doi:10.6041/j.issn.1000-1298.2026.05.016

基于 RGB - D 相机的苹果疏果期幼果与果柄识别定位方法

陈金成[1,2]　潘　峰[1,2]　王佰伟[1,2]　张　景[1,2]　纪　超[1,2]

(1. 新疆农垦科学院机械装备研究所, 石河子 832000; 2. 农业农村部西北农业装备重点实验室, 石河子 832000)

摘要: 苹果疏果作业中,幼果尺寸量化判断与果柄精准疏除位置定位是制约智能化疏除作业的核心难题,而以幼果轴向、径向尺寸及果柄直径为依据的智能化疏除技术亟待突破。为此,提出了一种基于 RGB - D 融合几何约束的苹果疏果期幼果及果柄三维识别定位与尺寸测量方法。首先,通过在 YOLO v8 模型中引入卷积注意力融合模块(CAFM)与 SIoU 损失函数,构建 YOLO - YF 检测模型,并结合 Mask R - CNN 实现幼果及果柄的高精度分割。其次,利用 D435i 型深度相机完成 RGB - D 图像对齐与三维坐标转换,建立幼果与果柄的空间从属关系判别算法。实验结果表明:YOLO - YF 模型在幼果簇检测中实现 86.80% 的精确率、78.70% 的召回率及 84.40% 的 mAP50;Mask R - CNN 精确率为 91.20%,交并比为 80.94%,可有效区分果柄与叶柄;在三维尺寸测量中,幼果径向、轴向尺寸均方根误差分别为 1.43、1.28 mm,果柄疏除位置定位误差为 1.20 mm。本研究为智能化疏果装备提供了"尺寸量化-位置定位"的技术路径。

关键词: 苹果; 疏果期; RGB - D; 识别定位; 三维测量

中图分类号: TP391.4　　**文献标识码**: A　　**文章编号**: 1000-1298(2026)05-0177-09　　**OSID**:

Recognition and Localization Method of Young Fruit and Fruit Stalk in Apple Fruit Thinning Period Based on RGB - D Cameras

CHEN Jincheng[1,2]　PAN Feng[1,2]　WANG Baiwei[1,2]　ZHANG Jing[1,2]　JI Chao[1,2]

(1. *Mechanical Equipment Research Institute*, *Xinjiang Academy of Agricultural and Reclamation Science*, *Shihezi* 832000, *China*
2. *Key Laboratory of Northwest Agriculture Equipment*, *Ministry of Agriculture and Rural Affairs*, *Shihezi* 832000, *China*)

Abstract: Accurate quantification of young fruit dimensions and precise localization of thinning points on fruit stalks represented core challenges for intelligent apple fruit thinning. To address this, a 3D recognition and sizing method integrating RGB - D data with geometric constraints was developed for young apples and fruit stalks during the thinning period. The YOLO v8 architecture was enhanced through integration of a convolutional attention fusion module (CAFM) and adoption of the SIoU loss function, yielding the proposed YOLO - YF detection model. Mask R - CNN was subsequently employed for high-precision instance segmentation of young fruits and fruit stalks. RGB - D image acquisition was performed by using an Intel RealSense D435i depth camera, followed by rigorous image alignment and 3D coordinate transformation. A spatial subordination discrimination algorithm based on geometric features was established to determine fruit stalk relationships. Experimental validation demonstrated that the YOLO - YF model achieved 86.80% precision, 78.70% recall, and 84.40% mAP50 in clustered young fruit clusters detection. Mask R - CNN attained 91.20% segmentation precision and 80.94% IoU, enabling reliable differentiation between fruit stalks and petioles. In 3D measurements, root mean square errors (RMSE) of 1.43 mm (radial) and 1.28 mm (axial) were obtained for young fruits. The positioning error at the optimal thinning point (fruit stalk midpoint) was constrained to 1.20 mm. This approach can provide a technical pathway characterized by quantitative sizing and millimeter-level positioning accuracy for intelligent thinning equipment. The methodology demonstrated extensibility to clustered fruit management applications.

Key words: apple; fruit thinning period; RGB - D; recognition and localization; three-dimensional measurement

收稿日期: 2025 - 07 - 10　修回日期: 2025 - 07 - 18

基金项目: 新疆生产建设兵团科技合作计划项目(2022BC007)和兵团英才青年项目

作者简介: 陈金成(1986—),男,副研究员,主要从事农业智能化检测与控制研究,E-mail: shznkycjc@163.com

通信作者: 纪超(1986—),男,研究员,主要从事大田智能农业装备关键技术及系统集成研究,E-mail: jicobear@163.com

0 引言

疏果作为苹果果园精细化管理的核心环节,其作业质量直接决定果实商品性指标及树体次年挂果能力[1]。当前生产中,化学疏果受气候、树势等因素影响显著,药剂浓度控制不当易导致疏除过度或不足,且存在农药残留风险;人工疏果虽能定向保留优质幼果,但面临劳动强度大、成本高(占果园管理成本 20% ~30%)及劳动力短缺等现实困境[2]。随着矮砧密植果园规模化发展,以幼果位置及尺寸为判断依据,实现疏果作业的智能化精准操作,已成为现代苹果产业提质增效的必然需求。

近年来,深度学习技术在图像识别领域取得了显著进展,并在农业领域得到了广泛应用。YOLO 系列[3-5]、Mask R-CNN[6]等深度学习框架及其改进的模型[7-10],在苹果[11-13]、葡萄[14-15]、脐橙[16-18]等采摘机器人的研究中得到了广泛的应用。尽管针对苹果疏果期幼果的识别研究相对较少,但研究者已经针对幼果体积小、颜色与叶片差异不明显、易受枝条或其他果实遮挡等特点,开展了针对性的研究。文献[19-20]通过融合注意力机制、改进损失函数及轻量化网络等方法,分别对疏果期幼果的检测与计数、苹果目标检测以及幼果生长角度预测等进行了研究。这些研究和应用表明,多尺度特征融合和注意力机制等是提升苹果幼果识别性能的重要方法,它们能够帮助模型更好地理解图像内容,区分目标与背景,并适应目标尺度的变化。

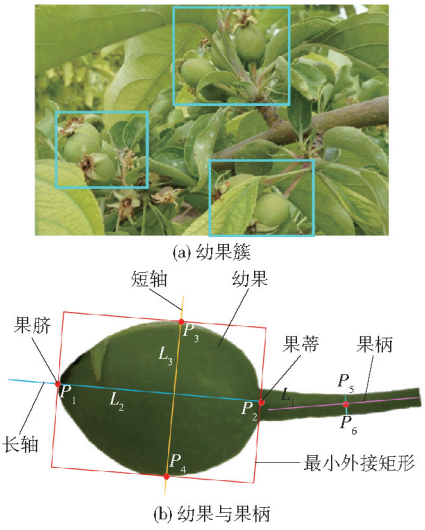
尽管苹果疏果期幼果识别定位技术在深度学习模型的推动下已经取得了一定的进展,但现有研究在处理一些关键性挑战时仍存在不足。一是当多个幼果紧密生长在一起时,模型难以准确分割和识别每一个被严重遮挡的幼果个体;二是由于果柄和叶柄在形态和颜色上高度相似,现有模型的区分精度仍有待进一步提高;三是如何准确判断每个果柄所对应的幼果,目前尚缺乏有效的自动化判别手段。这些局限性直接影响了自动化疏果系统在实际复杂果园环境中的鲁棒性和实用性。同样,如果不能精确区分果柄和叶柄,或者在判断果柄从属关系时出现错误,就可能导致错误的疏除操作,损伤果树或保留劣果。

针对上述难题,本文融合多尺度检测与几何约束,开展疏果期幼果及果柄的识别与定位方法研究,旨在突破粘连目标分割、从属关系建模及微细结构(如果柄)精准定位和尺寸估算等瓶颈。为此,构建跨模态特征融合框架并引入先验几何规则,进一步提升在复杂遮挡和尺度剧变场景下的鲁棒性,以期为智能化疏果装备提供技术支撑。

1 材料与方法

1.1 苹果簇、幼果、果柄形态及识别定位策略分析

本研究以疏果期幼果及其果柄为研究对象(图 1b),将幼果轴向尺寸 L_2、径向尺寸 L_3及果柄直径 $l_{P_5P_6}$作为衡量指标。由于现行园艺缺乏明确的果柄疏除位置,综合考虑机械作业空间适配性,将果柄中点的径向区域 P_5P_6确定为疏除位置。



(a) 幼果簇

(b) 幼果与果柄

图 1 疏果期幼果簇、幼果及果柄

Fig. 1 Young fruit clusters, young fruits and fruit stalks during fruit thinning period

针对幼果密集簇生遮挡、果柄纤细的作业难点,提出分层级识别策略。首先应用深度相机(RGB-D)在 1.0 ~3.0 m 远视场范围内进行快速扫描,识别并定位苹果簇分布区域(图 1a);随后移动至 0.3 ~1.0 m 近视场,融合 RGB-D 图像和深度信息实现幼果与果柄的实例分割及几何参数测量,为自动化疏除提供精准的空间拓扑关系与尺寸数据。该策略通过“远视场全局定位-近视场精细测量”的协同,解决了单一视场下“成像清晰但覆盖不足”或“范围全面但细节缺失”的矛盾。

1.2 数据集构建

本研究的数据采集工作在新疆维吾尔自治区石河子市 152 团苹果园开展。为构建适用于深度学习模型、兼具多样性与丰富信息量的图像数据集,整合了 3 种成像设备:Redmi Note 9 Pro 型智能手机(378 幅)、SONY ILCE-6000 型数码相机(418 幅)以及主要采集设备 Intel D435i 型深度相机(838 幅),共采集 1 634 幅苹果疏果期图像。在数据划分阶段,选取涵盖各种关键和挑战场景 50 幅图像作为测试集(含 203 个幼果簇及 532 个独立目标),剩余图像按

照比例 8∶2划分为原始训练集与验证集。此外，综合运用旋转、翻转、缩放、裁剪、色彩调整等多种技术手段，将训练集扩充至 4 085 幅图像，并借助 LabelImg 和 Labelme 工具，分别对幼果簇进行矩形标注用于目标检测，对幼果及果柄进行多边形标注以实现分割模型的训练。

1.3　苹果簇检测模型

YOLO v8 是 Ultralytics 公司开发的先进目标检测算法，它在检测速度和精度上实现了良好平衡，可同时完成图像中物体类别识别与边界框定位。但在苹果幼果簇检测场景中，该算法性能受果实高度重叠、复杂光线下色彩混淆以及幼果形态动态变化等因素影响。为此，本研究通过改进 YOLO v8 提出了 YOLO－Young Fruit（YOLO－YF）模型（图 2）。该模型整体由 Backbone、Neck 和 Head 3 部分构成。该模型在 Backbone 末端引入卷积注意力融合模块（Convolution and attention fusion module，CAFM）[23]，借助空间与通道注意力机制的协同强化，增强对重叠幼果簇特征的分辨能力；Neck 部分通过特征融合流程实现不同尺度特征的整合；Head 部分则承担检测头的分工任务。此外，将原有的 CIoU 损失函数替换为 SIoU 损失函数，进一步提升检测框的定位准确性，以此解决 YOLO v8 在苹果簇检测应用中存在的精度不足问题。

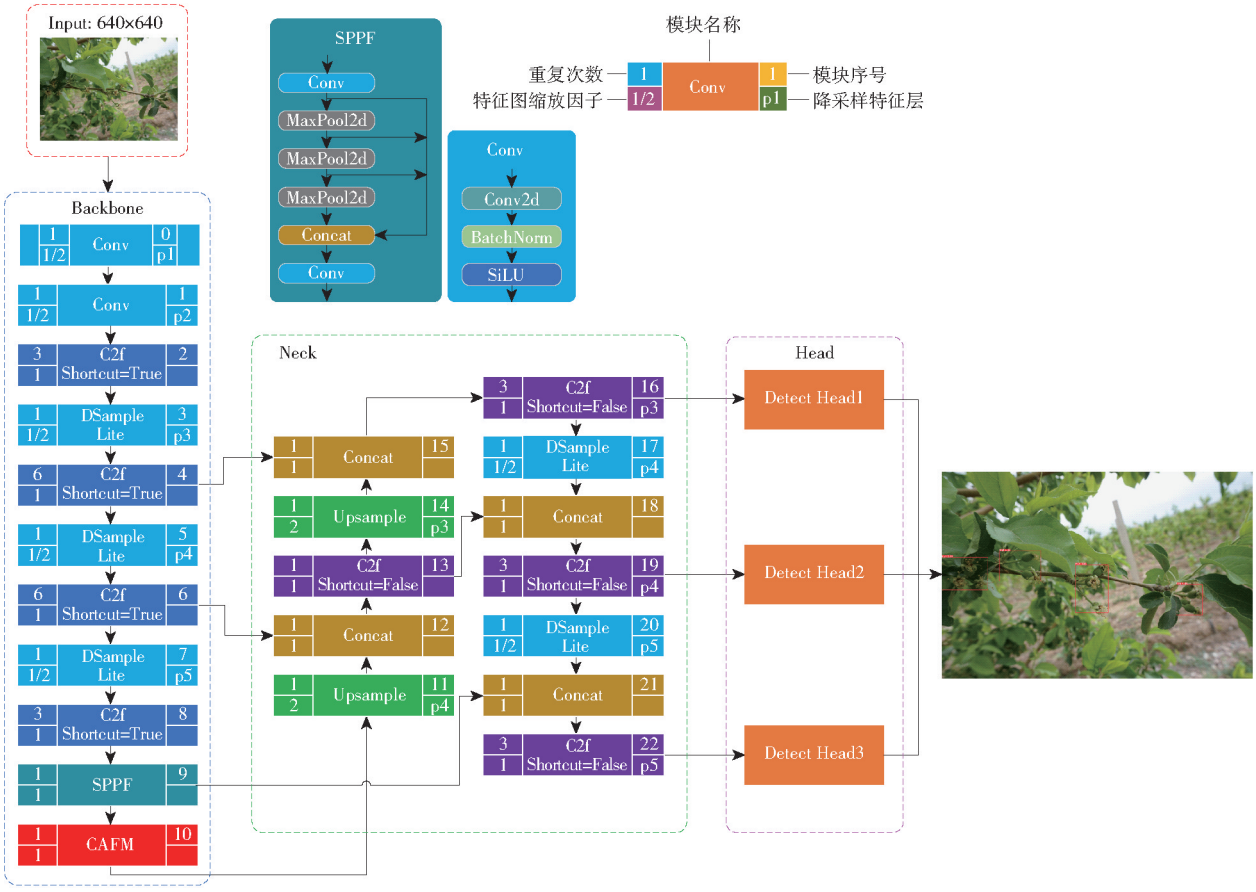


图 2　YOLO－YF 模型结构图

Fig. 2　Structure of YOLO－YF model

1.3.1　CAFM 注意力融合模块

CAFM 模块（图 3）由局部分支和全局分支组成，分别负责精细化特征处理和广泛上下文信息的捕获，克服了传统卷积网络仅针对局部特征提取的局限性。局部特征提取分支致力于挖掘和强化局部特征，首先使用 1×1 卷积调整输入特征图的通道数；接着，引入通道混洗操作；最后，应用 3×3×3 卷积层在空间维度上进一步细化和提取局部特征。其公式为

$$F_{Conv} = W_{3\times3\times3}(C_S(W_{1\times1}(X))) \tag{1}$$

式中　F_{Conv}——局部分支的输出

$W_{3\times3\times3}$——3×3×3 卷积的权重

$W_{1\times1}$——1×1 卷积的权重

C_S——通道混洗操作

X——输入特征图

全局注意力提取分支利用注意力机制来捕捉全局上下文特征，其构建模块是经典的 Transformer 架构[24]中的查询 Q、键 K 和值 V 机制。

1.3.2　SIoU 损失函数

SIoU 是一种用于解决目标检测任务的损失函

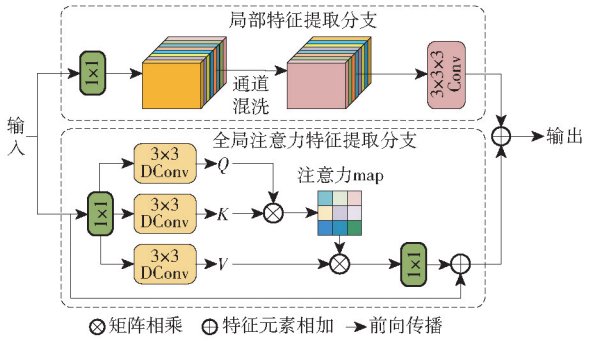


图 3 CAFM 网络模块结构图

Fig. 3 CAFM network module structure

数。它将预测框和基准真值框的角度、距离、形状和交点除以交并比(IoU)作为损失函数中的因子。在真值框和预测框之间建立角度匹配损失模型。图 4 中的真值框和预测框之间的几何关系定义了角度损失函数 Λ,公式为

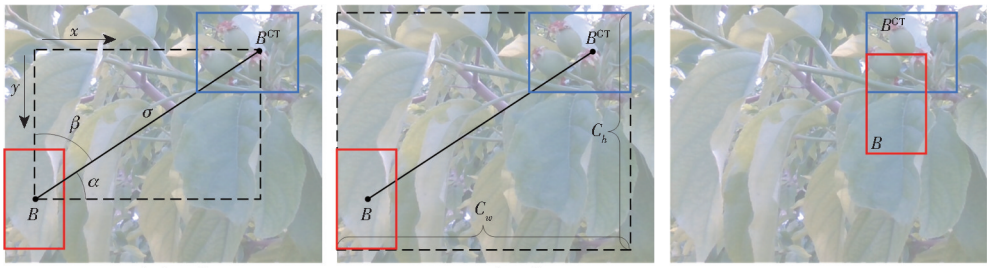
$$\Lambda = 1 - 2\sin^2\left(\alpha - \frac{\pi}{4}\right) \tag{2}$$

其中 $$\alpha = \arcsin\left(\frac{\max(b_{c_y}^{gt}, b_{c_y}) - \min(b_{c_y}^{gt}, b_{c_y})}{\sqrt{(b_{c_x}^{gt} - b_{c_x})^2 + (b_{c_y}^{gt} - b_{c_y})^2}}\right) \tag{3}$$

式中 $(b_{c_x}^{gt}, b_{c_y}^{gt})$——真实框中心 B^{GT} 的坐标

(b_{c_x}, b_{c_y})——预测框中心 B 的坐标

通过计算预测框的宽度与真实框的宽度之差来



(a) 角度函数 (b) 距离函数 (c) IoU

图 4 离散损失函数的几何表示

Fig. 4 Geometric representation of discrete loss functions

定义形状损失函数。此外,计算它们之间最大宽度的比率 Ω,以实现整体形状收敛。该模型定义为

$$\Omega = \sum_{t=w,h} (1 - e^{-\omega_t})^{\theta} \tag{4}$$

其中 $$\omega_w = \frac{|w - w^{gt}|}{\max(w, w^{gt})} \quad \omega_h = \frac{|h - h^{gt}|}{\max(h, h^{gt})}$$

式中 w、h——预测框宽、高

w^{gt}、h^{gt}——真实边界框宽、高

θ——对形状损失的关注程度

SIoU 损失函数为

$$S_{IoU} = 1 - IoU + \frac{\Lambda + \Omega}{2} \tag{5}$$

1.4 苹果幼果及果柄分割模型

针对苹果疏果作业中果柄尺寸微小且与叶柄形态相似的技术难点,本研究采用 Mask R-CNN 网络架构以实现高精度分割。相较于单阶段分割模型(如 YOLACT[25] 等),Mask R-CNN 作为典型的双阶段模型,由特征提取骨干网络(如 ResNet)、区域建议网络(RPN)、感兴趣区域对齐(RoIAlign)和分割掩模预测 4 个核心模块构成。该架构首先通过骨干网络提取多尺度特征,随后 RPN 生成候选区域,再经 RoIAlign 保持特征空间精度,最终完成目标分类、边界框回归和像素级分割。

1.5 模型训练

实验环境配置为 RTX 4090 GPU,内存容量为 16 GB,操作系统为 Ubuntu 20.04。模型的训练,采用 AdamW 优化器,学习率为 0.001,权重衰减系数为 0.000 5。同时,对模型进行了最多可达 500 次的训练周期。在训练过程中,若连续 50 次迭代未观察到明显的性能提升,则自动终止训练进程。

1.6 幼果与果柄从属关系判断

为建立幼果与果柄的从属关系,对于无遮挡场景(图 5a)通过果柄生长方向(主成分分析确定)的虚拟延伸线计算其与邻近幼果的交点及距离,选择距离果柄质心最近的幼果作为从属对象。在遮挡严重的情况下(图 5b、5c),由于多果柄可能与同一幼果相交,故引入幼果形态学特征辅助判断:图 5d~5f 中果柄①为真从属,由于果柄①延伸部分同幼果及其最小外接矩形的交点非常接近(近似重合),在此用 $l_{P_oP_{o'}}$ 来表示其长度;果柄②为非从属果柄,其与幼果边缘及其最小外接矩形相交的两个点 P_1 和 P_2 具有明显的分离,且明显大于果柄①与幼果之间紧密接近的距离。因此,通过量化交点与矩形边的距离差异($l_{P_oP_{o'}} < l_{P_1P_2}$),可有效区分从属关系。

1.7 像素点对应的三维坐标

本研究采用 Intel RealSense D435i 型深度相机获取苹果幼果及果柄的三维空间信息。深度相机(视场 86°×57°)与 RGB 相机(视场 69.4°×42.5°)需通过几何对齐实现两模态数据的像素级匹配。确保 RGB 图像中的每一个像素位置都能与深度图像中的相应像素点在空间上准确对应,从而实现深度

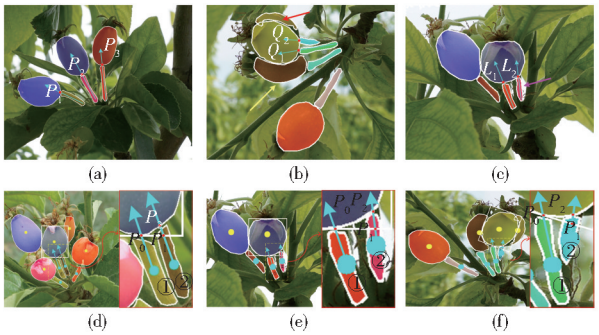


图 5　多个果柄与单个幼果相交的几何关系

Fig. 5　Multiple stems and single young fruit intersect geometrically

信息的精确匹配。不仅能够保留全部 RGB 图像的细节，还能为每个像素赋予深度维度。算法实现如下：

Local Variables：

1：depth_image：▷深度图

2：rgb_image：　▷RGB 图

3：intrinsic_depth：▷深度相机内参数

4：intrinsic_rgb：　▷RGB 相机内参数

5：extrinsic_transform：▷相机外参数

Initialize：

6：aligned_depth ← createArrayWithSameDimensionsAs(rgb_image)

7：width, height ← getSizeOf(rgb_image)

8：while y in range(height)：

9：　while x in range(width)：

10：　normalized_uv ← convertPixelCoordToNormalized(x, y, intrinsic_rgb)

11：　transformed_uv ← applyExtrinsicTransform(normalized_uv, extrinsic_transform.inverse())

12：　depth_val ← interpolateDepthValue(transformed_uv.u, transformed_uv.v, depth_image)

13：　aligned_depth[x, y] ← depth_val

通过上述对齐算法，能够得到与 RGB 对齐后的深度图，进而给每个像素点赋予深度。进行像素坐标(u,v)与世界坐标(x_w, y_w, z_w)之间转换的公式为

$$\begin{bmatrix} x_w \\ y_w \\ z_w \\ 1 \end{bmatrix} = \begin{bmatrix} \boldsymbol{R} & \boldsymbol{T} \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \frac{z_c d_x}{f} & 0 & \frac{-z_c c_x d_x}{f} \\ 0 & \frac{z_c d_y}{f} & \frac{-z_c c_y d_y}{f} \\ 0 & 0 & z_c \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} \tag{6}$$

式中　$\boldsymbol{R}$——相机的外参数旋转矩阵

$\boldsymbol{T}$——相机的外参数平移矩阵

f——相机焦距

c_x——像素坐标系下原点的 x 方向坐标

c_y——像素坐标系下原点的 y 方向坐标

d_x——图像传感器上像素尺寸的水平尺寸

d_y——图像传感器上像素尺寸的垂直尺寸

z_c——像素对应的场景深度

此外，D435i 型相机在出厂前已完成标定，并提供 API 接口用于获取内参。若图像中某像素坐标点(u,v)的深度为 z_c，则可计算出实际的三维坐标(x, y, z)，即

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 915.633 & 0 & -642.72 \\ 0 & 913.61 & -362.964 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} z_c \tag{7}$$

2　结果与分析

2.1　苹果簇检测效果评估

2.1.1　消融实验

本研究通过系统性的消融实验验证了 CAFM 模块和不同损失函数对模型性能的影响。实验结果如表 1 所示。原始 YOLO v8 模型（CIoU 损失）的 mAP50（当预测框与真实框的交并比（IoU）达到或超过 0.5 时，模型的平均精度）为 78.70%，F1 值为 75.55%。仅替换损失函数时，Shape－IoU 在精确率（78.30%）上略有提升，但 mAP50 有所下降，而 SIoU 显著提升 mAP50 至 79.0%，但 F1 值有所降低，说明单一损失函数调整难以平衡性能。引入 CAFM 模块后，模型性能得到改善：仅嵌入 CAFM 时，精确率提升至 84.10%，mAP50 增至 79.40%；结合 Shape－IoU 后，精确率进一步提高至 88.20%，mAP50 达 83.70%，但召回率略有下降。最优配置（CAFM＋SIoU）在各项指标上均表现最佳，mAP50 提升至 84.40%，F1 值达 82.55%，召回率提高至 78.70%，表明 CAFM 模块与 SIoU 损失函数的协同优化，能有效增强模型对目标的检测能力。

图 6 展示了不同损失函数的训练动态分析，SIoU 收敛最快（116 轮），Shape－IoU 最慢（207 轮），而 CIoU 居中（142 轮），验证了损失函数设计对优化效率的影响。加入 CAFM 模块后，Shape－IoU 的收敛轮次大幅减少（99 轮），且验证损失的波动有所降低，表明 CAFM 能提升模型泛化性。SIoU 在验证集上损失最低，结合 CAFM 后进一步优化了检测性能，说明该组合在边界框回归和特征提取方面具有优势。

2.1.2　对比实验

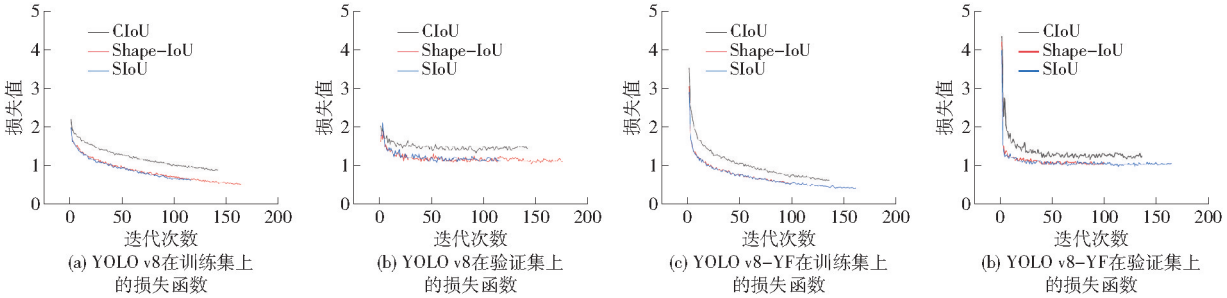
从表 2 的定量指标来看，YOLO－YF 在苹果幼果簇检测任务中展现明显优势。其精确率为 86.80%，召回率为 78.70%，对幼果簇的捕捉优于

表 1 改进组件消融实验效果评估

Tab. 1 Improvement of component ablation test effect evaluation %

模型	CAFM	损失函数			精确率	召回率	mAP50	F1 值
		CIoU	Shape－IoU	SIoU				
YOLO v8－S		√			77.60	73.60	78.70	75.55
模型 1			√		78.30	77.10	76.50	77.70
模型 2				√	78.10	74.10	79.00	76.05
模型 3	√	√			84.10	70.60	79.40	76.76
模型 4	√		√		88.20	72.10	83.70	79.34
YOLO－YF	√			√	86.80	78.70	84.40	82.55

注：√ 表示模块被包含在模型中。



(a) YOLO v8在训练集上的损失函数 (b) YOLO v8在验证集上的损失函数 (c) YOLO v8-YF在训练集上的损失函数 (b) YOLO v8-YF在验证集上的损失函数

图 6 训练与验证过程中损失函数趋势

Fig. 6 Trend of loss function during training and validation

表 2 不同模型的性能评估

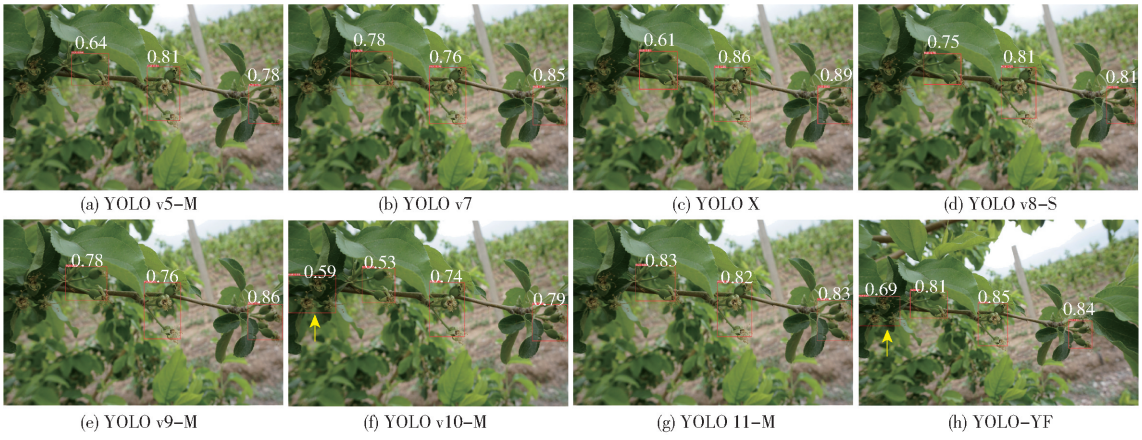
Tab. 2 Performance evaluation of different models %

模型	精确率	召回率	mAP50
YOLO v5－M	73.40	70.60	73.70
YOLO v7	76.10	75.60	76.30
YOLO X	73.50	75.10	76.40
YOLO v8－S	77.60	73.60	78.70
YOLO v9－M	84.10	69.80	80.20
YOLO v10－M	81.50	71.60	81.30
YOLO 11－M	82.10	77.70	84.00
YOLO－YF	86.80	78.70	84.40

对比模型。对比其他模型，YOLO－YF 在精确率、召回率协同提升下，体现其通过卷积注意力融合模块与 SIoU 损失函数优化，在复杂果园场景中对幼果簇检测的精准性与全面性。在检测效果上(图 7)，原始图像包含 4 个幼果簇，仅有 YOLO－YF 和 YOLO v10 实现了全部检出，YOLO－YF 具有较高的置信度，体现了其更强的目标判别能力，其他对比模型存在漏检。

2.2 幼果与果柄实例分割效果评估

表 3 结果表明，Mask R－CNN 展现出显著优势，其精确率达到 91.20%，召回率为 87.80%，交并比为 80.94%，Dice 系数为 89.47%。证明了 Mask R－CNN 在精细分割任务中的优越性。MaskLab 的表现仅次于 Mask R－CNN，其精确率(86.20%)和召回率(78.50%)均保持较高水平。相比之下，YOLO v8－m 的精确率(84.90%)虽优于其他单阶段模型，但其召回率和分割指标明显低于 Mask R－



(a) YOLO v5-M (b) YOLO v7 (c) YOLO X (d) YOLO v8-S

(e) YOLO v9-M (f) YOLO v10-M (g) YOLO 11-M (h) YOLO-YF

图 7 不同模型的检测效果

Fig. 7 Detection effects of different models

CNN；YOLACT 在复杂场景下表现欠佳；FastInst 虽引入注意力机制提升多尺度目标检测能力，但仍存在漏检和误检问题。

表 3　不同网络模型的性能

Tab. 3　Performance of different network models　%

模型	精确率	召回率	交并比	Dice 系数
YOLO v8－m	84. 90	80. 60	70. 49	82. 69
YOLACT	76. 50	68. 70	56. 73	72. 39
FastInst	79. 10	72. 30	60. 70	75. 55
MaskLab	86. 20	78. 50	69. 74	82. 17
Mask R－CNN	91. 20	87. 80	80. 94	89. 47

图 8 展示了各模型的分割效果。Mask R－CNN 在分割区域完整性和边缘细节处理上表现最优，其两阶段检测机制提升了边界分割的精确度。相比之下，YOLO v8－m 在处理细小果柄时出现锯齿状边缘，YOLACT 存在较高漏检率，FastInst 则在复杂场景下误检率较高。此外，YOLO v8－m、YOLACT 和 FastInst 均存在将叶柄误分类为果柄的问题，反映出单阶段模型在区分形态相似物体时的局限性。综合实验结果，Mask R－CNN 凭借其精细的分割能力，成为幼果及果柄识别定位任务的最优选。



(a) 原图　(b) YOLO v8－m　(c) YOLACT　(d) FastInst　(e) MaskLab　(f) Mask R－CNN

图 8　幼果及果柄分割对比效果

Fig. 8　Qualitative contrast effect of young fruit and fruit stem

2. 3　幼果及果柄从属算法判断评估

为验证幼果与果柄从属关系判断算法的准确性，本研究在选取近视场下 50 幅图像作为测试样本。实验结果表明（表 4），该算法整体判断准确率为 93. 7%，其中无遮挡场景的判断准确率为 97. 5%，仅因少数果柄断裂样本出现误判；遮挡场景的判断准确率为 88. 7%，误判主要源于叶柄与果柄形态相似性导致的干扰。该算法在果柄与幼果清晰分离场景（图 9a、9d）及相互遮挡的复杂场景（图 9b、9c、9e、9f）中均保持稳定性能；在数量不对等或形态相似性干扰（图 9c 青色箭头区域）的情况下，仍能通过 PCA 优化的延伸线距离匹配真实从属关系。

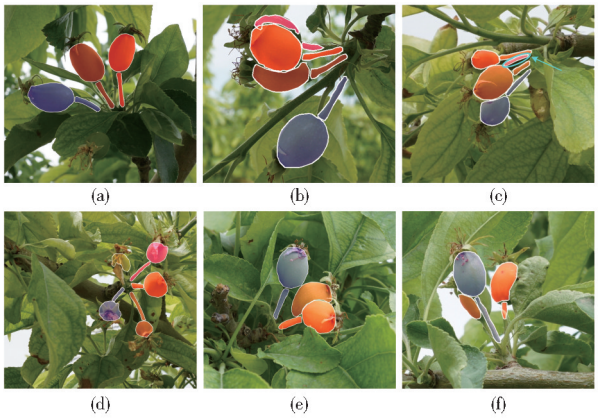
表 4　幼果及果柄从属关系判别准确率

Tab. 4　Accuracy rate of discrimination of relationship between young fruits and fruit stalks

场景	样本组数	正确组数	准确率/%
整体	273	256	93. 7
无遮挡	158	154	97. 5
遮挡	115	102	88. 7

2. 4　幼果及果柄三维尺寸评估

选取 30 幅包含深度信息的 RGB 图像作为基准样本，并与人工测量数据对比以评估算法效果。实验结果（表 5 和图 10）显示，幼果尺寸估计的成功率达到 100%，径向尺寸均方根误差（RMSE）为 1. 43 mm、平均绝对误差（MAE）为 1. 32 mm，轴向尺寸 RMSE 为 1. 28 mm、MAE 为 1. 17 mm，表明该方法对幼果尺寸测量具有较高的准确性。果柄直径估计的成功率为 90%，其中，3 个目标因深度信息丢失导致测量失败（图 10 红色圆圈标注），RMSE 为 1. 35 mm、MAE 为



(a)　(b)　(c)　(d)　(e)　(f)

图 9　幼果及果柄从属关系的判断效果

Fig. 9　Effect of determining subordination of young fruit and stalk

1.20 mm。尽管果柄测量的成功率略低于幼果，但整体误差仍在可接受范围内。

表 5 不同测量部位的指标评估

Tab. 5 Evaluation of indicators at different measurement sites mm

类别	均方根误差	平均绝对误差	标准差
幼果径向尺寸	1.43	1.32	1.20
幼果轴向尺寸	1.28	1.17	1.75
果柄直径	1.35	1.20	0.61

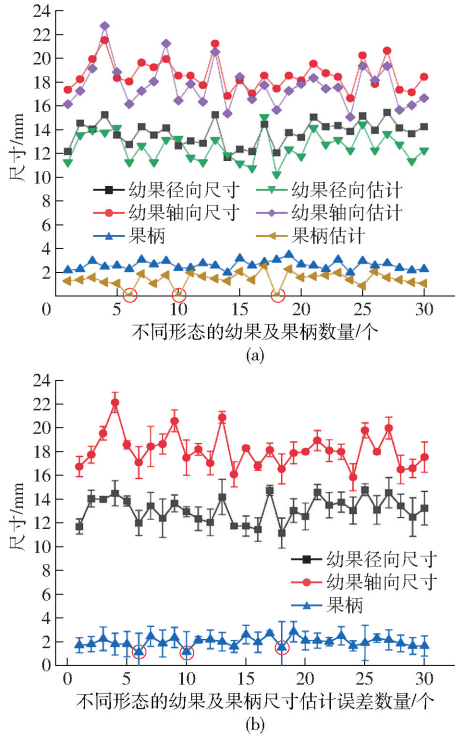


图 10 幼果及果柄三维尺寸评估

Fig. 10 Three-dimensional size assessment of young fruits and fruit stalks

此外，果柄直径估计中出现的深度信息丢失，主要是由于果柄细小的几何形态引发的“空洞现象”，即光线散射、相机分辨率限制或局部遮挡导致深度值缺失。此外，幼果轴向估计值的标准差为 1.75 mm，反映出幼果生长形态变异或环境扰动（如风力）对测量稳定性的影响。果柄尺寸 RMSE 与 MAE 分别为 1.35 mm 和 1.20 mm，但因果柄真实尺寸微小（通常 2 ~ 4 mm），其相对误差高达 33.8% ~ 60%，远超幼果（相对误差 6.4% ~ 7.2%）。根本问题在于果柄分割依赖 RGB 图像边缘检测，而细小目标的像素级分割偏差在三维重建中被显著放大，表明现有分割算法对边缘细节的捕捉能力还需进一步提升。

3 结论

（1）提出的 YOLO - YF 模型通过 CAFM 与 SIoU 损失函数的协同优化，在苹果幼果簇检测任务中实现了精确率（86.80%）与召回率（78.70%）的均衡提升，优于主流 YOLO 系列模型。

（2）采用 Mask R - CNN 实现幼果及果柄的实例分割。基于几何特征与空间分布规律的从属关系判定算法，整体判断准确率为 93.7%，为疏除点定位提供了拓扑支撑。

（3）基于 D435i 型深度相机测定的幼果径向/轴向尺寸均方根误差分别为 1.43 mm/1.28 mm；果柄直径均方根误差为 1.35 mm，虽受限于目标微小性导致的深度空洞现象，但绝对误差已可支撑机械疏除的位置定位。

参 考 文 献

[1] 唐金，马友福，卢磊，等. 伊犁河谷苹果化学疏花疏果效果初探[J]. 东北农业科学，2024，49(4)：41 - 46.
TANG Jin，MA Youfu，LU Lei，et al. Preliminary study on the effect of chemical flower and fruit thinning of apple in the Ili River valley[J]. Journal of Northeast Agricultural Sciences，2024，49(4)：41 - 46. (in Chinese)

[2] 赵保平，毋万来，陈永利. 一种便携式果树疏花疏果机械[J]. 西北园艺(果树)，2017(2)：50 - 51.

[3] WANG C Y，YEH I H，LIAO H Y M. YOLOv9：learning what you want to learn using programmable gradient information[J/OL]. https://arxiv.org/abs/2402.13616.

[4] WANG C Y，ALEXEY B，LIAO H Y M. YOLOv7：trainable bag-of-freebies sets new state-of-the-art for real-time object detectors[C]//IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)，2023：7464 - 7475.

[5] GE Z，LIU S，WANG F，et al. YOLOX：exceeding YOLO series in 2021[J/OL]. https://arxiv.org/abs/2107.08430.

[6] HE K，GKIOXARI G，DOLLÁR P，et al. Mask R - CNN[C]//2017 IEEE International Conference on Computer Vision (ICCV)，2017：2980 - 2988.

[7] 李翠明，杨柯，申涛，等. 基于改进 Faster R - CNN 的苹果采摘视觉定位与检测方法[J]. 农业机械学报，2024，55(1)：47 - 54.
LI Cuiming，YANG Ke，SHEN Tao，et al. Vision detection method for picking robots based on improved Faster R - CNN[J]. Transactions of the Chinese Society for Agricultural Machinery，2024，55(1)：47 - 54. (in Chinese)

[8] 李茂，肖洋轶，宗望远. 基于改进 YOLO 11 的多尺度板栗果实识别方法[J]. 农业机械学报，2025，56(5)：443 - 454.
LI Mao，XIAO Yangyi，ZONG Wangyuan. Multi-scale chestnut detection method based on improved YOLO 11 model[J]. Transactions of the Chinese Society for Agricultural Machinery，2025，56(5)：443 - 454. (in Chinese)

[9] 张震, 周俊, 江自真, 等. 基于改进 YOLO v7 轻量化模型的自然果园环境下苹果识别方法[J]. 农业机械学报,2023, 55(3): 231－242.
ZHANG Zhen, ZHOU Jun, JIANG Zizhen, et al. Lightweight apple recognition method in natural orchard environment based on improved YOLO v7 model[J]. Transactions of the Chinese Society for Agricultural Machinery, 2023, 55(3): 231－242. (in Chinese)

[10] LIANG Y, LIU Y, LIN Y, et al. Real-time citrus picking point localization guided by joint learning network[J]. Journal of Image and Graphics, 2024, 29(10): 3130－3143.

[11] 岳有军, 漆潇, 赵辉, 等. 基于改进 YOLOv8 的果园复杂环境下苹果检测模型研究[J]. 南京信息工程大学学报:自然科学版,2025, 17(1): 31－41.
YUE Youjun, QI Xiao, ZHAO Hui, et al. Apple detection in complex orchard environments based on improved YOLOv8[J]. Journal of Nanjing University of Information Science and Technology: Natural Science Edition, 2025, 17(1): 31－41. (in Chinese)

[12] JIANG Z, ZHOU J, HAN H, et al. An apple fruit localization system based on accurate and flexible hand-eye pose acquisition for robotic harvesting[J]. Computers and Electronics in Agriculture, 2025, 229: 109705.

[13] HUANG W, MIAO Z, WU T, et al. Design of and experiment with a dual-arm apple harvesting robot system[J]. Horticulturae, 2024, 10(12): 1268.

[14] COLL-RIBES G, TORRES-RODRÍGUEZ I J, GRAU A, et al. Accurate detection and depth estimation of table grapes and fruit stalks for robot harvesting, combining monocular depth estimation and CNN methods[J]. Computers and Electronics in Agriculture, 2023, 215: 108362.

[15] JIN Y, LIU J, WANG J, et al. Far-near combined positioning of picking-point based on depth data features for horizontal-trellis cultivated grape[J]. Computers and Electronics in Agriculture, 2022, 194: 106791.

[16] ZHOU B, WU K, CHEN M. Detection of gannan navel orange ripeness in natural environment based on YOLOv5－NMM[J]. Agronomy, 2024, 14(5):910.

[17] LI J, LU Y, LU R. Detection of early decay in navel oranges by structured-illumination reflectance imaging combined with image enhancement and segmentation[J]. Postharvest Biology and Technology,2023, 196: 112162.

[18] 王元昊,娄欢欢,罗红品,等. 基于改进 YOLOv8 算法对被遮挡柑橘的识别与定位优化[J]. 西南大学学报(自然科学版),2025, 47(2): 171－183.
WANG Yuanhao, LOU Huanhuan, LUO Hongpin, et al. Optimization of identification and localization of occluded citrus based on improved YOLOv8 algorithm[J]. Journal of Southwest University (Natural Science Edition), 2025,47(2):171－183. (in Chinese)

[19] 龙燕, 杨智优, 何梦菲. 基于改进 YOLOv7 的疏果期苹果目标检测方法[J]. 农业工程学报,2023, 39(14): 191－199.
LONG Yan, YANG Zhiyou, HE Mengfei. Recognizing apple targets before thinning using improved YOLOv7[J]. Transactions of the CSAE, 2023, 39(14): 191－199. (in Chinese)

[20] 姜国权, 杨正元, 霍占强, 等. 基于改进 YOLOv5 网络的疏果前苹果检测方法[J]. 江苏农业科学,2023, 51(14): 205－215.
JIANG Guoquan, YANG Zhengyuan, HUO Zhanqiang, et al. Apple detection method before fruit thinning based on the improved YOLOv5 network[J]. Jiangsu Agricultural Sciences,2023, 51(14): 205－215. (in Chinese)

[21] HU S, GAO F, ZHOU X, et al. Hybrid convolutional and attention network for hyperspectral image denoising[J]. IEEE Geoscience and Remote Sensing Letters,2024,21:1－5.

[22] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[C]// Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPs' 17). Red Hook: Curran Associates Inc., 2017:6000－6010.

[23] BOLYA D, ZHOU C, XIAO F, et al. YOLACT: real-time instance segmentation[C]// IEEE/CVP International Conference on Computer Vision (ICCV). Seoul,2019:9156－9165.