

中图法分类号: TP18; TP391 文献标识码: A 文章编号: 1006-8961(2026)02-0499-13

论文引用格式: Peng Z R and Pei Z B. 2026. Defect detection method for locomotive coil springs integrating features and deformable attention. Journal of Image and Graphics, 31(2):0499-0511(彭珍瑞, 裴志彪. 2026. 融合多尺度特征与可变形注意力的机车圆弹簧缺陷检测方法. 中国图象图形学报, 31(2):0499-0511)[DOI:10.11834/jig.250240]

融合多尺度特征与可变形注意力的机车 圆弹簧缺陷检测方法

彭珍瑞*, 裴志彪

兰州交通大学机电工程学院, 兰州 730070

摘要: 目的 机车圆弹簧缺陷检测对机车的安全运行至关重要, 受限于检测车间复杂环境的影响, 机车圆弹簧缺陷检测存在漏检、误检及检测效率低下的问题, 为此, 提出一种基于改进 RT-DETR(real-time detection Transformer)的机车圆弹簧缺陷检测方法。方法 首先, 应用结构重参数化(structural re-parameterization, Rep)方法来改进部分卷积(partial convolution, Pconv), 构建重参数化的部分卷积(Rep-Pconv)替换原始主干中的 Basic block; 其次, 在基于注意力的尺度内特征交互(attention-based intra-scale feature interaction, AIFI)模块中引入可变形注意力(deformable attention, DA)机制修正原有结构中的多头自注意力机制(multi-head self-attention), 提高模型对局部区域的关注度; 最后, 在颈部网络添加 P2 检测层, 结合尺度序列特征融合(scale-sequence feature fusion, SSFF)思想与三重特征编码器(triple feature encoder, TFE)结构, 构建轻量的跨尺度特征信息融合模块。**结果** 利用自建机车圆弹簧缺陷数据集对所提方法进行验证, 相较于原始的 RT-DETR 算法, 改进后 RT-DETR 算法在参数量减少 54% 的同时, 将 mAP50 提升至 97.2%, 提高 2.8%, 精度以及召回率分别提升 0.8% 和 1.2%; 与 YOLOv5s、YOLOv8s、YOLO11s 和 YOLO12s 相比, 所提算法在多项指标上表现出显著优势。**结论** 本文所改进的 RT-DETR 机车圆弹簧缺陷检测算法, 能够有效应对机车圆弹簧在车间复杂环境检测的要求, 通过实验验证并与当前主流目标检测算法进行对比分析, 结果表明所提算法在实验数据集上表现优异。

关键词: 机车圆弹簧; 缺陷检测; RT-DETR; 部分卷积(Pconv); 结构重参数化(Rep); 可变形注意力机制; 特征融合

Defect detection method for locomotive coil springs integrating features and deformable attention

Peng Zhenrui*, Pei Zhibiao

School of Mechanical Engineering, Lanzhou Jiaotong University, Lanzhou 730070, China

Abstract: Objective As critical components of a vibration reduction system, the primary and secondary coil springs of locomotives are vital to safe train operations. With the increasing demand for railway transportation, these springs are often subjected to alternating loads, environmental corrosion, and fatigue stress, making them highly susceptible to surface cracks. If not detected and addressed in time, such defects may lead to spring fractures, locomotive malfunctions, or even serious safety accidents. At present, the inspection of locomotive coil springs primarily relies on magnetic particle testing

收稿日期: 2025-05-25; 修回日期: 2025-07-16; 预印本日期: 2025-07-23

* 通信作者: 彭珍瑞 pzrui@163.com

基金项目: 甘肃省自然科学基金项目(24JRRA172); 中国铁路兰州局科技发展项目计划(LZJKY2024022-1)

Supported by: Natural Science Foundation of Gansu Province, China (24JRRA172); Science and Technology Development Project of China Railway Lanzhou Bureau (LZJKY2024022-1)

combined with manual visual inspection. While this approach can detect most apparent defects, fine or concealed cracks are often missed owing to dust interference, strong magnetic fields, and fluorescent reflection in the testing environment. Moreover, manual methods suffer from low efficiency and lack of consistency, failing to meet the demands of intelligent and highly reliable modern locomotive maintenance. Therefore, this study proposes a surface defect detection method for locomotive coil springs based on an improved real-time detection transformer (RT-DETR) model, aiming to leverage deep learning and computer vision technologies to enhance the automation and robustness of defect recognition, thereby ensuring spring operation safety and supporting intelligent maintenance. **Method** The RT-DETR framework was extensively enhanced in three key aspects to improve the detection capability for small surface defects under complex industrial conditions while ensuring real-time performance and model efficiency. First, in the backbone network, the original basic block was replaced with a re-parameterized partial convolution module (Rep-Pconv), constructed by applying structural reparameterization to partial convolution. This structure improves feature representation during training via a multibranch architecture and simplifies the inference phase through branch fusion, thus reducing computational cost while maintaining detection accuracy. Second, in the attention-based intrascale feature interaction module, the conventional multihead self-attention was replaced with a deformable attention (DA) mechanism. This modification allows the model to dynamically perceive spatial offsets and focus effectively on salient local regions, enhancing its ability to detect fine-grained defects in cluttered backgrounds. Third, a P2 detection layer was added to improve sensitivity to small targets. A lightweight cross-scale feature fusion module was also designed by incorporating scale-sequence feature fusion and a triple feature encoder, enabling efficient integration of shallow texture and deep semantic features across multiple scales. **Result** Comprehensive experiments were conducted on a self-constructed dataset of locomotive spring surface defects to validate the effectiveness and superiority of the proposed method. Results show that compared with the original RT-DETR, the improved model reduced the parameter count by approximately 54% and improves the mean average precision to 97.2%, representing a 2.8% increase. Additionally, precision and recall were improved by 0.8% and 1.2%, respectively, demonstrating enhanced performance in false positive suppression and missed defect reduction. Furthermore, comparative evaluations against mainstream lightweight object detection algorithms including YOLOv5s, YOLOv8s, YOLO11s, and YOLO12s revealed that the proposed method outperformed all baselines across key performance metrics. **Conclusion** The improved RT-DETR-based defect detection algorithm proposed in this paper introduces Rep-Pconv, deformable attention mechanisms, and a lightweight cross-scale feature fusion module, effectively addressing the challenges posed by complex industrial environments. Experimental results confirm that the model not only significantly reduces computational complexity but also achieves superior detection performance on the custom locomotive spring defect dataset. The method exhibits strong advantages in detection accuracy, efficiency, and robustness, offering a reliable technical solution for intelligent defect detection. It lays a solid foundation for the future development of automated and intelligent defect identification systems in practical industrial applications.

Key words: locomotive coil spring; defect detection; real-time detection Transformer (RT-DETR); partial convolution (Pconv); structural re-parameterization (Rep); deformable attention mechanism; feature fusion

0 引言

转向架一、二系圆弹簧作为机车减振系统的关键组件,对机车运行的平稳性、舒适性以及安全性具有重要作用(吴硕,2023)。然而,机车圆弹簧在服役过程中长期处于恶劣的运行环境,且受到振动影响,容易产生表面裂纹,从而严重影响机车健康运营。因此,有必要对圆弹簧进行表面缺陷检测,以保障行车安全。目前,检修车间以湿法荧光磁粉探伤(徐圣恩

和姚晓辰,2024)作为机车圆弹簧检测的主流方法,该方法的主要流程如下:首先将待检弹簧放至探伤机工作台,使用磁悬液喷洒,接着对弹簧进行磁化,最后由工作人员手持荧光灯对弹簧进行检测,观察其表面是否存在缺陷。然而,由于表面弯折、遮挡和光照不均等复杂因素,这种完全依赖于人工的检测手段在检测时难以覆盖所有可能存在的缺陷区域,且存在检测效率低、精度不足等问题。因此,亟需开发一种基于视觉处理技术的机车圆弹簧缺陷检测方法。

计算机视觉的飞速发展,为工业金属表面缺陷检测提供了新的解决思路,其中,基于深度学习的检测方法凭借其强大的特征提取与识别能力,在工业场景中得到了广泛应用(汤勃等,2017)。当前,主流的目标检测算法主要分为两大类:基于卷积神经网络(convolutional neural network, CNN)(Li等,2022)的检测算法和基于Transformer(Vaswani等,2017)的检测算法。前者包括两阶段检测算法和单阶段检测算法。两阶段检测算法有R-CNN(region-based convolutional neural network)(Girshick等,2014)系列,其核心思想是先生成候选区域,再对候选区域进行分类和回归。Yang等人(2023)针对数字胶片图像中焊缝感兴趣区域宽度和跨度分布不均匀的问题,首先提出一种两阶段定位方法,而后利用改进的Faster R-CNN(Ren等,2017)提高焊缝缺陷检测的准确性。常海涛等人(2018)提出一种基于Faster R-CNN的工业CT(computed tomography)缺陷检测方法,利用卷积神经网络自动提取目标的特征,避免了传统两阶段检测算法手动设计缺陷特征的问题。单阶段检测算法以YOLO(you only look once)(Redmon等,2016)为代表,直接进行端到端检测,速度快、适合实时应用。赖杰等人(2024)提出一种改进的YOLOv5算法,通过增加非显著目标检测层和嵌入CBAM(convolutional block attention module)注意力机制,成功提升了模型对信息不足目标的检测能力。万奥等人(2025)提出基于YOLOv8的模型YOLO-SF-TV,结合Swin Transformer的全局特征提取能力,设计了新的特征融合模块和优化特征金字塔结构,有效提升了经颅超声图像中三脑室的检测精度。尽管这些方法在各种目标检测任务中表现良好,但需手动设计锚框以及需要非极大值抑制进行后处理,这些限制增加了模型结构的复杂性以及推理速度,从而影响计算效率。

作为目标检测的新范例,DETR(detection Transformer)(Carion等,2020)首次将Transformer架构应用于端到端目标检测。它既不需要依赖于手动设计的组件,也不需要非极大值抑制。同时,DETR利用自注意力机制捕捉全局信息,能够更好地建模图像中不同位置之间的依赖关系,从而有更好的识别效果。RT-DETR(real-time detection Transformer)(Zhao等,2024)相对于DETR进行了多项优化,主要通过引入动态解码器设计和动态框匹配策略来加速收

敛,同时优化了特征提取和多尺度特征融合模块,提升了对小目标的检测效果。然而,RT-DETR因其较大的模型参数,存在计算时间和资源成本高昂的问题,这使其不能很好地适应资源受限的边缘设备。

为了解决上述问题,本文在RT-DETR算法的基础上,设计了一种机车圆弹簧表面缺陷检测算法,使用重参数化的部分卷积,重构原始ResNet18(residual neural network)主干中的basic block,以压缩模型体积,提升计算效率;利用可变形注意力机制替换尺度内交互模块中的多头注意力机制,提高模型对局部细节特征的关注度;在颈部网络添加P2检测层,并借助ASF-YOLO(attentional scale sequence fusion based YOLO)中的尺度序列特征融合思想和三重特征编码器结构,构建轻量化的跨尺度特征交互模块。通过以上改进,模型在保持检测精度的同时,可大幅减小参数量,提升检测效率,从而为机车圆弹簧提供新的技术支持。

1 RT-DETR 算法

RT-DETR由主干特征提取网络、高效的混合编码器,带有辅助预测头的解码器3部分组成。主干部分提供了4种残差网络(ResNet)(Wang等,2019),分别为ResNet18、ResNet34、ResNet50和ResNet101;混合编码器由基于注意力的尺度内特征交互(attention-based intra-scale feature interaction, AIFI)模块和跨尺度特征融合模块(cross-scale feature fusion module, CCFM)构成,AIFI中采用了一层Transformer结构,以ResNet18为主干的RT-DETR模型结构如图1所示,其中仅将主干部分输出的特征图 S_5 进行序列化和位置编码,之后送入AIFI模块输出 F_5 。而后将 S_3 、 S_4 和 S_5 这3部分送入CCFM进行处理,以更好地融合特征信息。在解码器和头部组件中,3个预测分支被转化为一组特征,利用交并比(intersection over union, IoU)感知查询进行解码操作,以获得对象的类别、置信度和位置信息。

2 RDP-DETR 算法

本文以RT-DETR-R18为基础,提出一种应用于机车圆弹簧缺陷检测算法RDP-DETR,对模型的改进主要分为3部分:1)应用结构重参数化(structural

re-parameterization, Rep)方法来改进部分卷积(partial convolution, Pconv),引入重参数化的部分卷积(Rep-Pconv)替换原有的basic block,压缩主干网络规模;2)将AIFI中的多头自注意力(multi-head self-attention)替换为可变形注意力(deformable attention, DA),增强模型对局部细节特征的表达能力;3)在颈部结构中加入P2检测层,并结合ASF-YOLO中的尺度序列特征融合思想与三重特征编码器结构优化多尺度特征交互方式。通过上述改进,模型在保持检测精度的基础上,实现了更优的轻量化设计,增强了对小目标的感知能力,并提高在复杂工

业场景中的实用性。RDP-DETR 网络结构如图2所示。

2.1 改进的主干网络

2.1.1 部分卷积

机车圆弹簧裂纹检测在荧光磁粉探伤环境下实施时,受工件表面油渍污垢、局部反光及车间强磁干扰等影响,易导致检测图像特征信息缺失,进而影响传统卷积对裂纹特征的提取效果。为增强模型对不完整或不规则区域的感知能力,本文引入了部分卷积(Chen等,2023)作为关键特征提取模块,图3解释了普通卷积及Pconv的原理。

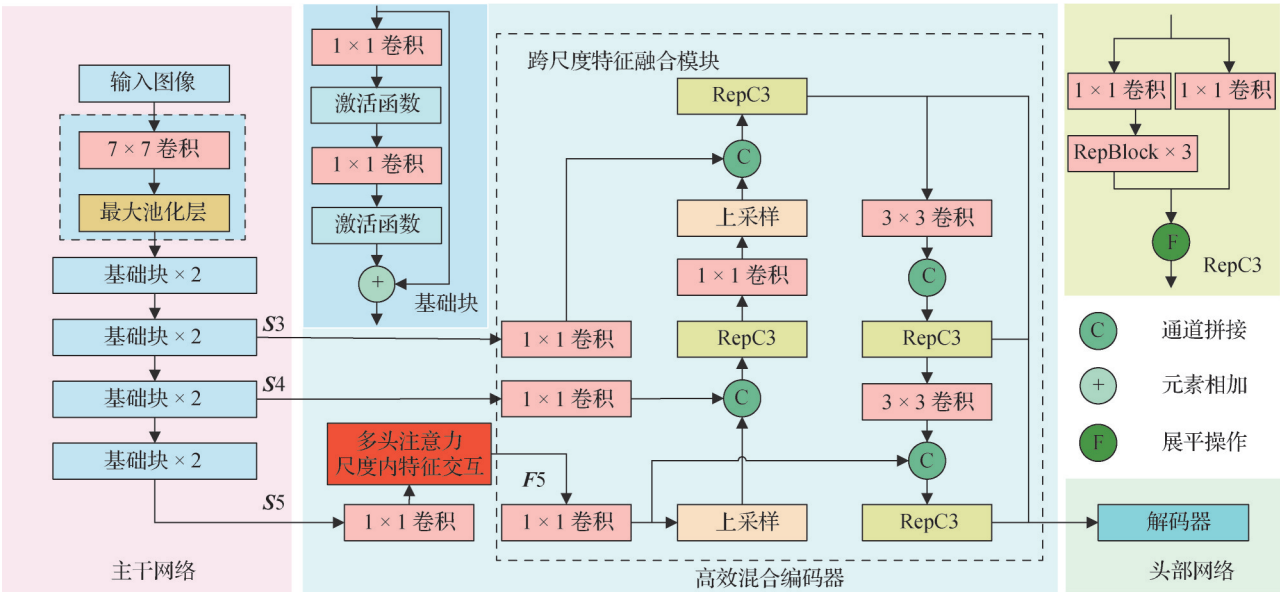


图1 RT-DETR 架构

Fig. 1 Structure of RT-DETR

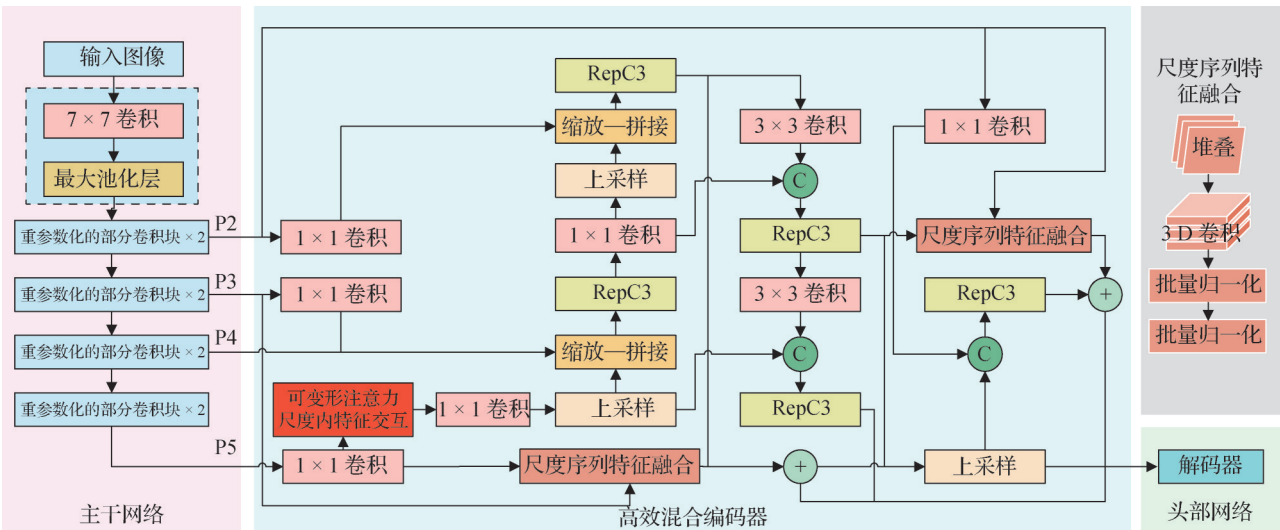


图2 RDP-DETR 架构

Fig. 2 Structure of RDP-DETR

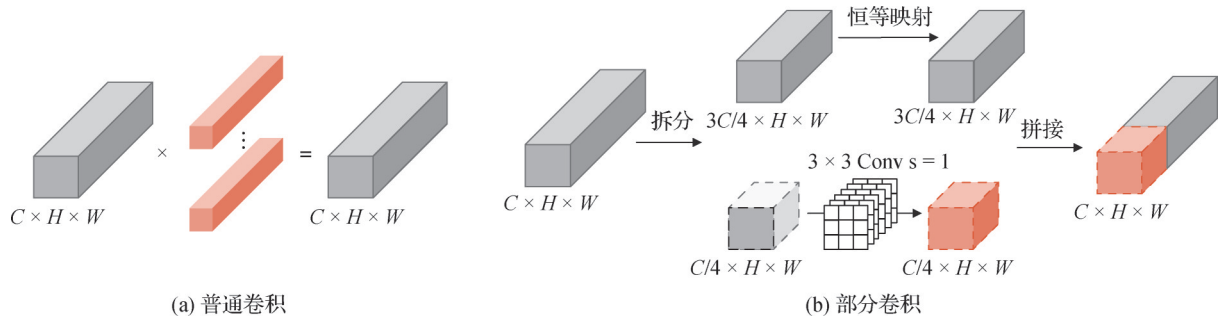


图3 普通卷积以及部分卷积

Fig. 3 Ordinary convolution and partial convolution ((a) ordinary convolution; (b) partial convolution)

由于图像中特定通道之间的特征映射高度相似,如果卷积仍然应用于特征图的每个通道,会出现特征冗余,进而显著增加计算成本,影响网络的实时性能。因此,为了降低计算成本,Pconv从输入通道的前1/4或后1/4选择一组连续的卷积通道来提取特征。此外,剩余通道的特征图将被映射到下一步,将卷积之后的特征图与未被卷积的特征图进行拼接(一般认为输入和输出特征映射具有相同数量的通道)。Pconv的每秒浮点运算次数(floating point operations per second, FLOPs)为

$$Pconv_{FLOPs} = H \times W \times k^2 \times c_p^2 \quad (1)$$

式中, H 为特征图的高度, W 为特征图的宽度, k 为卷积核大小, c_p 为通道数。

在比率 $ratio = c_p/c = 1/4$ 时, $Pconv_{FLOPs}$ 仅为常规卷积的1/16,此时Pconv的内存访问量为

$$H \times W \times 2c_p + k^2 \times c_p^2 \approx H \times W \times 2c_p \quad (2)$$

对于 $ratio = c_p/c = 1/4$,Pconv的内存访问量仅为常规卷积的1/4。

2.1.2 结构重参数化

部分卷积凭借对特征图冗余的利用,显著提升了模型的特征提取能力,降低了计算成本,但其计算效率在实际应用中仍存在提升空间。为进一步提升模型的推理速度和轻量化水平,本文在部分卷积的基础上引入了结构重参数化(Ding等,2021)技术。在训练阶段,网络中采用了多分支结构,包括 3×3 卷积、 1×1 卷积以及Identity分支,以增强特征表达能力。在推理阶段,首先将 1×1 卷积分支通过在卷积核周围填充为 3×3 大小的零扩展,使其尺寸与 3×3 卷积分支一致。其次,恒等映射分支则以 3×3 卷积形式表示,其中仅中心位置的权重为1,其余位置为0。最后,结构重参数化过程将各分支上的卷积层与其对应的批量归一化(batch normalization,

BN)(Ioffe和Szegedy,2015)层进行融合,进一步合并成1个单一的 3×3 卷积算子,实现多分支到单分支的转换,从而在保证特征提取能力的同时提升推理速度。第 i 个通道的BN计算为

$$y_i = \frac{x_i - \mu_i}{\sqrt{\sigma_i^2 + \varepsilon}} \cdot \gamma_i + \beta_i \quad (3)$$

式中, μ_i 是平均值; σ_i^2 是方差; γ_i 和 β_i 是学习参数; ε 是一个近乎为零的常数,防止分母为零; x_i 是输入特征图。卷积和BN融合后的输出为

$$y_i = \frac{x_i - \mu_i}{\sqrt{\sigma_i^2 + \varepsilon}} \cdot \gamma_i + \beta_i = x_i \cdot \frac{\gamma_i}{\sqrt{\sigma_i^2 + \varepsilon}} + \left(\beta_i - \frac{\mu_i \cdot \gamma_i}{\sqrt{\sigma_i^2 + \varepsilon}} \right) = x_i \cdot K + B \quad (4)$$

式中, K 是卷积和BN层融合后的新的卷积核参数矩阵; B 是融合后的偏置矩阵;假设生成 x_i 的卷积核是 k ,偏置是 b ,此时输入和输出为

$$x_i = \hat{x}_i \cdot k + b \\ y_i = x_i \cdot K + B = (\hat{x}_i \cdot k + b) \cdot K + B = \hat{x}_i \cdot k \cdot K + b \cdot K + B \quad (5)$$

式中, $\hat{x}_i$ 是输入的特征图; $k \cdot K$ 是融合后的卷积核; $b \cdot K + B$ 是新的偏置矩阵。最后,每个分支的新融合卷积核被添加到单个 3×3 卷积中。结构重参数化原理如图4所示。

2.1.3 重参数化的部分卷积

Pconv与Rep分别从不同角度提升模型在裂纹检测任务中的性能:前者利用特征图的冗余,对部分输入通道应用卷积,在保留特征提取能力的同时,减少了计算量和内存访问次数;后者则利用训练阶段的多分支结构与推理阶段单分支结构的结合;结合重参数化的部分卷积,可在压缩模型体积的同时减小模型计算量,兼顾了轻量化与计算效率的优势。

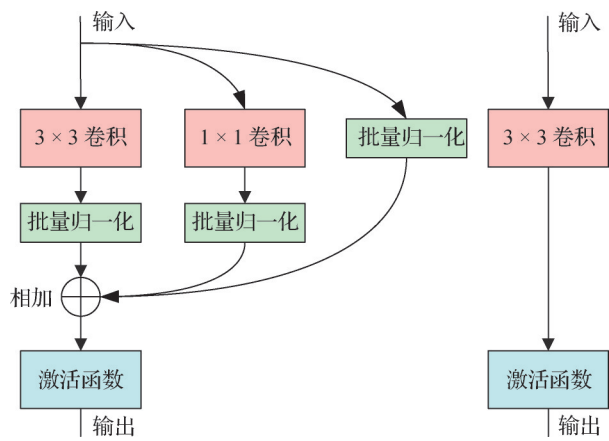


图4 结构重参数化

Fig. 4 Structural re-parameterization

Rep-Pconv 原理如图 5 所示。

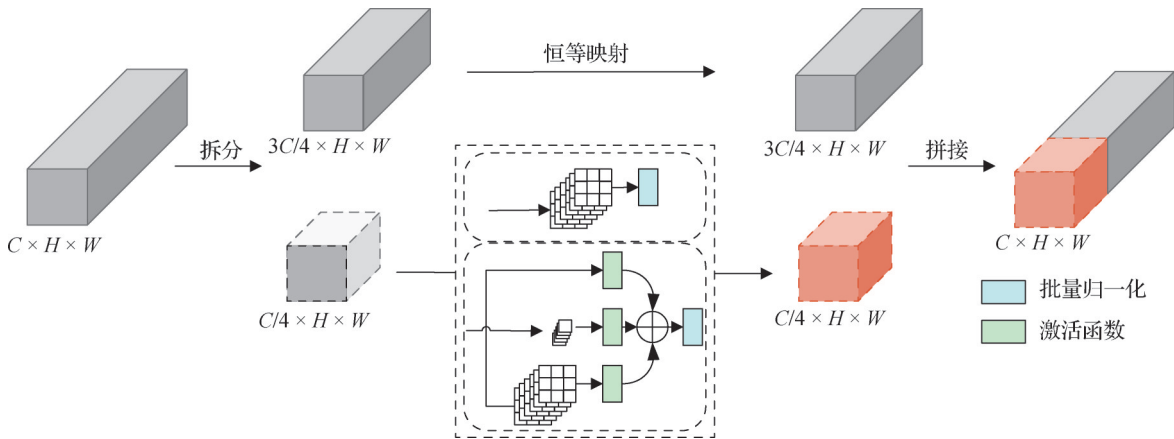


图5 重参数化的部分卷积

Fig. 5 Re-parameterized partial convolution

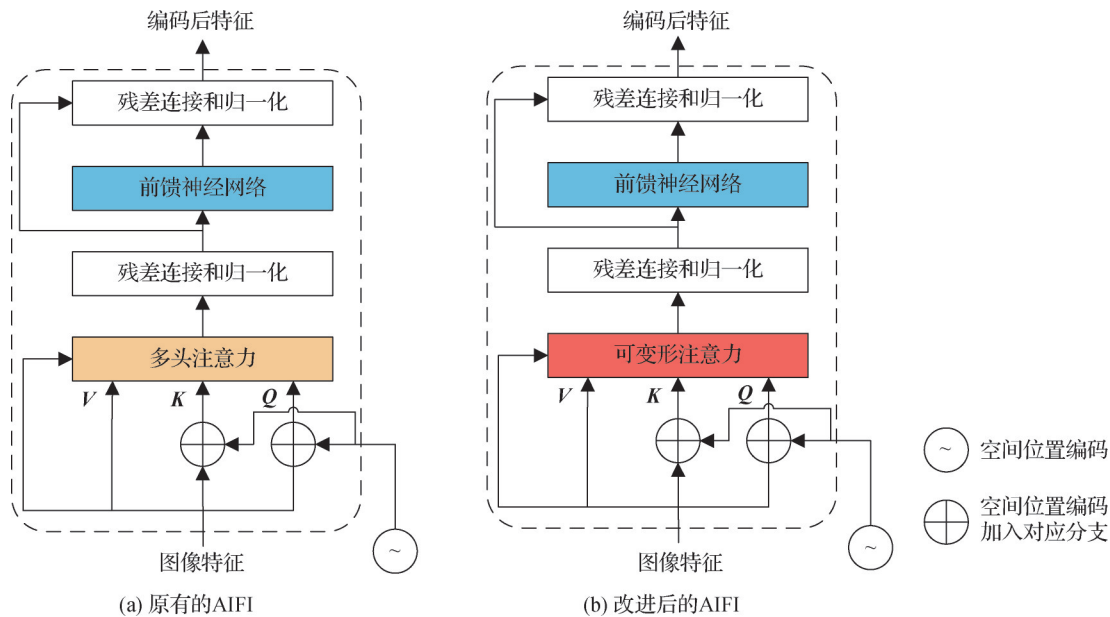


图6 原有的AIFI以及改进后的AIFI结构

Fig. 6 Original AIFI and improved AIFI structure ((a) original AIFI; (b) improved AIFI)

2.2 改进的AIFI

RT-DETR 利用 AIFI 对主干网络提取的特征图 S5 进行交互式编码,其通过使用多头自注意力机制计算最深特征图的注意力权重。图 6(a)所示为 RT-DETR 的 AIFI 结构。机车圆弹簧裂纹检测任务中,由于裂纹形态不规则、位置变化较大、目标区域尺寸较小且易受背景噪声干扰的特点,传统多头注意力机制可能导致注意力分布差异性不足,从而导致模型对裂纹的关注度偏移。为了解决这个问题,本文在 RT-DETR 中引入可变形注意力(DA)机制(Dai 等,2017),对其 AIFI 中的多头注意力结构进行了改进。改进后的 AIFI 结构如图 6(b)所示。

DA模块引入了由查询特征驱动的偏移网络,用以生成采样偏移量(Xia等,2022)。通过根据偏移值动态调整候选keys和values特征的位置,使得模型能够聚焦于重要信息区域以捕获更多特征,兼顾了自注意力机制的灵活性与效率。可变形注意力机制原理如图7所示。

首先,将输入特征 $\mathbf{x} \in \mathbf{R}^{H \times W \times C}$ 划分为网格作为参考点,其中网格的宽度为 r 。为了获得网格上参考点的偏移量,特征图被线性投影到查询标记 q 中。然后,将这些查询特征输入到一个轻量级偏移网络中(表示为 $\theta_{\text{offset}}(\cdot)$)来生成偏移量 ΔP 。基于生成的偏移量,使用双线性插值特征图进行采样,从而获得变形特征,采样得到的特征随后被分别投影至keys和values空间。queries、keys和values通过多头自注意力机制结合的同时考虑位置偏移。此过程会产生一个交互式特征序列,用于捕获不同元素之间的交互和关系,可变形注意力机制的特征交互过程为

$$\begin{cases} Q = \mathbf{x}W_q \\ \tilde{K} = \tilde{\mathbf{x}}W_k \\ \tilde{V} = \tilde{\mathbf{x}}W_v \end{cases} \quad (6)$$

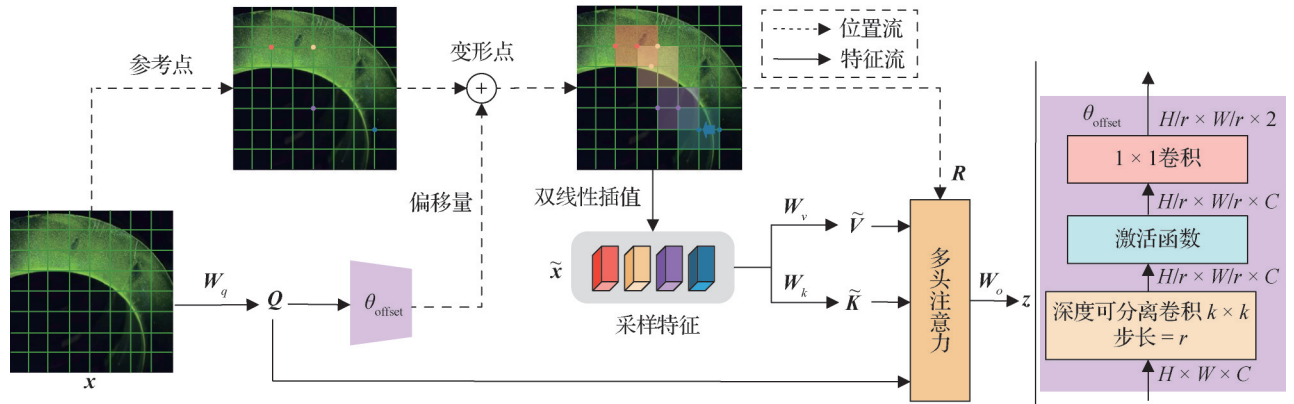


图7 可变形注意力

Fig. 7 Deformable attention

2.3 改进的CCFM

RT-DETR的CCFM模块由RepC3模块构建的特征金字塔网络(feature pyramid network, FPN)(Ghiasi等,2019)与原型对齐网络(prototype alignment network, PAN)(Wang等,2019)组成,该结构在保持模型整体轻量化的基础上,兼顾了多层特征信息的流通、细节增强和交互表达能力。然而,其融合方式主要通过上采样与拼接操作,特征之间的语义层次差异较大,信息融合时存在特征关联不强和重复的问题。

$$\Delta P = \theta_{\text{offset}}(q), \tilde{\mathbf{x}} = \mathcal{O}(\mathbf{x}; p + \Delta p) \quad (7)$$

式中, $\mathbf{x}$ 是输入特征图, $\tilde{K}$ 和 $\tilde{V}$ 是基于偏移特征 $\tilde{\mathbf{x}}$ 生成的keys和values; W_q 、 W_k 和 W_v 是可学习的线性变换矩阵,用于生成queries、keys和values。将采样函数 $\mathcal{O}(\cdot; \cdot)$ 转换成可微分的双线性插值函数后,表达式为

$$\mathcal{O}(z; (P_x, P_y)) = \sum_{(r_x, r_y)} g(P_x, r_x) g(P_y, r_y) z[r_x, r_y, :] \quad (8)$$

式中, $g(a, b) = \max(0, 1 - |a - b|)$ 是插值权重函数; (r_x, r_y) 表示 $z \in \mathbf{R}^{H \times W \times C}$ 的网格点坐标;融合位置偏移信息之后,对queries、keys和values应用多头自注意力机制计算最终特征表示。输出注意力表示为

$$z^{(m)} = \sigma \left(\frac{Q^{(m)} K^{(m)T}}{\sqrt{d}} + \mathcal{O}(\hat{B}; R) \right) \tilde{V}^{(m)} \quad (9)$$

式中, d 表示每个注意力头的维度; $Q^{(m)}$ 、 $K^{(m)}$ 和 $\tilde{V}^{(m)}$ 分别为第 m 个头的queries、keys和values;位置偏移增强函数 $\mathcal{O}(\hat{B}; R)$ 将头部特征连接在一起,最终通过投影 W_o 输出 z 。

在实际应用于机车圆弹簧的缺陷图像时,模型难以有效捕捉尺寸在 8×8 像素以下的细小裂纹,易出现漏检和误检。为增强网络对细小裂纹的敏感性,本文在原始主干提取结构基础上,新增一条P2小目标检测分支,引入来自主干中S2(即4倍下采样)阶段的浅层特征图,并将其输入至下游解码器中用于目标定位与分类。新增的P2层具备更高的分辨率和更多细节纹理信息,能够显著提升网络对细小裂纹的感知能力。改进前后的特征提取结构如图8所示。

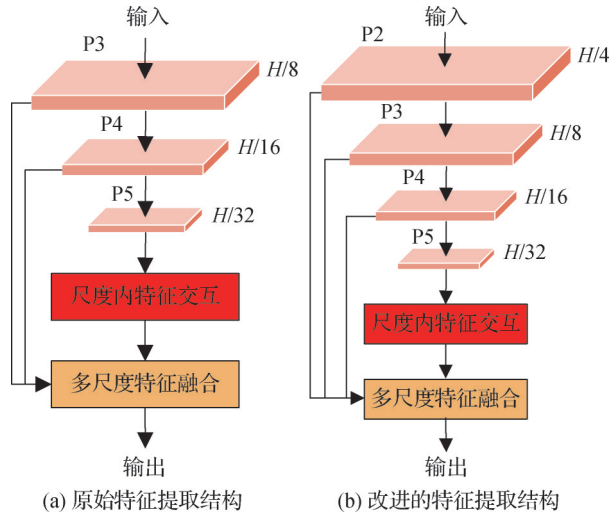


图8 改进前后的特征提取结构

Fig. 8 Feature extraction structure before and after improvement ((a) original feature extraction structure; (b) improved feature extraction structure)

P2检测层的引入提升了对小目标的感知能力,但也带来了特征融合难度上升、计算推理延迟和资源消耗等问题。故借鉴ASF-YOLO(Kang等,2024)中的尺度序列特征融合(scale-sequence feature fusion, SSFF)与三重特征编码器(triple feature encoder, TFE)思想,对CCFM结构进行重新设计,构建适用于多尺度、高分辨率特征融合的轻量化结构。具体实现方式如下:首先,将主干网络输出的多尺度特征图P2、P3、P4与P5按照空间尺度大小排序后,构造其尺度序列;其次,借助SSFF模块,对各层特征进行通道统一,并采用不同标准差的高斯核进行特征平滑,构建尺度感知序列;最后,将多尺度特征图通过水平堆叠操作组织为统一结构,并引入3D卷积提取其跨尺度语义关系,从而建立浅层高分辨率细节与深层抽象语义之间的关联。考虑到计算效率,本文在实际实现中对3D卷积进行了轻量化处理,仅保留特征堆叠结构和统一的空间分辨率(以P2为基准),尺度感知序列的高斯平滑特征图为

$$F_{\sigma}(i, j) = \sum_u \sum_v f(i - u, i - v) \times G_{\sigma}(u, v) \quad (10)$$

$$G_{\sigma}(x, y) = \frac{1}{2\pi\sigma^2} e^{-\frac{(x^2 + y^2)}{2\sigma^2}} \quad (11)$$

式中, f 表示二维特征图, (i, j) 表示输出特征图中的空间位置, σ 表示高斯函数的标准差, $G_{\sigma}(u, v)$ 表示以标准差 σ 为参数的二维高斯核, (u, v) 表示卷积窗口中相对于当前位置的偏移量, F_{σ} 是对二维特征图

使用具有递增标准差 σ 的二维高斯滤波器进行平滑卷积操作之后生成的结果。

在完成尺度序列建模之后,为进一步增强多尺度特征的细节表达能力,引入了TFE模块,对不同尺度的特征图进行结构化增强。TFE模块将大、中、小3个尺度的特征图分别处理,并在融合前统一通道维度,使其具备相同的结构基础。其中,大尺度特征图(如P5)通过最大池化与平均池化组合进行下采样,以保留长距离结构信息并提升对几何变化的鲁棒性;小尺度特征图(如P2)则采用最近邻插值方式上采样,有效增强局部边缘与细节响应,避免微小目标信息丢失。各尺度特征随后经过卷积编码模块提取结构特征,并在通道维度上进行拼接融合,进一步丰富了跨尺度的语义表达。TFE模块融合特征图的输出为

$$F_{TFE} = \text{Concat}(F_l, F_m, F_s) \quad (12)$$

式中, F_{TFE} 表示TFE模块输出的特征图。 F_l 、 F_m 和 F_s 分别表示大、中和小尺寸特征图。

综上,为了提升模型在小目标检测场景下的性能,本文在引入P2检测层的基础上,借鉴ASF-YOLO中注意力尺度序列融合的思想,对原CCFM模块进行了重新设计,构建了P2-ASF模块,该结构采用SSFF模块增强多尺度特征之间的表达与交互能力,同时引入TFE模块对不同尺度特征图进行细节增强与结构建模。

3 实验及结果分析

3.1 实验环境

本文实验具体配置如表1所示。训练超参数设置为:使用AdamW优化器,初始学习率0.0001,批处理大小为8,不使用预训练权重,迭代次数为300。

3.2 数据集构建

机车圆弹簧缺陷数据涉及到铁路行业的数据保密协议,当前尚无公开可用的机车圆弹簧表面缺陷图像数据集,在与某机务段的合作中,本研究获取了该段探伤车间近两年开展探伤作业所形成的相关日志资料,并从中筛选出具有代表性的圆弹簧裂损图像样本。为提升模型在真实工况下的适应能力,对数据集进行了多种形式的扩充操作,包括旋转、平移、随机裁剪、高斯模糊以及添加随机噪声。剔除无

表 1 实验环境配置
Table 1 Experimental environment configuration

名称	相关配置/版本
操作系统	Windows11
PyTorch	2.0.0
CPU	13th Gen Intel(R) Core(TM) i5-13600KF
GPU	NVIDIA GeForce RTX 4060Ti
CUDA	11.8
Python	3.10

效样本后,最终构建了包含 1 186 幅有效样本的数据集。使用 LabelImg 进行手动标注,标注为 .txt 格式,按照 7:2:1 的比例将数据划分为训练集 830 幅、验证集 237 幅与测试集 119 幅。

3.3 评价指标

为全面评估模型在机车圆弹簧表面裂纹识别任务上的性能及优势,本研究使用了目标检测中的标准指标精确率(precision,P)、召回率(recall,R)、平均精度均值(mean average precision, mAP)、参数量(Params)和计算量(GFLOPs(giga FLOPs))衡量模型性能。本文实验计算了 mAP50,其表示 IoU 阈值为 0.5 时的 AP 测量值,在目标检测模型中,参数量 Params 表示模型中可学习参数的总数,单位为百万(M),用于衡量模型的存储复杂度;计算量 GFLOPs 表示模型在处理一幅图像时的理论浮点运算次数,单位为 10 亿次浮点运算,用于估算模型的计算开销。

3.4 注意力机制对比

本文在 AIFI 模块中引入了 DA 注意力机制对原有的多头注意力机制进行改进,以提升模型对机车圆弹簧裂纹图像中关键局部区域的响应能力与特征建模精度。为验证其引入的有效性,本文对比了 4 种主流注意力机制:SimAM(simple attention module)(Yang 等,2021)、MSMHSA(multi-scale multi-head self-attention)(Guo 等,2020)、EAA(efficient additive attention)(Shaker 等,2023)以及 DHSA(dynamic-range histogram self-attention)(Sun 等,2024)并进行了对比实验,比较其在 mAP50 指标下的性能。

实验结果如表 2 所示,引入 SimAM、MSMHSA、EAA 和 DHSA 4 种注意力机制之后,mAP50 分别低于 DA 注意力机制 0.6%、0.8%、0.5% 和 0.7%。实

验结果表明在机车圆弹簧缺陷检测任务中,SimAM 作为一种无参数的注意力机制,虽然能提升部分特征通道的表达能力,但由于缺乏空间建模能力,在处理具有复杂纹理的图像时表现有限;MSMHSA 具备一定的特征感知能力,但其全局建模方式对局部高密度缺陷区域的响应不够敏感,当缺陷图像存在局部遮挡或缺陷面积较小时,无法准确聚焦目标;EAA 在低对比度裂纹区域的判别上能力有限,难以处理复杂的缺陷图像;DHSA 虽引入了层次化结构以增强上下文信息融合能力,但该方法结构复杂且对小目标或边缘模糊区域的响应不稳定。相比之下,DA 在保持自注意力机制全局建模能力的同时,能够使模型聚焦于弹簧缺陷图像的裂纹区域,有效提升了对局部细节的响应能力,从而提升检测精度。综合以上分析,本文所引入的注意力机制相较于所提 4 种注意力机制更适合嵌入 AIFI 模块。

表 2 注意力机制对比
Table 2 Performance comparison of attention mechanisms

注意力机制	mAP50/%
SimAM	95.5
MSMHSA	95.3
EAA	95.6
DHSA	95.4
DA	96.1

3.5 消融实验

本文对 RT-DETR-R18 模型进行了逐步优化改进,以提高其在复杂环境中面对机车圆弹簧检测的综合性能,改进分别为:在主干网络引入 Rep-Pconv;在 AIFI 模块引入可变形注意力机制;添加 P2 检测层,结合 ASF 思想,引入 P2-ASF。为验证本文所提模块改进的有效性,设置 6 组消融实验并加以分析,实验在同一数据集、环境配置下进行,具体结果如表 3 所示。

从表 3 可以看到,相较于基线模型,利用 Rep-Pconv 修正原有的卷积块之后,精度提升了 1%,召回率提升了 1.8%,mAP50 提升了 1.6%,参数量减少了 29.5%,GFLOPs 降低了 24.8%,说明 Rep-Pconv 中的部分卷积思想能在保证检测精度的同时压缩模型,减少计算冗余。利用 DA 注意力替换原有 AIFI 模块中的注意力之后,精度提升了 0.3%,召回率提升了

表 3 消融实验结果
Table 3 Results of ablation study

RT-DETR	Rep-Pconv	DA	P2-ASF	P/%	R/%	mAP50/%	Params/M	GFLOPs
√	–	–	–	95.5	93.4	94.8	19.87	56.9
√	√	–	–	96.5	95.2	96.4	14.00	42.8
√	–	√	–	95.8	94.5	96.1	19.87	57.2
√	–	–	√	95.6	94.5	96.7	15.00	57.6
√	√	√	–	96.5	95.1	95.7	14.00	43.0
√	√	√	√	96.3	94.6	97.2	9.13	43.7

注:加粗字体表示各列最优结果。“√”表示使用该模块,“–”表示未使用该模块。

1. 1%,mAP50 提升了 1. 3%,GFLOPs 无明显增加,表明模型提升了对局部细节特征的响应能力,验证了可变形注意力机制对特征表达能力的增强作用;引入 P2 检测层并在此基础上利用 ASF 尺度序列特征融合思想对 CCFM 模块重新设计之后,mAP50 提升了 1. 9%,达到了单项改进峰值,同时,模型参数量下降了 24. 5%,表明所提 P2-ASF 特征融合思想能有效提升模型对圆弹簧表面细小裂纹的感知能力,并在保持高分辨率特征表达的同时控制了模型复杂度。结合对主干网络以及 DA 模块的改进,在参数量以及 GFLOPs 都有明显改善的情况下,模型精度提升了 1%,召回率提升 1. 7%,mAP50 提升了 0. 9%,优于单项模型改进。得益于部分卷积实现的模型压缩效果,以及尺度序列特征融合阶级采用的特征图简单堆叠求和方法,虽然精度和召回率比起单模块引入时有略微损失,但模型参数下降 54. 1%,同时 mAP50 提升至 97. 2%,表明了 RDP-DETR 在检测精度以及轻量化之间取得了最佳平衡。

3. 6 模块引进可视化

为进一步验证所提引进模块的有效性,在消融实验的基础上,选取测试集中 3 幅有代表性的机车圆弹簧缺陷图像(单一线状裂纹标签为 0,连续线状裂纹标签为 1),对各模块引入后的模型检测性能进行可视化分析。如图 9 所示,原始 RT-DETR 模型在荧光灯反射和粉尘干扰明显的弱光环境下存在漏检现象,反映出其在低对比度和复杂背景条件下特征提取能力不足的问题。引入 Rep-Pconv 模块后,模型成功检测出此前漏检的裂纹,表明该模块通过优化卷积结构,有效提升了特征提取的稳定性与效率。引入 P2-ASF 模块后,模型对不同尺度的裂纹目标更

加敏感,检测框定位更加精确,验证了其在增强浅层细节信息表达与语义融合方面的优势。引入 DA 模块后,模型的检测效果提升尤为显著,裂纹均被精确识别且定位准确,检测框置信度普遍较高,表明所引进的可变形注意力机制显示出优异的局部细节聚焦能力,有效克服了多头注意力机制在复杂背景中对关键区域关注偏移的问题。在 RDP-DETR 模型中,检测结果最为理想。模型可准确识别全部裂纹,且检测框与缺陷边界高度贴合,置信度保持在较高水平。所示检测结果进一步验证了所提方法在工业实际应用中的优势与可行性。

3. 7 目标检测算法性能对比

YOLO 作为当前主流的目标检测算法,在复杂环境中小目标检测体现出优越的性能,在目标检测领域一直得到广泛使用。为验证所提改进的有效性,选取与本文模型体积持平的 YOLOv5s、YOLOv8s、YOLO11s 以及 YOLO12s 进行性能对比,结果如图 10 所示。结果表明,所提 RDP-DETR 算法精确率相比所提 YOLO 系列的最高值(92. 6%)高出 3. 7%,召回率相比所提 YOLO 系列最高值(92. 7%)高出 1. 9%,mAP50 相比所提 YOLO 系列的最高值(95. 8%)高出 1. 4%,上述 3 项指标均处于领先地位;从计算资源消耗来看,虽然 RDP-DETR 算法相比基础 RT-DETR 的 GFLOPs 降低了 23. 2%,但仍高于所提 YOLO 系列的最大值,这一额外开销主要用于提升模型的特征表达能力,使模型提升在复杂背景下对细小裂纹的检测效果,处于可接受范围之内。综合来看,RDP-DETR 在 3 项关键性能指标:精确率、召回率和 mAP50,均超越所提 YOLO 系列,同时保持了较好的轻量化特性。

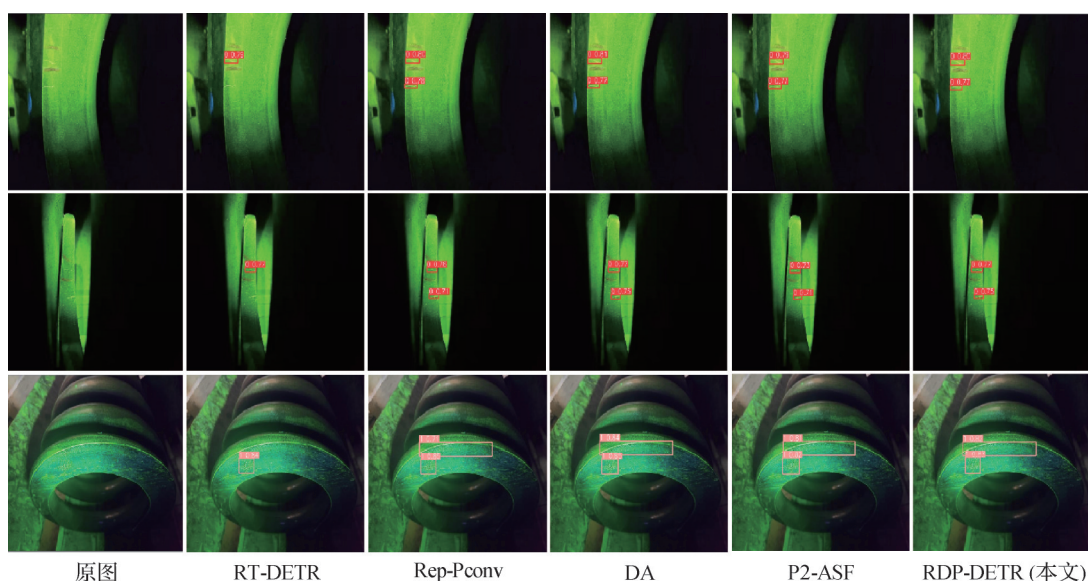


图9 模块引进可视化

Fig. 9 Visualization of module introduction

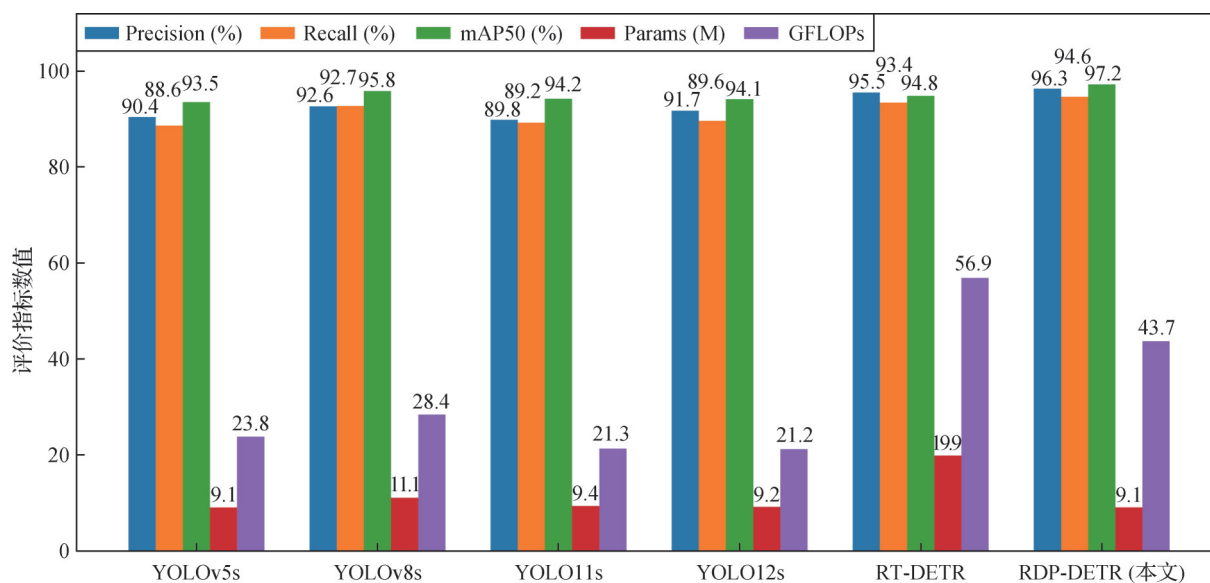


图10 算法性能对比

Fig. 10 Algorithm performance comparison

4 结论

本文针对机车圆弹簧人工检测效率低下的问题,在原有 RT-DETR-R18 的基础上,提出一种轻量的目标检测算法 RDP-DETR,主要贡献如下:1)针对 RT-DETR 模型计算量大的问题,在主干网络中应用了重参数化方法改进后的部分卷积,并将其引入到原始主干中的 basic block 中,从而有效减少了模型的参数量,提升了计算效率。2)为解决多头注意力

机制对裂纹的关注度偏移的问题,在 AIFI 模块中引入了可变形注意力机制,使得模型在处理弹簧表面局部细节特征时具有更强的响应能力。3)为更好地融合不同层级的语义信息,在原有的 CCFM 模块中引入了 P2 检测层,结合 ASF-YOLO 中的 SSFF 与 TFE 结构,构建了轻量高效的多尺度特征融合模块 P2-ASF,优化了不同尺寸特征图之间的交互方式。

本研究为机车圆弹簧缺陷检测提供了新的技术方案,也为其他领域小目标检测任务提供了新的思路和方法。虽然所提方法在特征提取、注意力机制

优化及多尺度特征融合方面具有一定的通用性,但仍有进一步优化空间。未来的研究可扩展数据来源与应用场景,增强模型在多样化工业环境下的鲁棒性与泛化能力,以满足更复杂检测任务的实际需求。

参考文献(References)

- Carion N, Massa F, Synnaeve G, Usunier N, Kirillov A and Zagoruyko S. 2020. End-to-end object detection with transformers//Proceedings of the 16th European Conference on Computer Vision. Glasgow, UK: Springer: 213-229 [DOI: 10.1007/978-3-030-58452-8_13]
- Chang H T, Gou J N and Li X M. 2018. Application of faster R-CNN in image defect detection of industrial CT. *Journal of Image and Graphics*, 23(7): 1061-1071 (常海涛, 荀军年, 李晓梅. 2018. Faster R-CNN在工业CT图像缺陷检测中的应用. *中国图象图形学报*, 23(7): 1061-1071) [DOI: 10.11834/jig.170577]
- Chen J R, Kao S H, He H, Zhuo W P, Wen S, Lee C H, et al. 2023. Run, don't walk: chasing higher FLOPS for faster neural networks//Proceedings of 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver, Canada: IEEE: 12021-12031 [DOI: 10.1109/CVPR52729.2023.01157]
- Dai J F, Qi H Z, Xiong Y W, Li Y, Zhang G D, Hu H, et al. 2017. Deformable convolutional networks//Proceedings of 2017 IEEE International Conference on Computer Vision. Venice, Italy: IEEE: 764-773 [DOI: 10.1109/ICCV.2017.89]
- Ding X H, Zhang X Y, Ma N N, Han J G, Ding J J and Sun J. 2021. RepVGG: making VGG-style convnets great again//Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville, USA: IEEE: 13728-13737 [DOI: 10.1109/CVPR46437.2021.01352]
- Ghiasi G, Lin T Y and Le Q V. 2019. NAS-FPN: learning scalable feature pyramid architecture for object detection//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, USA: IEEE: 7029-7038 [DOI: 10.1109/CVPR.2019.00720]
- Girshick R, Donahue J, Darrell T and Malik J. 2014. Rich feature hierarchies for accurate object detection and semantic segmentation//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Columbus: IEEE: 580-587 [DOI: 10.1109/CVPR.2014.81]
- Guo Q P, Qiu X P, Liu P F, Xue X Y and Zhang Z. 2020. Multi-scale self-attention for text classification//Proceedings of the 34th AAAI Conference on Artificial Intelligence. New York, USA: AAAI Press: 7847-7854 [DOI: 10.1609/aaai.v34i05.6290]
- Ioffe S and Szegedy C. 2015. Batch normalization: accelerating deep network training by reducing internal covariate shift//Proceedings of the 32nd International Conference on Machine Learning. Lille, France: JMLR.org: 448-456 [DOI: 10.5555/3045118.3045167]
- Kang M, Ting C M, Ting F F and Phan R C W. 2024. ASF-YOLO: a novel YOLO model with attentional scale sequence fusion for cell instance segmentation. *Image and Vision Computing*, 147: #105057 [DOI: 10.1016/j.imavis.2024.105057]
- Lai J, Peng R H, Sun D X and Huang J. 2024. Detection of camouflage targets based on attention mechanism and multi-detection layer structure. *Journal of Image and Graphics*, 29(1): 134-146 (赖杰, 彭锐晖, 孙殿星, 黄杰. 2024. 融合注意力机制与多检测层结构的伪装目标检测. *中国图象图形学报*, 29(1): 134-146) [DOI: 10.11834/jig.221189]
- Li Z W, Liu F, Yang W J, Peng S H and Zhou J. 2022. A survey of convolutional neural networks: analysis, applications, and prospects. *IEEE Transactions on Neural Networks and Learning Systems*, 33(12): 6999-7019 [DOI: 10.1109/TNNLS.2021.3084827]
- Redmon J, Divvala S, Girshick R and Farhadi A. 2016. You only look once: unified, real-time object detection//Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA: IEEE: 779-788 [DOI: 10.1109/CVPR.2016.91]
- Ren S Q, He K M, Girshick R and Sun J. 2017. Faster R-CNN: towards real-time object detection with region proposal networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(6): 1137-1149 [DOI: 10.1109/TPAMI.2016.2577031]
- Shaker A, Maaz M, Rasheed H, Khan S, Yang M H and Khan F S. 2023. SwiftFormer: efficient additive attention for transformer-based real-time mobile vision applications//Proceedings of 2023 IEEE/CVF International Conference on Computer Vision. Paris, France: IEEE: 17379-17390 [DOI: 10.1109/ICCV51070.2023.01598]
- Sun S Q, Ren W Q, Gao X W, Wang R and Cao X C. 2024. Restoring images in adverse weather conditions via histogram transformer//Proceedings of the 18th European Conference on Computer Vision (ECCV 2024). Milan, Italy. Cham: Springer: 111-129 [DOI: 10.1007/978-3-031-72670-5_7]
- Tang B, Kong J Y and Wu S Q. 2017. Review of surface defect detection based on machine vision. *Journal of Image and Graphics*, 22(12): 1640-1663 (汤勃, 孔建益, 伍世虔. 2017. 机器视觉表面缺陷检测综述. *中国图象图形学报*, 22(12): 1640-1663) [DOI: 10.11834/jig.160623]
- Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez A N, et al. 2017. Attention is all you need//Proceedings of the 31st International Conference on Neural Information Processing Systems. Long Beach, USA: Curran Associates Inc.: 6000-6010 [DOI: 10.5555/3295222.3295349]
- Wan A, Gao H L, Zhou X, Xue Z and Mou X G. 2025. YOLO-SF-TV: transcranial ultrasound images of the third ventricle as a detection model. *Journal of Image and Graphics*, 30(4): 977-988 (万奥, 高红铃, 周晓, 薛峥, 牟新刚. 2025. YOLO-SF-TV: 经颅超声图像三脑室检测模型. *中国图象图形学报*, 30(4): 977-988) [DOI: 10.11834/jig.240293]

- Wang K X, Liew J H, Zou Y T, Zhou D Q and Feng J S. 2019. PANet: few-shot image semantic segmentation with prototype alignment// Proceedings of 2019 IEEE/CVF International Conference on Computer Vision. Seoul, Korea (South): IEEE: 9196-9205 [DOI: 10.1109/ICCV.2019.00929]
- Wu S. 2023. Experimental Study on a New Nonlinear Secondary Suspension System for Railway Freight Car Bogies. Harbin, China: Institute of Technology (吴硕. 2023. 铁路货车转向架新型非线性二系悬挂系统实验研究. 哈尔滨: 哈尔滨工业大学) [DOI: 10.27061/d.cnki.ghgdu.2023.002283]
- Xia Z F, Pan X R, Song S J, Li L E and Huang G. 2022. Vision transformer with deformable attention//Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans, USA: IEEE: 4794-4803 [DOI: 10.1109/CVPR52688.2022.00475]
- Xu S E and Yao X C. 2024. Research on maintenance standards of primary suspension for HX_D2 locomotive C6 repairs. Railway Locomotive and Motor Car, (11): 38-41 (徐圣恩, 姚晓辰. 2024. HX_D2 系列机车C6修一系悬挂弹簧检修标准研究. 铁道机车与动车, (11): 38-41)
- Yang L X, Zhang R Y, Li L D and Xie X H. 2021. SimAM: a simple, parameter-free attention module for convolutional neural networks// Proceedings of 38th International Conference on Machine Learning. Virtual Event: PMLR: 11863-11874
- Yang W, Xiao Y C, Shen H K and Wang Z P. 2023. Generalized weld bead region of interest localization and improved faster R-CNN for weld defect recognition. Measurement, 222: #113619 [DOI: 10.1016/j.measurement.2023.113619]
- Zhao Y A, Lyu W Y, Xu S L, Wei J M, Wang G Z, Dang Q Q, et al. 2024. DETRs beat YOLOs on real-time object detection//Proceedings of 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA: IEEE: 16965-16974 [DOI: 10.1109/CVPR52733.2024.01605]

作者简介

彭珍瑞,男,教授,博士生导师,主要研究方向为计算机视觉、结构健康监测、机械故障诊断、结构动力学及模态分析。

E-mail: pzrui@163.com

裴志彪,男,硕士研究生,主要研究方向为计算机视觉和目标检测。E-mail: peizb28@163.com