面向不同粗糙程度地面的四足机器人自适应控制方法

张楠杰 [1] 陈玉全 [1] 季茂沁 [1] 孙运康 [1] 王 冰 [1]

摘 要 针对四足机器人在复杂环境中的高速稳定运动问题, 提出一种融合模型和学习的分层运动控制框架. 首先, 提出基于单次落足点偏差的惩罚机制, 实现对连续滑动状态的有效评估. 其次, 构建基于双曲正切函数的连续接触状态描述, 显著改善了传统离散方法中的相位切换冲击问题. 然后, 设计基于 LSTM 的地面特性实时估计网络, 实现质心位置的自适应调整. 最后, 提出基于执行层和决策层的分层控制框架, 提高系统的环境适应能力. 在 Isaac Gym 仿真环境中的实验表明, 该控制方法能够适应不同摩擦系数和运动速度条件. 特别是在极低摩擦环境 ($\mu = 0.05$) 下, 自适应控制策略通过 0.061 0 m 的质心高度调整, 在维持 1.428 4 m/s 运动速度的同时, 将足端滑动距离控制在 $0.308 \pm 0.005\,0$ cm, 充分验证了所提控制方法的有效性和实用价值.

关键词 四足机器人, 强化学习, 自适应控制策略, 奖励函数优化, 分层控制框架

引用格式 张楠杰, 陈玉全, 季茂沁, 孙运康, 王冰. 面向不同粗糙程度地面的四足机器人自适应控制方法. 自动化学报, 2025, **51**(7): 1585−1598

DOI 10.16383/j.aas.c240738 **CSTR** 32138.14.j.aas.c240738

Adaptive Control Method for Quadruped Robot Facing Floors of Different Roughness

ZHANG Nan-Jie[1] CHEN Yu-Quan[1] JI Mao-Qin[1] SUN Yun-Kang[1] WANG Bing[1]

Abstract Addressing the issue of high-speed stable motion for quadruped robots in complex environments, a hierarchical motion control framework integrating model and learning is proposed. First, a penalty mechanism based on single foot placement point deviation is introduced to effectively evaluate continuous sliding states. Second, a continuous contact state description based on hyperbolic tangent function is constructed, significantly improving the phase switching impact problem in traditional discrete methods. Then, a LSTM-based ground characteristics real-time estimation network is designed to achieve adaptive adjustment of the centroid of mass position. Finally, a hierarchical control framework based on execution and decision layers is proposed to enhance the system′s environmental adaptability. Experiments in the Isaac Gym simulation environment demonstrate that this control method can adapt to different friction coefficients and motion speeds. Particularly in an extremely low friction environment ($\mu = 0.05$), the adaptive control strategy adjusts the centroid of mass height by 0.061 0 m, while maintaining a motion speed of 1.428 4 m/s and controlling the sliding distance of the foot end to $0.308 \pm 0.005\,0$ cm. This outcome serves to provide a comprehensive demonstration of the effectiveness and practical value of the proposed control method.

Key words Quadruped robot, reinforcement learning, adaptive control strategy, reward function optimization, hierarchical control framework

Citation Zhang Nan-Jie, Chen Yu-Quan, Ji Mao-Qin, Sun Yun-Kang, Wang Bing. Adaptive control method for quadruped robot facing floors of different roughness. *Acta Automatica Sinica*, 2025, **51**(7): 1585−1598

四足机器人凭借其出色的环境适应性和运动灵活性, 在工业巡检、灾害救援、极地探索等领域具有广阔的应用前景. 然而, 随着应用场景的不断拓展, 四足机器人面临着严峻的控制挑战: 从崎岖不平的野外地形到潮湿光滑的室内地面, 从易变形的沙地到狭窄的障碍物密集区域. 这些复杂的环境条件对机器人的运动控制提出更高要求, 尤其在光滑地面环境下, 传统控制方法难以兼顾运动的稳定性和灵活性, 限制了四足机器人的实际应用. 相比之下, 四足动物能够在各种复杂环境中展现出优异的运动能力, 可以精确调整身体姿态, 灵活选择落脚点, 从容应对各类地形挑战[1].

为实现四足动物般的灵活运动能力, 研究人员对四足机器人控制进行广泛探索. 传统的基于模型的方法, 如模型预测控制 (Model predictive con-

收稿日期 2024-11-14 录用日期 2025-03-05
Manuscript received November 14, 2024; accepted March 5, 2025
国家自然科学基金 (51777058) 资助
Supported by National Natural Science Foundation of China (51777058)
本文责任编委 李鸿一
Recommended by Associate Editor LI Hong-Yi
1. 河海大学人工智能与自动化学院 常州 213000
1. College of Artificial Intelligence and Automation, Hohai University, Changzhou 213000

trol, MPC), 主要通过简化动力学模型实现在线轨迹优化[2−3]. 这类方法在理想环境下表现出色, 能够保证机器人的运动稳定性, 并具有一定的抗扰动能力[4]. 随着环境复杂度增加, 准确建模变得愈发困难. 特别是在低摩擦地面上, 由于摩擦力的不确定性, 基于模型的控制方法难以实现预期的控制效果[5].

近年来, 强化学习 (Reinforcement learning, RL) 方法为解决上述问题提供了新的思路. 与传统方法不同, RL 通过智能体与环境的持续交互来学习最优控制策略, 无需对复杂环境进行精确建模[6]. 研究表明, 基于 RL 的控制器能够自主学习复杂的运动策略, 优化步态参数, 提高能量效率, 并增强对外部扰动的鲁棒性[7−8]. 在步态优化方面, RL 方法具有显著优势: 训练得到的步态具有更强的泛化能力[9]、更高的学习效率[10] 以及更自然的运动特征[11].

尽管 RL 方法在四足机器人控制领域取得了显著进展, 但在实际应用中仍面临重大挑战. 首先, 仿真环境与实际环境存在显著的差异, 虽然研究者提出了快速运动适应 (Rapid motor adaptation, RMA) 等方法[7], 但其在低摩擦环境下的适应能力仍然有限. 其次, 现有研究往往忽视质心位置对运动稳定性的重要影响, 这导致机器人在不同摩擦系数环境下的控制性能显著下降. 特别是在未知的低摩擦环境中, 现有控制器无法及时调整机体姿态和步态参数, 容易导致运动失稳[12].

针对上述挑战, 本文提出一种融合模型驱动和强化学习的四足机器人运动控制框架, 主要创新点如下:

1) 奖励函数设计: 提出单次落脚点偏差惩罚机制, 克服传统累积惩罚方法对瞬时滑动的过度敏感问题, 实现对连续滑动状态的有效评估;

2) 接触状态建模: 提出基于双曲正切函数的连续接触状态描述方法, 有效解决传统离散描述方法中的接触冲击问题, 提升运动过程的连续性和稳定性;

3) 地面特性感知: 设计基于 LSTM 的地面特性实时估计网络, 结合机体姿态动态调整策略, 实现机器人在不同粗糙程度地面的鲁棒运动;

4) 分层控制架构: 构建基于执行层和决策层的分层控制框架, 实现模型驱动与强化学习的有效结合, 提高系统的环境适应能力.

本文结构如下: 第 1 节介绍四足机器人步态设计与训练方法; 第 2 节详细阐述所提出的控制方法; 第 3 节通过仿真实验验证方法的有效性; 第 4 节总结全文并讨论未来研究方向.

1 四足机器人步态设计与训练

步态作为四足机器人运动的基础模式, 其设计直接影响机器人的运动稳定性、能量效率和环境适应能力. 本节将从步态规划的基本原理和控制方法的实现两个方面, 系统地阐述四足机器人运动控制的关键技术. 重点介绍基于相位变量的步态生成方法, 通过构建统一的相位描述框架, 有效解决传统方法中相位不连续、参数优化困难等问题, 为实现稳定可靠的运动控制提供理论支撑.

1.1 四足机器人的典型步态及其应用

四足机器人具有多种基本步态模式 (如图 1 所示), 每种步态都有其特定的运动特征和应用场景:

1) 蹦跳步态: 四足同步的垂直运动模式, 主要应用于障碍跨越和机体高度调整. Chai 等[13] 通过将机器人简化为质点–弹簧–阻尼系统的虚拟模型控制方法, 显著提升了机器人在不平整地形中的适应性能.

2) 小跑步态: 对角腿同步的周期运动模式, 是最常用的稳定运动方式. Raibert[14] 的开创性工作和 Jiang 等[15] 基于弹簧–负载倒立摆 (Spring loaded inverted pendulum, SLIP) 模型的优化方法, 显著

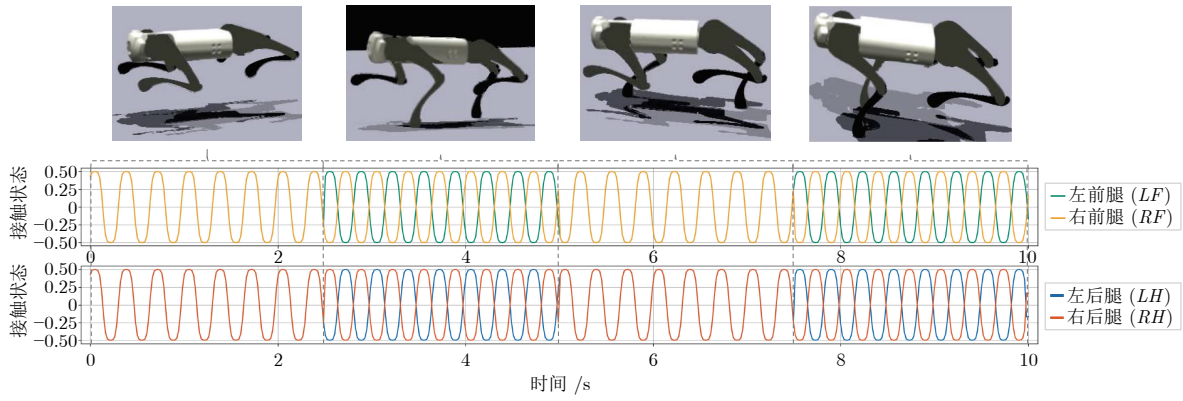


图 1 四种步态及其期望的接触状态图

Fig. 1 Diagram of the four gaits and the desired contact state for each gait

提高了小跑步态的稳定性和效率, 使其成为平坦地形移动的主要方案.

3) 侧步步态: 左右两侧腿交替的横向运动模式, 主要应用于横向移动和狭窄空间通过. Fukuoka 等[16] 通过模拟生物神经振荡器的中枢模式发生器 (Central pattern generators, CPG) 网络方法, 实现了机器人平稳的侧向运动控制.

4) 跃步步态: 前后腿对角同步的跃进运动模式, 主要应用于快速奔跑和高速移动. Adak 等[17] 通过优化足端接触力分布的规划方法, 有效提升了机器人在高速运动中的动态稳定性.

在实际应用中, 合理的步态设计对机器人的运动性能至关重要. 其中, 支撑相和摆动相的协调配合是保证运动稳定性的关键.

1.2　基于相位的步态训练方法

为实现稳定的步态控制, 本文采用基于 CPG 的相位变量和时间偏移量训练框架[18]. CPG 通过生成预期接触状态序列 (如图 1 所示), 结合统一的相位描述体系, 将步态控制问题转化为相位变量的优化问题, 简化控制策略的设计过程, 并提高运动的连续性.

1.2.1　步态建模与同步机制

步态建模的核心在于建立腿部运动与相位变量之间的映射关系. 四足机器人的每条腿都具有周期性运动特征, 可用相位变量 $\phi_i \in [0,\ 2\pi)$ 描述, 其中 $i \in \{LF,\ RF,\ LH,\ RH\}$ 分别表示左前腿、右前腿、左后腿和右后腿 (如图 2 所示). 运动周期包含两个主要阶段:

1) 摆动相 ($\phi_i \in [0,\ \phi_{\text{connect}})$): 完成腿部抬起和前向摆动;

2) 支撑相 ($\phi_i \in [\phi_{\text{connect}},\ 2\pi)$): 提供地面支撑力.

其中, ϕ_{connect} 表示从摆动相切换到支撑相的临界相位角. 为了使多条腿能够协调运动, 引入全局时间变量 t 来统一调节各条腿的运动相位. 每条腿的相位 ϕ_i 与全局时间的关系可表示为:

$$\phi_i = \frac{t + \delta_i}{2\pi} \tag{1}$$

其中, δ_i 表示第 i 条腿的相位偏移量, 用于调整各腿之间的相对运动时序. 全局时间变量 t 的更新遵循如下方程:

$$t_{\text{new}} = t_{\text{old}} + \frac{f^{\text{cmd}}}{f_\pi} \tag{2}$$

其中, t_{old} 表示当前时间步的全局时间, t_{new} 表示更新后的全局时间, f^{cmd} 表示目标运动频率, f_π 表示系统的更新频率. 相位偏移量 $\theta = \{\theta_1,\ \theta_2,\ \theta_3\}$, $\theta \in [0,\ 2\pi)$ 用于控制不同腿部之间的相位差, 从而实现特定的步态模式 (详见表 1). 具体地, 各腿的相位偏移量 δ_i 可由参数 θ 确定:

$$\delta_{LF} = 0,\ \delta_{RF} = \theta_1,\ \delta_{LH} = \theta_2,\ \delta_{RH} = \theta_3 \tag{3}$$

表 1　步态参数配置

Table 1　Configuration of gait parameters

步态类型	θ_1	θ_2	θ_3
蹦跳	0	0	0
小跑	π	π	0
侧步	π	0	π
跃步	0	π	π

1.2.2　接触状态建模与优化

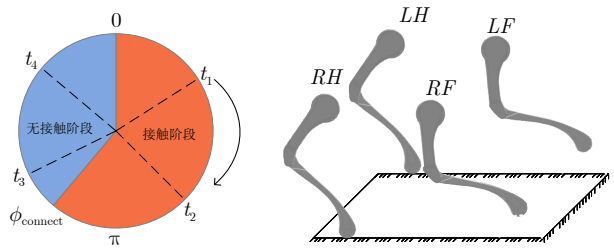
接触状态描述四足机器人各足端与地面的接触关系, 即足端是否与地面接触. 准确的接触状态建模是实现四足机器人稳定运动控制的关键环节. 为克服传统二值接触状态描述方法 (即仅用 0 和 1 表示无接触和完全接触) 在实际应用中的局限性, 本文提出一种基于双曲正切函数的连续状态描述方法. 首先, 构建基础平滑函数 $\Psi(x,\ \sigma)$:

$$\Psi(x,\ \sigma) = \frac{1}{2}\left(1 + \tanh\left(\frac{x}{\sigma}\right)\right) \tag{4}$$

其中, σ 为平滑度参数, 通过调整该参数可以控制状态过渡的平滑程度. 基于此函数, 构造单腿的连续接触状态描述:

$$C_{\text{foot}}(t) = \Psi(t,\ \sigma) \times (1 - \Psi(t - 0.5,\ \sigma)) + \Psi(t - 1,\ \sigma) \times (1 - \Psi(t - 1.5,\ \sigma)) \tag{5}$$

其中, $C_{\text{foot}}(t)$ 表示单腿在时间 t 时的连续接触状态值. 首先, 通过双曲正切函数的连续特性, 实现接触状态的平滑过渡, 避免传统二值方法的突变问题; 其次, 连续的状态反馈为后续的运动规划和控制优化提供可靠的数学基础. 如图 2(b) 所示, 该方法能



(a) 相位图　(b) 步态示意图

(a) Phase diagram　(b) Gait schematic diagram

图 2　相位变量与腿部运动的映射关系

Fig. 2　Mapping relationship between phase variables and leg motion

够准确描述机器人腿部与地面的动态交互过程, 为步态的稳定实现提供理论支撑. 该方法的有效性将在第 3.2.2 节中通过实验结果详细验证.

1.2.3 步态学习与优化策略

步态学习的核心是通过 RL 优化控制策略. 基于 Margolis 等[19] 的工作, 改进的奖励函数 (详见第 2.3.2 节) 包括运动跟踪、姿态稳定性以及运动约束三个部分的奖励. 在训练过程中, 采用 PPO[20] (Proximal policy optimization) 算法最大化累积奖励来优化控制策略, 并通过调整控制参数 c_t (包含前文定义的相位偏移量 θ 等步态参数) 实现运动精确控制.

2 四足机器人的分层控制框架

为有效应对不同粗糙程度地面带来的控制难题, 本文提出一种分层控制框架 (如图 3 所示). 该框架包含自适应调整的决策层和基于强化学习的执行层, 通过分层协同实现对环境变化的快速感知与响应.

2.1 分层控制框架的协同机制

为实现复杂环境下的稳定控制, 系统设计基于不同时间尺度的协同机制. 决策层以 10 Hz 频率运行, 保证环境感知的准确性; 执行层则以 100 Hz 频率运行, 确保控制的实时性. 这种时间尺度的差异使得系统能够同时兼顾环境适应性和控制稳定性.

该框架下, 决策层作为高层控制中枢, 进行环境感知和策略生成. 其输入包含机器人历史状态序列 $\{s_{t-49}, \cdots, s_t\}$ 以及运动指令 v^{cmd}. 通过 LSTM 网络和自适应控制, 决策层实现对地面特性的实时估计和控制参数生成:

$$\{\mu_{\text{est}}, d_{\text{slid}}, h_{\text{com}}^{\text{cmd}}\} = f_{\text{des}}(\{s_{t-49}, \cdots, s_t\}, v^{\text{cmd}}) \quad (6)$$

其中, f_{des} 是决策层的核心函数, 通过分析机器人过去 50 个时间步的状态数据和当前运动指令, 输出三个关键控制参数: μ_{est} 为地面摩擦系数估计值; d_{slid} 为预测滑动距离; $h_{\text{com}}^{\text{cmd}}$ 为质心目标高度.

基于决策层的输出, 执行层负责将包含高层指令的状态空间转化为实时的关节控制信号. 基础状态空间 s_t 定义如下:

$$s_t = [q_j, \dot{q}_j, \dot{q}_b, r_g, v^{\text{cmd}}, c_t] \quad (7)$$

其中, q_j, $\dot{q}_j$ 表示关节位置和速度; $\dot{q}_b$ 为机体角速度; r_g 为机体坐标系下的重力方向; c_t 为控制参数向量, 包含质心目标高度 $h_{\text{com}}^{\text{cmd}}$ 等控制量. 通过强化学习训练策略网络 π_θ, 执行层实时生成 12 个关节的目标位置信号 $a_t \in \mathbf{R}^{12}$:

$$a_t = \pi_\theta(s_t) \quad (8)$$

为实现稳定的关节控制, 采用 PD 控制器将位置指令转换为关节力矩.

2.2 决策层设计

决策层作为控制框架的核心, 负责感知环境变化并生成适应性策略. 本文设计基于质心自适应调整的控制策略, 通过实时调整机体质心高度来应对地面摩擦系数的变化, 提高机器人在不同粗糙程度地面的运动稳定性.

2.2.1 环境感知与估计

环境感知的核心任务是估计地面的摩擦特性和滑动距离. 相较于传统的基于接触力分析的方法, 本文提出一种基于历史状态序列的学习方法, 在不直接测量摩擦系数的情况下, 实现对地面特性的估计.

为实现对地面特性的估计, 本文构建一个包含

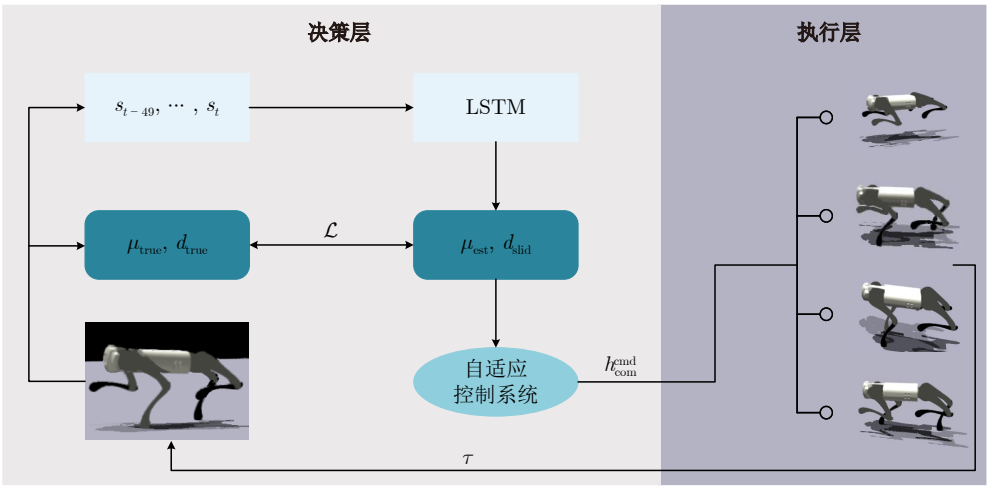


图 3 系统控制框架

Fig. 3 System control framework

机器人状态和环境交互信息的特征向量, 全面描述机器人与地面的交互过程:

$$s_t = [q_j,\ \dot{q}_j,\ f_{\text{contact}},\ p_{\text{foot}},\ \dot{p}_{\text{foot}}] \tag{9}$$

其中, f_{contact} 为足端接触力; p_{foot}, $\dot{p}_{\text{foot}}$ 分别代表足端位置和速度.

考虑到地面特性的时变性质, 本文设计一种基于 LSTM 的深度学习模型结构. 以连续 50 个时间步的状态序列 $\{s_{t-49}, \cdots, s_t\}$ 作为输入, 通过 128 个 LSTM 隐藏单元提取状态序列中的时序特征. 为增强关键时刻的状态信息, 引入自注意力机制, 并通过双层全连接网络 (64×64 维, ReLU 激活) 进行特征映射. 最终, 模型输出地面摩擦特性指标 μ_{est} 和预期滑动距离 d_{slid}.

模型的训练在摩擦系数 $\mu = 1.0$ 的地面环境中进行, 损失函数 $\mathcal{L}$ 如下:

$$\mathcal{L} = \alpha\|\mu_{\text{est}} - \mu_{\text{true}}\|^2 + \beta\|d_{\text{slid}} - d_{\text{true}}\|^2 \tag{10}$$

其中, μ_{true} 为环境中的真实摩擦系数, d_{true} 为实际测量的滑动距离. $\alpha = 0.2$, $\beta = 0.8$ 为损失权重系数, 用于平衡摩擦系数估计和滑动预测的精度. 同时, 为提高模型的泛化性能, 训练过程中引入高斯噪声扰动.

2.2.2　基于生物启发的姿态调整策略

基于对四足动物在低摩擦地面上运动特性的研究[21−22] 表明, 动物主要通过调整质心高度和躯干姿态来增强运动稳定性. 受此启发, 本文提出一种基于质心自适应调整的控制策略, 旨在提高机器人在不同粗糙程度地面的稳定性.

当四足机器人在低摩擦地面上运动时, 降低质心能够带来多方面的力学优势. 首先, 较低的质心位置可以减小倾覆力矩, 提升整体稳定性; 其次, 质心下降过程中腿部关节角度的变化会使足端与地面的接触面积增大, 从而提高可用摩擦力; 最后, 这种姿态调整还能优化关节力矩分布, 通过增加近端关节 (髋关节) 的负载来减轻远端关节 (膝关节和踝关节) 的压力, 这种力矩重分配特性与生物系统的运动模式高度一致.

基于环境感知提供的地面特性估计, 设计质心高度自适应调整算法:

$$h_{\text{com}}^{des} = \text{clip}(h_{\text{base}} - \Delta h_{\text{adapt}},\ 0,\ h_{\max}) \tag{11}$$

其中, h_{com}^{des} 为期望质心高度, h_{base} 为当前质心高度, Δh_{adapt} 为自适应调整量, $h_{\max}$ 为最大调整幅度, $\text{clip}(x,\ a,\ b)$ 函数将输入值 x 限制在区间 $[a,\ b]$ 内. 自适应调整量的计算考虑了滑动状态:

$$\Delta h_{\text{adapt}} = \frac{|d_{\text{slid}} - d_{\text{avg}}^{\text{slid}}|}{d_{\text{avg}}^{\text{slid}}} \times c \tag{12}$$

其中, $d_{\text{avg}}^{\text{slid}}$ 为历史平均滑动距离 (50 个时间步), c 为调整增益系数. 为确保运动的平滑性, 质心高度调整采用平滑过渡机制:

$$h_{\text{com}}^{\text{cmd}} = h_{\text{com}}^{prev} + \text{clip}(h_{\text{com}}^{des} - h_{\text{com}}^{prev},\ -\dot{h}_{\max}\Delta t,\ \dot{h}_{\max}\Delta t) \tag{13}$$

其中, h_{com}^{prev} 为上一时刻的质心高度, $\dot{h}_{\max}$ 为最大调整速率, Δt 为控制周期.

2.2.3　基础步态分析

基础步态的选择对自适应控制策略性能具有重要影响. 通过理论分析和初步验证, 本文采用小跑步态作为基础步态.

从力学特性来看, 小跑步态采用对角支撑模式, 在保证动态稳定性的同时, 提供最优的质心支撑, 确保质心始终位于支撑多边形内. 同时, 对角支撑模式下的地面反作用力通过两条对角线均匀分布, 降低单腿承受的最大力矩, 有助于降低滑动风险.

从控制特性来看, 小跑步态具有良好的相位连续性, 支撑相与摆动相之间过渡平稳. 不仅降低了基础运动控制的难度, 同时有利于自适应控制策略的实施. 对角支撑模式提供的稳定支撑条件, 能够确保调整过程的平稳性和可控性.

在第 3.2.2 节, 对小跑步态作为基础运动模式的有效性进行实验验证.

2.3　执行层设计

执行层的主要功能是将包含高层指令的状态空间转化为实时的关节控制信号. 机器人运动控制具有连续性和状态转移的不确定性特征, 本文将执行层的控制问题建模为马尔科夫决策过程[23] (Markov decision process, MDP), 并采用 RL 方法训练.

2.3.1　状态空间与动作映射

参考 Margolis 等[19] 的工作, 为满足自适应控制的需求, 状态空间 $s_t \in \mathbf{R}^{37}$ 设计如下 (详细参数及取值范围见表 2):

1) 机器人状态: 包括关节位置 $q_j \in \mathbf{R}^{12}$、关节速度 $\dot{q}_j \in \mathbf{R}^{12}$、机体角速度 $\dot{q}_b \in \mathbf{R}^3$ 以及机体坐标系下的重力方向 $r_g \in \mathbf{R}^3$.

2) 运动指令 $v^{\text{cmd}} \in \mathbf{R}^3$: 包含机体坐标系下的期望前向速度 v_x^{cmd}、侧向速度 v_y^{cmd} 和偏航角速度 ω_z^{cmd}.

3) 控制参数 $c_t \in \mathbf{R}^4$: 包括相位偏移量 θ、步态频率 f^{cmd}、质心目标高度 $h_{\text{com}}^{\text{cmd}}$ 和足端摆动高度 $h_{\text{foot}}^{\text{cmd}}$.

通过 PD 控制器将执行层输出 12 个关节的目

表 2 控制命令参数范围

Table 2 Range of control command parameters

参数	最小值	最大值	单位
v_x^{cmd}	−3.00	3.00	m/s
v_y^{cmd}	−1.00	1.00	m/s
ω_z^{cmd}	−1.00	1.00	rad/s
f^{cmd}	1.50	4.00	Hz
$h_{\text{com}}^{\text{cmd}}$	−0.45	0.10	m
$h_{\text{foot}}^{\text{cmd}}$	0.03	0.30	m

标位置信号 $a_t \in \mathbf{R}^{12}$ 转换为关节力矩:

$$\tau = k_p(a_t - q_{\text{current}}) + k_d(\dot{a}_t - \dot{q}_{\text{current}}) \tag{14}$$

其中, q_{current} 和 $\dot{q}_{\text{current}}$ 分别表示当前关节位置和速度; $k_p = 20$ N·m/rad, $k_d = 0.5$ N·m/rad 为比例和微分增益.

2.3.2 基于单次落足点偏差的奖励函数设计

在考虑滑动作为奖励时, 传统的方法通常对每个时间步的打滑行为进行惩罚, 不仅会导致策略过度保守, 影响运动效率, 并且难以有效区分瞬时打滑和持续性打滑, 影响学习效果.

为对运动过程中滑动的有效评估, 本文设计一种基于单次接触位置偏差的瞬时惩罚机制. 仅在足端与地面发生初始接触的瞬时时刻对落足点偏差进行离散化评估, 其数学表达为:

$$r_{\text{contact}} = \|p_{\text{foot}}^{c} - p_{\text{foot}}^{c,\,\text{cmd}}\| \times \mathbb{I}(t = t_{\text{contact}}) \tag{15}$$

其中, r_{contact} 表示单次落足点偏差的奖励值; p_{foot}^{c} 表示足端实际接触位置; $p_{\text{foot}}^{c,\,\text{cmd}}$ 为足端目标接触位置; t_{contact} 为足端与地面发生接触的时刻; $\mathbb{I}(\cdot)$ 为离散指示函数, 当且仅当 $t = t_{\text{contact}}$ 时输出 1, 其余时刻为 0. 该奖励函数具有以下特点: 1) 瞬时单次惩罚: 仅在足地接触初始时刻 t_{contact}, 避免传统方法对支撑相全过程连续惩罚导致的保守性约束; 2) 滑移误差量化: 采用接触点位置矢量的欧氏距离 $\|\cdot\|$ 直接表征滑移量, 建立滑动状态与运动精度的显式映射关系.

如表 3 所示 (表 3 中"公式"列所涉及的变量的含义见表 A1), 本文构建的复合奖励函数包含以下三组核心分量:

1) 运动跟踪奖励: 包括机体运动跟踪 (水平速度 $r_{v_{x,\,y}^{cmd}}$、垂直轴角速度 $r_{\omega_z^{cmd}}$)、姿态跟踪 (质心高度 $r_{h_z^{cmd}}$、俯仰角 $(r_{\rho^{cmd}})$)、腿部运动跟踪 (摆动相力 $r_{f_c^{cmd}}$、支撑相速度 $r_{v_f^{cmd}}$) 以及单次落足点偏差 r_{contact};

2) 姿态稳定性奖励: 包括垂直速度约束 v_z^2、横滚俯仰角速度约束 $\|\omega_{x,\,y}\|^2$ 以及足端滑动约束 $\|v_{\text{foot},\,x,\,y}\|^2$;

表 3 奖励函数

Table 3 Reward functions

类别	项目	公式	权重
运动跟踪	水平面速度跟踪 $r_{v_{x,\,y}^{cmd}}$	$\exp\left(-\frac{\|v_{x,\,y} - v_{x,\,y}^{cmd}\|^2}{\sigma_{v_{x,\,y}}}\right)$	0.02
	垂直轴角速度跟踪 $r_{\omega_z^{cmd}}$	$\exp\left(-\frac{\|\omega_z - \omega_z^{cmd}\|^2}{\sigma_{\omega_z}}\right)$	0.01
	质心高度跟踪 $r_{h_z^{cmd}}$	$\left(h_z - h_z^{cmd}\right)^2$	−0.08
	躯干俯仰角跟踪 $r_{\rho^{cmd}}$	$\left(\rho - \rho^{cmd}\right)^2$	−0.10
	摆动相力跟踪 $r_{f_c^{cmd}}$	$\sum_{\text{foot}}\left(1 - C_{\text{foot}}^{\text{cmd}}(t)\right) \times \exp\left(-\frac{\|f_{\text{foot}}^{\text{cmd}}\|^2}{\sigma_{cf}}\right)$	−0.08
	支撑相速度跟踪 $r_{v_f^{cmd}}$	$\sum_{\text{foot}} C_{\text{foot}}^{\text{cmd}}(t) \times \exp\left(-\frac{\|v_{\text{foot}}^{\text{cmd}}\|^2}{\sigma_{cv}}\right)$	−0.08
	单次落足点偏差 r_{contact}	$\|p_{\text{foot}}^{c} - p_{\text{foot}}^{c,\,\text{cmd}}\| \times \mathbb{I}(t = t_{\text{contact}})$	−0.10
姿态稳定性	垂直方向速度	v_z^2	-4×10^{-4}
	横滚和俯仰角速度	$\|\omega_{x,\,y}\|^2$	-2×10^{-5}
	足端滑动	$\|v_{\text{foot},\,x,\,y}\|^2$	-8×10^{-4}
运动约束	大腿/小腿碰撞	$1_{\text{collision}}$	−0.02
	关节限位违反	$1_{q_i > q_{\max} \,\|\|\, q_i < q_{\min}}$	−0.20
	关节力矩	$\|\tau\|^2$	-9×10^{-3}
	关节角速度	$\|\dot{q}\|^2$	-9×10^{-3}
	关节角加速度	$\|\ddot{q}\|^2$	-5×10^{-9}
	动作平滑度 (一阶)	$\|a_{t-1} - a_t\|^2$	-2×10^{-3}
	动作平滑度 (二阶)	$\|a_{t-2} - 2a_{t-1} + a_t\|^2$	-2×10^{-3}

3) 运动约束奖励: 包括安全约束 (碰撞检测 $1_{\text{collision}}$、关节限位 $1_{q_i>q_{\max}||q_i<q_{\min}}$)、能量约束 (关节力矩 $\|\tau\|^2$、角速度 $\|\dot{q}\|^2$、角加速度 $\|\ddot{q}\|^2$) 以及平滑约束 (一阶 $\|a_{t-1}-a_t|^2$ 和二阶 $\|a_{t-2}-2a_{t-1}+a_t\|^2$ 动作差分).

为协同优化运动性能与约束条件, 本文采用基于指数耦合的复合奖励计算架构:

$$r_{\text{total}} = r_{\text{task}} \times \exp(r_{\text{aux}}) \tag{16}$$

其中, r_{total} 为总奖励; r_{task} 为任务导向奖励项, 包含运动跟踪与姿态稳定性相关正权重项; r_{aux} 为辅助约束惩罚项, 集成运动约束相关负权重项.

通过指数变换实现两阶段优化: $\exp(r_{\text{aux}}) \in \mathbf{R}^+$ 确保总奖励 r_{total} 恒为正, 避免传统线性叠加中负奖励导致策略过早收敛; 指数放大效应对约束违反量 $\|r_{\text{aux}}\|$ 施加超线性惩罚, 增强策略对物理限制的敏感性; 允许任务奖励与约束惩罚独立调节, 通过乘积形式实现目标间自适应权衡.

2.3.3　训练策略优化

为提高训练的稳定性和策略的泛化能力, 执行层在高摩擦环境 ($\mu = 1.0$) 中完成基础运动策略训练, 并且采用教师–学生策略[24] (伪代码见算法 1) 与 PPO 算法 (伪代码见算法 2, 超参数设计见表 4) 相结合的训练方法. 随后, 通过选取 $\mu \in \{0.05, 0.2, 1.0\}$ 这三个典型值对决策层进行验证实验.

算法 1. 教师–学生网络训练算法

输入. 教师网络 π_{teacher}, 学生网络 π_{student}

输出. 优化后的学生网络 π^*_{student}

for 每个训练轮次 **do**
  收集完整状态轨迹 $\{s_t^{\text{full}}, a_t, r_t\}$
  更新 π_{teacher} 以最大化 $J(\theta)$
end for
for 每个训练轮次 **do**
  采样基础状态 s_t^{base}
  获取教师策略 $\pi^*_{\text{teacher}}(a_t|s_t^{\text{full}})$
  更新 π_{student} 以最小化 $\mathcal{L}_{\text{KD}}$
end for
return π^*_{student}

算法 2. PPO 算法

输入. 初始策略参数 θ_0, 值函数参数 ϕ_0

输出. 优化后的策略参数 θ^*

for 每个训练轮次 **do**
  使用当前策略 $\pi_{\theta_{\text{old}}}$ 收集轨迹 $\{s_t, a_t, r_t\}$
  计算时序差分误差 $\delta_t = r_t + \gamma V(s_{t+1}) - V(s_t)$
  计算广义优势估计 $\hat{A}_t = \sum_{l=0}^{T-t-1}(\gamma\lambda)^l\delta_{t+l}$
  for 每个优化轮次 **do**
    计算概率比 $\rho_t(\theta) = \frac{\pi_\theta(a_t|s_t)}{\pi_{\theta_{\text{old}}}(a_t|s_t)}$
    计算裁剪目标 $L^{\text{CLIP}}(\theta)$
    计算值函数损失 $L^{\text{VF}}(\theta)$
    计算策略熵 $S[\pi_\theta]$
    更新参数 θ 以最大化 $L^{\text{CLIP}}(\theta) - c_1L^{\text{VF}}(\theta) + c_2S[\pi_\theta]$
  end for
  $\theta_{\text{old}} \leftarrow \theta$
end for
return θ^*

表 4　PPO 超参数
Table 4　PPO hyperparameters

参数	值
批量大小	4 096 × 24
小批量大小	4 096 × 6
迭代次数	5
裁剪范围	0.20
熵系数	0.01
折扣因子	0.99
广义优势估计折扣因子	0.95
目标 KL 散度	0.01
学习率	自适应

教师–学生策略通过知识蒸馏框架实现从特权信息空间到观测空间的跨模态策略迁移, 具体实施包含两个阶段.

教师网络 π_{teacher} 获取包含特权信息的完整状态空间:

$$s_t^{\text{full}} = \left[s_t^{\text{base}},\ s_t^{\text{priv}}\right] \tag{17}$$

其中, s_t^{full} 为完整状态空间; s_t^{base} 为基础状态空间 (与上文状态空间一致); s_t^{priv} 为特权信息, 包含地面摩擦系数、足端接触力矢量 f_{foot} 及质心位置.

教师网络的训练目标是通过 PPO 算法最大化策略 π_θ 的期望折扣奖励. 期望折扣奖励定义为:

$$J(\theta) = \mathrm{E}_{\xi\sim\pi_\theta}\left[\sum_{k=0}^{\infty}\gamma^k r_{t+k}\right] \tag{18}$$

其中, $J(\theta)$ 表示期望的累积折扣奖励, r_{t+k} 是在时间步 $t+k$ 时的即时奖励, γ 是折扣因子 ($0 \le \gamma \le 1$), $\xi = (s_t, a_t, r_t, s_{t+1}, \cdots)$ 表示在策略 π_θ 下采样得到的轨迹. 通过优化参数 θ 以最大化 $J(\theta)$, 最终收敛得到面向完整状态空间的最优教师策略:

$$\pi^*_{\text{teacher}}(a_t|s_t^{\text{full}}) = \arg\max_{\pi}\mathrm{E}_{\xi\sim\pi}\left[J(\theta)\right] \tag{19}$$

学生网络 π_{student} 仅接收部分状态空间 s_t^{base}, 通过最小化策略分布 KL 散度实现知识迁移:

$$\mathcal{L}_{\text{KD}} = \mathrm{E}_{s_t^{\text{base}}}\left[D_{\text{KL}}\left(\pi^*_{\text{teacher}}(\cdot|s_t^{\text{full}})\|\pi_{\text{student}}(\cdot|s_t^{\text{base}})\right)\right] \tag{20}$$

其中, $\mathcal{L}_{\text{KD}}$ 为知识蒸馏损失; $D_{\text{KL}}(\cdot\|\cdot)$ 表示计算两个概率分布之间的 KL 散度. 该目标使学生策略在缺失 μ、f_{foot} 等不可测信息时, 仍能隐式学习教师网络对地面物理特性的响应模式.

PPO 算法通过交替优化策略网络和值函数网络, 实现策略的稳定且高效的学习. 其中即时奖励 r_t 与值函数 $V(s_t)$ 之差反映了策略评估中的偏差. 值函数 $V(s_t)$ 表示在状态 s_t 下策略预期获得的回报.

在每个训练轮次中, 首先利用旧策略 $\pi_{\theta_{\text{old}}}$ 与环境交互以采集一批轨迹数据. 基于采集的轨迹, 采用广义优势估计 (Generalized advantage estimation, GAE) 来计算优势函数 $\hat{A}_t$, 其形式为:

$$\hat{A}_t = \sum_{l=0}^{T-t-1}(\gamma\lambda)^l\delta_{t+l} \tag{21}$$

其中, T 表示一个完整的交互序列的终止时刻, 折扣因子 $\gamma\in[0,\,1]$ 用以衡量未来奖励的重要性, $\lambda\in[0,\,1]$ 控制估计的偏差与方差之间的权衡. 时序差分误差 δ_t 定义为:

$$\delta_t = r_t + \gamma V(s_{t+1}) - V(s_t) \tag{22}$$

在策略更新过程中, PPO 引入了对概率比的裁剪机制. 具体而言, 定义概率比为

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

其中, θ_{old} 为策略 π 的旧参数, θ 为新参数. 为了避免策略更新过大导致的训练不稳定, 算法设计了裁剪目标函数 $L^{\text{CLIP}}(\theta)$, 定义为:

$$L^{\text{CLIP}}(\theta) = \hat{\mathrm{E}}_t[\min(\rho_t(\theta)\hat{A}_t,\ \ \text{clip}(\rho_t(\theta),\ 1-\epsilon,\ 1+\epsilon)\hat{A}_t)] \tag{24}$$

其中, ϵ 为超参数. 另一方面, 值函数网络通过最小化均方误差损失进行优化, 值函数损失 $L^{\text{VF}}(\theta)$ 为:

$$L^{\text{VF}}(\theta) = \hat{\mathrm{E}}_t\left[\left(V(s_t) - V_t^{\text{target}}\right)^2\right] \tag{25}$$

其中, V_t^{target} 通常由实际回报或多步时序差分方法计算得到, 作为监督信号引导值函数的收敛.

算法的总损失函数 $L^{\text{total}}(\theta)$ 由策略损失、值函数损失和熵正则化项组成:

$$L^{\text{total}}(\theta) = L^{\text{CLIP}}(\theta) - c_1L^{\text{VF}}(\theta) + c_2S[\pi_\theta] \tag{26}$$

其中, 熵项 S 用于促进策略探索, 防止过早收敛到局部最优解, c_1 和 c_2 为权重系数. 每轮优化后, 更新目标策略参数 $\theta_{\text{old}}\leftarrow\theta$, 开始下一轮训练.

本文提出融合教师–学生策略与 PPO 算法的训练框架: 教师网络通过特权信息加速策略收敛; PPO 的裁剪机制保障训练稳定性; 学生网络基于实际观测数据决策, 提升实际部署能力. 如图 4 奖励曲线所示, 训练过程呈现典型三阶段演化特征: 初期探索阶段总奖励快速上升, 中期优化阶段子奖励

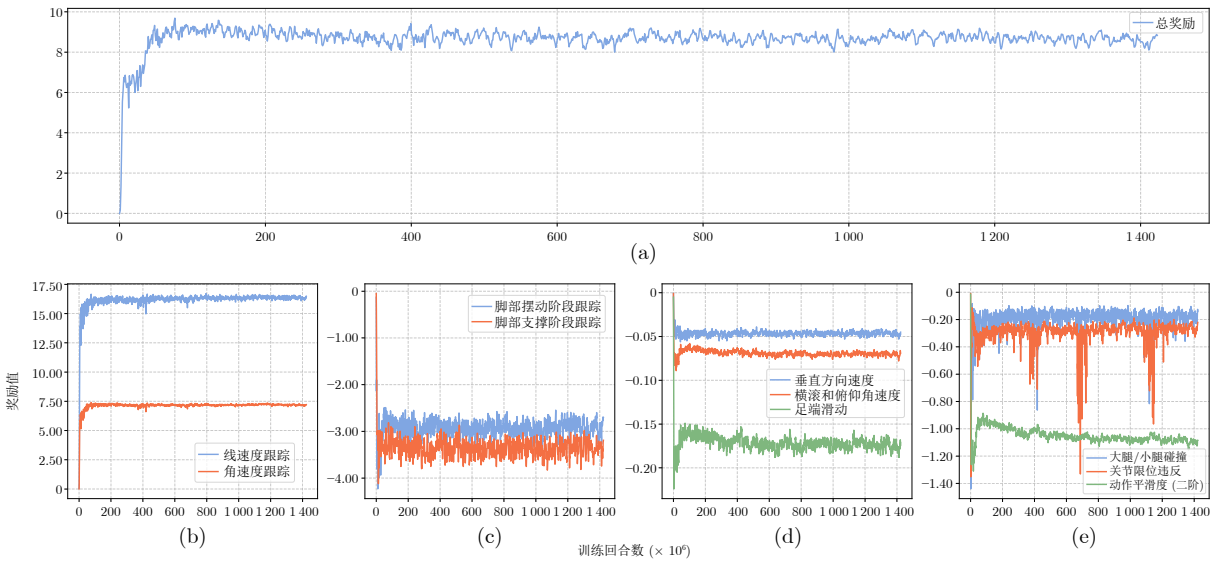


图 4 PPO 训练过程中部分奖励的变化趋势 ((a) 总奖励; (b) 运动学跟踪奖励; (c) 步态执行奖励; (d) 运动稳定性奖励; (e) 运动约束奖励)

Fig. 4 Trends in partial rewards during PPO training ((a) Total reward; (b) Kinematic tracking rewards; (c) Gait execution rewards; (d) Motion stability rewards; (e) Motion constraint rewards)

动态平衡, 后期稳定收敛.

3　实验验证与分析

3.1　实验环境与基础设置

本研究在 Isaac Gym[25] 物理仿真环境中开展实验, 该环境具有高度并行的仿真能力和精确的物理建模特性, 能够有效模拟四足机器人与地面的复杂交互过程. 实验设计一系列不同摩擦系数 ($\mu \in \{0.05, 0.2, 1.0\}$) 的地面环境, 并在低速 (1.5 m/s) 和高速 (2.0 m/s) 两种典型运动状态下进行测试, 以验证控制策略在不同动态条件下的适应性.

为进行系统性的性能评估, 本研究构建两种控制器: 自适应控制器和基准控制器. 自适应控制器采用本文提出的分层控制架构, 包含决策层和执行层两个主要组成部分, 能够根据地面条件动态调整机器人的质心高度, 从而实现更好的运动稳定性; 基准控制器采用文献 [19] 中的方法, 仅保留执行层结构, 其观测空间、动作空间、训练算法以及奖励函数设计与自适应控制器的执行层完全一致, 通过人为设定固定的质心高度进行控制.

3.2　基础性能验证实验

3.2.1　接触状态建模对比

为验证基于双曲正切函数的连续状态描述方法的有效性, 本研究构建基于基准控制器的接触状态建模对比实验体系. 实验设定摩擦系数 $\mu = 0.05$、质心高度 0 m 的基础参数, 选取小跑、蹦跳、跃步、侧步四种典型步态作为测试对象, 分别在 1.5 m/s (低速) 和 2.0 m/s (高速) 下开展对比测试. 通过计算足端接触序列的步态相似度指标, 定量评估实际足端接触序列与理想接触序列之间的一致性.

图 5 展示四种步态在不同速度条件下的足端接触序列差异. 图中采用颜色块机制表征足端触地状态: 实心色块表示触地阶段, 空白区域对应离地阶段. 每个子图均包含两方法对比框架, 其中深色区域代表本文提出的双曲正切连续描述方法, 浅色区域代表传统二值离散描述方法.

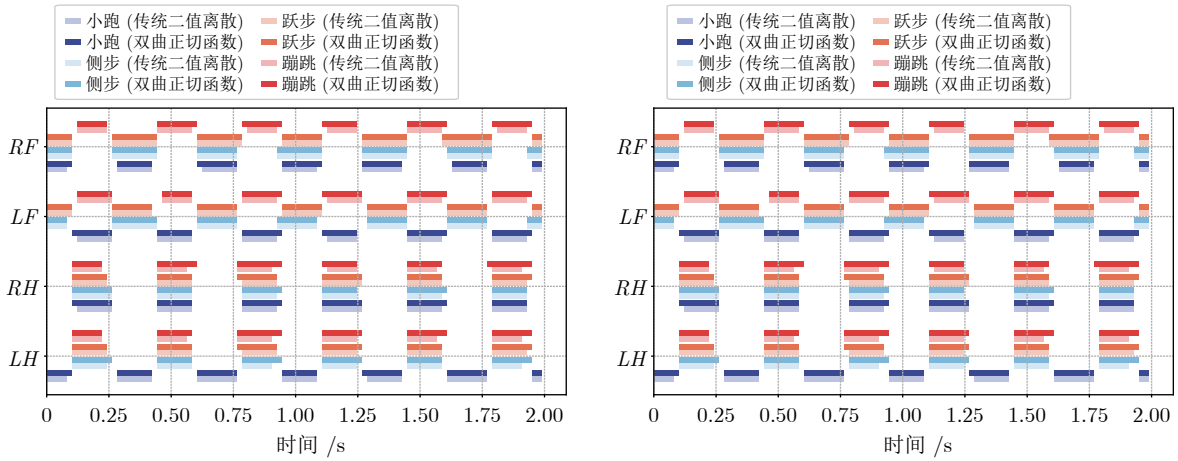
通过量化对比两种方法生成的足端接触序列与理想序列的步态相似度指标, 如表 5 所示. 在低速条件下, 该方法使四种步态的相似度分别提升: 小跑 (3.96%)、蹦跳 (8.93%)、跃步 (0.92%) 和侧步 (1.57%). 即便在控制难度更高的高速条件下, 性能提升依然保持在 3.72%、8.33%、2.23% 和 1.29% 的水平. 值得注意的是, 在蹦跳步态中, 该方法将步态相似度从 0.869 7 提升至 0.947 4, 这一显著改善主要归功于连续状态描述方法在处理同步接触转换时的独特优势.

3.2.2　基础步态性能对比分析

基于第 1.2.2 节的分析基础, 为验证方法在其他步态下的有效性, 本节在低摩擦环境 ($\mu = 0.05$)、质心高度设定为 0 m、目标速度为 1.5 m/s 的条件下, 对四种基础步态的运动特性进行对比研究, 分析基准控制器和自适应控制策略在四种步态下的表现. 实验结果如表 6 所示.

表 6 的数据显示, 相比基准控制器, 自适应控制策略在所有步态下均实现性能提升, 具体表现为:

1) 小跑步态: 通过降低质心高度 0.061 m, 足



(a) 低速条件下的足端接触序列 (1.5 m/s)

(a) Foot contact sequences under low-speed condition (1.5 m/s)

(b) 高速条件下的足端接触序列 (2.0 m/s)

(b) Foot contact sequences under high-speed condition (2.0 m/s)

图 5　基准控制器下不同建模方法的足端接触序列对比

Fig. 5　Comparison of foot contact sequences under different modeling methods with a benchmark controller

表 5 基准控制器下不同步态的接触状态建模方法对比
Table 5 Comparison of contact state modeling methods under different gaits with a benchmark controller

步态	1.5 m/s		2.0 m/s	
	二值离散	双曲正切	二值离散	双曲正切
小跑	0.9217	**0.9583**	0.8915	**0.9247**
蹦跳	0.8697	**0.9474**	0.8642	**0.9362**
跃步	0.9371	**0.9457**	0.9230	**0.9436**
侧步	0.9335	**0.9482**	0.9257	**0.9376**

端滑动距离从 0.343 cm 减少至 0.308 cm, 且保持了接近目标的运动速度 (1.428 m/s).

2) 蹦跳步态: 尽管需要较大的质心高度调整 (0.271 m), 但显著改善了运动性能, 速度从 0.898 m/s 提升至 1.116 m/s, 足端滑动距离减少了 32.5%.

3) 跃步步态: 通过适度降低质心高度 (0.119 m), 速度提升了 12.6% (从 1.219 m/s 到 1.372 m/s), 同时略微减少了足端滑动.

4) 侧步步态: 虽然需要最大的质心高度调整 (0.301 m), 但速度提升显著 (从 0.467 m/s 到 0.683 m/s), 足端滑动减少了 17.4%.

表 6 低摩擦环境下四种步态的性能对比
Table 6 Performance comparison of four gaits in a low friction environment

自适应控制	步态	质心高度变化 (m)	足端滑动距离 (cm)	运动速度 (m/s)
是	小跑	−0.061	0.308 ± 0.015	1.428
否	小跑	0	0.343 ± 0.081	1.455
是	跃步	−0.119	0.342 ± 0.032	1.372
否	跃步	0	0.354 ± 0.040	1.219
是	蹦跳	−0.271	0.423 ± 0.051	1.116
否	蹦跳	0	0.627 ± 0.046	0.898
是	侧步	−0.301	0.450 ± 0.004	0.683
否	侧步	0	0.545 ± 0.062	0.467

如图 6 所示, 小跑步态在保持最小质心高度调整的同时, 实现了较高的运动速度和最小的足端滑动距离. 这一结果表明, 小跑步态在低摩擦环境下具有独特的运动优势. 其优越性主要源于对角支撑配置所带来的固有稳定性.

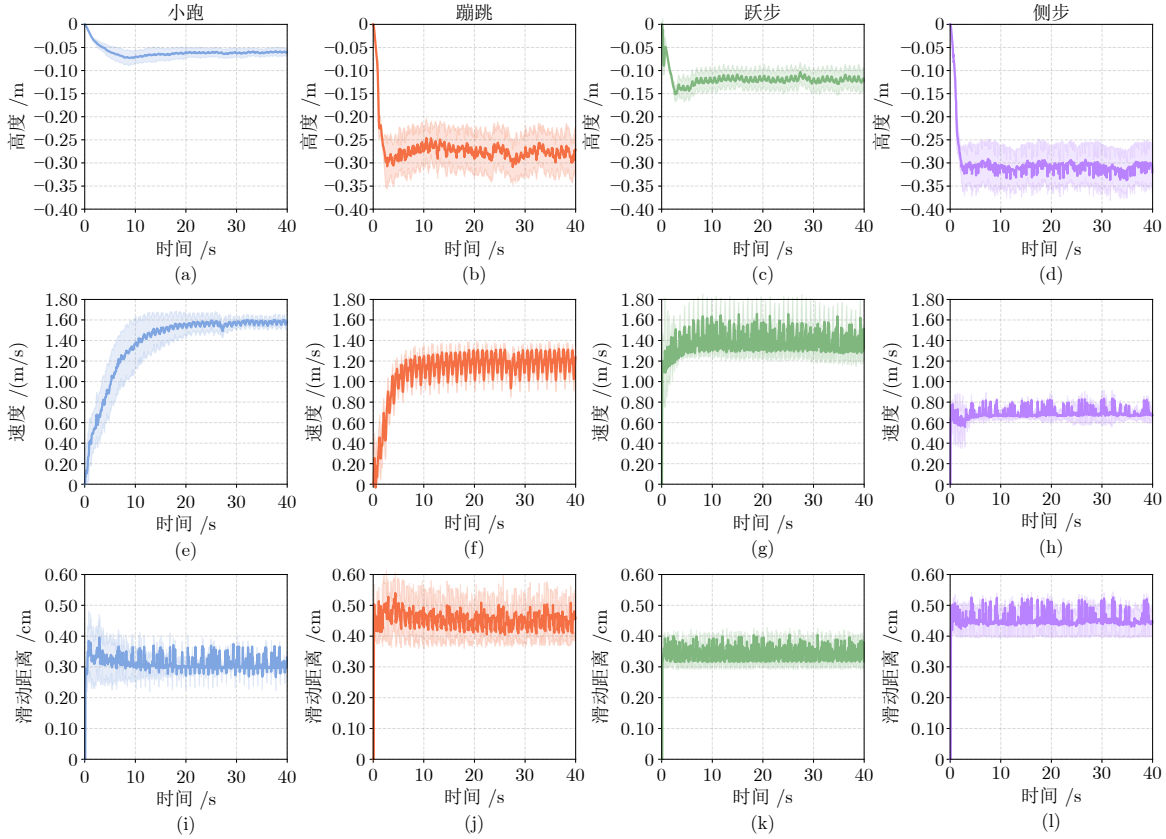


图 6 低摩擦环境下不同步态的运动特征对比 ((a) ~ (d) 为质心高度; (e) ~ (h) 为运动速度; (i) ~ (l) 为足端滑动距离. 从左至右分别为四种步态, 目标速度为 1.5 m/s, 摩擦系数 $\mu = 0.05$)

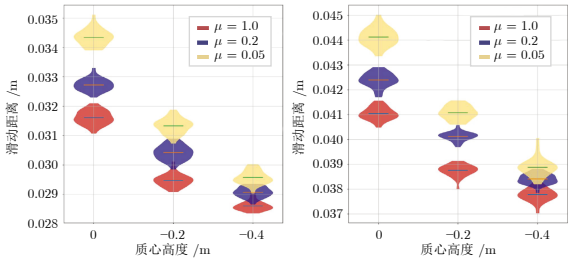
Fig. 6 Motion characteristics comparison of different gaits in a low friction environments ((a) ~ (d) is the centroid of mass height; (e) ~ (h) is the motion speed; (i) ~ (l) is the foot end sliding distance. From left to right, the four gaits have a target speed of 1.5 m/s and a friction coefficient $\mu = 0.05$)

通过与基准控制器的对比分析可以发现, 自适应控制策略通过合理调整质心高度, 有效提升了各步态在低摩擦环境下的运动性能. 其中, 小跑步态因其所需调整幅度最小且性能提升显著, 被确定为平坦地形下应对不同摩擦力系数地面最适合的基础步态. 这一结论与第 1.2.2 节的分析结果相互印证, 为后续控制策略的深入研究奠定基础.

3.3　自适应控制策略性能分析

本节将进一步系统分析自适应控制策略的性能特征. 首先, 对小跑步态在基准控制器下的实验进行评估, 实验涉及不同摩擦系数和期望速度条件下的质心高度与足端滑动, 如图 7 所示. 实验结果表明, 随着质心高度的逐步降低, 在各种摩擦系数条件下, 足端滑动距离均呈现出减小的趋势. 然而, 这种稳定性的提升通常是以牺牲运动速度为代价的.

为探究自适应控制器能否在保持较小滑动的同时维持较高的运动速度, 本研究设计摩擦系数 $\mu \in \{0.05, 0.2, 1.0\}$, 并针对低速 (1.5 m/s) 和高速 (2.0 m/s) 两种典型运动状态进行系统性测试.



(a) 速度指令为 1.5 m/s　(b) 速度指令为 2.0 m/s

(a) Speed command is 1.5 m/s　(b) Speed command is 2.0 m/s

图 7　基准控制器下不同摩擦系数和期望速度条件的质心高度与足端滑动分布图

Fig. 7　Distribution diagram of the centroid of mass height and foot end sliding under different friction coefficients and desired speeds with a benchmark controller

3.3.1　不同摩擦系数下的性能分析

为评估所提出控制策略在不同摩擦系数下的自适应性能, 在目标速度为 1.5 m/s、三种典型摩擦条件 ($\mu \in \{0.05, 0.2, 1.0\}$) 下进行对比实验. 图 8 和表 7 展示了自适应控制与基准控制器在质心高度变

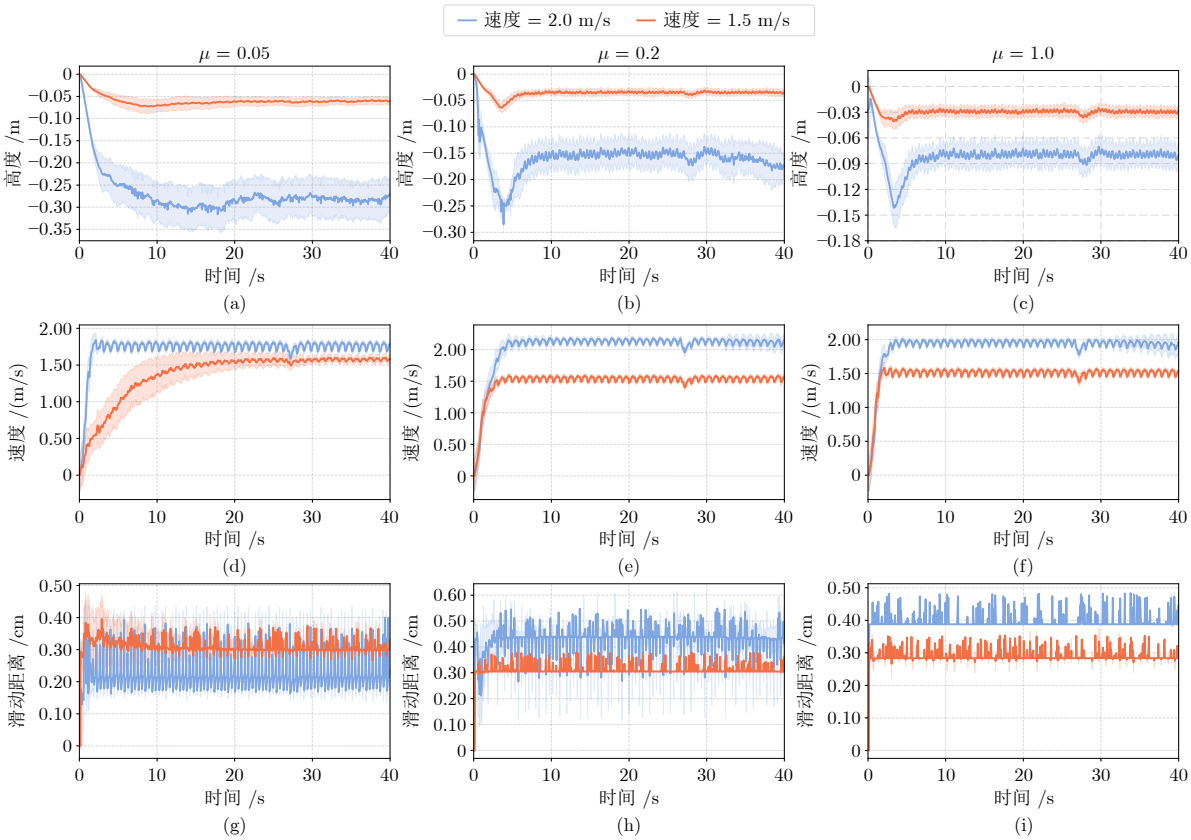


图 8　不同摩擦系数下机器人运动特性对比 ((a) ~ (c) 为质心高度; (d) ~ (f) 为运动速度; (g) ~ (i) 为足端滑动. 阴影区域表示标准差)

Fig. 8　Robot motion characteristics comparison under different friction coefficients ((a) ~ (c) is the centroid of mass height; (d) ~ (f) is the motion speed; (g) ~ (i) is the foot end sliding. Shaded areas indicate standard deviations)

表 7 不同期望速度和摩擦条件下小跑步态的控制性能对比

Table 7 Comparison of control performance under different desired speeds and friction conditions in the trotting gait

自适应控制	期望速度 (m/s)	摩擦系数	质心高度变化 (m)	足端滑动距离 (cm)	运动速度 (m/s)
是	1.5	1.0	−0.031 1	0.290 ± 0.000 4	1.490 0
否	1.5	1.0	0	0.316 ± 0.050 0	1.503 9
否	1.5	1.0	−0.200 0	0.295 ± 0.060 0	1.397 3
否	1.5	1.0	−0.400 0	0.286 ± 0.040 0	1.353 9
是	1.5	0.2	−0.038 0	0.312 ± 0.004 0	1.502 0
否	1.5	0.2	0	0.327 ± 0.060 0	1.548 1
否	1.5	0.2	−0.200 0	0.304 ± 0.050 0	1.436 7
否	1.5	0.2	−0.400 0	0.291 ± 0.030 0	1.372 7
是	1.5	0.05	−0.061 0	0.308 ± 0.005 0	1.428 4
否	1.5	0.05	0	0.343 ± 0.080 0	1.455 0
否	1.5	0.05	−0.200 0	0.313 ± 0.050 0	1.410 8
否	1.5	0.05	−0.400 0	0.296 ± 0.040 0	1.334 0
是	2.0	1.0	−0.083 6	0.396 ± 0.000 1	1.904 0
否	2.0	1.0	0	0.410 ± 0.051 0	1.963 9
否	2.0	1.0	−0.200 0	0.388 ± 0.037 0	1.835 8
否	2.0	1.0	−0.400 0	0.378 ± 0.055 0	1.792 2
是	2.0	0.2	−0.161 6	0.430 ± 0.006 0	2.060 0
否	2.0	0.2	0	0.424 ± 0.051 0	1.996 6
否	2.0	0.2	−0.200 0	0.401 ± 0.045 0	1.881 5
否	2.0	0.2	−0.400 0	0.384 ± 0.040 0	1.813 7
是	2.0	0.05	−0.275 0	0.239 ± 0.002 0	1.730 0
否	2.0	0.05	0	0.441 ± 0.089 0	1.871 8
否	2.0	0.05	−0.200 0	0.411 ± 0.049 0	1.719 3
否	2.0	0.05	−0.400 0	0.389 ± 0.011 6	1.495 7

化、足端滑动距离和实际运动速度关键指标上的表现. 实验结果表明, 自适应控制策略能够根据环境摩擦系数的变化, 动态调整质心高度以维持较好的性能.

在标准摩擦系数 ($\mu = 1.0$) 条件下, 自适应控制表现出明显的效率优势. 如表 7 和图 8(c) 所示, 自适应控制仅需降低质心高度 0.031 1 m, 即可将足端滑动距离控制在 $0.290 \pm 0.000\,4$ cm. 相比之下, 基准控制器在质心高度降低 0.400 0 m 时, 足端滑动距离为 $0.286 \pm 0.040\,0$ cm, 虽然滑动距离相近, 但需要付出更大的速度代价, 运动速度显著衰减至 1.353 9 m/s, 而自适应控制则保持在 1.490 0 m/s 的较高水平.

当摩擦系数降至 0.2 时, 自适应控制的优势进一步凸显. 如图 8(b) 所示, 控制器通过适度降低质心高度 0.038 0 m, 就实现了优秀的综合性能: 足端滑动距离仅为 $0.312 \pm 0.004\,0$ cm, 保持了 1.502 0 m/s 的理想运动速度. 相比之下, 基准控制器即使将质心降低 0.400 0 m, 其性能表现 (滑动距离为 $0.291 \pm 0.030\,0$ cm, 速度为 1.372 7 m/s) 仍不及自适应控制.

在极低摩擦环境 ($\mu = 0.05$) 下, 自适应控制的优势最为突出. 如图 8(a) 所示, 通过将质心高度降低 0.061 0 m, 自适应控制就实现了较小的足端滑动距离 ($0.308 \pm 0.005\,0$ cm) 和稳定的运动速度 (1.428 4 m/s). 而基准控制器即使采用 0.400 0 m 的最大质心调整量, 其性能 (滑动距离为 $0.296 \pm 0.040\,0$ cm, 速度为 1.334 0 m/s) 仍明显不如自适应控制.

3.3.2 不同运动速度下的性能分析

为进一步验证控制策略在不同运动速度下的优势, 本研究分别在低速 (1.5 m/s) 和高速 (2.0 m/s) 两种期望速度下开展系统性实验. 如图 8(d) ~ 8(f) 所示, 自适应控制在速度跟踪和运动稳定性方面均展现出显著优势.

低速运动条件下, 自适应控制通过精确的质心高度调整实现了优异的控制效果. 即使在低摩擦环境 ($\mu = 0.05$) 中, 控制器仅需降低质心高度 0.061 0 m, 就能将运动速度维持在 1.428 4 m/s, 达到期望速度的 95.2%. 而基准控制器虽然通过大幅降低质心高度 (0.400 0 m) 实现了相似的足端滑动控制效果, 但速度表现却明显不足.

高速运动条件下, 如表 7 所示, 自适应控制通过 0.275 0 m 的质心高度调整, 实现了 $0.239 \pm 0.002\,0$ cm 的极低足端滑动距离, 同时保持了 1.730 0 m/s 的较高运动速度. 相比之下, 基准控制器即使采用最大的质心高度调整 (0.400 0 m), 足端滑动距离仍达到 $0.389 \pm 0.011\,6$ cm, 且运动速度仅能维持在 1.495 7 m/s 的较低水平.

综合实验结果表明, 自适应控制策略相比基准控制器具有显著优势. 它能够通过更为合理的质心高度调整策略, 在保持较高运动速度的同时有效抑制足端滑动, 提升运动稳定性. 这种优势在高速低摩擦等极具挑战性的环境下表现得尤为突出, 充分证明了该控制策略的实用价值.

4 结束语

本文提出一种新型四足机器人运动控制方法, 通过融合基于模型和基于学习的方法, 实现了机器人在不同摩擦系数地面的稳定运动. 主要贡献包括: 1) 在奖励函数设计中, 提出基于单次落足点偏差的惩罚机制, 有效解决了传统方法对瞬时滑动过度惩罚的问题; 2) 在接触建模方面, 采用双曲正切函数构建连续状态描述, 显著改善了传统离散方法中的冲击问题; 3) 在环境感知上, 提出基于 LSTM 的地面摩擦特性估计方法, 实现了质心高度的实时自适

应调整; 4) 在控制架构上, 设计执行层与决策层协同的分层控制框架, 实现了基于模型的控制与强化学习方法的有效融合.

实验结果表明, 基于双曲正切函数的连续状态描述方法有效改善了传统离散方法中的相位切换冲击问题; 自适应控制策略通过动态调整质心高度, 成功应对了不同摩擦系数 ($\mu \in \{0.05,\ 0.2,\ 1.0\}$) 和运动速度 (1.5 m/s、2.0 m/s) 条件. 特别是在低摩擦环境 ($\mu = 0.05$) 下, 仅需 0.061 0 m 的质心调整就能将足端滑动距离控制在 $0.308 \pm 0.005\ 0$ cm, 同时保持 1.428 4 m/s 的较高运动速度, 验证了所提方法的有效性和实用价值.

未来工作将着重于以下方面: 研究针对不同地形特征的步态自适应选择策略, 扩展至多样化地形环境的实验验证、自适应策略参数的优化以及实物平台的部署测试, 以进一步提升控制方法的性能和实用性.

附录 A

表 A1　变量含义说明

Table A1　Description of the meaning of variables

类别	变量	含义
运动跟踪	$v_{x,\ y}$	质心在 x-y 平面上的实际速度
	$v_{x,\ y}^{cmd}$	期望的质心水平面速度
	ω_z	质心绕 z 轴的实际角速度
	ω_z^{cmd}	期望的质心偏航角速度
	h_z	质心的实际垂直高度
	h_z^{cmd}	期望的质心垂直高度
	ρ	躯干的实际俯仰角
	ρ^{cmd}	期望的俯仰角
	$C_{\text{foot}}^{\text{cmd}}(t)$	t 时刻, 期望的足端接触状态
	$f_{\text{foot}}^{\text{cmd}}$	期望的足端接触力
	$v_{\text{foot}}^{\text{cmd}}$	期望的足端速度
	p_{foot}^{c}	足端实际接触时的位置
	$p_{\text{foot}}^{c,\ \text{cmd}}$	期望落足点位置
	$\mathbb{I}(t = t_{\text{contact}})$	指示函数
	σ_*	缩放因子
姿态稳定性	v_z	质心在 z 轴上的实际速度
	$\omega_{x,\ y}$	质心绕 x-y 轴的实际角速度
	$v_{\text{foot},\ x,\ y}$	足端在 x-y 平面上的实际速度
运动约束	$1_{\text{collision}}$	是否发生碰撞
	q_i	第 i 个关节的实际角度
	$q_{\max}$, $q_{\min}$	关节角度的上限和下限
	τ	施加在关节上的力矩
	$\dot{q}$	关节角速度
	$\ddot{q}$	关节角加速度
	a_t, a_{t-1}	在 t 和 $t-1$ 时刻的动作

References

1 Shao Y, Jin Y, Liu X, He W, Wang H, Yang W. Learning free gait transition for quadruped robots via phase-guided controller. *IEEE Robotics and Automation Letters*, 2021, **7**(2): 1230−1237

2 Kang D, de Vincenti F, Adami N, Coros S. Animal motions on legged robots using nonlinear model predictive control. In: Proceedings of the International Conference on Intelligent Robots and Systems (IROS). Kyoto, Japan: IEEE, 2022. 11955−11962

3 Wensing P, Posa M, Hu Y, Escande A, Mansard N. Optimization-based control for dynamic legged robots. *IEEE Transactions on Robotics*, 2023, **40**: 43−63

4 Ding Y, Pandala A, Park H. Real-time model predictive control for versatile dynamic motions in quadrupedal robots. In: Proceedings of the International Conference on Robotics and Automation (ICRA). Montreal, Canada: IEEE, 2019. 8484−8490

5 Romero A, Song Y, Scaramuzza D. Actor-critic model predictive control. arXiv preprint arXiv: 2306.09852, 2023.

6 Hwangbo J, Lee J, Dosovitskiy A, Bellicoso D, Tsounis V, Koltun V, et al. Learning agile and dynamic motor skills for legged robots. *Science Robotics*, 2019, **4**(26): Article No. eaau5872

7 Kumar A, Fu Z, Pathak D, Malik J. RMA: Rapid motor adaptation for legged robots. arXiv preprint arXiv: 2107.04034, 2021.

8 Rudin N, Hoeller D, Reist P, Hutter M. Learning to walk in minutes using massively parallel deep reinforcement learning. In: Proceedings of the Conference on Robot Learning (CoRL). London, UK: PMLR, 2022. 91−100

9 Hasson C, Manczurowsky J, Yen S. A reinforcement learning approach to gait training improves retention. *Frontiers in Human Neuroscience*, 2015, **9**: 459−467

10 Haarnoja T, Ha S, Zhou A, Tucker G, Levine S. Learning to walk via deep reinforcement learning. arXiv preprint arXiv: 1812.11103, 2018.

11 Weng J, Hashemi E, Arami A. Natural walking with musculoskeletal models using deep reinforcement learning. *IEEE Robotics and Automation Letters*, 2021, **6**(2): 4156−4162

12 Shi H, Zhou B, Zeng H, Wang F, Dong Y, Li J. Reinforcement learning with evolutionary trajectory generator: A general approach for quadrupedal locomotion. *IEEE Robotics and Automation Letters*, 2022, **7**(2): 3085−3092

13 Chai H, Rong X, Tang X, Li Y. Gait-based quadruped robot planar hopping control with energy planning. *International Journal of Advanced Robotic Systems*, 2016, **13**(1): 20−32

14 Raibert M. Trotting, pacing and bounding by a quadruped robot. *Journal of Biomechanics*, 1990, **23**: 79−98

15 Jiang Z, Li M, Guo W. Running control of a quadruped robot in trotting gait. In: Proceedings of the IEEE 5th International Conference on Robotics, Automation and Mechatronics (RAM). Qingdao, China: IEEE, 2011. 172−177

16 Fukuoka Y, Kimura H. Dynamic locomotion of a biomorphic quadruped 'Tekken' robot using various gaits: Walk, trot, free-gait and bound. *Applied Bionics and Biomechanics*, 2009, **6**(1): 63−71

17 Adak O, Erbatur K. Bound gait reference generation of a quadruped robot via contact force planning. *International Journal of Mechanical Engineering and Robotics Research*, 2022, **11**(3): 129−137

18 Xiao W, Wang W. Hopf oscillator-based gait transition for a quadruped robot. In: Proceedings of the International Conference on Robotics and Biomimetics (ROBIO). Bali, Indonesia: IEEE, 2014. 2074−2079

19 Margolis G, Agrawal P. Walk these ways: Tuning robot control for generalization with multiplicity of behavior. In: Proceedings

of the Conference on Robot Learning (CoRL). Auckland, New Zealand: PMLR, 2023. 22−31

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

21 Heiden T, Sanderson D, Inglis J, Siegmund G. Adaptations to normal human gait on potentially slippery surfaces: The effects of awareness and prior slip experience. *Gait & Posture*, 2006, **24**(2): 237−246

22 Cappellini G, Ivanenko Y, Dominici N, Poppele R, Lacquaniti F. Motor patterns during walking on a slippery walkway. *Journal of Neurophysiology*, 2010, **103**(2): 746−760

23 Puterman L. Markov decision processes. *Handbooks in Operations Research and Management Science*, 1990, **2**: 331−434

24 Miki T, Lee J, Hwangbo J, Koltun K, Hutter M. Learning robust perceptive locomotion for quadrupedal robots in the wild. *Science Robotics*, 2022, **7**(62): Article No. eabk2822

25 Makoviychuk V, Wawrzyniak L, Guo Y, Storey K, Macklin M, Hoeller D, et al. Isaac Gym: High performance gpu-based physics simulation for robot learning. arXiv preprint arXiv: 2108.10470, 2021.

张楠杰 河海大学人工智能与自动化学院硕士研究生. 主要研究方向为无人系统控制和具身智能.
E-mail: 231622010045@hhu.edu.cn
(**ZHANG Nan-Jie** Master student at the College of Artificial Intelligence and Automation, Hohai University. His research interest covers unmanned system control and embodied artificial intelligence.)

陈玉全 河海大学人工智能与自动化学院副教授. 主要研究方向为具身智能, 智能优化, 分数阶系统与控制. 本文通信作者.
E-mail: cyq@mail.ustc.edu.cn
(**CHEN Yu-Quan** Associate professor at the College of Artificial Intelligence and Automation, Hohai University. His research interest covers embodied artificial intelligence, intelligent optimization, and fractional order system and control. Corresponding author of this paper.)

季茂沁 河海大学人工智能与自动化学院硕士研究生. 主要研究方向为非渐近收敛理论和无人系统控制.
E-mail: 231322010005@hhu.edu.cn
(**JI Mao-Qin** Master student at the College of Artificial Intelligence and Automation, Hohai University. His research interest covers non-asymptotic convergence theory and unmanned system control.)

孙运康 河海大学人工智能与自动化学院硕士研究生. 主要研究方向为多智能体控制, 无人系统控制, 自适应控制.
E-mail: sunyunkanghhu@163.com
(**SUN Yun-Kang** Master student at the College of Artificial Intelligence and Automation, Hohai University. His research interest covers multi-agent control, unmanned system control, and adaptive control.)

王 冰 河海大学人工智能与自动化学院教授. 主要研究方向为无人系统控制, 新能源发电控制, 电力系统调度. E-mail: icekingking@hhu.edu.cn
(**WANG Bing** Professor at the College of Artificial Intelligence and Automation, Hohai University. His research interest covers unmanned system control, new energy generation control, and power system scheduling.)