

基于改进 Vision Transformer 的水稻叶片 病害图像识别^{*}

朱周华 周怡纳 侯智杰 田成源

(西安科技大学通信与信息工程学院 西安 710061)

摘 要: 水稻叶片病害智能识别在现代农业生产中具有重要意义。针对传统 Vision Transformer 网络缺乏归纳偏置,难以有效捕捉图像局部细节特征的问题,提出了一种改进的 Vision Transformer 模型。该模型通过引入内在归纳偏置,增强了对多尺度上下文以及局部与全局依赖关系的建模能力,同时降低了对大规模数据集的需求。此外, Vision Transformer 中的多层感知器模块被 Kolmogorov-Arnold 网络结构取代,从而提升了模型对复杂特征的提取能力和可解释性。实验结果表明,所提模型在水稻叶片病害识别任务中取得了优异的性能,识别准确率达到 98.62%,较原始 ViT 模型提升了 6.2%,显著提高了对水稻叶片病害的识别性能。

关键词: 水稻叶片病害;图像识别;Vision Transformer 网络;归纳偏置;局部特征

中图分类号: TN919.8 **文献标识码:** A **国家标准学科分类代码:** 510.50

Rice leaf disease image recognition based on improved Vision Transformer

Zhu Zhouhua Zhou Yina Hou Zhijie Tian Chengyuan

(College of Communication and Information Technology, Xi'an University of Science and Technology, Xi'an 710061, China)

Abstract: Intelligent recognition of rice leaf diseases is of great significance in modern agricultural production. Focused on the issue that the traditional Vision Transformer network lacks inductive bias and is difficult to effectively capture the local detail features of the image, an improved Vision Transformer model was proposed. This model's ability to model multi-scale context as well as local and global dependencies was enhanced by introducing intrinsic inductive bias, while reduced the need for large-scale datasets. In addition, the multi-layer perceptron module in the Vision Transformer was replaced by the Kolmogorov-Arnold networks structure, thereby improving the model's ability to extract complex features and interpretability. Experimental results show that the proposed model achieved excellent performance in the rice leaf disease recognition task, with an accuracy of 98.62%, which was 6.2% higher than the original ViT model, effectively improving the recognition performance of rice leaf diseases.

Keywords: rice leaf disease; image recognition; Vision Transformer network; generalized bias; local feature

0 引 言

水稻是全球重要的粮食作物之一,也是我国以及世界许多国家的主要口粮^[1]。我国是世界上水稻种植量最多的国家之一,水稻主要分布在东北平原、长江上、中、下游地区及东南沿海五大优势产区^[2]。水稻生产在国民经济中具有重要地位,根据国家统计局的数据显示,截至 2023 年,我国水稻种植面积达 2 894.9 万公顷,稻谷总产量达到 20 660 万吨。水稻是一种极具经济价值的农作物,但在种植过程中,极易受到细菌、真菌和病毒等多种病害的侵扰,从而导致产量下降,影响经济效益^[3]。因此,及时发现和治疗水稻病害

对于保障水稻产量和品质至关重要。水稻病害大多表现在叶片上,而叶片长度较长并且形状较窄,病害不容易观察,尤其是当不同的病害呈现出相似的症状或同一种病害表现出多样化的病斑时^[4],由于许多种植者对水稻植株病害的认知程度不足,难以准确识别病害类型,人工肉眼识别通常会出现高概率的误判。有时,农民还会寻求专家帮助,但这种方式耗时费力且效率低下。

自 2006 年深度学习概念提出以来,深度学习技术逐渐在不同领域得到广泛应用,包括农业领域的病虫害检测和识别、农产品分级、农作物科学监管和动物识别等方面^[5]。与人工方法相比,基于深度学习的方法可以自动化农作物

病害识别,减少对人为干扰的需求,提高流程的精度和效率^[6]。卷积神经网络(convolutional neural networks, CNN)作为深度学习中的重要技术之一,因其在图像特征提取和数据分析方面的优秀表现,被广泛应用于农作物叶片病害的识别研究中^[7]。CNN 通过从输入的水稻叶片病害图像中自动提取病害的相关特征^[8]。在水稻叶片病害识别领域,Kumar 等^[9]提出了一种基于 YOLOv5 的检测方法,将模型的主干网络替换为 DenseNet-201,并融合双向特征注意金字塔网络。虽然该模型可以提取不同尺度的图像特征信息,但其检测的准确率不够高。Ni 等^[10]提出了一种水稻叶片病害分类模型 RepVGG_ECA,在 RepVGG 网络的 Block 层和 Head 层后添加高效通道注意力(efficient channel attention, ECA)模块,以提高模型对病害的特征提取能力,但该模型对 Hispa 病害的分类效果较差。由于 CNN 依赖于卷积操作,具有局部感受野,导致模型只能聚焦于图像局部特征的提取,而忽略全局特征信息,不能充分的利用全局上下文信息^[11]。因此,对于具有复杂背景信息的水稻叶片病害图像,和在不同光线和视觉影响下的病害图像,CNN 可能很难充分捕捉到图像的全局信息。

随着自然语言处理(natural language processing, NLP)领域中 Transformer^[12]模型的逐渐成熟,研究人员开始将其应用在计算机视觉(computer vision, CV)领域。2020 年,谷歌团队提出了 Vision Transformer(ViT)网络模型,主要将其应用在图像分类任务中。Vision Transformer 中的自注意力机制能够在全局范围内捕捉图像特征信息之间的关联,有效处理图像数据中的长距离依赖关系,弥补了 CNN 在全局信息提取方面的不足。例如, Li 等^[13]提出了一种植物病害识别模型(PMVT),该模型将 MobileNet 与 Vision Transformer 相结合,使用倒置残差结构替换了模型中的卷积块,并将卷积注意力模块(convolutional block attention module, CBAM)集成到模型中,该模型同时在小麦、水稻和咖啡等多个农业数据集上进行了实验,但在水稻叶片病害数据集上的准确率较低。He 等^[14]提出了一种水稻叶片病害识别模型 ECA-ViT,在 Vision Transformer 网络中添加 ECA 模块,弥补了 ViT 缺少对图像局部特征信息提取的能力,但模型的准确率还有待提高。Wu 等^[15]提出了一种基于 Vision Transformer 的多粒度特征提取模型进行番茄叶片病害识别任务,通过学习不同尺度的图像特征信息,使模型能够更精细地识别相似病害之间的细微差异,但模型的识别准确率不够高。Sharma 等^[16]提出了一种使用迁移学习的 Vision Transformer 模型,对香蕉叶片中缺乏营养物质的种类进行分类识别,通过预训练模型对 ViT 网络进行超参数的调节和网络层的冻结,但模型对图像局部特征的提取能力不强。由于水稻叶片的形状较窄较细,病害往往体现在叶片上的一小块区域,病害种类不容易识别,因此与其他的图像识别任务相比,模型需要更高效的特征提取能力。尽管 ViT 在 CV 领域展现出优异的效果,

但其对大规模训练数据的需求和产生大量计算成本^[17]的问题仍是 ViT 的主要缺点之一,其中一个原因是 ViT 缺乏 CNN 中固有的归纳偏置^[18],对局部特征信息的提取能力和多尺度信息处理方面表现较弱。

针对以上问题,本文提出了一种改进的 Vision Transformer 水稻叶片病害识别算法,对 5 种水稻叶片图像(白叶枯病、褐斑病、稻瘟病、东格鲁病和健康叶片)进行识别。将 CNN 的归纳偏置(局部性和尺度不变性)引入到 ViT 模型中,并将 Kolmogorov-Arnold 网络^[19](Kolmogorov-Arnold networks, KAN)应用到模型中,提高模型对水稻叶片病害识别的准确性。

1 水稻叶片病害识别算法

1.1 改进的 Vision Transformer 算法

本文在 Vision Transformer 的基础上进行了两方面的结构改进。首先,采用压缩单元(compression unit)模块替换 ViT 中的图像块编码(patch embedding)模块,引入 CNN 的固有归纳偏置,即局部性和尺度不变性^[20]。局部性是指模型可以提取输入图像相邻像素之间的局部特征,尺度不变性则是需要多尺度特征来有效地表示不同尺度的对象。相较于原始 ViT 模型中将图像分割和展平为 N 个一维向量输入编码层的方式^[21],本文提出的模型通过压缩单元模块增强了对水稻叶片病害局部特征和多尺度信息的提取能力,并缓解了模型对大规模数据集的依赖。其次,使用 KAN 模块替代 ViT 模型中的 MLP 模块,构建 kansformer 编码器(kansformer encoder)结构,并将 KAN 应用于最终的分头,以此提高模型对复杂特征的学习能力和增强模型的可解释性。为了更直观地分析比较改进前后模型的网络结构,原始的 Vision Transformer 网络模型和改进后的 Vision Transformer 网络模型分别如图 1 和 2 所示。

如图 2 所示,本文提出的改进 ViT 模型,首先将水稻叶片病害图像输送到 3 个依次堆叠的压缩单元模块中,对输入图像的空间维度分别压缩 4 倍、2 倍、2 倍,使得图像的特征信息具有尺度不变性和局部性。随后结合位置编码(position embedding)与分类标签(class token)将输入的特征序列送入 kansformer encoder 进行编码。最后使用 KAN 网络对水稻叶片病害进行分类,得到输出结果。

1.2 压缩单元(compression unit)模块

在原始的 Vision Transformer 模型中,输入图像(224×224 像素)首先被分割为 196 个大小为 16×16 的图像块(patch),相当于 Transformer 模型的输入序列长度为 196。接着,模型通过线性映射将每个 Patch 转换为一维向量,以便输入到 Transformer 编码器中。随后,引入分类标签与这些向量进行拼接,作为模型的输入序列。最后,这些拼接后的向量通过位置编码进行处理,以保留图像中各 patch 的位置信息,并将其送入到后续的 Transformer 编码层进行特征提取。

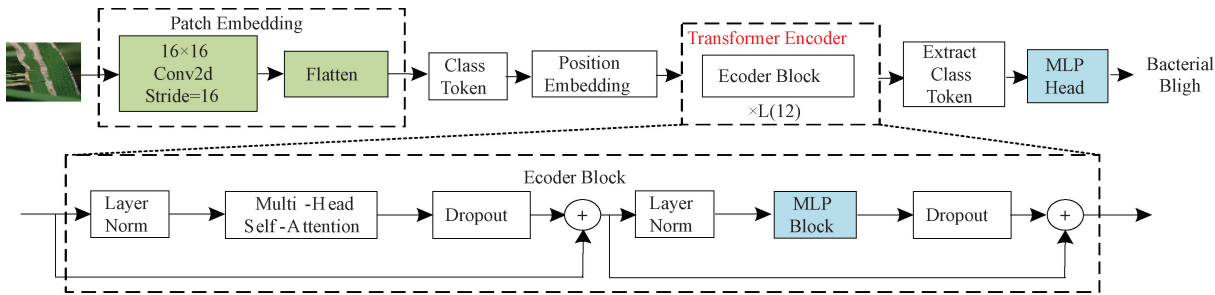


图 1 原始 Vision Transformer 网络结构

Fig.1 Original Vision Transformer network structure

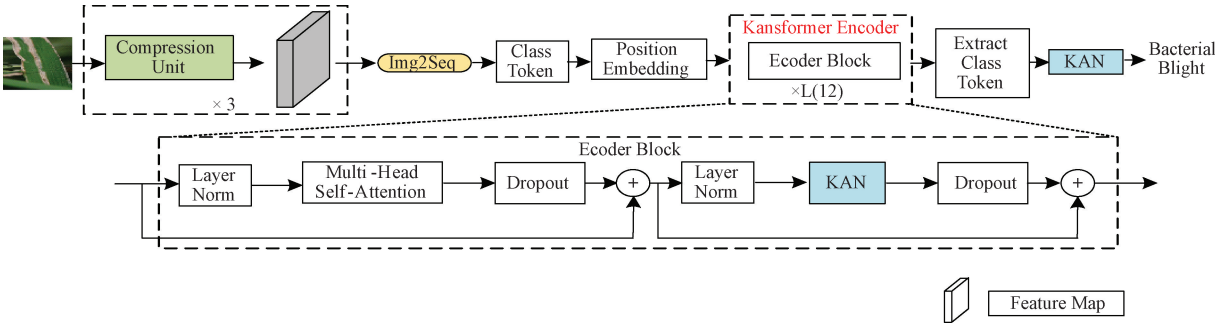


图 2 本文使用的改进 Vision Transformer 网络结构

Fig.2 Improved Vision Transformer network structure

本文使用 compression unit 模块替换了原始 ViT 模型中的图像块分割和展平操作。该模块由卷积分支和注意力分支组成,形成一个并行分支结构,如图 3 所示。注意力分支用于提取水稻叶片病害图像的全局特征信息,首先通过一个卷积金字塔模块 (pyramid reduction module, PRM),对输入图像进行多尺度上下文的提取,使得模型能够捕捉到不同尺度的特征信息,PRM 结构如图 4 所示。提取过程如式(1)所示。

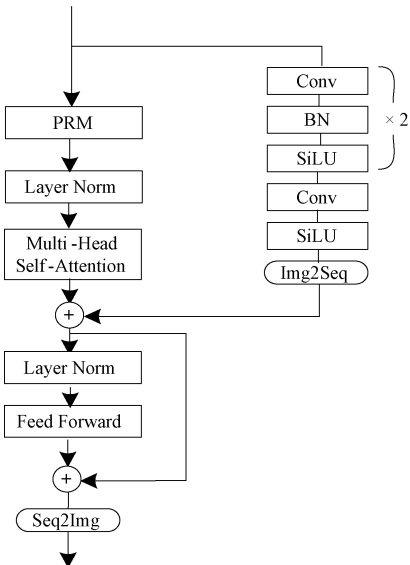


图 3 Compression Unit 结构

Fig.3 Compression unit structure

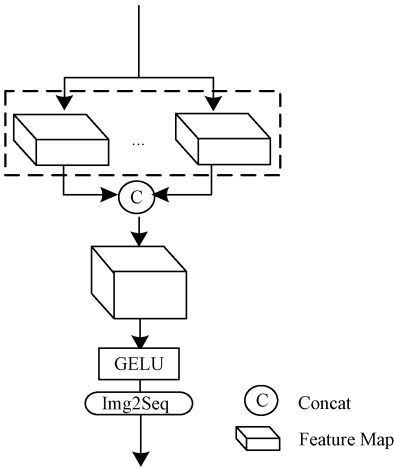


图 4 PRM 结构

Fig.4 PRM structure

$$PRM(f_i) = Cat([Conv_{ij}(f_i; s_{ij}, r_i) | s_{ij} \in S_i, r_i \in R]) \quad (1)$$

式中: f_i 为第 i 个输入的图像特征, $Conv_{ij}(\cdot)$ 为卷积金字塔模块的第 j 层卷积操作, s_{ij} 为卷积层的扩张速率, r_i 为卷积的缩减比率, $Cat(\cdot)$ 将图像的特征块拼接起来。拼接后的特征图通过 $Img2Seq(\cdot)$ 操作多尺度上下文嵌入到每个标记(token)中,并送入到多头自注意力(multi-head self-attention, MHSA)机制中,以建立模型对水稻叶片病害图像中的远程依赖性,从而更好的捕捉图像的全局信息,该

过程如式(2)所示。

$$X_i = \text{MHSA}_i(\text{Img2Seq}(f_i^c)) \quad (2)$$

式中: f_i^c 为拼接后的特征图, X_i 为经过注意力分支的输出, $\text{Img2Seq}(\cdot)$ 为将特征图展平为一维序列。

卷积分支是由堆叠的卷积层和 $\text{Img2Seq}(\cdot)$ 操作组成,该分支对输入的水稻叶片病害图像进行局部上下文信息的提取,以弥补 ViT 不能提取局部特征的缺陷。最终,将注意力分支和卷积分支的特征 token 进行融合,使得 token 既具有多尺度上下文特征信息又包含了局部上下文特征。融合后的 token 通过前馈网络(feed forward network, FFN)生成特征序列,并经过 $\text{Seq2Img}(\cdot)$ 操作重塑为特征图,之后再输入到下一个模块之中,重塑过程如式(3)所示。

$$f_{i+1} = \text{Seq2Img}(\text{FFN}_i(X'_i) + X'_i) \quad (3)$$

式中: f_{i+1} 为 compression unit 模块的输出, X'_i 为注意力分支和卷积分支融合的 token, $\text{Seq2Img}(\cdot)$ 为将 token 重塑为特征图的操作。经过 3 个堆叠的 compression unit 模块之后,将特征图展平为一维 token,并结合 position embedding 和 class token 输入到编码器之中,进行病害图像的特征提取和病害种类识别。

1.3 Kolmogorov-Arnold 网络

多层感知器(multi-layer perceptron, MLP),也被称为全连接前馈神经网络,是 ViT 模型中关键的组成部分,存在于两个部位:一是位于 Transformer 编码器内部的 MLP 层,二是在最终分类任务中作为输出层的 MLP Head。MLP 由多层“节点”(神经元)组成,每个节点对输入进行加权运算,并通过固定的激活函数生成输出^[22]。这种结构使得 MLP 逼近各种复杂的非线性函数,其逼近能力可以通过通用近似定理得以保证。具体来说,通用近似定理指出,只要具有足够数量的隐藏节点,MLP 可以逼近任意连续函数,如式(4)所示。

$$f(x) \approx \sum_{i=1}^{N(c)} a_i \sigma(w_i \cdot x + b_i) \quad (4)$$

式中: $w_i \in R^n$, $a_i, b_i \in R$, w_i 为权重, b_i 为偏置项, σ 为任意连续的非线性激活函数。MLP 因其强大的函数逼近能力,在图像分类、语音识别、自然语言处理等任务中表现出色。然而,MLP 也存在一定的局限性,例如 MLP 可解释性较低,难以直观理解模型的决策过程^[23],此外,MLP 在网络结构扩展时,深度和宽度的灵活性较差^[24],并且在处理长序列数据时,难以捕捉输入序列中长期依赖关系^[25]。

为了克服这一缺点,本文引入了 Kolmogorov-Arnold 网络,一种基于 Kolmogorov-Arnold 表示定理构建的神经网络。虽然 KAN 在外观上与 MLP 相似,均采用全连接结构,但其核心区别在于激活函数的设置方式。MLP 的激活函数是固定的,作用于“节点”(神经元)上,其结构如图 5(a)所示。而 KAN 则将可学习的激活函数置于“边”(权重)上,其结构如图 5(b)所示。因此,KAN 没有传统的

线性权重矩阵,取而代之的是参数化为样条函数的一维可学习函数。这种结构如式(5)所示。

$$f(x) = f(x_1, \dots, x_n) = \sum_{q=1}^{2n+1} \Phi_q \left(\sum_{p=1}^n \phi_{q,p}(x_p) \right) \quad (5)$$

式中: x_p 为输入向量 x 的第 p 个元素, n 为输入维度, q 为遍历外部函数 Φ_q 的每个组成部分, $\phi_{q,p}$ 为一元连续函数(或者称为单变量函数)。

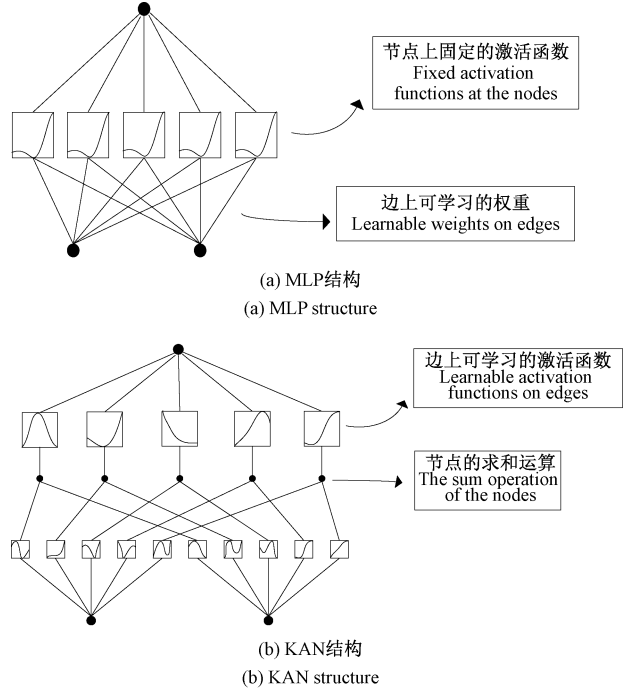


图 5 MLP 和 KAN 结构

Fig. 5 MLP and KAN structure

简而言之,KAN 可以将复杂的多元函数表示为多个简单的一元函数相加,从而更有效地表示图像的复杂特征。这使得 KAN 在处理图像变形、光照变化以及病害形态多样化影响的水稻叶片病害图像时,可以表现出更强的稳健性和适应性,从而提高模型的泛化性能。

从本质上讲,KAN 可以看作是样条函数和 MLP 的结合,兼具二者的优势。样条函数在低维度情况下表现出更高的精度,能够进行局部调整,并在不同分辨率之间切换。然而,它在单独使用时会产生维数缺陷的问题。相比之下,MLP 不容易受到维度的影响,但在低维度时精度不如样条函数。因此,KAN 通过在外层利用 MLP 学习特征信息,并在内层通过样条函数对这些特征进行优化,从而在高维空间中提供更精确的特征表示。

由图 2 可以看出,本文将 Transformer 编码器中的 MLP 模块更换为 KAN 网络,提出了 kansformer encoder 结构,并使用 KAN 网络替换 MLP Head。kansformer encoder 结构由交替的 MHSA 和 KAN 组成,在每个块之前都会进行一次归一化操作,并在每个块之后使用残差连接^[26]。compression unit 模块输出的图像序列首先经过位

置编码,随后将图像 tokens 输入到 kansformer encoder 中。在 kansformer encoder 中,输入序列经过 MHSA 进行特征提取,并在不同位置上联合关注信息,随后通过 KAN 网络进行非线性变换和特征融合。通过多个编码器模块的堆叠,模型能够更好地捕捉输入序列中的不同关系和特征。最终,kansformer encoder 输出一组嵌入向量,包括一个特殊 class token 和所有 patch。将这些向量通过 KAN 网络进行分类,最终得到水稻叶片病害的类别预测。

在 ViT 模型中,MLP 通过全连接层对输入信息进行线性变换,随后进行非线性激活。MLP 的学习能力受到全连接层线性变换的限制,而 KAN 则可以通过不同的一元函数构建多层次的非线性结构,使得模型可以更精准地捕捉不同维度输入变量之间的复杂关系。这种能力使得 KAN 在处理水稻叶片病害图像时,能够更细致地提取病害特征信息。此外,由于 KAN 结构紧凑,模型训练时相较于传统的 MLP 层收敛速度更快。因此,在水稻叶片病害识别任务中,KAN 可以加速模型学习病害相关特征,提高模型识别的准确率。

2 实验结果与分析

2.1 数据集构建与处理

本研究采用的水稻叶片病害数据集是在 Kaggle 网站中获得的,数据集共包括白叶枯病、稻瘟病、褐斑病、东格鲁病和健康叶片 5 种类型,其图像示例如图 6 所示。由于网络上获取的数据集样本数量不平衡,因此对数据集进行亮度变换、旋转和添加噪声等数据增强操作,并将图像像素大小统一调整为 224×224。最终,得到 6 982 张水稻叶片病害图像,具体组成如表 1 所示。按照 8:1:1 的比例将数据集随机划分为训练集、验证集和测试集,输入到模型中进行训练。

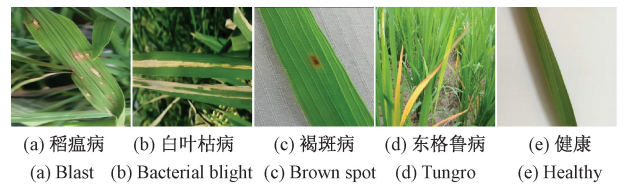


图 6 水稻叶片病害图像示例
Fig. 6 Example of rice leaf disease image

表 1 水稻叶片病害数据集组成

种类	数量/张
白叶枯病	1 309
稻瘟病	1 376
褐斑病	1 458
东格鲁病	1 444
健康叶片	1 395
总计	6 982

2.2 实验环境与训练参数

本研究的实验平台使用的操作系统为 Ubuntu 18. 04. 5 LTS,CPU 为 Intel(R) Xeon(R) Platinum 8255C, GPU 为 24 GB 显存的 NVIDIA GeForce RTX 3090。实验所采用编程语言为 Python3. 8,开发环境为 Pycharm,深度学习框架为 Pytorch。模型采用 Adam 优化器训练,训练轮数为 100,批大小(batch size)为 32,初始学习率为 0. 001,采用 LambdaLR 方法动态调整学习率。

2.3 训练评价指标

为了评估模型在水稻叶片病害识别任务中的性能,本文采取了准确率、精确率、召回率、F1 分数 4 个评价指标,其公式分别如式(6)~(9)所示。

$$Accuracy = \frac{TP + TN}{TP + FN + FP + TN}$$
 (6)

$$Precision = \frac{TP}{TP + FP}$$
 (7)

$$Recall = \frac{TP}{TP + FN}$$
 (8)

$$F1\text{-score} = 2 \times \frac{precision \cdot recall}{precision + recall}$$
 (9)

式中:TP(true positive)表示预测结果为正样本,实际结果也为正样本的样本数量;FP(false positive)表示预测结果为正样本,实际结果为负样本的样本数量;TN(true negative)表示预测结果为负样本,实际结果也为负样本的样本数量;FN(false negative)表示预测结果为负样本,实际结果为正样本的样本数量。

2.4 改进 ViT 模型的识别性能实验

为验证本文提出模型相较于原始 ViT 模型的有效性,在数据集上对二者的准确率和损失值进行比较,其曲线分别如图 7、8 所示。

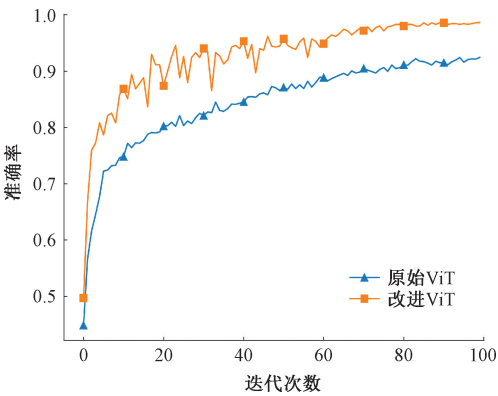


图 7 准确率对比曲线图
Fig. 7 Comparison curve of accuracy

如图 7 所示,当训练轮次为 60 轮左右时,本文提出的模型准确率曲线逐步趋于稳定,而原始 ViT 模型的准确率曲线还呈现上升趋势,直到 80 轮左右时才达到收敛。最终改进 ViT 模型的准确率达到 98.62%,原始 ViT 模型准

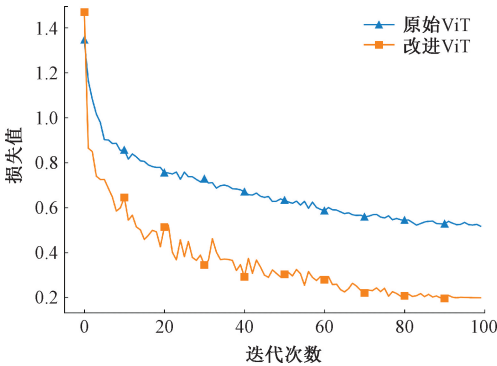


图 8 损失值对比曲线

Fig. 8 Comparison curves of loss values

确率为 92.42%。如图 8 所示,随着训练轮次的增加,模型的损失值均不断减小,并在 70 轮左右时逐步趋于稳定,但改进 ViT 的损失值曲线下落速度更快,最终达到 0.198,而原始 ViT 的最终损失值为 0.51,说明本文提出的模型更有效的最小化了训练误差,具有更好的学习能力。因此可以证明,本文的改进策略是合理可行的,提高了模型的稳定性,达到了较好的训练效果。

2.5 消融实验

为了验证本文对 Vision Transformer 的各个改进策略的有效性,开展了消融实验,在相同的训练参数和配置环境下,对 compression unit 和 KAN 网络进行测试,判断其在 ViT 模型上的有效性,消融实验的结果如表 2 所示。

表 2 消融实验						%
Table 2 Ablation experiments						
实验序号	Compression Unit	KAN	准确率	精确率	召回率	F1 分数
1			92.42	92.41	92.33	92.32
2		✓	96.60	96.51	96.53	96.51
3	✓		96.85	96.12	95.95	96.03
4	✓	✓	98.62	98.09	98.14	98.11

通过对表格中结果的分析,当引入 KAN 网络后,模型的性能指标均提高了约 4%,这说明 KAN 网络在提升模型性能上起到了积极作用;当模型单独使用 compression unit 模块后,模型的表现进一步提升,尤其是准确率达到了 96.85%。这一结果表明,compression unit 通过引入局部性和尺度不变性两种归纳偏置,减少了特征冗余和提高了模型对全局与局部特征信息的提取。当 compression unit 和 KAN 网络同时使用时,模型的性能达到最高,准确率、精确率、召回率和 F1 分数分别提高了 6.2%、5.68%、5.81%和 5.79%,充分验证了模型改进的有效性,有效证明了两者的结合能够最大化地提升 ViT 的识别效果。

2.6 对比实验

为了更直观地验证本文提出的改进 ViT 模型的性能

效果,与其他算法模型进行了对比,实验结果如表 3 所示。通过对表格的分析可知,本文提出的改进 ViT 模型在性能指标方面均优于其他对比网络模型。具体而言,改进 ViT 模型的准确率、精确率、召回率和 F1 分数分别达到了 98.62%、98.09%、98.14%和 98.11%,远高于其他模型。这一结果表明,本文的改进策略显著提升了 ViT 在水稻叶片病害识别任务中的表现。与性能最高的先进模型相比,本文提出的模型在准确率上提高了 1.56%,精确率提高了 0.96%,召回率提高了 1.06%,F1 分数提高了 1.02%。这得益于模型引入了固有的归纳偏置(局部性和尺度不变性)和 KAN 网络,增强了模型对局部叶片病害特征和多尺度上下文信息的提取能力,使得模型在处理复杂病害特征时表现更为出色,有效提升了水稻叶片病害图像的识别效果。

表 3 不同模型的性能对比					%
Table 3 Performance comparison of different models					
模型	准确率	精确率	召回率	F1 分数	
PMVT ^[13]	92.00	—	—	—	
DenseNet-BiFAPN-YOLOv5 ^[9]	94.87	—	75.81	92.00	
ECA-ViT ^[14]	95.41	95.86	95.79	94.47	
RepVGG_ECA ^[10]	97.06	97.13	97.08	97.09	
本文模型	98.62	98.09	98.14	98.11	

3 结 论

本文提出了一种基于改进 Vision Transformer 的水稻叶片病害识别方法,通过引入 CNN 的内在归纳偏置和

KAN 网络,显著提升了模型在局部特征与全局依赖性建模上的能力,增强了模型的非线性特征表达能力和可解释性。实验结果表明,改进后的 ViT 模型在水稻叶片病害识别任务中具有显著的优势。但 Vision Transformer 的模型

参数量较大,对硬件要求较高,难以部署在移动设备上。因此,未来工作希望在保证准确率的同时,降低模型的参数量,并尝试将该模型应用于其他作物的病害检测任务中,提高模型的泛化能力。

参考文献

- [1] 叶恭银,方琦,徐红星,等.我国水稻螟虫发生及治理研究进展[J].植物保护,2023,49(5):167-180.
YE G Y, FANG Q, XU H X, et al. Advances in the researches on the occurrence, damage and management of rice stem borers in China[J]. Plant Protection, 2023, 49(5): 167-180.
- [2] 严如玉,甘国渝,赵希梅,等.我国水稻优势产区生产格局及施肥现状研究[J].中国稻米,2023,29(3):1-8.
YAN R Y, GAN G Y, ZHAO X M, et al. Study on the production pattern and fertilization status of rice dominant production areas in China[J]. China Rice, 2023, 29(3): 1-8.
- [3] 魏赛金.有益微生物在水稻病害防治的研究进展与应用现状[J].生物灾害科学,2020,43(1):1-7.
WEI S J. Advance and status in the application of beneficial microorganisms in the control of rice diseases[J]. Biological Disaster Science, 2020, 43(1): 1-7.
- [4] 翁杨,曾睿,吴陈铭,等.基于深度学习的农业植物表型研究综述[J].中国科学:生命科学,2019,49(6):698-716.
WENG Y, ZENG R, WU CH M, et al. A survey on deep-learning-based plant phenotype research in agriculture[J]. Scientia Sinica(Vitae), 2019, 49(6): 698-716.
- [5] 兰玉彬,王天伟,陈盛德,等.农业人工智能技术:现代农业科技的翅膀[J].华南农业大学学报,2020,41(6):1-13.
LAN Y B, WANG T W, CHEN SH D, et al. Agricultural artificial intelligence technology: Wings of modern agricultural science and technology [J]. Journal of South China Agricultural University, 2020, 41(6): 1-13.
- [6] SHOAIB M, SHAH B, EI-SAPPAGH S, et al. An advanced deep learning models-based plant disease detection: A review of recent research[J]. Frontiers in Plant Science, 2023, 14: 1158933.
- [7] 周惠汝,吴波明.深度学习在作物病害图像识别方面应用的研究进展[J].中国农业科技导报,2021,23(5):61-68.
ZHOU H R, WU B M. Advances in research on deep learning for crop disease image recognition[J]. Journal of Agricultural Science and Technology, 2021, 23(5): 61-68.
- [8] 王东方,汪军.基于迁移学习和残差网络的农作物病害分类[J].农业工程学报,2021,37(4):199-207.
WANG D F, WANG J. Crop disease classification with transfer learning and residual networks [J]. Transactions of the Chinese Society of Agricultural Engineering, 2021, 37(4): 199-207.
- [9] KUMAR V S, JAGANATHAN M, VISWANATHAN A, et al. Rice leaf disease detection based on bidirectional feature attention pyramid network with YOLO v5 model[J]. Environmental Research Communications, 2023, 5(6): 065014.
- [10] NI H J, SHI ZH W, KARUNGARU S, et al. Classification of typical pests and diseases of rice based on the ECA attention mechanism. agriculture [J]. Agriculture, 2023, 13(5):1066.
- [11] 刘华咏,黄聪,金汉均.注意力增强的视觉Transformer图像检索算法[J].电子测量技术,2023,46(23):50-55.
LIU H Y, HUANG C, JIN H J. Image retrieval method with attention-enhanced visual Transformer [J]. Electronic Measurement Technology, 2023, 46(23): 50-55.
- [12] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need [J]. Advances in Neural Information Processing Systems, 2017, 30:03762.
- [13] LI G Q, WANG Y CH, ZHAO Q, et al. PMVT: A lightweight vision transformer for plant disease identification on mobile devices[J]. Frontiers in Plant Science, 2023, 14: 1256773.
- [14] HE F Y, LIU Y, LIU J F. ECA-ViT: Leveraging ECA and Vision Transformer for crop leaves diseases identification in cultivation environments[C]. 2024 4th International Conference on Machine Learning and Intelligent Systems Engineering (MLISE). IEEE, 2024: 101-104.
- [15] WU SH P, SUN Y Q, HUANG H. Multi-granularity feature extraction based on vision transformer for tomato leaf disease recognition [C]. 2021 3rd International Academic Exchange Conference on Science and Technology Innovation(IAECST). IEEE, 2021: 387-390.
- [16] SHARMA S K, VISHWAKARMA D K. Classification of banana plant leaves based on nutrient deficiency using Vision Transformer [C]. 2024 5th International Conference for Emerging Technology (INCET). IEEE, 2024: 1-6.

- [17] 陈晓, 夏颖. 基于改进 MobileViT 网络的番茄叶片病害识别[J]. 电子测量技术, 2023, 46(14): 188-196.
CHEN X, XIA Y. Improved MobileViT network for tomato leaf disease identification [J]. Electronic Measurement Technology, 2023, 46(14): 188-196.
- [18] ALEXEY D. An image is worth 16×16 words: Transformers for image recognition at scale[J]. ArXiv preprint arXiv: 2010.11929, 2020.
- [19] LIU Z M, WANG Y X, VAIDYA S, et al. KAN: Kolmogorov-arnold networks [J]. ArXiv preprint arXiv:2404.19756, 2024.
- [20] PENG ZH L, HUANG W, GU SH ZH, et al. Conformer: Local features coupling global representations for visual recognition[C]. IEEE/CVF International Conference on Computer Vision, 2021: 367-376.
- [21] 乔琦, 王红军, 马康, 等. 基于改进的 DenseNet-ViT 联合网络和迁移学习的燃气轮机转子故障诊断[J]. 电子测量与仪器学报, 2024, 38(11): 40-47.
QIAO Q, WANG H J, MA K, et al. Gas turbine rotor fault diagnosis based on improved DenseNet-ViT joint network and transfer learning[J]. Journal of Electronic Measurement and Instrumentation, 2024, 38(11): 40-47.
- [22] 邓聪颖, 邓子豪, 赵洋, 等. 有限样本下基于迁移学习的铣削稳定性预测方法[J]. 仪器仪表学报, 2023, 44(9): 313-321.
DENG C Y, DENG Z H, ZHAO Y, at al. Milling stability predictions under limited samples based on transfer learning [J]. Chinese Journal of Scientific Instrument, 2023, 44(9): 313-321.
- [23] MONTAVON G, SAMEK W, MULLER K R. Methods for interpreting and understanding deep neural networks[J]. Digital Signal Processing, 2018, 73: 1-15.
- [24] YU R P, YU W H, WANG X CH. KAN or MLP: A fairer comparison [J]. ArXiv preprint arXiv: 2407.16674, 2024.
- [25] 包强强, 唐思源, 李擎乾, 等. 改进 YOLOv8 模型实现多类型肺结节检测[J]. 计算机科学与探索, 2025, 19(2): 429-442.
BAO Q Q, TANG S Y, LI Q Q, et al. Improve the YOLOv8 model for multi-type lung nodule detection[J]. Journal of Frontiers of Computer Science and Technology, 2025, 19(2): 429-442.
- [26] CHEN Y F, ZHU ZH, ZHU SH H, et al. SCKansformer: Fine-grained classification of bone marrow cells via Kansformer backbone and hierarchical attention mechanisms [J]. ArXiv preprint arXiv: 2406.09931, 2024.

作者简介

朱周华, 副教授, 主要研究方向主要为信号与信息处理、机器视觉等。

E-mail: zhuzhouhua@xust.edu.cn

周怡纳(通信作者), 硕士研究生, 主要研究方向机器学习、目标检测等。

E-mail: 875321730@qq.com

侯智杰, 硕士研究生, 主要研究方向深度学习、目标检测等。

E-mail: sy913589711@163.com

田成源, 硕士研究生, 主要研究方向深度学习、图像处理等。

E-mail: 614489450@qq.com