

基于特征增强的无人机航拍光伏板轻量级双模态缺陷检测器

唐小煜^{1,2,3}, 何慧伊可^{1,3}, 黄晓丹^{1,3}, 吴文隶^{2,3}

(1. 华南师范大学材料与新能源学院, 汕尾 516600; 2. 华南师范大学数据科学与工程学院, 汕尾 516600; 3. 华南师范大学行知书院, 汕尾 516600)

摘要: 无人机航拍光伏板缺陷检测中, 红外图像与可见光图像分别反映热异常和纹理结构信息, 但局部模态错位、微小缺陷弱响应易衰减以及双模态模型推理开销较高等问题, 仍制约检测性能与实际应用。本文提出基于特征增强的轻量级双模态缺陷检测器 (Feature-enhanced lightweight dual-modal detector, FLD-DET)。以红外与可见光双模态检测框架为基础, FLD-DET 在骨干特征提取过程中嵌入动态稀疏窗口 (Dynamic sparse window, DSWin), 通过局部卷积归纳、通道压缩和自适应注意力分配增强微小缺陷弱响应的保持能力。针对双模态特征之间的局部空间偏移问题, 本文设计轻量级跨模态注意力机制 (Lightweight cross-modal attention, LCoA), 利用方向位置信息编码增强双模态特征的空间对应关系。在检测特征生成阶段, FLD-DET 引入区域感知特征聚合 (Regional-aware feature aggregation, RAFA) 头, 强化缺陷与背景之间的判别表达。实验结果表明, FLD-DET 在航拍双模态光伏板缺陷数据集上实现了检测性能与推理效率的较好平衡, 并在微小缺陷识别方面具有一定优势。FLD-DET 在 M³FD 数据集上也取得了较优的检测结果, 展现出一定的跨场景适用性。

关键词: 光伏板缺陷检测; 双模态融合; 微小缺陷; 轻量化模型; 深度学习

中图分类号: TP391.41

文献标志码: A

引用格式: 唐小煜, 何慧伊可, 黄晓丹, 等. 基于特征增强的无人机航拍光伏板轻量级双模态缺陷检测器[J]. 数据采集与处理, 2026, 41(5): 1333–1348. TANG Xiaoyu, HE Huiyike, HUANG Xiaodan, et al. A feature-enhanced lightweight dual-modal detector for UAV-based photovoltaic panel defect detection[J]. Journal of Data Acquisition and Processing, 2026, 41(5): 1333–1348.

引言

无人机航拍巡检为大规模光伏板缺陷检测提供了便利, 但航拍图像中缺陷响应区域通常较小, 且易受拍摄角度、光照变化和背景纹理重复等因素影响。近年来, 基于深度学习的目标检测方法已广泛应用于光伏板缺陷检测任务。Faster R-CNN^[1]、YOLO 系列^[2-3]以及 U-Net^[4]等方法为缺陷定位与识别提供了有效技术基础。其中, YOLO 等单阶段检测器^[5]因推理速度较快, 较适合无人机巡检等实时检测场景。特征金字塔网络 (Feature pyramid network, FPN)^[6]及其改进结构则常用于增强小目标的多尺度特征表达能力。围绕小目标检测这一问题, 李生辉等^[7]针对 YOLOv5 引入双向特征金字塔网络 (Bi-directional feature pyramid network, BiFPN), 通过增强网络的多尺度特征融合能力, 提升了小目标检测性能; 周璇等^[8]则在 YOLOv5 的基础上, 通过拆分合并梯度流、融合多池化的设计, 强化模型对小目标的特征提取能

力。然而,光伏缺陷通常同时涉及红外热异常和可见光纹理扰动,仅依赖单一模态难以稳定获得完整的缺陷表征。

红外-可见光双模态融合能够补充单一模态的信息不足。红外图像能够反映缺陷引起的热异常信息,但纹理细节不足,且易受热噪声影响;可见光图像能够提供表面纹理与结构信息,但对低对比度缺陷和内部热异常响应有限。因此,双模态协同有助于提高复杂巡检环境下的缺陷识别稳定性。现有融合方式主要包括像素级融合、决策级融合和特征级融合。像素级融合通常通过拼接原始图像实现^[9],实现方式简单,但对噪声和配准误差较为敏感;决策级融合主要在检测结果层面进行整合,难以充分建模跨模态互补信息;相比之下,特征级融合能够在网络内部建模跨模态交互,因此更适用于双模态检测任务,研究者已提出多种改进方法。Liu等^[10]提出COMO(Cross-mamba interaction and offset-guided fusion)多光谱目标检测网络,通过Cross-Mamba交互与偏移引导融合策略缓解模态空间错位问题;Xiao等^[11]提出GM-DETR(Generalized multispectral detection Transformer),降低了对时空对齐数据的依赖。然而,面向无人机航拍光伏板缺陷检测任务,现有双模态方法仍难以同时兼顾跨模态对齐、微小缺陷表征与轻量化部署需求。

具体而言,现有研究主要存在以下3个方面不足。(1)跨模态空间对应关系建模仍不充分。光伏板缺陷在红外图像中多表现为局部热异常,在可见光图像中则多表现为纹理或结构扰动;受成像视角差异、传感器误差及配准偏差影响,两种模态在局部区域内往往存在轻微错位。对于微小缺陷而言,其有效判别区域十分有限,局部错位会直接削弱双模态证据的一致性。(2)微小缺陷特征在传播过程中易被背景响应淹没。热点、微裂纹等缺陷通常目标范围小、对比度低,常规卷积或固定窗口注意力的特征提取过程中,局部异常信号易与大面积背景纹理及热噪声共同聚合,产生明显的平均化效应,导致细粒度缺陷特征在深层传播中逐渐衰减。(3)高精度双模态检测模型通常需要同时处理两路输入并进行跨模态交互,容易带来额外的参数量、计算量和推理延迟,难以满足无人机巡检与资源受限边缘设备下的实时检测需求。

围绕上述问题,已有研究分别从注意力机制、局部细节建模和轻量化网络设计等方面进行了探索。SE-Net^[12]、CBAM^[13]等注意力机制能够增强关键通道或空间响应,但主要面向单模态特征增强,对红外与可见光之间的空间对应关系建模不足。Shen等^[14]提出ICAFusion(Iterative cross-attention fusion),通过迭代交叉注意力促进双模态特征交互;Xu等^[15]引入双注意力机制,以增强不同模态在空间与通道维度上的协同融合。在微小缺陷表征方面,卷积归纳偏置与Transformer全局建模能力的结合为微小缺陷的检测提供了新的思路。卷积神经网络(Convolutional neural networks, CNNs)^[1,16]擅长提取局部纹理与边缘细节,Transformer则在长程依赖建模和全局上下文交互方面具有优势,二者的协同有助于提升复杂场景下的小目标表征能力。MobileFormer、ConvNeXt^[17]和RepLKNet^[18]等方法也从不同角度探索了局部特征提取与全局建模的结合。在轻量化部署方面,MobileNetV3^[19]、GhostNet^[20]和YOLOv7-tiny^[21]等通过改进卷积结构、减少冗余特征计算和简化网络组成,为资源受限场景下的实时检测提供了参考。在光伏缺陷检测领域,Pinho等^[22]将YUV像素级融合与YOLOv11相结合,以降低双模态融合的计算开销。现有研究虽然分别改进了特征融合、局部建模和轻量化结构,但在光伏板双模态微小缺陷检测任务中仍需同时处理跨模态校准、弱响应保持和推理开销。红外热异常与可见光纹理扰动若不能形成稳定对应,融合特征容易受到模态错位影响;微小缺陷响应本身较弱,若在特征传播过程中被背景信息稀释,即使引入双模态信息也难以稳定转化为有效检测结果;双模态交互还会增加模型计算开销,限制其在无人机巡检和边缘设备中的应用。因此,有必要从检测流程整体出发,将跨模态特征协同、弱缺陷信息保持和轻量化推理统一考虑。

基于此,本文提出基于特征增强的轻量级双模态缺陷检测器(Feature-enhanced lightweight

dual-modal detector, FLD-DET)。围绕无人机航拍光伏板红外-可见光缺陷检测任务,本文从弱缺陷特征保持、跨模态空间校准和检测特征判别增强 3 个方面进行设计,以缓解单模态信息利用不足、双模态局部错位、微小弱响应易衰减以及双模态结构推理开销较高等问题。本文的主要工作如下:

(1)构建动态稀疏窗口(Dynamic sparse window, DSWin),通过局部卷积归纳、通道压缩和自适应注意力分配,增强微小缺陷弱响应的保持能力,并降低窗口注意力计算开销。

(2)设计轻量级跨模态注意力机制(Lightweight cross-modal attention, LCoA),通过方向位置信息编码增强红外与可见光特征之间的空间对应关系,提高双模态特征融合稳定性。

(3)引入区域感知特征聚合(Regional-aware feature aggregation, RAFA)头,增强缺陷区域与背景区域之间的判别能力,提高微小缺陷的分类与定位可靠性。

(4)在航拍双模态光伏板缺陷数据集和公开数据集 M³FD 上开展实验验证,结果表明所提方法在保持较低计算开销的同时取得了较好的检测性能。

1 FLD-DET 算法

1.1 FLD-DET 整体架构

针对光伏板双模态微小缺陷检测中模态局部错位、弱缺陷特征易衰减以及边缘端推理开销受限等问题,本文构建了特征增强型轻量级双模态检测器 FLD-DET,其整体结构如图 1 所示。FLD-DET 以配准后的红外图像和可见光图像作为输入,采用双分支骨干网络分别提取两种模态的层级特征,经跨模态空间校准和颈部特征融合网络完成多尺度特征整合,最终由检测头输出缺陷类别、边界框位置和置信度。

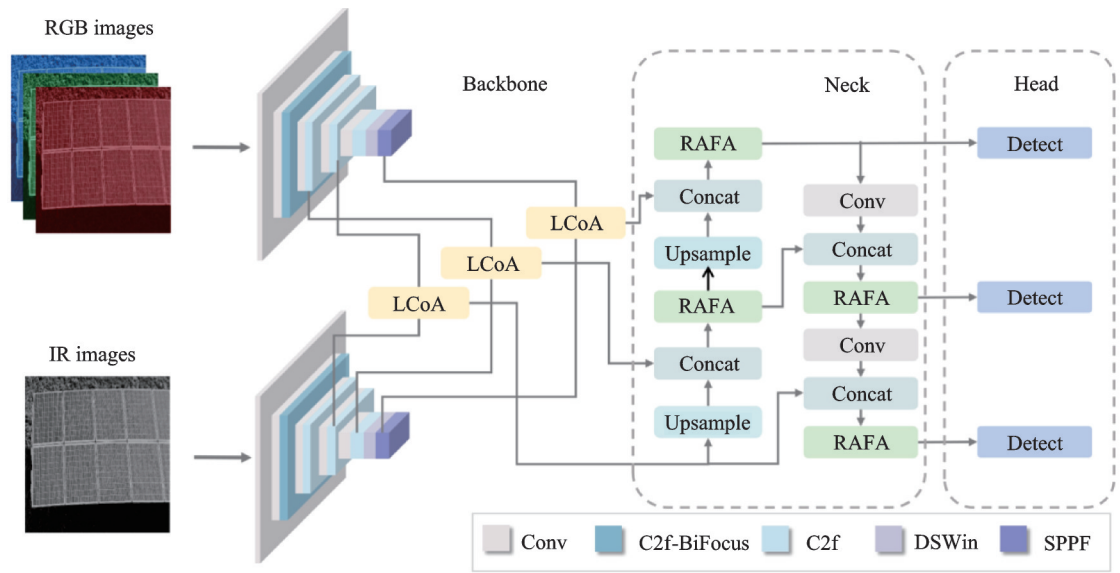


图1 FLD-DET 整体网络结构
Fig.1 Overall network structure of FLD-DET

常见双模态检测方法的骨干网络多沿用通用目标检测中的卷积结构或窗口注意力结构,能够提取多层次语义特征,但对光伏板微小缺陷的局部弱响应保持不足。无人机航拍图像中,热点、微裂纹等缺陷通常只在单块光伏板内部形成小范围热异常或细弱纹理扰动,其有效响应区域有限。随着骨干网络层级加深,常规卷积聚合和固定窗口注意力容易将局部异常响应与大面积背景纹理、热噪声共同建模,导致微小缺陷特征被逐渐稀释。为此,FLD-DET 在骨干特征提取过程中构建 DSWin,通过局部卷积归

纳增强细粒度异常响应,并结合通道压缩与自适应注意力头分配控制窗口注意力计算开销,使骨干网络在保持弱缺陷信息的同时兼顾轻量化推理需求。

双模态特征融合阶段的关键在于红外热异常信息与可见光纹理信息之间能否建立可靠的空间对应关系。现有融合方法常采用拼接、加权或常规注意力机制进行特征交互,这类方法在模态基本对齐时能够发挥互补作用,但对无人机航拍条件下的局部视角偏差、传感器误差和配准误差适应性不足。对于微小缺陷而言,红外图像中的热异常响应与可见光图像中的纹理扰动本身范围较小,一旦两种模态在局部区域内发生偏移,直接融合会降低双模态证据的一致性。为此,FLD-DET在双分支骨干网络与颈部特征融合网络之间设计 LCoA,通过方向位置信息编码增强红外与可见光特征之间的空间对应关系,使后续多尺度融合获得更稳定的双模态输入。

经过跨模态校准后的双模态特征仍需在颈部网络中进一步整合并形成检测特征。现有颈部特征融合结构通常依赖卷积、上采样和特征拼接完成多尺度信息传递,能够增强不同尺度目标的表达,但对微小缺陷区域与光伏板背景区域之间的细粒度判别仍不充分。光伏板表面纹理具有较强重复性,微小缺陷的边界与局部响应又较弱,若检测前特征缺乏区域级判别增强,容易造成缺陷区域与背景区域的响应混叠。为此,FLD-DET在颈部特征融合网络的多尺度融合路径中引入 RAFA,通过通道注意力与空间注意力协同建模,强化检测前特征中缺陷区域与背景区域之间的判别差异,提高微小缺陷分类与定位的可靠性。

综上,FLD-DET 围绕特征提取、模态融合和检测预测流程进行整体设计。骨干网络侧重保留微小缺陷的局部弱响应,跨模态校准结构用于增强红外与可见光特征之间的空间对应关系,颈部特征融合网络进一步强化缺陷区域与背景区域的判别差异。因此,FLD-DET 能够在提升微小缺陷表征能力的同时控制推理开销。下文将进一步介绍 DSWin、LCoA 和 RAFA 的具体结构与计算过程。

1.2 基于 DSWin 的骨干特征增强

基于 Swin Transformer 架构^[23],本文构建 DSWin,并将其用于骨干网络的层级特征增强,其结构如图 2 所示(图中物理量含义见后文解释)。

在注意力计算前,DSWin 首先对输入特征应用 3×3 深度可分离卷积,并通过残差连接保留原始特

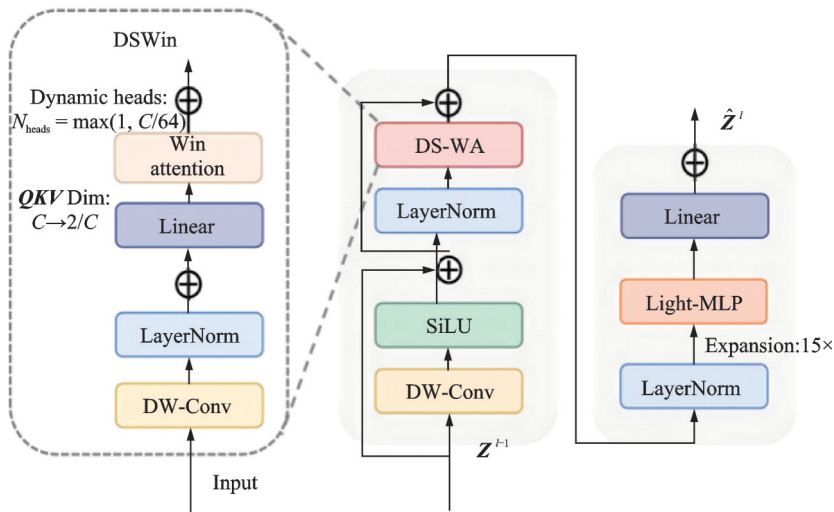


图2 动态稀疏窗口结构图

Fig.2 Structure of DSWin

征。其中,深度可分离卷积以较低计算代价引入局部卷积归纳偏置,使模型对裂纹边缘、局部热点等细粒度异常模式更敏感;残差路径可避免增强过程对原始弱信号造成过度重构,使缺陷相关信息在进入窗口注意力前保持较高保真度。其特征增强公式为

$$X^{l-1} = \text{DWConv}(X^{l-1}) + X^{l-1} \quad (1)$$

式中: X^{l-1} 表示增强后的特征图, X^{l-1} 为第 $l-1$ 层输入特征,DWConv表示 3×3 深度可分离卷积。

为降低窗口注意力的计算开销,本文将查询(Q)、键(K)、值(V)矩阵投影层的输出通道数从输入特征图的通道数 C 压缩至 $C/2$,并设置最小通道数下限为32。这一设计既通过将通道数减半实现了冗余计算的精简,又通过设置通道数下限,避免了通道压缩过强导致的特征表达退化,防止微小缺陷的低占比信号特征被丢弃。同时,本文将多层感知机(Multilayer perceptron, MLP)的特征扩展率由传统Transformer架构中的4倍降至1.5倍。由于MLP主要承担特征的非线性变换与维度映射功能,降低扩展率能在保留关键特征表达能力的同时大幅削减冗余计算量,提升模型推理速度。

固定数量的注意力头难以适应不同层级特征的代表需求。浅层特征以局部纹理为主,过多注意力头会造成冗余计算,深层特征则更依赖充分的表征容量,注意力头不足又会限制建模能力。由于在本网络结构中,不同层级特征通常对应不同的通道维度,而较高通道维度往往意味着更强的语义表征需求,因此本文采用输入通道数作为注意力头分配的近似依据,设计自适应注意力头分配策略,以实现窗口注意力计算资源的动态调整,其计算公式为

$$N_{\text{heads}} = \max\left(1, \left\lfloor \frac{C}{64} \right\rfloor\right) \quad (2)$$

式中: N_{heads} 为计算得到的注意力头数量, $\lfloor \cdot \rfloor$ 表示向下取整函数。该设计在降低窗口注意力开销的同时,为后续跨模态校准保留更多细粒度响应。

1.3 基于 LCoA 的跨模态空间校准

受坐标注意力机制(Coordinate attention, CA)^[24]启发,本文设计 LCoA 并将其嵌入双分支骨干网络与颈部特征融合网络之间,结构如图3所示(图中物理量含义见后文解释)。LCoA 保留了坐标注意力沿水平和垂直方向分解空间信息的基本思想,并将其扩展到红外与可见光双模态特征的协同建模中。具体而言,模块分别对两种模态在水平和垂直方向上的响应进行编码,以增强双模态特征之间的空间对应关系,提高局部异常区域的融合一致性。其水平和垂直特征响应计算公式分别为

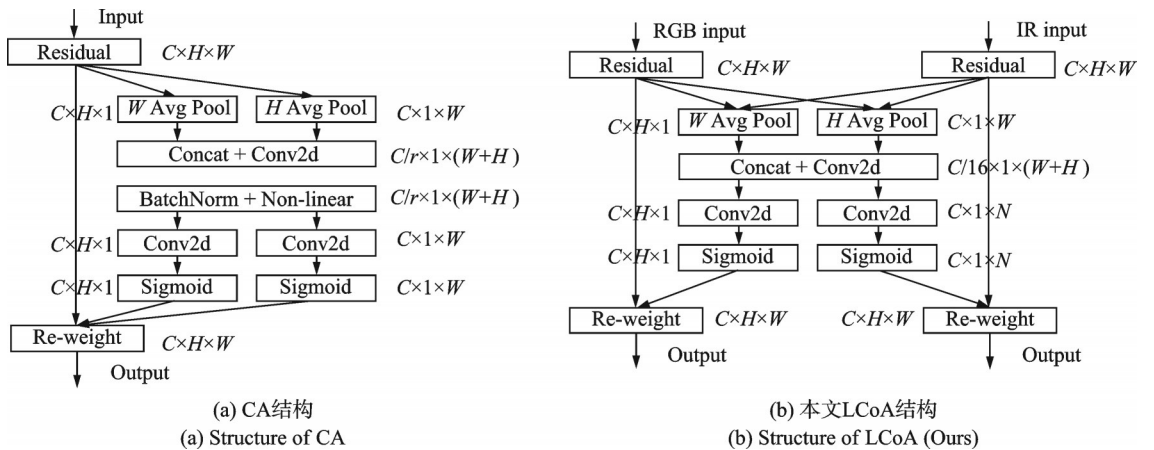


图3 CA与LCoA注意力机制结构对比图

Fig.3 Architecture comparison of CA and LCoA attention mechanisms

$$z_c^h(j) = \frac{1}{W} \sum_{0 \leq i < W} x_c(i, j) \quad (3)$$

$$z_c^w(i) = \frac{1}{H} \sum_{0 \leq j < H} x_c(i, j) \quad (4)$$

式中: $z_c^h(j)$ 表示第 c 个通道的水平特征响应, $z_c^w(i)$ 表示第 c 个通道的垂直特征响应, $x_c(\cdot)$ 为第 c 个通道的特征值, H 、 W 分别为特征图的高度和宽度。

为降低模块计算开销, LCoA 在特征变换中引入瓶颈结构, 并采用固定压缩比 ($r=16$) 对通道维度进行压缩。对于微小缺陷检测而言, 缺陷相关通道往往表现为低占比的异常响应, 若压缩过强, 这类弱判别信息会优先被舍弃, 导致融合特征对细粒度异常的感知下降。因此, 该压缩比是在模型复杂度与缺陷信息保留之间进行权衡后确定的, 可在降低计算开销的同时尽量保留关键跨模态上下文信息。考虑到光伏板缺陷图像背景均匀、缺陷对比度低的特点, 本文去除了初始特征变换后的批归一化 (Batch-Norm) 层。经 LCoA 校准后, 红外热异常信息与可见光纹理信息在空间维度上的对应关系得到增强, 有利于颈部特征融合网络进行稳定的双模态特征整合。

1.4 基于 RAFA 的区域感知特征聚合

受动态特征融合 (Dynamic feature fusion, DFF) 模块^[25]启发, 本文设计了 RAFA, 并将其嵌入颈部特征融合网络的多尺度融合路径中, 其结构如图 4 所示 (图中物理量含义见后文解释)。RAFA 从区域级特征建模出发, 通过协同引入通道注意力与空间注意力, 增强缺陷区域与背景区域之间的判别稳定性。在通道注意力分支中, 模块采用全局平均池化和全局最大池化并行的双池化架构。全局平均池化能够描述特征图的整体响应分布, 有助于建模缺陷区域的全局统计模式; 全局最大池化则更敏感于局部显著响应, 能够突出热点边缘、裂纹纹理等局部异常峰值。二者互补, 可同时保留全局分布信息与局部显著信息, 从而避免单一池化机制在细粒度缺陷建模上的偏置。随后, 模块通过元素相加融合两种池化输出, 经 1×1 卷积与 Sigmoid 激活生成通道注意力图, 实现对不同尺度缺陷特征的全覆盖表征, 计算公式为

$$w_{ch} = \sigma(\text{Conv}_{1 \times 1}(\text{AvgPool}(F) + \text{MaxPool}(F))) \quad (5)$$

式中: $w_{ch} \in \mathbb{R}^{2C \times 1 \times 1}$ 表示通道注意力图, σ 为 Sigmoid 激活函数, AvgPool、MaxPool 分别为全局平均池化与全局最大池化。

在空间分支中, 微小缺陷的边界与局部几何模式往往依赖邻域内的空间结构, 若仅采用 1×1 卷积进行通道线性组合, 难以充分利用局部空间连续性信息。为此, 本文将输入特征 F_1 、 F_2 所采用的普通 1×1 卷积替换为 3×3 深度可分离卷积 (Depthwise separable convolution, DWConv), 以在较低计算开销下增强局部邻域建模能力。深度可分离卷积由空间特征提取的深度卷积 (Depthwise convolution) 与通道维度融合的逐点卷积 (Pointwise convolution) 两阶段构成。深度卷积能有效扩展局

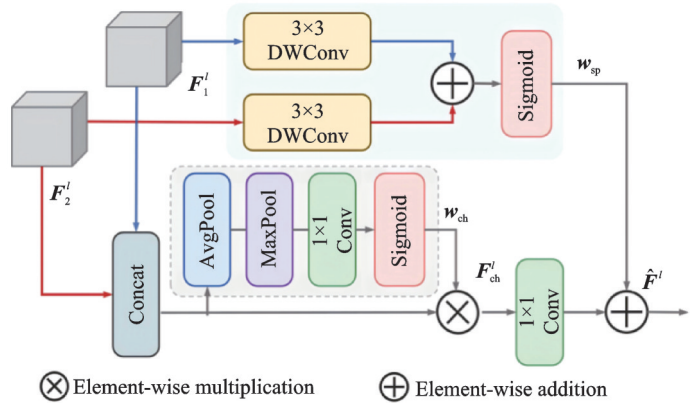


图 4 区域感知特征聚合头结构图

Fig.4 Structure of RAFA head

部感受野,更充分地捕捉缺陷边界的空间关联信息,逐点卷积则可在更低的计算开销下完成通道维度的特征融合。二者协同作用,在提升空间特征表征能力的同时有效降低了模型的计算复杂度。将这两个卷积阶段的输出进行逐元素相加后,再通过 Sigmoid 激活函数即可生成最终的空间注意力图,公式如下

$$\boldsymbol{w}_{\text{sp}} = \sigma(\text{DWConv}_{3 \times 3}(F_1) + \text{DWConv}_{3 \times 3}(F_2)) \quad (6)$$

式中: $\boldsymbol{w}_{\text{sp}} \in \mathbf{R}^{1 \times H \times W}$ 表示空间注意力图, F_1 、 F_2 为空间注意力分支的两路输入特征。随后,通过 1×1 卷积将通道校准后的特征恢复至原始维度,利用空间注意力图对其进行逐位置加权,实现通道权重调整与空间位置增强的协同作用。经通道校准与空间加权后的区域增强特征被送入检测头,用于缺陷类别预测和边界框回归,有助于提升微小缺陷区域的判别能力和定位可靠性。

2 实验结果与分析

2.1 数据集

本文采用课题组基于无人机采集的双模态光伏板缺陷数据集和公开数据集 M³FD^[26]对所提模型进行验证。其中,光伏板缺陷数据集包含 331 对配准的红外和可见光图像,涵盖大面积失效区域、热点、局部高风险热点、微裂纹、旁路二极管故障及部分遮挡 6 类典型光伏板缺陷,部分样本如图 5 所示。

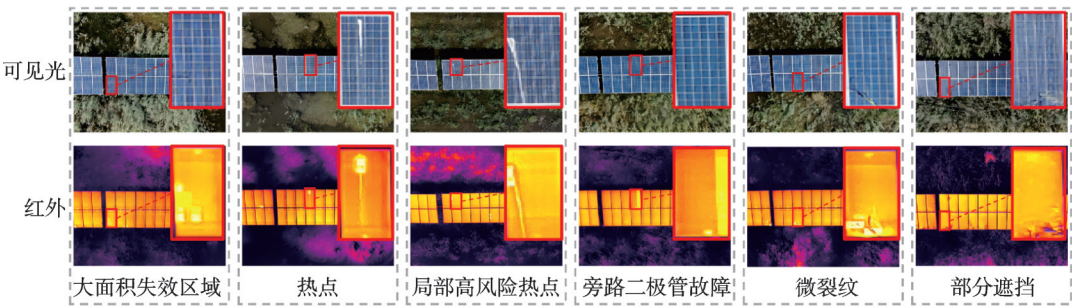


图 5 航拍双模态光伏板缺陷数据集样本图

Fig.5 Samples of aerial-captured dual-modal photovoltaic panel defect dataset

本文以缺陷响应区域在光伏板内部的相对占比作为微小缺陷的划分依据。设缺陷所在光伏板区域面积为 A_p , 板内缺陷响应区域面积为 A_d , 则缺陷相对尺度定义为

$$s = \frac{A_d}{A_p} \times 100\% \quad (7)$$

本文将 $s \leq 5\%$ 的局部缺陷定义为微小缺陷。结合缺陷形态特征与样本复核结果,热点、局部高风险热点和微裂纹主要表现为局部热异常或细微纹理扰动,归为微小缺陷类别,其余缺陷类别则归为较大缺陷类别。光伏板缺陷数据集中不同尺度缺陷的分布统计如表 1 所示。统计结果表明,微小缺陷占比为 60.5%,为后续分析模型对微小缺陷的检测能力提供了数据基础。

表 1 光伏板缺陷数据集不同尺度缺陷的分布统计

Table 1 Distribution of defect instances at different scales in the photovoltaic panel defect dataset

缺陷尺度	缺陷类别	实例数	占比/%
微小缺陷	热点、局部高风险热点、微裂纹	1 374	60.5
较大缺陷	大面积失效区域、旁路二极管故障、部分遮挡	897	39.5

此外,本文还引入公开数据集 M³FD 进行补充实验,评估所提方法在通用双模态目标检测任务中的适用性与泛化能力。M³FD 公开数据集包含 4 200 对图像,覆盖校园、景区、城市道路等多种真实场景,含行人、汽车、路灯、卡车、公交车和摩托车 6 大目标类别。

2.2 实验配置

实验在 Windows 与 CUDA 11.7 环境下进行,硬件平台为 NVIDIA RTX 3090 GPU(24 GB 显存),模型基于 PyTorch 1.13.0 和 Python 3.9 实现。所有实验训练 200 个 epoch,批大小设为 32,输入图像尺寸统一为 640 像素×640 像素,优化器采用 AdamW,动量参数设为 0.95。

2.3 评价指标

本文选用召回率 R 、均值平均精度(mean Average precision, mAP)作为核心评价指标。针对目标检测任务,分别计算交并比(Intersection over union, IoU)阈值为 0.5 时的 mAP,以及 IoU 阈值从 0.5 到 0.95、步长为 0.05 时的平均 mAP,分别记为 $\text{mAP}_{0.5}$ 和 $\text{mAP}_{0.5:0.95}$ 。计算公式如下

$$R = \frac{\text{TP}}{\text{TP} + \text{FN}} \times 100\% \quad (8)$$

$$\text{AP} = \sum_{i=1}^{n-1} (r_{i+1} - r_i) P(r_{i+1}) \quad (9)$$

$$\text{mAP} = \frac{\sum_{i=1}^K \text{AP}_i}{K} \quad (10)$$

式中:TP、FP 和 FN 分别表示在设定 IoU 阈值下的正确匹配预测框、错误预测框和漏检真实框; $r_1, r_2, \dots, r_n$ 为按升序排列的精确度 P 插值段第一个插值处对应的召回率值, K 为检测任务中的总类别数,AP 为单类别下精度-召回曲线的积分面积。

此外,本文采用参数量 Params、计算复杂度 FLOP 与单图像推理时间评价模型复杂度与推理效率。其中,模型参数量反映模型规模,FLOP 反映推理计算开销,单图像推理时间反映实际推理速度。

2.4 对比实验

2.4.1 光伏板双模态数据集上的对比实验

在航拍双模态光伏板缺陷数据集上,本文将 FLD-DET 与代表性单模态及双模态检测方法进行比较,以评估其在光伏板缺陷检测任务中的有效性。单模态方法包括 SparseR-CNN、YOLO 系列^[3,27,28]等;双模态方法包括 GM-DETR^[11]、IRFS^[29]、ICAFusion^[14]和 SeAFusion^[30]等。实验结果如表 2 所示,表中红色表示最佳结果,蓝色表示次佳结果。由表 2 可知,单模态检测器在可见光与红外模态下的性能差异较为明显。对于 YOLOv11、YOLOv12 等近年代表性单模态检测器,其在红外模态下的检测精度普遍高于可见光模态,说明红外图像中的热异常信息对光伏缺陷识别具有重要作用。然而,单一红外模态仍难以充分表征缺陷边界和局部纹理结构,可见光模态虽然能够提供表面纹理信息,但对低对比度热异常响应有限。因此,仅依赖单一模态难以同时兼顾热异常定位和纹理结构判别。

相比单模态检测器,双模态方法能够同时利用红外热异常信息和可见光纹理结构信息,在光伏缺陷检测中具有更好的信息互补优势。FLD-DET 在双模态条件下取得了 63.8% 的 $\text{mAP}_{0.5}$ 和 55.0% 的 $\text{mAP}_{0.5:0.95}$,整体优于多数单模态基线及现有双模态方法。

除检测精度外,模型复杂度和推理效率也是无人机巡检场景中的重要指标。FLD-DET 的参数量为 5.72×10^6 ,FLOP 为 8.14×10^9 ,单图像推理时间为 2.6 ms。与 YOLOv12 等单模态实时检测器相比,

表 2 FLD-DET 与其他先进算法在航拍双模态光伏板缺陷数据集上的检测性能对比

Table 2 Comparison of detection performance between FLD-DET and other advanced algorithms on the aerial dual-modal photovoltaic defect dataset

模型	模态	$P/\%$	$R/\%$	$mAP_{0.5}/\%$	$mAP_{0.5:0.95}/\%$	Params/ 10^6	FLOP/ 10^9	推理时间/ ms
Sparse-RCNN	可见光	—	—	46.5	46.4	—	—	—
	红外	—	—	56.8	56.7	—	—	—
YOLOv12	可见光	49.7	50.3	44.2	35.8	2.5	5.8	8.0
	红外	56.4	59.3	62.1	53.7	2.5	5.8	7.4
YOLOv11	可见光	51.8	48.0	41.5	33.1	2.6	6.3	1.5
	红外	59.8	60.9	61.7	52.9	2.6	6.3	1.5
YOLOv9	可见光	37.7	55.8	34.7	27.7	50.7	236.7	38.7
	红外	62.4	57.2	60.6	53.6	50.7	236.7	42.0
GM-DETR	可见光	—	—	51.6	41.6	72.17	—	249.17
	红外	—	—	55.5	44.8	72.17	—	249.17
	双模态	—	—	56.8	46.6	72.17	—	249.17
IRFS	双模态	70.4	53.3	54.7	45.2	7.03	15.08	9.7
CMADet	双模态	70.4	53.2	54.1	45.7	33.36	17.00	14.3
ICAFusion	双模态	57.8	57.8	56.5	44.3	7.08	8.26	11.3
SeAFusion	双模态	78.2	56.9	56.1	43.2	7.03	15.80	9.7
FLD-DET(Ours)	双模态	59.8	65.5	63.8	55.0	5.72	8.14	2.6

FLD-DET 在引入双模态信息后取得了更高的检测精度,说明红外热异常信息与可见光纹理结构信息的互补有助于提升光伏缺陷识别效果。与多数双模态方法相比,FLD-DET 在取得较高检测精度的同时具有更低的单图像推理时间,说明其在引入双模态信息后仍保持了较低的推理开销。

图 6 从检测精度和推理效率两个维度比较不同模型,图中散点面积表示模型推理时间。FLD-DET 位于散点图右上区域,且对应散点面积较小,表明其在取得较高检测精度的同时仍保持了较低的单图像推理时间。

为直观展示模型的检测效果,图 7 对比了 FLD-DET、YOLOv9、YOLOv12 与真实标注之间的可视化结果。结果显示,FLD-DET 的检测框位置与真实标注更加一致,尤其在微小缺陷样本下能够较完整地覆盖目标区域并保持较好的边界定位效果。相比之下,YOLOv9 和 YOLOv12 在部分样本中出现了漏检、误检以及框选范围偏移等现象。该可视化结果与表 2 相互印证,进一步说明 FLD-DET 在复杂航拍场景下具有较好的缺陷识别与定位能力。

2.4.2 微小缺陷检测能力分析

基于 2.1 节中的尺度划分标准,本文分别统计热点、局部高风险热点和微裂纹 3 类微小缺陷的

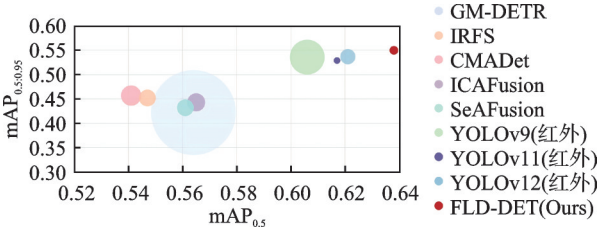


图 6 检测精度与推理时间散点图

Fig.6 Scatter plot of detection accuracy and inference time

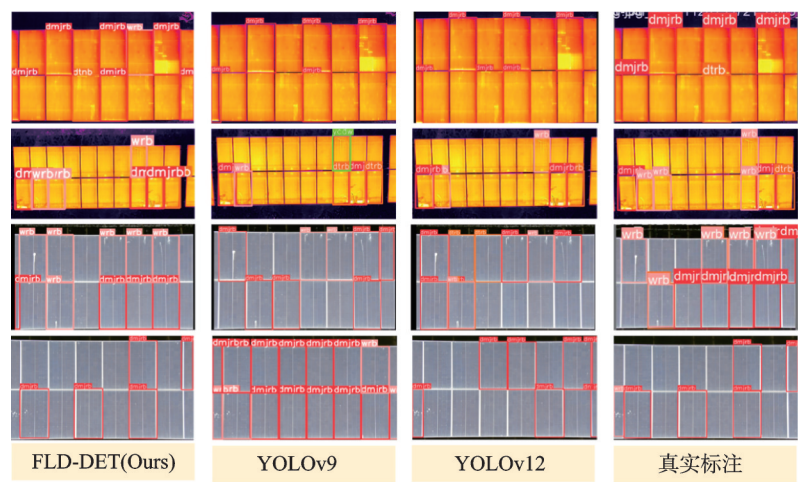


图7 不同模型检测效果与真实标注的可视化对比

Fig.7 Visual comparison of detection results and ground-truth annotations for different models

$AP_{0.5}$,并与DEYOLO、ICAFusion进行对比,以评估模型对不同微小缺陷类别的检测能力,结果如表3所示。

表3 不同微小缺陷类别检测性能对比

Table 3 Comparison of detection performance for different tiny defect categories				%
模型	热点 $AP_{0.5}$	局部高风险热点 $AP_{0.5}$	微裂纹 $AP_{0.5}$	微小缺陷 $mAP_{0.5}$
DEYOLO	26.2	56.3	32.0	38.2
ICAFusion	35.9	58.1	7.2	33.7
FLD-DET(Ours)	47.5	66.0	37.9	50.5

由表3可知,FLD-DET在3类微小缺陷上均取得了较优检测结果, $mAP_{0.5}$ 达到50.5%。其中,热点类别的提升较为明显。该类缺陷通常呈现为局部点状或小片状热异常,响应范围有限,容易在特征传播过程中受到光伏板背景纹理和热噪声干扰。FLD-DET在该类别上的提升表明,DSWin对局部弱响应信息的保持,以及RAFA对区域判别特征的增强,有助于改善小范围热异常特征表达不足的问题。对于局部高风险热点,其热异常区域相对更明显,但FLD-DET仍取得进一步提升,表明LCoA的跨模态空间校准有助于提高红外热异常与可见光结构信息的融合一致性。微裂纹类别主要表现为细弱纹理扰动,其边界连续性和局部对比度相对不足,对细粒度特征保持能力提出了更高要求。FLD-DET在该类别上仍取得了更高的精度,表明FLD-DET对细粒度弱纹理缺陷也具有一定适应能力。总体而言,FLD-DET在热异常型与纹理扰动型微小缺陷上均表现出较好的检测能力,说明所提模块在局部弱响应应保持、跨模态信息协同和区域判别增强方面具有一定互补作用。

2.4.3 泛化性验证

为评估FLD-DET在通用双模态目标检测任务中的适用性,本文在公开数据集M³FD上进行了对比实验,结果如表4所示,其中红色表示最佳结果,蓝色表示次佳结果。由表4可知,FLD-DET在M³FD上取得了87.1%的 $mAP_{0.5}$,为对比方法中的最高值。同时,在行人、公交车、摩托车和卡车等多个关键类别上均取得了较优检测结果。图8以雷达图形式呈现各模型在不同类别上的检测结果,用于比较模型在多类别场景下的性能分布。其中,半径方向表示检测精度,离中心越远代表在该类别上的性能越

表 4 FLD-DET 与其他先进算法在 M³FD 数据集上的检测性能对比

Table 4 Comparison of detection performance between FLD-DET and other advanced algorithms on the M³FD dataset

模型	mAP _{0.5}	mAP _{0.5:0.95}	检测精度					
			行人	汽车	公交车	摩托车	路灯	卡车
DEYOLO ^[31]	86.6	58.9	82.4	89.9	92.3	74.1	74.0	90.8
Fusion-Mamba ^[32]	85.0	57.5	80.3	91.9	92.8	73.0	84.8	87.1
IRFS ^[29]	81.2	55.8	78.8	90.7	92.2	69.5	72.1	83.9
Tardal ^[26]	81.0	54.9	81.1	89.8	90.3	71.0	70.0	83.8
CDDFusion ^[33]	80.3	54.9	77.7	89.7	90.2	69.3	68.5	86.6
U2Fusion ^[34]	79.2	53.8	77.5	90.1	91.4	64.3	67.7	84.1
SeAFusion ^[30]	80.7	55.4	79.5	90.4	90.9	67.6	70.5	85.2
SwinFusion ^[35]	80.2	54.7	78.4	89.9	90.0	67.5	67.9	87.6
DetFusion ^[36]	80.8	53.8	80.8	83.0	92.5	69.4	87.8	71.4
PIAFusion ^[37]	80.6	55.0	78.8	90.0	91.6	68.0	70.8	84.2
SuperFusion ^[38]	83.5	56.0	83.7	93.2	91.0	77.4	70.0	85.8
IGNet ^[39]	81.5	54.5	81.6	82.4	92.8	73.0	86.9	71.5
RI-YOLO ^[40]	83.6	56.6	84.1	91.6	88.5	73.5	75.7	87.8
FLD-DET(Ours)	87.1	55.0	84.9	93.0	94.6	80.0	76.4	93.9

优,FLD-DET 整体上占据最外层轮廓,在多数类别上的检测精度均处于较优水平。此外,FLD-DET 的 $mAP_{0.5:0.95}$ 也达到 55.0%,与其他模型精度相当,体现出模型在多种 IoU 阈值下均具备较稳定的检测一致性。结果表明,所提方法在外部公开双模态目标检测数据集上仍保持较强竞争力,具有一定的跨场景迁移潜力。

2.5 消融实验

2.5.1 模块消融

围绕跨模态空间校准、局部弱响应保持和区域判别增强 3 项设计,本文以 DEYOLO 为基准模型,在航拍双模态光伏板缺陷数据集上开展模块消融实验,分别分析 DSWin、LCoA 和 RAFA 单独引入及组合使用时对检测性能和模型复杂度的影响。具体而言,实验首先在基线模型中单独引入各模块,以评估其独立作用;随后对模块进行组合,以分析其协同增益。实验结果如表 5 所示,其中“√”表示加入相应改进点进行训练。由表 5 可知,模型(1)引入 LCoA 后, $mAP_{0.5}$ 提升 9.3%,表明引入的跨模态空间校准对双模态特征融合具有积极作用;模型(2)加入 DSWin 后, $mAP_{0.5}$ 提升 7.4%,原因是该模块通过增强局部细节建模与自适应计算分配,优化了小缺陷的表征效率;模型(3)使用 RAFA 后, $mAP_{0.5}$ 提升 11.4%,其优势在于通过双路径聚合机制强化了多尺度特征的判别力,改善了缺陷边界的识

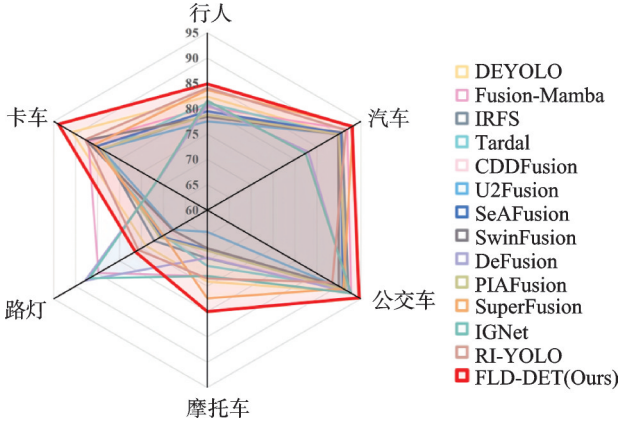


图 8 M³FD 数据集上各类别检测精度对比雷达图

Fig.8 Radar chart comparing detection accuracy across categories on the M³FD dataset

表 5 改进点消融实验结果对比

Table 5 Comparison of ablation results for different modules

模型	LCoA	DSWin	RAFA	mAP _{0.5} /%	mAP _{0.5:0.95} /%	Params/10 ⁶	FLOP/10 ⁹	推理时间/ms
DEYOLO				52.6	44.8	6.01	8.41	3.0
(1)	✓			57.5	50.6	6.02	7.00	3.5
(2)		✓		56.5	48.9	5.74	7.73	3.2
(3)			✓	58.5	51.2	5.89	7.64	5.1
(4)	✓	✓		58.2	50.5	5.79	8.32	2.7
(5)	✓		✓	60.4	51.7	5.75	8.32	3.4
(6)		✓	✓	59.9	51.1	5.66	7.55	3.1
FLD-DET	✓	✓	✓	63.8	55.0	5.72	8.14	2.6

别精度。

从模型(4~6)的组合结果可以看出,当两个模块联合使用时,模型性能较单模块方案仍有提升。当三者同时引入后,FLD-DET 的 mAP_{0.5} 和 mAP_{0.5:0.95} 分别达到 63.8% 和 55.0%,取得了最佳综合性能。该结果表明,DSWin、LCoA 与 RAFA 分别在跨模态融合、特征增强和最终预测阶段发挥作用,三者具有较好的互补性,协同设计有助于提升双模态光伏缺陷检测效果。

2.5.2 注意力机制对比

在相同基线模型与训练配置下,本文将 LCoA 与 SE、CBAM 和 CA 等轻量注意力机制进行比较,以评估其在跨模态特征校准中的有效性,实验结果如表 6 所示。对于光伏双模态缺陷检测任务,注意力机制不仅需要增强有效特征,还需要兼顾跨模态空间对应关系建模和弱缺陷响应保留。SE 主要通过通道重标定增强特征响应,但难以直接表征目标的空间位置信息;CBAM^[13] 进一步引入空间注意力,但其作用仍主要集中于单模态内部的显著区域增强,难以显式建立红外与可见光之间的空间对应关系。表 6 结果表明,相比无注意力、SE、CBAM 和 CA,LCoA 在 mAP_{0.5} 和 mAP_{0.5:0.95} 上均取得最高结果,同时计算复杂度降低,这说明 LCoA 在轻量化条件下能够更有效地兼顾跨模态对齐和细粒度缺陷表征。

表 6 不同轻量注意力机制的对比实验结果

Table 6 Comparison of different lightweight attention mechanisms

模型	注意力模块	mAP _{0.5} /%	mAP _{0.5:0.95} /%	Params/10 ⁶	FLOP/10 ⁹	推理时间/ms
DEYOLO	无	52.6	44.8	6.01	8.41	3.0
	CBAM	52.3	44.3	6.01	8.41	3.2
	SE	52.6	45.4	6.01	8.41	3.2
	CA	54.6	47.2	6.02	8.41	3.3
	LCoA	57.5	50.6	6.02	7.00	3.5

3 结束语

针对光伏板双模态微小缺陷检测中弱缺陷响应易衰减、跨模态空间对应不稳定以及检测前区域别特征不足等问题,本文提出了基于特征增强的无人机航拍光伏板轻量级双模态缺陷检测器 FLD-DET。FLD-DET 按照特征提取、模态校准和检测特征生成的流程进行设计,分别引入 DSWin、

LCoA 和 RAFA,以增强微小缺陷局部弱响应保持能力、提高红外与可见光特征的空间对应关系,并强化缺陷区域与背景区域之间的判别表达。实验结果表明,FLD-DET 在航拍双模态光伏板缺陷数据集上取得了较好的检测性能,并在微小缺陷检测任务中表现出较好的识别能力。在公开数据集 M³FD 上,FLD-DET 的 mAP_{0.5} 达到 87.1%,具备一定的跨场景迁移潜力。

本文实验仍主要基于规模有限的航拍双模态光伏板缺陷数据集,且尚未在真实机载边缘设备上完成部署验证。后续研究将扩充复杂环境下的双模态光伏缺陷样本,并围绕弱缺陷检测、小样本学习以及边缘部署条件下的推理效率与稳定性开展深入研究,以进一步提升模型在实际光伏巡检场景中的应用能力。

参考文献:

- [1] REN S, HE K, 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): 1137-1149.
- [2] WANG C Y, BOCHKOVSKIY A, LIAO H M. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors[C]//*Proceedings of 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Vancouver, BC, Canada: IEEE, 2023: 7464-7475.
- [3] WANG C Y, YE H I, MARK LIAO H Y. YOLOv9: Learning what you want to learn using programmable gradient information[C]//*Proceedings of Computer Vision—ECCV 2024*. Cham, Switzerland: Springer Nature, 2025: 1-21.
- [4] RONNEBERGER O, FISCHER P, BROX T. U-Net: Convolutional networks for biomedical image segmentation[EB/OL]. (2015-05-18). <https://doi.org/10.48550/arXiv.1505.04597>.
- [5] BOCHKOVSKIY A, WANG C Y, LIAO H M. YOLOv4: Optimal speed and accuracy of object detection[EB/OL]. (2020-04-23). <https://doi.org/10.48550/arXiv.2004.10934>.
- [6] LIN T Y, DOLLÁR P, GIRSHICK R, et al. Feature pyramid networks for object detection[C]//*Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Honolulu, HI, USA: IEEE, 2017: 936-944.
- [7] 李生辉, 李晓飞, 宋璋晗, 等. 基于改进 YOLOv5 的船舶多尺度 SAR 图像检测算法[J]. *数据采集与处理*, 2024, 39(1): 120-131.
LI Shenghui, LI Xiaofei, SONG Zhanghan, et al. Multi-scale SAR image detection algorithm for ships based on improved YOLOv5[J]. *Journal of Data Acquisition and Processing*, 2024, 39(1): 120-131.
- [8] 周璇, 葛琦, 邵文泽. 高分辨率特征增强的无人机航拍小目标检测[J]. *数据采集与处理*, 2024, 39(4): 908-921.
ZHOU Xuan, GE Qi, SHAO Wenze. Small target detection in UAV aerial images based on high resolution feature enhancement[J]. *Journal of Data Acquisition and Processing*, 2024, 39(4): 908-921.
- [9] KULKARNI S C, REGE P P. Pixel level fusion techniques for SAR and optical images: A review[J]. *Information Fusion*, 2020, 59: 13-29.
- [10] LIU C, MA X, YANG X, et al. COMO: Cross-Mamba interaction and offset-guided fusion for multimodal object detection[J]. *Information Fusion*, 2026, 125: 103414.
- [11] XIAO Y, MENG F, WU Q, et al. GM-DETR: Generalized multispectral detection Transformer with efficient fusion encoder for visible-infrared detection[C]//*Proceedings of 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. Seattle, WA, USA: IEEE, 2024: 5541-5549.
- [12] HU J, SHEN L, SUN G. Squeeze-and-excitation networks[C]//*Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Salt Lake City, UT, USA: IEEE, 2018: 7132-7141.
- [13] WOO S, PARK J, LEE J Y, et al. CBAM: Convolutional block attention module[C]//*Proceedings of Computer Vision—ECCV 2018*. Cham, Switzerland: Springer International Publishing, 2018: 3-19.
- [14] SHEN J, CHEN Y, LIU Y, et al. ICAFusion: Iterative cross-attention guided feature fusion for multispectral object detection[J]. *Pattern Recognition*, 2024, 145: 109913.

- [15] XU S, CHEN X, LI H, et al. Airborne small target detection method based on multimodal and adaptive feature fusion[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2024, 62: 5637215.
- [16] WANG A, CHEN H, LIN Z, et al. RepViT: Revisiting mobile CNN from ViT perspective[EB/OL]. (2024-03-14). <https://doi.org/10.48550/arXiv.2307.09283>.
- [17] LIU Z, MAO H, WU C Y, et al. A ConvNet for the 2020s[C]//*Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. New Orleans, LA, USA: IEEE, 2022: 11966-11976.
- [18] DING X, ZHANG X, HAN J, et al. Scaling up your kernels to 31×31 : Revisiting large kernel design in CNNs[C]//*Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. New Orleans, LA, USA: IEEE, 2022: 11953-11965.
- [19] HOWARD A, SANDLER M, CHEN B, et al. Searching for MobileNetV3[C]//*Proceedings of 2019 IEEE/CVF International Conference on Computer Vision (ICCV)*. Seoul, Korea: IEEE, 2019: 1314-1324.
- [20] HAN K, WANG Y, TIAN Q, et al. GhostNet: More features from cheap operations[C]//*Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Seattle, WA, USA: IEEE, 2020: 1577-1586.
- [21] HU S, ZHAO F, LU H, et al. Improving YOLOv7-tiny for infrared and visible light image object detection on drones[J]. *Remote Sensing*, 2023, 15(13): 3214.
- [22] PINHO L S, SOUSA T D, PEREIRA C D, et al. Anomaly detection for PV modules using multi-modal data fusion in aerial inspections[J]. *IEEE Access*, 2025, 13: 88762-88779.
- [23] LIU Z, LIN Y, CAO Y, et al. Swin Transformer: Hierarchical vision Transformer using shifted windows[C]//*Proceedings of 2021 IEEE/CVF International Conference on Computer Vision (ICCV)*. Montreal, QC, Canada: IEEE, 2021: 9992-10002.
- [24] HOU Q, ZHOU D, FENG J. Coordinate attention for efficient mobile network design[C]//*Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Nashville, TN, USA: IEEE, 2021: 13708-13717.
- [25] YANG J, QIU P, ZHANG Y, et al. D-Net: Dynamic large kernel with dynamic feature fusion for volumetric medical image segmentation[J]. *Biomedical Signal Processing and Control*, 2026, 113: 108837.
- [26] LIU J, FAN X, HUANG Z, et al. Target-aware dual adversarial learning and a multi-scenario multi-modality benchmark to fuse infrared and visible for object detection[C]//*Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. New Orleans, LA, USA: IEEE, 2022: 5792-5801.
- [27] KHANAM R, HUSSAIN M. YOLOv11: An overview of the key architectural enhancements[EB/OL]. (2024-10-23). <https://arxiv.org/abs/2410.17725>.
- [28] TIAN Y, YE Q, DOERMANN D. YOLOv12: Attention-centric real-time object detectors[EB/OL]. (2025-02-18). <https://arxiv.org/abs/2502.12524>.
- [29] WANG D, LIU J, LIU R, et al. An interactively reinforced paradigm for joint infrared-visible image fusion and saliency object detection[J]. *Information Fusion*, 2023, 98: 101828.
- [30] TANG L, YUAN J, MA J. Image fusion in the loop of high-level vision tasks: A semantic-aware real-time infrared and visible image fusion network[J]. *Information Fusion*, 2022, 82: 28-42.
- [31] CHEN Y, WANG B, GUO X, et al. DEYOLO: Dual-feature-enhancement YOLO for cross-modality object detection[C]//*Proceedings of Pattern Recognition*. Cham, Switzerland: Springer Nature, 2025: 236-252.
- [32] DONG W, ZHU H, LIN S, et al. Fusion-mamba for cross-modality object detection[J]. *IEEE Transactions on Multimedia*, 2025, 27: 7392-7406.
- [33] ZHAO Z, BAI H, ZHANG J, et al. CDDFuse: Correlation-driven dual-branch feature decomposition for multi-modality image fusion[C]// *Proceedings of 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Vancouver, BC, Canada: IEEE, 2023: 5906-5916.
- [34] XU H, MA J, JIANG J, et al. U2Fusion: A unified unsupervised image fusion network[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2022, 44(1): 502-518.
- [35] MA J, TANG L, FAN F, et al. SwinFusion: Cross-domain long-range learning for general image fusion via swin transformer

[J]. IEEE/CAA Journal of Automatica Sinica, 2022, 9(7): 1200-1217.

[36] SUN Y, CAO B, ZHU P, et al. DetFusion: A detection-driven infrared and visible image fusion network[C]//Proceedings of the 30th ACM International Conference on Multimedia. [S.l.]: ACM, 2022: 4003-4011.

[37] TANG L, YUAN J, ZHANG H, et al. PIAFusion: A progressive infrared and visible image fusion network based on illumination aware[J]. Information Fusion, 2022, 83/84: 79-92.

[38] TANG L, DENG Y, MA Y, et al. SuperFusion: A versatile image registration and fusion network with semantic awareness [J]. IEEE/CAA Journal of Automatica Sinica, 2022, 9(12): 2121-2137.

[39] LI J, CHEN J, LIU J, et al. Learning a graph neural network with cross modality interaction for image fusion[C]//Proceedings of the 31st ACM International Conference on Multimedia. Ottawa, ON, Canada: ACM, 2023: 4471-4479.

[40] CHEN Y, WANG B, ZHU W, et al. RGB-IR YOLO combining modality-specific reconstruction and information integration [C]//Proceedings of 2024 39th Youth Academic Annual Conference of Chinese Association of Automation (YAC). Dalian, China: IEEE, 2024: 2045-2050.

作者简介:



唐小煜(1980-),通信作者,男,副教授,研究方向:机器视觉、智能感知与控制, E-mail: tangxy@scnu.edu.cn。



何慧伊可(2006-),女,本科生,研究方向:计算机视觉与目标检测, E-mail: 202480313959@m.scnu.edu.cn。



黄晓丹(2005-),女,本科生,研究方向:深度学习与目标检测, E-mail: 202480322531@m.scnu.edu.cn。



吴文隶(2005-),男,本科生,研究方向:图像处理、计算机视觉, E-mail: 20238132076@m.scnu.edu.cn。

(编辑:张黄群)

A Feature-Enhanced Lightweight Dual-Modal Detector for UAV-Based Photovoltaic Panel Defect Detection

TANG Xiaoyu^{1,2,3*}, HE Huiyike^{1,3}, HUANG Xiaodan^{1,3}, WU Wenli^{2,3}

(1. School of Materials and New Energy, South China Normal University, Shanwei 516600, China; 2. School of Data Science and Engineering, South China Normal University, Shanwei 516600, China; 3. Xingzhi College, South China Normal University, Shanwei 516600, China)

Abstract: In UAV-based photovoltaic panel defect detection, infrared and visible images provide complementary information on thermal anomalies and surface texture structures. However, local modality misalignment, attenuation of weak responses from tiny defects, and the high computational cost of dual-modal models still limit detection performance and practical deployment. To address these challenges, this study proposes a feature-enhanced lightweight dual-modal detector for UAV-based photovoltaic panel defect detection, termed FLD-DET. Built on an infrared-visible dual-modal detection framework, FLD-DET incorporates a dynamic sparse window (DSWin) module into the backbone feature extraction process. DSWin enhances the retention of weak responses from tiny defects through local convolutional inductive bias, channel compression, and adaptive attention allocation. To mitigate local spatial discrepancies between dual-modal features, a lightweight cross-modal attention (LCoA) mechanism is designed to strengthen the spatial correspondence between infrared and visible features by encoding directional positional information. During the detection feature generation stage, FLD-DET introduces a regional-aware feature aggregation (RAFA) head to enhance the discriminative representation between defect regions and background regions. Experimental results show that FLD-DET achieves a favorable balance between detection performance and inference efficiency on the aerial dual-modal photovoltaic panel defect dataset and demonstrates good recognition capability for tiny defects. FLD-DET also obtains competitive detection results on the M³FD dataset, indicating potential cross-scenario generalization.

Highlights:

1. Construct DSWin to preserve weak responses of tiny photovoltaic defects while reducing computation.
2. Design LCoA to enhance spatial correspondence between infrared thermal anomalies and visible texture features.
3. Introduce RAFA to strengthen region-level discrimination between defect and background features before detection prediction.
4. Validate FLD-DET on an aerial dual-modal photovoltaic defect dataset and the M³FD dataset, demonstrating improved detection performance and cross-scene applicability.

Key words: photovoltaic panel defect detection; dual-modal fusion; tiny defect; lightweight model; deep learning

Foundation item: Guangdong Basic and Applied Basic Research Foundation Project (No.2024A1515012126).

Received: 2026-02-16; **Revised:** 2026-07-03

***Corresponding author, E-mail:** tangxy@scnu.edu.cn.