

基于改进 YOLACT++ 的成熟芦笋检测-判别-定位方法

汪小昆^{1,2} 李为民¹ 王琳³ 施印炎^{1,2} 武尧¹ 王得志¹

(1. 南京农业大学工学院, 南京 210031; 2. 江苏省现代设施农业技术与装备实验室, 南京 210031;
3. 智能农业动力装备全国重点实验室, 洛阳 471039)

摘要: 为解决芦笋采收机器人选择性采收过程中成熟芦笋的判别和采摘手准确定位难题,提出了一种改进 YOLACT++ (You only look at coefficients) 算法,利用该方法对成熟芦笋进行检测判别并定位采收切割。通过引入 CBAM (Convolutional block attention module) 注意力机制以及 SPP (Spatial pyramid pooling) 结构改进传统的 YOLACT++ 主干网络,提高了特征提取的有效性;设计了适用于芦笋目标检测的锚框长宽比以保证覆盖到不同姿态的芦笋,以提高网络检测速度和准确率。利用生成的芦笋掩膜分段计算芦笋长度和基部直径,来判定成熟芦笋,并通过空间位姿向量计算成熟芦笋基部区域切割点位置。采收机器人田间试验结果表明,经过训练的改进 YOLACT++ 模型的检测准确率为 95.22%,掩膜平均准确率为 95.60%,640 像素×480 像素图像检测耗时 53.65 ms,成熟芦笋判别准确率为 95.24%,在 X、Y、Z 方向的切割点定位误差小于 2.89 mm,滚转角和俯仰角误差最大为 7.17°;与 Mask R-CNN、SOLO 和 YOLACT++ 模型相比,掩膜平均准确率分别提高 2.28、9.33、21.41 个百分点,最大定位误差分别降低 1.07、1.41、1.92 mm,最大角度误差分别降低 1.81°、2.46°和 3.81°。使用该方法试制的芦笋采收机器人,采收成功率为 96.15%,单根芦笋采收总耗时仅为 12.15 s。本研究提出的检测-判别-定位方法在保证响应速度的前提下具有较高的检测精度和定位精度,为优化改进基于机器视觉的芦笋采收机器人提供了技术支持。

关键词: 芦笋; 采收机器人; 选择性采收; YOLACT++ 算法; 目标检测; 定位

中图分类号: TP391.4; S24

文献标识码: A

文章编号: 1000-1298(2023)07-0259-13

OSID:



Method of Detection – Discrimination – Localization for Mature Asparagus Based on Improved YOLACT++ Algorithm

WANG Xiaochan^{1,2} LI Weimin¹ WANG Lin³ SHI Yinyan^{1,2} WU Yao¹ WANG Dezhi¹

(1. College of Engineering, Nanjing Agricultural University, Nanjing 210031, China

2. Jiangsu Province Engineering Laboratory for Modern Facility Agriculture Technology and Equipment, Nanjing 210031, China

3. State Key Laboratory of Intelligent Agricultural Power Equipment, Luoyang 471039, China)

Abstract: Discrimination of ripe asparagus and accurate location of the picking hand is a challenge in the selective harvesting process of asparagus harvesting robots. To address this challenge, an improved you only look at coefficients (YOLACT++) based algorithm was proposed, which was used to detect and discriminate ripe asparagus and locate harvesting cuts. Improving the traditional YOLACT++ backbone feature extraction network, specifically including the introduction of a convolutional block attention module (CBAM) attention mechanism and a spatial pyramid pooling (SPP) module, to improve the effectiveness of the network for feature extraction and enhance its detection segmentation results. Asparagus have different sizes and postures, by designing different anchor frame sizes to ensure that they were covered, the adaptability of the anchor frame to the aspect ratio of the asparagus was improved, thus improving the detection accuracy and speed of the network. The skeleton was then fitted to asparagus with varying growth forms. Determination of asparagus maturity after calculating asparagus length and basal diameter in segments. Finally, the location of the cutting point in the bottom area of the mature asparagus was calculated, and its spatial location was determined by quantifying the roll angle and pitch angle to

收稿日期: 2022-12-21 修回日期: 2023-02-13

基金项目: 江苏省重点研发计划项目 (BE2021302)、拖拉机动力系统国家重点实验室开放课题 (SKT2022005) 和中国机械工业集团有限公司青年科技基金项目 (QNJJ-PY-2022-25)

作者简介: 汪小昆 (1968—), 男, 教授, 博士生导师, 主要从事设施农业装备智能化控制技术研究, E-mail: wangxiaochan@njau.edu.cn

locate the final harvesting cutting surface. The results of the harvesting robot field trials showed that the detection accuracy of the trained improved YOLACT + + model was 95.22% , the average accuracy of the mask was 95.60% , the detection time of 640 pixels × 480 pixels size image was 53.65 ms , the accuracy of mature asparagus discrimination was 95.24% , the error of cutting point positioning in X , Y and Z directions was less than 2.89 mm , and the maximum error in rotation and pitch angles was 7.17°. Compared with that of the Mask R - CNN , SOLO and YOLACT + + models , the average accuracy of the mask was improved by 2.28 , 9.33 and 21.41 percentage points , respectively ; the maximum positioning errors were reduced by 1.07 mm , 1.41 mm and 1.92 mm , respectively , and the maximum angle errors were reduced by 1.81° , 2.46° and 3.81° , respectively. The harvesting success rate of the trial asparagus harvesting robot was 96.15% , and that the total time taken to harvest a single asparagus was only 12.15 s. The detection - discrimination - location method proposed had high detection and location accuracy , which ensured detection speed on the premise. It can provide technical support for optimizing and improving the asparagus harvesting robot based on machine vision.

Key words: asparagus ; harvesting robot ; selective harvesting ; YOLACT + + algorithm ; object detection ; location

0 引言

芦笋具有丰富的营养价值,被誉为“蔬菜之王”^[1]。我国芦笋种植总面积达 $1.3206 \times 10^6 \text{ hm}^2$,占全球的 90%^[2]。目前芦笋采收基本靠人工进行,由于芦笋成熟期较长,期间需要进行多次采收,劳动强度大,采收成本高^[3]。随着机器人技术的成熟,采用机器人选择性采收芦笋是实现机器替代人工的必由之路^[4-5]。芦笋采收机器人的核心和关键是实现成熟芦笋的形状特征检测以及采收机械手的精确定位。

近年来,多家研究机构都对芦笋选择性采收机器人的芦笋成熟性检测与采收切割点定位进行了研究^[6-13],按其所使用的芦笋检测与定位方法大致可以分为基于传统的图像处理方法和基于目标检测框与图像处理相结合的方法。基于传统的图像处理方法通常需要先对 RGB 图像进行颜色空间转换、白平衡,或对 3D 点云进行滤波、RANSAC 聚类等预处理操作来获得目标芦笋的感兴趣区域,对该区域进行去噪处理后通过阈值分割、聚类等方法提取出目标芦笋^[6-10];基于目标检测框与图像处理相结合的方法首先使用 Faster R - CNN 等目标检测模型对待采区域进行芦笋检测,然后分别对检测框内图像进行后处理提取出目标芦笋^[11-13]。在分割出目标芦笋后,提取其表型信息进行成熟性判断,根据分割的位置信息定位成熟芦笋采收切割点。这两种方法都能获得较好的芦笋检测分割效果并定位成熟芦笋采收切割点,但前者的目标提取过程较为复杂,检测速度较慢^[14-15];后者提高了检测速度,而在对目标检测框内图像进行后处理时,如受到如隔垄芦笋等干扰,其分割效果并不理想。

融合色彩、形态以及纹理等多尺度特征建立的

实例分割模型,避免了复杂的构建过程,其作为“端到端”的像素级检测分割模型,可以有效改善芦笋的识别分割效果,并且具有更高的识别分割精度,被广泛应用于果蔬的识别与定位^[16-17]。YU 等^[18]提出了一种基于 Mask R - CNN 的草莓采摘点视觉定位方法,该方法使用残差网络和特征金字塔网络 (Feature pyramid network, FPN) 来进行特征提取,然后对区域建议网络进行了针对性训练,将实例分割的平均准确率提升至 95.78% ;贾伟宽等^[19]提出一种基于 SOLO (Segmenting objects by locations) 模型的绿色果实优化算法,通过在分离注意力网络的基础上融入 FPN 实现对绿色果实的多尺度特征提取,该方法的识别准确率为 96.16% ;ZHENG 等^[20]提出了一种基于 Mask R - CNN 的芒果采摘点定位方法,通过添加关键点检测分支实现端到端的芒果实例分割和采摘点定位,采摘点定位精度达 0.984。以上针对的目标对象均为形状相对规范且长宽比较为均衡的果实,而大部分芦笋在生长过程中呈现弯曲状态,并且作为细长物体,具有较大的长宽比,使用上述方法不一定适用于芦笋的检测与定位^[21]。

因此,为实现设施芦笋精准选择性收获,本研究提出一种成熟芦笋检测-判别-定位方法,拟在实例分割模型 YOLACT + + (You only look at coefficients) 的基础上融合 CBAM (Convolutional block attention module) 注意力机制以及 SPP (Spatial pyramid pooling) 结构,实现设施环境下成熟芦笋的准确检测与判别,最后定位其采收切割面。

1 芦笋采收机器人设计

基于垄作式栽植模式,本文设计了一种垄间行走式采收机器人(图 1)。该机器人包括 3 个主要模块:采收装置、视觉模块和行走装置。

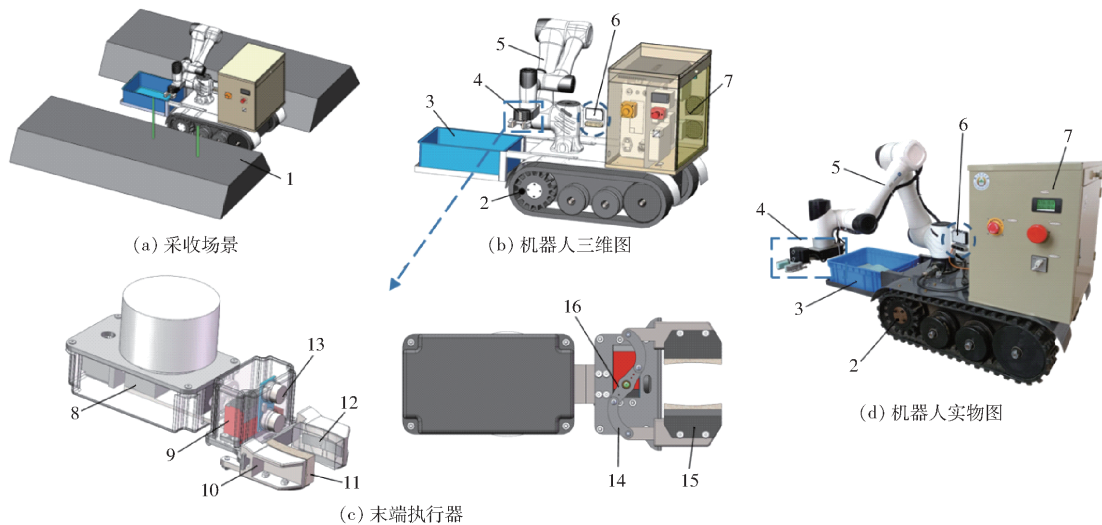


图 1 芦笋采收机器人设计三维图与实物图

Fig.1 Overall structure diagram of harvesting robot

1. 土垄 2. 履带底盘 3. 收集管 4. 末端执行器 5. 机械臂 6. 视觉模块 7. 控制柜 8. 丝杆直线模组 9. 舵机
10. 夹持手指 11. 夹持海绵 12. 压力薄膜传感器 13. 超声波传感器 14. 传动弧杆 15. 切割刀片 16. 舵机臂

采收装置主要由末端执行器和机械臂组成,如图 1c 所示,末端执行器的最大开口宽度为 45 mm,可切割的芦笋直径在 20 mm 范围以内;机械臂 (CR5) 具有 6 个自由度,末端执行器固定于机械臂法兰盘上,使得整机采收工作半径可达 108 cm。

视觉模块使用一个分辨率为 640 像素 × 480 像素的深度相机 (Intel RealSense D435i),其通过相机固定板固定于机械臂第一转动轴上,可以绕机械臂底座旋转 ±180°,可以使机器人实现垄间双边采收。

行走装置采用长 80 cm、宽 58 cm 的履带底盘,除了固定的六自由度机械臂,底盘上部还配备了收集管和 控制柜。控制柜包括主控制器 (LENOVO_LEGION_R7000P)、Arduino 控制板、机械臂控制柜、电压转换模块、驱动器和接线端子。

整机控制系统基于 Ubuntu 20.04 中的机器人操作系统 (Robot operating system, ROS) 进行构建。深度相机、CR5、末端执行器和主控制器之间的通讯由 ROS 实现。

2 成熟芦笋检测-判别-定位方法

成熟芦笋检测-判别-定位方法包括 2 个步骤:①使用改进的 YOLACT++ 算法对输入的 RGB 图像进行识别分割,提取出芦笋在图像中的掩膜。②对芦笋掩膜进行骨架拟合,利用相机的深度信息评估芦笋长度和基部直径,判别出成熟芦笋后定位其采收切割面。

2.1 基于改进 YOLACT++ 的芦笋检测

YOLACT++^[22] 的网络结构主要由主干网络 (Backbone)、掩膜原型生成分支 (Mask prototypes

generate network) 和预测分支 (Prediction head) 组成。为提高 YOLACT++ 模型对芦笋的检测精度并保证其检测速度,该研究对 YOLACT++ 模型进行改进。如图 2 所示,本文从 3 个方面进行改进:①在主干网络的 C3、C4 模块之后添加 CBAM 注意力机制,两个 CBAM 模块可以将网络通道与空间中有用的特征信息进行重用,减少输入图像中噪声对网络的干扰,强化对芦笋特征的提取。②在主干网络 C5 模块后面添加 SPP 模块,加强对芦笋局部特征的提取,有助于提升检测分割结果。③将 FPN 特征图中原始锚框长宽比 {1, 1/2, 2} 改为 {2, 4, 16},以提高锚框对芦笋长宽比的适应性,加快网络产生候选框的速度。

2.1.1 CBAM 注意力机制

主干特征提取网络在提取完图像特征之后将所得特征图 C3、C4、C5 直接进行拼接得到 FPN 中 P3、P4、P5 特征图,拼接过程中各特征图权重相同,特征图中各空间信息的权重也相同,由于芦笋的生长环境中存在如隔垄芦笋、杂草以及土垄等其他非检测目标,导致输入的图像中存在较多噪声,提取的特征图中也包含噪声干扰,随着网络的加深会影响模型精度^[23-24],因此,在 FPN 进行特征融合之前引入了通道注意力机制和空间注意力机制的 CBAM^[25] 模块。

CBAM 模块结构如图 3 所示,该模块不仅可以改变特征通道之间的权重差异,还可以让模型关注特征空间中的关键信息,提高模型的检测精度,而为了防止因加入注意力机制后 C5 特征图因通道数过

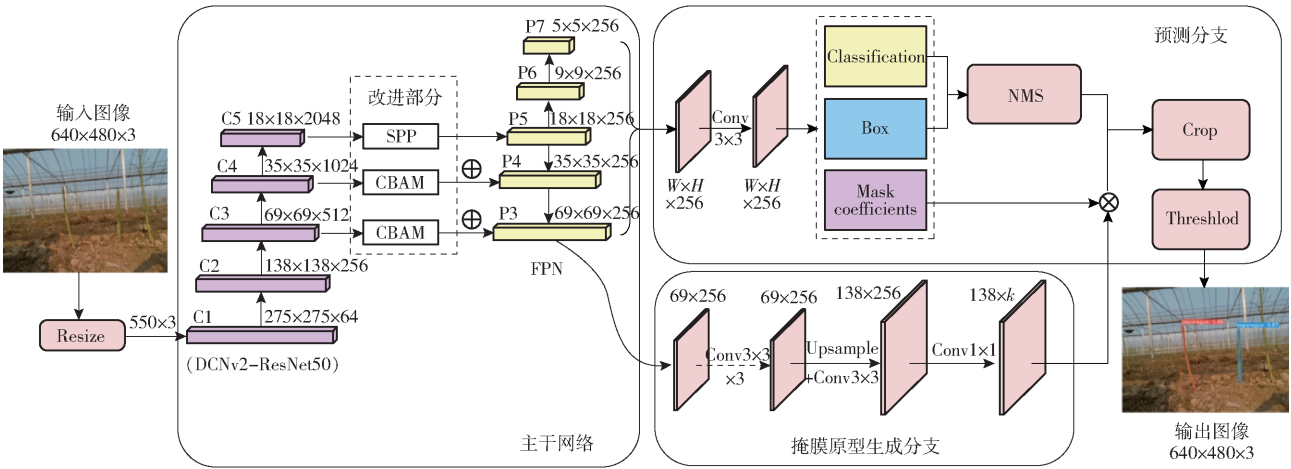


图2 改进的 YOLACT++ 模型
Fig.2 Model of improved YOLACT++

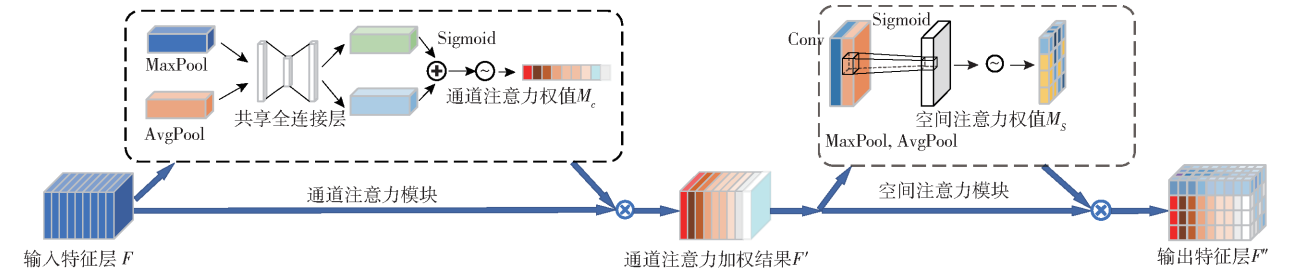


图3 CBAM 注意力机制
Fig.3 CBAM attention mechanism

多、尺度太小导致的模型过拟合问题,仅在 C3 和 C4 特征图后加入注意力机制。引入的 CBAM 模块定义为

$$F' = M_c(F) \otimes F \tag{1}$$

$$F'' = M_s(F') \otimes F' \tag{2}$$

式中 F ——输入特征层
 F' ——通道注意力加权结果
 F'' ——输出特征层
 M_c ——通道注意力模块
 M_s ——空间注意力模块

其中 $F, F', F'' \in \mathbf{R}^{W \times H \times C}, M_c \in \mathbf{R}^{1 \times 1 \times C}, M_s \in \mathbf{R}^{1 \times H \times C}$ 。

2.1.2 基于 SPP 结构的特征融合

模型对芦笋特征进行提取时,由于尺寸最小的一层特征通道数较多,存在卷积神经网络对图像相关特征重复提取的问题,YOLACT++ 网络中虽然有 FPN 作为特征融合结构,但是芦笋为细长物体,此问题会使得模型对芦笋的局部特征提取不够充分,降低模型的检测精度。通过在 C5 特征图后加入 SPP^[26] 结构实现局部特征和全局特征的融合,进而丰富特征图的表达能力,可以提高模型的检测精度^[27]。引入的 SPP 结构如图 4 所示,在特征层输入网络以及输出网络时使用卷积核大小均为 3×3 的卷积来改变通道数,中间所用池化核大小分别为 5×5、9×9、13×13。

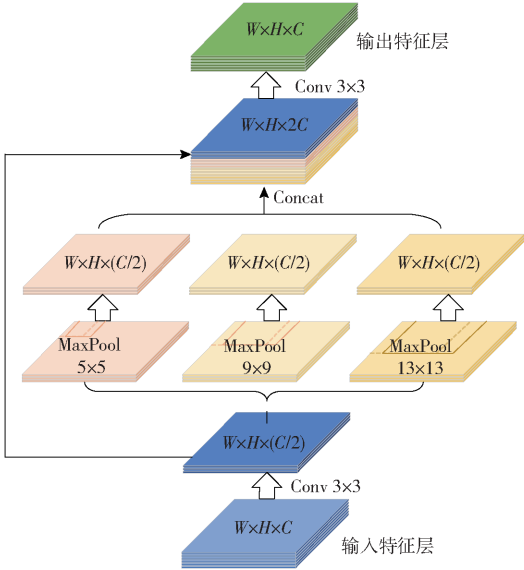


图4 SPP 结构
Fig.4 SPP structure

2.1.3 芦笋检测锚框长宽比设计

原始的 YOLACT++ 网络是基于单阶段的锚框机制检测器,生成的锚框用于前景与背景的判断及位置回归。图像在输入网络之前都会通过双线性插值法将尺寸大小改为 550 像素×550 像素,在 FPN 中采用 {24×24,48×48,96×96,192×192,384×384} 5 种尺度以及 {1,1/2,2} 3 种长宽比的锚框进行

目标区域的检测。为了验证该锚框比例是否适合芦笋检测,本研究随机选取混合数据集内 400 幅图像中共 947 根芦笋的目标框尺寸进行统计,计算统计结果如图 5 所示。

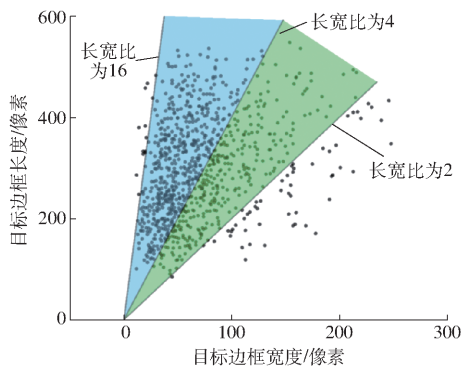


图 5 芦笋目标框统计

Fig. 5 Green asparagus target box statistics

由图 5 可知,芦笋目标边框的长宽比主要分布在 2~16 之间,因此,本研究将 P3~P7 这 5 层特征图的锚框长宽比改为 {2, 4, 16}, 锚框尺度保持不变,共生成 15 种不同的锚点,以保证覆盖到不同尺寸和姿态的芦笋。

2.2 成熟芦笋判别与切割面定位

成熟芦笋判别与切割面定位流程如图 6 所示,经过改进的 YOLACT++ 模型实例分割后,得到完整的芦笋掩膜(图 6b, ouv 表示像素坐标系)。对芦笋掩膜进行初始点位获取(图 6c),进而得到芦笋骨架关键点,拟合出芦笋骨架轮廓(图 6d)。然后进行芦笋长度、基部直径计算以及空间位姿估计(图 6e),进而判定出成熟芦笋,最后确定其基部区域切割点位置,定位最终采收切割面(图 6h)。

(1) 初始点获取。首先获得芦笋掩膜在像素坐标系 v 轴投影的最高点、最低点坐标值,计算式为

$$v_{\max}^{(i)} = \max(v^{(i)}) \quad (3)$$

$$v_{\min}^{(i)} = \min(v^{(i)}) \quad (4)$$

式中 $v_{\max}^{(i)}$ ——芦笋在 v 轴投影最低点坐标值

$v_{\min}^{(i)}$ ——芦笋在 v 轴投影最高点坐标值

$v^{(i)}$ ——芦笋在 v 轴投影轮廓坐标值

其中 $i=1, 2, 3, \dots$, 表示图像中芦笋数量。得到最高点坐标 $(u_n^{(i)}, v_{\min}^{(i)})$ 和最低点坐标 $(u_x^{(i)}, v_{\max}^{(i)})$ 。将目标芦笋在 v 轴投影方向上平均分成 k 等份(第 I 段为距土面最近的那段),各等分段交界处的 v 计算式为

$$v_j^{(i)} = v_{\max}^{(i)} - \frac{v_{\max}^{(i)} - v_{\min}^{(i)}}{k} \quad (5)$$

式中 $v_j^{(i)}$ ——芦笋在 v 轴投影中第 j 和第 $j+1$ 段交界处坐标值

根据预试验 k 通常设置为 6, $j=1, 2, \dots, k-1$ 。

根据 $v_j^{(i)}$ 遍历芦笋掩膜轮廓获取其对应目标轮廓上的初始点坐标 $(u_{j1}^{(i)}, v_j^{(i)})$ 和 $(u_{j2}^{(i)}, v_j^{(i)})$ 。

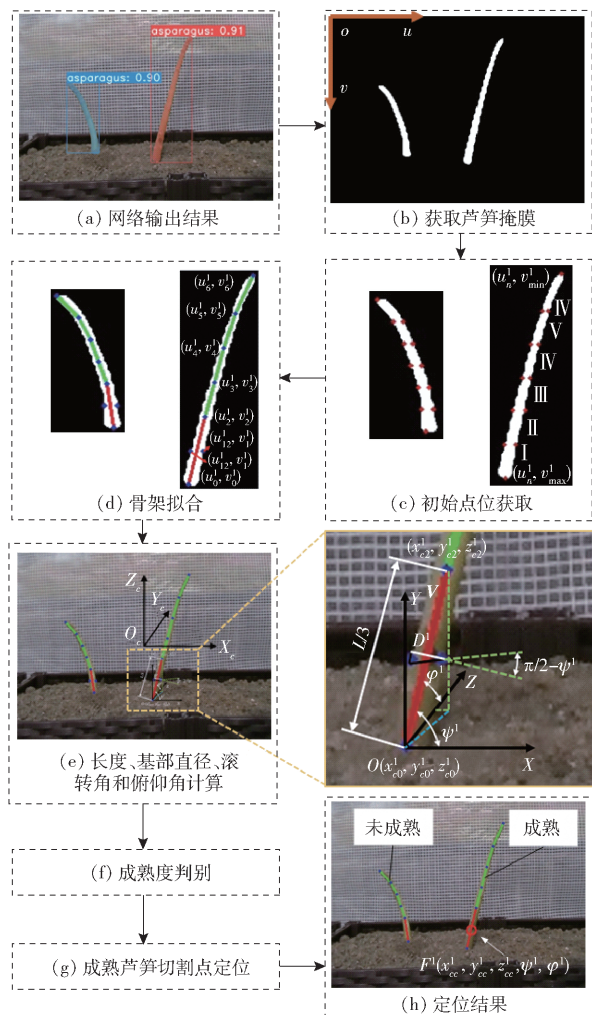


图 6 成熟芦笋判别与切割面定位步骤

Fig. 6 Steps for identifying mature asparagus and locating cutting surface

(2) 骨架拟合。通过初始点坐标计算各骨架关键点坐标 $(u_j^{(i)}, v_j^{(i)})$, 计算式为

$$u_j^{(i)} = \frac{u_{j1}^{(i)} + u_{j2}^{(i)}}{2} \quad (6)$$

式中 $u_j^{(i)}$ ——芦笋在第 j 和第 $j+1$ 段交界处骨架关键点坐标的 u 分量

由于芦笋是从土面长出,底部多被泥土遮挡为不规则形状,如果使用 v 轴投影最大值点作为最低点,在对芦笋轮廓进行长度测量和位姿估计时偏差较大,影响成熟芦笋的判断和切割面定位,故需重新选取最低点。通过对芦笋观察发现,其底部三分之一处韧性较大,弯曲度几乎为零(图 7a)。因此,使用第 I 段、第 II 段和第 III 段对应的两个关键点 $(u_1^{(i)}, v_1^{(i)})$ 和 $(u_2^{(i)}, v_2^{(i)})$ 建立直线(图 6d 中红色线段),取直线与芦笋掩膜轮廓 v 分量最大的交点作为新的最低点 $(u_0^{(i)}, v_0^{(i)})$ 。直线的建立公式为

$$\begin{cases} v = bu + a & (b \neq \infty) \\ u = u_1^{(i)} & (b = \infty) \end{cases} \quad (7)$$

其中 $b = \frac{v_2^{(i)} - v_1^{(i)}}{u_2^{(i)} - u_1^{(i)}} \quad a = v_1^{(i)} - bu_1^{(i)}$

式中 b ——斜率 a ——截距

芦笋的最高点不变,令 $u_6^{(i)} = v_n^{(i)}, v_6^{(i)} = v_{\min}^{(i)}$,则其坐标表示为 $(u_6^{(i)}, v_6^{(i)})$ 。

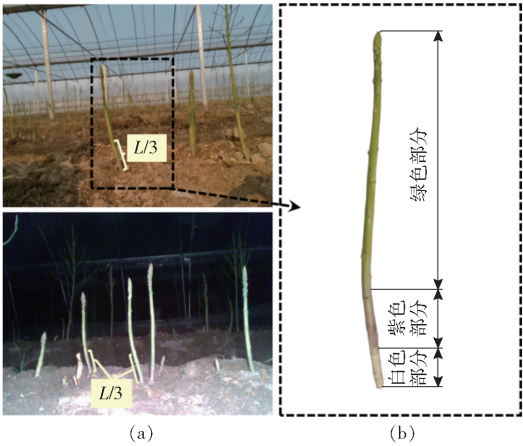


图 7 芦笋弯曲示例与颜色分布

Fig. 7 Asparagus bending example and color distribution

(3) 长度、基部直径以及位姿计算。使用 pyrealsense2 库中的 get_distance() 函数将像素坐标系 ouv 下的初始点、骨架关键点、最高点和最低点转换为相机坐标系 $O_cX_cY_cZ_c$ 下对应坐标,利用上述坐标分别计算出各段芦笋长度,芦笋总长度为 k 段长度之和。长度计算公式为

$$L^{(i)} = \sum_{n=0}^k \sqrt{(x_{c(n+1)}^{(i)} - x_{cn}^{(i)})^2 + (y_{c(n+1)}^{(i)} - y_{cn}^{(i)})^2 + (z_{c(n+1)}^{(i)} - z_{cn}^{(i)})^2} \quad (8)$$

式中 $L^{(i)}$ ——芦笋总长度

$x_{cn}^{(i)}, y_{cn}^{(i)}, z_{cn}^{(i)}$ ——相机坐标系下骨架关键点、最高点和最低点坐标值

芦笋生长形态各异,已知芦笋姿态可使采收机器人的末端执行器平行夹持芦笋,减少芦笋的损伤。在本研究中使用滚转角和俯仰角确定芦笋姿态,以芦笋底部最低点 $(x_{c0}^{(i)}, y_{c0}^{(i)}, z_{c0}^{(i)})$ 为原点建立空间直角坐标系 $OXYZ$,滚转角为步骤(2)中所建立直线在 OXY 平面上投影与 X 轴的夹角,俯仰角为其在 OYZ 平面上投影与 Z 轴的夹角。滚转角与俯仰角计算公式为

$$\psi^{(i)} = \arctan \frac{y_{c2}^{(i)} - y_{c0}^{(i)}}{x_{c2}^{(i)} - x_{c0}^{(i)}} \quad (9)$$

$$\varphi^{(i)} = \arctan \frac{y_{c2}^{(i)} - y_{c0}^{(i)}}{z_{c2}^{(i)} - z_{c0}^{(i)}} \quad (10)$$

式中 $\psi^{(i)}$ ——芦笋滚转角

$\varphi^{(i)}$ ——芦笋俯仰角
 $x_{c0}^{(i)}, y_{c0}^{(i)}, z_{c0}^{(i)}$ ——相机坐标系下最低点坐标
 $x_{c2}^{(i)}, y_{c2}^{(i)}, z_{c2}^{(i)}$ ——相机坐标系下第Ⅱ段与第Ⅲ段对应的骨架关键点坐标

基部直径如图 6e 中白色线段所示,使用第Ⅰ段与第Ⅱ段对应的芦笋轮廓初始点以及滚转角计算得出,计算公式为

$$D^{(i)} = \begin{cases} d^{(i)} & (\psi^{(i)} = 90^\circ) \\ d^{(i)} \sin(\pi/2 - \psi^{(i)}) & (0^\circ < \psi^{(i)} < 90^\circ) \\ d^{(i)} \sin(\psi^{(i)} - \pi/2) & (90^\circ < \psi^{(i)} < 180^\circ) \end{cases} \quad (11)$$

其中

$$d^{(i)} = \sqrt{(x_{c11}^{(i)} - x_{c12}^{(i)})^2 + (y_{c11}^{(i)} - y_{c12}^{(i)})^2 + (z_{c11}^{(i)} - z_{c12}^{(i)})^2}$$

式中 $D^{(i)}$ ——芦笋基部直径

$x_{c11}^{(i)}, y_{c11}^{(i)}, z_{c11}^{(i)}$ ——相机坐标系下芦笋底部第Ⅰ段与第Ⅱ段交界处轮廓左初始点坐标值

$x_{c12}^{(i)}, y_{c12}^{(i)}, z_{c12}^{(i)}$ ——相机坐标系下芦笋底部第Ⅰ段与第Ⅱ段交界处轮廓右初始点坐标值

(4) 成熟芦笋判别。芦笋雌雄粗细不同,对应成熟标准也不同,根据国家芦笋等级规格^[28]以及农户根据市场所确定的芦笋等级规格,成熟芦笋判定标准如表 1 所示。表中 L 表示芦笋长度, D 表示芦笋基部直径, $\times$ 表示芦笋未成熟, $\checkmark$ 表示芦笋成熟, 0 表示芦笋过成熟。

表 1 成熟芦笋判别标准

Tab.1 Criterion for mature asparagus

基部直径 D/mm	长度 L/cm				
	0 ~ 20	20 ~ 24	24 ~ 28	28 ~ 40	≥ 40
0 ~ 8 (细笋)	×	✓	✓	✓	0
8 ~ 13 (中笋)	×	×	✓	✓	0
≥ 13 (大笋)	×	×	×	✓	0

市面上鲜销的芦笋根据其直径主要分为 3 个等级,分别为细笋、中笋和大笋。芦笋直径小于 8 mm 为细笋,长度超过 20 cm 时成熟;芦笋直径介于 8 ~ 13 mm 之间为中笋,长度超过 24 cm 时成熟;芦笋直径不小于 13 mm 为大笋,长度超过 28 cm 为成熟。当芦笋长度超过 40 cm 为过成熟芦笋,此时芦笋出现散头现象,经济价值低下,故不作为选择性采收的目标。

(5) 切割面定位。与其他果蔬直接采摘果实不同,芦笋的可食用部分与不可食用部分没有明确的界限。如图 7b 所示,芦笋靠近土面一段为紫色,该

部分韧性较大,不具有食用价值,绿色部分脆嫩多汁,为芦笋的“果实”部分。本研究考虑到随着芦笋的生长,其底部的不可食用部分随之增加(图 7b 中紫色部分),采收切割点位置越偏上,将芦笋切割点位置定于芦笋总长 ε 处,即相当于底部三分之一段(第 I 和第 II 段) 3ε 处。以芦笋底部最低点 $(x_{c0}^{(i)}, y_{c0}^{(i)}, z_{c0}^{(i)})$ 为向量 $\mathbf{V}$ 起点,点 $(x_{c2}^{(i)}, y_{c2}^{(i)}, z_{c2}^{(i)})$ 为向量 $\mathbf{V}$ 终点,在向量 $\mathbf{V}$ 上求得芦笋切割点计算公式为

$$\begin{bmatrix} x_{cc}^{(i)} \\ y_{cc}^{(i)} \\ z_{cc}^{(i)} \end{bmatrix} = [3\varepsilon \quad 1-3\varepsilon] \begin{bmatrix} x_{c2}^{(i)} & y_{c2}^{(i)} & z_{c2}^{(i)} \\ x_{c0}^{(i)} & y_{c0}^{(i)} & z_{c0}^{(i)} \end{bmatrix} \quad (12)$$

式中 $x_{cc}^{(i)}, y_{cc}^{(i)}, z_{cc}^{(i)}$ ——相机坐标系下切割点 $P^{(i)}$ 坐标值

最终成熟芦笋采收切割面在相机坐标系下表示为 $F^{(i)}(x_{cc}^{(i)}, y_{cc}^{(i)}, z_{cc}^{(i)}, \psi^{(i)}, \varphi^{(i)})$ 。

3 模型训练与评价指标

3.1 数据集制作

本文使用的芦笋图像于 2022 年 3—5 月在江苏省红窑镇春华园芦笋种植基地(33°N, 119°E)分批采集,芦笋品种为翡翠明珠 F1。分别在不同天气光照条件下进行拍摄,拍摄相机为 Intel RealSense D435i,夜间拍摄采用 LED 补光灯进行补光,拍摄成像距离为 0.6~1 m,成像高度为 40 cm(成像距离及成像高度均与芦笋采收机器人实地工作时一致),相机连接计算机使用快门拍摄方式,共保存图像 1 918 幅,图像分辨率为 640 像素×480 像素,图像格式为 JPEG。对所有图像按照 COCO 数据集格式使用专业标注软件 Labelme 进行标注,标注文件格式保存为 JSON,共标注芦笋 4 730 根,形成原始数据集,用于训练生成芦笋掩膜。

为了避免训练过程中出现过拟合,提高模型对于芦笋识别分割的鲁棒性和泛化性,采用数据增强方法对数据集进行扩充。在扩充前将原始数据集按 3:1:1 的比例划分为初始训练集、初始验证集和测试集,之后分别通过水平翻转、平移、运动模糊、随机改变亮度、添加高斯噪声和椒盐噪声 6 种数据增强方法对初始训练集和初始验证集进行扩充。最终所得数据集图像共 11 122 幅,其中训练集图像 8 053 幅,验证集图像 2 685 幅,测试集图像 384 幅。

3.2 软硬件配置及训练策略

本研究模型训练与试验分别在 2 台计算机上进行,模型训练使用工作站硬件环境:CPU 型号为 Intel Xeon Gold 5220R,内存为 32 GB,频率为 2.2 GHz,GPU 型号为 PNY Quadro RTX5000,显存为 24 GB;试验

平台的硬件环境:CPU 型号为 AMD R7-5800H,内存为 16 GB,频率为 3.2 GHz,GPU 型号为 GeForce GTX 3060,显存为 6 GB。所使用的软件环境均为:Ubuntu20.04 LST 操作系统,Pytorch1.7.1 深度学习框架,NVIDIA 460.56 显卡驱动,CUDA 11.2 版本。

为了节省模型训练所用时间,本研究使用迁移学习方法进行模型训练,训练迭代次数 50 000 次,训练初始学习率为 0.001,分别在训练到 25 000、37 500、45 000 次时学习率调整为前一阶段学习率的 1/10。GPU 同时处理图形数为 64,优化器采用 Adam,参数设置为 $\beta_1 = 0.9, \beta_2 = 0.999, \varepsilon = 10 \times 10^{-8}$ 。

3.3 评价指标

采用检测准确率(Accuracy, A_{cc})、掩膜平均准确率(Mask average precision, AP_{mask})和平均检测时间(t_d)评价模型性能。

采用成熟芦笋判别准确率(M)、切割点在 $X/Y/Z$ 方向定位误差以及标准差($\sigma_x \pm s_x, \sigma_y \pm s_y, \sigma_z \pm s_z$)、俯仰角定位误差以及标准差($\sigma_\varphi \pm s_\varphi$)、滚转角定位误差以及标准差($\sigma_\psi \pm s_\psi$)、判别与定位时间(t_p)作为评估成熟芦笋判别与切割面定位方法性能的评价指标。其中 t_p 表示从获得芦笋识别分割结果到定位成熟芦笋切割面所用的时间。

4 试验结果与分析

4.1 芦笋检测试验

4.1.1 消融试验

改进后的 YOLACT++ 模型在 YOLACT++ 模型的基础上融合 CBAM 模块、SPP 结构,并且改变锚框长宽比,为了验证改进后模型中所改进模块结构的合理性以及有效性,设计消融试验,将 384 幅测试集图像放入训练好的模型中进行模型性能对比,最终得到不同模型下芦笋检测与分割的评价结果如表 2 所示。表中 Y 表示 YOLACT++ 模型,A 表示改变特征图锚框长宽比,S 表示添加 SPP 结构,C 表

表 2 模型性能对比结果

Tab.2 Model performance comparison results

模型	$A_{cc50}/\%$	$AP_{mask50}/\%$	$AP_{mask75}/\%$	检测总时间/s	t_d/ms
Y	69.75	74.19	64.24	19.27	50.19
A-Y	87.48	88.68	75.95	19.10	49.74
S-Y	75.30	83.56	69.84	19.93	51.90
C-Y	85.64	89.97	76.49	20.13	52.42
AS-Y	93.08	93.92	80.35	19.83	51.63
AC-Y	90.24	91.14	77.90	20.04	52.19
SC-Y	92.74	93.58	79.77	20.92	54.48
ASC-Y	95.22	95.60	81.32	20.60	53.65

示添加 CBAM 注意力机制, A_{cc50} 表示 IoU (Intersection over union) 为 0.5 时模型检测准确率, AP_{mask50} 和 AP_{mask75} 分别表示 IoU 为 0.5、0.75 时的掩膜平均准确率。

由表 2 可知,原始模型的 A_{cc50} 仅为 69.75%,融合 CBAM 注意力机制和 SPP 结构并且改变锚框长宽比的模型 A_{cc50} 最高,达到 95.22%。与原始模型相比,添加 CBAM 注意力机制后模型的 A_{cc50} 提升 15.89 个百分点,添加 SPP 结构后模型的 A_{cc50} 提升 5.55 个百分点,改变了 FPN 特征图锚框长宽比后模型的 A_{cc50} 提升效果最为明显,为 17.73 个百分点,表明所设计的锚框长宽比更加适合芦笋的检测。

在图像分割方面,添加 CBAM 注意力机制的模型在 AP_{mask50} 指标上明显优于未添加 CBAM 注意力机制的模型。A - YOLACT + +、C - YOLACT + +、AC - YOLACT + + 在添加了 SPP 结构后 AP_{mask50} 提升 5.24、3.61、4.46 个百分点,而 AP_{mask75} 则提升 4.40、3.28、3.42 个百分点,表明模型在继续添加了 SPP 结构后,FPN 减少了对芦笋图像相关特征重复提取的问题,并且分割性能的提升主要集中在 IoU 大于 0.75 时的高精度分割上,改善了模型对于芦笋局部细节区域的分割效果。

添加了 CBAM 注意力机制以及 SPP 模块后增加了计算量,但是改变了 FPN 特征图锚框长宽比提高了产生候选框的速度,模型可以更快更准确地定位到目标芦笋,最终改进后模型掩膜生成的准确率

($AP_{mask50} + AP_{mask75}$) 提升 27.80 个百分点,但是检测时间仅比原模型增加 3.46 ms,表明改进后的模型整体性能得到了提升。

图 8 为不同阈值下模型精准率与召回率变化曲线,从图中可以看出,在相同阈值的情况下,改进 YOLACT + + (ASC - YOLACT + +) 模型的精准率更高;当阈值大于 0.1 时,改进 YOLACT + + 的召回率较 YOLACT + + 更大,最大差值达 49.00 个百分点;改进 YOLACT + + 模型综合性能更好,泛化能力更强。

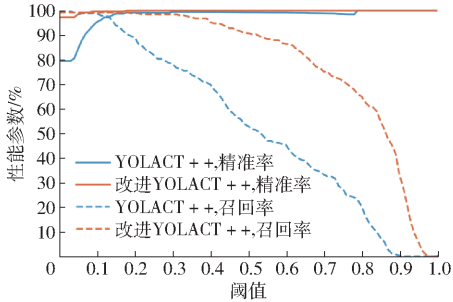


图 8 不同阈值下模型精准率与召回率变化曲线
Fig. 8 Variation curves of model accuracy and recall under different thresholds

4.1.2 模型改进前后识别分割效果

为了验证所改进模型在不同天气场景下的识别分割效果,本试验选取测试集中晴天顺光、晴天逆光、阴天和夜间补光 4 种不同天气光照条件下图像开展模型改进前后对比试验,结果如图 9 所示。

从图 9 可以看出,在检测方面,改进后的模型可

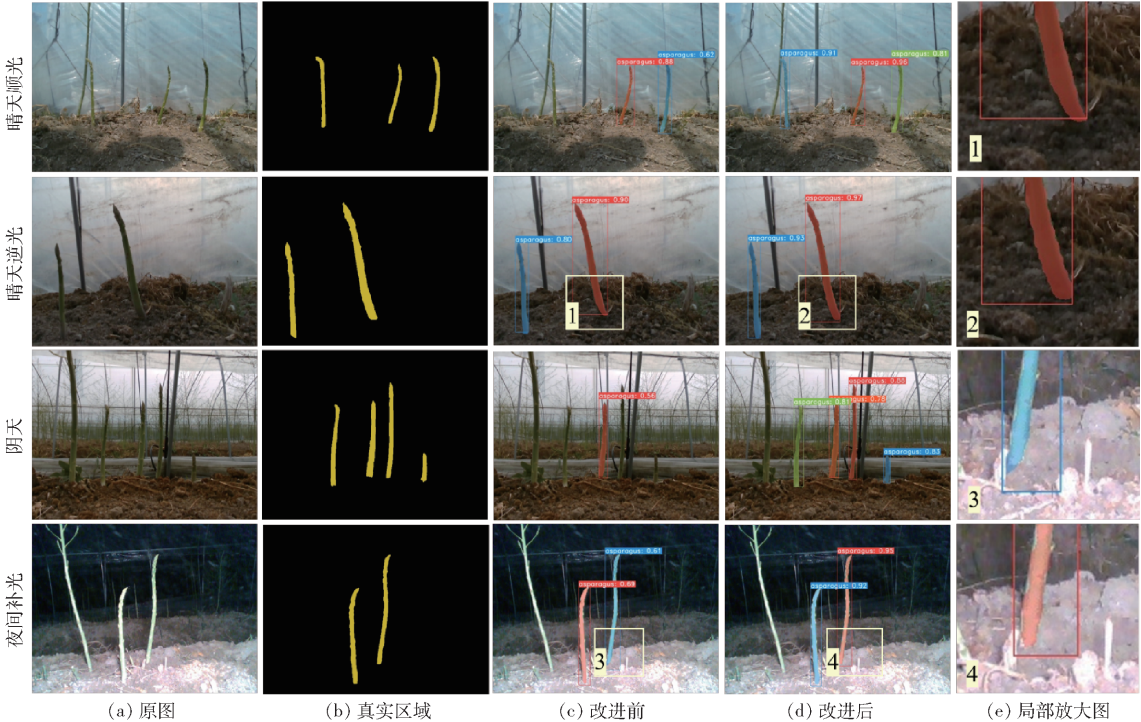


图 9 模型改进前后的检测结果
Fig. 9 Detection renderings before and after model improvement

以准确地检测出所有目标芦笋,没有出现漏检和误检的情况,并且相较于改进前检测精度更高。在分割方面,原模型出现分割不完整、过分割的情况,如图 9e 中框 1 与框 3 所示,前者所分割出来的芦笋基部掩膜与真实区域相差较大,这会使得芦笋基部直径的计算值远小于真实值;后者将部分土壤分割为芦笋,这会导致芦笋长度的计算值与真实值有较大偏差,分析其原因是由于芦笋纹理特征模糊以及土壤与芦笋底部颜色较为接近导致了该现象的发生。改进后模型对这些现象有明显改善,使得分割的芦笋掩膜较真实区域更为接近。

为了进一步验证 4 种天气光照条件下的识别分割效果,统计了改进前后模型的检测准确率、掩膜平均准确率和检测时间(表 3)。与原始模型相比,改进后模型在 4 种不同光照条件下的检测精度和分割精度均有较大提升,其中对阴天条件下提升最大, A_{cc50} 和 AP_{mask50} 分别提升 27.78、23.61 个百分点。在夜间补光条件下改进后模型的 A_{cc50} 最高,为 96.05%,在晴天顺光条件下改进后模型的 AP_{mask50} 最高,为 96.64%,说明在夜间补光以及晴天顺光条件下芦笋颜色、纹理以及轮廓更为清晰,较易被模型检测出。但是,在阴天光照条件下改进后模型的 A_{cc50} 和 AP_{mask50} 分别为 94.39%、95.04%,相比其他 3 种天气光照条件下,其准确率较低,可能的原因是在阴天光照条件下,光照强度较低,导致芦笋颜色与其他背景环境颜色相似,芦笋上半部分轮廓较为模糊。在晴天逆光条件下 AP_{mask50} 仅为 95.05%,分析其原因可能是因为芦笋在成像时呈现黑色,其下半部分轮廓较为模糊,与土壤连接处难以分辨导致掩膜平均准确率偏低。

表 3 模型改进前后对比结果

Tab.3 Comparison results before and after model improvement				
天气光照条件	模型	$A_{cc50}/\%$	$AP_{mask50}/\%$	t_d/ms
晴天顺光	改进前	69.97	74.76	50.09
	改进后	95.34	96.64	53.58
晴天逆光	改进前	70.33	75.16	50.15
	改进后	95.08	95.05	53.67
阴天	改进前	66.61	71.43	50.05
	改进后	94.39	95.04	53.63
夜间补光	改进前	72.10	75.41	50.45
	改进后	96.05	95.67	53.71
总计	改进前	69.75	74.19	50.19
	改进后	95.22	95.60	53.65

通过对不同天气光照条件下芦笋掩膜分割的可视化结果和性能指标统计结果的分析,本研究提出的融合 CBAM 注意力机制和 SPP 结构并改变锚框

长宽比的改进方法能有效解决输入芦笋图像中噪声干扰的问题,强化网络对芦笋局部特征的提取,提升检测分割结果。

4.1.3 不同算法对比试验

为进一步验证改进模型检测性能,本研究将改进 YOLACT++ 模型与 Mask R-CNN、SOLO 这 2 种常用实例分割网络进行对比试验,分别训练 2 种算法至收敛,然后对比这 3 种模型在测试集上的检测结果。图 10 为 3 种模型在训练过程中的掩膜平均准确率曲线,Mask R-CNN 模型在 15 000 ~ 30 000 次迭代时出现强烈振荡,30 000 次迭代后达到最大掩膜平均准确率;SOLO 模型大约在 12 000 次迭代后性能下降较为严重且振荡幅度相对较大,直到在 40 000 次迭代后才趋于稳定。与前两种模型相比,改进 YOLACT++ 模型的收敛速度更快,其在前 15 000 次迭代迅速拟合,掩膜平均准确率快速提升,然后逐渐趋于稳定,只有稍许的振荡。

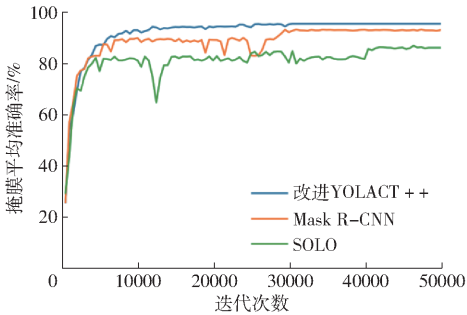


图 10 掩膜平均准确率曲线
Fig.10 Mask average accuracy curves

模型检测结果表明,Mask R-CNN 模型、SOLO 模型和改进 YOLACT++ 模型的 A_{cc50} 分别为 94.83%、88.71%、95.22%, AP_{mask50} 分别为 93.32%、86.27%、95.60%, 平均检测耗时分别为 238.01、108.69、53.65 ms。与 Mask R-CNN 模型相比,改进 YOLACT++ 模型的 A_{cc50} 和 AP_{mask50} 分别提高 0.39、2.28 个百分点;与 SOLO 模型相比,改进 YOLACT++ 模型的 A_{cc50} 和 AP_{mask50} 分别提高 6.51、9.33 个百分点;除此之外,改进 YOLACT++ 模型相较于 Mask R-CNN 模型和 SOLO 模型在检测速度上具有明显优势,单幅图像检测耗时可分别减少 184.36、55.04 ms。

通过对芦笋目标检测的可视化结果(图 11)可以看出,对于单根以及双根芦笋,3 种模型均能成功检测到目标芦笋;而对于多根芦笋,SOLO 模型则存在漏检现象,改进 YOLACT++ 模型与 Mask R-CNN 模型依然可以成功检测到所有芦笋。

综上所述,改进 YOLACT++ 模型在收敛速度、检测准确率、掩膜平均准确率以及检测耗时等性能

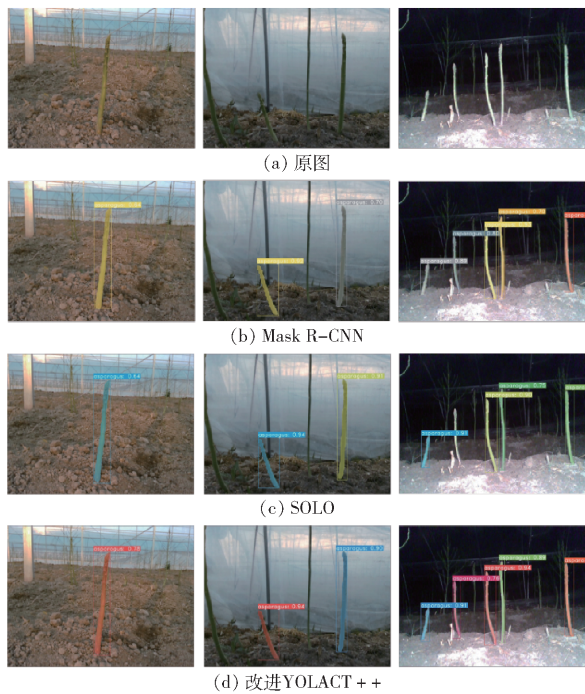


图 11 不同算法实际检测结果

Fig. 11 Actual detection effects of different algorithms

指标上均优于 Mask R - CNN 模型和 SOLO 模型。

4.2 成熟芦笋判别与切割面定位试验

为了验证成熟芦笋判别与切割面定位方法的性能,开展成熟芦笋判别与切割面定位试验。分别采用 Mask R - CNN 模型、SOLO 模型、YOLACT + + 模

型和改进 YOLACT + + 模型对芦笋进行检测分割后,判别出成熟芦笋并定位其采收切割面,每种模型共进行 21 次试验,通过人工测量芦笋长度和基部直径来判别成熟芦笋(图 12a、12b);在芦笋总长度 ε 处标记点位(图 12c),使用 D435i 型相机测得该点位在相机坐标系下坐标,操作时相机与莖面平行放置(图 12d),此处取 $\varepsilon = 3/20$ 进行切割点定位精度的评估;使用角度测量仪对芦笋滚转角和俯仰角进行测量(图 12e、12f)。通过人工获取的数据与本研究提出的成熟芦笋判别和切割面定位算法所评估出的结果进行对比来验证该方法的有效性。

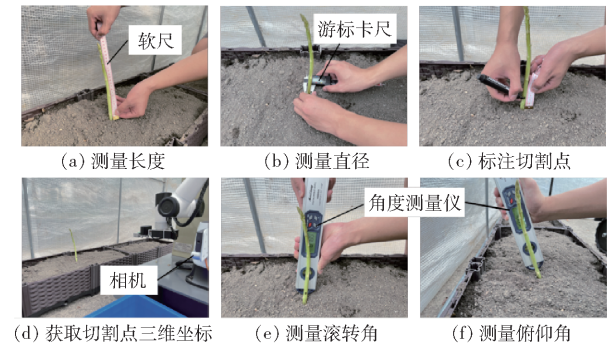


图 12 芦笋参数测量

Fig. 12 Asparagus parameter measurement

使用改进 YOLACT + + 模型进行成熟芦笋判别试验结果如表 4 所示,表中 0 为未成熟芦笋,1 为成

表 4 改进 YOLACT + + 模型成熟芦笋判别结果

Tab. 4 Mature asparagus discrimination results from improved YOLACT + + model								
试验 序号	测量长度 N_{mL}/cm	评估长度 N_{aL}/cm	长度评估误差 σ_L/cm	测量基部直径 N_{mD}/mm	评估基部直径 N_{aD}/mm	基部直径评估 误差 σ_D/mm	人工判定成熟 程度 M_m	算法判定成熟 程度 M_a
1	38.4	35.27	-3.13	13.9	11.85	-2.05	3	2
2	37.5	35.82	-1.68	18.5	18.72	0.22	3	3
3	36.5	35.75	-0.75	10.7	11.73	1.03	2	2
4	32.1	29.82	-2.28	14.8	12.70	-2.10	3	2
5	36.3	37.00	0.70	14.6	12.43	-2.17	3	2
6	39.2	36.17	-3.03	17.1	16.60	-0.50	3	3
7	34.9	32.82	-2.08	14.3	12.05	-2.25	3	2
8	27.3	26.85	-0.45	10.8	10.63	-0.17	2	2
9	26.5	25.29	-1.21	7.8	7.46	-0.34	1	1
10	31.5	30.87	-0.63	13.5	12.04	-1.46	3	2
11	33.7	31.89	-1.81	14.8	14.74	-0.06	3	3
12	33.4	32.11	-1.29	11.3	11.15	-0.15	2	2
13	29.7	28.47	-1.23	11.2	9.08	-2.12	2	2
14	25.2	27.01	1.81	12.0	13.15	1.15	2	0
15	35.7	33.23	-2.47	14.7	13.32	-1.38	3	3
16	35.0	34.18	-0.82	9.9	9.74	-0.16	2	2
17	20.9	19.85	-1.05	12.4	14.33	1.93	0	0
18	15.8	14.43	-1.37	8.5	10.27	1.77	0	0
19	19.8	19.98	0.18	16.4	16.06	-0.34	0	0
20	22.2	22.18	-0.02	11.3	10.79	-0.51	0	0
21	23.6	22.59	-1.01	12.1	9.20	-2.90	0	0
平均值	30.25	29.08	-1.17	12.89	12.29	-0.60		

熟细笋,2 为成熟中笋,3 为成熟大笋。21 次试验中,仅有试验 14 判别失败。试验 1 的长度评估误差最大,为 -3.13 cm ,试验 20 的长度评估误差最小,为 -0.02 cm ,平均评估误差为 -1.17 cm ,说明算法评估的大部分长度值小于真实值,这是由于芦笋嫩茎发生不同程度的弯曲导致的;除此之外,相较于长度较短的芦笋,长度较长的芦笋评估误差整体上偏大,说明其受到弯曲的影响更大。试验 21 的基部直径评估误差最大,为 -2.90 cm ,试验 16 的基部直径评估误差最小,为 -0.16 mm ,其中大笋被误判为中笋的情况出现最多,可能是在计算过程中受到滚转角测量误差的影响,导致算法评估的直径偏小。

为了进一步探究改进 YOLACT++ 模型的成熟

芦笋判别性能,对比该模型和其他 3 种模型的成熟芦笋判别与切割面定位的试验结果(表 5)。改进 YOLACT++ 模型的长度和基部直径的评估误差均为最小,分别为 $(-1.17 \pm 1.14)\text{ cm}$ 和 $(-0.60 \pm 1.34)\text{ mm}$,成熟芦笋判别准确率达到 95.24%。Mask R-CNN 模型的成熟芦笋判别准确率与改进 YOLACT++ 模型相同,但芦笋的成熟程度根据其长度以及直径进行判定,属于区间判断,后者的长度和直径评估误差均低于前者,容错区间更大,鲁棒性更强。与 SOLO、YOLACT++ 模型相比,改进 YOLACT++ 模型的成熟芦笋判别准确率分别提升 4.77、14.29 个百分点,并且长度和基部直径的评估误差均远低于前 2 种模型。

表 5 成熟芦笋判别与切割面定位结果
Tab.5 Identification of mature asparagus and positioning of cutting surface

模型	$\sigma_L \pm s_L /$ cm	$\sigma_D \pm s_D /$ mm	$M / \%$	$\sigma_X \pm s_X /$ mm	$\sigma_Y \pm s_Y /$ mm	$\sigma_Z \pm s_Z /$ mm	$\sigma_\psi \pm s_\psi /$ ($^\circ$)	$\sigma_\varphi \pm s_\varphi /$ ($^\circ$)	$t_p /$ ms
Mask R-CNN	-1.47 ± 1.21	-0.66 ± 2.04	95.24	2.07 ± 1.35	2.01 ± 1.58	2.07 ± 1.89	3.50 ± 2.89	4.47 ± 3.15	76.31
SOLO	-1.68 ± 1.52	-0.96 ± 2.20	90.47	2.53 ± 1.30	2.49 ± 1.81	2.51 ± 1.52	3.76 ± 3.28	4.73 ± 3.32	78.81
YOLACT++	-2.06 ± 1.89	-1.29 ± 2.28	80.95	3.18 ± 0.90	3.38 ± 1.27	3.48 ± 1.33	4.21 ± 3.64	6.47 ± 4.51	75.51
改进 YOLACT++	-1.17 ± 1.14	-0.60 ± 1.34	95.24	1.89 ± 0.92	1.82 ± 0.98	2.00 ± 0.89	2.71 ± 1.87	3.84 ± 3.33	76.28

改进 YOLACT++ 模型在 X、Y、Z 方向的最大定位误差不超过 2.89 mm,相比于其他 3 种模型分别降低 1.07、1.41、1.92 mm,并且在单方向定位上,改进 YOLACT++ 模型的定位误差以及定位标准差均为最小。

对于芦笋位姿估计,改进 YOLACT++ 模型的滚转角以及俯仰角最大评估误差分别为 4.58° 和 7.17° ,相比于其他 3 种模型,滚转角评估误差分别降低 1.81° 、 2.46° 、 3.27° ,在俯仰角评估误差上分别降低了 0.45° 、 0.88° 和 3.81° 。4 种模型的俯仰角评估误差整体上比滚转角评估误差大,可能是 Z 方向定位误差相较于 X 方向定位误差偏大造成的。相比于 Mask R-CNN 模型和 SOLO 模型,改进 YOLACT++ 模型在判别与定位时间上最大减少 2.53 ms,相比于 YOLACT++ 模型仅仅只增加 0.77 ms,而该方法的其他性能均高于 YOLACT++ 模型。

从成熟芦笋判别与切割面定位试验的综合效果来看,改进 YOLACT++ 模型在各方面的性能指标上均优于其他 3 种模型。

4.3 机器人采收试验

为了验证本研究所提算法的实地适用性,在江苏省红窑镇春华园芦笋种植基地进行了实地采收试验。机器人采收流程包括视觉定位、夹持剪切、运输和放置 4 个步骤(图 13)。

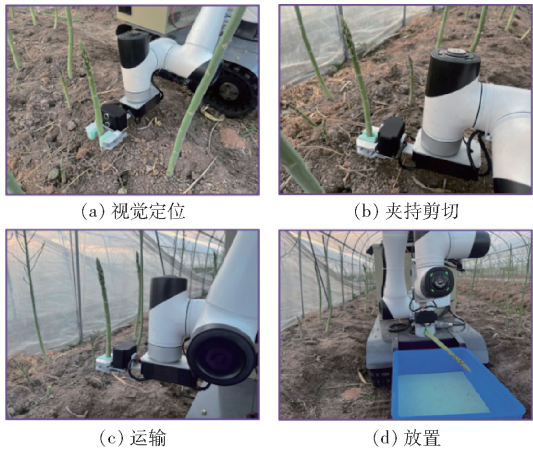


图 13 采收流程

Fig.13 Harvesting process

对 78 根芦笋进行采收,其中采收失败 3 根,采收成功率达 96.15%,平均采收单根芦笋耗时 12.15 s,表明该算法具有较高的鲁棒性和较强的适应能力。

5 结论

(1) 在芦笋检测试验中,与原始 YOLACT++ 模型相比,改进 YOLACT++ 模型对芦笋的检测准确率 $A_{\text{cc0.5}}$ 提升 25.47 个百分点,掩膜平均准确率 $AP_{\text{mask0.5}}$ 提升 21.41 个百分点;与 Mask R-CNN 模型、SOLO 模型相比,改进模型的 $A_{\text{cc0.5}}$ 分别提升 0.39、6.65 个百分点, $AP_{\text{mask0.5}}$ 分别提升 2.28、9.33

个百分点,检测耗时分别减少 184.36、55.04 ms。改进模型在检测精度以及分割精度方面都得到提升,并且在检测耗时上相较于其他算法优势显著,更有利于实现成熟芦笋的快速检测。

(2) 基于改进 YOLACT++ 模型的成熟芦笋判别准确率为 95.24%,与基于 Mask R-CNN 模型持平,相较于基于 SOLO 模型和传统的 YOLACT++ 模型分别提升 4.77、14.29 个百分点;其次,改进 YOLACT++ 模型在 X 、 Y 、 Z 方向定位误差最大仅为 2.89 mm,滚转角和俯仰角最大评估误差分别为 4.58°和 7.17°,相比于 Mask R-CNN 模型、SOLO 模

型和传统的 YOLACT++ 模型,最大定位误差分别降低 1.07、1.41、1.92 mm,滚转角误差分别降低 1.81°、2.46°、3.27°,俯仰角误差分别降低 0.45°、0.88°、3.81°。基于改进 YOLACT++ 模型的成熟芦笋判别准确率最高、定位误差最低,进一步降低了对成熟芦笋的定位误差以及位姿估计误差。

(3) 基于提出的方法试制的芦笋选择性采收机器人,采收成功率为 96.15%,采收单根芦笋平均耗时仅为 12.15 s,在设施环境下采用该方法能较好地进行成熟芦笋快速检测、判别与定位,实现芦笋的精准选择性收获。

参 考 文 献

- [1] 宋德庆,张劲,薛忠,等.我国芦笋产业发展现状与对策[J].广东农业科学,2011,38(21):187-190.
SONG Deqing, ZHANG Jing, XUE Zhong, et al. The development and countermeasures of asparagus industry in China[J]. Guangdong Agricultural Science, 2011, 38(21):187-190. (in Chinese)
- [2] 薛忠,宋德庆,刘恩平,等.国内外芦笋种植收获机械化研究进展[J].中国农业科技导报,2014(2):79-84.
XUE Zhong, SONG Deqing, LIU Enping, et al. Frontier of asparagus planting and harvesting mechanization in the world[J]. Journal of Agricultural Science and Technology, 2014(2):79-84. (in Chinese)
- [3] 陈度,ZHANG Qin,王书茂,等.芦笋机械化收获技术现状与发展分析[J].中国农业大学学报,2016,21(4):113-120.
CHEN Du, ZHANG Qin, WANG Shumao, et al. Current status and future solutions for asparagus mechanical harvesting[J]. Journal of China Agricultural University, 2016, 21(4):113-120. (in Chinese)
- [4] 刘继展.温室采摘机器人技术研究进展分析[J].农业机械学报,2017,48(12):1-18.
LIU Jizhan. Research progress analysis of robotic harvesting technologies in greenhouse[J]. Transactions of the Chinese Society for Agricultural Machinery, 2017, 48(12):1-18. (in Chinese)
- [5] 苑进.选择性收获机器人技术研究进展与分析[J].农业机械学报,2020,51(9):1-17.
YUAN Jin. Research progress analysis of robotics selective harvesting technologies[J]. Transactions of the Chinese Society for Agricultural Machinery, 2020, 51(9):1-17. (in Chinese)
- [6] IRIE N, TAGUCHI N, HORIE T, et al. Asparagus harvesting robot coordinated with 3-D vision sensor[C]//2009 IEEE International Conference on Industrial Technology. IEEE, 2009:1-6.
- [7] KENNEDY G, ILA V, MAHONY R. A perception pipeline for robotic harvesting of green asparagus[J]. IFAC-PapersOnLine, 2019, 52(30):288-293.
- [8] SAKAI H, SHIIGI T, KONDO N, et al. Accurate position detecting during asparagus spear harvesting using a laser sensor[J]. Engineering in Agriculture, Environment and Food, 2013, 6(3):105-110.
- [9] IVANOVA J F, KAMPE H, LIU X, et al. Vision based recognition and tracking of green asparagus in ambient light[J]. at-Automatisierungstechnik, 2017, 65(4):280-292.
- [10] LEU A, RAZAVI M, LANGSTÄDTLER L, et al. Robotic green asparagus selective harvesting[J]. IEEE/ASME Transactions on Mechatronics, 2017, 22(6):2401-2410.
- [11] PEEBLES M, LIM S H, DUKE M, et al. Overview of sensor technologies used for 3D localization of asparagus spears for robotic harvesting[J]. Applied Mechanics and Materials, 2018, 884:77-85.
- [12] PEEBLES M, LIM S H, STREETER L, et al. Ground plane segmentation of time-of-flight images for asparagus harvesting[C]//2018 International Conference on Image and Vision Computing New Zealand (IVCNZ). IEEE, 2018:1-6.
- [13] PEEBLES M, LIM S H, DUKE M, et al. Investigation of optimal network architecture for asparagus spear detection in robotic harvesting[J]. IFAC-PapersOnLine, 2019, 52(30):283-287.
- [14] NARANJO-TORRES J, MORA M, HERNÁNDEZ-GARCÍA R, et al. A review of convolutional neural network applied to fruit image processing[J]. Applied Sciences, 2020, 10(10):3443.
- [15] PEEBLES M, BARNETT J J, DUKE M, et al. Robotic harvesting of asparagus using machine learning and time-of-flight imaging-overview of development and field trials[C]//2020 IEEE 16th International Conference on Automation Science and Engineering (CASE). IEEE, 2020:1361-1366.
- [16] 孙红,李松,李民赞,等.农业信息成像感知与深度学习应用研究进展[J].农业机械学报,2020,51(5):1-17.
SUN Hong, LI Song, LI Minzan, et al. Research progress of image sensing and deep learning in agriculture[J]. Transactions of the Chinese Society for Agricultural Machinery, 2020, 51(5):1-17. (in Chinese)
- [17] HAMUDA E, GLAVIN M, JONES E. A survey of image processing techniques for plant extraction and segmentation in the field[J]. Computers and Electronics in Agriculture, 2016, 125:184-199.
- [18] 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.

- [19] 贾伟宽,李倩雯,张中华,等. 优化 SOLO 算法分割复杂环境下柿子和苹果绿色果实[J]. 农业工程学报,2021,37(18): 121–127.
JIA Weikuan, LI Qianwen, ZHANG Zhonghua, et al. Optimized SOLO segmentation algorithm for the green fruits of persimmons and apples in complex environments[J]. Transactions of the CSAE,2021,37(18):121–127. (in Chinese)
- [20] ZHENG C, CHEN P, PANG J, et al. A mango picking vision algorithm on instance segmentation and key point detection from RGB images in an open orchard[J]. Biosystems Engineering,2021,206:32–54.
- [21] WAN Z, CHEN Y, DENG S, et al. Slender object detection: diagnoses and improvements[J]. arXiv Preprint arXiv:2011.08529, 2020.
- [22] 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.
- [23] 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)
- [24] ZHU H, XIE C, FEI Y, et al. Attention mechanisms in CNN-based single image super-resolution: a brief review and a new perspective[J]. Electronics,2021,10(10):1187–1188.
- [25] WOO S, PARK J, LEE J Y, et al. CBAM: convolutional block attention module[C]//Proceedings of the European Conference on Computer Vision (ECCV),2018:3–19.
- [26] HE K, ZHANG X, REN S, et al. Spatial pyramid pooling in deep convolutional networks for visual recognition[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence,2015,37(9):1904–1916.
- [27] 刘慧,姜建滨,沈跃,等. 基于改进 DeepLab V3 + 的果园场景多类别分割方法[J]. 农业机械学报,2022,53(11):255–261.
LIU Hui, JIANG Jianbin, SHEN Yue, et al. Multi-category segmentation of orchard scene based on improved DeepLab V3 + [J]. Transactions of the Chinese Society for Agricultural Machinery,2022,53(11):255–261. (in Chinese)
- [28] 中华人民共和国农业部. NY/T 1585—2008 芦笋等级规格[S]. 北京:中国农业出版社,2008.

(上接第 258 页)

- [19] RUSSELL B C, TORRALBA A, MURPHY K P, et al. LabelMe: a database and web-based tool for image annotation[J]. International Journal of Computer Vision, 2008, 77(1/3): 157–173.
- [20] LIN T Y, MAIRE M, BELONGIE S, et al. Microsoft COCO: common objects in context[C]//European Conference on Computer Vision, 2014: 740–755.
- [21] FELZENSZWALB P F, GIRSHICK R B, MCALLESTER D, et al. Object detection with discriminatively trained part-based models[J]. IEEE Transactions on Software Engineering, 2010, 32(9): 1627–1645.
- [22] REN S, HE K, GIRSHICK R, et al. Faster R – CNN: towards real-time object detection with region proposal networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 39(6): 1137–1149.
- [23] HE K, GKIOXARI G, DOLLÁR P, et al. Mask R – CNN[C]//Proceedings of the IEEE International Conference on Computer Vision, 2017: 2961–2969.
- [24] LIU J, LI P. A Mask R – CNN model with improved region proposal network for medical ultrasound image[C]//Proceedings of the International Conference on Intelligent Computing, 2018: 26–33.
- [25] HUANG Z, ZHONG Z, SUN L, et al. Mask R – CNN with pyramid attention network for scene text detection[C]//Proceedings of 2019 IEEE Winter Conference on Applications of Computer Vision, 2019: 764–772.
- [26] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016: 770–778.
- [27] SZEGEDY C, IOFFE S, VANHOUCKE V, et al. Inception-v4, inception-resnet and the impact of residual connections on learning[C]//Proceedings of the 31st AAAI Conference on Artificial Intelligence, 2017: 4278–4284.
- [28] WU Z, SHEN C, HENGEL A V D. Wider or deeper: revisiting the resnet model for visual recognition[J]. Pattern Recognition, 2019, 90: 119–133.
- [29] LIN T Y, DOLLAR P, GIRSHICK R, et al. Feature pyramid networks for object detection[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 2117–2125.
- [30] KIRILLOV A, GIRSHICK R, HE K, et al. Panoptic feature pyramid networks[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2019: 6399–6408.
- [31] 王辉,韩娜娜,吕程,等. 基于 Mask R – CNN 的单株柑橘树冠识别与分割[J]. 农业机械学报,2021,52(5):169–174.
WANG Hui, HAN Na'na, LÜ Chengxu, et al. Recognition and segmentation of individual citrus tree crown based on Mask R – CNN[J]. Transactions of the Chinese Society for Agricultural Machinery, 2021, 52(5):169–174. (in Chinese)