

结构空间引导的复杂环境芒果叶片与病斑分割方法

王丹阳^{1,2}, 梁伟红^{1,2}, 叶露^{1,2}, 李海亮^{1,2}, 郑倩^{1,2}, 黄贵修^{3*}

1. 中国热带农业科学院科技信息研究所/海南省热带作物信息技术应用研究重点实验室, 海南海口 571101; 2. 海南省唐华俊院士工作站, 海南海口 571101; 3. 中国热带农业科学院环境与植物保护研究所, 海南海口 571101

摘要: 针对田间复杂场景下芒果叶片病害识别中, 复杂背景干扰与多叶片同时出现病症导致的特征识别困难和多任务分割问题, 本研究提出了结构空间引导的 G-Mask2Former 模型, 实现对叶片实例与病斑语义的协同分割。该模型在共享特征空间内同时学习叶片和病斑分割, 并辅助叶片结构先验对病斑进行空间约束。首先, 基于 Mask2Former 的多尺度像素表征与掩码注意力提取复杂场景中芒果叶片病害有效特征, 结合构建病斑伪实例, 在单一模型中实现叶片实例和病斑语义分割, 简化模型部署和推理流程; 然后, 通过叶片结构空间软门控模块, 引导模型关注叶片区域, 以减少复杂环境干扰; 最后在复杂场景下芒果叶片病害数据集上进行试验, 并与其他模型进行对比分析。结果表明: 该模型对复杂场景芒果叶片与病害分割的综合表现优于其他传统模型。在测试集上, 实例分割平均精度 (AP) 达到 87.37%, 语义分割平均交并比 (mIoU) 达到 82.83%。该方法可为复杂环境中芒果叶片病害的诊断提供有效的量化解决方案。

关键词: 结构空间引导; 芒果; 叶片病害; 实例分割; 语义分割

中图分类号: TP391.4

文献标志码: A

Structure-space-guided Method for the Segmentation of Mango Leaves and Lesions in Complex In-field Scenes

WANG Danyang^{1,2}, LIANG Weihong^{1,2}, YE Lu^{1,2}, LI Hailiang^{1,2}, ZHENG Qian^{1,2}, HUANG Guixiu^{3*}

1. Institute of Scientific and Technical Information, Chinese Academy of Tropical Agricultural Sciences / Key Laboratory of Applied Research on Tropical Crop Information Technology of Hainan Province, Haikou, Hainan 571101, China; 2. Hainan Tang Huajun Academician Workstation, Haikou, Hainan 571101, China; 3. Environment and Plant Protection Institute, Chinese Academy of Tropical Agricultural Sciences, Haikou, Hainan 571101, China

Abstract: To address mango leaf disease recognition in complex in-field scenes, where background interference and simultaneous infections across multiple leaves lead to feature recognition difficulties and multi-task segmentation problems, this paper proposes G-Mask2Former, a structure-space-guided model that achieves collaborative segmentation of leaf instances and lesion semantics. The method jointly learns leaf and lesion segmentation within a shared feature space, and leverages a leaf-structure prior to impose spatial constraints on the lesion predictions. First, based on Mask2Former, we use multi-scale pixel representations and mask attention to extract effective features of mango leaf diseases in complex scenes; by constructing pseudo-instances of lesions, we achieve leaf instance segmentation and lesion semantic segmentation within a single model, thereby simplifying model deployment and inference. Second, through a leaf structure-space soft gating module, the model is guided to focus on leaf regions, reducing interference from complex environments. Finally, we conduct experiments on a mango leaf disease dataset collected in complex in-field scenes and perform comparative analyses against other models. The proposed method achieves superior overall performance over classical baselines for segmenting mango leaves and diseases in complex in-field scenes. On the test set, the average precision (AP) for leaf instance segmentation reaches 87.37%, while the mean Intersection-over-Union (mIoU) for lesion semantic segmentation reaches 82.83%. The proposed method would provide an effective solution for mango leaf

收稿日期 2025-10-10; 接受日期 2025-11-20

基金项目 海南省自然科学基金青年项目 (No. 323QN300); 海南省科技人才创新项目 (No. KJRC2023D40)。

作者简介 王丹阳 (1993—), 女, 硕士, 助理研究员, 研究方向: 智能病害识别。*通信作者 (Corresponding author): 黄贵修 (HUANG Guixiu), E-mail: hgxiu@vip.163.com。

disease segmentation in complex environments and facilitates quantitative disease assessment.

Keywords: structure-space guidance; mango; leaf disease; instance segmentation; semantic segmentation

DOI: 10.3969/j.issn.1000-2561.2026.02.020

芒果是热带、亚热带重要果树之一，在我国广西、广东、云南、海南等省（区）广泛种植。据农业农村部农垦局统计，2021年我国芒果种植面积约37.46万 $\text{hm}^{2[1]}$ 。芒果产业在促进地方经济发展中发挥了重要作用，而病害是芒果生产上最为严重的生物限制因素之一，严重影响芒果的产量和品质。芒果叶部病害主要有炭疽病、藻斑病等，这些病害常破坏芒果树的正常生长和发育。植物叶片通过光合作用为植物提供必要的营养和能量，其健康状况直接影响作物的生长状态、产量和品质。当植物叶片受病害侵染后，病斑迅速扩散会导致有效光合作用面积减少，从而降低了光合效率，果实减产或品质下降。因此，及时精准识别并量化叶片病害程度，对病害早期诊断、精准防治及保障作物产量具有重要意义。

传统芒果叶片病害诊断及严重度评估主要依赖人工目测，该方法依靠植保专家的专业知识和经验判断，主观性强。在大规模种植环境下，人工巡查耗时耗力、诊断效率低，难以满足现代农业快速、精准监测的需求。因此，发展高效、自动化且客观准确的叶片病害分割与识别技术已成为亟待解决的科技难题。随着图像处理和智能计算技术的快速发展，作物病害自动化诊断得到了广泛关注^[2-3]。早期的研究工作多将病害严重程度评估视为典型的图像分类任务^[4-6]。依据专家经验先将病害划分为健康、发病早期、发病中期和发病晚期（或健康、一般和严重）等类别，再利用分类模型对图像特征进行提取和识别，实现病害自动分级和诊断。然而，专家在评估严重程度时因受人为主观性影响，常会造成病害等级评估的人为偏差，如在疾病严重程度较低时，高估病害等级，而在严重程度相同的情况下，病情较小的病害通常会被高估^[7]。

为了更科学地评估病害严重程度，研究人员尝试利用图像分割技术进行病害区域的精准识别与定量分析^[8]，通过计算病斑在叶片上的面积占比，客观量化病害严重程度。图像分割将像素划分为不同类别或实例，常见任务包括语义分割、实例分割和全景分割。其中，语义分割为每个像素分配类别，但不区分同类个体；实例分割在像

素级同时预测类别和掩码以区分不同实例；全景分割则对整幅图像所有像素进行互斥划分，兼顾可数目标实例与不可数区域的语义类别。在叶片病害程度评估方面，已有研究多采用语义分割模型实现叶片和病斑区域分割。如邓阳军等^[9]基于改进的UNet模型分割水稻叶片及病斑，并根据病斑面积占比对病害程度进行分级；DENG等^[10]提出跨层注意力融合机制结合多卷积的MC-UNet模型用于番茄叶片病害分割。然而，此类数据集多采集于受控的室内环境，背景单一，因而在自然场景中的可迁移性有限，如DENG等^[10]发现简单背景数据训练的模型在实际生产基地复杂环境病害图像中会将背景部分土壤错误分割为病害。针对复杂背景病害分割问题，LI等^[11]优化DeepLabV3+模型用于复杂场景下黄瓜叶片病害的分割与严重程度估计，LU等^[12]提出轻量混合网络分割模型MixSeg，以提升复杂环境下苹果叶片和病斑的分割性能。为降低复杂背景干扰，有些研究采用分阶段策略，即先从复杂背景中提取叶片，再进行病斑分割。如WANG等^[13]提出两阶段DUNet模型，第一阶段采用DeepLabV3+分割黄瓜叶片，第二阶段使用UNet提取病斑；CHENG等^[14]分别采用PBGNet和PDFNet完成叶片与病斑分割，实现苹果叶片病害严重程度分级；XU等^[15]基于改进的DeepLabV3+设计轻量级两阶段分割模型用于田间病叶和病斑提取。此外，有些研究结合实例分割与语义分割用于复杂场景下叶片病害检测。TASSIS等^[16]先用Mask-RCNN识别咖啡叶片位置，再用UNet和PSPNet提取咖啡叶片和病斑区域。LI等^[17]设计融合实例分割、分类和语义分割模型的集成框架，先用Mask R-CNN从复杂背景中分割出马铃薯叶片实例，提取单个叶片，再通过分类和分割模型完成病害识别与分割，实现复杂场景下多片叶片病害分割与检测。

与大多数默认单叶片无遮挡的研究不同，设施农业环境中，植株自然状态下采集的图像中常出现叶片相互遮挡及多片病叶并存情况，此时仅采用语义分割缺乏实例感知，难以区分不同叶片个体，无法按叶片统计病害严重程度。为适应真实生产条件，本研究采用叶片可见病斑面积占比

作为近似指标,以描述病害严重程度的大致范围。鉴于分段式方法往往依赖多套模型,模型维护与部署复杂,本研究在单一模型中同时实现叶片实例分割与病斑语义分割。针对复杂背景干扰和多叶片同时发生病害导致的特征提取困难和多任务分割问题,在 Mask2Former 框架上提出一种结构空间引导的实例-语义分割方法,在共享特征空间内同时学习叶片实例分割和病斑语义分割,并借助叶片结构先验引导模型关注叶片区域信息,从而提升在遮挡与杂乱背景中定位与分割质量;并在田间复杂场景下芒果叶片病害数据集上对模型进行验证分析,证实方法的有效性。

1 材料与方法

1.1 数据集与标注

芒果叶片病害数据集采自海南省三亚市和乐东黎族自治县的芒果种植基地,包含芒果叶片炭疽病和藻斑病。使用开源标注工具 LabelMe^[18]标注出包含病斑的叶片和病斑,得到炭疽病标注图像 153 幅,藻斑病标注图像 154 幅。图 1 为芒果叶片病害图像和标签的可视化示例。按 7 : 2 : 1 将数据集各个病害均衡划分为训练集、验证集和测试集,得到训练集 214 幅、验证集 60 幅、测试集 33 幅,并统一裁剪成分辨率为 512 像素×512 像素的图像。为了提高模型的泛化能力,采用翻转、几何变化、亮度变化和增加高斯噪声等数据增强技术,来增强数据的多样性。通过数据增强技术,将芒果叶片病害训练集由 214 幅图像扩增至 1284 幅;验证集与测试集分别包含 60 幅与 33 幅图像。对应的叶片实例数为:训练集 2496 个、验证集 135 个、测试集 55 个。

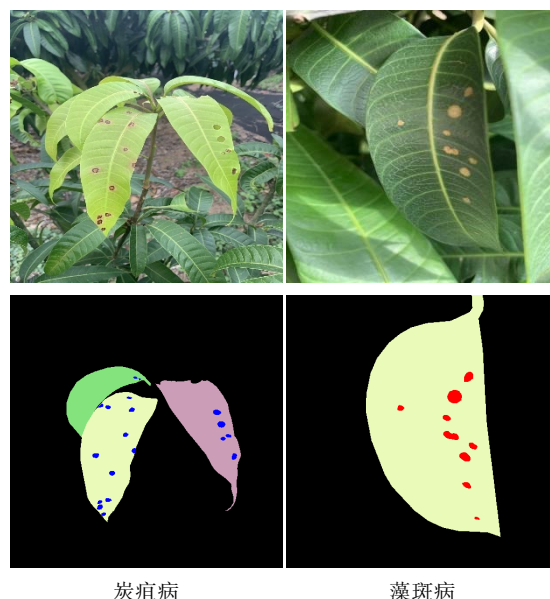


图 1 芒果叶片病害图像和标签可视化
Fig. 1 Mango leaf disease images and annotation visualization

1.2 模型框架

本研究在 Mask2Former^[19]框架基础上构建了结构空间引导的分割框架 G-Mask2Former,用于叶片实例分割和病斑语义分割任务。图 2 为 G-Mask2Former 的总体框架。为在单一架构中统一 2 个任务,我们采用共享查询空间的方法。首先,将病斑语义类构建为伪实例类标签,然后将病斑伪实例标签和叶片实例标签映射到一个合并类别集合中进行匹配计算损失。此外,通过叶片结构空间软门控模块引入“病斑应位于叶片区域内”结构空间先验,引导模型关注叶片区域。以下依次介绍 Mask2Former、多任务联合训练策略和叶片结构空间软门控模块。

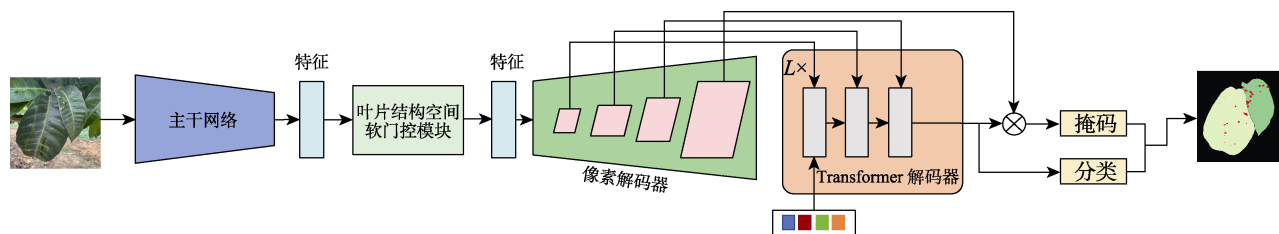


图 2 G-Mask2Former 总体框架
Fig. 2 Overall framework for G-Mask2Former

1.2.1 Mask2Former Mask2Former 是一种将语义分割、实例分割和全景分割统一为掩码分类问题的通用图像分割框架,由主干网络、像素解码器和 Transformer 解码器组成。主干特征提取器可

以基于 CNN (ResNet) 或 Transformer 模型构建,本研究使用基于 ResNet50^[20]构建的主干特征提取器。Mask2Former 默认采用多尺度可变形注意力 (MSDeformAttn) 的像素解码器,在控制计算

量的同时融合不同分辨率特征，形成多尺度特征表示，然后将同一个分辨率的特征输入到一个 Transformer 解码器层中。该设计利用高分辨率特征提升小目标与边界细节，同时借助低分辨率特征提供语义上下文，从而整体提升分割性能。

Transformer 解码器使用掩码注意力（masked attention）重点关注预测片段为中心的局部特征，与标准 Transformer 解码器中使用的关注整张特征图的交叉注意力相比，加快了收敛和提升了性能。

1.2.2 多任务联合训练策略 为在一个查询集合中同时学习叶片实例分割和病斑语义分割，本研究采用共享查询空间与统一集合预测范式。具体而言，将语义掩码转化为“伪实例”，并与叶片实例在统一类别空间中对齐后拼接为无序目标集。训练阶段在预测值与真实值之间建立一一对应关系，采用与 Mask2Former 一致的损失进行优化。该设计在单一模型中同时完成叶片实例分割和病斑语义分割，避免了为不同任务分别训练和部署多个模型，从而在一定程度上简化了部署与推理流程。

1.2.3 叶片结构空间软门控模块 在植物病害诊断中，病斑总发生在叶片区域内。为降低复杂环境的干扰，本研究引入病斑在叶片上这一结构先验知识。在训练阶段对多尺度特征图实施叶片引导的空间门控：在叶片内像素保持原有特征，叶片外像素按系数 $1-\tau$ 进行轻度抑制，从而将表达能力聚焦到叶片内部，同时保留必要上下文。

根据叶片实例构造二值并集 $U_{\text{leaf}} \in \{0,1\}^{H \times W}$ ，其中叶片内像素为 1，叶片外像素为 0。给定抑制强度 $\tau \in [0,1]$ ，生成叶片空间门控：

$$A_{\text{leaf}} = (1-\tau) + \tau U_{\text{leaf}}$$

设有多尺度特征表示为 $\{F_k\}_{k=1}^K$ ，对任意尺度特征（ F_k ），以双线性插值将 A_{leaf} 映射到对应尺度 A_{leaf}^k ，并对多尺度特征分别做逐点乘法门控：

$$F'_k = A_{\text{leaf}}^k \odot F_k$$

1.3 损失函数

本研究在训练过程中，采用与 Mask2Former 一致的损失，使用二元交叉熵损失（ L_{ce} ）和 Dice 损失（ L_{dice} ）来计算掩码损失（mask loss, L_{mask} ）： $L_{\text{mask}} = \lambda_{\text{ce}} L_{\text{ce}} + \lambda_{\text{dice}} L_{\text{dice}}$ ；使用交叉熵损失来计算分类损失（ L_{cls} ）。掩码和分类的损失组合表示为： $L = L_{\text{mask}} + \lambda_{\text{cls}} L_{\text{cls}}$ ，其中 λ 为损失权重。

1.4 试验环境及设置

试验在一台配备 CPU 为 Intel(R) Core(TM) i9-10900K CPU @3.70GHz，GPU 为 NVIDIA GeForce RTX 3090 的计算机上运行，使用 Python3.8 环境下的 PyTorch 深度学习框架。试验设置的批处理大小为 4，学习率为 0.0001，迭代次数为 5000，使用 ResNet50^[20]为模型主干网络。

1.5 评价指标

为了客观评价模型性能，我们采用平均交并比（mIoU）来评估语义分割，采用平均精度（AP）来评估实例分割。 p_{ii} 表示真实值和预测值均是类别 i 的像素总数， p_{ji} 表示预测像素类别为 j 真实像素类别为 i 的像素总数， p_{ij} 表示预测像素类别为 i 真实像素类别为 j 的像素总数。 $K+1$ 表示类别总数。交并比（IoU）是计算预测区域与真实区域的交集与并集的比值，单类 IoU 定义如下：

$$\text{IoU}_i = \frac{p_{ii}}{\sum_{j=0}^K p_{ij} + \sum_{j=0}^K p_{ji} - p_{ii}}$$

平均交并比（mIoU）是通过计算所有类 IoU 的平均值，对模型整体性能进行综合评价的指标。定义如下：

$$\text{mIoU} = \frac{1}{K+1} \sum_{i=0}^K \frac{p_{ii}}{\sum_{j=0}^K p_{ij} + \sum_{j=0}^K p_{ji} - p_{ii}}$$

实例分割任务中，AP 表示不同目标类别上的平均精度；AP₅₀ 表示 IoU 阈值为 50% 的平均精度，AP₇₅ 表示 IoU 阈值为 75% 的平均精度。

2 结果与分析

为评估 G-Mask2Former 在复杂场景下芒果叶片实例分割和病害语义分割性能，在芒果叶片病害数据集上开展一系列试验。首先对提出模块进行消融试验，并将所提模型与其他典型模型进行比较，最后得出严重程度计算量化结果的可视化。

2.1 消融试验

为评估各个模块的有效性，本研究设计了消融试验。其中，第一、第二个方案是直接使用 Mask2Former 分别进行实例分割和语义分割试验，第三个方案是在不增加参数的前提下通过构建病斑伪实例标签实现多任务联合训练，第四个方案在第三个方案的基础上加入结构空间软门控模块。通过这一系列试验，分析各个模块对平均精度（AP）、平均交并比（mIoU）和模型参数量

的影响。消融试验结果中仅实例分割任务无 mIoU 结果, 仅语义分割任务无 AP 结果。结果表明, 在总体参数量基本不变的情况下, 多任务联合训练在各项评价指标上取得了与单任务相当的性能。引入叶片空间软门控模块后, 验证集 AP 从 80.99%提升至 82.26%, mIoU 从 83.71%提升至 84.38%; 而在测试集上 AP 从 86.93%提升至

87.37%, mIoU 从 80.43%提升至 82.83% (表 1)。值得注意的是, AP₇₅ 提高了 2.96 个百分点, 表明该模块增强了边界定位能力, 而增加的参数开销可忽略不计。这些结果表明多任务联合训练与叶片结构空间软门控的有效性, 在不显著增加模型参数数量的前提下, 模型更专注叶片区域, 提升了病斑语义分割性能, 同时改善了叶片实例分割的精度。

表 1 消融试验结果
Tab. 1 Results of ablation study

试验方案 Experimental protocol	验证集 Validation set				测试集 Test set				参数量 Parameter/M
	AP/%	AP ₅₀ /%	AP ₇₅ /%	mIoU/%	AP/%	AP ₅₀ /%	AP ₇₅ /%	mIoU/%	
第一方案	79.58	83.31	79.96		86.94	90.49	87.02		43.95
第二方案				84.38				81.94	43.95
第三方案	80.99	87.43	81.32	83.71	86.93	90.20	86.15	80.43	43.95
第四方案	82.26	88.85	82.66	84.38	87.37	90.51	89.11	82.83	44.25

2.2 对比试验

所提出的 Mask2Former 模型可以在单一框架下实现叶片实例分割和病害语义分割, 为了验证所提出模型的有效性, 本研究在相同主干网络下评估所提出模型的性能, 并分别与典型的实例分割和语义分割模型进行对比。对于叶片实例分割, 使用经典实例分割模型 Mask R-CNN^[21]、SOLOv2^[22]、CondInst^[23]和 Mask2Former 模型进行分割试验。不同实例分割模型在芒果叶片病害数据集上叶片分割性能见表 2。G-Mask2Former 模型在芒果叶片病害验证集和测试集上叶片实例分割 AP 分别达到 82.26%和 87.37%。与 Mask R-CNN、SOLOv2、CondInst 和 Mask2Former 模型相比, G-Mask2Former 模型在测试集上叶片实例分割

AP 分别提高 9.52、3.77、4.00 和 0.43 个百分点。在测试集上, G-Mask2Former 模型的 AP₇₅ 相对于 Mask R-CNN、SOLOv2、CondInst 和 Mask2Former 分别提升了 9.55、4.44、5.52 和 2.09 个百分点。由于高 IoU 阈值的 AP (如 AP₇₅) 更强调掩码边界与精确定位, 该指标的增益表明, G-Mask2Former 模型在相互遮挡与边界细粒度刻画方面更具优势。与此同时, AP 与 AP₅₀ 取得不同程度提升, 体现出复杂背景多叶片同时发病场景中模型的总体匹配质量更加稳健。在模型参数量方面, G-Mask2Former 相较 Mask2Former 的模型参数量仅增加 0.30 M, 但在叶片实例分割上取得更好的整体性能。综合来看, 所提出模型在田间复杂场景下的芒果病害叶片实例分割中表现出较优的性能。

表 2 不同模型叶片实例分割试验结果
Tab. 2 Experimental results of leaf instance segmentation across models

模型 Model	验证集 Validation set			测试集 Test set			参数量 Parameter/M
	AP/%	AP ₅₀ /%	AP ₇₅ /%	AP/%	AP ₅₀ /%	AP ₇₅ /%	
Mask R-CNN	75.24	82.37	76.81	77.85	82.47	79.56	43.92
SOLOv2	75.79	81.73	78.20	83.60	87.37	84.67	46.18
CondInst	76.81	82.80	77.93	83.37	88.51	83.59	34.11
Mask2Former	79.58	83.31	79.96	86.94	90.49	87.02	43.95
G-Mask2Former	82.26	88.85	82.66	87.37	90.51	89.11	44.25

不同模型在复杂场景下芒果叶片实例分割可视化对比如图 3 所示。整体观察可见, 提出的 G-MasksFormer 模型在复杂场景中遮挡、贴边干扰条件下能更完整、准确地勾勒叶片轮廓, 与测

试集上性能领先表现一致。Mask R-CNN 模型易产生过分割和误检情况, 在未发生病害的健康叶片上常出现伪目标, 对应了其较低的 AP₇₅。SOLOv2 模型在边界刻画方面不足, 叶片边缘处

出现轮廓缺失与侵蚀, 遮挡处分割不完整。CondInst 模型同样存在结构不完整与健康叶片误识别问题。Mask2Former 整体优于上述模型, 但仍出现少量其他病害叶片误分割。相比之下, G-MasksFormer 模型在边界连续性与遮挡处实例

分割表现更稳健, 遮挡重叠的相邻叶片能被清晰分开。定量评估和定性可视化结果表明, 所提出的 G-MasksFormer 模型在复杂背景叶片相互遮挡干扰下的边界细粒度刻画与实例区分能力更强, 适用于田间复杂场景下芒果病害叶片实例分割。

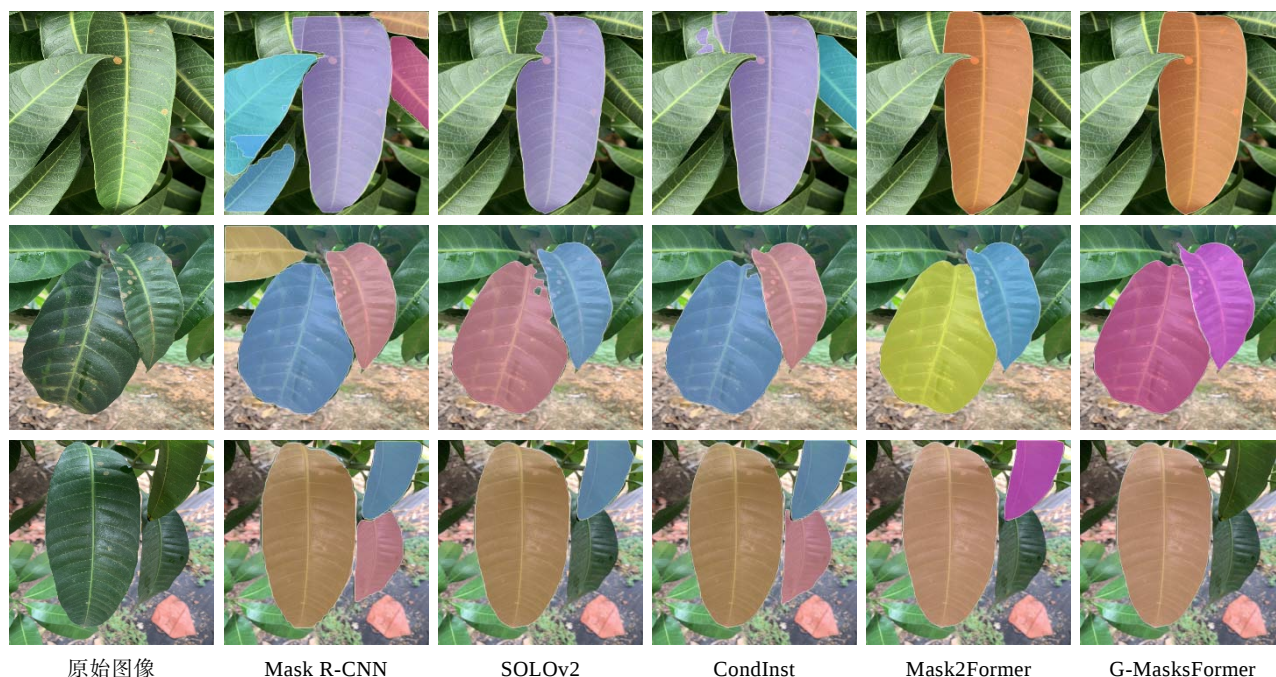


图3 不同模型叶片实例分割结果可视化

Fig. 3 Visualization of leaf instance segmentation results across models

对于病斑语义分割, 本研究使用 PSPNet^[24]、DeepLabV3+^[25]、ResNet-UNet^[20, 26]和 Mask2Former 模型进行试验。从不同模型语义分割试验结果可以看出, G-Mask2Former 模型在测试集上 mIoU 取得最高值, 达到 82.83%。在相同设置下, 该模型在测试集上的 mIoU 相较于 PSPNet、DeepLabV3+、ResNet-UNet 和 Mask2Former 模型分别提高了 8.14、2.75、0.97 和 0.89 个百分点, 对应的参数量分别增加 19.94、17.57、3.36、0.30 M (表 3)。值得注意的是, G-Mask2Former 模型采用实例分割和语义分割的多任务范式, 在与单任务语义模型相近的参数规模下可达到单任务语义分割模型性能。结果表明, G-Mask2Former 模型在复杂场景芒果病害诊断中具有一定应用潜力。按不同的类别测试集结果对比可以发现, 各模型背景交并比 ($\text{IoU}_{\text{背景}}$) 为 99.80% 左右, 基本饱和, 其改进主要来自藻斑病和炭疽病分割性能。对于藻斑病交并比 ($\text{IoU}_{\text{藻斑病}}$), G-Mask2Former 结果较优, 为 74.01%; ResNet-UNet 和 Mask2Former

的 $\text{IoU}_{\text{藻斑病}}$ 分别达到 73.74% 和 73.65%, 与最优值表现接近。对于炭疽病交并比 ($\text{IoU}_{\text{炭疽病}}$), G-Mask2Former 表现最佳, 达到 74.67%。综上, 在背景已饱和的条件下, 本方法的总体提升主要由前景类别带动, 尤其在炭疽病这一类别上优势明显, 表明 G-Mask2Former 模型在一定程度上能有效提高病斑分割质量。

不同模型三幅复杂场景下芒果叶片病害图像的语义分割结果的可视化见图 4。识别结果中, 黄色表示炭疽病病斑, 紫色表示藻斑病病斑。复杂背景纹理干扰增加了病斑的识别难度, PSPNet、DeepLabV3+ 和 Mask2Former 模型容易将炭疽病图像右上角的树枝截面处误识别为炭疽病, 而 G-Mask2Former 模型更关注叶片结构内信息, 能有效抑制该类误检。PSPNet 和 ResNet-UNet 模型将第一幅藻斑病图像枯萎叶缘误判为病斑, 反映出对前景组织形态变化的鲁棒性不足。而 DeepLabV3+ 模型未能识别出第二幅藻斑病图像中与叶片颜色对比度低的病斑, 存在漏检情况。

表 3 不同模型语义分割试验结果
Tab. 3 Experimental results of semantic segmentation across models

模型 Model	验证集 Validation set				测试集 Test set				参数量 Parameter/M
	mIoU/%	IoU 藻斑病/%	IoU 炭疽病/%	IoU 背景/%	mIoU/%	IoU 藻斑病/%	IoU 炭疽病/%	IoU 背景/%	
PSPNet	77.80	70.94	62.68	99.76	74.69	61.01	63.36	99.72	24.31
DeepLabV3+	81.19	74.06	69.70	99.81	80.08	70.17	70.27	99.80	26.68
ResNet-UNet	82.19	77.06	69.70	99.83	81.86	73.74	72.03	99.81	40.89
Mask2Former	84.38	78.04	75.27	99.84	81.94	73.65	72.38	99.81	43.95
G-Mask2Former	84.38	77.68	75.60	99.84	82.83	74.01	74.67	99.81	44.25

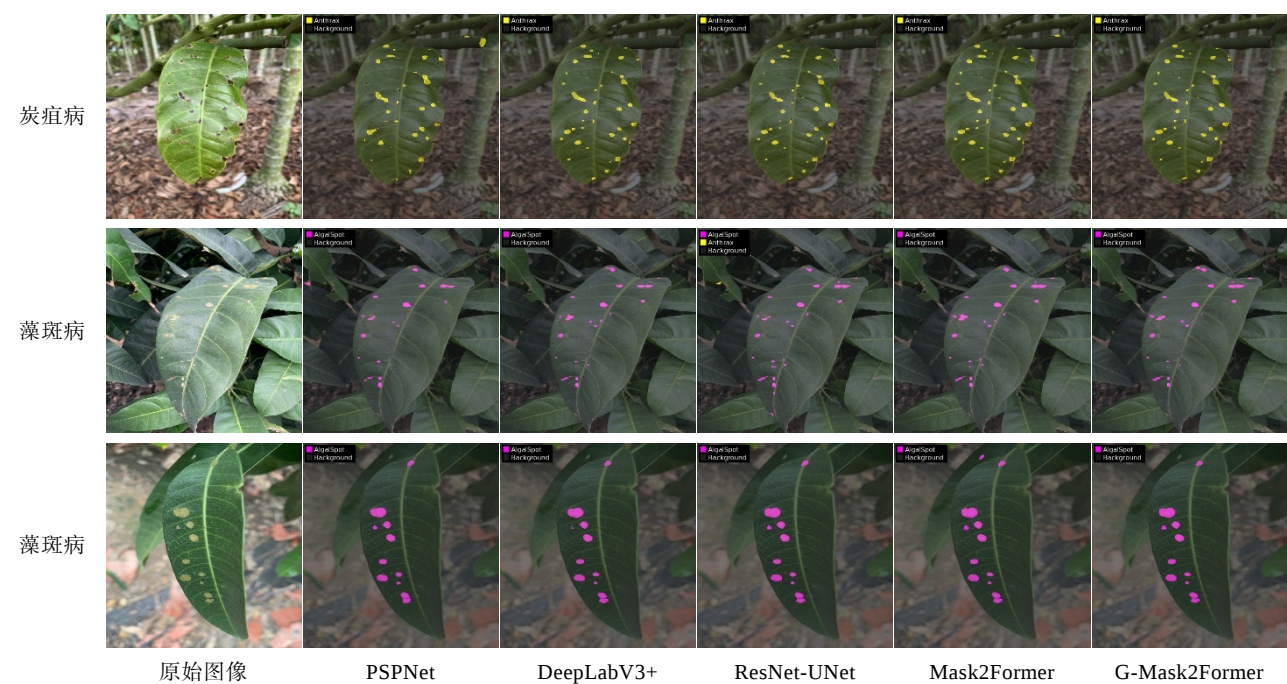


图 4 不同模型语义分割结果可视化
Fig. 4 Comparative visualization of semantic segmentation results

由此可见，G-Mask2Former 模型在复杂背景中能
有效抑制背景干扰，更适用于复杂场景下的芒果
叶片病斑分割。

目前，大多数叶片病害分割方法仅针对单叶
片图像，通常采用语义分割模型分别对病叶和病
斑进行分割。为与其他叶片病害分割方法进行对
比，本研究将叶片实例分割结果合并到叶片语义
掩码进行评估。选择两阶段方法 DUNet^[13]和
ALDNet^[14]作为对比模型。其中 DUNet 方法的第一
阶段采用以 MobileNet^[27]为主干网络的
DeepLabV3 进行叶片分割，第二阶段采用 U-Net
进行病斑分割。在实现过程中，为了简化过程，
本研究选择 MobileNetV2^[28]作为主干网络。ALDNet
方法的第一阶段设计 PBGNet (Res2Net^[29]) 进行叶
片分割，第二阶段设计 PDFNet 进行病斑分割。由
试验结果可以看出，在叶片分割任务上，ALDNet

的 mIoU 最高，G-Mask2Former 次之，DUNet 略
低。在病斑分割任务中，G-Mask2Former 模型达
到较好的效果。而在参数量方面，ALDNet 作为
两阶段方法，其总参数量达到 68.31 M，显著高
于 G-Mask2Former 模型（表 4）。DUNet 的总参
数量虽略低于 G-Mask2Former 模型，但其叶片分
割和病斑分割仍需分别训练两套模型。在工程实
现上，分阶段方法需要维护多个独立模型，部署
流程相对更加复杂。试验结果发现，所提出的模
型在分割精度与参数规模之间实现了良好平衡，对
实际叶片病害分割应用具有一定的参考价值。

综上所述，G-Mask2Former 模型在复杂场景
下实现了对芒果叶片实例和病斑区域的有效分
割，为后续病害定量评估提供了数据支撑。基于
分割结果，可进一步开展病害严重程度的定量估
算分析。

表 4 不同叶片病害分割模型对比试验结果
Tab. 4 Comparative experimental results of different leaf disease segmentation models

模型 Model	叶片分割/Leaf segmentation				病斑分割/Lesion segmentation					总参数量 Total parameter/M
	测试集 Test set			参数量 Parameter/M	测试集 Test set				参数量 Parameter/M	
	mIoU/%	IoU 叶片/%	IoU 背景/%		mIoU/%	IoU 藻斑病/%	IoU 炭疽病/%	IoU 背景/%		
DUNet	91.82	90.08	93.57	4.38	81.43	72.23	72.24	99.81	31.04	35.42
ALDNet	93.54	92.22	94.87	27.92	78.79	63.18	73.44	99.77	40.39	68.31
G-Mask2Former	92.62	91.27	93.97	44.25	82.83	74.01	74.67	99.81	44.25	44.25

2.3 芒果叶片病害量化

芒果叶片病斑占比是病害定量评估与病情指数计算的重要依据。通过计算同一片叶片上病斑面积占总叶面积的百分比，得到病害病斑面积占比。再依据分级标准可得到病害程度等级（病情指数）。单片叶片病斑占比（ P ）计算公式为：

$$P = \frac{S_{\text{disease}}}{S_{\text{leaf}}} \times 100\%$$

式中， S_{leaf} 为分割后的叶片实例面积， S_{disease} 表示该叶片上的病斑面积。为直观表征病害危害程度，图 5 为芒果叶片病斑面积占比的可视化结果。不同的叶片实例采用不同颜色表示，病斑也以轮廓线圈出以便识别。该可视化结果可直接呈现病斑在叶片内部的空间分布情况，配合每片叶片的病斑占比值，为病情指数计算提供定量依据，有助于开展病害的数字化诊断、监测与防控。



图 5 芒果叶片病斑面积占比可视化
Fig. 5 Visualization of mango leaf lesion area ratio

3 讨论

植物叶片病害分割与严重程度评估仍是农业研究的一项突出挑战。本研究提出的 G-Mask2Former 模型通过多任务联合训练策略和结构空间引导模块实现了叶片实例分割和病斑语义分割的协同优化，在复杂场景下芒果叶片病害分割任务中表现出良好的性能。现有研究多采用分阶段方

法处理叶片和病斑分割任务。如 WANG 等^[13]、CHENG 等^[14]均采用两阶段方法，先进行叶片语义分割再进行病斑语义分割。此类方法需要分别训练多个独立模型、模型维护和部署复杂。且叶片语义分割不能区分不同叶片实例，无法处理多叶片同时发生病害的情况。本研究通过构建病斑伪实例标签，在单一 Mask2Former 框架下实现了叶片实例分割和病斑语义分割的联合训练。这种统一框架的设计思路为多任务植物叶片病害分割提供了新的解决方案。此外，对叶片进行实例分割可支持多叶片同时发生病害的场景，便于统计每片叶片病斑面积占比，支持叶片级决策。LI 等^[17]在研究马铃薯叶片病害检测方法时设计了包含多叶片的分割研究，也说明多叶片分割更符合实际应用场景。

在复杂田间环境中，背景干扰是影响病害分类精度的主要原因之一。本研究利用“病斑位于叶片区域内”这一结构空间约束，设计了叶片结构空间软门控模块。该模块通过对多尺度特征图进行空间门控，引导模型关注叶片区域，有效抑制了背景干扰。消融试验显示，引入该模块后，测试集 AP 从 86.93%提升至 87.37%，mIoU 从 80.43%提升至 82.83%。上述结果表明，适度融入领域先验一定程度上会带来性能增益。

DENG 等^[10]通过研究发现，简单环境采集数据训练的病害识别模型在复杂背景下应用效果不佳。与使用受控环境或简单背景数据集不同，本研究所使用数据集采集于海南省芒果种植基地的真实田间环境，包含光照变化、遮挡、多叶片重叠等复杂情况。在这种真实应用场景下，本研究方法仍取得较好效果，证明了该方法对田间复杂场景芒果叶片病害分割有一定适用性。但是，受数据规模限制，虽然研究过程中通过数据增强扩充了数据集，但在强反光、强阴影等更复杂成像条件下，其稳健性仍需在更大规模、多场景的实

证中进一步验证,并开展有针对性优化。此外,数据集的病害多样性相对有限。目前样本仅覆盖芒果叶片炭疽病和藻斑病两类病害,缺乏对其他常见病害的表征,在实用性上受到一定约束。

考虑到病害多样性,进一步研究需采集多种病害、多场景、多生长阶段的数据,并进行细粒度标注,完善复杂场景下叶片病害数据集。对于早期病害,RGB 图像中病害往往不可见或对比度极不明显,可考虑融合高光谱、热红外图像等多模态信息,这些信息能够提供植物生理状态的丰富表征,有助于早期病害检测。同时可结合时间序列,捕获病害从潜伏期到显化期的特征演化,提升模型的检测能力。

4 结论

本研究针对复杂场景下芒果多叶片同时发生病害带来的特征学习困难和多任务分割挑战,提出结构空间引导的 G-Mask2Former 模型。该模型通过构建病斑伪实例标签,在单一 Mask2Former 框架下实现叶片实例分割和病斑语义分割的联合训练。与需要分别训练和部署多套子模型的分阶段病害分割方法相比,该设计在一定程度上简化了模型部署和推理流程。同时,通过引入叶片结构空间软门控模块,利用“病斑位于叶片区域内”的先验知识,有效抑制复杂背景干扰,提升了模型分割精度。试验结果表明,在田间复杂场景下芒果叶片病害分割测试数据集上, G-Mask2Former 模型的叶片实例分割平均精度(AP)达到 87.37%,病斑语义分割平均交并比(mIoU)达到 82.83%,证明了该方法在复杂环境下芒果叶片病害分割的有效性。基于分割结果,本研究进一步实现了叶片病斑面积占比的定量计算,为芒果叶片病害的严重程度评估提供了客观依据,有助于叶片病害数字化诊断、监测与精准防控。

参考文献

- [1] 王冰洁, 潘波, 杜长远, 姜蕾, 林勇, 张少帅. 植保无人机作业参数对雾滴在芒果冠层沉积分布的影响 [J]. 热带作物学报, 2024, 45(11): 2398-2406.
WANG B J, PAN B, DU C Y, JIANG L, LIN Y, ZHANG S S. Effects of spraying parameters for plant protection unmanned aerial vehicle on deposition distribution of droplets in mango canopy[J]. Chinese Journal of Tropical Crops, 2024, 45(11): 2398-2406. (in Chinese)
- [2] LEI L, YANG Q L, YANG L, SHEN T, WANG R X, FU C B. Deep learning implementation of image segmentation in agricultural applications: a comprehensive review[J]. Artificial Intelligence Review, 2024, 57(6): 149.
- [3] DANG M, WANG H X, LI Y F, NGUYEN T H, TIGHTIZ L, XUAN-MUNG N, NGUYEN T N. Computer vision for plant disease recognition: a comprehensive review[J]. The Botanical Review, 2024, 90(3): 251-311.
- [4] HU G S, WANG H Y, ZHANG Y, WAN M Z. Detection and severity analysis of tea leaf blight based on deep learning[J]. Computers & Electrical Engineering, 2021, 90: 107023.
- [5] LIANG Q K, XIANG S, HU Y C, COPPOLA G, ZHANG D, SUN W. PD2SE-Net: computer-assisted plant disease diagnosis and severity estimation network[J]. Computers and Electronics in Agriculture, 2019, 157: 518-529.
- [6] ESGARIO J G, KROHLING R A, VENTURA J A. Deep learning for classification and severity estimation of coffee leaf biotic stress[J]. Computers and Electronics in Agriculture, 2020, 169: 105162.
- [7] BOCK C H, CHIANG K S, DEL PONTE E M. Plant disease severity estimated visually: a century of research, best practices, and opportunities for improving methods and practices to maximize accuracy[J]. Tropical Plant Pathology, 2022, 47(1): 25-42.
- [8] SHI T T, LIU Y M, ZHENG X Y, HU K, HUANG H, LIU H L, HUANG H X. Recent advances in plant disease severity assessment using convolutional neural networks[J]. Scientific Reports, 2023, 13(1): 2336.
- [9] 邓阳君, 王旭, 龙陈锋, 刘金灵, 朱幸辉, 谭泗桥. 基于 VCDM-UNet 的水稻叶瘟病斑分割和病害程度分级[J]. 农业工程学报, 2024, 40(23): 190-198.
DENG Y J, WANG X, LONG C F, LIU J L, ZHU X H, TAN S Q. Segmenting and grading the blast disease of rice leaves using VCDM-UNet[J]. Transactions of the Chinese Society of Agricultural Engineering, 2024, 40(23): 190-198. (in Chinese)
- [10] DENG Y B, XI H R, ZHOU G X, CHEN A B, WANG Y F, LI L J, HU Y H. An effective image-based tomato leaf disease segmentation method using MC-UNet[J]. Plant Phenomics, 2023, 5: 0049.
- [11] LI K Y, ZHANG L X, LI B, LI S F, MA J C. Attention-optimized DeepLab V3+ for automatic estimation of cucumber disease severity[J]. Plant Methods, 2022, 18(1): 109.
- [12] LU B, LU J, XU X, JIN Y. MixSeg: a lightweight and accurate mix structure network for semantic segmentation of apple leaf disease in complex environments[J]. Frontiers in Plant Science, 2023, 14: 1233241.

- [13] WANG C S, DU P F, WU H R, LI J X, ZHAO C J, ZHU H J. A cucumber leaf disease severity classification method based on the fusion of DeepLabV3+ and U-Net[J]. *Computers and Electronics in Agriculture*, 2021, 189: 106373.
- [14] CHENG J X, SONG Z J, WU Y, XU J Y. ALDNet: a two-stage method with deep aggregation and multi-scale fusion for apple leaf disease spot segmentation[J]. *Measurement*, 2025, 253: 117706.
- [15] XU Y, MA B X, YU G W, ZHANG R Y, TAN H B, DONG F J, BIAN H T. Accurate cotton verticillium wilt segmentation in field background based on the two-stage lightweight DeepLabV3+ model[J]. *Computers and Electronics in Agriculture*, 2025, 229: 109814.
- [16] TASSIS L M, DE SOUZA J E T, KROHLING R A. A deep learning approach combining instance and semantic segmentation to identify diseases and pests of coffee leaves from in-field images[J]. *Computers and Electronics in Agriculture*, 2021, 186: 106191.
- [17] LI X D, ZHOU Y H, LIU J Y, WANG L B, ZHANG J, FAN X F. The detection method of potato foliage diseases in complex background based on instance segmentation and semantic segmentation[J]. *Frontiers in Plant Science*, 2022, 13: 899754.
- [18] RUSSELL B C, TORRALBA A, MURPHY K P, FREEMAN W T. LabelMe: a database and web-based tool for image annotation[J]. *International Journal of Computer Vision*, 2008, 77(1): 157-173.
- [19] CHENG B, MISRA I, SCHWING A G, KIRILLOV A, GIRDHAR R. Masked-attention mask transformer for universal image segmentation[C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). New Orleans, LA, USA: IEEE, 2022: 1280-1289.
- [20] HE K M, ZHANG X, REN S Q, SUN J. Deep residual learning for image recognition[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas, NV, USA: IEEE, 2016: 770-778.
- [21] HE K, GKIOXARI G, DOLLÁR P, GIRSHICK R. Mask R-CNN[C]//2017 IEEE International Conference on Computer Vision (ICCV). Venice, Italy: IEEE, 2017: 2961-2969.
- [22] WANG X L, ZHANG R F, KONG T, LI L, SHEN C H. Solov2: dynamic and fast instance segmentation[C]//Advances in Neural Information Processing Systems (NeurIPS), 2020, 33: 17721-17732.
- [23] TIAN Z, SHEN C H, CHEN H. Conditional convolutions for instance segmentation[C]//European Conference on Computer Vision. Glasgow, UK: Springer Cham, 2020: 282-298.
- [24] ZHAO H S, SHI J P, QI X J, WANG X G, JIA J Y. Pyramid scene parsing network[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu, HI, USA: IEEE, 2017: 6230-6239.
- [25] CHEN L C, ZHU Y, PAPANDREOU G, SCHROFF F, ADAM H. Encoder-decoder with atrous separable convolution for semantic image segmentation[C]//European Conference on Computer vision (ECCV). Munich, Germany: Springer-Verlag, 2018: 801-818.
- [26] RONNEBERGER O, FISCHER P, BROX T. U-net: convolutional networks for biomedical image segmentation[C]//Medical Image Computing and Computer-Assisted Intervention-MICCAI 2015. Munich, Germany: Springer, 2015: 234-241.
- [27] WU Z, SU L, HUANG Q M. Cascaded partial decoder for fast and accurate salient object detection[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach, CA, USA: IEEE, 2019: 3902-3911.
- [28] SANDLER M, HOWARD A G, ZHU M L, ZHMOGINOV A, CHEN L C. MobileNetV2: inverted residuals and linear bottlenecks[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, UT, USA: IEEE, 2018: 4510-4520.
- [29] HOWARD A G, ZHU M L, CHEN B, KALENICHENKO D, WANG W J, WEYAND T, ANDREETTO M, ADAM H. Mobilenets: efficient convolutional neural networks for mobile vision applications[J/OL]. (2017-04-17)[2025-09-17]. <https://doi.org/10.48550/arXiv.1704.04861>.