

基于多模态融合的农作物病害识别方法研究

——硕士学位论文开题报告

一、研究背景与意义

农业是国民经济的基础产业，粮食安全与农产品质量直接关系到国家发展和人民生活。农作物在生长发育过程中易遭受各类病原菌、病毒及害虫的侵袭，病害的发生不仅显著降低作物产量与品质，还常给农户带来巨大经济损失。据联合国粮农组织等机构估计，全球每年因植物病害造成的粮食损失约占作物总产量的10%—20%。及时、准确地识别农作物病害，是实现精准施药、减少农药盲目使用、保障粮食安全与农业绿色发展的关键前提。

传统的农作物病害识别主要依赖植保专家的目视观察与人工经验判断，存在主观性强、效率低、覆盖面有限等突出问题。在病害暴发初期，症状特征细微且极易与营养缺乏、药害等非生物胁迫相混淆，即便经验丰富的专家也常难以准确区分。同时，专业植保人才缺口较大，基层农技服务难以辐射到千家万户，导致许多病害往往发展至中后期才被发现，错过最佳防治窗口，造成难以挽回的损失。

近年来，随着计算机视觉与深度学习技术的快速发展，基于图像识别的农作物病害自动检测方法取得了显著进展，已成为智慧农业的重要技术支撑。然而，单一图像模态存在固有局限：其一，不同病害、不同发病阶段、不同光照与拍摄角度下，图像表观差异大，仅凭视觉特征难以完全区分表型相似或处于早期阶段的病害；其二，图像无法直接反映病原菌种类、田间温度、湿度等与病害发生密切相关的非视觉信息，识别结果缺乏足够的可解释性与鲁棒性，在复杂田间场景下识别性能易发生退化。

农作物病害的发生与发展是作物表型、病原菌与农田环境（温度、湿度、土壤、气象等）共同作用的结果，多种信息之间存在天然的互补关系。将病害图像与田间环境传感数据、病害发生的文本描述（如农技人员或农户的观察记录）、以及光谱数据等多源信息相融合，能够在视觉特征之外引入与病害发生机理直接相关的环境与语义线索，从而从整体上提升识别精度、鲁棒性与可解释性。多模态融合在自然语言处理、医学影像、遥感分析等领域已取得成熟应用，其在农业病害识别中的应用具备充分的技术可行性与广阔的研究前景。

从理论层面看，本研究探索面向农作物病害识别场景的多模态特征表征、跨模态对齐与融合方法，可为跨模态学习理论在农业视觉任务中的应用提供新的实证支撑，丰富多模态融合方法在类间相似度高、小样本、背景复杂的农业场景中的方法论体系，具有一定的理论意义。

从实践层面看，本研究构建的多模态病害识别方法有望直接服务于智慧农业中的田间智能巡检、远程诊断与精准施药决策，降低对人工经验的依赖，提升基层植保服务能力，对保障粮食安全、促进农业绿色可持续发展具有重要的现实意义。

二、国内外研究现状

（一）国外研究现状

国外对农作物病害识别的研究起步较早。早期研究主要依赖传统的图像处理与机器学习方法，通过颜色、纹理、形状等人工设计特征结合支持向量机、随机森林等分类器实现病害识别，例如基于颜色直方图、灰度共生矩阵纹理特征的稻瘟病、叶斑病识别等。此类方法的特征表达依赖人工设计，对光照变化与复杂背景干扰敏感，特征表达力有限，泛化能力不足。

随着深度学习的发展，卷积神经网络（CNN）被广泛应用于农作物病害识别。2016年，Mohanty等人利用公开的PlantVillage数据集，基于AlexNet与GoogLeNet在叶片图像病害分类任务上取得了超过99%的准确率，有力验证了深度卷积网络在该任务上的巨大潜力。此后，ResNet、DenseNet、EfficientNet以及引入注意力机制的各类网络被不断改进并用于病害识别，识别性能持续提升。近年来，Vision Transformer（ViT）及其变体凭借自注意力机制对长程依赖与全局上下文建模的优势被引入该领域，进一步推动了识别精度的提升。

在多模态融合方面，国外研究者率先探索了图像与光谱、图像与文本、图像与环境数据的融合框架。例如，有研究将高光谱成像与卷积网络结合，利用光谱信息辅助病害的早期检测；亦有工作将田间气象、温湿度等环境传感数据与图像特征进行拼接或注意力加权融合，以提升病害预测的准确性；还有学者尝试利用农作物病害的文本描述与图像构建图文联合表征。总体上，国外研究在多模态农业病害识别上已初步形成“以视觉为主、多源信息为辅”的融合思路，但针对农业场景的深度融合框架仍不成熟。

（二）国内研究现状

国内在农作物病害识别方面同样积累了丰富成果。基于深度学习的病害图

像识别研究近年快速增长，研究者围绕水稻、小麦、玉米、柑橘、茶叶等主要作物的叶部、穗部、果实病害开展了系统的研究，构建了多个公开或自建的病害图像数据集，并在田间复杂背景下实现了较高的识别精度。针对类间相似度高、样本不足等问题，国内学者引入了数据增强、迁移学习、生成对抗网络等策略加以缓解，有效提升了模型在小样本场景下的泛化能力。

在多模态与多源信息融合方面，国内研究正从单一图像识别向多源信息融合方向演进。已有工作将图像特征与农学领域知识、气象因子、植保文本等结合开展病害预测，在部分场景下表现出优于单模态方法的性能。但总体而言，国内多模态融合在农作物病害识别领域的研究仍处于起步阶段，融合框架的通用性、模态对齐的精细程度以及跨模态注意力机制的应用仍相对有限，尚未形成系统化的解决方案。

（三）研究述评

综合国内外研究现状可以看出：其一，基于深度学习的单模态图像识别已较为成熟，但受制于图像表现差异与非视觉信息的缺失，其鲁棒性与可解释性仍难以满足复杂田间场景的实用需求；其二，多模态融合虽然已在医疗、遥感等领域取得显著成效，但在农作物病害识别中仍以简单的特征拼接为主，缺乏针对“图像—文本—环境”异质模态的精细对齐与深度融合策略；其三，现有融合方法对跨模态注意力机制、模态缺失鲁棒性、多模态数据增强等关键问题的处理不够系统，融合带来的性能增益有待进一步挖掘与验证。上述不足构成本研究的主要切入点与可改进空间。

三、研究内容与目标

本研究以农作物病害识别为应用背景，面向图像单模态方法的固有局限，研究“图像—文本—环境传感”多模态信息的融合识别方法。总体目标是构建一套多模态特征提取、对齐与融合的病害识别框架，在复杂田间场景下实现比单模态方法更高的识别精度、更强的鲁棒性与更好的可解释性，完成系统原型验证，形成可推广的识别方法体系。具体研究内容包括以下五个方面。

（一）多模态数据集构建与增强。构建涵盖病害图像、病害文本描述与田间环境传感数据的多模态病害数据集。针对类间相似度高、样本不均衡与图像数量有限的问题，研究基于生成对抗网络、图像混合与颜色空间变换等的针对性数据增强方法，扩充视觉样本的多样性与鲁棒性。拟解决的关键问题：多模态数据在时间与空间上的对齐，以及小样本与类别不均衡下的数据增强。

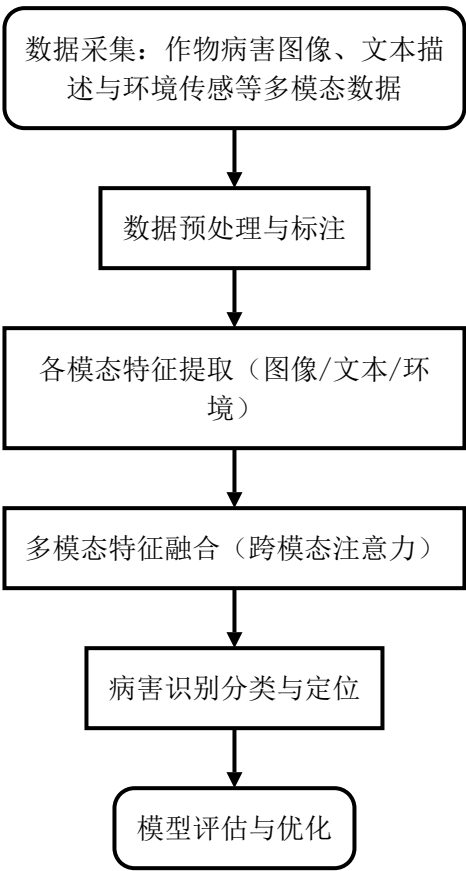
（二）各模态特征提取。分别设计图像、文本与环境传感三种模态的特征提取方法。图像模态利用改进的卷积神经网络与视觉Transformer提取多尺度视觉特征；文本模态利用预训练语言模型提取病害语义描述特征；环境模态对温度、湿度、气象等时序传感数据进行编码表示。拟解决的关键问题：各模态特征的判别力与尺度一致性。

（三）多模态特征融合策略设计。研究跨模态注意力机制与多层融合策略，设计"以图像为主、多源为辅"的深度融合框架，实现异质模态特征的自适应加权与互补性利用；同时探讨模态缺失条件下的鲁棒融合方法。拟解决的关键问题：异质模态特征间的语义鸿沟以及融合时机的选择。

（四）病害识别与定位模型。基于融合特征构建病害识别分类模型，并结合目标检测或注意力定位机制实现病害区域的定位与严重程度的估计。拟解决的关键问题：细粒度类间区分与病害发生位置的准确刻画。

（五）实验评估与优化。在公开数据集与自建数据集上开展系统的对比实验、消融实验与鲁棒性实验，科学验证多模态融合方法相对单模态基线的性能增益，并对模型进行工程优化与轻量化处理。拟解决的关键问题：融合增益的科学量化以及模型的工程实用性。

四、技术路线



五、创新点

针对现有研究在融合策略、模态对齐与数据增强等方面的不足，本研究的创新点主要体现在以下几个方面。

- 1、提出面向农作物病害识别的跨模态注意力融合策略，设计"图像—文本—环境"异质模态的自适应加权融合框架，有效缓解语义鸿沟，提升融合增益。
- 2、研究多模态特征的精细对齐方法，在时序与空间两个维度实现图像、文本与环境传感信息的协同对齐，增强多源信息的互补利用。
- 3、提出面向小样本与类别不均衡的多模态数据增强机制，并设计模态缺失条件下的鲁棒融合方法，提升模型在复杂田间场景下的泛化性与实用性。
- 4、构建"识别与定位一体化"的病害诊断方法，在病害分类的同时实现病害区域的准确刻画与严重程度估计。

六、预期成果

- 1、发表学术论文1—2篇，其中SCI/EI收录论文不少于1篇；
- 2、完成一套多模态融合农作物病害识别系统原型（模型原型），支持田间图像与多源信息的输入与诊断输出；
- 3、构建并开源一个多模态农作物病害数据集（含图像、文本与环境传感数据），为后续研究提供数据支撑。

七、三年研究进度

本研究总周期为三年，整体研究进度规划如下图所示。第一阶段（2026.09—2027.06）完成文献调研与多模态数据采集；第二阶段（2027.07—2028.06）完成数据预处理、特征对齐与融合模型设计；第三阶段（2028.07—2029.06）完成模型训练、调优与消融实验，并进行系统集成与田间验证；第四阶段（2029.07—2029.09）完成论文撰写与毕业答辩。

三年研究进度规划（2026. 09—2029. 09）

