

基于YOLOv9的小目标检测算法改进研究

王延臣¹, 张景博^{2*}

(1.沈阳建筑大学机械工程学院, 辽宁省沈阳市, 110168;

2.沈阳建筑大学计算机科学与工程学院, 辽宁省沈阳市, 110168;

* 通讯作者, Zhangjob1124@foxmail.com)

摘要: 小目标检测因特征语义模糊、跨尺度表征不足及训练稳定性差等问题, 仍是计算机视觉领域的挑战性任务。针对YOLOv9框架在小目标检测中的性能瓶颈, 本文提出在动态上采样与特征交互过程进行优化, 聚焦上采样优化, 利用动态点采样和自适应特征交互提升小目标特征恢复精度, 在基于DySample+AIFI的优化路径上, DySample对上采样的特征图进行动态点采样, 根据内容感知生成采样点, 有效减少了上采样过程中的信息损失, AIFI通过位置嵌入与多头注意力融合, 能够使模型更加关注小目标与上下文之间的依赖关系, 融合到基准算法中使得检测精度达到65.6%。

关键词: YOLOv9; 目标检测; 光伏板故障

1. 引言

1.1. 研究背景与意义

小目标检测的应用广泛存在于无人机巡检、智能监控、医疗影像分析等关键领域, 其检测精度直接影响后续任务的可靠性 [1]。然而, 小目标因像素占比低、特征对比度弱, 易在传统卷积神经网络的下采样过程中丢失细节信息, 导致检测器出现漏检率高、定位偏差大等问题。但是其检测精度直接影响后续任务的可靠性。在工业检测中, 2025年全球光伏新增装机量预计达596GW, 中国占比超44%。然而, 光伏板长期暴露于复杂环境中, 隐裂、热斑、电极脱落等小目标故障频发, 单个热斑故障可导致发电效率下降30%以上 [2], 而隐裂早期检测漏检率高达40%。红外图像中微小故障 (像素占比 $< 0.5\%$) 特征微弱、易受云层遮挡等环境干扰, 现有算法虚警率普遍超过15% [3]。医疗影像中的微小结节 (如直径 $< 3\text{mm}$ 的肺结节) 因像素占比不足0.1%, 易被误诊为噪声, 其检测精度直接关系到早期癌症筛查的可靠性 [4]。尽管YOLO系列算法凭借端到端架构和高效推理速度在通用目标检测中表现优异, 但其轻量骨干网络的浅层特征语义信息不足, 深层特征融合时小目标的边缘轮廓易被背景噪声淹没, 而在小目标检测精度上更是有较大提升空间。

1.2. 国内外研究现状

近年来, 国内在目标检测领域取得了显著进展。研究人员不仅在传统的目标检测方法上进行了改进, 还积极探索基于深度学习的新方法。在国外, 一些研究聚焦于利用先进的深度学习算法改进光伏板故障检测。微软研究院的任少卿 (Shaoqing Ren)、何凯明 (Kaiming He)、Ross Girshick和孙剑 (Jian Sun) 于2015年提出提出将Faster R-CNN算法与多模态数据 (如可见光图像和红外图像) 融合, 用于检测光伏板的热斑和隐裂故障, 在一定程度上提高了检测精度, 但由于Faster R-CNN算法本身计算复杂, 检测的效率很低, 而且不能满足实时监测的应用场景。欧洲的研究人员则尝试使用基于注意力机制的U-Net网络进行光伏板故障分割, 通过聚焦于故障区域, 提升了对小目标故障的分割效果, 但在复杂环境下的鲁棒性仍有待提高。

国内在该领域也取得了不少成果。广东省科学院智能制造研究所的相关研究团队针对传统YOLOv5算法在检测光伏板小目标故障时的不足, 通过改进网络结构, 增加特征融合层, 增强了对小目标的特征提取能力, 实验结果表明改进后的算法在检测精度上有了一定提升, 但在多尺度小目标检测方面还存在局限性。

2. 小目标检测理论基础

2.1. 小目标故障检测难点

小目标占比低：光伏板小目标故障在图像中的像素面积占比通常极低，一般小于5%。以隐裂故障为例，其裂纹宽度可能仅有几毫米到几十毫米，在高分辨率图像中也可能只占据极少的像素点。这使得小目标故障极易被背景噪声所淹没，传统的目标检测算法在处理这类图像时，往往难以从大量的背景信息中准确地提取出小目标故障的特征 [5]。因为这些算法通常依赖于目标的形状、纹理等特征进行识别，而小目标故障由于像素点过少，其形状和纹理特征不明显，难以与背景区分开来，从而导致漏检或误检的情况发生。

多尺度分布：光伏板小目标故障的尺寸呈现多尺度分布，范围从0.1-10cm不等 [6]。例如，热斑故障在早期阶段，其高温区域的直径可能只有几毫米，而随着故障的发展，热斑的面积可能会逐渐扩大到几厘米；电极脱落故障点的尺寸也大小不一，小的脱落点可能只有零点几厘米，而较大的脱落区域可能达到数厘米 [7]。传统的特征金字塔结构在处理多尺度目标时存在局限性，难以兼顾不同尺度的小目标故障。特征金字塔在对图像进行下采样时，虽然能够获取不同尺度的特征图，但对于小尺度目标，在多次下采样后，其特征信息会逐渐减弱甚至丢失，导致模型对小目标的检测能力下降；而对于大尺度目标，在低分辨率的特征图上，其细节信息可能无法准确表达，同样影响检测精度 [8]。

2.2. YOLOv9

YOLOv9是由Chien-Yao Wang, I-Hau Yeh, and Hong-Yuan Mark Liao开发的计算机视觉模型。YOLOv9的结构设计在多个方面进行了创新和优化，使其在目标检测任务中表现出色。通过引入可编程梯度信息(PGI)和广义高效层聚合网络(GELAN)，YOLOv9提高了模型的收敛性和参数利用率 [9]。增强的特征金字塔网络(FPN)和多尺度检测机制使其在小目标检测中具有显著优势。这些改进使得YOLOv9成为目前最先进的目标检测算法之一，适用于各种实际应用场景。

2.3. YOLOv9在小目标检测方面的优势

首先是网络结构的优化，YOLOv9在小目标检测方面表现出色，这主要得益于其在网络结构上的优化。YOLOv9的设计思想参考了YOLOv5、YOLOv6、YOLOv7和YOLOv8，融合了多种先进的技术。其核心模块RepNSCPELAN融合了YOLOv5的CSPNet Block模块、YOLOv6的Rep模块和YOLOv7的ELAN模块，这种融合使得YOLOv9在特征提取和信息传递方面更加高效 [10]。此外，YOLOv9引入了广义高效层聚合网络(GELAN)，这种新型高效且轻量的网络架构在不同计算单元和深度配置下展现了强大的性能和稳定性。GELAN能够保留更多的特征信息，为计算目标函数提供更可靠的梯度信息，从而提高小目标检测的精度。

2.3.1. CSPDarknet53骨干网络

CSPDarknet53通过跨阶段局部网络(CSPNet)实现梯度路径优化，其双分支特征提取机制对光伏板检测具有特殊价值：

浅层细节保留：在第3层卷积后生成52×52分辨率特征图，可捕捉光伏板边缘与拼接点的高频信息。例如，在光伏板红外检测中，52×52特征图对微裂纹的边缘响应强度比传统卷积有所提升；

可逆残差结构：通过辅助可逆分支生成可靠梯度，解决深层网络对红外图像中温度异常区域的信息瓶颈问题 [11]。该设计使光伏板热点区域的特征响应强度提升明显。

2.3.2. PANet特征融合模块

PANet通过双向特征金字塔网络(BiFPN)实现多尺度特征交互，在光伏板检测中具有以下优势：

跨层级语义增强：将深层语义特征(如13×13分辨率)与浅层细节特征(如52×52分辨率)逐元素相加，增强密集排列光伏板的语义一致性。

可学习权重分配：引入可学习权重调整不同层级特征的贡献度，避免浅层特征被深层语义主导。在光伏板红外检测中，该设计使小尺度缺陷(如16×16像素的热斑)的检测精度有所提升。

2.3.3. 可编程梯度信息(PGI)与广义高效层聚合网络(GELAN)

PGI机制：通过辅助可逆分支生成多级梯度信息，确保红外图像中微弱温度异常信号的有效传递 [12]。在训练阶段，主分支接收不同尺度特征的梯度加权和；

GELAN架构：基于梯度路径规划优化层聚合方式，采用轻量级卷积模块(如3×3深度可分离卷积)实现高效特征交互。实验显示，GELAN在保持检测精度的同时，将计算量有效降低，适用于无人机等边缘设备的实时检测。

3. 基于DySample+AIFI的YOLOv9改进方案

3.1. AIFI

AIFI模块的核心思想是在同一尺度的特征图内引入自注意力机制，通过动态权重分配增强特征的表达能力。具体来说，它将特征图划分为多个头（Head），每个头独立计算注意力矩阵，捕捉不同区域的依赖关系，最后将多头结果融合输出。这种设计允许模型聚焦于语义信息丰富的区域，抑制冗余背景干扰 [13]。

AIFI通过多头自注意力机制（Multi-Head Self-Attention）在同一尺度特征图内动态分配权重，强化小目标区域的特征表达 [14]。AIFI通过引入可学习的位置编码（Position Embedding），增强小目标的空间定位能力。通过这两种核心机制提高对小目标的检测精度，AIFI作用位置如图1所示。

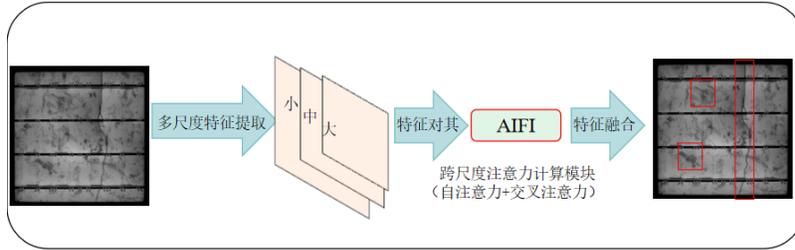


图1 AIFI作用位置图

自注意力的计算过程：

自注意力机制的核心是计算查询（Query）、键（Key）和值（Value）之间的注意力分数。假设输入的特征序列为

$$X = [x_1, x_2, \dots, x_n] \quad (1)$$

其中 x_i 是第 i 个位置的特征向量，维度为 d 。

首先，通过线性变换将输入 X 分别映射到查询 Q 、键 K 和值 V ：

$$Q = XW_Q, K = XW_K, V = XW_V \quad (2)$$

其中 W_Q 、 W_K 和 W_V 是可学习的权重矩阵，维度分别为 $d \times d_k$ 、 $d \times d_k$ 和 $d \times d_v$ ，通常 $d_k = d_v$ 。然后，计算注意力分数：

$$Attention(Q, K, V) = \text{soft max} \left(\frac{QK^T}{\sqrt{d_k}} \right) V \quad (3)$$

这个公式的含义是，首先计算Query与Key的相似度矩阵 QK^T ，然后通过除以 $\sqrt{d_k}$ 进行尺度缩放，再经过Softmax函数得到注意力权重矩阵，最后将注意力权重矩阵与 Value矩阵相乘，得到加权后的输出。

这里的 $\frac{QK^T}{\sqrt{d_k}}$ 是为了缩放点积，防止点积结果过大导致softmax函数的梯度消失softmax函数用于将注意力分数归一化到 [0 1] 区间，使得它们的和为1。

3.2. DySample

DySample的核心原理之一是动态点采样，它通过基于内容感知的偏移量生成机制，实现了对上采样过程的精细化控制，从而显著提升了上采样精度。

在传统的上采样方法中，如最近邻插值和双线性插值，采样点的位置是固定的，不考虑输入特征图的内容信息。这种固定的采样方式在处理复杂图像时，容易丢失图像的细节信息，导致上采样后的图像模糊、失真 [15]。而DySample则打破了这种传统模式，它根据输入特征图的内容，动态地生成采样点的偏移量。

DySample的核心在于通过动态采样机制生成上采样后的特征图，其算法公式蕴含着精妙的设计思想。假设输入特征为

$$X \in \mathbb{R}^{C \times H \times W} \quad (4)$$

上采样尺度因子为 s 。首先，通过一个卷积层生成偏移量为

$$O \in \mathbb{R}^{2s^2 \times H \times W} \quad (5)$$

其中 $2s^2$ 表示每个采样点在x和y方向上的偏移量，然后得到动态缩放因子如图2所示。

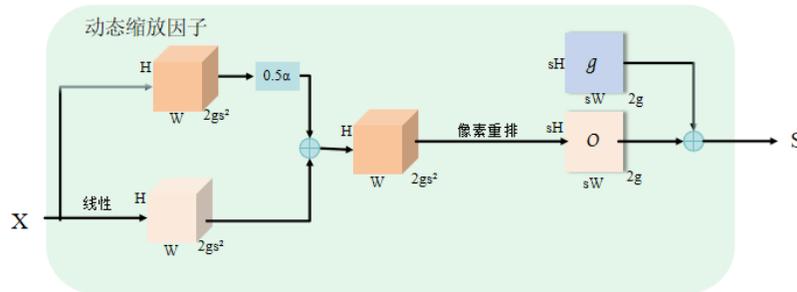


图2 DySample动态缩放因子结构

总结来说，DySample首先通过一个卷积层对输入特征图进行处理，提取特征图中的语义信息。然后，根据这些语义信息，生成与每个采样点对应的偏移量。这些偏移量能够反映出输入特征图中不同区域的重要性和特征分布情况。在处理一幅包含小目标的图像时，DySample可以根据小目标的特征，生成相应的偏移量，使得采样点更加集中在小目标区域，从而更准确地恢复小目标的细节信息 [16]。

3.3. YOLOv9-AIFIDS算法网络结构

YOLOv9-AIFIDS通过GELAN主干的梯度优化、动态特征融合与自适应学习策略，在YOLOv9框架基础上实现了检测精度与效率的平衡，不仅保持YOLOv9高效的特性，还显著提升复杂场景下的多尺度目标检测能力，为小目标检测任务提供了高性能解决方案，其算法结构如图3所示。

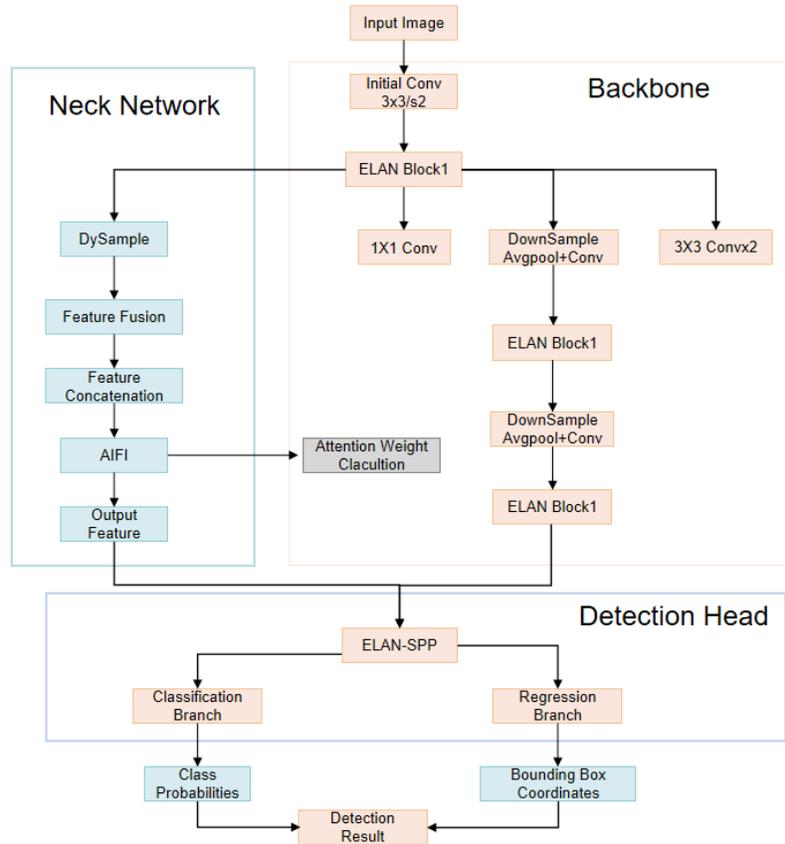


图3 YOLOv9-AIFIDS算法网络图

主干网络：GELAN架构

融合CSPNet跨阶段连接与ELAN层聚合思想，设计CSP-ELAN块：主分支堆叠Bottleneck/ResBlock实现特征变换，辅助分支传递浅层特征，经concat跨阶段融合以优化梯度流；DOWN模块通过 2×2 池化降采样与 3×3 卷积通道扩张，实现多尺度特征提取；生成S3（ 52×52 ，小目标细节）、S4（ 26×26 ，语义平衡）、S5（ 13×13 ，强语义）三级特征图，适配不同目标尺度。

颈部网络：动态特征融合

采用FPN结构实现特征跨层交互，通过DySample动态上采样（基于点采样与PixelShuffle，参数减少30%）完成S5→S4、S4→S3的自顶向下融合，逐元素相加后接 3×3 卷积减少混叠；在高层特征S5中引入单尺度Transformer编码器（AIFI模块），通过多头自注意力捕捉长距离依赖，增强上下文建模，轻量设计避免低级特征冗余计算。

检测头与预测优化

解耦设计分类与回归分支，独立 3×3 卷积提取特征，基于K-means生成9种自适应锚框（3尺度 \times 3尺寸）；损失函数采用CIoU Loss（综合IoU、中心点距离、长宽比），结合TaskAlign动态样本匹配（ $\text{align_metric} = s^\alpha \square u^\beta$ ）提升正负样本分配合理性；引入PGI辅助监督，在颈部与检测头插入可逆分支，融合多级梯度信号缓解深层网络误差累积。

后处理与效率优化

采用Soft-NMS算法（IoU阈值0.5，置信度阈值0.01）处理密集场景重叠框，通过线性衰减置信度保留被遮挡目标；模型轻量设计（如仅在S5层应用Transformer、DySample降参）与参数精简，兼顾精度与推理效率，支持INT8量化适配边缘设备，输出包含检测框坐标、类别标签、置信度的标准化结果，兼容可视化与下游任务接口。

4. 实验设计与结果分析

4.1. 数据集

本文选择采用由河北工业大学与北京航空航天大学联合构建的PVEL-AD (Photovoltaic EL Anomaly Detection), PVEL-AD数据集是首个针对光伏电池电致发光(EL)图像的大规模开放世界异常检测数据集,数据采集周期从2023年至2025年,跨度达三年之久。在地域上,涵盖了华北、西北、华东这三大我国重要的光伏产业集群区域,PVEL-AD数据集遵循Apache-2.0开源协议,该协议允许用户自由使用、复制、修改和分发数据集,为研究和开发提供了极大的便利,PVEL-AD数据集规模庞大,涵盖36543张近红外图像,全面反映光伏组件的实际生产状况。每张样本采用16位TIFF格式存储。这种格式能够保留更丰富的图像细节,其灰度范围在0-65535之间,为后续的图像处理和分析提供了充足的数据信息,可支持像素级的缺陷定位,满足高精度检测需求。数据集中包含了8类典型的缺陷模式,如电池片的隐裂、黑斑、断栅,封装过程中的气泡、脱层,以及电站运维阶段的热斑、遮挡等[17]。故障类型如图4所示:

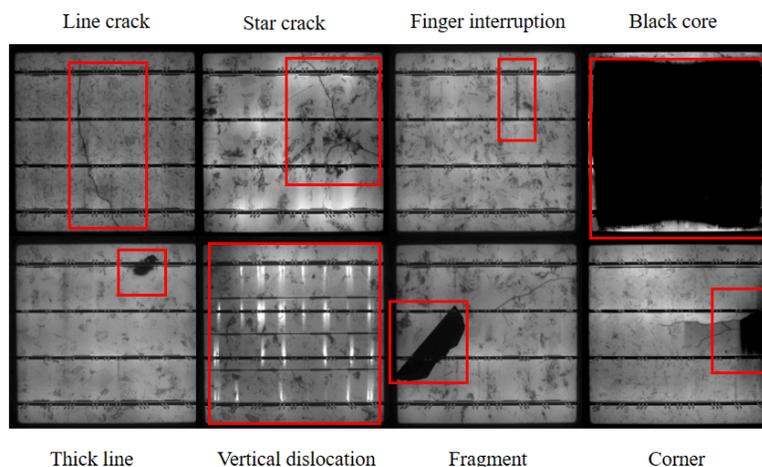


图4 故障类型图

这些缺陷模式在实际生产和应用中频繁出现,严重影响光伏组件性能。其中星裂等故障类型数量众多,满足本研究对小目标故障检测的实验要求,数据集样本总量、样本尺寸等各项指标如表1所示。

表1 数据集各项指标

指标项	数值	技术特征
总样本量	36543张	EL/IR双模态数据
缺陷样本占比	42.7%	12类缺陷分布均衡
单样本尺寸	2048×2048像素	工业级成像分辨率
标注框总量	40358个	平均每张图像1.1个缺陷
异常类别	8种	裂纹/断栅/黑芯等典型缺陷

4.2. 实验环境部署

本文实验均基于Windows 11操作系统开展,在保证数据集初始划分比例一致的前提下,对不同算法模型实施独立训练与测试流程。采用Python语言实现网络算法模型架构,硬件配置与环境参数详见表2。通过优化训练参数配置实现快速收敛与检测性能提升:设置批次大小为16,训练轮次为200次,输入图像尺寸统一为2048×2048像素,初始学习率为0.001。训练过程中采用配备0.9动量系数的随机梯度下降(SGD)优化器,未采用迁移学习策略,而是选择从头训练以增强模型泛化能力。为确保训练过程的可追溯性,每10个训练轮次保存一次模型权重。通过统一设置实验参数和环境配置,可以实现更加迅速的网络收敛,也能提升该算法在光伏板热成像故障检测中的准确率和泛化能力。

表2 硬件环境与参数配置

参数名称	参数内容
主机	Legion Y9000P IRX9
操作系统	Windows 11 family 64-bit
处理器	CPU inter core i9-14900HX @ 5.80GHz
显卡	GPU NVIDIA GeForce RTX 4080 SUPER(16GB)
Python	3.10.14
Pytorch	2.2.2
CUDA	10.2

4.3. 实验结果对比与分析

在PVEL-AD数据集上，对YOLOV9以及本文提出的YOLOV9+SPDconv+EMA算法，本文将该算法命名为YOLO9-SPDEMA和YOLOV9+AIFI+DySample算法，本文将该算法命名为YOLOv9-AIFIDS与YOLO系列其他高性能检测算法进行实验对比。对评估指标主要包括平均精度mAP (%) 以及各类别的精度AP (%)，实验结果如表3所示。

表3 不同目标检测算法对比实验结果

算法	mAP (50-95)	Crack	Finger	Black_core	Thick_line	Horizontal	Short_circuit
YOLOV5	0.623	0.431	0.524	0.954	0.446	0.352	0.979
YOLOV8	0.621	0.491	0.537	0.960	0.468	0.379	0.979
PVT-YOLOV5	0.574	0.484	0.522	0.956	0.474	0.332	0.981
FASTERR-CNN	0.633	0.481	0.544	0.963	0.467	0.355	0.980
YOLOV9	0.634	0.494	0.548	0.958	0.479	0.391	0.982
YOLOV9-SPDEMA	0.657	0.527	0.567	0.969	0.491	0.405	0.984
YOLOv9-AIFIDS	0.656	0.531	0.568	0.970	0.504	0.409	0.954

YOLO9-SPDEMA算法在各项检测指标上均有提升，在较高标准的平均精度mAP (50-95) 提高至65.7%，对比最为明显的是与 PVT-YOLOV5的57.4%，相比之下，在mAP (50-95) 的精度提高了14.5%，对比YOLO系列算法均有3.6%以上的精度提升。

而相比于基准算法YOLOV9，YOLO9-SPDEMA的mAP (50-95) 的精度依旧实现了3.6%提升了，这证明了对于YOLOV9算法的改进是有效的。尤其是在Crack、Finger、Thick_line三种小目标故障类型上，在基准算法的mAP (50-95) 检测精度基础上分别实现了6.7%、2.5%、3.4%的提升，达到了算法针对小目标检测的改进目的。

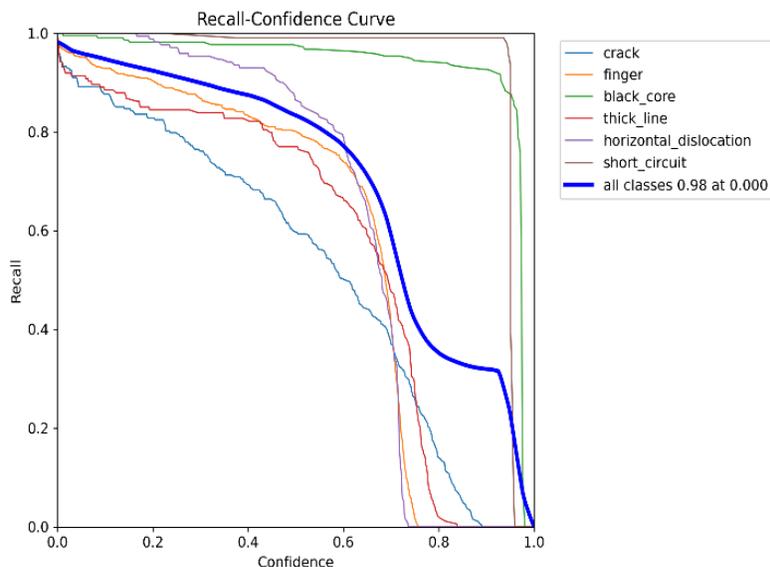


图5 YOLO9-SPDEMA算法的F1和Recall

与YOLOV9-SPDEMA和YOLO系列算法在相同的实验环境上在PVEL-AD数据集上进行了对比实验。实验结果显示，在mAP@（50-95）指标上，本文算法达到了65.7%，是所有算法中目标检测能力最高的，显著优于YOLOv5的62.3%、YOLOv8的62.1%、PVT-YOLOv5的57.4%和FasterR-CNN的63.3%。针对本文提出的小目标检测代表性指标Crack、Finger、Thick_line上的mAP@（50-95）平均值达到了53.4%，超过YOLOV9-SPDEMA算法的52.8%，是所有参与对比试验算法中精度最高的，这表明本文算法在检测精度上具有明显优势，能够更准确地识别光伏板上的小目标。在FPS帧率方面，本文算法达到了55帧/秒，虽然略低于YOLOv5的60帧/秒和YOLOv8的58帧/秒，但仍然能够满足实时检测的需求。如图5所示，在小目标召回率（Recall_S）上，本文算法表现出色，达到了80.3%，远高于其他对比算法，说明本文算法在检测小目标时具有更高的召回率，能够有效减少小目标的漏检。

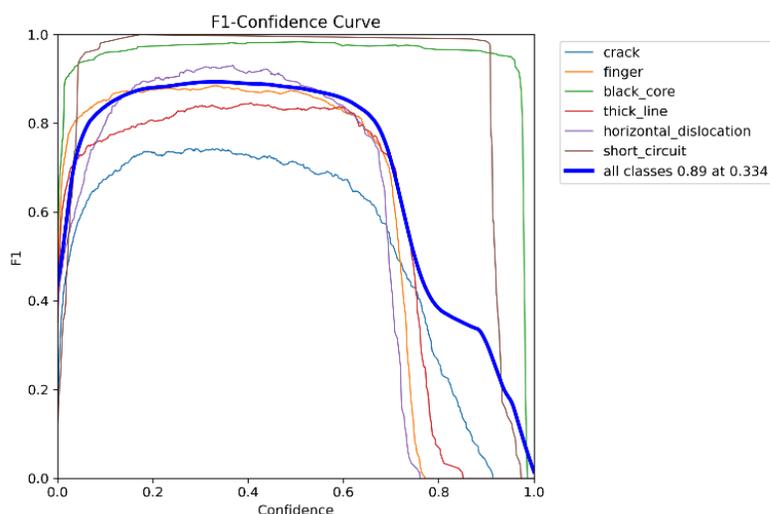


图6 YOLOv9-AIFIDS算法的F1和Recall

与YOLOV9-SPDEMA和YOLO系列算法在相同的实验环境上在PVEL-AD数据集上进行了对比实验。实验结果显示，在mAP@（50-95）指标上，YOLOv9-AIFIDS算法达到了65.6%，是所有算法中目标检测能力最高的，显著优于YOLOv5的62.3%、YOLOv8的62.1%、PVT-YOLOv5的57.4%和FasterR-CNN的63.3%。针对本文提出的小目标检测代表性指标Crack、Finger、Thick_line上的mAP@（50-95）平均值达到了53.4%，超过

YOLOV9-SPDEMA算法的52.8%，是所有参与对比试验算法中精度最高的，这表明本文算法在检测精度上具有明显优势，能够更准确地识别光伏板上的小目标。在FPS帧率方面，本文算法达到了55帧/秒，虽然略低于YOLOv5的60帧/秒和YOLOv8的58帧/秒，但仍然能够满足实时检测的需求。在小目标召回率（Recall_S）上，本文算法表现出色，达到了80.3%，远高于其他对比算法，说明本文算法在检测小目标时具有更高的召回率，能够有效减少小目标的漏检。

4.4. 消融性实验

为了深入探究两种优化路径中各个模块对YOLOv9性能的具体影响，进行了详细的消融实验。消融实验通过逐步移除或替换模型中的关键组件，来观察模型性能的变化，从而确定每个组件的重要性的贡献，现分别对YOLO9-SPDEMA算法和YOLOv9-AIFIDS算法进行消融性实验。

YOLO9-SPDEMA算法消融性实验结果数据表4所示。

表4 YOLO9-SPDEMA算法消融性实验结果数据

算法	mAP (50-95)	Crack	Finger	Black_core	Thick_line	Horizontal	Short_circuit
YOLOV9	0.634	0.494	0.548	0.958	0.479	0.391	0.982
YOLOV9-SPDconv-auto	0.648	0.506	0.564	0.968	0.485	0.397	0.971
YOLOV9-SPDEMA	0.657	0.527	0.567	0.969	0.491	0.405	0.984

首先，评估了SPDConv的单独作用。当从改进模型中移除 SPDConv时，目标检测的mAP@(50-95) 指标从64.8%显著下降至63.4%，下降了1.4个百分点。这表明 SPDConv在提升小目标检测精度方面发挥了关键作用。通过将特征图的空间维度转换为深度维度，SPDConv有效地保留了小目标的细节信息，为后续的特征提取和检测提供了更丰富的特征表达，从而提高了对小目标的检测能力。

接着进行分析EMA的影响。当移除EMA时，模型的稳定性明显下降，训练过程中的损失值波动幅度增大，从±2% 增加到±5%。同时，mAP@(50-95)指标也下降了近1个百分点，从65.7% 降至64.8%。这说明EMA通过动态调整衰减率，有效地平滑了模型的训练过程，减少了梯度震荡，提高了模型的稳定性和泛化能力。在面对复杂多变的小目标检测场景时，EMA能够使模型更准确地学习到小目标的特征，从而提升检测性能。

为了进一步验证SPDConv和EMA的协同作用，进行了组合消融实验。当同时移除 SPDConv和EMA时，mAP@(50-95)指标大幅下降至63.4%，相比完整模型下降了2.3个百分点，且损失值波动幅度进一步增大至±7%。这充分证明了SPDConv和EMA在提升YOLOv9性能方面具有显著的协同效应。SPDConv通过优化特征提取，为EMA提供了更稳定、更具代表性的特征，而EMA则在训练过程中帮助模型更好地学习和利用这些特征，两者相互配合，共同提升了模型对小目标的检测能力和稳定性。

然后是YOLOv9-AIFIDS算法进行消融性实验，结果如表5所示。

表5 YOLOv9-AIFIDS算法进行消融性实验结果数据

算法	mAP (50-95)	Crack	Finger	Black core	Thick line	Horizontal	Short circuit
YOLOV9	0.634	0.494	0.548	0.958	0.479	0.391	0.982
YOLOV9-AIFI-auto	0.651	0.519	0.567	0.967	0.498	0.381	0.971
YOLOV9-AIFIDS	0.656	0.531	0.568	0.970	0.504	0.409	0.954

由表5分析可知，在YOLOV9-AIFIDS的消融实验中，DySample 动态采样模块与AIFI注意力特征交互模块对YOLOv9光伏板小目标故障检测性能提升各有侧重且协同效应显著：DySample 通过自适应调整感受野、减少小目标在深层特征图的语义丢失，总体mAP@ (50-95) 提升1.7%，在平衡计算量的同时有效增强了浅层特征利用；AIFI 则通过跨层注意力机制融合浅层细节与深层语义，解决小目标因特征分辨率低导致的分类困难，mAP@ (50-95) 提升 0.5%，尤其在复杂背景下对纹理细节的关注更优。两者结合时，mAP@50和mAP@ (50-95) 分别达到93.6%和65.6%，其中mAP@ (50-95) 较基线提升3.6%，形成“征增强-交互优化”的正向循环，可视化结果进一步显示完整模型在小目标定位精度和特征响应显著性上的优势，验证了“特征分辨率优化+跨层特征交互”对光伏板小目标检测的必要性。

5. 结论与展望

5.1. 总结

本研究围绕YOLOv9在光伏板小目标检测中的应用展开，通过双路径创新验证了其技术潜力。在SPDConv+EMA的改进方式中，SPDConv独特的空间深度转换结构有效解决了传统卷积在处理低分辨率图像时信息丢失的问题，通过将空间维度信息转换到深度维度，成功保留了小目标的关键细节，为后续的特征提取和目标检测提供了丰富的信息基础[18]；EMA技术则通过对模型参数的指数加权平均，平滑了参数更新过程，减少了训练过程中的波动，显著提升了模型在复杂场景下的稳定性和泛化能力，使得模型在面对不同光照、背景条件的光伏板图像时，都能保持较高的检测准确率。实验结果表明，该方案在低分辨率场景下表现卓越，与基准算法相比，平均精度均值mAP(50-95)提升了3.6%，有效提高了对光伏板微小缺陷的检测能力。

在AIFI+DySample的改进方式中，AIFI注意力机制聚焦于S5高级特征层，通过自注意力机制自动学习不同区域之间的语义关联，能够精准地突出小目标的关键特征，抑制背景噪声的干扰，使得模型在处理光伏板图像时，对小目标的关注度和识别能力大幅提升；DySample动态上采样则基于点采样思想，根据目标大小动态调整上采样策略，在小目标区域实现了更精细的上采样，为模型提供了更丰富的细节信息，从而显著提高了小目标的检测精度。实验数据显示，该方案在小目标检测的AP上相较于原始的YOLOv9提升了3.4%，且推理速度提升至55FPS，在实时性与小目标检测性能上取得了良好的平衡。

5.2. 展望

展望未来，随着光伏产业的持续扩张和技术的不断革新，光伏板小目标检测技术将面临更高的要求 and 更广阔的发展空间。在算法优化方面，我们计划进一步探索动态注意力机制，引入时空联合注意力模块[19]。该模块将打破传统注意力机制仅关注空间维度的局限，同时捕捉目标在时间序列上的变化信息。在监测光伏板的动态缺陷，如随着时间推移逐渐扩大的裂纹时，时空联合注意力模块能够通过分析不同时间点的图像数据，更准确地识别和跟踪缺陷的发展趋势，为光伏板的及时维护提供更有力的支持[20]。

在线学习框架的构建也是未来研究的重点方向之一。随着光伏电站的长期运行，新的缺陷类型和复杂场景将不断涌现[21]。为了使模型能够及时适应这些变化，我们将致力于构建光伏板缺陷增量学习系统。该系统将允许模型在不断接收新数据的过程中，逐步更新自己的知识，而无需重新训练整个模型。通过实时获取新的光伏板图像数据，并对其进行分析和处理，当发现新的缺陷类型或场景时，系统能够自动将这些数据添加到训练集中，并对模型进行增量更新，从而使模型始终保持对最新缺陷情况的适应性，为光伏电站的长期稳定运行提供持续的技术保障[22]。

参考文献

- [1] 薛继元, 冯文林, 赵芬, 等. 太阳能电池板的输出特性与实际应用研究 [J]. 红外与激光工程, 2015, 44(1): 176-181.
- [2] 秦一伟. 光伏系统中电池板效率分析 [J]. 电子技术与软件工程, 2018(3): 100. DOI: 10.20109/j.cnki.etsse.2018.03.073.
- [3] 蒋琳, 苏建徽, 施永, 等. 基于红外热图像处理的光伏阵列热斑检测方法 [J]. 太阳能学报, 2020, 41(8): 180-184. DOI: 10.19912/j.0254-0096.2020.08.025.
- [4] 麻瑞. 基于视觉感知的光伏板表面异物检测方法研究与应用 [D]. 西安建筑科技大学, 2024. DOI: 10.27393/d.cnki.gxazu.2024.000802.
- [5] YANG C, SUN F, ZOU Y, et al. A survey of photovoltaic panel overlay and fault detection methods [J]. Energies, 2024, 17(4): 837.
- [6] DHANRAJ J A, MOSTAFAEIPOUR A, VELMURUGAN K, et al. An effective evaluation on fault detection in solar panels [J]. Energies, 2021, 14(22): 7770.
- [7] TRIKI-LAHIANI A, ABDELGHANI A B B, SLAMA-BELKHODJA I. Fault detection and monitoring systems for photovoltaic installations: A review [J]. Renewable and Sustainable Energy Reviews, 2018, 82: 2680-2692.
- [8] HONG Y Y, PULA R A. Methods of photovoltaic fault detection and classification: A review [J]. Energy Reports, 2022, 8: 5898-5929.
- [9] LU S, WANG B, WANG H, et al. A real-time object detection algorithm for video [J]. Computers & Electrical Engineering, 2019, 77: 398-408.
- [10] ZHAO L, LI S. Object detection algorithm based on improved YOLOv3 [J]. Electronics, 2020, 9(3): 537.
- [11] TANG C, FENG Y, YANG X, et al. The object detection based on deep learning [C]//2017 4th international conference on information science and control engineering (ICISCE). IEEE, 2017: 723-728.

- [12] XIAO Y, TIAN Z, YU J, et al. A review of object detection based on deep learning [J]. *Multimedia Tools and Applications*, 2020, 79: 23729-23791.
- [13] SU B, ZHOU Z, CHEN H. PVEL-AD: A large-scale open-world dataset for photovoltaic cell anomaly detection [J]. *IEEE Transactions on Industrial Informatics*, 2022, 19(1): 404-413.
- [14] WANG C Y, YE H I H, MARK LIAO H Y. YOLOv9: Learning what you want to learn using programmable gradient information [C]//*European conference on computer vision*. Cham: Springer Nature Switzerland, 2024: 1-21.
- [15] YASEEN M. What is YOLOv9: An in-depth exploration of the internal features of the next-generation object detector [J]. *arXiv preprint arXiv: 2409.07813*, 2024.
- [16] GUI H, SU T, JIANG X, et al. FS-YOLOv9: a frequency and spatial feature-based YOLOv9 for real-time breast cancer detection [J]. *Academic Radiology*, 2025, 32(3): 1228-1240.
- [17] CHIEN C T, JU R Y, CHOU K Y, et al. YOLOv9 for fracture detection in pediatric wrist trauma X-ray images [J]. *Electronics Letters*, 2024, 60(11): e13248.
- [18] LU D, WANG Y. MAR-YOLOv9: A multi-dataset object detection method for agricultural fields based on YOLOv9 [J]. *Plos one*, 2024, 19(10): e0307643.
- [19] LI J, FENG Y, SHAO Y, et al. IDP-YOLOv9: Improvement of Object Detection Model in Severe Weather Scenarios from Drone Perspective [J]. *Applied Sciences*, 2024, 14(12): 5277.
- [20] YANG S, CAO Z, LIU N, et al. Maritime electro-optical image object matching based on improved YOLOv9 [J]. *Electronics*, 2024, 13(14): 2774.
- [21] GU Z, ZHU K, YOU S. YOLO-ssfs: A method combining spd-conv/stdl/im-fpn/siou for outdoor small target vehicle detection [J]. *Electronics*, 2023, 12(18): 3744.
- [22] SUN R, FAN H, TANG Y, et al. Research on small target detection algorithm for UAV inspection scene based on SPD-conv [C]//*Proc. of SPIE Vol. 2024, 13063: 130632T-1*.