

植物病虫害分类轻量化与可解释性研究

张梅¹ 赖祥太¹

ZHANG Mei LAI Xiangtai

摘要

针对传统模型参数量大、田间实时检测难的问题,文章提出了一种嵌入坐标注意力(coordinate attention, CA)模块的轻量级 MobileNetV2 二分类模型,用于植物健康与病害叶片识别。模型通过水平/垂直池化编码空间坐标,构建双分支加权机制以精准捕捉病斑特征。同时,结合 Grad-CAM++ (gradient-weighted class activation mapping++) 可视化分析增强模型可解释性。在 New PlantVillage 数据集上,模型以 2.67×10^6 参数量、320 MFLOPs,实现 99.7% 的准确率,单图推理时间 27 ms。可视化结果显示其能聚焦病斑核心区域,相比 SE、CBAM 机制,在参数量、准确率和推理时间方面取得更优平衡,为田间实时病害检测提供了轻量化且可解释的解决方案,具备边缘设备快速部署潜力。

关键词

植物病虫害; MobileNetV2; 坐标注意力; 轻量化; Grad-CAM++; 可解释性

doi: 10.3969/j.issn.1672-9528.2025.11.008

0 引言

植物病虫害是影响农作物产量和质量的重要原因。全球每年约 40% 的农作物潜在产量因病虫害而损失^[1]。高效、准确地识别和诊断植物病害,对保障农业稳定发展具有重要意义。近年来,深度学习技术推动了基于图像识别的病虫害检测方法的显著突破。Mohanty 等人^[2]在 PlantVillage 数据集上利用深度卷积神经网络实现了 38 类植物病害的高精度识别,准确率高达 99.35%; Sladojevic 等人^[3]基于深度神经网络构建的模型,在 13 类植物病害识别上达到了 96.3% 的平均精度,充分验证了深度学习在植物病害图像识别任务中的高效性与广阔前景。

然而,现有主流深度网络如 ResNet、VGG 等模型结构复杂,例如,ResNet-50 参数量达 25.6×10^6 ,在边缘设备推理延迟超 200 ms,难以满足田间实时检测需求,因此轻量级网络成为必然选择^[4]。MobileNetV2 通过引入倒置残差结构与线性瓶颈设计,极大减少了模型参数与计算量,在保证精度的同时大幅提升了部署效率。多项研究也证实了 MobileNet 及其变体在病害识别任务中的优异表现^[5-6]。

同时,注意力机制作为提升网络特征表达能力的重要工

具,逐渐被广泛引入病虫害识别模型中。Hu 等人^[7]提出的 SE 模块通过建模通道间依赖实现特征自适应加权,Woo 等人^[8]进一步设计 CBAM 模块融合空间与通道注意力映射,均在医学和农业图像识别中取得良好效果。然而,这些模块对空间位置信息建模仍较弱。为此,Hou 等人^[9]提出坐标注意力(coordinate attention, CA)机制,在保持通道建模能力的同时编码特征图在水平与垂直方向上的位置信息。已有研究如 Wang 等人^[10]、Zhang 等人^[11]在多种植物病害图像分类中验证了该机制的有效性,特别适用于部署在低功耗设备上的轻量模型。

此外,深度模型的“黑箱”特性使其在实际农业应用中面临信任度问题。梯度加权类别激活映射(gradient-weighted class activation mapping, Grad-CAM)技术^[12]可用于通过可视化热力图展示模型决策依据。研究表明,Grad-CAM 可有效揭示模型关注的关键区域,提高用户对自动识别系统的接受程度^[13-14]。

综上,本文基于 MobileNetV2 构建轻量级植物病虫害识别模型,集成坐标注意力机制与梯度加权类激活映射增强版(gradient-weighted class activation mapping++, Grad-CAM++) 解释模块,实现高效、准确、可解释的图像识别性能,并在 New Plant Village (<https://www.kaggle.com/datasets/vip000001/new-plant-diseases-dataset>) 数据集上进行验证,为边缘部署提供实践依据。

1 相关工作

植物病害检测研究:深度学习在植物病虫害识别领域得

1. 贵州财经大学信息学院 贵州贵阳 550025

[基金项目] 2023 年度贵州省基础研究计划(自然科学类)项目“基于倾斜摄影测量的天然喀斯特景点 3D 建模研究”(黔科合基础-ZK[2023]一般 031); 2024 年贵州财经大学创新探索及学术新苗项目“基于多尺度感知和空间校正的实例分割研究”(2024XSXB21)

到广泛应用。如 Mohanty 等人^[2]基于深度网络在公开数据集上实现了高精度分类, Sladojevic 等人^[3]也采用 CNN 对不同植物病害进行识别, 均取得了 96% 以上的准确率。尽管现有方法精度高, 但多数模型参数庞大, 不适合移动设备实时应用。为此, 研究者们探索轻量级网络结构以降低计算成本。

轻量级网络结构: MobileNet 家族是移动视觉任务的代表。MobileNetV2 通过倒置残差块和线性瓶颈结构有效压缩模型复杂度^[4]。本研究选用 MobileNetV2 作为基线网络, 借助其高效特征提取能力实现模型轻量化。

注意力机制: Hu 等人^[7]提出的 SE 模块通过显式建模通道间依赖重标定特征响应, 取得了显著效果。Woo 等人^[8]的 CBAM 进一步将空间注意力引入, 依次在通道和空间维度生成注意力映射, 对特征进行自适应加权。然而, 这些方法仍然对空间位置信息关注不足。Hou 等人^[9]近期提出的坐标注意力模块在通道注意力中同时编码水平和垂直方向信息, 保留了精确位置关系, 在 MobileNetV2 等网络中提高了定位表现。本研究将坐标注意力模块嵌入 MobileNetV2 结构, 以增强网络对目标病斑的关注能力。

可解释性方法: 模型可解释性可增加用户信任度。在医学影像、疾病检测等领域, Grad-CAM 被用来可视化模型决策依据^[12]。本文使用改进版的 Grad-CAM++ 对植物叶片进行分析, 揭示模型对病斑的识别过程。

2 方法

本文设计的二分类模型基于 MobileNetV2 骨干, 并集

成坐标注意力模块^[4]。模型主要结构示意图如图 1 所示, MobileNetV2 基础网络包含倒置残差块, 倒置残差块 (inverted residual unit) 先通过 1×1 卷积扩张通道, 接着应用 3×3 深度可分离卷积, 最终利用 1×1 卷积将特征通道线性压缩至瓶颈通道维度。

坐标注意力模块嵌入倒置残差块以增强特征表示。Hou 等人^[7]提出的坐标注意力包含以下步骤:

- (1) 对输入特征图沿 X 、 Y 轴分别做全局平均池化, 获取方向特征;
- (2) 使用 1×1 卷积压缩通道并融合这两个特征;
- (3) 将融合后的信息解耦, 分别映射生成水平和垂直方向的空间注意力图。

最终通过坐标注意力图与原始特征相乘增强位置信息。CA 模块使得网络在捕获通道间依赖的同时拥有方向感知能力, 从而更准确地定位叶片上的病灶区域。模型的最后一层分类器修改为包含两个输出单元的全连接层, 并通过 Softmax 激活输出健康 / 病害概率。

2.1 坐标注意力模块 (CA) 原理

传统通道注意力 (如 SE 模块) 仅建模通道依赖, 空间注意力 (如 CBAM) 缺乏方向感知, 而 CA 模块通过水平 / 垂直分离池化实现空间坐标编码, 弥补了上述不足^[9]。给定输入特征图 $X \in \mathbb{R}^{C \times H \times W}$, 其中 C 表示输入特征的通道数; H 表示特征图的高度; W 表示特征图的宽度。

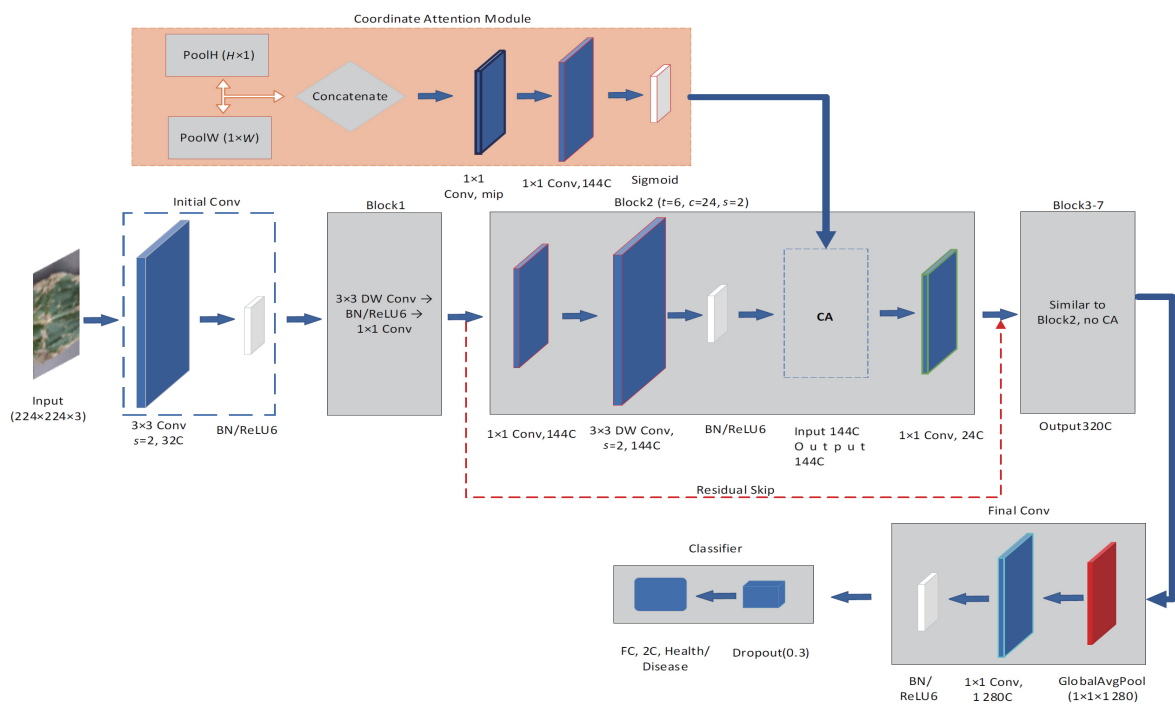


图 1 模型主要结构示意图

CA 模块的计算流程如下:

(1) 坐标信息编码

$$\text{水平方向编码: } \mathbf{X}_h = \text{Pool}_h(\mathbf{X}) \in \mathbb{R}^{C \times H \times 1} \quad (1)$$

$$\text{垂直方向编码: } \mathbf{X}_w = \text{Pool}_w(\mathbf{X}) \in \mathbb{R}^{C \times 1 \times W} \quad (2)$$

式中: Pool_h 和 Pool_w 分别为水平和垂直方向的全局平均池化, 提取水平 / 垂直方向的全局上下文信息。

(2) 特征融合与变换

$$\text{拼接操作: } \mathbf{Y} = \text{Concat}(\mathbf{X}_h, \mathbf{X}_w) \in \mathbb{R}^{C \times (H+W) \times 1} \quad (3)$$

将水平和垂直方向的特征描述在空间维度 (高度维度) 上拼接, 形成包含坐标信息的联合特征。

$$\text{降维卷积: } \mathbf{Y}' = \text{Conv}(\mathbf{Y}) \in \mathbb{R}^{r \times (H+W) \times 1} \quad (4)$$

$$\text{激活函数: } \mathbf{Y}'' = \text{h_swish}(\mathbf{Y}') \quad (5)$$

式中: $r = \max(8, \frac{C}{g})$ 为降维比例, 其中 g 为分组数 (代码中默认为 32); h_swish 是 MobileNetV2 等轻量化网络中提出的高效激活函数, 目的是在保持性能的同时降低计算成本^[4]。

(3) 方向感知权重生成

$$\text{水平权重: } \mathbf{w}_h = \text{Sigmoid}(\text{Conv2}(\mathbf{Y}_h'')) \in \mathbb{R}^{C \times H \times 1} \quad (6)$$

$$\text{垂直权重: } \mathbf{w}_w = \text{Sigmoid}(\text{Conv3}(\mathbf{Y}_w'')) \in \mathbb{R}^{C \times 1 \times W} \quad (7)$$

式中: $\mathbf{Y}_h''$ 和 $\mathbf{Y}_w''$ 是 $\mathbf{Y}''$ 沿空间维度分割后的结果。

(4) 特征重标定

$$\text{输出特征: } \mathbf{Z} = \mathbf{X} \times (\mathbf{w}_h \otimes \mathbf{w}_w) \in \mathbb{R}^{C \times H \times W} \quad (8)$$

式中: $\otimes$ 表示广播乘法, 将水平权重 $\mathbf{w}_h$ 和垂直权重 $\mathbf{w}_w$ 分别应用到输入特征上。最终输出 $\mathbf{Z}$ 为增强空间坐标信息后的特征图。

2.2 嵌入位置与残差协同

(1) 在倒置残差块中的位置

CA 模块嵌入于倒置残差块的 3×3 深度卷积与 1×1 投影卷积之间, 即: 1×1 扩张卷积 $\rightarrow 3 \times 3$ 深度卷积 + BN/ReLU6 $\rightarrow$ CA 模块 $\rightarrow 1 \times 1$ 投影卷积。

(2) 与残差连接的协同作用

当输入与输出通道数和空间尺寸匹配时, 残差连接启用, 形成 $\mathbf{Y} = \mathbf{X} + \text{CA}(\mathbf{F}(\mathbf{X}))$, 其中 $\mathbf{F}(\mathbf{X})$ 表示倒置残差块的主体操作, 包括扩张卷积、深度卷积和 CA 模块。残差连接通过叠加原始输入 $\mathbf{X}$, 缓解梯度消失问题, 同时保留底层特征与高层坐标感知特征的融合。

3 实验

3.1 数据集与预处理

实验采用公开植物病理数据集 New PlantVillage。该数据集在原始植物病害图像基础上通过数据增强技术扩充而成, 包含总计约 8.7 万张 RGB 图像, 覆盖 38 个细粒度类别 (12 类健康叶片和 26 类病害叶片)。所有图像预处理分辨率为 $224 \text{ px} \times 224 \text{ px}$, 通过类别归并构建健康 / 病害

二分类数据集。

训练 / 验证集采用 SubsetRandomSampler (Pytorch 子集随机采样器) 划分, 测试集通过手动均衡采样生成, 结合 WeightedRandomSampler (PyTorch 加权随机采样器) 实现训练阶段的类别平衡^[15], 通过设定正负样本比例确保每个训练批次中健康与病害样本数量均衡, 该方法在不平衡数据场景中的有效性已被广泛验证^[16]。

测试集构建流程如下:

(1) 对原始数据集的训练集、验证集, 均按类别名称是否含 “healthy” 标记为健康 (标签 0) 和病害 (标签 1), 构建二分类数据集;

(2) 将标记后的原始验证集按 1:1 比例拆分为最终验证集与测试集。最终训练集、验证集、测试集比例为 8:1:1, 数据划分统计结果如表 1 所示。

表 1 New PlantVillage 划分统计 (二分类)

类别	训练集	验证集	测试集
健康叶片	22 294	2 786	2 786
病害叶片	48 001	6 000	6 000

3.2 训练策略

训练以交叉熵损失 (Cross-Entropy Loss) 衡量预测与真实标签间误差, 并通过反向传播优化参数。具体设置如下:

(1) 优化器: Adam (adaptive moment estimation) (初始学习率 0.000 1, 权重衰减 1×10^{-4}), 自动调整学习率以加快收敛^[17];

(2) 学习率调度: 当验证集损失在连续 3 个 epoch 未下降时, 采用 ReduceLROnPlateau (PyTorch 学习率调整策略) 将学习率乘以 0.3;

(3) 早停机制: 当验证集损失在连续 5 个 epoch 无改善时, 触发早停;

(4) 训练轮次: 最多 30 个 epoch;

(5) 数据增强: 随机裁剪至 $224 \text{ px} \times 224 \text{ px}$ 、水平翻转 (概率 0.5) 及亮度 / 对比度抖动 (± 0.2) 以提高泛化能力。

训练过程监控训练 / 验证损失、准确率及学习率变化, 模型在训练至第 22 个 epoch 时触发早停, 此时训练损失降至 0.032, 验证损失稳定在 0.041, 两者差距小于 0.01, 无明显过拟合现象; 最终模型保持 99.7% 的验证准确率, 验证了动态学习率策略的有效性。

3.3 基础检测指标

准确率 (accuracy, Acc): 取值范围为 [0, 1], 值越高表示分类能力越强, 用公式表示为:

$$\text{Acc} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{TN} + \text{FN}} \quad (9)$$

式中: TP 为真正例 (true positive) 数量; TN 为真负例 (true

negative) 数量; FP 为假正例 (false positive) 数量; FN 为假负例 (false negative) 数量。

参数量: 值越小表明模型越轻量、计算成本越低。

百万次浮点运算 (million floating point operations, MFLOPs): 衡量模型计算复杂度的核心指标。

推理时间: 在 NVIDIA RTX 5090/32 GB 上实测的单张图像平均推理时间。

3.4 结果与分析

3.4.1 分类准确率与混淆矩阵分析

为了验证模型的泛化性能, 在测试集上进行了测试并得出混淆矩阵, 如图 2 所示。从混淆矩阵中可以看出, 两个类别的样本绝大多数被正确分类, 误分率仅为约 0.36% (健康→病害) 和 0.27% (病害→健康) (结合图 2 混淆矩阵数据), 表明模型区分健康叶片和病害叶片的能力较强。健康类正确分类率为 99.64%, 病害类正确分类率为 99.73%, 对角线上 (健康→健康、病害→病害) 均值较高, 说明模型二分类性能优异; 非对角线部分较小, 表示少量样本被错误分类。

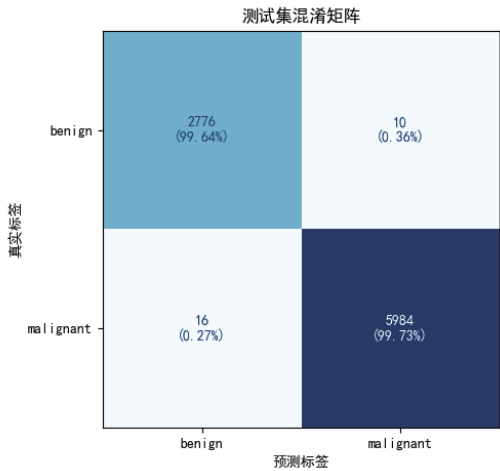


图 2 测试集上的混淆矩阵

3.4.2 注意力机制可视化与特征层级分析

为评估模型可解释性, 本文采用 Grad-CAM++ 对比基线模型与 CA 增强模型的注意力分布, 如表 2 所示。

表 2 基线模型与 CA 增强模型在各特征层的注意力分布对比

特征层级	基线模型表现	CA 增强模型注意力表现
浅层	注意力弥散, 背景干扰明显	初步聚焦病斑, 背景噪声降低
中层	病斑边界模糊, 焦点分散	注意力紧凑, 锁定核心区域
深层	病斑与背景区分度低	高层语义聚焦, 病斑突出性显著

结果表明, CA 模块显著提升各层级特征对病斑区域的聚焦能力: 浅层 (layer 7) 抑制背景干扰, 中层 (layer 10)

聚焦病斑核心, 深层 (layer 13) 增强语义区分度。可视化结果, 图 3 进一步显示, 基线模型热力图存在背景弥散现象, 而 CA 模型通过水平 / 垂直坐标编码, 使暖色区域 (高关注度) 精准覆盖病斑核心, 验证了其空间位置信息的增强作用。图 3 (a) 基线模型热力图显示背景干扰多; 图 3 (b) CA 模型热力图集中于病斑核心, 暖色区域与病害位置高度吻合。

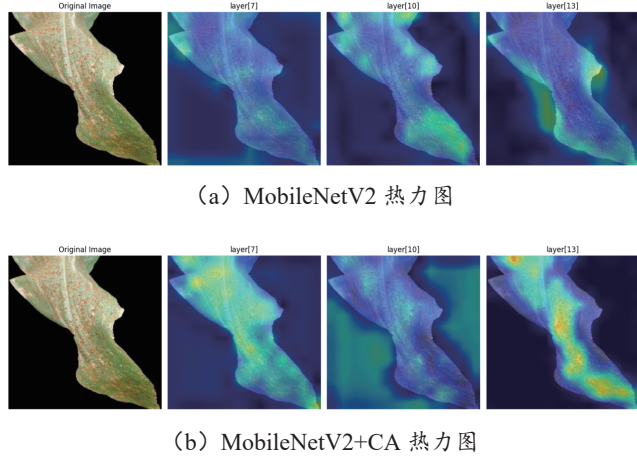


图 3 Grad-CAM++ 可视化对比示例

3.4.3 对比实验

为评估坐标注意力 (CA) 模块的效果, 实验设计了 4 组对比模型: ①基础 MobileNetV2; ② MobileNetV2+SE; ③ MobileNetV2+CBAM; ④ MobileNetV2+CA (本文方法)。在相同训练和测试条件下 (所有推理时间基于单张 224 px × 224 px 图像, 测试环境为 NVIDIA RTX 5090/32 GB, batch size = 32), 比较各模型的测试准确率、参数量、MFLOPs 和推理时间, 结果如表 3 所示。

表 3 各模型的测试集上的准确率、参数量和 MFLOPs 对比

模型	测试准确率 /%	参数量 /10 ⁶	MFLOPs	推理时间 /ms
MobileNetV2	97.5	2.50	300	15~18
MobileNetV2+SE	99.2	2.53	310	16~19
MobileNetV2+CBAM	99.5	2.62	312	20~23
MobileNetV2+CA	99.7	2.67	320	24~27

结果表明, 各注意力模块均能提升基线模型性能。SE 模块以最小开销 (参数量 +0.03×10⁶) 带来显著准确率提升 (+1.7%)。CBAM 进一步将准确率推至 99.5%, 但计算代价有所增加。本文的 CA 模块以适中的参数量与计算开销 (较基线 +6.8% 参数量, +6.7% MFLOPs), 实现了最高的 99.7% 准确率, 在性能与效率间取得了最佳平衡, 验证了其引入的坐标信息在轻量化模型中的独特价值。

4 结论与展望

本文提出了一种基于 MobileNetV2 和坐标注意力 (CA) 的轻量级植物病害二分类模型, 并采用 Grad-CAM++ 进行可视化分析。实验结果表明, 模型仅需 2.67×10^6 参数量、320 MFLOPs, 单图推理时间 27 ms, 分类准确率 99.7%。可视化结果验证了模型对叶片病斑区域的聚焦, 增强了结果可信度。与传统的 SE/CBAM 模块相比, 坐标注意力在提高准确率的同时只带来了极小的计算开销, 体现了轻量化设计的优势。本研究仅验证公开数据集二分类任务, 未涵盖多病害种类及复杂环境。未来将在更大规模多样化的数据集测试模型并探索多分类任务, 部署至边缘设备验证实时性与鲁棒性以提升性能。

参考文献:

- [1] HUGHES D P, SALATHÉ M. An open access repository of images on plant health to enable the development of mobile disease diagnostics[EB/OL]. (2016-04-12)[2025-05-12]. <https://doi.org/10.48550/arXiv.1511.08060>.
- [2] MOHANTY S P, HUGHES D P, SALATHÉ M. Using deep learning for image-based plant disease detection[J]. *Frontiers in plant science*, 2016, 7: 1419.
- [3] SLADOJEVIC S, ARSENOVIC M, ANDERLA A, et al. Deep neural networks based recognition of plant diseases by leaf image classification[J]. *Computational intelligence and neuroscience*, 2016: 3289801.
- [4] SANDLER M, HOWARD A, ZHU M L, et al. MobileNetV2: inverted residuals and linear bottlenecks[C]//IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2018: 4510–4520.
- [5] MUSTAK UN NOBI M, RIFAT M, MRIDHA M F, et al. GLD-Det: guava leaf disease detection in real-time using lightweight deep learning approach based on MobileNet[J]. *Agronomy*, 2023, 13(9): 2240.
- [6] ZHANG L Q, WU X. The lightweight deep learning model in sunflower disease identification: a comparative study[J]. *Applied sciences*, 2025, 15(4): 2104.
- [7] HU J, SHEN L, SUN G, et al. Squeeze-and-excitation networks[C]// IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2018: 7132–7141.
- [8] WOO S, PARK J, LEE J Y, KWEON I S, et al. CBAM: convolutional block attention module[C]//Computer Vision – ECCV 2018. Berlin: Springer, 2018: 3–19.
- [9] HOU Q B, ZHOU D Q, FENG J S. Coordinate attention for efficient mobile network design[C/OL]// 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2021[2025-09-15]. <https://ieeexplore.ieee.org/document/9577301>. DOI: 10.1109/CVPR46437.2021.01350.
- [10] WANG H, QIU S, YE H P, et al. A plant disease classification algorithm based on attention MobileNetV2[J]. *Algorithms*, 2023, 16(9): 442.
- [11] ZHANG M, LIN Z C, TANG S Q, et al. Dual-attention-enhanced MobileViT network: a lightweight model for rice disease identification in field-captured images[J]. *Agriculture*, 2025, 15(6): 571.
- [12] SELVARAJU R R, COGSWELL M, DAS A, et al. Grad-CAM: visual explanations from deep networks via gradient-based localization[C]//2017 IEEE International Conference on Computer Vision (ICCV). Piscataway: IEEE, 2017: 618–626.
- [13] TEWARI A S, KUMARI P. Lightweight modified attention based deep learning model for cassava leaf diseases classification[J]. *Multimedia tools and applications*, 2024, 83: 57983–58007.
- [14] CHEN J D, WANG W H, ZHANG D F, et al. Attention embedded lightweight network for maize disease recognition[J]. *Plant pathology*, 2021, 70(3): 630–642.
- [15] AGHAMOHAMMADESMAEILKETABFOROOSH K, NIKAN S, ANTONINI G, et al. Optimizing strawberry disease and quality detection with vision transformers and attention-based convolutional neural networks[J]. *Foods*, 2024, 13(12): 1869.
- [16] RÖBLE D, PREY L, RAMGRABER L, et al. Efficient noninvasive FHB estimation using RGB images from a novel multiyear, multirater dataset[J]. *Plant phenomics*, 2023, 5: 0068.
- [17] KINGMA D P, BA J. Adam: a method for stochastic optimization[EB/OL]. (2017-01-30)[2025-04-25]. <https://doi.org/10.48550/arXiv.1412.6980>.

【作者简介】

张梅 (1974—), 女, 贵州沿河人, 博士, 教授 (三级)、硕士研究生导师, 研究方向: 计算机软件与理论。

赖祥太 (1996—), 男, 四川达州人, 硕士研究生, 研究方向: 计算机软件与理论。

(收稿日期: 2025-06-09 修回日期: 2025-11-05)