局限性,无法有效捕捉图像中不同尺度的特征关系,进而影响了网络对复杂任务的适应性和泛化能力。

为了解决该问题,本文设计 BasicStar 模块,用来优化传统卷积操作。其整体结构如图 2 所示。BasicStar 首先通过 DWConv^[17]进行初步提取特征层的特征。与传统卷

积相比,DWConv 将标准卷积操作分解为两个独立的步骤:首先进行逐通道卷积,每个输入通道单独进行卷积操作;然后进行逐点卷积,即使用 1×1 卷积来组合这些逐通道卷积的输出。这种操作显著降低了计算复杂度和参数数量,特别是在输入通道数较大的情况下,能够大幅度减少模型的计算负担和存储开销,同时保留了网络的特征提取能力。

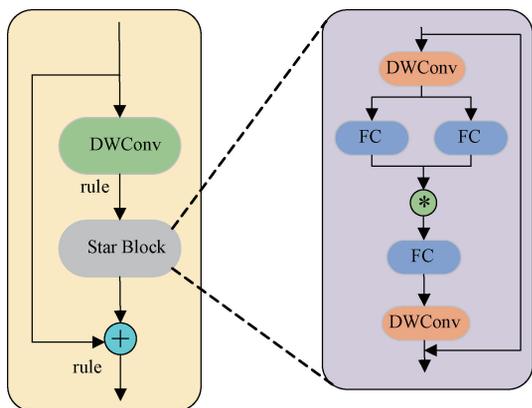


图 2 BasicStar 结构图

Fig. 2 BasicStar structure diagram

然后经过 ReLU 激活函数后输入至 StarBlock 模块再次降低模型复杂度,并且通过提升模型将输入映射到高维特征空间的能力,进而提升模型的特征提取能力。StarBlock 模块具体操作为首先采用深度可分离卷积 DWConv 进行特征的初步提取,然后对其进行批量归一化处理,接着将归一化处理后的特征进行星型操作,以及再次进行 DWConv 提取特征且将 DWConv 提取后的特征和原始的特征进行拼接。星型操作通过逐元素乘法融合两个线性变换的特征,计算公式如式(1)所示。

$$\begin{aligned} Output_{Star} &= \mathbf{w}_1^T X \times \mathbf{w}_2^T X = \left(\sum_{i=1}^{c+1} \omega_1^i x^i \right) \times \left(\sum_{j=1}^{c+1} \omega_2^j x^j \right) = \\ & \sum_{i=1}^{c+1} \sum_{j=1}^{c+1} \omega_1^i \omega_2^j x^i x^j = \alpha_{(1,1)} x^1 x^1 + \dots + \alpha_{(1,5)} x^1 x^5 + \dots + \\ & \alpha_{(c+1,c+1)} x^{c+1} x^{c+1} \end{aligned} \quad (1)$$

式中: $Output_{Star}$ 为星型操作后的结果, $\mathbf{w}_1, \mathbf{w}_2$ 为权重矩阵, $x \in R^{(c+1) \times 1}$, 其中 c 为输入通道数, i, j 为通道的下标, α 为各子项的系数,具体的定义如式(2)所示。

$$\alpha_{(i,j)} = \begin{cases} \omega_1^i \omega_2^j, & i = j \\ \omega_1^i \omega_2^j + \omega_1^j \omega_2^i, & i \neq j \end{cases} \quad (2)$$

星型运算最终可以展开有 $\frac{(c+2)(c+1)}{2}$ 种不同的式

子组合,约等于 $\left(\frac{c}{\sqrt{2}}\right)^2$ 的隐式维度特征空间。

BasicStar 模块最后将经过 DWConv 和 StarBlock 变换后的特征以及经过 ReLU 激活函数后的原始特征层进行拼接,输入给下一个特征层进行进一步的特征提取。

2.2 针对小目标的特征融合策略 SOPN

在 RT-DETR 模型中,虽然特征金字塔网络(feature pyramid network, FPN)已经能够在一定程度上处理不同尺度的特征,但对于深层特征的处理复杂性以及小目标检测能力仍有提升空间。通常针对小目标提升方法都是增

加一层 P2 检测层,但增加后会出现计算量过大、后处理更加耗时等问题。

因此,本文设计了一种新的改进模块 SOPN,目的是使每个尺度的特征都具备更详细、更丰富的上下文信息,从而提升模型的整体准确率和鲁棒性。SOPN 的结构图如图 3(a)所示。首先将 C2 特征层经过 SPDCConv^[18]操作,在特征层 $X(S, S, C)$ 进行下采样且设置下采样尺度 $scale=2$,特征层被分割为 4 个相同尺寸的子特征层,子特征层实现了对特征层 2 倍下采样,下采样层中包含了原图全局空间特征信息,增加了模型特征提取能力;在通道维度上对上述 4 个相同尺寸的子特征层进行拼接操作,获得中间特征层 $X'(\frac{S}{2}, \frac{S}{2}, 4C_1)$,通过卷积下采样来适应后续的拼接操作。具体映射如式(3)~(6)所示。

$$f_{(0,0)} = X[0:S:scale, 0:S:scale] \quad (3)$$

$$f_{(1,0)} = X[1:S:scale, 0:S:scale] \quad (4)$$

$$f_{(0,1)} = X[0:S:scale, 1:S:scale] \quad (5)$$

$$f_{(1,1)} = X[scale-1:S:scale-1, scale-1:S:scale] \quad (6)$$

接着将经过 SPDCConv 操作的特征层融入到 C3 特征层中。然后将融合后的特征层经过 CrossOpt 模块,如图 3(b)所示。CrossOpt 由 Conv 和 3 个分支组成,即全局分支、大分支和局部分支,这种设计可以有效地学习从全局到局部的特征表征,从而提高小目标的检测性能。CrossOpt 具体操作为将输入特征层以 0.25 的概率经过全局分支、大分支和局部分支操作,剩下的 0.75 的概率直接进行拼接操作,这种设计平衡了计算效率和模型性能,确保在处理复杂场景时,不会因为过度关注局部特征而导致计算量的急剧增加。通过这种设计,SOPN 特征融合在处理复杂背景下的小目标检测时,展现出更强的鲁棒性和精准度。

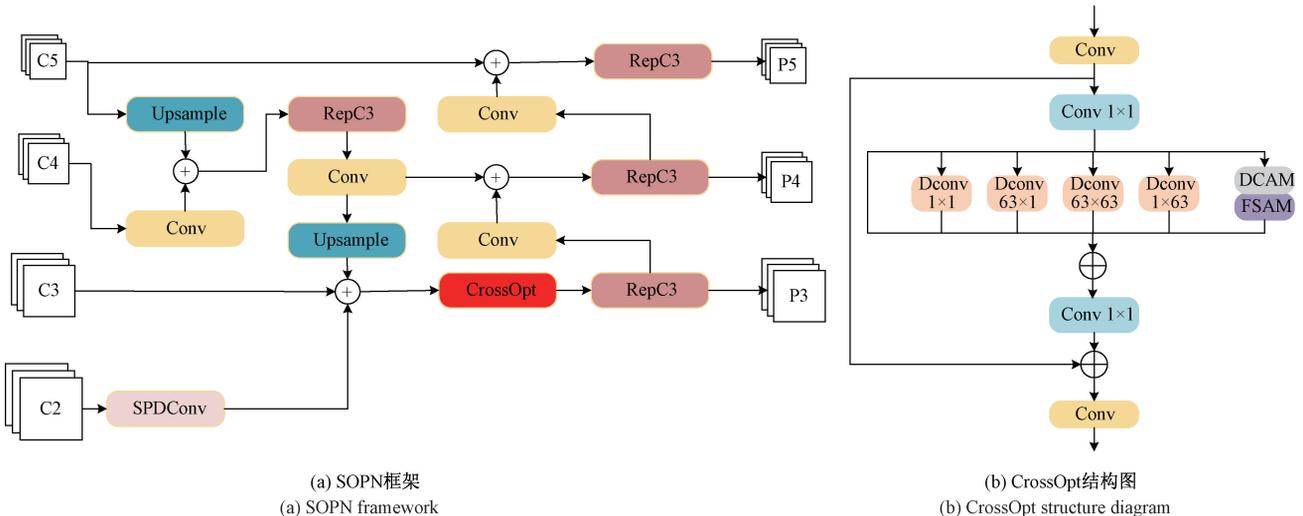


图 3 SOPN 结构图

Fig. 3 SOPN structure diagram

2.3 Focaler-SIoU 损失函数

在目标检测任务中,损失函数经常用于度量神经网络预测信息与期望标签的距离,预测信息越接近期望信息,损失函数值越小。在 RT-DETR 中使用 GIoU^[19],GIoU 不仅考虑了 IoU,还引入了两个边界框的外接最小矩形的概念,以此来衡量两个边界框之间的重叠程度和相对位置。但是 GIoU 对每个预测框与真实框均要计算最小外接矩形,从而计算和收敛的速度受到限制;并且在两个预测框完全重叠的情况下,不能反映出实际的情况,这时 GIoU 就会退化为 IoU,这就会在检测过程中产生一定的问题。

而 SIoU^[20]充分考虑到了真实框与预测框之间的向量角度问题,引入了角度损失,SIoU 相较于 GIoU,预测框回归只要确定了角度,即可沿着一条固定的线路回归,这可以有效地解决算法在训练过程中预测框即预测框在实际框周围震荡的问题,显著提升算法模型的收敛速度。SIoU 如图 4 所示。

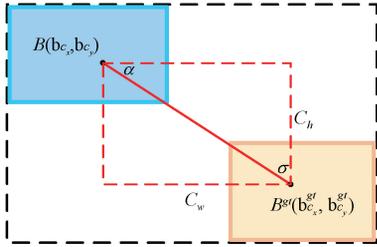


图 4 SIoU 工作示意图

Fig. 4 SIoU working diagram

式(7)~(10)为角度损失:

$$c_h = \max(b_{c_y}^{gt}, b_{c_y}) - \min(b_{c_y}^{gt}, b_{c_y}) \quad (7)$$

$$c_w = \max(b_{c_x}^{gt}, b_{c_x}) - \min(b_{c_x}^{gt}, b_{c_x}) \quad (8)$$

$$\sigma = \sqrt{(b_{c_x}^{gt} - b_{c_x})^2 + (b_{c_y}^{gt} - b_{c_y})^2} \quad (9)$$

$$\Delta = \cos\left(2 \times \left(\arcsin\left(\frac{c_h}{\sigma}\right) - \frac{\pi}{4}\right)\right) \quad (10)$$

其中, $b_{c_x}^{gt}$ 、 $b_{c_y}^{gt}$ 、 b_{c_x} 、 b_{c_y} 分别为真实框中心坐标及预测中心坐标, c_h 、 c_w 、 σ 分别为真实框和预测框中心点的高度差、宽度差及距离长度, Δ 为角度损失, 惩罚偏离对角线方向的角度误差。

式(11)~(16)为形状损失:

$$W_w = \frac{|\tau w - \tau w^{gt}|}{\max(\tau w, \tau w^{gt})} \quad (11)$$

$$W_h = \frac{|h - h^{gt}|}{\max(h, h^{gt})} \quad (12)$$

$$\Omega = (1 - e^{-W_w})^\theta + (1 - e^{-W_h})^\theta \quad (13)$$

$$\rho_x = \left(\frac{b_{c_x}^{gt} - d_{c_x}}{c_w}\right)^2, \rho_y = \left(\frac{b_{c_y}^{gt} - d_{c_y}}{c_h}\right)^2 \quad (14)$$

$$\gamma = 2 - \Delta \quad (15)$$

$$\Delta = 2 - e^{-\gamma \rho_x} - e^{-\gamma \rho_y} \quad (16)$$

其中, W_w 和 W_h 代表宽度和高度的归一化误差, 用于

衡量形状偏差; Ω 代表总体形状损失, θ 为控制损失曲线的超参数, 本文设置为 4; ρ_x 和 ρ_y 代表中心点偏移的归一化平方; γ 为调整惩罚强度的参数; Δ 代表位移惩罚项, 用于惩罚预测框中心与真实框中心之间的距离。

最终 SIoU 的损失定义如式(17)所示。

$$L_{SIoU} = 1 - IoU + \frac{\Delta + \Omega}{2} \quad (17)$$

GIoU 主要通过最大化预测框与真实框的整体 IoU 来衡量目标的检测精度, 而在小目标的场景中, 目标框的微小位置偏差难以通过 IoU 的变化体现, 导致其定位不够精确。其次, GIoU 在计算损失时, 对负样本的优化较为广泛, 这会使得模型在训练过程中更多关注背景区域, 而忽视了小目标的检测。背景噪声的干扰使得模型难以有效学习小目标的特征, 进一步降低了小目标的检测精度。为了解决这些问题, Focaler-IoU^[21] 损失函数被提出, 并且在小目标检测中展现出了显著的优势。Focaler-IoU 通过根据样本难度调整它们对损失函数的贡献来强调训练期间的困难样本, 优先加大对小目标和困难样本的关注, 使得模型在训练过程中能够更加专注于这些难以检测的目标, 进而提升其检测精度。具体公式如式(18)所示。

$$IoU^{focaler} = \begin{cases} 0, & IoU < d \\ \frac{IoU - d}{u - d}, & d \ll IoU \ll u \\ 1, & IoU > u \end{cases} \quad (18)$$

其中, d 、 u 分别为下限阈值和上限阈值。当 $IoU < d$ 时, 认为预测质量较差, 予以忽略; 当 $IoU > u$ 时, 认为预测质量较高, 赋予最大权重。本文中, d 设置为 0, u 设置为 0.95。

SIoU 通过引入角度损失约束预测框的回归方向, 减少训练中的边界框震荡现象, 同时利用形状损失精细化调整预测框与真实框的宽高匹配, 显著提升了定位精度; 然而, 其对所有样本赋予等权重的设计容易导致模型在复杂场景, 如小目标密集或背景干扰严重时被大量简单样本的损失所主导, 难以有效优化困难样本。为此, Focaler-IoU 通过动态权重分配机制, 依据 IoU 阈值 (d 和 u) 区分样本难度, 优先放大低 IoU 目标或小目标的损失贡献, 从而强化模型对困难样本的关注。二者的融合策略既保留了 SIoU 的几何约束优势, 确保回归路径的稳定性与定位精度, 又借助 Focaler-IoU 的样本筛选能力, 抑制背景噪声干扰并聚焦关键目标。具体的计算公式如式(19)所示。

$$L_{Focaler-SIoU} = L_{SIoU} + IoU - IoU^{Focaler} \quad (19)$$

3 实验结果分析

3.1 实验环境

本文的实验环境采用 Linux 操作系统, 具体配置如下: 中央处理器为 18 核的 AMD EPYC 9754 128-Core Processor, 具有强大的计算能力。图形处理单元 (GPU) 为 NVIDIA

RTX3090,显存容量为24 GB,确保了在大规模数据和复杂模型训练中的高效性能。CUDA版本为12.4。实验过程中,未使用任何预训练模型的权重,完全从头开始训练模型,确保实验结果的纯粹性和可靠性。

为了充分验证模型的性能,训练设置如下:采用AdamW优化器,超参数设置为 $\beta_1 = 0.9, \beta_2 = 0.999$,权重衰减为0.05。初始学习率为 2×10^{-4} ,其中Backbone的学习率为主干以外部分的0.1倍。训练轮次(epoch)设置为200次,以确保模型能够充分学习数据中的特征。学习率采用Step Decay策略,在第40和第50个epoch时分别衰减10倍,确保训练过程中模型收敛稳定且具备良好的泛化能力。批量大小(batch size)为16,这一设置在保证计算效率的同时,也考虑了GPU显存的使用情况。输入图片的尺寸统一调整为 640×640 ,以标准化输入数据,确保在训练和测试过程中模型能够一致地处理数据。

3.2 数据集介绍

本文引进数据集DBTT Computer Vision Project进行训练。数据集中包含两种训练类型:断裂、烧毁的问题钢缆,一共包括了3381张图像,并且按照70%、10%、20%的比例划分为训练集、验证集、测试集。两种钢缆类型如图5(a)、(b)所示。

在实际钢缆缺陷检测的工业场景中,检测任务面临诸多挑战,最常见的问题主要包括以下两类:首先,同一检测场景在不同时间段存在明显差异,尤其是在光照条件不佳或灯光不足的情况下,图像中缺陷区域的可见度大幅降低,导致目标难以识别,从而造成较高的误检率。针对这一问题,本文在数据预处理阶段引入了图像模糊增强技术,通过对原始图像施加一定程度的运动模糊,模拟真实场景中摄像模糊或视觉模糊情况,从而提升模型在复杂视觉条件下的鲁棒性和泛化能力。此外,为进一步增强模型对拍摄角度变化和图像倾斜的适应性,本文还引入了随机旋转增强,通过对图像进行小幅度的随机旋转,模拟钢缆在实际拍摄中产生的倾斜或姿态变化。具体增强如图5(d)、(f)所示。

3.3 模型评价指标

在实验中,采用的是目标检测任务中常用的评估指标:精度(Precision)、召回率(Recall)、mAP50(mean Average Precision)。其中精度(Precision)、召回率(Recall)计算方式如式(19)、(20)所示。

$$P = \frac{TP}{TP + FP} \quad (20)$$

$$R = \frac{TP}{TP + FN} \quad (21)$$

其中,TP(true positives)为模型检测为正样本且实际检测也为正样本的结果,FP(false positives)为模型检测为正样本但实际为负样本的结果,TN(true negatives)为模型检测为负样本且实际也为负样本的结果。

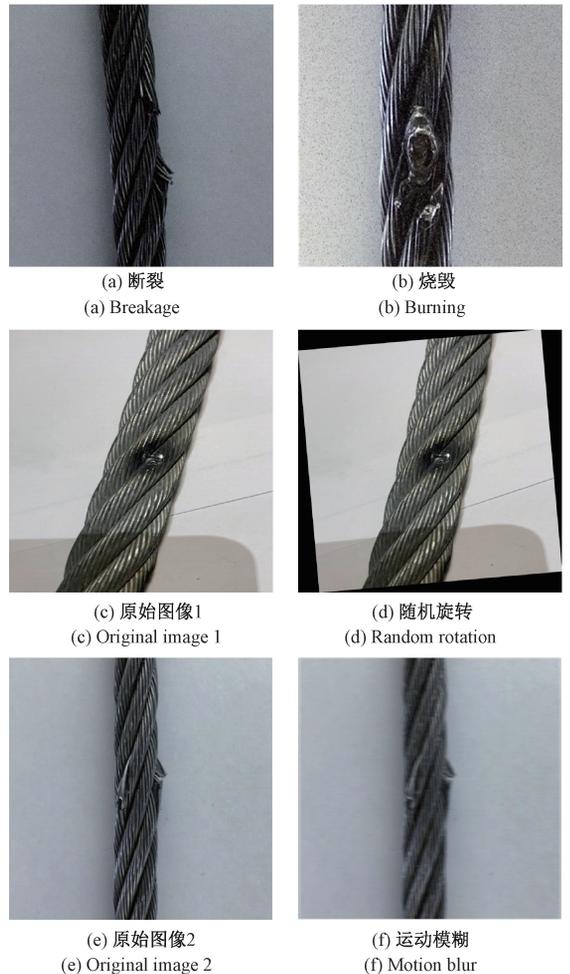


图5 数据集两种类型以及运动模糊与随机旋转增强效果
Fig. 5 Two types of datasets and enhanced effects of motion blur and random rotation

3.4 消融实验

为了验证提出的3种改进方法的有效性,共设计6组对照试验,结果如表1所示。首先本文以原始RT-DETR算法为基础,逐步引入一种改进方法,以单独评估每个改进方法的贡献。结果显示,引入BasicStar模块后,mAP50从83.5%提升至83.9%,说明其对特征提取能力的增强具有明显效果;引入SOPN特征融合策略后,mAP50提升至84.3%,参数量略增,说明其提升了小目标检测的能力但略微牺牲了速度;引入Focaler-SIoU损失函数后,mAP50提升至84.5%,在保持精度的同时提升了回归鲁棒性。然后,本文设计了组合改进的验证实验。当BasicStar与SOPN联合使用时,mAP50提升至84.8%;BasicStar与Focaler-SIoU组合时,mAP50达到84.9%,表现更为优异;SOPN与Focaler-SIoU组合时,mAP50提升至85.1%。

最后,本文将3种改进方法全部引入,模型的综合性能达到了最优,P为90.0%,R为82.9%,mAP50提升至

表 1 消融实验

Table 1 Ablation experiment

Baseline	BasicStar	SOPN	Focaler_SIoU	P/%	R/%	mAP50/%	Params/M	FPS/bs=1
✓				85.9	79.7	83.5	19.9	71.2
✓	✓			87.8	79.9	83.9	20.1	62.6
✓		✓		88.9	79.8	84.3	20.5	53.0
✓			✓	88.2	82.2	84.5	19.9	60.1
✓	✓	✓		88.5	82.2	84.8	20.7	58.1
✓	✓		✓	89.3	81.3	84.9	20.1	62.4
✓		✓	✓	86.7	81.3	85.1	20.5	65.5
✓	✓	✓	✓	90.0	82.9	85.6	20.7	56.9

85.6%，相较于原始 RT-DETR 算法分别提升了 4.1%、3.2%、2.1%。尽管参数量略增至 20.7 M，FPS 下降至 56.9，但检测精度和鲁棒性均实现了显著提升，验证了所提出改进方法的有效性和协同优势。

3.5 对比实验

为进一步分析本文改进模型在钢缆缺陷检测任务中的性能，将本文所提出的改进钢缆缺陷检测模型与 DINO^[22]、Faster RCNN、TOOD^[23]、DDQ^[24]、YOLOv8 主流目标检测模型进行了详细对比。所有模型均使用相同的数据集进行训练，并使用一致的训练参数和方法，以确保对比结果的公平性和可靠性。本文选取了 mAP50、mAP50:95、Params 和 GFLOPs 值作为主要评价指标，以全面评估各模型的性能。实验结果如表 2 所示，在略增参数量和计算量的情况下，本文改进的模型算法在 mAP50 和 mAP50:95 指标上远优于其他的模型算法。特别是，该模型在保持高精度的同时，符合实时检测的目标要求，展示了其在实际工业应用中的巨大潜力。通过这些对比实验，本文改进模型在钢缆缺陷检测任务中表现出了最优的综合性能，证明其有效性和实用性。

表 2 对比实验

Table 2 Comparative experiment

模型	mAP50/ %	mAP50:95/ %	Params/ M	GFLOPs
DINO	81.0	34.6	47.5	179
Faster RCNN	77.5	30.7	41.4	134
TOOD	75.3	30.0	32.0	123
DDQ	80.9	33.2	48.3	179
YOLOv8	79.7	32.8	3.0	8.1
YOLOv11	75.8	30.9	2.6	6.3
文献[25]	84.7	—	8.3	10.9
文献[26]	78.2	—	2.74	8.0
RT-DETR	83.5	38.2	19.9	56.9
本文算法	85.6	39.2	20.7	66.4

3.6 不同特征融合策略对比实验

为了进一步验证 SOPN 特征融合策略的有效性，本文设计了对比实验，将 SOPN 与其他几种常见的特征融合策略(如 BIFPN^[27]、GOLD-YOLO^[28]、HSFPN^[29]和 MAFPN^[30])进行了性能对比。实验结果表明，本文提出的 SOPN 特征融合模块在 mAP50 指标上达到 84.3%，分别超过了 BIFPN、GOLD-YOLO、HSFPN 和 MAFPN 的 83.3%、83.8%、84.2%、83.6% 等特征融合策略。综合来看，SOPN 特征融合模块在各项指标上均展现出最佳的综合性性能，证明了其在钢缆缺陷检测任务中的有效性和优越性。具体结果如表 3 所示。为了更直观反映本文模型的特征融合策略相较于其他特征融合策略的优越性，图 6 可视化展示了各个特征融合策略在数据集上面的检测结果。图 6 中第 1 行表示的是各个特征融合策略在断裂小目标情况下的表现。第 2 行表示的是各个特征融合策略在烧毁小目标情况下的表现。第 3 行表示的是模型在光线条件较差的情况下各个特征融合策略的表现。结果表明，SOPN 在小目标较多的情况下，准确率均高于其他的特征融合策略。

表 3 不同特征融合策略对比实验

Table 3 Comparative experiment of different feature fusion strategies

特征融合策略	P/%	R/%	mAP50/%
BIFPN	88.3	76.4	83.3
GOLDYOLO	89.6	78.5	83.8
HSFPN	88.9	81.1	84.2
MAFPN	89.3	77.2	83.6
SOPN	88.9	79.8	84.3

3.7 不同损失函数对比实验

为了验证本文改进的 Focaler-SIoU 损失函数的有效性，现选择了几种常见的损失函数进行对比实验，包括 Siou、Focaler-IoU、Wise_MPDIoU^[31-32]、Focaler-CIoU^[33]、Focaler-DIoU^[34]和 Focaler-GIoU^[35]。实验结果表明，本文

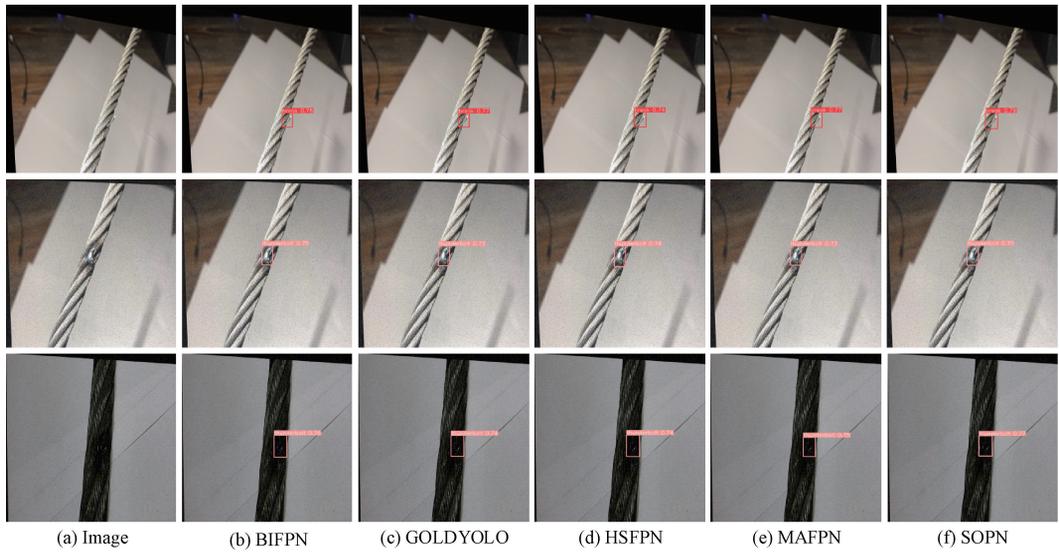


图6 不同特征融合策略在数据集上的检测结果图比较

Fig. 6 Comparison of detection results of different feature fusion strategies on the dataset

提出的 Focaler-SIoU 损失函数在 mAP50 指标上表现最优,达到了 84.5%,超过了其他损失函数的表现。通过对比实验,验证了 Focaler-SIoU 损失函数在钢缆缺陷检测任务中的有效性和优越性,证明了其在复杂工业场景中的实用价值。具体结果如表 4 所示。为了更直观反映本文模型所改进的损失函数表现优于其他损失函数,图 7 可视化展示了不同损失函数在数据集上面的检测结果。图中第 1 行和第 2 行表示的是不同损失函数在面对不同小目标时的准确率。第 3 行表示的是不同损失函数在面对小目标较多情况下的漏检情况。结果表明,本文所改进的 Focaler_SIoU 不存在漏检情况且准确率均高于其他损失函数。

表 4 不同损失函数对比实验

Table 4 Comparative experiments on different loss functions

损失函数	P	R	mAP50
SIoU	87.1	81.9	83.9
Focaler	87.5	81.0	84.2
Wise_MPDIoU	87.1	81.9	83.9
Focaler_CIoU	88.8	80.7	83.0
Focaler_DIoU	87.6	80.0	82.9
Focaler_GIoU	87.6	80.4	83.8
Focaler_SIoU	88.2	82.2	84.5

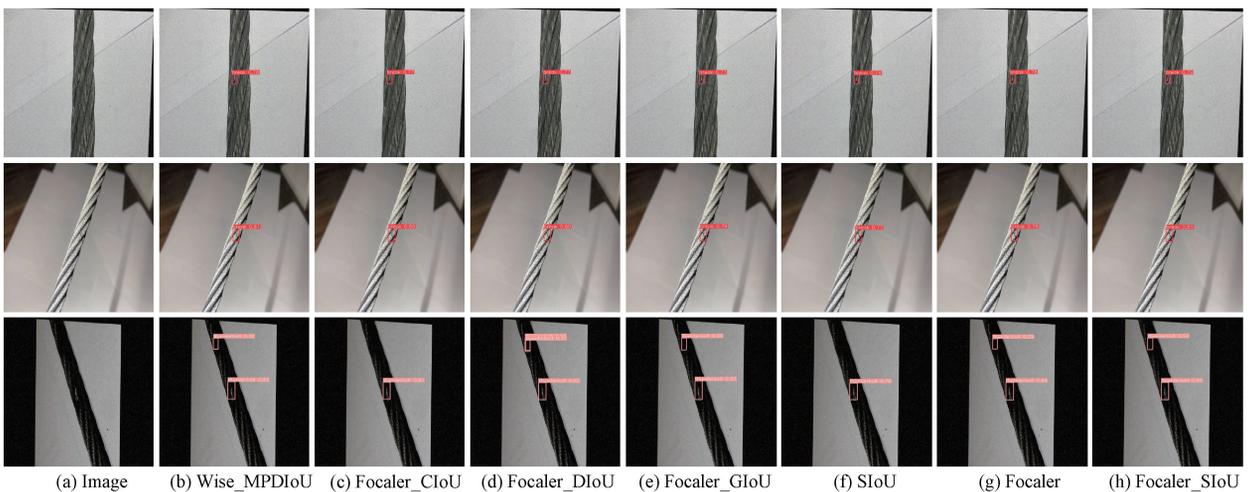


图7 不同损失函数在数据集上的检测结果图比较

Fig. 7 Comparison of detection results of different loss functions on the dataset

3.8 模型泛化与鲁棒性分析

为进一步验证本文改进模型的泛化能力,在一个包含 1 800 张图像的 Cable Damage Computer Vision Project 数据集上进行了实验。该数据集涵盖了 3 种常见的钢缆缺陷类型:损坏、电击和磨损。实验环境与 3.1 节中描述的配置保持一致,以确保实验条件的一致性。具体实验结果如表所示,展示了 RT-DETR 模型与本文改进算法在精度、召回率、mAP50 和 mAP50:95 等指标上的表现。

实验结果表明,本文改进算法在各项指标上均优于原始 RT-DETR 模型。具体来看,改进后的模型在精度上达到了 86.5%,召回率为 80.8%,mAP50 和 mAP50:95 分别为 86.6% 和 42.7%,均高于原始 RT-DETR 模型的 85.4%、81.0%、86.2% 和 41.4%。这些结果证明了本文改进算法在不同类型钢缆缺陷检测任务中的良好泛化能力,能够有效应对多种复杂场景下的检测需求。结果如表 5 所示。

表 5 泛性分析实验

Table 5 Pansexuality analysis experiment %

模型	P	R	mAP50	mAP50:95
RT-DETR	85.4	81.0	86.2	41.4
本文算法	86.5	80.8	86.6	42.7

此外,为进一步评估模型在真实工业场景中的鲁棒性表现,本文对测试数据进行了 3 类常见视觉干扰模拟:高斯噪声(均值为 0,标准差为 10)、模拟低光照以及颜色扰动。实验在原始测试集的基础上分别处理 3 类图像,再使用相同模型进行推理分析。结果显示,本文模型在干扰图像上的检测性能下降幅度明显小于原始 RT-DETR 模型,在低光照和颜色扰动条件下仍保持较高的 mAP50,表现出良好的抗干扰能力与鲁棒性。具体来看,改进后的模型在高斯噪声、低光照以及颜色扰动的 mAP50 分别为

72.8%、81.1%以及 84.6%,均优于 RT-DETR 的 69.6%、77.2%以及 82.3%。结果如表 6 所示。

表 6 不同干扰条件下检测性能对比

Table 6 Comparison of detection performance under different interference conditions %

干扰类型	RT-DETR mAP50	本文算法 mAP50	性能提升
原始图像	83.5	85.6	+2.1
高斯噪声	69.6	72.8	+3.2
低光照	77.2	81.1	+3.9
颜色扰动	82.3	84.6	+2.3

3.9 结果可视化

通过热力图对检测结果进行可视化,更直观地展示了改进模块在钢缆缺陷检测任务中的性能优势。热力图能够有效地反映不同模型对输入图像中各区域的关注程度,便于深入理解模型的检测机制及定位精度。图 8 展示了原始 ResNet 模型和改进的 BasicStar 模块在特征关注上的差异。从图 8(c)、(d)对比可以明显观察到,原始 ResNet 主干对缺陷区域关注度较低,而 BasicStar 模块显著增强了对缺陷区域的关注,表现出更高的局部响应强度。图 9 展示了 RT-DETR 原特征融合与改进的 SOPN 特征融合之间的性能差异。通过图 9(c)、(d)图对比可知,RT-DETR 原特征融合存在漏检情况,并且更多地关注了缺陷以外的不相关区域,而改进后的 SOPN 特征融合则有效避免了漏检现象,精准且集中的关注缺陷区域,体现了其更好的特征提取与融合能力。

综上所述,BasicStar 与 SOPN 模块在钢缆缺陷检测任务中展现出更精准、更有效的特征关注与响应性能,大幅提升了缺陷检测模型的准确性和稳定性。

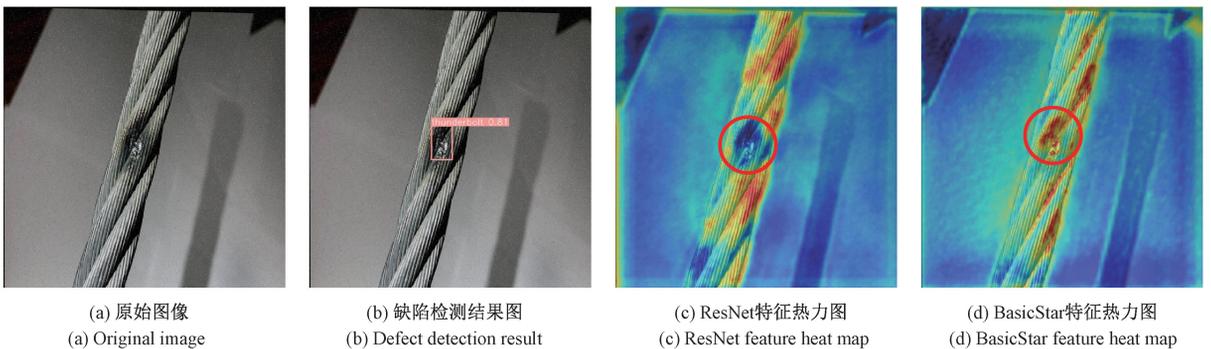


图 8 ResNet 与 BasicStar 模块特征响应热力图对比

Fig. 8 Comparison of ResNet and BasicStar module feature response heatmaps

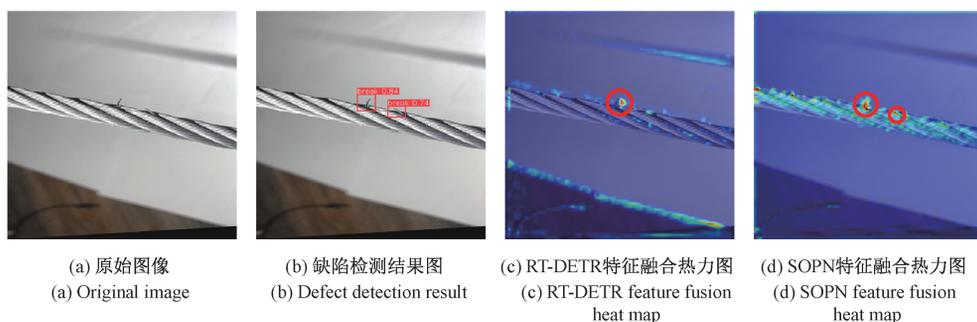


图9 RT-DETR特征融合与SOPN特征融合热力图对比

Fig.9 Comparison of RT-DETR feature fusion and SOPN feature fusion heat map

4 结 论

本文提出了一种基于改进 RT-DETR 模型的钢缆缺陷检测方法,针对传统方法在小目标检测精度上的不足进行了多项改进。首先,在模型主干中重新设计 BasicStar 特征提取模块,提升了模型在高维特征空间中的映射能力,增强了特征提取效果。其次,设计了一种新型 SOPN 特征融合策略,增强了模型对小目标的关注能力,提升了多尺度特征融合的效果。此外,本文融合 Focaler-IoU 和 SIoU 的优点,提出了 Focaler-SIoU 损失函数,进一步优化了模型在小目标检测中的表现。实验结果表明,改进后的模型在钢缆缺陷数据集上的检测精度相比原始 RT-DETR 算法提升了 2.1%,并且在检测精度上优于当前主流的目标检测算法,尤其适用于工业场景中包含大量小目标的钢缆缺陷检测任务。通过详细的对比实验和消融实验,验证了本文提出的各项改进措施对模型性能的提升作用。改进后的 RT-DETR 模型在钢缆缺陷检测任务中展示了卓越的性能,既提高了检测精度,又满足了工业应用中对小目标检测的需求,具有广泛的应用前景。未来的工作将继续优化模型结构和算法,以进一步提升检测效果和适应性,为钢缆检测领域提供更加高效和可靠的解决方案。

参考文献

- [1] KIM Y. Convolutional neural networks for sentence classification [EB/OL]. (2014-08-25) [2025-06-20]. <https://arxiv.org/abs/1408.5882>.
- [2] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need [C]. Advances in Neural Information Processing Systems, 2017, 30: 5998-6008.
- [3] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation [C]. IEEE Conference on Computer Vision and Pattern Recognition, 2014: 580-587.
- [4] REN SH Q, HE K M, GIRSHICK R, et al. Faster R-CNN: Towards real-time object detection with region proposal networks [C]. Advances in Neural Information Processing Systems, 2015, 28: 91-99.
- [5] LIU W, ANGUELOV D, ERHAN D, et al. SSD: Single shot multibox detector [C]. European Conference on Computer Vision (ECCV), 2016: 21-37.
- [6] KHANAM R, HUSSAIN M. YOLOv11: An overview of the key architectural enhancements [EB/OL]. (2024-10-24) [2025-06-20]. <https://arxiv.org/abs/2410.17725>.
- [7] CARION N, MASSA F, SYNNAEVE G, et al. End-to-end object detection with transformers [C]. European Conference on Computer Vision (ECCV), 2020: 213-229.
- [8] ZHU X, SU W, LU L, et al. Deformable DETR: Deformable transformers for end-to-end object detection [EB/OL]. (2020-10-08) [2025-06-20]. <https://arxiv.org/abs/2010.04159>.
- [9] ZHAO Y, LYU W, XU S, et al. DETRs beat YOLOs on real-time object detection [C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2024: 16965-16974.
- [10] 刘航,李明,刘志坚,等.基于轻量化改进 YOLOv8n 绝缘子自爆缺陷检测方法 [J]. 电子测量与仪器学报, 2025, 39(1): 57-69.
- [11] LIU H, LI M, LIU ZH J, et al. Lightweight insulator self-explosion defect detection method based on improved YOLOv8n [J]. Journal of Electronic Measurement and Instrumentation, 2025, 39(1): 57-69.
- [12] 吴葛,朱宇凡,叶天成.改进 YOLOv10 的钢材表面缺陷检测方法 [J]. 电子测量技术, 2025, 48(4): 158-168.
- [13] WU G, ZHU Y F, YE T CH. Steel surface defect detection method based on improved YOLOv10 [J]. Electronic Measurement Technology, 2025, 48(4): 158-168.
- [14] 吴斌斌,张礼华,刘军伟,等.基于改进的 EP-RTDETR 小目标 PCB 表面缺陷检测 [J]. 制造技术与机床, 2025(3): 139-148.
- [15] WU B B, ZHANG L H, LIU J W, et al. Small target PCB surface defect detection based on improved EP-RTDETR [J]. Manufacturing Technology & Machine Tool, 2025(3): 139-148.

- [13] HE K M, ZHANG X Y, REN SH Q, et al. Deep residual learning for image recognition [C]. IEEE Conference on Computer Vision and Pattern Recognition, 2016: 770-778.
- [14] ZHANG H, ZHANG S. Focaler-IoU: More focused intersection over union loss [EB/OL]. (2024-01-24) [2025-06-20]. <https://arxiv.org/abs/2401.10525>.
- [15] GEVORGYAN Z. SiO loss: More powerful learning for bounding box regression [EB/OL]. (2022-05-24) [2025-06-20]. <https://arxiv.org/abs/2205.12740>.
- [16] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition [EB/OL]. (2014-09-04) [2025-06-20]. <https://arxiv.org/abs/1409.1556>.
- [17] CHOLLET F. Xception: Deep learning with depthwise separable convolutions [C]. IEEE Conference on Computer Vision and Pattern Recognition, 2017: 1251-1258.
- [18] SUNKARA R, LUO T. No more strided convolutions or pooling: A new CNN building block for low-resolution images and small objects [C]. Joint European Conference on Machine Learning and Knowledge Discovery in Databases, 2022: 443-459.
- [19] REZATOFIGHI H, TSOI N, GWAK J Y, et al. Generalized intersection over union: A metric and a loss for bounding box regression [C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019: 658-666.
- [20] GEVORGYAN Z. SiO loss: More powerful learning for bounding box regression [EB/OL]. (2022-05-24) [2025-06-20]. <https://arxiv.org/abs/2205.12740>.
- [21] ZHANG H, ZHANG S. Focaler-IoU: More focused intersection over union loss [EB/OL]. (2024-01-24) [2025-06-20]. <https://arxiv.org/abs/2401.10525>.
- [22] ZHANG H, LI F, LIU S, et al. DINO: DETR with improved denoising anchor boxes for end-to-end object detection [EB/OL]. (2022-03-07) [2025-06-20]. <https://arxiv.org/abs/2203.03605>.
- [23] FENG C, ZHONG Y, GAO Y, et al. TOOD: Task-aligned one-stage object detection [C]. IEEE/CVF International Conference on Computer Vision (ICCV), 2021: 3490-3499.
- [24] ZHANG S, WANG X, WANG J, et al. Dense distinct query for end-to-end object detection [C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023: 7329-7338.
- [25] HANIF M W, LI ZH L, BASHIR R. An innovative approach to enhance the safety of elevator using steel cable damage detection model based on YOLO [J]. International Journal of Intelligent Systems and Applications in Engineering, 2024, DOI:10.1039/04/06/2024/OBJ.
- [26] 李志星, 杨啸龙, 李天昊, 等. 基于 YOLOv8 的煤矿用钢丝绳损伤检测算法 [J]. 电子测量技术, 2024, 47(9): 120-128.
- LI ZH X, YANG X L, LI T H, et al. Steel wire rope damage detection algorithm for coal mines based on YOLOv8 [J]. Electronic Measurement Technology, 2024, 47(9): 120-128.
- [27] TAN M, PANG R, LE Q V. EfficientDet: Scalable and efficient object detection [C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 10781-10790.
- [28] WANG C, HE W, NIE Y, et al. Gold-YOLO: Efficient object detector via gather-and-distribute mechanism [J]. Advances in Neural Information Processing Systems, 2023, 36: 51094-51112.
- [29] SHI Z, HU J, REN J, et al. HS-FPN: High frequency and spatial perception FPN for tiny object detection [EB/OL]. (2024-12-17) [2025-06-20]. <https://arxiv.org/abs/2412.10116>.
- [30] MA T, YIN H. MAFPN: A mixed local-global attention feature pyramid network for aerial object detection [J]. Remote Sensing Letters, 2024, 15(9): 907-918.
- [31] TONG Z, CHEN Y, XU Z, et al. Wise-IoU: Bounding box regression loss with dynamic focusing mechanism [EB/OL]. (2023-01-24) [2025-06-20]. <https://arxiv.org/abs/2301.10051>.
- [32] MA S, XU Y. MPDIoU: A loss for efficient and accurate bounding box regression [EB/OL]. (2023-07-14) [2025-06-20]. <https://arxiv.org/abs/2307.07662>.
- [33] 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.
- [34] ZHENG Z, WANG P, LIU W, et al. Distance-IoU loss: Faster and better learning for bounding box regression [C]. AAAI Conference on Artificial Intelligence, 2020, 34(7): 12993-13000.
- [35] REZATOFIGHI H, TSOI N, GWAK J Y, et al. Generalized intersection over union: A metric and a loss for bounding box regression [C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019: 658-666.

作者简介

杨昌昌, 硕士研究生, 主要研究方向为缺陷检测。

E-mail: yangchangchang88@163.com

张笃振(通信作者), 博士, 副教授, 主要研究方向为计算机视觉、人工智能、深度学习。

E-mail: zhduzhen@jsnu.edu.cn

王斯豪, 硕士研究生, 主要研究方向为图像分割。

E-mail: 1826653404@qq.com