doi:10.6041/j.issn.1000-1298.2024.04.011

基于 HER－TD3 算法的青皮核桃采摘机械臂路径规划

杨淑华[1,2]　谢晓波[1]　邴振凯[1]　郝建军[1,2]　张秀花[1,2]　袁大超[1,2]
（1.河北农业大学机电工程学院，保定 071001；2.河北省智慧农业装备技术创新中心，保定 071001）

摘要：针对青皮核桃和树枝等障碍物无序生长导致机械臂采摘环境复杂、训练任务量大、稳定性差等普遍存在的问题，本文设计了一种同步带模组与机械臂协作的采摘装置，并采用基于事后经验回放的双延迟深度确定性策略梯度算法（Twin delayed deep deterministic policy gradient with hindsight experience replay，HER－TD3）对采摘机械臂进行路径规划，通过 HER 算法提高智能体的探索能力，缓解稀疏奖励的问题；通过 TD3 算法提高智能体的稳定性，减少了训练中出现的震荡现象。为了证明 HER－TD3 算法的可行性和泛化能力，引入 TD3、HER－DDPG 算法进行对比，采用降维训练方法对 3 种深度强化学习智能体进行训练，结果表明 HER－TD3 算法模型在完成路径规划任务中成功率达到 98%，与 HER－DDPG 算法相比提高 4 个百分点，与 TD3 算法相比提高 19 个百分点；在 CoppeliaSim 软件中搭建三维模型仿真环境，设计初始姿态和碰撞检测，使用 YOLO v4 识别青皮核桃，通过该算法模型能够引导虚拟采摘机械臂避开树枝障碍物达到目标位置，完成无碰撞路径规划，无障碍物和有障碍物时路径规划成功率分别为 91% 和 86%；利用物理样机进行青皮核桃采摘试验时，仍能较好地完成路径规划任务，无障碍物时采摘路径规划成功率为 86.7%，平均运动时间为 12.8 s，有障碍物时采摘路径规划成功率为 80.0%，平均运动时间为 13.6 s，验证了 HER－TD3 算法对复杂环境具有较好的适应性和稳定性。

关键词：青皮核桃；采摘机器人；机械臂；HER－TD3 算法；路径规划
中图分类号：S225.93　　**文献标识码**：A　　**文章编号**：1000-1298(2024)04-0113-11　　**OSID**：

Path Planning of Green Walnut Picking Robotic Arm Based on HER－TD3 Algorithm

YANG Shuhua[1,2]　XIE Xiaobo[1]　BING Zhenkai[1]　HAO Jianjun[1,2]　ZHANG Xiuhua[1,2]　YUAN Dachao[1,2]
(1. *College of Mechanical and Electrical Engineering, Hebei Agricultural University, Baoding* 071001, *China*
2. *Technology Innovation Center of Intelligent Agricultural Equipment of Hebei Province, Baoding* 071001, *China*)

Abstract: In response to the common problems of complex environments, large training tasks, and poor stability caused by the disorder growth of green walnut and tree branches, etc., a harvesting device based on synchronous belt module and manipulator was designed, and the path planning of harvesting manipulator was carried out by using the twin delayed deep deterministic policy gradient with hindsight experience replay (HER－TD3) algorithm. HER algorithm was used to improve the agent's ability of exploration and alleviate the problem of sparse reward, and TD3 algorithm was used to improve the agent's stability and reduce the oscillation in training. In order to demonstrate the feasibility and generalization ability of the HER－TD3 algorithm, TD3 and HER－DDPG algorithms were introduced for comparison. Three deep reinforcement learning agents were trained by using dimensionality reduction training methods. The results showed that the success rate of the HER－TD3 algorithm model in completing path planning tasks reached 98%, which was 4 percentage points higher than that of the HER－DDPG algorithm and 19 percentage points higher than that of TD3. The 3D model simulation environment was built in CoppeliaSim software, and the initial attitude and collision detection were designed, YOLO v4 was used to recognize green walnuts, and used this algorithm model to guide the virtual harvesting robotic arm to avoid tree branches and obstacles to reach the target position, completing

收稿日期：2023－12－20　修回日期：2024－01－22
基金项目：河北省重点研发计划项目(21327211D)和河北省博士研究生创新能力培养项目(CXZZBS2022050)
作者简介：杨淑华(1975—)，女，讲师，博士生，主要从事智能化农业装备技术研究，E-mail：hbbdysh@163.com
通信作者：郝建军(1972—)，男，教授，博士生导师，主要从事农业生产机械化自动化装备研究，E-mail：hjjpaper@163.com

collision free path planning. The success rates of path planning were 91% in the absence of obstacles and 86% in the presence of obstacles. In the experiment of picking green walnut using a physical prototype, the path planning task was still well completed. The success rate of path planning for harvesting without obstacles was 86.7%, with an average motion time of 12.8 s, while the success rate in the presence of obstacles was 80.0%, with an average motion time of 13.6 s. It was verified that HER - TD3 algorithm had good adaptability and stability to complex environment.

Key words: green walnut; picking robot; robotic arm; HER - TD3 algorithm; path planning

0 引言

目前,核桃大多采用振动式采收机械采摘或人工采摘[1]。振动式采收宜在果实成熟后青皮开裂、果柄连接力小进行。然而我国除新疆外的大多数核桃种植区,核桃收获期恰逢阴雨气候,为避免采收不及时导致果仁发生霉变,需在核桃青皮未开裂前完成采摘。青皮核桃的果-柄连接力大,振动式采收不能很好满足采收需求。现有核桃园大多仍采用人工采摘,人工采摘工作量大、采摘效率低,采收问题已成为制约核桃产业发展的瓶颈,核桃智能化采摘已成为主要研究方向[2]。

在核桃园自然环境下进行青皮核桃智能化采摘,由于采摘目标物(青皮核桃)和障碍物(枝干、枝条、树叶等)具有随机性和不确定性,使采摘机械臂空间连续无碰撞作业难度增大[3]。对采摘机械臂进行路径规划,获取一条不与非目标障碍物发生碰撞的路径具有重要意义[4]。近年来,国内外诸多学者应用 Dijkstra 算法[5-6]、A* 算法[7-10]、人工势场法(APF)[11-12]等解决低自由度机械臂规划问题;在解决高维空间多自由度机械臂的运动规划问题时,快速搜索随机树算法(RRT)[3,12-16]被广泛使用;深度强化学习算法(DRL)在面对复杂环境和任务时表现突出,有助于机器人的自主学习和避障规划,该算法利用深度学习(DL)的端到端映射能力来获取较强的感知,并通过强化学习(RL)的自我学习能力从与环境的交互中获取反馈来优化策略,具有深度学习强大的感知能力和强化学习智能的决策能力。

HASSELT 等[17]提出了改进的深度双 Q 网络算法(DDQN),提高了神经网络在预测 Q 值时的准确性。为了解决样本采集过程中等概率选取样本而忽略优先级的问题,SCHAUL 等[18]提出了深度双 Q 网络以及优先级经验回放(Double deep Q - Network with prioritized experience replay),该方法对样本的选择进行重要性划分,并减少了训练时间。RAUBER 等[19]提出事后经验回放,用于在稀疏奖励环境中提高样本学习效率。LILLICRAP 等[20]提出了 DDPG 算法,使机械臂可以完成连续的行动空间任务。HAARNOJA 等[21]提出了双延迟深度确定性策略梯度(Twin delayed deep deterministic policy gradient, TD3)算法以及基于软策略和软 Q 值的算法。RAKELLY 等[22]使用基于 TD3 算法的元强化学习框架,引入概率性上下文变量机制来处理机械臂控制问题。李鹤宇等[23]通过深度强化学习算法控制了机械臂运动,并利用奖励函数进行神经网络参数的训练。熊俊涛等[4]提出基于深度强化学习虚拟机器人采摘路径避障规划方法,能够在采摘工作数量较多且障碍物位置不确定的情况下,快速进行路径规划,成功率达到 96.7% 以上。郑嫦娥等[24]提出了基于 DDPG 的渐进空间约束分步训练策略;利用迁移学习思想,将轨迹规划的最优策略迁移并进行了采摘轨迹规划仿真试验,结果表明分步迁移策略能够提高 DDPG 算法的训练效率与网络性能。LIN 等[25]提出的一种结合循环神经网络的 DDPG 算法,应用在番石榴采摘机械臂上获得了较好的鲁棒性。张强等[26]采用目标位置引导和 TD3 算法的联合方式进行轨迹优化,解决了高维动作空间中学习效率低下的问题。熊春源等[27]在柑橘果树非结构化环境的路径规划问题中,引入 LSTM 结构对 DDPG 和 SAC 两种深度强化学习算法进行改进,提升了收敛速度和采摘路径规划成功率。KIM 等[28]使用 TD3 算法使机械手运动具有连续动作,并采用事后经验回放(Hindsight experience replay, HER)算法来提高采样效率。

实际作业过程中,机械臂的状态与动作之间存在映射关系,即下一个状态取决于当前状态和当前动作,该映射关系可以用马尔可夫决策过程(Markov decision process, MDP)进行建模[20]。本文基于项目组前期对自然环境下青皮核桃识别与定位[29]的研究基础上,采用事后经验回放的双延迟深度确定性策略梯度算法(Twin delayed deep deterministic policy gradient with hindsight experience replay, HER - TD3)实现青皮核桃智能化采摘机械臂无碰撞路径规划。

1 整机结构

依据矮化密植核桃园栽植模式(行距 5 m,株距 3 m),设计一种青皮核桃采摘装置,如图 1 所示。主

要由机械臂、末端执行器、同步带模组等组成，并将其搭载在剪叉式升降作业平台上。机械臂底座下方布置双目相机，用于识别并获取核桃位置信息后，通过机械臂控制系统控制机械臂和末端执行器运动完成青皮核桃采摘路径规划和摘取。同步带模组水平交错布置，实现机械臂在株距方向的移动，剪叉式升降作业平台实现在树高方向的移动，履带式行走底盘实现整机沿行距方向移动。

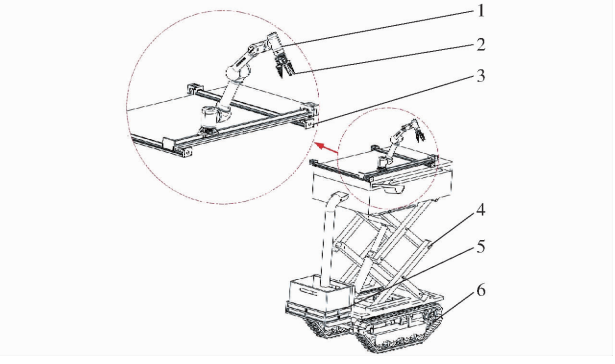


图 1　整机结构示意图

Fig. 1　Operation platform complete machine plan

1. 机械臂　2. 末端执行器　3. 同步带模组　4. 剪叉式升降作业平台　5. 输果收集装置　6. 履带式行走底盘

2　HER－TD3 算法

TD3 算法是在深度确定性策略梯度（Deep deterministic policy gradient，DDPG）算法上发展起来的[30]。而 DDPG 算法基于 DPG 算法[31]，使用 AC 算法框架[32]。假设给定任意的初始状态 q_0 和目标状态 q_{goal}，且回合的最大长度为 t。在状态 q_t 时，如果将动作 a_t 应用于马尔可夫决策过程中，则产生 q_{t+1}。根据奖励函数确定下一个状态 q_{t+1} 和 r_{t+1}，重复这个过程直到回合结束。

回合结束有两种可能：①结束状态为目标状态 $\boldsymbol{e}=[(q_0,a_0)\quad(q_1,r_1,a_1)\quad\cdots\quad(q_{\text{goal}},r_{\text{goal}})]$。②回合长度为 T，但没有达到目标状态 $\boldsymbol{e}=[(q_0,a_0)\quad(q_1,r_1,a_1)\quad\cdots\quad(q_T,r_T)]$ 且 $q_T\neq q_{\text{goal}}$。对于这两种情况，每个样本都以 (q_t,a_t,r_t,q_{t+1}) 的形式储存在经验池中。

处理系数奖励问题是深度强化学习最大的挑战之一[33]。针对此问题，OpenAI 提出了 HER 算法[34]，以路径规划为例，在智能体学习的过程中，起初智能体并不能准确了解如何到达指定的目标点，所以经验轨迹得到的奖励值都为 0，在后续训练中智能体并未用上这样的经验轨迹。而 HER 算法主要思想是将这些看似失败的经验轨迹转变为成功的经验。对于 q_t 不是目标状态经验池中的片段 $\boldsymbol{e}=[(q_0,a_0)\quad(q_1,r_1,a_1)\quad\cdots\quad(q_T,r_T)]$，经过 HER 算法的修改 q_t 状态在目标处成为结束状态，这样失败的状态可以用来训练智能体。

图 2 为 HER－TD3 算法的结构，TD3 算法有 2 个评论家目标深度神经网络（DNN）、1 个行动者目标深度神经网络，以用于每个评论家和行动者深度神经网络的 3 个目标深度神经网络，所以 TD3 有 6 个深度神经网络。评论家 DNN 生成状态-动作值函数的最优估计 $Q_{q_i}(s_t,a_t)$。评论家 DNN 的输入为经验池中的小批量样本，输出为 $Q_{q_i}(s_t,a_t)$，$i=1,2$，小批量样本为经验池的样本集 (s_t,a_t,r_t,s_{t+1})。在 TD3 中为了消除 $Q_{q_i}(s_t,a_t)$ 函数的高估偏差，当积累噪声导致不良状态被高估时，会出现高估偏差。为了解决这种情况，TD3 从 2 个评论家 DNN 输出中选择较小的 $Q_{q_i}(s_t,a_t)$ 值作为目标对象。θ_1、θ_2 为 2 个评价家 DNN 的参数，θ_1'、θ_2' 为对应 DNN 的参数。为了训练评论家 DNN，作为成本函数，将时间差误差 $\delta=Q_\theta(q,a)-y_{\text{target}}$ 最小化[26]，即

$$J_\delta(\theta)=\frac{1}{2}(Q_\theta(q,a)-y_{\text{target}})^2 \tag{1}$$

其中

$$y_{\text{target}}=r+\gamma\min_{i=1,2}Q_{\theta_i'}(q,\tilde{a}) \tag{2}$$

$$\tilde{a}=\mu_\varphi(q)+\bar{\varepsilon} \tag{3}$$

式中　$Q_\theta(q,a)$——参数 θ 的参数化状态-动作值函数 Q

y_{target}——目标动作（即评论家 DNN 中使用的动作），为函数 $Q_{\theta'}(q,\tilde{a})$ 的目标值

r——奖励函数

γ——折扣因子

$\mu_\varphi(q)$——策略

式（3）中噪声 $\bar{\varepsilon}$ 遵循正态分布 $\text{clip}[N(0,\sigma),-c,c]$，$c>0$，$\sigma>0$，表明噪声 $\bar{\varepsilon}$ 是一个 $N(0,\sigma)$ 的随机变量，区间为 $[-c,c]$。

行动者 DNN 的输入为评论家 DNN 的 $Q_{\theta_1}(s_t,a_t)$ 和经验池中的小样本，输出为动作。动作由 $a_t=\mu_\varphi(q_t)+\varepsilon$，其中 φ 为行动者 DNN 的输出，为一个确定的连续值。噪声 ε 服从正态分布 $N(0,\sigma)$ 并用于探索。为了调整参数 φ，取最小化，即

$$J_\mu(\varphi)=\sum_q d^\mu(q)Q_\theta(q,\mu_\varphi(q)+\varepsilon)$$

式中　$d^\mu(q)$——状态分布

$J_\mu(\varphi)$——策略梯度，用于更新参数 φ

在 2 个评论家 DNN 两个输出之间使用式（2）的目标值来更新 2 个评价家 DNN。TD3 算法每 d 步周期性更新行动者 DNN 和 3 个目标 DNN，以避免过快地收敛。此外，策略 $\mu_\varphi(q)$ 按 $Q_{\theta_1}(q,\mu_\varphi(q))$ 成比例更新，行动者目标 DNN 和评论家目标 DNN 的参数每

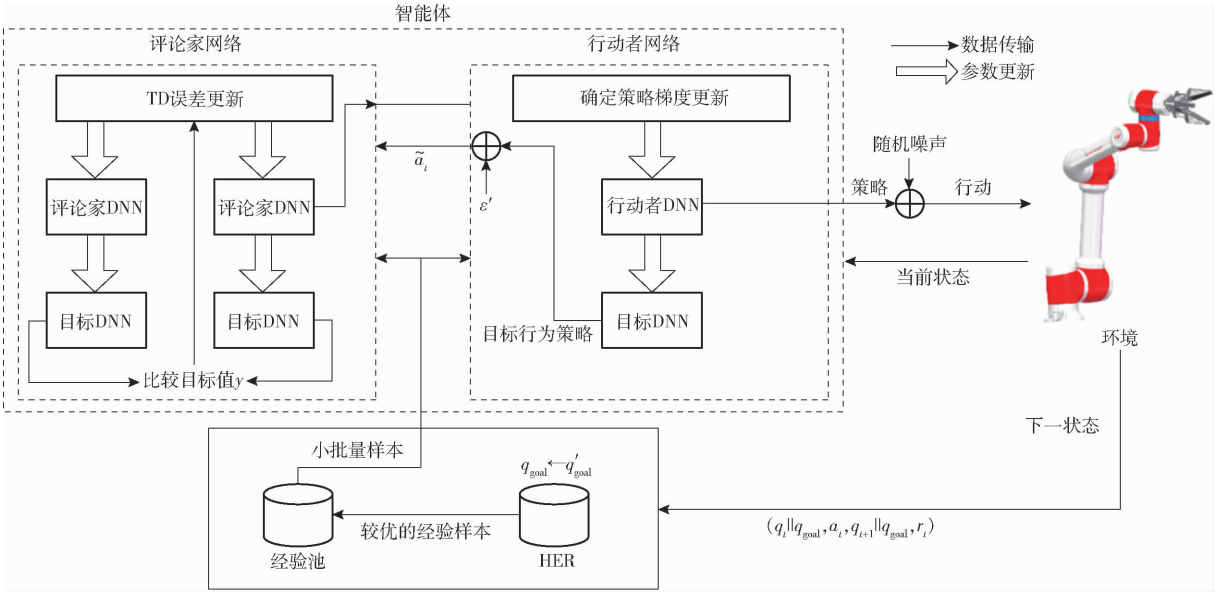


图 2 基于 HER－TD3 结构

Fig. 2 Structure diagram based on HER－TD3

d 步按 $\theta' \leftarrow \tau\theta + (1-\tau)\theta'$ 更新，这样不仅保证了较小的时间误差，而且使目标 DNN 中的参数更新变得缓慢。HER 算法从失败的经验获得更多的有效数据，从而提高智能体学习速率和网络收敛速度，进而加速了 TD3 算法累计足够的样本速度，在经验池中储存了大量样本内容，在每次迭代更新中，选取一定数量样本作为更新依据。HER 算法将经验池中的样本进行区分有利于算法训练，增加了模型训练成功的可能性。

采用 HER－TD3 算法解决自然环境下青皮核桃采摘机械臂的路径规划问题。智能体通过训练行动者网络获取控制策略并输出动作策略控制机械臂运动[33]。训练过程中，行动者网络根据评论家网络给出的动作评价和环境给的奖励不断迭代更新网络参数[35]，最终到达目标核桃。智能体依据奖励进行策略优化，奖励函数决定了算法收敛速度和收敛程度。将青皮核桃目标物位置中心点设置为(x_2, y_2)，机械臂末端中心点位置设置为(x_1, y_1)，根据机械臂末端中心点与青皮核桃目标物位置中心点的距离，设计奖励函数 r_1 为

$$r_1 = -\sqrt{(x_2-x_1)^2+(y_2-y_1)^2} \tag{4}$$

根据机械臂末端是否到达青皮核桃目标物，设置奖励函数 r_2 为

$$r_2 = \begin{cases} 10 & (\text{到达青皮核桃目标物}) \\ 0 & (\text{其他}) \end{cases} \tag{5}$$

根据机械臂末端是否触碰树枝障碍物，设置奖励函数 r_3 为

$$r_3 = \begin{cases} -10 & (\text{触碰树枝障碍物}) \\ 0 & (\text{其他}) \end{cases} \tag{6}$$

根据奖励函数 r_1、r_2、r_3 确定最终的奖励值 R 为

$$R = r_1 + r_2 + r_3 \tag{7}$$

通过设计奖励函数以期用较快收敛速度和较好收敛程度获取较高且稳定的奖励值。

3 机械臂仿真训练

在三维环境下对机械臂模型训练会导致解空间庞大，需要花费很长时间训练机械臂模型，且对计算机算力要求很高，采用降维训练方法，将机械臂抓取空间映射到二维空间。用 3 种不同的深度强化学习算法训练，并搭建三维仿真环境检验 HER－TD3 算法的训练效果。

3.1 Pygelt 模块二维平面仿真训练

3.1.1 机械臂降维至二维平面训练过程

机械臂模型的降维方法是将机械臂抓取任务分解为抓取方向和在抓取平面内的降维映射两部分[36]。图 3a 为 Jaka C5 机械臂的三维模型图，机械臂舵机可分为横向舵机（旋转轴垂直于水平面）和纵向舵机（旋转轴平行于水平面）两类。关节 1、5 为横向舵机，关节 2、3、4 为纵向舵机。按两类舵机不同的旋转方式将抓取任务分为：①确定抓取方向，通过一个或多个横向舵机的水平转动，令机械臂末端正对目标，并与目标点直线距离最短的位置为抓取平面。②根据选定的横向舵机确定好的抓取平面，在横向舵机保持不动的前提下，将纵向舵机、末端执行器及其连杆从三维空间映射（即投影）至二维空间。③在抓取平面内作机械臂的抓取训练。如图 3b 所示。

降维映射中需考虑两个问题：①臂杆长度比例

保持不变。②旋转角范围与二维仿真环境下的旋转角范围是否存在差异。通过以上分析，采摘机械臂在平面内可以简化为 3 个首尾相连的连杆，如图 3c 所示。通过映射简化机械臂模型、减少训练时间，降维后机械臂模型参数如表 1 所示。

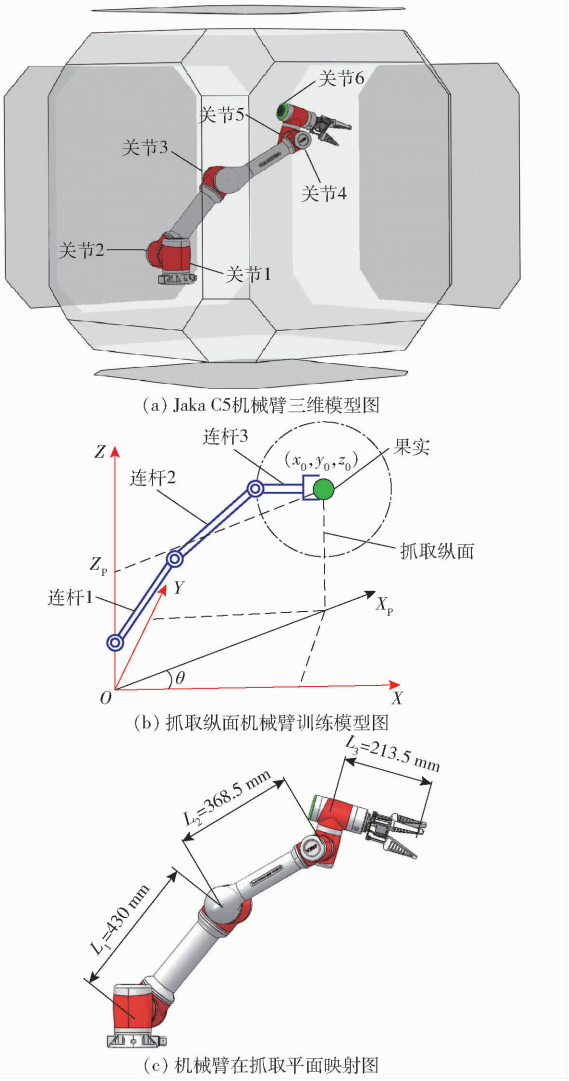


图 3　映射示意图

Fig. 3　Mapping diagrams

表 1　Jaka C5 降维后参数

Tab. 1　Parameters after dimensionality reduction of Jaka C5

连杆序号	长度/mm	旋转角/(°)	初始角/(°)
1	430.0	－360～360	0
2	368.5	－360～360	0
3	213.5	－360～360	0

为了验证深度强化学习在机械臂路径规划中的可行性，使用 Python 语言的 Pyglet 模块编写仿真环境。在模型训练过程中，使用 Windows 操作系统，并使用 Anaconda 4.5 环境管理器来管理 Python 3.7、Gym 0.19 和 Pyglet 1.5.26。

该仿真环境由机械臂（用红、黑、蓝色分别表示连杆 1、连杆 2、连杆 3）、树枝障碍物（用橙色长方块表示）和青皮核桃目标物（用绿色方块表示）组成，如图 4 所示。机械臂末端与目标物之间的欧氏距离小于等于 0.02 m 时，路径规划任务被认为是成功的。

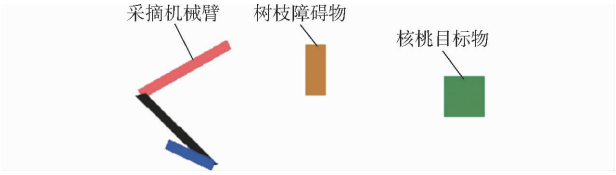


图 4　仿真环境

Fig. 4　Simulation environment

为了模拟自然环境条件下进行采摘工作，不固定青皮核桃目标物和树枝障碍物的位置，使其随机出现，训练参数：HER－TD3 算法迭代最大轮数设置为 10 000，每轮最大步数设置为 500 步，奖励折扣因子设置为 0.99，actor 部分学习率为 0.001，critic 部分学习率为 0.001，BATCH_SIZE 设置为 32。并设定一个 HER 算法经验池，其大小为 500，用于存储表现良好的样本。为了验证本文方法的有效性，引用 HER－DDPG 算法和 TD3 算法作为对比算法，训练参数设置与 HER－TD3 算法相同。训练过程如图 5 所示。图 5a 表示障碍物未遮挡核桃时路径规划开始、规划中和完成的过程，图 5b 表示障碍物遮挡核桃时路径规划开始、规划中和完成的过程。从图 5 可以看出，仿真过程中，无论有无障碍物遮挡，3 个连杆都能围绕连杆 1 的始端不断变换姿态，向目标核桃不断接近，直至成功到达目标点。

当深度强化学习算法训练完成后，可以利用 tensorboard 命令观察网络的损失值训练变化曲线来了解网络的训练情况。其中损失值指在线评论家网络输出，表示当前策略网络可以获得的奖励，优化此参数可以使行动者网络获得更高奖励。因此，通过 3 种算法在训练过程中损失值训练曲线来比较 3 种深度强化学习算法在采摘机械臂路径规划仿真环境下的实际训练效果。

3.1.2　仿真结果

如图 6 可知，3 种算法的损失值变化曲线都呈现下降趋势，说明通过训练使得采摘机械臂获得的奖励值得到了提高。但 TD3 算法在整个阶段都有大幅度震荡，说明其表现不稳定，相比较 HER－DDPG 算法和 HER－TD3 算法在 1 000 步时出现短暂的上升过程，到 3 000 步后呈急速下降趋势，说明其获得的奖励值明显得到提高，这个阶段采摘机械臂成功率也相应得到提高。HER－TD3 算法相较于 HER－DDPG 算法其损失值更低，为 0.04，说明该算法成功率更高。

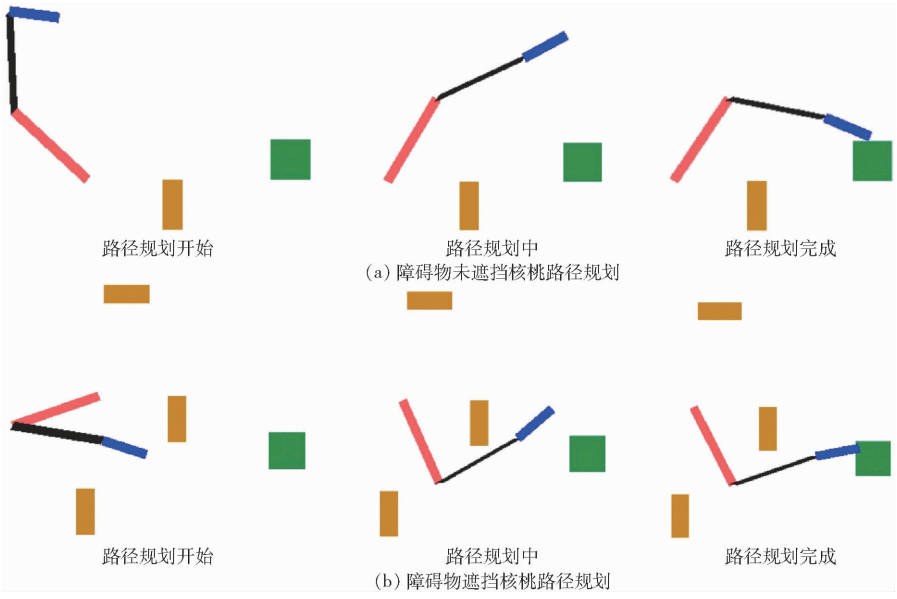


图 5 Pyglet 模块训练过程

Fig. 5 Pyglet module training process

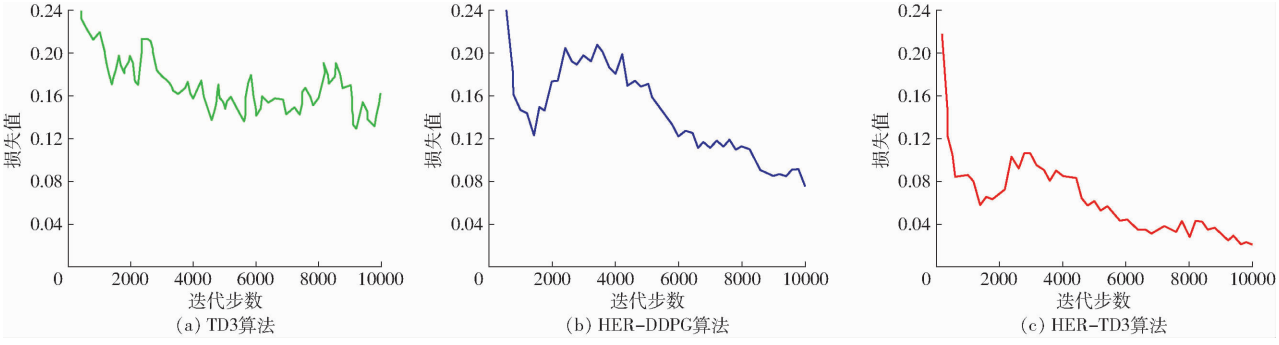


图 6 3 种算法损失值变化曲线

Fig. 6 Three algorithms loss change process diagrams

如图 7 所示,TD3 算法训练成功率增长缓慢,一直训练到 5 000 步左右时训练的成功率才开始呈明显的增长趋势,最后成功率停留在 79%。HER - DDPG 算法和 HER - TD3 算法相较于 TD3 算法有了较大提升,在 2 000 步左右时成功率开始快速增长,表明算法学习效率高,最后成功率为 94% 和 98%。综上,3 种算法在仿真的对比训练中,HER - TD3 算法在成功率方面分别提高 19 个百分点和 4 个百分点。由此证明,HER - TD3 算法对机械臂路径规划是有效的。

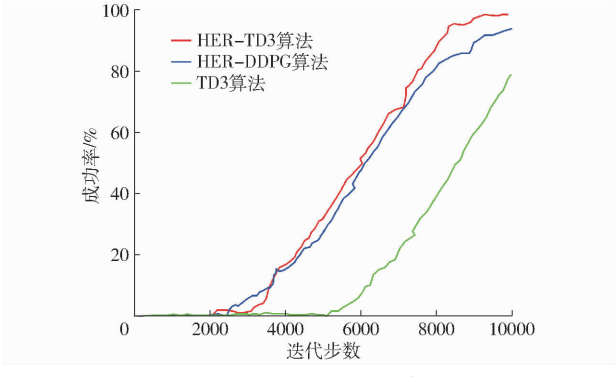


图 7 路径规划成功率

Fig. 7 Path planning success rate

3.2 CoppeliaSim 三维空间仿真训练过程

为确保仿真试验的真实性,选择在河北农业大学三分场核桃种植基地采摘的成熟中期青皮核桃(图 8a、8b)作为目标物进行仿真试验。挑选尺寸接近平均值的青皮核桃进行研究,采用 GD - 3DScan 型四目三维扫描仪(测量精度为 0.001 ~ 0.05 mm,单幅扫描点数为 2 620 000,点距为 0.15 ~ 0.30 mm)进行三维扫描,从而获得青皮核桃外表面的精确三维坐标,并生成点云数据。然后,使用 Geomagic Studio 软件对点云数据进行合并拼接处理,以获取完整的青皮核桃模型[37]。最后,利用 GOM Inspect 软件对尖锐部分和噪点进行锐化处理,得到最终的青皮核桃三维模型,如图 8c、8d 所示。

由于树叶轻、薄且有良好的退让性,对机械臂路径规划和采摘过程没有直接影响,为了简化仿真环境,减少计算复杂度,提高仿真效率,在核桃树模型中未添加树叶。

机械臂和末端执行器由 Solidworks 绘制完成,



图 8　青皮核桃图像与模型

Fig. 8　Green walnut photo and model

由于仿真中机械臂和末端执行器需要进行运动，使用 sw_urdf_exporter 插件导出为 URDF 文件。使用远程 API 控制机械臂的可视化仿真，Python 搭建和实现所需测试的强化学习算法，而 CoppeliaSim 提供物理交互环境，如图 9a 所示。将末端执行器放入 Jaka C5_connection 目录下，摄像头使用 CoppeliaSim 中的 Kinect 相机但外观使用普通摄像头，将以上各部分连接，本文初始姿态设计和环境如图 9b 所示。

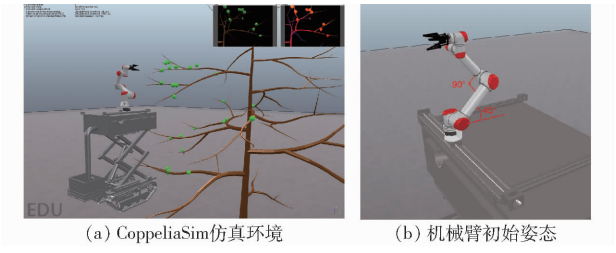


图 9　CoppeliaSim 仿真环境及初始姿态

Fig. 9　CoppeliaSim simulation environment and initial posture

通过 Kinect 相机获得深度图像(图 10a)和 RGB 图像(图 10b)，其中 kinect_depth 用来模拟真实 Kinect 相机的深度传感器获取空间点的深度；kinect_rgb 用来获取场景的彩色图。利用 pygame 模块按键事件功能调整机械臂位姿采集图像，制作成训练集，如图 10c、10d 所示。利用 LabelImg 标注软件人工进行青皮核桃图像数据标注，标注完成后按比例 8∶2 将图像数据和对应的标注文件随机划分为训练和测试数据集。

上述获得的训练集如图 11a、11c 所示，通过 YOLO v4 进行图像识别，训练后结果如图 11b、11d 所示。

选择使用基于包围层次盒的碰撞检测方法中的 AABB 包围盒法。为了检测采摘机械臂与环境以及

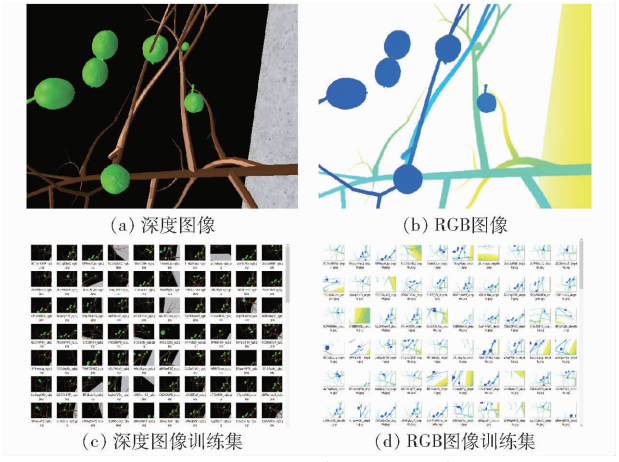


图 10　RGB 图像和深度图像

Fig. 10　RGB images and depth images

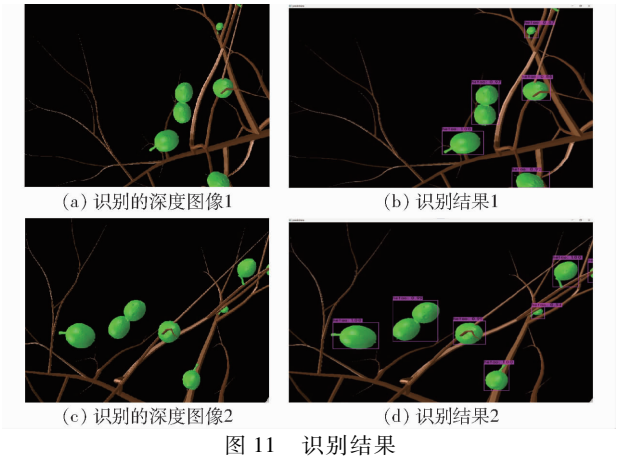


图 11　识别结果

Fig. 11　Identification results

机械臂自身各个关节之间是否发生碰撞，在 CoppeliaSim 的 Collections 设置中，将障碍物组合并命名为 Obstacle，同时将机械臂的各个关节按组进行划分。随后，在 CoppeliaSim 的 Calculation Modules 模块中设置碰撞和自碰撞检测的 Collision 表达式。当机械臂与物体发生碰撞或自碰撞时，机械臂整体颜色会变为红色，如图 12 所示。通过此设计，可以实时检测机械臂与环境以及机械臂自身各个关节之间的碰撞情况。

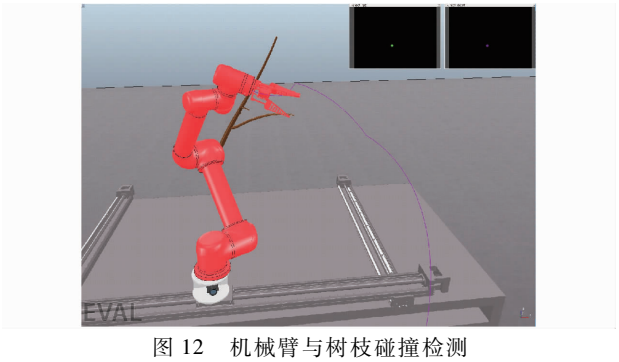


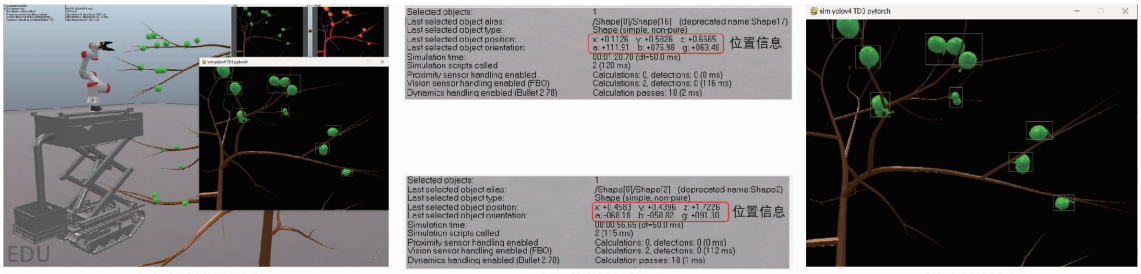
图 12　机械臂与树枝碰撞检测

Fig. 12　Collision detection

将训练好的识别模型导入 CoppeliaSim 中，可以实现机械臂实时识别功能(图 13a)，通过对物体的

检测可以得到抓取物体的二维包围框,如图 13c 所示,在 CoppeliaSim 中已经内置了对已识别目标的定位功能,识别到目标果实后,可以直接读取其位置信息,如图 13b 所示。

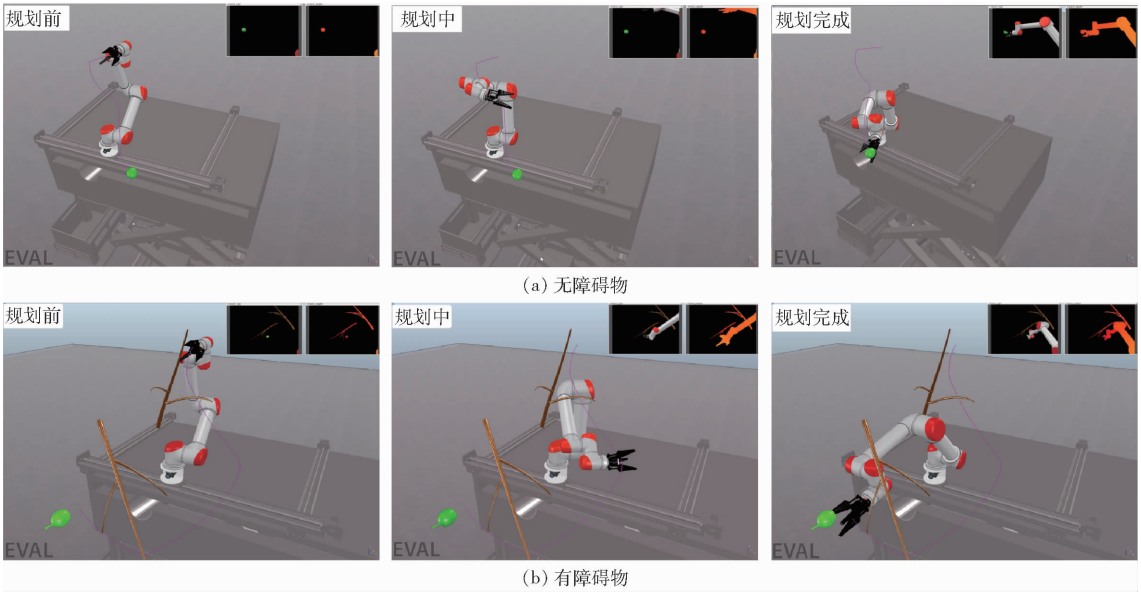
为检验仿真模型路径规划过程及效果,截取规划前、规划中及规划完成时 3 种状态,图 14a、14b 分别为无障碍物路径规划和有障碍物路径规划过程,可以看出机械臂的运动有明确的方向,虚拟机械臂能够根据采摘位置和障碍物位置,对于不同的情况,顺畅地完成无碰撞路径规划任务。



(a) 整体识别图 (b) 位置信息图 (c) 识别结果

图 13 实时识别

Fig. 13 Real time recognition



(a) 无障碍物

(b) 有障碍物

图 14 机械臂与目标果实间路径规划过程

Fig. 14 Path planning process between robotic arm and target fruit

无障碍物路径规划和有障碍物路径规划成功率如图 15 所示,可以看出无障碍物路径规划成功率明显高于有障碍物路径规划,当步数达到 2 000 步左右,成功率呈明显上升趋势,达到 8 000 步左右时趋于稳定,可以看出最终成功率为 91% 和 86%。

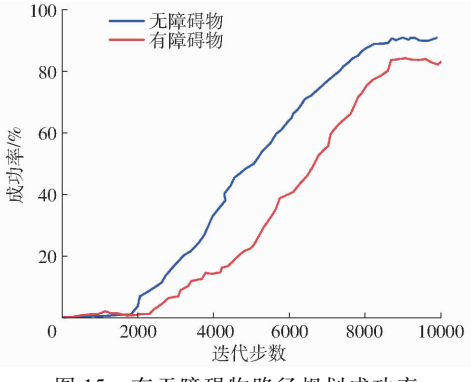
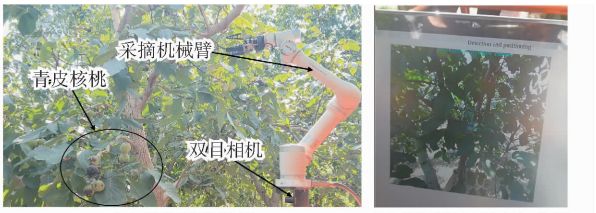


图 15 有无障碍物路径规划成功率

Fig. 15 Success rate of path planning with or without obstracles

4 样机试验验证

试验样机主要由 Jaka C5 机械臂、末端执行器、视觉定位系统、上位机、机械臂自带的控制系统组成,如图 16a 所示。其中,Jaka C5 采摘机械臂工作空间为:以基座为球心,半径为 1 050 mm 的近球形空间,最大负载 5 kg;末端执行器电机驱动 3 指联动鳍条柔性爪结构。上位机采用的操作系统为 Ubuntu 18.04,硬件配置为处理器 Intel(R) Xeon(R) CPU E5 - 2630 v3 @ 2.40 GHz,内存 32 GB,GPU 为 NVIDIA GeForce RTX2080Ti(11 GB)GPU × 2。使用 Anaconda 4.5 Python 包管理器,编程语言主要为 Python 3.8.10,深度学习框架为 Pytorch,使



(a) 采摘试验样机　(b) 识别目标果实与位置

图 16　采摘机械臂样机

Fig. 16　Picking machine robotic arm prototype

用 CUDA 11.1 并行计算架构，Cudnn 8.0 深度神经网络高性能库，实现多 GPU 并行加速；视觉定位采用 ZED 2i 型双目相机，分辨率调整为 2 ×（1 280 像素 ×720 像素），安装在机械臂底座下部的相机支架上，用于青皮核桃识别与定位。

为验证算法可行性，将视觉识别与定位算法及路径规划算法整合到上位机机械臂控制程序中进行联合调试，上位机通过 TCP 协议与机械臂自带的控制系统通讯，实时传输果实位置和机械臂运动控制参数，机械臂按照控制系统动作运动，进行采摘试验。

样机采摘试验于 2023 年 8 月 19—23 日在河北省邢台市临城县国家标准化矮砧密植核桃园进行，采摘品种为绿岭核桃，采收时天气晴、多云、阴等多种情况；于 2023 年 9 月 13 日在河北省保定市唐县西雹水核桃种植合作社进行，采摘品种为清香核桃，采摘时段为早、中、晚 3 个时段进行。如图 16a 所示，试验中，相机位置与仿真试验中保持一致，保证在 ZED 2i 型双目相机识别定位完成后，获取的青皮核桃位置信息处于机械臂和末端执行器的采摘空间内，采摘路径规划成功的判定标准为将机械臂从初始位置运动到目标果实位置，且运动中不发生碰撞或卡顿。统计的机械臂运动时间为从机械臂开始运动到达目标果实或机械臂中途停止运动时的时长。并且根据是否到达目标核桃将数据分为成功组和失败组，成功组记为“Y”，失败组记为“N”。试验中识别到目标果实并标注实时位置，如图 16b 所示。

试验中按照有障碍物和无障碍物分别进行 15 组采摘路径规划试验，每组取距离相近的 5 个青皮核桃，共 30 组。图 17a 为采摘过程中机械臂与目标果实间无障碍物的机械臂运动过程；图 17b 为采摘过程中机械臂与目标果实间有障碍物的机械臂运动过程。为了表达机械臂避障运动过程，图中目标果实用红色方框标识出，障碍物枝条用黄颜色标识出。



(a) 无障碍物

(b) 有障碍物

图 17　机械臂与目标果实间物理样机试验采摘路径规划

Fig. 17　Physical prototype testing of picking paht planing between robotic arm and target fruit

试验结果表明，在采摘过程中机械臂运动平稳、速度变化连续，能够顺利完成采摘工作。试验结果如表 2、3 所示，无障碍物青皮核桃采摘路径规划平均运动时间为 12.8 s，成功率为 86.7%，有障碍物青皮核桃采摘路径规划平均运动时间为 13.6 s，成功率为 80.0%。

表 2　无障碍物时采摘路径规划试验结果

Tab. 2　Experimental results of packing path planning without obstacles

试验组号	平均运动时间/s	是否到达目标果实
1	11.3	Y
2	12.4	Y
3	13.7	Y
4	13.1	Y
5	12.4	Y
6	8.7	N
7	13.8	Y
8	12.2	Y
9	12.6	Y
10	13.3	N
11	12.7	Y
12	14.1	Y
13	14.2	Y
14	12.9	Y
15	14.4	Y
平均值	12.8	

5　结论

（1）针对青皮核桃采摘过程中连续的空间任务，采用HER－TD3算法完成青皮核桃采摘无碰撞

表 3 有障碍物时采摘路径规划试验结果

Tab.3 Experimental results of picking path planning with obstacles

试验组号	平均运动时间/s	是否到达目标果实	试验组号	平均运动时间/s	是否到达目标果实
1	13.4	Y	9	10.9	N
2	13.8	Y	10	13.2	N
3	12.2	Y	11	14.7	Y
4	15.3	N	12	12.3	Y
5	11.4	Y	13	15.6	Y
6	12.7	Y	14	15.2	Y
7	16.9	Y	15	15.8	Y
8	10.7	Y	平均值	13.6	

路径规划,通过降维方法在 Pygelt 模块训练深度强化学习模型并设计对比试验,结果表明该算法的规划路径成功率达 98%,比 HER - DDPG 算法提高 4 个百分点,比 TD3 算法提高 19 个百分点。

(2)通过 CoppeliaSim 进行三维空间仿真试验模拟对青皮核桃的识别及机械臂路径规划过程,并制造物理样机进行采摘路径规划试验,结果表明 HER - TD3 算法在构建的虚拟仿真系统有障碍物和无障碍物时路径规划成功率 86% 和 91%,物理样机试验中有障碍物和无障碍物时路径规划成功率 80.0% 和 86.7%,实现采摘机械臂顺利采摘,环境适应性好。

参 考 文 献

[1] 祝前峰,陆荣鉴,刘彬,等. 核桃采摘机械研究现状与发展趋势[J]. 林业和草原机械, 2021, 2(1):45 - 53.
ZHU Qianfeng, LU Rongjian, LIU Bin, et al. Research status and development trends of walnut picking machinery[J]. Forestry and Grassland Machinery, 2021, 2(1): 45 - 53. (in Chinese)

[2] 刘成良,贡亮,苑进,等. 农业机器人关键技术研究现状与发展趋势[J]. 农业机械学报,2022,53(7):1 - 22,55.
LIU Chengliang, GONG Liang, YUAN Jin, et al. Current status and development trends of agricultural robots[J]. Transactions of the Chinese Society for Agricultural Machinery, 2022,53(7):1 - 22,55. (in Chinese)

[3] 荀一,李道政,王勇,等. 基于 VS - IRRT 算法的采摘机械臂路径规划[J]. 农业机械学报,2023,54(2):129 - 138.
XUN Yi, LI Daozheng, WANG Yong, et al. Path planning of picking robot arms based on VS - IRRT algorithm[J]. Transactions of the Chinese Society for Agricultural Machinery, 2023,54(2): 129 - 138. (in Chinese)

[4] 熊俊涛,李中行,陈淑绵,等. 基于深度强化学习的虚拟机器人采摘路径避障规划[J]. 农业机械学报,2020,51(增刊 2): 1 - 10.
XIONG Juntao, LI Zhongxing, CHEN Shumian, et al. Virtual robot picking path obstacle avoidance planning based on deep reinforcement learning[J]. Transactions of the Chinese Society for Agricultural Machinery, 2020,51(Supp. 2): 1 - 10. (in Chinese)

[5] RENGANATHAN V, FATHIAN K, SAFAOUI S, et al. Spoof resilient coordination in distributed and robust robotic networks [J]. IEEE Transactions on Control Systems Technology, 2022,30(2):803 - 810.

[6] LUO M, HOU X, YANG J. Surface optimal path planning using an extended Dijkstra algorithm[J]. IEEE Access,2020, 8: 147827 - 147838.

[7] VARGAS A N, FURLONI W, DOVAL J. Second moment constraints and the control problem of Markov jump linear systems [J]. Numerical Linear Algebra with Applications, 2013, 20(2): 357 - 368.

[8] LI Fenglei, HUANG Zexin, XU Lin. Path planning of 6 - DOF venipuncture robot arm based on improved A-star and collision detection algorithms[C]//2019 IEEE International Conference on Robotics and Biomimetics(ROBIO). IEEE,2019: 2971 - 2976.

[9] HENTEN E J V, HEMMING J, TUIJL B A J V, et al. Collision-free motion planning fora cucumber picking robot[J]. Biosystems Engineering, 2003, 86(2):135 - 144.

[10] 尹建军,武传宇,YANG Simon X,等. 番茄采摘机器人机械臂避障路径规划[J]. 农业机械学报, 2012, 43(12): 171 - 175,157.
YIN Jianjun, WU Chuanyu, YANG Simon X, et al. Obstacle avoidance path planning for tomato picking robot robotic arms [J]. Transactions of the Chinese Society for Agricultural Machinery, 2012, 43(12): 171 - 175,157. (in Chinese)

[11] 姬伟,程风仪,赵德安,等. 基于改进人工势场的苹果采摘机器人机械手避障方法[J]. 农业机械学报, 2013, 44(11): 263 - 269.
JI Wei, CHENG Fengyi, ZHAO Dean, et al. Obstacle avoidance method for apple picking robot manipulator based on improved artificial potential field[J] Transactions of the Chinese Society for Agricultural Machinery, 2013, 44 (11): 263 - 269. (in Chinese)

[12] 薛阳,俞志程,吴海东,等. 基于改进人工势场法的双机械臂避障路径规划[J]. 机械传动,2020,44(3):39 - 45.
XUE Yang, YU Zhicheng, WU Haidong, et al. Obstacle avoidance path planning for dual robotic arms based on improved artificial potential field method[J]. Journal of Mechanical Transmission, 2020,44(3): 39 - 45. (in Chinese)

[13] CAO Xiaoman, ZUO Xiangjun, JIA Chunyang, et al. RRT-based path planning for an intelligent litchi-picking manipulator [J]. Computers and Electronics in Agriculture,2019,156: 105 - 118.

[14] 王怀震,高明,王建华,等. 基于改进 RRT* - Connect 算法的机械臂多场景运动规划[J]. 农业机械学报,2022,53(4): 432 - 440.
WANG Huaizhen, GAO Ming, WANG Jianhua, et al. Multi scene motion planning of robotic arms based on improved RRT* -

Connect algorithm[J]. Transactions of the Chinese Society for Agricultural Machinery, 2022, 53(4): 432－440. (in Chinese)
[15] JEONG I B, LEE S J, KIM J H. Quick-RRT*: triangular inequality-based implementation of RRT* with improved initial solution and convergence rate[J]. Expert Systems with Applications, 2019, 123(6): 82－90.
[16] 张勤,乐晓亮,李彬,等.基于 CTB－RRT*的果蔬采摘机械臂运动路径规划[J].农业机械学报,2021,52(10):129－136.
ZHANG Qin, LE Xiaoliang, LI Bin, et al. Motion path planning of fruit and vegetable picking robot arm based on CTB－RRT*[J]. Transactions of the Chinese Society for Agricultural Machinery, 2021, 52(10): 129－136. (in Chinese)
[17] HASSELT H V. Double Q-learning[M]. Mit Press, 2010: 2613－2621.
[18] SCHAUL T, QUAN J, ANTONOGLOU I, et al. Prioritized experience replay[C]// ICLR. 4th International Conference on Learning Representations, 2016:1－21.
[19] RAUBER P, UMMADISINGU A, MUTZ F, et al. Hindsight policy gradients[C]// ICLR. 7th International Conference on Learning Representations, 2019:1－38.
[20] LILLICRAP T P, HUNT J J, PRITZEL A, et al. Continuous control with deep reinforcement learning[J]. Computer Science, 2015,8(6):A187.
[21] HAARNOJA T, ZHOU A, HARTIKAINEN K, et al. Soft actor-critic algorithms and applications[J]. arXiv:1812.05905, 2018.
[22] RAKELLY K, ZHOU A, QUILLEN D, et al. Efficient off-policy meta-reinforcement learning via probabilistic context variables [J]. arXiv:1903.08254, 2019.
[23] 李鹤宇,林廷宇,曾贲,等. 基于强化学习的空间机械臂控制方法[J].航天控制,2020,38(6):38－43.
LI Heyu, LIN Tingyu, ZENG Bi, et al. Control method of space manipulator by using reinforcement learning[J]. Aerospace Control, 2020,38(6):38－43. (in Chinese)
[24] 郑嫦娥,高坡,GAN Hao,等.基于分步迁移策略的苹果采摘机械臂轨迹规划方法[J].农业机械学报,2020,51(12):15－23.
ZHENG Chang'e, GAO Po, GAN Hao, et al. Trajectory planning method for apple picking manipulator based on stepwise migration strategy[J]. Transactions of the Chinese Society for Agricultural Machinery, 2020,51(12):15－23. (in Chinese)
[25] LIN G, ZHU L, LI J, et al. Collision-free path planning for a guava-harvesting robot based on recurrent deep reinforcement learning[J]. Computers and Electronics in Agriculture, 2021, 188: 106350.
[26] 张强,文闻,周晓东,等. 基于改进 TD3 算法的机械臂智能规划方法研究[J].智能科学与技术学报, 2022, 4(2): 223－232.
ZHANG Qiang, WEN Wen, ZHOU Xiaodong, et al. Research on the manipulator intelligent trajectory planning method based on the improved TD3 algorithm[J]. Chinese Journal of Intelligent Science and Technology,2022, 4(2): 223－232. (in Chinese)
[27] 熊春源,熊俊涛,杨振刚,等.基于深度强化学习的柑橘采摘机械臂路径规划方法[J].华南农业大学学报,2023,44(3):473－483.
XIONG Chunyuan, XIONG Juntao, YANG Zhengang, et al. Path planning method for citrus picking robotic arms based on deep reinforcement learning[J]. Journal of South China Agricultural University, 2023,44(3):473－483. (in Chinese)
[28] KIM M S, HAN D K, PARK J H,et al. Motion planning of robot manipulators for a smoother path using a twin delayed deep deterministic policy gradient with hindsight experience replay[J]. Applied Sciences, 2020, 10(2):575.
[29] 郝建军,邴振凯,杨淑华,等.采用改进 YOLO v3 算法检测青皮核桃[J].农业工程学报,2022,38(14):183－190.
HAO Jianjun, BING Zhenkai, YANG Shuhua, et al. Detection of green walnut with improved YOLO v3 algorithm[J]. Transactions of the CSAE, 2022,38(14): 183－190. (in Chinese)
[30] SILVER D, LEVER G, HEESS N, et al. Deterministic policy gradient algorithms[C]// International Conference on Machine Learning, 2014:387－395.
[31] XIE Jiexin, SHAO Zhenzhou, LI Yue, et al. Deep reinforcement learning with optimized reward functions for robotic trajectory planning[J]. IEEE Access, 2019,7:105669－105679.
[32] 陈佳盼,郑敏华.基于深度强化学习的机器人操作行为研究综述[J].机器人,2022,44(2):236－256.
CHEN Jiapan, ZHENG Minhua. A review of robot manipulation behavior based on deep reinforcement learning[J]. Robot, 2022,44(2): 236－256. (in Chinese)
[33] BAC C W, ROORDA T, RESHEF R, et al. Analysis of a motion planning problem for sweet-pepper harvesting in a dense obstacle environment[J]. Biosystems Engineering, 2016,146:85－97.
[34] ANDRYCHOWICZ M, WOLSKI F, RAY A, et al. Hindsight experience replay[J]. Advances in Neural Information Processing Systems, 2017(12):5049－5059.
[35] 李鹤宇,赵志龙,顾蕾,等.基于深度强化学习的机械臂控制方法[J].系统仿真学报,2019,31(11):2452－2457.
LI Heyu, ZHAO Zhilong, GU Lei, et al. Control method for robotic arms based on deep reinforcement learning[J]. Journal of System Simulation, 2019,31(11): 2452－2457. (in Chinese)
[36] 王敏.基于深度强化学习的机械臂快速训练方法研究[D].郑州:河南大学,2023.
WANG Min. Research on fast training method of robotic arm based on deep reinforcement learning[D]. Zhengzhou: Henan University,2023. (in Chinese)
[37] 郝建军,魏文波,黄鹏程,等.油葵籽粒离散元参数标定与试验验证[J].农业工程学报,2021,37(12):36－44.
HAO Jianjun, WEI Wenbo, HUANG Pengcheng, et al. Calibration and experimental verification of discrete element parameters for sunflower seeds[J]. Transactions of the CSAE, 2021,37(12): 36－44. (in Chinese)