

DOI:10.11918/202505007

ELT-RTDETR:轻量化变电站缺陷检测的 Transformer 增强模型

刘梓良¹, 尼鹿帕尔·艾克木², 伊力哈木·亚尔买买提³

(1. 新疆大学 智能科学与技术学院, 乌鲁木齐 830017; 2. 国网乌鲁木齐供电公司, 乌鲁木齐 830000;
3. 新疆大学 电气工程学院, 乌鲁木齐 830017)

摘要:为解决变电站缺陷检测任务中存在的实时性不足、计算资源消耗高,以及实际环境下检测精度低等问题,提出了一种轻量级目标检测模型 ELT-RTDETR。首先,采用 EfficientFormerV2 作为主干网络,结合局部卷积与轻量化 Transformer 设计,显著降低模型参数量与计算开销。其次,提出轻量化多尺度特征金字塔网络 LMSFPN,通过多尺度深度卷积、加权融合与高效上采样策略,增强多尺度缺陷特征的表达,同时减少冗余计算。最后,引入基于 TSSA 注意力机制的 TSSACFE 模块,通过局部统计建模与低维投影优化特征交互,有效提升微小缺陷的检测鲁棒性。结果显示,在自建的变电站设备缺陷数据集上,ELT-RTDETR 的检测精度达 82.1%,较传统 RT-DETR 提升 7.3%,同时模型计算量和参数量分别降低 63.2% 与 50.7%。消融实验与主流算法对比结果表明,该模型在精度、轻量化,以及推理效率上均优于 YOLO 系列与现有 RT-DETR 变体,尤其在表计外壳破损和硅胶桶变色等任务中表现突出。本研究为变电站环境下的实时缺陷检测提供了高效解决方案,具备显著的工程应用潜力。

关键词: 变电站缺陷检测;轻量化模型;RT-DETR;TSSACFE 模块;多尺度特征融合

中图分类号: TM63;TP183;TP391.41 文献标志码: A 文章编号: 0367-6234(2026)04-0141-15

ELT-RTDETR: lightweight Transformer-enhanced model for substation defect detection

LIU Ziliang¹, NILUPAER·Aikemu², YILIHAMU·Yaermainaiti³

(1. School of Intelligence Science and Technology, Xinjiang University, Urumqi 830017, China; 2. State Grid Urumqi Power Supply Company, Urumqi 830000, China; 3. School of Electrical Engineering, Xinjiang University, Urumqi 830017, China)

Abstract: To address the problems of insufficient real-time performance, high computational resource consumption, and low detection accuracy in practical environments for substation defect detection tasks, this paper proposed a lightweight object detection model, namely ELT-RTDETR. First, EfficientFormerV2 was adopted as the backbone network, combining local convolutions with a lightweight Transformer design to significantly reduce the number of model parameters and computational overhead. Second, a lightweight multi-scale feature pyramid network (LMSFPN) was proposed to enhance the expression capability of multi-scale defect features through multi-scale depth-wise convolutions, weighted fusion, and efficient upsampling strategies, while reducing redundant computations. Finally, a token statistics self-attention cross-feature enhancement (TSSACFE) module based on the token statistics self-attention (TSSA) mechanism was introduced. This module optimized feature interaction through local statistical modeling and low-dimensional projection, effectively improving the detection robustness of small defects. Results show that on a self-built substation equipment defect dataset, the detection accuracy of ELT-RTDETR reaches 82.1%, which is 7.3% higher than that of the traditional RT-DETR. Meanwhile, the model calculation volume and parameter count are reduced by 63.2% and 50.7%, respectively. Ablation experiments and comparisons with mainstream algorithms demonstrate that the proposed model outperforms the YOLO series and existing RT-DETR variants in terms of accuracy, light weight, and inference efficiency, especially in tasks such as meter shell damage and silica gel canister discoloration. This study provides an efficient solution for real-time defect detection in substation environments, possessing significant potential for engineering applications.

Keywords: substation defect detection; lightweight model; RT-DETR; TSSACFE module; multi-scale feature fusion

收稿日期: 2025-05-06; 录用日期: 2025-07-14; 网络首发日期: 2026-03-09

网络首发地址: <https://link.cnki.net/urlid/23.1235.T.20260309.1051.004>

基金项目: 新疆厅厅联动项目(2023B01006); 新疆露天矿智能生产与管控重点实验室项目(XJQY2007)

作者简介: 刘梓良(1999—),男,硕士研究生

通信作者: 伊力哈木·亚尔买买提,65891080@qq.com

变电站是电力系统中的重要环节,负责电能的转换、分配和传输,其安全稳定运行直接关系到电网的可靠性。然而,长期运行过程中,变电站设备容易受到环境因素以及电气负荷波动的影响,导致设备出现过热、放电、绝缘老化,以及机械损伤等缺陷。如果这些缺陷不能及时发现和处理,可能会引发设备故障,甚至导致大范围停电,严重影响社会经济运行,因此高效、智能的变电站缺陷检测技术对于保障电网安全至关重要。

近年来,变电站缺陷检测技术经历了从传统机器学习到深度学习的发展历程。从技术路线角度,当前研究主要分为 3 个发展阶段。早期研究主要采用人工特征提取与机器学习分类器相结合的方式,如支持向量机、随机森林等。这类方法虽然计算复杂度低,但特征表达能力有限,在复杂场景下检测精度不足。随着深度学习技术的发展,基于 CNN 的目标检测算法在变电站缺陷检测中得到广泛应用。Ren 等^[1]首次将 Faster R-CNN 应用于电网设备检测,实现了 90% 以上的检测精度。Liu 等^[2]采用 YOLOv3 实现绝缘子损坏检测,通过数据增强提升了模型鲁棒性。Wang 等^[3-4]分别基于 YOLOv5m 和 YOLOv8 提出了轻量化变电站缺陷检测模型,在保持检测精度的同时显著降低了计算复杂度。Feng 等^[5]提出了基于多尺度特征融合的变电站设备检测方法,在小目标检测方面取得了较好效果。然而,这类方法普遍存在网络结构复杂、参数量大、难以在资源受限设备上部署等问题。随着 Vision Transformer 在计算机视觉领域的成功应用,基于 Transformer 的目标检测方法逐渐兴起。李刚等^[6]将 DETR 应用于电力设备检测,利用其端到端训练的优势避免了复杂的后处理步骤。Xi 等^[7]提出了基于 Swin Transformer 的变电站缺陷检测框架,在复杂背景下表现出更强的鲁棒性。但现有基于 Transformer 的方法计算复杂度较高,实时性有待改善。

当前变电站缺陷检测研究多集中于特定场景下的单一缺陷识别,且现有模型普遍存在计算复杂度、部署困难等瓶颈。针对上述问题,本文提出轻量化变电站缺陷检测模型 ELT-RTDETR (efficient lightweight transformer-RTDETR)。首先,采用 EfficientFormerV2 作为 RT-DETR 骨干网络,通过层次化特征编码降低冗余计算。其次,引入基于 TSSA (token statistics self-attention) 注意力机制的 TSSACFE (token statistics self-attention cross-feature enhancement) 模块,提升密集缺陷的语义表征能力。最后,设计轻量化多尺度特征金字塔 LMSFPN (lightweight multi-scale feature pyramid network),融合深度可卷积与通

道重排技术,在减少计算量的同时保留多尺度特征完整性。通过实验验证 ELT-RTDETR 的有效性,旨在在复杂工况下实现多目标缺陷的实时精准定位,保障电网设备智能化高效运行。

1 相关工作

1.1 RT-DETR 网络

RT-DETR (real-time detection transformer)^[8] 是百度于 2023 年推出的实时端到端目标检测模型,它成功解决了传统模型 DETR (detection transformer) 推理速度慢的问题,在保持 Transformer 架构优势的同时实现了实时检测性能。RT-DETR 采用混合编码器架构,结合了 CNN 的局部特征提取能力和 Transformer 的全局建模能力。具体包括:尺度内特征交互 (attention-based intrascale feature interaction, AIFI) 模块,用于增强同一尺度内的特征表达;跨尺度特征融合模块 (cross-scale context fusion module, CCFM),实现不同尺度特征间的有效融合。RT-DETR 采用不确定性最小化策略来优化查询向量的选择。通过最小化预测结果的不确定性,模型能够更加专注于高置信度的检测结果,减少误检和漏检。

与 YOLO 系列模型相比,RT-DETR 具有以下优势:1) 无需 NMS 后处理,端到端训练避免了复杂的非极大值抑制步骤;2) 更强的小目标检测能力,Transformer 的全局建模能力使其在小目标检测方面表现更优;3) 更好的泛化性能,注意力机制使模型对不同场景具有更强的适应性。RT-DETR 的网络架构如图 1 所示,主要包括 3 个部分:Backbone 负责提取多尺度特征图、Neck 部分的混合编码器进行特征增强和融合,以及 Head 部分通过 Transformer 解码器完成最终的目标检测和分类。

1.2 ELT-RTDETR 网络结构

为适应变电站缺陷检测对实时性、高精度,以及轻量化的多重需求,本文提出了一种轻量化变电站缺陷检测模型—ELT-RTDETR。该模型在保留 RT-DETR 强大特征建模能力的基础上,针对变电站复杂环境下设备缺陷检测任务进行了系统性的轻量化与精细化改进,整体架构如图 2 所示。首先,在特征提取阶段,采用 EfficientFormerV2-S0 作为主干网络,以实现高效、低功耗的多尺度特征提取。其次,为提升模型对多尺度缺陷目标的检测能力,引入了自研的 LMSFPN 模块。该模块通过多尺度高效卷积、加权特征融合,以及轻量上采样等策略,能够在减少冗余计算的同时,实现不同尺度特征的高效融合与增强。最后,在特征交互阶段,针对传统 AIFI 模块在局部建模与计算开销方面不足的问题,提出

了基于 TSSA 注意力机制改进的 TSSACFE 模块。该模块通过局部统计特性建模,实现轻量、高效的跨尺度特征交互,进一步强化了模型在复杂场景下的

目标判别能力。综合上述改进,ELT-RTDETR 在保证检测精度的同时,显著降低了模型体积和计算负担,具备较强的实时推理能力和边缘部署适应性。

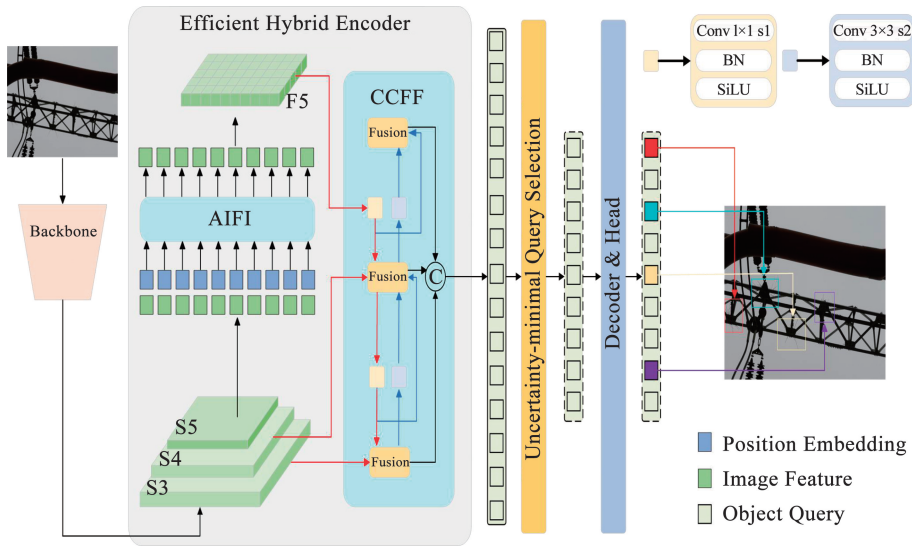


图 1 RT-DETR 网络结构

Fig. 1 RT-DETR network structure

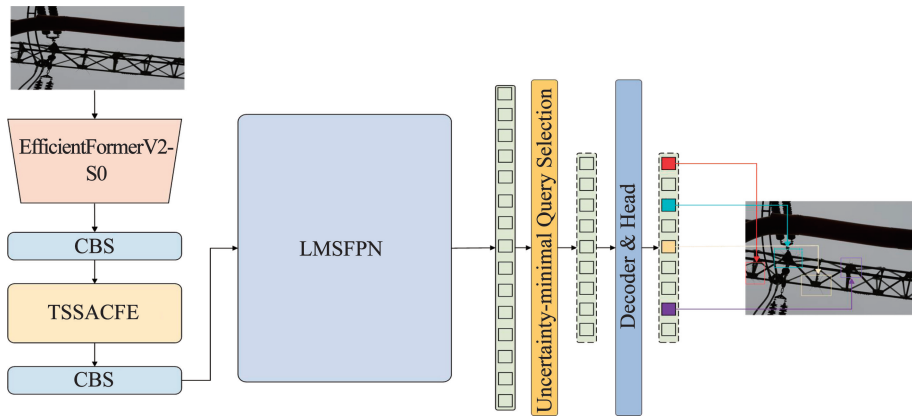


图 2 ELT-RTDETR 网络结构

Fig. 2 ELT-RTDETR network structure

1.3 改进主干网络

由于电力系统设备分布广泛,检测任务多部署于资源受限的边缘终端,如巡检机器人、无人机和手持设备等,因此模型不仅须具备较高的检测精度,还应具备低计算开销、低存储需求和快速响应能力。因此,亟须一种能够在保证检测性能的同时,满足轻量部署需求的目标检测模型,以实现对多类缺陷的高效、可靠识别。

为实现变电站场景下缺陷目标的高效检测,近年来广泛采用基于 Transformer 的目标检测框架如 RT-DETR。然而,RT-DETR 原生主干网络通常结构复杂、参数量大,难以满足边缘设备对模型体积、计

算量和实时性的要求,限制了其在实际工程中的部署与推广。与此同时,轻量级视觉模型的快速发展为解决该问题提供了新思路。EfficientFormerV2-S0^[9]作为一种专为资源受限环境设计的高效主干网络,兼顾卷积结构的局部特征提取能力与 Transformer 的全局建模能力,具备较低的参数量和浮点运算开销,同时能保持良好的检测性能。因此,本文提出将 RT-DETR 的主干替换为 EfficientFormerV2-S0,以实现对变电站复杂缺陷目标的轻量化检测,提升模型在实际场景中的实用性与可部署性。EfficientFormerV2 结构如图 3 所示。

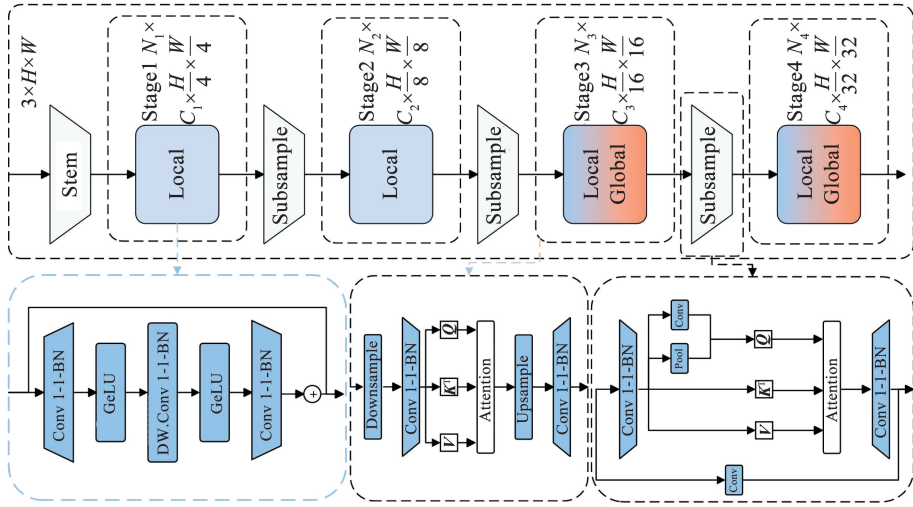


图 3 EfficientFormerV2 结构

Fig. 3 EfficientFormerV2 structure

EfficientFormerV2-S0 通过 4 阶段分层设计实现高效特征提取;第 1 阶段利用 Stem 和局部卷积提取 32 通道特征图并下采样至 160×160 ;第 2 阶段进一步降采样至 80×80 ,通过 Local Blocks 强化局部细节建模;第 3、4 阶段融合局部卷积与全局注意力机制,分别在 40×40 和 20×20 分辨率下增强多尺度感知能力,最终通过全局池化实现轻量分类。EfficientFormerV2-S0 骨干网络的具体参数见表 1。

表 1 EfficientFormerV2 骨干特征提取网络结构

Tab. 1 EfficientFormerV2 backbone feature extraction network structure

Stage	计算块	通道数	特征图尺寸($H \times W$)
Stage1	Stem + Local	32	160×160
Stage2	Local Blocks	48	80×80
Stage3	Local Global	96	40×40
Stage4	Local Global	176	20×20

1.4 轻量化多尺度特征金字塔

1.4.1 传统特征金字塔存在的局限性

传统特征金字塔在复杂场景检测中存在明显依赖逐层上采样与简单拼接或加法操作,缺乏对多尺度特征重要性的自适应建模,导致小目标特征易被淹没、大目标细节丢失的缺陷;此外,密集横向连接与重复特征处理引入冗余计算,使得参数量与计算复杂度大幅增加,难以满足边缘设备的实时性需求。这些问题共同导致传统 FPN 在变电站多尺度缺陷检测中面临精度低、效率差的双重挑战。

1.4.2 LMSFPN 模块设计

针对传统特征金字塔网络在变电站缺陷检测任务中存在的特征冗余、计算复杂度高,以及多尺度建模能力不足等问题,本文提出了一种轻量化多尺度

特征金字塔网络 LMSFPN,该模块通过融合 BIFPN (bidirectional feature pyramid network)^[10]的加权特征融合策略、MAF-YOLO (multi-branch auxiliary fusion YOLO)^[11]的多尺度卷积结构,以及 EMCAD (efficient multi-scale convolutional attention decoding)^[12]的高效上采样思想的基础上,构建了适用于边缘部署的轻量化特征融合框架,其中 BIFPN 结构如图 4 所示。

BIFPN (双向特征金字塔网络)不仅包含自顶向下的路径,还引入了自底向上的路径,使得底层特征能够直接影响高层特征的代表。其通过跳跃连接的方式,同一尺度的特征在不同处理阶段能够直接融合,增强了特征的表达能。同时每个特征融合节点都引入了可学习的权重参数,通过加权融合方式,自适应地调整不同输入特征的重要性。具体而言,BIFPN 的特征融合过程可以通过下式计算:

$$P_i^{\text{out}} = \text{Conv} \left(\frac{\sum_{j \in N(i)} w_{i,j} \cdot \text{Resize}(P_j^{\text{in}})}{\sum_{k \in N(i)} w_{i,k}} \right) \quad (1)$$

式中: P_i^{out} 、 P_j^{in} 分别为第 i 层的输出特征图和第 j 层的输入特征图; $N(i)$ 为节点 i 的邻接节点集合,即与节点 i 相连的所有节点; $w_{i,j}$ 为节点 i 对来自节点 j 的特征的可学习权重;Conv(\cdot) 为卷积操作,通常包含卷积层、批归一化和激活函数;Resize(\cdot) 为尺寸调整操作,当 $j < i$ 时执行上采样操作如双线性插值或转置卷积,当 $j > i$ 时执行下采样操作如最大池化、平均池化或步长卷积,当 $j = i$ 时则为恒等操作无需调整尺寸。这种设计使得 BIFPN 能够更好地平衡不同尺度特征的贡献。

LMSFPN 模块通过以下创新设计实现高效特征融合与多尺度感知能力的平衡:首先,借鉴 TridentNet^[13]思想,不同尺度的特征层适用于不同感受野的卷积核。通过并行部署不同尺寸的深度卷

积核,自适应捕获多尺度缺陷目标的局部细节与全局语义特征。同时结合通道重排机制,增强跨通道

信息交互能力,提升微小缺陷的表征鲁棒性。不同尺度的特征图对应的感受野分布如图5所示。

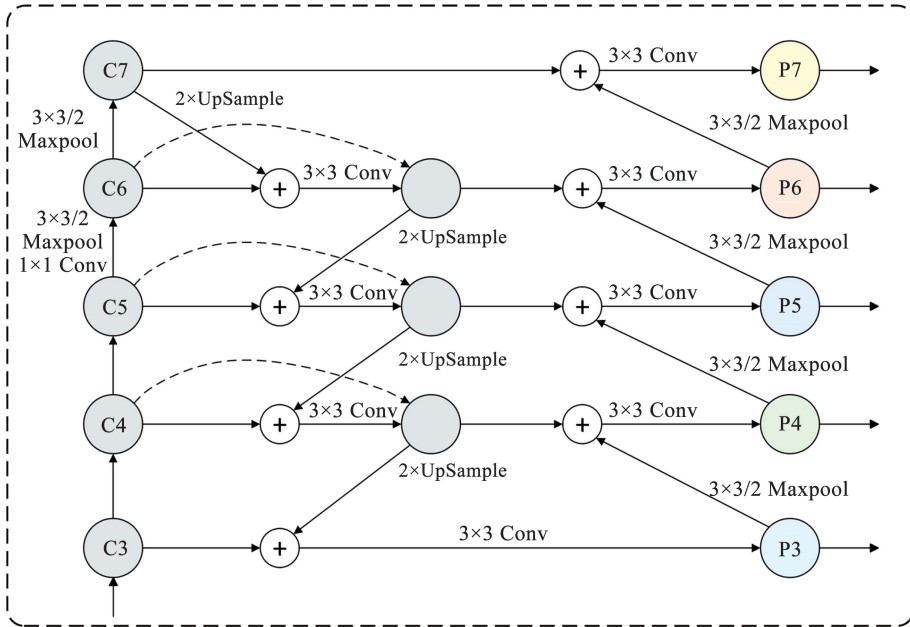


图4 BIFPN 网络结构

Fig. 4 BIFPN structure

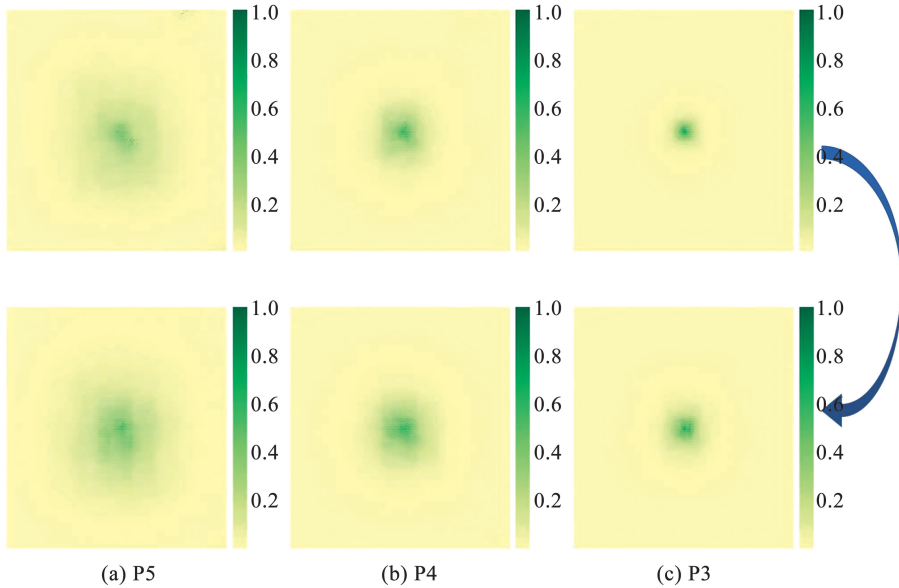


图5 通过不同大小和深度的卷积产生的有效感受野(ERF)(暗区表示更大的ERF)

Fig. 5 Effective receptive field (ERF) generated by convolutions of different sizes and depths (darker areas indicate larger ERF)

从图5可以看出,P5具有较大感受野的网络更适合检测较大的物体,反之,较小尺度的目标则从较小的感受野中受益,因此本文在FPN(feature pyramid network)阶段,对于不同尺度的特征层选择不同的多尺度卷积核适应并逐步获得多尺度感知场信息。

其次,摒弃传统拼接操作,提出基于可学习权重参数的轻量级加法融合策略。通过归一化权重动态平衡不同尺度特征的重要性,减少冗余计算的同时

优化特征表达。

再次,引入来自EMCAD的EUCB(efficient upsample convolution block)作为上采样结构,采用深度卷积与通道重排机制构建轻量上采样层,以低计算成本恢复高分辨率特征图,避免传统双线性插值导致的细节模糊问题,保留关键缺陷信息。

然后,引入跨阶段部分连接(cross stage partial, CSP)结构,将特征流拆分为主分支与残差分支,分别执行多尺度卷积与特征整合。该设计通过分离梯

度路径降低信息冗余,提升训练稳定性与特征复用效率。

述模块,在保证多尺度特征交互能力的前提下,显著降低模型参数量与计算开销。

如图 6 所示,LMSFPN 通过层级化结构集成上

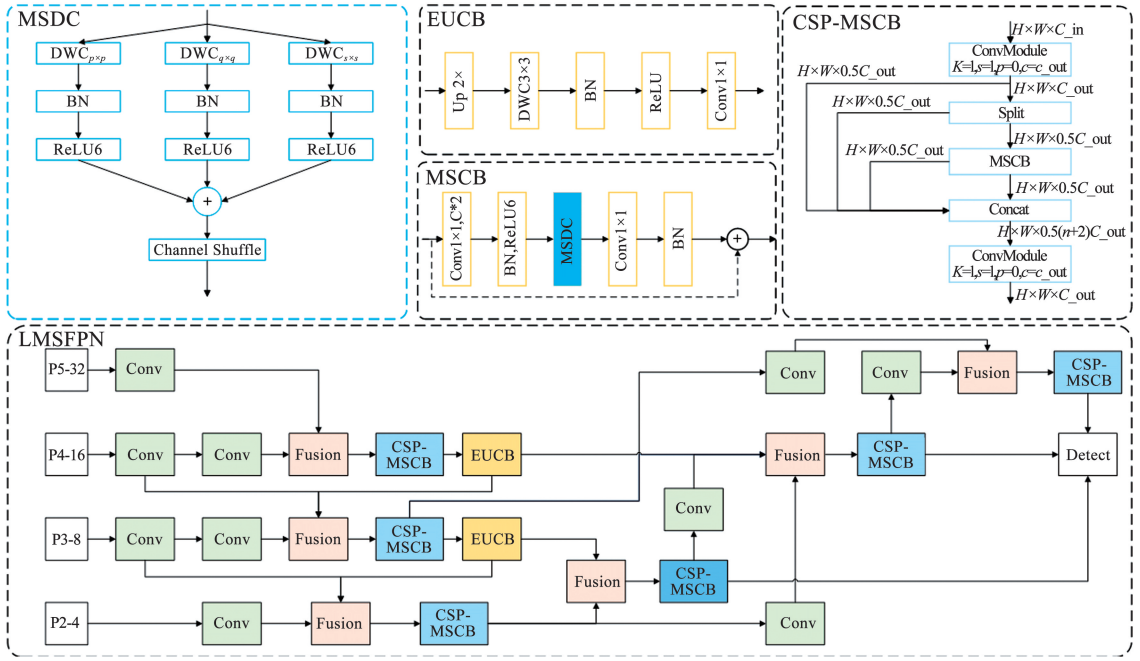


图 6 LMSFPN 网络架构

Fig. 6 LMSFPN architecture

1.5 TSSACFE 模块:基于 TSSA 的 AIFI 改进设计

1.5.1 AIFI 在变电站缺陷检测中的局限性

RT-DETR 中 AIFI 模块依赖全局注意力机制,在复杂变电站场景中存在对微小目标的局部细节捕捉能力不足,且高分辨率特征图上的密集计算导致复杂度与显存占用过高,难以满足边缘端实时性需求;同时,缺乏对局部统计特性的建模能力,在背景干扰或目标模糊时易引发误检与漏检。因此,亟须一种轻量化改进方案,以增强局部敏感性并优化计算效率。

1.5.2 引入 TSSA 的动机

为解决原始 AIFI 存在空间敏感性较弱,难以有

效聚焦于小尺度细粒度缺陷区域,且在高分辨率图像中存在计算与存储开销过大的问题,本文引入了一种基于 Token 统计的轻量注意力机制 TSSA^[14],其结构图如图 7 所示,该模块通过多头投影获取低维 Token 表征,并基于其统计特性构建对角注意力,从而在保持注意力建模能力的同时显著降低计算复杂度。因此,本文提出将 TSSA 融入 AIFI 增强其聚焦小目标特征的能力,构建全新的 TSSACFE 模块,提升特征交互建模能力,从而更高效地服务于变电站缺陷检测的目标识别与定位任务。

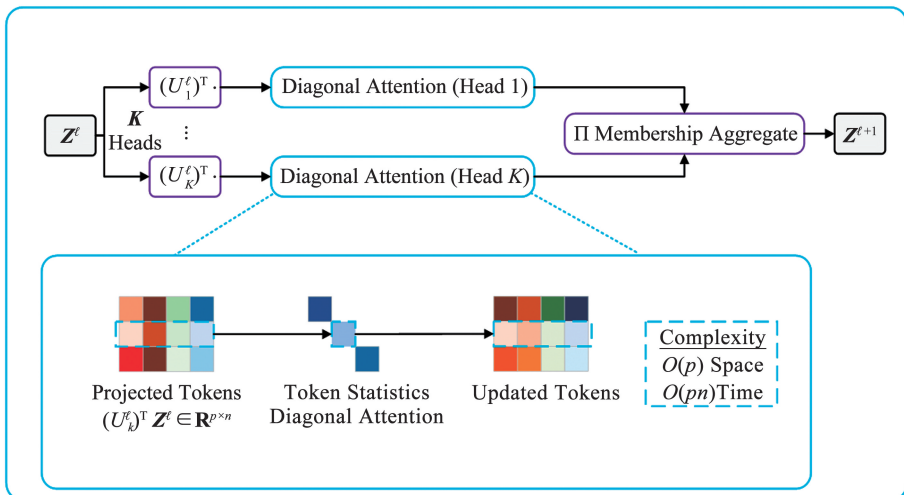


图 7 TSSA 网络结构

Fig. 7 TSSA network structure

1.5.3 TSSACFE 模块结构与工作流程

TSSACFE 模块通过轻量化统计注意力机制优化特征交互。首先,输入特征图被展平为 Token 序列,通过多头投影映射至低维空间;在每个子空间内,基于 Token 的均值、方差生成对角注意力矩阵,替代传统全局计算,聚焦微小缺陷区域并抑制噪声。随后,各子空间输出通过 Membership Aggregate 机制融合,整合多尺度语义信息,增强跨层级特征表达能力。

从理论角度分析,传统的全注意力机制计算复杂度为 $O(N^2)$,其中 N 为序列长度。TSSA 机制通过引入统计建模,将注意力计算转化为基于 Token 统计特性的对角矩阵运算,从而将复杂度降低至 $O(N)$ 。具体而言,对于输入特征 $X \in \mathbf{R}^{N \times D}$,TSSA 首先通过下式分别计算每个 Token 的统计特征:

$$\mu_i = \frac{1}{D} \sum_{j=1}^D X_{i,j} \quad (2)$$

$$\sigma^2 = \frac{1}{D} \sum_{j=1}^D (X_{i,j} - \mu_i)^2 \quad (3)$$

然后构建对角注意力矩阵 $A \in \mathbf{R}^{N \times N}$,其中 $A_{i,i} = \text{Softmax}(\mu_i / \sqrt{D})$,非对角元素为 0。相比原始 AIFI 模块,TSSACFE 的参数量减少主要体现在注意力计

算部分,具体减少量为 $(N^2 - N) \times H$ 个参数,其中 H 为注意力头数。在本文实验设置下,参数量减少 0.7%。

2 实验

2.1 数据集

本文使用的变电站电力设备缺陷数据集的图像采集自变电站现场,确保模型能适应实际应用场景,其中包含 6 004 张变电站设备缺陷图像,根据预定义的 7:2:1 比例,将数据集分成 3 个子集:4 202 张图像用于训练数据集,601 张图像用于验证数据集,1 201 张图像用于测试数据集,以确保模型的充分训练和泛化能力。

鉴于本文的主要重点在于分析计算机视觉算法在变电站设备缺陷检测中的作用,本文基于缺陷在外观特征上的差异,构建了针对 9 类典型缺陷的视觉识别体系。这 9 类常见缺陷包括:表计读数异常、表计外壳破损、异物鸟巢、空中漂浮物、表盘模糊、表盘破损、绝缘子破裂、地面油污,以及硅胶桶变色,其中各类缺陷具有可感知的图像特征。关于数据集标注的具体信息详见表 2。

表 2 变电站缺陷数据集标签信息

Tab. 2 Label information of substation defect dataset

数据集	数量	表计读数异常	表计外壳破损	异物鸟巢	表盘模糊	表盘破损	绝缘子破裂	地面油污	硅胶桶变色	空中漂浮物
Train	4 202	522	358	621	597	497	278	533	746	501
Val	601	83	50	73	93	76	43	99	98	61
Test	1 201	157	114	175	171	134	81	161	217	130
Total	6 004	762	522	869	861	707	402	793	1 061	692

2.2 实验配置及其参数选择

本实验操作系统采用 Windows11 家庭中文版,搭载了第 13 代 Intel Core i9-13980HX 处理器。GPU 方面使用 RTX4060 显卡,并运行在 CUDA12.1 版本上。实验基于深度学习框架 Pytorch2.10,并由 PyCharm 集成开发环境配合 Python3.9 进行设计开发,没有使用预训练模型进行迁移学习。详细的实验参数详见表 3。

表 3 训练参数设置

Tab. 3 Training parameter settings

参数	值	参数	值
epochs	150	Optimizer	AdamW
batch-size	4	image_size	640
lr0	1×10^{-4}	close_mosaic	0
worker	4	Weight_decay	1×10^{-4}

2.3 评价指标

为了准确评估模型改进效果,使用一些主流评估如 mAP50、mAP50-95、AP_s、AP_m、AP_l 等作为模型性能的评估指标,详细信息如下。

mAP50 是指当 IoU (intersection over union) 阈值设定为 0.50 时的平均精度均值。IoU 为预测框与真实框的重叠度,当 $\text{IoU} \geq 0.50$ 时认为检测正确。mAP50-95 为 IoU 阈值从 0.50 ~ 0.95 范围内所有阈值下 AP 值的平均。其中:AP_s 为小目标 (面积 < 32² 像素) 的平均精度,AP_m 为中等目标 (32² ≤ 面积 < 96² 像素) 的平均精度,AP_l 为大目标 (面积 ≥ 96² 像素) 的平均精度,这 3 个指标分别评估模型对不同尺寸目标的检测能力。

模型参数量 (Params) 指的是模型的总参数数量,反映模型的存储需求。计算量 (GFLOPs) 指的是 10 亿次浮点运算数,衡量模型的计算复杂度。每秒

帧数(FPS)指的是每秒处理帧数,反映模型的推理速度。

精确率 P (precision) 为正样本集中正确预测样本的比例,通过下式计算:

$$P = \frac{T_p}{T_p + F_p} \quad (4)$$

式中: T_p 为正确预测对象, F_p 为错误预测对象。

召回率 R (recall) 为实际阳性且预测正确的样本比例,通过下式计算:

$$R = \frac{T_p}{T_p + F_N} \quad (5)$$

式中 F_N 为存在但未被正确检测的对象。

2.4 对比实验

2.4.1 主干对比实验

变电站通常在工业环境中运行,模型的实际应用需要兼顾精度和实时性。如果使用的主干过大,部署到低算力设备上可能导致推理速度下降,影响实时检测。传统的 RT-DETR 采用 ResNet、Swin Transformer 等主干,虽然检测效果较好,但计算复杂度较高,不适用于边缘计算设备或嵌入式系统。因此,实验对比不同轻量化主干,找寻最优的轻量级方案,其中:不同 Backbone 结构参数量对比如图 8 所示,实验结果见表 4。

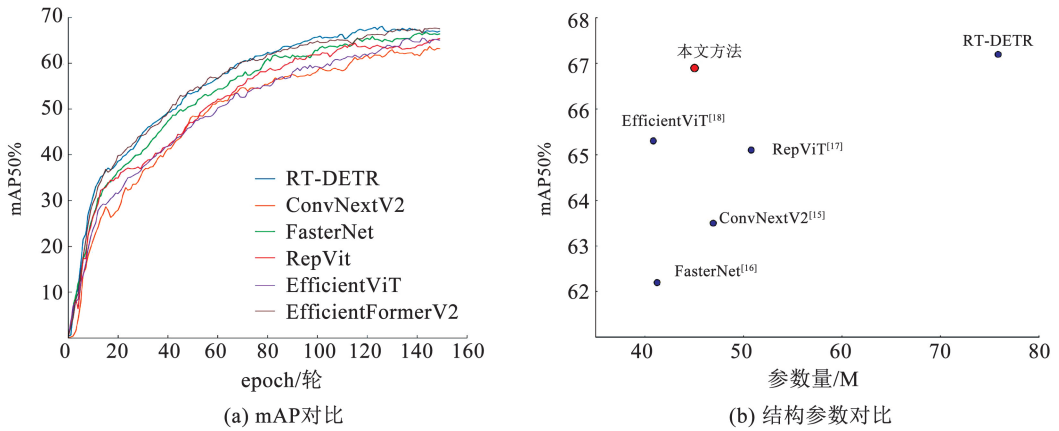


图 8 主干结构参数与 mAP50 的比较

Fig. 8 Comparison of backbone structure parameters with mAP50

表 4 更换骨干网后模型性能比较

Tab. 4 Comparison of model performance after replacing backbone network

模型	参数量/M	计算量/($10^9 s^{-1}$)	$AP_s/\%$	$AP_m/\%$	$AP_l/\%$	$F_1/\%$	$P/\%$	mAP50/%
RT-DETR	75.85	57.0	15.1	35.4	47.5	68.0	74.8	67.2
ConvNextV2 ^[15]	46.96	31.9	3.1	31.0	44.0	64.0	69.9	63.5
FasterNet ^[16]	41.24	28.5	10.9	30.8	47.0	67.0	75.9	62.2
RepVit ^[17]	50.79	36.4	8.7	30.0	45.4	66.0	76.5	65.1
EfficientViT ^[18]	40.87	27.3	11.8	24.7	46.4	65.0	76.0	65.3
本文	45.05	29.5	17.4	37.6	48.1	68.0	80.8	66.9

实验结果表明,采用提出的轻量化主干后,模型的参数量与计算量显著减少,同时在 AP_s 、 AP_m 、 AP_l 等指标上仍能保持竞争力甚至超越部分传统主干。这说明,在复杂的变电站缺陷检测场景中,轻量化主干不仅能够大幅降低模型的计算和存储成本,而且可以兼顾检测精度与速度,实现更优的综合性能。

2.4.2 特征金字塔对比实验

由于变电站环境中存在大量遮挡、光照变化和目標重叠现象,这对特征金字塔的上下文建模能力提出了更高要求。通过对比不同结构的表现,可以筛选出更适合复杂电力场景的特征融合方案。同

时,通过对比实验可以有效评估所提出的 LMSFPN 结构在轻量化方面的优势,从而验证其在资源受限设备上的部署潜力。实验结果如表 5、图 9 所示。

从表 5 与图 9 的实验结果可以看出,尽管部分方法在单项指标上略有提升,如 MAFPN 在 mAP50 上略高,但其参数量高达 87.49 M,计算量(GFLOPs)也达到 56.4,整体轻量化程度不及 LMSFPN。此外,与原始结构相比,LMSFPN 在大幅降低参数量与计算量的基础上,检测性能基本持平甚至有所提升,进一步验证了其在资源受限场景下的部署潜力。

表 5 特征金字塔替换的性能比较

Tab. 5 Performance comparison of feature pyramid replacement

模型	参数量/M	计算量/(10 ⁹ s ⁻¹)	AP _s /%	AP _m /%	AP _l /%	mAP50/%	推理速度/fps
RT-DETR	75.85	57.0	15.1	35.4	47.5	67.2	88.49
MAFPN ^[11]	87.49	56.4	25.6	32.7	49.3	70.1	84.76
HSFPN ^[19]	69.14	53.3	18.0	33.4	48.5	69.5	86.21
CA-HSFPN ^[20]	69.02	53.4	25.2	35.9	48.4	69.8	86.96
ELA-HSFPN ^[21]	76.06	53.7	23.0	30.8	49.1	69.6	86.21
LMSFPN	78.47	48.7	23.4	34.8	49.2	69.6	71.43

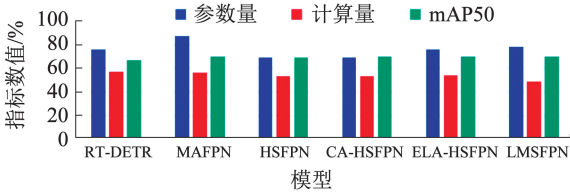


图 9 替换原始特征金字塔的影响

Fig. 9 Impact of replacing original feature pyramid

2.4.3 AIFI 对比实验

原始的 AIFI 虽具备较强的全局交互能力,但对于变电站缺陷检测这类目标种类少、差异小、缺陷区域尺寸小且背景复杂的任务,其特征提取能力可能存在不足。为了验证改进后的 AIFI 模块在变电站缺陷检测任务中的有效性,进行不同改进结构的对比实验。不同 AIFI 改进对比实验结果见表 6。

表 6 AIFI 对比实验结果

Tab. 6 Comparison experiment results of AIFI

模型	P/%	R/%	mAP50/%	mAP50-95/%	参数量/M	计算量/(10 ⁹ s ⁻¹)
RT-DETR	74.8	64.2	67.2	46.4	75.85	57.0
AIFI-DHSA ^[22]	81.3	65.5	70.6	49.4	76.39	57.3
AIFI-DTAB ^[23]	79.3	67.3	70.3	49.3	82.99	58.7
AIFI-EfficientAdditive ^[24]	84.2	62.0	68.6	48.0	75.85	57.2
AIFI-HiLo ^[25]	78.6	64.8	69.7	48.4	75.72	57.1
AIFI-LPE	76.6	65.6	69.2	47.5	76.24	57.0
AIFI-RepBN ^[26]	78.0	63.2	68.1	47.2	75.85	57.0
TSSACFE	82.7	65.8	70.6	49.0	75.34	57.1

从表 6 的实验结果可以看出,多数改进后的 AIFI 模块在精确率、召回率、mAP50 和 mAP50-95 等指标上均优于原始的 RT-DETR 架构,验证了改进 AIFI 模块在变电站缺陷检测任务中的有效性。其中,TSSACFE 模块在各项指标上表现尤为突出,在保持较小计算量和参数量增长的前提下,实现了并列最高的 mAP50,同时精度与召回率也处于领先水平。这表明 TSSACFE 能更有效地捕捉关键特征,进一步凸显其在轻量化与高精度之间的好权衡,具有较强的实际应用潜力。

为了全面评估所提出的 TSSACFE 模块在变电站缺陷检测任务中的性能优势,本文对比分析了原始 RT-DETR 与多种 AIFI 模块在不同训练指标下的表现,相关结果如图 10 所示。从精确率和召回率曲线可以看出,TSSACFE 均展现出更高的稳定性和准确性;在 mAP50 以及 mAP50-95 指标上,其曲线整体处于最高位置,验证了其更优的目标检测能力。

2.4.4 主流算法对比实验

为了验证所提出的 ELT-RTDETR 模型在真实场景中的改进性能,本文使用几种流行的目标检测模型,在相同的数据集和相同的实验条件下进行了对比实验。实验结果见表 7。

从表 7 可以看出,本文提出的 ELT-RTDETR 模型在轻量化变电站缺陷检测任务中表现出了明显的优越性。与同为轻量化设计的 YOLO 系列模型进行对比,ELT-RTDETR 在各项核心指标上均取得了更为优异的结果。具体来看,在检测准确性方面,ELT-RTDETR 的精确率达到 82.1%,远高于 YOLOv8s 以及其他轻量化模型;在 mAP50-95 指标上,本文模型取得了 46.4%,同样高于 YOLOv8s。在模型规模与计算复杂度方面,ELT-RTDETR 的参数量仅为 37.14 M,明显低于 YOLOv6s 和 YOLOv8s,同时计算量(GFLOPs)仅为 21.0,表明模型更适合边缘设备部署和实际应用场景。

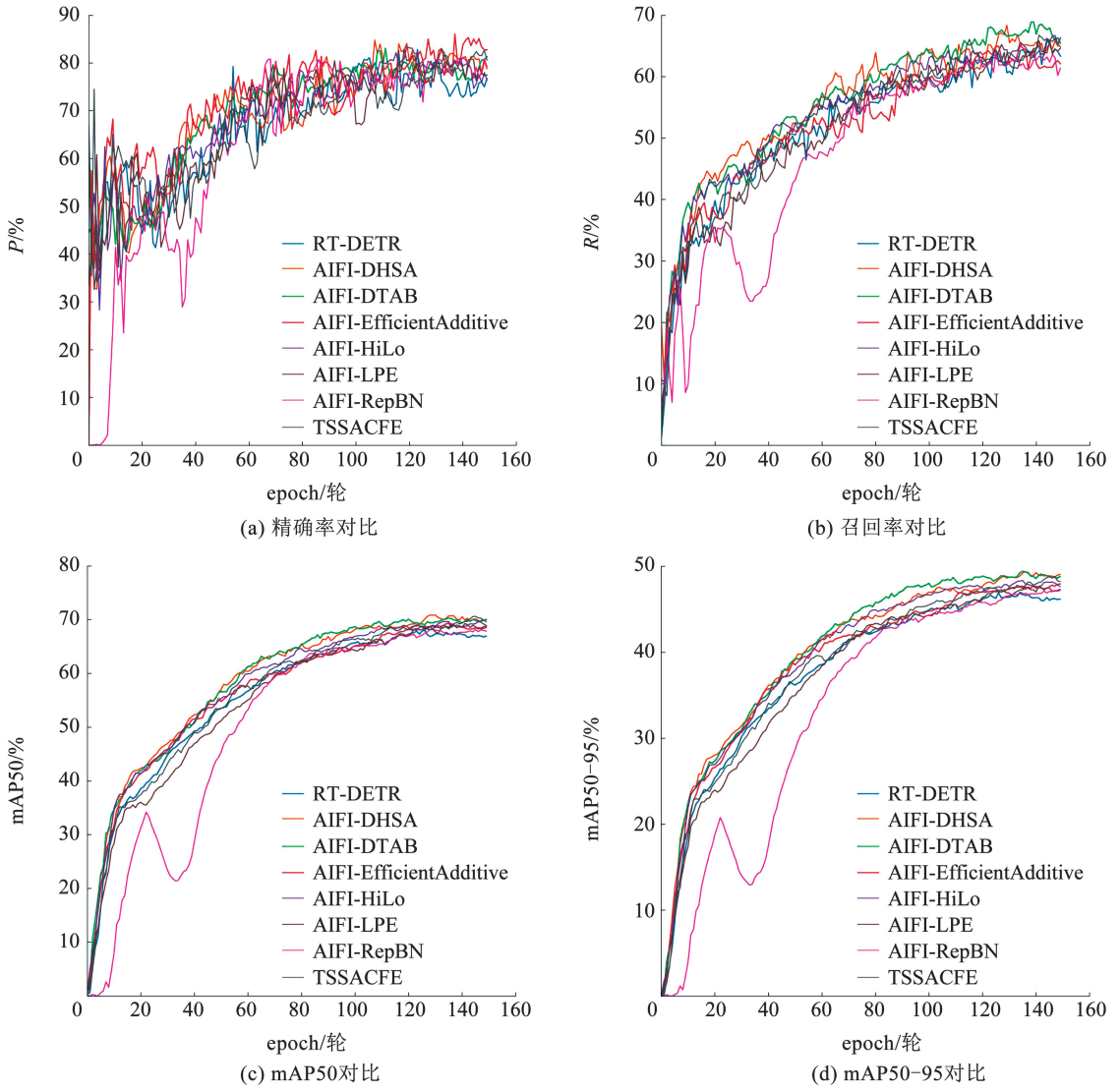


图 10 各种 AIFI 的性能曲线比较

Fig. 10 Performance curve comparison of various AIFIs

表 7 主流算法对比实验结果

Tab. 7 Comparison experiment results of mainstream algorithms

模型	P/%	R/%	mAP50/%	mAP50-95/%	参数量/M	计算量/(10 ⁹ s ⁻¹)	推理速度/fps
Heavyweight model							
YOLOv3	51.9	21.4	26.0	18.6	395.47	282.2	40.49
YOLOv5m	62.9	46.4	50.5	34.0	95.55	64.0	105.26
YOLOv6m	57.9	39.0	42.5	29.0	198.29	161.2	33.78
YOLOv8m	77.3	62.0	68.4	47.0	98.59	78.7	91.74
RTDETR-18	74.8	64.2	67.2	46.4	75.85	57.0	91.37
RTDETR-34	81.1	62.8	68.6	46.7	118.71	88.8	34.63
RTDETR-50	76.2	62.6	67.4	46.9	160.11	129.6	39.41
Lightweight model							
YOLOv3-tiny	79.5	64.1	69.0	47.5	46.38	18.9	212.76
YOLOv5s	62.3	56.7	57.3	38.8	34.77	23.8	163.93
YOLOv6s	70.8	51.7	58.1	39.9	62.18	44.0	161.29
YOLOv8s	76.0	51.2	59.8	40.5	42.45	28.5	161.29
本文	82.1	62.0	67.4	46.4	37.14	21.0	69.23

ELT-RTDETR 的推理速度为 69.23 fps,虽然能够满足实时检测需求,但相比 YOLOv8s 的推理速度为 161.29 fps 确实存在差距。推理速度相对较慢主要源于 Transformer 架构中注意力机制的矩阵运算开销、多尺度特征融合的计算复杂度,以及 Transformer 解码过程的并行度限制。为进一步提升推理速度,本文提出了多层次的优化策略。在模型结构层面,可采用深度可分离卷积替换、知识蒸馏加速和网络剪枝技术来减少计算负担;在推理引擎层面,可通过 TensorRT 优化、INT8 量化和图优化技术实现加速;在实际部署中,需根据应用场景在速度与精度间进行权衡,并针对不同硬件平台进行专门的适配优化。这些优化方向为后续研究提供了明确的改进路径,有助于在保持检测精度的前提下显著提升模型的实用性。

2.5 消融实验

为进一步验证所提出改进模块在变电站缺陷检测中的有效性和必要性,本文设计了系统性的消融实验,分别对主干网络、特征金字塔结构,以及特征交互模块进行替换或修改,分析每一处改动对模型性能的影响。消融实验结果见表 8,其中,实验编号对应于以下描述。1#表示原始 RT-DETR 模型。2#

表示使用 EfficientFormerV2 作为骨干网络的 RT-DETR模型。3#表示使用自研轻量化特征金字塔 LMSFPN 相结合的 RT-DETR 模型。4#表示将提出的 TSSACFE 模块相结合的 RT-DETR 模型。5#表示结合改进的 EfficientFormerV2、LMSFPN 的 RT-DETR 模型。6#表示结合改进的 EfficientFormerV2、LMSFPN 和 TSSACFE 的 RT-DETR 模型。

从表 8 可以看出,本文提出的各改进模块对模型性能均带来了不同程度的优化,具体分析如下:首先将原 RT-DETR 模型的主干替换为 EfficientFormerV2,虽然 mAP50 略微下降,但模型参数量和计算量大幅降低,验证了轻量化主干在保持检测精度的同时显著提升模型效率的有效性。在引入自研特征金字塔 LMSFPN 后,mAP50 提升至69.2%,进一步表明合理设计的轻量级特征融合模块有助于改善小目标检测性能;同时,计算量(GFLOPs)明显下降,模型部署友好性增强。将 AIFI 模块替换为 TSSACFE 后,模型在精度、召回率与综合指标上均得到显著提升,说明 TSSA 注意力机制能够有效增强特征提取能力。最终,在同时整合三大改进模块后,模型在参数量仅为 37.14 M、计算量(GFLOPs)仅为21.0的情况下,依然保持了良好的检测性能。

表 8 消融实验结果

Tab. 8 Results of ablation experiment

编号	EfficientFormerV2	LMSFPN	TSSACFE	P/%	R/%	mAP50/%	参数量/M	计算量/(10 ⁹ s ⁻¹)
1#				74.8	64.2	67.2	75.85	57.0
2#	✓			80.8	60.9	66.9	81.00	29.5
3#		✓		81.1	63.1	69.2	68.47	48.7
4#			✓	82.7	65.8	70.6	75.35	57.1
5#	✓	✓		82.6	62.6	66.8	37.65	20.9
6#	✓	✓	✓	82.1	62.0	67.4	37.14	21.0

2.6 改进模型分析

表 9 描述了改进模型在检测不同类型的变电站缺陷时的性能,展现出高效综合性能,整体准确率达 80.8%,召回率 62.0%,mAP50 和 mAP50-95 分别为 67.4%与 46.4%。在表计外壳破损和硅胶桶变色检测任务中表现尤为突出,精确率分别达 91.8%和 97.7%,mAP50 高达 87.1%和 92.8%,验证其对显著特征变化的强鲁棒性;然而,异物鸟巢和地面油污检测精度较低,主要因目标结构松散、边界模糊,以及环境干扰导致特征区分困难,需进一步优化复杂背景下的特征建模能力。

为了直观地展示模型对不同类型变电站缺陷检测的能力,本文可视化了几个评估指标,如图 11 所示。每个类别的变电站缺陷的性能曲线比较清楚地说明了改进模型在不同类型缺陷的检测能力。

表 9 改进模型在不同变电站缺陷检测中的应用

Tab. 9 Application of improved model in defect detection of different substations

类型	Instances	P/%	R/%	mAP50/%	mAP50-95/%
全部	676	80.8	62.0	67.4	46.4
表计读数异常	83	81.6	84.3	87.1	70.1
表计外壳破损	50	91.8	72.0	79.7	65.7
异物鸟巢	73	69.1	52.0	53.4	24.6
表盘模糊	93	79.1	77.2	74.7	52.5
表盘破损	76	81.3	67.1	71.2	50.0
绝缘子破裂	43	83.6	35.6	44.4	28.5
地面油污	99	69.4	35.4	45.6	22.6
硅胶桶变色	98	97.7	85.2	92.8	75.1
空中漂浮物	61	85.6	42.6	56.1	27.7

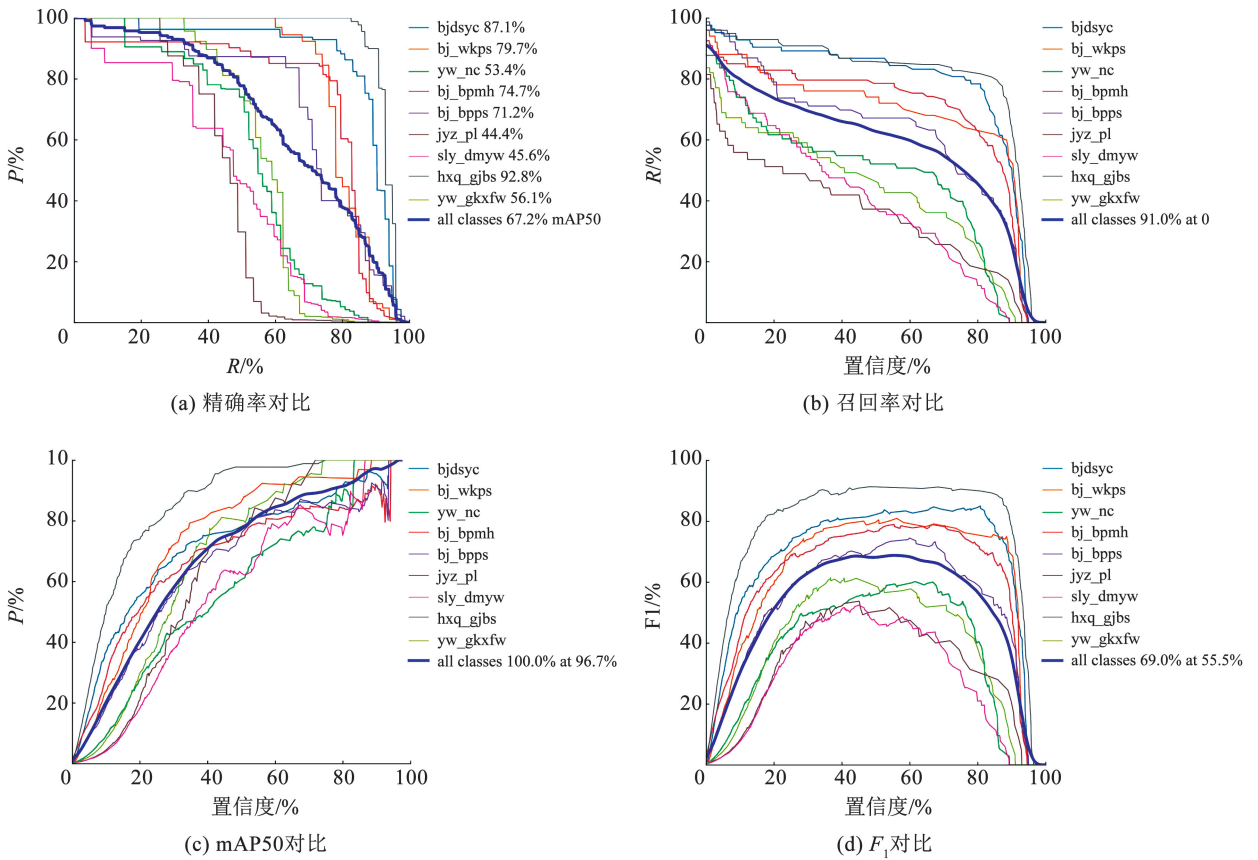


图 11 使用 ELT-RTDETR 模型在不同类型缺陷的精确率、召回率、mAP50 和 F_1 比较曲线

Fig. 11 Comparison of precision, recall, mAP50, and F_1 curves for different types of defects using ELT-RTDETR model

图 12 显示了 RT-DETR 和 ELT-RTDETR 模型在各自最佳性能下的混淆矩阵。通过可视化比较，

可以直观地观察到 ELT-RTDETR 模型在缺陷类别的分类性能上要优于原模型。

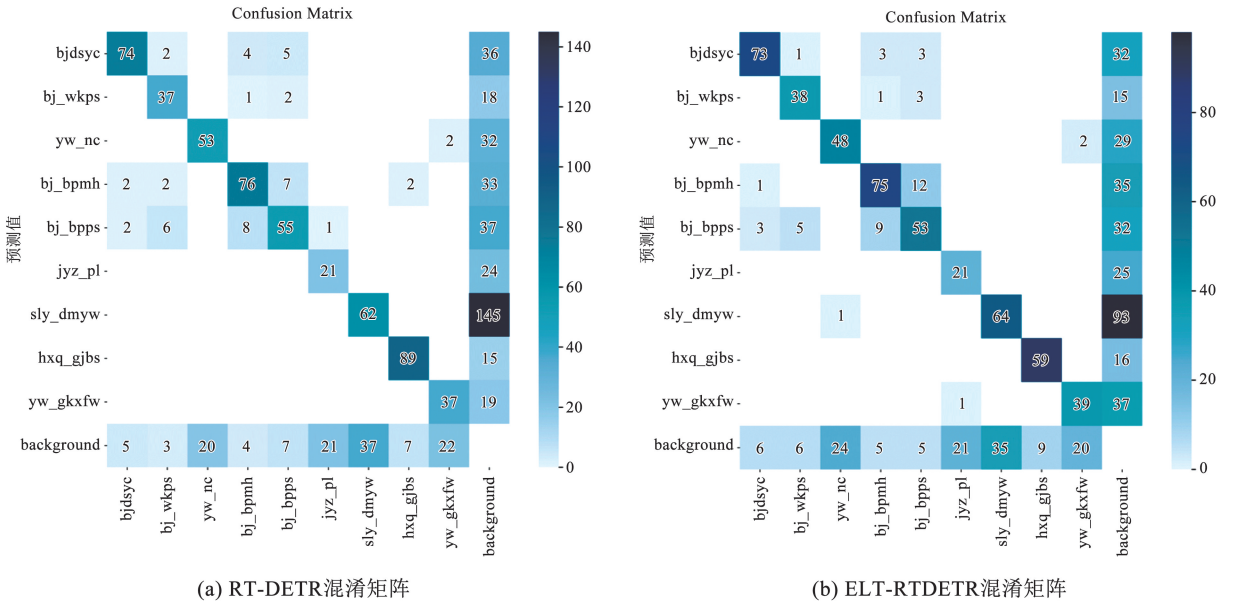


图 12 改进前、后模型混淆矩阵比较

Fig. 12 Comparison of model confusion matrix before and after improvement

2.7 泛化能力验证实验

为了验证 ELT-RTDETR 的泛化能力,本文在多

个公开数据集上进行了补充实验。首先,在 PASCAL VOC 数据集^[27]进行测试,该数据集包含了

与变电站设备结构相似的工业设备图像。其次,在工业缺陷检测公开数据集 NEU-DET(钢材表面缺陷)^[28]进行相关领域的迁移验证,该数据集由东北大学宋克臣团队制作,专注于钢材表面缺陷的检测与识别,包含了1 800张图片,涵盖了6种常见的钢

材表面缺陷类型。实验结果见表10。另外通过立体柱状图以及雷达图对性能指标进行可视化,以此能够更直观地展示改进模型在不同公共数据集上能力的提升,如图13所示。

表10 不同公共数据集结果对比

Tab.10 Comparison of results on different public datasets

模型	PASCAL VOC				NEU-DET			
	P/%	mAP50/%	计算量/(10 ⁹ s ⁻¹)	参数量/M	P/%	mAP50/%	计算量/(10 ⁹ s ⁻¹)	参数量/M
RT-DETR	67.4	56.9	57.0	75.9	79.4	74.6	57.0	75.8
ELT-RTDETR	70.0	60.5	21.0	37.2	83.9	78.8	21.0	37.1

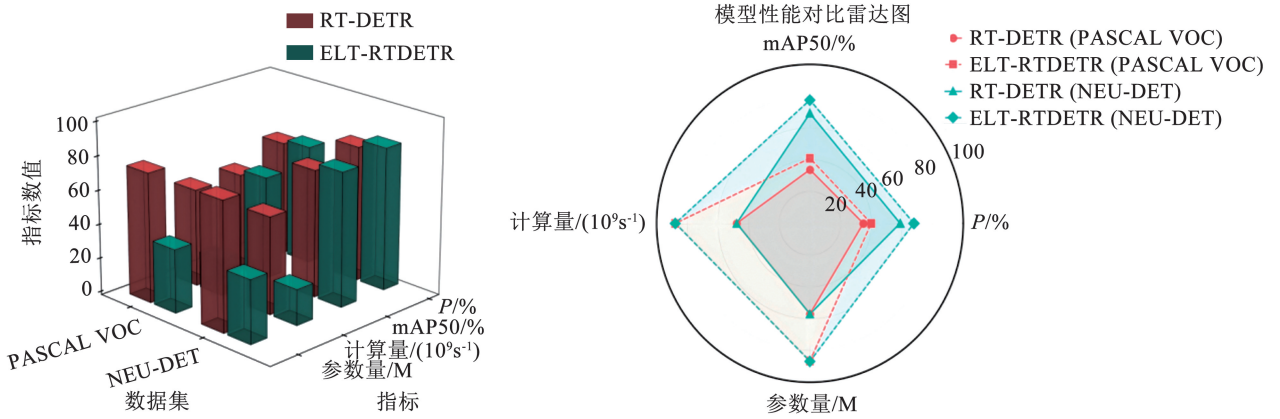


图13 性能指标可视化

Fig.13 Visualization of performance indicators

从表10可以看出,ELT-RTDETR相比RT-DETR在多个关键指标上都实现了显著的性能提升。在PASCAL VOC数据集上,精确率从67.4%提升至70.0%,mAP50从56.9%增长到60.5%;在NEU-DET工业缺陷检测数据集上表现更为突出,精确率从79.4%大幅提升至83.9%,mAP50也从74.6%增长到78.8%。特别值得注意的是,计算效率方面的巨大优化,ELT-RTDETR将计算量(GFLOPs)从57.0大幅削减至21.0,减少了约63.0%的计算量,同时模型参数量也从75.9 M压缩至37.2 M,实现了超过50%的模型体积缩减。这表明通过引入轻量化模块和优化的特征金字塔结构,模型能够更好地适应不同领域和场景下的目标特征,从而实现更强的泛化能力和检测性能提升。

另外,从跨域泛化能力来看,两个数据集代表了完全不同的应用场景:PASCAL VOC 考验模型对自

然场景中大幅度、易区分目标的检测能力,而NEU-DET则要求模型具备检测钢材表面细微、低对比度缺陷的精细化识别能力。两个模型在NEU-DET上的表现都明显优于PASCAL VOC,反映了工业缺陷检测任务目标类别相对简单但检测精度要求更高的特殊性。从图13中雷达图的可视化结果可以清晰地观察到,ELT-RTDETR在所有维度上都外扩于RT-DETR,特别是在计算效率相关的计算量和参数量维度上改进幅度最为显著,这种均衡的性能改进模式表明ELT-RTDETR在架构层面实现了根本性的优化,形成了效率与准确性的双重提升。

2.8 可视化分析

2.8.1 性能指标可视化

在变电站缺陷检测等特殊场景中,目标的尺度差异、类别复杂性、背景干扰等特点要求检测模型在特定任务上具有更好地表现,TIDE能够帮助更准确地捕捉这些任务特性^[29]。TIDE结果见表11。

表11 TIDE 结果可视化

Tab.11 TIDE result visualization

模型	E_{Cls}	E_{Miss}	E_{Bkg}	E_{Dupe}	E_{Both}	E_{Loc}	E_{FP}	E_{FN}
RT-DETR	4.11	0.97	1.97	1.41	0.68	4.07	22.24	6.63
ELT-RTDETR	4.04	0.97	2.46	0.95	0.33	4.72	20.87	6.90
Improvement	0.07 ↓	0	0.49 ↑	0.46 ↓	0.35 ↓	0.65 ↑	1.37 ↑	0.27 ↑

为了直观地展示改进模型能力的提升,本文可视化了几个评估指标,如图 14 所示。

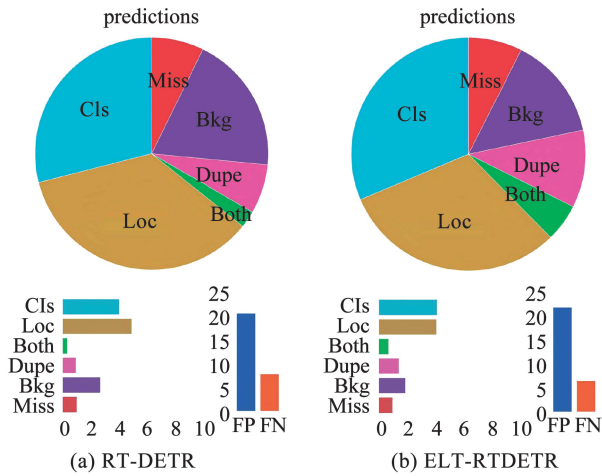


图 14 TIDE 结果可视化

Fig. 14 TIDE result visualization

根据表 11 和图 14 的 TIDE 结果可视化分析,

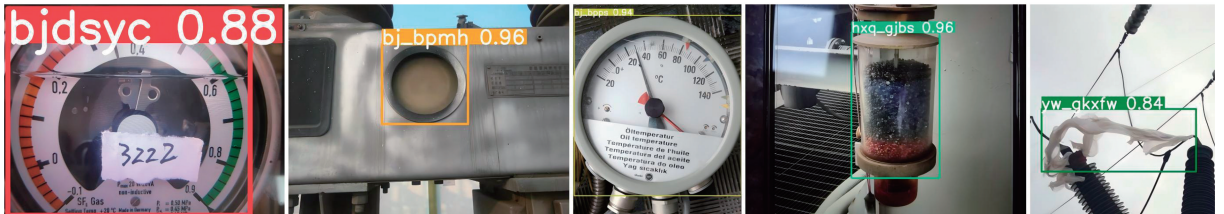
ELT-RTDETR 相较于 RT-DETR 在多个关键性能指标上有显著提升。具体来说,ELT-RTDETR 的分类错误率和漏检率分别下降 7% 和 35%,表现出更强的检测精度,尤其在小目标和复杂背景下的检测效果更佳。此外,重复误检和误检背景部分大幅降低,证明了其在多目标检测和去除冗余预测方面的优势。图 14 的可视化结果也进一步验证了这些改进,ELT-RTDETR 在定位精度和漏检控制方面显著优于 RT-DETR,整体提升了检测性能,展现出在变电站缺陷检测中的实际应用价值和工程可行性。

2.8.2 结果可视化分析

为了能更为直观地感受改进后的算法模型在实际对变电站缺陷检测应用中的效果,从数据集中选取了不同类型的变电站缺陷进行验证,结果如图 15 所示。图 15 中从左到右分别展示:表计读数异常、表盘模糊、表计外壳破损,以及硅胶桶变色与异物的对比效果。



(a) 原始RT-DETR检测结果



(b) ELT-RTDETR检测结果

图 15 不同方法的检测结果可视化对比

Fig. 15 Visual comparison of detection results from different methods

从图 15 的上、下两行对比结果可以看出,上行显示的 RT-DETR 检测结果与下行显示的 ELT-RTDETR 检测结果相比,ELT-RTDETR 相比于 RT-DETR 在变电站缺陷检测任务中的表现有了显著提升。具体来说,ELT-RTDETR 在各个目标检测的精度上均有显著改善,尤其是在表计读数异常和表盘破损这两个类别的检测效果上表现尤为突出。

3 结论

本文针对当前变电站缺陷检测研究多集中于特定场景下的单一缺陷识别,且现有模型普遍存在计算复杂度高、部署困难等问题,提出了 ELT-RTDETR 轻量化检测模型,通过理论分析、模型设计与实验

证,得出以下结论:

1)在模型架构优化方面,通过将 RT-DETR 主干网络替换为 EfficientFormerV2-S0,实现了参数量和计算量的显著降低,同时保持了良好的特征提取能力。引入 LMSFPN 轻量化多尺度特征金字塔网络,通过多尺度深度卷积、加权融合与高效上采样策略,在减少 63.2% 计算量和 50.7% 参数量的基础上,增强了多尺度缺陷特征的表达能力,所提出的 TSSACFE 模块有效提升了微小缺陷的检测鲁棒性。

2)在检测性能提升方面,模型在自建变电站设备缺陷数据集上的检测精度达到 82.1%,相比传统 RT-DETR 提升 7.3%,mAP50 指标达到 67.4%,mAP50-95 指标达到 46.4%。在表计外壳破损和硅

胶桶变色等特定缺陷检测任务中表现尤为突出,精确率分别达到91.8%和97.7%。

3)在工程应用价值方面,该模型实现了精度与效率的良好平衡,推理速度达到69.23 fps,满足实时检测需求,模型参数量大小仅为37.14 M,适合在资源受限的边缘设备上部署。消融实验和对比实验证明了各模块的有效性,为变电站智能运维系统的实际应用提供了高效解决方案,具有显著的工程应用潜力和推广价值。

参考文献

- [1] REN Shaoqing, HE Kaiming, GIRSHICK R, et al. Faster R-CNN: towards real-time object detection with region proposal networks[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017, 39(6): 1137. DOI:10.1109/TPAMI.2016.2577031
- [2] LIU Yunpeng, JI Xinxin, PEI Shaotong, et al. Research on automatic location and recognition of insulators in substation based on YOLOv3[J]. *High Voltage*, 2020, 5(1): 62. DOI:10.1049/hve.2019.0091
- [3] WANG Qian, YANG Lixin, ZHOU Bin, et al. YOLO-SS-large: a lightweight and high-performance model for defect detection in substations[J]. *Sensors*, 2023, 23(19): 8080. DOI:10.3390/s23198080
- [4] WANG Jianqiang, SUN Yiwei, LIN Ying, et al. Lightweight substation equipment defect detection algorithm for small targets[J]. *Sensors*, 2024, 24(18): 5914. DOI:10.3390/s24185914
- [5] FENG Qiaosheng, HUANG Li, SUN Ying, et al. Substation instrumentation target detection based on multi-scale feature fusion[J]. *Concurrency and Computation: Practice and Experience*, 2022, 34(23): e7177. DOI:10.1002/cpe.7177
- [6] 李刚, 张运涛, 汪文凯, 等. 采用DETR与先验知识融合的输电线路螺栓缺陷检测方法[J]. *图学学报*, 2023, 44(3): 438
- LI Gang, ZHANG Yuntao, WANG Wenkai, et al. Defect detection method of transmission line bolts based on DETR and prior knowledge fusion[J]. *Journal of Graphics*, 2023, 44(3): 438. DOI:10.11996/JG.j.2095-302X.2023030438
- [7] XI Yu, ZHOU Ke, MENG Lingwen, et al. Transmission line insulator defect detection based on swin transformer and context[J]. *Machine Intelligence Research*, 2023, 20(5): 729. DOI:10.1007/s11633-022-1355-y
- [8] ZHAO Yian, LV Wenyu, XU Shangliang, et al. DETRs beat YOLOs on real-time object detection[C]//2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle: IEEE, 2024: 16965. DOI:10.1109/CVPR52733.2024.01605
- [9] LI Yanyu, HU Ju, WEN Yang, et al. Rethinking vision transformers for MobileNet size and speed[C]//2023 IEEE/CVF International Conference on Computer Vision (ICCV). Paris: IEEE, 2024: 16843. DOI:10.1109/iccv51070.2023.01549
- [10] TAN Mingxing, PANG Ruoming, LE Q V. EfficientDet: scalable and efficient object detection[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle: IEEE, 2020: 10778. DOI:10.1109/CVPR42600.2020.01079
- [11] YANG Zhiqiang, GUAN Qiu, ZHAO Keer, et al. Multi-branch auxiliary fusion YOLO with re-parameterization heterogeneous convolutional for accurate object detection[C]//Pattern Recognition and Computer Vision (PRCV). Singapore: Springer, 2025: 492. DOI: 10.1007/978-981-97-8858-3_34
- [12] RAHMAN M M, MUNIR M, MARCULESCU R. EMCAD: efficient multi-scale convolutional attention decoding for medical image segmentation[C]//2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle: IEEE, 2024: 11769. DOI:10.1109/CVPR52733.2024.01118
- [13] LI Yanghao, CHEN Yuntao, WANG Naiyan, et al. Scale-aware trident networks for object detection[C]//2019 IEEE/CVF International Conference on Computer Vision (ICCV). Seoul: IEEE, 2019: 6053. DOI:10.1109/iccv.2019.00615
- [14] WU Ziyang, DING Tianjiao, LU Yifu, et al. Token statistics transformer: linear-time attention via variational rate reduction[PP/OL]. v1. arXiv (2024-12-23) [2025-03-15]. <https://doi.org/10.48550/arxiv.2412.17810>
- [15] WOO S, DEBNATH S, HU Ronghang, et al. ConvNeXt V2: Co-designing and scaling ConvNets with masked autoencoders[C]//2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition. (CVPR). Vancouver: IEEE, 2023: 16133. DOI: 10.1109/CVPR52729.2023.01548
- [16] CHEN Jierun, KAO S H, HE Hao, et al. Run, don't walk: chasing higher FLOPS for faster neural networks[C]//2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Vancouver: IEEE, 2023: 12021. DOI: 10.1109/CVPR52729.2023.01157
- [17] WANG Ao, CHEN Hui, LIN Zijia, et al. RepViT: revisiting mobile CNN from ViT perspective[C]//2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle: IEEE, 2024: 15909. DOI:10.1109/CVPR52733.2024.01506
- [18] CAI Han, LI Junyan, HU Muyan, et al. Efficientvit: lightweight multi-scale attention for high-resolution dense prediction[C]//2023 IEEE/CVF International Conference on Computer Vision (ICCV). Paris: IEEE, 2024: 17256. DOI:10.1109/ICCV51070.2023.01587
- [19] CHEN Yifei, ZHANG Chenyan, CHEN Ben, et al. Accurate leukocyte detection based on deformable-DETR and multi-level feature fusion for aiding diagnosis of blood diseases[J]. *Computers in Biology and Medicine*, 2024, 170: 107917. DOI:10.1016/j.combiomed.2024.107917
- [20] HOU Qibin, ZHOU Daquan, FENG Jiashi. Coordinate attention for efficient mobile network design[C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Nashville: IEEE, 2021: 13708. DOI:10.1109/CVPR46437.2021.01350
- [21] XU Wei, WAN Yi. ELA: Efficient local attention for deep convolutional neural networks[PP/OL]. v1. arXiv (2024-03-02) [2025-03-15]. <https://doi.org/10.48550/arxiv.2403.01123>
- [22] SUN Shangquan, REN Wenqi, GAO Xinwei, et al. Restoring images in adverse weather conditions via histogram transformer[C]//Computer Vision-ECCV 2024. Cham: Springer, 2025: 111. DOI: 10.1007/978-3-031-72670-5_7
- [23] LI Junyi, ZHANG Zhilu, ZUO Wangmeng. Rethinking transformer-based blind-spot network for self-supervised image denoising[J]. *Proceedings of the AAAI Conference on Artificial Intelligence*, 2025, 39(5): 4788. DOI:10.1609/aaai.v39i5.32506
- [24] SHAKER A, MAAZ M, RASHEED H, et al. SwiftFormer: efficient additive attention for transformer-based real-time mobile vision applications[C]//2023 IEEE/CVF International Conference on Computer Vision (ICCV). Paris: IEEE, 2024: 17379. DOI: 10.1109/ICCV51070.2023.01598
- [25] PAN Zizheng, CAI Jianfei, ZHUANG Bohan. Fast vision transformers with HiLo attention[PP/OL]. v1. arXiv (2022-01-01) [2025-03-15]. <https://doi.org/10.48850/arxiv.2205.13213>
- [26] GUO Jialong, CHEN Xinghao, TANG Yehui, et al. SLAB: Efficient transformers with simplified linear attention and progressive re-parameterized batch normalization[PP/OL]. v2. arXiv (2024-05-19) [2025-03-15]. <https://doi.org/10.48850/arxiv.2405.11582>
- [27] EVERINGHAM M, ALI ESLAMI S M, VAN GOOL L, et al. The pascal visual object classes challenge: a retrospective[J]. *International Journal of Computer Vision*, 2015, 111(1): 98. DOI:10.1007/s11263-014-0733-5
- [28] HE Yu, SONG Kechen, MENG Qinggang, et al. An end-to-end steel surface defect detection approach via fusing multiple hierarchical features[J]. *Transactions on Instrumentation and Measurement*, 2020, 69(4): 1493. DOI:10.1109/TIM.2019.2915404
- [29] BOLYA D, FOLEY S, HAYS J, et al. TIDE: a general toolbox for identifying object detection errors[M]//Computer vision-2020. Cham: Springer International Publishing, 2020: 558. DOI: 10.1007/978-3-030-58580-8_33