

中图分类号: TP391

单位代号: 10280

密 级: 公开

学 号: 23721573

上海大学



硕士学位论文

SHANGHAI UNIVERSITY
MASTER'S DISSERTATION

题 目	基于多尺度特征增强与融合的 植物叶片病害检测方法研究
--------	-------------------------------

作 者 马唯一

学科专业 软件工程

导 师 杨风雷

完成日期 二〇二六年四月

姓 名：马唯一

学号：23721573

论文题目：基于多尺度特征增强与融合的植物叶片病害检测方法研究

上海大学

本论文经答辩委员会全体委员审查，确认符合上海大学硕士学位论文质量要求。

答辩委员会签名：

主 席：

委 员：

导 师：

答辩日期：

马唯一

孙妍 陈俐

杨风雷

2026 年 6 月 9 日

姓 名：马唯一

学号：23721573

论文题目：基于多尺度特征增强与融合的植物叶片病害检测方法研究

上海大学学位论文原创性声明

本人郑重声明：所呈交的学位论文是本人在导师指导下，独立进行研究工作所取得的成果。除了文中特别加以标注和致谢的内容外，论文中不包含其他人已发表或撰写过的研究成果。参与同一工作的其他研究者对本研究所做的任何贡献均已在论文中作了明确的说明并表示了谢意。

学位论文作者签名：

马唯一

日期：2026 年 6 月 9 日

上海大学学位论文使用授权说明

本人完全了解上海大学有关保留、使用学位论文的规定，即：学校有权保留论文及送交论文复印件，允许论文被查阅和借阅；学校可以公布论文的全部或部分内容。

(保密的论文在解密后应遵守此规定)

学位论文作者签名：

马唯一

导师签名：

杨风雷

日期：2026 年 6 月 9 日

日期：2026 年 6 月 9 日

上海大学工学硕士学位论文

基于多尺度特征增强与融合的
植物叶片病害检测方法研究

作者: 马唯一

导师: 杨风雷

学科专业: 软件工程

计算机工程与科学学院

上海大学

2026 年 4 月

A Dissertation Submitted to Shanghai University for the
Degree of Master in Engineering

**Research on Plant Leaf Disease
Detection Methods Based on
Multi-Scale Feature Enhancement and
Fusion**

Candidate: Weiyi Ma

Supervisor: Fenglei Yang

Major: Software Engineering

**School of Computer Engineering and Science
Shanghai University
April, 2026**

摘 要

植物叶片病害的精准检测是保障粮食安全与推动智慧农业发展的重要手段。然而，在真实田间环境中，病害区域常呈现尺度差异大、形态复杂、背景干扰严重等特点，检测难度较高。现有目标检测方法应用于农业场景时，仍存在多尺度特征表达能力不足、不同层级特征融合不充分、语义信息衰减以及分类与回归任务之间的特征耦合等问题，限制了模型在小尺度及形态复杂病斑检测中的性能表现。

针对植物叶片病害目标尺度变化显著、病斑细节特征易丢失以及多尺度特征融合不充分的问题，本文提出一种基于分层递归对齐与全局协同融合的植物叶片病害目标检测算法（HRAGNet）。该方法构建多尺度特征增强编码器，通过交互式增强模块（IEM）实现不同层级特征之间的双向信息交互与递归语义强化；通过跨尺度语义对齐模块（CSAM）提升不同尺度特征在融合过程中的空间一致性，减少跨层融合带来的特征错位；并通过全局上下文聚合融合模块（GCAF）引入全局语义信息，对多尺度特征进行协同融合与统一建模，增强融合特征的判别能力与稳定性。

在多尺度特征增强与融合的基础上，针对融合特征在解码阶段利用不足以及分类与回归任务耦合过强的问题，本文进一步提出一种基于几何先验与任务特征协同优化的植物叶片病害检测算法（HRAGNet+）。该模型继承 HRAGNet 的多尺度特征增强与融合结构，并构建几何-任务引导增强解码器。几何先验增强注意力模块（GPEA）显式建模候选目标间的空间关系，增强模型对病斑几何结构与空间分布特征的感知能力；任务特征解耦与融合模块（TDFM）对分类特征与回归特征进行自适应分离与协同优化，缓解多任务优化过程中的特征冲突，提高分类与定位任务中的特征利用效率。

实验结果表明，在 PlantDoc、TLD 及 FD 等多个农业视觉数据集上，本文方法在平均精度等指标上均优于多种主流目标检测方法，尤其在小尺度病斑检测与复杂背景场景下表现出较好的检测性能与鲁棒性，验证了多尺度特征增强与融合策略在提升植物叶片病害检测性能方面的有效性。

关键词：叶片病害检测；目标检测；多尺度特征融合；特征增强；任务解耦

ABSTRACT

Accurate detection of plant leaf diseases is an important technical means for ensuring food security and promoting the development of smart agriculture. However, in real field environments, disease regions often exhibit large scale variations, complex morphologies, and severe background interference, making detection highly challenging. When existing object detection methods are applied to agricultural scenarios, they still suffer from insufficient multi-scale feature representation, inadequate fusion of features at different levels, semantic information attenuation, and feature coupling between classification and regression tasks. These issues limit model performance in detecting small-scale and morphologically complex lesions.

To address the problems of significant scale variations, loss of lesion detail features, and insufficient multi-scale feature fusion in plant leaf disease detection, this thesis proposes a plant leaf disease object detection algorithm based on hierarchical recursive alignment and global collaborative fusion, named HRAGNet. The proposed method constructs a multi-scale feature enhancement encoder. Specifically, the Interactive Enhancement Module (IEM) enables bidirectional information interaction and recursive semantic enhancement among features at different levels. The Cross-scale Semantic Alignment Module (CSAM) improves the spatial consistency of features at different scales during fusion and reduces feature misalignment caused by cross-level fusion. The Global Context Aggregation Fusion Module (GCAF) introduces global semantic information to collaboratively fuse and uniformly model multi-scale features, thereby enhancing the discriminative capability and stability of fused feature representations.

Based on multi-scale feature enhancement and fusion, this thesis further proposes an improved plant leaf disease detection algorithm based on geometric priors and task-feature collaborative optimization, named HRAGNet+, to address the insufficient utilization of fused features in the decoding stage and the strong coupling between classification and regression tasks. HRAGNet+ inherits the multi-scale feature enhancement and fusion struc-

ture of HRAGNet and constructs a geometric-task guided enhanced decoder. The Geometric Prior Enhanced Attention (GPEA) module explicitly models spatial relationships among candidate objects, enhancing the model's perception of lesion geometric structures and spatial distribution characteristics. The Task Decoupling and Fusion Module (TDFM) adaptively separates and collaboratively optimizes classification and regression features, alleviating feature conflicts in multi-task optimization and improving feature utilization efficiency in classification and localization tasks.

Experimental results on multiple agricultural vision datasets, including PlantDoc, TLD, and FD, show that the proposed methods outperform several mainstream object detection methods in terms of average precision and other metrics. In particular, the proposed methods achieve favorable detection performance and robustness in small-scale lesion detection and complex background scenarios, demonstrating the effectiveness of the multi-scale feature enhancement and fusion strategy in improving plant leaf disease detection performance.

Keywords: Plant leaf disease detection; Object detection; Multi-scale feature fusion; Feature enhancement; Task decoupling

目 录

摘 要	I
ABSTRACT	II
第一章 绪论	1
1.1 课题背景及意义.....	1
1.2 国内外研究现状.....	3
1.2.1 通用目标检测技术研究现状.....	3
1.2.2 植物叶片病害目标检测技术研究现状.....	7
1.3 本文主要研究内容.....	9
1.4 论文章节安排	10
第二章 相关理论和技术概述	12
2.1 深度学习目标检测技术基础.....	12
2.1.1 基于 CNN 的目标检测方法	12
2.1.2 基于 Transformer 的端到端目标检测方法	16
2.2 植物叶片病害目标检测技术基础	18
2.2.1 传统植物叶片病害视觉检测方法	18
2.2.2 深度学习下的植物叶片病害视觉检测方法	19
2.3 多尺度特征表示与融合技术基础	23
2.3.1 特征金字塔及其改进	23
2.3.2 多尺度特征融合策略	24
2.4 特征增强技术基础.....	26
2.4.1 卷积注意力机制	26
2.4.2 上下文建模方法	28
2.5 数据集与评价指标.....	29
2.5.1 数据集介绍	29
2.5.2 评价指标	32

第三章 基于特征交互与协调融合的叶片病害检测算法	34
3.1 模型总体设计	34
3.1.1 整体架构与模型输入输出	34
3.1.2 多尺度特征增强编码器	35
3.1.3 动态检测解码器	37
3.2 交互式增强模块 (IEM)	37
3.2.1 特征分组与子空间投影	38
3.2.2 结构-细节感知双路径并行特征增强	39
3.2.3 跨路径自适应聚合与特征整合阶段	42
3.3 跨尺度语义对齐模块 (CSAM)	43
3.3.1 跨尺度对齐建模	44
3.3.2 语义流场预测机制	45
3.3.3 可微分特征变形采样	46
3.4 全局上下文聚合融合模块 (GCAF)	47
3.4.1 全局上下文建模与多尺度联合表示	48
3.4.2 动态通道注意力权重生成	49
3.4.3 全局信息注入与残差增强	49
3.5 实验与结果分析	50
3.5.1 实验设置与评价指标	50
3.5.2 与主流检测算法的对比实验	51
3.5.3 消融实验	54
3.5.4 IEM 模块有效性分析	58
3.5.5 CSAM 模块有效性分析	60
3.5.6 GCAF 模块有效性分析	61
3.6 本章小结	62
第四章 基于几何先验与任务感知的叶片病害检测算法	63
4.1 模型总体设计	63
4.1.1 整体架构与模型输入输出	63
4.1.2 几何-任务协同增强解码器	65
4.1.3 优化目标与损失函数	66

4.2	几何先验增强注意力模块 (GPEA)	67
4.2.1	位置编码与位置感知提案特征的构建	68
4.2.2	基于 IoU 的几何关系建模机制	69
4.2.3	几何增强提案特征更新	71
4.3	任务特征解耦与融合模块 (TDFM)	72
4.3.1	分类-回归通道掩码解耦机制	73
4.3.2	自适应特征协同融合	75
4.4	编码器-解码器完整分析	76
4.4.1	完整模型架构与设计流程	76
4.4.2	编码器与解码器的协同优化机制	77
4.5	实验与结果分析	78
4.5.1	实验设置与评价指标	78
4.5.2	与主流检测算法的对比实验	78
4.5.3	消融实验	83
4.5.4	各组件有效性分析	85
4.6	本章小结	87
第五章	总结与展望	88
5.1	本文研究内容总结	88
5.2	未来研究展望	89
	插图索引	91
	表格索引	93
	参考文献	95
	攻读硕士学位期间取得的研究成果	104
	致 谢	105

第一章 绪论

1.1 课题背景及意义

粮食安全是关系国计民生的重大战略问题，病虫害的精准防控则是实现粮食稳产增产、保障农产品质量的关键环节。据联合国粮农组织（FAO）统计，植物病害是导致农作物减产和经济损失的重要因素之一，在全球范围内对农业生产造成了显著影响^[1]。通过早期病害检测与及时干预，可以在一定程度上降低农作物产量损失并提升农业生产效率。叶片是植物进行光合作用和物质代谢的重要器官，其健康状况直接影响作物生长，并最终影响产量与品质。在传统农业生产管理中，农作物病害的发现与诊断主要依赖人工巡检方式，即由植保人员或农业专家对田间作物叶片进行逐株观察与经验判断，从而识别潜在病害并采取相应防控措施。然而，该方式高度依赖专家经验与主观判断，不仅劳动强度大、检测效率低，而且具有明显滞后性，难以及时发现早期病害并实现精准定位。尤其在面对毫米级的初期病斑时，人工巡检极易出现漏检现象，难以满足现代规模化农业生产对病害实时监测与精准预警的需求。

随着智慧农业的不断发展，农业生产对植物病害的自动化与精准识别提出了更加迫切的需求。尽管计算机视觉与深度学习技术在通用目标检测领域已经取得显著进展，但在田间真实环境中开展植物叶片病害检测仍面临诸多挑战。其主要原因在于农业应用场景与通用视觉数据集之间存在显著的领域差异，使得在通用场景中训练得到的检测模型难以直接适应复杂农业环境。这种领域差异具体体现在植物叶片病害的目标表观特征与成像环境的差异上，主要体现在以下几个方面。第一，尺度差异显著。早期病斑通常仅占据图像中的少量像素，而在病害发展后期，病斑区域可能扩展并覆盖大面积叶面，呈现出显著的尺度变化。第二，形态复杂多变。病斑往往沿叶脉呈不规则扩展形态，其边界较为模糊，并伴随细粒度纹理变化；同时，同一类病害在不同作物品种或不同生育阶段下也可能表现出不同的视觉特征。第三，环境干扰显著。田间环境中的自然光照变化、复杂背景以及叶片之间的遮挡等因素相互叠加，使得病害区域与背景之间的视觉差异进一步减弱，从而增加检测任务难度。上述因素使得基于通用数据集训练的目标检测模型在农业场景中普遍出现性能退化，

难以满足田间实际应用对检测精度与鲁棒性的要求。

农业场景特性对目标检测模型的特征表达能力与环境适应能力提出了更高要求。然而,现有方法在多尺度特征建模与复杂背景下的鲁棒性方面仍存在不足,难以准确定位和识别不同尺度的病斑目标。首先,以特征金字塔网络 (Feature Pyramid Network, FPN)^[2]为代表的多尺度融合架构虽然通过自顶向下路径整合多层级特征,但其单向信息传播与静态融合机制容易导致高层语义信息在逐层传递过程中逐渐衰减,同时底层高分辨率的空间细节也难以形成有效反馈。其次,在面对大跨度尺度变化时,不同层级特征在强制融合过程中容易出现像素级空间对齐误差与语义冲突^[3],从而削弱模型对早期微小病斑的感知能力。此外,传统检测头通常采用共享特征进行分类与回归预测,而分类任务对平移不变性更为敏感,回归任务则更依赖空间结构信息,两者在特征需求上的差异导致任务间存在一定的耦合冲突^[4-5]。因此,现有代表性目标检测模型在微小病斑的局部特征强化以及叶片整体结构的全局上下文建模方面仍存在一定不足,使得病害特征在复杂背景中容易受到干扰,从而影响检测性能。

针对上述问题,当前植物叶片病害检测方法亟需从多尺度特征表达增强与空间先验引导的任务特征建模两个层面进行系统性改进。一方面,在编码阶段,现有特征金字塔结构对尺度变化的适应能力仍然有限,需引入更加灵活的多尺度特征增强机制,以缓解深层语义信息在传递过程中的逐层衰减问题,并强化浅层高分辨率特征对细粒度结构的表达能力,从而提升模型对微小病斑与大尺度病斑的协同感知能力。另一方面,在解码阶段,候选目标之间缺乏显式的空间关系建模,导致模型难以充分利用目标间的几何结构信息。为此,有必要引入空间先验信息来增强候选目标之间的交互能力,从而使模型能够更加准确地刻画植物病害在叶片上的空间分布特征。同时,针对分类与回归任务在特征需求上的差异,有必要构建任务驱动的特征建模机制,对不同任务分支的特征表示进行自适应调整,以缓解任务间的耦合冲突,从而提升整体检测性能。综上,从编码端特征增强到解码端任务优化的协同递进,是提升复杂田间场景下病害检测鲁棒性与精度的关键路径。

为此,本文围绕多尺度特征增强与跨尺度信息融合这一核心问题,提出一种面向复杂田间场景的植物叶片病害检测方法。在整体架构设计上,本文从编码端多尺度特征增强与解码端目标关系建模两个维度展开递进式优化,编码端通过引入结构化的多层级特征交互与语义对齐机制,构建表达能力更强、尺度一致性更优的特征表示;解码端则进一步考虑候选目标间的空间关联与分类、回归任务对特征需求的

差异性，通过目标关系重加权与任务自适应特征重组，提升提案生成质量与定位分类精度。通过上述编码与解码的协同优化，实现从特征表示到任务预测的逐级增强，最终显著提升模型在尺度变化剧烈、背景干扰复杂的田间环境下对植物叶片病害的检测精度与鲁棒性。

1.2 国内外研究现状

1.2.1 通用目标检测技术研究现状

通用目标检测是计算机视觉领域的重要研究方向，其技术发展经历了从手工特征方法向深度学习方法的转变，并逐渐由局部特征建模发展为全局关系建模^[6]。早期传统检测方法以手工设计的特征描述子为核心，Viola 等人提出的基于 Haar-like 特征的级联分类器框架，实现了人脸的实时检测^[7]；Dalal 等人提出的方向梯度直方图特征，通过统计局部梯度方向分布构建鲁棒的目标描述^[8]；Felzenszwalb 等人提出的可变形部件模型，将目标建模为根滤波器与部件滤波器的组合^[9]。此类方法通过人工设计特征表达方式，结合支持向量机分类器实现目标检测，但其特征表达能力受限于先验知识，泛化性能与鲁棒性难以适应复杂场景需求。

随着深度学习技术的发展，基于卷积神经网络（Convolutional Neural Network, CNN）的视觉模型在图像理解任务中取得了显著进展，并逐渐被引入到目标检测领域。基于 CNN 的目标检测框架通常分为两阶段与一阶段两类。两阶段方法以 R-CNN 系列为代表：Girshick 等 2014 年提出的 R-CNN^[10]首次将 CNN 引入目标检测，通过选择性搜索提取候选区域并逐区域分类；He 等 2015 年提出的 SPPNet^[11]引入空间金字塔池化层，实现任意尺寸输入的特征映射；Girshick 等 2015 年提出的 Fast R-CNN^[12]通过感兴趣区域池化实现特征共享，显著提升效率；Ren 等 2015 年提出的 Faster R-CNN^[4]创新性地引入区域提议网络（Region Proposal Network, RPN），将候选区域生成纳入网络端到端训练，实现了真正意义上的端到端检测。此后，两阶段检测框架进一步向精细化与高质量检测方向发展。He 等 2017 年提出的 Mask R-CNN^[13]，在检测基础上并行添加分割分支，实现实例分割与目标检测的统一；Cai 等 2018 年提出的 Cascade R-CNN^[14]，通过级联多阶段检测头逐步提高定位精度，进一步提升检测性能。总体而言，两阶段检测方法通过候选区域生成与精细分类回归的分阶段处理策略，通常能够获得较高的检测精度，但其复杂的检测流程也在一定程度上限

制了模型的实时性与计算效率。

为满足实时性需求，一阶段检测方法直接在密集网格上预测目标边界框与类别概率，从而避免候选区域生成过程。Redmon 等 2016 年提出的 YOLO (You Only Look Once)^[15] 将目标检测重构为统一的回归问题，实现了端到端的实时检测框架。随后 YOLO 系列模型通过引入锚点机制、多尺度预测结构以及更深层的特征提取网络，进一步提升了检测精度与鲁棒性^[16-17]。Liu 等 2016 年提出的 SSD (Single Shot Multi-Box Detector)^[18] 融合 YOLO 的回归思想与 Faster R-CNN 的锚点机制，在不同尺度特征图上设置锚点框，从而有效提升多尺度目标的检测能力。Lin 等 2017 年提出的 RetinaNet^[19] 揭示了一阶段检测中正负样本极度不平衡的问题，并设计焦点损失函数以强化困难样本学习，推动了一阶段检测方法在精度上的显著提升。总体而言，一阶段检测方法通过端到端的密集预测策略显著提升了检测速度。尽管这些方法在通用视觉场景中取得了较好效果，但在小目标检测与复杂背景环境下仍存在一定局限，而这正是农业图像分析中的典型挑战。

尽管上述检测方法在检测速度与效率方面取得了显著提升，但大多数一阶段检测器仍依赖预设锚点框来生成候选目标，这在一定程度上增加了模型设计复杂度，并对锚点尺寸与比例选择具有较强依赖性。为减少对人工先验的依赖，研究者进一步提出了无锚点检测方法。Law 等 2018 年提出的 CornerNet^[20] 通过检测目标的左上角与右下角关键点实现目标定位；Duan 等 2019 年提出的 CenterNet^[21] 进一步通过中心点表示目标，从而简化后处理流程；Tian 等 2020 年提出的 FCOS^[22] 直接进行逐像素预测，无需锚点即可获得与锚点方法相当的检测性能。YOLO 系列也在后续版本中进一步发展出无锚框机制，并逐步采取先进的标签分配策略、无锚框端到端检测以及高效的网络结构设计，持续提升检测精度与鲁棒性^[23-26]。总体来看，无锚点检测方法通过关键点建模或逐像素预测等方式替代传统锚点机制，在降低模型设计复杂度的同时增强了对目标尺度与形状变化的自适应能力，已成为目标检测领域的重要发展方向之一。

随着深度学习研究的不断发展，研究者开始探索能够建模全局上下文关系的新型检测框架。Transformer^[27] 在自然语言处理领域的成功为视觉任务提供了新的研究思路，并推动了目标检测方法从基于卷积网络的局部感知模式向全局关系建模范式的转变。Carion 等 2020 年提出的 DETR (DEtection TRansformer)^[28] 将目标检测重构为集合预测问题，利用全局注意力捕捉上下文依赖，彻底摒弃锚点与非极大值抑制，

实现端到端训练。这一框架为目标检测提供了新的研究范式，但在实际应用中仍面临训练收敛较慢以及小目标检测性能不足等问题，从而引发了一系列后续改进工作。Zhu 等 2021 年提出的 Deformable DETR^[3] 引入可变形注意力机制，仅对稀疏关键点进行注意力计算，大幅加速收敛并提升小目标检测性能。Liu 等 2022 年提出的 DAB-DETR^[29] 通过引入动态锚点框对查询进行初始化与更新，从而增强模型的空间位置感知能力；Zhang 等 2022 年提出的 DINO^[30] 融合对比学习与去噪训练策略，提高了模型对复杂样本的鲁棒性，并进一步优化了检测精度与收敛稳定性；Zong 等 2023 年提出的 CO-DETR^[5] 通过构建协同优化机制强化特征学习过程，有效缓解了训练过程中监督信号不足的问题；Cai 等于 2023 年提出的 AlignDETR^[31] 通过引入定位精度感知的分类损失，强化了分类分数与定位质量之间的相关性，有效缓解了 DETR 中分类与回归任务的不对齐问题；Zhao 等于 2024 年提出的 RT-DETR^[32] 通过轻量化结构设计与高效解码策略，实现了端到端检测框架在实时场景中的应用。这些模型通常以卷积神经网络或混合结构作为特征提取骨干，结合 Transformer 编码器—解码器结构进行全局关系建模，并通过引入位置编码与多尺度特征表示增强空间表达能力。

上述方法在全局建模与上下文信息利用方面取得了重要进展，但在处理真实场景中普遍存在的尺度变化问题时仍面临挑战。由于目标在自然场景中往往呈现出显著的尺度差异，如何在不同分辨率特征之间实现有效的信息交互与语义补充，成为目标检测模型设计中的关键问题。关于多尺度特征的早期研究主要围绕特征金字塔结构展开，形成了一系列具有代表性的经典方法。Lin 等于 2017 年提出特征金字塔网络^[2]，通过构建自顶向下的特征传播路径，将高层语义信息传递至低层特征，从而增强小目标检测能力；Liu 等在 2018 年提出路径聚合网络^[33]，在 FPN 基础上引入自底向上的信息传递路径，以进一步强化不同层级特征之间的信息流动；Ghiasi 等于 2019 年提出 NAS-FPN^[34]，利用神经结构搜索自动设计多尺度特征融合结构，提高了特征金字塔的表达能力；Schuster 等于 2020 年提出 ResFPN^[35]，通过引入残差跳跃连接强化不同分辨率特征之间的融合，有效提升了特征的空间定位能力与细节表达能力。这些研究从不同角度优化了多尺度特征的融合方式，为后续特征金字塔结构的改进提供了重要思路。

近年来，随着目标检测任务复杂度的不断提升，研究者对特征金字塔结构进行了进一步的优化与扩展，从跨尺度对齐机制、全局建模能力以及任务适应性融合策略等方面推动特征金字塔结构的发展。Guo 等于 2023 年提出 MSE-FPN^[36]，通过语

义增强与门控通道引导机制减小不同尺度特征之间的语义差异，从而提高跨尺度融合效果；Ke 等于 2024 年提出的 ProFPN^[37]，通过渐进式融合策略进一步增强特征金字塔中的上下文信息表达能力；Yang 等于 2025 年提出 PAMFPN^[38]，结合位置建模与多尺度卷积核设计，以增强对稀疏目标与复杂背景的适应能力；Mbietieu 等于 2025 年提出 ShipFPN^[39]，通过引入自适应融合与注意力增强机制，提高复杂场景下小目标检测的精度表现；Gu 等于 2026 年提出频域感知特征金字塔框架 FFPF^[40]，通过频率分解与双谱特征融合机制增强小目标的细粒度表达能力。这些研究表明，多尺度特征融合结构正朝着更精细的跨尺度交互与自适应融合方向发展，并持续推动目标检测性能的提升。

特征金字塔结构的改进在多尺度特征融合方面取得了一定进展，但其在特征选择与信息分配方面仍存在局限。为此，研究者从特征建模机制角度出发，引入注意力机制以提升特征融合过程中的信息建模能力与表达效果。Hu 等提出的 SE 模块^[41]通过通道注意力机制对特征响应进行自适应重标定；Woo 等提出的 CBAM 模块^[42]进一步结合通道注意力与空间注意力；Fu 等提出的 DANet^[43]联合建模空间与通道依赖关系，实现更充分的全局上下文建模。这些经典注意力结构为后续研究奠定了重要基础。近年来，研究者围绕注意力机制的表达能力与计算效率开展了进一步探索。Zhang 等提出的 HAM^[44]通过协同建模通道与空间注意力以实现更有效的特征重标定；Lau 等在视觉注意力网络中提出大卷积注意力机制^[45]及其改进结构，通过大卷积核注意力机制增强网络对长距离空间依赖的建模能力。这些研究表明，视觉注意力机制正从早期的简单通道或空间加权方式，逐步发展为能够同时建模局部细节与全局上下文关系的结构化注意力模块。

综上所述，通用目标检测技术在多尺度特征建模与注意力机制方面取得了显著进展。然而，现有通用目标检测方法大多针对通用视觉场景设计，在植物叶片图像分析等特定应用场景中仍面临诸多挑战。农业场景中常见的尺度剧变、小目标密集以及复杂背景干扰等问题会对目标检测模型的特征表达能力和定位精度提出更高要求。因此，如何在复杂自然环境中构建具有更强多尺度特征表达能力和上下文建模能力的检测框架，仍然是农业视觉检测领域亟待解决的重要问题。

1.2.2 植物叶片病害目标检测技术研究现状

在深度学习技术广泛应用之前,植物病害检测主要依赖人工判别与基于数字图像处理的传统机器学习方法^[46]。传统的人工肉眼判别通常由植物专家通过观察叶片病斑的颜色、形状与分布特征进行诊断。该方法受限于个体的经验差异,效率较低,且难以适应规模化农业生产的实时监测需求。随着数字图像处理技术的发展,研究者探索了自动化的检测方案。此类方法通常先通过滤波去噪与颜色空间变换等操作进行图像预处理,随后利用阈值分割、边缘检测等算法将疑似病斑区域从健康背景中分离,并提取人工设计的特征向量送入传统分类器进行识别。Arivazhagan 等利用颜色共生矩阵提取病害纹理特征并结合支持向量机进行植物病害分类^[47]; Mokhtar 等提出了基于 Gabor 滤波器提取番茄叶片纹理特征的方法^[48]; Phadikar 等利用 Otsu 阈值分割算法结合几何形状与颜色特征,实现了对水稻病害的初步识别^[49]。尽管传统机器学习方法在受控的实验室环境中取得了一定成效,但其有效性高度依赖于手工特征的设计质量。所提取的低阶特征难以完整刻画病斑复杂的形态变化,且算法的泛化能力有限。此外,在真实的田间场景中,光照变化、杂乱背景以及病斑尺度差异等因素极易导致特征提取失效,传统方法在此类复杂工况下的鲁棒性不足,难以满足实际农业环境的应用要求。

深度卷积神经网络的应用为植物病害检测提供了新的研究范式,其端到端的特征学习机制提升了模型在复杂环境下的特征表达能力。现有基于深度学习的植物病害检测研究主要分为分类任务与检测任务两类,前者关注图像是否患病及病害类别,后者则需要同时定位病斑位置并识别类别^[50]。主流目标检测框架被广泛迁移至植物病害检测任务中,并取得了优于传统方法的检测性能。深度学习早期阶段,具备候选区域生成机制的两阶段检测器因其较好的细粒度特征提取能力被广泛关注。Bari 等提出了一种基于 Faster R-CNN 的变体模型用于水稻病害检测,在公开数据集与特定场景下取得了较好的定位性能^[51]; Zhang 和 Yang 通过引入双层 RPN 架构,进一步改善了模型在复杂田间条件下的病害检测效果^[52]; Pallapothu 等基于 Mask R-CNN 构建棉花叶片检测模型,通过引入像素级分割分支对病斑区域进行精细刻画,在提升目标定位精度的同时增强了对不规则病斑形态的建模能力^[53]。

两阶段方法由于包含候选框生成与分类回归等串行步骤,推理速度受到一定限制,在算力受限的农业边缘设备上部署存在困难。为此,以 YOLO 系列为代表的一

阶段检测器因在检测精度与推理速度之间取得较好平衡, 被广泛应用于农业检测任务。**Wu** 等提出的 **DAONet-YOLOv8** 模型在 **YOLOv8** 基础上引入融合卷积与自注意力的双重注意力模块, 并结合遮挡感知检测头以及动态合成卷积, 在复杂茶园场景下提升了检测精度并降低了模型参数量^[54]; **Wang** 等提出了 **BED-YOLO** 模型, 在 **YOLOv10n** 基础上引入双向特征金字塔结构和高效多尺度注意力机制, 有效增强了复杂田间环境下小尺度与遮挡病斑的特征表达能力, 提升了番茄叶片病害检测的精度与鲁棒性^[55]; **Nguyen** 等提出了 **RiceLDD-YOLO** 模型, 在 **YOLOv13** 框架中引入 **ImDS-C3k2** 卷积模块以及融合 **SPPF** 的多尺度聚合检测头, 在提升模型对小尺度及不规则病斑检测能力的同时兼顾了实时性与轻量化部署需求^[56]。综上所述, 通过迁移通用目标检测框架并结合轻量化骨干网络、多尺度特征融合结构以及注意力机制等改进策略, 基于卷积神经网络的目标检测方法在检测精度与推理效率方面均取得了显著提升。然而, 由于卷积神经网络的感受野主要依赖局部卷积操作, 其对长距离上下文关系的建模能力仍然有限。在复杂田间环境中, 植物叶片病斑往往具有尺度变化剧烈、形态差异明显以及与背景纹理高度相似等特点, 单纯依赖局部特征建模的检测框架在全局语义理解方面仍存在一定不足。

为改善卷积神经网络在局部感受野与特征交互方面的局限性, 针对植物病害图像中病斑尺度变化剧烈、微小病斑检测困难的问题, 部分研究采取在特征金字塔不同层级引入通道注意力或空间注意力机制, 使网络能够根据病斑尺度自适应选择更具判别力的特征表示。**Luo** 等提出的 **SE-FPN** 金字塔架构通过结合感受野增强与高级语义提取模块, 有效改善了密集遮挡与复杂背景下的特征表达质量^[57]; **Wang** 等在骨干网络中引入 **CBAM** 增强病斑特征提取能力, 同时构建 **BiRepGFPN** 模块实现深层语义与浅层空间信息的有效融合^[58]; **Sahu** 等提出了一种基于空洞卷积与注意力机制的混合模型, 通过构建融合注意力的 **Mask R-CNN** 分割网络与多尺度 **DenseNet** 分类结构, 结合优化算法提升模型参数学习能力, 从而在复杂农业环境下实现了更高精度的植物病害检测与识别^[59]。在此基础上, 近年来研究逐步引入注意力机制与 **Transformer** 架构以增强特征建模能力。**Haruna** 等提出的 **SAR-PFF** 模型融合二维多头自注意力、注册标记与扩张卷积, 在捕获长距离依赖关系方面表现出较好效果, 从而降低了水稻叶部病害的漏检率^[60]; **Chai** 等提出的 **PlantAIM** 融合了 **ViT** 的全局注意力与 **CNN** 的局部特征提取能力, 实现了作物特异性与病害特异性特征的联合学习^[61]。

与此同时,随着 **Transformer** 架构的提出与不断发展,其在建模长距离依赖关系与全局上下文信息方面的优势逐渐凸显,使得基于 **Transformer** 的方法开始被引入目标检测任务,并逐步应用于植物叶片病害检测等复杂视觉场景中,以提升模型对遮挡、背景干扰及目标间关系的建模能力。在此基础上,DETR 首次将目标检测建模为集合预测问题,通过引入编码器—解码器结构与二分图匹配策略,实现了无需锚点框与非极大值抑制的端到端检测框架。这一范式也为植物叶片病害检测提供了新的技术路径。围绕 DETR 的性能优化与应用拓展,研究者提出了一系列改进方法。Fu 等提出 MHDI-DETR,通过引入 MobileNetv4 骨干与改进的特征融合模块,实现了对葡萄叶片病害的高效检测^[62];Huo 等提出 EBGC-RTDETR 模型,在 RT-DETR 基础上通过引入高效通道注意力以及 GSConv 双向特征金字塔,并结合剪枝与通道蒸馏策略,实现了精度提升与计算复杂度的显著降低^[63];Wang 等提出了 ERL-RTDETR 模型,在 RT-DETR 框架中引入轻量自适应特征提取模块,并结合 GELAN 结构优化特征表达,在保证模型轻量化的同时提升了苹果多病害检测的精度与细粒度特征表征能力^[64]。上述研究表明,基于 **Transformer** 的端到端检测方法在处理叶片遮挡与复杂背景问题方面具有良好的应用潜力。

综上所述,尽管深度学习方法在植物病害检测领域取得了一定进展,但在复杂田间环境中的应用仍存在明显局限。一方面,现有模型在多次下采样过程中容易丢失细粒度信息,导致对早期微小病斑的表征能力不足;另一方面,模型主要依赖数据驱动的特征学习,缺乏对病斑空间分布结构的显式建模,使其在复杂背景下的鲁棒性受限。此外,在端到端检测框架中,分类与定位任务共享特征进行联合优化,当目标尺度变化剧烈或类别分布不均时,易产生任务间特征竞争,从而影响检测性能。上述问题的本质在于现有方法在多尺度特征表达与跨尺度信息协同方面仍存在不足,难以同时兼顾细粒度特征保留与全局结构建模能力。因此,有必要从多尺度特征增强与融合的角度出发,对特征表达与检测机制进行协同优化,以提升模型在复杂农业场景中的适应能力。

1.3 本文主要研究内容

针对田间复杂环境下植物叶片病害检测中尺度变化大、背景干扰强及小目标占比高等特点,目标在不同空间尺度上的表现差异显著,模型通常需要依赖多尺度特

征对其进行表征与融合。然而，在实际建模过程中，多尺度特征在表达与利用上仍存在明显不足：一方面，不同尺度特征在融合过程中容易出现语义不一致与空间错位，导致特征表达不稳定；另一方面，融合后的特征在检测阶段仍以共享形式参与目标预测，缺乏对目标间空间关系与任务差异的进一步建模，从而限制了检测性能的提升。

为解决上述问题，本文首先提出一种基于特征交互与协调融合的叶片病害检测算法，用于实现高质量的多尺度特征增强与融合。在编码阶段，构建多尺度特征增强编码器，对不同层级特征进行协同建模与优化。该编码器主要由交互式增强模块、跨尺度语义对齐模块以及全局上下文聚合融合模块组成；其中，交互式增强模块通过跨层特征的双向交互与递归强化，使不同尺度特征能够充分融合细节信息与语义信息；跨尺度语义对齐模块对不同分辨率特征施加空间与语义一致性约束，以缓解多尺度融合过程中的特征错位问题；全局上下文聚合融合模块通过建模跨尺度的全局依赖关系，进一步生成语义一致且表达能力更强的多尺度融合特征，为后续检测阶段提供更加稳定且具有判别力的特征表示。

在此基础上，针对融合特征在检测阶段缺乏目标级结构建模以及分类与回归任务存在冲突的问题，进一步提出一种基于几何先验与任务感知的叶片病害检测算法。该方法在保留多尺度特征增强编码器的基础上，对解码阶段进行重构，提出几何-任务协同增强解码器，对多尺度融合特征进行目标级重建与任务驱动优化。该解码器主要由几何先验增强注意力模块和任务特征解耦与融合模块组成；几何先验增强注意力模块通过引入候选目标之间的几何关系，对提案交互过程进行约束，使模型能够显式刻画目标之间的空间分布关系；任务特征解耦与融合模块结合任务驱动的特征重组机制对分类与回归特征进行自适应调整，缓解多任务优化过程中的特征冲突，并生成更加稳定的提案表示用于后续迭代优化。

1.4 论文章节安排

本文共分为五章，各章内容安排如下。

第一章主要介绍田间植物叶片病害检测的研究背景与研究意义，从智慧农业与精准农业的发展需求出发，阐述利用计算机视觉技术实现农作物病害自动检测的重要价值。对复杂田间环境下农业视觉检测面临的主要挑战进行分析，并综述通用目

标检测技术及植物病害检测领域的国内外研究现状。在此基础上明确本文的研究目标与主要研究内容，并概述论文的整体技术路线。

第二章系统介绍目标检测任务涉及的基础理论与关键技术，为本文方法的研究奠定理论基础。本章首先阐述目标检测的基本任务形式与主流检测框架，包括两阶段检测器与一阶段检测器的基本结构及特点；随后梳理多尺度特征表示与融合方法、特征增强与上下文建模技术的发展思路与典型方法；最后结合农业视觉应用背景，总结植物叶片病害目标检测的相关研究进展。

第三章针对复杂田间场景下植物叶片病害目标尺度变化显著、背景干扰复杂以及多尺度特征语义不一致等问题，提出了一种基于特征交互与协调融合的叶片病害检测算法，并构建多尺度特征增强编码器。该编码器包含交互式增强模块、跨尺度语义对齐模块与全局上下文聚合融合模块；其中，交互式增强模块通过跨层级特征交互与多路径协同建模，实现多尺度特征在语义信息与结构细节层面的协同增强；跨尺度语义对齐模块用于缓解不同尺度特征之间的空间错位问题；全局上下文聚合融合模块则通过整合全局语义信息，进一步提升多尺度病害目标的特征表达能力与复杂场景下的检测鲁棒性。最后，在多个公开农业数据集上开展对比实验与消融实验，以验证所提模型的有效性与性能优势。

第四章在多尺度特征编码器的基础上，提出了一种基于几何先验与任务感知的叶片病害检测算法，并构建几何-任务协同增强解码器。针对检测过程中分类与定位任务之间的特征冲突以及模型对病斑几何结构感知能力不足的问题，引入几何先验增强模块以强化空间结构信息表达；同时设计任务特征解耦与融合模块，对分类与定位特征进行解耦以减少任务间干扰，从而提升病害区域的定位精度与类别识别能力，并结合自适应融合机制进一步完成任务特征的重组与优化。最后，通过多个公开农业数据集上的对比实验与消融实验验证模型的性能优势。

第五章为总结与展望。本章对全文研究工作进行总结，归纳本文在植物叶片病害目标检测模型设计方面的主要成果，并分析所提出方法的优势与不足。在此基础上，对未来植物叶片病害检测技术的发展方向进行展望。

第二章 相关理论和技术概述

植物叶片病害目标检测是计算机视觉在智慧农业领域的重要应用方向，其核心任务是在复杂田间环境下对叶片病斑区域进行准确定位与病害类别识别。随着深度学习技术的快速发展，农业目标检测方法逐渐从依赖人工特征的传统方法转向基于深度神经网络的检测框架。然而，农业场景普遍具有多尺度目标、复杂背景以及目标形态多样等特点，这些因素对现有检测技术提出了新的挑战。本章系统梳理目标检测相关的理论基础与关键技术，首先综述深度学习目标检测框架的演进脉络，重点阐述多尺度特征表示与融合、特征增强与上下文建模等核心技术原理，最后总结植物叶片病害目标检测的研究现状，为后续面向农业场景的多尺度特征融合与增强方法研究提供理论支撑。

2.1 深度学习目标检测技术基础

本节将现有主流深度学习目标检测方法划分为基于卷积神经网络（CNN）和基于 Transformer 两大类分别进行阐述。前者侧重介绍 CNN 检测器的典型范式与演进脉络，后者聚焦 Transformer 在检测任务中的核心原理与代表工作。

2.1.1 基于 CNN 的目标检测方法

卷积神经网络凭借其强大的局部特征提取能力和层级化的感受野结构，成为深度学习目标检测的主流技术基础。根据检测流程的设计差异，基于 CNN 的目标检测方法可进一步分为两阶段检测方法和一阶段检测方法。

（1）两阶段目标检测方法

两阶段目标检测方法是基于 CNN 的目标检测方法的重要研究方向，其核心思想是将检测任务划分为两个阶段，首先生成候选区域，然后对候选区域进行分类与边界框回归。主要代表工作为 R-CNN^[10]、Fast R-CNN^[12]和 Faster R-CNN 模型^[4]，R-CNN 系列算法具体工作流程如图 2.1 所示。

R-CNN^[10]算法由 Girshick 等人于 2014 年提出，该方法首次将卷积神经网络应用于目标检测任务，通过选择性搜索算法从输入图像中提取约两千个候选区域，并



图 2.1 R-CNN 系列算法对比

对每个候选区域进行缩放归一化后输入 CNN 提取特征，最后采用支持向量机进行分类。尽管 R-CNN 大幅提升了检测精度，但其逐候选区域的特征提取方式导致推理速度极慢且存储开销巨大。针对这一问题，2015 年提出的 Fast R-CNN^[12] 模型将整张图像一次性输入 CNN 得到共享特征图，所有候选区域通过感兴趣区域池化层映射到固定尺寸的特征向量上，实现了特征提取过程的计算共享，显著提升了训练与推理效率。真正实现端到端检测的突破性工作 Faster R-CNN^[4] 核心贡献在于提出了区域提议网络 (Region Proposal Network, RPN)。RPN 通过在共享特征图上滑动锚点框，对候选区域执行前景背景二分类以及边界框粗回归，从而生成候选区域。该设计实现了候选区域生成与检测网络在同一框架内的端到端训练，从而避免了对外部候选区域生成算法的依赖。

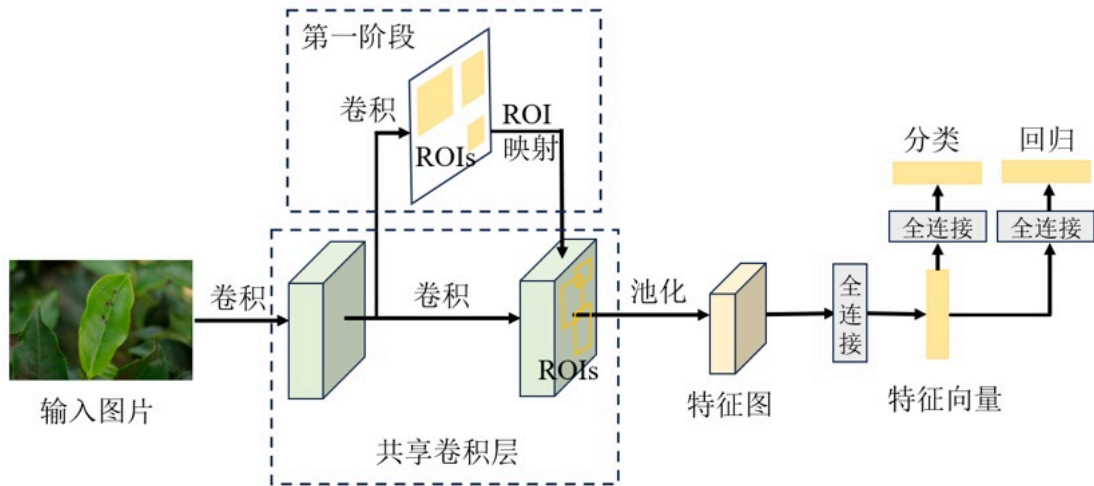


图 2.2 Faster R-CNN 算法的网络结构

Faster R-CNN 的网络结构如图 2.2 所示。模型首先通过骨干网络提取图像特征，随后 RPN 基于锚点框生成高质量的候选区域，最后由检测头对每个候选区域进行精细的分类与边界框回归。具体而言，RPN 在共享特征图上以滑动窗口的方式生成候选区域，在每个空间位置预定义多个具有不同尺度和长宽比的锚点框，并同时预测

每个锚点的前景或背景概率以及边界框偏移量，从而快速筛选出一组高质量候选区域。相比于传统方法中依赖选择性搜索的候选区域生成方式，RPN 实现了候选区域生成过程的卷积化与端到端学习，大幅提升了效率与候选区域质量。

在候选区域特征提取阶段，Faster R-CNN 采用 RoI Pooling 及其改进形式 RoI Align 对不同尺寸的候选区域进行特征重采样，将其映射为固定尺寸的特征表示，从而便于后续全连接层进行分类与回归预测。该过程有效解决了候选区域尺寸不一致的问题，同时保留了目标区域的判别性特征。

两阶段检测模型通过候选区域筛选与精细回归相结合的方式，使模型能够对目标区域进行更加精确的定位，尤其在复杂背景、小目标或目标间存在遮挡的场景中表现出较强的鲁棒性。然而，由于其在推理过程中需要对大量候选区域进行逐区域特征重采样与独立预测，计算开销较大，导致整体推理速度相对较慢，相比同期的一阶段方法在实时性方面存在一定不足，这也在对响应速度要求较高的农业部署场景中带来一定挑战。

(2) 一阶段目标检测方法

一阶段目标检测方法采用端到端的简洁架构设计，无需显式生成候选区域，其核心思想为在特征图上直接回归目标的类别概率与边界框坐标，通过单次前向传播即可完成目标检测任务。YOLO (You Only Look Once)^[15] 作为一阶段目标检测的代表方法，自 2016 年由 Joseph Redmon 等人提出后逐步发展形成系列模型。其核心思想是将输入图像划分为固定大小的网格，每个网格负责预测若干边界框及其置信度，并在统一网络中同时完成类别概率预测与边界框回归，实现了检测任务的端到端建模。其网络结构如图 2.3 所示。

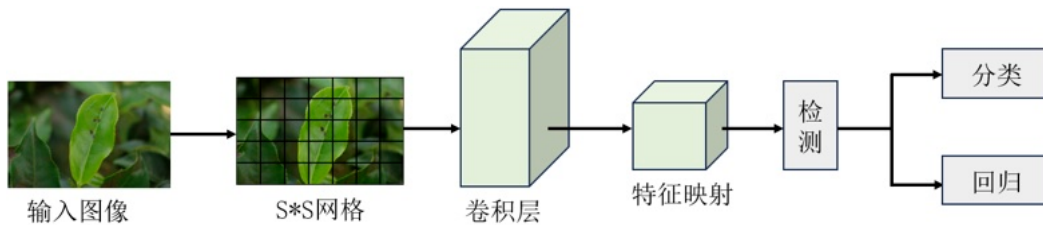


图 2.3 YOLO 算法的网络结构

YOLO 将目标检测问题转化为一个回归问题，通过直接学习从图像像素到边界框坐标与类别概率的映射关系，避免了候选区域生成与重复特征提取过程，从而显著提升了检测效率。这种统一化设计使 YOLO 在推理阶段能够达到每秒数十帧的处

理速度，首次实现了目标检测的实时性要求。与此同时，YOLO 通过多任务损失函数对位置误差、置信度误差以及分类误差进行联合优化，使模型能够在单次前向传播中同时学习定位与识别能力。然而，早期 YOLO 由于采用粗粒度网格划分，每个网格仅负责有限数量的目标预测，导致在处理密集目标或小目标时容易出现定位不准确或漏检问题。

针对上述局限，后续 YOLO 系列模型在多个方面进行了改进。一方面，通过引入锚点框机制与多尺度预测结构，如在不同层级特征图上进行检测，增强了对不同尺度目标的建模能力；另一方面，通过优化网络结构以及改进损失函数与训练策略，进一步提升了检测精度与鲁棒性。总体而言，YOLO 系列通过将目标检测统一为单阶段回归问题，在检测速度与系统简洁性方面具有显著优势，但其早期模型在小目标与密集场景中的表现相对受限，这也推动了后续多尺度特征融合与更精细预测机制的发展。

Liu 等人于 2016 年提出的 SSD^[18] 算法融合了 YOLO 的回归思想与锚点框机制，在多尺度特征图上引入不同尺度与长宽比的默认框，通过密集预测方式直接完成目标的类别分类与边界框回归，从而避免了候选区域生成过程。SSD 通过在骨干网络后构建多层逐级下采样的特征图，使不同层级分别对应不同感受野范围，浅层特征图更适合小目标检测，深层特征图则更关注大目标，从而实现多尺度目标的联合建模。然而，由于其对低层特征语义表达能力有限，小目标检测性能仍存在一定局限。

Lin 等人于 2017 年提出的 RetinaNet^[19] 也是一阶段目标检测的重要代表，其核心创新在于提出焦点损失，用于缓解一阶段检测中正负样本极度不平衡的问题。通过降低易分类负样本的损失权重，模型能够将学习重点集中于困难样本，从而显著提升检测性能。在结构上，RetinaNet 通常结合特征金字塔网络进行多尺度特征建模，并在不同尺度特征图上进行分类与回归预测，实现了速度与精度之间的良好平衡。然而，该方法仍依赖锚点机制，在复杂场景下仍可能存在冗余预测问题。

一阶段方法的核心优势在于其简洁的端到端流程带来的极快推理速度和较小的模型体积，这使得一阶段模型成为农业移动设备端和无人机载平台的理想选择。然而，由于缺少两阶段方法中候选区域的筛选与修正环节，一阶段方法直接面对全图所有位置进行密集预测，在小目标检测和多尺度目标检测任务上精度往往不及两阶段方法。农业场景中普遍存在微小病斑与尺度变化显著的目标，这对一阶段检测模型的特征表达能力提出了更高要求，而其在多尺度目标上的检测精度仍存在一定不

足。

2.1.2 基于 Transformer 的端到端目标检测方法

基于 Transformer 架构的目标检测方法是近年来计算机视觉领域的重要突破，其核心在于将自然语言处理中的自注意力机制与编码器-解码器结构迁移至视觉任务中。DETR (DEtection TRansformer)^[28] 由 Carion 等人于 2020 年提出，将 Transformer 架构引入目标检测任务，突破了传统检测器依赖锚框与非极大值抑制 (NMS) 的处理范式。其核心思想是将目标检测建模为一个集合预测问题，通过端到端的方式直接输出目标集合。其网络结构如图 2.4 所示。

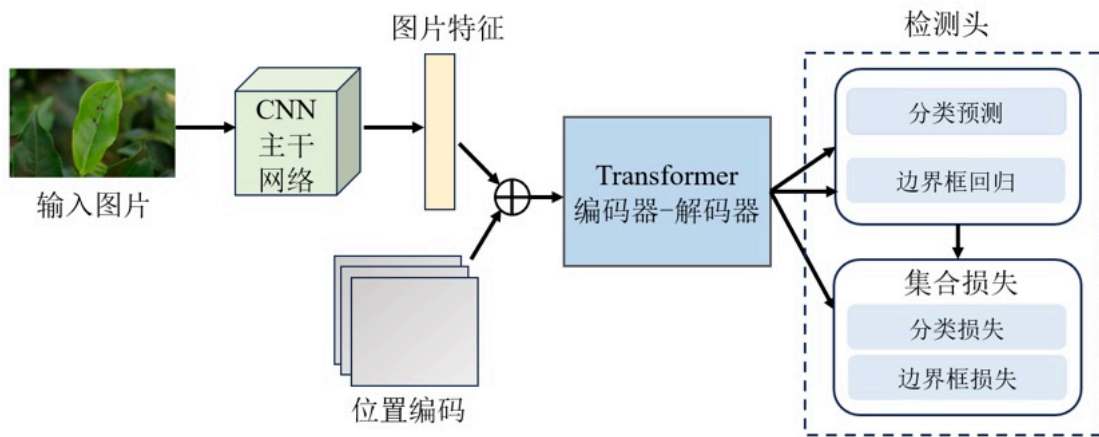


图 2.4 DETR 算法的网络结构

输入图像首先通过卷积神经网络提取二维特征图，并将其展平为一维序列，同时引入位置编码以保留空间位置信息。随后，序列特征被送入 Transformer 编码器，通过多头自注意力机制实现全局上下文建模，使每个位置的特征能够感知全局语义关系。在目标预测阶段，DETR 的解码器引入一组固定数量的可学习对象查询，作为目标的潜在表示，与编码器输出特征进行跨注意力交互，从全局特征中检索目标信息。每个查询最终对应一个预测结果，包括类别标签与边界框坐标。在训练过程中，DETR 采用基于匈牙利匹配算法的一对一匹配策略，将预测结果与真实标注进行最优匹配，从而保证每个真实目标仅对应一个预测结果，避免冗余检测并省去 NMS 后处理。

总体而言，DETR 通过全局自注意力建模与集合预测机制，将目标检测统一为端到端的序列到集合映射问题，实现了结构上的简洁统一，但在训练收敛速度与小

目标检测性能方面仍存在一定局限。然而，由于 **Transformer** 自注意力机制具有较高的计算复杂度，且缺乏对局部空间结构的归纳偏置，原始 **DETR** 在训练初期存在收敛速度较慢的问题，并对大规模训练数据具有较强依赖。同时，其固定数量的对象查询在处理小目标或密集目标场景时，容易出现匹配困难以及特征表达不足的情况。因此，如何在保持端到端建模优势的同时，提高训练效率并增强多尺度特征建模能力，已成为当前研究的重要方向。围绕这一问题，研究者提出了一系列改进方法，其中具有代表性的模型包括 **Deformable DETR**^[3]、**DAB-DETR**^[29] 以及 **DINO**^[30] 等，其中 **DINO** 通过引入去噪训练策略并结合对比学习机制，有效提升了查询匹配稳定性与训练收敛速度，已成为当前 **DETR** 系列中性能较强的代表模型之一。

在上述以 **DETR** 为代表的端到端集合预测框架及其改进方法基础上，部分工作从集合预测范式本身出发，对目标表示与优化方式进行了重新设计，其中具有代表性的模型为 **Sparse R-CNN**^[65]，网络结构如图 2.5 所示。该方法摒弃了传统检测器中依赖密集候选框或锚点机制的设计，直接引入一组固定数量的可学习建议框，并在端到端框架下对其进行迭代优化。在特征交互过程中，每个建议框通过动态实例交互头与骨干网络特征进行信息融合。具体而言，首先利用 **RoIAlign** 从特征图中提取对应的区域特征；随后，将该特征与建议框对应的实例特征进行融合，以生成动态卷积参数，从而实现对 **RoI** 特征的自适应变换。该动态卷积机制能够根据不同目标实例调整特征表达方式，有效提升特征的判别能力。

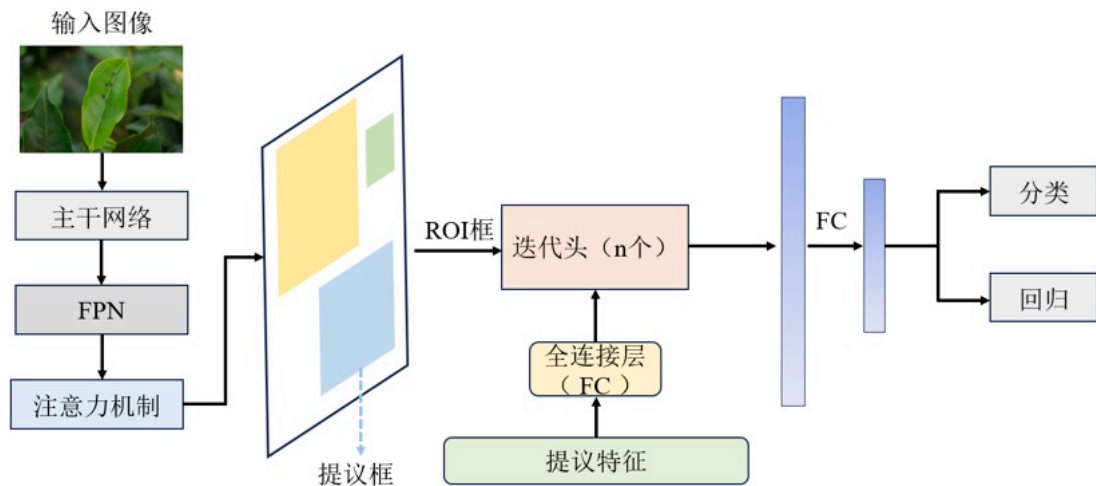


图 2.5 Sparse R-CNN 算法的网络结构

进一步地，**Sparse R-CNN** 通过级联堆叠多个动态交互头，对建议框的位置与特征表示进行逐步迭代更新，使预测结果不断逼近真实目标。该逐级精化过程被统一

整合于端到端框架之中，无需额外的候选区域生成与后处理操作。总体而言，Sparse R-CNN 从稀疏集合建模的角度对目标检测任务进行了重构，其设计理念与 DETR 中基于对象查询的集合预测思想具有内在一致性。尽管其未显式采用标准 Transformer 编解码结构，但通过引入可学习提案与迭代优化机制，在一定程度上弥合了传统检测器与 Transformer 检测器之间的差异。

综合而言，上述基于 Transformer 的目标检测方法主要以端到端集合预测或稀疏建模为核心思想，凭借全局特征建模能力、对锚框先验的摒弃以及端到端的输出形式，为目标检测技术提供了全新的设计空间。然而，这类方法的计算开销、训练数据需求以及对小目标的敏感性仍是其在农业领域广泛应用的制约因素。值得强调的是，各类目标检测框架的最终性能，在很大程度上依赖于模型的特征表达能力，这一共性基础决定了无论采用何种检测范式，特征表示的质量始终是影响检测精度的根本因素。

2.2 植物叶片病害目标检测技术基础

植物叶片病害目标检测是计算机视觉技术在智慧农业领域的重要应用，任务是在自然真实环境中对农作物叶片上的病斑区域进行精准定位与类别识别。随着深度学习技术的蓬勃发展，该任务已逐渐摆脱对人工先验特征的依赖，从传统方法演进而为基于数据驱动的端到端检测模型。然而，农业场景独有的极端复杂性使得该任务在底层特征表征与多尺度感知层面依然面临着严峻的技术挑战。

2.2.1 传统植物叶片病害视觉检测方法

在深度学习方法广泛应用之前，植物叶片病害检测主要依赖于传统图像处理技术与浅层机器学习模型。该类方法通常遵循图像预处理、病斑分割、特征提取和分类识别的基本流程，核心思想是通过人工设计的特征对病斑区域的颜色、纹理及形状等视觉属性进行描述。检测流程如图 2.6 所示。

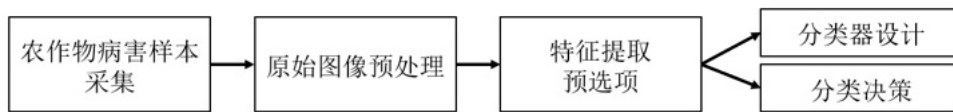


图 2.6 基于传统图像处理的植物叶片病害检测流程图

在图像预处理阶段，为减弱噪声干扰并增强病斑区域的可分性，通常采用中值

滤波、直方图均衡化以及色彩空间转换等方法，以提升病斑与背景之间的对比度^[66]。在此基础上，通过阈值分割、区域生长或边缘检测等技术实现病斑区域的初步定位与提取。

在特征提取阶段，传统方法主要依赖人工构造的低层视觉特征对病斑进行表征。其中，纹理特征方面常采用局部二值模式（LBP）描述局部结构信息，或利用灰度共生矩阵（GLCM）刻画像素间的统计关系；在形状特征方面，则通过面积、周长及不规则度等几何指标对病斑外观进行刻画；此外，还可结合颜色特征以增强对不同病害类型的区分能力。

在分类识别阶段，将上述提取的特征向量输入至浅层机器学习模型中进行判别，如支持向量机（SVM）^[67]、随机森林以及 K 近邻（KNN）等，从而实现病害类别的识别。

尽管该类方法在早期研究中取得了一定进展，但其局限性同样明显。一方面，人工设计特征依赖经验，难以全面刻画复杂多变的病斑形态；另一方面，该方法对光照变化、复杂背景以及尺度变化较为敏感，导致其在真实田间环境中的鲁棒性和泛化能力较弱。因此，传统方法逐渐难以满足实际农业应用的需求。

2.2.2 深度学习下的植物叶片病害视觉检测方法

随着深度学习技术的发展，基于数据驱动的端到端检测方法逐渐成为主流。与依赖人工特征的传统方法相比，深度学习模型能够自动学习多层次特征表示，大幅提升了病害检测的鲁棒性与精度。该流程主要包括数据准备、模型构建、训练、评估与识别等关键环节，各主要步骤及其对应的具体操作如表 2.1 所示。根据特征建模方式的不同，现有病害检测方法可分为基于 CNN 的方法与基于 Transformer 的方法两类。

（1）基于 CNN 的植物叶片病害检测方法

CNN 凭借局部连接、权值共享和层级特征提取等特点，长期以来一直是植物叶片病害检测任务中的主流技术路线。CNN 能够从原始图像中逐层提取由低级纹理至高级语义的层次化特征，有效表征叶片病斑在颜色、纹理、边缘及形状等方面的视觉差异。因此，大量植物叶片病害检测研究均建立在 CNN 目标检测框架基础之上，并逐步形成了以两阶段检测器和一阶段检测器为核心的技术体系。

两阶段检测方法以 R-CNN 系列^[4,10,12]及 Mask R-CNN^[13]等为代表，其核心思想

表 2.1 基于深度学习的农作物叶片病害检测主要步骤

主要步骤	具体操作或方法
数据准备	建立农作物病害图像数据库；开展图像预处理（包括样本筛选、数据扩充及降噪处理等）；进行病害信息标注（涵盖病斑类别、空间位置、边界轮廓等）；划分数数据集（一般按照训练集、验证集、测试集进行划分，常用比例有 8:1:1 或 7:2:1）
模型构建	针对农作物叶片病害识别任务的特点，选取合适的深度学习框架和具体模型
模型训练	采用标注好的病害图像样本对网络进行训练和优化
性能评估	选取科学的评价指标对模型效果进行检验，如识别准确率、精确率等
病害识别	将待检测的病害图像输入至已训练完成的模型中进行预测，并对检测结果进行可视化呈现

是先生成候选区域，再对候选区域进行精细分类和边界框回归。其中，Faster R-CNN 通过引入区域生成网络 RPN 实现候选框的端到端生成，显著提升了检测效率与精度，在植物叶片病害检测中被广泛应用于病斑定位、叶片病害类别识别以及复杂背景下的病害区域提取等任务。相较于早期方法，两阶段检测器通常具有更高的定位精度和更强的小目标识别能力，尤其适合病斑边界模糊、形态不规则、类别间差异细微的病害识别场景。此外，Mask R-CNN 还可进一步实现病斑区域的实例分割，为病害面积估计和严重程度分析提供支持。然而，这类方法由于检测流程相对复杂，推理速度较慢，难以完全满足田间移动终端和实时监测系统的应用需求。

为提高检测效率，一阶段检测方法逐渐成为研究热点。该类方法以 YOLO^[15] 系列、SSD^[18]、RetinaNet^[19] 和 EfficientDet^[68] 等为代表，直接在特征图上进行类别预测和边界框回归，将目标定位与分类统一到单一网络框架中，具有结构简洁、推理速度快和部署成本低等优势。其中，YOLO 系列在检测精度、速度和平衡性方面持续优化，已成为植物叶片病害实时检测任务中最常用的模型之一。SSD 通过多尺度特征图进行预测，在中等尺度目标检测中表现良好；RetinaNet 则针对一阶段检测器中正负样本极度不平衡的问题提出焦点损失，提高了困难样本的学习能力；EfficientDet 通过复合缩放策略和双向特征金字塔网络实现了精度与效率的良好折中。表 2.2 从实现方式、优缺点及代表算法四个维度，系统对比了两阶段与一阶段目标检测算法

在植物叶片病害检测领域的差异。

表 2.2 两阶段与一阶段目标检测算法的对比

对比维度	两阶段目标检测算法	一阶段目标检测算法
实现方式	基于 RPN 生成候选框，再完成分类与坐标回归	基于锚框密集预测，直接输出目标类别与位置
优势	检测精度较高，复杂田间环境下检测鲁棒性相对强	检测速率快，模型参数量少，适合边缘设备部署
局限	推理速度较慢，部署成本较高	密集小目标场景易漏检，同体量下精度略低于两阶段方法
代表算法	Faster R-CNN、Mask R-CNN	SSD、YOLO

针对植物叶片病害图像中病斑尺度差异大、分布离散、小目标占比较高等问题，多尺度特征建模也成为 CNN 检测方法的重要改进方向。研究者通常在基础检测框架中引入 FPN^[2]、PANet^[33]、BiFPN^[68] 等结构，以增强模型对不同层级语义信息的融合能力，从而提升对小病斑、密集病斑和早期病害症状的检测性能。此外，空洞卷积、可变形卷积、注意力模块以及 MobileNet^[69]、GhostNet^[70] 等轻量化骨干网络也被广泛用于改进 CNN 病害检测模型，以适应田间场景下复杂背景、遮挡干扰和边缘设备部署等实际需求。尽管 CNN 方法在植物叶片病害检测中取得了显著进展，但由于其本质上依赖局部卷积操作，感受野需通过网络加深和下采样过程逐步扩展，因此在建模远距离依赖关系和全局上下文信息方面存在天然局限。对于病斑区域较小、背景干扰较强或病害表征高度依赖叶片整体结构关系的场景，单纯依靠卷积特征往往难以充分捕捉全局语义信息，这在一定程度上限制了 CNN 方法的进一步性能提升。

(2) 基于 Transformer 的植物叶片病害检测方法

植物叶片病害检测的关键难点在于病斑尺度变化大、形状纹理多样，且常与叶片结构及复杂背景混合。传统卷积神经网络受限于局部感受野，难以建模病斑与叶片整体以及背景之间的长程依赖关系，在复杂场景下易产生误判。Transformer 凭借自注意力机制的全局建模能力，在处理植物叶片病害检测中复杂背景干扰、病斑与叶片整体结构关联以及跨区域语义交互等问题时展现出较强优势。

近年来，以 DETR^[28] 为代表的端到端检测框架被引入植物叶片病害检测研究。DETR 通过 Transformer 编码器-解码器结构对图像特征进行全局交互，并直接输出

目标框与类别，其天然的大感受野有助于模型在叶片图像中建立病斑区域与叶脉走向、叶片轮廓以及背景环境之间的语义关联，从而提高对背景干扰的鲁棒性。针对 DETR 小目标检测弱、收敛慢等问题而提出的 Deformable DETR^[3]、Conditional DETR、DINO^[30]等改进模型，进一步提升了在病害早期小斑点检测、跨尺度病斑定位等方面的性能。同时，以 Swin Transformer^[71]为代表的层次化视觉 Transformer 作为骨干网络，也被广泛应用于病害特征提取，其滑动窗口注意力机制在保持全局视野的同时兼顾局部细节，能更精细地刻画病斑纹理与边缘。

然而，Transformer 模型在植物叶片病害检测中的应用仍面临挑战。首先，标准 Transformer 缺乏 CNN 固有的局部归纳偏置，对训练数据的规模和质量要求较高，而农业病害数据集常存在样本有限、类别不均衡、标注成本高等问题，导致纯 Transformer 模型容易过拟合或泛化能力不足。其次，Transformer 在小目标检测、精确边界回归方面仍不及充分优化的卷积检测器，这在早期微小病斑识别中尤为明显。此外，部分 Transformer 模型参数量大、推理速度慢，限制了其在田间移动设备或边缘端的部署。

基于此，当前研究逐渐呈现出 CNN 与 Transformer 融合的趋势。CNN 负责提取局部病斑的纹理细节和边缘响应，提供稳定的归纳偏置；Transformer 则补充全局上下文建模与跨区域语义交互能力，增强对复杂背景和病斑整体结构的理解。这种混合架构既能稳定捕捉早期病斑的微弱特征，又能有效利用叶片全局结构辅助判别，已在多种作物病害识别任务中展现出优于单一模型的精度与泛化能力，成为该领域的重要研究方向。为更加直观地分析不同方法的特点，表 2.3 对植物叶片病害检测的不同方法进行了对比总结。

表 2.3 不同植物叶片病害检测方法对比

方法类别	特征类型	优势	局限性
传统方法	人工设计特征	计算开销低、实现简单	泛化能力弱、依赖人工经验
CNN 方法	局部层次特征	特征表达能力强、精度高	全局建模能力有限
Transformer 方法	全局注意力特征	全局建模能力强、结构灵活	数据需求大、小目标敏感性不足

2.3 多尺度特征表示与融合技术基础

2.3.1 特征金字塔及其改进

在目标检测任务中，目标尺度变化是影响检测性能的重要因素，尤其在植物叶片病害场景中，病斑区域通常同时包含大面积病变与微小斑点，呈现显著的尺度差异。传统单尺度检测方法难以同时兼顾不同尺度目标的特征表达需求，因此，多尺度特征表示成为提升检测性能的重要研究方向。

在深度卷积网络中，随着逐层下采样，特征逐渐形成不同语义层次：浅层特征具有较高空间分辨率，能够保留丰富的纹理与边缘信息，但语义表达能力较弱；深层特征语义表达能力较强，但空间分辨率较低，不利于小目标检测。因此，如何在增强语义表达的同时保留细粒度空间信息，成为多尺度目标检测的关键问题。为解决上述问题，Lin 等人提出了特征金字塔网络（Feature Pyramid Network, FPN）^[2]。FPN 通过构建自顶向下的特征传递路径并结合横向连接，实现不同层级特征的融合，从而生成具有较强语义表达能力的多尺度特征表示，其基本结构如图 2.7 所示。

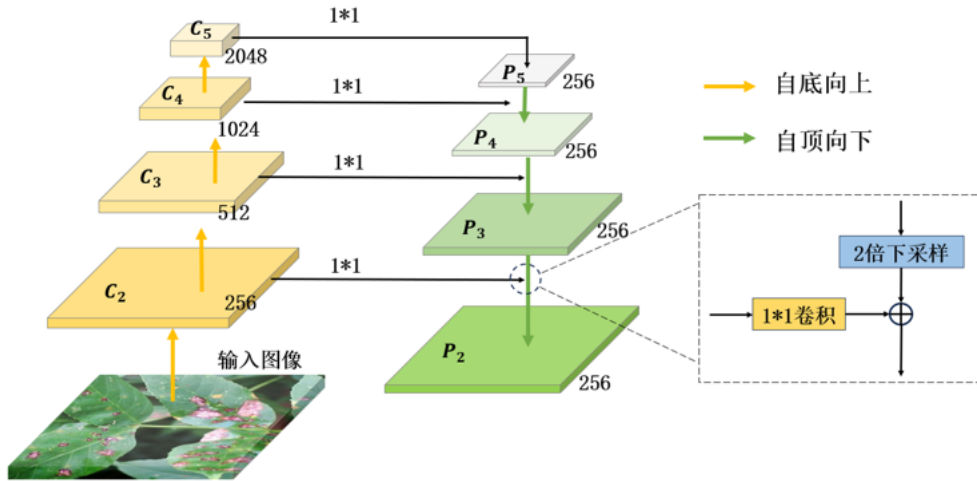


图 2.7 FPN 网络结构示意图

具体而言，FPN 首先通过骨干网络提取多个尺度的特征图 $\{C_2, C_3, C_4, C_5\}$ 。随后，从最高层特征开始逐层进行上采样，并与对应层的浅层特征通过横向连接进行融合，从而得到新的特征层 $\{P_2, P_3, P_4, P_5\}$ 。其融合过程可表示为：

$$\begin{cases} P_5 = \text{Conv}_{1 \times 1}(C_5) \\ P_l = \text{Up}(P_{l+1}) + \text{Conv}_{1 \times 1}(C_l), \quad l = 4, 3, 2 \end{cases} \quad (2.1)$$

其中 C_l 表示骨干网络第 l 层特征图, $\text{Up}(\cdot)$ 表示上采样操作, $\text{Conv}_{1 \times 1}$ 表示用于通道对齐的 1×1 卷积映射, 而 P_l 为融合后的金字塔特征层。通过这种自顶向下的特征融合机制, FPN 能够将高层语义信息逐级传递至低层特征, 从而在不同尺度上获得语义一致的特征表示。这一结构有效结合了深层语义信息与浅层空间细节, 使检测器能够在多个尺度上进行预测, 显著提升了模型对小目标与多尺度目标的检测能力。

在 FPN 基础上, 研究者进一步提出了多种改进结构。例如, PANet (Path Aggregation Network)^[33] 在 FPN 自顶向下路径之外增加了一条自底向上的增强路径, 从而缩短低层特征到高层特征之间的信息传递距离; BiFPN (Bidirectional Feature Pyramid Network)^[68] 通过构建高效的双向融合结构, 并引入可学习权重实现不同尺度特征的自适应加权融合; NAS-FPN^[34] 利用神经架构搜索自动优化跨尺度连接方式, 以获得更加高效的特征融合拓扑结构; 此外, AugFPN^[72] 通过引入残差增强与一致性约束机制, 进一步缓解了跨层特征之间的语义差异问题。

尽管上述方法在多尺度特征表示方面取得了显著进展, 但其融合方式大多依赖固定拓扑结构和简单算子操作, 缺乏对跨尺度特征之间语义一致性与空间对齐关系的深入建模。因此, 如何设计更加灵活且表达能力更强的特征融合机制, 仍然是目标检测领域的重要研究方向。

2.3.2 多尺度特征融合策略

多尺度特征融合的核心目标是在统一表示空间中整合不同层级的特征信息, 使其在空间分布与语义表达上实现互补与一致。一般而言, 浅层特征具有较高空间分辨率, 能够保留丰富的纹理与边缘信息, 但语义表达能力较弱; 深层特征则具有更强的语义表达能力, 但由于多次下采样, 其空间分辨率较低, 不利于小尺度目标的精细建模。因此, 如何有效融合不同层级特征以实现优势互补, 成为目标检测中的关键问题。

传统多尺度融合方法通常采用逐元素相加或通道拼接等简单操作进行特征整合。此类方法虽实现简洁、计算开销较低, 但由于未充分考虑不同尺度特征在语义层级与分布特性上的差异, 容易导致信息冲突或特征表达不足。因此, 围绕多尺度特征融合机制出现了多种改进策略。

(1) 加权融合机制

为了避免简单相加操作带来的信息干扰问题, 可以通过引入可学习权重对不同

尺度特征进行自适应调节。例如,自适应空间特征融合(Adaptive Spatial Feature Fusion, ASFF)^[73]通过在像素级别学习不同尺度特征的融合权重,使网络能够根据输入图像内容动态调整各尺度特征的重要性。其融合过程可以表示为:

$$F = \alpha_1 F_1 + \alpha_2 F_2 + \alpha_3 F_3 \quad (2.2)$$

其中 F_1, F_2, F_3 表示来自不同尺度的特征图,而 $\alpha_1, \alpha_2, \alpha_3$ 为通过网络学习得到的权重系数。此外, Huang 等人提出的自适应双级加权机制(Adaptive Dual-level Weighting Mechanism, ADWM)^[74]通过在通道维度与层级维度同时引入加权机制,实现对不同尺度特征贡献度的精细化建模,从而进一步提升特征融合的有效性。

(2) 跨尺度对齐融合

由于不同层级特征在空间分辨率与感受野方面存在差异,直接进行特征融合容易引入空间位置偏移问题。为缓解这一问题,研究者提出了多种特征对齐方法。其中,可变形卷积(Deformable Convolution)^[75]通过引入可学习偏移量,使卷积采样位置能够根据目标形状进行自适应调整,其卷积形式可表示为:

$$y(p_0) = \sum_{k=1}^K w_k \cdot x(p_0 + p_k + \Delta p_k) \quad (2.3)$$

其中, p_0 表示输出特征图中的当前位置, p_k 表示标准卷积核的第 k 个采样位置,即规则卷积中的固定偏移, Δp_k 为对应位置的可学习偏移量, w_k 为卷积核权重, $x(\cdot)$ 与 $y(\cdot)$ 分别表示输入特征图与输出特征图, K 为卷积核采样点的总数。通过引入可学习偏移量,卷积核的采样位置能够根据目标结构进行自适应调整,从而提升对目标区域的空间对齐能力。

进一步地,可变形注意力(Deformable Attention)^[3]将该思想扩展至注意力机制中,通过在多尺度特征图上进行稀疏采样,实现高效的跨尺度信息交互。此外, Align-Det^[76]通过引入专门的特征对齐模块,对不同层级特征进行重采样与空间校正,从而减少跨层特征之间的空间错位问题。

(3) 基于 Transformer 自注意力的融合机制

在多尺度特征融合过程中,传统方法通常依赖固定结构或局部操作,难以建模不同尺度特征之间的长距离依赖关系。为此,研究者引入自注意力机制,通过显式建模特征之间的全局关系,实现跨尺度信息的自适应聚合。自注意力机制最早由 Vaswani 等人于 2017 年提出^[27],并在自然语言处理领域取得了显著成功,随后被引入计算机

视觉任务，并逐渐发展为以 **Vision Transformer (ViT)**^[77] 为代表的视觉建模框架。

Transformer 通过计算特征之间的相关性权重，对不同位置的信息进行加权聚合，从而在全局范围内捕获长距离依赖关系。其基本计算过程可表示为：

$$\text{Attention}(Q, K, V) = \text{Softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (2.4)$$

其中， Q (Query)、 K (Key) 和 V (Value) 分别表示查询、键和值矩阵， QK^T 表示查询与键之间的相似度计算结果， $\sqrt{d_k}$ 为缩放因子，用于防止内积过大导致梯度不稳定， $\text{Softmax}(\cdot)$ 用于将相似度归一化为权重分布，最终通过与 V 的加权求和得到输出特征。

在多尺度场景中，自注意力机制能够跨越不同分辨率特征图，对各尺度信息进行统一建模，实现跨尺度信息的全局交互，从而提升特征融合的充分性与全局一致性。总体而言，基于自注意力的融合机制能够突破卷积操作的局部感受野限制，在全局范围内建模多尺度特征之间的依赖关系。然而，该类方法通常计算开销较大，如何在保证建模能力的同时提高计算效率，仍是当前研究的重要方向。

多尺度特征融合技术有效提升了不同层级特征之间的信息互补能力，然而，复杂农业场景中，植物叶片病害区域往往具有尺度差异大、形态不规则以及边界模糊等特点，传统融合方法仍难以充分捕捉其细粒度结构特征。因此，如何在多尺度融合过程中进一步强化特征表达能力，仍然是当前研究的重要方向。

2.4 特征增强技术基础

在目标检测任务中，仅依赖多尺度特征融合往往难以充分提升特征表达能力。由于复杂背景、目标形态变化以及尺度差异等因素的影响，部分关键特征在特征提取过程中容易被削弱甚至被背景信息所掩盖。因此，研究者提出了多种特征增强机制，通过对原始特征进行重加权或上下文建模，使网络能够更加关注与目标相关的重要信息。

2.4.1 卷积注意力机制

注意力机制最初来源于自然语言处理领域，其核心思想是通过对不同特征分配不同权重，使模型能够更加关注关键信息。在卷积神经网络中，注意力机制通常以轻量级模块的形式嵌入网络结构，通过对特征进行自适应重标定，从而增强有效信

息并抑制无关特征。根据建模维度的不同，卷积注意力机制主要可分为通道注意力、空间注意力以及混合注意力三类。

(1) 通道注意力

通道注意力通过建模不同通道之间的重要性差异，对各通道特征赋予不同权重。典型方法为 SENet (Squeeze-and-Excitation Network)^[41]。其核心思想是通过全局平均池化压缩空间信息，然后利用全连接层学习通道之间的非线性依赖关系。其过程可表示为：

$$s_c = \sigma(W_2 \delta(W_1 z_c)) \quad (2.5)$$

其中 z_c 表示通过全局平均池化获得的通道描述向量， W_1 和 W_2 为可学习参数， $\delta(\cdot)$ 表示 ReLU 函数， $\sigma(\cdot)$ 为 Sigmoid 函数，而 s_c 为通道权重。

在此基础上，ECA-Net (Efficient Channel Attention)^[78]通过局部跨通道交互替代全连接结构，在避免维度压缩的同时显著降低计算复杂度。

(2) 空间注意力

空间注意力机制关注特征图中不同空间位置的重要性，通过生成空间权重图来突出关键区域。通常通过对通道维度进行池化操作获得空间描述信息，然后利用卷积操作生成空间注意力权重。其计算过程可表示为：

$$M_s(F) = \sigma(f^{7 \times 7}([AvgPool(F); MaxPool(F)])) \quad (2.6)$$

其中 F 为输入特征图， $f^{7 \times 7}$ 表示卷积操作， $M_s(F)$ 为生成的空间注意力图。

(3) 混合注意力

混合注意力同时结合通道注意力与空间注意力，以实现更加全面的特征增强。CBAM (Convolutional Block Attention Module)^[42]通过串联通道注意力与空间注意力模块，对特征进行联合建模。此外，坐标注意力机制 (Coordinate Attention, CA)^[79]通过引入方向感知的位置编码，在保持通道依赖建模能力的同时保留精确的空间位置信息，从而进一步提升特征表达能力。

总体而言，卷积注意力机制通过对局部特征进行自适应加权，在较低计算开销下有效提升了特征表达能力，因此已成为目标检测网络中常用的特征增强模块。

2.4.2 上下文建模方法

除了注意力机制之外，引入上下文信息也是提升特征表达能力的重要研究方向。在复杂视觉场景中，目标的判别往往依赖于其与周围环境之间的语义关联，例如背景纹理、邻近目标以及整体场景结构等信息均可为目标识别提供辅助依据。围绕上下文信息的有效建模，已有研究从多个角度展开探索，形成了多种具有代表性的技术路径。一种典型方法是通过多尺度上下文聚合来增强特征表示。例如，PSPNet^[80]提出的金字塔池化模块通过在多个尺度上对特征图进行池化操作，然后将不同尺度的上下文信息进行融合，从而在保持空间分辨率的同时获得更大的感受野。类似地，ASPP^[81]通过在不同膨胀率的空洞卷积下提取多尺度上下文信息，使网络能够在多个感受野范围内捕获目标与背景之间的关系，该方法在语义分割与目标检测任务中得到广泛应用。

另一类重要方法是基于全局关系建模的上下文学习机制。其中，Non-local Network^[82]通过计算特征图中任意两个位置之间的相似度，实现全局信息交互，其基本计算过程可表示为：

$$y_i = \frac{1}{C(x)} \sum_j f(x_i, x_j) g(x_j) \quad (2.7)$$

其中 x_i 与 x_j 表示特征图中不同位置的特征向量， $f(\cdot)$ 为相似度函数， $g(\cdot)$ 为特征变换函数， $C(x)$ 为归一化系数。通过这种方式，模型能够在全局范围内聚合上下文信息，从而增强对复杂场景结构的理解能力。

在此基础上，随着自注意力机制的发展，基于 Transformer 的上下文建模方法逐渐成为新的研究热点。其中，Vision Transformer 将图像划分为固定大小的图像块序列，并通过全局自注意力机制对其进行建模，使网络能够直接捕获长距离依赖关系，从而显著增强特征的全局语义表达能力。然而，标准 Transformer 在高分辨率视觉任务中计算复杂度较高，且缺乏多尺度建模能力。为此，Swin Transformer^[71]通过引入分层结构与滑动窗口注意力机制，在降低计算开销的同时兼顾局部建模能力，并能够构建多尺度特征表示，使其更适用于目标检测等密集预测任务。相较于传统卷积神经网络依赖局部感受野逐层扩展的建模方式，基于 Transformer 的方法能够在更大范围内捕获特征之间的全局依赖关系，从而为复杂场景下的目标检测提供更加丰富且具有判别性的特征表示。

综上所述，上下文建模方法通过多尺度信息聚合与全局依赖关系建模，有效弥补了传统卷积神经网络在长距离语义关联表达方面的不足。相关研究涵盖了多尺度上下文融合、全局关系建模以及基于自注意力的建模方法，持续提升特征表达的范围与能力，为复杂场景下的目标检测提供了更加丰富且具有判别力的上下文信息。

2.5 数据集与评价指标

2.5.1 数据集介绍

植物叶片病害检测数据集的构建与公开，为深度学习技术在农业领域的应用奠定了重要基础。根据图像采集环境的不同，现有数据集通常可分为实验室受控环境数据集与真实田间场景数据集两类。实验室环境数据集一般在光照均匀、背景单一的条件下采集，图像质量较高且噪声干扰较小，常用于算法有效性的初步验证。

相比之下，真实田间场景数据集通常包含复杂背景、光照变化以及遮挡与重叠等干扰因素，更贴近实际农业生产环境，对模型的泛化能力提出了更高要求。因此，本文主要选用三种真实场景数据集作为实验基础，以系统评估所提方法在复杂环境下的检测性能。

(1) PlantDoc 数据集

PlantDoc 数据集^[83]由 Singh 等于 2020 年发布，是首个面向真实田间场景的植物叶片病害检测数据集。该数据集包含 2,598 张图像，涵盖 13 种植物物种（苹果、甜椒、蓝莓、樱桃、玉米、葡萄、桃子、土豆、覆盆子、大豆、南瓜、草莓和番茄），共计 27 个类别，其中 10 个为健康类别，17 个为病害类别。所有图像均在自然环境中拍摄，具有复杂的现实背景和多变的光照条件，适用于目标检测与分类任务。在本实验中，该数据集按 8:2 的比例随机划分为训练集和验证集，其中训练集包含 2,078 张图像，验证集包含 520 张图像。PlantDoc 数据集代表图像和数据集类别统计如图 2.8 和图 2.9 所示。

(2) 番茄叶片疾病数据集

番茄叶片疾病数据集 (Tomato Leaf Disease, TLD)^[84]由 Roboflow 于 2023 年发布，是一个专注于番茄病害的公开数据集。该数据集包含 4,936 张番茄叶片图像，涵盖九个类别，其中包含 8 个疾病类和 1 个健康状态，分别代表不同的病害状态与健康状态。数据集按 8:1:1 的比例随机划分为训练集、验证集和测试集。TLD 数据集代表图

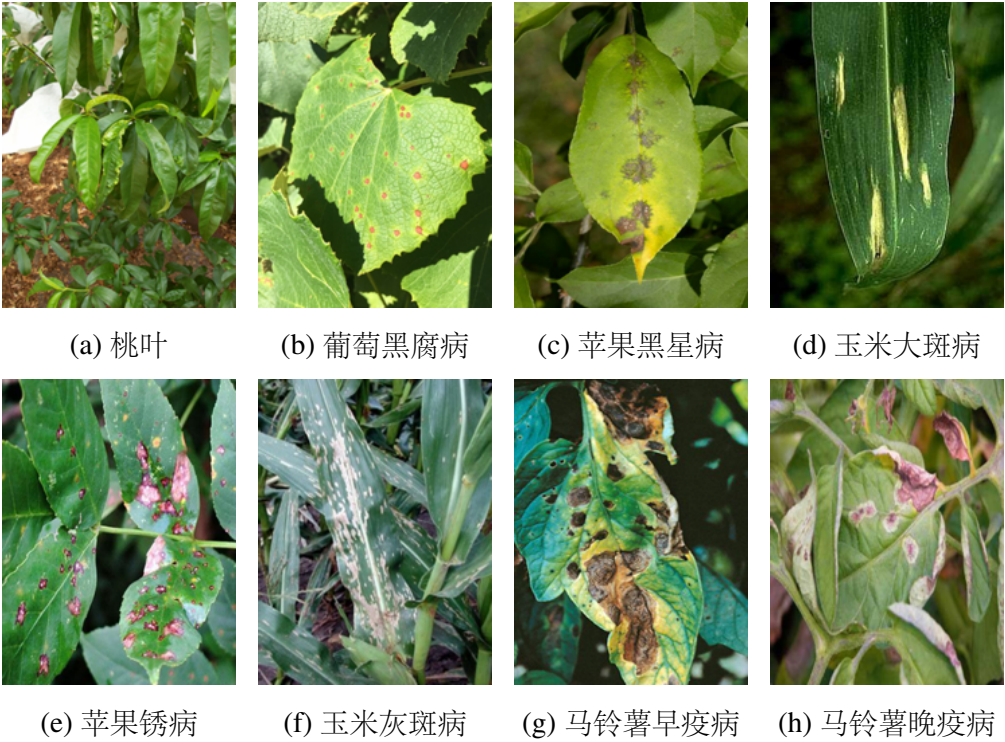


图 2.8 PlantDoc 数据集中的叶片图像示例

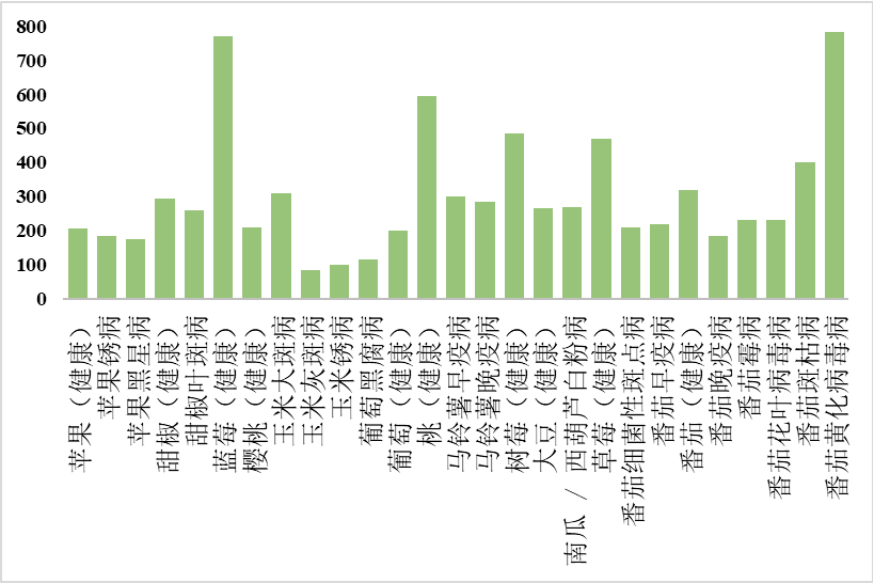


图 2.9 PlantDoc 数据集中不同类别的标注统计

像如图 2.10 所示；各疾病类别和标注统计如表 2.4 所示。



图 2.10 TLD 数据集中的叶片图像示例

表 2.4 TLD 数据集中标注数量统计

病害名称	标注数量
早疫病	1084
晚疫病	1450
花叶病毒病	1382
曲叶病毒病	1774
潜叶蝇危害	1091
叶霉病	1441
斑枯病	1343
红蜘蛛危害	916

(3) 水果检测数据集

为验证本文特征增强机制的跨场景通用性，补充引入水果检测数据集（Fruit Detection, FD）^[85]作为辅助验证数据集，用于评估模型在跨场景下的泛化能力。FD 数据集由 Tyagi 于 2023 年发布，面向日常场景下的水果检测任务。其所有图像均源自 COCO^[86]和 PASCAL VOC^[87]等广泛使用的基准数据集，并经过重新标注形成六种水

果类型，分别为苹果、葡萄、菠萝、橙子、香蕉和西瓜。该数据集共包含 8,479 张图像，图像内存在因光照变化、遮挡、尺度差异及背景杂乱等因素导致的显著差异。数据集按 8:1:1 的比例随机划分为训练集、验证集和测试集。FD 数据集代表图像如图 2.11 所示。



图 2.11 FD 数据集中的图像示例

2.5.2 评价指标

目标检测任务的性能评估通常需要综合考虑分类准确性、定位精度以及模型的计算效率等多个方面，常用评价指标包括交并比（Intersection over Union, IoU）、精确率（Precision）、召回率（Recall）、平均精度（Average Precision, AP）以及均值平均精度（mean Average Precision, mAP）等。

交并比（IoU）是评估目标检测算法定位准确性的关键指标，主要用于衡量预测边界框与真实标注框之间的重叠程度。该指标通过计算两个区域交集面积与并集面积的比值来反映检测结果的准确性。IoU 的取值范围为 $[0, 1]$ ，其值越大表示预测框与真实框的重合程度越高，从而说明检测结果越准确。其计算公式如下：

$$\text{IoU} = \frac{A \cap B}{A \cup B} \quad (2.8)$$

其中， A 表示预测框（Predicted Bounding Box）， B 表示真实标注框（Ground Truth Bounding Box）， $A \cap B$ 表示两者的交集区域面积， $A \cup B$ 表示两者的并集区域面积。

在实际评估过程中，当预测框与真实框类别一致且 IoU 大于设定阈值（如 0.5）时，该预测被视为真阳性（True Positive, TP），否则为假阳性（False Positive, FP），未被检测到的真实目标则记为假阴性（False Negative, FN），并通常采用基于置信度排序的一对一匹配策略完成预测框与真实框之间的分配。

在此基础上，精确率和召回率分别用于衡量模型的查准能力与查全能力，其定

义如下：

$$\text{Precision} = \frac{TP}{TP + FP} \quad (2.9)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (2.10)$$

其中， TP 、 FP 和 FN 分别表示真阳性、假阳性和假阴性的数量。为了综合评估模型在不同置信度阈值下的检测性能，引入平均精度（AP）指标。

具体而言，通过对预测结果按置信度排序，逐步计算对应的精确率与召回率，从而构建精确率-召回率（P-R）曲线，并通过对该曲线进行数值积分或插值计算得到 AP，其形式可表示为：

$$\text{AP} = \int_0^1 P(R) dR \quad (2.11)$$

在实际计算中，通常采用离散采样与插值方法对上述积分进行近似。

对于多类别检测任务，各类别 AP 的平均值定义为均值平均精度（mAP），用于整体评估模型性能。在 COCO 评价标准中，mAP 通常定义为在 IoU 阈值从 0.5 到 0.95（步长为 0.05）范围内各 AP 的平均值（记为 $\text{AP}_{50:95}$ ），此外还常采用 AP_{50} 和 AP_{75} 分别评估模型在不同定位精度要求下的检测性能，并通过 AP_s 、 AP_m 和 AP_l 衡量模型在不同尺度目标上的检测性能，其中小目标定义为 $A < 32^2$ ，中等目标为 $32^2 \leq A < 96^2$ ，大目标为 $A \geq 96^2$ ， A 表示目标边界框面积。

在本文实验中，采用 COCO 评价标准下的 mAP 及 AP_{50} 、 AP_{75} 、 AP_s 、 AP_m 和 AP_l 作为主要评价指标，以全面衡量模型在农业病害检测任务中的性能表现。

第三章 基于特征交互与协调融合的植物叶片病害检测算法

在真实田间环境中，植物叶片病害目标通常具有显著的尺度变化和复杂的形态结构，早期病斑仅占少量像素，而成熟病斑可能覆盖大面积叶片，并伴随细粒度纹理变化与复杂背景干扰。因此，检测模型需要具备较强的语义表达能力并有效保留空间细节信息，以实现多尺度病害目标的准确表征。然而，现有检测框架多依赖特征金字塔网络构建多尺度特征表示，其单向信息传播与固定融合策略在复杂农业场景中仍存在一定局限。针对上述问题，本章从检测模型编码器结构优化的角度出发，提出一种基于多尺度分层递归对齐与全局协同的植物叶片病害检测框架（Hierarchical Recursive Alignment and Global Collaborative Fusion Network, HRAGNet）。

HRAGNet 构建了一种多尺度特征增强编码器，通过三个功能互补且逐级递进的模块对特征金字塔结构进行优化。交互式增强模块（Interactive Enhancement Module, IEM）通过递归的特征交互机制增强层级特征之间的语义协同；跨尺度语义对齐模块（Cross-scale Semantic Alignment Module, CSAM）利用可学习的语义流场实现多尺度特征的像素级空间对齐；全局上下文聚合融合模块（Global Context Aggregation Fusion, GCAF）通过通道注意力机制对对齐后的特征进行自适应融合，从而构建兼具局部细节与全局语义信息的增强特征金字塔。通过上述设计，HRAGNet 能够有效提升检测模型在复杂农业场景中对多尺度叶片病害目标的表达能力。

3.1 模型总体设计

3.1.1 整体架构与模型输入输出

HRAGNet 以 Sparse R-CNN^[65] 为检测基础框架，整体采用编码器-解码器结构，由多尺度特征增强编码器与动态检测解码器两部分组成。其中，编码器通过递归增强、空间对齐与全局协同机制提升多尺度特征表达能力，解码器则基于迭代式预测策略逐步优化目标类别与边界框位置，从而实现高精度目标检测。具体模型架构如图 3.1 所示。

输入图像 I 首先经 ResNet-50 与 FPN 骨干网络提取多层级特征，构建基础特征金字塔 $\{C_1, C_2, C_3, C_4\}$ ，各层通道数统一为 256，对应空间分辨率分别为输入图像的

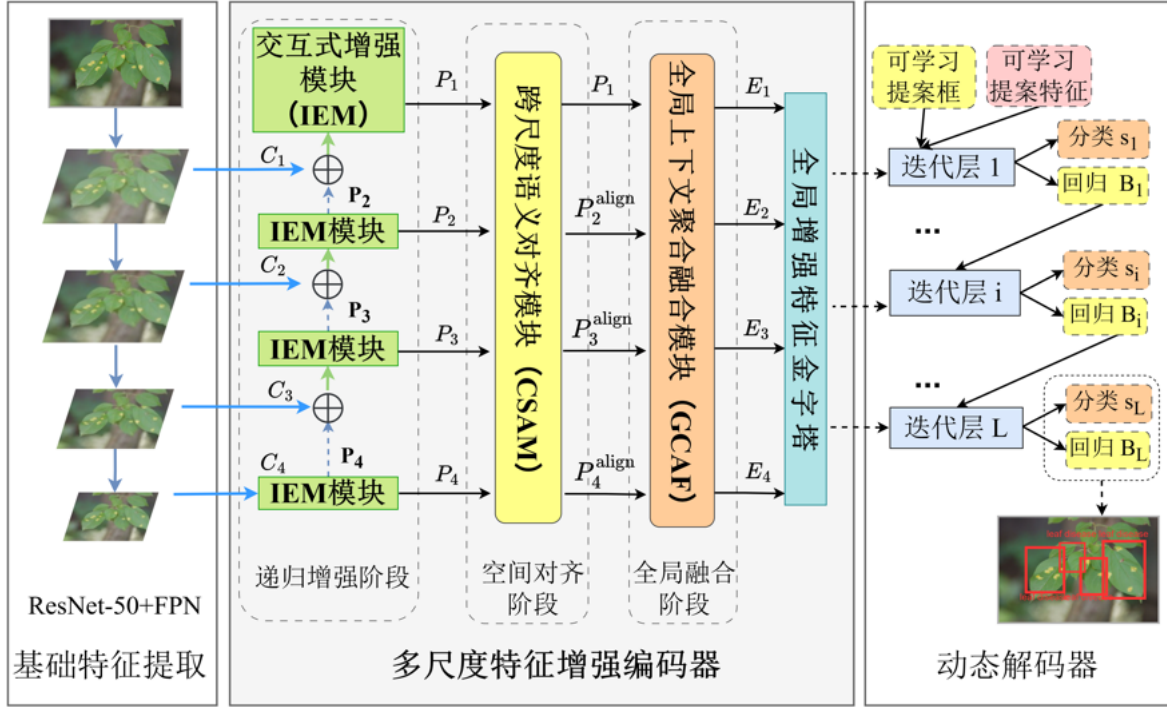


图 3.1 HRAGNet 模型整体框架图

1/4、1/8、1/16 和 1/32。随后，该特征金字塔被送入多尺度特征增强编码器，依次经过交互式增强模块（IEM）、跨尺度语义对齐模块（CSAM）以及全局上下文聚合融合模块（GCAF），分别从特征交互、空间对齐与全局上下文建模三个层面对多尺度特征进行协同优化，最终生成语义一致性与细粒度表征能力兼具的全局增强特征金字塔 $\{E_1, E_2, E_3, E_4\}$ 。最后，该增强特征被输入动态检测解码器，通过多轮迭代预测不断细化目标类别与边界框，实现对叶片病害目标的精准定位与识别。

3.1.2 多尺度特征增强编码器

多尺度特征增强编码器是 HRAGNet 的核心创新部分，其整体处理流程可概括为多尺度递归增强、跨尺度空间对齐和全局融合三个阶段，最终生成全局增强特征金字塔 $\{E_1, E_2, E_3, E_4\}$ 。其具体实现如图 3.2 所示。

(1) 基于 IEM 模块的多尺度递归增强阶段

传统特征金字塔网络主要采用自顶向下的单向路径与逐元素相加的方式进行多尺度融合，存在高层语义信息逐层衰减以及缺乏底层细节反馈的问题。为此，本章引入基于 IEM 模块的递归增强机制，通过跨层级特征交互完成多尺度信息的双向传播与逐级优化，从而实现对多尺度特征图充分的语义信息增强与细节信息补充。

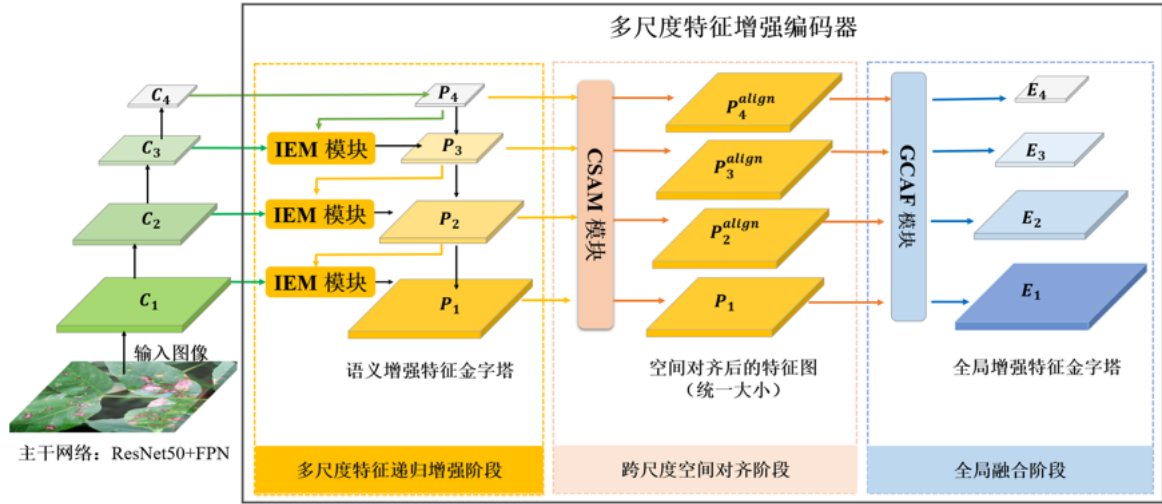


图 3.2 多尺度特征增强编码器结构图

给定基础特征金字塔 $\{C_1, C_2, C_3, C_4\}$ ，多尺度递归增强过程可表示为：

$$\begin{cases} P_4 = C_4 \\ P_i = \text{IEM}(C_i \oplus \text{Upsample}(P_{i+1})), \quad i = 3, 2, 1 \end{cases} \quad (3.1)$$

其中， $\oplus$ 表示逐元素相加，**Upsample** 表示上采样操作，**IEM**($\cdot$) 为交互式增强模块。

具体而言，每一层通过融合当前特征 C_i 与上层增强特征 P_{i+1} ，并经 **IEM** 进行特征重建，得到增强表示 P_i ，该结果进一步参与后续层级的递归更新。通过该机制，实现了跨层特征的逐级增强与协同优化，输出含有丰富语义信息与结构细节的语义增强特征金字塔 $\{P_1, P_2, P_3, P_4\}$ 。**IEM** 的具体结构将在第 3.2 节中详细介绍。

(2) 基于 CSAM 模块的跨尺度空间对齐阶段

递归增强尽管提升了层级内特征质量，但不同尺度特征间仍存在因多次下采样和非线性变换导致的空间错位。为此，本章引入 **CSAM** 模块完成跨尺度语义对齐，以最高分辨率特征 P_1 为空间基准，对 $\{P_2, P_3, P_4\}$ 进行像素级空间校准。**CSAM** 首先对低分辨率特征进行上采样至统一尺寸，随后基于 P_1 预测像素级语义流场，并通过可微重采样实现特征对齐，得到统一尺寸的 $\{P_1, P_2^{\text{align}}, P_3^{\text{align}}, P_4^{\text{align}}\}$ ，该过程有效缓解了多尺度特征间的空间偏移问题，为后续融合提供一致性表示。**CSAM** 模块的内部机制和详细实现将在第 3.3 节展开。

(3) 基于 GCAF 模块的全局融合阶段

尽管空间对齐机制已提升跨尺度特征的语义一致性，但不同层级之间的互补信息仍有待进一步融合。基于此，本章引入 **GCAF** 模块以实现多尺度特征的全局建模

与融合。GCAF 对 $\{P_1, P_2^{align}, P_3^{align}, P_4^{align}\}$ 进行联合建模，并引入基于全局统计信息的通道注意力机制，实现特征的自适应重标定，以突出关键语义并抑制冗余响应。随后，将融合后的全局特征以残差形式反馈至各尺度分支，生成全局增强特征金字塔 $\{E_1, E_2, E_3, E_4\}$ 。其中， E_1 具有最高空间分辨率，主要保留细粒度的边缘与局部纹理信息，有利于小目标病斑的精细定位； E_4 具有最强语义表达能力，能够提供全局上下文信息以增强类别判别能力； E_2 与 E_3 则在空间细节与语义抽象之间起到过渡作用，用于实现多尺度信息的平衡融合。GCAF 模块的内部机制和详细实现将在第 3.4 节展开。

经过上述三个阶段的逐步优化，HRAGNet 在统一框架下实现了语义信息增强、空间对齐优化与全局上下文建模的协同融合，使不同层级特征在保持尺度特异性的同时具备更强的语义一致性与判别能力，从而有效提升模型在复杂农业场景下的检测性能。

3.1.3 动态检测解码器

HRAGNet 解码器部分采用 Sparse R-CNN 的迭代精修架构，其核心思想是引入一组可学习的提案框与对应的提案特征，通过多轮迭代逐步完成实例级特征建模与边界框优化。

具体而言，模型初始化 N 个可学习提案框 $\{\mathbf{B}_i^{(0)}\}_{i=1}^N$ 与对应的提案特征 $\{\mathbf{q}_i^{(0)}\}_{i=1}^N$ ，其中 $\mathbf{B}_i^{(l)}$ 表示第 l 轮迭代中第 i 个提案框的坐标表示， $\mathbf{q}_i^{(l)}$ 表示其对应的特征向量。在第 l 轮迭代中，提案特征基于当前状态与全局增强特征金字塔 $\{E_1, E_2, E_3, E_4\}$ 进行交互更新，并预测该提案的类别概率 $\mathbf{s}_i^{(l)}$ 与边界框偏移量 $\Delta\mathbf{B}_i^{(l)}$ ，其中 $\mathbf{s}_i^{(l)}$ 表示分类结果， $\Delta\mathbf{B}_i^{(l)}$ 表示对当前提案框的回归修正量。

经过 L 轮迭代后，模型输出最终检测结果 $\{(\mathbf{B}_i^{(L)}, \mathbf{s}_i^{(L)})\}_{i=1}^N$ 。该解码机制通过动态参数化与迭代优化过程，避免了传统密集锚点与候选区域生成步骤，使 HRAGNet 能够在统一框架下实现高效的实例建模与定位优化，从而提升复杂农业场景下的检测精度与鲁棒性。

3.2 交互式增强模块 (IEM)

交互式增强模块 (Interactive Enhancement Module, IEM) 是 HRAGNet 编码器中的第一个核心组件，被嵌入于特征金字塔的自顶向下递归融合过程中。针对传统 FPN

在单向传递中存在的语义信息衰减与细节表达不足问题，**IEM** 模块在每一层对跨层融合特征进行重构与强化，从而逐层生成语义更加充分、结构细节更加清晰的特征表示，为后续跨尺度语义对齐提供高质量输入。

给定输入特征 $\mathbf{X} \in \mathbb{R}^{B \times C \times H \times W}$ (B 为批大小, $C = 256$ 为通道数, H, W 为空间尺寸), **IEM** 通过特征分组、结构与细节的双路径并行增强以及跨路径自适应融合三个步骤, 实现对多尺度信息的有效整合与递归优化, 从而输出结构与语义信息增强后的特征 $\mathbf{Y} \in \mathbb{R}^{B \times C \times H \times W}$ 。**IEM** 模块的具体实现如图 3.3 所示。

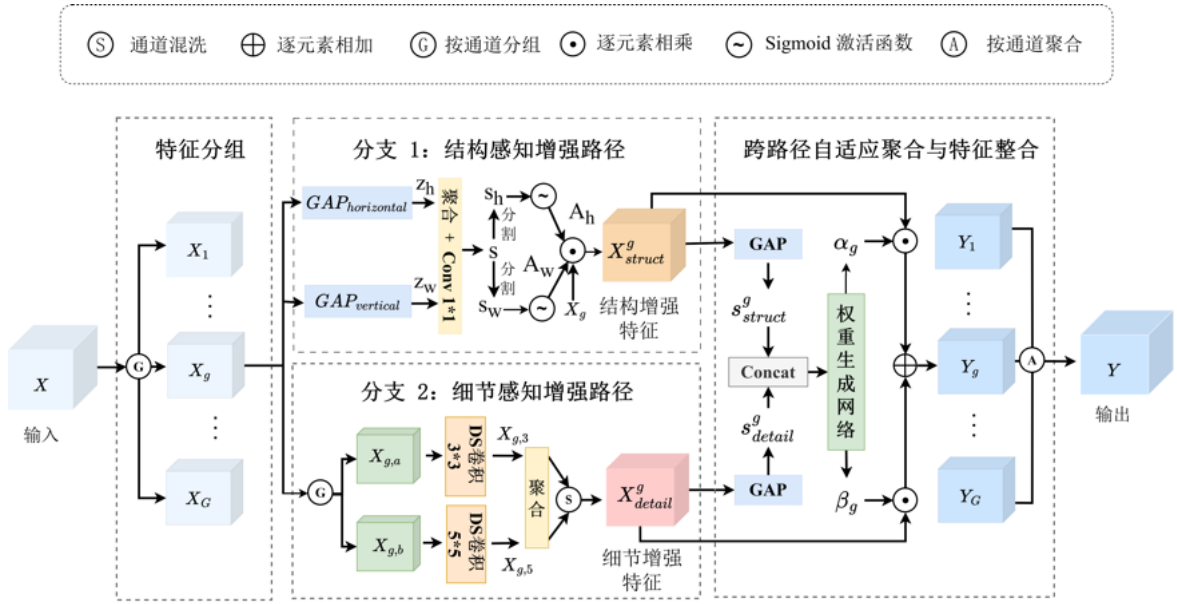


图 3.3 交互式增强模块 (IEM) 结构图

3.2.1 特征分组与子空间投影

深度卷积网络中学习到的特征通道往往存在较强的相关性与冗余性，这在一定程度上限制了特征表达能力。为此，**IEM** 模块通过通道分组策略，将输入特征划分为多个子空间，引导不同通道子集学习互补的视觉模式，从而提升特征表示的多样性与判别性。

在多尺度递归增强过程中，**IEM** 的输入特征 $\mathbf{X}$ 来源于当前层特征 C_i 与上层增强特征 P_{i+1} 上采样后的融合结果，即：

$$\mathbf{X} = C_i \oplus \text{Upsample}(P_{i+1}), \quad i = 1, 2, 3 \quad (3.2)$$

首先，将输入特征张量 $\mathbf{X} \in \mathbb{R}^{B \times C \times H \times W}$ 沿通道维度均匀分割为 G 组，得到分组

特征 $\mathbf{X}_g$:

$$\mathbf{X}_g = \text{Split}_g(\mathbf{X}), \quad \mathbf{X}_g \in \mathbb{R}^{B \times (C/G) \times H \times W}, \quad g = 1, 2, \dots, G \quad (3.3)$$

其中, G 为分组数。通过实验设置 $G = 8$, 此时每组包含 32 个通道, 在特征多样性与表达能力之间取得较优平衡。

分组后, 各子特征 $\mathbf{X}_g$ 在相对独立的子空间中进行建模, 使不同通道组能够关注输入数据的不同视觉属性, 如叶片的颜色、纹理与边缘等, 从而降低通道间冗余并增强特征表达的多样性, 为后续的专门化增强提供基础。

3.2.2 结构-细节感知双路径并行特征增强

在获得分组特征后, **IEM** 模块为每个子特征构建结构感知和细节感知双路径并行增强结构, 以分别建模全局结构信息与局部细节信息。该设计针对植物病害的多尺度特性, 使模型能够同时捕获病斑的整体形态与细粒度纹理特征。

(1) 结构感知增强路径

该路径旨在建模特征图中的长程空间依赖关系, 以捕获目标的整体结构与宏观分布模式。在叶片病害场景中, 病斑往往呈现大范围扩散或沿特定结构分布的特征, 仅依赖局部信息难以准确刻画其整体形态。为此, 受坐标注意力 (Coordinate Attention, CA)^[79] 机制启发, **IEM** 设计了方向敏感的结构感知增强路径, 以提升特征对全局结构信息的建模能力。

对于每个分组特征 $\mathbf{X}_g \in \mathbb{R}^{B \times (C/G) \times H \times W}$, 首先通过方向敏感的平均池化操作, 分别沿空间的高度 (垂直) 和宽度 (水平) 维度进行压缩, 生成一对方向特征描述子:

$$\mathbf{z}_h = \text{GAP}_{\text{horizontal}}(\mathbf{X}_g) = \frac{1}{W} \sum_{0 \leq w < W} \mathbf{X}_g[:, :, :, w] \in \mathbb{R}^{B \times (C/G) \times H \times 1} \quad (3.4)$$

$$\mathbf{z}_w = \text{GAP}_{\text{vertical}}(\mathbf{X}_g) = \frac{1}{H} \sum_{0 \leq h < H} \mathbf{X}_g[:, :, h, :] \in \mathbb{R}^{B \times (C/G) \times 1 \times W} \quad (3.5)$$

其中 $\text{GAP}_{\text{horizontal}}$ 和 $\text{GAP}_{\text{vertical}}$ 分别表示沿水平方向和垂直方向的全局平均池化。这一操作不仅保留了每个空间方向上的全局统计信息, 还维持了方向性的空间结构, $\mathbf{z}_h$ 保留高度维度的位置信息, $\mathbf{z}_w$ 保留了宽度维度的位置信息。这种分解策略使模型能够分别关注目标在不同方向上的延伸特性, 这对于许多具有方向性分布规律的叶片病害尤为有效。

获得方向描述子后，为建模不同方向之间的相关性，首先将 $\mathbf{z}_h$ 与 $\mathbf{z}_w$ 在空间维度进行拼接，并通过一个轻量的特征变换函数 $f(\cdot)$ 进行融合与压缩，得到融合后的方向信息中间表示 $\mathbf{s}$ ：

$$\mathbf{s} = f([\mathbf{z}_h, \mathbf{z}_w]) \in \mathbb{R}^{B \times (C/G)/r \times 1 \times (H+W)} \quad (3.6)$$

其中 $\mathbf{s}$ 用于后续生成空间注意力权重； r 为通道压缩比，实验中设为 8； $f(\cdot)$ 由 1×1 卷积及非线性激活构成，用于实现跨方向信息交互与通道压缩。该融合过程使模型能够在统一特征空间中建立水平方向与垂直方向之间的依赖关系，从而增强对整体结构模式的建模能力。

随后， $\mathbf{s}$ 按照空间维度重新划分为两个方向分支，并分别映射为对应的空间注意力权重：

$$\mathbf{A}_h, \mathbf{A}_w = \sigma(\text{Split}(\mathbf{s})) \quad (3.7)$$

其中， $\text{Split}(\cdot)$ 表示按照原始高度 H 与宽度 W 对特征进行划分； $\sigma(\cdot)$ 为 Sigmoid 激活函数。 $\mathbf{A}_h \in \mathbb{R}^{B \times (C/G) \times H \times 1}$ 和 $\mathbf{A}_w \in \mathbb{R}^{B \times (C/G) \times 1 \times W}$ 分别表示高度方向与宽度方向的空间注意力权重，其取值范围为 $[0, 1]$ ，用于刻画不同空间位置的重要性。

最后，将两个方向性的注意力权重作用于原始分组特征，实现空间结构的自适应增强，得到结构增强后的特征 $\mathbf{X}_{\text{struct}}^g$ ：

$$\mathbf{X}_{\text{struct}}^g = \mathbf{X}_g \odot \mathbf{A}_h \odot \mathbf{A}_w \quad (3.8)$$

其中 $\odot$ 表示逐元素乘法。这一操作在特征图中同时施加了水平方向和垂直方向的空间注意力，使模型能够突出在两个方向上均具有显著响应的区域，从而强化结构相关特征的表达能力。

(2) 细节感知增强路径

与结构路径关注宏观分布不同，细节感知路径专注于提取病变区域的微观局部特征，包括边缘、角点、纹理模式等。这对于精确边界定位、早期微小病变识别以及区分相似纹理背景至关重要。与此同时，细节增强后的特征有助于后续 CSAM 模块在像素级对齐时保留精细的边界信息，避免因对齐操作导致细节模糊。

为了在计算效率与特征表达能力之间取得平衡，模块采用深度可分离卷积 (Depthwise Separable Convolution, DSConv)^[88] 作为基础算子。考虑到不同局部模式

对感受野的需求存在差异，**IEM** 模块进一步设计了双分支多感受野卷积结构，实现对多尺度局部特征的有效建模。

首先，将每个分组特征 $\mathbf{X}_g$ 沿通道维度划分为两部分：

$$\mathbf{X}_{g,a}, \mathbf{X}_{g,b} = \text{Split}(\mathbf{X}_g) \quad (3.9)$$

随后分别输入核大小为 3×3 和 5×5 的深度可分离卷积层：

$$\mathbf{X}_{g,3} = \text{DSConv}_{3 \times 3}(\mathbf{X}_{g,a}) \quad (3.10)$$

$$\mathbf{X}_{g,5} = \text{DSConv}_{5 \times 5}(\mathbf{X}_{g,b}) \quad (3.11)$$

其中， $\text{DSConv}_{k \times k}$ 表示核大小为 $k \times k$ 的深度可分离卷积。通过引入不同尺度的卷积核，该结构能够在保持计算开销较低的前提下，同时建模细粒度局部纹理与更大范围的上下文信息，其中 3×3 分支侧重细粒度特征建模， 5×5 分支用于扩展感受野以捕获更丰富的局部上下文。

不同感受野卷积分支的输出在通道维度上往往呈现相对独立的分布。若仅通过拼接进行融合，这种分布差异将被直接保留，使来自不同分支的特征在通道维度上缺乏有效混合，从而限制跨分支的信息交互与多尺度特征融合效果。为缓解上述问题，引入通道混洗 (Channel Shuffle)^[89] 操作，对拼接后的特征进行重组，得到细节增强特征 $\mathbf{X}_{\text{detail}}^g$ ：

$$\mathbf{X}_{\text{detail}}^g = \text{ChannelShuffle}(\text{Concat}(\mathbf{X}_{g,3}, \mathbf{X}_{g,5})) \quad (3.12)$$

具体而言，通道重排通过对拼接后的通道进行重新排列并重新分组，使来自不同卷积分支的特征在各通道组中交错分布，从而打破原有的通道独立性，为后续特征融合提供更充分的跨通道信息交互条件。

(3) 双路径的互补性分析

表 3.1 展示了结构感知路径与细节感知路径在功能上的互补性。基于该特性，**IEM** 模块通过自适应机制对两条路径进行动态融合，使生成的特征在语义判别能力与空间精度之间取得良好平衡。经双路径协同增强后，特征表示在多尺度与多维度上得到有效提升，从而可以为后续 **CSAM** 与 **GCAF** 模块提供更加稳健且判别性更强的输入。

表 3.1 双路径增强机制的互补特性分析

特性维度	结构感知路径	细节感知路径	互补效应
感受野范围	全局（全图范围）	局部（ $3 \times 3, 5 \times 5$ 邻域）	覆盖从局部到全局的全尺度感知
信息类型	空间分布、方向性、宏观形态	边缘、纹理、微观模式	同时捕获宏观与微观特征
对病害的贡献	理解病害分布规律，定位病区	精确定位边界，识别早期病变	提升模型检测准确性
主要操作	轻量注意力机制	多尺度深度卷积	轻量级，高效利用资源

3.2.3 跨路径自适应聚合与特征整合阶段

获得结构增强特征 $\mathbf{X}_{\text{struct}}^g$ 和细节增强特征 $\mathbf{X}_{\text{detail}}^g$ 后，如何有效融合两者的互补信息成为关键。考虑到不同病害模式下结构信息与细节信息的重要性存在差异，IEM 模块采取轻量的自适应权重学习机制，让两条路径的输出通过自适应融合进行信息整合，生成兼具语义表达与细节保留能力的增强特征，为后续跨尺度对齐提供高质量输入。

IEM 设计了一个轻量的权重学习网络来动态决定两种特征的融合比例。首先，通过全局平均池化获取两种特征的全局显著性描述：

$$\mathbf{s}_{\text{struct}}^g = \text{GAP}(\mathbf{X}_{\text{struct}}^g) = \frac{1}{H \times W} \sum_{h=0}^{H-1} \sum_{w=0}^{W-1} \mathbf{X}_{\text{struct}}^g[:, :, h, w] \in \mathbb{R}^{B \times (C/G)} \quad (3.13)$$

$$\mathbf{s}_{\text{detail}}^g = \text{GAP}(\mathbf{X}_{\text{detail}}^g) = \frac{1}{H \times W} \sum_{h=0}^{H-1} \sum_{w=0}^{W-1} \mathbf{X}_{\text{detail}}^g[:, :, h, w] \in \mathbb{R}^{B \times (C/G)} \quad (3.14)$$

其中， $\mathbf{s}_{\text{struct}}^g$ 和 $\mathbf{s}_{\text{detail}}^g$ 分别表示结构特征与细节特征的全局通道描述向量，用于刻画各通道的重要性。

随后，将两个描述向量在通道维度进行拼接，并输入一个轻量的权重生成网络预测融合系数：

$$\alpha_g, \beta_g = \text{Softmax}(\text{FC}_2(\text{ReLU}(\text{FC}_1(\text{Concat}(\mathbf{s}_{\text{struct}}^g, \mathbf{s}_{\text{detail}}^g)))))) \quad (3.15)$$

其中， FC_2 和 FC_1 为两个全连接层， FC_2 的输出维度设置为 2，用于分别预测结构分支与细节分支的两个标量权重，经 Softmax 归一化后得到 α_g 与 β_g ，满足 $\alpha_g + \beta_g = 1$ ，从而实现对两个分支特征的自适应加权。

基于自适应学习得到的权重对两类特征进行加权融合，实现子组特征的增强表

达，从而得到增强后的子组特征 $\mathbf{Y}_g$ ：

$$\mathbf{Y}_g = \alpha_g \cdot \mathbf{X}_{\text{struct}}^g + \beta_g \cdot \mathbf{X}_{\text{detail}}^g \quad (3.16)$$

该自适应融合过程能够根据输入特征动态平衡结构信息与细节信息，从而提升特征表达的灵活性与针对性。

在获得增强后的各子组特征 $\{\mathbf{Y}_g\}_{g=1}^G$ 后，将其沿通道维度拼接，并通过 1×1 卷积进行统一整合与维度调整，得到 IEM 模块的输出特征：

$$\mathbf{Y} = \text{Conv}_{1 \times 1}(\text{Concat}(\mathbf{Y}_1, \mathbf{Y}_2, \dots, \mathbf{Y}_G)) \quad (3.17)$$

该过程在完成跨路径自适应融合的同时，实现多子组特征的协同整合，使输出特征在保持多样性的基础上具备更强的一致性与判别能力。最终得到的 $\mathbf{Y}$ 即为当前层级的增强特征 P_i ，并作为后续跨尺度对齐模块的输入。

综上所述，IEM 模块通过上述三个协同阶段，将输入特征转化为语义与结构信息增强的特征并嵌入 HRAGNet 的递归增强流程。通过递归增强流程输出的语义增强特征金字塔 $\{P_1, P_2, P_3, P_4\}$ 有效缓解了传统多尺度特征融合方法的语义衰减问题，为后续 CSAM 模块的特征对齐与 GCAF 模块的全局融合提供了高质量输入，为整体目标检测框架奠定了坚实基础。

3.3 跨尺度语义对齐模块 (CSAM)

跨尺度语义对齐模块 (Cross-scale Semantic Alignment Module, CSAM) 是 HRAGNet 编码器中的第二个核心组件，用于缓解 IEM 在特征增强过程中引入的空间不一致问题。尽管 IEM 模块已对多尺度特征进行了语义层面的增强，但其递归融合与尺度变换过程缺乏显式空间约束，容易导致不同层级特征在对应语义区域上的响应错位，从而影响后续融合效果。为此，CSAM 通过学习可变形的语义流场，对多尺度特征进行动态空间校准，使不同尺度特征能够在统一坐标空间下实现对齐。

具体而言，CSAM 以 IEM 递归增强阶段输出的多尺度特征 $\{P_1, P_2, P_3, P_4\}$ 为输入，选取高分辨率特征 P_1 作为参考，通过跨尺度建模将其余尺度特征映射至统一坐标空间。通过预测空间偏移并进行自适应重采样，实现跨尺度特征的像素级对齐，输出对齐后的特征集合 $\{P_1, P_2^{\text{align}}, P_3^{\text{align}}, P_4^{\text{align}}\}$ ，并作为后续 GCAF 模块的输入，以保证跨尺度融合的空间一致性。CSAM 的具体结构如图 3.4 所示。

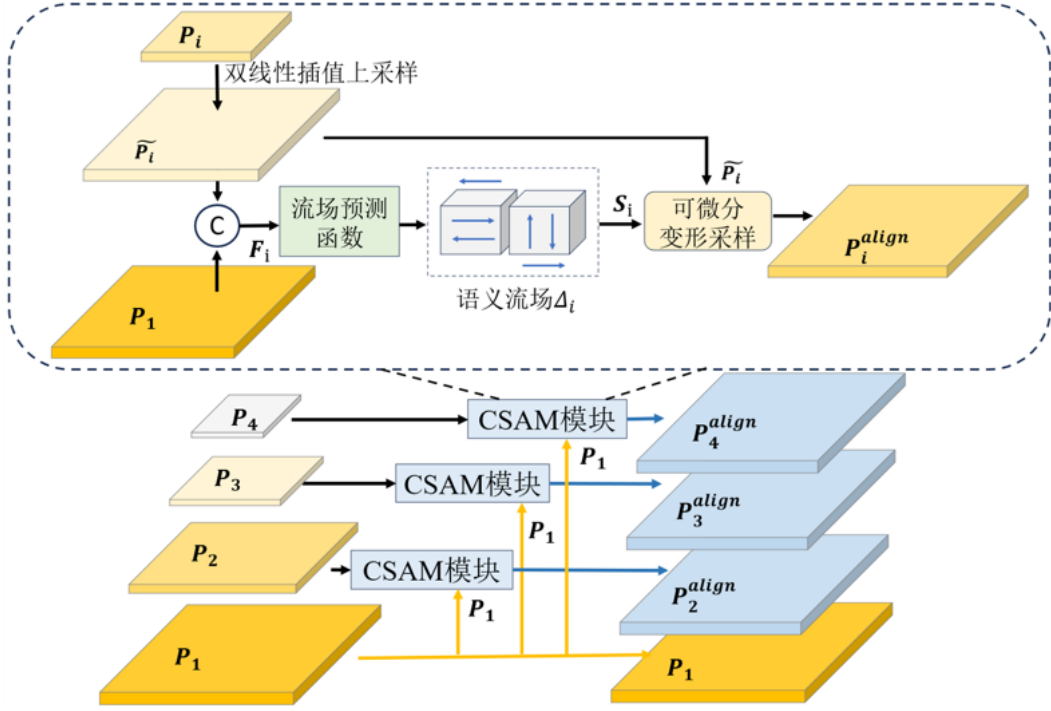


图 3.4 跨尺度语义对齐模块 (CSAM) 结构图

3.3.1 跨尺度对齐建模

对于 $P_i \in \mathbb{R}^{C \times H_i \times W_i}$ ($i = 1, 2, 3, 4$)，其中 C 表示特征通道数， H_i 与 W_i 分别表示第 i 个尺度特征图的空间尺寸。由于 P_1 具有最高空间分辨率，保留了更加丰富的边缘、纹理及局部结构信息，因此本章选取 P_1 作为跨尺度对齐的参考特征，并将其余尺度特征统一映射到 P_1 的坐标空间中。

对于待对齐特征 P_i ($i = 2, 3, 4$)，首先采用双线性插值将其上采样至与 P_1 相同的空间分辨率，得到：

$$\tilde{P}_i = \text{Up}(P_i; H_1, W_1) \quad (3.18)$$

其中， $\text{Up}(\cdot; H_1, W_1)$ 表示将输入特征重采样到参考尺度 $P_1(H_1, W_1)$ 的双线性上采样操作。

经过上述步骤后， $\tilde{P}_i$ 与 P_1 在空间尺寸上保持一致，但相同语义区域在不同尺度特征图中的激活位置仍可能存在偏移，因此仅依赖插值操作难以建立精确的跨尺度空间对应关系。因此，CSAM 的核心目标在于，在统一坐标空间下学习一种自适应的空间映射方式，使得不同尺度特征能够在像素级建立更加准确的语义对应关系。

3.3.2 语义流场预测机制

为实现上述跨尺度空间对齐，受光流方法在像素级对应建模中的启发^[90-91]，CSAM 模块引入语义流场以建模参考特征与待对齐特征之间的空间偏移关系。语义流场本质上是定义在参考特征 P_1 上的二维位移场，用于描述参考坐标在待对齐特征中的对应采样位置。

首先定义参考特征 P_1 的坐标集合为 Ω ：

$$\Omega = \{p = (x, y) \mid 1 \leq x \leq W_1, 1 \leq y \leq H_1\} \quad (3.19)$$

对于任意参考位置 $p \in \Omega$ ，语义流场 $\Delta_i(p)$ 被表示为二维位移向量：

$$\Delta_i(p) = (\Delta x_i(p), \Delta y_i(p)) \quad (3.20)$$

该语义流场用于表示该位置在待对齐特征 $\tilde{P}_i$ 中对应采样点相对于参考坐标 p 的偏移量。由此，参考位置 p 在 $\tilde{P}_i$ 上对应的连续采样坐标可写为：

$$S_i(p) = p + \Delta_i(p) \quad (3.21)$$

其中， $S_i(p)$ 表示参考位置 p 在源特征 $\tilde{P}_i$ 上的实际采样位置。

为使语义流场能够根据输入特征自适应调整，通过轻量级偏移预测函数对其进行预测。具体而言，首先将参考特征与待对齐特征在通道维度进行拼接，构造跨尺度联合表示 F_i ：

$$F_i = \text{Concat}(P_1, \tilde{P}_i), \quad F_i \in \mathbb{R}^{2C \times H_1 \times W_1} \quad (3.22)$$

随后，通过参数化映射函数 $\mathcal{F}_{\theta_i}(\cdot)$ 预测语义流场 Δ_i ：

$$\Delta_i = \mathcal{F}_{\theta_i}(F_i) \quad (3.23)$$

其中， $\mathcal{F}_{\theta_i}(\cdot)$ 表示由参数 θ_i 控制的流场预测函数，由两层 3×3 卷积构成的轻量级映射函数实现。第一层卷积用于编码跨尺度联合特征的局部结构信息，第二层卷积用于输出二维位移场，从而提升位移估计的稳定性与表达能力。

从位置层面看，对于任意参考位置 $p \in \Omega$ ， $\Delta_i(p)$ 反映了该位置在跨尺度语义匹配下应当施加的空间修正量。若某一病斑边缘在高分辨率特征中呈现更清晰的结构

响应，而在低分辨率特征中表现为模糊或位置偏移，则 $\Delta_i(p)$ 将引导对应采样点向更加符合边界分布的位置进行调整，从而减弱跨尺度语义错配带来的信息混叠。进一步地，语义流场的预测过程可以表示为：

$$S_i(p) = p + \mathcal{F}_{\theta_i}(F_i)(p) \quad (3.24)$$

这表明 CSAM 本质上是通过学习一个以联合特征为输入的空间偏移函数，对参考网格上的每个位置生成自适应采样坐标，从而建立从参考尺度到待对齐尺度的动态映射关系。

3.3.3 可微分特征变形采样

在获得采样坐标场 S_i 后，需要对特征 $\tilde{P}_i$ ($i = 2, 3, 4$) 进行空间重采样，以生成对齐后的特征表示 P_i^{align} ($i = 2, 3, 4$)。为保证模型的端到端可训练性，采用可微分双线性插值实现特征变形采样。

对于任意位置 $p \in \Omega$ ，其对应采样坐标为 $S_i(p)$ 。由于该坐标通常为连续值，需要通过双线性插值计算其特征响应：

$$P_i^{\text{align}}(p) = \sum_{q \in \mathcal{N}(S_i(p))} \omega(q, S_i(p)) \tilde{P}_i(q) \quad (3.25)$$

其中 $\mathcal{N}(S_i(p))$ 表示连续采样坐标 $S_i(p)$ 在特征图上的四邻域整数网格点集合， $q \in \mathcal{N}(S_i(p))$ 表示其中任意一个网格点， $\omega(q, S_i(p))$ 表示由空间距离确定的双线性插值权重。该过程能够在连续坐标空间中实现平滑且可微的特征采样，从而保证梯度能够通过采样操作反向传播至流场预测分支。

从整体形式上看，上述过程等价于在采样坐标场 S_i 的引导下，对特征图 $\tilde{P}_i$ 进行空间变形采样，其结果可统一表示为：

$$P_i^{\text{align}} = \text{Warp}(\tilde{P}_i, S_i) \quad (3.26)$$

其中， $\text{Warp}(\cdot)$ 表示基于采样坐标场的可微分双线性采样操作。该操作通过为输出特征图中的每个位置确定其在输入特征图中的连续对应点，并利用双线性插值获取该点的特征值，从而实现特征的空间变形与跨尺度对齐。

最终，CSAM 输出对齐后的特征集合：

$$\mathcal{P}^{\text{align}} = \{P_1, P_2^{\text{align}}, P_3^{\text{align}}, P_4^{\text{align}}\} \quad (3.27)$$

其中， P_1 为原始高分辨率参考特征， P_i^{align} ($i = 2, 3, 4$) 表示通过语义流场引导后对齐到 P_1 坐标空间中的特征表示。

经过空间对齐后，不同尺度特征能够在统一参考坐标系下建立更加准确的语义对应关系，并作为后续全局上下文聚合融合模块（GCAF）的输入。相较于传统基于固定插值的对齐方式，CSAM 模块能够联合建模参考特征与待对齐特征之间的语义关系，自适应预测空间位移，从而建立更加精确的跨尺度映射。同时，所学习的语义流场允许特征在空间维度进行非刚性调整，使模型能够更好地适应植物叶片病斑的复杂形态，并通过可微分采样实现端到端优化。在 HRAGNet 框架中，CSAM 作为连接 IEM 与 GCAF 的关键环节，通过显式空间对齐提升多尺度特征的一致性，从而为后续融合提供更可靠的结构基础。

3.4 全局上下文聚合融合模块（GCAF）

全局上下文聚合融合模块（Global Context Aggregation Fusion, GCAF）是 HRAGNet 编码器中的第三个核心组件，其目标是在完成跨尺度空间对齐之后，对多尺度特征进行全局上下文感知的自适应融合。尽管 CSAM 已在空间层面实现了特征对齐，但不同尺度特征在语义表达上仍存在不一致与冗余问题。FPN 等传统融合方法依赖逐层信息传递，难以建模全局上下文，同时缺乏对多尺度特征重要性的自适应建模能力。针对上述不足，GCAF 引入全局上下文建模与通道重标定机制，以实现更加有效的多尺度特征融合。

具体而言，GCAF 以对齐后的特征集合 $\{P_1, P_2^{\text{align}}, P_3^{\text{align}}, P_4^{\text{align}}\}$ 为输入，首先通过通道拼接构建跨尺度联合表示；随后在全局特征空间中建模通道间依赖关系，生成自适应权重对特征进行重标定；最后结合残差连接，将增强后的全局语义信息反馈至各尺度特征，输出融合后的全局增强特征金字塔 $\{E_1, E_2, E_3, E_4\}$ ，为后续检测解码器提供统一且语义增强的多尺度表示。GCAF 模块的具体实现如图 3.5 所示。

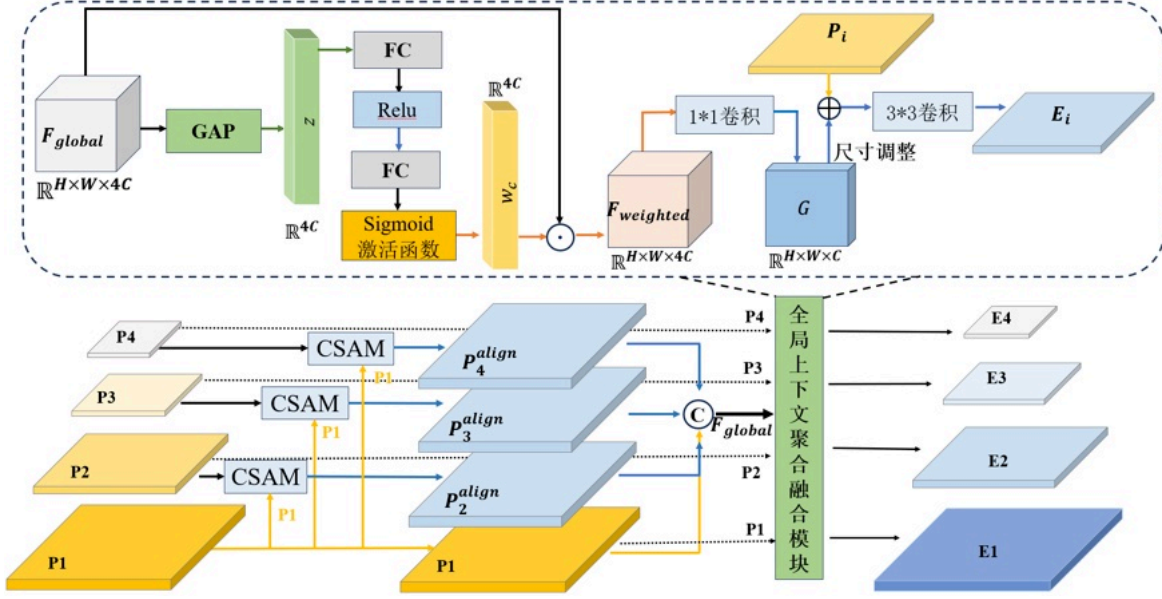


图 3.5 全局上下文聚合融合模块 (GCAF) 结构图

3.4.1 全局上下文建模与多尺度联合表示

为充分利用不同尺度特征中的互补信息，首先将对齐后的特征沿通道维度拼接，构建跨尺度联合表示 F_{global} ：

$$F_{global} = \text{Concat}(P_1, P_2^{align}, P_3^{align}, P_4^{align}), \quad F_{global} \in \mathbb{R}^{B \times 4C \times H \times W} \quad (3.28)$$

其中， $\text{Concat}(\cdot)$ 表示在通道维度上进行拼接操作。该操作将原本分布于不同层级的特征统一映射至同一表示空间，使模型能够在全局范围内同时感知不同尺度的语义信息。低层特征提供精细的纹理与边缘细节，高层特征提供抽象的语义结构信息，多尺度信息在统一空间中的协同表达为后续全局建模奠定基础。相比传统逐层融合方式，该策略实现了跨尺度信息的一步式聚合，有效缩短信息传递路径。

随后，对 F_{global} 进行全局平均池化，得到通道描述向量 $z \in \mathbb{R}^{4C}$ ：

$$z = \frac{1}{H \times W} \sum_{h=1}^H \sum_{w=1}^W F_{global}(:, :, h, w) \quad (3.29)$$

向量 z 对跨尺度联合特征在空间维度上的整体响应进行压缩，其每一维均反映对应通道在全局范围内的激活强度。由于该描述子来源于融合后的多尺度表示，因此其中不仅包含单一尺度的局部信息，还隐式编码了不同尺度之间的语义互补关系，从而为后续注意力权重学习提供全局上下文依据。

3.4.2 动态通道注意力权重生成

通道描述子 z 输入一个由两个全连接层构成的瓶颈结构进行非线性变换，生成通道注意力权重 $w_c \in \mathbb{R}^{4C}$ ：

$$w_c = \sigma(W_2 \cdot \delta(W_1 \cdot z)) \quad (3.30)$$

其中 $W_1 \in \mathbb{R}^{(4C/r) \times 4C}$ 和 $W_2 \in \mathbb{R}^{4C \times (4C/r)}$ 为可学习权重矩阵； $\delta(\cdot)$ 表示 **ReLU** 激活函数，引入非线性变换； $\sigma(\cdot)$ 表示 **Sigmoid** 函数，将输出约束至 $(0, 1)$ 区间，形成软注意力权重。

将 w_c 重塑为与 F_{global} 相同的维度，然后进行逐通道乘法，实现对全局特征的自适应重校准：

$$F_{\text{weighted}} = w_c \odot F_{\text{global}} \quad (3.31)$$

其中 $\odot$ 表示逐通道乘法。

通过该过程，**GCAF** 能够增强与病斑颜色、纹理、边缘形状及区域结构相关的判别性通道，同时抑制由复杂背景、光照变化和叶脉纹理干扰所带来的冗余响应。相较于固定融合策略，该方法具有更强的输入自适应能力，能够针对不同样本动态调整多尺度特征的融合权重，从而提升特征表达的有效性与稳健性。经过重校准的全局特征 F_{weighted} 已经具备了跨尺度的通道重要性判别能力。

3.4.3 全局信息注入与残差增强

将加权后的全局特征 F_{weighted} 通过 1×1 卷积进行通道压缩，得到紧凑的全局特征图 $G \in \mathbb{R}^{B \times C \times H \times W}$ ：

$$G = \text{Conv}_{1 \times 1}(F_{\text{weighted}}) \quad (3.32)$$

其中， 1×1 卷积能够在不改变空间分辨率的前提下实现跨通道信息交互，并将联合特征重新映射回与单尺度特征一致的通道维度。此时， G 已经聚合了来自不同尺度分支的上下文信息，可视为对当前输入图像全局语义结构的紧凑表征。

为了在保留各尺度原有表征特性的同时，将全局上下文信息反馈到各层特征中，模块采用残差式信息注入策略。具体而言，令 P_i 表示 **IEM** 模块输出的第 i 层语义增

强特征，则第 i 个尺度的增强输出定义为：

$$E_i = \text{Conv}_{3 \times 3} (P_i + \text{Resize}_i(G)), \quad i = 1, 2, 3, 4, \quad (3.33)$$

其中， $\text{Resize}_i(\cdot)$ 用于将全局特征调整至与 P_i 相同的空间分辨率。

基于上述残差式融合机制，全局语义信息以逐层回注的方式与局部特征进行融合，在保持各尺度特征分布稳定的同时实现跨尺度上下文信息的有效补充。融合结果进一步通过 3×3 卷积进行平滑处理，以减弱跨尺度信息叠加带来的不连续性。最终，GCAF 输出全局增强特征金字塔 $\{E_1, E_2, E_3, E_4\}$ ，该特征在各尺度上具有更强的语义一致性与判别能力，为后续检测解码器提供稳定的候选区域表示。相较于传统逐层传播式融合方法，该模块能够在统一特征空间内直接建模跨尺度关联，并有效强化关键语义信息，从而提升复杂田间环境下叶片病害目标的检测性能。

3.5 实验与结果分析

3.5.1 实验设置与评价指标

(1) 模型基线与环境配置

HRAGNet 基于 PyTorch 2.0 实现，运行于 NVIDIA RTX 4090 GPU (24 GB 显存)，采用 Sparse R-CNN 作为基本检测框架，ImageNet 预训练的 ResNet-50 为骨干网络。具体设置如表 3.2 所示。

表 3.2 实验设置

配置项	具体设置
CPU	Intel(R) Xeon(R) CPU E5-2650 v3
GPU	NVIDIA RTX 4090 GPU 24 GB
深度学习框架	PyTorch 2.0
操作系统	Ubuntu 18.04.2 LTS (64-bit)
内存	128 GB
硬盘容量	2 TB

(2) 数据增强与训练参数

训练阶段采用多维度数据增强策略，具体配置及核心训练参数如表 3.3 所示。优化器选用 AdamW，共训练 36 个 epoch，学习率在第 24 和 33 个 epoch 处衰减至原值的 1/10。测试阶段 NMS 阈值设为 0.5，保留置信度最高的 100 个检测框。

表 3.3 数据增强策略与训练参数配置

配置项	具体设置
随机缩放	0.8~1.2
随机翻转	概率 0.5
输入尺寸	短边 800, 长边 ≤ 1333
优化器	AdamW
初始学习率	2.5×10^{-5}
权重衰减	1×10^{-4}
批次大小	4
训练轮数	36 epoch
学习率衰减	第 24、33 epoch
颜色抖动	± 0.2
高斯模糊	概率 0.3, 3×3
标准化	ImageNet 均值/方差

(3) 评估指标

实验采用 MS COCO 标准评估体系。综合性能指标包括平均精度 AP ($\text{IoU} \in [0.5, 0.95]$)、 AP_{50} 和 AP_{75} ，分别衡量不同定位精度要求下的检测能力。多尺度目标性能指标涵盖小目标 AP_s (面积 $< 32^2$ 像素)、中目标 AP_m ($32^2 \leq \text{面积} \leq 96^2$ 像素) 和大目标 AP_l (面积 $> 96^2$ 像素)。

3.5.2 与主流检测算法的对比实验

为验证本章提出的 HRAGNet 模型在叶片病害检测任务中的有效性与泛化能力，选取多种主流目标检测算法作为对比基线，涵盖两阶段检测方法 Faster R-CNN^[4]、一阶段检测方法 FCOS^[22] 与 YOLOv10^[23]、基于 Transformer 的端到端检测方法 DETR^[28] 与 DINO^[30] 等，在 PlantDoc 与 TLD 等典型植物叶片病害数据集上进行主要性能评估，并在 FD 数据集上进一步验证模型在跨类别农业目标检测任务中的泛化能力。

(1) PlantDoc 数据集上的结果

PlantDoc 数据集包含复杂背景和多变光照下的多种作物病害，小尺度病斑占比较高，对模型的空间细节感知能力提出了严峻挑战。表 3.4 展示了各模型在该数据集上的性能对比。

从表 3.4 可见，HRAGNet 在 AP 达到 43.8%，其中， AP_{50} 和 AP_{75} 均达到当前最

表 3.4 在 PlantDoc 数据集上与其他先进算法的比较

模型	骨干网络	AP	AP ₅₀	AP ₇₅	AP _s	AP _m	AP _l
Faster R-CNN ^[4]	ResNet-50	36.6	57.3	40.5	2.2	19.0	38.6
Mask R-CNN ^[13]	ResNet-50	33.3	52.3	36.3	1.8	22.8	34.4
FCOS ^[22]	ResNet-50	34.5	53.2	36.5	2.6	26.9	34.9
YOLOv10 ^[23]	CSPNet	39.8	56.1	—	—	—	—
Sparse R-CNN ^[65]	ResNet-50	35.9	52.4	39.5	3.3	18.1	37.8
DETR ^[28]	ResNet-50	34.6	51.3	37.3	1.8	22.5	34.2
Deformable-DETR ^[3]	ResNet-50	40.0	56.3	45.1	1.7	25.2	41.8
DINO ^[30]	ResNet-50	40.2	53.0	43.2	2.3	28.5	41.1
HDINO ^[92]	ResNet-50	33.7	32.0	35.5	1.5	27.8	35.1
DDQ ^[93]	ResNet-50	38.0	54.1	43.2	1.8	27.2	40.0
RecursiveDet ^[94]	ResNet-50	42.2	60.4	46.2	3.5	30.3	44.3
AlignDETR ^[31]	ResNet-50	43.4	62.7	47.3	3.4	29.1	45.4
HRAGNet	ResNet-50	43.8	64.4	47.8	3.9	29.9	46.2

注：AP 为 COCO 标准下的平均精度均值。

佳水平，同时在小目标和大目标检测性能上也表现突出，整体检测性能取得最优结果。该结果表明，所提出方法在不同尺度目标的建模上具有更强的适应能力。这一优势主要归因于三个模块的协同作用。首先，**IEM** 通过递归增强机制强化语义信息在层级间的传递，使低层特征获得高层语义指导，从而提升对微小病斑的响应强度；其次，**CSAM** 以高分辨率特征为基准进行可变形对齐，有效修正因下采样导致的病斑边缘错位，避免小病斑在融合过程中被模糊化；最后，**GCAF** 利用全局注意力融合多尺度上下文，使模型在面对复杂背景时能够聚焦真正的病害区域，抑制噪声干扰。尽管在中尺度 AP_m 上略低于 RecursiveDet，但整体 AP 的提升证明了 HRAGNet 在处理尺度多样性方面的综合优势。

(2) 番茄叶片疾病数据集（TLD）上的结果

TLD 数据集包含九类病害，图像中的病斑尺度分布广泛，且存在早期微小病斑与晚期大面积病斑共存的情况。表 3.5 汇总了各模型在该数据集上的检测性能。

HRAGNet 在各项评价指标上均取得最佳性能表现，其中 AP 达到 81.9%，AP₇₅ 以及不同尺度目标的检测性能均达到当前最优水平。小尺度目标检测方面也优于 RecursiveDet 和 AlignDETR，体现出更强的尺度适应能力。这表明 HRAGNet 能够同时兼顾不同尺度的病斑检测，尤其在中尺度目标上表现突出。**IEM** 的双路径设计使结构感知路径捕获病斑整体形态，细节感知路径强化早期病斑的微观纹理；**CSAM** 通

表 3.5 在 TLD 数据集上与其他先进算法的比较

模型	骨干网络	AP	AP ₅₀	AP ₇₅	AP _s	AP _m	AP _l
Faster R-CNN ^[4]	ResNet-50	76.3	90.3	84.3	5.6	44.5	78.8
Mask R-CNN ^[13]	ResNet-50	75.2	90.0	83.5	3.3	44.8	78.2
FCOS ^[22]	ResNet-50	75.8	91.5	84.2	7.2	43.9	78.2
YOLOv10 ^[23]	CSPNet	81.5	91.3	—	—	—	—
Sparse R-CNN ^[65]	ResNet-50	78.4	88.9	83.8	9.5	43.8	81.7
DETR ^[28]	ResNet-50	70.2	78.9	75.4	5.5	34.6	75.9
Deformable-DETR ^[3]	ResNet-50	76.9	89.8	83.5	12.1	39.8	79.1
DINO ^[30]	ResNet-50	81.2	90.2	86.1	8.2	48.2	84.1
HDINO ^[92]	ResNet-50	70.5	77.4	74.3	10.2	44.1	73.9
DDQ ^[93]	ResNet-50	81.1	91.4	83.2	9.8	55.2	83.0
RecursiveDet ^[94]	ResNet-50	81.4	91.5	86.9	11.2	44.8	84.1
AlignDETR ^[31]	ResNet-50	80.0	89.3	85.0	11.0	48.1	82.8
HRAGNet	ResNet-50	81.9	91.2	86.9	11.5	56.8	84.5

注：AP 为 COCO 标准下的平均精度均值。

过学习语义流场，灵活应对番茄叶片卷曲、重叠导致的非线性空间偏移，确保低分辨率特征中的语义信息准确映射到高分辨率空间；GCAF 的全局注意力融合整合来自不同层级的上下文，帮助模型区分相似病害，从而提升分类准确率。尽管 AP₅₀ 略低于另外三个主流模型，但 AP₇₅ 的领先说明 HRAGNet 生成的边界框更加精确，定位质量更高。

(3) 水果检测数据集 (FD) 上的结果

为验证 HRAGNet 在通用目标检测任务中的泛化能力，在 FD 数据集上进行对比实验。该数据集包含六类日常水果，图像来源于 COCO 和 PASCAL VOC，具有遮挡、光照变化和复杂背景等挑战。实验结果如表 3.6 所示。需要说明的是，FD 数据集主要面向水果目标检测任务，与叶片病害检测在目标形态与语义上存在一定差异，因此本文将作为辅助验证数据集，用于评估模型在跨场景下的泛化能力，而非作为主要性能评价依据。

可以看到，HRAGNet 以 40.6% 的 AP 领先所有对比方法，不同尺度目标的检测性能也达到最优，证明了其跨领域迁移能力。尽管 AP₇₅ 略低于 AlignDETR，但整体 AP 的提升说明 HRAGNet 在不同交并比阈值下具有更稳定的性能。这一结果验证了所提模块并非仅针对植物病害设计，而是能够普遍提升多尺度特征的表示质量。具体而言，IEM 的分组卷积与双路径增强能够提取水果的纹理和形状；CSAM 通过可

表 3.6 在 FD 数据集上的检测性能比较

模型	骨干网络	AP	AP ₅₀	AP ₇₅	AP _s	AP _m	AP _l
Faster R-CNN ^[4]	ResNet-50	35.8	51.3	37.2	9.8	21.6	47.8
Sparse R-CNN ^[65]	ResNet-50	37.5	52.8	39.6	10.4	23.1	49.8
Deformable-DETR ^[3]	ResNet-50	37.2	57.8	38.8	9.9	23.8	49.3
DINO ^[30]	ResNet-50	38.6	54.5	39.7	12.4	21.9	51.3
HDINO ^[92]	ResNet-50	39.1	55.3	40.6	12.3	22.7	51.7
RecursiveDet ^[94]	ResNet-50	39.3	54.5	41.8	12.0	22.8	52.3
AlignDETR ^[31]	ResNet-50	39.9	55.1	41.9	13.2	23.2	52.4
HRAGNet	ResNet-50	40.6	56.9	41.8	13.5	24.1	52.6

注：AP 为 COCO 标准下的平均精度均值。

变形对齐解决了因遮挡或视角变化引起的特征错位；GCAF 的全局上下文融合则有助于在杂乱背景中突出目标物体。这些设计共同作用，使 HRAGNet 在通用检测任务中同样表现出色。

HRAGNet 与两种性能较优的目标检测方法 RecursiveDet^[94] 和 AlignDETR^[31] 在 PlantDoc、TLD 及 FD 数据集上的检测结果可视化对比如图 3.6 所示。综合三个数据集的实验结果和可视化结果可以看出，HRAGNet 在多数评价指标上均优于现有主流目标检测方法，验证了其基于 IEM、CSAM 与 GCAF 的递进式协同优化设计的有效性。IEM 通过递归增强解决了语义衰减问题，为后续对齐和融合提供高质量输入；CSAM 通过可学习流场实现了像素级空间对齐，消除了跨尺度特征错位；GCAF 通过全局注意力与残差注入，使各层级特征在保留自身尺度的同时融入全局上下文。三者共同构建了兼具语义判别性、空间精确性与全局一致性的多尺度特征表示。此外，在通用水果检测数据集上的良好表现也证明了该框架具有良好的泛化能力，为农业场景及其他领域的多尺度目标检测提供了一种有效解决方案。

3.5.3 消融实验

为验证本章提出 IEM、CSAM 和 GCAF 模块在 HRAGNet 中的有效性，以 Sparse R-CNN 作为基准模型，在 PlantDoc、番茄叶片疾病（TLD）和水果检测（FD）三个数据集上进行系统的消融实验。逐步添加各模块并观察其对检测性能的影响。主要评价指标为 AP（平均精度均值），同时辅以 AP₅₀ 和 AP₇₅ 进行对比分析。

（1）PlantDoc 数据集上的消融实验

PlantDoc 数据集包含复杂背景和多尺度病斑，对模块的语义增强、空间对齐和

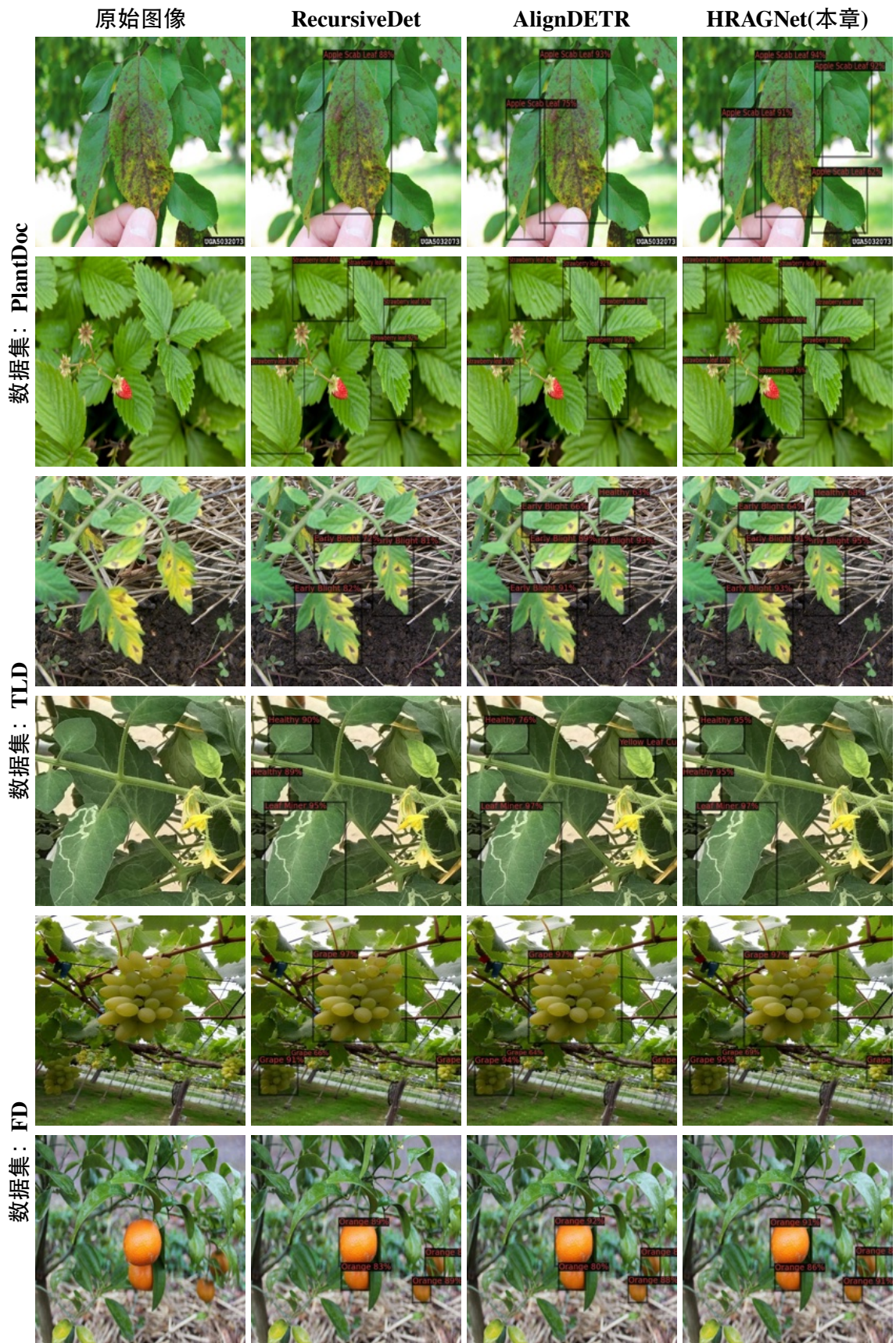


图 3.6 不同模型在 PlantDoc、TLD 和 FD 数据集上的检测结果可视化对比

表 3.7 各组件在 PlantDoc 数据集上的消融实验结果

IEM	CSAM	GCAF	AP	AP ₅₀	AP ₇₅
			35.9	52.4	39.5
✓			39.3	56.3	42.5
	✓		38.9	56.8	41.1
		✓	39.2	56.6	40.8
✓		✓	40.5	59.9	45.3
✓	✓		40.2	59.4	44.1
	✓	✓	42.8	62.0	46.8
✓	✓	✓	43.8	64.4	47.8

全局融合能力均有较高要求。表 3.7 展示了不同模块组合下的检测结果。可以看出，基准模型 Sparse R-CNN 的 AP 为 35.9%。单独添加 IEM 后，AP 提升至 39.3%，相比基准模型提高了 3.4%，验证了递归增强机制在语义传递中的有效性。IEM 通过融合、增强与传递机制，使高层语义信息在层级传播过程中不断得到强化，从而显著提升特征的判别能力。单独添加 CSAM 带来 38.9% 的 AP，表明可学习的语义流场能够有效缓解跨尺度特征错位问题，增强空间精确性。单独添加 GCAF 获得 39.2% 的 AP，说明全局注意力融合能够整合多尺度上下文，抑制背景噪声。

进一步观察两两组合的结果可以看到，IEM + CSAM 组合达到 40.2% 的 AP，较单模块进一步提升，证明语义增强后的特征经过空间对齐能更好地保留细节；IEM + GCAF 组合达到 40.5% 的 AP，显示递归增强与全局融合的协同作用；CSAM + GCAF 组合更是达到 42.8% 的 AP，显著高于其他两两组合，说明在空间对齐基础上进行全局融合是提升性能的关键。最终，三个模块共同作用使 AP 达到 43.8%，较基准提升 7.9%，充分证明了模块间的互补性与递进式设计的合理性。

(2) 番茄叶片疾病数据集 (TLD) 上的消融实验

TLD 数据集中的病害尺度分布广泛，且存在相似类别混淆的挑战。表 3.8 展示了各模块在 TLD 上的消融结果。可以看到，基准模型 AP 为 78.4%。单独添加 IEM 使 AP 提升至 80.2%，验证了递归增强对病害区域语义信息的强化作用。单独添加 CSAM 的 AP 值为 80.5%，AP 提升 2.1%，同时 AP₇₅ 从 83.9% 提升至 86.9%，表明精确的空间对齐对精细定位尤为关键。单独添加 GCAF 获得 79.9% 的 AP，说明全局上下文融合有助于区分相似病害。

在两两组合中，CSAM + GCAF 组合达到 81.2% 的 AP，高于组合 IEM + CSAM

和 IEM + GCAF 的 AP 值，再次凸显了空间对齐与全局融合协同的重要性。值得注意的是，IEM + GCAF 组合的 AP 低于单独添加 IEM 和 CSAM，这可能是因为在 TLD 数据集中，空间错位问题较为突出，仅靠语义增强和全局融合而缺乏对齐会导致信息叠加错误，反而影响性能。最终，三个模块共同作用使 AP 达到 81.9%，较基准提升 3.5%，同时 AP_{50} 和 AP_{75} 均达到最优，验证了模块组合的全面优势。

表 3.8 各组件在 TLD 数据集上的消融实验结果

IEM	CSAM	GCAF	AP	AP_{50}	AP_{75}
			78.4	88.9	83.9
✓			80.2	89.6	83.6
	✓		80.5	91.2	86.9
		✓	79.9	89.6	84.6
✓		✓	80.1	89.9	83.7
✓	✓		80.9	90.7	86.7
	✓	✓	81.2	91.0	86.3
✓	✓	✓	81.9	91.2	86.9

(3) 水果数据集 (FD) 上的消融实验

为验证模块在通用目标检测任务中的泛化能力，在 FD 数据集上进行消融实验。该数据集包含来自 COCO 和 PASCAL VOC 的图像，具有遮挡、光照变化等挑战。表 3.9 展示了各模块的贡献。可以看到，基准模型 Sparse R-CNN 的 AP 为 37.5%。单独添加 IEM 后 AP 提升至 39.2%，提升 1.7%，说明递归增强机制对水果的纹理和形状特征提取同样有效。单独添加 CSAM 带来 38.8% 的 AP，提升 1.3%，表明可变形对齐能够修正因遮挡或视角变化导致的空间偏移。单独添加 GCAF 获得 38.4% 的 AP，提升 0.9%，验证了全局注意力融合在复杂背景中突出目标的能力。三个模块共同作用使 AP 达到 40.6%，较基准提升 3.1%，超过任意单模块或两两组合的简单叠加，证明模块之间具有强互补性，能够协同提升多尺度特征表示的质量。

三个数据集上的消融实验表明，IEM、CSAM 与 GCAF 均能有效提升模型性能。其中，IEM 通过递归增强缓解语义衰减，在各数据集上均带来稳定增益，在小目标占比较高的 PlantDoc 上提升更为明显；CSAM 通过语义流场实现跨尺度对齐，在 AP_{75} 上提升突出，体现出更强的定位能力；GCAF 通过全局上下文建模改善特征融合效果，虽单独增益有限，但与其他模块结合时可进一步提升整体性能。组合实验进一

表 3.9 各组件在 FD 数据集上的消融实验结果

模块	AP	AP 提升量
基准	37.5	-
+IEM	39.2	+1.7
+CSAM	38.8	+1.3
+GCAF	38.4	+0.9
本章方法	40.6	+3.1

步表明，多模块协同优于单一模块，三者联合时性能最佳，验证了所提递进式特征优化设计的有效性。

3.5.4 IEM 模块有效性分析

为了进一步论证 IEM 模块内部机制设计的科学性，本章在 PlantDoc 数据集上进行了模块的有效性分析。

(1) IEM 模块与其他注意力机制的有效性验证

为验证 IEM 模块的特征增强能力，将 ECA、CBAM 与 CA 三种代表性注意力机制嵌入基线模型特征提取阶段，并与 IEM 进行对比实验。结果如表 3.10 所示，各方法均带来不同程度的性能提升，其中 IEM 在 AP、AP₅₀ 与 AP₇₅ 指标上均取得最优表现，体现了其在细粒度特征建模与表达方面的优势。该结果进一步验证了 IEM 在特征增强任务中的有效性。

表 3.10 不同注意力模块与 IEM 模块的对比结果

注意力模块	AP	AP ₅₀	AP ₇₅
+ ECA ^[78]	37.7	55.1	40.8
+ CBAM ^[42]	37.2	53.7	40.2
+ CA ^[79]	38.0	55.5	41.6
+ IEM(本章)	38.3	55.9	41.9

(2) IEM 模块中特征分组数的影响分析

为验证分组数 G 对模型性能的影响，本章进行对比实验，实验结果如表 3.11 所示。随着 G 从 1 增大至 8，模型性能持续提升，AP 由 34.1% 提升至 39.3%，表明适当的通道分组有助于增强特征表达能力。当 $G = 8$ 时模型取得最优性能，此时在特征多样性与表征能力之间达到较好平衡。然而，当 G 进一步增加至 16 时，性能出现

明显下降 (35.1%)，说明过细的通道划分会导致特征表达碎片化，从而影响检测效果。实验结果表明，在 $G = 8$ 的配置下，IEM 的双分支结构能够更有效地协同工作，从而进一步提升模型的整体检测性能。

表 3.11 IEM 模块中不同特征分组数的消融实验

G	AP	AP ₅₀	AP ₇₅
1	34.1	48.1	35.4
2	36.2	53.8	38.9
4	37.9	55.2	40.6
8	39.3	56.3	41.1
16	35.1	50.9	38.2

(3) IEM 模块中感受野配置分析

本章在 PlantDoc 数据集上，评估了 IEM 模块中细节感知增强路径内不同多尺度卷积核配置的效果，结果如表 3.12 所示。结果表明，仅使用 3×3 卷积核能取得 37.1% 的 AP，而将 3×3 与 5×5 结合后 AP 提升至 37.5%，AP₅₀ 和 AP₇₅ 分别达到 55.6% 和 40.0%。实验表明， 3×3 卷积核擅长捕捉细粒度边缘和点状纹理，而 5×5 卷积核能够感知稍大范围的局部上下文模式。两者的组合能够在保持计算效率的同时，覆盖从微观到中观的局部特征尺度，这对于区分真实病斑与类似背景干扰（如叶脉阴影、水滴反光等）更有效。引入 7×7 卷积核后 AP 降至 36.8%，说明过大的感受野会引入背景噪声并过度平滑小病斑。因此， $k = \{3, 5\}$ 的配置在局部细节与上下文信息间实现了最优平衡。

表 3.12 IEM 不同感受野的消融实验

k	AP	AP ₅₀	AP ₇₅
3	37.1	55.1	39.4
3, 5	37.5	55.6	40.0
3, 5, 7	36.8	54.2	39.2

(4) 基于 IEM 模块的特征图可视化

为验证所提 IEM 模块在增强特征聚焦能力上的有效性，本章以 ResNet-50 为骨干网络，在叶片图像上进行注意力热力图可视化。如图 3.7 所示，原始骨干网络与其

他对比注意力机制的激活区域分布较为分散，通常无法精准定位病斑区域。相比之下，融入 **IEM** 后，模型即使在复杂背景下也能将注意力高度集中在病害区域，同时有效抑制无关的叶脉与周围纹理干扰。这种聚焦式激活特征证明 **IEM** 模块能够增强判别性特征提取能力，提升病斑区域的定位精度。

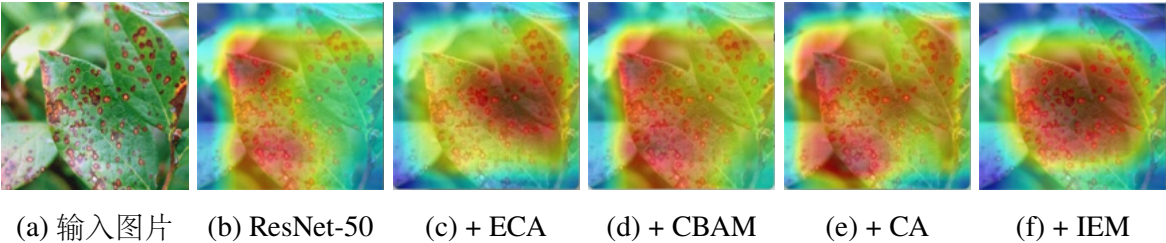


图 3.7 不同注意力机制的特征图可视化

3.5.5 CSAM 模块有效性分析

为了进一步论证 **CSAM** 模块内部机制设计的科学性，本章在 **PlantDoc** 数据集上进行了模块的有效性分析。

(1) **CSAM** 模块与其他常用特征对齐的实验

为了进一步验证 **CSAM** 模块在跨尺度特征对齐方面的优势，本章将其与多种常见特征对齐策略进行比较，包括最近邻插值 (**Nearest Neighbor**)、双线性插值 (**Bilinear Interpolation**) 以及基于可变形卷积的对齐方法 (**Deformable Convolution**)。实验中所有对齐方法均以 **IEM** 输出的多尺度语义增强特征图为输入，输出对齐后的特征送入相同的 **GCAF** 模块和检测头进行预测。

表 3.13 不同跨尺度特征对齐策略性能比较

对齐方法	mAP (%)	参数量 (M)
最近邻插值对齐	36.4	0
双线性插值对齐	36.6	0
基于可变形卷积对齐	37.2	2.5
本章方法	37.7	3.1

如表 3.13 所示，最近邻插值对齐与双线性插值对齐均为固定采样策略，不具备学习能力，无法适应叶片弯曲、拍摄角度变化及病斑形态不规则带来的复杂空间偏移，**mAP** 最低。基于可变形卷积对齐作为局部可学习的对齐方法，能够在感受野范围内自适应调整特征位置，**mAP** 提升至 37.2%，但在跨尺度、全局几何变形方面仍

存在局限。**CSAM** 通过学习跨尺度语义流场实现全局像素级对齐, **mAP** 达到 37.7%, 较基于可变形卷积对齐方法提升 0.5%, 参数量仅增加 0.6M, 证明了其在跨尺度特征对齐和保留病斑细粒度结构方面的优势。综上所述, **CSAM** 能够显著增强多尺度特征的协同表达能力, 为后续 **GCAF** 模块提供更准确的输入, 从而提升叶片病害检测的整体性能。

(2) 选择不同对齐参考层对 **CSAM** 模块性能影响的实验

为验证 **CSAM** 模块中不同对齐参考层的选择对检测性能的影响, 本实验以 **IEM** 输出的多尺度特征金字塔 $\{P_1, P_2, P_3, P_4\}$ 为输入, 分别选取 P_1 、 P_2 、 P_3 作为空间对齐的参考基准, 对其余尺度特征执行 **CSAM** 变形对齐。具体而言, 以参考层为目标分辨率, 通过预测语义流场将其他层特征变形至该分辨率, 随后经 **GCAF** 全局融合及检测头完成预测。实验保持网络其他组件及超参数一致, 仅改变参考层选择。

表 3.14 不同对齐参考层对 **CSAM** 模块性能的影响

对齐参考层	mAP (%)	AP₅₀ (%)	AP₇₅ (%)
P_1 (最高分辨率, 1/4)	38.9	56.8	41.1
P_2 (1/8)	38.1	55.9	39.2
P_3 (1/16)	37.6	54.5	38.6

如表 3.14 所示, 以最高分辨率层 P_1 作为参考特征时模型取得最优性能, **mAP** 为 38.9%, 相较于 P_2 与 P_3 均有明显提升。该结果表明, 高分辨率特征在跨尺度对齐中更具优势, 其能够保留更丰富的空间细节与边界信息, 从而为其他尺度特征提供更精确的对齐基准。同时, 在以小目标病斑为主的检测场景中, 高分辨率特征有助于减少细粒度信息的丢失, 并提升语义流场估计的稳定性。因此, 本章选择 P_1 作为 **CSAM** 的参考层, 以实现更精确的跨尺度空间对齐。

3.5.6 **GCAF** 模块有效性分析

为了验证 **GCAF** 模块在特征融合过程中的作用, 本章在 **PlantDoc** 数据集上设计实验, 以 **IEM** 与 **CSAM** 模块输出的多尺度对齐特征为输入, 在保持网络结构和训练参数一致的情况下, 仅改变全局融合阶段的通道权重生成方式, 对比四种策略: (1) 无注意力机制, 仅进行全局特征拼接与卷积融合; (2) 仅全局聚合, 使用全局特征融合但不生成通道权重; (3) 静态通道注意力, 采用固定结构进行特征加权; (4) 动态通道注意力 (本章方法), 通过全局上下文描述子自适应生成通道权重。实验结果

如表 3.15 所示。

表 3.15 通道注意力消融实验

模型结构	全局特征聚合	注意力机制	mAP
基础融合结构	否	无	37.5
全局聚合结构	是	无	38.4
静态注意力结构	是	静态通道注意力	38.7
本章方法 (GCAF)	是	动态通道注意力	39.2

从表 3.15 可以看出，全局特征聚合使 mAP 提升 0.9%，验证了跨尺度语义信息对病害检测的重要价值。进一步引入通道注意力机制后，模型性能持续提升，表明对不同通道重要性进行建模能够有效改善多尺度特征融合质量。其中，本章提出的动态通道注意力机制取得最佳检测效果，其通过全局上下文描述子自适应生成通道权重，使模型更加关注病斑纹理、颜色及结构形态等关键特征，同时抑制背景叶片纹理的干扰。这一结果说明，动态通道注意力机制能够在特征融合过程中实现更精细的通道选择，从而显著提升特征表达能力与整体检测性能。

3.6 本章小结

本章针对植物叶片病害检测任务中尺度变化显著、空间结构复杂以及多尺度特征融合不足等问题，提出了一种分层递归对齐与全局协同融合网络 HRAGNet。该方法从多尺度特征金字塔的信息交互机制出发，对传统单向传递的特征融合方式进行改进，构建了递归增强、空间对齐与全局融合相结合的层次化特征优化框架。在模型设计上，以 Sparse R-CNN 为基础检测框架，并构建由 IEM 模块、CSAM 模块以及 GCAF 模块组成的特征增强编码器。三个模块通过递进式协同作用，实现多尺度特征的语义强化、空间对齐与上下文融合，从而显著提升特征表达的判别能力与结构一致性。在实验验证方面，通过在多个植物病害数据集上的对比实验与消融分析，对所提出方法进行了系统评估。实验结果表明，HRAGNet 在整体检测性能上优于多种主流目标检测方法，验证了所提特征增强策略在农业视觉场景中的有效性。

第四章 基于几何先验与任务感知的叶片病害检测算法

第三章提出的 HRAGNet 通过构建多尺度增强编码器，结合交互式增强模块、跨尺度语义对齐模块与全局上下文聚合融合模块三个核心组件实现了多尺度特征的高效增强与融合，有效提升了叶片病害检测的特征表达能力。但在复杂农业场景下，仅依靠编码阶段的特征融合和增强仍难以充分建模候选目标的空间结构，且无法缓解分类与回归任务间的特征冲突，检测性能仍有提升空间。现有解码器通常缺乏对病斑空间形态的显式建模，且多采用任务共享特征，易在相似病害类别间产生误检与漏检。

为此，本章在 HRAGNet 的基础上提出一种增强型植物叶片病害检测框架 HRAGNet+ (Hierarchical Recursive Alignment and Global Collaborative Fusion Network Plus)。在保留 HRAGNet 编码器的基础上对解码环节进行重构，提出一种几何-任务协同增强解码器。其中，几何先验增强注意力模块 (Geometric Prior Enhancement Attention, GPEA) 将候选框空间关系与交并比约束引入注意力机制，强化不规则病斑的定位精度；任务特征解耦与融合模块 (Task Decoupling and Fusion Module, TDFM) 利用动态通道掩码实现分类与回归特征解耦以减少任务干扰，并通过自适应融合对解耦后的特征进行重构，生成下一阶段迭代所需的提案特征。通过上述设计，HRAGNet+ 在保持多尺度特征融合优势的基础上，进一步引入目标级结构建模与任务级特征重构机制，使模型能够在复杂叶片病害场景下同时提升空间定位精度与语义判别能力，从而获得更加稳定且鲁棒的检测性能。

4.1 模型总体设计

4.1.1 整体架构与模型输入输出

HRAGNet+ 采用编码器-解码器式目标检测框架。该框架结合了第三章提出的多尺度特征增强编码器，以及本章提出的几何-任务协同增强解码器。其核心思想是在高质量特征表达的基础上，通过引入空间结构建模与任务解耦机制，实现对目标提案特征的逐阶段精细化优化，从而在保持多尺度特征表达能力的同时，进一步提升了目标定位精度与分类判别能力。模型整体框架如图 4.1 所示。

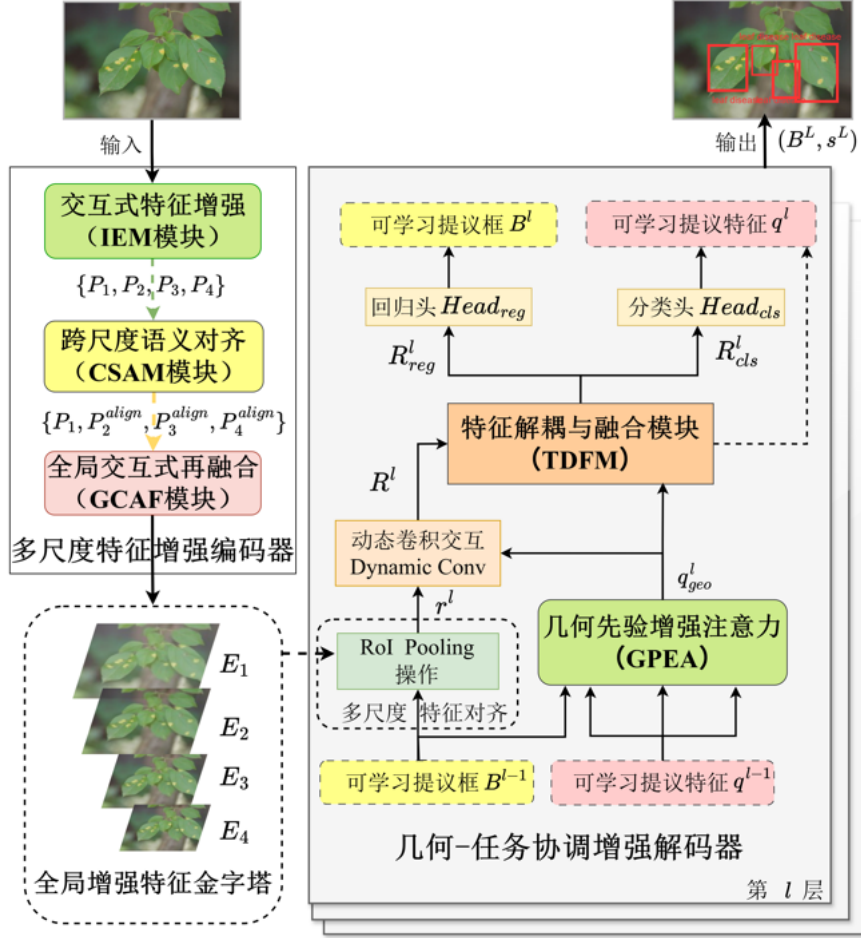


图 4.1 HRAGNet+ 模型整体框架图

给定输入图像 $I \in \mathbb{R}^{3 \times H_0 \times W_0}$ ，模型首先通过骨干网络以及第三章提出的多尺度特征增强编码器进行特征提取与增强，得到全局增强特征金字塔 $\{E_1, E_2, E_3, E_4\}$ 。该特征表示在语义上具有一致性，同时在空间上保持良好的对齐特性，为后续目标检测提供统一且高质量的特征基础。

在解码阶段，模型以一组可学习的提案表示作为初始输入，包括提案框 $B^0 \in \mathbb{R}^{N \times 4}$ 与对应的提案特征 $q^0 \in \mathbb{R}^{N \times d}$ ，其中 N 表示提案数量， d 表示特征维度。通过多阶段级联结构，模型对提案进行迭代更新，即上一阶段的输出 (B^{l-1}, q^{l-1}) 被用于当前第 l 阶段的输入，并与全局增强特征金字塔 $\{E_1, E_2, E_3, E_4\}$ 共同输入几何-任务协同增强解码器，得到更新后的提案表示 (B^l, q^l) ，使提案在空间位置与语义表示上逐步逼近真实目标。

在该过程中，模型通过引入几何关系建模与任务解耦机制，使提案特征在空间约束与任务引导下逐步优化。经过 L 轮迭代后，模型输出最终的目标检测结果，包括预测边界框集合 $B^L \in \mathbb{R}^{N \times 4}$ 以及对应的类别置信度 $s^L \in \mathbb{R}^{N \times C}$ ，其中 L 表示解码

阶段的总迭代次数, C 表示类别数。

4.1.2 几何-任务协同增强解码器

几何-任务协同增强解码器以级联形式构建。以第 l 个解码阶段为例, 每个阶段的解码器接收三类输入: 上一阶段的提案框 $B^{l-1} \in \mathbb{R}^{N \times 4}$ 、提案特征 $q^{l-1} \in \mathbb{R}^{N \times d}$, 以及多尺度特征增强编码器 (第三章) 输出的多尺度特征 $\{E_1, E_2, E_3, E_4\}$ 。经过几何先验增强注意力模块 (GPEA) 和任务特征解耦与融合模块 (TDFM) 处理后, 输出更新后的提案框 B^l 与提案特征 q^l , 并传递至下一阶段, 每个阶段依次包含如下四个步骤。

(1) RoIAlign 操作下的区域特征提取

在第 l 个解码阶段中, 模型首先根据当前提案框 B^{l-1} 的空间位置, 通过 RoIAlign 操作从多尺度特征金字塔 $\{E_i\}$ 中提取对应的区域特征 r^l , 为每个候选目标提供与其空间位置对齐的局部视觉信息。具体而言, RoIAlign 操作根据提案框 B^{l-1} 的坐标信息, 在多尺度特征图上进行双线性插值采样, 生成固定尺寸的区域特征表示 $r^l \in \mathbb{R}^{N \times C \times H_r \times W_r}$, 其中 C 表示通道数, H_r 和 W_r 分别表示 RoIAlign 输出的空间尺寸, 采用 7×7 大小。该操作确保了提案特征与其对应的视觉信息在空间上的精确对齐, 为后续的特征交互与任务解耦提供了可靠的输入基础。

(2) 基于几何先验的增强注意力构建

本章设计了一种基于几何先验的注意力模块 GPEA, 用于建模目标之间的空间结构关系。该模块以提案框 B^{l-1} 与提案特征 q^{l-1} 为输入, 通过显式建模提案之间的几何关系, 对提案特征进行空间感知增强。具体而言, 首先基于提案框计算两两之间的 IoU 关系, 构建几何先验矩阵; 随后, 将该几何先验引入注意力权重计算过程, 对原始注意力进行重加权, 从而强化空间上相关提案之间的特征交互, 抑制无关干扰。最终输出融合几何信息的提案特征 q_{geo}^l , 该几何增强的特征不仅保留原始语义信息, 同时显式编码了目标间的空间关系。

在获得几何增强后的提案特征 q_{geo}^l 后, 首先通过动态卷积机制实现提案特征与区域特征的交互。具体地, 以 q_{geo}^l 为条件生成动态卷积参数, 并作用于 RoIAlign 操作提取到的区域特征 r^l , 得到动态交互特征 R^l 。该特征融合了提案的语义信息与局部视觉信息, 具备更强的表达能力。

(3) 动态特征交互与任务特征解耦与融合

在上述几何增强建模的基础上，本章进一步设计 **TDFM** 模块，用于缓解分类与回归任务之间的特征竞争问题。该模块以动态交互特征 R^l 与几何增强提案 q_{geo}^l 为联合输入，通过学习任务特定的通道权重，对共享特征进行自适应重标定，从而生成解耦后的两类特征 R_{cls}^l 和 R_{reg}^l 。其中， R_{cls}^l 强调语义判别信息，用于分类任务； R_{reg}^l 突出空间结构信息，用于边界框回归。进一步地，为形成统一的提案表示并支持级联解码过程，解耦后的特征将通过自适应协同融合机制进行重组，生成当前阶段的提案特征 q^l ，并作为下一阶段解码器的输入。

(4) 并行预测与特征迭代更新

基于解耦后的特征，分别通过分类头与回归头进行预测：

$$\Delta s^l = \text{Head}_{cls}(R_{cls}^l), \quad \Delta b^l = \text{Head}_{reg}(R_{reg}^l) \quad (4.1)$$

其中 Δs^l 表示由分类头预测得到的类别置信度分布， Δb^l 表示边界框回归偏移量。

进而基于回归结果对提案框进行迭代更新：

$$B^l = B^{l-1} + \Delta b^l \quad (4.2)$$

同时，上一阶段通过自适应协同融合机制生成的提案特征 q^l ，将与更新后的提案框 B^l 一同传递至下一阶段解码器，将作为下一阶段解码器的输入，实现提案特征在多阶段中的逐层演化。

通过多阶段迭代，几何-任务协同增强解码器能够逐步优化提案框位置并提升特征表达能力，从而获得更加精确的检测结果。

4.1.3 优化目标与损失函数

HRAGNet+ 的优化目标是一个作用于最终阶段输出的多任务联合损失函数，用于同时约束分类准确性与定位精度。遵循 **Sparse R-CNN** 的集合预测方式，首先通过匈牙利二分匹配为预测框分配唯一真实标签，再计算加权联合损失，其形式如下：

$$\mathcal{L} = \lambda_{cls} \mathcal{L}_{cls}(s^L, \hat{s}) + \lambda_{reg} \mathcal{L}_{reg}(B^L, \hat{B}) + \lambda_{giou} \mathcal{L}_{giou}(B^L, \hat{B}) \quad (4.3)$$

其中， $\hat{s}$ 和 $\hat{B}$ 分别表示真实的类别标签与边界框坐标， s^L 和 B^L 为模型最终阶段的分类预测与定位预测，超参数 λ_{cls} 、 λ_{reg} 和 λ_{giou} 用于平衡各项损失的贡献。参照 **Sparse**

R-CNN 等目标检测模型的通用超参数设定并结合 HRAGNet+ 模型结构特性, 实验设置 $\lambda_{\text{cls}} = 2.0$ 、 $\lambda_{\text{reg}} = 5.0$ 、 $\lambda_{\text{giou}} = 2.0$, 以保证分类与定位任务优化权重配比合理。

分类损失 $\mathcal{L}_{\text{cls}}$ 采用 Focal Loss, 用于缓解类别不平衡问题, 其定义为:

$$\mathcal{L}_{\text{cls}} = -\alpha(1 - p_t)^\gamma \log(p_t) \quad (4.4)$$

其中 p_t 表示模型对真实类别的预测概率, α 与 γ 分别用于调节类别不平衡程度以及困难样本的关注程度。参考 Focal Loss 中的典型设置, 并结合目标检测任务中前景与背景类别分布不均的特点, 实验设置 $\alpha = 0.25$ 、 $\gamma = 2.0$ 。

定位回归损失记为 $\mathcal{L}_{\text{reg}}$, 由平滑 L1 损失 (Smooth L1 Loss) 构成。平滑 L1 损失定义为:

$$\mathcal{L}_{\text{smoothL1}}(b, \hat{b}) = \sum_{i \in \{x, y, w, h\}} \text{smooth}_{L1}(b_i - \hat{b}_i) \quad (4.5)$$

$$\text{smooth}_{L1}(x) = \begin{cases} 0.5x^2, & |x| < 1, \\ |x| - 0.5, & \text{otherwise.} \end{cases} \quad (4.6)$$

广义交并比损失 (Generalized IoU Loss) 定义为:

$$\mathcal{L}_{\text{giou}} = 1 - \text{GIoU}(b, \hat{b}) = 1 - \text{IoU}(b, \hat{b}) + \frac{|C \setminus (b \cup \hat{b})|}{|C|} \quad (4.7)$$

其中 C 表示能够同时覆盖预测框 b 与真实框 $\hat{b}$ 的最小闭合区域。

4.2 几何先验增强注意力模块 (GPEA)

在基于 Transformer 的目标检测框架中, 解码器通常通过自注意力机制建模提案之间的关系。然而, 标准自注意力主要依赖特征语义相似性进行信息交互, 缺乏对提案间空间几何关系的显式建模能力, 难以有效刻画目标之间的结构约束。在植物叶片病害检测任务中, 病斑区域常呈现局部聚集或沿叶脉分布的空间特性, 单纯依赖语义相似性容易导致空间邻近病斑之间的特征混淆, 并受到复杂背景纹理的干扰, 从而影响检测性能。

针对上述问题, 本章提出几何先验增强注意力模块 (Geometric Prior Enhancement Attention, GPEA), 在解码阶段显式引入提案之间的空间几何关系, 对自注意力权重

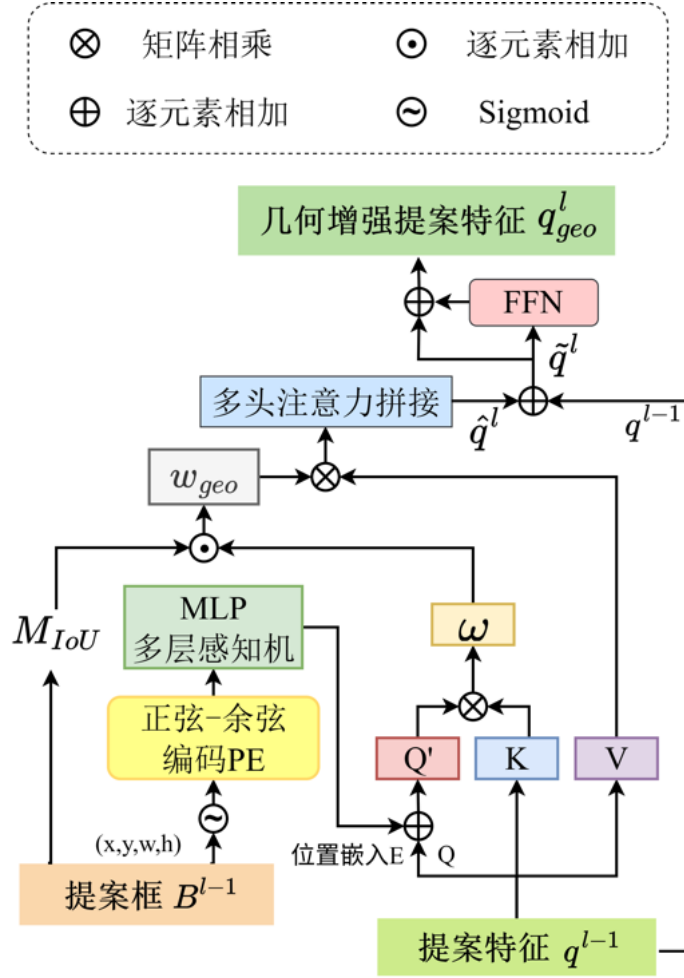


图 4.2 几何先验增强注意力模块（GPEA）结构图

进行约束，使模型在建模语义关联的同时兼顾空间结构信息，从而提升 HRAGNet+ 对病斑分布模式的表达能力。在第 l 个解码阶段中，GPEA 的输入包括上一阶段的提案特征 $q^{l-1} \in \mathbb{R}^{N \times d}$ 和提案框 $B^{l-1} \in \mathbb{R}^{N \times 4}$ ；输出为新的几何增强的提案特征 q_{geo}^l 。GPEA 模块的具体实现如图 4.2 所示。

4.2.1 位置编码与位置感知提案特征的构建

在第 l 个解码阶段中，输入的提案查询特征记为 $q^{l-1} \in \mathbb{R}^{N \times d}$ ，其中 N 表示提案数量， d 表示特征维度。为了构建自注意力计算所需的查询特征，首先通过线性投影将上一阶段的查询特征映射为查询矩阵：

$$Q = q^{l-1} W_Q \quad (4.8)$$

其中 $W_Q \in \mathbb{R}^{d \times d}$ 为可学习的线性映射参数。

为了将提案的绝对空间信息引入特征表示,需要对边界框参数进行位置编码。设第 i 个提案框为

$$B_i^{l-1} = (x_i, y_i, w_i, h_i) \quad (4.9)$$

其中 (x_i, y_i) 表示提案框中心坐标, w_i 和 h_i 分别表示边界框的宽度和高度。

为使模型能够感知提案在图像中的具体位置,本章采用正弦-余弦位置编码方式对边界框参数进行映射。对于任意坐标分量 $p \in \{x_i, y_i, w_i, h_i\}$, 其位置编码定义为:

$$\text{PE}(p) = \left[\sin \left(\frac{p}{10000^{2k/d}} \right), \cos \left(\frac{p}{10000^{2k/d}} \right) \right] \quad (4.10)$$

其中 k 表示编码维度索引, d 为嵌入维度。该编码方式能够将连续的坐标信息映射到高维空间,从而增强模型对空间位置变化的表达能力。

随后,将边界框四个参数的编码结果进行拼接,并通过多层感知机 (MLP) 进行非线性映射,得到第 i 个提案的空间位置嵌入:

$$e_i = \text{MLP}(\text{PE}(b_i^{l-1})) \quad (4.11)$$

设所有提案的空间嵌入为:

$$E = [e_1, e_2, \dots, e_N] \quad (4.12)$$

其中 N 表示提案数量。

将位置嵌入 E 与查询特征 Q 进行逐元素相加,得到位置感知的提案特征:

$$Q' = Q + E \quad (4.13)$$

通过引入位置编码,查询特征 Q' 同时包含语义信息和绝对空间位置信息,为后续几何关系建模提供基础。在植物病害检测任务中,病斑往往在叶片的特定区域出现,例如部分病害更容易分布在叶缘或叶尖位置。引入位置感知特征能够帮助模型更好地捕捉这些空间分布规律,从而提升对病斑区域的识别能力。

4.2.2 基于 IoU 的几何关系建模机制

在获取每个提案框的绝对位置信息后,提案之间的相对空间关系对于理解全局场景并区分独立病斑同样具有重要作用。若两个提案框在空间上高度重叠,则它们

很可能对应同一病斑区域的不同部分；而空间距离较远的提案则更可能属于不同病斑或背景区域。为提高注意力机制对空间关系的感知能力，并减少无关区域之间的干扰，本章引入 IoU（Intersection over Union, IoU）几何先验，对注意力权重进行空间约束。

首先计算提案之间的几何关系矩阵 $M_{\text{IoU}} \in \mathbb{R}^{N \times N}$ ，其中矩阵元素 $M_{\text{IoU}}(i, j)$ 表示第 i 个与第 j 个提案框之间的 IoU：

$$M_{\text{IoU}}(i, j) = \frac{|B_i^{l-1} \cap B_j^{l-1}|}{|B_i^{l-1} \cup B_j^{l-1}|} \quad (4.14)$$

该矩阵刻画了提案之间的空间重叠程度，可作为几何先验引导注意力权重的分布。当两个提案具有较高的 IoU 时，它们更可能对应同一病斑区域或相邻病斑，应加强信息交互；反之，若 IoU 较低，则表明两者空间相关性较弱，其特征交互应适当抑制。

在此基础上，首先由提案特征 q^{l-1} 通过线性投影生成键和值向量：

$$K = q^{l-1} W_K, \quad V = q^{l-1} W_V \quad (4.15)$$

其中 W_K 与 W_V 为可学习的投影矩阵。

随后，利用位置增强后的查询向量 Q' 与键向量 K 计算标准缩放点积注意力权重矩阵：

$$\omega = \text{Softmax} \left(\frac{Q' K^T}{\sqrt{d}} \right) \quad (4.16)$$

其中 $\omega \in \mathbb{R}^{N \times N}$ 表示基于特征相似性计算得到的注意力权重矩阵， d 为键向量的维度，缩放因子 $\sqrt{d}$ 用于稳定训练过程。

由于上述注意力权重仅依赖特征相似性而未考虑空间关系，因此进一步利用 IoU 几何先验对其进行引导。具体地，将原始注意力权重矩阵 ω 与 IoU 矩阵 M_{IoU} 进行逐元素相乘，得到引入几何关系约束后的注意力权重矩阵 ω_{geo} ：

$$\omega_{\text{geo}} = \omega \odot M_{\text{IoU}} \quad (4.17)$$

其中 $\odot$ 表示逐元素乘法运算。

基于几何增强后的注意力权重，对值向量进行加权聚合，得到单头几何增强的

输出表示：

$$Z = \omega_{\text{geo}} V \quad (4.18)$$

在此基础上，该机制可自然扩展至多头注意力形式。对于第 h 个注意力头，其几何增强表示为：

$$Z_h = \omega_{\text{geo}}^h V_h \quad (4.19)$$

最终，将所有头的输出进行拼接并通过线性映射得到第 l 层更新后的提案特征 $\hat{q}^l$ ：

$$\hat{q}^l = \text{Concat}(Z_1, Z_2, \dots, Z_H) W_O \quad (4.20)$$

其中 H 表示注意力头的数量， W_O 为输出投影矩阵。

通过上述几何先验引导机制，模型能够强化空间上相关提案之间的信息交互，同时抑制无关区域之间的噪声传播。在植物病害检测任务中，该机制有助于促进同一病斑区域内部特征的聚合，并减少来自叶片背景纹理或其他区域的干扰，从而提升模型对聚集性病斑结构的整体感知能力。

4.2.3 几何增强提案特征更新

在获得几何增强的上下文特征后，需要对提案表示进行更新。为保证信息传递的稳定性与特征表达能力，模块采用残差连接与前馈网络（Feed-Forward Network, FFN）相结合的方式完成特征更新。

首先，将几何注意力输出 $\hat{q}^l$ 与上一阶段的提案特征 q^{l-1} 进行残差融合，得到初步融合特征 $\tilde{q}^l$ ：

$$\tilde{q}^l = q^{l-1} + \hat{q}^l \quad (4.21)$$

随后，将初步融合特征 $\tilde{q}^l$ 输入前馈网络进行非线性变换，并再次通过残差连接得到最终的提案特征 q_{geo}^l ：

$$q_{\text{geo}}^l = \tilde{q}^l + \text{FFN}(\tilde{q}^l) \quad (4.22)$$

其中, $\text{FFN}(\cdot)$ 通常由两层全连接网络及中间的非线性激活函数构成, 其作用是进一步提升特征的表达能力, 使其更适合后续的动态卷积与任务特征解耦与融合模块进行处理。

最终得到的几何增强提案特征 q_{geo}^l 作为第 l 个解码阶段的中间表示, 用于后续模块进行进一步处理。通过更新提案特征, 模型能够在语义信息与几何约束的共同作用下不断细化提案表示, 实现对目标位置与类别的逐步逼近, 从而提升整体检测性能。

通过上述设计, **GPEA** 实现了对解码器中提案关系建模机制的系统性增强, 在植物叶片病害检测任务中表现出显著优势。首先, 通过将边界框的绝对位置信息与提案之间的相对空间关系显式引入注意力计算过程, 模型获得了对空间结构的感知能力, 使其能够更好地刻画病斑在叶片中的分布规律, 从而为病害类型判别提供重要的几何依据。其次, 基于 **IoU** 构建的几何先验有效强化了空间相关提案之间的信息交互, 提升了模型在密集病斑场景下的特征表达能力, 能够缓解由于目标重叠带来的漏检与误检问题。

综上, **GPEA** 有效弥补了传统自注意力在空间建模方面的不足, 为后续动态卷积与任务解耦提供了更加可靠的特征基础, 从整体上提升了植物叶片病害检测的精度与稳定性。

4.3 任务特征解耦与融合模块 (TDFM)

在目标检测任务中, 分类与边界框回归通常共享同一特征表示, 但两类任务在语义判别与空间定位上的需求存在显著差异, 容易在联合优化过程中产生任务不一致问题, 从而限制检测性能的提升。该问题在植物叶片病害检测场景中尤为突出, 不同病害类别外观相似, 对语义判别能力要求较高, 而病斑区域形状不规则且尺度变化较大, 又对空间定位精度提出了更高要求。针对上述问题, 本文在 **GPEA** 模块生成的提案特征基础上, 设计了任务特征解耦与融合模块 (**Task Decoupling and Fusion Module, TDFM**)。该模块首先通过构建独立的分类与回归特征子空间, 实现任务特征的解耦表示; 随后引入自适应融合机制对多任务信息进行重组, 以缓解任务间的相互干扰。由此, 模块能够在保持语义判别能力的同时增强空间定位的一致性, 从而有效提升检测性能。

在第 l 个解码阶段中, TDFM 接收几何增强提案特征 $q_{geo}^l \in \mathbb{R}^{N \times d}$ 以及动态交互特征 $R^l \in \mathbb{R}^{N \times C \times 7 \times 7}$ 。其中, q_{geo}^l 来源于 GPEA 模块, 而 R^l 为提案特征与区域特征 r^l 通过动态卷积交互得到的区域表示, 用于后续任务解耦建模。基于上述输入, TDFM 通过任务解耦与特征重组生成分类特征 R_{cls}^l 、回归特征 R_{reg}^l , 以及用于下一阶段迭代更新的提案特征 $q^l \in \mathbb{R}^{N \times d}$ 。TDFM 模块的整体结构如图 4.3 所示。

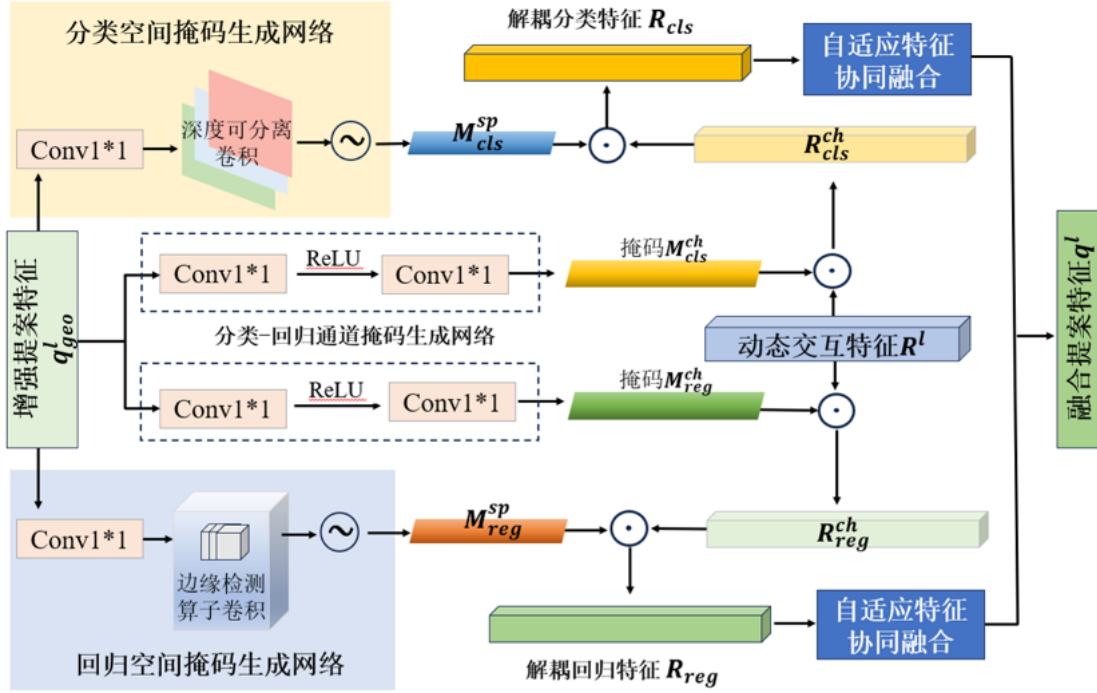


图 4.3 任务特征解耦与融合模块 (TDFM) 结构图

4.3.1 分类-回归通道掩码解耦机制

不同任务在特征关注与信息需求上存在本质差异。分类任务更依赖高层语义信息以判别物体类别, 回归任务则更需要空间结构与边界细节信息以实现精确定位。然而, 在 Sparse R-CNN^[65] 等端到端模型中, 动态交互特征 $R^l \in \mathbb{R}^{C \times H \times W}$ 同时服务于目标分类与边界回归两个任务。若直接共享同一特征, 两类任务的优化目标可能相互干扰。因此, 有必要在特征层面对任务信息进行显式解耦, 以构建任务专属的表示空间。

为此, TDFM 引入任务感知掩码机制, 在通道与空间两个维度上分别生成任务自适应掩码, 逐步将 R^l 解耦为分类特征 R_{cls}^l 与回归特征 R_{reg}^l , 从而在共享特征空间内显式构建分类与回归任务的解耦表示, 实现两类任务的独立优化与互补建模。

(1) 通道维度掩码生成

通道掩码旨在依据任务需求，自适应筛选对特定任务响应更强的特征通道。**TDFM** 利用提案特征 q_{geo}^l 作为引导，将 q_{geo}^l 分别输入两个轻量级多层感知机，生成分类掩码向量 $M_{cls}^{ch} \in \mathbb{R}^C$ 和回归掩码向量 $M_{reg}^{ch} \in \mathbb{R}^C$ ：

$$M_{cls}^{ch} = \sigma(W_2 \delta(W_1 q_{geo}^l)), \quad M_{reg}^{ch} = \sigma(W_4 \delta(W_3 q_{geo}^l)) \quad (4.23)$$

其中， W_1 至 W_4 为可学习的投影矩阵， $\delta(\cdot)$ 为 **ReLU** 激活函数， $\sigma(\cdot)$ 为 **Sigmoid** 函数，将输出值压缩到 $(0, 1)$ 区间，表示每个通道被保留的权重。

随后，通过通道广播机制对动态交互特征 R^l 进行逐通道加权，从而获得经过通道筛选的分类中间特征 $R_{cls}^{ch,l}$ 和回归中间特征 $R_{reg}^{ch,l}$ ：

$$R_{cls}^{ch,l} = M_{cls}^{ch} \odot R^l, \quad R_{reg}^{ch,l} = M_{reg}^{ch} \odot R^l \quad (4.24)$$

其中 $\odot$ 表示沿通道维度的逐元素乘法。

(2) 空间维度掩码生成

即使完成了通道筛选，分类与回归任务所关注的空间位置仍有不同，分类要求整体物体区域响应一致，而回归需要强化边界附近的精细定位。因此，**TDFM** 模块对上述中间特征分别施加空间掩码，进一步调整特征的空间分布。

对于分类空间掩码，为扩大感受野、捕捉物体整体上下文，采用膨胀率 $d = 2$ 的深度可分离卷积 $\text{DWConv}_{3 \times 3}^{d=2}$ 处理 $R_{cls}^{ch,l}$ ，再经 **Sigmoid** 激活得到空间权重图 $M_{cls}^{sp} \in \mathbb{R}^{1 \times H \times W}$ ：

$$M_{cls}^{sp} = \sigma(\text{DWConv}_{3 \times 3}^{d=2}(R_{cls}^{ch,l})). \quad (4.25)$$

对于回归空间掩码，为增强对边缘、角点等局部结构的敏感度，引入基于 **Sobel** 梯度的边缘感知卷积算子 Conv_{edge} ，以显式提取特征图中的一阶导数信息，从而增强对目标边界与局部结构变化的感知能力。再经 **Sigmoid** 得到 $M_{reg}^{sp} \in \mathbb{R}^{1 \times H \times W}$ ：

$$M_{reg}^{sp} = \sigma(\text{Conv}_{edge}(R_{reg}^{ch,l})). \quad (4.26)$$

最后，将空间掩码逐空间位置乘到对应的通道筛选特征上：

$$R_{cls}^l = M_{cls}^{sp} \odot R_{cls}^{ch,l}, \quad R_{reg}^l = M_{reg}^{sp} \odot R_{reg}^{ch,l}. \quad (4.27)$$

至此，原始共享特征 R^l 被显式解耦为分类专用特征 R_{cls}^l 与回归专用特征 R_{reg}^l 。前者通过通道筛选与上下文感受野扩大，突出语义判别性区域；后者通过通道筛选与边缘强化，提升定位精度。该机制有效缓解了多任务学习中的特征冲突，同时保持了整体表示的紧凑性。

4.3.2 自适应特征协同融合

在级联解码框架中，当前阶段经解耦得到的分类特征 R_{cls}^l 与回归特征 R_{reg}^l 不仅用于本阶段的分类与回归预测，还需进一步融合为统一的提案表示，以供下一阶段解码器进行迭代优化。为此，本章设计自适应特征协同融合机制，在保持任务解耦的同时实现跨任务信息的有效整合。

首先，对第 l 阶段的分类特征与回归特征 $R_{\text{cls}}^l, R_{\text{reg}}^l \in \mathbb{R}^{N \times C \times H \times W}$ 进行全局平均池化，提取每个提案在空间维度上的全局汇聚表示，用于分别建模分类特征的语义响应与回归特征的结构响应：

$$z_{\text{cls}}^l = \text{GAP}(R_{\text{cls}}^l), \quad z_{\text{reg}}^l = \text{GAP}(R_{\text{reg}}^l) \quad (4.28)$$

其中 $z_{\text{cls}}^l, z_{\text{reg}}^l \in \mathbb{R}^{N \times C}$ 表示分类与回归特征的全局描述向量。

随后，通过轻量级映射函数估计两类任务对下一阶段推理的重要性权重，并进行归一化：

$$\alpha_{\text{cls}} = \frac{\sigma(\text{FC}(z_{\text{cls}}^l))}{\sigma(\text{FC}(z_{\text{cls}}^l)) + \sigma(\text{FC}(z_{\text{reg}}^l))}, \quad \alpha_{\text{reg}} = \frac{\sigma(\text{FC}(z_{\text{reg}}^l))}{\sigma(\text{FC}(z_{\text{cls}}^l)) + \sigma(\text{FC}(z_{\text{reg}}^l))} \quad (4.29)$$

其中 $\alpha_{\text{cls}} + \alpha_{\text{reg}} = 1$ ，用于在全局层面调节分类特征与回归特征的相对贡献。

为进一步建模分类与回归特征之间的互补关系，引入空间级软门控交互机制。具体地，将两类特征在通道维度拼接后，通过 1×1 卷积生成门控图：

$$G_{\text{gate}} = \sigma(\text{Conv}_{1 \times 1}(\text{Concat}[R_{\text{cls}}^l, R_{\text{reg}}^l])) \quad (4.30)$$

其中 $G_{\text{gate}} \in \mathbb{R}^{N \times 1 \times H \times W}$ 表示空间位置上的融合权重。

基于该门控图，在空间维度对两类特征进行加权融合：

$$R_{\text{fuse}}^l = G_{\text{gate}} \odot R_{\text{cls}}^l + (1 - G_{\text{gate}}) \odot R_{\text{reg}}^l \quad (4.31)$$

随后对融合特征进行全局平均池化，得到交互信息向量：

$$z_{\text{inter}}^l = \text{GAP}(R_{\text{fuse}}^l) \quad (4.32)$$

最终，将多源信息进行加权融合，得到第 l 阶段的提案表示：

$$q^l = \alpha_{\text{cls}} z_{\text{cls}}^l + \alpha_{\text{reg}} z_{\text{reg}}^l + \eta z_{\text{inter}}^l \quad (4.33)$$

其中 η 为可学习参数，用于调节跨任务交互信息的贡献； q^l 为融合后的提案特征，并作为下一阶段解码器的输入。

该机制通过全局权重调节与空间门控交互的协同建模，在保持分类与回归特征解耦的基础上，实现跨任务信息的自适应融合，从而生成兼具语义判别能力与几何表达能力的提案表示。

4.4 编码器-解码器完整分析

4.4.1 完整模型架构与设计流程

本文围绕多尺度特征增强与融合这一核心问题，构建了从特征表示到目标推理的统一优化框架 **HRAGNet+**。整体模型采用端到端目标检测结构，由多尺度特征增强编码器与几何-任务协同增强解码器两部分组成，实现从特征建模到检测的逐阶段优化。模型以输入图像为起点，首先通过编码器对多尺度特征进行增强与融合，构建语义一致且空间对齐的特征表示；随后在解码阶段，通过引入空间结构建模与任务解耦机制，对候选目标进行逐阶段提案优化，从而实现复杂植物叶片病害目标的精准检测。

在特征编码阶段，模型以 **ResNet-50** 作为骨干网络提取基础视觉特征，并结合第三章提出的多尺度特征增强编码器，通过 **IEM**、**CSAM** 和 **GCAF** 模块，对不同层级特征进行逐层增强与跨尺度融合，从而生成多尺度特征金字塔 $\{E_i\}$ 。该过程有效提升了特征表达能力，为后续检测提供高质量输入。

在检测阶段，对解码结构进行重新设计，提出几何-任务协同增强解码器。在第 l 个解码阶段中，首先基于当前提案框 B^{l-1} 从特征金字塔中提取区域特征，通过 **GPEA** 模块显式建模提案之间的空间关系，使特征交互过程具备几何约束。随后，通过动态特征交互机制融合局部视觉信息，并引入 **TDFM** 模块对共享特征进行重标定，将

其映射至分类与回归两个子空间；在此基础上，通过自适应融合策略对两类特征进行整合，生成用于下一阶段迭代的提案特征，从而在缓解多任务学习特征冲突的同时提升提案表示的判别能力。

通过多阶段级联优化，模型实现提案框与提案特征的同步更新，使目标表示逐步由粗粒度语义信息演化为同时具备空间定位能力与语义判别能力的高质量检测表示。**HRAGNet+** 在编码—解码协同框架下，对多尺度特征增强与融合、空间结构建模以及任务解耦优化进行统一建模，从而使模型在复杂农业场景中具备更高的检测精度与鲁棒性。

4.4.2 编码器与解码器的协同优化机制

HRAGNet+ 在编码器与解码器之间建立了紧密的交互关系，使编码阶段得到的多尺度特征能够在解码过程中被充分利用，并在训练过程中不断得到优化。

在前向传播过程中，多尺度特征增强编码器通过跨尺度特征增强与语义对齐，生成语义一致且空间对齐的特征金字塔 $\{E_i\}$ 。这些特征在解码阶段通过 **RoIAlign** 转换为区域特征，并与提案特征共同参与后续建模。在此基础上，**GPEA** 模块利用提案之间的 **IoU** 关系对查询特征进行更新，使不同提案之间能够感知彼此的空间位置关系。由于编码器已经对不同尺度特征完成了对齐，解码阶段的几何关系建模不再依赖额外的对齐过程，从而在更稳定的特征基础上进行。

在任务建模方面，编码器输出的融合特征作为共享表示输入几何-任务协同增强解码器，同时服务于分类与回归任务。针对两类任务对特征需求的差异，**TDFM** 模块对共享特征进行拆分，得到分类特征 R_{cls}^l 与回归特征 R_{reg}^l 。其中，分类分支侧重语义信息，而回归分支更关注目标的空间结构与边界细节。通过这种方式，编码器生成的统一特征可以在解码阶段被分别利用，避免不同任务之间的相互干扰。

在反向传播过程中，分类与回归损失通过解码器逐层传递至编码器，对特征金字塔 $\{E_i\}$ 进行联合优化。编码器不仅提供初始特征，还会根据任务需求不断调整其表示方式。同时，**GPEA** 中引入的几何关系也通过梯度传递影响编码器，使其在空间结构表达上更加一致。

综上所述，多尺度特征增强编码器负责提供融合空间与语义信息的多尺度特征，几何-任务协同增强解码器在此基础上通过几何关系建模与任务解耦，对提案表示进行逐阶段细化，并在训练过程中对编码器进行反馈更新，从而形成前向建模与反向

优化相互耦合的协同机制。

4.5 实验与结果分析

4.5.1 实验设置与评价指标

本章的底层训练策略、数据增强方式、学习率调度以及优化器配置等基础实验设置，均与第三章保持严格一致。所有实验均采用 **ResNet-50** 作为统一的骨干网络，并在 **PlantDoc**（多种植物病害）、**TLD**（番茄叶片病害）以及 **FD**（多类别水果）三个典型的农业视觉数据集上进行训练与验证。评价指标采用 **COCO** 官方评估指标。

4.5.2 与主流检测算法的对比实验

为全面评估 **HRAGNet+** 模型的检测性能，本章选取多种前沿目标检测算法作为对比基线。此外，第三章提出的 **HRAGNet** 亦作为重要对比对象，以直观反映第四章增强解码器带来的性能增益（注：**YOLOv10**^[23] 因其底层评估框架限制，未提供尺度细分指标，故以“-”表示）。

(1) PlantDoc 数据集上的性能比较

表 4.1 不同检测算法在 PlantDoc 数据集上的性能比较

模型	骨干网络	AP	AP ₅₀	AP ₇₅	AP _s	AP _m	AP _l
Faster R-CNN ^[4]	ResNet-50	36.6	57.3	40.5	2.2	19.0	38.6
Mask R-CNN ^[13]	ResNet-50	33.3	52.3	36.3	1.8	22.8	34.4
FCOS ^[22]	ResNet-50	34.5	53.2	36.5	2.6	26.9	34.9
YOLOv10 ^[23]	CSPNet	39.8	56.1	—	—	—	—
Sparse R-CNN ^[65]	ResNet-50	35.9	52.4	39.5	3.3	18.1	37.8
DETR ^[28]	ResNet-50	34.6	51.3	37.3	1.8	22.5	34.2
Deformable-DETR ^[3]	ResNet-50	40.0	56.3	45.1	1.7	25.2	41.8
DINO ^[30]	ResNet-50	40.2	53.0	43.2	2.3	28.5	41.1
HDINO ^[92]	ResNet-50	33.7	32.0	35.5	1.5	27.8	35.1
DDQ ^[93]	ResNet-50	38.0	54.1	43.2	1.8	27.2	40.0
RecursiveDet ^[94]	ResNet-50	42.2	60.4	46.2	3.5	30.3	44.3
AlignDETR ^[31]	ResNet-50	43.4	62.7	47.3	3.4	29.1	45.4
HRAGNet (第三章)	ResNet-50	43.8	64.4	47.8	3.9	29.9	46.2
HRAGNet+ (本章)	ResNet-50	45.7	66.3	49.8	4.3	30.5	47.1

表 4.1 汇总了各模型在 PlantDoc 数据集上的表现。可以看出，传统两阶段检测

器 Faster R-CNN 在 PlantDoc 数据集上的 AP 为 36.6%，Deformable DETR 和 DINO 等近年来提出的 Transformer 检测器通过注意力机制建模全局信息，使检测性能得到一定提升。在此基础上，第三章提出的多尺度特征增强编码器通过多尺度特征递归增强与全局上下文聚合，有效提升了特征表达能力，使模型 AP 提升至 43.8%，在各对比方法中取得较优性能。第四章进一步引入增强解码器结构，通过几何先验引导的注意力机制以及任务特征解耦与协同优化，对提案特征进行精细建模，使模型整体性能提升至 45.7%，相比 HRAGNet 提高了 1.9%。综合来看，本文所提出的编码器增强与解码器优化策略能够在特征表达与目标建模两个层面协同提升检测性能，验证了整体框架设计的有效性。

(2) 番茄叶片病害数据集（TLD）上的性能比较

表 4.2 不同检测算法在 TLD 数据集上的性能比较

模型	骨干网络	AP	AP ₅₀	AP ₇₅	AP _s	AP _m	AP _l
Faster R-CNN ^[4]	ResNet-50	76.3	90.3	84.3	5.6	44.5	78.8
Mask R-CNN ^[13]	ResNet-50	75.2	90.0	83.5	3.3	44.8	78.2
FCOS ^[22]	ResNet-50	75.8	91.5	84.2	7.2	43.9	78.2
YOLOv10 ^[23]	CSPNet	81.5	91.3	—	—	—	—
Sparse R-CNN ^[65]	ResNet-50	78.4	88.9	83.8	9.5	43.8	81.7
DETR ^[28]	ResNet-50	70.2	78.9	75.4	5.5	34.6	75.9
Deformable-DETR ^[3]	ResNet-50	76.9	89.8	83.5	12.1	39.8	79.1
DINO ^[30]	ResNet-50	81.2	90.2	86.1	8.2	48.2	84.1
HDINO ^[92]	ResNet-50	70.5	77.4	74.3	10.2	44.1	73.9
DDQ ^[93]	ResNet-50	81.1	91.4	83.2	9.8	55.2	83.0
RecursiveDet ^[94]	ResNet-50	81.4	91.5	86.9	11.2	44.8	84.1
AlignDETR ^[31]	ResNet-50	80.0	89.3	85.0	11.0	48.1	82.8
HRAGNet (第三章)	ResNet-50	81.9	91.2	86.9	11.5	56.8	84.5
HRAGNet+ (本章)	ResNet-50	83.6	93.4	88.7	12.9	58.3	86.1

TLD 数据集具有病灶密集、类间相似度高、尺度跨度极大的特点，对模型的精细化感知能力提出挑战。表 4.2 汇总了各模型在该数据集上的表现。可以看出，在 TLD 数据集上，传统检测器如 Faster R-CNN 和 Mask R-CNN 的检测性能 AP 分别达到 76.3% 和 75.2%，DINO 与 AlignDETR 等基于 Transformer 结构的检测器如通过引入全局注意力机制能够获得更高检测性能。第三章提出的 HRAGNet 模型通过多尺度特征递归增强与上下文聚合机制，在该数据集上 AP 达到 81.9，在中等尺度目标检

测指标 AP_m 上达到 56.8%，表现出较强的特征表达能力。在此基础上，本章提出的 HRAGNet+ 进一步引入增强解码器，通过几何先验增强注意力模块对提案之间的空间关系进行建模，并通过任务特征解耦与融合模块优化分类与回归任务的特征表示，使模型整体性能进一步提升，AP 达到 83.6%，相比 HRAGNet 提升 1.7%。同时，在小目标检测指标 AP_s 上，HRAGNet+ 达到 12.9%，说明增强解码器结构能够有效提升模型对密集病害区域的检测能力。

(3) 水果检测 (FD) 数据集上的性能比较

FD 数据集涵盖多种具有复杂几何形态与重叠现象的农作物目标，实验结果如表 4.3 所示。实验结果表明，在 FD 数据集上，传统两阶段检测器 Faster R-CNN 的 AP 为 35.8%，而基于 Transformer 架构的检测器（如 DINO 与 AlignDETR）能够将 AP 提升至接近 40%，体现出全局建模能力在复杂场景中的优势。第三章提出的 HRAGNet 模型通过增强编码器特征表达能力，在该数据集上取得 40.6% 的 AP 性能，在小目标检测指标 AP_s 上达到 13.5%，表明该模型能够较好地处理农业场景中的小尺度目标检测问题。在此基础上，HRAGNet+ 通过引入增强解码器进一步优化检测阶段的特征表示，使模型 AP 提升至 42.2%。同时，在 AP_{50} 、 AP_{75} 以及 AP_s 指标上均取得明显提升，说明增强解码器能够有效增强模型对复杂农业目标的定位能力与分类能力。

表 4.3 不同检测算法在 FD 数据集上的性能比较

模型	骨干网络	AP	AP_{50}	AP_{75}	AP_s	AP_m	AP_l
Faster R-CNN ^[4]	ResNet-50	35.8	51.3	37.2	9.8	21.6	47.8
Sparse R-CNN ^[65]	ResNet-50	37.5	52.8	39.6	10.4	23.1	49.8
Deformable-DETR ^[3]	ResNet-50	37.2	57.8	38.8	9.9	23.8	49.3
DINO ^[30]	ResNet-50	38.6	54.5	39.7	12.4	21.9	51.3
HDINO ^[92]	ResNet-50	39.1	55.3	40.6	12.3	22.7	51.7
RecursiveDet ^[94]	ResNet-50	39.3	54.5	41.8	12.0	22.8	52.3
AlignDETR ^[31]	ResNet-50	39.9	55.1	41.9	13.2	23.2	52.4
HRAGNet (第三章)	ResNet-50	40.6	56.9	41.8	13.5	24.1	52.6
HRAGNet+ (本章)	ResNet-50	42.2	58.7	43.4	14.8	25.7	54.1

(4) 可视化分析

HRAGNet (第三章) 和 HRAGNet+ (本章) 在 PlantDoc、TLD 以及 FD 数据集上的检测结果可视化对比如图 4.4 所示。可以看出，在继承 HRAGNet 编码端多尺度特

征增强与融合能力的基础上，HRAGNet+ 在目标定位与类别判别方面表现更优。对于小尺度病斑，HRAGNet+ 能够给出更加准确的边界范围，减少漏检现象；在复杂背景和目標密集场景下，其对相邻病斑或相似目标的区分更加清晰，误检和重叠预测有所减少。

上述改进主要得益于解码阶段的几何先验建模与任务特征解耦机制。几何先验增强注意力模块通过建模候选目标间的空间关系，提高了目标定位的稳定性；任务特征解耦与融合模块则缓解了分类与回归任务之间的特征冲突，使预测结果更加可靠。总体来看，HRAGNet+ 在保持 HRAGNet 多尺度特征增强与融合优势的同时，进一步优化了融合特征在解码阶段的利用方式，从而提升了复杂场景下的检测质量。但在极小病斑、严重遮挡或病害纹理与背景高度相似的情况下，模型仍可能出现漏检或定位偏差，说明复杂田间场景下的小目标病害检测仍有进一步提升空间。

(5) 极端光照条件下的鲁棒性测试

为评估 HRAGNet+ 在复杂光照条件下的鲁棒性，本文基于 PlantDoc 验证集构建极端光照测试集。具体而言，对原始图像施加三种光照扰动（每种 300 张图片）：过曝（亮度增加 +50%）；欠曝（亮度降低 -50%）；随机阴影遮挡，以模拟真实农业场景中的强对比与低可见度情况，共生成 900 张测试图像。在正常光照与极端光照测试集上，对 HRAGNet+ 与六种代表性检测方法进行了对比实验。如表 4.4 所示，HRAGNet+ 在极端光照条件下取得最高检测性能，AP 为 41.6%，且性能下降最小， ΔAP 仅为 -4.1%。这一结果表明，HRAGNet+ 在光照变化场景下具有更强的鲁棒性与稳定性。

表 4.4 极端光照子集下的实验结果

模型	正常光照测试集 AP	极端光照测试集 AP	ΔAP
Faster R-CNN ^[4]	36.6	31.1	-5.5
Sparse R-CNN ^[65]	35.9	30.7	-5.2
Deformable-DETR ^[3]	40.0	35.4	-4.6
DINO ^[30]	40.2	34.6	-5.6
RecursiveDet ^[94]	42.2	37.8	-4.4
AlignDETR ^[31]	43.4	38.8	-4.6
HRAGNet+（本章）	45.7	41.6	-4.1

图 4.5 给出了不同光照扰动下的定性结果。可以观察到，各方法在极端光照下均出现不同程度的性能下降，而 HRAGNet+ 在过曝、欠曝及阴影遮挡三种情况下均保

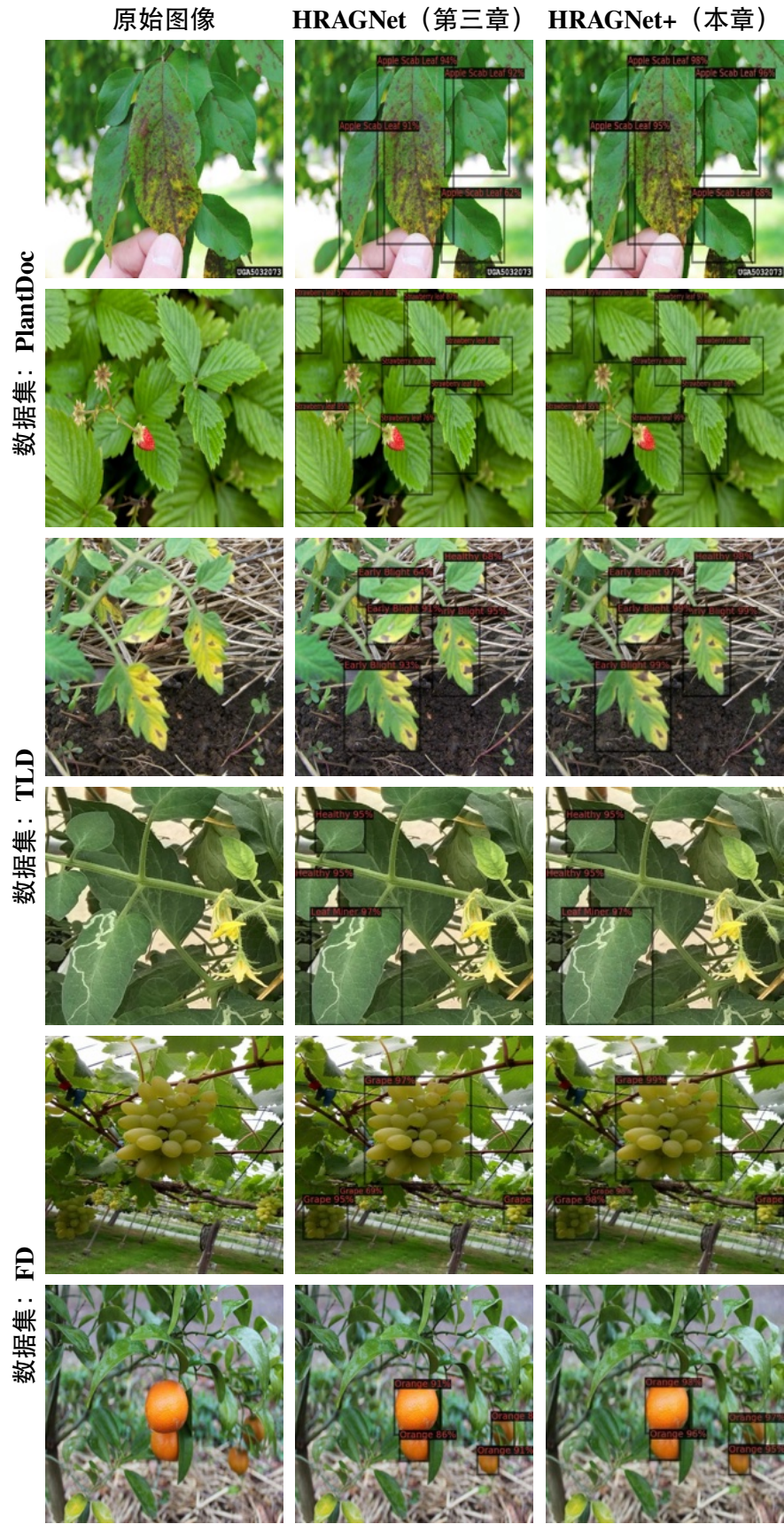


图 4.4 HRAGNet (第三章) 和 HRAGNet+ (本章) 在 PlantDoc、TLD 以及 FD 数据集上的检测结果可视化对比

持较为完整的检测结果，表现出更好的目标定位能力与稳定性。其中，过曝对检测性能影响最为明显，主要由于强光削弱了病斑的纹理与边界信息；相比之下，欠曝与阴影遮挡的影响更具局部性，因此模型仍能通过上下文信息进行一定程度的恢复。综上，实验结果验证了 HRAGNet+ 在复杂光照条件下的鲁棒性优势。

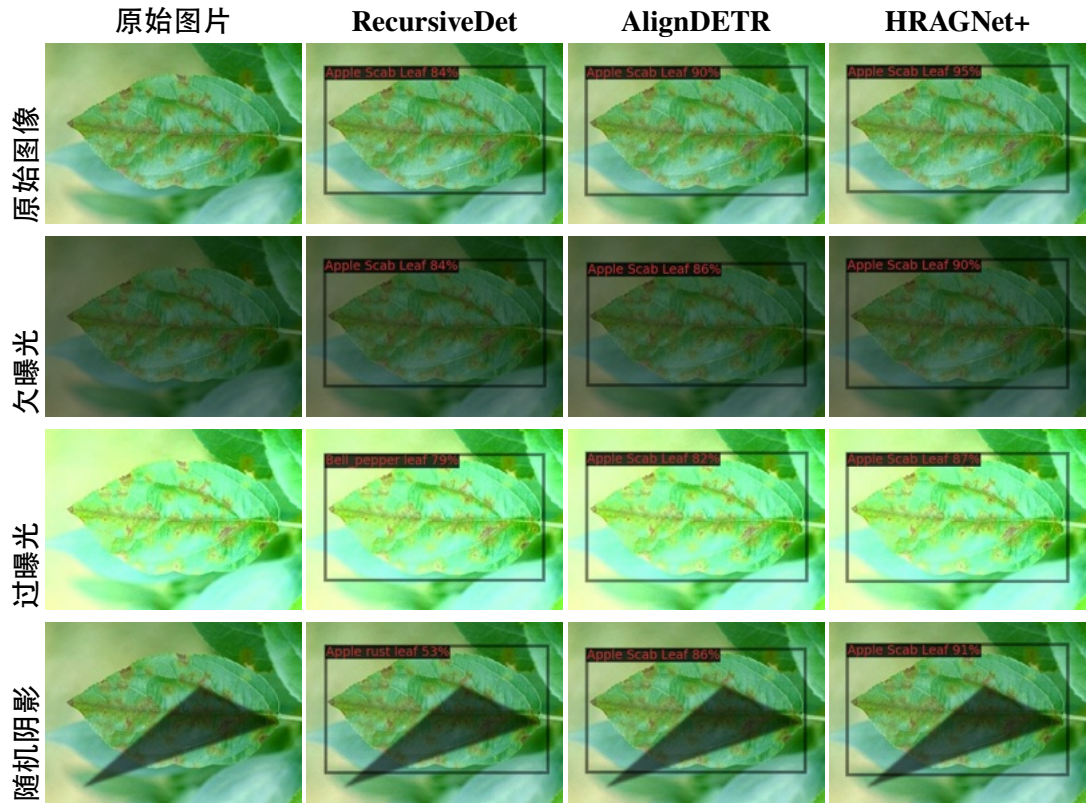


图 4.5 HRAGNet+ 及对比模型在极端光照场景下预测结果的可视化展示

4.5.3 消融实验

为了验证 HRAGNet+ 中各模块的有效性，本章在 HRAGNet 基础上逐步加入不同模块进行消融实验。实验分别评估 GPEA 与 TDFM 对模型性能的影响。

表 4.5 PlantDoc 数据集上的消融实验结果

GPEA	TDFM	AP	AP ₅₀	AP ₇₅
		43.8	64.4	47.8
✓		44.7	65.6	48.9
	✓	45.2	66.1	49.6
✓	✓	45.7	66.3	49.8

从表 4.5 可以看出, 在 HRAGNet 基础上逐步引入增强解码器模块后, 模型检测性能得到持续提升。当仅引入 GPEA 时, 模型 AP 从 43.8% 提升至 44.7%, 提升了 0.9%。这表明通过建模检测提案之间的空间几何关系, 能够有效增强模型对目标空间结构信息的理解能力, 从而提升目标定位精度。当仅引入 TDFM 时, 模型 AP 提升至 45.2%, 相比基线模型提升 1.4%。该结果表明, 通过对分类任务与边界框回归任务进行特征解耦, 可以有效缓解两类任务在共享特征空间中的优化冲突, 从而提升整体检测性能。当同时引入 GPEA 与 TDFM 模块时, 模型最终达到最佳 AP 性能, 这表明两个模块在功能上具有良好的互补性, 其中 GPEA 主要增强空间关系建模能力, 而 TDFM 则优化任务特征表达, 两者协同作用进一步提升了模型的检测性能。

为了进一步验证 HRAGNet+ 在不同农业场景中的稳定性, 本章在 TLD 和 FD 数据集上进行了消融实验, 实验结果如表 4.6 和表 4.7 所示。实验结果表明, 在 TLD 和 FD 数据集上, 单独添加 GPEA 或 TDFM 均能带来稳定的性能提升, 且两者联合使用效果最佳。TLD 数据集上, HRAGNet+ 相比基线提升 1.7%; FD 数据集上同样提升 1.7%。这充分说明增强解码器结构在不同农业数据集上均能够稳定提升检测性能, 具有良好的泛化能力。

表 4.6 TLD 数据集上的消融实验结果

GPEA	TDFM	AP	AP ₅₀	AP ₇₅
		81.9	91.2	86.9
✓		82.6	92.1	87.8
	✓	83.0	92.5	88.1
✓	✓	83.6	93.4	88.7

表 4.7 FD 数据集上的消融实验结果

GPEA	TDFM	AP	AP ₅₀	AP ₇₅
		40.6	56.9	41.8
✓		41.2	57.6	42.3
	✓	41.6	58.0	42.7
✓	✓	42.2	58.7	43.4

4.5.4 各组件有效性分析

为了进一步论证 GPEA 与 TDFM 内部机制设计的科学性，本章在 PlantDoc 数据集上进行了模块的有效性分析。

(1) GPEA 空间建模机制对比

表 4.8 不同空间关系建模方式的性能比较

空间关系建模方式	AP	AP ₅₀	AP ₇₅	AP _s
无空间关系建模 (Baseline)	43.8	64.4	47.8	3.9
传统相对位置编码 (Relative Pos)	44.2	65.0	48.3	4.1
GPEA (绝对定位 + IoU 掩码引导)	44.7	65.6	48.9	4.5

表 4.8 对比了三种不同的空间处理策略。实验可见，若仅采用 Transformer 中常见的标准相对位置编码 (Relative Pos)，AP 仅微弱提升 0.4%。这是由于常规相对编码缺乏对病斑真实物理边界和遮挡情况的感知。而本章设计的 GPEA 将绝对坐标感知与基于 IoU 矩阵的重校准相结合，为注意力机制提供了软门控机制，从而将 AP_s 显著提升至 4.5%。这说明在密集农业场景中，显式的几何形态约束远优于隐式的相对位置偏置。

(2) TDFM 模块中不同任务特征建模方式的实验

表 4.9 不同任务特征建模方式的性能比较

任务特征建模方式	AP	AP ₅₀	AP ₇₅	AP _m
完全共享特征检测头 (Baseline)	43.8	64.4	47.8	26.9
简单双分支分离 (Naive Decoupled Head)	44.5	65.4	48.7	27.6
TDFM (双维度掩码解耦 + 自适应协同)	45.2	66.1	49.6	28.3

表 4.9 对比了不同程度的检测头解耦策略。当前目标检测方法多采用分类与回归分支的简单解耦结构，相比完全共享特征结构，该方式可在一定程度上缓解任务冲突，在实验中带来 0.7% 的性能提升。然而，该方法仍属于静态结构设计，缺乏对不同任务特征需求的自适应建模能力。TDFM 模块通过引入通道与空间双维度特征掩码，实现对 RoI 特征的自适应分解，并进一步通过跨任务信息交互机制实现分类与回归特征的协同优化。在保持任务特异性的同时增强必要的信息共享，从而提升整体检测性能。实验结果表明，TDFM 在 AP、AP₅₀ 与 AP₇₅ 等指标上均取得最优表

现 (45.2% AP)，验证了其在任务解耦与特征建模方面的有效性。

(3) 基于 TDFM 的任务专属特征可视化分析

为定性验证 TDFM 的有效性，对其生成的任务专属特征图进行可视化，如图 4.6 所示。结果表明，分类分支特征更关注具有判别性的语义区域（如颜色与纹理变化），而回归分支特征则更多响应于目标边界及结构细节区域。这种显著的注意区域差异说明 TDFM 能够有效实现分类与回归任务的特征解耦，并缓解二者之间的优化冲突，从而提升检测性能与稳定性。

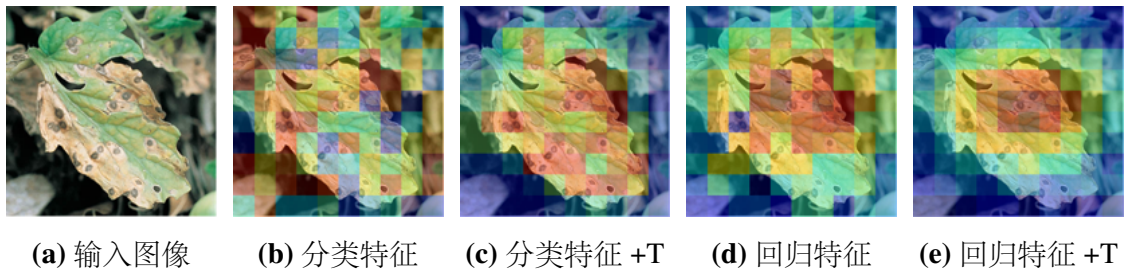


图 4.6 有无 TDFM 模块的特征图可视化

(4) 解码器迭代次数验证

为分析迭代解码阶段数对模型性能的影响，本文对解码器的迭代层数进行了消融实验，结果如表 4.10 所示。实验结果表明，当迭代阶段数由 2 增加至 4 时，模型性能持续提升，说明逐步细化的解码过程有助于目标位置回归与类别预测的优化。当阶段数增加至 6 时，模型取得最佳性能，表明适当的迭代解码能够有效增强特征表达与预测精度。然而，当迭代阶段进一步增加至 8 或 10 时，模型性能出现轻微下降。这可能是由于过多的迭代带来冗余计算，并引入一定的误差累积，从而影响模型泛化能力。综合性能与计算开销，本文最终将解码阶段数设置为 6。

表 4.10 解码器迭代阶段数量对检测性能的影响

解码阶段数	AP	AP ₅₀	AP ₇₅
2	33.5	48.7	36.9
4	39.5	56.7	43.2
6	45.7	66.3	49.8
8	40.8	61.3	45.4
10	38.6	57.4	43.9

4.6 本章小结

本章在第三章提出的 **HRAGNet** 多尺度特征增强编码器基础上, 进一步针对目标检测解码阶段中空间结构建模不足以及分类与回归任务耦合过强的问题, 提出了一种几何先验与任务解耦协同优化的检测框架 **HRAGNet+**。该框架通过构建几何-任务协同增强解码器, 在级联解码过程中引入 **GPEA** 模块与 **TDFM** 模块。其中, **GPEA** 通过显式建模候选框之间的空间关系增强目标结构感知能力, 而 **TDFM** 通过动态生成任务特异性通道权重实现分类与回归特征的有效分离, 从而在多阶段迭代过程中逐步优化目标提案的特征表示与空间位置。

在整体框架层面, **HRAGNet+** 实现了编码器与解码器之间的协同优化: 编码器通过多尺度递归增强与全局上下文融合提供稳定的特征表示, 而解码器则在此基础上进行几何关系建模与任务解耦优化, 从而形成完整的端到端检测流程。实验结果表明, 在 **PlantDoc**、**TLD** 及 **FD** 等多个农业视觉数据集上, **HRAGNet+** 在平均精度及各尺度检测指标上均优于多种主流检测方法; 消融实验进一步验证了 **GPEA** 与 **TDFM** 模块的有效性及其协同增益作用, 说明所提出方法能够在复杂植物病害场景下取得更加稳定和准确的检测效果。

第五章 总结与展望

5.1 本文研究内容总结

本文围绕植物叶片病害目标检测任务中多尺度特征表达不足、跨层语义不一致以及复杂田间场景适应性差等关键问题，系统研究了多尺度特征增强与融合机制，并在此基础上构建了递进式优化检测框架 HRAGNet 及其增强版本 HRAGNet+。针对病斑尺度变化剧烈、小尺度目标占比高以及背景干扰显著等特点，本文从特征建模与检测解码两个层面对目标检测过程进行了系统性改进。

在编码端，围绕多尺度特征表达能力不足的问题，构建了分层递进的特征增强机制。通过引入交互式增强模块，实现不同尺度特征之间的双向信息交互与递归融合，使高层语义能够有效指导低层特征表达；进一步通过跨尺度语义对齐机制，利用可学习的语义流场建立像素级空间对应关系，缓解多尺度特征在融合过程中的空间错位问题；在此基础上，引入全局上下文聚合机制，将不同层级特征进行统一建模，从而增强模型在复杂背景下的整体判别能力。

在解码端，针对传统检测头对目标结构信息利用不足以及分类与回归任务耦合的问题，提出了几何-任务协同增强解码方法。一方面，通过几何先验增强机制显式建模候选框之间的空间关系，对其进行结构约束与特征重加权，提升模型对病斑分布形态的感知能力；另一方面，通过任务驱动特征重组机制，对分类与回归特征进行自适应解耦，并在解耦基础上实现动态融合，从而缓解多任务学习中的冲突问题。同时，结合多阶段迭代优化策略，使候选框在逐层迭代过程中不断逼近真实目标，提高检测结果的稳定性与精度。

总体而言，本文通过编码端多尺度特征增强与解码端几何关系建模及任务协同优化的递进式设计，实现了从特征表达、空间对齐到目标预测的全流程优化，有效提升了模型在复杂田间环境下对多尺度病斑目标的建模能力，为植物叶片病害检测任务提供了一种系统性的改进思路。

5.2 未来研究展望

尽管本文方法在植物叶片病害检测任务中取得了较好的检测精度与鲁棒性，但在实际农业场景部署和更广泛应用中仍存在一定局限。未来研究可从模型轻量化、跨场景泛化、数据高效学习以及真实环境鲁棒性等方面进一步展开。

(1) 模型计算效率与边缘端部署适配

当前 HRAGNet+ 在引入多级特征增强、跨尺度融合以及迭代解码模块后，模型的特征表达能力和检测性能得到提升，但相较基线模型，其计算开销可能有所增加。本文主要关注复杂田间环境下病害检测精度的提升，尚未对模型的参数量、FLOPs、推理速度和显存占用等计算效率指标进行系统统计与分析，因此目前仍难以全面评估其在无人机、移动终端和嵌入式边缘设备上的实时部署可行性。在实际农业应用中，边缘端设备通常对模型复杂度、推理延迟和资源占用具有较高要求。未来工作将进一步补充模型复杂度和推理效率评估，并结合轻量化骨干网络、知识蒸馏、结构化剪枝和模型量化等方法，在尽量保持检测精度的前提下降低模型参数量与计算量，提升 HRAGNet+ 在田间边缘设备上的部署适配能力。

(2) 跨数据集与跨作物泛化能力验证

本文实验主要在 PlantDoc、TLD 及 FD 三个公开数据集上开展，虽然涵盖了多类作物与不同场景，但不同数据源在光照条件、拍摄视角、病害类型、背景复杂度以及标注粒度等方面仍存在明显差异，模型在未知作物类型或新田间环境下可能出现性能波动。未来研究可进一步引入跨数据集测试、跨作物迁移实验和跨区域采集数据验证，系统评估模型在不同作物、不同拍摄设备和不同自然环境下的泛化能力。同时，可结合领域自适应、领域泛化和风格迁移等策略，减少数据分布差异对检测性能的影响。

(3) 数据高效学习方法探索

植物病害图像的高质量标注通常需要农业专家参与，标注成本较高，尤其对于早期微小病斑、稀有病害类别以及复杂遮挡场景，获取充足且准确的标注样本较为困难。本文方法仍主要依赖有监督训练方式，模型性能在一定程度上受限于标注数据规模和类别分布。未来可借鉴半监督学习、自监督表征学习、小样本目标检测和主动学习等方法，充分利用未标注或少标注的农业图像数据，降低模型对大规模人工标注数据的依赖，提升模型在数据受限条件下的学习效率与检测稳定性。

（4）复杂真实田间环境下的鲁棒性提升

真实农业场景中常伴随强光照、阴影遮挡、叶片重叠、运动模糊、背景杂草干扰以及拍摄角度变化等复杂因素，这些因素可能导致病害区域与背景之间的视觉差异进一步减弱，从而影响检测模型的稳定性。虽然本文通过多个数据集验证了方法的有效性，但对于极端天气、不同生长阶段以及连续田间巡检条件下的鲁棒性仍有进一步提升空间。未来可构建更加贴近真实生产环境的病害检测数据集，并结合数据增强、鲁棒特征学习和不确定性建模等方法，提高模型在复杂自然环境下的稳定检测能力。

综上所述，未来围绕模型轻量化部署、跨场景泛化、数据高效学习以及真实环境鲁棒性等方向开展深入研究，将有助于进一步提升植物叶片病害检测技术的实用价值与推广潜力。

插图索引

图 2.1	R-CNN 系列算法对比.....	13
图 2.2	Faster R-CNN 算法的网络结构.....	13
图 2.3	YOLO 算法的网络结构	14
图 2.4	DETR 算法的网络结构	16
图 2.5	Sparse R-CNN 算法的网络结构.....	17
图 2.6	基于传统图像处理的植物叶片病害检测流程图	18
图 2.7	FPN 网络结构示意图	23
图 2.8	PlantDoc 数据集中的叶片图像示例	30
图 2.9	PlantDoc 数据集中不同类别的标注统计	30
图 2.10	TLD 数据集中的叶片图像示例	31
图 2.11	FD 数据集中的图像示例	32
图 3.1	HRAGNet 模型整体框架图	35
图 3.2	多尺度特征增强编码器结构图	36
图 3.3	交互式增强模块 (IEM) 结构图	38
图 3.4	跨尺度语义对齐模块 (CSAM) 结构图	44
图 3.5	全局上下文聚合融合模块 (GCAF) 结构图	48
图 3.6	不同模型在 PlantDoc、TLD 和 FD 数据集上的检测结果可视化对比	55
图 3.7	不同注意力机制的特征图可视化	60
图 4.1	HRAGNet+ 模型整体框架图	64
图 4.2	几何先验增强注意力模块 (GPEA) 结构图	68

图 4.3	任务特征解耦与融合模块 (TDFM) 结构图	73
图 4.4	HRAGNet (第三章) 和 HRAGNet+ (本章) 在 PlantDoc、TLD 以及 FD 数据集上的检测结果可视化对比	82
图 4.5	HRAGNet+ 及对比模型在极端光照场景下预测结果的可视化展示	83
图 4.6	有无 TDFM 模块的特征图可视化	86

表格索引

表 2.1	基于深度学习的农作物叶片病害检测主要步骤	20
表 2.2	两阶段与一阶段目标检测算法的对比	21
表 2.3	不同植物叶片病害检测方法对比	22
表 2.4	TLD 数据集中标注数量统计	31
表 3.1	双路径增强机制的互补特性分析	42
表 3.2	实验设置	50
表 3.3	数据增强策略与训练参数配置	51
表 3.4	在 PlantDoc 数据集上与其他先进算法的比较	52
表 3.5	在 TLD 数据集上与其他先进算法的比较	53
表 3.6	在 FD 数据集上的检测性能比较	54
表 3.7	各组件在 PlantDoc 数据集上的消融实验结果	56
表 3.8	各组件在 TLD 数据集上的消融实验结果	57
表 3.9	各组件在 FD 数据集上的消融实验结果	58
表 3.10	不同注意力模块与 IEM 模块的对比结果	58
表 3.11	IEM 模块中不同特征分组数的消融实验	59
表 3.12	IEM 不同感受野的消融实验	59
表 3.13	不同跨尺度特征对齐策略性能比较	60
表 3.14	不同对齐参考层对 CSAM 模块性能的影响	61
表 3.15	通道注意力消融实验	62
表 4.1	不同检测算法在 PlantDoc 数据集上的性能比较	78

表 4.2	不同检测算法在 TLD 数据集上的性能比较	79
表 4.3	不同检测算法在 FD 数据集上的性能比较	80
表 4.4	极端光照子集下的实验结果	81
表 4.5	PlantDoc 数据集上的消融实验结果	83
表 4.6	TLD 数据集上的消融实验结果	84
表 4.7	FD 数据集上的消融实验结果	84
表 4.8	不同空间关系建模方式的性能比较	85
表 4.9	不同任务特征建模方式的性能比较	85
表 4.10	解码器迭代阶段数量对检测性能的影响	86

参考文献

- [1] Food and Agriculture Organization of the United Nations. Plant health and food security[EB/OL]. 2020. <https://www.fao.org/plant-health>.
- [2] LIN T Y, DOLLÁR P, GIRSHICK R, et al. Feature pyramid networks for object detection[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 2117-2125.
- [3] ZHU X, SU W, LU L, et al. Deformable detr: Deformable transformers for end-to-end object detection[C/OL]//International Conference on Learning Representations. 2021. <https://openreview.net/forum?id=gZ9hCDWe6ke>.
- [4] REN S, HE K, GIRSHICK R, et al. Faster R-CNN: Towards real-time object detection with region proposal networks[C]//Advances in Neural Information Processing Systems (NeurIPS): Vol. 28. 2015.
- [5] ZONG Z, SONG G, LIU Y. Detsr with collaborative hybrid assignments training[A]. 2023.
- [6] ZHAO Z Q, ZHENG P, XU S T, et al. Object detection with deep learning: A review [J]. IEEE Transactions on Neural Networks and Learning Systems, 2019, 30(11): 3212-3232.
- [7] VIOLA P, JONES M. Rapid object detection using a boosted cascade of simple features[C]//Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR): Vol. 1. IEEE, 2001: I-511-I-518.
- [8] DALAL N, TRIGGS B. Histograms of oriented gradients for human detection[C]//Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR): Vol. 1. IEEE, 2005: 886-893.
- [9] FELZENSZWALB P F, MCALLESTER D, RAMANAN D. A discriminatively trained, multiscale, deformable part model[J]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2008: 1-8.

- [10] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2014: 580-587.
- [11] HE K, ZHANG X, REN S, et al. Spatial pyramid pooling in deep convolutional networks for visual recognition[J]. IEEE transactions on pattern analysis and machine intelligence, 2015, 37(9): 1904-1916.
- [12] GIRSHICK R. Fast R-CNN[C]//Proceedings of the IEEE International Conference on Computer Vision (ICCV). 2015: 1440-1448.
- [13] HE K, GKIOXARI G, DOLLÁR P, et al. Mask R-CNN[C]//Proceedings of the IEEE International Conference on Computer Vision (ICCV). 2017: 2961-2969.
- [14] CAI Z, VASCONCELOS N. Cascade R-CNN: Delving into high quality object detection[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2018: 6154-6162.
- [15] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: Unified, real-time object detection[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 779-788.
- [16] REDMON J, FARHADI A. Yolo9000: better, faster, stronger[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 7263-7271.
- [17] REDMON J, FARHADI A. YOLOv3: An incremental improvement[A]. 2018.
- [18] LIU W, ANGUELOV D, ERHAN D, et al. Ssd: Single shot multibox detector[C]//European conference on computer vision. Springer, 2016: 21-37.
- [19] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal loss for dense object detection[C]//Proceedings of the IEEE international conference on computer vision. 2017: 2980-2988.
- [20] LAW H, DENG J. Cornernet: Detecting objects as paired keypoints[C]//Proceedings of the European conference on computer vision (ECCV). 2018: 734-750.
- [21] DUAN K, BAI S, XIE L, et al. Centernet: Keypoint triplets for object detection[C]//Proceedings of the IEEE/CVF international conference on computer vision. 2019: 6569-6578.

- [22] TIAN Z, SHEN C, CHEN H, et al. FCOS: Fully convolutional one-stage object detection[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). 2019: 9627-9636.
- [23] WANG A, CHEN H, LIU L, et al. Yolov10: Real-time end-to-end object detection[J]. Advances in neural information processing systems, 2024, 37: 107984-108011.
- [24] KHANAM R, HUSSAIN M. Yolov11: An overview of the key architectural enhancements[A]. 2024.
- [25] TIAN Y, YE Q, DOERMANN D. Yolov12: Attention-centric real-time object detectors[A]. 2025.
- [26] LEI M, LI S, WU Y, et al. Yolov13: Real-time object detection with hypergraph-enhanced adaptive visual perception[A]. 2025.
- [27] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[J]. Advances in Neural Information Processing Systems (NeurIPS), 2017, 30: 5998-6008.
- [28] CARION N, MASSA F, SYNNAEVE G, et al. End-to-end object detection with transformers[C]//European conference on computer vision. Springer, 2020: 213-229.
- [29] LIU S, LI F, ZHANG H, et al. DAB-DETR: Dynamic anchor boxes are better queries for DETR[C/OL]//International Conference on Learning Representations (ICLR). 2022. <https://openreview.net/forum?id=oMI9PjOb9Jl>.
- [30] ZHANG H, LI F, LIU S, et al. Dino: Detr with improved denoising anchor boxes for end-to-end object detection[C/OL]//International Conference on Learning Representations (ICLR). 2023. <https://openreview.net/forum?id=3mRwyG5one>.
- [31] CAI Z, LIU S, WANG G, et al. Align-detr: Improving detr with simple iou-aware bce loss[A]. 2023.
- [32] ZHAO Y, LV W, XU S, et al. Detsr beat yolos on real-time object detection[A]. 2024.
- [33] LIU S, QI L, QIN H, et al. Path aggregation network for instance segmentation [C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 8759-8768.

- [34] GHIASI G, LIN T, LE Q V. NAS-FPN: Learning scalable feature pyramid architecture for object detection[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach, CA, USA: IEEE, 2019: 7036-7045.
- [35] SCHUSTER R, BATTRAUWY R, WASENMÜLLER O, et al. Resfpn: Residual skip connections in multi-resolution feature pyramid networks for accurate dense pixel matching[C]//2020 25th International Conference on Pattern Recognition (ICPR). IEEE, 2021: 180-187.
- [36] GUO D, WU Z, FENG J, et al. Multi-scale semantic enhancement network for object detection[J]. Scientific Reports, 2023, 13(1): 7178.
- [37] KE J, HE L, HAN B, et al. Profpn: Progressive feature pyramid network with soft proposal assignment for object detection[J]. Knowledge-Based Systems, 2024, 299: 112078.
- [38] YANG X, XUE S, LI L, et al. Pamfpn: Position-aware multi-kernel feature pyramid network with adaptive sparse attention for robust object detection in remote sensing imagery[J]. Remote Sensing, 2025, 17(13): 2213.
- [39] MB M A, HM T K, TSOPZE N, et al. Shipfpn: New feature pyramid network architecture for object detection in high-resolution satellite images: Application to ship detection[J]. Science of Remote Sensing, 2025: 100270.
- [40] GU L, FANG Q, POPOV E, et al. Frequency-aware feature pyramid framework for contextual representation in remote sensing object detection[J]. Astronautics, 2026, 1(1): 5.
- [41] HU J, SHEN L, SUN G. Squeeze-and-excitation networks[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 7132-7141.
- [42] WOO S, PARK J, LEE J Y, et al. Cbam: Convolutional block attention module[C]//Proceedings of the European conference on computer vision (ECCV). 2018: 3-19.
- [43] FU J, LIU J, TIAN H, et al. Dual attention network for scene segmentation[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2019: 3146-3154.

- [44] LI G, FANG Q, ZHA L, et al. Ham: Hybrid attention module in deep convolutional neural networks for image classification[J]. *Pattern Recognition*, 2022, 129: 108785.
- [45] LAU K W, PO L M, REHMAN Y A U. Large separable kernel attention: Rethinking the large kernel attention design in cnn[J]. *Expert Systems with Applications*, 2024, 236: 121352.
- [46] MURUGAN T, BADUSHA N A N M, MUSA N S, et al. Research advances in maize crop disease detection using machine learning and deep learning approaches[J]. *Computers*, 2026, 15(2): 99.
- [47] ARIVAZHAGAN S, SHEBIAH R N, ANANTHI S, et al. Detection of unhealthy region of plant leaves and classification of plant leaf diseases using texture features [J]. *Agricultural Engineering International: CIGR Journal*, 2013, 15(1): 211-217.
- [48] MOKHTAR U, ALI M A, HASSANIEN A E, et al. Tomato leaves diseases detection approach based on support vector machines[C]//2015 11th International Computer Engineering Conference (ICENCO). IEEE, 2015: 246-250.
- [49] PHADIKAR S, SIL J, DAS A K. Rice diseases classification using feature selection and rule generation techniques[J]. *Computers and Electronics in Agriculture*, 2016, 127: 349-359.
- [50] EWIEDA M, MOHAMED HASSAN T, ROUSHDY M I. Deep learning architectures and imaging modalities for plant disease detection: A systematic review[J]. *Journal of Artificial Intelligence in Engineering Practice*, 2026, 3(1): 159-213.
- [51] BARI B S, ISLAM M N, RASHID M, et al. A real-time approach of diagnosing rice leaf disease using deep learning-based faster r-cnn framework[J]. *PeerJ Computer Science*, 2021, 7: e432.
- [52] ZHANG Y, YANG L. Incremental learning for plant disease detection in complex field conditions with two-layer rpn[J]. *Computers and Electronics in Agriculture*, 2024, 217: 108567.
- [53] PALLAPOTHU T, SINGH M, SINHA R, et al. Cotton leaf disease detection using mask rcnn[C]//AIP Conference Proceedings: Vol. 2393. AIP Publishing LLC, 2022: 020114.

- [54] WU Y, WAN S, WU L, et al. Daonet-yolov8: An occlusion-aware dual-attention network for tea leaf pest and disease detection[A/OL]. 2025. arXiv: 2511.23222. <https://arxiv.org/abs/2511.23222>.
- [55] WANG Q, YAN N, QIN Y, et al. Bed-yolo: an enhanced yolov10n-based tomato leaf disease detection algorithm[J]. *Sensors*, 2025, 25(9): 2882.
- [56] NGUYEN P T, HUYNH D C, HO L D, et al. Riceldd-yolo: Optimizing the yolov13 model to improve the efficiency of rice leaf disease detection[J]. *IEEE Access*, 2026, 14: 51282-51293.
- [57] LOU X, HUANG D, FAN J. Object detection for cucumber leaf disease under occlusion and clutter using se-fpn[J]. *Computers and Electronics in Agriculture*, 2022, 198: 107076.
- [58] WANG Y, ZHANG P, TIAN S. Tomato leaf disease detection based on attention mechanism and multi-scale feature fusion[J]. *Frontiers in Plant Science*, 2024, 15: 1382802.
- [59] SAHU K, TIWARI S, SINGH M K, et al. Plant disease detection using a hybrid dilated cnn with attention mechanisms and optimized mask rcnn segmentation[J]. *Scientific Reports*, 2025, 15(1): 42008.
- [60] HARUNA Y, QIN S, CHUKKOL A H A, et al. SaRPFF: A self-attention with register-based pyramid feature fusion module for enhanced rice leaf disease (RLD) detection [J/OL]. *Multimedia Tools and Applications*, 2025, 84(31): 38017-38043. DOI: 10.1007/s11042-025-20696-3.
- [61] CHAI A Y H, LEE S H, TAY F S, et al. Plantaim: A new baseline model integrating global attention and local features for enhanced plant disease identification[J]. *Smart Agricultural Technology*, 2025, 10: 100813.
- [62] FU H, WANG J, CHEN X. Lightweight grape leaf disease detection via mhdi-detr with mobilenetv4[J]. *Computers and Electronics in Agriculture*, 2024, 220: 108765.
- [63] HUO Z, YANG X, SUN X, et al. EBGC-RTDETR: A lightweight model for eggplant disease detection with enhanced RT-DETR architecture[J/OL]. *Smart Agricultural Technology*, 2025, 13: 101761. DOI: 10.1016/j.atech.2025.101761.

- [64] WANG S, LIU M, DONG S, et al. Ertl-rtdetr: A lightweight transformer-based framework for high-accuracy apple disease detection in precision agriculture[J]. Concurrency and Computation: Practice and Experience, 2025, 37(23-24): e70276.
- [65] SUN P, ZHANG R, JIANG Y, et al. Sparse R-CNN: End-to-end object detection with learnable proposals[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2021: 14454-14463.
- [66] KINDALKAR A A. A research review and perspective toward plant leaf disease detection using image processing techniques[J]. International Journal of Image and Graphics, 2026, 26(04): 2650032.
- [67] XUE H, YANG Q, CHEN S. Svm: Support vector machines[M]//The top ten algorithms in data mining. Chapman and Hall/CRC, 2009: 51-74.
- [68] TAN M, PANG R, LE Q V. Efficientdet: Scalable and efficient object detection[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020: 10781-10790.
- [69] HOWARD A G, ZHU M, CHEN B, et al. Mobilenets: Efficient convolutional neural networks for mobile vision applications[A]. 2017.
- [70] HAN K, WANG Y, TIAN Q, et al. Ghostnet: More features from cheap operations[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020: 1580-1589.
- [71] LIU Z, LIN Y, CAO Y, et al. Swin transformer: Hierarchical vision transformer using shifted windows[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). 2021: 10012-10022.
- [72] GUO C, FAN B, GU J, et al. AugFPN: Improving multi-scale feature learning for object detection[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle, WA, USA: IEEE, 2020: 7313-7322.
- [73] LIU S, HUANG D, WANG Y. Learning spatial fusion for single-stage object detection[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle, WA, USA: IEEE, 2020: 3071-3080.

- [74] HUANG J, CHEN H, REN J, et al. A general adaptive dual-level weighting mechanism for remote sensing pansharpening[C]//Proceedings of the Computer Vision and Pattern Recognition Conference. 2025: 7447-7456.
- [75] DAI J, QI H, XIONG Y, et al. Deformable convolutional networks[C]//Proceedings of the IEEE International Conference on Computer Vision (ICCV). 2017: 764-773.
- [76] LI L, YOU H, WANG Z, et al. Aligndet: Bridging the modality and domain gaps for open-vocabulary object detection[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2024: 17034-17044.
- [77] DOSOVITSKIY A, BEYER L, KOLESNIKOV A, et al. An image is worth 16x16 words: Transformers for image recognition at scale[C]//International Conference on Learning Representations (ICLR). 2021.
- [78] WANG Q, WU B, ZHU P, et al. Eca-net: Efficient channel attention for deep convolutional neural networks[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2020: 11531-11539.
- [79] HOU Q, ZHOU D, FENG J. Coordinate attention for efficient mobile network design[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2021: 13713-13722.
- [80] ZHAO H, SHI J, QI X, et al. Pyramid scene parsing network[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2017: 2881-2890.
- [81] CHEN L C, PAPANDREOU G, SCHROFF F, et al. Rethinking atrous convolution for semantic image segmentation[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2017.
- [82] WANG X, GIRSHICK R, GUPTA A, et al. Non-local neural networks[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2018: 7794-7803.
- [83] SINGH D, JAIN N, JAIN P, et al. Plantdoc: A dataset for visual plant disease detection [M]//Proceedings of the 7th ACM IKDD CoDS and 25th COMAD. 2020: 249-253.

- [84] Roboflow. Tomato leaf diseases dataset[EB/OL]. 2023. <https://universe.roboflow.com/dyploma/tomato-leaf-diseases-4xa5i>.
- [85] Lakshay Tyagi. Fruit detection dataset[EB/OL]. 2023. <https://www.kaggle.com/datasets/lakshaytyagi01/fruit-detection>.
- [86] LIN T Y, MAIRE M, BELONGIE S, et al. Microsoft coco: Common objects in context [C]//European conference on computer vision. Springer, 2014: 740-755.
- [87] EVERINGHAM M, VAN GOOL L, WILLIAMS C K, et al. The pascal visual object classes (voc) challenge[J]. International journal of computer vision, 2010, 88(2): 303-338.
- [88] CHOLLET F. Xception: Deep learning with depthwise separable convolutions [C/OL]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017: 1800-1807. DOI: 10.1109/CVPR.2017.195.
- [89] ZHANG X, ZHOU X, LIN M, et al. Shufflenet: An extremely efficient convolutional neural network for mobile devices[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 6848-6856.
- [90] LI X, ZHANG J, YANG Y, et al. Sfnet: Faster and accurate semantic segmentation via semantic flow[J]. International Journal of Computer Vision, 2024, 132(2): 466-489.
- [91] HUANG Z, SHI X, ZHANG C, et al. Flowformer: A transformer architecture for optical flow[C]//European conference on computer vision. Springer, 2022: 668-685.
- [92] JIA D, YUAN Y, HE H, et al. Detrs with hybrid matching[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2023: 19702-19712.
- [93] ZHANG S, WANG X, WANG J, et al. Dense distinct query for end-to-end object detection[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2023: 7329-7338.
- [94] ZHAO J, SUN L, LI Q. Recursivedet: end-to-end region-based recursive object detection[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. 2023: 6307-6316.

攻读硕士学位期间取得的研究成果

一、论文

[1] Fenglei Yang, Weiyi Ma, Qiaochuan Chen, Yan Sun, Yuexing Han*, PDDNet: An end-to-end object detection framework for real-world plant leaf disease diagnosis, Expert Systems with Applications, Volume 312, doi: <https://doi.org/10.1016/j.eswa.2026.131294>. 2026. (中科院一区, 导师一作, 本人二作)

二、软件著作权

[1] 基于深度学习的植物叶病智能检测软件 V1.0, 开发人: 杨风雷, 马唯一, 韩越兴, 登记号: 2025SR2154393, 登记日期: 2025 年 11 月 5 日, 申请人: 上海大学。(导师一作, 本人二作)

致 谢

时光荏苒，三年研究生生活即将画上句点。回首求学之路，有迷茫探索，有潜心钻研，更有无数师长与同窗的陪伴与扶持，心中满是感恩与不舍。

首先，我谨向我的导师杨风雷老师致以最诚挚的谢意。从论文选题、框架搭建，到研究推进、反复修改，杨老师始终以严谨的治学态度、深厚的学术素养悉心指导，为我指明研究方向，耐心解答困惑。当我的研究陷入困境时，杨老师总能一针见血地指出问题所在，又给予我充分的信任和空间去探索解决之道。杨老师不仅教会我科研的方法，更以踏实谦逊、精益求精的治学精神为我树立榜样，这份教诲与关怀，我将永远铭记于心。

同时，我要特别感谢韩越兴老师。研究生三年里，韩老师给予了我极为重要的科研支持与思路启发。无论是研究方法的选择、实验设计的优化，还是理论分析的深化，韩老师都毫无保留地分享经验，提出诸多极具建设性的意见，帮助我突破研究瓶颈、完善研究内容。韩老师开阔的学术视野和务实的科研作风都让我深受触动，也让我在科研道路上走得更加顺利，这份帮助与指引，我由衷感念。

感谢陈侨川老师、孙妍老师在日常科研探讨与学术交流中给予的指导与启发。在我的研究遇到瓶颈时，老师们总能以独特的视角提出宝贵的建议。每一次与老师们的讨论都让我茅塞顿开，老师们渊博的知识、认真负责的教学态度帮助我拓宽思路，让我在专业学习中不断夯实基础、拓宽视野，为论文的顺利完成奠定了坚实基础；老师们的关心和鼓励也让我在紧张的科研之余感受到如沐春风的温暖。

感谢各位好朋友们，求学路上我们相互鼓励、彼此扶持，一同探讨问题、分享心得，在枯燥的科研时光里增添了许多温暖与欢乐。感谢各位师兄师姐在实验操作与论文写作上的倾囊相授，这份珍贵的友谊将成为我人生中宝贵的财富。

最后，我要向默默支持我的家人致以最深的谢意。家人始终是最坚实的后盾，他们的理解、包容与无条件的支持，让我能够心无旁骛地完成学业，勇敢面对求学路上的每一次挑战。研究生阶段的结束，亦是人生新征程的开始。前路漫漫，我将带着所有的感恩与期许，不忘初心，未来的路上，我会带着所有的感恩与收获，继续砥砺前行。