

DOI:10.13718/j.cnki.zwyx.2025.06.008

基于 Cu-ViT 深度学习的烟草气候斑病害分级识别模型的开发应用

金亚波¹, 孙佳照², 罗锦舟²,
吴修森², 丁伟², 罗建钦¹

1. 广西中烟工业有限责任公司, 南宁 400000;
2. 西南大学 植物保护学院, 重庆 400715

摘要: 对烟草气候斑病害的准确分级识别在农业生产、病害防治和环境保护方面具有多维价值。人工识别存在成本高、主观性强、效率低等问题, 可借助图像识别技术解决这一需求。本研究在 Vision Transformer(ViT)框架的基础上, 将压缩单元(Compression Unit, Cu)替代补丁嵌入, 提出 Cu-ViT 模型, 系统地提升了 ViT 模型对图像捕捉的识别能力。Cu-ViT 模型在模拟测试中达到了 91.23% 的准确率, 其 F1 分数、精确率和召回率均超过了 ViT 以及 ResNet152、InceptionResNetV2、Swin Transformer (SwinT) 和 VGGNet19 等先进识别模型, 每张图像的平均识别时间为 104.23 ms。此外, Cu-ViT 模型在实际生产环境下验证的准确率优于人工识别($p < 0.01$)。结果表明, Cu-ViT 模型能够对烟草气候斑病害进行分级识别。

关键词: 烟草气候斑病; 病害分级;
图像识别; 深度学习

中图分类号: S435.72; TP181 文献标识码: A
文章编号: 2097-1354(2025)06-0065-12

开放科学(资源服务)标识码(OSID):



Development and Application of Tobacco Climate Spot Disease Grading Recognition Model Based on Cu-ViT Deep Learning

JIN Yabo¹, SUN Jiazhao², LUO Jinzhou²,
WU Xiumiao², DING Wei², LUO Jianqin¹

收稿日期: 2025-10-20

基金项目: 广西中烟工业有限责任公司项目“基于烟叶质量保障的生产技术数字化模型构建”(0633-224042118J00)。

作者简介: 金亚波, 博士, 农艺师。

通信作者: 罗建钦, 高级农艺师。

1. China Tobacco Guangxi Industrial Co. Ltd., Nanning 530000, China;

2. School of Plant Protection, Southwest University, Chongqing 400715, China

Abstract: Accurate grading and identification of tobacco climate spot disease hold multidimensional value in agricultural production, disease control, and environmental protection. Manual identification suffers from issues such as high costs, strong subjectivity, and low efficiency, which can be addressed through image recognition technology. This study, based on the Vision Transformer (ViT) framework, replaces patch embedding with compression units to propose the Cu-ViT model, systematically enhancing the ViT model's capability in image capture and recognition. In simulated tests, the Cu-ViT model achieved an accuracy of 91.23%, with its F1 score, precision, and recall all surpassing those of ViT as well as advanced recognition models such as ResNet152, InceptionResNetV2, Swin Transformer (SwinT) and VGGNet19. The average recognition time per image was 104.23 milliseconds. Furthermore, the Cu-ViT model's accuracy, validated in real production environments, outperformed manual identification ($p < 0.01$). These results indicate that the Cu-ViT model is capable of grading and identifying tobacco climate spot disease.

Key words: tobacco climate spot disease; disease grading; image recognition; deep learning

目前,全球有100多个国家和地区种植烟草,烟草种植行业已经成为一些国家的经济支柱^[1]。在中式卷烟的生产中,烟草原料的质量与烟叶的工业可用性直接挂钩,会影响后续的卷烟配方加工和卷烟产品的质量^[2]。由于病虫害的侵扰和危害,烟草的产量和质量逐年下降^[3]。在烟草种植过程中,叶部病害历来是导致作物减产和品质下降的主要原因。烟草在田间生长阶段,烟草气候斑病害是主要发生的非侵染性叶部病害之一。气候变化引起的大气臭氧浓度失衡已被广泛证实是该病害的诱发因素^[4]。在生产过程中,病害种类得以明确后,实施与病害特征相适配的补救方案。目前,烟草气候斑病害的诊断主要依靠传统经验,在病害类型和损害严重程度方面容易出现误判,迫切需要用数字化方法代替传统经验判别,以提高诊断的准确性和效率。

近年来,计算机技术与机器人技术的飞速发展,为烟草病害的智能识别提供了可能。人工智能(AI)与先进分类技术,尤其是基于图像的方法的应用,大幅提升了病害检测的效率^[5]。在玉米病害诊断领域,研究人员采用了WG-MARNet模型并结合数据增强技术,平均分类准确率达到97.96%^[6]。同样,在水稻病害识别方面,改进后的VGG网络架构结合ResNet模型成功检测出稻瘟病、纹枯病和白叶枯病,准确率达98.64%,展现出精准的病害分类能力^[7]。VGGNet19网络用于对番茄成熟度等级进行分类,分类效率达92%^[8]。在烟草病害研究方面,已有研究提出一种名为CAMIFFNet的卷积神经网络模型,用于在田间条件下识别烟草花叶病和烟草赤星病。该模型通过多特征融合模块和坐标注意力机制,能有效提取病害特征并降低环境干扰。实验结果表明,在烟草病害图像分类任务中,CAMIFFNet模型的准确率达到89.71%^[9]。现有模型尽管在病害类型分类方面已取得显著成果,但它们均侧重于识别不同种类病害,若能对特定病害的严重程度进行分级识别,将进一步提升在农业生产中的实际应用价值。因此,对烟叶病害感染程度进行精准量化,将为针对性防治措施的实施提供更科学的指导,进而优化病害管理策略。

2020年,谷歌团队提出了ViT网络模型,该模型主要应用于图像分类任务。近年来,ViT模型通过各种视觉任务推进的最新技术,取得了显著的成功^[10]。作为一种图像识别模型,与卷积神经网络(Convolutional Neural Network, CNN)模型相比,ViT模型可以通过自注意力机

制捕获图像中不同区域之间的全局依赖关系,从而弥补了 CNN 在全局信息提取方面的不足,同时,其模块化结构使其易于扩展^[11],因此,ViT 越来越受到研究人员的青睐。例如,He 等^[12]提出了一种用于水稻叶片病害识别的 ECA-ViT 模型,该模型将 ECA 模块集成到 ViT 模型的网络中,以弥补 ViT 模型在提取图像局部特征信息方面的不足。Schmidt^[13]开发了一种使用迁移学习的 ViT 模型,用于分类和识别香蕉叶片中营养缺乏的类型。尽管使用了预训练模型来调整超参数并冻结 ViT 模型的网络层,但该模型在提取局部图像特征方面仍然表现出较弱的能力。上述两项研究都证明了 ViT 模型在作物病害识别中的效用。在本研究中,需要对受烟草气候斑病害不同程度影响的烟叶进行分类和识别,因此,将 ViT 模型作为基础模型采用。尽管 ViT 模型在作物病害图像识别领域表现出色,但它缺乏 CNN 固有的归纳偏置,并且在提取局部特征信息和处理多尺度信息方面存在不足。

为解决上述问题,本文提出一种基于 ViT 模型的改进型烟草病害识别模型,即 Cu-ViT 模型,用于烟草气候斑严重程度的分级识别。在 ViT 模型的基础上,引入压缩单元替代补丁嵌入,通过卷积层提取局部特征,以提升模型的识别性能。实验结果显示,Cu-ViT 模型在模拟测试中的准确率达 91.23%,显著高于人工识别准确率($p < 0.05$),单张图像平均识别时间为 104.23 ms。

1 材料与方法

1.1 数据采集

本研究采用智能手机摄像头进行图像采集,最大限度地提高研究图像源与实际应用场景之间的一致性,确保模型在移动设备应用中的适用性。为获得模型最佳的识别性能,在可见光条件下使用随机抽样法收集图像样本,覆盖了多背景拍摄场景。在拍摄时,拍摄对象统一选取中国主要栽培品种“云烟 87”作为研究对象,不使用微距镜头来放大。总共使用了 6 种不同品牌的智能手机拍摄,包括 iPhone、三星、小米、华为、OPPO、VIVO,分辨率从 0.6 万~2 000 万像素不等。考虑到自然光的昼夜变化,收集数据的时间均在早上 7:00 到晚上 7:00 之间。

气候变化导致田间条件下烟草气候斑病害的发生。烟草气候斑病害初期叶片上会出现小白点,随后逐渐扩散,在此过程中偶见褐色斑块。当病情发展至严重阶段时,主脉及叶尖两侧的侧脉会出现叶片枯萎现象。叶片表面形成密集斑块,中心坏死区域逐渐塌陷,叶片组织呈现黄化特征。根据研究描述,正常烟草与病害烟草叶片被划分为四个等级,具体分级如图 1 所示。



图 1 烟草叶片正常与病害分级图

1.2 ViT 模型框架的改进

本研究提出了针对 ViT 模型架构的双重优化策略。采用压缩单元模块代替传统的补丁嵌入机制，旨在引入 CNN 所具备的内在归纳偏置特性，具体包括局部感知能力与尺度不变性。局部感知能力赋予模型从输入图像邻近像素区域捕获局部特征的功能，尺度不变性则确保多尺度特征能够充分表征不同尺寸目标的特性。相较于 ViT 模型架构将图像分块并线性化为 N 个一维特征向量供后续编码器处理的方式，本研究设计的压缩单元模块能够有效强化烟草气候斑病害病斑相关局部细节特征及多尺度信息的捕获能力，从而减轻模型对大规模训练数据的依赖程度。该优化方案不仅提升了模型对复杂特征的学习性能，还增强了模型的可解释性。为便于直观对比架构优化前后的差异，图 2 与图 3 分别展示了 ViT 模型架构与 Cu-ViT 模型架构的详细结构。

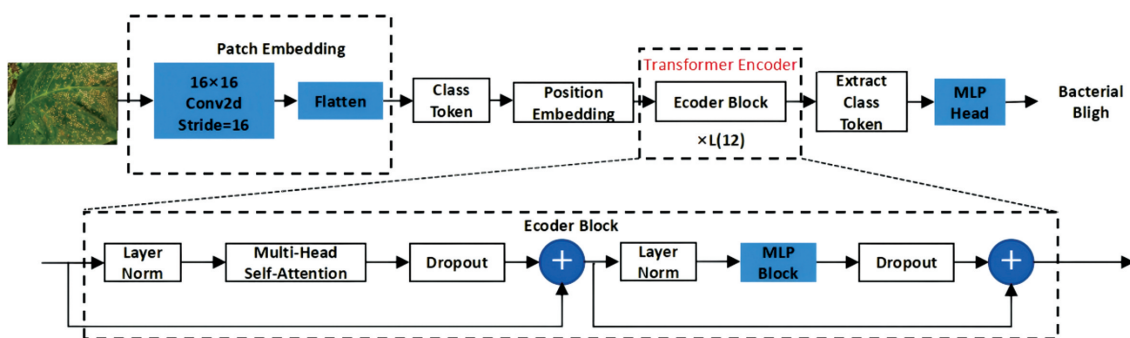


图 2 ViT 模型网络结构

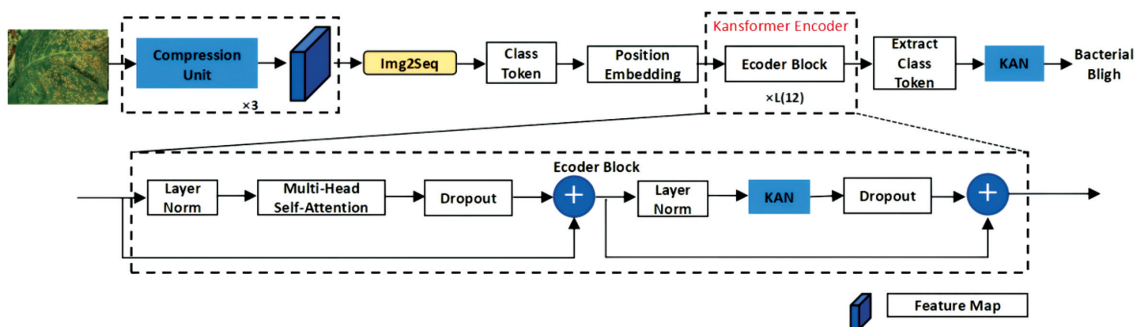


图 3 Cu-ViT 模型网络结构

1.3 压缩单元模块

ViT 模型架构的前端处理采用如下流程：对于尺寸为 224×224 像素的输入图像，系统首先将其均匀切分为 196 个规格为 16×16 像素的子区域，由此构成一个长度为 196 的序列化输入。然后，通过线性映射操作将各个图像子区域转换为对应的一维特征向量，以满足 Transformer 编码器的输入要求。在此基础上，系统引入一个可学习的类别标记，将其与上述特征向量序列进行拼接，构成完整的模型输入。最终，整个向量序列经由位置编码机制获得空间位置信息，并传递至后续的 Transformer 编码器进行深层特征学习。本研究对上述处理范式进行了结构性优化，采用一种新型压缩单元模块^[14]替代了 ViT 模型中的图像分块与线性展平策略。如图 4 所示，该压缩单元采用双通道并行设计，包含卷积通道与注意力通道两个处理路径。其

中,注意力通道专门负责从烟草叶部病害图像中挖掘全局语义特征。为实现多尺度信息的有效捕获,该通道内嵌了金字塔缩减模块,其详细架构见图 5。通过金字塔缩减模块的多层次特征提取机制,使模型获得了在不同空间尺度上识别病害特征的能力。

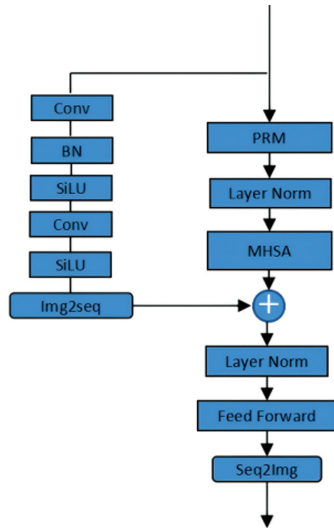


图 4 压缩单元双通道结构

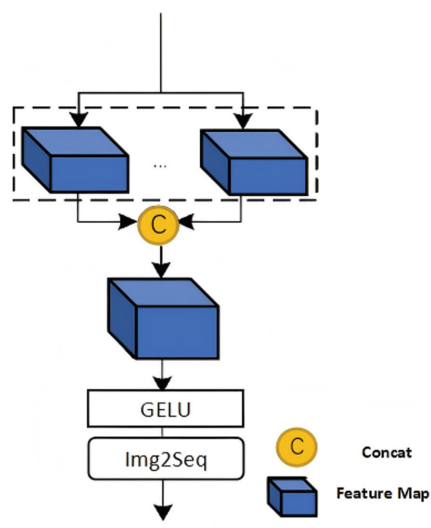


图 5 金字塔缩减模块结构

多层感知机(Multi-Layer Perceptron, MLP), 又称全连接前馈网络, 作为 ViT 模型架构中的核心构件, 主要分布于两个关键节点: 一是嵌入在 Transformer 编码器内部的 MLP 层, 二是承担最终分类功能的 MLP 分类头。从架构层面审视, MLP 本质上是由若干全连接神经元层级联构成的网络结构。在该结构中, 每个神经元单元对前层输出实施线性加权变换, 随后通过预设的激活函数完成非线性映射^[15]。这种分层递进的处理机制赋予了 MLP 卓越的非线性函数拟合能力, 其理论依据源于通用逼近定理。该定理明确阐述了一个重要结论: 当 MLP 网络拥有充分数量的隐藏层神经元时, 其理论上能够以任意精度逼近任何形式的连续函数。

MLP 在计算机视觉分类、语音信号识别以及自然语言理解等方面表现出优异效果。然而, MLP 也存在一定的局限性。例如, 它的可解释性较低, 难以直观地理解模型的决策过程^[16]; 此外, MLP 在网络规模扩展方面存在结构性约束, 面对长序列数据处理任务时, 其在捕捉序列内部长距离依赖模式方面能力有限^[17]。

基于上述分析, 本研究提出一种改进策略: 通过在网络内部引入样条函数机制对学习得到的特征进行精细化调节, 旨在高维特征空间中实现更为精确的特征刻画, 提高模型对烟草气候斑病害病斑的识别能力。

1.4 评估与验证

在模型研究中, 将 Cu-ViT 模型应用于图像识别测试, 使用 7 个指标评估模型性能: 训练集上的损失值, 测试集上的损失值, 训练集上的准确率, 测试集上的准确率、精确率、召回率和 F1 分数, 以证明 Cu-ViT 模型的优越性。一般来说, 较低的损失值和较高的准确率是模型性能优异的指标。每个模型都经过 100 次迭代训练, 并绘制迭代曲线以评估模型的稳定性。同时, 绘制了模型的分类混淆矩阵, 以比较 Cu-ViT 模型与 ViT 模型在分类准确率方面的改进。在改进评估方面, 采用注意力可视化分析模型在病斑的聚焦点。考虑到模型在部署过程中的可移植

性,研究将模型大小和推理时间作为应用程序的评估指标,基于 Python 的 Tkinter 库开发了一个图像识别图形用户界面(GUI)应用程序,以测试模型的识别准确率。该界面可以方便地实现上传、显示和预测烟草气候斑病害图像的功能。通过将模型性能与人工识别的结果进行比较,验证了该模型的实际应用价值。

2 结果与分析

2.1 构建训练平台

用于图像识别训练的计算机采用 Windows 11 X64 操作系统作为实验软件环境,基于开源深度学习框架 PyTorch 构建。主机配备 32 GB 内存和第 13 代英特尔® 酷睿™ i9-13900H 处理器(主频 2.60 GHz),支持 GPU 加速和动态神经网络技术。其中,Nvidia RTX 4060Ti 显卡用于加速深度学习模型运行。模型部署基于开源 mmdetection 框架和 PyTorch 框架,配置包含 Python 3.7 环境、Cuda 11.0 计算架构及 Cudnn 7.6 加速库。最终训练好的模型将保存为完整的 PyTorch 模型文件(.pt),用于后续评估和部署。

2.2 采集数据划分数据集

本研究共采集了 1 860 张烟草叶片图像,包含 419 张健康叶片(无病害)、529 张轻度气候斑叶片、516 张中度气候斑叶片以及 396 张重度气候斑叶片。所有图像均取自真实烟田。在后续模型训练中,80% 的数据集被随机分配为训练集,用于模型训练,剩余 20% 作为验证集使用。

2.3 模型结果分析

2.3.1 确定基础模型

为了确定最佳性能的基础模型,对图像大小、批量大小、初始学习率和优化器进行了测试。针对模型所包含的各类参数,通过逐步调整其数值并验证性能反馈,最终使模型达到最优性能,最佳参数调整设置如表 1 所示。将图像大小调整为 96×96 像素,批量大小设置为 32,初始学习率为 0.01,训练轮数为 100,选择 AdamW 优化器,并采用 StepLR 学习率调度策略实现了模型的最佳性能。

表 1 通用模型参数的最优结果

| | 图像尺寸 | 批量大小 | 初始学习率 | 训练轮数 | 优化器 |
|------|------------------|------|-------|------|--------|
| 通用参数 | 64×64 | 32✓ | 0.01✓ | 50 | AdamW✓ |
| | 96×96 ✓ | 64 | 0.05 | 100✓ | Nadam |
| | 144×144 | 128 | 0.1 | 150 | SGD |
| | 224×224 | 256 | 0.2 | 200 | Adam |

注:✓为最优性能选择。

2.3.2 Cu-ViT 模型的测定

以 Cu-ViT 模型对烟草气候斑病害进行分级识别,表 2 展示了消融实验的结果,该实验旨在评估不同模型组件对整体性能的影响。在缺失数据的情况下,比较了 Cu-ViT 模型和基线模型在 4 个关键指标上的表现:F1 分数、召回率、精确率和准确率。将压缩单元整合到基线模型后,F1 分数提高了 10.69%,召回率提高了 10.15%,精确率提高了 11.27%,准确率提高了 4.42%。

表 2 消融实验结果

| 模型/指标 | F1 分数 | 召回率 | 精确率 | 准确率 |
|-----------|---------|---------|---------|---------|
| 基线模型 | 0.813 4 | 0.796 1 | 0.831 4 | 0.685 2 |
| Cu-ViT 模型 | 0.900 4 | 0.876 9 | 0.925 1 | 0.715 5 |
| 提高率/% | 10.69 | 10.15 | 11.27 | 4.42 |

图 6 展示了 Cu-ViT 模型与基线模型之间的混淆矩阵。对于健康叶片(标签 0)，基线模型做出了 189 次正确预测，错误预测 41 次；而 Cu-ViT 模型做出了 222 次正确预测以及 8 次错误预测，性能得到显著提升。在患病叶片中，所有等级下 Cu-ViT 模型预测正确次数皆超过基线模型：在 1 级为 200 次，基线模型为 195 次；在 2 级为 214 次，基线模型为 213 次；在 3 级为 212 次，基线模型为 208 次。与基线模型相比，Cu-ViT 模型在多个病害等级预测中表现出优越性能。Cu-ViT 模型在所有 4 个指标上均优于基线模型，其特点是更高的准确率和更低的误分类率。尤其在 1 级和 2 级之间，在较轻的病害等级和中度病害等级下依靠经验较难区分，而此时病害的及时识别与防治，对于抑制病害进一步发展和挽回叶片损失至关重要。

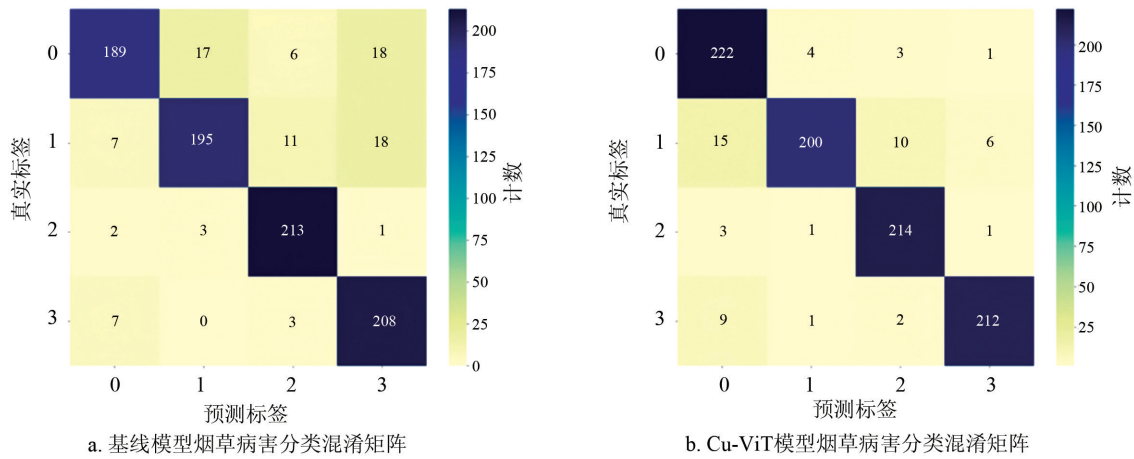


图 6 基线模型与 Cu-ViT 模型混淆矩阵的对比分析

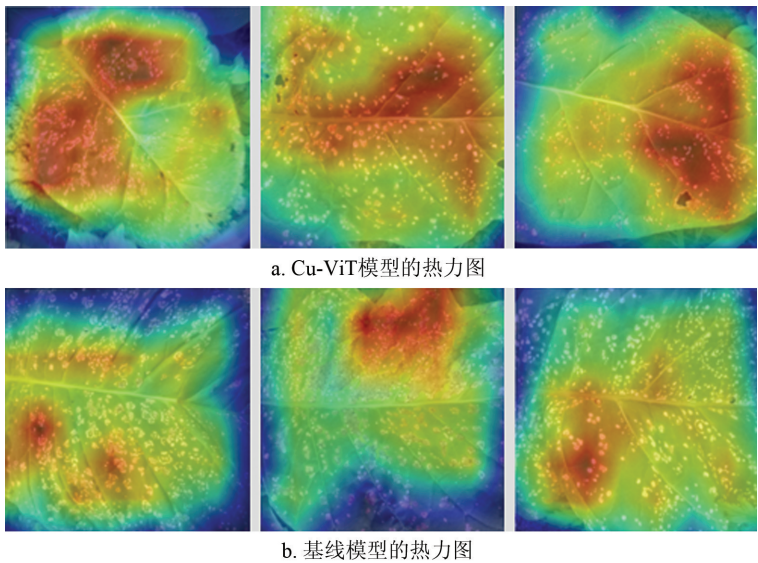


图 7 Cu-ViT 模型与基线模型的聚焦情况对比

为了进一步验证 Cu-ViT 模型的注意力,通过热力图对比 Cu-ViT 模型与基线模型在患病烟叶图像上的聚焦情况。图 7 展示 Cu-ViT 模型在烟草气候斑病害的病斑识别中,相比基线模型更能聚焦于病斑数量多和危害较严重的关键区域,有效减少背景干扰,对病斑部位捕捉性更好。

图 8 显示了在 50 次迭代中,每种算法与 Cu-ViT 模型之间在训练集和验证集上的损失值和准确率的变化。在训练过程中,Cu-ViT 模型的训练损失值迅速下降,并在早期阶段达到较低水平,表明其收敛速度快,学习能力强。Cu-ViT 模型的验证损失值也呈现快速下降的趋势,并在后期保持在较低且稳定的水平,这表明该模型对未见过的数据具有良好的泛化能力。Cu-ViT 模型的训练准确率在训练初期迅速上升,并在后期超过 90%,表明其对训练数据的分类效果非常出色。Cu-ViT 模型的验证准确率在训练过程中持续上升,并在后期保持在较高且稳定的水平,这进一步证明了其对未见过的数据具有出色的分类能力。与几种已成熟的图像识别模型性能进行比较(表 3),尽管 InceptionResNetV2 和 SwinT 在精确率、召回率和 F1 分数上也表现出良好的性能,但与 Cu-ViT 模型相比仍然有差距。

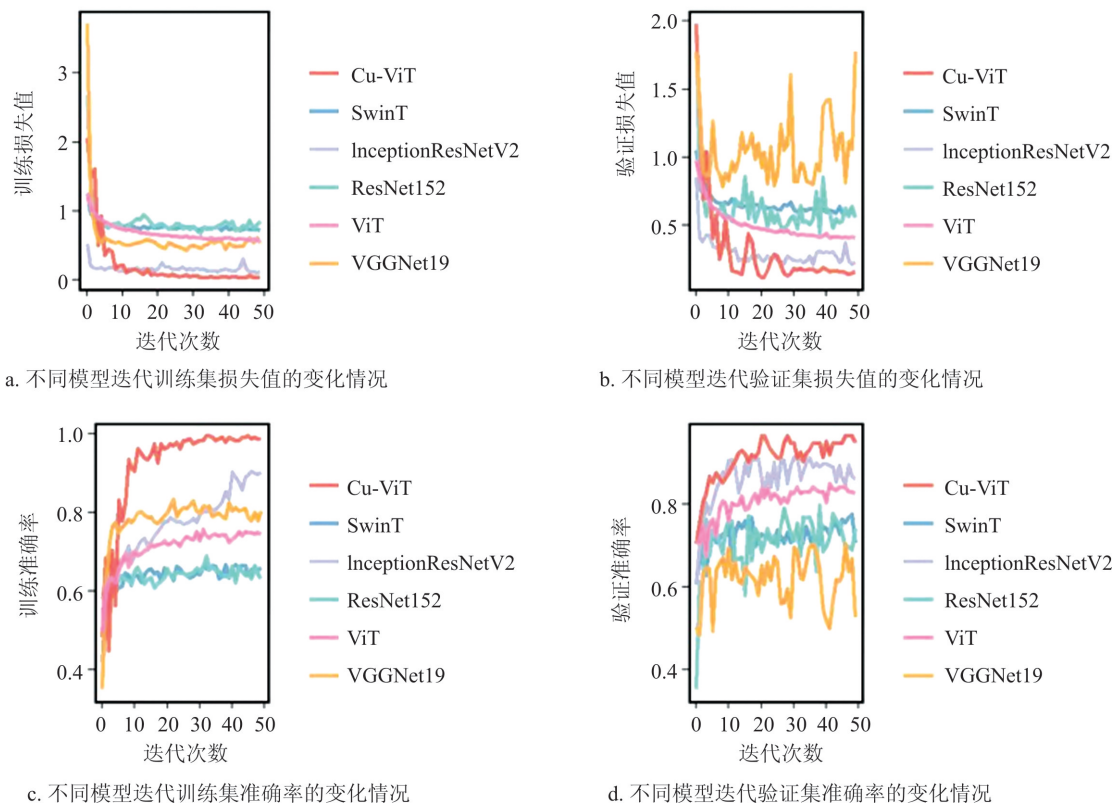


图 8 烟草气候斑病害分层识别中不同深度学习模型的训练和验证性能比较

表 3 不同模型在烟草气候斑病害图像分层识别的平均性能参数

| 模型/参数 | 训练损失值 | 训练准确率 | 验证损失值 | 验证准确率 | 精确率 | 召回率 | F1 分数 |
|-------------------|---------|---------|---------|---------|---------|---------|---------|
| Cu-ViT | 0.221 8 | 0.923 2 | 0.311 8 | 0.935 8 | 0.925 1 | 0.876 9 | 0.900 4 |
| EfficienNetV2 | 0.782 3 | 0.638 2 | 0.653 5 | 0.722 7 | 0.712 9 | 0.701 8 | 0.707 3 |
| InceptionResNetV2 | 0.169 2 | 0.774 9 | 0.295 4 | 0.863 1 | 0.840 6 | 0.837 9 | 0.839 2 |
| ResNet152 | 0.848 8 | 0.633 3 | 0.622 9 | 0.709 4 | 0.685 7 | 0.669 2 | 0.677 3 |
| SwinT | 0.690 5 | 0.711 8 | 0.501 3 | 0.806 7 | 0.789 3 | 0.775 2 | 0.782 2 |
| VGGNet19 | 0.605 7 | 0.781 3 | 1.052 7 | 0.618 6 | 0.653 1 | 0.628 7 | 0.640 7 |

使用 PyTorch 构建的模型进行性能评估后,将 Cu-ViT 模型用于烟草气候斑病害的等级识别,并保存为“best.pt”文件,表 4 展示了 Cu-ViT 模型的训练参数。随后,利用 Python 中的 Tkinter 库创建窗口和控件,其中包括加载背景图像、显示上传图像、执行预测和显示预测结果等功能。图 9 展示了一个基于 GUI 窗口创建的用于识别患有烟草气候斑病害烟叶的小程序,用户可以点击指令选择图像进行识别,识别结果包括烟草气候斑病害的不同严重程度以及所需的识别时间。

表 4 Cu-ViT 模型推断出最优参数

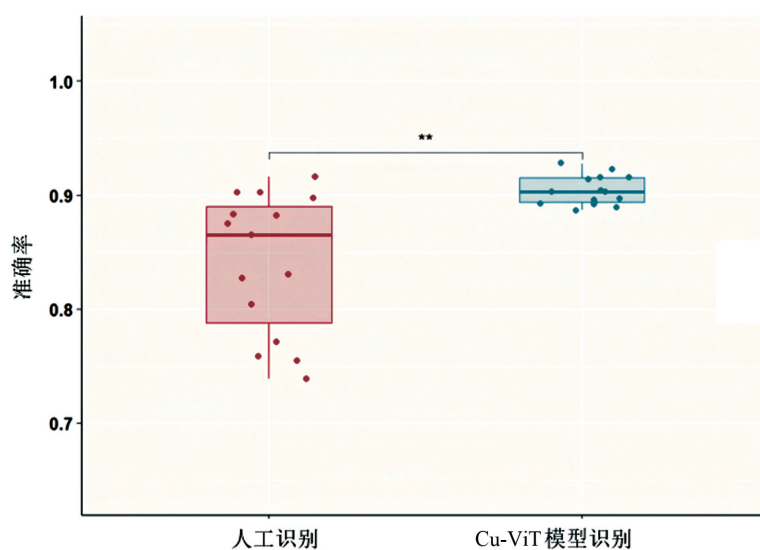
| 类别 | 参数 | 符号/键 | 值/设置 |
|---------|---------------|-------|-------------------------------|
| 模型结构 | 主干网络 | — | Vision-Transformer (ViT-Base) |
| | 卷积层 1 通道数 | C1 | 64 |
| | 卷积层 2 通道数 | C2 | 128 |
| | 卷积层 3 通道数 | C3 | 768 |
| | Transformer 层 | L | 12 |
| 正则化与初始化 | Dropout | pdrop | 0.1 |



图 9 用于识别烟草气候斑病害的模型

对不同模型在部署后识别烟草气候斑病害的性能进行了分析,涵盖准确率、识别时间和模型大小。在病害识别领域,关键指标的优先级排序如下:识别准确率、识别时间和模型大小。因此,本研究按照“准确率→时间消耗→大小”的顺序评估了这些模型。基于对 100 张烟草斑害目标图像的测试结果:Cu-ViT 模型明显优于其他模型,准确率达到 91.23%;EfficientNetV2 模型的时间消耗最短,仅为 68.52 ms;ResNet152 模型的体积最小,为 107.53 MB。在实际应用中,几十毫秒的差异几乎可以忽略不计,158.53 MB 的模型大小可以满足通用平台的部署要求。综合考虑各个模型的整体性能,Cu-ViT 模型被认为是最佳选择。

为了评估所提出的深度学习模型在烟草气候斑病害分级识别任务中的有效性,我们在实际生产中进行测试,将其识别准确率与人工识别的结果进行了比较。图 10 显示了模型识别和人工识别两种识别方法的数值分布,并使用统计检验来验证它们之间的显著差异。此外,表 5 详细列出了模型识别和人工识别在不同病害程度下的具体准确率。从图中可以看出,模型识别的数值范围主要集中在 0.9 左右,而人工识别的值则较为分散,范围为 0.8~0.9。值得注意的是,模型识别和人工识别之间的差异具有极高的统计学意义($p<0.01$),表明模型识别总体上表现出更高的准确性和一致性。在没有患病的烟叶情况下,模型识别的准确率略低于人工识别,但在其他分类中,模型识别的准确率明显高于人工识别。这表明 Cu-ViT 模型在识别中等程度病害方面具有更强的能力,并且模型识别的优势在不同程度的烟叶病害中尤为显著。



注：* 表示 $p < 0.05$ ，** 表示 $p < 0.01$ 。

图 10 Cu-ViT 模型识别与人工识别准确率比较

表 5 对烟草气候斑病害不同级别的识别具体准确率

| 模型/分级 | 级别 0 | 级别 1 | 级别 2 | 级别 3 |
|-------------|---------|---------|---------|---------|
| Cu-ViT 模型识别 | 0.923 1 | 0.875 4 | 0.888 1 | 0.902 5 |
| 人工识别 | 0.967 0 | 0.829 5 | 0.754 5 | 0.817 8 |

3 讨论

不同类型的烟草病害之间存在显著差异，而同一种病害的不同严重程度等级之间的差异相对较小，开发用于烟草气候斑病害的等级识别模型具有挑战性。在本研究中，ViT 模型被选为基线模型，其在农业分类和识别方面的优越性已得到广泛证实^[18-20]。烟草气候斑病害病变颜色多样且大小不均匀，直接使用 ViT 模型识别会导致准确率降低，无法作为有效的识别模型。近年来，许多研究人员对 ViT 模型进行了修改，以适应各种场景需求^[21]。

本研究以 ViT 模型为基础，考虑到不同严重程度的烟草气候斑病害特征主要位于烟叶的局部区域，因此，为了保留图像所携带的详细特征，将 ViT 模型的补丁嵌入(使用步长为 16 的单个 16×16 卷积)替换为压缩单元。该单元采用 3 个小型卷积层级联，然后进行 ReLU 激活，以逐步扩大感受野，从而能够从局部到中局和全局尺度提取多尺度特征。烟草气候斑病害早期(如白点)和晚期(如棕色坏死区域)的形态差异需要根据局部细节进行区分，而压缩单元可以有效地保留这些特征。与单尺度补丁嵌入相比，多尺度特征融合可以提供更丰富的表示，这在区分相似的病害等级(如轻度和中度)时尤为重要。类似的研究方法已应用于其他研究^[22]。在集成压缩单元后，该模型在所有指标上都表现出改进，准确率提高了约 5.59%。因此，Cu-ViT 模型在烟草气候斑病害图像的分级和识别任务中表现出色。与基线模型相比，它不仅提高了分类准确率，而且有效地减少了不同类别之间的混淆。这一结果对于烟草病害自动识别技术的应用具有重要意义，为针对不同病害等级实施的防控策略提供了精确的判别支持。

随着中国人口老龄化的加速和农业人口的急剧下降，劳动力的严重短缺已成为制约农业生产和可持续发展的主要因素。因此，本研究选取烟草气候斑病害不同损害程度的叶片图像，进

行人工识别和 Cu-ViT 模型识别的比较分析。结果表明, Cu-ViT 模型识别的平均准确率接近 0.9, 大多数数据点都聚集在这个水平附近, 表明 Cu-ViT 模型在识别过程中能保持足够的稳定性和可靠性。相比之下, 人工识别的平均准确率略低于模型, 而且数据点的分布更为分散。这些测试结果表明, Cu-ViT 模型已在实际场景中实现了初步的应用性能, 可以在一定程度上替代传统的人工识别。

4 结论

在本研究中, 旨在对烟草气候斑病害的严重程度进行分级和识别。提出的 Cu-ViT 模型, 基于 ViT 模型框架, 通过将压缩单元替代补丁嵌入, 提升了 ViT 模型多尺度和多层次特征提取的能力, 并加强了非线性表达。结果表明, Cu-ViT 模型在测试中达到了 91.23% 的准确率, 其综合性能优于 ResNet152、InceptionResNetV2、SwinT 和 VGGNet19 等先进的图像分类模型。在本研究中, 研究对象采用中国主要栽培品种“云烟 87”, 未来, 可以收集更多不同品种的烟草图像作为训练样本, 以提高模型的泛化能力。

参考文献:

- [1] 孙佳照, 冉渝澳, 冯俊, 等. 西南地区烟草潜在适生区预测[J]. 中国烟草科学, 2023, 44(5): 37-44, 61.
- [2] WANG G L, ZHU Q K, SONG C D, et al. MedKAFormer: When Kolmogorov-Arnold Theorem Meets Vision Transformer for Medical Image Representation[J]. IEEE Journal of Biomedical and Health Informatics, 2025, 29(6): 4303-4313.
- [3] WU Y H, LIU Y, ZHAN X, et al. P2T: Pyramid Pooling Transformer for Scene Understanding[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2023, 45(11): 12760-12771.
- [4] HARIDASAN A, THOMAS J, RAJ E D. Deep Learning System for Paddy Plant Disease Detection and Classification[J]. Environmental Monitoring and Assessment, 2023, 195(1): 120.
- [5] 孙佳照, 李群岭, 林小兴, 等. 基于 Resnet-101 模型的烟蚜数量图像识别系统开发[J]. 植物医学, 2024, 3(4): 26-31.
- [6] LI Z C, ZHOU G X, HU Y W, et al. Maize Leaf Disease Identification Based on WG-MARNet[J]. PLoS One, 2022, 17(4): e0267650.
- [7] LIU H, CUI Y D, WANG J M, et al. Analysis and Research on Rice Disease Identification Method Based on Deep Learning[J]. Sustainability, 2023, 15(12): 9321.
- [8] BEGUM N, HAZARIKA M K. Prediction of Physico-Chemical Properties in Tomatoes Using Deep Neural Architecture[J]. Agricultural Research, 2024: 1-11.
- [9] LIN J W, CHEN Y, PAN R Y, et al. CAMFFNet: a Novel Convolutional Neural Network Model for Tobacco Disease Image Recognition[J]. Computers and Electronics in Agriculture, 2022, 202: 107390.
- [10] WU T N, ZHANG Y W, GONG Z W, et al. Quantification of Tobacco Leaf Appearance Quality Index Based on Computer Vision[J]. IEEE Access, 2022, 10: 120352-120368.
- [11] HAN K, WANG Y H, CHEN H T, et al. A Survey on Vision Transformer[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2023, 45(1): 87-110.
- [12] HE F Y, LIU Y, LIU J F. ECA-ViT: Leveraging ECA and Vision Transformer for Crop Leaves Diseases Identification in Cultivation Environments[C]//2024 4th International Conference on Machine Learning and Intelligent Systems Engineering (MLISE). June 28-30, 2024. Zhuhai, China. IEEE, 2024: 101-104 DOI: 10.1109/mlise62164.2024.10674238.
- [13] SCHMIDT-HIEBER J. The Kolmogorov-Arnold Representation Theorem Revisited[J]. Neural Networks, 2021, 137: 119-126.

- [14] SILVEIRA A C C, DO CARMO D S, UEDA L H, et al. VITMST++: Efficient Hyperspectral Reconstruction through Vision Transformer-Based Spatial Compression[J]. IEEE Open Journal of Signal Processing, 2025, 6: 398-404.
- [15] ZENG Z H, LIU C B, TANG Z, et al. AccTFM: an Effective Intra-Layer Model Parallelization Strategy for Training Large-Scale Transformer-Based Models[J]. IEEE Transactions on Parallel and Distributed Systems, 2022, 33(12): 4326-4338.
- [16] MACDOWALL F D H. Predisposition of Tobacco to Ozone Damage[J]. Canadian Journal of Plant Science, 1965, 45(1): 1-12.
- [17] XUE W X, XU P J, WANG X F, et al. Natural-Enemy-Based Biocontrol of Tobacco Arthropod Pests in China [J]. Agronomy, 2023, 13(8): 1972.
- [18] HAQUE M A, DEB C K, GOLE P, et al. An Enhanced Vision Transformer Network for Efficient and Accurate Crop Disease Detection[J]. Expert Systems with Applications, 2025, 283: 127743.
- [19] MONTAVON G, SAMEK W, MÜLLER K R. Methods for Interpreting and Understanding Deep Neural Networks[J]. Digital Signal Processing, 2018, 73: 1-15.
- [20] SHARMA S K, VISHWAKARMA D K. Classification of Banana Plant Leaves Based on Nutrient Deficiency Using Vision Transformer[C]//2024 5th International Conference for Emerging Technology (INCET), May 24-26, 2024, Belgaum, India. IEEE, 2024: 1-6.
- [21] 冉渝澳, 金亚波, 王振国, 等.烟草靶斑病预测模型构建及数字化应用研发[J].植物医学, 2024, 3(4): 40-49.
- [22] SHINODA R, KATAOKA H, HARA K, et al. Transformer-Based Ripeness Segmentation for Tomatoes[J]. Smart Agricultural Technology, 2023, 4: 100196.

责任编辑 杨光明