

马柏雄, 刘成忠, 韩俊英, 等. 基于 Swin Transformer 的马铃薯植株和品种识别[J]. 智能计算机与应用, 2025, 15(9): 12–18.  
DOI: 10.20169/j. issn. 2095-2163. 250903

# 基于 Swin Transformer 的马铃薯植株和品种识别

马柏雄<sup>1</sup>, 刘成忠<sup>1</sup>, 韩俊英<sup>1</sup>, 曲亚英<sup>2</sup>, 邢 雪<sup>1</sup>

(1 甘肃农业大学 信息科学技术学院, 兰州 730070; 2 甘肃省农业科学院 马铃薯研究所, 兰州 730070)

**摘要:** 马铃薯在国内的农业资源中占据着重要地位, 拥有丰富多样的品种。准确识别马铃薯品种对于推动马铃薯育种发展至关重要。因此, 本研究基于深度学习网络提出了一种多品类马铃薯植株识别模型。该模型采用了基于 Swin Transformer 的架构, 并通过对注意力机制的巧妙改进, 有效提升了模型的特征提取能力。与此同时, 通过减少模型参数量, 该模型的准确率得到了显著提升。原始的 Swin Transformer 在对 30 个马铃薯品种进行植株识别时的准确率为 95.0%, 而改进后的 Swin Transformer 达到了 97.1%, 提升了 2.1 个百分点。研究结果明确显示, 改进后的 Swin Transformer 模型在对马铃薯植株进行识别分类方面优于原始 Swin Transformer 模型。深度学习网络模型在马铃薯植株种类识别方面展现出可行性, 为其在实际生产中的推广应用提供了有力支持。

**关键词:** 马铃薯植株; 品种分类; 深度学习; Swin Transformer

中图分类号: S532; TP391.4 文献标志码: A 文章编号: 2095-2163(2025)09-0012-07

## Potato plants and variety identification based on Swin Transformer

MA Baixiong<sup>1</sup>, LIU Chengzhong<sup>1</sup>, HAN Junying<sup>1</sup>, QU Yaying<sup>2</sup>, XING Xue<sup>1</sup>

(1 College of Information Science and Technology, Gansu Agricultural University, Lanzhou 730070, China;

2 Potato Research Institute of Gansu Academy of Agricultural Sciences, Lanzhou 730070, China)

**Abstract:** Potatoes play a vital role in China's agricultural landscape, boasting a diverse array of varieties. The accurate identification of these varieties is crucial for advancing potato breeding. This study proposes a multi-variety potato plant identification model based on the deep learning network. The model leverages the architecture of the Swin Transformer, enhancing feature extraction through clever improvements in the attention mechanism. Simultaneously, a reduction in the number of model parameters results in a significant boost in accuracy. The original Swin Transformer achieves 95.0% accuracy in identifying 30 potato varieties. In contrast, the enhanced Swin Transformer achieves an impressive 97.1%, marking a substantial 2.1 percentage point improvement. These results unequivocally demonstrate the superiority of the enhanced Swin Transformer model in the identification and classification of potato plants over its original counterpart. The deep learning network model has shown feasibility in potato plant species identification, which provides strong support for its popularization and application in actual production.

**Key words:** potato plants; breed classification; deep learning; Swin Transformer

## 0 引言

马铃薯是全球重要的粮食作物之一<sup>[1]</sup>, 马铃薯品种的识别对马铃薯的育种与推广有着重要的意义。目前, 对于马铃薯的种类还主要依靠人工参与的方式鉴别, 使用人工参与的方法来识别大量相似度极高的图像并实现分类则需要消耗极大的时间成本和人力资源, 且效率低下、带有较高的主观性、难

以适应时代发展的需要<sup>[2-3]</sup>。近年来, 大量的研究表明, 利用深度学习技术可以高效地完成植物的种类识别<sup>[4-5]</sup>。然而, 在品种识别任务中, 该方法仍然面临较大的挑战。这主要是因为植物在生长过程中经历的形态变化以及不同发育情况导致同一物种的不同品种之间存在微小但显著的差异, 同时还存在大量的种内变异。在当前追求高效率和高信息化的时代, 国内外学者及专业研究机构等对以图像分析及图像处理为基础的植物识别展开了广泛的研究,

基金项目: 国家自然科学基金(32360437); 甘肃省高等学校产业支撑计划项目(2021CYZC-57); 甘肃省高等学校创新基金项目(2021A-056)。

作者简介: 马柏雄(1998—), 男, 硕士研究生, 主要研究方向: 农业工程与信息技术。

通信作者: 刘成忠(1969—), 男, 教授, 硕士生导师, 主要研究方向: 人工智能, 农业信息技术。Email: liucz@gsau.edu.cn。

收稿日期: 2023-12-26

哈尔滨工业大学主办 ◆ 学术研究与应用

并取得了一些成果<sup>[7]</sup>。2006年,Hinton等学者就提出了深度学习的概念,同时提出了有效的方法用于解决训练深度神经网络的难点,这一举动不但吸引了国内外学者的广泛关注,也推动了深度学习的发展<sup>[7]</sup>。2014年,邓丽苗等学者<sup>[8]</sup>针对玉米叶片单独做了48个叶片特征的提取,通过支持向量机算法研发了识别系统,达到了高达96%的识别率。2015年,王丽君等学者<sup>[9]</sup>利用图像多特征融合方法,成功开发了一套观叶植物种类识别系统。通过对观叶植物的形态等特征进行详尽分析,该系统提取了叶片图像的颜色、形状和纹理等共计26个特征。运用支持向量机(SVM)算法进行分类识别,最终取得了显著成果,识别率达到了91.41%。2015年,邹秋霞等学者<sup>[10]</sup>提出了基于Android手机的植物叶片分类系统,直接利用移动终端完成对叶片识别代表了新的发展趋势,该系统对样本的识别取得了不错的结果。2018年,陈文根等学者<sup>[11]</sup>基于深度卷积网络对9种小麦进行品种识别研究,采用Softmax分类器进行品种分类,达到97.78%的平均识别准确率。2017年,丁秋等学者<sup>[12]</sup>采集了15个品种小麦籽粒在388~1 009 nm波段的近红外高光谱图像,通过主成分分析优选特征波长,应用贝叶斯判别分析进行多元分析,建立判别函数判别回代准确率为99.9%,交叉验证的准确率为98%,模型取得了不错的效果。2020年,李鸿强等学者<sup>[13]</sup>采用高光谱分析技术结合模式识别,对8种马铃薯微型薯种分别使用线性判别、BP神经网络和支持向量机来搭建分类模型,最终建立了分步骤、分层分类识别的分类模型,其平均正确识别率为89.75%。表明光谱分析可用于马铃薯微型薯种分类检测。2017年,袁亮等学者<sup>[14]</sup>在对番茄实时分级研究中应用机器视觉技术进行建模,实现了自动化,且模型精度达到92%。2022年,孟志超等学者<sup>[15]</sup>通过网络结构调整提出了基于VGG模型的油茶品种分类模型Enhanced VGG16,其对油茶叶的品种分类准确率达到98.44%,模型综合性能得到优化。2021年,韩斌等学者<sup>[16]</sup>在深度学习方法中引入了传统特征,提出了一种基于多特征融合的叶片图像识别方法。该方法通过提取叶片的LBP特征和Gabor特征,将两者相加融合后输入卷积神经网络进行叶片分类,取得了平均正确识别率为96.37%的显著成果。

国内外学者在作物品种分类<sup>[17-18]</sup>领域做了深入研究,并取得了不错的成果,但将该技术应用到马铃薯品种分类鉴别中的鲜少见到。本研究将对不同品种马铃薯植株照片进行探索和研究,其目的是研

究出一种高效的马铃薯品种分类模型来满足当前生产应用的需求,为马铃薯品种的快速鉴别提供理论支撑和技术支持。

## 1 材料与方法

### 1.1 材料

实验采用的马铃薯品种共30个,来源于甘肃省农业科学院马铃薯研究所在榆中试验站的推广品种(系),主要品种有陇薯19号、陇薯20号、陇薯23号、陇薯25号、陇薯26号、陇薯12号等。图像数据于2023年7月底采集,在自然光照条件下使用Nikon COOLPIX B700全天候拍摄,拍摄过程中主要以晴天为主,有3~4天为阴雨天,这使得数据有更好的健壮性。本试验所构建的马铃薯植株图像数据集中,对30个品种共采集马铃薯植株图像样本27 300张,平均每个品种约900张。因多数马铃薯长势是匍匐于地表且长势较密集,在拍摄图像时,采用了多个角度,包括顶部、侧面、整体和部分等视角。图像的命名方式为“马铃薯品种编号\_植株编号\_相机视图编号.jpg”。例如,“1\_1\_1.jpg”表示第一个品种第一株的第一视角图像。图1(a)~(f)展示了马铃薯植株的6种视角图,其中(a)~(d)组采用了蓝色背景,(e)~(f)组模拟了自然田间拍摄。

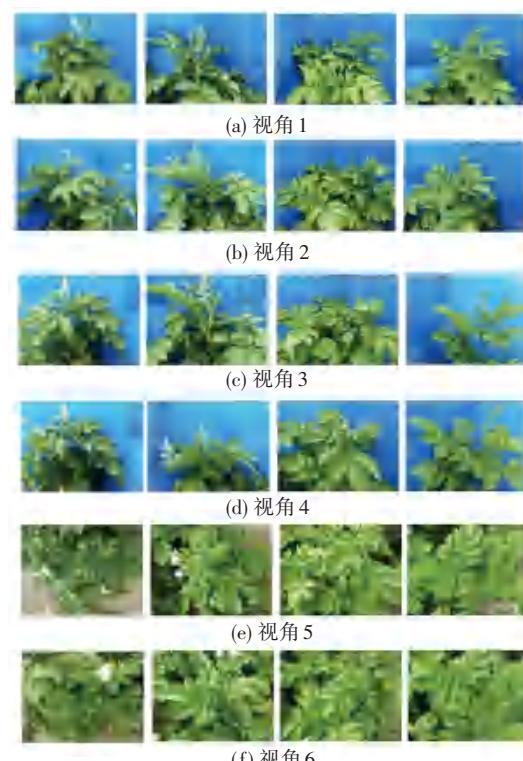


图1 马铃薯植株数据集部分图

Fig. 1 Partial plot of the potato plant dataset

## 1.2 Swin Transformer 简介

Swin Transformer 是一种基于 Vision Transformer 的计算机视觉模型,由微软亚洲研究院(MSRA)提出。Swin Transformer 通过将局部自注意力与分层

窗口移位相结合,实现了对图像的全局和局部表示的学习<sup>[19-20]</sup>。Swin Transformer 在多个计算机视觉任务性能上取得了显著提升,如图像分类、目标检测和语义分割等。其模型网络结构如图 2 所示。

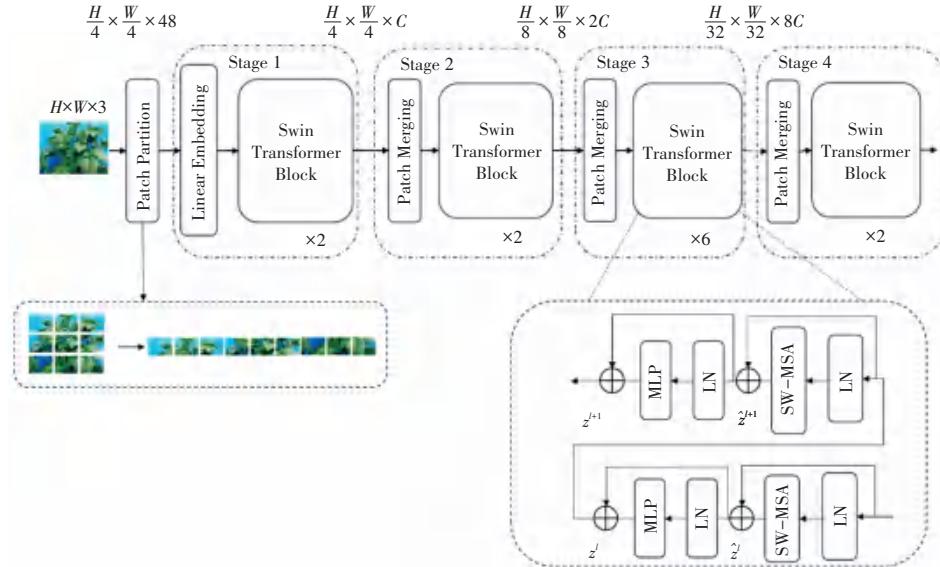


图 2 Swin Transformer 网络结构

Fig. 2 Swin Transformer network structure

Swin Transformer 模型的迭代过程主要包括以下几个步骤:首先,在 Patch Partition 层通过一个二维卷积将输入的图片切成多个 patch,然后进行嵌入向量,接着输入到 Linear Embedding 层做线性变换,稍后就是通过 4 个 Stage 构建不同大小的特征图。Stage1 中先通过一个 Linear Embedding 层外,剩下 3 个 Stage 都是先通过一个 Patch Merging 层进行下采样,继而都是重复堆叠 Swin Transformer Block。其输出的特征图如图 3 所示。由图 2 可知,每个 Block 通过窗口多头注意力(W-MSA)机制来计算复杂度,同时,通过位移窗口的多头自注意力机制(SW-

MSA)来实现窗口之间的信息交互,并保证高效的计算效率。研究推得的公式具体如下:

$$z^l = \text{MLP}(\text{LN}(z^l) + z^l) \quad (1)$$

$$\hat{z}^l = \mathbf{W} - \text{MSA}(\text{LN}(z^{l-1}) + z^{l-1}) \quad (2)$$

$$z^{l+1} = \mathbf{W} - \text{MSA}(\text{LN}(z^{l-1}) + z^{l-1}) \quad (3)$$

$$\hat{z}^{l+1} = \mathbf{SW} - \text{MSA}(\text{LN}(z^l) + z^l) \quad (4)$$

其中,  $z^l, \hat{z}^l, z^{l+1}, \hat{z}^{l+1}$  表示输出。

基于此,再续接上一个 LayerNorm 层、全局池化层以及全连接层得到最终输出。

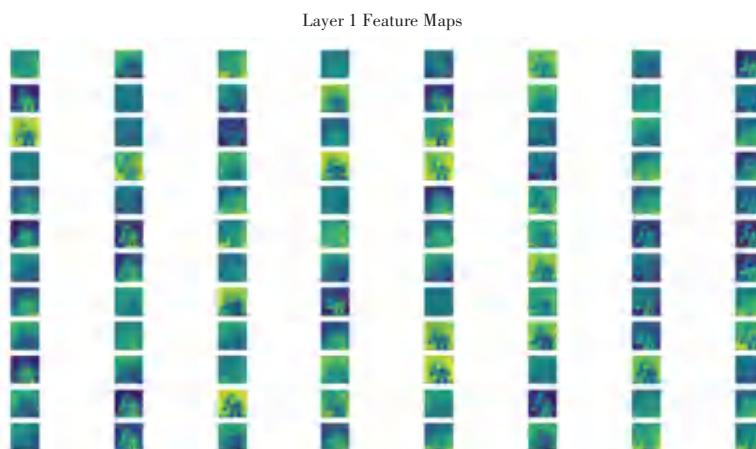


图 3 输出特征图

Fig. 3 Output feature map

### 1.3 模型的改进

Swin Transformer<sup>[19-20]</sup>引入了Shifted Windows机制,将全局自注意力机制改进为多尺度的局部自注意力,使用多头自注意力机制,通过引入多个独立的注意力头,使得模型能够在不同的子空间中学习不同类型的关系。每个注意力头都对输入序列进行不同的投影和关注,以捕捉多样性的信息<sup>[21-22]</sup>。在计算注意力时,输入序列首先通过每个注意力头的线性投影,并将这些头的输出拼接在一起。随后,通过线性转换将多头注意力的输出融合为最终的输出,同时Swing Transformer引入了相对位置编码,以捕捉不同位置之间的相对关系。总体而言,Swin Transformer的注意力机制通过Shifted Windows的局部自注意力机制实现了对图像中不同位置关系的建模,提高了计算效率,同时保持对图像全局信息的充分感知。但相对位置编码的引入,模型的计算参数量随即增加,加大了模型的计算复杂度,同时使得模型数据过拟合,达不到泛化效果。因此,本试验做了以下几个调整:取消了相对位置编码机制;使用了线性层self.linear\_1计算最终输出;Dropout应用在输出上,通过self.attn\_drop实现。

## 2 试验结果与分析

试验算法平台为曙光云计算服务平台,模型训练与测试的软件环境为Linux 64位系统,选择Python3.8编程语言,加速卡为国产dtk,深度学习框架为Pytorch2.0。

### 2.1 模型评价标准

仅使用识别准确率作为评价指标存在较大缺陷,往往无法全面评估模型在各个时期的详细分类效果等有价值的信息。为了更全面地评估模型性能,本试验采用了准确率、精确度、召回率和特异度等4个指标<sup>[23]</sup>。其中,准确率(Accuracy)表示分类正确的马铃薯品种占所有品种的比例;精确度(Precision)表示正确分类的马铃薯种类数量;召回率(Recall)表示某一品种中分类正确的数量比例;特异度(Specificity)指真正不属于某一种类的图像中,被判断为非此类的比例。计算公式如下:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \times 100\% \quad (5)$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \times 100\% \quad (6)$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \times 100\% \quad (7)$$

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}} \times 100\% \quad (8)$$

### 2.2 训练结果

#### 2.2.1 准确率与损失分析

模型准确率与训练损失结果如图4所示。需要提及的是,loss与acc是改进后,loss00与acc00是原始模型。从图4中可以看出,原模型的初始损失约为3.00左右,前30次迭代下降快速,30~90次下降缓慢,90次以后趋于稳定约为0.15;改进后初始损失约为2.7,前20次迭代快速下降,20~60次缓慢下降,60次以后基本趋于平稳,稳定值约0.11左右。由此可以看出,改进后的模型损失下降速度快,且值较为稳定。就其准确率来看,原模型在40次迭代后趋于稳定且最终最高准确率为95.0%;改进后模型在30次迭代后就增长缓慢,最终最高准确率为97.1%。通过与原模型对比,改进后模型收敛速度明显变快,且准确率高于原模型。

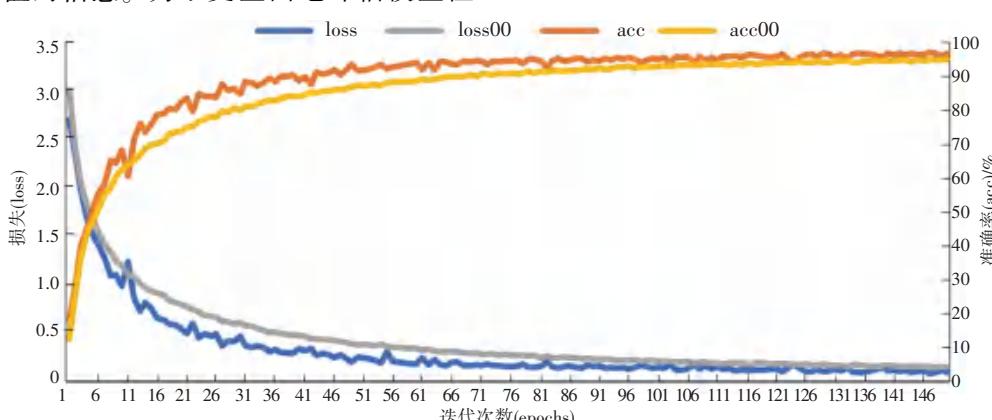


图4 网络训练准确率与损失

Fig. 4 Network training accuracy and loss

#### 2.2.2 混淆矩阵分析

混淆矩阵作为一种基本的性能评价工具,可以将

分类模型的预测结果和真实类别之间的关联用矩阵表示出来。本试验用5443张马铃薯植株图片对改

进后的模型进行了测试,其分类结果混淆矩阵如图 5 所示,横轴表示马铃薯品种的正确类别,纵轴表示模型预测值。矩阵对角线表示其正确分类值,其余为混淆分类。由图 5 中可以看出,模型总体分类效果比较理想,L14140-5 和 1520-1 分类效果最好,所有测试图片均被正确分类;L12116-19、LS12、LS17、LS25、Ly1916-3、Ly1624-2、Ly1722-1、Ly1727-3 和 Ly1733-16 分类效果较差,其中 Ly1722-1 分类准确率最低,共 183 张照片中有 13 张未正确分类;其次就是

Ly1727-3 与 Ly1733-16, Ly1727-3 实际 203 张图片,只有 192 张正确分类,Ly1733-16 实际 183 张图片,正确分类 172 张。通过混淆矩阵可以看出,Ly1722-1 易被误认为 Ly1615-3, Ly1619-3 易认成 L12116-19, Ly1733-16 易被认成 Ly1628-1。容易误分的原因可能在于皆同属于一个大类,且又可能是同一亲本,导致其植株纹理极为相似,给网络模型的分类识别增加了一定的难度。总体来讲,改进后的模型总体分类效果较好,可以通过植株对马铃薯品种进行准确分类。

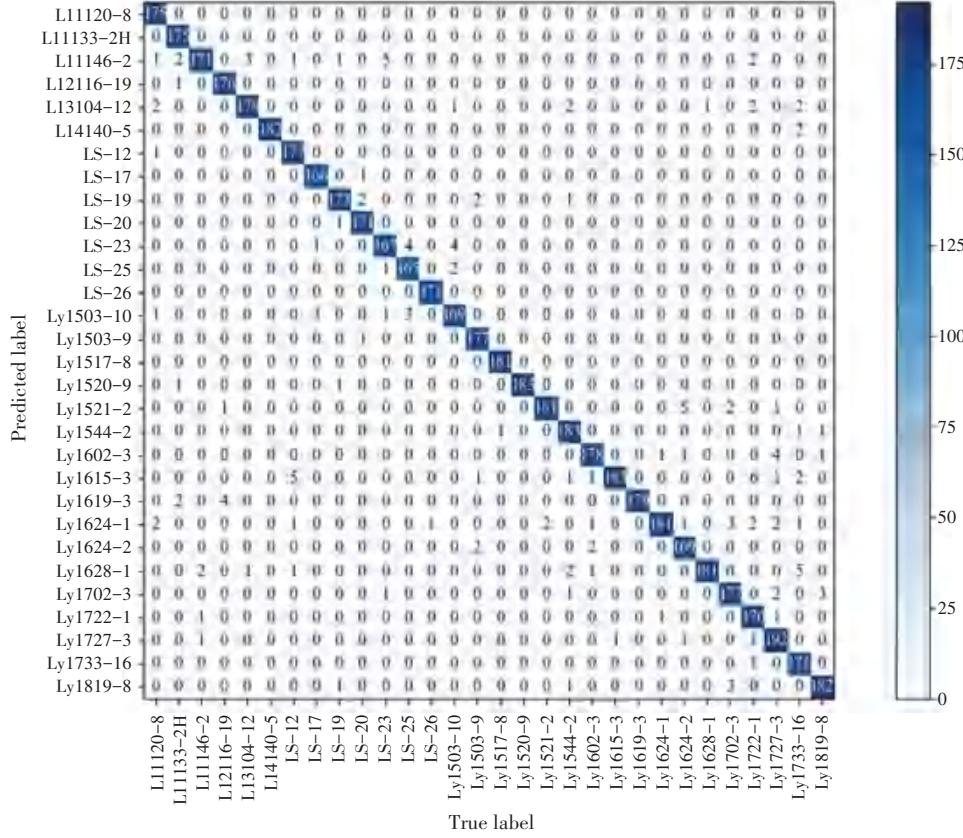


图 5 混淆矩阵

Fig. 5 Confusion matrix

### 3 精确度、召回率和特异度

改进后模型与原模型的精确度、召回率和特异度如图 6 所示。需要提及的是,Precision、Recall 与 Specificity 是改进后, Precision 0、Recall 0 与 Specificity 0 是原始模型。精确度率是分类正确马铃薯种类占全种类比例,由图 6(a)可知,改进后模型精确率较高,各品种已均超 90%,其中 L11120-8、L11133-2H、LS-26、Ly1517-8 的精确度达 100%,平均精确率为 97.2%;改进前模型型精确率也比较高,均在 85%之上,平均精确率达 96.1%。相比之下,改进后模型的精确率基本上都比原先高,各个分

类之间精确率差距较小,相对比较稳定。召回率(Recall)是对某一品种中分类正确数量的比例。由图 6(b)中可以看出改进后各品种之间的召回率差距较小,进一步说明了改进后模型识别的稳定性。从特异度上看,改进前后均高于 99.5%,进一步说明此模型适用于马铃薯品种的分类。

为了验证改进后模型的优越性,与其他主流模型进行了对比试验,其结果见表 1。所有模型在同一环境下进行训练,采用相同的参数。由表 1 可以看出,改进后模型的平均准确率与最高准确率均高于原始模型、GoogLeNet 和 ResNet34,说明了改进后模型的优越性,改进是比较成功的。

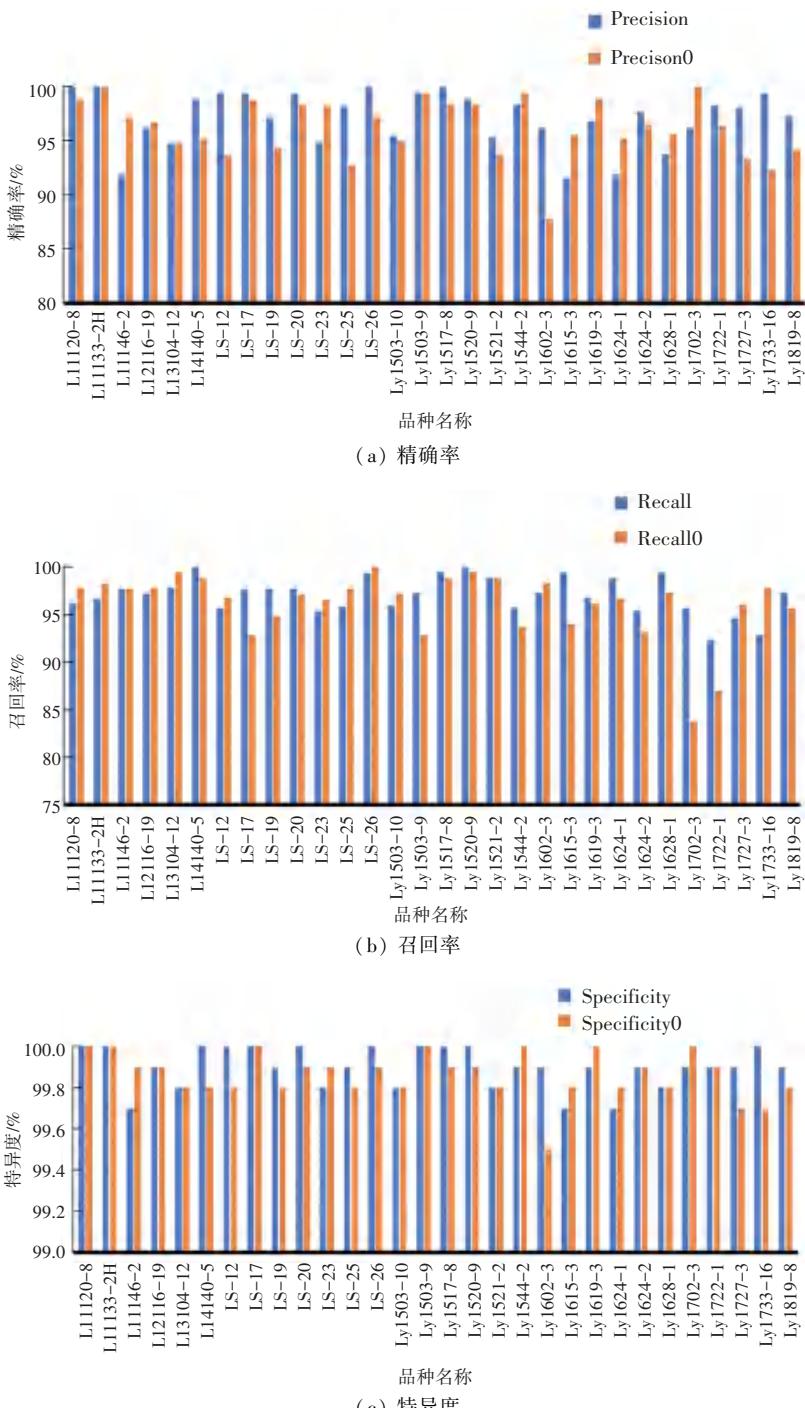


图 6 精准率、召回率与特异度

Fig. 6 Accuracy, recall, and specificity

表 1 不同模型识别结果

Table 1 Identification results of different models

| 模型               | 迭代次数 | 平均准确率   | 最高准确率 |
|------------------|------|---------|-------|
| Swin Transformer | 150  | 0.857 8 | 0.953 |
| GoogLeNet        | 150  | 0.862 3 | 0.939 |
| ResNet34         | 150  | 0.894 4 | 0.960 |
| 本文模型             | 150  | 0.894 9 | 0.973 |

## 4 讨论

本试验以深度学习网络为基础, 改进了 Swin Transformer 模型, 旨在提高对多品类马铃薯植株的准确识别。通过对比原始模型和改进后模型损失下降速度和准确率方面的表现, 结果显示改进后的模型损失下降速度更快, 准确率更高。混淆矩阵分析

展示了改进后模型对各个马铃薯品种的分类效果, 总体表现较为理想。在精确度、召回率和特异度方面的对比分析中, 改进后模型呈现出更高的精确度, 各品种精确率均超过 90%, 召回率差距较小, 显示出模型的分类稳定性。在特异度方面, 改进后模型与原始模型均高于 99.5%, 验证了模型在马铃薯品种分类上的适用性。最后, 通过与其他主流模型进行对比试验, 结果显示改进后模型的平均准确率和最高准确率均优于原始模型、GoogLeNet 和 ResNet34, 证明了改进的有效性。

## 5 结束语

改进后的 Swin Transformer 模型在马铃薯植株多品种识别任务中表现出色。相对于原始模型, 改进后模型在损失下降速度、准确率、精确度、召回率等方面均取得显著提升。混淆矩阵的分析结果显示, 改进后模型能够更准确地对各个品种进行分类, 证明了其在实际应用中的可行性和优越性。

## 参考文献

- [1] 施维, 才颖, 黄缨. 中华颐养书水果蔬菜颐养方 [M]. 上海: 上海科学技术文献出版社, 2020.
- [2] 郭跃明. 马铃薯高产栽培技术 [J]. 农民致富之友, 2019(14): 32.
- [3] KUMAR N, BELHUMEUR P N, BISWAS A, et al. Leafsnap: A computer vision system for automatic plant species identification [C]// Proceedings of the European Conference on Computer Vision. Cham: Springer, 2012: 502–516.
- [4] 付雪婷, 王新鑫, 杨凡凡, 等. 基于卷积神经网络的植物品种识别研究 [J]. 南方农机, 2023, 54(17): 65–69.
- [5] 袁培森, 黎薇, 任守纲, 等. 基于卷积神经网络的菊花花型和品种识别 [J]. 农业工程学报, 2018, 34(5): 152–158.
- [6] 杨新宇, 冯全, 张建华, 等. 基于对比学习的植物叶片病害识别 [J]. 浙江农业学报, 2024, 36(1): 215–224.
- [7] NAEIMIE Z, PRINDLE K, 汪忠德. 机器学习和向机器学习 [J]. 世界地震译丛, 2019, 50(5): 442–452.
- [8] 邓立苗, 唐俊, 马文杰. 基于图像处理的玉米叶片特征提取与识别系统 [J]. 中国农机化学报, 2014, 35(6): 72–75.
- [9] 王丽君, 淮永建, 彭月橙. 基于叶片图像多特征融合的观叶植物种类识别 [J]. 北京林业大学学报, 2015, 37(1): 55–61.
- [10] 邹秋霞, 鄂鲁涛, 盛立冲. 基于 Android 手机和图像特征识别技术的植物叶片分类系统的研究 [J]. 安徽农业科学, 2015, 43(11): 367–369.
- [11] 陈文根, 李秀娟, 吴兰. 基于深度卷积网络的小麦品种识别研究 [J]. 粮食储藏, 2018, 47(2): 1–4.
- [12] 丁秋, 舒在习, 赵会义, 等. 基于小麦籽粒高光谱特征的品种鉴别研究 [J]. 粮食储藏, 2017, 46(2): 30–35.
- [13] 李鸿强, 孙红, 李民赞. 基于高光谱的马铃薯微型种薯分类检测 [J]. 分析测试学报, 2020, 39(11): 1421–1426.
- [14] 袁亮, 涂雪灌, 巨刚, 等. 基于机器视觉的番茄实时分级系统设计 [J]. 新疆大学学报(自然科学版), 2017, 34(1): 11–16.
- [15] 孟志超, 贺磊盈, 杜小强, 等. 基于 Enhanced VGG16 的油茶品种分类 [J]. 农业工程学报, 2022, 38(10): 176–181.
- [16] 韩斌, 曾松伟. 基于多特征融合和卷积神经网络的植物叶片识别 [J]. 计算机科学, 2021, 48(S1): 113–117.
- [17] 阳灵燕, 张红燕, 陈玉峰, 等. 机器学习在农作物品种识别中的应用研究进展 [J]. 中国农学通报, 2020, 36(30): 158–164.
- [18] 尹显明, 棘玉, 张日清, 等. 深度学习在基于叶片的油茶品种识别中的研究 [J]. 南京林业大学学报(自然科学版), 2023, 47(3): 29–36.
- [19] LIU Ze, LIN Yutong, CAO Yue, et al. Swin Transformer: Hierarchical vision transformer using shifted windows [C]// Proceedings of 2021 IEEE/CVF International Conference on Computer Vision (ICCV). Piscataway, NJ: IEEE, 2021: 9992–10002.
- [20] LIANG Jingyun, CAO Jiezhang, SUN Guolei, et al. SwinIR: Image restoration using Swin Transformer [C]// Proceedings of 2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW). Piscataway, NJ: IEEE, 2021: 1833–1844.
- [21] 田战胜, 刘立波. 基于改进 Transformer 的细粒度图像分类模型 [J]. 激光与光电子学进展, 2023, 60(2): 171–178.
- [22] 田永林, 王雨桐, 王建功, 等. 视觉 Transformer 研究的关键问题: 现状及展望 [J]. 自动化学报, 2022, 48(4): 957–979.
- [23] 扶兰兰, 黄昊, 王恒, 等. 基于 Swin Transformer 模型的玉米生长期分类 [J]. 农业工程学报, 2022, 38(14): 191–200.