

# 基于 AgriSwin 的植物病虫害检测算法

刘微 张傲

(沈阳理工大学信息科学与工程学院 沈阳 110159)

**摘要:** 针对现代农业中植物病虫害检测所面临的多尺度特征和复杂背景处理难题,本文提出了一种高效且精准的检测模型 AgriSwin,以提升农业病虫害检测的精度和效率。AgriSwin 模型在 Swin Transformer 的基础上,融合了扩张特征聚合模块与自适应空间卷积模块。扩张特征聚合模块通过不同扩张率的卷积层实现多尺度特征提取,并利用全局特征信息的自适应加权机制优化了特征融合效果。自适应空间卷积模块则通过生成自适应权重,对特征图进行动态加权,从而在复杂背景下增强局部和全局信息的捕捉能力。实验结果表明,AgriSwin 模型在 PlantDoc、PlantVillage 和自建数据集上的检测精确率分别达到 79.65%、99.90% 和 95.08%。此外,该模型的参数量比 Swin Transformer-T 减少了 25.63%,在保持高精确率的同时显著降低了内存和计算资源的占用,展示了在大规模农业应用中的广泛潜力。

**关键词:** 植物病虫害检测;深度学习;卷积神经网络;多尺度卷积;自适应空间卷积;特征聚合;农业自动化

中图分类号: TN911.7 文献标识码: A 国家标准学科分类代码: 460.6

## Plant disease and pest detection algorithm based on AgriSwin

Liu Wei Zhang Ao

(College of Information Science and Engineering, Shenyang Ligong University, Shenyang 110159, China)

**Abstract:** To address the challenges of multi-scale features and complex background processing in plant pest and disease detection in modern agriculture, this paper proposes an efficient and accurate detection model, AgriSwin, to improve the precision and efficiency of agricultural pest and disease detection. The AgriSwin model is based on the Swin Transformer and integrates a dilated feature aggregation module and an adaptive spatial convolution module. The dilated feature aggregation module extracts multi-scale features through convolutional layers with different dilation rates and optimizes feature fusion using an adaptive weighting mechanism for global feature information. The adaptive spatial convolution module generates adaptive weights to dynamically weight the feature maps, enhancing the ability to capture both local and global information in complex backgrounds. Experimental results show that the AgriSwin model achieves detection accuracies of 79.65%, 99.90%, and 95.08% on the PlantDoc, PlantVillage, and custom datasets, respectively. Additionally, the model's parameter count is reduced by 25.63% compared to Swin Transformer-T, significantly lowering memory and computational resource requirements while maintaining high accuracy, demonstrating its broad potential for large-scale agricultural applications.

**Keywords:** plant disease and insect pest detection; deep learning; convolutional neural network; multiscale convolution; adaptive spatial convolution; feature aggregation; agricultural automation

## 0 引言

植物病害对全球农业生产构成了严重威胁,直接影响作物的产量和质量。随着全球气候变化和农业生产规模的扩大,病虫害的频率和严重性不断增加,给植物健康管理带来了更大的挑战。传统的病害检测依赖人工观察,效率低且难以满足大规模农业生产的需求<sup>[1]</sup>。因此,开发高效、自

动化的病害检测技术尤为重要。

深度学习技术,特别是卷积神经网络(convolutional neural network, CNN)和Transformer,在植物病害检测领域取得了显著进展。文献[2]提出了基于 CNN 模型的植物病害检测方法,有效提高了检测的准确度;文献[3]提出了混合卷积自动编码器与 CNN 相结合的模型,在复杂背景下具备较强的识别能力。然而,CNN 在全局特征提取和

复杂背景处理方面仍存在一定局限性。文献[4-5]指出, CNN 在多尺度特征提取方面的表现不尽如人意,特别是在复杂场景中表现较弱。

为了应对这些问题, Thakur 等<sup>[6]</sup>提出的 PlantViT 模型结合了 Vision Transformer, 利用自注意力机制显著提升了全局信息的捕捉能力, 特别是在复杂背景下。文献[7]的研究进一步证实, Transformer 在复杂场景中的病害分类能力优于传统方法。Yu 等<sup>[8]</sup>通过将 Inception 卷积与 Transformer 结合, 进一步优化了多尺度特征处理。文献[9]提出了一种改进算法, 针对小区域病害的检测, 在复杂环境下展现了更强的适应性。然而, 现有研究仍在计算资源消耗和实时性方面面临挑战, 这也是植物病害检测在大规模农业应用中的主要瓶颈。

本文提出了一种基于改进 Transformer 架构的植物病害检测模型——农业 Swin Transformer(agricultural swin transformer, AgriSwin), 旨在优化多尺度特征提取和复杂背景处理能力, 并降低计算资源占用, 实现大规模农业中的实时检测。AgriSwin 结合了 CNN 的局部特征提取优势与 Transformer 的全局信息捕捉能力。通过扩展特征聚合模块, 强化了多尺度信息的提取; 自适应空间卷积模块则动态调整特征图的权重分配, 增强了模型在复杂背景下的鲁棒性。

为验证 AgriSwin 的有效性, 本文使用 PlantDoc、

PlantVillage 和自建数据集进行多场景实验, 测试模型在复杂背景下的检测精度、鲁棒性及资源消耗情况。通过与现有主流模型的对比, 展示 AgriSwin 在复杂农业场景中的优越性。

AgriSwin 有望显著提升植物病害检测的精度与效率, 并通过降低计算资源需求, 实现实时检测。该模型将为智能农业的推广提供技术支持, 助力全球粮食安全的保障。

## 1 AgriSwin 模型

尽管卷积神经网络和视觉 Transformer 在视觉任务中表现出色, 但在处理复杂背景时, CNN 难以捕捉全局信息, 而 Transformer 对细节特征的敏感度不足。为了解决上述问题, 本文提出了 AgriSwin 模型(如图 1 所示)。该模型包括 4 个主要阶段, 其中 Stage 1~Stage 3 采用了 Swin Transformer Block, 在每个阶段之间通过 Patch Merging 操作进行降采样和特征融合。在 Stage 4 中, 首先, 扩张特征聚合模块通过引入扩张卷积, 扩大感受野, 增强模型对不同尺度目标的感知能力, 特别是在处理大小和形态变化显著的病斑时, 该模块能够有效聚合多尺度特征信息, 确保模型在复杂场景下的鲁棒性。其次, 自适应空间 Transformer 模块结合了自适应空间卷积, 通过动态调整权重, 提升局部特征提取的能力。自适应机制使得模型能够根据不同区域的特征需求, 自动调整提取过程, 进一步提升检测精度。最终, 模型通过分类层输出检测结果。

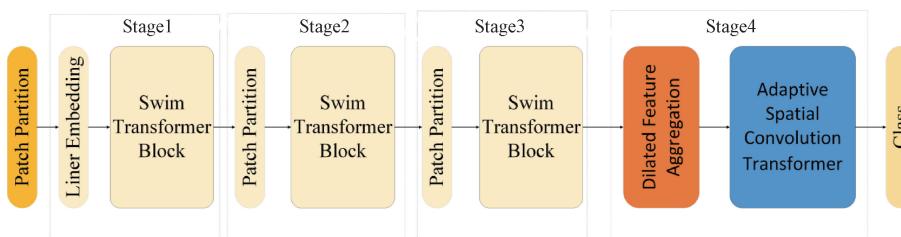


图 1 AgriSwin 结构图  
Fig. 1 AgriSwin model architecture

### 1.1 扩张特征聚合模块

为解决植物病虫害检测任务中的多尺度特征提取问题, 本文提出了扩张特征聚合模块(如图 2 所示)。该模块通过不同扩张率的卷积层实现多尺度特征提取, 并结合全局特征信息的自适应加权机制, 优化了不同尺度特征的融合效果<sup>[10]</sup>。通过将多尺度扩张卷积与自适应加权相结合, 模型的病虫害检测能力得到了增强, 尤其是在应对不同大小、形态和复杂背景的目标时表现优异。

#### 1) 特征转换与初步特征提取

输入图像首先经过标准卷积操作进行初步特征提取。在保持输入与输出空间维度一致的前提下, 卷积操作为后续的多尺度卷积处理提供了有效的初始特征表达。通过以下步骤, 确保特征转换的准确性:

$$X_{conv} = \text{Conv2D}(X, k=3, s=1, p=1) \quad (1)$$

其中,  $X_{conv}$  为转换后的特征图,  $k$  为卷积核大小,  $s$

为步长,  $p$  为填充, 确保输出与输入的空间维度一致。此操作确保输入特征具有适合后续多尺度处理的表达形式。

#### 2) 多尺度扩张卷积层

为提取不同尺度的特征, 本模块采用多个具有不同扩张率的卷积层, 每个卷积层通过调整扩张率  $r$  来改变卷积核的感受野。具体来说, 扩张率  $r=1, 3, 6, 9$  分别对应 4 个卷积层:

$$Y_i = \text{DilatedConv2D}(X_{conv}, r=i) \quad (2)$$

这里,  $Y_i$  代表扩张卷积后的特征图,  $r$  为扩张率, 用于定义卷积操作的感受野大小。通过采用不同的扩张率, 模型能够有效捕捉从局部到全局的多尺度特征, 同时不增加参数量。将扩张卷积与标准卷积结合使用, 使模型在保持较低计算量的前提下, 增强了感受野的灵活性和多样性, 从而更好地适应病斑或病害目标的多样化表现。

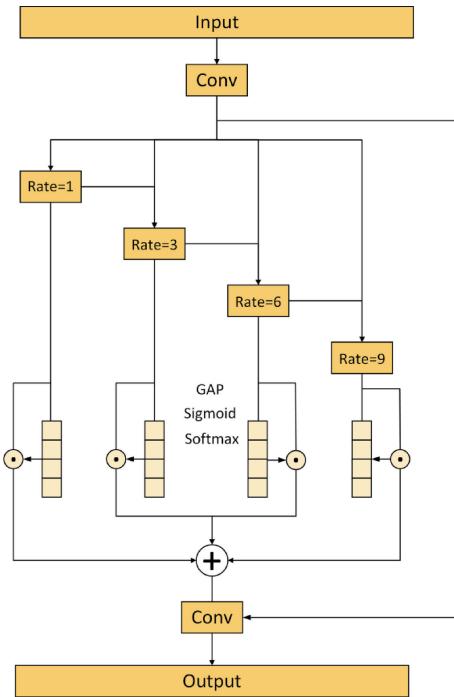


图 2 扩张特征聚合模块

Fig. 2 Dilated feature aggregation module diagram

### 3) 全局特征提取与自适应权重生成

为了进一步增强多尺度特征的表达能力,本文设计了基于全局特征的自适应加权机制<sup>[11]</sup>。对于每个尺度的特征图  $Y_i$ , 通过全局平均池化(global average pooling, GAP)和全局最大池化(global max pooling, GMP)提取其全局信息,接着,将这些全局信息拼接在一起:

$$W_{avg,i} = \text{GAP}(Y_i) \quad (3)$$

$$W_{max,i} = \text{GMP}(Y_i) \quad (4)$$

$$W_i = \text{Concat}(W_{avg,i}, W_{max,i}) \quad (5)$$

其中,  $W_{avg,i}$  表示通过全局平均池化提取的全局信息,而  $W_{max,i}$  则表示通过全局最大池化提取的全局信息,  $W_i$  为融合后的全局特征信息。该操作通过整合每个尺度的全局特征,获取不同尺度的重要性。

然后,通过一个卷积操作生成自适应权重矩阵:

$$\alpha_i = \text{Sigmoid}(\text{Conv2D}(W_i, k=7, s=1, p=3)) \quad (6)$$

在此处  $\alpha_i$  为生成的自适应权重,用于对特征进行动态加权, Sigmoid 函数用于生成在(0,1)之间的非线性权重值,确保特征权重的可训练性和有效性。

### 4) 权重归一化与多尺度特征融合

为了确保不同尺度特征在融合时具有合理的平衡性,本文通过 Softmax 函数对生成的权重进行归一化处理:

$$\alpha_{norm,i} = \text{Softmax}(\alpha_i) \quad (7)$$

这里,  $\alpha_{norm,i}$  表示归一化后的权重,用于控制不同尺度特征的权重分配。通过归一化,保证每个尺度的权重总和为 1,进而防止某一尺度特征对融合结果的过度影响。加

权后的多尺度特征融合表示为:

$$Y_{fused} = \sum_{i=1}^4 \alpha_{norm,i} \cdot Y_i \quad (8)$$

通过对不同尺度特征进行加权求和,得到最终融合的特征图  $Y_{fused}$ , 该特征图包含了来自不同尺度的丰富上下文信息,增强了模型对多尺度目标的鲁棒性。

### 5) 残差连接与输出

为了保留输入特征的原始信息,本文通过残差连接将融合后的特征图与输入特征图相加,形成最终输出:

$$Y_{out} = \text{Conv2D}(Y_{fused} + X, k=1) \quad (9)$$

其中,  $Y_{out}$  为最终输出的特征图,  $Y_{fused}$  是经过多尺度特征融合后的特征图,  $X$  为输入特征图,  $k$  表示卷积核大小,此处设置为 1,以确保融合后的特征能够被有效提取和处理。残差连接不仅能够防止梯度消失,还能够保留输入特征的信息,实现更为高效的特征表达<sup>[12]</sup>。

## 1.2 自适应空间 Transformer

自适应空间 Transformer(如图 3 所示)结合了 CNN 和 Transformer 的优势,通过将自适应空间卷积模块引入到 Transformer 架构中,有效增强了特征提取能力。

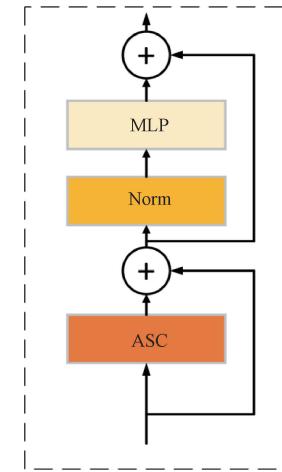


图 3 自适应空间 Transformer

Fig. 3 Adaptive spatial transformer diagram

传统卷积神经网络在处理多尺度特征时,通常依赖固定大小的卷积核。尽管感受野可以通过网络层次的叠加逐渐扩大,但在应对复杂场景或多尺度目标时仍可能出现问题。例如,在植物病虫害检测任务中,病斑或病害的大小和形状高度多样,固定大小的卷积无法同时有效捕捉细小病斑和大范围病害。此外,卷积网络在特征融合时缺乏动态调整的能力,导致不同尺度特征难以充分结合。

为克服这些问题,本文提出了自适应空间卷积模块(如图 4 所示),通过以下几个方面有效处理了多尺度上下文中的丰富空间特征,并兼顾计算效率:

首先,该模块引入了不同扩张率的卷积核,能够动态调整感受野的大小,以便同时处理不同尺度的特征。这种方法使模型既能捕捉细小病斑,又能识别大范围病害目

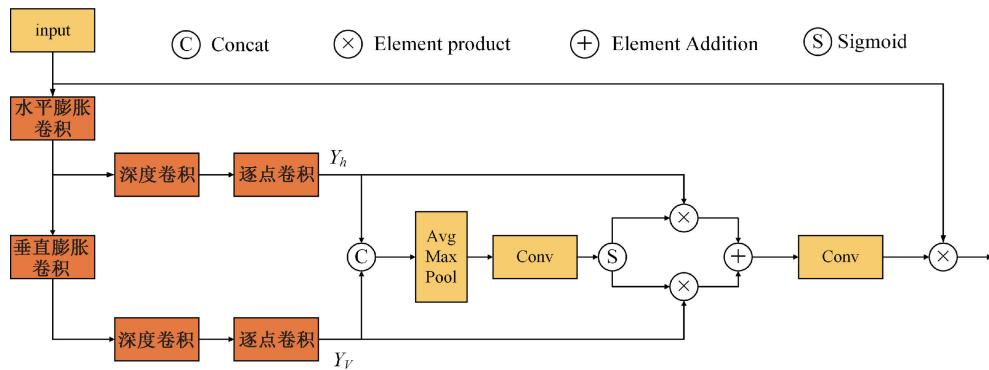


图 4 自适应空间卷积模块

Fig. 4 Adaptive spatial convolution module diagram

标,增强了对多尺度特征的感知能力,避免了固定卷积核的局限性。

其次,模块采用自适应特征增强机制,结合全局平均池化和最大池化提取全局特征信息,并生成自适应权重。这一机制能够动态调整特征图中不同区域的权重,确保多尺度特征在融合过程中合理加权,从而提升复杂背景下的检测精度。

最后,为了减少计算开销,该模块结合了深度卷积和逐点卷积(depthwise convolution 和 pointwise convolution)。深度卷积在每个通道上独立操作,降低了计算复杂度,而逐点卷积通过  $1 \times 1$  卷积核将各通道特征进行压缩与融合,有效降低计算量,同时确保多尺度特征的高效融合。

### 1) 多方向卷积处理

在传统卷积中,固定方向的卷积核往往难以充分捕捉不同方向上的细节特征。为解决这一问题,模块采用了水平方向和垂直方向的膨胀卷积操作相结合的方式。水平卷积核的操作为:

$$Y_{\text{horizontal}} = \text{DilatedConv2D}_{\text{horizontal}}(X) \quad (10)$$

垂直方向的卷积核为:

$$Y_{\text{vertical}} = \text{DilatedConv2D}_{\text{vertical}}(X) \quad (11)$$

其中,  $Y_{\text{horizontal}}$  表示通过水平膨胀卷积生成的特征图,  $Y_{\text{vertical}}$  表示通过垂直膨胀卷积生成的特征图。水平膨胀卷积和垂直膨胀卷积的卷积核形状分别为  $(1, 3)$  和  $(3, 1)$ , 膨胀因子为 2。这种设计不仅能够有效捕捉水平方向的长条形特征,还能在垂直方向上保持更高的特征完整性,提升模型在不同场景下的适应性。

### 2) 深度卷积与逐点卷积的结合

传统的标准卷积在计算量与参数量上存在较大的开销,为了进一步提升特征融合效率,本文在水平和垂直方向膨胀卷积后引入了深度卷积与逐点卷积的结合<sup>[13]</sup>。首先,深度卷积被用于每个通道上,以保留局部细节特征,避免特征在初步融合时丢失。随后,逐点卷积通过  $1 \times 1$  的卷积核对通道进行压缩,并将通道间不同通道的特征进行融合,提升整体特征表达能力。具体的计算过程如下:

$$Y_{\text{depthwise}} = \text{DepthwiseConv2D}(Y) \quad (12)$$

$$Y_{\text{pointwise}} = \text{PointwiseConv2D}(Y_{\text{depthwise}}) \quad (13)$$

其中,  $Y_{\text{depthwise}}$  和  $Y_{\text{pointwise}}$  分别表示经过深度卷积与逐点卷积处理后的特征图。这种组合方式不仅减少了计算量,还提升了多通道特征的融合效率,确保了卷积操作的轻量化与高效性。

然后,将  $Y_{\text{horizontal}}$  经过深度卷积和逐点卷积处理后生成的水平卷积特征图  $Y_h$ , 和  $Y_{\text{vertical}}$  经过深度卷积和逐点卷积处理后生成的垂直卷积特征图  $Y_v$  进行 concat 拼接操作生成  $Y_{\text{concat}}$ 。

### 3) 自适应特征增强机制

为了更加有效地动态调整特征图中不同位置的重要性,本文设计了一种改进的全局信息提取与权重生成机制,旨在提升模型在多尺度特征提取中的表现。该机制结合了全局平均池化与最大池化操作,生成全局特征,并通过自适应权重对不同尺度特征进行增强,以应对复杂的背景和不同大小的目标<sup>[14]</sup>。

首先,通过全局平均池化和最大池化操作从输入特征图中提取全局特征。这两个操作分别提取输入特征的全局平均信息  $\text{avg}_{\text{attn}}$  与最大值信息  $\text{max}_{\text{attn}}$ , 能够综合捕捉不同区域的全局特征。为了进一步增强特征表达能力,本文将这两种全局特征拼接在一起,生成融合后的特征表示  $\text{agg}$ 。具体的计算过程如下:

$$\text{avg}_{\text{attn}} = \text{GAP}(Y_{\text{concat}}) \quad (14)$$

$$\text{max}_{\text{attn}} = \text{GMP}(Y_{\text{concat}}) \quad (15)$$

$$\text{agg} = \text{Concat}(\text{avg}_{\text{attn}}, \text{max}_{\text{attn}}) \quad (16)$$

接下来,将拼接后的特征通过一个  $7 \times 7$  的卷积层生成自适应权重:

$$\sigma = \text{Sigmoid}(\text{Conv2D}(\text{agg}, \text{kernelsize} = 7, \text{padding} = 3)) \quad (17)$$

其中,  $\sigma$  表示生成自适应权重,应用 Sigmoid 门控函数以获得水平自适应权重  $\sigma_{\text{horizontal}}$  和垂直自适应权重  $\sigma_{\text{vertical}}$ 。该自适应权重机制能够动态调整特征图中不同区域的重要性,从而更好地捕捉细节特征与全局信息的平衡。

这一步操作通过逐点相乘的方式,将自适应权重分别应用到水平卷积特征图  $Y_h$  和垂直卷积  $Y_v$  后的特征图上,增强不同方向上的特征融合效果:

$$Y_{final} = \sigma_{horizontal} \cdot Y_h + \sigma_{vertical} \cdot Y_v \quad (18)$$

其中,  $Y_{final}$  表示最终的融合特征图。通过这种动态调整机制,模型能够在复杂背景和多尺度目标的场景下,更精确地提取关键特征,提升整体检测精度和鲁棒性。

## 2 实验

### 2.1 实验环境与参数设置

本次实验在以下硬件和软件环境中进行:CPU 采用 14 核的 Intel(R) Xeon(R) Platinum 8362 @ 2.80 GHz; GPU 为 NVIDIA RTX 4090,显存容量为 24 GB。软件环境包括 Python 3.8、Pytorch 1.11.0 及 CUDA 11.3。

实验中,输入图像的大小设置为  $224 \times 224$ ,训练轮次 (epochs) 为 30,批处理大小 (batch size) 为 16,初始学习率 (learning rate) 为 0.005,权重衰减系数 (weight decay) 为 0.0005。此外,实验采用随机梯度下降 (stochastic gradient descent, SGD) 优化器,并在网络训练过程中对学习率进行了动态调整,以进一步提高模型的收敛速度和整体性能。

### 2.2 实验数据集

本研究使用了两个公共数据集来验证所提出模型的有效性:PlantDoc 数据集<sup>[15]</sup>、PlantVillage 数据集<sup>[16]</sup>和自建数据集。这 3 个数据集涵盖了丰富的植物病害信息,并包含不同的植物种类、病害类型以及复杂背景条件。

#### 1) PlantDoc 数据集

PlantDoc 是一个专注于在复杂条件下进行植物病害检测的公开数据集。此数据集包含了 2 598 张从实际农田采集的图片,覆盖了 13 种不同类型的植物及其对应的 27 种病害类型。PlantDoc 的特点在于其图像的多样性,包括变化的光照条件、复杂的背景以及多种类别的植物,这使得它成为一个特别适合用于多类别病害检测研究的数据集。图 5 展示的是 PlantDoc 数据集中的一些图像示例。



Tomato Early Blight leaf

Peach leaf

Blueberry leaf

图 5 PlantDoc 数据集样本图像

Fig. 5 Sample images from PlantDoc dataset

#### 2) PlantVillage 数据集

PlantVillage 是一个广泛应用的开源数据集,用于植物病害的自动化检测与分类任务。该数据集由 20 638 张图像组成,其中包括健康叶片与受病害影响的叶片图片,涉及的作物种类如胡椒、马铃薯和番茄等。PlantVillage

数据集以其高质量的图像而著称,常被用来训练和评估机器学习及深度学习模型在农业应用中的性能。在本研究中,本文使用了该数据集中胡椒、马铃薯和番茄叶片的健康与病害图像作为训练和测试数据。图 6 呈现了 PlantVillage 数据集中的一些代表性图像样本。



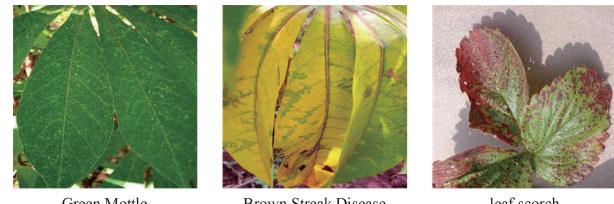
Pepper Bell Bacterial spot Tomato Yellow Leaf Curl Virus Potato Late blight

图 6 PlantVillage 数据集样本图像

Fig. 6 Sample images from PlantVillage dataset

#### 3) 自建数据集

自建数据集包含 5 077 张在不同光照条件下采集的作物病害图像,涵盖了木薯、草莓和葡萄多种作物的实际病害情况,充分反映了真实农业环境中的病害特征。数据集中共有 10 个类别,每个类别对应一种特定的植物病害或健康叶片状态。图 7 展示了自建数据集中的部分代表性图像样本。



Green Mottle Brown Streak Disease leaf scorch

图 7 自建数据集样本图像

Fig. 7 Sample images from the custom dataset

### 2.3 评价指标

在本次实验中,为了全面评估 AgriSwin 模型在植物病害检测任务中的表现,采用了以下关键评价指标:精确率 (Precision)、召回率 (Recall)、F1 分数 (F1 Score) 和 GMACs。每个指标的具体定义和公式如下:

1) 精确率 (Precision): 精确率用于衡量模型预测为正样本的实例中,实际为正样本的比例。它反映了模型预测的准确性。

$$Precision = \frac{TruePositives}{TruePositives + FalsePositives} \quad (19)$$

2) 召回率 (Recall): 召回率表示模型能从实际的正样本中正确识别的比例,反映了模型的敏感性。

$$Recall = \frac{TruePositives}{TruePositives + FalseNegatives} \quad (20)$$

3) F1 分数 (F1 Score): F1 分数是精确率和召回率的调和平均数,用于平衡模型的精确率和召回率,特别适用于正负样本不平衡的情况下。

$$F1 Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (21)$$

4) GMACs (giga multiply-accumulate operations): GMACs 是衡量模型计算复杂度的一个重要指标, 它表示模型在进行一次前向传播时所需的乘加运算次数, 以十亿次为单位。GMACs 越低, 模型的计算效率越高。

这些指标的综合分析可以帮助全面评估 AgriSwin 模型在植物病害检测任务中的性能。

## 2.4 模型性能与分析

本实验旨在验证 AgriSwin 在植物病害识别任务中的性能表现, 并评估其在实际应用中的有效性。实验结果展示了该模型在不同类别的识别精确率、召回率和 F1 分数方面的优越性, 具体结果(如表 1 所示)和混淆矩阵(如图 8 所示)。

表 1 PlantDoc 数据集分类性能对比

Table 1 Classification performance comparison on

PlantDoc dataset

Class	Precision/%	Recall/%	F1 Score/%
0	100	90.00	94.74
1	66.66	88.88	76.18
2	81.81	90.00	85.71
3	71.42	62.50	66.66
4	77.77	77.77	77.77
5	55.55	90.90	68.96
6	100	50.00	66.67
7	14.28	25.00	18.18
8	66.66	50.00	57.14
9	90.00	90.00	90.00
10	85.71	66.66	74.99
11	54.54	75.00	63.15
12	60.00	75.00	66.67
13	100	57.14	72.72
14	60.00	75.00	66.67
15	100	100	100
16	80.00	100	88.89
17	87.50	77.77	82.35
18	61.11	100	75.86
19	100	37.50	54.55
20	100	22.22	36.36
21	85.71	60.00	70.59
22	100	50.00	66.67
23	60.00	100	75.00
24	71.42	83.33	76.92
25	85.71	100	92.31
26	100	87.50	93.33
ALL	79.65	74.15	76.80

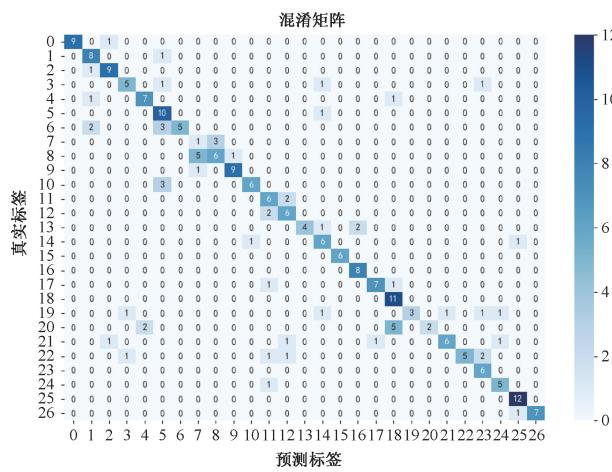


图 8 PlantDoc 数据集分类混淆矩阵

Fig. 8 Confusion matrix for PlantDoc dataset

实验结果显示, 本文提出的 AgriSwin 在 PlantDoc 数据集上的整体精确率达到了 79.65%, 召回率为 74.15%, F1 分数为 76.8%。这些结果表明, 该模型在病虫害检测任务中具有良好的整体性能, 尤其在复杂环境和多种病害类型的检测中表现出色。

类别 0 的检测精确率达到 100%, 召回率为 90%, F1 分数为 94.74, 显示出极高的识别能力。对于类别 2 和类别 9, 精确率分别为 81.81% 和 90%, F1 分数分别为 85.71% 和 90%, 说明模型在这些类别的检测中表现较为稳定。类别 5 的召回率较高, 达到 90.90%, 但精确率较低, 仅为 55.55%, 显示出较高的误报率。未来可以通过优化特征提取策略, 进一步提升模型在该类别的表现。类别 6 的精确率为 100%, 但召回率为 50%, 表明模型在检测某些类别时存在一定的难度。

如图 8 所示, 混淆矩阵展示了 AgriSwin 在 PlantDoc 数据集上的分类表现。从矩阵可以看出, 模型在多个类别表上的表现较为稳定, 尤其在类别 0、1 和 9 上, 显示出较高的识别准确率, 主要集中在对角线的区域。然而, 某些类别, 如类别 5 和类别 6, 在检测中表现出了一定的误差, 存在较多的误分类。类别 5 的误报率较高, 这在矩阵中表现为它的预测结果分散在多个其他类别, 说明该类别的特征在复杂背景下容易混淆。此外, 类别 6 的召回率较低, 这表明模型在检测此类时存在一定的难度, 可能需要进一步优化特征提取策略来提升识别能力。

根据图 9 的结果可以看出, 随着训练轮次的增加, 模型的验证损失逐渐下降, 表明模型持续学习并有效降低误差。训练初期, 损失下降较快, 说明模型迅速抓住了数据的关键特征。随着训练的推进, 验证损失逐渐趋于平稳。图 10 则展示了 AgriSwin 模型在训练过程中验证精确率的变化情况。可以看到, 验证精确率在初期波动较大, 但经过几轮训练后, 精确率逐步上升, 并趋于稳定, 最高达到约 79.65%。这一结果表明, 随着训练的深入, 模型的泛化能

力逐渐增强,能够更好地适应验证集数据,在植物病害识别任务中取得了较优的分类效果。

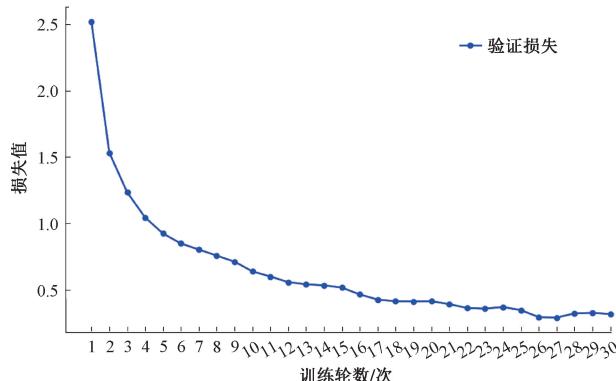


图 9 验证集损失变化曲线

Fig. 9 Validation set loss change curve

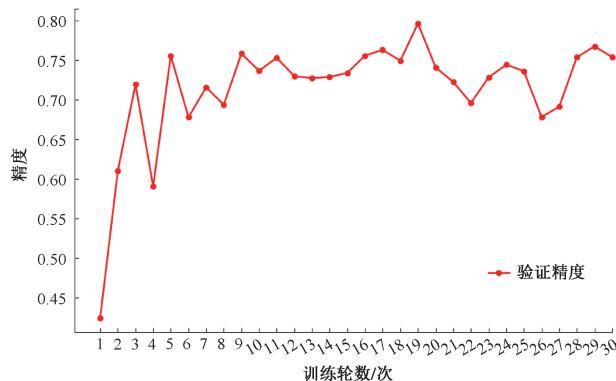


图 10 验证集精确率变化曲线

Fig. 10 Validation set loss change curve

## 2.5 对比实验

为了全面评估本文提出的 AgriSwin 模型在植物病害检测任务中的表现,本文对比了多个主流深度学习模型,包括 ResNet 系列<sup>[17]</sup>、DenseNet<sup>[18]</sup>、ShuffleNet<sup>[19]</sup>、MobileNet<sup>[20]</sup>、Swin Transformer、Vision Transformer、GoogleNet<sup>[21]</sup>、ConvNeXt<sup>[22]</sup>、EfficientNet<sup>[23]</sup>、MobViT<sup>[24]</sup>,以及其他植物病虫害检测文献 T-cnn(ResNet-101)<sup>[25]</sup>、ICVT 模型<sup>[26]</sup>做对比,表 2 中展示了各模型的 Precision、Recall、F1 分数以及模型参量。

实验结果显示,本文提出的 AgriSwin 模型在 3 个性能指标上均表现出明显优势, Precision 达到 79.65%, Recall 为 74.15%, F1 分数为 76.84%, 并且模型参量为 21.04 M, 实现了在性能和计算效率之间的良好平衡。相比之下,ResNet 系列(ResNet-34 和 ResNet-50)的 Precision 分别为 67.35% 和 71.08%, F1 分数分别为 66.33% 和 68.75%, 虽然模型参数较小,适合计算资源有限的场景,但其整体性能明显低于 AgriSwin。GoogleNet 的 F1 分数为 70.60%, 尽管模型参数量为 10.04 M, 计算效率较高,但性能未能超越 AgriSwin。ConvNeXt 的 Precision 为

表 2 模型性能对比结果

Table 2 Comparison of model performance results

模型	Precision/ %	Recall/ %	F1/ %	模型参量/M
ResNet-34	67.35	65.32	66.33	21.30
ResNet-50	71.08	66.52	68.75	25.60
GoogleNet	73.12	68.24	70.60	10.04
ConvNeXt	70.41	70.52	70.46	28.59
EfficientNet	73.25	69.81	71.49	5.30
DenseNet-161	74.51	71.63	73.03	29.00
ShuffleNetV2-X1.5	72.61	70.34	71.46	4.40
MobileNetV3-L	73.24	70.35	71.77	5.40
Swin Transformer-T	75.15	68.61	71.80	28.29
Swin Transformer-B	74.16	69.91	72.98	86.77
Vision Transformer-B	54.35	56.77	55.54	85.83
MobViT-M	74.36	69.92	72.09	23.00
T-cnn(ResNet-101)	74.44	—	—	—
ICVT	77.23	—	—	—
本文模型	79.65	74.15	76.80	21.04

70.41%, Recall 为 70.52%, F1 分数为 70.46%, 但模型参数量为 28.59 M, 整体表现仍逊于 AgriSwin。

EfficientNet 的 F1 分数为 71.49%, 模型参数量为 5.3 M, 性能虽好但不及 AgriSwin。DenseNet-161 的 F1 分数为 73.03%, 性能稳定, 但模型参数量为 29 M, 计算效率较低。ShuffleNetV2-X1.5 和 MobileNetV3-L 的模型参数量分别为 4.4 M 和 5.4 M, F1 分数为 71.46% 和 71.77%, 在计算效率上表现良好, 但性能仍不及 AgriSwin。MobViT-M 的 F1 分数为 72.09%, 模型参数量为 23 M, 性能接近但略低于 AgriSwin。

Swin Transformer 系列中, Swin Transformer-T 的 Precision 为 75.15%, Swin Transformer-B 的 F1 分数为 72.98%, 但模型参量明显高于 AgriSwin, 尤其是 Swin Transformer-B 的模型参量达到 86.77 M, 计算效率较低。Vision Transformer 基础模型的表现较为逊色, Precision 仅为 54.35%, F1 分数为 55.54%, 尽管模型参数较大(85.83 M), 但其性能远低于其他模型。

T-cnn(ResNet-101) 和 ICVT 模型的 Precision 分别为 74.44% 和 77.23%, 但两者未提供 Recall 和 F1 分数。虽然它们在某些方面表现较好, 但未能全面超越本文提出的模型。

实验结果表明, 本文提出的 AgriSwin 在 Precision、Recall 和 F1 分数等性能指标上均优于其他对比模型, 并且在模型参量方面表现出更高的计算效率, 特别适用于植物病害检测任务。

## 2.6 消融实验

为了评估 AgriSwin 模型中不同模块对整体分类性能

的影响,本研究设计了 7 组对比实验,通过逐步移除或替换模型的关键模块,来验证各个模块对模型表现的贡献。实验结果展示了不同模块的去除或替换对模型参数、计算复杂度以及 F1 分数的影响。

第 1 组实验采用 Swin Transformer-T 作为基线模型。该模型具备良好的全局特征提取能力,但由于其复杂的计算结构,参数量较高 (28.29 M),计算复杂度为 4.5 GMACs。实验结果显示(如表 3 所示),基线模型的 F1 分数为 72.76%。虽然基线模型能够捕捉全局特征,但在处理多尺度病害特征时表现欠佳,尤其是在面对复杂背景时,模型的表现存在一定的局限性。

表 3 消融实验结果对比

Table 3 Ablation experiment results comparison

对比实验	Precision/ %	Recall/ %	F1/ %	模型参量/M	GMACs
1	75.55	68.61	72.76	28.29	4.5
2	73.53	69.49	71.45	12.16	2.6
3	76.78	70.76	73.64	15.75	3.37
4	74.10	69.91	71.94	15.75	3.37
5	75.02	69.49	72.14	14.99	3.15
6	75.51	69.91	72.60	20.59	3.67
7	79.65	74.15	76.80	21.04	4.23

在第 2 组实验中,模型仅保留了 Swin Transformer-T 的前 3 个阶段。去除了第 4 阶段的模块导致模型的全局特征提取能力下降,尽管计算成本有所降低,但 F1 分数也随之下降。具体结果显示,参数量减少至 21.03 M,F1 分数为 69.45%。这一实验表明,第 4 阶段对捕捉复杂的多尺度信息起到了关键作用。

第 3 组实验将自适应空间卷积模块替换为标准的多头注意机制。该替换使模型在处理局部特征时缺乏灵活性,导致对细粒度特征的捕捉能力下降。实验结果显示,F1 分数为 71.32%,这表明自适应空间卷积模块在模型中对局部信息的精确捕捉具有重要贡献,而多头注意机制未能有效替代该模块的作用。

在第 4 组实验中,模型的自适应空间卷积模块被自注意机制替代。这一实验展示了自注意机制对局部特征提取的影响。尽管自注意机制擅长全局信息捕捉,但在本次实验中,模型的 F1 分数仅为 70.11%,且计算成本较高。这表明,自适应空间卷积模块在处理复杂背景下的细粒度信息提取上更具优势。

第 5 组实验中,移除了扩张特征聚合模块,仅保留自适应空间卷积模块。实验结果显示,模型的 F1 分数下降至 73.45%,表明扩张特征聚合模块在增强多尺度特征提取方面起到了关键作用,尤其是对大范围病害特征的捕捉。

第 6 组实验中,移除了自适应空间卷积模块,仅保留扩张特征聚合模块。模型的 F1 分数为 72.83%,与第 5 组实验相比有所下降。这表明,虽然扩张特征聚合模块在多尺度特征融合方面效果显著,但缺乏自适应空间卷积模块的局部信息增强机制,模型在细粒度特征提取方面有所不足。

最后,第 7 组实验为本文提出的 AgriSwin 模型。该模型同时结合了扩张特征聚合模块和自适应空间卷积模块,实现了对局部与全局特征的有效提取。实验结果表明,模型在多尺度病害检测任务中的表现最佳,F1 分数达到了 76.84%,同时参数量较为适中 (21.04 M),计算复杂度为 4.23 GMACs。这一结果验证了本文提出模块设计的有效性。

## 2.7 正则化方法对模型性能的影响

在深度学习模型中,正则化方法通过抑制过拟合,提升模型的泛化能力。为了研究正则化对 AgriSwin 模型在植物病害检测任务中的性能影响,本文对比了 L1 正则、L2 正则以及不使用正则化时的模型表现,具体实验结果如表 4 所示。

表 4 正则化方法对 AgriSwin 模型性能的影响

Table 4 Impact of regularization methods on AgriSwin model performance

正则化方法	Precision/ %	Recall/ %	F1/ %
L1	79.65	74.15	76.80
L2	78.03	72.44	75.10
None	76.65	72.34	74.43

使用 L1 正则化时,模型在 Precision、Recall 以及 F1 分数方面均表现最佳,Precision 达到了 79.65%,F1 分数为 76.77%。L1 正则通过在损失函数中引入权重绝对值的惩罚项,促使部分权重趋于零,从而实现了特征选择的效果。这有助于模型在病害检测任务中抑制不必要的复杂特征,提升模型的泛化能力,同时有效避免了过拟合。

L2 正则化的模型性能次于 L1 正则化。L2 正则化通过在损失函数中引入权重平方的惩罚项,使得权重整体更小,且趋于平滑分布。尽管 L2 正则在一定程度上抑制了模型的过拟合,但其作用相较于 L1 正则化略显不足,Precision 为 78.03%,F1 分数为 75.10%。这种情况表明,对于植物病害检测任务,L2 正则虽然降低了部分权重的数值,但未能实现更有效的特征选择,导致模型的综合性能略有下降。

在不使用正则化的情况下,模型的泛化能力较差,Precision 为 76.65%,F1 分数为 74.43%。没有正则化的模型容易出现过拟合现象,即在训练集上表现良好,但在测试集上性能有所下降。这是由于模型在训练过程中捕

捉到了大量的噪声或无效特征,导致其在实际检测任务中无法很好地处理复杂背景和多尺度病害特征。

从表 4 可以看出,正则化方法在提升模型泛化能力、抑制过拟合方面起到了重要作用。相比于不使用正则化的模型,L1 和 L2 正则化都显著提升了模型的 Precision、Recall 和 F1 分数。特别是 L1 正则化,通过特征选择减少了模型的复杂性,在多尺度病害检测任务中表现最为优异。因此,在植物病害检测这种需要处理复杂特征的任务中,使用 L1 正则化可以有效提升模型的综合性能,同时确保计算开销适中。

## 2.8 多数据集的系统性能验证

为了验证 AgriSwin 模型的泛化能力和稳定性,本文分别使用了 PlantVillage 数据集和自建数据集进行了实验,并将结果与其他主流模型进行了对比分析。实验不仅验证了 AgriSwin 模型在检测精度和效率方面的表现,还通过多个数据集的验证展示了模型的实际应用价值。

从表 5 可以看出,AgriSwin 模型在 PlantVillage 数据集上取得了 99.90% 的精确率、99.90% 的召回率和 99.90% 的 F1 分数,其检测能力显著优于其他模型,这表明 AgriSwin 模型不仅具有强大的识别能力,还能够在多个维度上稳定保持高效性能。

表 5 PlantVillage 数据集实验结果

Table 5 Generalization experiment results for AgriSwin model

模型	Precision/ %	Recall/ %	F1-Score/ %
Machine learning model using LGBM model <sup>[27]</sup>	94.75	94.76	94.72
CNN <sup>[28]</sup>	96.46	96.25	96.25
Attention and the residual network <sup>[29]</sup>	96.20	95.60	95.89
Pre-trained MobileNet-V2 and attention mechanism <sup>[30]</sup>	97.49	95.83	96.64
本文模型	99.90	99.90	99.90

在自建数据集上,AgriSwin 模型同样表现出强大的稳定性和泛化能力,如表 6 所示,其精确率达到了 95.08%,召回率为 94.95%,F1 分数为 95.01%。同时,与其他模型相比,AgriSwin 的参数量明显较低,仅为 21.04 M,远低于 Vision Transformer 的 85.83 M。这表明 AgriSwin 在计算资源的使用上更具优势。

通过以上结果可以看出,AgriSwin 模型在精确率、召回率、F1 分数等主要指标上均优于其他模型,尤其是在自建数据集上的表现更加突出。同时,AgriSwin 模型的参数量较小,计算资源占用低,展示了该模型在实际应用中的巨大潜力。与其他传统模型相比,AgriSwin 在保持高精度

表 6 自建数据集实验结果

Table 6 Generalization experiment results for AgriSwin model on custom Dataset

模型	Precision/ %	Recall/ %	F1-Score/ %	模型参量/M
Swin Transformer-T	93.36	92.61	92.98	28.29
Vision Transformer-B	90.91	89.34	90.12	85.83
本文模型	95.08	94.95	95.01	21.04

的同时具备低计算资源消耗,特别适用于检测病虫害的农业场景。

## 3 结 论

本文提出了 AgriSwin 模型,以应对植物病虫害检测中的多尺度特征提取和复杂背景处理难题。通过融合扩张特征聚合和自适应空间卷积模块,成功地将卷积神经网络(CNN)在局部特征捕捉方面的强大功能与 Transformer 架构在全局信息处理上的优势结合起来,在多尺度病害识别任务上实现了高效且准确的表现。测试结果显示,AgriSwin 在 PlantDoc、PlantVillage 和自建数据集上的表现超越了现有的主要模型,证明了它在复杂环境和多样化目标检测中的出色性能。

随着农业应用场景的增加以及病虫害种类与环境条件的变化,对于模型的泛化能力提出了更高的需求。因此,AgriSwin 有潜力应用于更为广泛的植物病虫害识别任务中。此外,探索该模型与其他先进深度学习技术的整合,或是进一步优化其组件以增强其实时响应能力和适应性,均为未来的改进方向。AgriSwin 的开发不仅为植物病虫害检测提供了一种创新的方法,也为推动农业自动化监测技术的发展开辟了新的研究领域。

## 参考文献

- [1] 金婷婷,房建东,赵于东.注意力特征融合的番茄叶部早期病斑诊断算法[J].电子测量技术,2024,47(4): 156-164.  
JIN T T, FANG J D, ZHAO Y D. Tomato leaf early disease spot diagnosis algorithm based on attention feature fusion [J]. Electronic Measurement Technology, 2024, 47(4): 156-164.
- [2] ISLAM M T. Plant disease detection using CNN model and image processing[J]. International Journal of Engineering Research & Technology, 2020, 9(5): 212-220.
- [3] BEDI P, GOLE P. Plant disease detection using hybrid model based on convolutional autoencoder and convolutional neural network[J]. Artificial Intelligence in Agriculture, 2021, 5: 140-149.
- [4] SUN X, LI G, QU P, et al. Research on plant disease

- identification based on CNN[J]. *Cognitive Robotics*, 2022, 5: 100015.
- [5] LU J, TAN L, JIANG H. Review on convolutional neural network (CNN) applied to plant leaf disease classification[J]. *Agriculture*, 2021, 11(8): 707.
- [6] THAKUR P S, KHANNA P, SHEOREY T, et al. Vision transformer for plant disease detection: PlantViT[C]. 2021 Springer International Conference on Computer Vision. Springer, 2021: 331-340.
- [7] PAREZ S, DILSHAD N, ALGHAMDI N S, et al. Visual intelligence in precision agriculture: Exploring plant disease detection via efficient vision transformers[J]. *Sensors*, 2023, 23(15): 6949.
- [8] YU S, XIE L, HUANG Q. Inception convolutional vision transformers for plant disease identification[J]. *Internet of Things*, 2023, 23: 100124.
- [9] YANG H, DENG X, SHEN H, et al. Disease detection and identification of rice leaf based on improved detection transformer [J]. *Agriculture*, 2023, 13(7): 1361.
- [10] 卢浩龙, 朱彦敏. 基于多尺度图 Transformer 的滚动轴承故障诊断 [J]. 国外电子测量技术, 2023, 42(12): 186-194.
- LU H L, ZHU Y M. Fault diagnosis of rolling bearing based on multi-scale graph transformer[J]. *Foreign Electronic Measurement Technology*, 2023, 42(12): 186-194.
- [11] 吴兴辉, 何赟泽, 周辉, 等. 改进 YOLO v7 算法下的监控水域环境人员识别研究[J]. *电子测量与仪器学报*, 2023, 37(5): 20-27.
- WU X H, HE Y Z, ZHOU H, et al. Research on personnel identification in water environment monitoring based on improved YOLO v7 algorithm[J]. *Journal of Electronic Measurement and Instrumentation*, 2023, 37(5): 20-27.
- [12] 王子辰, 陈晓艳, 王倩, 等. 基于残差自注意力连接的深度电学层析成像方法[J]. *仪器仪表学报*, 2023, 44(5): 288-301.
- WANG Z CH, CHEN X Y, WANG Q, et al. Deep electrical tomography method based on residual self-attention connection[J]. *Chinese Journal of Scientific Instrument*, 2023, 44(5): 288-301.
- [13] 贾晓芬, 吴雪茹, 赵佰亭. 绝缘子自爆缺陷的轻量化检测网络 DE-YOLO[J]. *电子测量与仪器学报*, 2023, 37(5): 28-35.
- JIA X F, WU X R, ZHAO B T. Lightweight detection network DE-YOLO for insulator explosion defects[J]. *Journal of Electronic Measurement and Instrumentation*, 2023, 37(5): 28-35.
- [14] 马敏, 梁雅蓉. 基于多尺度自适应特征聚合网络的 ECT 图像重建[J]. *仪器仪表学报*, 2023, 44(6): 264-272.
- MA M, LIANG Y R. ECT image reconstruction based on multi-scale adaptive feature aggregation network[J]. *Chinese Journal of Scientific Instrument*, 2023, 44(6): 264-272.
- [15] SINGH D, JAIN N, JAIN P, et al. PlantDoc: A dataset for visual plant disease detection[C]. 7th ACM IKDD CoDS and 25th COMAD, 2020: 249-253.
- [16] KAGGLE. The dataset is taken from the Kaggle Opensource link[EB/OL]. [2024-08]. <https://www.kaggle.com/datasets/emmarex/plantdisease>.
- [17] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[C]. IEEE Conference on Computer Vision and Pattern Recognition(CVPR), 2016: 770-778.
- [18] HUANG G, LIU Z, LAURENS V D M, et al. Densely connected convolutional networks[C]. IEEE Conference on Computer Vision and Pattern Recognition(CVPR), 2017: 4700-4708.
- [19] ZHANG X, ZHOU X, LIN M, et al. ShuffleNet: An extremely efficient convolutional neural network for mobile devices [C]. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018: 6848-6856.
- [20] HOWARD A G, ZHU M, CHEN B, et al. MobileNets: Efficient convolutional neural networks for mobile vision applications [J]. ArXiv preprint arXiv:1704.04861, 2017.
- [21] SZEGEDY C, LIU W, JIA Y, et al. Going deeper with convolutions[C]. IEEE Conference on Computer Vision and Pattern Recognition(CVPR), 2015: 1-9.
- [22] LIU Z, LIN Y, CAO Y, et al. A ConvNet for the 2020s [C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition(CVPR), 2022: 11976-11986.
- [23] TAN M, LE Q V. EfficientNet: Rethinking model scaling for convolutional neural networks[C]. 36th International Conference on Machine Learning (ICML), 2019: 6105-6114.
- [24] MEHTA S, RASTEGARI M. MobileViT: Light-weight, general-purpose, and mobile-friendly vision transformer [J]. ArXiv preprint arXiv: 2110.02178, 2021.
- [25] WANG J, WANG W, LI P, et al. T-CNN: Trilinear convolutional neural networks model for visual

- detection of plant diseases [J]. Computers and Electronics in Agriculture, 2021, 190: 106468.
- [26] TABBAKH A, BARPANDA S S. A deep features extraction model based on the transfer learning model and Vision Transformer “TLMViT” for plant disease classification[J]. IEEE Access, 2023, 11: 45377-45392.
- [27] TABBAKH A, BARPANDA S S. Evaluation of machine learning models for plant disease classification using modified GLCM and wavelet-based statistical features[J]. Traitement Du Signal, 2022, 39 (6): 1893-1905.
- [28] MOHANA R, REDDY C K K, ANISHA P R. A study and early identification of leaf diseases in plants using convolutional neural network [C]. 4th International Conference on Smart Computing and Informatics, Cham: Springer, 2021: 693-709.
- [29] KARTHIK R. Attention embedded residual CNN for disease detection in tomato leaves[J]. Applied Soft Computing, 2020, 86: 105933.
- [30] CHEN J, ZHANG D, ZEB A, et al. Identification of rice plant diseases using lightweight attention networks [J]. Expert Systems with Applications, 2021, 169: 114514.

### 作者简介

刘微(通信作者),教授,硕导,博士研究生,主要研究方向为智能算法。

E-mail:liuwei19781020@126.com

张傲,硕士研究生,主要研究方向为数据安全与数据处理。

E-mail:1797718602@qq.com