文章编号:1007-2780(2026)08-1175-12

基于改进RT-DETR的光伏红外缺陷检测算法

邓万宇，彭鑫亮*

（西安邮电大学 计算机学院，陕西 西安 710121）

摘要:光伏板表面缺陷检测是保障光伏系统发电效率与运行安全的关键。针对无人机航拍红外图像中航拍小目标检测难、户外热背景干扰强及缺陷目标尺度跨度大等挑战，本研究提出一种结合多尺度特征与动态卷积的光伏红外缺陷检测模型。首先，在浅层路径采用空间到深度卷积(SPD-Conv)替代传统下采样，规避二极管异常、热斑等微小目标的细粒度特征与边缘信息丢失；其次，引入动态全维卷积(CSPOmniKernel)模块，通过多分支机制自适应调整感受野，增强低对比度缺陷特征并抑制复杂热背景干扰；最后，构建高效多尺度特征融合模块(MFM)，借助跨层级信息交互与自适应聚合，提升模型对不同尺度目标的感知鲁棒性。在包含5类典型缺陷的光伏红外航拍数据集上，改进模型的mAP@0.5达到84.12%，较基准RT-DETR-r18提升5.16%，且推理速度达48 FPS，满足工业实时巡检需求。结果表明，该方案在复杂户外场景下能有效提升光伏缺陷检测的精度与可靠性。

关　键　词:光伏缺陷检测；航拍红外图像；RT-DETR算法；动态卷积；小目标检测

中图分类号:TP391.4　　**文献标识码**:A　　**doi**:10.3724/CJLCD.2026-0079　　**CSTR**:32172.14.CJLCD.2026-0079

Photovoltaic infrared defect detection algorithm based on improved RT-DETR

DENG Wanyu, PENG Xinliang*

(*School of Computer Science and Technology, Xi'an University of Posts and Telecommunications, Xi'an* 710121, *China*)

Abstract: Surface defect detection of photovoltaic panels is of paramount importance for ensuring the power generation efficiency and operational safety of photovoltaic systems. To tackle challenges such as the difficulty in detecting small aerial targets in infrared images captured by drones, significant interference from outdoor thermal backgrounds, and the broad scale range of defect targets, this study introduces a photovoltaic infrared defect detection model that integrates multi-scale features with dynamic convolution. First, in the shallow pathway, spatial-to-depthwise convolution(SPD-Conv) is employed to substitute traditional downsampling methods, thereby preventing the loss of fine-grained features and edge information pertaining to small targets, including diode anomalies and hot spots. Second, a dynamic full-dimensional convolution(CSPOmniKernel) module is incorporated, which adaptively modulates the receptive field through

收稿日期:2026-05-11；**修订日期**:2026-06-24.

基金项目:陕西省科学家工程师创新团队项目（No. 2023KXJ-091）；秦创原总窗口“四链融合”计划重点项目（No. 2025PT-ZCK-67）

Supported by Shaanxi Scientist & Engineer Innovation Team (No. 2023KXJ-091); Key Project of Qinchuangyuan General Window “Four-Chain Integration” Program (No. 2025PT-ZCK-67)

*通信联系人，E-mail:1802368981@stu.xupt.edu.cn

a multi-branch mechanism, thereby enhancing low-contrast defect features while mitigating interference from complex thermal backgrounds. Finally, an efficient multi-scale feature fusion module(MFM) is devised, leveraging cross-level information interaction and adaptive aggregation to bolster the model's robustness in perceiving targets of varying scales. On a photovoltaic infrared aerial dataset encompassing five typical defect types, the enhanced model attains an mAP@0.5 of 84.12%, marking a 5.16% improvement over the baseline RT-DETR-r18, and achieves an inference speed of 48 FPS, fulfilling the demands of industrial real-time inspection. The results demonstrate that this approach significantly elevates the accuracy and reliability of photovoltaic defect detection in intricate outdoor environments.

Key words: photovoltaic defect detection; aerial infrared image; RT-DETR algorithm; dynamic convolution; small target detection

1 引 言

光伏发电作为清洁能源的重要组成部分,其组件的质量监控是保障电站发电效率与系统运行安全的关键环节。在长期的户外运行过程中,受环境老化、机械应力及热应力等因素影响,光伏板不可避免地会产生玻璃破损、局部热斑、电势诱导衰减(Potential Induced Degradation,PID)效应、二极管失效以及表面遮挡等多种缺陷[1]。陈天祥等人[2]的研究指出,光伏缺陷形态的多样性对自动检测算法的泛化性提出了严苛要求。传统的人工巡检方式主要依赖运维人员携带红外热像仪进行地面手工检测[3],不仅效率低下,且在面对大规模山地或屋顶电站时,受视觉疲劳与视角受限影响,漏检率居高不下[4],难以适配现代光伏电站对智能化运维的需求。因此,利用无人机搭载红外相机进行自动化实时检测已成为行业转型的迫切需求[5-7]。

早期的缺陷检测算法多基于红外图像的阈值分割、频谱分析或人工设计的纹理特征[8]。这类方法在环境单一的情况下表现尚可,但在面对光照变化引起的表面反光、环境温度波动造成的低对比度背景时,算法的稳健性显著不足。随着深度学习的兴起,以Faster R-CNN[9]和YOLO系列[10]为代表的卷积神经网络(CNN)逐渐成为主流,并在光伏缺陷检测任务中得到广泛应用[11-12]。然而,CNN固有的局部感受野限制了其对长距离依赖关系的建模能力[13],导致在处理红外图像中大面积的PID效应或极微小的热斑缺陷时,常因全局上下文信息缺失而发生误判。

为突破CNN的局限性,DETR(Detection Transformer)引入了自注意力机制[14],实现了对图像全局内容的端到端建模。随后的RT-DETR通过优化编码器与混合查询选择策略[15-16],在大幅提升推理速度的同时保持了Transformer的高精度特性,展现出在实时巡检领域的巨大潜力[17-18]。然而,将RT-DETR直接迁移至航拍红外光伏巡检场景仍面临实质性障碍:其一,无人机航拍视角下,二极管异常等缺陷呈现极小尺寸特征,标准模型的连续下采样操作容易导致这类细粒度信息的特征湮灭[19-20];其二,红外热成像反映的是物体的温度与热辐射分布,在户外巡检场景下,被太阳辐射加热的屋顶、地面、墙面,以及玻璃盖板的镜面反射、正常发电组件自身的整体温升等,均会在红外图像中形成与热斑缺陷相近的高亮区域,加之缺陷目标与背景的温度对比度本就极低,致使常规注意力机制难以在如此复杂的热背景中精准聚焦于真实缺陷[21-22];其三,航拍图像中不同缺陷的尺度差异极大[23],从占据数个像素的热斑到贯穿整个组件阵列的遮挡,标准Transformer在捕捉跨度如此之大的多尺度特征时,往往缺乏足够的空间感知灵敏度[24]。

针对上述挑战,本研究在RT-DETR基础上提出一种融合多尺度特征与动态卷积的光伏检测模型。首先,利用SPD-Conv改进下采样机制,通过将空间信息映射至通道维度,有效规避了细粒度特征湮灭,提升了航拍微小缺陷的检出率;其次,构建动态全维卷积(CSPOmniKernel)模块,利用多维注意力动态校准权重,实现复杂背景下热目标的自适应提取与噪声抑制;最后,设计高效多尺度特征融合模块(MFM)进行跨层级语义对齐,强化了多尺度表征并降低了尺度敏感性。

本研究通过特征保真、动态感知与多尺度融合的协同优化，显著提升了复杂工业场景下光伏检测的精度与鲁棒性。

2　相关理论与模型改进

2.1　RT-DETR 基准算法原理

RT-DETR 是首个实时端到端目标检测器，其整体架构如图 1 所示。

该模型主要由骨干网络（Backbone）、高效混合编码器（Efficient Hybrid Encoder）以及 Transformer 解码器组成。图像输入后，首先经由骨干网络提取出 3 个不同尺度的特征图{S_3、S_4、S_5}。随后，特征进入混合编码器：其中基于注意力的尺度内特征交互（AIFI）模块仅对包含丰富语义的高层特征 S_5 进行自注意力运算，以低计算成本捕捉全局上下文；而跨尺度跨特征融合（CCFF）模块则负责将多尺度特征进行融合。随后使用 IoU 感知从编码器输出序列中选择固定数量的图像特征作为解码器的初始对象查询。最后，带有辅助预测头的解码器迭代优化对象查询以生成方框和置信度分数。

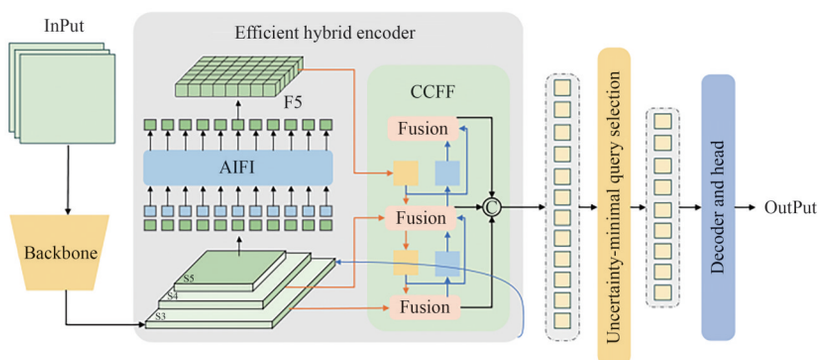


图 1　RT-DETR 模型结构图

Fig. 1　RT-DETR model architecture diagram

2.2　MFK-DETR 模型整体架构

针对无人机巡检光伏板过程中存在的“航拍小目标检测难”、“户外环境热背景干扰强”以及“缺陷目标尺度跨度大”3 大难题，本研究提出一种基于改进 RT-DETR 的光伏缺陷检测模型（MFK-DETR）。该模型以 RT-DETR 为基础架构，通过重构特征提取路径、引入动态感知机制并优化多尺度融合策略，实现了对复杂红外环境下光伏缺陷的高效感知。模型主要由骨干网络（Backbone）、高效混合编码器（Efficient Hybrid Encoder）以及 Transformer 解码器（Decoder）3 部分组成。本研究通过在高效混合编码器中嵌入 SPD-Conv 以保留微小缺陷特征，并引入 CSPOmniKernel 动态全维卷积模块增强模型对低对比度异常目标的辨别力，如图 2 所示。

改进模型的处理流程首先从骨干网络开始，输入的红外图像经过一系列卷积层（ConvNorm）、最大池化层（MaxPool）及基础残差块（BasicBlock）完成特征提取。为了克服传统下采样造成的微小缺陷信息丢失，在骨干网络的浅层分支（即第一个 BasicBlock 输出处）引入了空间到深度卷积（Space-to-Depth Convolution，SPD-Conv）模块。该分支不经过传统的步长卷积，而是直接将高分辨率特征图的空间维度映射至深度维度，从而完整保留了二极管异常、微小热斑等细小目标的特征信息与边缘轮廓，并将其直接注入到后续的特征融合网络中，有效解决了航拍视角下极小目标难以被深层网络捕获的问题。

随后，提取的多尺度特征进入增强型混合编码器（Enhanced Hybrid Encoder），这是模型的核心改进部分。深层特征首先通过 AIFI（Intra-scale Feature Interaction）模块，利用自注意力机制捕获图像的全局上下文信息，为大面积的遮挡（Covering）或 PID 效应等显著热异常区域建立语义基础。在跨尺度融合方面，级联的多尺度特征融合模块（Multi-scale Feature Fusion Module，MFM）替代了原有的融合节点，通过跨层级连接将 SPD-Conv 提取的浅层细节与 AIFI 提取的深层语义进行动

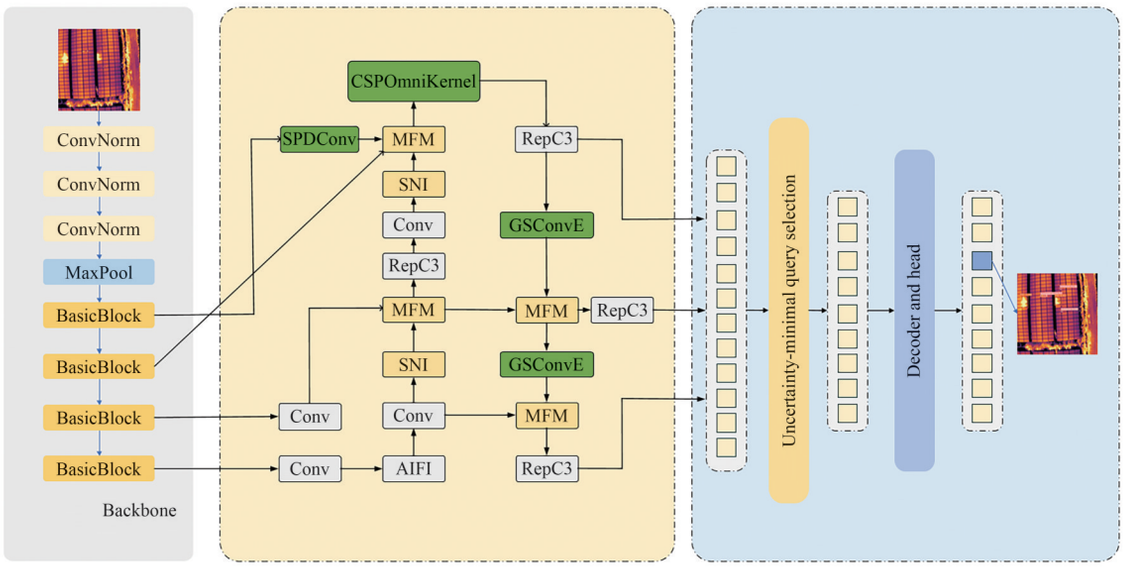


图 2 MFK-DETR 模型结构图

Fig. 2 MFK-DETR model architecture diagram

态聚合。为了保证特征传输的高效性与稳定性，在不同尺度的 MFM 之间采用 GSConvE（Ghost Shuffle Convolution Enhanced）模块进行高效下采样，并利用 SNI（Scale-Normalized Interaction）模块对特征进行归一化与交互增强。

此外，在编码器的顶层路径中嵌入了 CSPOmniKernel 动态卷积模块。该模块接收来自 MFM 的高层融合特征，利用全维动态卷积核自适应地调整感受野和通道权重，专门用于在深层特征空间中剥离复杂的户外环境热背景干扰，从而增强模型对低对比度或模糊边缘缺陷目标的辨识度。经过各层级处理后的特征，最终通过 RepC3（Re-parameterized Convolution with 3 Branches）结构进行重参数化整理，输出 3 个针对不同尺度缺陷设计的特征图。

最后，融合后的多尺度特征图被送入“最小不确定性查询选择”（Uncertainty-minimal Query Selection）模块。该模块优选出置信度最高且位置不确定性最低的特征点作为初始对象查询（Object Queries），随后送入 Transformer 解码器（Decoder and Head）。模型利用二分图匹配机制（Bipartite Matching）计算预测框与真实框的损失，直接预测出光伏组件缺陷的类别（如热斑、PID、遮挡等）与精确边界框。

2.3 空间深度转换（SPD-Conv）模块

在传统的卷积神经网络中，通常使用步长为 2 的卷积层或池化层进行下采样，以增大感受野并降低计算量。然而，在无人机航拍红外巡检场景下，由于拍摄距离较远，图像中如二极管异常、微小热斑等缺陷的像素占比极小。这种粗粒度的下采样操作往往导致“细粒度特征丢失”问题，使得这些关键的小目标特征在深层网络中逐渐模糊甚至完全消失，严重影响了模型的召回率。

为解决这一问题，本研究引入 SPD-Conv 模块。SPD-Conv 完全摒弃了传统的步长卷积和池化操作，由空间到深度转换（SPD）层和非步长卷积层共同组成，如图 3 所示。

其核心原理是将特征图在空间维度（高和宽）上的信息切片并堆叠到通道维度上。假设输入特征图 X 的大小为 $S\times S\times C_1$，其中 S 为空间尺寸，C_1 为通道数。SPD 层按照缩放因子 scale = 2 对 X 进行切片采样，生成 4 个子特征图，其索引过程可表示为：

$$\begin{cases} f_{0,0}=X[0:S:2,0:S:2] \\ f_{1,0}=X[1:S:2,0:S:2] \\ f_{0,1}=X[0:S:2,1:S:2] \\ f_{1,1}=X[1:S:2,1:S:2] \end{cases} \tag{1}$$

随后，这 4 个子特征图在通道维度上进行拼接（Concatenation），得到中间特征图 X'，其维度变为 $\frac{S}{2}\times\frac{S}{2}\times 4C_1$。这一过程在降低空间分辨率的同时，保留了全部的像素信息，没有任何信

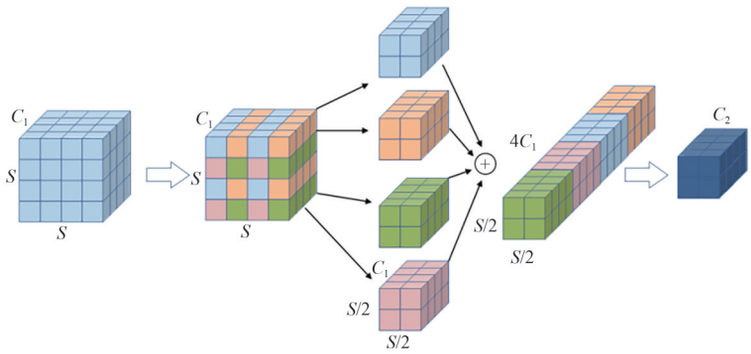


图 3　空间到深度卷积

Fig. 3　Space-to-depth convolution

息损失。最后，为了保持通道数的恒定并融合信息，通过一个标准的 1 × 1 或 3 × 3 卷积层将通道数从 $4C_1$ 调整为目标维度 C_2。在 MFK-DETR 中，我们在骨干网络的低层级阶段应用 SPD-Conv，有效保留了红外热异常特征，为后续特征融合提供了丰富的信息流。

2.4　动态全维卷积模块

为了有效应对红外光伏缺陷形态的多样性，以及复杂户外环境（如建筑物反光、地面植被热辐射等）产生的强背景干扰，本文提出了动态全维卷积（CSPOmniKernel）模块。区别于传统的静态卷积或仅在核空间维度进行动态调整的卷积算子，CSPOmniKernel 创新性地构建了“局部-上下文-全局”3 分支并行处理架构，并将频域注意力机制引入特征提取过程，以实现对低对比度异常目标的动态感知。输入红外特征首先经过卷积层进行降维映射，随即被分流至 3 个并行的处理路径中，通过多维度权重校准显著增强模型在复杂热背景下剥离缺陷特征的能力，如图 4 所示。

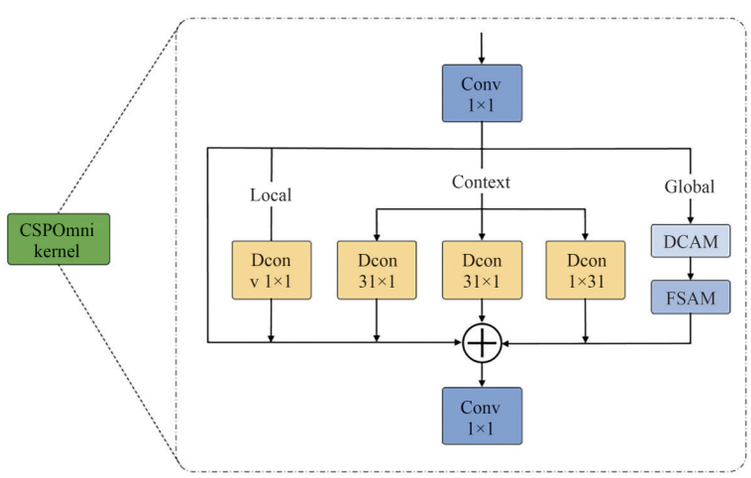


图 4　动态全维卷积模块

Fig. 4　CSPOmniKernel module

在空间特征提取层面，局部特征分支（Local Branch）与上下文特征分支（Context Branch）协同工作以适应不同尺度的缺陷。其中，局部特征分支利用动态卷积（DConv）专注于捕捉像素级的细微特征，这对于识别红外图像中微小的二极管异常或极小点状热斑至关重要。与此同时，上下文特征分支针对光伏缺陷几何形态变化剧烈的特点（如细长的玻璃破损裂纹、不规则的表面遮挡或大面积 PID 效应），设计了多组多尺度、非对称的动态卷积核（例如 3 × 1、1 × 3 等）。这种非对称卷积核组合策略使得模型能够根据缺陷的物理形态自适应地调整有效感受野，从而在复杂

的户外背景中精准界定缺陷目标的几何边界。

为了进一步抑制户外复杂热背景产生的噪声干扰并增强全局感知能力，全局特征分支（Global Branch）引入了基于频域分析的注意力机制，该机制由串联的双域上下文注意力模块（Dual-Domain Context Attention Module，DCAM）和频率选择注意力模块（Frequency-Selection Attention Module，FSAM）构成。

如图 5 所示，DCAM 旨在通过构建频域与空域的交互通道来增强特征的上下文表达。在无人机航拍红外巡检中，缺陷目标（如热斑）与复杂的背景（如屋顶、地面）往往存在不同的频率分布特征。输入特征分流后，一路经由快速傅里叶变换（FFT）转换至频域，与保留在空域的特征进行跨域交互融合，最后通过快速傅里叶逆变换（IFFT）还原。这一过程利用了频域的全局感受野特性，能够更有效地聚合图像的全局上下文信息，从而在复杂的背景辐射中精准锁定潜在的热异常区域。

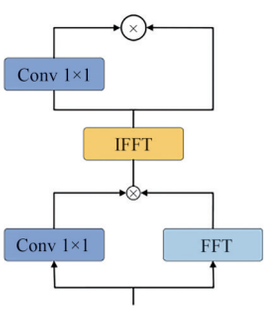


图 5 双域上下文注意力模块

Fig. 5 Dual-domain context attention module

随后如图 6 所示，FSAM 模块进一步利用频域信息生成注意力掩码。在光伏红外图像中，由于缺陷区域（如玻璃破损边缘或热斑中心）往往伴随剧烈的温度梯度变化，在频域上表现为显著的高频特征；而背景（如平展的光伏面板表面）则多表现为平滑的低频纹理。该模块通过并行的空域路径和频域变换路径提取特征，将两者融合后生成空间注意力图，并以此对主特征流进行加权。这一过程实现了对特定频率成分（即缺陷目标的温度突变边缘）的定向增强，以及对背景低频热噪声的有效抑制。

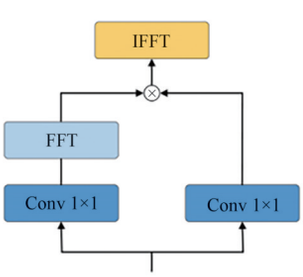


图 6 频率选择注意力模块

Fig. 6 Frequency-selection attention module

最终，上述 3 个分支提取的特征经过逐元素相加（Element-wise Summation）策略进行自适应融合，并通过末端的卷积层输出，形成了对光伏缺陷具有高度辨识力和鲁棒性的特征表达。

2.5 多尺度特征融合模块

针对无人机巡检视角下光伏缺陷尺度跨度大（如微小热斑与大面积遮挡并存）以及户外热背景复杂的特性，现有 RT-DETR 所采用的简单求和融合策略难以兼顾深层全局语义与浅层细粒度细节。为此，本文设计了高效多尺度特征融合模块（MFM），其结构如图 7 所示。该模块通过引入跨层级信息交互机制，旨在实现不同尺度特征的自适应聚合，从而提升模型在复杂巡检环境下对多尺度缺陷目标的捕捉能力。

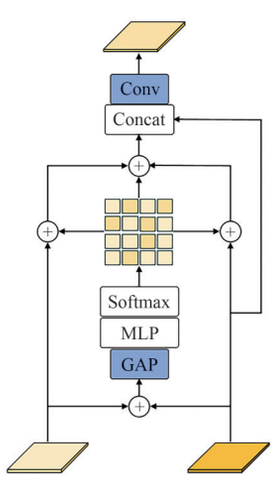


图 7 多尺度特征融合模块

Fig. 7 Multi-scale feature fusion module

如图 7 所示，MFM 采用“注意力引导的自适应聚合”策略。首先，设浅层几何特征为 $F_{low}\in R^{C\times H\times W}$，深层语义特征为 $F_{high}\in R^{C\times H\times W}$，两者经上采样对齐后逐元素相加得到融合特征。

$$F=F_{low}+F_{high},F\in R^{C\times H\times W}. \tag{2}$$

为解决无人机巡检中复杂的背景热辐射干

扰，模块核心引入了通道注意力机制，如图 7 中 GAP-MLP-Softmax 路径所示：首先通过全局平均池化（GAP）将每个通道压缩为单一描述值，得到通道描述向量 $z\in R^C$，其第 c 个分量为：

$$z_c=\frac{1}{H\times W}\sum_{i=1}^{H}\sum_{j=1}^{W}F_c(i,j). \tag{3}$$

随后由多层感知机（MLP）与 Softmax 函数生成各尺度分支的动态通道权重：

$$w=\mathrm{Softmax}\left(W_2\delta(W_1z)\right),$$
$$W_1\in R^{\frac{C}{r}\times C},\ W_2\in R^{C\times\frac{C}{r}}, \tag{4}$$

式中：$\delta(\cdot)$ 为 ReLU 激活，其定义为 $\delta(x)=\max(0,x)$，用于为通道权重的生成引入非线性；$w\in R^C$ 为归一化后的通道权重。该权重通过广播机制分别作用于不同尺度的输入分支，对特征通道进行重加权：

$$\hat{F}_{\mathrm{low}}=w\odot F_{\mathrm{low}},\ \hat{F}_{\mathrm{high}}=w\odot F_{\mathrm{high}}, \tag{5}$$

式中，$\odot$ 表示沿通道维度的逐通道乘法。最后采用拼接结合残差连接融合并经 1×1 卷积输出。

在计算效率方面，MFM 通过将特征交互约束于通道维度，实现了融合质量与计算开销的平衡。由式（5）可知，全局平均池化使注意力权重的生成仅依赖于通道向量 $z\in R^C$，与特征图的空间分辨率 $H\times W$ 完全解耦，权重生成的计算复杂度为 $O\left(\frac{2C^2}{r}\right)$（$r$ 为降维比），从而避免了像素级注意力中随分辨率平方增长 $\left(O\left(H^2W^2C\right)\right)$ 的高昂开销；而跨尺度特征的聚合仅通过逐通道乘法 $w\odot F$ 实现，无需大核卷积或多分支密集运算。因此，MFM 在几乎不增加参数量与计算量的前提下完成了跨尺度信息的自适应融合，既强化了对多尺度目标的感知能力，又缓解了二极管、热斑等小目标在深层特征传递中的信息丢失问题。

3　实验设计与结果分析

3.1　实验环境配置

实验在 Windows 平台下进行，硬件环境为 AMD Ryzen5 5600 处理器、16 GB 内存及 NVIDIA GeForce RTX 3080Ti 显卡。模型基于 Python3.11 和 PyTorch2.0.1 框架实现，并利用 CUDA11.8 进行加速计算。所有模型的输入图像分辨率均统一为 640×640，训练与测试阶段保持一致。训练共进行 200 个 Epoch，批量大小（Batch Size）设为 8。

3.2　评价指标定义

为了客观评估改进后的 RT-DETR 模型与原模型及其他模型的性能，在同等实验条件下对缺陷检测结果进行对比。本研究选取精确率（Precision）、召回率（Recall）、平均精度均值（mAP@0.5）、每秒检测帧数（FPS）、参数量（Params）和浮点运算量（GFLOPs）作为评价指标。其中，精确率、召回率与平均精度均值的计算公式如式（6）～（9）所示：

$$\mathrm{Precision}=\frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FP}}, \tag{6}$$

$$\mathrm{Recall}=\frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FN}}, \tag{7}$$

$$\mathrm{AP}=\int_0^1 P(R)\mathrm{d}R, \tag{8}$$

$$\mathrm{mAP}=\frac{1}{N}\sum_{i=1}^{N}\mathrm{AP}_i, \tag{9}$$

式中：TP 为正确检出的缺陷数量，FP 为将背景误判为缺陷的数量（误检），FN 为未能检出的真实缺陷数量（漏检），$P(R)$ 为精确率 P 作为召回率 R 的函数，AP_i 为第 i 个类别的平均精度，N 为总类别数。本实验中 mAP 主要关注 IoU 阈值为 0.5 时的平均精度均值（mAP@0.5）。

3.3　数据集构建与预处理

本研究通过无人机搭载红外热成像相机，在多种时段（早、中、晚）及飞行高度（15～50 m）下实地采集并构建了光伏红外巡检数据集。经筛选清洗，最终保留 1 430 张高质量图像，涵盖玻璃破损、热斑、PID 效应、二极管异常及表面遮挡 5 类典型缺陷，并采用 LabelImg 进行精准标注。实验按照 8∶1∶1 的比例随机划分训练集、验证集与测试集，以确保模型在复杂户外巡检场景下的泛化能力与鲁棒性。

3.4　数据集实验结果

3.4.1　消融实验

为验证各模块的有效性及其协同作用，本文采用单模块验证与模块组合相结合的方式设计消融实验（表 1）：A、B、C 三组分别在基线模型上单独嵌入 SPD-Conv、CSPOmniKernel、MFM，用于考察各模块的独立贡献；D、E 两组在此基础上进行模块组合，其中 D 组同时嵌入 SPD-Conv 与 CSPOmniKernel，E 组进一步融合 MFM 构成

表1 消融实验结果

Tab. 1 Ablation experiment result

组别	SPD-Conv	CSPOmniKernel	MFM	mAP@0.5/%	*P*/%	*R*/%	Params/M	GFLOPs/G
Baseline				78.96	76.92	73.12	20.09	60.25
A	√			82.30	80.21	76.16	20.80	62.14
B		√		81.67	78.89	79.90	20.58	61.35
C			√	82.83	80.24	**81.36**	**20.02**	**59.86**
D	√	√		83.10	79.19	77.52	20.40	63.28
E	√	√	√	**84.12**	**82.05**	78.79	20.62	64.70

注:粗体数字表示最优值

完整模型,用于验证模块间的协同增益。

单模块方面,引入SPD-Conv(A组)后mAP@0.5提升3.34%,有效缓解了下采样过程中的细粒度信息丢失;引入CSPOmniKernel(B组)使召回率提升6.78%,显著增强了对低对比度、弱特征缺陷的感知能力;引入MFM(C组)在3个单模块中综合表现最优,且在提升精度的同时参数量与计算量均低于基准网络,实现了精度与效率的兼顾。

在模块组合方面,SPD-Conv与CSPOmniKernel联合使用(D组)后性能进一步提升;3模块融合的完整模型(E组)取得最优结果,mAP@0.5较基准提升5.16%,而参数量与计算量相比基准仅分别增加约0.53M与4.45G。结果表明,各模块间具有显著的协同增强作用,能够在大幅提升检测精度的同时保持较低的计算开销,可满足光伏电站户外巡检的实际部署需求。

3.4.2 对比实验

为验证改进算法的先进性,将MFK-DETR与当前主流的目标检测模型(包括YOLO系列,RT-DETR系列)在测试集上进行同条件对比。

从表2可以看出,改进模型在mAP@0.5和召回率指标上均取得了最优数值。与基准模型RT-DETR-r18相比,所提模型的mAP@0.5提升了5.16%,召回率提高了5.67%。即便与参数量更大的RT-DETR-r34相比,mAP@0.5仍提升了2.7%。在实时性方面,所提模型的检测速度达到48 FPS,高于工业实时检测普遍要求的30 FPS门槛,可满足光伏电站红外巡检的实时处理需求。实验结果表明,该算法在保持均衡检测精度的同时,显著改善了红外图像中微小目标的漏检情况,证明了所提改进方案在处理低对比度、尺度跨度大的光伏缺陷时的有效性,能够更好地满足工业实时巡检对于高精度与高可靠性的双重需求。

表2 不同模型的对比实验结果

Tab. 2 Comparative experiment results of different models

Models	mAP@0.5/%	*P*/%	*R*/%	FPS
YOLOv5s	72.45	78.12	65.34	102
YOLOv8s	76.88	79.45	71.20	86
YOLOv11n	75.12	81.33	68.56	**105**
YOLOv12n	80.55	82.70	75.42	82
RT-DETR-r18	78.96	76.92	73.12	52
RT-DETR-r34	81.42	82.15	76.88	34
RT-DETR-C2f-Mamba	80.12	83.45	74.56	40
RT-DETR-DCNV4	82.35	**84.78**	77.10	37
Ours	**84.12**	82.05	**78.79**	48

注:粗体数字表示最优值

3.4.3 检测结果可视化

为直观验证改进MFK-DETR模型在复杂户外巡检场景下的实际表现,本研究从验证集中选取5组具有代表性的光伏红外图像进行可

视化对比分析。将原图（Original images）、基准模型 RT-DETR-r18（Baseline）预测结果与本文改进模型（Ours）预测结果进行横向对比，如图 8 所示。

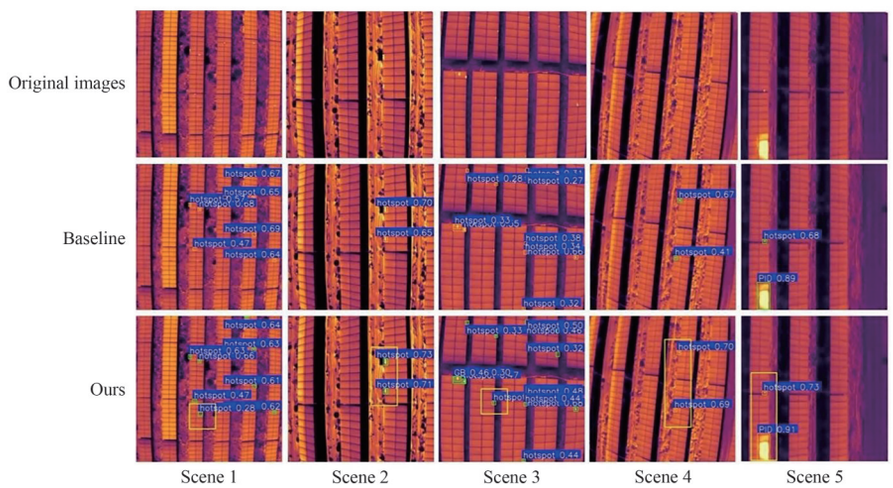


图 8　检测结果对比图

Fig. 8　Comparison of test results

整体来看，本文模型的检测结果与缺陷实际分布更为吻合。在 Scene1、Scene3 等缺陷较密集的场景中，基准模型存在部分目标漏检、定位框偏移的现象，而本文模型检出目标更为完整，预测框与缺陷区域贴合更紧致。在 Scene2、Scene4 等热背景复杂的场景中，基准模型易受背景高温区域干扰，出现框定位松散的情况，本文模型则能更准确地框定缺陷边界，有效抑制了热背景噪声的干扰。综合各场景可见，本文模型在小目标检出完整性与复杂热背景下的定位准确性上均优于基准模型，在多尺度与模糊细节挑战下表现出更强的鲁棒性。

为从定量角度评估改进模型对各类缺陷的检测性能，本文绘制了基准模型与改进模型在各缺陷类别上的精确率-召回率（Precision-Recall，*P-R*）曲线，如图 9 所示。改进后各类别 *P-R* 曲线整体向右上方偏移，曲线下面积显著增大。这表明改进模块在维持较高精确率的同时提升了召回水平，实现了精确率与召回率的协同优化。

由图 9 可见，模型误差主要表现为漏检与误检两类：微小热斑因目标尺度极小、像素占比低，其曲线在高召回率区间精确率骤降，AP 仅为 0.699，是漏检与误检的主要来源；PID 效应与微小热斑在红外图像中均呈现局部高温特征，热分布相近，边界模糊时易产生类间混淆。相比之下，玻璃破损、二极管异常等大尺度、热特征显著

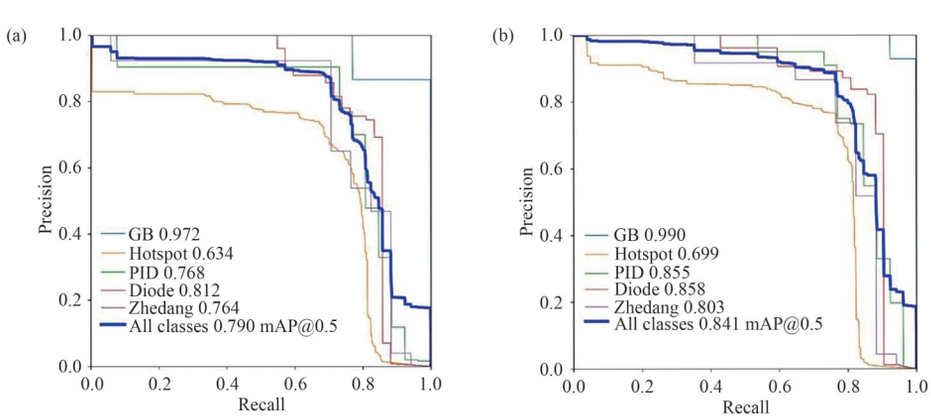


图 9　模型改进前（a）和改进后（b）的 *P-R* 曲线

Fig. 9　*P-R* curves of the model before (a) and after (b) improvement

的缺陷曲线饱满、AP较高(分别为0.990、0.858),误差较低。这表明模型误差集中于极小尺度目标及热特征相近类别,也印证了引入SPD-Conv保留细粒度特征、CSPOmniKernel抑制热背景干扰的必要性。

3.5 公开数据集泛化性测试

为进一步验证改进模型在异构成像设备、复杂气象条件及多样化目标分布下的泛化性能与鲁棒性,本研究选取光伏热斑检测红外基准数据集PV-HSD-2025进行跨场景迁移验证实验。该数据集采用R-JPEG格式存储,完整保留了高辐射灵敏度的原始温度数据,涵盖了分布式屋顶、地面电站等多种复杂背景下的热斑缺陷,具有极高的行业代表性与算法挑战性。

实验从该数据集中筛选标注完备的2 808张图像,严格按照8∶1∶1的比例划分训练集、验证集与测试集,且所有对比模型的超参数配置与输入分辨率均保持一致。

表3所示的量化评估结果表明,在面对全新的红外成像风格和环境背景时,各模型指标受数据集间域偏移(Domain Shift)的影响均有不同程度的波动。实验结果显示,改进模型MFK-DETR表现出显著的鲁棒性,其mAP@0.5达到了78.08%,较基准模型RT-DETR-r18大幅提升了3.67%。

表3 泛化实验结果

Tab. 3 Generalization experiment results

Models	mAP@0.5/%	*P*/%	*R*/%	FPS
YOLOv5s	65.25	72.54	63.12	97
YOLOv8s	69.80	73.92	66.45	85
YOLOv11n	71.12	75.18	68.20	**101**
YOLOv12n	75.12	77.33	72.85	83
RT-DETR-r18	74.41	75.12	72.69	50
RT-DETR-r34	75.32	74.55	**75.95**	33
RT-DETR-C2f-Mamba	73.85	74.12	71.88	37
RT-DETR-DCNV4	76.22	76.45	74.31	32
Ours	**78.08**	**77.42**	75.48	43

注:粗体数字表示最优值

此外,与目前主流的改进方案进行纵向对比,MFK-DETR的综合性能不仅优于引入Mamba结构的变体(73.85%),也超过了基于DCNV4算子的增强版RT-DETR(76.22%)。在精确率指标上,本研究模型达到了77.42%的最优水平;在召回率指标上(75.48%),虽然略低于参数量更大的RT-DETR-r34,但仍优于绝大多数对比模型。这有力地证明了MFK模块在处理异构红外特征时具有更高的特征聚合效率与抗干扰能力,能够有效降低环境差异导致的漏检与误报风险,在实际巡检工程中具有良好的应用潜力。

4 结 论

本文针对无人机航拍光伏红外图像中航拍小目标检测难、户外热背景干扰强以及缺陷目标尺度跨度大等挑战,提出了一种改进的RT-DETR光伏缺陷检测模型(MFK-DETR)。通过引入SPD-Conv模块替代传统下采样,模型有效规避了微小目标在特征提取过程中的细粒度信息丢失,显著提升了对二极管异常及热斑的捕捉能力。在此基础上,结合动态全维卷积(CSPOmniKernel)与多尺度特征融合模块(MFM),实现了对不同尺度目标的感知鲁棒性,并在增强低对比度缺陷特征的同时有效抑制了背景热噪声。实验结果表明,改进后的模型在自建数据集上的mAP@0.5达到84.12%,较基准模型提升了5.16%,且在公开数据集上表现出良好的泛化性能,推理速度完全满足工业实时巡检需求。综上所述,该方案在复杂户外场景下兼顾了检测精度与实时性能,为光伏电站的智能化运维提供了可靠的技术支撑。未来的研究将聚焦于极端气象条件下的性能优化,并进一步探索算法在嵌入式端的轻量化部署。

参　考　文　献:

[1] 王小宇,刘波,孙凯,等. 光伏阵列故障诊断技术综述[J]. 电工技术学报,2024,39(20):6526-6543.
WANG X Y, LIU B, SUN K, *et al*. A review of photovoltaic array fault diagnosis technology [J]. *Transactions of China Electrotechnical Society*, 2024, 39(20): 6526-6543. (in Chinese)

[2] 陈天祥,陈志聪,吴丽君,等. 面向边缘设备的EL图像光伏电池缺陷检测模型[J]. 福州大学学报(自然科学版),2026,54(1):1-9.
CHEN T X, CHEN Z C, WU L J, *et al*. An EL image photovoltaic cell defect detection model for edge devices [J]. *Journal of Fuzhou University* (*Natural Science Edition*), 2026, 54(1): 1-9. (in Chinese)

[3] 翟金星,王力平,张艳辉. 基于无人机红外视觉的光伏电站热斑智能巡检[J]. 红外技术,2026,48(1):95-104.
ZHAI J X, WANG L P, ZHANG Y H. Unmanned-aerial-vehicle infrared-vision-based intelligent inspection of hotspots in photovoltaic power station [J]. *Infrared Technology*, 2026, 48(1): 95-104. (in Chinese)

[4] 张绘敏,赵扬,康会峰. 基于卷积神经网络算法的光伏组件热斑图像检测方法研究[J]. 计算机测量与控制,2024,32(7):57-63.
ZHANG H M, ZHAO Y, KANG H F. Research on thermal spot image detection method for photovoltaic modules based on convolutional neural network algorithms [J]. *Computer Measurement & Control*, 2024, 32(7): 57-63. (in Chinese)

[5] 赵玲玲,汪烨,刘俊. 基于无人机与HSV空间的光伏电池板检测分析[J]. 红外技术,2020,42(10):978-982.
ZHAO L L, WANG Y, LIU J. Detection and analysis of photovoltaic panels based on UAV and HSV space [J]. *Infrared Technology*, 2020, 42(10): 978-982. (in Chinese)

[6] 张志豪,厉小润,陈淑涵. 基于改进YOLO11的无人机航拍图像小目标检测算法[J]. 液晶与显示,2025,40(6):915-930.
ZHANG Z H, LI X R, CHEN S H. Small object detection algorithm in UAV aerial images based on improved YOLO11 [J]. *Chinese Journal of Liquid Crystals and Displays*, 2025, 40(6): 915-930. (in Chinese)

[7] 王浩,闫号,叶海瑞,等. 基于无人机的光伏电站智能巡检[J]. 红外技术,2022,44(5):537-542.
WANG H, YAN H, YE H R, *et al*. Intelligent patrol inspection of photovoltaic power station based on UAVs [J]. *Infrared Technology*, 2022, 44(5): 537-542. (in Chinese)

[8] 赵强,刘胜杰,韩东成,等. 基于改进K均值聚类的光伏板缺陷检测方法[J]. 红外技术,2024,46(4):475-482.
ZHAO Q, LIU S J, HAN D C, *et al*. Improved K-means clustering-based defect detection method for photovoltaic panels [J]. *Infrared Technology*, 2024, 46(4): 475-482. (in Chinese)

[9] XU X Y, ZHAO M, SHI P X, *et al*. Crack detection and comparison study based on faster R-CNN and mask R-CNN [J]. *Sensors*, 2022, 22(3): 1215.

[10] JIANG P Y, ERGU D, LIU F Y, *et al*. A review of Yolo algorithm developments [J]. *Procedia Computer Science*, 2022, 199: 1066-1073.

[11] 孙建波,王丽杰,麻吉辉,等. 基于改进YOLO v5s算法的光伏组件故障检测[J]. 红外技术,2023,45(2):202-208.
SUN J B, WANG L J, MA J H, *et al*. Photovoltaic module fault detection based on improved YOLOv5s algorithm [J]. *Infrared Technology*, 2023, 45(2): 202-208. (in Chinese)

[12] 杨彦萍,高军伟,刘兆龙,等. 基于改进YOLO11的印刷电路板表面缺陷检测算法[J]. 液晶与显示,2025,40(10):1545-1556.
YANG Y P, GAO J W, LIU Z L, *et al*. Surface defect detection on printed circuit boards based on improved YOLO11 [J]. *Chinese Journal of Liquid Crystals and Displays*, 2025, 40(10): 1545-1556. (in Chinese)

[13] 闫号,戴佳佳,龚小溪,等. 基于多源图像融合的光伏面板缺陷检测[J]. 红外技术,2023,45(5):488-497.
YAN H, DAI J J, GONG X X, *et al*. Defect detection of photovoltaic panel based on multisource image fusion [J]. *Infrared Technology*, 2023, 45(5): 488-497. (in Chinese)

[14] CHEN S H, LU Y, QIN G X, *et al*. CSPD-DETR: real-time silicon crystalline photovoltaic cell surface defect detection transformer for building photovoltaic systems [J]. *Journal of Building Engineering*, 2025, 108: 112810.

[15] CHEN Y L, SHAN M T, ZHOU W, *et al*. RT-DETR-Pro: an enhanced RT-DETR model for visual detection of

photovoltaic module defects [J]. *IEEE Access*, 2025, 13: 116473-116479.

[16] JEONG H, YOON C, LIM H, *et al*. MT-former: multi-task hybrid transformer and deep support vector data description to detect novel anomalies during semiconductor manufacturing [J]. *Light: Advanced Manufacturing*, 2025, 6(2): 306-318.

[17] 邱刚,张廷龙,柏仓,等. 基于红外与可见光图像匹配的巡检机器人障碍识别与分类[J]. 红外技术,2025,47(1): 81-88.
QIU G, ZHANG N L, BAI C, *et al*. Inspection robot obstacle recognition and classification based on infrared and visible light image matching [J]. *Infrared Technology*, 2025, 47(1): 81-88. (in Chinese)

[18] WU N F, SUN Y X, HU J T, *et al*. Intelligent nanophotonics: when machine learning sheds light [J]. *eLight*, 2025, 5(1): 5.

[19] Mao W Z, Luan H Y, Ma R M. Singulonics: narwhal-shaped wavefunctions for sub-diffraction-limited nanophotonics and imaging [J]. *eLight*, 2025, 5(1): 27.

[20] LI J L, TONG T, LI D Q, *et al*. A dynamically adaptive and high-efficiency small object detection network for infrared thermographic images in online monitoring of solar photovoltaic panel defects [J]. *Energy*, 2025, 335: 138129.

[21] LI Z P, YUAN J Q, ZHANG Z Z. SolarIR-DETR: a lightweight transformer for UAV-assisted thermal hotspot detection in solar photovoltaic panels [J]. *Scientific Reports*, 2026, 16(1): 17172.

[22] 任鹏百,雷慧云,党建武,等. 基于HMSD与改进PCNN的红外与可见光图像融合[J]. 光学 精密工程,2025, 33(9):1481-1495.
REN P B, LEI H Y, DANG J W, *et al*. Infrared and visible image fusion based on HMSD and improved PCNN [J]. *Optics and Precision Engineering*, 2025, 33(9): 1481-1495. (in Chinese)

[23] YU S S, GAN F B, HAN T, *et al*. Multimodal fault detection model for photovoltaic modules [J]. *Scientific Reports*, 2026, 16(1): 3278.

[24] CAO H, HU H L, LU M, *et al*. LWMF-DETR: a lightweight transformer network for photovoltaic hot-spot fault detection with multi-scale wavelet feature fusion [J]. *Solar Energy Materials and Solar Cells*, 2026, 295: 114005.

作者简介:



邓万宇,男,博士,教授,2010年于西安交通大学获得博士学位,主要从事计算机视觉及图像处理方面的研究。E-mail: dengwanyu@126.com



彭鑫亮,男,硕士研究生,2024年于西安建筑科技大学获得学士学位,主要从事图像处理算法的研究。E-mail: 1802368981@stu.xupt.edu.cn