doi:10.6041/j.issn.1000-1298.2026.05.007

基于 ALH－SAC 算法的柑橘采摘机械臂无碰撞抓取路径规划

兰玉彬[1,2]　宋皓昱[1,2]　王　莹[1,2]　刘星宇[1,2]　李　想[1,2]　黎源鸿[1,2]

(1. 华南农业大学人工智能与低空技术学院, 广州 510630; 2. 国家精准农业航空施药技术国际联合研究中心, 广州 510630)

摘要: 在果园等复杂农业环境中,机器人果实采摘面临枝叶遮挡、目标到达不准以及采摘效率低等技术难题。本文提出了一种基于深度强化学习的机械臂无碰撞高效抓取方法。该方法融合了基于 YOLO 11 的柑橘果实分类与定位的感知模型,以及基于压力传感器的柔性夹具压力预测模块,并改进了一种基于深度强化学习的路径规划算法 Attention－LSTM－HER－SAC(ALH－SAC),以实现稳定且无损的抓取。在运动规划方面,ALH－SAC 算法在软演员-评论家(SAC)框架中引入了事后经验回放(HER)、长短期记忆网络(LSTM)编码器以捕捉机械臂关节动作的时间依赖性,以及注意力编码器(Attention Encoder)以自适应聚焦关键空间特征,从而在遮挡环境下提升策略的感知能力与决策精度。设计了自定义奖励函数,引导智能体实现快速到达目标、避障及优化抓取姿态。此外,提出的混合引导机制结合启发式先验与基于笛卡尔空间线性插值的先验策略,加速了训练收敛并提升了策略鲁棒性。仿真试验结果表明,ALH－SAC 的定位精度达(10.35±4.18) mm,较 HER－TD3 和 HER－DDPG 平均提升 12.5 mm。果园试验进一步验证了方法的有效性,平均采摘成功率达 84%,单次操作平均耗时 10.58 s,充分证明了该方法在柑橘采摘场景中的有效性。

关键词: 柑橘采摘; 机械臂; 路径规划; SAC 算法; 柑橘检测; YOLO 11

中图分类号: S24; TP242　　**文献标识码**: A　　**文章编号**: 1000-1298(2026)05-0082-13　　**OSID**:

Collision-free Grasp Path Planning for Citrus-picking Manipulator Based on ALH－SAC Algorithm

LAN Yubin[1,2]　SONG Haoyu[1,2]　WANG Ying[1,2]　LIU Xingyu[1,2]　LI Xiang[1,2]　LI Yuanhong[1,2]

(1. *College of Artificial Intelligence and Low-altitude Technology*, *South China Agricultural University*, *Guangzhou* 510630, *China*
2. *National Center for International Collaboration Research on Precision Agriculture Aviation Pesticides Spraying Technology*, *Guangzhou* 510630, *China*)

Abstract: In complex orchard environments, robotic fruit picking is challenged by branch occlusions, positioning inaccuracies, and low operational efficiency. A deep reinforcement learning-based framework for collision-free and efficient citrus harvesting with a robotic manipulator was proposed. The system integrated a YOLO 11 perception model for fruit classification and localization, together with a pressure prediction module for a flexible gripper to enable stable and damage-free grasping. An improved path planning algorithm, termed Attention－LSTM－HER－SAC (ALH－SAC), was developed to achieve adaptive and reliable grasp control. The proposed algorithm extended the soft actor-critic framework by incorporating hindsight experience replay, an LSTM encoder for modeling temporal joint dependencies, and an attention mechanism to emphasize critical spatial features under occlusion. A customized reward function was designed to balance rapid target approach, obstacle avoidance, and optimal grasp orientation. In addition, a hybrid guidance strategy combining heuristic priors with a Cartesian-space linear interpolation prior was introduced to accelerate training convergence and enhance robustness. Simulation results demonstrated that ALH－SAC achieved a positioning accuracy of (10.35±4.18) mm,

收稿日期: 2025－12－01　修回日期: 2025－12－24

基金项目: 国家自然科学基金项目(32301708)、国家重点研发计划项目(2023YFD2000200)、高等学校学科创新引智基地项目(D18019)和广东省科技创新战略专项资金项目(pdjh2025bc042)

作者简介: 兰玉彬(1961—),男,教授,博士,主要从事精准农业航空研究,E-mail: ylan@scau.edu.cn

通信作者: 黎源鸿(1990—),男,副教授,博士,主要从事具身智能采摘机器人关键技术研究,E-mail: liyuanhong@stu.scau.edu.cn

outperforming HER－TD3 and HER－DDPG baselines by over 12.5 mm on average. Orchard experiments further confirmed the effectiveness of the proposed method, achieving an average picking success rate of 84% and an average operation time of 10.58 s, validating its practical applicability in real-world citrus harvesting.

Key words: citrus picking; robotic arm; path planning; SAC algorithm; citrus detection; YOLO 11

0　引言

温室栽培改善了柑橘果实的生长环境，减少了病虫害的发生率[1]。随着人工智能与机器人技术的快速发展，机械臂在农业领域中发挥着越来越重要的作用，尤其是在自动化果实采摘方面[2]。近年来，目标检测算法取得了显著进展，即使在遮挡等复杂环境下，果实也能被较为准确地识别[3-4]。然而，尽管三维视觉技术可以提供果实的空间位置信息，机械臂在结构复杂的果园环境中，仍难以通过固定路径准确完成采摘任务。

高效的采摘过程需要最优的抓取姿态与无碰撞路径规划。在抓取姿态方面，KANG 等[5]利用霍夫变换与三维建模规划，将遮挡果实的采摘成功率从 61% 提升至 81%。MOHYUDDIN 等[6]通过三维点云模型实现了 88% 的抓取精度，但这两种方法均依赖精确的姿态估计，难以适应动态果园环境。在路径规划方面，传统避障算法快速探索随机树（RRT）已被广泛用于机械臂路径规划中[7]。鲍秀兰等[8]提出的改进型 RRT 算法在柑橘采摘中实现了 88% 的路径规划成功率，平均耗时为 10.5 s。SHANG 等[9]将基于五次多项式插值的 RRT 算法部署于番茄采摘机器人，虽取得了 88% 的成功率，但每次规划耗时达 20 s。尽管上述方法在果实抓取任务中表现良好，但其对预设环境模型依赖性强，且在高自由度系统中存在计算复杂度激增的问题[10]。

深度强化学习（Deep reinforcement learning, DRL）[11]具有无需显式环境模型和计算轻量化的特点，在解决抓取姿态与路径规划问题中展现出巨大潜力[12-14]。基于 Q-learning[15]、策略梯度（Policy gradient, PG）[16]和深度 Q 网络（Deep Q－network, DQN）[17]的理论，DRL 已被广泛应用于机器人运动控制[18-19]。LIN 等[20]提出的基于深度确定性策略梯度算法（Deep deterministic policy gradient, DDPG）结合循环神经网络，在 29 ms 内生成无碰撞路径，成功率达到 90.90%。LI 等[21]对 HER－SAC 算法进行了改进，用于番茄茎部抓取，抓取成功率提高了 14% 以上。

本文基于深度强化学习算法提出 ALH－SAC 算法，实现抓取姿态自适应与避障路径生成，避免传统方法需分别求解姿态与路径导致的高计算复杂度问题。ALH－SAC 以 HER－SAC 为结构基础，引入 LSTM[22]与注意力编码器[23]。其中，LSTM 用于捕捉机械臂关节动作的时间依赖性，增强策略对连续动作序列的建模能力；注意力机制则自适应地聚焦于复杂遮挡环境中的关键空间特征，从而显著提升策略在动态环境下的感知与决策精度。此外，本文将设计基于笛卡尔空间线性插值的混合路径引导机制，通过引入启发式先验策略，有效加速训练收敛。视觉感知与压力预测模块作为系统的辅助单元，分别用于果实检测与柔性夹持力调节，以保障采摘过程的稳定性与果实完整性。

1　路径规划方法

1.1　采摘机器人系统

该机器人用于低矮柑橘果实的采摘，目的是能够无碰撞地将果实从果树上分离。采摘系统由 1 个六轴机械臂、1 个夹持型末端执行器、2 个位于末端执行器内侧的压力传感器、1 个英特尔 D435i RGB－D 相机、1 台工业个人计算机等组成。

该 6－DOF 采摘机械臂安装在移动底盘中间，有效载荷为 1 kg，工作半径为 350 mm，机械臂基座的高度为 500 mm，采用德纳维特–哈滕伯格（D－H）方法开发的 6－DOF 机械臂的运动学模型如图 1 所

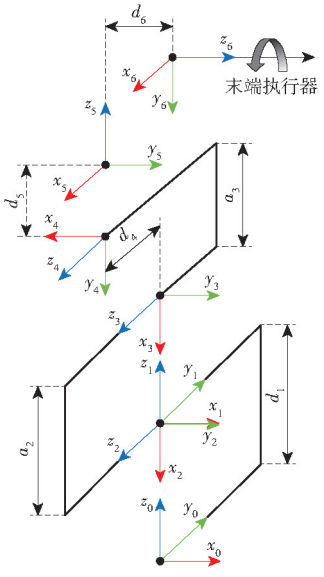


图 1　6－DOF 机械臂运动学模型

Fig. 1　Kinematic model of proposed 6－DOF robotic manipulator

示，$d_1=173.90$ mm，$a_2=135.00$ mm，$a_3=120.00$ mm，$d_4=88.78$ mm，$d_5=95.00$ mm，$d_6=65.50$ mm。该六轴机械臂的 D－H 变换矩阵为

$$\boldsymbol{A}_i=\begin{bmatrix}\cos\theta_i & -\sin\theta_i\cos\alpha_i & \sin\theta_i\sin\alpha_i & a_i\cos\theta_i\\ \sin\theta_i & \cos\theta_i\cos\alpha_i & -\cos\theta_i\sin\alpha_i & a_i\sin\theta_i\\ 0 & \sin\alpha_i & \cos\alpha_i & d_i\\ 0 & 0 & 0 & 1\end{bmatrix}\quad(i=1,2,\cdots,6)\tag{1}$$

式中 α_i——绕 x_i 轴的旋转角

θ_i——绕 z_{i-1} 轴的旋转角

RGB－D 相机安装在机械臂的末端执行器上，以捕捉彩色图像和对齐深度图像。RGB 图像的分辨率为 1 280 像素×720 像素。捕获的图像随后被 IPC 接收和处理，其中，果实的三维位置与姿态作为机械臂运动规划的目标，用作强化学习算法训练的输入数据，检测到的果实边界框和深度数据则与压力传感器数据共同用于柔性果实夹具控制网络设计。具体而言，在强化学习模型生成机械臂到达目标点的路径后，系统利用柔性果实夹具控制网络生成末端执行器的开合角度，共同提高采摘成功率。整体作业流程包括果实检测、运动规划、压力预测、果实分离等模块，具体流程如图 2 所示。

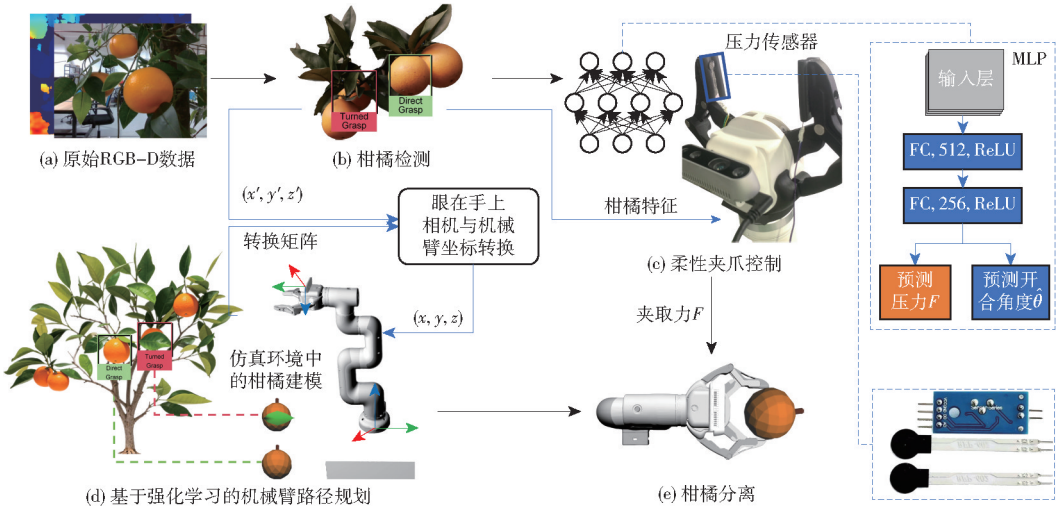


图 2 采摘机器人系统工作流程图

Fig. 2 Workflow diagram of picking robotic system

1.2 先验策略指导下的 ALH－SAC 算法

在机械臂面对复杂的果园环境时，最主要的工作就是避开障碍并准确到达目标，在此过程中，每一个关节转动的角度不能超过其机械限制，理论上，该机械臂工作时观测空间的约束条件可以描述为

$$\begin{cases}p,p_{target},p_{obstacle}\in U_{et}\\ \|\theta_i\|\in U_{\theta_i}\quad(i=1,2,\cdots,6)\\ if_direct_grasp\end{cases}\tag{2}$$

式中 U_{et}——可达空间三维坐标集合

U_{θ_i}——机械臂关节值范围

p——机械臂末端执行器位姿

p_{target}——目标水果位姿

$p_{obstacle}$——障碍物位姿

if_direct_grasp 表示检测到的果实种类。

在实际应用中，通常一个六自由度机械臂能够通过多种姿态来实现避障和到达目标，然而并非每一种姿态都符合物理合理性，其中一些姿态可能导致机械臂关节受力不均匀或机械装置损耗增加，同时，在机械臂进行目标探索时，存在无意义的探索空间。于是，在定义观测空间时，并非使用每个关节所能转动的角度阈值，而是通过先验策略来重新定义每个关节的转动角度，由此来避免一些不合理的姿态，具体细节将在 1.2.1 节详细阐述。在观测空间中，使用关节转动角度的正弦和余弦值来替代角度值，将角度信息映射到［－1，1］的范围内，在先验策略的前提下，这种映射被证明更加高效。这为提出的算法提供了更平滑和连续的状态表示，有助于减少状态空间的复杂性，使得算法能够更快速地收敛到合适的策略。于是，在此基础上，将观测空间定义为

$$s=(p,p_{target},p_{obstacle},\cos\theta_i,\sin\theta_i,if_direct_grasp)\tag{3}$$

式中 s——观测空间

考虑到观测空间的高维度特性和目标的随机性，选择最大化熵的 SAC 算法[24]进行改进，提出 ALH－SAC 算法，算法整体架构如图 3 所示。

1.2.1 具有先验知识的混合策略

由于机械臂的路径规划问题同时考虑了末端执行器的位置和方向，可探索空间太大，通过引入特定问题的先验知识，可以将可探索的空间限制在最优

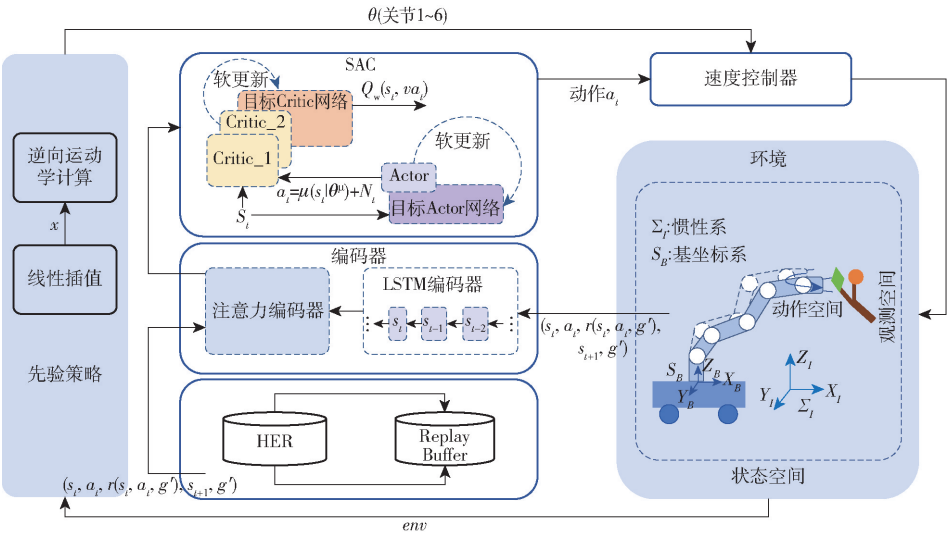


图 3　基于 ALH－SAC 的机械臂路径规划架构

Fig. 3　Robotic arm path planning architecture based on ALH－SAC

策略的附近，从而提高了强化学习的样本效率。具有先验知识的混合策略[25]可以定义为

$$u_k(s)=\omega u_\theta(s)+(1-\omega)u_{prior}(s) \tag{4}$$

式中　ω——权重

u_θ——强化学习策略

u_{prior}——先验策略

u_k——混合策略

使用混合策略相当于在 RL 策略 u_{θ_k}上放置一个功能性正则化器的上游。

图 4 表明，在训练过程中，混合策略和最优策略之间的偏差受到先验偏差的限制。与无限更新相比，混合策略使早期的探索更加高效，因此，利用笛卡尔空间中的线性插值作为规划算法的先验策略。相比于使用经典的全局路径规划算法如 RRT、A* 等引入较强的路径先验，线性插值能够提供低复杂度的方向性引导，使强化学习模型仍可基于环境反馈自主调整动作输出，从而保留充分的探索能力。首先从机械臂末端执行器到目标点进行线性插值，然后利用逆运动学方法计算两个相邻插值点之间的关节角误差，得到所期望的关节角度。然而当直接在线性插值点上利用逆运动学时，机械臂很可能通过奇异姿态接近目标或达到关节极限。因此在进行线性插值时，只考虑末端执行器在笛卡尔空间中的位置，在插值位置上可以选择任何非奇异方向来计算关节角误差。

1.2.2　ALH－SAC 算法

本研究提出的 ALH－SAC 算法在 SAC 框架上引入了 Attention Encoder、LSTM Encoder 以及 HER[26] 3 个模块，以同时提升策略的状态表达能力、时序建模能力以及稀疏奖励场景下的学习效率。

SAC 在目标函数中引入了熵正则化项以鼓励探索，这一特性在高维或稀疏奖励环境中尤为有益，其最优策略需同时最大化预期累积奖励和策略熵值，表达式为

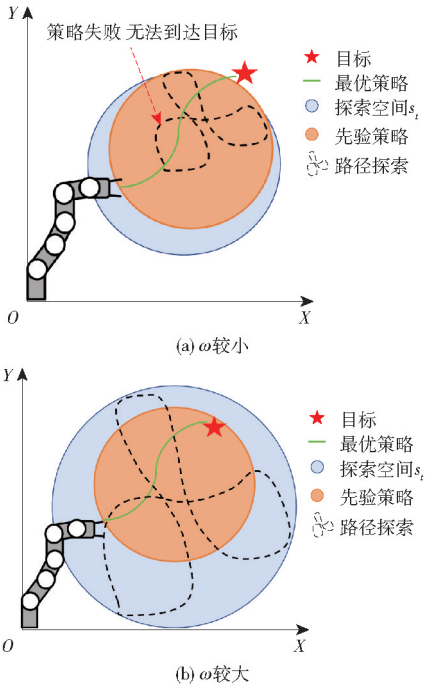


图 4　先验策略在可探索空间中的应用示意图

Fig. 4　Illustration of using prior policy in explorable space

$$\pi^*=\arg\max_\pi \sum_t E_{(s_t,a_t)\sim\rho_\pi}[r(s_t,a_t)+\alpha H(\pi(\cdot|s_t))] \tag{5}$$

式中　$r(s_t,a_t)$——当代理在状态s_t下执行动作a_t时获得的奖励

π^*——最优策略

π——候选策略

$E_{(s_t,a_t)\sim\rho_\pi}$——数学期望算子

ρ_π——由 π 引导的状态和动作的联合分布

α——温度权重

H——熵值

HER 是由 ANDRYCHOWICZ 等[26]提出的一种强化学习经验增强方法,旨在解决稀疏奖励场景中学习效率低的问题。在传统 off-policy 强化学习中,每次交互得到的元组为$(s_t, a_t, r_t, s_{t+1}, g)$,其中 g 是当前回合中的目标。在稀疏奖励情形下,只有当代理精确达到目标 g 时,才能获得正奖励,这使得探索极其困难。HER 的核心思想是:即使最终没有完成原始目标 g,也可以将实际达到的某个状态 $g' = s_t'$当作一个“虚拟目标”来重新计算奖励。于是,HER 将原始轨迹中的每一个状态转移,额外转换为$(s_t, a_t, r(s_t, a_t, g'), s_{t+1}, g')$,由此大大提高了强化学习的训练效率。

在上述 HER - SAC 的基础上,ALH - SAC 整体结构仍保持 SAC 的 Actor - Critic 框架,但在输入端增加两级编码,使策略网络能更准确地从复杂观测中提取有效信息。首先,Attention 编码器对输入的机械臂状态与目标向量进行特征重加权,使网络能够自动关注与果实位姿相关的分量。随后,编码后的特征进入 LSTM 编码器,用于捕获机械臂运动过程中的时序相关性。在每个时刻 t,注意力编码器对当前状态s_t与目标位姿p_{target}进行联合编码,得到目标相关的状态特征

$$z_t^{att} = \mathrm{softmax}(\mathrm{ReLU}(\boldsymbol{\omega}_s \boldsymbol{s}_t + \boldsymbol{b}_s, \boldsymbol{\omega}_p \boldsymbol{p}_{target} + \boldsymbol{b}_p)) \odot (\boldsymbol{\omega}_s \boldsymbol{s}_t + \boldsymbol{b}_s) \tag{6}$$

式中 $\boldsymbol{\omega}_s$、$\boldsymbol{\omega}_p$——可学习权重

$\boldsymbol{b}_s$、$\boldsymbol{b}_p$——偏置向量

z_t^{att}——注意力特征

首先,将原始状态向量$\boldsymbol{s}_t$和目标向量$\boldsymbol{p}_{target}$通过全连接层映射到相同的特征维度空间,并利用 softmax 函数对状态特征进行加权,计算每个状态分量的注意力分数,将注意力分数作用于原始状态特征,得到目标导向更强的编码特征z_t^{att}。对于状态中的离散变量 *if_direct_grasp*,在编码前进行归一化处理,以保证连续特征与离散特征在神经网络中统一表示。

随后,LSTM 编码器对相邻时刻的状态序列进行建模,以提取机械臂动作与末端位姿变化的动态相关性,表达式为

$$z_t^{lstm} = f_{lstm}(s_t, s_{t-1}) \tag{7}$$

式中 f_{lstm}——标准 LSTM 递归计算过程

z_t^{lstm}——时序特征

输出的z_t^{lstm}与目标相关特征z_t^{att}拼接后,构成融合的高维状态s_t' 替代原始环境状态s_t作为 Actor 与 Critic 网络的输入,Actor 网络接收 s_t' 并输出动作向量$\boldsymbol{a}_t$,Critic 网络同样使用s_t' 作为输入,对$\boldsymbol{a}_t$进行价值评估。同时,训练阶段继续结合 HER 进行目标重标注,从失败轨迹中构造额外的成功样本,以显著缓解采摘任务中稀疏奖励带来的训练困难。先验策略下的 ALH - SAC 算法代码为:

初始化$\boldsymbol{\pi}_\phi(a_t|s_t)$、$Q_{\theta_i}(s_t, a_t)$ 和目标网络
初始化经验回放缓冲区 D
for episode $m = 1$ to M do
  初始化环境并获取初始状态 s_0
  从目标空间 G 生成随机目标p_{target}
  生成线性插值轨迹 $X = \{X_0, \cdots, X_n\}$
  for step $t = 0$ to $T - 1$ do
    通过逆运动学从X_e计算θ_e
    计算角度差 $\Delta\theta = \theta_e - \theta_t$
    从 $\Delta\theta$ 推导先验关节角度θ_{prior}
    if not *if_direct_grasp* then
      compute_downward_heuristic_velocity()
        has_redirected = True
    end if
    从策略 $\boldsymbol{\pi}_\phi(a_t|s_t)$ 采样动作$\boldsymbol{a}_t$
    状态特征 = Attention_encoder(s_t, p_{target})
    时序特征 = LSTM_encoder(s_t, s_{t-1})
    更新历史状态
    基于时序特征从策略中采样动作
      计算最终混合动作 $a_t' = \mathrm{mix}(a_t, \theta_{prior})$
    将动作 a_t'应用到环境
    获取新状态 s_{t+1}和奖励 r_t
    存储 $(s_t, a_t, r_t, s_{t+1}, p_{target})$ 到重放缓冲区 D
    采样额外目标 g'
    for each p_{target}' do
      计算 hindsight reward $r_t' = r(s_t, a_t, p_{target}')$
      存储 $(s_t, a_t, r(s_t, a_t, g'), s_{t+1}, p_{target}')$ 到 D
    end for
  end for
  for gradient step $n = 1$ to N do
    从 D 采样 mini-batch
    更新 $Q_{\theta_1}, Q_{\theta_2}$
    更新 $\boldsymbol{\pi}_\phi$
    使用 soft target update 更新目标网络
  end for
end for

1.2.3 奖励函数设计

基于机械臂末端执行器与目标中心之间的欧几里得距离误差,所构建的非线性奖励函数r_{target}定义为

$$r_{target}=\begin{cases}-\dfrac{1}{2}|d_t|^2 & (|d_t|\leqslant\phi)\\ -\phi\left(|d_t|-\dfrac{1}{2}\phi\right) & (|d_t|>\phi)\end{cases} \tag{8}$$

式中　ϕ——常量

由于机械臂末端执行器的位置变化范围要比方向变化范围小，在训练过程中能够进行有效的探索。然而，末端执行器的方向变化范围为 $[-\pi,\pi]$，这个范围使得末端执行器在方向上的探索效率较低，如果直接在 3 个方向上分别衡量误差，无法有效引导网络收敛。为提高方向空间上的探索效率，仅考虑三轴误差上的最大值来计算奖励。这种设计将三维方向误差简化为单一的标量值，避免了在三维方向误差上逐轴优化的复杂性，使得神经网络能够更有效地学习到使方向误差最小化的策略。于是方向奖励函数 $r_{orientation}$ 仅考虑 3 个姿态误差中的最大值

$$r_{orientation}=\cos(\max(\varepsilon_r,\varepsilon_p,\varepsilon_y)) \tag{9}$$

式中　ε_r、ε_p、ε_y——翻滚、俯仰和偏航的方向误差

碰撞惩罚被定义为一个分段函数 $r_{obstacle}$，即

$$r_{obstacle}=\begin{cases}-\kappa e^{-\eta|d_o|} & (|d_o|<|d_{threshold}|)\\ \delta & (|d_o|\geqslant|d_{threshold}|)\end{cases} \tag{10}$$

式中　$|d_o|$——末端执行器与障碍物之间的距离

κ、η、δ——碰撞参数　　$d_{threshold}$——阈值

阈值 $|d_{threshold}|$ 是通过在实际应用场景中进行测试后凭经验确定的，碰撞检测仅在以目标为中心的扇形区域内进行。将机械臂的操作性函数定义为优化操作性[27]的奖励函数 $r_{manipulability}$，即

$$r_{manipulability}=\sqrt{\det(\boldsymbol{J}(\theta_i)\boldsymbol{J}^{\mathrm{T}}(\theta_i))} \tag{11}$$

式中　$\boldsymbol{J}$——对应的雅各比矩阵

最终奖励函数为加权求和，即

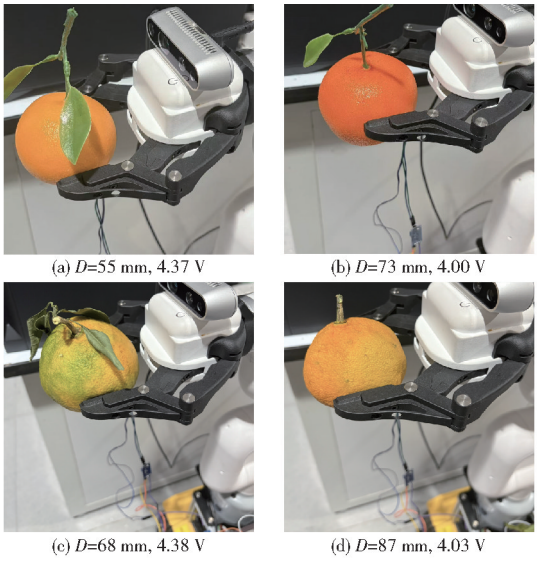
$$r(s,a)=\lambda_1 r_{target}+\lambda_2 r_{orientation}+\lambda_3 r_{obstacle}+\lambda_4 r_{manipulability} \tag{12}$$

式中　λ_1、λ_2、λ_3、λ_4——常数缩放因子，各类奖励函数在训练过程中的权重

1.3　基于深度相机与压力传感器的柔性夹持器控制网络

在机器人采摘作业前，研究团队从柑橘园实景及室内搭建的仿真果园中共采集了 2 000 幅 RGB－D 图像，用于分析果实的遮挡情况与深度信息。基于 YOLO 11 算法，根据果实的遮挡程度（小于 40%）及其深度对比，将检测到的果实划分为可直接抓取型与需调整姿态型 2 类（如图 2b 所示）。在该数据集上，YOLO 11 对可直接抓取型果实的平均检测精度达 90.5%，对需调整姿态型果实的平均检测精度为 87.7%，能够为后续采摘策略的决策提供可靠的视觉输入。

为模拟果园真实环境，搭建了具备果实位置可调功能的半自然平台。在末端夹爪上安装薄膜压力传感器，并通过已知重力进行标定，建立电压与夹持力的映射关系。采集 300 组抓取样本（包含检测框、深度信息、夹爪控制角、压力、果实直径、图像特征及成功标签），研究发现不同大小和硬度的柑橘在相同控制角下夹爪施加的压力存在差异，这里的“控制角”并非夹爪的物理开合角，而是控制系统输入的数值（通常在 30～80 之间），用于调节夹爪开合。当夹爪闭合到能够稳固夹紧柑橘时，夹爪施加的力主要表现为压力方向的正向力。为了防止柑橘沿重力方向滑落，夹爪的压力通过与果实接触面的摩擦力提供支撑。因此，质量较大的柑橘需要更大的摩擦力才能稳定夹持，从而夹爪施加的压力也随之增大；不同硬度的柑橘因接触面的摩擦因数不同，也会导致压力传感器读数（电压）出现差异，如图 5 所示。为此开发了双分支神经网络结构，如图 2c 所示，该网络以果实特征与直径为输入，可同步预测最优夹持力与自适应抓取控制参数。



(a) D=55 mm, 4.37 V　(b) D=73 mm, 4.00 V　(c) D=68 mm, 4.38 V　(d) D=87 mm, 4.03 V

图 5　不同直径 D 和粗糙程度柑橘在夹爪夹紧下压力传感器读数示例

Fig. 5　Pressure sensor readings sample of citrus fruits with varying diameters and surface roughness during gripper grasping

2　试验结果与分析

2.1　采摘机械臂仿真试验

2.1.1　仿真训练试验设置

进行仿真试验来评估 ALH－SAC 的机械臂姿态控制性能，本研究中所有识别过程、模拟和训练都是在 Ubuntu 22.04 上使用 Python 3.7 和 PyTorch 框架进行的，运行在一台配备英特尔 i7－10700 K CPU、32 GB 内存和 Nvidia GeForce RTX 3060 GPU 的计算机上。

仿真环境主要由基地平台、操纵器和柑橘果实模型 3 个对象组成,如图 6 所示。该机械臂的设计使用了 YahBoom 官方的 myCobot 320 M5 机械臂 URDF 模型,搭载了一个带有夹具的末端执行器。柑橘果实模型分为可直接采摘和转向采摘 2 个类别,并利用几何先验对该模型的位置和姿态进行重构,在该模型中,果实的三维坐标、类别和果实的生长位姿均设置为变量。

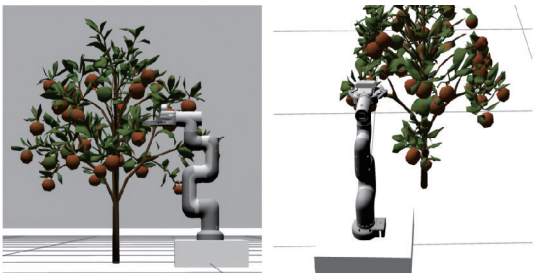


图 6 Gazebo 中的仿真采摘环境

Fig. 6 Simulated picking environment in Gazebo

在本研究中,对 ALH - SAC 进行了消融试验,对比测试了 3 种为连续动作空间设计的最先进的 DRL 算法:DDPG[28]、TD3[29]、SAC。所有的算法都遵循相同的模拟环境设置和数据集。采用其默认的初始化参数和权值,以保证算法的最优性能。

ALH - SAC 的关键训练参数见表 1。在仿真环境中,将机械臂的初始关节角度(θ_1,θ_2,θ_3,θ_4,θ_5,θ_6)设置为 (0,π/2,-π/2,0,0,0),将柑橘模型随机放在 300 mm×300 mm×300 mm 的范围内。在评估阶段,分析了奖励函数曲线和成功率曲线来评估模型是否已经收敛。成功的判定标准为:①末端执行器参考系原点与抓取点的欧氏距离小于 0.03 m,即$d_{target} = |p - p_{target}| < 0.03$ m。②末端执行器姿态与果实最佳收获姿态的四元数误差小于 0.05 rad,即$q_{diff} = q \otimes q_{target}^{-1} < 0.05$ rad,果实的最佳收获姿态参考文献[19]中的姿态约束方法。③在机械臂的运动过程中没有发生碰撞。

表 1 ALH - SAC 的关键参数

Tab. 1 Key parameters of ALH - SAC

参数	数值
学习率λ_π、λ_Q、λ	0.000 3
集数 M	10^4
折扣因子 γ	0.99
目标平滑系数 τ	0.005
每集最大时间步 T	50
小批量大小 N	256
缓冲区大小	10^6
额外目标数(HER)	8
目标熵	-6
目标奖励参数 φ	0.05
惩罚强度 κ	3.0
指数衰减系数 η	50
碰撞参数 δ	100
碰撞阈值$d_{threshold}$	0.02
奖励加权因子λ_1	20
奖励加权因子λ_2	1.0
奖励加权因子λ_3	1.2
奖励加权因子λ_4	0.9

此外,考虑使用末端执行器与目标的三维坐标误差及姿态误差,结合 LINDNER 等[30]提出的指标来评估 DRL 算法的性能。在表 2 中,给出了 100 个测试柑橘模型的结果,该测试的 100 个柑橘模型均为可直接采摘类别。比较的参数为末端执行器与目标果实的三维坐标误差 e_x、e_y、e_z及其标准差σ_x、σ_y、σ_z;末端执行器与目标果实的姿态坐标误差 θ_r、θ_p、θ_y及其标准差 σ_{θ_r}、σ_{θ_p}、σ_{θ_y};末端执行器参考系原点到抓取位置的距离误差 e,误差 e 的标准差 σ_e;机械臂轨迹距离起始位置到目标位置的最短路径的增量距离 d_s,距离 d_s的标准差 σ_s;实现目标所需的步数 n_s,步数的标准差 σ_n。当满足上述决策成功条件或机械臂姿态不再有任何变化时,计数步骤即完成。

表 2 6 种不同模型在测试集上的仿真结果

Tab. 2 Simulation results of six distinct models on test set

评价指标	ALH - SAC	HER - DDPG	HER - TD3	SAC	DDPG	TD3
$e \pm \sigma_e$/mm	10.35 ±4.18	31.50 ±12.90	22.85 ±10.71	19.70 ±9.82	64.30 ±28.58	70.80 ±20.48
$e_x \pm \sigma_x$/mm	8.15 ±4.28	16.70 ±2.82	15.20 ±9.12	8.95 ±11.37	39.30 ±35.58	43.10 ±25.86
$e_y \pm \sigma_y$/mm	7.30 ±3.80	18.60 ±10.12	14.85 ±8.91	9.10 ±10.50	42.20 ±37.32	46.40 ±27.84
$e_z \pm \sigma_z$/mm	5.05 ±2.53	11.80 ±6.49	8.75 ±5.25	16.35 ±9.81	27.30 ±28.38	35.90 ±21.54
$\theta_r \pm \sigma_{\theta_r}$/rad	0.048 ±0.022	0.49 ±0.025	0.038 ±0.020	0.051 ±0.027	0.35 ±0.18	0.27 ±0.14
$\theta_p \pm \sigma_{\theta_p}$/rad	0.041 ±0.029	0.52 ±0.025	0.046 ±0.030	0.076 ±0.030	0.39 ±0.23	0.30 ±0.27
$\theta_y \pm \sigma_{\theta_y}$/rad	0.038 ±0.020	0.39 ±0.020	0.027 ±0.013	0.062 ±0.029	0.28 ±0.14	0.22 ±0.09
$d_s \pm \sigma_d$/m	0.05 ±0.01	0.20 ±0.04	0.07 ±0.02	0.06 ±0.04	0.14 ±0.05	0.09 ±0.04
$n_s \pm \sigma_n$	8.54 ±1.35	30.78 ±2.92	13.23 ±2.57	7.38 ±1.70	20.73 ±5.87	16.35 ±5.68
成功率/%	90	20	54	72	43	40

2.1.2　仿真试验结果

图 7 对比展示了使用先验策略对算法学习阶段的影响，结果表明，先验策略对算法学习到最优策略具有指导作用，通过引入先验策略，算法在初始阶段获得了更优的起点。同时，先验策略加速了算法的初期探索过程，并提高了最终策略的性能。具有 HER 的算法强调了先验策略的作用，HER－SAC 在先验策略的指导下相比于随机探索的 HER－SAC 收敛速度提升了 33%，最终成功率提升了 5 个百分点。由此可见，优先经验回放机制通过强调关键经验的重放，进一步强化了先验策略的指导作用。

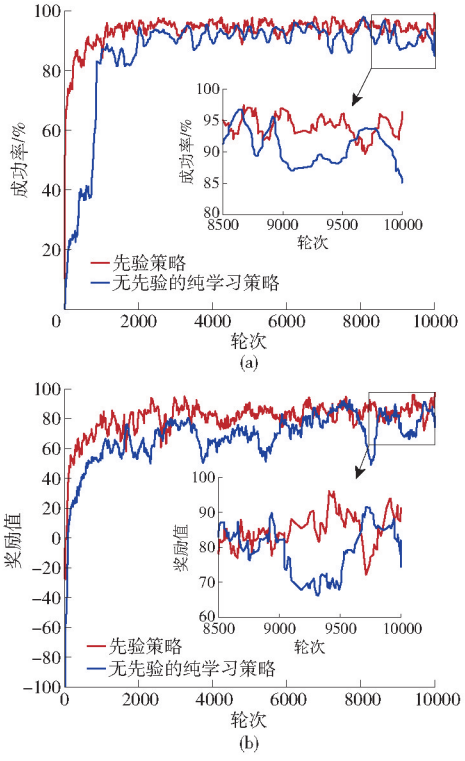


图 7　HER－SAC 在先验策略与纯学习策略下的学习曲线

Fig. 7　Learning curves of HER－SAC under prior policy and pure learning policy

进一步研究了公式(4)中先验策略权重 ω 的设置对算法的影响，通过分别对比固定引导权重($\omega=0.5$)、线性衰减权重 $\left(\omega=\omega_0\max\left(1-\frac{n_{step}}{n_{stepsmax}},0\right)\right)$ 以及指数衰减权重($\omega=\omega_0 e^{-kn_{step}}$) 3 种策略发现：固定权重在整个训练过程中提供持续引导，有利于提升最终策略的稳定性；线性衰减策略在保持初期高引导效果的同时，逐步释放学习自由度，使策略能够充分优化；指数衰减策略则更强调初期引导与后期自主学习的分离，更适用于在初期迅速逼近目标区域。试验结果表明，动态衰减策略相较于固定引导具有更好的训练效果。其中，采用线性衰减权重的 SAC 和 ALH－SAC 算法，最终在平均成功率上优于固定权重设定。

图 8 展示了不同先验权重下在训练 500 次随机采样的 50 个末端执行器位置分布轨迹，可以看出，较大的初始权重值使末端轨迹在初期更快速向目标集中分布，而纯学习策略在初期的运动轨迹则相对分散，存在较多无效尝试。这说明先验策略在训练早期对动作空间的有效约束，能够提升探索效率，加快学习过程。逐步减小的引导权重则使策略能够过渡到自主控制阶段，直观地验证了先验策略在强化学习训练中能提供方向性指导。

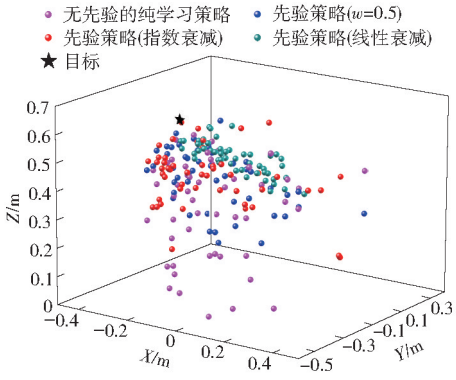


图 8　训练 500 回合先验策略与纯学习策略下末端执行器位置分布

Fig. 8　Distribution of end-effector positions during initial 500 training episodes under prior-guided strategy and pure learning strategy

为评估无碰撞轨迹生成性能，对 6 种算法在模拟柑橘采摘环境中进行了测试。图 9 为在同一模拟环境中 6 个不同模型的学习曲线：ALH－SAC、HER－DDPG、HER－TD3、SAC、DDPG 和 TD3 在学习阶段的表现，ALH－SAC 算法实现了最快的收敛速度和最佳最终性能，其稳定速度比 HER－TD3 快约 75%，在平均奖励和成功率上均超越所有基线算法。

表 2 展示了训练完成的路径规划算法在采摘仿真环境中针对 100 个无遮挡柑橘模型的测试结果。从结果可以看出，提出的 ALH－SAC 算法采摘成功率达 90%，相比其他算法至少提升了 18 个百分点；同时，其末端执行器与目标之间的平均距离误差仅为 (10.35 ± 4.18) mm，相较其他算法至少提高了 9.35 mm 的精度，表明该方法在引导机械臂末端准确接近目标方面具有显著优势。在对比算法中，DDPG 和 TD3 在高维动作空间中的训练过程中出现了收敛失败的现象，验证了其在默认训练参数设置下难以有效完成目标接近任务。SAC 虽在某些测试中表现良好，但存在临界成功与超时并存的情况，部分测试的距离误差高达 31.52 mm。

为了更直观地展示出算法在接近目标准确度和效率方面的性能，选择位于机械臂左下方与右上方的目标 1(0.253，－0.272，0.279) m 和目标 2(－0.208，

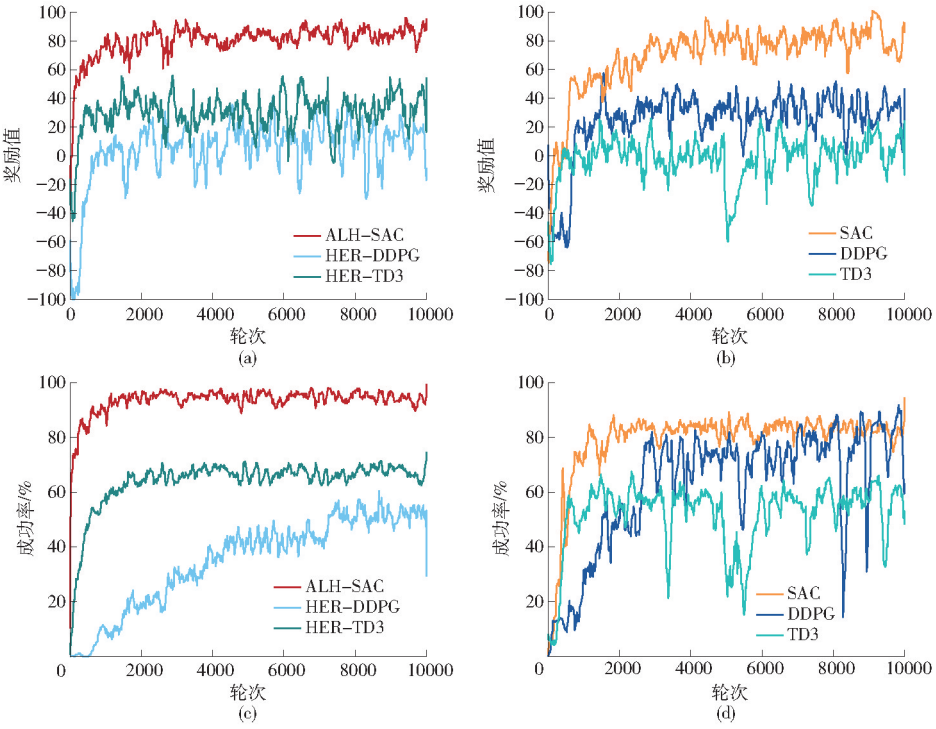


图 9 不同 DRL 算法在机械臂姿态控制任务训练过程中的学习曲线

Fig. 9 Learning curves of training process for manipulator posture control task which different DRL algorithms

-0.108,0.508) m 绘制了测试过程中末端执行器移动轨迹图,如图 10 所示。从轨迹图可以看出,在两个目标点下,ALH - SAC 算法能够稳定地产生收敛轨迹,且最终误差在两个目标下均为最小(目标 1 为 0.007 m,目标 2 为 0.001 m)。在分维度误差分析上,图 11a ~ 11d(目标 1)和图 11e ~ 11h(目标 2)显示了各算法在 X、Y、Z 3 个方向上的定位精度。ALH - SAC 在所有维度上的误差下降速度较快,在 Y 轴方向上表现尤为稳定,几乎在第 6 步后保持在 0.01 m 以下。而 HER - DDPG 和 HER - TD3 虽然整体趋势也在收敛,但在 Y 轴和 Z 轴上波动相对较大,特别是在图 11c 中,HER - TD3 在第 9、10 步误差出现反弹。Z 轴误差方面,ALH - SAC 在两个目标上都能快速将误差控制在 ±0.01 m 内,说明其在高度控制上的精度也优于其他算法。

2.2 采摘机器人样机试验

2.2.1 样机采摘试验设置

在进行真实采摘系统对比试验之前,为验证所提出的夹具压力预测网络在实际应用中的预测精度,本文首先对压力预测模块进行了独立评估试验。

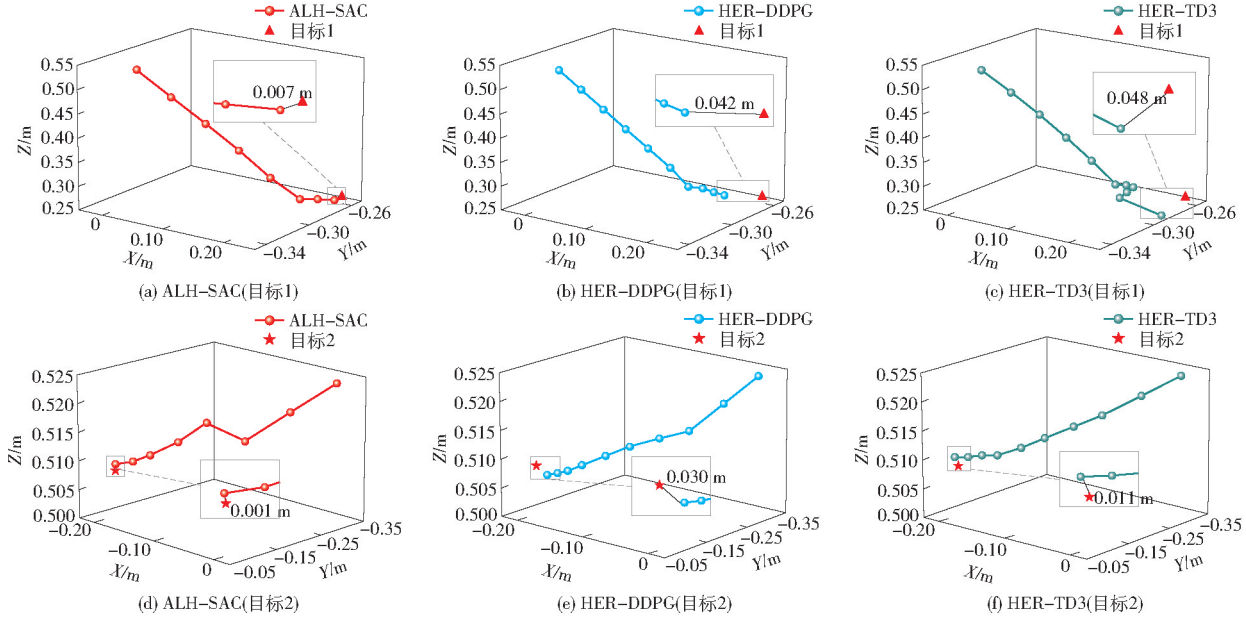


图 10 不同算法采摘过程中末端执行器在工作空间路径上的位置

Fig. 10 Position of end-effector on workspace path during picking process based on different algorithms

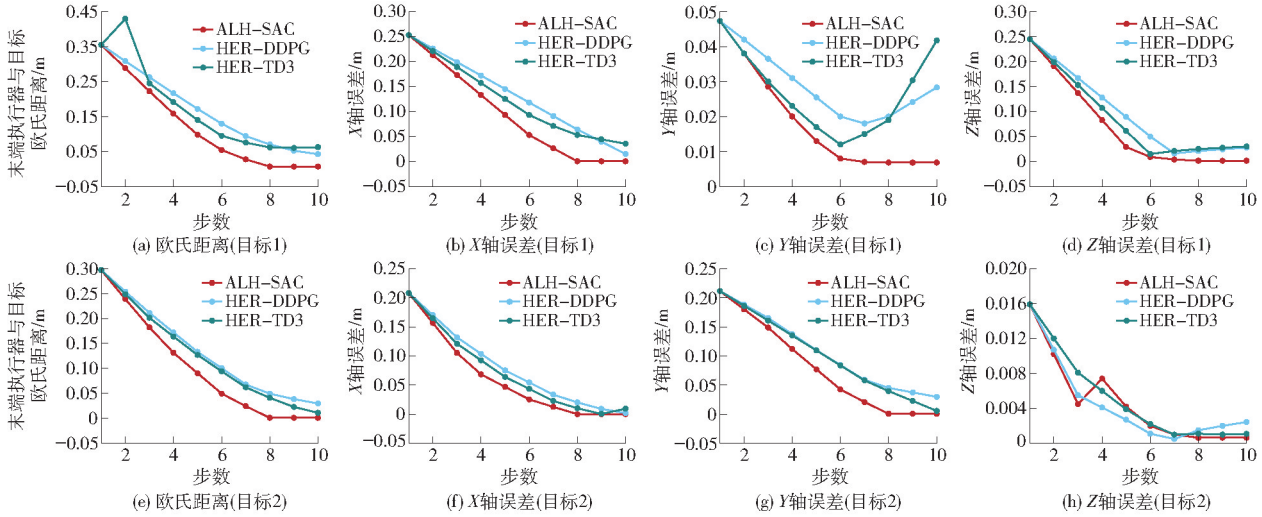


图 11　ALH－SAC、HER－DDPG 和 HER－TD3 算法在两个目标上的欧几里得距离与轴向距离误差变化曲线

Fig. 11　Euclidean and axis-wise distance errors of ALH－SAC, HER－DDPG and HER－TD3 on two goals

试验过程中，使用不同大小和种类的柑橘随机采集了 50 组夹取过程数据，每组数据均包含：网络预测的夹取压力和通过压力传感器实时测得的真实压力，网络预测的夹具控制角度和正好夹紧时的夹具角度。

为验证所提出的 ALH－SAC 算法及压力预测网络在真实环境中的采摘性能，在上述试验基础上，设计了 3 组对比试验，使用机器人采摘样机结合不同的路径规划方法和夹具控制策略，评估机械臂在实际采摘任务中的表现。试验共采用两种路径规划方法，分别为 ALH－SAC 算法和传统的双向快速随机树（Bidirectional RRT）算法；同时结合两种夹具开合控制方式进行配套测试，分别为：固定张开角度（依据柑橘的平均尺寸预设一个固定值）和自适应张开角度控制（根据视觉感知的自适应夹取力控制网络动态调整夹爪张开角度）。这两种路径规划方法与两种夹具控制策略组合，共构成 3 种试验设置：Bidirectional RRT 算法＋固定夹具角度控制；ALH－SAC 算法＋固定夹具角度控制；ALH－SAC 算法＋自适应夹具角度控制。在每组试验中，机械臂系统执行多次真实采摘任务，记录总尝试次数与成功夹取次数，从而计算成功率。所有采摘任务均在室内仿真柑橘树上进行，果实具有不同的空间位置和一定程度的枝叶遮挡，其三维位姿由 RGB－D 视觉系统识别获得，末端执行器根据路径规划结果接近果实并完成夹取操作。

2.2.2　样机采摘试验结果

图 12 展示了 ALH－SAC 策略控制下机械臂执行柑橘采摘的过程。根据策略输出，机械臂沿计算轨迹向目标果实移动，同时末端执行器根据策略调整张开角度。当达到预估抓取力时，机械臂将果实从枝条分离并返回初始位姿。



图 12　实验室测试过程（在模拟棚架结构上对柑橘进行采摘）

Fig. 12　Laboratory test procedure for citrus picking on simulated trellis structure

图 13 给出了预测压力与真实测量压力以及预测角度与真实控制角度之间的对比散点图。可以观察到，大部分数据点分布在 $y = x$ 参考线附近，表明预测结果与真实压力具有较高的一致性，未出现明显系统性偏差。

表 3 为上述自主采摘过程的试验结果。通过对比不同试验配置下机器人系统的成功率和平均操作时间，对系统性能进行了系统评估。其中采摘成功率定义为成功采收果簇数占尝试总数的比例，平均操作时间涵盖从机械臂初始定位到完成采摘的全程，包括检测、运动序列生成与执行。

在试验 1 中，采用 Bidirectional RRT 方法取得

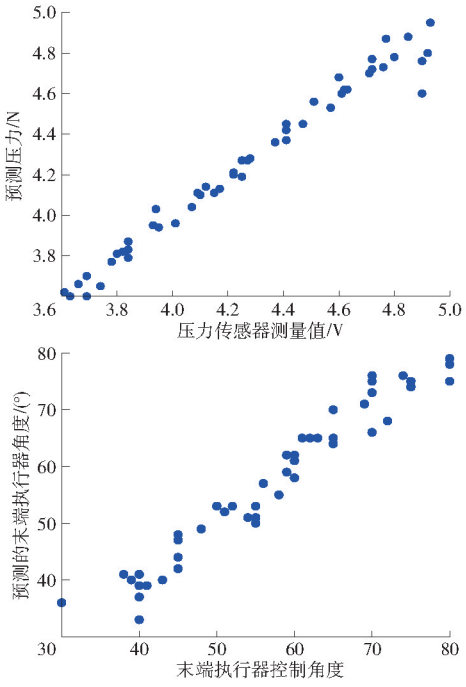


图 13 预测值与真实值的对比散点图

Fig. 13 Scatter plots comparing predicted values with ground-truth measurements

了 52% 的成功率,该较低数值表明该方法在具有遮挡的复杂果园环境中生成稳定无碰撞轨迹的能力有限,虽然 Bidirectional RRT 在结构化场景中可能有效,但其对预设模型的严重依赖导致适应动态环境变化的能力不足。此外,固定角度夹爪的使用削弱了对不同尺寸果实的适应性,进一步降低了整体成功率。

表 3 不同算法的采摘系统试验结果

Tab. 3 Experimental results of picking systems with different algorithms

参数	试验 1(Bidirectional RRT + 固定开合角度)	试验 2(ALH − SAC + 固定开合角度)	试验 3(ALH − SAC + 自适应角度控制)
总次数	100	100	100
成功次数	52	78	84
成功率/%	52	78	84
平均操作时间/s	9.51	9.74	10.58

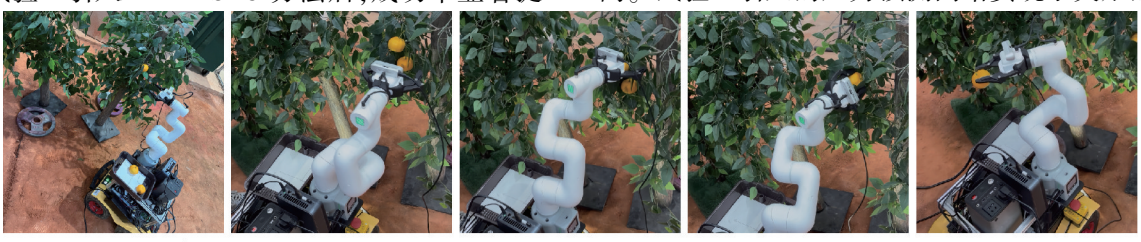
试验 2 引入 ALH − SAC 方法后,成功率显著提升至 78%。该策略通过经验回放与目标重标注机制增强机械臂路径规划的适应性,即使在部分遮挡或末端执行器朝向受限情况下仍能生成有效采摘轨迹。但由于继续采用固定夹爪角度,抓取误差仍是主要失败原因,特别是在处理小型果实或非标准朝向果梗时,抓取角度偏差可能导致果实滑脱或损伤。试验 3 将 ALH − SAC 方法与基于实时果实信息的自适应夹爪控制策略相结合,成功率显著提升至 84%,为所有试验中的最高值,印证了感知驱动抓取控制在非结构化环境中的实用价值。

虽然试验 3 的平均操作时间略增至 10.58 s,较试验 2 增加了 0.84 s,但结合 16 个百分点的成功率提升,该时间成本具有合理性,尤其对于高价值经济作物而言。相较于 XIAO 等[31] 在室内无遮挡柑橘采摘场景中 87.15% 的成功率,本方法在遮挡环境中仅低 3.15 个百分点,凸显了其在真实环境中处理遮挡与避障问题的鲁棒性。

图 14 展示了基于 ALH − SAC 策略的样机采摘路径规划全过程,机械臂首先移动至果树区域并识别目标果实,随后规划并执行路径,在接触果实后,基于预设的力控策略将其分离,最后规划完成并将果实放入果篮。

2.2.3 失败情况及分析

结合表 4 对现实场景中的机器人采摘失败案例进行分析。首先,当果实位于工作空间 Y 轴远端且 Z 轴较高位置时易发生采摘失败,此类位姿要求机械臂执行大幅度的前伸与上移动作,超出其机械极限。为避免训练过程中出现奇异位姿,工作空间边界区域的探索受到限制。在茂密果园环境中,末端执行器在运动过程中与枝条的碰撞可能导致果实移动或脱离摄像视野,从而引发抓取失败。传统规划器(如 Bi − RRT)常采用自上而下的接近方式,增加了碰撞风险。相比之下,试验 2 和试验 3 采用的 ALH − SAC 方法通过自下而上的接近策略,能更好规避遮挡并提升成功率。夹爪控制是另一关键问题。试验 1 和试验 2 采用固定开口度的夹爪导致适应性不足:大型果实易受损,小型果实则难以有效分离。试验 3 引入的压力预测网络实现了夹爪的自适



(a) 底盘定位到达果树 (b) 初始化位姿,柑橘识别 (c) 路径规划到达目标 (d) 依据预测力夹取果实 (e) 放入果篮,规划完成

图 14 ALH − SAC 策略下样机试验采摘路径规划过程

Fig. 14 ALH − SAC-based harvesting path planning in physical robot tests

表 4　不同试验设置的失败原因及出现次数

Tab. 4　Failure causes and frequencies on different experiment setups

失败原因	试验 1	试验 2	试验 3
因工作区边界约束而不可达	16	8	8
末端执行器与树枝碰撞	22	4	4
固定夹持角无法适应果实大小变化	6	6	0
相机检测错误导致的目标定位失败	4	4	4

应调节，提升了对不同尺寸果实的处理能力。视觉误差同样导致失败案例，特别是当果实出现在图像边缘区域时。深度精度不足及遮挡问题会引起定位偏差和抓取错误。后续工作将采用数据增强和聚焦边缘的采样策略以改进图像边界附近的检测精度。

3　结论

（1）本研究提出一种基于 SAC 算法的深度强化学习算法 ALH－SAC，引导六自由度机械臂以无碰撞的方式到达柑橘果实的抓取位置，并实现柔性力度控制以防止果实在采摘过程中损坏和掉落。通过在算法中引入注意力编码器、LSTM 编码器和 HER，并结合先验策略和姿态约束的奖励函数，在模拟温室果园场景的仿真环境进行强化学习训练，并在搭建的温室果园中验证该算法提升了采摘精度。

（2）所提方法首先在仿真平台上进行了广泛测试。消融试验表明，引入路径引导先验策略显著加速了早期策略收敛，并提升了末端定位精度。在随机目标测试中，ALH－SAC 的成功率达到 90%，较 HER－TD3（54%）和 HER－DDPG（20%）至少高出 36 个百分点。在关键指标上，该方法末端距离误差（10.35 ± 4.18）mm、轨迹平滑度（0.05 ± 0.01）mm 和执行步数（8.54 ± 1.35）均表现优异。

（3）将所提方法部署于果园实地采摘测试中，结合夹具控制网络的机械臂，平均采摘成功率达 84%，显著优于传统规划方法的 52%，平均采摘时间为 10.58 s，展现出良好的实际应用性能与实时性。

参　考　文　献

[1] QASIM M, BABAR S, YOUNAS M, et al. Exploration of biochars for enhancing soil health under citrus cultivation[J]. European Journal of Agronomy, 2025, 168: 127661.

[2] 陈天赐, 张世昂, 付根平, 等. 热带水果采摘机器人研究进展综述[J]. 农业机械学报, 2025, 56(5): 184－201.
CHEN Tianci, ZHANG Shiang, FU Genping, et al. Review of research progress of tropical fruit harvesting robots[J]. Transactions of the Chinese Society for Agricultural Machinery, 2025, 56(5): 184－201. (in Chinese)

[3] LU J, CHEN W, LAN Y, et al. Design of citrus peel defect and fruit morphology detection method based on machine vision[J]. Computers and Electronics in Agriculture, 2024, 219: 108721.

[4] 李丽, 梁继元, 张云峰, 等. 基于改进 YOLO v5 的复杂环境下柑橘目标精准检测与定位方法[J]. 农业机械学报, 2024, 55(8): 280－290.
LI Li, LIANG Jiyuan, ZHANG Yunfeng, et al. Accurate detection and localization methed of citrus targets in complex environments based on improved YOLO v5[J]. Transactions of the Chinese Society for Agricultural Machinery, 2024, 55(8): 280－290. (in Chinese)

[5] KANG H, ZHOU H, CHEN C, et al. Visual perception and modeling for autonomous apple harvesting[J]. IEEE Access, 2020, 8: 62151－62163.

[6] MOHYUDDIN G, KHAN M, HASEEB A, et al. Evaluation of machine learning approaches for precision farming: a review[J]. IEEE Access, 2024, 12: 60155－60184.

[7] KUFFNER J, LAVALLE S M. RRT-connect: an efficient single-query path planning approach[C]// Proceedings of the IEEE International Conference on Robotics and Automation, 2000: 995－1001.

[8] 鲍秀兰, 包有刚, 马萧杰, 等. 自然环境下柑橘采摘机器人避障规划研究[J]. 农业机械学报, 2025, 56(2): 420－428.
BAO Xiulan, BAO Yougang, MA Xiaojie, et al. Obstacle avoidance planning of citrus picking robot in natural environment[J]. Transactions of the Chinese Society for Agricultural Machinery, 2025, 56(2): 420－428. (in Chinese)

[9] SHANG H, CHI X, LI R, et al. Obstacle-aware path planning for robotic manipulators in factory farming[J]. IEEE Transactions on Industrial Informatics, 2025, 21(10): 7733－7741.

[10] MALIK A, LISCHUK Y, HENDERSON T, et al. Deep reinforcement learning approach for inverse kinematics of robotic manipulator[J]. Robotics, 2022, 11(2): 44.

[11] LADOSZ P, WENG L, KIM M, et al. Exploration in deep reinforcement learning: a survey[J]. Information Fusion, 2022, 85: 1－22.

[12] QU M, WANG Y, PHAM D. Robotic disassembly training using reinforcement learning[J]. IEEE Transactions on Industrial Informatics, 2023, 19(11): 10934－10943.

[13] CUI Y, XU Z, ZHONG L, et al. Task-adaptive deep reinforcement learning for dual-arm robot manipulation[J]. IEEE Transactions on Automation Science and Engineering, 2025, 22: 466－479.

[14] LI J, SHI H, HWANG K. Goal-conditioned reinforcement learning with deep imitation for robot arm control[J]. IEEE

Transactions on Automation Science and Engineering, 2024, 21(4): 6217 - 6228.
[15] WATKINS C J C H, DAYAN P. Q-learning[J]. Machine Learning, 1992, 8(3 - 4): 279 - 292.
[16] SUTTON R S, MCALLESTER D, SINGH S, et al. Policy gradient methods for reinforcement learning[C]// Advances in Neural Information Processing Systems, 2000, 12: 1057 - 1063.
[17] MNIH V, KAVUKCUOGLU K, SILVER D, et al. Human-level control through deep reinforcement learning[J]. Nature, 2015, 518(7540): 529 - 533.
[18] JOSHI S, KUMRA S, SAHIN F. Robotic grasping using deep reinforcement learning[C]// Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics, 2020: 3215 - 3220.
[19] GAUTIER P, LAURENT J, DIGUET J P, et al. Deep Q-learning-based dynamic management of robotic cluster[J]. IEEE Transactions on Automation Science and Engineering, 2022, 19(4): 2475 - 2486.
[20] LIN G, ZHU L, LI J, et al. Collision-free path planning for guava-harvesting robot with recurrent deep reinforcement learning[J]. Computers and Electronics in Agriculture, 2021, 188: 106350.
[21] LI Y, FENG Q, ZHANG Y, et al. Peduncle collision-free grasping for tomato harvesting robot[J]. Computers and Electronics in Agriculture, 2024, 216: 108488.
[22] HOCHREITER S, SCHMIDHUBER J. Long short-term memory[J]. Neural Computation, 1997, 9(8): 1735 - 1780.
[23] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[C]// Advances in Neural Information Processing Systems, 2017: 5998 - 6008.
[24] HAARNOJA T, ZHOU A, ABBEEL P, et al. Soft actor-critic: maximum entropy deep reinforcement learning with a stochastic actor[C]// Proceedings of the International Conference on Machine Learning, 2018: 1861 - 1870.
[25] CHENG R, VERMA A, OROSZ G, et al. Control regularization for reduced variance reinforcement learning[C]// Proceedings of the International Conference on Machine Learning, 2019: 1141 - 1150.
[26] ANDRYCHOWICZ M, WOLSKI F, RAY A, et al. Hindsight experience replay[C]// Advances in Neural Information Processing Systems, 2017: 30.
[27] YOSHIKAWA T. Manipulability of robotic mechanisms[J]. International Journal of Robotics Research, 1985, 4(2): 3 - 9.
[28] LILLICRAP T P, HUNT J J, PRITZEL A, et al. Continuous control with deep reinforcement learning[J]. arXiv Preprint, arXiv:1509.02971, 2015.
[29] FUJIMOTO S, HOOF H, MEGER D. Addressing function approximation error in actor-critic methods[J]. arXiv Preprint, arXiv:1802.09477, 2018.
[30] LINDNER T, MILECKI A, WYRWAŁ D. Positioning of robotic arm using different reinforcement learning algorithms[J]. International Journal of Control, Automation and Systems, 2021, 19: 1661 - 1676.
[31] XIAO X, WANG Y, ZHOU B, et al. Flexible hand claw picking method for citrus-picking robot[J]. Agriculture, 2024, 14(8): 1227.

(上接第 70 页)

[16] RONG J, DAI G, WANG P, et al. A peduncle detection method of tomato for autonomous harvesting[J]. Complex & Intelligent Systems, 2022, 8(4): 2955 - 2969.
[17] 王亚薇, 何津立, 林熙淼, 等. 温室番茄采摘机器人伸缩式机械臂设计与试验[J]. 农业机械学报, 2024, 55(增刊1): 18 - 28.
WANG Yawei, HE Jinli, LIN Ximiao, et al. Design and experiment of tomato picking robot based on telescopic robotic arm[J]. Transactions of the Chinese Society for Agricultural Machinery, 2024, 55(Supp. 1): 18 - 28. (in Chinese)
[18] RONG J, ZHOU H, ZHANG F, et al. Tomato cluster detection and counting using improved YOLOv5 based on RGB - D fusion[J]. Computers and Electronics in Agriculture, 2023, 207: 107741.
[19] RONG J, LI X, ZHENG W, et al. An end-to-end maturity prediction and hierarchical counting method for cherry tomatoes[J]. Biosystems Engineering, 2025, 256: 104191.
[20] ZHANG Y, SUN P, JIANG Y, et al. Bytetrack: multi-object tracking by associating every detection box[C]// European Conference on Computer Vision (ECCV), 2022: 1 - 21.
[21] RONG J, ZHENG W, QI Z, et al. RTMFusion: an enhanced dual-stream architecture algorithm fusing RGB and depth features for instance segmentation of tomato organs[J]. Measurement, 2025, 239: 115484.
[22] RONG J, WANG P, YUAN T. An analytical inverse kinematics optimization method of 7-DOF anthropomorphic manipulators with joint limits[C]// International Conference on Intelligent Robotics and Applications, 2025: 231 - 245.
[23] RONG J, LIU Y, LI X, et al. Decoupled motion planning method for 7-DOF manipulator and lifting joint in automated tomato harvesting[J]. Computers and Electronics in Agriculture, 2025, 237: 110693.
[24] MAHMOUDI S, DAVAR A, SOHRABIPOUR P, et al. Leveraging imitation learning in agricultural robotics: a comprehensive survey and comparative analysis[J]. Frontiers in Robotics and AI, 2024, 11: 1441312.