

清 华 大 学

综 合 论 文 训 练
文献综述与选题报告

题目：基于人形机器人本体感觉的
多地形强化学习步态生成方法

院 系：自动化系

专 业：控制科学与工程

学 号：2021013445

姓 名：李昭阳

指导教师：张涛

目录:

第一章 研究背景和意义.....	3
第二章 文献综述.....	3
2.1 人形机器人.....	3
2.2 人形机器人步态规划方法.....	4
2.3 基于本体感知的地形预测方法.....	6
第三章 研究内容.....	7
3.1 基于强化学习的步态库生成.....	7
3.2 基于本体感觉的地形预测.....	7
3.3 自适应时变步态顺滑切换.....	8
第四章 方案设计.....	8
4.1 Isaac Gym 仿真环境搭建.....	8
4.2 PPO 强化学习算法部署与数据收集.....	9
4.3 分类神经网络搭建.....	9
4.4 测试实验.....	9
第五章 日程计划与安排.....	10
参考文献.....	11

第一章 研究背景和意义

人类的运动能力涵盖了广泛的范围[1]，拥有适应各种任务所需的不同步态，例如站立、行走、快步走、小跑、奔跑和跳跃[2]。这种适应性甚至包括蹲姿行走，使人类能够在广泛的动作范围中表现出卓越的灵活性和机动性。

人类生活中的环境主要为双足行走而设计，其丰富多样性使得双足机器人领域的总体目标是开发能够在这些环境中可靠运行的机器人。其中，多地形步态规划是一个复杂且关键的问题。人形机器人需要能够在不同的地形上稳定行走，包括但不限于平坦地面、斜坡、不平坦的地面等。多地形步态规划的目标是确保机器人在这些多变环境中保持平衡并更高效地移动。同时，人形机器人在夜间、雨雪、火场等低能见度场景有广泛的应用前景，不依赖外部视觉输入来感知和预测地形变化至关重要[3]。仅依靠本体感觉来识别和适应各种地形，可以在保持稳定性和效率的同时，增强其在极端环境中的自主性和适应性。

本毕业设计项目旨在开发一套基于本体感觉的自适应强化学习步态选择执行算法。具体而言，首先通过强化学习预训练一个丰富的步态库；然后，仅利用关节角度、姿态信息和平衡状态等本体感觉数据（不依赖视觉和雷达信息），预测当前地形；最后，依据预测结果自适应选择合适的步态并进行微调，实现高效的多地形行走。

本毕业设计项目的主要预期创新点为，设计人形机器人基于本体感觉的地形预测算法，在此基础上调用步态库并实现自适应时变步态顺滑切换。

第二章 文献综述

2.1 人形机器人

近年来，人形机器人领域备受关注，众多研究机构和公司相继发布了尖端创新成果和研究进展，标志着该领域的快速发展与崛起。波士顿动力公司的 Atlas 机器人已展示出跑酷级的运动能力；Tesla 的 Optimus[4]和 Figure 的人形机器人[5]利用人类数据学习并执行复杂的桌面操作任务；双足机器人 Cassie[5][6][7]及其人形版本 Digit[6]，依靠电机驱动成功适应多种地形；知名的腿足机器人公司 Unitree 推出了其人形机器人产品 H1、G1[7]；Apptronik 公司研发了完全由推杆

电机驱动的人形机器人 Apollo[8]；在通用人工智能领域备受瞩目的 OpenAI[9]收购了 1X 机器人公司[10]，并提出了面向具身智能的开发计划。以上种种表明，人形机器人正逐渐成为研究者和企业关注的关键方向之一，掌握人形机器人核心技术对于弥合数字化通用人工智能与实体硬件之间的鸿沟具有重要意义。



图 1 宇树科技 H1、G1 型号人形机器人

2.2 人形机器人步态规划方法

大多数人类运动可以被视为周期性或准周期性[11][12]，这为人形机器人运动控制器的设计提供了极大的便利。在过去几十年中，学术界已经开发了许多机器人及相应的控制算法，以实现周期性步态[13][14][15]。尽管取得了显著进展，人形机器人在运动范围方面仍无法与人类相提并论。因此，尤其是像跑步这样的高动态步态的实现，已引起越来越多的关注。

通常有两种方法可以生成稳定的周期性步态，其中之一是基于模型的方法[16]。这种方法通过建立简化或精确的模型来分析运动学或动力学，从而生成理想的参考轨迹，并通过基于模型计算的反馈控制来跟踪该轨迹，以确保机器人的行走模式稳定[17][18]。线性倒立摆模型（LIPM）是人形和四足机器人中应用最广泛的简化模型之一[19][20]。Kajita 等人[21]将机器人建模为固定高度的倒立摆，将复杂的动力学简化为一个质心点质量，并使用预览控制方法结合零力矩点（ZMP）控制。基于 LIPM，Takenaka 等人[22]将其扩展为可变高度倒立摆模型（VHIP），并分析了运动的发散分量（DCM），以生成周期性跑步步态。Han 等人[23]结合 3D 弹簧加载倒立摆（3D-SLIP）模型与任务空间非线性最小二乘优化器，实现了 6.5 m/s 的稳定跑步。单刚体模型（SRBM）在基于优化的方法中被广泛用于质心运动规划[24][25][26]。通过手动指定接触顺序，并参考 SRBM 在线

求解优化问题，Bellicoso 等人[27]在四足机器人上实现了多种步态及平滑过渡。García 等人[28]提出了结合 SRBM 和方向变量的 MPC 运动控制方法。通过模仿生物体的神经机制，中央模式发生器（CPG）使用耦合振荡器模型生成可调节频率和幅度的步态周期，以响应感官反馈和环境变化[29][30][31][32]。

模型方法的一个主要缺点是，这些简化模型通常假设机器人某些关节是被动的或锁定的[32][33]，只允许一种或有限的步态模式。这导致运动空间仅代表了机器人全部潜力的一小部分，限制了对所有驱动关节所提供的完整运动范围和灵活性的探索和利用。

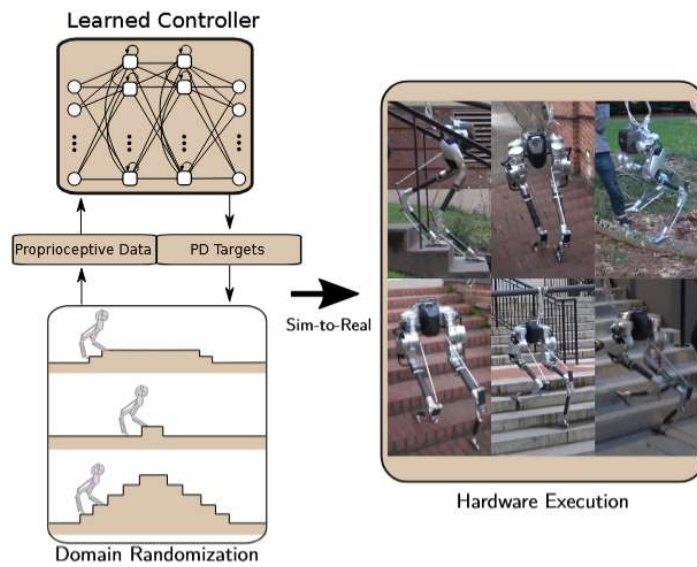


图2 [40]文章描述方法结构图

另一种实现人形机器人稳定周期性步态的方法是强化学习。在过去十年中，强化学习越来越多地应用于人形机器人，实现了如爬台阶[33]和后空翻[34]等复杂动作。这种方法不需要精确的机器人建模，而是通过智能体与环境的交互逐步学习运动控制策略[35][36]，从而缩短了开发周期[37]。通过在 GPU 上进行大规模并行计算，平地行走的训练时间已缩短至仅 5 分钟[37]。强化学习更高效地生成了周期性步态。Wu 等人[38]表明，通过使用简单的正弦信号来获得预期的摆动-支撑时序并将其作为观测的一部分，并设计奖赏函数以跟踪该时序，机器人可以实现固定频率的行走。Li 等人[39]提出了傅里叶潜在动力学（FLD）方法，从人类运动中提取周期特征，通过傅里叶变换将动作捕捉的关节角度转化为频域信息，并利用解码器重建这些特征，以模仿原始人类动作。即使直接在奖励函数中引入周期性的脚部提升和落地，强化学习也能够在不依赖周期性参考信号的情

况下生成自适应步态模式。[40] 修改现有的平地训练 RL 框架，加入楼梯状地形的随机化，无需改变奖励函数，首次展示双足机器人仅使用本体感知穿越真实世界的楼梯。[41]描述了第一个通过端到端强化学习掌握真实世界挑战性地形的人形机器人，提出了一个新颖的强化学习框架，以弥合模拟到现实的差距，并实现鲁棒的泛化性能。尽管 RL 目前仅限于特定步态，覆盖了运动能力的一小部分，但其简单性和自主学习特性为探索人形机器人的运动能力和扩展步态空间提供了潜力。

2.3 基于本体感知的地形预测方法

由于视觉、雷达系统的许多局限性，依赖本体感觉的系统在自主机器人研究中日益受到欢迎。Brooks 等[42]使用车载加速度计，通过测量振动并采用波形表示进行在线处理，同时应用主成分分析（PCA）识别地形类别，实验将地形分为四类：砾石、泥土、沙地和未分类地形，分类的置信水平达到 96%至 100%。Dupont 等[43]提出了一种基于频率响应的振动分析方法，用于通过快速傅里叶变换（FFT）分类地形。Giguire 等[44]针对常规触觉传感器和视觉传感器存在的局限性，开发了一种安装在车轮上的触觉探针，无需依赖机器人设计，灵活性较高，并能以 94.6%的准确率分类地形。Chengchao Bai 等[45]提出了一种基于多传感器融合的地形识别方法，将地形分为砖块、沙地、平地、水泥和土壤五种类型。Nampoothiri 等[46]开发了一种基于惯性测量单元（IMU）的未知地形导航机器学习方法，通过 23 种不同算法对机器人导航地形进行实时分类。Mariela 等[47]利用了神经网络，从本体感觉数据中增强不确定性，提高了分类的透明度和可靠性，为太空探索机器人的地形分类提供了一种新的方法。

文章索引	机器人类型	使用非本体感觉	具体方法
[42]	四轮月球车	x	主成分分析
[43]	四轮车	x	快速傅里叶变换
[44]	扫地机器人	√	触觉探针
[45]	四轮车	√	多传感器融合
[46]	三轮车	x	机器学习
[47]	异形月球车	x	机器学习

表 1 基于本体感知的地形预测方法

第三章 研究内容

在第二章中已经提到，针对机器人在复杂环境中高效行走的挑战，现有的步态生成和调整方法存在一定的局限性。例如，传统基于强化学习的步态生成方法通常追求极致的泛化性以适应多变的地形条件，导致对单一地形的步态缺乏“专业”；而基于视觉输入的地形感知方法虽然能够提高环境感知能力，但也增加了对外部传感器的依赖，限制了机器人在极端、动态环境中的自适应能力。因此，本项目的目标是，通过结合强化学习与本体感觉系统，实现一个自适应、多地形环境下的步态库生成与切换方法。该方法的核心是基于 Proximal Policy Optimization (PPO)[48]算法训练步态库，利用本体感觉数据进行地形预测，并通过自适应的时变步态顺滑切换，提高机器人的稳定性和舒适性。其关键点和难点在于，如何构建多地形步态库的生成框架、如何进行地形预测和分类、以及如何平滑地切换不同步态，以保证机器人在各种复杂环境中的高效和稳定运动。

3.1 基于强化学习的步态库生成

强化学习通过智能体与环境的交互，逐步学习运动控制策略[35][36]。Isaac Gym[49]能够模拟复杂的多地形环境，包括平坦地面、斜坡和不平坦地面等关键地形条件。根据本毕业设计项目的目标，需要对预测的地形状态应用相应步态，因此设立步态库并预先训练所需步态以备调用。PPO 算法能够有效处理连续动作空间的问题，并具有较快的收敛速度。基于此，计划搭建 PPO 强化学习神经网络与 Isaac Gym 仿真环境，利用相同的神经网络结构在不同地形上分别训练，获得并存储相应的网络参数，最终实现多地形步态库。

3.2 基于本体感觉的地形预测

如第一章所述，减少对外部视觉输入的依赖，改用本体感觉来预测地形变化，能够提高机器人的自适应能力，尤其在复杂的动态环境中。支持向量机[50]、决策树[51]以及深度学习[52]方法在分类问题中显示出较强的潜力，特别是在模式识别和特征分类方面，能够从复杂数据中提取有价值的信息，进而实现精准的分类与预测。利用 3.1 节训练过程中积累的本体感觉数据，可以为分类方法提供高质量的数据集。因此，初步设想是将机器人本体感觉系统采集的参数作为输入，通过上述模型进行处理，完成地形分类。该方法将根据不同地形条件下机器人本

体的动态响应，自动提取特征并进行分类，实现地形预测与辨识。

3.3 自适应时变步态顺滑切换

本项目的机器人在不同地形上使用不同的神经网络参数来适应复杂环境，因此步态切换时可能会出现较为明显的突变。这些突变表现在关节角度、步幅、步态周期等运动参数上，会导致步态变化不够平滑，从而影响机器人运动的舒适性和稳定性。为了解决这一问题，初步考虑在步态切换过程中引入 PID 控制[35]等传统控制方法。具体来说，PID 控制可以在过渡阶段精细调节各关节的运动，通过调节比例、积分和微分项平衡系统的响应速度与稳定性，从而避免步态切换时产生过大的波动或震荡。

第四章 方案设计

4.1 Isaac Gym 仿真环境搭建

在机器人控制算法的开发与验证之前，构建一个高效且可靠的仿真环境至关重要。Isaac Gym 可以根据实际应用需求创建一个符合实验条件的仿真环境，包括不同地面摩擦、障碍物和高度变化，模拟真实世界中的复杂情况。此外，仿真环境中的物理引擎参数，如重力、摩擦力等，也需要根据控制需求进行调整。我计划搭建包括粗糙平面、粗糙斜坡、粗糙崎岖、光滑平面等多种训练环境。Isaac Gym 能够实时获取仿真环境中的传感器数据，如机器人的关节角度、速度、加速度、姿态、位置等信息，并将控制算法的输出指令传递给仿真环境。因此，我将调配 Isaac Gym 的传感器接口，以便用于状态和观测接入强化学习神经网络，同时测试控制算法接口，以便于强化学习动作输出进行仿真。

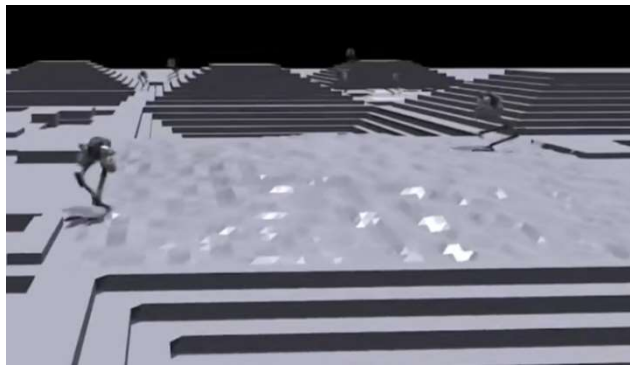


图 3 双足机器人在 Isaac Gym 中

4.2 PPO 强化学习算法部署与数据收集

PPO 算法的核心优势在于通过使用“剪切”目标函数来限制策略的更新幅度，有效避免由于过大更新导致的训练不稳定问题。PPO 算法的流程包括数据收集、策略更新和价值函数更新三个主要步骤。在数据收集阶段，机器人通过与环境交互，获取当前状态、所执行的动作及其相应的奖励信息。在策略更新阶段，PPO 利用当前收集的数据优化目标函数，更新策略，使其更加符合期望的行动目标。在价值函数更新阶段，通过时间差分法（TD）更新价值函数，以准确评估不同状态的期望奖励。

对于本研究中的机器人控制任务，PPO 算法被用来优化机器人在不同地形的步态。在完成算法编写后，接入仿真环境进行训练。一方面在不同地形上进行不同网络参数的基础训练，另一方面在仿真环境中引入干扰，训练机器人在遭遇外部扰动（如碰撞、推力等）时的反应能力和恢复能力。收集仿真过程中机器人的关节角度、力矩、速度、姿态变化等数据，为地形预测分类积累数据集。

4.3 分类神经网络搭建

在建立分类神经网络时，首先需要对数据集进行预处理，确保其适合神经网络训练。我会将收集到的关节数据、力矩数据和姿态数据整理成适合输入到神经网络的格式，同时，考虑到不同地形的特征差异，采用适当的标签进行标注，确保每种数据能够与对应的地形类型准确匹配。神经网络将通过这些输入数据学习地形的特征，从而实现分类功能。

在训练过程中，网络将通过反向传播算法优化，逐步调整权重和偏置，从而最小化分类误差。为了提高模型的稳定性和准确度，我会采用交叉验证方法，确保网络的表现不会因为过拟合或欠拟合而受到影响。最终目标是构建一个高效、可靠的地形分类预测器，能够基于机器人感知的本体数据准确判断当前地形类型，为机器人在复杂多变环境中的自适应行走提供有力支持。

4.4 测试实验

本项目的实验旨在通过一系列精心设计的测试场景，模拟现实世界中的复杂挑战，以评估算法在复杂地形中行走及步态切换的能力。仿真实验过程中，将对步态的关键参数进行持续监控，包括步长、步速及姿态调整等，以确保机器人能

够实现稳定且高效的行走表现。此外，将记录机器人的平衡状态和行走效率，以评估训练算法的性能。

为确保实验结果的可靠性，我将在虚拟环境中设置多个实验副本，确保每个副本具有一致的初始条件和参数设置。通过对比不同副本的实验结果，分析算法的稳定性和可靠性。实验数据将采用统计方法进行处理，以确保结果的有效性。

在条件允许的情况下，计划将经过充分验证的训练算法应用在实体机器人上进行测试，并根据实际情况进行调整和优化。我将设计多样化的实地测试场景，通过收集和分析机器人的关节角度、速度、力矩、电池消耗和与环境的交互数据，评估算法的实际表现，并与仿真结果进行对比，最终评估算法的泛化能力与适应性。

第五章 日程计划与安排

内容	时间安排
搭建步态规划的模拟环境，继续调研强化学习算法	2024.12.01-2024.12.31
通过 PPO 强化学习算法训练步态库	2024.01.01-2025.02.15
设计并训练基于本体感觉的地形分类预测算法	2025.02.16-2025.03.15
对未预期的异常进行处理	2025.03.16-2025.03.31
仿真和实机实验	2025.04.01-2025.05.15
论文撰写	2025.05.16-2025.05.30

参考文献

- [1] J. K. Aggarwal and Q. Cai, "Human motion analysis: A review," **Computer Vision and Image Understanding**, vol. 73, no. 3, pp. 428 – 440, 1999.
- [2] J. P. Gupta, D. Polytool, N. Singh, and V. B. Semwal, "Analysis of gait pattern to recognize the human activities," **IJIMAI**, vol. 2, no. 7, pp. 7 – 16, 2014.
- [3] Michele Focchi, Romeo Orsolino, Marco Camurri, Victor Barasuol, Carlos Mastalli, Darwin G. Caldwell, and Claudio Semini. Heuristic Planning for Rough Terrain Locomotion in Presence of External Disturbances and Variable Perception Quality, volume 132, pages 165 – 209. Springer International Publishing, Cham, 2020. ISBN 978-3-030-22327-4. doi: 10.1007/ 978-3-030-22327-4 9.
- [4] Wikipedia, "Optimus (robot)," [https://en.wikipedia.org/wiki/Optimus \(robot\)](https://en.wikipedia.org/wiki/Optimus_(robot)), 2024, accessed: 2024-02-28.
- [5] Y. Gong, R. Hartley, X. Da, A. Hereid, O. Harib, J.-K. Huang, and J. Grizzle, "Feedback control of a cassie bipedal robot: Walking, standing, and riding a segway," in 2019 American Control Conference (ACC). IEEE, 2019, pp. 4559 – 4566.
- [6] Z. Li, X. B. Peng, P. Abbeel, S. Levine, G. Berseth, and K. Sreenath, "Reinforcement learning for versatile, dynamic, and robust bipedal locomotion control," arXiv preprint arXiv:2401.16889, 2024.
- [7] Z. Li, X. Cheng, X. B. Peng, P. Abbeel, S. Levine, G. Berseth, and K. Sreenath, "Reinforcement learning for robust parameterized locomotion control of bipedal robots," in IEEE International Conference on Robotics and Automation (ICRA), Xi' an, China, June 2021.
- [8] C. E. Garcia, D. M. Prett, and M. Morari, "Model predictive control: Theory and practice—a survey," *Automatica*, vol. 25, no. 3, pp. 335348, 1989.
- [9] N. Rudin, D. Hoeller, P. Reist, and M. Hutter, "Learning to walk in minutes using massively parallel deep reinforcement learning," in Conference on Robot Learning. PMLR, 2022, pp. 91 – 100.
- [10] Z. Li, X. B. Peng, P. Abbeel, S. Levine, G. Berseth, and K. Sreenath, "Robust and versatile bipedal jumping control through reinforcement learning."
- [11] S. R. Simon, "Quantification of human motion: gait analysis—benefits and limitations to its

- application to clinical problems," **Journal of Biomechanics**, vol. 37, no. 12, pp. 1869 – 1880, 2004.
- [12] S. Shahid, A. Nandy, S. Mondal, M. Ahamad, P. Chakraborty, and G. C. Nandi, "A study on human gait analysis," in **Proceedings of the Second International Conference on Computational Science, Engineering and Information Technology**, 2012, pp. 358 – 364.
- [13] Y. Sakagami, R. Watanabe, C. Aoyama, S. Matsunaga, N. Higaki, and K. Fujimura, "The intelligent Asimo: System overview and integration," in **IEEE/RSJ International Conference on Intelligent Robots and Systems**, vol. 3, IEEE, 2002, pp. 2478 – 2483.
- [14] M. Chignoli, D. Kim, E. Stanger-Jones, and S. Kim, "The MIT humanoid robot: Design, motion planning, and control for acrobatic behaviors," in **2020 IEEE-RAS 20th International Conference on Humanoid Robots (Humanoids)**, IEEE, 2021, pp. 1 – 8.
- [15] S. Kuindersma, R. Deits, M. Fallon, A. Valenzuela, H. Dai, F. Permenter, T. Koolen, P. Marion, and R. Tedrake, "Optimization-based locomotion planning, estimation, and control design for the Atlas humanoid robot," **Autonomous Robots**, vol. 40, pp. 429 – 455, 2016.
- [16] K. Yamamoto, T. Kamioka, and T. Sugihara, "Survey on model-based biped motion control for humanoid robots," **Advanced Robotics**, vol. 34, no. 21-22, pp. 1353 – 1369, 2020.
- [17] H. Chen, X. Chen, C. Dong, Z. Yu, and Q. Huang, "Online running pattern generation for humanoid robot with direct collocation of reference tracking dynamics," **IEEE/ASME Transactions on Mechatronics**, 2023.
- [18] A. Escande, N. Mansard, and P.-B. Wieber, "Hierarchical quadratic programming: Fast online humanoid-robot motion generation," **The International Journal of Robotics Research**, vol. 33, no. 7, pp. 1006 – 1028, 2014.
- [19] S. Kajita, F. Kanehiro, K. Kaneko, K. Yokoi, and H. Hirukawa, "The 3D linear inverted pendulum mode: A simple modeling for a biped walking pattern generation," in **Proceedings 2001 IEEE/RSJ International Conference on Intelligent Robots and Systems**, vol. 1, IEEE, 2001, pp. 239 – 246.
- [20] S. Kajita, M. Morisawa, K. Miura, S. Nakaoka, K. Harada, K. Kaneko, F. Kanehiro, and K. Yokoi, "Biped walking stabilization based on linear inverted pendulum tracking," in **2010 IEEE/RSJ International Conference on Intelligent Robots and Systems**, IEEE, 2010, pp. 4489 – 4496.
- [21] S. Kajita, F. Kanehiro, K. Kaneko, K. Fujiwara, K. Harada, K. Yokoi, and H. Hirukawa, "Biped

- walking pattern generation by using preview control of zero-moment point," in *2003 IEEE International Conference on Robotics and Automation*, vol. 2, IEEE, 2003, pp. 1620 – 1626.
- [22] T. Takenaka, T. Matsumoto, T. Yoshiike, and S. Shirokura, "Real time motion generation and control for biped robot-2nd report: Running gait pattern generation," in *2009 IEEE/RSJ International Conference on Intelligent Robots and Systems*, IEEE, 2009, pp. 1092 – 1099.
- [23] B. Han, H. Yi, Z. Xu, X. Yang, and X. Luo, "3D-slip model based dynamic stability strategy for legged robots with impact disturbance rejection," *Scientific Reports*, vol. 12, no. 1, p. 5892, 2022.
- [24] P. M. Wensing and D. E. Orin, "Improved computation of the humanoid centroidal dynamics and application for whole-body control," *International Journal of Humanoid Robotics*, vol. 13, no. 01, p. 1550039, 2016.
- [25] J. Shen and D. Hong, "Convex model predictive control of single rigid body model on $SO(3)$ for versatile dynamic legged motions," in *2022 International Conference on Robotics and Automation (ICRA)*, IEEE, 2022, pp. 6586 – 6592.
- [26] R. Batke, F. Yu, J. Dao, J. Hurst, R. L. Hatton, A. Fern, and K. Green, "Optimizing bipedal maneuvers of single rigid-body models for reinforcement learning," in *2022 IEEE-RAS 21st International Conference on Humanoid Robots (Humanoids)*, IEEE, 2022, pp. 714 – 721.
- [27] C. D. Bellicoso, F. Jenelten, C. Gehring, and M. Hutter, "Dynamic locomotion through online nonlinear motion optimization for quadrupedal robots," *IEEE Robotics and Automation Letters*, vol. 3, no. 3, pp. 2261 – 2268, 2018.
- [28] G. García, R. Griffin, and J. Pratt, "MPC-based locomotion control of bipedal robots with line-feet contact using centroidal dynamics," in *2020 IEEE-RAS 20th International Conference on Humanoid Robots (Humanoids)*, IEEE, 2021, pp. 276 – 282.
- [29] J. Nassour, P. Hénaff, F. Benouezdou, and G. Cheng, "Multi-layered multi-pattern CPG for adaptive locomotion of humanoid robots," *Biological Cybernetics*, vol. 108, pp. 291 – 303, 2014.
- [30] J. Or, "A hybrid CPG – ZMP control system for stable walking of a simulated flexible spine humanoid robot," *Neural Networks*, vol. 23, no. 3, pp. 452 – 460, 2010.
- [31] C. Yao, C. Liu, L. Xia, M. Liu, and Q. Chen, "Humanoid adaptive locomotion control through a bioinspired CPG-based controller," *Robotica*, vol. 40, no. 3, pp. 762 – 779, 2022.
- [32] C. Li, R. Lowe, and T. Ziemke, "Humanoids learning to walk: a natural CPG-actor-critic

architecture," **Frontiers in Neurorobotics**, vol. 7, p. 5, 2013.

[33] J. Siekmann, K. Green, J. Warila, A. Fern, and J. Hurst, "Blind bipedal stair traversal via sim-to-real reinforcement learning," **arXiv preprint arXiv:2105.08328**, 2021.

[34] Unitree, Able: "Unitree H1," 2024.[Online]. Available: https://www.bilibili.com/video/BV19p421U727/?vd_source=82b909d5abbb8108ea2220646b7b0467

[35] J. Peters, S. Vijayakumar, and S. Schaal, "Reinforcement learning for humanoid robotics," in **Proceedings of the Third IEEE-RAS International Conference on Humanoid Robots**, 2003, pp. 1 - 20.

[36] L. C. Melo and M. R. O. A. Máximo, "Learning humanoid robot running skills through proximal policy optimization," in **2019 Latin American Robotics Symposium (LARS)**, 2019 Brazilian Symposium on Robotics (SBR) and 2019 Workshop on Robotics in Education (WRE), IEEE, 2019, pp. 37 - 42.

[37] N. Rudin, D. Hoeller, P. Reist, and M. Hutter, "Learning to walk in minutes using massively parallel deep reinforcement learning," in **Conference on Robot Learning**, PMLR, 2022, pp. 91 - 100.

[38] Q. Wu, C. Zhang, and Y. Liu, "Custom sine waves are enough for imitation learning of bipedal gaits with different styles," in **2022 IEEE International Conference on Mechatronics and Automation (ICMA)**, IEEE, 2022, pp. 499 - 505.

[39] C. Li, E. Stanger-Jones, S. Heim, and S. Kim, "FLD: Fourier latent dynamics for structured motion representation and learning," **arXiv preprint arXiv:2402.13820**, 2024.

[40]

[41]

[42] Brooks CA, Iagnemma K (2005) Vibration-based terrain classification for planetary exploration rovers. *IEEE Trans Robot* 21(6):1185–1190. <https://doi.org/10.1109/TRO.2005.855994>

[43] DuPont EM, Moore CA, Collins EG, Coyle E (2008) Frequency response method for terrain classification in autonomous ground vehicles. *AutonRobots* 24(4):337 - 347. <https://doi.org/10.1007/s10514-007-9077-0>

[44] Giguere P, Dudek G (2011) A simple tactile probe for surface identification by mobile robots. *IEEE Trans Robot* 27(3):534 - 44

[45] Bai C, Guo J, Guo L, Song J (2019) Deep multi-layer perception based terrain classification

for planetary exploration rovers. *Sensors (Switzerland)* 19(14):1 – 18. [https:// doi. org/ 10. 3390/ s19143102](https://doi.org/10.3390/s19143102)

[46] Nampoothiri MGH, Anand PSG, Antony R (2020) Real time terrain identification of autonomous robots using machine learning. *J Intell Robot Appl Int.* [https:// doi. org/ 10. 1007/ s41315-020-00142-3](https://doi.org/10.1007/s41315-020-00142-3)

[47] Alvarez, Mariela De Lucas, et al. "Terrain Classification Enhanced with Uncertainty for Space Exploration Robots from Proprioceptive Data." *arXiv preprint arXiv:2407.03241* (2024).

[48] Schulman, John, et al. "Proximal policy optimization algorithms." *arXiv preprint arXiv:1707.06347* (2017).

[49] Makoviychuk, Viktor, et al. "Isaac gym: High performance gpu-based physics simulation for robot learning." *arXiv preprint arXiv:2108.10470* (2021).

[50] Cortes, Corinna. "Support-Vector Networks." *Machine Learning* (1995).

[51] Quinlan, J. Ross. *C4. 5: programs for machine learning*. Elsevier, 2014.

[52] LeCun, Yann, Yoshua Bengio, and Geoffrey Hinton. "Deep learning." *nature* 521.7553 (2015): 436-444.

综合论文训练记录表

学生姓名	李昭阳	学号	2021013445	班级	自 11														
论文题目	基于人形机器人本体感觉的多地形强化学习步态生成方法																		
主要内容以及进度安排	<p>本项目的目标是，通过结合强化学习与本体感觉系统，实现一个自适应、多地形环境下的步态库生成与切换方法。该方法的核心是基于 Proximal Policy Optimization (PPO)算法训练步态库，利用本体感觉数据进行地形预测，并通过自适应的时变步态顺滑切换，提高机器人的稳定性和舒适性。其关键点和难点在于，如何构建多地形步态库的生成框架、如何进行地形预测和分类、以及如何平滑地切换不同步态，以保证机器人在各种复杂环境中的高效和稳定运动。</p> <table><tr><td>内容</td><td>时间安排</td></tr><tr><td>搭建步态规划的模拟环境，继续调研强化学习算法</td><td>2024.12.01-2024.12.31</td></tr><tr><td>通过 PPO 强化学习算法训练步态库</td><td>2024.01.01-2025.02.15</td></tr><tr><td>设计并训练基于本体感觉的地形分类预测算法</td><td>2025.02.16-2025.03.15</td></tr><tr><td>对未预期的异常进行处理</td><td>2025.03.16-2025.03.31</td></tr><tr><td>仿真和实机实验</td><td>2025.04.01-2025.05.15</td></tr><tr><td>论文撰写</td><td>2025.05.16-2025.05.30</td></tr></table> <p>指导教师签字：_____</p> <p>考核组组长签字：_____</p> <p>年 月 日</p>					内容	时间安排	搭建步态规划的模拟环境，继续调研强化学习算法	2024.12.01-2024.12.31	通过 PPO 强化学习算法训练步态库	2024.01.01-2025.02.15	设计并训练基于本体感觉的地形分类预测算法	2025.02.16-2025.03.15	对未预期的异常进行处理	2025.03.16-2025.03.31	仿真和实机实验	2025.04.01-2025.05.15	论文撰写	2025.05.16-2025.05.30
	内容	时间安排																	
	搭建步态规划的模拟环境，继续调研强化学习算法	2024.12.01-2024.12.31																	
	通过 PPO 强化学习算法训练步态库	2024.01.01-2025.02.15																	
	设计并训练基于本体感觉的地形分类预测算法	2025.02.16-2025.03.15																	
对未预期的异常进行处理	2025.03.16-2025.03.31																		
仿真和实机实验	2025.04.01-2025.05.15																		
论文撰写	2025.05.16-2025.05.30																		
中期考核意见	<p>考核组组长签字：_____</p> <p>年 月 日</p>																		

指导教师评语	<div>指导教师签字：_____</div> <div>年 月 日</div>
评阅教师评语	<div>评阅教师签字：_____</div> <div>年 月 日</div>
答辩小组评语	<div>答辩小组组长签字：_____</div> <div>年 月 日</div>

总成绩：_____

教学负责人签字：_____

年 月 日