doi:10.6041/j.issn.1000-1298.2026.12.005

基于改进 YOLO v8 - Pose 的串番茄采收果柄二次剪切点识别与定位

孙　腾[1]　王一帆[1,2]　何　梁[2]　单明宇[1,2]　刘佳希[1,2]　李亚军[2]

（1.上海大学机电工程与自动化学院，上海 200444；2.北京市农林科学院智能装备技术研究中心，北京 100097）

摘要：针对串番茄二次剪柄作业中存在的剪切点易被端部果实遮挡、定位精度不高和剪切易损伤果实的问题，本文提出基于改进 YOLO v8 - Pose 的解决方法：利用主动识别的思想，依据 4 分类 + 空间平面法向量的方法旋转果串以确定最佳观测角度；利用关键点信息定位二次剪切点。在网络结构方面，引入 PSA 部分自注意力机制以增强全局上下文建模能力；设计具有多尺度卷积核的 Dyn_GSConv 模块，并基于 Slim - Neck 结构构建 Dyn_VoVGSCSP 模块以替代原颈部的 C2f 结构，在降低模型复杂度的同时兼顾局部细节与全局特征；最后采用 VCM 模块替换特征提取层的标准卷积，进一步精简网络。试验结果表明，改进后网络在类别检测与关键点识别的 mAP50 均达到 97.4%，较原模型提升 1.4 个百分点；精确率分别为 93.3% 和 90.9%，提升 5.3、5.0 个百分点。同时，浮点运算量降低 9×10^8，参数量减少 1.9×10^5。在最佳观测视角选取方面，根据网络类别输出将随机姿态的果串调整至正面，并结合深度信息构建串番茄空间平面法向量，当法向量与相机 yz 面差角小于 10°，即确定最佳观测视角，为关键点识别和二次剪切操作提供清晰观测条件。本研究为实现串番茄自动化、高精度、低损伤的二次剪柄作业提供了有效的视觉感知解决方案。

关键词：串番茄；YOLO v8 - Pose；主动识别；空间平面法向量

中图分类号：TP242.6　　**文献标识码**：A　　**文章编号**：1000-1298(2026)12-0058-11　　**OSID**：

Active Recognition and Localization of Secondary Cutting Point for Cherry Tomato in Intelligent Harvesting Based on Improved YOLO v8 - Pose

SUN Teng[1]　WANG Yifan[1,2]　HE Liang[2]　SHAN Mingyu[1,2]　LIU Jiaxi[1,2]　LI Yajun[2]

(1. *School of Mechanical, Electrical and Automation Engineering, Shanghai University, Shanghai* 200444, *China*
2. *Intelligent Equipment Technology Research Center, Beijing Academy of Agriculture and Forestry Sciences, Beijing* 100097, *China*)

Abstract: Aiming to address issues in the secondary stem-cutting operation of cherry tomato, such as occlusion of cutting points by terminal fruits, low positioning accuracy, and risk of fruit damage, a solution based on an improved YOLO v8 - Pose was proposed. By introducing an active recognition strategy, the tomato was rotated to determine the optimal viewing angle according to a four-category classification and spatial plane normal vector method, while key points were used to locate secondary cutting positions. In terms of network architecture, the PSA partial self-attention mechanism was integrated to enhance global context modeling. A Dyn_GSConv module with multi-scale convolutional kernels was designed, and a Dyn_VoVGSCSP module was constructed based on a Slim - Neck structure to replace the original C2f neck, reducing model complexity while preserving local details and global features. Meanwhile, a VCM module was adopted to replace standard convolution in the feature extraction layer, further lightweighting the network. Experiments showed that the improved network achieved an mAP50 of 97.4% in both detection and keypoint recognition, an improvement of 1.4 percentage points over the original model. The precision rates reached 93.3% and 90.9%, increasing by 5.3 percentage points and 5.0 percentage points, respectively. Furthermore, GFLOPs were reduced by 9×10^8, and the number of parameters was decreased by 1.9×10^5. For optimal viewing angle selection, the network's

收稿日期：2025 - 12 - 02　修回日期：2026 - 01 - 08

基金项目：北京市农林科学院探索项目与创新能力建设专项(TSXM202514)

作者简介：孙腾(1989—)，男，助理教授，博士，主要从事机器人主动感知和农业机器人关键技术研究，E-mail：sunny315@shu.edu.cn

通信作者：李亚军(1995—)，男，助理研究员，博士，主要从事人工智能与农业机器人技术研究，E-mail：liyj@nercita.org.cn

category output was used to adjust randomly posed tomatoes to a front view, and depth information was employed to construct spatial plane normals of fruit clusters. When the angle between the normal vector and the camera's *yz*-plane was less than 10°, the optimal viewing angle was determined, providing a clear observation view for keypoint recognition and secondary cutting operations. The research result can provide an effective visual perception solution for achieving automated, high-precision, and low-damage secondary stem-cutting in cherry tomato production.

Key words: cherry tomato; YOLO v8 - Pose; active recognition; spatial plane normal vector

0　引言

串番茄凭借其口感优良、营养丰富/外观整齐和极高的耐贮运特性具有优秀的商品价值，广受消费者青睐[1]。目前，国内外不断提出并逐步普及了各式的串番茄采摘机器人[2]，其大多采用剪切果柄的方式完成采摘[3-4]。串番茄成串采收和销售，机器人采摘后的串番茄端部存在残余果柄，在装箱运输过程中易对其他串番茄果实造成损伤。传统人工采收流程需对串番茄多余果柄进行二次剪切，但人工剪柄操作导致大量资源和成本浪费，亟需引入一种基于机械臂的串番茄自动化二次剪柄方法。

串番茄的自动化二次剪柄操作，核心在于实现对二次剪切点的精准识别与定位。理想的二次剪切需满足商业化标准，将果柄修剪至长度小于等于5mm 并尽量贴近果实基部。然而，该任务面临多重挑战：一方面，串番茄果柄极为纤细，视觉辨识度低；另一方面，剪切点紧贴果实，加之果串在采摘后姿态不确定、生长形态多样，导致其极易被顶部果实遮挡，进一步增加了识别难度。为有效应对上述问题，需结合主动视觉的思想，寻找能够最大限度暴露果柄-果实连接区域的最佳观测视角，并借助高精度关键点检测算法，实现对剪切点的精准识别。一个合理的观测视角不仅能清晰呈现剪切点位置信息，也有助于简化机械臂运动轨迹规划；而识别准确性将直接决定自动化剪柄操作的完成度。

主动识别是一种任务驱动型的视觉探索方法，其通过不停改变观测视角，以最大化获取信息来解决目标任务[5-6]。LEHNERT 等[7]将末端执行工具上的成像传感器增加至 9 个，对比排成 3D 阵列的图像引导末端工具绕过遮挡区域，获得更清晰的果实视野，该方法需要处理大量图像数据和极高的硬件支持。SUN 等[8]在提取番茄和枝梗区域后，利用番茄区域的计算遮挡比例，指导机械臂运动选择新的视角重复识别直到视角满足剪切要求。主动识别解决了“看得到”的问题，而基于神经网络的关键点识别技术则解决了“看得准”的问题。刘莫尘等[9]基于 YOLO v8 - Pose 训练出一种有较强鲁棒性的红熟期草莓采摘点识别网络，该网络可在受光照、遮挡和角度的影响下直接准确输出采摘点位置，识别精度分别达 94.52%、95.48% 和 94.63%。RONG 等[10]使用语义分割算法识别番茄果实、花萼和茎后经过形态学后处理提取茎骨架的每个端点，根据番茄花萼的位置约束筛选出最佳采摘点，其采摘点识别算法精确率达 97.43%。

针对上述问题，本文提出一种基于改进 YOLO v8 - Pose 网络的串番茄二次剪切点识别方法。结合生产实际，选择串番茄正面（可见花穗的一面）作为最佳观测视角选取的先验条件：具体由改进的多分类 YOLO v8 - Pose 网络识别采摘完成后随机姿态的番茄果串；依据网络输出的类别信息决策机械臂第 7 关节运动直至达到果串正面；随后基于串番茄空间平面法向量确定正视视角，即最佳观测角度；最后依据网络输出的关键点信息定位二次剪切点和剪切执行器姿态。

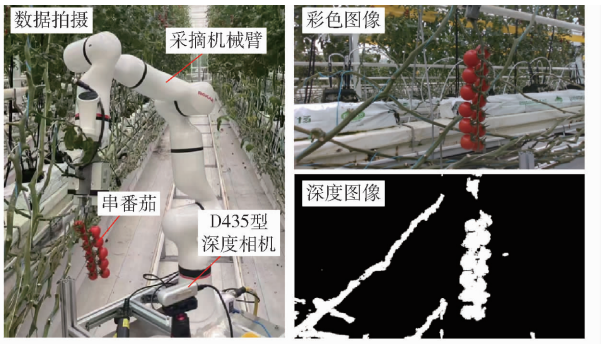
1　数据集构建

1.1　数据集采集

串番茄数据均采自首农集团北京海淀区翠湖温室，使用珞石 xMate ER3 Pro 型 7 自由度柔性协作机械臂夹持串番茄，选取 Intel RealSense D435 型深度相机以分辨率 1 280 像素 ×720 像素进行拍摄获取 RGB 彩色图像和深度图像。本数据集采集耗时 5 d，于 10:00—16:00 完成拍摄。所选用果串皆属于成熟红果，且保证每串果实数量大于 8 个。拍摄时，将机械臂置于温室大棚垄间，其末端夹持番茄果串（串番茄完成采摘后的状态），果串与相机距离保持在 40 ~ 60 cm 之间。设置机械臂前 6 个关节角度为 77°、0°、-73°、65°、2°、100°，这个姿态保证机械臂只需运动第 7 关节就可以带动串番茄自转，机械臂第 7 关节运动角度范围为 0° ~ 360°，每转动 45°拍摄 1 幅果串图像，每个果串只拍摄 8 幅图像。本数据集共 1 738 幅图像，按照比例 7:2:1将数据集随机划分为训练集、验证集和测试集。数据集采集如图 1 所示。

1.2　数据集标注

使用 X - AnyLabeling 标注工具进行标注。由于随机采摘完成的果串较相机有 4 种位置可能，所

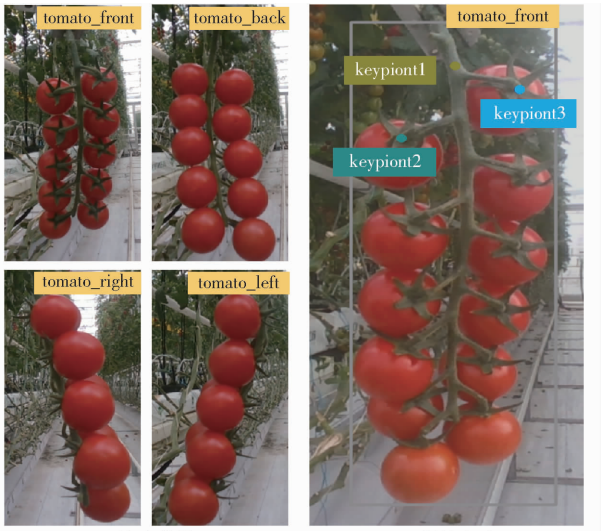


(a) 数据采集场景 (b) RGB彩色图像和深度图像

图 1 数据集采集示意图

Fig. 1 Schematics of data set collection

以将数据分成 4 个类别表示串番茄 4 个姿态，分别为 tomato_front（正面/果穗面）、tomato_left（左面）、tomato_right（右面）和 tomato_back（背面）。经统计，训练集与验证集标注的各类别样本分布为 377∶436∶500∶253，对应正面∶左面∶右面∶背面。关键点标注时，关键点 1 位于果柄基部与果实连接点上方约 0.1 cm 处；关键点 2 位于左侧端部第 1 个果实的果蒂处；关键点 3 位于右侧端部第 1 个果实的果蒂处。关键点 1 直接给出二次剪切的位置，关键点 2、3 连线的向量与相机坐标系的夹角为后续剪切操作的末端工具提供旋转角度。关键点 1 只标注完全可见情况，关键点 2、3 可见度超过 50% 时进行标注。串番茄姿态分类与标注方法如图 2 所示。



(a) 姿态分类 (b) 标注方法

图 2 串番茄姿态分类及标注方法

Fig. 2 Classification and labeling method of cherry tomatoes

2 串番茄姿态和关键点识别模块

2.1 YOLO v8 - Pose 网络结构

YOLO v8 - Pose 延续 YOLO 系列的端到端检测思路，但在模型结构、训练策略与推理效率等方面进行了显著改进[11-13]，其主要结构可以分为主干层、颈部和检测头 3 部分。主干网络部分主要负责特征提取，即通过一系列卷积操作降低输入的分辨率、增加通道数，同时模型加入了 C2f 模块、SPPF 模块以提升特征融合与表达能力。颈部网络结构采取 FPN + PAN 结构[14-15]，这样的双向特征融合特点实现了语义与细节的双向融合。检测头层采用解耦的方式将分类和回归分支分开[16]，从而提高了检测精度。此外，其将 anchor - based 检测框架改为 anchor - free 框架，提高了模型的鲁棒性与泛化能力[17]，改进传统 NMS 算法以减少误检和漏检，提出自动混合精度训练以加快训练速度同时减少显存占用量。

YOLO v8 - Pose 网络最初应用于人体关键点检测，对于串番茄适用度不高，在串番茄的关键点识别和状态预测上存在不够精准的问题，因此对原网络进行了一些改进：在主干网络最后一层引入 YOLO v10 的 PSA 注意力模块，增强了模型全局建模的能力；使用 Vision - Clue - Merge 模块替换特征提取的 Conv 模块，为模型训练保留了更多的视觉信息；提出 Dyn_GSConv 卷积，并根据 Slim - Neck 的结构提出 Dyn_VoVGSCSP 替换颈部的 C2f 模块，利用多尺度卷积核强化了模型对不同尺寸感受野的捕捉能力，动态权重分配实现了输入自适应的信息融合。改进后网络大幅提高了串番茄 4 个状态和 3 个关键点识别成功率和准确度，其结构如图 3 所示。

2.2 PSA 部分自注意力模块

凭借卓越的全局特征捕捉特性，自注意力机制已成为众多计算机视觉任务中的重要技术手段[18]，但其往往带来大量冗余计算量。PSA[19] 为了解决注意力头的冗余问题，采取双流结构，将特征通道分为 2 部分：一部分通道参与多头自注意力（MHSA）和前馈网络（FFN）的计算，另一部分保留原始特征并最终与前者融合。其中 MHSA 采用自适应注意力头设计，在保证每个头有 64 维的基础上计算数量，在本文模型中 MHSA 有 4 个注意力头；FFN 采用 2 层瓶颈结构：第 1 层将通道数扩展 2 倍，第 2 层又压缩回原始维度，以增强网络的非线性能力。为进一步降低计算负担，MHSA 采用以下优化策略：①缩减查询/键的维度至值维度的一半。②用批归一化（BatchNorm）替代层归一化（LayerNorm）以提升推理速度。这种设计以极小的计算开销为模型赋予了显著的全局上下文建模能力，有效提升了模型精度。为防止自注意力二次计算的庞大数据量增长，仅在特征图分辨率最低层（P5 层）使用该模块，具体添加在 SPPF 模块后。PSA 结构如图 4 所示，其中 h 和 w 分别表示特征图在空间维度上的高度、宽度，c 表示特征通道数。

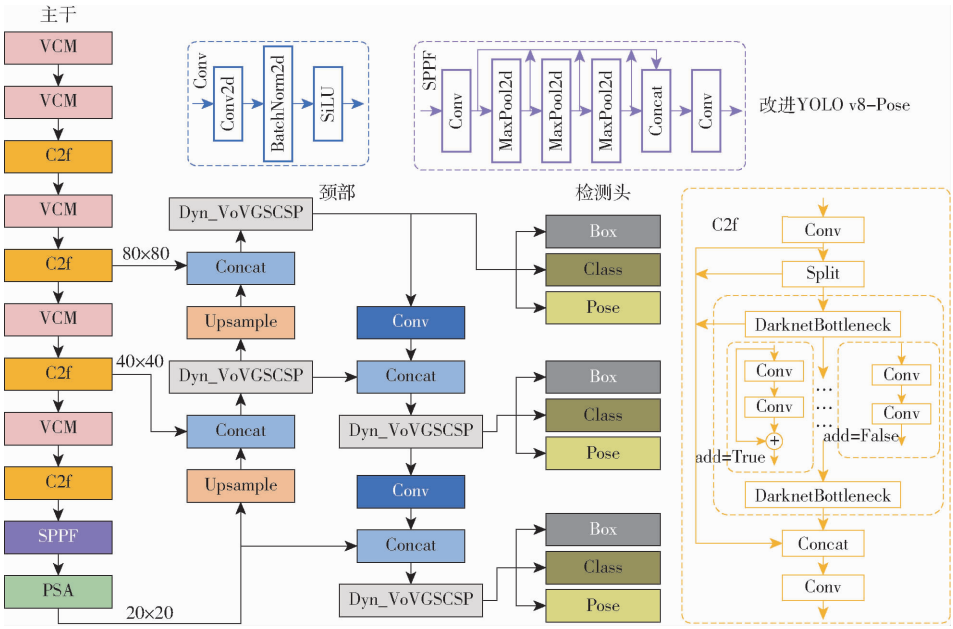


图 3　改进 YOLO v8－Pose 网络结构

Fig. 3　Network structure of improved YOLO v8－Pose

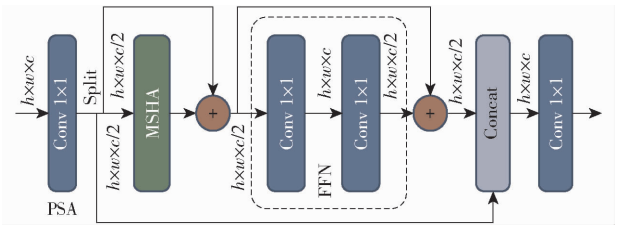


图 4　PSA 模块结构

Fig. 4　Module structure of PSA

特征图通过 SPPF 模块进行多尺度池化以提取多层级特征，随后利用 PSA 模块捕获跨通道的全局依赖关系[20]，从而挖掘特征中的长距离关联信息，增强了不同区域间的特征交互与融合能力，有效提升了网络对串番茄图像的深层感知能力。

2.3　Dyn_GSConv 和 Dyn_VoVGSCSP

GSConv[21]是一个结合普通卷积 SC 和深度可分离卷积 DSC 的轻量化卷积模型，其综合了 GhostNet 和 ShuffleNETv2 的架构思想，使用少量标准卷积核生成初始特征图，再将包含大量相似信息的初始特征图输入到低成本的深度可分离卷积扩展特征数量，将初始特征图和经过深度可分离卷积的次级特征图进行拼接，以增加特征多样性；最后将拼接后的特征图进行随机通道顺序操作，打破固定通道位置带来的信息壁垒，强制不同通道特征在后续层中交叉融合，以此提升全局表征能力[22]。GSConv 的计算成本约为 SC 的 50%，但对模型学习能力的贡献与 SC 相当。GSConv 结构如图 5 灰色背景部分所示。

GSConv 具有强大的降低计算复杂程度和推理时间的能力，但仍采用固定形状的卷积核，只能获取固定的局部感受野[23]，而串番茄关键点识别和最佳视角选择问题依赖对果串整体和局部细节不同感受野的多方面把握，所以引入 Inception[24]模块的多尺度特征并行提取与融合思想，在 GSConv 的深度可分离卷积阶段，将原始单一尺度的深度可分离卷积核替换为 3 个不同尺寸深度可分离卷积核（3×3、5×5、7×7）的并行结构，以增强网络对不同尺度感受野的适应性。3 个深度可分离卷积分支的输入通道数与原 GSConv 保持一致，均设为最终输出通道数的一半，即 $c_2/2$。

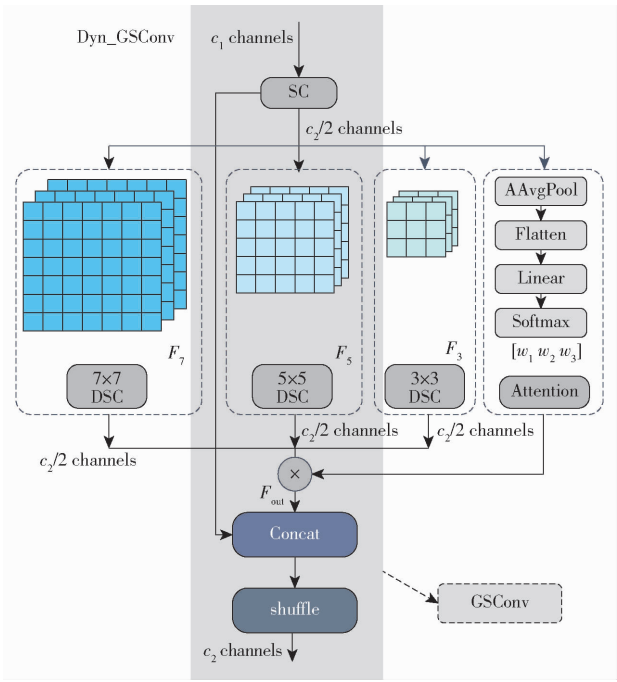


图 5　GSConv 和 Dyn_GSConv 模块结构

Fig. 5　Module structure of GSConv and Dyn_GSConv

与传统 Inception 模块不同，额外引入 1 个并行 Attention 模块（池化层 + 展平层 + 全连接层 + 归一化层）用于自适应地融合不同大小卷积核的特征，具体来说：使用 1 个自适应平均池化将空间维度压缩到 1 × 1，再使用以 Flatten 将 4D 张量转换为 2D 矩阵，之后使用 1 个全连接层学习 3 个卷积核的权重分配，最后进行归一化操作。Attention 模块接受与 3 个深度可分离卷积相同的输入通道数，输出 3 维的权重[w_1 w_2 w_3]。给出了 3 个特征图在最终融合中的相对重要性，权重与 3 个特征图加权求和得到输出，公式为

$$F_{out} = w_1 F_3 + w_2 F_5 + w_3 F_7 \tag{1}$$

式中 F_{out}——加权求和结果

F_3、F_5、F_7——3 个不同尺寸卷积核特征图输出

Attention 模块为 3 个不同尺寸卷积核输出自适应提供了权重，让网络自主地、有针对性地获取最优的感受野信息。该机制能有效抑制由非目标植株枝叶及温室行间固定设施（如灌溉管道、支撑架等）的复杂背景干扰，从而增强网络对目标果串区域的聚焦与识别能力，同时增强了关键点定位精度，显著提升了网络鲁棒性。改进的 GSConv 卷积重新命名为 Dyn_GSConv，结构如图 5 所示（虚线包围的整体）。

VoVGSCSP 是 Slim - Neck 中基于 GSConv 设计的跨层级单次聚合模块[25]。在此基础上，本文将原 GSConv 替换为具有多尺度卷积核的 Dyn_GSConv，构建出 Dyn_VoVGSCSP 模块，其结构如图 6 所示。YOLO v8 - Pose 原颈部所使用的 C2f 模块通过融合不同分支的梯度流增强特征表示能力，但结构相对复杂、计算成本较高。以 Dyn_VoVGSCSP 模块替换 C2f 模块，在显著降低计算复杂度和提升推理速度的同时，使网络能够融合更多尺度的视觉信息，从而增强对串番茄整体结构与局部关键点细节特征感知能力。

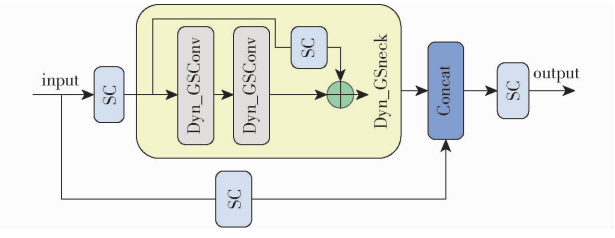


图 6 Dyn_VoVGSCSP 模块结构

Fig. 6 Module structure of Dyn_VoVGSCSP

2.4 Vision Clue Merge 模块

VCM[26-27]模块摒弃了传统 CNN 和 ViTs 一直使用卷积进行下采样的惯例，而采用像素重排和 1 × 1 的卷积特征融合方式实现空间降采样。具体而言，输入特征图先经过像素重排操作，将空间信息重新排列为多通道特征表示，再通过 1 × 1 卷积、BatchNorm 和 SiLU 激活函数进行特征压缩与融合。同使用滑动窗口的传统卷积操作相比，VCM 模块避免了因卷积核滑动窗口所造成的信息丢失，能够更完整地保留结构信息，为后续网络提供更丰富的视觉线索。此外，VCM 仅使用 1 × 1 卷积，可大幅减少参数量，提高模型效率。本研究使用 VCM 模块替换特征提取层的 Conv 模块，可保证串番茄图像在降低分辨率同时保留其空间结构信息，对串番茄局部关键点信息更加敏感。其结构如图 7 所示。

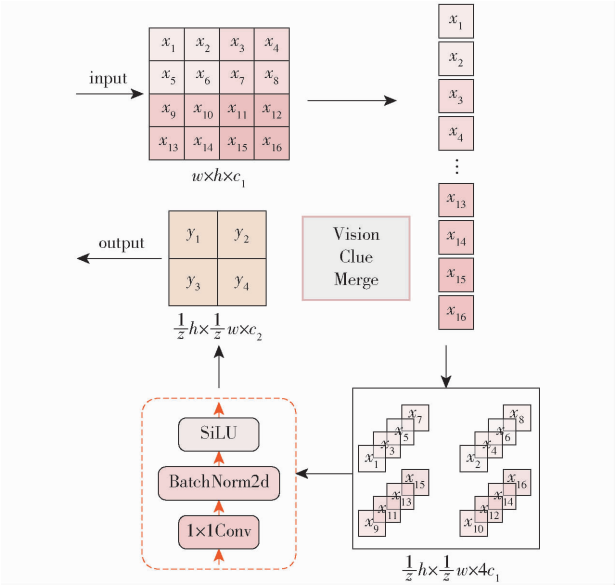


图 7 VCM 模块结构

Fig. 7 Module structure of VCM

3 串番茄最佳观测视角确定

3.1 果穗面确定

使用串番茄果穗面作为寻找最佳观测视角的先验条件，既降低最佳视角界定的复杂度，又符合人工剪柄的传统角度和串番茄一般生长特性，同时以正面角度完成二次剪柄工作便于后续串番茄收纳装框操作，是符合实际生产需要和机械化自动化作业流程的有效策略。

改进 YOLO v8 - Pose 网络输出串番茄姿态类别和关键点位置，依据其类别输出自适应选择末端旋转方向，可使采摘后姿态各异串番茄快速到达果穗面。对于果穗朝左的 tomato_left 类串番茄，控制机械臂末端逆时针旋转；同理当串番茄果穗朝右，则顺时针旋转；对于背面情况则规定末端顺时针转动。直到网络的类别输出结果变为 tomato_front，此时即到达果穗面，需依靠果串空间平面法向量具体确定最佳观测角度。

3.2　基于空间平面法向量最佳角度确定

串番茄跟随末端关节的转动而改变姿态，当神经网络输出为 tomato_front 时，因其生长姿态随机，当前视角下图像仍可能会出现二次剪切点被遮挡、剪切运动复杂、易损伤果实的问题。而对于类别和关键点置信度，其仅能反应识别结果可靠性，与果串最佳图像采集姿态并没有直接关系。所以仅依靠神经网络难以精准确定串番茄最佳的视角，需继续运动机械臂末端第 7 关节调整果串姿态。由于串番茄果穗呈鱼骨状且果实呈 2 列排列在果梗两侧[28]，其正视于相机的视角可使二次剪切点完全暴露且便于后续机械臂剪切的运动轨迹规划。

在 2D 图像基础上，引入深度信息，提出一种基于空间平面法向量的方法寻找串番茄正视相机视角：根据网络输出的 Box 框位置初步筛选出串番茄的深度信息；使用改进 SOR 统计离群点滤波算法剔除噪声，得到没有异常深度点的、干净完整的串番茄深度值；应用 RANSAC 随机抽样一致算法对串番茄整体进行空间平面拟合操作并计算其平面法向量；依据法向量与相机坐标系 yz 平面的夹角自适应控制末端的转动，直至夹角 θ 小于误差，即确定最佳观测视角。

在点云数据预处理中，为有效过滤异常深度，改进了传统统计离群点滤波（SOR）算法。先通过全局深度一致性检测，剔除深度值超出预设阈值的离群点；再执行传统 SOR 滤波以进一步去除边界区域的离群深度点。为尽可能保留有效点云数据，在 SOR 阶段将邻居数量和标准差比例参数设置为较大值，从而在滤除噪声的同时使点云边界过渡更为平滑。

图 8 为使用 RANSAC 算法拟合平面示意图，左侧为经过 Box 框和改进 SOR 算法处理的串番茄正面点云图，右侧为其侧视图。依据算法，随机取 3 个不共线的点云生成平面，根据剩余点云至平面的距离将其分为内点和外点，距离 d 小于阈值 σ 的点为内点，计算内外点比值，循环迭代至比值达到预设标准，确定为串番茄空间平面，计算其法向量为寻找最佳观测视角奠基。

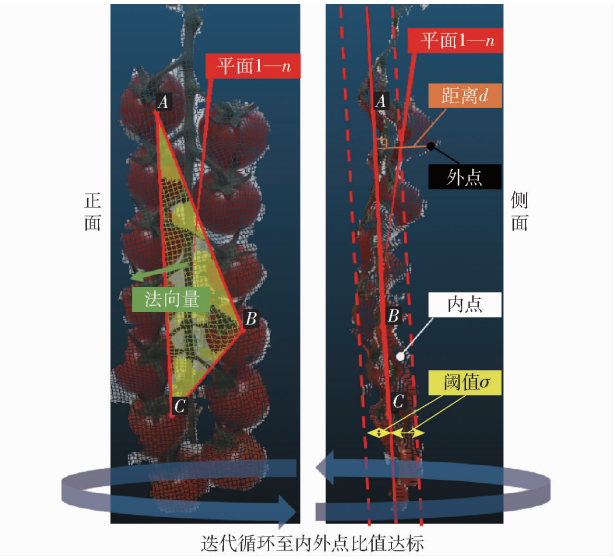


图 8　串番茄空间平面拟合

Fig. 8　Spatial plane fitting of cherry tomato

串番茄属于柔性物体，且多数采摘采用机械臂末端垂直果柄方式进行，即使调整机械臂姿态也无法使末端与果柄成 180°，所以末端旋转角度难以等同于串番茄果柄实际旋转角度。在确保视觉观测效果的同时，通过将 θ 设定在合理范围内，可提高机械臂运动的容错性，避免了因过度微调造成的时间消耗，缩短调姿时间，提升了整体作业效率。

4　试验

4.1　试验环境

模型训练所用计算机硬件配置为：处理器为 Intel（R）Core i5－12600KF，内存为 32 GB，显卡为 NVIDIA GeForce RTX 4070S，采用 Windows 11（64 位）操作系统。训练的软件环境如下：Conda 版本为 24.9，Python 版本为 3.9.21，Pytorch 版本为 2.5.1，采用 CUDA 12.4 和 cuDNN 9.1.0。训练时，将图像尺寸压缩至 640 像素 ×640 像素，并使用随机梯度下降算法（SGD）作为优化器训练和更新网络参数，训练轮次 300，批次大小为 16。训练时开启 YOLO 自带的数据增强，以提升模型鲁棒性和泛化能力。

最佳视角试验所用边缘设备为：处理器为 AMD Ryaen7 3750H，内存为 16 GB，显卡为 NVIDIA GeForce RTX 1660Ti，采用 Windows 11（64 位）操作系统。软件环境与上述网络训练环境一致。

4.2　评价指标

目标关键点相似度（OKS）是衡量预测关键点与真实关键点接近程度的评价指标，类似目标检测中的交并比（IoU）；平均精度均值（mAP50）为 OKS 大于等于 0.5 的情况下，所有预测结果的平均精度（AP）均值。

同时关注检测和关键点 2 项任务，以 mAP50、精确率（Precision）、浮点运算量和模型参数量作为评价指标。mAP50 和精确率可反映网络对串番茄姿态和关键点位置估计的准确程度，而浮点运算量和模型参数量可以体现模型复杂程度和大小。由于串番茄在自转过程中，不同的面可能造成关键点遮挡，对训练结果产生消极影响，因此重点考察 tomato_front（果穗面）的关键点识别结果。

4.3　消融试验

为验证各模块对网络最终优化效果，分别将

3 个模块添加进模型进行消融试验,消融试验结果如表 1 所示(其中Ⅰ表示 PSA 部分自注意力模块、Ⅱ表示 Dyn_VoVGSCSP 模块、Ⅲ表示 VCM 模块)。改进的网络在检测、关键点和轻量化上较原始 YOLO v8 - Pose 网络都有不错提升:对于识别的准确度,改进的网络在检测(Detect)和关键点(Pose)平均精度均值都达到 97.4%,精确率分别达到 93.3% 和 90.9%,较原始模型分别提升 1.4、5.3、5.0 个百分比;对于轻量化,网络的优化使浮点运算量降低 9×10^8,参数量减少 1.9×10^5,在保证性能的同时显著降低了计算开销和存储成本。

表 1 消融试验结果

Tab.1 Ablation test results

模型	平均精度均值/%		精确率/%		参数量	浮点运算量
	Detect	Pose	Detect	Pose		
YOLO v8 - Pose	96.0	96.0	88.0	85.9	3.08×10^6	8.3×10^9
YOLO v8 - Pose + Ⅰ	96.6	96.6	90.1	89.1	3.33×10^6	8.5×10^9
YOLO v8 - Pose + Ⅰ + Ⅱ	97.2	97.2	91.7	87.6	3.11×10^6	7.7×10^9
YOLO v8 - Pose + Ⅰ + Ⅱ + Ⅲ	97.4	97.4	93.3	90.9	2.89×10^6	7.4×10^9

由表 1 可知,引入 PSA 自注意力模块虽增加了一定参数量和浮点运算量,但平均精度均值、检测精确率和关键点精确率分别提升 0.6、2.1、3.2 个百分点。Dyn_VoVGSCSP 模块通过多尺度深度卷积有效提升了网络性能,在平均精度均值和检测精确率分别提高 0.6、1.6 个百分点的同时,显著降低了计算复杂度,参数量和浮点运算量分别减少 2.2×10^5、8×10^8。然而,其关键点精确率较 PSA 略有下降,可能由于多尺度分支通过注意力权重进行融合,使结果更偏向于整体结构特征,对关键点的细小目标的响应不足,从而影响关键点定位精度。相比之下,VCM 模块对平均精度均值提升有限,可能由于像素重排更多作用于中低层特征信息,对于高层语义理解不深。但其在 Dyn_VoVGSCSP 基础上进一步提高了检测与关键点精确率(1.6、3.3 个百分点),并继续降低网络复杂度。综合来看,3 种模块协同使用在保证模型轻量化的同时,有效提升了串番茄状态与关键点识别性能。

特别的,由于网络训练采用高性能显卡,其对于帧率敏感度较低,所以消融试验只使用最终改进网络在边缘设备进行帧率试验,帧率为 125 f/s,满足机器人二次剪柄的实时性要求。

4.4 对比试验

为进一步验证算法性能,在相同训练环境和参数配置下对多个 YOLO 关键点检测网络进行对比,对比试验结果如表 2 所示。

表 2 对比试验结果

Tab.2 Comparative test results

模型	平均精度均值/%		精确率/%		参数量	浮点运算量
	Detect	Pose	Detect	Pose		
YOLO v8s - Pose	95.5	94.3	92.1	88.1	1.141×10^7	2.94×10^{10}
YOLO 11n - Pose	96.5	95.7	90.9	85.5	2.650×10^6	6.60×10^9
YOLO 11s - Pose	96.4	94.8	93.1	89.1	9.700×10^6	2.23×10^{10}
YOLO 12n - Pose	96.3	95.3	90.7	89.2	2.630×10^6	6.60×10^9
本文算法	97.4	97.4	93.3	90.9	2.890×10^6	7.40×10^9

由表 2 可知,本文算法在预测精度方面展现出显著优势,其平均精度均值(97.4%)和精确率(93.3%、90.9%),在所有对比模型中均为最高。验证了本文在特征融合与关键点表征设计上改进的有效性,能够更精准地识别类别、定位关键点。在模型效率方面,本文算法相较于 YOLO 11n - Pose 和 YOLO 12n - Pose,参数量分别增加 2.4×10^5、2.6×10^5,浮点运算量均增加 8×10^8,但仍为同一数量级。综合来看,本文算法在精度-效率两方面取得了良好平衡,为串番茄自动化二次剪柄提供了精度和实时双保障。

在实际识别测试中,上述网络均对串番茄状态做出了正确预测,测试结果如图 9 所示。在关键点识别中,除本文算法和 YOLO 11n - Pose 网络外,其余算法均出现关键点 1 脱离主茎的问题,关键点 1 作为二次剪切点,其脱离主茎会导致剪切失败。此外 YOLO 11n - Pose 和 YOLO 12n - Pose 网络还出现关键点 1 位置过于靠上(易导致二次剪切后所剩端部多余果柄较多)的问题。本文算法与 4 个其余算法相比,虽模型轻量化不及 YOLO 11n - Pose 和



图 9　不同模型识别结果

Fig. 9　Recognition results of different models

YOLO 12n－Pose 算法，但实际测试的关键点检测准确度更高，更适合实际生产需求。

4.5　基于空间平面法向量最佳视角试验

当算法输出为果穗面时，仍不适宜完成二次剪柄（图 10）。由图 10 可知，即使果串 tomato_front 的类别置信度高达 0.97、0.95，实际因串番茄随机生长姿态，依然会出现二次剪切点被遮挡或当前姿态剪切易损伤果实的情况，证明了本文基于空间平面法向量的最佳视角判定方法的必要性。

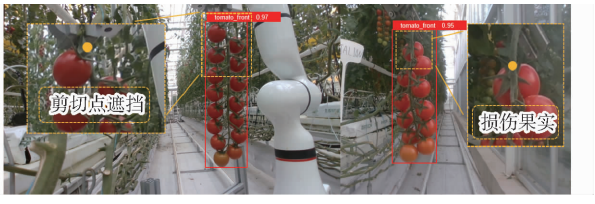


图 10　果穗面下不适宜姿态

Fig. 10　Inappropriate posture under tomato_front

经过对所采集数据集统计分析，将 RANSAC 的内外点最低比例设置为 70%。当比例过高时，多数果串无法拟合出有效平面；当比例过低时，拟合平面的随机性将被放大，导致算法的不确定性。

在内外点比例为 70% 的基础上，测试 RANSAC 对同一深度图像拟合平面的随机性误差。取 10 幅 tomato_front 类别串番茄深度图像，对每幅图像进行 5 次独立的平面拟合并计算空间法向量；计算这 5 个法向量平均值，定义为“标准法向量”；计算 5 个法向量与标准法向量偏差，并取其最大偏差，其结果如表 3 所示。

由表 3 可知，多数情况下，平面法向量绝对值越大，算法最大偏差越大。10 组样本最大偏差平均值为 2.22°。为减轻RANSAC算法的随机性误差对试验的影响，并基于对数据集的分析（当夹角 $\theta \leqslant 10°$ 时二次剪切点特征清晰），后续试验将夹角阈值 θ 设定为 10°。

表 3　RANSAC 算法随机性误差试验

Tab. 3　RANSAC algorithm random error experiment

（°）

果串序号	标准法向量	最大偏差
1	－12.3	1.5
2	－7.6	0.8
3	－17.5	3.1
4	－6.0	2.4
5	－11.8	3.2
6	15.6	4.1
7	10.7	2.7
8	9.8	1.5
9	14.3	1.9
10	8.8	1.0

为验证空间平面法向量方法的有效性，取 50 个果串进行试验，分别以顺逆时针接近正视视角，共试验 100 次。其中末端以 TCP 线速度 100 mm/s、空间旋转速度 20°/s 转动，试验均在果串为 tomato_front 类别下开始。

试验结果如图 11 所示，当选取合适的内外点比例确定果串空间平面时，法向量与 yz 平面的夹角均可落于 ±10°之间，且保证二次剪切点的清晰可见。从折线图并无法得出 RANSAC 的内外点比例与最佳角度的直接关系。从算法理论角度分析，初始角度和最佳角度的符号应保持一致，但串番茄的果柄多朝向果穗面弯曲，在夹持状态下果串与机械臂末端无法保持共线，导致其在转动过程中法向量的变化存在不连续性。如图 11 中存

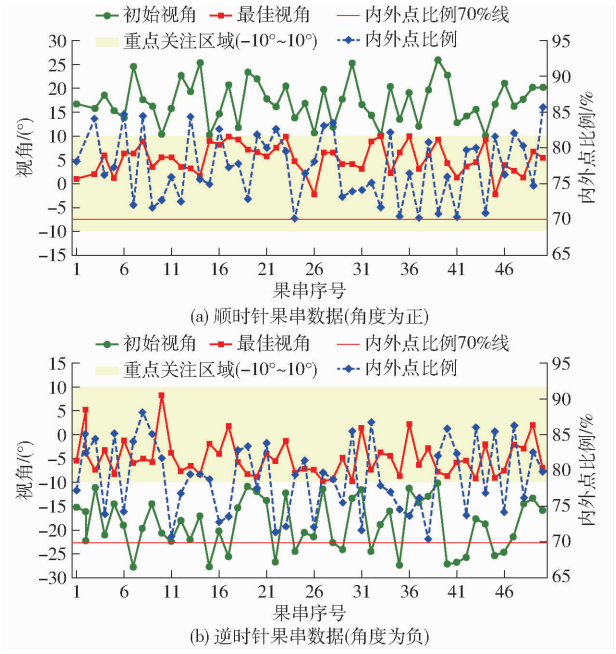


图 11 最佳观测视角试验结果

Fig. 11 Optimal observation perspective test results

在少许果串的最佳视角超过 0°,初始角度和最佳角度位于标准向量的两侧,出现了较大的角度跳变现象。

图 12 为本文算法在果串 1、2 上的完整算法流程结果,包括改进 YOLO v8 - Pose 网络识别结果和最佳视角选取示意。图 12a 为改进网络的识别检测结果;图 12b 为检测框和改进 SOR 算法过滤后的串番茄深度信息,蓝色点云是深度一致性过滤的异常深度点,红色点云是传统 SOR 算法过滤的边界离群点云;图 12c 为空间平面的拟合和最终的可视化结果,淡黄色平面即为 RANSAC 算法拟合的空间平面,绿色箭头为法向量,将法向量标注在二次剪切点位置,其底部显示了法向量与 yz 平面的夹角。2 个果串分别以顺时针和逆时针方向到达最佳视角,第 1 行为初始状态,第 2 行为最终最佳观测角。由图 12b 可知,改进 SOR 算法不仅过滤了检测框内不属于串番茄的点云,也去除了果串边缘的离群点云。图 12 完整展示了改进 YOLO 网络识别结果和基于空间平面法向量的最佳视角确定流程。

需要指出的是,试验对象具有特定性——果柄弯曲度小。对于弯曲度大的果串,RANSAC 算法会因内点不足而无法在限次迭代内收敛,致使平面拟合失败,并最终阻碍后续算法的推进;而为了拟合出

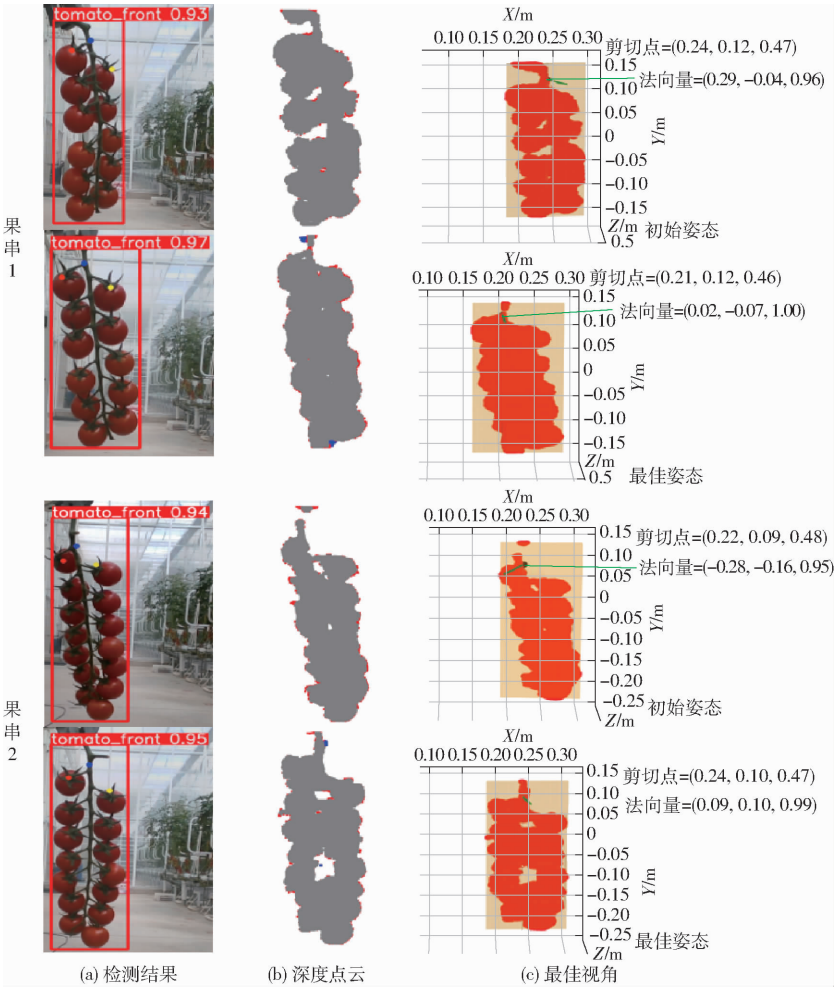


图 12 本文算法完整流程结果示意图

Fig. 12 Diagrams of complete algorithm process result

空间平面过度降低内外点比例将导致本算法的不准确性。如图 13 所示，此时 RANSAC 内外点比例仅为 55.61%，但其差角仅为 3.905°，仅由彩色图就可判断当前姿态剪切易损伤果实，并非为最佳观测视角。

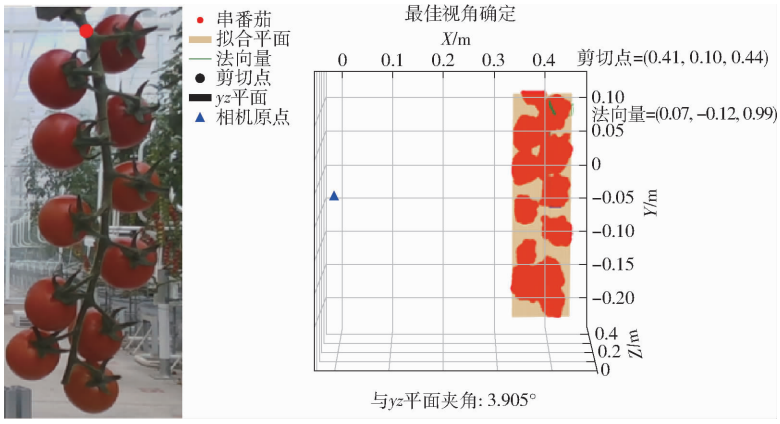


图 13　算法失效情况示意

Fig. 13　Diagram of algorithm failure scenarios

5　结论

（1）根据串番茄二次剪柄的实际需求改进 YOLO v8－Pose 网络：引入 PSA 部分自注意力机制强化网络学习能力，提高识别精度；在 GSConv 的基础上增加多尺寸深度卷积核并自适应分配权重，形成 Dyn_GSConv，依据 Slim－Neck 结构提出 Dyn_VoVGSCSP 模块，轻量化网络的同时增强视野感受能力；应用 VCM 模块降低网络复杂度。改进的网络在检测和关键点 mAP50 都达到 97.4%，较原模型提升 1.4 个百分点；精确率分别为 93.3% 和 90.9%，提升 5.3、5.0 个百分点。同时，浮点运算量降低 9×10^8、参数量减少 1.9×10^5，实现精度和轻量化双提升。这些改进使网络可以准确提供关键点位置，同时正确预测串番茄状态，为寻找最佳视角做基础。

（2）基于空间平面法向量的串番茄最佳视角判定方法经过试验证明了其的必要性和有效性。对于大多数番茄果串，依靠平面法向量可以是二次剪切点完整可见且姿态适宜切剪，满足自动化剪柄实际需求。

参　考　文　献

[1] 李静. 串收樱桃番茄新品种选育[D]. 武汉：华中农业大学，2024.
LI Jing. Breeding of new varieties of truss cherry tomato[D]. Wuhan: Huazhong Agricultural University, 2024. (in Chinese)

[2] 窦汉杰，陈震宇，翟长远，等. 果园智能化作业装备自主导航技术研究进展[J]. 农业机械学报，2024，55(4)：1－22.
DOU Hanjie, CHEN Zhenyu, ZHAI Changyuan, et al. Research progress on autonomous navigation technology for orchard intelligent equipment[J]. Transactions of the Chinese Society for Agricultural Machinery, 2024, 55(4): 1－22. (in Chinese)

[3] JUN J, KIM J, SEOL J, et al. Towards an efficient tomato harvesting robot: 3D perception, manipulation, and end-effector [J]. IEEE Access, 2021, 9: 17631－17640.

[4] 赵博，柳苏纯，张巍朋，等. 面向樱桃番茄采摘识别的轻量化 Transformer 架构优化研究[J]. 农业机械学报，2024，55(10)：62－71，105.
ZHAO Bo, LIU Suchun, ZHANG Weipeng, et al. Performance optimization of lightweight Transformer architecture for cherry tomato picking[J]. Transactions of the Chinese Society for Agricultural Machinery, 2024, 55(10): 62－71, 105. (in Chinese)

[5] FOIX S, ALENYÀ G, TORRAS C. Task-driven active sensing framework applied to leaf probing[J]. Computers and Electronics in Agriculture, 2018, 147: 166－175.

[6] 苗中华，周子斐，孙腾，等. 基于强化学习的苹果采摘机器人遮挡抑制与机械臂路径规划方法[J]. 农业机械学报，2026，57(5)：95－104.
MIAO Zhonghua, ZHOU Zifei, SUN Teng, et al. Occlusion suppression and robotic arm path planning for apple-picking robots based on reinforcement learning[J]. Transactions of the Chinese Society for Agricultural Machinery, 2026, 57(5): 95－104. (in Chinese)

[7] LEHNERT C, TSAI D, ERIKSSON A, et al. 3D move to see: multi-perspective visual servoing towards the next best view within unstructured and occluded environments[C]// 2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). IEEE, 2019: 3890－3897.

[8] SUN T, ZHANG W, MIAO Z, et al. Object localization methodology in occluded agricultural environments through deep learning and active sensing[J]. Computers and Electronics in Agriculture, 2023, 212: 108141.

[9] 刘莫尘，褚镇源，崔明诗. 基于改进 YOLO v8－Pose 的红熟期草莓识别和果柄检测[J]. 农业机械学报，2023，54(增刊

2):244 - 251.
LIU Mochen, CHU Zhenyuan, CUI Mingshi. Red ripe strawberry recognition and stem detection based on improved YOLO v8 - Pose[J]. Transactions of the Chinese Society for Agricultural Machinery, 2023, 54(Supp.2):244 - 251. (in Chinese)
[10] RONG Q, HU C, HU X, et al. Picking point recognition for ripe tomatoes using semantic segmentation and morphological processing[J]. Computers and Electronics in Agriculture, 2023, 210: 107923.
[11] SOHAN M, SAI T, RAMI C V. A review on YOLO v8 and its advancements[C]// International Conference on Data Intelligence and Cognitive Informatics, 2024: 529 - 545.
[12] VARGHESE R, SAMBATH M. YOLOv8: a novel object detection algorithm with enhanced performance and robustness[C]// 2024 International Conference on Advances in Data Engineering and Intelligent Computing Systems (ADICS). IEEE, 2024: 1 - 6.
[13] HUSSAIN M. YOLO v5, YOLO v8 and YOLO v10: the go-to detectors for real-time vision[J]. arXiv Preprint arXiv:2407.02988, 2024.
[14] 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, 2017: 2117 - 2125.
[15] 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, 2018: 8759 - 8768.
[16] CHEN F, DENG M, GAO H, et al. NHD - YOLO: improved YOLO v8 using optimized neck and head for product surface defect detection with data augmentation[J]. IET Image Processing, 2024, 18(7): 1915 - 1926.
[17] ZHANG Y, ZHANG H, HUANG Q, et al. DsP - YOLO: an anchor-free network with DsPAN for small object detection of multiscale defects[J]. Expert Systems with Applications, 2024, 241: 122669.
[18] LIU Z, LIN Y, CAO Y, et al. Swin transformer: hierarchical vision transformer using shifted windows[C]// Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021: 10012 - 10022.
[19] WANG A, CHEN H, LIU L, et al. Yolov10: real-time end-to-end object detection[J]. Advances in Neural Information Processing Systems, 2024, 37: 107984 - 108011.
[20] ZHOU Y, ZHAO Z. MPA - YOLO: steel surface defect detection based on improved YOLO v8 framework[J]. Pattern Recognition, 2025,168: 111897.
[21] LI H, LI J, WEI H, et al. Slim-neck by GSConv: a better design paradigm of detector architectures for autonomous vehicles[J]. arXiv Preprint arXiv:2206.02424, 2022.
[22] ZHAO X, SONG Y. Improved ship detection with YOLOv8 enhanced with MobileViT and GSConv[J]. Electronics, 2023, 12(22): 4666.
[23] 喻陈楠,伍永红,周杰,等. 基于改进 YOLO v8n 的非结构环境下杭白菊检测方法[J]. 农业机械学报,2025,56(5):405 - 414.
YU Chennan, WU Yonghong, ZHOU Jie. Improved YOLO v8n for detection of Hangzhou white chrysanthemum in unstructured environments[J]. Transactions of the Chinese Society for Agricultural Machinery, 2025,56(5):405 - 414. (in Chinese)
[24] SZEGEDY C, LIU W, JIA Y, et al. Going deeper with convolutions[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015: 1 - 9.
[25] LIN B. Safety helmet detection based on improved YOLOv8[J]. IEEE Access, 2024, 12: 28260 - 28272.
[26] WANG Z, LI C, XU H, et al. Mamba yolo: a simple baseline for object detection with state space model[C]// Proceedings of the AAAI Conference on Artificial Intelligence, 2025: 8205 - 8213.
[27] SUN S, DENG M, YU X, et al. Self-adaptive gamma context-aware ssm-based model for metal defect detection[J]. arXiv Preprint arXiv:2503.01234, 2025.
[28] 吴志刚,冯辉,徐娜,等. 串番茄品种的特征特性和应用状况[J]. 长江蔬菜,2006(2):36 - 38.