

基于自适应卷积注意力机制的茶叶嫩芽分割与采摘点定位方法

杨长辉 张垚鑫 张尧尧 程露 徐春雨 刘彧希

(重庆理工大学机械工程学院, 重庆 400054)

摘要:为解决自然环境下茶叶嫩芽密集分布、多尺度特征以及小目标导致的识别与定位难题,本文提出融合改进注意力机制的茶叶嫩芽实例分割模型,并基于轮廓最低点查找算法(Contour bottom search algorithm, CBS)精准定位茶芽采摘点。首先,在YOLO v8n-seg主干网络引入改进的自适应卷积注意力机制(Selective kernel attention mechanism, SKAM),将空间金字塔快速池化(Spatial pyramid pooling-fast, SPPF)模块替换为感受野模块(Receptive field block, RFB),通过扩展感受野,增强模型对多尺度茶芽特征的捕获能力。其次,在颈部网络设计浅层特征融合结构,采用动态上采样算子(Up-sampling by dynamic sampling, Dysample)保留小目标细节,提高密集嫩芽分割精度。最后,基于mask轮廓分析提出最低点定位算法。试验结果表明,本文模型边界框检测指标(准确率93.9%、召回率87.2%、平均精度均值(mAP₅₀)93.6%)较基准模型分别提升3.7、6.5、4.0个百分点;掩膜分割指标(90.9%、84.1%、88.6%)相应提升4.9、7.4、6.0个百分点。在密集与尺度多变场景下,提出的茶芽采摘点定位算法平均定位准确率达到76.4%,验证了本文方法能够稳定识别密集多尺度茶芽并精确定位采摘点,为智能采茶装备开发提供了理论依据。

关键词: 茶叶嫩芽; 实例分割; 采摘点定位; 自然环境; YOLO v8

中图分类号: TP391.41

文献标识码: A

文章编号: 1000-1298(2026)14-0267-11

OSID:



Selective Kernel Attention Mechanism-based Method for Tea Buds Segmentation and Picking Point Localization

Yang Changhui Zhang Yaixin Zhang Yaoyao Cheng Lu Xu Chunyu Liu Yuxi

(College of Mechanical Engineering, Chongqing University of Technology, Chongqing 400054, China)

Abstract: Aiming to address the challenges of recognition and localization caused by the dense distribution, multi-scale features, and small target characteristics of tea buds in natural environments, a tea bud instance segmentation model that incorporated an improved attention mechanism was proposed. Additionally, a contour bottom search algorithm (CBS) was employed to accurately locate the picking points of tea buds based on the analysis of the contour's lowest point. Firstly, an improved selective kernel attention mechanism (SKAM) was introduced into the backbone network of YOLO v8n-seg. The spatial pyramid pooling-fast (SPPF) module was replaced with the receptive field block (RFB) module. By expanding the receptive field, the model's ability to capture multi-scale features of tea buds was enhanced. Secondly, a shallow feature fusion structure was designed in the neck network. The up-sampling by dynamic sampling (Dysample) operator was adopted to preserve the details of small targets, thereby improving the segmentation accuracy of densely distributed tea buds. Finally, a lowest point localization algorithm was proposed based on the analysis of the mask contour. The experimental results showed that the bounding box detection indicators of the proposed model (precision of 93.9%, recall of 87.2%, and mAP₅₀ of 93.6%) were improved by 3.7, 6.5, and 4.0 percentage points, respectively compared with that of the baseline model. The mask segmentation indicators (90.9%, 84.1%, and 88.6%) were correspondingly improved by 4.9, 7.4, and 6.0 percentage points. In scenarios with

收稿日期: 2025-04-02 修回日期: 2025-04-30

基金项目: 国家自然科学基金项目(52075062)、重庆市技术创新与应用发展项目(cstc2021jsex-gksbX0039)和重庆理工大学校级创新项目(gzlex20252014)

作者简介: 杨长辉(1976—),男,教授,博士,主要从事智能农业机械和机器视觉研究,E-mail: yangchanghui@cqut.edu.cn

dense distribution and varying scales, the average localization accuracy of the proposed tea bud picking point localization algorithm reached 76.4%. This verified that the method can stably identify dense tea buds of multiple scales and accurately locate the picking points, providing a theoretical basis for the development of intelligent tea picking equipment.

Key words: tea buds; instance segmentation; picking point location; natural environment; YOLO v8

0 引言

目前,传统的大宗茶叶采摘大多采用往复切割方式,但易出现芽叶破碎、误采、漏采等问题^[1]。然而,对于嫩芽品质要求高、采摘条件严苛的名优茶,机械化“一刀切”采摘方式难以满足精细化需求,名优茶嫩芽采摘仍普遍依赖人工方式,存在劳动强度大、采摘成本高等问题^[2]。

因此,研究茶芽准确识别与采摘点定位技术,实现名优茶嫩芽的智能无人化采摘,对降低劳动强度、提升采摘效率及推动我国智慧农业发展具有重要意义。然而茶叶嫩芽的识别场景通常具有复杂背景、光照变化、茶芽尺度不一以及分布密集等特点,传统图像识别方法难以适应上述采摘场景。随着深度学习技术快速发展,基于深度卷积神经网络的茶叶嫩芽识别方法逐渐成为研究热点^[3]。

Chen 等^[4]提出一种结合 YOLO v3 算法、语义分割、骨架提取和最小外接矩形的采摘点定位方法,试验结果表明,该方法在茶叶嫩芽识别任务中平均准确率达到 71.96%,采摘点提取成功率为 83%。Li 等^[5]开发的田间采茶机器人,由于受限于作业空间和定位误差,优质茶叶识别成功率为 85.16%,定位准确率为 78.9%。Li 等^[6]通过改进 YOLO v4 模型,将茶芽识别准确率提升 1.08 个百分点,计算复杂度降低 89.11%,但检测准确率为 85.15%。张晴晴等^[7]基于 YOLO v3 优化的识别模型在简单背景下识别精度达到 91%。朱红春等^[8]采用 Faster R-CNN 和 VGG19 网络结合的方法识别茶芽,试验结果表明,在复杂场景下识别性能显著优于传统图像处理方法。毛腾跃等^[9]提出显著性检测与 Grabcut 算法组合,在局部稀疏和目标背景相似茶芽场景中表现良好。周宇杰等^[10]基于 YOLO v3 和形态学方法确定茶芽采摘位置,采摘成功率为 75.53%。朱铭敏等^[11]开发的轻量化 YOLO v5s 模型成功部署于树莓派,平均精度均值为 73%。吕军等^[12]提出的光照自适应校正模型,在不同光照条件下检测精度达到 92.4%。黄海松等^[13]结合多尺度注意力机制与知识蒸馏的茶芽分级模型,对单芽、一芽一叶、一芽两叶的识别精度分别达到 100%、92.7%、89.89%。Wang 等^[14]采用 Mask R-CNN 方法,准

确率和召回率分别为 93.95% 和 92.48%,但采摘点微小特征仍影响定位效果。Gui 等^[15]通过轻量化改进 YOLO v5l 模型,茶芽识别精度为 92.66%,但计算量(5.689×10^{10})仍然较大。Chen 等^[16]改进 YOLO v7,引入多尺度自适应注意力机制和多特征融合模块,改进模型召回率和准确率分别达到 84.95% 和 90.99%。黄家才等^[17]和胡和平等^[18]分别优化 YOLO v4 和 YOLO v5s 模型,前者实现移动端部署,模型平均精度均值达 72.93%,后者改进 YOLO v5s 应用于嫩叶分级识别。蔡其锴^[19]改进 YOLO v5s 模型,参数量达到 1.32×10^7 ,由于茶芽生长较快,实际采摘场景复杂,田间试验采摘完好率为 60%。王梦妮等^[20]针对复杂背景、多目标场景改进 YOLO v5s 模型,茶芽识别平均精度达到 84.3%。宋彦等^[21]基于多头自注意力机制提出茶叶采摘点语义分割算法,在简单背景下茶芽检测像素准确率 达到 75.2%。林桂潮等^[22]设计优化的茶叶采摘末端执行器,采摘成功率达 93.3% ~ 95.0%,但未涉及茶芽识别与定位研究。

尽管上述研究在茶叶嫩芽识别与定位领域取得阶段性进展,但现有方法大多集中在尺度固定和局部稀疏场景,尚未解决茶叶嫩芽密集分布、尺度变化多样导致的特征提取困难,以及小目标场景下的识别精度不足与采摘点定位存在偏差的问题。针对上述茶叶嫩芽识别定位问题,本文提出一种融合改进注意力机制的实例分割模型与精准采摘点定位方法,基于改进 YOLO v8n-seg 模型提取嫩芽 mask 轮廓,并通过查找轮廓最低点实现采摘点精确定位。

1 材料与数据

1.1 数据采集

采集的茶叶图像分别来自重庆市璧山区三合茶园、綦江区雾垭茶园、巴南区定心茶园,采集对象为黄金茶芽、龙井长叶、巴南银针、定心绿茶等。图像采集设备为 Redmi Note 12 Turbo 型和 RealSense D435i 型深度相机,图像分辨率分别为 4 624 像素 × 3 472 像素和 640 像素 × 480 像素。对于自然生长茶叶,侧面识别茶芽效果优于顶部识别^[23],因此拍摄设备与茶树距离设置为 10 ~ 60 cm,同时相机光轴与竖直方向夹角为 10° ~ 90°,样本采集过程如

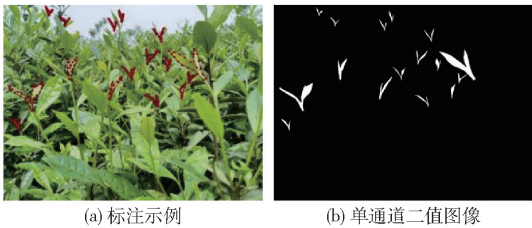
图 1 所示。为确保数据多样性,采集图像包括简单背景、复杂背景、茶叶嫩芽相互遮挡、光照强度变化以及茶叶嫩芽密集生长场景,共计 1 186 幅。图 2 为部分茶叶嫩芽图像示例。

1.2 数据集构建

为提取茶叶嫩芽边缘轮廓信息,以便采摘点定位,采用 labelme 工具对茶叶嫩芽进行标注,其中包括一芽一叶和一芽两叶,得到茶叶嫩芽训练样本和对应的单通道二值图,如图 3 所示。



图 2 茶叶图像样本多样性示例
Fig. 2 Tea image sample diversity example



(a) 标注示例
(b) 单通道二值图像
图 3 标注示例和单通道二值图像

Fig. 3 Labeled example and single-channel binary diagram

为增强模型泛化能力,从原始图像中随机选取 80% 进行数据增强,采用旋转、平移、添加噪声、调整

亮度和镜像处理等方法,将图像数据集扩充至 2 505 幅。增强后数据按比例 9: 1 随机划分为训练集和验证集。其余 20% 原始图像(共 237 幅)作为测试集。

2 茶叶嫩芽实例分割模型

2.1 YOLO v8n-seg 模型改进

基于 YOLO v8n-seg 模型,对主干网络和颈部网络进行改进,构建茶叶嫩芽实例分割模型 SRN-YOLO v8n-seg,网络结构如图 4 所示。主要改进包括:在主干网络,引入改进自适应卷积注意力机制

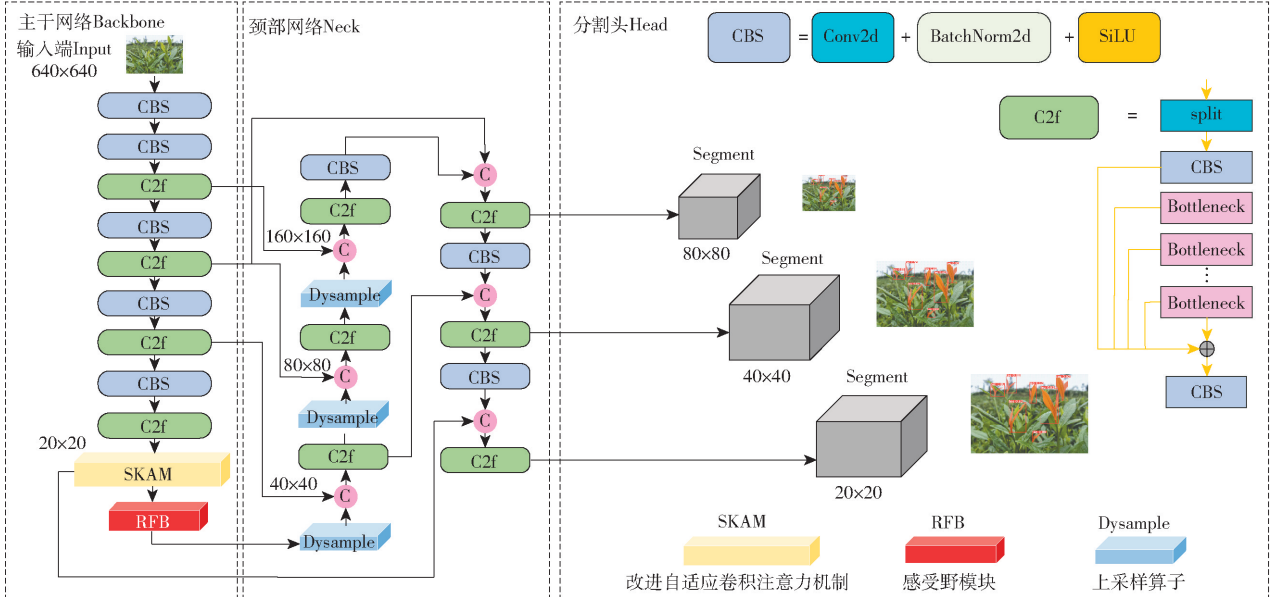


图 4 SRN-YOLO v8n-seg 模型结构
Fig. 4 Structure of SRN-YOLO v8n-seg

(Selective kernel attention mechanism, SKAM) 和感受野模块 (Receptive field block, RFB); Neck (颈部网络) 部分, 增加 1 个上采样层以融合浅层特征, 从而提升模型对小目标特征的提取能力。同时, 采用动态上采样算子 (Upsampling by dynamic sampling, Dysample) 替代模型原有的上采样器, 以提高模型在

密集场景下的分割性能。

2.1.1 改进自适应卷积注意力机制

针对自然环境中茶叶嫩芽密集分布和形态尺度多变的问题, 在主干网络中引入自适应卷积注意力机制 (SKAM), 并对 Split、Fuse 和 Select 3 个阶段进行改进, 改进后 SKAM 结构如图 5 所示。

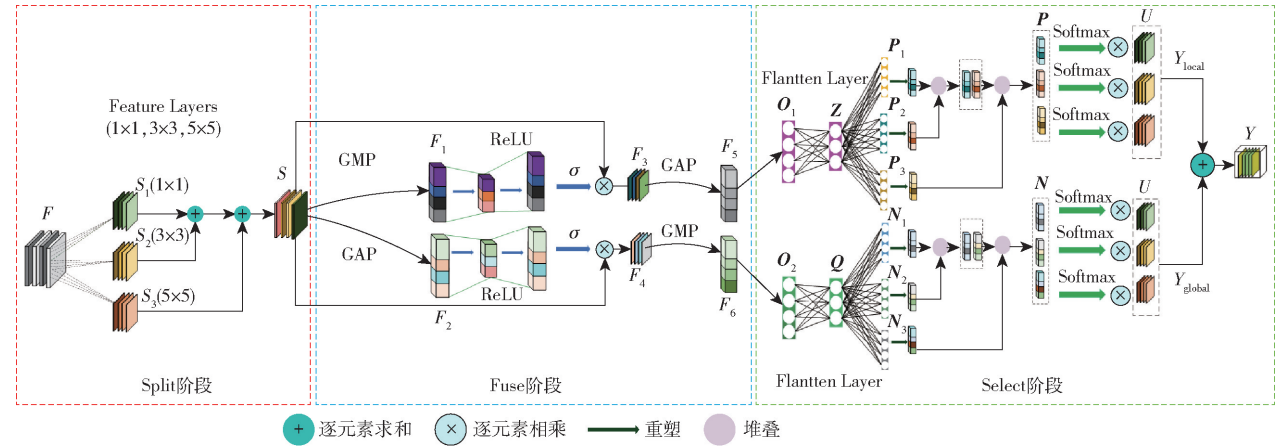


图 5 SKAM 结构

Fig. 5 SKAM structure

为有效捕捉茶叶嫩芽的多尺度特性并避免单一卷积核导致的特征局限, 采用多尺度卷积核集 K 对输入特征图 F 进行卷积处理, 分别生成不同尺度的特征图 S_1 、 S_2 、 S_3 , 随后采用逐元素求和融合多尺度特征图 S , 如图 5 Split 阶段。获得的特征图 S 维度为 $B \times C \times H \times W$, 其中 B 表示批次大小, C 为通道数, H 和 W 分别表示特征图高度和宽度。

SKAM 通过全局平均池化融合特征图 S 以获取全局上下文信息, 但可能丢失重要的局部细节特征。为解决这一问题, 对 SKAM 中 Fuse 操作进行改进, 如图 5 Fuse 阶段。具体采用双路径池化策略: 对多尺度特征图 S 分别进行全局最大池化 (Global max pooling, GMP) 和平均池化 (Global average pooling, GAP) 操作, 其中 GMP 生成的特征图 F_1 保留了局部细节信息, GAP 生成的特征图 F_2 获取了全局上下文信息。该设计有效避免单一全局平均池化可能导致的局部信息丢失问题。

针对池化后的特征图 F_1 和 F_2 , 采用 1×1 卷积层对特征图通道维度进行变换, 构建类多层感知器 MLP 模块。具体处理流程为: 首先使用 1×1 卷积层对特征图 F_1 和 F_2 进行通道降维。其次, 通过 ReLU 激活函数引入非线性表示, 并将通道数恢复至原始维度。最后, 应用激活函数 σ 对特征图通道进行归一化处理, 从而获得局部和全局信息通道权重 $W_{\text{local}} \in \mathbf{R}^{B \times C \times 1 \times 1}$ 和 $W_{\text{global}} \in \mathbf{R}^{B \times C \times 1 \times 1}$ 。

为获得局部和全局特征信息, 分别对局部通道权重 W_{local} 和全局通道权重 W_{global} 进行广播操作, 使维

度与原始多尺度特征图 S 保持一致, 然后将调整后的权重与特征图 S 进行逐元素相乘, 分别得到局部特征图 $F_3 \in \mathbf{R}^{B \times C \times H \times W}$ 和全局特征图 $F_4 \in \mathbf{R}^{B \times C \times H \times W}$ 。

全局平均池化能抑制噪声并平滑特征, 而全局最大池化有助于提取显著特征。在局部特征图 F_3 上应用 GAP 操作, 生成平滑特征图 F_5 维度为 $B \times C \times 1 \times 1$ 以抑制异常值干扰, 使 SKAM 更专注于关键特征表达; 同时在全局特征图 F_4 上应用 GMP 操作, 使 SKAM 能在抑制噪声后的特征上激活显著特征, 进而提取全局特征最大值, 生成特征图 F_6 维度为 $B \times C \times 1 \times 1$ 。

文献[24]中自适应卷积注意力机制通过单一全连接层实现特征自适应选择, 而本文则采用 2 个独立的全连接层实现自适应选择, 并通过逐元素相加的方式进行特征融合, 如图 5 Select 阶段所示。为实现 SKAM 自适应特征选择能力, 将特征图 F_5 和 F_6 分别重塑为 $B \times C$ 维特征张量, 记为 O_1 和 O_2 。对特征张量 O_1 和 O_2 , 分别经过 1 个全连接层处理后, 在空间维度上输出 $B \times L$ 维特征张量 Z 和 Q , 计算式为

$$Z = [p_1 \quad p_2 \quad p_L] \tag{1}$$

$$Q = [n_1 \quad n_2 \quad n_L] \tag{2}$$

$$p_i = \sum_{h=1}^C w_{i,h}^1 O_1 + b \tag{3}$$

$$n_i = \sum_{h=1}^C w_{i,h}^2 O_2 + b \tag{4}$$

$$L = \text{Max}(d, C/r) \tag{5}$$

式中 b ——偏置项
 $w^1、w^2$ ——隐藏层权重, $w^1、w^2 \in \mathbf{R}^{L \times C}$
 p_i ——特征图 $\mathbf{O}_1$ 第 i 个神经元输出
 n_i ——特征图 $\mathbf{O}_2$ 第 i 个神经元输出
 L ——输出张量通道数

式中压缩率 r 取 16, d 取 32。该设计在降低特征维度的同时减少了计算复杂度,并有效捕捉不同层级特征之间的交互关系,使 SKAM 能够协同关注全局和局部特征。

采用相同的全连接操作,从特征张量 $\mathbf{Z}$ 生成多个分支张量 $\mathbf{P}_1、\mathbf{P}_2、\mathbf{P}_3$,同时将其重塑为 $B \times C \times 1 \times 1$ 维。使用特征图堆叠操作,将 $\mathbf{P}_1、\mathbf{P}_2、\mathbf{P}_3$ 堆叠成特征张量 $\mathbf{P}$,其中 $\mathbf{P}$ 维度为 $K \times B \times C \times 1 \times 1$ 。类似地,从特征张量 $\mathbf{Q}$ 生成多个分支张量 $\mathbf{N}_1、\mathbf{N}_2、\mathbf{N}_3$,同时重塑为 $B \times C \times 1 \times 1$ 维,并将其堆叠为特征张量 $\mathbf{N}$,维度为 $K \times B \times C \times 1 \times 1, K$ 为 3。利用激活函数 Softmax 对特征张量 $\mathbf{P}$ 和 $\mathbf{N}$ 的每个通道进行加权处理,将权重压缩到区间 $[0, 1]$,以获得局部注意力和全局注意力权重,计算式为

$$y_{\text{local}} = \text{Softmax}(\mathbf{P}) \tag{6}$$

$$y_{\text{global}} = \text{Softmax}(\mathbf{N}) \tag{7}$$

式中 y_{local} ——局部注意力权重
 y_{global} ——全局注意力权重

将不同尺度卷积核提取的特征图 $S_1、S_2、S_3$ 堆叠为维度为 $K \times B \times C \times H \times W$ 的多尺度特征图 U 。为获得具有多尺度的局部特征图 and 全局上下文信息特征图,将注意力权重 y_{local} 和 y_{global} 分别与多尺度特征图 U 进行逐元素相乘,得到局部多尺度特征图 Y_{local} 和全局多尺度特征图 Y_{global} 。

分别对局部多尺度特征图 Y_{local} 和全局多尺度特征图 Y_{global} 沿着第 0 维进行求和操作,获得 $B \times C \times H \times W$ 维 2 个特征图,然后将 2 个特征图采用逐元素相加方式进行融合,获得一个具有多尺度的特征图 Y 。

2.1.2 感受野模块

如图 6 所示,感受野模块(RFB)采用多尺度卷积核(如 $3 \times 3、5 \times 5$)进行并行处理,并结合跳跃连接机制,通过逐层扩展感受野将底层特征传递至高

层,从而增强多尺度上下文信息的提取能力^[25]。针对茶叶嫩芽识别任务中尺度多变的特点,采用 RFB 替代 YOLO v8n-seg 中空间金字塔快速池化模块(Spatial pyramid pooling-fast, SPPF),以增强主干网络捕捉多尺度特征的能力,此外采用 2 个级联的 3×3 卷积核替代 5×5 卷积核,在保持等效感受野(5×5)的同时,减少计算量。

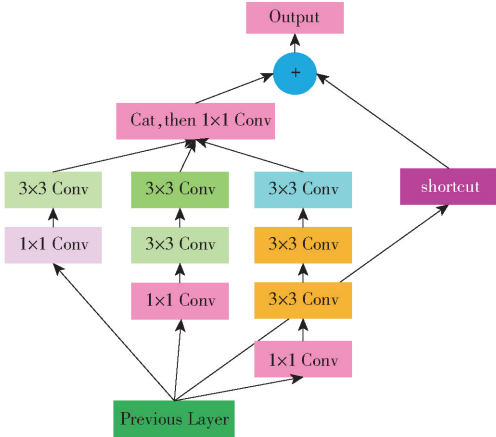


图 6 RFB 结构
Fig. 6 Receptive field block structure

2.1.3 Neck 网络优化

为提升模型对小目标茶芽识别能力,避免因过多卷积操作导致小目标特征丢失,在 Neck 部分引入主干网络的 P2 层特征图。P2 层保留丰富的小目标边缘、纹理和细节信息,从而改善模型分割性能,改进后颈部网络结构如图 7 所示。同时采用动态上采样算子(Dysample)替换颈部网络常规上采样器,网络结构如图 8 所示。该算子基于网络采样函数 Grid sample,通过结合动态采样集 γ 中位置信息,将维度为 $C \times H \times W$ 的特征图 χ 重新采样为 $C \times sH \times sW$ 的特征图 χ' 。其中,动态采样集 γ 是原始采样网络 ε 和偏移量 O 的和。偏移量 O 通过静态范围因子和像素打乱的方式生成^[26]。

2.2 结果与分析

2.2.1 模型训练环境

模型训练工作站配置:Inter(R) Xeon(R) W-2255 处理器,NVIDIA GeForce RTX 3090 GPU,Win10 系统,64 位,Python 3.8.19,Torch 1.13.1,Cuda

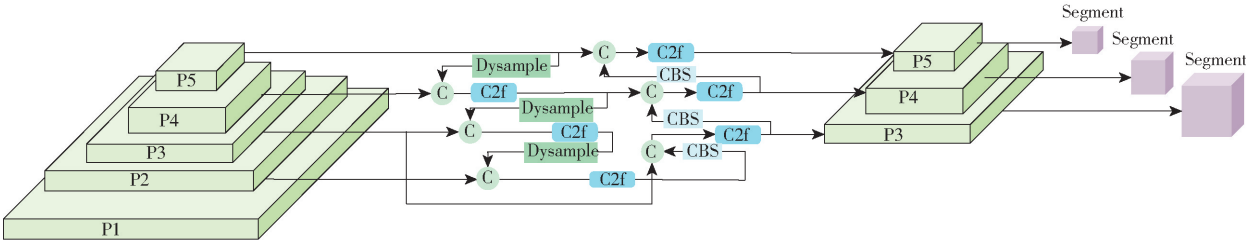


图 7 改进颈部网络结构
Fig. 7 Improved neck network structure

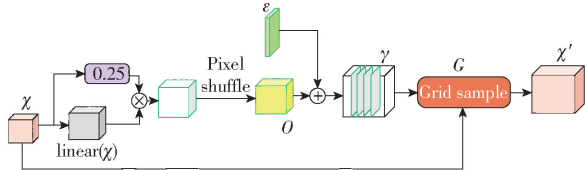


图8 Dysample 采样网络结构

Fig. 8 Structure of dynamic sampling network

11. 6. 模型训练总迭代次数为 300, 初始学习率设置为 0. 01, 优化器算法使用随机梯度下降, 设置学习率动量为 0. 9。此外, 优化器权重衰减系数为 0. 000 5, 不采用预训练权重。

2. 2. 2 模型评价指标

采用准确率 (Precision, P)、召回率 (Recall, R)、

平均精度均值 (mean Average Precision, mAP) 作为模型边界框 (box) 和掩码 (mask) 精度评价指标, 其中 mAP_{50} 为 IoU 为 0. 5 的平均精度均值, mAP_{50-95} 为区间 $[0. 5, 0. 95]$ 、步长 0. 05 的平均精度均值。同时以每秒帧数 (FPS) 作为推理速度指标, 以计算量和参数量作为模型复杂度指标。

2. 2. 3 消融试验

为验证改进方法有效性并分析各模块对模型的贡献, 以 YOLO v8n-seg 为基准模型进行消融试验。采用准确率、召回率、平均精度均值、推理速度和计算量作为主要评价指标, 结果如表 1 所示。

表 1 消融试验结果

Tab. 1 Results of ablation experiments

模型	SKAM	RFB	改进颈部网络	准确率/%		召回率/%		mAP ₅₀ /%		推理速度/ (f·s ⁻¹)	计算量
				边界框	掩码	边界框	掩码	边界框	掩码		
基准模型 (YOLO v8n-seg)	-	-	-	90. 2	86. 0	80. 7	76. 7	89. 6	82. 6	55. 2	1. 20 × 10 ¹⁰
模型 1	✓	-	-	90. 7	88. 4	83. 0	77. 2	90. 0	83. 4	46. 9	1. 39 × 10 ¹⁰
模型 2	-	✓	-	91. 5	87. 8	82. 2	78. 0	89. 4	82. 8	49. 8	1. 20 × 10 ¹⁰
模型 3	-	-	✓	91. 0	88. 5	83. 8	80. 1	91. 1	85. 7	45. 7	1. 31 × 10 ¹⁰
模型 4	✓	✓	-	93. 4	90. 6	85. 9	81. 8	91. 6	86. 4	48. 8	1. 39 × 10 ¹⁰
模型 5	-	✓	✓	88. 5	86. 6	85. 4	82. 2	90. 8	85. 6	42. 0	1. 31 × 10 ¹⁰
模型 6	✓	-	✓	89. 3	87. 4	84. 8	79. 1	90. 7	85. 2	39. 2	1. 51 × 10 ¹⁰
本文模型	✓	✓	✓	93. 9	90. 9	87. 2	84. 1	93. 6	88. 6	37. 4	1. 51 × 10 ¹⁰

注: ✓表示使用该模块, -表示不使用该模块。

由表 1 可知, 与基准模型相比, 单独引入 SKAM 模块后, 掩码准确率和边界框召回率分别提升 2. 4、2. 3 个百分点。模型 2 边界框准确率、召回率和掩码准确率、召回率相比基准模型分别提升 1. 3、1. 5 个百分点和 1. 8、1. 3 个百分点, 表明采用 RFB 替换 SPPF 模块能够扩大感受野, 有效提升模型的多尺度特征捕捉能力。在引入改进 Neck 结构后相比基准模型, 边界框准确率略有提高, 边界框召回率、mAP₅₀和掩码准确率、召回率、mAP₅₀分别提升 3. 1、1. 5 个百分点和 2. 5、3. 4、3. 1 个百分点, 证明浅层网络特征的引入显著改善了小目标识别性能。模型 4 结合 SKAM 和 RFB 模块后, 表现出更显著提升, 边界框准确率、召回率、mAP₅₀和掩码准确率、召回率、mAP₅₀相比基准模型分别提升 3. 2、5. 2、2 个百分点和 4. 6、5. 1、3. 8 个百分点, 说明结合 SKAM 和 RFB 可有效增强模型对多尺度特征的提取能力。而本文模型在所有评估指标上均展现出最显著的提升, 边界框准确率、召回率、mAP₅₀和掩码准确率、召回率、mAP₅₀相比基准模型分别提高 3. 7、6. 5、4 个百分点和 4. 9、7. 4、6 个百分点, 其中边界框召回率和掩码召回率显著提升, 表明模型漏检率大幅降低, 虽然计算量略有增加, 推理速度降低为 37. 4 f/s, 但仍

完全满足实时采摘作业要求。

2. 2. 4 不同模型性能

为验证本文模型性能优势, 将与当前主流的实例分割模型, 包括 YOLACT、YOLO v5n-seg、YOLO v7-seg、YOLO v8n-seg 进行对比试验。试验采用统一的数据集划分, 训练集 2 255 幅图像, 验证集 250 幅图像, 测试集 237 幅图像, 各模型性能结果如表 2 所示。

表 2 不同模型性能结果

Tab. 2 Performance results of different models

模型	mAP ₅₀ /%		参数量	计算量	推理速度/ (f·s ⁻¹)
	边界框	掩码			
YOLACT	77. 7	55. 2	3. 14 × 10 ⁷	6. 110 × 10 ¹⁰	33. 5
YOLO v5n-seg	88. 8	76. 6	1. 90 × 10 ⁶	6. 700 × 10 ⁹	42. 4
YOLO v7-seg	92. 8	85. 3	3. 78 × 10 ⁷	1. 419 × 10 ¹¹	36. 8
YOLO v8n-seg	89. 6	82. 6	3. 30 × 10 ⁶	1. 200 × 10 ¹⁰	55. 2
本文模型	93. 6	88. 6	5. 70 × 10 ⁶	1. 510 × 10 ¹⁰	37. 4

表 2 表明, 本文模型边界框和掩码 mAP₅₀ 最高, 分别达到 93. 6% 和 88. 6%, 相较于 YOLO v8n-seg 模型分别提升 4、6 个百分点。计算量明显低于 YOLO v7-seg 模型, 比 YOLO v8n-seg 模型略高, 计算量较小。

为定量描述茶叶嫩芽稀疏程度,引入茶叶嫩芽之间平均距离 $\bar{d}$ 、嫩芽数量 n 和尺度面积 s , 图像总面积 A (取 4 624 像素 $\times$ 3 472 像素), 茶叶嫩芽覆盖率 δ 和疏密度 D 可表示为

$$\delta = \frac{ns}{A}$$

(8)

$$D = 10\,000 \frac{\delta s}{\bar{d}^2}$$

(9)

多尺度是指图像中茶叶嫩芽的尺寸差异。为量化表征这一特征,通过计算嫩芽掩码 (mask) 像素面积定义目标尺度,具体尺度划分标准如表 3 所示。

表 3 不同尺度划分

Tab.3 Divisions of different scales 像素

尺度	面积
大尺度	$>3 \times 10^5$
中尺度	$2 \times 10^4 \sim 3 \times 10^5$
小尺度	$\leq 2 \times 10^4$

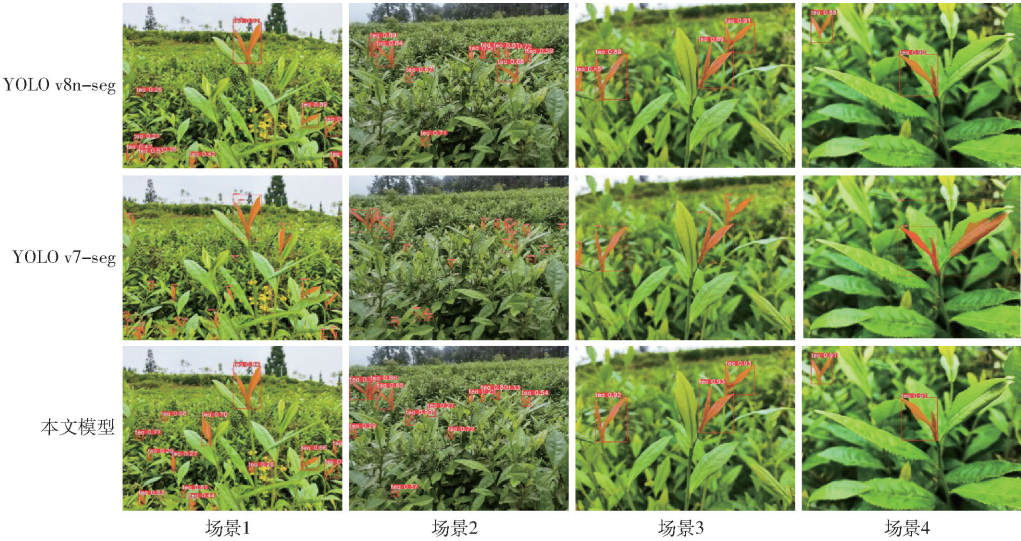


图 9 尺度变化条件下不同模型识别结果

Fig. 9 Identification results of different models under condition of scale change

在密集和遮挡条件下,本文模型也表现优异。如图 10 所示,场景 5 茶叶嫩芽数量超过 12 个,且茶芽间平均距离较小,疏密度较大,属于密集场景, YOLO v8n-seg 出现了漏分割和过分割, YOLO v7-seg 存在过分割情况,而采用 Dysample 替换 YOLO v8n-seg 的上采样器,增强了本文模型在密集场景中的分割能力。在场景 6 中,嫩芽数量少于 5 个,茶芽间平均距离较大,疏密度较小,属于稀疏场景, YOLO v7-seg 和本文模型分割表现较好, YOLO v8n-seg 在小目标嫩芽分割上存在漏分割现象,表明通过改进颈部网络增强了本文模型小目标的分割能力。场景 7 和场景 8 分别为茶芽相互遮挡和未遮挡情形。在茶芽相互遮挡的场景 7 中, YOLO v8n-seg 出现了过分割和误分割, YOLO v7-seg 出现了误分割现象。在未遮挡的场景 8 中, YOLO v8n-seg 出

图 9 为不同模型在尺度固定和尺度变化场景下的识别结果。在场景 1 和场景 2 中,目标具有不同尺度分布,包括大、中、小目标。例如,场景 1 中大目标茶芽 mask 面积大于 3×10^5 像素,小目标茶芽 mask 面积小于等于 2×10^4 像素,且这些目标呈混合分布;而在场景 3 和场景 4 中,茶叶嫩芽 mask 面积集中在 $2 \times 10^4 \sim 3 \times 10^5$ 像素之间,呈尺度固定分布,并且目标间存在部分遮挡或光照变化;在尺度变化场景 1 中,本文模型通过引入 SKAM 和 RFB,增强了多尺度目标识别能力。在尺度变化且以小目标为主的场景 2 中,本文模型漏分割最少,表明通过改进颈部网络后,增强了本文模型对小目标的识别能力。在场景 3 和场景 4 中, YOLO v7-seg 分别出现误分割和过分割情况, YOLO v8n-seg 在场景 3 中出现误分割情形,而本文模型在以上场景中识别效果较好。

现了误分割和漏分割,而 YOLO v7-seg 出现了误分割。上述结果表明,相比于 YOLO v7-seg 和 YOLO v8n-seg,在茶芽密集生长和遮挡场景中,本文模型对茶芽的特征提取和分割优势明显。

2.2.5 公开数据集验证

为进一步验证本文模型泛化能力,选择 Kaggle 公开数据集 Tea Buds。该数据集由浙江省杭州市茶叶研究所于 2022 年收集,图像内容涵盖近景、远景、俯视和侧视等多种拍摄角度,以及密集复杂场景。数据集包括 5 055 幅训练图像、562 幅验证图像和 625 幅测试图像^[27]。该数据集不适用于实例分割任务,但本文模型仅对 YOLO v8n-seg 主干网络和颈部网络进行改进,因此可以采用目标检测模型 YOLO v8n 和改进 SRN-YOLO v8n 验证本文模型的泛化能力。表 4 为 SRN-YOLO v8n 模型在公开数

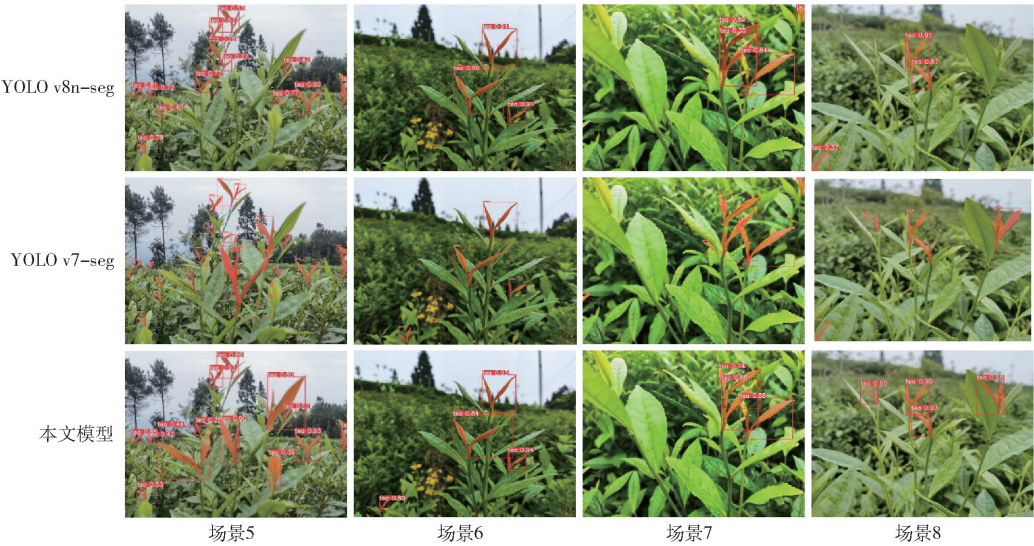


图 10 密集和遮挡条件下不同模型识别结果

Fig. 10 Identification results of different models under dense and occlusion conditions

表 4 不同模型在公开数据集性能

Tab. 4 Performance of different models on public datasets

模型	边界框准	边界框召	mAP ₅₀	mAP ₅₀₋₉₅
	准确率	回率		
YOLO v8n	83.6	73.6	82.7	59.7
SRN-YOLO v8n	85.8	74.1	83.0	63.1

据集性能,与 YOLO v8n 相比,SRN-YOLO v8n 边界框准确率提升 2.2 个百分点,表明预测结果可靠,误检少。边界框召回率和 mAP₅₀ 略有提高,而 mAP₅₀₋₉₅ 提升 3.4 个百分点,说明 SRN-YOLO v8n 在不同 IoU 阈值下综合性能更强,进一步验证了本文改进方法的优越性。

3 茶叶嫩芽采摘点定位方法

3.1 采摘点定位算法

根据茶叶嫩芽生长特性,采摘位置通常位于分割掩码(mask)最底部区域。因此,基于所提出的 SRN-YOLO v8n-seg 嫩芽分割模型,设计 mask 轮廓最低点查找算法(Contour bottom search algorithm, CBS),实现嫩芽采摘点定位。首先,通过分割模型生成 mask 并提取轮廓。然后,在像素坐标系 $O-x-y$ 中查找每个轮廓 y 坐标最大的点,该点对应于轮廓最底部,茶叶嫩芽采摘位置,图 11 为采摘点定位过程。

算法具体实现过程为:

设第 i 个 mask 轮廓由 N_i 个点组成,每个点坐标为 (x_k, y_k) ,则该轮廓可表示为

$$C_i = \{ (x_k, y_k) | k = 1, 2, \dots, N_i \}$$

式中 C_i ——第 i 个轮廓点集

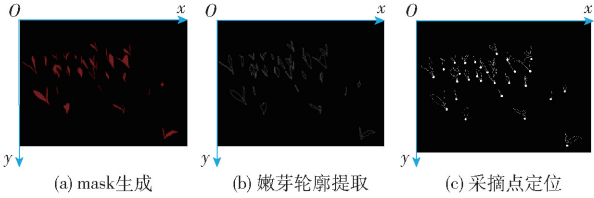


图 11 茶叶嫩芽采摘点定位过程

Fig. 11 Positioning process of tea bud picking points

N_i ——第 i 个轮廓点总数

查找 C_i 中纵坐标 y_k 最大的点索引,记为 $k_{\max}$,计算式为

$$k_{\max} = \operatorname{argmax}_k y_k \tag{11}$$

第 i 个轮廓上最低点对应的茶叶嫩芽采摘点 (x, y) 计算式为

$$(x, y) = (x_{k_{\max}}, y_{k_{\max}}) \tag{12}$$

3.2 采摘点定位结果

图 12 为在实际茶叶嫩芽图像中,采用本文定位方法定位的多个茶芽采摘点。每个采摘点用蓝色点标记,并按顺序标注采摘点像素坐标。通过该方式,能够精准地定位茶芽采摘的分布位置,为自动化采摘提供重要参考。

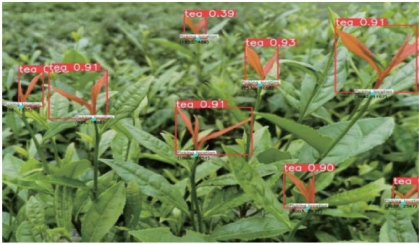


图 12 茶叶嫩芽采摘点定位结果

Fig. 12 Localization results of tea bud picking points

3.3 定位精度评价

采用终点误差 (Endpoint error, EPE) 和定位准

确率作为评估定位精度指标,终点误差定义为茶叶嫩芽理论采摘点(茶叶嫩芽外接矩形与嫩茎的交点位置)与预测点间欧氏距离^[28]。为更好地评估算法性能,将茶芽理论采摘点与预测点间横坐标距离定义为水平误差,纵坐标距离定义为垂直误差,计算式为

$$E_x = \sum_{k=1}^N \frac{\|x^t - x^p\|}{N_{\text{total}}} \tag{13}$$

$$E_y = \sum_{k=1}^N \frac{\|y^t - y^p\|}{N_{\text{total}}} \tag{14}$$

式中 x^t, y^t ——理论采摘点横、纵坐标
 x^p, y^p ——预测采摘点横、纵坐标
 E_x, E_y ——水平、垂直误差
 N_{total} ——统计茶芽数

平均定位准确率定义为成功定位的茶芽数量与茶芽总数比值。当预测采摘点的水平误差小于茶芽平均直径(50 像素)且垂直误差小于平均嫩茎长度(200 像素)时,视为定位成功。

从茶芽密集分布和尺度多变的图像中随机选取 35 幅,对其中的 250 个茶芽进行采摘点定位试验,统计结果如表 5 所示。

表 5 表明,本文提出的 CBS 定位方法能够在茶芽密集分布和尺度多变的场景下实现有效定位。与采用 YOLO v8n-seg 模型的定位方法(CBS + YOLO

表 5 密集和尺度多变场景下采摘点定位试验结果

Tab.5 Experiments on location results of picking points in dense and variable scale scenarios

定位方法	目标数/ 个	正确数/ 个	E_x / 像素	E_y / 像素	定位准 确率/%
YOLO v8n pose	250	87	39.2	80.8	34.8
CBS + YOLO v8n-seg	250	172	17.6	38.2	68.3
本文方法(CBS + SRN YOLO v8n-seg)	250	191	17.4	40.3	76.4

v8n-seg)相比,本文方法平均定位准确率提升 8.1 个百分点。虽然本文方法水平误差为 17.4 像素,但仍小于平均茶芽直径。垂直误差达到 40.3 像素,同样小于平均嫩茎长度,其整体误差保持在有效范围内。相比之下,YOLO v8n pose 模型平均定位准确率较低,仅为 34.8%,主要原因为该模型预测点超出了规定的可采摘区域,且垂直和水平误差均显著大于 CBS 定位方法。

图 13 为不同定位方法在稀疏和密集场景中定位效果。YOLO v8n pose 模型在稀疏场景中能够准确定位茶芽采摘点,但在密集场景中表现欠佳,出现定位点超出可采摘区域情况。基于 YOLO v8n-seg 模型的 CBS 定位方法在场景 9 和场景 12 中存在误

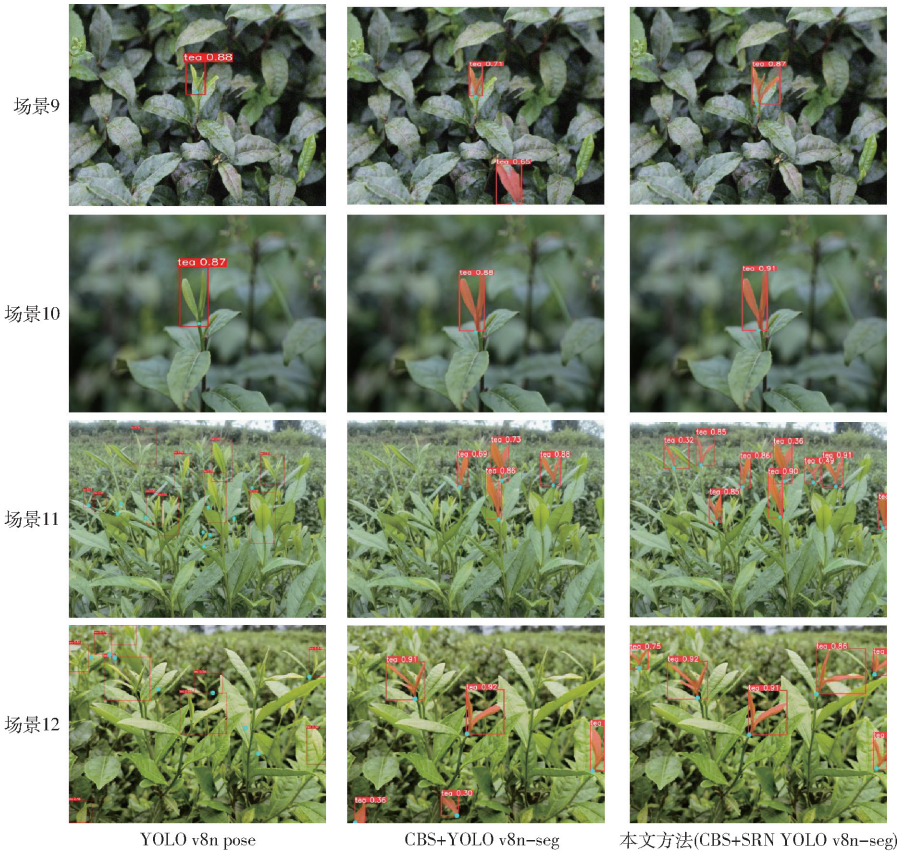


图 13 不同定位方法定位结果

Fig. 13 Location results of different localization methods

定位问题,在场景 11 中出现漏定位现象。相比之下,本文方法在所有测试场景中都表现出更高的定位精度。

尽管本文模型在密集场景和多尺度分割任务中表现出较好的鲁棒性和泛化性,但面向动态茶园采摘场景的工程需求,后续研究将围绕光照适应性增强、小样本学习与模型轻量化部署 3 方面开展,从而提升模型在采摘机器人实际作业中的适用性。

4 结论

(1) 在 YOLO v8n-seg 模型中引入 SKAM 和 RFB,有效增强了模型感受野,提升了模型多尺度感

知能力。通过在颈部网络引入浅层特征,并采用动态上采样算子替代原模型上采样器,使模型性能显著提升。与基准模型相比,本文模型边界框准确率、召回率、mAP₅₀和掩码准确率、召回率、mAP₅₀分别提升 3.7、6.5、4 个百分点和 4.9、7.4、6 个百分点。

(2) 通过对模型分割得到的 mask 进行轮廓提取,设计了轮廓最低点查找算法,定位茶叶嫩芽采摘点。试验结果表明:与基于 YOLO v8n-seg 模型的 CBS 定位方法相比,本文方法在平均定位准确率上提高 8.1 个百分点;与 YOLO v8n pose 模型的采摘点定位结果相比,本文方法在密集场景和多尺度场景下均展现出更好的适应性。

参 考 文 献

- [1] 李杨,董春旺,陈建能,等. 茶叶智能采摘技术研究进展与展望[J]. 中国茶叶,2022,44(7):1-9.
Li Yang, Dong Chunwang, Chen Jianneng, et al. Research progress and prospect of intelligent tea picking technology[J]. China Tea,2022,44(7):1-9. (in Chinese)
- [2] 陆德彪,尹军峰. 我国名优绿茶机械化采摘的意义与实现路径[J]. 中国茶叶,2018,40(1):1-4.
Lu Debiao, Yin Junfeng. The significance and realization path of mechanized picking of famous and high-quality green tea in China[J]. China Tea,2018,40(1):1-4. (in Chinese)
- [3] 高一聪,许晨,林琼,等. 茶叶生产装备自动化与智能化技术研究进展与展望[J]. 农业机械学报,2024,55(7):1-14.
Gao Yicong, Xu Chen, Lin Qiong, et al. Tea production equipment automation and intelligent technology research progress and prospects[J]. Transactions of the Chinese Society for Agricultural Machinery,2024,55(7):1-14. (in Chinese)
- [4] Chen C L, Lu J Z, Zhou M C, et al. A YOLO v3-based computer vision system for identification of tea buds and the picking point[J]. Computers and Electronics in Agriculture,2022,198:107116.
- [5] Li Y T, Wu S K, He L Y, et al. Development and field evaluation of a robotic harvesting system for plucking high-quality tea[J]. Computers and Electronics in Agriculture,2023,206:107659.
- [6] Li J, Li J H, Zhao X, et al. Lightweight detection networks for tea bud on complex agricultural environment via improved YOLO v4[J]. Computers and Electronics in Agriculture,2023,211:107955.
- [7] 张晴晴,刘连忠,宁井铭,等. 基于 YOLOV3 优化模型的复杂场景下茶树嫩芽识别[J]. 浙江农业学报,2021,33(9):1740-1747.
Zhang Qingqing, Liu Lianzhong, Ning Jingming, et al. Tea buds recognition under complex scenes based on optimized YOLOV3 model[J]. Acta Agriculturae Zhejiangensis,2021,33(9):1740-1747. (in Chinese)
- [8] 朱红春,李旭,孟炀,等. 基于 Faster R-CNN 网络的茶叶嫩芽检测[J]. 农业机械学报,2022,53(5):217-224.
Zhu Hongchun, Li Xu, Meng Yang, et al. Tea bud detection based on Faster R-CNN network[J]. Transactions of the Chinese Society for Agricultural Machinery,2022,53(5):217-224. (in Chinese)
- [9] 毛腾跃,张雯娟,帖军. 基于显著性检测和 Grabcut 算法的茶叶嫩芽图像分割[J]. 中南民族大学学报(自然科学版),2021,40(1):80-88.
Mao Tengyue, Zhang Wenjuan, Tie Jun. Image segmentation of tea buds based on salient object detection and Grabcut[J]. Journal of South-Central University for Nationalities(Natural Science Edition),2021,40(1):80-88. (in Chinese)
- [10] 周宇杰,吴强,贺磊盈,等. 名优茶采摘机器人的系统设计与试验[J]. 机械工程学报,2022,58(19):12-23.
Zhou Yujie, Wu Qiang, He Leiyong, et al. Design and experiment of intelligent picking robot for famous tea[J]. Journal of Mechanical Engineering,2022,58(19):12-23. (in Chinese)
- [11] 朱铭敏,张国平,谭建军,等. 基于 YOLO v5s 的轻量级茶叶嫩芽终端检测模型[J]. 浙江农业学报,2024,36(6):1413-1424.
Zhu Mingmin, Zhang Guoping, Tan Jianjun, et al. A lightweight tea buds terminal detection based on YOLO v5s[J]. Acta Agriculturae Zhejiangensis,2024,36(6):1413-1424. (in Chinese)
- [12] 吕军,方梦瑞,姚青,等. 基于区域亮度自适应校正的茶叶嫩芽检测模型[J]. 农业工程学报,2021,37(22):278-285.
Lü Jun, Fang Mengrui, Yao Qing, et al. Detection model for tea buds based on region brightness adaptive correction[J]. Transactions of the CSAE,2021,37(22):278-285. (in Chinese)
- [13] 黄海松,陈星燃,韩正功,等. 基于多尺度注意力机制和知识蒸馏的茶叶嫩芽分级方法[J]. 农业机械学报,2022,53(9):399-407,458.
Huang Haisong, Chen Xingran, Han Zhenggong, et al. Tea buds grading method based on multiscale attention mechanism and knowledge distillation[J]. Transactions of the Chinese Society for Agricultural Machinery,2022,53(9):399-407,458. (in Chinese)

[14] Wang T,Zhang K M,Zhang W, et al. Tea picking point detection and location based on Mask – RCNN [J]. Information Processing in Agriculture,2023,10(2):267 – 275.

[15] Gui Z Y,Chen J N,Li Y, et al. A lightweight tea bud detection model based on Yolov5 [J]. Computers and Electronics in Agriculture,2023,205:107636.

[16] Chen T C,Li H X,Chen J Z, et al. Detection network for multi-size and multi-target tea bud leaves in the field of view via improved YOLO v7 [J]. Computers and Electronics in Agriculture,2024,218:108700.

[17] 黄家才,唐安,陈光明,等. 基于 Compact – YOLO v4 的茶叶嫩芽移动端识别方法 [J]. 农业机械学报,2023,54(3):282 – 290.
Huang Jiakai, Tang An, Chen Guangming, et al. Mobile recognition solution of tea buds based on Compact – YOLO v4 algorithm [J]. Transactions of the Chinese Society for Agricultural Machinery,2023,54(3):282 – 290. (in Chinese)

[18] 胡和平,吴明晖,洪孔林,等. 基于改进 YOLO v5s 的茶叶嫩芽分级识别方法 [J]. 江西农业大学学报,2023,45(5):1261 – 1272.
Hu Heping, Wu Minghui, Hong Konglin, et al. Classification and recognition method for tea buds based on improved YOLO v5s [J]. Acta Agriculturae Universitatis Jiangxiensis,2023,45(5):1261 – 1272. (in Chinese)

[19] 蔡其锴. 名优茶采摘机器人的视觉识别和采摘技术研究 [D]. 成都:电子科技大学,2024.
Cai Qikai. Research on visual recognition and picking technology of tea picking robot [D]. Chengdu:University of Electronic Science and Technology of China,2024. (in Chinese)

[20] 王梦妮,顾寄南,王化佳,等. 基于改进 YOLO v5s 模型的茶叶嫩芽识别方法 [J]. 农业工程学报,2023,39(12):150 – 157.
Wang Mengni, Gu Jìnan, Wang Huajia, et al. Method for identifying tea buds based on improved YOLO v5s model [J]. Transactions of the CSAE,2023,39(12):150 – 157. (in Chinese)

[21] 宋彦,杨帅,郑子秋,等. 基于多头自注意力机制的茶叶采摘点语义分割算法 [J]. 农业机械学报,2023,54(9):297 – 305.
Song Yan, Yang Shuai, Zheng Ziqiu, et al. Semantic segmentation algorithm based multi-headed self-attention for tea picking points [J]. Transactions of the Chinese Society for Agricultural Machinery,2023,54(9):297 – 305. (in Chinese)

[22] 林桂潮,郑晓鑫,曾文勇,等. 欠驱动式茶叶采摘末端执行器优化设计与试验 [J]. 农业机械学报,2025,56(1):133 – 140.
Lin Guichao, Zheng Xiaoxin, Zeng Wenyong, et al. Optimization design and testing of under-actuated tea harvesting end-effector [J]. Transactions of the Chinese Society for Agricultural Machinery,2025,56(1):133 – 140. (in Chinese)

[23] Xu W K,Zhao L G,Li J, et al. Detection and classification of tea buds based on deep learning [J]. Computers and Electronics in Agriculture,2022,192:106547.

[24] Li X, Wang W H, Hu X L, et al. Selective kernel networks [C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition(CVPR),2019:510 – 519.

[25] Liu S T, Huang D, Wang Y H, et al. Receptive field block net for accurate and fast object detection [C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition(CVPR),2018:404 – 419.

[26] Liu W Z, Lu H, Fu H T, et al. Learning to upsample by learning to sample [C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition(CVPR),2023:6004 – 6014.

[27] Xie S,Sun H W. Tea – YOLO v8s: a tea bud detection model based on deep learning and computer vision [J]. Sensors, 2023, 23:6576.

[28] Wang X M,Wu Z L,Fang C. TeaPoseNet: a deep neural network for tea leaf pose recognition [J]. Computers and Electronics in Agriculture,2024,225:109278.