基于 PIRL 的空间机械臂仿生智能抓取方法

李连鹏 [1] 郭 航 [1] 李明洋 [2, 3] 张海博 [2, 3] 徐拴锋 [2, 3] 张冬浩 [1]

摘 要 针对空间机械臂在微重力环境下执行漂浮目标自主抓取任务时存在的样本获取难、泛化能力弱、动态扰动适应差的问题, 提出一种融合仿生智能的渐进式模仿强化学习方法. 首先, 基于遥操作采集的人类臂手协同操作专家示范数据, 构建多层感知机 (MLP) 初始抓取策略模型, 并通过行为克隆完成仿生抓取训练; 然后, 将该初始模型嵌入 Genesis 高保真空间操作仿真环境, 采用近端策略优化空间抓取算法开展抓取策略在线微调, 依托叠加式动作空间与分阶段奖励机制实现专家先验知识与环境自主探索的协同优化, 有效解决模仿学习的分布偏移缺陷与强化学习样本效率瓶颈. 实验结果表明, 所提方法在目标随机位姿扰动下抓取成功率达 89.5%, 较 MLP 模仿学习提升 14.5%, 显著增强了策略在目标位姿偏差下复杂空间场景中的鲁棒性与环境适应能力, 为微重力环境下空间机械臂漂浮目标自主抓取提供新的技术方案.

关键词 空间机械臂; 漂浮目标自主抓取; 仿生智能; 强化学习

引用格式 李连鹏, 郭航, 李明洋, 张海博, 徐拴锋, 张冬浩. 基于 PIRL 的空间机械臂仿生智能抓取方法. 自动化学报, 2026, **52**(5): 1101−1115

DOI 10.16383/j.aas.c250686 **CSTR** 32138.14.j.aas.c250686

Bionic Intelligent Grasping Method of Space Manipulators Based on Progressive Imitation-reinforcement Learning

LI Lian-Peng[1] GUO Hang[1] LI Ming-Yang[2, 3] ZHANG Hai-Bo[2, 3] XU Shuan-Feng[2, 3] ZHANG Dong-Hao[1]

Abstract To address the challenges faced by space manipulators in performing autonomous grasping tasks of floating targets in microgravity environments, specifically the difficulties in sample acquisition, weak generalization capability, and poor adaptation to dynamic disturbances, a bionic intelligence-integrated progressive imitation-reinforcement learning method is proposed. First, based on expert demonstration data of human arm-hand collaborative operations collected through teleoperation, a multi-layer perceptron (MLP) initial grasping strategy model is constructed, and bionic grasp training is conducted through behavior cloning; Next, the initial model is embedded into the high-fidelity Genesis space operation simulation environment, and the proximal policy optimization for grasping in space algorithm is employed for online fine-tuning of the grasping strategy. By leveraging a stacked action space and a staged reward mechanism, the method achieves collaborative optimization between expert prior knowledge and autonomous environmental exploration, effectively addressing the distribution shift defect in imitation learning and the sample efficiency bottleneck in reinforcement learning. Experimental results indicate that the proposed method achieves a grasp success rate of 89.5% under random target pose disturbances, an improvement of 14.5% compared to MLP-based imitation learning, significantly enhancing the robustness and environmental adaptability of the strategy in complex spatial scenes with target pose deviations. This provides a new technical solution for autonomous grasping of floating targets by space manipulators in microgravity environments.

Keywords space manipulator; autonomous grasping of floating targets; bionic intelligence; reinforcement learning

Citation Li Lian-Peng, Guo Hang, Li Ming-Yang, Zhang Hai-Bo, Xu Shuan-Feng, Zhang Dong-Hao. Bionic intelligent grasping method of space manipulators based on progressive imitation-reinforcement learning. *Acta Automatica Sinica*, 2026, **52**(5): 1101−1115

收稿日期 2025-11-30 录用日期 2026-03-31
Manuscript received November 30, 2025; accepted March 31, 2026
国家自然科学基金 (62406032), 北京市自然科学基金 (4242036), 空间智能控制技术全国重点实验室基金 (HTKJ2025KL502016, 2025-JCJQ-LB-065) 资助
Supported by National Natural Science Foundation of China (62406032), Beijing Natural Science Foundation (4242036), and Fund of National Key Laboratory of Space Intelligent Control (HTKJ2025KL502016, 2025-JCJQ-LB-065)
本文责任编委 刘志杰
Recommended by Associate Editor LIU Zhi-Jie

随着民用航天、深空探测和空间作业等领域的快速发展, 对在轨服务中空间目标进行高效、精确的自主操作需求日益迫切[1]. 空间环境具有微重力、

1. 北京信息科技大学自动化学院 北京 100192 2. 北京控制工程研究所 北京 100094 3. 空间智能控制技术全国重点实验室 北京 100094
1. College of Automation, Beijing Information Science and Technology University, Beijing 100192 2. Beijing Institute of Control Engineering, Beijing 100094 3. National Key Laboratory of Space Intelligent Control, Beijing 100094

强扰动等特性, 使得目标动力学行为与地面环境差异显著, 为空间机械臂自主抓取带来巨大挑战[2]. 特别是在卫星修复、轨道垃圾清理等典型任务场景中, 需要对漂浮或缓慢旋转的目标进行动态抓取, 这对空间机械臂系统的定位精度、关节灵活性与环境适应性提出极高要求[3].

当前, 目标自主抓取作为空间操作任务的核心基础, 已成为航空领域的重要研究方向. 受生物感知与运动协调机制启发, 空间仿生智能方法[4] 试图从人类或其他生物的高效操作行为中提取鲁棒的策略先验, 以应对非结构化太空环境中的动态不确定性. 经典主流方法通常基于物体几何或动力学模型[5], 对抓取位姿进行显式建模与优化, 部分研究也开始尝试融入仿生智能的核心思想, 以提升控制策略的适配性. 例如, 原劲鹏等[6] 围绕双臂六自由度空间机器人目标协同搬运任务, 借鉴生物协同运动的协调机制, 采用机器人与目标间的虚切断铰接结构及主臂刚性运动、从臂柔性运动的作业模式, 建立闭链系统动力学模型并设计关节跟踪控制律与柔顺阻尼控制律. Jiang 等[7] 针对存在建模不确定性的双臂机器人双边作业任务, 借鉴仿生智能动态适配复杂环境的核心思想, 提出基于显式动力学模型的自适应控制方法, 通过系统分解与命令滤波实现轨迹跟踪与接触力协同控制, 并采用径向基函数 (radial basis function, RBF) 神经网络结合复合学习律补偿系统动态不确定性, 构建适配复杂场景的自适应补偿机制. 张孟旭等[8] 设计机器视觉的机械臂抓取系统, 结合颜色识别与逆运动学 (inverse kinematics, IK), 构建具有旋转不变性的视觉抓取系统, 模拟生物视觉感知与动作联动的特性, 提升了抓取系统的环境适配能力. 然而, 现有研究多集中于静态目标或地面模拟环境, 针对空间微重力条件下漂浮目标的动态抓取方法仍不成熟. 传统基于模型的控制策略高度依赖精确的动力学参数, 在面对未知质量分布、姿态不确定或接触力学复杂的非合作目标时, 存在适应性差、鲁棒性弱等问题, 难以满足空间目标抓取任务需求.

近年来, 数据驱动方法凭借强大的特征学习与泛化能力备受重视. 其中, 模仿学习 (imitation learning, IL) 与强化学习 (reinforcement learning, RL) 作为两类主流智能决策方法, 在机械臂自主抓取任务中展现出卓越潜力. IL 方法通过专家示范快速构建初始策略, 有效降低策略学习难度[9]. 例如, Odesanmi 等[10] 提出一种基于可穿戴设备的人类演示学习框架, 通过实时映射人体运动数据实现对机器人手臂的直观、低成本远程操控, 提升了人–机器人技能转移的交互性、适应性与易用性; Kota 等[11] 通过引入分割图像和触觉信息改进端到端模仿学习模型, 增强了远程操控机器人在不同背景和光照条件下的任务成功率, 并有效降低过强的抓取力; Zhang 等[12] 提出一种基于卷积神经网络 (convolutional neural network, CNN) 的模仿学习框架, 通过观测人类抓取操作快速习得抓取点选择策略, 可使双臂机器人在 20 min 内掌握新任务, 其抓取规划效率显著优于传统几何建模方法. 然而, 该方法本质为监督学习, 在小样本条件下泛化能力有限. 尤其在空间任务中, 高质量遥操作数据获取成本高、覆盖范围窄, 导致策略在面对未见目标位姿或动力学扰动时易发生分布外崩溃, 难以维持稳定抓取. RL 方法具备自主探索与策略优化能力, 在机器人抓取复杂动态任务中展现出强大潜力[13]. Liu 等[14] 提出基于数字孪生的深度强化学习抓取策略, 实现从虚拟环境到实际机器人物理空间的转移; Shukla 等[15] 提出一种基于进化计算和深度强化学习的机器人抓取操作方法, 通过分解位置和方向学习问题, 显著提升机器人对物体的智能抓取能力; Hu 等[16] 研究一种基于 RL 的两阶段抓取算法, 通过逆向 RL 建模活体物体的对抗行为并结合图卷积网络优化灵巧手控制, 在预抓取和手内把持阶段分别达到较高的成功率, 有效减小接触力, 实现对动态目标的安全抓取. 但强化学习普遍存在样本效率低、收敛速度慢、奖励信号稀疏等问题, 尤其在高维连续动作空间中难以快速实用化; 在缺乏先验引导时从零开始训练需要海量交互样本, 在真实空间系统中难以落地应用.

人类在执行抓取任务时, 可自然融合视觉引导、多阶段轨迹规划与接触柔顺性, 展现出对位姿偏差和动力学扰动的高度适应能力. 该能力源于长期演化的感知、动作耦合机制, 可为空间机械臂系统提供高质量的行为原型[17]. 为此, 本文采用人类遥操作获取专家示范数据, 将生物启发式操作特性融入策略学习过程, 并结合模仿学习的高样本效率与强化学习的环境探索优化能力, 成为提升空间机械臂抓取鲁棒性与环境适应能力的有效途径.

为兼顾样本效率与策略鲁棒性, 近年来, 结合 IL 与 RL 优势的混合学习框架成为研究热点. 通过模仿学习提供策略初值, 再以强化学习进行环境驱动优化. 例如, Li 等[18] 提出仅需单次示范的多指抓取学习方法, 结合模仿学习与强化学习并引入课程学习, 提升了模型泛化能力; 申珅[19] 构建一种融合编程示教、模仿学习与强化学习的机械臂抓取控制方法, 通过构建基于概率模型的算法优化示教数据, 并设计奖励与策略双优化的算法提升训练效率;

Pereira 等[20] 提出融合强化学习与模仿学习的残差手势校正模型, 将含噪声三维手势估计映射为合理的虚拟手姿态, 在无真实交互数据条件下生成高质量手–物交互样本. 然而, 上述研究主要面向地面静态或低动态目标, 未充分考虑空间微重力环境下漂浮目标的独特挑战, 如无重力约束下的自由运动、接触力难以建模以及任务对姿态–位置协同控制要求高等难题.

因此, 本文提出一种面向空间漂浮目标的智能抓取方法, 构建渐进式模仿强化学习 (progressive imitation-reinforcement learning, PIRL) 智能抓取策略. 利用模仿学习初始化的样本高效性优势搭建优质策略初值, 结合强化学习微调的环境探索能力实现策略动态优化, 二者协同融合提升从接近到稳定抓握全过程的仿生智能抓取能力. 本文主要贡献如下:

1) 搭建面向空间机械臂的 real-to-sim 遥操作系统, 实现人手姿态到灵巧手动作的实时映射, 高效构建高质量专家示范数据集;

2) 提出基于相对位姿建模与移动–抓取–后抓取移动的仿生序贯轨迹解耦的样本生成方法, 通过单次演示生成多样化训练序列, 显著提升小样本下的泛化能力;

3) 构建融合模仿学习与强化学习的渐进式智能抓取策略, 针对 Genesis 仿真平台中漂浮目标的自主抓取任务, 验证了所提方法在微重力动态场景下的有效性.

1　PIRL 方法

PIRL 是面向空间微重力环境漂浮目标自主抓取任务设计的混合学习方法, 融合仿生智能先验, 以模仿学习构建基础策略初值, 强化学习实现场景化微调优化, 在提升样本效率的同时增强策略对动态扰动和位姿不确定性的鲁棒性.

1.1　PIRL 方法框架

在空间微重力环境下, 抓取目标常处于自由漂浮或缓慢旋转状态, 其动力学行为高度非线性且难以精确建模[21]. 基于模型的控制方法与单一数据驱动方法均难以兼顾适应性与学习效率. 为此, 本文构建的渐进式混合学习框架, 融合遥操作专家经验与仿真环境自主探索能力, 形成高效、鲁棒的抓取策略生成机制. 该框架由数据采集层、样本生成层、模型训练层、任务执行层四大模块构成, 各模块协同完成从数据获取到策略验证的全流程学习. 整体算法流程如图 1 所示.

仿生智能抓取学习过程在 Genesis 仿真平台中开展, 该平台可精准模拟微重力环境, 同时支持刚体动力学求解、碰撞检测与实时控制接口, 为抓取策略训练提供高保真的空间操作环境支撑. 首先通过键盘控制器与单目摄像头搭建的 real-to-sim 遥操作系统, 采集机械臂末端位姿与灵巧手 20 个关节角度数据, 构建含 560 个状态–动作对的原始专家数据集, 并基于接近、抓取、后抓取移动的轨迹解耦方法生成多样化训练样本; 再依托 Genesis 微重力仿真平台, 采用多层感知机 (multi-layer perceptron, MLP) 结合行为克隆 (behavior cloning, BC) 算法完成模仿学习初始化, 得到基础抓取策略; 通过近端策略优化空间抓取 (proximal policy optimization for grasping in space, PPO-GS) 算法开展在线微调, 最终融合得到鲁棒的漂浮目标抓取策略; 最后由机械臂–灵巧手协同系统在仿真环境中执行抓取任务, 操控者通过仿真实时画面监督全过程, 实现微重力场景下对漂浮目标的精准自主抓取.

与现有模仿学习与强化学习的混合方法相比, 本研究提出的 PIRL 方法是面向空间微重力漂浮目标抓取设计的定制化协同架构, 其核心创新均围绕图 1 各层级设计实现:

1) 在样本生成层, 基于 SE(3) 齐次变换矩阵开展相对位姿建模, 摆脱了绝对坐标系的限制, 可从单次人类遥操作演示中生成覆盖目标 ±0.1 m 位置、±5° 姿态扰动的多样化随机位姿样本, 有效解决空间任务中高质量演示样本获取难、覆盖范围窄的问题, 突破了现有方法直接复用原始演示数据导致泛化能力受限的局限.

2) 在模型训练层, 设计叠加式动作空间融合机制, 由 MLP 网络基于人类专家经验输出基础抓取位姿, PPO-GS 算法仅针对分布外偏差输出增量修正量, 而非直接替换原有策略, 既保留了专家先验知识带来的策略稳定性, 又有效规避了传统替换式融合易引发的策略漂移问题; 同时构建适配空间微重力场景的分阶段奖励函数, 采用稠密奖励与稀疏奖励相结合的设计思路, 引入 PPO 增量正则化项抑制动作突变, 引导策略严格遵循接近、抓取、稳定持有的空间抓取任务逻辑学习, 契合微重力环境下漂浮目标的动力学运动特性, 提升策略在动态扰动下的操作鲁棒性, 以此实现专家先验知识与环境自主探索的协同优化, 有效解决模仿学习的分布偏移缺陷与强化学习的样本效率瓶颈.

1.2　基于行为克隆的模仿学习

行为克隆作为一种模仿学习方法, 本质是一种监督学习, 常用于机器人精准作业与智能交互等场

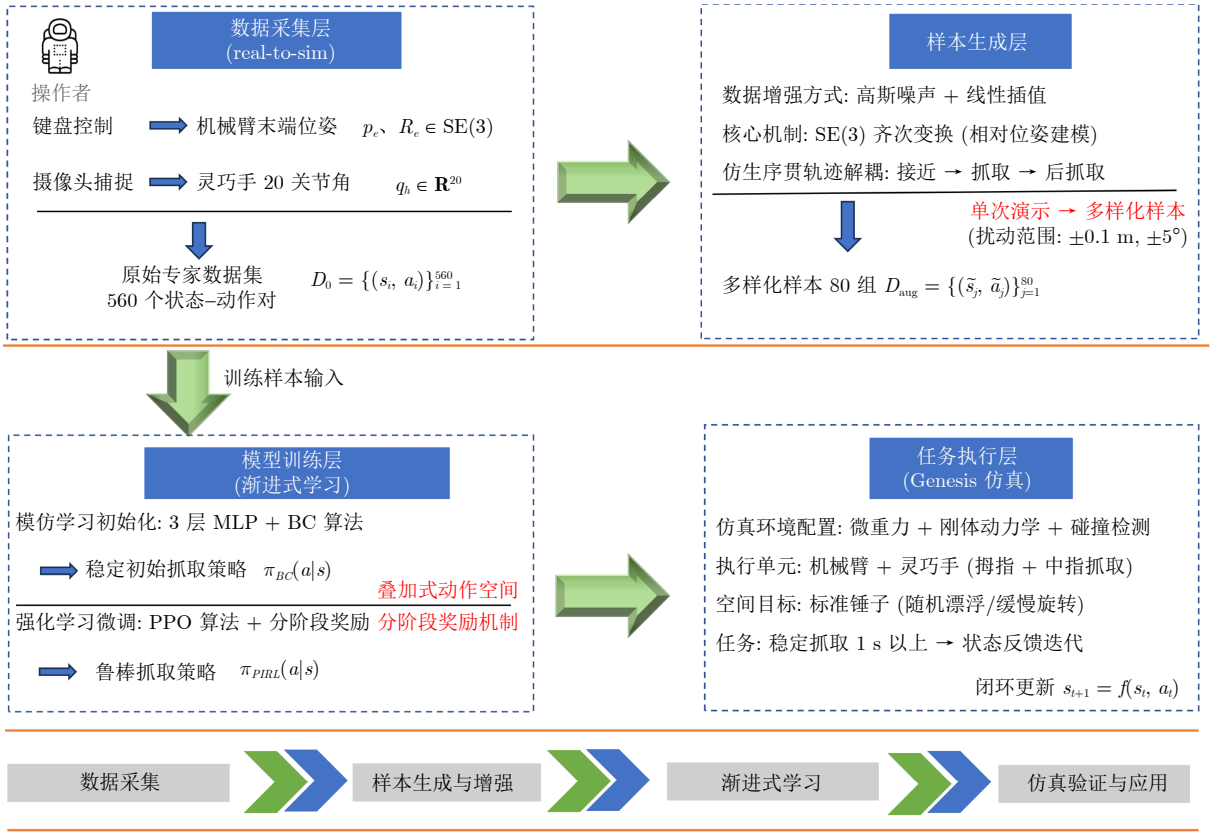


图 1 PIRL 方法框架图

Fig. 1 Framework diagram of the PIRL method

景, 如精密零件的抓取和装配任务以及语音肢体回应等[22].

为依托人类遥操作的专家经验构建稳定的抓取策略初值, 解决强化学习从零探索导致收敛慢的问题, 同时为后续强化学习微调提供先验知识支撑, 利用 BC 方法开展模仿学习建模与训练. 首先设计虚实映射的遥操作系统, 采集高质量专家示范数据; 其次提出相对位姿建模与轨迹解耦方法, 生成泛化性强的训练样本; 最终构建轻量化仿生模仿网络, 实现专家策略的精准复刻与高效学习.

1.2.1 遥操作系统设计

为构建高质量的专家示范数据集, 本文采用 real-to-sim 映射方法, 将操作者在真实世界中的操控行为实时迁移至仿真环境中执行.

遥操作系统的设计目标是高效获取人类操作者在抓取任务中的臂手协同运动经验. 系统采用混合输入模式, 分别基于机械臂映射器和灵巧手映射器控制仿真环境中的机械臂和灵巧手进行运动, 最终在仿真环境中完成工具抓取任务.

机械臂映射器[23] 是基于 PyGame 库开发的轻量级、低延迟交互界面, 支持操作者通过键盘控制仿真环境中机械臂末端的六自由度运动. 映射器核心涵盖末端空间平移、姿态旋转、操作目标位姿重置、抓取过程数据实时记录存储四个核心功能, 如图 2 所示.

操作端通过设置 w、s、a、d、q、e 组合按键实现机械臂末端沿三维坐标系的平移调节, z、x、c、v、r、f 组合按键完成三维姿态的旋转控制, Space 键实现目标工具的位姿, b 键触发数据存储指令. 所有运动调节均采用固定步长增量式执行, 确保机械臂操作精度与稳定性.

灵巧手映射器[24] 依托 Google MediaPipe Hands 视觉识别模型搭建, 通过单目摄像头实时感知操作者手部姿态, 提取 21 个人手关键点的三维坐标, 并将其映射为灵巧手的关节角度指令. MediaPipe 输出 21 个关键点, 定义如图 3 所示. 其中, f_0 为手腕根部, f_1 为大拇指, f_2 为食指, f_3 为中指, f_4 为无名指, f_5 为小拇指. 为统一手部姿态表示, 首先构建一个以手腕为原点的局部坐标系 O_0:

通过点 f_0、$f_{1\text{-}1}$、$f_{2\text{-}1}$、$f_{3\text{-}1}$、$f_{4\text{-}1}$、$f_{5\text{-}1}$ 进行最小二乘拟合平面 p_0; 取向量 $f_0f_{2\text{-}1} \times f_0f_{4\text{-}1}$ 为平面 p_0 的法线 z_0; 取向量 $f_0f_{3\text{-}1}$ 方向为 x_0 轴, 由右手定则 $y_0 = z_0 \times x_0$ 得到 y_0 轴.

图 2　机械臂映射器

Fig. 2　Robot arm mapper

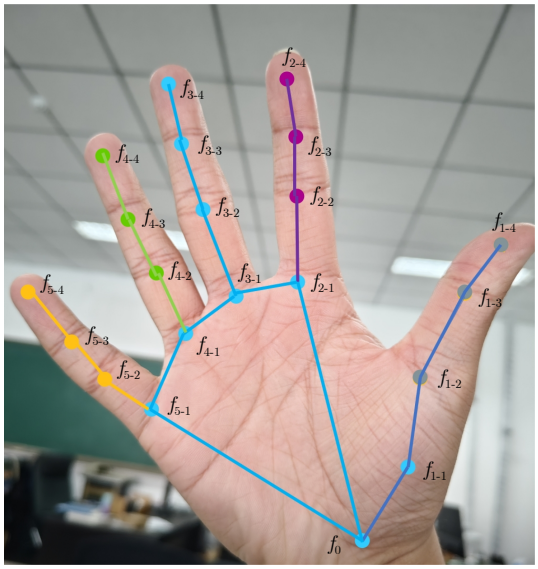


图 3　手部关键点定义

Fig. 3　Hand key point definition

由此构成右手坐标系 $\{O_0 : x_0,\ y_0,\ z_0\}$, 作为后续角度计算的参考系. 下面给出关节角度映射算法:

1) 大拇指关节定义

根部旋转角 $q_{1\text{-}1}$ 公式为:

$$q_{1\text{-}1} = \angle\left(\text{proj}_{x_0y_0}(f_{1\text{-}1}f_{1\text{-}2}),\ -y_0\right) \tag{1}$$

其中, proj 表示投影, 式 (1) 代表向量 $f_{1\text{-}1}f_{1\text{-}2}$ 向平面 x_0-y_0 的投影与 $-y_0$ 的夹角, 指向 z_0 方向为正, 反之为负.

根部侧摆角 $q_{1\text{-}2}$ 为:

$$q_{1\text{-}2} = \angle\left(f_{1\text{-}2}f_{1\text{-}3},\ \text{proj}_{p_1}(f_{1\text{-}2}f_{1\text{-}3})\right) \tag{2}$$

其中, p_1 为过点 $f_{1\text{-}1}$ 与 $f_{1\text{-}2}$ 且与平面 $\{f_{1\text{-}1},\ f_{1\text{-}2},\ f_{2\text{-}1}\}$ 相垂直的平面.

屈伸角 $q_{1\text{-}3}$ 与 $q_{1\text{-}4}$ 为:

$$q_{1\text{-}3} = \angle\left(\text{proj}_{p_1}(f_{1\text{-}2}f_{1\text{-}3}),\ f_{1\text{-}1}f_{1\text{-}2}\right) \tag{3}$$

$$q_{1\text{-}4} = \angle\left(f_{1\text{-}2}f_{1\text{-}3},\ f_{1\text{-}3}f_{1\text{-}4}\right) \tag{4}$$

其中, 构成夹角的两个向量的叉乘结果与向量 $f_{1\text{-}1}f_{1\text{-}2} \times f_{1\text{-}1}f_{2\text{-}1}$ 相反为正, 反之为负.

2) 食指关节定义

侧摆角 $q_{2\text{-}1}$ 指向 $+y_0$ 为正:

$$q_{2\text{-}1} = \angle\left(f_{2\text{-}1}f_{2\text{-}2},\ \text{proj}_{x_0z_0}(f_{2\text{-}1}f_{2\text{-}2})\right) \tag{5}$$

屈伸角 $q_{2\text{-}2}$、$q_{2\text{-}3}$、$q_{2\text{-}4}$:

$$q_{2\text{-}2} = \angle\left(\text{proj}_{x_0z_0}(f_{2\text{-}1}f_{2\text{-}2}),\ x_0\right) \tag{6}$$

$$q_{2\text{-}3} = \angle\left(f_{2\text{-}1}f_{2\text{-}2},\ f_{2\text{-}2}f_{2\text{-}3}\right) \tag{7}$$

$$q_{2\text{-}4} = \angle\left(f_{2\text{-}2}f_{2\text{-}3},\ f_{2\text{-}3}f_{2\text{-}4}\right) \tag{8}$$

其中, 构成夹角的两个向量的叉乘结果与坐标轴 $+z_0$ 相反为正, 反之为负.

3) 中指、无名指、小拇指关节定义

中指、无名指、小拇指的映射方式与食指相同, 分别计算其侧摆角 $q_{i\text{-}1}$ 与三级屈伸角 $q_{i\text{-}2}$、$q_{i\text{-}3}$、$q_{i\text{-}4}$ $(i = 3,\ 4,\ 5)$.

最终, 灵巧手共输出 20 个关节角度, 构成动作向量的一部分.

灵巧手映射过程存在多源误差, 直接影响动作复现精度. 需对误差源进行理论建模并基于统计方法估计总误差, 为系统优化提供依据. 映射链路如图 4 所示.

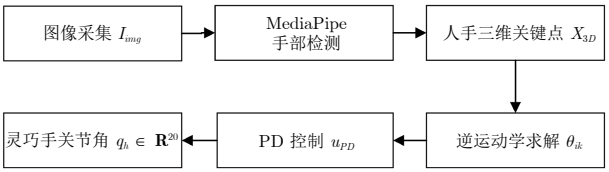


图 4　关节映射链路

Fig. 4　Joint mapping link

1) 视觉感知误差

视觉感知误差来源于图像采集 I_{img} 与手部关键点 X_{3D} 检测过程. 通过相似三角形原理, 可将像素级检测误差转换为三维空间误差, 计算公式为:

$$\sigma_{vision} = \frac{L}{W} \times p \tag{9}$$

其中, L 为相机有效视野; W 为图像横向分辨率; p 为 2D 关键点像素检测误差.

2) 逆运动学求解误差

采用伪逆最小二乘法进行逆运动学求解, 将手部三维关键点 X_{3D} 转换为灵巧手目标关节角度 θ_{ik}, 求解模型为:

$$J\dot{\theta}_{ik} = \dot{x} + \alpha\theta_{ik} \tag{10}$$

其中, J 为雅可比矩阵; α 为正则化系数; $\dot{x}$ 为手部三维关键点的目标位移向量; θ_{ik} 为逆运动学求解得到的灵巧手目标关节角度向量; $\dot{\theta}_{ik}$ 为关节角速度向量. 该方法产生的关节角度残差会等效为指尖空间定位误差.

3) 时间延迟误差

系统总时延由图像采集、模型推理、通信与 PD 控制 u_{PD} 延迟叠加而成, 对应空间误差为:

$$\sigma_{delay} = v_{hand} \times \Delta t \tag{11}$$

其中, v_{hand} 为手部运动平均速度; Δt 为系统累积总时延.

基于上述三类误差分量, 根据方差叠加原理, 总误差的标准差估计为:

$$\sigma_{total} = \sqrt{\sigma_{vision}^2 + \sigma_{IK}^2 + \sigma_{delay}^2} \tag{12}$$

其中, σ_{IK} 为逆运动学求解误差.

通过上述方法设计遥操作系统, 操作者通过观察 Genesis 仿真环境中的实时画面, 在真实世界中同步执行抓取动作. 其手臂运动通过键盘指令控制虚拟机械臂, 手部姿态通过摄像头捕捉并驱动虚拟灵巧手, 最终在仿真环境中完成工具抓取任务. 系统同步记录以下信息:

1) 机械臂末端执行器的位姿 p_e、R_e;

2) 灵巧手各关节的角度 q_h;

3) 目标工具 (锤子) 的位姿 T_{ob}.

1.2.2 样本生成方法

在空间机械臂遥操作任务中, 操作人员通过示教标定关键位姿时, 标准抓取目标处于特定位姿. 然而, 实际执行时目标物体的位置和姿态是随机变化的. 传统的绝对坐标记录方法无法应对这种变化, 如果目标偏离标定位置, 机械臂仍按原坐标移动将导致任务失败.

本文所提的样本生成方法基于相对位姿建模与轨迹解耦, 通过提取并保存空间机械臂末端执行器相对于目标物体的空间几何关系. 该相对几何关系具有坐标系内在不变性, 可泛化应用于任意初始条件.

1) 工具坐标系的标定标准

工具坐标系的定义直接影响相对变换的计算精度: 原点为抓取目标的几何中心, 即灵巧手三指 (食指、中指、无名指) 闭合时的协同接触中心点, 该位置通过仿真环境中工具 CAD 模型的几何参数直接提取; 坐标轴定义为 Z 轴沿抓取目标中心线指向头部, X 轴垂直于头部平面且指向手指张开方向, Y 轴由右手定则确定, 构成右手正交坐标系.

2) 相对位姿建模与轨迹插值

在三维刚体运动学中, 物体的位姿表示为 SE(3) 特殊欧氏群上的齐次变换矩阵:

$$T = \begin{bmatrix} R_{3\times3} & b_{3\times1} \\ \mathbf{0}_{1\times3} & 1 \end{bmatrix} \tag{13}$$

其中, 旋转部分 R 描述物体的姿态, 为正交矩阵; 平移部分 b 描述物体的位置; 齐次化行, 即底部的 [0, 0, 0, 1] 使得连续变换可通过矩阵乘法复合.

给定标准抓取目标位姿 T_A (人类标定时抓取目标的位置) 和目标抓取位姿 T_B (手部应到达的位置), 相对变换定义为:

$$T_{rel} = T_A^{-1} \cdot T_B \tag{14}$$

式 (14) 表示从工具坐标系的视角观察, 机械臂末端执行器的位置和姿态. 由于采用了坐标系变换, T_{rel} 不依赖于全局参考系的选择, 具有内在不变性.

齐次变换矩阵的逆运算遵循以下解析形式:

$$T^{-1} = \begin{bmatrix} R^{\mathrm{T}} & -R^{\mathrm{T}} \cdot b \\ \mathbf{0} & 1 \end{bmatrix} \tag{15}$$

当目标物体的位姿从标准 T_A 变化为随机 T_D 时, 新的目标抓取位姿通过以下公式计算:

$$T_F = T_D \cdot T_{rel} \tag{16}$$

该操作为左平移, 其物理意义是将标定时确定的相对几何关系应用到目标的新位姿. 这保证了无论目标物体位于何处, 机器人末端与物体之间的空间构型保持一致, 从而实现抓取动作的泛化.

同理, 预抓取位姿 (接近但未接触工具的安全点) 遵循相同计算逻辑:

$$T_G = T_A^{-1} \cdot T_C \tag{17}$$

$$T_H = T_D \cdot T_G \tag{18}$$

其中, T_C 是标定时的预抓取位姿; T_G 是相对位姿关系; T_H 是计算出的实际预抓取的位姿. 通过这种方式, 仅需标定三个关键位姿 (物体标准位姿、预抓取位姿、抓取位姿), 系统即可自动生成适配任意工具位置的完整轨迹.

预抓取位姿 T_H 和抓取点 T_F 之间, 需要生成满足刚体运动约束的平滑中间轨迹, 分别处理平移和旋转分量.

平移插值采用空间的标准线性插值, 公式为:

$$p(t) = (1-t)\cdot p_H + t p_F \tag{19}$$

其中, $p(t)$ 代表插值过程中任意时刻的平移向量; p_F 和 p_H 分别表示最终抓取位姿和预抓取位姿的平移向量; 插值参数 $t \in [0,\ 1]$.

旋转插值采用球面线性插值 (spherical linear interpolation, SLERP). 旋转用单位四元数 q 表示, 所有合法旋转构成四维单位超球面 S^3. SLERP 定义为:

$$\mathrm{SLERP}(q_H,\ q_F,\ t) = \frac{\sin((1-t)\theta)}{\sin\theta} q_H + \frac{\sin(t\theta)}{\sin\theta} q_F \tag{20}$$

其中, q_H 和 q_F 分别为预抓取位姿和最终抓取位姿对应的单位四元数; $\theta = \arccos(q_H \cdot q_F)$ 为两个四元数的夹角.

上述轨迹插值的核心目的是在预抓取位姿到最终抓取位姿的运动过程中, 生成平滑、连续、符合刚体运动约束的中间位姿序列, 为机械臂提供可执行的运动指令.

3) 仿生序贯轨迹解耦与样本生成

基于相对位姿和目标关节角构建一条完整的抓取任务轨迹, 分为接近、抓取和后抓取移动三个环节, 实现从单次演示生成多样化样本.

a) 接近环节: 随机采样满足物理约束的工具位姿, 计算对应目标抓取位姿, 对机械臂从当前位姿到目标抓取位姿的路径进行插值, 灵巧手保持张开, 要求逆运动学求解末端残差 $\Delta e < 2$ cm, 姿态残差 $\Delta\theta_{ee} < 1°$; 生成接近−定位样本, 增强模型对不同工具位姿的适应能力.

b) 抓取环节: 机械臂保持目标抓取位姿不动, 将灵巧手关节角从张开状态线性插值至闭合状态, 要求相邻时间步关节角位移 $\Delta q_h < 15°$; 无机器人自碰撞、环境碰撞及非意图的目标物体碰撞, 生成中间抓握样本, 学习稳定抓握的时机与力度.

c) 后抓取移动环节: 灵巧手保持闭合状态, 对机械臂从抓取位姿到目标放置位姿的路径进行插值, 要求抓取后两指 (拇指与中指) 到物体中线点距离和 $d < 8$ cm 且抬起阶段物体 Z 轴抖动 $\Delta z < 1$ cm, 生成负载状态下的控制样本, 提升机械臂稳定携带工具的能力.

通过在不同随机工具位姿下重复上述流程, 可从单次人类演示生成多条多样化轨迹, 结合高斯噪声增强与线性插值补充中间样本, 进一步拓宽数据分布范围. 通过多次重复采样, 循环加载单个样本数据、拼接特征, 再纵向合并所有样本, 最终得到一个包含所有数据的完整数据集.

1.2.3　仿生模仿网络

为复现操作人员在抓取任务中展现的高效感知与动作映射特性, 提出轻量级的 3 层全连接前馈神经网络 MLP 作为仿生模仿策略网络. 通过监督学习方式实现从环境状态到抓取动作的端到端映射, 网络结构如表 1 所示.

行为克隆的核心训练目标是最小化网络预测动作与专家示范动作的偏差. 针对空间机械臂臂手协同的连续位姿控制需求, 采用均方误差 MSE 作为损失函数, 实现对专家抓取策略的精准拟合, 损失函数定义为:

$$\mathrm{MSE} = \frac{1}{N}\sum_{i=1}^{N}\left(\beta_i - \hat{\beta}_i\right)^2 \tag{21}$$

其中, β_i 为专家示范的真实动作; $\hat{\beta}_i$ 为网络预测动作; N 为专家数据集的状态−动作对总数. 该损失函数为连续凸函数, 保证参数空间存在全局最优解, 可避免训练陷入局部最优; 同时梯度计算复杂度为线性阶, 适配轻量化网络的实时训练要求, 为后续强化学习阶段提供稳定的策略初值.

1.3　PPO-GS 算法

尽管行为克隆能够从专家示范中快速学习初始控制策略, 然而存在分布偏移问题. 在推理阶段, 策略可能进入训练数据中未覆盖的状态区域, 导致动作预测失效, 进而引发误差累积与任务失败.

基于训练得到的基础抓取策略, 以搭载 PIRL 策略的空间机械臂−灵巧手一体化系统为智能体, 面向空间微重力漂浮目标抓取场景, 提出面向空间

表 1　MLP 网络结构

Table 1　MLP network structure

网络层级	维度/数量	核心内容描述
输入层	15	机械臂当前末端位姿 (7 维) + 工具当前位姿 (7 维) + 灵巧手状态 (1 维)
隐藏层 1	128	特征提取层, 处理输入层 15 维数据, 初步压缩冗余信息
隐藏层 2	64	特征优化层, 进一步提炼关键操作特征
输出层	8	机械臂期望末端位姿 (7 维) + 灵巧手期望运动状态 (1 维)

抓取的 PPO-GS 算法. 该算法是用于空间抓取的近端策略优化算法, 在标准 PPO 框架基础上, 优化设计了叠加式动作空间、分阶段奖励机制与增量正则化约束, 通过智能体与仿真环境的持续交互优化策略, 使其能够在位姿扰动、动态漂浮等复杂场景下稳定执行抓取任务, 具体流程如图 5 所示.

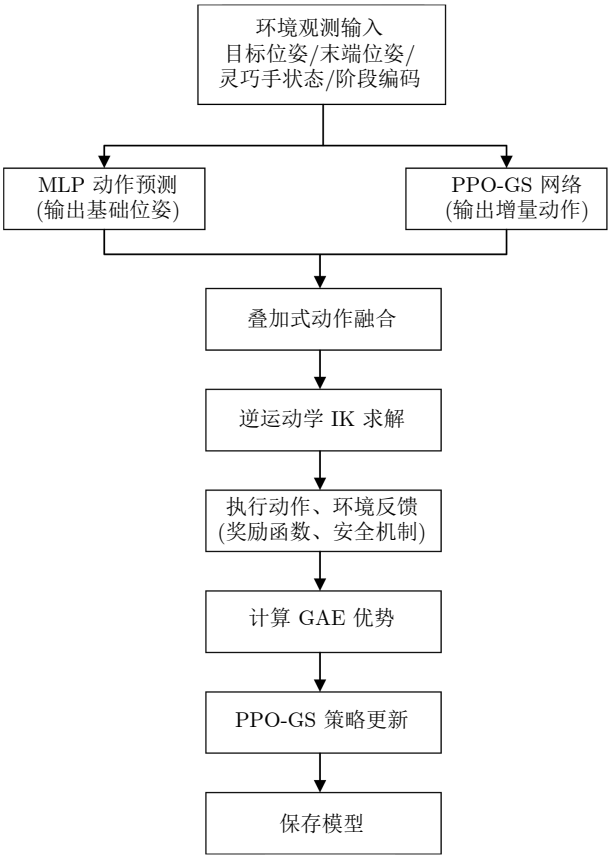


图 5 PPO-GS 算法执行与训练流程

Fig. 5 Flowchart of PPO-GS algorithm execution and training

PPO-GS 训练过程中, 观测输入同步送入 MLP 网络与 PPO-GS 网络, 输出的基础位姿与动作增量叠加并随阶段动态缩放, 生成最终动作; 经逆运动学求解得到关节控制指令后驱动系统执行, 动作执行后环境反馈新状态与多目标奖励, 经广义优势估计 (generalized advantage estimation, GAE) 与 PPO-GS 策略更新完成网络优化, 最终保存鲁棒模型.

空间机械臂–灵巧手一体化系统与环境持续交互, 感知工具状态并输出动作决策, 环境根据决策反馈新的状态与对应奖励, 系统以获取最大化累积奖励为核心优化目标.

1.3.1 近端策略优化 PPO

本文采用近端策略优化 PPO[25] 进行策略微调. PPO 是一种基于策略梯度的强化学习算法, 具有训练稳定、样本效率高、适用于连续动作空间等优点, 满足机械臂–灵巧手系统的高维连续控制任务需求. PPO 的核心思想是通过限制策略更新的步长来避免太大的性能下降. 这是通过引入一种特殊的目标函数实现的, 该目标函数包含一个剪辑 (clipping) 项来限制策略的改变程度, 其目标函数定义为:

$$\mathcal{L}^{\mathrm{CLIP}}(\theta) = \mathrm{E}_t\left[\min\left(r_t(\theta)\hat{A}_t,\ \mathrm{clip}\left(r_t(\theta),\ 1-\epsilon,\ 1+\epsilon\right)\hat{A}_t\right)\right] \tag{22}$$

其中,

$$r_t(\theta) = \frac{\pi_\theta(a_t|s_t)}{\pi_{\theta_{\mathrm{old}}}(a_t|s_t)} \tag{23}$$

表示更新后的策略 π_θ 与更新前的旧策略 $\pi_{\theta_{\mathrm{old}}}$ 在状态 s_t 下采取动作 a_t 的概率比; $\hat{A}_t$ 为优势函数, 由 GAE 方法计算得到; ϵ 为裁剪范围; clip 为剪切函数.

PPO 使用标准的 actor-critic 架构, 用于策略学习与价值估计, 网络结构如表 2 所示.

动作空间采用叠加式增量控制, 由基础位姿和相对增量叠加得到. 其中基础位姿由独立的 ActionPredictor 网络提供, 该网络以 15 维观测状态为输入, 输出 8 维控制指令, 运行时取前 7 维作为初始目标位姿. 为实现分阶段精细化控制, 在 PPO 网络的 15 维观测输入中, 第 15 维为阶段编码, 用于区分抓取过程中的接近与抓取阶段.

PPO 输出的 7 维动作含义为: 前 3 维是末端执行器相对位置增量 Δpos; 中间 3 维是欧拉角增量 ΔRPY, 分别对应翻滚角 (roll)、俯仰角 (pitch)、偏航角 (yaw); 最后 1 维是抓取强度命令, 映射至手指开合插值.

执行机制是将 ActionPredictor 输出的绝对位姿与 PPO 输出的相对增量融合, 通过逆运动学 (IK) 求解关节目标, 实现平滑、可控的末端轨迹跟踪.

1.3.2 奖励机制

为引导空间机械臂–灵巧手一体化系统学习高效、稳定的抓取策略, 建立了分阶段、稠密与稀疏结合的奖励机制. 该机制以任务完成为核心导向, 通过差异化正向激励与约束性负向惩罚, 引导策略逐步完成接近、抓取、稳定持有的完整操作流程. 奖励函数融合基础距离奖励、进度奖励、时间惩罚、PPO 增量正则化、阶段特异性奖励、失败惩罚和稀疏成功奖励, 实现对抓取过程的精细化引导. 总奖励函数 R_t 定义为:

$$R_t = R_{dist} + R_{progress} + R_{time} + R_{reg} + R_{phase} + R_{fail} + R_{success} \tag{24}$$

表 2　PPO 网络结构
Table 2　PPO network structure

网络层级	维度/数量	核心内容描述
输入层	15	观测向量: 抓取目标位姿 (3 维 + 4 维)、末端执行器位姿 (3 维 + 4 维)、阶段编码 (1 维)
actor 隐藏层 1	128	全连接层, ELU 激活
actor 隐藏层 2	64	全连接层, ELU 激活
actor 输出层	7	相对增量动作: Δpos 位置, ΔRPY 欧拉角, 抓取强度命令
critic 隐藏层 1	128	全连接层, ELU 激活
critic 隐藏层 2	64	全连接层, ELU 激活
critic 输出层	1	标量状态价值估计值

1) 基础距离奖励 R_{dist}

$$R_{dist} = k_1 \times \mathrm{e}^{-k_2 d} \tag{25}$$

其中, d 为两指的距离和; k_1、k_2 为可调正常数. 距离奖励作为核心驱动项, 通过指数衰减形式, 让智能体持续向工具靠近.

2) 进度奖励 $R_{progress}$

$$R_{progress} = \begin{cases} \zeta, & d_{t-1} - d_t > 0 \\ \kappa, & d_{t-1} - d_t \le 0 \end{cases} \tag{26}$$

其中, ζ 为正常数; κ 为负常数; d_t 和 d_{t-1} 分别表示当前时刻和上一时刻系统与目标的距离. 该奖励用于判别每一步的动作方向, 鼓励系统持续逼近目标, 抑制系统在局部区域震荡.

3) 时间惩罚 R_{time}

每步数值为 $-\phi$, $\phi > 0$ 为可调常数, 每步施加固定惩罚, 引导系统高效完成任务, 避免耗时.

4) PPO 增量正则化 R_{reg}

$$R_{reg} = -\varphi \times ||\Delta pos||_2 \tag{27}$$

其中, $\varphi > 0$ 为可调常数. 此项机制抑制策略更新的突变幅度过大, 保证轨迹平滑性.

5) 阶段特异性奖励 R_{phase}

a) 接近阶段触发奖励

$$R_{approach} = \begin{cases} \rho, & d < \tau_1 \\ 0, & \text{其他} \end{cases} \tag{28}$$

其中, ρ 为可调正常数; τ_1 为距离阈值. 该项引导系统主动进入抓取准备范围.

b) 抓取阶段子奖励

距离奖励为:

$$R_{shape} = \begin{cases} q, & d < \tau_2 \\ r, & \tau_2 \le d < \tau_3 \\ s, & \tau_3 \le d < \tau_4 \\ 0, & \text{其他} \end{cases} \tag{29}$$

其中, $q > r > s > 0$ 为可调参数; $\tau_2 < \tau_3 < \tau_4$ 为距离阈值. 该奖励以阶梯形式引导系统从粗对齐到精准抓取. 阈值间隔匹配灵巧手关节角分辨率, 数值梯度契合抓取精度需求.

闭合度奖励为:

$$R_{grip} = n \times G \tag{30}$$

其中, $n > 0$ 为常数; $G \in (0, 1)$ 为手指闭合程度. 该奖励线性鼓励手指持续闭合, 接近目标握姿.

稳定性奖励定义为:

$$R_{stable} = \begin{cases} w, & G > s_{th} \text{ 且持续 } T_{st} \text{ 步} \\ 0, & \text{其他} \end{cases} \tag{31}$$

其中, w 为正常数; s_{th} 为闭合度阈值; T_{st} 为持续步数阈值. 该奖励保证灵巧手稳定包裹工具, 避免瞬时闭合的误触发.

6) 失败惩罚 R_{fail}

$$R_{fail} = \begin{cases} -u, & \text{工具掉落/飞出/超时} \\ 0, & \text{其他} \end{cases} \tag{32}$$

其中, $u > 0$ 为可调常数. 该惩罚约束错误行为, 避免策略陷入循环失败.

7) 稀疏成功奖励 $R_{success}$

$$R_{success} = \begin{cases} k, & \text{抓取成功} \\ 0, & \text{其他} \end{cases} \tag{33}$$

其中, k 为正常数, 是一个高额终极奖励值. 该奖励明确任务终极目标, 驱动智能体完成有效抓取与稳定持有.

所有奖励项的设计以任务完成为核心导向, 同时兼顾操作流程的流畅性与效率.

2　实验

为验证所提的 PIRL 方法在空间机械臂抓取任务中的有效性, 本文基于 Genesis 仿真平台搭建了微重力环境下灵巧手抓取漂浮目标的实验环境, 分

别进行遥操作实验与抓取实验.

2.1 仿真系统搭建

2.1.1 Genesis 平台

Genesis[26] 是一个物理平台, 专为通用机器人、具身 AI、物理 AI 应用而设计, 该平台具备高精度刚体动力学求解、接触力计算与多自由度机器人建模能力, 支持通过 API 动态设置物理参数, 适用于空间机械臂在微重力环境下的操作任务仿真. 其中微重力模拟是通过将系统重力设为 0 m/s^2, 模拟空间中漂浮目标的操作场景;

物理参数设置可通过 Genesis 仿真接口动态调整物体质量、惯性张量、摩擦系数、阻尼等参数;

控制器支持底层关节位置/速度/力矩控制, 本文采用关节位置控制模式.

本文实验部分采用的软硬件配置如表 3 所示.

表 3 实验平台软硬件配置

Table 3 Experimental platform software and hardware configuration

配置项	具体规格
CPU	Core i5-10200H
GPU	NVIDIA GTX 1650 Ti
显存	4 GB
内存	16 GB
操作系统	Ubuntu 22.04
Python	3.10.12
PyTorch	2.4.0
CUDA	12.4
仿真引擎	Genesis 0.3.4

2.1.2 空间机械臂系统

机械臂系统采用机械臂 + 灵巧手的组合构型. 其中, 机械臂采用宇树科技 Z1 机械臂, 采用类人手臂构型, 具有 6 自由度, 关节为全旋转设计, 具备高扭矩密度与良好动态性能, 可实现末端执行器在工作空间内的任意位姿, 其作为灵巧手的载体. 灵巧手采用 5 指仿人手, 共 20 自由度, 具体构型如下:

大拇指, 每指 4 自由度, 包括旋转、侧摆和屈伸;

其余四指, 每指 4 自由度, 包括屈伸和侧摆.

所有自由度在仿真中均开启, 支持全手协同操作.

2.1.3 操作目标

抓取目标物体采用标准锤子工具, 质量 0.8 kg, 质心偏移明显, 模拟空间中非合作目标.

初始状态包括随机位姿、零初速度, 漂浮于工作空间内; 任务目标包括灵巧手从任意方向接近并稳定抓取锤柄, 成功抓取后保持工具稳定 1 s 即视为完成.

可通过 Genesis 仿真接口进行物体位姿等物理特性以及底层控制器参数的设置. 将每个物体的 URDF 文件导入即可构建仿真系统, 完整仿真系统如图 6 所示.

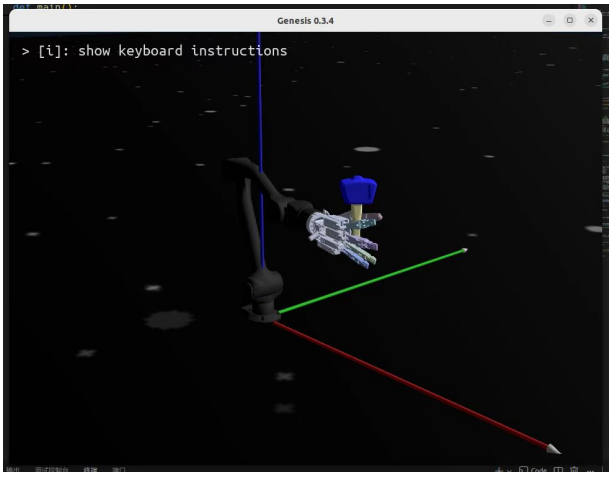


图 6 仿真系统搭建图

Fig. 6 Simulation system construction diagram

2.2 仿真实验

仿真实验包括遥操作实验和抓取实验, 核心思想是 MLP 基于人类经验提供粗定位, PPO-GS 基于环境反馈优化提供微调.

部分仿真物理参数如表 4 所示.

表 4 仿真物理参数

Table 4 Simulated physical parameters

参数	配置/说明
仿真时间步长 dt	50 ms
重力加速度 g	(0, 0, 0) m/s^2
仿真子迭代数 n	2
工具质量 m	0.8 kg
约束求解器 S	Newton 法
碰撞与关节限位 F	启用
工作空间半径 D	0.5 m
Z 轴高度限制 Z_m	[0.1, 0.6] m

2.2.1 遥操作实验

遥操作实验的主要目的是验证所提出的灵巧手映射方法[27]. 在人手自由运动任务中进行实验, 验证手是否可以从现实世界映射到仿真环境中, 如图 7 所示.

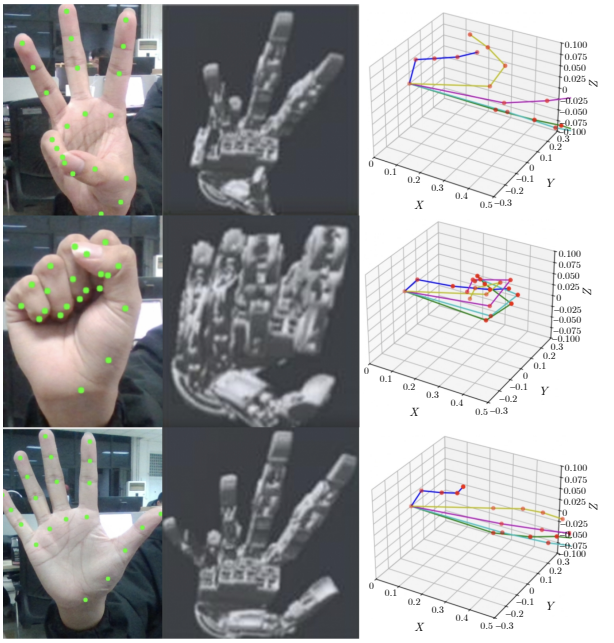


图 7　灵巧手映射

Fig. 7　Dexterous hand mapping

由图 7 的实验结果可见, 本研究测试了三指张开、手掌攥拳以及五指张开等多种典型手势, 不同类型的人手动作均可成功映射到仿真环境的灵巧手中, 图中右侧的部分展示了各个关键点的实时预测结果. 从该实验中可以定性看到, 遥操作灵巧手映射器具有复现人手的能力, 可以实现从现实空间动作到仿真空间动作的映射.

通过遥操作示教进行数据采集, 采用轨迹数据增强方法拓宽初始策略的覆盖范围. 具体而言, 在原始示教轨迹的状态参数 (末端位姿、关节角) 上添加小幅高斯噪声, 同时通过线性状态插值补充中间样本, 从数据层面丰富状态–动作对的分布特征, 初步缓解人类操作舒适域带来的分布窄化问题. 高斯噪声的添加采用如下公式:

$$x_{\text{aug}} = x_{\text{raw}} + \mathrm{N}(0,\ \sigma^2 I) \tag{34}$$

其中, x_{aug}、x_{raw}、$\mathrm{N}(0,\ \sigma^2 I)$ 分别表示添加高斯噪声后的数据增强样本、原始遥操作示教轨迹的状态向量和零均值高斯噪声; σ 为噪声标准差, 取值为 0.005 m; I 为单位矩阵.

生成的原始专家数据集 D_0 包括 560 个有效的状态–动作对, 数据有效率 100%, 平均轨迹长度 7 步且无碰撞事件. 不同 episode 几何路径的平均 Fréchet 距离达 0.12 m. 该指标用于衡量两条连续轨迹间的相似程度. 此数据作为原始专家数据集, 为后续抓取实验中的多样化样本生成提供基础.

为进一步验证遥操作映射精度, 通过 100 次重复性实验对灵巧手映射误差进行量化测试, 实验系统参数如表 5 所示.

由式 (9) 计算得视觉感知误差约等于 2.30 mm; 由关节角残差换算得逆运动学等效指尖误差约等于 3.00 mm; 由式 (11) 得时延对应误差约等于 5.00 mm; 由式 (12) 计算得总误差理论值约等于 6.27 mm. 经 100 次重复性实验实测误差约等于 6.10 mm, 与理论值偏差小于 3%, 验证了误差模型的合理性.

2.2.2　抓取实验

抓取实验包含遥操作映射及初始示教、样本增强生成、模仿学习网络训练以及推理测试四个阶段. 通过机械臂映射器和灵巧手映射器实时映射真实环境的动作, 在仿真环境中进行工具的抓取操作, 遥操作系统平台如图 8 和图 9 所示.

平台是遥操作实验的数据采集核心设备, 负责机械臂的键盘控制与灵巧手的姿态映射, 采集的原始专家数据集用于样本生成. 图 8、图 9 分别展示了抓取任务执行前与执行后的系统状态. 其中, 左上角为基于 Genesis 的仿真环境, 其功能是接收遥操作指令并执行相应的仿真过程; 右上角为机械臂映射单元, 该单元接收键盘输入指令, 进而控制机械臂完成平移、旋转等动作, 并支持工具重置与数据记录功能; 左下角显示的是人手实时操作的视觉信息流, 该信息流由 RGB 相机实时采集, 并经过图像算法处理; 右下角呈现了根据人手图像计算得出的各关节点三维坐标信息.

仿真结果表明, 构建的遥操作系统平台能够有效地完成指定的工具抓取任务, 验证了其可行性与有效性.

基于原始专家数据集, 采用所提出的样本生成方法在 Genesis 仿真平台中自动生成多组操作轨迹. 工具的空间随机扰动范围设定为 X、Y、Z 三轴方向位置扰动 ± 0.1 m, 姿态角扰动 $\pm 5°$, 实现目标

表 5　灵巧手映射误差参数

Table 5　Dexterous hand mapping error parameter

误差项目	数值
相机有效视野 L	0.3 m
图像横向分辨率 W	640
关键点检测误差 p	5 pixel
逆运动学关节角残差 e_q	0.05 rad
图像采集时延 t_1	约 33 ms
模型推理时延 t_2	约 15 ms
通信传输时延 t_3	约 5 ms
系统总时延 Δt	约 50 ms
手部运动平均速度 v_{hand}	0.1 m/s

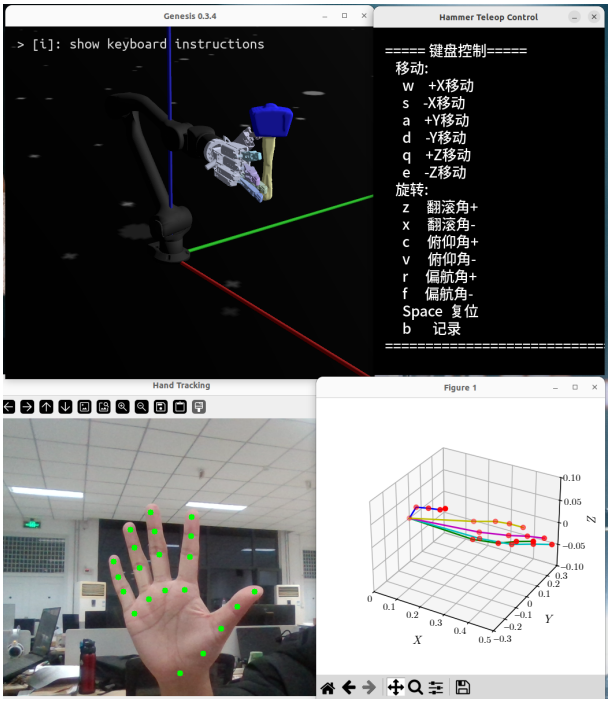


图 8 遥操作系统平台 (a)

Fig. 8 Teleoperation system platform (a)

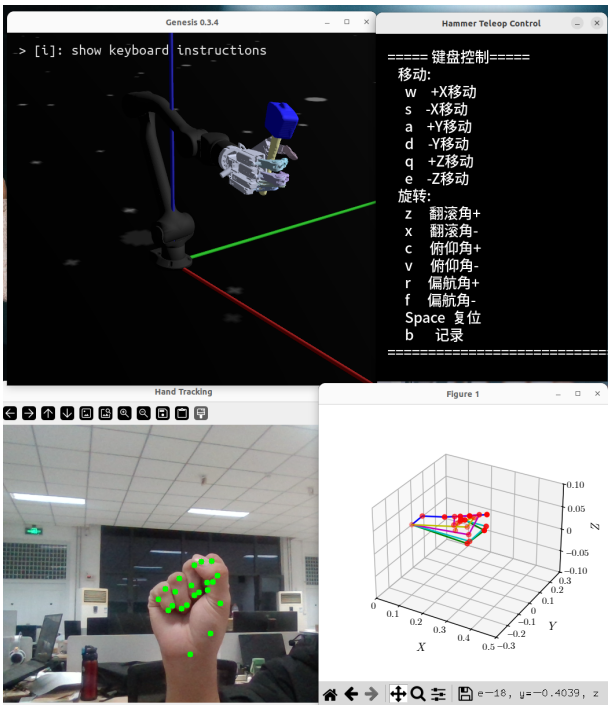


图 9 遥操作系统平台 (b)

Fig. 9 Teleoperation system platform (b)

位姿的全维度随机扰动. 累计采集并记录 80 组有效操作样本, 构成增强数据集 D_{aug}. 部分典型操作样本的示意图如图 10 所示.

基于上述采集的有效操作样本, 开展模型训练实验. MLP 动作预测网络使用 MSE 损失函数进行监督学习, 其具体训练过程如图 11 所示.

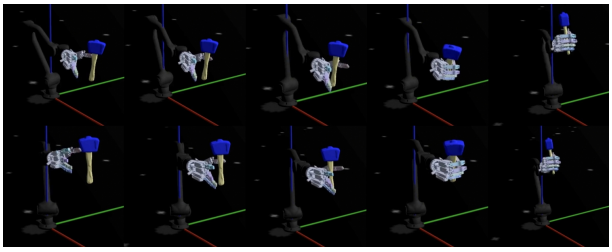


图 10 仿真系统典型操作样本生成效果

Fig. 10 Simulation system typical operation sample generation effect

实验结果表明, 当训练迭代至 400 个 epoch 时, 模型训练损失 (loss) 逐渐趋于收敛, 验证了该 3 层 MLP 架构在本任务中的有效性, 最终实现了初始训练效果.

基于训练完成的模仿学习初始策略 π_{BC}, 本研究开展推理测试实验以验证所提方法的泛化性能. 为模拟更贴近实际应用的场景, 测试设定工具初始运动状态: 三轴初始线速度均为 0.002 m/s, 三轴初始角速度均为 0.02 rad/s. 灵巧手完成闭合后, 锤子工具在机械臂末端保持稳定且未发生滑落或碰撞失效为成功. 具体推理测试结果如图 12 所示, 实验共开展 200 次独立测试, 最终抓取成功率达 75.0%. 尽管测试场景与训练样本的抓取环境存在轻微差

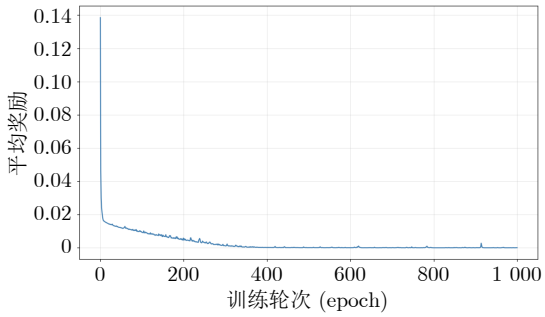


图 11 模仿学习模型训练过程

Fig. 11 Imitation learning model training process

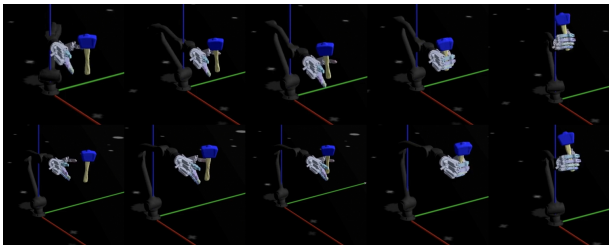


图 12 工具抓取推理测试结果

Fig. 12 Tool grasping reasoning test results

异, 模型仍能稳定完成工具抓取任务. 结果表明, 所采用的模仿学习方法能够有效实现目标抓取任务, 并具备一定的泛化能力.

实验过程中部分抓取失败案例如图 13 所示. 这类失败情况主要发生于机械臂工作空间的边界区域, 该区域的逆运动学求解结果中, 关节角变化幅度显著增大, 导致底层控制器无法实现稳定跟踪, 或因手指闭合时机过早导致锤子偏移, 进而造成误差失控. 下面进行强化学习 PPO 调优, 对通过模仿学习训练获得的模型进行二次优化训练, 以进一步提升模型在边界工况下的泛化性与鲁棒性.

强化学习将模仿学习训练得到的模型作为初始化权重加入 PPO-GS 网络中, 迭代次数为 1 000 轮, 每轮步数设为 30, 学习率 η 为 0.000 1, 折扣因子 γ 取 0.95, GAE 系数 λ 为 0.9, PPO 裁剪系数 ε 为 0.1, 熵正则系数 β 为 0.01, KL 散度目标 δ 为 0.02, 梯度裁剪阈值 c 为 0.5.

训练经历 1 000 轮迭代后, 平均奖励缓慢上升并收敛, 训练过程奖励函数如图 14 所示.

由图 14 可知, 模型的平均奖励曲线在训练 600 轮后逐渐趋于稳定并收敛. 相关时间性能如表 6 所示.

表 6 指标表明, PIRL 策略 π_{PIRL} 在线推理延迟小于 5 ms、系统控制频率达 20 Hz, 完全满足空间机械臂自主抓取任务对实时性的严格约束; 同时, 单步仿真、单动作执行的耗时设计兼顾了仿真精度与策略执行的动态响应能力, 适配微重力环境下漂浮目标抓取的实时控制需求. 此外, 推理阶段仅包含 MLP 前向与 PPO 增量修正的轻量化计算逻辑, 可有效适配空间系统的硬件计算资源约束.

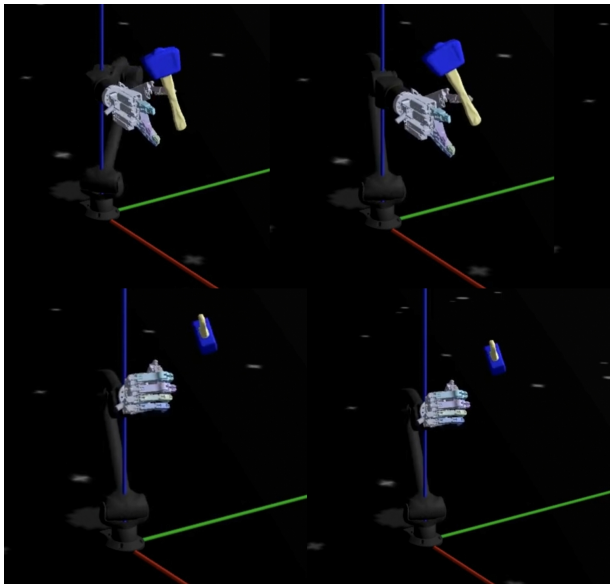


图 13　工具抓取测试失败情况

Fig. 13　Tool grasping test failure cases

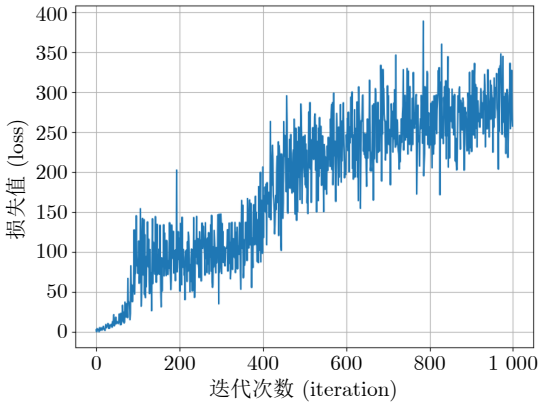


图 14　训练平均奖励随轮次迭代图

Fig. 14　Training average reward versus iteration epochs

表 6　时间性能

Table 6　Time performance

指标	数值	计算方法
仿真时间步长 dt	0.05 s	$dt = 0.05$ s
单动作执行时长 T_{act}	1.25 s	25 次物理步 × 0.05 s/次
episode 平均时长 T_{epi}	8.75 ~ 22.5 s	7 ~ 18 步 × 1.25 s/步
单轮训练时长 T_{train}	约 37.50 s	30 步 × 1.25 s
总训练时长 T	约 10.4 h	1 000 × 37.50 s
在线推理延迟 Δt	小于 5 ms	MLP 前向 + PPO 前向
控制频率 f	20 Hz	1/0.05 s

基于强化学习训练完成的模仿学习模型, 采用模仿学习验证相同的初始运动状态和成功判定规则进行 200 次独立测试, 最终成功 179 次, 优化成功率为 89.5%.

2.2.3　仿真结果与分析

模仿学习凭借遥操作专家数据快速构建初始策略, 实现 75.0% 的初始成功率, 但受限于训练数据的分布覆盖范围, 在工具位姿偏离演示分布时易出现分布外崩溃, 如机械臂工作空间边界区域的抓取失败率偏高.

PIRL 方法通过渐进式学习, 既利用 BC 算法的样本高效性构建接近最优的初始策略, 又通过 PPO-GS 算法的数万次在线交互优化分布外状态的决策能力, 最终优化成功率提升至 89.5%, 较初始成功率提升了 14.5%. 尤其在机械臂工作空间边界区域, PPO-GS 微调后的策略通过动态调整关节角跟踪误差补偿机制, 使失败率大幅降低, 验证了所提方法的有效性.

该方法有效性的核心逻辑在于初始策略已经具备较好的性能, 避免了强化学习从随机策略开始探索时效率低、收敛慢的问题, 强化学习具有主动探

索机制, 通过在线交互和奖励信号, 能发现比专家更好的策略.

PIRL 方法的 10% 失败率主要集中于两种场景: 1) 极端位姿扰动下, 机械臂工作空间易触及关节运动极限, 导致逆运动学求解出现奇异点, 末端位姿跟踪误差超过 0.03 m, 灵巧手无法精准对齐锤柄抓取区域, 最终引发抓取失效; 2) 目标工具初始位姿靠近机械臂基座时, 因奖励函数中未设置碰撞规避的强约束项, 策略优先追求抓取成功率而非避障安全, 易出现工具与基座碰撞的情况, 导致任务中断.

针对上述问题, 后续可从两方面优化: 一是在 PPO 网络输入特征中增加机械臂关节角限位约束与奇异点预判特征, 通过设计关节角安全裕度指标, 引导策略避开奇异点区域; 二是引入碰撞惩罚项优化奖励函数, 通过负向激励强化避障意识, 进一步提升极端场景下的鲁棒性.

3 结束语

本文针对空间微重力环境下漂浮目标自主抓取任务中样本获取难、泛化能力弱、动态适应差的核心问题, 提出 PIRL 方法, 通过理论构建、系统设计与仿真验证, 形成一套完整的空间机械臂仿生智能抓取方法, 实现了微重力环境下工具抓取任务的鲁棒执行. 具体工作包括: 1) 设计了融合键盘控制与单目视觉的遥操作系统, 实现真实人手姿态到仿真灵巧手动作的实时映射; 2) 基于相对位姿建模与三阶段轨迹解耦方法, 从单次遥操作演示生成 80 组多样化专家数据集, 构建含 560 个有效状态−动作对的训练样本库; 3) 构建了基于 3 层 MLP 网络的仿生模仿网络, 得到抓取成功率 75.0% 的基础策略; 4) 在 Genesis 高保真仿真环境中, 通过 PPO-GS 算法对策略进行强化学习微调, 在随机扰动下, 优化成功率提升至 89.5%. 最终实现了任务成功率的有效提升, 验证了方法在动态微重力场景下的有效性.

未来工作将聚焦以下三个方面: 一是拓展多结构、多约束、多目标的通用抓取策略, 结合空间非合作目标的几何与动力学特性构建工具特征库, 搭建统一的抓取决策理论框架, 从策略泛化的理论机制层面提升方法对不同类型空间目标的适配能力. 通过构建工具特征库与统一抓取框架, 提升方法的任务泛化能力. 二是融合计算机视觉与机器人感知理论, 采用基于深度学习的视觉位姿估计算法实现目标位姿的端到端感知与求解, 替代本文采用的真值获取方法. 三是结合实际灵巧手的机械驱动特性与运动学约束, 开展关节自由度的理论建模与优化分析, 建立适配实际机构的动作空间约束模型, 同时搭建地面微重力模拟实验平台, 开展物理实验验证与算法迭代, 实现从仿真理论验证到物理实验落地的方法迁移.

参考文献

1 Li Lin-Feng, Xie Yong-Chun. Space robotic manipulation: A multi-task learning perspective. *Chinese Space Science and Technology*, 2022, **42**(3): 10−24
(李林峰, 解永春. 空间机器人操作: 一种多任务学习视角. 中国空间科学技术, 2022, **42**(3): 10−24)

2 Chihi M, Hassine C B, Hu Q. Segmented hybrid impedance control for hyper-redundant space manipulators. *Applied Sciences-Basel*, 2025, **15**(3): Article No. 1133

3 Xie Fang-Lin, Wang Ling-Xin, Zhang Ya-Hang, Wang Yao-Bing, Wang Jie. Obstacle avoidance motion planning of manipulator for space autonomous assembly validation and evaluation. *Spacecraft Engineering*, 2025, **34**(2): 82−89
(谢芳霖, 汪凌昕, 张亚航, 王耀兵, 王捷. 面向空间自主装配验证评估的机械臂避障运动规划. 航天器工程, 2025, **34**(2): 82−89)

4 Si Y F, Wang D, Jiang Y Z, Zhu H, Shi S, Tan L, et al. Bionic intelligent clothing. *Advanced Materials*, 2025, **38**(5): Article No. e14621

5 Li M G, Zhang N, Xing Y, Liu B Y, Su W Y, Li S Y, et al. Design, analysis, and experimental research of flexible multi-constraint gripper for nest frames. *Journal of Mechanical Design*, 2026, **148**(2): Article No. 023301

6 Yuan Jin-Peng, Ge Lian-Zheng, Li De-Lun. Cooperative compliance control strategy for dual arm space robot with closed chain system. *Aerospace Control and Application*, 2023, **49**(2): 42−50
(原劲鹏, 葛连正, 李德伦. 双臂空间机器人闭链系统的协同柔顺控制策略研究. 空间控制技术与应用, 2023, **49**(2): 42−50)

7 Jiang Y M, Wang Y N, Miao Z Q, Na J, Zhao Z J, Yang C G. Composite-learning-based adaptive neural control for dual-arm robots with relative motion. *IEEE Transactions on Neural Networks and Learning Systems*, 2022, **33**(3): 1010−1021

8 Zhang Meng-Xu, Gao Xiang-Chuan, Yin Li-Nan, Wang Jian-Hui. Design of a robotic arm grasping system based on machine vision. *Computer Applications and Software*, 2024, **41**(8): 22−27
(张孟旭, 高向川, 尹丽楠, 王建辉. 基于机器视觉的机械臂抓取系统设计. 计算机应用与软件, 2024, **41**(8): 22−27)

9 Huang Yan-Long, Xu De, Tan Min. On imitation learning of robot movement trajectories: A survey. *Acta Automatica Sinica*, 2022, **48**(2): 315−334
(黄艳龙, 徐德, 谭民. 机器人运动轨迹的模仿学习综述. 自动化学报, 2022, **48**(2): 315−334)

10 Odesanmi G A, Wang Q N, Mai J G. Skill learning framework for human-robot interaction and manipulation tasks. *Robotics and Computer-Integrated Manufacturing*, 2023, **79**: Article No. 102444

11 Kota I, Yasutake T, Satoki T, Masaki H. Autonomous teleoperated robotic arm based on imitation learning using instance segmentation and haptics information. *Journal of Advanced Computational Intelligence and Intelligent Informatics*, 2025, **29**(1): 79−94

12 Zhang S, Liu S Q, Li Y, Li X, Wang Z G. A visual imitation learning algorithm for the selection of robots′ grasping points. *Robotics and Autonomous Systems*, 2024, **172**: Article No. 104600

13 Wang Xue-Song, Wang Rong-Rong, Cheng Yu-Hu. Safe reinforcement learning: A survey. *Acta Automatica Sinica*, 2023, **49**(9): 1813−1835
(王雪松, 王荣荣, 程玉虎. 安全强化学习综述. 自动化学报, 2023, **49**(9): 1813−1835)

14 Liu Y K, Xu H, Liu D, Wang L H. A digital twin-based sim-to-real transfer for deep reinforcement learning-enabled industrial robot grasping. *Robotics and Computer-Integrated Manufacturing*, 2022, **78**: Article No. 102365

15 Shukla P, Kumar H, Nandi G C. Robotic grasp manipulation using evolutionary computing and deep reinforcement learning. *Intelligent Service Robotics*, 2021, **14**(1): 61−77

16 Hu Z, Zheng Y, Pan J. Grasping living objects with adversarial

behaviors using inverse reinforcement learning. *IEEE Transactions on Robotics*, 2023, **39**(2): 1151−1163

17 Yagna J, Mahmoud S, Paul W, Aaisha M. A comprehensive review of robotics advancements through imitation learning for self-learning systems. In: Proceedings of the 9th International Conference on Mechanical Engineering and Robotics Research. Barcelona, Spain: ICMERR, 2025. 1−4

18 Li Y H, He H Y, Chai J, Bai G R, Dong E B. Grasping unknown objects with only one demonstration. *IEEE Robotics and Automation Letters*, 2025, **10**(2): 987−994

19 Shen Shen. Research on Robotic Arm Grasping Control Based on the Combination of Reinforcement Learning and Imitation Learning [Master thesis], North University of China, China, 2023.
(申珅. 基于强化学习与模仿学习结合的机械臂抓取控制研究 [硕士学位论文], 中北大学, 中国, 2023.)

20 Pereira M, Dimou D, Moreno P. In-hand manipulation of unseen objects through 3D vision. In: Proceedings of the 5th Iberian Robotics Conference. Zaragoza, Spain: ROBOT, 2022. 163−174

21 Yuan Li, Jiang Tian-Tian, Wei Chun-Ling, Yang Meng-Fei. Advances and perspectives of space control technology. *Acta Automatica Sinica*, 2023, **49**(3): 476−493
(袁利, 姜甜甜, 魏春岭, 杨孟飞. 空间控制技术发展与展望. 自动化学报, 2023, **49**(3): 476−493)

22 Yang Y C, Li R J, Wang L F, Zheng S, Ma S Z, Zhang K Y, et al. Scalable dexterous robot learning with AR-based remote human-robot interactions. arXiv preprint arXiv: 2602.07341, 2026.

23 Lin Qi-Guang, Liu Yu, Li Jie, Liu Xiao-Feng. Motion tracking model of 6-DOF manipulator based on trajectory measurement and human-machine mapping. *Journal of Electronic Measurement and Instrumentation*, 2023, **37**(3): 102−110
(林麒光, 刘宇, 李杰, 刘小峰. 基于轨迹测量与人机映射的六自由度机械臂运动追踪模型. 电子测量与仪器学报, 2023, **37**(3): 102−110)

24 Zhang Ling-Jun, Tang Liang, Liu Lei. Target position-guided in-hand reorientation of five-fingered dexterous hands. *Robotics*, 2025, **47**(1): 10−21
(张玲俊, 汤亮, 刘磊. 目标位置引导的五指灵巧手手内重定向. 机器人, 2025, **47**(1): 10−21)

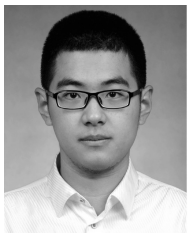
25 Schulman J, Wolski F, Dhariwal P, Radford A, Klimov O. Proximal policy optimization algorithms. arXiv: 1707.06347, 2017.

26 Genesis: A generative and universal physics engine for robotics and beyond [Online], available: https://genesis-world.readthedocs.io/zh-cn/latest/, April 20, 2026
(Genesis: 面向机器人及具身智能的生成式通用物理引擎 [Online], available: https://genesis-world.readthedocs.io/zh-cn/latest/, 2026-04-20)

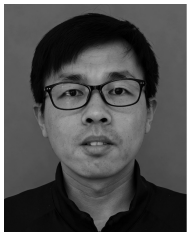
27 Li M Y, Du Z J, Ma X X, Dong W, Gao Y Z. A robot hand-eye calibration method of line laser sensor based on 3D reconstruction. *Robotics and Computer-Integrated Manufacturing*, 2021, **71**: Article No. 102136

李连鹏 北京信息科技大学自动化学院副教授. 主要研究方向为多智能体协同控制, 机器人智能控制.
E-mail: llp@bistu.edu.cn
(**LI Lian-Peng** Associate professor at the College of Automation, Beijing Information Science and Technology University. His research interests include multi-agent cooperative control and robot intelligent control.)

郭 航 北京信息科技大学自动化学院硕士研究生. 主要研究方向为机器人安全控制.
E-mail: 2024020508@bistu.edu.cn
(**GUO Hang** Master student at the College of Automation, Beijing Information Science and Technology University. His main research interest is robot security control.)



李明洋 北京控制工程研究所高级工程师. 主要研究方向为机器人智能操作控制. 本文通信作者.
E-mail: lmy_hit@163.com
(**LI Ming-Yang** Senior engineer at Beijing Institute of Control Engineering. His main research interest is intelligent operation control of robots. Corresponding author of this paper.)



张海博 北京控制工程研究所研究员. 主要研究方向为空间智能操作控制.
E-mail: zhanghb502@163.com
(**ZHANG Hai-Bo** Researcher at Beijing Institute of Control Engineering. His main research interest is space intelligent operation control.)

徐拴锋 北京控制工程研究所高级工程师. 主要研究方向为空间机器人智能控制.
E-mail: xushuanfeng2003@163.com
(**XU Shuan-Feng** Senior engineer at Beijing Institute of Control Engineering. His main research interest is intelligent control of space robots.)

张冬浩 北京信息科技大学自动化学院副教授. 主要研究方向为机器人智能控制, 具身智能.
E-mail: Donghaozhang@bistu.edu.cn
(**ZHANG Dong-Hao** Associate professor at the College of Automation, Beijing Information Science and Technology University. His research interests include robot intelligent control and embodied intelligence.)