



西北农林科技大学

博士学位论文

基于深度学习的葡萄霜霉病害表型诊断
——从实验室到田间的应用

培 养 单 位 资源环境学院

学 科 专 业 土地资源与空间信息技术

论 文 作 者 田琪

指 导 教 师 于强研究员

合作指导教师 赵刚研究员

2025 年 9 月

Dissertation Submitted to Northwest A&F University
in Partial Fulfillment of the Requirements
Doctor of Philosophy

Deep Learning-Based Phenotyping Diagnosis of
Grape Downy Mildew——
Application in Lab and Field

Colleges: Department of Resources and Environment

Major: Land Resources and Spatial Information Technology

Candidate: Qi Tian

Supervisor: Qiang Yu

Co- Supervisor: Gang Zhao

September,2025

分类号：S33

UDC：634.8

密级：

学校代码：10712

研究生学号：2021060317

西北农林科技大学博士学位论文

基于深度学习的葡萄霜霉病害表型诊断 ——从实验室到田间的应用

论文作者：田琪

指导教师：于强

指导小组：赵刚

答辩委员会：

西北农林科技大学水利与建筑工程学院何建强教授（主席）

西北农林科技大学水利与建筑工程学院李毅研究员（委员）

中国科学院大学水土保持研究所郭胜利研究员（委员）

中国农业大学资源与环境学院王靖教授（委员）

兰州大学资源环境学院马轩龙教授（委员）

答辩日期：2025年8月15日

本研究得到陕西省重点产业链项目：农作物病虫害人工智能
诊断预警研究与示范（编号：2023-ZDLNY-64）以及
广西壮族自治区重点研发项目：葡萄病虫害绿色综合防控技术
应用与示范（编号：GuiKeAB21076001）资助。

研究生学位论文的独创性声明

本人声明：所呈交的学位论文是我个人在导师指导下独立进行的研究工作及取得的研究成果；论文中的研究数据及结果的获得完全符合学术道德的有关规定，如有违反，一切后果与法律责任均由本人承担。

尽我所知，除了文中特别加以标注和致谢的地方外，论文中不包含其他人已经发表或撰写过的研究成果，也不包含为获得西北农林科技大学或其它教育机构的学位或证书而使用过的材料。与我一同工作的同事对本研究所做的任何贡献均已在论文的致谢中作了明确的说明并表示了谢意。

研究生签名：

时间： 2025 年 8 月 13 日

导师指导研究生学位论文的承诺

本人承诺：我的研究生 所呈交的学位论文是在我指导下独立开展研究工作及取得的研究成果，并严格遵守学术道德的有关规定。如有违反，我愿接受按学校有关规定的处罚处理并承担相应导师连带责任。

导师签名：

时间： 2025 年 8 月 13 日

关于研究生学位论文使用授权的说明

本学位论文的知识产权归属西北农林科技大学。本人及导师同意学校保存或向国家有关部门或机构送交论文的纸质版和电子版，允许论文被查阅和借阅；同意学校将本学位论文的全部或部分内容授权汇编录入《中国博士/硕士学位论文全文数据库》和《中国学位论文全文数据库》进行出版，并享受相关权益。

本人保证，在毕业离开（或者工作调离）西北农林科技大学后，发表或者使用本学位论文及其相关研究成果时，将以西北农林科技大学为第一署单位名称，否则，愿意按《中华人民共和国著作权法》等有关规定接受处理并承担法律责任。

任何收存和保管本论文各种版本的其他单位和个人(包括研究生本人)未经本论文作者的导师同意,不得有对本论文进行复制、修改、发行、出租、改编等侵犯著作权的行为,否则,按违背《中华人民共和国著作权法》等有关规定处理并追究法律责任。

(保密的学位论文在保密期限内,不得以任何方式发表、借阅、复印、缩印或扫描复制保存)

研究生签名: 时间: 2025 年 8 月 13 日

导师签名: 时间: 2025 年 8 月 13 日

摘要

葡萄霜霉病是影响葡萄产量与品质的主要病害之一，严重时将显著削弱植株活力、降低果实品质，并威胁葡萄产业的可持续发展。当前防治手段高度依赖化学药剂，不仅成本高昂，还可能导致环境污染。抗病品种的选育与病害实时监测是实现绿色、高效防控的关键路径，而二者均依赖对病害类型的准确识别与病情程度的量化。然而，现有方法多依赖人工判读，效率低下且主观性强，难以满足高通量表型分析需求。传统图像处理方法泛化能力差、鲁棒性弱。近年来，卷积神经网络（CNN）及其衍生的目标检测（如 YOLO v8）与图像分割模型（如 UNet）在图像识别与分割领域取得显著进展，同时，轻量化视觉分割基础模型（如 Mobile SAM）和视觉语言模型（如 Qwen-2.5-VL-3B）也为农业表型智能分析提供了新思路。然而，这些模型在病害智能识别中的应用仍面临如表型数据稀缺、模型解释性差、田间适应性弱等挑战。

本研究围绕葡萄霜霉病的智能识别与定量评估，分析了深度学习技术在实验室与田间环境中的适用性与优化策略，重点探究了以下三个问题：

（1）抗性筛选任务中的高通量检测问题：针对传统抗性鉴定手段成本高、效率低的问题，我们基于手机拍摄的叶盘霜霉病图像，开发了一种低成本、高通量的抗性检测系统。该系统集成 YOLO v8n 进行葡萄叶盘定位，Mobile SAM 实现背景剔除，UNet 完成霜霉病产孢区域的精准分割。其中，YOLO v8n 在实验图像中的 mAP@50 超过 99%，优于 Faster RCNN；UNet 分割模型在 mIoU@50 上达到 96% 以上，显著优于其他 CNN 模型。与端到端的 Mask RCNN 相比，序列化模型的检测率与分割精度更高，病害严重程度预测结果与人工标注高度一致（ $R^2=0.99$ ），显著减少主观误差与人力成本，为葡萄抗病育种与药效评价提供了高效工具。

（2）模型可解释性问题的解决路径：为提升病害识别模型的可解释性与用户信任度，我们整合田间实拍与公开葡萄病害数据，共收集 2213 张葡萄病害图像，借助视觉语言模型（Visual-Language Model, VLM）生成并专家校正病害特征描述文本，构建高质量图文配对数据集。在此基础上，我们采用 LoRA 技术对 Qwen-2.5-VL-3B 进行微调，构建定制化模型 FT-Qwen-VL。实验结果表明，FT-Qwen-VL 在 BLEU-4 (67.82%)、ROUGE-1 (71.54%)、ROUGE-2 (48.18%)、ROUGE-L (61.49%) 等指标上，均显著优于 BLIP2、DeepSeek-VL2、Describe Anything、Kimi-VL-A3B 等主流 VLM。其在病斑属性提取（颜色、形状、位置、面积）上的准确率分别达到 80.04%、80.49%、29.27%、86.92%，验证了微调 VLM 在农业场景下的可定制化能力，为可解释的田间病害识别提供了关键技术参考。

(3) 田间复杂背景下的鲁棒分割问题：我们采集 687 张葡萄高质量田间叶片图像（其中 338 张感病、349 张健康），构建用于田间霜霉病严重程度评估的图像数据集。结合 CNN 分割模型与 Mobile SAM，我们设计了一个基于提示的分割框架。实验显示，UNet++在田间数据集上的 mIoU@50 达到 97.49%（验证集）和 96.93%（测试集），F1 得分分别为 98.72%和 98.42%，在所有对比模型中表现最佳。引入 Mobile SAM 进一步提升了分割精度，边界框提示方式对模型性能改善最为显著。基于优化后的分割结果，构建了田间病害严重程度评估模型，在验证集与测试集上的 R^2 分别为 0.93 与 0.94，显示出良好的拟合与泛化能力，证明基于视觉基础模型的提示机制可显著提升病害分割与量化性能，具有良好田间适应性。

本研究基于手机图像，在实验室与田间环境中分别采集、标注葡萄霜霉病叶片数据，构建覆盖实验室与田间表型的图像数据集，并以深度学习技术的应用与评估为核心，系统分析了相关模型在抗性筛选、特征解释与田间分割任务中的性能与适用性。研究成果可为葡萄抗病育种和精准农药施用提供关键技术支撑，有助于推动葡萄病害智能管理体系的构建，实现农业绿色高效发展目标。

关键词：葡萄霜霉病；表型鉴定；深度学习；田间监测

ABSTRACT

Downy mildew is one of the most significant diseases affecting grape yield and quality. In severe cases, it can drastically weaken plant vigor, degrade fruit quality, and threaten the sustainable development of the grape industry. Current disease control strategies rely heavily on chemical treatments, which are not only expensive but also pose environmental risks. Breeding resistant cultivars and implementing real-time disease monitoring are critical for achieving green and efficient disease management. However, both approaches depend on accurate identification of disease types and quantification of disease severity. Existing methods are largely based on manual inspection, which is inefficient and highly subjective, making them unsuitable for high-throughput phenotyping. Traditional image processing techniques generally suffer from poor generalizability and weak robustness. In recent years, convolutional neural networks (CNNs) and their derivatives—such as YOLO, UNet—have made substantial progress in image recognition and segmentation. Meanwhile, lightweight visual foundation segmentation models like Mobile SAM, as well as visual-language models (e.g., Qwen-2.5-VL-3B), offer promising new approaches for intelligent agricultural phenotyping. However, their application to plant disease recognition still faces challenges such as limited annotated phenotypic data, poor model interpretability, and insufficient adaptability to field conditions.

This study focuses on the intelligent identification and quantitative evaluation of grape downy mildew, analyzing the applicability and optimization of deep learning techniques in both laboratory and field settings. Specifically, the contributions of this research can be summarized in three aspects:

(1) High-throughput detection for resistance screening: To address the high cost and inefficiency of traditional resistance evaluation methods, we developed a low-cost, high-throughput detection system based on smartphone-acquired images of leaf discs inoculated with downy mildew. The system integrates YOLO v8n for leaf disc localization, Mobile SAM for background removal, and UNet for precise segmentation of the sporulation regions. YOLO v8n achieved a mAP@50 exceeding 99% in laboratory images, outperforming Faster RCNN. UNet achieved a mIoU@50 above 96%, significantly surpassing other CNN segmentation models. Compared with the end-to-end Mask RCNN model, this modular pipeline delivered superior detection and segmentation accuracy. The predicted disease severity showed strong agreement with manual annotations ($R^2 = 0.99$), significantly reducing labor and subjective

bias. This system offers an effective tool for resistance evaluation and pesticide efficacy testing in grape breeding programs.

(2) Enhancing model interpretability through visual-language alignment: To improve interpretability and user trust in disease recognition models, we collected a total of 2,213 grape disease images from both in-field photography and public datasets. Using a visual-language model (VLM), we generated disease feature descriptions that were then refined and validated by domain experts to build a high-quality image-text paired dataset. Based on this dataset, we fine-tuned Qwen-2.5-VL-3B using LoRA, resulting in a customized model named FT-Qwen-VL. The model achieved substantial improvements over existing VLMs—including BLIP2, DeepSeek-VL2, Describe Anything, and Kimi-VL-A3B—on multiple benchmarks: BLEU-4 (67.82%), ROUGE-1 (71.54%), ROUGE-2 (48.18%), and ROUGE-L (61.49%). In attribute extraction tasks, FT-Qwen-VL achieved accuracies of 80.04% (color), 80.49% (shape), 29.27% (position), and 86.92% (area). These results validate the customizability and effectiveness of fine-tuned VLMs for domain-specific knowledge extraction in agriculture, providing key support for interpretable disease recognition in field settings.

(3) Robust segmentation under complex field conditions: To address the challenge of accurate segmentation in field environments with complex backgrounds, we collected 687 high-resolution grape leaf images in the field, including 338 diseased and 349 healthy samples, and constructed a dataset for evaluating disease severity. A prompt-based segmentation framework was developed by integrating CNN models with Mobile SAM. Among all tested models, UNet++ showed the best performance, achieving mIoU@50 scores of 97.49% (validation set) and 96.93% (test set), with F1 scores of 98.72% and 98.42%, respectively. The incorporation of Mobile SAM further enhanced segmentation accuracy, with bounding box prompts demonstrating the most significant improvement. Based on the optimized segmentation results, we constructed a field-based severity assessment model, which achieved R^2 values of 0.93 and 0.94 on the validation and test sets, respectively, demonstrating strong model fit and generalization. These findings confirm that prompt-guided vision foundation models can significantly improve lesion segmentation and quantification performance under field conditions, supporting robust in-field deployment.

In conclusion, this study leveraged smartphone-captured images to construct comprehensive laboratory and field datasets for grape downy mildew, and evaluated deep learning models for resistance screening, lesion description, and field segmentation. The results provide critical technical support for resistant cultivar development and precision fungicide application, promoting the development of intelligent and sustainable disease

ABSTRACT

management in viticulture.

KEY WORDS: Grape downy mildew; Disease phenotyping; Deep learning; Field scouting

目录

摘要	I
ABSTRACT	III
第一章 绪论	1
1.1 研究背景、目的和意义	1
1.2 国内外研究进展	3
1.2.1 葡萄霜霉病侵染过程与防控技术	3
1.2.2 高通量葡萄病害表型鉴定研究进展	8
1.2.3 田间病害识别与诊断研究进展	10
1.2.4 田间病害特征描述研究进展	11
1.2.5 田间病害严重程度评估研究进展	13
1.3 研究主要存在的问题	14
1.4 研究内容和技术路线	15
1.4.1 研究内容	15
1.4.2 技术路线	16
第二章 材料和方法	18
2.1 数据来源	18
2.1.1 实验室葡萄叶盘接种霜霉病照片数据	18
2.1.2 田间葡萄健康叶片和霜霉病叶片照片数据	20
2.1.3 开源病害数据集中葡萄病害照片数据	23
2.2 数据预处理过程	23
2.2.1 叶盘接种霜霉病照片标注	23
2.2.2 葡萄病害特征图文对比数据标注	24
2.2.3 田间葡萄健康和霜霉病叶片标注	27
2.3 训练和评估的数据集划分	28
2.3.1 葡萄叶盘霜霉病数据集划分	28
2.3.2 葡萄病害图文配对数据集划分	29
2.3.3 田间葡萄健康和霜霉病叶片数据集划分	30
2.4 研究中主要采用的方法和模型	30
2.4.1 深度学习基本理论与模块设计	30

2.4.2 YOLO v8 系列目标检测模型	36
2.4.3 UNet 以及 UNet++ 图像分割模型	39
2.4.4 Mobile SAM 提示分割模型	40
2.4.5 Qwen-2.5-VL-3B 视觉语言模型	42
第三章 SporeScan: 高通量叶盘霜霉病表型鉴定的序列化模型	43
3.1 概述	44
3.1.1 SporeScan 模型开发	44
3.1.2 试验平台和配置	44
3.2 模型开发中的对比实验设计	45
3.2.1 比较 YOLO v8 系列模型与其他目标检测模型定位叶盘的表现	45
3.2.2 比较 UNet 模型与其他图像分割模型分割叶盘和产孢区域的的表现	46
3.2.3 比较使用 Mobile SAM 去除背景后分割的对不同部位分类的影响	47
3.2.4 比较 SporeScan 与 Mask RCNN 对叶盘霜霉病检测分割的表现	47
3.3 叶盘霜霉病严重程度的评估	48
3.4 模型预测的产孢分布与人工标注的双盲实验	48
3.5 结果和分析	49
3.5.1 YOLO v8 系列模型在叶盘定位的结果	49
3.5.2 分割模型性能: UNet 和 Mobile SAM	50
3.5.3 SporeScan 与 Mask R-CNN 的性能对比	53
3.5.4 SporeScan 模型在病害严重程度预测中的性能表现	54
3.5.5 手动标注与模型生成的偏好双盲试验结果	55
3.6 讨论	56
3.6.1 序列化 SporeScan 优于端到端的 Mask R-CNN	56
3.6.2 SporeScan 在自动化葡萄霜霉病严重程度评估中的优越性	57
3.6.3 研究的不足与展望	58
3.7 本章小结	58
第四章 基于微调 Qwen-VL 的葡萄病害图像描述生成方法比较	59
4.1 概述	60
4.1.1 Qwen-VL 模型介绍	60
4.1.2 使用 LoRA 进行模型微调得到 FT-Qwen-VL	60
4.1.3 模型评估指标	60
4.1.4 实验平台和配置	62
4.2 对比实验设计	62

4.2.1 不同微调参数对 Qwen-VL 病害特征描述生成的性能影响	62
4.2.2 比较 FT-Qwen-VL 与其他视觉语言模型对病害特征描述生成的表现	62
4.2.3 比较视觉语言模型对葡萄病害不同的特征描述的准确率	63
4.3 结果和分析	63
4.3.1 不同参数微调 Qwen-VL 对病害特征描述生成的表现	63
4.3.2 FT-Qwen-VL 对比其他视觉语言模型病害特征描述生成的结果	65
4.3.3 FT-Qwen-VL 与其他视觉语言模型病害特征描述生成的准确率	65
4.3.4 可视化分析示例展示和分析	66
4.4 讨论	68
4.4.1 定制微调视觉语言模型可以显著提升病害特征描述以增强解释性	68
4.4.2 多模态视觉语言大模型辅助农业大数据信息提取	69
4.5 本章小结	69
第五章 基于提示引导的葡萄病害分割方法及严重程度评估	71
5.1 概述	72
5.1.1 DMS-Prompter: 整合 UNet++与 Mobile SAM 的霜霉病害分割模型	72
5.1.2 训练 UNet++模型预测叶片和掩膜	73
5.1.3 模型评估指标	73
5.1.4 实验平台和配置	74
5.2 对比实验设计	74
5.2.1 DMS-Prompter 与经典分割模型的性能比较	74
5.2.2 不同提示策略对 DMS-Prompter 分割表现的影响	75
5.2.3 对比 DMS-Prompter 对葡萄叶片霜霉病严重程度评估	75
5.3 实验结果和分析	76
5.3.1 DMS-Prompter 与经典分割模型的性能比较	76
5.3.2 不同提示策略对 DMS-Prompter 分割表现的影响	77
5.3.3 病害可视化示例展示和分析	78
5.3.4 模型对葡萄叶片霜霉病严重程度评估	80
5.4 讨论	80
5.4.1 整合 CNN 和 Mobile SAM 可以实现病斑的细致分割	81
5.4.2 优化提示策略可以增强分割效果以提高病害评估	81
5.4.3 准确的病害记录有助于提升病害预警水平和优化田间管理	82
5.5 本章小结	83
第六章 结论与展望	84

6.1 主要结论	84
6.2 主要进展和创新点	85
6.3 研究局限与展望	85
6.3.1 研究局限	85
6.3.2 未来展望	85
参考文献	87
附录	98
致谢	103
个人简历	105

绪论

1.1 研究背景、目的和意义

葡萄是一种多年生藤本果树，被广泛应用于酿酒和鲜食消费，是全球最重要的经济作物之一，至2020年，全球种植面积达约740万公顷，产量约为8000万吨(OIV 2022)。近年来，我国葡萄种植面积持续增长，至2020年已达70万公顷以上，年产量约为1400至1450万吨，为全球葡萄种植面积最大的国家(李小红等, 2021; 刘柯含等, 2024)。葡萄霜霉病是一种普遍且具有显著破坏性的葡萄病害，自1834年在美国东部地区被发现后，随着葡萄品种的引种和贸易活动（例如，为防治根瘤蚜病而引进抗根瘤蚜砧木）的进行，逐渐蔓延至全球，对葡萄种植业及葡萄酒产业造成了重大影响(Fontaine et al. 2020)。我国首次对该病害的记录可追溯至1899年，此后，该病在我国主要葡萄产区逐渐流行，尤其是在多雨潮湿的地区，由于环境条件有利于病原菌的传播与侵染，常引发大面积暴发，有时候造成20%-80%的产量损失（张玮等 2020）。

在全球气候变化背景下，特别是高温高湿等气候条件的增强，进一步加剧了葡萄霜霉病的发生风险，包括首次发生时间提前、发生频率增加，加快病情发展等，显著提升了葡萄产量和品质的威胁（Van Leeuwen et al. 2024）。气候变化的影响可能导致平均增加2次的杀菌剂施用次数，不仅将进一步提高生产成本，也带来了潜在的环境污染和食品安全风险（Francesca et al. 2006）。培育具有抗霜霉病特性的葡萄品种，并优化杀菌剂施用频率与田间病害管理措施，对于提升产量、改善果实品质，实现葡萄产业的可持续发展具有重要意义（Chen et al. 2020; Magon et al. 2023）。

在实验室中，常采用叶盘试验法来评估不同葡萄品种的抗霜霉病能力及各类杀菌剂的防效（Qiu et al. 2022; Qu et al. 2021）。而在田间环境下，则需实现对霜霉病发生的精准识别与严重程度评估，以便据此优化病害管理策略，包括杀菌剂类型的选择（如保护性或治疗性药剂）（Gutiérrez et al. 2021; Kunova et al. 2021）。进一步而言，若能在田间实现对病害发生与发展进程的动态监测，并结合预测模型进行分析，可有效指导施药时机的决策，从而最大限度保障葡萄的产量与品质（Bove et al. 2020; Rossi et al. 2008）。综上所述，无论是抗性筛选、药剂评估，还是田间管理策略的制定与优化，均依赖于对霜霉病类型的准确识别及其严重程度的量化评估。然而，在实际应用中，如何高效、准确地获取病害表型信息，仍是当前面临的重要技术挑战（Song et al. 2021）。

传统的葡萄霜霉病抗性筛选通常依赖人工表型鉴定，过程繁琐、效率较低，且结果易受人为主观因素影响，难以满足大规模育种筛选的需求。近年来，高通量表型分析系统为病害表型的自动化与精准量化提供了有效手段。这些系统通过集成多种传感器，获取病害部位的光学信息，从而实现准确表征，并可结合基因组学促进抗病育种（Song

et al. 2021; Yang et al. 2020)。其中,图像处理算法是表型数据分析的核心环节。然而,在葡萄叶盘霜霉病抗性鉴定过程中,传统的目标定位与图像分割算法,如霍夫圆监测和阈值分割等,往往依赖手工参数设置,泛化能力较弱(Hernández et al. 2022)。在利用深度学习技术处理葡萄叶盘霜霉病的相关研究中,普遍采用显微镜拍摄的高分辨率图像作为数据源。尽管此类图像能够提供丰富的细节信息,但同时也导致了显著的计算资源消耗及数据采集和标注成本的增加(Qiu et al. 2022; Zender et al. 2021)。

在田间环境下,实现对霜霉病发生的精准识别具有较大挑战,主要由于自然光照变化、复杂背景干扰以及病斑形态相似等因素影响,导致识别难度显著增加(Barbedo 2018a)。传统病害识别主要依赖有经验的农户或专家进行目测判断,方法主观且准确性受限,常导致杀菌剂喷洒时机延误。传统的图像处理算法在特征选择上依赖手动设计,且在复杂背景条件下的特征提取效果不佳,限制了其在实际田间环境中的应用表现(Waghmare et al. 2016; Wang et al. 2012)。近年来,利用手机等移动设备和深度学习技术进行病害自动识别成为研究热点,特别是卷积神经网络(CNNs),在背景简单的数据集上取得了较高的准确率(Ramcharan et al. 2019; Too 2019)。然而,在面对复杂背景的田间照片时,CNNs病害识别准确率显著下降(Ferentinos 2018; Tian et al. 2024)。更重要的是,在现实世界中,农业专家通常根据作物的物理特征(即颜色、形状、纹理等)进行病害诊断,然而,基于CNNs病害识别模型仅提供病斑的位置信息和类别,缺乏对病斑物理特征的详细描述,导致用户无法完全信任检测结果(Sun et al. 2023; Zhang et al. 2024)。尽管一些研究尝试结合图像字幕模型和图文生成技术以增强病害特征的识别与解释能力,但由于病害图文数据稀缺及文本生成模型能力有限,限制了病害识别模型的田间的可解释性和适用性(Liang et al. 2024; Sun et al. 2023)。

在田间诊断葡萄霜霉病时,除了准确识别病害类型外,还需对其严重程度进行精确量化。传统方法多采用阈值分割、区域生长等图像处理技术,以区分病斑区域与健康叶片区域(Jothiaruna et al. 2019; Ma et al. 2018)。近年来,尽管CNNs在分割性能上取得了显著提升,但其效果仍然受到复杂背景的影响,如叶片重叠和背景中相似特征的干扰,进而增加了严重程度评估的不确定性(Rossi et al. 2022; Zhang et al. 2024)。因此,许多研究采用了两阶段的方法,即先对图像中的叶片区域进行精确分割,再在叶片区域内进行病斑识别与分割。该策略在一定程度上降低了背景干扰的影响,相对提升了整体分割性能和表型评估(Liu et al. 2024; Wang et al. 2021)。最近,视觉基础模型,特别是Segment Anything Model (SAM),由于具备提示性引导及强大的零样本分割泛化能力,为解决复杂背景下的病斑分割问题提供了潜在的解决方案(Kirillov et al. 2023a)。已有部分研究开始探索 SAM 在农业病害识别中的应用潜力,尤其是在复杂背景下的目标分割任务中展现出良好性能(Carraro et al. 2023; Nguyen et al. 2023;

Williams et al. 2024)。然而，由于其模型体量较大、计算资源需求高，难以直接部署于资源受限的田间环境，限制了其在实际场景中的广泛应用。

因此，当前病害表型研究主要面临以下三个方面的挑战：（1）传统算法泛化能力有限，且病害表型数据的采集与处理成本较高，亟需开发轻量级且高精度的自动诊断模型，以降低计算资源消耗与数据依赖，提升系统在实际应用中的效率。（2）基于CNNs的田间病害识别方法往往缺乏对病斑特征的可解释性描述，导致用户无法完全信任检测结果，需要进一步评估大视觉语言模型对病害特征描述的生成能力和适用性，以增强病害识别结果的可解释性和实用性。（3）田间病害分割受到复杂背景的影响，基于CNNs的分割网络对边界处理方面能力受限，尽管SAM提供了潜在的工具但受限于其高计算开销，田间部署困难，需要进一步评估其轻量化版本的在田间实际表型诊断中的适用性、稳定性与性能表现。因此，针对上述问题，开展葡萄霜霉病的识别与表型评估研究，对于提升抗性品种筛选效率和优化精准田间管理策略具有重要意义，并可为相关研究提供可靠的数据支持与方法借鉴。

1.2 国内外研究进展

1.2.1 葡萄霜霉病侵染过程与防控技术

1.2.1.1 葡萄霜霉病侵染过程及危害

葡萄霜霉病是由葡萄生单轴霉（*Plasmopara viticola*）引起的，它的生活史如所图1-1所示。该病原菌以卵孢子的形式在葡萄落叶和葡萄藤的凋落层中越冬。春季来临时，卵孢子萌发产生动孢子囊。当动孢子囊在湿润叶片表面停留一定时间后，便释放出具有鞭毛的游动孢子。游动孢子能主动游动至叶片气孔附近，萌发并形成芽管，随后穿透气孔进入叶片组织，形成附着胞和吸器，进而发展成菌丝体，完成初次侵染。菌丝体在叶片细胞间扩展，并通过瘤状吸器侵入寄主细胞以吸收养分。经过10至20天的潜育期，随着菌丝体的生长繁殖，在叶片背面形成白色的霜霉状霉层，即病原菌的孢子囊及其孢子囊梗。在多雨高湿的环境中，孢子囊不断释放游动孢子，引发继发侵染，导致该病害在整个生长季节内多次暴发流行（Burruano 2000）。

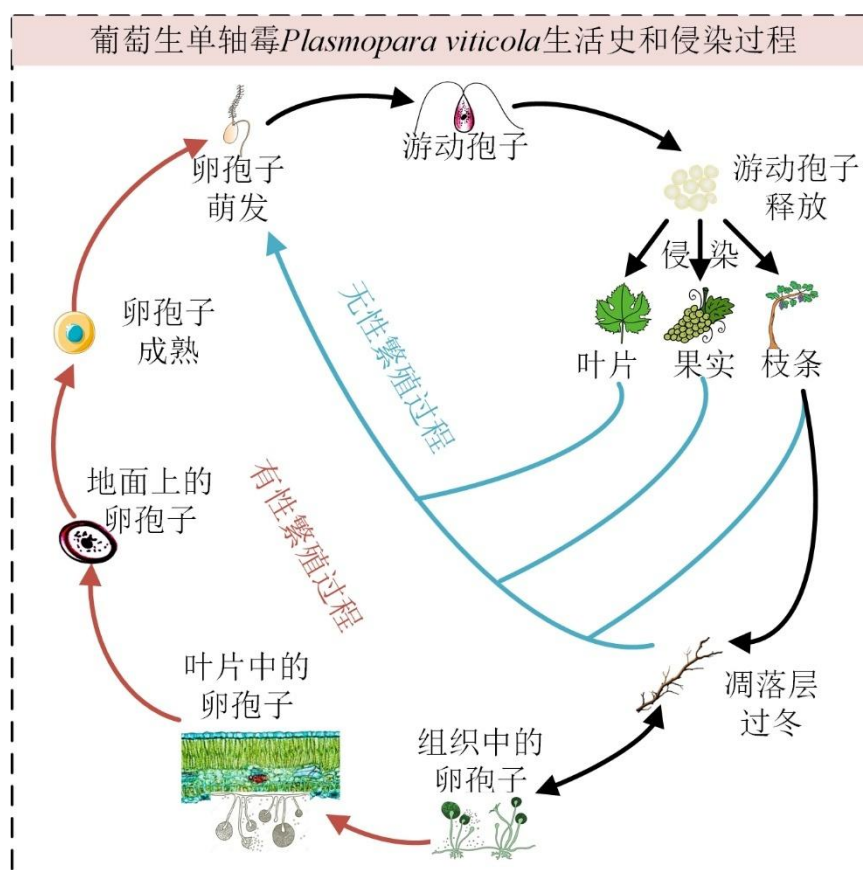


图 1-1 由葡萄生单轴霉 (*Plasmopara viticola*) 引起的霜霉病的病害循环示意图。红色箭头指的是有性繁殖过程，蓝色箭头指的是无性繁殖过程。

Figure 1-1 The diagram illustrating the life cycle of downy mildew caused by *Plasmopara viticola*. Red arrows indicate the sexual cycle, while blue arrows indicate the asexual cycle.

葡萄生单轴霉包含有性繁殖和无性繁殖两个过程。有性繁殖主要发生在寄主叶片的海绵组织中，偶尔也可在栅栏组织中进行，形成卵孢子。卵孢子在适宜条件下萌发，产生孢子囊，并释放游动孢子完成初次侵染。无性繁殖则通过孢子囊直接释放游动孢子，实现快速传播和再侵染 (Bove et al. 2020)。该病菌生长的最低温度为 9℃-10℃，高于 34℃-35℃ 则抑制病原体 (Francesca et al. 2006)。在阳光直射的条件下，孢子囊在 15 分钟内就会死亡。

葡萄霜霉病的发生显著受到气象因素的影响，其中，降雨量与气温是影响病害发展的关键环境因子。在葡萄的生长周期内，一旦存在适宜病原菌生长与扩散的气象条件，病害便可能经历多次侵染事件，进而导致疾病的广泛流行 (Rossi et al. 2008)。其主要危害葡萄叶片，使得叶片变枯萎，同时也会对新梢、花序、穗轴及幼果造成危害 (王颖 2021)。叶片感染霜霉病会导致早期枯萎脱落、新梢生长受阻、果实产量减少以及品质劣化 (Burruano 2000)。

霜霉病对葡萄叶片和果实的感染如图 1-2 所示。对于叶片而言，侵染初期会在

叶片正面出现边缘不清晰的油渍状小斑点，这些斑点会逐渐扩大成不规则形黄褐色的大斑。在潮湿条件下，病叶的背面会长出明显的霜霉菌层，这是病原菌的孢囊梗和孢囊。随着病情的发展，叶片上会出现浅黄褐色斑块，多个病斑可能会连成一片，导致叶片焦枯，最终导致叶片提前脱落。在生长季节后期，病斑多而小，受叶脉的限制而成多角形，使葡萄叶片变成“花叶”。随着病情的发展，叶片上出现浅黄褐色斑块，病斑背面附着许多浅白色的霜霉物质，最终导致叶子逐渐变黄、枯死。这一侵染循环可以在短时间内重复发生，导致病害的迅速蔓延。对于果实，在适宜的温湿环境下，霜霉病的游动孢子先吸附在幼果表皮，并萌发形成芽管，穿透角质层后在果肉开始定殖。随后，菌丝在细胞间隙中延伸，早期果实表面无明显病变，但病原体已迅速扩散。随着感染进一步发展，果皮逐渐显现出由黄褐至黑褐的斑块，果实停止发育，质地变软、水分流失，最终干瘪甚至脱落。此时病斑部位再次形成孢子，借助风雨传播至邻近未成熟组织，反复侵染并扩展为显著病害中心，若在果粒膨大阶段遭受侵染，常表现为软腐褐变，并于短时间内干缩脱落(安玉善等, 2018)；当果实进入着色期，角质层增厚，且抗性增强，通常不再发生感染。

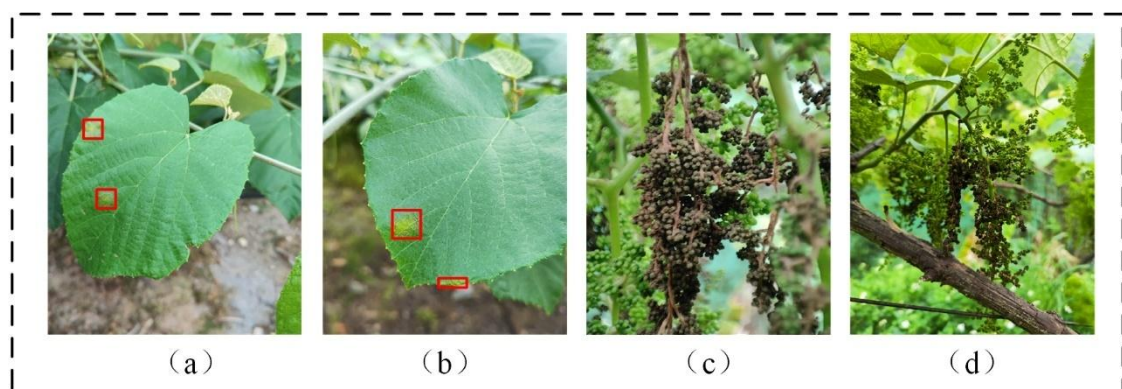


图 1-2 田间采集的葡萄霜霉病叶片病斑数据以及霜霉病危害果实的情况。(a) 和 (b) 表示早期叶片上的黄色油斑。(c) 和 (d) 显示了霜霉病导致葡萄幼果果皮褐变、组织软化甚至脱落。

Figure 1-2 Field-collected data of grape downy mildew symptoms on leaves and infected fruits. (a) and (b) show yellowish oil spots on grape leaves at the early stage of infection. (c) and (d) illustrate the browning of the skin, tissue softening, and even abscission of grape berries caused by downy mildew.

在田间调查中，常常根据病叶感染的不同程度，可以分级别评估其发生严重程度。当叶片无任何病斑时，判定为 0 级；病斑面积占整片叶面积的比例在 5% 以内时为 1 级；若病斑覆盖范围在 6% 至 25% 之间，则归为 3 级；病斑面积达 26% 至 50% 时为 5 级；当其占比升至 51% 至 75%，划分为 7 级；而超过 75% 的病斑面积则定为 9 级（刘会彪 2024）。在本研究中，我们收集了不同感染程度的葡萄霜霉病叶片，如图 1-3 所示。

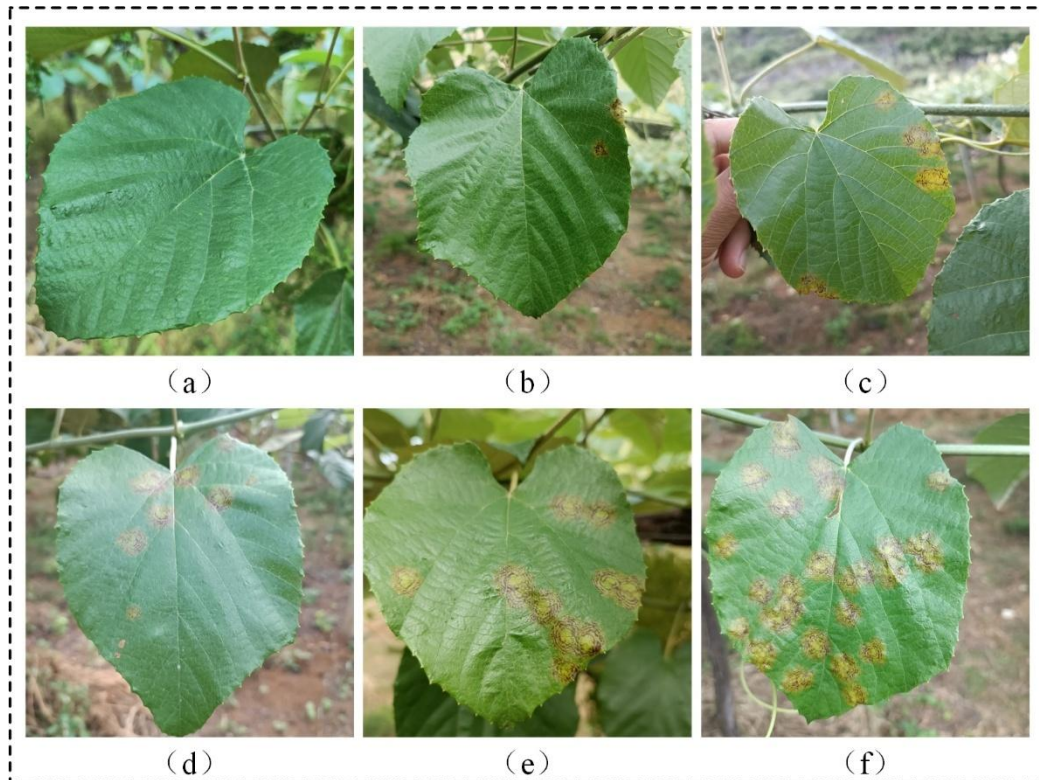


图 1-3 毛葡萄霜霉病不同发展阶段的黄色油斑照片。从上至下，从左至右黄色油斑面积逐渐增大，表示严重程度也逐渐增大。

Figure 1-3 Images of yellowish oil spots at different developmental stages of downy mildew.

1.2.1.2 霜霉病抗性育种与田间病害管理优化

当前，针对霜霉病的防治策略主要依赖于化学杀菌剂的喷洒，涵盖了保护性杀菌剂（例如波尔多液、代森锰锌等）与内吸性（治疗性）杀菌剂（例如甲霜灵、氟噻唑吡乙铜等）(李博雅等, 2023; 毕秋艳等, 2018)。保护性杀菌剂需在病害发生前施用，其主要作用机制是在植物叶片表面形成一层防护膜。而内吸性杀菌剂则主要用于病原体侵入并定植后。然而，化学杀菌剂的频繁和大量使用，若缺乏科学管理，易导致毒性残留、环境污染以及病原体抗药性问题的产生，这些因素严重威胁土壤生态健康、自然环境及人类健康。此外，农药的过度或不当使用不仅会提高农业生产成本，还可能造成环境污染，带来潜在的健康风险（Richard et al. 2022）。

葡萄霜霉病的绿色防控技术主要包括培育抗性品种、农业管理防治、化学防治以及综合病害管理（Integrated Pest Management, IPM）等(于舒怡等, 2018; 李博雅等, 2023)。抗性品种的培育基于对不同品种抗性差异的评估，进而开展基因调控工作。农业防控措施则侧重于优化栽培管理，如改善通风透光条件、采用遮雨棚、科学排水与施肥以及及时清理果园残叶，以减少病原菌传播。化学防治方面，则强调使用新型绿色环保药剂，逐步替代传统的高毒性化学农药。

在抗性品种培育过程中，利用抗病种质通过常规杂交育种仍然是目前的主要途径

(付晴晴等, 2019)。虽然该方法技术成熟, 但存在周期长、效率低、依赖经验等不足。近年来, 随着二代测序技术、表型组学和机器学习等多学科技术的发展, 融合多领域手段的现代育种正成为趋势 (杨文庆等 2022; Yang et al. 2020)。其中, 数量性状基因座定位 (QTL) 和全基因组关联分析 (GWAS) 技术的应用, 为抗性品种培育提供了更为高效的策略。GWAS 利用全基因组范围的标记, 快速筛选与抗性相关的基因; 而 QTL 定位则通过基因组标记与性状关联分析, 精确定位控制抗性状的基因区域。例如, 在葡萄抗霜霉病品种的培育中, 实验室的病菌接种试验是鉴定不同品种叶片抗性的重要手段, 通过人工接种霜霉病菌, 可以准确评估品种抗性, 为基因工程改良和育种工作提供依据 (Qu et al. 2021)。这些新兴技术的结合极大加快了葡萄抗霜霉病品种的培育进程。

基于数据驱动模型的田间病害管理策略优化, 也是当前研究的热点方向之一, 通过可靠的模型来预测葡萄霜霉病发生的规律和风险, 对于葡萄病害管理至关重要 (Velásquez-Camacho et al. 2022; Volpi et al. 2021)。早期的经验模型通过统计资料来进行预测, 如 3-10rules, 而机理模型则考虑了病原菌、寄主植物和环境的交互, 来模拟和预测葡萄霜霉病的发生和发展, 包括瑞士的 Vinemild 模型、美国的 DMCAST 模型, 以及意大利的 PLASMO 模型(张玮等, 2020)。而国内的研究起步较晚, 目前主要使用气象因子和空间信息技术 (GIS), 或者结合机器学习如支持向量机等对葡萄霜霉病进行预测 (李嘉泓等, 2019; 于舒怡等, 2024; 李文学等, 2017; 吴宁等, 2020)。近年来, Rossi 等开发了霜霉病感染过程的病害模型, 模拟了初侵染过程中的各个阶段, 包括卵孢子的成熟与萌发、游动孢子释放、和扩散, 游动孢子定殖和侵染等 (Bove et al. 2020; Rossi et al. 2008)。该模型在欧洲很多地区得到相对较好的应用, 同时 Rossi 模型在中国北方地区也能够及时的预测霜霉病的发生, 并强调了控制始发感染对葡萄霜霉病管理具有重要的作用, 由于在继发感染中的病菌多数由始发感染转变 (Gobbin et al. 2005; Hui et al. 2022)。

这两类病害防控技术的实施均依赖于对病害发生与感染过程的精准量化与诊断。然而, 在当前无论是实验室研究还是田间应用中, 病害的监测与评估仍主要依靠人工完成。该方式不仅耗时耗力, 而且存在显著的主观性。人工监测的不确定性可能导致对品种抗性水平的评估不准确, 进而影响育种过程, 同时也会干扰田间病害发生情况的客观判断。此外, 实验室和田间的表型数据获取与解析明显滞后于基因组数据和环境数据的发展 (Hein et al. 2021; Song et al. 2021)。表型采集面临缺乏专业设备、成本高昂, 以及解析算法鲁棒性差等挑战。比如, 传统的特征提取算法通过手动来设计, 然后输入随机森林或者支持向量机进行预测, 这些手动的特征包括颜色指数、边缘信息的, 虽然在特定条件下能够反映植物的表型特征, 但这些方法依赖于人工经验, 且在不同环境和条件下可能表现出较大的差异 (Hamuda et al. 2016; Ngugi et al.

1. 2021)。近年来,随着深度学习技术的快速发展,特别是 CNNs,病害图像识别与分类的准确性显著提升(Barbedo 2018b; Chen et al. 2020)。再加上迁移学习和预训练模型的广泛应用,为病害自动监测与高通量表型数据获取提供了全新的解决思路。这些模型的自动化系统能够在多种环境条件下保持较高的鲁棒性和泛化能力,有望突破传统方法在效率与准确性方面的限制。

1.2.2 高通量葡萄病害表型鉴定研究进展

农作物病害对全球粮食安全构成严重威胁,导致产量损失,并且每年都对农作物生产构成危害(Bebber et al. 2013; Savary et al. 2019)。为了有效控制病害,农业生产中通常依赖频繁使用杀菌剂,这不仅带来高昂的经济成本,还会对环境造成显著污染(Carisse 2010; Spencer-Phillips et al. 2002)。减少对杀菌剂的依赖,培育抗病品种成为一直以来的重要研究方向(Boso and Kassemeyer 2015; Gessler et al. 2011)。无论是新型杀菌剂的筛选,还是抗病品种的开发,都需在实验室和田间条件下对病害严重程度进行系统评估,特别是针对受感染器官的病斑面积等关键表型特征(Kunova et al. 2021)。这些严重程度评估对于遗传分析和育种发展至关重要,例如 QTL 作图或 GWAS(Amo and Soriano 2022; Qiu et al. 2022)。传统上,严重程度是由技术人员手动估计的,这是一个劳动密集型的过程,由于经验和个人偏见的差异,容易受到显著的不确定性影响(Nutter et al. 2006)。

为了应对这些挑战,基于图像的高通量表型(HTP)技术的最新进展已经证明了其在疾病严重程度评估领域的潜力,它能够快速、准确、大规模地量化与疾病相关的特征(Araus and Kefauver 2018; Yang et al. 2020)。这些 HTP 系统首先尝试采用了各种不同的光源来检测表型特征,如高光谱成像(Sarić et al. 2022)、热红外荧光(Pineda et al. 2020)、叶绿素荧光(Pérez-Bueno et al. 2019)等。其中,RGB 成像因其能够以较低成本非侵入性地识别而被广泛应用。在表型解析算法方面,研究人员也已开发出多种用于病害严重程度评估的算法,涵盖传统图像处理与基于 CNNs 的深度学习技术(Ngugi et al. 2021; Singh et al. 2020)。传统方法如霍夫圆检测、阈值分割等,常用于叶片的检测与分离,进而在去除背景干扰后对霜霉病感染区域进行定量评估(Hernández et al. 2022; Ma et al. 2018)。然而,这些方法由于参数敏感,在多样的环境下泛化性差,不能够稳定的应用在 HTP 系统中。基于 CNNs 的病害识别方法,目前多采用了高分辨率的显微图像作为数据源,导致模型计算量增加。为了减少计算,一些研究将高分辨图像裁剪为若干子图,再利用 CNNs 对子图进行感染与否的分类或者进一步使用类激活图(Class Activation Map, CAM),从而确定感染面积(Bierman et al. 2019; Qiu et al. 2022; Zender et al. 2021)。此外,也有研究直接采用编码器-解码器结构的 CNNs,对叶片及其病变区域进行端到端的语义分割(Li

u W et al. 2024)。基于 CNNs 的方法在分割表现和泛化性能上显著优于传统的方法。

然而, CNNs 在替代人工评估实际应用中仍面临多项关键挑战。首先, 由于所使用的显微图像分辨率高, 这些模型通常仅在单个叶盘上进行叶片与病斑的分割。这不仅显著增加了图像处理的计算成本, 也限制了 HTP 系统的抗性鉴定的效率。其次, 一些适用于目标检测或实例分割的模型, 如 Faster RCNN (Ren et al. 2016) 和 Mask RCNN (He et al. 2018), 尽管有潜力同时完成多目标检测和分割问题, 却缺乏对其在病害表型数据集上的系统适用性评估, 尚不明曲是否其性能限于参数量大部署推理慢, 还受到数据限制导致模型性能不足的影响。最后, 即使是实验室照片, 也不可避免会受到背景干扰。特别是在实验室中, 叶盘接种霜霉病可能需放在白色 A4 纸上, 导致其与白色霜霉层特征相似, 从而可能会增加 HTP 系统的误差。

在检测任务中, 一些轻量级的模型, 如 YOLO 系列已经从 V1 发展到 V8, 带来了更少的模型权重和实时对象检测能力 (Jocher et al. 2023)。这些 YOLO 模型已经被广泛应用于表型定位, 如杂草检测和番茄表型定位, 实现了高精度的性能 (Casado-García et al. 2020; Fu et al. 2021; Jiang et al. 2022)。对于分割任务, 编码器-解码器模型, 如 UNet (Ronneberger et al. 2015), 通常用于实例分割, 参数较少, 在表型提取和疾病分割方面取得了强大的结果, 特别是在背景干净和受控的情况下 (Li et al. 2022)。此外, 视觉基础模型, 如 Segmentation Anything Model (SAM) (Kirillov et al. 2023a), 在许多下游表型任务中展示了卓越的性能 (Williams et al. 2024; Zhang et al. 2024)。这些视觉基础模型可以通过输入提示为一些特定目标生成更精确的轮廓。最近的研究开始结合新开发的 CNNs 以序列化方式获得表型特征 (Mejias et al. 2024; Morales et al. 2024; Yang et al. 2024), 被证明是一种有效且易于利用它们在实际应用中的特殊功能的方法。这些新的技术为实现高通量准确的葡萄病害表型诊断提供了潜在的工具。然而, 由于其巨大的图像编码器, SAM 在资源受限的移动设备上的部署仍是挑战。为了解决这个问题, 轻量级版本, 如 Mobile SAM (Zhang et al. 2023), 通过用轻量级视觉变换器替换编码器蒸馏, 得到参数量更少、推理速度更快、精度损失小的轻量级模型。然而, 目前的研究缺乏对 SAM 轻量版本的适用性评估, 而这正是田间病害表型监测的关键基础。

综上, HTP 是连接基因型和环境型的桥梁, 可以作为加速作物抗性品种培育的高效工具。传统的表型分析算法参数敏感, 基于 CNNs 的模型需要大量数据训练, 成本高。基于预训练的 SAM 为增强模型泛化性提供了工具, 基于提示性的轻量级视觉模型可能有潜力处理复杂的情景, 需要进一步评估其适用性。

1.2.3 田间病害识别与诊断研究进展

病虫害的发生会影响农作物的正常生长,造成粮食产量降低和质量下降,从而引发经济损失和粮食危机 (Strange and Scott 2005)。联合国粮农组织 (FAO) 估计,全球每年因为病虫害导致的粮食产量损失超过 40%以上,植保投入占据了种植成本的 50%以上 (FAO 2021)。早期的准确诊断和预警是能够及时的采取植保措施以减少这种损失的关键,这依赖于农户对于病害发生的早期认识 (Vishnoi et al. 2021)。对于大多数的农户而言,由于缺少专业的植保知识,通常会延误采取植保措施以保护作物的时机。对于病虫害自的识别,即便对于农艺专家而言,人工识别病虫害也会相当花费时间,有时候会出现主观的错误,导致杀菌剂的错误使用 (Ma et al. 2018)。因此,自动准确的病虫害识别和记录是做好农业植保管理的有效手段,特别是在智能手机相对普及的情况下,利用图像处理技术进行病虫害的自动识别成为可能。

对病虫害特征的提取和分类是实现病虫害自动识别的两个核心步骤。传统的特征提取算法通通常需要人工设计特征提取器,然后将其输入到分类器中,实现对病虫害的自动识别分类 (Ngugi et al. 2021)。这些手动的特征提取器包括颜色指数、边缘信息、特征纹理以及阈值分割等 (Kataoka et al. 2003; Ojala et al. 2002; Otsu 1979)。常见的分类算法包括人工神经网络 (Artificial Neural Network, ANN)、随机森林 (Random Forest, RF) 以及支持向量机 (Support Vector Machine, SVM) 等。研究中通常使用这些特征提取算法和分类器的组合,来比较病虫害分类的准确性 (Ahmed et al. 2021; Shruthi et al. 2019; Waghmare et al. 2016)。这些传统的病虫害识别算法,在田间复杂的条件下鲁棒性较差,在实际中难以应用 (Kamilaris 2018)。

近年来, CNNs 在图像识别领域取得了显著成果。例如,在包含 1000 类图像的 ImageNet 分类任务中, CNN 将 Top-5 分类错误率降低至 3.6%,已低于人类的识别误差。这一突破推动了基于深度学习的病虫害自动识别成为具有应用潜力的研究方向。已有大量研究验证了 CNN 在病虫害识别任务中的鲁棒性。Ramcharan et al., (2019) 比较了几种机器学习算法,如 RF, SVM, 和 CNNs, 结果表明 CNNs 在识别精度方面具有明显优势。Ghosal et al., (2018) 等通过使用可解释的神经网络方法,比较了 CNN 模型在大豆胁迫症状识别中所关注的图像区域与植保专家诊断区域之间的一致性,结果显示两者高度重合,进一步证明了 CNNs 模型的可用性和可信度。越来越多的 CNN 架构被应用于病虫害诊断任务中。Too (2019) 使用了多个 CNN 模型 (VGG-16 Net, ResNet-50/101/152, Inception-V4, 和 DenseNet), 在 PlantVillage 数据集上测试了模型性能, 结果表明几乎所有模型的准确率均超过 97%。Brahimi et al., (2018) 同样基于 PlantVillage 数据集, 比较了六种经典 CNN 架构 (AlexNet, VGG-13, DenseNet, Inception-V3 和 ResNet-34) 在三种训练策略 (从头训练、仅微调全连接层、微调全部层) 下的表现, 其中 Inception-V3 取得了最高的准确率 (99.76%)。

尽管许多基于 PlantVillage 数据集的研究在病虫害识别任务中取得了超过 97% 的分类准确率，但在实际田间环境下，这些模型的识别性能却大幅下降，表明在干净背景下训练的模型存在明显的泛化能力不足问题（Mohanty et al. 2016; Tian et al. 2024）。在田间环境下的识别准确率受到多种因素影响，如图像采集角度、光照条件不均、病害不同阶段的症状差异，以及病斑特征的不明显性（Barbedo 2018a）。有研究表明，在实验室图像上训练的 CNN 模型，其在田间图像上的准确率下降了约 30%；而用田间图像训练的模型在实验室数据上的表现则更为稳健，显示出更好的泛化能力（Ferentinos 2018）。然而，田间图像数据的获取与标注过程极其耗费人力、物力与时间，因此，如何在小样本数据集上提升田间识别准确率成为一个关键挑战（Barbedo 2018b）。之前的研究中，数据增强与迁移学习是缓解小样本问题的主要手段，即通过图像翻转、旋转、仿射变换、加噪、随机裁剪等方式生成新样本，从而扩大训练集规模（Abayomi-Alli et al. 2021; Tsai and Lee 2020）。目近年来，也有研究尝试使用生成对抗网络（Generative Adversarial Nets, GAN）进行图像生成，以进一步丰富训练样本（Chen and Wu 2022）。然而，该方法也存在潜在风险，若在数据增强前对数据集进行划分，可能导致增强后的样本在训练集和测试集之间高度相似，从而引发数据泄漏问题。虽然这种方法可能提升模型在测试集上的准确率，但在真实应用中模型性能可能不升反降（Barbedo 2018b; Ferentinos 2018）。此外，基于图像增强生成的数据，其特征空间往往与原始图像保持一致，尤其在背景不变的情况下，这进一步限制了模型在复杂田间环境中的泛化能力。

综上，田间病虫害识别是一个复杂且具有挑战性的科学技术问题，尤其是在数据集规模受限的情况下。迁移学习技术为该问题提供了重要解决路径。通过利用在大型通用数据集（如 ImageNet）上预训练的模型权重，并结合数据增强等策略，可在小样本病虫害数据集上有效提升识别精度，从而增强模型在田间环境下的实用性与泛化能力。

1.2.4 田间病害特征描述研究进展

田间病害识别是一个复杂且具有挑战性的问题，主要受到背景干扰复杂和数据稀缺等因素的影响（Barbedo 2018b）。尽管已有研究在实验条件下取得了较高的识别准确率，但这些模型在田间环境中的实际应用效果却不理想（Ferentinos 2018）。更重要的是，基于CNNs的病害识别方法通常仅关注图像的标签分类，缺乏对图像内容的深入解释（Chen et al. 2020; Too 2019）。现有的分类可视化方法，如类激活（CAM）解释性等，虽可标示病害部位和类别，但无法像人类专家那样，基于颜色、形状、纹理等物理特征进行全面诊断。模型缺乏对决策依据的清晰解释，即“黑箱”问题，已成为限制病害识别模型实际部署与应用的重要因素之一（Sun et al. 2023）。

为了解决上述问题，越来越多的研究尝试模拟人类在病害观察和识别中的认知过程，这通常涉及多模态数据的融合，尤其是病害特征的文本描述信息（Liu et al. 2025）。然而，作物病害的图文配对数据相对稀缺，且获取成本高昂。当前多采用手工标注或借助语言模型生成病害描述文本，但这些描述往往缺乏结构化特征，关键病害特征表达不充分，限制了模型的表现。因此，许多研究仍依赖小样本进行训练和评估（Arshad et al.; Zhong et al. 2020）。

在多模态数据基础上，现有研究多数聚焦于图像与文本间的语义对齐，通过计算图文间的特征距离来提升分类精度（Cao et al. 2023; Zhou et al. 2021）。尽管该方法可在一定程度上实现病害的文字描述，但由于这些描述多为检索生成，缺乏针对具体叶片的真实解释，导致图文特征之间存在语义偏差。另一些研究则尝试使用生成模型直接生成病害描述句子，模仿人类专家的诊断逻辑，以增强模型的解释性（Wang et al. 2023）。然而，使用如LSTM、RNN 或 Transformer 等生成模型通常需要大量多样化的数据进行重新训练，在数据有限的条件下，这些方法往往表现出文本生成能力不足（Zhou et al. 2021）。

2023年以来，大语言模型（Large Language Models, LLMs），比如ChatGPT（Open AI 2023）、BERT（Devlin et al. 2019）等，在自然语言处理领域取得了显著进展，展现出强大的语言理解与生成能力，尤其是ChatGPT，首次展示了通用人工智能在语言处理中的广泛适应性与通用性。这一进步显著增强了下游任务的泛化能力，并为病害识别任务中的文本生成提供了新契机。视觉语言模型（Vision-Language Models, VLMs）的发展为突破传统模型的限制提供了可能。近期已有研究将 LLM 应用于病害图像描述任务，如将BERT或ChatGPT 等模型与视觉特征结合，展现出良好的问答能力与图文理解效果（Liang et al. 2024; Zhao et al. 2024）。此外，一些专为农业设计的视觉语言模型，如AgriCLIP，CDIPChatGLM3，仅在零样本分类中提升了识别精度，还能生成详细的管理建议，从而显著增强了模型的实用性与可解释性（Awais et al. 2025; Yan et al. 2025; Zhang K et al. 2024）。这进一步凸显了视觉语言模型，与表型信息结合，在田间的潜在的应用潜力。随着技术发展，越来越多先进的视觉语言模型被陆续开源并投入使用，包括Qwen-2.5-VL（Bai et al. 2025），以及DeepSeek-VL2（Wu et al. 2024）等，其中Qwen-2.5-VL在中文语境下表现出更好的适配性，为中文作物病害识别任务提供了更为理想的语言理解能力。

综上所述，构建一个具备良好解释性的田间病害识别系统，需要模型能够模拟人类专家对病害的识别过程，输出合理且结构化的诊断性描述。然而，以往的方法多依赖图像的显著区域定位或基于相似度的检索，难以在实际场景中提供诊断依据。视觉语言模型为病害识别的解释性突破提供了新的可能，通过融合视觉特征与表型描述，增

强生成病害特征描述的能力。评估视觉语言模型在田间环境下的泛化能力与实用价值是必要的研究工作。

1.2.5 田间病害严重程度评估研究进展

在农业场景中，为了准确的表型监测，需要对作物的各个参数进行定位、识别、分割和测量，如叶片胁迫、胚芽鞘长度、气孔数量等，因此表型系统对图像处理算法的鲁棒性和泛化性要求较高（郝倩琳等 2024; Casado-García et al. 2020; Liu et al. 2024; Qiu et al. 2022）。在之前的研究中，传统的图像处理算法在田间情况下通常表现不佳，近年来随着深度学习技术的进步，基于卷积神经网络的田间表型鉴定逐渐成为流行（Jiang and Li 2020）。但这种监督学习的方式，往往需要采集大量的数据来训练，同时需要专业的人员来进行标注，导致较高的成本。而且即使有田间数据，由于背景信息的复杂性，深度学习算法常常达不到应用要求。一些研究直接使用 CNNs 分割网络来对目标进行提取，然而却无法避免背景的干扰，如临近叶片的病斑等（Zhang et al. 2024）。一些研究则通过序列化分割来，先分割叶片，在分割病斑，来实现田间病害表型诊断（Liu et al. 2024; Wang et al. 2021）。

在计算机视觉领域，同样也出现了大模型，如 Segment Anything Model（SAM）（Kirillov et al. 2023a），并被广泛的应用在各种分割任务中，表现出了强大的零样本泛化能力。SAM 的结构主要包含三个，基于 Vision Transformer 的照片编码器，以及提示线索编码器以及掩膜生成的解码器。SAM 使用了超过 1000 万张标注的图像来训练，从而具有出色的零样本迁移能力，能够很好地泛化到以前看不到的数据。同时，SAM 支持输入不同的提示，如目标对象的点和边界框等，从而促进精确和多样的对象分割，这使得它在具有挑战性的下游场景中表现出强大的泛化性能。2024 年以来，SAM 也被广泛应用于表型特征的提取中。比如杂草识别中，Nguyen et al. (2023) 使用 PhenoBench 数据，结合 SAM、DINO 和 YOLO-v8 模型，在作物-杂草分割竞赛上得到 81.33 的准确率。在多种植物表型参数测量任务上，Zhang et al. (2024) 使用 SAM 和 CLIP 方法，无需训练和标签数据，对于大多数测试样本，均得到小于 0.05 的平均绝对误差。在水稻穗部表型分析任务上，Yang et al. (2024) 利用智能手机和组合 SAM、YOLOv8 及 NeRF 方法，实现了高精度 3D 重建和特征提取，与人工测量相比高度一致。这些基于视觉基础模型的算法显示出巨大的应用潜力（Li Y et al. 2023; Yang et al. 2024）。

然而，由于 SAM 的图像编码使用的参数较多的 Transformer 结构，在移动设备上变得困难。同时由于在特定任务中，模型并没有得到太多的专业的数据训练，导致其在泛化能力上有时候会出现差错，比如病斑位置的分割，以及隐藏的目标分割等（Chen et al. 2023, 2023; Ji et al. 2023）。为了解决这些问题，一些轻量级的模型

被开发出来用于模型的移动设备部署, 比如 Mobile SAM (Zhang et al. 2023b) 和 Edge SAM (Zhou et al. 2024) 等。这些模型采用了替换图像编码器的方式, 以及进一步结合预训练模型, 对新的模型进行蒸馏或者重新训练, 以降低模型的参数量。另一些模型使用了量化、剪枝或局部优化以及图像后处理等技术对原始模型进行压缩。比如通过在模型中插入 Adapter 层, 在该层进行模型的微调, 同时冻结其他层的参数, 以达到与微调整个大模型的一样的效果 (Chen et al. 2023)。然而, 可提示分割是 SAM 的基本任务, 其在任意给定提示(例如, 一个点、一个边界框、一个面具和文本)的情况下返回一个有效的掩码, 这不仅是 SAM 解决各种下游任务的基础, 也是数据引擎最后一个阶段的关键步骤。因此, 采用一种有效且高效的采样策略进行提示生成是必不可少的 (Sun et al. 2024)。

提示学习 (Prompt Learning) 是一种最新的预训练模型范式, 通过在预训练过程中提供特定任务的提示信息, 来指导模型学习, 帮助模型更好地利用任务的上下文信息, 从而提高模型的性能, 也使得模型可以在少样本、零样本等低资源场景下保持良好的表现 (Schulhoff et al. 2024)。最近的研究开始结合 SAM 和其他模型, 来进一步生成提示, 以满足下游任务的需求, 增强其应用能力。比如 Uygun et al., (2024) 研究创建了一个包含 800 张健康和受损图像的数据集, 使用 SAM 模型将图像标记的边界框自动转换为掩码标签, 然后用 YOLOv8-Seg 模型进行训练, 结果显示, YOLOv8l-Seg 模型在 mAP0.5 指标下性能最佳, 达到了 92.4% 的 mAP。在葡萄病害抗性鉴定上, Lin et al. (2024) 通过 SAM-CLIP 和 PM-SAM-CLIP 对两个数据集进行分割, 生成葡萄树冠层和 PM 感染的掩模, 然后用于量化感染的严重程度。Zhang et al. (2024) 提出的 YOLOv8n-DDA-SAM 模型通过在目标检测中添加语义分割分支 SAM, 有效地获取了采摘点的掩码, 并通过对简单形状中心点计算精确获得了采摘点, 相比于以前的模型有了显著提升。

综上, 基于轻量级参数的视觉大模型相比如传统的卷积神经网络, 具有显著的优势, 不仅可以降低数据标注成本, 也具有很强的零样本泛化能力。如何增强这些模型在农业场景下游任务中的应用, 是重要的研究课题。模型微调、提示学习是两个有效的方式, 需要做进一步的评估。

1.3 研究主要存在的问题

本研究以葡萄霜霉病表型为研究对象, 主要针对以下三个技术应用问题展开研究:

(1) 在实验室中, 传统的葡萄霜霉病抗性鉴定中, 往往依赖人工标注, 一些研究虽然使用了如霍夫变换以及阈值分割等传统算法, 但是模型泛化性能差。近年来基于深度学习的方法, 使用的均为高分辨率的显微镜照片, 不仅获取成本高, 且消耗计

算资源。因此，开发适用于葡萄抗性鉴定的低成本高通量系统是现实的需求。

(2) CNN 已经被广泛的应用病害的分类和识别，且取得了较高的准确率。然而，CNN 往往依赖 CAM（类激活图）做可解释性分析。虽然指示了病害发生的部位，却缺乏对病害特征的描述性解释，导致用户无法完全信任检测结果。因此，模型无法清晰解释决策依据的“黑箱”是制约各类病害识别模型实际应用的关键原因。尽管视觉语言模型已经开始被用于病害特征描述，然而表型数据稀缺，模型专业化程度低，模型训练成本高，制约了其性能以及适用性。因此，本研究利用田间采集和开源的葡萄病害数据集，通过微调定制化葡萄病害描述生成的多模态模型，来评估其对病害特征的描述准确性。

(3) 准确的田间病害记录对于优化病害管理至关重要以往的记录往往需要人工完成，费时费力。近年来，基于 CNN 的分割网络被应用于田间病害分割。然而，病害田间背景复杂多样，使得即使是先进的卷积神经网络也很难达到准确的分割。因此，如何实现避免田间背景的影响，以及准确细致的分割叶片和病斑，是重要的技术问题。需要进一步评估视觉基础模型提示策略在田间表型自动分割的表现，准确量化病害发生严重程度。

1.4 研究内容和技术路线

1.4.1 研究内容

(1) 研发高通量叶盘霜霉病表型鉴定模型

使用实验室采集葡萄叶盘接种霜霉病菌的图像数据集，通过 AnyLabeling 软件结合 Mobile SAM 实现叶盘的半自动标注以及霜霉病产孢区域则以手动方式精细标注。通过整合了 YOLOv8n、Mobile SAM 和 UNet，开发该低成本高精度的自动化模型。主要试验设计包括：(a) 基于 YOLOv8 和 Faster R-CNN 选择最佳的叶盘检测模型；(b) 评估使用 Mobile SAM 进行背景移除的有效性；(c) 对比 8 种主流编码器-解码器分割模型（如 UNet、DeepLabV3+等）在叶盘和产孢区域的分割表现。(d) 整个模型性能进一步与经典的实例分割模型 Mask R-CNN 进行对比，以验证其在病害检测中的优势。研究目标是利用该模型预测掩码计算产孢区域面积占叶盘面积的比例，作为葡萄抗病性的重要指标，实现霜霉病严重度的自动化评估。

(2) 比较微调 Qwen-VL 的葡萄病害图像描述生成表现

使用实验室葡萄病害照片数据，包括霜霉病、黑腐病、叶斑病、黑麻疹等，通过视觉语言模型进行标注，然后进一步通过专家校正，以得到高质量的病害照片以及对应的病害特征描述文本数据。然后通过 LoRA 微调 Qwen-2.5-VL-3B 模型，以评估其在病害特征描述生成和病害特征区分准确性的表现。主要试验设计包括：(a) 比较了两种 LoRA 微调参数设置对 Qwen-2.5-VL-3B 的模型性能的影响；(b) 对比了微调后

的模型与其他视觉语言模型的表现，包括 BLIP2、DeepSeek-VL2、Describe anything model、Kimi-VL-A3B；（c）比较了不同病害特征的准确率，包括病斑颜色、病斑形状、病斑位置、病斑面积。定制化葡萄病害特征描述生成的视觉语言模型，评估其表现，是否可以推动田间病害识别的可解释性和高可信度。

（3）评估提示引导的 Mobile SAM 在葡萄病害分割方法及严重程度的表现

使用田间采集的葡萄霜霉病病害叶片和健康叶片，通过手动和半自动的标注，制作了用于训练、验证和测试的数据集。利用传统的 CNN 分割网络得到叶片和病斑的粗略掩膜，然后生成不同的点提示和边界框提示，进一步结合 Mobile SAM 做分割细化。主要试验设计包括：（a）比较 6 个经典的图像分割网络对叶片和病斑分割表现，选择最好的模型采用；（b）比较了单点提示、3 点提示和单个边界框提示的不同组合对分割结果的影响；（c）分别评估了最好模型在验证集和测试集严重程度评估表现，证明其泛化能力。评估提示性的轻量视觉基础模型 Mobile SAM 在田间病害分割的应用潜力。

1.4.2 技术路线

本研究围绕葡萄霜霉病叶片尺度的抗性评估与病害识别和分割，构建了一套多维度的技术路线。首先，基于实验室叶盘接种图像，结合 YOLO v8n、Mobile SAM 和 UNet，开发了一个低成本高精度的自动化模型 SporaScan，用于实现叶盘检测、背景移除和孢子囊区域分割，并通过产孢面积比例评估抗病性，同时与 Mask R-CNN 等方法进行性能对比验证。其次，利用多类葡萄病害图像及专家校正的描述文本，构建了图文配对数据集，并采用 LoRA 技术对 Qwen-2.5-VL-3B 视觉语言模型进行微调，以提升其在病害特征描述生成中的准确性和区分能力，并与多种主流视觉语言模型进行比较。最后，面向田间真实场景，采集霜霉病与健康叶片图像，整合 CNN 分割网络与 Mobile SAM，探索不同提示方式（点提示与边界框提示）对叶片与病斑分割精度的影响，构建田间病害分割与严重程度评估模型，评估其泛化性能与应用潜力。整体技术路线融合检测、分割与多模态生成，旨在实现葡萄病害从实验室到田间的全流程智能评估与应用。总体技术路线如图 1-4 所示。

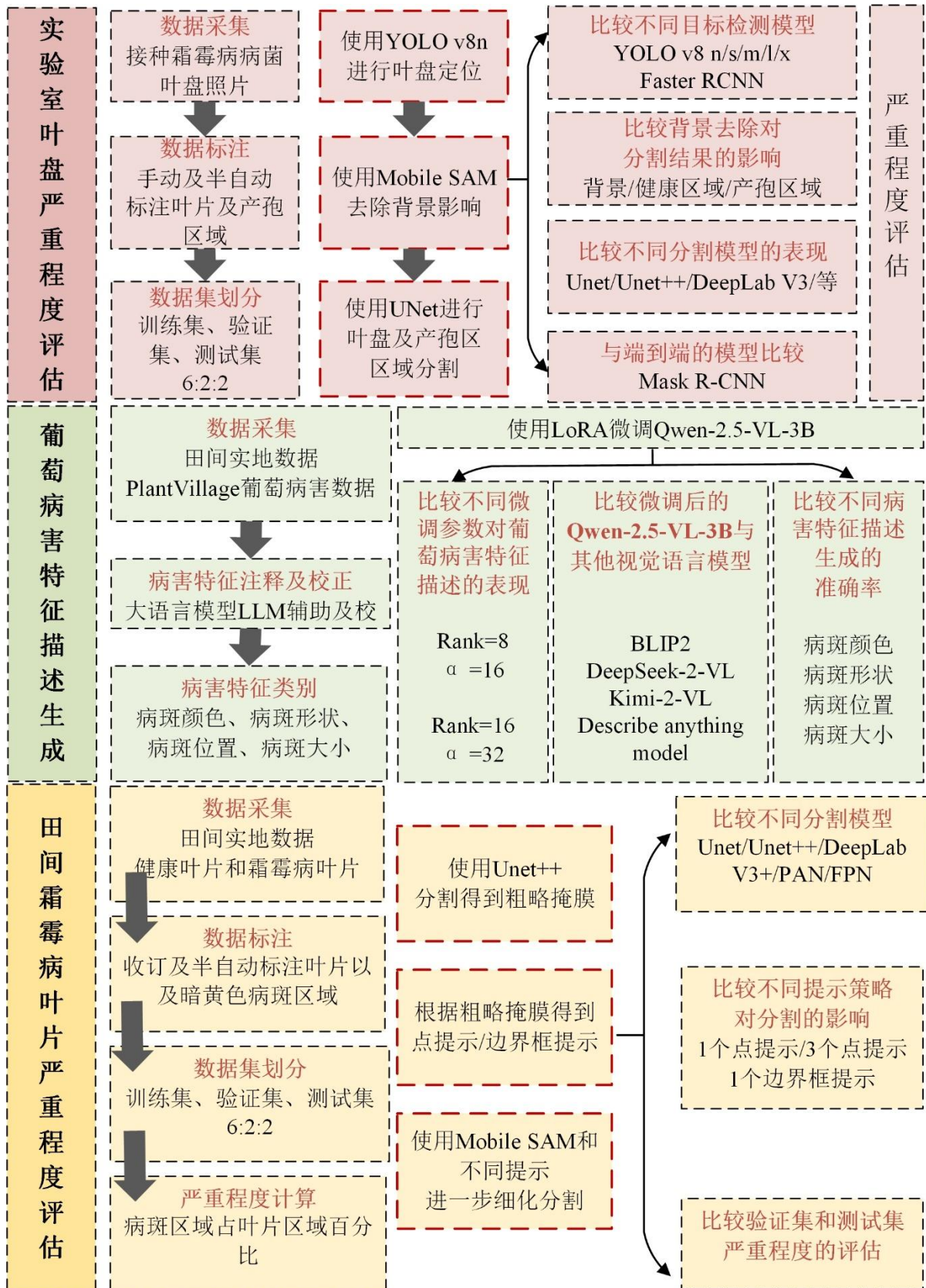


图 1-4 总体技术路线图

Figure 1-4 Overall workflow

材料和方法

1.5 数据来源

1.5.1 实验室葡萄叶盘接种霜霉病照片数据

2.1.1.1 葡萄叶盘接种霜霉病试验介绍

葡萄叶盘试验被广泛的应用于病原菌的抗性检测、病害防治药剂的筛选以及与植物免疫反应等研究 (Gómez-Zeledón and Kaiser 2016; Štambuk et al. 2021)。在病原菌的抗性检测中, 叶盘法可以通过模拟自然感染环境, 测试不同病原菌对葡萄的致病性, 并评估葡萄不同品种的抗性。通过在葡萄叶片上接种病原菌, 能够直观地观察到病害的发展过程, 为葡萄品种的抗病性筛选提供基础数据。在本研究内容中, 主要关注葡萄霜霉病。

葡萄叶盘接种霜霉病试验的主要流程包括: (1) 材料准备: 选择健康的葡萄叶片, 通常选择生长旺盛、无明显病害的叶片。根据需要, 可以选择不同品种、不同生长阶段的葡萄叶片。然后, 制备所需浓度的霜霉病菌的孢子悬浮液, 原始孢子通常通过培养霜霉病菌并收集其孢子囊来获得。(2) 叶片处理: 对健康的葡萄叶片进行表面消毒, 一般使用 75%乙醇或 1%次氯酸钠溶液, 消毒后再用蒸馏水洗涤干净, 再用某一直径大小的打孔器, 通常为 1mm-2mm, 将健康的叶片打成小叶盘。(3) 病原菌接种: 使用制备霜霉病孢子悬浮液对叶盘进行接种, 常见的接种方式包括滴加和喷雾。

(4) 培养和观察: 将处理后的叶盘放置在适宜的培养条件下 (一般为湿度较高、温度适宜的环境), 在培养期间, 观察并记录叶片上病斑的发展情况, 包括病变的大小、形状和颜色等, 通过数字成像、目视评分等方法进行量化分析。(6) 根据试验结果, 评估不同药不同病原菌或不同处理条件下的防治效果。目前, 叶盘法通常与基因组学、代谢组学相结合, 以进一步分析其抗病机制, 包括基因的表达、突变以及代谢产物的机制分析 (Qiu et al. 2022; Yang et al. 2020)。通过全基因组关联分析 (GWAS)、数量基因座定位分析 (QTL) 以及 CRISPR-Cas9 基因编辑等, 寻找与相关抗病基因的对应位点, 并通过编辑 (敲除) 相关基因并测试, 以进一步用于品种改良, 服务于精准育种。

2.1.1.2 葡萄叶盘接种霜霉病数据采集

在本研究中使用的数据集来源于葡萄叶盘试验, 该试验旨在评估葡萄品种对霜霉病的抗性 (Qu et al. 2021)。在霜霉病感染发展后, 通过智能手机拍摄了培养皿中的叶盘图像。叶盘试验和图像采集的主要过程图 1-1 所示。

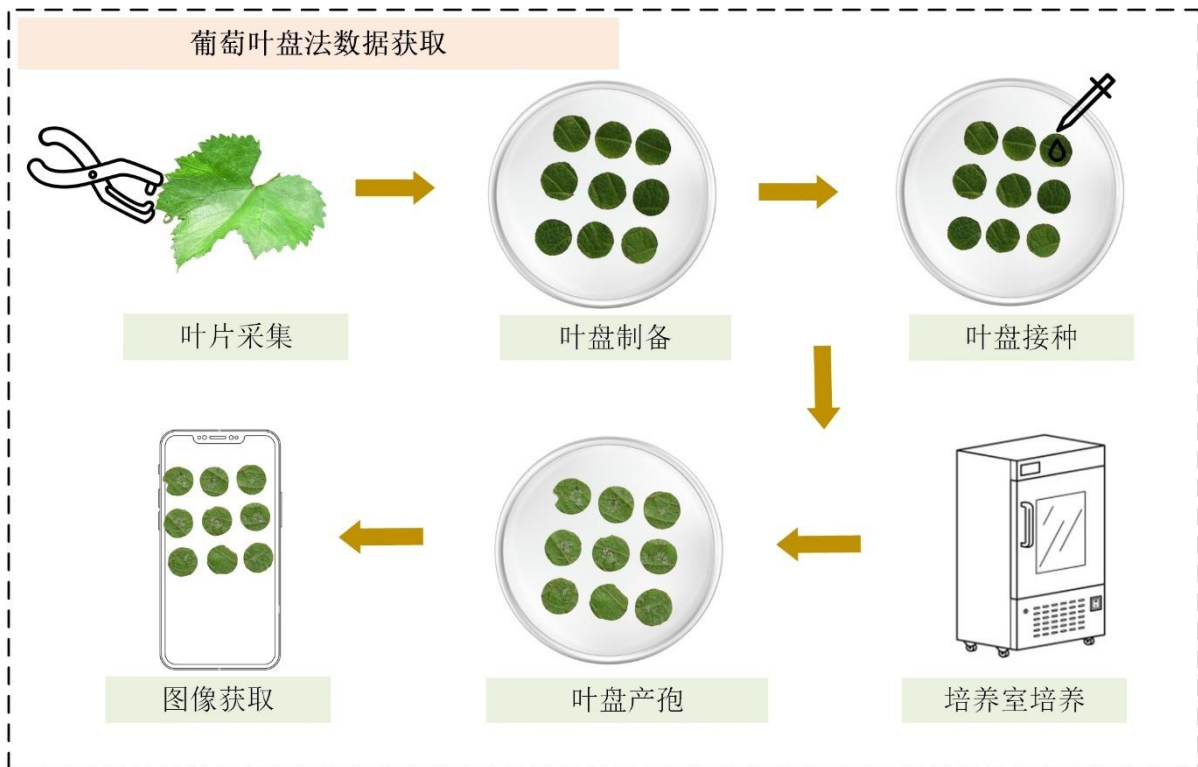


图 1-1 叶盘试验和图像采集的示意图，详细说明了以下步骤：叶片准备、圆盘放置、用葡萄霜霉（*Plasmopara viticola*）孢子囊接种、在控制条件下培养、感染发展以及图像采集用于病害严重度评估。

Figure 1-1 Schematic representation of the leaf disc assay and image acquisition, illustrating the sequential steps of leaf preparation, disc placement, inoculation with *Plasmopara viticola* sporangia, incubation under controlled conditions, infection development, and image acquisition for disease severity assessment.

葡萄叶片采自广西都安瑶族自治县农业合作社的葡萄试验园（地理位置 22.47° N, 108.21° E, 海拔 142.3 米）。本研究中评估的葡萄品种包括霞多丽、叶酿六号、桂葡六号和夏黑。在实验室中，收集的葡萄叶片首先用蒸馏水冲洗，以去除表面杂质，然后使用吸水纸轻轻擦干。从完全展开且健康的叶片上，用无菌打孔器切取直径为 1.5 厘米的叶盘。根据实验设计，每个培养皿中放置了 6 至 25 个叶盘，并确保其背面朝上。每个叶盘均用含有葡萄霜霉（*Plasmopara viticola*）孢子囊悬浮液（浓度为 1×10^6 孢子囊/mL）的无菌水接种，通过将液滴滴在每个叶盘的表面进行接种。葡萄霜霉孢子囊已在实验室中分离、准备并保存在合适条件下。接种后，培养皿被置于植物生长箱中，在 22° C 下，采用 16 小时光照/8 小时黑暗的周期进行培养。培养期为五到七天，期间叶盘上开始出现霜霉病的可见症状。未接种的叶盘未表现出任何症状。

使用智能手机拍摄这些叶盘的图像，以记录感染情况。共收集了 432 张图像，代表了 6,972 个叶盘，涵盖多个实验重复。由于图像采集时相机与物体之间的距离不同，叶盘的分辨率有所不同。每个叶盘的图像分辨率如图 1-2 所示。

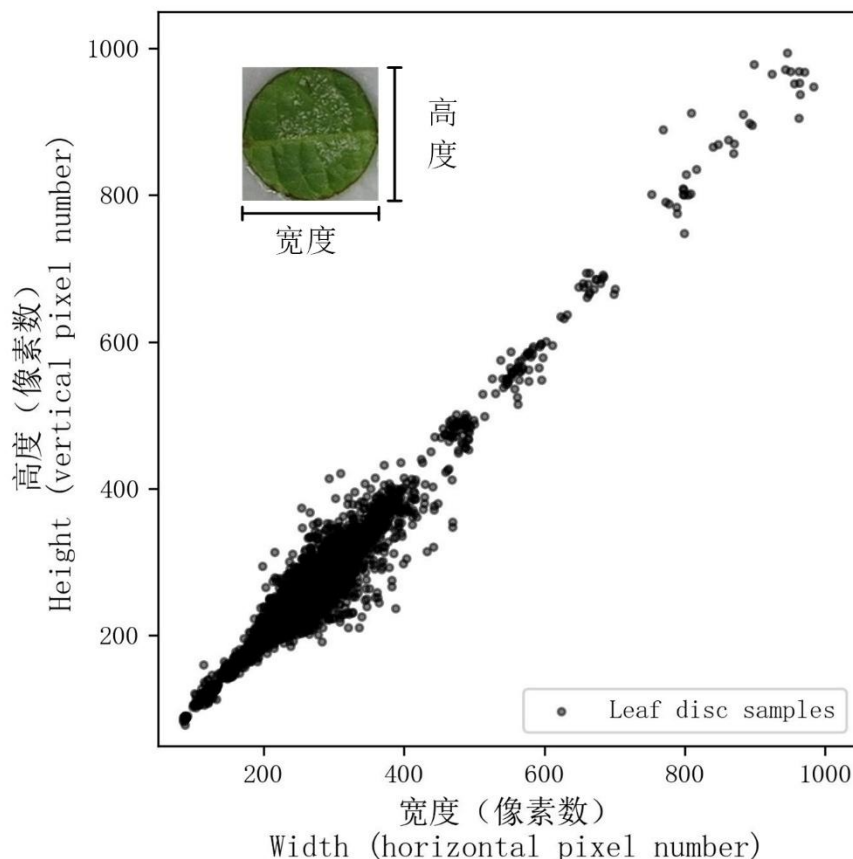


图 1-2 数据集中葡萄叶盘的不同分辨率。宽度表示水平像素数，且高度表示垂直像素数。

Figure 1-2 Different resolutions of grapevine leaf discs in the dataset. “Width” represents the horizontal pixel number, and “Height” represents the vertical pixel number

1.5.2 田间葡萄健康叶片和霜霉病叶片照片数据

2.1.2.1 数据采集地点简介

本研究的数据采集自中国广西壮族自治区都安瑶族自治县的葡萄种植园（地理坐标为 23.829N, 108.110E），如图 1-3 所示。该葡萄园的种植面积大约为 300 亩，采用露天栽培技术，主要种植品种为野酿 2 号，该品种对霜霉病具有较高的易感性。在都安地区，毛葡萄霜霉病发病时间主要在 4 月中旬到 5 月上旬，取决于当年和当地的气候条件。2023 年，由于天气较为干旱，首次观测到霜霉病的时间为推迟到 5 月 8 号，发病点在果园中地势较为低洼的地区。毛葡萄霜霉病发病初期，在叶片上出现黄色油斑，在叶片背面，出现白色霉层，随着病情的发展，黄色油斑逐渐增多，继而油斑中间变为褐色，而后枯萎，最终导致整个叶片的枯萎和凋落。在 2023 年，随着天气条件的变化，毛葡萄经历了多轮霜霉病流行。

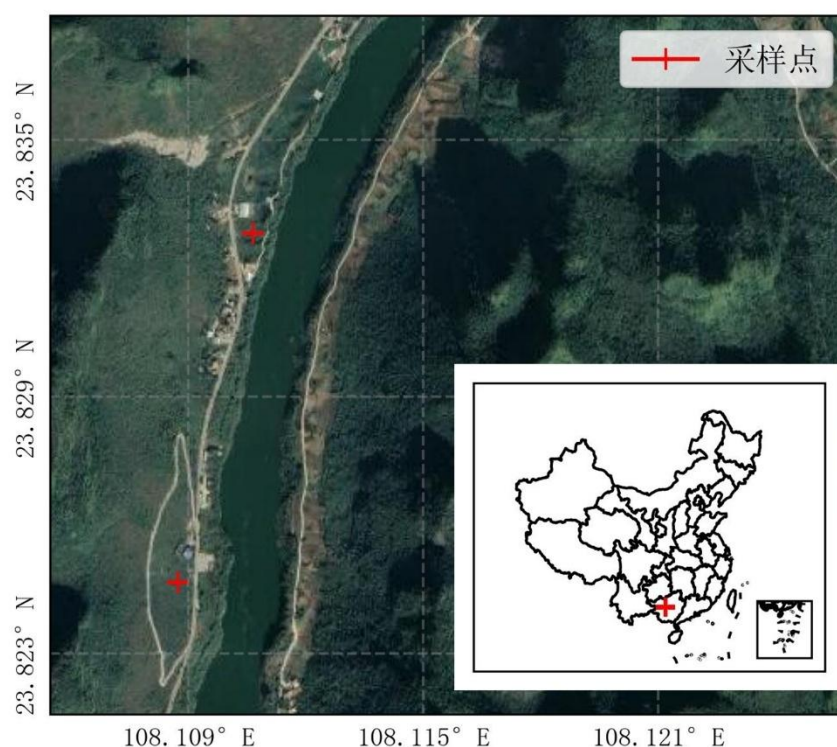


图 1-3 试验地位置，具体为广西壮族自治区都安瑶族自治县安阳镇下梯屯，
该果园主要种植野生毛葡萄（*V. quinquangularis* Rehd），面积约 350 亩。

Figure 1-3 The experimental grape yard is located in Xiatituan, Anyang Town, Du'an Yao Autonomous County, Guangxi Zhuang Autonomous Region. This grape yard primarily cultivates wild varieties (*Vitis quinquangularis* Rehd), covering an area of approximately 350 mu.



图 1-4 葡萄试验园的实地照片

Figure 1-4 Field photograph of the grape experimental yard.

2.1.2.2 健康、霜霉病以及药害照片

自首次发病后，在后续的发病周期内，我们使用了不同类型的手机随机进行毛葡萄霜霉病照片的拍摄，拍摄手机型号包括小米 12S，苹果 12 以及红米 Note 11 等，拍摄果园位置图 1-3 以及图 1-4 所示。在拍摄的过程中，在试验小区的不同位置，选择发病的叶片和健康的叶片作为主要对象。由于葡萄叶片相对较大，我们没有拍多个叶片病害的照片，也就是说每张拍摄的照片中只包含一个健康的叶片或者有病斑的叶片。拍摄叶片时，我们也没有刻意调整角度，只为确保叶片为主体包含在照片中，并可以看到病斑位置，以模拟实际拍摄情况。

在监测霜霉病的发展过程中，我们发现一种与霜霉病初期感染非常相似的叶片症状（实际为药害）。尽管有经验的农技人员能够识别两种症状的差别（药害的黄斑症状明度较高），使用已有的病害识别软件，如植小保、植保家、耘眼等，以及专家系统等，均无法正确的识别该类型症状。因此，我们使用手机设备，同时收集了包含该症状的照片。

我们使用将存储在手机的葡萄叶片照片数据，进一步转移到电脑上进行分类整理，得到了田间的葡萄病害叶片数据集。其中，毛葡萄霜霉病叶片 2483 张，健康叶片 1408 张。与霜霉病类似，由于其他原因导致的黄斑（药害）叶片 1743 张。尽管照片数据量大，相当多的照片并未包含整个的叶片，因此大部分没有被应用到田间病害叶片分割中。

数据收集完毕后，我们对所有照片进行了严格的质量筛选，最终获得了 687 张高质量的照片数据集，其中包含 338 张带有霜霉病症状的叶片照片和 349 张健康叶片的照片，确保了研究数据的平衡性。采集的健康叶片和霜霉病叶片照片样本如图 1-5 所示。药害照片如图 1-7 所示。通过田间采集的葡萄霜霉病叶片、健康叶片，药害照片数据，分别应用于第 4、5 章节，田间病害描述生成、分割诊断模型的研发和评估。

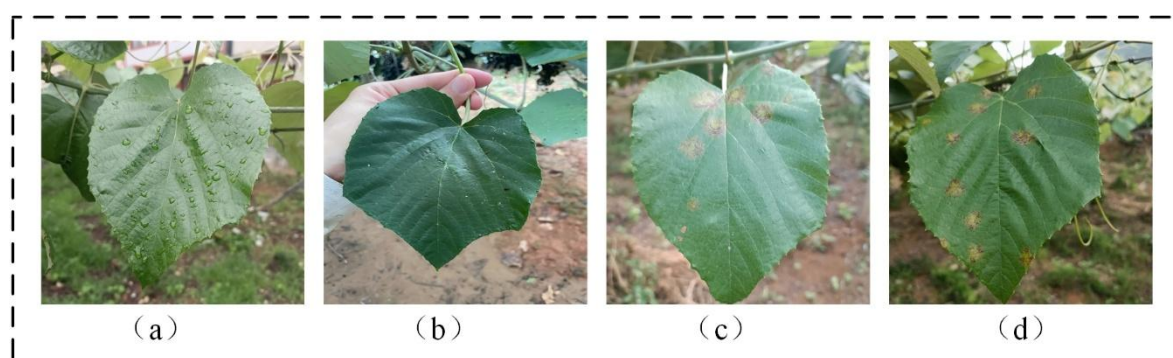


图 1-5 田间采集的葡萄健康叶片和感染霜霉病叶片照片。其中（a）和（b）为健康叶片，（c）和（d）为霜霉病感染的叶片。

Figure 1-5 Field-collected images of healthy and downy mildew-infected grape leaves. Images (a) and (b) show healthy leaves, while (c) and (d) depict leaves infected with downy mildew.

1.5.3 开源病害数据集中葡萄病害照片数据

我们使用了开源病害数据集 Plant Village (Mohanty et al. 2016)。Plant Village 数据集是一个广泛用于植物病害识别和分类的图像数据集，由宾夕法尼亚州立大学的 Plant Village 项目团队创建 (<https://www.kaggle.com/datasets/emmarex/plantdisease>)。该数据集被广泛用于病害识别，并且能够取得较高的分类准确率 (Brahimi et al. 2018; Too 2019)。数据集中包括超过 54,000 张病害图像，涵盖 14 种作物类型，38 种病害类型。每张病害照片均在实验室环境下拍摄，因此背景较为干净。同时，这些照片都有着较小的分辨率 (256×256)，因此所需的计算资源也相对较小。

我们从这些数据中，选取了葡萄的 3 种病害以及健康的叶片数据，包括葡萄黑腐病、叶斑病以及黑麻疹病（以下统称病害）。在原始 Plant Village 数据集中，葡萄黑腐病有 1180 张照片，黑麻疹病为 1383 张照片，叶斑病为 1076 张，健康的叶片为 423 张。为了保证数量的平衡，我们通过 Python 脚本从这些数据中随机选了 400 多张照片，已匹配健康叶片和田间采集的数据。这四种葡萄病害照片如图 1-7 所示。这些数据与田间采集的霜霉病叶片数据和药害照片数据，一起组成用于图像描述生成的数据集，用于第四章。

1.6 数据预处理过程

1.6.1 叶盘接种霜霉病照片标注

我们使用了开源工具 AnyLabeling (<https://anylabeling.nrl.ai/>)，对葡萄叶盘及其对应的霜霉病产孢区域进行标注。该软件可以使用预训练视觉基础模型，如 Mobile SAM，支持半自动标注。在此过程中，我们首先手动绘制一个矩形框架，使其能够覆盖叶盘范围，随后矩形框的四至坐标会自动输入到 Mobile SAM，生成对应的叶盘掩膜轮廓。霜霉病产孢区域由于轮廓不规则，我们使用了多个连接点形成的线进行手动标注。在标注过程中，如果叶片轮廓未能准确的标注出来，则重新输入提示框，直至轮廓完整。如果不能得到完整的轮廓，则手动标注。对于病斑的标注，我们主要依靠手动标注，只对明显的产孢区域进行了边界框提示分割。图像中的其他区域则被归类为背景。

标注后的图像被分配了数值标签：背景为 0，叶盘为 1，霜霉病产孢区域为 2。叶盘和霜霉病产孢区域的手动标注掩膜是通过 Python 脚本生成的 (<https://github.com/wkentaro/labelme>)。为确保标注质量的一致性并减少个人偏差，我们找到了八名有经验的葡萄育种实验员参与了标注工作，图像在他们之间随机交换。由于焦距变化会导致某些照片变得模糊模糊，使得即便是经过训练的实验员也难以准确识别，因而对于霜霉病孢子囊不清晰的叶盘，未进行标注。图 1-6 展示了标注后的叶盘和霜霉病孢子囊的代表性样本。

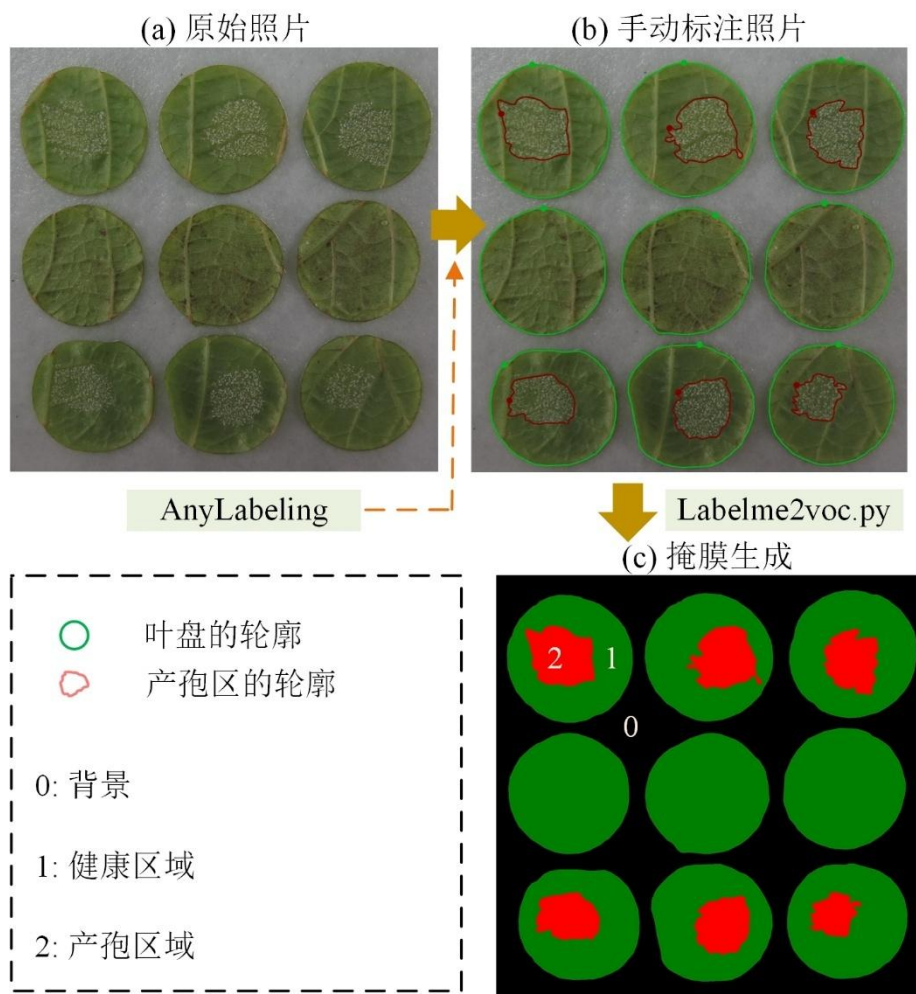


图 1-6 叶盘和霜霉病孢子囊的标注与掩膜生成过程概述。(a) 原始图像展示了健康或感染的叶盘。(b) 使用 AnyLabeling (<https://anylabeling.nrl.ai/>) 进行叶盘和霜霉病孢子囊的标注。绿色轮廓表示由 Mobile SAM 生成的叶盘轮廓，而红色轮廓则是手动标注的霜霉病孢子囊轮廓。(c) 生成的分割掩膜，并为不同类别分配了数值标签。

Figure 1-6 Overview of the annotation and mask generation process for leaf discs and downy mildew sporulation. (a) The original image depicts healthy or infected leaf discs. (b) Annotation for each leaf disc and sporulation using AnyLabeling (<https://anylabeling.nrl.ai/>). Green contours for leaf discs were generated by Mobile SAM, while red contours were manually annotated. (c) Generated segmentation masks with numerical labels assigned to different classes.

1.6.2 葡萄病害特征图文对比数据标注

2.2.2.1 数据组成

我们搜集并整理了葡萄 4 种不同病害的照片数据集，包括霜霉病、葡萄叶斑病、葡萄黑麻疹病和葡萄黑腐病。此外，我们还同时采集了健康葡萄叶片及不当农药使用导致的药害图像照片数据。因此，整个数据集由两部分构成：(1) 田间采集霜霉病与药害图像数据，采集地点为中国广西壮族自治区都安瑶族自治县（地理坐标：23.829 N, 108.110E），我们使用智能手机，如红米 12 以及苹果 12 等，在霜霉病发生以后实

地进行拍摄，通过整理和筛选后得到。（2）另外 3 种病害图像及健康叶片数据则来源于 Plant Village 公开数据集（章节 1.5）。为保证数据集的平衡性，减少对模型训练的影响，我们对 Plant Village 数据集和收集的田间数据进行了随机筛选，以保证每个病害类型中的照片数量与健康叶片、实地采集的两类数据数量相当，便于后续的分析与研究工作。因此，我们最后得到 1 套数据平衡，包含葡萄霜霉病在内的病害照片数据集，其具体病害类型以及照片数据的数量详见表 1-1。

表 1-1 葡萄病害类型、英文名称、致病菌及对应的照片数量。

Table 1-1 Grape disease types, English names, pathogens, and the corresponding number of data.

类型	英文名称	致病菌	照片数量	数据源
霜霉病	Downy mildew	<i>Plasmopara viticola</i>	345	田间采集
药害	Chemical damage	-	368	田间采集
叶斑病	Leaf blight	<i>Cercospora viticola</i>	384	Plant Village
黑麻疹病	Black measles	<i>Elsinoe ampelina</i>	372	Plant Village
黑腐病	Black rot	<i>Guignardia bidwellii</i>	368	Plant Village
健康	Healthy	-	376	Plant Village

我们使用手机拍摄的病害照片分辨率相对较高，为了减少后续分析中数据对显存的占用，我们对采集的高分辨率的田间病害及药害照片进行了重采样。由于在后续分割中的模型输入尺寸问题，我们将田间采集的照片重采样为 224×224 的分辨率，与 Plant Village 数据集的照片分辨率接近。另外，由于 Plant Village 数据集中照片背景相对简单，我们也同样对田间的数据集进行了背景去除处理。类似的，我们使用了 Any Labeling (<https://github.com/vietanhdev/anylelabeling>)，并借助视觉基础模型，Mobile SAM 作为辅助，经过手动提示框对照片中的叶片进行了轮廓的生成，然后使用了 OpenCV4.2 包 (Bradski 2000) 进行位操作对背景进行去除，从而保留了叶片及病斑的区域。我们选择使用背景简单的照片数据，可以明确避免复杂背景的干扰，如土壤以及人手等，使得模型专注以病害叶片及其病斑特征的信息提取。最终，我们得到了 4 种不同葡萄病害类型、1 种药害类型以及健康叶片的数据集（以下统称为病害照片），总共为 2213 张照片。处理后的照片如图 1-7 所示。

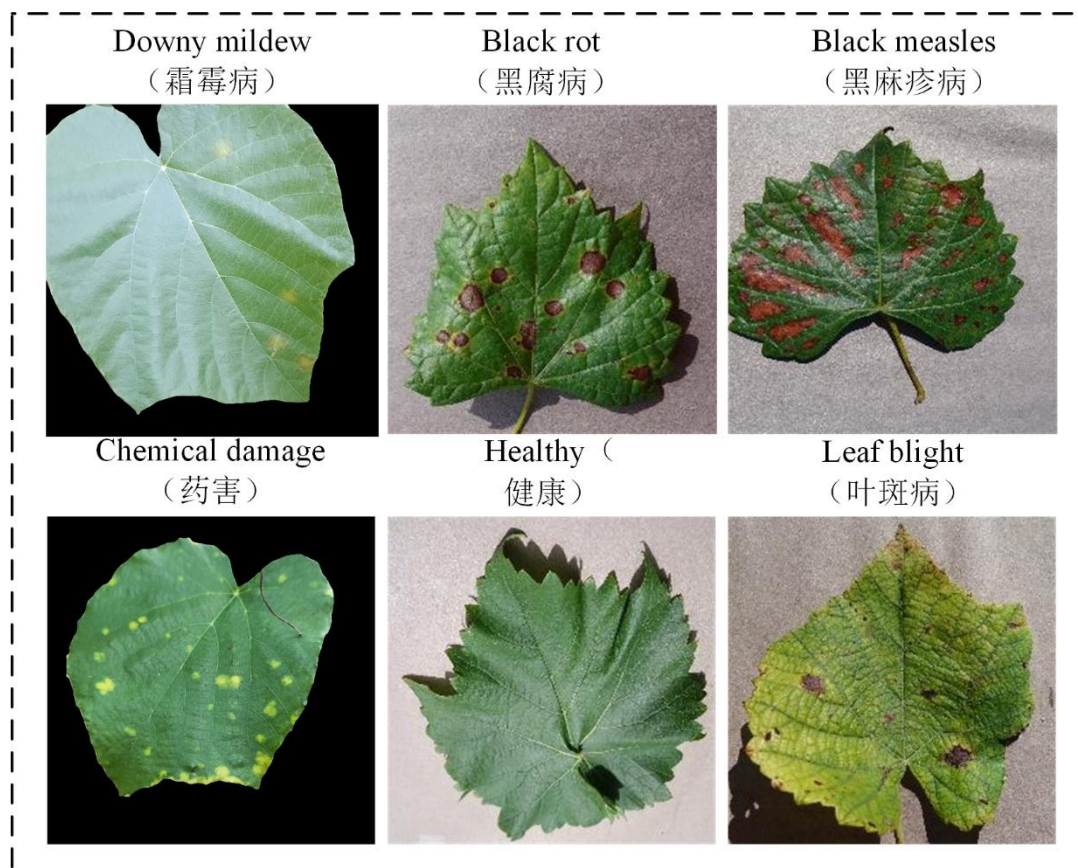


图 1-7 6 种不同类型的病害叶片照片示意图，图片均移除背景以提高对病斑描述的准确性。

Figure 1-7 Illustration of six different types of grape leaf disease images, with backgrounds removed to enhance the accuracy of lesion description by model.

2.2.2.2 大语言模型辅助病害文本注释生成

为了降低人工标注的成本，我们采用了大语言模型（LLM）对不同种类的葡萄病害照片进行了初步的注释，随后由植保专家进行进一步修正，主要关注其病斑特征（病斑颜色、病斑形状、病斑位置、病斑面积），以确保图片描述文本的准确性，流程如图 1-8 所示。利用 LLM 进行图片的初步标注，已被证明在图像标注任务中具有显著效果（Liu Q et al. 2024）。本研究中，我们采用的模型是来自 Kimi 的 moonshot-8k-visual-preview，通过使用 Python 接口 API (<https://kimi.moonshot.cn/>) 将病害照片传输到模型中接口，接受模型返回的响应数据，从而获取到病害照片中病害特征的描述。

对于 LLM 而言，模型的结构化提示非常重要，因此，为了获取到我们需要对病害的颜色、形状、位置和面积四个特征描述，我们设定了特定的提示词，用于得到结构化的病害描述返回。以霜霉病为例，设计的提示词为：“请描述这张显示葡萄霜霉病症状的照片，包括叶片上病斑的颜色、形状、位置以及面积，并以结构化按照顺序输出。”对于无病害的健康叶片，我们直接定义描述信息为：“此为健康的葡萄叶片，未发现病斑。”通过 LLM 的初步标注，我们成功获取了每张葡萄病害照片的详细

描述信息。

预标注完成后，这些描述信息被保存为 CSV 文件，形成了初步的病害文本注释数据集。然而，由于 LLM 模型可能存在理解偏差或错误，生成的描述信息并不总是完全准确。因此，我们组织了一组专业人员对这些描述信息进行人工修正，以确保其准确性和可靠性。修正过程中，专业人员会根据照片的实际内容，对描述信息中的病斑颜色、病斑形状、病斑位置以及病斑面积等细节进行核对和调整。经过人工修正后，我们得到了一个高质量、结构化的病害特征图文对比数据集，为后续的研究和应用提供了有力的支持。

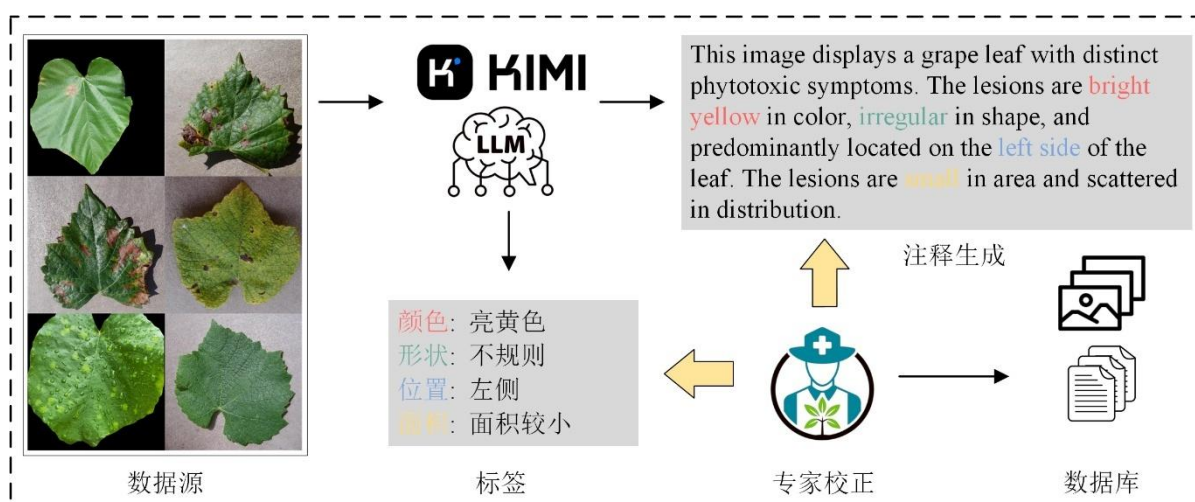


图 1-8 Kimi 模型在植物病害识别中的应用流程。图中展示了从数据源获取葡萄叶图像，通过 KIMI 模型进行图像分析，提取特征标签（如颜色、形状、位置和面积），生成描述性标题，并由专家进行校正后存储到数据库中的整个过程。

Figure 1-8 The workflow of the large language model in grape disease caption. The diagram illustrates the process of acquiring grape leaf images from the data source, analyzing the images through Kimi model to extract the tags of lesions (such as color, shape, location, and area), generating captions, and storing them in the database after expert correction.

1.6.3 田间葡萄健康和霜霉病叶片标注

对于田间健康和霜霉病叶片照片，我们同样采用 AnyLabeling (<https://github.com/vietanhdev/anylabeling>)，进行了详尽的手动以及半自动标注。AnyLabeling 具备利用预训练视觉基础割模型辅助数据标注的功能，可以显著提升标注工作的效率与精确度。为节省计算资源并确保标注流程的顺畅性，我们使用了选取了 Mobile SAM 模型作为基础模型，以辅助进行霜霉病斑及叶片的标注工作。然而，鉴于模型生成的叶片轮廓和病斑的轮廓并不是总符合，我们仍通过人工细化来保证标注保住数据的准确性。本研究的标注数据主要涵盖两大类别：病斑与叶片。图 1-9 展示经过标注处理的图像样本。

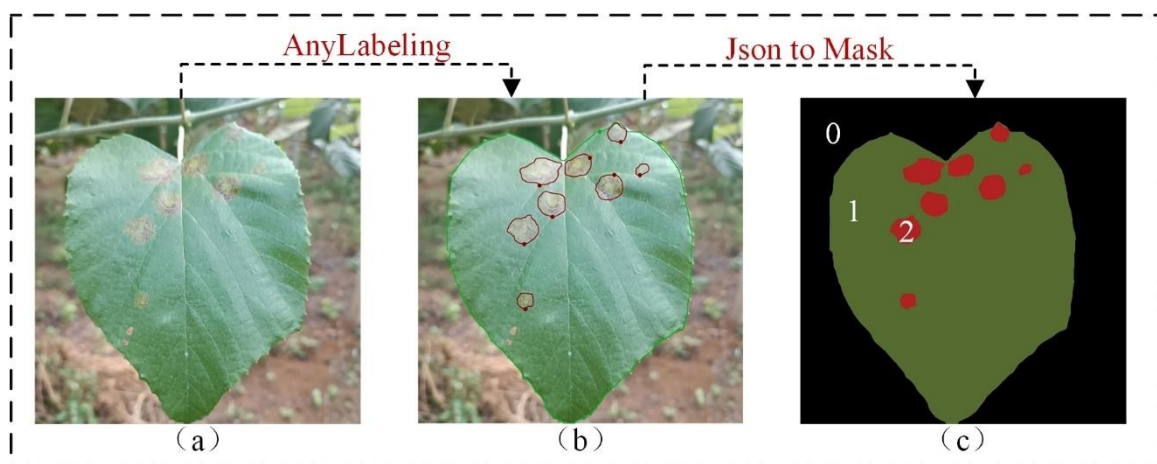


图 1-9 数据标注流程。原始数据使用 AnyLabeling 进行半自动的标注，生成的标注文件使用脚本转化为掩膜标签文件。其中 (a) 原始照片，(b) 为标注的照片，(c) 是生成的掩膜，0 表示背景，1 表示叶片，2 表示病斑。

Figure 1-9 Image annotation workflow. The original images were semi-automatically annotated using AnyLabeling, and the generated annotation files were converted into mask label files via Python script. (a) Original image, (b) Annotated image, and (c) Generated mask, where 0 represents the background, 1 represents the leaf, and 2 represents the lesion.

1.7 训练和评估的数据集划分

1.7.1 葡萄叶盘霜霉病数据集划分

在葡萄叶盘接种霜霉病试验中，我们通过使用手机拍摄和收集叶盘试验中的产孢区域照片数据，并经过半自动的标注后，得到了叶盘及其产孢区域的细致标注数据。其中，照片数据被统一转为 JPG 格式，而掩膜文件则被统一转为 npy 格式，方便后续使用 python 的 Numpy 包 (Harris et al. 2020) 读取，以计算严重程度。最终，我们得到了 432 张培养皿的照片数据，每个培养皿中叶盘数量不同，我们使用标注的叶盘掩膜，通过图像处理将每个叶盘单独提取出来，最后得到约 6000 张叶盘照片。其中，由于拍摄焦距的问题，一些叶盘比较模糊，我们只用来做叶盘检测。

在叶盘检测任务中，我们将提取出来的叶盘，按照 6:2:2 的比例随机划分为了训练集、验证集和测试集 (章节 1.10)。其中，训练集中包括 3582 张照片，训练集和测试集照片数量则比例一致。我们使用验证集监督模型训练，将模型参数固定后，在测试集上进行预测推理，以检验模型的泛化性能。

在叶盘和产孢区域分割任务中，我们首先手动的将模糊的叶盘照片移除数据集，然后将剩下的数据集，重新按照 6:2:2 的比例随机划分为训练集、验证集和测试集 (章节 1.10)。与叶盘检测任务一致，我们使用验证集监督模型训练，将模型参数固定后，在测试集上进行预测推理，以检验模型的泛化性能。

由于需要评估背景去除对叶盘及其产孢区域的分割表现的影响，我们将移除模糊

叶盘照片的数据集，使用标注的掩膜，去除了背景信息，填充为 0 值。然后再次划分为 6:2:2 的训练集、验证集和测试集。因此，在叶盘及其病斑分割任务中，我们使用了两套数据来训练模型，分别是去除背景的，和保留背景的照片数据集（章节 1.10）。

1.7.2 葡萄病害图文配对数据集划分

我们通过田间采集和使用公开的葡萄病害数据集，整理了一套不同葡萄病害的照片数据集。预处理重采样和去除背景后，我们使用了 LLM 进行粗略标注和人工校正后，得到了一套结构化的葡萄病害图文对比数据集。示例图如图 1-10 所示。

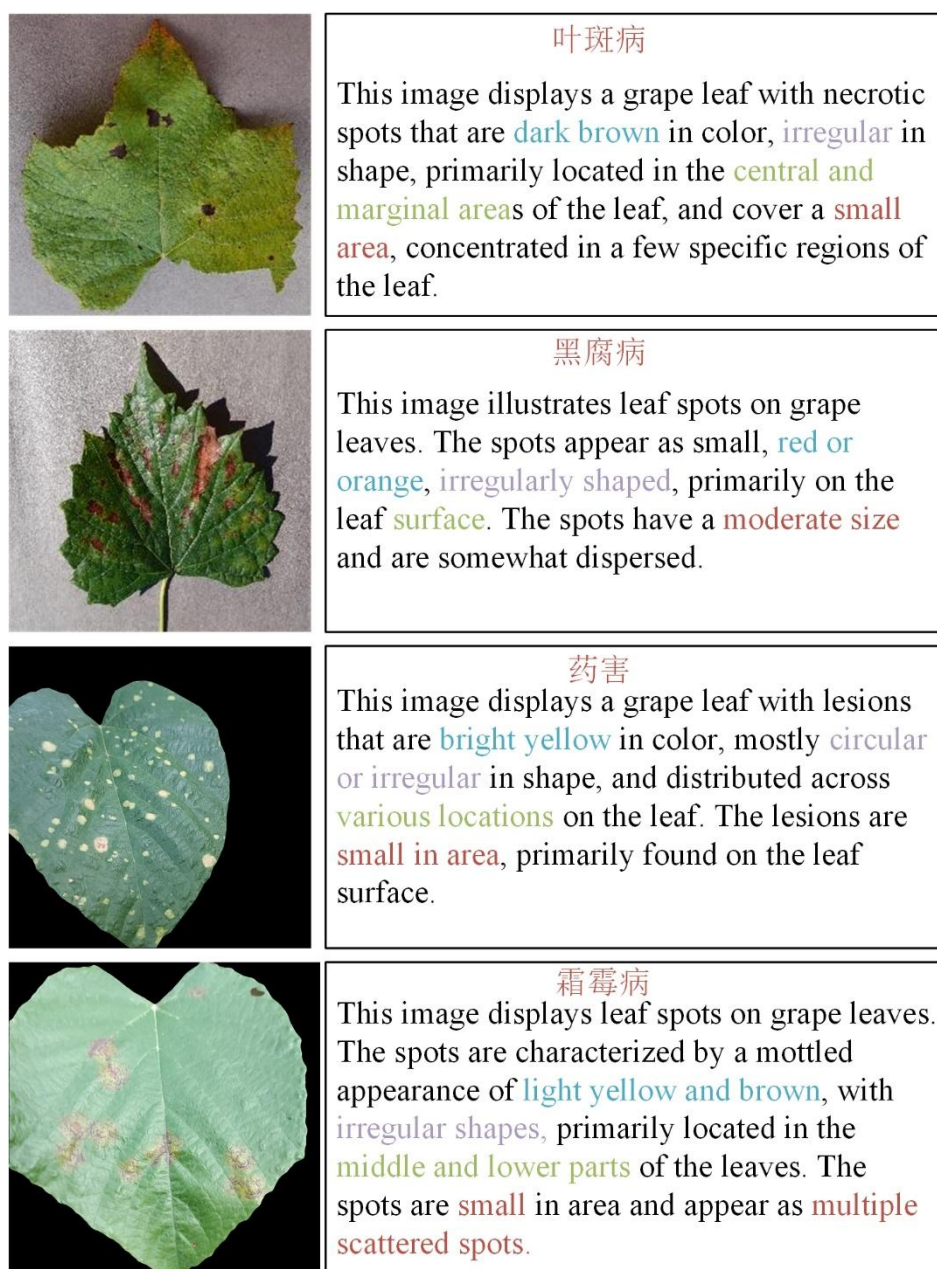


图 1-10 葡萄病害图文配对数据集

Figure 1-10 Image-Text examples of grape diseases

由于微调模型没有用到验证集，我们将数据集分为了训练和测试集，训练集用于微调视觉语言模型，测试集用于评估视觉语言模型的病害描述生成表现。这些数据集中，还包括同时得到的不同病害特征描述数据，如病斑颜色、病斑形状、病斑位置、病斑面积。照片存储路径、照片的病害描述以及各个病斑的特征描述使用 CSV 文件存储。

1.7.3 田间葡萄健康和霜霉病叶片数据集划分

通过在实际葡萄园环境中进行田间实地拍摄，我们采集到了毛葡萄健康叶片和霜霉病叶片数据，经过精细筛选和质量控制，我们最终构建了一套高质量、具有代表性的田间数据集。我们通过了人工标注与算法辅助分割，最终为每张病叶图像生成了对应的掩膜，并同样存储为 `numpy` 格式，以方便计算严重程度，即病斑面积占整张叶片面积的比例。

训练 CNN 分割模型时，如 UNet (Ronneberger et al. 2015)，我们将整理好的数据集按照 6:2:2 的比例随机划分为训练集、验证集和测试集（章节 1.22）。具体来说，训练集包含 412 张图像，验证集和测试集分别包含 137 张和 138 张图像。为确保模型输入的一致性，并与后续轻量化模型 Mobile SAM 读取，我们对所有图像及其掩膜进行了统一采样，调整为 1024×1024 的分辨率。标准化后的尺寸有助于模型高效提取特征，同时避免了由于非标准输入尺寸所引发的空值填充问题，否则可能导致特征图中出现条带状无效区域。

1.8 研究中主要采用的方法和模型

1.8.1 深度学习基本理论与模块设计

卷积神经网络适合用于处理呈矩阵形式排列的数据结构，广泛应用于图像、视频等以矩阵形式输入数据的处理与分析任务。典型的卷积神经网络结构由多个关键模块组成，通常由卷积、池化、激活函数、归一化以及全连接层等模块组成，其中激活函数赋予了网络非线性特征，归一化层可以缓解梯度消失或者爆炸问题，而全连接层则用于分类任务 (O'Shea and Nash 2015)。

2.4.1.1 卷积和池化

卷积层作用于图像特征的提取过程，其计算方式如图 1-11 所示。

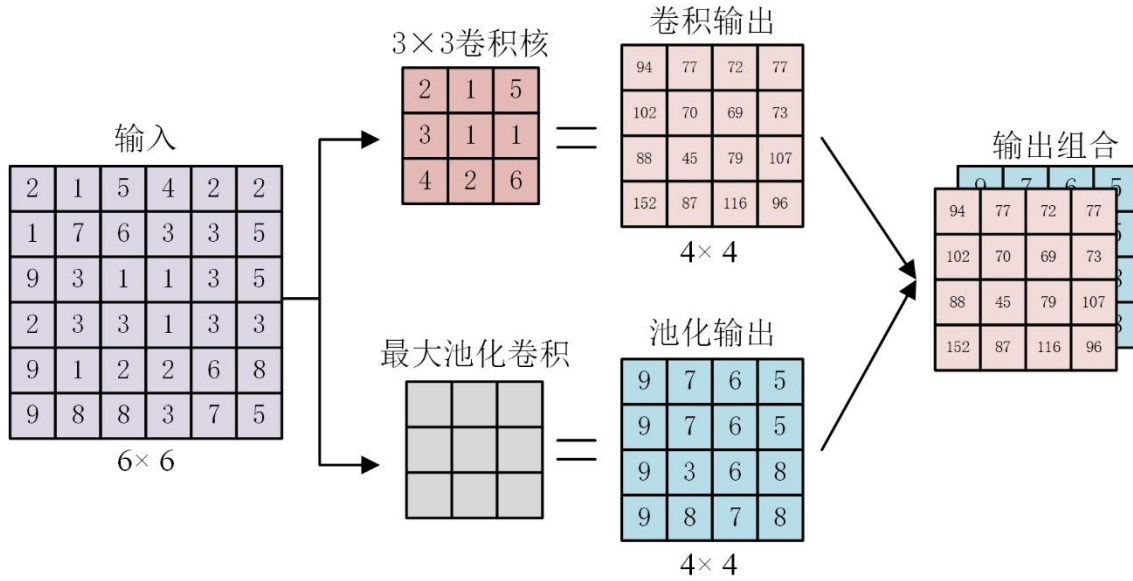


图 1-11 卷积计算示意图

Figure 1-11 Schematic diagram of convolution operation

卷积操作涉及两个输入对象，这两个对象通常是“实数值函数”，例如图像（可以视作二维实值输入）和卷积核（滤波器函数）。通过特定的滑动和加权相乘方式，将两者组合起来，得到一个新的函数或结果。在计算机视觉中，由于图像由有限数量的像素构成，卷积操作就变成了在这些离散像素位置上进行的一种加权求和运算，而不再是数学中那种需要积分的连续卷积形式，其公式：

$$y(t) = \sum_{\tau} x(t+\tau) \cdot a(\tau) \quad (1-1)$$

其中 $a(\tau)$ 表示在位置 τ 处的权重， $x(t+\tau)$ 表示在 t 时刻位置 τ 处的输入，最后得到求和结果。

在实际的二维卷积运算中，通常使用以下函数表征卷积过程：

$$Y(i, j) = \sum_m \sum_n X(i+m, j+n) K(m, n) + b \quad (1-2)$$

上述公式说明，在图像特征提取的过程中，卷积核会在图像上逐像素（或按照步长）滑动，每滑动到一个位置，就把卷积核里的每个数与图像中对应位置的像素值相乘，然后把所有乘积加起来，在这个求和结果上加一个偏差值。训练开始时，卷积核的值不是手动设定的，而是随机赋值，然后通过反向传播和梯度下降等优化算法，根据误差反馈，一步步调整这些参数，使得网络的预测越来越准确。

在进行卷积时，如果不对图像边缘做处理，图像会逐层“缩小”。为了避免图像尺寸过快变小，可以在图像四周补上一圈（或多圈）零，以保持特征图的尺寸。同时，设

置步长可以降低特征图尺寸,从而减少计算量。实际上,输入的 RGB 图像是三维向量,卷积核被设置为三维的向量,因此通过设置多个卷积核,每个核会生成一个输出特征图,最终输出的通道数就是卷积核的数量。池化层可以主动的降低特征图的维度,提取特征中的摘要信息,从而减少计算的参数数量。它通过使用 $k \times k$ 的卷积核,来提取对应像素的信息,常用的池化操作,包括最大值池化和均值池化,即保留了图像特征信息,也能降低计算量。

2.4.1.2 激活函数

卷积运算是一个连续的线性计算过程,不能够拟合非线性关系。因此,通过引入非线性的函数,来建造模型的非线性建模能力。这些函数即为激活函数。激活函数的优势在于不仅使得模型的特征值非线性变化,也可以通过降低不重要的特征值大小,来实现决策功能。在深度学习中,常用的激活函数为 Sigmoid, Tanh、ReLU 以及 Leaky ReLU,其公式如下:

$$\text{Sigmoid}(x) = \frac{1}{1+e^{-x}} \quad (1-3)$$

$$\text{Tanh}(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \quad (1-4)$$

$$\text{Relu}(x) = \max(0, x) \quad (1-5)$$

$$f(x) = \begin{cases} x, & \text{if } x \geq 0 \\ 0.01x, & \text{if } x < 0 \end{cases} \quad (1-6)$$

式中, x 为输入特征变量。这四个激活函数如图 1-12 所示。

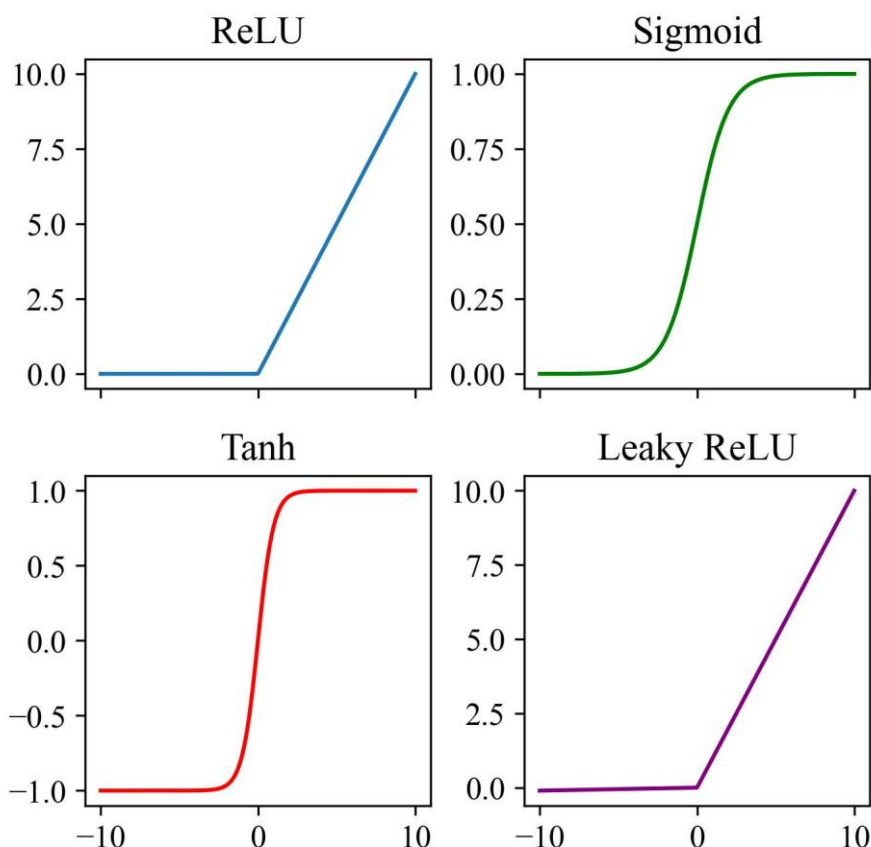


图 1-12 不同激活函数的示意图，激活函数使得神经网络具有非线性能力

Figure 1-12 Different activation functions, which enable neural networks to model nonlinear ability.

2.4.1.3 损失函数

损失函数衡量神经网络的预测结果与真实值之间的差距。网络预测出结果后，通过损失函数对比真实值与预测值的“差距”，然后利用反向传播算法调整网络参数，使损失变小，模型更准确。在分类任务中，*Softmax* 函数配合交叉熵损失函数常被用作标准组合，前者输出预测概率，后者计算损失值，以实现高效准确的模型训练，其定义如下：

$$y_i = \frac{e^{z_i}}{\sum_j e^{z_j}} \quad (1-7)$$

式中， z_j 为模型输出最后一层在位置 i 的值，分母为每个位置值相加求和。*Softmax* 将每个值归一化到 0 至 1 的区间，代表每一类的概率值。因此，只需要使用交叉熵函数将生成的概率值向量与类别标签的编码进行点积计算，即可得到损失值。交叉熵函数公式定义如下：

$$L = -\sum_{i=1}^C y_i \log y_i \quad (1-8)$$

式中 $\hat{y}$ 表示正确的分类标签， y_i 则为模型的输出，其中 C 表示类别正确的索引。

2.4.1.4 优化器

在卷积神经网络的训练过程中，通常采用反向传播算法来计算损失函数对各可学习参数的梯度，随后利用不同的梯度优化算法（如梯度下降）对模型的参数进行迭代更新，从而引导模型逐步收敛至最优点（LeCun et al. 2015）。模型的权重更新可表示为，当模型进行第 t 次迭代训练时，权重参数 w_t 的更新方式为：

$$W_t = W_{t-1} - \alpha * \nabla W_{t-1} \quad (1-9)$$

式中， α 是学习率，一般在初始阶段设置为 0.001。优化器可以根据损失值的变化调整学习率。

在实践中，神经网络通常不会在整个训练集上计算梯度（太慢），而是在每次迭代中使用一个小批量的数据来估算梯度。小批量样本具有随机性，导致梯度方向不稳定，可能使优化过程震荡、收敛速度变慢，甚至陷入局部最优，因此被称为随机梯度下降（Stochastic Gradient Descent, SGD）。通过引入历史梯度的动量（momentum），多个方向更新的基础上，保留之前更新的趋势，可以加快收敛、减小震荡。加入 momentum 的训练方式变成：

$$\nabla W' = \eta \nabla W_{t-1} + (1-\eta) \nabla W_t \quad (1-10)$$

$$W_{t+1} = W_t - \alpha \nabla W' \quad (1-11)$$

式中， η 为动量系数，一般为 0.9，表示 90% 的梯度来自于迭代前积累的历史梯度，剩下 10% 来自于本批次。其他的优化器则在此基础上进行了进一步的改进，本章不在赘述。

2.4.1.5 反向传播算法

反向传播算法是深度神经网络中用于计算参数梯度的标准方法，为模型训练提供了高效且可扩展的优化路径。该算法基于链式法则（Chain rule），将整体的梯度计算过程分解为网络各层的局部计算，从而显著降低了直接计算高维梯度所带来的计算成本（Rumelhart et al. 1986）。现代深度学习框架（如 PyTorch）的自动微分功能直接基于反向传播算法实现（Paszke et al. 2017）。反向传播在中间层的计算公式为：

$$\delta^{(l)} = \left(W^{(l+1)\top} \delta^{(l+1)} \right) \odot f' \left(z^{(l)} \right) \quad (1-12)$$

式中， $\delta^{(l)}$ 表示第 l 层的误差项，反映该层中每个神经元对整体损失的贡献，本质

是损失函数关于该层激活前输出 $z^{(l)}$ 的梯度。 $W^{(l+1)}$ 表示第 $l+1$ 层的权重矩阵，描述从第 l 层到第 $l+1$ 层神经元间的连接强度。通过矩阵计算将它传递到第 $l+1$ 层。 $f'(z^{(l)})$ 为激活函数的导数，衡量神经元输出对输入变化的敏感程度。 $z^{(l)}$ 表示为第 l 层的输入向量。 $\oplus$ 表示逐元素乘积，为误差项与激活导数的逐元素乘积。

2.4.1.6 Transformer 结构

Transformer 结构在自然语言处理中应用广泛，其核心的结构为自注意力机制。该模型由两大部分组成：编码器（Encoder）和解码器（Decoder）。其中，编码器的功能是将原始输入序列转换为潜在的表示向量；而解码器则基于该潜在表示生成对应的目标输出序列（Vaswani et al. 2017）。Transformer 结构如图 1-13 所示。每个编码器和解码器模块由多头自注意力机制、前馈神经网络、层归一化和残差连接组成。其中，自注意力机制采用相关性计算了输入序列中每个位置与其他位置的权重，可以捕捉全局上下文信息，从而实现并行处理。其计算公式如下

$$Attention(Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right) \quad (1-13)$$

$$\begin{aligned} Q &= XW^Q \\ K &= XW^K \\ V &= XW^V \end{aligned} \quad (1-14)$$

式中， X 是输入序列的表示（通常是 embedding 或前一层的输出）， W^Q ， W^K 和 W^V 是可训练的权重矩阵， K ， Q ， V 分别是 Query、Key 和 Value， d_k 是 Key 的维度，用于缩放。

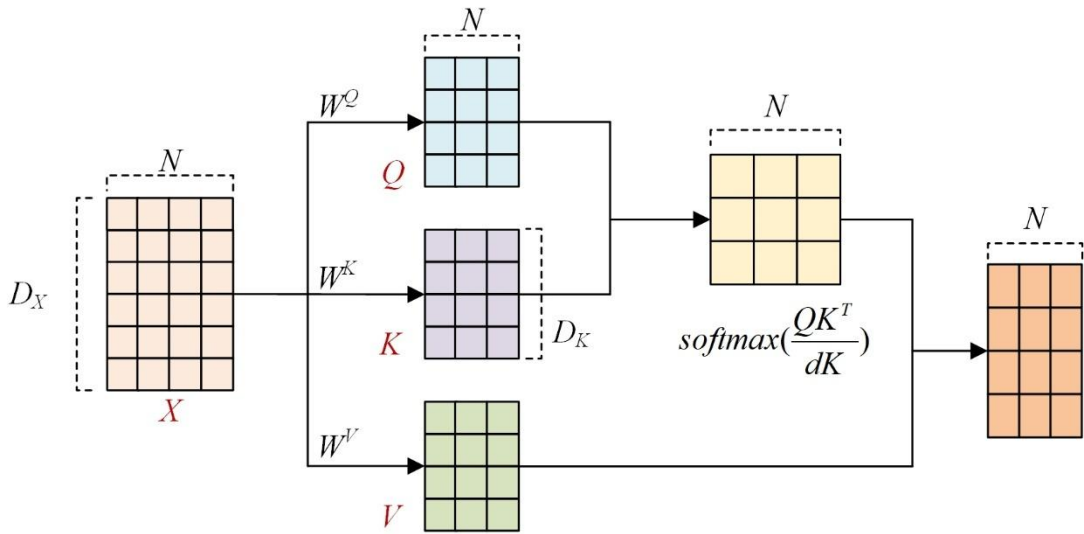


图 1-13 Transformer 结构中的自注意力模块，包括三个部分：K (key)，Q(query)，V(value)。

Figure 1-13 Self-attention module in the Transformer, consisting of three components: K (Key), Q (Query), and V (Value).

Vision Transformer (ViT) 是将 Transformer 应用于图像任务的模型，其结构基于对图像进行分块处理。首先，ViT 通过卷积过程，将输入图像分割为固定大小的图像块，对每个块进行编码以生成特征向量，再加上位置编码以保持块之间的相对位置关系。这些嵌入作为输入，通过 Transformer 编码器，利用多头自注意力机制和前馈神经网络进行全局特征提取，最终通过分类头（如 MLP 层）完成任务（Dosovitskiy et al. 2021）

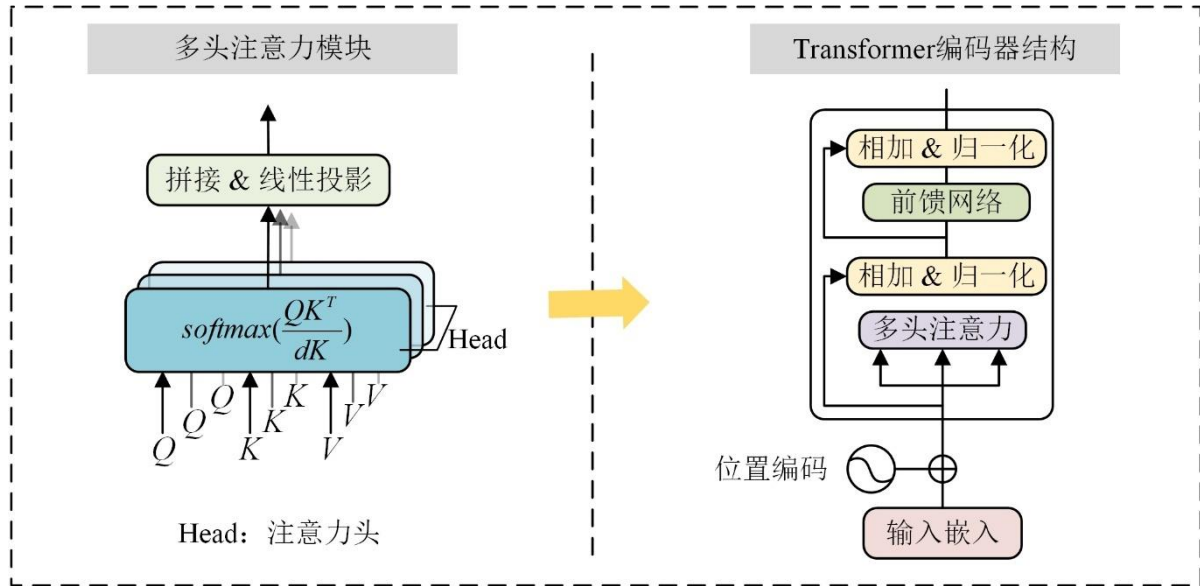


图 1-14 Transformer 结构中的多头注意力和其编码器结构。左侧为多头注意力，右侧为编码器结构，包括位置编码、多头注意力以及前馈神经网络。

Figure 1-14 Multi-head attention and encoder structure in the Transformer. The left side shows the multi-head attention mechanism, while the right side illustrates the encoder structure, which includes positional encoding, multi-head attention, and a feed-forward neural network.

1.8.2 YOLO v8 系列目标检测模型

目标检测是机器视觉的核心任务之一，主要目的是定位和分类图像中的目标。传统的方法通常依赖手工设计特征提取算子和机器学习方法。这些特征提取算子包括颜色特征、质地特征以及边缘特征等，将这些特征输入随机森林，或者支持向量机等算法中，以对图像中的目标进行定位和分类（Govardhan and M B 2019）。深度学习的兴起推动了目标检测的革命性发展，主要是得益于 CNN 的成功。在这个过程中，主要发展了单阶段和双阶段的目标检测模型。

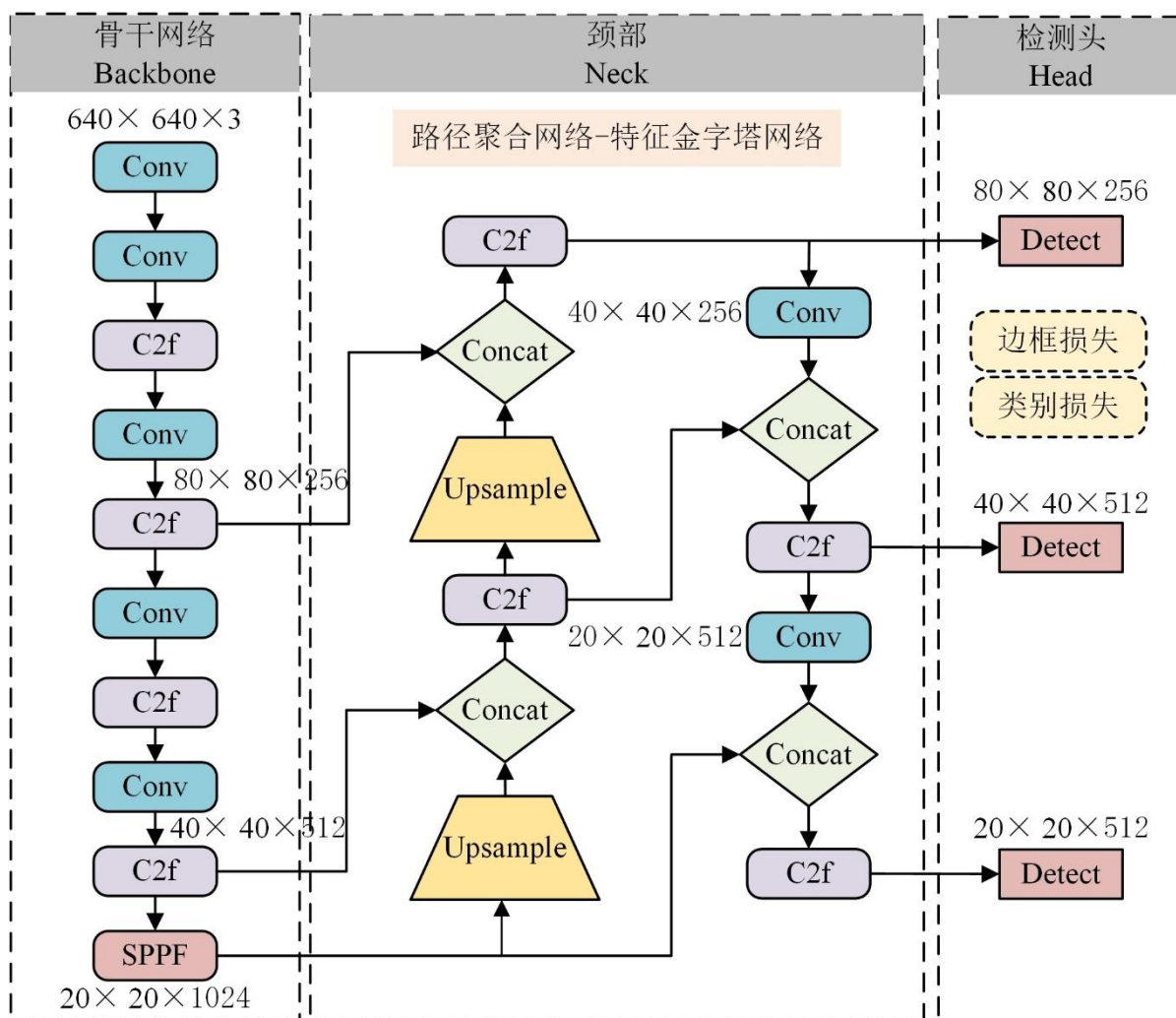


图 1-15 YOLO v8 模型的结构示意图，其中不同颜色部分代表了不同的卷积模块。

Figure 1-15 Diagram of the YOLO v8 architecture, where different colored sections represent different convolution modules.

单阶段模型直接在整个图像上进行目标检测，速度快，缺点是可能存在精度不足。典型的单阶段模型包括 YOLO (Jocher et al. 2023)、SSD (Liu et al. 2016)、RetinaNet (Lin et al. 2018) 等。这里介绍两类常用的单阶段目标检测网络 YOLO 和 SSD。

YOLO 模型在 2015 年首次提出，将预测边框和分类整合到一个网络中，并将图像分割为固定大小的网格，在每个网格直接预测目标类别、边界框和置信度，从而减小模型参数和计算量。此后，YOLO 系列模型不断进化，目前已经到 V11 版本。YOLO v2 引入了预定义的锚框，支持检测 9000 多种物体类别，显著增强了多功能性和准确性。YOLO v3 结合特征金字塔 (FPN) 引入多尺度预测和更深的网络架构 (DarkNet-53)，进一步提高了小目标的检测效果。YOLO v4 加入 Cross-Stage Partial (CSP) 连接和 Mosaic 数据增强。YOLO v5 由 Ultralytics 开发，提供四种不同规模

的模型，在嵌入式设备上进一步优化了易用性和性能。YOLO v6 使用了 RepVGG 结构，并减少了锚框设计的复杂性，在部署性能（如实时性和硬件适配性）方面优于 YOLOv5。YOLO v7 实现了目标检测、目标分割和实例分割的统一框架。YOLO v8 则进一步优化了其多任务能力。本部分主要展示了 YOLO v8，其结构如图 1-15 所示。

双阶段模型则通过设置特征提取器，在第一阶段，通过区域建议网络来生成潜在的候选框，在第二阶段，候选区域选择来进行目标分类和定位，如 RCNN，Fast RCNN 和 Faster RCNN。其中，RCNN 利用选择性搜索算法（Selective Search）在图像中提取 2000 个左右可能的候选框，然后输出到 CNN 和分类器中进行预测。Faster RCNN 在候选区生成之前使用 CNN 进行特征提取，并使用感兴趣区池化（ROI pooling）进行固定大小的特征映射，然后，使用多任务损失对模型进行端对端的训练。Faster RCNN 则基于 Fast RCNN 进一步优化。基于本研究内容，主要介绍 Faster RCNN（Ren et al. 2016），其结构如图 1-16 所示。

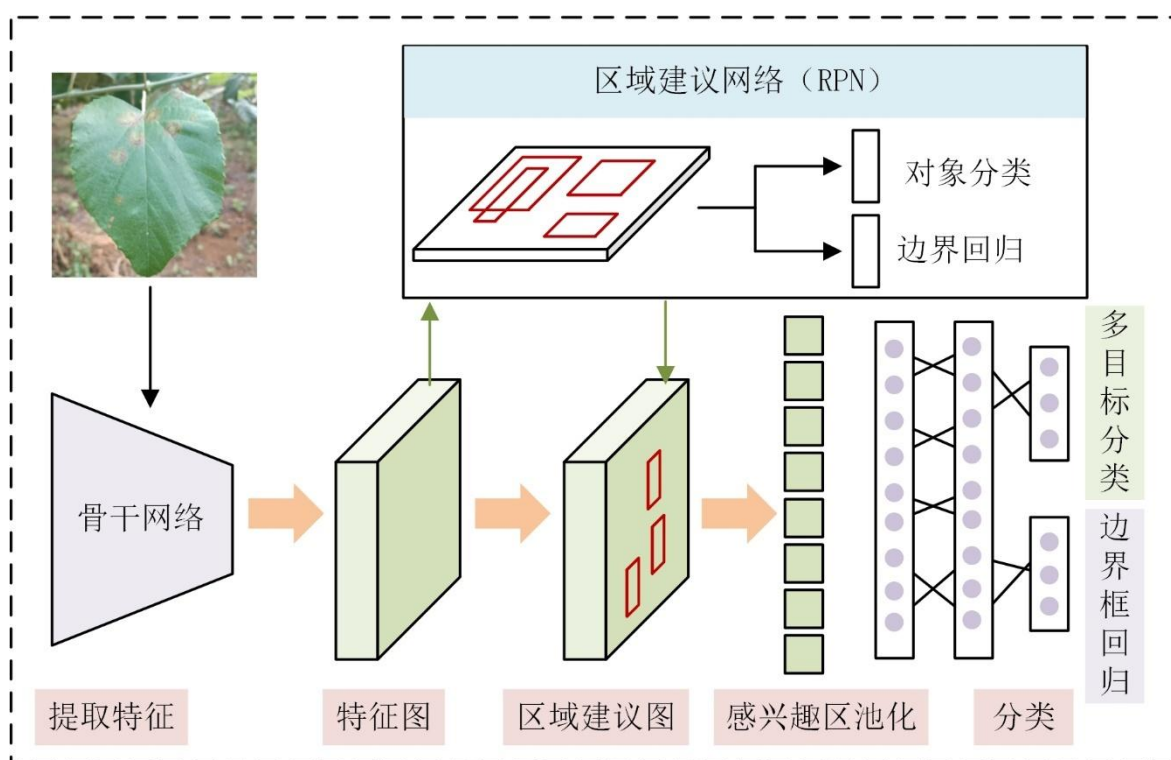


图 1-16 Faster RCNN 的模型结构示意图，其包含两个分支：区域建议网络和分类网络。

Figure 1-16 Illustrative diagram of the Faster R-CNN, which includes two branches: the Region Proposal Network (RPN) and the classification network.

Faster RCNN 主要包含 4 部分：骨干网络，区域建议网络，感兴趣区池化以及分类网络。骨干网络用于提取图像的特征图，然后输入到 RPN 中。RPN 对特征图进行处理，主要为在每一个 3×3 的卷积窗口上根据锚点（Anchors）生成 9 个不同比例的候选区（proposals），然后根据 ROI pooling 进行特征映射，将候选区的特征固定大小

后，通过全连接层进行目标分类，并再次调整边界框的位置和大小。生成的候选区通过非极大值抑制（Non-maximum suppression, NMS）筛选出最优的预测框。

1.8.3 UNet 以及 UNet++ 图像分割模型

传统的图像分割算法包括阈值分割，迭代算法以及颜色指数，这些算法在图像标准的情况下可以实现较好的表现，但是泛化性差（Ahmed et al. 2021; Jothiaruna et al. 2019）。近年来，深度学习技术为该问题提供了有效的解决方案。语义分割主要包括全卷积网络 FCN（Long et al. 2015）、DeepLab（Chen et al. 2017）系列、UNet（Ronneberger et al. 2015）系列等。实例分割包括 Mask RCNN（He et al. 2018）等，全景分割主要包括 Panoptic FPN 等。根据本研究内容，主要介绍语义分割和实例分割任务，包括 UNet 系列模型，DeepLab 系列模型，和 Mask RCNN。

FCN 是使用深度学习进行语义分割的开山模型，之后的分割模型大都采用的编码器-解码器结构。FCN 网络首先通过 VGG 等网络对图像进行编码，提取到图像的特征图。不同的是，在网络的最后一层，FCN 将全连接层替换为上采样层（双线性插值的反卷积操作），从而在整个图像上进行 *softmax* 运算，对图像进行解码，得到语义分割图。同时，通过设置不同的上采样比例，并结合跳跃连接与编码器特征图融合，FCN 可以得到更精细的分割结果。

UNet 则继承了 FCN 的编码器-解码器结构，最初为医疗影像分割产生，但在其他分割任务中也取得了优异的表现（Ronneberger et al. 2015）。与 FCN 相比，UNet 同样采用了跳跃连接和上采样，主要的区别为 UNet 的编码器和解码器是对称结构，深度一致，以及 UNet 采用了将编码器的特征图与相应的解码器特征图拼接，而不是相加。UNet 的结构如图 1-17 所示。而 UNet++ 是一种具有密集跳跃连接和多层嵌套结构的改进型（Zhou et al. 2018）。

图 1-17 典型的 UNet 模型示意图，包括下采样层和上采样层，以及跳跃连接。

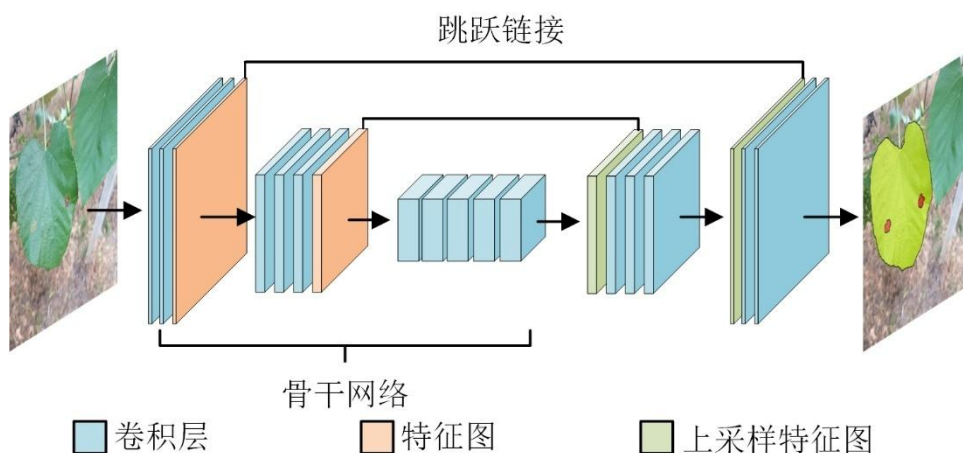


Figure 1-17 Schematic diagram of a typical U-Net model, including down-sampling and up-sampling layers, as well as skip connections

1.8.4 Mobile SAM 提示分割模型

SAM 的出现,使得视觉任务零样本泛化能力大大增强,因此该模型也被称为视觉基础模型 (Kirillov et al. 2023b)。由于 SAM 使用的参数量较大的图像编码器 (ViT-H),在后续的研究中,出现了众多的 SAM 变体。这些变体包括 Mobile SAM (Zhang et al. 2023b),EdgeSAM (Zhou et al. 2024) 等,均在尽可能减少精度损失的前提下,降低了模型的参数量。根据本论文的内容,这里主要介绍 SAM 以及 Mobile SAM 的结构和应用。

SAM 的基本结构包含三部分:图像编码器,提示编码器,掩膜解码器,如图 1-18 所示。SAM 使用基于 Vision Transformer (ViT) 的模型作为图像编码器,利用 ViT 强大的全局特征捕获能力,使得模型能够理解复杂的上下文信息,将输入图像 (1024×1024) 嵌入到高维特征空间,生成全局的特征图。提示编码器接受用户提供的多模态提示信息,如点、边界框、文本等,其中点提示包括前景点和背景点,编码为二维坐标,并通过正余弦位置编码;边界框提示使用坐标位置编码,并额外处理边界框的形状和尺寸信息,文本提示则通过文本嵌入器处理文本提示。掩膜解码器通过文本嵌入器处理文本提示,它是一个轻量级的 Transformer 模块,具有多头注意力机制,输入的图像特征与提示特征经过多次注意力计算后生成掩码,能够高效地融合多模态信息。

SAM 的训练数据来自 Meta AI 构建的大规模分割数据集 SA-1B,该数据集包含超过 1 亿个高质量分割掩码。该数据集涵盖多种场景和物体类别,多样的数据保证了模型的泛化性能。SAM 在大规模分割数据集上进行预训练,从而学习通用分割特征。尽管 SAM 在分割任务中表现出色,但它仍然存在一些局限性,主要包括对小目标和边界细节的处理有限,以及对提示质量的依赖性强,而且计算开销较高,存在偏向某些特定场景或类别的问题 (Ji et al. 2023)。这意味在特定领域或下游任务中,需要进一步优化 SAM 的提示,或者微调其参数,以及减小计算开销。

SAM 的低效主要源于其重量级的图像编码器,因此针对 SAM 的变体主要在于设计新的或者替换为更高效的架构,通过模型蒸馏、模型剪枝等方法来实现。其中,Mobile SAM 使用了知识蒸馏的技术,在保证精度的前提下,模型参数减少了 60 倍。

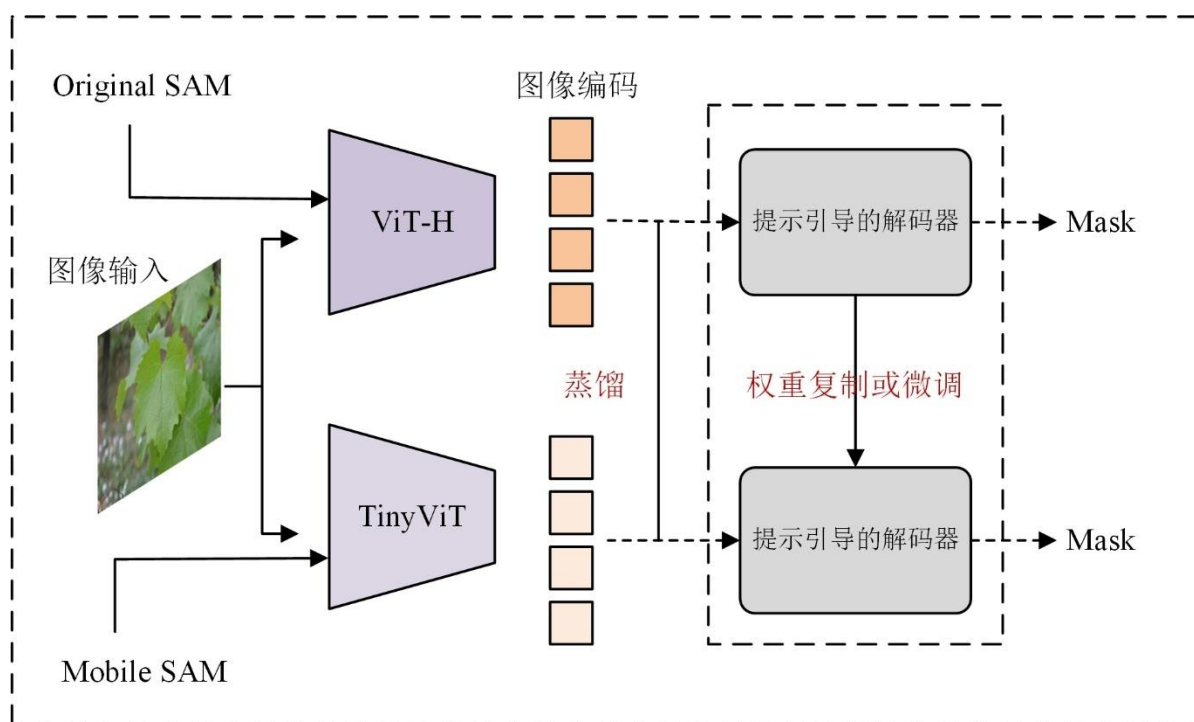


图 1-18 Segment Anything Model (SAM) 的主要模型结构图，

以及 Mobile SAM 通过替换其 ViT-H 编码器为 TinyViT 得到轻量级提示分割模型。

Figure 1-18 Diagram of the main model structure of the Segment Anything Model (SAM), along with the lightweight Mobile SAM, which is obtained by replacing the ViT-H encoder with TinyViT to create an efficient prompt-based segmentation model.

Mobile SAM 是一种针对移动设备优化的轻量化实例分割模型，基 SAM 的设计思想，旨在实现高效、实时的分割性能。Mobile SAM 的架构在保留 SAM 核心功能的同时，重点进行模型压缩与计算优化，其架构如图 1-18 所示。

Mobile SAM 使用了轻量化图像编码器 (ViT-Tiny) 替代 ViT-H，减少计算量和参数量。先是对模型进行解耦，即在图像编码器上进行知识蒸馏，使用均方误差损失来保证蒸馏得到的特征图与原始 SAM 模型接近。接下来，保留掩码解码器结构，使用蒸馏得到的特征图，结合提示编码器，对掩码解码器进行微调。模型结果表明，通过解耦蒸馏的训练方法，实现了比原始 SAM 更快的推理速度。

Mobile SAM 的轻量化设计使其适用于资源受限的环境，可以被广泛的应用于实时图像分割，适合实时、低功耗应用场景，为实例分割技术的实际落地提供了新的可能性。

1.8.5 Qwen-2.5-VL-3B 视觉语言模型

视觉语言模型是一类能够同时处理图像和文本信息的多模态人工智能模型，其主要结构如图 1-19 所示。它们是连接计算机视觉（CV）和自然语言处理（NLP）两个领域的关键技术，在图文理解、图文生成、图像问答、图像描述生成等任务中表现出色。视觉语言模型主要连接了视觉编码器和语言生成器，通过对图像的编码和文本的编码融合，来达到学习对齐关系的目的。比如 BLIP2，通过一个 Q-former 结构连接视觉编码器和大语言模型如 Vicuna，作为桥梁来实现图文理解（Li et al. 2023）。Qwen-2.5-VL-3B 是阿里云推出的 Qwen-2.5 系列中的多模态视觉语言模型（VLM），其具备较强的图文理解与对话能力，模型规模为 3B（30 亿参数），属于轻量级多模态大模型（Bai et al. 2025）。其语言能力来源于 Qwen-2.5 LLM 系列，支持中文和英文等多语言输入，尤其对中文有良好适配。本文选用 Qwen-2.5-VL-3B 是因为考虑到其参数量相对较小，以及适配中文。

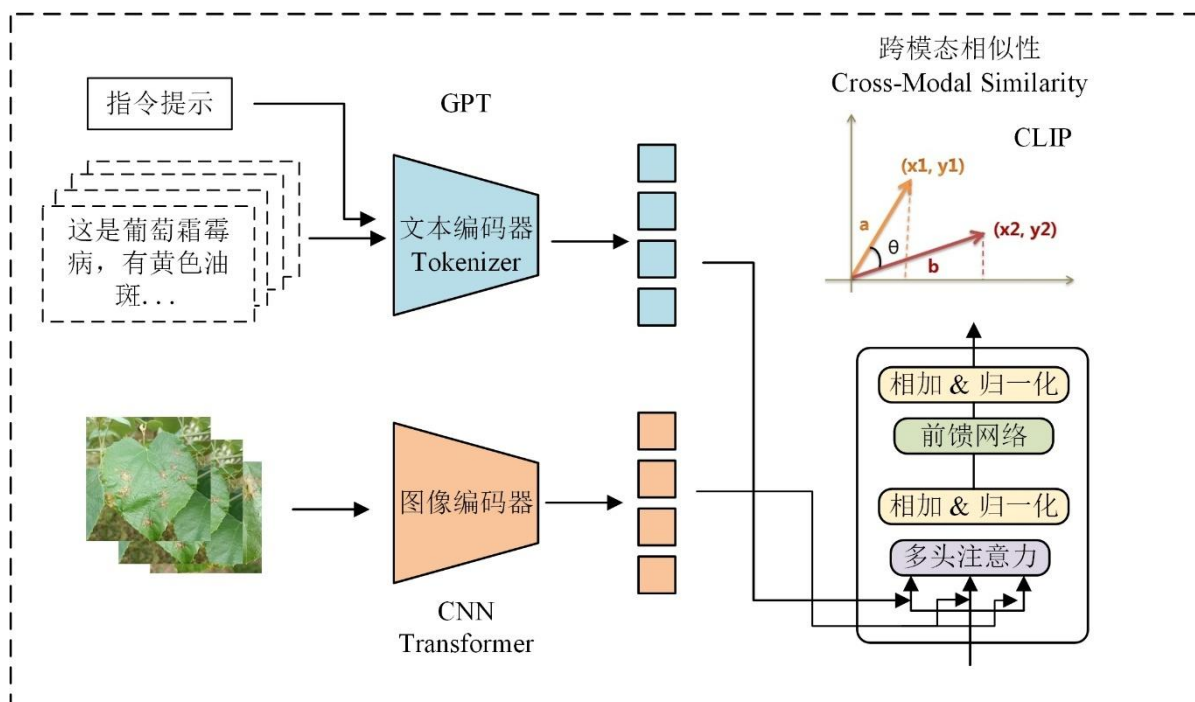


图 1-19 视觉语言模型的基本组成

Figure 1-19 The basic architecture of visual-language models

SporeScan: 高通量叶盘霜霉病表型鉴定的序列化模型

在实验室中，通常使用人工来鉴定葡萄霜霉病感染的严重程度，不仅费时费力，还存在一定的主观性。近年来，随着计算机视觉和深度学习的进步，很多研究使用了图像处理或者卷积神经网络来完成这项任务。然而，这些算法面临这数据采集成本高、泛化性弱的问题（Hernández et al. 2022）。因此，在本章节中，我们使用了手机拍摄的不同分辨率的葡萄叶盘霜霉病照片数据，开发了一个组合的网络，即 YOLO v8n（叶盘定位），Mobile SAM（背景去除）和 UNet（叶盘和病斑分割），来实现高通量、低成本、高精度的叶盘霜霉病自动诊断模型（图 1-1）。

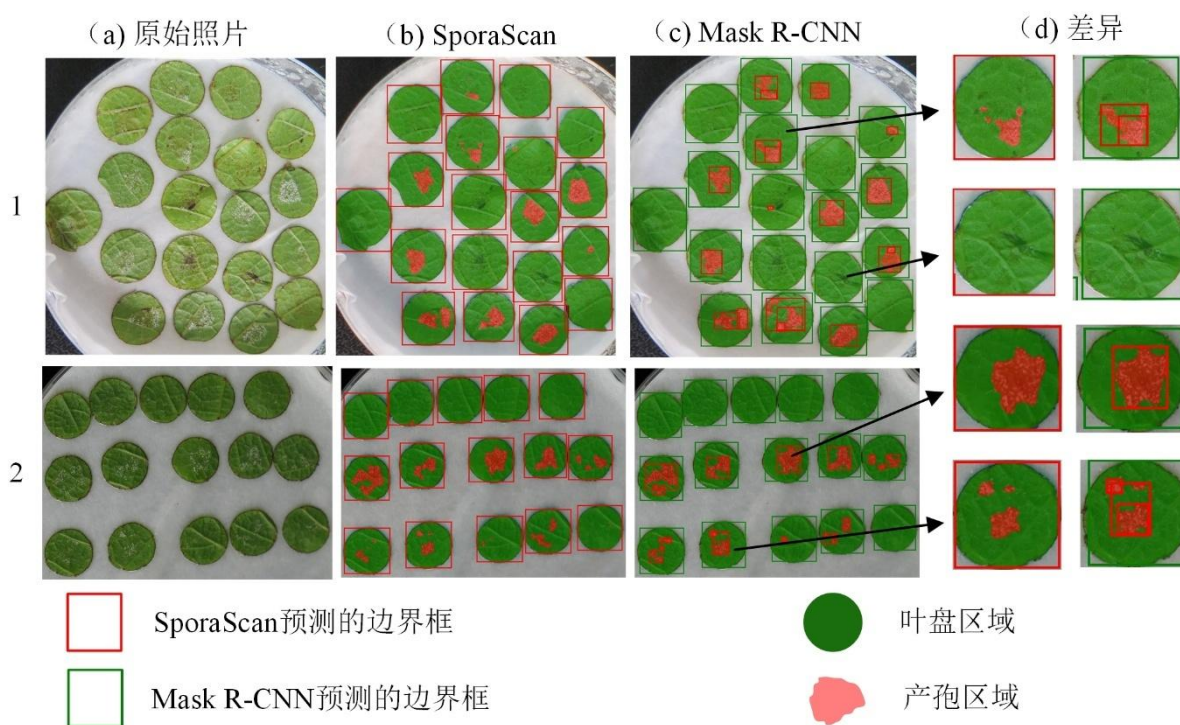


图 1-1 在培养皿中放置叶盘以评估病害严重程度。(a) 原始图像显示每个叶盘上的孢子繁殖区域或健康区域。(b) 每个叶盘的严重程度评估结果，展示了感染区域和健康区域，其中黄色数值表示相应计算出的病害严重程度。传统的计算机视觉技术受限于通用性较差，而以往采用卷积神经网络（CNNs）的研究主要集中于单个叶盘的评估。

Figure 1-1 Assessment of disease severity in leaf discs placed within a Petri dish. (a) Original image depicting sporulation or healthy regions on each leaf disc. (b) Severity assessment results for each individual leaf disc, illustrating the areas of infection and healthy regions, with yellow value representing the corresponding calculated severity. Traditional computer vision techniques are constrained by limited generalisability, while previous studies employing convolutional neural networks (CNNs) have predominantly focused on the evaluation of a single leaf disc.

1.9 概述

1.9.1 SporeScan 模型开发

为了高效准确进行葡萄叶盘上霜霉病严重度的评估，我们开发了自动化模型 SporaScan。完整的工作流程如图 1-2 所示。SporaScan 开发的主要目标如下：（1）从 YOLO v8 系列和 Faster R-CNN 中选择最适合的检测模型进行叶盘定位，（2）评估使用 Mobile SAM 进行背景移除对分割性能的影响，以及（3）选择适当的编码器-解码器模型，用于叶盘和霜霉病孢子囊的分割。每个阶段的具体细节将在以下小节中进行描述。

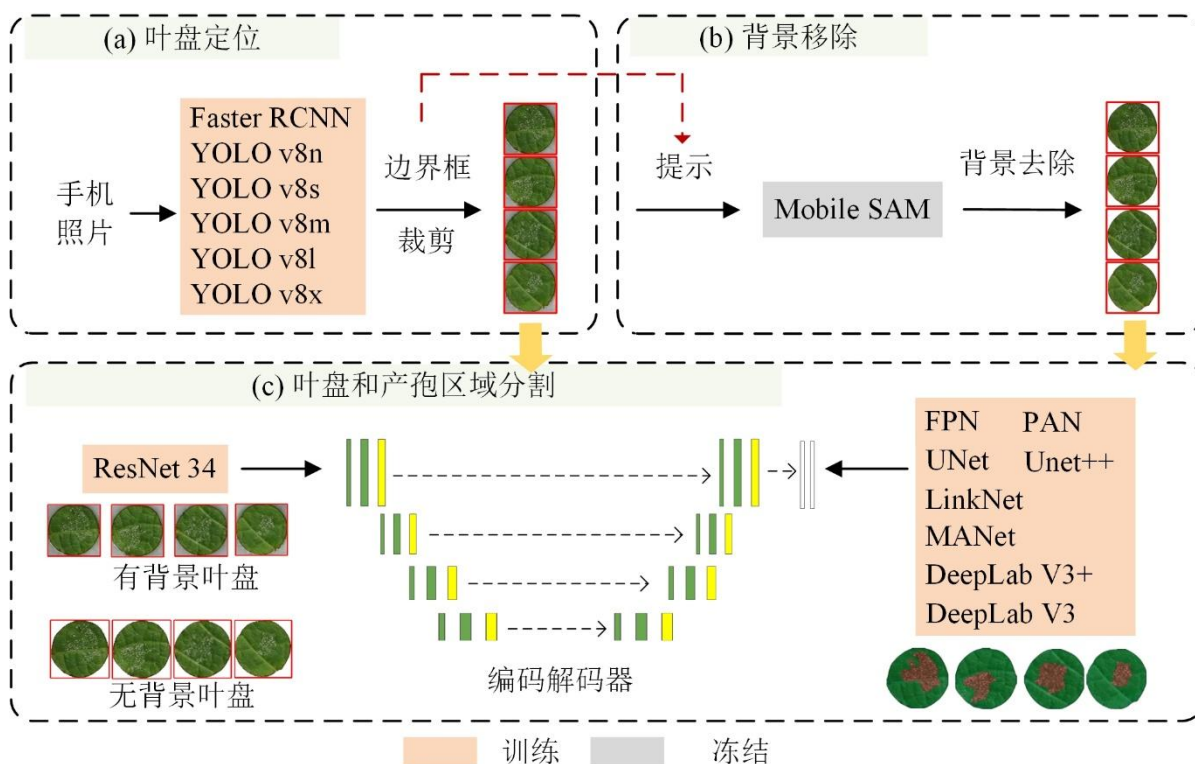


图 1-2 葡萄叶盘霜霉病严重度自动评估的完整工作流程。(a)用于在智能手机图像中定位叶盘的模型。(b)使用 Mobile SAM 进行背景移除，并提供边界框作为输入提示。(c)用于分割叶盘和霜霉病孢子囊的模型。

Figure 1-2 The complete workflow for automated severity assessment of grapevine downy mildew on leaf discs. (a) Models employed for leaf discs localization in smartphone images, (b) Background removal using Mobile SAM, with bounding boxes provided as input prompts, (c) Models utilized for segmenting leaf discs and downy mildew sporulation.

1.9.2 试验平台和配置

实验所用硬件包括英特尔酷睿 i7-13700KF CPU、128GB 内存以及配备 CUDA 12.2 驱动和 24GB 显存的 NVIDIA GeForce GTX 4090 GPU。YOLO 系列的训练

和验证使用了 Ultralytics 提供的脚本，而 Faster R-CNN 是基于 TorchVision 实现的。Mask R-CNN 的实现代码来源于 <https://github.com/WZMIAOMIAO/deep-learning-for-image-processing>。

1.10 模型开发中的对比实验设计

1.10.1 比较 YOLO v8 系列模型与其他目标检测模型定位叶盘的表现

为了准确高效地定位不同感染状态下的叶盘，进行了 YOLO v8 系列与 Faster R-CNN 的对比评估。这两个模型在农业物体检测任务中广泛应用，并且已证明具有较高的精度（Di et al. 2023; Tu et al. 2020）。YOLO v8 系列作为一种单阶段模型，旨在有效平衡推理时间和准确性。与之对比，Faster R-CNN 是一种两阶段模型，在更复杂的条件下通常提供更高的精度，但具有更大的参数量和较慢的推理速度（Ren et al. 2016）。

为了训练和评估这些模型，432 张图像被随机分为训练集（260 张）、验证集（86 张）和测试集（86 张），按照 6:2:2 的比例进行分配。YOLO v8 系列和 Faster R-CNN 分别使用 CSPNet 和 ResNet50 作为特征提取的骨干网络，促进了模型的开发。所有其他参数在模型之间保持一致。具体来说，初始权重来自于 COCO 数据集的预训练（Lin et al. 2015）。优化器使用了随机梯度下降（SGD），初始学习率为 0.001，权重衰减为 0.0005。训练过程持续 150 个周期，并在 10 个周期内无改进时实施早期停止。批处理大小为 16。对于验证集上表现最佳的模型权重，进行了保留以便进一步评估。

为了评估叶盘的定位效果，采用了 0.5 阈值下的平均精度均值（mAP@50）作为评估指标。mAP@50 的计算步骤如下：（1）根据每个类别在 0.5 的交并比（IoU）阈值下生成精度-召回曲线（Precision-Recall curve）。（2）计算每个类别的平均精度（AP），即在插值精度-召回曲线下的面积。（3）对所有类别的 AP 值进行平均，以获得 mAP@50。由于本研究只考虑了一个类别（叶盘），因此 mAP@50 等价于叶盘的 AP。mAP@50 的计算如公式所示：

$$\text{mAP@50} = \frac{1}{N} \sum_{i=1}^N \text{AP}_i \quad (1-15)$$

式中， N 表示类别总数， AP_i 是第 i 个类别的平均精度。mAP@50 的取值范围为 0% 到 100%，数值越高表明模型在目标检测中的性能越好。此外，还使用每秒帧数（FPS）来评估推理时间和计算效率。

1.10.2 比较 UNet 模型与其他图像分割模型分割叶盘和产孢区域的的表现

为了选择最合适的模型，我们使用了 8 个最先进的模型来进行叶盘和孢子形成的分割。这些模型包括 UNet、UNet++、DeepLabV3、DeepLabV3+、FPN、PAN、MA Net 和 LinkNet。这些模型采用了编码器-解码器架构，已被证明在准确识别叶盘上的孢子形成方面非常有效（Lin et al. 2019）。这些模型的关键特征及其对应的参考文献总结在表 1-1。

表 1-1 于叶盘和霜霉菌孢子形成分割的八个分割模型概述，
突出显示其关键特征和对应参考文献

Table 1-1 Overview of the eight segmentation models employed for leaf disc and downy mildew sporulation segmentation, highlighting their key features and corresponding references

模型	关键特征	参考文献
UNet	基础的编码器-解码器架构，将编码器中的特征图与解码器中相应深度的特征图相结合。	(Ronneberger, Fischer, & Brox, 2015)
UNet++	增强版 UNet，通过密集连接的路径促进不同深度之间的特征传播。	(Zhou, Siddiquee, Tajbakhsh, & Liang, 2020)
DeepLabV3	使用扩张卷积来扩大感受野，并采用 Atrous 空间金字塔池化（ASPP）进行多尺度特征融合。	(Chen, Papandreou, Schroff, & Adam, 2017)
DeepLabV3+	在 ASPP 模块中引入 Xception 和深度可分离卷积。	(Chen, Zhu, Papandreou, Schroff, & Adam, 2018)
FPN	构建多尺度特征图的特征金字塔网络。	(T.-Y. Lin et al., 2017)
PAN	在不同尺度的特征图融合前引入注意力模块，扩展了 FPN。	(H. Li, Xiong, An, & Wang, 2018)
MANet	引入多核注意力模块以优化特征图。	(R. Li et al., 2022)
LinkNet	一种轻量级分割模型，优化了效率，同时保持了良好的性能。	(Chaurasia & Culurciello, 2017)

对于这两个数据集，叶盘图像及其对应的掩码按照 6:2:2 的比例随机划分为训练集（1,595 张图像）、验证集（531 张图像）和测试集（532 张图像）。所有八个分割模型都采用了相同的超参数设置。具体来说，使用 ResNet-34 作为特征提取的骨干网络，并使用在 ImageNet 数据集上预训练的初始权重（Deng et al. 2009; He et al. 2015）。图像编码器的模型深度设置为 5。模型训练的迭代次数设置为 300 次，在验证集上连续 10 次迭代没有改进时提前停止。使用了 64 的批量大小。采用 Adam 优化器，初始学习率为 0.0001，并选择 Dice 系数作为损失函数。保留了在验证集上表现最佳的权重。

在阈值为 0.5 的情况下，使用平均交并比（mIoU@50）来评估这些分割模型的性能。交并比（IoU）被定义为手动标注掩码与模型预测掩码之间重叠区域与联合区域的比率。具体来说，重叠区域是真正例像素，联合区域是真正例像素（TP）、假正例像素（FP）和假负例像素（FN）的总和。因此，mIoU@50（公式 3-2）表示所有类别

上的平均 IoU。

$$mIoU@50 = \frac{1}{N} \sum_{i=1}^N \frac{TP_i}{TP_i + FP_i + FN_i} \quad (3-1)$$

式中, N 表示类别总数, TP_i 是第 i 类别中正确预测的真正例像素数, FP_i 表示第 i 类别中错误预测为正例的像素数, FN_i 表示第 i 类别中未预测为正例的像素数。mIoU@50 的取值范围为 0% 到 100%, 数值越高表示模型在分割任务中的性能越好。同样, 还使用每秒帧数 (FPS) 来评估推理时间和计算效率。

1.10.3 比较使用 Mobile SAM 去除背景后分割的对不同部位分类的影响

准确的背景移除对于提高后续分析的精度至关重要 (Xu et al. 2024)。为了将叶盘从背景中分离, 本研究直接使用了 Mobile SAM, 无需额外训练。Mobile SAM 是一个轻量级的视觉基础模型, 经过优化以适应移动设备, 并源自于 Segmentation Anything Model (Kirillov et al. 2023b; Zhang et al. 2023b)。它能够将输入的提示信息, 如点、边界框和掩膜, 编码为图像嵌入, 从而在各种任务中具备强大的零-shot 泛化能力, 进行目标分割。

在本研究中, YOLO v8n 生成的边界框坐标直接作为 Mobile SAM 的提示输入。这使得能够精确地生成叶盘的轮廓 (附图 2)。随后, 利用这些轮廓通过位运算去除背景, 位运算是通过 OpenCV 4.2 (Bradski 2000) 实现的。最终, 背景被清除后的每个叶盘有效地从输入图像中分离出来。

为了评估背景去除的影响, 这些模型在两个不同的叶盘图像数据集上进行了训练: 一个包含干净背景的图像, 另一个包含原始背景的图像。未标注孢子形成的图像被排除在外, 以确保模型评估的准确性。因此, 每个数据集都包含了 2,658 张叶盘图像。

1.10.4 比较 SporeScan 与 Mask RCNN 对叶盘霜霉病检测分割的表现

如前所述, Mask R-CNN 被广泛认为是最有效的实例分割模型之一, 并已展示出在自动化病害严重度评估中的端到端潜力 (He et al. 2018; Kumar et al. 2023)。因此, SporeScan 的性能进一步与 Mask R-CNN 进行了比较, 评估了叶盘定位以及叶盘和霜霉病孢子囊的分割效果。

Mask R-CNN 使用原始的 432 张图像进行训练, 模型从在 COCO 数据集上预训练的权重初始化 (Lin et al. 2015)。优化器采用 SGD, 初始学习率为 0.001, 动量为 0.9。由于内存限制, 批处理大小设置为 2。在评估过程中, 叶盘定位使用 mAP@50 指标, 分割性能则使用 mIoU@50。

需要注意的是, 由于某些霜霉病孢子囊样本的轮廓未被标注, Mask R-CNN 的

分割性能可能低于 SporaScan。然而，准确的叶盘定位是有效分割的前提，因此，这里比较的主要关注点是定位性能。

1.11 叶盘霜霉病严重程度的评估

为了评估 SporaScan 的效率，基于叶盘和孢子形成的预测掩码计算了严重程度。严重程度被定义为叶盘被孢子覆盖的面积比例，是评估葡萄品种抗性的一个关键指标（Hernández et al. 2021; Zendler et al. 2021）。它可以使用以下公式（公式（3））进行计算：

$$\text{严重程度} = \frac{\text{霜霉病产孢区域面积}}{\text{叶盘面积}} \times 100\% \quad (3-2)$$

对于每张图像，有两个掩码：一个用于叶盘，另一个用于孢子形成。首先，这两个掩码通过 0.5 的阈值转换为二值掩码，其中像素被分类为 1 表示正确分类，被分类为 0 表示错误分类。孢子形成的面积是通过正确分类为孢子形成的像素之和来确定的，而叶盘的面积则是通过正确分类为叶盘和孢子形成的像素之和来确定的。手动标注的严重程度是直接基于标注的叶盘和孢子形成的轮廓面积计算的。

为了评估手动标注和模型预测之间的一致性，我们使用了决定系数（ R^2 ）。 R^2 的公式如下：

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (3-3)$$

式中， y_i 表示每个叶圆盘上霜霉病孢子繁殖的手动标注区域， $\hat{y}_i$ 表示预每个叶圆盘上霜霉病孢子繁殖的预测严重程度， $\bar{y}$ 表示所有叶圆盘上霜霉病手动标注的平均严重程度， n 表示样本数量。我们仅计算了测试集中每个叶圆盘的严重程度。

1.12 模型预测的产孢分布与人工标注的双盲实验

由于像素级别的手动标注工作量大且耗时，因此在生成的掩码中常常会出现一些小的不准确之处（附图 3）。因此，在实际应用中，手动标注的掩码并不总是比 SporaScan 的预测质量更高。为了评估 SporaScan 相对于手动标注的实际性能，我们开展了一项双盲实验，以捕捉葡萄育种技术人员的偏好。

我们开发了一个在线平台，用于展示每个样本的三种图像：（1）原始叶盘图像，（2）以 50% 透明度叠加在该图像上的手动标注掩码，以及（3）以 50% 透明度叠加在该图像上的 SporaScan 预测掩码。后两种图像以随机顺序呈现。在没有任何事先提示

或指导的情况下, 11 名技术人员分别从数据集中随机分配到 100 张图像, 并被要求选择最准确的标注。提供了三种选项: (1) 手动标注更优, (2) 模型预测更优, 或 (3) 两种标注同样准确。选择选项 (2) 和 (3) 的结果被解释为表明 SporaScan 的性能优于或与手动标注相当。

在收集技术人员的反馈后, 我们对每种选项的选择比例进行了统计分析。我们还进行了 T 检验, 以评估手动标注与模型预测之间的差异, 并将显著性阈值设定为 $p < 0.05$ 。这种统计评估为比较模型预测与手动标注在实际应用中的可接受性提供了宝贵的见解。

1.13 结果和分析

1.13.1 YOLO v8 系列模型在叶盘定位的结果

表 1-2 总结了 YOLO v8 系列和 Faster R-CNN 在叶盘定位任务中的性能表现。所有模型在验证集和测试集上均达到了超过 99% 的 $mAP@50$ (平均精度均值), 显示出卓越的准确性和强大的泛化能力。在 YOLO v8 系列中, 参数量最小的 YOLO v8n (3.0M) 与参数量最大的 YOLO v8x (68.2M) 实现了相当的 $mAP@50$ 。此外, YOLO v8n 展现出最高的推理速度, 达到 212.8 FPS, 而 YOLO v8l 的推理速度最低 (109.8 FPS)。尽管 Faster R-CNN 在 $mAP@50$ 上略优于 YOLO v8 系列, 但由于其较大的参数量和计算复杂性, 其推理速度显著较低。图 1-3 展示了模型在不同分辨率和条件下对叶盘的精确定位能力, 证实了其能够准确检测健康和感染的葡萄叶盘。

表 1-2 葡萄叶盘定位任务中各模型的性能表现。 $mAP@50$ 表示在阈值为 0.5 时, 每个类别的平均精度均值。Params 表示模型参数的大小 (单位为百万)。FPS 指每秒处理的帧数。

Table 1-2 Models performance for grapevine leaf discs localization. $mAP@50$ represents the mean of average precision for each class at a threshold of 0.5. Params denotes the size of models' parameters in million. FPS refers to the number of frames processed per second.

目标检测模型		$mAP@50^{val}$	$mAP@50^{test}$	Params(M)	FPS
Faster R-CNN		99.80%	99.90%	41.3	19
YOLO v8 系列	YOLO v8n	99.40%	99.50%	3	213
	YOLO v8s	99.40%	99.40%	11.1	196
	YOLO v8m	99.40%	99.50%	25.8	185
	YOLO v8l	99.40%	99.50%	43.6	110
	YOLO v8x	99.40%	99.50%	68.2	120

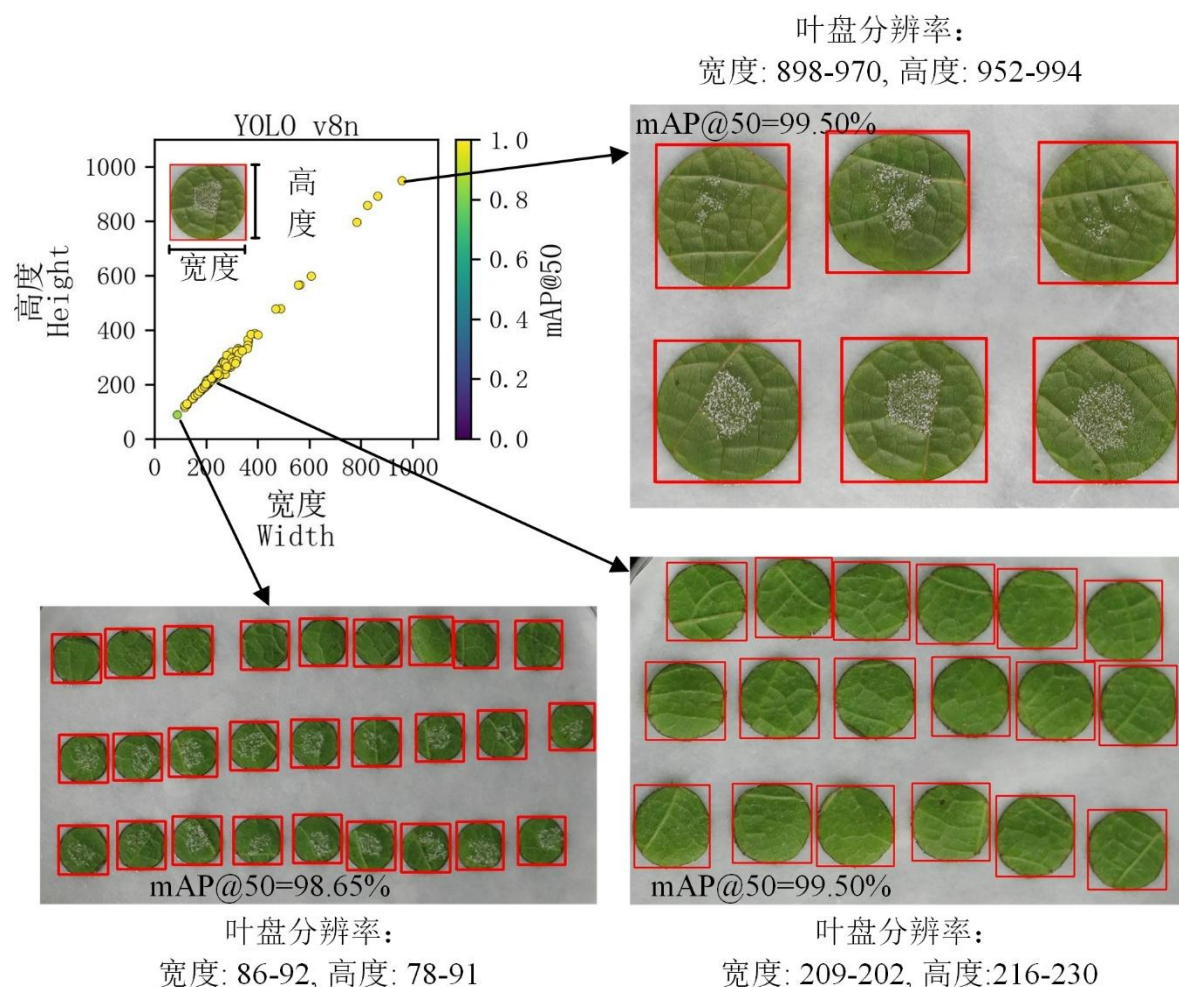


图 1-3 YOLO v8n 在不同分辨率下对葡萄藤叶盘的预测定位。红色框表示葡萄藤叶盘的预测边界框。“Width”表示水平像素数，“Height”表示垂直像素数。背景为培养皿。

Figure 1-3 Predicted localization of grapevine leaf discs at different resolutions by the YOLO v8n. Red boxes represent the predicted bounding boxes of grapevine leaf discs. “Width” is the horizontal pixel number, and “Height” is the vertical pixel number. The background is a Petri dish.

1.13.2 分割模型性能：UNet 和 Mobile SAM

表 1-3 总结了八个分割模型在葡萄藤叶盘和霜霉菌孢子形成分割任务中的性能表现。在去除背景后，所有分割模型在验证集和测试集上均显示出一致的性能提升（+0.4%），达到了超过 96% 的高 mIoU@50（平均交并比），并表现出可靠的泛化能力。验证集上的平均 mIoU@50 达到 96.2%，测试集上为 96.4%。这些模型之间的性能差异很小。UNet 略微优于其他模型，在验证集（96.44%）和测试集（96.53%）上均实现了最高的 mIoU@50。大多数模型的数量在 21.47M 到 26.08M 之间，性能表现非常接近。在推理时间方面，FPN 实现了最高的 FPS（289），而 DeepLab V3 的 FPS 最低，为 234。尽管 UNet++ 的 mIoU@50 与 UNet 相似，但其 FPS 显著较低。UNet 在高 mIoU@50 的同时，还保持了 276 的竞争力 FPS。

表 1-3 8 个分割模型在葡萄藤叶盘和霜霉菌孢子形成分割任务中的性能表现。mIoU@50 表示在阈值为 0.5 时，每个类别的平均交并比（Intersection over Union, IoU）。Params 表示模型参数的大小（单位为百万）。FPS 指每秒处理的帧数。

Table 1-3 The performance of the eight segmentation models in segmenting grapevine leaf discs and downy mildew sporulation. mIoU@50 represents the mean of IoU (Intersection over Union) for each class at a threshold of 0.5. Params denotes the size of models' parameters in million. FPS refers to the number of frames processed per second.

模型	背景未移除		移除背景		参数量	FPS
	mIoU@50 ^{val}	mIoU@50 ^{test}	mIoU@50 ^{val}	mIoU@50 ^{test}		
DeepLab V3	95.84%	96.02%	96.22%	96.37%	26	234
DeepLab V3+	95.68%	95.79%	96.08%	96.25%	22.44	280
FPN	95.84%	95.95%	96.25%	96.44%	23.15	289
LinkNet	95.77%	95.93%	96.17%	96.31%	21.77	274
MANet	95.83%	95.94%	96.33%	96.53%	31.78	237
PAN	95.68%	95.82%	96.21%	96.38%	21.47	262
UNet++	95.96%	96.10%	96.33%	96.48%	26.08	236
UNet	95.82%	96.07%	96.42%	96.54%	24.44	276

图 1-4 展示了 Mobile SAM 在背景去除与否的情况下，对背景、叶盘和孢子形成三个类别分类性能的混淆矩阵。矩阵中的每个单元格的值按行归一化，表示类别之间的误分类率。背景去除后，各分类之间的误分类率总体下降，但叶盘与孢子形成之间的误分类率略有增加。叶盘与背景之间的误分类率降低了 2.77%。孢子形成与背景之间的误分类率降为零。然而，孢子形成与叶盘之间的误分类率增加了约 1%。其他类别之间的误分类率差异较小。

图 1-5 展示了八种分割模型生成的叶盘和霜霉菌孢子形成的分割实例。这些示例突出了这些模型的优势。前两行强调了模型预测与手动标注在不同轮廓上的高度一致性。第 3 至第 5 行展示了模型对大面积和多个小面积霜霉菌孢子形成的分割能力。最后一行则展示了对健康叶盘的准确分割。在所有实例中，这些模型均表现出一致的分割准确性，仅在叶盘和霜霉菌孢子形成的分割与手动标注的对比中观察到细微的视觉差异。

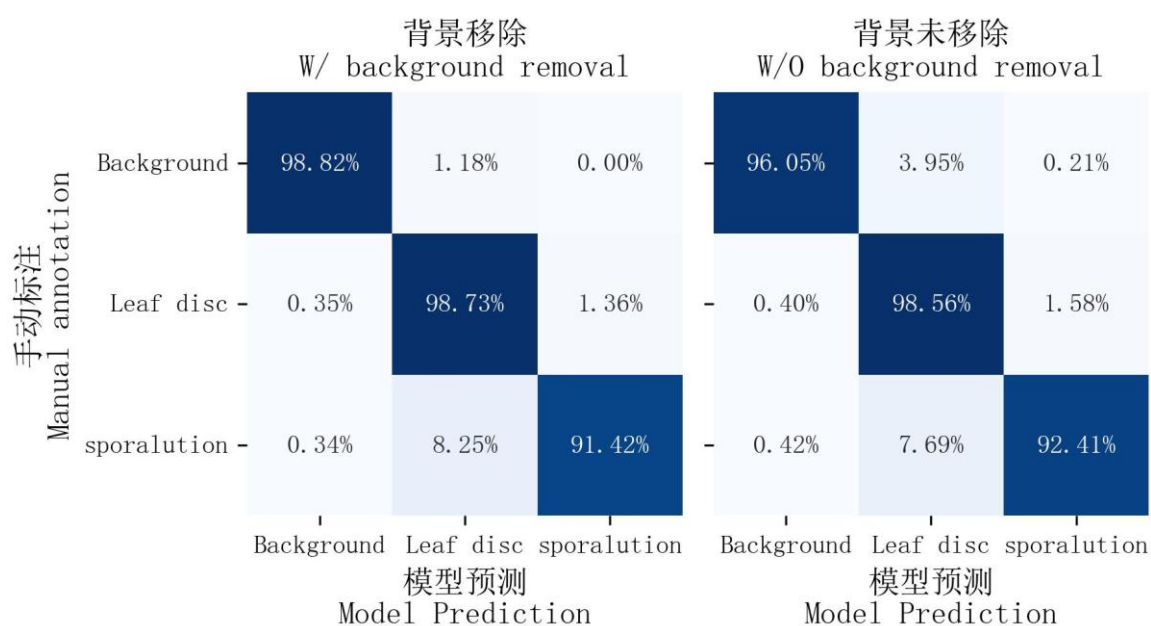


图 1-4 展示了 Mobile SAM 在背景去除与否的情况下，对背景、叶盘和孢子形成三个类别分类性能的混淆矩阵。矩阵中的每个单元格的值按行归一化，表示类别之间的误分类率。背景去除后，各分类之间的误分类率总体下降，但叶盘与孢子形成之间的误分类率略有增加。

Figure 1-4 The confusion matrix of the classification performance among the three categories: background, leaf disc and sporulation with and w/o background removal by Mobile SAM. The values in each cell are normalized by row, representing the inter-class misclassification rates.

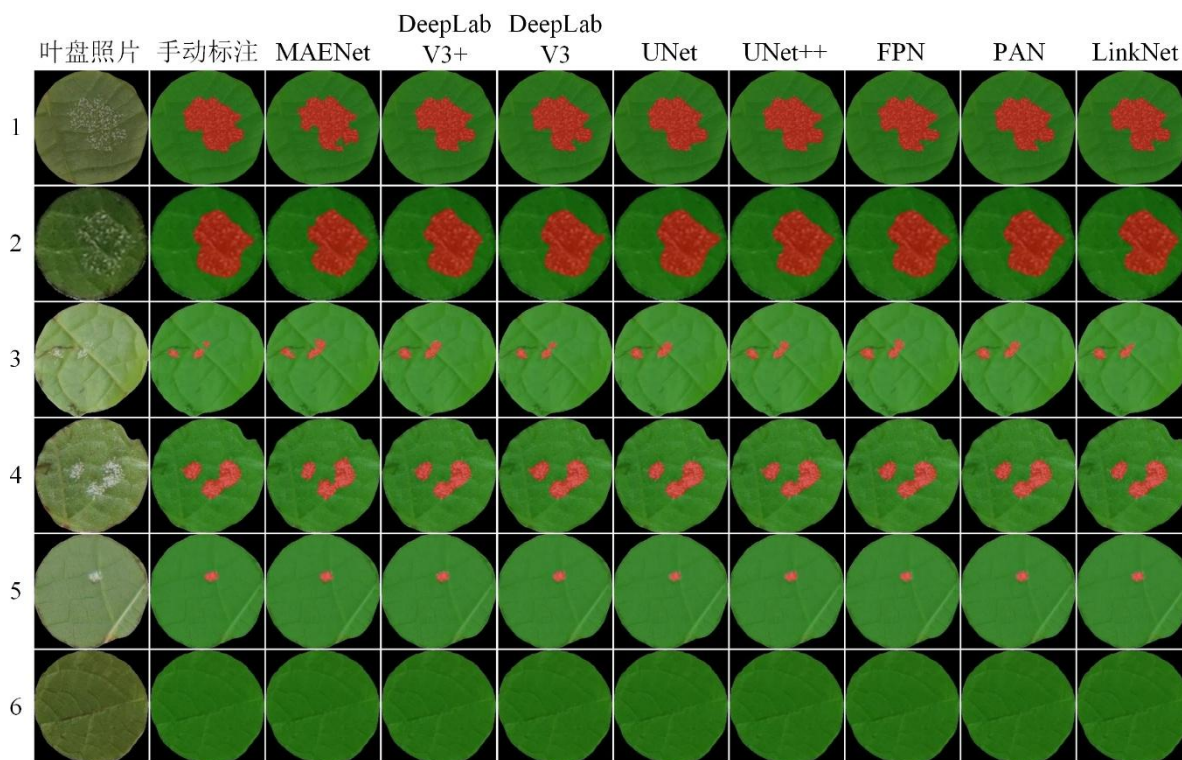


图 1-5 葡萄叶盘上霜霉菌孢子形成的代表性分割示例。第一列显示原始图像，而第二列展示了手

动标注的掩码。其他列则展示了由八种分割模型生成的预测掩码。

Figure 1-5 Representative examples of downy mildew sporulation segmentation on leaf discs. The first column displays the original image, while the second column shows the masks from manual annotation.

The other columns present the predicted masks generated by the eight segmentation models.

1.13.3 SporaScan 与 Mask R-CNN 的性能对比

表 1-4 对 Mask R-CNN 和 SporaScan 在叶盘定位以及叶盘和霜霉菌孢子形成分割任务中的性能进行了对比分析。SporaScan 在叶盘定位任务中实现了更高的 $mAP@50$ (平均精度均值), 在分割任务中也展现了更高的 $mIoU@50$ (平均交并比)。具体而言, 在验证集和测试集上, SporaScan 在叶盘定位任务中分别比 Mask R-CNN 高出 4.51%和 5.90%。在叶盘和孢子形成的分割任务中, SporaScan 在验证集和测试集上的性能提升分别为 8.04%和 9.27%。这表明 SporaScan 在性能和泛化能力上均优于 Mask R-CNN。

表 1-4 Mask R-CNN 和 SporaScan 的性能表现。 $mAP@50$ 表示在阈值为 0.5 时的平均精度均值; $mIoU@50$ 表示在阈值为 0.5 时的平均交并比 (Intersection over Union)

Table 1-4 The performance of Mask R-CNN and the SporaScan. $mAP@50$ represents the mean average precision at a threshold of 0.5. $mIoU@50$ represents the mean of IoU (Intersection over Union) at a threshold of 0.5.

模型	叶盘定位		叶盘和产孢区域分割	
	$mAP@50^{val}$	$mAP@50^{test}$	$mIoU@50^{val}$	$mIoU@50^{test}$
Mask R-CNN	94.89%	94.60%	88.36%	87.23%
SporaScan	99.40%	99.50%	96.40%	96.50%

图 1-6 展示了 Mask R-CNN 和 SporaScan 的模型预测代表性实例。Mask R-CNN 的主要失败案例包括叶盘定位不完整以及孢子形成分割的冗余。值得注意的是, 未定位的叶盘部分大多位于图像的左侧。相比之下, SporaScan 采用顺序处理方法, 能够准确地定位叶盘, 并有效避免孢子形成掩码的冗余, 显示出其在分割任务中的稳健性和精确性。

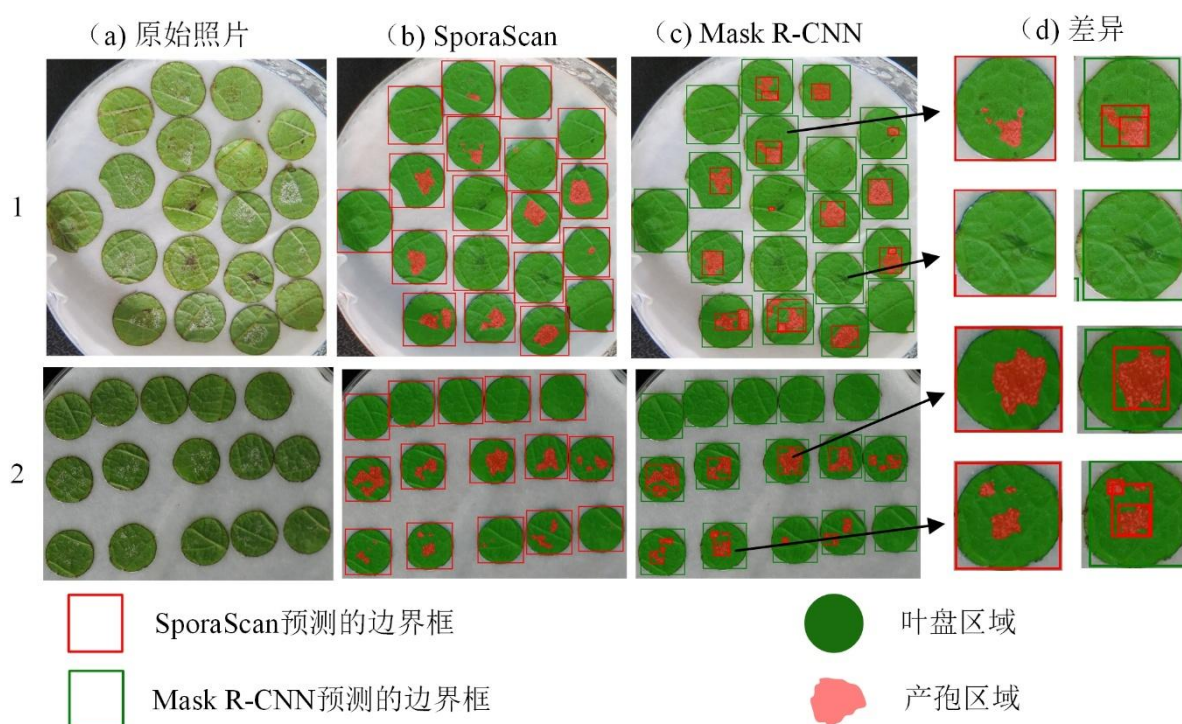


图 1-6 葡萄叶盘定位和叶盘及孢子形成分割的实例。(a) 原始图像, (b) SporaScan 预测的边界框和掩码, (c) Mask R-CNN 预测的边界框和掩码, (d) 两种模型之间的差异。

Figure 1-6 The instances of grapevine leaf disc localization and the segmentation of both leaf discs and sporulation. (a) original images, (b) predicted bounding boxes and masks by SporaScan, (c) predicted bounding boxes and masks by Mask R-CNN, (d) differences between the two models.

1.13.4 SporaScan 模型在病害严重程度预测中的性能表现

图 1-7 显示了 SporaScan 对霜霉病严重程度的手动标注与模型预测之间的 R^2 值, 以及图像中每个叶片圆片的预测实例。在测试集中, 大多数严重程度值集中在 0% 到 40% 之间, 40% 到 60% 之间的样本较少。SporaScan 与手动标注表现出高度一致性, R^2 达到 0.992。该模型的这种一致性凸显了其在评估霜霉病等病害严重程度时表现出的稳健且可靠的性能, 仅存在轻微差异。

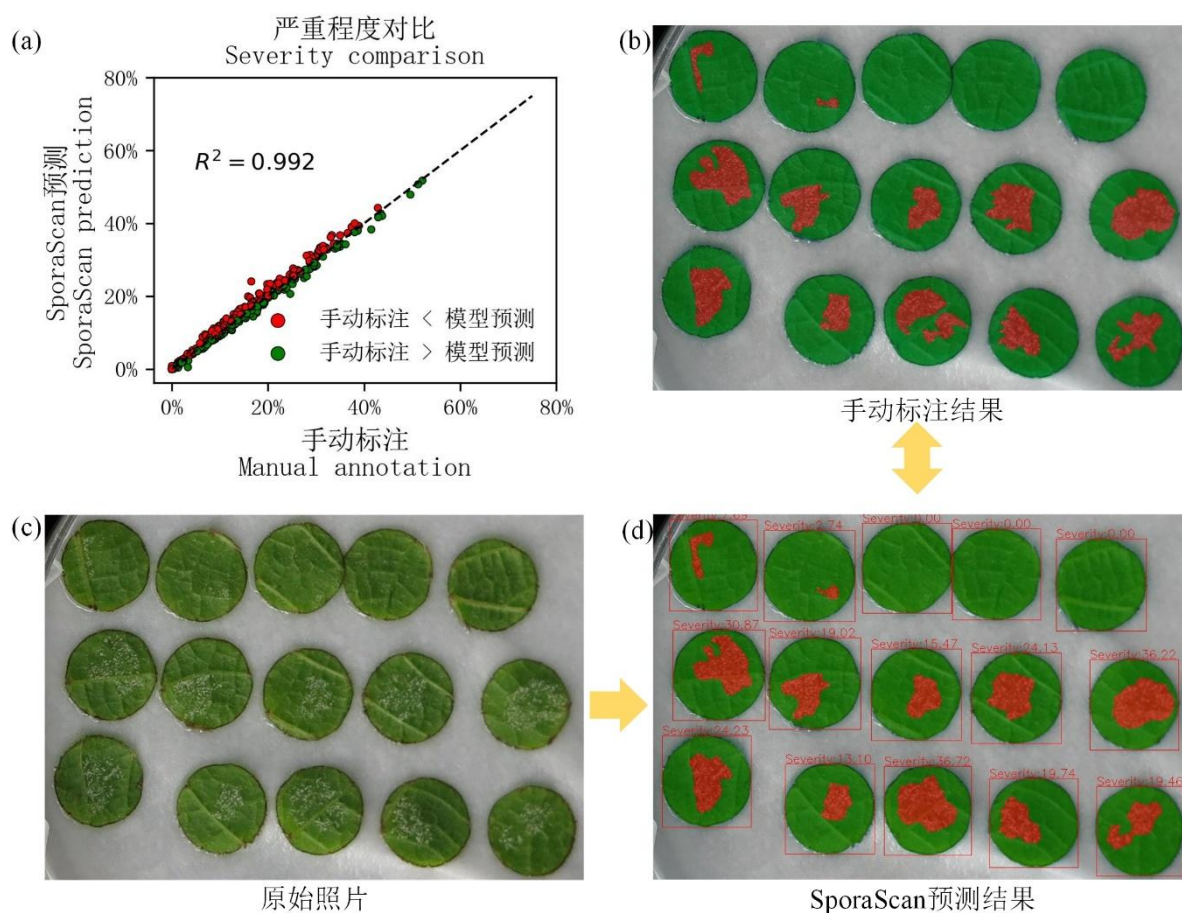


图 1-7 霜霉病严重程度的手动标注与 SporaScan 预测之间的一致性 (R^2 值) 及预测实例。(a) 手动标注与 SporaScan 预测的严重程度对比, (b) 手动标注的图像, (c) 原始图像, (d) SporaScan 对每个叶盘的预测掩码和严重程度。

Figure 1-7 The coefficient of determination (R^2) of downy mildew severity between manual annotation and SporaScan, alongside a predicted instance. (a) severity comparison between manual annotation and SporaScan, (b) manual annotated image, (c) original image, (d) predicted masks and severity by SporaScan for each leaf disc.

1.13.5 手动标注与模型生成的偏好双盲试验结果

图 1-8 展示了技术人员在 SporaScan 预测与手动标注之间偏好的双盲试验结果。手动标注的中位选择比例约为 20%, 而 SporaScan 预测则获得了大约 80% 的偏好选择。在两种选项之间观察到了统计学上的显著差异 ($p < 0.05$), 这表明技术人员对 SporaScan 生成的预测结果有强烈的偏好。

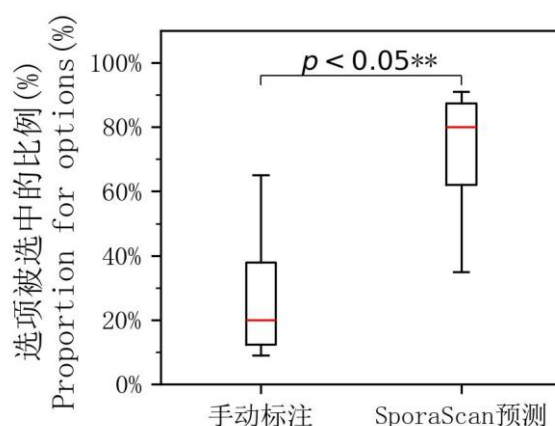


图 1-8 展示了技术人员在 SporaScan 预测与手动标注之间偏好的双盲实验结果。手动标注的中位选择比例约为 20%，而 SporaScan 预测则获得了大约 80% 的偏好选择。在两种选项之间观察到了统计学上的显著差异 ($p < 0.05$)，这表明技术人员对 SporaScan 生成的预测结果有强烈的偏好。

Figure 1-8 The results of T-test between SporaScan predictions and manual annotations based on technicians' preferences. Manual annotation refers to cases where technicians rated manual annotation as superior to SporaScan prediction, while "SporaScan" indicates cases where SporaScan prediction was preferred over manual annotation. The red line represents the median, with whiskers denoting the first and third quartiles. The symbol ** denotes the significance level as determined by T-test.

1.14 讨论

评估新开发的杀菌剂和品种对于病害抗性至关重要，而病害抗性通常通过病害严重程度来量化。传统的叶盘实验中病害严重程度的评估方法劳动强度大、耗时长，并且由于技术人员的专业水平差异和主观解读，容易出现变异性和偏差。在本研究中，我们引入了一种基于深度学习的流程—SporaScan，它整合了 YOLOv8n、Mobile SAM 和 UNet，以自动化执行关键任务，包括叶盘检测、背景去除和孢子形成分割。通过利用智能手机拍摄的低成本数据集，SporaScan 在准确性方面优于 Mask R-CNN 等端到端模型，并且在设计的双盲试验中甚至超过了手动评估的表现。SporaScan 为培育抗病作物品种和杀菌剂效果评估提供了一种实用且可扩展的工具，有助于推动植物病害管理实践的进步。

1.14.1 序列化 SporaScan 优于端到端的 Mask R-CNN

Mask R-CNN 是一种广泛采用的端到端实例分割模型，理论上可以通过同时检测和分割叶片及病害区域来计算病害严重程度 (Afzaal et al. 2021; Kumar et al. 2023)。尽管这种方法提供了一种吸引人的端到端解决方案，但本研究的实验结果表明，Mask R-CNN 的性能不如 SporaScan 所实现的顺序处理流程。这种性能差距可以归因于几个技术因素。

首先, Mask R-CNN 依赖于区域提议网络 (Region Proposal Network, RPN) 进行目标检测。虽然叶盘与背景之间的颜色对比有利于检测, 但复杂的背景以及叶盘与孢子形成之间的微妙边界常常会干扰区域提议的准确性, 导致叶盘的定位不够精确。相比之下, SporaScan 采用的 YOLOv8n 专门针对单目标检测进行了训练, 通过稳健的定位能力实现了超过 99% 的 $mAP@50$, 从而最大限度地减少了误分类。

其次, 端到端框架中任务的集成化导致了任务专业化的固有挑战。Mask R-CNN 需要同时执行检测、背景区分和细粒度分割, 这使得模型无法针对单一任务进行优化 (Yang et al. 2024)。特别是对于离散但位置接近的孢子形成区域, Mask R-CNN 可能会将它们检测为一个较大的区域和多个较小的区域, 从而导致孢子形成的重复分割。相反, SporaScan 通过解耦工作流程来规避这一限制: 它使用 Mobile SAM 进行有效的背景去除, 并利用 UNet 进行精确的分割。这种模块化方法使每个组件都能专注于其特定功能, 从而实现了超过 96% 的分割 $mIoU@50$, 并减少了通常因单一模型处理多个任务而产生的不确定性。

此外, Mask R-CNN 架构的复杂性、较大的参数量以及增加的计算开销导致其推理速度较慢, 这对于高通量表型分析应用来说是一个显著的缺点。而 SporaScan 则利用轻量级但功能强大的模型, 不仅提供了高准确性, 还实现了更优的推理速度, 使其更适合实时应用。

综上所述, 尽管 Mask R-CNN 提供了一种端到端的病害严重程度评估方案, 但其在定位精度、任务专业化和计算效率方面的局限性限制了其整体性能。SporaScan 的顺序处理方法通过分离和优化每个独立任务, 展示了更高效且可扩展的解决方案, 能够更准确地评估葡萄霜霉病的严重程度。

1.14.2 SporaScan 在自动化葡萄霜霉病严重程度评估中的优越性

准确且自动化的病害严重度量对于抗性品种的培育至关重要, 需要一个高效且稳健的模型以最小化手动标注引入的偏差。在本研究中, SporaScan 的严重程度评估与手动标注表现出极好的一致性, 并且在技术人员偏好的双盲试验中甚至超过了手动标注。这一表现表明, SporaScan 不仅能够以高精度复制手动评估, 还克服了手动标注中固有的不一致性和主观偏差。

SporaScan 的一个关键优势在于其能够执行精确的像素级分割——这是技术人员通过手动标注无法始终如一地实现的精度水平。手动方法通常会导致对病害区域的近似估计, 而像 UNet 这样的模型则是为了以高保真度描绘这些区域而设计的。因此, SporaScan 能够有效地从结构化的输入数据中分离并量化病害症状, 尤其是在背景干扰最小的受控条件下。最近的研究表明, 基于 CNN 的模型在使用高质量显微图像评估病害严重程度时可以达到与手动标注相当的精度 (Lin et al. 2019; Qiu et al. 202

2)。

从实际角度来看, 这些发现突出了基于 CNN 的模型在病害严重程度评估中的优势和局限性。技术人员更倾向于选择模型预测而非手动标注, 这可能是由于基于 CNN 的模型的低错误率和完全自动化, 即使在使用智能手机拍摄的低成本图像上应用时也是如此。观察到的局限性可能归因于小而离散的孢子形成引入的噪声, 以及叶片表面的光照。未来的研究可能会专注于通过解决复杂背景、重叠叶片和多样化植物生长阶段等挑战, 进一步推进病害严重程度的评估 (Barbedo 2018b)。

1.14.3 本章节研究的局限性

本研究存在一些潜在的改进空间: (1) 依赖现有的模型, 而不是开发专门为该任务定制模型; (2) 在叶盘实验中, 离散霜霉菌孢子照片的数量有限。然而, 现有的卷积神经网络 (CNN) 已在各种下游应用中表现出色, 包括农业任务, 通常超过了传统方法。本研究的发现表明, 这些模型已经非常熟练, 这从它们在不同条件下的稳定表现中得到了证明。此外, 最近的研究强调了整合多种架构的精简模型的优势, 这些模型相比单一模型能够产生更优越的结果 (Xu et al. 2024; Yang et al. 2024)。

对于表现出离散孢子分布的叶盘, 手动标注仍然特别具有挑战性, 这给标注的准确性以及模型训练都带来了复杂性。输入图像的质量可能影响了整体的分割性能, 这也凸显了进一步优化的途径。未来的研究应探讨增强离散霜霉菌孢子形成分割的策略, 包括实施先进的模型集成技术以及使用更高质量的数据集。

1.15 本章小结

本研究研发了 SporaScan, 这是一个自动化的流程, 整合了 YOLO v8n 用于葡萄藤叶盘的定位、Mobile SAM 用于背景去除以及 UNet 用于霜霉病分割。使用 432 张智能手机拍摄的图像, 覆盖超过 6000 个叶盘, SporaScan 实现了超过 99% 的 mAP@50、高于 96% 的 mIoU@50 以及 0.99 的 R^2 值, 其性能与手动标注相当, 同时大大减少了人工劳动和主观性。

这些结果突出了深度学习在培育抗病作物品种、杀菌剂评估及相关应用中简化工作流程的潜力, 通过将复杂任务分解为专业化的模块来实现。未来的研究可能会致力于将这种方法扩展到多样化的条件和作物中, 增强对复杂病害模式的分割能力, 并纳入更高质量的数据集和田间试验, 以进一步提高模型的稳健性和泛化能力。

基于微调 Qwen-VL 的葡萄病害图像描述生成方法比较

在田间观测中，准确的识别病害类型，是病害管理的前提和关键步骤。错误的识别通常导致错误的药物喷洒，可能导致更多的药害，也提高了管理的成本。基于卷积神经网络的病害识别尽管已经显示出非常高的准确率，特别是在实验室条件下。然而这些算法可解释性依赖类激活图算法，如 Grad-CAM 等，仅仅指示出图像哪一部分影响分类结果，模型对每张病害图像的语义理解仍然不够清楚（Rahman et al. 2020; Yan et al. 2025）。对图像病害特征细节的描述缺失，可能导致应用场景下的信任不足。这表明，目前的模型需要充分的理解病害叶片的语义细节特征。最近的视觉语言模型，已经展示了将图像和文本结合的能力，可以生成对图像内容的详细且可解释的描述。在田间葡萄病害识别的背景下，引入视觉语言模型可以提供一个解决方案，以生成对病害特征的详细解释性描述。这种描述可以包括病害的位置、大小、颜色、形状等特征。通过这种方式，农民或病害管理专家可以更准确地理解病害状况，制定更有效的管理策略，同时减少因错误识别而导致的药害风险和管理成本图 1-1。因此本章节评估了微调视觉语言模型，如 Qwen-2.5-VL-3B，对不同病害特征描述生成的表型，并与其他 BLIP2、DeepSeek-2-VL、Kimi-2-VL 以及 Describe Anything Model (DAM)做了比较。

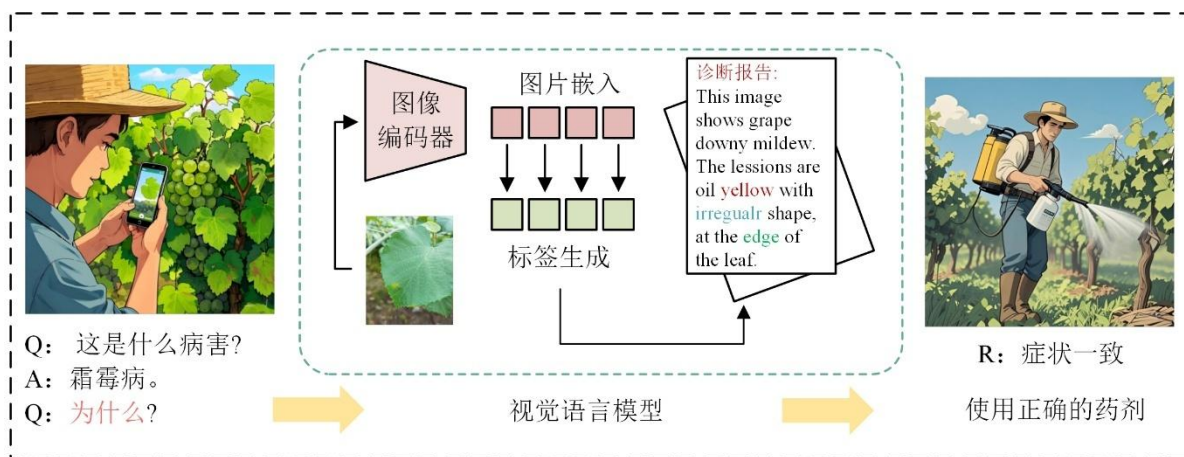


图 1-1 视觉语言模型在植物病害诊断中的应用。图中展示了农民通过手机拍摄受病害影响的葡萄叶，利用视觉语言模型进行图像编码和描述生成，最终得到病害诊断报告的过程。模型输出的报告帮助农民识别病害并采取适当的防治措施。

Figure 1-1 Application of visual language models in plant disease diagnosis. The diagram illustrates the process where a farmer takes a photo of disease-affected grape leaves with a smartphone, uses a visual language model to encode the image and generate a description, and ultimately receives a disease diagnosis report. The report generated by the model assists farmers in identifying the disease and taking appropriate control measures.

1.16 概述

1.16.1 Qwen-VL 模型介绍

为了定制化葡萄病害描述生成模型，我们使用 Qwen-2.5 VL-3B（VL, Vision Language）大模型作为基准模型（以下成 Qwen-VL）。Qwen-VL 是由阿里巴巴通义团队开源的一种多模态模型，结合了视觉 Transformer（ViT）和 Qwen-2.5 大语言模型（Bai et al. 2025）。它采用了一种支持动态分辨率的设计，能够无需图像归一化步骤，就可以直接处理不同尺寸的图像。该模型利用了窗口注意力机制、绝对时间编码和改进的位置嵌入方法，在多个视觉语言任务上，如文档理解、目标定位均取得了优异效果。整体性能在已与 GPT-4o 等相当。Qwen-VL 提供了 3B/7B/72B 三个不同参数量的模型，考虑到显存的限制以及推理速度影响，我们选择了 3B 的模型作为基准模型，用于定制化葡萄病害特征描述生成模型。

1.16.2 使用 LoRA 进行模型微调得到 FT-Qwen-VL

LoRA（Low-Rank Adaptation）是一种参数高效的微调方法，广泛应用于大规模预训练模型的下游微调任务中（Hu et al. 2021; Zhong et al. 2024）。其核心思想是通过在固定的预训练权重基础上，引入两个低秩矩阵以近似表示模型参数的更新，从而显著减少训练参数的数量。LoRA 的基本形式可以表示为：

$$\Delta W \approx BA \quad (4-1)$$

式中， ΔW 表示对原始权重矩阵 W 的增量调整， $A \in \mathbb{R}^{r \times d}$ ， $B \in \mathbb{R}^{d \times r}$ ， $r \ll d$ 示两个可学习的低秩矩阵，构成了一个秩为 r 的近似更新。在实际应用中，LoRA 将模型的线性层进行如下替换：

$$Wx \rightarrow (W + \alpha \cdot BA)x \quad (4-2)$$

式中， x 为输入向量， α 是一个缩放因子，用于控制 LoRA 引入的权重调整幅度。参数 α 的引入有助于平衡预训练权重与微调权重之间的影响，防止低秩矩阵更新对原始模型结构产生过大扰动。通过上述设计，LoRA 在保持模型性能的同时，可以大幅减少微调所需的参数数量，是实现高效迁移学习的一种有效手段。

1.16.3 模型评估指标

为了评估模型输出的病害特征描述质量，我们使用了 BLEU-4、ROUGE-1、ROUGE-2、ROUGE-L 四个指标。其中，BLEU-4（Bilingual Evaluation Understudy, 4-gram）是基于 4-gram 匹配的指标，它会将模型生成的句子和参考答案进行比较，主要关注四个连续词（4-gram）的匹配情况，用于衡量生成文本与参考文本之间的相似度（Papineni et al. 2001）。BLEU-4 主要强调短语层次匹配和语言流畅度。越多的 n-gram 匹配，得分越高表示生成内容越接近参考文本。BLEU-4 的计算公式下：

$$BLEU_4 = BP \times \exp \left(\sum_{n=1}^4 w_n \times \log(p_n) \right) \quad (4-3)$$

式中, P_n 为第 n -gram 的精确率, n -gram 为 n -gram 权重 (通常平均分配即 0.25), BP 为惩罚因子 (brevity penalty), 用于惩罚生成过短的文本。

ROUGE 用于评估文本生成与参考文本之间相似度的指标, 更侧重于召回率 (Ganesan 2018)。ROUGE-1 统计生成文本与参考文本中相同的“单个词”(1-gram) 的数量, 占参考文本中所有词的比例, 衡量关键词覆盖程度, 反映生成内容是否包含了参考答案的主要词汇。ROUGE-2 计算生成文本和参考文本之间相同的“连续两个词”(2-gram) 对的数量, 评估文本的连贯性和局部语言结构, 比 ROUGE-1 更严格, 关注短语的匹配。ROUGE-1 和 ROUGE-2 的计算公式分别如公式 4-2、方程 4-3 所示:

$$ROUGH-1 = \frac{\sum_{w \in \text{参考文本}} \text{匹配次数}(w)}{\sum_{w \in \text{参考文本}} \text{出现次数}(w)} \quad (4-4)$$

$$ROUGH-2 = \frac{\text{生成文本与参考文本相同的二元词组数}}{\text{参考文本中的所有二元词组总数}} \quad (4-5)$$

ROUGE-L 基于最长公共子序列 (LCS) 来评估生成文本与参考文本之间的重合度。该指标通过同时考虑精确率和召回率, 并使用调和平均 (F1-score) 来综合评估生成内容的覆盖度和准确度, 常用衡量整体结构和语义的一致性, 评估模型是否抓住了句子的主要结构或语义骨架。其计算公式如方程 4-4 所示:

$$ROUGE-L = \frac{(1 + \beta^2) \times Precision \times Recall}{Recall + \beta^2 \times Precision} \quad (4-6)$$

在本研究中, 为了评估模型对不同病害特征的描述信息提取的准确率, 我们采用了准确率来评估和衡量结果。病害特征主要包括病斑颜色、病斑形状、病斑位置以及病斑面积。准确率为模型能够正确预测出的样本数量与所有参与预测的样本总数之间的比例关系。准确率的计算公式可以表示方程 4-5:

$$\text{准确率}_i = \frac{\text{生成文本中正确分类的样本数}_i}{\text{所有样本}} \quad (4-7)$$

式中, i 为不同病害特征属性, 分别为病害颜色、病斑形状、病斑位置以及病斑面积。

1.16.4 实验平台和配置

实验所采用的硬件配置由 Intel(R) Core(TM) i7-13700KF 中央处理器、128GB 内存、以及 24GB 显存的 NVIDIA GeForce GTX 4090 图形处理单元组成。。此外，实验环境还包括 Python 3.11、Pytorch 2.0 以及 CUDA toolkits 12.2。模型的训练与验证过程均在 Ubuntu 22.04 操作系统环境下进行。

1.17 对比实验设计

1.17.1 不同微调参数对 Qwen-VL 病害特征描述生成的性能影响

为了比较微调模型对病斑特征生成描述的影响，我们使用了 LoRA 对 Qwen-VL 进行了微调。在 LoRA 微调的过程中，我们设置了两组不同的参数设置，具体为秩为 8, α 因子为 16, 以及秩为 16, α 因子为 32。这两组参数是微调大语言模型中常用的参数设置。结合冻结的 Qwen-VL 模型，我们总共得到三组模型的比较。在训练的过程中，我们记录了模型的损失值变化，用于监督模型的训练效果。在训练和验证过程中，使用的数据均保持一致，以保证比较的公平性。微调后的模型经过性能比较后，我们选择了表现较好的模型作为目标模型，命名为 FT-Qwen-VL。Qwen-VL 模型的微调使用了 LLaMa-factory 来完成（Zheng et al. 2024）。冻结模型的参数来源于 Hugging Face 的 transformers 包（Wolf et al. 2020）。

1.17.2 比较 FT-Qwen-VL 与其他视觉语言模型对病害特征描述生成的表现

为了模型生成的病斑特征描述信息中的不同属性，我们使用了 LLM 来辅助提取不同病害特征的描述信息。LLM 同样使用了 Kimi，模型选择为 moonshot-v1-8k，生成时的温度设置为 0.3，以减少模型的随机性，使得更倾向于选择概率最高的词。为了量化不同病斑属性的描述与参考描述的结果，我们设计了特定的提示词，具体为“从病斑颜色、病斑形状、病斑位置和病斑大小四个方面比较以下两个句子,如果颜色描述类似,则返回 1,否则返回 0,其他三个方面同理,缺少描述则返回 0,最终得到类似[1, 1,0,1]四个值。句子 1 为:{caption},句子 2 为:{prediction}理由放在前面,最后的四个值,并用[]输出。”其中 caption 为手动标注的病情特征描述，prediction 是不同模型预测的病情描述。同时，为了避免大语言模型的上下文影响和长距离记忆导致幻觉现象，我们对每个句子对都重新调用了 Kimi 模型 API，并且温度设置为 0.3 以保守估计，尽可能减少不确定性。模型输出的结果如图附表 1 所示。最终，我们得到了不同病斑特征的量化结果用于分析其准确率。

1.17.3 比较视觉语言模型对葡萄病害不同的特征描述的准确率

为了评估微调后的 FT-Qwen-VL 模型在解析葡萄病斑特征方面的表现，我们评估了其他 4 个不同的视觉语言模型，这些模型包括 BLIP2（Li et al. 2023）、Describe Anything Model（DAM）（Lian et al. 2025）、DeepSeek-VL2（Wu et al. 2024）、以及 Kimi-VL-A3B（KIMI-VL）（Team et al. 2025）。这些模型均为开源项目，因此在学术界和工业界都受到了广泛的关注和应用。它们在性能上被认为与闭源模型 GPT-4o 相媲美。为了更好地理解这些模型的性能，我们整理了它们的基本信息，如表 1-1 所示。由于 GPT-4o 是闭源模型，我们无法对其进行直接评估，因此在本研究中并未考虑。

表 1-1 用于评估的 4 个视觉语言模型，输入葡萄病害照片，使用一致的提示词，生成相应的病斑特征描述文本。

Table 1-1 Four visual-language models used for evaluation. Grape disease images are input into each model using specific prompts, and the models generate corresponding textual descriptions of lesion features

模型名	类型	参数量	主体架构	应用场景
BLIP2	图文生成	~220M / ~340M	ViT Encoder + Transformer Decoder	图像字幕生成、视觉 问答、图文检索
Describe Anything Model	图像分割 图文生成	~400M	ViT + SAM + Text Encoder	基于任意文本的语义 分割和描述
DeepSeek-VL2	多模态	~7B	InternLM + 多模态融合	图文问答、图文生 成、描述
Kimi-VL2	多模态	~30B	DeepSeekV3+图像编码 器	图文理解、图文问 答、生成

为了评估这些模型，我们遵循了标准的对话结构，并将一系列带有病害的葡萄叶片图片逐一输入到这些模型中。我们为每个模型提供了一致的提示词，即“请描述叶片病斑的颜色、形状、分布位置和面积大小”（中文翻译）。通过这种方式，我们获得了每张病害图片对应的病斑特征描述文本。在获取了这些描述文本之后，我们采用了以上四个评估指标，对生成的文本进行了评估和对比。同时，我们也计算了这些模型对不同病斑特征属性提取的准确率。

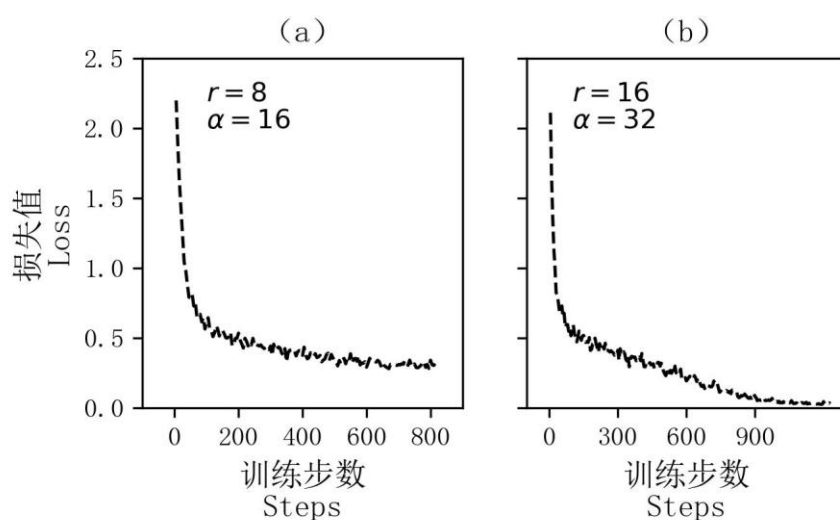
1.18 结果和分析

1.18.1 不同参数微调 Qwen-VL 对病害特征描述生成的表现

图 1-2 显示了两组不同的 Lora 参数设置下，模型训练过程中的损失值变化，最

终我们选择表现较好的模型用于比较试验分析。当参数为 $r=8$, $\alpha=16$ 时, 模型训练了约 800 步达到稳定, 同时评估指标已经超过另一组参数 ($r=16$, $\alpha=32$) 训练的模型表 1-2。当参数设置为 $r=16$, $\alpha=32$ 时, 模型训练超过 900 步。

同时, 表 1-2 表明 LoRA 微调显著优于参数冻结模型。与参数“冻结”相比, 应用 LoRA 微调的两个实验设置在各项指标上都实现了大幅度提升。以 BLEU-4 为例, 冻结条件下仅为 25.39%, 而微调后最高提升至 67.82%, 增幅达 +42.43 %。ROUGE 指标均得到提升, 表明微调不仅提升了文本生成的一致性和覆盖率, 还提高了长距离依赖的信息捕捉能力, ROUGE-1 从冻结的 31.54% 提升至最高 71.54% ($r=8$, $\alpha=16$), 提升 40 %, ROUGE-2 从 7.86% 升至 48.18%, 增长幅度尤为显著 (+40.32%)。ROUGE-L 从 19.77% 升至 61.49%。尽管两组 LoRA 参数设置都远超冻结效果, 但 $r=$



8, $\alpha=16$ 在所有指标上都略优于 $r=16$, $\alpha=32$, 说明较小的秩参数 ($r=8$) 和注意力维度 ($\alpha=16$) 可能在病斑特征描述生成任务中更适合模型泛化。

图 1-2 训练过程中不同训练参数微调 Qwen-2.5-VL-3B 模型的损失变化

Figure 1-2 Loss variation during the training process when fine-tuning the Qwen-2.5-VL-3B model with different training parameters.

表 1-2 微调后的 Qwen 模型在测试集中的表现

Table 1-2 Performance of the fine-tuned Qwen model on the test dataset.

微调参数	模型	BLEU-4	ROUGE-1	ROUGE-2	ROUGE-L
冻结	Qwen-2.5-VL	25.39%	31.54%	7.86%	19.77%
$r=8$, $\alpha=16$	Qwen-2.5-VL	67.82%	71.54%	48.18%	61.49%
$r=16$, $\alpha=32$	Qwen-2.5-VL	67.37%	70.66%	46.85%	60.52%

1.18.2 FT-Qwen-VL 对比其他视觉语言模型病害特征描述生成的结果

表 1-3 显示出 FT-Qwen-VL 在各项指标上表现最优，展现出较强的病害特征描述生成能力，BLEU-4 达 67.82%，ROUGE-1 为 71.54%，ROUGE-2 高达 48.18%，ROUGE-L 为 61.49%，远超 DAM（BLEU-4 为 33.30%）。相比之下，BLIP2 表现最差，BLEU-4 仅为 3.90%，ROUGE-2 仅为 2.63%。DeepSeek-VL2 和 Kimi-VL 表现居中，其中 Kimi-VL 的 BLEU-4 为 23.41%，略优于 DeepSeek-VL2 的 14.99%。从不同指标看，FT-Qwen-VL 在句子覆盖、关键信息匹配和语义保留方面均显著领先，显示其在病斑特征描述生成任务中具备更强的能力。

表 1-3 与其他视觉语言模型对葡萄病害注释生成，在测试集上的表现。DAM 是 Describe Anything Model。FT-Qwen-VL 指的是微调后的 Qwen-VL 模型

Table 1-3 Performance comparison for grape disease annotation generation among various visual-language models. DAM refers to the Describe Anything Model, and FT-Qwen-VL denotes the fine-tuned Qwen-VL model.

模型	BLEU-4	ROUGE-1	ROUGE-2	ROUGE-L
BLIP2	3.90%	22.49%	2.63%	15.94%
DAM	33.30%	34.42%	6.88%	23.33%
DeepSeek-VL	14.99%	27.80%	7.73%	15.43%
Kimi-VL	23.41%	32.68%	9.17%	20.96%
FT-Qwen-VL	67.82%	71.54%	48.18%	61.49%

1.18.3 FT-Qwen-VL 与其他视觉语言模型病害特征描述生成的准确率

表 1-4 显示了 FT-Qwen-VL 在病斑识别四个属性（颜色、形状、位置、面积）上全面领先，整体表现最优，分别达到 80.04%、80.49%、29.27%、86.92%。相比之下，BLIP2 各项指标最低，尤其在形状和位置识别上仅为 7.47% 和 4.07%。Kimi-VL 次之，在颜色和形状方面表现较强（分别为 76.47% 和 71.95%），但在位置识别上为 16.06%，低于 FT-Qwen-VL。DeepSeek-VL2 在形状和面积上表现良好（66.74%、69.00%），但位置识别仅为 9.50%。Qwen-VL 在形状属性上接近 DeepSeek-VL2 和 Kimi-VL（67.42%），位置表现也较平衡（21.49%）。整体来看，FT-Qwen-VL 在精细化病斑特征提取上具备较强的综合能力。

表 1-4 多模型在病斑特征识别任务中的表现（准确率 %）

Table 1-4 Model Performance on Lesion Feature Recognition Task (Accuracy %)

模型	病斑颜色	病斑形状	病斑位置	病斑面积
BLIP2	33.26%	7.47%	4.07%	24.21%
DAM	65.16%	40.27%	21.95%	54.52%
DeepSeek-VL	66.06%	66.74%	9.50%	69.00%
Kimi-VL	76.47%	71.95%	16.06%	69.91%
Qwen-VL	57.24%	67.42%	21.49%	59.28%
FT-Qwen-VL	80.04%	80.49%	29.27%	86.92%

1.18.4 可视化分析示例展示和分析

图 1-3 显示了微调后的 FT-Qwen-VL 模型与冻结的 Qwen-VL 模型，与标注的病斑特征描述生成的示例比较。在第一个示例中，FT-Qwen-VL 准确的识别到 4 个病斑属性，而 Qwen 仅识别出 3 个病斑属性，且病斑形状、病斑颜色识别不对。第二个示例中，FT-Qwen-VL 仍然准确的识别到 4 个病斑属性，而 Qwen 仅识别出 2 个病斑属性，且病斑位置识别不对。



病害特征标注:

This image displays a leaf spot disease on a grape leaf. The lesions are **irregular** in shape, **tan in color**, and have indistinct margins. They are primarily distributed in the **middle and lower parts of the leaf**, covering a **small area**.

Qwen-VL回答:

The image shows a leaf with visible signs of a disease. The leaf appears to be green, but there are noticeable **dark spots** or lesions on it. It has a typical oval shape, which is common for many plant species. Several dark spots are scattered across the surface of the leaf. These spots **vary in size and shape**, indicating different stages of infection or damage. The lesions are distributed unevenly **across the leaf**, suggesting that the disease may have affected various parts of the leaf at different times.

FT-Qwen-VL回答:

This image displays a grape leaf with distinct lesions. The color of the lesions is **light yellow**, the shape is **irregular**, and they are primarily distributed along the **edges and center of the leaf**. The area of the lesions is **small**, presenting a scattered distribution.



病害特征标注:

The grapevine leaf lesions in this image are **irregular** in shape, **light brown** in color, and located at the **marginal areas of the leaves**. The lesions are **small in size** and scattered in distribution.

Qwen-VL回答:

The image shows a leaf with a distinctive heart-shaped form. The leaf has a green color with some lighter areas, indicating possible damage or disease. There are visible lesions on the leaf, which appear as **brown spots** or areas of discoloration. These lesions are located along the **edges and veins of the leaf**, suggesting that the disease may be affecting the vascular system or the outer surface of the leaf. The overall condition of the leaf suggests it might be suffering from a **fungal or bacterial infection**.

FT-Qwen-VL回答:

This image displays a grape leaf with some lesions. The lesions are **light brown** in color, **irregular** in shape, primarily located on the **edges and center of the leaf**. The lesions are **small in area**, numerous, and scattered in distribution.

图 1-3 微调后的 FT-Qwen-VL 模型与冻结的 Qwen-VL 模型，
与标注的病斑特征描述生成的示例

Figure 1-3 Examples of lesion feature descriptions generated by the fine-tuned FT-Qwen-VL model and the frozen Qwen-VL model, compared with the annotated references

1.19 讨论

准确的田间葡萄病害识别对于农户实施精细化病害管理至关重要，有助于确保作物的产量与品质。先前的研究尝试采用单一模态数据（如手机拍摄的照片），通过卷积 CNN 进行病害类型的识别。尽管此方法在识别准确性方面取得了显著成效，但其仅提供了分类标签信息及 CAM 解释图，未能深入描述和理解病害叶片的具体内容，导致这种分类结果的可信度解释存在困难。在本研究中，我们收了 6 种葡萄病害叶片的共 2213 张照片数据，并借助视觉语言模型和专家校正，得到了每张病害照片的细粒度的文本描述。利用 Qwen-VL 模型，我们比较了不同参数设置的 LoRA 微调技术对模型病害特征描述生成的表现。微调后的模型 FT-Qwen-VL，明显优于其他冻结的模型和其他视觉语言模型。另外，我们也评估了 FT-模型病斑的不同属性（颜色、形状、位置、面积）的准确率。FT-Qwen-VL 为葡萄病害识别提供了可解释性强的工具，有助于推动深度学习技术在田间病害识别中的应用，助力田间病害管理。

1.19.1 定制微调视觉语言模型可以显著提升病害特征描述以增强解释性

增强视觉语言模型对病害特征描述的理解，对于可解释的病害识别应用具有重要的现实意义。然而现有的图像分类系统只能预测害虫图像的标签，缺乏对图像内容的解释。使用 CAM 来实现对病害的判断，在实际应用中无法清晰解释决策依据，导致用户无法完全信任检测结果。这成为制约各类病害识别模型实际应用的关键限制原因。因此，如何模拟人类专家的病害诊断过程是亟待解决的关键挑战（Sun et al. 2023）

在本研究中，我们使用了经过专家标注的葡萄六种病害的照片和病害特征标注数据，微调了 Qwen-VL 模型用于生成葡萄病害特征描述生成。研究结果表明，微调后的 FT-Qwen-VL 模型相较于冻结的 Qwen-VL 模型，对于病害描述生成的能力得到显著的提升。以 BLEU-4 为例，冻结条件下仅为 25.39%，而微调后最高提升至 67.82%，增幅达 +42.43 %。同时，对于不同病害特征（颜色、形状、位置和面积），FT-Qwen-VL 也表现优于其他视觉模型。这表明通过微调视觉语言模型，可能实现较为准确的病害特征描述生成能力，从而能够为田间病害识别提供关键的数据支持和显式的可解释性。微调后的模型之所以能够学习到病害特征的描述，可能源于病害训练数据中的专业化知识。这些领域知识强化了视觉语言模型对于病害特征的关键特点表达，从而能够更加符合人工标注的输出结果。同时，微调参数对于模型的性能也有一定的影响。在这些病害特征的描述中，位置信息表现非常差。不同病害特征的准确率一方面是由于模型本能对空间信息的理解能力不足（Tong et al. 2024），另一方面可能是由于拍摄叶片的方向与照片方向不同导致的（Fig-S4-2）。之前的研究有些尝试直接用视觉语言模型对病害进行分类，这样同样导致了模型缺乏对病害特征的描述（Cao et

al. 2023; Lee et al. 2023; Zhang et al. 2024)。另外一些研究采用了对于病害特征描述的学习，特别是应用不同属性的标签来分类，这增加了模型的文本可解释性，也更加符合人类专家对病害分类的认知（Xie et al. 2025; Zhao et al. 2024）。这些研究加深了视觉语言模型在病害识别上的潜在应用价值，特别是对于病害标签的提取，是未来可能得研究方向。

1.19.2 多模态视觉语言大模型辅助农业大数据信息提取

病害特征的提取是非常具有挑战性的工作，需要对相似语义的特征进行判别，以及对关注不同的病害特征。如何得到大量标注准确的信息，以及分析大量的文本数据，对于未来的农业数据研究和应用至关重要。

我们的研究中，借助与视觉语言模型，对病害特征进行了初步的标注，经过人工校正后，得到大量较为准去的病害特征描述数据。这不仅大大降低了人工的工作量，同时也为视觉语言模型进行病害特征描述学习构建了高质量的数据库。在表型数据稀缺且标注成本较高的情况下，视觉语言模型结合预训练权重，为克服这些局限提供了有力的工具。另外，在分析 FT-Qwen-VL 预测的病害描述文本与人工标注作对比时，我同样使用了大语言模型作为工具，设计特定的提示，依赖其强大的文本理解能力，我们可以得到不同病害特征的量化数据。由此可见，基于大语言模型或者视觉语言模型，可以作为强有力的工具，来来生成、分析和管理数据，从而形成循环，服务于数据驱动的农业数据提取和解析。最近的研究已经使用了这样的方式，来循环的生成分割的掩膜数据，从而显著提升了模型的水平（Kirillov et al. 2023b）。利用大语言模型的重要性，已经被广泛的应用于其他下游任务中，显著提升了下游数据的提取和分析，为推动科学研究提供了强有力的支持（Ma et al. 2025; Open AI 2023）。未来的研究中，可以应用定制的大语言模型，来服务专业领域知识的生成、管理以及解析。

1.20 本章小结

本研究中，我们通过在田间和公开数据，收集和整列 2213 张葡萄病害照片，通过借助视觉语言模型，生成和校正了葡萄不同病害特征文本描述数据。在 Qwen-VL 的基础上，通过 LoRA 微调得到了 FT-Qwen-VL。FT-Qwen-VL 在病害特征提取表现上，显著的由于其他视觉语言模型和冻结的 Qwen-VL，如 BLIP2、DeepSeek-VL2、DAM 以及 Kimi-VL 等，同时对不同病害特征提取的准确率，也显著高于其他模型。

我们的研究表明基于微调的视觉语言模型可以定制化的提高专业领域的知识提取，为可解释的田间病害识别提供了关键的参考。未来的研究可以通过应用这种方式，并且结合病害特征，加强使用深度学习技术的田间病害分类模型的应用。进一

步，在未来下游任务中，基于视觉语言模型或者大语言模型，通过微调定制化，可以推动数据驱动的场景的数据生产、管理和解析，解决数据稀缺、解析困难等共性问题。

基于提示引导的葡萄病害分割方法及严重程度评估

去除背景能够显著提高田间病害识别准确性的提升（Tian et al. 2024; Too 2019），然而，由于田间病害叶片数据通常包含其他叶片（相似特征），传统的图像处理和深 CNN 分割模型均存在泛化性差的问题。2023 年以来，视觉基础模型 Segment Anything Model（SAM），通过提示引导，显示出了鲁棒的零样本泛化能力（Kirillov et al. 2023b）。然而，SAM 面临着参数量大，在特定领域（农业）专业性不足，以及在小目标检测效果较差的问题，因此轻量化的视觉基础模型，如 Mobile SAM 等被开发并有望成为终端应用的模型图 1-1。因此，在本章节中，基于田间实地采集的葡萄霜霉病叶片数据，利用 CNN 分割模型，通过结合 Mobile SAM（Zhang et al. 2023 a），以及进一步改进其提示优化策略，评估了其在田间病害叶片分割及病害严重程度量化的表现。

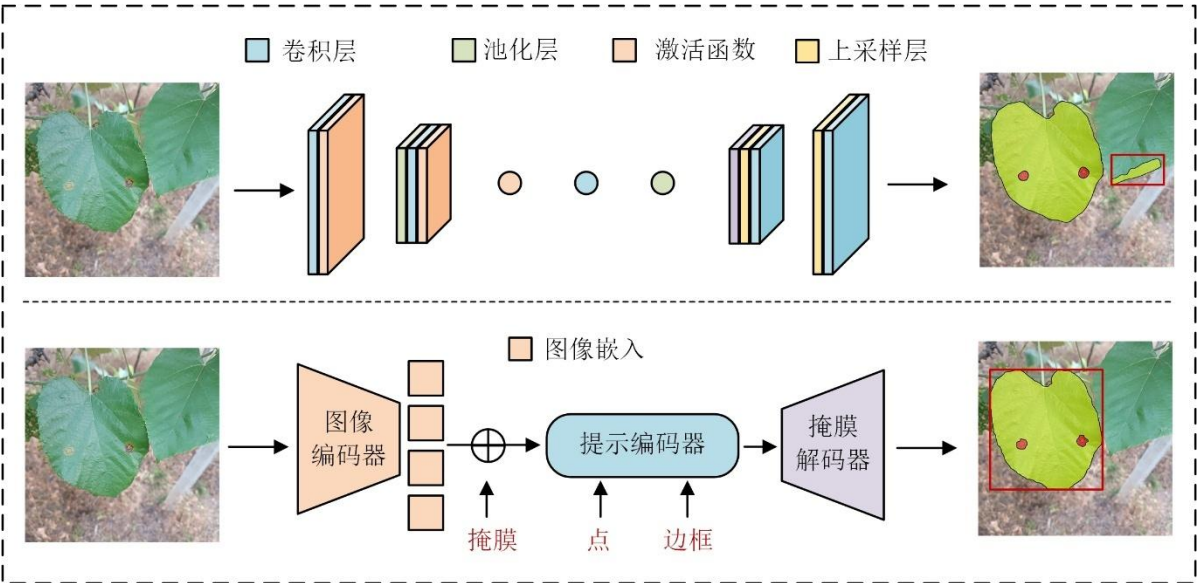


图 1-1 植物病害检测模型的架构示意图。上图展示了一个典型的卷积神经网络（CNN）流程，包括卷积层、池化层、激活函数和上采样层，用于从原始图像中提取特征并进行病害区域的定位。下图展示了一个基于提示的图像分割模型，该模型通过图像编码器、提示编码器和掩码解码器来精确识别和分割病害区域

Figure 1-1 Schematic diagram of plant disease detection model architecture. The upper part shows a typical convolution neural network (CNN) process, including convolution layers, pooling layers, activation function, and up-sampling layers, used for feature extraction from raw images and localization of diseased areas. The lower part illustrates a prompt-based image segmentation model that uses an image encoder, prompt encoder, and mask decoder to accurately identify and segment diseased areas.

1.21 概述

1.21.1 DMS-Prompter: 整合 UNet++与 Mobile SAM 的霜霉病害分割模型

在本研究中，我们整合了 CNN 分割模型和 Mobile SAM 用于田间葡萄霜霉病叶片的粗略分割和细化。该方法的架构设计如图 1-2 所示，其工作流程如下：首先，输入的葡萄叶片图像通过 Mobile SAM 图像编码器，即 TinyViT，进行特征提取。与此同时，该图像也被送入卷积网络，以生成一个粗略的掩膜。本研究使用的卷积网络为 UNet++。这个粗略掩膜随后被用来生成提示信息，这些提示信息包括不同数量的点和边界框，旨在辅助模型更精确地识别病害区域。接下来，这些提取的图像特征和生成的提示信息一同输入到掩膜解码器中。掩膜解码器利用这些信息来生成最终的分割掩膜。通过这种方式，模型能够更准确地定位和分割出叶片上的霜霉病区域。最终得到的模型我们命名为 DMS-Prompter。在图示的模型中，图像编码器主要提取目标图片的语义信息，提示编码器则接收点输入、边框输入和掩膜输入，来生成提示嵌入，再通过解码器对目标进行分割。

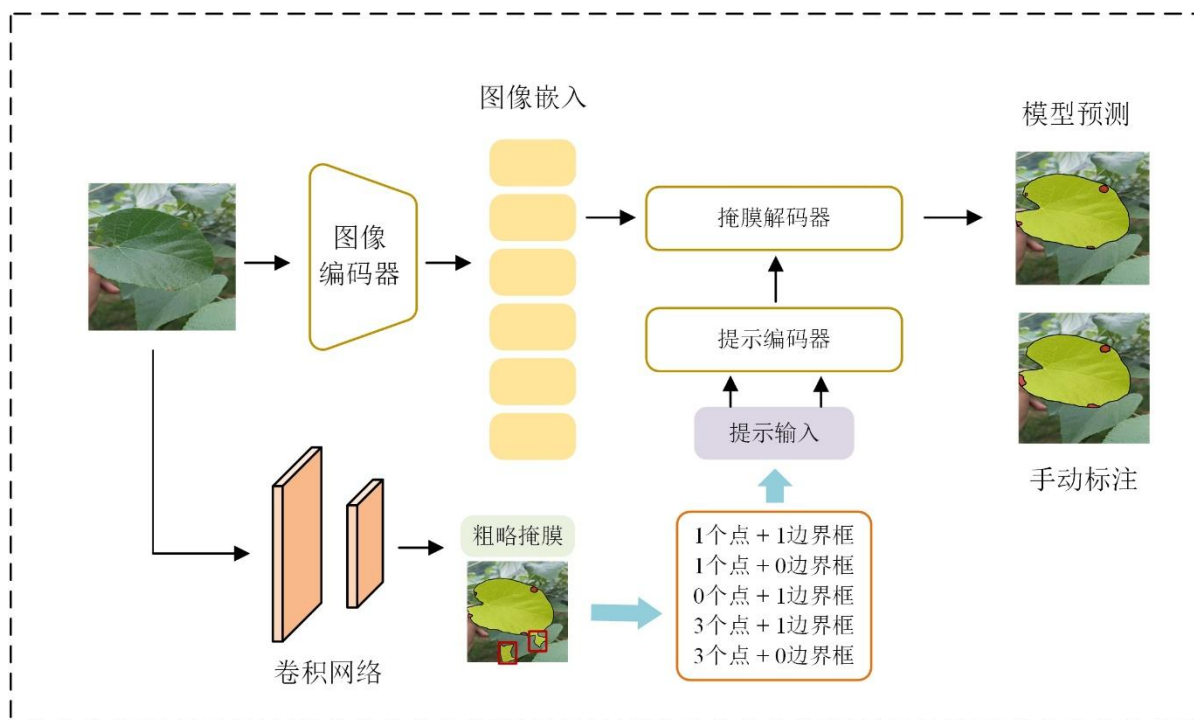


图 1-2 研究中使用的图像分割框架示意图，图像经过卷积神经网络得到粗略掩膜，然后生成不同的边界框提示，输入 Mobile SAM 中进行提示分割。

Figure 1-2 Schematic diagram of the image segmentation framework used in the study. Images are processed by a convolutional neural network to obtain coarse masks, which are then used to generate various bounding box prompts. These prompts are input into Mobile SAM for prompt-based segmentation.

1.21.2 训练 UNet++模型预测叶片和掩膜

利用田间收集的葡萄健康和霜霉病叶片数据，我们对模型进行训练和提示优化。首先，由于提示输入模块需要手动的提示输入，我们先训练了传统的 CNN 模型来得到叶片和病斑粗略的掩膜。在训练过程中，我们使用了以下模型参数，见表 1-1。为了增强模型对边缘的提取能力，我们使用了 Dice 损失。我们使用了 Adam 优化器以增强模型的收敛速度，优化器的学习率设置为 0.0001。此外，我们使用了早停机制，当验证集的得分在连续 10 个训练周期内没有进一步提高，训练过程即终止。

表 1-1 模型训练参数

Table 1-1 hyper-parameters for training DMS-Prompter

模型参数	参数值
损失函数	Dice
优化器	Adam
学习率	1e-4
Early stoping patience	10
训练批次	8

在粗略掩膜的基础上，我们进一步设计了不同的提示策略，用于评估提示类型对分割结果的影响，主要为提示点和边界框的组合。在 CNN 预测生成的掩膜的基础上，我们计算了最大面积叶片的边界框，以及在边界框内随机取点，从而得到具有正标签的提示点（负标签代表背景点，不是分割对象）。得到的提示点或者提示框，输入到 Mobile SAM 的提示编码器中，结合其图像嵌入，得到新的掩膜。

1.21.3 模型评估指标

为了评估叶片和并病斑分割的表现，我们使用了 mIoU@50 作为评估指标。IoU（Intersection over Union，交并比）是衡量预测结果与真实结果重叠程度的指标，其计算公式为方程 5-1：

$$\text{IoU} = \frac{|\text{模型预测} \cap \text{手动标注}|}{|\text{模型预测} \cup \text{手动预测}|} \quad (5-1)$$

式中，模型预测表示预测为正的像素，手动标注表示为标注为正的像素。

mIoU@50 表示在 IoU 大于等于 0.5 的情况下，计算所有样本的平均 IoU：

$$mIoU @ 50 = \frac{1}{N} \sum_{i=1}^N \text{IoU}_i > 0.5 \quad (5-2)$$

此外，我们还使用了 Precision（精确率）、Recall（召回率）和 F1 分数来评估模型性能：

$$Precision = \frac{TP}{TP + FP} \quad (5-3)$$

$$Recall = \frac{TP}{TP + FN} \quad (5-4)$$

$$F1 = 2 \cdot \frac{Precision \times Recall}{Precision + Recall} \quad (5-5)$$

式中，TP（True Positive）表示正确识别的正样本数量，FP（False Positive）表示错误识别为正样本的负样本数量，FN（False Negative）表示错误识别为负样本的正样本数量。在所有评估过程中，仅考虑验证集和测试集中的图像，并基于模型预测的分割掩膜与人工标注的真实掩膜进行计算。

1.21.4 实验平台和配置

实验运行的硬件配置包括 Intel(R) Core(TM) i7-13700KF CPU、128G 内存以及 24G 显存的 NVIDIA GeForce GTX 4090 GPU。软件环境包括 Mobile SAM 模型，来源于 <https://github.com/ChaoningZhang/MobileSAM>，以及 Python 3.10, Pytorch 2.0 和 CUDA 12.2。操作系统为 Ubuntu22.04。经典的图像分割模型如 Unet 等，使用了 segment-models-pytorch 来完成（https://github.com/qubvel-org/segmentation_models_pytorch）。

1.22 对比实验设计

1.22.1 DMS-Prompter 与经典分割模型的性能比较

我们使用了 5 个经典的图像分割模型用于比较分割效果，包括 UNet, UNet++, DeepLabV3, DeepLabV3+, FPN（Feature Pyramid Network），这些模型被广泛的应用于图像分割任务中（Liu et al. 2024; Wang et al. 2021）。这些模型的关键特性及参考文献总结如表 1-2。

表 1-2 使用的分割模型特性

Table 1-2 Key features of the segmentation models used

模型	关键特性	参考文献
UNet	基础的编码-解码架构，编码器的特征图与解码器相同深度的特征图进行融合	(Ronneberger et al. 2015)
UNet++	增强版 UNet，采用密集连接以融合不同深度的特征图	(Zhou Z et al., 2018)
DeepLabV3	采用膨胀卷积扩展感受野，并利用 Atrous Spatial Pyramid Pooling (ASPP) 进行多尺度特征融合	(Chen et al. 2018)
DeepLabV3+	在 ASPP 模块中引入 Xception 及深度可分离卷积，提高分割性能	(Chen et al. 2018)
FPN	构建多尺度特征图的特征金字塔网络。	(Lin et al. 2017)
PAN	在不同尺度的特征图融合前引入注意力模块，扩展了 FPN。	(Lin et al. 2017)

在训练这些模型时，采用 ResNet-34 作为特征提取的主干网络（He et al. 2015），并使用来自 ImageNet 数据集的预训练权重（Deng et al. 2009）。编码器深度设为 5，最大训练轮次（epoch）设为 300，并使用提前停止机制（early stopping），设为 10 轮。批量大小（batch size）设为 64，优化器采用 Adam，初始学习率设为 0.0001。损失函数选择 Dice 系数（Dice coefficient）。最终，保存在验证集上表现最佳的模型权重。为了保证模型对比的公平，训练这些模型所使用的参数，与训练 DSPrompter 保持一致。输出的 Mask 大小，同样为 224×224 。

1.22.2 不同提示策略对 DMS-Prompter 分割表现的影响

为了评估不同的优化策略对田间叶片分割的影响，我们设计了不同的提示组合，主要包括点提示（point）以及框提示（boxes）。在点提示中，我们设计了三种情况，没有点提示，1 个点提示以及 3 个点提示。在边界框提示中，我们只限制了叶片的边界，设计了有边界框提示和无边界框提示两种情况。除了无边界框和无点，总共得到了 5 种情况，分别为 1 个提示点和 1 个边界框，1 个提示点和 0 个边界框，3 个提示点，3 个提示点和 1 个边界框。1 个边界框 5 种情景。五种情景得到的掩膜被才养成与 CNN 模型输出的掩膜大小，即 224×224 ，从而进行公平的比较，评估指标同样为 mIoU@50 以及 F1 分数。

1.22.3 对比 DMS-Prompter 对葡萄叶片霜霉病严重程度评估

为了准确的量化霜霉病严重程度，我们对不同模型的输出，进行了严重程度计算。霜霉病严重程度为感染面积占整个叶片面积的比例。计算公式如方程 5-6 下：

$$\text{严重程度} = \frac{\text{病斑面积}}{\text{叶片面积}} \times 100\% \quad (5-6)$$

在本研究中，整个叶片的面积等于病斑面积和叶片面积的总和。具体计算过程如

下：（1）对于手动标注的照片，我们直接使用标注的病斑的轮廓面积，除以标注的叶片的轮廓面积，来计算得到严重程度。（2）对于预测的病斑和叶片的 Mask，设置 0.5 的阈值，分别将预测的病斑的 Mask 和叶片的 Mask，转化成为二值图像，此时，病斑和叶片的面积分别为二值图像中预测为真的像素数量。预测的严重程度则为对应的像素数量比值，其中叶片的面积等于预测的病斑像素和叶片像素的总和。

为了比较预测可靠性，我们使用了决定系数(R^2)来评估手动标注的严重程度与预测的严重程度是否一致。 R^2 的计算公式如方程 5-7 下：

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (5-7)$$

式中， y_i 代表手动标注的严重程度， $\hat{y}_i$ 代表预测的严重程度， $\bar{y}$ 代表手动标注严重程度的平均值， n 代表样本数量。 R^2 的值越接近于 1，表示预测结果与手动标注结果的一致性越高。

1.23 实验结果和分析

1.23.1 DMS-Prompter 与经典分割模型的性能比较

我们对比评估了 7 种语义分割模型在验证集和测试集上的性能表现，包括 PAN、FPN、DeepLab V3、DeepLab V3+、UNet、UNet++ 以及我们提出的 DMS-Prompter 模型。性能评估主要基于 mIoU@50 和 F1 score 两个指标，结果如表 1-3 所示。

从总体上看，DMS-Prompter 在所有模型中取得了最优性能，其在验证集上 mIoU@50 为 97.61%，F1 score 达到 98.77%，在测试集上 mIoU@50 和 F1 score 分别为 97.64% 和 98.79%，显著优于其他基线模型。这表明 DMS-Prompter 不仅在训练阶段表现优异，在泛化能力方面也更具优势。具体来看，传统架构如 PAN 与 FPN 在性能上相对靠后，测试集 mIoU@50 分别为 94.81% 和 95.09%，而 F1 score 分别为 97.25% 和 97.40%。中等水平模型如 DeepLab V3 和 DeepLab V3+ 在两个指标上均有一定提升，尤其是 DeepLab V3+ 在测试集上的 F1 score 达到了 98.15%。进一步提升的是 UNet 系列模型，其中 UNet++ 略优于 UNet，在测试集上的 mIoU@50 和 F1 score 分别为 96.93% 和 98.42%。然而，尽管 UNet++ 已表现出较高的精度，DMS-Prompter 仍在所有指标上全面领先，充分验证了本方法在高精度语义分割任务中的有效性与先进性。

表 1-3 与基准模型比较的结果，mIoU@50 指的是在 0.5 阈值条件下预测的平均交并比。

Table 1-3 The results compared with the baseline model. mIoU@50 is the mean Intersection over Union under a threshold of 0.5.

模型名称	验证集		测试集	
	mIoU@50	F1 score	mIoU@50	F1 score
PAN	95.28%	97.50%	94.81%	97.25%
FPN	95.66%	97.72%	95.09%	97.40%
DeepLab V3	96.67%	98.29%	96.15%	98.00%
DeepLab V3+	96.85%	98.39%	96.46%	98.15%
UNet	97.45%	98.69%	96.90%	98.38%
UNet++	97.49%	98.72%	96.93%	98.42%
DMS-Prompter	97.61%	98.77%	97.64%	98.79%

1.23.2 不同提示策略对 DMS-Prompter 分割表现的影响

我们进一步研究了不同提示（Prompt）组合对模型性能的影响，探讨“Points”和“Box”两类提示在语义分割任务中的作用与协同效果。实验结果如表 1-4 所示。

从单一提示类型来看，仅使用“Points”时，模型性能较低，验证集和测试集的 mIoU@50 分别为 70.26%和 70.07%，F1 score 为 73.26%和 72.54%，说明单点提示无法提供足够的空间约束信息。而当仅使用“Box”提示时，模型性能显著提升，验证集 mIoU@50 达 97.61%，F1 score 为 98.77%，测试集指标亦相应达到 97.64%和 98.79%，充分显示出 Box 提示在提供形状边界信息方面的重要性。当同时使用“Points+Box”组合时，模型在验证集和测试集上均表现出稳定且高精度的性能，其中 mIoU@50 分别为 97.29%和 97.45%，F1 score 为 98.59%和 98.68%，虽然略低于仅用 Box 提示的表现，但依旧优于大多数传统分割方法。进一步地，我们引入密集提示（即多个点提示，表示为“√√√”）进行测试。结果表明，密集的点提示在没有 Box 信息时的性能虽优于单点提示，但在测试集上 mIoU@50 仅为 75.84%，F1 score 为 79.74%，表明即使增加点的数量，也难以替代 Box 的空间上的引导。当密集点提示与 Box 结合使用时，模型性能接近最优，在验证集和测试集上 mIoU@50 分别为 96.53%和 95.73%，F1 score 分别为 98.18%和 97.40%。

表 1-4 不同提示组合对 DMS-Prompter 模型语义分割性能的影响。

×表示没有使用，√表示单个使用。

Table 1-4 Analysis of the Impact of Different Prompt Combinations on the Semantic Segmentation Performance of the DMS-Prompter Model. × indicates not used, √ indicates single use.

提示组合		验证集		测试集	
Points	Box	mIoU@50	F1 score	mIoU@50	F1 score
√	×	70.26%	73.26%	70.07%	72.54%
×	√	97.61%	98.77%	97.64%	98.79%
√	√	97.29%	98.59%	97.45%	98.68%
√√√	×	85.28%	88.45%	75.84%	79.74%
√√√	√	96.53%	98.18%	95.73%	97.40%

1.23.3 病害可是化示例展示和分析

图 1-3 展示了不同模型与人工标注的分割效果对比。总体而言，DMS-Prompter 的集成模型在限制对背景叶片的分割方面表现出色，而传统模型在面对复杂背景时则容易受到干扰。具体来说，在第一行主要叶片之间并没有实际的联通关系，因此，在引入了提示信息之后，我们能够更加精确地消除背景因素的干扰。在图像处理的第二行和第三行中，可以观察到，这些干扰部分与主要叶片相连通。在这种复杂的情况下，传统的图像分割模型往往难以准确区分叶片的边缘部分。然而，通过使用 DMS-Prompter，我们可以显式地识别并区分出叶片的边缘部分，从而实现更精细的分割效果。因此，DMS-Prompter 能够在现有的 CNN 的基础上，进一步对分割的掩膜进行细化处理，显著提高了图像分割的精度和效果。

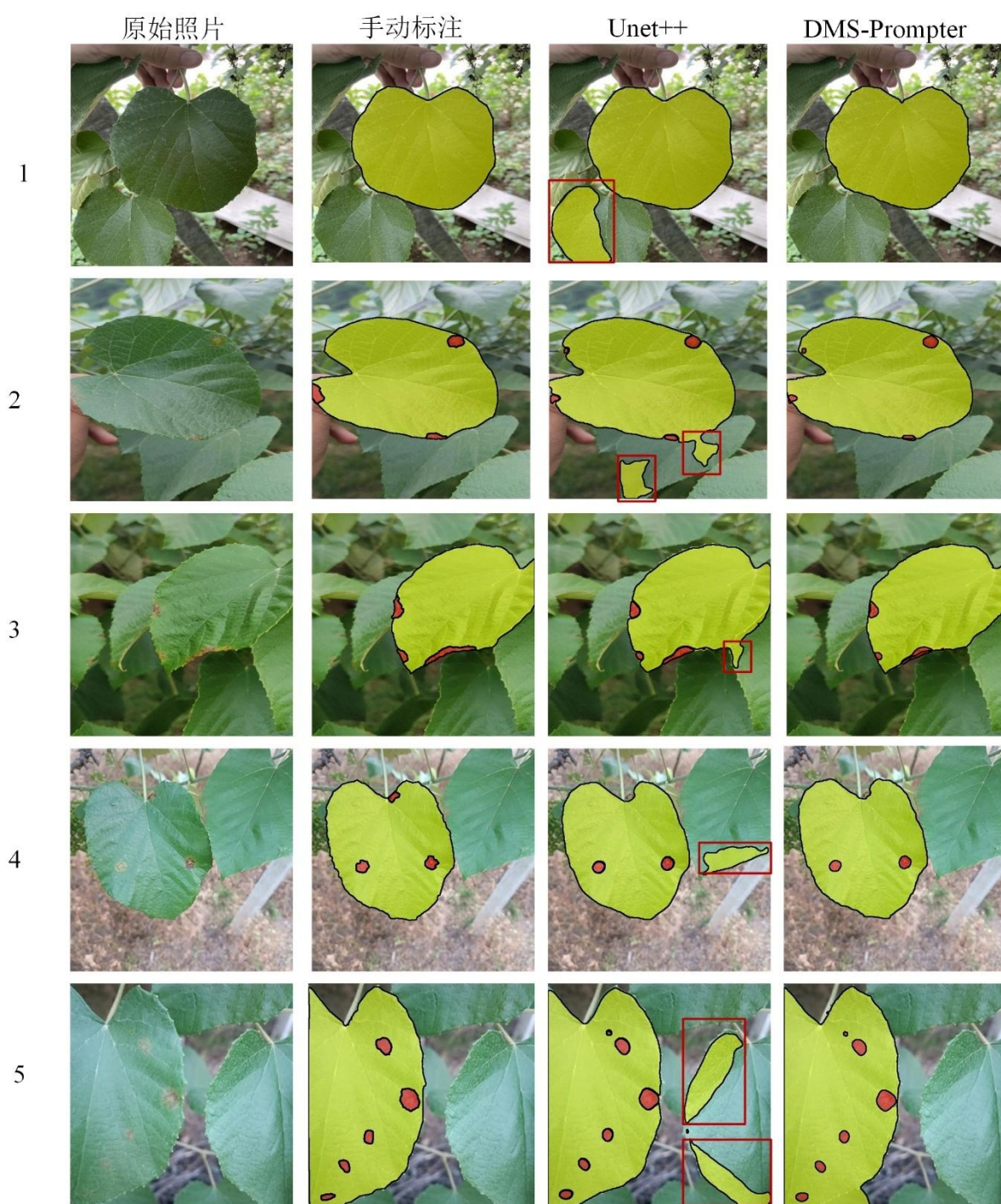


图 1-3 田间葡萄霜霉病叶片的分割效果示意图。从左到右依次为原始照片、手动标注的掩膜，UNet++预测的掩膜，以及 DMS-Prompter 生成的掩膜。红色边框为错误分割的部分。

Figure 1-3 Illustration of segmentation results for grape downy mildew leaves in the field. From left to right: original image, manually annotated mask, mask predicted by UNet++, and mask generated by DMS-Prompter. Red boxes indicate incorrectly segmented regions.

1.23.4 模型对葡萄叶片霜霉病严重程度评估

图 1-4 中可以看出, 在验证集和测试集中, 模型预测的田间葡萄霜霉病的严重程度与人工手动标注的严重程度之间存在高度的一致性。具体来说, 在测试集中, 葡萄霜霉病的严重程度分布范围在 0 到 0.15 之间, 这一分布情况与验证集中的分布情况保持一致。进一步地, 在验证集中, 模型的 R^2 值达到了 0.93, 而在测试集中, R^2 值提升到了 0.94。这一结果不仅表明了模型预测结果与人工标注结果之间具有高度的一致性, 而且还充分证明了该模型在不同数据集上具有较强的泛化能力, 能够准确地预测葡萄霜霉病的严重程度。

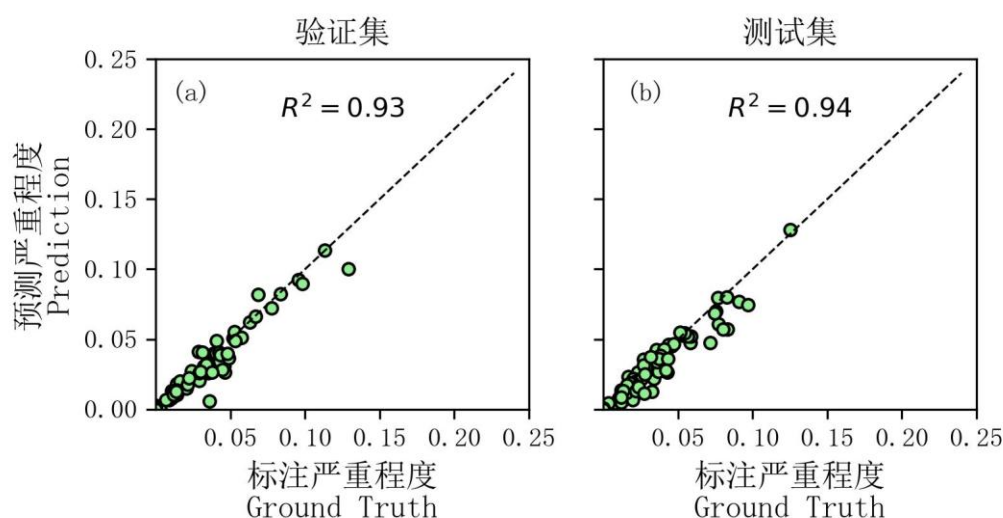


图 1-4 在测试集和验证集中, 手动标注的葡萄霜霉病严重程度与模型预测的严重程度的一致性。

(a) 和 (b) 分别为在验证集和测试集的 R^2 。 R^2 为决定系数。

Figure 1-4 Consistency between the manually annotated severity of grape downy mildew and the model-predicted severity in the validation and test sets. (a) and (b) show the R^2 values for the validation and test sets, respectively. R^2 is the coefficient of determination.

1.24 讨论

准确评估葡萄田间霜霉病的严重程度对于制定精细化的防控措施, 例如农药施用剂量等, 具有显著的应用价值。先前研究通过应用 CNN 实现了对病斑与叶片的分割, 取得了较好的分割效果。然而, 在面对复杂背景, 尤其是与叶片颜色相似的背景时, 传统 CNN 分割模型的泛化性能就会下降, 导致分割结果的不精确。因此, 本研究通过田间采集的葡萄霜霉病叶片和健康叶片数据, 进行手动和半自动的标注后, 得到田间的病害分割数据集。我们首先使用了 CNN (UNet++) 对田间叶片进行了训练和验证, 得到叶片和病斑的粗略掩膜, 并根据粗略掩膜生成了不同组合的点提示和边界框提示。接下来, 我们整合了基于提示分割的 Mobile SAM 模型, 利用其在预训练权重基础上得到的图像特征, 对田间叶片进一步分割, 以增强对相似背景的抑制效

果。结果表明，整合的模型能够更加细化分割效果。这个整合的模型被命名为 DMS-Prompter。DMS-Prompter 为田间病害的精确量化提供了有效的工具，以帮助农户更准确地监测田间病害的发生情况，从而有助于病害的防控工作，并确保作物的产量与品质。

1.24.1 整合 CNN 和 Mobile SAM 可以实现病斑的细致分割

准确的分割田间叶片和病斑对于评估病情严重程度和田间病害管理至关重要。使用 CNN 网络虽然能够得到较高的分割效果，然而田间复杂的背景以及早期病斑面积小等因素，制约了 CNN 模型在田间的实际应用。我们的研究中，通过结合 CNN 和基于提示学习的 Mobile SAM 模型，实现了葡萄叶片及霜霉病斑的精细分割，从而为田间病害评估和管理提供了有效的工具。

我们的研究表明，整合 CNN 和 Mobile SAM 以后，田间叶片和病斑分割 mIoU @50 在验证集和测试集上分别达到了 97.71% 和 97.74%，F1 得分分别达到了 98.77% 和 98.79%。研究结果意味着 DMS-Prompter 模型可以实现对叶片和病斑的细化分割，从而可以进一步做严重程度的精准评估。这种准确的分割效果，可能主要来源于以下三个方面：（1）CNN 模型对于局部信息提取具有显著的优势，因此具有识别病斑的能力，（2）Mobile SAM 以及 SAM 模型，模型架构基于 Transformer 中，Transformer 对于形状信息具有提取优势，因此可以提取出叶片的边缘信息（Naseer et al. 2021）。（3）Mobile SAM 结合了提示编码，从而限制了背景的干扰。然而，CNN 对于小目标的病斑难以提取到，而基于 Transformer 的模型，则对于病斑难以准确的获取。主要原因可能是当使用 CNN 提取特征时，输入图像尺寸被压缩，在卷积过程中，小目标信息丢失，而 Transformer 结构在由于更加关注全局信息，没有办法在少量数据集的情况下，提取出病斑的特征。另外，仅仅使用 Transformer 的模型，在数据集多的情况下表现较好，数据量少则难以达到 CNN 的效果，而表型数据的稀缺则面临着数据量少的情况。之前的研究中，背景的干扰是主要的影响因素，因此去除背景是关键步骤（Tian et al. 2024; Yang et al. 2024）。也有一些研究使用了自提示的策略，来生成目标对象的掩膜（Lin et al. 2023; Wu et al. 2023）。然而，这些自提示的方法往往需要在 SAM 的图像嵌入中存在目标对象的特征，我们的分析表明 SAM 及 Mobile SAM 在田间病害叶片的图像编码中，均没有办法得到其特征图（附图 5、附图 6）。因此，未来的研究中，可以考虑设计将 CNN 和 Transformer 的优势结合，以实现端到端的病害准确分割。

1.24.2 优化提示策略可以增强分割效果以提高病害评估

使用点提示、边界框提示以及掩膜提示，可以显式的为视觉基础模型，如 SAM 等，提供引导，从而实现特定目标的准确分割。然而，仅仅基于点提示或者边界框提

示，不能很好的实现目标的分割（Chen et al. 2023; Ji et al. 2023），因此，评估不同提示组合对分割效果的影响，能够帮助我们更好的使用大模型来适应下游任务，如田间病害叶片的分割。

在本研究中，我们评估不同的点提示和边界框提示对模型表现的影响，结果表明，使用边界框提示，在高质量分割中起到了关键性作用，而点提示则在特定情境下提供了有效补充。这可能是由于点提示仅提供一个像素位置，在目标较大、边缘模糊或重叠复杂时，模型不知道目标的真正范围。如果点提示偏离目标或点在背景区域，模型可能会输出错误结果。而边界框提示则通过基于 Transformer 的视觉分割模型，可以从提示中提取区域上下文，引导注意力机制专注目标范围，从而提供目标的位置、大小、形状范围。最近的研究同样表明框提示有着明显的优势（Ali et al. 2024; Mazurowski et al. 2023）。然而，尽管边界框提示具有显著的作用，边界框的生成仍然是一个具有挑战性的问题。比如基于 YOLO 系列的目标检测模型，被广泛的应用于提示框的生成（Xu et al. 2024; Yang et al. 2024; Zhang et al. 2024）。然而在实际应用中，仍旧面临小目标检测以及背景干扰问题。因此，在未来的研究中，可能需要进一步考虑优化小目标检测，以及结合其他提示策略辅助分割。

1.24.3 准确的病害记录有助于提升病害预警水平和优化田间管理

对田间病害的准确记录，对于优化田间病害管理具有重要的意义，如药剂使用量和喷洒策略（Hui et al. 2022; Puellas et al. 2024）。当下田间的病害严重程度判断，往往以来农户经验判断或者分级操作，存在一定的主观性。准确的田间病害评估，不仅可以评价不同品种葡萄抗性水平，也可以作为药剂效用评价的关键指标。因此，定量的严重程度评估对于田间病害管理具有实际意义（Carraro et al. 2023）。

在本研究中，我们开发了用于准确分割田间葡萄霜霉病叶片和病斑的模型 DMS-Prompter，该模型可以实现叶片尺度的病害严重程度准确计算，模型结果与人类标注结果相关性达到 0.98。对于霜霉病叶片的准确分割，可能取决于 Mobile SAM 模型，提供了显式的提示分割功能，避免了田间照片复杂背景的影响，因为复杂背景中可能包括其他叶片，以及与病斑颜色相似的土壤特征，导致分割的不准确。之前的研究，无论是使用 CNN，或者基于 Transformer 模型的病斑点分割，都不可避免的将背景分割出来，或者将其他叶片上的病斑分割出来（Chen and Wu 2022; Zhang et al. 2024），从而导致病害评估的误差。然而，对于早期病斑，其颜色特征不明显，目标面积较小，因此对于早期病斑的分割，仍旧是目前的挑战，而早期的病斑恰恰是进行田间病害管理的关键时机（Cooke et al. 2006; Richard et al. 2022）。这不仅表明我们需要采集更多的早期田间病害表型数据，同时也需要进一步优化模型，来实现早期病害的监测和评估。因此，在未来的研究中，可能需要提升早期病害数据的搜集和建

模，以进一步提高田间病害预警水平，从而服务于田间病害管理。

1.25 本章小结

本研究中，我们在田间手机了葡萄霜霉病的实地照片，687 张高质量的照片数据，其中包含 338 张带有霜霉病症状的叶片照片，以及另外 349 张健康叶片的照片。通过半自动的标注叶片和病斑区域，我们生成了一套用于田间霜霉病严重程度评估的数据集。基于该数据集，我们设计了整合 CNN 和 Mobile SAM 的田间病害分割框架。结果显示，在使用的 CNN 分割模型中，UNet++ 优于其他分割模型，如 UNet、DeepLab V3+ 等，在验证集和测试集上 mIoU@50 分别达到了 97.49% 和 96.93%，F1 得分达到了 98.72% 和 98.42%。使用 Mobile SAM 提示分割后，这两个指标分别得到了提高。同时，我们比较了多种提示组合对分割结果的影响，结果表明边界框提示对于准确的分割具有重要的影响。基于优化后的分割模型，我们评估了田间葡萄霜霉病严重程度，在验证集和测试集上表现出了较高的一致性，R2 分别为 0.93 和 0.94。

我们的研究表明基于提示的视觉基础模型可以有效的细化田间病害表型的分割，对于田间病害的管理具有显著的应用潜力。未来的研究可以进一步提高提示的准确性，以及探索更多类型的提示，如语义分割提示、关键点提示等，以进一步提升病害分割的精度和效率。同时，也可以考虑将本研究的方法拓展到其他田间病害的识别与管理中，为田间病害记录和管理优化提供更多的技术支持。

结论与展望

1.26 主要结论

本研究聚焦于葡萄霜霉病在叶片尺度上的抗性评估与病害识别，设计并实施了一套多层次的技术方案。首先，针对实验室条件下的叶盘接种图像，结合 YOLO v8n、Mobile SAM 与 UNet，开发了自动化模型 SporaScan，实现叶盘检测、背景剔除和孢子囊区域分割，并以孢子覆盖面积比例量化抗病性，进一步与 Mask R-CNN 等方法进行性能对比分析。其次，借助多种葡萄病害图像与经专家修订的描述文本，构建高质量图文对齐数据集，利用 LoRA 技术对 Qwen-2.5-VL-3B 模型进行轻量微调，评估其在病害特征描述生成与细粒度区分方面的表现，并与其他主流视觉语言模型进行对照实验。最后，面向实际田间环境，采集霜霉病与健康叶片图像，结合 CNN 分割模型与 Mobile SAM，系统分析点提示与边界框提示在叶片和病斑分割任务中的差异，构建适用于野外场景的病害严重程度评估系统，验证其泛化能力与实用价值。该研究整合图像检测、精细分割与多模态理解，形成从实验室到田间的完整智能化病害识别与评估流程。主要结果如下：

（1）序列化模型精准高效的完成实验室叶盘霜霉病抗性鉴定

本研究开发了 SporaScan，一个整合 YOLO v8n、Mobile SAM 和 UNet 的序列化模型，能够高效完成葡萄叶盘的检测、背景剔除与霜霉病区域分割。相比于 Faster R-CNN，YOLO v8n 实现了更高的检测效率与准确度（ $mAP@50 > 99\%$ ），而 UNet 模型分割效果优越（ $mIoU@50 > 96\%$ ）。该模型在严重程度评估上与人工标注高度一致（ $R^2=0.99$ ），不仅提升了精度，也显著减少了人工标注的劳动强度和主观误差。

（2）定制化视觉语言模型提升了对病害特征描述的生成能力

我们基于田间与公开病害图像，构建了葡萄病害图文配对数据集，并通过 LoRA 方法微调 Qwen-VL 模型，形成定制化模型 FT-Qwen-VL。该模型在病斑颜色、形状、位置与面积等特征的描述生成上，显著优于原始冻结模型与其他主流视觉语言模型（如 BLIP2、DeepSeek-VL2、Describe Anything Model 和 Kimi-VL-A3B），展现出更强的表征能力与任务适配性。

（3）提示引导的视觉基础模型可以提升和细化田间分割表现

我们构建了基于 687 张田间图像的病害分割数据集，并提出结合 UNet++ 与 Mobile SAM 的提示引导式分割框架。实验证明，边界框提示在提升分割准确性方面效果最优，显著优于点提示组合。优化模型在验证集和测试集上的 $mIoU@50$ 均超过 96%，F1 得分超过 98%，严重程度预测 R^2 分别达 0.93 与 0.94，表明其在田间复杂背

景下具备良好的泛化性能与实际应用潜力。

1.27 主要进展和创新点

(1) 本研究基于低成本的手机照片数据，开发了基于深度学习的序列化的叶盘图像分析模型，该模型可以实现高效精准的实验室霜霉病抗性鉴定。

(2) 本研究借助视觉语言模型，建立了葡萄病害照片及病害特征描述的文本数据库，并使用 LoRA 微调定制化了 Qwen-2.5-VL-3B 模型。定制化模型提升了病害特征描述生成的准确性。

(3) 本研究探索了提示机制优化田间分割效果，表明通过优化提示策略，可以提升模型在复杂环境下的泛化能力，增强分割性能。

1.28 研究局限与展望

1.28.1 研究局限

(1) 病害的特征表现形式多种多样，而其严重程度仅仅是众多评估指标中的一种。在实验室环境中，除了严重程度，研究人员还会关注其他一些关键指标，例如产孢密度和坏死率等 (Buonassisi et al. 2018)。产孢密度是指单位面积内产生的孢子数量，这一指标能够反映出病原菌的繁殖能力和传播潜力。而坏死率则是指植物组织因病害而死亡的比例，这一指标能够直观地展示病害对植物组织的破坏程度。在田间监测过程中，除了实验室中的指标外，还会关注一些更为直观的田间指标，如病叶率。病叶率是指受病害影响的叶片数量占总叶片数量的比例，这一指标能够反映出病害在田间的扩散程度和对作物整体健康状况的影响 (Rossi and Caff 2009)。通过对这些多样化的指标进行综合评估，才能够更全面地了解病害的发生、发展情况，从而为病害的防控提供更为科学的依据。

(2) 在本研究中，我们采用了多种模型的融合，并对学习策略进行了优化。尽管如此，在新模型的设计方面，我们仍然存在一些研究不足之处。具体来说，我们结合了多种先进的模型，通过融合这些模型的优势，以期达到更好的性能。同时，我们还对学习策略进行了深入的优化，以提高模型的训练效率和泛化能力。然而，在新模型的设计和创新方面，我们还需要进一步加强研究，以弥补现有的不足，确保研究的全面性和深入性。

1.28.2 未来展望

(1) 在未来的研究中，将进一步加强田间病害表型数据的系统化收集。通过构建标准化的数据采集流程，结合多模态传感器设备，可获得更全面、高质量的病害图像与环境信息，为下游任务数据分析提供数据支持。在数据处理方面，可以基于先进

的视觉基础模型开展自动化图像标注研究。利用大规模预训练模型的表征能力，结合少量人工校正，提高标注效率与准确性，缓解当前农业图像标注成本高、速度慢的问题。同时，将探索视觉语言模型在农作物病害识别中的应用潜力。通过融合图像特征与语义信息，有望提升模型对复杂、多样化病害类型的识别能力，特别是在少样本或多发病共存场景下的表现。

（2）进一步地，未来研究将侧重于构建具备零样本泛化能力的视觉模型体系，以适应实验室与田间多变环境下的病害识别需求。依托预训练视觉基础模型，通过跨场景迁移与任务适配，可实现对新病害类别或新作物类型的快速识别，降低对人工标注数据的依赖。在此基础上，可以将表型组学与基因组学数据融合进视觉识别流程，实现表型到基因层面的深度学习模型驱动，助力高抗病作物的精准筛选与育种进程。同时，通过提升模型对病害时空分布的解析能力，可为病害管理系统提供更高效率的决策支持，推动农药使用的精量化和减量，促进农业绿色发展。

参考文献

- 安玉善,闫德祺.2018.葡萄霜霉病发病规律及防治技术[J].黑龙江农业科学(12):176-178.
- 毕秋艳,韩秀英,马志强,赵建江,贾海民,王文桥.2018.寡雄腐霉与烯酰吗啉互作防治葡萄霜霉病和替代部分化学药剂减量用药应用[J].植物病理学报,48(5):675-681.
- 付晴晴,褚燕南,王跃进,徐炎.2019.葡萄霜霉病抗性育种研究进展[J].中外葡萄与葡萄酒(5):69-75.
- 郝倩琳,杨廷志,吕新茹,秦慧敏,王亚林,贾晨飞,夏先春,马武军,徐登安.2024.小麦胚芽鞘长度 QTL 定位和 GWAS 分析[J].作物学报,50(3):590-602.
- 李博雅,冉隆贤,施李鸣,张克诚,葛蓓亭.2023.葡萄霜霉病研究进展[J].中国农学通报,39(4):125-131.
- 李嘉泓,李文学,顾沛雯.2019.基于 ArcGIS 的贺兰山东麓酿酒葡萄霜霉病发生流行的预测预报技术初步探究[C].119.
- 李文学,郭琰杰,胡婧桢,顾沛雯.2017.贺兰山东麓酿酒葡萄霜霉病孢子囊时空动态与田间病情及环境因子相关分析[C].70.
- 李小红,李运景,马晓青,郭军,刘海礁,郑国清,陶建敏.2021.我国葡萄产业发展现状与展望[J].中国南方果树,50(5):161-166.
- 刘会彪.2024.河南地区露地栽培葡萄霜霉病发生规律及不同药剂防治效果研究[D].郑州:河南农业大学.
- 刘柯含,张芮,高彦婷,张红娟,温健,薛莲,李青青.2024.中国葡萄产业现状分析及其发展对策[J].中国果树(7):132-138.
- 王颖.2021.葡萄霜霉病发生特点与绿色防控技术[J].农村新技术(4):21-23.
- 吴宁,陈天恩.2020.基于灰色关联和 PSO-SVM 的葡萄霜霉病的短期预测[J].湖南农业大学学报(自然科学版),46:02.
- 杨文庆,刘天霞,唐兴萍,徐国富,马喆,杨贺凯,吴文斗.2022.智慧农业背景下的植物表型组学研究进展[J].河南农业科学,51(7):1-12.
- 于舒怡,李柏宏,刘丽,王辉,关天舒,刘长远.2024.葡萄霜霉病流行传播模型和扩散规律研究[J].中国植保导刊,44(1):46-55.
- 于舒怡,王辉,刘丽,关天舒,李柏宏,刘长远.2018.辽宁葡萄霜霉病流行动态及防控对策[J].湖北农业科学,57(13):34-36.
- 张玮,燕继晔,刘梅,彭军波,邢启凯,李兴红.2020.葡萄霜霉病流行与预测研究进展[J].中国果树(3):11-15.
- Abayomi-Alli OO, Damaševičius R, Misra S, Maskeliūnas R. 2021. Cassava disease recognition from low-quality images using enhanced data augmentation model and deep learning[J]. Expert Systems.
- Afzaal U, Bhattarai B, Pandeya YR, Lee J. 2021. An Instance Segmentation Model for Strawberry Diseases Based on Mask R-CNN[J]. Sensors, 21(19): 6565.
- Ahmed N, Asif HMS, Saleem G. 2021. Leaf Image based Plant Disease Identification using Color and Texture Features[J]. Wireless Personal Communications, 121(2): 29.
- Ali L, Alnajjar F, Swavaf M, Elharrouss O, Abd-alrazaq A, Damseh R. 2024. Evaluating segment anything model (SAM) on MRI scans of brain tumors[J]. Scientific Reports, 14(1).

- Amo A, Soriano JM. 2022. Unravelling consensus genomic regions conferring leaf rust resistance in wheat via meta-QTL analysis[J]. *Plant Genome*, 15(1).
- Araus JL, Kefauver SC. 2018. Breeding to adapt agriculture to climate change: affordable phenotyping solutions[J]. *Current Opinion in Plant Biology*, 45: 237-247.
- Arshad MA, Jubery TZ, Singh AK, Singh A, Hegde C, Ganapathysubramanian B, Balu A, Krishnamurthy A, Sarkar S. Assisted few-shot learning for vision-language models in agricultural stress phenotype identification[J].
- Awais M, Alharthi AHSA, Kumar A, Cholakkal H, Anwer RM. 2025. AgroGPT: efficient agricultural vision-language model with expert tuning[A]. arXiv.
- Bai S, Chen K, Liu X, Wang J, Ge W, Song S, Dang K, Wang Peng, Wang S, Tang J, Zhong H, Zhu Y, Yang M, Li Z, Wan J, Wang Pengfei, Ding W, Fu Z, Xu Y, Ye J, Zhang X, Xie T, Cheng Z, Zhang H, Yang Z, Xu H, Lin J. 2025. Qwen2.5-VL technical report[A]. arXiv.
- Barbedo JGA. 2018a. Factors influencing the use of deep learning for plant disease recognition[J]. *Biosystems Engineering*, 172: 84-91.
- Barbedo JGA. 2018b. Impact of dataset size and variety on the effectiveness of deep learning and transfer learning for plant disease classification[J]. *Computers and Electronics in Agriculture*, 153: 46-53.
- Bebber DP, Ramotowski MAT, Gurr SJ. 2013. Crop pests and pathogens move polewards in a warming world[J]. *Nature Climate Change*, 3(11): 985-988.
- Bierman A, LaPlumm T, Cadle-Davidson L, Gadoury D, Martinez D, Sapkota S, Rea M. 2019. A High-Throughput Phenotyping System Using Machine Vision to Quantify Severity of Grapevine Powdery Mildew[J]. *Plant Phenomics*, 2019: 2019/9209727.
- Boso S, Kassemeyer HH. 2015. Different susceptibility of European grapevine cultivars for downy mildew[J]. *VITIS - Journal of Grapevine Research*: 39 Pages.
- Bove F, Savary S, Willocquet L, Rossi V. 2020. Designing a modelling structure for the grapevine downy mildew pathosystem[J]. *European Journal of Plant Pathology*, 157(2): 251-268.
- Bradski G. 2000. The OpenCV Library[M].
- Brahimi M, Arsenovic M, Laraba S, Sladojevic S, Boukhalfa K, Moussaoui A. 2018. Deep Learning for Plant Diseases: Detection and Saliency Map Visualisation[M]. In: Zhou J, Chen F (editors), *Human and Machine Learning*. Cham: Springer International Publishing, pp. 93-117.
- Buonassisi D, Cappellin L, Dolzani C, Velasco R, Peressotti E, Vezzulli S. 2018. Development of a novel phenotyping method to assess downy mildew symptoms on grapevine inflorescences[J]. *Scientia Horticulturae*, 236: 79-89.
- Burruano S. 2000. The life-cycle of *Plasmopara viticola*, cause of downy mildew of vine[J]. *Mycologist*, 14(4): 179-182.
- Cao Y, Chen L, Yuan Y, Sun G. 2023. Cucumber disease recognition with small samples using image-text-label-based multi-modal language model[J]. *Computers and Electronics in Agriculture*, 211: 107993.
- Carisse O. 2010. Fungicides[M]. Erscheinungsort nicht ermittelbar: IntechOpen.
- Carraro A, Sozzi M, Marinello F. 2023. The Segment Anything Model (SAM) for accelerating the smart

- farming revolution[J]. Smart Agricultural Technology, 6: 100367.
- Casado-García A, del-Canto A, Sanz-Saez A, Pérez-López U, Bilbao-Kareaga A, Fritschi FB, Miranda-Apodaca J, Muñoz-Rueda A, Sillero-Martínez A, Yoldi-Achalandabaso A, Lacuesta M, Heras J. 2020. LabelStoma: A tool for stomata detection based on the YOLO algorithm[J]. Computers and Electronics in Agriculture, 178: 105751.
- Chen J, Chen Jinxiu, Zhang D, Sun Y, Nanekaran YA. 2020. Using deep transfer learning for image-based plant disease identification[J]. Computers and Electronics in Agriculture, 173: 105393.
- Chen LC, Papandreou G, Schroff F, Adam H. 2017. Rethinking atrous convolution for semantic image segmentation[A]. arXiv.
- Chen LC, Zhu Y, Papandreou G, Schroff F, Adam H. 2018. Encoder-decoder with atrous separable convolution for semantic image segmentation[M]. In: Ferrari V, Hebert M, Sminchisescu C, Weiss Y (editors), Computer Vision – ECCV 2018: Vol. 11211. Cham: Springer International Publishing, pp. 833-851.
- Chen M, Brun F, Raynal M, Makowski D. 2020. Delaying the first grapevine fungicide application reduces exposure on operators by half[J]. Scientific Reports, 10(1): 6404.
- Chen T, Zhu L, Ding C, Cao R, Wang Y, Li Z, Sun L, Mao P, Zang Y. 2023. SAM Fails to Segment Anything? -- SAM-Adapter: Adapting SAM in Underperformed Scenes: Camouflage, Shadow, Medical Image Segmentation, and More[A]. arXiv.
- Chen Y, Wu Q. 2022. Grape leaf disease identification with sparse data via generative adversarial networks and convolutional neural networks[J]. Precision Agriculture.
- Cooke BM, Jones DG, Kaye B. 2006. The epidemiology of plant diseases[M]. 2nd ed. Dordrecht, the Netherlands: Springer.
- Deng J, Dong W, Socher R, Li LJ, Kai Li, Li Fei-Fei. 2009. ImageNet: A large-scale hierarchical image database[C]. 2009 IEEE Conference on Computer Vision and Pattern Recognition, pp. 248-255.
- Devlin J, Chang MW, Lee K, Toutanova K. 2019. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding[A]. English: arXiv.
- Di Y, You N, Dong J, Liao X, Song K, Fu P. 2023. Recent soybean subsidy policy did not revitalize but stabilize the soybean planting areas in Northeast China[J]. European Journal of Agronomy, 147: 126841.
- Dosovitskiy A, Beyer L, Kolesnikov A, Weissenborn D, Zhai X, Unterthiner T, Dehghani M, Minderer M, Heigold G, Gelly S, Uszkoreit J, Houlsby N. 2021. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale[A]. arXiv.
- FAO.2021.New standards to curb the global spread of plant pests and diseases[N].FAO.
- Ferentinos KP. 2018. Deep learning models for plant disease detection and diagnosis[J]. Computers and Electronics in Agriculture, 145: 311-318.
- Fontaine MC, Labbé F, Dussert Y, Delière L, Richart-Cervera S, Giraud T, Delmotte F. 2020. Europe as a bridgehead in the worldwide invasion history of grapevine downy mildew, *Plasmopara viticola*[R]. Evolutionary Biology.
- Francesca S, Simona G, Francesco Nicola T, Andrea R, Vittorio R, Federico S, Cynthia R, Maria Lodovica

- G. 2006. Downy mildew (*Plasmopara viticola*) epidemics on grapevine under climate change: grapevine under climate change[J]. *Global Change Biology*, 12(7): 1299-1307.
- Fu L, Feng Y, Wu J, Liu Z, Gao F, Majeed Y, Al-Mallahi A, Zhang Q, Li R, Cui Y. 2021. Fast and accurate detection of kiwifruit in orchard using improved YOLOv3-tiny model[J]. *Precision Agriculture*, 22(3): 754-776.
- Ganesan K. 2018. ROUGE 2.0: updated and improved measures for evaluation of summarization tasks[A]. arXiv.
- Gessler C, Ilaria P, Michele P. 2011. *Plasmopara viticola*: a review of knowledge on downy mildew of grapevine and effective disease management[J]. *Phytopathologia Mediterranea*, 1(50): 3-44.
- Ghosal S, Blystone D, Singh AK, Ganapathysubramanian B, Singh A, Sarkar S. 2018. An explainable deep machine vision framework for plant stress phenotyping[J]. *Proceedings of the National Academy of Sciences*, 115(18): 4613-4618.
- Gobbin D, Jermini M, Loskill B, Pertot I, Raynal M, Gessler C. 2005. Importance of secondary inoculum of *Plasmopara viticola* to epidemics of grapevine downy mildew[J]. *Plant Pathology*, 54(4): 522-534.
- Gómez-Zeledón J, Kaiser M. 2016. An Extended Leaf Disc Test for Virulence Assessment in *Plasmopara Viticola* and Detection of Downy Mildew Resistance in *Vitis*[J]. *J Plant Pathol Microbiol*, 07(05).
- Govardhan M, M B V. 2019. Diagnosis of Tomato Plant Diseases using Random Forest[C]. 2019 Global Conference for Advancement in Technology (GCAT), pp. 1-5.
- Gutiérrez S, Hernández I, Ceballos S, Barrio I, Díez-Navajas AM, Tardaguila J. 2021. Deep learning for the differentiation of downy mildew and spider mite in grapevine under field conditions[J]. *Computers and Electronics in Agriculture*, 182: 105991.
- Hamuda E, Glavin M, Jones E. 2016. A survey of image processing techniques for plant extraction and segmentation in the field[J]. *Computers and Electronics in Agriculture*, 125: 184-199.
- Harris CR, Millman KJ, Van Der Walt SJ, Gommers R, Virtanen P, Cournapeau D, Wieser E, Taylor J, Berg S, Smith NJ, Kern R, Picus M, Hoyer S, Van Kerkwijk MH, Brett M, Haldane A, Del Río JF, Wiebe M, Peterson P, Gérard-Marchant P, Sheppard K, Reddy T, Weckesser W, Abbasi H, Gohlke C, Oliphant TE. 2020. Array programming with NumPy[J]. *Nature*, 585(7825): 357-362.
- He K, Gkioxari G, Dollár P, Girshick R. 2018. Mask R-CNN[A]. arXiv.
- He K, Zhang X, Ren S, Sun J. 2015. Deep Residual Learning for Image Recognition[A]. arXiv.
- Hein NT, Ciampitti IA, Jagadish SVK. 2021. Bottlenecks and opportunities in field-based high-throughput phenotyping for heat and drought stress[J]. *Journal of Experimental Botany*, 72(14): 5102-5116.
- Hernández I, Gutiérrez S, Ceballos S, Iñíguez R, Barrio I, Tardaguila J. 2021. Artificial Intelligence and Novel Sensing Technologies for Assessing Downy Mildew in Grapevine[J]. *Horticulturae*, 7(5): 103.
- Hernández I, Gutiérrez S, Ceballos S, Palacios F, Toffolatti SL, Maddalena G, Diago MP, Tardaguila J. 2022. Assessment of downy mildew in grapevine using computer vision and fuzzy logic. Development and validation of a new method[J]. *OENO One*, 56(3): 41-53.
- Hu EJ, Shen Y, Wallis P, Allen-Zhu Z, Li Y, Wang S, Wang L, Chen W. 2021. LoRA: low-rank adaptation

- of large language models[A]. arXiv.
- Hui W, Wei Z, Bregaglio S, Shuyi Y, Xinghong L, Yonghua L, Chunhao L, Jiye Y. 2022. Adoption of Sobol's analysis method improved the application of a coupled primary and secondary infection grape downy mildew model in northern China[J]. *Computers and Electronics in Agriculture*, 199: 107154.
- Ji GP, Fan DP, Xu P, Cheng MM, Zhou B, Gool LV. 2023. SAM Struggles in Concealed Scenes -- Empirical Study on Segment Anything[J]. *Sci China Inf Sci*, 66(12): 226101.
- Jiang M, Song L, Wang Y, Li Z, Song H. 2022. Fusion of the YOLOv4 network model and visual attention mechanism to detect low-quality young apples in a complex environment[J]. *Precision Agriculture*, 23(2): 559-577.
- Jiang Y, Li C. 2020. Convolutional Neural Networks for Image-Based High-Throughput Plant Phenotyping: A Review[J]. *Plant Phenomics*, 2020: 2020/4152816.
- Jocher G, Chaurasia A, Qiu J. 2023. Ultralytics YOLOv8[J].
- Jothiaruna N, Joseph Abraham Sundar K, Karthikeyan B. 2019. A segmentation method for disease spot images incorporating chrominance in Comprehensive Color Feature and Region Growing[J]. *Computers and Electronics in Agriculture*, 165: 104934.
- Kamilaris A. 2018. Deep learning in agriculture_ A survey[J]. *Computers and Electronics in Agriculture*: 21.
- Kataoka T, Kaneko T, Okamoto H, Hata S. 2003. Crop growth estimation system using machine vision[C]. *Proceedings 2003 IEEE/ASME International Conference on Advanced Intelligent Mechatronics (AIM 2003)*: Vol. 2, pp. b1079-b1083.
- Kirillov A, Mintun E, Ravi N, Mao H, Rolland C, Gustafson L, Xiao T, Whitehead S, Berg AC, Lo WY, Dollár P, Girshick R. 2023a. Segment Anything[A]. arXiv.
- Kirillov A, Mintun E, Ravi N, Mao H, Rolland C, Gustafson L, Xiao T, Whitehead S, Berg AC, Lo WY, Dollár P, Girshick R. 2023b. Segment anything[A]. arXiv.
- Kumar M, Chandel NS, Singh D, Rajput LS. 2023. Soybean disease detection and segmentation based on mask-RCNN algorithm[J]. *J Exp Agric Int*, 45(5): 63-72.
- Kunova A, Pizzatti C, Saracchi M, Pasquali M, Cortesi P. 2021. Grapevine Powdery Mildew: Fungicides for Its Management and Advances in Molecular Detection of Markers Associated with Resistance[J]. *Microorganisms*, 9(7): 1541.
- LeCun Y, Bengio Y, Hinton G. 2015. Deep learning[J]. *Nature*, 521(7553): 436-444.
- Lee DI, Lee JH, Jang SH, Oh SJ, Doo IC. 2023. Crop disease diagnosis with deep learning-based image captioning and object detection[J]. *Applied Sciences*, 13(5): 3148.
- Li J, Li D, Savarese S, Hoi S. 2023. BLIP-2: Bootstrapping Language-Image Pre-training with Frozen Image Encoders and Large Language Models[A]. arXiv.
- Li S, Yan Z, Guo Y, Su X, Cao Y, Jiang B, Yang F, Zhang Z, Xin D, Chen Q, Zhu R. 2022. SPM-IS: An auto-algorithm to acquire a mature soybean phenotype based on instance segmentation[J]. *The Crop Journal*, 10(5): 1412-1423.
- Li Y, Wang D, Yuan C, Li H, Hu J. 2023. Enhancing Agricultural Image Segmentation with an Agricultural

- Segment Anything Model Adapter[J]. *Sensors*, 23(18): 7884.
- Lian L, Ding Y, Ge Y, Liu S, Mao H, Li B, Pavone M, Liu MY, Darrell T, Yala A, Cui Y. 2025. Describe anything: detailed localized image and video captioning[A]. arXiv.
- Liang F, Huang Z, Wang W, He Z, En Q. 2024. Dynamic text prompt joint multimodal features for accurate plant disease image captioning[J]. *Visual Computer*.
- Lin K, Gong L, Huang Y, Liu C, Pan J. 2019. Deep Learning-Based Segmentation and Quantification of Cucumber Powdery Mildew Using Convolutional Neural Network[J]. *Frontiers in Plant Science*, 10: 155.
- Lin TY, Dollár P, Girshick R, He K, Hariharan B, Belongie S. 2017. Feature pyramid networks for object detection[A]. arXiv.
- Lin TY, Goyal P, Girshick R, He K, Dollár P. 2018. Focal loss for dense object detection[A]. arXiv.
- Lin TY, Maire M, Belongie S, Bourdev L, Girshick R, Hays J, Perona P, Ramanan D, Zitnick CL, Dollár P. 2015. Microsoft COCO: Common Objects in Context[A]. arXiv.
- Lin X, Xiang Y, Zhang L, Yang X, Yan Z, Yu L. 2023. SAMUS: Adapting Segment Anything Model for Clinically-Friendly and Generalizable Ultrasound Image Segmentation[A]. arXiv.
- Lin Y, Underhill A, Cadle-Davison L, Jimenez A, Riaz S, Jiang Y. 2024. Effective integration of vision foundational models for semantic segmentation to quantify grape foliage powdery mildew infection[C]. 2024 Anaheim, California July 28-31, 2024.
- Liu Q, Wu W, Zheng K, Tong Z, Liu J, Liu Y, Chen W, Wang Z, Shen Y. 2024. TagAlign: improving vision-language alignment with multi-tag classification[A]. arXiv.
- Liu W, Anguelov D, Erhan D, Szegedy C, Reed S, Fu CY, Berg AC. 2016. SSD: single shot MultiBox detector[M]. pp. 21-37.
- Liu W, Chen Y, Lu Z, Lu X, Wu Z, Zheng Z, Suo Y, Lan C, Yuan X. 2024. StripeRust-Pocket: A Mobile-Based Deep Learning Application for Efficient Disease Severity Assessment of Wheat Stripe Rust[J]. *Plant Phenomics*, 6: 0201.
- Liu X, Liu Z, Hu H, Chen Z, Wang Kohou, Wang Kai, Lian S. 2025. A multimodal benchmark dataset and model for crop disease diagnosis[M]. In: Leonardis A, Ricci E, Roth S, Russakovsky O, Sattler T, Varol G (editors), *Computer Vision – ECCV 2024: Vol. 15144*. Cham: Springer Nature Switzerland, pp. 157-170.
- Long J, Shelhamer E, Darrell T. 2015. Fully Convolutional Networks for Semantic Segmentation[A]. arXiv.
- Ma J, Du K, Zheng F, Zhang L, Gong Z, Sun Z. 2018. A recognition method for cucumber diseases using leaf symptom images based on deep convolutional neural network[J]. *Computers and Electronics in Agriculture*, 154: 18-24.
- Ma Z, Peng B, Yue Z, Zeng H, Pan M, Wu X, Yang J, Mai L, Guan K. 2025. Embracing large language model (LLM) technologies in hydrology research[A]. *California Digital Library (CDL)*.
- Magon G, De Rosa V, Martina M, Falchi R, Acquadro A, Barcaccia G, Portis E, Vannozzi A, De Paoli E. 2023. Boosting grapevine breeding for climate-smart viticulture: from genetic resources to predictive genomics[J]. *Frontiers in Plant Science*, 14: 1293186.
- Mazurowski MA, Dong H, Gu H, Yang J, Konz N, Zhang Y. 2023. Segment Anything Model for Medical

- Image Analysis: an Experimental Study[A]. arXiv.
- Mejias J, Kozłowski DK, Goshon J, Maier TR, Baum TJ. 2024. NemaCounter: A user-friendly software to accurately phenotype SCN cysts[A].
- Mohanty SP, Hughes DP, Salathé M. 2016. Using Deep Learning for Image-Based Plant Disease Detection[J]. *Frontiers in Plant Science*, 7: 1419.
- Morales MA, Worral H, Piche L, Adeniyi AS, Dariva F, Ramos C, Hoang K, Yan C, Flores P, Bandillo N. 2024. High-Throughput Phenotyping of Seed Quality Traits Using Imaging and Deep Learning in Dry Pea[A].
- Naseer M, Ranasinghe K, Khan S, Hayat M, Khan FS, Yang MH. 2021. Intriguing Properties of Vision Transformers[A]. arXiv.
- Ngugi LC, Abelwahab M, Abo-Zahhad M. 2021. Recent advances in image processing techniques for automated leaf pest and disease recognition-A review[J]. *Information Processing in Agriculture*, 8(1): 27-51.
- Nguyen KD, Phung TH, Cao HG. 2023. A SAM-based Solution for Hierarchical Panoptic Segmentation of Crops and Weeds Competition[A]. arXiv.
- Nutter FW, Esker PD, Netto RAC. 2006. Disease Assessment Concepts and the Advancements Made in Improving the Accuracy and Precision of Plant Disease Data[J]. *Eur J Plant Pathol*, 115(1): 95-103.
- OIV.2022.State of the World Vine and Wine Sector in 2022[M].
- Ojala T, Pietikainen M, Maenpää T. 2002. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns[J]. *IEEE Trans Pattern Anal Machine Intell*, 24(7): 971-987.
- Open AI.2023.ChatGPT[CP/OL].Open AI.<https://chat.openai.com>.
- O'Shea K, Nash R. 2015. An introduction to convolutional neural networks[A]. arXiv.
- Otsu N. 1979. A Threshold Selection Method from Gray-Level Histograms[J]. *IEEE Trans Syst, Man, Cybern*, 9(1): 62-66.
- Papineni K, Roukos S, Ward T, Zhu WJ. 2001. BLEU: a method for automatic evaluation of machine translation[C]. *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics - ACL '02*, p. 311.
- Paszke A, Gross S, Chintala S, Chanan G, Yang E, DeVito Z, Lin Z, Desmaison A, Antiga L, Lerer A. 2017. Automatic differentiation in PyTorch[J]. 4.
- Pérez-Bueno ML, Pineda M, Barón M. 2019. Phenotyping Plant Responses to Biotic Stress by Chlorophyll Fluorescence Imaging[J]. *Frontiers in Plant Science*, 10: 1135.
- Pineda M, Barón M, Pérez-Bueno ML. 2020. Thermal Imaging for Plant Stress Detection and Phenotyping[J]. *Remote Sensing*, 13(1): 68.
- Puelles M, Arbizu-Milagro J, Castillo-Ruiz FJ, Peña JM. 2024. Predictive models for grape downy mildew (*Plasmopara viticola*) as a decision support system in Mediterranean conditions[J]. *Crop Protection*, 175: 106450.
- Qiu T, Underhill A, Sapkota S, Cadle-Davidson L, Jiang Y. 2022. High throughput saliency-based quantification of grape powdery mildew at the microscopic level for disease resistance breeding[J].

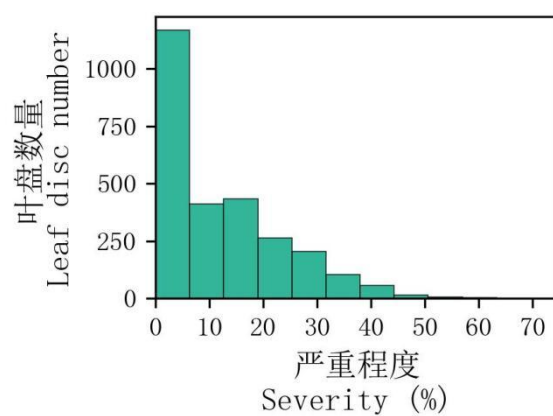
- Horticulture Research, 9: uhac187.
- Qu J, Dry I, Liu L, Guo Z, Yin L. 2021. Transcriptional profiling reveals multiple defense responses in downy mildew-resistant transgenic grapevine expressing a TIR-NBS-LRR gene located at the MrRUN1/MrRPV1 locus[J]. Horticulture Research, 8(1): 161.
- Rahman CR, Arko PS, Ali ME, Iqbal Khan MA, Apon SH, Nowrin F, Wasif A. 2020. Identification and recognition of rice diseases and pests using convolutional neural networks[J]. Biosystems Engineering, 194: 112-120.
- Ramcharan A, McCloskey P, Baranowski K, Mbilinyi N, Mrisho L, Ndalawa M, Legg J, Hughes DP. 2019. A Mobile-Based Deep Learning Model for Cassava Disease Diagnosis[J]. Front Plant Sci, 10: 272.
- Ren S, He K, Girshick R, Sun J. 2016. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks[J]. arXiv:1506.01497 [cs].
- Richard B, Qi A, Fitt BDL. 2022. Control of crop diseases through Integrated Crop Management to deliver climate-smart farming systems for low- and high-input crop production[J]. Plant Pathology, 71(1): 187-206.
- Ronneberger O, Fischer P, Brox T. 2015. U-Net: Convolutional Networks for Biomedical Image Segmentation[J]. arXiv:1505.04597 [cs].
- Rossi L, Valenti M, Legler SE, Prati A. 2022. LDD: A Grape Diseases Dataset Detection and Instance Segmentation[M]. In: Sclaroff S, Distant C, Leo M, Farinella GM, Tombari F (editors), Image Analysis and Processing – ICIAP 2022: Vol. 13232. Cham: Springer International Publishing, pp. 383-393.
- Rossi V, Caff T. 2009. Modelling the dynamics of infections caused by sexual and asexual spores during Plasmopara viticola epidemics[J]. JOURNAL OF PLANT PATHOLOGY.
- Rossi V, Caffi T, Giosuè S, Bugiani R. 2008. A mechanistic model simulating primary infections of downy mildew in grapevine[J]. Ecological Modelling, 212(3-4): 480-491.
- Rumelhart DE, Hinton GE, Williams RJ. 1986. Learning representations by back-propagating errors[J].
- Sarić R, Nguyen VD, Burge T, Berkowitz O, Trtílek M, Whelan J, Lewsey MG, Čustović E. 2022. Applications of hyperspectral imaging in plant phenotyping[J]. Trends in Plant Science, 27(3): 301-315.
- Savary S, Willocquet L, Pethybridge SJ, Esker P, McRoberts N, Nelson A. 2019. The global burden of pathogens and pests on major food crops[J]. Nature Ecology & Evolution, 3(3): 430-439.
- Schulhoff Sander, Ilie M, Balepur N, Kahadze K, Liu A, Si C, Li Y, Gupta A, Han H, Schulhoff Sevien, Dulepet PS, Vidyadhara S, Ki D, Agrawal S, Pham C, Kroiz G, Li F, Tao H, Srivastava A, Da Costa H, Gupta S, Rogers ML, Goncarenco I, Sarli G, Galynker I, Peskoff D, Carpuat M, White J, Anadkat S, Hoyle A, Resnik P. 2024. The Prompt Report: A Systematic Survey of Prompting Techniques[A]. arXiv.
- Shruthi U, Nagaveni V, Raghavendra BK. 2019. A Review on Machine Learning Classification Techniques for Plant Disease Detection[C]. 2019 5th International Conference on Advanced Computing & Communication Systems (ICACCS), pp. 281-284.
- Singh D, Jain N, Jain P, Kayal P, Kumawat S, Batra N. 2020. PlantDoc: A Dataset for Visual Plant Disease Detection[J]. Proceedings of the 7th ACM IKDD CoDS and 25th COMAD: 249-253.

- Song P, Wang J, Guo X, Yang W, Zhao C. 2021. High-throughput phenotyping: Breaking through the bottleneck in future crop breeding[J]. *The Crop Journal*, 9(3): 633-645.
- Spencer-Phillips PTN, Gisi U, Lebeda A. 2002. *Advances in Downy Mildew Research*[M]. 1st ed. 2002. Dordrecht: Springer Netherlands.
- Štambuk P, Šikuten I, Preiner D, Nimac A, Lazarević B, Marković Z, Maletić E, Kontić JK, Tomaz I. 2021. Screening of Croatian Native Grapevine Varieties for Susceptibility to *Plasmopara viticola* Using Leaf Disc Bioassay, Chlorophyll Fluorescence, and Multispectral Imaging[J]. *Plants*, 10(4): 661.
- Strange RN, Scott PR. 2005. Plant Disease: A Threat to Global Food Security[J]. *Annual Review of Phytopathology*, 43(1): 83-116.
- Sun W, Wang C, Wu H, Miao Y, Zhu H, Guo W, Li J. 2023. DFYOLOv5m-M2transformer: interpretation of vegetable disease recognition results using image dense captioning techniques[J]. *Comput Electron Agric*, 215: 108460.
- Sun X, Liu J, Shen HT, Zhu X, Hu P. 2024. On Efficient Variants of Segment Anything Model: A Survey[A]. arXiv.
- Team K, Du A, Yin B, Xing B, Qu B, Wang B, Chen C, Zhang C, Du C, Wei C, Wang C, Zhang D, Du D, Wang D, Yuan E, Lu E, Li F, Sung F, Wei G, Lai G, Zhu H, Ding H, Hu H, Yang H, Zhang H, Wu H, Yao H, Lu H, Wang H, Gao H, Zheng H, Li J, Su J, Wang Jianzhou, Deng J, Qiu J, Xie J, Wang Jinhong, Liu J, Yan J, Ouyang K, Chen L, Sui L, Yu L, Dong Mengfan, Dong Mengnan, Xu N, Cheng P, Gu Q, Zhou R, Liu S, Cao S, Yu T, Song T, Bai T, Song W, He W, Huang W, Xu W, Yuan X, Yao X, Wu X, Li X, Zu X, Zhou X, Wang X, Charles Y, Zhong Y, Li Y, Hu Y, Chen Yanru, Wang Yejie, Liu Yibo, Miao Y, Qin Y, Chen Yimin, Bao Y, Wang Yiqin, Kang Y, Liu Yuanxin, Dong Y, Du Y, Wu Y, Wang Yuzhi, Yan Y, Zhou Z, Li Z, Jiang Z, Zhang Z, Yang Z, Huang Zhiqi, Huang Zihao, Zhao Z, Chen Z, Lin Z. 2025. Kimi-VL Technical Report[A]. arXiv.
- Tian Q, Zhao G, Yan C, Yao L, Qu J, Yin L, Feng H, Yao N, Yu Q. 2024. Enhancing practicality of deep learning for crop disease identification under field conditions: insights from model evaluation and crop-specific approaches[J]. *Pest Management Science*, 80(11): 5864-5875.
- Tong S, Liu Z, Zhai Y, Ma Y, LeCun Y, Xie S. 2024. Eyes wide shut? Exploring the visual shortcomings of multimodal LLMs[C]. 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 9568-9578.
- Too EC. 2019. A comparative study of fine-tuning deep learning models for plant disease identification[J]. *Computers and Electronics in Agriculture*, 161: 272-279.
- Tsai TH, Lee YC. 2020. A Light-Weight Neural Network for Wafer Map Classification Based on Data Augmentation[J]. *IEEE Trans Semicond Manufact*, 33(4): 663-672.
- Tu S, Pang J, Liu H, Zhuang N, Chen Y, Zheng C, Wan H, Xue Y. 2020. Passion fruit detection and counting based on multiple scale faster R-CNN using RGB-D images[J]. *Precision Agriculture*, 21(5): 1072-1091.
- Uygun T, Ozguven MM. 2024. Determination of tomato leafminer: *Tuta absoluta* (Meyrick) (Lepidoptera: Gelechiidae) damage on tomato using deep learning instance segmentation method[J]. *European Food Research and Technology*, 250(6): 1837-1852.
- Van Leeuwen C, Sgubin G, Bois B, Ollat N, Swingedouw D, Zito S, Gambetta GA. 2024. Climate change

- impacts and adaptations of wine production[J]. *Nat Rev Earth Environ*, 5(4): 258-275.
- Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser L, Polosukhin I. 2017. Attention Is All You Need[J]. *arXiv:1706.03762 [cs]*.
- Velasquez-Camacho L, Otero M, Basile B, Pijuan J, Corrado G. 2022. Current Trends and Perspectives on Predictive Models for Mildew Diseases in Vineyards[J]. *Microorganisms*, 11(1): 73.
- Vishnoi VK, Kumar K, Kumar B. 2021. Plant disease detection using computational intelligence and image processing[J]. *Journal of Plant Diseases and Protection*, 128(1): 19-53.
- Volpi I, Guidotti D, Mammini M, Marchi S. 2021. Predicting symptoms of downy mildew, powdery mildew, and gray mold diseases of grapevine through machine learning[J]. *IJA*(2): 57-69.
- Waghmare H, Kokare R, Dandawate Y. 2016. Detection and classification of diseases of Grape plant using opposite colour Local Binary Pattern feature and machine learning for automated Decision Support System[C]. 2016 3rd International Conference on Signal Processing and Integrated Networks (SPIN), pp. 513-518.
- Wang C, Du P, Wu H, Li J, Zhao C, Zhu H. 2021. A cucumber leaf disease severity classification method based on the fusion of DeepLabV3+ and U-Net[J]. *Computers and Electronics in Agriculture*, 189: 106373.
- Wang H, Li G, Ma Z, Li X. 2012. Image recognition of plant diseases based on principal component analysis and neural networks[C]. 2012 8th International Conference on Natural Computation, pp. 246-251.
- Wang S, Zeng Q, Ni W, Cheng C, Wang Y. 2023. ODP-transformer: interpretation of pest classification results using image caption generation techniques[J]. *Computers and Electronics in Agriculture*, 209: 107863.
- Williams D, Macfarlane F, Britten A. 2024. Leaf only SAM: A segment anything pipeline for zero-shot automated leaf segmentation[J]. *Smart Agricultural Technology*, 8: 100515.
- Wolf T, Debut L, Sanh V, Chaumond J, Delangue C, Moi A, Cistac P, Rault T, Louf R, Funtowicz M, Davison J, Shleifer S, Platen P von, Ma C, Jernite Y, Plu J, Xu C, Scao TL, Gugger S, Drame M, Lhoest Q, Rush AM. 2020. HuggingFace's transformers: state-of-the-art natural language processing[A]. *arXiv*.
- Wu Q, Zhang Y, Elbatel M. 2023. Self-prompting large vision models for few-shot medical image segmentation[A]. *arXiv*.
- Wu Z, Chen X, Pan Z, Liu X, Liu W, Dai D, Gao H, Ma Y, Wu C, Wang B, Xie Z, Wu Y, Hu K, Wang J, Sun Y, Li Y, Piao Y, Guan K, Liu A, Xie X, You Y, Dong K, Yu X, Zhang H, Zhao L, Wang Y, Ruan C. 2024. DeepSeek-VL2: mixture-of-experts vision-language models for advanced multimodal understanding[A]. *arXiv*.
- Xie Y, Dong X, Zhao K, Sirishantha GMAD, Xiao Y, Yu P, Zhai C, Wang Q. 2025. Plant disease phenotype captioning via zero-shot learning with semantic correction based on llm[A]. *SSRN*.
- Xu S, Shen J, Wei Y, Li Y, He Y, Hu H, Feng X. 2024. Automatic plant phenotyping analysis of melon (*cucumis melo* L.) germplasm resources using deep learning methods and computer vision[J]. *Plant Methods*, 20(1): 166.
- Yan C, Liang Z, Cheng H, Li S, Yang G, Li Z, Yin L, Qu J, Wang J, Wu G, Tian Q, Yu Q, Zhao G. 2025. CDIP-ChatGLM3: a dual-model approach integrating computer vision and language modeling for

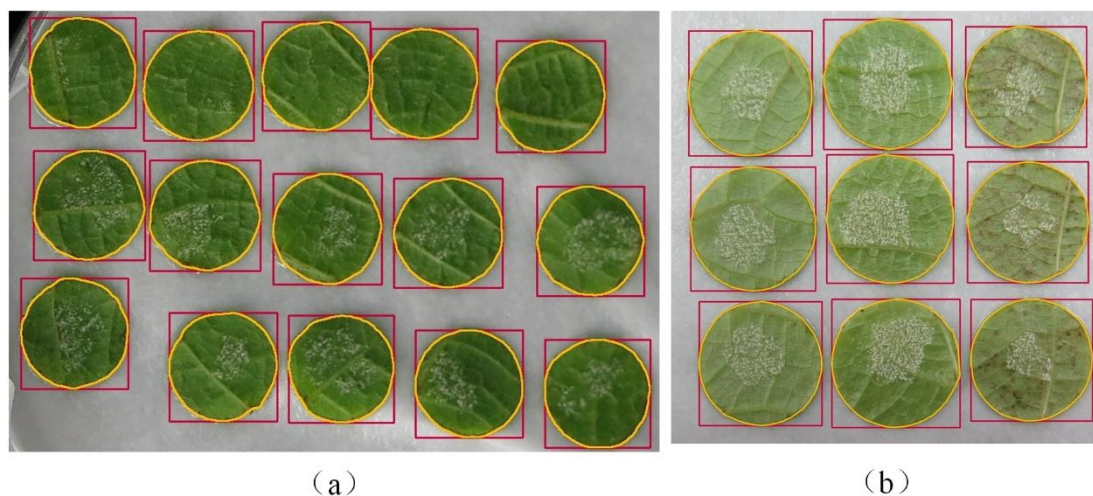
- crop disease identification and prescription[J]. *Comput Electron Agric*, 236: 110442.
- Yang W, Feng H, Zhang X, Zhang J, Doonan JH, Batchelor WD, Xiong L, Yan J. 2020. Crop Phenomics and High-Throughput Phenotyping: Past Decades, Current Challenges, and Future Perspectives[J]. *Molecular Plant*, 13(2): 187-214.
- Yang Xiao, Bahadur Bist R, Paneru B, Liu T, Applegate T, Ritz C, Kim W, Regmi P, Chai L. 2024. Computer Vision-Based cybernetics systems for promoting modern poultry Farming: A critical review[J]. *Computers and Electronics in Agriculture*, 225: 109339.
- Yang Xin, Lu X, Xie P, Guo Z, Fang H, Fu H, Sun Z, Cen H. 2024. PanicleNeRF: low-cost, high-precision in-field phenotyping of rice panicles with smartphone[J]. *Plant Phenomics*, 6: 279.
- Zendler D, Malagol N, Schwandner A, Töpfer R, Hausmann L, Zyprian E. 2021. High-throughput phenotyping of leaf discs infected with grapevine downy mildew using shallow convolutional neural networks[R]. *Plant Biology*.
- Zhang C, Han D, Qiao Y, Kim JU, Bae SH, Lee S, Hong CS. 2023a. Faster segment anything: towards lightweight SAM for mobile applications[A]. *arXiv*.
- Zhang C, Han D, Zheng S, Choi J, Kim TH, Hong CS. 2023b. MobileSAMv2: Faster Segment Anything to Everything[A]. *arXiv*.
- Zhang G, Cao H, Jin Y, Zhong Y, Zhao A, Zou X, Wang H. 2024. YOLOv8n-DDA-SAM: Accurate Cutting-Point Estimation for Robotic Cherry-Tomato Harvesting[J]. *Agriculture*, 14(7): 1011.
- Zhang K, Ma L, Cui B, Li X, Zhang B, Xie N. 2024. Visual large language model for wheat disease diagnosis in the wild[J]. *Computers and Electronics in Agriculture*, 227: 109587.
- Zhang W, Dang LM, Nguyen LQ, Alam N, Bui ND, Park HY, Moon H. 2024. Adapting the Segment Anything Model for Plant Recognition and Automated Phenotypic Parameter Measurement[J]. *Horticulturae*, 10(4): 398.
- Zhang X, Li F, Zheng H, Mu W. 2024. UPFormer: U-sharped perception lightweight transformer for segmentation of field grape leaf diseases[J]. *Expert Systems with Applications*, 249: 123546.
- Zhao K, Wu X, Xiao Y, Jiang S, Yu P, Wang Y, Wang Q. 2024. PlanText: gradually masked guidance to align image phenotypes with trait descriptions for plant disease texts[J]. *Plant Phenomics*, 6: 272.
- Zheng Y, Zhang R, Zhang J, Ye Y, Luo Z, Feng Z, Ma Y. 2024. LlamaFactory: unified efficient fine-tuning of 100+ language models[A]. *arXiv*.
- Zhong F, Chen Z, Zhang Y, Xia F. 2020. Zero- and few-shot learning for diseases recognition of citrus aurantium L. using conditional adversarial autoencoders[J]. *Comput Electron Agric*, 179: 105828.
- Zhong Z, Tang Z, He T, Fang H, Yuan C. 2024. Convolution Meets LoRA: Parameter Efficient Finetuning for Segment Anything Model[A]. *arXiv*.
- Zhou C, Li X, Loy CC, Dai B. 2024. EdgeSAM: Prompt-In-the-Loop Distillation for On-Device Deployment of SAM[A]. *arXiv*.
- Zhou J, Li J, Wang C, Wu H, Zhao C, Teng G. 2021. Crop disease identification and interpretation method based on multimodal deep learning[J]. *Computers and Electronics in Agriculture*, 189: 106408.
- Zhou Z, Siddiquee MMR, Tajbakhsh N, Liang J. 2018. UNet++: a nested U-net architecture for medical image segmentation[A]. *arXiv*.

附录



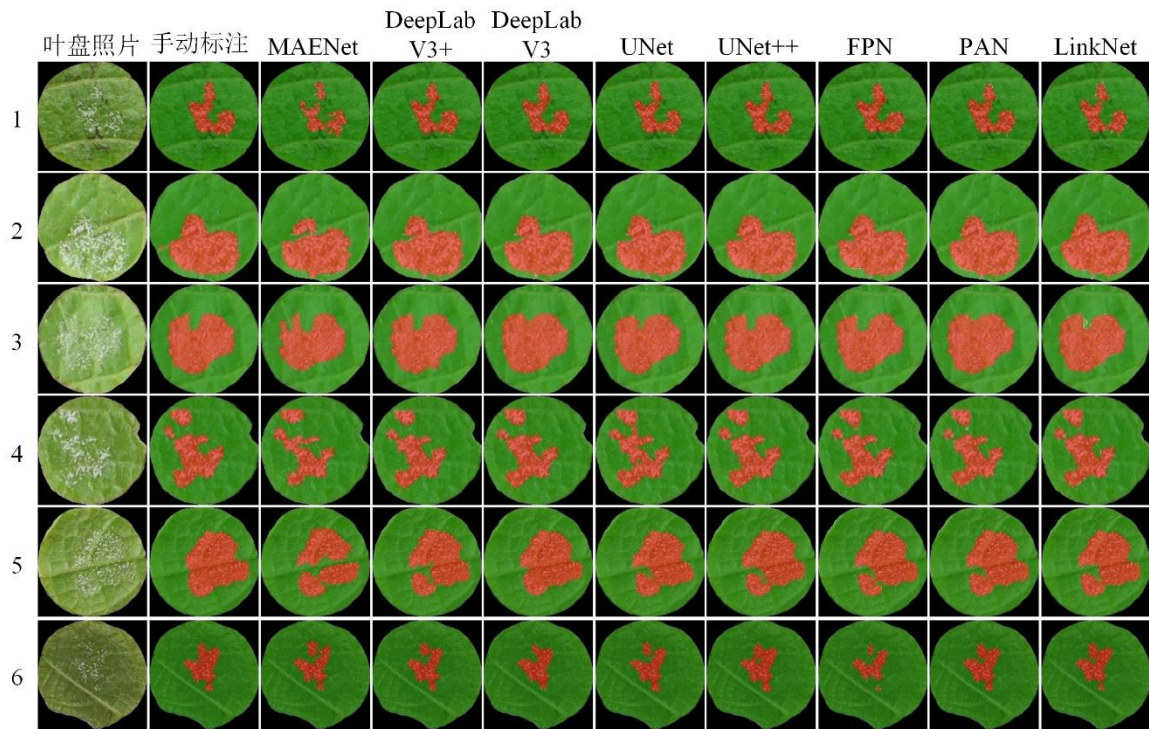
附图 1 感染霜霉病的叶片圆片按感染严重程度分类的数量。

Fig.S 3-1 The number of leaf discs with varying severity infected by downy mildew.



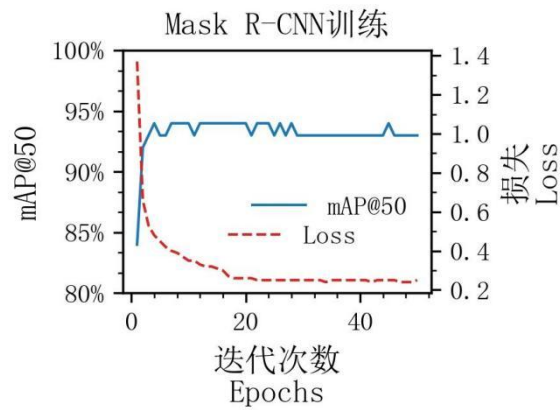
附图 2 由 Mobile SAM 生成的叶片圆片轮廓示例。红色框为 YOLO 生成的边界框，黄色轮廓由 Mobile SAM 提取。

Fig.S 3-2 Examples of leaf disc contours generated by Mobile SAM. Red boxes represents the bounding box generated by YOLO, and yellow contours are generated by Mobile SAM.



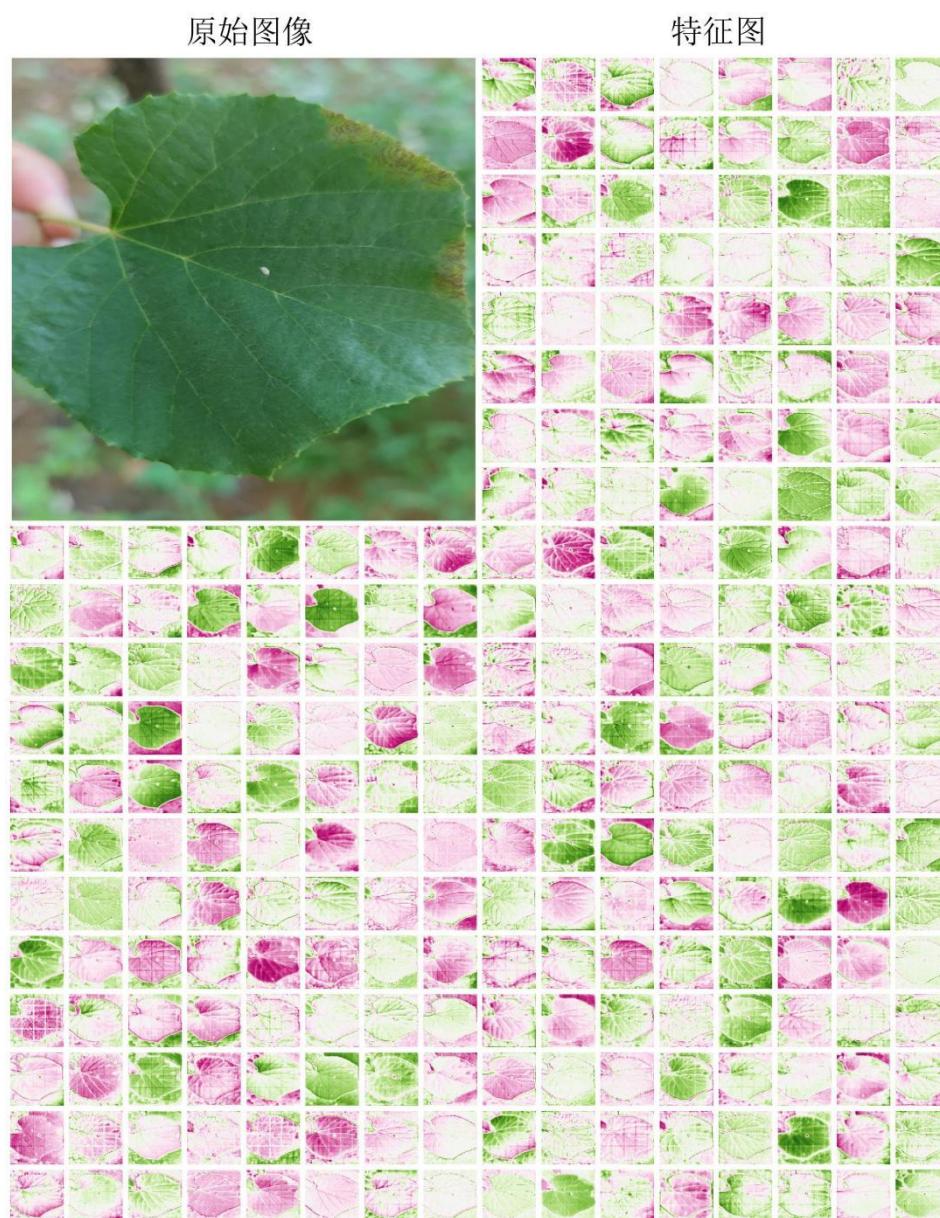
附图 3 手动标注与模型预测之间存在轻微差异的示例。每一行表示一个叶片圆片实例，每一列表示对应分割模型的预测结果。

Fig.S 3-3 Examples with slight difference between manual annotation and models' prediction. each row represents an instance of leaf disc, and each column represents the prediction of each segmentation model.



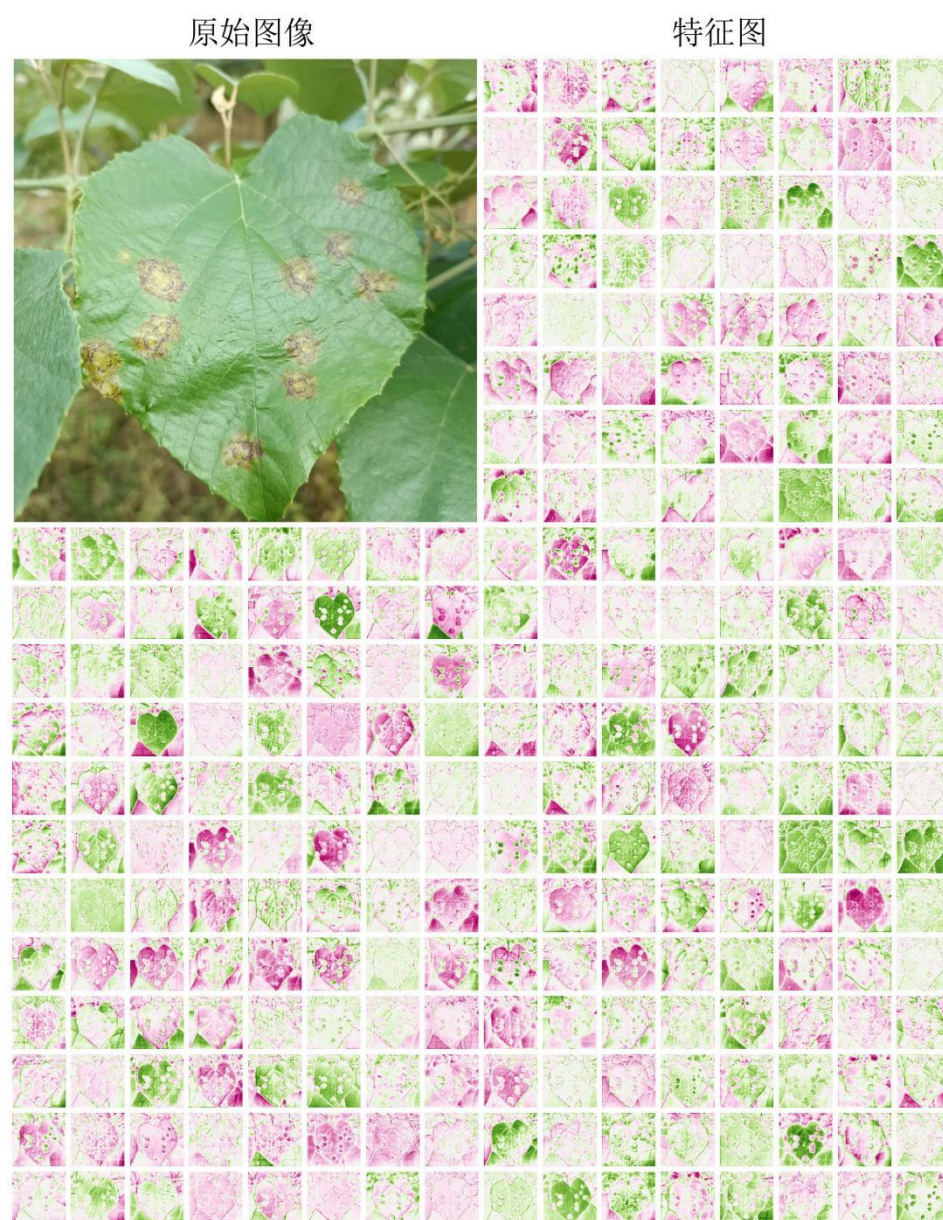
附图 4 Mask R-CNN 训练过程中 mAP@50 和损失值的变化情况。

Fig.S 3-4 The mAP@50 and loss during training process of Mask RCNN.



附图 5 由 Mobile SAM 得到的图像嵌入的特征图，可以看出不能提取到病斑的信息

Fig.S 3-5 The feature map of image embeddings generated by Mobile SAM shows that lesion information cannot be effectively extracted.



附图 6 由 Mobile SAM 得到的图像嵌入的特征图，可以看明显的病斑的信息

Fig.S 3-6 The feature map of image embeddings generated by Mobile SAM clearly shows the lesion information.

附表 1 用于分析不同病害特征准确率的提示模板以及回复结构

Table S 1 Prompt templates and response structures for analyzing the accuracy of different disease feature detections.

提示词:

从病斑颜色、病斑形状、病斑位置和病斑大小四个方面比较以下两个句子,如果颜色描述类似,则返回 1,否则返回 0,其他三个方面同理,缺少描述则返回 0,最终得到类似[1,1,0,1]四个值。句子 1 为:{caption},句子 2 为:{prediction}。理由放在前面,最后的四个值,并用[]输出。

标注:

This image displays a grape leaf with necrotic spots that are dark brown in color, irregular in shape, primarily located in the central and marginal areas of the leaf, and cover a small area, concentrated in a few specific regions of the leaf.

预测:

The image shows a single, large, green leaf with a slightly irregular shape. The leaf has a prominent central vein and several smaller veins branching out. There are several dark brown spots scattered across the surface of the leaf, with one larger spot near the top left and another smaller spot towards the bottom right. The leaf's edges are slightly wavy and uneven.

理由:

1. 病斑颜色: 句子 1 中提到病斑是“dark brown”, 句子 2 中提到病斑是“dark brown”, 描述类似, 因此返回 1。
2. 病斑形状: 句子 1 中提到病斑是“irregular in shape”, 句子 2 中提到病斑是“slightly irregular”, 虽然表述略有差异, 但都提到病斑形状不规则, 因此返回 1。
3. 病斑位置: 句子 1 中提到病斑“primarily located in the central and marginal areas of the leaf”, 而句子 2 中提到病斑“scattered across the surface of the leaf, with one larger spot near the top left and another smaller spot towards the bottom right”, 句子 1 强调病斑主要集中在中央和边缘区域, 而句子 2 描述病斑分布在叶片表面, 且具体位置不同, 因此返回 0。
4. 病斑大小: 句子 1 中提到病斑“cover a small area, concentrated in a few specific regions of the leaf”, 句子 2 中提到病斑有“one larger spot near the top left and another smaller spot towards the bottom right”, 虽然句子 1 没有明确提到单个病斑的大小, 但从“small area”可以推断病斑整体较小, 而句子 2 提到有“larger spot”和“smaller spot”, 说明病斑大小有明显差异, 因此返回 1。

结果:

[1, 1, 0, 1]

致谢

在博士论文的章节中，最有感情的莫过于致谢部分。心里构思了很多的话，在键盘前却一再停顿。在西农学习和生活的十年里，经历了许多相聚与离别，此刻却轮到我。何老师说读博是人生的修炼，如今深有体会，回望过往，心中唯有感激。

感谢我的博士导师于强研究员。于老师胸怀宽广，让人如沐春风。每一次交流，都能感受到亦师亦友的温暖与快乐。于老师总能以开阔的视野审视科学问题，鼓励我们自上而下地思考和写作。在他的影响下，我学会了倾听与理解，这将使我一生受益匪浅。

感谢我的博士指导老师赵刚研究员。赵老师如于老师般为学生着想，积极的帮助大家成长和进步。每次和赵老师散步的时候，边走边聊，总是最放松的时刻。在他的指导下，我学会了论文写作与修改，更加思考如何将研究与田间实践结合，以及如何高效交流与沟通。赵老师还为我们提供了舒适的办公环境和英文学习氛围，带我们一起打球锻炼，这些都让我受益无穷。

感谢我的硕士生导师岳超教授。岳老师认真严谨，辛勤工作，高标准要求自己。硕士期间，我深入理解了操作系统、编程思想、数据分析以及搜索技巧，也慢慢体会到批判性思维的重要性。

感谢何建强教授、李毅研究员、吴耕泓教授、冯璞玉教授和石浩研究员，帮助我指正了博士论文中的内容、逻辑、结构、语病和格式等问题。我感觉到很羞愧，因为我在预答辩前还认真的修改了论文结构和格式，结果错误却是最多的。

感谢广西农业科学院的尹玲研究员和曲俊杰研究员。广西项目的支持，让我得以采集完成论文所需的数据，在项目困难时期他们也给予了大量的帮助和包容。感谢韦淑梅在日常工作辛勤付出。

感谢广西壮族自治区都安瑶族自治县安阳镇梁耀康老板和韦春妮女士，在广西的三个月时光，是一段难忘的经历。

感谢山东科技大学的闫长青教授，以及梁泽芸、李馨、李书洋、蔺鑫、黄振鹏和李志伟同学等。

感谢何亮、靳宁、刘准桥、姚宁、他志杰、庄伟、张亚捷（男）、何杰、刘亚东、郑磊、蒋腾聪以及和志豪等师兄，在生活与工作中的支持与帮助。

特别感谢赵杰师兄和梁楚涛师兄。赵杰师兄仗义真诚，在很多时候总会给我信心和鼓励，让我感觉到信任的力量。梁楚涛师兄知识面广，无论是科研上的问题，还是生活中的琐事，我们几乎无话不谈。

感谢张艳、李林超、何沁锬、施宇、戎亮秉、贾倩兰、弓开元、任凯、康晓凤、王雪彤、李鹏、姚林佳、谭姣姣、牛文浩、陈森、陈彬、王猛帅等。感谢 Benaly Mohanmod Amine，来华交流期间，我的英语听力和口语得到了明显提升。感谢汤显辉、赵洪飞、杨怡、刘锦月、侯鑫、马慧、叶俊志、吴婕、孟怡凡、李昱、王乐乐、王家铭等。感谢杨晶晶、王亚凯、郭晨辉。感谢同窗好友徐梦洋、刘斌斌、陈晓凯、温志豪、汤媛媛等。正是和大家不断地交流过程中，我增长了知识和技能，以及感受到合作的重要性。

感谢本科同学和舍友彭子恒、崔保中、王浩然、张桠婕（女）、骆琛、闫通、孙成章、张宇航、赵震霆、土哥（李广垚）。感谢孟梓安、陈祥、王旭鹏。感谢刘海江。感谢好朋友李焱鑫、杨首浩、田亚豪、王鹏伟、王毅。感谢陈晨、袁佳阳、李杨、申佳宝。感谢宋志阳。感谢杨丹、武晓宇。感谢邢凯、任晓东、张敬奇。感谢广西大学刘文清博士，在广西期间对我的照顾和帮助。

感谢我的父母田杏林和宋秀平，以及妹妹田晓雪，父母的默默支持和辛苦付出，以及家庭的温馨，为我提供了巨大的支持 and 安全感。感谢我的亲戚家人，对我父母的帮助和照顾，让我在上大学也能有一份安心。感谢田红兵大伯在我升学时提供的指导。

十年时间，经历了很多幸福的时刻。2019 年和赵杰去甘南出差，回兰州的路上，看见漫山大雪的欣喜，以及跟着岳老师和赵杰在洋县看油菜花和大熊猫的回忆。2020 年，与孟总、老王、陈祥一起去珠三角游玩，320 块钱从佛山打车到深圳经过虎门大桥的激动。在北京，和土哥、王毅一起参观清北校园合影留念。2022 年世界杯，克罗地亚补时绝平巴西的感动。和梁楚涛师兄骑行关中环线、在广西红水河里畅游、疫情期间过年没回家在 413 包饺子、汤显辉师兄给我讲生态学实验原理、设计和理论推导，和赵洪飞、杨怡、侯鑫、马慧、吴婕在红河谷登山，以及和刘博等登顶冰晶顶。2024 年，跟着于老师去青海贵德参观，跟着赵老师去芜湖、长沙水稻基地，以及和彭子恒自驾去银川找保中等。此外，每一次在渭河边散步，每一次回家与朋友聚餐，每一次叶俊志从西安过来杨凌我们散步聊天时，每一个好友的婚礼，每一年送别毕业的师兄师姐，都是快乐而珍贵的片段……这些美好的记忆为焦虑的博士生活带来了诸多安慰。

时间过得非常快，有幸在这几年，认识很多厉害且善良的老师 and 同门，你们是我学习的榜样，真心的祝愿大家越来越幸福。尽管博士阶段的学习已经结束，但这段过程却令我受益终身，我将带着这份感恩继续前行。

田琪
2025 年立秋

个人简历

田琪，男，河南安阳人。本科毕业于西北农林科技大学创新实验学院和资源环境学院，指导老师为朱求安教授。硕士保送至西北农林科技大学，指导老师为岳超研究员，开发了森林火线扩散算法。博士毕业于资源环境学院，指导老师为水土保持与工程学院于强研究员和赵刚研究员，主要方向为作物病害表型。

1. 主要学习经历:

2019–2025 硕博 西北农林科技大学 资源环境学院 土地资源与空间信息技术

2015–2019 本科 西北农林科技大学 创新实验学院和资源环境学院 资源环境科学

2. 论文发表

[1] **Qi Tian**; Gang Zhao*; Changqing Yan; ...; Qiang Yu*. Enhancing practicality of deep learning for crop disease identification under field conditions: insights from model evaluation and crop-specific approaches. *Pest Management Science* 2024.

[2] **Qi Tian**; Junjie Qu; Ling Yin; Shumei Wei; ...; Jing Wang; Qiang Yu; Gang Zhao*. SporaScan: An Automated Pipeline for Precise Disc-Based Disease Severity Assessment in Grapevine Downy Mildew. *Scientia Horticulturae* (under review)

[3] Yuanyuan Tang; Xiangyun Yang; Xinru Wang; ...; **Qi Tian**; Yanbin Qi*. Soil Inorganic Carbon Losses Counteracted Soil Organic Carbon Increases in Deeper Soil Layers Over 30 Years in North China. *Land* 2025

[4] Peng Li; Liang He; Xuotong Wang; ...; **Qi Tian**; ...; Qiang Yu*. The Forecasting Yield of Highland Barley and Wheat by Combining a Crop Model with Different Weather Fusion Methods in the Study of the Northeastern Tibetan Plateau. *Atmosphere* 2025.

[5] Changqing Yan; Zeyun Liang; Han Cheng; ...; **Qi Tian**; ...; Gang Zhao*. CDIP-ChatGLM3: A dual-model approach integrating computer vision and language modeling for crop disease identification and prescription. *Computers and Electronics in Agriculture* 2025.

[6] Bin Chen; Gang Zhao*; **Qi Tian**; ...; Qiang Yu*. Climate-driven shifts in suitable areas of *Alternaria* leaf blotch (*Alternaria mali* Roberts) on apples: Projections and uncertainty analysis in China. *Agricultural and Forest Meteorology* 2025.

[7] Wenhao Niu; Lan Luo; Yu Shi; ...; **Qi Tian**; ...; Qiang Yu. Impacts of “One Household One Plot” and “One Village Group One Plot” fragmentation consolidation models on cultivated land use transition from perspective of human-land system. *Habitat International* 2024.

[8] Xiaofeng Kang; Dingrong Wu*; Jiaojiao Tan; ...; **Qi Tian**; Qiang Yu*. Performance of nine maize phenology models in China under historical climate change conditions. *Agricultural and Forest Meteorology* 2024.

[9] Changqing Yan; Zeyun Liang; Lin Yin; ...; **Qi Tian**; ...; Junjie Qu*. AFM-YOLOv8s: An Accurate, Fast, and Highly Robust Model for Detection of Sporangia of *Plasmopara viticola* with Various Morphological Variants. *Plant Phenomics* 2024.

[10] Jiaojiao Tan; Gang Zhao*; **Qi Tian**; ...; Qiang Yu*. Overcoming mechanistic limitations of process-based phenological models: A data clustering method for large-scale applications. *Agricultural and Forest Meteorology* 2024.

[11] Junzhi Ye; Yunfeng Hu*; Zhiming Feng; ...; **Qi Tian**; Yunzhi Zhang. Monitoring of Cropland Abandonment and Land Reclamation in the Farming-Pastoral Zone of Northern China. *Remote Sensing* 2024.

[12] Ziheng Peng; Yu Liu; Jiejun Qi; ...; **Qi Tian**; ...; Shuo Jiao*. The climate driven distribution and response to global change of soil-borne pathogens in agroecosystems. *Global Ecology and Biogeography* 2023.

[13] Jie Zhao; Jiaming Wang; Yifan Meng; ...; **Qi Tian**; ...; Chao Yue*. Spatiotemporal patterns of fire-driven forest mortality in China. *Forest Ecology and Management* 2023.

[14] Qinsi He; De Li Liu; Bin Wang; ...; **Qi Tian**; ...; Qiang Yu*. Identifying effective agricultural management practices for climate change adaptation and mitigation: A win-win strategy in South-Eastern Australia. *Agricultural Systems* 2022.

[15] Jie Zhao; Liang Wang; Xin Hou; ...; **Qi Tian**; ...; Chao Yue*. Fire Regime Impacts on Postfire Diurnal Land Surface Temperature Change Over North American Boreal Forest. *JGR: Atmospheres* 2021.

[16] 康晓凤, 郭定荣*, 田琪,..., 于强*. 玉米发育模式在我国主要产区的适用性比较研究. *中国生态农业学报* 2022

3. 合作在审论文

[1] Linjia Yao; Gang Zhao*; Changqing Yan; ...; **Qi Tian**; ...; Qiang Yu*. Information Sources and Technology Adoption in Apple Production in China and Its Implications. 2025

[2] Linjia Yao; Genghong Wu; **Qi Tian**; ...; Qiang Yu*; Gang Zhao*. Survey Data on Apple Farming in China: Agronomic Management, Advisory Channels, and Profitability. 2025

[3] Mengyang Xu; Zhang, Pengyi; He, Junhao; Chao Yue*; ...; **Qi Tian**; ...; Wang, Lele. Critical role of land management in augmenting China's terrestrial carbon sink up to 2060. 2025

[4] Wei Zhuang; Ying-Ping Wang*; ...; **Qi Tian**; Qiang Yu*. Mineral Adsorption in a Microbial-Explicit Soil Carbon Model and Insights for Soil Carbon Stabilization. 2025

[5] Liangbing Rong; Yiwen Zhou; Fengying Duan; ...; **Qi Tian**; ...; Qiang Yu*; Wenbin Zhou*. Optimizing site-specific management practices to achieve agricultural sustainability across China. 2025

[6] Xuetong Wang; Liang He*; Peng Li; ...; **Qi Tian**; ...; Qiang Yu*. Spatially adaptive estimation of multi-layer soil temperature at a daily time-step across China during 2010–2020. *Earth System Science Data* preprint. 2025

[7] Yuanyuan Tang, Xiangyun Yang, Xinru Wang, ..., Qi Tian, Yanbing Qi*. Monitoring vertical soil carbon storage variation based on GWRK model and environmental covariates in North China

4. 学术交流

(1) 生态模拟与智慧农业学术研讨会暨第四届“植被动力学”暑期学校会议

5. 其他

- (1) 基于野火行为计算机模拟的林草火灾风险预警系统 2021R11L2159454
- (2) 中国国际大学生创新大赛 2024 陕西赛区省级复赛 银奖 第三
- (3) 农业环境监测数据采集软件 v1.0