doi:10.6041/j.issn.1000-1298.2026.12.025

基于 YOLO 26n - seg 和沙漏神经网络模型的草莓相对轮廓预测

洪　旗[1,2]　孔鸣宇[1]　贺磊磊[1]　孔朔琳[1]　杨蜀秦[1,2]　傅隆生[1,2]

(1. 西北农林科技大学机械与电子工程学院, 陕西杨凌 712100; 2. 农业部农业物联网重点实验室, 陕西杨凌 712100)

摘要: 针对非结构化农业环境下草莓果实常受枝叶、果柄及相邻果实遮挡,导致机器视觉系统难以获取完整几何轮廓,进而影响采摘机器人定位精度与抓取成功率的问题,本研究提出一种融合改进 YOLO 26 实例分割与沙漏网络的草莓相对轮廓级联预测方法。首先采用最新的 YOLO 26n - seg 模型作为视觉感知主干,通过引入轻量化 GhostNetV2 重构主干网络并嵌入注意力机制,在复杂背景下精准分割出果实可见区域的掩码。其次针对遮挡区域的形状推断难题,结合主成分分析旋转校正算法对轮廓特征进行标准化处理,并构建基于 Hourglass Network 的相对轮廓预测模型。该模型利用其对称的自底向上与自顶向下结构,有效整合果实可见部分的局部细节与全局语义特征,建立从可见区域特征到整体轮廓坐标的映射关系。试验结果表明,改进分割模型平均精度均值达到 91.3%,参数量降至 3×10^6;轮廓预测模型在 4 个遮挡方向的交叉测试中,测试集平均 R^2 稳定在 0.650 0 ~ 0.698 7,平均绝对误差(MAE)降至 9 像素左右,显著优于未校正及传统 MLP(Multi-layer perceptron)模型。实现了对不同遮挡程度草莓果实精准形态补全,为果园智能采摘机器人避障规划与无损抓取提供了可靠的空间几何信息。

关键词: 草莓轮廓预测; YOLO 26n - seg; Hourglass Network; 实例分割; PCA; 旋转校正

中图分类号: TP391.4; S24　　**文献标识码**: A　　**文章编号**: 1000-1298(2026)12-0264-11　　**OSID**:

Strawberry Relative Contour Prediction Based on YOLO 26n - seg and Hourglass Network Model

HONG Qi[1,2]　KONG Mingyu[1]　HE Leilei[1]　KONG Shuolin[1]　YANG Shuqin[1,2]　FU Longsheng[1,2]

(1. *College of Mechanical and Electronic Engineering, Northwest A&F University, Yangling, Shaanxi* 712100, *China*
2. *Key Laboratory of Agricultural Internet of Things, Ministry of Agriculture and Rural Affairs, Yangling, Shaanxi* 712100, *China*)

Abstract: Addressing the issue that in unstructured agricultural environments, strawberry fruits are often obscured by leaves, stems, and adjacent fruits, making it difficult for machine vision systems to obtain complete geometric contours, which in turn affects the positioning accuracy and grasping success rate of picking robots, a relative contour cascade prediction method was proposed for strawberries that integrated an improved YOLO 26 instance segmentation with a Hourglass Network. This method first used the latest YOLO 26n - seg model as the visual perception backbone. By introducing a lightweight GhostNetV2 to reconstruct the backbone network and embedding an attention mechanism, it accurately segmented the mask of the fruit's visible area against complex backgrounds. Secondly, to address the challenge of shape inference in occluded areas, principal component analysis (PCA) rotation correction was applied to standardize contour features, and a relative contour prediction model based on the Hourglass Network was constructed. This model leveraged its symmetric bottom-up and top-down structure to effectively integrate local details of the fruit's visible parts with global semantic features, establishing a mapping from visible area features to the coordinates of the entire contour. Experimental results showed that the proposed improved segmentation model achieved an average precision (AP) of 91.3% with parameters reduced to 3×10^6, and the contour prediction model maintained an average R^2 between 0.650 0 and 0.698 7 with a mean absolute error (MAE) of around 9 pixels across different occlusion directions on the test set,

收稿日期: 2026 - 03 - 02　修回日期: 2026 - 03 - 23

基金项目: 国家自然科学基金项目(32572201)

作者简介: 洪旗(1989—),男,副教授,主要从事农业智能化技术与装备研究,E-mail: h407460979@163.com

通信作者: 傅隆生(1984—),男,教授,博士生导师,主要从事智慧农业技术与装备研究,E-mail: fulsh@nwafu.edu.cn

significantly outperforming uncorrected and traditional multi-layer perceptron (MLP) models. This method enabled accurate shape completion of strawberry fruits under various degrees of occlusion, providing reliable spatial geometric information for obstacle avoidance planning and non-destructive grasping by intelligent orchard picking robots.

Key words: strawberry contour prediction; YOLO 26n－seg; Hourglass Network; instance segmentation; PCA; rotation correction

0　引言

草莓作为一种高经济价值的浆果类作物，采收环节主要依赖人工，劳动强度大且成本高。随着智慧农业发展，研发能够适应非结构化田间环境的智能采摘机器人可以提高规模化生产采摘效率[1]。在机器人视觉感知系统中，精准获取果实空间位置与完整几何形态是实现无损抓取与避障规划的前提[2]。然而，实际生产环境中光照条件多变，且草莓植株生长密集，果实常被叶片、果柄或相邻果实遮挡[3]，这使得视觉系统往往仅能获取果实局部信息，导致基于可见区域计算的几何质心显著偏离真实物理中心[4]，进而引发机械手定位误差大、抓取失败率高及果实损伤等问题[5]。因此，如何在遮挡条件下精准推断出果实完整轮廓，是当前农业机器人视觉领域的痛点与难点。

针对果实检测与分割问题，近年来国内外学者开展了大量研究。以 YOLO 系列为代表的单阶段检测算法因优异的实时性，被广泛应用于果实识别任务。张毅等[6]针对柑橘果实检测问题在 YOLO v8 模型引入 CBAM 注意力机制，通过自适应调节特征图权重来抑制背景噪声，有效解决了果实重叠导致的漏检问题；LI 等[7]针对火龙果采摘机器人在夜间补光环境下的作业需求，提出了一种轻量化的 YOLO v5s 模型，通过将主干网络 CSPDarknet53 替换为 ShuffleNetV2 大幅降低了模型参数量，在保证低计算负载的同时实现了对重叠果实的精准识别。然而，上述方法本质上属于“模态分割”，即仅能分割出图像中可见的像素区域。当果实受遮挡比例较大时，残缺的分割掩码无法反映果实真实大小与形状，会直接影响后续机械臂末端执行器姿态控制[8]。

为了获取被遮挡物体的完整信息，研究者开始探索“非模态分割”与形状补全技术。传统的解决方法主要依赖几何拟合，如 LI 等[9]利用最小二乘法对残缺边缘进行椭圆拟合，该方法计算简单，但难以适应草莓这种心形、圆锥形等非规则几何体的形态变化，拟合误差较大。随着深度学习的发展，生成式模型逐渐得到广泛应用。CHEN 等[10]提出了一种基于条件生成对抗网络的果实图像修复方法，通过学习大量无遮挡样本的纹理分布，成功生成了被遮挡区域的像素信息；MAGISTRI 等[11]构建了大规模水果 3D 形状补全数据集，并利用 Transformer 网络从点云数据中推断果实的完整三维模型。尽管基于 GAN 或 Transformer 的方法在视觉效果上表现出色，但其网络结构复杂、计算量高，且容易生成虚假的纹理细节，难以满足农业嵌入式设备对实时性与鲁棒性的要求[12]。

目前，在非结构化农业环境下，草莓果实常因枝叶、果柄及相邻果实的遮挡，导致机器视觉系统难以获取完整的几何轮廓，进而严重影响采摘机器人的定位与抓取精度。如何在有限算力下，利用局部可见特征精准推理出全局几何轮廓，成为解决这一问题的关键。针对上述问题，本文提出一种融合改进 YOLO v8n－seg 实例分割与沙漏神经网络(Hourglass network)的草莓相对轮廓级联预测方法。首先，利用改进的 YOLO v8n－seg 模型，在复杂田间环境下对果实可见区域进行高精度实例分割以获取 RoI(Region of interest)。然后，使用基于主成分分析的草莓旋转校正算法对提取的轮廓坐标进行数据预处理，实现特征的标准化。其次，将轮廓补全建模为从“可见特征”到“相对轮廓坐标”的直接回归问题，构建基于沙漏网络的形状预测模型。模型利用自底向上与自顶向下的多尺度特征融合机制，捕捉并整合草莓可见部分的局部细节、边缘曲率与全局语义信息，进而预测出果实边缘相对于质心的完整轮廓点集，且无需生成冗余的像素纹理[13]。通过该级联架构，在不显著增加计算成本的前提下，以期有效解决遮挡导致的定位偏移与形状感知缺失难题，为果园智能采摘机器人的避障规划与精准无损抓取提供可靠的视觉支撑技术。

1　材料与方法

1.1　试验环境

1.1.1　试验场景构建

为验证算法在非结构化(如枝叶、果柄和果实遮挡等)农业环境下的鲁棒性，在陕西省咸阳市杨凌区草莓种植基地构建了模拟采摘试验场景，如

图 1 所示。试验采用履带式智能巡检机器人作为移动载体,搭载高清摄像模组,安装位置模拟机械臂末端执行器视角,保持主要作业距离为 20 ~ 50 cm。

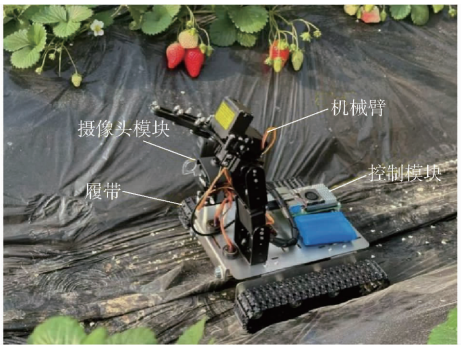


图 1 实地试验场景

Fig. 1 Field experiment scenario

试验涵盖晨间、正午及傍晚等不同时段,通过多时序采集引入强光直射、阴影变化及枝叶重叠等复杂环境干扰,以充分测试改进的 YOLO 26n - seg 模型的边缘分割性能。在试验过程,机器人沿垄间匀速巡航,系统对实时视频流执行“检测—分割—识别”级联处理。

1.1.2 软硬件环境配置

试验计算平台配置:处理器为 Intel Core i5 - 14600KF,显卡为 NVIDIA GeForce RTX 5060 Ti(16 GB显存),内存 32 GB。软件环境基于 Windows 11 操作系统,深度学习框架采用 PyTorch 2.0 配合 CUDA 11.8 加速,目标检测与分割算法基于 Ultralytics YOLO 26 框架部署。具体配置见表 1。

表 1 试验环境配置

Tab. 1 Experimental environment configuration

类别	配置
GPU	NVIDIA GeForce RTX 5060 Ti(16 GB)
CPU	Intel Core i5 - 14600KF
内存	32 GB DDR4
操作系统	Windows 11
深度学习框架	PyTorch 2.0 + Ultralytics YOLO 26
Python 版本	3.9
其他	OpenCV、NumPy、PyQt5

1.2 试验材料与数据集构建

1.2.1 分割模型数据集构建

采用“公共数据 + 针对性实地采集”的双源数据构建策略,共构建包含 1 500 幅原始图像的混合数据集。其中,800 幅基础样本精选自 StrawDI 公共数据集(https://strawdi.github.io/)[14],旨在使模型学习田间自然生长环境下果实的一般形态特征。针对枝叶密集遮挡、果实簇拥重叠等难识别场景[15-16],在陕西省西安市某草莓种植基地(34°05′N,108°83′E)利用搭载于机器人末端的 Raspberry Pi Camera Module v3 型高清摄像头补充采集 700 幅图像,分辨率为 4 608 像素 × 2 592 像素。如图 2 所示,采集过程严格覆盖晨间强光、正午阴影及傍晚弱光等不同时段,通过多时序采集引入环境干扰,重点捕获了受果柄缠绕、叶片遮挡等各种遮挡情况的复杂样本,以充分测试模型的边缘分割性能。



(a) 不同角度及光照条件

(b) 各遮挡条件

图 2 各类条件下采集数据示意图

Fig. 2 Diagrams of data collection under various conditions

数据集标注采用 Labelme[17],由具备农业视觉研究经验的人员对果实可见区域边缘进行逐像素勾勒。针对非结构化环境下的标注难题,本研究制定了严格的标注规范:对于受遮挡果实,仅标注其可见区域的多边形掩码,确保掩码边缘与真实物理边缘高度契合。为保障数据质量,团队实施了双人交叉回检机制,对重叠区域的界限界定及微小目标的漏检情况进行严格校验,确保标注的准确性,为模型训练提供高质量的数据集。

最终,将标注后数据集按比例 8∶1∶1 随机划分为训练集、验证集和测试集。在模型输入前,对训练集执行了包括随机旋转、亮度调整及添加高斯噪声在内的离线数据增强操作,以模拟实际作业中摄像头可能遇到的运动模糊与传感器噪声。这种多源数据融合与针对性增强策略,显著提升了模型对边缘模糊分割难题的挖掘能力,确保了改进 YOLO 26n - seg 模型在复杂场景下的鲁棒性。

1.2.2 预测模型数据集构建

为构建草莓轮廓预测的专用数据集,以 YOLO 26n - seg 模型分割得到的草莓可见区域图像为基础。首先通过 OpenCV 读取图像并将其从默认的 RGB 色彩格式转换为 HSV 色彩空间,再依据草莓红色在 HSV 空间中分布于 2 个连续区间的特性,通过阈值筛选函数提取符合该颜色特征的像素以生成初始掩码,随后针对初始掩码中存在的孤立白色噪点与草莓区域内的黑色孔洞问题,采用形态学操作对掩码进行优化以消除干扰并补全轮廓区域,接着从优化后的掩码中提取由连续白色像素构成的边界点集,即草莓轮廓,并通过形状特征筛选出符合草莓轮廓特征的有效轮廓,之后对该有效轮廓点集进行均

匀采样以获取 15 个特征点，同时将 OpenCV 默认的坐标系转换为“左下角为坐标原点”的新坐标系以实现坐标标准化，最终将标准化后的草莓轮廓点坐标存储至 Excel 表格，图 3 为提取的轮廓特征点示例，图 4 为基于该特征点绘制的某草莓样本轮廓效果图。

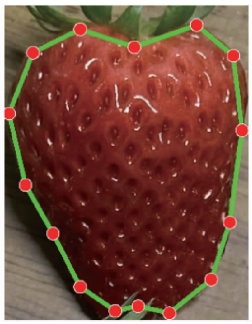


图 3　相关代码提取的轮廓点

Fig. 3　Contour points extracted from relevant code

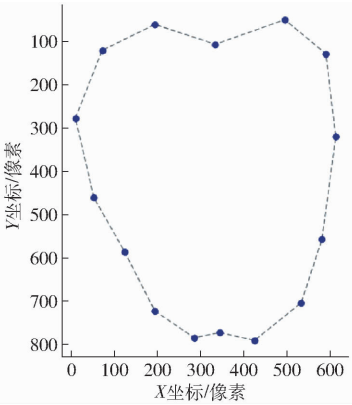


图 4　绘制的样本草莓轮廓图

Fig. 4　Detailed outline drawing of a particular strawberry sample

1.3　草莓果实分割算法

1.3.1　分割算法选择

针对复杂田间环境下草莓果实精准分割的难题，算法模型需要在检测精度和推理效率之间取得较好的平衡。为此，本研究首先针对当前主流的实例分割算法进行了对比。

以 Mask R－CNN 为代表的两阶段分割算法虽在边缘分割精度上表现优异，但其参数量大且计算流程复杂，导致推理延迟较高[18]，难以满足农业采摘机器人对作业实时性的要求，而基于 Transformer 架构的 RT－DETR－seg 模型虽在全局特征捕获方面具有显著优势[19]，但其计算复杂度较高，不仅导致训练收敛周期延长，且在算力受限的嵌入式设备上部署难度较大。

相比之下，YOLO 系列模型参数量较少，对算力要求较低。然而，传统的 YOLO v5、YOLO v7 和 YOLO v8 等模型依赖非极大值抑制（Non-maximum suppression，NMS）进行检测框的后处理，在处理草莓簇拥生长的密集场景时，极易将紧邻的真实果实误判为冗余框而漏检。鉴于此，选取 YOLO 26n－seg 作为基准框架。模型采用无 NMS 的端到端检测架构，一定程度上解决了密集重叠目标的漏检难题，但在处理严重枝叶遮挡时，其卷积结构仍存在全局语义感知不足的问题。因此，针对非结构化田间作业环境，对其进行针对性的结构优化，构建一种改进的果实分割模型。

1.3.2　分割算法改进

改进的 YOLO 26n－seg 模型拓扑结构如图 5 所示，其中 GhostNetV2 Block 表示引入 DFC 注意力的

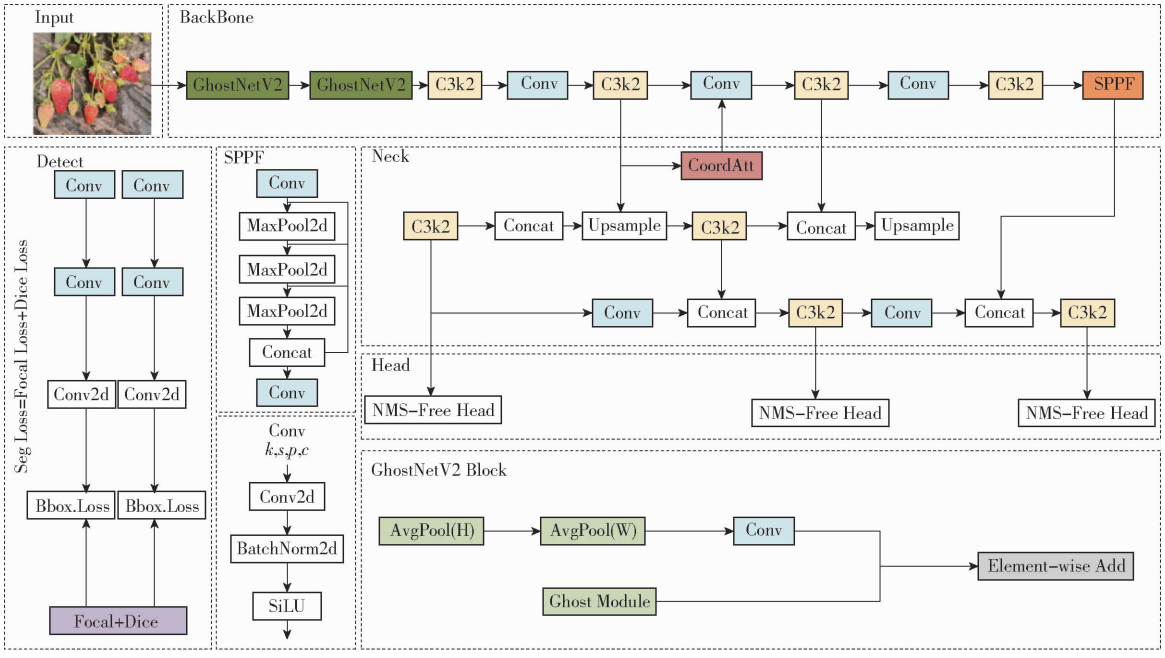


图 5　改进的 YOLO 26n－seg 模型结构

Fig. 5　Structure of improved YOLO 26n－seg model

轻量化主干模块,用于捕捉长距离依赖并降低计算开销;CoordAtt 为坐标注意力模块,用于强化对遮挡区域的空间感知;NMS - Free Head 为无非极大值抑制检测头,实现端到端推理。其余模块包括 C3k2(跨阶段特征融合)、SPPF(快速空间金字塔池化)及组合损失函数(Focal + Dice),具体改进策略涵盖 3 个维度。

(1)主干网络轻量化与全局感知重构。引入 GhostNetV2[20] 对主干网络进行重构。尽管 YOLO 26n 在轻量化方面表现优异,但在极端遮挡场景下对丰富语义信息的获取能力仍有提升空间。为此,利用 GhostNetV2 特有的 DFC 注意力机制,将全连接运算解耦为水平与垂直方向的特征聚合。

输入特征图为 $X_T \in \mathbf{R}^{H\times W\times C}$,DFC 机制通过在水平方向 W 和垂直方向 H 分别进行长距离卷积聚合,生成注意力图。其解耦过程可表示为

$$A_h = \sum_{j=1}^{W} K_h(j) X_{:,j,:} \tag{1}$$

$$A_w = \sum_{i=1}^{H} K_w(i) X_{i,:,:} \tag{2}$$

式中 K_h、K_w——水平、垂直方向变换权重

A_h、A_w——聚合后的特征描述子

i、j——水平和垂直方向空间坐标索引

最终经 DFC 注意力机制处理后的输出特征图 Y_T 可表示为

$$Y_T = X_T \odot \sigma(A_h + A_w) \tag{3}$$

式中 σ——Sigmoid 激活函数

这一设计将全局上下文信息注入原始特征,使模型在维持低计算成本的同时,具备了捕捉长距离空间依赖的能力,从而能通过周围叶片等环境上下文有效推断被遮挡果实的整体位置。

(2)特征融合注意力增强。在 Neck 端的特征融合模块嵌入 CoordAtt[21-22] 机制。针对草莓果实空间分布不均的特点,CoordAtt 摒弃了会丢失位置信息的传统全局平均池化,改为沿 X 轴和 Y 轴 2 个方向分别聚合特征。

对于输入特征图第 c 个通道,输出 $z_c^h(h)$ 与 $z_c^w(w)$ 可分别表示为

$$z_c^h(h) = \frac{1}{W} \sum_{0\leqslant i<W} x_c(h,i) \tag{4}$$

$$z_c^w(w) = \frac{1}{H} \sum_{0\leqslant j<H} x_c(j,w) \tag{5}$$

式中 H、W——输入特征图高度、宽度

通过式(4)、(5),模型分别生成了一对方向感知的特征图,强化了模型对果实可见区域空间位置与关键通道信息的感知敏感度,有效抑制背景噪声干扰。

(3)损失函数策略优化。引入 Focal Loss 与 Dice Loss 构建组合损失函数[23-24],总损失 L_{total} 定义为两者加权和,即

$$L_{total} = \lambda_{fl} L_{fl} + \lambda_{dice} L_{dice} \tag{6}$$

式中 λ_{fl}、λ_{dice}——Focal Loss 和 Dice Loss 的权重

利用 Focal Loss 解决正负样本不平衡问题,挖掘难分遮挡样本,公式为

$$L_{fl}(p_t) = -\alpha_t (1-p_t)^{\gamma} \ln p_t \tag{7}$$

式中 p_t——预测概率

α_t——平衡因子

γ——聚焦参数

同时,利用 Dice Loss 直接优化预测掩码与真实掩码的几何重合度,重点攻克边缘模糊分割的难题,公式为

$$L_{dice} = 1 - \frac{2\sum_{i}^{N} p_i y_i + \dot{o}}{\sum_{i}^{N} p_i + \sum_{i}^{N} y_i + \dot{o}} \tag{8}$$

式中 p_i——预测像素值

y_i——真实像素值

$\dot{o}$——平滑项

N——像素总数

该组合损失为后续轮廓补全提供高质量的几何输入。

1.4 草莓相对轮廓预测模型

1.4.1 预测模型输入输出确定

在草莓采摘场景,采集到的草莓图像常存在严重遮挡问题,会导致机械臂无法获取被遮挡草莓的完整信息,进而大幅降低采摘效率。因此,核心目标是解决被遮挡草莓的识别难题,为机械臂采摘提供精准依据。

基于 1.2.2 节获取的草莓相对轮廓坐标,设计了针对性的分割策略,以模拟草莓被遮挡的场景,实现对遮挡草莓识别能力的优化。对所有样本的草莓相对轮廓坐标,分别沿 X 轴和 Y 轴进行排序,据此将坐标集合划分为沿 X 轴方向的左右 2 部分和沿 Y 轴方向上下 2 部分。在取点规则上,左部分与上部分各选取 8 个坐标点,右部分与下部分各选取 7 个坐标点,图 6 为此次采集的草莓轮廓离散点分布图。从草莓图像中提取草莓边缘轮廓像素点,以此作为草莓轮廓坐标,因此草莓坐标的单位为像素。

在真实的草莓采摘场景,果实受遮挡方位具有高度随机性(如枝叶可能遮挡果实的上方、侧方或下方)。为了全面评估模型在不同遮挡模式下的识别与形状推断能力,基于提取的草莓相对轮廓坐标,

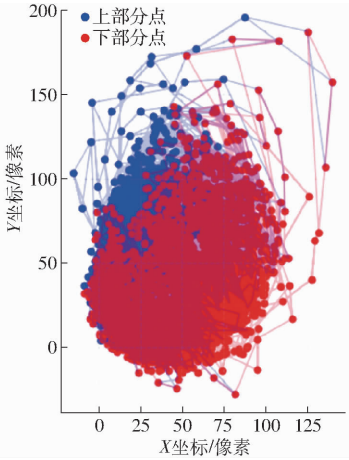


图 6　草莓轮廓离散点分布图

Fig. 6　Visualization of discrete contour points sampling

设计了 4 种不同方向的遮挡模拟策略。

具体而言，针对标准化后的 15 个草莓轮廓离散特征点，分别沿 X 轴和 Y 轴对其坐标集合进行排序与划分。在取点规则上，分别构建 4 组独立的输入–输出映射关系：输入“上部分点”预测“下部分点”、输入“下部分点”预测“上部分点”、输入“左部分点”预测“右部分点”，以及输入“右部分点”预测“左部分点”。通过这种多向交叉组合策略，全面模拟了草莓在非结构化环境中可能遭遇的各类遮挡工况，旨在验证并提升模型对任意局部可见区域的全局形状推理鲁棒性。

1.4.2　基于主成分分析的草莓旋转校正算法

相关研究表明，草莓可近似视为具有对称特性的三维结构。然而在实际草莓采摘场景，草莓的空间姿态存在显著差异，导致提取的轮廓坐标分布无序，使得神经网络对其特征的提取能力显著下降。主成分分析（Principal component analysis，PCA）是一种从离散数据提取主导分布方向的统计方法，无需人工设定规则，能客观量化草莓的对称方向，自动找到草莓轮廓的对称轴。为此，提出基于 PCA 的草莓旋转校正算法[25]，对草莓轮廓坐标进行旋转校正，以此来标准化草莓姿态，消除不同草莓之间和草莓因摆放角度差异而导致的轮廓坐标分散性，使所有草莓对称轴统一平行于 Y 轴，为后续特征提取、形状分析或机器学习建模提供一致的数据输入。图 7 为基于主成分分析草莓旋转校正算法的基本流程。

草莓旋转校正算法基本步骤为：

（1）设草莓由 N 个点 (x_i, y_i)（$i = 1, 2, \cdots, N$）组成，通过轮廓几何矩求解质心，即

$$\begin{cases} x_0 = \dfrac{\boldsymbol{M}_{10}}{\boldsymbol{M}_{00}} \\ y_0 = \dfrac{\boldsymbol{M}_{01}}{\boldsymbol{M}_{00}} \end{cases} \tag{9}$$

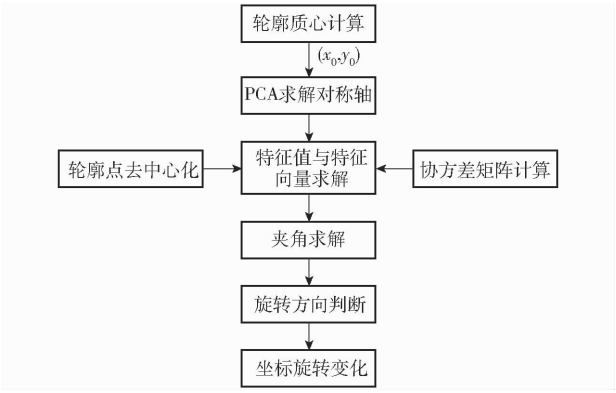


图 7　基于 PCA 草莓旋转校正算法

Fig. 7　PCA-based strawberry rotation correction algorithm

其中

$$\boldsymbol{M}_{00} = \sum_{i=1}^{N} 1 \tag{10}$$

$$\boldsymbol{M}_{10} = \sum_{i=1}^{N} x_i \tag{11}$$

$$\boldsymbol{M}_{01} = \sum_{i=1}^{N} y_i \tag{12}$$

式中　x_0、y_0——草莓轮廓质心坐标

$\boldsymbol{M}_{00}$——轮廓零阶矩

$\boldsymbol{M}_{10}$——一阶 x 矩

$\boldsymbol{M}_{01}$——一阶 y 矩

（2）通过 PCA 分析求解其对称轴。为消除质心位置影响，将所有轮廓点减去质心坐标，得到去中心化点，即

$$\begin{cases} x_i' = x_i - x_0 \\ y_i' = y_i - y_0 \end{cases} \tag{13}$$

式中　x_i'、y_i'——去中心化后坐标

x_i、y_i——原始轮廓点坐标

并计算去中心化点集的协方差矩阵 $\boldsymbol{\Sigma}$ 为

$$\boldsymbol{\Sigma} = \begin{bmatrix} Cov(x', x') & Cov(x', y') \\ Cov(y', x') & Cov(y', y') \end{bmatrix} \tag{14}$$

协方差计算式为

$$Cov(a, b) = \frac{1}{N} \sum_{y=1}^{N} (a_i - \bar{a})(b_i - \bar{b}) \tag{15}$$

通过求解协方差矩阵 $\boldsymbol{\Sigma}$ 的特征值和对应的特征向量，可得

$$\boldsymbol{\Sigma} \boldsymbol{v}_i = \lambda_i \boldsymbol{v}_i \quad (i = 1, 2) \tag{16}$$

式中　λ_i——协方差矩阵特征值，设 $\lambda_1 \geqslant \lambda_2$

$\boldsymbol{v}_i$——特征值对应特征向量

最大特征值 λ_1 对应的特征向量 $\boldsymbol{v}_1$，即为草莓轮廓的对称轴方向。

（3）旋转角度计算。先将向量 $\boldsymbol{v}_1$ 单位化，由于需要将对称轴 $\boldsymbol{v}_1$ 旋转至 Y 轴，即与单位向量 $\boldsymbol{e}_y = (0, 1)$ 平行，因此需要计算两者夹角 θ。

根据向量点积几何意义，可以得到

$$\cos\theta = \boldsymbol{v}_1 \cdot \boldsymbol{e}_y = v_{1x} 0 + v_{1y} 1 = v_{1y} \tag{17}$$

因此角度可表示为

$$\theta = \arccos v_{1y} \tag{18}$$

根据向量叉积确定旋转方向，即

$$\boldsymbol{v}_1 \times \boldsymbol{e}_y = v_{1x}1 - v_{1y}0 = v_{1x} \tag{19}$$

若 $v_{1x} < 0$，角度取 $-\theta$；若 $v_{1x} \geqslant 0$，角度取 θ。

最后，基于图像坐标系和旋转矩阵公式，可得

$$\begin{bmatrix} x_r' \\ y_r' \end{bmatrix} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta y \end{bmatrix} \tag{20}$$

式中 x_r'、y_r'——去中心化后轮廓点经过旋转矩阵变换得到的相对坐标

Δx、Δy——去中心化后坐标

旋转后点坐标可表示为

$$\begin{cases} x_r' = \Delta x\cos\theta - \Delta y\sin\theta \\ y_r' = \Delta x\sin\theta + \Delta y\cos\theta \end{cases} \tag{21}$$

再将坐标转换回原来坐标系，即

$$\begin{cases} x' = x_r' + x_0 \\ y' = y_r' + y_0 \end{cases} \tag{22}$$

式中 x'、y'——旋转校正后草莓轮廓点坐标

由此得到旋转后坐标。

该草莓旋转校正算法基于 PCA 解析草莓轮廓的主轴方向，并结合二维平面旋转变换的数学关系完成轮廓姿态校正，有效消除了草莓生长异质性带来的轮廓提取结果无序问题，实现了轮廓特征标准化，为后续神经网络特征建模与模型训练提供了关键的数据预处理支撑，真实数据集在校正前后的空间分布对比如图 8 所示。

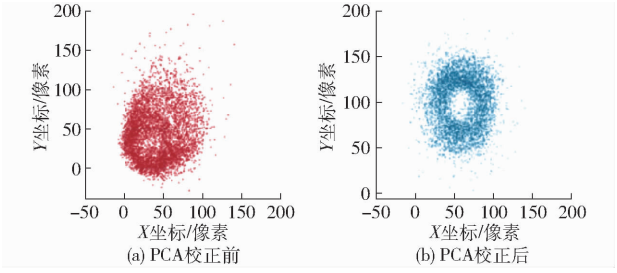


图 8 真实数据集在 PCA 校正前后的空间分布

Fig. 8 Spatial distribution of real dataset before and after PCA correction

1.4.3 基于沙漏神经网络的草莓相对轮廓预测模型

沙漏网络凭借对称式结构可实现多尺度特征的全面捕捉与融合，突破了传统网络在全局-局部信息平衡方面的局限性，该网络已广泛应用于机械零件边缘预测等轮廓预测类任务，因此本研究选用该网络构建草莓相对轮廓预测模型。

设计的沙漏神经网络为适配一维坐标数据的全连接版本，核心遵循“编码-瓶颈融合-解码”的对称架构，并引入残差连接机制缓解梯度消失问题，最终实现从 16 维草莓上轮廓（或左轮廓）坐标到 14 维草莓下轮廓坐标的非线性回归预测。图 9 为沙漏神经网络的整体框架，其核心组成与结构如下：

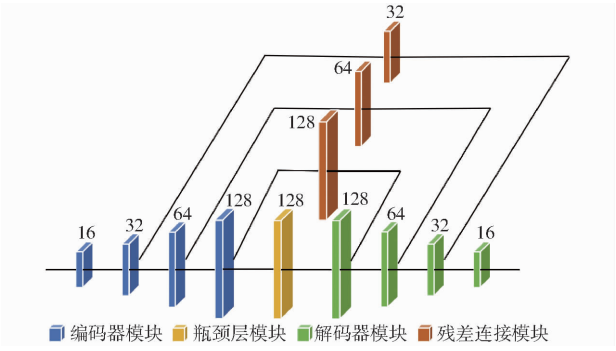


图 9 沙漏神经网络基本框架

Fig. 9 Hourglass neural network basic framework

（1）编码器模块

编码器由 3 层全连接层构成，特征维度从输入的 16 维逐步升维至 32、64 维，最终达到 128 维；每层全连接层后均配置 LeakyReLU 激活函数与 BatchNorm1d 批量归一化层。

该层级化结构可将原始坐标的浅层特征（如单个轮廓点的位置信息）逐步抽象为高维语义特征（如轮廓整体形态特征），契合特征提取的层级性规律。针对坐标预测任务的复杂非线性映射需求，选用 LeakyReLU 激活函数可避免普通 ReLU 易出现的“死亡神经元”问题；而 BatchNorm1d 模块则能解决小批次训练场景下的内部协变量偏移问题，有效加速模型收敛并提升训练过程的稳定性。

（2）瓶颈层模块

瓶颈层仅包含 1 层全连接层，特征维度保持 128 维不变，且同样配置 LeakyReLU 激活函数与 BatchNorm1d 批量归一化层。

作为编码器与解码器的中间枢纽，瓶颈层可对编码器输出的高维抽象特征进行进一步非线性变换，提炼轮廓核心特征并剔除冗余信息；同时保持维度稳定，为后续解码器对称降维过程提供统一且高辨识度的特征输入。

（3）解码器模块

解码器由 3 层全连接层构成，特征维度从 128 维逐步降维至 64、32 维，最终还原为 16 维输出维度；其中前 2 层全连接层配置与编码器一致的 LeakyReLU 激活函数和 BatchNorm1d 批量归一化层，最后 1 层仅保留线性变换。

解码器采用与编码器对称的降维策略，契合沙漏网络“收缩-扩张”的核心设计理念，可实现特征从高维抽象形态到低维具体坐标的精准还原；逐步降维方式避免了直接从 128 维降至 16 维导致的信息过度压缩，有效保留轮廓点的细节特征。

此外，最后一层采用纯线性变换，适配坐标回归任务对连续值输出的需求，避免激活函数引入额外非线性误差。

(4)残差连接模块

残差连接模块包含 32、64、128 维 3 个维度的线性层，各线性层维度与编码器对应层级输出维度完全匹配，实现编码器各层特征与解码器或瓶颈层对应层特征的叠加融合。

该模块可使梯度通过“短路路径”直接回传至网络浅层，有效缓解深层全连接网络的梯度消失问题，保障深层网络的训练有效性；同时让解码器在降维过程中融合编码器的低层细节特征，如单个轮廓点的位置信息，避免解码阶段丢失关键的轮廓细节，显著提升坐标预测精度。

综上，该沙漏神经网络架构能够在小样本训练场景下稳定拟合草莓上轮廓或左轮廓到下轮廓的坐标映射关系，既可保证模型训练稳定性，又可有效提升了轮廓坐标预测精度。

1.5　所提方法基本流程

图 10 为草莓轮廓预测流程图，先在草莓采摘试验场景中采集草莓整体图像，借助 YOLO－seg 算法将其分割为单个草莓图像；再通过轮廓提取算法获取草莓轮廓数据，经旋转校正与排序得到草莓左上、右下轮廓数据并组成轮廓数据集；最终用该数据集训练沙漏神经网络，以此实现草莓轮廓预测。

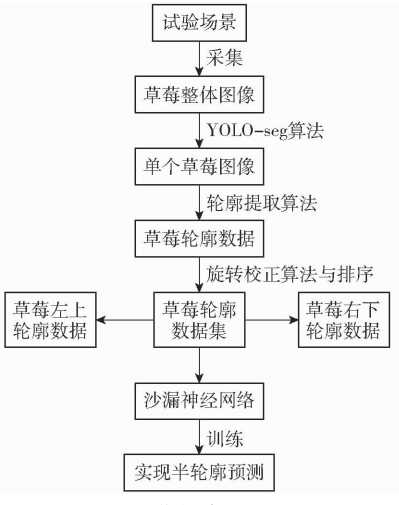


图 10　草莓轮廓预测流程图

Fig. 10　Strawberry contour prediction flowchart

2　结果与分析

2.1　评价指标

为客观评估各模型在草莓实例分割任务中的综合性能，选取精确率(P)、召回率(R)及平均精度均值(mAP)作为精度评价指标。其中，mAP@0.5 表示交并比(IoU)阈值为 0.5 时的平均精度均值，主要用于衡量模型基础的定位与分割准确度；mAP@0.5:0.95 表示 IoU 阈值从 0.5 递增至 0.95 时全部平均精度均值的算术平均值，该指标对空间对齐的精度要求更为严苛，能够更全面地反映模型对果实边缘的高精度贴合能力。选取参数量(Params)与推理速度作为模型复杂度与实时性的衡量标准。

针对几何形状补全任务的特殊性，单一的决定系数 R^2 虽能反映坐标数值的拟合方差，但难以全面衡量形状拓扑结构与几何相似度。为此，在 R^2 的基础上，采用平均绝对误差(Mean absolute error, MAE)评估预测轮廓与真实物理边界的整体贴合精度，并利用均方误差(Mean square error, MSE)量化如突刺或内凹等极端几何畸变，以综合检验草莓形状推断的平滑度与拓扑结构稳定性。

综合上述指标，从数值回归拟合度、边界贴合距离与拓扑结构稳定性 3 个维度对模型进行全面评估。

2.2　不同分割模型性能测试

为了全面验证所提算法在复杂农业环境下的性能优势，本研究在统一的草莓数据集与 GPU 试验配置下，将改进的 YOLO 26n－seg 模型与当前主流实例分割算法进行对比。选取的对比模型涵盖了不同架构，具体包括：二阶段算法 Mask R－CNN、基于 Transformer 架构的 RT－DETR－seg，以及轻量化单阶段算法 YOLO v5n－seg[26]、YOLO v7－tiny－seg[27]和 YOLO v8n－seg。

试验结果如表 2、3 所示。由表 2 可知，Mask R－CNN 得益于 Transformer 主干网络对特征的深层提取能力，在 mAP@0.5 上表现优异，达到 87.3%。但其复杂的两阶段架构导致参数量高达 8.6×10^7，推理速度仅为 18.0 f/s，难以满足农业机器人实时作业要求。同理，RT－DETR－seg 虽在精度上保持了较高水准，但其计算量过大，限制了其在嵌入式终端上的部署潜力。相比之下，本文模型通过引入注意力机制与轻量化重构策略，成功实现了精度与效率的双重优化。模型 mAP@0.5 达到 91.3%，在分割精度上优于 Mask R－CNN，且参数量仅为 3×10^6，同时，推理速度达到 92.0 f/s，能够流畅处理高帧率视频流。结果表明，本文模型在降低计算负载的同时维持了分割性能，平衡了检测速度与精度，为草莓自动采摘机器人视觉系统提供了可行的技术方案[28]。

2.3　预测方法性能测试

为验证所提草莓旋转校正算法与改进沙漏神经网络在各种随机遮挡工况下的有效性与鲁棒性，针对 4 种遮挡进行了全面测试。所有模型均训练 200 轮，结果如表 4 所示。由表 4 可知，PCA 旋转校正对

表 2 不同分割模型在草莓数据集上性能测试结果

Tab. 2 Performance test results of different segmentation models

模型	主干网络	mAP@0.5/%	mAP@0.5:0.95/%	参数量	推理速度/(f·s^{-1})
Mask R-CNN	Swin-T	87.3	63.2	8.60×10^7	18
RT-DETR-seg	HGNetv2-L	86.9	62.4	3.25×10^7	32
YOLO v5n-seg	CSPDarknet	83.3	58.2	2.10×10^6	98
YOLO v7-tiny-seg	ELAN	85.2	60.2	6.80×10^6	76
YOLO v8n-seg	CSPDarknet	84.4	59.5	3.40×10^6	95
YOLO 26n-seg	Native-Nano	86.7	61.0	3.20×10^6	95
YOLO 26n-seg(本文)	GhostNetV2	91.3	66.5	3.00×10^6	92

表 3 消融试验结果

Tab. 3 Ablation experiment results

模型	mAP@0.5/%	mAP@0.5:0.95/%	精确率/%	召回率/%	参数量	推理速度/(f·s^{-1})
YOLO 26n-seg	86.7	61.0	81.7	79.0	3.2×10^6	95
YOLO 26n-seg+CoordAtt	87.3	62.0	83.7	81.0	3.4×10^6	88
YOLO 26n-seg+GhostNetV2	87.2	61.6	83.0	79.9	2.8×10^6	110
YOLO 26n-seg+Focal+Dice	87.3	62.0	83.0	80.0	3.2×10^6	95
YOLO 26n-seg+CoordAtt+GhostNetV2	88.5	63.7	85.0	82.0	3.0×10^6	98
YOLO 26n-seg+CoordAtt+Focal+Dice	88.9	64.0	85.0	81.8	3.4×10^6	88
YOLO 26n-seg+GhostNetV2+Focal+Dice	88.7	64.0	84.8	81.0	2.8×10^6	110
YOLO 26n-seg+CoordAtt+GhostNetV2+Focal+Dice	91.3	66.5	87.5	84.7	3.0×10^6	92

表 4 不同遮挡下预测模型在测试集上的性能

Tab. 4 Performance of prediction models on test set under different occlusion directions

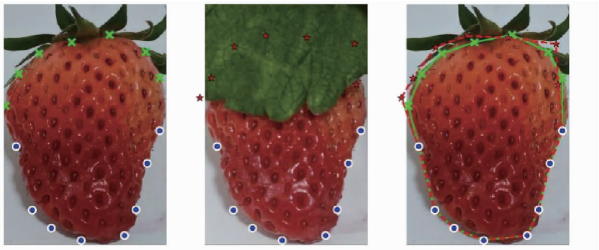
模型输入	模型输出	数据处理	模型	MSE/像素^2	MAE/像素	R^2
上部分点	下部分点	未经校正	MLP	463.41	16.74	0.274 6
			沙漏网络	414.60	15.22	0.349 0
		经 PCA 校正	MLP	274.82	13.07	0.427 2
			沙漏网络	147.65	9.49	0.650 0
下部分点	上部分点	未经校正	MLP	514.25	18.04	0.156 3
			沙漏网络	523.11	17.12	0.128 2
		经 PCA 校正	MLP	237.59	12.01	0.505 6
			沙漏网络	137.14	8.88	0.694 6
左部分点	右部分点	未经校正	MLP	565.80	17.98	0.222 9
			沙漏网络	516.27	16.45	0.325 7
		经 PCA 校正	MLP	265.34	13.24	0.465 3
			沙漏网络	140.77	9.16	0.698 7
右部分点	左部分点	未经校正	MLP	540.32	16.85	0.113 2
			沙漏网络	495.07	15.15	0.300 0
		经 PCA 校正	MLP	237.99	11.94	0.541 4
			沙漏网络	143.16	9.11	0.664 9

缓解遮挡带来的姿态分布混乱起到了一定积极作用。在 4 种遮挡方向下,未经校正的原始轮廓数据使得模型拟合效果普遍较差(测试集 R^2 均不大于 0.349,MAE 较高)。引入 PCA 校正进行初步标准化后,各模型的 MSE 和 MAE 均出现明显下降,表明标准化坐标输入有助于减弱草莓生长姿态异质性带来的干扰,在一定程度上降低了神经网络对几何形变学习的难度。

其次,改进的沙漏神经网络展现出了较好的形状拓扑重构能力与一定的方向鲁棒性。在经 PCA 校正数据集上,传统 MLP 模型测试集 R^2 为 0.427 2 ~ 0.541 4,平均 MAE 为 12 ~ 13 像素,预测误差相对较大。相比之下,本文提出的沙漏网络在 4 个不同遮挡方向上的测试集 R^2 提升至 0.650 0 ~ 0.698 7,MAE 降低至约 9 像素。相对较小的 MAE 与 MSE 表明,面对不同方位遮挡,沙漏网络能够较好地依据局部特征推断出缺失部位的坐标空间分布。同时,MSE 降低说明该网络在一定程度上减少了导致拓扑破裂的极端异常预测点,输出的轮廓相对更为平滑合理,具备了一定实际应用潜力。

2.4　预测结果可视化

为了直观验证所提模型在形状补全任务中的有效性，选取典型草莓样本在遮挡情况下进行轮廓预测可视化分析。结果如图 11 所示，图中蓝色标记表示提取的可见区域轮廓点，绿色标记表示被遮挡区域真实轮廓点，红色标记表示沙漏神经网络输出的预测点。



(a) 真实轮廓点　(b) 预测上轮廓点　(c) 预测轮廓与真实轮廓

图 11　基于沙漏神经网络的草莓轮廓预测可视化结果

Fig. 11　Visualization of strawberry contour prediction

从图 11a、11b 点位分布可看出，模型能够基于可见部分的几何特征，合理推断出被遮挡区域的空间位置。图 11c 通过闭合连线进一步展示了整体形态的拟合情况，预测轮廓与真实轮廓在几何形状与曲率变化上较为一致。这表明该方法能够应对遮挡造成的形状缺失问题，且所预测的轮廓具有一定平滑性与结构合理性，能够为采摘机器人的无损抓取规划提供有效的空间几何信息。

3　结论

(1)构建了轻量化改进的 YOLO 26n－seg 果实分割模型。针对遮挡与实时性需求，利用 GhostNetV2 的 DFC 注意力增强全局感知，结合 CoordAtt 与组合损失函数改善了密集场景下的边缘分割效果。试验表明，改进的模型推理速度达到 92.0 f/s，且引入 GhostNetV2 后模型参数量降至 3×10^6，体现了其在轻量化方面的有效性。最终，改进的模型在测试集上的 mAP@0.5 达到 91.3%，较好地平衡了农业机器人对实时性与精度的需求，能够为后续处理提供相对准确的果实可见区域掩码。

(2)验证了基于 PCA 的草莓旋转校正预处理策略在规范轮廓特征空间分布方面的积极作用。针对草莓生长姿态各异导致的坐标分布无序问题，利用主成分分析解析轮廓对称主轴，并将坐标统一旋转至标准化方向。多向遮挡模拟试验表明，该预处理步骤有效减弱了姿态异质性干扰，使得各模型在测试集上的 MSE 与 MAE 均出现明显下降，R^2 得到普遍改善，证实了姿态标准化是降低神经网络几何学习难度、提升形状推断稳定性的重要步骤。

(3)设计了基于沙漏神经网络的相对轮廓预测模型，探索了多向遮挡区域形状推断的可行性。利用沙漏网络的对称“编码-解码”结构与残差连接机制，融合草莓局部细节与整体形态特征。在经 PCA 校正的数据集上针对 4 个不同遮挡方向进行交叉测试，该模型的 MAE 约稳定在 9 像素，测试集平均 R^2 保持在 0.650 0 ~ 0.698 7，整体表现优于传统 MLP 模型。这表明该模型具备一定的方向鲁棒性，能够基于局部可见特征相对平滑地推断出缺失部位的坐标空间分布，未产生严重的拓扑结构破裂，为解决遮挡条件下的形状推理难题提供了一种具有潜力的参考思路。

(4)提出的“分割—校正—预测”协同策略，实现了复杂环境下草莓果实的精准识别与分割，完成了被遮挡果实的形态补全。该方法在保证系统实时运行的前提下，为果园智能采摘机器人避障规划与无损抓取提供了相对可靠的空间几何信息支撑。

参　考　文　献

[1]　傅隆生，家博，刘小娟，等. 基于文献计量学的农业工程技术发展动态分析[J]. 农业机械学报，2025，56(10)：249－276.
FU Longsheng，JIA Bo，LIU Xiaojuan，et al. Development trends of agricultural engineering technology based on bibliometrics [J]. Transactions of the Chinese Society for Agricultural Machinery，2025，56(10)：249－276.（in Chinese）

[2]　WANG C，LIU Y，ZHANG Z，et al. Fruit localization and picking point estimation for strawberry harvesting robot in complex environments[J]. Computers and Electronics in Agriculture，2024，216：108523.

[3]　WU D，LV C，JIANG M，et al. Strawberry maturity recognition and localization based on improved YOLO v8[J]. International Journal of Agricultural and Biological Engineering，2024，17(2)：189－198.

[4]　AFONSO M，FONTEIJN H，FIORENTIN F，et al. Amodal instance segmentation for overlapped fruits in unstructured environments[J]. AgriEngineering，2023，5(3)：1425－1442.

[5]　LYU S，LI M，HE Z，et al. Occluded apple fruit detection and localization using an improved YOLO v8 algorithm and depth information[J]. Biosystems Engineering，2024，238：125－139.

[6]　张毅，李斌，陈浩. 基于 YOLO v8 与注意力机制的柑橘果实检测[J]. 农业工程学报，2024，40(12)：77－85.
ZHANG Yi，LI Bin，CHEN Hao. Citrus fruit detection based on YOLO v8 and attention mechanism[J]. Transactions of the CSAE，2024，40(12)：77－85.（in Chinese）

[7]　LI H，LI J，WEI H，et al. A lightweight improved YOLO v5s model and its deployment for detecting pitaya fruits in daytime and nighttime light-supplement environments[J]. Computers and Electronics in Agriculture，2024，220：108914.

[8] GENÉ-MOLA J, SANZ-CORTÉS M, MARTIN-GORRIZ B, et al. Fruit detection, segmentation and 3D localisation using the MobileNetv3 - YOLO v4 architecture[J]. Precision Agriculture, 2023, 24(1): 278 - 299.
[9] LI D, WANG R, XIE C. A recognition method for *Camellia oleifera* fruit based on improved YOLO v5 and ellipse fitting[J]. Computers and Electronics in Agriculture, 2023, 209: 107812.
[10] CHEN Z, WU Y, TING D, et al. Fruit image restoration and occlusion handling based on conditional generative adversarial networks (cGAN)[J]. IEEE Access, 2022, 10: 45231 - 45242.
[11] MAGISTRI F, STACHNISS C. A dataset and benchmark for shape completion of fruits for agricultural robotics[C]// 2024 IEEE International Conference on Robotics and Automation (ICRA). IEEE, 2024: 1123 - 1129.
[12] MIN X, MEHDIPOUR S. Computer vision meets generative models in agriculture: advances, challenges, and opportunities [J]. Applied Sciences, 2025, 15(14): 7663.
[13] NEWELL A, YANG K, DENG J. Stacked hourglass networks for human pose estimation[C]// European Conference on Computer Vision (ECCV), 2016: 483 - 499.
[14] PÉREZ-BORRERO I, MARÍN D, GEGÚNDEZ-ARIAS M E, et al. StrawDI: the strawberry digital images data set[C]// International Conference on Image Analysis and Recognition (ICIAR), 2020: 711 - 723.
[15] LU Y, YOUNG S. A survey of public datasets for computer vision tasks in precision agriculture[J]. Computers and Electronics in Agriculture, 2020, 178: 105760.
[16] LIU X, HE L, NIU S, et al. Opportunities and challenges in the application of Digital Twins for orchard management[J]. Computers and Electronics in Agriculture, 2025, 239: 111104.
[17] LIU Y, ZHANG K, YANG J, et al. A scalable and efficient image annotation framework for agricultural computer vision[J]. Computers and Electronics in Agriculture, 2023, 207: 107705.
[18] KIRILLOV A, HE K, GIRSHICK R, et al. Segment anything[C]// Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), 2023: 4015 - 4026.
[19] ZHANG H, ZHAO Y, LI S, et al. RT - DETR: real-time detection transformer[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2024: 15256 - 15265.
[20] TANG Y, HAN K, GUO J, et al. GhostNetV2: enhance cheap operation with long-range attention[J]. Advances in Neural Information Processing Systems, 2022, 35: 9969 - 9981.
[21] LIU S, CHEN Y, WANG J, et al. Attention-enhanced lightweight network for fruit detection in complex orchards[J]. Biosystems Engineering, 2024, 234: 79 - 92.
[22] ZHOU L, HUANG X, QIN Y, et al. Spatial-aware attention network for instance segmentation in agricultural environments [J]. Computers and Electronics in Agriculture, 2023, 204: 107522.
[23] WANG J, LIU H, SUN Y, et al. Improved focal loss for imbalanced fruit segmentation tasks[J]. IEEE Access, 2023, 11: 118945 - 118956.
[24] ZHANG Y, CHEN M, LI Q, et al. A hybrid focal-dice loss for accurate segmentation of occluded fruits[J]. Computers and Electronics in Agriculture, 2024, 221: 109030.
[25] JOLLIFFE I T, CADIMA J. Principal component analysis: a review and recent developments[J]. Philosophical Transactions of the Royal Society A, 2016, 374(2065): 20150202.
[26] CHEN Z, TING D, WU Y, et al. Real-time strawberry detection and segmentation in complex environments using a lightweight YOLO v5n - seg model[J]. Computers and Electronics in Agriculture, 2024, 218: 108688.
[27] GE Z, LIU S, WANG F, et al. YOLOX: exceeding YOLO series in 2021[C]// Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), 2021: 439 - 446.
[28] TANG Y, ZHOU H, WANG H, et al. Fruit detection and positioning for picking robot based on improved YOLO v8 and binocular stereo vision[J]. Computers and Electronics in Agriculture, 2024, 215: 108426.
[29] PARSA S, DEBNATH B, KHAN M A, et al. Modular autonomous strawberry picking robotic system[J]. Journal of Field Robotics, 2024, 41(7): 2226 - 2246.
[30] ZHOU H, WANG X, AU W, et al. Intelligent robots for fruit harvesting: recent developments and future challenges[J]. Precision Agriculture, 2024, 25(1): 128 - 156.