doi:10.6041/j.issn.1000-1298.2026.02.018

基于 DenseNet 优化 Swin Transformer 模型的苹果叶部病害多尺度特征分类

谷　伟[1,2]　叶　峥[1]　矫桂娥[2]

（1. 上海海洋大学信息学院，上海 201306；2. 上海建桥学院信息技术学院，上海 201306）

摘要：针对人工检测苹果病害效率低、成本高且准确性差的问题，本文以 Swin Transformer 作为基础模型，在核心模块中引入 DenseNet 思想，增强特征传递和梯度流动；使用 Outlook Attention 捕捉图像中细节特征，提升模型细粒度信息提取能力。为了进一步优化模型性能，引入了深度可分离卷积和膨胀卷积，实现在较小参数量前提下捕捉不同尺度的特征；在模型中引入 Non－Local，以整合全局上下文信息，进一步提高模型的综合性能。以上改进共同作用，使得本文模型在多个任务上表现出了优异的性能和鲁棒性。实验结果显示，苹果叶部病害分类识别准确率达到 95.8%，精确率、召回率和 F1 分数分别达到 95.80%、95.74%、95.76%，均超过基线模型。基于改进 Swin Transformer 的苹果叶部病害分类模型能够有效实现苹果叶部病害的种类识别及其严重程度分类，为大规模作物病害监测提供了理论支持和研究基础，助力精准防控与绿色农业。

关键词：苹果叶部病害；深度可分离卷积；细粒度信息；全局上下文信息；特征传递

中图分类号：S432.1　　**文献标识码**：A　　**文章编号**：1000-1298(2026)02-0181-12　　**OSID**：

Swin Transformer Model Optimized with DenseNet for Multi-scale Feature Classification of Apple Leaf Diseases

GU Wei[1,2]　YE Zheng[1]　JIAO Guie[2]

(1. *School of Information*, *Shanghai Ocean University*, *Shanghai* 201306, *China*

2. *School of Information Technology*, *Shanghai Jianqiao University*, *Shanghai* 201306, *China*)

Abstract: Given the limitations of manual apple disease detection, including inefficiency, high cost, and low accuracy, aiming to propose a more effective solution that can improve detection accuracy while reducing time and costs. The Swin Transformer was utilized as the base model and the DenseNet framework was integrated into its core modules to enhance feature propagation and improve gradient flow. Additionally, the Outlook Attention module captured fine-grained image details, enhancing the model's ability to extract intricate features. To further optimize the model's performance, the depthwise separable and dilated convolutions were introduced, enabling the capture of multi-scale features while reducing the parameter size. Finally, the Non－Local module was integrated into the model to incorporate global context information, thereby further enhancing overall performance. These improvements collectively enabled the model to exhibit superior performance and robustness across multiple tasks. Experimental results indicated that the accuracy for classifying apple leaf diseases reached 95.8%, with precision, recall, and F1 score values of 95.80%, 95.74%, and 95.76%, respectively, all surpassing those of the baseline model. The proposed Swin Transformer-based model, optimized for apple leaf disease classification, efficiently identified both the type and severity of apple leaf diseases. This served as a theoretical foundation and provided critical support for large-scale crop disease monitoring, facilitating precise disease prevention and control in sustainable agriculture. Moreover, compared with existing deep learning models like ResNet and standard Swin Transformer, the proposed model exhibited superior accuracy and computational efficiency. Future research would focus on further optimizing the model architecture to address more complex agricultural scenarios, such as classifying co-occurring diseases,

收稿日期：2024－10－18　修回日期：2024－12－02

基金项目：国家自然科学基金项目(42376194)和上海市晨光计划项目(AASH2004)

作者简介：谷伟(1978—)，男，教授，博士，主要从事计算机应用技术、数据处理与分析和智能教育应用研究，E-mail：2958671279@qq.com

通信作者：矫桂娥(1976—)，女，教授，博士，主要从事数据挖掘及可视化、数字媒体技术及应用研究，E-mail：jiaoguie@gench.edu.cn

and integrating drone-based image acquisition technologies for real-time disease detection and prediction.

Key words: apple leaf diseases; depthwise separable convolution; fine-grained information; global context information; feature propagation

0 引言

苹果是我国主要的经济作物之一,其产量与品质易受病虫害影响。但传统的人工病害检测方法存在效率低、成本高和误用农药的风险,基于图像处理和机器视觉技术的苹果叶部病害识别研究为病虫害防控提供了更高效的解决方案[1]。随着卷积神经网络(Convolutional neural networks, CNN)和 Transformer[2]模型的广泛应用,植物病害的自动化识别得到了显著进展。传统机器学习方法[3]如支持向量机(Support vector machine,SVM)[4]和主成分分析[5]等,尽管在特定场景下有效,但依赖于手工特征提取,难以应对复杂多变的病害形态[6]。传统机器学习依赖手工特征,难以应对复杂病害,分类精度和效率受限。

深度学习和卷积神经网络的应用推动了叶片病害检测与识别研究的发展。,如通过引入通道和空间注意力模块[7]、改进 Inception 结构[8]以及采用多扩张率并行卷积[9]。增强病斑关注与多尺度特征提取。尽管 CNN 在简单背景下表现良好,但在复杂背景或光照变化时易混淆相似病害。为此,CHEN 等[10]引入多尺度与通道增强,TANG 等[11]采用动态协调结构,YU 等[12]结合子卷积与双线性 CNN,以提升细粒度特征提取与鲁棒性。

近年来,融合 CNN 与 Transformer 优势的混合架构成为研究热点。相关工作通过在 Transformer 中引入卷积模块或特征交互机制,有效兼顾局部细节与全局语义信息,从而在病害识别任务中实现更高的精度与泛化能力[13-18]。

ViT[19]在视觉任务中表现突出,但因难以捕捉局部结构,在中小数据集上性能不及 CNN。T2T[20]通过重叠切分与迭代聚合增强局部特征,虽提升性能但仍逊于先进的 CNN。深度可分离卷积与空洞卷积被广泛用于提升模型效率与特征提取能力[21-24]。研究者通过结合残差结构、金字塔架构及并行空洞卷积,实现通道与空间解耦,在降低参数与计算量的同时增强多尺度特征提取和分类精度。非局部注意力模块(Non-Local)通过建模全局关系增强特征表示[25-27],可自适应捕捉长距离依赖,提升复杂场景下细粒度特征和微小病变的识别能力,同时抑制无关信息。

本研究将 Swin Transformer[28]作为主干网络,并结合 DenseNet[29]的特征复用机制,以增强梯度流动和特征传递。为进一步提升模型性能,集成深度可分离卷积和膨胀卷积,以捕捉不同尺度病害特征,并在此基础上引入 Outlook Attention[30],用于细粒度信息提取。最终,通过 Non-Local 整合全局上下文信息,以提高模型综合分类性能。为进一步增强 Transformer 在小数据集上的表现,采用数据增强(如翻转、旋转和颜色抖动)和卷积模块扩展样本多样性与提取局部特征,提升模型泛化能力及细节识别精度。

1 材料与方法

1.1 实验数据

1.1.1 数据获取

所使用数据集为 AppleLeaf 9,其为一个专门用于苹果叶片病害分类的综合数据集,由多个来源的数据集整合而成,其来源包括 PlantVillage 数据集(PVD)、苹果树叶片病害分割数据集(ATLDSD)、以及 PPCD2020 和 PPCD2021 数据集。图 1 展示了数据集中部分图像。

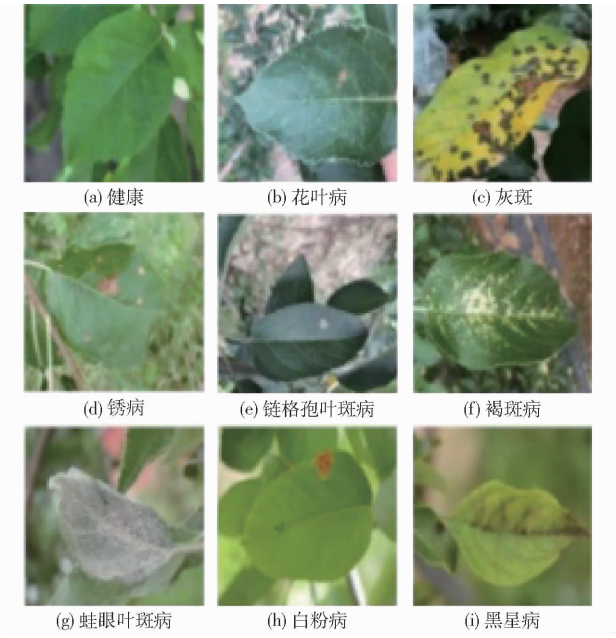


图 1 数据集部分图像示例

Fig. 1 Example images from dataset

1.1.2 数据预处理

通过分析数据集的分布和数量,发现数据集存在严重的不平衡问题,且某些类别的图像数量明显过多,这导致模型在训练过程对样本数量多的类别进行过度训练,而忽视样本数量少的类别,从而影响

了模型整体识别效果。为此，采用 K-means 聚类算法[31]对样本数量较多的类别进行下采样，通过将相似样本聚集在一起，然后从每个聚类中选择代表性样本进行下采样。这种方法可减少数据集的整体大小，提高训练速度，节约了时间成本，同时保留关键样本信息。

1.1.3　数据增强

运用多种数据增强技术，模拟采集图像时受角度、遮挡、天气等外界因素干扰的情况，以防止模型过拟合[32]，并提升模型的鲁棒性与泛化能力[33]。图像相关变换方法及具体的数据增强途径为：①应用饱和度、对比度、翻转、亮度、锐化和旋转等图像变换方法，针对样本数量少的类别进行随机变换，增加样本多样性。②部分小类别数据集中存在多幅叶片的图像，如图 2 所示。鉴于针对其数量较少，采用单独离线处理，先将图像均等剪切为左上、左下、右上、右下 4 个部分，再对剩余单一叶片图像随机裁剪，以此增加小类别样本数量。③小类别数据中部分图像背景复杂，增加了识别难度，而大部分目标图像集中在中心区域，故对图像的左上、左下、右上、右下 4 个区域随机放大 2 倍，以减少复杂背景及非目标区域干扰。④采用自适应直方图均衡化方法进行数据增强，随机选取不同的 ClipLimit 值（1、2、3、4）重新分配灰度，使图像灰度分布更均匀，增强细节与对比度效果，生成不同对比度水平的样本，提升模型在复杂背景下分类稳健性与泛化能力。



图 2　多叶样例

Fig. 2　Multileaf sample

经数据预处理和多种数据增强处理后，最终得到增强图像效果如图 3 所示，最终得到 7 989 幅图像，并按比例 14∶3∶3将数据集划分为训练集、验证集和测试集。增强后图像分布如表 1 所示。

1.2　苹果叶部病害分类模型

改进 Swin Transformer 模型主要包含 4 个模块：① Swin Transformer 作为骨干网络模块，用于特征提

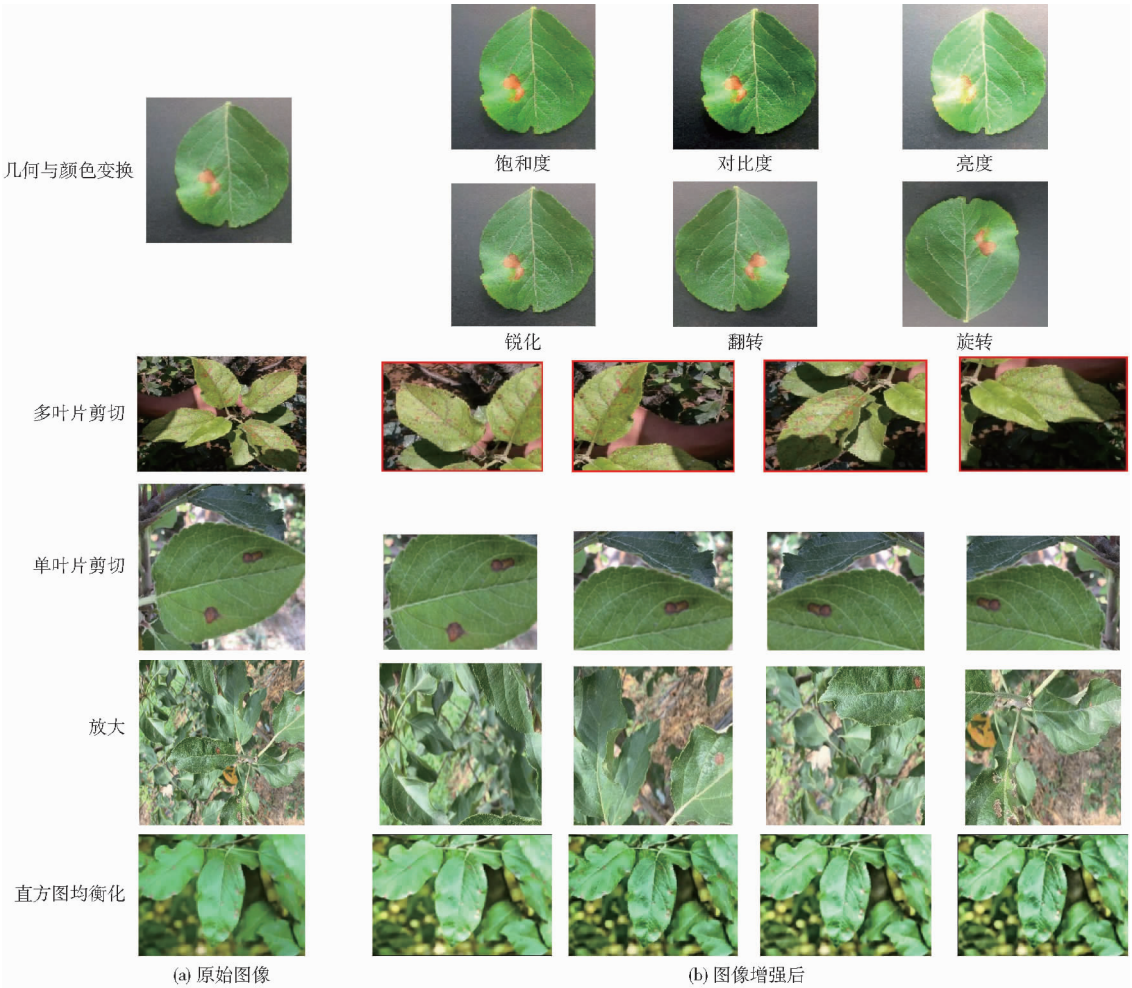


图 3　图像增强效果示例

Fig. 3　Image enhancement effect example

表 1 增强前后图像数量对比

Tab. 1 Comparison of number of images before and after enhancement

类别	初始数量	增强后数量
健康	516	800
花叶病	317	749
灰斑	339	687
锈病	2 753	1 000
链格孢叶斑病	417	842
褐斑病	411	821
蛙眼叶斑病	3 181	960
白粉病	1 184	1 184
黑星病	5 410	1 000

取和信息传递。在其核心模块中引入 DenseNet 思想，以减少参数量并实现特征复用。② Outlook Attention 用于捕捉图像的细节特征，从而增强模型提取细粒度信息的能力。③ 深度可分离卷积和膨胀卷积可以在保持较低计算复杂度的同时，有效捕捉不同尺度特征。④ 非局部模块用于整合全局上下文信息，从而提升模型整体表现。改进模型整体架构如图 4 所示。

1.2.1 Swin Transformer 网络模型

为了确保苹果病害分类任务的高准确性和实时性，选择 Swin Transformer 作为骨干网络，该网络通过引入滑动窗口机制和层次化结构，能够有效捕捉图像的多尺度特征。相比于传统的卷积神经网络，Swin Transformer 在保证病害特征提取能力的同时，具有更为出色的多尺度特征表达能力和更高的计算效率。

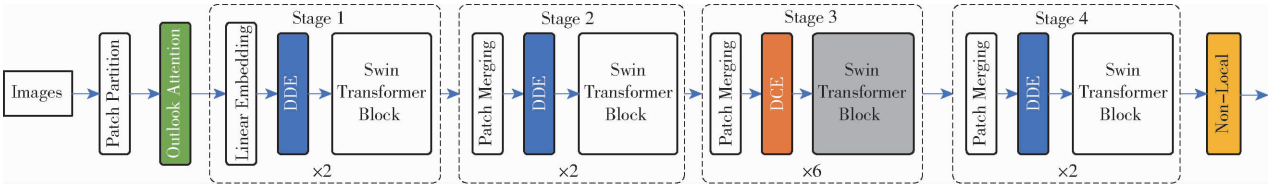


图 4 改进模型整体架构图

Fig. 4 Improved model overall architecture diagram

分层的 Swin Transformer，采用了一种名为移动窗口机制，使其效率提高，并显著降低序列长度。模型整体架构如图 5 所示。其中滑动移动操作，使得相邻窗口间有了交互。这种层级结构允许相邻窗口间交互，不仅灵活，能提供多尺度的特征信息，而且由于自注意力仅在小窗口内计算，计算复杂度随图像尺寸线性增长，而非平方增长。因此，它还适用于超高分辨率图像预训练。

ViT 将图像分割成 16 × 16 的补丁，相当于 16 倍下采样，整个过程中图像尺寸保持不变。而 Swin Transformer 中下采样从 4 倍逐步增加到 16 倍，从而能够捕捉到多尺度特征。Swin Transformer 核心在于其对移动窗口的巧妙使用，Transformer 块是先使用基于窗口的多头注意力，后面紧跟基于移动窗口的多头注意力，这样不仅能关注到全局的特征，而且可以实现不同窗口之间的信息交互。通过窗口偏移操作使 4 个窗口变为 9 个窗口，此前蓝色区域的图像处于分离状态，无法实现信息交互，而后续基于移动窗口的使用使得蓝色区域转换为红色区域，绿色区域变成灰色区域，从而实现不同窗口之间的信息交互。窗口变换过程如图 6 所示。

Swin Transformer 在苹果病害分类任务中展现出相较于其他常见骨干网络（如 ResNet[34] 和 DenseNet）的性能优势。该模型不仅在分类准确率和召回率方面显著提升，同时在处理复杂背景和不同光照条件下图像时表现出更强的泛化能力。因此，选择 Swin Transformer 作为骨干网络能够在保证病害特征提取能力的同时，实现高效的训练和推理，满足实际应用需求。

1.2.2 结合 DenseNet 思想的 Swin Transformer 核心模块

以 Swin Transformer 模型为基础，其核心架构包

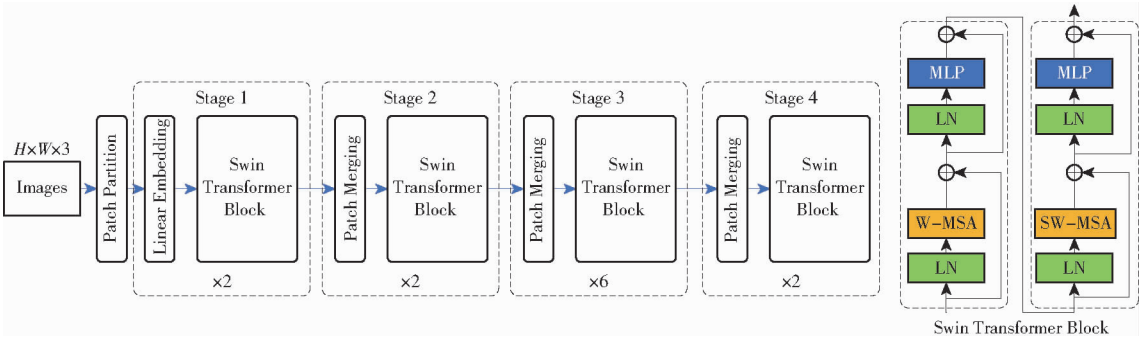


图 5 Swin Transformer 模型架构

Fig. 5 Swin Transformer model architecture

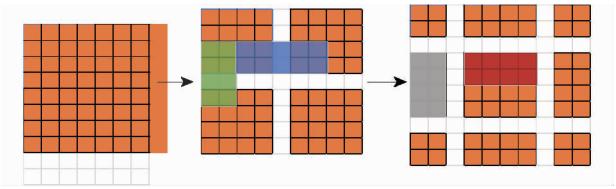


图 6　窗口变换

Fig. 6　Window transform

含 4 个阶段，每个阶段的块由具有固定窗口注意力机制模块和具有移动窗口注意力机制模块组成，且块数为 2、2、6、2。本研究在第 3 阶段做出了模型的改进，其他 3 个阶段保持原模型不变，图 7 为 Swin Transformer 核心模型的第 3 阶段。该阶段分为 3 个相同的部分进行串联，每个部分先经过固定窗口注意力机制（Window based self-attention，W－MSA），再经过可移动窗口注意力机制（Shifted window based self-attention，SW－MSA），这样处理不仅提取了图像的局部信息和全局信息，而且显著降低了计算复杂度。本研究将 DenseNet 的特征重用思想引入 Swin Transformer 模型，以提升其在苹果叶部病害分类任务中的性能。在 Swin Transformer 的各个阶段中增加了跨层连接，使得后期层能够直接利用早期层的特征。这不仅确保了特征的高效重用，还避免了信息丢失。在原模型的基础上，将 6 个模块分为奇偶组合，1、3、5 为一组，2、4、6 为一组。通过图 8 可以看到，在每个阶段的块中，固定窗口和移动窗口的输入不仅来源于上一层输出，还包括进入块之前的所有输入。这一设计实现了特征的有效重用，并通过跨层连接增强了信息流动。即

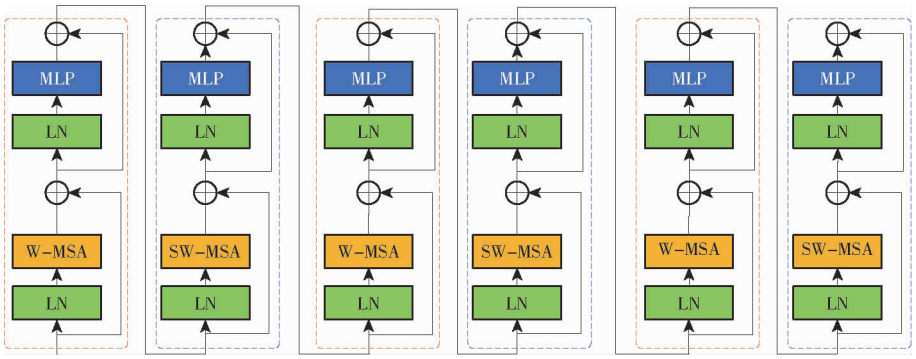


图 7　Swin Transformer 核心模型

Fig. 7　Swin Transformer core model

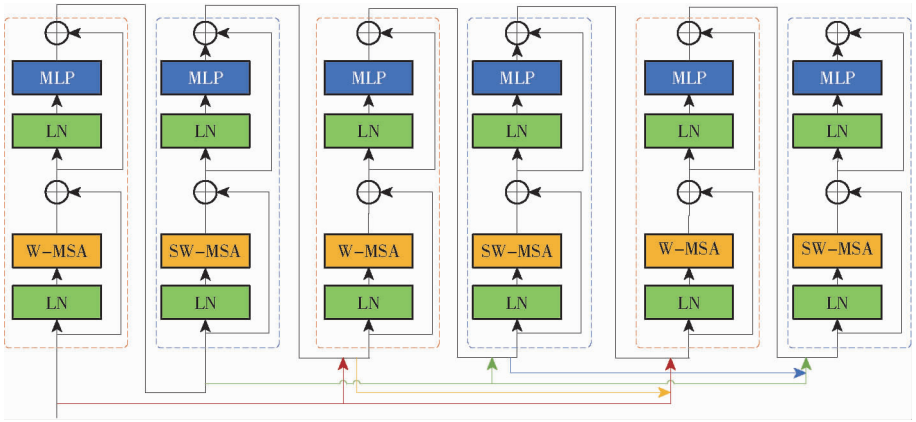


图 8　改进 Swin Transformer 核心模型

Fig. 8　Improved Swin Transformer core model

$$\boldsymbol{X}_i=\begin{cases}\boldsymbol{X}_1 & (i=1)\\ \boldsymbol{Y}_1 & (i=2)\\ \sum_{j=1}^{k}\boldsymbol{X}_{2j-1}+\boldsymbol{Y}_{2k} & (i=2k+1)\\ \sum_{j=2}^{k}\boldsymbol{X}_{2j-2}+\boldsymbol{Y}_{2k-1} & (i=2k)\end{cases}\tag{1}$$

式中　$\boldsymbol{X}_i$、$\boldsymbol{Y}_i$——第 i 层输入、输出

i——模型当前层数

k、j——临时变量

当输入为奇数层时，当前输入来自上层的输出和前面奇数层输入之和；当输入为偶数层时，当前输入来自上层的输出和前面偶数层输入之和。对于模型输出 $\boldsymbol{Y}$ 计算式有

$$\boldsymbol{Y}_m=W-MSA(LN(\boldsymbol{X}_i))+\boldsymbol{X}_i\quad(i\text{ 为奇数})\tag{2}$$

$$\boldsymbol{Y}_n=SW-MSA(LN(\boldsymbol{X}_i))+\boldsymbol{X}_i\quad(i\text{ 为偶数})\tag{3}$$

$$\boldsymbol{Y}_t=MLP(LN(\boldsymbol{Y}_k))+\boldsymbol{Y}_k\quad(k\text{ 为 }m\text{ 或 }n)\tag{4}$$

式中　$\boldsymbol{Y}_m$、$\boldsymbol{Y}_n$——奇数层、偶数层中间输出结果

W——窗口多头注意力

SW——移位窗口多头注意力

m、n、t——临时变量

通过这种方式能够在不同层次上同时提取和重用局部与全局特征，显著提升了模型特征表示能力。在改进中，特征重用机制不仅通过常规的前向传递，还通过跨层连接进行特征有效传播。这种机制不仅丰富了每一层特征输入，还显著提高了模型在处理复杂图像任务时的整体性能。在保留 Swin Transformer 核心架构（包括 W－MSA、SW－MSA）的基础上，通过引入跨层连接与特征重用机制，成功将 DenseNet 的特征重用思想与 Swin Transformer 模型有机融合。

在苹果叶部病害分类任务中，这些改进使模型能够更精确地捕捉叶片的细微特征和病害模式，从而显著提升分类准确率。特征重用和跨层连接确保了每个阶段都能利用丰富的上下文信息，从而更准确地识别病害特征。

1.2.3 Outlook 注意力机制

为了准确分类苹果叶部病害，某些病害图像特征并不明显，而提取细粒度特征相对较为困难。为此，对 Swin Transformer 模型进行重要改进，即在其块前嵌入 Outlook 注意力机制模块。Outlook 注意力机制通过在每个空间位置(i,j)上计算以该位置为中心的 $K\times K$ 局部窗口内所有邻近点的相似度来实现。给定一个输入 $\boldsymbol{X}$，使用 2 个线性权重 $\boldsymbol{W}_A\in\boldsymbol{R}^{C\times K^4}$ 和 $\boldsymbol{W}_V\in\boldsymbol{R}^{C\times C}$，分别投影到权重 $\boldsymbol{A}\in\boldsymbol{R}^{H\times W\times K^4}$ 和 $\boldsymbol{V}_{(i,j)}\in\boldsymbol{R}^{H\times W\times C}$（其中 $\boldsymbol{R}$ 为实数矩阵，C、H、W 为特征图特征通道数、高度、宽度），即

$$\boldsymbol{V}_{i,j}=\left\{\boldsymbol{V}_{i+p-\frac{K}{2},j+q-\frac{K}{2}}\right\}\quad(0\leqslant p,q<K)\tag{5}$$

$$\boldsymbol{Y}_{i,j}=matmul(\mathrm{softmax}(\boldsymbol{A}_{i,j}),\boldsymbol{V}_{i,j})\tag{6}$$

式中 $\boldsymbol{V}_{i,j}$——局部窗口(i,j)中所有值

$\boldsymbol{A}_{i,j}$——(i,j)位置权重

$\boldsymbol{Y}_{i,j}$——(i,j)位置处输出结果

K——窗口大小

p、q——高度、宽度偏移量

i、j——高度、宽度维度索引

不同于自我注意力机制所需的复杂 Query－Key 矩阵乘法，Outlook 注意力机制通过简单的重塑操作（Reshape）简化了复杂的 Query－Key 矩阵乘法。具体来说，输入 $\boldsymbol{X}$ 经过 2 个线性层得到 Outlook 权重 $\boldsymbol{A}$ 和数值 $\boldsymbol{V}$，模型结构如图 9 所示。通过局部窗口内的相似度计算直接获得注意力权值，简化了计算流程，同时保留了局部特征提取的有效性。此改进在细粒度特征提取和计算效率上具有显著优势。首先，Outlook 注意力机制显著提升了细粒度局部特征的提取能力。尽管传统的 Swin Transformer 在整合全局和局部特征方面表现良好，但在细粒度特征提取上仍有不足。通过在 Patch Partition 模块后加入 Outlook Block，模型能够更好地捕捉图像中微小差异和细节，尤其需要精细辨别的任务，如病害分类中表现尤为突出。Outlook 注意力机制在局部特征提取方面具有更高的敏感度，能够有效提升模型细粒度特征识别能力，从而提高分类准确性。

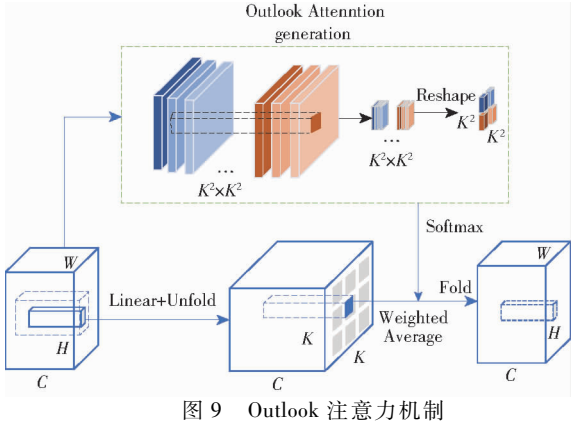


图 9 Outlook 注意力机制

Fig. 9 Attention mechanism

相比于传统的全局注意力机制，Outlook 注意力机制计算效率更高。全局注意力机制虽能捕获图像的全局信息，但计算量大，可能导致资源浪费和处理速度慢。Outlook 注意力机制通过局部注意力提取特征，可显著降低计算复杂度。本研究将 Outlook 注意力机制与 Swin Transformer 的层次结构结合，既保持高效计算，又利用图像的多层次信息，优化计算效率和模型性能。

图 10 为基准模型在添加 Outlook 注意力机制前后的热力图效果。由图 10 可知，新增模块提升了图像关键区域的识别精度，热度分布更加集中于重要特征，形成明显焦点，热度差异更加显著，有助于清晰地识别图像中最关键的部分。

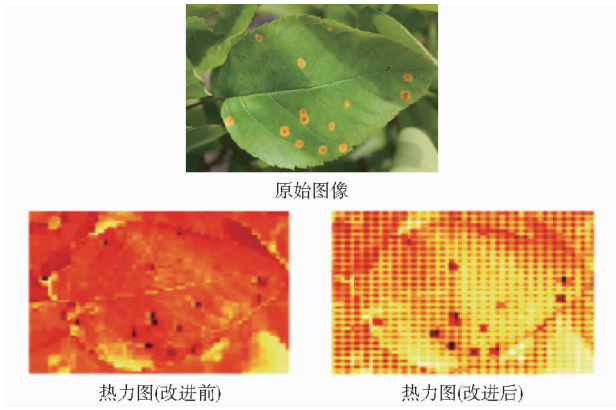


图 10 改进前后热力图效果

Fig. 10 Heat map effects before and after improvement

综上所述，通过在 Swin Transformer 模型中引入 Outlook 注意力机制，显著提升了细粒度局部特征的提取能力和计算效率。这种改进不仅提升了苹果叶部病害分类的准确性和鲁棒性，也使得模型在处理大规模数据时更加高效，为实际应用提供了更强有力的技术支持。

1.2.4　深度可分离卷积和空洞卷积

为了在保持高特征提取能力的同时降低计算量，深度可分离卷积[35]将标准卷积分解为深度卷积和逐点卷积，大幅减少了计算复杂度和参数量，从而有效减轻计算负担并降低过拟合的风险。空洞卷积[36]通过在卷积核元素之间引入空洞来扩展感受野，而无需增加计算量，从而捕捉更广泛的上下文信息。与通过增加卷积核大小以扩展感受野的方式相比，空洞卷积在不增加参数量和计算量的情况下实现了这一目标，在计算效率上展现出显著优势。通过设置不同的膨胀率，能够在同一层同时提取不同尺度特征，从而增强模型对多尺度信息的敏感度。

基于 SwiftFormer[37]中 Conv 以及空洞卷积的思想，提出 DDE 和 DCE，以提升模型的准确性。DDE 包括 1 个 3×3 深度可分离卷积（DWConv）、归一化层（Norm）、3×3 卷积（Conv），以及 1 个残差连接（Residual connection），模块详细结构如图 11 所示。

SwiftFormer 模型在每一个阶段的 SwiftFormer Encoder 前面添加深度可分离卷积模块，提取局部特征，受此启发，将 DDE 添加到 Swin 核心模块的 1、2、4 阶段之前，以便更有效地提取局部特征。其中，深度可分离卷积能够在保持特征提取能力的同时降低计算成本，使模型更轻量化且更高效。DDE 能够提取细粒度的局部特征，同时保留丰富的上下文信息，从而显著提升模型的整体表现和泛化能力。

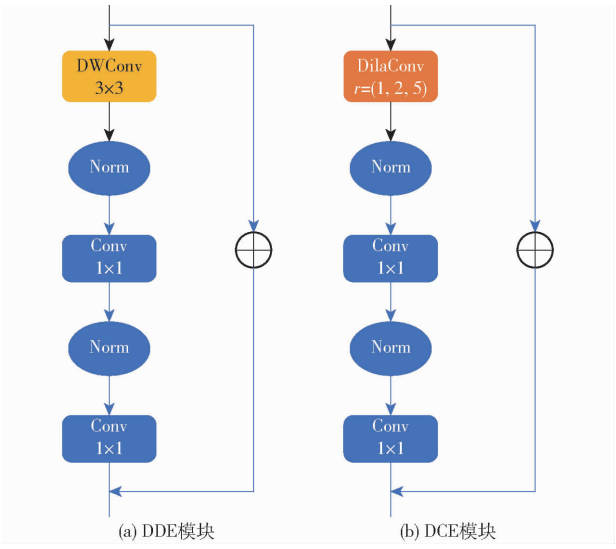


图 11　DDE 模块与 DCE 模块结构

Fig. 11　DDE and DCE module structures

DCE 包括一个空洞卷积（Dilated convolution，空洞数分别为 1、2、5）、归一化层、3×3 卷积，以及 1 个残差连接，模块结构如图 11b 所示。在模型推进至第 3 阶段时，已积累一定量的特征表示，但为更好地捕捉上下文信息并增强复杂场景下的适应性和稳定性，扩展感受野变得关键。为此，DCE 模块利用空洞卷积，通过空洞数 1、2、5 的组合达到最优效果：空洞数 1 增强对局部细节的捕捉，空洞数 2 适度扩展感受野以丰富局部信息，空洞数 5 则捕获更广泛的上下文信息。这样设计不仅在不增加计算量的情况下扩展了感受野，还显著提升了特征表示能力和模型鲁棒性。具体嵌入位置如图 12 所示。

通过将 DDE 和 DCE 模块结合使用，Swin Transformer 能够在保持高计算效率的同时，实现细粒度特征与全局信息的协同提取。这种特征

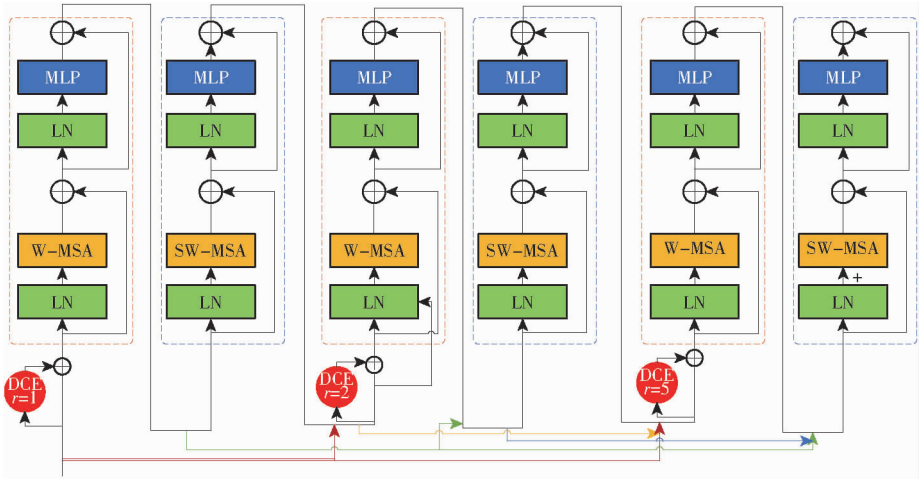


图 12　DCE 详细嵌入结构图

Fig. 12　Embed structure diagram in detail

重用机制能够有效增强模型在不同尺度下的特征表示能力,进而提升对复杂图像的识别与分类性能。

1.2.5 注意力 Non-Local 模块

Non-Local 模块能够捕获时间(如一维时序信号)、空间(如图像)以及时空(如视频序列)中的长程依赖。不同于反复堆叠卷积层和循环神经网络(Recurrent neural network, RNN)算子,Non-Local 操作通过直接计算不同位置之间的关系,高效捕获长程依赖。本文在 Swin Transformer 的核心模块后添加了 Non-Local 模块,以进一步优化模型的最终表现。该模块能够有效表达目标信息,抑制无关信息,帮助模型聚焦于重要特征,从而更好地适应复杂背景,特别是在处理小目标、目标重叠或稀疏地理对象分布不均等问题时表现尤为出色。

具体实现细节如图 13 所示。首先,对输入特征图 x 进行线性映射(1×1 卷积),以减少通道数并生成特征 θ、ϕ 和 $\boldsymbol{g}$;通过 Reshape 操作合并这 3 个特征的非通道维度;对和进行矩阵点积,得到类似协方差矩阵的结果;对自相关特征应用 Softmax 操作,生成取值范围在[0,1]之间的 Self-attention 系数。最后,将 Attention 系数与特征矩阵 $\boldsymbol{g}$ 逐元素相乘,再通过 1×1 卷积扩展通道数,并与原始特征图 x 进行残差连接,最终得到 Non-Local Block 的输出,即

$$y_i = \frac{1}{C(x)} \sum_{\forall j} f(x_i, x_j) \boldsymbol{g}(x_j) \tag{7}$$

式中 x——输入特征图

y——输出结果

i——时间、空间或时空上一个点

j——所有可能位置的索引

f——配对计算函数

$C(x)$——归一化函数

f 用于计算位置 i 与所有其他位置之间的相关性。$\boldsymbol{g}$ 负责信息变换,而 $C(x)$ 确保信息在变换前后保持一致。

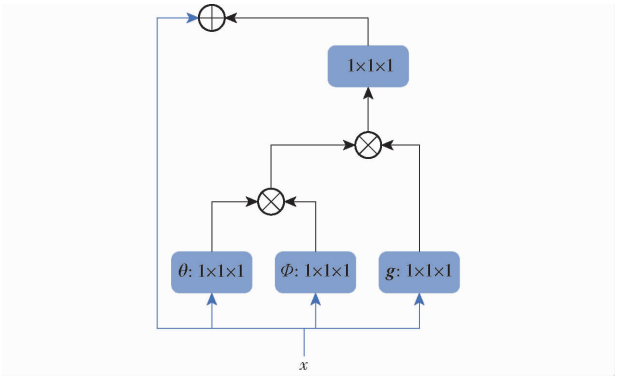


图 13 非局部注意力模块

Fig. 13 Non-partial attention module

2 结果与分析

2.1 实验环境

网络模型训练实验环境配置如表 2 所示。

表 2 实验环境配置

Tab. 2 Experimental environment configuration

实验环境	型号及版本
操作系统	Windows 10
编程语言	Python 3.8.18
深度学习框架	Pytorch 2.0.1
GPU 加速库	CUDA 12.2, CUDNN 8.9.5
CPU	Intel(R) Core(TM) i7-9700 CPU @ 3.00GHz
GPU	NVIDIA GeForce GTX 1660

2.2 超参数设置

为了实现最佳分类效果,通过调整超参数进行了多次实验。首先确定每批次训练样本的数量(Batch size)和模型的学习率(Learning rate),然后调整其他参数。每次实验运行 100 轮(epochs),即使用训练集全部数据进行一次完整训练。批量大小通常设置为 2 的次幂,以便并行计算。测试了 4、8、16 3 种批量,结果发现较小的批量会导致收敛较慢,而较大的批量内存占用量过高,训练速度反而较慢。因此,最终选择批量为 4。学习率影响网络权重更新速度,合理设置能使目标函数更快收敛。实验测试了 0.01、0.001 和 0.000 1 3 种学习率,结果显示学习率 0.000 1 效果最佳。最终采用上述默认参数设置进行模型训练。

2.3 评价指标

采用召回率(Recall)、F1 分数、精确率(Precision)、推理时间、帧率、浮点运算量和准确率(Accuracy)全面评估模型性能。

2.4 苹果叶部病害分类模型消融实验

为评估各模块对苹果叶片病害分类模型的影响,设计多组实验,结果如表 3 所示。

首先构建基准模型(Swin Transformer 原始模型),其分类准确率为 90.7%,为后续对比提供基础。加入 Swin(DenseNet)模块后,分类准确率提升至 92.5%,表明其层次化注意力机制增强了特征提取。基于此模型增加 DCE 模块后,分类准确率进一步提升至 94.1%,且推理时间增加较少。DCE 模块通过优化特征表达效率,显著提升了分类性能。加入 DDE 模块后,准确率提高至 94.6%,帮助模型更好适应数据多样性;当 DCE 和 DDE 模块联合使用时,准确率达到 95.2%,性能显著提升。引入 Non-Local 模块后,准确率达到 94.8%,通过全局特征捕捉增强了特征表示能力。集成 Outlook 模块后,准

表 3　消融实验结果

Tab. 3　Ablation experiment results

Swin(DenseNet)	DCE	DDE	Non－Local	Outlook	准确率/%	参数量	推理时间/s	帧率/(f·s^{-1})
					90.7	2.75×10^7	35	33.76
√					92.5	2.75×10^7	41	29.48
	√				94.1	3.94×10^7	46	26.03
		√			94.6	3.19×10^7	43	27.29
	√	√			95.2	4.38×10^7	47	25.25
			√		94.8	2.87×10^7	41	28.90
				√	94.7	2.77×10^7	43	27.75
√	√	√	√	√	95.8	4.51×10^7	48	24.93

注：√表示采用此模块。

确率为 94.7%，提升了特定类别分类效果，但略逊于单独加入 DCE 或 DDE 模块。最终，集成所有模块的模型准确率达到最高(95.8%)，表明各模块联合能实现最优特征表达与分类性能。

综上所述，消融实验显示各模块对模型性能提升均有贡献。Swin(DenseNet)、DCE、DDE、Non－Local 和 Outlook 模块组合显著优化了模型表现，为模型设计提供了重要参考依据。

2.5　模型对比

2.5.1　改进模型与基准模型对比

为了评估改进 Swin－Transformer 模型在苹果叶片病害分类中的表现，与 Swin－Transformer 模型进行了对比。表 4 展示了两者在不同病害类别上的精确率、召回率和 F1 分数的对比。Swin－Transformer 模型在各病害类别上的表现较为稳定，其中花叶病精确率最高，达到 99%，健康叶片精确率较低，为

表 4　改进模型对比基准模型

Tab. 4　Improved model *vs* benchmark model　　%

模型	评价指标	链格孢叶斑病	褐斑病	蛙眼叶斑病	灰斑病	健康	花叶病	白粉病	锈病	黑星病
Swin Transformer	精确率	90	93	92	88	86	99	94	89	87
	召回率	81	98	93	88	92	88	91	94	89
	F1 分数	85	96	93	88	89	93	92	91	88
改进 Swin Transformer	精确率	94	100	91	92	95	100	97	97	96
	召回率	96	100	98	94	98	96	99	93	88
	F1 分数	95	100	94	93	97	98	98	95	92

86%。整体上，该模型对锈病识别效果较好，召回率和 F1 分数均为 91%，但对灰斑病识别效果较差，F1 分数仅为 88%。

改进 Swin－Transformer 模型在所有病害类别上均表现出显著提升。精确率方面，灰斑病从 88% 提升至 92%，健康叶片从 86% 提升至 95%，整体模型精确率显著提高。召回率方面，改进模型在所有病害类别上均超过 94%，特别是健康叶片的召回率从 92% 提升至 98%，表明改进模型在各类别病害的识别上更加全面且准确。F1 分数作为综合精确率和召回率的指标，改进 Swin－Transformer 模型在所有类别上表现优异，其中白粉病和花叶病 F1 分数均为 98%，表明模型不仅能准确识别这些病害，还能在各类别中保持高水平的分类一致性。然而，灰斑病精确率较低，经分析数据集发现，这 2 类病害叶片样本数量在所有类别中最少，因此认为精确率较低的原因是数据量较少。综上所述，改进 Swin－Transformer 模型在精确率、召回率和 F1 分数等方面均优于基准模型，展示了更强的病害分类能力和更好的泛化性能。由图 14(其中 0 表示链格孢叶斑病，1 表示褐斑病，2 表示蛙眼叶斑病，3 表示灰斑病，4 表示健康，5 表示花叶病，6 表示白粉病，7 表示锈病，8 表示黑星病)可以看到，改进模型在不同病害类别之间的分类错误率大幅减少，尤其是在灰斑病和健康叶片之间的误分类问题得到了有效缓解。改进模型的表现证明了在 Swin－Transformer 模型中进行优化设计的有效性，为进一步提升苹果叶片病害分类的准确性提供了重要参考。

2.5.2　不同模型性能对比

为了评估改进 Swin Transformer 模型在苹果叶片病害分类中的表现，对 ViT、Swin Transformer(Swin)、Swin Transformer－V2(Swin V2)、VGG16[38]、ResNet101 和改进 Swin Transformer 模型进行了对比实验。图 15 展示了各模型在训练过程

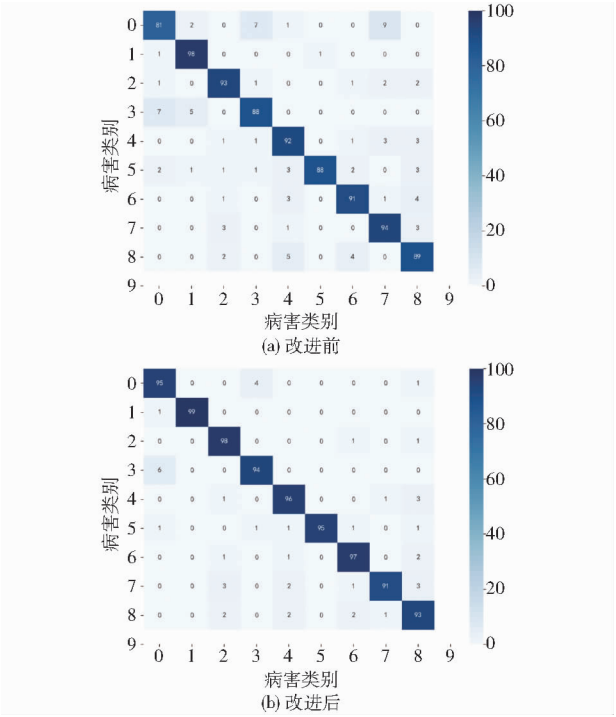


图 14 改进前后 Swin Transformer 混淆矩阵

Fig. 14 Swin Transformer confusion matrix before and after improvement

中的损失值和准确率变化曲线,且不同模型在训练表现上有显著差异,ResNet101 和改进 Swin Transformer 表现较好。相较之下,改进 Swin Transformer 的表现更加平稳,损失曲线下降趋势更加一致,并在后期收敛至较低的损失值。同时,尽管改进 Swin Transformer 在训练过程中的准确率略低于 ResNet101,但由图 15 可以看出,改进 Swin Transformer 展现出更好的泛化能力,其整体性能优于 ResNet101。表明改进 Swin Transformer 在处理复杂数据和提高模型鲁棒性方面具有显著优势。表明通过对 Swin 模型优化,增强了其在训练过程中的稳定性和最终分类性能。

表 5 展示了各模型在准确率、精确率、召回率和 F1 分数等指标的对比。改进 Swin Transformer 模型在这些核心指标上表现最佳,准确率达到 95.8%,远高于 ViT(81.5%)和 Swin(90.7%)。其精确率、召回率和 F1 分数也分别为 95.80%、95.74% 和 95.76%,均优于其他模型。尽管该模型参数量较多(4.51×10^7),但其计算复杂度相对较低,浮点运算量为 5.92×10^9。

综上所述,改进 Swin Transformer 模型在多个关键指标上表现突出,且在计算复杂度上保持合理平衡,验证了其在苹果叶片病害分类中的有效性。

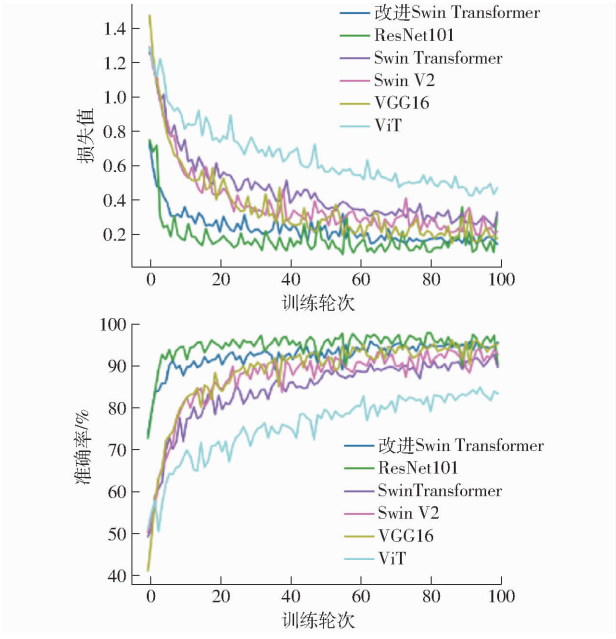


图 15 不同模型损失值和准确率变化曲线

Fig. 15 Loss and accuracy curves of different models

表 5 不同模型效果对比

Tab. 5 Comparison of different models

模型	准确率/%	精确率/%	召回率/%	F1 分数/%	浮点运算量	参数量
ViT	81.50	81.87	81.37	81.07	1.69×10^{10}	8.58×10^7
Swin	90.70	90.91	90.60	90.60	4.38×10^9	2.75×10^7
Swin V2	93.30	93.40	93.27	93.27	3.35×10^9	2.76×10^7
VGG16	94.60	94.83	94.55	94.57	1.55×10^{10}	1.34×10^8
ResNet101	90.80	92.17	90.62	90.64	7.86×10^9	4.25×10^7
改进 Swin Transformer	95.80	95.80	95.74	95.76	5.92×10^9	4.51×10^7

3 结束语

提出了一种基于改进 Swin Transformer 的苹果叶部病害分类模型,通过引入 DenseNet 特征重用、Outlook 注意力机制、深度可分离卷积、膨胀卷积和 Non-Local,本模型在捕捉细粒度特征和全局上下文信息方面表现优异,且在多个任务上展现了出色的性能和鲁棒性。实验结果显示,改进模型在苹果叶片病害分类中准确率达到 95.8%,精确率、召回率和 F1 分数显著优于基线模型,验证了方法有效性。

参 考 文 献

[1] 樊东燕. 基于 CNN 的植物叶片病害智能检测系统算法设计与仿真[J]. 农业技术与装备, 2022(11): 36 – 37,40.
FAN Dongyan. Design and simulation of plant leaf disease intelligent detection system algorithm based on CNN[J]. Agricultural Technology & Equipment, 2022(11): 36 – 37,40. (in Chinese)

[2] VASWANI A. Attention is all you need[J]. Advances in Neural Information Processing Systems, 2017,30:5998 – 6008.

[3] 魏丽冉, 岳峻, 李振波, 等. 基于核函数支持向量机的植物叶部病害多分类检测方法[J]. 农业机械学报, 2017,48(增刊): 166 – 171.
WEI Liran, YUE Jun, LI Zhenbo, et al. Multi-classification detection method of plant leaf disease based on kernel function SVM [J]. Transactions of the Chinese Society for Agricultural Machinery, 2017,48(Supp.): 166 – 171. (in Chinese)

[4] 王建玺, 宁菲菲, 鲁书喜. 基于支持向量机的苹果叶部病害识别方法研究[J]. 山东农业科学, 2015,47(7): 122 – 125,141.
WANG Jianxi, NING Feifei, LU Shuxi. Study on apple leaf disease recognition method based on support vector machine[J]. Shandong Agricultural Sciences, 2015,47(7): 122 – 125,141. (in Chinese)

[5] 师韵, 王旭启, 张善文. 基于主分量分析的苹果叶部 3 种常见病害识别方法[J]. 江苏农业科学, 2016,44(9): 337 – 340.

[6] SI H, LI M, LI W, et al. A dual-branch model integrating CNN and Swin Transformer for efficient apple leaf disease classification[J]. Agriculture, 2024,14(1): 142.

[7] 姜红花, 杨祥海, 丁睿柔, 等. 基于改进 ResNet18 的苹果叶部病害多分类算法研究[J]. 农业机械学报, 2023,54(4): 295 – 303.
JIANG Honghua, YANG Xianghai, DING Ruirou, et al. Identification of apple leaf diseases based on improved ResNet18[J]. Transactions of the Chinese Society for Agricultural Machinery, 2023,54(4): 295 – 303. (in Chinese)

[8] 董跃华, 彭辉林. 改进 Inception 结构的图像分类方法[J]. 软件导刊, 2023,22(2): 41 – 46.
DONG Yuehua, PENG Huilin. Image classification method based on improved Inception structure[J]. Software Guide, 2023, 22(2):41 – 46. (in Chinese)

[9] LUO Y, SUN J, SHEN J, et al. Apple leaf disease recognition and sub-class categorization based on improved multi-scale feature fusion network[J]. IEEE Access, 2021,9: 95517 – 95527.

[10] CHEN X, XING X, ZHANG Y, et al. MSCR – FuResNet: a three-residual network fusion model based on multi-scale feature extraction and enhanced channel spatial features for close-range apple leaf diseases classification under optimal conditions[J]. Horticulturae, 2024,10(9): 953.

[11] TANG L, YI J, LI X. Improved multi-scale inverse bottleneck residual network based on triplet parallel attention for apple leaf disease identification[J]. Journal of Integrative Agriculture, 2024,23(3): 901 – 922.

[12] YU H J, SON C H. Apple leaf disease identification through region-of-interest-aware deep convolutional neural network[J]. arXiv Preprint,arXiv: 1903.10356, 2019.

[13] 刘拥民, 刘翰林, 石婷婷, 等. 一种优化的 Swin Transformer 番茄叶片病害识别方法[J]. 中国农业大学学报, 2023, 28(4):80 – 90.
LIU Yongmin, LIU Hanlin, SHI Tingting, et al. Tomato leaf disease recognition based on an optimized Swin Transformer[J]. Journal of China Agricultural University, 2023,28(4): 80 – 90. (in Chinese)

[14] LIU Z, HU H, LIN Y, et al. Swin Transformer v2: scaling up capacity and resolution[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022: 12009 – 12019.

[15] 方俊泽, 郭正, 李歌, 等. 基于改进 Swin – Transformer 的柑橘病叶分类模型[J]. 中国农机化学报, 2024,45(1): 252 – 258.
FANG Junze, GUO Zheng, LI Ge, et al. Classification model of citrus disease leaf based on improved Swin – Transformer[J]. Journal of Chinese Agricultural Mechanization, 2024,45(1): 252 – 258. (in Chinese)

[16] TOUVRON H, CORD M, DOUZE M, et al. Training data-efficient image transformers & distillation through attention[C]// International Conference on Machine Learning, PMLR,2021: 10347 – 10357.

[17] GUO J, HAN K, WU H, et al. Cmt: convolutional neural networks meet vision transformers[C]// CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2022: 12165 – 12175.

[18] PENG Z, HUANG W, GU S, et al. Conformer: local features coupling global representations for visual recognition[C]// Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021: 367 – 376.

[19] DOSOVITSKIY A. An image is worth 16 × 16 words: Transformers for image recognition at scale[J]. arXiv Preprint,arXiv: 2010.11929, 2020.

[20] YUAN L, CHEN Y, WANG T, et al. Tokens-to-token vit: training vision transformers from scratch on imagenet[C]// Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021: 558 – 567.

[21] DANG L, PANG P, LEE J. Depth-wise separable convolution neural network with residual connection for hyperspectral image classification[J]. Remote Sensing, 2020,12(20): 3408.

[22] WANG X, YUAN L, XU H, et al. CSDS: end-to-end aerial scenes classification with depthwise separable convolution and an attention mechanism[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2021,14: 10484 – 10499.

[23] LEI X, PAN H, HUANG X. A dilated CNN model for image classification[J]. IEEE Access, 2019,7: 124087 – 124095.

[24] ZHANG Q. A novel ResNet101 model based on dense dilated convolution for image classification[J]. SN Applied Sciences,

2022,4: 1 - 13.
[25] WANG X, GIRSHICK R, GUPTA A, et al. Non-local neural networks[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 7794 - 7803.
[26] 李薇, 樊彦国, 周培希. 改进 Transformer 的高光谱图像地物分类方法——以黄河三角洲为例[J]. 自然资源遥感, 2024,36(3): 137 - 145.
LI Wei, FAN Yanguo, ZHOU Peixi. Improved Transformer-based hyperspectral image classification method for surface features: a case study of the Yellow River delta[J]. Remote Sensing for Natural Resources, 2024,36(3): 137 - 145. (in Chinese)
[27] WEN Y, CHEN L, CHEN H, et al. Non-local attention learning for medical image classification[C]// 2021 IEEE International Conference on Multimedia and Expo (ICME). IEEE, 2021: 1 - 6.
[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] GAO H, ZHUANG L, MAATEN L, et al. Densely connected convolutional networks[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 4700 - 4708.
[30] YUAN L, HOU Q, JIANG Z, et al. Volo: vision outlooker for visual recognition[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022,45(5): 6575 - 6586.
[31] 李贞, 刘海燕, 刘策, 等. 基于机器学习的 K-Means 聚类优化算法研究[J]. 数据挖掘, 2022,12(1):20 - 26.
LI Zhen, LIU Haiyan, LIU Ce, et al. Research on K-Means clustering optimization algorithm based on machine learning[J]. Hans Journal of Data Mining, 2022,12(1):20 - 26. (in Chinese)
[32] 管正雄. 基于深度生成模型的数据增强方法研究[D]. 南京:南京信息工程大学, 2019.
GUAN Zhengxiong. Data augmentation model based on deep generative model[D]. Nanjing: Nanjing University of Information Science & Technology, 2019. (in Chinese)
[33] 马晓, 董天亮, 钟闻宇, 等. 基于改进 ConvNeXt 的大豆叶片病害分类研究[J]. 大豆科学, 2023,42(6): 733 - 741.
MA Xiao, DONG Tianliang, ZHONG Wenyu, et al. Soybean leaf diseases classification method based on improved ConvNeXt [J]. Soybean Science, 2023,42(6): 733 - 741. (in Chinese)
[34] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016: 770 - 778.
[35] SIFRE L, MALLAT S. Rigid-motion scattering for texture classification[J]. arXiv Preprint, arXiv: 1403.1687, 2014.
[36] YU F. Multi-scale context aggregation by dilated convolutions[J]. arXiv Preprint, arXiv:1511.07122, 2015.
[37] SHAKER A, MAAZ M, RASHEED H, et al. Swiftformer: efficient additive attention for transformer-based real-time mobile vision applications[C]// Proceedings of the IEEE/CVF International Conference on Computer Vision, 2023: 17425 - 17436.
[38] KAREN S. Very deep convolutional networks for large-scale image recognition[J]. arXiv Preprint, arXiv: 1409.1556, 2014.

(上接第 180 页)

[28] 刘琳,郑兴明,姜涛,等.无人机遥感植被覆盖度提取方法研究综述[J].东北师大学报(自然科学版),2021,53(4):151 - 160.
LIU Lin, ZHENG Xingming, JIANG Tao, et al. Extraction method of fractional vegetation cover from unmanned aerial vehicle remote sensing: a review[J]. Journal of Northeast Normal University (Natural Science Edition), 2021,53(4):151 - 160. (in Chinese)
[29] 赵燕红,侯鹏,蒋金豹,等.植被生态遥感参数定量反演研究方法进展[J].遥感学报,2021,25(11):2173 - 2197.
ZHAO Yanhong, HOU Peng, JIANG Jinbao, et al. Progress in quantitative inversion of vegetation ecological remote sensing parameters[J]. National Remote Sensing Bulletin, 2021,25(11):2173 - 2197. (in Chinese)
[30] 韩文霆,彭星硕,张立元,等.基于多时相无人机遥感植被指数的夏玉米产量估算[J].农业机械学报,2020,51(1):148 - 155.
HAN Wenting, PENG Xingshuo, ZHANG Liyuan, et al. Summer maize yield estimation based on vegetation index derived from multi-temporal UAV remote sensing[J]. Transactions of the Chinese Society for Agricultural Machinery, 2020,51(1):148 - 155. (in Chinese)
[31] 姚付启,蔡焕杰,王海江,等.基于平稳小波变换的冬小麦覆盖度高光谱监测[J].农业机械学报,2012,43(3):173 - 180.
YAO Fuqi, CAI Huanjie, WANG Haijiang, et al. Monitoring winter wheat percentage vegetation cover based on stationary wavelet transformation derived from hyperspectral reflectance[J]. Transactions of the Chinese Society for Agricultural Machinery, 2012,43(3):173 - 180. (in Chinese)
[32] 虞连玉,蔡焕杰,姚付启,等.植被指数反演冬小麦植被覆盖度的适用性研究[J].农业机械学报,2015,46(1):231 - 239.
YU Lianyu, CAI Huanjie, YAO Fuqi, et al. Applicability of vegetation indices to estimate fractional vegetation coverage[J]. Transactions of the Chinese Society for Agricultural Machinery, 2015,46(1):231 - 239. (in Chinese)