

尹书宇, 阮宁君. 基于 YOLO11n 的轻量级草莓成熟度检测方法[J]. 江苏农业学报, 2025, 41(10): 1997-2008.  
doi: 10.3969/j.issn.1000-4440.2025.10.013

## 基于 YOLO11n 的轻量级草莓成熟度检测方法

尹书宇, 阮宁君

(长江大学电子信息与电气工程学院, 湖北 荆州 434020)

**摘要:** 在田间环境下,为解决草莓采摘机器人设备计算资源受限、草莓果实目标小以及叶片遮挡和草莓果实重叠导致难以精确检测的问题,提出了一种基于 YOLO11n 的改进模型,对未成熟期、转变期、成熟期的草莓进行检测。首先,利用 MobileNetV4 替换原模型的主干网络来降低参数量和计算量。其次,提出了一种新的特征融合方法 Bi-Freq,替代原模型中颈部网络的特征融合方法,提升了特征表示能力和鲁棒性。最后,在目标检测的输出层添加 SEAM 注意力机制来提高模型对空间维度和通道的处理能力。改进后的模型(YOLO11n-MFBS)参数量为 1.713 M,浮点计算量为 4.7 G,相比于原始模型 YOLO11n 参数量和浮点计算量分别减少 33.9%和 26.6%。与其他主流检测模型相比,YOLO11n-MFBS 模型在轻量化和检测精度上的综合表现更好。

**关键词:** 目标检测; YOLO11n; 轻量化; 特征融合; 注意力机制; 草莓

**中图分类号:** S126;S668.4 **文献标识码:** A **文章编号:** 1000-4440(2025)10-1997-12

## A lightweight strawberry ripeness detection method based on YOLO11n

YIN Shuyu, RUAN Ningjun

(School of Electronic Information and Electrical Engineering, Yangtze University, Jingzhou 434020, China)

**Abstract:** In field environments, to address the challenges of limited computing resources in strawberry-picking robots, small target size, as well as occlusions and overlaps caused by leaves and fruits that hinder accurate detection, an improved model based on YOLO11n was proposed to detect strawberries in the immature, turning, and mature stages. Firstly, MobileNetV4 was used to replace the original backbone network to reduce the number of parameters and computational cost. Secondly, a new feature fusion method, Bi-Freq, was proposed to replace the neck network's original feature fusion strategy, enhancing feature representation and robustness. Finally, the SEAM attention mechanism was added to the detection head to improve the model's capability in processing spatial and channel information. The improved model (YOLO11n-MFBS) achieved 1.713 M parameters and 4.7 G FLOPs, reducing parameters and floating-point computation by 33.9% and 26.6% respectively compared to the original YOLO11n. Compared with other mainstream detection models, YOLO11n-MFBS demonstrates superior performance in both lightweight design and detection accuracy.

**Key words:** object detection; YOLO11n; lightweight design; feature fusion; attention mechanism; strawberry

收稿日期: 2025-04-25

基金项目: 国家自然科学基金青年项目(52205587); 湖北省自然科学基金青年项目(2022CFB734)

作者简介: 尹书宇(2001-), 男, 湖北襄阳人, 硕士研究生, 研究方向为深度学习、图像处理、人工智能等。(E-mail) ysy04292001@163.com

通讯作者: 阮宁君, (Tel) 18607219955; (E-mail) ruanj@yangtzeu.edu.cn

草莓果实色泽诱人,营养价值高,向来深受大众喜爱,调查结果显示,中国是目前世界上最大的草莓生产国,草莓产量约占全球的 1/3<sup>[1]</sup>。随着草莓产量的不断提高,传统人工采摘方式的效率和成本已经无法满足现在的需求,因此研发草莓采摘机器人已经逐渐成为未来发展的趋势,而在田间环境下快速准确地识别出草莓的成熟度和确定草莓的位置是草莓采摘

机器人视觉系统研发过程中需要解决的关键问题。

近年来,针对果实成熟度的识别与草莓的定位问题,传统的机器学习检测方法通常先对果实样本进行人工提取物理特征,然后利用传统机器学习算法进行模式识别。Zhang 等<sup>[2]</sup>对草莓的高光谱图像进行光谱数据和纹理特征提取,并结合支持向量机对草莓成熟度进行分类预测。Shao 等<sup>[3-5]</sup>利用高光谱成像技术对田间和实验室 3 个不同成熟阶段的草莓进行图像采集,通过支持向量机结合竞争性自适应重加权采样方法评估草莓成熟度。Zhang 等<sup>[6]</sup>利用 HSV(H 指色相,S 指饱和度,V 指明度)颜色空间特性结合绿色检测、最大类间方差法(OTSU)算法<sup>[7]</sup>及支持向量机方法<sup>[8]</sup>,实现了稻瘟病慢型、急型和白型的分类识别。随着深度学习的快速发展,对于水果目标检测方法的研究取得了显著进展。Jing 等<sup>[9]</sup>提出了一种基于 YOLOv8s<sup>[10]</sup>结合倒置残差移动块(iRMB)和渐近特征金字塔网络(AF-PN)<sup>[11]</sup>的桃果轻量级检测模型,平均精度为 94.40%,模型大小为 10.9 MB,推理时间为 6.3 ms,实现了模型的轻量化。Vasquez 等<sup>[12]</sup>通过对比不同卷积神经网络在水果检测和计数问题上的表现并进行综合评估,研究了以 MobileNet<sup>[13]</sup>为基础的 SSD<sup>[14]</sup>目标检测算法,结果表明,设计的网络模型在检测和水果汁数方面的平均精度达到了 90%,花费时间最短。方国文等<sup>[15]</sup>利用 ContextGuide 模块替换 YOLOv8n 模型主干网络中的 Bottleneck 模块,并引入加权双向特征金字塔网络(BiFPN)<sup>[16]</sup>模块替代 PANet 模块,同时采用 MPDIoU 损失函数代替 CIoU 损失函数,实现了对巴旦木果实的轻量化识

别,改进后模型参数量减少 49.3%,检测精度提高 0.5 个百分点。

在实际的田间水果检测任务中,光照变化、环境干扰等因素通常会导致识别精度下降。为提高模型性能,基于卷积神经网络的深度学习模型逐渐复杂化,模型参数、计算量以及内存需求逐渐增大。然而移动设备的计算资源和存储空间有限,因此,如何设计轻量化的目标检测模型以平衡检测精度和效率,已成为该领域研究的关键方向。基于以上问题,本研究拟在 YOLO11n<sup>[17]</sup>模型的基础上利用 MobileNetV4<sup>[18]</sup>替换原模型的主干网络,融合 FreqFusion<sup>[19]</sup>和 BiFPN 2 种特征融合方法替换原模型中的 PANet 模块,在目标检测层引入 SEAM<sup>[20]</sup>注意力机制,构建轻量化模型 YOLO11n-MFBS,为田间草莓自动化采摘机器人的视觉系统研究提供一种新的方法。

## 1 材料与方法

### 1.1 数据集的获取

本次研究选取草莓品种情珠为研究对象。数据集包含 3 100 张涵盖未成熟期、转变期、成熟期 3 个时期的田间草莓图片。图片格式为 jpg,图片分辨率为 1 008×756。采用 LabelImg 数据集标注工具标注草莓处于哪种成熟度,并绘制相应的标注框。同时为了防止模型过拟合以及提升模型鲁棒性和数据样本的多样性,采用旋转、曝光度、模糊、随机遮挡、噪声 5 种增强方式对原始数据集进行数据增强,增强后共得到 15 500 张图片。然后将增强后的数据集按照 8 : 1 : 1 随机划分为训练集、验证集和测试集。增强后的部分数据集图像如图 1 所示。

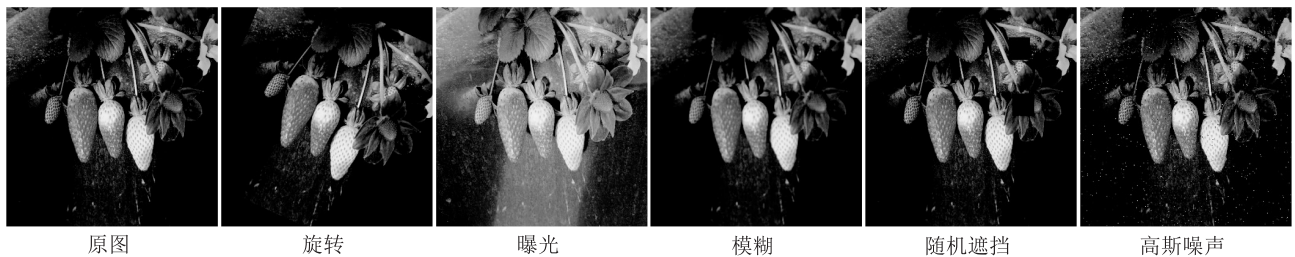


图 1 数据集图像示例及其增强后的图像示例

Fig.1 Dataset image examples and examples of five data augmentation methods

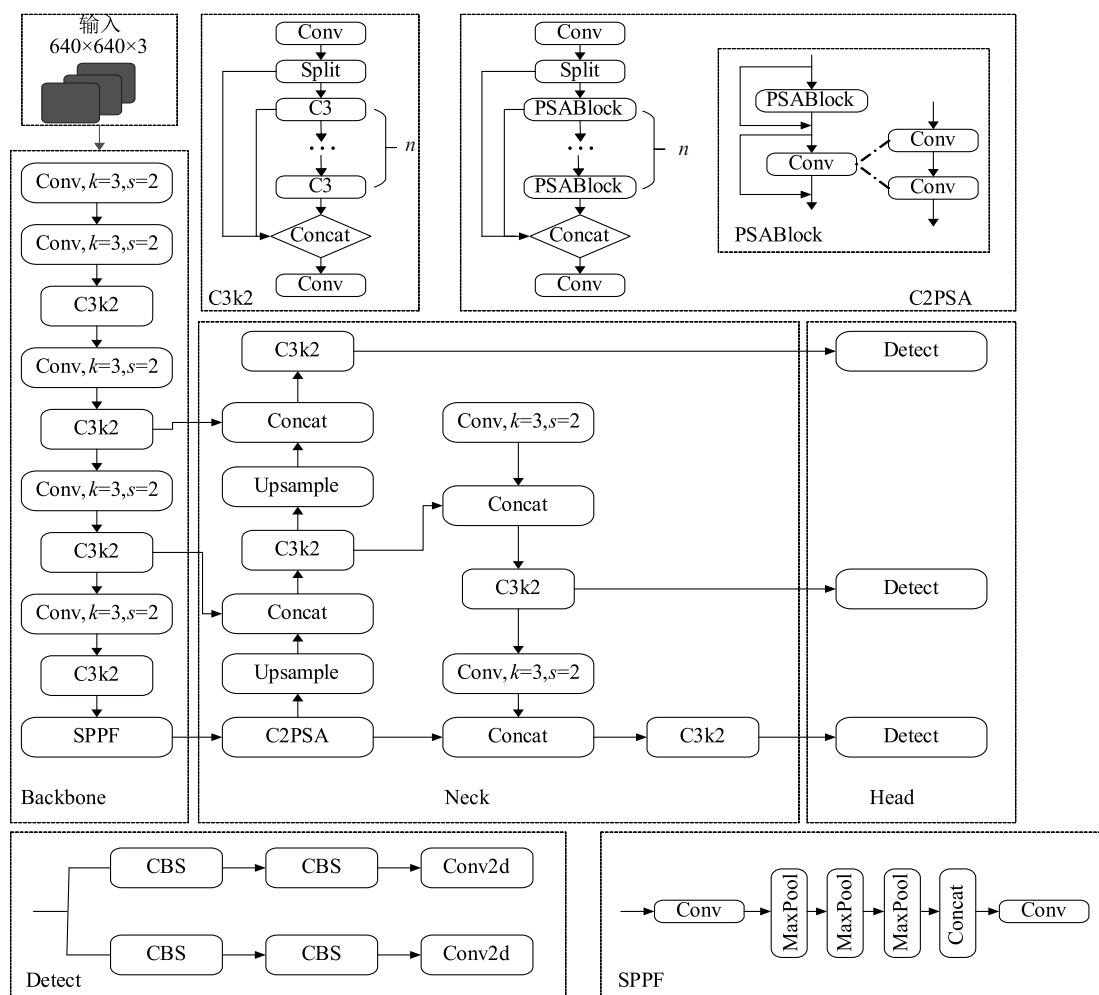
### 1.2 YOLO11n 模型及改进

YOLO11n 模型是 Ultralytics 公司推出的 YOLO 系列的最新版本,结合了尖端的准确性、速度和效

率,被用于目标检测、分割、分类、定向边界框和姿态估计。如图 2 所示,YOLO11n 模型采用 C3k2 模块替换 YOLOv8 系列模型中的 C2f 模块,在提升感受

野和特征表达能力的同时,进一步增强了梯度流信息的丰富性。C3k2 模块通过动态卷积、自适应权重调整和分组卷积的设计,有效捕获全局上下文信息,并优化了多尺度特征融合能力。此外,模型继续沿用无锚框思想,通过中心点、关键点与边界信息的组合表示物体,显著提升了复杂场景下的检测精度与

鲁棒性。YOLO11 模型在 SPPF 模块后面添加了 1 个 C2PSA 模块,它结合了逐点空间注意力 PSA 模块,用于增强特征提取和注意力机制,提高了模型对重要特征的捕捉能力。与 YOLOv8 系列模型相比,它具有更少的参数量和更好的检测结果。



Conv:卷积; $k$ :卷积核大小; $s$ :步长;C3k2:特征提取模块;SPPF:空间金字塔池化模块;Split:分割模块;PSABlock:逐点空间注意力模块;Concat:通道拼接;Upsample:上采样;C2PSA:融合跨阶段部分连接和空间注意力机制的模块;Detect:检测头;Backbone:主干网络;Neck:颈部网络;Head:头部网络;CBS:用于特征提取和增强的模块;Conv2d:2维卷积层;MaxPool:最大池化层;C3:跨阶段部分连接结构模块; $n$ :重复次数。

图2 YOLO11n 模型结构

Fig.2 YOLO11n model structure diagram

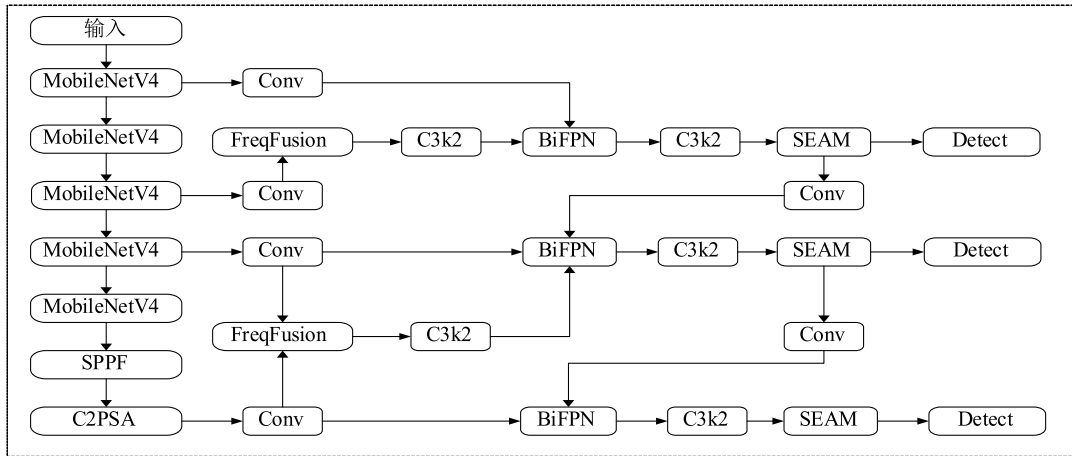
但是在实际应用移动到移动设备时,原始的 YOLO11n 模型还是相对复杂,计算量较大,对设备的计算能力和存储资源有较高的要求。本研究在原始 YOLO11n 模型的基础上进行改进,构建了一种新的轻量化目标识别检测模型——YOLO11n-MFBS,其整体网络结构如图 3 所示。模型主要进行了 3 个方面的改进。首先是用 MobileNetV4 替换 YOLO11n

模型的主干网络,其次是采用 Bi-Freq 模块替换 YOLO11n 模型的特征融合模块,最后是在 YOLO11n 模型的目标检测层中引入 SEAM 注意力机制。

1.2.1 MobileNetV4 MobileNetV4 是一种专为轻量化设计的卷积神经网络,适用于移动设备的高效架构,其核心思想在于利用深度可分离卷积和逐点卷积高效提取特征,同时大幅降低计算复杂度和参数

量<sup>[21-22]</sup>。具体来说,深度卷积在每个通道上独立提取空间信息(图 4a),而逐点卷积则在通道维度上实

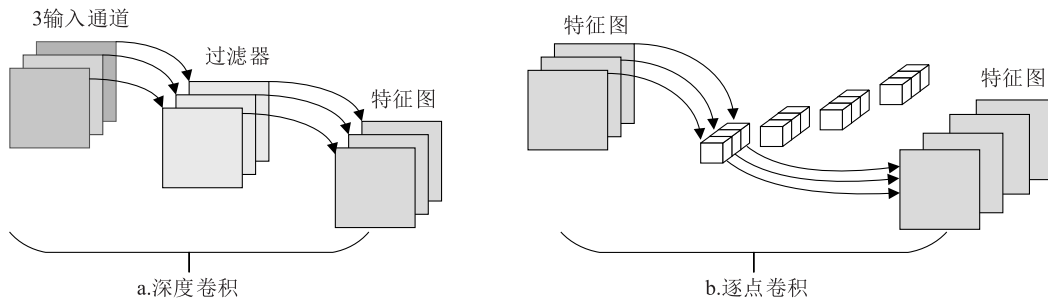
现特征融合和加权(图 4b),从而保证了在降低运算量的同时,仍能有效捕捉局部与全局特征。



MobileNetV4:一种轻量化主干网络架构;FreqFusion:频域感知特征融合模块;BiFPN:加权双向特征金字塔网络;Conv:卷积;C3k2:特征提取模块;SPPF:空间金字塔池化模块;C2PSA:融合跨阶段部分连接和空间注意力机制的模块;SEAM:注意力机制模块;Detect:检测头。

图 3 YOLO11n-MFBS 模型结构

Fig.3 YOLO11n-MFBS model structure diagram



a.深度卷积

b.逐点卷积

图 4 深度卷积与逐点卷积

Fig.4 Depthwise convolution and pointwise convolution

为进一步提升网络的表达能力和灵活性,MobileNetV4 还引入了通用倒置瓶颈模块(UIB),该模块通过引入额外的深度卷积(如 Extra dw、Inverted Bottleneck、Conv next 及 FFN 结构),实现了更细粒度的特征提取与多路径信息融合,有效增强了网络的非线性表达能力,UIB 结构<sup>[23-24]</sup>如图 5 所示。同时,MobileNetV4 还通过优化的神经结构搜索(NAS)策略,在粗粒度搜索与细粒度搜索的结合下,进一步改善了网络结构的设计。

因此,将 MobileNetV4 作为 YOLO11n 的主干网络可以使模型在保持较高检测性能的同时,极大地降低参数量和计算负担。非常适用于移动设备和嵌入式系统等资源受限的场景。

1.2.2 基于 FreqFusion 与 BiFPN 的高效自适应特征融合方法 密集图像预测任务对特征类别一致

性和空间边界精度提出了极高的要求。传统特征融合方法在上采样过程中常会引入高频扰动,导致类别内的一致性边界位移问题。针对这一问题,Chen 等<sup>[19]</sup>提出了 FreqFusion 方法,该方法通过集成自适应低通滤波器(ALPF)生成器、偏移生成器和自适应高通滤波器(AHPF)生成器,显著提升了特征一致性和边界清晰度,核心公式可以表示为:

$$Y_l(i, j) = \tilde{Y}_{l+1}(i+u, j+v) + \tilde{X}_l(i, j) \quad (1)$$

式中,  $\tilde{Y}_{l+1} = \text{FUP}[\text{FLP}(\tilde{Y}_{l+1})]$ , 表示经过自适应低通滤波器(ALPF)生成器和上采样后的高层特征;  $\tilde{X}_l = \text{FHP}(X_l) + X_l$ , 表示低层特征经过自适应高通滤波器(AHPF)生成器增强后与原始特征的叠加;  $(u, v)$  为偏移生成器预测的局部偏移量,用于精确校正边界。

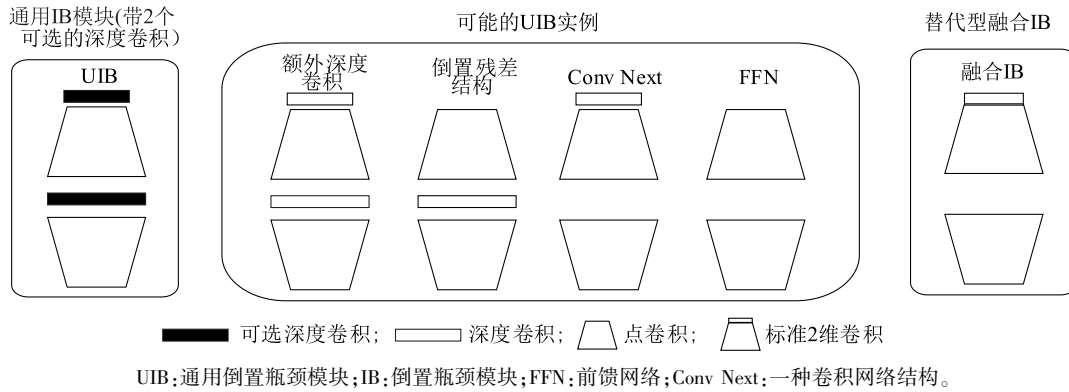


图5 通用倒置瓶颈(UIB)模块结构

Fig.5 Universal inverted bottleneck (UIB) block structure

ALPF生成器通过低通滤波来平滑特征,以减少类别内的不一致性,同时优化上采样过程中引入的高频扰动,其结构如图6a所示。偏移生成器通过重新采样,将低类别内的相似度特征替换为高类别内的相似度邻近特征,从而修复大范围不一致区域并精确边界,其结构如图6b所示。AHPF生成器则通过高通滤波增强边界细节信息,恢复在下采样过程中丢失的高频细节,其结构如图6c所示。这3种生成器协同工作,显著提升了特征提取的准确性,有效解决了标准特征融合技术中常见的特征模糊和边界偏移问题<sup>[25]</sup>。FreqFusion在语义分割、目标检测、实例分割和全景分割等多种密集预测任务中,均可以显著提升模型性能,在保持较低参数数量的同时,实现更高的准确性和鲁棒性。

在多尺度特征融合方面,BiFPN模块采用双向(自顶向下和自底向上)融合策略,同时引入可学习权重对各尺度特征进行加权融合。对于某一融合节点,其输出 $O$ 可表示为:

$$O = \frac{\sum_i w_i \cdot I_i}{O + \sum_i w_i} \quad (2)$$

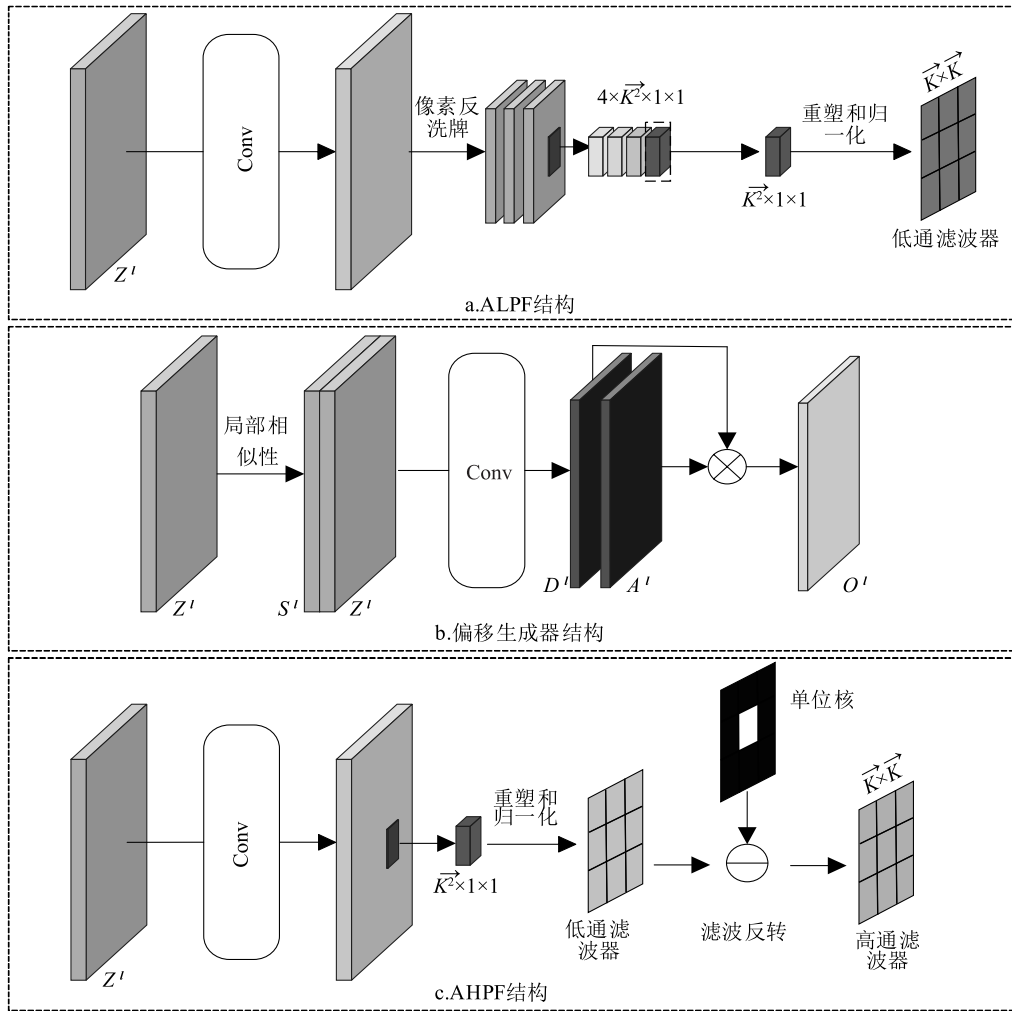
式中, $I_i$ 表示不同尺度的输入特征; $w_i$ 为对应的可学习权重; $O$ 是一个很小的常数,用于防止除零错误。此公式确保了融合过程中的数值稳定性和各尺度特征的合理贡献。

BiFPN解决了传统FPN中单向信息流动和特征贡献不均等问题。在BiFPN中,每个输入特征的权重是可学习的,网络能够自动调整不同尺度特征对融合输出的贡献比例。此外,BiFPN还优化了连接结构,通过去除单输入边节点和添加同层节点间

的直连边,使网络结构更加简洁高效<sup>[26-27]</sup>,其结构如图7所示。BiFPN在多个目标检测任务中具有卓越的性能,其特征融合能力显著优于传统的FPN<sup>[28]</sup>、PANet<sup>[29]</sup>,并以更少的参数量和浮点计算量实现了更高的检测精度。

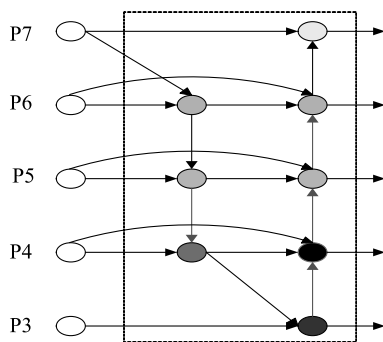
在本研究中,为实现模型的轻量化部署,我们将YOLO11n的原始主干网络替换为参数量更少、计算量更低的MobileNetV4。然而,主干网络的轻量化不可避免地导致特征提取能力有所下降。为弥补轻量化主干网络在特征提取方面的不足,提升模型性能,本研究将FreqFusion与BiFPN结合,设计了一种新的既能捕捉频域特征又能高效融合多尺度信息的特征融合模块Bi-Freq,其结构如图8所示。该设计旨在解决草莓成熟度检测任务中存在的未成熟期阶段草莓目标小、草莓边界特征细微、背景干扰等问题。

Bi-Freq模块首先通过若干个卷积层对来自不同尺度的特征图进行通道数对齐和降维处理,得到初步融合特征;接下来,在较高尺度的特征图上应用FreqFusion模块,对特征进行低通平滑和高通增强,同时利用偏移生成器校正边界,形成调整后的特征 $\tilde{Y}_{i+1}$ 和 $\tilde{X}_i$ ;然后,将经FreqFusion调整后的特征与低层特征进行拼接,输入BiFPN模块,通过公式(2)实现双向特征交互,生成最终的融合特征;最后,以经过BiFPN融合的特征图作为检测头的输入,用于后续任务。Bi-Freq模块不仅能够在特征融合过程中充分保留和强化高频细节信息,同时可以通过BiFPN的双向加权机制实现多尺度特征的高效交互,在草莓成熟度检测任务中以更少的参数量和计算量来提升模型的整体检测精度和定位精度。



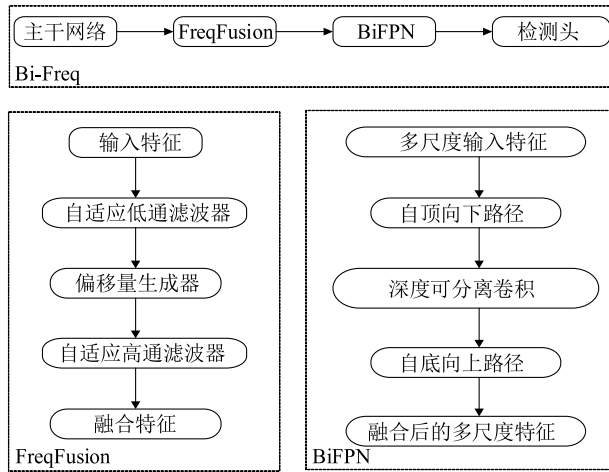
ALPF: 自适应低通滤波器; AHPF: 自适应高通滤波器;  $Z^l$ :  $l$  层的输入特征图;  $S^l$ :  $l$  层局部相似性特征图;  $D^l$ :  $l$  层卷积处理后的特征图;  $A^l$ :  $l$  层注意力图;  $O^l$ :  $l$  层输出结果; Conv: 卷积;  $4 \times \vec{K}^2 \times 1 \times 1$ : 表示生成了 4 个  $\vec{K} \times \vec{K}$  的卷积核;  $\vec{K} \times \vec{K}$ : 最终生成的低卷积核大小;  $\vec{K}^2 \times 1 \times 1$ : 表示一个大小为  $\vec{K} \times \vec{K}$  的卷积核, 并以  $1 \times 1$  的方式使用。

图 6 自适应低通滤波器 (a)、偏移生成器 (b) 和自适应高通滤波器 (c) 结构  
Fig.6 Adaptive low-pass filter and adaptive high-pass filter architecture diagram



P3~P7 表示第 3 层到第 7 层的模型特征。  
图 7 加权双向特征金字塔网络 (BiFPN) 结构  
Fig.7 Weighted bidirectional feature pyramid network (BiFPN) architecture diagram

1.2.3 SEAM 注意力机制 为了提高模型处理复杂背景和被遮挡目标的能力,本研究在 YOLO11n 模型的目标检测层中引入了一种空间注意力机制 SEAM,其结构如图 9 所示。首先采用带残差连接的深度可分离卷积对输入特征进行处理,有效提取每个通道的空间信息。然后将深度卷积的输出通过逐点卷积进行通道间信息融合,弥补深度卷积中通道独立处理带来的信息损失。随后 SEAM 注意力机制使用 2 个全连接层进一步聚合空间与通道特征,增强不同区域之间的上下文联系。最后将 SEAM 注意力机制的输出作为注意力权重与原始输入特征图进行逐点加权融合,从而引导模型更加关注目标区



Bi-Freq:本研究提出的特征融合模块;FreqFusion:频域特征融合模块;BiFPN:加权双向特征金字塔网络。

图8 Bi-Freq 结构

Fig.8 Bi-Freq architecture diagram

域<sup>[30]</sup>。在模型检测过程中,SEAM 注意力机制不仅能够特征提取过程中有效地聚焦于目标区域,还能根据其他视角推断出被遮挡区域的特征,从而提升模型在复杂背景和被遮挡情况下的检测精度<sup>[31]</sup>。

### 1.3 试验环境及参数设置

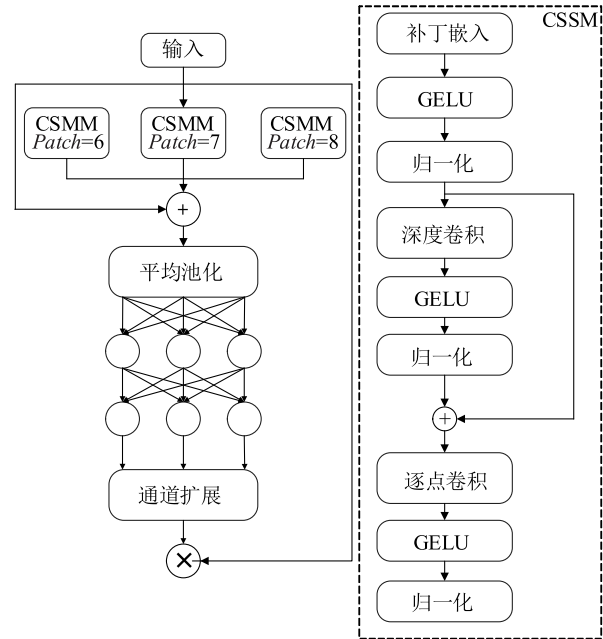
本研究的试验环境为 64 位 Windows11 操作系统,中央处理器(CPU)为 AMD Ryzen 7 8745H with Radeon 780M Graphics,图形处理器(GPU)为 NVIDIA GeForce RTX4060,深度学习框架为 Pytorch,编程环境为 Python3.9,软件平台为 PyCharm,搭配环境为 CUDA12.3+CuDNN8.9.7.29。初始化学学习率设置为 0.01,动量参数设置为 0.937,批量大小设置为 16,权重衰减率设置为 0.0005,迭代次数为 150 次,图片尺寸为 640×640 像素。

### 1.4 轻量化模型对比试验

在本研究中采用 MobileNetV4 替换 YOLO11n 的主干网络以达到轻量化的作用。为了进一步探究不同轻量化改进方法对草莓成熟度目标检测性能的影响,本研究选择将 ShuffleNetV2、EfficientNetV2、GhostnetV2、MobeileNetV4 4 种轻量化网络分别替换 YOLO11n 的主干特征提取网络,进行对比试验。

### 1.5 特征融合模块对比试验

为验证本研究提出的 Bi-Freq 模块在轻量化主干网络下进行草莓成熟度检测任务时的性能,本研究在 YOLO11n 模型的基础上,设计了 4 组对比试



CSMM:卷积注意力子模块;Patch:感受野大小;GELU:激活函数。

图9 SEAM 注意力机制示意

Fig.9 SEAM attention mechanism diagram

验;方案 1 为仅更换主干网络为 MobileNetV4,保留原始特征融合结构;方案 2 和方案 3 分别为在方案 1 的基础上使用 BiFPN 和 FreqFusion 替换原 Neck 中的特征融合模块;方案 4 则在方案 1 基础上替换原 Neck 中的特征融合模块为 Bi-Freq。

### 1.6 消融试验设计

为了更清楚地了解每个改进模块对草莓成熟度检测效果的影响,本研究在原始的 YOLO11n 模型基础上设计 5 组消融试验:方案 1 是采用 MobileNetV4 替换 YOLO11n 模型的主干网络;方案 2 是采用 Bi-Freq 替换 Neck 中的特征融合模块;方案 3 是在目标检测层添加 SEAM 注意力机制;方案 4 是采用 MobileNetV4 替换主干网络和使用 Bi-Freq 替换 Neck 中的特征融合模块;方案 5 是采用 MobileNetV4 替换主干网络和使用 Bi-Freq 替换 Neck 中的特征融合模块以及在目标检测层添加 SEAM 注意力机制。

### 1.7 改进模型与其他主流模型的对比试验

为了进一步验证本研究所改进的模型 YOLO11n-MFBS 的性能,设计对比试验,分别采用 YOLO11n-MFBS、YOLOv5n、YOLOv8n、YOLO11n、YOLOv12n、Faster-RCNN 进行草莓成熟度检测,最后将检测结果与本研究提出的模型进行对比。

## 1.8 评价指标

本研究的目的是更好地让草莓采摘机器人在田间环境下快速准确地检测出草莓果实成熟度,模型更注重于精度检测和效率。因此选择参数量、浮点计算量( $FLOPs$ )<sup>[32]</sup>、平均精度值(交并比阈值为 0.5 的  $mAP_{0.50}$  和交并比阈值为 0.50~0.95 的  $mAP_{0.50:0.95}$ ),  $mAP$  用以下方程式来计算。

$$mAP = \frac{\sum AP}{N} = \frac{\int_0^1 P(R) dR}{N} \quad (3)$$

式中,  $AP$  指针对单一类别的平均精度;  $mAP$  为平均精度值, 在多类别目标检测任务中  $mAP$  是所有类别  $AP$  的算术平均值, 它是衡量模型在所有类别上整体性能的一个重要指标;  $N$  为样本数量;  $P$  为精确率;  $R$  为召回率。

## 2 结果与分析

### 2.1 轻量化模型对比试验结果

本次主干网络轻量化对比试验结果如表 1 所示。整体来看, 4 种轻量化改进方案在检测精度方面差异较小。其中, ShuffleNetV2 在精度上表现相对较好, 但其参数量和计算量相比原始模型分别仅减少了 2.5% 和 6.3%, 轻量化效果有限; EfficientNetV2 与 MobileNetV4 的  $mAP_{0.50:0.95}$  均为 85.9%, 但 MobileNetV4 在参数量和计算量上均明显优于前者; GhostNetV2 虽然精度略高于 MobileNetV4, 但其参数量和计算复杂度也显著增加。MobileNetV4 在精度略降的情况下, 将模型参数量和浮点计算量分别降低至 1.800 M 和 4.3 G, 相较原始模型分别减少 30.5% 和 32.8%, 展现出最优的轻量化性能。因此, 本研究最终选用 MobileNetV4 对 YOLO11n 模型进行轻量化改进。

表 2 特征融合模块对比试验结果

Table 2 Experimental results of feature fusion module comparison

改进方法	试验方案				参数量 (M)	浮点计算量 (G)	$mAP_{0.50}$ (%)	$mAP_{0.50:0.95}$ (%)
	MobileNetV4	BiFPN	FreqFusion	Bi-Freq				
方案 1	√	×	×	×	1.800	4.3	97.1	85.9
方案 2	√	√	×	×	1.885	4.9	97.3	86.6
方案 3	√	×	√	×	1.688	4.9	97.5	86.7
方案 4	√	×	×	√	1.642	4.6	97.6	87.3

√表示使用所对应的改进方法替换 YOLO11n 模型中的相应模块; ×表示不使用所对应的改进方法, YOLO11n 中相应的模块不改变;  $mAP_{0.50}$  表示在交并比为 0.50 的情况下的平均精度值;  $mAP_{0.50:0.95}$  表示分别计算 0.50~0.95 区间内(步长 0.05)的 10 个交并比阈值对应的平均精度均值, 再取平均值。

为了进一步验证 Bi-Freq 模块的性能, 分别用以上 4 种模型对测试集中的部分图片进行检测, 从图

表 1 轻量化模型对比试验结果

Table 1 Experimental results of lightweight model comparison

改进方法	参数量 (M)	浮点计算量 (G)	$mAP_{0.50}$ (%)	$mAP_{0.50:0.95}$ (%)
YOLO11n	2.590	6.4	98.0	87.1
YOLO11n-ShuffleNetV2	2.525	6.0	97.5	86.3
YOLO11n-EfficientNetV2	2.092	5.3	97.2	85.9
YOLO11n-GhostNetV2	2.384	5.6	97.3	86.1
YOLO11n-MobileNetV4	1.800	4.3	97.1	85.9

YOLO11n 为原始模型; YOLO11n-ShuffleNetV2、YOLO11n-EfficientNetV2、YOLO11n-GhostNetV2、YOLO11n-MobileNetV4 分别表示使用 ShuffleNetV2、EfficientNetV2、GhostNetV2、MobileNetV4 替换 YOLO11n 的主干网络;  $mAP_{0.50}$  表示在交并比为 0.50 的情况下的平均精度值;  $mAP_{0.50:0.95}$  表示分别计算 0.50~0.95 区间内(步长为 0.05)的 10 个交并比阈值对应的平均精度均值, 再取平均值。

### 2.2 特征融合模块对比试验结果

从表 2 可以看出, 在方案 1 中, 仅将 YOLO11n 的主干网络替换为 MobileNetV4, 虽然模型参数量与计算量分别降至 1.800 M 和 4.3 G, 具备较高的轻量化优势, 但由于缺乏有效的特征融合策略, 其  $mAP_{0.50}$  和  $mAP_{0.50:0.95}$  分别仅为 97.1% 和 85.9%。方案 2 在此基础上引入 BiFPN, 通过双向多尺度特征融合机制增强了高低层语义信息的交互,  $mAP_{0.50}$  和  $mAP_{0.50:0.95}$  分别提升至 97.3% 和 86.6%。方案 3 将 Neck 部分替换为 FreqFusion 模块, 着重加强了特征边缘与纹理细节的表达能力, 在参数量降至 1.688 M 的前提下, 实现了 97.5% 和 86.7% 的平均精度, 展现出更优的轻量化效果。方案 4 引入本研究提出的 Bi-Freq 模块, 综合利用频域增强与多尺度融合的优势, 在计算量为 4.6 G 和最低参数量 1.642 M 的情况下将  $mAP_{0.50}$  和  $mAP_{0.50:0.95}$  提升至最高值, 分别为 97.6% 和 87.3%, 综合表现最优。

10 中可以看出, 当草莓特征明显、背景干扰较少时 4 种模型均能够较好地完成任务。但是当草莓果

实个体尺寸较小或与背景颜色相近时,方案 1、方案 2 和方案 3 均出现不同程度的漏检现象,相比之下,方案 4 所采用的 Bi-Freq 融合模块在此类复杂环境下表现更为稳定,这也进一步说明,Bi-Freq 通过将

FreqFusion 的频域特征增强能力与 BiFPN 的多尺度语义融合能力有效结合,弥补了轻量化主干在复杂场景下特征提取能力不足的问题,显著提升了模型的检测精度。



方案 1~方案 4 见表 2。raw:未成熟期;turing:转变期;ripe:成熟期;圆圈表示漏检果实;矩形表示正确检测的果实。

图 10 4 种模型检测效果

Fig.10 Detection images of four models

### 2.3 消融试验结果

本次消融试验结果如表 3 所示,当采用原始 YOLO11n 模型时模型的参数量和浮点计算量分别为 2.590 M 和 6.4 G,  $mAP_{0.50}$  和  $mAP_{0.50:0.95}$  分别为 98.0% 和 87.1%;当单独采用 MobileNetV4 替换主干网络时模型参数量和浮点计算量分别减少了 30.5% 和 32.8%,但由于轻量化主干网络在特征提取能力上存在一定局限性,导致  $mAP_{0.50}$  和  $mAP_{0.50:0.95}$  分别下降了 0.9 个百分点和 1.2 个百分点;当采用 Bi-Freq 模块替换原始模型 Neck 模块时,检测精度提升至 98.5% 和 88.4%,验证了替换模块在特征提取方面的优势;当单独引入 SEAM 注意力机制时,参数量和浮点计算量分别上升至 2.693 M 和 6.6 G,检测精度分别达到 98.2% 和 87.4%,表明注意力机制对目标特征的强化有一定的效果,但是会少量增加模型的参数量和计算量;在同时采用 MobileNetV4 与 Bi-Freq 模块的情况下,模型参数量和计算量分别减少 36.6% 和 28.1%,  $mAP_{0.50}$  和

$mAP_{0.50:0.95}$  分别为 97.6% 和 87.3%,这表明 Bi-Freq 模块在保持轻量化特性的同时,能够有效提升模型的特征提取能力;而在方案 4 的基础上添加 SEAM 注意力机制后,模型参数量略增至 1.713 M,浮点计算量为 4.7 G,同时  $mAP_{0.50}$  和  $mAP_{0.50:0.95}$  分别提升至 97.8% 和 87.6%。这一结果表明,方案 5 在显著降低模型复杂度的同时,保持了较高的检测性能,综合表现最佳。因此,方案 5 作为最终改进方案,能够在有效降低计算资源消耗的同时,保持良好的草莓成熟度检测效果,为实际应用提供了更加高效、轻量的解决方案。

### 2.4 不同模型对草莓成熟度检测性能对比试验结果

不同模型对草莓成熟度检测性能的对比试验结果如表 4 所示。与本研究提出的 YOLO11n-MFBS 模型相比,双阶段目标检测模型 Faster-RCNN 的参数量和浮点计算量高且检测精度相对较低;单阶段目标检测模型 YOLOV5n 虽然参数量和浮点计算量

较低,但是  $mAP_{0.50}$  和  $mAP_{0.50:0.95}$  分别比 YOLO11n-MFBS 低 4.6 个百分点和 8.0 个百分点;YOLOv8n、YOLO11n、YOLO12n 的检测精度虽然与 YOLO11n-MFBS 相差不大,但是参数数量和浮点计算量远高于

本研究提出的模型;在这些主流模型中本研究提出的 YOLO11n-MFBS 参数量为 1.713 M,  $mAP_{0.50:0.95}$  值最高,为 87.6%。改进后的模型实现了精度与效率的同步优化,因此更适合应用在移动端设备。

表 3 消融试验结果

Table 3 Ablation study results

试验方案	改进方法			参数量 (M)	浮点计算量 (G)	$mAP_{0.50}$ (%)	$mAP_{0.50:0.95}$ (%)
	MobileNetV4	Bi-Freq	SEAM				
原始模型	×	×	×	2.590	6.4	98.0	87.1
方案 1	√	×	×	1.800	4.3	97.1	85.9
方案 2	×	√	×	2.434	6.8	98.5	88.4
方案 3	×	×	√	2.693	6.6	98.2	87.4
方案 4	√	√	×	1.642	4.6	97.6	87.3
方案 5	√	√	√	1.713	4.7	97.8	87.6

√表示使用所对应的改进方法替换 YOLO11n 模型中的相应模块;×表示不使用所对应的改进方法,YOLO11n 中相应的模块不改变;×表示在交并比为 0.50 的情况下的平均精度值; $mAP_{0.50}$  表示分别计算 0.50~0.95 区间内(步长 0.05)的 10 个交并比阈值对应的平均精度均值,再取平均值。

表 4 不同模型检测性能对比试验

Table 4 Comparison of detection performance across different models

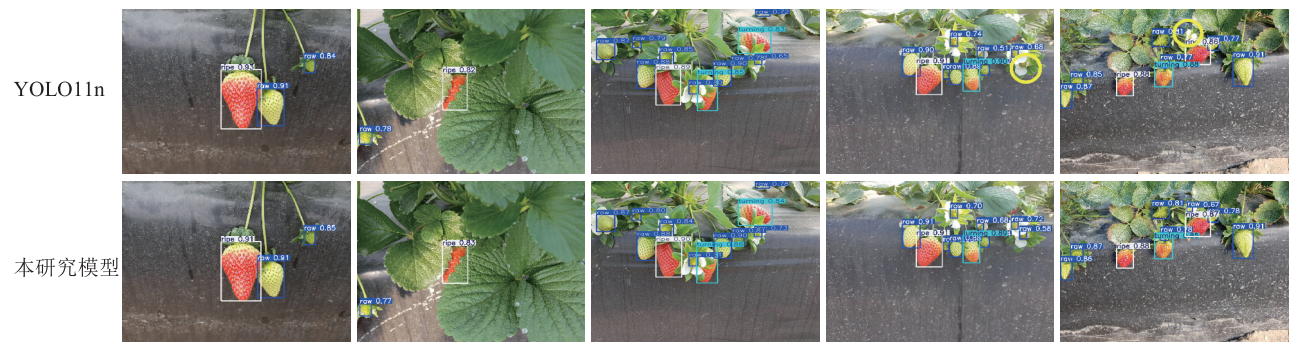
模型	参数量 (M)	浮点计算量 (G)	$mAP_{0.50}$ (%)	$mAP_{0.50:0.95}$ (%)
YOLOv5n	1.767	4.2	93.2	79.6
YOLOv8n	3.011	8.2	97.5	86.2
YOLO11n	2.590	6.4	98.0	87.1
YOLOv12n	2.520	6.0	97.8	86.6
Faster-RCNN	41.760	134.3	94.6	83.5
YOLO11n-MFBS	1.713	4.7	97.8	87.6

YOLOv5n、YOLOv8n、YOLOv12n、Faster-RCNN 分别为当前的主流的卷积神经网络模型;YOLO11n 为原始模型;YOLO11n-MFBS 为本研究改进后的模型; $mAP_{0.50}$  表示在交并比为 0.50 的情况下的平均精度值; $mAP_{0.50:0.95}$  表示分别计算 0.50~0.95 区间内(步长 0.05)的 10 个交并比阈值对应的平均精度均值,再取平均值。

为了进一步验证 YOLO11n-MFBS 模型的性能,用 YOLO11n 模型和 YOLOv11n-MFBS 模型分别对测试集中的部分图片进行检测,检测结果如图 11 所示。从图 11 可以看出,改进后的模型对 3 种成熟度的草莓检测的置信度更高,特别是对遮挡目标和小目标的检测能力要高于原始 YOLO11n 模型。同时在进行密集任务检测时,改进后的模型漏检率比 YOLO11n 更低,对不同成熟度的草莓检测效果更好。

### 3 结论

针对在田间草莓成熟度检测任务中,现有模型参数量大,浮点计算量高,不能满足移动端设备部署



raw: 未成熟期;turning: 转变期;ripe: 成熟期;圆圈表示漏检果实;矩形表示正确检测的果实。

图 11 2 种模型检测效果对比

Fig.11 Comparison of detection performance between two models

条件的问题,本研究在YOLO11n模型的基础上,利用MobileNetV4替换原主干网络,大幅减少参数量和计算量,设计BiFreq特征融合模块,强化多尺度细节表达,并在目标检测层引入SEAM注意力机制,进一步提升对遮挡和小目标的识别能力,构建了轻量化检测模型YOLO11n-MFBS,其参数量降至1.713 M,浮点计算量降至4.7 G,相较原始YOLO11n模型参数量和计算量分别减少33.9%和26.6%。

本研究结果表明,不同的改进模块对模型性能的影响各不相同,当综合使用3种改进模块时,模型的 $mAP_{0.50}$ 和 $mAP_{0.50:0.95}$ 分别达到了97.8%和87.6%,综合性能优于其他方案。与YOLOv5n、YOLOv8n、YOLOv11、YOLOv12n、Faster-RCNN等主流检测模型<sup>[33]</sup>对比,YOLO11n-MFBS在保持低参数量和浮点计算量的情况下, $mAP_{0.50:0.95}$ 最高,达到了87.6%。在实际的测试中,YOLO11n-MFBS模型与原始的YOLO11n模型相比漏检率更低,检测的置信度也更高。

综上所述,YOLO11n-MFBS模型在大幅降低计算资源消耗的同时,依然保持了较高的检测精度和推理速度,具备良好的实时性和鲁棒性,适合部署于移动端和嵌入式草莓采摘机器人,为田间草莓自动化采摘提供了高效可靠的视觉方案。

## 参考文献:

- [1] 张丽勃,邹小花,邓波,等. 草莓产业标准化现状及对推进上海草莓全产业链标准化的思考[J]. 上海农业学报,2024,40(5):80-88.
- [2] ZHANG C, GUO C T, LIU F, et al. Hyperspectral imaging analysis for ripeness evaluation of strawberry with support vector machine[J]. Journal of Food Engineering,2016,179:11-18.
- [3] SHAO Y Y, WANG Y X, XUAN G T, et al. Assessment of strawberry ripeness using hyperspectral imaging[J]. Analytical Letters,2021,54(10):1547-1560.
- [4] 化春键,黄宇峰,蒋毅,等. 基于改进YOLOv7的田间环境下食用玫瑰检测方法[J]. 南京农业大学学报,2025,48(3):714-723.
- [5] 化春键,黄宇峰,蒋毅,等. 基于改进YOLOv5s模型的田间食用玫瑰花检测方法[J]. 江苏农业学报,2024,40(8):1464-1472.
- [6] ZHANG J, YAN L F, HOU J Y. Recognition of rice leaf diseases based on salient characteristics[C]//2018 13th World Congress on Intelligent Control and Automation (WCICA). July 4-8,2018, Changsha, China; IEEE,2018:801-806.
- [7] OTSU N. A threshold selection method from gray-level histograms[J]. IEEE Transactions on Systems, Man, and Cybernetics,1979,9(1):62-66.
- [8] CORTES C, VAPNIK V. Support-vector networks[J]. Machine Learning,1995,20(3):273-297.
- [9] JING J P, ZHANG S J, SUN H X, et al. Detection of maturity of "Okubo" peach fruits based on inverted residual mobile block and asymptotic feature pyramid network[J]. Journal of Food Measurement and Characterization,2025,19(1):682-695.
- [10] YASEEN M. What is YOLOv8: an in-depth exploration of the internal features of the next-generation object detector[J]. arXiv e-prints,2024,arXiv:2408.15857.
- [11] YANG G Y, LEI J, ZHU Z K, et al. AFPN: asymptotic feature pyramid network for object detection[C]//2023 IEEE International Conference on Systems, Man, and Cybernetics (SMC). October 1-4,2023, Honolulu, Oahu, HI, USA; IEEE,2023:2184-2189.
- [12] VASCONEZ J P, DELPIANO J, VOUGIOUKAS S, et al. Comparison of convolutional neural networks in fruit detection and counting: a comprehensive evaluation[J]. Computers and Electronics in Agriculture,2020,173:105348.
- [13] HOWARD A G, ZHU M L, CHEN B, et al. MobileNets: efficient convolutional neural networks for mobile vision applications[J]. arXiv preprint,2017,arXiv:1704.04861.
- [14] LIU W, ANGUELOV D, ERHAN D, et al. SSD: single shot MultiBox detector[M]. Cham: Springer International Publishing,2016:21-37.
- [15] 方国文,何超,王鑫泽. 基于YOLOv8n的轻量化巴旦木果实识别方法[J]. 江苏农业学报,2024,40(9):1662-1670.
- [16] TAN M X, PANG R M, LE Q V. EfficientDet: scalable and efficient object detection[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). June 13-19,2020, Seattle, WA, USA. IEEE,2020:10778-10787.
- [17] KHANAM R, HUSSAIN M. YOLOv11: an overview of the key architectural enhancements[J]. arXiv preprint,2024,arXiv:2410.17725.
- [18] QIN D F, LEICHTNER C, DELAKIS M, et al. MobileNetV4: universal models for the mobile ecosystem[C]//Computer Vision-ECVCV 2024. Cham: Springer Nature Switzerland,2025:78-96.
- [19] CHEN L W, FU Y, GU L, et al. Frequency-aware feature fusion for dense image prediction[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence,2024,46(12):10763-10780.
- [20] WANG Y D, ZHANG J, KAN M N, et al. Self-supervised equivariant attention mechanism for weakly supervised semantic segmentation[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). June 13-19,2020, Seattle, WA, USA; IEEE,2020:12272-12281.
- [21] 陈仁凡,谢知,林晨. 基于YOLO-ODM的温室草莓成熟度的快速检测[J]. 华中农业大学学报,2023,42(4):262-269.
- [22] 赵佰亭,程瑞丰,贾晓芬. 融合多尺度特征的YOLOv8裂缝缺陷检测算法[J]. 计算机工程与应用,2024,60(22):261-270.
- [23] 杨信廷,胡焕,陈晓,等. 多源场景下粘虫板小目标害虫轻量化检测识别模型[J]. 智慧农业(中英文),2025,7(1):111-

- 123.
- [24] 郑红娜,周理想,王 强. 基于 YOLO-PS 的马铃薯幼苗检测方法研究[J]. 中国农机化学报,2024,45(12):245-250.
- [25] 黎祖胜,唐吉深,匡迎春. 基于改进 YOLOv10n 的轻量化荔枝虫害小目标检测模型[J]. 智慧农业(中英文),2025,7(2):146-159.
- [26] 奚小波,丁杰源,翁小祥,等. 基于轻量化 YOLOv5s-MCA 的番茄成熟度检测方法[J]. 农业机械学报,2025,56(3):383-391,436.
- [27] 曾孟佳,徐 欢,黄 旭. 一种基于改进 YOLOv5n 的黄桃虫害检测方法[J]. 安徽农业科学,2025,53(3):236-242.
- [28] LIN T Y, DOLLÁR P, GIRSHICK R, et al. Feature pyramid networks for object detection[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). July 21-26,2017, Honolulu, HI, USA:IEEE,2017:936-944.
- [29] LIU S, QI L, QIN H F, et al. Path aggregation network for instance segmentation[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. June 18-23,2018,Salt Lake City, UT, USA:IEEE,2018:8759-8768.
- [30] 韩科立,王振坤,余永峰,等. 基于改进 YOLO 11n 模型的棉花田间复杂环境障碍物检测方法[J]. 农业机械学报,2025,56(5):111-120.
- [31] 李凯璇,刘晓锋,陈 强,等. YOLOv8-GAIS:一种改进的无人机航拍目标检测算法[J]. 光电工程,2025,52(4):77-90.
- [32] 陈仁凡,孔彦琪,钟雅露,等. 基于改进 YOLO 的轻量级草莓成熟度检测方法研究[J]. 江苏农业科学,2024,52(20):138-147.
- [33] 姜业帆,江朝晖,朱 瑞,等. 基于 PCIA-YOLO 的复杂环境下草莓成熟度轻量化检测方法[J]. 南京农业大学学报,2025,48(4):1-15.

(责任编辑:陈海霞)