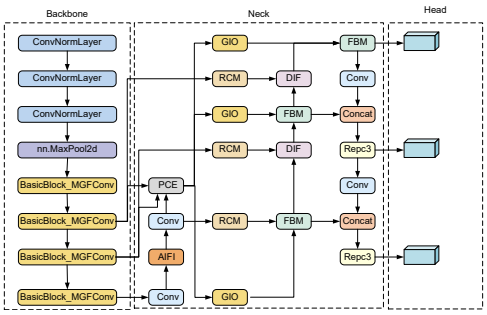
OEE | 光电工程 Opto-Electronic Engineering

Article

2026 年 2 月, 第 53 卷, 第 2 期

自适应多尺度融合的 Transformer 锻件表面缺陷检测算法



梁　丹[1,2]，张　上[1,2*]，管慧攀[1,2]

[1] 三峡大学计算机与信息学院，湖北 宜昌 443002；
[2] 三峡大学湖北省建筑质量检测装备工程技术研究中心，湖北 宜昌 443002

摘要：针对复杂背景下锻件表面缺陷形态多样、尺度多变及检测精度与效率难以兼顾等问题，本文提出一种基于 RT-DETR 架构改进的自适应多尺度融合的上下文检测模型 (MCF-RTDETR)。首先，采集重型卡车转向节生产流水线中的磁粉探测图像，构建锻件表面裂纹缺陷数据集。其次，在骨干网络中引入多向梯度融合卷积 (MGFConv)，通过多方向卷积自适应增强细长裂纹及微小缺陷的方向感知与边缘提取能力；同时，在颈部网络中嵌入上下文感知递归特征金字塔 (CARFPN)，根据局部上下文语义强度为不同尺度特征分配动态权重，实现对尺度变化与背景复杂性的自适应融合。最后，采用 Focaler-MPDIoU 边界框回归损失函数，对样本难度与几何比例进行自适应调节，提高长条状缺陷的定位精度与边界拟合能力。实验结果表明，MCF-RTDETR 相较基准模型 mAP 提升 2.3%，计算复杂度由 58.3 G 降低至 33.5 G，推理速度达到 137.2 f/s；在 GC10-DET 数据集上 mAP 提升了 2.8%，验证了其检测精度与泛化性能。

关键词：缺陷检测；RT-DETR；特征融合；损失函数

The English title and abstract appear at the end of the paper.

DOI: 10.12086/oee.2026.250242 | CSTR: 32245.14.oee.2026.250242 | 中图分类号：TP391 | 文献标识码：A

梁丹，张上，管慧攀. 自适应多尺度融合的 Transformer 锻件表面缺陷检测算法 [J]. 光电工程，2026，**53**(2): 250242

Liang D, Zhang S, Guan H P. Adaptive multi-scale fusion transformer for forging surface defect detection[J]. *Opto-Electron Eng*, 2026, **53**(2): 250242

1　引　言

锻造作为关键的金属塑性成形工艺，广泛应用于汽车、航空航天、船舶制造和石油化工等工业领域。该工艺通过在高温下对金属坯料施加外力实现塑性变形，不仅能优化材料的微观组织，消除内部缺陷，还能显著提升锻件的力学性能[1]。然而，由于原材料不均匀、模具磨损及工艺参数波动等因素，锻件表面常常出现裂纹、夹杂、气泡、斑点及氧化层等缺陷。这些缺陷不仅降低了锻件的机械性能，还可能影响产品结构与使用安全。因此，提升锻件表面缺陷检测技术对保障制造质量与生产可靠性具有重要意义。

当前，无损检测技术已广泛应用于锻件表面缺陷识别，常用方法包括磁粉检测、超声检测、射线检测等[2]。

收稿日期：2025-08-18；修回日期：2025-10-23；录用日期：2025-10-28；网络出版日期：2026-02-13

基金项目：湖北省国家级大学生创新创业训练计划 (S202311075047)；国家级大学生创新创业训练计划 (202111075012, 202011075013)；湖北省数字经济试点示范建设专项 (2312-420625-04-02-996363)

*通信作者：张上，zhangshang@ctgu.edu.cn

版权所有©2026 中国科学院光电技术研究所

其中，磁粉检测因其对铁磁性材料表面和近表面微小裂纹的高灵敏度，被广泛应用于裂纹检测[3]。然而，传统磁粉检测依赖人工视觉判断，存在检测效率低、主观性强和误判率高等问题，尤其在检测大型复杂结构件如转向节等关键部件时，人工检测难以满足现代制造业对自动化和高精度的需求[4]。随着智能制造的加速发展，基于计算机视觉的缺陷检测技术逐渐成为替代传统方法的主流趋势[5]。与传统方法相比，深度学习模型不仅能够自动提取图像特征，还能在处理复杂多样的缺陷类型时表现出更强的适应性，提高检测的准确率和鲁棒性[6]。目前，基于深度学习的目标检测算法主要分为两大类：一类是基于卷积神经网络 (convolutional neural network, CNN) 的检测方法，另一类是基于 Transformer 架构的检测方法。CNN 类方法可进一步细分为两阶段检测 (如 Fast R-CNN[7]、Faster R-CNN[8]) 与单阶段检测 (如 YOLO 系列[9]、SSD[10])。前者通过候选区域生成与分类回归两个阶段实现目标检测，具有较高的检测精度，尤其在小目标检测中表现优异。例如，冷岳峰等[11] 提出一种基于 Fast R-CNN 的钢材缺陷检测方法，通过引入特征融合与通道注意力模块，有效提升了小目标的检测能力。后者则在结构简化与检测速度方面更具优势，适用于工业场景下的实时检测任务。例如，果泽萱等[12] 提出的 MSFT-YOLO 模型，通过引入 TRANS 模块与多尺度特征金字塔结构，实现全局语义与局部特征的有效融合，提高了检测性能和检测速度。然而，YOLO 类模型普遍依赖非极大值抑制 (non-maximum suppression, NMS) 去除冗余框，该过程需设定多个超参数，增加了模型的复杂性，并可能影响检测结果的稳定性。

近年来，Transformer 架构在自然语言处理领域取得了突破性进展，其自注意力机制具备良好的全局建模能力，被广泛应用于计算机视觉任务中。2023 年，百度飞桨团队提出了基于 Transformer 的实时目标检测模型 RTDETR (real-time detection transformer)。该模型在 COCOval2017 数据集上实现了优于同规模 YOLO 系列模型的检测性能[13]，且推理过程中无需依赖 NMS 处理，依然能够实现较高的检测精度，具备更强的可扩展性与部署适应性。例如，Tian 等[14] 提出了 PDP-RTDETR 模型，通过引入轻量化模块 (PSConv) 降低模型计算复杂度，并结合可变形注意力机制与小目标检测层，有效提升了对微小缺陷的检测效果。Yu 等[15] 提出 MCG-RTDETR 模型，结合双可变形卷积、级联群注意力机制和上下文引导下采样特征融合结构，显著提升了模型在复杂背景的目标检测精度。

尽管现有基于 YOLO 或 Transformer 架构的目标检测算法在通用视觉任务中取得了显著成果，但在工业锻件表面缺陷检测中仍面临三方面挑战： 1) 缺陷形态多样、尺度差异显著且边缘模糊，尤其在复杂背景或弱光条件下，细长裂纹等小目标极易被遗漏； 2) 多尺度特征融合过程中，浅层高分辨率特征与深层语义特征在空间与语义上的对齐不足，导致细节信息传递受损； 3) 现有回归损失函数对长条状、不规则边缘缺陷的几何敏感性有限，易出现定位偏差和漏检。

针对上述问题，本文基于 RT-DETR 架构提出一种自适应多尺度融合的上下文实时检测模型 MCF-RTDETR，主要研究工作包括： 1) 设计多向梯度融合卷积模块 (multi-gradient fusion convolution, MGFConv)，采用五分支多方向差分卷积结合重参数化技术，使模型能够根据缺陷边缘形态自适应选择最具判别力的方向特征，从而增强对细长裂纹及微小缺陷的边缘方向敏感性，在不显著增加计算量的情况下提升特征表达能力； 2) 提出上下文感知递归特征金字塔网络 (context-aware recursive feature pyramid network, CARFPN)，通过结合长条卷积的递归上下文模块 (RCM) 与双输入融合模块 (DIF)，在多尺度特征传递中自适应调整不同层次特征的权重比例，实现浅层高分辨率特征与深层语义特征的精确融合。 3) 构建融合难样本聚焦机制与多点距离约束的损失函数 (Focaler-MPDIoU)，通过双重约束自适应调整不同几何形态缺陷的回归误差，有效提升长条状缺陷的定位精度并降低边缘不连续目标的漏检率。

2 RTDETR-r18 网络结构

RTDETR 是首个实现端到端实时目标检测的 Transformer 模型，该模型主要基于 DETR 架构进行优化[16]。DETR 模型在检测过程中依赖非极大值抑制 (NMS) 来消除冗余的目标框。然而，在处理小目标检测或缺陷区域重叠时，NMS 容易出现误检与漏检的情况，从而影响检测精度。RTDETR 则是通过引入改进的解码器结构，直接从输入特征图中生成目标框，在避免使用 NMS 的同时提高检测精度并降低推理延迟。本文选择相对轻量级的 RTDETR-r18 作为基准模型，该模型由主干 (backbone) 网络、颈部 (neck) 网络和检测头 (head) 三部分组成，如图 1 所示。主干部分采用 ResNet-18 作为

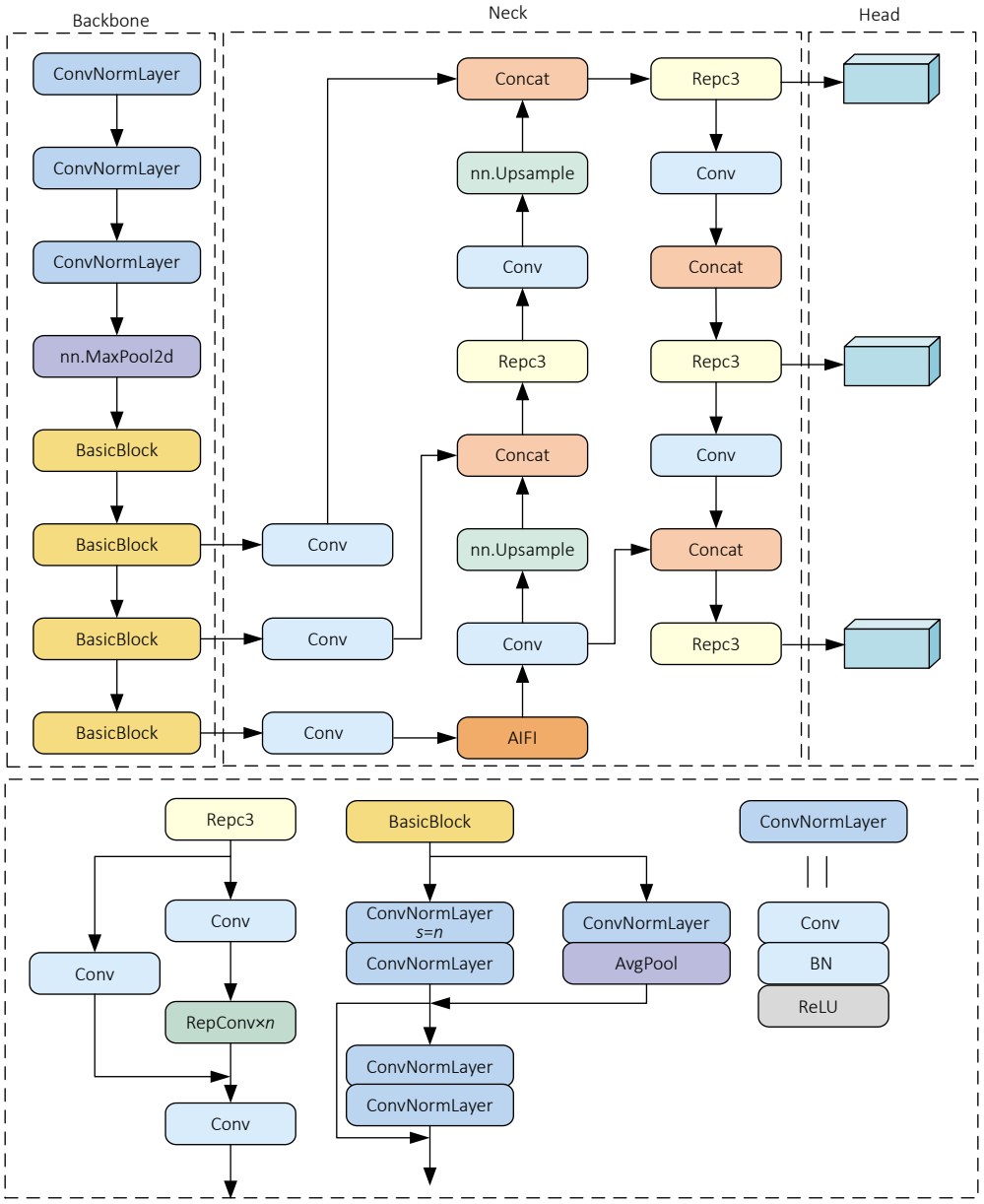


图 1 RTDETR-r18 网络结构

Fig. 1 RTDETR-r18 network architecture

特征提取骨干网络，其中 ConvNormLayer 模块包含卷积层、归一化层和激活函数，用于增强网络的感受野和非线性表达能力；nn.MaxPool2d 模块通过最大池化操作降低特征图的空间尺寸，实现特征图的空间下采样；BasicBlock 模块由两个卷积层和一个残差连接组成，有助于缓解梯度消失的问题。颈部部分采用特征金字塔网络 (FPN) 结构，用于融合多尺度信息，提高对锻件表面缺陷的感知能力。检测头引入 Transformer 结构，结合注意力机制实现端到端的目标定位与分类。

3 MCF-RTDETR 网络结构

为提高锻件表面缺陷检测的整体性能，本文针对锻件裂纹多呈长条状、分布不规律、边缘模糊等特点，基于 RTDETR-r18 算法提出以下改进：对于主干网络，将多向梯度融合卷积 (MGFConv) 集成到普通卷积层中，通过多方向差分卷积融合机制提升模型对缺陷边缘区域的感知能力，使浅层特征图能够更准确地描述细微裂纹、模糊轮廓等小目标结构，从而增强整体检测精度；在颈

部网络中，引入基于上下文感知递归特征金字塔网络结构 (CARFPN)，通过集成矩形自校准模块、金字塔上下文提取模块以及多尺度动态融合模块，实现不同尺度特征信息的有效对齐，提升模型在复杂背景下的小目标检测能力和上下文建模能力；在损失函数方面，采用融合 Focaler-IoU 动态聚焦策略与多点距离约束的 Focaler-MPDIoU 损失函数，替代原有 GIoU 损失函数，从而增强训练过程中的样本难度感知与边界回归精度，提升模型在小目标场景下的收敛速度与泛化能力。MCF-RTDETR 网络结构如图 2 所示。

3.1 MGFConv 模块

在实际工业场景中，锻件表面缺陷在生产过程中会出现形态复杂、尺度差异大、边缘模糊等特征。传统卷积结构由于感受野固定且缺乏方向性，在捕获微小目标的细节边缘与全局结构方面存在局限性，限制了小目标缺陷的检测性能。为增强特征提取网络的多尺度感知和多方向边缘的建模能力，本文设计了一种具有自适应方向感知能力的多向梯度融合卷积 (MGFConv) 模块，用于替代骨干网络中的标准卷积结构，其结构如图 3 所示。

MGFConv 采用五分支并行结构，由四个方向性差分卷积 (difference convolution, DC) 与一个标准卷积 (vanilla convolution, VC) 组成，具体包括： 1) 中心差分卷积 (center DC, CDC)，用于增强局部区域的梯度变化感知，有助于识别裂纹等细微缺陷； 2) 水平差分卷积 (horizontal DC, HDC)，突出水平方向的纹理与边缘特征； 3) 垂直差分卷积 (vertical DC, VDC)，强化垂直方向的结构响应； 4) 对角线差分卷积 (diagonal DC, DDC)，捕获斜向梯度信息，提升对复杂边界形态的适应能力； 5) 标准卷积 (VC)，用于保留全局语义信息，实现细节与整体信息的协同建模。通过多方向差分卷积与标准卷积的融合，MGFConv 能够根据缺陷在不同方向和不同尺度上的特征分布，自适应地突出最具判别力的边缘与纹理信息，从而在复杂背景下保持对细长裂纹和微小缺陷的敏感性。其计算形式可表示为

$$P_{\text{out}} = MGFConv(P_{\text{in}}) = \sum_{i=1}^{5} P_{\text{in}} * F_i = P_{\text{in}} * F_{\text{cvt}}, \quad (1)$$

式中：P_{out}为输出的特征图；P_{in}为输入的特征图；MGFConv(·)为差分卷积操作；F_i为第 i 个卷积核，*为卷积运算，F_{cvt}为经过重参数化的等效卷积核。

然而，并行部署五个卷积进行特征提取将会额外增加参数量和计算开销，降低了模型在实际工业场景中的可部署性。为此，本文提出卷积重参数化 (convolutional re-parameterization) 技术，利用卷积层的可加性，将并行的五个卷积简化成单个标准卷积核，如图 4 所示。具体计算过程如式 (2)~(4) 所示。

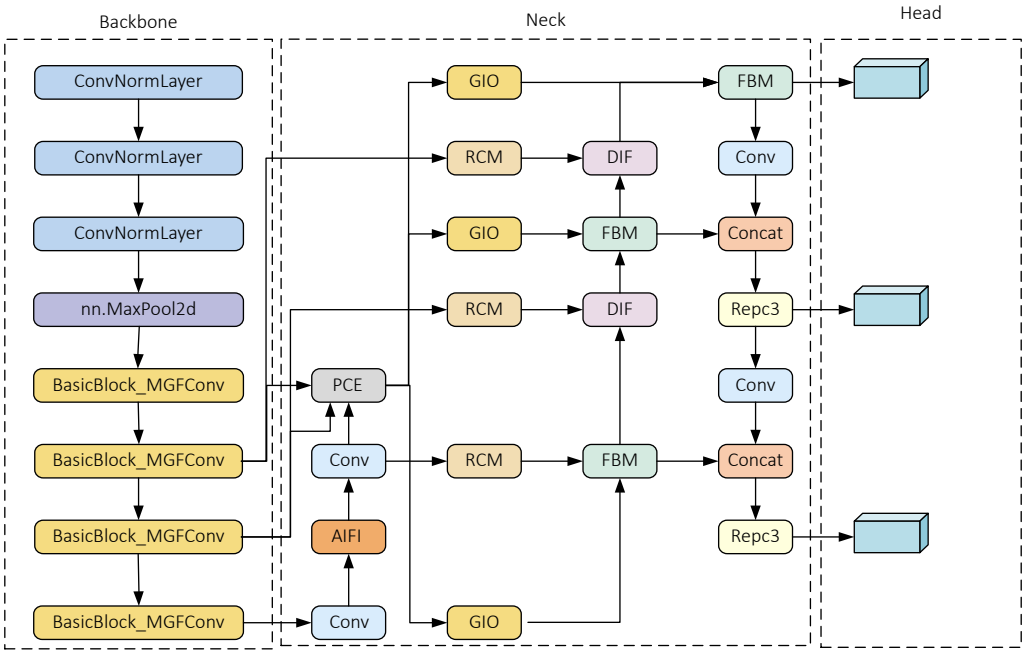


图 2 MCF-RTDETR 网络结构

Fig. 2 MCF-RTDETR network architecture

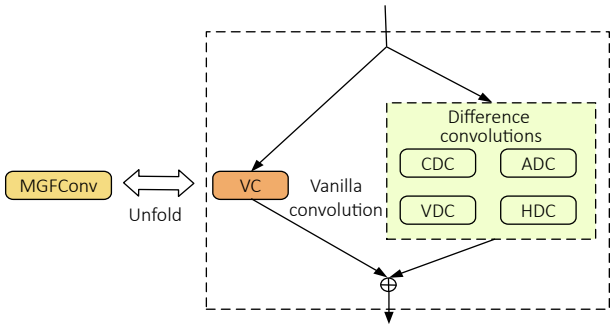


图 3 MGFConv 模块结构

Fig. 3 MGFConv module architecture

若第 i 个分支卷积核为 F_i，偏置为 b_i，分支后接 BatchNorm，其参数为 $\gamma_i, \beta_i, \mu_i, \sigma_i^2$，$\hat{F}_l$ 表示经过 BatchNorm 融合后，第 i 个分支的等效卷积核权重，$\hat{b}_l$ 表示第 i 个分支经过 BatchNorm 融合后的等效偏置项，则该分支可等效为

$$\hat{F}_l = \frac{\gamma_i}{\sqrt{\sigma_i^2+\varepsilon}} F_i,\ \hat{b}_l = \frac{\gamma_i}{\sqrt{\sigma_i^2+\varepsilon}} (b_i-\mu_i)+\beta_i\ . \tag{2}$$

在对齐所有卷积核尺寸并按元素相加后，得到最终的等效卷积核 $F_{\rm eq}$ 与偏置 $b_{\rm eq}$：

$$F_{\rm eq} = \sum_{i=1}^{5} \hat{F}_i, b_{\rm eq} = \sum_{i=1}^{5} \hat{b}_i\ . \tag{3}$$

因此，并行五分支的线性组合与单个等效卷积完全等价：

$$\boldsymbol{P}_{\rm out} = \sum_{i=1}^{5} (P_{\rm in} * F_i + b_i) = P_{\rm in} * F_{\rm eq} + b_{\rm eq}\ . \tag{4}$$

在训练过程中，五个并行卷积分支独立进行特征提取，并通过反向传播的链式法则分别更新其卷积核权重，使模型能够在整体训练中自动学习各方向分支的重要性，从而对不同方向的响应强弱进行隐式选择。在推理阶段，这些卷积核的参数保持固定，通过卷积重参数化技术将五个分支等效折叠为单个标准卷积核，该等效核继承了训练过程中形成的方向偏置，无需额外超参数或阈值即可实现方向自适应，减少模型的结构冗余。由于训练阶段和推理阶段均包含前向传播计算，重参数化技术能够同时提高训练效率和推理速度，增强模型在实际工业场景中部署的灵活性。在给定输入特征图的情况下，以与标准卷积相同的计算成本和推理时间进行输出特征图。

3.2 CARFPN 模块

在传统特征金字塔网络 (feature pyramid network, FPN) 中，特征融合主要依赖自顶向下的特征传递和横向连接，但这种结构在空间特征重建及多尺度上下文信息提取方面存在一定局限性。为提升多尺度特征的表示能力，本文提出一种具备自适应融合机制的上下文感知递归特征金字塔网络 (CARFPN)，其结构如图 5 所示。

CARFPN 采用逐层下采样的编码器-解码器架构，输入四个尺度的特征图 $\{P_1, P_2, P_3, P_4\}$，其分辨率分别为 $\left[\frac{H}{4}\times\frac{W}{4},\ \frac{H}{8}\times\frac{W}{8},\ \frac{H}{16}\times\frac{W}{16},\ \frac{H}{32}\times\frac{W}{32}\right]$。为降低计算复杂度，解码阶段舍弃最高分辨率 P_1，仅对 P_2、P_3 和 P_4 进行平均池化，使其统一至 $\frac{H}{16}\times\frac{W}{16}$ 分辨率后进行拼接，

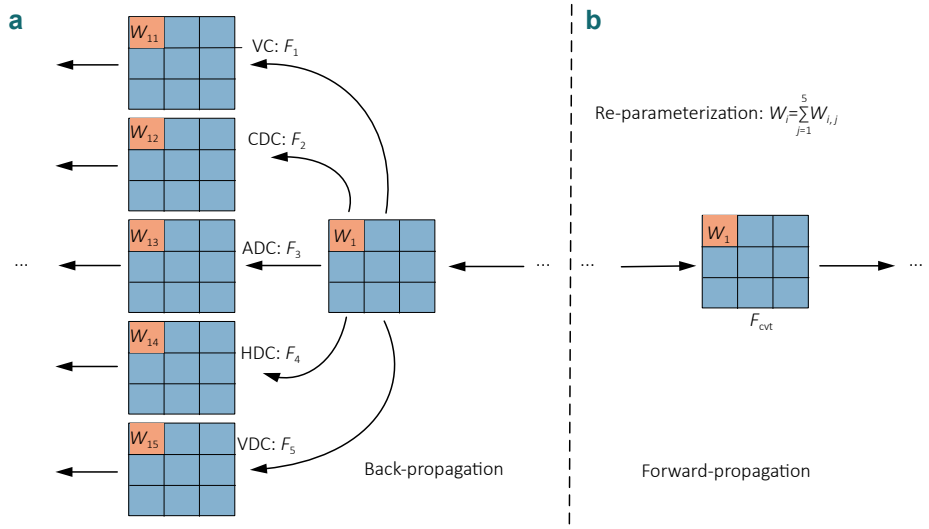


图 4 重参数化技术

Fig. 4 Re-parameterization technique

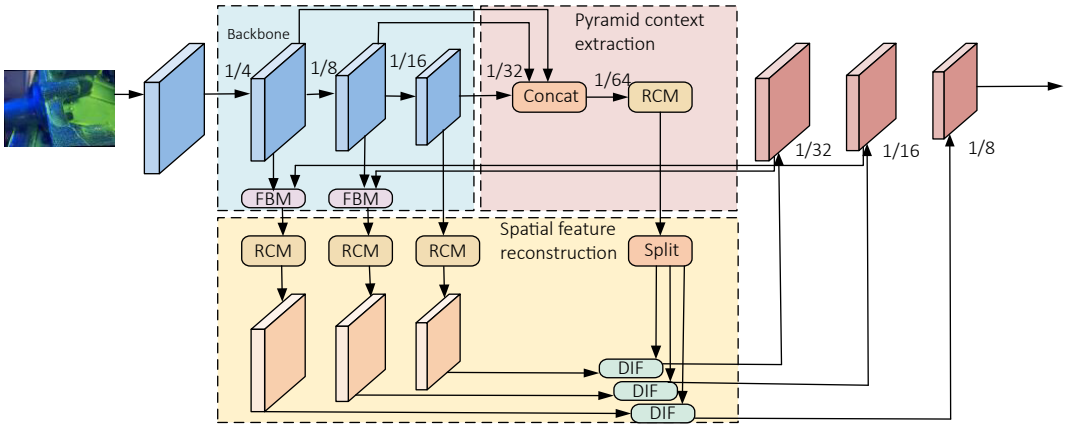


图 5 CARFPN 模块结构

Fig. 5 CARFPN module architecture

得到融合特征P_5。随后，P_5通过矩形自校准模块(rectangular self-calibration module, RCM)增强上下文提取能力，并最终恢复至原始通道数，实现空间特征重建。RCM 模块采用大核条卷积分别在水平与垂直方向建模方向感知信息，从而自适应强化对长条裂纹、划痕等方向性缺陷的建模能力，如图 6 所示。大核条卷积能够有效扩展感受野、增强方向敏感性，同时兼顾计算效率与建模深度，满足工业图像中复杂纹理的表达需求。计算过程如下所示：

$$\xi_{\mathrm{C}}(\bar{y})=\delta(\psi_{k\times1}(\phi(\psi_{1\times k}(\bar{y})))), \tag{5}$$

$$\xi_{\mathrm{P}}(x,y)=\psi_{3\times3}(x)\odot y, \tag{6}$$

$$\boldsymbol{P}_{\mathrm{out}}=\rho(\xi_{\mathrm{P}}(x,\xi_{\mathrm{C}}(H_{\mathrm{F}}(x)\oplus V_{\mathrm{F}}(x))))+\boldsymbol{P}_{\mathrm{in}}, \tag{7}$$

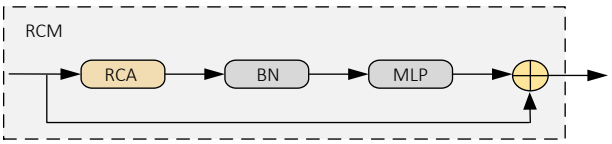


图 6 RCM 模块

Fig. 6 RCM module

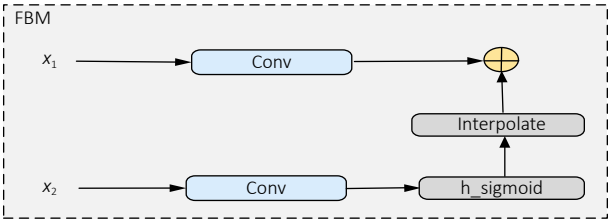


图 7 FBM 模块

Fig. 7 FBM module

式中：$\xi_{\mathrm{C}}(\cdot)$表示上下文权重生成函数；$\xi_{\mathrm{P}}(\cdot,\cdot)$表示特征调制函数；$\delta$表示 h_sigmoid 函数；$\psi$表示大核条卷积；$k$为条卷积长度；$\phi$由 BN 与 ReLU 组成；$\odot$代表逐元素乘法；$H_{\mathrm{F}}(x)$和$V_{\mathrm{F}}(x)$分别为水平和垂直方向的池化；$\rho$表示 BN 和 MLP 构成的线性映射函数；$\oplus$代表逐元加法。

为进一步增强不同尺度特征间的信息交互与空间对齐能力，CARFPN 在编码器路径中引入 FuseBlockMulti (FBM) 模块，如图 7 所示。FBM 模块通过采用 1×1 卷积对浅层和深层特征进行通道压缩与维度统一，降低冗余信息的干扰。随后，深层特征经过激活函数处理，并通过双线性插值上采样至浅层特征的空间尺寸，作为引导权重参与融合。在融合过程中，深层语义信息以动态加权方式调节浅层空间细节响应，从而实现自适应的语义一致性与边缘特征增强。在解码器路径中采用 DIF (dynamic interpolation fusion) 模块，如图 8 所示。该模块通过双线性插值上采样至浅层特征的空间分辨率，并通过 1×1 卷积完成通道数匹配。随后，两者通过逐元素相加实现高低层特征的融合，有效缓解尺度不一致带来的语义错配问题。DIF 模块能够在不同尺度特征间自适

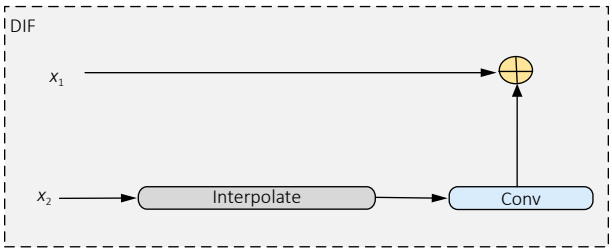


图 8 DIF 模块

Fig. 8 DIF module

应调整权重比例，从而增强多尺度语义建模能力，并提升特征在复杂背景下的鲁棒性。CARFPN 的整体计算过程如下：

$$F = RCM(AF(P_2,8),AF(P_3,4),AF(P_4,2)), \tag{8}$$

式中：RCM 表示矩形自校准模块；AF 表示下采样函数；P 表示对特征图 P 进行 x 倍平均池化下采样。

3.3 Focaler-MPDIoU

RTDETR-r18 采用 Generalized IoU (GIoU) 通过预测框与真实框之间最小外接矩形面积的差异来衡量其偏离程度。然而，GIoU 未能有效建模目标的长宽比例与中心点偏移，且缺乏对回归样本难度的感知能力，容易导致梯度分布不均和回归不稳定等问题。为此，本文提出一种融合 Focaler-IoU 的动态加权机制与 MPDIoU 多点结构约束的边界框回归损失函数—Focaler-MPDIoU，如图 9 所示。

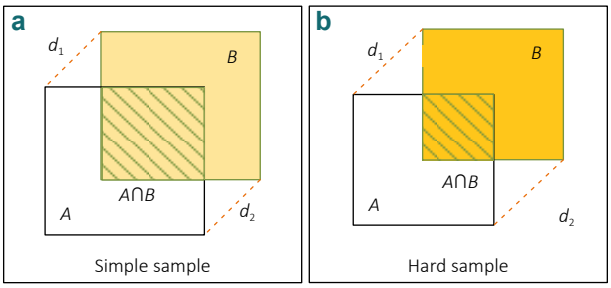


图 9 Focaler-MPDIoU 边框处理

Fig. 9 Bounding box processing with Focaler-MPDIoU

Focaler-IoU[17] 通过引入预测误差的动态加权机制，并结合线性区间映射策略，能够自适应调整不同 IoU 区间样本的回归损失聚焦程度，使模型能够根据预设阈值主动筛选难易样本。其加权函数定义如下：

$$IoU^{\text{focaler}} = \begin{cases} 0, & IoU < d \\ \dfrac{IoU-d}{u-d}, & d \ll IoU \ll u \\ 1, & IoU > u \end{cases}, \tag{9}$$

式中：d 与 u 分别为下限与上限阈值。通过调整这两个参数，可控制损失函数对不同难度样本的聚焦程度：增大 d 会提升对易样本的权重，而增大 u 则增强对难样本的权重。其对应损失函数定义为

$$L_{\text{Focaler - IoU}} = 1 - IoU^{\text{focaler}}. \tag{10}$$

Multi-point distance IoU (MPDIoU)[18] 在传统 IoU 损失基础上，引入了预测框与真实框多点之间的空间距离对齐约束，综合考虑了交并比、中心距离及宽高偏差等因素。其计算方式如下：

$$MPDIoU = \frac{A \cap B}{A \cup B} - \frac{d_1^2}{w^2+h^2} - \frac{d_2^2}{w^2+h^2}, \tag{11}$$

$$L_{MPDIoU} = 1 - MPDIoU = 1 + \frac{d_1^2}{d^2} + \frac{d_2^2}{d^2}, \tag{12}$$

式中：A 与 B 分别表示真实框面积和预测框面积；d_1 与 d_2 分别表示预测框中心点与两端角点之间的欧氏距离；w 与 h 为目标框的宽度与高度；d 为归一化尺度因子。两者融合后的改进损失函数定义如下：

$$L_{\text{Focaler}}\text{-MPDIoU} = L_{\text{MPDIoU}} + IoU - IoU^{\text{Focaler}}. \tag{13}$$

综上所述，Focaler-MPDIoU 在强化边界框几何一致性约束的基础上，引入了样本难度感知的自适应聚焦调控机制，有效提升了模型在复杂场景下对锻件表面微小缺陷的检测性能，增强了模型的泛化能力与适应性。

4 实验设计与分析

4.1 实验数据集

本文选取湖北三环锻造有限公司探伤车间的重型卡车转向节磁粉探测图像作为数据集，用以验证本文提出的改进模型的有效性[19]。图像由固定式工业相机采集，分辨率为 1920 pixel×1080 pixel，采集过程中综合考虑不同角度、距离与光照条件，尽可能还原实际检测环境。原始图像数量为 362 张，为解决数据量不足问题，在标注后采用旋转、翻转和亮度变化等方法进行数据增强，增强过程使用 Augly 图像增强库[20]，参数设置包括：旋转角度为±20°，亮度调整范围为原图的 0.8 ~ 1.2 倍，上下/左右翻转各占 50%。经过增强处理后，共得到 4622 张包含缺陷的图片，其中训练集包含 3206 张图片，标注 11700 个缺陷目标，验证集与测试集相同包含 1416 张图片，标注 3792 个缺陷目标，所有缺陷类型均统一标注为 defect 类别。数据集中部分图片如图 10 所示。

4.2 实验环境

本文实验配置环境为 Windows 10 操作系统，CPU 使用 Intel Xeon E5-2680，GPU 采用双 NVIDIA GeForce RTX 3060 (12 GB 显存)；深度学习框架使用 Pytorch 2.1.0 版本，Python 版本为 3.9，CUDA 版本为 11.8。在实验参数设置方面，输入图像尺寸为 640 pixel×640 pixel，

训练轮数设为 *epochs*=300，批处理大小设为 s_{batch}=16，IoU 设置为 0.5，其余均保持基准模型默认参数。

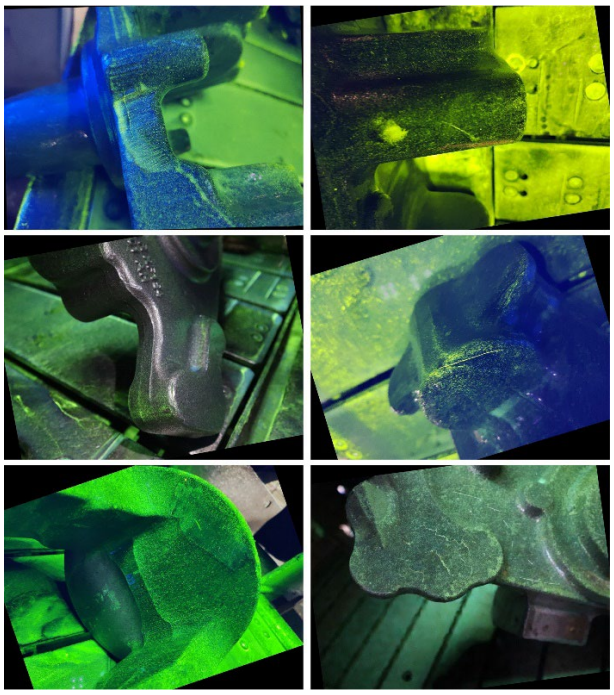


图 10 重型卡车转向节磁粉探测数据集

Fig. 10 Magnetic particle inspection dataset of heavy truck steering knuckles

4.3 评估指标

为分析改进后的模型在锻件表面缺陷检测任务中的性能，本文选取平均精度均值 (mean average precision, mAP)、准确率 (precision, *P*) 和召回率 (recall, *R*) 等作为评估 MCF-RTDETR 模型性能的关键指标。其中，mAP 通过计算所有类别的 AP@0.5 值的平均值来评估检测精度，能够综合反映模型在检测任务中的整体性能，准确率用于衡量模型预测结果的准确性，召回率则反映了模型检测的完整性。各项指标的计算公式如下：

$$P=\frac{TP}{TP+FP}, \tag{14}$$

$$R=\frac{TP}{TP+FN}, \tag{15}$$

$$AP=\int_0^1 P(R)\mathrm{d}R, \tag{16}$$

$$mAP=\frac{1}{n}\sum_{i=1}^{n}AP(R)_i, \tag{17}$$

式中：TP表示预测正确的目标样本个数；FP表示预测错误的目标样本个数；FN表示没有检测出的目标；n 表示类别总数；AP_i表示第 i 类别的平均精度。

4.4 验证不同条形卷积核参数

为验证矩形自校准模块 (RCM) 中条形卷积核长度参数 k 对模型性能的影响，本文基于 MCF-RTDETR 模型，在保持其他结构与训练策略不变的前提下，分别设置 k=3、5、7、9、11、13 六种条形卷积核长度，分析其对模型检测精度、推理速度及计算复杂度的影响。该参数将直接影响卷积操作在不同方向上的感受野大小与方向感知能力，进而影响模型对细长裂纹、划痕等典型缺陷目标的检测能力。实验结果如表 1 所示。

由表 1 可知，k=7 时，模型在精度和效率之间取得了最佳平衡。此时 mAP 达到 88.5%，准确率和召回率分别为 88.2% 和 83.1%，推理速度维持在 137.2 f/s，计算资源消耗处于合理范围内 (33.5 GFLOPs)。相比之下，较小卷积核 (k=3、5) 的推理速度更快和计算复杂度更低，但检测精度下降明显；较大卷积核 (k= 9、11、13) 虽在精度上略有提升，但计算复杂度和推理延迟显著增加，不利于在资源受限的工业场景中部署。

综上所述，设定卷积核长度为 k=7 可在精度与效率之间取得更优，兼顾方向感知能力与计算代价。因此，本文在 RCM 模块中采用 k=7 作为默认配置参数，用于后续模型的训练与推理。

表 1 不同条形卷积核参数对比试验

Table 1 Comparison of different bar-shaped convolution kernel lengths

k	*P*/%	*R*/%	mAP/%	FPS	Parameters/M	FLOPs/G
3	85.9	80.1	86.2	145.6	23.8	30.1
5	86.6	80.9	86.8	142.3	24.2	31.2
7	88.2	83.1	88.5	137.2	25.7	33.5
9	88.2	83.3	88.7	128.5	27.1	36.4
11	88.4	83.6	88.9	118.7	28.3	38.1
13	88.9	83.8	89.1	110.2	31.4	41.3

4.5 验证不同损失函数有效性

本文根据数据集特性，修改了骨干网络和颈部网络，分别引入 MGFConv 模块和 CARFPN 模块，从而使得准确率提高到 87.9%，召回率提高到 83.5%，平均检测精度提高到 88.1%。为评估不同边界框回归损失函数对模型优化效果的影响，本文选取了多种 IoU 损失函数进行对比实验，实验结果如表 2 所示。结果表明，引入不同损失函数后，各项指标均有所变化，其中 Focaler-MPDIoU 的准确率和平均准确率表现最优，分别达到了 88.2% 和 88.5%。基于表 2 实验结果，本文最终选用 Focaler-MPDIoU 作为边界框回归损失函数。

4.6 消融实验

为了验证本文所提出的各个改进模块的有效性，设计了 7 组消融实验来评估不同改进模块对模型性能的影响。通过依次添加各个改进模块进行实验，结果如表 3 所示，其中“√”表示该实验中引入了相应模块，“—”则表示未使用该模块。

实验结果如表 3 所示，引入 MGFConv 模块增强骨干网络后，与基准模型相比，精确率、召回率和 mAP 分别提高了 0.7%、0.3% 和 1.4%。该模块通过多方向差分卷积分支显式建模水平、垂直及对角方向的梯度变化，使模型在训练过程中能够自适应调整不同方向分支的重要性，从而强化了对小目标、细长裂纹以及微小缺陷边缘的表征能力。尽管该模块在训练阶段引入了多分支卷积，导致参数量上升至 26.8 M，但借助卷积重参数化技术，在推理阶段将多分支折叠为单核，使计算复杂度从 58.3 GFLOPS 降低至 43.2 GFLOPS，保证了部署效率。CARFPN 模块的引入使精确率和 mAP 指标分别提高了 0.6% 和 0.7%，其递归上下文建模与动态融合机制有效缓解了多尺度特征在空间和语义上的对齐不足，使浅层细节与深层语义实现更充分的互补，在小目标检测和复杂背景下表现出更强的鲁棒性。替换 Focaler-MPDIoU 为损失函数后，召回率提升至 83.2%，mAP 提高至 87.1%，该损失函数融合了难样本聚焦机制与多点几何约束，能够自适应调整不同 IoU 区间样本的回归权重，

表 2　不同损失函数对比试验

Table 2　Comparative experiments with different loss functions

Loss function	*P*/%	*R*/%	mAP/%
SIoU	85.8	83.0	87.6
DIoU	86.0	**83.9**	87.3
Focaler-GIoU	87.3	81.7	87.0
Focaler-Shape-IoU	84.7	82.4	87.7
Inner-MPDIoU	86.6	81.3	85.9
Inner-SIoU	87.4	83.1	87.3
Inner-Shape-IoU	87.2	83.7	87.9
Focaler-MPDIoU	**88.2**	83.1	**88.5**

表 3　消融实验

Table 3　Results of ablation detection experiment

Algorithms	Network			*P*/%	*R*/%	mAP/%	FPS	Parameters/M	FLOPs/G
	MGFConv	CARFPN	Focaler-MPDIoU						
A	—	—	—	87.0	82.8	86.2	120.7	20.1	58.3
B	√	—	—	87.7	83.1	87.6	139.2	26.8	43.2
C	—	√	—	87.6	82.7	86.9	**143.1**	**19.3**	48.6
D	—	—	√	87.2	83.2	87.1	127.0	20.1	58.3
E	√	√	—	87.9	**83.5**	88.1	136.4	25.7	**33.5**
F	√	—	√	87.6	82.3	88.3	138.1	26.8	43.2
G	—	√	√	87.8	83.3	88.0	140.5	**19.3**	48.6
MCF-RTDETR	√	√	√	**88.2**	83.1	**88.5**	137.2	25.7	**33.5**

并同时考虑预测框与真实框的多点距离一致性，从而提升了边界回归的稳定性与整体检测精度。此外，该损失函数仅在训练阶段参与优化，而不涉及推理阶段的前向计算，因此在提升精度的同时并未增加推理阶段的计算复杂度。

综上所述，MCF-RTDETR 各模块的引入显著提升了锻件表面裂纹的检测精度，在将 mAP 提升至 88.5% 的同时，仍保持较低的计算复杂度与优异的推理速度(*FPS*>100)。该模型在精度与速度之间实现了良好平衡，具备良好的工程实用性。

4.7 对比实验

为了验证所提出的改进模型 MCF-RTDETR 在锻件表面缺陷检测任务中的综合性能，本文将与其他几种目前应用最广泛的检测模型进行对比实验，包括 Faster R-CNN、YOLOv5s、YOLOv8s、YOLOv11s、YOLOv11、DETR、Deformable-DETR 和 RTDETR-r18，以及工业缺陷检测算法 SLF-YOLO[21]、PCBA-YOLO[22]、LW-DETR[23] 和 SCP-DETR[24] 进行性能对比。所有实验均在相同的数据集上进行训练，训练过程中保持统一的训练策略、超参数配置及硬件环境，以确保结果的公平性和可靠性。各模型的具体实验结果如表 4 所示。

由表 4 可知，MCF-RTDETR 在各项主要性能指标上均优于其他对比模型。其中，精确率和召回率分别达到了 88.2% 和 83.1%，平均检测精度提升至 88.5%。相较于 Faster R-CNN、YOLOv5s、YOLOv8s、YOLOv11s、YOLO11l、DETR、Deformable-DETR 和 RTDETR-r18，MCF-RTDETR 的平均检测精度分别提高了 4.5%、3.7%、4.0%、2.9%、2.3%、3.3%、1.8% 和 2.3%。与同样应用于工业缺陷检测模型 SLF-YOLO、PCBA-YOLO、LW-DETR 和 SCP-DETR 相比，MCF-RTDETR 的平均精度分别提升了 3.9%、3.2%、3.6% 和 2.5%。尽管因引入 MGFConv 的方向性卷积分支使得 MCF-RTDETR 模型的参数量和原模型相比有所增加，但其计算复杂度由 58.3 GFLOPS 下降至 33.5 GFLOPS，在保证检测精度的同时有效降低了计算开销，提升了检测效率。此外，MCF-RTDETR 的推理速度达到 137.2 f/s，远高于 Faster R-CNN、DETR、Deformable-DETR 等目标检测算法，具有较强的实时检测性能。综上所述，MCF-RTDETR 模型在保持较高检测精度的情况下，兼顾了模型的效率和部署成本，满足实际工业场景中的实时检测需求。

如图 11 所示，对比了 YOLOv11s、Deformable DETR、RTDETR-r18 和 MCF-RTDETR 四种模型在测试集上的检测效果。由图 11(d) 可知，MCF-RTDETR 在多种复杂场景下的检测效果最稳定和准确。尤其是在锻件表面存在微小裂纹、纹理复杂及背景干扰显著的情况下，MCF-RTDETR 的检测框能够更精确地贴合缺陷区域，并保持较高的置信度。相比之下，YOLOv11s (图 11(a)) 在部分

表 4 对比实验结果

Table 4 Comparative experimental results

Algorithms	*P*/%	*R*/%	mAP/%	FPS	Parameters/M	FLOPs/G
Faster R-CNN	86.1	79.4	84.0	18.6	40.0	207.0
YOLOv5s	85.7	80.1	84.8	124.5	**9.1**	24.0
YOLOv8s	83.7	79.1	84.5	128.2	11.1	28.4
YOLOv11s	84.2	83.5	85.6	132.0	9.4	21.5
YOLOv11l	85.0	**84.1**	86.2	105.3	25.4	87.6
DETR	86.0	80.5	85.2	22.0	40.8	225.0
Deformable DETR	87.4	82.0	86.7	24.5	40.0	196.0
RTDETR-r18	87.0	82.8	86.2	120.7	20.1	58.3
SLF-YOLO[21]	85.9	80.7	84.6	125.0	9.7	24.6
PCBA-YOLO[22]	86.4	82.5	85.3	122.4	12.4	**20.8**
LW-DETR[23]	85.3	80.2	84.9	30.0	28.2	42.8
SCP-DETR[24]	86.8	82.3	86.0	47.8	20.2	66.1
MCF-RTDETR	**88.2**	83.1	**88.5**	**137.2**	25.7	33.5

样本中出现漏检现象，Deformable DETR (图 11(b)) 的检测框出现不同程度的偏移，而 RTDETR-r18 (图 11(c)) 在微小缺陷识别上的检测效果不如 MCF-RTDETR。可视化结果与表 4 的量化指标一致，进一步验证了 MCF-RTDETR 在锻件表面缺陷检测任务中的优越性能。

4.8 泛化性实验

为了进一步验证本文提出的改进模型在实际工业场景中的泛化性能，本文选取在对比实验中整体 mAP 表现较优的几类模型，包括 YOLOv11l、Deformable DETR、RTDETR-r18 和 SCP-DETR[24]，与所提出的 MCF-RTDETR 在真实工业数据集 GC10-DET[25] 上进行泛化性对比实验。该数据集包含 3570 张灰度图像，涵盖十种典型表面缺陷类型，分别为冲孔 (Pu)、焊缝 (Wl)、新月形缝隙 (Cg)、水斑 (Ws)、油斑 (Os)、夹杂物 (In)、丝斑 (Ss)、轧坑 (Rp)、折痕 (Cr) 和腰部折痕 (Wf)，如图 12

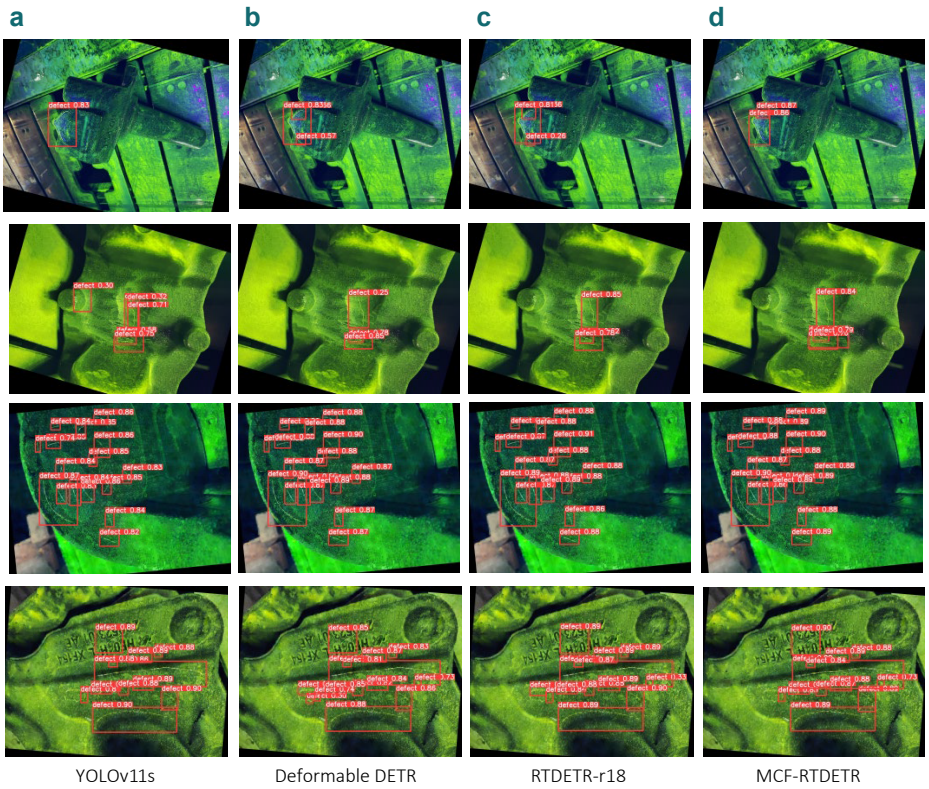


图 11 检测效果对比

Fig. 11 Comparison of detection effects

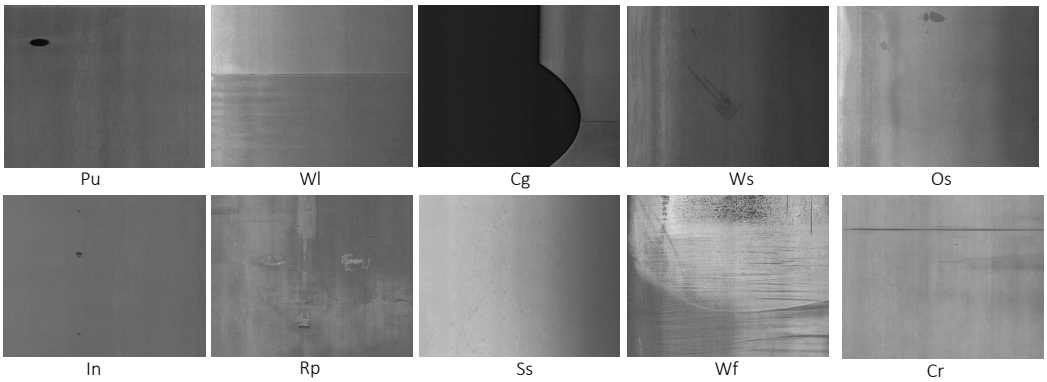


图 12 GC10-DET 数据集示例图

Fig. 12 Sample diagram of the GC10-DET dataset

所示。

实验结果如表 5 所示，MCF-RTDETR 在 GC10-DET 数据集的大多数缺陷类别中均优于对比模型，整体 mAP 由 69.4% 提升至 72.2%。其中，夹杂物 (In) 缺陷通常呈现不规则小块或点状区域，而 MCF-RTDETR 所引入的改进结构更擅长捕获方向性边缘特征，因此在该类别上的 AP 为 41.3%，略低于 RTDETR-r18 的 42.5%。相比之下，对于折痕 (Cr) 和丝斑 (Ss) 等长条状或边缘特征突出的缺陷类别，检测精度提升最为明显，因为 MGFConv 模块中的多方向差分卷积结构能够增强模型对梯度变化的敏感性，从而有效捕获复杂边缘信息，提升对细长裂纹和微小缺陷的识别能力。对于焊缝 (Wl) 与水斑 (Ws) 等大面积或尺度变化明显的缺陷类别，主要由于 CARFPN 模块的上下文递归融合机制，该机制通过动态调节不同尺度特征权重，有效缓解了背景复杂性与尺度不均衡问题。此外，自适应损失函数在训练阶段针对难检样本和类别不均衡进行了差异化加权，进一步提升了整体检测精度与鲁棒性。整体实验表明，所提出的方法不仅在局部缺陷检测上表现出明显优势，而且在整体性能上展现出更强的泛化能力，验证了其在真实工业缺陷检测场景中的应用潜力。

为更直观地对比 MCF-RTDETR 与 RTDETR-r18 在 GC10-DET 数据集上的检测效果，引入热力图可视化方法对两类模型进行分析，如图 13 所示。结果表明，MCF-RTDETR 在相同测试场景下展现出更加清晰且准确的响应区域，其注意力热力图分布能够更好地贴合缺陷真实轮廓。可视化结果显示，RTDETR-r18 在 Wl 与 In 缺陷类别上存在漏检现象，主要由缺陷边缘模糊或尺度过小导致的特征混淆；在 Cg 缺陷类别上则出现误检，将背景纹理错误识别为缺陷区域。相比之下，MCF-RTDETR 通过 MGFConv 模块的多方向差分卷积结构与 CARFPN 的上下文递归融合机制，能够更有效地捕获细长裂纹和复杂边缘特征，同时动态抑制背景干扰，从而显著降低了误检与漏检的发生率。

5 结　论

针对传统人工磁粉检测在重卡转向节表面裂纹检测中存在效率低、依赖人工经验强和安全性差等问题，本文提出一种改进的 RTDETR 检测模型—MCF-RTDETR。通过引入多向梯度融合卷积 MGFConv 模块，有效提升了对缺陷边缘信息的提取能力，提高模型对小尺度缺陷区域的感知精度；通过在颈部网络中集成 CARFPN 模块，重构特征金字塔结构，强化上下文引导的多尺度特征融合，进一步提升了模型在复杂背景下的检测精度；同时，通过融合 Focaler-IoU 与 MPDIoU 损失函数，提出 Focaler-MPDIoU 边界框回归损失函数，增强了模型对困难样本的聚焦能力，从而提高了回归精度。

实验结果表明，MCF-RTDETR 用于锻件表面缺陷检测，在准确率、召回率及平均检测精度方面均优于基准模型和主流检测算法，且在真实工业数据集上的泛化能力得到有效验证。结合工业部署需求，在当前配置环境下，模型在输入尺寸为 640×640 时可达 137.2 FPS/s，平均单帧推理延迟约为 7.4 ms，能够满足工业相机每秒 30 帧的实时检测需求，具备在实际工业边缘设备上部署的可行性与实用性。

尽管 MCF-RTDETR 模型在多数工业检测任务中表现出良好的检测性能与实时性，但由于引入多分支特征提取与上下文增强模块，模型参数量有所增加。未来研究将围绕模型轻量化展开，重点探索结构剪枝、知识蒸馏、轻量卷积替代等技术路径，以进一步提升模型部署灵活性与跨平台兼容性。同时，计划将模型拓展应用至

表 5　泛化实验结果

Table 5　Results of generalization experiments

Algorithms	mAP/%	AP@0.5/%									
		Pu	Wl	Cg	Ws	Os	In	Ss	Rp	Cr	Wf
YOLOv11l	66.9	81.5	96.0	95.1	73.4	54.2	41.1	52.3	36.9	49.8	88.6
Deformable DETR	68.7	83.4	97.3	95.8	75.6	55.9	43.7	53.8	39.5	51.6	90.2
RTDETR-r18	69.4	84.5	98.2	96.1	78.4	56.0	42.5	55.4	38.8	54.8	89.6
SCP-DETR[24]	70.0	85.4	97.6	95.9	77.8	56.8	**44.6**	55.3	42.8	53.4	90.7
MCF-RTDETR	**72.2**	**89.0**	**98.6**	**96.7**	**79.2**	**59.7**	41.3	**58.3**	**49.3**	**57.5**	**92.8**

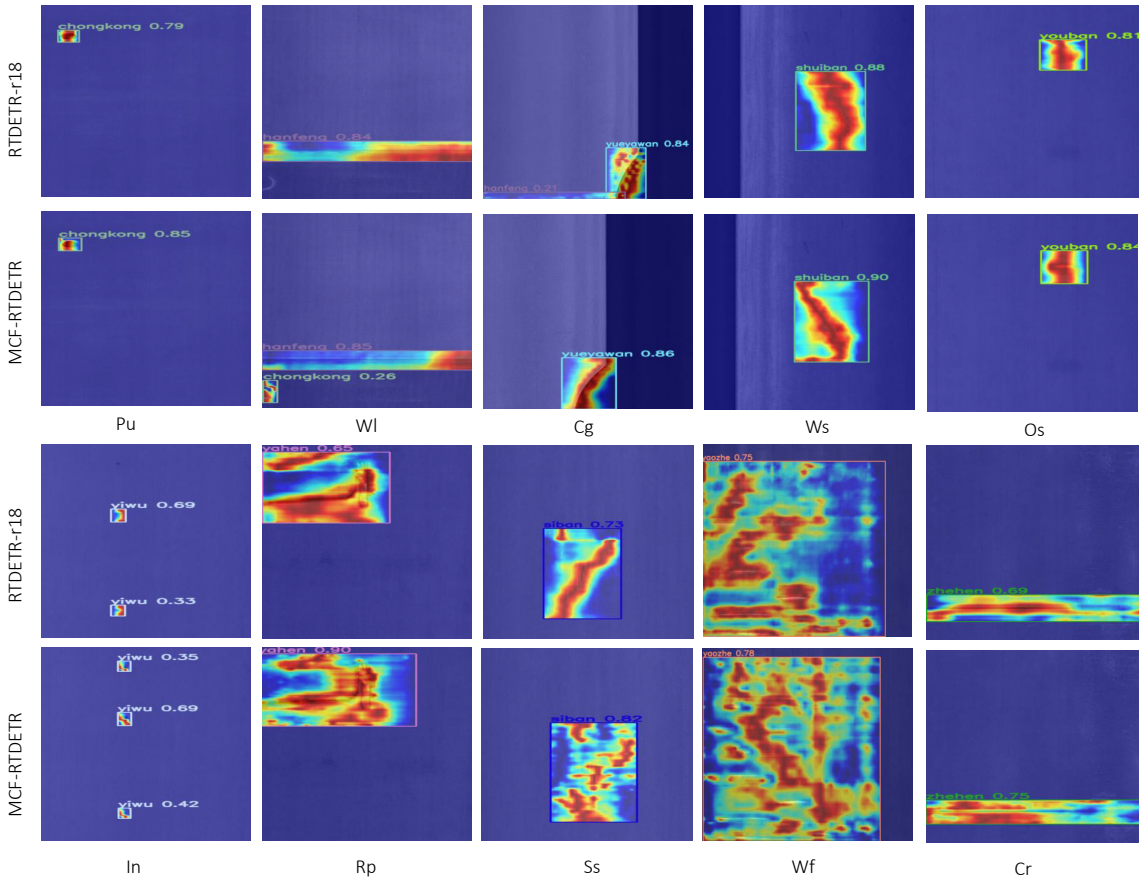


图 13 在 GC10-DET 数据集上的可视化对比

Fig. 13 Visualization comparison on the GC10-DET dataset

其他锻造部件与检测模态中，增强其在多源感知融合下的通用性与实用性。

作者贡献

梁丹：研究资料的收集、分析，论文写作；张上、管慧攀：指导论文选题、论文修改。

利益冲突

所有作者声明无利益冲突

参考文献

1. Jiang P, Fang G, Cao S J, et al. FEM analysis of the extrusion-forging compound process for knuckle of Benz heavy truck[J]. *Forg Stamp Technol*, 2006, **31**(1): 22−27.
蒋鹏, 方刚, 曹世金, 等. 奔驰重卡转向节挤压锻造复合工艺的有限元分析[J]. 锻压技术, 2006, **31**(1): 22−27.
2. Zhang H, Song Y N, Wang Y N, et al. Review of rail defect non-destructive testing and evaluation[J]. *Chin J Sci Instrum*, 2019, **40**(2): 11−25.
张辉, 宋雅男, 王耀南, 等. 钢轨缺陷无损检测与评估技术综述[J]. 仪器仪表学报, 2019, **40**(2): 11−25.
3. Wang C, Tang Y, Zhang X F, et al. An intelligent magnetic particle testing method for forgings based on the improved EfficientNet[J]. *Chin J Sci Instrum*, 2021, **42**(9): 89−96.
王宸, 唐禹, 张秀峰, 等. 基于改进 EfficientNet 的锻件磁粉探伤智能检测方法研究[J]. 仪器仪表学报, 2021, **42**(9): 89−96.
4. Zhao Z K. Review of non-destructive testing methods for defect detection of ceramics[J]. *Ceram Int*, 2021, **47**(4): 4389−4397.
5. Zhao L Y, Wu Y Q. Research progress of surface defect detection methods based on machine vision[J]. *Chin J Sci Instrum*, 2022, **43**(1): 198−219.
赵朗月, 吴一全. 基于机器视觉的表面缺陷检测方法研究进展[J]. 仪器仪表学报, 2022, **43**(1): 198−219.
6. Tao X, Hou W, Xu D. A survey of surface defect detection methods based on deep learning[J]. *Acta Autom Sin*, 2021, **47**(5): 1017−1034.
陶显, 侯伟, 徐德. 基于深度学习的表面缺陷检测方法综述[J]. 自动化学报, 2021, **47**(5): 1017−1034.
7. Girshick R. Fast R-CNN[C]//*Proceedings of 2015 IEEE International Conference on Computer Vision*, Santiago, Chile, 2015: 1440–1448. https://doi.org/10.1109/ICCV.2015.169.
8. Ren S Q, He K M, Girshick R, et al. Faster R-CNN: towards real-time object detection with region proposal networks[J]. *IEEE Trans Pattern Anal Mach Intell*, 2017, **39**(6): 1137−1149.
9. Redmon J, Divvala S, Girshick R, et al. You only look once: unified, real-time object detection[C]//*Proceedings of 2016 IEEE*

Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 2016: 779–788. https://doi.org/10.1109/CVPR.2016.91.

10. Liu W, Anguelov D, Erhan D, et al. SSD: single shot MultiBox detector[C]//*Proceedings of the 14th European Conference on Computer Vision*, Amsterdam, The Netherlands, 2016: 21–37. https://doi.org/10.1007/978-3-319-46448-0_2.
11. Leng Y F, Liu Z, Xu B Y, et al. Detection of steel surface defect based on improved Faster R-CNN[J]. *Mech Sci Technol Aerosp Eng*, 2025, **44**(1): 75−83.
冷岳峰, 刘正, 徐宝祎, 等. 改进 Faster R-CNN 的钢材表面缺陷检测[J]. 机械科学与技术, 2025, **44**(1): 75−83.
12. Guo Z X. Surface defect detection based on deep learning methods[D]. Beijing: Beijing University of Posts and Telecommunications, 2022.
果泽萱. 基于深度学习的表面缺陷检测[D]. 北京: 北京邮电大学, 2022.
13. Zhao Y, Lv W Y, Xu S L, et al. DETRs beat YOLOs on real-time object detection[C]//*Proceedings of 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Seattle, WA, USA, 2024: 16965–16974. https://doi.org/10.1109/CVPR52733.2024.01605.
14. Tian F, Song C, Liu X P. Small target detection in coal mine underground based on improved RTDETR algorithm[J]. *Sci Rep*, 2025, **15**(1): 12006.
15. Yu C W, Shin Y. MCG-RTDETR: multi-convolution and context-guided network with cascaded group attention for object detection in unmanned aerial vehicle imagery[J]. *Remote Sens*, 2024, **16**(17): 3169.
16. Carion N, Massa F, Synnaeve G, et al. End-to-end object detection with transformers[C]//*Proceedings of the 16th European Conference on Computer Vision*, Glasgow, UK, 2020: 213–229. https://doi.org/10.1007/978-3-030-58452-8_13.
17. Zhang H, Zhang S J. Focaler-IoU: more focused intersection over union loss[Z]. arXiv: 2401.10525, 2024. https://doi.org/10.48550/arXiv.2401.10525.
18. Ma S L, Xu Y. MPDIoU: a loss for efficient and accurate bounding box regression[Z]. arXiv: 2307.07662, 2023. https://doi.org/10.48550/arXiv.2307.07662.
19. Zhang Y, Zhang S, Wang H T, et al. Crack detection algorithm for forging surface based on cross-scale feature extraction[J]. *Comput Integr Manuf Syst*, 2025, **31**(2): 499−511.
张岳, 张上, 王恒涛, 等. 基于跨尺度特征提取的锻件表面裂纹检测算法[J]. 计算机集成制造系统, 2025, **31**(2): 499−511.
20. Papakipos Z, Bitton J. AugLy: data augmentations for robustness[Z]. arXiv: 2201.06494, 2022. https://doi.org/10.48550/arXiv.2201.06494.
21. Liu Y, Liu Y L, Guo X Y, et al. Metal surface defect detection using SLF-YOLO enhanced YOLOv8 model[J]. *Sci Rep*, 2025, **15**(1): 11105.
22. Shen M H, Liu Y J, Chen J, et al. Defect detection of printed circuit board assembly based on YOLOv5[J]. *Sci Rep*, 2024, **14**(1): 19287.
23. Chen Q, Su X B, Zhang X Y, et al. LW-DETR: a transformer replacement to YOLO for real-time detection[Z]. arXiv: 2406.03459, 2024. https://doi.org/10.48550/arXiv.2406.03459.
24. Wang Y Y, Yin T T, Chen X C, et al. SCP-DETR: a efficient small-object-enhanced feature pyramid approach for PCB defect detection[J]. *PLoS One*, 2025, **20**(8): e0330039.
25. Lv X M, Duan F J, Jiang J J, et al. Deep metallic surface defect detection: the new benchmark and detection network[J]. *Sensors*, 2020, **20**(6): 1562.

作者简介



梁丹 (1999-)，女，硕士研究生，主要研究方向为目标检测、深度学习。

E-mail：15679411084@163.com



【通信作者】张上 (1979-)，男，博士，副教授，主要研究方向为计算机视觉、嵌入式。

E-mail：zhangshang@ctgu.edu.cn

Adaptive multi-scale fusion transformer for forging surface defect detection

Liang Dan[1,2], Zhang Shang[1,2*], Guan Huipan[1,2]

[1] College of Computer and Information Technology, China Three Gorges University, Yichang, Hubei 443002, China;
[2] Hubei Province Engineering Technology Research Center for Construction Quality Testing Equipment, China Three Gorges University, Yichang, Hubei 443002, China

Abstract

Objective Forged parts are widely used as key load-bearing components in automotive, aerospace, shipbuilding, and heavy equipment industries. Their surface quality directly affects structural reliability and service safety. During forging and subsequent processing, surface defects such as cracks, inclusions, pores, oxide layers, and spot-like imperfections are frequently generated due to material heterogeneity, mold wear, and process instability. Magnetic particle inspection is commonly adopted for ferromagnetic forged parts because of its high sensitivity to surface and near-surface cracks. However, magnetic particle inspection images usually exhibit complex backgrounds, uneven illumination, magnetic trace interference, and weak texture contrast. These characteristics cause surface defects to present blurred boundaries, irregular shapes, and large scale variation. Slender cracks and tiny defects are especially difficult to detect accurately under real-time constraints. Existing convolutional neural network–based and Transformer-based detection models often show insufficient directional perception, limited contextual representation, and inadequate geometric sensitivity in bounding-box regression. As a result, detection accuracy and computational efficiency are difficult to balance in practical industrial inspection scenarios. The objective of this study was to develop a real-time forged-part surface defect detection model that enhances directional feature perception and contextual fusion capability, improves multi-scale feature alignment, and strengthens localization robustness for elongated defects, while maintaining high inference efficiency suitable for industrial deployment.

Methods A real-time surface defect detection framework, named MCF-RTDETR, was constructed based on an improved RT-DETR architecture. Magnetic particle inspection images of heavy-duty truck steering knuckles were collected from an industrial production environment and manually annotated to establish a forged-part surface crack dataset. The dataset contained complex background structures and fine-scale defect patterns, providing realistic samples for model evaluation.

In the backbone network, a multi-gradient fusion convolution module was introduced to enhance directional edge representation. This module integrated four directional difference convolutions with one standard convolution to capture gradient responses along multiple orientations. Such design improved feature sensitivity to slender cracks and weakly textured defects. To avoid increased inference cost, convolutional re-parameterization was adopted during the inference stage, allowing the multi-branch structure to be equivalently transformed into a single convolution layer while preserving feature enhancement capability.

In the neck network, a context-aware recursive feature pyramid network was designed to address insufficient spatial–semantic alignment in conventional multi-scale fusion schemes. A rectangular self-calibration module was first employed to model directional contextual information by expanding receptive fields along horizontal and vertical directions, which strengthened anisotropic defect representation. Subsequently, a FuseBlockMulti module was used to realize semantic-guided alignment between shallow detail features and deep semantic features, enabling adaptive information interaction across different feature levels. Furthermore, a dynamic interpolation fusion mechanism was introduced to assign adaptive fusion weights to multi-scale features according to local contextual semantic strength. Through recursive fusion, feature consistency across scales was progressively refined, improving robustness under complex backgrounds and large scale variation.

For bounding-box regression, the original loss function was replaced with a Focaler-MPDIoU loss. This loss combined difficulty-aware sample weighting with multi-point distance geometric constraints, enhancing localization accuracy and boundary fitting ability for elongated and irregular defects. The entire detection framework was optimized in an end-to-end manner and evaluated under unified training and inference configurations.

Foundation item: Hubei Provincial National College Student Innovation and Entrepreneurship Training Program of China (S202311075047), the National College Student Innovation and Entrepreneurship Training Program of China (202111075012, 202011075013), and the Hubei Provincial Digital Economy Pilot Demonstration Construction Special Project (2312-420625-04-02-996363).

*Correspondence: Zhang S, E-mail: zhangshang@ctgu.edu.cn

Results and Discussions Experimental results demonstrated that the proposed MCF-RTDETR achieved superior performance in forged-part surface defect detection. On the magnetic particle inspection dataset of heavy-duty truck steering knuckles, the proposed model obtained a mean average precision of 88.5%, which exceeded the RTDETR-r18 baseline by 2.3%. Meanwhile, computational complexity was reduced from 58.3 G to 33.5 G, and inference speed reached 137.2 f/s, indicating an effective balance between detection accuracy and computational efficiency.

Performance improvement was particularly evident for slender cracks and small-scale defects. Enhanced directional gradient perception reduced missed detections caused by weak edge responses, while context-aware recursive fusion suppressed background interference and improved defect-background discrimination. Ablation experiments confirmed the effectiveness of each proposed component. The multi-gradient fusion Convolution strengthened edge sensitivity without increasing inference overhead. The Context-Aware Recursive Feature Pyramid Network improved multi-scale feature consistency and spatial-semantic alignment. The Focaler-MPDIoU loss further enhanced bounding-box regression accuracy for defects with large aspect ratios.

Generalization capability was validated on the public GC10-DET dataset. The proposed model achieved a 2.8% increase in mean average precision compared with the baseline detector, demonstrating that the proposed feature enhancement and fusion strategies are transferable to different industrial surface defect detection scenarios. These results indicate that the proposed framework effectively integrates accuracy improvement with computational reduction, which is essential for real-time industrial inspection systems.

Conclusions A context-aware multi-scale fusion real-time detection model for forged-part surface defects is presented. Directional gradient fusion, adaptive contextual feature integration, and geometry-sensitive regression jointly enhance detection robustness under complex backgrounds and severe scale variation. Experimental results confirm that the proposed approach improves detection accuracy while reducing computational complexity and maintaining high inference speed. The model shows strong generalization capability and practical applicability, providing an effective solution for automated forged-part surface defect inspection in industrial environments.

Keywords: defect detection; RT-DETR; feature fusion; loss function

Liang D, Zhang S, Guan H P. Adaptive multi-scale fusion transformer for forging surface defect detection[J]. *Opto-Electron Eng*, 2026, **53**(2): 250242; DOI: 10.12086/oee.2026.250242

Opto-Electronic Journals Group
www.oejournal.org

扫描二维码，获取 PDF 全文