

doi:10.6041/j.issn.1000-1298.2026.21.031

基于 ASP - F - YOLO 的自然环境下疏果期玫瑰香葡萄果粒识别方法

张小花^{1,2} 黄兆轩³ 林桂潮³ 张文峰^{2,4} 刘 芹^{1,2} 钟婷婷¹ 林钊昊⁵
(1. 仲恺农业工程学院自动化学院, 广州 510550; 2. 广东省智能海洋牧场与装备重点实验室, 广州 510550;
3. 仲恺农业工程学院机电工程学院, 广州 510550; 4. 广东省农产品冷链运输与物流工程技术研究中心, 广州 510225;
5. 仲恺农业工程学院人工智能学院, 广州 510225)

摘要: 利用计算机视觉技术实现疏果期果粒自动检测,对于葡萄生长状态智能化监测及疏果期表型分析等研究具有重要意义。然而,自然环境下疏果期玫瑰香葡萄果粒相比于成熟期,具有小尺度、密集分布、遮挡粘连严重及前景对比度低等特点,识别难度较大,针对此问题,本文提出一种基于改进 YOLO v8n 的果粒检测模型 ASP - F - YOLO,以提升复杂场景下疏果期果粒检测的精度。首先,面向密集小目标检测任务对网络架构进行优化,通过引入高分辨率特征层参与特征融合,提升模型对小尺度果粒的感知能力。其次,提出自适应空间提示模块,利用频率动态卷积自适应生成空间提示,增强浅层细粒度特征对复杂纹理和高频细节的表征能力,从而提升目标区域响应并抑制背景干扰。最后,引入自适应超图结构模块,实现跨尺度特征增强和建模果粒与果穗之间的高阶关系,提高模型对密集遮挡果粒的判别能力。试验结果表明,ASP - F - YOLO 模型的 mAP_{0.5} 达到 95.4%,较 YOLO v8n、YOLO 12n、YOLO 13n、RT - DETR、CenterNet、RTMDet 和 RetinaNet 分别提升 4.3、4.6、5.0、2.2、12.8、14.7、29.3 个百分点; mAP_{0.5~0.95} 达到 74.1%,较上述模型分别提升 4.4、5.2、5.4、5.8、8.1、11.0、30.1 个百分点;召回率达到 88.0%,较上述模型分别提升 4.0、5.1、5.6、0.7、19.5、22.1、37.5 个百分点。部署试验表明,该模型在 NVIDIA Jetson AGX Orin 平台上可实现 21.05 f/s 的检测速度,模型内存占用量为 7.2 MB,具备良好的边缘部署能力。本文方法可为复杂自然环境下疏果期密集果粒检测提供有效技术支撑。

关键词: 玫瑰香葡萄; 疏果; YOLO v8; 密集小目标; 自适应空间提示模块; 频率动态卷积

中图分类号: TP391.4; S24 **文献标识码:** A **文章编号:** 1000-1298(2026)21-0336-11

OSID:



Muscat Hamburg Grape Berry Recognition Method Based on ASP - F - YOLO during Berry Thinning Stage under Natural Environments

Zhang Xiaohua^{1,2} Huang Zhaoxuan³ Lin Guichao³ Zhang Wenfeng^{2,4} Liu Qin^{1,2} Zhong Tingting¹ Lin Fanhao⁵
(1. College of Automation, Zhongkai University of Agriculture and Engineering, Guangzhou 510550, China
2. Guangdong Province Key Laboratory of Intelligent Ocean Ranching and Equipment Technology, Guangzhou 510550, China
3. College of Mechanical and Electrical Engineering, Zhongkai University of Agriculture and Engineering, Guangzhou 510550, China
4. Guangdong Engineering Research Center for Agricultural Product Cold Chain Transportation and Logistics, Guangzhou 510225, China
5. College of Artificial Intelligence, Zhongkai University of Agriculture and Engineering, Guangzhou 510225, China)

Abstract: The use of computer vision technology for automatic grape berry detection during the berry thinning stage is of great significance for the intelligent monitoring of grape growth status and phenotypic analysis during berry thinning. However, compared with those at the mature stage, Muscat Hamburg grape berries at the berry thinning stage in natural environments are characterized by small size, dense distribution, severe occlusion and adhesion, and low foreground contrast, making accurate detection challenging. To address these problems, an improved YOLO v8n-based grape berry detection model, termed ASP - F - YOLO, was proposed to improve the detection accuracy of grape berries during the berry thinning stage in complex environments. Firstly, the network architecture was optimized for dense small-object detection by introducing a high-resolution feature layer into the feature fusion process, thereby enhancing the model's ability to perceive small-scale grape berries. Secondly, an adaptive

spatial prompt module was proposed, in which frequency dynamic convolution was employed to adaptively generate spatial prompts. This module enhanced the representation capability of shallow fine-grained features for complex textures and high-frequency details, thereby strengthening responses to target regions while suppressing background interference. Finally, an adaptive hypergraph structure module was introduced to achieve cross-scale feature enhancement and model the high-order relationships between grape berries and clusters, thereby improving the model's discrimination ability for densely occluded grape berries. Experimental results showed that ASP - F - YOLO achieved an mAP_{0.5} of 95.4%, which was 4.3, 4.6, 5.0, 2.2, 12.8, 14.7, and 29.3 percentage points higher than those of YOLO v8n, YOLO 12n, YOLO 13n, RT - DETR, CenterNet, RTMDet, and RetinaNet, respectively. The mAP_{0.5~0.95} reached 74.1%, representing improvements of 4.4, 5.2, 5.4, 5.8, 8.1, 11.0, and 30.1 percentage points over the above models, respectively. The recall reached 88.0%, representing corresponding improvements of 4.0, 5.1, 5.6, 0.7, 19.5, 22.1, and 37.5 percentage points, respectively. Deployment experiments showed that the model achieved a detection speed of 21.05 f/s on the NVIDIA Jetson AGX Orin platform, with a model memory footprint of 7.2 MB, demonstrating favorable edge deployment capability. The proposed method can provide effective technical support for dense grape berry detection during the berry thinning stage in complex natural environments.

Key words: Muscat Hamburg grape; berry thinning; YOLO v8; dense small objects; adaptive spatial prompt module; frequency dynamic convolution

0 引言

葡萄作为中国重要水果之一,既可鲜食,也广泛用于深加工。在果粒发育阶段,若不进行人工调控,过高的挂果量易导致果粒畸形、色泽不佳及糖度不均等问题^[1-2],因此疏花疏果作业尤为关键。通常果粒长至黄豆粒大小时进行疏果,且不同葡萄品种对果粒保留数量的要求存在差异^[3-5]。在实际生产中,疏果合理性不仅取决于果粒保留数量,还与果粒分布均匀性和整体结构特征密切相关。在自然环境下依赖人工经验判断疏果时机,不仅效率低、劳动强度大,而且易受主观因素影响,难以满足规模化、精细化及智能化管理需求。基于此,利用计算机视觉技术实现高精度果粒检测,不仅能够提高果粒信息获取效率,还可为葡萄生长状态监测及疏果期表型分析等研究提供可靠的数据支撑。

随着计算机视觉技术的发展,其在智慧农业领域应用取得了显著进展,其中 YOLO 系列模型因兼具较高检测精度和实时性,逐渐成为农作物和果实检测的主流方法。在葡萄领域的研究也逐步展开,主要集中在葡萄采摘、葡萄病害检测、葡萄整形等方面。

在葡萄采摘方面,已有研究多基于 YOLO 系列模型开展关键点定位与目标检测,如采用改进 YOLO v8 模型实现采摘点检测与果梗定位,并取得较高精度与成功率^[6]。在病害检测方面,相关研究通过引入轻量化网络结构或注意力机制,对葡萄叶片病害识别进行了优化,同样获得了较好的检测性能^[7]。这些研究多针对特征边界清晰、目标尺度较

大的检测对象,其特征表达与判别难度相对较低。在葡萄整形方面,葡萄果粒呈现出小尺度、密集分布及遮挡严重等特点,使得目标检测与分割面临较大挑战。为此,研究者引入特征提取能力较强的模型以提升检测与分割性能^[8-12]。但这些方法多依赖参数量大、计算复杂度高的模型。

此外,自然环境下目标与背景颜色趋同导致的前景对比度不足,也是影响检测性能的重要因素。已有研究方法^[13-14]在果簇或较大目标检测中表现较好,但在果粒级密集小目标场景中仍存在不足,尤其在目标粘连及边界模糊情况下,难以实现稳定且高精度的检测。

综上,疏果期玫瑰香葡萄果粒处于早期发育阶段,表现为单果尺寸小、果粒分布密集,且易发生遮挡与粘连,同时前景对比度较低,导致目标边界模糊、特征易混淆。相较于成熟期或采摘期,该阶段果粒检测面临更为复杂的识别问题,现有目标检测方法在此类场景下的检测精度仍有待提升。为此,本文以 YOLO v8n 为基线模型,针对疏果期葡萄果粒检测任务对网络结构进行优化,引入自适应空间提示机制和跨尺度特征关联建模方法,构建 ASP - F - YOLO 果粒检测模型,以提升模型在复杂自然环境下的检测性能,并验证其部署可行性。在自建疏果期玫瑰香葡萄果粒数据集上开展对比试验和消融试验,以评估所提方法的有效性。

1 材料与方法

1.1 数据采集与制作

本研究中疏果期玫瑰香葡萄果粒数据集收集自

广东省广州市从化区康城葡萄园,使用 iPhone 13 与 iPhone XS Max 手机对整株植株进行多角度拍摄,共获得约 1 050 幅图像,图 1 展示了部分样本图像。为贴近实际应用场景,去除了因模糊导致无法有效标注的图像,而对轻度模糊但仍具备标注可行性的样本予以保留,以保证数据集的多样性并增强模型的泛化能力。最终共保留 920 幅疏果期玫瑰香葡萄图像。在数据标注过程中,采用矩形框对目标进行逐一标注。对于完整可见的果粒,按其实际边缘进行精确框选;对于存在部分遮挡的果粒,统一仅对其清晰可辨识的可见区域进行标注;对于因严重遮挡或图像模糊而导致无法通过肉眼可靠识别的目标,则不纳入标注范围,以避免因主观估计造成的标注误差。该标注策略能够有效保证标注结果的一致性和客观性,从而提高数据集的标注质量和后续模型训练的可靠性。

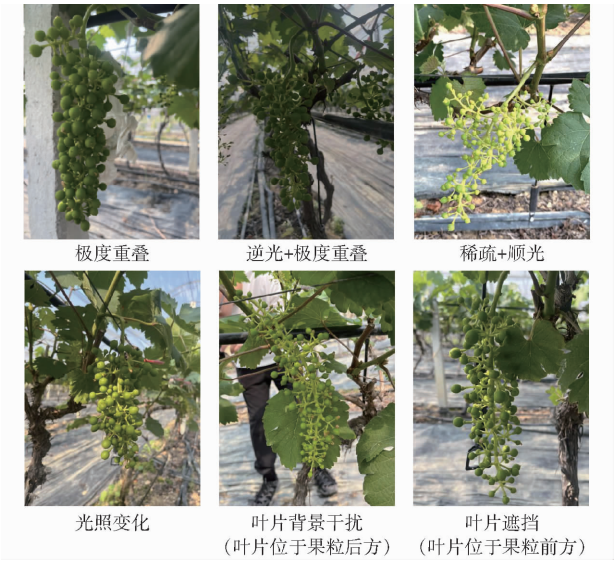


图 1 疏果期玫瑰香葡萄样本数据

Fig. 1 Sample images of Muscat Hamburg grape at berry thinning stage

由于每幅图像中约包含 100 粒葡萄果粒,完全采用人工矩形框标注平均需约 30 min,耗时较长。为提高标注效率,本文首先人工标注 150 幅图像,并基于 YOLO v8n 训练初步模型。随后,将训练得到的权重导入 X - AnyLabeling 进行半自动标注,并对遗漏及不准确目标进行人工修正,使单幅图像的标注时间缩短至约 15 min。

本研究先将原始数据集按比例 8 : 1 : 1 划分为训练集、验证集和测试集,然后分别进行翻转与亮度调整,进行离线数据增强,最终得到 2 760 幅图像。同时在训练过程中采用 Mosaic 在线增强策略,以提升模型在复杂场景中的适应性。

1.2 目标尺度分布统计

为分析疏果期玫瑰香葡萄果粒的目标尺度分布

特征,对已标注数据集中的目标框尺寸进行了统计分析。鉴于模型输入尺寸统一为 640 像素 × 640 像素,本文将标注框尺寸统一换算至图像缩放后的对应尺度进行统计,并参考 COCO 数据集的目标尺度划分标准,将面积小于 32 像素 × 32 像素的目标定义为小目标。其统计结果及目标尺度分布如图 2 所示。结果表明,原始数据集共包含 93 431 个目标,其中小目标 87 642 个,占目标总数的 93.80%。该结果表明,疏果期玫瑰香葡萄果粒检测任务中小目标占比极高,具有显著的小目标密集分布特征。

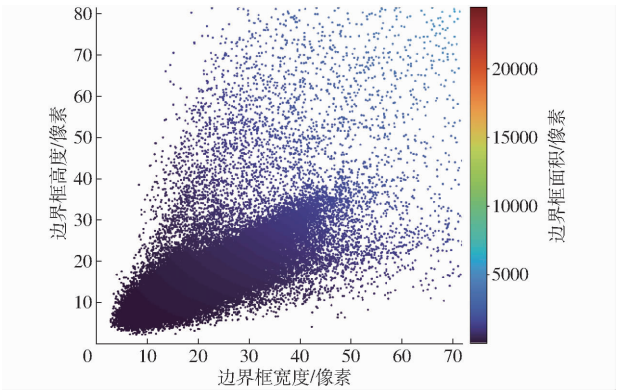


图 2 目标尺度分布图

Fig. 2 Distribution of grape berry target scales during thinning period

2 自然环境下疏果期玫瑰香葡萄果粒识别 ASP - F - YOLO 模型

2.1 改进的 YOLO v8n 模型

YOLO v8n 是 YOLO 系列中最具代表性的模型之一,在检测精度与检测速度之间实现了良好的平衡,且在预试验表现最优,因此本研究基于 YOLO v8n 进行改进。

模型首先重构为适用于密集小目标检测的网络架构,并提出自适应空间提示模块 (Adaptive spatial prompt-frequency, ASP - F)。为增强提示模板对复杂纹理和高频细节的建模能力,ASP - F 中的 Prompt Generator 采用频率动态卷积^[15] (Frequency dynamic convolution, FDConv) 组合模块,并将 ASP - F 模块嵌入至 C2f 结构中形成 C2f_ ASP 模块,以提升复杂场景下密集小目标的特征表达能力。引入 YOLO 13^[16] 所提出的自适应超图结构模块 (Hypergraph-based adaptive correlation enhancement, HyperACE), 接收 3 层不同分辨率的特征图,通过自适应高阶关系建模实现跨尺度特征增强,改进的模型结构如图 3 所示。

2.2 基于小目标的网络架构优化

根据 1.2 节目标尺度分布分析,疏果期玫瑰香

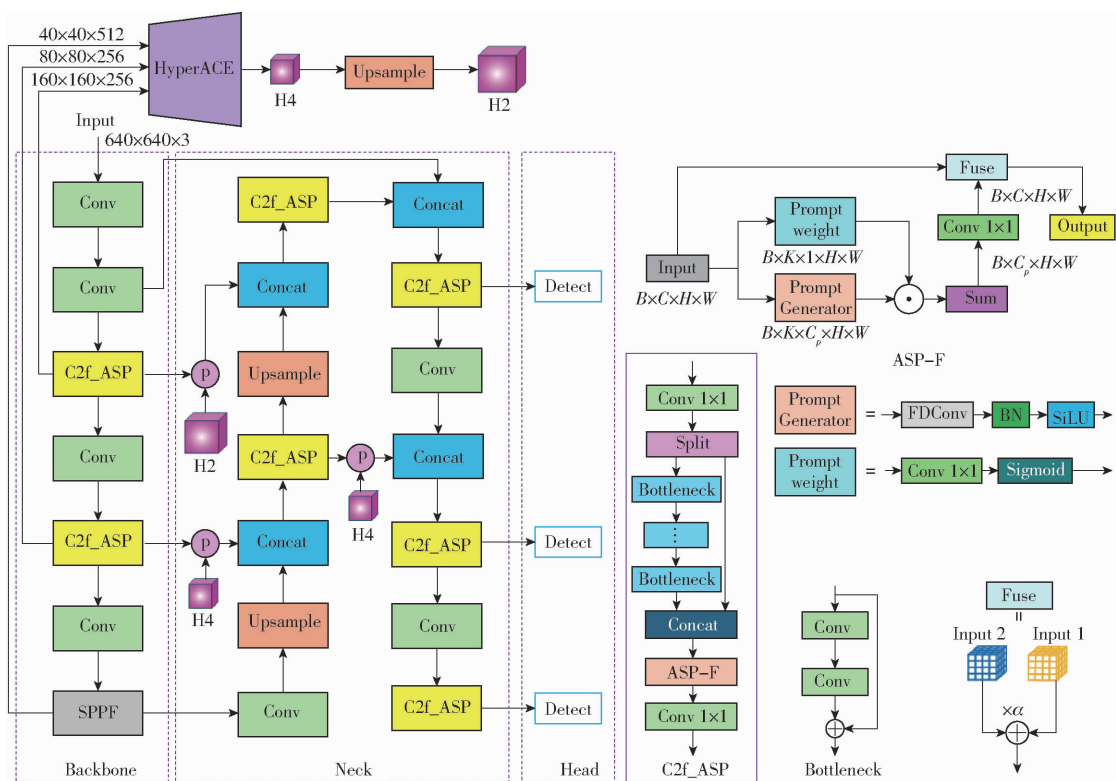


图 3 ASP-F-YOLO 模型整体结构图

Fig. 3 Overall structure diagram of ASP-F-YOLO model

葡萄果粒以小尺度目标为主,小目标占比高达 93.8%,且在图像中呈现密集分布特征。在 YOLO v8n 原始网络架构中,20 像素 $\times$ 20 像素低分辨率特征图主要用于大尺度目标检测,其感受野较大但空间分辨率较低,在多次下采样过程中,小目标的边缘与纹理信息发生丢失,导致定位精度下降。同时,该分支具有较高的通道数与参数规模,但在以小目标为主的检测任务中贡献有限,存在计算资源利用率不高的问题。

为提升密集小目标的检测性能,本文从目标尺度分布特征出发对网络架构进行优化,通过特征尺度调整移除低贡献高参数量的 20 像素 $\times$ 20 像素特征分支,使 160 像素 $\times$ 160 像素高分辨率特征参与特征融合主导检测任务,并通过跨尺度融合将浅层细节与深层语义信息结合,实现特征在高分辨率尺度上的有效流动,从而增强小目标的特征表达能力并降低模型复杂度。进一步地,该结构优化实现了检测尺度与目标尺度分布之间的匹配,使网络感受野与目标尺寸范围更加一致,从而提升模型整体对密集小目标的感知能力。

2.3 ASP-F 模块设计

针对果粒级密集小目标检测中存在的边界模糊、目标特征表达微弱、传统多尺度特征融合与注意力机制对非目标抑制能力有限的问题,本研究结合提示学习机制^[17]和动态提示模块^[18]提出

ASP-F 模块,其结构如图 3 所示。与现有注意力方法主要基于单一特征表达进行加权不同,ASP-F 模块在保留通道表达能力的基础上,引入提示生成机制,自适应地构建多个具有空间分布特性的提示模板及其对应的空间权重生成方式。通过在空间位置上引入可学习的提示信息,使不同区域能够自适应选择融合不同的特征表达,从而实现特征图的细粒度空间调制,增强目标区域响应并有效抑制背景干扰。设输入特征 $F \in \mathbf{R}^{C \times H \times W}$, C 为特征图通道数, W 为特征图的宽度, H 为特征图的高度。ASP-F 模块首先通过 Prompt Generator 动态生成初始特征,并经 Reshape 操作将其重构为空间提示模板,计算式为

$$\tilde{P} = g(F) = \phi(\text{BN}(\text{FDConv}(F))) \quad (1)$$

$$P = \text{Reshape}(\tilde{P}) \quad (2)$$

$$\begin{cases} \tilde{P} \in \mathbf{R}^{C_h \times H \times W} \\ P \in \mathbf{R}^{K \times C_p \times H \times W} \\ C_h = KC_p \end{cases} \quad (3)$$

式中 $g(\cdot)$ ——Prompt Generator 的映射函数

$\text{FDConv}(\cdot)$ ——频率动态卷积

$\text{BN}(\cdot)$ ——批量归一化

$\tilde{P}$ ——Prompt Generator 生成的初始提示特征

$\phi(\cdot)$ ——SiLU 激活函数

$\text{Reshape}(\cdot)$ ——维度重构

C_h ——特征提取后的通道数
 C_p ——每个提示模板的通道数
 P —— K 个空间提示模板

同时, Prompt weight 模块通过 1×1 卷积动态和 Sigmoid 函数生成权重响应, 并经 Reshape 操作重构为 K 个提示模板对应的空间权重图 $W_0 \in \mathbf{R}^{K \times 1 \times H \times W}$, 使多个提示模板能够在同一空间位置同时发挥作用, 从而增强复杂场景下密集小目标的表征能力。在获得空间权重后, 各提示模板与其对应权重进行逐元素加权, 并对所有加权后的提示特征进行融合, 计算式为

$$W_0 = \text{Reshape}(\sigma(\text{Conv}_{1 \times 1}(F))) \quad (4)$$
$$P'_k = W_k \odot P_k \quad (5)$$
$$P_{\text{fusion}} = \sum_{k=1}^K P'_k \quad (6)$$

式中 $\text{Conv}_{1 \times 1}(\cdot)$ ——卷积核为 1×1 的卷积层
 $\sigma(\cdot)$ ——Sigmoid 激活函数
 P_k ——第 k 个原始提示模板
 W_k ——第 k 个提示模板的空间权重图
 $\odot$ ——逐元素乘法
 P'_k ——第 k 个加权后的提示特征
 P_{fusion} ——融合后的提示特征
最后, 通过 1×1 卷积将融合后的提示特征映射

至与原始特征一致的通道数并以残差方式将调整后的提示信息引入主干特征中, 计算式为

$$F_p = \text{Conv}_{1 \times 1}(P_{\text{fusion}}) \quad (7)$$
$$Y = F + \alpha F_p \quad (8)$$

式中 F_p ——调整后的提示特征
 α ——可学习参数
 Y ——ASP - F 模块输出

2.4 频率动态卷积

在上述 ASP - F 模块提示模板生成过程中, Prompt Generator 是关键模块, 其作用是生成具有判别性的空间提示模板。为增强提示模板对复杂纹理和高频细节信息的建模能力, 本文采用 FDConv、批量归一化 (Batch normalization, BN) 以及 SiLU 激活函数构成的结构。其中 FDConv 由 3 部分组成, 分别是傅里叶解耦权重 (Fourier disjoint weight, FDW)、核空间调制 (Kernel spatial modulation, KSM) 和频带调制 (Frequency band modulation, FBM), FDConv 结构如图 4 所示, 图中 iDFT 表示离散傅里叶逆变换; DFT 表示离散傅里叶变换; FC 表示全连接层; Sigmoid 为激活函数; AvgPool 表示平均池化; Conv2D 和 Conv1D 表示卷积; Kernel DFT 表示对卷积核的傅里叶变换; Frequency Decompose 表示对卷积核的频率分解; π_n 表示第 n 个核贡献度。

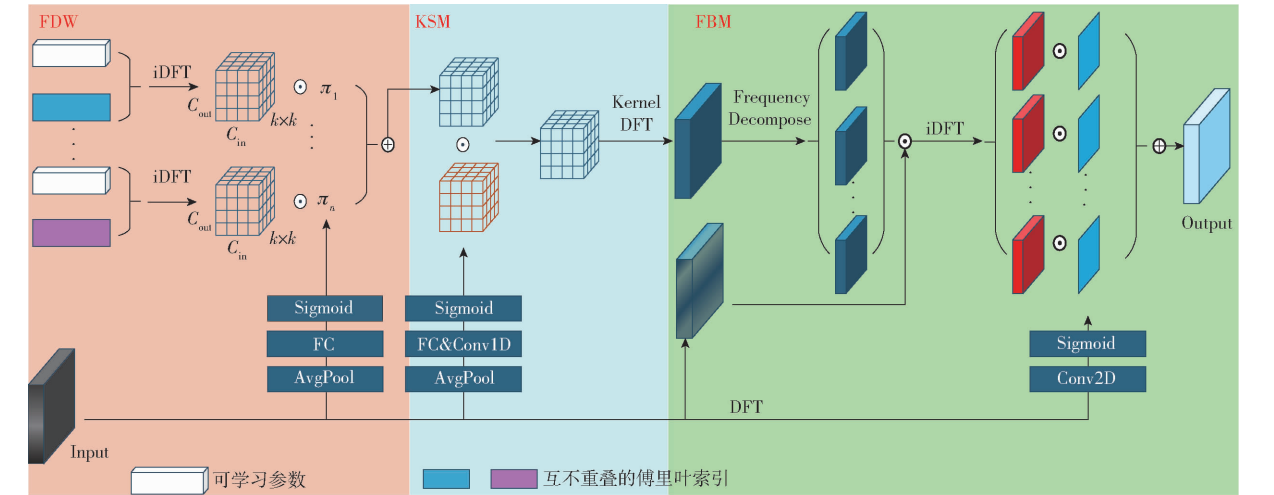


图 4 FDConv 模块结构图
Fig.4 FDConv module architecture

FDW 将卷积核参数通过傅里叶变换映射到频域, 按频率划分不相交的组, 通过逆傅里叶变换回到空间域, 得到频率互补的卷积核。KSM 通过全局通道分支和局部通道分支为卷积核的每一个元素单独调制。FBM 通过在频域中分解卷积核并为各频带生成空间自适应的调制图, FBM 能够针对经过傅里叶变换后的特征图的不同空间位置动态地调整各频率成分的响应, 从而实现空间位置相关的频率选择性调制。

2.5 HyperACE 模块

HyperACE 的设计思想与图神经网络相似, 其核心在于构建特征点之间的关联关系, 实现信息在特征空间中的高效传播与融合。然而, 与依赖人工规则构建图结构的传统方法不同, HyperACE 不显式构建固定图结构, 而是在不同尺度的特征图上自适应学习像素点之间的连接关系, 从而动态建模跨位置、跨尺度特征间的潜在依赖。在上述机制作用下, 在目标粘连及遮挡严重的情况下,

模型不仅依赖局部纹理特征,还能够利用目标间的空间关系与整体结构信息进行辅助判别。设 3 层输入分别为 $B_1 \in \mathbf{R}^{C_1 \times H_1 \times W_1}$ 、 $B_2 \in \mathbf{R}^{C_2 \times H_2 \times W_2}$ 和 $B_3 \in \mathbf{R}^{C_3 \times H_3 \times W_3}$,且满足以下关系

$$\begin{cases} H_1 = 2H_2 = 4H_3 \\ W_1 = 2W_2 = 4W_3 \\ C_1 = C_2 = \frac{C_3}{2} \end{cases} \quad (9)$$

3 个输入通过 HyperACE 模块输出为 $A \in \mathbf{R}^{C_2 \times H_2 \times W_2}$ 的整体映射过程可表示为

$$A = H(B_1, B_2, B_3) \quad (10)$$

式中 $H(\cdot)$ ——HyperACE 映射函数

为实现对原始特征的高阶相关建模与自适应调制,引入门控机制即图 3 中 P 模块,将 HyperACE 输出的相关增强特征注入原始特征 F 中,则门控融合过程可表示为

$$F' = F + \gamma \text{Resize}(A) \quad (11)$$

式中 γ ——可学习参数

$\text{Resize}(\cdot)$ ——特征尺度对齐操作

F' ——融合后的增强特征

γ 作为可学习参数,用于自适应调节相关增强信息的注入强度。本研究所改进的 YOLO v8n 框架中,将主干网络输出的空间分辨率分别为 160 像素 $\times$ 160 像素、80 像素 $\times$ 80 像素和 40 像素 $\times$ 40 像素的 3 层特征图作为 HyperACE 模块输入。

3 试验与结果分析

3.1 试验环境和参数配置

为保证试验环境的一致性,本研究的所有试验均在 Ubuntu 20.04.6 LTS Desktop 64 位操作系统上进行。硬件环境包括 Intel Core i7-14700KF 处理器、NVIDIA GeForce RTX 3090 GPU、32GB 运行内存。深度学习软件环境配置: CUDA 12.4、cuDNN 8.9.7、PyTorch 2.3.1、Python 3.11.13。模型的关键训练参数:输入图像为 640 像素 $\times$ 640 像素、batchsize 为 8、初始学习率为 0.01、epoch 为 300 轮次,每 1 个 epoch 结束之后在验证集上验证一次。

3.2 评价指标

为全面评估模型性能,本文选取以下指标进行评价: Precision(P) 表示预测结果的精确率;召回率表示所有真实目标中被模型正确检测出来的比例; mAP 为各类别平均精度均值,其中 $\text{mAP}_{0.5}$ 、 $\text{mAP}_{0.75}$ 及 $\text{mAP}_{0.5 \sim 0.95}$ 分别表示在 IoU 阈值为 0.5、0.75 及 0.5~0.95 时计算的平均精度均值;参数量表示模型可学习参数数量;浮点运算量表示单次推理的计算量;检测速度表示模型每秒处理的帧数;模型内存占用量表示训练后的权重文件所占的存储空间。

3.3 目标检测模型性能比较

为充分验证改进模型的性能,本研究在测试集上将 ASP-F-YOLO 与多种主流目标检测模型进行了对比,结果如表 1 所示。

表 1 不同模型对比试验结果

Tab. 1 Performance comparison of different models

模型	精确率/ %	$\text{mAP}_{0.5}/$ %	$\text{mAP}_{0.75}/$ %	$\text{mAP}_{0.5 \sim 0.95}/$ %	召回率/ %	浮点运算量	参数量	检测速度/ ($\text{f} \cdot \text{s}^{-1}$)	内存占用量/ MB
YOLO v8n	93.0	91.1	79.9	69.7	84.0	8.1×10^9	3×10^6	107.5	6.3
YOLO 12n	93.1	90.8	76.9	68.9	82.9	6×10^9	2.5×10^6	105.3	5.5
YOLO 13n	93.0	90.4	76.9	68.7	82.4	6.2×10^9	2.4×10^6	103.0	5.4
RT-DETR	93.3	93.2	77.9	68.3	87.3	1.034×10^{11}	3.19×10^7	40.0	66.2
CenterNet	66.0	82.6	77.0	66.0	68.5	1.47×10^{10}	1.42×10^7	9.0	119.9
RTMDet	63.1	80.7	72.6	63.1	65.9	8×10^9	4.8×10^6	39.9	86.8
RetinaNet	44.0	66.1	49.8	44.0	50.5	7.8×10^{10}	1.98×10^7	17.3	162.5
ASP-F-YOLO	94.4	95.4	83.9	74.1	88.0	1.33×10^{10}	1.5×10^6	77.5	7.2

结果表明,ASP-F-YOLO 模型的 $\text{mAP}_{0.5}$ 达到 95.4%,相较于 YOLO v8n、YOLO 12n、YOLO 13n、RT-DETR、CenterNet、RTMDet 和 RetinaNet 分别提升 4.3、4.6、5.0、2.2、12.8、14.7、29.3 个百分点。ASP-F-YOLO 的召回率达到 88.0%,相较于上述模型分别提升 4.0、5.1、5.6、0.7、19.5、22.1、37.5 个百分点。ASP-F-YOLO 的 $\text{mAP}_{0.75}$ 达到 83.9%,相较于上述模型分别提升 4.0、7.0、7.0、6.0、6.9、11.3、34.1 个百分点。ASP-F-YOLO

的 $\text{mAP}_{0.5 \sim 0.95}$ 达到 74.1%,相较于上述模型分别提升 4.4、5.2、5.4、5.8、8.1、11.0、30.1 个百分点。ASP-F-YOLO 模型的参数量,相比上述模型分别降低 50.0%、40.0%、37.5%、95.3%、89.4%、68.8% 和 92.4%。尽管由于引入高分辨率特征层及 ASP-F 模块,模型检测速度由 107.5 f/s 下降至 77.5 f/s,但仍显著高于 RT-DETR、CenterNet、RTMDet 和 RetinaNet,能够满足实际应用中的实时检测需求。

3.4 改进模型消融试验

为验证 ASP - F - YOLO 中各改进策略的有效性,本文以 YOLO v8n 为基线模型,分别引入网络结

构优化、ASP - F 模块及 HyperACE 模块进行消融试验,试验结果如表 2 所示。

结果表明,与原始 YOLO v8n 相比,通过网络架

表 2 消融试验结果

Tab.2 Ablation test results

网络架构优化	ASP - F	HyperACE	精确率/%	mAP _{0.5} /%	mAP _{0.75} /%	mAP _{0.5-0.95} /%	召回率/%	浮点运算量	参数量
			93.0	91.1	79.9	69.7	84.0	8.1 × 10 ⁹	3 × 10 ⁶
√			93.5	93.6	81.1	71.6	85.9	9.9 × 10 ⁹	9 × 10 ⁵
	√		93.6	92.4	79.6	71.1	85.1	1.82 × 10 ¹⁰	6 × 10 ⁶
		√	93.7	92.1	79.4	70.8	84.6	1.16 × 10 ¹⁰	3.7 × 10 ⁶
√	√		93.7	94.5	82.4	72.8	87.6	1.28 × 10 ¹⁰	1.7 × 10 ⁶
√		√	93.4	94.6	82.7	72.9	87.6	1.09 × 10 ¹⁰	9 × 10 ⁵
	√	√	94.0	92.8	80.4	71.8	85.7	2.57 × 10 ¹⁰	7.8 × 10 ⁶
√	√	√	94.4	95.4	83.9	74.1	88.0	1.33 × 10 ¹⁰	1.5 × 10 ⁶

注:“√”表示采用该模块,下同。

构优化后,mAP_{0.5}、mAP_{0.75}及召回率分别提升 2.5、1.2、1.9 个百分点,表明高分辨率特征分支能够有效保留小目标的细粒度特征信息,从而提升密集小目标的感知能力。与此同时,模型参数量由 3 × 10⁶降至 9 × 10⁵。参数量的下降主要来源于低分辨率特征层的移除,由于该分支通常具有更大的通道宽度,其卷积核权重参数占比较高,因此删除后能够显著减少模型体积。对于浮点运算量,其大小不仅与通道数相关,更与特征图的空间尺寸密切相关。高分辨率特征层参与后续特征融合后,模型需要在更大的特征图上进行卷积计算,导致乘加操作次数增加,因此整体浮点运算量由 8.1 × 10⁹上升至 9.9 × 10⁹。在基线模型中单独引入 ASP - F 模块后,mAP_{0.5}、mAP_{0.5-0.95}及召回率分别提升至 92.4%、71.1% 和 85.1%,说明像素级空间提示模板与动态权重调制能够增强目标区域响应并抑制背景干扰,从而提升密集果粒的特征表达能力。然而,由于该模块在各分辨率特征图上进行像素级调制,模型计

算量有所增加。单独引入 HyperACE 模块后,mAP_{0.5}、mAP_{0.5-0.95}及召回率分别提升至 92.1%、70.8% 和 84.6%,通过跨尺度特征之间的高阶相关建模,使浅层纹理信息与深层语义信息得到更加充分的融合。当 3 个模块同时引入时,模型性能达到最佳,mAP_{0.5}、mAP_{0.75}、mAP_{0.5-0.95}及召回率分别达到 95.4%、83.9%、74.1% 和 88.0%。结果表明,各改进策略在密集小目标检测任务中具有良好的协同增益效果。

3.5 低分辨率特征层消融分析

为分析 20 像素 × 20 像素低分辨率特征层及各模块对高占比小目标的疏果期玫瑰香葡萄果粒检测性能的影响,本文在保留 40 像素 × 40 像素、80 像素 × 80 像素和 160 像素 × 160 像素高分辨率特征层特征融合的基础上,开展了以 20 像素 × 20 像素低分辨率特征层是否参与后续特征融合为变量的消融试验,系统评估该分支及不同模块组合对模型检测性能、参数量和计算复杂度的影响,试验结果如表 3 所示。

表 3 低分辨率检测分支对模型性能的影响试验结果

Tab.3 Ablation study on impact of low-resolution detection branch

低分辨率特征层	ASP - F	HyperACE	精确率/%	mAP _{0.5} /%	mAP _{0.75} /%	mAP _{0.5-0.95} /%	召回率/%	浮点运算量	参数量
			93.5	93.6	81.1	71.6	85.9	9.9 × 10 ⁹	9 × 10 ⁵
√			93.4	93.6	80.7	71.6	86.0	1.22 × 10 ¹⁰	3 × 10 ⁶
	√		93.7	94.5	82.4	72.8	87.6	1.28 × 10 ¹⁰	1.7 × 10 ⁶
		√	93.4	94.6	82.7	72.9	87.6	1.09 × 10 ¹⁰	9 × 10 ⁵
	√	√	94.4	95.4	83.9	74.1	88.0	1.33 × 10 ¹⁰	1.5 × 10 ⁶
√	√		93.5	94.0	81.7	72.5	86.8	2.36 × 10 ¹⁰	6.3 × 10 ⁶
√		√	92.9	93.8	81.6	72.1	86.5	1.45 × 10 ¹⁰	3.1 × 10 ⁶
√	√	√	94.4	94.6	82.6	73.2	87.3	2.63 × 10 ¹⁰	6.5 × 10 ⁶

表 3 结果表明,在疏果期玫瑰香葡萄果粒密集小目标检测任务中,在引入相同模块或模块组合的情况下,移除 20 像素 × 20 像素低分辨率特征层模

型的多数评价指标优于保留该分支的模型。当同时引入 ASP - F 与 HyperACE 模块时,无低分辨率特征层的网络 mAP_{0.5}、mAP_{0.75}、mAP_{0.5-0.95}和召回率分别

达到 95.4%、83.9%、74.1% 和 88.0%，而保留低分辨率特征层时分别为 94.6%、82.6%、73.2% 和 87.3%。同时，移除低分辨率分支后，模型参数量由 6.5×10^6 降至 1.5×10^6 ，浮点运算量由 2.63×10^{10} 降至 1.33×10^{10} ，分别下降约 76.9% 和 49.4%。该性能提升主要源于结构优化带来的特征表达效率提升：通过移除低贡献的低分辨率特征分支，并强化高分辨率特征在网络中的流动与融合，使模型能够更加聚焦于小目标区域，从而在减少冗余参数的同时提升检测性能。综上，在本文疏果期玫瑰香葡萄果粒密集小目标检测场景下，低分辨率特征层对小目标的感知能力有限，其较大的感受野容易引入过多背景信息，从而造成小目标特征混淆；同时，该分支较高的通道数也会带来额外的计算开销。因此，在该类场景下移除低分辨率特征层能够在降低模型复杂度的同时提升检测性能。

3.6 可视化分析

3.6.1 改进模型与基准模型热力图响应对比

本研究采用 Grad-CAM^[19] 方法生成特征热力图，对改进模型与基准模型的关注区域进行了可视化分析。热力图中，暖色区域表示模型响应强和对应预测贡献大的区域，冷色区域则对应响应较弱的区域。模型均选取自 3 个检测层前一层的特征进行热力图生成，以全面反映网络在不同尺度下对目标的关注情况，可视化图如图 5 所示。

可视化结果表明，相比于基准模型 YOLO v8n，ASP-F-YOLO 的热力响应主要集中于每个果粒的中心位置，而背景及果梗区域响应较弱，处理重叠目标时，其漏检率更低，体现模型对单个果粒的敏感性、准确性及抗干扰能力。同时，其冷色区域占比相对较少，进一步验证了改进的模型在小目标检测任务中的鲁棒性与可靠性。

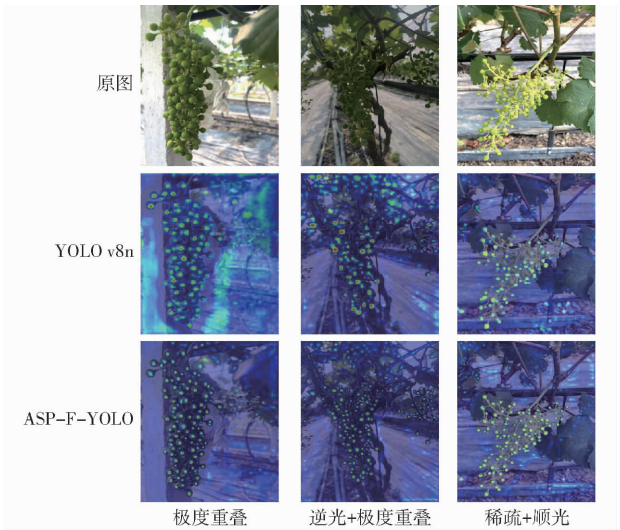


图 5 ASP-F-YOLO 与 YOLO v8n 热力图可视化

Fig. 5 Visualization of heatmaps for ASP-F-YOLO and YOLO v8n

3.6.2 不同注意力机制与 ASP-F 模块性能对比

为检验模块设计的合理性，在保持 ASP-F-YOLO 结构与训练策略一致的条件下，本文依次将 C2f ASP 中的 ASP-F 模块替换为不同注意力机制空间注意力(SA)、通道注意力(CA)、卷积块注意力(CBAM)进行对比，并采用 Grad-CAM 对各模块输出特征进行可视化分析。同时，设计了 ASP-C 作为对照版本，其 Prompt Generator 由卷积(Conv)、BN 和 ReLU 激活函数构成；在保持 ASP-F-YOLO 结构与训练策略一致的条件下，将 ASP-F 模块替换为该对照结构，并分别输出两类可视化结果：一方面对该模块输出特征采用 Grad-CAM 生成特征热力图；另一方面对两者中 Prompt Generator 中 FDConv 和 Conv 输出的特征进行 FFT 频谱分析，以验证 ASP-F 模块引入 FDConv 的设计合理性，试验结果与可视化结果分别如表 4、图 6 及图 7 所示。

表 4 不同注意力机制、ASP-C 模块与 ASP-F 性能对比

Tab.4 Performance comparison of different attention mechanisms, ASP-C module and ASP-F module										
SA	CA	CBAM	ASP-C	ASP-F	精确率/%	mAP _{0.5} /%	mAP _{0.75} /%	mAP _{0.5-0.95} /%	召回率/%	浮点运算量
√					93.5	94.4	82.6	72.9	87.0	1.09×10^{10}
	√				93.5	94.6	82.9	73.1	87.5	1.1×10^{10}
		√			94.2	94.6	82.8	73.3	87.1	1.1×10^{10}
			√		93.3	94.9	82.7	73.1	88.2	1.21×10^{10}
				√	94.4	95.4	83.9	74.1	88.0	1.33×10^{10}

试验表明，相比 SA、CA、CBAM，ASP-F 模块的 mAP_{0.5} 分别提升 1.0、0.8、0.8 个百分点，mAP_{0.75} 提升 1.3、1.0、1.1 个百分点，mAP_{0.5-0.95} 分别提升 1.2、1.0、0.8 个百分点，体现出更强的检测性能与特征表达能力。可视化结果表明，在极度重叠场景下，本文选取不同模型热力图中的相同区域进行对

比，其中严重重叠区域以粉色圆标出。ASP-C 与 ASP-F 在遮挡程度较高的果粒区域均产生较强响应，而 SA、CA 及 CBAM 模块在该区域表现为无响应或响应较弱，同时在周围背景区域存在较明显的非目标响应。验证了 ASP-F 和 ASP-C 模块的结构能够有效增强特征图中目标区域的响应能力，并抑

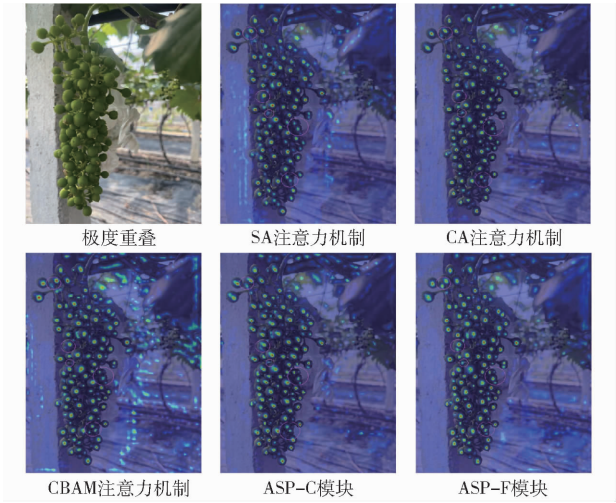


图 6 不同注意力机制、ASP-C 模块与 ASP-F 模块热力图可视化

Fig. 6 Heatmap visualization of different attention mechanisms, ASP-C module and ASP-F module

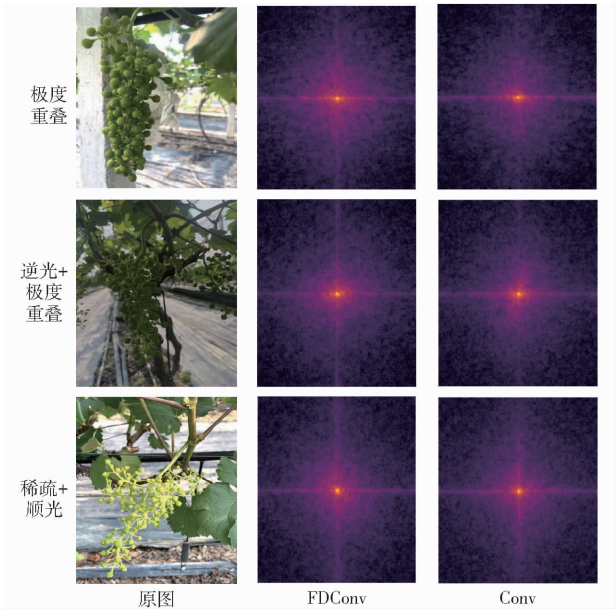


图 7 ASP-F 中 FDConv 的频域响应分析

Fig. 7 Frequency-domain response analysis of FDConv in ASP-F

制无关背景干扰。

进一步对比发现,与 ASP-C 相比,ASP-F 模块在遮挡程度较高目标上的响应范围更广、响应强度更高,相比 ASP-C,ASP-F 的 $mAP_{0.5}$ 仅提升 0.5 个百分点,但更严格的 $mAP_{0.75}$ 与 $mAP_{0.5-0.95}$ 分别提升 1.2、1.0 个百分点,表明 ASP-F 更能提升预测框的定位质量与遮挡目标的判别可靠性。同时,频谱可视化显示,FDConv 作为 Prompt Generator 时相较 Conv 在中心外侧的中高频范围响应更强、分布更丰富,表明 FDConv 能够有效增强与果粒边界及密集纹理相关的高频细节信息。在此基础上,ASP-F 模块使模型能够将频域建模与动态调制机制相结

合,更有效地挖掘遮挡区域中的判别性特征,从而增强复杂场景下目标细节与边界信息的表征能力,提升对密集遮挡目标的特征表征效果。

3.6.3 HyperACE 模块融合前后特征热力图对比

为进一步分析 HyperACE 模块的作用,本文对融合前后的特征热力图进行了对比,可视化结果如图 8 所示。

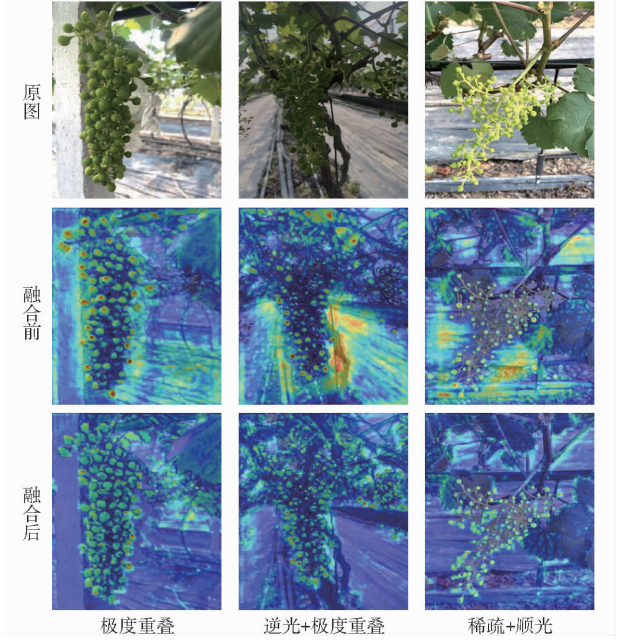


图 8 HyperACE 模块融合前后热力图可视化

Fig. 8 Heatmap visualization of features before and after HyperACE integration

结果表明,模型在模块融合前热力响应主要集中在果穗与背景之间,对果粒呈离散分布,网络更多依赖背景与果穗整体结构的差异进行判别。因此,果穗边缘的果粒响应较强,而处于果穗中间的果粒响应较弱,对果穗整体结构及目标间关系的建模能力有限;融合 HyperACE 输出后,高响应区域从果穗边缘延伸至果穗整体,中间果粒的响应增强,同时背景区域的有效激活明显减弱。这表明 HyperACE 模块不仅能够增强局部细节信息,还可通过高阶相关自适应建模强化果粒与果穗之间的特征关联,从而提高模型对密集遮挡场景中目标的表征与判别能力。

3.6.4 模型改进前后检测效果对比

为验证 ASP-F-YOLO 在自然环境下对疏果期玫瑰香果粒识别效果,选择测试集中图像进行检测并和原始 YOLO v8n 作对比,果粒识别效果和混淆矩阵分别如图 9、10 所示,深蓝色框为模型检测出的目标,红色和粉色圆分别为漏检和误检的目标。

从图 9 可知,在相同场景下相较于 YOLO v8n,ASP-F-YOLO 的漏检与误检更少;图 10 混淆矩阵

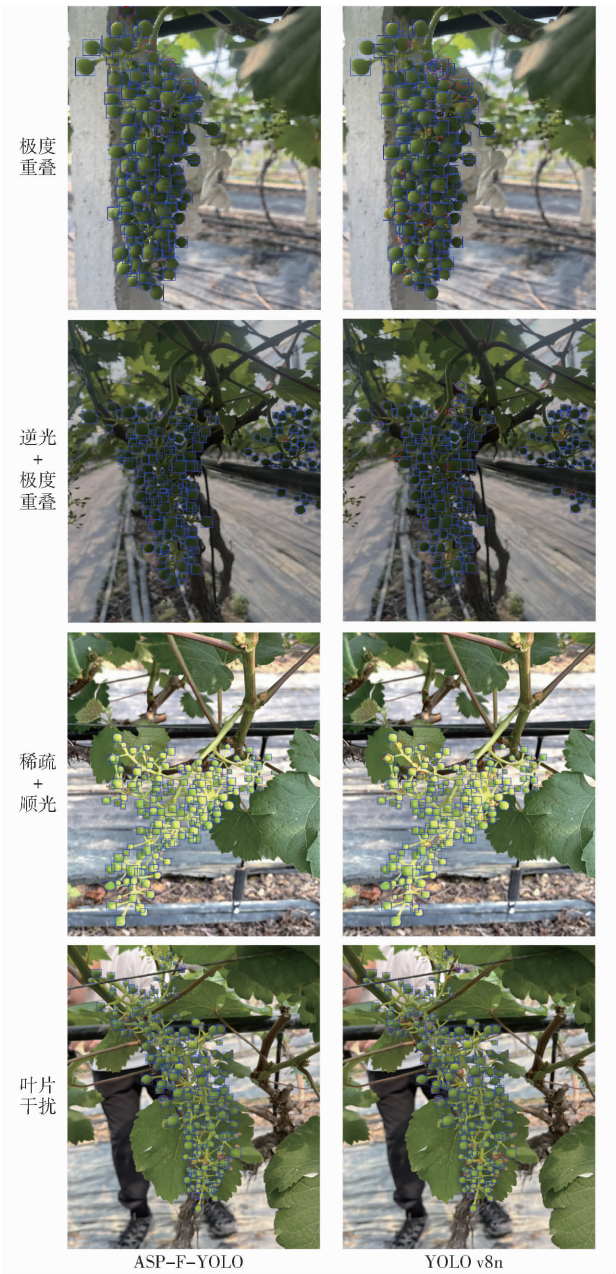


图 9 YOLO v8n 模型改进前后检测效果图

Fig. 9 Detection effect diagram of YOLO v8n model before and after improvement

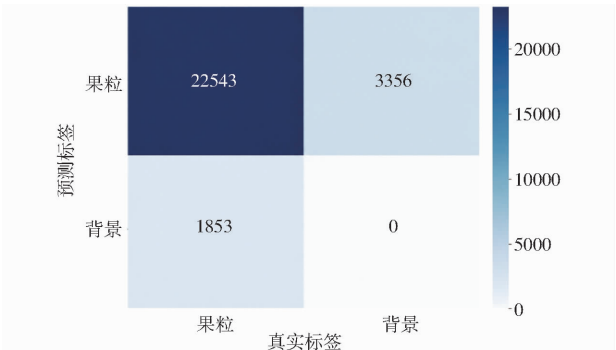


图 10 ASP-F-YOLO 混淆矩阵

Fig. 10 Confusion matrix of ASP-F-YOLO

统计结果可知,模型的误检数为 3 356,高于漏检数的 1 853。综合检测效果图与混淆矩阵可知,主要误

差来源是将背景误判为果粒,误检主要发生在两类情景:①果粒分布密集、相互遮挡严重时,目标边界与纹理信息受到干扰,误检多出现在被果粒遮挡后的果梗区域。②当果穗位于叶片前方时,果粒间重叠形成的局部空隙在绿色叶片背景作用下呈现出与果粒相近的颜色与形状特征,从而引发误检。与此同时,漏检主要出现在极端遮挡、果粒在重度重叠下仅微小可见,导致有效特征不足的情况下,模型更易将其判为背景。尽管在极端重叠情况下仍存在少量漏检及误检,但相比于 YOLO v8n, ASP-F-YOLO 对遮挡目标的识别更加充分,对复杂背景下伪目标的抑制能力更强,整体检测结果更加完整和稳定。这表明所提出的改进方法能够有效增强模型在密集遮挡场景中对小目标的特征提取与判别能力,从而提升复杂自然环境下的检测精度与鲁棒性。

3.7 边缘设备部署

为验证改进算法在实际应用场景中的可行性与工程部署能力,本文在 NVIDIA Jetson AGX Orin 嵌入式平台上对 ASP-F-YOLO 进行推理测试。部署结果表明,该模型可实现 21.05 f/s 的实时推理能力,单幅图像预处理、推理与后处理耗时分别为 0.4、44.6、2.5 ms,同时模型参数量为 1.48×10^6 ,内存占用量为 7.2 MB,计算量为 1.33×10^{10} ,综合体现了其在边缘设备上的部署可行性与实时应用潜力。

4 讨论

尽管本文方法在果粒检测任务中取得了较好的性能,但仍存在一定局限性。后续研究可从以下几个方面进一步改进:①在数据方面,涵盖更多不同品种葡萄疏果期数据集,以提升模型在不同果穗结构与外观特征下的泛化能力。②在模块结构方面,探索 ASP-F 模块与通道提示机制的协同建模,实现空间与通道维度的自适应联合优化,期望进一步提高模型的召回率和精度。③在模型优化层面,引入知识蒸馏方法,以提升轻量级模型性能。

此外,未来研究可进一步结合双目摄像头获取疏果期果粒直径等表型信息及深度距离等三维空间信息,以增强对果粒生长状态的量化表征能力,为后续葡萄生长状态监测及疏果作业时机判断提供依据。同时,结合果粒空间分布特征为人工疏果策略制定提供定量参考,并为后续疏果机器人作业提供数据支持。

5 结论

(1) 针对疏果期玫瑰香葡萄在自然环境下存在

果粒尺寸小、分布密集、前景对比度不足及遮挡严重等检测难点,本文提出了一种面向密集小目标检测的改进模型 ASP-F-YOLO。该模型在 YOLO v8n 基础上,针对小目标检测任务进行网络结构优化,以提升模型对小尺度目标的感知能力;同时,引入融合提示学习机制、动态提示模块及频域信息的 ASP-F 模块,通过自适应生成提示模板以增强目标区域响应并抑制非目标区域干扰,强化小目标的特征表达;

结合 HyperACE 高阶相关建模方法,建模果粒与果穗整体之间的高阶关系。试验结果表明,ASP-F-YOLO 的 $mAP_{0.5}$ 、 $mAP_{0.75}$ 、 $mAP_{0.5-0.95}$ 、召回率分别较基线模型 YOLO v8n 提升 4.3、4.0、4.4、4.0 个百分点。

(2)在 NVIDIA Jetson AGX Orin 平台上的测试结果表明,ASP-F-YOLO 可实现 21.05 f/s 的实时推理速度,具备实时部署能力。

参 考 文 献

[1] Preszler T, Schmit T M, Vanden Heuvel J E. A model to establish economically sustainable cluster-thinning practices[J]. American Journal of Enology and Viticulture, 2010, 61(1): 140-146.

[2] Han W, Han N, He X, et al. Berry thinning to reduce bunch compactness improves fruit quality of Cabernet Sauvignon (*Vitis vinifera* L.)[J]. Scientia Horticulturae, 2019, 246: 589-596.

[3] 郑碧霞,李长林,戢小梅,等.不同疏果处理对阳光玫瑰葡萄产量及果实品质的影响[J].湖北农业科学,2025,64(1): 104-108.

Zheng Bixia, Li Changlin, Ji Xiaomei, et al. Effect of different fruit thinning treatments on yield and quality of Shine Muscat grape[J]. Hubei Agricultural Sciences,2025,64(1): 104-108. (in Chinese)

[4] Roberto S R, Mashima C H, Colombo R C, et al. Berry-cluster thinning to reduce compactness of ‘Black Star’ table grapes[J]. Ciência Rural, 2017, 47: e20160661.

[5] Piernas J, Giménez M J, Noguera-Artiaga L, et al. Influence of bunch compactness and berry thinning methods on wine grape quality and sensory attributes of wine in *Vitis vinifera* L. cv. ‘Monastrell’[J]. Agronomy, 2022, 12(3): 680.

[6] Wu Z, Xia F, Zhou S, et al. A method for identifying grape stems using keypoints[J]. Computers and Electronics in Agriculture, 2023, 209: 107825.

[7] 张惠莉,代晨龙,任景龙,等.基于 YOLO v8-GSGF 模型的葡萄病害识别方法研究[J].农业机械学报,2024,55(11): 75-83.

Zhang Huili, Dai Chenlong, Ren Jinglong, et al. Grape disease identification method based on YOLO v8-GSGF model[J]. Transactions of the Chinese Society for Agricultural Machinery, 2024, 55(11):75-83. (in Chinese)

[8] Chen Y, Li X, Jia M, et al. Instance segmentation and number counting of grape berry images based on deep learning[J]. Applied Sciences, 2023, 13(11): 6751.

[9] Du W, Liu P. Instance segmentation and berry counting of table grape before thinning based on AS-SwinT[J]. Plant Phenomics, 2023, 5: 0085.

[10] Du W, Cui X, Zhu Y, et al. Detection of table grape berries need to be removed before thinning based on deep learning[J]. Computers and Electronics in Agriculture, 2025, 231: 110043.

[11] Buayai P, Saikaew K R, Mao X. End-to-end automatic berry counting for table grape thinning[J]. IEEE Access, 2020, 9: 4829-4842.

[12] Buayai P, Yok-In K, Inoue D, et al. Supporting table grape berry thinning with deep neural network and augmented reality technologies[J]. Computers and Electronics in Agriculture, 2023, 213: 108194.

[13] Liu Q, Lv J, Zhang C. MAE-YOLOv8-based small object detection of green crisp plum in real complex orchard environments [J]. Computers and Electronics in Agriculture, 2024, 226: 109458.

[14] 张传栋,高鹏,亓璐,等.基于 SAW-YOLO v8n 的葡萄幼果轻量化检测方法[J].农业机械学报,2024,55(10): 286-294.

Zhang Chuandong, Gao Peng, Qi Lu, et al. Lightweight detection method for young grape cluster fruits based on SAW-YOLO v8n[J]. Transactions of the Chinese Society for Agricultural Machinery, 2024, 55(10): 286-294. (in Chinese)

[15] Chen L, Gu L, Li L, et al. Frequency dynamic convolution for dense image prediction[C]//Proceedings of the Computer Vision and Pattern Recognition Conference, 2025: 30178-30188.

[16] Lei M, Li S, Wu Y, et al. YOLOv13: real-time object detection with hypergraph-enhanced adaptive visual perception[PP/OL]. V2. arXiv (2025-09-05)[2026-08-17]. <https://doi.org/10.48550/arXiv.2506.17733>.

[17] Potlapalli V, Zamir S W, Khan S H, et al. Promptir: prompting for all-in-one image restoration[J]. Advances in Neural Information Processing Systems, 2023, 36: 71275-71293.

[18] Zheng J, Fu Y, Lu J, et al. RetinalCoNet: underwater fish segmentation network based on bionic retina dual-channel and multi-module cooperation[J]. Fishes, 2025, 10(9): 424.

[19] Selvaraju R R, Cogswell M, Das A, et al. Grad-cam: visual explanations from deep networks via gradient-based localization [C]//Proceedings of the IEEE International Conference on Computer Vision, 2017: 618-626.