DOI: 10.12398/j.issn.2096-7217.2026.02.014

融合MobileViT-CA机制与多任务学习的农业图像智能识别技术研究

常清雅

(西安航空职业技术学院自动化工程学院,陕西西安,710089)

摘要:针对当前农业图像识别任务中模型分散、多任务协同性弱、实时性与泛化能力不足的问题,研究提出一种基于深度学习与多任务融合的农业图像智能识别系统。该系统以"端—边—云"3级协同为核心架构,通过构建多任务联合学习框架,实现作物株高测量与病害识别的协同训练与推理。融合移动端卷积神经网络(mobile convolutional neural network, MobileNet)、视觉转换器(vision transformer, ViT)特征提取能力与坐标注意力(coordinate attention, CA)机制,在强化关键特征提取、抑制农田背景干扰的同时,大幅降低计算负担。试验结果表明,该系统在株高测量与病害识别两项核心任务中均取得显著成效。在株高测量方面,算法测量结果与人工实测值之间的决定系数R^2最高达99.98%,在病害识别方面,改进后的轻量化模型平均检测精度达到98.76%,模型参数量仅为12.64 M,单张图像平均检测时间最低仅为9.1 ms,具备良好的嵌入式部署能力。此外,系统在复杂环境中仍保持稳定性能,株高测量偏差均在±2 cm范围内,可无缝嵌入无人机等低空监测装备,适配低空农业作业场景的实时数据处理需求。综上所述,研究系统兼具良好的可扩展性、实时性与鲁棒性,有效破解传统技术在多任务协同、轻量化部署与复杂环境适配中的瓶颈,为低空经济赋能下的智慧农业多维度图像识别提供一体化解决方案,助力精准农业规模化落地与农业生产提质增效。

关键词:农业图像识别;多任务学习;轻量化网络;注意力机制;YOLO算法;低空农业监测

中图分类号:TP751;S24 **文献标识码:**A **文章编号:**2096-7217(2026)02-0153-08

常清雅. 融合MobileViT-CA机制与多任务学习的农业图像智能识别技术研究[J]. 智能化农业装备学报(中英文), 2026, 7(2): 153-160

Chang Qingya. Research on agricultural image intelligent recognition technology integrating MobileViT-CA mechanism and multi-task learning [J]. Journal of Intelligent Agricultural Mechanization, 2026, 7(2): 153-160

0 引言

低空经济是我国战略性新兴产业的重要发展方向,以无人机为核心的低空遥感装备已成为智慧农业田间信息采集的核心载体,推动传统人工田间调查向大范围、高效率、非接触式的智能化监测转型[1]。在规模化粮食作物与经济作物种植中,作物株高是表征生长态势、预估产量的核心农艺指标,作物病害是影响产量与品质的关键限制因素[2]。二者的快速、精准、无损监测是实现水肥精准调控、病虫害绿色防控、田间精细化管理的核心前提,更是保障粮食安全、推动农业生产"提质增效、减损增收"的关键技术支撑。农业图像智能识别技术是实现株高与病害自动化监测的核心手段,近年来随着深度学习技术的快速发展,已成为精准农业领域的研究热点[3]。

在作物病害图像识别领域,相关学者围绕深度学习模型优化开展了系统性研究,例如李康顺团队[4]提出基于卷积注意力机制改进的YOLOX-Nano模型,优化损失函数并采用迁移学习,实现轻量化部署与农作物病害高精度检测,适配移动端设备。陆娇娇等[5]通过优化去雾算法与注意力机制改进YOLOv5模型,有效提升复杂光照环境下农作物虫害图像的识别精度与效率。Fu等[6]提出改进视觉转换器(vision transformer, ViT)模型实现作物害虫图像的高精度识别,验证了Transformer架构在农业图像全局特征提取中的有效性。在作物株高图像测量领域,现有研究主要分为立体视觉测量与单目视觉测量两类技术路线。立体视觉方案通过双目相机或多视角图像重建作物三维结构实现株高测量,如姜友谊等[7]基于无人机多视立体影像构建了玉米株高反演模型,实现了大田作物株高的批

收稿日期:2026-02-02 修回日期:2026-04-10

基金项目:西安航空职业技术学院2025年度教育教学改革研究项目(25XHJG025);陕西省教育科学"十四五"规划2025年度一般课题(SGH25Y4086)

作者简介:常清雅,女,1996年生,陕西渭南人,讲师,硕士;研究方向为电子技术、数据处理。E-mail: qingya216280@126.com

量测量。但该类方案对采集设备、飞行航线与数据处理算力要求高,处理周期长,难以满足田间实时监测需求。单目视觉方案凭借设备成本低、处理效率高的优势,成为田间实时监测的主流研究方向,例如南德旺等[8]基于改进Yolov8模型实现了大田烟株株高的快速测量。但现有株高测量研究依赖固定拍摄场景与人工地面基准标记,对逆光、多株重叠等复杂田间环境的鲁棒性不足,并且由于株高测量(空间回归任务)与病害识别(细粒度分类任务)的优化目标、特征偏好存在本质冲突,联合训练易出现精度互损,现有方案难以兼顾双任务性能,未与病害识别任务进行深度融合,部署时需多套算法并行,大幅增加了嵌入式设备的算力负担。

为了解决上述问题,研究以多任务联合学习为核心,融合MobileNet轻量化特性与ViT特征提取能力,嵌入CA机制,针对无人机等低空装备的算力约束与数据传输需求优化部署方案,最终搭建AgriViT-MT端边云协同农业图像智能识别系统。

研究创新点在于深度融合轻量化网络、注意力强化与多任务协同技术,针对农田嵌入式设备约束与多维度监测需求,构建"多源采集、实时处理、协同推理、全局管理"全链路方案。应用层面可提供兼具部署灵活性与识别精准性的技术支撑,为农业管理提供精准决策依据,助力精准农业规模化落地与农业生产提质增效。

1 融合多任务学习与轻量化网络的识别系统设计

1.1 多任务联合学习框架构建

农业图像的识别需要支撑株高测量、病害识别等多类核心监测任务,既要保证各任务的识别精度,又要适配农田嵌入式设备的实时性需求与多场景泛化能力[9-10]。多任务联合学习框架通过共享特征提取与个性化任务分支设计,既能减少冗余计算、降低部署开销,又能挖掘跨任务关联特征强化各任务性能[11]。因此,研究首先构建适配农业场景的多任务联合学习框架,其核心结构如图1所示。

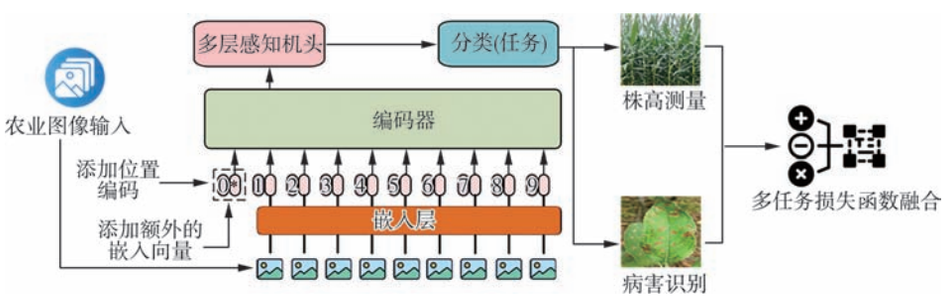


图1 多任务联合学习架构

Figure 1 Multi-task joint learning architecture

多任务联合学习架构以ViT为核心特征提取组件,将原始农业图像进行分块处理,随后传入嵌入层,同时为其添加位置编码与额外的嵌入向量,将分块后的图像转换为结构化的嵌入向量序列。这些嵌入向量序列会被输入至ViT编码器。编码器输出的特征传递至多层感知机头,同步支撑株高测量、病害识别农业监测任务的推理计算。最后,各任务的推理结果会汇入多任务损失函数融合模块,通过动态加权机制平衡不同任务的训练进度,保障各任务精度的协同提升,ViT中初始输入的运算如式(1)所示[12]。

$$z_0=\left(x_{\text{class}},\ x_{\text{p}}^1E,\ x_{\text{p}}^2E,\ \cdots,\ x_{\text{p}}^NE\right)+E_{\text{pos}} \tag{1}$$

式中:x_{class}——多任务全局特征嵌入;

x_{p}^i——第i个农业图像分块的展平向量;

E——农业图像分块的特征投影矩阵;

E_{pos}——农业图像分块的位置编码向量;

N——农业图像的分块总数。

全局窗口多头自注意力的计算复杂度$Complexity_{\text{global}}$的运算流程如式(2)所示[13]。

$$Complexity_{\text{global}}=4ND^2+2N^2D \tag{2}$$

式中:D——分块特征维度。

基于窗口的自注意力复杂度$Complexity_{\text{window}}$的计算如式(3)所示[14]。

$$Complexity_{\text{window}}=4ND^2+2MW^2D \tag{3}$$

式中:M——窗口的数量;

W——单个窗口包含的分块数量。

1.2 轻量化网络与注意力机制优化

研究构建的多任务联合学习框架,依托ViT实现了农业图像多任务特征的协同提取。但农田场景下的嵌入式监测设备算力与功耗资源受限,原ViT的参数量较大、计算开销较高,难以满足实时识别与轻量化部署的需求。将轻量级的卷积神经网络MobileNet与ViT结合,引入CA机制定向强化农业图像关键目标特征、抑制背景噪声,构建出适配农业图像识别的MobileViT-CA算法,其优化算法具体结构如图2所示。

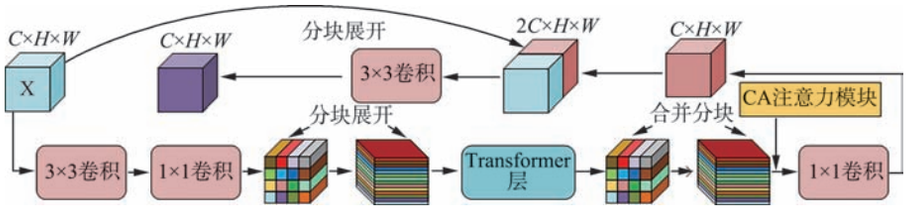


图 2　MobileViT-CA 优化算法结构

Figure 2　Structure of the MobileViT-CA optimization algorithm

MobileViT-CA 优化算法结构先经 3×3 卷积、1×1 卷积完成初步特征提取，再通过分块展开处理为适配 Transformer 层的特征格式。随后由 Transformer 层捕捉农业图像分块间的关联特征，经合并分块恢复空间维度后，插入 CA 注意力模块。接着经 1×1 卷积细化特征，再与初始农业图像经 3×3 卷积得到的残差特征进行拼接，最后通过 3×3 卷积融合输出增强后的农业图像特征。其中，初始卷积特征提取如式(4)所示[15]。

$$\begin{cases} x_1 = \mathrm{Conv}_{3\times3}(x,\ \theta_1) \\ x_2 = \mathrm{Conv}_{1\times1}(x_1,\ \theta_2) \end{cases} \tag{4}$$

式中：x_1——3×3 卷积输出的农业图像特征图；

x_2——1×1 卷积输出的农业图像特征图；

θ_1——3×3 卷积的可学习参数；

θ_2——1×1 卷积的可学习参数；

$\mathrm{Conv}_{3\times3}$——3×3 卷积操作，用于提取农业图像的基础特征；

$\mathrm{Conv}_{1\times1}$——1×1 卷积操作，用于提取农业图像的基础特征。

分块展开与合并分块步骤的运算如式(5)所示[16]。

$$\begin{cases} z = \mathrm{Unfold}(x_2,\ HW) \\ x_3 = \mathrm{Fold}(z'',\ HW) \end{cases} \tag{5}$$

式中：z——分块展开后的农业图像特征序列；

Unfold——分块展开操作，将二维特征图划分为局部分块并展平为序列格式；

z''——Transformer 层输出的农业图像分块特征序列；

Fold——合并分块操作；

x_3——合并分块后输出的农业图像特征图。

1.3　农业图像智能识别系统的实现与部署

为满足农业监测场景“多源采集、实时处理、精准识别、全局管理”的实际化需求，最终以多任务联合学习框架为核心支撑，通过整合适配农田场景的 MobileViT-CA 优化算法及多终端适配组件，构建一种覆盖农业图像“采集—预处理—特征提取—多任务推理—结果输出—全局调度”全链路的智能识别系统，并将其命名为 AgriViT-MT，该系统具体架构如图 3 所示。

AgriViT-MT 农业图像智能识别系统以“端—边—云”3 级协同为核心，首先由终端采集层的多旋翼无人机设备、农田监测站及地面移动终端 3 类装备，完成多高度、多视角的农业图像采集，适配低空作业场景的大范围监测需求。随后各类图像数据传入边缘处理层，先执行数据预处理操作，输出格式统一的农业图像数据。接着预处理后的数据传入 MobileViT-CA 特征提取模块，再输入多任务推理子模块，同步完成株高测量、病害识别的推理运算，最终输出本地检测结果。边缘处理层的检测结果与原始数据会同步存储并上传至云端管理层，采集终端调度中心负责协调各采集设备的任务调度，监测结果管理可视化模块生成株高趋势、病害分布等可视化报表，模型迭代训练模块基于新增数据优化算法参数，农业图像数据库则存储全链路数据以支撑后续分析。

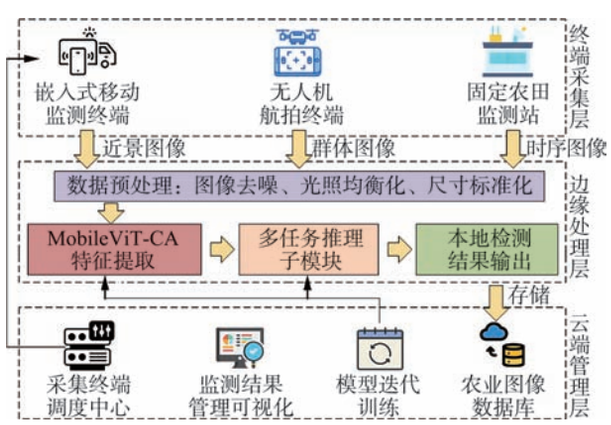


图 3　AgriViT-MT 农业图像智能识别系统架构

Figure 3　AgriViT-MT agricultural image intelligent recognition system architecture

其中多任务联合损失函数用于平衡株高测量、病害识别任务的训练权重，如式(6)所示。

$$L_{\mathrm{total}} = \omega_{\mathrm{h}} \cdot L_{\mathrm{height}} + \omega_{\mathrm{d}} \cdot L_{\mathrm{disease}} \tag{6}$$

式中：L_{total}——总的均方根误差损失；

L_{height}——株高测量任务；

L_{disease}——病害识别任务的均方根误差损失；

ω_{h}、ω_{d}——基于任务验证精度与收敛速度的动态加权系数，二者和为 1，训练后根据验证集精度自动更新，对精度更低、收敛更慢的任务赋予更高权重。

该机制的优势在于:无须人工预设固定权重、无额外计算开销,可自适应适配农业场景数据波动,鲁棒性强。可解决株高检测—病害识别两类任务损失量级不匹配、收敛速度不同步的核心问题,避免训练中单一任务主导优化,有效抑制异质任务联合训练的负迁移现象,实现双任务精度协同提升。

2 试验结果与分析

2.1 试验环境与验证标准

为全面验证AgriViT-MT农业图像智能识别系统的综合性能,研究设计3组对照试验,分别从多任务识别精度、系统实时运行效率、复杂农田环境适应性3个维度展开验证。

试验硬件环境采用嵌入式终端(NVIDIA Jetson Xavier NX,8 GB显存)、PC端(Intel i7-12700H,32 GB内存,RTX 3060显卡),软件环境为Python 3.8,PyTorch 1.12.0,OpenCV 4.5.5。试验数据包含PlantVillage公开农业图像数据集与实地采集自建数据集。株高测量任务全部采用实地采集所得800组图像—株高配对标注数据,无跨数据集分布偏移干扰。病害识别数据集构建过程中,融合PlantVillage公开数据集与1 200组实地采集病害样本;两类数据集交集涵盖玉米、小麦、番茄等6种作物,对应8类相同病害类别,完成数据集间标签体系的统一规范。PlantVillage为实验室可控环境下的叶片病害样本,实地数据通过M300 RTK无人机低空航拍(飞行高度5~10 m)及地面采集获取,包含复杂田间环境下的病害样本、逆光/雾天/多株重叠等复杂环境样本600组。

为缓解两类数据的分布差异,预处理阶段统一执行尺寸归一化、光照标准化、背景抑制操作,完成数据分布的初步对齐。数据划分采用随机分层划分策略,按7:3比例分为训练组与验证组,划分过程中保证两类数据集中的作物品种占比、株高区间分布完全一致,避免抽样偏差。训练组用于模型参数迭代优化与多任务损失函数调优,验证组全程独立于训练过程,仅用于评估模型的泛化能力与测量精度,确保试验结果的客观性与可靠性。

对比系统基于近年来广泛应用的YOLOv8n、李康顺团队[4]改进的YOLOX-Nano、陆娇娇等[5]改进的AgriPest-YOLOv5搭建,确保试验对比的客观性与结果的实时有效性。

2.2 多任务识别精度性能验证

基于PlantVillage公开数据集中的数据,研究首先针对株高测量、病害识别两大核心任务,验证AgriViT-MT系统的识别精度,试验结果如图4所示。

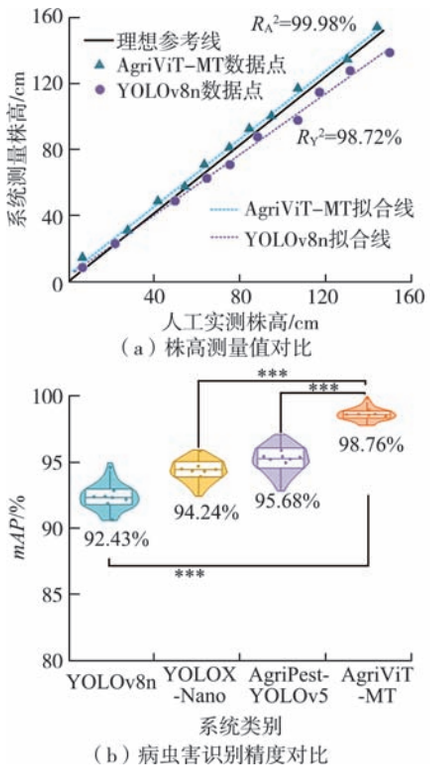


图4 多任务识别精度性能对比验证

Figure 4 Comparison and validation of multi-task recognition accuracy performance

注:***表示$p<0.001$,差异极显著

由图4(a)可见,随人工实测株高从0 cm增长至160 cm,AgriViT-MT与YOLOv8n模型的系统测量株高均呈线性增长趋势,且AgriViT-MT的测量数据点更贴近理想参考线,其拟合决定系数R^2达99.98%,显著高于YOLOv8n模型的98.72%。这表明AgriViT-MT的株高测量值与人工实测值高度吻合,可实现自然光照条件下的无损精准测量。产生这一结果的原因在于,系统融合了ViT特征提取能力,并嵌入CA注意力机制,强化了作物株高的空间特征表征,有效抑制了农田背景干扰。图4(b)中,AgriViT-MT模型的病害识别平均精度(mean average precision, mAP)最高,达到98.76%,显著高于3类对比系统($p<0.001$)。这是由于AgriViT-MT采用多任务联合学习框架,协同提取株高与病害特征,同时通过轻量化网络优化与CA注意力机制强化了病害纹理特征的识别能力,在保证多任务协同推理的同时,提升了病害识别的精准性与稳定性。为进一步量化研究模型的综合性能,明确与当前农业图像识别领域相关研究的性能差异,研究选取同领域、同任务的主流优化模型与权威研究方法,从株高测量、病害识别两大核心任务的关键量化指标进行横向对比,结果如表1所示。

由表1横向对比结果可见,研究提出的AgriViT-MT模型在3项核心分割量化指标上均显著优于当前

相关研究方法，性能差异的核心原因可归纳为两点：

1）对比方法多为病害识别单任务模型，仅能学习病害局部纹理特征，对复杂田间背景下的小病灶、模糊边界的特征捕捉能力不足。而研究的多任务联合学习框架，通过株高测量任务的全局空间特征学习，辅助强化了病害区域的空间边界表征，有效提升了病灶区域的分割完整性。

2）对比的轻量化模型多采用基础通道注意力，对病害边缘的位置信息感知不足，易出现边界偏移、漏分割。而研究引入的 CA 机制，可精准锁定病害区域的空间位置与边缘细节，有效抑制田间背景噪声干扰，大幅提升了病灶边界的识别精准度。

表 1　相关研究的性能量化对比

Table 1　Quantitative comparison of sexual energy in related studies

对比方法	任务类型	病害识别平均交并比 *mIoU*/%	病害识别 Dice 系数/%	病害识别边界 *F*1 分数/%
AgriViT-MT	株高、病害多任务协同	92.35	93.18	94.27
YOLOX-Nano[4]	病害识别单任务	84.62	85.19	86.34
改进 AgriPest-YOLOv5[5]	病害识别单任务	87.34	88.07	89.52
改进 YOLOv8[8]	作物生长多任务监测	83.26	84.15	85.09

2.3　系统实时性能评估

考虑到农田嵌入式监测设备的算力与功耗约束，系统轻量化程度与推理速度是决定其能否落地部署的关键指标。因此，须进一步从参数量、单图检测时间两个核心维度，评估系统的实时性能与嵌入式适配能力，结果如图 5 所示。

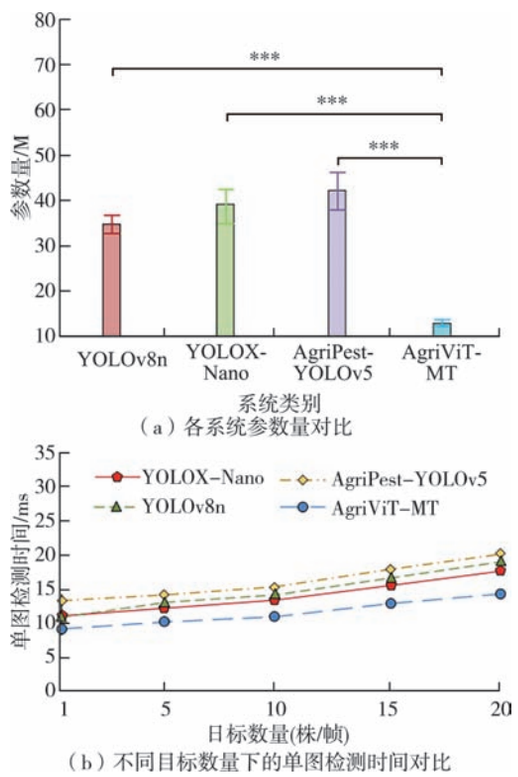


图 5　系统实时性能对比验证

Figure 5　Comparison and validation of system real-time performance

AgriPest-YOLOv5 参数量最大，为 41.2 M，而 AgriViT-MT 参数量最小，仅为 12.64 M，显著低于对比系统（$p<0.001$）。产生这一结果的原因在于，AgriViT-MT 融合了 MobileNet 的轻量化卷积结构与 ViT 的特征共享机制，通过多任务特征复用与通道剪枝优化，大幅压缩了模型参数量。图 5(b)中，随单帧目标数量从 1 株增长至 20 株，各模型的单图检测时间均呈上升趋势，但 AgriViT-MT 的增长最为平缓，检测时间仅从 9.1 ms 增长至 14.5 ms，始终稳定在 15 ms 以内。这是由于 AgriViT-MT 嵌入了坐标注意力机制强化关键特征提取，多任务协同推理框架减少了重复计算，即使在多目标并行检测场景下，仍能保持高效的推理速度，适配农田嵌入式终端的实时处理需求。

2.4　复杂环境下 AgriViT-MT 系统可靠性验证

为验证 AgriViT-MT 系统在真实农田复杂场景下的运行稳定性，研究选取自然逆光、薄雾遮挡、多株重叠、动态视频监测 4 类典型复杂场景，4 类场景下部分样本及识别如图 6 所示。

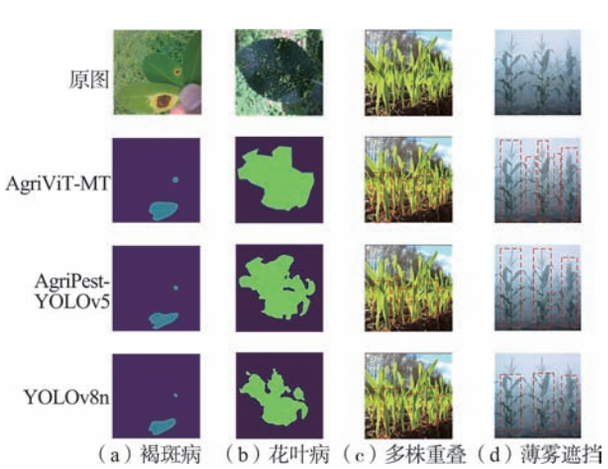


图 6　系统识别可视化对比

Figure 6　Visual comparison of system recognition

如图 6(a)褐斑病场景、图 6(b)花叶病场景所示，3 类模型均完成病变区域识别，但 AgriViT-MT 的分割结

果与病变标注值重合度更高,无漏分割、过分割现象,对比模型的分割结果存在边缘偏移、局部区域漏检。

图 6(c)多株重叠场景、图 6(d)薄雾遮挡场景中,AgriViT-MT能够完成所有作物目标的精准框选,无漏检、重复框选,而对比模型存在目标漏检、检测框重叠偏移。上述结果与量化指标形成互证,体现了研究模型对病害特征的强化表征能力与复杂环境抗干扰能力。研究进一步以株高测量精度、病害识别精度、单帧处理时间为核心指标,与 YOLOv8n、AgriPest-YOLOv5 展开具体的量化对比验证,结果如表 2 所示。

表 2 复杂环境下系统性能对比

Table 2 System performance comparison in complex environments

试验场景	评估指标	AgriViT-MT	YOLO v8n	AgriPest-YOLOv5
自然逆光场景	株高测量决定系数 R^2/%	99.62	98.15	98.03
	株高测量平均偏差/cm	1.23	2.78	1.95
	病害识别平均精度/%	98.15	91.02	95.37
薄雾遮挡场景	病害识别平均精度/%	97.82	94.48	94.71
	单帧处理时间/ms	10.3	17.5	14.8
多株重叠场景	株高测量平均偏差/cm	1.87	2.69	2.44
	多株并行检测帧率/(帧·s^{-1})	66.7	35.8	42.3
动态视频监测场景	株高测量平均偏差/cm	1.95	2.86	2.58

AgriViT-MT在各类复杂农田场景下均展现出良好的稳定性与可靠性。在自然逆光场景中,株高测量决定系数 R^2 达 99.62%,平均偏差仅 1.23 cm,病害识别精度为 98.15%,显著优于 YOLOv8n 与 AgriPest-YOLOv5。薄雾遮挡场景下,病害识别精度仍保持在 97.82%,单帧处理时间仅 10.3 ms,远低于对比模型。多株重叠场景中,株高测量平均偏差 1.87 cm,多株并行检测帧率达 66.7 帧/s,动态视频监测场景下株高测量平均偏差仅 1.95 cm,整体性能优势明显。上述优势源于 AgriViT-MT 系统中 CA 机制与多任务联合学习框架,CA 机制可强化作物关键特征提取,有效抑制逆光、薄雾带来的背景干扰。多任务特征共享与轻量化设计则减少了重复计算,提升了多株并行检测的效率与动态场景下的稳定性。既保障了复杂环境下监测的精准性与实时性,又契合农田实战化监测的可靠性需求。

3 结论

为解决传统农业图像识别技术难以兼顾多任务精度、轻量化部署与复杂环境适应性的问题,研究融合 MobileNet 轻量化卷积结构、ViT 特征提取能力与 CA 机制,构建 AgriViT-MT 端边云协同多任务农业图像智能识别系统。试验验证结果显示,AgriViT-MT 在株高测量任务中拟合决定系数 R^2 达 99.98%,病害识别 mAP 值达 98.76%。模型参数量仅 12.64 M,单图检测时间在 1~20 株多目标场景下仅从 9.1 ms 增长至 14.5 ms,始终稳定在 15 ms 以内。在自然逆光、薄雾遮挡等复杂农田场景中,株高测量平均偏差最低仅为 1.23 cm,病害识别精度大于 97.82%,整体性能优势显著。

尽管研究系统展现出优异的精准性、实时性与可靠性,但尚未验证系统在跨作物品类与极端灾害场景下的适配性。后续研究将纳入多品类作物样本与极端场景数据,优化动态算力调度模块与低空装备数据传输接口,提升系统的场景普适性与协同效率,为低空经济赋能下的精准农业规模化落地提供更坚实的技术支撑。

参 考 文 献

[1] 何勇, 王月影, 何立文, 等. 低空经济政策和技术在农业农村的应用现状与前景[J]. 农业工程学报, 2025, 41(8): 1−16.
He Yong, Wang Yueying, He Liwen, et al. Current status and prospects of low-altitude economy policies and technologies in agriculture and rural areas [J]. Transactions of the Chinese Society of Agricultural Engineering, 2025, 41(8): 1−16.

[2] 赵泽行, 吴晓鹏, 王怡馨, 等. 基于知识图谱的农作物病虫害问答系统研究[J]. 智能化农业装备学报(中英文), 2024, 5(4): 39−50.
Zhao Zexing, Wu Xiaopeng, Wang Yixin, et al. Research on question answering system for crop diseases and pests based on knowledge graph [J]. Journal of Intelligent Agricultural Mechanization, 2024, 5(4): 39−50.

[3] 李家辉, 李晗, 牛永泽, 等. 精准农业技术对农户生产效率和家庭收入的协同影响及机制[J]. 资源科学, 2025, 47(11): 2466−2479.

[4] 李康顺, 杨振盛, 江梓锋, 等. 基于改进 YOLOX-Nano 的农作物叶片病害检测与识别方法[J]. 华南农业大学学报, 2023, 44(4): 593−603.
Li Kangshun, Yang Zhensheng, Jiang Zifeng, et al. Detection and identification of crop leaf diseases based on

improved YOLOX-Nano [J]. Journal of South China Agricultural University, 2023, 44(4): 593－603.

[5] 陆娇娇, 段峰峰. 改进YOLOv5模型下农作物虫害图像识别算法仿真[J]. 计算机仿真, 2023, 40(11): 156－160.

[6] Fu X, Ma Q, Yang F, et al. Crop pest image recognition based on the improved ViT method [J]. Information Processing in Agriculture, 2024, 11(2): 249－259.

[7] 姜友谊, 刘博伟, 张成健, 等. 利用无人机多光谱影像的多品种玉米成熟度监测[J]. 农业工程学报, 2023, 39(20): 84－91.
Jiang Youyi, Liu Bowei, Zhang Chengjian, et al. Monitoring the maturity of multiple varieties of corn using unmanned aerial vehicle multispectral images [J]. Transactions of the Chinese Society of Agricultural Engineering, 2023, 39(20): 84－91.

[8] 南德旺, 李军营, 梁虹, 等. 基于无人机图像和改进Yolov8的烟草植株株高检测[J]. 计算机测量与控制, 2025, 33(7): 19－26.
Nan Dewang, Li Junying, Liang Hong, et al. Tobacco plant stem height detection based on UAV images and improved Yolov8 [J]. Computer Measurement and Control, 2025, 33(7): 19－26.

[9] 王娜, 孟庆辉, 魏宏飞. 基于嵌入式S3C6410的采摘机器人自动导航控制系统[J]. 农机化研究, 2023, 45(10): 182－186.
Wang Na, Meng Qinghui, Wei Hongfei. Design and experimental research of intelligent pipe vegetable planter [J]. Journal of Agricultural Mechanization Research, 2023, 45(10): 182－186.

[10] 卢中华, 袁灿, 马征, 等. 国内外中耕除草装备及其智能化技术现状及趋势[J]. 智能化农业装备学报(中英文), 2026, 7(1): 129－135.
Lu Zhonghua, Yuan Can, Ma Zheng, et al. Current status and trends of intercultural weeding equipment and its intelligent technology at home and abroad [J]. Journal of Intelligent Agricultural Mechanization, 2026, 7 (1): 129－135.

[11] 王新蕾, 王硕, 翟嘉政, 等. 多任务联合学习下的复杂天气航拍图像目标检测算法[J]. 计算机工程与应用, 2025, 61(2):97－111.
Wang Xinlei, Wang Shuo, Zhai Jiazheng, et al. Object detection algorithm of aerial image in complex weather based on multi-task joint learning [J]. Computer Engineering and Applications, 2025, 61(2): 97－111.

[12] 张强, 张灿智, 曹恒, 等. 融合ViT与多尺度注意力的改进YOLOv8飞鸟识别算法[J]. 科学技术与工程, 2025, 25(19): 8151－8157.
Zhang Qiang, Zhang Canzhi, Cao Heng, et al. Improved YOLOv8 bird recognition algorithm integrates ViT and multi-scale attention [J]. Science Technology and Engineering, 2025, 25(19): 8151－8157.

[13] Zhang Q, Xu Y, Zhang J, et al. Vitaev2: Vision transformer advanced by exploring inductive bias for image recognition and beyond [J]. International Journal of Computer Vision, 2023, 131(5): 1141－1162. .

[14] Ammar L B, Gasmi K, Ltaifa I B. ViT-TB: ensemble learning based ViT model for tuberculosis recognition [J]. Cybernetics and Systems, 2024, 55(3): 634－653.

[15] 杨森森, 张昊, 兴陆, 等. 改进MobileViT网络识别轻量化田间杂草[J]. 农业工程学报, 2023, 39(9): 152－160.
Yang Sensen, Zhang Hao, Xing Lu, et al. Light weight recognition of weeds in the field based on improved MobileViT network [J]. Transactions of the Chinese Society of Agricultural Engineering, 2023, 39 (9): 152－160.

[16] 张欢, 周毅, 王克俭, 等. 基于MobileViT-PC-ASPP和迁移学习的果树害虫识别方法[J]. 农业机械学报, 2024, 55(11): 57－67.
Zhang Huan, Zhou Yi, Wang Kejian, et al. Fruit tree pest identification method based on MobileViT-PC-ASPP and transfer learning [J]. Transactions of the Chinese Society for Agricultural Machinery, 2024, 55(11): 57－67.

Research on agricultural image intelligent recognition technology integrating MobileViT-CA mechanism and multi-task learning

Chang Qingya

(*School of Automation Engineering, Xi'an Aeronautical Polytechnic Institute, Xi'an* 710089, *China*)

Abstract: To address the limitations of current agricultural image recognition tasks, such as isolated models, weak multi-task

synergy, insufficient real-time performance and generalization capability, this study develops an intelligent agricultural image recognition system based on deep learning and multi-task fusion. This system adopts a "device-edge-cloud" three-level collaborative architecture. By constructing a multi-task joint learning framework, the system enables concurrent training and collaborative inference for both crop height measurement and disease recognition. The system fuses the feature extraction capabilities of mobile convolutional neural networks (MobileNet) and visual transformers (ViT) with a coordinate attention (CA) mechanism; this design enhances key feature extraction, suppresses complex farmland background interference while significantly alleviating computational burden. Experimental results show that the system achieves remarkable performance across both core tasks. For crop height measurement, the determination coefficient R2 between the algorithm's measurement results and the manual measured values reaches up to 99.98%. For disease recognition, the optimized lightweight model achieves an average detection accuracy of 98.76%, with only 12.64 M parameters. The average single-image inference latency is as low as 9.1 ms, demonstrating excellent potential for embedded deployment. Moreover, the system exhibits high stability in complex environments, maintaining measurement deviation of crop height within ±2 cm. This precision allows for seamless integration into low-altitude monitoring platforms, such as unmanned aerial vehicles (UAVs), to satisfy the real-time data processing requirements of low-altitude agricultural operations. In summary, the proposed system delivers excellent scalability, real-time efficiency, and robustness, effectively overcoming traditional bottlenecks in multi-task synergy, lightweight deployment, and adaptation to complex environments. It provides an integrated solution for multi-dimensional image recognition in smart agriculture empowered by the low-altitude economy, thereby facilitating the large-scale implementation of precision agriculture while improving agricultural production efficiency.

Keywords: agricultural image recognition; multi-task learning; lightweight networks; attention mechanisms; YOLO algorithm; Agricultural monitoring at low altitudes

Foundation items: 2025 Education & Teaching Reform Research Project of Xi'an Aeronautical Polytechnic Institute (25XHJG025); 2025 General Project of the 14th Five-Year Plan for Educational Science of Shaanxi Province (SGH25Y4086)

Author Biography: Chang Qingya, E-mail: qingya216280@126.com