清 华 大 学

**综 合 论 文 训 练**

**文献翻译与选题报告**

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](#_Toc152191184)

[**第二章** **文献综述** 4](#_Toc152191185)

[**2.1大语言模型** 4](#_Toc152191186)

[**2.2 多模态感知模型** 5](#_Toc152191187)

[**2.3 基于大语言模型的具身智能工作** 7](#_Toc152191188)

[**2.4 大语言模型微调方法** 8](#_Toc152191189)

[**第三章** **研究内容** 9](#_Toc152191190)

[**3.1 环境视觉感知** 9](#_Toc152191191)

[**3.2 基于大语言模型的决策生成技术** 10](#_Toc152191192)

[**3.3对预训练大语言模型的微调策略** 10](#_Toc152191193)

[**第四章** **方案设计—具身可执行命令生成器** 10](#_Toc152191194)

[**4.1 感知并记录环境视觉信息** 10](#_Toc152191195)

[**4.2 构建视觉信息—指令—操作步骤数据集** 11](#_Toc152191196)

[**4.3 微调预训练大模型** 11](#_Toc152191197)

[**4.4 测试** 12](#_Toc152191198)

[**4.5 方案设计的总流程框图** 12](#_Toc152191199)

[**第五章** **日程计划与安排** 13](#_Toc152191200)

[**参考文献** 13](#_Toc152191201)

1. **研究背景和意义**

人类的运动能力涵盖了广泛的范围 [1]，能够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。不过，大多数人类运动可以被视为周期性或准周期性 [3] [4]，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 [5] [6] [7]。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 [8]。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 [9] [10]。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 [11] [12]。Kajita等人 [13] 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 [14] 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 [15] 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 [16] [17] [18]。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 [19] 在四足机器人上实现了多种步态及平滑过渡。García等人 [20] 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 [21] [22] [23] [24]。

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 [24] [25]，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。

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 [25] 和后空翻 [26] 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 [27] [28]，从而缩短了开发周期 [29]。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 [29]。强化学习更高效地生成了周期性步态。Wu等人 [30] 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 [31] 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。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。

综上而言，在人形机器人领域，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。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强化学习的自适应步态选择算法。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，在此基础上训练步态库并实现自适应时变步态顺滑切换。

1. **研究内容**

在第一章中已经提到，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。

**2.1 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[27][28]。Isaac Gym[111]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。Proximal Policy Optimization (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。

**2.2 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[222]、决策树[333]以及深度学习[444]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。

**2.3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。

1. **方案设计**

**3.1 Isaac Gym仿真环境搭建**

在进行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是至关重要的。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，即一个包含不同地面摩擦、障碍物和高度变化的环境，用以模拟真实世界中可能遇到的复杂情况。此外，仿真环境中的物理引擎参数也需要根据控制需求进行调整，如重力、摩擦力等因素。我考虑搭建粗糙平面、粗糙斜坡、粗糙崎岖、光滑平面等多种训练环境。

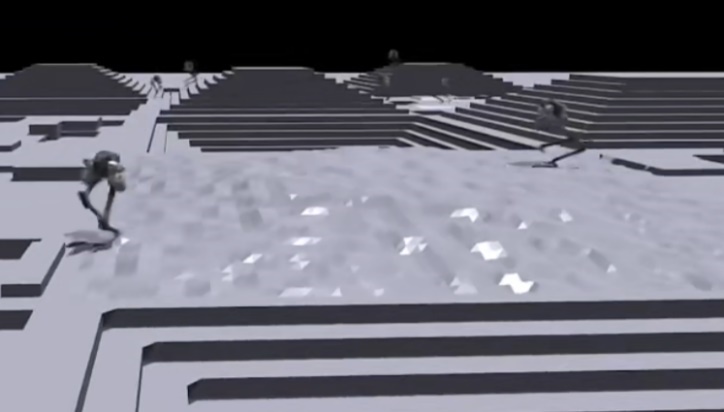
同时，Isaac Gym能够实时获取仿真环境中的传感器数据，如机器人的关节角度、速度、加速度、姿态、位置等信息，并将控制算法的输出指令传递给仿真环境。因此我将调配Isaac Gym的传感器接口，以便用于作为状态和观测接入强化学习神经网络，同时将测试控制算法接口，以便于强化学习动作输出进行仿真。

图1 双足机器人在Isaac Gym中

**3.2 PPO强化学习算法部署与数据收集**

PPO的核心优势在于它通过使用“剪切”目标函数来限制策略的更新幅度，这样能够有效地避免由于过大更新导致的训练不稳定问题。这种方法在实践中表现出了较高的训练效率和较好的稳定性。在具体应用时，PPO算法的流程通常包括数据收集、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的在不同地形的步态。

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。对已训练好的模型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。

**3.3 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，我会采用交叉验证技术，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，通过优化模型结构和训练策略，期望能够实现一个可靠的步态分类规划器，能够准确地根据机器人所感知的本体数据，判断当前地形的类型，从而为机器人在复杂环境中的自适应行走提供支持。

**3.4 测试**

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，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于实体机器人上进行测试，以评估算法在现实环境中的表现，并根据实际情况进行调整和优化。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。

1. **日程计划与安排**

|  |  |
| --- | --- |
| **内容** | **时间安排** |
| 构建或寻找真实场景的图片数据集 | 2023.12.1-2023.12.15 |
| 测试并选取提取视觉信息的多模态感知模型 | 2023.12.16-2023.12.31 |
| 构建详细视觉信息—指令—操作步骤数据集 | 2024.1.1-2024.2.15 |
| 对选定的预训练大模型进行微调 | 2024.2.16-2024.4.21 |
| 对训练得到的专用大模型进行测试、评估、比较 | 2024.4.22-2024.5.15 |
| 论文撰写 | 2024.5.16-2024.5.30 |

**参考文献**

[1] Brown T, Mann B, Ryder N, et al. Language models are few-shot learners[J]. Advances in neural information processing systems, 2020, 33: 1877-1901.

[2] Touvron H, Lavril T, Izacard G, et al. Llama: Open and efficient foundation language models[J]. arXiv preprint arXiv:2302.13971, 2023.

[3] Ouyang L, Wu J, Jiang X, et al. Training language models to follow instructions with human feedback[J]. Advances in Neural Information Processing Systems, 2022, 35: 27730-27744.

[4] Taori, R., Gulrajani, I., Zhang, T., Dubois, Y., Li, X., Guestrin, C., Liang, P., Hashimoto, T.B.:

Stanford alpaca: An instruction-following llama model. <https://github.com/tatsu-lab/stanford_alpaca>, 2023.

[5] Zhang S, Roller S, Goyal N, et al. Opt: Open pre-trained transformer language models[J]. arXiv preprint arXiv:2205.01068, 2022.

[6] Zhang R, Han J, Zhou A, et al. Llama-adapter: Efficient fine-tuning of language models with zero-init attention[J]. arXiv preprint arXiv:2303.16199, 2023.

[7] Zeng A, Liu X, Du Z, et al. GLM-130B: An Open Bilingual Pre-trained Model[C]//The Eleventh International Conference on Learning Representations. 2022.

[8] Chatgpt. https://chat.openai.com (2023)

[9] Hu Q, Li K, Zhao X, et al. InstructCoder: Empowering Language Models for Code Editing[J]. arXiv preprint arXiv:2310.20329, 2023.

[10] Zheng Q, Xia X, Zou X, et al. Codegeex: A pre-trained model for code generation with multilingual evaluations on humaneval-x[J]. arXiv preprint arXiv:2303.17568, 2023.

[11] Nijkamp E, Pang B, Hayashi H, et al. CodeGen: An Open Large Language Model for Code with Multi-Turn Program Synthesis[C]//The Eleventh International Conference on Learning Representations. 2022.

[12] Saidi H, Jha S, Sahai T. math-PVS: A Large Language Model Framework to Map Scientific Publications to PVS Theories[J]. arXiv preprint arXiv:2310.17064, 2023.

[13] Polu S, Han J M, Zheng K, et al. Formal Mathematics Statement Curriculum Learning[C]//The Eleventh International Conference on Learning Representations. 2022.

[14] Lewkowycz A, Andreassen A, Dohan D, et al. Solving quantitative reasoning problems with language models[J]. Advances in Neural Information Processing Systems, 2022, 35: 3843-3857.

[15] Vemprala S, Bonatti R, Bucker A, et al. Chatgpt for robotics: Design principles and model abilities[J]. Microsoft Auton. Syst. Robot. Res, 2023, 2: 20.

[16] Li C, Wong C, Zhang S, et al. Llava-med: Training a large language-and-vision assistant for biomedicine in one day[J]. arXiv preprint arXiv:2306.00890, 2023.

[17] Wu J, Antonova R, Kan A, et al. Tidybot: Personalized robot assistance with large language models[J]. arXiv preprint arXiv:2305.05658, 2023.

[18] Bender E M, Koller A. Climbing towards NLU: On meaning, form, and understanding in the age of data[C]//Proceedings of the 58th annual meeting of the association for computational linguistics. 2020: 5185-5198.

[19] Wu Z, Wang Z, Xu X, et al. Embodied task planning with large language models[J]. arXiv preprint arXiv:2307.01848, 2023.

[20] Jang E, Irpan A, Khansari M, et al. Bc-z: Zero-shot task generalization with robotic imitation learning[C]//Conference on Robot Learning. PMLR, 2022: 991-1002.