冗余人工肌肉驱动的仿生机器人强化学习控制

牛鹏军[1] 程屹涛[1] 朱彦臣[2] 厉侃[2] 刘珂[1]

摘 要 人工肌肉是仿生机器人的核心驱动部件, 然而当前人工肌肉的应用与真实生物相差甚远, 缺乏像生物一样的冗余多肌肉协同. 针对上述问题, 围绕仿生机器人的复杂人工肌肉驱动与协同, 本文提出一种由多股人工肌肉并联驱动的软体机器人设计, 并围绕这种设计建立基于强化学习的运动控制策略. 研制了以柔性十字形电路板为主体, 集成六路液晶弹性体人工肌肉与驱动电路的原型样机, 并测试获得其应变特性与响应性能; 针对原型样机形变–运动特点, 在仿真环境中构建基于绳腱驱动的简化模型. 通过合理设计状态空间、动作空间及奖励函数等, 以 soft Actor-Critic 算法进行强化学习并行训练, 得到平移与旋转运动肌肉协同策略. 将运动策略中稳定周期段以离线方式驱动实物样机, 实现有效的多向平移与旋转运动, 验证了采用强化学习控制复杂人工肌肉系统的可行性.

关键词 仿生机器人; 人工肌肉; 强化学习; 软体机器人; 运动控制

引用格式 牛鹏军, 程屹涛, 朱彦臣, 厉侃, 刘珂. 冗余人工肌肉驱动的仿生机器人强化学习控制. 自动化学报, 2026, **52**(5): 953−965

DOI 10.16383/j.aas.c250508 **CSTR** 32138.14.j.aas.c250508

Reinforcement Learning Control for Bioinspired Robots Driven by Redundant Artificial Muscles

NIU Peng-Jun[1] CHENG Yi-Tao[1] ZHU Yan-Chen[2] LI Kan[2] LIU Ke[1]

Abstract Artificial muscles are key actuation components for bioinspired robots. However, their current applications remain far from the capabilities of biological muscle systems, particularly due to the lack of redundant and coordinated multi-muscle actuation similar to that found in living organisms. To address the challenge of complex artificial-muscle actuation and coordination in bioinspired robots, this study proposes a soft robotic design driven by multiple artificial muscles arranged in parallel and develops a reinforcement learning-based locomotion control strategy for this design. A prototype was developed using a flexible cross-shaped printed circuit board as the main body, integrating six liquid crystal elastomer artificial muscles and their driving circuits. Its strain characteristics and dynamic response performance were experimentally characterized. Considering the deformation and locomotion characteristics of the prototype, a simplified tendon-driven model was constructed in a simulation environment. By properly designing the state space, action space, and reward functions, parallel reinforcement learning training was conducted using the soft Actor-Critic algorithm to obtain coordinated muscle activation strategies for translational and rotational locomotion. The stable periodic segments of the learned locomotion policies were then extracted and used to drive the physical prototype offline. The robot achieved effective multidirectional translation and rotation, demonstrating the feasibility of using reinforcement learning to control complex artificial-muscle-driven systems.

Keywords bionic robot; artificial muscles; reinforcement learning; soft robot; locomotion control

Citation Niu Peng-Jun, Cheng Yi-Tao, Zhu Yan-Chen, Li Kan, Liu Ke. Reinforcement learning control for bioinspired robots driven by redundant artificial muscles. *Acta Automatica Sinica*, 2026, **52**(5): 953−965

自然界的运动体系, 是“形态–行为–智能”三位一体的优化成果[1−5]: 从尺蠖的波形蠕动[6], 到猛禽的扑翼–滑翔切换[7−8], 再到鱼类依靠鳍条与尾柄协同交替产生推力[9−10], 不同环境中的生物体, 往往能够在能耗、噪声与鲁棒性之间达成近乎最优的平衡. 受此启发, 研究者针对仿生机器人运动控制首先需要解决“如何驱动”这一底层核心问题. 目前, 工程领域已形成几条主流技术路线:

1) 流体驱动. 这种方式依托气动或液压腔体通过泵–阀网络构建压差, 在单位质量下可输出较大的力/位移, 性能比肩部分节肢动物的爆发式弹跳[11−12].

收稿日期 2025-09-29 录用日期 2026-01-04
Manuscript received September 29, 2025; accepted January 4, 2026
国家重点研发计划 (2022YFB4701900) 资助
Supported by National Key Research and Development Program of China (2022YFB4701900)
本文责任编委 刘志杰
Recommended by Associate Editor LIU Zhi-Jie
1. 北京大学先进制造与机器人学院 北京 100871 2. 华中科技大学智能制造装备与技术全国重点实验室 武汉 430074
1. School of Advanced Manufacturing and Robotics, Peking University, Beijing 100871 2. State Key Laboratory of Intelligent Manufacturing Equipment and Technology, Huazhong University of Science and Technology, Wuhan 430074

2) 磁场驱动. 在结构内部预埋硬磁或软磁颗粒, 借助亥姆霍兹线圈或稀土磁体即可实现远程致动, 能够完成旋转、翻滚乃至多达十二种运动模态的多场景运动切换[13−17].

3) 智能材料驱动. 直接利用材料内部的能量转换实现驱动, 例如形状记忆合金 (热−相变转换)[18−21]、介电弹性体 (电−静电转换)[22−24] 或离子聚合物 (化−电转换)[25] 等, 其结构可与仿生形态实现高度耦合.

若要构造一台能够自主运动并长期在非结构化环境中执行任务的仿生机器人, 驱动系统除高功率密度外, 更须满足轻量化、外场独立与可编程形变等综合指标[26−28]. 与依赖外部泵阀 (气动/液压) 或场源 (磁、光、电) 的方案相比, 智能材料驱动将能量转换单元嵌入结构内部, 可在无附属设备的情况下输出可控形变, 因此通常认为是实现仿生机器人自主运动控制的最具潜力驱动路径.

进一步聚焦到能够直接模拟生物体肌肉收缩−舒张机理的“人工肌肉”, 液晶弹性体 (liquid crystal elastomer, LCE) 凭借可编程取向、高比功率和大可逆应变等优势, 可在无泵、无阀、无线圈的极简架构下, 复现生物肌肉“静默输出、高功率密度”的核心特性, 成为当前人工肌肉候选材料中的优先选择[29−31].

然而, LCE 致动过程伴随显著的热−力耦合、迟滞与速率依赖; 多束人工肌肉的串并联进一步产生高维耦合与参数漂移, 使传统 PID、增益调度或模型预测控制难以在安全边界内保持精确、快速响应. 深度强化学习 (deep reinforcement learning, DRL) 近年来已在刚体臂[32−34]、腿式机器人[35−38] 上展示了对复杂动力学的自适应优势[39−41], 但在复杂人工肌肉系统中仍面临样本效率低和仿真−现实落差两大瓶颈: 1) 在线试错易导致过流或过温烧毁; 2) 材料本构与迟滞模型的不确定性削弱了策略迁移性能.

针对上述人工肌肉驱动系统控制难题, 本文以自研的复杂人工肌肉驱动软体机器人为研究载体, 提出一种基于软演员−评论家 (soft Actor-Critic, SAC) 的强化学习控制框架, 并通过“仿真−实物”离线策略执行实验验证其有效性, 具体研究工作与成果如下:

1) 提出并完成一种采用 LCE 人工肌肉驱动的仿生软体机器人的研发, 同步设计配套柔性驱动电路, 完成对该机器人控制性能与驱动性能的系统性测试;

2) 针对该系统的高度非线性特性, 在 MuJoCo 环境中建立简化桁架模型, 明确机器人的状态空间与动作空间, 结合软体驱动特性设计 Episode 流程, 构建与空间自由度相匹配的奖励函数, 基于 SAC 算法训练得到 x、y、yaw 运动模式;

3) 通过实物驱动实验, 验证了深度强化学习方法在复杂人工肌肉软体机器人运动控制中的可行性, 为该类仿生机器人的技术落地与实用化推进提供支撑.

1 仿生软体机器人样机及性能测试

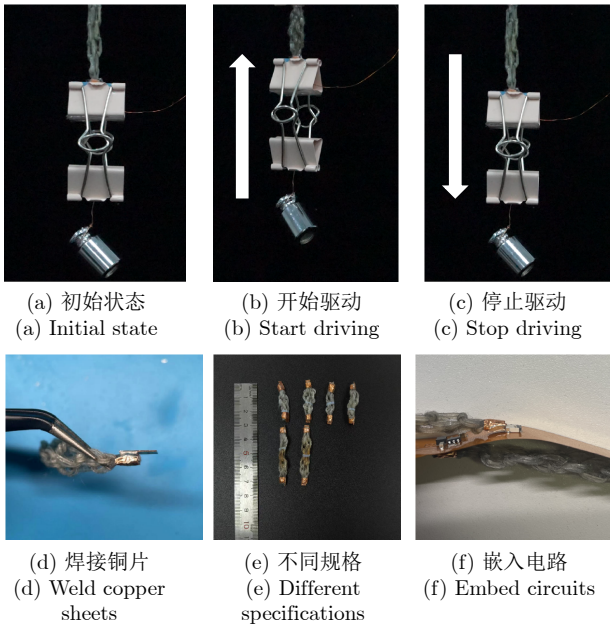
1.1 机器人物理样机设计与制作

本研究提出的仿生软体机器人如图 1 所示, 总体设计采用多股人工肌肉并联驱动与单片柔性印制电路 (flexible printed circuit board, FPCB) 一体化的思路: 机器人以十字形 FPCB 作为力学承载结构, 将驱动电路、信号采样与结构本体集成于同一柔性基材, 实现整机尺度上的结构、线路与功能一体化; 样机布置六条人工肌肉, 顶面四肢各布置一条, 底面则以两条人工肌肉交叉于中心位置, 通过选择性激活不同的人工肌肉组合, 机器人能够在横向、纵向以及扭转方向上产生连续的形变序列, 从而完成多模态弯曲、扭转和平移, 为实现复杂轨迹运动提供结构基础; FPCB 的柔性基材在弯折过程中可提供可重复、受控的形变路径, 与人工肌肉的轴向收缩形成明确的驱动−响应关系, 便于后续在 MuJoCo 环境中建立简化模型进行强化学习训练. 上述结构布局与驱动方式共同构成本文控制框架的物理基础.

图 1 原型样机

Fig. 1 The prototype

在人工肌肉的制备上, 根据 Chen 等[30] 报道的工艺方法: 首先采用可紫外固化的 LCE 墨水在基板上逐层打印细丝状预制体; 随后沿预制细丝轴向均匀缠绕直径 25 μm 的不锈钢纤维, 使其具备整体导电通路; 最后将多根细丝以锁扣−并联方式编织, 得到承载拉力更大的人工肌肉成品. 人工肌肉驱动时效果如图 2(a) ~ 2(c) 所示, 通电时, 不锈钢纤维内电流产生欧姆热, 驱动 LCE 取向相变并产生轴



(a) 初始状态 (a) Initial state　(b) 开始驱动 (b) Start driving　(c) 停止驱动 (c) Stop driving

(d) 焊接铜片 (d) Weld copper sheets　(e) 不同规格 (e) Different specifications　(f) 嵌入电路 (f) Embed circuits

图 2　人工肌肉制备与封装

Fig. 2　Fabrication and encapsulation of artificial muscles

向收缩, 进而牵拉 FPCB 产生可控弯曲; 断电后, 人工肌肉依靠与环境的热交换完成冷却, 恢复原始形态. 为实现在 FPCB 上的即插即用, 人工肌肉两端采用 3 mm × 3 mm 铜片手工烙铁焊接, 再通过连接器嵌入电路.

通过激励不同组合的人工肌肉, 可使机器人本体呈现多种形变构型, 如图 3 所示. 当按照特定时序循环驱动时, 上述离散形态会在时间维度上拼接为连续的形变轨迹, 进而驱动机器人产生宏观的运动如平移或旋转等[42−43].

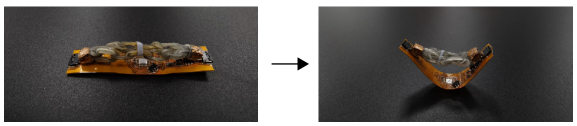


图 3　形变特征

Fig. 3　Deformation characteristic

在实物样机的运动中, 驱动的核心问题可归纳为: 1) 在给定电流波形下, 定量刻画并预测人工肌肉的输出力、应变与响应速率, 以确保形变幅值与相位的可重复性; 2) 设计一套稳定的人工肌肉驱动电路, 使全过程驱动信息始终处于可控范围.

上述问题对应于人工肌肉的力电响应特性与 FPCB 电路设计, 并共同构成运动控制的物理基础. 为此, 本文首先对锁扣并联封装的 LCE 人工肌肉进行系统力学表征, 给出恒流条件下的拉力−时间曲线与应变−时间曲线等; 随后结合 STM32G491KEU6 + TPS92200D2DDCR 驱动电路, 在机器人本体上实现稳定的人工肌肉驱动, 进而建立完整的力学性能数据库与电致驱动指标体系, 为后续控制策略设计及策略迁移奠定坚实基础.

1.2　力学性能测试

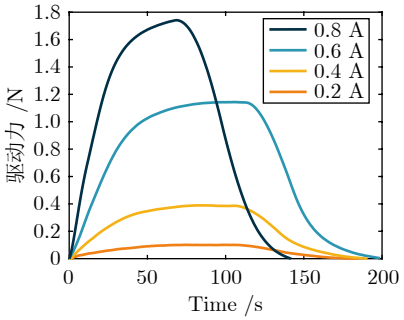
为验证人工肌肉能够提供足够的驱动力以实现预期构型变形, 本文分别对“致动端−输出力”与“被牵引端−反作用力”进行对标测试 (图 4). 首先, 将锁扣式并联封装的 LCE 人工肌肉两端夹持于拉压试验机 (图 4(a)), 在 0.2 ~ 0.8 A 区间施加恒流, 实时记录轴向拉力随时间变化的曲线, 所得数据用以刻画电致输出力随电流的标定关系. 随后, 在同一设备中将 FPCB 十字臂的两侧连接器固定 (图 4(d) 和图 4(e)), 逐步施加压缩位移以模拟肌肉收缩时的弯折, 并同步测量反作用力. 通过将两组曲线叠加比较, 可以发现, 2.9 cm 与 4.2 cm 尺寸的 FPCB 压缩形变阈值分别约为 0.35 N 与 0.18 N, 在对应尺寸的人工肌肉电流输入分别大于 0.4 A 时, 均可达到人工肌肉的稳态拉力超过 FPCB 对应形变所需的临界力这一要求.

实际运动的形成依赖于人工肌肉的收缩量. 为获得驱动电流与人工肌肉轴向应变的映射关系, 设计如图 5 所示实验: 在人工肌肉自由端加载约 40 g 负载 (与单条 FPCB 工作情况相当), 通过录制不同驱动恒流档位下的收缩视频, 并结合图像追踪方法获取人工肌肉的实时位移, 归一化计算其应变加以分析. 结果如图 5(c) 所示, 在恒流驱动中, 肌肉经历“启动加热−快速收缩−平台稳态”三阶段: 启动延迟随电流减小而加长, 平台段应变随电流增大而增大. 在输入电流为 0.6 A 时, 稳态应变超过 35%. 此外, 采用 0.5 Hz、0.2 A 方波电流驱动时, 在 1.13 g 载荷的情况下, 人工肌肉应变在 10% ~ 20% 之间波动. 该结果表明, 相较于缓慢的全程启停控制, 在未达到稳态平台期前进行周期性开关量驱动, 同样能够实现高效的周期形变控制.

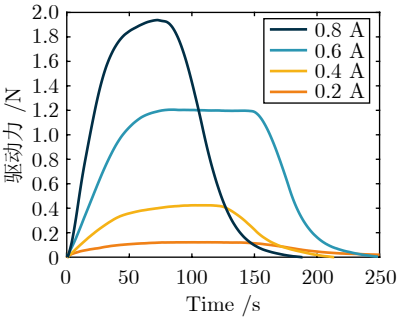
1.3　电路设计与测试

FPCB 是以柔性材料 (通常为聚酯薄膜或聚酰亚胺) 为基材制成, 具有高度可挠性的印刷电路板, 并且具有重量轻、厚度薄、弯曲性能好等特点[44]. 本文机器人采用 FPCB 集成式设计, 直接以聚酰亚胺基材作为软体机器人本体, 在其上预留人工肌肉接口, 并将器件、线路、结构一体化布局在单片柔性电路板上. 为避免柔性基材与硬质器件混装可能引起的局部刚度突变与弯折中的连接稳定性, 在布局过程中将芯片、电阻、连接端子等硬封装器件在 FPCB 形变较小的区域进行集中布局, 使关键走线尽量远

(a) 人工肌肉力输出实验
(a) Experiment on force output of artificial muscle



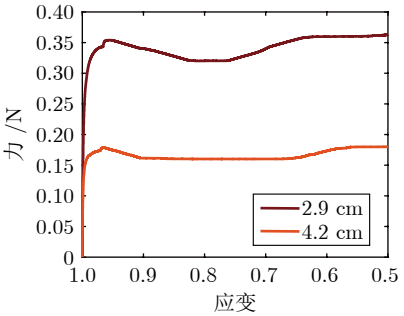
(b) 2.9 cm 人工肌肉力输出曲线
(b) Force output curve of 2.9 cm artificial muscle



(c) 4.2 cm 人工肌肉力输出曲线
(c) Force output curve of 4.2 cm artificial muscle

(d) FPCB 压缩形变实验 (压缩前)
(d) FPCB compression deformation experiment (before compression)

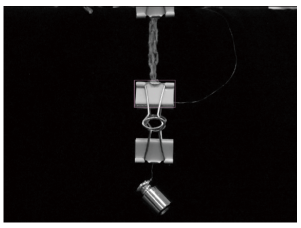
(e) FPCB 压缩形变实验 (压缩后)
(e) FPCB compression deformation experiment (after compression)



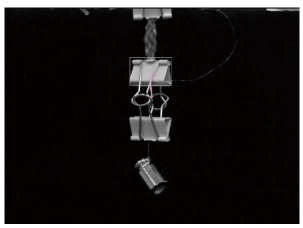
(f) FPCB 压缩
(f) FPCB compression

图 4 拉压力学性能测试

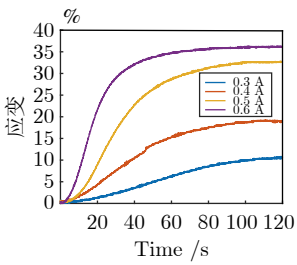
Fig. 4 Tension-compression mechanical performance testing



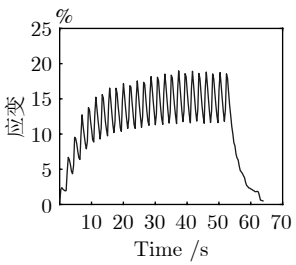
(a) 驱动前人工肌肉形态
(a) Artificial muscle before actuation



(b) 驱动后人工肌肉形态
(b) Artificial muscle after actuation



(c) 恒流驱动响应
(c) Constant-current actuation response



(d) 周期驱动响应
(d) Cyclic actuation response

图 5 人工肌肉驱动响应

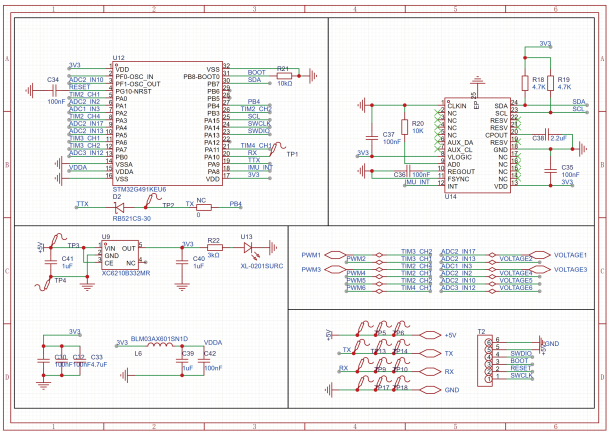
Fig. 5 Artificial muscle actuation response

离 FPCB 边缘并适当增加线宽, 从而减少在弯折过程中因硬质元器件导致的 FPCB 整体折弯性能下降, 并增强内部走线的连接可靠性. 电路板如图 6 所示.

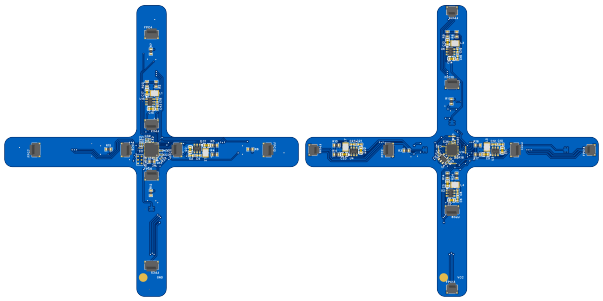
在电路逻辑方面, 人工肌肉的驱动流程如图 7 所示. 系统以 STM32G491KEU6 微控制器作为核心控制单元, 其内部定时器配置为高速脉冲宽度调制 (pulse-width modulation, PWM) 模式, 通过六路通用输入输出端口 (general-purpose input/output, GPIO) 输出 100 kHz 的脉冲信号. PWM 信号作为控制量输入至 TPS92200D2DDCR, 该芯片依据占空比将逻辑级信号转换为 0 ~ 0.8 A 的可控恒流输出, 使得人工肌肉实际收缩量由 PWM 信号占空比控制, 从而实现“弱电控强电”的驱动方式, 驱动六条人工肌肉. 人工肌肉两端布置分压采样通路, 片上模数转换器 (analog-to-digital converter, ADC) 周期性读取端电压, 以实时监测致动状态.

这种电压采样机制不仅能够实时反映致动过程中的通断状态, 而且还可用于后续扩展, 如闭环功率调节、过流保护或健康状态监测等功能. 此外, 在时序驱动中, 多通道 PWM 信号的同步由微控制单元 (microcontroller unit, MCU) 内部时钟统一管理, 可确保六路驱动指令保持稳定的相位关系, 使复合人工肌肉的协同收缩在动作节律上具有可重复性.

基于上述柔性电路结构与驱动逻辑, 机器人能



(a) 部分电路装置选择
(a) Partial circuit device selection

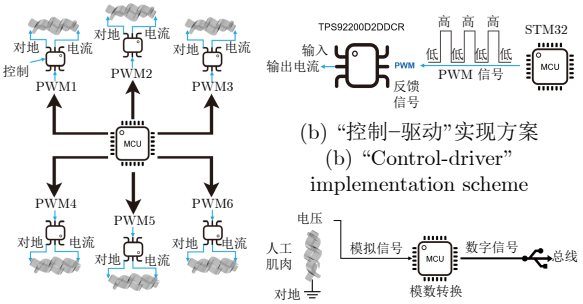


(b) 顶面电路
(b) Top surface circuit

(c) 底面电路
(c) Bottom surface circuit

图 6　器件选择与电路设计

Fig. 6　Component selection and circuit design



(a) “一对多”控制关系
(a) “One-to-many” control relationship

(b) “控制–驱动”实现方案
(b) “Control-driver” implementation scheme

(c) “驱动–反馈”实现方案
(c) “Driver-feedback” implementation scheme

图 7　人工肌肉驱动方案

Fig. 7　Artificial muscle actuation scheme

够在本体上直接实现六路人工肌肉的时序驱动. FPCB 的柔性基材在形变过程中提供可重复且受控的弯折路径, 与人工肌肉的轴向收缩形成稳定的驱动–响应关系, 使机器人最终能够实现连续的形变拼接与整体运动. 多轮实验中未观察到因器件布置或局部刚度差异导致的运动异常, 说明该电路设计对于整体运动性能影响有限且具有良好的工程可行性.

2　基于 MuJoCo 的简化建模与深度强化学习训练

2.1　基于 MuJoCo 仿真环境的模型建立

本文的训练平台选用 MuJoCo (multi-joint dynamics with contact), 该引擎基于解析力学框架, 通过稀疏雅可比与隐式积分在 CPU 即可提供 kHz 级的仿真步频, 能够在保证数值稳定的同时快速积累强化学习所需的大规模交互样本[45]. 与基于 ODE/Bullet 的 Gazebo 或依赖 NVIDIA PhysX 的 Isaac Gym 相比, MuJoCo 在刚体–关节–软接触的统一求解上具有更高的精度; 其软接触模型允许对铰链、滑动以及绳腱等多种约束进行解析计算, 恰好契合本研究“LCE 牵拉 + FPCB 弯曲”所呈现的转动、伸缩及耦合接触行为. 此外, MuJoCo 提供可定义的刚度、阻尼、迟滞等驱动特性, 据此将实验获得的等效模型嵌入求解, 从而缩小仿真–实物间的差距[46].

MuJoCo 与 OpenAI Gym、Stable-Baselines3 等深度强化学习生态已实现原生对接, 环境向量化、重放缓存管理与策略并行更新均可无缝调用, 大幅降低了算法开发与调试成本[47]. 综合计算效率、动力学保真度以及与强化学习 (reinforcement learning, RL) 框架的兼容性, MuJoCo 为复杂人工肌肉系统提供了一条“高采样–高保真–易迁移”的训练路径, 因而确定为本工作的物理仿真基础.

受外力与复杂环境接触影响, 软体机器人变形状态难以进行精确估计, 这为其控制带来诸多困难[48–51]. 为将实物运动时“人工肌肉收缩–FPCB 弯曲”的形变特征映射到仿真环境, 建模时对样机进行结构抽象与自由度压缩, 如图 8 所示. 针对 FPCB 在人工肌肉的牵拉下形变形式主要为平面弯曲这一特点, 简化模型中使用有限角位移的铰链来近似, 据此特征将其简化为刚体模型; 关于并联封装的人工肌肉产生可逆伸缩时应变沿轴向均匀分布的特性, 借助 MuJoCo 中的 tendon 元件进行等效建模, 根据实物样机设计, 在仿真样机的顶面与底面分别布置 4 条与 2 条 tendon; 简化模型驱动结构包含 8 个旋转关节与 6 条 tendon, 在空间布置上, 顶面 4 条 tendon 对应驱动四肢“远中心侧–近中心侧”铰链, 底面 2 条 tendon 对应驱动四肢“近中心侧–中心”铰链. 实物样机与简化模型形变特征对应如图 9 所示.

2.2　基于软演员–评论家算法的深度强化学习

根据深度强化学习 (DRL) 框架, 本文将复杂人工肌肉系统驱动的机器人运动过程视为一个马尔科

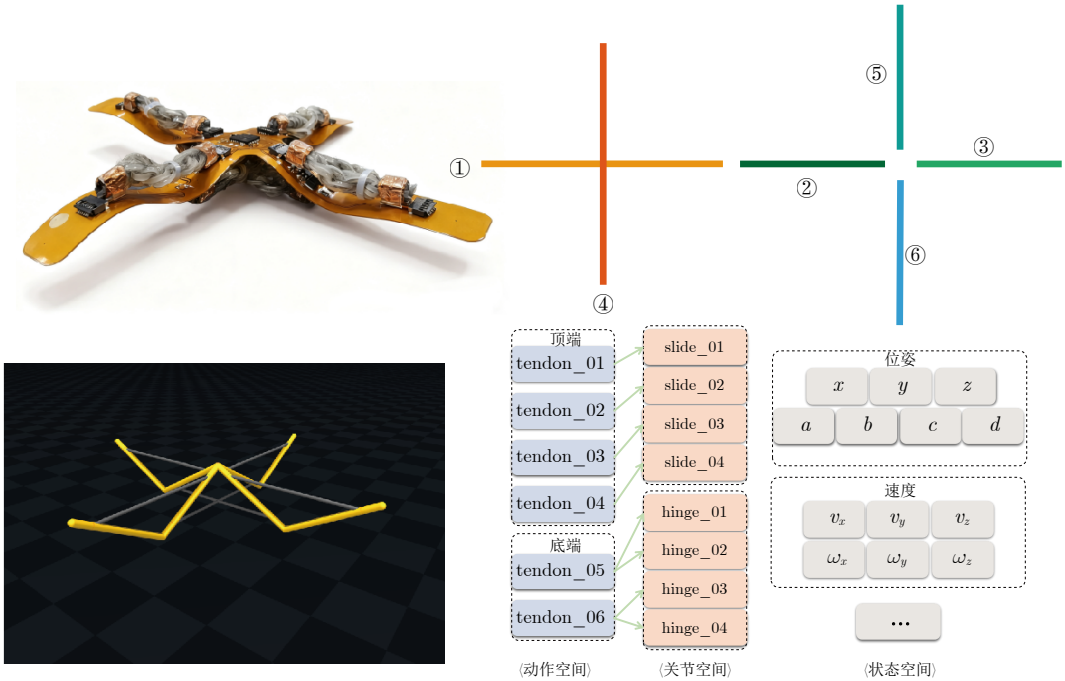


图 8 简化模型结构

Fig. 8 Simplified model structure

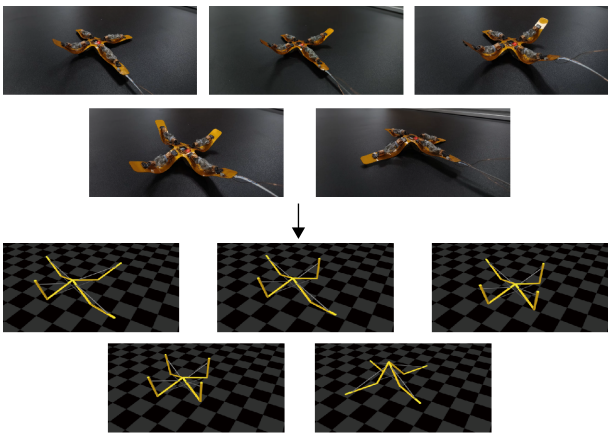


图 9 实物–仿真形变对应

Fig. 9 Physical-simulation deformation correspondence

夫决策过程 (Markov decision process, MDP), 用四元组 (S, A, P, R) 描述[52]. 其中, S 表示状态空间, 包含中心节点的空间坐标与四元数、线速度与角速度, 以及 6 条 tendon 的实时收缩率等信息; A 表示动作空间, 具体为 6 条 tendon 的控制量; P 为状态转移函数, 由 MuJoCo 引擎给出; R 为奖励函数, 综合目标位移与能量惩罚等信息, 由具体控制目标决定.

实际的仿真行走过程可表述为: 在 t 时刻, 状态为 S_t 的智能体根据策略网络 $\pi(\cdot|S_t)$ 生成动作 A_t; MuJoCo 在 S_t 与 A_t 的条件下, 根据 P 转移至新状态 S_{t+1}, 在此过程中计算获得奖励 R_t. 智能体通过最大化折扣累计回报 $\sum\gamma^t R_t$ ($\gamma \in [0, 1]$) 优化策略网络参数, 从而学习到使机器人在平面上完成 x、y 平移和 yaw 旋转的高效控制策略.

在完成 MDP 建模后, 核心问题转化为如何在连续、高维且受安全约束的动作空间内高效求得最优策略. 传统的值迭代或离散动作 DQN (deep Q-network) 难以直接处理 6 路 tendon 这类连续控制量; 而单纯的策略梯度方法又常因探索不足和收敛不稳定而难以在强非线性、迟滞显著的 LCE 动力学中获得可部署的解. 基于上述考虑, 本文选用软演员–评论家 (SAC) 算法作为学习主体. SAC 是一种面向连续动作空间的最大熵强化学习算法, 其思想是在最大化期望回报的同时引入熵正则项, 鼓励策略保持足够随机性, 从而在训练早期加速探索与在收敛后提供更强的鲁棒性. 其基于经验回放的 Off-policy 机制显著提高样本效率, 非常适合在 MuJoCo 中并行采样的大规模交互场景[53−54]. 下面将首先回顾 SAC 的基本原理和网络结构, 随后给出针对本平台的部署情况.

SAC 算法结构如图 10 所示, 神经网络包括一个 Actor 网络 (负责根据状态 s_t 输出 a_t, 并依据概率采取具体动作)、Critic1/Critic2 两个 Q 网络 (输入值是状态 s_t 和动作 a_t, 输出是 Q 值) 和 Critic target 1/2 两个目标 Q 网络 (用于稳定训练).

设 Q 网络的参数为 φ_i, 两个 Critic 网络的损失函数为

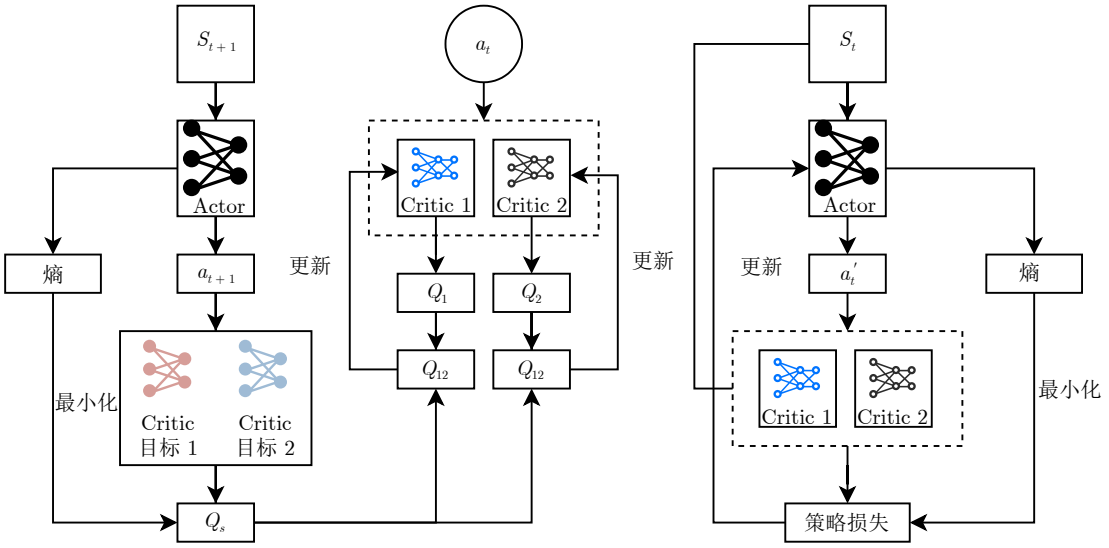


图 10　SAC 网络结构

Fig. 10　SAC network structure

$$L(\varphi_i, D) = \mathrm{E}_{(s_t,\, r,\, s_{t+1},\, a_t)\sim D}\Big[(Q_{\varphi_i}(s_t,\, a_t) - Q_s(s_{t+1}))^2\Big],\ i \in \{1,\, 2\} \tag{1}$$

式中, $(s_t,\, r,\, s_{t+1},\, a_t) \sim D$ 表示 $(s_t,\, r,\, s_{t+1},\, a_t)$ 经验来自经验池 D; $Q_{\varphi_i}(s_t,\, a_t)$ 为在参数 φ_i 下, 网络对在状态 s_t 下实施动作 a_t 的 Q 值估计.

设 Actor 网络的参数为 θ, 其更新式为

$$\max_{\theta} \mathrm{E}_{s\sim D}\left[\min_{j=1,\,2} Q_{\varphi_j}(s_t,\, a_t') - \alpha \log \pi_\theta(a_t'|s_t)\right] \tag{2}$$

目标 Critic 网络更新式为

$$\varphi_{ti} \leftarrow \rho\varphi_{ti} + (1-\rho)\varphi_{ti},\quad i \in \{1,\, 2\} \tag{3}$$

式中, $\rho \in (0,\, 1)$ 为超参数, φ_{ti} 为目标 Critic 网络参数.

在 SAC 中, 最优策略表示为

$$J(\pi) = \mathrm{E}_{\pi}\left[\sum_{t=0}^{\infty} \gamma^t \left(R(s_t,\, a_t) + \alpha H(\pi(\cdot|s_t))\right)\right] \tag{4}$$

其中, R 是奖励, γ 为折扣因子, H 为熵, 即

$$H(\pi(\cdot|s_t)) = -\mathrm{E}_{a\sim\pi(\cdot|s_t)}[\log \pi(a|s_t)] \tag{5}$$

熵值的大小反映策略 $\pi(\cdot|s_t)$ 的随机程度. $\alpha \in (0,\, 1)$ 为熵系数, 决定熵对于奖励的重要性, 通过调整 α 可以控制最优策略的探索随机性, 在获得最大回报和随机探索之间寻求平衡, 使得策略可以兼顾实时回报与探索性.

SAC 算法的伪代码如算法 1 所示.

算法 1. SAC 算法

输入. 折扣因子 γ, 经验回放池容量 P, 训练次数 M, 经验回放池采样的 Batch size 为 K.

输出. 优化后的策略 $\pi(\cdot|s_t)$.

初始化各网络参数: Actor 参数 θ, Critic 参数 φ_1, φ_2, 目标 Critic 参数 $\varphi_{t1} \leftarrow \varphi_1$, $\varphi_{t2} \leftarrow \varphi_2$
初始化温度参数 α, 经验回放池 $D \leftarrow \emptyset$
for 训练次数 $= 1,\, 2,\, \cdots,\, X$ **do**
  for 训练步数 $= 1,\, 2,\, \cdots,\, Y$ **do**
    根据当前策略 $\pi(\cdot|s_t)$ 采样动作 $a_t \sim \pi(\cdot|s_t)$
    在环境中执行 a_t, 得到下一个状态 s_{t+1} 和奖励 r_{t+1}
    将 $(s_t,\, a_t,\, r_t,\, s_{t+1})$ 存入经验池 D
  end for
  if 经验池大小 $\geq P$ **then**
    for 迭代次数 $= 1,\, 2,\, \cdots,\, Z$ **do**
      从经验池中随机采样一批数据 $(s_t,\, a_t,\, r_t,\, s_{t+1})$
      计算目标 Q 值
      更新 Critic 网络参数 $\varphi_1,\, \varphi_2$
      更新 Actor 网络参数 θ
      更新目标 Critic 网络参数 $\varphi_{t1},\, \varphi_{t2}$
      if α 可学习 **then**
        更新熵系数 α
      end if
    end for
  end if
end for

实际训练时, 为高效利用计算资源, 在多核 CPU 上并行运行 128 个独立 MuJoCo 环境, 各环境负责生成交互数据; 环境的状态、奖励与动作反馈回传主

进程, 而策略网络 (Actor/Critic) 的前向推理与梯度更新则全部在一块 5090D GPU 上执行. 由于 SAC 采用 Off-policy 结构, 所有环境回传的数据均可存入经验池并多次复用, 从而显著提高单位样本的利用率. 在这种“CPU 端高并行采样 + GPU 端集中更新”的架构下, 结合本文简化模型接触少、自由度低、动态行为稳定等特点, 在无任何基于经验的基础设定情况下, 可以在 1 小时内获得稳定收敛的 x、y 与 yaw 三类运动策略, 训练效率相比传统软体机器人 RL 场景显著提高. 并行训练的流程如图 11 所示.

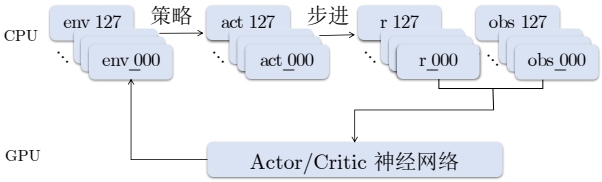


图 11 并行训练

Fig. 11 Parallel training

2.3 运动控制策略训练

在实际的深度强化学习训练时, 机器人的状态空间 S 与动作空间 A 如图 8 所示, 状态转移函数 P 由 MuJoCo 引擎给出, 决定获得何种策略的关键在于奖励函数 R 的设置. 本文将奖励设计为“速度奖励−能量惩罚−防停滞项”的组合, 具体为

$$r_{(\xi,\,\sigma)}(t)=\sigma w_\xi u_\xi(t)-w_E E_t-w_s\exp\left(-\frac{|u_\xi(t)|}{w_s}\right),$$
$$\sigma\in\{+1,\,-1\},\quad \xi\in\{x,\,y,\,\psi\} \tag{6}$$

其中, ξ 的取值对应于 x、y、yaw 三种平面任务中的情况, 情况分别为

$$u_x(t)=v_x(t),\ u_y(t)=v_y(t),\ u_\psi(t)=\omega_z(t) \tag{7}$$

$$\begin{cases} w_x=w_y=w_v,\ w_\psi=w_\omega \\ s_x=s_y=v_s,\ s_\psi=\omega_s \end{cases} \tag{8}$$

式 (7) 和式 (8) 中, ω_z 表示沿 z 轴旋转的角速度, 即为平面任务 yaw 的角速度; w_v 表示平面移动运动的权重; w_ω 表示平面旋转运动的权重; v_s 表示平面移动运动的防停滞权重相关线速度; ω_s 表示平面旋转运动的防停滞权重相关角速度; σ 为方向系数, 其取值对应不同运动方向, 与 ξ 共同决定当前是何训练任务, 可分为以下六种情况, 即

$$\begin{aligned} &x^+:(\xi=x,\,\sigma=+1), && x^-:(\xi=x,\,\sigma=-1)\\ &y^+:(\xi=y,\,\sigma=+1), && y^-:(\xi=y,\,\sigma=-1)\\ &yaw^+:(\xi=\psi,\,\sigma=+1), && yaw^-:(\xi=\psi,\,\sigma=-1) \end{aligned} \tag{9}$$

另外, 式 (6) 中 $\sigma w_\xi u_\xi(t)$ 为速度奖励项, 促使机器人以更高的速度运动. 其中, w_ξ 为速度权重, $u_\xi(t)$ 为 t 时刻的速度值. $w_E E_t$ 为能量惩罚项, 促使机器人获取较低能耗策略, 其中 w_E 为能耗权重, E_t 为等效能耗, 其计算方法为

$$E_t=\sum_{i=1}^{6}k_{\ell,\,i}\left(\delta\ell_i(t)\right)^2\Delta t^2,$$
$$\delta\ell_i(t)=[\ell_i^0-\ell_i(t)]_+=\max\left(0,\,\ell_i^0-\ell_i(t)\right) \tag{10}$$

式 (10) 中, ℓ_i^0 为第 i 条人工肌肉的松弛长度, $\ell_i(t)$ 为其实时长度, $k_{\ell,\,i}\geq 0$ 为收缩能量惩罚系数, Δt 为仿真步长. 该定义使惩罚随“实时收缩量”增大而增大, 并按 Δt^2 进行时间尺度加权.

最后, 式 (6) 中 $w_s\exp(-\frac{|u_\xi(t)|}{w_s})$ 为防停滞项, 防止机器人原地不动. 其中, w_s 为停滞权重, 该防停滞项在 $u_\xi(t)=0$ 时等于 w_s, 惩罚最大; 而当 $|u_\xi(t)|>0$ 时, 快速衰减到接近 0, 既不推动过高速度抢占速度奖励项作用, 又能可靠避免“停着不动”.

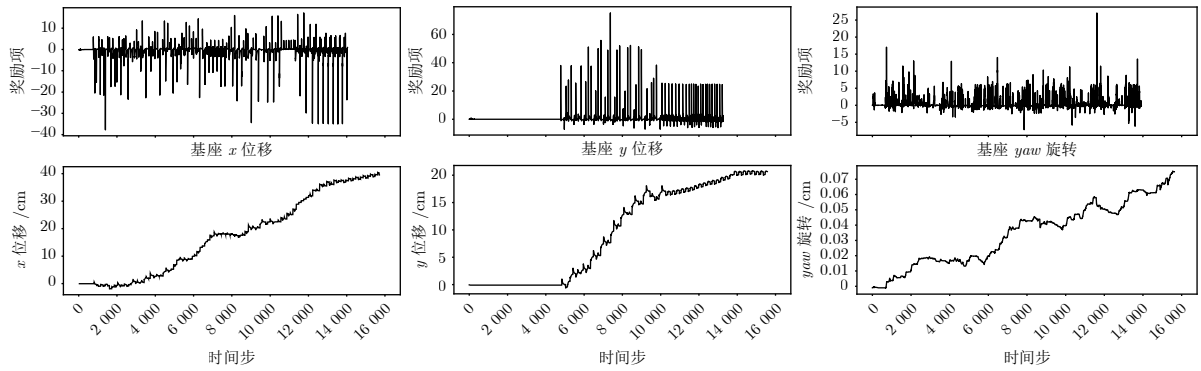


图 12 强化学习训练成果

Fig. 12 Reinforcement learning training results

三种运动模式训练结果如图 12 所示, 通过对不同人工肌肉的差异化时序驱动, 机器人可分别实现 x、y 平移和 yaw 旋转. 图 12 的每个训练任务中, 第 1 行表示全过程奖励获取情况; 第 2 行表示仿真样机在目标运动方向上的位移 (或旋转角) 随时间的变化.

3　机器人样机运动实验

按照第 1 节中所述制备样机的组件并完成样机搭建. 采用第 2 节中图 12 所示的策略流程, 从三种运动模式的时序序列中提取完整周期段作为离线驱动方案, 并根据人工肌肉的通道对应关系, 将其以电流波形形式输入至相应驱动通道.

实物样机采用该策略的离线驱动效果如图 13 ~ 15 所示. 可见以此方案进行驱动, 机器人样机成功实现了 x、y 平移和 yaw 旋转, 其运动与旋转速度分别约为 1.5 mm/周期与 1.2 (°)/周期. 图 13 ~ 15 分别展示了在执行三种任务时, 仿真中样机运动情况、仿真中实时驱动情况与实物运动情况, 人工肌肉对应见图 8.

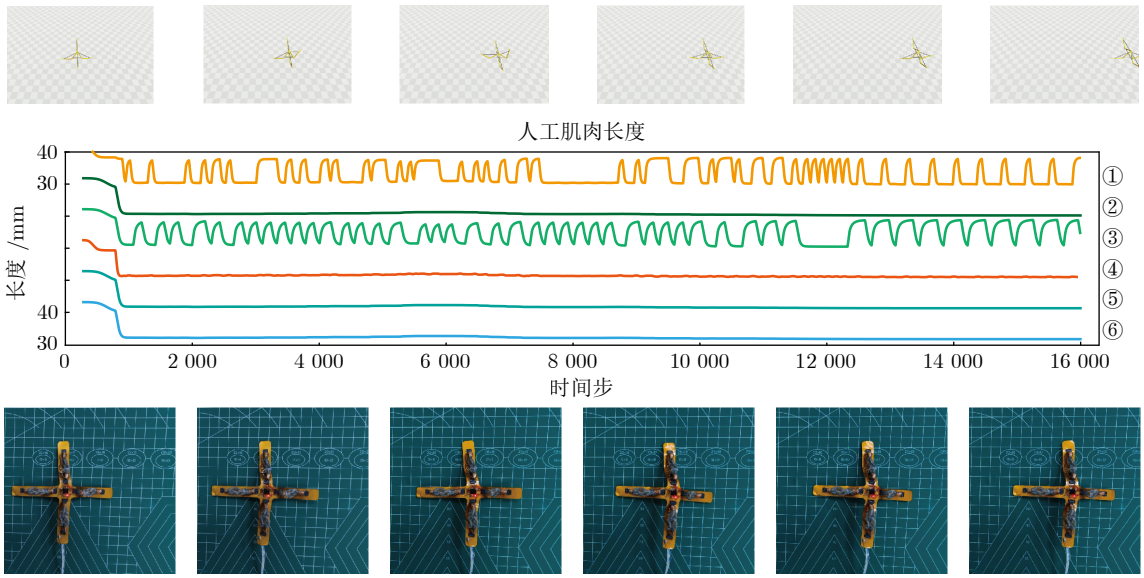


图 13　离线 x 方向运动

Fig. 13　Offline translation along the x-axis

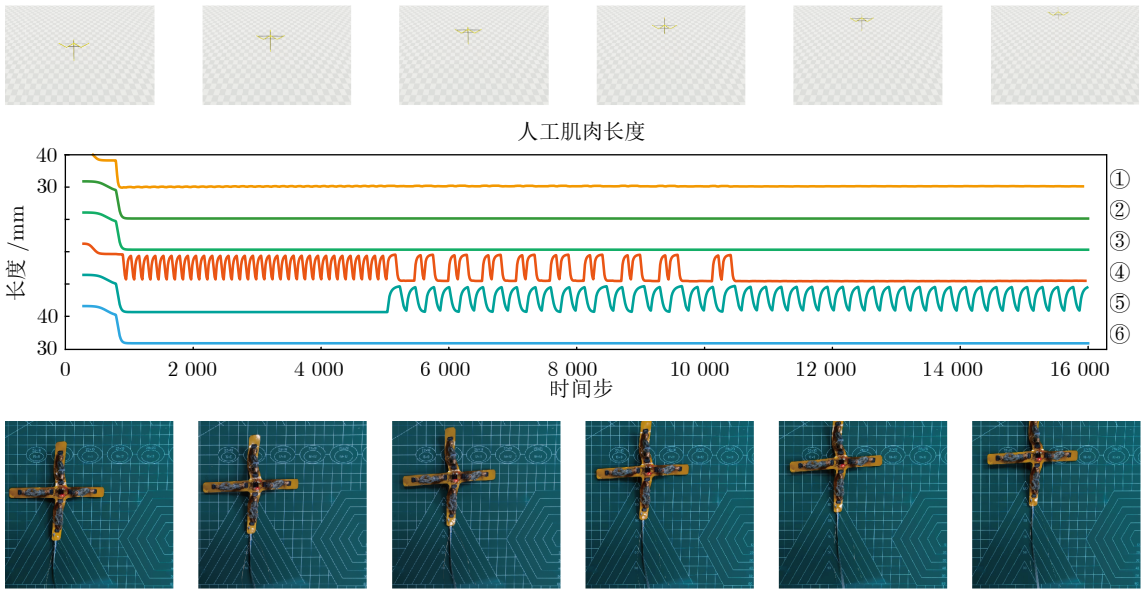


图 14　离线 y 方向运动

Fig. 14　Offline translation along the y-axis

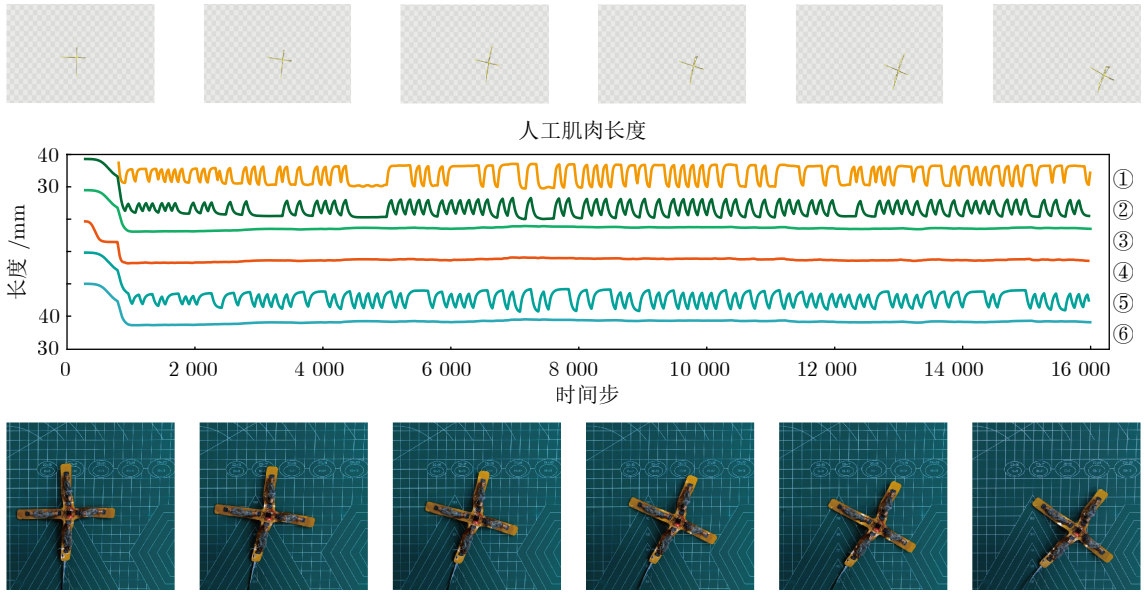


图 15 离线 *yaw* 方向旋转

Fig. 15 Offline rotation about the *yaw* axis

需要说明的是, 由于本研究以运动模式验证为核心目标, 上述策略在仿真中基于简化桁架模型获得, 策略部署采用离线周期波形输入, 而非依赖实时仿真推理, 因此对模型细节一致性的依赖性较弱. 此外, 人工肌肉驱动采用开环电流控制方式, 通过统一配方、工艺与规格的批量制备, 使材料离散性对整体运动的影响保持在可接受范围内. 实际运行中, 通过多轮周期驱动实验, FPCB 未出现焊点开裂、线路损坏或接触不良现象, 说明当前驱动频率与形变幅值处于安全工作范围内.

在图 13 所示的机器人离线 x 方向运动实验中, 截取仿真过程第 14 000 ~ 16 000 时间步之间的信号段作为驱动策略. 根据运动过程的观察, 可将该策略分为三个阶段: 首先, 3 号人工肌肉 (顶面 x 方向) 收缩, 使机器人前肢沿 x 方向抬起, 此时, 抬起侧中段与地面的接触增强, 形成稳定支点; 随后, 1 号人工肌肉 (底面 x 方向) 收缩, 由于前肢已处于抬起状态, 前后肢与地面之间的摩擦条件产生差异, 因此推动整体结构向 x 正方向发生位移; 最后, 两条人工肌肉同时舒张, 机器人出现一定幅度的回缩, 但该阶段产生的位移量小于前一阶段的正向推进量. 由此, 单个周期内的净位移沿 x 正方向累积, 实现沿该方向的连续推进.

在图 14 所示的机器人离线 y 方向运动实验中, 截取仿真过程第 6 000 ~ 8 000 时间步的信号段作为驱动策略. y 正方向的运动机制与 x 正方向基本一致, 仅人工肌肉的激活顺序发生对应变化: 由 4 号人工肌肉 (底面 y 方向) 与 5 号人工肌肉 (顶面 y 方向) 按时序交替收缩与释放, 以生成沿 y 方向的周期性形变序列. 由于该形变过程在不同阶段产生的推进量与回缩量不对称, 单个周期内的净位移沿 y 正方向累积, 从而实现沿该方向的连续运动.

在图 15 所示的机器人离线 *yaw* 方向旋转实验中, 截取仿真过程第 9 000 ~ 10 000 时间步作为驱动策略. 根据对运动过程的观察, 同样将旋转策略分为三个阶段: 首先, 2 号人工肌肉 (顶面 x 方向) 与 5 号人工肌肉 (顶面 y 方向) 同时收缩, 机器人 x、y 方向前肢均抬起, 中段紧压地面; 随后, 1 号人工肌肉 (底面 x 方向) 收缩, 在原本沿 x、y 双轴对称的驱动布局中引入非对称量, 使整体结构产生绕竖直轴的偏转, 从而形成旋转趋势; 最后, 各人工肌肉释放, 机器人完成一次完整的旋转周期. 多周期的连续驱动使得净角位移累积, 实现了沿 *yaw* 方向的持续旋转运动.

4 结束语

本文根据自然界生物体运动中“形态–行为–智能”思想, 聚焦于冗余人工肌肉驱动的仿生机器人, 提出并制作一款由基于 LCE 的人工肌肉驱动的仿生软体机器人及其配置电路. 针对该机器人运动控制这一难点, 本文选用 MuJoCo 仿真环境, 建立简化模型, 并采用 SAC 算法进行强化学习, 获得了 x、y、*yaw* 三个方向的可控运动策略. 结合上述两个部分, 本文将运动策略中的周期部分作为离线策略部署于机器人实物样机, 驱动中实现了 1.5 mm/周期的平移与 1.2 (°)/周期的旋转. 在未来研究中, 我们将进行在线策略部署工作, 形成更加智能的仿生运动方案.

参考文献

1 Joachimczak M, Suzuki R, Arita T. Improving evolvability of morphologies and controllers of developmental soft-bodied robots with novelty search. *Frontiers in Robotics and AI*, 2015, **2**: Article No. 00033

2 Woodward M A, Sitti M. Morphological intelligence counters foot slipping in the desert locust and dynamic robots. *Proceedings of the National Academy of Sciences of the United States of America*, 2018, **115**(36): E8358−E8367

3 Ghazi-Zahedi K, Haeufle D F B, Montufar G, Schmitt S, Ay N. Evaluating morphological computation in muscle and DC-motor driven models of human hopping. arXiv preprint arXiv: 1512.00250, 2015.

4 Uppington M, Gobbo P, Hauert S, Hauser H. Evolving and generalising morphologies for locomoting micro-scale robotic agents. *Journal of Micro and Bio Robotics*, 2022, **18**: 37−47

5 Wang Jiu-Bin, He Wei, Meng Ting-Ting, Zou Yao, Fu Qiang. System design of dove-like flapping-wing flying robot based on highly bionic morphological layout. *Acta Automatica Sinica*, 2024, **50**(2): 308−319
(王久斌, 贺威, 孟亭亭, 邹尧, 付强. 基于高仿生形态布局的仿鸽扑翼飞行机器人系统设计. 自动化学报, 2024, **50**(2): 308−319)

6 Wang T Y, Pierce C, Kojouharov V, Chong B X, Diaz K, Lu H, et al. Mechanical intelligence simplifies control in terrestrial limbless locomotion. *Science Robotics*, 2023, **8**: Article No. eadi2243

7 Chen A, Song B F, Liu K, Wang Z H, Xue D, Qi H D. Flapping-wing robot achieves bird-style self-takeoff by adopting reconfigurable mechanisms. *Science Advances*, 2025, **11**: Article No. eadx0465

8 Zhong Q, Zhu J, Fish F E, Kerr S J, Downs A M, Bart-Smith H, et al. Tunable stiffness enables fast and efficient swimming in fish-like robots. *Science Robotics*, 2021, **6**: Article No. eabe4088

9 Wen L, Ren Z Y, Di Santo V, Hu K N, Yuan T, Wang T M, et al. Understanding fish linear acceleration using an undulatory biorobotic model with soft fluidic elastomer actuated morphing median fins. *Soft Robotics*, 2018, **5**(4): 375−388

10 Wu Zheng-Xing, Yu Jun-Zhi, Tan Min. Comparison of two methods to implement backward swimming for a carangiform robotic fish. *Acta Automatica Sinica*, 2013, **39**(12): 2032−2042
(吴正兴, 喻俊志, 谭民. 两类仿鲹科机器鱼倒游运动控制方法的对比研究. 自动化学报, 2013, **39**(12): 2032−2042)

11 Liu Z M, Liu J Q, Wang H, Yu X, Yang K, Liu W B, et al. A 1 mm-thick miniatured mobile soft robot with mechanosensation and multimodal locomotion. *IEEE Robotics and Automation Letters*, 2020, **5**(2): 3291−3298

12 Zhang Y F, Yang D Z, Yan P N, Zhou P W, Zou J, Gu G Y. Inchworm inspired multimodal soft robots with crawling, climbing, and transitioning locomotion. *IEEE Transactions on Robotics*, 2022, **38**(3): 1806−1819

13 Ren Z Y, Sitti M. Design and build of small-scale magnetic soft-bodied robots with multimodal locomotion. *Nature Protocols*, 2024, **19**: 441−486

14 Hu W Q, Lum G Z, Mastrangeli M, Sitti M. Small-scale soft-bodied robot with multimodal locomotion. *Nature*, 2018, **554**: 81−85

15 Niu J W, Zhang F W, Liu C L, Xie K R, Zhang Y X, Zhang J, et al. Magnetically driven biomimetic microrobot loaded with eleutheroside B for targeted delivery and neural repair in spinal cord injury. *ACS Applied Materials and Interfaces*, 2025, **17**(30): 42688−42705

16 Yu S M, Zhang W W, Feng Y Z, Zhang X, Li C H, Shi S J, et al. Magnetic cell-mimetic droplet microrobots with division and exocytosis capabilities. *Research*, 2025, **8**: Article No. 0730

17 Li T L, Yu S M, Sun B, Li Y L, Wang X L, Pan Y L, et al. Bioinspired claw-engaged and biolubricated swimming microrobots creating active retention in blood vessels. *Science Advances*, 2023, **9**: Article No. eadg4501

18 Pan F, Liu J Q, Zuo Z H, He X, Shao Z Y, Chen J Y, et al. Miniature deep-sea morphable robot with multimodal locomotion. *Science Robotics*, 2025, **10**: Article No. eadp7821

19 Feng R Y, He Y M, Feng S Y, Li S G. Impulsive actuation for soft robots. *npj Robotics*, 2025, **3**: Article No. 27

20 Xu Y, Zhuo J S, Fan M Y, Li X, Cao X N, Ruan D R, et al. A bioinspired shape memory alloy based soft robotic system for deep-sea exploration. *Advanced Intelligent System*, 2024, **6**: Article No. 2300699

21 Huang X N, Kumar K, Jawed M K, Nasab A M, Ye Z S, Shan W L, et al. Chasing biomimetic locomotion speeds: Creating untethered soft robots with shape memory alloy actuators. *Science Robotics*, 2018, **3**: Article No. eaau7557

22 Gu G Y, Zou J, Zhao R K, Zhao X H, Zhu X Y. Soft wall-climbing robots. *Science Robotics*, 2018, **3**: Article No. eaat2874

23 Wang X X, Pei X, Wang X Y, Hou T G. Lightweight untethered soft robotic fish. In: Proceedings of the IEEE International Conference on Robotics and Automation (ICRA). Yokohama, Japan: IEEE, 2024. 669−675

24 Shintake J, Cacucciolo S, Shea H, Floreano D. Soft biomimetic fish robot made of dielectric elastomer actuators. *Soft Robotics*, 2018, **5**(4): 466−474

25 Wang T L, Joo H J, Song S Y, Hu W Q, Keplinger C, Sitti M. A versatile jellyfish-like robotic platform for effective underwater propulsion and manipulation. *Science Advances*, 2023, **9**: Article No. eadg0292

26 Tao Zi-Chen, Liu Song-Yuan, Gui Yun, Hao Si-Yuan, Fang Hao, Yang Qing-Kai. Design and control of tensegrity based cross-domain robot. *Robot*, 2025, **47**(3): 338−347, 360
(陶子辰, 刘松源, 桂昀, 郝思远, 方浩, 杨庆凯. 张拉整体跨域机器人的设计与控制. 机器人, 2025, **47**(3): 338−347, 360)

27 Mo J X, Gao C Q, Fang H, Yang Q K. Design and locomotion characteristic analysis of a novel tensegrity hopping robot. In: Proceedings of the IEEE International Conference on Robotics and Biomimetics (ROBIO). Koh Samui, Thailand: IEEE, 2023. 1−8

28 Mo J X, Fang H, Yang Q K. Design and locomotion characteristic analysis of two kinds of tensegrity hopping robots. *iScience*, 2024, **27**(3): Article No. 109226

29 Chen Wen-Hui, Zhou Xiao-Hang, Liu Ke. Application of liquid crystal elastomers in the development of artificial muscles. *Chinese Journal of Liquid Crystals and Displays*, 2025, **40**(2): 201−217
(陈雯慧, 周晓航, 刘珂. 液晶弹性体在人工肌肉领域的研究进展. 液晶与显示, 2025, **40**(2): 201−217)

30 Chen W H, Tong D Z, Meng L H, Tan B W, Lan R C, Zhang Q F, et al. Knotted artificial muscles for bio-mimetic actuation under deepwater. *Advanced Materials*, 2024, **36**: Article No. 2400763

31 Chen W H, Yang S A, Zhu C, Cheng Y T, Shi Y T, Yu C P, et al. Scalable jet swimmer driven by pulsatile artificial muscles and soft chamber buckling. *Advanced Materials*, 2025, **37**: Art-

icle No. 2503777

32 Lai M, Go K, Li Z B, Kroger T, Schaal S, Allen K, et al. RoboBallet: Planning for multirobot reaching with graph neural networks and reinforcement learning. *Science Robotics*, 2025, **10**: Article No. eads1204

33 Cao S J, Sun L, Jiang J J, Zuo Z Y. Reinforcement learning-based fixed-time trajectory tracking control for uncertain robotic manipulators with input saturation. *IEEE Transactions on Neural Networks and Learning Systems*, 2023, **34**(8): 4584−4595

34 Pavlichenko D, Behnke S. Real-robot deep reinforcement learning: Improving trajectory tracking of flexible-joint manipulator with reference correction. In: Proceedings of the IEEE International Conference on Robotics and Automation (ICRA). Philadelphia, PA, USA: IEEE, 2022. 2671−2677

35 He J Z, Zhang C, Jenelten F, Grandia R, Bacher M, Hutter M. Attention-based map encoding for learning generalized legged locomotion. *Science Robotics*, 2025, **10**: Article No. eadv3604

36 Huang H D, Sun S L, Zhao Z D, Huang H L, Shen C Q, Xu W F. PTRL: Prior transfer deep reinforcement learning for legged robots locomotion. arXiv preprint arXiv: 2504.05629, 2025.

37 Li Yuan-Chao, Tao Chong-Ben, Wang Chen. Gait control method based on maximum entropy deep reinforcement learning for biped robot. *Journal of Computer Applications*, 2024, **44**(2): 445−451
(李源潮, 陶重犇, 王琛. 基于最大熵深度强化学习的双足机器人步态控制方法. 计算机应用, 2024, **44**(2): 445−451)

38 Wu Xiao-Guang, Liu Shao-Wei, Yang Lei, Deng Wen-Qiang, Jia Zhe-Heng. A gait control method for biped robot on slope based on deep reinforcement learning. *Acta Automatica Sinica*, 2021, **47**(8): 1976−1987
(吴晓光, 刘绍维, 杨磊, 邓文强, 贾哲恒. 基于深度强化学习的双足机器人斜坡步态控制方法. 自动化学报, 2021, **47**(8): 1976−1987)

39 Ma J C, Lu H M, Xiao J H, Zeng Z W, Zheng Z Q. Multi-robot target encirclement control with collision avoidance via deep reinforcement learning. *Journal of Intelligent and Robotic Systems*, 2020, **99**: 371−386

40 Zhou Z Q, Zhu P M, Zeng Z W, Xiao J H, Lu H M, Zhou Z T. Robot navigation in a crowd by integrating deep reinforcement learning and online planning. *Applied Intelligence*, 2022, **52**: 15600−15616

41 Hua H, Wang Y N, Zhong H, Zhang H, Fang Y C. Deep reinforcement learning-based hierarchical motion planning strategy for multirotors. *IEEE Transactions on Industrial Informatics*, 2025, **21**(6): 4324−4333

42 Zhu Ya-Zhou, Liu Yu-Ying, Wang Ya-Dong, Xie Hui-Ting, Li Gong-Xin. Inchworm-like soft robot based on liquid crystal elastomer. *Chinese Journal of Liquid Crystals and Displays*, 2025, **40**(4): 527−535
(朱亚洲, 刘煜莹, 王亚东, 谢慧婷, 李恭新. 基于液晶弹性体的仿尺蠖软体机器人. 液晶与显示, 2025, **40**(4): 527−535)

43 Wu S, Hong Y Y, Zhao Y, Yin J, Zhu Y. Caterpillar-inspired soft crawling robot with distributed programmable thermal actuation. *Science Advances*, 2023, **9**: Article No. eadf8014

44 Rogers J A, Someya T, Huang Y. Materials and mechanics for stretchable electronics. *Science*, 2010, **327**: 1603−1607

45 Todorov E, Erez T, Tassa Y. MuJoCo: A physics engine for model-based control. In: Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). Vilamoura-Algarve, Portugal: IEEE, 2012. 5026−5033

46 Kumar S, Narayanan M S, Singhal P, Corso J J, Krovi V. Surgical tool attributes from monocular video. In: Proceedings of the IEEE International Conference on Robotics and Automation (ICRA). Hong Kong, China: IEEE, 2014. 4887−4892

47 Brockman G, Cheung V, Pettersson L, Schneider J, Schulman J, Tang J, et al. OpenAI Gym. arXiv preprint arXiv: 1606.01540, 2016.

48 Ma J, Han Z J, Yang L S, Min G C, Liu Z J, He W. Dynamics modeling of a soft arm under the Cosserat theory. In: Proceedings of the IEEE International Conference on Real-time Computing and Robotics (RCAR). Xining, China: IEEE, 2021. 87−90

49 Li J Z, Ma J, Hu Y J, Zhang L, Liu Z J, Sun S Y. Vision-based reinforcement learning control of soft robot manipulators. *Robotic Intelligence and Automation*, 2024, **44**(6): 783−790

50 Yang Yan, Liu Yun-Peng, Han Jiang-Tao, Liu Zhi-Jie, Han Zhi-Ji. Modeling and neural network control of a soft manipulator. *Chinese Journal of Engineering*, 2023, **43**(3): 454−464
(杨妍, 刘运鹏, 韩江涛, 刘志杰, 韩志冀. 软体机械臂的建模与神经网络控制. 工程科学学报, 2023, **43**(3): 454−464)

51 Cheng Yi-Tao, Yang Huang-Yu, Liu Ke. Reduced order model for soft robotic surface based on beam elements. *Robot*, 2025, **47**(5): 646−656
(程屹涛, 杨焕煜, 刘珂. 基于梁单元的曲面软体机器人简化力学模型. 机器人, 2025, **47**(5): 646−656)

52 Sutton R S, Barto A G. *Reinforcement Learning: An Introductions, 2nd ed.* Cambridge, MA: MIT Press, 2018.

53 Haarnoja T, Zhou A, Hartikainen K, Tucker G, Ha S, Tan J, et al. Soft actor-critic algorithms and applications. arXiv preprint arXiv: 1812.05905, 2018.

54 Haarnoja T, Zhou A, Abbeel P, Levine S. Soft actor-critic: Off-policy maximum entropy deep reinforcement learning with a stochastic actor. arXiv preprint arXiv: 1801.01290, 2018.

牛鹏军 北京大学先进制造与机器人学院博士研究生. 2025 年获得北京航空航天大学机械工程及自动化学院学士学位. 主要研究方向为机器人仿真与控制.
E-mail: pjniu25@stu.pku.edu.cn
(**NIU Peng-Jun** Ph.D. candidate at the School of Advanced Manufacturing and Robotics, Peking University. He received his bachelor degree from the School of Mechanical Engineering and Automation, Beihang University in 2025. His research interests include robot simulation and control.)

程屹涛 北京大学先进制造与机器人学院博士研究生. 2023 年获得北京大学工学院学士学位. 主要研究方向为软体机器人, 机器人感知与控制, 人机交互.
E-mail: chengyitao@pku.edu.cn
(**CHENG Yi-Tao** Ph.D. candidate at the School of Advanced Manufacturing and Robotics, Peking University. He received his bachelor degree from the College of Engineering, Peking University in 2023. His research interests include soft robot, robot perception and control, and human-machine interaction.)

朱彦臣 华中科技大学智能制造装备与技术全国重点实验室博士研究生. 2023 年获得四川大学机械工程学院学士学位. 主要研究方向为柔性电子, 软体机器人.
E-mail: yanchenshizhu@hust.edu.cn
(**ZHU Yan-Chen** Ph.D. candidate at the State Key Laboratory of Intelligent Manufacturing Equipment and Technology, Huazhong University of Science and Technology. He received his bachelor degree from the School of Mechanical Engineering, Sichuan University in 2023. His research interests include flexible electronics and soft robotics.)

厉 侃 华中科技大学智能制造装备与技术全国重点实验室研究员. 2019 年获得美国西北大学理论与应用力学专业博士学位. 主要研究方向为三维柔性微飞行器, 三维柔性可拉伸电子器件.
E-mail: kanli@hust.edu.cn
(**LI Kan** Research fellow at the State Key Laboratory of Intelligent Manufacturing Equipment and Technology, Huazhong University of Science and Technology. He received his Ph.D. degree in theoretical and applied mechanics from Northwestern University, USA, in 2019. His research interests include three-dimensional flexible micro aerial vehicles and three-dimensional flexible and stretchable electronic devices.)

刘 珂 北京大学先进制造与机器人学院研究员. 2019 年获得美国佐治亚理工学院博士学位. 主要研究方向为柔性结构与软体机器的设计、分析与应用. 本文通信作者.
E-mail: liuke@pku.edu.cn
(**LIU Ke** Research fellow at the School of Advanced Manufacturing and Robotics, Peking University. He received his Ph.D. degree from Georgia Institute of Technology, USA, in 2019. His research interests include design, analysis and application of flexible structures and soft machines. Corresponding author of this paper.)