草业科学
Pratacultural Science
中文核心期刊
中国科学引文数据库(CSCD) 核心期刊
英国CABI文摘数据库来源期刊
中国科技核心期刊
中国精品科技期刊

基于ResViT模型的城市草地草本植物智能分类

南喆 杨宏伟 杨梦鹭

Intelligent classification of urban grassland herbs based on ResViT modeling

NAN Zhe, YANG Hongwei, YANG Menglu

在线阅读 View online: https://doi.org/10.11829/j.issn.1001-0629.2024-0551

您可能感兴趣的其他文章

Articles you may be interested in

基于卷积神经网络的牧草种子图像识别

Convolutional neural network-based image recognition of forage seeds

草业科学. 2022, 39(11): 2338 https://doi.org/10.11829/j.issn.1001-0629.2022-0504

基于无人机遥感的多特征组矿区草本植物地上生物量反演

Inversion of herbage aboveground biomass in a multi-feature group mining area based on UAV remote sensing

草业科学. 2024, 41(1): 35 https://doi.org/10.11829/j.issn.1001-0629.2023-0005

基于CNN和SVM的地面高光谱遥感草地植物识别

Identification of grassland plants using hyperspectral remote sensing based on convolutional neural network and support vector machine

草业科学. 2023, 40(2): 394 https://doi.org/10.11829/j.issn.1001-0629.2022-0457

新疆草地生态健康智能监测网络体系构建

The systematic construction of a smart network for ecological health observation of grassland in Xinjiang

草业科学. 2023, 40(5): 1420 https://doi.org/10.11829/j.issn.1001-0629.2022-0658

基于GEE的香格里拉草地分类及其生物量遥感估算

Classification of Shangri-La grasslands based on Google Earth Engine and remote sensing estimation of their biomass

草业科学. 2024, 41(10): 2250 https://doi.org/10.11829/j.issn.1001-0629.2024-0069

草地资源调查与智能分析系统简介

Introducing a grassland resource survey and intelligent analysis system

草业科学. 2023, 40(8): 2171 https://doi.org/10.11829/j.issn.1001-0629.2021-0525

DOI: 10.11829/j.issn.1001-0629.2024-0551
南喆，杨宏伟，杨梦鹭. 基于 ResViT 模型的城市草地草本植物智能分类. 草业科学, 2025, 42(3): 628-637.
NAN Z, YANG H W, YANG M L. Intelligent classification of urban grassland herbs based on ResViT modeling. Pratacultural Science, 2025, 42(3): 628-637.

基于 ResViT 模型的城市草地草本植物智能分类

南 喆，杨宏伟，杨梦鹭

（蒙古高原生态学与资源利用教育部重点实验室 / 省部共建草地生态学国家重点实验室培育基地 /
内蒙古大学生态与环境学院, 内蒙古 呼和浩特 010021）

摘要：在干旱半干旱城市草地监测领域中，草本植物分类识别的用处及贡献不容小觑，但当前的深度学习模型在样本数据偏重且规模较小的任务中尚有不足之处。城市草地监测能够有效地评估草地的生长状况，并根据草本植物的分类，对当地生态系统的潜在危害提供判断信息。基于 ViT (Vision Transformer) 和 ResNet50 (深度残差网络) 构建了混合神经网络模型 ResViT。ResViT 在测试集准确率上最高，优于 AlexNet、ResNet50 和 VGG19 模型，在平均召回率和 F1 评分上也均优于 AlexNet、ResNet50 和 VGG19 模型。ResViT 的训练时间约是 VGG19 的一半。ResViT 在 16 分类任务中测试集达到了 95.45% 的准确率和 0.95 的 F1 评分。综上所述，ResViT 模型可以准确高效地完成草本植物分类的图像识别任务，比其他 3 种模型均有优势。其在偏重的小规模数据集上展现出优异的性能，显著降低了前期数据准备的成本，同时提升了训练效率，减少了训练时间。因此，ResViT 的建立为草本植物分类领域的研究提供了新视角，并有望在干旱半干旱城市草地监测的广泛应用中发挥重要作用。

关键词：草本植物分类；卷积神经网络；ResViT 模型；干旱半干旱；草地监测
文献标识码：A **文章编号：**1001-0629(2025)03-0628-10

Intelligent classification of urban grassland herbs based on ResViT modeling

NAN Zhe, YANG Hongwei, YANG Menglu

(Ministry of Education Key Laboratory of Ecology and Resource Use of the Mongolian Plateau / Inner Mongolia Key Laboratory of Grassland Ecology / School of Ecology and Environment, Inner Mongolia University, Hohhot 010021, Inner Mongolia, China)

Abstract: In the field of grassland monitoring in arid and semi-arid areas, the utility and contribution of the classification and recognition of herbaceous plants cannot be underestimated. However, current deep learning models continue to have shortcomings with respect to tasks involving substantial sample data and small scale. Urban grassland monitoring can effectively enable assessments of the growth status of grasslands and provide information for evaluating the potential harm to local ecosystems based on the classification of herbs. On the basis the ViT (Vision Transformer) and ResNet50 (Residual Network 50 layers) models, in this study, we constructed a hybrid neural network model referred to as ResViT, which is superior to the AlexNet, ResNet50, and VGG19 models in terms of test set accuracy, average recall rates, and F1 scores. ResViT can be trained within half the time needed for VGG19, and achieved an accuracy of 95.45% and an F1 score of 0.95 when used to perform a test set of 16 classification tasks. To summarize, the ResViT model can accurately and efficiently accomplish image recognition tasks for the classification of herbs and has distinct advantages compared with the AlexNet, ResNet50, and VGG19 models. It has shown excellent performance when used to assess heavily small-scale datasets, significantly reducing the cost of preliminary data preparation, whilst also improving training efficiency and reducing

收稿日期：2024-08-23 **接受日期：**2024-11-23
基金项目：内蒙古大学“一区两基地”超算能力建设项目 (21300-231510)；国家自然科学基金项目 (42165003)；内蒙古自治区科技重大专项“内蒙古风光资源开发对生态环境影响及其应对策略研究”
第一作者：南喆 (1999-)，男，吉林长春人，在读硕士生，主要从事人工智能在环境生态工程中的应用研究。E-mail: mrlanf@163.com
通信作者：杨宏伟 (1972-)，男，江苏徐州人，特聘研究员，博士，主要从事人工智能在环境生态工程中的应用研究。E-mail: hyangimu@163.com

training time. Consequently, the establishment of ResViT offers a novel perspective for research in the field of herb classification, and it is anticipated that this model will play key roles in extensive applications for grassland monitoring in arid and semi-arid areas.

Keywords: classification of herbs; convolutional neural network; ResViT model; arid and semi-arid; grassland monitoring

Corresponding author: YANG Hongwei　E-mail: hyangimu@163.com

干旱与半干旱地区正面临着前所未有的挑战，尤其是城市草地的维持与管理。这些地区因自然降水稀少和蒸发强烈，城市草地生态系统往往处于脆弱状态。随着城市化进程的加速，城市草地的作用日益凸显，它们不仅是“城市之肺”，更是城市生态系统的重要组成部分。在这一背景下，高效和准确地监测城市草地的变化，对于草地退化的早期预警和有效管理至关重要，而草本植物的识别可以为草地监测提供重要的判断信息。

传统的草本植物识别依赖于专家的目视判断和经验积累，这种方法不仅效率低下且成本高昂，难以满足大规模和高精度的监测需求。随着模式识别技术的发展，基于机器的图像分类方法得到广泛应用。这些方法基于特定的图像处理算法提取特征，并利用分类器对这些特征进行数学分析以得出分类结果[1]。然而，这种方法不仅需要丰富的专业知识，还要求具备一定的分类经验和实验研究。因此，在数据集的制作和特征提取方面都面临一定的挑战，缺乏普适性。

随着人工智能技术的发展，深度学习为草本植物识别提供了新的解决方案。近年来，基于 Transformer 模型[2] 的视觉模型，如 ViT (Vision Transformer)[3]，通过多头注意力机制获取全局信息，在图像分类任务中展现出卓越的性能。例如，刘金宇等[4] 首次将 ViT 应用于荒漠草原微斑块识别，有效提升了识别精度。王杨等[5] 改进了 ViT 模型缺乏局部归纳偏置和自注意力过于关注自身的问题，提出了更适合农作物病害识别的方法。此外，陈少真等[6] 将知识蒸馏与改进的 ViT 网络相结合，实现了高精度的花卉图像细粒度分类。Testagrose 等[7] 比较了 ViT 模型和两种不同的 EfficientNet 模型在自动沼泽草识别中的表现，结果显示使用 ViT 可以提高沼泽草识别的准确性。这些研究表明，ViT 及其衍生模型在图像识别领域具有强大的泛化能力和应用潜力。然而，ViT 在细节特征提取方面存在局限，对小数据集的泛化能力可能较弱，且计算量较大[8]。Lee 等[9] 结合卷积神经网络和 ViT，增强了单一模型的能力，在植物分类领域取得了较高的准确率。

与此同时，深度残差网络 (residual network 50 layers，ResNet50 [10]) 等卷积神经网络 (convolutional neural networks，CNNs) 架构也是计算机视觉领域的强大工具。Mukti 等[11] 利用 ResNet50 对 38 种健康叶子和病害叶子进行分类，取得了高达 99.80% 的准确率。Al-Gaashani 等[12] 提出的基于 ResNet50 的自注意力网络 (self-attention net，SANET) 在水稻 (*Oryza sativa*) 病害诊断中提高了效率和有效性。Liang 等[13] 采用 ResNet50 网络结合双路径注意力机制和随机池化方法，有效消除了非最大值的影响，使模型在番茄 (*Solanum lycopersicum*) 叶病分类任务中达到了 99.28% 的准确率。Du 等[14] 利用改进的 ResNet50 网络对棉 (*Gossypium*) 籽质量进行检测，其模型平均检测精度达到 97.23%，单张图像处理时间缩短至 0.11 s，满足了对棉籽品质进行快速准确检测的需求。Wang 等[15] 使用改进的 ResNet50 对玉米 (*Zea mays*) 病害进行了分类研究，分别在数据集和农田图像上的识别准确率达到 98.52% 和 97.83%，平均识别速度为 204 ms，为玉米田喷洒设备的研发提供了技术支持。这些研究证实了 ResNet50 在特征提取方面的强大能力，尤其在植物分类和识别任务中表现出色。然而，深度残差网络的计算效率较低，在训练数据不足时易发生过拟合，且需要大量标注数据[16]。为了解决这些问题，Abdulsalam 等[17]结合预训练的 ResNet50 和 YOLO v2 对象检测器，组成混合网络来改进农田杂草检测和分类的效果。Han 等[18] 提出了边缘自编码网络 (edge autoencoder network，E-A-Net) 算法，与原始的 ResNet50 网络结合，进一步提高了 ResNet50 的准确率。

尽管现有研究在草本植物识别方面取得了一定进展，但仍面临诸多挑战，如模型训练周期过长、对大规模数据集的依赖、识别精度提升空间有

限，以及在实际应用中数据集偏重的问题。本研究旨在结合 ViT 和 ResNet50 这两种先进的图像识别模型，构建一种针对干旱和半干旱城市草地监测的高效模型。该模型结合了 ViT 的全局注意力机制和 ResNet50 的强大特征提取能力，希望在保持高准确率的同时，提高模型的训练效率和适应性。希望此模型为干旱和半干旱城市草地的监测和管理提供新的解决方案，并促进类似区域的生态环境保护和改善。

1 数据与研究方法

1.1 数据

采用原创数据集进行试验，数据集由 1 100 张图片组成，拍摄于内蒙古自治区呼和浩特市内蒙古大学校园内，于北京时间 2023 年 4 月 20 日 12:30 开始拍摄，此时光线充足，拍摄效果良好。拍摄设备是佳能 EOS80D 相机搭配佳能 EF-S 18-200 mm f3.5-5.6 IS 变焦镜头。

选择的草本植物是在内蒙古大学校园内普遍存在的草本植物，从这些草本植物的数量来说，是校园内生态系统影响的主要因素。其次，选择的种类也是基于对照片内容的全面分析和领域专业知识，其在形态学上差异较为明显，方便在分类时和预测后通过目视的方式进行分类及验证结果的正确性。同时也希望通过这些具有代表性的草本植物来验证和测试 ResVit 模型的分类性能。最后，考虑到计算资源的限制和模型训练的时间成本，将类别控制在一个合理的范围内，以确保模型能够高效地训练并保持良好的泛化能力。

为了模拟真实情况，采用随机拍摄的方式进行照片采集。收集到的照片被整理并分为 16 类：A 类齿果酸模 (*Rumex crispus*) 108 张，B 类狭苞斑种草 (*Spodiopogon stenochloa*) 76 张，C 类早开堇菜 (*Viola prionantha*) 74 张，D 类扁秆荩草 (*Arthraxon compressus*) 68 张，E 类细叶荩草 (*A. hispidus*) 67 张，F 类金焰绣线菊 (*Spiraea bumalda* ‘Gold Flame’) 64 张，G 类宽叶荩草 (*Arthraxon macrophyllus*) 62 张，H 类珍珠梅 (*Sorbaria sorbifolia*) 61 张，I 类老鹳草 (*Geranium wilfordii*) 31 张，J 类黄花蒿 (*Artemisia annua*) 51 张，K 类马蔺 (*Iris lactea*) 48 张，L 类蜀葵 (*Alcea rosea*) 52 张，M 类北美独行菜 (*Lepidium virginicum*) 37 张，N 类蒲公英 (*Taraxacum mongolicum*) 49 张，O 类独行菜 (*Lepidium apetalum*) 28 张和其他类 (P 类) 224 张，总计 1 100 张照片。这些照片被汇总生成数据表单，作为原始数据集。其中，其他类照片是数据集中数量最多的类别 (优势类)，用于在数据集存在偏重的情况下测试模型的分类能力。A－O 类照片如图 1 所示。将原始数据集按照 8∶1∶1 的比例划分为训练集、验证集和测试集，分别包含 880 张、110 张和 110 张图片。

将 1 100 张照片进行数据增强，采用的增强方式包括缩放、旋转、平移和混合模式 (前 3 种模式的组合)。以 A 类照片为例，4 种增强方式的示例如图 2 所示。单张照片通过 1 种增强方式生成 3 张不同的增强照片 (图 3)，4 种增强方式最终生成 12 张增强照片。1 100 张照片总计生成 13 200 张增强照片。这些增强照片被汇总生成数据表单，作为增强数据集。增强数据集的目的是通过多样化数据来提升模型的泛化能力，使其能够更好地识别不同形式和角度的同类照片。为了保证测试集的严谨性，防止产生泄题，导致模型提前学习到要测试的内容，将原数据集中划分出来的测试集部分和测试集产生的增强数据集部分进行隔离，使其不加入到训练集中。

1.2 研究方法

1.2.1 ResNet50 网络

ResNet 在 2015 年被提出[19]，并在 ImageNet 比赛的分类任务中获得了第 1 名。ResNet 是一种经典的卷积神经网络，它的核心思想是引入残差学习，解决深层网络训练中的梯度消失和梯度爆炸问题。通过在网络中添加直接的前向连接 (即残差连接)，ResNet 使得梯度可以直接通过这些连接流动，从而支持训练更深的网络。

本研究主要参考了 ResNet 系列模型中的 ResNet50[11] 模型，该模型在植物分类方面具有出色的特征提取能力。ResNet50 的网络结构主要由卷积块 (convolutional block) 和恒等块 (identity block) 两个模块组合而成。其中，卷积块由 3 层卷积层构成。第 1 层使用 1 × 1 卷积核，主要用于降低维度，有时也用于改变维度；第 2 层使用 3 × 3 卷积核，进行特征提取，并设置步长为 2 以实现下采样；第 3 层使用 1 × 1 卷积核，用于恢复维度。而恒等块也

图 1　数据集中 A—O 类的图片示例

Figure 1　Illustration of classes A—O in the dataset

由 3 层卷积层组成。第 1 层同样使用 1 × 1 卷积核，主要用于降低维度；第 2 层使用 3 × 3 卷积核，是主要的特征提取层；第 3 层使用 1 × 1 卷积核，用于恢复维度。这样的设计使得整体网络更易于训练，并且相较于同系列的其他模型。ResNet50 拥有更深的网络结构，因此在图像分类任务上表现优异。本研究主要使用了 ResNet50 模型的特征提取部分，舍弃了后半部分的全连接层。为了更好地反映其来源和特性，将该特征提取部分命名为 Res50Ftr。

1.2.2　ViT 模型及其变种

1) ViT 模型。

ViT (vision transformer) 模型[3] 是 Google 团队于

图 2 单张原图及其 4 种增强方式增强结果演示图

Figure 2 A representative original image and the results in four different augmentation methods

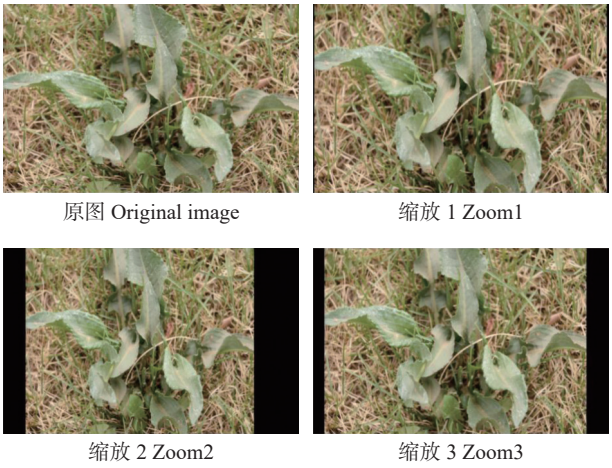

图 3 单张原图及其 3 张缩放增强图示例

Figure 3 A representative original image and three scaled augmentation images

2020 年提出的一种将 Transformer [2] 应用于图像分类的模型。ViT 的主要特点是简单且高效。它通过将图像分割成小块 (patches)，并将每个 patch 视为序列的输入，从而将图像分类任务转化为序列处理任务。这种方法不仅使模型能够捕捉到图像中的空间关系，还允许 Transformer 模型直接在图像数据上工作，而不依赖于卷积神经网络的归纳偏差。在大规模数据集上训练时，ViT 模型的性能可以达到或超过传统的卷积神经网络。由于其自注意力机制，ViT 在处理不同尺寸和比例的图像时展现出较好的灵活性和泛化能力。

2) ViT 模型变种。

ViT 模型在图像分类方面表现出色，但也存在一些局限性，如需要大规模数据集进行训练。为了解决这一问题，研究人员开发了专为小数据集优化的衍生 ViT 模型[20]，称为 ViTSmall。这款模型在保持良好分类性能的同时，降低了对数据集规模的要求，从而减少了各领域从业者在前期数据准备方面的难度，并节省了大量时间。

ViTSmall 模型在设计上与原始的 ViT 模型有所不同，它采用了局部化自注意力 (locally sensitive attention, LSA) 和平移分块嵌入 (shifted patch tokenization，SPT) 类，这两个类在参数数量和计算效率方面进行了优化。LSA 类是一种改良的注意力机制，旨在增强模型对局部特征的敏感性，通过对点积结果应用温度缩放来提高注意力机制的稳定性和效果。SPT 类则通过对图像进行空间位移来丰富 patch 的表示，这不仅简化了图像的分块处理，还增强了模型对位置信息的感知能力。

3) ResViT 模型。

本研究将 Res50Ftr 和 ViTSmall 模型结合，形成一个新的模型，命名为 ResViT。在这个组合中，Res50Ftr 模型负责提取全局特征，然后将这些特征输入到 ViTSmall 模型中进行局部特征的提取。这样，整个模型的特征提取能力得到了最大化，增强了模型的学习能力和分类准确性。即使在小规模数据集上，ResViT 仍能保持对图像分类的高精度。因此，ResViT 不仅弥补了原始 ViT 在数据集规模方面的不足，还确保了图像分类的准确性。

ResViT 模型的输入是像素为 244 × 244 × 3 的

RGB 图像，范围为 0～255 像素。在训练前，需要对图像进行预处理，将其缩放至像素为 244 × 244 的大小。图像进入模型后，首先通过 Res50Ftr 部分，该部分包括一个卷积核大小为 7 × 7、步长为 2 的卷积层，接着是 4 个瓶颈卷积块 (bottleneck block)。每个瓶颈卷积块由多个 1× 1 和 3× 3 的卷积层以及 ReLU 激活函数组成，旨在减少网络参数并提高模型性能。然后，通过全局池化层生成一个像素为 7 × 7 × 2 048 的特征图。接下来，这个特征图输入到 SPT 部分。SPT 部分对特征图进行上下左右 4 个方向的平移，目的是增强局部上下文信息。通过这种平移操作，每个图像块可以获取更多的邻近信息，从而提高模型对局部特征的捕捉能力。然后，将平移后的特征图与原特征图进行拼接，生成一个像素为 49 × 1 024 的新特征图，这个新特征图将输入到 ViTSmall 中。在 ViTSmall 中，新特征图首先被分割，并编入位置信息。在 Transformer 中，通过 LSA 机制强调局部特征并限制注意力的计算范围，以减少计算复杂度和过拟合的风险。最终，模型输出分类结果。整体模型流程如图 4 所示。

2　实验结果

本研究除了测试 ResViT 模型外，还选择 AlexNet、ResNet50 和 VGG19 作为对比实验，以验

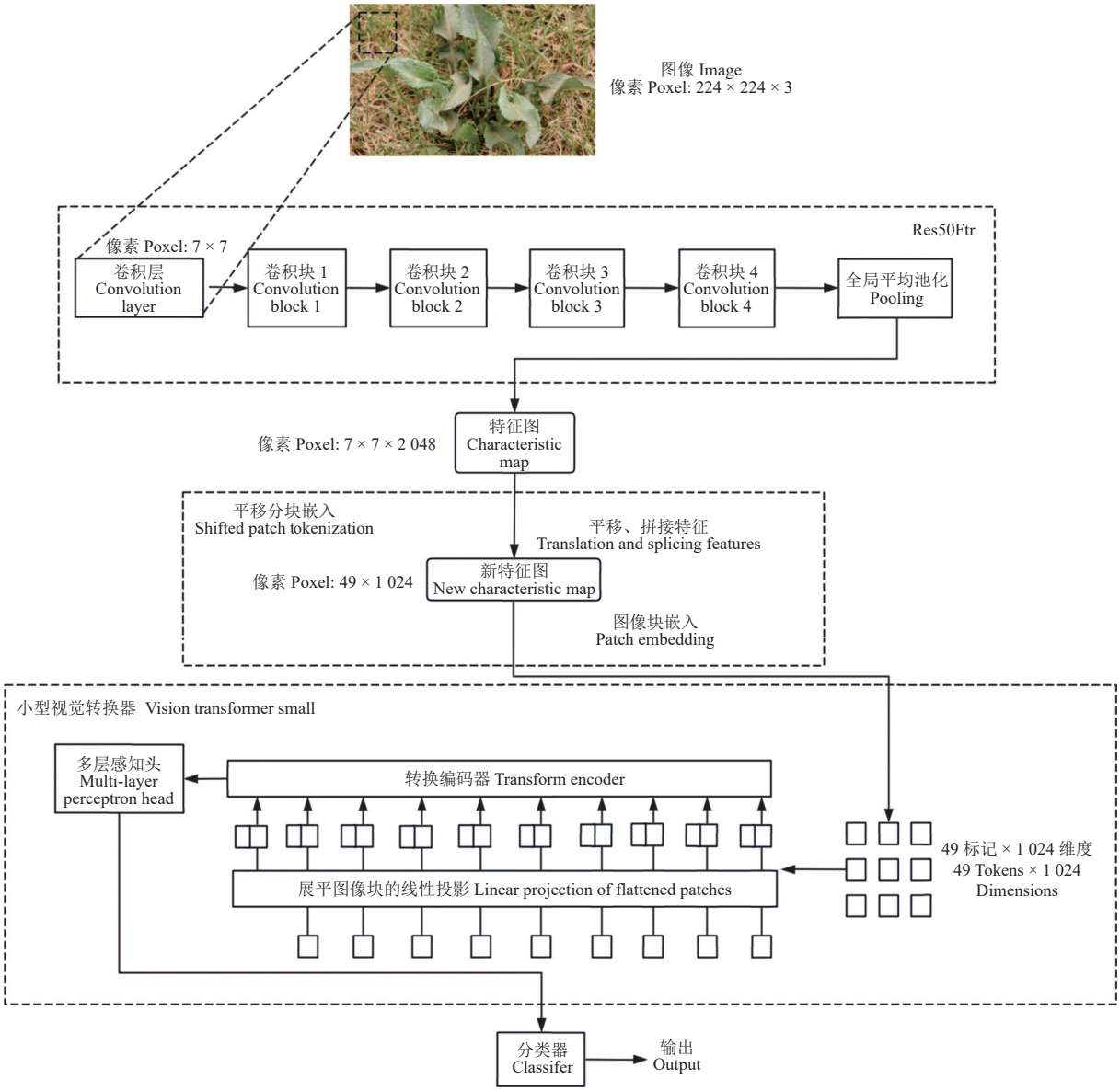

图 4　ResViT 模型流程

Figure 4　A flow chart of the ResViT model

证 ResViT 模型在草地分类识别中的准确性和训练时长是否具有优势。在实验中，将预处理后的 224 × 224 像素大小的 RGB 图像输入到 ResViT 模型，并按照 **1.2.3** 节中描述的流程开展实验。

2.1 参数设置

图像块的大小 (patch_size) 设置为 7，类别数 (num_classes) 为 16，通道数 (channels) 为 3，模型隐藏维度 (dim) 为 1 024，Transformer 块的数量 (depth) 为 6，注意力头数 (heads) 为 8，多层感知器 (multi-layer perceptron，MLP) 的隐藏层维度 (mlp_dim) 为 2 048，训练轮数 (epoch) 设置为 70。使用了 Adam 优化器，初始学习率为 0.003，并采用 CrossEntropyLoss 函数作为损失函数。

2.2 评价指标

计算准确率 (Accuracy)：

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN}。$$

式中：TP表示实际为正类且被模型正确预测为正类的样本数，TN 表示实际为负类且被模型正确预测为负类的样本数，FP 表示实际为负类但被模型错误预测为正类的样本数，FN表示实际为正类但被模型错误预测为负类的样本数。

为更好地评估模型性能，在训练集和验证集评价过程中使用了平均准确率 (Average_accuracy)。

$$Average\_accuracy = \frac{Accuracy}{Epoch}。$$

式中：$Epoch$ 为实验轮数，本研究中各个模型训练集和验证集的实验轮数均为 70 轮，测试集轮数为 1 轮，即测试集的准确率应等同于其平均准确率。通过平均准确率更可以反映出模型的学习效率。

由于本试验涉及多分类问题，因此使用的召回率 (Recall) 计算方法是先计算单个类别的召回率，然后取这些召回率的平均值作为最终的召回率。单个类别的召回率的计算公式为：

$$Recall_i = \frac{TP_i}{TP_i+FN_i}。$$

式中：$Recall_i$表示第i类的召回率；TP_i表示实际为第i类且被模型正确预测为第 i类的样本数；FN_i表示实际为第 i 类但被模型错误预测为其他类别的样本数。平均召回率 (Recall_mean) 的计算公式为：

$$Recall\_mean = \frac{\sum_{i=1}^{n} Recall_i}{num\_classes}。$$

本研究采用宏平均 (Macro-average) F1评分的计算方法，首先计算每个类别的精确率 (Precision) 和召回率。每个类别的精确率计算公式为：

$$Precision_i = \frac{TP_i}{TP_i+FP_i}。$$

式中：$Precision_i$表示第 i 类的精确率；FP_i表示实际不为第 i 类但被模型错误预测为第 i 类的样本数。然后计算平均精确率 (Precision_mean)：

$$Precision\_mean = \frac{\sum_{i=1}^{n} Precision_i}{num\_classes}；$$

$$F1 = 2 \times \frac{Precision\_mean \times Recall\_mean}{Precision\_mean + Recall\_mean}。$$

将经过 70 轮训练后轮次模型达到最优时的参数作为最终的模型参数。使用这些参数在测试集上进行测试，并将测试结果作为模型的最终准确率标准。

2.3 实验环境

本实验在 Linux 操作系统 (CentOS 7.6) 上进行，使用 Python 3.8.8 和 PyTorch 2.2.2 深度学习框架。硬件配置为：Intel Xeon Gold 6254 处理器 (2 颗，18 核，3.1 GHz) 和 192 GB 内存。

3 结果

如表 1 所列，经过 70 轮训练后，ResViT 模型训练集和验证集的平均准确率均为最高，分别达到了 97.02% 和 97.25%，其测试集的准确率也为 4 个模型中最高 (95.45%)。由此可见，ResViT 模型在 16 类分类的任务中仍然可以出色的完成任务。此外，AlexNet 模型的准确率最低，在训练集和验证集上的平均准确率仅有 78.02% 和 81.05%。ResNet50 模型的准确率排名第二，分别为 96.78% 和 96.97%。在召回率方面，ResViT 模型的平均召回率为 96.32%，显示出它在检测目标类别时的优秀能力。相比之下，ResNet50 的召回率为 93.75%，与 ResViT 模型的召回率相比略逊一筹。关于 F1 评分，ResViT 和 ResNet50 模型分别达到了 0.95 和 0.93，表明它们在精确度和召回率之间取得了良好的平衡，能够在保持高精度的同时减少误报和漏报的概率。这也表明即使数据集中

表 1　4 种模型实验结果

Table 1　Experimental results obtained using the four models

模型 Model	训练集平均准确率 Training set average accuracy/%	验证集平均准确率 Verification set average accuracy/%	测试集准确率 Test set accuracy/%	平均召回率 Average recall rate/%	F1评分 F1 rating	训练时长 Training duration
AlexNet	78.20	81.05	19.09	6.25	0.02	66 h 28 min
ResNet50	96.78	96.97	92.73	93.75	0.93	135 h 28 min
VGG19	91.78	94.22	80.91	79.82	0.81	300 h 56 min
ResViT	97.02	97.25	95.45	96.32	0.95	158 h 09 min

可能存在数据偏差，仍能准确完成分类任务。VGG19 模型的 F1 评分也较高，但和 ResViT 与 ResNet50 相比仍然存在差距。在训练时长方面，ResViT 稍比 ResNet50 用时长，但只有 VGG19 的一半。AlexNet 用时最短。总体上，ResViT 模型在各项指标上表现均突出，显示了较高的实用价值。虽然在训练时长方面不是最快的，但从准确率等其他结果来看，其优势比较全面。

如图 5 所示，ResViT 模型能基本准确地区分各类样本，仅有少量样本出现错误。如图 6 所示，ResViT 模型的表现最佳，可以快速高效的学习，准确率在 10 轮以内就超过了 90%，并在第 20 轮左右趋于稳定；ResNet50 模型略晚于 ResViT 模型 90% 以上的准确率，且波动始终较大，并不平稳；VGG19 模型在第 10 轮左右达到 90% 的准确率；而 AlexNet 模型仅有少数几次在验证集上准确率在 90% 以上，且波动很大，尤其是在第 67 到 70 轮，训练损失增大且测试集准确率下降。这种现象通常是过拟合的特征。因此推断其可能出现了过拟合的现象。综上所述，ResViT 模型在验证集上的表现优于其他 3 种模型，其次是 ResNet50，表现最差的是 AlexNet 模型。

综上所述，ResViT 模型在草本植物识别分类任务中表现出色，能够达到高准确度，并在短时间内取得良好的训练效果。因此，本研究构建的 ResViT 模型在小数据集上能够出色地完成分类任务，同时展现出良好的鲁棒性和处理负例的能力。

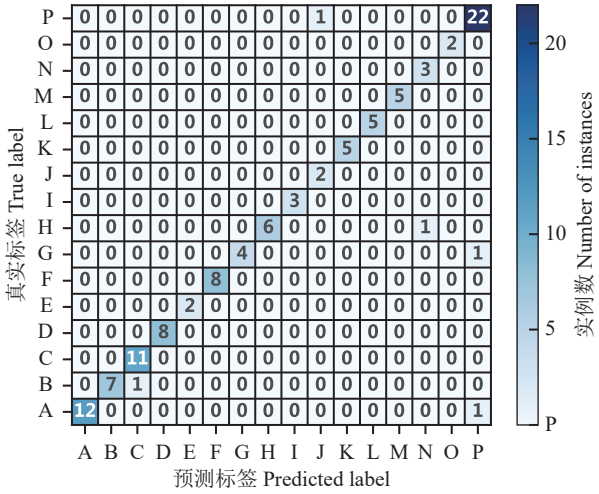

图 5　ResViT 模型的混淆矩阵

Figure 5　Confusion matrix for the mixed model

A－O 类同图 1，P 为其他类。

A－O are similar to Figure 1, P is other classes.

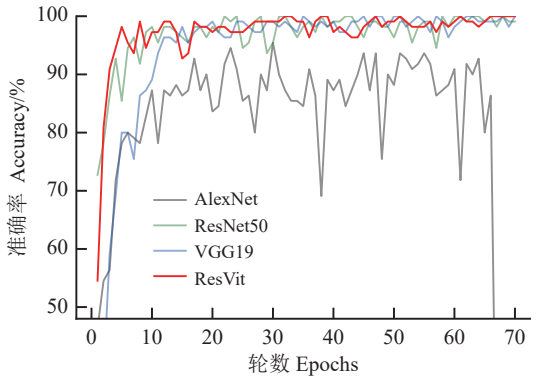

图 6　4 种模型的准确率曲线

Figure 6　Accuracy of the four models when used to assess the test set

4　讨论

Res50Ftr 是一种深度卷积神经网络，专注于提取图像的局部特征。通过训练，该模型能够有效地提取图像的边缘信息，并利用卷积块生成高质量的局部特征图。ViTSmall 则采用了自注意力机制，通过平移操作，确保即使特征图被分割后仍能保留强大的全局信息。这一机制不仅能够联系附近的上下文，还能捕捉图像中的长距离特征信息。通过独特的 Transformer 层，ViTSmall 可以学习更加全面的特征和高级别的语义信息。ResViT 模型将 Res50Ftr

和 ViTSmall 的优点结合在一起，既保留了局部特征信息，又不忽视全局特征。因此，它具备强大的学习能力，继承了两个模型的优秀特性。借助 ViTSmall 的多头自注意力机制，ResViT 能够处理不同的特征子空间，增强模型的表达力。优秀的特征提取能力和强大的模型表达力使得 ResViT 模型在仅需 30 轮或更少的训练后即可达到一个稳定的状态。因此，ResViT 模型展现了优异的性能。

本研究使用的数据集模拟了用户使用手机拍摄的视角和效果，用户可以将训练集中的图像按类别分类后直接输入模型，模型会自动提取和学习相应特征，在测试集上实现准确预测。另一方面，ResViT 即使在偏重的小规模数据集上也能保持高效的学习能力和分类准确度，克服了现有模型依赖中、大规模数据集的缺陷。因此，ResViT 显著降低了前期数据准备的成本。同时显著提升了训练效率，减少了训练时间。ResViT 的建立弥补了草类分类领域研究的不足，并有望在干旱和半干旱城市草地监测中发挥重要作用。

本研究结果还存在一些有待进一步研究的问题。首先，ResViT 模型的最大分类能力尚未确定，目前的 16 类分类是受限于现有设备条件的上限。如果设备性能提升，可以进一步增加分类数量，以测试 ResViT 模型的分类能力极限。其次，本研究使用的是强优势类占主导的数据集，因此模型在强优势类上的学习能力较为准确。然而，研究结果虽然在弱势类上表现完美，但未能充分测试强势类数据对弱势类学习的影响，可能导致弱势类的学习不足或强势类的学习过剩。此外，由于许多草种的叶片性状和颜色相似，难以通过肉眼区分，需要借助触感、嗅觉等。这些因素是否能够与现有模型结合，以提高特征提取的精度，实现更高水平的草类分类，仍需进一步研究验证。

5 结论

为了在干旱和半干旱城市草地监测任务中精准快速地识别草的种类，结合 ViT 和 ResNet50 的优势构建了混合神经网络模型 ResViT。得到以下结论：1) ResViT 模型在准确率上超过了 AlexNet、ResNet50 和 VGG19 模型。2) ResViT 模型在平均召回率和 F1 评分上优于 AlexNet 和 ResNet50 模型。3) ResViT 模型的训练时间是 VGG19 模型的一半。4) ResViT 模型在 16 分类任务中测试集达到了 95.45% 的准确率和 0.95 的 F1 评分。5) ResViT 模型在偏重的小规模数据集上展现了优异的性能，显著降低了前期数据准备的成本，同时提升了训练效率减少了训练时间。

参考文献 References:

[1] DALAL N, TRIGGS B. Histograms of Oriented Gradients for Human Detection. San Diego, CA, USA: 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2005.

[2] VASWANI A, SHAZEER N, PARMAR N, USZKOREIT J, JONES L, N. GOMEZ A, KAISER L, POLOSUKHIN I. Attention is all you need. Advances in Neural Information Processing Systems, 2017, 30: 6000-6010.

[3] DOSOVITSKIY A, BEYER L, KOLESNIKOV A, WEISSENBORN D, ZHAI X, UNTERTHINER T, DEHGHANI M, MINDERER M, HEIGOLD G, GELLY S, USZKOREIT J, HOULSBY N. An image is worth 16 × 16 words: Transformers for image recognition at scale. arXiv preprint, arXiv: 2010. 11929, 2020.

[4] 刘金宇, 杜健民. 基于视觉 Transformer 的荒漠草原微斑块识别. 信息技术与信息化, 2023, 30(12): 200-203.
LIU J Y, DU J M. Desert steppe micro-patch recognition based on vision Transformer. Information Technology and Informatization, 2023, 30(12): 200-203.

[5] 王杨, 李迎春, 许佳炜, 王傲, 马唱, 宋世佳, 谢帆, 赵传信, 胡明. 基于改进 Vision Transformer 网络的农作物病害识别方法. 小型微型计算机系统, 2024, 45(4): 887-893.
WANG Y, LI Y C, XU J W, WANG A, MA C SONG S J, XIE F, ZHAO C X, HU M. Crop disease recognition method based on improved vision transformer network. Journal of Chinese Computer Systems, 2024, 45(4): 887-893.

[6] 陈少真, 叶武剑, 刘怡俊. 基于知识蒸馏与改进 ViT 网络的花卉图像细粒度分类. 光电子·激光, 2024, 35(1): 29-40.
CHEN S Z , YE W J , LIU Y J. Flower fine-grained images classification based on the knowledge distillation and improved vision transformer. Journal of Optoelectronics·Laser, 2024, 35(1): 29-40.

[7] TESTAGROSE C, SHABBIR M, WEAVER B, LIU X. Comparative study between vision transformer and efficientnet on marsh grass classification. The International FLAIRS Conference Proceedings, 2023, 36: 1-5.
[8] RAGHU M, UNTERTHINER T, KORNBLITH S, ZHANG C, DOSOVITSKIY A. Do vision transformers see like convolutional neural networks?. Advances in Neural Information Processing Systems, 2021, 34: 12116-12128.
[9] LEE C P, LIM K M, SONG Y X, ALQAHTANI A. Plant-CNN-ViT: Plant classification with ensemble of convolutional neural networks and vision transformer. Plants, 2023, 12(14): 2642.
[10] MASCARENHAS S, AGARWAL M. A comparison between VGG16, VGG19 and ResNet50 architecture frameworks for Image Classification. //2021 International Conference on Disruptive Technologies for Multi-Disciplinary Research and Applications (CENTCON). Bengaluru, India: IEEE, 2021: 96-99.
[11] MUKTI I Z, BISWAS D. Transfer learning based plant diseases detection using ResNet50. //2019 4th International Conference on Electrical Information and Communication Technology (EICT). Khulna, Bangladesh: IEEE, 2019: 1-6.
[12] AL-GAASHANI M S A M, SAMEE N A, ALNASHWAN R, KHAYYAT M, MUTHANNA M S A. Using a ResNet50 with a kernel attention mechanism for rice disease diagnosis. Life, 2023, 13(6): 1277.
[13] LIANG J, JIANG W. A ResNet50-DPA model for tomato leaf disease identification. Frontiers in Plant Science, 2023, 14: 1258658.
[14] DU X, SI L, LI P, YUN Z. A method for detecting the quality of cotton seeds based on an improved ResNet50 model. PLoS One, 2023, 18(2): e0273057.
[15] WANG G, YU H, SUI Y. Research on maize disease recognition method based on improved ResNet50. Mobile Information Systems, 2021, DOI:10.1155/2021/9110866.
[16] SHAFIQ M, GU Z. Deep residual learning for image recognition: A survey. Applied Sciences, 2022, 12(18): 8972.
[17] ABDULSALAM M, AOUF N. Deep weed detector/classifier network for precision agriculture. //2020 28th Mediterranean Conference on Control and Automation (MED). Saint-Raphaël, France: IEEE, 2020: 1087-1092.
[18] HAN D, TIAN M, GONG C, ZHANG S, JI Y, DU X, WEI Y, CHEN L. Image classification of forage grasses on Etuoke Banner using edge autoencoder network. PLoS One, 2022, 17(6): e0259783.
[19] LI S, JIAO J, HAN Y, WEISSMAN T. Demystifying ResNet. arXiv preprint, arXiv: 1611. 01186, 2016.
[20] LEE S H, LEE S, SONG B C. Vision transformer for small-size datasets. arXiv preprint, arXiv: 2112. 13492, 2021.

（责任编辑　王芳）

2025 年第 3 期《草业科学》审稿专家

蔡延江	曹文侠	常生华	陈　骥	陈丽娟	陈　焘	陈有军	崔治家
董　瑞	范树高	方强恩	房建东	冯琦胜	干友民	郭晓军	韩云华
侯扶江	胡　涛	黄晓东	寇建村	李　飞	李君风	李隆云	李万宏
李　渊	李志刚	李　舟	刘建斌	刘文献	刘兴元	刘　秀	刘永杰
刘玉冰	娄燕宏	鲁旭阳	罗栋梁	马红媛	马　啸	马欣荣	马亚玲
满都呼	毛培胜	秦立刚	尚占环	孙洪仁	孙　建	孙丽娟	王虎成
王金牛	王丽佳	王　强	王晓波	武俊喜	谢　燕	徐浩杰	闫世程
杨培志	杨宪龙	于应文	袁　帅	张宝成	张　程	张红瑞	张金鑫
张　林	张美玲	张晓东	张　勇	朱剑霄			

承蒙以上专家对《草业科学》期刊稿件的审阅，特此表示衷心的感谢！