工程科学学报
Chinese Journal of Engineering

异构三机器人协同搬运的高柔顺性研究

张树忠 齐春雨 张弓 苏佳鸿 邱伟前 阮玉镇

High flexibility of heterogeneous tri-robot collaborative handling

ZHANG Shuzhong, QI Chunyu, ZHANG Gong, SU Jiahong, QIU Weiqian, RUAN Yuzhen

引用本文:
张树忠, 齐春雨, 张弓, 苏佳鸿, 邱伟前, 阮玉镇. 异构三机器人协同搬运的高柔顺性研究[J]. 北科大：工程科学学报, 2025, 47(10): 2049–2058. doi: 10.13374/j.issn2095–9389.2024.12.16.002
ZHANG Shuzhong, QI Chunyu, ZHANG Gong, SU Jiahong, QIU Weiqian, RUAN Yuzhen. High flexibility of heterogeneous tri–robot collaborative handling[J]. *Chinese Journal of Engineering*, 2025, 47(10): 2049–2058. doi: 10.13374/j.issn2095–9389.2024.12.16.002

在线阅读 View online: https://doi.org/10.13374/j.issn2095–9389.2024.12.16.002

工程科学学报, 第 47 卷, 第 10 期: 2049−2058, 2025 年 10 月
Chinese Journal of Engineering, Vol. 47, No. 10: 2049−2058, October 2025
https://doi.org/10.13374/j.issn2095-9389.2024.12.16.002; http://cje.ustb.edu.cn

异构三机器人协同搬运的高柔顺性研究

张树忠[1)], 齐春雨[1)], 张　弓[2,3)✉], 苏佳鸿[1,2)], 邱伟前[2)], 阮玉镇[1)]

1) 福建理工大学福建省智能加工技术及装备重点实验室, 福州 350108　2) 华南理工大学超级机器人研究院(黄埔), 广州 510700　3) 广东技术师范大学自动化学院, 广州 510665

✉通信作者, E-mail:gong_zhang@foxmail.com

摘　要　针对异构三机器人系统的协同搬运柔顺性问题, 提出基于近端策略优化(Proximal policy optimization)的强化学习控制方法. 在 CoppeliaSim 机器人仿真器中建立了异构三机器人协同搬运的仿真环境, 分别开展了力控制与强化学习控制的对比仿真. 仿真结果表明: 强化学习控制下, 物体质心的轨迹误差在 *Z* 方向上最优, 仅为力控制的 4.7%, 机器人 2 的末端速度变化和其典型关节的角速度变化更为平滑. 采用 sim2real 的方法, 将两种控制方法部署到三机器人协同搬运实验中. 实验结果表明: 强化学习控制下, *Z* 方向的物体轨迹跟踪误差同样最优, 仅为力控制的 5.4%. 机器人 2 在 *X* 方向上的速度变化仅为力控制的 20.7%, 其典型关节展现出更好的柔顺性, 角速度变化仅为力控制下的 35.2%. 仿真与实验结果表明: 强化学习的控制效果更优, 也具备从仿真到现实迁移的可行性.

关键词　异构的; 三机器人; 协同搬运; 强化学习; 柔顺性

分类号　TP242

High flexibility of heterogeneous tri-robot collaborative handling

ZHANG Shuzhong[1)], *QI Chunyu*[1)], *ZHANG Gong*[2,3)✉], *SU Jiahong*[1,2)], *QIU Weiqian*[2)], *RUAN Yuzhen*[1)]

1) Fujian Key Laboratory of Intelligent Machining Technology and Equipment, Fujian University of Technology, Fuzhou 350108

2) Institute for Super Robotics (Huangpu), South China University of Technology, Guangzhou 510700

3) School of Automation, Guangdong Polytechnic Normal University, Guangzhou 510665

✉Corresponding author, E-mail: gong_zhang@foxmail.com

ABSTRACT　This paper proposes a reinforcement learning (RL)-based control framework utilizing the proximal policy optimization (PPO) algorithm to address compliance issues in cooperative transportation tasks for heterogeneous tri-robot systems. The focus is on enhancing motion coordination and force adaptability in three heterogeneous robots during collaborative object transportation. A high-fidelity simulation environment was first constructed in the CoppeliaSim robotic simulator, where the tri-robot with distinct kinematic and dynamic configurations was programmed to collaboratively manipulate a shared object. Comparative simulations were conducted between traditional force control methods and the proposed RL-based approach to evaluate the robot performance in trajectory tracking accuracy, motion smoothness, and system compliance. Under the RL control framework, the PPO algorithm was trained to optimize the robots' joint actions by maximizing a reward function designed to penalize trajectory deviations, excessive contact forces, and abrupt velocity changes. The simulation results demonstrate that the RL-controlled system achieves remarkable improvements in vertical (*Z*-axis) trajectory tracking precision. Specifically, the trajectory error of the object's center of mass in the *Z*-direction was reduced to 4.7% of that observed under conventional force control. Furthermore, Robot 2—selected as a representative agent owing to its central role in the task—exhibited significantly smoother motion characteristics under RL control. Its end-effector velocity variations in the horizontal

收稿日期: 2024−12−16

基金项目: 国家自然科学基金资助项目(62073092); 福建省智能加工技术及装备重点实验室开放基金项目资助项目(KF-01-22005); 广东省自然科学基金资助项目(2021A1515012638); 福建省自然科学基金资助项目(2025J01985)

(*X*–*Y*) plane were attenuated by 82% compared to force control, while angular velocity fluctuations in its primary rotational joint were reduced to 35% of the baseline values, indicating enhanced mechanical compliance and reduced oscillatory behavior. To validate the real-world applicability of the learned policies, a sim2real transfer methodology was implemented. The control strategies were deployed on a physical tri-robot platform comprising one six-degrees-of-freedom (DOF) industrial manipulator and two customized four-DOF collaborative robots, tasked with synchronously transporting a deformable payload. The experimental results agreed with simulation predictions: the RL-based controller maintained superior *Z*-direction trajectory tracking performance, limiting errors to 5.4% of those under force control. Robot 2's motion compliance showed further improvement in physical experiments, with its *X*-direction velocity variations reduced to 20.7% of the force control benchmark. Critical joint-level analyses revealed that the angular velocity variations of Robot 2's third joint—a pivotal component for vertical motion compensation—were suppressed to 35.2% of the force control values, confirming the RL controller's ability to mitigate mechanical vibrations and adapt to dynamic payload interactions. The study also investigates the robustness of the RL framework to real-world uncertainties, including sensor noise, communication latency, and payload deformation. Despite these challenges, the RL controller maintained stable performance, achieving a 92% reduction in peak contact forces compared to force control during sudden payload shifts. Statistical analyses of motion data further indicated that the RL-based system reduced the standard deviation of inter-robot coordination errors by 76% and 68% in simulation and physical experiments, respectively, underscoring its consistency across domains. Both simulation and experimental findings conclusively demonstrate that the PPO-based RL framework not only surpassed traditional force control in precision and compliance but also successfully bridged the sim2real gap. The framework's ability to learn adaptive policies in simulation and transfer them to physical robots with minimal fine-tuning highlights its potential for deployment in industrial applications requiring heterogeneous multi-robot collaboration. This work advances the field of compliant robotic control by providing a scalable, data-driven solution that harmonizes trajectory accuracy, motion smoothness, and real-world adaptability in complex cooperative tasks.

KEY WORDS heterogeneous; tri-robot; collaborative handling; reinforcement learning; flexibility

面向机器人技术的典型应用场景, 如大尺寸或者大负载的物体搬运, 目前主要由单个机器人来执行, 其能力受到其固有机制和预定义程序的限制. 采用多个机器人之间的协同配合来执行任务, 能够可靠地完成单个机器人无法完成的高精度作业[1]. 这种分布式操作方法可以通过冗余提高其鲁棒性, 并通过使用多个简单的机器人, 而不是单个强大的机器人来降低成本. 多机器人协同可广泛用于机器人协同搬运、装配、冲压、焊接等领域, 具有高灵活性和环境适应性等特点[2], 已成为构建智能无人生产线的研究热点.

目前, 已有不少关于多机器人协同作业的运动控制和路径规划研究. Solanes 等[3] 采用滑模方法设计机器人的力控制系统, 提出基于任务优先级的机器人位置力混合控制方法. 毛欢[4] 针对双机器人协同夹持物体与外界环境接触的力控作业任务, 提出内外双环阻抗控制策略, 实现了双机器人协作系统的内外力跟踪. 段晋军[5] 提出自适应变阻抗双臂力/位协调控制方法, 完成了空间复杂焊缝的多机器人协同焊接任务. 苏牧青等[6] 针对多无人车围捕问题, 提出了基于 SAC 算法的协同围捕算法, 通过加入长短期记忆及注意力机制提升了多设备协同效率.

Lan 等[7] 针对智能制造领域多机器人系统协调拾取与放置, 引入深度强化学习算法优化多机器人协同拾取和放置系统, 仿真结果表明该方法能有效提升生产效率. Perrusquia 等[8] 提出基于强化学习的阻抗控制方法, 对阻抗控制下, 机器人搬运作业生成的期望力进行学习, 从而实现机器人的柔顺控制. Roveda 等[9] 提出将强化学习(Q 学习)算法用于预测与调整机器人的阻抗控制参数, 提高了机器人协同作业中的力柔顺控制. Zhang 等[10] 提出基于强化学习的双机器人力/位多元数据驱动方法, 主机器人采用理想位置元控制, 通过强化学习算法来学习期望位置; 从机器人采用基于主机器人位置偏差的力元控制, 通过强化学习算法来学习期望作用力, 可解决力/位控制中的参数优化问题.

Liu 等[11] 通过近端策略优化算法(Proximal policy optimization, PPO)控制框架, 仿真验证该方法的有效性, 为异构多机器人协作任务提供了数据驱动解决方案, 并引入非线性扩张状态观测器(NNESO)以提升系统抗干扰能力[12], 并通过将分数阶多智能体系统(FOMASs)从单积分拓展至多积分动态[13], 为多设备协同的实际场景提供了理论支持.

多机器人协同搬运同一个物体时, 各机器人之间具有物理链接和内力约束, 要实现紧耦合必

须通过实施有效的力—位置协同控制策略, 从而有效分配载荷, 进而提升多机器人协同作业的柔顺性[14]. 为此, 本文聚焦异构多机器人系统[15], 即具有不同品牌、不同构型和不同能力的多个机器人, 面向三机器人协同搬运场景, 通过两种控制方法的仿真分析与实验研究对比, 来解决更广泛的异构多机器人协同搬运的柔顺性问题.

1 力控制策略

力控制策略的核心是通过力传感器测量机器人与环境之间的力和力矩, 并将其作为控制输入进行实时调整. 通过感知和响应外部力的大小和方向, 使机器人适应不同的工作环境和任务需求.

力控制策略架构如图 1 所示, f_d为期望接触力, f_s为机器人末端力传感器测量力, f_e为力跟踪误差. 力控制策略根据力跟踪误差产生期望轨迹x_d, 通过机器人逆运动学关系, 输出期望关节角度q_d到机器人进行运动, 机器人输出关节位置q.

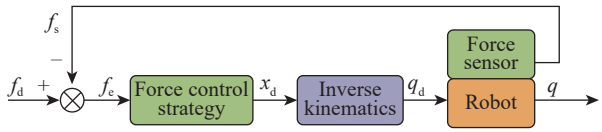

图 1 力控制策略基本架构

Fig.1 Basic structure of force control strategy

2 基于 PPO 的强化学习控制策略

强化学习[16](Reinforcement learning, RL)因其强大的探索能力与自主学习能力, 在机器人控制[17]领域应用广泛. 多机器人协同搬运的强化学习的基本架构, 如图 2 所示.

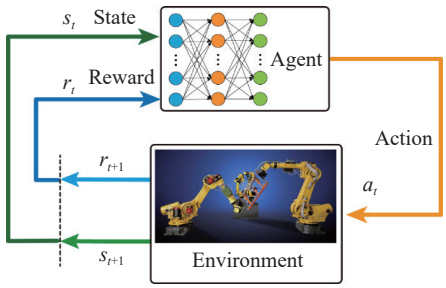

图 2 多机器人协同搬运的强化学习基本架构

Fig.2 Basic structure of reinforcement learning for multirobot collaborative handling

强化学习过程中, 智能体与环境一直交互. 在某一时刻t, 智能体从环境中获得到该时刻下状态信息(States)s_t, 智能体会根据该状态输出动作(Action)a_t作用于环境, 环境会根据被执行的动作a_t的输出下一个状态s_{t+1}, 并将当前动作的奖励(Reward)r_t反馈给智能体. 智能体的目的就是不断调整自身决策, 尽可能多的获取奖励.

PPO[18] 由 OpenAI 在 2017 年提出, 并将其作为强化学习的 baseline 算法[19].

近端策略优化 PPO 算法基于 Actor−Critic 架构, 由策略梯度算法发展而来, 其 Critic 网络采用时间差分法对$t \in [0,N]$值函数进行估计:

$$A_t = \delta_t + (\gamma\lambda)\delta_{t+1} + \cdots + \cdots + (\gamma\lambda)^{N-t+1}\delta_{t-1} \tag{1}$$

$$\delta_t = r_t + \gamma V(s_{t+1}) - V(s_t) \tag{2}$$

式中: A_t为t时刻下的状态−动作对的优势(Advantage), γ为折扣因子, λ为衰减因子, δ_t为时间差分误差(TD-Error), r_t为t时刻下执行当前动作的奖励, $V(s_t)$为t时刻下价值函数, δ_t为r_t与价值函数与折扣因子的乘积$\gamma V(s_{t+1})$跟价值函数$V(s_t)$的差值$\gamma V(s_{t+1}) - V(s_t)$的和.

PPO 算法中的 Actor 网络为 On-Policy 算法, 引入重要性采样(Importance Sampling), 即可以使用θ'采样到的数据去训练θ, 这样就可以实现采样一次, 更新多次[20].

为满足重要性采样条件, PPO 算法需要对策略进行约束保证其差异很小. 对策略的约束多采用裁剪方式进行, 可通过设置调节超参数, 通过对超出限定前后策略差异上限范围的差异进行裁剪基于 PPO 网络结构的多机器人协同搬运, 如图 3 所示.

文中所述的多机器人强化学习控制系统设定, 主要包括: 状态空间、动作空间以及奖励函数. 状态空间为机器人提供清晰和高效的决策基础, 直接关系到强化学习算法学习效率和效果. 为了分析多机协同搬运柔顺性问题, 本文将机器人的各关节角度、末端位置、关节角速度、关节角加速度、末端速度、末端加速度和末端 6 维力数据作为观测值.

动作空间作为强化学习动作神经网络的输出结果, 主要对机器人进行运动控制. 动作空间首先要提供实现预期目标的可能性, 避免在任务空间中出现无法到达的奇异点. 另外动作空间应该尽量简单, 降低训练的难度, 提升算法的性能.

奖励函数作为被控对象运行状态的检验, 是设定的评价控制策略好坏的标准. 奖励函数的值与算法控制效果正相关, 即算法控制越好, 奖励函数的值越大. 在多机器人协同搬运中, 首先要保证机器人能够完成完整的搬运轨迹, 轨迹完成度越

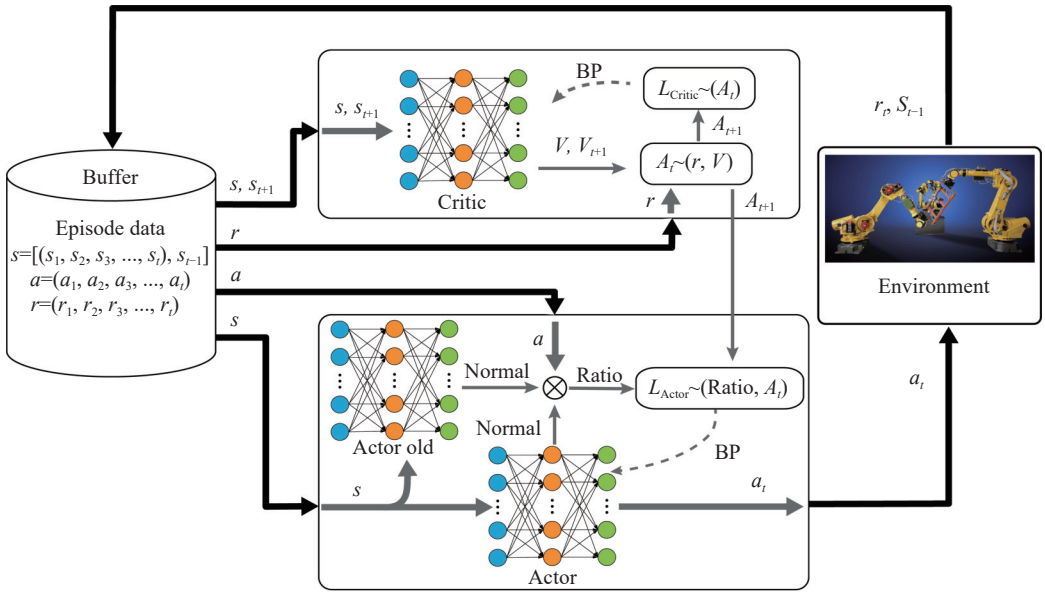

图 3　基于 PPO 算法的多机器人协同搬运

Fig.3　PPO algorithm structure for multi-robot collaborative handling

高, 获得的奖励越大, 如式 (3) 所示:

$$\text{reward} = r_{\text{Traj}} + r_{\text{Pos}} + r_{\text{v}} + r_{\omega} + r_{\text{force}} \tag{3}$$

式中, r_{Traj}为轨迹完成度奖励; r_{Pos}为位置奖励函数, 与机器人末端位置差值相关; r_{v}为机器人末端速度奖励函数, 与机器人末端速度相关, 反应运动平稳性; r_{ω} 为关节速度奖励函数, 与机器人各关节速度变化相关, 表征机器人运动平顺性; r_{force}为内力控制奖励函数.

除了轨迹完成度奖励为正外, 其余均为负值. 正轨迹完成度为期望值, 有利于引导机器人实现完整轨迹. 其余指标为非期望值, 负值(惩罚)有利于机器人快速收敛.

通过设置合适的随机噪声, 不仅可以模拟真实环境, 从而提升算法的实验可迁移性, 还能提升强化学习算法训练后的鲁棒性[21]. 因此本文添加两种噪声. 一是力信号噪声, 二是延迟噪声. 基于被搬运物体的受力在 10 N 左右, 因此按受力 5% 大小选择力信号噪声在 [–0.25 N, 0.25 N] 内随机取值. 自上位机发送动作空间内最长位置移动信号计时, 到机器人完成运动结束计时, 计时时长为 100 ms 左右, 因此延迟噪声根据动作空间内选择概率以 50%、30%、20% 的概率选择 0、50、100 ms 的延迟.

3　三机器人协同搬运仿真分析

3.1　仿真环境设置

采用 Gym[22] 创建机器人强化学习环境, 调用 SB3(Stable Baselines 3)部署 PPO 算法, PPO 算法的主要超参数设置如下, 学习率为 3×10^{-4}, γ=0.99, clip_range=0.2, 训练总时长 179 h. 通过通讯接口实现 Gym 环境与机器人仿真器 CoppeliaSim[23] 进行通讯与控制. 采用 Bullet Physics v2.83 物理模拟引擎作为机器人的动力学模拟计算[24], 并使用 CoppeliaSim 的逆运动学模块进行运动规划.

在 CoppeliaSim 中建立异构三机器人协同搬运的仿真环境, 如图 4 所示. 设置三个异构机器人共同夹持一块边长为 30 cm, 厚度 1.5 cm, 质量为 1 kg 的正三角体. 机器人 2 和机器人 3 的末端分别加装力传感器, 其末端最大速度约为 0.7 m·s^{-1}, 同时设定控制周期为 0.05 s.

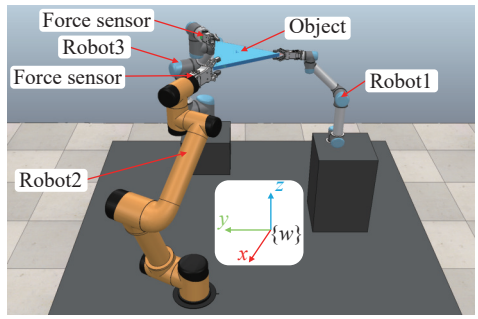

图 4　三机器人协同搬运仿真环境

Fig.4　Tri-robot collaborative handling simulation scenario

设置被搬运物体质心的期望轨迹, 如图 5 所示. 轨迹的前半段为矩形下降, 后半段为圆形上升, 从而检验三机器人协同搬运的柔顺性.

3.2　仿真结果及分析

仿真中, 设定机器人 1 采用位置控制, 机器人

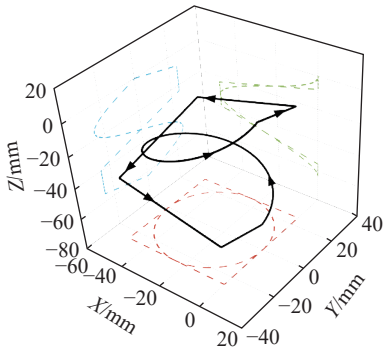

图 5 三机协同搬运的期望轨迹

Fig.5 Desired trajectory of tri-robot collaborative handling

2 和机器人 3 对比采用力控制和强化学习两种控制策略. 鉴于三机器人之间的力关联复杂, 而且各个机器人与被搬运物体之间并非刚性约束, 力控制下无法完成期望轨迹. 因此, 在轨迹前半部分对机器人 2 和机器人 3 末端力传感器采集的数据, 添加不大于 1 的安全系数, 从而降低因非运动方向受力对运动的影响, 后半段则取消该系数.

限于篇幅, 本文仅展示部分仿真分析结果, 即在不同控制策略下, 被搬运物体的质心位置曲线(图 6), 机器人 2 的末端速度曲线(图 7)和典型关节的角速度曲线(图 8), 以及三机器人协同搬运的仿真结果对比(表 1).

具体地, 两种控制方式下, 针对被搬运物体的质心位置与期望轨迹如图 6 所示. 可以看出, 强化学习下质心位置变化最为平顺, 与期望轨迹最接近.

X 方向上, 力控制的轨迹误差相对于期望轨迹, 在 0 ~ 10 s, 由 2% 逐渐增加到约 16.2%; 10 ~ 17 s, 在 8% ~ 30% 内波动; 17 s 之后逐步从约 2% 扩大到 40% 以上. 而强化学习控制的轨迹误差, 仅在 2 ~ 7 s 和 25 s 之后为 2% ~ 3%, 其余均在 2% 以下. 总之, 强化学习的最大轨迹误差仅为力控制下的 7.1%, 力控制下最大误差 25.5 mm, 平均误差 8.4 mm, 强化学习下最大误差 1.8 mm, 平均误差 0.7 mm.

Y 方向上, 0 ~ 15 s, 力控制的轨迹误差由最大约 38% 下降到约 5%; 15 s 之后轨迹误差从 0.2% ~ 20% 内波动逐步增加到 40%, 之后逐步衰减到 10% 以内. 而强化学习的轨迹误差在 6 ~ 10 s 和 22 ~ 25 s 时在 2% ~ 3% 内波动, 其余均在 1% 以内, 其最大值仅为力控制的 8.4%, 力控制下最大误差 20.2 mm, 平均误差 8.8 mm, 强化学习下最大误差 1.7 mm, 平均误差 0.8 mm.

Z 方向上, 力控制下的轨迹误差从约 2% 逐步

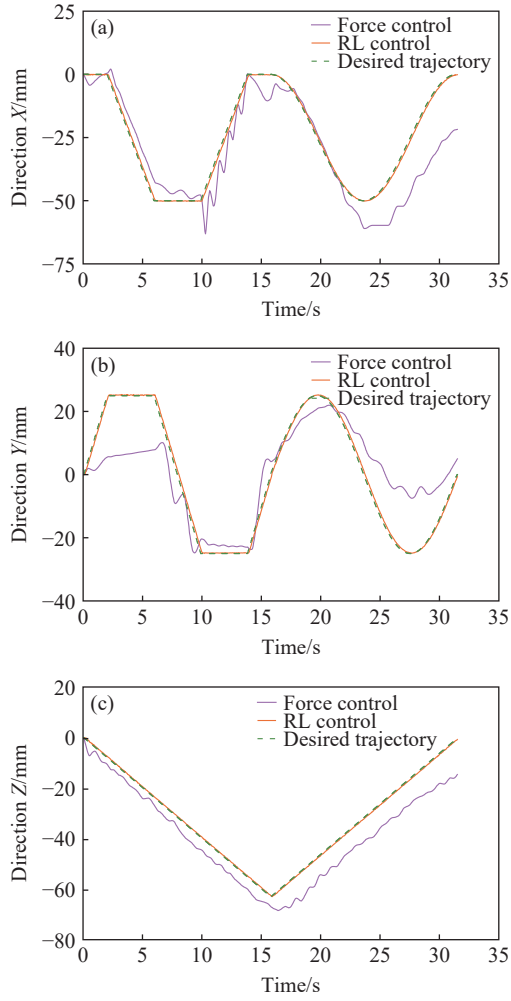

图 6 被搬运物体质心位置仿真图. (a) *X* 方向; (b) *Y* 方向; (c) *Z* 方向

Fig.6 Simulated centroid position of the handled object: (a) direction *X*; (b) direction *Y*; (c) direction *Z*

增加到约 23%. 而强化学习的轨迹误差均在 1% 以内, 其幅值仅为力控制下的 5.4%, 效果更明显, 力控制下最大误差 14.9 mm, 平均误差 7.1 mm, 强化学习下最大误差 0.8 mm, 平均误差 0.4 mm

对于机器人 2 的末端速度变化幅度, 从图 7 可知, 强化学习的速度变化最平缓, 控制效果更优. 具体地, 在 *X* 方向上, 强化学习的速度最大变化幅度仅为力控制下的 20.7%, 强化学习下与期望速度平均误差为 2.4 mm·s^{-1} 优于力控制下 7.0 mm·s^{-1}; *Y* 方向上, 强化学习的速度最大变化幅度仅为力控制的 35.9%, 强化学习下与期望速度平均误差为 2.0 mm·s^{-1} 优于力控制下 5.3 mm·s^{-1}; *Z* 方向上, 强化学习下与期望速度平均误差为 2.2 mm·s^{-1} 优于力控制下 2.7 mm·s^{-1}, 虽然强化学习下的速度最大变化幅度仅为力控制的 63.9%, 但需要指出的是, 其仅出现在物体运动的初始阶段, 且后续运动的速度变化仍然展现出良好的优势.

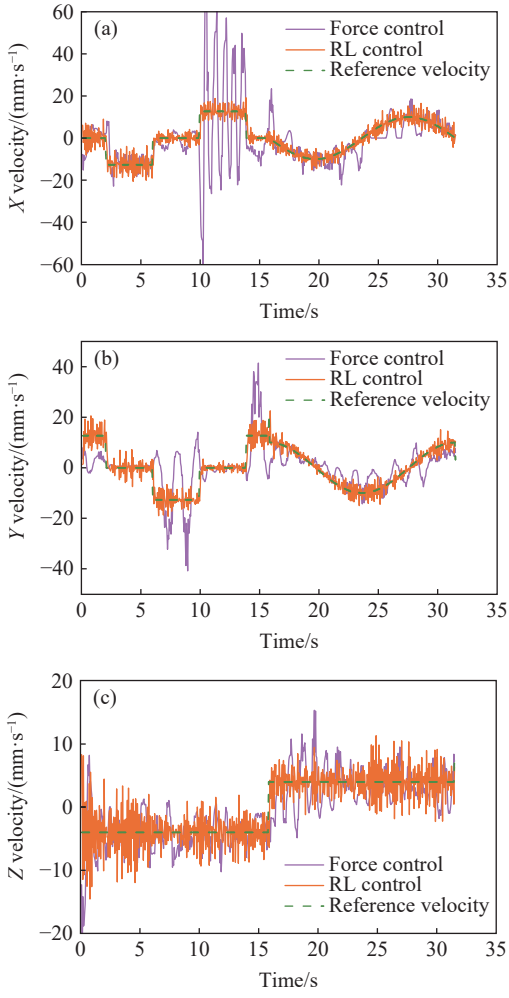

图 7　机器人 2 的末端速度仿真图. (a) *X* 方向; (b) *Y* 方向; (c) *Z* 方向

Fig.7　Simulated velocity of robot 2's end effector: (a) direction *X*; (b) direction *Y*; (c) direction *Z*

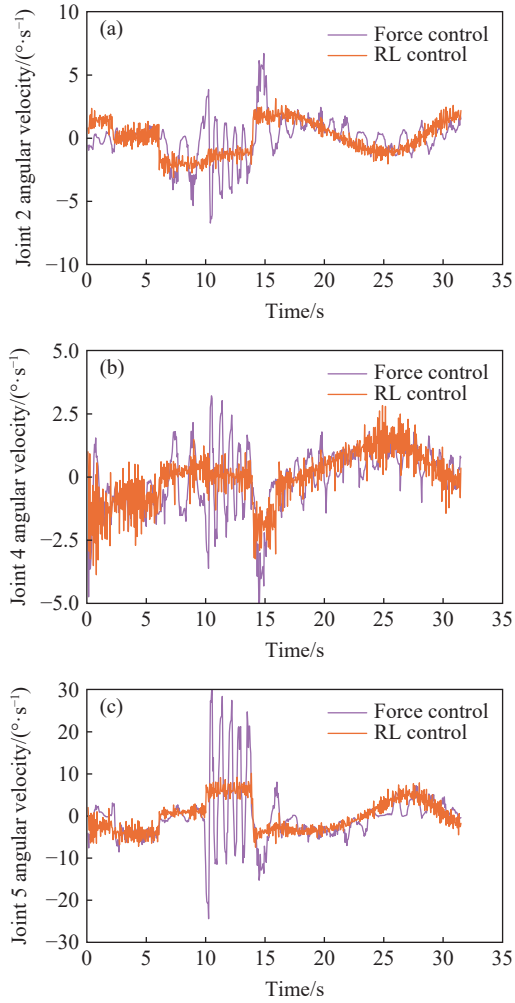

图 8　机器人 2 典型关节角速度仿真图. (a) 关节 2; (b) 关节 4; (c) 关节 5

Fig.8　Simulated angular velocity of robot 2's typical joint: (a) joint 2; (b) joint 4; (c) joint 5

表 1　三机器人协同搬运仿真结果对比

Table 1　Comparison of simulation results of tri-robot collaborative handling

Method	Position error of object/mm			Position average error of object/mm			Fluctuation of robot 2 terminal velocity/(mm·s^{-1})			Fluctuation of robot 2 of average error terminal velocity/(mm·s^{-1})			Fluctuation of angular velocity of robot 2/(°·s^{-1})		
	X	*Y*	*Z*	*X*	*Y*	*Z*	*X*	*Y*	*Z*	*X*	*Y*	*Z*	Joint 2	Joint 4	Joint 5
Force control	25.5	20.2	14.9	8.4	8.8	7.1	59.8	42.6	19.7	7.0	5.3	2.7	7.1	4.9	29.9
Reinforcement learning	1.8	1.7	0.8	0.7	0.8	0.4	12.4	15.3	12.6	2.4	2.0	2.2	2.5	2.6	11.2

关节角速度的变化同样是三机器人协同搬运系统柔顺性的重要指标, 针对机器人 2, 以典型关节 2、4、5 为例. 从图 8 所示的关节数据可以看出, 力控制下的关节角速度, 大幅变化持续存在于整个搬运过程. 强化学习控制下, 无论哪个关节都展现出更好的柔顺性, 特别是对于关节 2 的速度变化, 强化学习的速度变化在 0.9°·s^{-1} ~ 2.5°·s^{-1} 之间, 力控制的速度变化在 0.6°·s^{-1} ~ 7.1°·s^{-1} 之间, 由此可见强化学习的最大幅度仅为力控制下的 35.2%.

4　三机器人协同搬运实验研究

4.1　实验平台搭建

当前, 强化学习从仿真到现实的 Sim-to-Real 问题, 尚未有简而有效的方法, 其迁移效果较差. 为能在现实中复现其效果, 本文首先依照仿真环境搭建实验平台, 然后采用行为克隆[25](Behavior cloning)仿真中的信息, 再映射到执行动作信息中, 从而共同生成有效的控制信息, 以对强化学习

多机协同算法进行可行性验证. 具体方法为当实际输出与期望轨迹进行实时比对, 当数值差距过大时切换仿真数据驱动. 所搭建的三机器人协同搬运实验平台, 如图 9 所示.

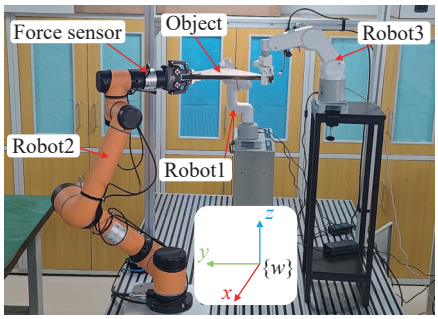

图 9 三机器人协同搬运实验平台

Fig.9 Tri-robot collaborative handling experimental platform

在上述实验平台中, 机器人 1 采用大象六自由度协作机器人(ER) myCobot 280, 末端负载 0.25 kg, 重复定位精度±0.5 mm; 机器人 2 采用遨博(AUBO)六自由度协作机器人 i5, 末端负载 5 kg, 重复定位精度±0.02 mm; 机器人 3 采用大象四自由度协作机器人(ER) myPalletizer, 末端负载 0.25 kg, 重复定位精度±0.5 mm; 采用坤维科技的六维力传感器(KWR75B)实时采集机器人 2 的末端接触力; 通过上位机程序, 机器人的实时控制周期, 经实验测试约为 500 ms.

用于实验的被搬运物体与仿真中一致, 均为边长 30 cm、厚度 1.5 cm、质量 1 kg 的正三角体.

4.2 实验结果及分析

因实验条件及机器人兼容性所限, 对机器人 3 采用仿真数据驱动. 同样限于篇幅, 仅展示部分实验研究结果, 即主要分析被搬运物体和机器人 2 在力控制与强化学习下的代表性实验结果.

两种控制方式下, 被搬运物体的质心位置与期望轨迹, 如图 10 所示.

由图 10 可知, X 方向上, 0 ~ 50 控制周期内, 力控制的轨迹误差从 4% 逐渐增加到 17.4%. 50 ~ 200 控制周期后逐步减小到 5% 内, 200 ~ 400 控制周期内在 1% ~ 30% 内波动, 400 控制周期后逐步扩大到 50% 左右. 而强化学习的轨迹误差在 40 ~ 120 和 500 ~ 630 控制周期内的误差在 3% ~ 4%, 200 ~ 280 控制周期内的误差约为 5%. 可以看出, 强化学习的位置变化更为平顺, 与期望轨迹跟踪误差在 2.4 mm 内, 仅为力控制的 9.3%, 力控制下最大误差 25.9 mm, 平均误差 8.6 mm, 强化学习下最大误差 2.4 mm, 平均误差 1.1 mm, 效果非常明显.

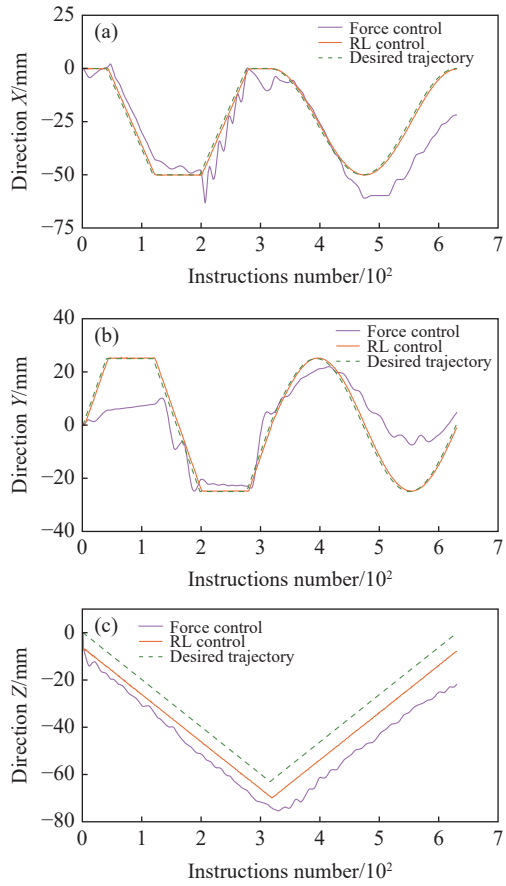

图 10 被搬运物体质心位置实验图. (a) X 方向; (b) Y 方向; (c) Z 方向

Fig.10 Measured centroid position of the handled object: (a) direction X; (b) direction Y; (c) direction Z

Y 方向上, 0 ~ 40 控制周期内, 轨迹误差逐步增加到 39%. 40 ~ 120 控制周期内, 轨迹误差在 30% 以上, 120 ~ 400 控制周期的轨迹误差在 10% ~ 30% 内波动. 400 控制周期后, 从 2% 逐步增加到 40%, 随后衰减到约 10%. 而强化学习下在 0 ~ 40 控制周期的轨迹误差约为 3% ~ 5%, 120 ~ 200 控制周期内高于 4%, 280 ~ 360 控制周期内高于 3%, 其余控制周期误差在 2% 以内, 力控制下最大误差 20.3 mm, 平均误差 8.8 mm, 强化学习下最大误差 2.3 mm, 平均误差 1.1 mm. 可以看出, 强化学习下位置变化相对小, 其幅值仅为力控制的 11.3%.

Z 方向上, 力控制下轨迹误差由 10% 逐步增加到约 35%. 而强化学习下轨迹误差维持在 9% ~ 13%. 力控制下最大误差 22.4 mm, 平均误差 14.0 mm, 强化学习下最大误差 8.1 mm, 平均误差 7.0 mm 总之, 强化学习下位置变化相对较多, 其幅值为力控制的 36.2%, 但也展现出一定的有效性.

另外, 两种控制方式下, 机器人 2 的末端速度实验对比, 如图 11 所示.

从图 11 可知, 强化学习的速度变化最平缓, 控

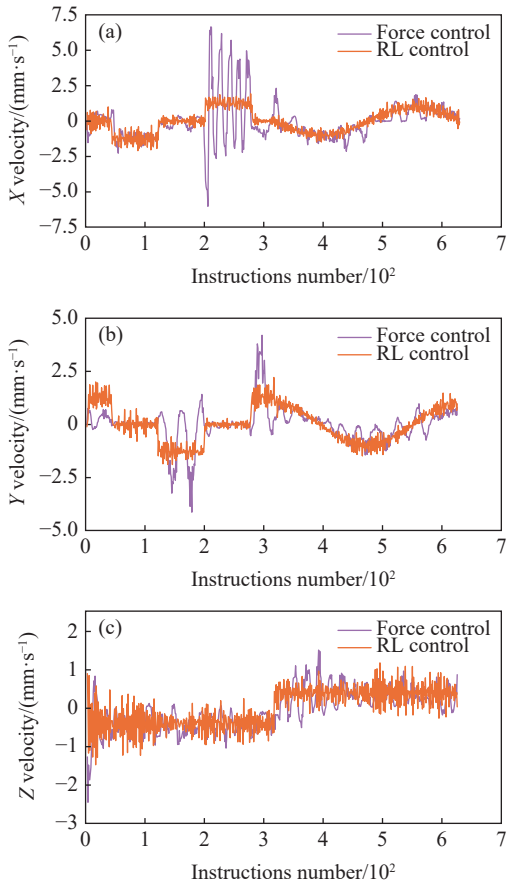

图 11　机器人 2 的末端速度实验图. (a) *X* 方向; (b) *Y* 方向; (c) *Z* 方向

Fig.11　Measured velocity of robot 2's end effector: (a) direction *X*; (b) direction *Y*; (c) direction *Z*

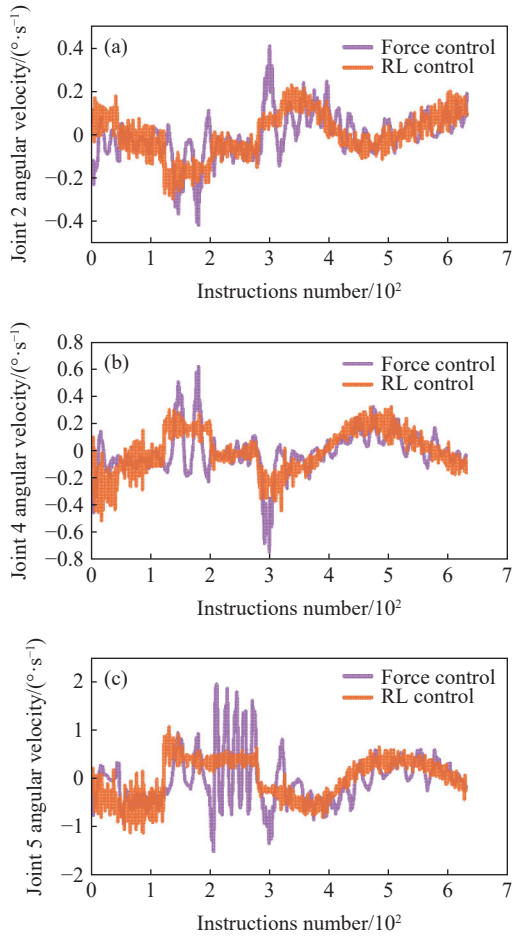

图 12　机器人 2 典型关节角速度实验图. (a) 关节 2; (b) 关节 4; (c) 关节 5

Fig.12　Measured angular velocity of robot 2's typical joint: (a) joint 2; (b) joint 4; (c) joint 5

制效果更优. 具体地, *X* 方向上, 强化学习的速度变化范围普遍在 2.3 mm·s^{-1} 以内, 力控制下的速度变化在 0.9 ~ 7.2 mm·s^{-1} 之间. 由此可见, 强化学习的速度最大变化幅度仅为力控制下的 31.9%.

Y 方向上, 强化学习的速度最大变化幅度仅为力控制下的 60.61%; *Z* 方向上, 强化学习的速度变化在 1.4 mm·s^{-1} 以内, 力控制的速度变化在 2.6 mm·s^{-1} 以内. 虽然强化学习下的速度最大变化幅度仅为力控制的 53.8%, 但需要指出的是, 其仅出现在物体运动的初始阶段.

同样的, 选择机器人 2 的典型关节 2、4、5 进行分析. 两种控制方式下机器人 2 关节数据对比, 如图 12 所示. 从图 12 所示的关节数据可以看出, 力控制下的关节角速度, 大幅度振荡变化, 强化学习控制下, 无论哪个关节都展现出更好的柔顺性. 特别是关节 2, 强化学习的速度变化在 0.08°·s^{-1} ~ 0.1°·s^{-1} 之间, 力控制的速度变化在 0.12°·s^{-1} ~ 0.4°·s^{-1} 之间, 强化学习下速度变化普遍是力控制下 47% 左右, 其最大幅度仅为力控制下的 25.0%.

三机器人协同搬运的实验结果, 对比于表 2 所示. 从表中可以看出, 强化学习的控制效果更优.

表 2　三机器人协同搬运实验结果对比

Table 2　Comparison of experiment results of tri-robot collaborative handling

Method	Position error of object/mm			Position average error of object/mm			Fluctuation of robot 2 terminal velocity/(mm·s^{-1})			Fluctuation of robot 2 of average error terminal velocity/(mm·s^{-1})			Fluctuation of angular velocity of robot 2/(°·s^{-1})		
	X	*Y*	*Z*	*X*	*Y*	*Z*	*X*	*Y*	*Z*	*X*	*Y*	*Z*	Joint 2	Joint 4	Joint 5
Force control	25.9	20.3	22.4	8.6	8.8	14.0	7.2	3.3	2.6	0.7	0.5	0.3	0.4	0.6	1.9
Reinforcement learning	2.4	2.3	8.1	1.1	1.1	7.0	2.3	2.0	1.4	0.3	0.2	0.2	0.1	0.3	1.1

5 结论

面向异构三机器人协同搬运柔顺性问题, 提出了基于近端策略优化(PPO)的强化学习控制方法, 开展了两种控制方法(力控制与强化学习控制)的异构三机器人协同搬运仿真分析与实验研究, 实现了异构多机器人协同搬运的柔顺性提升.

基于强化学习 PPO 算法, 设计出三机器人协同搬运的强化学习控制策略, 通过建立 Gym 与 CoppeliaSim 仿真环境, 分别进行了力控制和强化学习控制的对比仿真. 仿真结果表明强化学习控制效果更优.

通过搭建异构三机器人协同搬运的实验平台, 将力控制与强化学习控制部署到实验平台上进行可行性验证. 实验结果表明强化学习在本文搭建的实验条件和场景下多机协同搬运中的有效性, 复杂条件及多场景、复杂任务中仍需继续探索.

下一步研究, 将强化学习方向提出的新算法应用到机器人控制领域, 并对比分析不同强化学习算法的有效性; 在强化学习迁移到现实空间方向上可以结合数字孪生, 使用虚拟空间数据训练, 物理空间控制应用. 特别地, 将丰富异构多机器协同搬运的实验对象、实验场景. 尤其是搭建与仿真环境完全一致的实验场景, 从而增强研究分析结果的可比性.

参 考 文 献

[1] Qu D K, Tan D L, Zhang C J, et al. A control system for two robots coordination. *Robot*, 1991, 13(3): 6
(曲道奎, 谈大龙, 张春杰, 等. 双机器人协调控制系统. 机器人, 1991, 13(3): 6)

[2] Zhang S Z, Zhu Q, Zhang G, et al. Intelligent human–robot collaborative handover system for arbitrary objects based on 6D pose recognition. *Chin J Eng*, 2024, 46(1): 148
(张树忠, 朱祺, 张弓, 等. 基于 6D 位姿识别面向任意物体的智能人–机协同递送. 工程科学学报, 2024, 46(1): 148)

[3] Solanes J E, Gracia L, Muñoz-Benavent P, et al. Robust hybrid position-force control for robotic surface polishing. *J Manuf Sci Eng*, 2019, 141(1): 011013

[4] Mao H. *Research on Force Compliance Control for Multi-Robot Collaborative Grinding* [Dissertation]. Harbin: Harbin Institute of Technology, 2021
(毛欢. 面向多机器人协同打磨的力柔顺控制研究[学位论文]. 哈尔滨: 哈尔滨工业大学, 2021)

[5] Duan J J. *Trajectory Planning and Position Force Coordination Control in Multi-Robot Cooperative Welding Process* [Dissertation]. Nanjing: Southeast University, 2019
(段晋军. 多机器人协作焊接中的轨迹规划和位置力协调控制研究[学位论文]. 南京: 东南大学, 2019)

[6] Su M Q, Wang Y, Pu R M, et al. Cooperative encirclement method for multiple unmanned ground vehicles based on reinforcement learning. *Chin J Eng*, 2024, 46(7): 1237
(苏牧青, 王寅, 濮锐敏, 等. 基于强化学习的多无人车协同围捕方法. 工程科学学报, 2024, 46(7): 1237)

[7] Lan X, Qiao Y S, Lee B. Coordination of a multi robot system for pick and place using reinforcement learning // 2022 2*nd International Conference on Computers and Automation* (*CompAuto*). Paris, 2022: 87

[8] Perrusquía A, Yu W, Soria A. Position/force control of robot manipulators using reinforcement learning. *Ind Robot Int J Robot Res Appl*, 2019, 46(2): 267

[9] Roveda L, Testa A, Ali Shahid A, et al. Q-Learning-based model predictive variable impedance control for physical human-robot collaboration. *Artif Intell*, 2022, 312: 103771

[10] Zhang G, Wu Y Y, Zhu Q, et al. *Dual-Robot Position/Force Multivariate-Data-Driven Method Using Reinforcement Learning*: US Patent, US17751024. 2024−09−18

[11] Liu J, Li P, Chen W, et al. Distributed formation control of fractional-order multi-agent systems with relative damping and nonuniform time-delays. *ISA Trans*, 2019, 93: 189

[12] Liu J, Tan J H, Li H B, et al. Active disturbance rejection consensus control of multi-agent systems based on a novel NESO. *IEEE/ASME Trans Mechatron*, 2025, 30(1): 634

[13] Liu J, Chen W, Qin K Y, et al. Consensus of multi-integral fractional-order multiagent systems with nonuniform time-delays. *Complexity*, 2018, 2018(1): 8154230

[14] Wang X Y, Wu Y Y, Zhang G, et al. Three-dimensional trajectory planning for multi-robot collaboration in complex components with heterogeneous materials // 2023 3*rd International Conference on Robotics, Automation and Artificial Intelligence* (*RAAI*). Singapore, 2023: 134

[15] Rizk Y, Awad M, Tunstel E W. Cooperative heterogeneous multi-robot systems. *ACM Comput Surv*, 2020, 52(2): 1

[16] Sutton R S, Barto A G. Reinforcement learning: An introduction. *IEEE Trans Neural Networks*, 1998, 9(5): 1054

[17] Duan Y, Li C, Xie M C. One fast RL algorithm and its application in mobile robot navigation // 2012 *IEEE International Conference on Computer Science and Automation Engineering* (*CSAE*). Zhangjiajie, 2012: 552

[18] Fan D D, Ding S, Zhang H T, et al. A novel proximal policy optimization approach for filter design. *ACES Journal*, 2024: 390

[19] Yao S Y, Chen G D, Pan L F, et al. Multi-robot collision avoidance with map-based deep reinforcement learning // 2020 *IEEE 32nd International Conference on Tools with Artificial Intelligence* (*ICTAI*). Baltimore, 2020: 532

[20] Luo J L, Li H. A learning approach to robot-agnostic force-guided high precision assembly // 2021 *IEEE/RSJ International Conference on Intelligent Robots and Systems* (*IROS*). Prague,

2021: 2151

[21] Wang Y Z. *Research on Robotic Grasping and Intelligent Assembly Based on Deep Reinforcement Learning* [Dissertation]. Shenyang: Shenyang University of Technology, 2022
(王永志. 基于深度强化学习的机器人抓取及智能装配研究[学位论文]. 沈阳: 沈阳工业大学, 2022)

[22] Le Gléau T, Marjou X, Lemlouma T, et al. A multi-agent OpenAI gym environment for telecom providers cooperation // 2021 24*th Conference on Innovation in Clouds, Internet and Networks and Workshops* (*ICIN*). Paris, 2021: 28

[23] Kyrylovych V, Kravchuk A, Dobrzhanskyi O, et al. Automation of the process of attestation of metrics for industrial robots using software products CoppeliaSim and MATLAB. *Eng Proc*, 2024, 70(1): 9

[24] Julius Fusic S, Ramkumar P, Hariharan K. Path planning of robot using modified dijkstra Algorithm // 2018 *National Power Engineering Conference* (*NPEC*). Madurai, 2018: 1

[25] Ly A O, Akhloufi M. Learning to drive by imitation: An overview of deep behavior cloning methods. *IEEE Trans Intell Veh*, 2020, 6(2): 195