

基于改进 YOLOv8 的草莓识别与果梗采摘关键点检测

杨震宇¹, 汪小昆^{1,2*}, 祁子涵¹, 王得志¹

(1. 南京农业大学工学院, 南京 210031; 2. 江苏省现代农业设施农业技术与装备试验室, 南京 210031)

摘要: 为解决草莓采摘机器人工作过程中果梗采摘点定位精度低和遮挡草莓识别困难等问题, 该研究提出一种改进后 YOLOv8 算法与 pose 关键点检测算法相结合的草莓识别定位方法。通过对 YOLOv8 进行优化, 引入 BiFPN (bidirectional feature pyramid network) 和 GAM (generalized attention module) 模块以强化模型的双向信息流, 动态分配特征权重, 并专注于小目标特征的提取和强化被遮挡区域特征, 旨在提升模型复杂环境中采摘点定位准确率和遮挡识别的预测准确性。试验结果显示, 相比于原始模型, 经过改进的 YOLOv8-pose 模型在草莓识别准确率 (precision, P)、召回率 (recall, R)、平均精度 (mean average precision, mAP) 及关键点平均精度 (mean average precision - key point, mAP_{kp}) 上分别提高 6.01、1.98、6.67 和 7.85 个百分点, 基于关键点检测的果梗采摘点在 X 、 Y 、 Z 方向定位误差分别为 1.4、1.4 和 2.2 mm。此外, 根据草莓遮挡重叠区域面积对草莓遮挡程度分类, 利用不同遮挡程度对模型性能验证, 在遮挡情况下改进后 YOLOv8-pose, mAP_{kp} 比原模型提高 9.78 个百分点。基于该研究所提出的视觉模型, 机器人在田间试验下的采摘成功率为 95%, 单个草莓的采摘耗时 10 s, 可为实现机器人精准采摘提供重要的技术支持。

关键词: 图像识别; 草莓; YOLOv8; 采摘; 遮挡

doi: 10.11975/j.issn.1002-6819.202405044

中图分类号: TP391.4; S24

文献标志码: A

文章编号: 1002-6819(2024)-18-0167-09

杨震宇, 汪小昆, 祁子涵, 等. 基于改进 YOLOv8 的草莓识别与果梗采摘关键点检测[J]. 农业工程学报, 2024, 40(18): 167-175. doi: 10.11975/j.issn.1002-6819.202405044 <http://www.tcsae.org>

YANG Zhenyu, WANG Xiaochan, QI Zihan, et al. Recognizing strawberry to detect the key points for peduncle picking using improved YOLOv8 model[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2024, 40(18): 167-175. (in Chinese with English abstract) doi: 10.11975/j.issn.1002-6819.202405044 <http://www.tcsae.org>

0 引言

草莓是一种典型的季节性水果, 其果实易在抓取采摘过程中遭受损伤^[1-2]。目前, 草莓采摘机器人的采收方式主要是通过剪切果梗实现^[3-4]。然而, 草莓的果梗直径较小, 加之农业环境中光照等多种因素导致的复杂背景干扰, 识别定位难度较大。此外, 高架种植模式下果实遮挡使得草莓的准确检测变得困难, 这对草莓机器人的精准识别与定位技术提出了挑战。

近年来, 国内外研发人员已经在草莓的识别方面进行大量的研究。基于传统的图像处理方法通常需要通过 RGB 图像进行颜色空间转换、白平衡等方式识别^[5-9]。CAO 等^[10]采用 Lab 颜色模型将草莓的原始图像转化为灰色图像, 并通过边缘检测 Canny 技术分析草莓轮廓来判断其形状。XU 等^[11]采用 HOG 特征, 通过计算图像局部区域的梯度方向直方图描述草莓的形状与外观信息。以上图像处理技术算法虽然容易实现, 但是对环境的变化不够敏感, 处理遮挡问题时鲁棒性较差。

随着计算机技术与机器视觉技术的发展, 深度学习

逐渐被应用到农业果实收获领域, 极大提高果实识别准确率^[12]。YU 等^[13]提出一种基于 Mask R-CNN 草莓识别方法, 采用残差网络与特征金字塔网络进行提取特征, 平均准确率达到 95.78%。贾伟宽等^[14]提出一种基于 SOLO 模型的绿色果实优化算法, 该算法能有效提取绿色果实的多尺度特征, 展现出较高的识别准确率。YU 等^[15]提出基于 YOLO 检测技术, 结合旋转自适应边界框方法来预测草莓的位姿, 同时通过提取草莓中轴线顶点上方 2~3 cm 处的特征点预判采摘点。然而, 在复杂的农业环境下, 仅通过成熟草莓进行识别及几何估算采摘点无法满足准确剪切果梗的要求。

关键点检测技术起初在人体姿态估计领域得到了广泛应用。JEONG 等^[16]采用了 Open Pose 算法, 通过分析人体骨骼的关键点来判断是否存在吸烟行为。目前, 该技术已被广泛引入到农业生产领域。WU 等^[17]提出了一种自上而下葡萄茎定位策略, 该策略借鉴了人体姿势分析的概念。通过采用目标识别算法来锁定葡萄串, 结合关键点检测技术以实现茎部精确定位。

针对上述问题, 本文将 YOLOv8 和关键特征点 (pose) 检测算法相结合, 在 YOLOv8-pose 基础上融合 GAM (generalized attention module) 和 BiFPN (bidirectional feature pyramid network) 模块, 实现成熟草莓的识别和关键点检测, 并在遮挡情况下实现对草莓的准确识别。成功匹配到准确的关键点后, 选择适当的关键点

收稿日期: 2024-05-09 修订日期: 2024-08-13

基金项目: 江苏省农业科技自主创新资金 (CX (21) 2006)

作者简介: 杨震宇, 研究方向为农业机器人技术。

Email: 304144369@qq.com

*通信作者: 汪小昆, 教授, 博士生导师, 研究方向为设施农业装备智能化控制技术。Email: wangxiaochan@njau.edu.cn

作为草莓的剪切位置。

1 成熟草莓的识别与关键点检测方法

成熟草莓的识别和关键点的检测主要通过以下步骤实现：首先，使用改进的 YOLOv8 目标检测网络识别可采摘的成熟草莓；其次，通过边界框信息结合草莓的形状特征，拟合出草莓果实和果梗的关键点；最后，对关键点进行拼接，获取果梗合适采摘点的位置。

1.1 YOLOv8 网络模型

YOLOv8 算法由 Glenn-Jocher 提出，与 YOLOv3 算法、YOLOv5 算法一脉相承^[18-19]。其网络结构主要由输入端 (Input)、骨干网络 (Backbone)、颈部网络 (Neck) 和检测头 (Head) 组成。在 Input 部分，YOLOv8 引用 Mosaic 图像增强技术，使模型能够在各个尺度和方向上进行训练学习，有助于提高模型检测的泛化能力；在 Backbone 部分，YOLOv8 引入 C2F 和 SPPF (spatial pyramid pooling fast) 等模块对目标特征进行提取。其中 C2F 模块融合 YOLOv7 中 ELAN 思想，通过对不同层次的特征进行有效传播，提高网络模型的收敛性能；在

Neck 部分，YOLOv8 引入特征金字塔网络 FPN (feature pyramid network)^[20] 和路径聚合网络 PAN (path aggregation network)^[21] 对不同层的特征信息进行尺度融合，提高模型的检测性能；在 Head 部分，YOLOv8 引入前沿的解耦头结构，将分类和检测头分别进行处理。为适应不同形状和大小的物体以及提高模型检测的鲁棒性，该模型采用无锚框 (Anchor-Free) 设计。然而，该设计在识别小目标或被遮挡下的目标时精度较低。

1.2 YOLOv8 模型改进

本研究在 YOLOv8 模型框架上进行了拓展，设计了一种优化的 pose 算法，该算法不仅有效集成目标检测功能，而且还实现对草莓关键部位坐标的精确预测。在 YOLOv8 的 Neck 部分加入 BiFPN 模块，该模块能进行高层次特征和低层次特征的双向交互和融合，增强对小目标和被遮挡目标的识别精度，使模型实现更准确的预测。另外，为进一步优化关键点局部特征的估计性能，本研究还引入了 GAM 机制，对空间和通道的注意力进行调整，提高被遮挡关键点的检测精度。改进 YOLOv8 网络结构如图 1 所示。

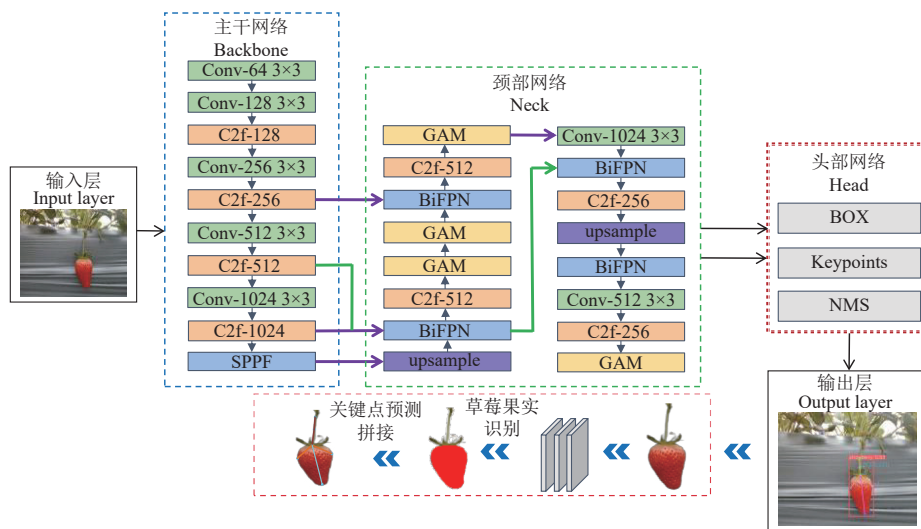


图 1 改进 YOLOv8 的模型结构
Fig.1 Improving the YOLOv8 architecture

1.2.1 GAM 模块

由于草莓与其周围环境 (如相邻草莓果实) 的视觉相似性较高，草莓关键点的定位容易出现误差和不确定性。为此，本文在 YOLOv8 模型中集成了 GAM，通过引入 GAM 能够增强特征的表达和选择，使模型更好地聚焦于草莓的遮挡关键点特征^[21]。

注意力机制分为通道注意力机制和空间注意力机制，但只将这两种注意力机制进行融合放置，如 CBAM，虽然能够增加模型检测性能，却丢失了跨维度的信息^[22]。GAM 注意力机制是在 CBAM 的基础上，采用序贯的通道-空间注意力机制来重新设计 CBAM 模块，过程结构图如图 2 所示。在模型中 GAM 被放在每个特征图层级上的 C2f 操作之后，通过在每个特征图层级上使用

GAM，模型可以更好地对各种不同尺度的目标特征进行处理、筛选和强化。GAM_attention 的计算过程如下：

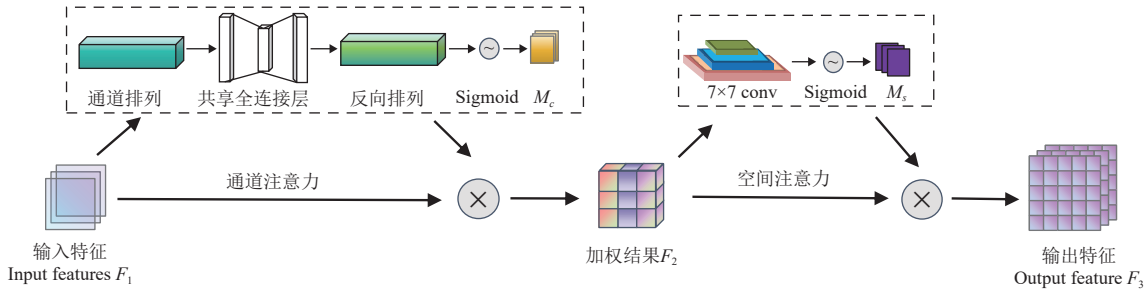
$$F_2 = M_c(F_1) \otimes F_2 \quad (1)$$

$$F_3 = M_s(F_2) \otimes F_1 \quad (2)$$

式中 F_1 输入图片的特征信息， $F_1 \in R^{C \times H \times W}$ ，其中 C 为通道数， H 为图像的宽度， W 为图像的高度， F_2 为中间加权结果， F_3 为输出图片的特征信息， M_c 和 M_s 分别为通道注意力和空间注意力的输出结果， \otimes 表示矩阵运算。

1.2.2 BiFPN 模块

草莓果梗的直径较小，其不同果梗的形态存在较大的差异，并受到果实间相互遮挡等多种复杂环境因素的影响，这些因素共同增加了草莓识别任务的复杂性。



注： M_c 表示通道注意力输出结果； M_s 表示空间注意力输出结果，下同。
 Note: M_c represents channel attention output results; M_s represents spatial attention output results, the same below.

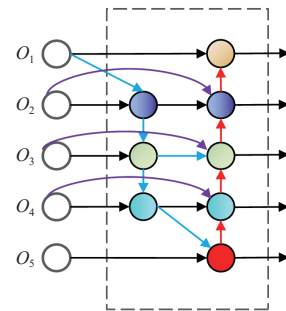
图 2 GAM 注意力机制
 Fig.2 GAM attention mechanism

针对这种情况，本研究引入 BiFPN，它是一种特征金字塔网络，能够在不同尺度和层次之间完成双向的信息流动，可以实现较好地特征融合^[23]，有助于提高对小目标任务的检测精度。该模块引入一种自适应权重分配机制，根据特征重要性动态分配权重，能在遮挡情况下强化被遮挡部分的特征，同时抑制遮挡区域和背景噪声，有利于模型在遮挡情况下准确识别草莓，其结构如图 3。

1.3 基于关键点检测的草莓果梗定位

为减轻机器人采摘草莓过程中带来的机械损伤和采摘后的果实附着果梗过长等问题。本文提出关键点检测的方法来对草莓采摘点进行定位。首先，利用 YOLOv8 网络对成熟草莓进行关键点检测；其次，从相应先验区域的特征图中解码出关键点的位置信息；最后，通过分析果梗上关键点信息，验证采摘位置的合理性，并选取

合适的位置作为果梗采摘点，定位流程图如图 4 所示。



注： $O_1 \sim O_5$ 为不同层级特征，下同
 Note: O_1 to O_5 represent different levels of feature, the same below.

图 3 BiFPN 模块
 Fig.3 BiFPN module

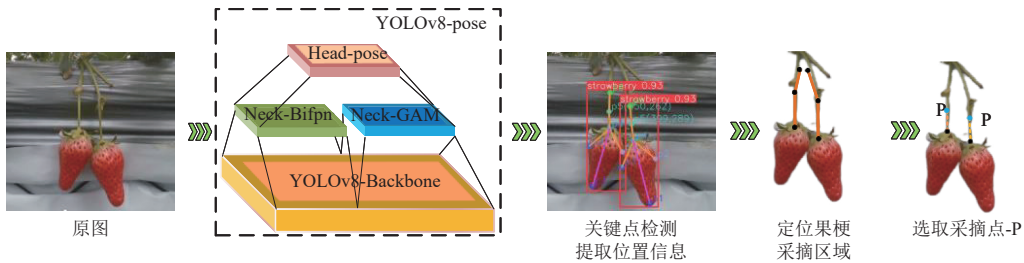


图 4 果梗采摘点定位流程图
 Fig.4 Flowchart for locating stem picking points

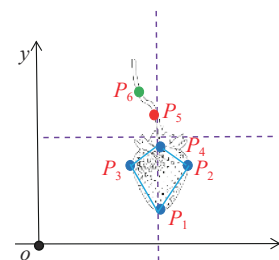
1.3.1 草莓关键点选取

由于草莓的农艺环境复杂，会出现果实倾斜、果梗细且短小等问题^[24]。因此，选取适宜的关键点来表征草莓形态特性，是视觉定位系统中的关键环节。本研究针对草莓形态特性，提出一种草莓关键点选取方法，如图 5 所示。

根据草莓近似椭圆的形状特征，在其轮廓上设置 4 个关键顶点 P_1 、 P_2 、 P_3 、 P_4 。其中 P_1 和 P_4 位于草莓的中部轴线上，反映草莓中部的宽度变化， P_2 和 P_3 位于草莓底部的两侧，反映草莓底部的弧度变化。经草莓果实关键点的位置特征描述，体现出草莓外形的左右对称和整体曲率特性。草莓实与果梗交界处上方选取 P_5 和 P_6 和作为果梗关键点。

通过上述草莓果实关键点的对称性和曲率有助于确

定草莓的位姿，确保视觉模型能够准确区分果实和果梗，为后续果梗采摘点准确定位提供基础。



注： $P_1 \sim P_6$ 为草莓上 6 个关键点，下同。
 Note: P_1 to P_6 are six key points on the strawberry, the same below.

图 5 草莓关键点选取结构图
 Fig.5 Strawberry key points selection structure

1.3.2 定位果梗采摘区域

如前所述, 果梗部分包含 3 个关键点分别是 P_4 (果实的顶点)、 P_5 和 P_6 , 这 3 点连线距离近似表示为草莓果梗的长度即采摘区域。当 YOLOv8 网络检测到成熟草莓及果梗的关键点后, 提取 P_4 、 P_5 和 P_6 三点的像素坐标值, 使用 RealSense D435i 相机中 pyrealsense2 库的 `get_distance` 函数将像素坐标系 ouv 下果梗的 3 个关键点 P_4 、 P_5 和 P_6 转换为相机坐标系下 $O_cX_cY_cZ_c$ 的真实坐标, 其中深度信息 z 由相机内部的深度传感器获取, 由上述关键点的真实坐标来计算果梗各段的长度。长度计算式为

$$S^i = \sum_{n=0}^2 \sqrt{(X_{c(n+1)}^i - X_{c(n)}^i)^2 + (Y_{c(n+1)}^i - Y_{c(n)}^i)^2} \quad (3)$$

式中 S^i 为草莓果梗的总长度, $X_{c(n)}^i$ 、 $Y_{c(n)}^i$ 分别为相机坐标系下的 P_4 、 P_5 关键点的坐标值, $i=1$, 表示 P_4 到 P_5 的长度, $i=2$ 时, 表示 P_5 到 P_6 的长度。

1.3.3 选取果梗采摘点

不同草莓的果梗长度存在差异, 在果梗采摘区域内进行剪切时, 部分草莓果实会附带过长的果梗, 对草莓销售的经济性产生不利影响。

根据草莓果农的人工采摘经验以及本次试验末端执行器的刀刃厚度判定草莓果梗采摘点, 判定标准如表 1 所示。如图 6 所示。

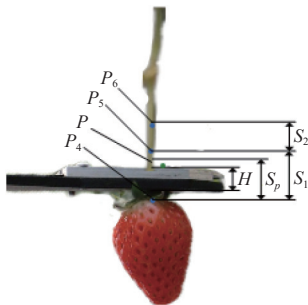
表 1 采摘点选取标准

Table 1 Criteria for selecting pick points

| 采摘点选取 Picking point selection | $S_p < S_1$ | $S_1 < S_p < S^t$ | $S_p > S^t$ |
|----------------------------------|-------------|-------------------|-------------|
| $S_1 \leq H$ | × | √ | ○ |
| $S_1 > H$ | √ | ○ | ○ |

注: S_1 表示果梗的第一段区域, H 表示刀刃厚度, S^t 表示果梗的总长度, S_p 表示采摘点到 P_4 的距离, √表示剪切后的果梗合理, ×表示剪切后的果梗过长, ○表示剪切后的果梗过短, 下同。

Note: S_1 is the first segment of the fruit stem region, H denotes the blade thickness, S^t is the total length of the fruit stem, and S_p is the distance from the picking point to P_4 , √ indicates a properly trimmed peduncle, × indicates a peduncle that is too short, and ○ indicates a peduncle that is too long after cutting, the same below.



注: S_2 表示果梗第二段区域, 下同。

Note: S_2 represents the second segment of the peduncle, the same below.

图 6 草莓采摘结构示意图

Fig.6 Diagram of strawberry picking

当果梗的第一段长度 S_1 小于刀刃厚度 H 时, 为避免刀刃接触到果实, 采摘点选取在果梗第二段区域位置的三分之一处 (β); 当果梗的第一段长度 S_1 大于刀刃厚

度 H 时, 为避免采摘后剩余果梗较长, 采摘点选取在果梗的第一段区域位置的三分之一处 (β)。因此, 草莓果梗采摘点位置定于果梗总长区域, 以果梗底部 P_4 点 (x_c^4, y_c^4, z_c^4) 为向量 \vec{v} 起点, P_6 (x_c^6, y_c^6, z_c^6) 为向量 \vec{v} 终点, 在 \vec{v} 上求解草莓果梗采摘点计算式为

$$\begin{bmatrix} x_c^p \\ y_c^p \\ z_c^p \end{bmatrix} = \begin{cases} \begin{bmatrix} \frac{1}{3}\beta & 1 - \frac{1}{3}\beta \end{bmatrix} \begin{bmatrix} x_c^6 & y_c^6 & z_c^6 \\ x_c^4 & y_c^4 & z_c^4 \end{bmatrix}, S_1 \leq H \\ \begin{bmatrix} \frac{1}{3}\alpha & 1 - \frac{1}{3}\alpha \end{bmatrix} \begin{bmatrix} x_c^6 & y_c^6 & z_c^6 \\ x_c^4 & y_c^4 & z_c^4 \end{bmatrix}, S_1 > H \end{cases} \quad (4)$$

式中 (x_c^p, y_c^p, z_c^p) 为相机坐标系下采摘点 P 坐标值。

1.4 遮挡程度分类

为精准评估模型对草莓果实不同遮挡程度下的识别性能, 本研究通过估计草莓遮挡区域的面积来量化草莓果实遮挡程度。根据人工经验选取 150 张不同果实遮挡程度下的草莓图像, 其中局部遮挡图像 83 张、半遮挡图像 49 张、全遮挡图像 18 张, 分别计算被遮挡草莓的重叠面积: 首先, 通过 Open CV 提取草莓图像的完整轮廓; 其次, 用 Photoshop 软件手工标注被遮挡的重叠区域轮廓, 记录被遮挡草莓轮廓内所有像素的总数 N_m 和重叠轮廓内所有像素的总数 N_r 。定义 N_m 和 N_r 的像素比值为遮挡比 D :

$$D = \frac{N_r}{N_m}$$

不同遮挡程度下的遮挡比, 如表 2 所示。通过分析可知: $5\% < D \leq 20\%$ 为局部遮挡; $20\% < D \leq 70\%$ 为半遮挡; $70\% < D \leq 100\%$ 为全遮挡。

表 2 不同遮挡程度下遮挡比

Table 2 Occlusion ratio under different degree

| 遮挡程度 Degree of occlusion | 遮挡区域像素值 Occluded region pixel value N_m | 重叠区域像素值 Overlapping region pixel value N_r | 遮挡比 Occlusion ratio/% |
|-----------------------------|---|--|-----------------------------|
| 局部遮挡 Partial occlusion | 3 056 2 120 | 157 415 | 5.14 19.6 |
| 半遮挡 Half occlusion | 3 200 3 543 | 748 2 532 | 23.4 71.48 |
| 全遮挡 Full occlusion | 2 945 2 453 | 2 158 2 361 | 73.29 96.26 |

1.5 图像采集和数据集处理

本次试验数据采集于 2024 年 2—3 月在江苏省南京市江宁区丰硕农场高架草莓种植基地, 草莓品种为章姬草莓。

本次试验采用 Intel Realsense D435i 深度相机, 以机械臂工作视域设定拍摄角度, 拍摄时的相机与被摄物体间距范围在 0.4~0.7 m, 采集高度为 0.5 m。该深度相机连接 Linux 操作系统工控机, 通过调用官方 API 编程接口方式进行工作, 所拍摄图片的分辨率为 640 像素×480 像素, 照片储存格式为 JPEG, 试验期间共收集白天、夜间、不同光照、以及果实遮挡、叶片遮挡等共 1 570 张草莓图像。为确保模型训练过程中的稳定性和增强其泛化能力, 将标记的 JSON 格式文件转换为 TXT 格式, 数据集按照 8:1:1 的比例分配至训练集、

验证集及测试集。

使用 Labelme 标注软件对成熟草莓进行关键点标注。以草莓果实果梗整体 (rectangle) 作为标签对象, 命名

为 strawberry, 草莓果实关键点 label 命名为 P_1 、 P_2 、 P_3 , 果梗处标注 P_4 、 P_5 、 P_6 , 其中被遮挡关键点部分在标注后缀 hide, 如图 7 所示。

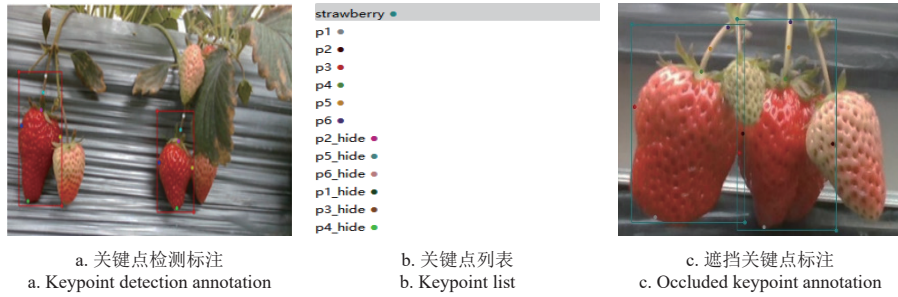


图 7 Label 图像标注

Fig.7 Label image annotation

1.6 模型训练配置

本研究使用模型训练和预测硬件环境: CPU 型号为 Intel Xeon Gold 522R, GPU 型号为 PNY Quadro RTXA 5000, 显存为 24GB, 训练周期设为 300, 批量大小设为 32, 早停策略值设为 2 防止过拟合。为确保模型训练和预测的稳定性, 所需软件环境均为: Ubuntu 20.04.06 LST 系统, 深度学习框架为 Pytorch1.11.1。

1.7 关键点检测评价指标

YOLOv8 的评价方式不同于传统的目标检测, 它通过对目标关键点相似度 OKS_p (object key point similarity) 来评价模型的性能。通过检测获得的 OKS_p 值来估算平均准确率 AP (average precision), 进而利用 AP 值计算得出关键点评估的指标 mAP_{kp} (mean average precision for key points) [25], 采用准确率 (precision, P)、召回率 (recall, R) 和草莓识别的平均准确率均值 (mean average precision, mAP) 作为草莓的识别指标, 用帧率 (frame per second, FPS) 来评价模型的检测速度, 计算式如下:

$$OKS_p = \frac{\sum_i \exp\{-d_{pi}^2 / 2s_p^2\sigma_i^2\} \delta}{\sum_i \delta} \quad (5)$$

式中 d_{pi} 为第 i 个关键点与目标对应的实际关键点的距离, s_p 为 p 点的尺度因子, σ_i 为 i 类型关键点的归一化因, δ 为检验关键点有效性的指示函数。

2 结果与分析

2.1 消融试验结果

改进后的 YOLOv8 在原有模型的基础融合了 BiFPN 模块、GAM 模块, 为验证改进后模型的合理性和有效性, 本文设计消融试验来进行模型性能, 验证结果如表 3 所示。

由表 3 可知, 原 YOLOv8 的 mAP_{kp} 、 P 、 R 、 mAP 分别为 88.69%、91.46%、90.14%、88.76%, 融合了 GAM 注意力机制后提升到 94.14%, 精确率 P 、召回率 R 和平均准确率均值 mAP 值也有小幅度提升, 试验结果表明引入 GAM 能改善模型的关键点准确率。融合 BiFPN 模块后, YOLOv8-B 的精确率 P 值从原模型的 91.46% 提升到 95.56%, 表明添加该模块后, 模型的识别准确率明显

改善。当模型同时融合 GAM 和 BiFPN 后, YOLOv8-G-B 的关键点平均准确率均值 mAP_{kp} 由 88.69% 提升到 96.54%, 精确率 P 从 91.46% 提升到了 97.48%, 召回率 R 从 90.14% 提升到了 92.12%, 平均准确率均值 mAP 从 88.76% 提升到 95.43%, 试验结果表明, GAM 和 BiFPN 模块具备更强的特征表示能力, 从而提高了对草莓关键点检测能力和遮挡草莓的分类效果。

表 3 消融试验

Table 3 Ablation experiments

| 模型 Models | $mAP_{kp}/\%$ | $P/\%$ | $R/\%$ | $mAP/\%$ | FPS/帧·s ⁻¹ |
|------------|---------------|--------|--------|----------|-----------------------|
| YOLOv8 | 88.69 | 91.46 | 90.14 | 88.76 | 85 |
| YOLOv8-G | 94.14 | 92.43 | 91.48 | 90.12 | 83 |
| YOLOv8-B | 92.45 | 95.56 | 91.94 | 91.79 | 81 |
| YOLOv8-G-B | 96.54 | 97.48 | 92.12 | 95.43 | 73 |

注: G、B 分别表示添加了 GAM 注意力机制和 BiFPN 结构; mAP_{kp} 表示关键点平均准确率均值; P 表示精确率; R 表示召回率; mAP 表示识别平均准确率均值; FPS 表示帧率, 下同。

Note: G adds a GAM attention mechanism; B incorporates a BiFPN structure. mAP_{kp} represents the mean average precision of keypoints; P stands for precision; R stands for recall; mAP represents the mean average precision for detection; FPS stands for frames per second, the same below.

2.2 不同模型检测效果对比

为验证改进后的模型性能, 对 YOLOv5、YOLOv7 以及改进前后的 YOLOv8 模型进行验证, 得到的评价结果如表 4 所示。

表 4 模型性能对比

Table 4 Model performance comparison

| 模型 Models | 内存 Memory/MB | $mAP_{kp}/\%$ | $P/\%$ | $R/\%$ | $mAP/\%$ | FPS/帧·s ⁻¹ |
|-----------------|--------------|---------------|--------|--------|----------|-----------------------|
| YOLOv5 | 89 | 73.68 | 72.31 | 71.34 | 82.19 | 90 |
| YOLOv7 | 53 | 82.32 | 86.32 | 81.57 | 87.13 | 84 |
| YOLOv8 | 52 | 88.69 | 91.46 | 90.14 | 88.76 | 85 |
| 改进 YOLOv8 | | | | | | |
| Improved YOLOv8 | 52 | 96.54 | 97.48 | 92.12 | 95.43 | 73 |

从不同模型的验证结果对比中可以发现, YOLOv8 在关键点和目标评价上的提升效果较为明显, 优于其他 YOLO 系列的关键点检测模型。改进 YOLOv8 相比原有 YOLOv8 的 mAP_{kp} 提升 7.85 个百分点, P 提升 6.02 个百分点, R 提升 1.98 个百分点, mAP 提升 6.67 个百分点。在引入 GAM 和 BiFPN 后虽然略微增模型内存 0.2 M, 但关键点定位精度和目标检测性能的显著改善对于精准采摘更为关键。

通过草莓关键点检测可视化结果图 8, 可以看出:

YOLOv5 进行检测时存在较为严重的漏检现象, YOLOv7 在检测时关键点拟合位置存在偏差, 而改进的 YOLOv8

相较其他模型其置信度提高效果明显, 可以满足机器人精准定位采摘点的工作要求。

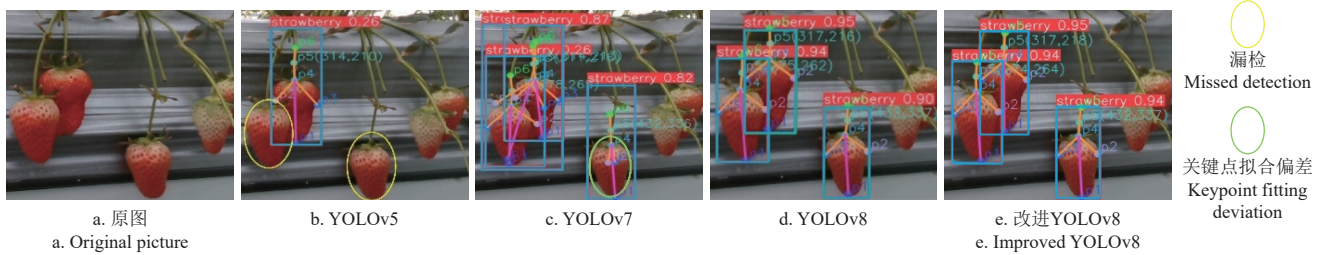


图 8 改进 YOLOv8 与不同模型检测对比
Fig.8 Comparison of YOLOv8 with different modules

2.3 模型改进前后遮挡检测效果对比

为验证改进后模型对于高架草莓果实遮挡复杂情况的关键点检测性能, 综合考虑机器人采摘过程中可能会遭遇的多种特殊遮挡情况, 本试验不仅从验证集中选取局部遮挡、半遮挡、完全遮挡这 3 种不同程度的果实遮挡场景进行验证, 还扩展叶片遮挡、果柄遮挡两种情况的模型检测, 以确保改进后的 YOLOv8 在各种复杂遮挡情况下的有效性和鲁棒性, 如图 9 所示。

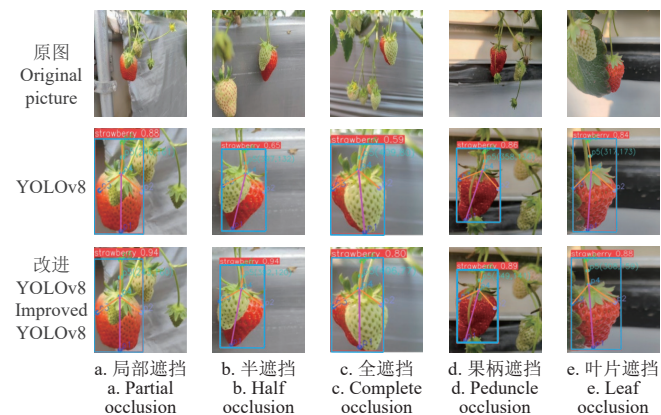


图 9 不同遮挡条件下的不同模型关键点检测对比
Fig.9 Comparison of keypoint detection by different models under various occlusion conditions

由对比结果可知: 在成熟草莓识别方面, 改进后的

YOLOv8 没有出现漏检或错检的情况, 并与改进前的 YOLOv8 相比置信度得到提升; 在关键点检测方面, 完全遮挡和果柄遮挡, 原模型检测的果梗关键点拟合错误, 分析原因是遮挡原本特征模糊的小目标果梗从而导致信息通道流失。当引入 BiFPN 和 GAM 模块之后会对这类情况有明显改善, 加强对果梗信息的聚焦, 减少关键点预测错误的情况出现, 识别结果如表 5 所示。

表 5 遮挡情况下关键点检测结果

| 模型 Models | mAP _{kp} /% | P/% | R/% | mAP/% | FPS |
|-----------|----------------------|-------|-------|-------|-----|
| YOLOv8 | 82.34 | 84.26 | 82.33 | 82.49 | 57 |
| 改进 YOLOv8 | 92.48 | 93.72 | 90.15 | 90.27 | 51 |

2.4 草莓果梗采摘点定位试验

本试验使用深度相机来验证采摘点位于相机坐标系下的真实坐标, 通过人工标定的数据和基于关键点检测算法估计出的结果进行对比来验证草莓果梗采摘点定位方法的有效性。

定位试验结果如表 6 所示。在 x 方向上采摘点定位最大误差为 2.4 mm, 平均误差为 1.4 mm; 在 y 方向上采摘点定位最大误差为 2.3 mm, 平均误差为 1.4 mm; 在 z 方向上采摘点定位最大误差为 4.8 mm, 平均误差为 2.2 mm, 满足草莓采摘机器人果梗采摘定位需求。

表 6 采摘点检测估计值及其误差

Table 6 The estimated values and errors of picking point detection

| 试验序号 Experiment No. | 采摘点模型检测估计值 Picking point from model detected value/mm | | | 采摘点人工标定值 Picking point from Manually labeled value/mm | | | 采摘点误差值 Picking point error value/mm | | |
|------------------------|--|-------|-------|--|-------|-------|--|------------|------------|
| | x | y | z | x' | y' | z' | δ_x | δ_y | δ_z |
| 1 | 176.5 | 143.1 | 507.3 | 177.9 | 140.4 | 505.4 | -1.4 | 2.3 | 1.9 |
| 2 | 104.8 | 79.6 | 430.7 | 107.2 | 77.1 | 425.9 | -2.4 | 2.5 | 4.8 |
| 3 | 53.4 | 102.7 | 560.9 | 52.1 | 101.5 | 558.2 | 1.3 | 1.2 | 2.7 |
| 4 | 101.9 | 49.7 | 516.8 | 100.3 | 49.0 | 517.2 | 1.6 | 0.7 | -0.4 |
| 5 | 3.7 | 87.3 | 480.2 | 3.1 | 86.3 | 482.1 | 0.6 | 1 | -1.9 |
| 6 | -17.7 | 115.9 | 500.6 | -17.3 | 115.1 | 501.3 | -0.4 | 0.8 | -0.7 |
| 7 | -120.4 | 13.8 | 600.4 | -121.9 | 11.6 | 598.1 | 1.5 | 2.2 | 1.9 |
| 8 | 98.4 | 120.7 | 536.4 | 99.1 | 119.3 | 535.1 | -0.7 | 1.4 | 1.3 |
| 9 | -199.4 | 12.4 | 400.1 | -196.7 | 12.1 | 397.2 | -2.7 | 0.3 | 2.8 |
| 10 | 198.1 | 147.3 | 601.8 | 199.7 | 149.4 | 598.3 | -1.6 | -2.1 | 3.5 |
| 平均值 Average | | | | | | | 1.4 | 1.4 | 2.2 |

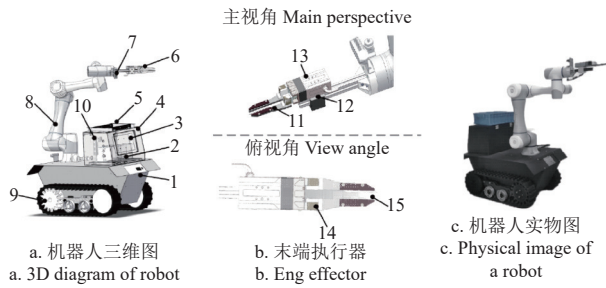
注: x 、 y 、 z 为模型检测采摘点的横向、纵向、深度估计值; x' 、 y' 、 z' 为人工标定采摘点横向、纵向、深度估计值; δ_x 、 δ_y 、 δ_z 为模型检测值与人工标定值之间的差值, 下同
Note: x , y , and z are the estimated horizontal, longitudinal, and depth values of the picking point from model detection; x' , y' , and z' are the manually labeled horizontal, longitudinal, and depth values; δ_x , δ_y , and δ_z are the differences between the model detection values and the manually labeled values, the same below.

2.5 采摘机器人田间试验

为了验证改进后的 YOLOv8 算法能否满足草莓采摘

机器人视觉系统对识别和定位的高效准确要求, 本研究将该算法部署在草莓采摘机器人上进行测试。该机器人

配备了六自由度协作机械臂，并选用了内嵌 RTX 3060Ti 的嵌入式工控机作为控制平台，能够满足草莓采摘机器人实时检测，如图 10 所示。



1. 工控机 2. 机械臂控制器 3. 电子触摸屏 4. 集成控制柜 5. 草莓收集篮 6. 末端执行器 7. Intel Realsense D435i 深度相机 8. 六自由度机械臂 9. 履带式移动平台 10. 安全控制开关 11. 切剪刀刃 12. 移动滑块 13. 机械手驱动器 14. 舵机 15. 夹持海绵
1. Industrial PC 2. Robot arm controller 3. Electronic touchscreen 4. Integrated control cabinet 5. Strawberry collection basket 6. End effector 7. Intel realsense D435i depth camera 8. Six-degree-of-freedom Robot Arm 9. Tracked mobile platform 10. safety control switch 11. Cutting shear blades 12. Moving slider 13. robot hand actuator 14. Servo motor 15. Sponge gripper

图 10 草莓采摘机器人实物图与三维建模图

Fig.10 Actual and 3D modeled images of the strawberry picking robot

本次试验在江苏省南京市江宁区丰硕农场高架草莓采摘园进行。机器人采摘顺序包括视觉识别、机械手剪切、运输收集箱等步骤。如图 11a 所示，机器人通过拍照姿态对获取到的草莓图像进行处理，识别并定位采摘点，识别完成后，将采摘点的坐标赋值给机械臂进行工作；当机械臂运动到采摘点后，末端执行器进行剪切果梗操作，如图 11b 所示；在末端执行器内部设计海绵体夹持装置，如图 10b 的 15 所示，当采摘草莓时，海绵体先对草莓果梗进行柔性夹持，随后刀刃将剩余果梗进行剪切完成采摘，通过夹剪采摘可以有效防止草莓果实剪切过程中发生掉落。如表 7 所示，田间试验的采摘结果通过三组试验数据进行统计分析，共计 60 次采摘循环。

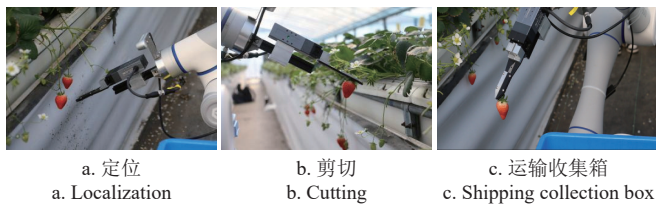


图 11 采摘流程示意图

Fig.11 Diagram of picking process

表 7 基于改进 YOLOv8 的田间试验采摘结果

| 循环次数 Number of iterations | 平均单颗识别时间 Average recognition time per item/s | 平均单颗采摘时间 Average picking time per item/s | 定位失败的草莓数量 Number of unidentified strawberry | 识别错误的草莓数量 Number of misidentified strawberry |
|------------------------------|---|---|--|---|
| 16 | 1.7 | 13 | 1 | 0 |
| 23 | 1.6 | 8 | 0 | 1 |
| 21 | 1.2 | 9 | 1 | 0 |

在本次试验中，共有 3 颗草莓采摘失败，采摘成功率为 95%。其中两次由于定位失败，1 次由于识别错误。

单颗草莓平均识别过程耗时 1.5 s，机器人采摘单颗草莓平均耗时 10 s，结果表明改进后的 YOLOv8 算法可以满足草莓采摘机器人的采摘性能要求。

3 结论

本研究采用 YOLOv8 算法和 pose 关键点检测算法相结合的方法，用于草莓识别和果梗采摘点定位，引入 BiFPN (bidirectional feature pyramid network) 和 GAM (generalized attention module) 模块，加强对小目标的特征提取，量化草莓果实间的遮挡程度并进行分类，最后与其他检测模型进行对比，得出主要结论如下：

1) 通过消融试验，与原 YOLOv8 模型相比，改进 YOLOv8 的关键点平均精度均值 (mean average precision-key point, mAP_{kp})、准确率 (Precision, P)、召回率 (Recall, R)、识别平均精度均值 (mean average precision, mAP) 分别提升 7.85 个百分点、6.02 个百分点、1.98 个百分点和 6.67 个百分点，与现有 YOLOv8 检测算法对比，本研究提方法实现更准确地识别定位果梗采摘点。

2) 与 YOLOv5、YOLOv7 等主流检测相比，本文所提出的改进 YOLOv8 模型关键点平均精度均值 mAP_{kp} 分别提升 22.86、14.22 和 7.85 个百分点。在不同遮挡条件下与原有模型 YOLOv8 相比，其关键点平均精度均值 mAP_{kp} 提升 10.14 个百分点。试验表明，改进后模型在不同遮挡条件下能够准确的识别定位果梗采摘点。

3) 经过人工标定和检测估计试验对比，本研究提出的采摘点定位算法误差在 x 、 y 、 z 方向的误差为 1.2 mm、1.6 mm、1.9 mm，为实现草莓采摘机器人精准采收提供有效支持。

4) 本文将 YOLOv8 算法改进与关键点识别技术，成功应用到高架草莓采摘机器人中，通过田间试验采摘成功率达到 95%，单颗采摘草莓耗时 10 s。

本研究采集并制作章姬等成熟表皮呈红色的草莓图像，但缺乏对其他品种草莓图像数据的训练，因此可能会限制模型在识别和定位其他颜色或不同特征草莓时的准确性和鲁棒性。后续研究中，将引入白珍珠、粉珍珠等成熟表皮呈淡粉或其他颜色的草莓，丰富草莓图像的样本数据集，提高模型对不同品种草莓的泛化能力和适应性。

[参考文献]

- [1] 谷青青, 柯裴蓓, 黄欣, 等. 基于 Web of Science 的草莓研究热点与趋势的可视化分析[J/OL]. 分子植物育种 (2023-11-20)[2024-05-07]. <http://kns.cnki.net/kcms/detail/46.1068.S.20231120.0845.002.html>.
GU Qingqing, KE Peibei, HUANG Xin, et al. Visualization analysis of research hotspots and trends in strawberry studies based on Web of Science[J/OL]. Molecular Plant Breeding: (2023-11-20)[2024-05-07]. <http://kns.cnki.net/kcms/detail/46.1068.S.20231120.0845.002.html>.

- [2] 孙兴民, 陈巍, 吕慧. 南京市草莓生产成本和综合收益研究[J]. *中国果业信息*, 2020, 37(8): 26-28, 31.
SUN Xingmin, CHEN Wei, LV Hui. Study on the production costs and comprehensive benefits of strawberry cultivation in Nanjing[J]. *China Fruit Industry Information*, 2020, 37(8): 26-28, 31. (in Chinese with English abstract)
- [3] 刘晓敏, 田德宝, 宋懋征, 等. 气动球果采摘柔性手爪设计与实验[J]. *农业机械学报*, 2021, 52(2): 30-43.
LIU Xiaomin, TIAN Debao, SONG Maozheng, et al. Design and experiment of a pneumatic flexible gripper for cone picking[J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2021, 52(2): 30-43. (in Chinese with English abstract)
- [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. Analysis of research progress on greenhouse harvesting robot technology[J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2017, 48(12): 1-18. (in Chinese with English abstract)
- [5] IRIE N, TAGUCHI N, HORIE T, et al. Asparagus picking robot coordinated with 3-D vision sensor[C]//2009 IEEE International Conference on Industrial Technology, Geelong, Victoria, Australia: IEEE, 2009: 1-6.
- [6] KENNEDY G, ILA V, MAHONY R. A perception pipeline for robotic picking of green asparagus[J]. *IFAC-PapersOnline*, 2019, 52(30): 288-293.
- [7] SAKAI H, SHIIGI T, KONDO N, et al. Accurate position detecting during asparagus spear picking using a laser sensor[J]. *Engineering in Agriculture, Environment and Food*, 2013, 6(3): 105-110.
- [8] 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.
- [9] LEU A, RAZAVI M, LANGSTÄD L, et al. Robotic green asparagus selective picking[J]. *IEEE/ASME Transactions on Mechatronics*, 2017, 22(6): 2401-2410.
- [10] CAO Qixin, NAGATA M, GEJIMA Y, et al. Basic study on strawberry picking robot[J]. *IFAC Proceedings Volumes*, 2000, 33(29): 49-54.
- [11] XU YW, IMOU K, KAIZU Y, et al. Two-stage approach for detecting slightly overlapping strawberries using HOG descriptor[J]. *Biosystems Engineering*, 2013, 115(2): 144-153.
- [12] 刘小刚, 范诚, 李加念, 等. 基于卷积神经网络的草莓识别方法[J]. *农业机械学报*, 2020, 51(2): 237-244.
LIU Xiaogang, FAN Cheng, LI Jianian et al. Strawberry recognition method based on convolutional neural networks[J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2020, 51(2): 237-244. (in Chinese with English abstract)
- [13] YU Y, ZHANG K, YANG L, et al. Fruit detection for strawberry picking robot in non-structural environment based on Mask-RCNN[J]. *Computers and Electronics in Agriculture*, 2019, 163(3): 104846.
- [14] 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. Optimization of SOLO algorithm for segmentation of green persimmon and apple fruits in complex environments[J]. *Transactions of the Chinese Society of Agricultural Engineering*, 2021, 37(18): 121-127. (in Chinese with English abstract)
- [15] YU Y, ZHANG K L, LIU H, et al. Real-time visual localization of the the picking for a ridge-planting strawberry harvesting robot [J]. *IEEE Access*, 2020, 8: 116556-116568.
- [16] JEONG T Y, HA L K. Open pose based smoking gesture recognition system using artificial neural network[J]. *Technical Journal*, 2023, 17(2): 251-259.
- [17] WU Z W, XIA F, ZHOU S Y, et al. A method for identifying grape stems using key points[J]. *Computers and Electronics in Agriculture*, 2023, 209(6): 107-125.
- [18] REDMON J, FARHADI A. YOLOv3: An Incremental Improvement[EB/OL]. arXiv preprint arXiv: 1804.02767, (2018-04-08)[2024-09-04]. Available: <https://arxiv.org/abs/1804.02767>.
- [19] JOHNSON G. , ULTRALYTICS. YOLOv8[EB/OL]. GitHub repository, (2023)[2024-09-04]. Available: <https://github.com/ultralytics/ultralytics>.
- [20] LIN T Y, DOLLÁR P, GIRSHICK R, et al. Feature pyramid networks for object detection[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu: IEEE, 2017: 936-944.
- [21] LIU S, QI L, QIN H, et al. Path aggregation network for instance segmentation[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018(4): 8759-8768.
- [22] 鲍禹辰, 徐增波, 田丙强. 基于 YOLOv8 改进的服装疵点检测算法[J]. *东华大学学报(自然科学版)*, 2024, 50(4): 49-56.
BAO Yuchen, XU Zengbo, TIAN Bingqiang. Improved YOLOv8-based algorithm for clothing defect detection[J]. *Journal of Donghua University (Natural Science Edition)*, 2024, 50(4): 49-56. DOI: 10.19886/j.cnki.dhdz.2023.0296.
- [23] 周鑫, 郝万君, 卞长庚, 等. 基于 g~nConv 和 GAM 的 YOLOv5 钢管焊接缺陷检测方法[J]. *微电子学与计算机*, 2023, 40(09): 29-37.
ZHOU Xin, HAO Wanjun, BIAN Changgeng, et al. YOLOv5-based steel pipe welding defect detection method using g~n Conv and GAM[J]. *Microelectronics & Computers*, 2023, 40(09): 29-37. DOI: 10.19304/J.ISSN1000-7180.2022.0778.
- [24] 张凯良, 杨丽, 王粮局, 等. 高架草莓采摘机器人设计与试验[J]. *农业机械学报*, 2012, 43(9): 165-172.
ZHANG Kailiang, YANG Li, WANG Liangju, et al. Design and experiment of an overhead strawberry picking robot[J]. *Transactions of the Chinese Society for Agricultural Machinery*,

2012, 43(9): 165-172. (in Chinese with English abstract)

[25] DEBAPRIYA M J, SOYEB N, MANU M, et al. YOLO-Pose: enhancing YOLO for multi person pose estimation using object

key point similarity loss[J]. *Computer Vision and Pattern Recognition*, 2022, arXiv: 2204. 06806. Available: <https://arxiv.org/abs/2204.06806>

Recognizing strawberry to detect the key points for peduncle picking using improved YOLOv8 model

YANG Zhenyu¹, WANG Xiaochan^{1,2✉}, QI Zihan¹, WANG Dezhi¹

(1. College of Engineering, Nanjing Agricultural University, Nanjing 210031, China; 2. Jiangsu Province Engineering Lab for Modern Facility Agriculture Technology, Nanjing 210031, China)

Abstract: Robotic harvesting had been constrained by the low positioning accuracy of strawberry stem picking points and the significant challenge of identifying occluded strawberries. In this study, we proposed an improved YOLOv8 model combined with Pose key-point detection for enhanced strawberry recognition and localization. The accuracy of picking point localization was also improved, especially for occluded strawberries in complex environments. To optimize the YOLOv8 model, we introduced the Bidirectional Feature Pyramid Network (BiFPN) and the Generalized Attention Module (GAM), which enhanced bidirectional information flow, dynamically allocated feature weights, and focused on extracting features of small targets and enhancing the features of occluded regions. As a result, the model's ability to accurately detect and localize strawberries in complex environments was significantly improved. Experimental results showed that the improved YOLOv8-pose model outperformed the original model in several metrics: the Precision (P) increased by 6.01 percentage points, Recall (R) by 1.98 percentage points, mean Average Precision (mAP) by 6.67 percentage points, and mean Average Precision for key points (mAP_{kp}) by 7.85 percentage points. The positioning accuracy for strawberry stem picking points, based on key-point detection, achieved errors of just 1.4 mm in both the x and y directions and 2.2 mm in the z direction. Additionally, the occlusion level was classified according to the overlap area of occluded strawberries, and the model's performance under varying occlusion conditions was assessed. Under these conditions, the mAP_{kp} of the improved YOLOv8-pose model increased by 9.78 percentage points compared to the original model. Field trials further validated the model's effectiveness, with the strawberry-picking robot achieving a 95% success rate, picking each strawberry within 10 seconds. The high success rate and short picking time demonstrated the practicality of the model in real-world agricultural settings, indicating its high efficiency and accuracy. The improved YOLOv8 model with key-point detection accurately and robustly recognized strawberries, leveraged multi-scale features with the BiFPN architecture, and focused attention on relevant regions with the GAM, especially for occluded strawberries. These advancements significantly improved overall performance in precision, recall, and average precision, particularly under occlusion scenarios. In conclusion, these advanced techniques were integrated into a more capable strawberry-picking robot system. The enhanced accuracy and efficiency achieved in recognizing and localizing strawberries, even in challenging occlusion scenarios, highlighted the system's potential for practical agricultural applications. The findings contributed significantly to automated strawberry harvesting in agricultural robotics, paving the way for more efficient and cost-effective farming solutions in sustainable production.

Keywords: image recognition; strawberry; YOLOv8; picking; occlusion