

# Crop Pests and Diseases Detection Based on UAV Remote Sensing and Improved VMamba-UNet

Xijun Wang<sup>1</sup> Xinhua Xu<sup>2</sup> Ping Li<sup>2</sup>

1. iFlytek Co., Ltd. Zhengzhou, Henan, 451150, China

2. College of Engineering, Sias University, Zhengzhou, Zhengzhou, Henan, 451150, China

## Abstract

In response to the challenges of weak small target features, complex backgrounds, low computational efficiency of traditional models, and insufficient global modeling ability in crop pest and disease detection in unmanned aerial vehicle remote sensing images, an improved multi-scale visual state space U-Net model (MVMamba UNet) is proposed. This method introduces a multi-scale state space module (MSSM) to model the long-range spatial dependencies of pests and diseases with linear computational complexity, and integrates a channel space attention mechanism (CSA) to adaptively focus on salient regions and suppress background interference. At the same time, it combines an optimized U-shaped encoding and decoding architecture to achieve efficient fusion of multi-scale features. The experiments on the self built unmanned aerial vehicle remote sensing pest and disease dataset show that the accuracy, recall, and F1 score of the proposed model reach 0.942, 0.937, and 0.939, respectively, and the comprehensive performance and training efficiency are better than the comparison model. The research provides an effective solution for precise and real-time intelligent detection of crop diseases and pests in complex field environments, which has practical application value for promoting the development of smart agriculture.

## Keywords

Crop pest and disease detection; UAV remote sensing; VMamba-UNet; Improve VMamba UNet

# 基于无人机遥感与改进 VMamba-UNet 的作物病虫害检测方法

王喜军<sup>1</sup> 许新华<sup>2</sup> 李萍<sup>2</sup>

1. 科大讯飞股份有限公司, 中国·河南 郑州 451150

2. 郑州西亚斯学院工学部, 中国·河南 郑州 451150

## 摘要

针对无人机遥感图像中作物病虫害检测存在的小目标特征弱、背景复杂、传统模型计算效率低且全局建模能力不足的挑战, 提出了一种改进的多尺度视觉状态空间U-Net模型(MVMamba-UNet)。该方法通过引入多尺度状态空间模块(MSSM)以线性计算复杂度建模病虫害的长程空间依赖, 并集成通道-空间注意力机制(CSA)以自适应聚焦于显著区域、抑制背景干扰, 同时结合优化的U型编解码架构实现多尺度特征的高效融合。在自建无人机遥感病虫害数据集上的实验表明, 所提模型的精确率、召回率和F1分数分别达到0.942、0.937和0.939, 综合性能与训练效率均优于对比模型。研究为田间复杂环境下的作物病虫害精准、实时智能检测提供了有效的解决方案, 对推动智慧农业发展具有实际应用价值。

## 关键词

作物病虫害检测; 无人机遥感; VMamba-UNet; 改进VMamba-UNet

## 1 引言

粮食安全是国家发展的基石, 而病虫害是导致作物减产的主要生物灾害。传统田间调查方法效率低、覆盖面窄, 难以满足大范围农情监测需求<sup>[1]</sup>。无人机(UAV)遥感凭借

其灵活、低成本、高分辨率的优势, 已成为病虫害动态监测的重要工具<sup>[2,3]</sup>。其获取的影像包含丰富的作物冠层信息, 为病虫害的早期精准诊断提供了数据基础<sup>[4]</sup>。对无人机遥感图像进行智能解译的核心在于病虫害区域的自动检测。显著性目标检测(SOD)技术能够快速定位图像中的突出区域, 有助于从复杂农田背景中提取病虫害斑块<sup>[5,6]</sup>。当前方法主要基于深度学习。基于CNN的方法(如YOLO、U-Net系列)局部特征提取能力强, 但感受野有限, 难以建模病虫害的离散分布与长程关联, 易受复杂背景干扰<sup>[7,8]</sup>。基于ViT的方法通过自注意力机制实现全局上下文建模, 但其计算复

【基金项目】河南省科技攻关项目(252102110343); 河南省校企协同创新项目(26AXQXT120)。

【作者简介】王喜军(1982—), 男, 中国河南平顶山人, 工程师, 硕士研究生, 研究方向计算机应用技术。

杂度随图像尺寸呈二次增长,难以满足无人机平台对高分辨率影像的实时处理要求 [9]。新兴的状态空间模型 (SSM),特别是 Mamba 架构,为解决上述矛盾提供了新思路。它能在保持线性计算复杂度的同时有效捕获长程依赖,已在遥感目标检测等任务中展现出潜力 [10]。然而,将其直接应用于农业病虫害显著性检测仍面临独特挑战:病虫害目标 (尤其是早期症状) 常为特征微弱的“小目标”,易丢失细节或与背景混淆;田间光照变化、遮挡等动态环境干扰大;无人机载平台计算资源严格受限,要求模型必须在精度、速度和轻量化间取得最佳平衡 [11]。因此,本文提出一种基于改进 VMamba-UNet(MVMamba-UNet)的无人机遥感图像重大作物病虫害检测方法。旨在融合全局上下文建模与局部细节提取的优势,构建一个面向农田场景、轻量化、高性能的检测网络,以期作为作物病虫害的自动化、精准化监测提供有效的技术解决方案。

## 2 改进 VMamba-UNet(MVMamba-UNet)

本文提出了一种用于无人机遥感图像作物病虫害检测的多尺度视觉状态空间 U-Net 模型 (MVMamba-UNet)。其整体架构如图 1 所示,采用经典的编码器-解码器 U 型结构,并集成了通道-空间注意力 (CSA) 跳跃连接机制。编码器与解码器主要由多尺度状态空间模块 (MSSM)、改进的块合并 (IPM) 与改进的块扩展 (IPE) 操作构成。

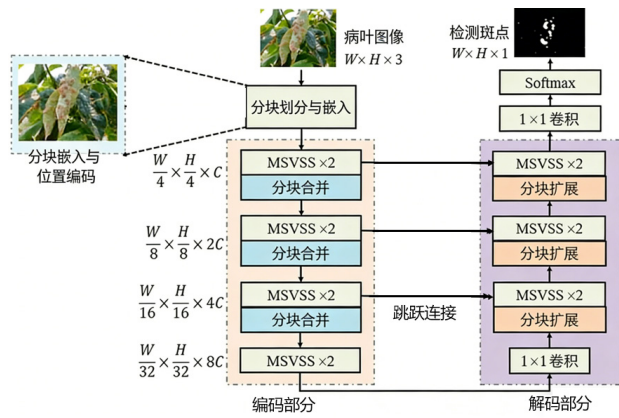


图 1 MVMamba-UNet 的架构图

整体架构: 给定输入图像, 首先将其划分为不重叠的  $4 \times 4$  图像块, 并通过线性投影嵌入到  $C$  维特征空间 (默认  $C=96$ ), 得到块嵌入特征。该特征经层归一化 (LayerNorm) 后输入编码器。编码器包含四个渐进阶段, 依次使用  $[2,2,2,2]$  个 MSSM 模块, 通道数按几何级数从  $C$  增至  $8C$  ( $C \rightarrow 2C \rightarrow 4C \rightarrow 8C$ )。每两个阶段之间通过 IPM 操作进行 2 倍空间下采样并使通道数翻倍。解码器对称地包含  $[2,2,2,1]$  个 MSSM 模块, 通道数从  $8C$  缩减至  $C$  ( $8C \rightarrow 4C \rightarrow 2C \rightarrow C$ )。除最后阶段外, 每个上采样阶段起始于一个 IPE 操作。通过逐元素相加的方式融合编码器与解码器对应层的特征, 以保持参数效率。最终, 一个由 4 倍

块扩展层和  $1 \times 1$  卷积投影层构成的重建模块将特征恢复至原始空间分辨率和通道维度。

借鉴 VMamba 思想, 设计“行优先”与“列优先”的双向扫描策略, 将 2D 图像特征图转换为两条互补的 1D 序列, 确保每个像素都能充分融合全局行、列上下文信息, 避免因序列化造成空间结构信息丢失。在每个 Mamba 块前加入扫描模块, 对输入特征图进行两个方向的展开与处理, 再将结果融合。为有效分离结构性目标信息与纹理噪声, 研究引入频域特征自推理模块。拟通过在频域维度对特征进行分解与推理, 为后续多尺度特征融合提供更加判别性和鲁棒性的语义表示。给定输入特征  $F \in R^{C \times H \times W}$ , 通过离散小波变化 (DCT) 可以获得不同频率分量特征  $F_{LL}$ 、 $F_{HL}$ 、 $F_{LH}$  和  $F_{HH}$ 。

$$F_{LL}, F_{HL}, F_{LH}, F_{HH} = DWT(F) \quad (1)$$

研究通过将分解的频域特征输入卷积层进行进一步提取与编码, 以捕获频域空间中的局部模式与结构细节。

$$F_1 = Conv_{3 \times 3}(F_{LL}), F_2 = Conv_{3 \times 3}(F_{LL}) \quad (2)$$

$$F_3 = Conv_{3 \times 3}(F_{LL}), F_4 = Conv_{3 \times 3}(F_{LL}) \quad (3)$$

为进一步降低模型参数量与计算开销, 在频域特征提取中引入深度可分离卷积进行处理如下:

$$F'_1 = D_e(F_1), F'_2 = D_e(F_2) \quad (4)$$

$$F'_3 = D_e(F_3), F'_4 = D_e(F_4) \quad (5)$$

将各分支处理后的特征经过上采样和沿通道维度进行拼接, 以融合不同频率、不同尺度的特征信息, 形成包含丰富结构细节与全局语义的多尺度频域特征表示。

## 3 实验与结果分析

### 3.1 实验设置

为系统评估 VMamba-UNet 在无人机遥感作物病虫害检测任务中的性能, 本研究基于自建数据集开展了对比与消融实验。实验数据通过搭载高清 RGB 相机的 DJI S1000+ 无人机采集, 飞行高度 25 米, 地面采样距离约 1.0 厘米/像素。原始图像经裁剪、标准化预处理后, 通过添加噪声、旋转及裁剪等数据增强技术, 构建了包含 1000 张  $512 \times 512$  像素图像及其像素级标注的数据集。按 6:4 的比例随机划分为训练集 (600 张) 和测试集 (400 张)。

所有实验在配备 NVIDIA Tesla V100 GPU 的工作站上完成, 使用 Keras (基于 TensorFlow 后端) 框架。为保证公平比较, 所有对比模型均采用 Kaiming 方法初始化参数, 使用随机梯度下降 (SGD) 优化器, 设置初始学习率为 0.001, 动量为 0.4, 批量大小为 15, 共训练 1000 轮。针对病虫害斑块像素占比较小的类别不平衡问题, 训练中采用加权交叉熵损失函数。采用像素精度 (Precision)、召回率 (Recall) 和 F1 值作为主要评价指标, 计算如下:

$$\begin{aligned}
 Precision &= \frac{TP}{TP + FP} \\
 Recall &= \frac{TP}{TP + FN} \\
 F1 &= \frac{Precision \cdot Recall}{Precision + Recall}
 \end{aligned}
 \tag{6}$$

### 3.2 对比实验结果

将改进 VMamba-UNet 与 2 种模型进行对比: U-Net、VM-UNet。所有模型在相同条件下训练至收敛。检测结果如图 2 所示。

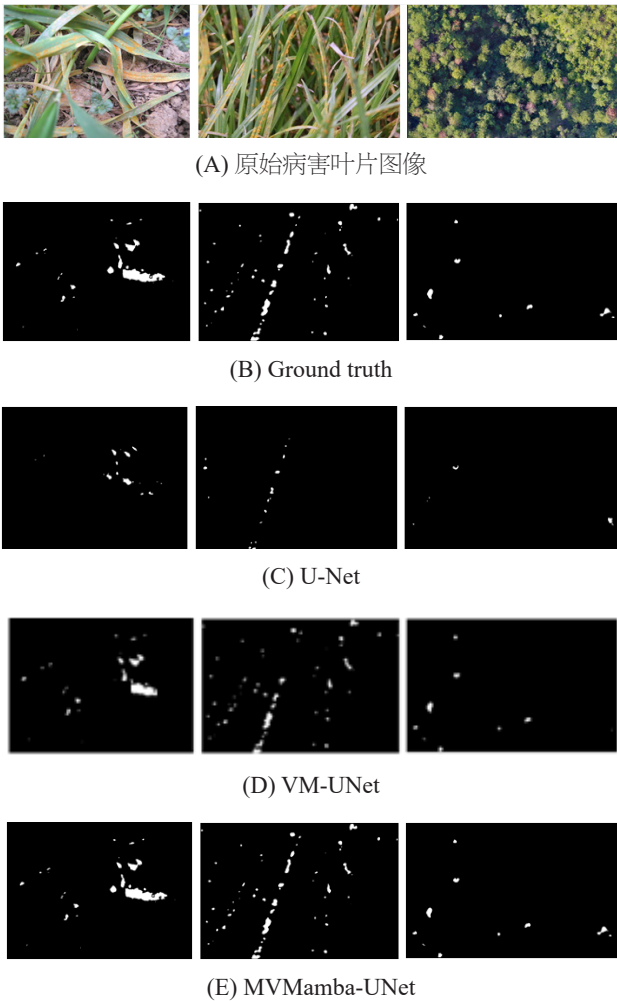


图 2 由三种模型得到的病害检测结果

表 1 不同模型在作物病害检测任务上的性能对比

模型	精确率 (Precision)	召回率 (Recall)	F1 分数 (F1-Score)	训练时间 (分钟)
U-Net	0.721	0.758	0.739	126
VM-UNet	0.913	0.938	0.925	79
MVMamba-UNet	0.942	0.937	0.939	87

图 3 为害虫图像数据库上的检测结果。

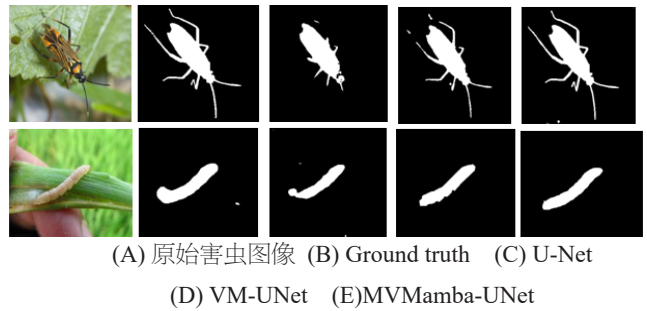


图 3 由三种模型得到的害虫检测结果

### 3.3 结果分析

实验结果表明, 本文提出的 MVMamba-UNet 在无人机遥感作物病虫害检测任务中取得了较好的性能。其在精确率 (0.942) 与 F1 分数 (0.939) 上均优于对比模型, 且在召回率 (0.937) 与训练效率 (87 分钟) 上达到优异平衡。可视化结果进一步证实, 该模型对复杂背景下的微小、离散病虫害目标具有更完整、精准的定位与分割能力, 显著减少了误检与漏检。

### 4 结语

针对无人机遥感图像中作物病虫害检测所面临的小目标特征微弱、背景复杂且计算资源受限的核心挑战, 提出了一种基于改进 VMamba-UNet (MVMamba-UNet) 的检测方法。该方法创新性地融合了多尺度视觉状态空间模块与通道-空间注意力机制, 在保持线性计算复杂度的同时, 实现了对病虫害区域长程语义依赖的高效建模与精准定位。在自建数据集上的实验表明, MVMamba-UNet 在精确率 (0.942)、召回率 (0.937) 和 F1 分数 (0.939) 上均达到最优, 且训练效率高, 显著优于传统 U-Net 与基准 VM-UNet 模型。该研究为复杂田间环境下的病虫害自动化、精准化监测提供了一套切实可行的轻量化解决方案, 对推动智慧农业与精准植保实践具有重要应用价值。未来工作将聚焦于模型的进一步轻量化设计以及在更多元病虫害类型与多模态数据上的泛化能力验证。

### 参考文献

- [1] 方晨晨,石繁槐.基于改进深度残差网络的番茄病害图像识别[J].计算机应用,2020,40(S1):203-208.
- [2] 曾伟辉,李淼,李增,等.基于高阶残差和参数共享反馈卷积神经网络的农作物病害识别[J].电子学报, 2019,47(09):1979-1986.
- [3] 王美华,吴振鑫,周祖光.基于注意力改进CBAM的农作物病虫害细粒度识别研究[J].农业机械学报,2021,52(04):239-247.
- [4] 张善文,邵或,齐国红,等.基于多尺度注意力卷积网络的作物害虫检测[J].江苏农业学报,2021,37(03):579-588.
- [5] He H M, Liu L N, Munir S, et al. Crop diversity and pest management in sustainable agriculture. Journal of Integrative Agriculture, 2018, 18(9): 1945-1952. DOI:10.1016/S2095-3119(19)62689-4.

- [6] Xavier T W F, Souto R N V, Statella T, et al. Identification of Ramularia leaf blight cotton disease infection levels by multispectral, multiscale UAV Imagery. *UAVs*, 2019, 3(2): 33-33. DOI:10.3390/UAVs3020033.
- [7] Zhang P, Huang Y, Pu R. Monitoring plant diseases and pests through remote sensing technology: A review. *Computers and Electronics in Agriculture*, 2019, 165, 104943. DOI:10.1016/j.compag.2019.104943
- [8] Dehkordi R, Jarroudi M, Kouadio L, et al. Monitoring wheat leaf rust and stripe rust in winter wheat using high resolution UAV based red-green-blue imagery. *Remote Sensing*, 2020, 12(22): 3696. DOI:10.3390/rs12223696.
- [9] Narmilan A, Gonzalez F, Salgadoe A, et al. Detection of White Leaf Disease in Sugarcane Using Machine Learning Techniques over UAV Multispectral Images. *UAVs*, 2022, 6,230. DOI:10.3390/UAVs6090230
- [10] Gao D, Sun Q, Hu B, et al. A framework for agricultural pest and disease monitoring based on internet of things and unmanned aerial vehicles. *Sensors*, 2020, 20(5): 1487-1487. DOI:10.3390/s20051487.
- [11] Amarasingham N, Gonzalez F, Salgadoe A, et al. Detection of White Leaf Disease in Sugarcane Crops Using UAV-Derived RGB Imagery with Existing Deep Learning Models. *Remote Sens.*2022,14,6137.