

基于实例分割的大场景下茶叶嫩芽轮廓提取与采摘点定位

李翰林，高延峰^{*}，熊根良，李昀谦，杨雅心

(上海工程技术大学机械与汽车工程学院，上海 201620)

摘要：为了解野外大场景下茶叶嫩芽识别与采摘点精确定位问题，提出了一种基于实例分割的 Yolov5s-segment 改进算法。该算法首先引入 P2 微小目标检测层，解决原始 Yolov5s-segment 网络 P3、P4、P5 检测层对于小目标检测能力不佳的问题。其次，在主干网络末端增加 CBAM(convolutional block attention module) 注意力机制模块，提升模型的抗干扰能力，实现野外自然光照环境下茶叶嫩芽轮廓特征提取。最后，根据嫩芽轮廓特征进行采摘点精确定位。研究结果表明，相较于原始 Yolov5s-segment 模型，改进模型的精确度、召回率、F1 分数、平均精度均值 mAP_{50} 和 mAP_{50-95} 分别提升了 7.0、8.9、8.1、8.3、7.3 个百分点。使用该方法可以准确提取大场景下的单芽、一芽一叶、一芽两叶三种类型茶叶嫩芽轮廓，并且实现采摘点的精确定位。研究结果为名优茶智能化快速采摘提供了一定的理论基础。

关键词：图像处理；图像识别；茶叶；嫩芽；Yolov5s-segment；轮廓提取；采摘点定位

doi: 10.11975/j.issn.1002-6819.202404016

中图分类号: S23

文献标志码: A

文章编号: 1002-6819(2024)-15-0135-08

李翰林，高延峰，熊根良，等. 基于实例分割的大场景下茶叶嫩芽轮廓提取与采摘点定位[J]. 农业工程学报，2024，40(15): 135-142. doi: 10.11975/j.issn.1002-6819.202404016 <http://www.tcsae.org>

LI Hanlin, GAO Yanfeng, XIONG Genliang, et al. Extracting tea bud contour and location of picking points in large scene using case segmentation[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2024, 40(15): 135-142. (in Chinese with English abstract) doi: 10.11975/j.issn.1002-6819.202404016 <http://www.tcsae.org>

0 引言

目前，名优茶主要依赖于人工采摘^[1]。名优茶采摘一般集中在清明前后较短的时期内，人工采摘效率较低，难以适应短期内的集中采摘需求。而机械大批量采摘只适用于大宗茶的采摘，无法实现精确采摘。因此，实现名优茶的智能化采摘成为当前急需解决的问题。

目前，在名优茶自动化采摘研究中，茶叶嫩芽识别是主要的研究内容之一^[2-3]。杨福增等^[4]通过提取图像的 G 分量，对图像进行阈值分割，最后基于嫩芽的形状特征进行边缘检测，达到了 94% 的准确率，但该方法易受到外界环境的影响。吴雪梅等^[5]使用 RGB 颜色模型，利用目标和背景的颜色信息差异实现嫩芽的识别。SHAO 等^[6]通过改进的 K-means 聚类算法实现嫩芽的分割和识别。龙樟等^[7]设计了基于超绿特征的茶叶嫩芽图像分割方法，提取茶丛图像的超绿特征，采用大津法 (OTSU) 进行阈值分割，并通过闭运算去除噪声，经色彩合并获得嫩芽分割图像。

随着人工智能的发展，基于深度学习的茶叶嫩芽识别也取得了一定的进展。基于深度学习的目标识别算法可分为单阶段识别算法和两阶段识别算法，单阶段识别算法主要有 SSD 系列和 Yolo 系列算法，两阶段识别算法主要是 R-CNN 系列算法。张雯娟^[8]通过相邻高低层特

征融合方式对传统 SSD 网络进行改进，提高了传统 SSD 模型对茶叶嫩芽识别的精度。吕军等^[9]通过计算图像的平均灰度值进行亮度自适应校正，结合 Yolov5 识别模型实现嫩芽识别，提高了不同光照条件下的识别精度以及召回率。CAO 等^[10]提出了一种结合 Yolov5 和 GhostNet 的嫩芽识别模型，引入 GhostNet 模块提高了识别速度。胡和平等^[11]基于改进 Yolov5s 模型实现了茶叶的分级识别，实现了识别效果和识别速度的均衡。洪孔林等^[12]提出了一种基于 Yolov7-tiny 的嫩芽分级识别模型，提升了复杂环境下模型的识别能力。CHEN 等^[13]采用 Faster R-CNN 深度学习模型对不同品种的茶叶嫩芽进行识别，达到了一定的识别精度。朱红春等^[14]研究了基于 Faster R-CNN 模型和 VGG16 特征提取网络在复杂背景下茶叶嫩芽类型（单芽、一芽一叶、一芽两叶）的识别精度，结果表明深度学习算法具有良好的识别效果。YAN 等^[15]提出一种基于 Mask R-CNN 的识别模型，进行一芽一叶的目标识别，对于目标重叠任务达到了一定的效果。

在识别茶叶嫩芽之后，进一步确定茶叶嫩芽的采摘位置是完成嫩芽采摘的重要环节。张浩等^[16]提出了一种基于主动视觉的嫩梢定位方法，通过色彩因子法和光栅投影轮廓术实现了自然环境下茶叶嫩梢识别以及嫩梢高度的获取。YAN 等^[15]提出了一种基于 Mask R-CNN 的识别模型，实现了嫩芽的分割，通过获得的嫩芽掩膜信息以及最小外接矩形，完成了采摘点的定位。李亚涛^[17]基于 Yolov3 算法进行嫩梢区域目标框定位，并采用聚类算法分离非嫩梢区域的点云信息，最后通过圆柱包络方法得到嫩梢的采摘点三维坐标信息。

然而，上述研究主要集中在局部视场环境下茶叶嫩芽的识别与定位方面。为了提高采摘效率，如何实现大

收稿日期: 2024-04-02 修订日期: 2024-06-22

基金项目: 上海市地方高校建设项目 (23010501600)

作者简介: 李翰林，研究方向为机器视觉和深度学习。

Email: lihanlin3833@163.com

^{*}通信作者: 高延峰，博士，教授，研究方向为机器视觉，移动操作机器人，机器人传感与智能控制等。Email: gyf_2672@163.com

场景下多个茶叶嫩芽的同时识别与采摘点定位是需要解决的关键问题。为此,本文提出一种基于实例分割的 YOLOv5s-segment 改进网络模型,通过增加小目标检测层和 CBAM 注意力模块解决图像中多个密集小目标的识别问题,实现大场景自然光照环境下茶叶嫩芽识别与定位。

1 数据集的建立

1.1 茶叶嫩芽状态特征分析

自然环境下的嫩芽形态多种多样,根据茶叶生长周期,可以将茶叶嫩芽分为单芽、一芽一叶、一芽两叶等^[18]多种状态,如图 1 所示。在实际采摘中,单芽并非意味着最好,本文以名优茶采摘过程中需求较高的单芽、一芽一叶、一芽两叶为例进行研究。

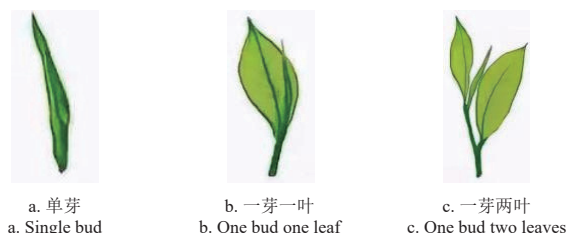


图 1 茶叶嫩芽状态

Fig.1 The states of tea bud

1.2 图像采集

为保证数据集的真实有效性,本文数据集图像于 2023 年和 2024 年 4 月中下旬在上海市佘山茶园以及上海沪枫茶园实地拍摄,茶叶品种主要为龙井和大红袍。拍摄使用高清相机固定拍摄,照片像素为 3 000×4 000 像素,为了减少训练内存消耗,将图片统一缩小为 640×640 像素。拍摄时间选择了中午 11 点至下午 1 点和下午 4 点至 5 点两个时间段。拍摄的茶叶图像如图 2 所示。



图 2 茶叶嫩芽实景拍摄

Fig.2 Tea buds under real scene shooting

图 2a 和图 2b 表示从不同距离下的拍摄效果,近景除了嫩芽的颜色特征外,还能让模型更好的获取嫩芽纹理特征。图 2c 和图 2d 表示在不同光照条件下的拍摄效果,前者光线明亮,后者光线较暗,可以增强模型对不

同天气条件的适应能力。图 2e 和图 2f 表示从不同角度下的拍摄效果,其中 0°和 45°拍摄更符合实际场景应用,嫩芽的轮廓在这两个角度也是最为完整和清晰的。

为了提高训练模型的准确度,将采集图像进行适当筛选,去除画质模糊和嫩芽区域有明显反光等因素造成目标不清晰的图像,最终保留 524 张原始图像。

1.3 数据集的建立和标注

采用图像增强的方法对数据集进行扩展。本研究使用的图像增强方法主要包括几何变换(旋转)、颜色空间变换(亮度变化)和像素点操作(添加噪声和随机点),其中几何变换模拟不同角度拍摄效果,颜色空间变换模拟不同天气条件下的拍摄效果,像素点操作模拟相机不稳定而引入噪声的效果。采用这几种变换的随机组合进行图像增强,变换后的示例如图 3 所示。



图 3 图像增强后的示例

Fig.3 Examples of enhanced images

经过图像增强处理后,数据集的图像一共有 2 096 张,将图像数据集随机分为 3 组,分别构成训练、验证和测试数据集。利用 Python 语言编写的轻量化标注工具 labelme 对茶叶图像数据进行标注,其中单芽、一芽一叶、一芽两叶标签分别为“SINGLE”、“YI”、“LIANG”,表 1 是最终建立的数据集具体信息,图 4 是原始图像和标注后图像的示例。

表 1 数据集具体信息

Table 1 Data set specific information

类别	图像	标签	单芽标	一芽一叶标签	一芽两叶标签
Category	Images/张	Labels	签 Labels of single bud	Labels of one bud with one leaf	Labels of one bud with two leaves
数据集	2 096	10 135	4 025	4 207	1 903
训练集	1 689	8 722	3 484	3 655	1 583
验证集	307	1 413	541	552	320
测试集	100	×	×	×	×

注: ×表示未对测试集进行标注。

Note: × indicates that the test set is not annotated.

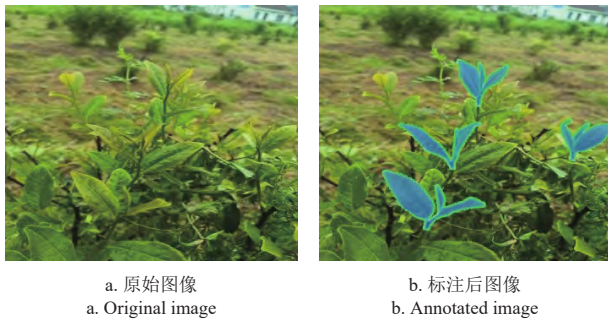


图 4 原始图像和标注后图像
Fig.4 Original image and annotated image

2 茶叶嫩芽图像分割与采摘点定位

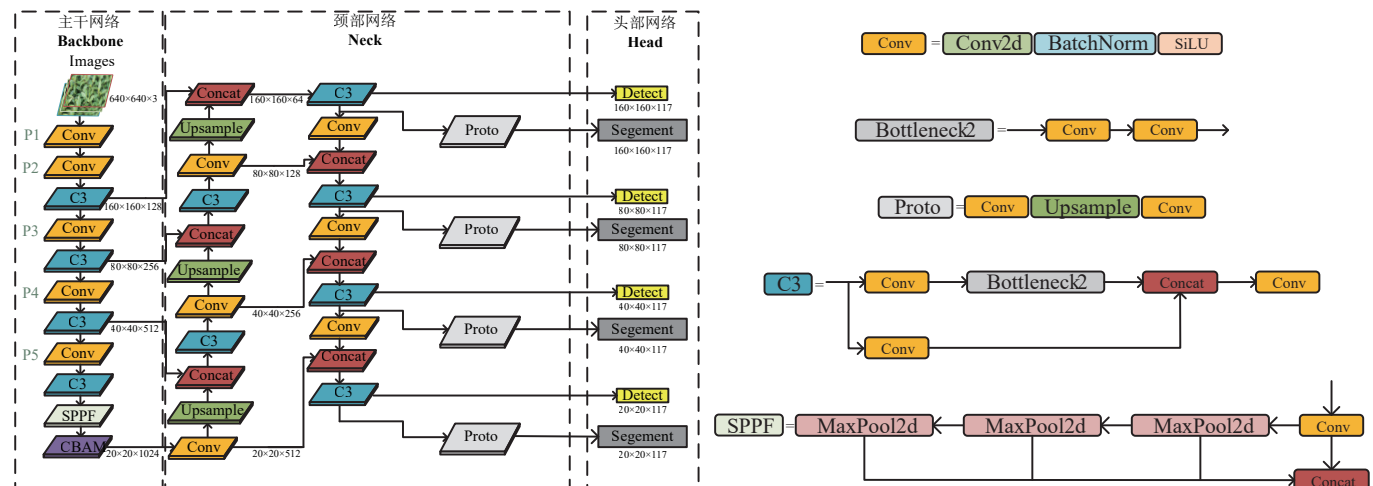
2.1 Yolov5s-TB 实例分割模型

Yolov5s-segment 的分割模块采用了和 Yoloact 类似的思想^[19]，其具有较高的成熟性和稳定性。与双阶段实例分割算法相比，其在保证一定精度的同时具有更快的检测速度并且能够实现视频实时检测。在实际的应用场景中，Yolov5s-segment 网络模型凭借其易于部署、集成简单的设计结构以及检测实时性好等优点成为了一个较好的选择^[20-21]。

Yolov5s-segment 整体由主干网络（Backbone）、颈部网络（Neck）和头部预测网络（Head）构成。Yolov5s-

segment 使用 CSPDarknet53 作为其主干网络，其具有较强的特征提取能力和计算效率，但深层网络的感受野有限，特别是对小目标检测与分割的准确性有一定的影响，需要多尺度的特征融合来解决这个问题。颈部网络（Neck）使用的是 FPN（feature pyramid network）和 PAN（path aggregation network）结构，FPN 结构和 PAN 结构可以增强模型的感知能力和多尺度特征的融合能力。与 Yolov5s 的不同之处是，Yolov5s-segment 使用与 Yoloact 类似的思想，在头部预测网络（Head）中增加了掩膜分割生成分支（Proto），生成 k 个 mask 原型图（prototype masks），然后将 mask 原型图与目标检测分支生成的 mask 原型图置信度（mask coefficients）进行线性组合，生成实例分割结果，将检测任务划分为目标检测和实例分割两个并行任务。

为了解决原始 Yolov5s-segment 网络模型在野外大场景下茶叶嫩芽检测方面的不足，本文提出了 Yolov5s-TB 实例分割模型。首先引入 P2 微小目标检测层，解决原始 Yolov5s-seg 网络 P3、P4、P5 检测层对于小目标检测能力不佳的问题。其次，在 SPPF 空间金字塔池化结构后，引入融合了通道注意力机制和空间注意力机制的 CBAM 模块，对嫩芽特征图进行更好的融合，加强特征的表示，提高模型的检测性能。本文提出的 Yolov5s-TB 网络模型如图 5 所示。



注：Conv 为卷积操作，BatchNorm 为归一化操作，SiLU 为激活函数，UpSample 为上采样操作，MaxPool2d 为最大池化操作，Concat 为拼接操作，Proto 为掩膜生成分支，P1~P5 为主干网络不同尺度的特征层。

Note: Conv is the convolution operation, BatchNorm is the normalization operation, SiLU is the activation function, UpSample is the upsample operation, MaxPool2d is the maximum pooling operation, Concat is the splicing operation, Proto is the mask generation branch, P1-P5 are the characteristic layers of different scales in the Backbone.

图 5 Yolov5s-TB 网络模型
Fig.5 Yolov5s-TB network model

2.1.1 增加微小目标检测层

传统检测方法对微小目标可能会出现误检或者漏检的情况。在 Yolov5s-seg 中有 P3、P4、P5 三个检测层，它们分别代表对小尺度（8×8 像素）、中尺度（16×16 像素）、大尺度（32×32 像素）物体的检测能力，其特征层大小分别为 [80,80]、[40,40]、[20,20]。输入特征图尺寸为 [640,640]，P3、P4、P5 三个检测层相较于输入特

征图尺寸分别下采样了 8 倍、16 倍、32 倍，下采样的倍数越大，特征图就会存在特征少以及语义信息不显著的问题，对于小目标的检测能力会降低。因此在原始网络结构中引入了 P2 检测层，其特征层大小为 [160,160]，相比于输入特征图尺寸下采样了 4 倍，可以加强对特征的提取，保留更多的特征和语义信息。

原始颈部网络（Neck）没有 160×160 像素的特征图，

因此对 Neck 中 80×80 像素的特征图进行一次卷积 (Convolution) 和上采样 (Upsample) 操作得到 160×160 像素的特征图, 并与 Backbone 中 P2 检测层生成的特征图进行一次张量拼接 (Concat) 操作, 最后通过 C3 模块进行处理, 最终获得小目标的检测和分割结果。微小目标检测层在提出模型中的结构如图 6 所示。

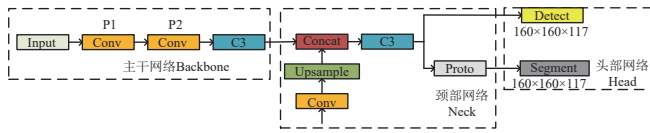


图 6 微小目标检测层
Fig.6 Micro target detection layer

2.1.2 CBAM(convolutional block attention module) 注意力机制

CBAM^[22] 是一种简单而有效的前馈卷积神经网络注意力模块。给定一个中间特征映射, CBAM 模块沿着通道和空间两个独立的维度依次推断注意力映射, 然后将注意力映射乘以输入特征映射进行自适应特征细化。它是一个轻量级的通用模块, 可以集成到任何一个 CNN 架构中。CBAM 由通道和空间两个顺序子模块构成, 其整体结构如图 7 所示。

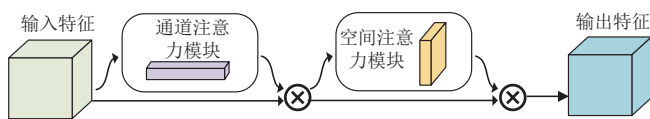


图 7 CBAM 整体结构
Fig.7 Overall structure of CBAM(convolutional block attention module)

给定一个中间特征图作为输入, CBAM 依次导出一维通道注意图和二维空间注意图, 整个注意力机制的过程可以被概括为式 (1):

$$\begin{aligned} F' &= M_c(F) \otimes F \\ F'' &= M_s(F') \otimes F' \end{aligned} \quad (1)$$

式中 F 为输入特征图, M_c 为一维通道注意图, M_s 为二维空间注意图, \otimes 代表逐元素的乘法, F' 是一维通道注意图和输入特征图逐元素相乘的结果, F'' 是 CBAM 模块最终的精炼输出。

通道注意力模块首先通过平均池化 (average pooling) 和最大池化 (max pooling) 操作聚合输入特征图 F 的空间信息, 生成特征向量, 将特征向量输入共享网络 (Shared MLP), 得到减少参数后的特征向量, 然后通过 softmax 函数进行特征合并, 得到包含通道注意力的特征向量, 最后将其与输入特征图 F 逐元素相乘得到 F' , 传递给空间注意力模块。通道注意力模块可以被概括为式 (2)。

$$\begin{aligned} M_c(F) &= \sigma(\text{MLP}(\text{AvgPool}(F)) + \text{MLP}(\text{MaxPool}(F))) \\ &= \sigma(W_1(W_0(F_{\text{avg}}^C)) + W_1(W_0(F_{\text{max}}^C))) \end{aligned} \quad (2)$$

式中 σ 表示 Sigmoid 激活函数, AvgPool 是平均池化操作, MaxPool 是最大池化操作, MLP 是一种应用于平均池化特征的多层感知器, W_0 和 W_1 是其操作权重, F_{avg}^C 和 F_{max}^C

是通道注意力模块平均池化和最大池化结果。

空间注意力模块首先对特征图 F' 使用平均池化和最大池化操作, 得到两个 $1 \times H \times W$ 的特征图, 将它们拼接在一起得到一个 $2 \times H \times W$ 的特征图。然后通过一个 7×7 的卷积重新得到 $1 \times H \times W$ 的特征图。最后通过一个 sigmoid 函数, 得到包含空间注意力的特征图, 将其与特征图 F' 逐元素相乘, 就得到了与输入特征图大小相同且包含通道和空间注意力的特征图 F'' 。空间注意力模块可以被概括为式 (3)。

$$\begin{aligned} M_s(F) &= \sigma(f^{7 \times 7}([\text{AvgPool}(F); \text{MaxPool}(F)])) \\ &= \sigma(f^{7 \times 7}(F_{\text{avg}}^S; F_{\text{max}}^S)) \end{aligned} \quad (3)$$

式中 $f^{7 \times 7}$ 表示 7×7 的卷积核, F_{avg}^S 和 F_{max}^S 是空间注意力模块平均池化和最大池化结果。

2.2 采摘点定位方法

由于嫩芽自下而上的生长特性, 采摘点一般都位于分割出的嫩芽的底端, 为了更精准的获取嫩芽二维采摘点, 根据分割模型获取的嫩芽轮廓信息, 取轮廓最底部的 3 个像素点作为定位关键点, 这些像素点一般都位于茶叶的茎, 以这 3 个像素点的均值作为最终的采摘点, 单芽、一芽一叶、一芽两叶定位关键点的选取如图 8 所示。

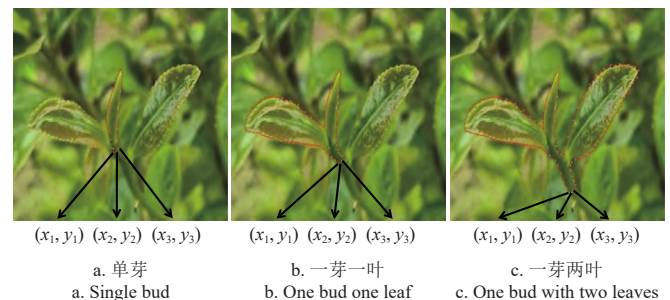


图 8 关键点的选取

Fig.8 The selection of key points

嫩芽二维坐标的计算式可以被概括为式 (4):

$$(X, Y) = \left(\frac{\sum_{i=1}^n x_i}{n}, \frac{\sum_{i=1}^n y_i}{n} \right), n = 3 \quad (4)$$

式中 X 、 Y 表示采摘点的纵横像素值, x_i 、 y_i 表示轮廓最底部三个像素点的纵横像素值, n 为所取像素点个数。

2.3 试验条件与评价指标

2.3.1 试验条件

试验基于本文第一部分所建立的数据集进行模型训练和测试, 所有试验均在相同的硬件条件下进行。

试验平台硬件配置为 Intel(R) Xeon(R) Platinum 8338C CPU, NVIDIA GeForce RTX 4090 GPU。软件环境为 Python 3.9 版本, Pytorch2.1.1 框架, Cuda 12.1 版本。

对于试验参数的选取, 图片大小统一调整为 640×640 像素, 与模型输入尺寸一致, 减少训练内存消耗。学习率均设为 0.01, 防止过大或过小无法得到最优解。优化器选取 SGD 优化器, 相比 ADAM 优化器其训练效果更好。经过试验, 模型在训练 500 个周期左右趋于稳定, 因此将训练周期设为 600。根据显卡性能以及训练

内存消耗，将批量大小设置为 32。

2.3.2 评价指标

为了验证模型的有效性，选取精确度（precision, P ）、召回率（recall, R ）、F1 分数（F1）、平均精度均值（mean average precision, mAP）作为精度评价指标。同时，引入模型权重大小（weights）和帧率（frames per second, FPS）作为其他评价指标，来衡量模型部署在不同设备的能力以及检测速度。

3 结果和分析

3.1 消融试验结果

为了验证改进模型对于原 Yolov5s-segment 模型的提升，通过消融试验逐步在原始模型中加入相应的模块来判断每个模块的有效性并结合相应评价指标进行分析。逐步加入微小目标检测层和 CBAM 注意力机制的掩膜（Mask）各项评价指标如表 2 所示。

表 2 Yolov5s-seg 消融试验各项评价指标
Table 2 Evaluation indexes of Yolov5s-seg ablation test

模型编号 Model NO.	4p	CBAM	SE	ECA	SimAM	精确度 Precision P /%	召回率 Recall R /%	F1 分数 F1 score $F1$ /%	平均精度均值		权重 Weights /MB	帧率 Frames per second FPS/(帧·s ⁻¹)
									Mean average precision/%	mAP ₅₀₋₉₅		
1	×	×	×	×	×	86.6	75.4	80.6	79.2	48.3	14.5	63.3
2	√	×	×	×	×	92.4	81.6	86.7	85.5	54.6	15.3	62.5
3	√	√	×	×	×	93.6	84.3	88.7	87.5	55.6	16.6	62.1
4	√	×	√	×	×	93.3	82.8	87.7	86.4	53.9	15.6	61.5
5	√	×	×	√	×	93.3	83.2	88.0	86.9	55.5	15.6	63.0
6	√	×	×	×	√	93.9	81.9	87.5	86.6	54.6	15.6	60.6

注：表中 4p 为微小目标检测层；mAP₅₀ 为使用 IoU 阈值为 0.50 来计算平均精度均值，而 mAP₅₀₋₉₅ 为使用 IoU 阈值从 0.50 到 0.95，每隔 0.05 一个间隔，计算 10 个 mAP 值，对这 10 个值取平均得到的平均精度；√表示使用该模块，×表示不使用该模块。
Note: The 4p in table represent the small target detection layer; mAP₅₀ represents the average accuracy calculated using IoU threshold of 0.50, while mAP₅₀₋₉₅ represents the average accuracy obtained by averaging 10 mAP values using IoU threshold of 0.50 to 0.95 at an interval of 0.05; √ means to use the module, × means not to use the module.

通过表 2 模型 1-3 可以看出，在增加了微小目标检测层之后，模型各项精度指标均有显著提升，证明了加入微小目标检测层对于茶叶嫩芽这种小目标具有良好的效果。同时，在增加了微小目标检测层的基础上，再在主干网络末端加入 CBAM 注意力机制。结果显示，相较于只添加了微小目标检测层的模型，精度指标均有一定的提升。

同时为了验证不同注意力机制对于模型改进的影响，在添加了微小目标检测层网络结构的基础上，分别在主干网络 SPPF 模块后添加 SE^[23]、ECA^[24]、SimAM^[25] 等注意力机制模块与 3 号模型进行比较，以选取最适用于茶叶嫩芽分割的注意力机制模块。通过表 2 模型 3-6 可以看出，加入了 CBAM 注意力机制的模型除了精确度这一项指标略低于加入了 SimAM 注意力机制的模型之外，

其他指标均高于加入了其他注意力机制的模型。F1 分数也是精确度和召回率这两项指标的综合反映，所以 CBAM 注意力机制对于茶叶嫩芽分割模型具有更好的效果，其融合通道注意力与空间注意力可以对嫩芽特征更好的表示。因此，添加了微小目标检测层和 CBAM 注意力机制后的模型取得了更好的效果，相较于原始模型，精确度、召回率、F1 分数、mAP₅₀、mAP₅₀₋₉₅ 分别提升了 7.0、8.9、8.1、8.3、7.3 个百分点。对于其他评价指标，模型权重大小相较于原始模型增加了 2.1 MB，FPS 下降了 1.2 帧/s，在可以接受的范围之内，可以满足对于茶叶嫩芽检测的需求。基于 Yolov5s-segment 模型逐步增加了微小目标检测层和 CBAM 注意力机制的 2 号和 3 号模型在训练过程中的关键性能参数变化曲线如图 9 所示。本文将验证集上表现最好的 3 号模型命名为 Yolov5s-TB。

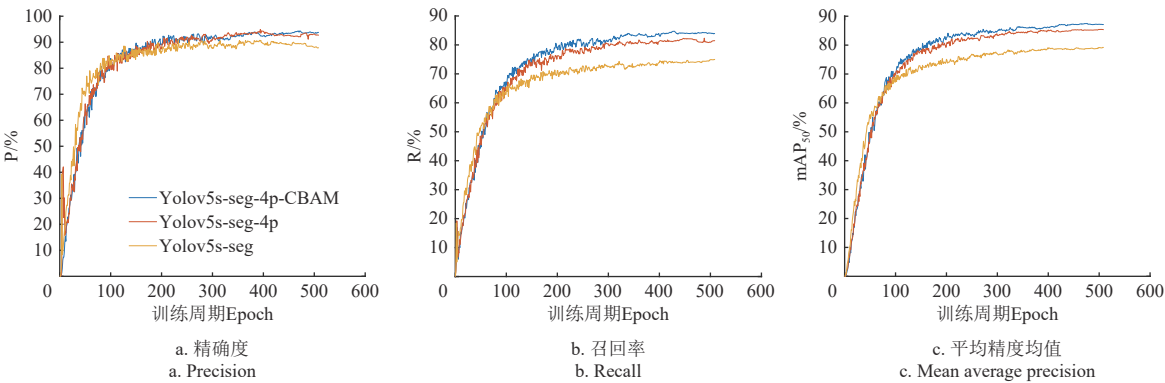


图 9 消融试验训练过程关键性能参数变化曲线
Fig.9 The changes of key performance parameters during ablation test training

3.2 不同实例分割模型的对比试验结果

为了评估本文算法的检测性能，将几种典型的实例分割算法应用于茶叶嫩芽分割与检测，并与本文的算法

进行对比，对比算法包括 Yolact、Yolov8s-seg^[26]。为了保证结果的公平性，所有算法均采用相同的试验配置，在本文自构建的茶叶数据集上进行性能对比，各项评价指

标见表 3。

表 3 不同实例分割模型对比试验评价指标

Table 3 Comparative test evaluation indexes of different case segmentation models

模型 Model	mAP ₅₀ /%	权重 Weights/MB	FPS/(帧·s ⁻¹)
Yolov5s-seg	79.2	14.5	63.3
Yolov8s-seg	83.3	22.8	68.0
Yolact	78.0	189.0	4.2
Yolov5s-TB	87.5	16.6	62.1

表 3 中, Yolov5s-TB 模型的 mAP₅₀ 相较于 Yolov8s-seg 和 Yolact 分别提升了 4.2 和 9.5 个百分点, 模型大小适中, 检测速度能够实现实时检测需求。改进后模型的部分识别结果如图 10 所示, 同时采用 Grad-CAM^[27] (gradient-weighted class activation mapping) 绘制热力图, 热力图中颜色越红说明区域对识别结果贡献较大, 越蓝说明区域对识别结果贡献较少, 通过热力图可以较为直观的看到改进前后模型对于嫩梢区域的关注情况。

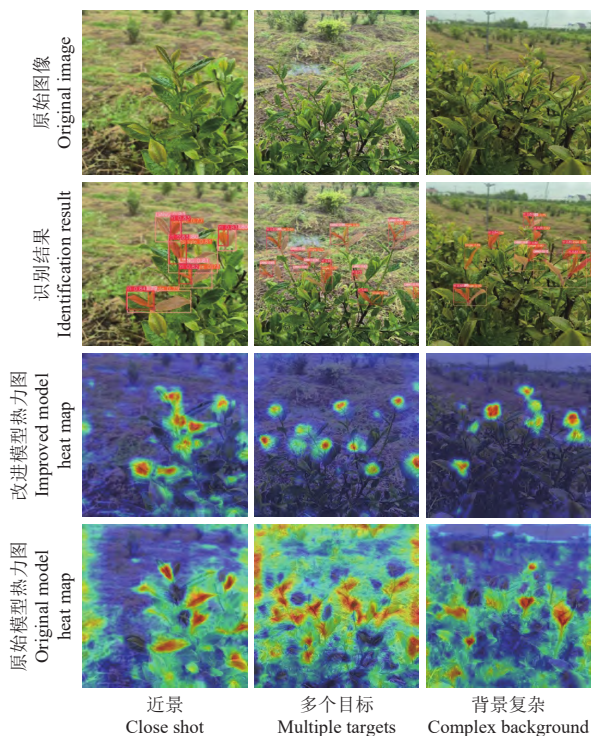


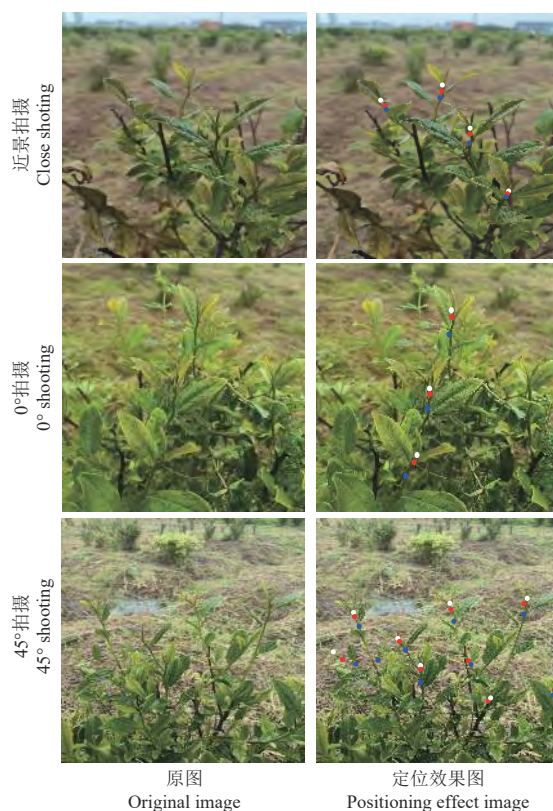
图 10 改进后模型识别结果
Fig.10 The recognition results of improved model

从图 10 可以看出, Yolov5s-TB 模型在预测方面, 可以在复杂背景下识别茶叶嫩梢, 并且对于芽部分被遮挡、叶片间互相遮挡或者部分叶片不在图像范围内的情况也具有一定的识别能力。热力图可视化也验证了模型关注的重点区域, 相较于原始模型, 改进后模型可以更好的关注到嫩梢区域, 提升了模型的检测能力。

3.3 采摘点定位结果

根据本文第二部分的茶叶嫩芽分割模型以及采摘点定位方法, 从近景、0°、45°拍摄图像的部分定位结果如图 11 所示。根据嫩芽状态将采摘点分为单芽、一芽一叶、和一芽两叶三种采摘点, 图中白色圆点为算法找到的单

芽采摘点, 红色圆点为一芽一叶采摘点, 蓝色圆点为一芽两叶采摘点。从图中可以看出, 定位方法可以较为准确的得到不同状态嫩芽的采摘点像素坐标, 并且对于最符合实际应用的 0°以及 45°拍摄也达到了一定的定位效果。



注: 白色圆点为单芽采摘点, 红色圆点为一芽一叶采摘点, 蓝色圆点为一芽两叶采摘点。
Note: The white dots are the picking points of single bud, the red dots are the picking points of a bud with one leaf, and the blue dots are the picking points of a bud with two leaves.

图 11 不同拍摄角度采摘点定位结果

Fig.11 Location results of picking points under different camera shooting angles

4 结 论

本文针对名优茶的采摘需求, 提出了一种基于改进 Yolov5s-segment 的茶叶嫩芽检测与分割模型。在原网络模型基础上加入了微小目标检测层, 增强了模型对小目标的检测能力。在主干网络末端添加了 CBAM 注意力机制模块, 可以对嫩芽特征图更好的融合, 加强特征的表示, 提高检测精度和检测效率。

1) 根据茶园实际环境, 制作了不同场景下的茶叶嫩芽数据集用于模型训练、验证和测试, 并与 Yolact、Yolov8s-seg 两个实例分割模型进行试验对比, 结果表明, 本研究提出的 Yolov5s-TB 精度指标相较于原始模型均有提升, 其中精确度、召回率、F1 分数、mAP₅₀、mAP₅₀₋₉₅ 分别提升了 7.0、8.9、8.1、8.3、7.3 个百分点。随着改进后模型复杂度的提高, 权重大小增加了 2.1 MB, 帧率下降了 1.2 帧/s, 但可以满足实时检测的需求。在今后仍需要进一步改进模型提升检测精度和检测速度, 并且压缩权重大小。

2) 基于本文提出的 YOLOv5s-TB 模型, 提取茶叶嫩芽根茎位置的坐标均值作为采摘点, 实现单芽、一芽一叶、一芽两叶三种嫩芽状态的采摘点定位, 结果表明, 定位方法在各种场景下均有一定的定位效果。在今后需要采用 3D 相机实现三维采摘点的定位并且验证定位精度。

[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(07): 1-9.
- [2] 陈光明, 毛奔, 章永年. 名优茶采摘机器人关键技术研究[J]. 中国农机化学报, 2023, 44(12): 174-179.
CHEN Guangming, MAO Ben, ZHANG Yongnian. Research on key technologies of famous tea picking robot[J]. Journal of Chinese Agricultural Mechanization, 2023, 44(12): 174-179. (in Chinese with English abstract)
- [3] 张昆, 袁博涵, 崔静莹, 等. 基于计算机视觉的茶叶嫩芽识别方法研究进展[J]. 山东农业科学, 2024, 56(5): 163-170.
- [4] 杨福增, 杨亮亮, 田艳娜, 等. 基于颜色和形状特征的茶叶嫩芽识别方法[J]. 农业机械学报, 2009, 40(S1): 119-123.
YANG Fuzeng, YANG Liangliang, TIAN Yanna, et al. Recognition of the tea sprout based on color and shape features[J]. Transactions of the Chinese Society for Agricultural Machinery, 2009, 40(S1): 119-123. (in Chinese with English abstract)
- [5] 吴雪梅, 张富贵, 吕敬堂. 基于图像颜色信息的茶叶嫩芽识别方法研究[J]. 茶叶科学, 2013, 33(6): 584-589.
WU Xuemei, ZHANG Fugui, LV Jingtang. Research on recognition of tea tender leaf based on image color information[J]. Journal of Tea Science, 2013, 33(6): 584-589. (in Chinese with English abstract)
- [6] SHAO P D, WU M H, WANG X W, et al. Research on the tea bud recognition based on improved k-means algorithm[C]// Proceedings of 2018 2nd International Conference on Electronic Information Technology and Computer Engineering, Shanghai, 2018.
- [7] 龙樟, 姜倩, 王健, 等. 茶叶嫩芽视觉识别与采摘点定位方法研究[J]. 传感器与微系统, 2022, 41(2): 39-41, 45.
- [8] 张雯娟. 基于计算机视觉的茶叶嫩芽识别方法研究[D]. 武汉: 中南民族大学, 2022.
ZHANG Wenjuan. Research on Recognition Method of Tea Bud Based on Computer Vision[D]. Wuhan: South-Central University for Nationalities, 2022. (in Chinese with English abstract)
- [9] 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.
LYU Jun, FANG Mengrui, YAO Qing, et al. Detection model for tea buds based on region brightness adaptive correction[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2021, 37(22): 278-285. (in Chinese with English abstract)
- [10] CAO M L, FU H, ZHU J Y, et al. Lightweight tea bud recognition network integrating GhostNet and YOLOv5[J]. Mathematical Biosciences and Engineering, 2022, 19(12): 12897-12914.
- [11] 胡和平, 吴明晖, 洪孔林, 等. 基于改进 YOLOv5s 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 YOLOv5s[J]. Acta Agriculturae Universitatis Jiangxiensis, 2023, 45(5): 1261-1272. (in Chinese with English abstract)
- [12] 洪孔林, 吴明晖, 高博, 等. 基于改进 YOLOv7-tiny 的茶叶嫩芽分级识别方法[J]. 茶叶科学, 2024, 44(1): 62-74.
- [13] CHEN Y T, WU C C, CHEN S F. Application of deep learning algorithm on tea shoot identification and localization[C]// Annual International Meeting of the American Society of Agricultural and Biological Engineers, Detroit, Michigan USA, 2018.
- [14] 朱红春, 李旭, 孟炆, 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 with English abstract)
- [15] YAN L, WU K, LIN J, et al. Identification and picking point positioning of tender tea shoots based on MR3P-TS model[J]. Frontiers in Plant Science, 2022, 13: 962391.
- [16] 张浩, 陈勇, 汪巍, 等. 基于主动计算机视觉的茶叶采摘定位技术[J]. 农业机械学报, 2014, 45(9): 61-65.
ZHANG Hao, CHEN Yong, WANG Wei, et al. Positioning Method for Tea Picking Using Active Computer Vision[J]. Transactions of the Chinese Society for Agricultural Machinery, 2014, 45(9): 61-65. (in Chinese with English abstract)
- [17] 李亚涛. 茶叶采摘机器人的视觉检测与定位技术研究[D]. 杭州: 浙江理工大学, 2023.
LI Yatao. Research on the Visual Detection and Localization Technology of Tea Harvesting Robot[D]. Hangzhou: Zhejiang Sci-Tech University, 2023. (in Chinese with English abstract)
- [18] 骆耀平, 王永镜, 张兰兰等. 名优茶鲜叶原料分级机研究[J]. 茶叶, 2012, 38(1): 27-33.
- [19] BOLYA D, ZHOU C, XIAO F, et al. Yolact++ Better Real-Time Instance Segmentation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022, 44(2): 1108-1121.
- [20] JOCHER G. YOLOv5 by Ultralytics (Version 7.0) [Computer software][Z/OL]. (2022-11-22)[2024-04-02]. <https://doi.org/10.5281/zenodo.3908559>.
- [21] 黄涛, 李华, 周桂等. 实例分割方法研究综述[J]. 计算机科学与探索, 2023, 17(4): 810-825.
- [22] WOO S, PARK J, LEE J Y, et al. CBAM: Convolutional block attention module[J]. Springer, 2018, 11211: 3-19.
- [23] HU J, SEHN L, ALBANIE S, et al. Squeeze-and-Excitation Networks[J]. IEEE Transactions on Pattern Analysis and

- Machine Intelligence*, 2020, 42(8): 2011-2023.
- [24] WANG Q, WU B, ZHU P. ECA-Net: Efficient channel attention for deep convolutional neural networks[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, Washington USA, 2020.
- [25] YANG L X, ZHANG R Y, LI L D, et al. SimAM: A simple, parameter-free attention module for convolutional neural networks[C]//International Conference on Machine Learning, Vienna, Austria, 2021.
- [26] JOCHER G. Ultralytics YOLO (Version 8.0. 0) [Computer software][Z/OL].(2023-01-10)[2024-04-02].<https://github.com/ultralytics/ultralytics>.
- [27] SELVARAJU R R, COGSWELL M, DAS A, et al. Grad-cam: Visual explanations from deep networks via gradient-based localization[J]. *International Journal of Computer Vision*, 2020, 128(2): 336-359.

Extracting tea bud contour and location of picking points in large scene using case segmentation

LI Hanlin , GAO Yanfeng^{*} , XIONG Genliang , LI Yunqian , YANG Yaxin

(College of Mechanical and Automotive Engineering, Shanghai University of Engineering Science, Shanghai 201620, China)

Abstract: Automatic picking tea buds is ever increasing in the continuous expansion of famous brand tea market at present. However, manual picking cannot fully meet the needs of short-term concentrated harvesting in the large-scale tea production. Mechanical mass picking can also share the very short window period of famous brand tea. Therefore, it is the urgent demand to realize the intelligent and accurate picking of famous tea. Among them, visual identification has been confined to the small target of tea buds and the complex background under the natural environment in the field. Fortunately, Yolov5s-segment network model is suitable to identify tea buds, due to the easy deployment, simple integrated structure and real-time detection. In this study, a novel model was proposed to extract the contour and then locate the picking point of tea bud using improved Yolov5s-segment. The rapid recognition and accurate location of picking points were realized for the tea buds in large field scenes. Firstly, P2 micro-target detection layer was imported into the Neck and Head of the original network. The small targets detection of P3, P4 and P5 layers was achieved in the improved Yolov5s-segment network. Secondly, the contour features of tea bud were extracted in the natural lighting environment. CBAM (convolutional block attention module) was added into the end of Backbone of Yolov5s-segment network for the anti-interference of the improved model. Finally, the position coordinates of the bud stem were extracted, according to the bud contour. The picking points were located accurately for the single bud, one bud with one leaf, and one bud with two leaves. Furthermore, the micro target detection layer and CBAM attention mechanism were gradually added into the original Yolov5s-segment model. In ablation test, the CBAM attention mechanism was replaced with SE, ECA and SimAM. After that, the optimal model was compared with Yolov8s-segment and Yoloact. Finally, the positioning test was carried out using the optimal model. The results showed that the best performance was achieved in the model with the micro-target detection layer and CBAM attention mechanism. Compared with the original Yolov5s-segment model, the accuracy, recall rate, F1 score, mean average precision mAP_{50} and mAP_{50-95} were improved by 7.0, 8.9, 8.1, 8.3 and 7.3 percentage points, respectively. In the comparison test, the mAP_{50} of the improved model increased by 4.2 and 9.5 percentage points, respectively, compared with the Yolov8s-segment and Yoloact. The positioning test showed that the pixel coordinates of picking points were accurately obtained in the 0° and 45° shooting angles of camera under different states, indicating the high positioning accuracy. The improved model can be used to accurately extract the outline of single bud, one bud with one leaf and one bud with two leaves in large field scenes. The precise location of picking points can be realized at the same time. The findings can provide a theoretical basis for intelligent and rapid picking of famous tea.

Keywords: image processing; image recognition; tea; bud; Yolov5s-segment; contour extraction; picking point location