doi:10.6041/j.issn.1000-1298.2026.05.008

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

苗中华　周子斐　孙　腾　高　旋　李　楠　张　伟
（上海大学机电工程与自动化学院，上海 200444）

摘要：针对复杂农业场景中果树枝叶遮挡导致采摘机器人定位精度下降与采摘困难的问题，本研究提出一种基于强化学习的苹果采摘机器人遮挡抑制与路径规划方法。通过构建视觉-运动协同的主动感知系统，在视觉感知层基于 YOLO v8 模型设计动态遮挡抑制计算模型（OSCM），融合颜色直方图区域分割与边界拟合估计算法，实时生成果实遮挡率及遮挡抑制热力图；在运动决策层，将深度相机固联于机械臂末端，建立三维遮挡抑制矩阵映射机制，将相机至目标的深度空间离散化为三维栅格矩阵，每个栅格包含 OSCM 输出的遮挡率与抑制优先级；在路径规划层，提出了一种深度强化学习驱动的机械臂避障决策模型，通过融合遮挡率惩罚与采摘时效性的复合奖励函数，引导机械臂在三维栅格中自主搜索最优抑制路径。实验结果表明，系统在枝叶遮挡率不小于 60% 的高遮挡场景下，果实遮挡面积平均下降 33%，采摘成功率从 67% 提升至 94.7%，且较启发式方法平均单果耗时降低 3.2 s。本研究将遮挡抑制机制与强化学习决策闭环融合，为农业机器人动态环境适应性研究提供了方法。

关键词：苹果采摘机器人；强化学习；遮挡抑制；机械臂；路径规划
中图分类号：TP242.2　　**文献标识码**：A　　**文章编号**：1000-1298(2026)05-0095-10　　**OSID**：

Occlusion Suppression and Robotic Arm Path Planning for Apple-picking Robots Based on Reinforcement Learning

MIAO Zhonghua　ZHOU Zifei　SUN Teng　GAO Xuan　LI Nan　ZHANG Wei
(*School of Mechatronic Engineering and Automation*, *Shanghai University*, *Shanghai* 200444, *China*)

Abstract: Aiming to address the dual challenges of declining localization accuracy and harvesting difficulties caused by fruit branch occlusion in complex agricultural environments, a reinforcement learning-based occlusion suppression and path planning method for apple picking robots was proposed. By constructing a vision-motion collaborative active perception system, a dynamic occlusion suppression calculation model (OSCM) at the visual perception layer was developed based on YOLO v8 model. This model integrated color histogram-based region segmentation and boundary fitting estimation algorithms to generate real-time occlusion rates and suppression weights heatmaps for target fruits. At the motion decision layer, a depth camera was rigidly mounted on the robotic arm's end-effector, establishing a three-dimensional occlusion suppression matrix mapping mechanism. This mechanism discretized the depth space from the camera to the target into a three-dimensional grid matrix, where each grid cell contained the OSCM-derived occlusion rate and suppression priority. Building on this framework, a deep reinforcement learning-driven obstacle avoidance decision method was proposed. It employed a composite reward function incorporating occlusion rate penalties and harvesting timeliness, enabling the double deep Q-network (DDQN) to search for the optimal suppression path within the three-dimensional grid. Experimental results demonstrated that in high-occlusion scenarios with branch-leaf density no less than 60%, the system reduced the average occlusion rate by 33%, increased the harvesting success rate from 67% to 94.7%, and decreased the average picking time per fruit by 3.2 s. The research pioneered the closed-loop integration of occlusion suppression mechanisms with reinforcement learning decision-making,

收稿日期：2025-07-13　修回日期：2025-07-23
基金项目：国家重点研发计划项目(NK2023150202)
作者简介：苗中华(1977—)，男，教授，博士，主要从事智能装备与机器人技术研究，E-mail：zhhmiao@shu.edu.cn
通信作者：孙腾(1989—)，男，讲师，博士，主要从事机器感知理论和采摘机器人技术研究，E-mail：sunny315@shu.edu.cn

providing a novel methodology for enhancing agricultural robots' adaptability in dynamic environments.

Key words: apple picking robot; reinforcement learning; occlusion suppression; robotic arm; path planning

0 引言

苹果采摘机器人是农业智能化发展的关键装备,其技术进展直接影响果园的生产效率与产业可持续发展,是应对劳动力短缺与作业瓶颈的重要路径[1-3]。然而,复杂采摘场景中密集枝叶对目标苹果的随机遮挡导致现有采摘系统存在"感知-决策双退化"现象[4]:一方面,常用目标检测深度模型(如 YOLO 系列)因遮挡干扰产生定位漂移,平均误差达 3.5 cm[5];另一方面,机械臂运动规划器忽视或难以适应环境状态变化,致使采摘成功率普遍低于 70%[6]。视觉伺服技术(Visual servoing)通过视觉反馈实现目标追踪与控制闭环,而主动视觉(Active vision)则强调通过控制视角提升感知效率和信息获取能力。尽管视觉伺服控制和主动视觉等技术已在农业机器人感知与控制任务中得到广泛应用,相关研究在动态感知、多源融合与任务导引方面已有较为成熟的理论积累[7],这些方法已被用于果实识别、抓取定位、动态避障等任务中[8-10]。然而,在复杂果树环境中,尤其面对枝叶遮挡率超过 60% 的高遮挡场景时,传统视觉伺服研究多单一聚焦改进深度学习网络结构增强遮挡感知[11],策略多基于固定视角或单次动作反馈[12],难以动态建模遮挡分布、预估多步路径并结合抑制策略进行联合优化[13]。目前仍缺乏针对"遮挡抑制-路径规划-动作选择"一体化的视觉-运动协同决策系统,尤其缺少强化学习与遮挡注意机制融合的研究框架。这种解耦架构式研究导致两大技术缺陷:环境状态更新频率受限于固定观测逻辑,难以应对复杂遮挡状态变化;视觉特征与运动状态缺乏协同映射,决策滞后引发路径失效[14]。

本文突破传统"感知-决策"解耦式架构,提出一种视觉-运动协同的主动感知采摘新范式:构建遮挡状态动态计算模型(OSCM),融合颜色直方图区域分割与边界拟合算法,实现果实遮挡率与遮挡抑制热力图的实时生成;设计三维空间遮挡权重注意力驱动的机械臂运动决策模型,通过自注意力网络动态聚焦关键遮挡区域,相比传统全局搜索策略,显著减少无效空间探索;提出 DDQN(Double deep Q-network)驱动的三维遮挡矩阵路径寻优机制,将相机-目标空间及其邻域映射为离散化栅格矩阵,通过双重 Q 网络解耦动作选择与价值评估,使机械臂在三维状态空间中实现最小遮挡路径快速收敛。

1 系统总体设计

苹果采摘机器人的硬件层由 JAKA 六自由度机械臂、末端固联的 RealSense D435i 型深度相机及 DH-AG95 型夹爪执行器构成,通过 ROS Melodic 实现多模态数据同步。主动感知采摘方法的主要原理流程如图 1 所示。

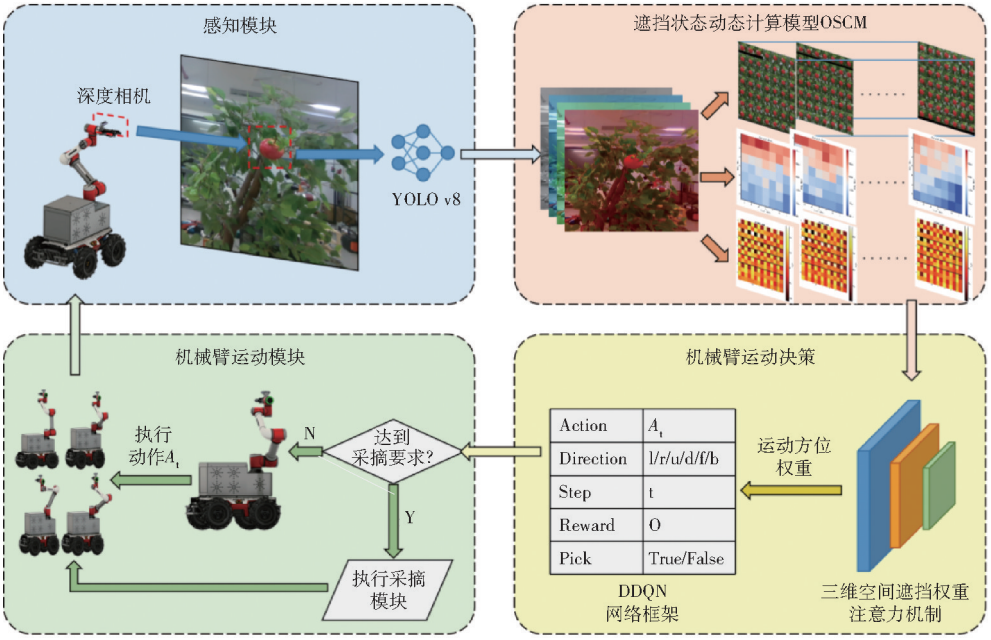


图 1 主动感知采摘方法的主要原理流程

Fig. 1 Main principle and process of active perception picking method

方法运行流程如下：

(1)初始感知：利用 YOLO v8 模型获取目标苹果当前观测位置的置信度 C、遮挡率 O 和遮挡抑制状态，其中 C 直接通过 YOLO v8 模型获得[15]，而 O 和遮挡抑制状态通过集成深度学习和图像处理方法的 OSCM 计算获得。

(2)动态观测：机械臂带动相机移动到新的位置观测获取数据，重复“移动-观测”行为构建以目标为中心的三维栅格矩阵($x \times y \times z$)，本文采用 $x = y = 7$，深度维度 $z = 3$，每层栅格边长 10 cm，生成大小为 $x \times y \times z$ 的置信度和遮挡情况矩阵。

(3)构建奖励：将获取到的三维栅格矩阵作为强化学习训练环境，将矩阵中 C 和 O 作为阈值和输入构建奖励函数，遮挡抑制状态输入基于三维空间的遮挡抑制注意力机制作为局部动作选择的权重。

(4)决策训练：将其整合输入强化学习 DDQN 进行动作选择与路径优化，通过对目标的精准遮挡状态识别和将相机到目标间的三维空间栅格化，机械臂最优轨迹规划问题转换为在三维空间矩阵中寻找最小遮挡抑制的走迷宫问题。

基于主动感知策略结合 OSCM 和 DDQN，机械臂执行移动行为，更新感知状态，重复进行“移动感知-决策规划-移动逼近”的闭环操作机制，实现了精准高效的采摘策略。

2　遮挡状态动态计算模型(OSCM)构建

OSCM 模型的感知建立在改进型 YOLO v8 - L 实例分割网络基础之上，该模型在识别准确率和速度之间取得了较好的平衡[16]，且拥有多个预训练模型，方便定制化部署[17]。通过对预训练的检测模型 YOLO v8 - L 进行了调整适配以满足本研究的需求：将安装在机械臂末端关节上的深度相机获取的 RGB 图像作为输入，输入中可能包含子实体、叶子、茎和环境背景等。网络输入为经过自适应直方图均衡化与伽马校正的预处理图像，以应对果园环境光照不均的问题。

此外，YOLO v8 模型使用了实例分割技术，网络输出包含目标检测框(Bounding box)、实例分割掩膜(mask)及置信度三元组。其中，检测框用于初步定位果实 ROI 区域，掩膜则通过二值化处理生成精细化目标轮廓，结合可以计算目标的遮挡率[18]。针对部分遮挡场景，收集了来自不同温室和各种光照条件下的苹果视频数据。从这个全面的数据集中，提取了大量的图像，主要关注被部分遮挡的情况。这是为了增强深度学习模型的稳健性，确保即使只有一部分检测目标存在(如置信度低至 0.3)，它仍然可以提供检测到的边界框。

2.1　区域分割遮挡率计算

区域分割法采用了 YOLO v8 模型输出自带的位于边界框内的目标 mask，来区分属于水果本身的像素和不属于水果本身的像素。具体来说，在使用分割模型分割水果区域后，为该区域分配较深的颜色。然后，提取同时输出的实例分割结果 mask，生成二值掩膜 $M(x,y)$，计算公式为

$$M(x,y) = \begin{cases} 1 & (\text{像素}(x,y) \in \text{mask}) \\ 0 & (\text{像素}(x,y) \notin \text{mask}) \end{cases} \tag{1}$$

其中 mask 内像素显示为白色。最后基于分割框的长 H、宽 W，计算 BOX 中黑色和白色像素的总数，相应的遮挡率O_{seg}计算公式为

$$O_{\text{seg}} = \left(1 - \frac{\sum_{(x,y) \in B} M(x,y)}{N_B}\right) \times 100\% \tag{2}$$

式中　B——检测框区域

N_B——框内总像素数

区域分割法具有处理速度快、定位精准且易于解释和调整等优点，例如加入补偿系数或误差修正。但应注意，提取的区域仅为近似苹果的形状，不仅包含实际水果区域，还可能包含背景和其他物体[19]。此外，在遮挡程度较高的情况下，提取的水果可能不完整，导致估算遮挡率出现偏差。为了解决这些问题，本文提出了一种基于侵蚀膨胀算法的偏差处理方法[20]，首先对 ROI 区域周围进行如图 2 所示的处

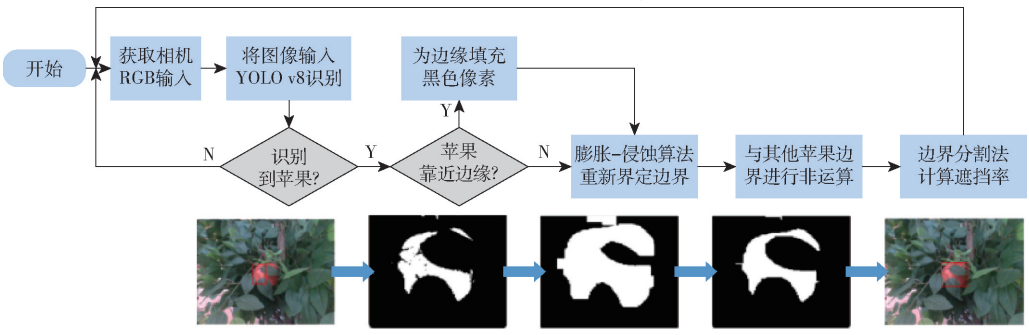


图 2　边界分割侵蚀膨胀算法遮挡率计算流程

Fig. 2　Calculation process of occlusion rate for boundary segmentation erosion dilation algorithm

理步骤。将 ROI 周边区域提取后首先进行 HSV 阈值处理,然后进行相同系数的先膨胀后侵蚀操作。为避免聚合后将附近其他水果误识别为当前水果的情况,需在重构 ROI 区域后与其他 ROI 区域进行逻辑非运算,再与原 ROI 区域进行逻辑或运算。

2.2 边界拟合遮挡率计算

边界拟合法主要基于苹果本身的规则集合形状可以近似为圆形。具体而言,由 RealSense D435i 型深度相机获得到的目标点云集 $P=\{p_i \mid p_i \in \mathbf{R}^3\}$,可将其映射至 RGB 图像后预估出当前苹果的最大近似直径,以获得目标半径 r 及其中心点。在获得近似的半径和中心点后,同区域分割算法一样提取实例分割结果 mask,生成二值掩膜 $M(x,y)$,并累加和计算这些像素占拟合圆形总像素数的比例[21]。最终的遮挡率 O_{geo} 计算公式为

$$O_{\text{geo}}=1-\frac{\sum M(x,y)\psi(\|\text{proj}(x,y)-C\|\leqslant r)}{\pi r^2} \tag{3}$$

式中 C——圆心坐标坐标

r——苹果拟合圆半径

$\text{proj}(x,y)$——像素点(x,y)在拟合球面上的投影坐标

ψ——指示函数,满足条件时取 1,否则取 0

边界拟合法对部分遮挡如枝叶边缘更加敏感,并且能够更接近苹果本身轮廓计算遮挡率。但实际检测时,近似圆不一定能完美地包围整个果实区域,而且随着遮挡的增加,果实区域预测的长轴会小于其真实长度,导致计算出的遮挡率与实际的遮挡率出现明显差异,因此当遮挡率与实际有较大偏差时,需要进行补偿调整。另一方面,在预先给定不同深度的情况下,通过获取真实苹果在二维平面上所占像素数的平均数,可以预先确定苹果的平均尺寸。这可以作为补偿由于角度扭曲或遮挡而导致的苹果感知尺寸与实际尺寸差异的指标。

2.3 动态拟合计算优化及遮挡抑制热力图生成

综上所述,区域分割法使用圆形表示苹果的尺寸并指示其果实面积,而边界拟合法则将实际果实以外的区域包含在矩形边界框内。这 2 种方法都有其局限性,因此,对 2 种方法的结果进行了整合。为了得到真实的遮挡率,对于每幅图像,手动将遮挡物体的遮挡区域用黑色标出,将水果区域用白色标出,利用颜色阈值计算其面积,并计算遮挡率作为真实值。随后选取 600 个样本数据,将真值曲线与 2 种遮挡方式所测遮挡率通过最小二乘法拟合近似曲线,最终得到公式为

$$O=aO_{\text{seg}}+bO_{\text{geo}}+c \tag{4}$$

式中 a——区域分割法调整系数,取 0.597 6

b——边界拟合法调整系数,取 0.614 3

c——补偿系数,取 −0.178 7

融合 2 种遮挡计算方法后更接近真值,遮挡率平均绝对误差(MAE)下降 4.6%。如图 3 所示,以果实中心(x_c,y_c)为原点,检测区域划分为 4 个 45°扇区,分别对应上、下、左、右四方向遮挡分布区间,配对遮挡抑制热力图以表示不同方向的遮挡强度。每个扇形区遮挡率通过极坐标像素点计数计算,最终得到对应位置的遮挡抑制热力图 O_{heatmap},重复机械臂移动带动相机获取新图像行为,最终构建出三维栅格空间的遮挡抑制热力图。

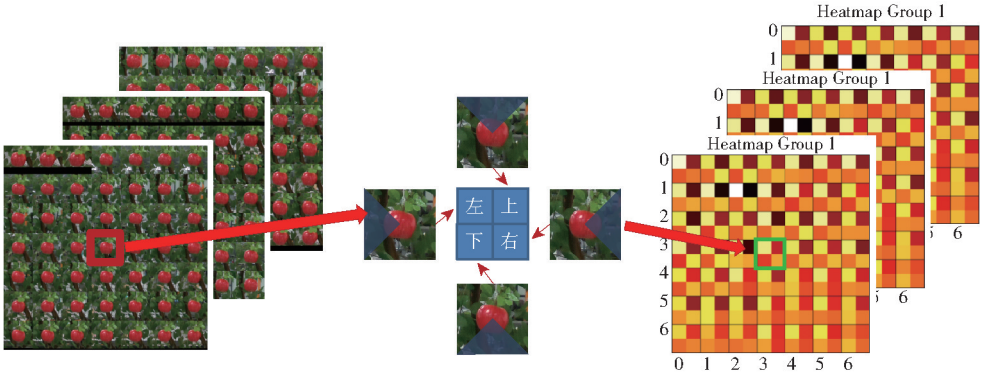


图 3 遮挡抑制热力图生成方法

Fig. 3 Occlusion suppression heatmap generation method

3 机械臂运动决策模型

本研究旨在提出一种高效的主动感知遮挡抑制策略,根据当前的遮挡情况快速准确地确定机械臂自主移动的优化路径以抑制遮挡影响,提高识别准确率和采摘效率。多次单步移动相机进行多视角观察,可以提高识别准确率并在一定程度上减少遮挡影响,但花费时间过多且效率低下,这对于自主采摘作业并不实用[22]。虽然固定多位置观察策略或简单的启发式感知方法可以在相对较低的遮挡水平下提高识别准确率,但它们可能会在运动过程中错过最佳采摘方位,并且过于依赖遮挡物体的检测和定位。

当人类和动物观察到被遮挡的物体时会根据当前的遮挡状态快速调整视角，避开障碍物并重新识别物体。为了使机器人实现这种智能，引入了强化学习 DDQN 进行全局采摘路径规划以达到最优采摘位置，再辅以注意力机制进行动作权重训练实现快速局部避障，令机器人能够最大化抑制遮挡实现精准采摘，并在每次感知后迅速决策并选择下一步动作。

3.1　三维遮挡抑制矩阵映射机制策略

为解决枝叶遮挡对目标果实检测与路径规划造成的干扰问题，本研究在运动决策层引入了基于注意力机制的三维遮挡抑制矩阵[23]，通过将目标果实与相机之间的空间区域离散为若干三维栅格单元，并在每个栅格中赋予感知遮挡率与抑制优先级信息，从而实现遮挡抑制与识别路径的联合局部优化。每个栅格包括置信度 C、遮挡率 O、遮挡抑制热力图 O_{heatmap}，置信度 C 通过 YOLO v8 模型深度学习方法获得，表示在物体周围特定范围内识别物体的确定性水平，遮挡率 O 和遮挡抑制热力图 O_{heatmap} 由 OSCM 获得，表示目标物体在周围受到遮挡的程度和抑制遮挡方位。该结构形成一个包含多维感知信息的三维张量，用于构建强化学习的状态空间。将上述三者以及当前深度 d 作为输入，构造三维栅格化结构化特征张量，将其展开输入至空间注意力决策网络公式为

$$W^i_{\text{spatial}} = \frac{1}{1+\mathrm{e}^{k(O^i-\tau)}} \sum_{\text{dir}} \frac{O^i_{\text{dir}}}{1+\alpha d^i} \tag{5}$$

式中　k——Sigmoid 曲线陡度系数，取 0.2

τ——遮挡率阈值，取 0.8

O^i——第 i 层深度上目标投影面积的平均遮挡率

O^i_{dir}——不同深度下的深度遮挡抑制

α——深度衰减因子，取 0.33

d^i——第 i 层的深度值

i——深度层数，取 1 ~ 3

第 1 项通过 Sigmoid 函数抑制高遮挡（$O>80\%$）区域的权重：当某方向预测单元的遮挡率大于 80% 时，强制置 $W_{\text{action}}(a)=0$，避免机械臂进入不可通行区域；第 2 项融合四向热力分布与深度衰减效应，强化近端低遮挡区域的决策影响。为避免不同深度栅格因尺度变化导致遮挡率估算误差，本文引入基于距离深度的前后遮挡抑制衰减修正机制，计算公式为

$$O^i_{\text{dir}}(d) = O^i\left(1-\frac{d\alpha}{z}\right) \tag{6}$$

不同深度下的深度遮挡抑制 O^i_{dir} 由当前深度 d 确定：由于深度增加时目标投影面积会缩小，影响遮挡率及遮挡抑制计算，且由经验可得采摘和避障路径规划成功率与目标距离成反比，因此深度遮挡抑制由远及近减少，以鼓励机械臂向前移动，具体数值依当前遮挡率、深度 d、栅格化深度维度 z 及深度衰减因子 α 确定。本文深度设置为 3 层，因此深度抑制由当前遮挡率乘以深度层数及衰减因子得到，而深度遮挡因子会逐层等比衰减至 0，保证了在最远和最近层不会选择后退和前进动作。同时基于 DDQN 的动作空间，将遮挡权重投影至 6 个基础动作方向，其中各方向权重 W_{action} 计算公式为

$$W_{\text{action}}(A)=\begin{cases}\displaystyle\sum_{i\in\Omega_A} W^i_{\text{spatial}}\,O^i_{\text{dir}} & (A\in\{\text{上},\text{下},\text{左},\text{右}\})\\ \displaystyle\sum_{i\in\Omega_A}\frac{W^i_{\text{spatial}}}{1+Ad^i} & (A\in\{\text{前},\text{后}\})\end{cases} \tag{7}$$

式中　A——所选动作

由此得到当前状态下的动作权重，将其与 DDQN 耦合得到公式

$$Q(s,A)=Q_{\text{base}}(s,A,\theta)+\gamma W_{\text{action}}(A) \tag{8}$$

式中　θ——DDQN 权重参数，包括所有隐藏层和输出层的连接权重及偏置项

γ——注意力增益系数，本文设定为 0.3

s——强化学习网络状态空间

Q_{base}——DDQN 的动作选择公式，即强化学习在状态 s 下所给出的动作选择

第 1 项为原 DDQN 的 Q 值，第 2 项注意力增益系数由 action 空间确定，并且为注意力权重引入动作屏蔽机制。前者通过时序差分法（TD 误差）学习全局策略以捕获长期奖励，$W_{\text{action}}(A)$ 则反映实时环境变化，使机械臂根据当前感知的局部信息（如右侧明显遮挡），调整动作优先级，最终 Q 值由 θ 的全局策略和 $W_{\text{action}}(A)$ 的局部调整共同决定。通过这种设计，Q 函数既能利用历史经验学习全局最优策略，又能根据实时感知灵活调整动作选择。

与传统路径规划方法不同，本文提出的主动感知采摘策略不仅针对单步动作的优化，同时通过强化学习 $Q(s,A)$ 对整条路径累计奖励进行评估，从而实现路径序列的最优性。这使得机械臂既能选择局部最优动作，并且每一步动作均基于当前状态下的未来最优回报预测，从而实现“逐步逼近最优采摘点”的多步决策链。

3.2　DDQN 框架设计

为增强系统对动态遮挡环境的适应能力，本文设计基于 DDQN 的运动控制模型，构建“感知-决策-移动”的闭环优化流程。在三维空间中，机械臂的每一个动作都被视为强化学习中的一个动作选

择过程,目标是在感知到该栅格的状态信息后,预测一个最优路径以趋近于低遮挡、高置信的采摘位置。

模型的训练遵循以下流程:首先,初始化经验回放缓冲区 D、Q 网络(参数为 θ)及其同步更新的目标网络 Q′(参数为 θ')。在每一轮训练中,环境被重置,智能体获得初始状态 $\boldsymbol{s}_0$。在每一步 t 内,智能体基于当前状态 $\boldsymbol{s}_t$,通过 Q 网络选择具有最大 Q 值的动作 A_t,并由机械臂执行该动作。执行动作后,系统观察到新状态 $\boldsymbol{s}_{t+1}$ 和即时奖励 r_t。若 r_t 已达到预设的最大奖励(即到达最优采摘点),则额外添加 50 的高额奖励并终止当前轮次;否则,将状态转移样本 $(\boldsymbol{s}_t, A_t, r_t, \boldsymbol{s}_{t+1})$ 存入经验回放缓冲区 D。随后,从 D 中随机采样小批量样本进行网络训练:对于每个样本,使用在线网络 Q 选择下一状态 $\boldsymbol{s}'$ 下的最优动作 A',然后利用目标网络 Q′评估该状态-动作对的 Q 值,计算目标 Q 值 y_t。最后,通过最小化预测 Q 值与目标 Q 值之间的均方误差(MSE)来更新在线网络 Q 的参数 θ,并定期将 θ 复制到目标网络 Q′以稳定训练。由于苹果采摘环境的复杂与非结构空间特性,针对单次信息感知难以获取目标周边空间信息致使采摘成功率低的问题,依据主动感知策略将目标及其邻域的空间分割为非连续的等量区域(本文设定为 7×7×3,步长为 10 cm)作为 DDQN 网络的环境,使机器人系统在尽可能贴近真实情况的条件下进行训练[24]。

主动感知的状态更新策略如下:输入状态 $\boldsymbol{s}_t$ 为 7×7×3 的三维栅格矩阵,每个单元包含 6 维特征:总置信度 C、遮挡率 O 以及四向热力分布 $[O_{\text{up}}, O_{\text{down}}, O_{\text{left}}, O_{\text{right}}]$。DDQN 基于三维空间矩阵定义了 6 个基础移动方向作为动作空间:上、下、左、右、前、后。将当前状态的环境感知特征输入三维空间遮挡权重注意力机制得到一组遮挡抑制热力图作为动作权重。每次动作选择后,机械臂执行移动,由末端深度相机重新采集新视角图像并经 OSCM 模块处理,获得更新后的状态 $\boldsymbol{s}_{t+1}$。此过程构成了一个典型的“感知-决策-执行”闭环,其机制用公式表示为

$$\boldsymbol{s}_{t+1} = \boldsymbol{\Phi}(\boldsymbol{s}_t, A_t) = \text{OSCM}(I_{rgbd}(P_t + A_t)) \quad (9)$$

式中 A_t——t 状态选择的动作

P_t——t 状态的空间坐标

I_{rgbd}——深度相机图像

$\boldsymbol{\Phi}$——状态空间更新操作

机械臂执行动作 A_t 后,末端执行器移动至新坐标 $P_t + A_t$ 后,深度相机重新采集环境信息,触发 OSCM 模型动态更新三维栅格特征。

主动感知策略的 DDQN 奖励函数聚焦遮挡规避与路径效率,其表达公式为

$$R(\boldsymbol{s}_t, A_t) = -O_t + \Pi(1,0) + \mathrm{e}^{-0.05l} - 1 \quad (10)$$

式中 O_t——t 状态的遮挡率

l——t 状态时移动步数

$\Pi(1,0)$——轮次终止判断条件

由于遮挡率取值范围为[0,1],将其映射至[-1,0]区间,随后加入采摘效率惩罚($\mathrm{e}^{-0.05l} - 1$),采用指数衰减系数取 -0.05 以平衡探索与利用,从而标准化奖励于[-1,1]。最后设定完成目标时获得奖励 50 以鼓励网络到达最优采摘点。网络训练过程中采用经验回放机制,以打破状态序列相关性,提高学习效率。

以三维栅格矩阵为状态空间,每个状态含有当前位置的遮挡率、置信度及四向热力信息,DDQN 在策略收敛后可输出从任意初始位置至最优采摘点的完整动作序列,即使在高遮挡环境下,系统可以通过 3~5 步连续动作实现遮挡最小化路径搜索的典型实例。此外,动作权重由注意力机制动态调节,使得在不同遮挡状态下的路径规划具有全局一致性和方向性,避免陷入梯度消失或过拟合情况。对比传统固定视角规划,本方法在多层遮挡场景中的路径长度缩短 33%,因为机械臂可通过主动移动获取更优观测角度。最终,当遮挡率达到最小值时则奖励函数为最大值,即成功到达最佳采摘位置。

4 实验与结果分析

为验证本文提出的视觉-运动协同采摘系统的综合性能,进行了数据采样、模型训练和真实世界实验。通过使用所提出的遮挡状态计算模型和主动感知采摘算法,测试了农业环境中遇到的各种遮挡情况,并在室内仿生果园环境进行了实验。

4.1 数据集建立

为了训练 YOLO v8 模型识别和分割网络,在实验室中拍摄了包含多幅苹果图像的视频。具体步骤如下:首先将苹果随机放置在模型树上,并在实验室中使用树枝和树叶创建不同程度的遮挡状态。随后,使用安装在机械臂上的深度摄像头获取了大量视频。此外,为了增加数据集的数据量,除了使用安装在移动机器人机械臂上的深度摄像头外,还使用手持摄像头从不同角度拍摄了图像和视频。获得的数据涵盖了广泛的遮挡情况,有利于构建数据库和训练深度网络[25]。

由于视频中存在非常相似的图像,因此以特定的帧间隔从视频中抽选图像。然后,调整每幅图像的亮度,并将其尺寸调整为 640 像素×480 像素。总共获得了 7 350 幅苹果图像。在对图像进行标记后,创建了训练和验证数据集。

为了获得主动感知决策网络所需的训练数据集

（如第 3 节所述），本文设置了不同实验场景，这些场景具有不同的遮挡程度，范围从 0% 到 80%，每 20% 的间隔内至少包含一个场景。对于每个场景，通过移动机械臂和相机获取与场景相对应的所需数据，形成“遮挡抑制-置信度图”和“遮挡抑制热力图”数据集。对于每个场景，收集 7 ×7 ×3 的三位网格矩阵图，其中沿每个轴采取 $d_x = d_y = 10$ cm 的步长，并使用 $z_{off} = 10$ cm 来保持深度方向偏移。最终，在沿 x 和 y 轴采取 6 个 10 cm 的步长以及在深度方向上采取 2 个 10 cm 的步长后，在 60 cm × 60 cm ×20 cm 的范围内获得具有 7 ×7 ×3 个网格的三维地图。总共收集了 600 组数据图。

4.2　遮挡场景测试

为系统评估 OSCM 模型在复杂遮挡环境下的感知能力与机械臂运动决策的实时性，本研究设计了真实世界模态测试场景。测试平台采用模块化仿真苹果树结构，其枝叶密度可通过增减枝条动态调整。在测试中，设置逐渐增加的典型枝叶遮挡率测试，遮挡率由 0 逐渐提升至 80%。实验结果表明：在 40% 遮挡率以下，OSCM 的 MAE 为 1.2%，主要误差来源于枝叶边缘的微小遮挡；当遮挡率增至 40% 以上时，MAE 上升至 2.7%，系因高密度枝叶导致 YOLO v8 模型分割掩膜出现孔洞，对比单纯的区域分割和边界拟合法 MAE 分别下降了 5.3% 和 3.8%，平均 MAE 降至 1.8%。同时选择不同方位的枝叶遮挡以测试遮挡抑制热力图生成，以果实中心为原点建立极坐标系，热力图的抑制平均误差随密度增加呈线性上升趋势。两组实验结果如图 4 所示。

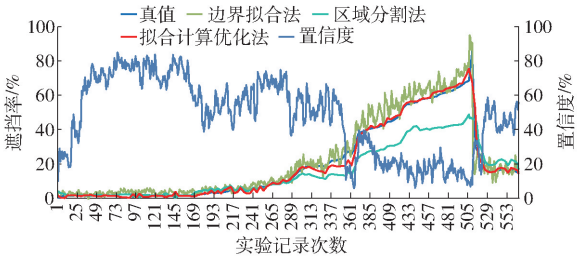


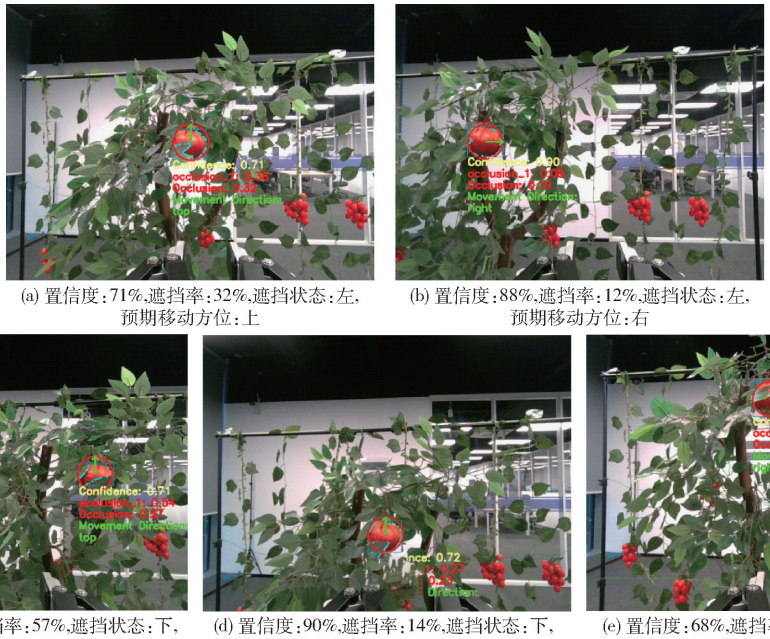
图 4　低/高遮挡情况下不同遮挡率实验曲线

Fig. 4　Comparison chart of calculation results for different occlusion rates under low/high occlusion conditions

4.3　采摘对比实验结果分析

三维空间遮挡权重注意力机制能够有效判断出当前状态下不同动作对遮挡状态的变化趋势，并辅助 DDQN 框架决策机械臂运动，其效果如图 5 所示。其中，图 5a 为栅格化三维矩阵的中心点，将 OSCM 得到遮挡抑制数据输入注意力机制中得到动作权重，图 5b ~5e 为选择向右、向左、向上、向下 4 个动作后新的状态下的情况，可以看到基于遮挡状态和预期移动方位采取的动作能够有效提高置信度并降低遮挡率。

为评估本文 DDQN 模型在不同遮挡场景下的收敛性表现，本文在仿真环境中设置了 4 组典型遮挡等级（0 ~20%，20% ~40%，40% ~60%，60% ~80%），分别记录每组下的累积奖励变化趋势。图 6 展示了 4 组实验中 DDQN 在训练迭代过程中的总奖励变化曲线。结果显示：在遮挡率较低的前 2 组中，DDQN 模型均能快速收敛，总奖励值稳定在 50 左右，表明低遮挡场景下动作策略学习稳定性良好，模型训练过程鲁棒性高；在遮挡率 40% ~60%



(a) 置信度：71%，遮挡率：32%，遮挡状态：左，预期移动方位：上

(b) 置信度：88%，遮挡率：12%，遮挡状态：左，预期移动方位：右

(c) 置信度：41%，遮挡率：57%，遮挡状态：下，预期移动方位：上

(d) 置信度：90%，遮挡率：14%，遮挡状态：下，预期移动方位：上

(e) 置信度：68%，遮挡率：38%，遮挡状态：左，预期移动方位：右

图 5　主动感知采摘方法结果展示

Fig. 5　Results of active perception picking method

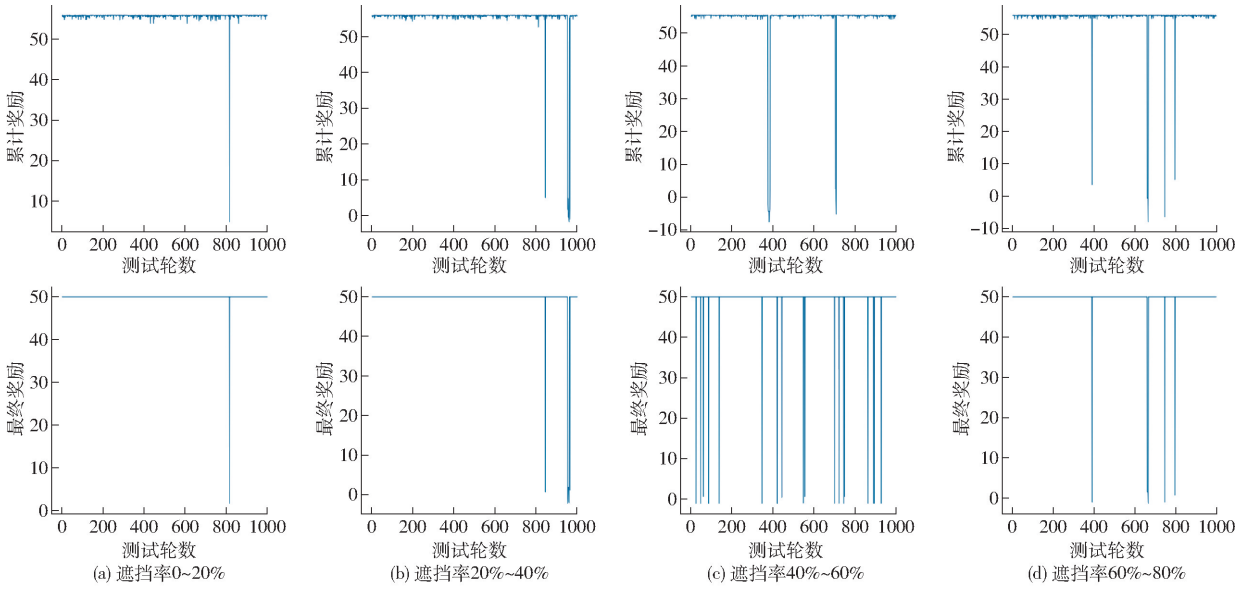


图 6 4 组 DDQN 框架仿真测试结果

Fig. 6 Four sets of DDQN framework simulation test results

组中,模型依然可以收敛至最优策略,但收敛速度略有下降,训练过程中偶有波动。这主要由于中度遮挡对置信度评估和动作选择造成局部干扰,导致策略网络需更多迭代周期完成收敛。第 4 组实验中,模型训练曲线明显波动较大,平均奖励值呈现上下收敛过程变得缓慢,虽然最终仍可达到近似最小遮挡路径,但策略收敛不稳定性加剧,且训练周期显著延长。

进一步分析表明,DDQN 收敛性能主要受 4 个因素制约:高遮挡率导致环境不确定性与感知误差增大,增加稳定学习所需样本量;遮挡抑制热力图在高遮挡区易失真,影响动作价值评估准确性;动作屏蔽机制在高遮挡下限制有效动作空间,加大搜索与训练难度;高遮挡状态中奖励值普遍为负、反馈稀疏,亦延缓收敛进程。

为进一步评估所提方法的计算效率与模型复杂度,本文将其与传统连续空间强化学习方法 DDPG 进行了对比。在相同的仿真训练环境与计算硬件下,本文基于离散三维栅格与注意力机制的 DDQN 方法,其训练策略收敛至稳定性能所需的平均迭代轮数较 DDPG 方法减少了约 25%。这得益于状态空间的离散化降低了学习复杂度,以及注意力机制帮助网络更聚焦于关键遮挡区域,从而加速了策略探索。然而,引入的空间注意力模块也增加了网络的参数量,经统计,本文 DDQN 网络的总参数量较未引入注意力机制的基线 DDQN 网络增加了约 18%。这反映了系统在提升决策精度与收敛速度的同时,所付出的模型复杂度代价。

最后为全面评估系统的综合性能,选取 3 类主流方法作为对比实验:无移动和随机移动,单次主动感知和简单的启发式方法[26];主动深度探索(ADS)[27];本文的基于主动感知策略的 DDQN 方法。第 1 类为无遮挡抑制或随机遮挡抑制下的采摘,第 2 类为单步非深度学习下的遮挡抑制采摘,第 3 类为深度学习下的遮挡抑制采摘。其中,单次主动感知为移动至目标一半深度位置以实现近距离再识别定位,启发式的原理为向最小遮挡抑制方向移动固定步长距离,ADS 依靠卷积神经网络预估该环境下的最佳采摘位置作为遮挡抑制与路径规划方法。在实验室仿真果树上进行实验,涵盖 5 种典型采摘遮挡情况,每组实验重复 20 次以消除随机误差后取平均值,最终实验结果如表 1 所示。

表 1 各方法对比实验结果

Tab. 1 Results of comparative experimental for various methods

方法	遮挡率变化率/%	采摘成功率(模拟)/%	采摘成功率(实验)/%	单果采摘时间/s
无遮挡抑制	0	31	29	12.1
随机遮挡抑制	7	46	40	24.6
单次主动感知	-10	53	51	20.8
启发式感知	-18	71	67	18.9
ADS	-25	90	88	15.4
主动感知 DDQN	-33	98	95	15.7

表 1 结果表明,本文所提主动感知 DDQN 方法在各项性能指标上均表现最优。该方法在遮挡率超过 60% 的高遮挡场景下,能使目标果实的视觉遮挡率平均降低 33%,并将采摘成功率从传统启发式方法的 67% 显著提升至 95%。具体而言,在遮挡率 80% 的极端条件下,本方法相较于单次主动感知

方法，采摘成功率提升了 44 个百分点；相较于启发式感知方法，成功率提升了 28 个百分点。在采摘效率方面，本方法平均单果采摘时间为 15.7 s，虽因执行“感知决策移动”多步闭环决策而较无移动感知策略耗时有所增加（平均增加约 33%），但相较于启发式方法仍实现了约 29% 的效率提升。这体现了本系统在复杂遮挡环境下，为达成高成功率而对感知决策深度与采摘时效所进行的有效权衡。

为进一步直观展示本文方法在实际采摘过程中的轨迹选择与视场变化情况，图 7 总结了一次采摘过程的机械臂运动规划过程。图 7a 中浅蓝色区域为机械臂的起始位置，机械臂首先通过 OSCM 和三维空间注意力机制主动避让高遮挡区域，选择向左移动到达更优观察角度从而动态获取目标的最优可见面。然后图 7b、7c 中重复了“观测-决策-移动”这一闭环过程，其中图 7b 虽然 OSCM 推荐向下或向左，但注意到遮挡率的下降，基于 DDQN 网络和注意力机制选择向前动作；接着图 7c 继续向前后到达位置“3”，由运动决策网络模块决定采摘，最终完成路径收敛至采摘位点的动作链并到达满足采摘条件的采摘点位。

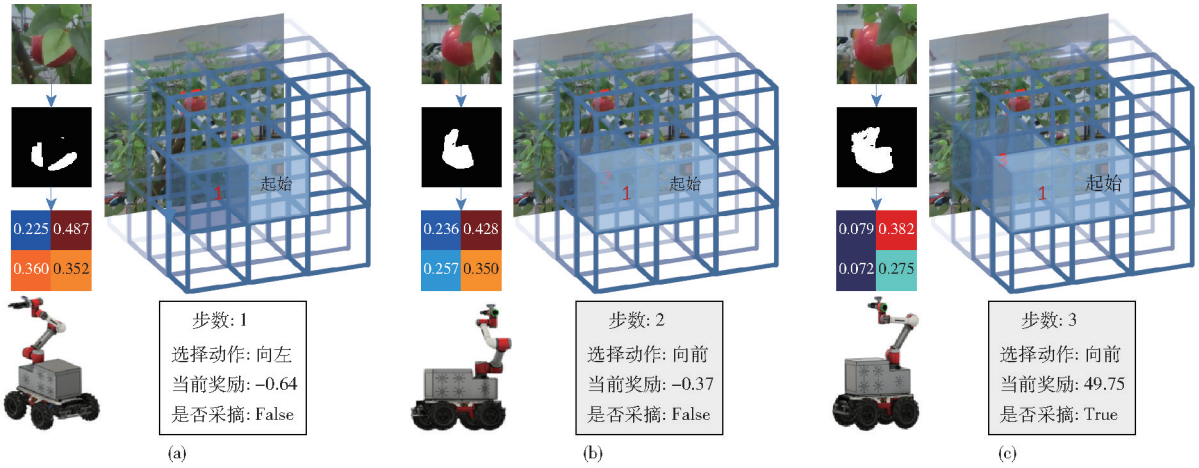


图 7　采摘过程中的机械臂运动轨迹规划

Fig. 7　Trajectory planning of robotic arm movement during harvesting process

综上，本文通过主动感知强化学习方法，不仅优化了单步动作选择，还在三维空间中实现了路径规划的全局性能提升。实验结果表明，DDQN 框架在不同遮挡密度下均可学习稳定、收敛的多步动作序列，其路径遮挡率下降趋势清晰、动作数量稳定，展示出良好的全局抑制路径搜索能力。此外，结合遮挡抑制热力图与注意力机制，有效避免了传统贪心策略易陷入局部最优的问题，增强了路径规划策略在复杂农业环境中的适应性与鲁棒性。当然，算法仍有更多改进空间以增加实用性。例如，当前使用的强化学习网络相对简单，可以在示教模式下采用 DDPG 网络作为连续移动方法[28]，在自主移动过程中改变机械臂的方向，以探索所提方法的可行性。并且，对于 OSCM 的障碍物识别也可进行分类，例如对于树干树枝等“硬遮挡”必须考虑避障，但对于树叶藤蔓等“软遮挡”也许可以考虑更少的移动距离[29]。

5　结论

（1）提出了一种基于主动感知策略的苹果采摘机器人动态决策方法。将 YOLO v8 模型网络与深度相机点云数据融合构建遮挡状态动态计算模型（OSCM），通过 HSV 颜色空间下的实例分割掩膜实现区域分割与边界拟合双模态融合，使遮挡率计算平均绝对误差（MAE）降至 1.8%；同时在栅格化三维空间矩阵中引入空间注意力机制，基于 OSCM 建立遮挡抑制热力图与机械臂动作的映射关系，将路径规划问题转换为离散空间的最优路径搜索；最后设计双重深度 Q 网络解耦动作选择与奖励评估，结合优先级经验回放机制提升收敛速度。实验表明，在 80% 遮挡密度下，本文方法相较于传统单次主动感知、启发式方法，采摘成功率分别提升 44 个百分点与 28 个百分点，平均采摘耗时缩短至 15.7 s，平均采摘效率提升 29%。

（2）基于主动感知策略的视觉-运动协同系统架构通过遮挡状态动态计算模型、栅格化三维空间注意力机制和 DDQN 网络实现了计算效率与决策精度的平衡，实验表明，相较于传统连续空间强化学习方法，训练收敛速度增加 25%，但因引入注意力机制及多步决策流程，网络参数量增加 18%，且单果采摘总耗时平均增加约 33%。

参 考 文 献

[1] MARINOUDI V, SØRENSEN C G, PEARSON S, et al. Robotics and labour in agriculture. A context consideration[J]. Biosystems Engineering, 2019, 184: 111 - 121.

[2] 汪一昕, 贺磊盈. 基于点云的果园环境苹果定位技术研究[J]. 建模与仿真,2023, 12(3): 1976 - 1986.
WANG Yixin, HE Leiying. Research on apple positioning technology in orchard environment based on point cloud[J]. Modeling and Simulation, 2023, 12(3): 1976 - 1986. (in Chinese)

[3] 赵辉,乔艳军,王红君,等. 基于改进 YOLOv3 的果园复杂环境下苹果果实识别[J]. 农业工程学报,2021,37(16):127 - 135.
ZHAO Hui, QIAO Yanjun, WANG Hongjun, et al. Apple fruit recognition in complex orchard environment based on improved YOLOv3[J]. Transactions of the CSAE, 2021,37(16):127 - 135. (in Chinese)

[4] GE Y, XIONG Y, FROM P J. Three-dimensional location methods for the vision system of strawberry-harvesting robots: development and comparison[J]. Precision Agriculture, 2023, 24(2): 764 - 782.

[5] 刘彦清. 基于 YOLO 系列的目标检测改进算法[D]. 长春:吉林大学,2021.
LIU Yanqing. An improved algorithm for object detection based on YOLO series[D]. Changchun: Jilin University, 2021. (in Chinese)

[6] 刘双喜,张玮平,胡宪亮,等. 基于改进 YOLO v8 模型 s 的水稻种植机械作业质量检测[J]. 农业机械学报,2024,55(增刊 1):61 - 70.
LIU Shuangxi, ZHANG Weiping, HU Xianliang, et al. Quality inspection of rice planting machinery operations based on improved YOLO v8 model[J]. Transactions of the Chinese Society for Agricultural Machinery, 2024,55(Supp. 1):61 - 70. (in Chinese)

[7] CHUNJIANG Z, BEIBEI F, JIN L, et al. Agricultural robots: technology progress, challenges and trends[J]. Smart Agriculture, 2023, 5(4): 1 - 15.

[8] LI Y, FENG Q, LIU C, et al. MTA - YOLACT: multitask-aware network on fruit bunch identification for cherry tomato robotic harvesting[J]. European Journal of Agronomy, 2023, 146: 126812.

[9] FENG Q, ZOU W, FAN P, et al. Design and test of robotic harvesting system for cherry tomato[J]. International Journal of Agricultural and Biological Engineering, 2018, 11(1): 96 - 100.

[10] LI T, XIE F, ZHAO Z, et al. A multi-arm robot system for efficient apple harvesting: perception, task plan and control[J]. Computers and Electronics in Agriculture, 2023, 211: 107979.

[11] LI Y, FENG Q, LI T, et al. Advance of target visual information acquisition technology for fresh fruit robotic harvesting: a review[J]. Agronomy, 2022, 12(6): 1336.

[12] DAS A, PATHAN F, JIM J, et al. XLTLDisNet: a novel and lightweight approach to identify tomato leaf diseases with transparency[J]. Heliyon, 2025, 11:4.

[13] KOK E, CHEN C. Occluded apples orientation estimator based on deep learning model for robotic harvesting[J]. Computers and Electronics in Agriculture, 2024, 219: 108781.

[14] JI W, HUANG X, WANG S, et al. A comprehensive review of the research of the "Eye - Brain - Hand" harvesting system in smart agriculture[J]. Agronomy, 2023, 13(9): 2237.

[15] DROUKAS L, DOULGERI Z, TSAKIRIDIS N, et al. A survey of robotic harvesting systems and enabling technologies[J]. Journal of Intelligent & Robotic Systems, 2023, 107(2): 21.

[16] MIAO Z, YU X, LI N, et al. Efficient tomato harvesting robot based on image processing and deep learning[J]. Precision Agriculture, 2023, 24(1): 254 - 287.

[17] ZUO G, TONG J, WANG Z, et al. A graph-based deep reinforcement learning approach to grasping fully occluded objects[J]. Cognitive Computation, 2023, 15(1): 36 - 49.

[18] LI T, WANG C, MENG M, et al. Attention-driven active sensing with hybrid neural network for environmental field mapping[J]. IEEE Transactions on Automation Science and Engineering, 2021, 19(3): 2135 - 2152.

[19] 袁杰,谢霖伟,郭旭,等. 基于改进 YOLO v7 的苹果叶片病害检测方法[J]. 农业机械学报,2024,55(11):68 - 74.
YUAN Jie, XIE Linwei, GUO Xu, et al. Apple leaf disease detection method based on improved YOLO v7[J]. Transactions of the Chinese Society for Agricultural Machinery, 2024,55(11):68 - 74. (in Chinese)

[20] DU X, CHENG H, MA Z, et al. DSW - YOLO: a detection method for ground-planted strawberry fruits under different occlusion levels[J]. Computers and Electronics in Agriculture, 2023, 214: 108304.

[21] BLOK P, VAN HENTEN E, VAN EVERT F, et al. Image-based size estimation of broccoli heads under varying degrees of occlusion[J]. Biosystems Engineering, 2021, 208: 213 - 233.

[22] 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.

[23] TANG Y, CHEN M, WANG C, et al. Recognition and localization methods for vision-based fruit picking robots: a review[J]. Frontiers in Plant Science, 2020, 11: 510.

[24] PARK J, LEE S, LEE J, et al. GadgetArm—automatic grasp generation and manipulation of 4-DOF robot arm for arbitrary objects through reinforcement learning[J]. Sensors, 2020, 20(21): 6183.

[25] MOHAMMED M, KWEK L C, CHUA S, et al. Deep reinforcement learning-based robotic grasping in clutter and occlusion[J]. Sustainability, 2021, 13(24): 13686.

[26] SUN T, LIU H, MIAO Z. Active object perception using Bayesian classifiers and haptic exploration[J]. Autonomous Robots, 2023, 47(1): 19 - 36.

[27] SUN T, ZHANG W, GAO X, et al. Efficient occlusion avoidance based on active deep sensing for harvesting robots[J]. Computers and Electronics in Agriculture, 2024, 225: 109360.

[28] 王晶,苏工兵,袁梦,等. 基于深度强化学习的采摘机械臂研究[J]. 数字农业与智能农机,2023(6):13 - 15.
WANG Jing, SU Gongbing, YUAN Meng, et al. Research on deep reinforcement learning-based harvesting manipulators[J]. Digital Agriculture and Intelligent Agricutural Machinery, 2023(6):13 - 15. (in Chinese)

[29] SUN T, ZUO H, ZHOU Z, et al. Enhancing safety during harvesting with a deep learning-based action determination method for picking robots[C]//International Conference on Optical and Photonic Engineering (icOPEN 2024). SPIE, 2025: 248 - 253.