



DOI:10.13364/j.issn.1672-6510.20250017

网络首发日期: 2025-06-20; 网络首发地址: <http://link.cnki.net/urlid/12.1355.N.20250619.1604.009>

基于图文模型和对比学习的水稻病害识别方法

杨巨成, 沈 杰, 刘建征, 吴 超

(天津科技大学人工智能学院, 天津 300457)

摘 要: 针对传统的基于单一水稻图像的水稻病害识别方法存在识别精度不高、过度依赖标签数据训练的问题, 本文提出基于图文模型和对比学习的水稻病害识别方法。文本模态的信息可以与图像模态信息形成互补, 在一定程度上弥补图像训练样本不足的问题, 而对比学习采用无监督学习方式, 改善现有的深度学习模型在水稻叶部病害的识别中过度依赖标签数据训练的情况。利用 GPT-3.5 大模型图文生成技术, 构建水稻图文对数据集; 在图文模型的水稻病害检测模型中, 对文本编码器模块进行优化, 从而缩短训练时长; 基于对比学习的水稻病害识别方法, 对上述图文模型中的图像编码器实施难负样本策略改进, 更有效地学习各类别特征表示, 从而具备更强的鲁棒性。实验结果表明, 本文模型在水稻病害任务上识别准确率优于其他模型。

关键词: 水稻病害识别; 对比学习; 图文模型; 无监督学习

中图分类号: TP391

文献标志码: A

文章编号: 1672-6510(2026)03-0010-08

A Method for Rice Disease Recognition Based on Image-Text Model and Contrastive Learning

YANG Jucheng, SHEN Jie, LIU Jianzheng, WU Chao

(College of Artificial Intelligence, Tianjin University of Science and Technology, Tianjin 300457, China)

Abstract: To address the issues of low recognition accuracy and over-reliance on labeled data in traditional rice disease recognition methods based solely on rice images, this article proposes a rice disease recognition method based on an image-text model and contrastive learning. The text modality complements image information, partially mitigating the problem of insufficient image training samples. Contrastive learning, an unsupervised learning approach, improves the current problem of over-reliance on labeled data in training deep learning models for rice leaf disease recognition. Firstly, a rice image-text pair dataset is constructed using the text generation from image technology of large models, GPT-3.5. Then, within the image-text model for rice disease recognition, the text encoder module is optimized to reduce training time. Additionally, a rice disease recognition method based on contrastive learning employs a hard negative sampling strategy within the image encoder of the image-text model. This allows it to effectively learn feature representations of various categories, thereby enhancing model robustness. Experimental results demonstrate that the proposed model outperforms other models in recognition accuracy for rice disease tasks.

Key words: rice disease recognition; contrastive learning; image-text model; unsupervised learning

引文格式:

杨巨成, 沈杰, 刘建征, 等. 基于图文模型和对比学习的水稻病害识别方法[J]. 天津科技大学学报, 2026, 41(3): 10-17.
YANG J C, SHEN J, LIU J Z, et al. A method for rice disease recognition based on image-text model and contrastive learning[J]. Journal of Tianjin university of science and technology, 2026, 41(3): 10-17.

收稿日期: 2025-01-25; 修回日期: 2025-05-06

基金项目: 天津市自然科学基金重点项目(18JCZDJC32100)

作者简介: 杨巨成(1980—), 男, 湖北天门人, 教授; 通信作者: 刘建征, 讲师, jz_leo@tust.edu.cn

农作物病害是主要农业灾害之一,其发生范围和严重程度对国民经济特别是农业生产常造成重大损失。及时监测作物病害是科学有效防控农作物病虫害、提高作物单产水平、保障粮食稳产丰收的农业生产实践中最重要的举措之一。水稻是人类重要的粮食作物之一,其总产量居世界粮食作物产量第三位。据全国农技中心预测分析,2024 年我国水稻病虫害将呈偏重发生态势,发生面积 12.4 亿亩^[1]。因此,快速、准确识别水稻病害种类,有利于针对性地采取防治措施,对中国粮食安全具有重要的现实意义。

随着计算机技术的不断发展,深度学习被广泛应用于图像分类^[2-3]、目标检测、自然语言处理^[4-6]中,植物病害也逐渐从人工识别转向智能识别,大大减少了人力、物力。深度学习算法通过学习大量的病害图像数据,能够自动提取病害特征,提高识别的准确性,显著优于传统方法。通过准确识别病害类型和发生程度,可以为农民提供精准的防治建议,减少农药的使用量,降低环境污染,同时提高水稻产量和质量。现代信息技术与农业生产相结合,不仅提高了农业生产效率,还为农业的可持续发展提供技术支持,同时进一步优化了病害识别模型。这有助于推动农业领域的技术创新和知识共享,对保障粮食安全具有重要的现实意义。

尽管水稻病害识别单模态技术取得了显著的进步,但在实际应用中仍面临诸多挑战。由于田间采集的病害图像通常背景复杂、病症间差异细微且多变,水稻病害在实际田间环境下识别困难。为解决上述问题,现有研究主要采用两种途径:第一种是增加训练数据规模,使模型充分学习各类病害的特征及变化;第二种则是设计性能更强的模型,从图像中提取更精准的视觉特征,提升分类预测的准确性。此外,现有水稻病害识别方法存在共性,即传统卷积网络的有监督学习使用大量标注图像样本,模型的性能过度依赖已标注的训练集,并且植物病害症状可能由不同疾病、不同环境引起,但在视觉上却几乎完全相同,仅靠局部特征无法区分。

针对上述问题,本文提出基于图文模型和对比学习的水稻病害识别方法。文本模态的信息可以与图像模态信息形成互补,在一定程度上弥补图像训练样本不足的问题,而对比学习采用无监督学习方式,改善现有的深度学习模型在水稻叶部病害识别中过度依赖标签数据训练的情况。本文以 4 类常见的水稻病害为研究对象进行实验,改变以往单一图像模

态,使用以 ResNet101 与 Transformer 为骨干网络的图文模型对水稻病害进行检测,通过对比学习计算图像文本特征相似度,得到最终预测结果。实验首先针对水稻病害识别领域任务,在数据准备阶段构建水稻图文对数据集;其次在大型植物病害图像库 Plant Wild 上进行特征提取,两次训练得到预训练模型;最后在自制的水稻病害数据集上进一步训练微调,得到水稻病害分类结果,从而更具针对性地进行水稻病害防治工作。

1 相关工作

1.1 图文模型

图文模型可以同时从图像和文本中学习,以处理多种多模态任务。图像模态通常采用卷积神经网络(CNN)构造,Kawasaki 等^[7]最早将基于 CNN 的图像分类技术引入黄瓜叶病识别领域,训练时用 4-fold 交叉验证,其识别准确率最终达到 94.9%。孙俊等^[8]在 AlexNet 模型的基础上,采用批归一化方法以及加入全局池化层和缩减特征图数目的方法,得到 8 种改进模型,对 PlantVillage 中的 14 种植物共 26 种病害进行识别,其最优模型的平均测试识别准确率达到 99.56%。Lu 等^[9]针对大田环境的复杂性,采集 9 230 张小麦病害图像,构建多实例弱监督框架。该模型结合全卷积网络(FCN)的像素级分割能力,同步实现病害区域定位〔平均交并比(IoU)达 92.7%〕与类别诊断(分类精度为 97.95%),为复杂背景下的精准识别提供了技术范式。

文本模态通常是人们对图像的文本描述,它们往往包含丰富的语义信息。这种文本模态信息可以与图像模态信息形成互补,在一定程度上解决图像训练样本不足的问题。与传统的仅依赖图像特征的分类方法相比,多模态分类能够捕捉更丰富的信息,提高分类的鲁棒性和准确性。

多模态融合技术主要包括特征融合和决策融合两种方法。特征融合(feature fusion)是将图像特征和文本特征在特征层面进行结合,形成新的特征表示。常见的特征融合方法包括拼接融合(concatenation fusion)和加权融合(weighted fusion)。决策融合(decision fusion)是在分类决策层面进行融合,即将图像分类器和文本分类器的输出进行结合,得到最终的分类结果。常见的决策融合方法包括平均融合(average fusion)、投票融合(voting fusion)和逻辑融

合(logic fusion)。

Zhou 等^[10]以番茄和黄瓜常见侵袭性疾病识别为研究对象,针对现有模型中多模态数据利用不足的问题,构建基于图像-文本多模态协同表示和知识辅助(ITK-Net)的疾病识别模型,利用两类模态数据之间的相关性和互补性,实现疾病特征的联合表征学习。在由图像-文本对组成的数据集上,该模型的精确度达到了 99%。周一帆等^[11]针对现有农作物病害叶片检测方法图像特征定位病害区域精确率不高的问题,提出基于多模态特征对齐的作物病害叶片检测方法,该方法在马铃薯、番茄、苹果和草莓 4 种类型病害叶片数据集上的精确率分别达到 0.9574、0.9611、0.9580 和 0.9502。He 等^[12]以联合的端到端方式将视觉和语言双流结合起来,保留通道内和通道间的信息,生成互补的细粒度表示,在加州鸟数据集上的实验结果表明,模型充分融合文本特征和视觉特征,实现更好的分类精度。

1.2 对比学习

对比学习是自监督学习中的重要方法,其核心思想是通过比较样本之间的相似性和差异性学习数据的表示,即最大化正样本(相似样本)之间的相似性,同时最小化负样本(不相似样本)之间的相似性。这种方法可以帮助模型捕捉到数据的内在结构和分布。通过该学习机制,使模型能够学习到不同类别的特征表示,同时保持对输入扰动的鲁棒性。

目前大量的研究工作针对正对采样的策略,并在

多视图和对比学习领域进展迅猛。对图像数据而言,正对采样策略通常为保持语义内容的变换,例如抖动、随机剪切、分离颜色通道等。在原始像素数据中,这种变换学习控制策略对最终结果影响显著,而在对比学习中,负对的选择却很少受到关注。通常,给定一个锚点 x ,从训练数据中均匀采样一个负样本(x^-),而不管它对学习到的表示可能有多大的信息量。在监督和度量学习设置中,“难”(真负)示例可以帮助指导学习方法更快地纠正错误。Chen 等^[13]提出基于对比学习的自监督学习方法 SimCLR,它使用小批量数据中其他项的增强视图作为负样本。He 等^[14]提出 MoCo 方法,它使用动量更新旧负表征存储体库的方法,解决使用大批量的负样本困难的问题。Chuang 等^[15]随后提出了“除偏”的方法,使用正-负标记学习的方法,利用真实负分布的分解,纠正了并非所有的负对都可能是真正否定事实。Kalantidis 等^[16]考虑应用 Mixup^[17]在潜在空间生成难负样本, Jin 等^[18]利用视频的特定时间结构生成难负样本用于视觉任务中的目标检测。

2 基于图文模型和对比学习的水稻病害识别

本文提出基于图文模型和对比学习的水稻病害识别方法,采用无监督学习方式,改善现有的深度学习模型在水稻叶部病害的识别中过度依赖标签数据训练的情况。本文方法示意图如图 1 所示。

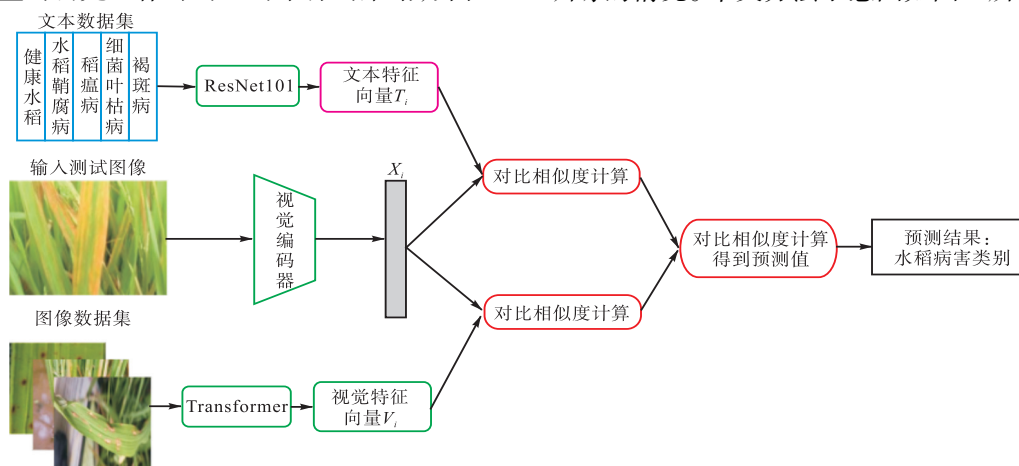


图 1 本文方法示意图

Fig. 1 Schematic diagram of proposed method

首先,根据公开数据集,针对水稻病害识别领域任务,在数据准备阶段构建水稻图文对数据集。然后,构造以 ResNet101 与 Transformer 为骨干网络的水稻病害检测图文模型,改进文本编码器模块,在保

证模型识别准确率的同时缩短模型训练时间。最后,针对植物病害分类任务在传统模式下过度依赖大量标签数据的问题,基于对比学习思想,对上述图文模型中的图像编码器进行难负样本策略改进。通过改进

学习机制,模型能够在学习到不同类别的特征表示的同时,保持良好的鲁棒性。

2.1 骨干网络

ResNet 引入残差学习(residual learning)机制,解决传统 CNN 在增加网络深度时导致网络性能下降的梯度消失和表示瓶颈等问题。ResNet101 包含 101 层网络结构,残差学习机制引入的恒等映射(identity mapping)和残差块(residual block),使网络可以学习输入和输出之间的残差,网络更利于提取丰富的特征信息。Transformer 的自注意力(self-attention)机制,在模型处理序列的每个元素时动态地聚焦于序列中其他重要元素,从中获取长距离依赖关系,而其在不同下游任务中的成功也很好证明了该架构在捕获语言的深层语义和句法结构方面的有效性。

本文方法由基于 ResNet101 的图像编码器和基于 Transformer 的文本编码器构成。图像模态通道输入图像,ResNet101 提取图像特征,Transformer 获取文本提示,通过 k 均值聚类(k-means)方法构造图像原型和文本原型,通过对比学习将其映射到同一空间计算特征相似度,通过对角线上的标签引导编码器对

齐,加速模型收敛,得到预测结果。

2.2 数据集准备

在植物病害领域,某些病害可能呈现出相似的整体表现,为了捕获这些植物病害的细粒度特征,在数据集中为每个类别提供丰富的文本描述,以便提高识别准确度。在此使用可控文本生成技术,采用大规模预训练模型 GPT-3.5 生成图像对应文本内容。可控文本生成任务期望生成的文本符合给定的属性,其生成过程为

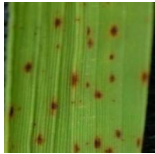
$$p(y) = \sum_{i=1}^r p(y_i | y_{<i}, x, c) \quad (1)$$

其中: x 为输入文本, y 为生成文本, $p(y)$ 为生成 y 的概率, y_i 为第 i 生成步生成的词, $y_{<i}$ 为前 i 步已经生成的词序列, r 为生成文本长度, c 为要控制的文本属性(情感、主题、风格、人物角色等)。

本文构建水稻数据集 Rice Leaf Diseases。该数据集有水稻鞘腐病、稻瘟病、细菌性叶枯病、褐斑病和健康水稻叶片这 5 种类别,共有 3 874 张水稻叶片图像。使用大规模语言模型 GPT-3.5 为每类图像生成相应文字描述,图像-文本对示例见表 1。

表 1 图像-文本对示例

Tab. 1 Image-text example

类别	图像	文本描述
稻瘟病 (rice blast)		稻瘟病的病斑在叶片上表现为长形、菱形,边缘为褐色或红褐色,中心为灰色或白色。稻瘟病的病斑最初出现为微小的棕色斑点,然后成长为纺锤形,两端尖细。病斑中心通常是灰色或白色,边缘为褐色或红褐色。
水稻鞘腐病 (rice sheath blight)		稻鞘腐病初期表现为叶鞘上出现褐色小斑点,这些斑点边缘不明显,随后扩展形成深褐色的菱形斑块。剑叶叶鞘上会出现暗褐色小斑,边缘模糊,逐渐融合形成云纹状大斑块。湿度较大时,病斑内外会出现白色至粉红色的霉状物。
细菌性叶枯病 (bacterial blight leaf)		细菌性叶枯病初期在叶片上出现圆形小水渍状褪绿斑,逐渐扩大呈近圆形或多角形的褐色斑。病斑直径通常在 1~2 mm,周围有褪绿晕圈,病叶背面不易见到菌脓。叶片上的病斑以叶脉为界线形成多角形,初为水渍状,后中部组织坏死呈褐色至黑色。
褐斑病 (brown spot)		细菌性褐斑病的病斑初为褐色水浸状小点,向上、下扩展后形成菱形深褐色斑,边缘浅褐色。叶片与叶脉交界处多现褐色大片病斑。剑叶叶鞘先发病且受害严重,叶鞘上生褐色至暗褐色不规则病斑,中间色浅,边缘黑褐色较清晰。
健康水稻叶片 (rice healthy leaf)		叶片颜色鲜绿,色泽均匀,没有黄化或白斑。叶片完整,无缺损,边缘平滑无锯齿。叶片挺直,直立生长,表现出良好的生长态势。叶脉清晰,分布均匀,为绿色或淡绿色。

由于数据集不同分类下的图像采集自不同设备, 所以需要对原始图像进行统一裁剪, 将图像大小统一调整至 224 像素 × 224 像素。数据集中病害图像的文本描述均为中文, 在输入网络之前需要对中文文本进行归一化、分词、构建词表、文本向量化等操作, 其中分词使用 jieba 分词工具, 综合考虑原始文本的初始文本长度以及切分后文本的病害特征保留程度, 并且经过大量实验验证, 向量化的文本长度为 20 个字符时效果最佳, 原始文本超过此长度则切除, 不足此长度则补充为 0。

2.3 图像编码器

图像编码器以 ResNet101 为基础骨架, 提取图像特征 $F_v \in \mathbf{R}^{N_{\text{train}} \times D}$, 其中 N_{train} 是所有训练样本的个数, D 是每个特征向量的维数。对每一类别迭代应用 k-means 得到 K 个簇, 从而将具有相似表示的图像特征聚类在一起。包含训练样本中每一个类别的聚类中心为该类别的图像原型 $V \in \mathbf{R}^{M \times K \times D}$ 。为使水稻检测任务得到提升, 使用难采样方法^[19], 设计分布 q , 用来取样负样本, 具体采样分布定义为

$$\begin{cases} q_{\beta}(x^-) := q_{\beta}(x^- | h(x) \neq h(x^-)) \\ \text{where } q_{\beta}(x^-) \propto e^{\beta f(x) - f(x^-)} \cdot p(x^-) \end{cases} \quad (2)$$

该策略通过指数项控制采样难度, 使难负样本同原始样本尽可能相似, 这有助于指导模型学习, 更快地纠正其错误。

2.4 文本编码器

文本编码器采用 1 个 63 M 参数、12 层、512 维且带有 8 个注意力头的 Transformer 模型, 该模型宽度随着图形编码器尺寸的增加而增加。为了提高效率, 本研究改进文本编码器, 在编码器块中用 1 对 Macaron-Like Feed Forward Modules 替代 Transformer 模型中单个 Feed Forward Module, 用 Strang 算子分裂代替 Lie-Trotter 算子分裂, 将 Transformer 模块的 Feed Forward 层改为 2 个半步 Feed Forward 层, 一个在 Attention 层之前, 另一个放在最后。该方法可以让文本编码器中 Encoder Block 的效率得到提升, 具体结构如图 2 所示。利用文本编码器从描述性提示 $P \in \mathbf{R}^{M \times J}$ 中提取文本特征, 其中 J 表示每个类的提示数量。通过对每一类文本提示, 使用文本特征构造文本原型 $T \in \mathbf{R}^{M \times J \times D}$ 。

2.5 对比学习相似度计算

对比学习相似度计算可以衡量两个样本之间的相似性。将两个样本的表示向量映射到一个共同的

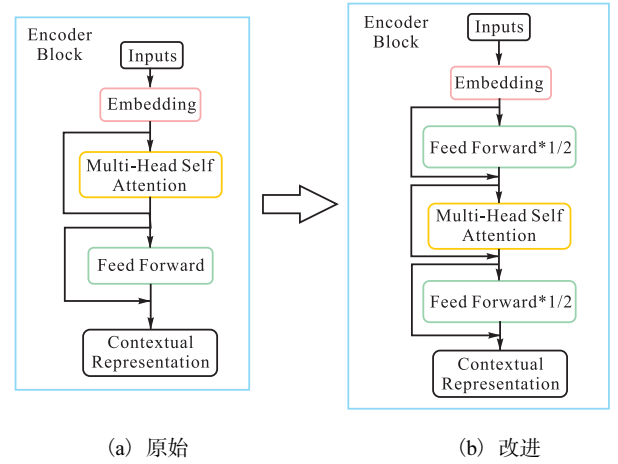


图 2 文本编码器

Fig. 2 Text encoder

嵌入空间, 然后计算它们之间的相似度。具体公式为

$$s = \frac{f(x_1) \cdot f(x_2)}{|f(x_1)| \cdot |f(x_2)|} \quad (3)$$

使用大量的图像-文本对进行对比学习训练。在这个阶段, 模型学习将图像和文本嵌入同一特征空间中, 使相似的图像-文本对在特征空间中更接近, 而不相似的对则更远离。使用图像原型 V 和文本原型 T 作为分类器的初始化权重。为了对图像 I 进行分类, 使用图像编码器提取图像特征向量 $X \in \mathbf{R}^D$ 。然后, 获取特征向量 X 与图像原型 V 和文本原型 T 之间的余弦相似度矩阵: $\cos(x, V_i^k)$ 和 $\cos(x, T_i^j)$, 其中 V_i^k 和 T_i^j 分别表示第 i 个类别的第 k 个图像原型和第 j 个文本原型。归一化处理后, 使用 Softmax 函数将模型原始输出 (logit) 转化为概率分布, 每个类别的概率表示对这个类别的置信度。第 i 类概率为

$$p_{vi} = \frac{\exp\left(\sum_{k=1}^K \cos(x, V_i^k) / \tau\right)}{\sum_{m=1}^M \exp\left(\sum_{k=1}^K \cos(x, V_i^m) / \tau\right)} \quad (4)$$

对文本原型每个类别分别进行最大化操作和平均化操作, 分别得到两个原始输出 (logit)。

最大化概率计算公式为

$$p_{\max i} = \frac{\exp\left(\cos(x, \tilde{T}_i^j) / \tau\right)}{\sum_{m=1}^M \exp\left(\cos(x, \tilde{T}_i^m) / \tau\right)} \quad (5)$$

平均概率计算公式为

$$p_{\text{avg} i} = \frac{\exp\left(\frac{1}{J} \sum_{j=1}^J \cos(x, T_i^j) / \tau\right)}{\sum_{m=1}^M \exp\left(\frac{1}{J} \sum_{j=1}^J \cos(x, T_i^m) / \tau\right)} \quad (6)$$

为进一步提高准确率, 将原型设置为可学习的参数。利用交叉熵损失对获得的 logit 进行训练, 同时

保持骨干网络的参数冻结。相应的对比损失为

$$L_v = -\frac{1}{N_{\text{train}}} \sum_{n=1}^{N_{\text{train}}} \sum_{i=1}^M y_{n,i} \log p_{n,v,i} \quad (7)$$

经过训练得到预测值, 调整 3 个超参数, 整合每个输出概率。总体 logit 为

$$p_i = \alpha \cdot p_{v,i} + \beta \cdot p_{t_{\max},i} + \gamma \cdot p_{t_{\text{avg}},i} \quad (8)$$

其中: p_i 表示第 i 个类别的最终预测概率; α 、 β 和 γ 是平衡每个概率权重的超参数; 在 M 个类中, 选择概率值最高的类作为最终的预测。

3 实验

实验使用的软件为 Python 3.8 和 PyTorch 1.12.2 深度学习框架, 运行内存为 32 GB, 搭载 AMD Ryzen 7 3800X 8-Core Processor 3.90 GHz 处理器, 配备 NVIDIA GeForce GTX1060, 操作系统为 Windows 10, CUDA 11.7.102。

3.1 数据集

使用 2 个数据集, 分别为公共数据集 Plant Wild^[20]和自制数据集 Rice Leaf Diseases。

Plant Wild 数据集是一个大规模的野生植物病害识别数据集, 涵盖了 89 类, 包括 33 个健康植物类和 56 个病害类, 共 18 542 张图像。基于先前的植物病害数据集^[21]和其他常见植物种类, Wei 等^[20]通过从 Google Images、Ecosia 和百度图像下载图像进行数据收集, 图像数量最多的类别有 589 张图像, 最少的类别有 44 张图像。

本文提出的水稻数据集 Rice Leaf Diseases 包含水稻鞘腐病、稻瘟病、细菌性叶枯病、褐斑病和健康水稻叶片这 5 种类别共 3 874 张水稻叶片图像, 使用大规模语言模型 GPT-3.5 为每类图像生成相应文字描述, 以此作为图像-文本对, 最终按照 7 : 2 : 1 划分原始数据集为训练集、验证集和测试集, 具体数量见表 2。

表 2 数据集样本数量

Tab. 2 Sample sizes of datasets

病害类别	训练集/张	验证集/张	测试集/张
水稻鞘腐病	175	50	25
稻瘟病	103	30	15
细菌性叶枯病	1 123	321	160
褐斑病	1 134	324	162

3.2 评价指标

从准确率 (accuracy, 用 A 表示)、精确率 (precision, 用 P 表示)、召回率 (recall, 用 R 表示) 和 F_1 这 4

个方面进行模型之间的比较, 分别按照式 (9) — 式 (12) 计算。准确率是识别病害正确的样本数量占总样本数量的比例, 是最直观的评价指标。精确率是对某单一病害识别正确的样本数量占所有识别正确的样本总数的比例。召回率是对某单一病害识别正确的样本数量占该病害样本总数的比例。 F_1 是精确率和召回率的调和平均值, 体现模型的稳健性^[22]。

$$A = (N_{\text{TP}} + N_{\text{TN}}) / (N_{\text{TP}} + N_{\text{TN}} + N_{\text{FP}} + N_{\text{FN}}) \quad (9)$$

$$P = N_{\text{TP}} / (N_{\text{TP}} + N_{\text{FP}}) \quad (10)$$

$$R = N_{\text{TP}} / (N_{\text{TP}} + N_{\text{FN}}) \quad (11)$$

$$F_1 = 2 \times (P \cdot R) / (P + R) \quad (12)$$

式中: N_{TP} 为模型将正样本正确预测为正例的样本数量, N_{TN} 为模型将负样本正确预测为负例的样本数量, N_{FP} 为模型将负样本错误预测为正例的样本数量, N_{FN} 为模型将正样本错误预测为负例的样本数量。

3.3 实验结果分析

实验使用 PyTorch 框架构建 ResNet101 与 Transformer 图像文本对比学习模型, 在 Plant Wild 数据集上训练 2 次后得到预训练模型, 然后在水稻病害数据集 Rice Leaf Diseases 上进行微调训练, 与以往提出的对比学习模型进行对比, 结果见表 3。4 项指标较其他对比学习模型均有所提升。召回率衡量的是模型识别正类 (positive class) 样本的能力, 即模型正确识别出的正样本占所有实际正样本的比例。本文模型的召回率优于其他模型, 说明采用难负样本策略有效。

表 3 不同对比学习方法在水稻数据集上的结果对比

Tab. 3 Comparison of comparative learning methods on rice disease dataset

模型	准确率/%	精确率/%	召回率/%	F_1 /%
MoCo-v2 ^[23]	91.47	92.25	90.25	91.24
BYOLTMS ^[24]	93.02	93.54	92.07	92.80
DeepCluster ^[25]	90.70	90.28	91.03	90.56
本文模型	97.74	97.38	97.84	97.61

本文模型在针对水稻病害识别任务的自制数据集上, 对各类别水稻病害的识别准确率见表 4。由表 4 可知, 模型对细菌叶枯病的识别准确率最高, 达到 99.62%, 而其他 3 类识别准确率不高的原因可能是原始数据集小以及水稻病害在不同时期染病呈现相似症状不易区分, 比如水稻鞘腐病与稻瘟病在高湿度条件下都可能产生霉层, 但前期状态都为褐色斑点, 区别仅仅是前者后期为紫褐色云纹状, 后者病斑中央为灰白色, 边缘褐色。图 3 为水稻鞘腐病叶片图, 但在识别过程中, 模型将其识别为稻瘟病。

表 4 各类别水稻病害准确率
Tab. 4 Accuracy of various rice diseases

类别	准确率/%
水稻鞘腐病	98.12
稻瘟病	97.37
细菌性叶枯病	99.62
褐斑病	96.51



图 3 水稻鞘腐病叶片图
Fig. 3 Rice sheath blight leaf image

3.4 消融实验

本文提出的基于对比学习的水稻叶片病害检测模型主要包括图像模态和文本模态两部分,为了验证图像模态和文本模态在模型性能提升中的作用,设计3组消融实验,结果见表5。

表 5 消融实验结果
Tab. 5 Results of ablation experiment

图像模态	文本模态	准确率/%	精确率/%	召回率/%	F_1 /%
√	×	92.37	91.38	92.06	91.72
×	√	85.88	86.52	85.31	85.91
√	√	97.74	97.38	97.84	97.61

与文本模态相比,图像模态携带的信息更有助于提升分类的性能。最好的检测结果由图文原型模型得到,该模型可以实现 97.38%的精确率、97.84%的召回率和 97.61%的 F_1 。

4 结 语

本文提出基于图文模型和对比学习的水稻病害识别方法,该模型较单一图像模态学习到更多语义信息,显著提高了模型的识别准确度。该模型通过对比学习思想,采用无监督学习方法,缓解了传统卷积网络过分依赖数据集标签问题。实验结果证明了本文方法的有效性。在未来工作中,将尝试使用多种模态构造方式实现模态间语义对齐,挖掘模态间的共有语义信息,进一步提升模型对水稻病害任务的识别能力,同时提高泛化能力,以适用于其他下游任务。

参考文献:

[1] 全国农业技术推广服务中心. 2024 年全国农作物重大病虫害发生趋势预报[J]. 中国植保导刊, 2024, 44(1): 37–40.

[2] YOUNG T, HAZARIKA D, PORIA S, et al. Recent trends in deep learning based natural language processing[J]. IEEE Computational intelligence magazine, 2018, 13(3): 55–75.

[3] OTTER D W, MEDINA J R, KALITA J K. A survey of the usages of deep learning for natural language processing[J]. IEEE Transactions on neural networks and learning systems, 2021, 32(2): 604–624.

[4] LI H. Deep learning for natural language processing: advantages and challenges[J]. National science review, 2018, 5(1): 24–26.

[5] SLADOJEVIC S, ARSENOVIC M, ANDRLA A, et al. Deep neural networks based recognition of plant diseases by leaf image classification[J]. Computational intelligence and neuroscience, 2016(1): 3289801.

[6] ZHOU Y F, LIU D Y, ZHOU Y P. Detection of crop disease leaf based on multi-modal feature alignment[J]. Journal of Chinese agricultural mechanization, 2024, 45(7): 180.

[7] KAWASAKI Y, UGA H, KAGIWADA S, et al. Basic study of automated diagnosis of viral plant diseases using convolutional neural networks[C]//ISVC. Advances in Visual Computing: 11th International Symposium. Berlin: Springer International Publishing, 2015: 638–645.

[8] 孙俊, 谭文军, 毛罕平, 等. 基于改进卷积神经网络的多种植物叶片病害识别[J]. 农业工程学报, 2017, 33(19): 1–10.

[9] LU Y, YI S J, ZENG N Y, et al. Identification of rice diseases using deep convolutional neural networks[J]. Neurocomputing, 2017, 267: 378–384.

[10] ZHOU J, LI J X, WANG C S, et al. Crop disease identification and interpretation method based on multimodal deep learning[J]. Computers and electronics in agriculture, 2021, 189: 106408.

[11] 周一帆, 刘东洋, 周宇平. 基于多模态特征对齐的作物病害叶片检测[J]. 中国农机化学报, 2024, 45(7): 180–187.

[12] HE X T, PENG Y X. Fine-grained image classification via combining vision and language[C]//IEEE. Proceedings of the IEEE Conference on Computer Vision and

- Pattern Recognition. New York: IEEE, 2017: 5994–6002.
- [13] CHEN T, KORNBLITH S, NOROUZI M, et al. A simple framework for contrastive learning of visual representations[C]//PMLR. International Conference on Machine Learning. New York: PMLR, 2020: 1597–1607.
- [14] HE K N, FAN H Q, WU Y X, et al. Momentum contrast for unsupervised visual representation learning[C]//IEEE. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. New York: IEEE, 2020: 9729–9738.
- [15] CHUANG C Y, ROBINSON J, LIN Y C, et al. Debiased contrastive learning[J]. Advances in neural information processing systems, 2020, 33: 8765–8775.
- [16] KALANTIDIS Y, SARIYILDIZ M B, PION N, et al. Hard negative mixing for contrastive learning[J]. Advances in neural information processing systems, 2020, 33: 21798–21809.
- [17] ZHANG H Y, CISSE M, DAUPHIN Y, et al. MIXUP: beyond empirical risk management[C]//ICLR. 6th International Conference on Learning Representations (ICLR). New Orleans: ICLR, 2018: 1–13.
- [18] JIN S Y, ROYCHOWDHURY A, JIANG H Z, et al. Unsupervised hard example mining from videos for improved object detection[C]//IEEE. Proceedings of the European Conference on Computer Vision (ECCV). New York: IEEE, 2018: 307–324.
- [19] LASKIN M, SRINIVAS A, ABBEEL P. CURL: contrastive unsupervised representations for reinforcement learning[C]//PMLR. International Conference on Machine Learning. New York: PMLR, 2020: 5639–5650.
- [20] WEI T Q, CHEN Z, HUANG Z, et al. Benchmarking in the wild multimodal disease recognition and a versatile baseline[C]//ACM. Proceedings of the 32nd ACM International Conference on Multimedia. Kerrville: ACM, 2024: 1593–1601.
- [21] HUGHES D P, SALATHÉ M. An open access repository of images on plant health to enable the development of mobile disease diagnostics[EB/OL]. [2024–12–01]. <http://doi.org/10.48550/arXiv.1511.08060>.
- [22] FIALLOS J J F, BOSQUEZ G L M, BARCENES V A B, et al. Deep neural networks based recognition of paddy crop diseases by image classification[J]. Annals of forest research, 2023, 66(1): 3857–3867.
- [23] CHEN X L, FAN H Q, GIRSHICK R, et al. Improved baselines with momentum contrastive learning[EB/OL]. [2024–12–01]. <https://doi.org/10.48550/arXiv.2003.04297>.
- [24] GRILL J B, STRUB F, ALTCHÉ F, et al. Bootstrap your own latent: a new approach to self-supervised learning[J]. Advances in neural information processing systems, 2020, 33: 21271–21284.
- [25] CARON M, BOJANOWSKI P, JOULIN A, et al. Deep clustering for unsupervised learning of visual features[C]//Proceedings of the European Conference on Computer Vision (ECCV). Berlin: Springer, 2018: 132–149.

责任编辑: 郎婧