

融合多尺度特征的枸杞果实颗粒分割方法

姚明诚¹, 卫建国², 张华科¹, 吴门新³, 范锦龙⁴, 王静萍¹, 范鑫愉¹, 张承明^{1*}, 杨晓霞^{1*}

1. 山东农业大学信息科学与工程学院, 山东 泰安 271018

2. 宁夏气象局, 中国气象局旱区特色农业气象灾害监测预警与风险管理重点实验室,
宁夏回族自治区 银川 750002

3. 国家气象中心, 北京 100081

4. 北京师范大学地理科学学部, 北京 100875

摘要: 精确的果实颗粒形状是评估枸杞生长状态、测量果实尺寸的重要基础。本文提出了一种融合多尺度特征的枸杞果实颗粒分割网络(Chinese Wolfberry Fruitlet Particles Instance Segmentation Network, CWISNet), 用于从图像中提取精确的枸杞果实颗粒形状。CWISNet融合Transformer的全局特征提取能力和卷积神经网络的局部细节特征能力提取逐像素语义特征, 融合ResNet-34的深层特征提取能力和空洞卷积的大感受野能力生成边缘信息, 利用分级融合策略对逐像素语义特征和边缘信息进行融合, 生成高质量的逐像素语义特征, 最后利用多级候选框提取出精细的枸杞果实颗粒形状。综合考虑形状损失和边界损失构建损失函数, 用于对模型进行训练。选择宁夏回族自治区中宁县为实验区, 选择Mask R-CNN、YOLACT、HTC、SOLOv2作为对比模型开展实验。实验结果表明, 本方法的边界框平均精度均值和掩码平均精度均值在IoU=0.5阈值下分别达到了92.5%和92.3%, 明显优于对比模型, 在密集遮挡和边界模糊场景下同样表现出了卓越的性能, 可为科研和生产实际提供技术和数据支撑。

关键词: 枸杞; 实例分割; 边缘检测; 混合架构; 深度学习

中图分类号: TP 391.413

文献标识码: A

文章编号: 1000-2324(2026)05-0907-11

An Instance Segmentation Method for Chinese Wolfberry Fruit Particles Based on Multi-scale Feature Fusion

YAO Ming-cheng¹, WEI Jian-guo², ZHANG Hua-ke¹, WU Men-xin³,
FAN Jin-long⁴, WANG Jing-ping¹, FAN Xin-yu¹, ZHANG Cheng-ming^{1*},
YANG Xiao-xia^{1*}

1. College of Information Science and Engineering/Shandong Agricultural University, Tai'an 271018, China

2. Key Laboratory of Meteorological Disaster Monitoring, Early Warning and Risk Management for Characteristic Agriculture in Arid Areas/China Meteorological Administration, Ningxia Hui Autonomous Region, Yingchuan 750002, China

3. The National Meteorological Centre, Beijing 100081, China

4. Faculty of Geographical Sciences/Beijing Normal University, Beijing 100875, China

Abstract: Accurate fruit particle shape is an important basis for evaluating the growth status of Chinese Wolfberry and measuring fruit size. This paper proposes a Chinese Wolfberry Fruitlet Particles Instance Segmentation Network (CWISNet) that fuses multi-scale features to extract accurate shapes of Chinese Wolfberry fruitlet particles from images. CWISNet combines the global feature extraction capability of Transformer with the local detail feature extraction capability of the convolutional neural network to obtain pixel-wise semantic features. It also fuses the deep feature extraction capability of ResNet-34 with the large receptive field capability of dilated convolution to generate edge information. A hierarchical fusion strategy is used to merge the pixel-wise semantic features and edge information, producing high-quality pixel-wise semantic features. Finally, multi-level candidate boxes are used to extract fine Chinese Wolfberry fruitlet particle shapes. Considering both shape loss and boundary loss, CWISNet constructs a loss function to train the model. Zhongning County in Ningxia Hui Autonomous Region is selected as the experimental area, and Mask R-CNN, YOLACT, HTC, and SOLOv2 are used as comparative models for the experiments. The results show that the proposed method achieves a box mAP of 92.5% and a

收稿日期: 2025-11-06

修回日期: 2026-09-14

基金项目: 宁夏回族自治区自然科学基金(2024AAC03419); 山东省气象局创新团队专项(2024sdcxtd04); 风云卫星应用先行计划(三期)项目(FY-APP-2024.0301); 中央引导地方科技发展专项项目(2024FRD05056); 山东省自然科学基金(ZR2021MD097)

第1作者简介: 姚明诚(2000-), 男, 硕士研究生, 研究方向: 深度学习与智慧农业。E-mail: yaomingcheng@qq.com

***通讯作者:** Author for correspondence. E-mail: chming@sdau.edu.cn; yangxx@sdau.edu.cn

mask mAP of 92.3% at an IoU threshold of 0.5, outperforming the comparative models by a clear margin. It also demonstrates excellent performance in scenes with dense occlusions and blurred boundaries, providing technical and data support for both research and field applications.

Keywords: Chinese Wolfberry; instance segmentation; edge detection; hybrid architecture; deep learning

枸杞是宁夏等地农民增收和经济发展的支柱产业,但传统生产管理方式劳动强度较大^[1,2]。利用图像处理技术准确提取枸杞果实颗粒信息,是枸杞生产管理中进行机械作业时的关键环节^[3]。但由于枸杞果实体积小、果实分布密集,果实之间、果实与枝叶之间常发生相互遮挡,导致分割难度较大,如何从图像中提取出准确的果实颗粒信息,成为目前枸杞产业迫切需要解决的关键技术难题^[4]。

高质量特征是准确提取枸杞果实的基础^[5,6]。近年来,以卷积神经网络(Convolutional Neural Network, CNN)为基础的目标检测与实例分割技术在农业领域得到广泛应用,如YOLOv3^[7]、YOLOv5^[8]、YOLOv8^[9]、Faster R-CNN^[10]、Mask R-CNN^[11]、RetinaNet^[12]。其中,Mask R-CNN是性能较为出色的模型,应用于油茶果^[13]、葡萄^[14]、草莓^[15]、苹果^[16]、番茄^[17]等作物时都取得了较理想的效果。CNN的突出特点是具有平移不变性和局部性,但由于CNN较为关注局部信息,感受野通常较小,在捕捉长距离依赖和建模全局上下文方面能力不足,难以平衡全局语义与局部细节的需求^[18,19],影响了复杂条件下特征提取结果的质量。

现有研究发现,CNN的全局特征提取能力仍然不够理想^[19],针对这一问题,研究者使用Transformer结构去捕捉全局特征,如Swin Transformer通过引入层次化设计和基于移位窗口的自注意力机制,实现了高效的全局特征提取^[20,21]。但Transformer架构缺乏CNN所具有的平移不变性和局部性,导致在表达精细纹理与局部特征时效果不够理想。

为了综合利用两者的优势,研究者开始尝试使用Transformer-CNN混合架构来构建能够兼顾全局与局部的特征提取方法^[22]。但从公开文献来看,现有方法多采用简单的特征拼接或串联方式^[23,24],这类方法应用于目标结构复杂的场景时,由于现有混合方式未能充分挖掘两种架构在特征层面的互补潜力,仍难以提取出理想的高质量特征。

由于颜色、亮度的变化,图像上的果粒间通常会呈现出较明显的边界^[25,26]。有研究者发现在提取过程中融入边界信息后,可明显提高提取结果的精度^[27],即使在复杂背景下也能取得较理想的效果^[28]。在多任务学习联合优化轮廓的工作中,边界信息的作用更为明显^[29,30]。但现有研究中,边界提取与语义分割通常共享同一骨干网络,这种结构可能会导致在逐级抽象生成高级语义特征的过程中,细微的边缘信息被平滑掉^[31,32];另一方面,边界与语义特征的融合策略较为单一,多在单一尺度上进行简单拼接,不能兼顾不同尺度目标的边界细化^[33,34]。为了保证提取过程中能够生成精细的果粒边界,需要针对果粒的实际情况设计相应的模型结构和融合策略。

基于上述分析,本文针对科研和生产实际对提取精细果实颗粒信息的迫切需求,提出了一种融合多尺度特征的枸杞果实颗粒分割网络(Chinese Wolfberry Fruitlet Particles Instance Segmentation Network, CWISNet),该网络综合利用Transformer和CNN在特征提取方面的优势建立特征提取模块,生成兼顾全局与细节的综合语义特征;利用多个层叠加的方式建立了边缘信息提取模块,生成枸杞果实颗粒边缘信息;最后融合边界信息与逐像素特征生成精细的枸杞果实轮廓。

1 数据获取与预处理

选择宁夏回族自治区中宁县作为研究区,中宁是我国重要的高品质枸杞产区,该县位于干旱地带,少雨、多日照、大温差的气象条件非常适合枸杞的生长,其地理位置如图1所示。

图像采集工作于2024年和2025年的5至10月进行,使用搭配有定焦镜头的专业相机进行拍摄,在拍摄过程中,注重从多个角度取景,充分获取了多类型的田间环境、不同程度遮挡等情况的数据资料,共采集3 000余张有效图像。

图2展示了三张图像,从图中可以看出,存在多种导致枸杞果实颗粒分割难度较大的因素,如枸杞果实形态差异较大;自然光照在果实上产

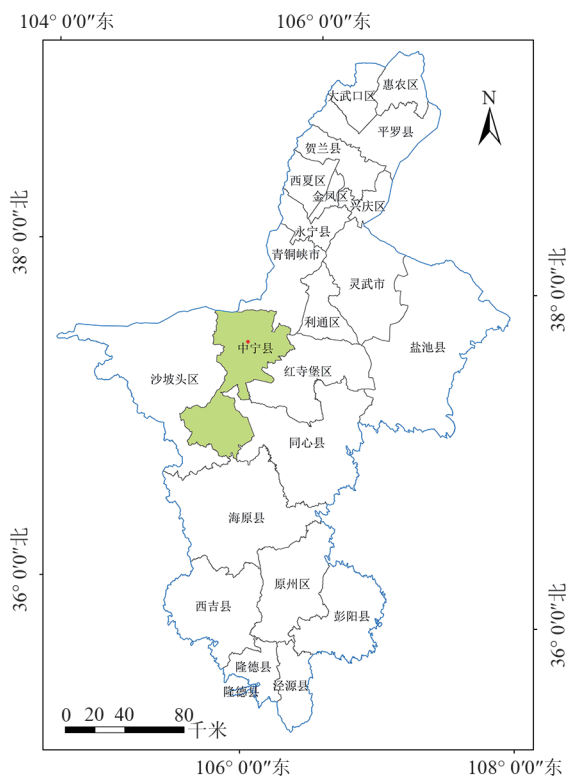


图1 中宁县位置

Fig. 1 Location of Zhongning County

生了高光 and 阴影;枸杞果实以簇生形式密集分布;密集的枝叶与果实重叠(图2(a),图2(b)),

土壤、枯枝等形成的复杂背景(图2(c))。

利用 LabelMe 图像标注软件标记出每个枸杞果实颗粒的轮廓,共标绘枸杞实例 63 756 个。为平衡计算效率与模型精度,将所有图片尺寸调整至 800×800 像素。使用随机划分方式,选择出训练集 1 600 张图像(46 736 个果实),验证集 300 张图像(8 357 个果实),测试集 300 张图像(8 663 个果实)。对训练集采用了多种离线数据增强技术,包括水平翻转、垂直翻转、颜色抖动(亮度变化范围为±50%,对比度变化范围为±50%,色调偏移范围为±20%)、添加高斯噪声(均值 $\mu=0$ 、标准差 $\sigma=50$)和随机旋转(角度范围为±25°),最终训练集图像数量为 9 600 张。

2 方法

CWISNet 的结构如图 3 所示,主要包含(a)输入,(b)语义特征提取模块,(c)边缘信息提取模块,(d)特征融合模块,(e)分割生成模块,(f)输出六个部分。其中特征提取模块用于从图像中提取逐像素语义特征和边缘信息;特征融合模块用于对语义特征和边缘特征进行融合、优化;生成分割模块用于生成枸杞果实的分割结果;输入输出均为图像。

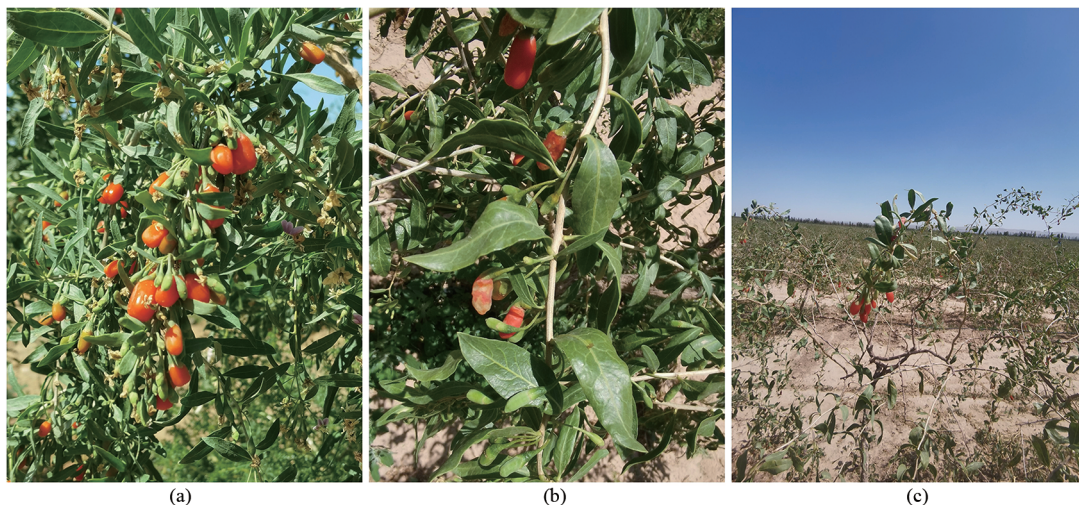


图2 田间自然场景下的枸杞图像

Fig. 2 Wolfberry images in natural field scenes

2.1 语义特征提取模块

语义特征提取子模块用于获取图像的全局特征和局部细节特征,生成逐像素的语义特征,由多个使用残差连接的 Swin-Conv 混合块组成。从图 4 所给出的 Swin-Conv 混合块结构图可以看

出,一个 Swin-Conv 混合块由两个连续的层构成,主要组件为位置编码生成器(Position Encoding Generator, PEG)、窗口多头自注意力(Window-based Multi-Head Self-Attention, W-MSA)、移位窗口多头自注意力(Shifted Window-

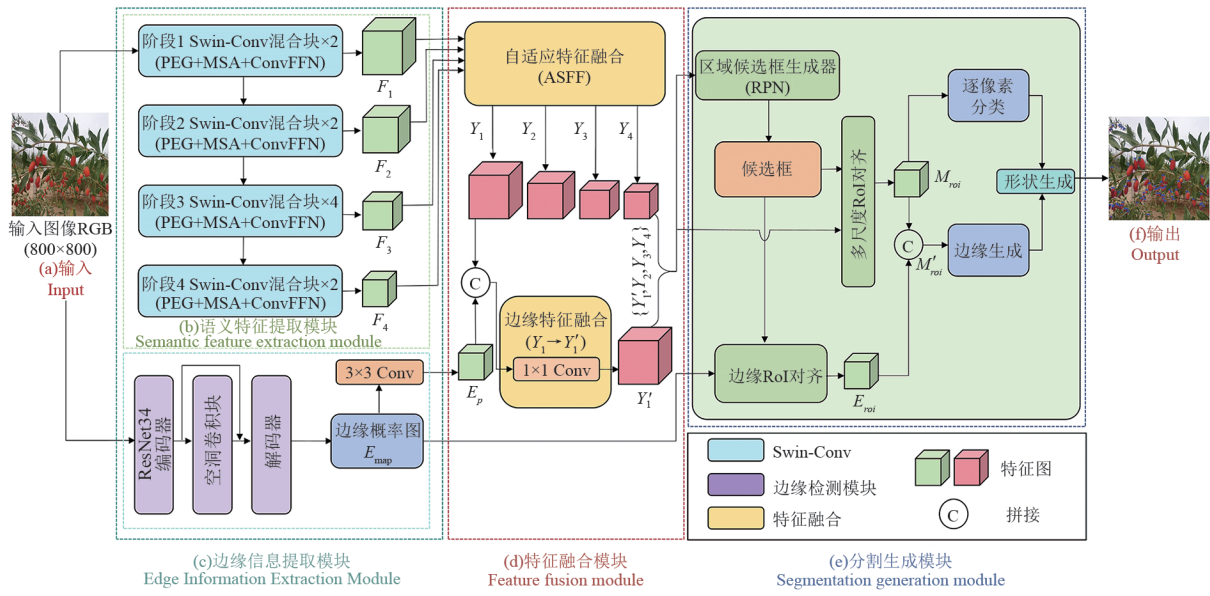


图3 CWISNet总览

Fig. 3 CWISNet overview

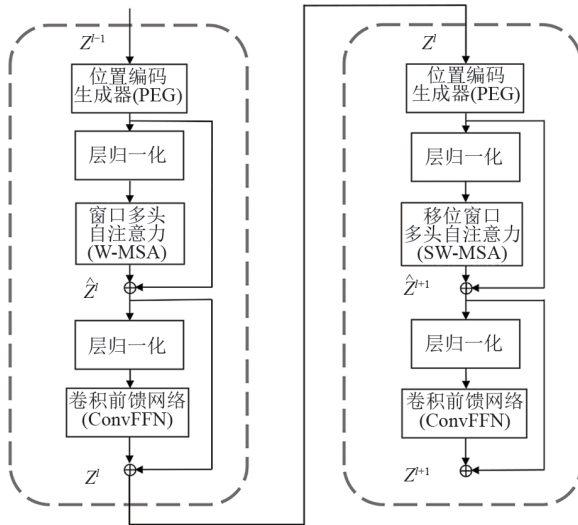


图4 Swin-Conv块结构示意图

Fig. 4 Swin-Conv block structure diagram

based Multi-Head Self-Attention, SW-MSA) 和卷积前馈网络(Convolutional Feed-Forward Network, ConvFFN)。Swin-Conv混合块中,内容感知的条件位置编码和卷积化的前馈网络是专门针对枸杞果实个体小、簇生分布的特点设计的,可有效提升网络对纹理和轮廓的捕捉能力。

Swin-Conv的第一层和第二层均设置了PEG,PEG能够利用深度卷积生成包含语义特征和位置相关信息的位置编码。位置编码计算方法如公式(1)所示。

$$PEG(z) = z + DWConv_{3 \times 3}(\text{reshape}(z)) \quad (1)$$

其中 z 是输入的特征图, $DWConv_{3 \times 3}$ 代表一个核大小为 3×3 的深度卷积。在图像内容变化剧烈的区域,位置编码的差异较大,而在图像内容变化较小的区域,位置编码变化也相对平滑;在两个枸杞颗粒的交界处,PEG将生成具有明显梯度的位置编码。

Swin-Conv第一个子层使用W-MSA作为注意力机制,用于对来自PEG的特征图进行自注意力处理,W-MSA首先将输入划分为多个不重叠的局部窗口,然后在每个窗口内部独立进行自注意力计算,在捕捉有效特征依赖的同时显著减少了计算量。W-MSA输出的结果通过残差与PEG的输出连接,生成中间特征 $\hat{Z}^l$,如公式(2)所示。

$$\hat{Z}^l = W\text{-MSA}(\text{LN}(PEG(Z^{l-1}))) + PEG(Z^{l-1}) \quad (2)$$

其中LN表示层归一化,LN通过统计单个样本在所有特征维度上的均值 μ 和方差 σ^2 来对输入特征 x 进行标准化,加速收敛并提高训练稳定性,其计算过程如公式(3)所示。

$$\text{LN}(x) = \frac{x - \mu}{\sqrt{\sigma^2 + \epsilon}} \cdot \gamma + \beta \quad (3)$$

其中, ϵ 为极小常数(本文中为 10^{-5}), γ 和 β 分别为模型在训练过程中学习到的缩放因子和偏置因子。

前馈网络(Feed-Forward Network, FFN)是Transformer的重要组件,对于获取局部细节特征

中结构实现的。

边缘信息融合子模块将枸杞果实颗粒边缘信息图与优化后的逐像素语义特征进行融合,一方面为边缘像素附加上语义信息,便于在分割时将同一果实颗粒的边缘点形成一个整体;另一方面增强边缘两侧像素的特征差异,使逐像素分类器能够更好地完成分类工作,提高逐像素分类结果的精度。

2.4 分割生成模块

分割生成模块由区域候选框生成器(Region Proposal Network, RPN)、多尺度感兴趣区(Region of Interest, RoI)对齐、边缘对齐、逐像素分类、边缘生成、形状生成等组件构成。

RPN 组件是构成分割生成模块最重要的组件,负责从输入的特征图中生成可能包含枸杞果实的候选框。通过对枸杞果实颗粒图像进行统计分析,发现枸杞颗粒的宽高比可分为 1:1, 1:2 和 2:1 等三种结果,根据这一统计知识,RPN 首先设计了 8×8, 16×16, 32×32 和 64×64 四种尺寸的正方形基础锚点框,再对每种基础锚点框进行三种变化,共形成了 12 种不同尺寸的锚点框,以适应不同尺寸的枸杞颗粒。

RPN 生成的锚点框经过筛选后,包含完整枸杞颗粒的候选框内的语义特征子图将被送到多尺度感兴趣区对齐组件进行尺寸对齐处理;然后与经过边缘对齐的边缘信息进行拼接,输入到逐像素分类组件进行分类处理,初步获取枸杞颗粒所在区域;提取出的初始枸杞颗粒区域与边缘生成组件生成的边缘图一起作为形状生成组件的输入,经形状组件处理后,生成最终的高质量枸杞颗粒形状。

2.5 损失函数

损失函数由形状损失和边界损失构成,如公式(8)所示。

$$L_{\text{total}} = (L_{\text{RPN}} + L_{\text{cls}} + L_{\text{reg}} + L_{\text{mask}}) + \lambda_b L_b \quad (8)$$

其中, L_{total} 表示综合损失, L_{RPN} 表示区域候选框生成器损失, L_{cls} 表示分类损失, L_{reg} 表示边界框回归损失, L_{mask} 表示掩码分割损失, L_b 表示边界辅助损失, λ_b 为权重参数表示边界辅助损失在综合损失中的占比,其值根据验证集上的对比实验确定,在候选取值集合 {0.05, 0.1, 0.2, 0.5, 1.0} 中分别训练并比较验证集效果后选取性能最优者,

最终确定该参数的值为 0.2。

L_{RPN} 的计算公式如公式(9)所示:

$$L_{\text{RPN}} = -\frac{1}{N_a} \sum_{i=1}^{N_a} [y_i \cdot \log(p_i) + (1 - y_i) \cdot \log(1 - p_i)] + \lambda_{\text{rpn_reg}} \frac{1}{N_{\text{pos}}} \sum_{i \in \rho_a} \text{smooth}_{\text{L1}}(t_i - t_i^*) \quad (9)$$

其中, N_a 为锚框总数, y_i 为第 i 个锚框的真实标签, p_i 为模型预测第 i 个锚框为枸杞果实的概率, ρ_a 为枸杞果实锚框的集合, N_{pos} 为枸杞果实锚框总数, t_i 和 t_i^* 为第 i 个锚框的预测参数和真实边界框参数。

L_{cls} 的计算公式如公式(10)所示:

$$L_{\text{cls}} = -\frac{1}{M_r} \sum_{j=1}^{M_r} \log(p_{j,c_j}) \quad (10)$$

其中, M_r 为 RoI 的总数, p_{j,c_j} 为模型预测第 j 个 RoI 属于其真实类别 c_j 的概率。

L_{reg} 的计算公式如公式(11)所示:

$$L_{\text{reg}} = \frac{1}{M_{\text{pos}}} \sum_{j \in \rho_{\text{pos}}} \sum_{k=1}^4 \text{smooth}_{\text{L1}}(t_{j,k} - t_{j,k}^*) \quad (11)$$

其中, ρ_{roi} 为正样本 RoI 集合, M_{pos} 为枸杞果实 RoI 的数量, $t_{j,k}$ 和 $t_{j,k}^*$ 为第 j 个 RoI 的第 k 个边界框参数的预测值和真实值。

L_{mask} 的计算公式如公式(12)所示:

$$L_{\text{mask}} = -\frac{1}{M_{\text{pos}}} \sum_{j \in \rho_{\text{pos}}} \frac{1}{H \times W} \sum_{h=1}^H \sum_{w=1}^W [m_{j,h,w}^* \cdot \log(m_{j,h,w}) + (1 - m_{j,h,w}^*) \cdot \log(1 - m_{j,h,w})] \quad (12)$$

其中, $m_{j,h,w}$ 和 $m_{j,h,w}^*$ 为第 j 个 RoI 在位置 (h, w) 的预测掩码值和真实掩码值。

L_b 的计算公式如公式(13)所示:

$$L_b = \frac{1}{N} \sum_{i=1}^N [w_p \cdot q_i^* \cdot \log(q_i) + w_n (1 - q_i^*) \cdot \log(1 - q_i)] \quad (13)$$

其中, N 是像素总数, q_i 和 q_i^* 是第 i 个像素的边缘预测概率和真实标签, w_p 和 w_n 分别代表正负样本的权重,用于缓解正负样本数量不均衡对训练的影响,一般可设置 $w_p = 0.9$ 、 $w_n = 0.1$ 。

3 实验结果与分析

3.1 对比实验设计

选择 PyTorch 作为深度学习框架,选择 OpenCV 等开源组件作为基础库,使用 Python 语

言开发了CWISNet的代码。

选择Mask R-CNN、YOLACT、HTC、SOLOv2作为对比模型。Mask R-CNN是两阶段实例分割的经典方法,用于验证CWISNet相较于标准两阶段方法在掩码质量与边界细节刻画方面的提升效果;YOLACT使用两个分支生成原型掩码和掩码系数,通过线性组合原型掩码与掩码系数得到分割结果,用于验证CWISNet在相近推理效率约束下对密集小目标场景的分割精度与轮廓完整性的优势;HTC通过在级联阶段中交替执行边界框回归和掩码预测,并引入语义分割分支辅助,用于验证CWISNet与强两阶段级联方法对比时对重叠遮挡实例区分及边界精细化的有效性;SOLOv2作为一种先进的单阶段的实例分割方法,直接根据实例的位置和尺度预测掩码,用于验证CWISNet相较于单阶段、无候选框方法在复杂背景下的鲁棒性与边界定位精度表现。

对比实验在一台配备Intel Xeon Silver 4210R CPU和NVIDIA GeForce RTX 3090 GPU

的工作站上完成,使用梯度下降算法对模型进行训练,使用AdamW作为优化器,在训练过程中,使用了自动混合精度技术^[35],该技术能够显著加快训练速度,并降低对显存的需求。为防止梯度爆炸,应用梯度裁剪将模型参数梯度的L2范数上限限制为5.0。

3.2 对比实验结果及评价

使用边界框平均精度均值(Mean Average Precision for Box, mAP^{box})评价分类与定位效果, mAP_{50}^{box} 和 mAP_{75}^{box} 分别表示IoU阈值为0.5和0.75时的边界框检测精度,值越高表示检测框与真实框的重合度更高、检测更准确;使用掩码平均精度均值(Mean Average Precision for Mask, mAP^{mask})评价实例分割效果,该指标在类别预测正确的前提下衡量掩码重合度, mAP_{50}^{mask} 和 mAP_{75}^{mask} 分别表示IoU阈值为0.5和0.75时的掩码分割精度,值越高表示预测掩码与真实掩码的重合度更高、果实轮廓分割更精确。对比结果如表1所示,因为SOLOv2的输出没有边界框,所以在表1中相应指标值为空。

表1 不同模型的mAP对比(%)

Table 1 Comparison of mAPs for different models

模型 Model	边界框 mAP^{box}	边界框@0.5 mAP_{50}^{box}	边界框@0.75 mAP_{75}^{box}	掩码 mAP^{mask}	掩码@0.5 mAP_{50}^{mask}	掩码@0.75 mAP_{75}^{mask}
Mask R-CNN	63.3	91.0	72.5	62.0	90.1	71.7
YOLACT	60.5	85.8	71.2	59.3	85.5	66.9
HTC	65.8	91.4	73.6	64.7	90.5	73.3
SOLOv2	\	\	\	61.2	87.3	68.7
Ours	71.7	92.5	80.6	71.0	92.3	80.2

图6给出了一组对比实验的可视化结果,图

中用黄色框出了识别结果差异明显的区域。



图6 对比模型结果可视化

Fig. 6 Visualization of comparison model results

对比实验结果可以发现,虽然Mask R-CNN使用了多尺度特征融合机制,但当图像中枸杞果实颗粒尺寸差异较大时,Mask R-CNN的多尺度特征融合模块无法为尺寸小的枸杞颗粒生成有

效的特征信息,从图6可以看出,Mask R-CNN的结果中,枸杞果实颗粒密集处的边界定位不够准确。CWISNet通过引入ASFF模块,能够合理设置不同尺度的权重系数,克服了枸杞颗粒目标尺

寸差异带来的影响。

与 CWISNet 相比, HTC 在确定边界时往往因特征差异较小而出现错误, 尤其是当枸杞果实因遮挡而不规则、边界模糊、密集粘连时, 这种现象尤为突出。CWISNet 通过引入边缘概率图, 使得网络对高频区域给予充分关注, 从而能够较好地将相连的颗粒分离开。

与单阶段模型相比, SOLOv2 直接利用特征图分割出目标, 但由于缺少两阶段方法中 RoI 特征对齐和边界细化等步骤生成的提示信息, 导致分割结果并不理想。从图 6 也可以看出, SOLOv2 生成的结果中存在明显漏检现象, 且生

成的形状的贴合度也比两阶段方法差。

从图 6 可以看出, 当因光照变化导致图像亮度存在差异时, YOLACT 的结果中存在较明显的漏检现象, 这种情况是由于 YOLACT 基于原型掩码线性组合的机制难以表达复杂边缘导致的。CWISNet 为每个实例独立预测掩码, 并融合边缘概率图提供的边界信息, 能够更好地适应光照变化和复杂形态。

图 7 展示了 CWISNet 在不同场景下的分割结果, 可以看出, 在高密度簇生(图 7(a))、复杂背景(图 7(b))以及光照多变(图 7(c))等场景下, CWISNet 均能较好地完成准确检测和分割任务。

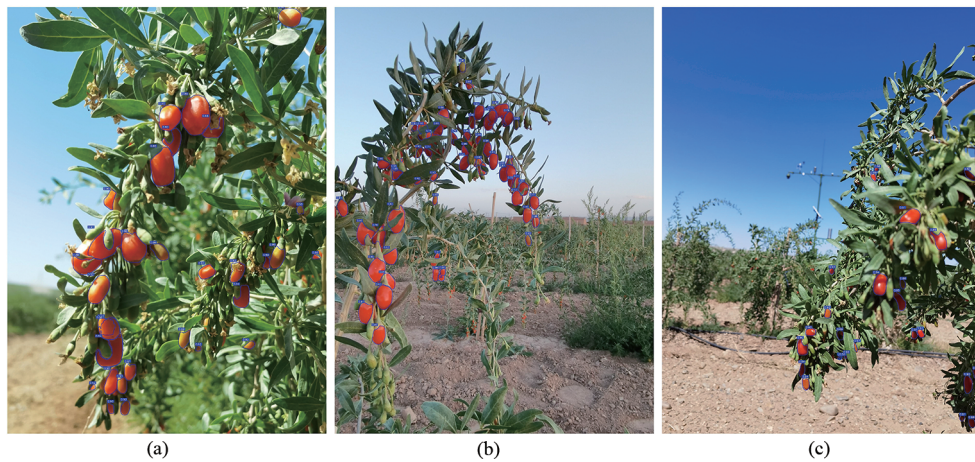


图 7 CWISNet 在不同场景下的结果
Fig. 7 Results of CWISNet in different scenes

3.3 消融实验

消融实验设置如表 2 所示, 用来定量评估 PEG、ConvFFN、ASFF 和边缘信息提取模块对模型性能的贡献。实验采用 Swin Transformer 作为基线模型, 其输出的多尺度特征直接送入分割生成模块。在相同数据和硬件条件下, 基线模型的 mAP^{mask} 为 62.0%, 这一数值将作为后续比较的基

础。在表 2 中, “×”表示未使用对应模块, “√”表示在基线模型上引入对应模块, 不同字母(A-E)表示不同的消融配置。其中 A 表示不引入 PEG、ConvFFN、ASFF 与边缘检测模块的基线模型, B 表示在基线模型上仅引入 PEG, C 表示在 B 的基础上进一步引入 ConvFFN, D 表示在 C 的基础上加入 ASFF 模块, E 表示四个模块全部引入。

表 2 消融实验配置
Table 2 Ablation experiment configuration

消融配置 Ablation setting	位置编码生成器 PEG	卷积前馈网络 ConvFFN	自适应特征融合 ASFF	边缘检测 Edge detection
A	×	×	×	×
B	√	×	×	×
C	√	√	×	×
D	√	√	√	×
E	√	√	√	√

表 3 汇总了消融实验在不同消融配置下的边界框检测与掩码分割性能,可以看出各模块的引入对两类任务均带来稳定增益。相较于基线配置 A,加入 PEG(配置 B)后边界框与掩码指标均有所提升,说明内容相关的位置编码有助于密集区域的实例区分。进一步加入 ConvFFN(配置 C)后, mAP^{box} 提升至 67.3%, mAP^{mask} 提升至

66.4%,表明局部感知能力增强后,边界框贴合度与掩码轮廓质量同步改善。在此基础上引入 ASFF(配置 D),指标继续提升,说明 ASFF 有效增强了不同尺寸目标的定位与分割能力。最终加入边缘信息提取模块形成完整模型(配置 E),指标达到最高,表明边缘监督为模型提供了更明确的轮廓约束,有利于密集粘连场景下的精细分割。

表 3 消融实验的平均精度均值结果(%)
Table 3 mAP results of ablation experiments (%)

消融配置 Ablation setting	边界框 mAP^{box}	边界框@0.5 mAP_{50}^{box}	边界框@0.75 mAP_{75}^{box}	掩码 mAP^{mask}	掩码@0.5 mAP_{50}^{mask}	掩码@0.75 mAP_{75}^{mask}
A	63.8	90.7	72.1	62.0	89.3	71.3
B	65.9	91.1	73.2	63.9	89.7	72.0
C	67.3	91.8	74.6	66.4	90.8	74.3
D	70.5	91.9	77.7	69.1	91.2	78.4
E	71.7	92.5	80.6	71.0	92.3	80.2

图 8 给出了一组对比结果,从图 8 可以看出,基线模型在密集区域存在较多漏检且边界粗糙。引入 PEG 后漏检情况有所改善,但分割边界仍不够精细;继续引入 ConvFFN 后,大部分果实能

被正确识别,粘连果实间开始出现分离趋势;加入 ASFF 后,大小果实的分割完整性进一步提高;完整的 CWISNet 能够正确识别大部分果实,粘连果实之间开始出现较清晰的分离趋势。

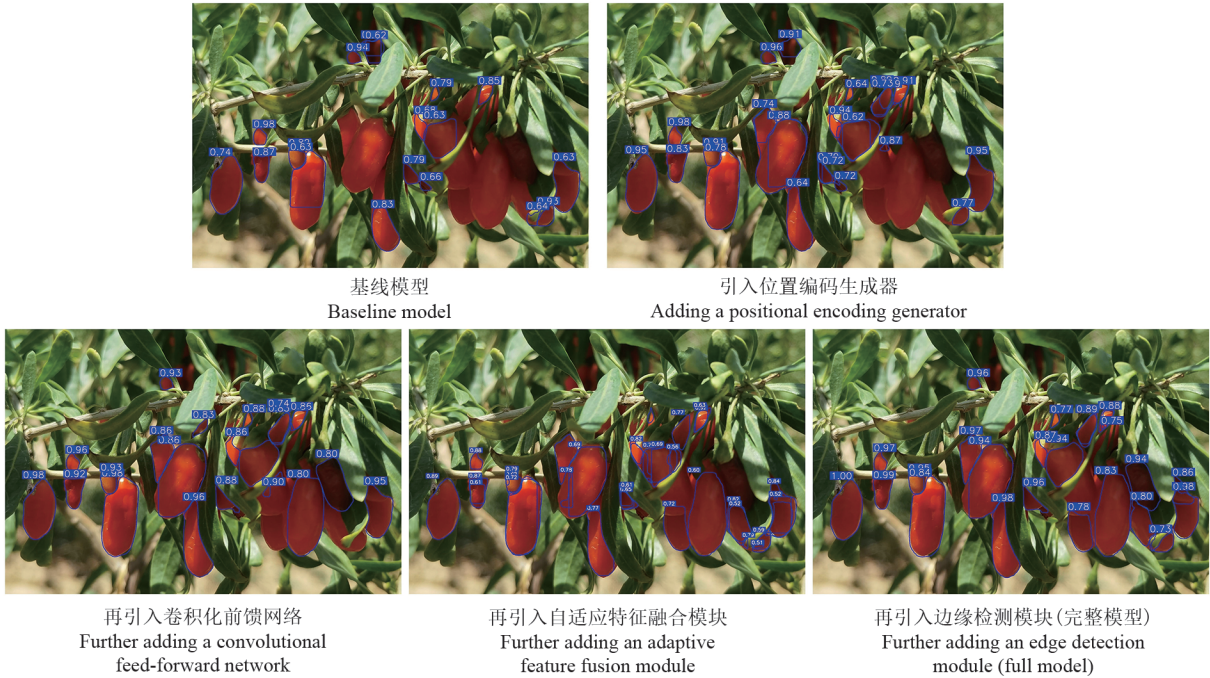


图 8 消融实验结果可视化
Fig. 8 Visualization of ablation experiment results

4 结论

本文从科研和生产实践出发,针对如何从图像中提取出完整、精细的单个枸杞果实颗粒形状这一需求,提出了一种融合多尺度特征的枸杞果

实颗粒分割网络 (Chinese Wolfberry Fruitlet Particles Instance Segmentation Network, CWISNet), 用于从图像中提取精确的枸杞果实颗粒形状。CWISNet 在特征捕捉方面,引入 Transformer 的全局建模能力与卷积网络的局部细节刻画能力,

获得了更具判别力的语义特征;在边缘信息刻画方面,结合 ResNet-34 的深层特征提取优势与空洞卷积的扩展感受野能力,生成稳定的边缘信息;随后将语义特征与边缘信息融合以构建更高质量的特征,最终通过多级候选框输出精细的枸杞果实轮廓。

在后续工作中,将针对 CWISNet 的模型结构进行优化,引入深度信息或多视角图像等,从三维视角缓解完全遮挡问题,进一步提升分割结果的质量,为现代农业领域提供高质量的技术和数据服务。

参考文献

- [1] Dai G, Fan J, Dewi C. ITF-WPI: Image and text based cross-modal feature fusion model for wolfberry pest recognition[J]. *Computers and Electronics in Agriculture*, 2023, 212:108129.
- [2] Xing Z, Wang Y, Qu A, et al. MFENet: Multi-scale feature extraction network for images deblurring and segmentation of swinging wolfberry branch[J]. *Computers and Electronics in Agriculture*, 2023, 215: 108413.
- [3] Wang J, Tan D, Sui L, et al. Wolfberry recognition and picking-point localization technology in natural environments based on improved YOLOv8n-Pose-LBD [J]. *Computers and Electronics in Agriculture*, 2024, 227:109551.
- [4] Gongal A, Amatya S, Karkee M, et al. Sensors and systems for fruit detection and localization: A review [J]. *Computers and Electronics in Agriculture*, 2015, 116:8-19.
- [5] 苗中华,沈一筹,王小华,等. 自然环境下重叠果实图像识别算法与试验[J]. *农业机械学报*, 2016, 47(06): 21-26.
- [6] 陈浩,任奕林. 大田经济作物性状研究中的图像分割技术应用[J]. *广东农业科学*, 2021, 48(11): 123-135.
- [7] Tian Y, Yang G, Wang Z, et al. Apple detection during different growth stages in orchards using the improved YOLO-V3 model[J]. *Computers and Electronics in Agriculture*, 2019, 157:417-426.
- [8] Lv J, Xu H, Han Y, et al. A visual identification method for the apple growth forms in the orchard[J]. *Computers and Electronics in Agriculture*, 2022, 197: 106954.
- [9] Wang C, Wang H, Han Q, et al. Strawberry detection and ripeness classification using yolov8+ model and image processing method[J]. *Agriculture*, 2024, 14 (5):751.
- [10] Wan S, Goudos S. Faster R-CNN for multi-class fruit detection using a robotic vision system[J]. *Computer Networks*, 2020, 168:107036.
- [11] Shen L, Su J, Huang R, et al. Fusing attention mechanism with Mask R-CNN for instance segmentation of grape cluster in the field[J]. *Frontiers in Plant Science*, 2022, 13:934450.
- [12] 姚青,谷嘉乐,吕军,等. 改进 RetinaNet 的水稻冠层害虫为害状自动检测模型[J]. *农业工程学报*, 2020, 36(15):182-188.
- [13] 王梁,侯义锋,贺杰. 基于 Mask-RCNN 的自然场景下油茶果目标识别与检测[J]. *中国农机化学报*, 2022, 43(12):148-154.
- [14] Huang X, Peng D, Qi H, et al. Detection and instance segmentation of grape clusters in orchard environments using an improved mask R-CNN model[J]. *Agriculture*, 2024, 14(6):918.
- [15] Yu Y, Zhang K, Yang L, et al. Fruit detection for strawberry harvesting robot in non-structural environment based on Mask-RCNN[J]. *Computers and Electronics in Agriculture*, 2019, 163:104846.
- [16] Jia W, Tian Y, Luo R, et al. Detection and segmentation of overlapped fruits based on optimized mask R-CNN application in apple harvesting robot [J]. *Computers and Electronics in Agriculture*, 2020, 172:105380.
- [17] Huang Y, Wang T, Basanta H. Using fuzzy mask R-CNN model to automatically identify tomato ripeness[J]. *IEEE Access*, 2020, 8:207672-207682.
- [18] Liu Y, Sun P, Wergeles N, et al. A Survey and Performance Evaluation of Deep Learning Methods for Small Object Detection[J]. *Expert Systems with Applications*, 2021, 172:114602.
- [19] Cong P, Li S, Zhou J, et al. Research on instance segmentation algorithm of greenhouse sweet pepper detection based on improved mask RCNN[J]. *Agronomy*, 2023, 13(1):196.
- [20] Liu H, Wang X, Zhao F, et al. Upgrading swin-B transformer-based model for accurately identifying ripe strawberries by coupling task-aligned one-stage object detection mechanism[J]. *Computers and Electronics in Agriculture*, 2024, 218:108674.
- [21] Xiao B, Nguyen M, Yan W. Fruit ripeness identification using transformers[J]. *Applied Intelligence*,

- 2023, 53(19):22488-22499.
- [22] 陈鸿坤,罗会兰. 多尺度语义信息融合的目标检测[J]. 电子与信息学报, 2021, 43(07):2087-2095.
- [23] 郑剑,贺朝辉,于祥春. 基于相邻特征融合与特征解耦的一阶段目标检测[J]. 北京航空航天大学学报, 2025, 51(04): 1205-1214.
- [24] Liu G, Zhang Y, Liu J, et al. An improved YOLOv7 model based on Swin Transformer and Trident Pyramid Networks for accurate tomato detection[J]. *Frontiers in Plant Science*, 2024, 15:1452821.
- [25] Charisis C, Argyropoulos D. Deep learning-based instance segmentation architectures in agriculture: A review of the scopes and challenges[J]. *Smart Agricultural Technology*, 2024, 8:100448.
- [26] Yang J, Jiao L, Shang R, et al. EPT-Net: Edge perception transformer for 3d medical image segmentation[J]. *IEEE Transactions on Medical Imaging*, 2023, 42(11):3229-3243.
- [27] He C, Li S, Xiong D, et al. Remote sensing image semantic segmentation based on edge information guidance[J]. *Remote Sensing*, 2020, 12(9):1501.
- [28] Sheng X, Kang C, Zheng J, et al. An edge-guided method to fruit segmentation in complex environments [J]. *Computers and Electronics in Agriculture*, 2023, 208:107788.
- [29] Jia W, Cao K, Liu M, et al. FCOS-EAM: An accurate segmentation method for overlapping green fruits[J]. *Computers and Electronics in Agriculture*, 2024, 226: 109392.
- [30] Zimmermann R S, Siems J N. Faster training of Mask R-CNN by focusing on instance boundaries[J]. *Computer Vision and Image Understanding*, 2019, 188:102795.
- [31] Du W, Liu P. Instance segmentation and berry counting of table grape before thinning based on AS-SwinT[J]. *Plant Phenomics*, 2023, 5:0085.
- [32] Li H, Xiong X, Liu C, et al. SFFNet: Staged feature fusion network of connecting convolutional neural networks and graph convolutional neural networks for hyperspectral image classification[J]. *Applied Sciences*, 2024, 14(6):2327.
- [33] Pei F, Li Z, Mei S, et al. Recognition of Chinese wolfberry images with windy and sandy noises using improved YOLOv8[J]. *International Journal of Agriculture and Biological Engineering*, 2025, 18(2): 239-247.
- [34] Wang R, Tan D, Ju X, et al. An object feature-based recognition and localization method for wolfberry[J]. *Sensors*, 2025, 25(11):3365.
- [35] Micikevicius P, Narang S, Alben J, et al. Mixed precision training[J]. *arXiv preprint arXiv: 1710.03740*, 2017.