



XXXX

基于多尺度特征融合的变电站设备缺陷检测

邢壮¹, 温书凯²

(1. 西安大麦网络科技有限公司, 陕西 西安 710016;

2. 西安交通大学, 电气工程学院, 陕西 西安 710049)

摘要: 针对变电站设备缺陷目标尺度跨度大及背景干扰严重等问题, 提出一种基于多尺度特征融合的变电站设备缺陷检测模型 (Substation Equipment Defect Detection Transformer, SED-DETR)。模型在编码阶段设计分组级联自注意力模块 (Grouped Cascaded Self-Attention, GCSA), 增强对关键区域的聚焦能力; 在颈部阶段构建跨阶段多分支特征融合网络 (Cross-Scale Multi-Branch Fusion Feed-Forward Network, CSMB-FFN), 通过改进单位步长融合卷积 (Unit Step Fusion Convolution, US-Conv) 实现多尺度特征精准对齐, 引入跨阶段多分支建模模块 (Cross-Stage Partial Omni-Kernel Block, CSP-OKB) 加强全局感知与局部纹理信息融合; 训练阶段采用 MPDIoU (Minimum Pairwise Distance Intersection over Union) 替换边界框回归损失中的 GIoU (Generalized Intersection over Union), 增强高重叠样本条件下的回归监督敏感性。SED-DETR 在自建数据集和 IGV Dataset 上均取得优于当前主流检测模型的检测结果, mAP@0.5 分别为 94.1% 和 95.4%, mAP@0.5:0.95 分别达到 77.6% 和 85.2%。端侧部署结果说明, 该模型在 Jetson Orin Nano 4GB 平台上具备良好的实时推理能力, 经 FP16 加速后, 推理时间缩短至 23.9ms, 且 mAP@0.5 达到 92.5%。结果表明, SED-DETR 在保证较高检测精度的同时兼顾计算效率, 为变电站设备缺陷检测任务提供可行方案。

关键词: 变电站设备缺陷检测; 深度学习; 多尺度特征融合; 自注意力机制; RT-DETR; 跨阶段特征融合; 特征对齐

中图分类号: TP391.4; TM63

文献标志码: A

doi: 10.11959/j.issn.1000-0801.

Substation Equipment Fault Detection Based on Multi-scale Feature Fusion

Xing Zhuang¹, Wen Shukai²

1. Xi'an Dama Network Technology Co., Ltd., Xi'an 710016, China

2. School of Electrical Engineering, Xi'an Jiaotong University, Xi'an 710049, China

Abstract: To address the large-scale variation of defect targets and severe background interference in substation equipment inspection images, this paper proposes a substation equipment defect detection model, termed the Substation Equipment Defect Detection Transformer (SED-DETR), based on multi-scale feature fusion. In the encoder, a

收稿日期: XXXX-XX-XX; 修回日期: XXXX-XX-XX

基金项目: 国家自然科学基金 (12301559)



Grouped Cascaded Self-Attention (GCSA) module is introduced to enhance focus on critical regions and improve cross-region contextual modeling. In the neck, a Cross-Scale Multi-Branch Fusion Feed-Forward Network (CSMB-FFN) is constructed, where an improved Unit Step Fusion Convolution (US-Conv) is employed for precise multi-scale feature alignment, and a Cross-Stage Partial Omni-Kernel Block (CSP-OKB) module is incorporated to strengthen the fusion of global context and local texture information. During training, the Generalized Intersection over Union (GIoU) term in the bounding-box regression loss is replaced with the Minimum Pairwise Distance Intersection over Union (MPDIoU) to improve the sensitivity of regression supervision under highly overlapped samples. Experimental results show that SED-DETR achieves mAP@0.5 values of 94.1% and 95.4%, and mAP@0.5:0.95 values of 77.6% and 85.2%, on the self-collected dataset and the public IGV Dataset, respectively, outperforming current mainstream detection models. Edge-side deployment results further demonstrate that the proposed model maintains favorable real-time inference capability on the Jetson Orin Nano 4GB platform; after FP16 acceleration, the inference time is reduced to 23.9 ms while the mAP@0.5 remains 92.5%. These results indicate that SED-DETR achieves a favorable balance between detection accuracy and computational efficiency, providing a practical solution for substation equipment defect detection.

Key words: Substation Equipment Fault Detection, Deep Learning, Multi-scale Feature Fusion, Self-attention Mechanism, RT-DETR, Cross-stage Feature Fusion, Feature Alignment

0 引言

变电站设备作为电力系统的重要组成部分,其运行状态直接影响到电力系统的安全性与稳定性^[1]。随着设备规模扩张与服役年限增加,设备运行过程中的缺陷风险与故障隐患相应增大^[2]。传统依赖人工的巡检方式在巡检时效、成本投入与结果一致性等方面均受到限制,且容易受人员经验、环境光照等因素影响,出现漏检、误判等问题^[3]。因此,基于视觉的智能缺陷检测逐步成为变电站设备检修与智能运维的重要技术路径^[4]。

变电站设备缺陷具有形状不规则、对比度低等特征,在复杂背景和遮挡、反光等成像干扰条件下检出与定位难度较大,通常采用两阶段检测框架^[5]。张铭泉等^[6]以快速区域卷积神经网络(Faster Region-based Convolutional Neural Network, Faster R-CNN)为基础,引入特征金字塔网络(Feature Pyramid Network, FPN)与空间金字塔池化(Spatial Pyramid Pooling, SPP),经自建数据集测试优于主流检测模型。不同于改进检测框架, Xin等^[7]将精细定位扩展到掩膜基于区

域的卷积神经网络(Mask Region-based Convolutional Neural Network, Mask R-CNN),引入全局特征金字塔网络(Global Feature Pyramid Network, GFPN)与卷积注意力模块(Convolutional Block Attention Module, CBAM),提升缺陷和背景相近场景下的可分离性。然而,两阶段方法对候选阶段召回与感兴趣区域(Region of Interest, RoI)较为敏感,候选漏召回与边界细节表征不足导致小缺陷漏检与定位不稳。

相较于两阶段检测方法,一阶段密集预测在多尺度特征图上直接进行回归与类别,更容易覆盖不同尺度缺陷^[8]。Wang等^[9]提出YOLO-SS-Large,通过改进骨干网络、回归损失和动态检测头增强多尺度感知能力,在其构建的17类缺陷数据集上AP较YOLOv5m提高0.3个百分点。在增强小目标分辨能力方向上, Wang等^[10]进一步提出EAL-YOLO,在高分辨率层级引入P2层融合,与基线模型相比mAP@0.5提升5.3%。此外, Sun等^[11]从轻量骨干与特征重标定角度对YOLOv8进行补强,证明优化骨干表征与特征重标定有助于抑制背景噪声。但是,密集预测依赖非极大值抑

制 (Non-Maximum Suppression, NMS) 等启发式后处理与阈值策略进行冗余筛选, 在目标密集或边界模糊场景下, 误抑制与定位漂移更容易出现^[12]。

为降低对启发式框筛选的依赖, 端到端 Transformer 检测提供了新的研究方式^[13]。DETR^[14] (Detection Transformer, DETR) 通过一对一匹配集合损失实现唯一分配, 从框架层面削弱对锚框设计以及人工设定先验与后处理策略的依赖。在此基础上, RT-DETR^[15] (Real-Time Detection Transformer) 提出高效混合编码器以实现多尺度特征交互, 通过查询选择策略提升初始查询质量, 减少重复预测带来的筛选压力。

上述研究均在一定程度上提高变电站设备缺陷检测精度, 但面对复杂的变电站场景仍存在以下瓶颈: 1) 变电站设备密集且局部纹理相近。跨区域信息交互和显著区域聚合不足时, 模型对关键区域的响应分散, 容易引发判别不稳, 造成定位一致性下降。2) 缺陷目标尺度分布离散, 简单的金字塔融合难以在不同分辨率上兼顾语义表达与细节保留。3) 变电站设备构件排布紧凑, 遮挡与包含关系较为常见, 训练阶段预测框与真实框易处于高重叠状态。

针对上述局限, 提出以 RT-DETR 为基础的 SED-DETR 检测框架, 主要工作内容如下: 1) 设计分组级联自注意力模块 GCSA, 通过分组注意交互与级联信息传递实现特征更新, 缓解相似背景下响应分散带来的判别不稳。2) 构建跨阶段多分支特征融合网络 CSMB-FFN, 在主干语义路径基础上引入高分辨率分支, 通过对齐、筛选与融合等步骤实现多尺度信息的可控输入。3) 引入最小成对距离交并比 MPDIoU 作为训练阶段边界框回归损失中的几何约束项, 将位置偏移与尺度变化纳入回归监督, 增强重叠样本条件下的几何约束表达。

1 模型介绍

变电站设备检测图像中, 缺陷目标尺度差异显著且细节信息有限, 易与背景构件纹理产生混淆。为此, 以 RT-DETR 为基础构建端到端检测框架 SED-DETR。模型整体由骨干网络、颈部特征融合网络、混合编码器与解码器组成, 结构如图 1 所示。骨干网络对输入图像进行分层特征提取, 形成 P2-P5 多尺度特征表示。在骨干网络输出的多尺度特征基础上, 设计 GCSA 模块对高层语义特征开展跨区域上下文交互与特征更新。随后, 在骨干网络之后构建 CSMB-FFN, 对不同分辨率特征进行尺度对齐与信息聚合, 结合 US-Conv 实现高效特征变换。在融合单元中采用 CSP-OKB 增强多分支信息交互。融合后的特征输入混合编码器进行统一建模。最后, 解码器基于查询机制完成分类与边界框回归, 输出缺陷类别及其对应边界框。

1.1 关键区域表征增强

变电站设备缺陷检测场景中, 目标区域常与背景构件纹理相近, 特征建模过程中容易将无关区域响应引入, 干扰对有效区域的表征。为缓解复杂背景带来的冗余关联, 现有方法通常采用自注意机制进行全局信息交互增强上下文表达^[16], 但在高分辨率特征上进行全局自注意会带来较高计算开销。为解决上述问题, 在编码端的特征交互阶段设计 GCSA, 通过分组建模降低冗余, 利用级联传递逐级细化, 在保持全局建模能力的同时增强对关键区域的聚焦。结构如图 2 所示。

GCSA 将来自骨干网络的高层语义特征沿通道维划分为 h 个分组。随后, 在各个分组内完成组内上下文交互。设第 i 个注意力头的第 j 组特征为 X_{ij} , 查询、键和值由线性投影得到。在此基础上, 通过自注意力算子获得更新后的分组表示。计算公式为

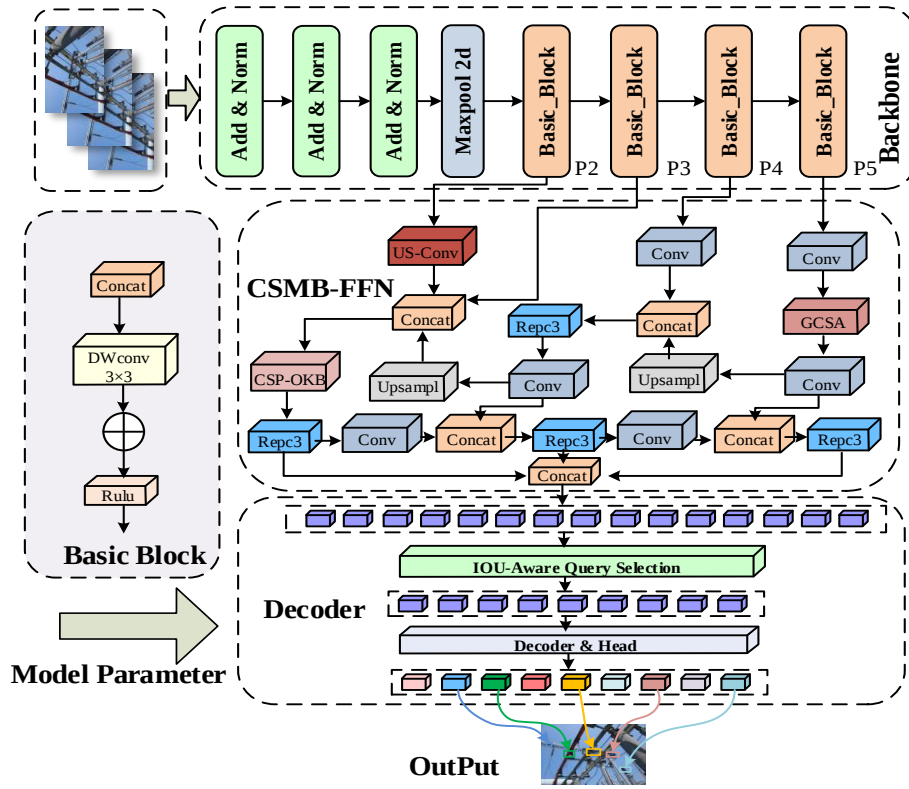


图1 SED-DETR 模型框架图

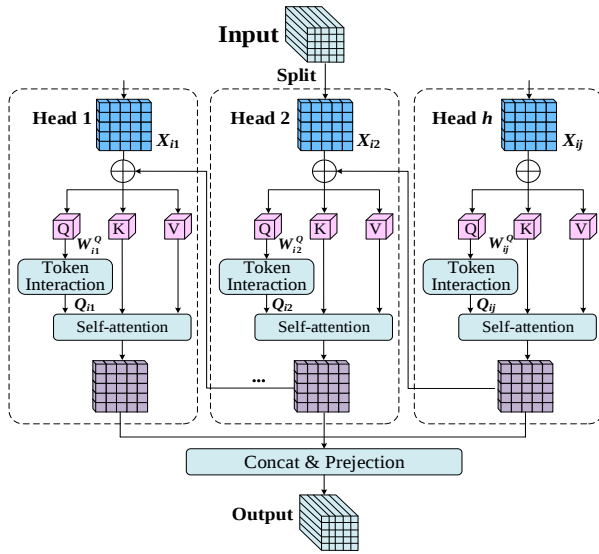


图2 GCSA 结构框架

$$\tilde{X}_{ij} = \text{Attn}(X_{ij}W_{ij}^Q, X_{ij}W_{ij}^K, X_{ij}W_{ij}^V) \quad (1)$$

其中, $\text{Attn}(\cdot)$ 表示自注意力算子, W_{ij}^Q, W_{ij}^K 和 W_{ij}^V 为可学习投影矩阵。

得到各组的注意力更新结果后, GCSA将各

组输出的注意力进行通道维拼接, 通过投影矩阵回映射到目标维度。计算公式为

$$\tilde{X}_{i+1} = \text{Concat}[\tilde{X}_{ij}]_{j=1:h} W_i^P \quad (2)$$

其中, $\text{Concat}[\cdot]$ 表示拼接操作, W_i^P 为输出投影矩阵。

在组内注意力计算与组间汇聚的基础上, 引入级联传递机制, 将后一组的输入在进入注意力计算前与前一组的注意力输出进行显式叠加。更新后的输出特征为

$$X'_{ij} = X_{ij} + \tilde{X}_{i(j-1)} \quad (3)$$

1.2 多尺度信息融合

受成像条件和设备结构复杂性影响, 巡检图像中的缺陷区域存在边界模糊、易被背景纹理淹没等问题。仅依靠深层特征, 多次下采样后空间分辨率下降, 使边缘裂纹、局部破损等细粒度信息衰减。高分辨率特征携带大量低层噪声与冗余纹理, 直接引入会削弱模型前景与背景区分能

力^[17]。为在细节保留与噪声抑制之间取得平衡，在颈部融合阶段设计 CSMB-FFN，在兼顾融合效率的基础上实现对多尺度信息的充分整合与有效筛选。模块架构如图 3 所示。

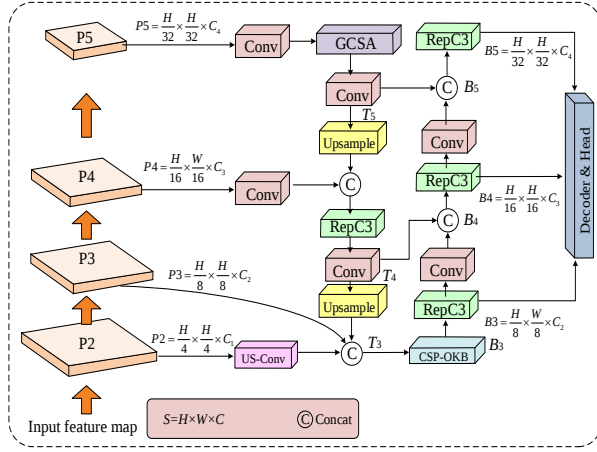


图3 CSMB-FFN 网络结构

CSMB-FFN 在骨干网络层级特征输出基础上引入高分辨率分支 P2 参与融合，形成四层输入特征 {P2, P3, P4, P5}。设定输入图像尺寸为 $H \times W$ ，各层输入特征表示为 $P2 \in \mathbb{R}^{H/4 \times W/4 \times C1}$ 、 $P3 \in \mathbb{R}^{H/8 \times W/8 \times C2}$ 、 $P4 \in \mathbb{R}^{H/16 \times W/16 \times C3}$ 和 $P5 \in \mathbb{R}^{H/32 \times W/32 \times C4}$ ， $C1$ 、 $C2$ 、 $C3$ 和 $C4$ 分别表示对应层的通道数。

为缓解跨层融合的维度不匹配并提升融合稳定性，CSMB-FFN 对 P4 和 P5 进行卷积对齐，可表示为

$$F_i = \psi_i(P_i), i \in \{4, 5\} \quad (4)$$

其中， $\psi_i(\cdot)$ 表示卷积映射算子。

针对高分辨率分支直接降尺度会出现信息丢失的问题，通过 US-Conv 对 P2 进行尺度匹配与信息筛选得到 $P2'$ 。其中，US-Conv 为本文设计的单位步长融合卷积，详见 1.2.1。CSMB-FFN 以顶层特征为起点进行增强，通过自顶向下路径进行逐级传递与融合。融合过程为

$$T_5 = \varepsilon(F_5) \quad (5)$$

$$T_4 = \text{Concat}(\text{Up}(T_5), F_4) \quad (6)$$

$$T_3 = \text{Concat}(\text{Up}(T_4), P3, P2') \quad (7)$$

其中， $\varepsilon(\cdot)$ 为顶层增强算子， $\text{Concat}(\cdot)$ 表示特征拼接， $\text{Up}(\cdot)$ 为上采样算子， T_5 、 T_4 和 T_3 分别为不同尺度的融合中间特征。

仅依赖自顶向下路径容易使信息流呈单向传播，使细节信息难以充分反馈至更高层特征。为此，CSMB-FFN 构建自底向上回流路径，在关键尺度处引入 CSP-OKB 进行多分支整合。CSP-OKB 为跨阶段多分支融合块，详见 1.2.2。回流融合过程为

$$B_3 = \text{CSP-OKB}(T_3) \quad (8)$$

$$B_4 = \text{Concat}(\text{Down}(B_3), T_4) \quad (9)$$

$$B_5 = \text{Concat}(\text{Down}(B_4), F_5) \quad (10)$$

其中， B_3 、 B_4 、 B_5 为回流聚合后的多尺度输出特征， $\text{Concat}(\cdot)$ 为特征拼接， $\text{Down}(\cdot)$ 为降尺度算子，用于尺度对齐。

1.2.1 特征对齐

针对颈部特征融合过程中高分辨率分支直接并入主融合链路产生冗余响应、增加融合开销等问题。在 CSMB-FFN 中设计 US-Conv，用于高分辨率分支进入融合节点前进行尺度对齐与信息整合。US-Conv 以单位步长卷积为核心，通过连续邻域的逐位置聚合完成空间信息重整，使每个输出位置均由局部连续区域参与计算。将 P2 层特征变换与中层特征一致的融合尺度后接入主干语义路径，与自顶向下传递的语义信息在同一尺度进行拼接与融合，在不改变后续融合的前提下实现细节补充与计算开销的平衡。

二维卷积计算

$$Y(o, i, j) = \sum_{c=1}^{C_{in}} \sum_{u=1}^k \sum_{v=1}^k W(o, c, u, v) X(c, i+u, j+v) \quad (11)$$

其中， X 表示输入特征张量， Y 表示卷积输出特征张量， c 为通道索引， o 为输出通道索引， i, j 分别为输出特征在空间维上的行、列坐标， C_{in} 为输入通道数， k 为卷积核大小， u, v 为卷积核在空间



维的偏移索引。

1.2.2 跨阶段协同建模

变电站设备缺陷在形态上既包含整体结构层面的异常，也伴随局部细碎纹理和弱对比细节，仅依赖单一感受野的卷积特征，容易出现全局上下文刻画不足和局部细节响应不稳的问题^[18]。为此，在CSMB-FFN的关键融合尺度处设置CSP-OKB模块，融合全局感知、大感受野建模和局部边缘细节三类互补信息。模块如图4所示。

全局分支输入特征为 $X_{\text{Global}} \in \mathbb{R}^{C \times H \times W}$ 。双通道注意力机制DCAM首先执行频域通道注意力FCA，表达式为

$$X_{\text{FCA}} = f^{-1} \left(f(X_{\text{Global}}) \otimes (W_{1 \times 1} \text{GAP}(X_{\text{Global}})) \right) \quad (12)$$

其中， $f(\cdot)$ 和 $f^{-1}(\cdot)$ 分别表示快速傅里叶变换与逆变换（IFFT），GAP为全局平均池化， $W_{1 \times 1}$ 为 1×1 卷积层。

随后，DCAM继续执行空间通道注意SCA，得到

$$X_{\text{DCAM}} = X_{\text{FCA}} \otimes W_{1 \times 1}(\text{GAP}(X_{\text{FCA}})) \quad (13)$$

其中， X_{DCAM} 为双域通道注意输出。

在空间维重建细粒度频谱信息，特征选择注意力模块FSAM的表达式为

$$X_{\text{FSAM}} = f^{-1} \left(f(W_{1 \times 1}^1(X_{\text{DCAM}})) \otimes W_{1 \times 1}^2(X_{\text{DCAM}}) \right) \quad (14)$$

其中， $W_{1 \times 1}^1, W_{1 \times 1}^2$ 为两个独立的 1×1 卷积层， X_{FSAM} 为频域空间注意输出。

通过式（12）-（14），全局分支将频域信息与空间注意联合引入特征增强。

为获得更大的感受野提升特征提取能力，大核分支采用超大核深度卷积与多形状卷积核进行建模，局部分支通过 1×1 深度卷积补充局部边缘与细节响应。为便于统一描述，记 $X_g = X_{\text{FSAM}}$ 全局分支、 X_b 大核分支输出和 X_l 局部分支输出。聚合输出为

$$X_{\text{OK}} = \psi_{1 \times 1}(\text{Concat}(X_g, X_b, X_l)) \quad (15)$$

其中，Concat表示通道维拼接， $\psi_{1 \times 1}(\cdot)$ 为 1×1 卷积映射，用于跨分支信息融合控制通道规模。

在CSP结构中，输入特征 X 沿通道维拆分为两部分 $(X^{(a)}, X^{(b)})$ ，一部分进入上述三分支增强得到 X_{OK} ，另一部分保持直连促进跨阶段信息流动。最终输出可写为

$$Y = \psi_{1 \times 1}(\text{Concat}(X_{\text{OK}}, X^{(b)})) \quad (16)$$

1.3 边界框回归约束

RT-DETR在边界框回归中引入GIoU作为几

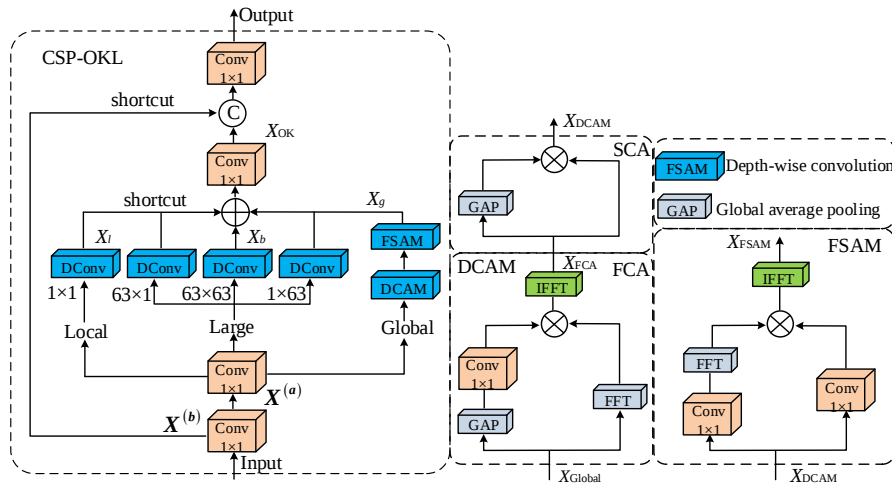


图4 CSP-OKB模块结构

何约束项，用于刻画预测框与真实框的空间重叠关系。变电站设备缺陷中小尺度目标和重叠目标占比较高，预测框与真实框易出现较高重叠甚至包含关系。为增强模型对高重叠样本中细微几何差异的约束表达，在 SED-DETR 训练阶段的回归损失中引入 MPDIoU^[19]。MPDIoU 以点距离度量补充重叠度信息，将位置偏移与尺度变化显式纳入回归监督过程。

MPDIoU 损失函数表达式如下所示

$$L_{MPDIoU} = 1 - MPDIoU \quad (17)$$

$$MPDIoU = \frac{A \cap B}{A \cup B} - \frac{b_1^2}{W^2 + L^2} - \frac{b_2^2}{W^2 + L^2} \quad (18)$$

$$b_1^2 = (x_1^t - x_1^p)^2 - (y_1^t - y_1^p)^2 \quad (19)$$

$$b_2^2 = (x_2^t - x_2^p)^2 - (y_2^t - y_2^p)^2 \quad (20)$$

其中， b_1 和 b_2 为惩罚项， (x_1^t, y_1^t) 为真实框左上角坐标， (x_2^t, y_2^t) 为真实框右下角坐标， (x_1^p, y_1^p) 为预测框左上角坐标， (x_2^p, y_2^p) 为预测框右下角坐标。

2 实验设置

2.1 数据集介绍

为验证模型的有效性与泛化能力，在公开数据集和自建变电站设备缺陷数据集上开展实验。公开数据集 IGV Dataset^[20]共包含 6131 张有效图像，其中鸟巢 1094 张、表盘模糊 1143 张、绝缘子自爆 769 张、腐蚀 658 张、绝缘子破损 1039 张和硅胶变色 1428 张。自建数据集来源于西安市某电力公司变电站巡检现场采集的可见光图像，使用无人机和数码相机采集变电站设备缺陷数据。经人工审核与清洗后剔除模糊、重复及无效样本，最终得到 9144 张有效图像。

使用 LabelImg 工具对图像中的缺陷目标进行精细化矩形边界框标注。将检测目标划分为异物、绝缘子破损、绝缘子自爆等 9 类。各类别的标签定义和样本数量统计见表 1。数据集按照 7:1:

2 的比例划分为训练集、验证集和测试集，在划分过程中避免同一设备或巡检序列中的高度相似图像同时出现在训练集与测试集中。训练阶段对输入图像进行尺度缩放并采用随机数据增强策略如随机翻转、颜色扰动，测试阶段采用单尺度输入完成检测。

表 1 样本类别和数量

样本类别	标签	数量
异物	YW	2572
绝缘子破损	JS	1895
绝缘子自爆	ZB	874
硅胶变色	GJ	661
腐蚀	FS	716
渗漏油	LY	544
箱门异常	YC	621
表盘破损	BP	597
观察窗破裂	PL	634

2.2 实验环境设置

所有实验均在 Ubuntu 20.04 LTS 操作系统下完成，服务器内存为 128 GB，GPU 采用 NVIDIA GeForce RTX 4090 用于模型训练与推理。算法使用 Python 3.9 实现，网络模型基于 PyTorch 2.1.0 搭建，GPU 计算环境为 CUDA 11.8。主要参数设置如表 2 所示。

表 2 模型训练参数设置

参数名称	设置
优化器(Optimizer)	AdamW
学习率(Learning Rate)	0.0001
批次大小(Batch Size)	16
训练周期(Epochs)	200
动量(Momentum)	0.937
输入分辨率 (Input Resolution)	1280×720

3 实验验证

3.1 评价指标

为验证模型有效性，使用精确率 (Precision, P)、召回率 (Recall, R) 和平均精度均值



(mAP) 作为模型评价指标。为衡量边界框的回归质量使用 $\text{mAP}@0.5$ 与 $\text{mAP}@0.5:0.95$ 进行综合评估, 0.5 与 0.50:0.95 表示预测框与真实框匹配所使用的重叠率阈值。指标公式如下

$$P = \frac{TP}{TP + FP} \quad (21)$$

$$R = \frac{TP}{TP + FN} \quad (22)$$

$$\text{mAP} = \frac{1}{C} \sum_{i=1}^C \int_0^1 P(R) dR \quad (23)$$

式中, TP 表示正确检出的目标数, FP 表示误检目标数, FN 表示漏检目标数, C 为类别数, i 为类别。

3.2 US-Conv 嵌入位置

为分析 US-Conv 在不同融合层级中的有效性, 保持 SED-DETR 其余网络结构与训练策略不变, 仅改变 US-Conv 的嵌入位置, 在自建数据集上展开对比实验, 分别嵌入到 P3、P4、P5 对应的融合节点进行对比, 实验结果如表 3 所示。

表 3 样本类别和数量

	P%	R%	mAP@0.5/%	mAP@0.5:0.95/%
P2	93.1	95.7	94.1%	77.6%
P3	92.5	94.6	93.4%	76.2%
P4	91.2	93.9	92.5%	72.7%
P5	90.6	92.1	91.3%	68.4%

由表 3 可知, 随着 US-Conv 接入层级由 P2 逐步向后至 P5, 指标均呈下降趋势。mAP@0.5 下降 2.8%, 这表明在宽松 IoU 阈值下, 模型的检测能力受接入位置影响相对有限。mAP@0.5:0.95 下降更为明显, 由 77.6% 降至 68.4%, 说明更严

格阈值下的定位与边界拟合质量对嵌入层级更敏感。原因在于越靠前如 P2 和 P3 层接入越能保留细粒度结构线索并在后续融合中形成有效补充, 当嵌入为 P4 或 P5 层时, 由于特征分辨率更低、空间细节被进一步压缩, US-Conv 带来的细节优势难以转化为高 IoU 下的精确回归, 导致 mAP@0.5:0.95 出现更明显数值下滑。

3.3 CSP-OKB 消融实验

为分析颈部融合模块中 CSP-OKB 各组成策略对检测性能的影响, 以 SED-DETR 完整结构作为参照组开展消融对比。在保持骨干网络、CSMB-FFN 其余结构、解码器及训练配置一致的前提下, 仅在融合节点对 CSP 与 OKB 相关结构进行替换。对比设置三种变体: 1) 将 CSP-OKB 替换为常规 3×3 卷积融合块, 2) 去除 CSP 直连分流, 3) 移除 OKB 多分支结构。实验结果如表 4 所示, 在不引入 CSP-OKB 的情况下, 模型的数量和计算量最小且推理时间最短, 为 14.2ms, 但检测性能相对较弱, 相较于完整模型 mAP@0.5 降低 2.3%。当仅移除 CSP、保留 OKB 时, mAP@0.5:0.95 降低至 76.5%, 降幅最小, 表明 OKB 多分支建模与注意力重标定有效提升高阈值定位质量, 但代价是参数量与计算量显著增加。当仅移除 OKB、保留 CSP 时, 相比于移除 CSP-OKB 检测性能有所提升, 但相较 OKB 分支的增益幅度较小, 表明 CSP 主要作用在于增强信息传递稳定性并抑制冗余特征干扰, 对检测性能的直接贡献相对有限。

表 4 CSP-OKB 消融实验结果

CSP-OKB	CSP	OKB	P%	R%	mAP@0.5/%	mAP@0.5:0.95/%	推理时间/ms	Params/M	FLOPs/G
×			92.4	93.7	91.8	72.9	14.2	16.3	46.2
	×		93.4	94.7	93.9	76.5	14.3	19.4	57.7
		×	93.2	94.1	93.7	75.3	15.0	17.9	51.9
√	√	√	93.7	95.4	94.1	77.6	14.5	18.7	55.4

3.4 损失函数对比

为分析不同边界框回归损失对变电站设备缺陷检测任务的适配性,在相同网络结构与训练策略下开展损失函数对比实验。对比对象包括 GIoU、CIoU^[21]、N-IoU^[22]、InterpIoU^[23] 和 MPDIoU 五种损失函数。训练过程中的损失变化如图5所示。

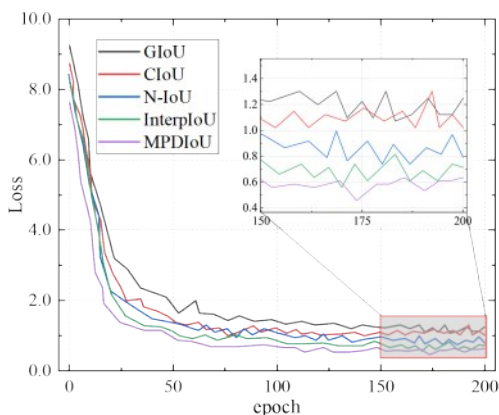


图5 损失函数对比

分析图5可知,不同损失函数在训练初期均呈现较快下降趋势,说明各类IoU改进策略均能对边界框回归产生有效约束。随着训练进行,GIoU曲线整体较高且后期波动更明显,表明对高重叠样本中细微几何偏差的约束相对有限。相比之下,CIoU、N-IoU和InterpIoU的损失下降更充分,后期稳定性也有所改善。结合150-200轮的局部放大曲线可以看出,MPDIoU在训练后期

始终保持更低且更平稳的损失水平,说明能够在高重叠样本条件下提供更稳定的梯度反馈,有利于提升边界框拟合能力和定位一致性。

3.5 对比实验

为验证SED-DETR模型的有效性,与主流目标检测方法进行对比。包括Faster R-CNN、YOLOv12-m、RT-DETR、OHEH-RTDETR^[24]、RAP-DETR^[25]和CPF-DETR^[17]等。为保证实验的公平性,所有对比实验均在相同的数据集和实验设置下进行。自建数据集模型性能对比结果如表5所示。公开数据集模型性能对比结果如表6所示。

由表5可知,SED-DETR在自建数据集上取得了较优的综合性能。 $mAP@0.5$ 达到94.1%,比Faster R-CNN提升8.6%,较RT-DETR提升4.4%,表明所提模型在总体检出能力上具有明显优势。在更严格的 $mAP@0.5:0.95$ 指标上,SED-DETR取得77.6%,相较于RAP-DETR提升10.2%,同当前对比方法中表现最优的CPF-DETR相比提升5.1%,说明在高IoU阈值下本文模型定位质量与回归可靠性更好。检测效率方面,SED-DETR的推理时间达到14.5ms,显著优于Faster R-CNN,同样低于CPF-DETR,并与速度最快的RAP-DETR相近。资源开销方面,SED-DETR参数量仅18.7M,相比于RT-DETR略有降低,计算量

表5 自建数据集模型性能对比

模型	P%	R%	$mAP@0.5/\%$	$mAP@0.5:0.95/\%$	推理时间/ms	Params/M	FLOPs/G
Faster R-CNN	84.2	86.7	85.5	52.6	107.5	42.0	189.5
YOLOv10-m	86.3	88.4	87.6	55.8	20.3	15.4	63.2
YOLOv11-m	87.6	89.1	88.7	58.3	19.4	20.1	72.8
YOLOv12-m	88.3	90.5	89.4	59.2	18.6	20.2	71.7
RT-DETR	88.6	90.7	89.7	60.7	17.1	19.9	62.3
OHEH-RTDETR	89.1	91.2	90.2	65.9	16.3	14.3	48.1
RAP-DETR	87.5	89.8	88.3	67.4	14.2	13.4	56.4
CPF-DETR	90.4	92.9	91.6	72.5	16.1	15.2	71.3
SED-DETR	93.7	95.4	94.1	77.6	14.5	18.7	55.4



FLOPs 为 55.4G，低于 RT-DETR 及多数对比方法。综合精度、定位质量与计算代价，SED-DETR 在自建数据集上表现出更优的性能权衡，验证所提模型的有效性。

由表 6 可知，SED-DETR 在 IGV Dataset 上取得了最优的检测精度和定位质量。SED-DETR 的 $mAP@0.5$ 达到 95.4%，相比两阶段的 Faster R-CNN 提升 6.2%，在同类端到端框架中，也明显优于 RAP-DETR 与 CPF-DETR，表明所提出方法对多类别缺陷的识别与检出有更稳定判别能力。在 $mAP@0.5:0.95$ 上，相对最优对比模型 CPF-DETR 提升 3.4%，说明模型在高 IoU 阈值下仍能保持更好的边界框贴合度与定位一致性。此外，SED-DETR 在 P 与 R 上分别达到 94.5% 和 96.3%，均为表中最高，反映其在提升召回的同时兼顾精确率，整体检测输出更为均衡。

在效率和复杂度方面，SED-DETR 的推理时间为 13.1ms，与 YOLO 系列模型相比，推理时间更短。资源开销方面，SED-DETR 的参数量为 18.7M，与 RT-DETR 接近，计算量为 54.3GFLOPs，低于 RT-DETR。综合来看，SED-DETR 在 IGV Dataset 上实现了更高的精度上限与更稳定的定位表现，同时保持了较为可控的计算代价。

3.6 消融实验

为评估各模块对模型性能的影响，在自建数据集上开展四组消融实验。第一组去除 GCSA，第二组去除 CSMB-FFN，第三组去除 MPDIoU，第四组为完整模型。实验结果如表 7 所示，SED-DETR 中 GCSA、CSMB-FFN 与 MPDIoU 三项改进对检测精度与计算开销均产生影响。移除 GCSA 后 $mAP@0.5$ 降低 1.7%， $mAP@0.5:0.95$ 降低 4.0%，高阈值指标下降幅度更为显著，说明 GCSA 在增强特征交互与细粒度定位方面具有更关键的作用。移除 CSMB-FFN 后模型精度退化最为明显， $mAP@0.5$ 降至 90.3%， $mAP@0.5:0.95$ 降至 71.2%，尽管推理时间缩短了 0.7ms，但跨尺度融合能力的削弱导致定位质量与整体检出显著下降，说明 CSMB-FFN 对多尺度信息整合与稳定回归具有基础性支撑作用。将回归损失由 MPDIoU 替换回 GIoU 后， $mAP@0.5$ 与 $mAP@0.5:0.95$ 分别下降 0.4% 和 3.5%。

3.7 实验结果可视化

为展示不同模型在公开数据集和自建数据集上的检测效果，通过可视化实验结果来对比模型性能。自建数据集检测效果如图 6 所示，公开数据集检测效果如图 7 所示。

分析图 6 和图 7 可知，在小目标、同类目标尺度差异较大或线缆密集的场景中，YOLOv12-

表 6 IGV Dataset 模型性能对比

模型	P%	R%	$mAP@0.5/\%$	$mAP@0.5:0.95/\%$	推理时间/ms	Params/M	FLOPs/G
Faster R-CNN	88.3	90.2	89.2	78.5	73.5	42.0	179.3
YOLOv10-m	89.4	91.3	90.5	79.3	15.8	15.4	59.1
YOLOv11-m	90.7	92.4	91.1	79.6	16.3	20.1	68.5
YOLOv12-m	90.9	92.8	91.4	79.9	15.2	20.2	67.4
RT-DETR	91.2	92.9	91.8	80.2	15.0	19.9	61.1
OHEH-RTDETR	91.5	93.1	91.9	80.7	14.2	14.3	46.6
RAP-DETR	91.6	93.4	92.3	81.5	12.8	13.4	56.5
CPF-DETR	91.8	93.7	92.7	81.8	13.6	15.2	62.6
SED-DETR	94.5	96.3	95.4	85.2	13.1	18.7	54.3

表7 消融实验结果

GCSA	CSMB-FFN	MPDIoU	P%	R%	mAP@0.5/%	mAP@0.5:0.95/%	推理时间/ms	Params/M	FLOPs/G
×	√	√	91.5	93.2	92.4	73.6	14.4	17.2	51.3
√	×	√	89.4	91.6	90.3	71.2	13.8	16.5	49.6
√	√	×	92.1	94.2	93.7	74.1	14.6	18.5	55.3
√	√	√	93.7	95.4	94.1	77.6	14.5	18.7	55.4



图6 自建数据集可视化结果

(a) YOLOv12-m (b) RT-DETR (c) CPF-DETR (d) SED-DETR

m与RT-DETR普遍存在置信度偏低、边界框贴合不足等问题，CPF-DETR虽检测较稳定，但在结构重复和背景纹理干扰较强的区域仍出现定位误差与置信度波动。相比之下，SED-DETR在不同数据集上均表现出更一致的检测结果，预测框与目标边缘贴合更紧，置信度整体更高，说明在复杂背景抑制和跨尺度特征表征方面具有更好的稳定性。

3.8 嵌入式部署实验

为验证模型在实际应用中的适用性，将RT-DETR与SED-DETR部署至Jetson Orin Nano 4GB平台。部署时，首先将训练完成的模型转换为ONNX格式，随后利用NVIDIA TensorRT生成推

理引擎，并结合FP16半精度推理与算子融合策略进行加速。在自建变电站设备缺陷数据集上开展推理实验，实验结果如表8所示。

由表8可知，SED-DETR在Jetson Orin Nano 4GB平台上具有较好的精度与效率优势。与RT-DETR相比，未加速条件下，SED-DETR的mAP@0.5和mAP@0.5:0.95分别提高4.4个百分点和16.9个百分点，推理时间缩短17.2 ms。采用FP16优化后，两种模型的推理时间均出现下降，但SED-DETR仍保持更高检测精度，mAP@0.5和mAP@0.5:0.95分别为92.5%和72.3%，推理时间为23.9ms。表明模型具有较好的嵌入式部署适用性，能够在资源受限平台上实现较优的精度与



图7 公开数据集可视化结果

(a) YOLOv12-m (b) RT-DETR (c) CPF-DETR (d) SED-DETR

表8 嵌入式实验结果

模型	mAP@0.5/%	mAP@0.5: 0.95/%	推理时间/ms
RT-DETR	89.7	60.7	47.8
RT-DETR (FP16)	86.2	53.4	35.2
SED-DETR	94.1	77.6	30.6
SED-DETR (FP16)	92.5	72.3	23.9

效率平衡。

4 结论

围绕变电站设备缺陷检测中的复杂背景干扰、多尺度目标表征不足以及高重叠样本定位不稳定等问题，提出 SED-DETR 模型，主要结论如下：

1) 模型从特征表达、多尺度融合和回归约束三个方面进行优化。GCSA 提升复杂背景下关键区域的响应能力，CSMB-FFN 结合 US-Conv 与 CSP-OKB 增强不同尺度特征之间的融合效果，MPDIoU 改善高重叠样本条件下的边界框回归质

量，提高模型整体检测精度与定位稳定性。

2) 在检测性能方面，SED-DETR 在自建数据集和公开 IGV Dataset 上均取得最优检测结果。在自建数据集上 mAP@0.5 较 RT-DETR 提高 4.4 个百分点，在公开 IGV Dataset 上提高 3.6 个百分点。消融实验表明，GCSA、CSMB-FFN 和 MPDIoU 均对性能提升具有积极作用，其中 CSMB-FFN 的贡献最为显著。

3) 嵌入式部署结果显示，SED-DETR 在 Jetson Orin Nano 4GB 平台上具备较好的边缘端适应能力。经 FP16 加速后推理时间缩短至 23.9 ms，模型在保持较高检测精度的同时兼顾推理效率，具有一定的边缘端部署潜力和工程应用价值。

未来工作将通过优化特征融合机制和计算效率，提升模型在恶劣环境下的适应能力，推动变电站设备检测向更加智能和自动化的方向发展。

参考文献：

- [1] 于景飞, 刘俊俊. 基于动态贝叶斯网络的变电站运维风险评估[J]. 科学技术与工程, 2025, 25(33): 14507-14516.

- Yu J F, Liu J J. Risk assessment of substation operation and maintenance based on dynamic Bayesian network[J]. *Science Technology and Engineering*, 2025, 25(33): 14507-14516. [维普]
- [2] 张宗包, 韩茂林, 王业震, 等. 基于模糊生成对抗网络的变电站数字孪生自动化建模方法[J]. *电力自动化设备*, 2025, 45(11): 52-58.
Zhang Z B, Han M L, Wang Y Z, et al. Automation modeling method for substation digital twin based on fuzzy generative adversarial networks[J]. *Electric Power Automation Equipment*, 2025, 45(11): 52-58. [万方]
- [3] 叶远波, 王吉文, 孙振兴, 等. 分区协同下适应新一代智能变电站的通信链路故障定位方法[J]. *中国电力*, 2025, 58(9): 79-87.
Ye Y B, Wang J W, Sun Z X, et al. Communication link fault localization method adapted to new smart substations under partition collaboration[J]. *Electric Power*, 2025, 58(9): 79-87. [维普]
- [4] 强仁, 龚钢军, 王大为, 等. 基于变电站信息物理耦合特征的扰动建模、韧性评估和提升策略[J/OL]. *中国电机工程学报*, 2025: 1-22. Qiang R, Gong G J, Wang D W, et al. Substation Disturbance Modeling, Resilience Evaluation Method and Enhancement Strategy Based on the Characteristics of Cyber Physical Coupling[J]. *Proceedings of the CSEE*, 2025: 1-22. [知网]
- [5] Yang Q, Ma S, Guo D Q, et al. A small object detection method for oil leakage defects in substations based on improved faster-RCNN[J]. *Sensors*, 2023, 23(17): 7390. [https://doi.org/10.3390/s23177390]
- [6] 张铭泉, 邢福德, 刘冬. 基于改进Faster R-CNN的变电站设备外部缺陷检测[J]. *智能系统学报*, 2024, 19(2): 290-298.
Zhang M Q, Xing F D, Liu D. External defect detection of transformer substation equipment based on improved Faster R-CNN[J]. *CAAI Transactions on Intelligent Systems*, 2024, 19(2): 290-298. [万方]
- [7] Xin M Y, Xu C B, Gao J P, et al. High-precision recognition algorithm for equipment defects based on mask R-CNN algorithm framework in power system[J]. *Processes*, 2024, 12(12): 2940. [https://doi.org/10.3390/pr12122940]
- [8] Jiang J J, Zhang Y Q, Wan A P, et al. Multi-target detection for safety monitoring in complex substation environments using YOLO-DySE[J]. *Signal, Image and Video Processing*, 2025, 19(10): 804. [https://doi.org/10.1007/s11760-025-04425-9]
- [9] Wang Q, Yang L X, Zhou B, et al. YOLO-SS-large: a lightweight and high-performance model for defect detection in substations[J]. *Sensors*, 2023, 23(19): 8080. [https://doi.org/10.3390/s23198080]
- [10] Wang J Q, Sun Y W, Lin Y, et al. Lightweight substation equipment defect detection algorithm for small targets[J]. *Sensors*, 2024, 24(18): 5914. [https://doi.org/10.3390/s24185914]
- [11] Sun Y W, Sun X R, Lin Y, et al. Substation equipment defect detection based on improved YOLOv8[J]. *Sensors*, 2025, 25(11): 3410. [https://doi.org/10.3390/s25113410]
- [12] Jiang J J, Zhang Y Q, Wan A P, et al. Multi-target detection for safety monitoring in complex substation environments using YOLO-DySE[J]. *Signal, Image and Video Processing*, 2025, 19(10): 804. [https://doi.org/10.1007/s11760-025-04425-9]
- [13] Yu L, Tang L, Mu L S. A review of detection transformer: From basic architecture to advanced developments and visual perception applications[J]. *Sensors*, 2025, 25(13): 3952. [https://doi.org/10.3390/s25133952]
- [14] Carion N, Massa F, Synnaeve G, et al. End-to-end object detection with transformers[C]//*Computer Vision - ECCV 2020*. Cham: Springer, 2020: 213-229. [https://doi.org/10.1007/978-3-030-58452-8_13]
- [15] Zhao Y A, Lv W Y, Xu S L, et al. DETRs beat YOLOs on real-time object detection[C]//*Proceedings of the 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Piscataway: IEEE Press, 2024: 16965-16974. [https://doi.org/10.1109/CVPR52733.2024.01605]
- [16] 徐程, 郑春厚, 胡祥, 等. 基于TLI-DETR的输电线路巡检图像小目标检测方法[J]. *智慧电力*, 2025, 53(9): 117-125.
Xu C, Zheng C H, Hu X, et al. A small object detection method for transmission line inspection images based on TLI-DETR[J]. *Smart Power*, 2025, 53(9): 117-125. [维普]
- [17] Fang H, Lin S, Hu J W, et al. CPF-DETR: an end-to-end DETR model for detecting complex patterned fabric defects[J]. *Fibers and Polymers*, 2025, 26(1): 369-382. [https://doi.org/10.1007/s12221-024-00809-9]
- [18] Lan A R, He W, Wang T. Research on substation equipment defect detection technology based on an improved RT-DETR model[C]//*Proceedings of the 2024 6th International Conference on Robotics, Intelligent Control and Artificial Intelligence (RICAI)*. Piscataway: IEEE Press, 2024: 680-686. [https://doi.org/10.1109/RICAI64321.2024.10911708]
- [19] Xiong C Q, Zayed T, Jiang X Y, et al. A novel model for instance segmentation and quantification of bridge surface cracks: the YOLOv8-AFPN-MPD-IoU[J]. *Sensors*, 2024, 24(13): 4288. [https://doi.org/10.3390/s24134288]
- [20] Hu M T, Zhuang Y, Wang J H, et al. PBZGNet: a novel defect detection network for substation equipment based on gradual parallel branch architecture[J]. *Sensors*, 2026, 26(1): 300. [https://doi.org/10.3390/s26010300]



- [21] Du S J, Zhang B F, Zhang P, et al. An improved bounding box regression loss function based on CIOU loss for multi-scale object detection[C]//Proceedings of the 2021 IEEE 2nd International Conference on Pattern Recognition and Machine Learning (PRML). Piscataway: IEEE Press, 2021: 92-98. [https://doi.org/10.1109/PRML52754.2021.9520717]
- [22] Su K K, Cao L H, Zhao B T, et al. N-IoU: better IoU-based bounding box regression loss for object detection[J]. Neural Computing and Applications, 2024, 36(6): 3049-3063. [https://doi.org/10.1007/s00521-023-09133-4]
- [23] Liu H, Watanabe H. InterIoU: Rethinking Bounding Box Regression with Interpolation-Based IoU Optimization[EB/OL]. arXiv: 2507.12420, 2025-07-16 [2026-05-13]. [https://doi.org/10.48550/arXiv.2507.12420]
- [24] Wang H C, Zhang Y G, Wu W H. OHEH-RTDETR: an improved RT-DETR detection model based on frequency layered processing and advanced feature selection[J]. Measurement Science and Technology, 2025, 36(4): 046211. [https://doi.org/10.1088/1361-6501/adc1f3]
- [25] Xie H, Zhou H B, Chen R L, et al. RAP-DETR: enhancing RT-DETR for railway track defect detection[J]. Measurement, 2026, 258: 119058. [https://doi.org/10.1016/j.measurement.2025.119058]

[作者简介]

邢壮, 1991, 男, 西安大麦网络科技有限公司, 总经理, 副高, 主要从事电力设备故障诊断方面的研究。

温书凯, 1987, 男, 西安交通大学电气工程学院, 教授, 主要从事计算机应用、电力设备故障诊断技术方面的研究。

邢壮(1991-), 男, 山东威海人, 本科, 副高级工程师, 研究方向为电力设备故障诊断;
温书凯(1987-), 男, 河南郑州人, 博士, 研究方向为计算机应用、电力设备故障诊断技术。E-mail:wenshukai187@163.com

