

施 武, 袁伟皓, 杨梦道, 等. 一种基于改进 YOLOv8n-seg 的轻量化茶树嫩芽的茶梗识别模型[J]. 江苏农业学报, 2025, 41(1): 75-86.
doi:10.3969/j.issn.1000-4440.2025.01.010

一种基于改进 YOLOv8n-seg 的轻量化茶树嫩芽的茶梗识别模型

施 武, 袁伟皓, 杨梦道, 许高建
(安徽农业大学信息与人工智能学院, 安徽 合肥 230036)

摘要: 茶树嫩芽茶梗识别对实现茶叶采摘的自动化和智能化具有重要意义。然而, 现有的目标检测算法检测茶树嫩芽茶梗存在精度较低、计算量大、模型体积庞大等问题, 限制了其在终端设备上的部署。因此, 本研究基于 YOLOv8n-seg 模型, 提出一种轻量化的茶树嫩芽茶梗识别模型 YOLOv8n-seg-VLS, 并在以下 3 个方面进行了改进: 引入 VanillaNet 轻量化模块替代原有卷积层, 以降低模型的复杂程度; 在颈部引入大型可分离核注意力模块 (LSKA), 以降低存储量和计算资源消耗; 将 YOLOv8 的损失函数从中心点与边界框的重叠联合 (CIoU) 替换为边界框自身形状与自身尺度之间的损失 (Shape-IoU), 从而提高边界框的定位精度。在采集的茶叶数据集上进行测试, 结果表明, 改进后获得的 YOLOv8n-seg-VLS 模型的平均精度值 (mAP) 方面表现较好, 交并比阈值为 0.50 的平均精度值 ($mAP_{0.50}$) 为 94.02%, 交并比阈值为 0.50 至 0.95 的平均精度值 ($mAP_{0.50:0.95}$) 为 62.34%; 模型的准确度 (P) 为 90.08%, 召回率 (R) 为 89.96%; 改进模型的每秒传输帧数 (FPS) 为 245.20 帧, 模型的大小为 3.92 MB, 仅为 YOLOv8n-seg 大小的 57.39%。研究结果为后续茶叶智能化采摘装备的研发提供了技术支持。

关键词: 图像识别; 茶叶采摘; 轻量化模型; YOLOv8n-seg; VanillaNet

中图分类号: TP212; S571.1 **文献标识码:** A **文章编号:** 1000-4440(2025)01-0075-12

A lightweight model for identifying the stalks of tea buds based on the improved YOLOv8n-seg

SHI Wu, YUAN Weihao, YANG Mengdao, XU Gaojian

(School of Information and Artificial Intelligence, Anhui Agricultural University, Hefei 230036, China)

Abstract: Identifying the stalks of tea buds is of great significance for achieving automated and intelligent tea picking. However, existing object detection algorithms face significant challenges in terms of low detection accuracy, high computational demands, and large model sizes, which collectively limit their deployment on edge devices. To address these challenges, we proposed a lightweight tea stalk detection model, YOLOv8n-seg-VLS, which was based on the YOLOv8n-seg framework. The model incorporated three significant enhancements. First, the VanillaNet lightweight module was introduced to replace traditional convolutional layers, thereby reducing the model's complexity. Second, a large separable kernel attention (LSKA) module was incorporated into the neck section of the network to minimize memory usage and resource consumption. Third, the

收稿日期: 2024-08-23

基金项目: 安徽省高校自然科学研究重点项目 (KJ2020A0106); 安徽省重大科技专项 (202103b06020013); 安徽省大学生创新创业计划项目 (S202310364126)

作者简介: 施 武 (2004-), 男, 安徽六安人, 本科, 研究方向为计算机视觉。 (Tel) 15656038325; (E-mail) wshi@stu.ahau.edu.cn

通讯作者: 许高建, (E-mail) xugj@ahau.edu.cn

loss function of YOLOv8 was modified from center intersection over union (CIoU) to shape- and scale-aware intersection over union (Shape-IoU), thereby enhancing the precision of bounding box localization. The experimental results on a collected tea dataset demonstrated that YOLOv8n-seg-VLS achieved a mean average precision (mAP) of 94.02% at $mAP_{0.50}$ and 62.34% at $mAP_{0.50:0.95}$, with a precision of 90.08% and a recall of 89.96%. In com-

parison to the original YOLOv8n-seg, the proposed model demonstrated an improvement in frame rate, reaching 245.20 frames per second (FPS). Moreover, the model size was 3.92 MB, which was only 57.39% of the size of YOLOv8n-seg. These results provide technical support for further development of intelligent tea harvesting equipment.

Key words: image recognition; tea harvesting; lightweight model; YOLOv8n-seg; VanillaNet

茶叶产业作为农业经济的重要组成部分,在推动农村发展和扶贫方面发挥着至关重要的作用。随着茶叶产量的提升,采摘难度也相应增加。目前,茶树嫩芽的采摘主要依赖人工^[1],这一过程既费时又费力。尽管手工采摘能够确保茶芽的高质量和高完整率,但其高成本、低效率以及错过最佳采摘时机的问题使对机械化采摘的研究成为必要。传统机械化采摘因识别精度不高,导致嫩芽完整率偏低。为提升采摘效率和茶芽完整性,集成视觉识别和精细操作的智能采茶机器人已成为学术界和产业界广泛关注的研究热点^[2]。

在智能采茶机器人系统中,准确识别茶叶采摘点是高效且高质量采摘的关键。随着计算机视觉技术的兴起,科研工作者越来越多地将其应用于茶树嫩芽的采摘中。计算机视觉技术在嫩芽采摘中的应用主要经历了 2 个阶段:传统图像识别阶段和深度学习阶段。在传统图像识别阶段,裴伟等^[3]、邵佩迪等^[4]、张博^[5]使用阈值分割法直接分割出茶树嫩芽整体图像,并使用最小外包斜矩形的中心点来替代茶叶的位置。这种方法虽然简单,但采摘点可能定位在嫩芽或者空隙中,不宜用于需要精细定位的采摘末端执行器。罗坤等^[6]针对人工采摘困难和专用采摘器缺乏的问题,设计了低振动、准确采摘的捏切组合式采摘器。以符合茶园嫩梢采摘的要求。龙樟等^[7]利用边缘检测的结果进行最小外接矩形提取,结合骨架化处理找到茶树嫩芽的最低采摘点,但这种算法在不同尺度的嫩芽的采摘中表现不佳。

有关深度学习的研究也取得了诸多成果。李翰林等^[8]提出一种基于实例分割的 YOLOv5s-seg 改进算法来实现野外自然光照环境下嫩芽轮廓的提取,根据嫩芽轮廓特征进行采摘点的准确定位。改进后的模型 $mAP_{0.50}$ (交并比阈值为 0.50 的平均精度值) 和 $mAP_{0.50:0.95}$ (交并比阈值为 0.50 至 0.95 的平均精度值) 分别提高了 8.3 个和 7.3 个百分点。吕军等^[9]基于 YOLOv5 识别模型,计算嫩芽图像的平均灰度值并进行亮度自适应校正,提升了在不同光照条件下的识别精度。杨大勇等^[10]基于 YOLOv8n 模型,在主干网

络中引入动态蛇形卷积,并将颈部的路径聚合网络替换为加权双向特征金字塔网络,提高了模型对茶叶嫩梢的识别率。Shuai 等^[11]使用基于容积扩展的上采样算子(CARAF)完成茶芽的上采样操作,并引入卷积注意力机制模块(CBAM),从而在通道和空间 2 个维度上提高模型对茶树嫩芽和关键点检测的平均精度值(mAP)。Wang 等^[12]提出了一种基于 R-CNN 及 Mask-RCNN 的茶叶采摘点定位方法,通过残差网络(ResNet)和特征金字塔(FPN)进行特征提取,并通过区域提议网络(RPN)进行特征初步分类和候选框回归训练。Yan 等^[13]在 Mask-RCNN 中扩展了掩码(Mask)分支,通过计算多个连通域的面积来识别茶梢的主体部分,通过计算主体部分的最小外接矩形来确定茶梢轴线,进而获得采摘点坐标位置。Xu 等^[14]利用 YOLOv8 进行茶叶检测,并通过开源计算机视觉库(OpenCV)进行红绿蓝-色调饱和度明度(RGB-HSV)颜色转换得到茶叶轮廓,最后通过形态学算法进行关键点选取和定位。

目前,茶树嫩芽采摘点的定位方法大多采用传统图像处理方法与深度学习相结合,主要存在以下不足:识别精度低,检测速度慢,计算量大,不利于终端部署等问题。然而,准确识别茶树嫩芽茶梗分割区域对于智能化茶叶采摘至关重要。研究结果表明,可通过双目相机获取左右图像,然后利用立体匹配算法计算深度信息,再通过对茶梗嫩芽识别框的二维坐标与茶梗分割区域的坐标点,逐一对比得到交界点,并结合深度信息获得茶叶采摘点的三维坐标^[15]。这一过程中,准确识别茶梗分割区域是获得精确采摘点坐标的关键步骤。

鉴于此,本研究拟以茶树嫩芽茶梗为研究对象,设计一种基于 YOLOv8n-seg 的轻量级茶树嫩芽茶梗识别模型,为智能化采摘提供参考。

1 材料与方法

1.1 数据采集

本研究数据集为在安徽省合肥市大阳镇的高科技农业园区拍摄的 4 325 张茶树嫩芽的图片。拍摄

过程中经历了阴天和晴天,光照条件较为复杂。由于大多数茶树的形状被修剪为圆拱形,为了获取尽可能多的不同形态和角度的茶树嫩芽图像,本研究选择从 3 个角度进行拍摄:茶树顶部垂直 90°拍摄,茶树左侧倾斜 45°拍摄,以及茶树右侧倾斜 45°拍

摄。这些角度选择旨在减少视角依赖并捕捉更多的特征。此外,团队在顺光和逆光条件下进行拍摄,同时也考虑了茶叶间相互遮挡因素。采集到的茶树嫩芽图像如图 1 所示。

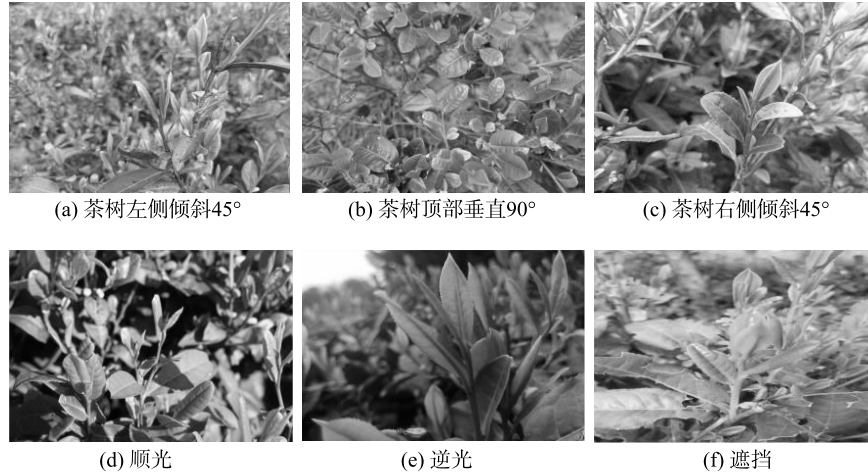


图 1 茶树嫩芽图像采集实例

Fig.1 Example of image acquisition of tea buds

1.2 数据集构建

1.2.1 数据标注 本研究开始前剔除与研究无关或可能破坏模型训练的嫩芽图片,剔除前景或背景模糊、无嫩芽或嫩芽被遮挡的面积超过 50% 的图片。构建训练嫩芽茶梗识别模型所需要的数据集(共 3 725 张图片),然后使用开源图像标注工具 LabelMe^[16] 对图像中的茶树嫩芽梗部进行标注(图 2),标签命名为“Stem”,将其余部分视为背景,标注信息保存至与图像名称对应的 json 文件中,通过 txt 文件转换,将 json 文件转换为模型训练需要的 txt 文件格式。

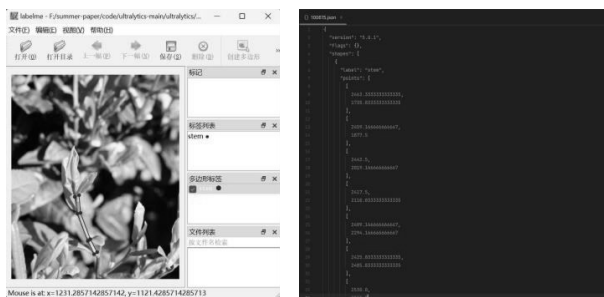


图 2 数据标注过程

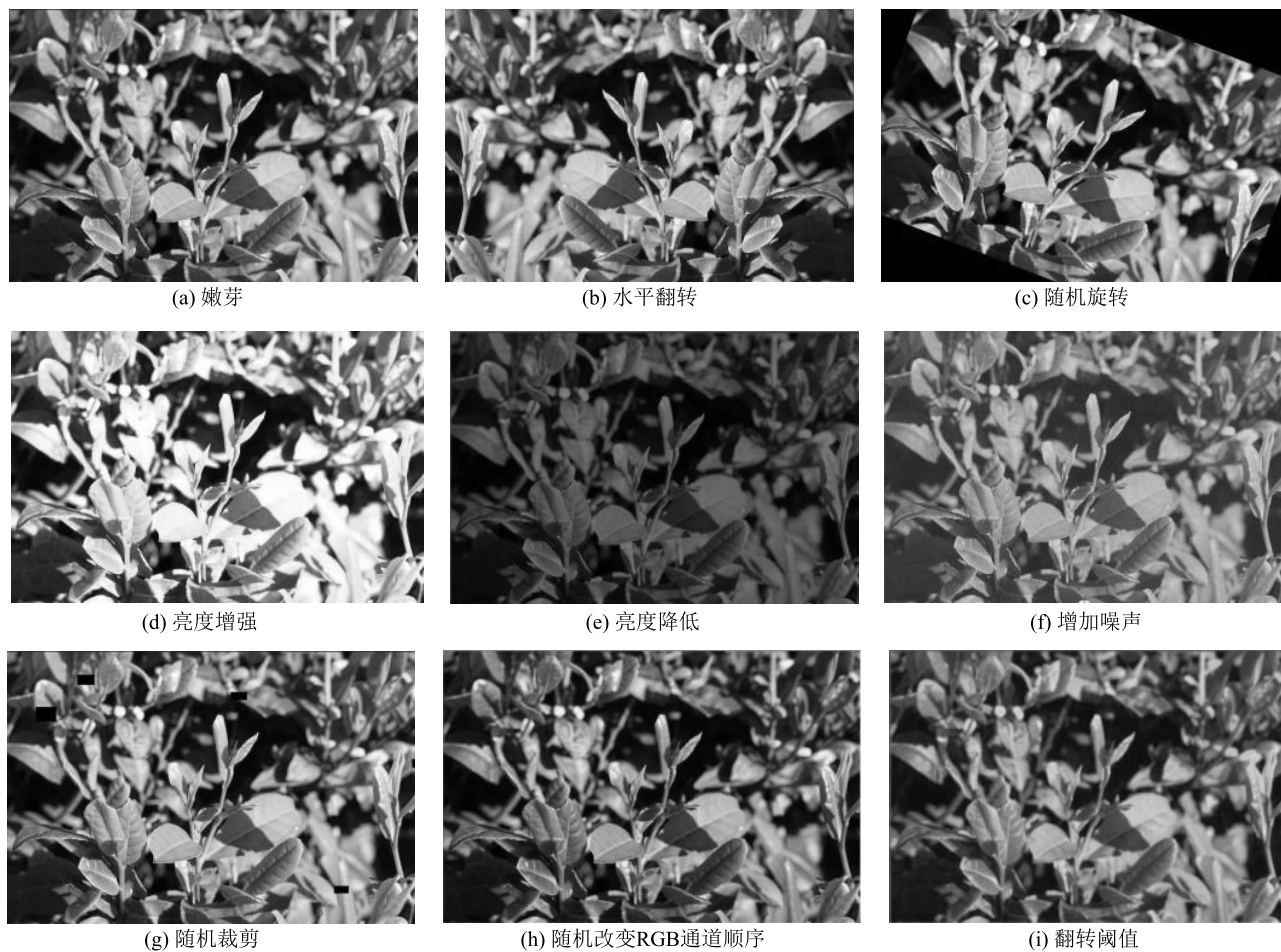
Fig.2 Data annotation process

1.2.2 数据增广 考虑到茶园的环境复杂多变,茶树芽叶的生长角度呈现多样性,加之天气和光照等环境因素影响,给嫩芽的识别与定位带来挑战。由

于采集的茶树嫩芽图像数量有限,难以涵盖所有特征,再加上图像采集和标注过程既耗时又费力,难以构建大规模的数据集,因此需要通过图像处理对数据进行增强,以提高模型的泛化能力。

本研究通过多种数据增强技术扩展训练集,以提高模型的多样性和鲁棒性。图 3(a) 为未经数据增广的原始图片。如图 3(b) 和图 3(c) 所示,通过水平翻转和随机旋转来模拟不同角度下的芽叶特征。如图 3(d) 和图 3(e) 所示,通过增强图像亮度^[17] 和降低图像亮度模拟茶园中光照和天气变换导致的嫩芽颜色变化。如图 3(f) 所示,通过在图像中增加噪声^[18],模拟采集过程中不可避免的噪声干扰,以增强模型在现实应用中的鲁棒性。如图 3(g) 所示,考虑到茶园环境中相互遮挡的情况,对部分嫩芽进行随机裁剪^[19]。此外,如图 3(h) 和图 3(i) 所示,通过随机改变红绿蓝(RGB)通道顺序和翻转阈值,模拟茶园中光照变化和植被反射特性对嫩芽颜色和亮度的影响。

1.2.3 数据集构建 本研究共计采集茶树嫩芽图像样本 3 725 张。根据 1.2.1 节所示的方法,对图像中的嫩芽茶梗进行标注,标注完成后,将标注后的数据集按照比例随机划分为训练集和验证集,训练集的样本数量为 2 600 张,验证集的样本数量为 1 125 张。



RGB 分别表示红色 (Red)、绿色 (Green)、蓝色 (Blue)。

图 3 数据增广效果

Fig.3 Data augmentation effect

将训练集样本按照 1.2.2 节的方法进行数据随机增广,得到增广后的训练集图像样本 7 800 张,对验证集不进行数据增广。验证集样本数量仍为 1 125 张。

1.2.4 基于茶梗识别的采摘点的计算 关于茶叶采摘点三维坐标的确定过程,文献[15]提出了一种基于双目相机获取深度信息的方法,通过嫩芽识别框与茶梗识别框的交界点来确定精确的采摘位置。具体而言,对于嫩芽识别框 S 中的每一个顶点 $P(x_i, y_i)$,判断其是否位于茶梗识别框 T 的边界上(图 4),判断条件表示为:

$$P(x_i, y_i) \in T_{\text{boundary}}, i = 1, 2, \dots, n \quad (1)$$

其中, n 表示嫩芽识别框中的顶点数量。 T_{boundary} 表示茶梗识别的边界框。在此基础上,通过双目相机获取嫩芽中心区域的深度信息 z 。具体表示为:

$$z = f_{\text{depth}}(x_{\text{center}}, y_{\text{center}}) \quad (2)$$

最终,采摘点的三维坐标可确定为 (x_i, y_i, z) 。



S:嫩芽识别框;T:茶梗识别框; $P(x_i, y_i)$:嫩芽识别框 S 中的顶点。

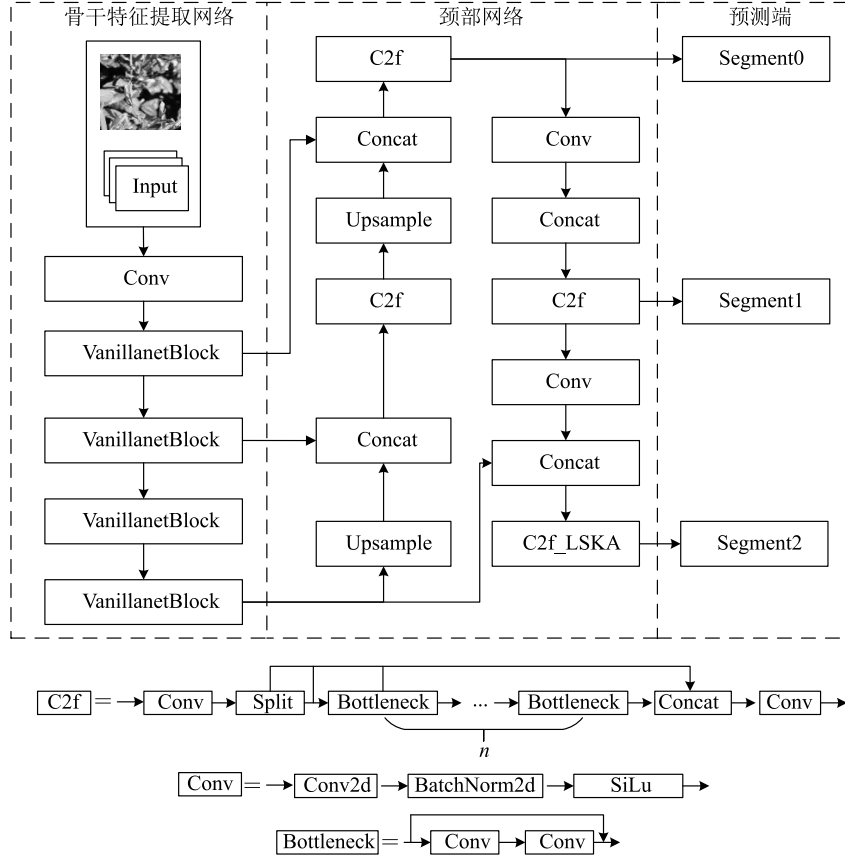
图 4 采摘点计算模拟图

Fig.4 Picking point calculation simulation diagram

1.3 网络模型及轻量化改进

本研究使用 VanillaNet 网络结构来替换 YOLOv8n-seg 中的骨干特征提取网络,在保证精度的前提下,减少模型参数和权重,并将大型可分离核注意力模块 (LSKA) 注意力机制引入到 YOLOv8n-seg 的特征融

合网络中来减少模型的复杂性,考虑到光照、遮挡、背景等因素的影响,本研究采用 Shape-IoU 损失函数来代替原有的 CIoU 损失函数,改进后的模型结构如图 5 所示。



Input 表示图像输入;Conv 表示卷积模块;VanillanetBlock 表示 Vanillanet 轻量化卷积模块;C2f 表示跨阶段特征融合模块;Upsample 表示上采样操作;Concat 表示特征通道拼接操作;C2f_LSKA 表示带 LSKA 的跨阶段特征融合模块;Segment0、Segment1、Segment2 分别表示多尺度分割预测层 0、1、2;Split 表示通道分割模块;Bottleneck 表示一部分通过一系列的卷积、归一化和激活操作进行处理,最后生成的特征图与另一部分直接传递的特征图在连接模块进行拼接;Conv2d 表示二维卷积层;BatchNorm2d 表示二维批量归一化层;SiLu 表示激活函数。

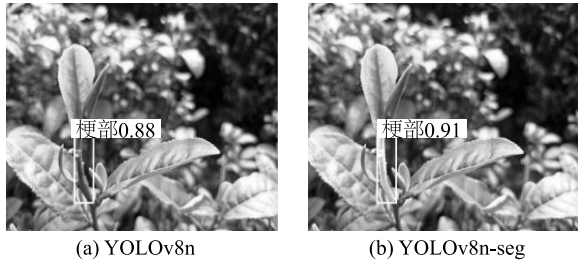
图 5 改进后的模型结构图

Fig.5 Improved model structure diagram

1.3.1 YOLOv8n-seg 分割网络 对于茶树嫩芽茶梗的识别,考虑到茶树嫩芽采摘点的定位需要将茶树嫩芽识别框和茶树嫩芽梗部的分割区域的坐标点进行对比,如果采用 YOLOv8n 模型,结果如图 6(a)所示,只能识别出嫩芽茶梗的边界框,不利于茶树嫩芽采摘点的计算。故本研究采用 YOLOv8n-seg 算法对茶梗区域进行识别[图 6(b)]。YOLOv8n-seg^[20-21]是一种高效的实例分割算法,融合了目标检测和实例分割功能。其网络结构由输入层、主干网

络、颈部网络和分割层组成,每一部分都有其特定的作用。输入层接收图像数据,主干网络通过多层卷积和池化提取图像特征,颈部网络整合这些特征以增强模型的理解能力,而分割层则执行精细的像素级分类和实例分割。处理过程分为 2 个阶段:首先进行目标检测,识别图像中的目标并确定其位置和类别;然后进行实例分割,对每个检测到的目标进行像素级的精确分割。在此基础上引入 YOLOX^[22]中的做法,将马赛克数据增强(Mosaic)在训练最后 10

轮关闭,从而提高检测模型的准确性。



0.88 和 0.91 为识别的平均准确度值。

图 6 YOLOv8n (a) 与 YOLOv8n-seg (b) 茶树嫩芽梗部识别效果对比

Fig.6 Comparison of YOLOv8n (a) and YOLOv8n-seg (b) for tea bud and stem identification

1.3.2 VanillaNet 轻量化网络 由于模型复杂度的增加,对训练环境配置的要求也随之提高,这不利于终端设备上的部署,因此本研究在 YOLOv8n-seg 的骨干特征提取网络中引入了 VanillaNet^[23] 网络,网络结构如图 7 所示。VanillaNet 由主干、主体和全连接层构成。每一个阶段仅用一层卷积模块就可以构建一个非常简单的网络,适合用于检测设备性能受限的情况。

VanillaNet 网络在主干部分使用一个步幅为 4 的 $4 \times 4 \times 3 \times C$ (C 表示输出特征图的通道数) 卷积层,将 3 通道输入图像映射为 C 通道特征图。主体部分包含 4 个阶段。在阶段 I、阶段 II 和阶段 III 中,通过步幅为 2 的最大池化层进行下采样,并在每个阶段中将通道数加倍。在阶段 IV 中,通道数保持不变,接着进行全局平均池化。最后一层为全连接层,用于输出对图像进行分类的结果。为确保计算量最小化,所有卷积层均采用 1×1 大小。每个卷积层后均跟随批量归一化 (BN) 层和激活函数。这种设计在保证网络性能的同时,通过减少模型复杂度和参数数量来降低模型的整体大小。

1.3.3 LSKA 注意力机制 本研究通过引入 LSKA 模块^[24] 来优化模型的特征提取能力,传统的大内核注意力机制虽然能够捕捉到更大范围的上下文信息,但是计算复杂度较高,容易导致过拟合。通过卷积分解策略,LSKA 不仅保留了大内核的优势,还显著减少了计算量,能够适应不同尺度和形状的目标特征,提高了对茶树嫩芽梗部识别的准确性和鲁棒性。

LSKA 注意力机制的网络结构如图 8 所示。LS-

KA 将 $k \times k$ 的大卷积核分解成 $(2d-1) \times (2d-1)$ 的深度卷积、 $k/d \times k/d$ 的深度扩张卷积和 1×1 的卷积。将深度卷积和深度扩张卷积分解成横向和纵向 2 个层次,最后通过连接卷积核来构建完整的网络架构。对于一个输入特征图 F^C ,其 LSKA 的输出如下所示:

$$\bar{Z}^C = \sum_{H,W} W_{(2d-1) \times 1}^C * (\sum_{H,W} W_{1 \times (2d-1)}^C \times F^C) \quad (3)$$

$$Z^C = \sum_{H,W} W_{\left[\frac{k}{d}\right] \times 1}^C \times (\sum_{H,W} W_{1 \times \left[\frac{k}{d}\right]}^C \times \bar{Z}^C) \quad (4)$$

$$A^C = W_{1 \times 1} \times Z^C \quad (5)$$

$$\bar{F}^C = A^C \otimes F^C \quad (6)$$

其中 “ $*$ ” 和 “ \otimes ” 分别代表卷积和哈达玛积。 F^C 是输入特征图, \sum 表示对高度 (H) 和宽度 (W) 维度求和, $W_{(2d-1) \times 1}^C$ 和 $W_{1 \times (2d-1)}^C$ 分别表示大小为 $(2d-1) \times 1$ 和大小为 $1 \times (2d-1)$ 的卷积核, \bar{Z}^C 表示中间计算结果, $W_{\left[\frac{k}{d}\right] \times 1}^C$ 和 $W_{1 \times \left[\frac{k}{d}\right]}^C$ 分别表示大小为 $\left[\frac{k}{d}\right] \times 1$ 和 $1 \times \left[\frac{k}{d}\right]$ 的卷积核, Z^C 表示另一个中间计算结果, $W_{1 \times 1}$ 表示 1×1 的卷积核, A^C 表示注意力权重图, \bar{F}^C 为输出特征图。

1.3.4 Shape-IoU 损失函数 YOLOv8 采用 CIoU^[25] 损失函数作为边界框的损失函数,该函数主要关注了检测框的尺度损失,由于没有考虑到真实框和预测框之间的方向,导致模型收敛速度较慢,由于田间茶叶密集相互遮挡,茶树嫩芽梗部的识别变得困难。为了提高识别效果,本研究采用 Shape-IoU^[26] 损失函数替换 CIoU 损失函数,Shape-IoU 通过整合形状适应性损失和尺度适应性惩罚来优化边界框的形状和尺度,增强模型对不同尺寸目标的适应性和精度,特别是对于像嫩芽茶梗这种精细和细长形状的目标。这一改变不仅解决了 CIoU 敏感的弱点,还提升了模型在处理不同角度和尺寸的茶叶采摘点时的表现,具体公式如下:

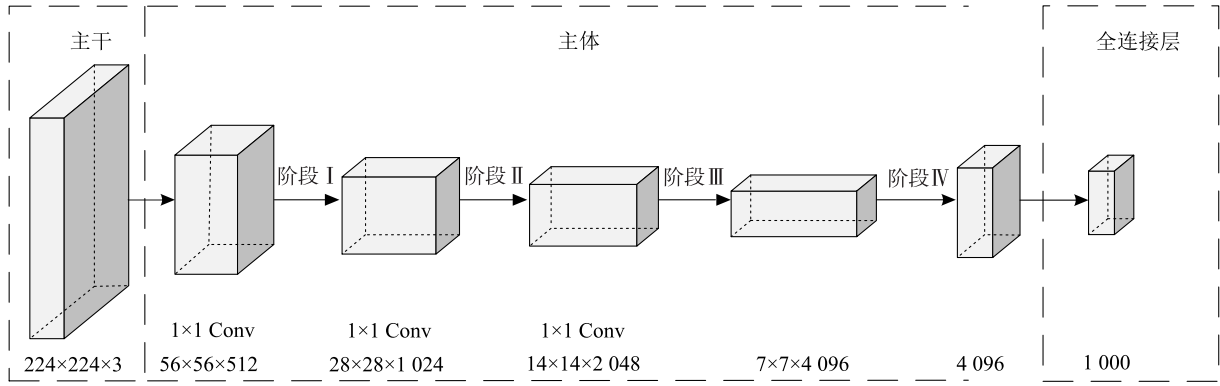
$$IoU = \frac{B^{\text{pred}} \cap B^{\text{gt}}}{B^{\text{pred}} \cup B^{\text{gt}}} \quad (7)$$

$$w_w = \frac{2 \times (w^{\text{gt}})^s}{(w^{\text{gt}})^s + (h^{\text{gt}})^s} \quad (8)$$

$$h_h = \frac{2 \times (h^{\text{gt}})^s}{(w^{\text{gt}})^s + (h^{\text{gt}})^s} \quad (9)$$

$$d^s = h_h \times \frac{(x_c - x_c^{\text{gt}})^2}{c^2} + w_w \times \frac{(y_c - y_c^{\text{gt}})^2}{c^2} \quad (10)$$

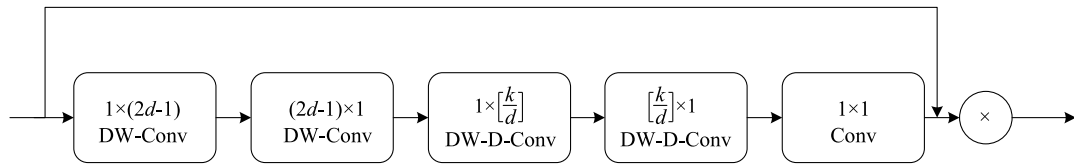
$$\Omega^s = \sum_{t=w,h} (1 - e^{-w^t}) \theta, \theta = 4 \quad (11)$$



Conv 表示卷积模块。

图 7 VanillaNet 网络结构图

Fig.7 VanillaNet network structure diagram



⊗表示哈达玛积; d 和 k 表示不同的卷积核大小;Conv表示卷积模块;DW-Conv表示深度卷积模块;DW-D-Conv表示深度扩张卷积模块。

图 8 LSKA 的网络架构图

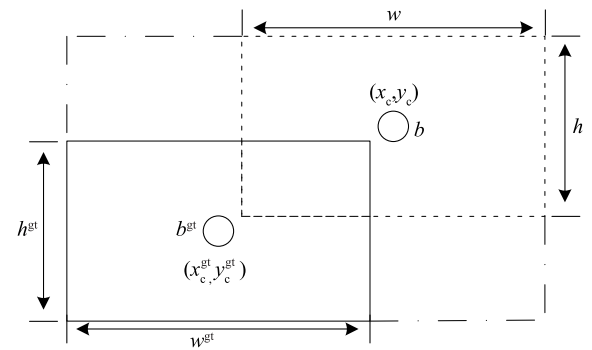
Fig.8 Network architecture diagram of LSKA

$$\begin{cases} \omega_w = h_h \times \frac{|w - w^{gt}|}{\max(w, w^{gt})} \\ \omega_h = w_w \times \frac{|h - h^{gt}|}{\max(h, h^{gt})} \end{cases} \quad (12)$$

$$L_{\text{Shape-IoU}} = 1 - f_{\text{IoU}} + d^s + 0.5 \times \Omega^s \quad (13)$$

以上公式中, B^{pred} 表示预测的边界框; B^{gt} 表示真实的边界框; IoU 表示预测边界框与真实边界框的交集与并集的比值; s 表示比例因子,其与数据集中的目标的比例相关; w^{gt} 表示真实框的宽度; h^{gt} 表示真实框的高度; w_w 和 h_h 分别表示水平和垂直方向的权重系数; (x_c, y_c) 为预测框中心点的坐标; $(x_c^{\text{gt}}, y_c^{\text{gt}})$ 为真实框中心点的坐标; c 表示用于规范化的常数; d^s 表示形状适应性损失; w^l 表示宽度或者高度的权重; θ 为超参数; Ω^s 表示尺度适应性惩罚; ω_w 、 ω_h 分别表示宽度和高度的加权函数; w 和 h 表示预测框的宽度和高度; $\max(\cdot)$ 为求最大值操作; f_{IoU} 为IoU损失函数; $L_{\text{Shape-IoU}}$ 为最终的损失函数。

如图9所示,Shape-IoU专注于调整预测的边界框,使其在形状和大小上更精确地匹配真实框。通过计算边界框形状和尺度的匹配程度,实现对边界框形状和大小更精细的调整。



□真实框; □预测框

b^{gt} 表示真实中心点; b 表示预测中心点; w 和 h 分别表示预测框的宽度和高度; w^{gt} 表示真实框的宽度; h^{gt} 表示真实框的高度。

图 9 Shape-IoU 损失函数示意图

Fig.9 Shape-IoU loss function diagram

2 结果与分析

2.1 训练环境和方法

本研究中的模型训练环境为 Ubuntu 20.04 操作系统,内存为 80 GB,显卡为 NVIDIA RTX 4090 (24 GB),中央处理器为 Intel (R) Xeon (R) Platinum 8481C,配备 16 V 中央处理器。使用的软件环境包括 PyTorch 2.0.0 深度学习框架、Python (版本为

3.8), CUDA(版本为 11.8)。训练参数如表 1 所示。

表 1 模型训练超参数配置

Table 1 Model training hyperparameter configuration

训练超参数	数值
输入大小	640×640
初始学习率	0.01
优化器	随机梯度下降(SGD)
优化器动量	0.937
优化器权重衰减系数	0.000 5
训练批次数	64
迭代周期数	300

2.2 评价指标

本研究采用 $mAP_{0.50}$ 和 $mAP_{0.50:0.95}$ 来衡量模型和掩码精度,模型大小和每秒传输帧数(FPS)用来评估模型的复杂度。 $mAP_{0.50}$ 和 $mAP_{0.50:0.95}$ 可以用以下公式来表示:

$$mAP_{0.50} = \frac{1}{n_c} \int_0^1 P(R) dR \quad (14)$$

$$mAP_{0.50:0.95} = \text{avg}(mAP_i) \quad (15)$$

式中, n_c 表示类别的数量, dR 表示对 R 的微分量, i 表示一系列从 0.50 到 0.95 步长为 0.05 的阈值, P 表示准确度, R 表示召回率, 并且它们满足:

$$P = \frac{TP}{TP+FP} \quad (16)$$

$$R = \frac{TP}{TP+FN} \quad (17)$$

式中, TP 代表真阳性, 即被模型预测为正类的样本; FP 代表假阳性, 即被模型预测为正类的负样

本; FN 代表假阴性, 即被模型预测为负类的正样本。

2.3 消融试验

由表 2 可知, 同时进行 3 种算法改进的 YOLOv8n-seg-VLS 模型的 $mAP_{0.50}$ 为 94.02%, $mAP_{0.50:0.95}$ 为 62.34%, FPS 为 1 s 245.20 帧, 模型的大小为 3.92 MB。使用 VanillaNet 网络替换 YOLOv8n-seg 的骨干特征提取网络, 可以在略微降低 $mAP_{0.50}$ 、 $mAP_{0.50:0.95}$ 、 P 、 R 的情况下, 提高模型的检测速度和减少模型的大小, 这是因为 VanillaNet 在每一个阶段使用单层构建网络, 同时保留了特征映射信息, 在实现轻量化的同时保持了较高的精度。引入 LSKA 模块, 与原有的 YOLOv8n-seg 网络相比, 模型大小有些提升, 这主要是因为 LSKA 通过精简附加模块, 降低了存储需求和计算资源消耗, 从而缩小了模型大小; 此外, 尽管引入 Shape-IoU 损失函数, 在与原 YOLOv8n-seg 模型结果对比上不占优势, 但同时进行 3 种算法改进的 YOLOv8n-seg-VLS 与仅不使用 Shape-IoU 损失函数的 YOLOv8n-seg-VL 的对比结果表明, Shape-IoU 损失函数的加入可以有效提高 FPS , 原因是 Shape-IoU 损失函数可以更好地聚焦于普通质量的边界框来提高模型的总体性能。

由表 2 可见, 轻量化改进后的模型 YOLOv8n-seg-VLS 在 $mAP_{0.50}$ 、 $mAP_{0.50:0.95}$ 、 P 、 R 指标上较 YOLOv8n-seg 略有下降, 这种下降在一定程度上是因为网络复杂度的降低, 这在轻量化过程中难以避免, 然而, 本研究的重点在于茶树嫩芽梗部的识别, 旨在实现模型的轻量化便于在终端部署, 从而提高茶叶采摘点的定位效率, 因此 $mAP_{0.50}$ 、 $mAP_{0.50:0.95}$ 、 P 、 R 略微降低带来的影响甚微。

表 2 消融试验结果

Table 2 Ablation experiment results

模型	VanillaNet	LSKA	Shape-IoU	$mAP_{0.50}$ (%)	$mAP_{0.50:0.95}$ (%)	P (%)	R (%)	FPS (帧, 1 s)	模型大小 (MB)
YOLOv8n-seg	×	×	×	95.42	62.83	91.22	91.02	228.56	6.83
YOLOv8n-seg-V	√	×	×	92.93	61.34	90.05	89.45	234.33	4.36
YOLOv8n-seg-L	×	√	×	93.45	62.44	91.24	90.43	211.09	6.34
YOLOv8n-seg-S	×	×	√	93.67	62.78	90.92	90.43	214.36	6.83
YOLOv8n-seg-VL	√	√	×	93.44	62.14	90.02	89.89	237.46	3.92
YOLOv8n-seg-VS	√	×	√	93.86	62.16	89.98	90.08	239.42	4.36
YOLOv8n-seg-VLS	√	√	√	94.02	62.34	90.08	89.96	245.20	3.92

√表示原始模型 YOLOv8n-seg 中相应的模块被替换; ×表示原始模型 YOLOv8n-seg 相应模块未被替换; $mAP_{0.50}$ 表示交并比阈值为 0.50 的平均精度值; $mAP_{0.50:0.95}$ 表示交并比阈值为 0.50 至 0.95 的平均精度值; P 表示准确度; R 表示召回率; FPS 表示每秒传输帧数。

2.4 其他轻量化网络对比

本研究将 YOLOv8n-seg-VLS 中的 VanillaNet 轻量化模块替换为主流轻量化特征提取主干网络,如 MobileNetV3^[27]、MobileNetV2^[28]、GhostNet^[29]、ShuffleNetV2^[30],在保证除主干网络外其他参数一致的基础上,对比不同主干网络对茶树嫩芽梗部的识别效果。

如表 3 所示,与 MobileNetV3、MobileNetV2、GhostNet 和 ShuffleNetV2 的训练结果相比, VanillaNet 在评价指标上表现出较大的优势。VanillaNet $mAP_{0.50}$ 较其他网络分别提高了 3.65 个、8.60 个、7.65 个、9.14 个百分点, $mAP_{0.50:0.95}$ 较其他网络分

别提高了 3.46 个、5.67 个、5.22 个、5.85 个百分点, P 较其他网络分别提高了 4.40 个、5.76 个、4.64 个、5.79 个百分点, R 较其他网络分别提高了 2.09 个、7.22 个、6.18 个、5.98 个百分点, FPS 较其他网络 1 s 分别提高了 26.44 帧、17.77 帧、6.82 帧和 7.74 帧。虽然 GhostNet 和 ShuffleNetV2 网络的模型大小和 VanillaNet 几乎相同,但与 VanillaNet 网络相比,它们的 $mAP_{0.50}$ 、 $mAP_{0.50:0.95}$ 、 P 、 R 均较低,导致茶树嫩芽梗部的识别效果不理想,因此,本研究采用 VanillaNet 网络作为 YOLOv8n-seg 的主干特征提取网络。

表 3 不同轻量化网络对比

Table 3 Comparison of different lightweight networks

主干网络	$mAP_{0.50}(\%)$	$mAP_{0.50:0.95}(\%)$	$P(\%)$	$R(\%)$	$FPS(\text{帧}, 1\text{ s})$	模型大小(MB)
MobileNetV3	90.37	58.88	85.68	87.87	218.76	5.81
MobileNetV2	85.42	56.67	84.32	82.74	227.43	4.25
GhostNet	86.37	57.12	85.44	83.78	238.38	3.98
ShuffleNetV2	84.88	56.49	84.29	83.98	237.46	4.01
VanillaNet	94.02	62.34	90.08	89.96	245.20	3.92

MobileNetV3、MobileNetV2、GhostNet、ShuffleNetV2、VanillaNet 为当前主流轻量化特征提取主干网络; $mAP_{0.50}$ 表示交并比阈值为 0.50 的平均精度值; $mAP_{0.50:0.95}$ 表示交并比阈值为 0.50 至 0.95 的平均精度值; P 表示准确度; R 表示召回率; FPS 表示每秒传输帧数。

2.5 不同模型对比试验

将 YOLOv8n-seg-VLS 与 Mask-RCNN^[31]、YOLOv5-seg^[32]、YOLOv7-seg^[33]、YOLOv9-seg^[34] 进行对比试验,试验结果如表 4 所示。与其他模型相比, YOLOv8n-seg-VLS 在多个关键指标上均取得了优势。具体而言, YOLOv8n-seg-VLS 的 $mAP_{0.50}$ 和 $mAP_{0.50:0.95}$ 分别达到 94.02% 和 62.34%, 超越了包括 Mask R-CNN 和 YOLO

系列在内的所有对比模型。同时, YOLOv8n-seg-VLS 展现出较高的实时性能,以 1 s 245.20 帧的处理速度领先其他模型。本模型仅占用 3.92 MB 存储空间,大幅降低了对计算资源的需求。这些结果充分证明了本研究提出的改进策略在平衡检测精度、推理效率和模型轻量化方面的有效性,为在资源受限环境下实现茶树嫩芽梗部的精准识别提供了新的可能性。

表 4 不同模型对比

Table 4 Comparison of different models

模型	$mAP_{0.50}(\%)$	$mAP_{0.50:0.95}(\%)$	$P(\%)$	$R(\%)$	$FPS(\text{帧}, 1\text{ s})$	模型大小(MB)
Mask-R-CNN	85.38	54.43	79.98	81.02	150.37	173.30
YOLOv5s-seg	88.71	56.43	85.42	85.51	195.03	13.80
YOLOv7-seg	89.72	58.84	87.14	88.52	202.15	74.40
YOLOv9c-seg	91.22	59.14	89.34	88.87	210.74	56.20
YOLOv8n-seg-VLS	94.02	62.34	90.08	89.96	245.20	3.92

Mask-R-CNN、YOLOv5s-seg、YOLOv7-seg、YOLOv9c-seg 为当前主流卷积神经网络模型; YOLOv8n-seg-VLS 为本研究提出的模型; $mAP_{0.50}$ 表示交并比阈值为 0.50 的平均精度值; $mAP_{0.50:0.95}$ 表示交并比阈值为 0.50 至 0.95 的平均精度值; P 表示准确度; R 表示召回率; FPS 表示每秒传输帧数。

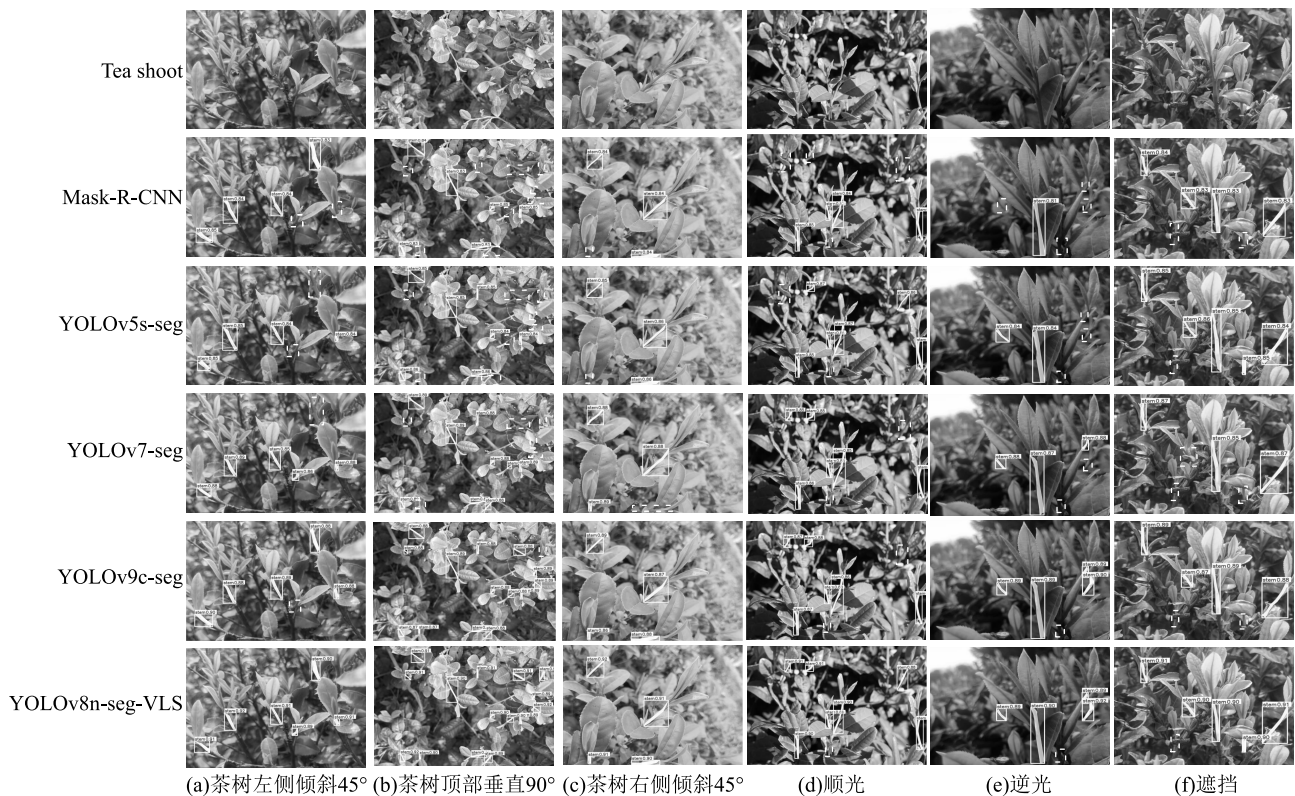
为了验证 YOLOv8n-seg-VLS 模型的实际识别

性能,本研究在相同的试验条件下使用嫩芽图像对

Mask-R-CNN、YOLOv5s-seg、YOLOv7-seg、YOLOv9c-seg 模型进行了测试,识别结果如图 10 所示,该图从左至右依次展示了不同模型在茶树左侧倾斜 45°、茶树顶部垂直 90°、茶树右侧倾斜 45°、顺光、逆光和遮挡条件下对茶树嫩芽梗部的识别效果。

图 10 显示,Mask-R-CNN、YOLOv5s-seg、YOLOv7-seg 和 YOLOv9c-seg 模型在茶树嫩芽梗部的识别上,整体效果不及 YOLOv8n-seg-VLS 模型。在茶树左侧倾

斜 45°、茶树顶部垂直 90°以及右侧倾斜 45°的条件下,YOLOv8n-seg-VLS 模型能够全面识别嫩芽梗部区域,且其置信度高于其他模型。在顺光、逆光和遮挡的条件下,YOLOv8n-seg-VLS 模型也出现了个别漏检现象。这些漏检可能是由于光照变化导致目标特征不够显著,或是遮挡物干扰了模型对目标的识别能力。尽管存在漏检情况,YOLOv8n-seg-VLS 模型在整体识别性能上仍优于其他模型。



Tea-shoot 表示原始茶树嫩芽图像;Mask-R-CNN、YOLOv5s-seg、YOLOv7-seg、YOLOv9c-seg 为当前主流卷积神经网络模型;YOLOv8n-seg-VLS 表示本研究提出的模型;实线边框表示正确识别结果,虚线边框表示漏检实例。

图 10 不同模型对茶树嫩芽梗部识别的测试效果

Fig.10 Recognition effects of different models on the stalks of tea buds

3 结论

针对现有的茶叶智能化采摘研究中茶叶采摘点识别模型精度差及模型较大的问题,本研究提出了一种适用于现代茶园系统中茶树嫩芽梗部识别的模型,实现了茶树不同角度的识别。本研究提出了一种基于 YOLOv8n-seg 模型的轻量化茶树嫩芽梗部识别模型 YOLOv8n-seg-VLS。通过引入 VanillaNet 模块、LSKA 卷积模块以及 Shape-IoU 损失函数,显著

提高了模型的检测精度和处理速度,同时有效减少了模型的大小。试验结果显示,改进获得的 YOLOv8n-seg-VLS 模型的 $mAP_{0.50}$ 、 $mAP_{0.50:0.95}$ 、 P 、 R 分别为 94.02%、62.34%、90.08%、89.96%。改进模型的 FPS 为 1 s 245.20 帧,而模型的大小为 3.92 MB,仅为 YOLOv8n-seg 的 57.39%。

本研究的算法还有一些不足,改进后的 YOLOv8n-seg-VLS 模型的 $mAP_{0.50}$ 、 $mAP_{0.50:0.95}$ 、 P 、 R 较原本 YOLOv8n-seg 模型略有降低,这种精度与模

型大小的权衡是轻量化过程中不可避免的挑战,反映了在资源受限环境下平衡模型性能和计算效率的复杂性,后续的研究将优化后的模型部署到终端设备中,并探索更先进的轻量化方法,在保持甚至提高模型精度的同时,进一步减小模型体积,提高模型在实际应用中的鲁棒性和泛化能力,旨在为茶叶智能采摘系统提供更加精确、高效且易于部署的技术支持,进一步推动茶叶产业的智能化和现代化进程。

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.
- [2] 金寿祥,周宏平,姜洪喆,等. 采摘机器人视觉系统研究进展[J]. 江苏农业学报,2023,39(2):582-595.
- [3] 裴 伟,王晓林. 基于图像信息的茶叶二维采摘坐标的提取[J]. 浙江农业学报,2016,28(3):522-527.
- [4] 邵佩迪,吴明晖,季亚波,等. 茶叶嫩芽机器视觉识别算法研究[J]. 农业装备与车辆工程,2020,58(4):34-36,45.
- [5] 张 博. 基于 RGB-D 的茶叶识别与定位技术研究[D]. 沈阳:沈阳工业大学,2020.
- [6] 罗 坤,吴正敏,曹成茂,等. 茶鲜叶嫩梢捏切组合式采摘器设计与试验[J]. 农业工程学报,2022,38(13):1-9.
- [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] 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.
- [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.
- [10] 杨大勇,黄正栋,郑昌贤,等. 基于改进 YOLOv8n 的茶叶嫩梢检测方法[J]. 农业工程学报,2024,40(12):165-173,313.
- [11] SHUAI L Y, MU J, JIANG X Q, et al. An improved YOLOv5-based method for multi-species tea shoot detection and picking point location in complex backgrounds[J]. Biosystems Engineering,2023(231):117-132.
- [12] 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.
- [13] YAN L J, WU K H, 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.
- [14] XU F, LI B, XU S. Accurate and rapid localization of tea bud leaf picking point based on YOLOv8[C]//MENG X F, CHEN Y, SUO L M, et al. 2023 China National Conference on Big Data and Social Computing. Urumqi, China:Springer,2023:261-274.
- [15] 高 洋. 基于深度学习的茶叶采摘点定位[D]. 合肥:安徽农业大学,2023.
- [16] RUSSELL B C, TORRALBA A, MURPHY K P. LabelMe;a data-base and web-based tool for image annotation[J]. International Journal of Computer Vision,2008,77(1):157-173.
- [17] WANG Y, LI B, YUAN X L. BrightFormer;a transformer to brighten the image[J]. Computers & Graphics,2023(110):49-57.
- [18] XIAO T, XIA T, YANG Y, et al. Learning from massive noisy labeled data for image classification[C]//HORST B, DAVID F, CORDELIA S, et al. 2015 IEEE conference on computer vision and pattern recognition. Boston, MA, USA: IEEE, 2015: 2691-2699.
- [19] ZHONG Z, ZHENG L, KANG G L, et al. Random erasing data augmentation[J]. Proceedings of the AAAI Conference on Artificial Intelligence,2020,34(7):13001-13008.
- [20] YUE X, QI K, NA X Y, et al. Improved YOLOv8-seg network for instance segmentation of healthy and diseased tomato plants in the growth stage[J]. Agriculture,2023,13(8):1643.
- [21] LI H W, HUANG J Z, GU Z N, et al. Positioning of mango picking point using an improved YOLOv8 architecture with object detection and instance segmentation[J]. Biosystems Engineering, 2024(247):202-220.
- [22] GE Z, LIU S T, WANG F, et al. YOLOX:exceeding YOLO series in 2021[EB/OL]. (2021-07-18)[2024-08-20]. <https://doi.org/10.48550/arXiv.2107.08430>.
- [23] CHEN H T, WANG Y H, GUO J Y, et al. VanillaNet;the power of minimalism in deep learning[EB/OL]. (2023-05-23)[2024-08-20]. <https://doi.org/10.48550/arXiv.2305.12972>.
- [24] LAU K W, PO L M, REHMAN Y A. Large separable kernel attention;rethinking the large kernel attention design in CNN[J]. Expert Systems with Applications,2024(236):121352.
- [25] ZHENG Z H, WANG P, REN D W, et al. Enhancing geometric factors in model learning and inference for object detection and instance segmentation[J]. IEEE Transactions on Cybernetics,2022,52(8):8574-8586.
- [26] ZHANG H, ZHANG S J. Shape-IoU;more accurate metric considering bounding box shape and scale[EB/OL]. (2023-12-29)[2024-08-20]. <https://doi.org/10.48550/arXiv.2312.17663>.
- [27] LU J B, YU M M, LIU J Y. Lightweight strip steel defect detection algorithm based on improved YOLOv7[J]. Scientific Reports, 2024,14(1):13267.
- [28] SANDLER M, HOWARD A, ZHU M L, et al. MobileNetV2;inverted residuals and linear bottlenecks[C]//BROWN M S, MORSE B, PELLEGRINI S, et al. 2018 IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City,USA:IEEE,2018:4510-4520.
- [29] PAOLETTI M E, HAUT J M, PEREIRA N S, et al. GhostNet for hyperspectral image classification[J]. IEEE Transactions on Geoscience and Remote Sensing,2021,59(12):10378-10393.
- [30] MA N, ZHANG X, ZHENG H T, et al. ShuffleNetV2;practical guidelines for efficient CNN architecture design[C]//LIN T Y, PATTERSON G, RONCHI M R, et al. 2018 European Conference on Computer Vision. Munich,Germany:Springer,2018:116-131.
- [31] WANG D D, HE D J. Fusion of mask R-CNN and attention mech-

- anism for instance segmentation of apples under complex background[J]. Computers and Electronics in Agriculture, 2022, 196: 106864.
- [32] WANG M J, LI Y, MENG H W, et al. Small target tea bud detection based on improved YOLOv5 in complex background[J]. Frontiers in Plant Science, 2024, 15: 1393138.
- [33] WANG C Y, BOCHKOVSKIY A, LIAO H Y. YOLOv7: trainable bag-of-freebies sets new state-of-the-art for real-time object detectors[C]//BROWN M S, LI F F, MORI G, et al. 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver, BC, Canada: IEEE, 2023: 7464-7475.
- [34] YE R, SHAO G Q, GAO Q, et al. CR-YOLOv9: improved YOLOv9 multi-stage strawberry fruit maturity detection application integrated with CRNET[J]. Foods, 2024, 13(16): 2571.

(责任编辑:陈海霞)