光电技术及应用

DOI：10.16818/j.issn1001－5868.20260424002

融合空频特征的RT-DETR单晶硅电池缺陷检测算法

胡　铂[1]，代　盼[1*]，孙玉润[2]

（1. 湖州师范大学 信息工程学院，浙江 湖州 313000；2. 中国科学院 苏州纳米技术与纳米仿生研究所，江苏 苏州 215123）

摘　要： 针对单晶硅太阳能电池电致发光（EL）图像中缺陷形态多样、对比度低、尺度差异大等问题，提出一种基于RT-DETR的缺陷检测算法GAE-DETR。首先设计了GDSANet主干网络模块，通过动态空间注意力机制实现缺陷显著区域的自适应聚焦，并引入频域增强前馈网络来捕获EL图像周期性栅格结构的异常变化。进一步提出ESA-PGE轻量化Transformer编码器，采用部分通道自注意力与动态稀疏门控机制，在降低计算复杂度的同时保持全局上下文建模能力。最终构建DHAF双层次注意力特征融合模块，实现了跨尺度特征的精细化整合。实验结果表明，在自建的单晶硅太阳能电池片EL图像数据集上，GAE-DETR的mAP@50达到93.6%、mAP@50-95达到71.2%，相比基线算法原始RT-DETR分别提升5.2%和8.9%，参数量减少17.4%，计算量降低19.2%，实现了检测精度与计算效率的双重优化。

关键词： 缺陷检测；单晶硅太阳能电池；RT-DETR；注意力机制；频域增强；特征融合

中图分类号： TP391.4　**文章编号：** 1001－5868(2026)04－0773－07

Defect Detection Algorithm for Monocrystalline Silicon Solar Cell EL Images Based on Improved RT-DETR

HU Bo[1], DAI Pan[1], SUN Yurun[2]

（1. School of Information Engineering, Huzhou Normal University, Huzhou 313000, CHN; 2. Suzhou Institute of Nano-Tech and Nano Bionics, Chinese Academy of Sciences, Suzhou 215123, CHN）

Abstract: To overcome the challenges of diverse defect morphologies, low contrast, and large-scale variations in the electroluminescence (EL) images of monocrystalline silicon solar cells, this paper presents an improved RT-DETR-based defect detection algorithm, namely GAE-DETR. First, a GDSANet backbone module is designed that adaptively focuses on salient defect regions through a dynamic spatial attention mechanism and incorporates a frequency-domain-enhanced feed-forward network to capture abnormal changes in the periodic grid structure of EL images. Second, an ESA-PGE lightweight Transformer encoder is proposed, which uses partial channel self-attention and dynamic sparse gating mechanisms to reduce computational complexity while preserving global context modeling capability. Furthermore, a DHAF dual-level attention feature fusion module is constructed to achieve fine-grained integration of cross-scale features. Experimental results on a self-built monocrystalline silicon solar cell EL image dataset show that GAE-DETR achieves 93.6% mAP@50 and 71.2% mAP@50-95, which are 5.2 and 8.9 percentage points higher than those of the baseline model, respectively. Meanwhile, the number of parameters is reduced by 17.4%, and the computational complexity (FLOPs) is reduced by 19.2%, achieving dual optimization of detection accuracy and computational efficiency.

收稿日期：2026－04－25.

基金项目：湖州市科技计划项目（2025GY009）.

*通信作者：代　盼

Key words: defect detection; monocrystalline silicon solar cell; RT-DETR; attention mechanism; frequency-domain enhancement; feature fusion

0 引言

太阳能光伏产业作为全球清洁能源转型的核心驱动力,其产品质量直接关系到光伏发电系统的长期稳定性与经济效益。单晶硅太阳能电池片在生产过程中不可避免地会产生裂痕、暗斑、瑕疵等多种类型的缺陷,这些缺陷将显著降低电池的光电转换效率并缩短使用寿命。电致发光(EL)检测技术因其非接触、高灵敏度的特性,已成为工业生产中电池片缺陷检测的主流手段。然而,EL图像中的缺陷往往呈现形态多样、对比度低、尺度差异大等特点,传统的人工目视检测方式不仅效率低下,且容易受到主观因素的影响,难以满足现代化产线对高通量、高一致性质量检测的需求。

近年来,基于深度学习的目标检测算法在工业缺陷检测领域展现出强大的潜力。以YOLO系列为代表的单阶段检测器凭借其高效的推理速度被广泛应用于在线检测场景[1-3],但其在小目标和低对比度缺陷的检测精度上仍有提升空间。Faster R-CNN等两阶段检测器虽然在精度上表现较好[4],但较高的计算开销限制了其在实时检测中的应用。DETR作为一种端到端的目标检测范式[5],通过全局注意力机制有效建模特征间的长程依赖关系,在处理复杂背景下的目标检测任务时展现出独特优势。RT-DETR在DETR基础上引入高效的混合编码器设计,在保持较高检测精度的同时显著提升了推理速度,为实时检测应用提供了新的技术路线[6]。

然而,将RT-DETR直接应用于单晶硅太阳能电池片EL图像缺陷检测仍面临若干挑战。原始ResNet主干网络的标准卷积操作仅能在固定局部感受野内进行特征聚合,难以适应EL图像中缺陷形态的高度多样性[7]。而且标准多头自注意力机制的二次复杂度在处理高分辨率特征图时带来巨大计算负担[8],简单的通道拼接融合策略无法有效对齐深层语义特征与浅层细节特征间的分布差异[9]。

针对上述问题,本文提出一种GAE-DETR的改进RT-DETR缺陷检测算法,从特征提取、特征编码和特征融合三个维度进行系统性优化。首先设计GDSANet主干网络模块,融合动态空间注意力与频域增强前馈网络,实现空间-频域双域特征增强。随后提出ESA-PGE轻量化Transformer编码器,通过部分通道自注意力与动态稀疏门控降低计算复杂度。最终构建DHAF双层次注意力融合模块,借助语义提示向量实现任务导向的跨尺度特征整合。在自建数据集上的实验结果,初步表明所提方法的有效性。

1 相关工作

1.1 太阳能电池缺陷检测

太阳能电池缺陷检测是光伏制造质量控制的关键环节。早期方法主要基于传统图像处理技术,如边缘检测、阈值分割和形态学运算等[10],但这类方法对成像条件敏感且泛化能力有限。随着深度学习的快速发展,基于卷积神经网络的方法逐渐成为主流。Chen等[11]利用改进的YOLOv5实现了光伏电池EL图像中多类型缺陷的自动检测。Su等[12]提出融合注意力机制的Faster R-CNN变体,提升了对微裂纹的检测灵敏度。然而,现有方法在处理EL图像特有的低对比度和周期性栅格背景干扰方面仍存在不足。

1.2 DETR系列检测器

DETR开创性地将Transformer架构引入目标检测领域,通过二分匹配实现端到端的集合预测,消除了传统检测器对锚框和非极大值抑制的依赖。Deformable DETR[13]通过可变形注意力机制加速了收敛。DINO[14]引入对比去噪训练和混合查询选择策略进一步提升了性能。RT-DETR提出高效混合编码器和IoU感知查询选择机制,首次实现了DETR架构的实时检测,为工业在线检测应用提供了新的可能性。

1.3 注意力机制与特征融合

注意力机制能够自适应地聚焦于输入中的重要信息。SE-Net[15]通过通道注意力实现特征校准,CBAM[16]联合通道与空间注意力提升特征表达。在特征融合方面,FPN[17]和PANet[9]分别实现了自顶向下和双向的多尺度特征传递。然而,现有融合方法多采用简单的元素级运算或通道拼接策略,缺乏对跨尺度语义差异的显式建模,限制了多尺度特征的充分利用。

2　GAE-DETR 检测算法

2.1　整体网络结构

针对单晶硅太阳能电池 EL 图像缺陷检测的特殊需求，本文提出一种 GAE-DETR 的目标检测算法，如图 1 所示。从特征提取、特征编码和特征融合三个层面对原始 RT-DETR 进行系统性改进。

首先，采用 GDSANet 模块重构主干网络，该模块内部集成动态空间注意力（DSA）与频域增强前馈网络（FEDF）两个核心组件，通过空间-频域双域特征增强机制有效捕获 EL 图像中的判别性缺陷信息。其次，提出 ESA-PGE 轻量化 Transformer 编码器替代原始 AIFI 模块，通过部分通道自注意力与动态稀疏门控机制，在大幅降低计算开销的同时保持全局上下文建模能力。最后，设计 DHAF 双层次注意力特征融合模块替代原始 CCFM 中的通道拼接操作，通过双尺度局部-全局注意力与可学习语义提示向量实现跨尺度特征的精细化整合。三个模块沿特征传递路径依次作用，最终融合特征经 IoU-aware Query Selection 机制筛选后送入 Decoder 完成端到端检测。

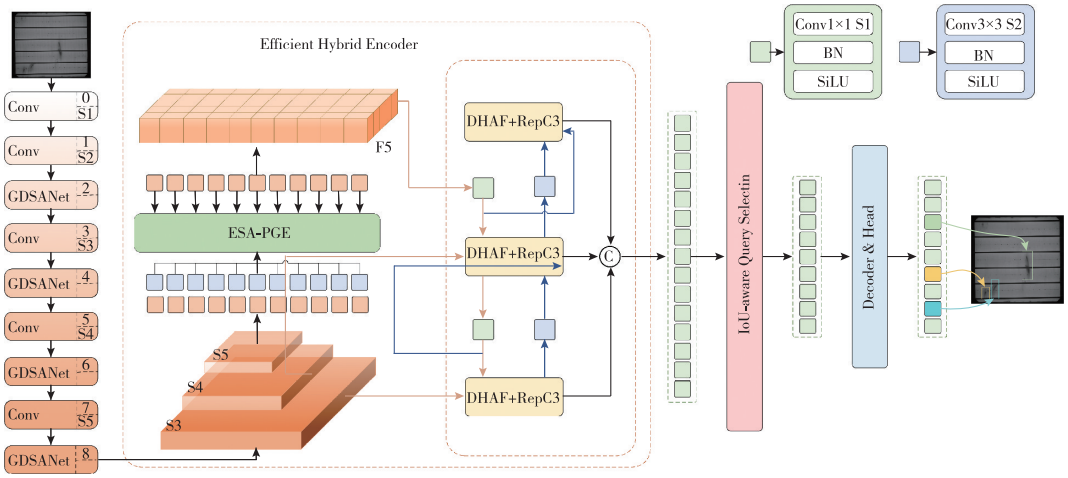


图 1　GAE-DETR 网络模型结构图

2.2　GDSANet 主干网络模块

原始 RT-DETR 采用基于 ResNet 架构的主干网络，其标准 3×3 卷积操作仅能在固定局部感受野内进行特征聚合，难以适应 EL 图像中缺陷形态的高度多样性，且无法充分挖掘频域空间中蕴含的判别性信息。

为此，本算法提出 GDSANet 特征提取模块。GDSANet 继承了 C2f 的跨阶段部分连接思想，通过密集特征复用促进梯度有效传播。如图 2 所示，对于输入特征图 $\boldsymbol{X}\in\mathbb{R}^{C\times H\times W}$，模块通过输入投影卷积将通道维度扩展至 $2c$（其中 $c=\lfloor C\cdot e\rfloor$ 为隐藏通道数，e 为扩展系数），沿通道维度将特征均分为两个并行分支。

其中，第一分支 $\boldsymbol{X}_1$ 作为恒等路径直接参与后续的特征融合，保留原始输入的底层信息；第二分支 $\boldsymbol{X}_2$ 则流经 n 个串联的 GDSA Block 进行深度非线性变换，逐级提取高层语义特征。各计算阶段的输出沿通道维度进行拼接聚合，最终通过输出投影卷积实现跨层特征的融合与通道维度的恢复。上述过程的数学表达式为：

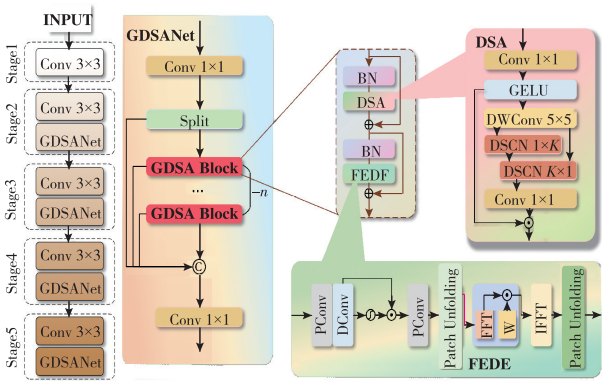


图 2　GDSANet 模块结构图

$$Y=\mathcal{W}_{\text{out}}\left(\oplus_{i=0}^{n+1}F_i\right),\quad F_i=\begin{cases}\boldsymbol{X}_1 & i=0\\ \boldsymbol{X}_2 & i=1\\ \mathcal{G}_{i-1}(F_{i-1}) & i\geqslant 2\end{cases}\tag{1}$$

式中，$[\boldsymbol{X}_1,\boldsymbol{X}_2]=\text{Split}(\mathcal{W}_{\text{in}}(\text{X}))$ 表示输入投影与特征分割的复合操作，$\mathcal{G}_i(\cdot)$ 表示第 i 个 GDSA Block 的映射函数，$\oplus$ 表示沿通道维度的拼接操作，$\mathcal{W}_{\text{in}}$ 与 $\mathcal{W}_{\text{out}}$ 分别对应输入和输出的 1×1 卷积投影矩阵。

GDSA Block 作为 GDSANet 的核心计算单元，

采用Pre-Norm残差连接范式，内部串联DSA分支与FEDF分支，并引入可学习的层尺度（Layer Scale）参数实现训练动态的精细调控。GDSA Block的前向传播过程可形式化为复合变换：

$$Z^{\text{out}}=Z+\gamma_2\odot\Phi_{\text{FEDF}}\times\left\{\text{BN}\left\{Z+\gamma_1\odot\Phi_{\text{DSA}}\left[\text{BN}(Z)\right]\right\}\right\} \tag{2}$$

式中，$\Phi_{\text{FEDF}}(\cdot)$与$\Phi_{\text{DSA}}(\cdot)$分别表示FEDF与DSA的映射函数，BN$(\cdot)$为批归一化操作，$\gamma_1,\gamma_2\in\mathbb{R}^C$为可学习的层尺度参数向量。

DSA模块通过动态门控机制实现对显著区域的自适应聚焦：首先输入特征经1×1逐点卷积进行通道混合并施加GELU激活引入非线性，随后通过5×5和7×7双尺度深度可分离卷积组进行多尺度空间上下文聚合，最后经另一1×1卷积恢复通道维度并建立残差连接。大核卷积的引入使DSA能够在单层内建立较大范围的空间关联。

太阳能电池的规则栅格结构在频域空间中表现为特定频率处的能量峰值，缺陷的存在会导致周期性规律的局部破坏。基于此观察，FEDF模块将频域滤波操作嵌入前馈网络框架。空间域阶段采用1×1卷积投影、3×3深度卷积与分裂-门控（Split-Gate）机制进行特征调制。

频域增强阶段将特征图划分为$p\times p$的非重叠图像块，对每个块施加2D-DFT获得局部频谱表示，通过可学习的频域权重矩阵进行自适应频率调制，再经逆傅里叶变换重建空间域特征。

DSA与FEDF协同工作，形成空间-频率双域的全面特征刻画能力。

2.3 ESA-PGE轻量化编码器

原始RT-DETR的AIFI模块采用标准多头自注意力机制实现尺度内特征全局交互，其$O(N^2)$计算复杂度在处理高分辨率特征图时带来显著计算负担。此外，多头注意力对所有通道进行同质化处理，忽视了通道间语义重要性差异。

针对上述问题，本算法提出ESA-PGE编码器，如图3所示。ESA-PGE包含两项关键创新：部分通道自注意力（PCSA）机制与动态稀疏门控（DSG）机制。

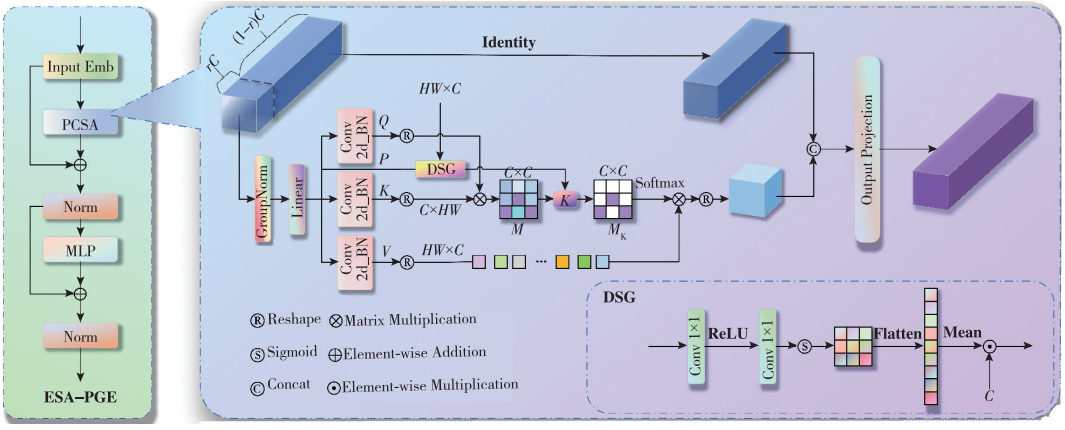


图3 ESA-PGE模块结构图

PCSA将输入特征沿通道维度显式分割为激活通道$\boldsymbol{X}_a\in R^{(C_p\times H\times W)}$与保留通道$\boldsymbol{X}_r\in R^{[(C-C_p)\times H\times W]}$两个功能互补的子空间。在激活通道子空间内采用单头自注意力架构，对$\boldsymbol{X}_a$施加分组归一化后通过线性投影生成Query，Key，Value三元组。保留通道直接旁路传递，避免了对所有通道进行统一注意力计算的冗余开销。

DSG机制与采用固定稀疏度的传统方法不同，DSG通过轻量级门控网络根据输入内容动态预测最优的稀疏保留比例，从而实现样本自适应的注意力聚焦。门控网络$\mathcal{G}(\cdot)$以原始输入特征$\boldsymbol{X}$为条件，输出一个标量值$\rho\in(0,1)$表征当前样本的激活比例：

$$\rho=\sigma\left\{\boldsymbol{W}_2\cdot\text{ReLU}\left[\boldsymbol{W}_1\cdot\text{GAP}(\boldsymbol{X})\right]\right\} \tag{3}$$

式中，GAP$(\cdot)$表示全局平均池化操作，$\boldsymbol{W}_1\in\mathbb{R}^{(C/2)\times C}$和$\boldsymbol{W}_2\in\mathbb{R}^{1\times(C/2)}$为门控网络的可学习参数，$\sigma(\cdot)$为Sigmoid激活函数。动态稀疏度$k$由$\rho$与总位置数$N$共同确定，即$k=\lfloor\rho\cdot N\rfloor$。

获得动态k值后，对注意力矩阵的每一行执行Top-k选择操作，仅保留注意力得分最高的k个位置，其余位置被掩码为负无穷，以在Softmax归一化

后趋近于零。

稀疏掩码操作可形式化为：

$$\mathcal{M}_{\text{sparse}}(A)_{ij}=\begin{cases}A_{ij} & \text{if } j\in \text{TopK}_i(A,k)\\ -\infty & \text{otherwise}\end{cases} \tag{4}$$

其中，$\text{TopK}_i(A,k)$为返回矩阵 A 第 i 行中得分最高的 k 个列索引集合。这种动态稀疏策略使得模型能够根据输入图像的具体内容自适应调整注意力的聚焦范围：对于缺陷分布稀疏的样本，门控网络倾向于预测较低的 ρ 值以实现更精准的定位；对于缺陷分布密集或形态复杂的样本，则预测较高的 ρ 值以保留更广泛的上下文信息。

2.4　DHAF 双层次注意力融合模块

原始 RT-DETR 的 CCFM 结构在跨尺度特征融合时采用简单通道拼接操作，深层语义特征与浅层细节特征在表征空间中存在显著分布差异，缺乏有效对齐机制将导致融合结果中的语义冲突与信息冗余。

因此，本算法提出 DHAF 模块替代通道拼接操作，通过双层次注意力机制实现自适应特征融合。DHAF 接收相邻尺度的两路特征图 $\boldsymbol{X}_1$ 和 $\boldsymbol{X}_2$ 作为输入，部署于 CCFM 的四个关键融合节点，如图 4 所示。

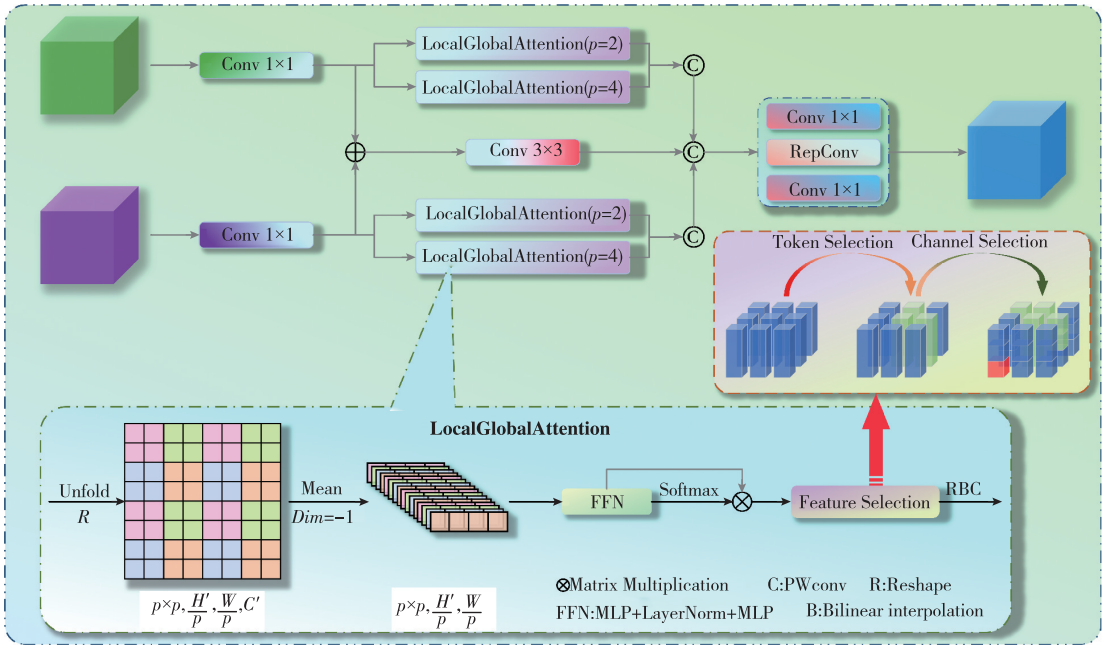


图 4　DHAF 模块结构图

核心组件为局部-全局注意力（LGA）模块。给定输入特征 Z 和图像块尺寸 P，LGA 将特征图划分为 N_p 个非重叠的 $P\times P$ 图像块（其中 $N_p=HW/P^2$），每个块内特征经通道维度平均池化形成紧凑描述子，通过含层归一化的两层 MLP 映射至高级语义空间。

创新性引入的可学习语义提示向量 $\boldsymbol{p}\in R^C$ 在训练过程中自动习得缺陷检测任务的语义原型表示，通过余弦相似度计算生成软掩码，实现任务导向的特征筛选。DHAF 采用 $P=2$ 和 $P=4$ 的双尺度设计，小尺度精准捕获局部细小缺陷边缘细节，大尺度完整建模区域性缺陷空间分布。

多路径聚合模块接收三路输入：$\boldsymbol{X}_1$ 的双尺度注意力拼接特征、$\boldsymbol{X}_2$ 的双尺度注意力拼接特征和基础路径输出。三路特征通道拼接后依次执行 1×1 通道压缩、重参数化卷积特征精炼与 1×1 输出投影。

3　实验与结果分析

3.1　数据集与实验设置

实验采用的单晶硅太阳能电池片 EL 图像数据集由公开网络资源获取的开源图像库和实验室自主采集样本构成，共计 6 500 幅图像（训练/验证/测试集按 8∶1∶1 划分）。其中合格样本 1 821 幅，含缺陷样本 4 679 幅，图像分辨率统一为 1 088×1 088 像素。缺陷类型划分为裂痕（Crack）、暗斑（Dark）和瑕疵（Defect）三类，如图 5 所示，使用 LabelImg 软件以矩形边界框方式进行标注。

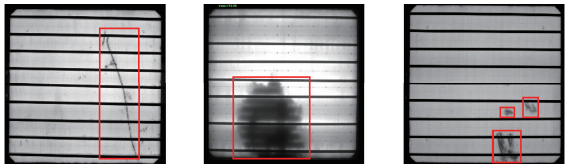


图5 缺陷检测任务数据集包含的典型缺陷

实验环境配置如下：操作系统 Windows 10，处理器 Intel Core i9-10900X，32 GB 内存，NVIDIA GeForce RTX 3090 显卡；编程语言 Python 3.9，深度学习框架 PyTorch；训练超参数设置批处理大小为 8，训练轮次 300，优化器 AdamW，初始学习率 0.000 1。

评价指标包括精确率（P）、召回率（R）、mAP@50 和 mAP@50-95 用于衡量检测性能，参数量（Params）和浮点运算次数（GFLOPs）用于衡量模型复杂度。

3.2 消融实验

为验证各改进模块的有效性，在单晶硅太阳能电池 EL 图像数据集上进行消融实验，以基线算法 RT-DETR 模型为标准，结果如表 1 所示。

表1 消融实验结果

Methods	GDSANet	ESA-PGE	DHAF	*P*/%	*R*/%	*mAP*@50/%	*mAP*@50-95/%	*GFLOPs*/G	*Params*/M
Base				87.2	84.6	88.4	62.3	103.5	42.1
2	√			89.4	86.8	88.7	65.8	98.2	38.6
3		√		88.1	85.9	89.6	64.1	86.4	35.8
4			√	88.6	86.2	90.1	64.9	105.8	43.7
5	√	√		90.2	87.5	91.4	67.2	81.3	32.4
6	√		√	90.8	88.1	92.0	68.4	100.5	40.2
7		√	√	89.5	87.0	91.0	66.5	88.7	37.4
Ours	√	√	√	92.3	89.7	93.6	71.2	83.6	34.8

实验结果表明，三种改进模块均能有效提升模型性能。单独引入 GDSANet 使 *mAP*@50-95 提升 3.5%，这得益于空间-频域双域特征增强机制对 EL 图像周期性栅格异常的敏感捕获。ESA-PGE 通过部分通道自注意力和动态稀疏门控机制，在计算量大幅降低 17.1G 的同时实现 *mAP*@50 提升 1.2%。DHAF 借助双层次注意力融合有效解决跨尺度语义冲突，mAP@50-95 提升 2.6%。两两组合实验验证了模块间的协同增益效应。最终本文完整模型相比基线取得 5.2% 和 8.9% 的 mAP@50 和 mAP@50-95 提升，参数量减少 17.4%，计算量降低 19.2%，实现了检测精度与计算效率的双重优化。

3.3 对比实验

基线算法为原始 RT-DETR 模型，通用算法选取 YOLO 系列、Faster-R-CNN、SSD 等主流目标检测模型，所有模型采用统一训练参数，保证对比公平性。为全面评估 GAE-DETR 的性能优势，将其与当前主流目标检测模型进行系统性对比，结果如表 2 和表 3 所示。

表2 不同检测模型对比结果

Model	*P*/%	*R*/%	*mAP*@50/%	*mAP*@50-95/%	*GFLOPs*/G	*Params*/M
YOLOv5m	86.8	83.9	88.6	64.2	64	25
YOLOv8m	87.5	84.2	89.1	64.8	78.7	25.8
YOLOv9m	88.9	83.7	89.5	65.1	77	20.1
YOLOv10m	86.7	85.1	89.9	66.0	58.9	15.3
YOLOv11m	88.4	83.8	89.7	65.5	67.6	20.2
SSD	83.2	74.6	82.1	49.8	271	23.7
Faster R-CNN	86.9	79.4	86.2	61.7	169	41.3
GAE-DETR	92.3	89.7	93.6	71.2	83.6	34.8

表3 DETR 系列模型对比结果

Model	*P*/%	*R*/%	*mAP*@50/%	*mAP*@50-95/%	*GFLOPs*/G	*Params*/M
RT-DETR-L	87.9	85.7	89.8	63.8	103.4	31.9
RT-DETR-r50	89.1	86.3	90.8	65.4	129.5	41.9
RT-DETR-r34	88.6	85.0	90.3	64.9	88.8	31.1
RT-DETR-r18	86.5	83.4	88.5	62.1	56.9	19.8
GAE-DETR	92.3	89.7	93.6	71.2	83.6	34.8

实验结果表明，GAE-DETR 在检测精度方面全面领先。与 YOLO 系列经典模型 YOLOv11m 相比，GAE-DETR 在 mAP@50 和 mAP@50-95 上分别提升 3.9% 和 5.7%，验证了 Transformer 架构在 EL 图像复杂缺陷检测场景中的优势。与两阶段代表 Faster R-CNN 相比，mAP@50 提升 7.4%，计算量降低 50.5%。

在 DETR 系列内部对比中，GAE-DETR 相较于 RT-DETR-r50 在 mAP@50-95 上提升 5.8%，计算量降低 35.4%，参数量减少 17.0%，充分验证了三项改进在提升检测性能方面的有效性。

3.4 可视化分析

为直观展示 GAE-DETR 的检测能力，采用 Grad-CAM 技术绘制热力图进行可视化分析。如图 6 所示，与 RT-DETR 和 YOLOv11n 相比，GAE-DETR 在缺陷目标区域呈现出更为准确的高亮响应，表明模型能够更精准地聚焦于缺陷关键特征。特别是在微裂纹和隐裂等细小缺陷区域，GDSANet 的频域增强机制使模型敏感捕获周期性栅格结构

的局部破坏。而DHAF的双层次注意力机制确保了多尺度缺陷特征的有效整合，共同提升了模型对缺陷区域的精准定位能力。

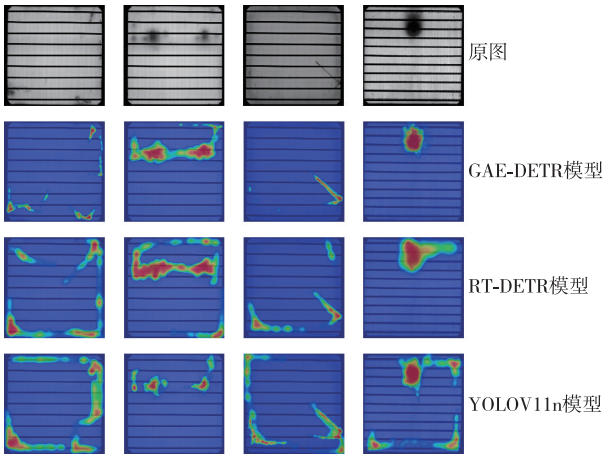


图6　不同算法的可视化检测结果

4　总结

针对单晶硅太阳能电池EL图像缺陷检测中缺陷形态多样、对比度低、尺度差异大等难题，本文提出一种基于改进RT-DETR的GAE-DETR检测算法。通过设计GDSANet主干网络，ESA-PGE轻量化编码器，以及DHAF双层次注意力融合模块实现精细化跨尺度特征整合，三个模块协同工作形成了面向EL图像缺陷检测的全方位优化框架。在自建数据集上，GAE-DETR取得了93.6%的mAP@50和71.2%的mAP@50-95，相比基线模型分别提升5.2和8.9个百分点，同时参数量减少17.4%、计算量降低19.2%。

参考文献：

[1] Redmon J, Divvala S, Girshick R, et al. You only look once: unified, real-time object detection[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition, 2016: 779-788.

[3] Wang C Y, Yeh I H, Mark Liao H Y. YOLOv9: learning what you want toLearn using programmable gradient information[C]//Computer Vision - ECCV, 2025: 1-21.

[4] 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.

[5] Carion N, Massa F, Synnaeve G, et al. End-to-end object detection with transformers[C]//Computer Vision - ECCV, 2020: 213-229.

[6] Zhao Y, Lü W, Xu S, et al. DETRs beat YOLOs on real-time object detection[C]//2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2024: 16965-16974.

[7] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition, 2016: 770-778.

[8] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need[C]//Advances in neural information processing systems. 2017, 30: 5998-6008.

[9] Liu S, Qi L, Qin H, et al. Path aggregation network for instance segmentation[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018: 8759-8768.

[10] Tsai D M, Wu S C, Li W C. Defect detection of solar cells in electroluminescence images using Fourier image reconstruction[J]. Solar Energy Materials and Solar Cells, 2012, 99: 250-262.

[11] Chen H, Pang Y, Hu Q, et al. Solar cell surface defect inspection based on multiscale deep convolutional network[J]. Journal of Visual Communication and Image Representation, 2020, 73: 102897.

[12] Su B, Chen H, Zhou Z. BAF-detector: an efficient CNN-based detector for photovoltaic cell defect detection[J]. IEEE Transactions on Industrial Electronics, 2022, 69(3): 3161-3171.

[13] Zhu X, Su W, Lu L, et al. Deformable DETR: deformable transformers for end-to-end object detection [C]// International Conference on Learning Representations, 2021: 1-16.

[14] Zhang H, Li F, Liu S, et al. DINO: DETR with improved denoising anchor boxes for end-to-end object detection[C]// International Conference on Learning Representations, 2023: 1-23.

[15] Hu J, Shen L, Sun G. Squeeze-and-excitation networks[C]// 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018: 7132-7141.

[16] Woo S, Park J, Lee J Y, et al. CBAM: convolutional block attention module[C]//Computer Vision - ECCV, 2018: 3-19.

[17] Lin T Y, Dollár P, Girshick R, et al. Feature pyramid networks for object detection[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition, 2017: 936-944.

作者简介：

胡　铂(1999—)，男，甘肃省天水市人，硕士研究生，主要研究领域为机器视觉；

代　盼(1987—)，女，安徽宿州市人，博士，副教授，主要研究领域为光电器件以及人工智能在光电器件的应用。