doi:10.6041/j.issn.1000-1298.2026.23.029

基于 MLS－Net 模型的复杂苹果园环境下疏果前幼果检测方法

贾伟宽[1]　张凤萍[1]　杨瑜婧[1]　阮承治[2]　谢士琨[1]　印　祥[3]

（1. 山东师范大学计算机与人工智能学院，济南 250358；
2. 武夷学院农机智能控制与制造技术福建省高校重点实验室，武夷山 354300；
3. 山东理工大学农业工程与食品科学学院，淄博 255000）

摘要：针对非结构化果园环境下，疏果前幼果检测面临着果实尺寸小、遮挡严重以及背景复杂等难题，本文基于改进的 YOLO 11 架构，提出了轻量化检测模型 MLS－Net。研究设计了多尺度卷积模块 MSCV，该模块借助并行多尺度特征提取技术，提升了模型对目标尺寸变化的适应性，进而增强了模型的检测能力。同时，引入优化的 SEA 注意力机制，强化关键特征通道表达，有效抑制无关背景。此外，采用轻量化共享检测头 Detect_LSCD，在提升检测精度的同时，显著降低了模型参数量。为提高模型的泛化能力，研究构建了涵盖多种光照条件、背景复杂度以及果实形态的数据集。实验结果显示，MLS－Net 在测试集上表现出色，AP_{50} 达到 88.2%，精确率和召回率分别为 93.6% 和 78.5%，检测帧率为 89.1 f/s，参数量降低至 2.34×10^6。与 YOLO 11、FCOS 等主流模型相比，MLS－Net 在精度与效率的平衡性方面更具优势，AP_{50} 提升了 1.6～7.3 个百分点，参数量减少了 9.3%，推理速度提升了 25%。这表明 MLS－Net 能够有效应对果园环境中的光照波动、枝叶遮挡和果实密集等复杂情况，为智能疏果装备的研发提供了可靠的技术支撑。

关键词：苹果园；幼果检测；MLS－Net；多尺度卷积；通道注意力机制

中图分类号：TP391.4　　**文献标识码**：A　　**文章编号**：1000-1298(2026)23-0307-09　　**OSID**：

Young Fruit Detection Model Based on MLS－Net for Pre-thinning in Complex Apple Orchards

Jia Weikuan[1]　Zhang Fengping[1]　Yang Yujing[1]　Ruan Chengzhi[2]　Xie Shikun[1]　Yin Xiang[3]

(1. *School of Computer Science and Artificial Intelligence*, *Shandong Normal University*, *Ji'nan* 250358, *China*
2. *Key Laboratory of Agricultural Machinery Intelligent Control and Manufacturing of Fujian Province University*, *Wuyi University*, *Wuyishan* 354300, *China*
3. *School of Agricultural Engineering and Food Science*, *Shandong University of Technology*, *Zibo* 255000, *China*)

Abstract: In unstructured orchard environments, pre-thinning young fruit detection faces numerous challenges, including small fruit size, severe occlusions, and complex backgrounds. To effectively address these issues, a lightweight detection model, MLS－Net, was proposed based on an improved YOLO 11 architecture. A multi-scale convolution module (MSCV) was designed, leveraging parallel multi-scale feature extraction to enhance the model's adaptability to variations in target sizes, thereby improving detection performance. At the same time, an optimized squeeze-and-excitation attention (SEA) mechanism was introduced to strengthen the representation of critical feature channels and effectively suppress irrelevant background interference. In addition, a lightweight shared detection head, Detect_LSCD, was employed to improve detection accuracy while significantly reducing the number of model parameters. To further enhance generalization capability, a dataset was constructed, covering diverse illumination conditions, complex backgrounds, and various fruit morphologies. Experimental

收稿日期：2025－08－26　修回日期：2025－10－15

基金项目：国家自然科学基金项目(62372278)、山东省自然科学基金项目(ZR2020FM076)、南平市科技特派员资源产业科技创新联合资助项目(N2023Z003)、福建省科技厅引导性项目(2026N0015)和农机智能控制与制造技术福建省高校重点实验室开放基金项目(AMICM202501)

作者简介：贾伟宽(1982—)，男，副教授，博士，主要从事人工智能和智慧农业研究，E-mail：jwk_1982@163.com

通信作者：阮承治(1984—)，男，教授，博士，主要从事智能控制和智能农业机械研究，E-mail：ruanczhi@163.com

results demonstrated that MLS - Net achieved excellent performance on the test set, with an AP_{50} of 88.2%, precision of 93.6%, recall of 78.5%, and an inference speed of 89.1 f/s, while reducing the number of parameters to 2.34×10^6. Compared with mainstream models such as YOLO 11 and FCOS, MLS - Net demonstrated a superior balance between accuracy and efficiency, with AP_{50} improvements of 1.6 ~ 7.3 percentage points, a 9.3% reduction in model parameters, and a 25% increase in inference speed. These results indicated that MLS - Net could effectively handle variations in illumination, branch and leaf occlusions, and dense fruit distributions in orchard environments, providing a reliable technical foundation for the development of intelligent fruit-thinning equipment.

Key words: apple orchards; young fruit detection; MLS - Net; multi-scale convolution; channel attention mechanism

0 引言

我国果园种植面积与水果产量持续保持全球领先地位[1],庞大的消费市场需求推动着水果产业稳步发展。然而,在传统果蔬生产体系中,疏果、采摘等关键环节仍高度依赖人工作业,存在劳动强度大、作业效率低等突出问题。其中,疏果作业作为果树栽培管理的重要技术措施,其执行效率直接影响果实商品率和果园经济效益。研发基于智能技术的疏果前幼果检测方法,对于提升作业效率、稳定果实产量、降低生产成本以及推动果蔬产业智能化转型升级具有重要现实意义[2]。近年来,随着深度学习技术的突破性发展,果实识别研究取得显著进展,但复杂果园场景下的果实精准检测仍面临多重技术挑战。在复杂果园环境中,首先,果实与背景间存在的色度相近、边缘模糊等特征干扰严重影响检测精度;其次,果园环境动态变化显著,光照强度波动、夜间低照度条件以及雨雾等恶劣天气因素对检测系统构成额外干扰;此外,果实品种的形态多样性、不同生长期的表型特征变化以及枝叶遮挡效应等因素进一步增加了检测难度。因此,开发高精度、强鲁棒性的果实检测算法已成为满足现代农业自动化采收、产量预估和生长监测需求的迫切任务,也是当前农业人工智能领域的研究热点[3]。

在果实检测技术演进过程中,早期研究主要基于传统机器学习方法,学者提出了包括 SVM、随机森林等在内的多种解决方案。虽然此类传统机器学习方法在特定场景下取得突破,但其特征表达能力有限且环境适应性不足,难以满足复杂果园环境下的检测需求。随着深度学习技术的突破性发展,基于深度神经网络的检测方法逐渐成为主流研究方向。

卷积神经网络凭借其强大的特征自主学习能力,在图像识别领域展现出显著优势[4]。研究者针对果园复杂环境特点,从模型结构优化[5-7]、多模态数据融合[8-12]、轻量化设计[13-15]等维度开展创新研究。在特征表征能力、环境适应性及检测效率等方面的突破,为疏果前幼果的精准检测提供了重要技术参考。

基于深度学习的果实检测技术,特别是卷积神经网络架构,显著提升了复杂场景下的检测性能。然而,实际应用仍面临三大技术瓶颈:①疏果期幼果与枝叶背景存在色度趋同性及形态相似性,导致特征混淆与误判。②果实间相互重叠以及枝叶遮挡等复杂场景下,易产生误检和冗余检测问题。③非结构化果园环境的动态多变特性,检测模型鲁棒性与泛化能力不足,从而影响其在实际生产环境中的稳定部署与推广应用。

近年来,YOLO 系列目标检测算法因结构简洁、运算高效,已在智慧农业领域得到广泛应用。尽管近年来轻量化目标检测网络在果园场景中实现了实时检测,但在处理小目标及复杂背景时仍存在性能瓶颈。针对果园场景下环境复杂、光照变化大以及小目标难以识别的问题,为进一步提升疏果前幼果检测的精度与效率,本研究以 YOLO 11[16] 为基线,构建轻量化疏果前幼果检测模型 MLS - Net。本研究在保持 YOLO 11 原有轻量化特性的基础上,对网络结构进行针对性优化。提出 MSCV 模块,在特征提取阶段替代原有 C3k2 结构,通过并行多尺度卷积增强特征提取能力,有效捕捉不同尺度感受野信息,从而更好地适应幼果多尺度特性;添加优化后的 SEA 注意力机制,结合通道注意力,自适应调整通道权重,突出目标区域特征、抑制背景噪声,有效增强小目标响应能力,并提升果实与枝叶颜色相近或被遮挡情况下的区分能力;采用 Detect_LSCD 结构实现轻量化检测头设计,通过共享卷积减少计算冗余,并结合分布焦点损失优化边界框预测,在保证检测性能的前提下进一步降低模型复杂度,提升部署效率与实用性。

1 数据集

1.1 图像采集

本研究的图像采集工作于山东省淄博市沂源县

悦庄镇张家庄村(北纬 36°23′,东经 118°29′)进行,采集苹果品种包括红富士、金帅及美国 8 号。数据采集时间为 2024 年 5 月初至 5 月中旬,每日 06:00 至 22:00 进行。采集设备采用 HUAWEI Nova7 型智能手机,图像分辨率为 3 456 像素 ×4 608 像素,存储格式为 JPG。不同苹果品种幼果图像如图 1 所示。



图 1　不同苹果品种幼果图像

Fig. 1　Young fruit images of different apple varieties

本研究所有图像均采集于自然果园环境,背景包括土壤、天空及其他自然环境要素,采集过程中采用多维度策略:空间维度上,通过多角度(俯角、仰角)和多距离进行拍摄,分别从近距离(10 ~ 30 cm)和远距离(50 ~ 100 cm)获取图像,拍摄距离由操作人员根据相机镜头与目标果实之间的空间位置关系进行人工估测,并通过保持相机与果实之间的相对尺度差异来控制,从而确保在不同拍摄条件下获得具有代表性的数据;光照维度上,涵盖顺光、逆光等多种光照条件;时间维度上,采集不同天气(晴天、阴天、雾天、雨天、雨后)及不同时间(白天、夜间)条件下的图像数据。不同拍摄条件下部分幼果图像示例如图 2 所示。



图 2　不同环境下的幼果图像

Fig. 2　Young apple images in different environments

1.2　数据集制作

对采集的原始图像进行筛选,去除重复图像及果实边缘过度模糊的图像后,最终保留 1 000 幅有效图像,其中红富士、美国 8 号、金帅苹果分别占比 52%、41%、7%。所有保留图像均使用 LabelMe 进行标注,并按照比例 7∶3 随机划分为训练集和测试集,分别统计标签数量,具体统计结果如表 1 所示。

表 1　苹果数据集组成

Tab. 1　Composition of apple dataset

数据集	图像数量/幅	标签数量
训练集	700	4 942
测试集	300	2 779
总计	1 000	7 721

2　MLS－Net 模型

在非结构化果园环境中,高重叠的小苹果检测具有较大挑战,模型的参数量与计算复杂度也影响检测的效率与实时性。本研究基于 YOLO 11 框架,提出了改进模型 MLS－Net。其网络结构如图 3 所示,主要由 Backbone 模块、Neck 模块和 Head 模块 3 部分组成。在 Backbone 模块中,设计了 MSCV 模块,利用并行多尺度卷积结构扩展感受野范围,有效增强了疏果前幼果的特征提取能力。引入改进的 SEA 机制,通过自适应地调整通道权重,以精准聚焦目标的关键信息。检测头采用轻量化共享卷积架构,并通过参数共享策略显著降低模型参数量,提高检测效率,实现高效、精准的疏果前幼果检测。

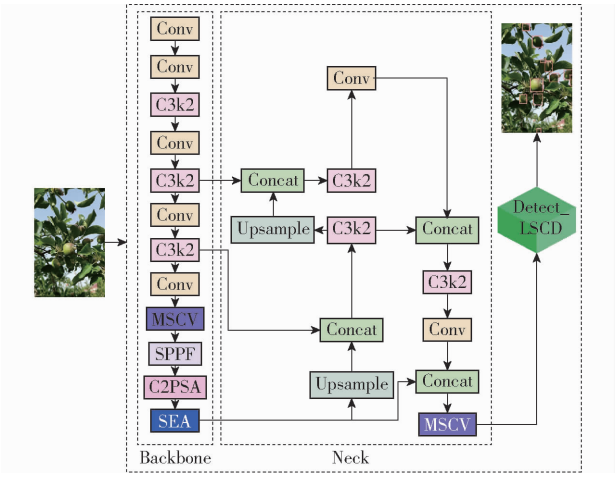

图 3　MLS－Net 总体结构图

Fig. 3　Overall architecture diagram of MLS－Net

2.1　MSCV 特征增强

为了提升模型对不同尺度幼果的特征表达能力,本文设计了一种多尺度卷积(Multi-scale convolution, MSCV)模块。在深度神经网络的特征提取过程中,通常会生成多个存在较高相似性的特征图,即存在特征冗余现象。受 GhostNetv1[17] 启发,通过充分利用多尺度并行卷积与通道划分策略,减少计算开销的同时提升了特征表达的丰富性与网络的表示能力。此外,考虑到传统特征提取操作在单一特征通道上独立进行,各通道间缺乏有效信息交互,受 MobileNetv1[18] 的启发,采用 1 ×1 卷积机制

以实现逐点卷积功能,促进通道特征的有效融合。通过多尺度特征交互与融合,模型能够同时保留低层细粒度特征与高层语义信息。针对疏果前果实普遍较小的特性,MSCV 模块有助于提升小尺度苹果特征的显著性,增强模型对多尺度目标的适应性。MSCV 模块结构如图 4 所示,整体计算公式为

$$Y = \mathrm{Conv}_{1\times1}\{\mathrm{Concat}(X_{\mathrm{cheap}}, \mathrm{Concat}(X_{\mathrm{group},G}))\} \tag{1}$$

式中 X_{cheap}——低成本特征通道

$X_{\mathrm{group},G}$——多尺度特征通道

G——低成本特征通道分支的子分组数量

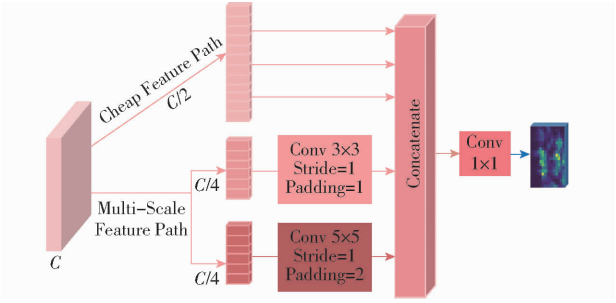

图 4 MSCV 结构图

Fig. 4 MSCV architecture diagram

MSCV 模块整体由两条并行通道构成。首先,输入特征图按通道维度划分为两部分:其中一半通道直接保留,作为低成本特征通道 X_{cheap},该分支不经过额外卷积操作,能够有效保留输入特征中的基础语义信息,同时避免冗余计算,保证了整体计算量的可控性。另一半通道划分为 G 个子分组,形成多尺度特征通道 X_{group}。在本研究中取 $G=2$,分别采用 3×3 与 5×5 卷积核进行特征提取,各分支的 Stride 设为 1,Padding 设置为 kernel/2 以保证空间尺寸一致。每个分支的输出通道数统一为输入通道的 1/4。不同尺度卷积分支既能提取局部细节特征,又能通过较大卷积核扩大感受野、捕捉全局上下文信息,从而增强模型对目标尺寸变化的适应性。随后,将多尺度分支的输出在通道维度上拼接,并与低成本特征通道进行融合。最后,通过一个 1×1 卷积实现通道压缩与特征整合,使输出保持与输入相同的通道数 C。

MSCV 模块在多尺度感知能力与计算效率之间达成了良好的平衡,其设计巧妙地融合了多尺度卷积与低成本特征通道的优势。这种组合不仅保留了输入特征图的基础语义信息,还通过多尺度卷积提取了丰富的细节特征,同时确保了模型在复杂场景下仍能保持较高的运行速度和较低的参数量。

2.2 SEA 注意力机制优化

针对果园中幼果体积小、易被枝叶遮挡,以及果实与枝叶颜色相近等复杂背景问题,本研究引入优化后的 SEA (Squeeze-and-excitation attention)[19] 模块。如图 5 所示,SEA 模块对输入特征进行处理,主要包括 3 个关键操作:全局信息压缩(Squeeze)、两层全连接映射(Excitation)以及特征重标定(Recalibration)。

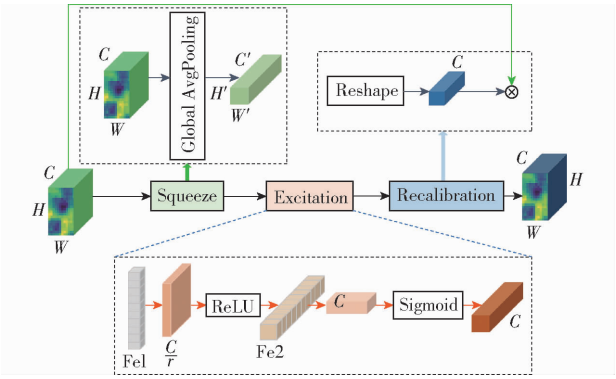

图 5 SEA 结构图

Fig. 5 SEA architecture diagram

给定输入特征图 $X \in \mathbf{R}^{B\times C\times H\times W}$,其中 B 表示批次大小,C 表示通道数,H、W 分别表示特征图的高度和宽度,模块首先通过一个自适应平均池化层对各通道的空间信息进行聚合。与传统固定核大小的池化方式不同,该层可根据输入特征图尺寸自适应地确定池化核大小和步长,以保证输出空间尺寸固定为 1×1。在实际计算中,每个通道独立进行池化操作,相当于在该通道上使用一个尺寸与特征图完全一致的池化窗口(高为 H、宽为 W),将该通道所有像素取平均后输出为一个标量值。该操作可将二维空间特征压缩为单一的统计值,实现从“空间域”到“通道域”的特征聚合。信息压缩操作生成的通道描述符记为 Z,用于表征各通道的全局响应强度。对于第 c 个通道,其全局描述符 Z_c 的计算公式为

$$Z_c = \frac{1}{|\Omega|}\iint_{\Omega} x_c(p)\,\mathrm{d}p \tag{2}$$

式中 $x_c(p)$——输入特征图 x 在第 c 个通道上位置 $p=(i,j)$ 的特征值

Ω——特征图的空间区域,即高度 H 和宽度 W 的二维平面

$|\Omega|$——空间区域总面积

激励操作通过两层全连接映射对通道间的依赖关系进行自适应建模,生成通道注意力权重。第 1 层线性变换将通道数从 C 减少到 C/r(其中 r 是缩减比例,本研究中 r 取值为 8),实现信息的紧凑表达并减少计算量;随后,通过 ReLU 激活函数引入非线性能力,以提升特征区分度。第 2 层线性映射再将通道维度恢复至原始大小 C,并通过 Sigmoid 函数将输出范围约束在[0,1],生成各通道的权重系数

向量。最终，将该权重向量调整为 $\mathbf{R}^{B\times C\times 1\times 1}$ 的形状，并与原输入特征图逐通道相乘，实现特征的自适应重标定。权重较大的通道代表与苹果幼果特征关联性更强的区域，而权重较低的通道则对应于噪声或背景特征。

SEA 模块能够显著提升模型的特征辨识能力，使网络在果园复杂光照、遮挡及同色干扰背景下更有效地捕获幼果关键特征，缓解幼果与背景之间的同色干扰，同时强化幼果目标的特征表征能力。

2.3　Detect_LSCD 检测头设计

传统检测头通常采用多个独立卷积层分别处理不同尺度的特征图，存在明显的计算冗余。为解决这一问题，本研究引入轻量化共享卷积检测头 Detect_LSCD[20]，通过共享卷积权重减少计算量。同时，结合分布焦点损失 DFL[21] 优化边界框回归，使模型在降低复杂度的同时进一步提升目标定位精度。Detect_LSCD 具体架构如图 6 所示。

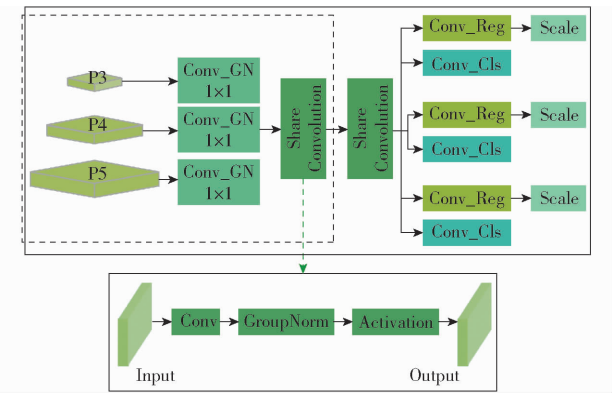

图 6　Detect_LSCD 结构图

Fig. 6　Detect_LSCD architecture diagram

Detect_LSCD 模块接收来自主干网络和颈部网络输出的多尺度特征图作为输入，分别对应不同的空间分辨率与语义层次。针对每个尺度的特征图，首先引入带归一化的卷积模块进行特征提炼。该模块使用 3×3 卷积核，步长为 1，在边缘处填充 1 像素以保持空间尺寸不变。卷积后通过组归一化实现特征分布的稳定化，再经过非线性激活函数增强特征的区分能力。经过这一过程，各尺度特征的通道数被统一为设定的隐藏通道维度。在特征提炼之后，Detect_LSCD 引入了共享卷积结构以实现跨尺度特征的统一建模：第 1 层为 3×3 深度可分离卷积，每个通道独立进行卷积操作（分组数与通道数相同），极大地减少了计算量。第 2 层为 1×1 卷积，用于跨通道特征的重新映射与信息融合。在输出阶段，Detect_LSCD 将共享卷积后的特征分为两个预测分支：边界框回归分支，采用 1×1 卷积层输出 $4\times R_{\max}$ 个通道，用于预测目标边界框的分布参数；分类分支，同样通过 1×1 卷积层输出与类别数 N_c 对应的通道数，用于预测目标的类别概率。两分支的输出在通道维度上进行拼接，形成综合预测特征图，其通道总数为 $4\times R_{\max}+N_c$，包含完整的位置信息与类别信息。

在边界框回归过程中，Detect_LSCD 使用离散分布回归（DFL）机制。该机制将每个边界框回归任务离散化为多个候选偏移值，并通过 Softmax 函数计算每个候选偏移值的概率分布。对于第 i 个输入特征图，其边界框预测可表示为

$$B_i=\mathrm{Softmax}(W_bG_i+b_b)D \tag{3}$$

式中　B_i——第 i 个输入特征图的边界框偏移值

D——离散的偏移量集合

W_b——回归头权重

b_b——回归头偏置

G_i——输入特征图的卷积输出

随后，通过边界框解码函数将离散偏移值转换为实际图像坐标，实现最终的边界框预测，计算公式为

$$\hat{B}_i=d(B_i,A_i,s) \tag{4}$$

式中　A_i——第 i 个输入特征图锚框

s——步长

d——边界框解码函数

Detect_LSCD 模块的共享卷积操作提高了检测精度，减少了模型的训练难度和对大规模数据集的依赖，尤其在疏果前幼果检测场景下表现出更优的鲁棒性和精度。

3　实验

3.1　实验环境与设置

本研究中进行的所有实验都在同一服务器系统上完成，服务器系统主要配置的是 Ubuntu 18.04 64 位操作系统、NVIDIA GeForce RTX 3090 24 GB 和 V11.4 CUDA 环境。所有模型均使用 Python 3.8 版本和 Pytorch 1.10.0 版本进行开发，模型组件利用 MMdetection v3.0.0 学习库构建。初始学习率设置为 0.01，并采用 0.000 5 的权重衰减来缓解过拟合。为优化网络参数并有效地最小化损失函数，使用动量为 0.937 的随机梯度下降优化器（SGD）来辅助训练。最后，对模型进行了 200 轮的训练，最终得到模型在数据集上的 P－R 曲线，如图 7 所示。

3.2　评价指标

为验证 MLS－Net 在疏果前幼果检测任务中的有效性与准确性，本研究采用不同置信度阈值下的精确率-召回率曲线（P－R 曲线）作为模型的性能评估指标。

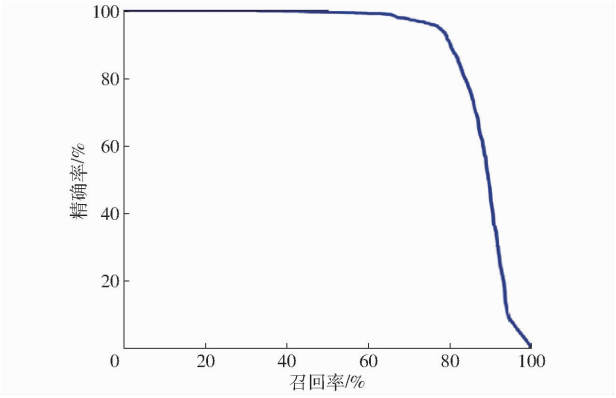

图 7 P－R 曲线

Fig. 7 P－R curve

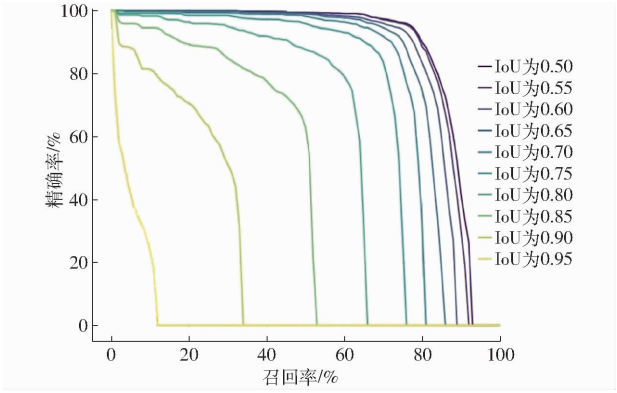

图 8 不同 IoU 阈值下的精确率与召回率曲线

Fig. 8 Precision－recall curves at different IoU thresholds

通过分析 P－R 曲线的形态，可以全面评估模型在不同置信度阈值下的检测性能，为模型优化提供量化依据，平均精度（AP）能够有效衡量模型在精确率与召回率之间的权衡关系。

FLOPs 用于衡量模型的计算复杂度，即网络在一次前向推理过程中所执行的浮点运算次数，较低的 FLOPs 表明模型计算轻量。IoU 用于量化目标检测任务中预测框与真实框之间的空间重叠程度。P－R 曲线与置信度阈值之间的关系能够有效揭示 MLS－Net 模型在不同 IoU 阈值下检测性能的变化规律，如图 8 所示。模型在 IoU 为 0.50～0.75 范围内保持了极高的精确率与召回率，P－R 曲线表现优异；即便在 IoU 为 0.90 以上的极端严格条件下，模型仍展现出可接受的性能边界，证明了其在常规及高要求场景下的综合检测能力。

3.3 消融实验

如表 2 所示，本研究采用模块化引入策略，系统分析不同组件对目标检测模型性能的影响。基线模型 YOLO 11 的 AP_{50} 为 86.3%，精确率和召回率分别为 91.7% 和 77.1%，模型参数量为 2.58×10^6，浮点运算次数为 6.3×10^9。引入 MSCV 模块后，模型参数量略微减少至 2.53×10^6，AP_{50} 提升至 86.8%。单独引入 Detect_LSCD 模块，AP_{50} 进一步优化至 86.9%，同时参数量减少至 2.42×10^6，浮点运算次数降至 5.6×10^9。联合使用 MSCV 和 Detect_LSCD 模块，AP_{50} 提高至 87.3%，参数量进一步压缩至 2.32×10^6，浮点运算次数降至 5.5×10^9，表明模块协同可显著优化精度与效率。最终，本研究提出的改进模型 MLS－Net 在融合 MSCV、Detect_LSCD 和 SEA 模块后实现最优性能，AP_{50} 提升至 88.2%，精确率和召回率分别达到 93.6% 和 78.5%，相较于基准模型分别提升了 1.9 个百分点和 1.4 个百分点，参数量为 2.34×10^6，浮点运算次数保持 5.5×10^9，验证了模块化设计在平衡精度与资源效率方面的有效性。

表 2 消融实验结果

Tab. 2 Ablation study results

序号	MSCV	Detect_LSCD	SEA	精确率/%	召回率/%	AP_{50}/%	参数量	浮点运算次数
1	×	×	×	91.7	77.1	86.3	2.58×10^6	6.3×10^9
2	√	×	×	91.0	79.1	86.8	2.53×10^6	6.3×10^9
3	×	√	×	91.9	78.5	86.9	2.42×10^6	5.6×10^9
4	√	√	×	92.8	77.8	87.3	2.32×10^6	5.5×10^9
5	√	√	√	93.6	78.5	88.2	2.34×10^6	5.5×10^9

注：“√”表示采用该模块，“×”表示不采用该模块。

3.4 结果分析

为了进一步验证 MLS－Net 模型的检测性能，将其与经典和最先进的检测算法进行比较，包括 Backbone 方法 EfficientNet[22]，两阶段方法 Faster R－CNN、Cascade[23]、DINO[24]，一阶段方法 FCOS[25]、YOLO v10[26]、YOLO 11、Hyper－YOLO[27]。实验在相同的环境下使用相同的设备进行。从表 3 的结果可以看出，改进后的模型在 AP_{50}、AP_{75}、AP_{50-95} 等指标分别达到了 88.2%、72.7%、63.2%。与经典模型 Cascade 相比，AP_{50} 提升了 1.6 个百分点。与 FCOS 相比，AP_{50} 显著提升了 7.3 个百分点，AP_{75} 提升了 9.9 个百分点，AP_{50-95} 提升了 7.7 个百分点。与 YOLO 11 相比，AP_{75} 提升了 1.4 个百分点，AP_{50-95} 提升了 1.3 个百分点且参数量减少 9.3%。在一阶段方法中，相较于 Hyper－YOLO，在 AP_{50} 和 AP_{75} 上分别提升 1.8 个百分点和 3.0 个百分点。

综合表明，本模型通过高效结构设计，在检测精

表 3　不同算法对幼果果实检测结果的对比

Tab. 3　Comparison of fruit detection results for young fruits using different algorithms

模型	锚框	AP_{50}/%	AP_{75}/%	AP_{50-95}/%	参数量	浮点运算次数	帧率/(f·s^{-1})	推理延时/ms
EfficientNet	×	85.5	67.4	59.0	1.836×10^{7}	5.44×10^{10}	17.9	55.9
Faster R－CNN	√	82.2	65.6	56.9	4.176×10^{7}	9.15×10^{10}	18.3	54.6
Cascade	√	86.6	70.5	60.7	6.930×10^{7}	1.19×10^{11}	16.6	60.2
DINO	√	84.6	70.0	60.7	4.754×10^{7}	1.19×10^{11}	17.4	57.5
FCOS	×	80.9	62.8	55.5	3.230×10^{7}	8.01×10^{10}	17.4	57.5
YOLO v10	×	84.7	70.1	60.6	2.690×10^{6}	8.20×10^{9}	70.7	14.1
YOLO 11	×	86.3	71.3	61.9	2.580×10^{6}	6.30×10^{9}	71.3	14.0
Hyper－YOLO	×	86.4	69.7	61.5	2.680×10^{6}	7.60×10^{9}	78.6	12.7
MLS－Net(本文)	×	88.2	72.7	63.2	2.340×10^{6}	5.50×10^{9}	89.1	11.2

度、资源占用与实时性之间实现了更优平衡。传统检测器的帧率均在 16～18 f/s 之间，推理延时维持在 54～60 ms，难以满足实时性需求。而 YOLO 系列及其改进模型整体表现显著优于传统方法，帧率均超过 70 f/s，延时控制在 11～14 ms。本研究提出的模型在此基础上进一步提升，帧率和延时分别达到 89.1 f/s 与 11.2 ms，帧率提升 25%，在所有对比模型中推理效率最高，充分验证了其在实时检测任务中的优势。

通过对样本图像分析可知，疏果前幼果目标检测面临以下主要技术挑战：检测目标果实大小和形状的变异性较大；复杂的果园环境中果实与枝叶之间频繁遮挡和重叠，影响了边界特征的完整性；背景环境复杂且枝叶纹理干扰明显。本研究从测试集中选取了 6 幅具有代表性的疏果前幼果图像样本，以客观反映模型在未知数据上的检测性能。这些样本涵盖了同色系背景环境、不同光照条件、夜间环境以及遮挡等典型场景。图 9 展示了各检测器在上述场景下的可视化结果，其中一阶段检测模型检测框为粉色，EfficientNet 和两阶段检测模型检测框为红色，所有模型误检和漏检检测框为黄色。相比之下，MLS－Net 在处理果实大小差异、叶片遮挡、目标重叠及背景相似等复杂情况时，仍表现出较为稳定的检测性能。然而，从图 9 中第 2 列样例黄色框可以观察到，当树干对果实形成显著遮挡时，所有检测器均未能识别被遮挡目标。这主要因为数据集中枝干遮挡样本数量有限，模型难以充分学习相关特征，检测效果因此受限。这也表明后续研究需在数据采集环节增加多样化场景样本，并在模型结构中强化对遮挡与复杂背景特征的建模能力。



图 9　数据集检测结果对比

Fig. 9　Comparison of detection results for dataset

4　讨论

消融实验结果表明，所提出的 MSCV 模块与改进后的 SEA 注意力机制在增强多尺度特征表达、抑制复杂背景干扰方面具有显著效果；Detect_LSCD 模块在保持检测精度的同时显著降低了参数量与计算复杂度，验证了轻量化结构设计的合理性。然而，

在极端光照条件(如强逆光、低照度夜间场景)下,模型的特征响应仍存在不稳定现象,说明其对光照变化的鲁棒性仍需进一步提升。后续研究可考虑引入光照不变特征提取或多模态信息融合(如红外或深度图像)等策略,以增强模型在复杂环境下的稳健性。与两阶段检测器相比,MLS-Net 在保持相当检测精度的同时,显著减少了模型参数量和计算开销,体现了一阶段轻量化检测器在农机视觉任务中的应用潜力。然而,与基于 Transformer 的检测器相比,虽然 MLS-Net 在实时性方面具有优势,但其对高遮挡场景的全局特征建模能力相对较弱。未来研究可探索引入轻量化自注意力编码器或改进的视觉 Transformer 模块,以进一步提升复杂目标间的特征关联建模能力。数据集涵盖多种光照与天气条件,但样本主要集中于单一地区与少数苹果品种,数据分布的局限性可能影响模型的泛化性能。从可视化结果(图 9)可见,当果实发生严重遮挡或重叠时,模型易出现漏检或误检现象,说明其在学习复杂遮挡特征方面仍有不足。

5 结论

(1)为解决果园复杂环境下疏果前幼果的检测难题,本研究构建了一个疏果前幼果检测数据集,提出了一种优化 YOLO 11 算法,通过设计特征提取模块 MSCV,引入轻量化检测头增强算法对疏果前幼果检测性能。

(2)构建了涵盖红富士、金帅和美国 8 号苹果的数据集,包含多种光照条件、背景复杂度和果实形态,总计 1 000 幅标注图像,为复杂果园环境下的疏果前幼果检测研究提供了数据支撑。

(3)设计 MSCV 模块在 YOLO 11 特征提取阶段替换 C3k2 模块,提升模型对不同尺度目标的感受能力;添加优化的 SEA 注意力机制,抑制无关信息;采用 Detect_LSCD 实现轻量化设计。实验结果表明 MLS-Net 模型相较于基准模型精确率和召回率分别提升了 1.9 个百分点和 1.4 个百分点,且模型较小,参数量为 2.34×10^6,便于部署。

(4)本模型疏果前幼果检测的 AP_{50} 为 88.2%,平均召回率和精确率分别达到 78.5% 和 93.6%,优于原始 YOLO 11 网络模型。帧率也有显著提升,从 71.3 f/s 提升到 89.1 f/s。相较于经典模型与主流模型,AP_{50} 提升了 1.6 ~ 7.3 个百分点,AP_{75} 提升了 1.4 ~ 9.9 个百分点,AP_{50-95} 提升了 1.3 ~ 7.7 个百分点,展现出更强的特征提取和鲁棒性能。

参 考 文 献

[1] 张放. 2022 年我国水果生产变化简析[J]. 中国果业信息,2024,41(2):35-43.
Zhang Fang. Brief analysis of fruit production changes in China in 2022[J]. China Fruit News, 2024, 41(2): 35-43. (in Chinese)

[2] 张哲, 马斌, 罗娜, 等. 果蔬农产品产地智能化处理技术研究进展与展望[J]. 农业机械学报, 2025, 56(6): 1-16.
Zhang Zhe, Ma Bin, Luo Na, et al. Research progress and prospects of key technologies for intelligent treatment of fruits and vegetables agricultural products at origin[J]. Transactions of the Chinese Society for Agricultural Machinery, 2025, 56(6): 1-16. (in Chinese)

[3] 李振波,刘欣,竺明昊,等. 面向种植农业复杂场景的 YOLO 算法应用综述[J]. 农业机械学报,2026,57(13):45-67.
Li Zhenbo, Liu Xin, Zhu Minghao, et al. Survey of YOLO algorithm applications for complex scenarios in crop agriculture[J]. Transactions of the Chinese Society for Agricultural Machinery, 2026,57(13):45-67. (in Chinese)

[4] Ahmad M, Hameed A, Ullah F, et al. Adaptation of vehicular ad hoc network clustering protocol for smart transportation[J]. Computers, Materials & Continua, 2021, 67(2): 1353-1368.

[5] Zhang J, Karkee M, Zhang Q, et al. Multi-class object detection using Faster R-CNN and estimation of shaking locations for automated shake-and-catch apple harvesting[J]. Computers and Electronics in Agriculture, 2020, 173: 105384.

[6] Ren S, He K, Girshick R, et al. Faster R-CNN: towards real-time object detection with region proposal networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(6): 1137-1149.

[7] Zhang L, Wang M, Ding Y, et al. MS-FRCNN: a multi-scale Faster RCNN model for small target forest fire detection[J]. Forests, 2023, 14(3): 616.

[8] Jia W, Zhang Z, Shao W, et al. FoveaMask: a fast and accurate deep learning model for green fruit instance segmentation[J]. Computers and Electronics in Agriculture, 2021, 191: 106488.

[9] 朱立成, 王文贝,赵博,等. 基于 SDE-YOLO 的矮砧密植化果园苹果检测方法[J]. 农业机械学报,2025, 56(9): 638-647.
Zhu Licheng, Wang Wenbei, Zhao Bo, et al. Detection method of apples in dwarf rootstock densification orchards based on SDE-YOLO[J]. Transactions of the Chinese Society for Agricultural Machinery, 2025, 56(9): 638-647. (in Chinese)

[10] 宋怀波, 马宝玲, 尚钰莹, 等. 基于 YOLO v7-ECA 模型的苹果幼果检测[J]. 农业机械学报,2023, 54 (6): 233-242.
Song Huaibo, Ma Baoling, Shang Yuying, et al. Detection of young apple fruits based on YOLO v7-ECA model[J].

Transactions of the Chinese Society for Agricultural Machinery, 2023, 54(6): 233－242. (in Chinese)

[11] 莫恒辉, 魏霖静. 基于改进 YOLOv7 的复杂环境下苹果目标检测[J]. 浙江大学学报(工学版),2024, 58(1): 2447－2458.
Mo Henghui, Wei Linjing. Improved YOLOv7 based apple target detection in complex environment[J]. Journal of Zhejiang University (Engineering Science), 2024, 58(1): 2447－2458. (in Chinese)

[12] Wen S, Zhou J, Hu G, et al. PcMNet: an efficient lightweight apple detection algorithm in natural orchards[J]. Smart Agricultural Technology,2024,9:100623.

[13] Zhao B, Guo A, Ma R, et al. YOLOv8s－CFB: a lightweight method for real-time detection of apple fruits in complex environments[J]. Journal of Real-Time Image Processing, 2024, 21(5): 164.

[14] Han B, Lu Z, Zhang J, et al. Rep-ViG-Apple: a CNN－GCN hybrid model for apple detection in complex orchard environments[J]. Agronomy, 2024, 14(8): 1733.

[15] Jia W, Wang Z, Zhang Z, et al. A fast and efficient green apple object detection model based on Foveabox[J]. Journal of King Saud University-Computer and Information Sciences, 2022, 34(8): 5156－5169.

[16] Khanam R, Hussain M. YOLO11: an overview of the key architectural enhancements[PP/OL]. V4. arXiv(2022－07－11)[2025－08－25]. https://arxiv.org/abs/2203.03605.

[17] Han K, Wang Y, Tian Q, et al. GhostNet: more features from cheap operations[C]// Computer Vision and Pattern Recognition, 2020: 1580－1589.

[18] Howard A G, Zhu M, Chen B, et al. MobileNets: efficient convolutional neural networks for mobile vision applications[PP/OL]. V1. arXiv(2017－04－17)[2025－08－25]. https://arxiv.org/abs/1704.04861.

[19] Hu J, Shen L, Albanie S, et al. Squeeze-and-excitation networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2019, 42(8): 2011－2023.

[20] Sun Y, Meng X, Guo P, et al. Extracting the navigation center line for fishery complementary photovoltaic boat using improved YOLOv8n[J]. Transactions of the CSAE, 2024, 40(22): 173－182.

[21] Li X, Wang W, Wu L, et al. Generalized focal loss: learning qualified and distributed bounding boxes for dense object detection[J]. Advances in Neural Information Processing Systems, 2020, 33: 21002－21012.

[22] Tan M, Le Q V. EfficientNet: rethinking model scaling for convolutional neural networks[PP/OL]. V5. arXiv(2020－09－11)[2025－08－25]. https://arXiv.org/abs/1905.11946.

[23] Cai Z, Vasconcelos N. Cascade R－CNN: delving into high quality object detection[C]// In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 6154－6162.

[24] Zhang H, Li F, Liu S, et al. DINO: DETR with improved denoising anchor boxes for end-to-end object detection[PP/OL]. V4. arXiv(2022－07－11)[2025－08－25]. https://arxiv.org/abs/2203.03605.

[25] Tian Z, Shen C, Chen H, et al. FCOS: fully convolutional one-stage object detection[C]//2019 IEEE/CVF International Conference on Computer Vision. Seoul: IEEE, 2019: 9626－9635.

[26] Wang A, Chen H, Liu L, et al. YOLOv10: Real-time end-to-end object detection[PP/OL]. V2. arXiv(2024－10－30)[2025－08－25]. https://arxiv.org/abs/2405.14458.

[27] Feng Y, Huang J, Du S, et al. Hyper－YOLO: when visual object detection meets hypergraph computation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2025, 47(4): 2388－2401.