LiteFocus-Net：面向玉米叶片病虫害小目标检测的三点式轻量化增强框架

杨 玉[1], 张轶博[2], 毛 波[2], 章 磊[2*]

(1. 南京林业大学机械电子工程学院,江苏南京 210037,中国; 2. 南京财经大学计算机与人工智能学院,江苏南京 210023,中国)

摘 要：［**目的/意义**］玉米叶片病虫害早期目标通常具有尺度小、纹理弱、背景复杂以及类间差异不明显等特点，现有轻量化检测模型在田间复杂环境和边缘设备部署条件下难以兼顾检测精度与计算效率。为提高复杂田间环境下玉米叶片病虫害小目标识别能力，提出一种三点式轻量化增强框架LiteFocus-Net。［**方法**］以YOLOv11n为基础模型，从特征提取、细节恢复和损失优化3个方面进行改进。首先，设计自适应核轻量化模块（Adaptive Kernel Lightweight Block, AKL-Block），利用轻量化门控机制在3×3、5×5和7×7深度可分离卷积核之间进行动态选择，提高模型对不同尺度病虫害目标的特征提取能力；其次，提出特征分解重构模块（Feature Decomposition and Reconstruction, FDR），将深层特征划分为结构分支和细节分支，并采用轻量化亚像素重构方式恢复小目标边缘、纹理和局部轮廓信息；最后，提出尺度感知梯度增强损失（Scale-Aware Gradient Boosting Loss, SAGB-Loss），结合特征层级权重和目标面积连续增强因子，提升小目标样本在训练过程中的梯度贡献。［**结果和讨论**］在包含10 324张田间玉米叶片图像、8类病虫害目标的数据集上进行实验验证。LiteFocus-Net的mAP@0.5达到74.97%，相比YOLOv11n提升2.77个百分点；小目标检测精度AP_s由28.10%提升至32.77%，提升4.67个百分点。消融实验表明，AKL-Block、FDR和SAGB-Loss分别从多尺度特征表达、深层细节恢复和训练梯度优化方面发挥作用，三者协同提升了模型对微小病斑、虫卵和孢子簇等弱特征目标的检测能力。同时，模型参数量仅为2.52 M，计算量为5.3 GFLOPs，较YOLOv11n降低15.9%，且中、大目标检测性能保持稳定。［**结论**］LiteFocus-Net能够在较低计算开销下提升玉米叶片病虫害小目标检测性能，在精度和轻量化之间取得较好平衡，可为无人机巡检、农业机器人和边缘智能植保设备提供有效的视觉检测方法。

关键词：病虫害检测；小目标检测；轻量化模型；YOLOv11n；自适应卷积核选择；特征重构

中图分类号：TP391.41 **文献标识码：**A **文章编号：**SA202605017

引用格式:杨玉, 张轶博, 毛波, 章磊. LiteFocus-Net:面向玉米叶片病虫害小目标检测的三点式轻量化增强框架[J]. 智慧农业(中英文), 2026, 8(4): 217-237. DOI：10.12133/j.smartag.SA202605017

YANG Yu, ZHANG Yibo, MAO Bo, ZHANG Lei. LiteFocus-Net: A Three-Point Lightweight Enhancement Framework for Small-Target Detection of Corn Leaf Diseases and Pests[J]. Smart Agriculture, 2026, 8(4): 217-237. DOI：10.12133/j.smartag.SA202605017 (in Chinese with English abstract)

0 引 言

玉米病虫害的早期检测对于保障作物产量、减少农药使用、降低环境负担以及提高农业生产效率具有重要意义[1,2]，也为无人机撒药、智能采摘设备、自动巡检机器人等智慧农业设备的实时监控工作提出了更高要求[3-5]。早期病虫害通常呈现小尺度、低对比度、局部纹理弱、分布不均以及背景复杂等特点，农业视觉感知系统不仅需要处理高分辨率田间图像，还需在端侧设备上实现快速、低延迟的目标检测，这对模型的轻量化设计、精度兼顾及

收稿日期：2026-05-12

基金项目：国家重点研发计划项目（2024YFF1105504）

作者简介：1.杨 玉，硕士研究生，研究方向为智慧农业及装备。E-mail：2514056148@qq.com；2.张轶博，研究方向为人工智能应用。E-mail：feptsss@gmail.com；杨玉、张轶博为并列第一作者

*通信作者：章 磊，硕士，实验师，研究方向为智慧农业及装备。E-mail：zhanglei@nufe.edu.cn

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

部署效率提出了更高的技术要求[6,7]。现有轻量化目标检测方法如YOLO（You Only Look Once）系列，具备推理速度快、部署方便等优势，而YOLOv11采用C3k2模块，其轻量化版本（n/s）表现尤为突出[8-10]。研究者提出ShuffleNet[11]、MobileNet[12]、GhostNet[13]等轻量化主干网络，通过深度可分离卷积等操作进一步压缩计算量。而上述方法直接应用于玉米叶片病虫害检测仍存在三个主要问题。一是尺度适应能力不足。主干网络中的固定卷积核无法同时覆盖微小病斑、虫卵和较大叶斑区域，导致多尺度病灶特征提取不充分。二是深层特征细节丢失。经过多次下采样，小目标在深层特征图中的响应点极少，边缘、纹理和局部轮廓容易衰减或消失。三是小目标梯度易被淹没。训练阶段中大目标通常占据回归损失的主导地位，小目标对位置偏差更敏感，其梯度信号容易被中、大目标掩盖，从而影响优化效果。

现有研究提出了多种特征金字塔网络（Feature Pyramid Network, FPN）[14]、路径聚合网络（Path Aggregation Network, PANet）[15]、双向特征金字塔网络（Bidirectional FPN, BiFPN）[16]等多尺度特征融合策略，通过跨层信息交互增强小目标特征表达。然而，这类方法存在计算冗余，难以在端侧高效部署，且深层特征中小目标响应稀疏、浅层语义信息有限的问题仍未得到解决。在损失函数方面，Focal Loss通过降低易分样本权重来缓解类别和正负样本不均衡[17,18]，但它主要作用于分类分支，无法处理回归分支上的尺度失衡。广义交并比（Generalized Intersection over Union, GIoU）、距离交并比（Distance IoU, DIoU）、完全交并比（Complete IoU, CIoU）和高效交并比（Efficient IoU, EIoU）不断对边界框回归施加更精细的几何约束，定位精度虽然提升[19]，但小目标与大目标之间固有的损失幅值差距没有改善。也有研究根据特征层级分配不同的损失权重，让负责检测小目标的浅层获得更大的权重[20]。但这些改进多数只在单一维度（层级或面积）上调整，容易导致阈值边界附近梯度不连续。

近年来，面向农业场景的小目标检测研究逐渐由单一网络压缩转向多尺度特征提取、浅层细节保留、注意力增强和轻量化部署等协同优化方向。针对玉米叶片病害检测，MENG等[4]提出多尺度选择性机制YOLO，通过多尺度可变核卷积扩大不同尺度病斑的感受野，并结合边界框回归优化提升密集病斑定位能力。针对茶园害虫尺度小和目标不完整等问题，ZHANG等[21]提出AgriPest-YOLO，利用分组空间金字塔池化和协调局部注意力增强复杂背景下的多尺度害虫特征，在24类害虫数据集上测得交并比阈值为0.5时的平均精度均值（Mean Average Precision, mAP@0.5）较YOLOv5提升3.5%。ZHENG等[22]提出Rice-YOLO，通过多分支分组检测头、深度监督和改进动态上采样提升复杂背景下的水稻害虫检测能力，mAP@0.5达到78.1%。HUANG等[23]面向复杂环境下的青椒目标检测提出Pepper-YOLO，通过双金字塔可逆检测头降低特征传输过程中的信息损失，提高了遮挡条件下的目标检测与定位能力。LU等[24]设计MA-YOLO，通过轻量化卷积、多尺度空洞卷积和自适应特征融合增强农业害虫特征表达，该模型在Pest24数据集上的mAP@0.5达到73.9%，参数量与计算量分别降低28.6%和19.6%。CEN等[25]提出YOLO-LCE，通过轻量化残差模块、高效下采样卷积和Ghost检测头降低模型复杂度，mAP@0.5相比基线提升了1.7个百分点，参数量降低了43.9%。YANG等[26]针对玉米叶片病害目标尺度变化大和细节特征不足的问题，提出YOLO-SDW（SPD-Conv and Deformable Attention），引入空间深度转换卷积和可变形注意力机制，并采用Wise-IoU V3优化边界框回归，mAP@0.5相比原始YOLOv5提升6.4%，同时推理速度提高3.2%。

上述研究表明，多尺度建模、细节保留、注意力增强和轻量化设计能够在一定程度上改善农业小目标检测性能，但复杂融合结构增加了端侧部署开销，浅层增强难以低成本恢复深层已衰减的边缘纹理，单一层级或面积加权也未同时解决回归梯度失衡。在玉米叶片病虫害检测中，微小锈斑、虫卵和孢子簇常与叶脉、阴影、杂草等复杂背景混杂，并可能与大面积病斑同时出现，进一步增加了检测难度。因此，如何在较低计算开销下兼顾多尺度特征适应、深层细节恢复和小目标训练监督，仍是需要进一步解决的问题。

针对上述问题，本研究提出LiteFocus-Net，一种面向玉米叶片病虫害小目标的三点式轻量化增强框架，目的在于提高小目标检测精度，同时保持低计算量和端侧部署可行性。具体创新点如下：一是设计了自适应核轻量化模块（Adaptive Kernel Lightweight Block, AKL-Block），通过全局平均池化和轻量化门控机制为特征图动态选择3×3、5×5

或7×7的最优深度可分离卷积核，在提升多尺度特征表征能力的同时降低计算开销；二是构建了特征分解重构模块（Feature Decomposition and Reconstruction, FDR），将深层特征分为结构分支和细节分支，利用轻量化亚像素卷积重构细节分支，实现深层细节补偿与低计算成本的平衡；三是提出了尺度感知梯度增强损失（Scale-Aware Gradient Boosting Loss, SAGB-Loss），结合特征层级权重和连续指数型面积增强因子，对小目标回归梯度进行平滑放大，提升训练阶段小目标的贡献，同时避免对中、大目标精度产生负面影响。

1　数据与预处理

1.1　数据来源与标注

本研究数据由课题组于2023年6月至2025年9月在吉林、黑龙江、山东和河南等玉米主产区采集，覆盖多种天气、光照、叶片姿态、作物密度和杂草背景。采集设备为12 MP手机相机和16 MP便携式工业相机。剔除低质量样本后，对病斑、虫体等目标进行边界框标注，最终获得10 324张图像、8类目标；小目标定义为边界框面积小于32×32像素的实例，统计见表1。

表 1　田间玉米叶片病虫害数据集中8类目标的图像数、实例数及小目标占比

Table 1　Image counts, instance counts, and small-object proportions for eight categories in the field corn leaf disease and pest dataset

类别	图像数/张	目标实例数/个	小目标实例数/个	小目标占比/%
玉米锈病	1 503	3 964	2 761	69.7
灰斑病	1 247	2 638	1 316	49.9
玉米大斑病	1 189	2 147	758	35.3
叶枯病	1 083	1 841	551	29.9
玉米螟	1 164	2 509	1 623	64.7
草地贪夜蛾幼虫	1 287	2 895	1 738	60.0
虫卵	1 538	4 621	4 074	88.2
孢子簇	1 313	3 521	3 125	88.8
合计	10 324	24 136	15 946	66.1

1.2　数据划分

为避免相似图像跨集合造成数据冗余，依据采集地点、拍摄时段、设备和连续拍摄序列进行分组，并将同一叶片或同一病斑区域的高度相似图像归入同一组。随后在组层面进行分层抽样，使8类目标及不同尺度目标在训练集、验证集和测试集中的分布尽可能一致，具体划分见表2。

表 2　田间玉米叶片病虫害数据集的分层分组训练集、验证集和测试集划分

Table 2　Stratified group-wise division of the field corn leaf dataset into training, validation, and test subsets

类别	训练集/张	验证集/张	测试集/张	合计/张
玉米锈病	902	301	300	1 503
灰斑病	749	250	248	1 247
玉米大斑病	714	238	237	1 189
叶枯病	650	217	216	1 083
玉米螟	699	233	232	1 164
草地贪夜蛾幼虫	772	258	257	1 287
虫卵	923	308	307	1 538
孢子簇	788	263	262	1 313
合计	6 147	2 128	2 049	10 324

2　研究方法

2.1　实验平台、训练策略与评价指标

实验平台配置为：AMD Ryzen 7 7800X3D处理器、32 GB内存和NVIDIA RTX 3080显卡（10 GB显存），软件环境为PyTorch 2.0.1加CUDA 11.8。

训练时，输入图像分辨率设为640×640；采用动量因子0.937、权重衰减5e-4的随机梯度下降

(Stochastic Gradient Descent, SGD)优化器；初始学习率0.01，按余弦退火衰减至0.001；共训练200轮，批次大小取16。训练阶段为了增强模型对不同拍摄角度和叶片姿态的鲁棒性，在保持Mosaic增强的基础上加入随机旋转增强(旋转角度范围±15°)。AKL-Block中Gumbel-Softmax的温度系数τ在前100轮从1.0线性退火到0.1，后100轮固定为0.1。

在评价指标方面，使用交并比阈值为0.5时的平均精度均值(Mean Average Precision, mAP@0.5)、小目标精度AP_s(目标面积小于32×32像素)、中等目标精度AP_m(目标面积为32×32～96×96像素)、大目标精度AP_l(目标面积大于等于96×96像素)、参数量(Params)、计算量以及每秒处理帧数(Frames Per Second, FPS)。消融实验和本地复现模型均以随机种子42、123和2 024独立训练3次，实验结果统一表示为均值±标准差。多模型比较先采用单因素方差分析，当总体差异达到统计显著水平时，进一步采用Welch双侧t检验比较改进模型与基线模型之间的差异，并使用Holm-Bonferroni方法进行多重比较校正。显著性水平设置为$p<0.05$，其中$p<0.01$表示差异高度显著。

2.2 整体架构

本研究提出的LiteFocus-Net整体结构如图1所示，以YOLOv11n为基线，在三个关键阶段做了改进。

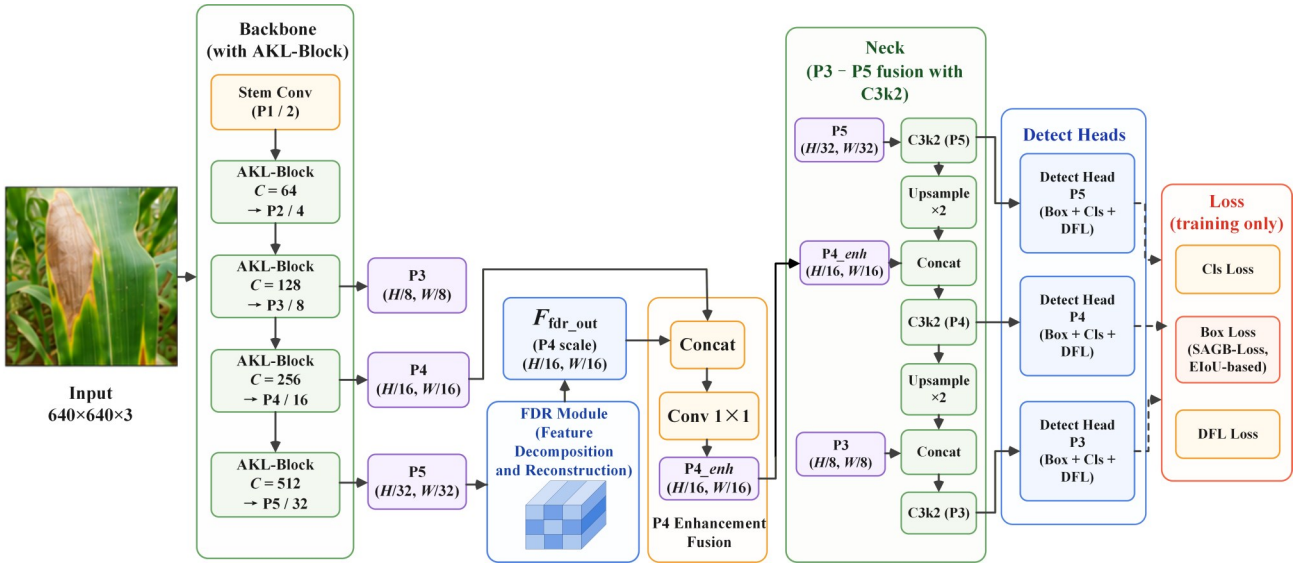


图1 LiteFocus-Net整体架构

Fig. 1 Overall architecture of LiteFocus-Net

1)主干网络。将原C3k2模块换为自适应核轻量化模块(Adaptive Kernel Lightweight Block, AKL-Block)，由核选择器完成输入特征图级的自适应感受野分配，仅以单核推理的代价获得多尺度表征能力。

2)特征增强层。在P5特征层后插入特征分解重构模块(Feature Decomposition and Reconstruction, FDR)，将深层语义特征划分为结构分支和细节分支，并将重构后的细节响应补偿到P4，再与Neck中的原P4特征融合后送入后续网络。

3)检测头与损失函数。检测头沿用了YOLOv11n的解耦头加分布焦点损失(Distribution Focal Loss Module, DFL)设计；训练时把原回归损失替换为尺度感知梯度增强损失(Scale-Aware Gradient Boosting Loss, SAGB-Loss)，通过二维联合的方式同时引入层级权重和面积增强因子，精准放大小目标训练梯度信号。

2.3 AKL-Block:自适应核轻量化模块

AKL-Block通过轻量门控在3×3、5×5和7×7候选深度可分离卷积核中选择最适合当前输入的单一路径，并结合Ghost通道扩展、SE通道重标定和残差连接，实现低开销多尺度特征提取，结构如图2所示。

模块主要实现过程如下。

1) Ghost通道扩展。给定输入特征$\boldsymbol{X}\in R^{C\times H\times W}$，首先用标准卷积进行计算，输出的固有特征图$\boldsymbol{X}_{\text{intrinsic}}$通道为输入的一半，再用深度可分离卷积从每个固有特征图生成一个幽灵特征图

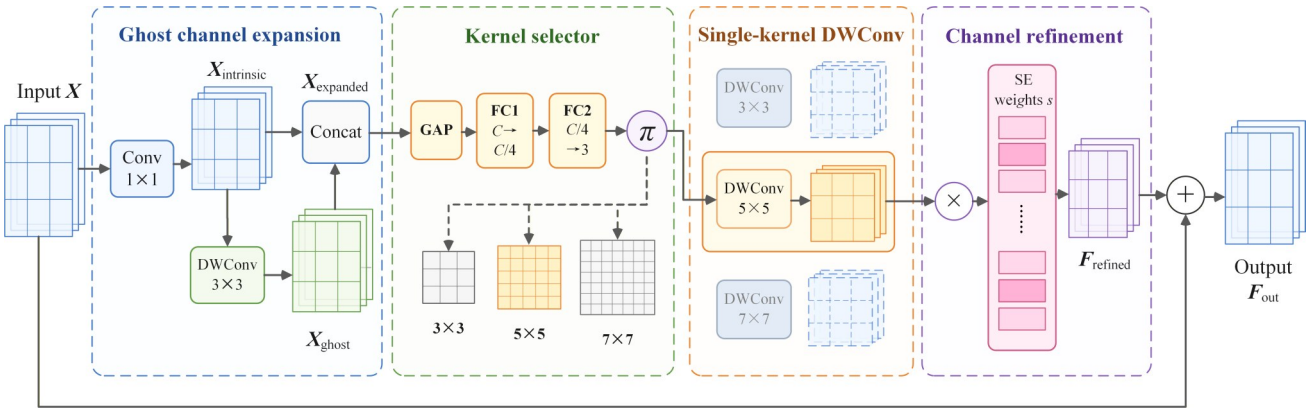


图 2　AKL-Block 结构图

Fig. 2　Structure of the proposed AKL-Block

$\boldsymbol{X}_{\text{ghost}}$，最后把两类特征图拼接起来，从而使总通道数保持不变[13]。具体计算过程如公式（1）～公式（3）所示。

$$\boldsymbol{X}_{\text{intrinsic}} = \text{Conv}_{1\times1}(\boldsymbol{X}) \in R^{C/2\times H\times W} \tag{1}$$

$$\boldsymbol{X}_{\text{ghost}} = \text{DWConv}_{3\times3}(\boldsymbol{X}_{\text{intrinsic}}) \in R^{C/2\times H\times W} \tag{2}$$

$$\boldsymbol{X}_{\text{expanded}} = \text{Concat}(\boldsymbol{X}_{\text{intrinsic}}, \boldsymbol{X}_{\text{ghost}}) \in R^{C\times H\times W} \tag{3}$$

式中：$\text{Conv}_{1\times1}$表示 1 × 1 标准卷积；C、H、W分别表示输入特征的通道数、高度和宽度；$\text{DWConv}_{3\times3}$表示 3 × 3 深度可分离卷积；Concat 表示在通道维度上的拼接操作；$\boldsymbol{X}_{\text{expanded}}$为拼接后输出的扩展特征张量。

在 Ghost 操作中，其所用深度可分离卷积的计算量比标准卷积低得多，而固有特征与幽灵特征的拼接又能维持总通道数不变。

2）卷积核选择器。选择器的任务是为整张特征图预测三个候选核的被选概率。它先对扩展特征做全局平均池化（Global Average Pooling, GAP），再经过两层全连接（Fully Connected, FC）层映射到 3 维通道，最后 Softmax 生成归一化概率，输出一个概率向量$\boldsymbol{\pi}\in R^3$，这部分的总参数量大约只有 0.005 M。其具体计算过程如公式（4）所示。

$$\boldsymbol{\pi} = \text{Softmax}(\text{FC}_2(\text{ReLU}(\text{FC}_1(\text{GAP}(\boldsymbol{X}_{\text{expanded}}))))) \tag{4}$$

式中：FC_1和FC_2分别表示第一层和第二层全连接操作；ReLU 为激活函数；$\boldsymbol{\pi}$为最终输出的 3 维归一化概率向量。

如图 2 所示，FC_1将C维特征压缩至$C/4$维以剔除冗余信息，FC_2进一步将特征映射到 3 维，分别对应三个候选卷积核尺寸。概率向量$\boldsymbol{\pi}$的三个分量$\boldsymbol{\pi}_1$、$\boldsymbol{\pi}_2$、$\boldsymbol{\pi}_3$，依次表示 3×3、5×5、7×7 深度可分离卷积核的选择概率。

3）可微分卷积核选择与推理。核选择本质上是离散决策，对 3 维概率向量$\boldsymbol{\pi}$进行 argmax 操作后，在推理阶段每张输入特征图只选择一个候选卷积核。但 argmax 操作是不可微分的，在训练阶段会把梯度传播路线切断。为了解决这个问题，这里使用了 Gumbel-Softmax 重参数化[27]，如公式（5）所示。

$$\boldsymbol{g} = \text{GumbelSoftmax}(\boldsymbol{\pi}, \tau) \in [0, 1]^3 \tag{5}$$

式中：$\boldsymbol{g}$为经过重参数化生成的软权重向量；τ为控制软分布平滑程度的温度系数。训练初期τ取 1.0，此时$\boldsymbol{g}$近似均匀软权重如［0.30，0.40，0.30］，3 个卷积核的输出按比例加权融合，梯度可以平稳流动。随着训练推进，τ逐步降到 0.1，$\boldsymbol{g}$逐渐收窄为独热向量，比如变化为［0.01，0.98，0.01］，选择器从软加权平滑过渡到离散决策。这种退火策略让训练过程始终保持连续可微，又在最终推理时实现了确定性的单核选择。

之后的特征图经由选中的卷积核计算得到公式（6）：

$$\begin{aligned}\boldsymbol{F}_{\text{selected}} = {} & g_1\text{DWConv}_{3\times3}\left(\boldsymbol{X}_{\text{expanded}}\right) + \\ & g_2\text{DWConv}_{5\times5}\left(\boldsymbol{X}_{\text{expanded}}\right) + \\ & g_3\text{DWConv}_{7\times7}\left(\boldsymbol{X}_{\text{expanded}}\right)\end{aligned} \tag{6}$$

式中：$\boldsymbol{F}_{\text{selected}}$为根据权重融合后的特征张量；$g_1$、$g_2$、$g_3$分别为软权重向量$\boldsymbol{g}$的三个标量分量；$\text{DWConv}_{3\times3}$、$\text{DWConv}_{5\times5}$、$\text{DWConv}_{7\times7}$分别表示三种尺寸的候选深度可分离卷积操作。

公式（6）对应训练阶段的软选择过程，3 个候选深度卷积分支均参与前向传播，并按照门控权重对输出进行加权融合以保证梯度可传递。而推理阶段则采用硬选择策略，根据门控网络输出的概率向量$\boldsymbol{k}^* = \text{argmax}_i(\boldsymbol{\pi}_i)$确定最优卷积核，仅保留概率

最大的单一路径。因此，下文结果中的运算量和推理延迟均基于推理阶段的单核路径计算，训练阶段的额外分支开销不计入边缘部署复杂度。3种深度可分离卷积的运算量分别为：3×3约$9C\times H\times W$，5×5约$25C\times H\times W$，7×7约$49C\times H\times W$。在等概率选择条件下，平均计算量约为$\frac{83}{3}C\times H\times W$，只占三路并行融合总计算量$83C\times H\times W$的33.4%。

4）为了加强所选卷积核输出的通道间交互，后面接了一个压缩激励（Squeeze-and-Excitation，SE）模块沿通道维度做自适应特征重标定[28]。同时引入残差连接，确保模型训练时数值稳定性。具体特征提炼与融合过程如公式（7）所示。

$$\boldsymbol{F}_{\text{out}} = \boldsymbol{X} + \text{SE}(\text{Conv}_{1\times 1}(\boldsymbol{F}_{\text{selected}})) \tag{7}$$

式中：$\boldsymbol{F}_{\text{out}}$为AKL-Block的最终输出特征张量；SE表示压缩激励操作。AKL-Block的候选卷积核设置为3×3、5×5和7×7，主要原因为：

1）覆盖典型目标尺度。在玉米叶片病虫害数据集中，微小病斑和孢子簇通常只有数像素，而大叶斑可达到几十像素。3×3卷积适合捕捉微小目标的局部细节，5×5提供中尺度特征感受野，7×7适合较大目标的上下文信息。3种核能有效覆盖实际目标尺度范围，保证多尺度特征表征能力。

2）计算开销与部署约束。更大卷积核（如9×9或11×11）会显著增加计算量和推理延迟，不利于轻量化和边缘设备部署；而更小卷积核（如1×1）主要用于通道混合，对空间多尺度信息捕获作用有限。综合考虑性能与计算成本，3×3、5×5、7×7是折中最优选择。

3）AKL-Block的训练阶段采用Gumbel-Softmax可微分核选择，保证在训练过程中3个卷积核的梯度可以平稳传播，推理阶段只激活概率最高的核，实现单核推理的低计算量特性。

2.4 FDR:特征分解与重构模块

YOLOv11n里P5特征图的尺寸，经过5次2倍下采样，缩小到输入分辨率的1/32。对于小于32×32像素的目标，在其输出的特征图上可能变为1×1的响应点，甚至消失。常规做法是对P5做双线性上采样，恢复到更高分辨率再与浅层特征融合。但双线性插值本质上是邻域像素的加权平均，难以充分恢复用于区分小目标和复杂背景的边缘、纹理等细节响应。另一种方案是用亚像素卷积（PixelShuffle）对全部P5特征做可学习上采样[29]，让网络自己从下采样特征中把丢失的细节重建出来。但P5特征的通道数通常有256或512，完整的亚像素卷积要先用1×1卷积把通道数扩大到4倍，其相应的参数量和计算量都会上升。

对于玉米叶片病虫害检测，下采样过程中更容易丢失的是病斑边缘、虫体轮廓、孢子纹理等局部细节响应；而叶片形状、大病灶位置、背景排布等全局结构信息，通常可通过双线性上采样进行近似恢复。因此，没有必要对所有通道都执行可学习上采样，而应将有限计算资源集中用于细节分支的恢复。

基于上述思想，设计了一个FDR模块，其用1×1卷积将P5特征分成两支：结构分支保留$C/4$个通道，主要用于传递叶片形状、病斑区域位置等全局结构信息，由双线性插值处理；细节分支保留$3C/4$个通道，并通过轻量化亚像素重构恢复边缘、虫体轮廓和孢子堆纹理等局部响应，再通过1×1卷积扩展至$4C_h^{\text{out}}$，经PixelShuffle重排后得到$C_h^{\text{out}} = C/8$通道的细节重构特征。其中可学习上采样的开销只发生在细节分支上。FDR具体结构如图3所示。

输入为P5特征$\boldsymbol{F}_{\text{P5}} \in R^{C\times(H/32)\times(W/32)}$，其中$C$为P5特征的通道数（YOLOv11n中为256）。其具体过程如下。

1）用1×1卷积对通道进行压缩后，将结构和细节分支分开，如公式（8）和公式（9）所示：

$$\boldsymbol{F}_{\text{global}} = \text{Conv}_{1\times 1}(\boldsymbol{F}_{\text{P5}}) \in R^{C/4\times H/32\times W/32} \tag{8}$$

$$\boldsymbol{F}_{\text{detail}} = \text{Conv}_{1\times 1}(\boldsymbol{F}_{\text{P5}}) \in R^{3C/4\times H/32\times W/32} \tag{9}$$

式中：C、H、W分别为P5特征的通道数及对应原图的高度和宽度；$\boldsymbol{F}_{\text{global}}$为输出的结构分支特征张量；$\boldsymbol{F}_{\text{detail}}$为输出的细节分支特征张量。

2）结构分支直接用双线性插值做2倍上采样，如公式（10）所示：

$$\boldsymbol{F}_{\text{global_up}} = \text{BilinearUpsample}(\boldsymbol{F}_{\text{global}}, \text{scale} = 2) \in R^{C/4\times H/16\times W/16} \tag{10}$$

式中：BilinearUpsample为双线性插值上采样操作；scale = 2表示空间分辨率放大2倍；$\boldsymbol{F}_{\text{global_up}}$为上采样后的结构特征张量。双线性上采样不引入额外可学习参数，计算开销相对较低，仅靠像素索引和加权求和就能完成。

3）对细节分支特征做亚像素上采样。亚像素卷积要求输入通道数为$r2 \times C_{\text{out}}$，这里$r = 2$，因此先通过1×1卷积把通道扩充至4倍，如公式（11）所示：

$$\boldsymbol{F}_{\text{detail_expand}} = \text{Conv}_{1\times 1}(\boldsymbol{F}_{\text{detail}}) \in R^{4C_{\text{detail_out}}\times H/32\times W/32} \tag{11}$$

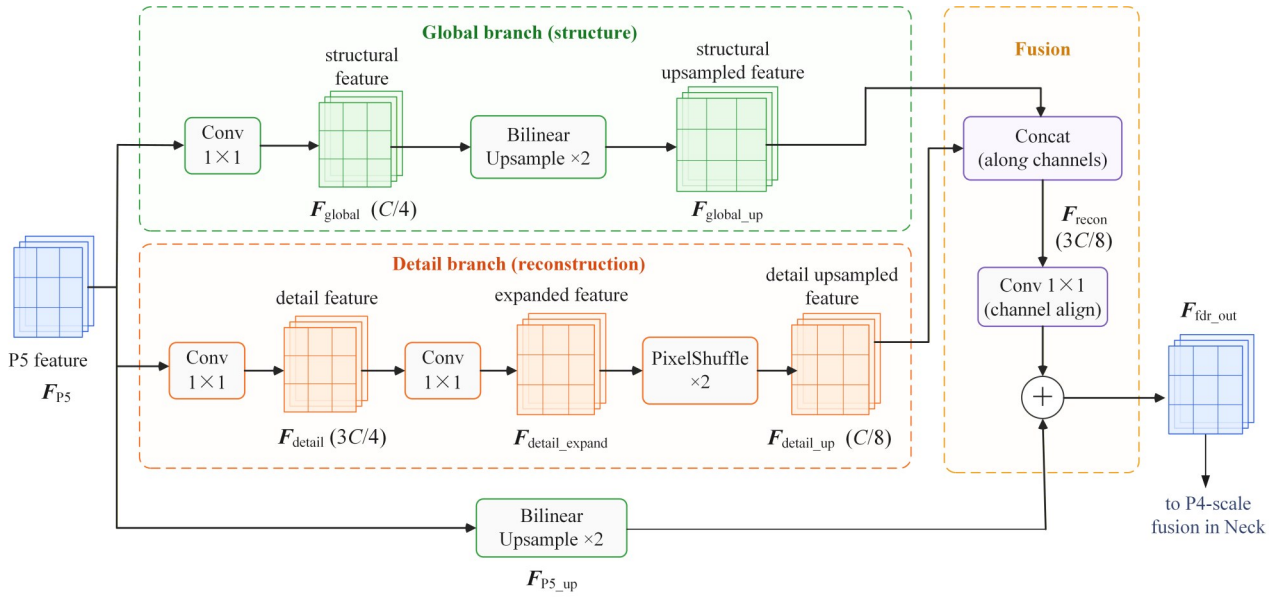


图3　FDR模块结构

Fig. 3　Structure of the FDR module

式中：C_{detail_out}为细节分支输出的通道数；$\boldsymbol{F}_{detail_expand}$为通道扩展后的细节特征张量。

细节分支输出通道数C_{detail_out}设为$C/8$，仅为结构分支的一半。随后通过PixelShuffle操作将扩展后的通道维度重排至空间维度，如公式（12）所示：

$$\boldsymbol{F}_{detail_up}=\text{PixelShuffle}(\boldsymbol{F}_{detail_expand}, r=2)\in \mathrm{R}^{(C_{detail_out})\times H/16\times W/16} \tag{12}$$

式中：PixelShuffle为亚像素卷积中的像素重排操作；$r=2$为上采样因子；$\boldsymbol{F}_{detail_up}$为重排后得到的高分辨率细节特征张量。

亚像素卷积在数学上可以描述为：输入张量维度为$(r^2\times C_{out})\times H\times W$，输出为$C_{out}\times rH\times rW$，映射关系为公式（13）所示：

$$\text{PixelShuffle}(\boldsymbol{F})_{c,y,x}=\boldsymbol{F}_{C_{out}\times r\times(y \bmod r)+C_{out}\times(x \bmod r)+c,\lfloor y/r\rfloor,\lfloor x/r\rfloor} \tag{13}$$

式中：$\boldsymbol{F}$代表输入的扩展特征张量；c、y、x分别表示输出张量在通道、高度和宽度维度上的索引；C_{out}为输出特征的通道数；mod为取余运算；$\lfloor\cdot\rfloor$为向下取整符号。

4）把两分支的上采样结果沿通道维度拼接，如公式（14）所示。

$$\boldsymbol{F}_{recon}=\text{Concat}(\boldsymbol{F}_{global_up},\boldsymbol{F}_{detail_up})\in R^{(C/4+C/8)\times H/16\times W/16} \tag{14}$$

式中：$\boldsymbol{F}_{recon}$为拼接重构后的特征张量。拼接后总通道数仅为$3C/8$。再接一个1×1卷积将其调整到与Neck中P4特征相同的通道数，为后续融合做准备。同时加入经过双线性上采样的P5特征作为残差连接，保证基础信息的传递不中断，如公式（15）所示。

$$\boldsymbol{F}_{fdr_out}=\text{Conv}_{1\times1}(\boldsymbol{F}_{recon})+\text{BilinearUpsample}(\boldsymbol{F}_{P5}) \tag{15}$$

式中：$\boldsymbol{F}_{fdr_out}$为FDR模块最终输出的融合特征张量。这条残差连接相当于给了一个从P5到P4的恒等映射，使得FDR只需要去学习相对双线性上采样的细节增强部分。后续FDR的输出特征$\boldsymbol{F}_{fdr_out}$与Neck中对应尺度的P4特征逐元素相加，送入后面的C3k2模块和检测头。

5）FDR输出还需与Neck中同尺度的P4特征进行融合。常见的候选融合方式有3种：逐元素相加（Add）、通道拼接和可学习加权相加。设P4层特征为$\boldsymbol{F}_{P4}$，最终输出的融合特征为$\boldsymbol{F}_{fuse}$，三种方案计算过程如公式（16）～公式（18）所示。

$$\boldsymbol{F}_{fuse}=\boldsymbol{F}_{P4}+\boldsymbol{F}_{fdr_out} \tag{16}$$

$$\boldsymbol{F}_{fuse}=\text{Conv}_{1\times1}(\text{Concat}[\boldsymbol{F}_{P4},\boldsymbol{F}_{fdr_out}]) \tag{17}$$

$$\boldsymbol{F}_{fuse}=\lambda\boldsymbol{F}_{P4}+(1-\lambda)\boldsymbol{F}_{fdr_out} \tag{18}$$

公式（16）表示Add方式。该方式将FDR输出视为细节残差，不增加额外通道；公式（17）表示通道拼接方式。该方式保留两路特征后再进行通道压缩，1×1卷积用于将通道数与P4对齐。公式（18）为可学习加权相加方式。其中λ为可学习权重，由Sigmoid函数约束在0与1之间。

以256通道P5特征做2倍上采样为例，把全亚像素卷积方案与FDR方案的计算量做一对比。全亚

像素卷积要先从C=256扩展到4×256=1 024通道，再做亚像素卷积得到256 × H/16 × W/16的特征，其复杂度$\mathrm{FLOPs}_{\mathrm{full}}$约为公式（19）所示。

$$\mathrm{FLOPs}_{\mathrm{full}} \approx C \times 4C \times HW/32^2 + (\text{Subsequent convolution FLOPs}) \quad (19)$$

FDR方案的计算量则由3部分组成：生成结构分支的降维卷积$C \times C/4 \times HW/32^2$、生成细节分支的降维卷积$C \times 3C/4 \times HW/32^2$、细节分支做亚像素卷积前的通道扩展卷积$3C/4 \times 4(C/8) \times HW/32^2$。此外仅涉及融合等低开销操作，总开销远低于全通道亚像素上采样。

2.5 SAGB-Loss：尺度感知梯度增强损失函数

小目标在训练中面临的层级和面积比例的双重信号衰减问题，已有方法通常只从其中一个维度去缓解。层级加权策略给不同特征层分配差异化的权重，用于跨层的粗粒度调配，但同一层内大目标和小目标仍然参数一致；而面积加权策略给面积小于预设阈值的锚框乘以固定放大系数，这类硬阈值方法在阈值附近梯度不连续，训练后期容易造成中等目标的震荡。

SAGB-Loss的思路是把层级权重和连续型面积增强因子结合起来。层级权重负责强化浅层特征图，指数型面积因子则在同层特征内，随着目标尺寸的减小而平滑地放大其梯度权重。这种设计使得小目标获取更高权重的过程是连续可导的，避免了硬阈值带来的突变问题。

对第l层特征图上的第k个锚框，记其匹配的真实框面积为A_k，面积增强因子w_k^{area}定义为公式（20）：

$$w_k^{\mathrm{area}} = 1 + \alpha \cdot \exp(-A_k/A_0) \quad (20)$$

式中：A_0为归一化面积常数，这里A_0=0.002 5对应约32×32像素的小目标面积阈值，α=2.0用来控制最大放大倍数，此时小目标损失的上限约为原始的3倍。当一个目标的面积远小于32×32像素时，指数项趋近于1，回归损失被放大至约3倍；当面积远超过这个阈值后，增益回落到0，损失不再被额外放大。变化全程是平滑的，权值不会在某一点突然跳变。

本研究选用EIoU作为基础回归损失。EIoU对边界框的宽高误差是分开建模的，相比CIoU对小目标的宽高偏差更敏感，其计算如公式（21）所示：

$$L_{\mathrm{EIoU}} = 1 - \mathrm{IoU} + \frac{\rho^2(b, b^{\mathrm{gt}})}{c^2} + \frac{(w - w^{\mathrm{gt}})^2}{c_w^2} + \frac{(h - h^{\mathrm{gt}})^2}{c_h^2} \quad (21)$$

式中：$\rho(\cdot)$表示欧氏距离；b和b^{gt}分别为预测框与真实框的中心点；c为最小外接矩形的对角线长度；c_w和c_h分别为最小外接矩形的宽度与高度；IoU为预测框与真实框的交并比；w和h为预测框的宽度和高度；w^{gt}和h^{gt}为真实框的宽度和高度。

结合层级权重w_l^{layer}与面积增强因子w_k^{area}，总回归损失定义如公式（22）所示：

$$L_{\mathrm{reg}} = \frac{\sum_l \sum_k \mathrm{w}_l^{\mathrm{layer}} \mathrm{w}_k^{\mathrm{area}} L_{\mathrm{EIoU}}}{\sum_l \sum_k \mathrm{w}_l^{\mathrm{layer}} \mathrm{w}_k^{\mathrm{area}}} \quad (22)$$

式中：L_{reg}为加权后的总回归损失；$\sum_l$和$\sum_k$分别表示对所有特征层l和所有正样本锚框k求和；L_{EIoU}为对应的基础回归损失。

层级权重w_l^{layer}随特征层级升高而减小，并依据特征图的步长进行定义，如公式（23）所示：

$$w_l^{\mathrm{layer}} = 1 + \beta \times \frac{S_{\max}}{S_l} \quad (23)$$

式中：第l层特征图的步长记为S_l，而$S_{\max}$表示网络中最大步长。YOLOv11中$S_{\max}$取值为32，β为可调节超参数。当β = 4时，计算可得P3层层级权重$w_{\mathrm{P3}}^{\mathrm{layer}}$ = 17、P4层$w_{\mathrm{P4}}^{\mathrm{layer}}$ = 9、P5层$w_{\mathrm{P5}}^{\mathrm{layer}}$ = 5。

分布焦点损失L_{DFL}保留原形式不作改动。模型总损失L_{total}为三项的加权求和，如公式（24）所示：

$$L_{\mathrm{total}} = \lambda_{\mathrm{box}} \times L_{\mathrm{reg}} + \lambda_{\mathrm{cls}} \times L_{\mathrm{cls}} + \lambda_{\mathrm{dfl}} \times L_{\mathrm{DFL}} \quad (24)$$

式中：λ_{box}、λ_{cls}、λ_{dfl}分别为边界框回归损失、分类损失和分布焦点损失的权重系数，L_{cls}和L_{DFL}分别对应分类损失项和分布焦点损失项。λ_{box} = 7.5，λ_{cls} = 0.5，λ_{dfl} = 1.5，全部沿用YOLOv11n的默认设置。

为了在后续消融实验中验证面积增强函数形式对回归优化的具体影响，本研究将公式（20）中的指数增强策略泛化为通用函数$g(A)$。在保持层级权重不变的条件下，进一步定义无面积增强、硬阈值增强、线性衰减、Sigmoid衰减和指数衰减5种候选形式。设目标归一化面积为A，面积阈值A_0 = 0.002 5，增强强度α = 2.0，各函数形式如公式（25）～公式（29）所示。

无面积增强：

$$g_{\mathrm{none}}(A) = 1 \quad (25)$$

硬阈值增强：

$$g_{\mathrm{hard}}(A) = 1 + \alpha I(A < A_0) \quad (26)$$

线性衰减增强：

$$g_{\text{linear}}(A) = 1 + \alpha\max(0, 1 - A/A_0) \tag{27}$$

Sigmoid衰减增强：

$$g_{\text{sigmoid}}(A) = 1 + \frac{\alpha}{1 + \exp(\tau(A - A_0))} \tag{28}$$

指数衰减增强（本研究采用形式）：

$$g_{\text{exp}}(A) = 1 + \alpha\exp(-A/A_0) \tag{29}$$

式中：$I(\cdot)$为指示函数，当括号内条件成立时取值为1，否则取0；$\max(\cdot)$为最大值函数；τ为平滑超参数，用于控制Sigmoid函数在阈值A_0附近的过渡速度。

3　结果与讨论

3.1　消融实验

为验证LiteFocus-Net中各模块的有效性，开展了消融实验。同时为进一步验证LiteFocus-Net中关键结构和参数配置的合理性，针对AKL-Block候选核集合、FDR分支通道比例、FDR与P4特征融合方式以及SAGB-Loss面积增强函数形式进行了相关消融实验。

3.1.1　模块消融实验

以YOLOv11n作为基线，逐模块进行消融，结果汇总在表3中。

表3　LiteFocus-Net在玉米叶片病虫害数据集上的完整消融实验及统计显著性分析

Table 3　Complete ablation study and statistical significance analysis of LiteFocus-Net on the corn leaf disease and pest dataset

Scheme	AKL	FDR	SAGB	mAP@0.5/%	AP_s/%	AP_m/%	AP_l/%	Params/M	计算量/GFLOPs	调整后p值（AP_s）	调整后p值（mAP）
YOLOv11n Baseline	—	—	—	72.20±0.20	28.10±0.50	55.80±0.30	64.30±0.20	2.58	6.3	—	—
only AKL-Block	√	—	—	73.50±0.20	30.10±0.40	56.20±0.20	64.10±0.30	2.50	5.0	0.026	0.005
only FDR	—	√	—	72.80±0.20	28.70±0.40	55.90±0.30	64.00±0.20	2.58	6.3	0.367	0.043
only SAGB-Loss	—	—	√	72.70±0.20	28.50±0.40	55.90±0.30	64.10±0.20	2.58	6.3	0.367	0.043
AKL-Block + FDR	√	√	—	74.30±0.30	31.40±0.50	56.10±0.30	63.90±0.30	2.52	5.3	0.008	0.005
AKL-Block + SAGB-Loss	√	—	√	73.70±0.20	30.30±0.40	56.30±0.20	64.10±0.30	2.50	5.0	0.023	0.005
FDR + SAGB-Loss	—	√	√	73.50±0.20	29.20±0.50	56.00±0.30	63.80±0.20	2.58	6.3	0.163	0.005
LiteFocus-Net	√	√	√	74.97±0.32	32.77±0.71	56.37±0.40	64.03±0.15	2.52	5.3	0.004	0.009

由表3可见，AKL-Block、FDR和SAGB-Loss均能够在YOLOv11n基线模型上带来一定增益，但三者的作用位置和提升幅度存在差异。YOLOv11n基线模型的mAP@0.5为72.2%，小目标AP_s为28.1%。单独引入AKL-Block后，mAP@0.5提升至73.5%，AP_s提升至30.1%，较基线提高2.0个百分点；同时，计算量由6.3 GFLOPs降低至5.0 GFLOPs。该结果说明，自适应核选择机制能够在减少冗余计算的同时，提高模型对不同尺度病斑、虫卵和孢子簇的表征能力，是三个模块中独立增益最明显的结构改进。

在不引入AKL-Block的情况下，单独加入FDR模块和单独使用SAGB-Loss对mAP@0.5和AP_s均有不同程度的增益。这表明FDR和SAGB-Loss并不依赖AKL-Block才能发挥作用。FDR可以在原始YOLOv11n主干网络基础上增强P5至P4的细节补偿能力，SAGB-Loss也可以通过调整不同尺度目标的回归梯度分布，改善小目标训练不足的问题。当FDR与SAGB-Loss组合使用时，mAP@0.5提升至73.5%，AP_s提升至29.2%，较基线提高1.1个百分点。该增益略高于两个模块独立增益之和，说明深层细节补偿与小目标梯度增强之间具有一定互补性。FDR使弱小目标在特征层中获得更清晰的局部响应，SAGB-Loss则在训练阶段进一步提高这些小目标样本的回归权重，两者分别作用于特征表达和监督优化环节。

AKL-Block与FDR联合使用时，mAP@0.5达到74.3%，AP_s达到31.4%，较基线提高3.3个百分点。该提升高于AKL-Block和FDR各自独立增益之和，表明两者之间存在较明显的正向协同作用。其原因在于，AKL-Block在主干网络中通过动态选择不同感受野的卷积核，改善了多尺度目标的初始特征提取质量；FDR则在此基础上对深层特征中的边缘、纹理和局部轮廓进行轻量化重构。换言之，AKL-Block为FDR提供了质量更高的可恢复特征，而FDR进一步减少了多次下采样造成的细节衰减。

在AKL-Block和FDR的基础上加入SAGB-Loss后，LiteFocus-Net的mAP@0.5为74.97%，

AP_s达到32.77%，较YOLOv11n基线提高4.7个百分点。与此同时，AP_m为56.37%，AP_l为64.03%，与基线相比未出现明显下降；模型参数量为2.52 M，计算量为5.3 GFLOPs，仍低于YOLOv11n基线的6.3 GFLOPs。该结果说明，LiteFocus-Net的性能提升来自3个模块在不同阶段的协同优化，三者共同作用，使模型在保持轻量化优势的同时，提高了玉米叶片病虫害小目标的检测精度。

由表4方差分析表明，各模块组合在mAP@0.5和AP_s上存在显著差异；经Holm–Bonferroni校正后的双侧t检验进一步表明，完整模型的AP_s提升达到统计显著水平。与此同时，AP_m和AP_l未出现明显下降。

表4 LiteFocus-Net完整消融矩阵在各精度指标上的单因素方差分析结果

Table 4 One-way ANOVA results of the complete ablation matrix of LiteFocus-Net on various precision metrics

指标	ANOVA结果	p值	结论
mAP@0.5	F(7,16)=44.78	<0.001	差异显著
AP_s	F(7,16)=32.25	<0.001	差异显著
AP_m	F(7,16)=1.47	0.246	差异不显著
AP_l	F(7,16)=1.19	0.361	差异不显著

重复性实验用了3组不同随机种子（seed=42、123、2024），结果见表5所示。

表5 LiteFocus-Net在3种不同随机种子下的重复性实验结果

Table 5 Repeatability results of LiteFocus-Net under three different random seeds

随机种子	mAP@0.5/%	AP_s/%	AP_m/%	AP_l/%
42	75.1	33.4	56.6	64.0
123	74.6	32.0	55.9	63.9
2 024	75.2	32.9	56.6	64.2
均值±标准差	74.97±0.32	32.77±0.71	56.37±0.40	64.03±0.15
95%置信区间	[74.17,75.77]	[31.00,34.53]	[55.36,57.37]	[63.65,64.41]

3组不同随机种子下LiteFocus-Net的标准差为0.32。模型在不同随机初始化条件下保持了较稳定的整体检测性能，但小目标指标仍表现出相对明显的波动，AP_s的标准差为0.71，高于AP_m和AP_l。其主要原因包括：小目标通常仅占据少量像素，预测框发生轻微偏移即可引起较明显的IoU变化；AKL-Block在训练阶段采用Gumbel-Softmax动态核选择，不同随机种子会对候选卷积核选择路径产生一定影响；SAGB-Loss根据目标面积动态调整回归梯度，不同训练批次中的小目标数量和尺度分布变化会放大局部波动。

表6统计了各模块在AP_s上的单独增益和组合增益，并计算了协同度，其中Overlap=联合增益−各模块单独增益之和。

表6 LiteFocus-Net各模块在小目标精度(AP_s)上的单独及组合增益对比

Table 6 Gain comparison of individual and combined modules on small object precision (AP_s) for LiteFocus-Net

模块组合	AP_s/%(均值±标准差)	相对基线增益	单独增益之和	交互增益
Baseline	28.10±0.50	—	—	—
only AKL	30.10±0.40	+2.0	—	—
only FDR	28.70±0.40	+0.6	—	—
only SAGB	28.50±0.40	+0.4	—	—
AKL + FDR	31.40±0.50	+3.3	+2.6	+0.7(正向)
AKL + SAGB	30.30±0.40	+2.2	+2.4	−0.2(轻微负向)
FDR + SAGB	29.20±0.50	+1.1	+1.0	+0.1(中性)
AKL + FDR + SAGB	32.77±0.71	+4.7	+3.0	+1.7(正向)

从结果可见，由于AKL保留的丰富特征细节正好为FDR的重构提供了更好的基础，AKL与FDR之间出现了正向协同（+0.7）。而AKL与SAGB之间则有轻微的负向协同（−0.2），FDR与SAGB的增益基本满足线性叠加。三模块全部叠加后的协同增益达到+1.7。

3.1.2　卷积核消融实验

AKL-Block默认采用｛3×3，5×5，7×7｝这3种深度可分离卷积核，为分析1×1卷积的作用，在默认集合基础上加入1×1候选核，并进一步增加9×9卷积作为大感受野对照，实验结果如表7所示。

表7　AKL-Block不同候选核集合的消融实验

Table 7　Ablation study of different candidate kernel sets in AKL-Block

卷积核选择方案	mAP@0.5/%	AP_s/%	AP_m/%	AP_l/%	Params/M	计算量/GFLOPs
｛3×3｝	72.80±0.20	29.00±0.40	55.90±0.30	63.70±0.30	2.47	4.6
｛3×3，5×5｝	73.20±0.20	29.70±0.40	56.00±0.30	63.90±0.30	2.49	4.8
｛3×3，5×5，7×7｝	73.50±0.20	30.10±0.40	56.20±0.20	64.10±0.30	2.50	5.0
｛1×1，3×3，5×5，7×7｝	73.40±0.20	29.90±0.40	56.10±0.30	64.00±0.30	2.51	4.9
｛3×3，5×5，7×7，9×9｝	73.60±0.20	30.20±0.40	56.20±0.30	64.20±0.30	2.52	5.4

由结果可见，仅保留3×3卷积时，模型计算量最低，但中、大尺度目标的上下文信息提取不足，mAP@0.5和AP_s均低于默认方案。加入5×5卷积后，检测性能有所改善。｛3×3，5×5，7×7｝组合获得73.5%的mAP@0.5和30.1%的AP_s，此时AP_s的单因素方差分析结果为：$F(4, 10)=4.26$，$p=0.029$，其中4和10分别为组间自由度和组内自由度；$p<0.05$表明该组候选核设置在5组中的AP_s均值具有统计学显著性差异，说明该组合在精度与计算量之间取得较好平衡。加入1×1候选核后，计算量略有下降，但AP_s由30.1%降至29.9%。这是因为1×1卷积缺少空间邻域建模能力，难以独立提取小病斑边缘和虫体轮廓。加入9×9卷积后，AP_s仅提高0.1个百分点，而计算量增加至5.4 GFLOPs。

3.1.3　通道比例消融实验

FDR将P5特征划分为结构分支和细节分支。为分析两类特征的合理分配比例，本研究设置1∶1、3∶5、1∶3和1∶7共4组结构分支与细节分支通道比例。P5输入通道数C取256，其余模块和融合方式保持不变，结果如表8所示。

表8　FDR模块不同通道比例的消融实验

Table 8　Ablation study of different channel ratios in the FDR module

结构分支∶细节分支	结构分支通道数	细节分支通道数	mAP@0.5/%	AP_s/%	AP_m/%	AP_l/%	Params/M	计算量/GFLOPs
1∶1	128	128	74.00±0.20	30.80±0.40	56.20±0.30	64.10±0.30	2.51	5.2
3∶5	96	160	74.10±0.20	31.10±0.40	56.20±0.30	64.00±0.30	2.51	5.2
1∶3	64	192	74.30±0.30	31.40±0.50	56.10±0.30	63.90±0.30	2.52	5.3
1∶7	32	224	74.10±0.20	31.50±0.50	55.80±0.30	63.60±0.30	2.54	5.4

由表8可见，随着细节分支通道占比增加，AP_s逐步提高，说明将更多计算资源分配给细节重构有利于增强微小病斑和孢子簇的局部响应。当结构分支与细节分支比例由1∶1调整为1∶3时，AP_s由30.8%提升至31.4%，而计算量仅由5.2 GFLOPs增加至5.3 GFLOPs。继续将比例调整为1∶7后，AP_s虽小幅提高至31.5%，但AP_m和AP_l分别下降至55.8%和63.6%，表明结构分支通道过少会削弱整体语义信息。综合考虑不同尺度目标性能和计算开销，本研究采用1∶3作为默认通道比例。

3.1.4　FDR融合方式消融实验

为验证2.4节所定义的FDR与P4特征融合策略，本研究在保持输出通道数及其余训练配置不变的条件下，对逐元素相加、通道拼接后接1×1卷积映射和可学习加权相加3种方案进行消融比较，结果如表9所示。

由表9可见，通道拼接后接1×1卷积的方式获得最高AP_s，但提升幅度仅为0.2个百分点，同时参数量由2.52 M增加至2.58 M，计算量由

表9 FDR与P4特征不同融合方式的消融实验

Table 9 Ablation study of different fusion strategies between FDR and P4 features

Scheme	mAP@0.5/%	AP_s/%	AP_m/%	AP_l/%	Params/M	计算量/GFLOPs
逐元素相加(Add)	74.30±0.30	31.40±0.50	56.10±0.30	63.90±0.30	2.52	5.3
通道拼接 + 1×1 Conv	74.40±0.30	31.60±0.50	56.20±0.30	64.00±0.30	2.58	5.6
可学习加权相加	74.40±0.30	31.50±0.50	56.20±0.30	64.00±0.30	2.53	5.3

5.3 GFLOPs增加至5.6 GFLOPs。可学习加权相加的性能与拼接方式接近，但相较默认的逐元素相加(Add)方案，增益仍较小。逐元素相加无需额外通道压缩操作，能够将FDR输出作为P4特征的细节残差进行直接补偿，更符合轻量化部署目标。因此，本研究采用逐元素相加作为默认融合方式。

3.1.5 面积增强函数消融实验

为验证2.5节所定义的面积增强函数形式，本研究在保持层级权重、A_0和α不变的条件下，对5种候选方案进行消融比较，结果如表10所示。

表10 SAGB-Loss不同面积增强函数形式的消融实验

Table 10 Ablation study of different area-aware enhancement functions in SAGB-Loss

Scheme	mAP@0.5/%	AP_s/%	AP_m/%	AP_l/%
无面积增强	74.70±0.30	32.20±0.50	56.40±0.30	64.00±0.30
硬阈值增强	74.60±0.30	32.80±0.50	55.80±0.30	63.80±0.30
线性衰减增强	74.70±0.30	32.50±0.50	56.00±0.30	63.90±0.30
Sigmoid衰减增强	74.80±0.30	32.60±0.50	56.20±0.30	64.00±0.30
指数衰减增强	74.97±0.32	32.77±0.71	56.37±0.40	64.03±0.15

由表10可见，不使用面积增强时，AP_s为32.2%。硬阈值增强可将AP_s提高至32.8%，但AP_m略有下降。这是因为硬阈值在尺度边界两侧产生不连续权重，容易导致接近阈值的中等目标训练波动。线性函数和Sigmoid函数通过平滑过渡缓解了该问题，但整体mAP@0.5仍略低于指数衰减方案。指数衰减函数在不同面积区间内均保持连续变化，最终获得75.0%的mAP@0.5、32.8%的AP_s和56.4%的AP_m。因此，采用连续指数型面积增强函数作为SAGB-Loss的默认设置。

3.2 模型对比

3.2.1 AKL-Block与现有多尺度模块对比

在相同的训练配置下(不引入FDR和SAGB-Loss，仅替换骨干网络结构)，将AKL-Block与固定7×7卷积的C3k2、C3k2-Inception，以及C3k2-SK模块进行对比，结果见表11。

表11 固定大核、Inception、SKNet和AKL-Block骨干模块的性能与复杂度比较

Table 11 Performance and complexity comparison of fixed-large-kernel, Inception, SKNet, and AKL-Block backbone modules

Scheme	mAP@0.5/%	AP_s/%	AP_m/%	AP_l/%	Params/M	计算量/GFLOPs
YOLOv11n Baseline	72.20±0.20	28.10±0.50	55.80±0.30	64.30±0.20	2.58	6.3
C3k2-Fixed 7×7	72.10±0.20	27.60±0.40	55.60±0.30	64.50±0.30	2.55	6.8
C3k2-Inception	73.70±0.30	30.40±0.50	56.30±0.30	64.80±0.30	2.68	7.9
C3k2-SK	73.60±0.30	30.20±0.50	56.20±0.30	64.60±0.30	2.66	7.3
AKL-Block(Ours)	73.50±0.20	30.10±0.40	56.20±0.20	64.10±0.30	2.50	5.0

由结果可见，固定7×7卷积模型的AP_s为27.6%，低于基线值，说明大核对微小病斑细节确实造成了过度平滑。Inception和SKNet的小目标精度分别达到30.4%和30.2%，高于AKL-Block的30.1%，这反映出并行多分支结构在捕获多尺度表达上更全面。但两者的计算量分别是7.9和7.3 GFLOPs，而表中AKL-Block的计算量为5.0 GFLOPs，mAP为73.5%，在精度损失极小的前提下显著降低了计算复杂度。选取同时包含大小尺度目标样本的图像进行测试，结果如表12所示。

表12数据也能反映出AKL-Block的局限。AKL-Block的小目标AP_s为30.1%，低于Inception

表 12　AKL-Block与Inception在同图大小尺度目标混合样本上的性能比较

Table 12　Performance comparison of AKL-Block and Inception on images containing mixed small and large targets

Scheme	mAP@0.5/%	AP_s/%	AP_l/%
C3k2-Inception	73.00±0.30	30.20±0.50	63.80±0.30
AKL-Block（Ours）	73.10±0.30	30.10±0.40	64.10±0.30

的30.2%。这是由于图像级的核选择无法在同一张图上同时兼顾到细小和大面积目标。而AKL-Block的大目标AP_l为64.1%，略高于Inception的63.8%，整体mAP为73.1%，与Inception基本相当。

3.2.2　FDR与其他特征增强方法对比

将FDR与多种用于P5→P4特征重构的主流方案进行对比，结果见表13。

表 13　FDR与其他特征增强方法对比

Table 13　Comparison of FDR with other feature enhancement methods

Scheme	mAP@0.5/%	AP_s/%	AP_m/%	AP_l/%	Params/M	计算量/GFLOPs	Extra Params/M
No P5 enhancement	73.50±0.20	30.10±0.40	56.20±0.20	64.10±0.30	2.50	5.0	0
Bilinear-only	73.60±0.20	30.30±0.40	56.10±0.20	64.10±0.30	2.50	5.0	0
Deconv	74.10±0.30	31.00±0.50	56.20±0.30	64.00±0.30	2.68	5.6	0.18
Full-PixelShuffle	74.80±0.30	32.20±0.50	56.40±0.30	64.20±0.30	2.62	5.8	0.12
FDR（Ours）	74.30±0.30	31.40±0.50	56.10±0.30	63.90±0.30	2.52	5.3	0.02

Full-PixelShuffle获得最高的AP_s（32.2%），比FDR高出0.8个百分点。原因在于它对所有通道都做了无差别的可学习上采样，能保留更多特征信息，但成本也明显更高，额外参数量为0.12 M，计算量提升了16%。FDR参数仅增加了0.02 M，计算量上升了6%，而AP_s从30.1%升到了31.4%，同时AP_1从64.1%降到63.9%，该结果表明FDR在轻量化约束下取得了较好的精度与计算量的折中。

选取包含超小目标（小于64像素）的图像进行实验，其召回率（Recall）对比见表14。FDR的Recall@0.5为33.0%，低于Full-PixelShuffle的34.8%。差距的根源在于FDR依赖P5残差特征，当原始P5信号已经非常微弱时，重构能带来的增益有限。而把IoU阈值放宽到0.3之后，两者的差距随之缩小，分别为50.4%和52.1%。

表 14　FDR与Full-PixelShuffle在超小目标上的召回率对比

Table 14　Recall comparison between FDR and Full-PixelShuffle on extreme small objects

Scheme	Recall@0.5/%	Recall@0.3/%
No P5 enhancement	31.20±0.50	48.50±0.60
Full-PixelShuffle	34.80±0.60	52.10±0.60
FDR（Ours）	33.00±0.50	50.40±0.50

3.2.3　SAGB-Loss与其他损失加权策略对比

在AKL-Block+FDR的网络框架下，把SAGB-Loss与几种主流加权策略进行对比，结果见表15。

表 15　SAGB-Loss与其他损失加权策略对比

Table 15　Comparison of SAGB-Loss with other loss weighting strategies

Loss Strategy	mAP@0.5/%	AP_s/%	AP_m/%	AP_l/%
CIoU（Baseline）	74.30±0.30	31.40±0.50	56.10±0.30	63.90±0.30
EIoU	74.50±0.30	31.70±0.50	56.30±0.30	64.00±0.30
EIoU + Layer-wise weighting	74.70±0.30	32.20±0.50	56.40±0.30	64.00±0.30
EIoU + Hard area threshold weighting	74.60±0.30	32.80±0.50	55.80±0.30	63.90±0.30
SAGB-Loss（Ours）	74.97±0.32	32.77±0.71	56.37±0.40	64.03±0.15

从CIoU换为EIoU之后，小目标AP_s涨了0.3个百分点，在此基础上加入层级加权，AP_s上升到32.2%。硬面积阈值方法则达到了最高的AP_s，但AP_m从56.3%降到了55.8%，其原因正是阈值边界梯度在中等目标上不连续而引发震荡。SAGB-Loss的AP_s为32.8%，接近硬阈值的水平，同时AP_m为56.3%，对中、大目标的负面影响微乎其微。

如前所述，超参数α用于控制小目标梯度的最大放大倍数。表16通过将α在0～4.0范围内遍历取值，进行超参数敏感性分析。

如结果所示，AP_s随α增大单调上升，从32.2%提升至32.9%。但当α超过2.0之后，AP_m开始下降，到α=4.0时降到了55.7%；AP_l也从64.0%回落到63.7%。mAP在α=2.0时达到峰值74.97%。综合考虑mAP@0.5、AP_s、AP_m和AP_l，本研究选取α=2.0。

表16 SAGB-Loss增强系数α对小、中、大目标检测性能的敏感性分析

Table 16 Sensitivity analysis of small-, medium-, and large-object detection to enhancement coefficient α in SAGB-Loss

单位：%

α	mAP@0.5	AP_s	AP_m	AP_l
0.0	74.70±0.30	32.20±0.50	56.40±0.30	64.00±0.30
1.0	74.82±0.30	32.50±0.55	56.40±0.30	64.05±0.25
2.0	74.97±0.32	32.77±0.71	56.37±0.40	64.03±0.15
3.0	74.75±0.30	32.85±0.60	56.00±0.30	63.90±0.30
4.0	74.45±0.30	32.90±0.60	55.70±0.30	63.70±0.30

3.2.4 LiteFocus-Net整体与现有方法对比

将本研究所提LiteFocus-Net与主流目标检测模型进行性能对比。所有模型均在同一数据集、相同实验配置下完成训练与测试，表17中为统一设置下的代表性结果。

表17 LiteFocus-Net与现有模型的性能对比

Table 17 Performance comparison between LiteFocus-Net and other models

Method	mAP@0.5/%	AP_s/%	AP_m/%	AP_l/%	Params/M	计算量/GFLOPs	FPS/(帧/s)
Faster R-CNN (ResNet50)	66.9	19.3	48.2	60.4	137.10	370.20	21.0
SSD (VGG16)	72.3	24.5	53.1	63.8	26.29	62.74	41.0
RT-DETR (R18)[30]	75.8	31.2	57.0	65.8	20.00	71.10	24.0
YOLOv5n	62.1	22.3	48.5	56.2	1.90	4.50	141.0
YOLOv8n	71.1	26.5	54.2	63.1	3.20	8.70	124.0
YOLOv11n (Baseline)	72.2	28.1	55.8	64.3	2.58	6.30	136.0
YOLOv11s	74.5	30.1	56.6	65.2	9.42	21.90	86.0
YOLOv11m	75.3	31.0	57.0	65.8	20.00	67.70	51.0
YOLOv11l	76.7	32.3	58.2	66.5	25.29	86.60	43.0
EfficientViT[31]	73.3	29.5	55.9	64.8	3.56	6.90	128.0
Gold-YOLO[32]	74.8	31.5	56.9	65.6	21.50	46.00	38.0
YOLO-MS (XS)[33]	73.8	30.2	56.1	64.9	4.50	8.70	118.0
YOLO-MS (S)[33]	75.6	32.0	57.2	66.0	8.10	15.40	82.0
DAMO-YOLO (Tiny)[34]	73.1	29.0	55.5	64.5	6.80	14.10	95.0
DAMO-YOLO(Small)[34]	74.3	30.6	56.3	65.2	11.20	23.40	68.0
RTMDet-Tiny[35]	73.5	29.8	55.7	64.7	4.00	14.20	72.0
RTMDet-Small[35]	74.9	31.2	56.8	65.3	8.89	14.80	68.0
MobileViT-YOLO[36]	72.8	28.6	54.8	64.0	8.60	10.20	105.0
EMO-YOLO[37]	73.4	29.7	55.6	64.5	5.10	8.50	122.0
FasterNet-YOLO[38]	72.5	28.9	55.0	64.2	3.96	7.80	136.0
YOLO-MSM[4]	74.0	30.2	55.8	64.5	5.40	18.50	69.6
MA-YOLO[24]	73.9	32.1	56.8	65.2	6.42	17.20	67.2
YOLO-LCE[25]	73.9	30.1	56.2	64.0	1.69	5.40	130.0
LiteFocus-Net (Ours)	75.0	32.8	56.4	64.0	2.52	5.30	135.0

其中，LiteFocus-Net的mAP@0.5为75.0%，略逊于YOLOv11m（75.3%）和YOLOv11l（76.7%），但参数量仅2.52 M、计算量5.3 GFLOPs，参数量约为YOLOv11m的八分之一，计算量约为其十三分之一。AP_s为32.8%，在所有对比模型中排在第一位，相比YOLOv11n，AP_s提高4.7个百分点，mAP上升2.8个百分点，计算量下降15.9%。AP_l为64.0%，低于部分更大规模模型，其原因在于

LiteFocus-Net把更多表征能力分配到小目标上。

由表18结果可见，与YOLOv11n基线模型相比，LiteFocus-Net在整体检测精度和小目标检测性能方面均获得了稳定提升，经Holm-Bonferroni方法校正后，p值为0.002，差异达到统计显著水平。LiteFocus-Net的AP_s校正后的p值为0.004，同样达到统计显著水平。相比之下，AP_m和AP_l两项指标的差异均未达到统计显著水平。结果表明，LiteFocus-Net能够在不显著降低中、大目标检测性能的前提下，有效提升模型对玉米叶片病虫害小目标的识别能力。

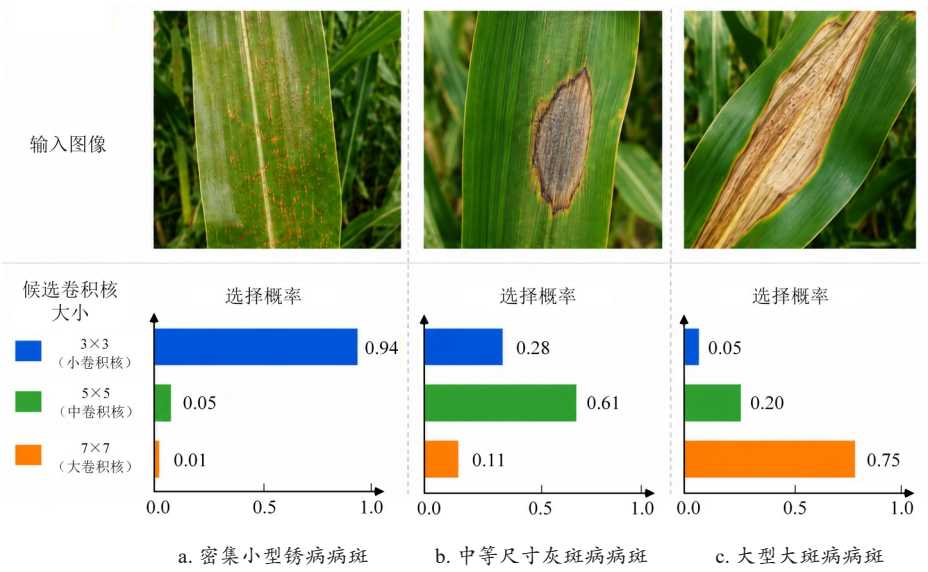
表18　YOLOv11n与LiteFocus-Net的统计显著性对比

Table 18　Statistical comparison between YOLOv11n and LiteFocus-Net

指标	YOLOv11n	LiteFocus-Net	提升/百分点	调整后p值	结论
mAP@0.5/%	72.20±0.20	74.97±0.32	+2.77	0.002	显著提升
AP_s/%	28.10±0.50	32.77±0.71	+4.67	0.004	显著提升
AP_m/%	55.80±0.30	56.37±0.40	+0.57	0.250	差异不显著
AP_l/%	64.30±0.20	64.03±0.15	-0.27	0.250	差异不显著

3.3　可视化分析

从测试集中挑出包含不同尺度病斑的图像，对推理阶段卷积核的选择概率进行可视化。蓝色对应3×3，绿色对应5×5，橙色对应7×7，结果如图4所示。



a. 密集小型锈病病斑　　b. 中等尺寸灰斑病病斑　　c. 大型大斑病病斑

图4　AKL-Block卷积核选择行为可视化

Fig. 4　Visualization of kernel selection behavior of AKL-Block

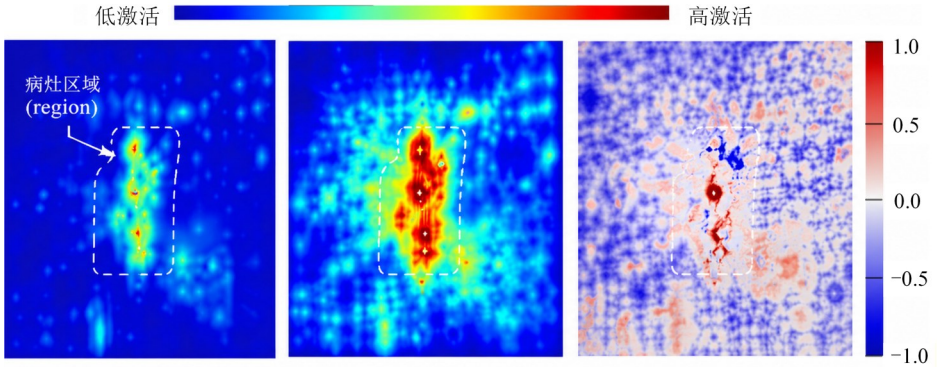
在小锈病目标占主导的图像中，3×3卷积核的选择概率接近1.0；中等尺寸灰斑病病灶的图像中，选择概率在3×3和5×5之间均匀分布；大面积叶斑病图像中，7×7的选择概率则明显抬升，这与核选择的预期一致。

图5为FDR对小目标特征响应的增强效果的热力图，目标对象为小型玉米锈病孢子簇。图5a为未加FDR时颈部P4特征的激活热力图，图5b为FDR输出特征的激活热力图，图5c为两者差值，即FDR额外引入的细节残差增强信息。

图5a未增强的P4热力图中，微小病斑的激活信号既弱又散，没有明显的峰值，和周围叶片纹理的区分度很低，后续处理中很容易被淹没。图5b与原始P4特征相比，FDR在微小病斑的边界和纹理区域显著增强了细节响应，FDR重构后的热力图显示，同一病斑区域的激活强度明显增大，响应峰值变得集中且突出，病斑特征被凸显出来。图5c残差热力图展现了FDR额外引入的细节信息，病斑位置呈现清晰的高激活信号。

3.4　边缘部署性能评估

为评估LiteFocus-Net在边缘设备上的实际推理



a. 原始 P4 特征激活热力图 b. FDR 增强后特征激活热力图 c. 高频残差热力图

图 5 FDR 模块对微小目标的特征增强效果

Fig. 5 Feature enhancement results of the FDR module

性能，在瑞芯微 RK3588 和英伟达 Jetson Orin 两类常用边缘平台上进行了部署实验。模型采用了 8 位整型（8-bit Integer, INT8）和 16 位浮点型（16-bit Floating Point, FP16）量化方法以降低计算复杂度并适配端侧硬件，量化过程分别基于 RK3588 官方提供的神经网络处理器软件开发工具包（Neural Processing Unit Software Development Kit, NPU SDK）和英伟达张量推理引擎（NVIDIA TensorRT）。其中，INT8 量化采用对称整数量化（映射至［−128，127］）及最小-最大范围校准以逼近 FP32 特征分布。针对硬件特性，模型在 RK3588 NPU 上仅激活 AKL-Block 单核卷积路径以避免多分支计算冗余，在 Jetson Orin 上则通过 TensorRT 离线校准保证 INT8 模型在推理中保持接近 FP32 的精度。实验结果如表 19 所示。

表 19 LiteFocus-Net 与 YOLOv11n 在 RK3588 和 Jetson Orin 上的量化推理性能及能效

Table 19 Quantized inference performance and energy efficiency of LiteFocus-Net and YOLOv11n on RK3588 and Jetson Orin

平台	模型	精度	延迟/ms	功耗/W	内存/MB	mAP@0.5/%	单帧能耗/(mJ/frame)	能效比/(帧/s/W)
RK3588 CPU	YOLOv11n	INT8	12.7	0.9	92	69.8	11.43	87.5
RK3588 CPU	LiteFocus-Net	INT8	12.1	0.9	88	71.5	10.89	91.8
RK3588 NPU	LiteFocus-Net	INT8	7.5	1.2	152	71.5	9.00	111.1
Jetson Orin	YOLOv11n	INT8	5.3	5.1	124	69.7	27.03	37.0
Jetson Orin	LiteFocus-Net	FP16	6.8	6.2	198	74.7	42.16	23.7
Jetson Orin	LiteFocus-Net	INT8	4.8	4.9	122	71.5	23.52	42.5

LiteFocus-Net 在 RK3588 NPU 上推理仅需 7.5 ms，换算成帧率约 133 帧/s；在 Jetson Orin 上做了 INT8 量化后延迟只有 4.8 ms，帧率可达 208 帧/s。量化后的 mAP 为 71.5%，明显高于同等条件下量化的 YOLOv11n（69.8%）。

3.5 复杂田间条件下的鲁棒性测试

为进一步验证 LiteFocus-Net 在复杂田间环境中的鲁棒性，从测试集中筛选并标注强光、弱光、逆光、局部遮挡和不同拍摄角度条件下的图像子集。考虑到一张田间图像可能同时具有遮挡、逆光和倾斜拍摄等特征，各子集允许存在少量交叉，因此子集图像数量之和不等于测试集图像总数。所有模型均采用相同训练参数和置信度阈值，在原始测试图像上进行评估，并采用 seed=42 条件下训练获得的代表性模型权重进行测试。结果如表 20 所示。

由结果可见，复杂环境会不同程度地降低模型检测精度。逆光和局部遮挡条件下的性能下降较为明显。逆光会使叶片局部区域过曝或形成明显阴影，降低病斑纹理与叶片背景之间的对比度；遮挡则会破坏病斑、虫体和孢子簇的完整轮廓，增加定位难度。

五类复杂条件下，LiteFocus-Net 的 mAP@0.5 均高于 YOLOv11n。其中，在局部遮挡条件下，LiteFocus-Net 的 mAP@0.5 由 68.7% 提升至 72.6%，AP_s 由 25.1% 提升至 30.4%。在逆光条件下，AP_s

表 20　LiteFocus-Net与YOLOv11n在不同复杂田间条件下的鲁棒性测试结果

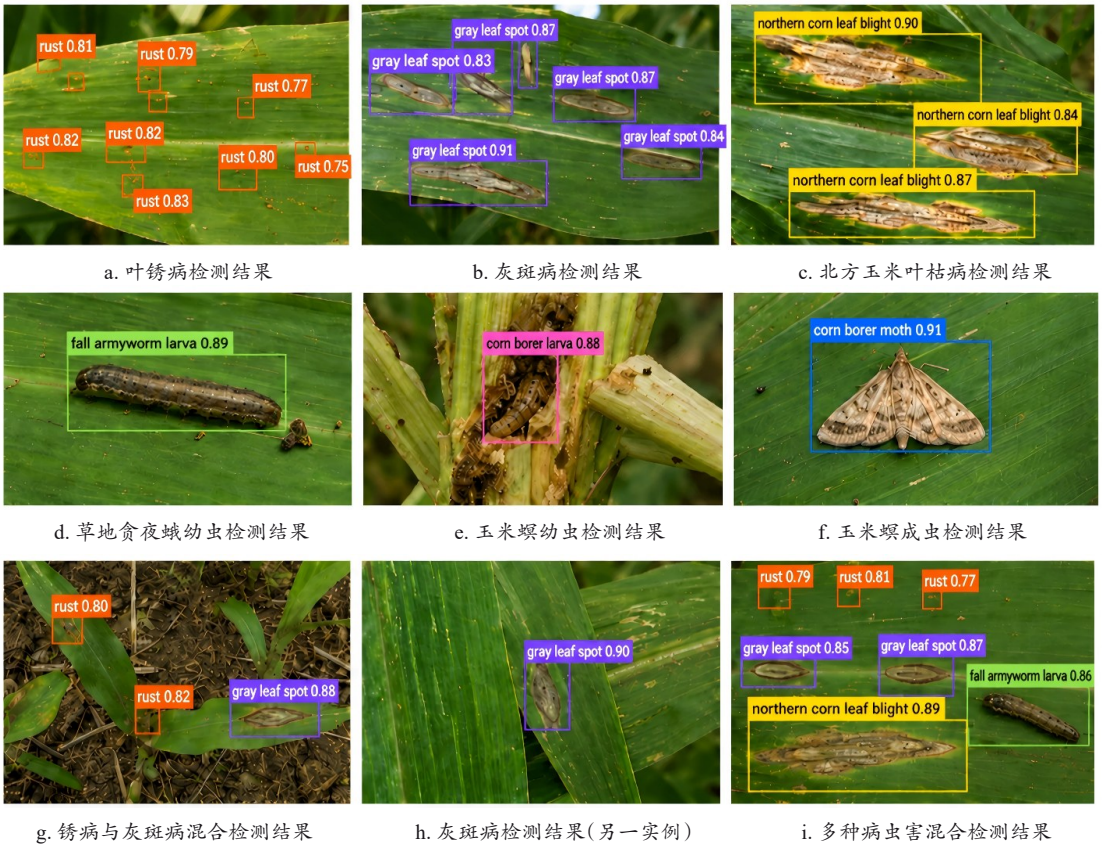
Table 20　Robustness evaluation of LiteFocus-Net and YOLOv11n under different complex field conditions

测试条件	图像数/张	目标实例数/个	YOLOv11n mAP@0.5/%	LiteFocus-Net mAP@0.5/%	YOLOv11n AP_s/%	LiteFocus-Net AP_s/%
全部测试集	2 049	4 827	72.2	75.0	28.1	32.8
强光条件	312	708	70.8	74.1	26.9	31.7
弱光条件	286	655	69.6	73.0	25.8	30.8
逆光条件	241	536	67.9	71.7	24.6	29.8
局部遮挡	354	812	68.7	72.6	25.1	30.4
不同拍摄角度	328	741	71.0	74.4	27.2	32.0

由 24.6% 提升至 29.8%。结果表明，AKL-Block 能够根据输入特征分布调整有效感受野，FDR 可以补偿复杂光照和遮挡条件下部分衰减的边缘与纹理信息，SAGB-Loss 则提高了小目标样本在回归训练中的贡献。因此，LiteFocus-Net 在复杂田间环境下具有较好的鲁棒性。同时，当逆光导致目标区域严重过曝，或者遮挡面积较大且极小孢子簇与叶脉纹理高度重叠时，模型仍存在少量漏检。这与本研究定性检测分析中的结果一致。后续研究将通过更丰富的光照增强策略、遮挡模拟增强以及跨区域数据扩展进一步提升模型在极端条件下的稳定性。

3.6　检测效果定性分析

图 6 展示了模型在密集细小锈斑、草地贪夜蛾幼虫、叶枯病复杂背景等典型田间场景下的检测结果。对密集锈斑和叶片边缘处的小病斑，模型能够保持较高召回率；但当孢子面积小于约 10×10 像素且与叶脉高亮纹理重叠时，仍存在少量漏检。



a. 叶锈病检测结果　b. 灰斑病检测结果　c. 北方玉米叶枯病检测结果

d. 草地贪夜蛾幼虫检测结果　e. 玉米螟幼虫检测结果　f. 玉米螟成虫检测结果

g. 锈病与灰斑病混合检测结果　h. 灰斑病检测结果（另一实例）　i. 多种病虫害混合检测结果

图 6　LiteFocus-Net在田间复杂场景下的定性检测玉米叶片结果

Fig. 6　Qualitative detection results of LiteFocus-Net in complex field scenarios

在田间复杂环境下评估 LiteFocus-Net 时，模型对密集锈斑、草地贪夜蛾幼虫及叶枯病区域表现出较高的检测精度，但仍存在漏检和误检现象。

如图 7 可见，图中绿色框表示成功检测目标，

红色圆圈或箭头标示漏检位置，直观展示模型在极小目标、背景干扰、尺度混合、遮挡和光照变化等条件下的局限性。结合AKL-Block、FDR和SAGB-Loss模块分析可知，AKL-Block的单核激活限制了不同尺度目标的同时响应，FDR在极小目标上的重建能力有限，而SAGB-Loss的面积增强因子和层级权重依赖于训练集目标分布，当迁移至不同作物或拍摄条件时，可能引起中、大目标定位精度下降。

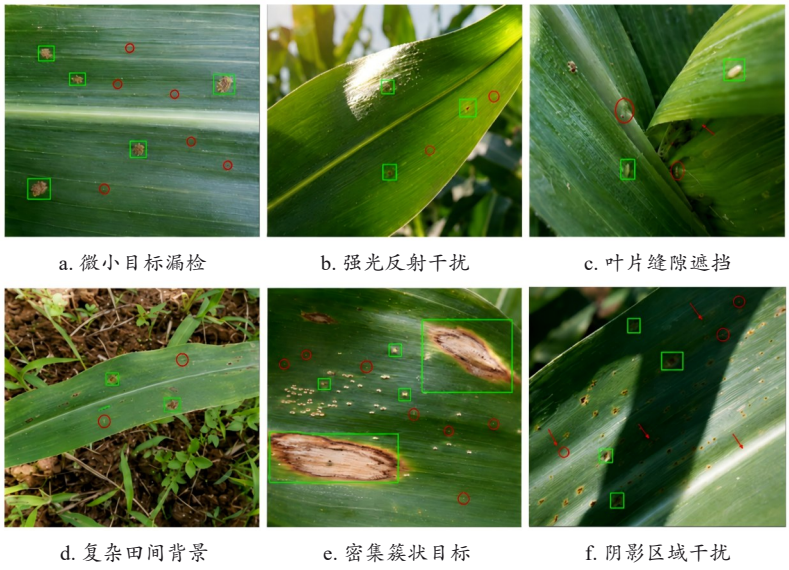


图7 LiteFocus-Net玉米病虫害检测典型漏检异常分析

Fig. 7 Typical missed-detection anomaly analysis of Litefocus-Net for maize disease and pest detection

4 讨 论

LiteFocus-Net在保持轻量化规模的同时，有效提高了玉米叶片病虫害小目标检测精度。其主要原因在于，模型改进并不是简单增加网络深度或检测头数量，而是针对小目标检测中的3个关键环节进行了约束：输入尺度差异、深层细节衰减和训练梯度分配。与直接扩大模型规模相比，这种定向优化方式更适合边缘设备算力有限的田间应用场景。

4.1 模块有效性分析

实验结果表明，LiteFocus-Net的各模块设计不仅在各自环节发挥了定向优化作用，也展现出明显的正向协同作用。

传统多尺度卷积结构通常通过并行分支获得不同感受野，但在推理阶段同时计算多个分支容易增加部署负担。AKL-Block采用输入级卷积核选择方式，使每次推理只激活1个候选卷积核，动态适应玉米锈病小斑点、虫卵和较大病斑等尺度差异明显的目标。消融实验显示，引入AKL-Block后，为主干网络提供了丰富的多尺度特征表达，小目标AP_s提升了2.0个百分点；同时计算量由6.3 GFLOPs下降至5.0 GFLOPs，在保留多尺度适应能力的同时有效减少了冗余计算。

FDR模块针对深层特征中微小目标的细节衰减问题，没有对全部P5通道进行高成本重构，而是将特征分为结构分支和细节分支，仅对细节分支进行轻量化重构，从而恢复病斑边缘、虫体轮廓及孢子纹理等关键信息。当其与AKL-Block结合后，AKL-Block提供的多尺度感受野增强了FDR对微小细节的可恢复性。消融实验中，单独使用FDR小目标AP_s提升0.6个百分点；而组合使用AKL-Block与FDR后，AP_s由30.1%提升至31.4%，在展现出+0.7的正向协同增益的同时，计算量仅由5.0增加至5.3 GFLOPs，说明该设计在精度提升和计算代价之间取得了较合理的平衡。

SAGB-Loss损失函数则从训练监督层面增强了小目标权重。其通过层级权重和面积增强因子共同调节回归损失，采用连续指数形式减少了尺度边界附近的权重突变。与硬阈值加权相比，这种设计使小目标样本在训练中获得更充分优化的同时，没有明显牺牲AP_m和AP_l。引入SAGB-Loss后，完整LiteFocus-Net的AP_s达到32.8%，较YOLOv11n基线提高了4.7个百分点，进一步验证了结构模块设计与损失函数优化在小目标检测上的系统协同效应。

边缘部署结果进一步说明了该模型的应用前

景。在RK3588和Jetson Orin等平台上，LiteFocus-Net经过INT8量化后仍保持较高推理速度，能够满足田间巡检、移动端识别和精准防控场景中的实时性要求。结合定性检测结果可以看出，模型对密集锈斑、草地贪夜蛾幼虫和复杂背景下的叶枯病区域具有较稳定的检测能力，说明其具备一定的端侧应用潜力。

4.2 局限性分析

尽管LiteFocus-Net在小目标检测精度、参数量和计算复杂度之间取得了较好的平衡，但该方法仍存在一定局限性。主要体现在以下几个方面。

1）AKL-Block采用图像级或特征图级的动态核选择策略，其根本局限在于：单次核选择是针对整张特征图作出的整体决策，无法在同一幅图像的不同空间区域分别采用不同大小的感受野。前文尺度混合样本实验也表明，AKL-Block在小目标精度上略低于完整并行分支结构，这说明图像级单核选择虽然具有较好的轻量化优势，但在多尺度目标同时出现的复杂场景中仍存在表达能力上限。

2）FDR模块的细节恢复能力受到深层特征信息保留程度的限制，当目标尺寸极小，或者目标在连续下采样过程中已经被严重压缩时，P5特征中可能不再保留足够的原始信息。在这种情况下，FDR即使采用可学习重构，也只能放大已有的微弱响应，无法恢复已经完全消失的边缘和纹理信息。

3）SAGB-Loss的效果与目标尺度分布密切相关，当模型应用于其他作物、不同拍摄设备或不同作业场景时，目标在图像中的像素尺度可能发生明显变化。例如，无人机远距离巡检图像中的病斑尺度通常更小，而近距离手持设备采集图像中的虫体和病斑可能占据更大的像素区域。如果直接沿用当前参数，模型可能对部分小目标增强不足，或者对中等目标赋予过高权重。过度放大小目标梯度虽然可能进一步提高AP_s，但也可能干扰中、大目标的回归优化，造成AP_m或AP_l下降。

5　结　论

本研究提出了面向玉米叶片病虫害小目标检测的轻量化增强框架LiteFocus-Net。该方法以YOLOv11n为基线，通过AKL-Block、FDR模块和SAGB-Loss分别改善尺度适应能力、深层细节恢复能力和小目标回归监督强度，从而提升轻量化模型对微小病斑、虫体、虫卵和孢子簇等目标的检测效果。

在包含10 324张田间玉米叶片图像和8类病虫害目标的数据集上，LiteFocus-Net取得了75.0%的mAP@0.5，参数量为2.52 M，计算量为5.3 GFLOPs。与YOLOv11n相比，AP_s提高4.7个百分点，同时计算量降低15.9%。实验结果表明，该方法能够在不扩大模型规模的情况下提升小目标检测性能，并具备在农业边缘设备上部署的可行性。后续研究将重点从区域级动态核选择、跨区域样本扩展和视频连续检测三个方向进一步提升模型的泛化能力和长期应用稳定性。

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] KALUNGA P, KUNDA D. Deep learning applications in maize disease detection: A systematic review of trends, gaps, and future research[J]. Zambia ICT Journal, 2025, 9(2): 31-37.

[2] 李董, 汤启国, 王红波, 等. 农作物重大病虫害预警与应急防控技术研究现状与趋势[J]. 智能化农业装备学报(中英文), 2025, 6 (1): 25-40.
LI D, TANG Q G, WANG H B, et al. Current status and trends of research on early warning and emergency control technology for major crop pests and diseases[J]. Journal of Intelligent Agricultural Mechanization, 2025, 6(1): 25-40.

[3] FENG Z R, SHI R M, JIANG Y H, et al. SPD-YOLO: A method for detecting maize disease pests using improved YOLOv7[J]. Computers, Materials & Continua, 2025, 84(2): 3559-3575.

[4] MENG Y, ZHAN J W, LI K S, et al. A rapid and precise algorithm for maize leaf disease detection based on YOLO MSM[J]. Scientific Reports, 2025, 15: 6016.

[5] NAKATUMBA-NABENDE J, MURINDANYI S. Deep learning models for enhanced in-field maize leaf disease diagnosis[J]. Machine Learning with Applications, 2025, 20: 100673.

[6] KHAN Z, SHEN Y, LIU H. ObjectDetection in agriculture: A comprehensive review of methods, applications, challenges, and future directions[J]. Agriculture, 2025, 15(13): 1351.

[7] SAPKOTA R, FLORES-CALERO M, QURESHI R, et al. YOLO advances to its genesis: A decadal and comprehensive review of the You Only Look Once (YOLO) series[J]. Artificial Intelligence Review, 2025, 58(9): 274.

[8] WANG C Y, YEH I H, LIAO H Y M. YOLOv9: Learning what you want to learn using programmable gradient information[C]// Computer Vision-ECCV 2024. Cham, Germany: Springer, 2025: 1-21.

[9] WANG A, CHEN H, LIU L H, et al. YOLOv10: Real-time end-to-end object detection[C]// Advances in Neural Information Processing Systems 37. Vancouver, BC, Canada: Neural Information Processing Systems Foundation, Inc. (NeurIPS), 2024: 107984-108011.

[10] KHANAM R, HUSSAIN M. YOLOv11: An overview of the key architectural enhancements[PP/OL]. arXiv: 2410.17725, 2024.

[11] ZHANG X Y, ZHOU X Y, LIN M X, et al. ShuffleNet: An extremely efficient convolutional neural network for mobile devices [C]// 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway, New Jersey, USA: IEEE, 2018: 6848-6856.

[12] HOWARD A G, ZHU M L, CHEN B, et al. MobileNets: Efficient convolutional neural networks for mobile vision applications[PP/OL]. arXiv: 1704.04861, 2017.

[13] HAN K, WANG Y H, TIAN Q, et al. GhostNet: More features from cheap operations[C]// 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, New Jersey, USA: IEEE, 2020: 1577-1586.
[14] LIN T Y, DOLLAR P, GIRSHICK R, et al. Feature pyramid networks for object detection[C]// 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, New Jersey, USA: IEEE, 2017: 936-944.
[15] LIU S, QI L, QIN H F, et al. Path aggregation network for instance segmentation[C]// 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway, New Jersey, USA: IEEE, 2018: 8759-8768.
[16] TAN M X, PANG R M, LE Q V. EfficientDet: Scalable and efficient object detection[C]// 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, New Jersey, USA: IEEE, 2020: 10778-10787.
[17] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal loss for dense object detection[C]// 2017 IEEE International Conference on Computer Vision (ICCV). Piscataway, New Jersey, USA: IEEE, 2017: 2999-3007.
[18] ZHANG H, ZHANG S J. Focaler-IoU: More focused intersection over union loss[PP/OL]. arXiv: 2401.10525, 2024.
[19] ZHANG Y F, REN W Q, ZHANG Z, et al. Focal and efficient IOU loss for accurate bounding box regression[J]. Neurocomputing, 2022, 506: 146-157.
[20] TONG Z J, CHEN Y H, XU Z W, et al. Wise-IoU: Bounding box regression loss with dynamic focusing mechanism[PP/OL]. arXiv: 2301.10051, 2023.
[21] ZHANG W, HU H, SUN Y Q, et al. AgriPest-YOLO: A lightweight detection network for dense small agricultural pests under complex field backgrounds[J]. Agriculture Communications, 2022, 1(1): 100017.
[22] ZHENG Y, ZHENG W H, DU X. A lightweight rice pest detection algorithm based on improved YOLOv8[J]. Scientific Reports, 2024, 14: 29888.
[23] HUANG Y K, ZHONG Y L, ZHONG D C, et al. Pepper-YOLO: An lightweight model for green pepper detection and picking point localization in complex environments[J]. Frontiers in Plant Science, 2024, 15: 1508258.
[24] LU Y Z, LIU P F, TAN C. MA-YOLO: A pest target detection algorithm with multi-scale fusion and attention mechanism[J]. Agronomy, 2025, 15(7): 1549.
[25] CEN X Y, LU S L, QIAN T T. YOLO-LCE: A lightweight YOLOv8 model for agricultural pest detection[J]. Agronomy, 2025, 15(9): 2022.
[26] YANG H, SHENG S Y, JIANG F, et al. YOLO-SDW: A method for detecting infection in corn leaves[J]. Energy Reports, 2024, 12: 6102-6111.
[27] JANG E, GU S X, POOLE B. Categorical reparameterization with gumbel-softmax[PP/OL]. arXiv: 1611.01144 , 2017.
[28] HU J, SHEN L, SUN G. Squeeze-and-excitation networks[C]// 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway, New Jersey, USA: IEEE, 2018: 7132-7141.
[29] SHI W Z, CABALLERO J, HUSZAR F, et al. Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network[C]// 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, New Jersey, USA: IEEE, 2016: 1874-1883.
[30] ZHAO Y A, LV 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.
[31] CAI H, LI J Y, HU M Y, et al. EfficientViT: Lightweight multi-scale attention for high-resolution dense prediction[C]// 2023 IEEE/CVF International Conference on Computer Vision (ICCV). Piscataway, New Jersey, USA: IEEE, 2023: 17256-17267.
[32] WANG C C, HE W, NIE Y, et al. Gold-YOLO: Efficient object detector *via* gather-and-distribute mechanism[C]// Advances in Neural Information Processing Systems 36. Vancouver, BC, Canada: Neural Information Processing Systems Foundation, Inc. (NeurIPS), 2023: 51094-51112.
[33] CHEN Y M, YUAN X B, WANG J B, et al. YOLO-MS: Rethinking multi-scale representation learning for real-time object detection[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2025, 47(6): 4240-4252.
[34] XU X Z, JIANG Y Q, CHEN W H, et al. DAMO-YOLO: A report on real-time object detection design[PP/OL]. arXiv: 2211.15444, 2022.
[35] LYU C Q, ZHANG W W, HUANG H A, et al. RTMDet: An empirical study of designing real-time object detectors[PP/OL]. arXiv.2212.07784, 2022.
[36] MEHTA S, RASTEGARI M. MobileViT: Light-weight, general-purpose, and mobile-friendly vision transformer[PP/OL] arXiv: 2110.02178, 2021.
[37] ZHANG J N, LI X T, LI J, et al. Rethinking mobile block for efficient attention-based models[C]// 2023 IEEE/CVF International Conference on Computer Vision (ICCV). Piscataway, New Jersey, USA: IEEE, 2023: 1389-1400.
[38] CHEN J R, KAO S H, HE H, et al. Run, don't walk: Chasing higher FLOPS for faster neural networks[C]// 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, New Jersey, USA: IEEE, 2023: 12021-12031.

LiteFocus-Net: A Three-Point Lightweight Enhancement Framework for Small-Target Detection of Corn Leaf Diseases and Pests

YANG Yu[1], ZHANG Yibo[2], MAO Bo[2], ZHANG Lei[2*]

(1. *College of Mechanical and Electronic Engineering*, *Nanjing Forestry University*, *Nanjing* 210037, *China*; 2. *School of Computer Science and Artificial Intelligence*, *Nanjing University of Finance and Economics*, *Nanjing* 210023, *China*)

Abstract: [Objective] Early detection of corn leaf diseases and pests is important for field monitoring, precision pesticide application, and yield protection. In practical field environments, early lesions, insect bodies, eggs, and spore clusters often occupy only a very

small part of the image. Their visual appearance is affected by weak texture, irregular shape, and complex backgrounds, making the visual boundary between disease targets and normal leaf tissue less distinguishable. When lightweight detectors based on the YOLO series are applied, fixed convolution kernels lack flexibility for targets with large scale differences. After repeated downsampling, small objects lose representation in deep feature maps, while larger targets dominate the regression loss. These factors make it difficult to balance detection accuracy and deployment efficiency on edge devices. To improve the recognition capability for small targets under complex field conditions, LiteFocus-Net, a lightweight enhancement framework, was developed based on YOLOv11n. **[Methods]** The framework improved the baseline model from three closely related aspects: adaptive backbone feature extraction, deep feature detail recovery, and scale-aware regression supervision. In the backbone, an Adaptive Kernel Lightweight Block AKL-Block was designed to replace part of the original feature extraction structure. Instead of computing multiple convolution branches simultaneously during inference, AKL-Block used global average pooling and a lightweight gating module to estimate the selection probabilities of three candidate depthwise separable convolution kernels with different receptive fields. A Gumbel-Softmax strategy was used during training to keep the kernel selection process differentiable, while only the kernel with the highest selection probability is activated during inference. In this way, the model could adjust its effective receptive field according to the input feature distribution without introducing the redundant computation usually caused by parallel multi-branch structures. To alleviate the loss of small-object details in deep layers, a Feature Decomposition and Reconstruction module (FDR) was introduced after the P5 feature layer. The deep feature was divided into a structure branch and a detail branch. The structure branch was upsampled by bilinear interpolation to preserve global semantic information, such as leaf shape, lesion distribution, and large disease regions. The detail branch used a lightweight pixel-shuffle reconstruction operation to enhance local responses related to lesion edges, insect contours, and spore textures. The reconstructed feature was then fused with the corresponding P4 feature in the neck network, allowing the model to reuse deep semantic information while selectively strengthening detail-sensitive responses. For the regression loss, a Scale-Aware Gradient Boosting Loss, termed SAGB-Loss, was constructed to increase the training contribution of small targets. The loss combined feature-level weighting and target-area-aware weighting. The feature-level term assigned larger weights to shallow layers that are more closely related to small-object detection, while the area-aware term used a continuous exponential function to smoothly enhance the regression gradients of small targets. Experiments were carried out on a field corn leaf image dataset containing 10 324 images and eight categories of disease and pest targets. The dataset was divided into training, validation, and test subsets at a ratio of 6:2:2. **[Results and Discussions]** LiteFocus-Net achieved 74.97% mean average precision (mAP) at an Intersection over Union (IoU) threshold of 0.5, with 2.52 million parameters and 5.3 GFLOPs. Compared with the YOLOv11n baseline, the average precision for small objects (AP_s) increased from 28.1% to 32.77%, giving an improvement of 4.7 percentage points, while the computational cost decreased by 15.9%. Ablation experiments confirmed the contribution of each component. AKL-Block reduced the overall computational cost and improved small-target accuracy, FDR further enhanced detail-sensitive features from deep layers, and SAGB-Loss increased the regression contribution of small targets without causing an obvious decline in medium- and large-object detection. Comparisons with Inception-style and selective-kernel multi-scale modules showed that AKL-Block achieved a more favorable accuracy-complexity trade-off under lightweight deployment constraints. Comparisons with full-channel pixel-shuffle reconstruction also indicated that FDR provided a practical balance between detail recovery and additional computational cost. Deployment tests on RK3588 and Jetson Orin NX further showed that LiteFocus-Net maintained real-time inference capability after INT8 quantization, which indicated its potential for field inspection and agricultural edge intelligence applications. **[Conclusions]** The results suggest that LiteFocus-Net improves small-target detection of corn leaf diseases and pests without relying on a larger model scale. Future work will focus on expanding cross-region, cross-variety, and cross-growth-stage samples, and on exploring region-level dynamic kernel selection, video-based continuous detection, and weakly supervised annotation strategies to improve robustness in long-term field deployment.

Keywords: disease and pest detection; small-target detection; lightweight model; YOLOv11n; adaptive kernel selection; feature reconstruction

Foundation items: National Key Research and Development Program of China (2024YFF1105504)

First author: 1. YANG Yu, E-mail: 2514056148@qq.com; 2. ZHANG Yibo, E-mail: feptsss@gmail.com
***Corresponding author:** ZHANG Lei, E-mail: zhanglei@nufe.edu.cn