

基于全局-局部注意力机制的甘蔗病害分类算法

邓健志¹ 黄福兴² 王泽平³ 罗丽平⁴ 井佩光⁵ 李云^{4,6}

(1. 桂林理工大学物理与电子信息工程学院, 桂林 541004; 2. 桂林理工大学计算机科学与工程学院, 桂林 541004;
3. 广西农业科学院, 南宁 530007; 4. 广西民族大学物理与电子信息学院, 南宁 530006;
5. 天津大学电气自动化与信息工程学院, 天津 300192; 6. 广西智能视觉协作机器人工程研究中心, 南宁 530006)

摘要: 针对复杂自然场景下, 甘蔗病斑受光照不均等干扰, 识别难度大, 检测效率低等问题, 本文提出了一种基于全局-局部注意力机制的甘蔗病虫害分类算法 (Sugarcane disease classification algorithm using global-local attention mechanism, SDCA-GLAM)。为扩充模型容量, 将改进的 Vision Transformer (ViT) 模型线性投影层替换为可变形的卷积模块, 自适应地提取病斑纹理与叶片边缘信息; 引入可重参数化的卷积结构以增强空间位置信息表达能力, 在多层感知机环节融合深度卷积模块, 用于挖掘高维空间特征; 为减少模型参数量并提升检测准确率, 设计全局-局部自注意力并行学习支路, 局部支路采用窗口注意力细化高频纹理特征, 全局支路引入池化策略压缩向量 K/V 的空间维度, 并通过超参数 α 聚合关键区域信息; 将层归一化操作替换为批归一化, 以降低频繁 reshape 带来的内存开销和时间损耗。实验结果表明, SDCA-GLAM 在包含 11 个类别的甘蔗叶片数据集上准确率达到 88.26%, 吞吐量达 1 620 幅/s, 模型参数量为 2.758×10^7 , 显著优于对比的主流模型。本文算法在准确率与效率之间取得了良好平衡, 可为甘蔗病害移动端快速识别提供有效技术支撑。

关键词: 甘蔗; 分类算法; 轻量化模型; ViT; 全局-局部注意力机制

中图分类号: TP391.4 文献标识码: A 文章编号: 1000-1298(2026)06-0300-11

OSID:



Sugarcane Disease Classification Algorithm Based on Global-local Attention Mechanism

DENG Jianzhi¹ HUANG Fuxing² WANG Zeping³ LUO Liping⁴ JING Peiguang⁵ LI Yun^{4,6}

(1. College of Physics and Electronic Information Engineering, Guilin University of Technology, Guilin 541004, China
2. College of Computer Science and Engineering, Guilin University of Technology, Guilin 541004, China
3. Guangxi Academy of Agricultural Sciences, Nanning 530007, China
4. School of Physics and Electronic Information, Guangxi Minzu University, Nanning 530006, China
5. School of Electrical Automation and Information Engineering, Tianjin University, Tianjin 300192, China
6. Guangxi Engineering Research Center for Intelligent Vision and Collaborative Robotics, Nanning 530006, China)

Abstract: Sugarcane is a globally important crop for both sugar production and bioenergy, and it is widely cultivated in tropical and subtropical regions. Effective disease diagnosis is essential to ensuring agricultural productivity and economic returns. In response to challenges posed by complex field environments, such as uneven lighting, low recognition accuracy, and limited detection efficiency, a novel algorithm: sugarcane disease classification algorithm using global-local attention mechanism (SDCA-GLAM) was proposed. To enhance model capacity, the linear projection layers in a modified Vision Transformer (ViT) were replaced with deformable convolution modules, enabling adaptive extraction of lesion textures and leaf-edge information. Re-parameterized convolution was incorporated to strengthen spatial positional encoding, and deep convolutional modules were embedded in the multilayer perceptron to extract high-dimensional semantic features. To improve both accuracy and model efficiency, a parallel global-local self-attention architecture was designed. The local branch leveraged window attention to

收稿日期: 2025-06-08 修回日期: 2025-08-12

基金项目: 广西民族大学引进人才科研启动(2024KJQD218)、广西科技重大专项(桂科 AA24263038)、国家自然科学基金项目(62361002)和广西农业科学院基本科研业务专项(桂农科 2025YP065)

作者简介: 邓健志(1982—),男,教授,主要从事大数据与人工智能研究,E-mail: dengjzh@glut.edu.cn

通信作者: 李云(1978—),女,教授,博士,主要从事大数据与人工智能研究,E-mail: 20240165@gxmzu.edu.cn

refine fine-grained textures, while the global branch reduced the spatial dimensions of key/value vectors via pooling and aggregating critical region information using a hyperparameter α . Finally, LayerNorm was replaced with BatchNorm to reduce the memory and time overhead caused by frequent reshaping. Experimental results on an 11-class sugarcane leaf dataset demonstrated that SDCA - GLAM achieved an accuracy of 88.26%, a throughput of 1 620 images per second, and a model size of 2.76×10^7 . The proposed method outperformed mainstream models in both accuracy and efficiency, making it suitable for real-time mobile deployment in field diagnosis of sugarcane conditions.

Key words: sugarcane; classification algorithm; lightweight model; Vision Transformer; global-local attention mechanism

0 引言

根据国际糖业组织及近年来农业统计数据,甘蔗为重要的热带与亚热带经济作物,全球糖产量的80%源自甘蔗^[1],种植区域超过100个国家,直接支撑超1亿人口的生计,部分国家农村人口的25%依赖甘蔗种植的收入生活^[2-3]。然而,甘蔗种植常受多种病害威胁,严重影响产量与品质^[4]。例如,甘蔗黄叶病可导致减产20%~50%,而梢腐病等真菌性病害则会直接破坏甘蔗糖分积累,进而波及制糖精炼及生物乙醇生产等下游产业^[5]。部分地区在病害暴发后减产幅度可达40%以上^[6],加剧了对化学农药的依赖,带来环境与健康风险。因此,准确及时地识别甘蔗病虫害对实现精准防治、减少农药使用和推动可持续农业至关重要^[7-8]。

随着遥感成像与机器学习技术的不断发展,基于图像的农作物叶片病害识别逐渐成为精准农业中的重要研究方向。JAVIDAN等^[9]提出了一种结合自动K-means聚类与多类支持向量机的算法,用于葡萄叶片病害的识别,通过自动分离病斑区域并提取多种颜色空间特征,实现了对黑斑病、黑腐病和枯叶病的高效分类。NI等^[10]利用傅立叶变换红外光谱技术结合变量重要性投影(VIP)与KNN模型,成功识别了玉米叶片上的灰斑病和北方玉米叶斑病,有效提升了特征维度压缩后的分类准确率。FENG等^[11]则融合高光谱、热红外与RGB图像信息,结合随机森林回归算法,构建了小麦白粉病病害指数的预测模型,实现了不同波段信息的协同建模。这些研究为农作物病害识别提供了有效途径,然而,传统机器学习方法仍依赖人工设计特征,泛化能力有限。

近年来,深度学习技术在图像识别与分类领域取得重要进展,为农业病害智能检测提供了新的解决方案。LIU等^[12]提出了一种基于PSOC - DRCNet的水稻病害识别模型,引入双模注意力机制(DMA)和残差自适应模块(RAB),有效缓解了图像背景干扰和边缘模糊的问题,并通过粒子群优化算法

(PSOC)提升训练策略优化效果。在自建数据集上的平均识别准确率达92.65%。WANG等^[13]针对番茄病害检测中的小目标识别困难、遮挡和复杂背景问题,提出了TomatoDet模型,结合Swin - DDETR注意力机制与改进型特征金字塔网络,实现了较高的识别准确率,平均准确率达92.3%。此外,AZGOMI等^[14]提出了一种基于图像处理和多层感知机神经网络(MLP)的苹果病害分类方法,该方法结构简单、成本较低,但分类准确率相对较低(73.7%)。

尽管现有算法一定程度上解决了病害分类问题,但甘蔗病斑识别仍存在边界模糊、背景干扰强及模型难以边缘部署的难点^[15]。为兼顾准确率与效率,解决上述特征提取难题,本文提出一种改进的轻量化模型SDCA - GLAM,以期用于田间自然场景下对11类甘蔗病害进行精准识别。

1 数据采集与处理

1.1 数据集构建

所用甘蔗样本来自国家级示范性平台和广西核心农作物种植基地:广西农业科学院(南宁市),甘蔗病害数据集采集于2024年5月,采用佳能EOS 600D型相机和iPhone13型手机,分早、中、晚3个时段拍摄,模拟真实田间全天候监测环境。为增强样本多样性,从门德利数据网站获取了部分甘蔗叶片图像数据。

所有原始图像统一预处理为768像素 \times 1 024像素,最终构建包含6 708幅图像的11类样本数据集,涵盖9类主要甘蔗病害^[16],额外引入“健康叶片”、“甘蔗与杂草混杂”2类非病害类别,用于增强模型对病害特征的辨别能力,具有良好的泛化能力。各病害类别及图像数量:带状萎黄病431幅、褐斑病1 722幅、褐锈病314幅、枯叶343幅、甘蔗与杂草混杂346幅、健康叶片430幅、梢腐病297幅、枯萎病652幅、黑穗病316幅、病毒663幅、黄叶病1 194幅。甘蔗叶片图像示例如图1所示。

数据集存在如下挑战:健康叶片与病害叶片混

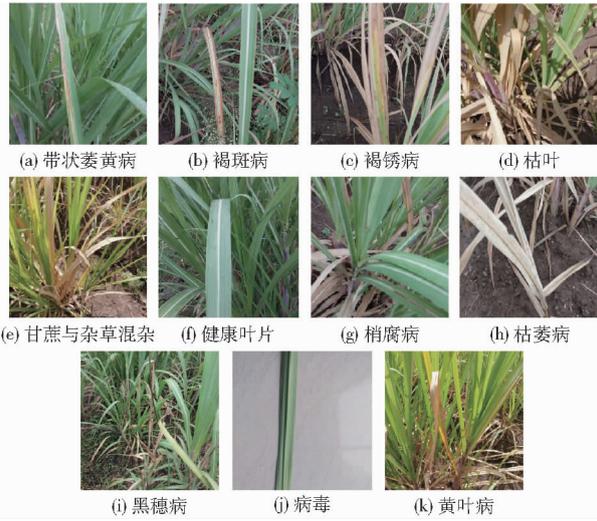


图1 甘蔗叶片图像

Fig. 1 Sugarcane leaf images

杂共存;不同类特征高度相似(例如病毒病与健康叶片、杂草苗与黄叶病);显著的光照干扰。这些复杂因素在提升识别难度的同时,也强化了数据集对田间真实场景的表征能力,为模型泛化性能提供有效支撑。

1.2 数据集预处理与数据增强

将原始数据集以比例8:2划分为训练集和测试集,确保训练集和测试集独立且没有重复。为了增加训练数据多样性,减少模型过拟合的可能,对训练集进行数据增强,方式有:随机裁剪、颜色抖动、随机增强、随机擦除、Mixup等,随机增强参照文献[17]的方法,包括位置调整(图像旋转、水平平移、垂直平移、水平剪切、垂直剪切)、颜色调整(直方图均衡化、颜色反转、曝光反转、对比度增强等)共15种增强方式,从其中随机挑选2种组合进行数据增强。部分数据增强操作生成的样本如图2所示。最终调整图像尺寸为224像素×224像素。

考虑到固定的数据增强方式可能导致模型过拟合特定的增强模式,采用在线数据增强策略。即在

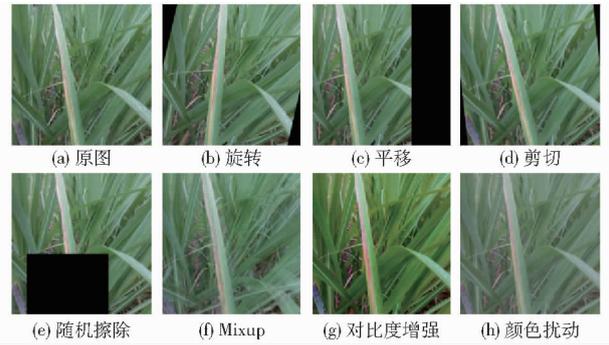


图2 图像增强示例

Fig. 2 Image enhancement examples

训练过程中,每幅训练图像在被加载时都会随机应用不同的增强操作,从而在每一训练轮次中生成多样化的样本,提升模型的泛化能力与鲁棒性。对验证集不进行任何图像增强操作,仅在转化为张量后对图像张量的均值和标准差进行标准化,以保证模型在评估阶段的输入一致性。

2 识别模型构建

2.1 SDCA-GLAM 网络模型结构

复杂自然环境下的田间甘蔗数据集存在不同类特征相似性高、背景噪声干扰强、病斑形态不规则等问题。为了更好地提取甘蔗叶片纹理和像素信息,以及提高模型推理速度,设计了一个轻量级混合架构网络SDCA-GLAM,该模型如图3所示。模型采用5级处理流:Patch编码层自适应捕捉病斑形变特征,特征提取模块DWMLP(Depthwise Conv MLP)强化局部纹理表达,下采样层压缩空间分辨率,自注意力层实现多尺度特征互补,最终由分类头完成高层语义聚合与类别预测。此架构在保留Transformer全局建模优势的同时,深度融合卷积操作的局部感知与计算效率特性。

2.2 位置投影

模型入口结构借鉴ViT框架,采用可变形卷积

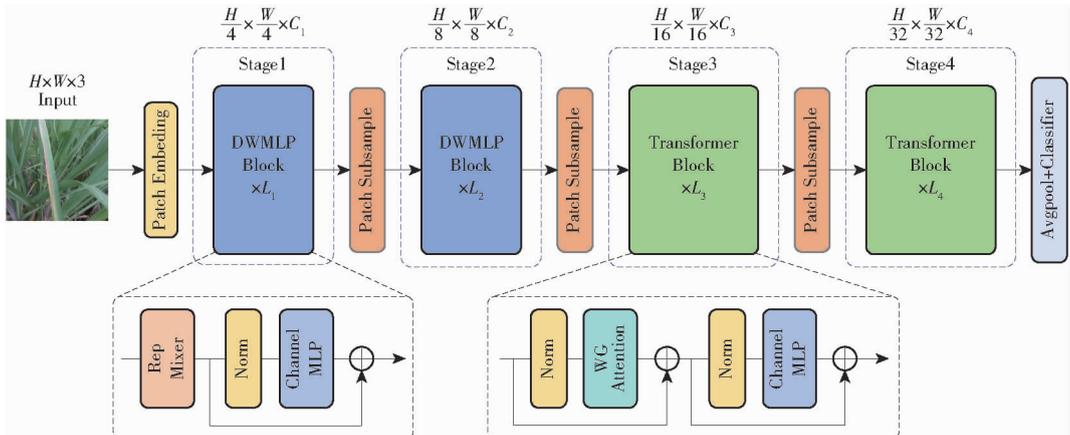


图3 模型框架图

Fig. 3 Model framework

Patch Embedding 层替代标准卷积操作。模块将输入图像分割为固定尺寸的 Patch 序列,并映射至高维特征空间。针对甘蔗病虫害图像中病斑形态不规则及背景干扰严重的特性,采用可变型卷积增强空间自适应能力。其结构由基础卷积核($K \times K$)与偏移量预测分支组成,参数量计算式为

$$M = C_{in} K^2 2K_0^2 \quad (1)$$

式中 M ——参数量

C_{in} ——输入通道数

K ——标准卷积尺寸

K_0 ——偏移量卷积尺寸

该设计通过动态调整采样位置,使模型聚焦于病斑关键区域而非背景噪声。部署层面,模块使用 C++ 编译实现核心运算,相较于解释型语言 Python,推理延迟降低,满足边缘设备实时性需求。

2.3 DWMLP 模块

ViT 基于注意力机制关注全局特征,与之相比,文献[18]提出的 PoolFormer 架构中替换了注意力模块,将 Self-Attention 替换为简单的空间池化算子,实验表明替换后即使使用不含任何参数的池化模块模型,在 ImageNet1K 上准确率也能达到 77.2%。ViT 整体架构设计是性能优越的关键因素,而非完全依赖于注意力模块。受此启发,提出移除注意力模块,采用可重参数化的深度卷积模块替代原 Self-Attention 结构,如图 4 所示。

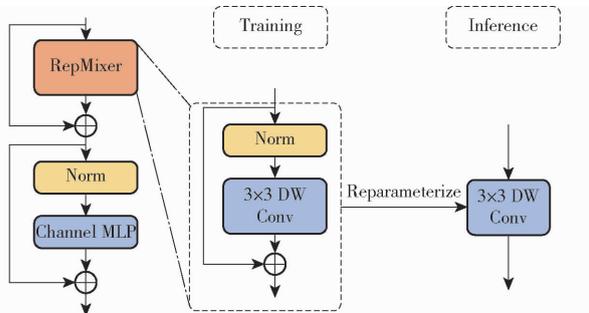


图 4 可重参数化的 RepMixer 模块

Fig. 4 RepMixer module with reparameterization capability

深度卷积通过解耦通道间的关联,仅对单组进行卷积运算。这一机制在保持特征提取能力的同时,大幅降低了内存占用量,即

$$M = C_{in} K^2 \quad (2)$$

为了提高模型容量,让模型能够自适应地提取甘蔗叶片信息,使用 MobileOne 作为深度卷积的骨干,如图 5 所示。

为了减少参数量,采用 N 为 1 的轻量化 MobileOne 配置。同时为了提高差异性,增强特征差异性,保留右侧 BN (BatchNorm) 分支,并引入 N 为 0 的相同模块进行相减操作执行差分运算,即

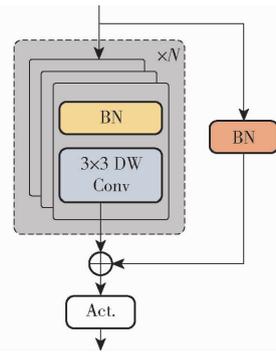


图 5 MobileOne 结构

Fig. 5 MobileOne architecture

$$F(x) = x + Conv_1(BN_1(x)) \times_{N_1} + BN_1(x) - (Conv_2(BN_2(x)) \times_{N_2} - BN_2(x)) \quad (3)$$

式中 $F(x)$ ——经过多个 MobileOne 模块得到的特征图

$Conv_i$ ——第 i 个卷积核

BN_j ——第 j 个 BN 层

N_k ——重复卷积 k 次

2 个 BN 层各自有独立的参数,使用减法操作可抵消相同部分,从而近似为只含深度卷积和残差连接模块。这样做可以保留 2 个 BN 层中的差异并提高模型容量,从而让模型学习到更多特征,即

$$F(x) = x + Conv(BN(x)) \quad (4)$$

算法将恒等映射替换为相同尺寸的卷积,卷积核权重矩阵为 $C_{out} \times C_{in} \times K^2$,人为创建 1 个相同矩阵,设定第 i 个 C_{out} 的第 j 个输入通道 (C_{out_i-j}) 的卷积核最中心为 1,其余输入通道和卷积核参数均为 0,这样在卷积时每个卷积核只有 1 个输入通道有输出,共输出 C_{out} 个通道,最终等效为一个恒等映射卷积,即

$$W_{i,j,u,v} = \begin{cases} 1 & (i=j \text{ 且 } u=v = \lfloor \frac{K}{2} \rfloor) \\ 0 & \text{(其他)} \end{cases} \quad (5)$$

式中 $W_{i,j,u,v}$ ——恒等映射卷积核权重矩阵元素

i ——卷积核输入通道数

j ——卷积核输出通道数

u ——卷积核横坐标

v ——卷积核纵坐标

对于含 BN 的分支,算法可以将 BN 与卷积进行融合,将式(4)的 $Conv(BN(x))$ 部分转换为

$$Conv(BN(x)) = x \frac{\gamma}{\sigma} W + \left(\beta - \mu \frac{\gamma}{\sigma} \right) W + b \quad (6)$$

式中 μ ——均值 γ ——缩放系数

σ ——标准差

β ——BN 偏置项

b ——卷积核偏置项

令 W' 为 $\frac{\gamma}{\sigma}W$, b' 为 $(\beta - \mu \frac{\gamma}{\sigma})W + b$, 构造一个与之等价的卷积, 完成该支路重参数化。

最终, 通过将多个支路重参数化后的卷积核进行加权合并, 得到一个等效的单一卷积核。这样在推理阶段, 模型无需保留多分支结构, 显著提升了计算效率。

在 MLP 模块优化方面(图 6), 特征图经升维操作后输入深度卷积层进行空间特征变换, 并采用 GeLU 激活函数引入非线性表达能力。该设计适配

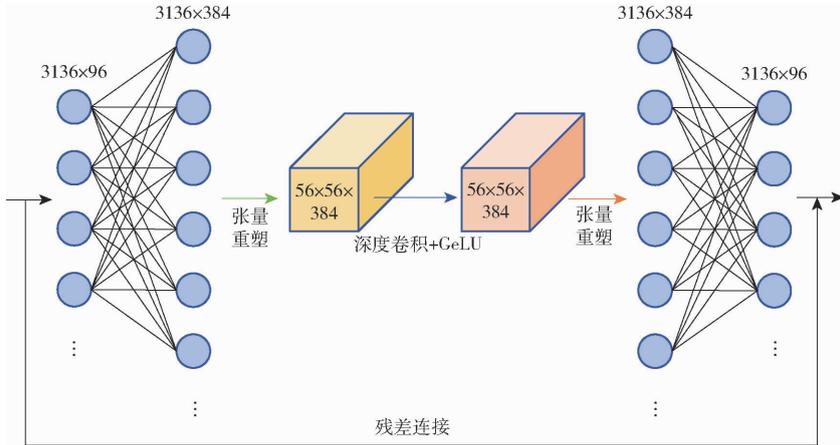


图 6 Channel MLP 模块

Fig. 6 Channel MLP module

2.4 Transformer 模块

当特征图经前级特征提取进入到后 2 个 stage 时, 模型采用 Transformer 架构进行像素序列建模。

为减少参数量并提升计算速度, 设计如图 7 所示的多分支注意力结构。

假定输入为 224 像素 × 224 像素, 在经历前 2

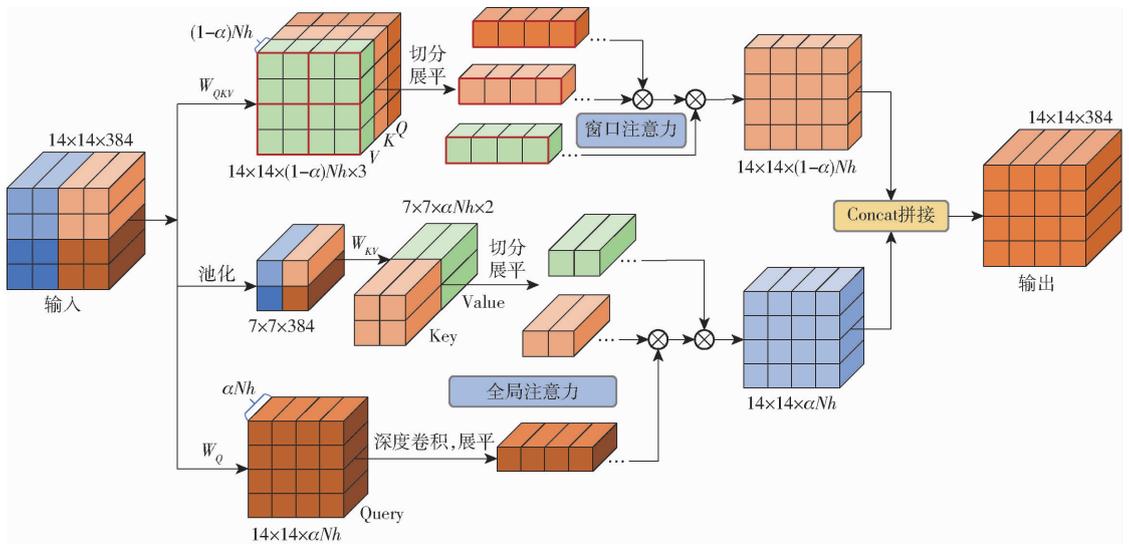


图 7 窗口-全局注意力

Fig. 7 Window - global attention

个下采样阶段后, 特征图尺寸降为 14 像素 × 14 像素, 此时引入注意力机制实现长距离建模。文献[19]的余弦相似度分析表明, 标准多头注意力 MSA (Multi-head self-attention) 中各头部存在高度相似性, 导致特征关注冗余。为了降低高额的注意力

成本, 算法将注意力头进行拆分, 采用 2 种不同注意力方式, 使用多支路形式进行计算。

局部支路[占比 $(1 - \alpha) \times N_h$ 个注意力头]采用窗口注意力机制, 在 $R \times R$ 局部区域内计算自注意力, 捕获细粒度纹理特征; 全局支路[占比 $\alpha \times N_h$ 个

注意力头]实施池化压缩注意力:先对特征图进行 $k \times k$ 空间池化压缩,再通过线性投影生成键/值矩阵(Key/Value),最终与完整分辨率的查询矩阵(Query)计算全局语义关联,显著降低KV维度带来的计算负担。标准MSA计算量为

$$\Omega(MSA) = 4hwC^2 + 2(hw)^2C \quad (7)$$

式中 $\Omega(MSA)$ ——标准多头注意力计算量

h, w ——特征图高度、宽度

C ——特征图通道数

而多支路计算量 Ω 为

$$\Omega = \Omega(W - MSA) + \Omega(P - MSA) \quad (8)$$

其中

$$\Omega(W - MSA) = hwC(1 - \alpha) [(4 - \alpha)C + 2R^2] \quad (9)$$

$$\Omega(P - MSA) = hwC\alpha \left[\left(\frac{2}{k^2} + \alpha + 1 \right) C + \frac{2}{k^2} hw \right] \quad (10)$$

式中 $\Omega(W - MSA)$ ——基于窗口的注意力计算量

$\Omega(P - MSA)$ ——基于池化的全局注意力计算量

α ——分配各支路注意力头数量超参数

R ——窗口注意力窗口尺寸

k ——池化步长

W-MSA(Window-MSA)算法仅在对应窗口内计算,与标准的注意力相比(假设 α 为0)。计算量减少 $2hwC(hw - M^2)$,W-MSA投影矩阵为 $(C, 3(1 - \alpha)N_h)$,生成QKV向量维度为

$$Q, K, V = hw(1 - \alpha)N_h \quad (11)$$

式中 Q ——注意力查询向量

K ——注意力键向量

V ——注意力值向量

N_h ——注意力头总数量

P-MSA(Pooling-MSA)首先将特征图进行池化(下采样)然后经过线性投影(维度为 $(C, 2\alpha N_h)$)生成KV向量的维度为

$$K, V = \frac{h}{k} \frac{w}{k} \alpha N_h \quad (12)$$

接着使用投影(维度为 $(C, \alpha N_h)$)生成Q矩阵,维度为

$$Q = hw\alpha N_h \quad (13)$$

再用深度卷积继续提取Q矩阵特征,并引入1个残差连接防止梯度消失。最后进行多头注意力将QKV矩阵融合,输出维度与Q矩阵相同。随后进行上下2条支路的融合,为了减少计算开销,采用Concat拼接张量,将W-MSA和P-MSA输出维度直接拼接,具体为

$$\text{OutPut}(x) = \text{Concat}(W - MSA(x), P - MSA(x)) \quad (14)$$

式中 $\text{OutPut}(x)$ ——特征图输出向量

Concat——沿着通道维度进行拼接

采用2种不同的注意力机制,可以有效帮助模型同时学习并捕捉全局与局部层面的特征信息。甘蔗病害通常表现为形态不一、分布不均的病斑,既有小面积的局部病斑,也可能存在大范围的感染区域。通过引入多分支结构融合,窗口注意力能够聚焦于局部区域,提取更加细致的病斑特征,而全局注意力则关注病斑与其周围环境之间的关联,从而识别出对分类或检测任务具有决定性作用的关键特征。

3 结果分析

3.1 实验环境及参数设置

实验在Ubuntu 20.04 LTS(64位)操作系统环境下进行,硬件平台搭载Intel Core i9-11900K处理器、NVIDIA GeForce RTX 3090显卡(24 GB显存)及64 GB DDR4运行内存。软件框架采用PyTorch 2.1.0深度学习库,基于CUDA 12.1加速架构,编程语言为Python 3.10。模型训练参数配置如下:使用AdamW优化器进行权重更新,初始学习率设置为0.001并附加0.05权重衰减系数,批大小(Batch size)固定为256样本,总训练轮次为300。在训练前20轮中采用线性学习率预热策略,使学习率逐步升至设定初始值,以避免训练初期梯度不稳定问题。SDCA-GLAM模型主要模块参数配置如表1所示。

3.2 评价指标

为了验证模型的分类性能,参照SAJITHA等^[20]的方法,采用准确率(Accuracy)、精确率(Precision)、召回率(Recall)和F1值(F1 score)作为模型的精度评价指标。同时,为了衡量模型实际应用适用性,引入吞吐量(Throughput)、参数量(Parameters)和浮点运算量作为模型计算效率与复杂度评估指标。

3.3 消融实验

在田间甘蔗数据集上,对SDCA-GLAM模型中的可变量卷积、RepMixer模块、窗口-全局注意力模块、深度卷积模块分别进行渐进式消融实验,结果见表2。基线模型LiT准确率为84.55%。针对甘蔗叶片中健康与病害区域混杂的特性,在模型前半部分引入可变量卷积改进Patch划分过程,通过自适应纹理提取使准确率提升至86.00%。随后添加RepMixer模块解决ViT模型的位置信息缺失问题,

表1 SDCA - GLAM 主要模块参数配置

Tab.1 Parameters settings of main modules in SDCA - GLAM

模块	输入尺寸	子模块类型	参数配置	输出尺寸	使用次数
Patch Embedding	224 × 224 × 3	DeConv	$K = 4, S = 4$	56 × 56 × 96	1
DWMLP Block 1	56 × 56 × 96	RepMixer	$K = 3, S = 1$	56 × 56 × 96	2
DWMLP Block 1	56 × 56 × 96	Channel MLP	$Exp. = 4$	56 × 56 × 96	2
Patch Subsample	56 × 56 × 96	Conv	$K = 2, S = 2$	28 × 28 × 192	1
DWMLP Block 2	28 × 28 × 192	RepMixer	$K = 3, S = 1$	28 × 28 × 192	2
DWMLP Block 2	28 × 28 × 192	Channel MLP	$Exp. = 4$	28 × 28 × 192	2
Patch Subsample	28 × 28 × 192	Conv	$K = 2, S = 2$	14 × 14 × 368	1
Transformer Block 1	14 × 14 × 368	Window Attention	$H = 1$	14 × 14 × 368	6
Transformer Block 1	14 × 14 × 368	Global Attention	$H = 11$	14 × 14 × 368	6
Transformer Block 1	14 × 14 × 368	DWConv	$K = 5, S = 1$	14 × 14 × 368	6
Transformer Block 1	14 × 14 × 368	Channel MLP	$Exp. = 4$	14 × 14 × 368	6
Patch Subsample	14 × 14 × 368	Conv	$K = 2, S = 2$	7 × 7 × 768	1
Transformer Block 2	7 × 7 × 768	Window Attention	$H = 0$	7 × 7 × 768	2
Transformer Block 2	7 × 7 × 768	Global Attention	$H = 24$	7 × 7 × 768	2
Transformer Block 2	7 × 7 × 768	DWConv	$K = 5, S = 1$	7 × 7 × 768	2
Transformer Block 2	7 × 7 × 768	Channel MLP	$Exp. = 4$	7 × 7 × 768	2
Avgpool	7 × 7 × 768	Avgpool, Flatten	$K = 7, S = 7$	768	1
Classifier	768	Linear, softmax	$W = 768 \times 11$	11	1

注: S 为卷积步长 (Stride), $Exp.$ 为 MLP 扩展比例 (Expansion), H 为注意力头数 (Head), W 为线性层权重矩阵维度 (Weight)。

表2 渐进式消融实验结果

Tab.2 Results of progressive ablation experiment

模块	参数量	吞吐量/(幅·s ⁻¹)	准确率/%
LiT	2.733×10^7	1 718	84.55
可变型卷积	2.731×10^7	1 783	86.00
RepMixer	2.734×10^7	1 734	86.82
深度卷积	2.744×10^7	1 710	87.44
MLP 卷积	2.780×10^7	1 614	88.22
多分支结构	2.758×10^7	1 518	88.56
BN	2.758×10^7	1 620	88.26

准确率进一步提升到 86.82%。在模型后半部分,首先考虑到查询向量 (Query) 维度显著高于键/值 (Key/Value) 向量,在生成查询向量后引入卷积操作以进一步提炼其关键特征。随后在 MLP 模块中融合深度卷积,构建 MBCConv 结构^[21],有效捕获高维空间中的重要语义信息,使准确率提升至 88.22%。接着引入多分支结构,使模型能够同时建模窗口级细粒度特征与全局长距离依赖关系^[22],进一步将准确率提升至 88.56%。为提升推理效率,最终统一采用 BatchNorm 替代原有 LayerNorm,避免重参数化过程中因归一化切换带来的频繁 reshape 操作,吞吐量提升 6.7%,准确率仅下降 0.3 个百分点。

3.4 不同模型性能对比实验

为验证 SDCA - GLAM 模型有效性并探究 CNN - Transformer 混合架构优势,选取 7 类基准模型进行对比: ResNet50、MobileOne、EfficientNetV2 构成的纯

CNN 模型组; ViT - B、Swin - Tiny 组成的纯 Transformer 模型组; 以及 FastViT - SA24、LITv2 和 RepViT 3 种混合架构模型。如表 3 所示, SDCA - GLAM 的准确率 (88.26%) 优于所有对比模型: 相较于纯 CNN 模型 ResNet50、EfficientNetV2 和 MobileOne 分别提升 3.17、4.52、8.63 个百分点; 较混合架构 LiTv2、FastViT、RepViT 分别提升 3.71、2.05、8.31 个百分点。推理效率方面,其吞吐量达 1 620 幅/s,较 ViT - B 与 Swin - Tiny 分别提升 1 227、759 幅/s,在混合架构中也保持优势,领先 FastViT、RepViT (分别为 341、840 幅/s),同时与纯 CNN 模型保持相当的计算效率。实验结果表明, SDCA - GLAM 在准确率 (88.26%) 与速度 (1 620 幅/s) 间实现了最优平衡,成为适用于田间复杂环境的高效检测解决方案。

3.5 可视化分析

为验证实验的有效性,对图像进行可视化分析,选取 6 个差异显著的病害类别进行可视化对比,结果如图 8 所示。由图 8 可以看出,原模型在甘蔗叶片密集区域存在显著识别盲区: 其纹理特征提取能力明显不足 (如腐烂叶片仅被局部识别),且过度依赖色彩信息导致枯叶与黄叶类别误判,这种纹理-色彩特征失衡现象揭示了模型将多维度病斑识别简化为色彩分类的本质缺陷。可视化结果证实了纹理特征缺失与识别准确率不足的因果关系,凸显 SDCA - GLAM 通过可变形卷积与 RepMixer 模块增强纹

表 3 对比实验结果

Tab. 3 Comparative experiment results

模型	参数量	浮点运算量	吞吐量/(幅·s ⁻¹)	准确率/%	精确率/%	召回率/%	F1 值/%
RepViT ^[23]	2.24×10^7	4.53×10^9	780	79.95	78.34	79.02	78.68
LiTv2 ^[24]	2.73×10^7	3.74×10^9	1 718	84.55	83.66	83.57	83.12
MobileOne ^[25]	1.29×10^7	3.03×10^9	1 823	79.63	80.21	79.97	80.09
ResNet50 ^[26]	2.56×10^7	4.14×10^9	1 546	85.09	84.76	84.73	84.74
ViT-B ^[27]	8.58×10^7	1.76×10^{10}	393	76.00	75.42	75.67	75.54
Swin-Tiny ^[28]	2.75×10^7	4.52×10^9	861	83.82	83.54	83.89	83.71
EfficientNetV2 ^[29]	2.14×10^7	7.52×10^9	1 468	83.74	83.88	83.76	83.82
FastViT ^[30]	2.06×10^7	3.76×10^9	1 279	86.21	86.34	86.72	86.53
SDCA-GLAM(本文)	2.76×10^7	3.77×10^9	1 620	88.26	87.54	88.52	88.02

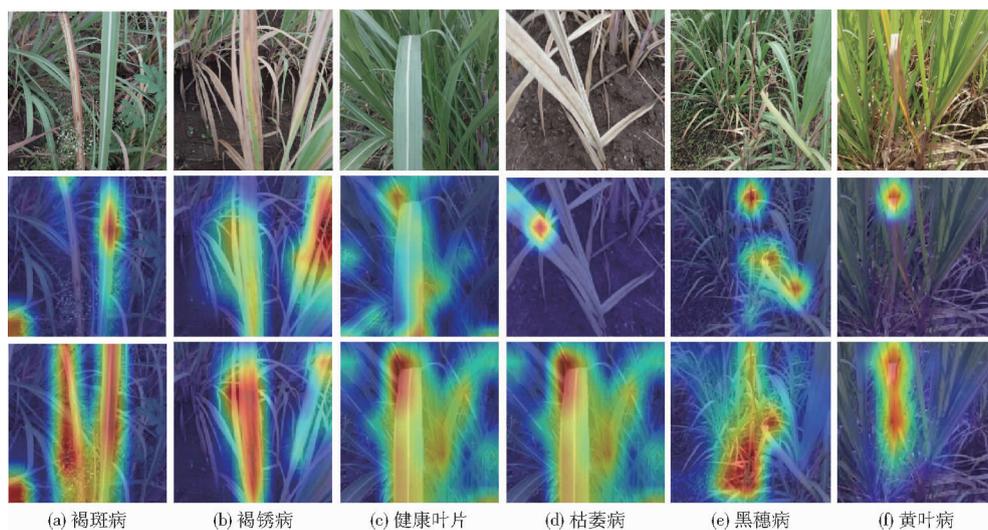


图 8 可视化结果

Fig. 8 Visualization results

理建模的必要性。

为深入分析模型在各类别上的分类性能,基于混淆矩阵进行实例化诊断(图 9)。由图 9 可知,基线模型对枯萎病(类别 7)和病毒病(类别 9)存在显著识别困难:由于特征提取能力不足,模型将枯萎病识别为褐斑病(类别 1)比例高达 61.7%,由图 7 可知模型关注区域过小导致易误判。与之形成鲜明对比的是在模型增加各种复杂模块后,枯萎病识别率高达 98.1%,充分验证了提取纹理特征对复杂病斑鉴别的关键作用。

3.6 超参数 α 对模型的影响

采用多支路的设计,使用超参数 α 人为地给 2 条支路分配不同的头数(0~12)。当窗口大小为 2 时,发现全局注意力的头数越多,模型吞吐量越大,这是因为窗口注意力中 reshape 等操作的计算花费了大部分时间,且这种操作不利于 GPU 的密集矩阵乘法操作。最终使用准确率最高的 α 为 0.95,如果完全使用全局注意力,那么模型可能存在每个查询向量都关注一个地方,而忽略自身和

周围的情况,导致出现冗余进而进度下降,结果如图 10 所示。

3.7 跨平台模型可迁移性验证

为了验证所提模型可迁移性,分别采用 ONNX (Open neural network exchange) 和 NCNN (N convolutional neural network) 2 种开源框架对模型进行推理测试。ONNX 和 NCNN 均具备良好的跨平台特性,可部署 Android、iOS、Windows、Linux 等操作系统。其中,ONNX 具有更强的兼容性,而 NCNN 则针对移动端进行了专门优化,具备更高的推理速度^[31-34]。

测试平台为 Windows 系统,CPU 为 Intel i5-12400F,开发环境采用 C++ 语言和 Visual Studio 工具。将模型在 Ubuntu 系统上完成训练后,分别使用 ONNX 和 NCNN 框架进行导出,并迁移至 Windows 平台进行 CPU 推理测试^[35-36],结果如表 4 所示。由表 4 可知,模型在跨平台部署后依然保持良好的性能。在 NCNN 框架中,即使仅使用 CPU 进行推理,处理速度仍能达到 67.9 幅/s,验证了模型在移动端部署的可行性和高效性。

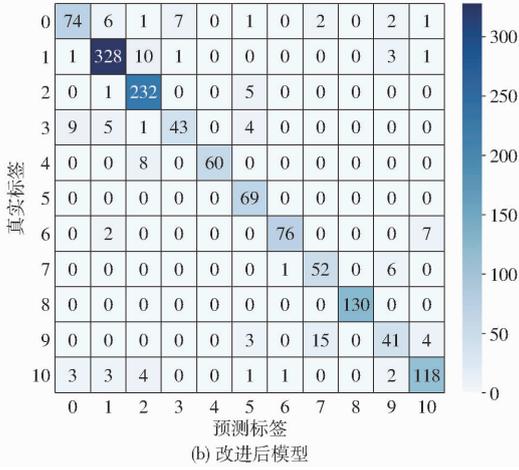
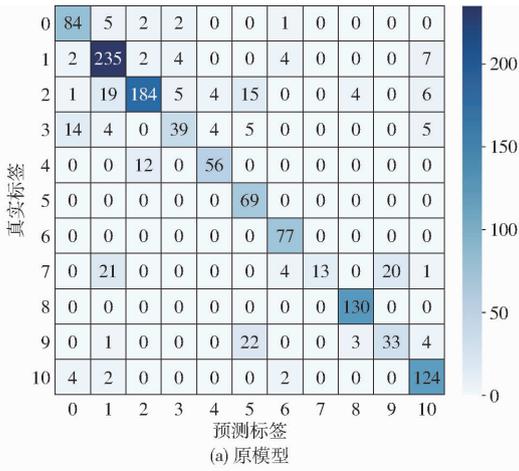


图9 混淆矩阵

Fig.9 Confusion matrix

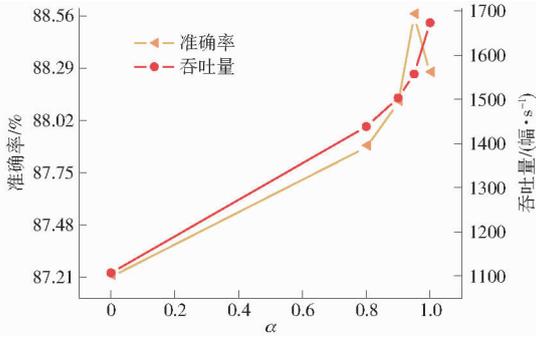


图10 超参数 α 对模型的影响

Fig.10 Influence of hyperparameter α on model

表4 模型推理速度

Tab.4 Model inference speed

幅/s

模型	ONNX	NCNN
ResNet50	49.3	66.1
EfficientNetV2	38.4	81.2
MobileOne	37.3	74.9
ViT - B	15.7	-
Swin - Tiny	33.4	-
FastViT	20.5	30.8
RepViT	24.2	29.3
LiTv2	29.4	69.0
SDCA - GLAM	28.7	67.9

注：- 表示该模型暂不支持导出。

4 结论

(1)通过引入可变形卷积模块、可重参数化的RepMixer 模块、MLP 中的深度卷积结构以及查询向量(Query)路径中的深度卷积模块,构建了多分支注意力融合机制,并将层归一化替换为批归一化,提出了一种轻量化的甘蔗病害识别模型 SDCA - GLAM。

(2)结果表明,该模型在保持较高吞吐量的同时,能够有效提升对 11 类甘蔗叶片样本的识别准确率。模型在测试集上分类准确率达到 88.26%,吞吐量为 1 620 幅/s。与改进前基线模型相比,识别准确率提升 3.71 个百分点;相较于常用轻量级 CNN 模型 ResNet50、EfficientNetV2 和 MobileOne,准确率分别提高 3.17、4.52、8.63 个百分点;相比 Transformer 类模型 ViT、Swin - Tiny、FastViT、RepViT 和 LiTv2,准确率分别提高 12.26、4.44、2.05、8.31、3.71 个百分点,且在吞吐量方面仍处于较高水平。此外,在移动端平台 NCNN 上的部署测试表明,模型在该平台实现了较高推理速度(67.9 幅/s),展现出良好的应用性能。

(3)该模型具备良好的跨平台兼容性,能够在不同系统与框架中高效运行,识别性能稳定,适用于移动端部署场景。研究成果为甘蔗病害智能诊断及轻量化模型的实际落地提供了新思路与参考。

参 考 文 献

[1] LI X, LI X, ZHANG S, et al. SLViT: shuffle-convolution-based lightweight vision transformer for effective diagnosis of sugarcane leaf diseases[J]. Journal of King Saud University-Computer and Information Sciences, 2023, 35(6): 101401.

[2] 黄亦其,刘祥焕,黄震宇,等.基于高光谱成像的甘蔗叶片早期轮斑病与锈病识别技术[J].农业机械学报,2022,54(4):259-267.

HUANG Yiqi, LIU Xianghuan, HUANG Zhenyu, et al. Identification of early wheel spot and rust on sugarcane leaves based on spectral analysis[J]. Transactions of the Chinese Society for Agricultural Machinery, 2022, 54(4): 259 - 267. (in Chinese)

[3] 李尚平,郑创锐,文春明,等.基于改进 YOLO v5s 的甘蔗切种茎节特征识别定位技术[J].农业机械学报,2023,54(10):234-245.

LI Shangping, ZHENG Chuangrui, WEN Chunming, et al. Stem node feature recognition and positioning technology for transverse cutting of sugarcane based on improved YOLO v5s [J]. Transactions of the Chinese Society for Agricultural Machinery, 2023, 54(10): 234 - 245, 293. (in Chinese)

- [4] ONG P, JIAN J, LI X, et al. New approach for sugarcane disease recognition through visible and near-infrared spectroscopy and a modified wavelength selection method using machine learning models [J]. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, 2023, 302: 123037.
- [5] 黄海荣,徐林,李翔,等. 274个甘蔗常用亲本主要叶部病害和黑穗病的抗性评价[J]. *热带作物学报*, 2022, 43(7): 1485 - 1496.
HUANG Hairong, XU Lin, LI Xiang, et al. Resistance evaluation of main leaf diseases and smut diseases in 274 commonly-used sugarcane parents [J]. *Chinese Journal of Tropical Crops*, 2022, 43(7): 1485 - 1496. (in Chinese)
- [6] 周成,陈章彬,杜雅刚,等. 面向边缘计算的水稻病害检测方法 with 装置研究[J]. *农业机械学报*, 2025, 56(4): 353 - 362.
ZHOU Cheng, CHEN Zhangbin, DU Yagang, et al. Development of rice disease detection methods and devices for edge computing [J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2025, 56(4): 353 - 362. (in Chinese)
- [7] 李云红,张蕾涛,李丽敏,等. 基于 CycleGAN - IA 方法和 M - ConvNext 网络的苹果叶片病害图像识别[J]. *农业机械学报*, 2024, 55(4): 204 - 212.
LI Yunhong, ZHANG Leitao, LI Limin, et al. Image recognition of apple leaf disease based on CycleGAN - IA method and M - ConvNext network [J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2024, 55(4): 204 - 212. (in Chinese)
- [8] 张林鎰,巴音塔娜,曾庆松. 基于 StyleGAN2 - ADA 和改进 YOLO v7 的葡萄叶片早期病害检测方法[J]. *农业机械学报*, 2024, 55(1): 241 - 252.
ZHANG Linxuan, BA Yintana, ZENG Qingsong. Grape disease detection method based on StyleGAN2 - ADA and improved YOLO v7 [J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2024, 55(1): 241 - 252. (in Chinese)
- [9] JAVIDAN S M, BANAKAR A, VAKILIAN K A, et al. Diagnosis of grape leaf diseases using automatic K-means clustering and machine learning [J]. *Smart Agricultural Technology*, 2023, 3: 100081.
- [10] NI Q, ZUO Y, ZHI Z, et al. Diagnosis of corn leaf diseases by FTIR spectroscopy combined with machine learning [J]. *Vibrational Spectroscopy*, 2024, 135: 103744.
- [11] FENG Z, SONG L, DUAN J, et al. Monitoring wheat powdery mildew based on hyperspectral, thermal infrared, and RGB image data fusion [J]. *Sensors*, 2021, 22(1): 31.
- [12] LIU Z, ZHOU G, ZHU W, et al. Identification of rice disease under complex background based on PSOC - DRCNet [J]. *Expert Systems with Applications*, 2024, 249: 123643.
- [13] WANG X, LIU J. An efficient deep learning model for tomato disease detection [J]. *Plant Methods*, 2024, 20(1): 61.
- [14] AZGOMI H, HAREDA SHT F R, MOTLAGH M R S. Diagnosis of some apple fruit diseases by using image processing and artificial neural network [J]. *Food Control*, 2023, 145: 109484.
- [15] 韦金菊,邓展云,黄诚华,等. 广西甘蔗主要真菌病害调查初报 [J]. *南方农业学报*, 2012, 43(9): 1316 - 1319.
WEI Jinju, DENG Zhanyun, HUANG Chenghua, et al. Preliminary report on investigating major sugarcane fungal diseases in Guangxi [J]. *Journal of Southern Agriculture*, 2012, 43(9): 1316 - 1319. (in Chinese)
- [16] VISWANATHAN R, RAO G P. Disease scenario and management of major sugarcane diseases in India [J]. *Sugar Tech.*, 2011, 13: 336 - 353.
- [17] CUBUK E D, ZOPH B, SHLENS J, et al. Randaugment: practical automated data augmentation with a reduced search space [C] // *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2020: 702 - 703.
- [18] YU W, LUO M, ZHOU P, et al. Metaformer is actually what you need for vision [C] // *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022: 10819 - 10829.
- [19] LIU X, PENG H, ZHENG N, et al. Efficientvit: memory efficient vision transformer with cascaded group attention [C] // *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023: 14420 - 14430.
- [20] SAJITHA P, ANDRUSHIA A D, ANAND N, et al. A review on machine learning and deep learning image-based plant disease classification for industrial farming systems [J]. *Journal of Industrial Information Integration*, 2024, 38: 100572.
- [21] QIN D, LEICHTNER C, DELAKIS M, et al. MobileNetV4: universal models for the mobile ecosystem [C] // *European Conference on Computer Vision*, 2024: 78 - 96.
- [22] XIONG Y, LI Z, CHEN Y, et al. Efficient deformable convnets: rethinking dynamic and sparse operator for vision applications [C] // *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024: 5652 - 5661.
- [23] WANG A, CHEN H, LIN Z, et al. Repvit: revisiting mobile CNN from ViT perspective [C] // *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024: 15909 - 15920.
- [24] PAN Z, CAI J, ZHUANG B. Fast vision transformers with holo attention [J]. *Advances in Neural Information Processing Systems*, 2022, 35: 14541 - 14554.
- [25] VASU P K A, GABRIEL J, ZHU J, et al. Mobileone: an improved one millisecond mobile backbone [C] // *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023: 7907 - 7917.
- [26] 付新阳,崔利华,董雨昕,等. 去除土壤背景影响的多光谱遥感影像玉米叶面积指数估算 [J]. *农业机械学报*, 2025, 56(5): 384 - 394.
FU Xinyang, CUI Lihua, DONG Yuxin, et al. Estimation of maize leaf area index from multi-spectral remote sensing with soil background effects removed [J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2025, 56(5): 384 - 394. (in Chinese)

- [27] DOSOVITSKIY A, BEYER L, KOLESNIKOV A, et al. An image is worth 16x16 words: transformers for image recognition at scale[J]. arXiv Preprint, arXiv:2010.11929, 2020.
- [28] LIU Z, LIN Y, CAO Y, et al. Swin transformer: hierarchical vision transformer using shifted windows[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021: 10012 – 10022.
- [29] ZHU L, WANG X, KE Z, et al. Biformer: vision transformer with bi-level routing attention[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023: 10323 – 10333.
- [30] VASU P K A, GABRIEL J, ZHU J, et al. Fastvit: a fast hybrid vision transformer using structural reparameterization[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision, 2023: 5785 – 5795.
- [31] 范耀冰, 吴尚蓉, 匡炜, 等. 基于无人机高光谱影像的稻谷氮含量估算研究[J]. 农业机械学报, 2025, 56(1): 332 – 343, 423.
FAN Yaobing, WU Shangrong, KUANG Wei, et al. Estimation of nitrogen content in rice grains based on UAV hyperspectral imagery[J]. Transactions of the Chinese Society for Agricultural Machinery, 2025, 56(1): 332 – 343, 423. (in Chinese)
- [32] TOLSTIKHIN I O, HOULSBY N, KOLESNIKOV A, et al. Mlp-mixer: an all-mlp architecture for vision[J]. Advances in Neural Information Processing Systems, 2021, 34: 24261 – 24272.
- [33] DONG X, BAO J, CHEN D, et al. Cswin transformer: a general vision transformer backbone with cross-shaped windows[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022: 12124 – 12134.
- [34] DAI J, QI H, XIONG Y, et al. Deformable convolutional networks[C]//Proceedings of the IEEE International Conference on Computer Vision, 2017: 764 – 773.
- [35] 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. (in Chinese)
- [36] ZENG T, LI S, SONG Q, et al. Lightweight tomato real-time detection method based on improved YOLO and mobile deployment[J]. Computers and Electronics in Agriculture, 2023, 205: 107625.

(上接第 186 页)

- [28] 陈松林, 王天星. 等间距法和均值标准差法界定城市热岛的对比研究[J]. 地球信息科学学报, 2009, 11(2): 145 – 150.
CHEN Songlin, WANG Tianxing. Comparison analyses of equal interval method and mean-standard deviation method used to delimitate urban heat island[J]. Journal of Geo-information Science, 2009, 11(2): 145 – 150. (in Chinese)
- [29] SHIH M Y, JHENG J W, LAI L F. A two-step method for clustering mixed categorical and numeric data[J]. Journal of Applied Science and Engineering, 2010, 13(1): 11 – 19.
- [30] 储岳中. 一类基于贝叶斯信息准则的 k 均值聚类算法[J]. 安徽工业大学学报(自然科学版), 2010, 27(4): 409 – 412.
CHU Yuezhong. An k means clustering algorithm based on Bayesian information criterion[J]. Journal of Anhui University of Technology (Natural Science), 2010, 27(4): 409 – 412. (in Chinese)
- [31] WANG J F, LI X H, CHRISTAKOS G, et al. Geographical detectors-based health risk assessment and its application in the neural tube defects study of the Heshun Region, China[J]. International Journal of Geographical Information Science, 2010, 24(1): 107 – 127.
- [32] 王劲峰, 徐成东. 地理探测器: 原理与展望[J]. 地理学报, 2017, 72(1): 116 – 134.
WANG Jinfeng, XU Chengdong. Geodetector: principle and prospective[J]. Acta Geographica Sinica, 2017, 72(1): 116 – 134. (in Chinese)
- [33] 张英, 冯雪珂, 任少宝, 等. 耕地质量和产能评价指标体系研究——以广西宾阳县为例[J]. 农业资源与环境学报, 2021, 38(6): 1039 – 1050.
ZHANG Ying, FENG Xueke, REN Shaobao, et al. Evaluation index system of cultivated land quality and productivity: a case study of Binyang County, Guangxi[J]. Journal of Agricultural Resources and Environment, 2021, 38(6): 1039 – 1050. (in Chinese)
- [34] 陈卓. 邓州市耕地质量评价与分区保护对策研究[D]. 郑州: 河南农业大学, 2023.
CHEN Zhuo. Research on evaluation and zoning protection strategy of Dengzhou Country's arable land quality[D]. Zhengzhou: Henan Agricultural University, 2023. (in Chinese)
- [35] 中国土壤调查办公室. 全国第二次土壤普查养分分级标准[M]. 北京: 中国农业出版社, 1979.
- [36] 唐炬彪, 朱庆伟, 董士伟, 等. 农业用地土壤重金属样本点数据精化方法——以北京市顺义区为例[J]. 农业环境科学学报, 2020, 39(10): 2288 – 2296.
TANG Guibiao, ZHU Qingwei, DONG Shiwei, et al. Data refinement method for sampling sites of agricultural soil heavy metals: a case study in Shunyi District, Beijing, China[J]. Journal of Agro-Environment Science, 2020, 39(10): 2288 – 2296. (in Chinese)
- [37] DE LEEUW J, MAIR P. Multidimensional scaling using majorization: SMACOF in R[J]. Journal of Statistical Software, 2009, 31(3): 1 – 30.