多尺度异构特征协同的棉花叶片病害检测模型

沈学利, 张 越*, 金海波, 张旭旭

(辽宁工程技术大学软件学院,辽宁葫芦岛 125105,中国)

摘 要: [**目的/意义**] 面对自然田间场景中背景干扰强、病斑多尺度分布,以及边缘设备算力受限等多重挑战,现有的农业计算机视觉应用亟需高效的解决方案。基于改进 RT-DETR (Real-Time Detection Transformer) 架构,提出了多尺度异构特征协同的棉花叶片病害检测模型 (Multi-Scale Heterogeneous Synergistic Feature DETR, MHSF-DETR)。[**方法**] 其核心在于重构了特征提取与融合的内在机制。该模型构建了分层上下文选择感知主干网络,在浅层特征阶段嵌入微宏观空间上下文注意力,防止微小病斑细节在深层网络中被淹没;在深层特征阶段集成竞争性选择融合模块,提升语义定位准确性;在特征融合层面,引入可学习加权上下文融合模块,通过动态学习权重解决层级间的语义错位,并结合边界感知重构机制,以反射填充和部分卷积策略有效抑制了边缘伪影,从而在保证特征鲁棒性的同时削减计算开销。[**结果和讨论**] 消融实验验证了各模块的结构协同效应。与基准模型相比,该模型参数量与计算量分别降低 22.42% 和 13.29%,平均精度均值 (mAP50) 提高了 3.2 个百分点。与主流算法横向对比,该模型实现精度与效率的更优平衡;在外部数据集的泛化测试下,其平均精度均值分别可达 80.3% 与 86.9%,跨域检测精度优于同规模主流算法。此外,同等条件的边缘仿真测试中,该模型推理速度最优,单帧耗时约 1454 ms,具备低功耗部署潜力。上述结果证实,该模型的分层特征提取与动态融合机制的协同作用效果显著,架构层面的优化可在低算力约束下实现精度与效率的统一。[**结论**] 本研究提出的模型 MHSF-DETR 能够有效执行复杂场景下的低功耗检测任务,为构建高效的田间实时监测系统提供技术支撑。

关键词: 棉花病害检测;RT-DETR;轻量化模型;注意力机制;特征融合

中图分类号: TP391.4;S435 **文献标识码:** A **文章编号:** SA202601027

引用格式:沈学利, 张越, 金海波, 张旭旭 . 多尺度异构特征协同的棉花叶片病害检测模型[J]. 智慧农业(中英文), 2026, 8(4): 204-216. DOI: 10.12133/j.smartag.SA202601027

SHEN Xueli, ZHANG Yue, JIN Haibo, ZHANG Xuxu. Multi-Scale Heterogeneous Feature Synergistic Model for Cotton Leaf Disease Detection[J]. Smart Agriculture, 2026, 8(4): 204-216. DOI: 10.12133/j.smartag.SA202601027 (in Chinese with English abstract)

0 引 言

棉花是中国重要的经济作物和战略物资,其生产发展对国民经济具有显著影响。中国不仅是世界棉花生产水平最高的国家,同时也是需求量最大的国家[1,2]。然而在种植过程中,棉花全生育期均可遭受病菌侵害,而病害的严重发生会直接导致产量和品质的损失[3]。面对这一难题,传统的病虫害识别方法包括人工识别和仪器识别,在识别效率、识别准确性及应用场景等方面已无法满足科学研究和生产的需要[4]。随着国家智慧农业的深入推进,计算机视觉技术替代传统方法,已成为实现农作物病害智能感知的关键[5]。在大田作业场景下,针对边缘设备算力受限的瓶颈,构建轻量化、高精度的实时检测模型,已成为当前农业智能装备研究的热点与难点[6]。

当前,主流的目标检测算法根据处理流程可划分为两类架构,一类以 R-CNN (Region-Based Convolutional Neural Network)[7]、Faster R-CNN[8]、Mask R-CNN[9] 为代表的两阶段检测器,通过先粗选候选区,再精细分类与定位的分步策略,实现了高精度检测,但牺牲了实时性。另一类则是以 SSD

收稿日期:2026-01-22

基金项目:国家自然科学基金项目 (62173171)

作者简介:沈学利,博士,教授,研究方向为图像与视觉信息计算技术。E-mail:Shenxueli@lntu.edu.cn

*通信作者:张 越,硕士研究生,研究方向为图像与视觉信息计算技术。E-mail:zyue535@163.com

copyright©2026 by the authors (登录 www.smartag.net.cn 免费获取电子版全文)

（Single Shot Multi Box Detector）[10] 和YOLO（You Only Look Once）[11] 等为代表的单阶段检测方法，通过将目标检测视为1个统一的密集预测任务，实现了1次前向传播同时完成定位与分类，从而在速度上取得了革命性优势。在实际应用研究中，龚昌智和郭丹丹[12] 针对番茄叶片检测，提出了一种融合注意力机制的轻量化卷积网络，但其对小尺寸病斑的感知能力较弱。王俏等[13] 虽然通过优化锚框生成与分类机制提升了葡萄叶片病害的检测精度，但在应对复杂田间背景干扰时适应性仍显不足。

这些研究反映出基于卷积神经网络（Convolutional Neural Network, CNN）的病害检测模型在农业应用中仍面临模型效率与精度难以兼顾、复杂场景适应性不足等共性挑战。鉴于此，研究者们开始探索Transformer[14] 架构的全局建模能力与目标检测任务更高效结合的途径。CARION等[15] 首次提出了一个完全端到端的目标检测模型DETR（DEtection TRansformer），它摒弃了传统检测器中复杂的锚框设计与非极大值抑制后处理步骤，通过Transformer编码器与解码器结构直接预测目标集合，在概念上实现了重大革新。然而，标准DETR因其训练收敛速度慢、对数据要求高、计算复杂度高，难以满足实时检测需求[16]。为了突破DETR在实际部署中的实时性瓶颈，ZHAO等[17] 提出了RT-DETR（Real-Time Detection Transformer）。其核心创新在于采用高效的混合编码器重构了多尺度特征融合流程，并引入了基于交并比（Intersection over Union, IoU）感知的查询选择机制。

RT-DETR架构有着出色的泛化性能，成为智慧农业视觉感知领域的研究热点，目前的研究主要遵循两条改进路线：其一主要为提升计算效率。例如，FU等[18] 利用MobileNetv4重构残差主干，并规划轻量级特征金字塔，实现葡萄病害在端侧的低功耗检测。其二侧重于加强复杂场景中的特征提取。XIN和LI[19] 采用Faster Net结合级联注意力机制，提升了小样本在复杂背景下的识别稳定性。WU等[20] 凭借坐标注意力改良可变形卷积，有效地解决了相似病害之间特征混乱的问题。

尽管已有许多改良策略，但在面对棉花病害这一特定任务时，仍面临显著挑战。与其他常见作物不同，棉花病害在发生初期往往难以察觉，其病斑尺寸微小，极易受到田间复杂背景与多变光照条件的干扰。现有的轻量化模型普遍因计算资源受限，难以在特征信息稀疏且干扰众多的条件下实现有效识别。这导致模型在追求轻量化的过程中，不可避免地牺牲了对小尺寸病斑特征的检测精度。

针对上述挑战，本研究在RT-DETR-R18的基础上提出了改进的多尺度异构特征协同的棉花叶片病害检测模型（Multi-Scale Heterogeneous Synergistic Feature DETR, MHSF-DETR），主要从以下3个维度进行改进：首先，为解决微小病斑特征易丢失与复杂背景干扰问题，构建了分层上下文选择感知网络（Hierarchical Context-Selective Perception Network, HCSP-Net），通过微宏观空间上下文注意力模块（Micro-Macro Spatial Context Attention, M^2-SCA）与竞争性选择融合模块（Competitive Selection Fusion, CSF）协同增强特征判别力；其次，针对跨层级特征融合时的语义错位，设计了可学习加权上下文融合模块（Learnable Weighted Context Fusion, LWC-Fusion）实现动态特征对齐；最后，为克服传统卷积在叶片边缘产生的伪影并降低计算负担，引入边界感知重构机制（Edge-Aware Reconstruction Mechanism, EARM），重构特征聚合路径。

1 数据与预处理

本研究使用的图像数据来源于Roboflow Universe平台公开的棉花病害数据集[21]。该数据已预划分为训练集、验证集和测试集，三者数量遵循约8∶1∶1的比例，总计2 319张原始图像。数据集涵盖6种类别，包括卷叶病（Curl）、灰霉病（Grey Mildew）、叶斑病（Leaf Spot）、枯萎病（Wilt）、疫病（Blight）以及健康叶片（Healthy）。病害类别如图1所示。

为缓解因样本数量有限可能导致的模型过拟合问题，对预划分后的训练集和测试集进行数据扩增，使训练集样本量扩充至3 708张，测试集扩充至466张。此外，为保证评估的可靠性，232张验证集未经过任何处理。最终，构建了1个包含4 406张图像的棉花叶片病害数据集，用于后续模型训练与评估。

2 MHSF-DETR模型架构设计

2.1 网络整体架构

MHSF-DETR在保留RT-DETR高效混合编码器的基础上，对特征提取与融合路径进行了重构。如图2所示，整体数据流向如下：首先，图像输入HCSP-Net网络，该网络凭借分层异构结构，能够协同提取微观纹理与宏观语义特征，为后续处理提



a. 叶斑病 b. 卷叶病 c. 灰霉病 d. 枯萎病 e. 疫病 f. 健康

图 1 棉花病害类别

Fig. 1 Cotton diseases category

供丰富的多尺度信息。其次，提取的特征进入颈部网络，经由 LWC-Fusion 模块进行跨层级的动态加权融合。最后，融合后的特征流通过边缘重构 C3 模块（Edge-Refined Convolution C3 Module, ER-ConvC3）进行聚合，在有效抑制边缘噪声后，输入 Transformer 解码器完成端到端的预测。

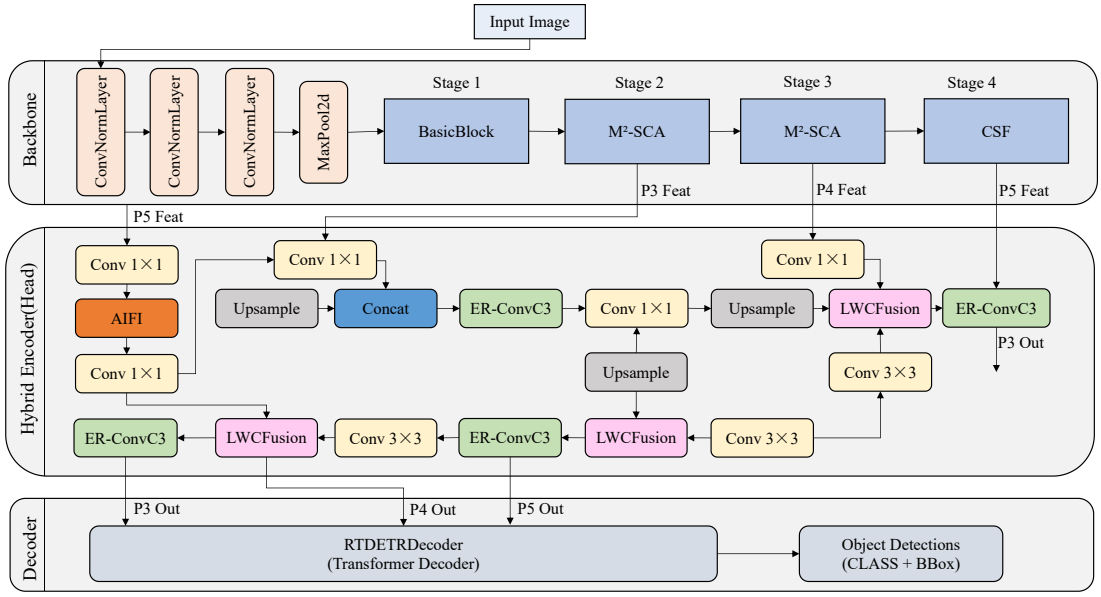


图 2 MHSF-DETR 网络结构图

Fig. 2 MHSF-DETR network architecture diagram

2.2 HCSP-Net 主干网络

传统 ResNet-18 结构采用均匀的层级堆叠与固定尺寸的卷积核进行特征提取，限制了其对剧烈尺度变化特征的感知能力。因此，本研究提出 HCSP-Net 主干网络。该网络依据特征层级的深度特性进行了差异化配置：Stage 1 保持标准卷积结构，以保留基础纹理信息；Stage 2 与 Stage 3 嵌入 M²-SCA 模块，专注于保留浅层特征中的微观病斑细节；在 Stage 4 集成 CSF 模块，旨在深层特征中强化语义定位并过滤背景干扰。这种分层递进的设计确保了网络能根据不同深度的特征需求进行针对性感知。

2.2.1 M²-SCA 模块

不同时期棉花叶片病斑在图像中呈现出显著的频域差异性：早期微小病斑表现为高频的纹理细节，而晚期大面积病变则表现为低频的形态结构。这种差异导致单一尺度的卷积核难以在统一特征空间内兼顾不同的表征需求。针对此问题，设计了 M²-SCA 模块，如图 3 所示。该模块通过通道语义过滤、双流空间感知与自适应融合的 3 阶段处理机制，在特征编码阶段实现对微观细节与宏观语义的并行捕获与自适应融合，有效解决了跨尺度特征的协同表征难题。

通道语义过滤阶段为防止背景噪声在后续空间卷积中放大，模块引入通道重校准机制对输入信号进行纯化。首先，利用全局平均池化（Global Average Pooling, GAP）将二维空间特征映射为一维通道描述符，随后通过一维卷积捕获跨通道的相互依赖关系，生成能够抑制无关背景通道的注意力权重。该权重经 Sigmoid 函数激活后，与输入特征逐

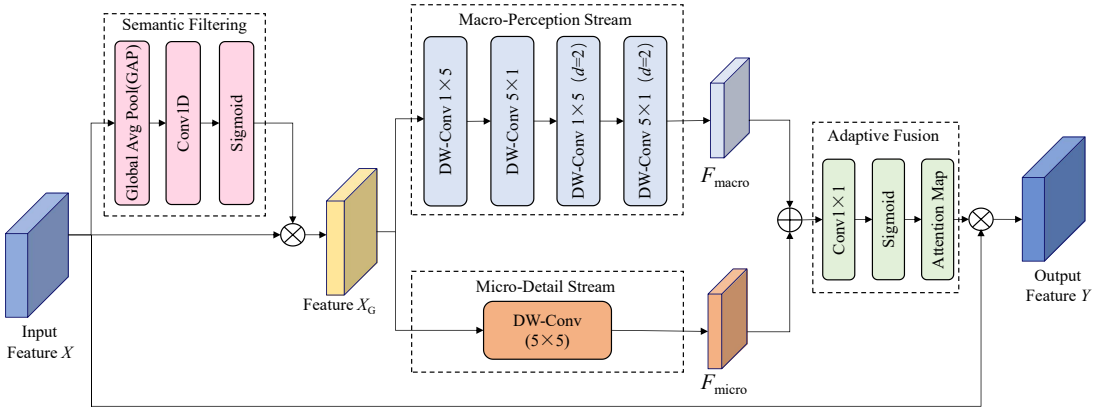


图 3　M^2-SCA 模块结构

Fig. 3　M^2-SCA architecture diagram

通道相乘，以此增强关键通道的响应并抑制无关背景信息。计算过程如公式（1）所示。

$$X_{\mathrm{G}} = X \odot \mathrm{Sigmoid}\left(\mathrm{Conv1D}_k\left(\mathrm{GAP}(X)\right)\right) \tag{1}$$

式中：X表示原始输入特征；X_{G}表示增强后特征；$\mathrm{Conv1D}_k$表示卷积核大小为k的一维卷积；$\odot$表示通道维度的逐元素相乘。此过程为后续的双流空间感知提供了1个高信噪比的纯净输入。

为达成空间信息的正交化解耦，双流空间感知阶段设计了一种微观细节与宏观感知并行的双通道结构。微观流利用深度可分离卷积提取局部高频特征，精准锚定微小病灶边缘梯度，规避下采样过程中的特征湮灭；宏观流则采用大核分解策略，以序列化操作单元替代常规大核，在遏制计算开销的同时赋予模型广阔的全局感受野与连贯特征。具体过程如公式（2）所示。

$$\begin{cases} F_{\mathrm{macro}} = \mathrm{D}^{d}_{K_{\mathrm{v}} \times 1}\left(\mathrm{D}^{d}_{1 \times K_{\mathrm{h}}}\left(\mathrm{D}_{K_{\mathrm{v}} \times 1}\left(\mathrm{D}_{1 \times K_{\mathrm{h}}}\left(X_{\mathrm{G}}\right)\right)\right)\right) \\ F_{\mathrm{micro}} = \mathrm{DWConv}_{s \times 5}\left(X_{\mathrm{G}}\right) \end{cases} \tag{2}$$

式中：F_{macro}和F_{micro}分别表示宏观和微观的输出特征；D表示空洞卷积；d为膨胀率；K_{v}和K_{h}分别表示垂直和水平方向的卷积核大小；DWConv表示深度可分离卷积；s为卷积核大小。

考虑到简单堆叠不同来源的特征可能扰乱其原始分布，模块引入了基于动态权重的自适应融合机制。该机制通过卷积压缩与Sigmoid函数生成空间注意力图，再以此为自适应掩码对输入特征进行加权融合。这样的设计实现了局部病灶细节与整体形态信息有效互补，从而增强了模型在复杂田间场景下对多尺度病斑的识别鲁棒性。

2.2.2　CSF模块

随着网络层级的加深，特征图分辨率的急剧下降常导致模型难以区分高度相似的背景与病理目标。为应对这一语义混叠难题，HCSP-Net在深层嵌入了CSF模块，其结构如图4所示。该模块突破了传统静态融合的维度局限，通过构建动态竞争仲裁机制，使网络在局部显著性与全局一致性之间自适应寻优，显著提升了对关键病灶区域的语义定位精度。

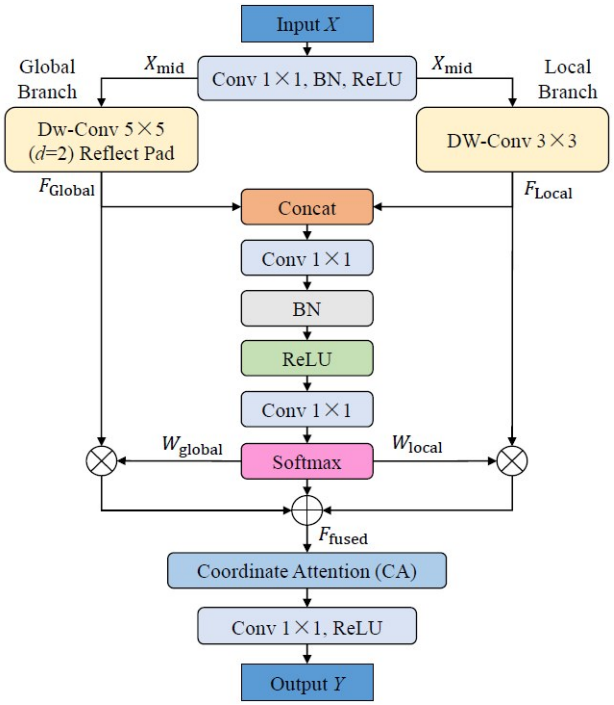


图 4　CSF 示意图

Fig. 4　CSF schematic

为打破单一感受野对特征表达的限制，模块首先构建了双路解耦架构，分别表征不同的语义维度。输入特征X首先经1 × 1卷积降维并激活，得到特征X_{mid}。随后，该特征被送入两个并行的特征

提取分支：局部细节分支采用3 × 3深度可分离卷积提取特征F_{Local}。该分支专注于捕获图像的高频分量，旨在保留微小病斑的边缘纹理信息，全局语境分支则使用5 × 5深度可分离卷积配合扩张率d = 2的空洞卷积，以扩大感受野提取特征F_{Global}。其计算过程如公式（3）所示。

$$\begin{cases} F_{\text{Local}} = \text{DWConv}_{3\times3}(X_{\text{mid}}) \\ F_{\text{Global}} = \text{DWConv}_{5\times5}^{d=2}(\text{ReflectPad}(X_{\text{mid}})) \end{cases} \tag{3}$$

式中：X_{mid}表示经卷积降维后的中间特征；F_{Local}和F_{Global}分别表示局部细节分支和全局语境分支提取的特征；Reflect Pad表示反射填充机制。

为实现局部与全局特征的动态选择，模块并未采用简单的元素级加法，而是构建了1个自适应的门控系统。首先将F_{Local}和F_{Global}在通道维度进行拼接，融合后的特征流经由全局平均池化与全连接层映射至权重空间。在此空间内，Softmax函数被用于激发两个分支之间的软竞争，生成两个互斥的权重掩码W_{local}与W_{global}（满足$W_{\text{local}} + W_{\text{global}} = 1$）。最终，融合特征$F_{\text{fused}}$通过加权求和得到。具体计算如公式（4）所示。

$$F_{\text{fused}} = F_{\text{Local}} \odot W_{\text{local}} + F_{\text{Global}} \odot W_{\text{global}} \tag{4}$$

式中：F_{fused}表示加权融合后的特征；W_{local}和W_{global}分别表示对应分支的自适应权重掩码。

该机制使模型具备了出色的动态适应能力：当图像中的病斑较为清晰时，模型会更多依赖局部纹理信息；而在遇到光照不均或边界模糊的情况时，则会侧重全局上下文进行判断。

为弥补竞争融合过程中可能损失的空间位置信息，本研究在模块末端引入了坐标注意力机制（Coordinate Attention）[22]，其通过沿水平和垂直方向聚合特征，对空间坐标响应进行重校准。该机制能对病灶中心二次定位，使输出特征在保持强语义判别能力的同时，增强位置感知精度，实现语义与空间信息的平衡。

2.3 边界感知重构机制

在处理不规则形态的棉花叶片时，传统卷积通常采用零填充策略以保持特征图尺寸，但易在叶片边缘区域引入无效噪声并产生边界伪影[23]。这种边界效应在经过多次卷积与池化叠加后，会导致边缘语义信息被稀释或失真，进而影响模型对位于图像边缘区域病斑的检测鲁棒性。为此，本研究提出边界感知重构机制，该机制主要由两个核心组件构成：边缘重构卷积（Edge-Refined Convolution,ER-Conv）与ER-ConvC3，旨在从信号延拓的角度出发，通过重构边缘像素的几何连续性，修复受损的语义特征。

2.3.1 ER-Conv卷积

ER-Conv卷积采用分割-变换-聚合的轻量化范式，在降低计算冗余的同时，引入反射填充机制来维持空间语义的完整性。其结构如图5所示。为兼顾计算效率与特征修复能力，模块采用差异化通道处理策略。将输入特征$X \in R^{C\times H\times W}$在通道维度上划分为3个独立的子集：局部特征子集$X_{\text{loc}}$、全局特征子集$X_{\text{glb}}$和恒等映射子集$X_{\text{id}}$。与传统卷积对所有通道执行相同操作不同，本模块通过可配置的通道分配比例参数实现差异化处理，其中通道按1∶1∶2的默认比例进行分配。

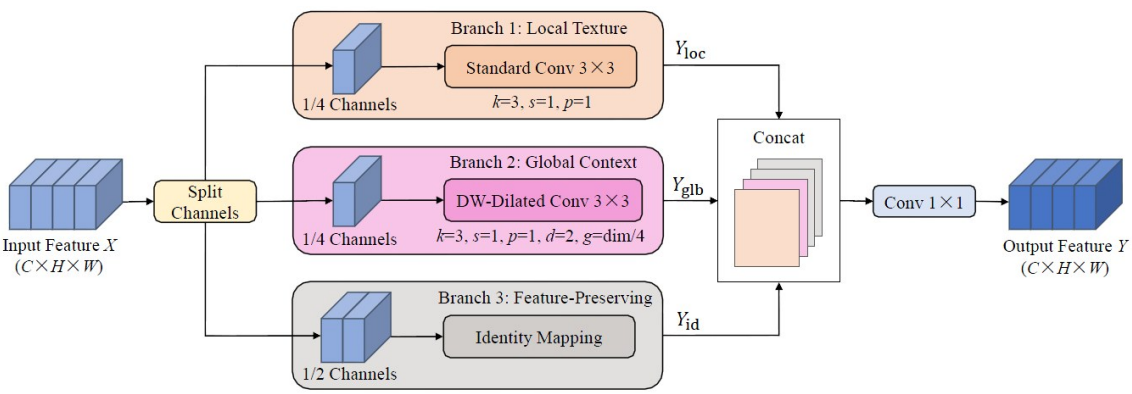


图5 ER-Conv结构图

Fig. 5 ER-Conv architecture diagram

模块通过3分支并行架构实现多源特征交互：局部分支利用标准卷积提取病斑纹理；为根除边界效应导致的漏检问题，全局分支采用基于反射填充的空洞卷积策略，通过镜像延拓取代传统补零，确保卷积核在处理边缘区域时仍能捕获有效的上下文信息，显著提升了模型对图像边缘病斑的检测性

能；恒等分支通过直接传递原始特征确保信息的完整性，并充当梯度缓冲路径，有效缓解了深层网络中的梯度消失问题。最后，通过通道拼接与卷积操作实现多源信息交互。该设计引入部分通道卷积（Partial Convolution）[24] 策略，在增强跨通道信息融合的同时显著降低了运算成本，实现了模型轻量化与边缘鲁棒性的平衡。具体计算过程如公式（5）所示。

$$\begin{cases} Y_{\text{loc}} = \text{Conv}_{3\times 3}(X_{\text{loc}}) \\ Y_{\text{glb}} = \text{DWConv}^{\text{reflect}}_{3\times 3, d=2}(X_{\text{glb}}) \\ Y_{\text{id}} = X_{\text{id}} \\ Y = \text{Conv}_{1\times 1}(\text{Concat}([Y_{\text{loc}}, Y_{\text{glb}}, Y_{\text{id}}])) \end{cases} \tag{5}$$

式中：X_{loc}、X_{glb}和X_{id}分别表示划分后的局部纹理、全局语境和恒等映射分支的输入特征子集；Y_{loc}、Y_{glb}和Y_{id}分别表示各分支处理后的输出特征；$\text{DWConv}^{\text{reflect}}_{3\times 3,\ d=2}$表示采用反射填充且膨胀率为2的3 × 3深度可分离卷积。

2.3.2　ER-ConvC3特征聚合模块

传统RepC3（Reparameterized Convolution C3）模块在多尺度特征聚合过程中容易忽视叶片病斑的边界一致性，尤其针对小目标病害检测时，其特征易受噪声干扰，导致微弱的病理特征被淹没。为了克服RepC3模块在多尺度语义特征融合中的有限全局感知能力[25]，本研究使用ER-ConvC3模块替代跨尺度通道融合模块（Cross-Scale Channel Fusion Module, CCFM）结构中的RepC3，其结构如图6所示。

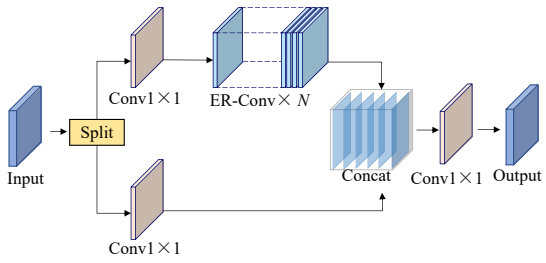


图6　ER-ConvC3结构图

Fig. 6　ER-ConvC3 architecture diagram

ER-ConvC3在CSP（Cross Stage Partial）网络的残差路径的基础上，通过级联N个ER-Conv卷积构建了边缘修复流，确保了在下采样与融合过程中病斑轮廓的清晰度，同时避免微弱边缘特征弥散。该模块利用灵活的通道分配参数n_{div}，以缓解特征分布的层级化差异。在默认均衡模式下，局部、全局与恒等分支按1∶1∶2的比例分配计算资源，兼顾融合效率与多尺度信息保留。当处于深层语义层级时，模块则转而采用高精细模式，舍弃恒等映射并将通道平均分配至局部与全局分支，以强化非线性变换，提升语义特征提取的准确性。

2.4　LWC-Fusion模块

深层特征语义性强而空间性弱，浅层特征则相反，两者在特征金字塔内部形成语义势差。传统的静态融合方法（如线性叠加、通道拼接）往往忽略层级间的分布差异，导致信息混淆与对齐偏差。为此，本研究设计了LWC-Fusion模块，该模块能够自适应地重构特征分布，进而实现跨层级信息在语义和空间维度上的精确对齐。如图7所示。

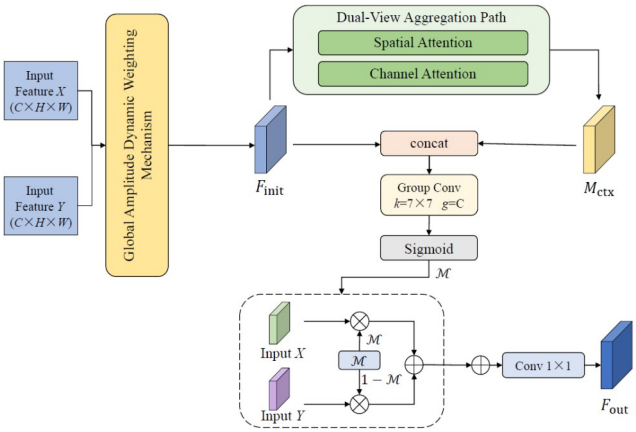


图7　LWC-Fusion结构图

Fig. 7　LWC-Fusion architecture diagram

为缩小跨层级特征的数值分布差异，本模块设计了全局幅值动态加权机制，具体过程如图8所示。该机制通过为特征引入可学习的分布调节因子W_1，W_2，替代了传统的静态超参数。为稳定训练，调节因子需经软归一化处理，再与对应特征逐元素相乘并累加，生成初步融合特征F_{init}。调节因子计算如公式（6）所示。

$$\hat{W}_k = \frac{W_{k,c}}{\sum_{j=1}^{2}\sum_{c=1}^{C} W_{j,c} + \epsilon}, \quad k \in \{1, 2\} \tag{6}$$

式中：$\hat{W}_k$表示归一化后第k个分支的学习因子；$W_{k,\,c}$表示第k个分支中第c个通道的原始权重值；C表示特征图的通道总数；ϵ为极小值常数。这种设计允许模型在通道维度上自适应地调X和Y的贡献比例，从而能够有效保留高响应区域的特征。

在完成初步数值对齐后，模块构建了双视点聚合路径以生成高精度的上下文注意力图。该路径包含两个并行分支：空间感知分支利用平均与最大池

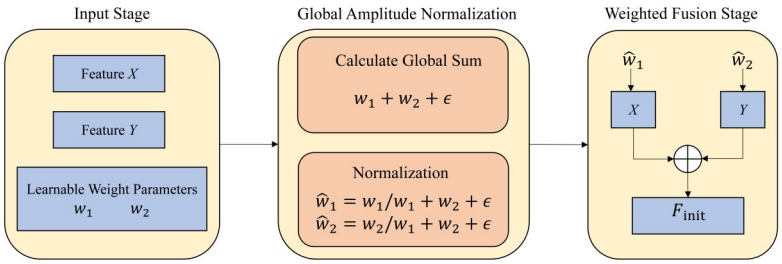


图8 LWC-Fusion模块全局幅值动态加权机制流程

Fig. 8 Process of the global amplitude dynamic weighting mechanism in LWC-Fusion

化压缩特征，配合反射填充卷积策略生成空间注意力图M_s，有效消除了边界伪影；通道感知分支则通过多层感知机制捕获全局依赖以生成通道注意力图M_c。最终，将两者输出进行叠加，以获得综合上下文图M_{ctx}。

最后，为了实现通道间的独立交互，将特征F_{init}与M_{ctx}进行堆叠，并采用分组卷积处理，再经Sigmoid激活生成像素选择掩码$\mathcal{M} \in [0, 1]^{C \times H \times W}$。同时，引入初始特征$F_{init}$作为残差项，以兼顾梯度稳定与基础信息的保留。最终通过1 × 1卷积融合输出特征F_{out}，计算过程为公式（7）。

$$F_{out} = \text{Conv}_{1\times1}(F_{init} + \mathcal{M} \odot F_{high} + (1 - \mathcal{M}) \odot F_{low}) \quad (7)$$

式中：F_{init}为初始残差项特征；F_{out}表示最终融合输出特征；F_{high}和F_{low}分别表示深层语义特征和浅层细节特征；$\mathcal{M}$表示生成的像素级选择掩码。

在该机制下，选择掩码$\mathcal{M}$被用于动态平衡深层语义特征F_{high}与浅层细节特征F_{low}的贡献权重：当掩码值趋近于1时，表明该位置语义相关性较强，侧重保留F_{high}；反之，当掩码值趋近于0时，侧重保留F_{low}。

3 结果与分析

3.1 实验设置与评估指标

本实验采用Windows 10操作系统，硬件平台采用Intel（R）Xeon（R）Silver 4216 CPU和NVIDIA GeForce RTX 3080（10 GB）GPU；深度学习框架采用PyTorch，版本为2.5.1+cu121，Python版本为3.9.20，CUDA版本为12.1。为确保实验结果的公平性与可比性，本研究中所有模型均在相同的软硬件条件下进行训练与测试，并使用统一的参数设置，关键参数设置如表1所示。

本研究采用的评估指标主要有准确率（Precision, P）、召回率（Recall, R）、平均精度均值（Mean Average Precision at IoU=0.5, mAP50）、模型参数量（Parameters）、计算量（GFLOPs）、推理耗时（Inference Time）、帧率（Frames Per Second, FPS）。P、R、mAP50、FPS计算如公式（8）～公式（11）所示。

$$P = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (8)$$

$$R = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (9)$$

$$\text{mAP50} = \frac{1}{N}\sum_{i=1}^{N}\int_0^1 P_i(R)\,dR|_{\text{IoU}=0.5} \quad (10)$$

$$\text{FPS} = \frac{1}{T_{\text{inference}}} \quad (11)$$

式中：TP、FP和FN分别表示正确检测出的目标数量、误检目标的数量和未检测出目标的数量；N为检测的类别总数；$P_i(R)$表示第i个类别在不同召回率下的准确率函数；$T_{\text{inference}}$表示模型处理单张图像所需的平均推理耗时。

表1 实验参数设置

Table 1 Experimental parameter settings

参数名称	参数数值
输入图像尺寸	640×640
初始学习率	0.000 1
权重衰减	0.000 5
批次大小	4
总训练轮次	200
优化器	AdamW

3.2 消融实验

为验证本研究MHSF-DETR模型各个核心模块是否有效，制定了8组消融实验方案，采用逐步集成的方式进行消融实验，各模块依次引入并评估其贡献，具体实验结果如表2所示。

实验数据显示，各个模块在棉花病害检测任务中表现出良好的协同效果。采用HCSP-Net主干网络，模型在参数减少19.6%的情况下，将mAP50提升至81.1%，这显示出轻量化设计和性能优化可以同时达成。这一提升主要源于两个方面：其一，

表2 核心模块消融实验结果

Table 2 Ablation study results of core modules

实验	LWC-Fusion	EARM	HCSP-Net	准确率/%	召回率/%	mAP50/%	参数量/M	计算量/GFLOPs
1	×	×	×	88.6	76.1	79.3	19.89	57.2
2	√	×	×	86.4	75.5	79.8	20.44	57.6
3	×	√	×	85.8	78.2	80.2	19.13	52.5
4	×	×	√	84.8	77.8	81.1	16.00	55.0
5	√	√	×	87.1	78.9	80.9	19.66	53.2
6	√	×	√	88.2	78.5	81.8	16.48	55.4
7	×	√	√	88.4	79.2	82.1	15.01	49.4
8	√	√	√	90.2	79.6	82.5	15.43	49.6

注：√表示引入该模块；×表示未引入该模块。

CSF模块通过削减通道多余部分减轻了计算压力；其二，M^2-SCA模块利用宏观和微观注意力机制，有效地加强了对病害微小纹理的识别能力。针对跨层级融合的难点，LWC-Fusion模块在独自使用时，其性能表现虽有波动，但与HCSP-Net协同工作时，召回率显著提升至78.5%，这证明了其能够修正语义不一致并减少漏检。EARM模块采用独特的反射填充策略，既能抑制边缘伪影和噪声，又大幅缩减了计算开销。当HCSP-Net与EARM结合时，EARM节省的计算资源与HCSP-Net提供的语义信息形成互补，使得模型以仅15.01 M的参数量便达成了82.1%的mAP50，后续加入LWC-Fusion模块之后，深层语义特征与浅层细节信息之间的交互得以加强，准确率与召回率同步提升，这验证了多尺度融合对模型判别能力改善的效果。

当MHSF-DETR模型完整构建后，其综合性能达到最优。与基准模型相比，参数量和计算量分别降低22.42%和13.29%，同时平均精度均值提高了3.2个百分点。结果表明，本研究设计的各个模块不但能够各自执行特定的功能，还从系统角度做到了有机融合，协同促进了模型整体性能的全面提升。

3.3 与主流模型对比实验

为全面评估MHSF-DETR在棉花叶片病害检测任务中的综合性能，实验将其与Faster R-CNN、SSD等经典架构，YOLOv5/v8/v10、RT-DETR-50等主流算法，EdgeNeXt-B（Edge-oriented Next Generation Networks）、MobileViT-S（Mobile Vision Transformer）等轻量化模型进行了横向对比。详细实验数据如表3所示。

表3 MHSF−DETR与不同检测模型性能对比实验结果

Table 3 Performance experimental results between MHSF-DETR and different detection models

算法	准确率/%	召回率/%	mAP50/%	参数量/M	计算量/GFLOPs
YOLOv5m	87.7	77.4	81.9	21.32	49.20
YOLOv8m	89.1	78.2	83.1	25.84	78.70
YOLOv10m	85.3	74.5	79.6	15.41	59.30
RT-DETR-R18	88.6	76.1	79.3	19.89	57.20
RT-DETR-R50	90.3	80.5	83.8	42.65	110.50
SSD	78.5	68.2	74.3	26.28	62.40
Faster R-CNN	81.3	76.9	80.1	136.02	358.50
EdgeNeXt-B	82.4	71.5	76.9	18.51	3.84
MobileViT-S	76.5	64.2	70.8	5.63	2.03
MHSF-DETR	90.2	79.6	82.5	15.43	49.60

实验结果表明，MHSF-DETR的性能提升并非源于简单的参数堆叠，而是通过深度的架构重构与轻量化设计实现的。相比于基线模型RT-DETR-R18，MHSF-DETR模型在参数量减少了22.42%的

同时，将mAP50提升至82.5%，大幅降低了计算成本，并增强了对病害特征的捕捉能力，在检测精度和推理效率之间达成了兼顾。同YOLO系列模型进行比较，MHSF-DETR同样表现出不错的综合性能，尽管YOLOv8m在mAP50上稍占优势，但在计算量和参数量上分别超出58.7%和67.5%，实际部署成本较高。与同规模模型YOLOv10m相比，本模型mAP50领先2.9个百分点，这表明本模型采用融合宏观与微观注意力的混合结构，对于提取复杂多尺度病斑特征更具潜力。

从轻量化角度看，相较于RT-DETR-R50，MHSF-DETR的参数量和计算量分别缩减63.8%和55.1%，mAP50仅下滑1.3个百分点，进一步对比Transformer轻量级变体，MobileViT-S因空间细节丢失导致召回率不足，而EdgeNeXt-B在参数量多出约3 M的前提下，精度仍落后本模型5.6个百分点，表明通用轻量化骨干难以兼顾定位与分类需求。此外，与传统检测架构SSD和Faster R-CNN对比，本模型在参数量与检测精度指标上均显著提升。这些对比结果充分展示了本模型在模型轻量化与特征表达能力上的双重优势。

综合各项评估指标，MHSF-DETR在检测精度、推理速度与模型轻量化程度上取得了出色平衡。其在多种对比实验中所表现出的强大综合适应能力，验证了其作为一款极具竞争力的轻量级病害检测解决方案的有效性。

3.4 泛化实验

为进一步验证MHSF-DETR模型在不同数据分布下的鲁棒性与泛化能力，本研究引入了两个具有不同分布特征的外部数据集进行测试。数据集1来源于Plant Village公开数据集，实验数据如表4所示。数据集2来源于Roboflow Universe平台上的公开数据集，实验数据如表5所示。

表4 不同模型在Plant Village数据集上的泛化实验结果

Table 4 Generalization experiment results of different models on the Plant Village dataset

算法	准确率/%	召回率/%	mAP50/%	参数量/M	计算量/GFLOPs
YOLOv5m	84.2	74.1	78.5	21.32	49.20
YOLOv8m	86.5	75.3	80.1	25.84	78.70
YOLOv10m	82.1	71.8	78.4	15.41	59.30
RT-DETR-R18	85.4	76.8	76.8	19.89	57.20
RT-DETR-R50	87.2	80.5	80.7	42.65	110.50
SSD	75.2	70.1	70.2	26.28	62.40
Faster R-CNN	78.6	76.5	76.5	136.02	358.50
EdgeNeXt-B	80.2	69.6	74.1	18.51	3.84
MobileViT-S	74.5	62.8	68.6	5.63	2.03
MHSF-DETR	87.8	76.9	80.3	15.43	49.60

表5 不同模型在Roboflow Universe数据集上的泛化实验结果

Table 5 Generalization experiment results of different models on the Roboflow Universe dataset

算法	准确率/%	召回率/%	mAP50/%	参数量/M	计算量/GFLOPs
YOLOv5m	89.5	78.4	84.2	21.32	49.20
YOLOv8m	92.2	82.5	89.6	25.84	78.70
YOLOv10m	92.8	81.9	88.1	15.41	59.30
RT-DETR-R18	88.2	77.5	82.5	19.89	57.20
RT-DETR-R50	93.5	83.1	88.6	42.65	110.50
SSD	82.4	72.1	76.8	26.28	62.40
Faster R-CNN	87.5	80.2	81.9	136.02	358.50
EdgeNeXt-B	83.2	73.3	78.5	18.51	3.84
MobileViT-S	80.5	69.4	73.1	5.63	2.03
MHSF-DETR	92.5	81.1	86.9	15.43	49.60

实验结果表明，面对不同来源与分布的数据集，MHSF-DETR均保持了稳定且极具竞争力的检测性能。即使在参数量低至15.43 M的条件下，其综合表现仍优于YOLOv8、YOLOv10及RT-DETR等先进模型，实现了精度与效率的平衡优化。这有力验证了MHSF-DETR在复杂农业场景下优异的泛化性能。

3.5 边缘计算下的模型部署性能评估

为验证模型在资源受限边缘设备上的实际部署潜力，本研究搭建了基于ARM架构的单核仿真测试环境，以最大程度还原农业物联网边缘节点的算力瓶颈。测试平台搭载ARMv8架构处理器，并严格限制为单核vCPU运行，内存配置为4 GB，操作系统为Ubuntu 22.04.5 LTS。推理测试基于ONNX Runtime引擎执行，未启用GPU加速，并强制设定为单线程模式，结果如表6所示。

表6　MHSF−DETR与不同算法在边缘设备上的推理效率对比结果

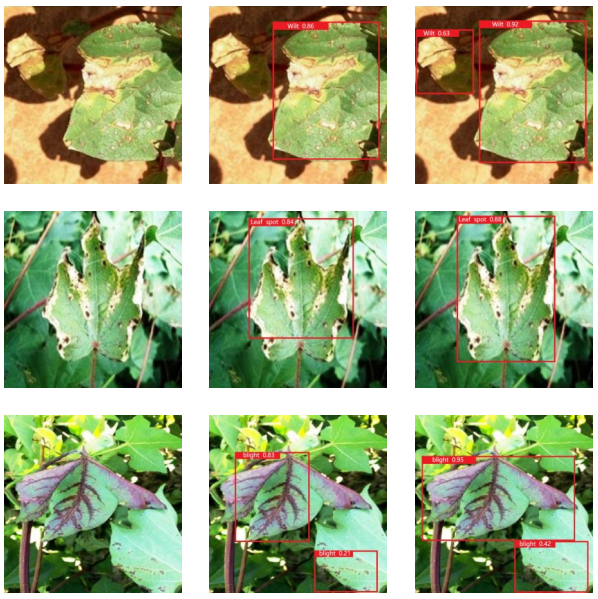
Table 6　Comparison of inference efficiency between MHSF-DETR and different algorithms on edge devices

算法	参数量/M	mAP50/%	推理耗时/ms	FPS/(帧/s)
YOLOv8m	25.84	83.1	2 054.56	0.49
RT-DETR-R50	42.65	83.8	2 833.34	0.35
RT-DETR-R18	19.89	79.3	1 685.23	0.59
MHSF-DETR	15.43	82.5	1 454.04	0.69

实验结果表明，在纯CPU推理环境下，YOLOv8m检测器表现出显著的计算负担。相比之下，MHSF-DETR凭借轻量化的骨干网络与高效的CSF模块，将单帧推理耗时压缩至1 454.04 ms，较RT-DETR-R50实现近2倍的加速。RT-DETR-R50虽语义提取能力强，但其模型冗余在ARM平台上引发了严重的延迟，难以支撑实际应用。综合而言，在极低算力约束下，MHSF-DETR在精度与效率之间取得了更优的平衡，展现出更强的边缘部署适配性。

3.6 可视化分析

图9的检测结果对比表明，相较于基线RT-DETR-R18，MHSF-DETR在识别准确率与目标定位的贴合度方面均有明显改善。该模型不仅能精准锚定叶片边缘的早期微小病斑，缓解小目标检测的困难，还在土壤、阴影等颜色相似的背景干扰下保持了稳定的判别能力，有效降低了误检现象。



a. 原图　　b. RT-DETR-R18　　c. MHSF-DETR

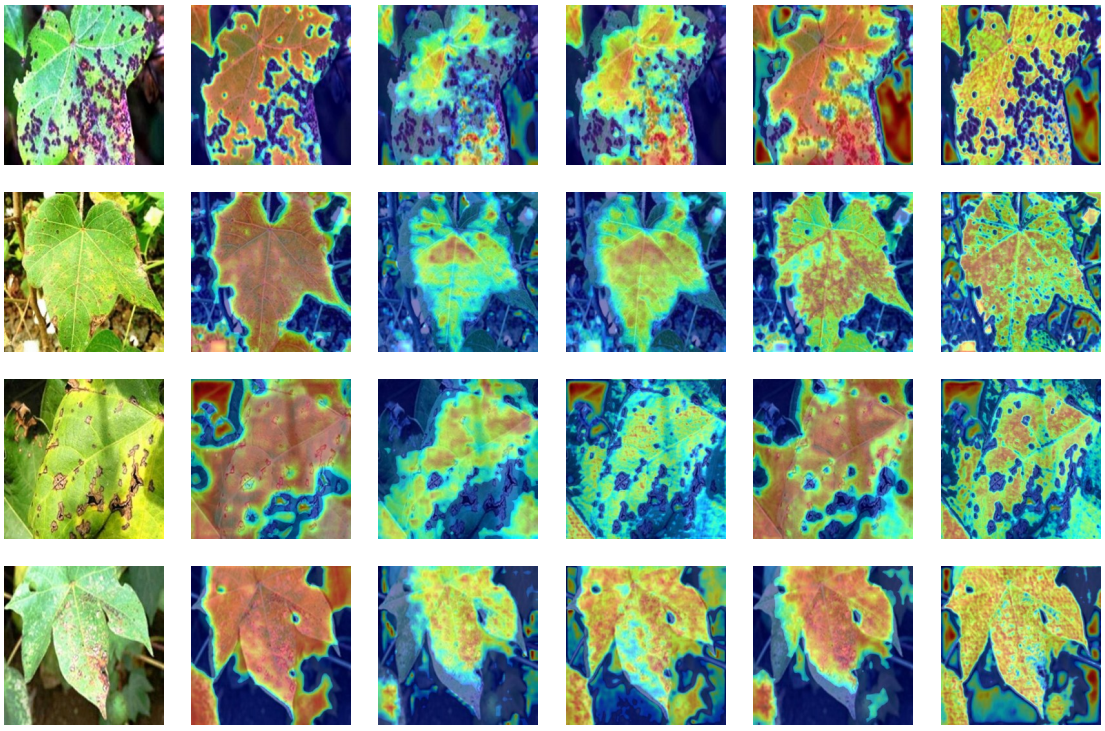
图9　MHSF-DETR与基准模型RT-DETR-R18的检测结果对比图

Fig. 9　Comparison of detection results between MHSF-DETR and baseline model RT-DETR-R18

为直观地阐释MHSF-DETR模型在棉花叶片病害检测中的决策依据，并验证其对关键病害特征的识别能力，采用Grad-CAM（Gradient-Weighted Class Activation Mapping）[26]技术对多种模型的特征提取结果进行了可视化分析。图10展示了原图、基线模型RT-DETR-R18，以及RT-DETR、YOLOv8m、YOLOv10m和本研究提出的MHSF-DETR模型在病害样本上的热力图对比。

如图10所示，针对棉花病害呈现的密集且微小的病斑，对比模型的热力图往往呈现出弥散状，难以精确区分独立的病灶点，甚至出现对健康叶肉区域的错误关注。相比之下，MHSF-DETR的热力图焦点高度集中于病斑中心，且边界清晰。这主要得益于骨干网络中引入的M^2-SCA模块，其微观细节流有效保留了高频纹理特征，使得模型在深层网络中依然能够保持对细小病灶的敏锐感知，有效解决了小目标特征丢失的问题。

对于分布在叶片边缘病斑或复杂背景干扰，传统模型因卷积操作中的零填充策略，在处理图像边缘时极易产生边界效应，导致热力图在叶片边缘出现虚假激活或对边缘病灶的漏检。而MHSF-DETR模型生成的注意力热图则能完整覆盖边缘处的病斑，且边界轮廓清晰。这一改进得益于ER-Conv模块的设计，有效消除了边界效应带来的噪声干扰，



a. 原图 b. RT-DETR-R18 c. RT-DETR d. YOLOv8m e. YOLOv10m f. MHSF-DETR

图 10 不同模型进行棉花叶片病害检测的热力图对比结果

Fig. 10 Comparative heatmap results of different models on cotton leaf diseases detection

显著提升了模型对边缘区域病害的检测鲁棒性。

在强反射光环境下，MHSF-DETR 凭借 CSF 模块的竞争性选择机制表现出良好的稳定性。该机制通过动态评定与重新校准局部与全局特征的置信度，有效抑制了非语义环境干扰造成的误激活。当叶片表面存在高光反射时，对比模型 RT-DETR 和 YOLOv8m 的注意力容易在反光区域产生虚假反应，造成错误的检测结果，而本研究方法通过自适应特征加权策略，实现了对真实病斑区域的精准关注。

综上所述，MHSF-DETR 不仅在特征关注合理性与抗干扰能力上表现优异，在检测精度等量化指标上也具备明显优势，这验证了其在复杂田间场景下相较于主流模型的先进性与实用性。

4 结 论

本研究针对智慧农业场景下高精度与低功耗难以兼顾的难点，构建了 MHSF-DETR 棉花病害检测模型。研究表明，通过微宏观双流感知与竞争性融合模块的协同，模型克服了多尺度病斑的特征提取难题；而 ER-Conv 与 LWC-Fusion 的引入，则从信号处理与特征对齐的维度，有效解决了边缘伪影与语义混叠问题。实验证实，MHSF-DETR 在参数量减少 22.42% 的前提下，取得了 82.5% 的 mAP50，在综合效能上优于当前主流的 YOLO 系列模型与 RT-DETR 模型，为深度学习模型在资源受限的田间移动终端部署提供了具备推广价值的轻量化范式。然而，本研究仍存在一定局限性。首先，模型目前主要聚焦于棉花叶片表面的常见病害检测，尚未涵盖棉铃、茎秆等其他部位的病理特征。其次，本研究已通过模拟边缘设备环境初步验证了模型的轻量化部署潜力，但其在真实的嵌入式硬件上进行实地测试与能耗分析，仍需在未来的工作中进一步开展，以全面评估模型在复杂农田环境下的运行稳定性与资源消耗表现。

利益冲突声明：本研究不存在研究者以及与公开研究成果有关的利益冲突。

参考文献：

[1] 赵卫松, 郭庆港, 鹿秀云, 等. 我国棉花主要病虫草害农药登记现状及存在问题与展望[J]. 农药学学报, 2026, 28(1): 64-75.
ZHAO W S, GUO Q G, LU X Y, et al. Current status, challenges and prospects of pesticide registration for major diseases, insect pests and weeds of cotton in China[J]. Chinese Journal of Pesticide Science, 2026, 28(1): 64-75.

[2] 魏梦婷. 中国棉花国际竞争力及影响因素分析[C]// 高质量伙伴关系与全球可持续发展论文集(下). 2022: 197-205. DOI: 10.26914/c.cnkihy.2022.079923.
WEI M T. Analysis on international competitiveness and influenc-

ing factors of Chinese cotton [C]// Proceedings of High-quality Partnership and Global Sustainable Development (Volume II). 2022: 197-205. DOI:10.26914/c.cnkihy.2022.079923.
[3] 徐虹, 李林峰, 杨昊, 等. 基于深度学习的植物叶病斑精细化分割方法[J]. 沈阳农业大学学报, 2025, 56 (1): 92-107.
XU H, LI L F, YANG H, et al. Fine segmentation method for plant leaf disease spots based on deep learning[J]. Journal of Shenyang Agricultural University, 2025, 56(1): 92-107.
[4] 王晋伟, 赵丽红, 师勇强, 等. 棉花病害全程防治技术研究初报[J]. 中国棉花, 2020, 47(5): 20-22, 46.
WANG J W, ZHAO L H, SHI Y Q, et al. Preliminary report on the whole process control techniques of cotton diseases[J]. China Cotton, 2020, 47(5): 20-22, 46.
[5] 曹冰雪, 赵春江, 李瑾, 等. 中国智慧农业技术发展现状、挑战与展望[J]. 农业工程学报, 2025, 41(21): 1-10.
CAO B X, ZHAO C J, LI J, et al. Current status, challenges and prospects of smart agriculture technology development in China [J]. Transactions of the Chinese Society of Agricultural Engineering, 2025, 41(21): 1-10.
[6] 赵法川, 徐晓辉, 宋涛, 等. 融合多头注意力的轻量级作物病虫害识别[J]. 华南农业大学学报, 2023, 44(6): 986-994.
ZHAO F C, XU X H, SONG T, et al. A lightweight crop pest identification method based on multi-head attention[J]. Journal of South China Agricultural University, 2023, 44(6): 986-994.
[7] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation [C]// 2014 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, New Jersey, USA: IEEE, 2014: 580-587.
[8] REN S Q, HE K M, 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.
[9] HE K M, GKIOXARI G, DOLLAR P, et al. Mask R-CNN[C]// 2017 IEEE International Conference on Computer Vision (ICCV). Piscataway, New Jersey, USA: IEEE, 2017: 2980-2988.
[10] LIU W, ANGUELOV D, ERHAN D, et al. SSD: Single shot MultiBox detector[C]// Computer Vision – ECCV 2016. Cham, Germany: Springer, 2016: 21-37.
[11] 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 (CVPR). Piscataway, New Jersey, USA: IEEE, 2016: 779-788.
[12] 龚昌智, 郭丹丹. 基于深度学习的番茄叶片病害检测研究[J]. 现代农业科技, 2025(10): 159-164.
GONG C Z, GUO D D. Tomato leaf disease detection based on deep learning[J]. XianDai NongYe KeJi, 2025(10): 159-164.
[13] 王俏, 张彪, 刘鑫. 基于改进行锚分类的快速葡萄叶片病害检测算法[J]. 江苏农业科学, 2024, 52(23): 206-213.
WANG Q, ZHANG B, LIU X. Rapid grape leaf disease detection algorithm based on modified anchor classification[J]. Jiangsu Agricultural Sciences, 2024, 52(23): 206-213.
[14] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[PP/OL]. arXiv: 1706.03762, 2017.
[15] CARION N, MASSA F, SYNNAEVE G, et al. End-to-end object detection with transformers [C]//Computer Vision ECCV 2020: 16th European Conference, Glasgow, UK, August 23-28, 2020, Proceedings, Part I. Cham: Springer, 2020: 213-229.
[16] DAI Z G, CAI B L, LIN Y G, et al. Unsupervised pre-training for detection transformers[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2023, 45(11): 12772-12782.
[17] ZHAO Y A, LYU W Y, XU S L, et al. DETRs beat YOLOs on real-time object detection[C]// 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, New Jersey, USA: IEEE, 2024: 16965-16974.
[18] FU Z L, YIN L F, CUI C, et al. A lightweight MHDI-DETR model for detecting grape leaf diseases[J]. Frontiers in Plant Science, 2024, 15: 1499911.
[19] XIN D Y, LI T Q. Revolutionizing tomato disease detection in complex environments[J]. Frontiers in Plant Science, 2024, 15: 1409544.
[20] WU M Y, QIU Y, WANG W Y, et al. Improved RT-DETR and its application to fruit ripeness detection[J]. Frontiers in Plant Science, 2025, 16: 1423682.
[21] DetectionDisease. Cotton disease detection dataset[DB/OL]. Roboflow Universe, 2024. [2026-01-11]. https://universe.roboflow.com/disease-detection-wounm/cotton-disease-detection-xevxs.
[22] HOU Q B, ZHOU D Q, FENG J S. Coordinate attention for efficient mobile network design[C]// 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, New Jersey, USA: IEEE, 2021: 13708-13717.
[23] GE X, ZHU Y, QI L P, et al. Enhancing border learning for better image denoising[J]. Mathematics, 2025, 13(7): 1119.
[24] LIU G L, REDA F A, SHIH K J, et al. Image inpainting for irregular holes using partial convolutions[C]// Computer Vision – ECCV 2018. Cham, Germany: Springer, 2018: 89-105.
[25] 李江, 骆炜, 陈豪, 等. 基于改进RT-DETR的PCBA管脚焊点缺陷检测方法[J]. 液晶与显示, 2025, 40(10): 1532-1544.
LI J, LUO W, CHEN H, et al. PCBA pin solder defect detection method based on improved RT-DETR[J]. Chinese Journal of Liquid Crystals and Displays, 2025, 40(10): 1532-1544.
[26] SELVARAJU R R, COGSWELL M, DAS A, et al. Grad-CAM: Visual explanations from deep networks *via* gradient-based localization[J]. International Journal of Computer Vision, 2020, 128 (2): 336-359.

Multi-Scale Heterogeneous Feature Synergistic Model for Cotton Leaf Disease Detection

SHEN Xueli, ZHANG Yue*, JIN Haibo, ZHANG Xuxu

(*School of Software*, *Liaoning Technical University*, *Huludao* 125105, *China*)

Abstract: [Objective] Detecting cotton leaf diseases in natural field environments is challenging due to diverse image interferences, variable leaf-spot sizes, and the demand for fast computation on mobile phones and other small-scale devices. Nevertheless, existing lightweight models cannot well balance detection accuracy and computational efficiency, especially for detecting small lesions and suppressing noise surrounding leaves. To address these challenges, MHSF-DETR (Multi-Scale Heterogeneous Synergistic Feature DETR), an improved detection model based on the RT-DETR framework, is proposed. It aims to achieve high-precision, low-power diagnosis in complex agricultural scenarios. **[Methods]** The primary innovation of this study consisted of the complete reconfiguration of

the feature extraction and fusion architectures. Firstly, a hierarchical context-selective perception network (HCSP-Net) was constructed as the backbone to replace conventional architectures. This backbone employed a differentiated processing strategy tailored to the depth of the feature maps: In the early parts of the process where the features were simple, it used something called micro-macro spatial context attention (M^2-SCA). This module used a channel semantic filter then a dual stream spatial perception structure to actively capture high frequency textures of micro-lesions and preserve macro semantics so that fine details were not lost when downsampled. At the deep feature stage, a competitive selection fusion (CSF) module was added. Unlike the traditional static summing approach, CSF created a dynamic competition arbitration system that flexibly balances local importance versus overall coherence via soft competition gates, making the semantics sharper and filtering away irrelevant background noise. Secondly, to tackle the spatial and semantic misalignment that was commonly seen in cross-level feature fusions, a learnable weighted context fusion (LWC-Fusion) module was created inside the neck network. This module used global amplitude dynamic weighting to learn autonomously the best blending ratios, so that deep semantic features were aligned precisely with shallow geometry. Moreover, to solve the problem of artifacts appearing at irregular leaf boundaries caused by traditional zero-padding convolutions, an edge-aware reconstruction mechanism (EARM) was proposed. By using edge-refined convolution (ER-Conv) and the edge-refined convolution C3 module (ER-ConvC3), which integrated reflection padding and partial convolution techniques, the model successfully curtailed invalid edge noise and diminished computational redundancy without sacrificing the geometric continuity of features. **[Results and Discussions]** Empirical benchmarks demonstrated that the proposed MHSF-DETR achieved a superior balance between detection performance and computational efficiency. Compared to the RT-DETR-R18 baseline, MHSF-DETR yielded a significant 3.2 percentage points increase in mean average precision (mAP), while simultaneously reducing parameters by 22.42% and GFLOPs by 13.29%. When benchmarked against mainstream detectors, MHSF-DETR consistently outperformed models such as YOLOv5m, YOLOv10m, and RT-DETR-R50. Although YOLOv8m maintained a marginal mAP50 lead in specific scenarios, its exorbitant computational overhead rendered it less practical for real-time deployment compared to MHSF-DETR. It successfully matched the lean efficiency of YOLOv10m but excelled in detection accuracy. Furthermore, extensive ablation studies confirmed that these performance gains stemmed from the structural synergy among the HCSP-Net backbone, the LWC-Fusion neck, and specialized reconstruction modules, rather than isolated component upgrades. These results validated the effectiveness of MHSF-DETR design in optimizing feature extraction and fusion, offering a highly efficient solution for resource-constrained object detection tasks. **[Conclusions]** MHSF-DETR addresses the long-standing accuracy-efficiency trade-off in cotton disease monitoring. By integrating hierarchical perception, adaptive fusion and edge refinement, the model mitigates scale disparity and resource-constraint challenges. It provides a feasible lightweight template for real-time diagnosis on agricultural edge devices to support deployment in smart-farming systems. Future work will expand validation to other plant organs (bolls, stems) and conduct rigorous embedded-hardware field tests to assess real-world robustness.

Key words: cotton disease detection; RT-DETR; lightweight model; attention mechanism; feature fusion

Foundation items: National Natural Science Foundation of China (62173171)

First author: SHEN Xueli, E-mail: Shenxueli@lntu.edu.cn
***Corresponding author:** ZHANG Yue, E-mail: zyue535@163.com