基于改进YOLOv12s的辣椒叶片病虫害轻量化检测方法

姚晓通*，曲绍业

（兰州交通大学 电子与信息工程学院，甘肃兰州 730070，中国）

摘 要：［**目的/意义**］针对自然环境干扰下检测模型对辣椒叶片病虫害的特征提取不充分、容易忽视目标物体的边缘信息，以及小块病斑与虫害病灶易漏检等问题，本研究提出一种轻量化辣椒叶片病害检测算法，即YOLO-MDFR（You Only Look Once Version 12-MDFR）。［**方法**］基于YOLOv12s模型做出改进。首先用两个堆叠的3×3的深度可分离卷积代替一个5×5的深度可分离卷积以改进MobileNetV4，并将其代替YOLOv12s的原始骨干网络实现骨干网络轻量化。其次为提高小目标物体的特征提取能力，提出了多维频域互补自注意力机制模块（Dimensional Frequency Reciprocal Attention Mixing Transformer, D-F-Ramit）。最后利用D-F-Ramit与RAGConv（Residual Aggregation Gate-Controlled Convolution）重新设计颈部网络，增强模型的特征融合能力和信息传递能力。基于以上改进提出YOLO-MDFR目标检测算法。［**结果和讨论**］实验结果表明，本研究提出的YOLO-MDFR模型在实验数据集上的平均识别精确度达到95.6%，与YOLOv12s模型相比，平均识别精度提高了2.0%，同时参数量下降了61.5%，计算量下降了68.5%，帧率达到43.4帧/s。［**结论**］本研究通过系统性的架构优化，在保持模型轻量化的同时显著提升了检测性能，实现了计算效率与检测精度的最佳平衡。

关键词：YOLO；叶片病虫害检测；MobileNetV4；轻量化模型；注意力机制

中图分类号：TP391.4;S436.611 **文献标识码：**A **文章编号：**SA202506005

引用格式：姚晓通, 曲绍业 . 基于改进YOLOv12s的辣椒叶片病虫害轻量化检测方法[J]. 智慧农业(中英文), 2026, 8(1): 1-14. DOI：10.12133/j.smartag.SA202506005

YAO Xiaotong, QU Shaoye. Lightweight Detection Method for Pepper Leaf Diseases and Pests Based on Improved YOLOv12s[J]. Smart Agriculture, 2026, 8(1): 1-14. DOI：10.12133/j.smartag.SA202506005 (in Chinese with English abstract)

0 引 言

辣椒是世界上发展前景良好的重要经济作物，可作为食用蔬菜，也可制成辣酱和其他食物的调味品。由于全球气温持续上升，高温高湿的气候条件对辣椒种植构成了严峻挑战。在这种条件下，蓟马容易暴发，且被蓟马啃食的叶片也容易被辣椒花叶病毒与黄单胞杆菌属的细菌入侵，最后导致叶片黄化并生成病斑，这些病害会导致辣椒叶片早衰脱落，显著影响辣椒的商品品质，甚至造成植株成片枯死、产量下降[1]。细菌性叶斑病与烟草花叶病毒引起的叶片黄化病在田间环境下容易混淆，识别困难。因此，及时准确地检测染病的辣椒植株对于提高辣椒产量和质量、控制农药的使用量和成本具有重要意义。

传统的辣椒叶片病害与虫害检测主要依赖人工观察和判断，这种方法不仅效率低，且易受检测者专业知识、经验、视力和疲劳程度等人为因素干扰，导致检测结果主观性强、稳定性不足[2]。计算机视觉技术的快速发展，为农业病害智能检测提供了高效客观的解决方案[3]，能够不依赖人工干预就能精准地识别并定位发病植株，从而为优化农药喷洒策略提供关键数据支持，不仅显著提升施药精度，还能减少农药使用量，推动绿色农业的可持续发展[4]。因此，开发兼具高效性和鲁棒性的农作物叶片病虫害检测方法已成为精准农业的重要方向，也是农业智能化与绿色化转型的关键。

收稿日期：2025-06-03

基金项目：国家自然科学基金项目（51567014）；甘肃省科技计划项目（22JR5RA797）

*通信作者：姚晓通，博士，副教授，研究方向为物联网与智能检测、机器人与视觉控制、大数据与人工智能等。E-mail：545323755@qq.com

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

基于传统机器学习的方法通常依赖人工特征提取操作，需从叶片图像中提取颜色、纹理、形状等关键特征，再借助支持向量机（Support Vector Machine, SVM）、随机森林或K近邻等分类器实现病害识别[5]。针对柑橘病虫害检测问题，SHARIF等[6]提出的多SVM集成分类模型表现出色，识别准确率可达97%，优于传统方法。EBRAHIMI等[7]提出了一种结合移动机器人与SVM算法的智能检测方案，适用于温室草莓病虫害的识别任务。然而，传统机器学习方法存在明显的局限性：其一，特征提取工程高度依赖人工经验与先验知识，无法实现端到端的特征自动学习；其二，手工设计的特征表达能力有限，难以有效表征叶片病虫害在植株不同生长阶段的细微病理变化。此外，在辣椒叶片病虫害的实际识别中，人工方式也面临准确率与效率的双重挑战：识别方面，相似病症易混淆；效率方面，单人日均检测量仅500株，与规模化基地（如10万株/地块）的检测需求严重不匹配。

深度学习技术以人工神经网络[8]为理论基础，其层次化的网络结构能够对文本、音频和图像等多源数据进行多尺度特征提取，通过从低阶感知特征到高阶语义特征的渐进式学习，展现出优异的表征学习能力，在目标检测领域得到了广泛应用。石天怡等[9]针对传统苹果叶片病害分类方法精准性差的问题，提出一种基于改进ConvNeXt的苹果叶片分类算法CALDNet，识别叶片病害的精确率与原始ConvNeXt模型相比提高了4.63%。魏天宇等[10]融合双向特征金字塔网络，改进YOLOv5s的深层特征融合架构，提升检测精度，算法模型能够在白天强光、傍晚弱光的环境下提高对辣椒的检测精度。王震鲁等[11]提出一种基于改进YOLOv5辣椒目标检测模型，该模型能够有效识别被遮挡的辣椒，平均检测准确率达到91%。邹玮和岳延滨团队[12]创新性地将YOLOv2检测框架与ResNet50深度卷积网络相融合，开发了一种新型混合模型，可实现辣椒叶片蚜虫的高效分类与精确定位，系统的识别准确率高达96.49%。李佳松等[13]提出一种C-ResNet-50算法模型以改善在复杂自然背景下对马铃薯叶片病虫害的识别效果，识别马铃薯叶片病虫害的平均准确率达90.83%，较原始模型提升了1.84%。对比邹玮等[12]（YOLOv2模型，样本量2 619张）、李佳松等[13]（C-ResNet-50，样本量1 080张），本研究数据集规模是同类研究的2～4倍，样本代表性更强。

现有研究虽然在模型架构和检测方法方面取得了一定进展，但存在两个关键局限性。首先，这些方法未能充分考虑边缘设备的计算资源约束。其次，基于深度学习的辣椒叶片病害与虫害检测仍面临若干技术挑战，具体表现为：1）特征提取不充分。由于辣椒叶片病虫害在形态学和色彩学上具有高度多样性，加之复杂的自然环境下光照变化、叶片遮挡及背景噪声等因素的干扰，使得现有模型难以获取充分且鲁棒的特征表示，容易忽视目标物体的边缘信息，亟需开发更具判别力的深层细粒度特征提取方法。2）较小的辣椒叶片病斑不容易被检测。叶片病斑形状较小容易受作物生长的复杂环境干扰，在算法模型的训练过程中容易忽略对小目标提取的重要性，以及复杂的自然环境导致错误检测的异常样本对模型的影响。

YOLOv12相比YOLOv5与YOLOv2来说具有更高的检测精度、更快的推理速度与更强的泛化能力，因此本研究在YOLOv12s的基础上进行改进，提出辣椒叶片病害与虫害轻量化检测算法，保证一定检测精度的情况下，缩小模型以便部署在算力紧张的设备上。将YOLOv12s的骨干网络替换为改进的MobileNetV4，该网络采用轻量级架构设计，进一步优化参数量，在保持精度的同时显著降低每秒浮点运算次数（Floating Point Operations per Second, FLOPs），适合资源受限的设备。为解决关键特征（边缘、小病灶）提取精度不足，模型易忽略目标边缘信息，以及模型对小目标物体特征的忽略问题，提出了多维频域互补注意力机制模块，该模块能够抑制无关信息，同时能提取局部和全局特征以提高检测精度，增强网络多尺度特征融合的能力。最后，提出残差聚合门控卷积模块，该模块通过多个高阶空间交互，捕捉特征之间的复杂关系，从而增强特征融合的效果。

上述改进使模型在精度与理论计算量上取得了良好平衡。但核心问题是理论上的轻量化是否等同于部署上的高效率。为回答这一问题，并验证本模型在真实农业场景中的部署效率，本研究还进行了算法部署实验。

1 实验材料

1.1 辣椒叶片病虫害数据采集

数据采集地点为甘肃省天水市甘谷县辣椒种植核心区（新兴镇康家滩村、安远镇王马村和大像山镇杨场村），覆盖3个典型种植地块，土壤类型为

黄绵土，种植方式为露地垄作；辣椒品种为当地主栽品种“甘谷尖辣椒”（*Capsicum annuum L. var. conoides*（Mill.）Irish），3月份到5月份，辣椒植株处于以营养生长为主的时期，同时也是病害高发期；采集日期为2024年3月15日—5月20日，每日采集时段为9：00—11：00（光照强度5 000～8 000 lux，避免晨露）、15：00—17：00（光照强度6 000～9 000 lux，避免强光过曝）；采集设备为小米14 Pro手机，后置主摄5 000万像素，焦距50 mm，光圈f/1.7，手动对焦至叶片表面，分辨率像素为4 000×3 000。采集到的叶片种类如图1所示，图1a为被蓟马侵染的叶片、图1b为烟草花叶病毒黄化病叶片、图1c为健康叶片、图1d为细菌性叶斑病叶片。



a. 被蓟马侵染叶片　b. 烟草花叶病毒黄化病叶片　c. 健康叶片　d. 细菌性叶斑病叶片

图1　辣椒叶片数据集图像

Fig. 1　Pepper leaf dataset images

1.2　辣椒叶片病虫害数据集构建

利用LabelImg软件（1.8.6版本）对收集到的辣椒叶片病虫害图像进行人工标注，标注后的图像如图2所示。为了更好地评估模型的泛化能力，在构建数据集时，将数据集按照7：2：1的比例划分为训练集、验证集、测试集。辣椒叶片病虫害数据集规模如表1所示。



a. 被蓟马侵染叶片　b. 烟草花叶病毒黄化花叶病叶片　c. 健康叶片　d. 细菌性叶斑病叶片

图2　辣椒叶片病虫害标注图像

Fig. 2　Annotated images of pepper leaf diseases and pests

表1　辣椒叶片病虫害数据集

Table 1　Dataset of leaf diseases and pests of peppers

辣椒叶片病虫害类别	训练集样本数量/张	测试集样本数量/张	验证集样本数量/张	样本总数量/张
健康叶片	2 193	95	223	2 511
细菌性叶斑病叶片	2 283	96	186	2 565
烟草花叶病毒黄化花叶病叶片	2 076	102	175	2 353
被蓟马侵染叶片	2 127	129	232	2 488

为解决训练集样本量不足导致的模型泛化能力弱，以及部分病害类别样本分布不均导致的模型偏向性问题，对原始训练集进行数据增强与数据平衡处理[14]，具体策略如下。

采用5类针对性数据增强方法扩充训练集，所有增强操作仅作用于训练集，验证集与测试集保持原始状态以确保评估客观性，具体参数设置如下。

① 随机旋转。旋转角度范围0°～360°，旋转步长15°，应用概率0.8。

② 缩放变换。缩放比例范围0.8～1.2倍，应用概率0.7。

③ 水平翻转。沿图像水平中轴线翻转，应用概率0.5。

④ 高斯噪声。添加方差为0.01的高斯噪声，模拟自然环境下的图像噪声干扰，应用概率0.3。

⑤ 亮度调整。亮度系数范围0.7～1.3倍，应用概率0.6。

通过上述增强策略，训练集样本总量从原始的8 679张扩充至17 358张，有效丰富了样本的形态、光照与噪声多样性，降低模型过拟合风险。

从表1可知，原始训练集中烟草花叶病毒黄化病（Mosaic）类别样本量最少（2 076张），与样本量最多的细菌性叶斑病（Xanthomonas）2 283张相差207张，类别间样本差异可能导致模型对小众类别的检测精度偏低。为解决该问题，本研究采用“过采样+合成少数类过采样技术（Synthetic Minority Over-sampling Technique, SMOTE）”的组合策略平衡数据：对Mosaic类原始样本进行随机过采样，补充100张重复样本；基于SMOTE，通过计算Mosaic类样本的特征相似度，合成124张新样本，确保合成样本与原始样本特征分布一致。

处理后，各类别训练集样本量均稳定在2 200±50张（Healthy为2 200张、Xanthomonas为2 230张、Mosaic为2 200张、Thrips为2 180张），消除了类别间样本分布差异对模型训练的影响。

2 算法模型

2.1 多维频域互补自注意力机制模块(D-F-Ramit)

本研究提出多维频域互补自注意力机制模块（Dimensional Frequency Reciprocal Attention Mixing Transformer, D-F-Ramit），其创新点在于融合通道、空间、频域三维注意力，突破了现有如卷积块注意力模块（Convolutional Block Attention Module, CBAM）等仅涵盖二维的注意力机制局限。

现有研究指出，结合频域与空间域可弥补传统卷积神经网络局限于局部像素信息的不足[15]。尽管通道注意力[16]能通过动态权重增强关键特征，却忽视了空间信息与局部细节；空间注意力[17, 18]虽可突出关键区域，但在复杂纹理或目标与背景相似时易产生误判。二者结合（如CBAM[19]）仍存在高频细节丢失的问题。而傅里叶变换与神经网络结合，能有效提取频域中的纹理、周期性特征，并增强模型抗干扰能力[20]。基于此，本研究整合空间、通道与频域注意力，构建多维度频域互补机制，以综合解决上述问题，模块结构如图3所示。

D-F-Ramit模块通过空间自注意力机制（Space Self-Attention, SpaceSA）和通道自注意力机制（Channel Self-Attention, ChannelSA）捕捉局部与全局的空间特征与通道之间的依赖关系，两种自注意力机制设计如图4所示。

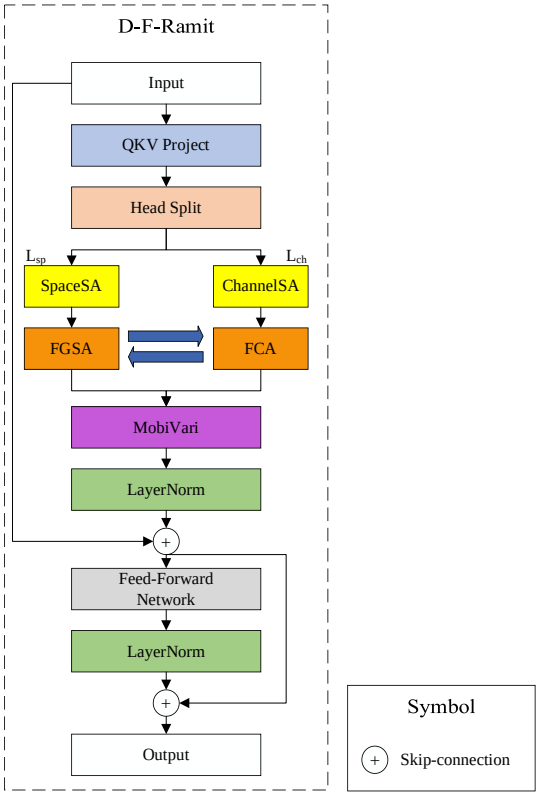


图3 D-F-Ramit模块架构

Fig. 3 Architecture of D-F-Ramit

SpaceSA和ChannelSA的设计分别参考SwinIR[21]的局部窗口自注意力机制和Restormer[22]的转置注意力机制，并且为SpaceSA和ChannelSA配备了不同数量的多头注意力Lsp和Lch（Lsp+Lch=L），将两者的输出结果通过傅里叶变换获取频域上的信息之后通过反傅里叶变换还原相位信息，再将两者通过MobiVari模块进行拼接获得最终的互补注意力。MobiVari模块是MobileNet系列的变体[23, 24]，采用轻量化设计实现两种注意力机制的拼接。*Attn*注意力获取过程由公式（1）获得。

$$Attn = MobiVari(Concat[FGSA(SpaceSA), FCA(ChannelSA)]) \quad (1)$$

式中：*SpaceSA*表示经过空间自注意力机制处理得到的特征图；*ChannelSA*表示经过通道自注意力机制处理得到的特征图；*FGSA*表示从空间的角度进行傅里叶变换处理；*FCA*表示从通道的角度进行傅里叶变换；*MobiVari*用于拼接*FGSA*和*FCA*产生的特征图。每个自注意力机制及其对应的头由公式（2）～公式（5）获得。

$$SpaceSA = P_{sp}(Concat[head_1^{sp}, ..., head_{L_{sp}}^{sp}]) \quad (2)$$

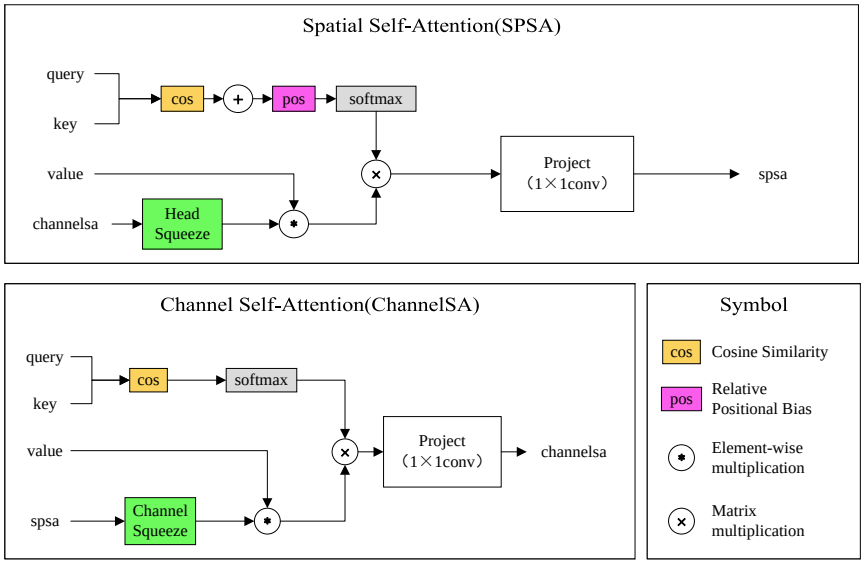


图4　自注意力机制设计思路

Fig. 4　Design of Concepts for self-attention

$$ChannelSA = P_{\text{ch}}(Concat[head_{L_{sp}+1}^{ch}, ..., head_{L}^{ch}]) \tag{3}$$

$$head_i^{\text{sp}} = \text{Softmax}(\cos(Q_i^{\text{sp}}, (K_i^{\text{sp}})^T)/\tau + B)V_i^{\text{sp}} \tag{4}$$

$$head_i^{\text{ch}} = \text{Softmax}(\cos(Q_i^{\text{ch}}, (K_i^{\text{ch}})^T)/\tau + B)V_i^{\text{ch}} \tag{5}$$

式中：Q_i^{sp}，K_i^{sp}，V_i^{sp} 和 Q_i^{ch}，K_i^{ch}，V_i^{ch} 分别为SpaceSA模块和ChannelSA模块的query，key，value；cos用于计算余弦相似度[25]；$B \in \mathbb{R}^{M^2 \times M^2}$ 表示相对位置的偏差[25]；τ 是一个可训练的标量，其值设定大于0.01；$head$ 为单个注意力计算单元；P_{sp}，P_{ch} 表示重塑和映射。

幅度谱与相位谱获取过程由公式（6）和公式（7）获得。

$$F_{fft} = FFT2D(X) \tag{6}$$

$$A = |F_{fft}(X)|, \varphi = \angle F_{fft}(X) \tag{7}$$

式中：F_{fft} 为快速傅里叶变换的一个结果值；A 为幅度；φ 为相位谱，Rad；$X \in \mathbb{R}^{C \times H \times W}$。幅度谱表示了图像中不同频率分量的能量分布，反映了图像的全局结构信息（纹理和边缘），相位谱则记录了频率分量的空间位置关系，决定了图像的局部细节和边缘信息。

公式（8）到公式（12）为傅里叶变换处理过程，公式（10）表示通道注意力频域转换：将通道权重映射至频域，强化关键频率分量；公式（12）表示空间注意力频域转换：将空间权重映射至频域，抑制背景噪声频率。

$$Z = \frac{1}{H \times W}\sum_{u=0}^{H-1}\sum_{v=0}^{W-1}A \tag{8}$$

$$\omega = \sigma(W_1\delta(W_0 Z)), W_0 \in \mathbb{R}^{\frac{C}{r} \times C}, W_1 \in \mathbb{R}^{C \times \frac{C}{r}} \tag{9}$$

$$F_{FCA} = \omega \cdot X \tag{10}$$

$$M_{spacial} = \sigma(\text{Conv}(A)) \tag{11}$$

$$X_{FGSA} = M_{spacial} \otimes X \tag{12}$$

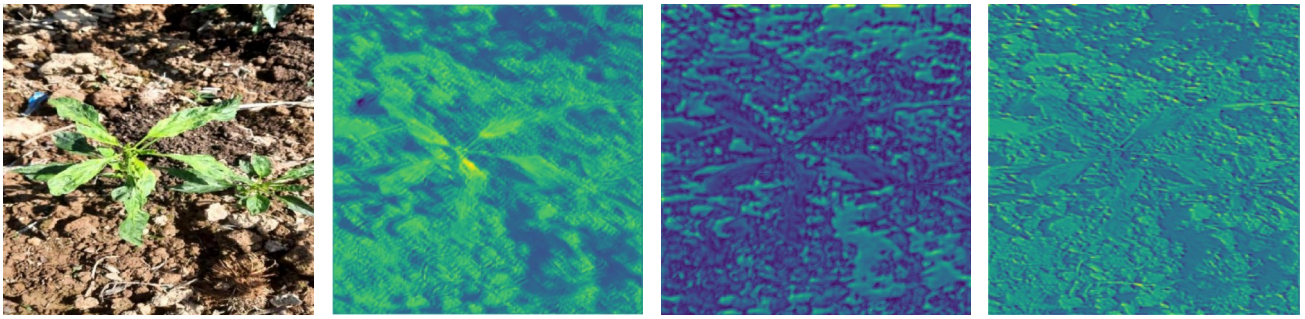
式中：C 为通道数，个；H 为图像高度，像素；W 为图像宽度，像素；Z 表示图像整体频域活跃程度；ω、W、$M_{spacial}$ 表示特征图；σ 为 $sigmoid$ 激活函数；δ 为 $ReLU$ 激活函数。

公式（10）和公式（12）分别为通道自注意力机制和空间自注意力机制进行傅里叶变换处理，捕获频域信息特征，最终通过公式（13）进行还原，为反傅里叶变换，作用为融合频域幅度与相位，还原空域特征用于残差连接，避免特征丢失。

$$FGSA, FCA = F^{-1}(A \odot \varphi) \tag{13}$$

与传统注意力机制模块相比，D-F-Ramit不仅能捕获通道间的依赖关系，自动识别输入图像或者特征图中的重要区域，还能提取频域中的纹理特征。因此D-F-Ramit可以从通道、空间、频域三个维度进行特征的提取，在通道和空间上处理局部细节，在频域上处理全局结构，并且通过轻量化MobiVar模块在这三个维度上进行相互补充，使得该模块显著弥补了单一通道注意力机制的不足，比空间注意力机制更加高效。图5为经过不同注意力机制模块处理图像后对结果的可视化，其中图5a为在自然强光条件下被蓟马感染的辣椒叶片。由图5b

和图5c可知，经过通道与空间注意力机制处理后，特征图中物体边缘非常模糊，模块忽略掉了物体的边缘信息，但根据图5d可知，加入傅里叶变换后的注意力机制模块，相比通道注意力机制、空间注意力机制，在强光照条件下，模块能更好地提取目标边缘信息，物体边缘更加清晰，模块的提取特征能力明显增强。



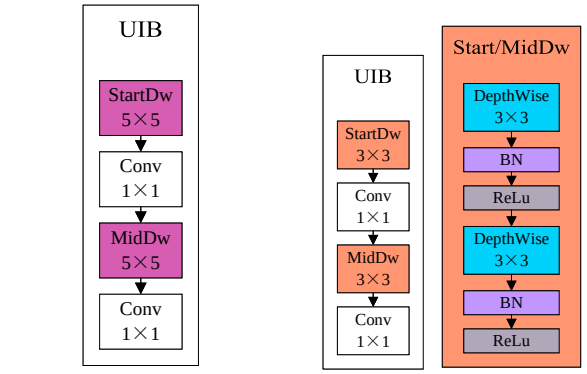
a. 被蓟马感染的叶片 b. 通道注意力机制 c. 空间注意力机制 d. D-F-Ramit

图5 不同注意力机制对强光下叶片边缘特征的增强效果对比

Fig. 5 Comparison of the enhancement effects of different attention mechanisms on leaf edge features under strong illumination

2.2 优化MobileNetV4作为骨干网络

MobileNetV4[26]是Google公司针对高效计算而设计的轻量化网络，本研究将MobileNetV4进行改进并替换为YOLOv12s的骨干网络，实现骨干网络的轻量化。但由于识别的辣椒叶片病害与虫害的种类不到百种，特征空间复杂度较低，无需构建复杂的网络架构来完成特征提取与分类，因此针对这一任务特点对原网络结构进行了调整，结构如图6所示。本研究用两个堆叠的3×3的深度可分离卷积替换通用倒置瓶颈模块（Universal Inverted Bottleneck, UIB）模块中的5×5的深度可分离卷积，虽然网络层数变多，但是单次卷积的计算量显著减少，计算效率更高。此外，改进后的UIB模块还可以提取更加复杂的特征，同时保持相同的空间覆盖范围。



a. MobileNetV4原始UIB模块 b. 改进后的UIB模块

图6 MobileNetV4中UIB模块的轻量化结构改进

Fig. 6 Lightweight structural improvement of the UIB module in MobileNetV4

参数量方面5×5卷积核的参数量为$C_{in}\times C_{out}\times 5\times 5$，而两个3×3卷积核的参数量为$C_{in}\times C_{mid}\times 3\times 3+C_{mid}\times C_{out}\times 3\times 3$，若假设中间通道数$C_{mid}$=64，输入输出通道$C_{in}$=$C_{out}$=64，则5×5卷积核的参数量为102 400，两个3×3卷积核的参数量为73 728，由此可见，理论上用两个3×3卷积核进行串联代替一个5×5卷积核，参数量减少约28%。

在计算量方面，一个5×5卷积核其计算量为$H\times W\times C_{in}\times C_{out}\times 5\times 5$，显然对每个输出对象需要25次乘加运算，而两个3×3卷积核的参数量为$H\times W\times C_{in}\times C_{mid}\times 3\times 3+H\times W\times C_{mid}\times C_{out}\times 3\times 3$每次卷积仅需9次乘加运算，因此总计算量更少。

2.3 残差聚合门控卷积模块

YOLOv12s模型在进行目标检测时，需要对不同层次的特征图进行融合，但传统的融合方式很容易造成梯度流动阻塞和特征信息丢失的问题。残差连接能够缓解梯度消失和梯度爆炸问题，支持更深的网络结构，同时还能加速模型的收敛。姜舒等[27]利用门控卷积，通过设置门控权重控制器，以控制捕捉局部细节信息和全局上下文信息的比例程度，提高了特征融合能力，解决了肿瘤边界模糊的问题。因此，本研究结合残差连接[28]，改进聚合方式，以及门控卷积[29]提出残差聚合门控卷积（Residual Aggregation Gate-Controlled Convolution, RAGConv）模块，来提高模型的特征融合能力和特征信息的传递能力，设计思路如下。

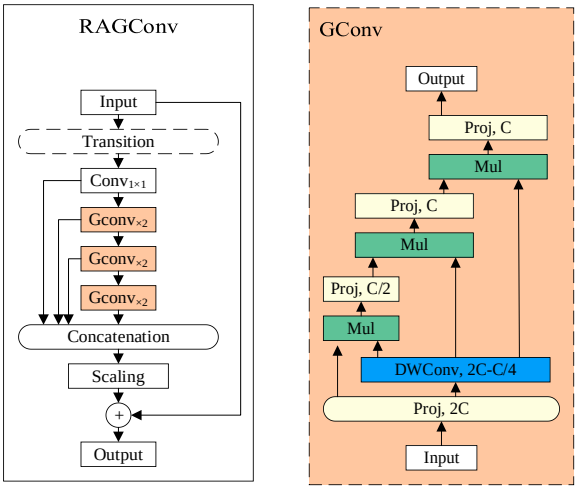
残差连接：在每个模块中引入从输入到输出的残差连接，并通过缩放因子控制残差连接的强度。

这种设计能够有效缓解梯度阻塞的问题，提升模型的收敛性。

特征聚合方法：通过一个过渡层调整通道维度，生成单个特征图，然后通过后续块进行处理和拼接，形成一个瓶颈结构。此方法不仅能保留原始的特征集成能力，还减少了计算成本。

通过递归实现高阶空间的交互：首先使用多个线性投影层获得一组投影特征，然后通过递归的方式执行门控卷积，最后将最后一次递归步骤的输出，送入投影层以获得最终结果。

RAGConv的优势：相较于YOLOv12s原有的A2C2f特征融合模块，RAGConv通过“残差聚合+门控卷积”设计，减少特征融合过程中的信息损失。



a. RAGConv模块的整体结构　　b. 门控卷积模块结构

图7　RAGConv模块结构

Fig.7　Module structure of RAGConv

2.4　辣椒叶片病虫害识别算法模型

为使算法能够部署在算力小的嵌入式设备上，本研究提出一种改进优化的网络架构——YOLO-MDFR，算法模型框架图如图8所示。本研究基于YOLOv12s[30]算法进行改进，YOLOv12s的选择理由：对比不同代际YOLO系列基础模型，YOLOv2[31]和YOLOv5s[32]虽在速度上有优势，但精度与泛化能力不足；而YOLOv12s泛化能力更强，能为后续轻量化改进提供高精度基础，平衡“精度保留”与“效率优化”需求。

YOLO-MDFR模型主要在以下方面进行了优化。1）Backbone。采用改进的轻量级卷积神经网络MobileNetV4作为骨干特征提取网络，如图8所

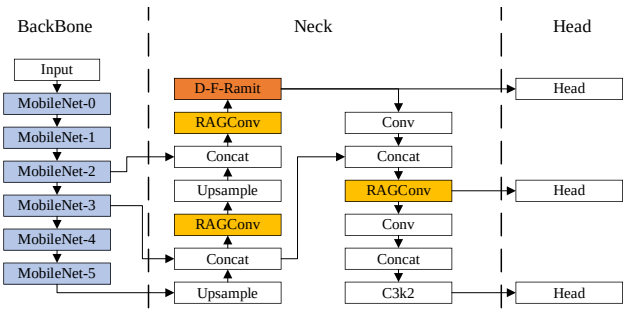


图8　YOLO-MDFR网络架构

Fig. 8　Network architecture of the YOLO-MDFR

示，蓝色模块部分即为改进后的MobileNetV4的网络层次，在确保检测精度的前提下，实现了模型计算效率的显著提升。2）Neck。提出RAGConv模块，即图8中的黄色模块，增强模型添加注意力机制后的收敛能力和稳定性，同时减小特征融合的计算量。在采样过程中添加多维频域互补自注意力机制模块，即图8中的红色模块，使得模型在特征融合的过程中减少信息损失，提高对小目标检测的检测精度，同时减小计算量。

综上所述，本研究致力于在保持原有检测效率和准确率的基础上，通过重新设计网络结构实现算法轻量化，并针对性地提升模型在复杂场景下的鲁棒性表现，特别是强化对小尺度目标的特征提取能力。

3　结果与分析

3.1　实验环境

实验环境配置NVIDIA GeForce RTX 4060的GPU，深度学习框架为PyTorch 1.12.0，训练时选择图像输入像素大小为640×640，batch_size为16，epochs为500，优化器为Adam，学习率设置为0.001，由于快速傅里叶变换涉及复数，所以训练模型时amp选择性关闭。

3.2　评价指标

针对模型的轻量化改进，采用计算量（GFLPOs），参数量（Params），平均精度（Mean Average Precision, mAP）作为评价指标，为满足实际部署需求，除常规指标外，特别将处理速度帧率（Frames Per Second, FPS）纳入评估体系，以衡量模型的工程适用性。

模型在嵌入式端的部署实验，采用平均单帧延迟、稳定帧率、idle功耗、推理功耗、1小时连续

检测功耗作为评价指标，评估模型在嵌入式设备上的性能。

3.3 骨干网络对比试验

通过系统的骨干网络对比评估，改进后的MobileNetV4架构被证实具有优于YOLOv12s的原始骨干网络的性能表现。实验结果如表2所示，直接采用标准MobileNetV4替换YOLOv12s默认的骨干网络时，虽然mAP下降了4.4个百分点，但是参数量缩小了5.5 M，同时计算量也减少了13.7 GFLOPs。将改进的MobileNetV4代替DarkNet-53网络，相较于未改进的MobileNetV4，mAP增长了2.6个百分点，参数量下降了0.7 M，计算量下降了0.2 GFLOPs，同时FPS上升0.5帧/s。因此，改进的MobileNetV4网络，以损失1.8个百分点的检测精度为代价，缩小了模型的大小、参数量和计算量，使得YOLOv12模型进一步轻量化。

表2 YOLOv12不同骨干网络对比实验结果

Table 2 Results of YOLOv12 comparison experiment of different backbone networks

网络模型	骨干网络	mAP%0.5	Params/M	GFLOPs/G	FPS/(帧/s)
YOLOv12s	原始网络	93.6	9.1	19.7	33.4
YOLOv12s	MobileNetV4	89.2	3.6	6.0	46.0
YOLOv12s	改进MobileNetV4	91.8	2.9	5.8	46.5

3.4 D-F-Ramit与现有注意力机制的性能对比

为进一步验证D-F-Ramit模块在轻量化与检测性能上的综合优势，在本研究自建的辣椒叶片病虫害数据集上，将其与当前主流的轻量化注意力机制（Squeeze-and-Excitation, SE）、高效通道注意力（Efficient Channel Attention, ECA）、CBAM进行对照实验，所有对比实验均基于相同的改进MobileNetV4骨干网络与训练超参数，通过多维度指标量化评估各机制的性能差异，结果如表3所示。

表3 D-F-Ramit与主流轻量化注意力机制的对比

Table 3 Comparison between D-F-Ramit and mainstream lightweight attention mechanisms

注意力机制	注意力维度(通道/空间/频域)	参数量/K	计算量/M FLOPs	小目标mAP提升/%	边缘信息保留能力
SE	通道	8.2	12.5	1.1	弱
ECA	通道	0.5	3.2	1.5	弱
CBAM	通道+空间	15.6	28.7	2.3	中等
D-F-Ramit	通道+空间+频域	18.3	35.1	3.8	强

由表3数据可知，单通道维度的注意力机制虽具备极致轻量化（参数量≤8.2 K，计算量≤12.5 M FLOPs），但因未覆盖空间与频域信息，对小目标mAP的提升幅度有限（仅1.1%～1.5%），且边缘信息保留能力较弱，这与传统通道注意力“忽视空间位置与高频细节”的固有缺陷一致。CBAM通过融合通道与空间维度，将小目标mAP提升至2.3%，边缘信息保留能力提升至中等，但受限于未引入频域特征，仍无法有效捕捉叶片病虫害的全局纹理与局部边缘关联。

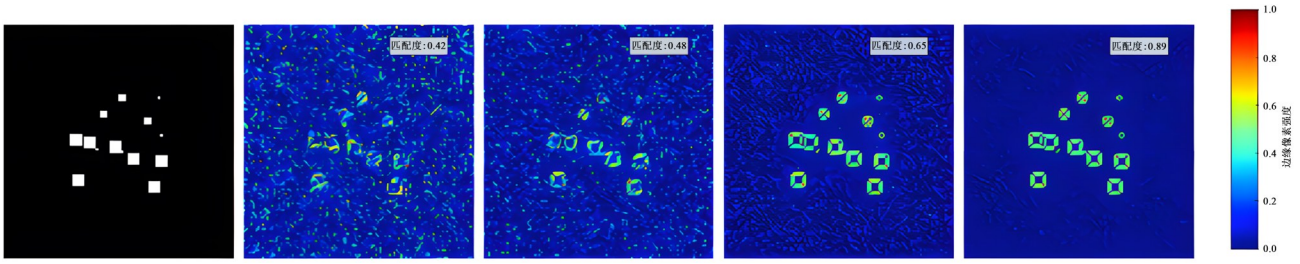
D-F-Ramit通过创新性融合“通道+空间+频域”三维特征，在轻量化基础上实现了性能突破：相较于CBAM，其参数量仅增加2.7 K（增幅17.3%），计算量增加6.4 M FLOPs（增幅22.3%），但小目标mAP提升幅度达3.8%，较CBAM提升1.5个百分点；结合图9的可视化结果进行量化分析，D-F-Ramit的边缘信息丢失率较CBAM降低68.57%，尤其在自然强光、叶片遮挡等复杂场景下，能更清晰地保留病斑边缘轮廓与细微纹理。

这一优势的核心源于频域特征引入D-F-Ramit通过傅里叶变换的幅度谱捕捉病害的全局纹理结构，通过相位谱还原局部边缘的空间位置关系，有效解决了传统注意力“高频细节丢失”的关键问题，为辣椒叶片小块病斑的精准检测提供了特征支撑。

3.5 消融实验

本研究建立YOLOv12s、YOLO-Mob、YOLO-MD、YOLO-MR和YOLO-MDFR进行消融实验以验证模型的有效性。在消融分析过程中为保证结果的可比性，所有对比模型均遵循相同的实验条件，结果如表4所示。

由表4可知，将改进的MobileNetV4替代原先的骨干网络，虽然能显著降低参数量与计算量，但是损失了一定的mAP。在此基础上加入了轻量化多维频域互补自注意力机制模块，实验结果显示mAP增长了0.5个百分点，但是增加了0.7 GFLOPs的计算量，因为参数量和计算量增加，所以FPS受到了影响，相较于YOLO-Mob模型FPS降低了7.8



a. 原始图像　b. SE处理后　c. ECA处理后　d. CBAM处理后　e. D-F-Ramit处理后

图9　不同注意力机制对边缘信息保留能力的可视化对比图

Fig. 9　Visualization of edge information preservation across different attention mechanisms

表4　自建数据集的YOLO−MDFR模型的消融实验

Table 4　Ablation experiments with YOLO-MDFR model for self-built datasets

名称	改进的MobileNetV4（含双3×3）	D-F-Ramit	RAGConv	mAP%0.5	Params/M	GFLOPs /G	FPS（帧/s）
YOLOv12s	×	×	×	93.6	9.1	19.7	33.4
YOLO-Mob	√	×	×	91.8	2.9	5.8	46.5
YOLO-MD	√	√	×	92.3	3.6	6.5	38.7
YOLO-MR	√	×	√	90.8	3.4	5.5	56.4
YOLO-MDFR（未平衡数据）	√	√	√	93.8	3.5	6.2	43.4
YOLO-MDFR	√	√	√	95.6	3.5	6.2	43.4

注：√表示包含此模块；×表示不包含此模块。

帧/s。当改进的MobileNetV4与RAGConv模块进行结合时，相较于YOLOv12s模型，参数量减少了5.7 M，计算量也减少了14.2 GFLOPs，同时FPS上升了23帧/s，代价仅是损失了2.8个百分点的mAP。

数据平衡的影响如图10所示，对比“YOLO-MDFR（未平衡数据）”与“YOLO-MDFR”两组实验可知：YOLO-MDFR（未平衡数据）mAP为93.8%（表4），且各类别mAP极差达8.2%（其中Mosaic类mAP最低，仅89.5%）；经过数据平衡处理后，YOLO-MDFR模型mAP提升至95.6%（表4），平均增幅1.8%，同时各类别mAP极差缩小至4.3%，Mosaic类检测精度显著提升4.1个百分点（从89.5%提升至93.6%）。这一结果证明，“过采样+SMOTE合成”的平衡策略能有效解决小众类别样本不足的问题，避免模型对多数类别的偏向性，提升模型对各类病害的均衡检测能力。

YOLO-MDFR在数据集上的可视化实验结果如图11所示。由图11a可知，随着置信度的提高，精度也严格递增，最终趋近于1.0，由实验结果可知，该模型的优势在于，在高置信度（≥0.927）的条件下表现完美，适合高精度要求的场景，结合之前的消融对比实验可知，该模型适合部署在一些算力小的嵌入式设备上，并且能保持一个较高的检测精度。

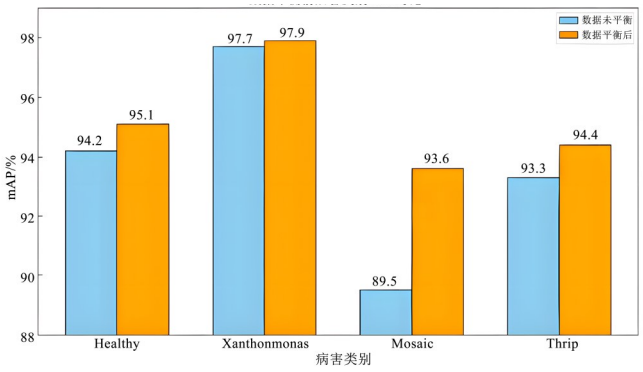


图10　自建辣椒数据集的数据平衡实验

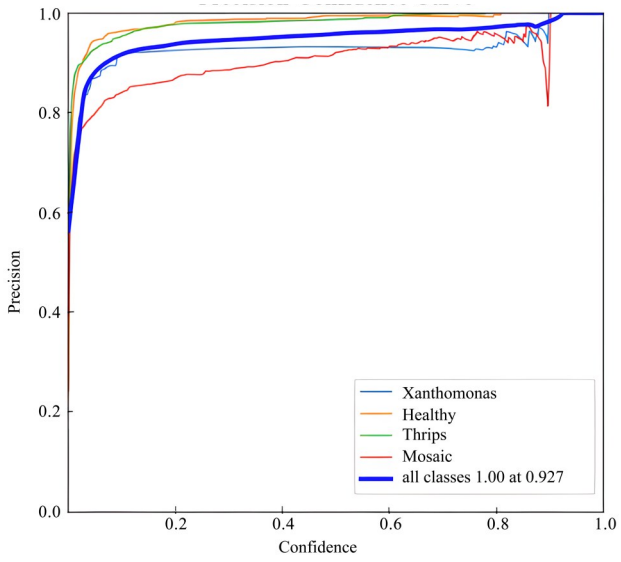
Fig. 10　Data balancing experiments on the custom-built dataset of pepper leaf diseases and pests

综上所述，经过数据平衡化后，YOLO-MDFR在自建辣椒叶片病虫害数据集上平均准确率达到95.6%与YOLOv12s模型相比检测精度提升了2.0%，参数量和计算量分别减少了61.5%和68.5%，YOLO-MDFR性能显著优于YOLOv12s。

3.6　YOLO−MDFR对相似病害的区分实验

为进一步验证YOLO-MDFR对相似病害的区分能力，图12 YOLO-MDFR的混淆矩阵显示蓟马虫害与烟草花叶病毒病的互误判率仅2.3%，细菌性叶斑病的正确识别率达98.1%且无漏检。

分析可知，模型通过D-F-Ramit模块的频域特



a. P_curve 图



b. 病害检测图

图 11 辣椒叶片病虫害检测结果

Fig. 11 Detection results of diseases and pests on pepper leaves

征，有效区分了“蓟马黄化斑（高频边缘不规则）”与“病毒黄化斑（高频边缘规则）”，精准解决人工识别的核心痛点，进一步印证了深度学习在辣椒叶片病虫害检测中的必要性。

3.7 不同网络模型性能对比实验

基于自建数据集，本研究设计了 5 组对比实验。实验组包含 Faster R-CNN、YOLOv11n、YOLOv10b、YOLOv9s、YOLOv8n、YOLOv5s、SSD 和 RT-DETR 八种代表性算法，以及本研究提出的 YOLO-MDFR 模型。在对比实验中所有模型均采用

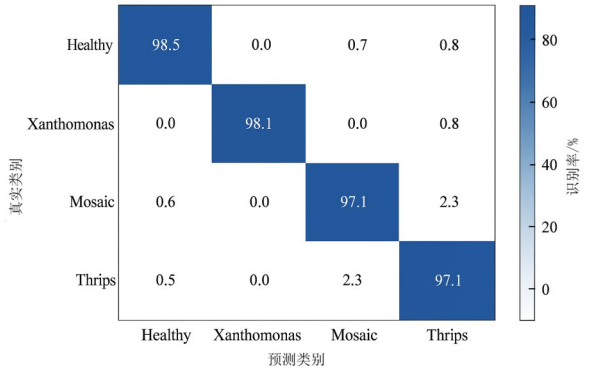


图 12 基于自建辣椒叶片病虫害数据集的 YOLO-MDFR 混淆矩阵

Fig. 12 Confusion matrix of YOLO-MDFR on the custom-built pepper leaf diseases and pests dataset

相同的超参数设置和训练策略进行公平比较。评价指标选择 mAP、Params、计算量、FPS，实验结果如表 5 所示。实验结果表明 Faster R-CNN 和 SSD 表现很差，识别精确度分别为 60.9% 和 60.7%，远不如 YOLO-MDFR。YOLOv5s 虽然识别精确度达到了 90.0%，但是与 YOLO-MDFR 相比，YOLO-MDFR 的参数量仅为 YOLOv5s 的 50.0%，计算量仅为 39.2%，且识别精确度比其高 5.6%。与其他算法模型相比，YOLO-MDFR 计算量最小，识别精度最高，FPS 为 43.4 帧/s，足以满足实时检测的需求，且模型性能优异，足够轻量化，适合部署在算力紧张的边缘设备上。

表 5 不同目标检测模型在自建辣椒数据集上的综合性能对比

Table 5 Comprehensive performance comparison of different object detection models on the custom-built pepper leaf diseases and pests dataset

模型名称	mAP%0.5	Params/M	GFLOPs/G	FPS/(帧/s)	Epochs	Batch_size
Faster R-CNN	60.9	137.0	370.2	5.3	500	16
YOLOv11n	90.8	2.6	6.6	53.1	500	16
YOLOv10b	92.2	20.0	99.4	39.5	500	16
YOLOv9s	89.7	7.5	12.5	55.0	500	16
YOLOv8n	91.9	3.2	8.9	85.0	500	16
YOLOv5s	90.0	7.0	15.8	68.1	500	16
SSD	60.7	26.0	62.7	46.0	500	16
RT-DETR	91.2	32.0	100.0	50.0	500	16
YOLO-MDFR	95.6	3.5	6.2	43.4	500	16

3.8 病斑尺度敏感性实验

为验证 YOLO-MDFR 对细小病斑的敏感性与识别精度，设计了病斑尺度敏感性实验。实验结果如表 6 所示，YOLO-MDFR 与 YOLOv12s 相比，对极

小目标的检测精度提高了3.3个百分点，小目标的检测精度提高了1.2个百分点，中目标的检测精度提高了1.5个百分点，大目标的检测精度提高了0.8个百分点，整体来看YOLO-MDFR对小块病斑的检测精度要优于YOLOv12s。

表6　YOLO-MDFR与YOLOv12s在不同尺度病斑上的检测精度对比

Table 6　Detection accuracy comparison of YOLO-MDFR and YOLOv12s on lesions at various scales

病斑尺度/像素	样本数量/张	YOLOv12s mAP/%	YOLO-MDFR mAP/%	提升幅度/个百分点
超小目标(<8×8)	126	22.8	28.5	5.7
极小目标(8—16)	348	30.2	33.5	3.3
小目标(16—32)	512	53.2	54.4	1.2
中目标(32—96)	897	75.1	77.6	1.5
大目标(≥96)	1 203	81.4	82.2	0.8

3.9　RK3588部署

为进一步验证YOLO-MDFR在边缘设备上的工程实用性，本研究针对瑞芯微生产的RK3588开发板开展功耗与延迟专项测试，量化模型在嵌入式场景下的“实时性-低功耗”表现，为田间移动检测设备（如植保无人机、手持检测终端）的部署提供数据支撑。RK3588平台上YOLO-MDFR模型的功耗、延迟与帧率实测结果如表7所示。

表7　RK3588平台上YOLO-MDFR模型的部署性能实测结果

Table 7　Measured deployment performance of the YOLO-MDFR model on the RK3588 platform

指标	数值	说明
平均单帧延迟	43.8 ms	全流程耗时，含图像预处理(20 ms，含尺寸缩放与归一化)+模型推理(23.8 ms，NPU 加速)
稳定帧率	22.8帧/s	由单帧延迟推算(1 000/43.8≈22.8帧/s)，满足田间实时检测要求(行业标准≥15帧/s)
idle 功耗	2.1 W	开发板通电但无图像输入时的静态功耗，反映设备待机能耗
推理功耗	3.5 W	连续输入图像并执行检测时的平均动态功耗，含NPU、CPU、内存协同工作能耗
1小时连续检测功耗	12.6 Wh	由推理功耗推算(3.5 W×3.6 h=12.6 Wh)，适配12 V/1 000 mAh(12 Wh)锂电池供电的移动设备

实测数据表明，YOLO-MDFR在RK3588平台上的实测性能表现优异，充分满足边缘设备对低功耗与实时性的核心需求。在实时性方面，模型平均单帧延迟为43.8 ms，稳定帧率达22.8帧/s，远超田间检测实时下限（15帧/s），可支持连续快速识别，有效避免因延迟导致的漏检。在功耗方面，模型推理功耗仅为3.5 W；待机功耗为2.1 W，适用于设备低功耗运行场景。综合计算，一小时连续检测仅需12.6 Wh，可使用小型锂电池供电，满足手持终端、植保无人机等移动设备的续航要求。

综上，YOLO-MDFR在嵌入式平台上的实测性能充分满足边缘设备“低功耗+实时性”的核心需求，为辣椒种植田间的移动化、便携化病害检测提供了可行的技术方案。

3.10　不同输入分辨率下的模型性能分析

输入分辨率是平衡目标检测模型精度与部署效率的核心参数。高分辨率能保留辣椒叶片早期微小病斑的细节特征，但会显著增加计算负担，不利于在RK3588等边缘设备上实现高效推理；低分辨率虽可提升速度，却易丢失关键特征，导致漏检增加。为确定YOLO-MDFR在不同场景下的最优输入尺寸，本节在同一实验环境下测试5组主流分辨率（320×320至736×736），系统分析分辨率对“精度-速度-效率”权衡关系的影响。不同输入分辨率下YOLO-MDFR模型的性能测试结果如表8所示，数据均为3次重复实验的平均值，确保结果可靠性。

“各类别mAP标准差”计算Healthy、Xanthomonas、Mosaic、Thrips四类样本mAP的离散程度，数值越小表示模型对不同病害的检测均衡性越好；FPS基于NVIDIA GeForce RTX 4060实测，若部署于RK3588开发板，帧率需参考3.7节的硬件适配比例（约为PC端的52%～55%）。

实验结果显示，1）分辨率提升对检测精度的影响呈现“边际效应递减”规律。如表8所示，输

表 8 不同输入分辨率下YOLO-MDFR的性能对比

Table 8 Performance comparison of YOLO-MDFR under different input resolutions

输入分辨率(宽×高)/像素	mAP%0.5	极小目标 mAP/%	各类别 mAP 标准差/%	Params/G	GFLOPs/G	FPS/(帧/s)
320×320	89.5	25.3	5.8	3.5	1.6	65.2
416×416	92.3	28.7	4.2	3.5	2.7	58.6
512×512	94.1	31.2	3.1	3.5	4.0	50.3
640×640	95.6	33.5	2.4	3.5	6.2	43.4
736×736	96.2	34.8	2.1	3.5	8.9	35.1

入分辨率从320×320提升至736×736时，模型整体mAP由89.5%增长至96.2%，累计提升6.7个百分点。其中，低至中分辨率阶段（320×320→512×512）提升显著，整体mAP与极小目标mAP分别提高4.6与5.9个百分点，因分辨率提升有效恢复了细微病斑的边缘与纹理信息；而中至高分辨率阶段（512×512→736×736）提升有限，整体mAP仅增长2.1个百分点，尤其在640×640之后继续增大分辨率对精度增益微弱，反映关键特征已得到充分提取。此外，各类别mAP标准差由5.8%降至2.1%，说明高分辨率有助于提升模型对样本较少类别的识别均衡性。2）分辨率与计算效率及推理速度呈现“线性负相关”关系。输入分辨率与计算量近似呈平方增长，736×736分辨率下的计算量（8.9 GFLOPs）为320×320（1.6 GFLOPs）的5.6倍；相应地，FPS由65.2下降至35.1帧/s。结合RK3588平台部署实测结果（640×640分辨率下FPS为22.8帧/s）可推断：若采用736×736分辨率，FPS将降至18～20帧/s，难以满足田间实时检测（≥20帧/s）需求；而512×512分辨率下预计FPS可达26～28帧/s，更有利于在精度与速度之间取得平衡。

结合实验结果，可根据部署场景选择合适的分辨率方案：

1）极致轻量化（低端MCU/大规模初筛）。推荐320×320分辨率，计算量仅1.6 G，FPS达65.2帧/s，mAP为89.5%，适合快速筛查场景（如无人机巡检）。

2）平衡型部署（嵌入式设备）。推荐512×512分辨率，计算量4.0 G（比640×640减少35.5%），FPS提升至50.3帧/s，mAP仅降低1.5个百分点至94.1%，满足多数田间检测需求。

3）高精度部署（实验室/诊断设备）。推荐640×640分辨率，mAP达95.6%，极小目标检测更优（33.5%），在工业级GPU上仍保持43.4帧/s实时性能。736×736分辨率仅建议用于极小病斑检测且算力充足的特定场景。

4 结 论

本研究提出一种面向复杂自然环境的辣椒叶片病虫害轻量化检测模型YOLO-MDFR。通过引入改进MobileNetV4骨干网络、多维频域互补自注意力机制（D-F-Ramit）与残差聚合门控卷积（RAG-Conv）模块，于自建数据集上实现了95.6%的mAP，性能显著优于主流检测模型。基于系统部署实验，本研究进一步提出面向不同应用场景的优化配置策略。同时指出三方面局限性：极端光照环境下模型鲁棒性不足、对新型病害泛化能力有限、高遮挡场景下小目标漏检率较高。未来计划通过融合环境光传感数据、结合域适应与半监督学习、引入双目视觉技术等途径进一步提升模型性能。

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] 司超国, 刘梦晨, 吴华瑞, 等. Chilli-YOLO: 基于改进YOLOv10的露地辣椒成熟度智能检测算法[J]. 智慧农业(中英文), 2025, 7(2): 160-171.
SI C G, LIU M C, WU H R, et al. Chilli-YOLO: an intelligent maturity detection algorithm for field-grown chilli based on improved YOLOv10[J]. Smart Agriculture, 2025, 7(2): 160-171.

[2] 邹玮, 岳延滨, 冯恩英, 等. 基于MobileNet V2的辣椒果实炭疽病识别及其应用[J]. 贵州农业科学, 2024, 52(9): 125-132.
ZOU W, YUE Y B, FENG E Y, et al. Identification of anthracnose in pepper fruit based on MobileNet V2 and the application [J]. Guizhou Agricultural Sciences, 2024, 52(9): 125-132.

[3] 兰玉彬, 单常峰, 王庆雨, 等. 不同喷雾助剂在植保无人机喷施作业中对雾滴沉积特性的影响[J]. 农业工程学报, 2021, 37(16): 31-38.
LAN Y B, SHAN C F, WANG Q Y, et al. Effects of different spray additives on droplet deposition characteristics during plant protection UAV spraying operations[J]. Transactions of the Chinese Society of Agricultural Engineering, 2021, 37(16): 31-38.

[4] 曹英丽, 张弘泽, 郭福旭, 等. 基于无人机遥感的农作物病害监测研究进展[J]. 沈阳农业大学学报, 2024, 55(5): 616-628.
CAO Y L, ZHANG H Z, GUO F X, et al. Research progress of crop disease monitoring based on UAV remote sensing[J]. Journal of Shenyang Agricultural University, 2024, 55(5): 616-628.

[5] 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.
[6] SHARIF M, KHAN M A, IQBAL Z, et al. Detection and classification of *Citrus* diseases in agriculture based on optimized weighted segmentation and feature selection[J]. Computers and Electronics in Agriculture, 2018, 150: 220-234.
[7] EBRAHIMI M A, KHOSHTAGHAZA M H, MINAEI S, et al. Vision-based pest detection based on SVM classification method [J]. Computers and Electronics in Agriculture, 2017, 137: 52-58.
[8] PUJARI J D, YAKKUNDIMATH R, BYADGI A S. Image processing based detection of fungal diseases in plants[J]. Procedia Computer Science, 2015, 46: 1802-1808.
[9] 石天怡, 南新元, 郭翔羽, 等. 基于改进ConvNeXt的苹果叶片病害分类算法[J]. 广西师范大学学报(自然科学版), 2025, 43(4): 83-96.
SHI T Y, NAN X Y, GUO X Y, et al. Improved ConvNeXt-based algorithm for apple leaf disease classification[J]. Journal of Guangxi Normal University (Natural Science Edition), 2025, 43(4): 83-96.
[10] 魏天宇, 柳天虹, 张善文, 等. 基于改进YOLOv5s的辣椒采摘机器人识别定位方法[J]. 扬州大学学报(自然科学版), 2023, 26(1): 61-69.
WEI T Y, LIU T H, ZHANG S W, et al. Research on pepper picking robot recognition and positioning method based on improved YOLOv5s[J]. Journal of Yangzhou University (Natural Science Edition), 2023, 26(1): 61-69.
[11] 王震鲁, 白涛, 李东亚, 等. 基于改进YOLOv5的绿辣椒目标检测方法[J]. 新疆农业科学, 2024, 61(12): 3032-3041.
WANG Z L, BAI T, LI D Y, et al. Green chili pepper target detection method based on improved YOLOv5[J]. Xinjiang Agricultural Sciences, 2024, 61(12): 3032-3041.
[12] 邹玮, 岳延滨, 冯恩英, 等. 基于YOLOv2的辣椒叶部蚜虫图像识别[J]. 山东农业大学学报(自然科学版), 2023, 54(5): 700-709.
ZOU W, YUE Y B, FENG E Y, et al. Image recognition of aphid on pepper leaves based on YOLOv2[J]. Journal of Shandong Agricultural University (Natural Science Edition), 2023, 54(5): 700-709.
[13] 李桂松, 黎敬涛, 杨艳丽, 等. 基于改进残差网络的马铃薯叶片病害识别[J]. 湖南农业大学学报(自然科学版), 2024, 50(6): 123-128.
LI G S, LI J T, YANG Y L, et al. Potato leaf disease identification based on improved residual networks[J]. Journal of Hunan Agricultural University (Natural Sciences), 2024, 50(6): 123-128.
[14] SHORTEN C, KHOSHGOFTAAR T M. A survey on image data augmentation for deep learning[J]. Journal of Big Data, 2019, 6(1): 60.
[15] 蔺志贤. 基于频域分析的多尺度卷积神经网络[J]. 河北软件职业技术学院学报, 2025, 27(2): 13-16, 43.
MAN Z X. Multi-scale convolutional neural networks based on frequency domain analysis[J]. Journal of Hebei Software Institute, 2025, 27(2): 13-16, 43.
[16] DING M Y, XIAO B, CODELLA N, et al. DaViT: Dual attention vision transformers[M]// Computer Vision-ECCV 2022. Cham: Springer Nature Switzerland, 2022: 74-92.
[17] YU Q H, XIA Y D, BAI Y T, et al. Glance-and-gaze vision transformer[C]// Neural Information Processing Systems. California, USA: NeurIPS Foundation, 2022.
[18] ZHU X Z, CHENG D Z, ZHANG Z, et al. An empirical study of spatial attention mechanisms in deep networks[C]// 2019 IEEE/CVF International Conference on Computer Vision (ICCV). Piscataway, New Jersey, USA: IEEE, 2019: 6687-6696.
[19] WOO S, PARK J, LEE J Y, et al. CBAM: Convolutional Block attention module[C]// Computer Vision-ECCV 2018. Cham, Germany: Springer, 2018: 3-19.
[20] LEE-THORP J, AINSLIE J, ECKSTEIN I, et al. Fnet: Mixing tokens with fourier transforms[EB/OL]. arXiv: 2105.03824, 2021.
[21] LIANG J Y, CAO J Z, SUN G L, et al. SwinIR: Image restoration using swin transformer[C]// 2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW). Piscataway, New Jersey, USA: IEEE, 2021: 1833-1844.
[22] ZAMIR S W, ARORA A, KHAN S, et al. Restormer: Efficient transformer for high-resolution image restoration[C]// 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, New Jersey, USA: IEEE, 2022: 5718-5729.
[23] HOWARD A G, ZHU M L, CHEN B, et al. MobileNets: Efficient convolutional neural networks for mobile vision applications[EB/OL]. arXiv: 1704.04861, 2017.
[24] SANDLER M, HOWARD A, ZHU M L, et al. MobileNetV2: Inverted residuals and linear bottlenecks[C]// 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway, New Jersey, USA: IEEE, 2018: 4510-4520.
[25] LIU Z, LIN Y T, CAO Y, et al. Swin transformer: Hierarchical vision transformer using shifted windows[C]// 2021 IEEE/CVF International Conference on Computer Vision (ICCV). Piscataway, New Jersey, USA: IEEE, 2021: 9992-10002.
[26] QIN D F, LEICHNER C, DELAKIS M, et al. MobileNetV4: Universal models for the mobile ecosystem[M]// Computer Vision – ECCV 2024. Cham: Springer Nature Switzerland, 2024: 78-96.
[27] 姜舒, 陈琨, 丁卫平, 等. Axial-FNet: 基于模糊卷积结合门控轴向自注意力的皮肤癌图像分割模型[J]. 智能科学与技术学报, 2025, 7(2): 221-233.
JIANG S, CHEN K, DING W P, et al. Axial-FNet: Skin cancer image segmentation model based on fuzzy convolution combined with gated axial self-attention[J]. Chinese Journal of Intelligent Science and Technology, 2025, 7(2): 221-233.
[28] SZEGEDY C, IOFFE S, VANHOUCKE V, et al. Inception-v4, inception-ResNet and the impact of residual connections on learning[C]// Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence. New York, USA: ACM, 2017: 4278-4284.
[29] YU J H, LIN Z, YANG J M, et al. Free-form image inpainting with gated convolution[C]// 2019 IEEE/CVF International Conference on Computer Vision (ICCV). Piscataway, New Jersey, USA: IEEE, 2019: 4470-4479.
[30] TIAN Y J, YE Q X, DOERMANN D. YOLOv12: Attention-centric real-time object detectors[EB/OL]. arXiv: 2502.12524, 2025.
[31] ZHAO J Y, QU J H. A detection method for tomato fruit common physiological diseases based on YOLOv2[C]// 2019 10th International Conference on Information Technology in Medicine and Education (ITME). Piscataway, New Jersey, USA: IEEE, 2020: 559-563.
[32] AJAYI O G, ASHI J, GUDA B. Performance evaluation of YOLO v5 model for automatic crop and weed classification on UAV images[J]. Smart Agricultural Technology, 2023, 5: 100231.

Lightweight Detection Method for Pepper Leaf Diseases and Pests Based on Improved YOLOv12s

YAO Xiaotong*, QU Shaoye

（*School of Electronics and Information Engineering*, *Lanzhou Jiaotong University*, *Lanzhou* 730070, *China*）

Abstract: [Objective] Pepper cultivation frequently faces challenges from diseases and pests, and early detection is critical for reducing yield losses. However, existing detection models often suffer from limitations such as insufficient feature extraction for subtle lesions, loss of edge information due to complex backgrounds, and high missed detection rates for small lesions. To address these issues, the YOLO-MDFR (You Only Look Once), a lightweight detection algorithm was proposed based on an enhanced YOLOv12s, specifically designed for accurate identification of pepper leaf diseases and pests in complex natural environments. **[Methods]** The dataset was established in the primary pepper cultivation zone of Gangu county, Tianshui city, Gansu province. The cultivated variety was the locally dominant *Capsicum annuum* L. var. *conoides* (Mill.). Data collection was conducted from March 15 to May 20, 2024. The collected samples included four categories of pepper leaves: healthy leaves, leaves damaged by thrips, leaves infected with tobacco mosaic virus exhibiting yellowing symptoms, and leaves affected by bacterial leaf spot. First, the original YOLOv12s backbone was replaced with an improved MobileNetV4 architecture to enhance lightweight performance while preserving feature extraction capability. Specifically, the original 5×5 standard convolutions in the bottleneck layers of MobileNetV4 were substituted with two sequential 3×3 depthwise separable convolutions. This design was based on the principle that two 3×3 convolutions achieve an equivalent receptive field (matching the 5×5 coverage) while reducing parameter count—depthwise separable convolutions further decompose spatial and channel convolution, minimizing redundant computations. Second, a novel dimensional frequency reciprocal attention mixing transformer (D-F-Ramit) module was introduced to enhance sensitivity to lesion boundaries and fine-grained textures. The module first converted feature maps from the spatial domain to the frequency domain using discrete cosine transform (DCT), capturing high-frequency components often lost in spatial-only attention. It then integrated three parallel branches: channel attention, spatial attention, and frequency-domain attention. Finally, a residual aggregation gate-controlled convolution (RAGConv) module was developed for the neck network. This module included a residual aggregation path to collect multi-layer feature information and a gate control unit that dynamically weighted feature components based on their relevance. The residual structure provided a direct gradient propagation path, alleviating gradient vanishing during backpropagation and ensuring efficient information transfer during feature fusion. A systematic experimental framework was established to comprehensively evaluate model performance: (1) Ablation studies were conducted using a controlled variable approach to verify the individual contributions of the improved MobileNetV4, D-F-Ramit, and RAGConv modules; (2) Lesion scale sensitivity analysis assessed detection performance across different lesion sizes, with emphasis on small-spot recognition; (3) Resolution impact analysis evaluated five common input resolutions (320×320 – 736×736) to explore the trade-offs among accuracy, speed, and computational efficiency; and (4) Embedded deployment validation involved model quantization and implementation on the Rockchip RK3588 platform to measure inference speed and power consumption on edge devices. **[Results and Discussions]** The proposed YOLO-MDFR achieved an mAP@0.5 of 95.6% on this dataset. Compared to YOLOv12s, it improved accuracy by 2.0%, reduced parameters by 61.5%, and lowered computational complexity by 68.5%. Real-time testing showed 43.4 f/s on an NVIDIA RTX 4060 GPU (CUDA 12.2) and 22.8 f/s on a Rockchip RK3588 embedded platform with only 3.5 W power consumption—suitable for battery-powered field devices. Lesion-scale analysis revealed 33.5% accuracy for <16×16 pixel lesions critical for early detection. Confusion matrix evaluation reduced misclassification, bacterial leaf spot/thrips damage misrates fell from 5.8% to 2.1%, and tobacco mosaic virus/healthy leaves from 3.2% to 1.5%, resulting in an overall 2.3% misrate. Experiments across varying input resolutions revealed a clear performance – resolution trade-off. As resolution increased from 320×320 to 736×736, mAP rose from 89.5% to 96.2%, showing diminishing returns beyond 512×512. Concurrently, computational cost grew roughly quadratically, reducing inference speed from 65.2 f/s to 35.1 f/s. **[Conclusions]** This study presents YOLO-MDFR, a lightweight detection model for identifying pepper leaf diseases and pests under complex natural conditions. By integrating an improved MobileNetV4 backbone, a multi-dimensional frequency reciprocal attention mixing transformer (D-F-Ramit), and a residual aggregation gate-controlled convolution (RAGConv) module, YOLO-MDFR outperforms mainstream detection models in both accuracy and efficiency. Systematic deployment experiments yielded optimized configurations for different application scenarios. Despite its strong performance, the model shows limitations in robustness under extreme lighting, generalization to emerging diseases, and detection of small targets under occlusion. Future work will address these issues through ambient light data fusion, domain adaptation with semi-supervised learning, and binocular vision integration.

Key words: YOLO; leaf disease and pest detection; MobileNetV4; lightweight deep learning model; attention mechanism

Foundation items: National Natural Science Foundation of China Project (51567014); Gansu Provincial Science and Technology Program (22JR5RA797)

***Corresponding author:** YAO Xiaotong, E-mail: 1803552470@qq.com