

钱思远

北京, 中国 [✉ qiansiyuan@stu.pku.edu.cn](mailto:qiansiyuan@stu.pku.edu.cn) [☎ 135-0139-3580](tel:135-0139-3580)

[🔍 SiriYep](#) [📖 Google Scholar](#)



个人简介

北京大学计算机学院 HMI 实验室博士生, 导师为仇尚航教授。研究聚焦具身智能 (Embodied AI), 围绕 VLA、World Model 与 VLM, 探索其在移动操纵等真实机器人场景中的建模、推理与泛化能力。在 NeurIPS、RSS 以及 Nature Computational Science 等顶级国际会议与期刊发表论文, 并担任 NeurIPS、ECCV 等会议审稿人。

教育经历

- | | | |
|-----------|--|-------------------|
| 博士 | 北京大学, 计算机学院, 计算机应用技术专业
研究方向: 具身智能场景下的多模态感知融合与自适应行动决策
关注领域: Embodied AI (VLA / World Model / VLM) | 2022.09 – 2027.06 |
| 本科 | 北京航空航天大学, 电子信息工程学院, 电子信息工程专业
荣誉奖项: 信息大类本科生最高荣誉 (大类前 0.3%)、学院本科生最高荣誉 (学院前 3%)、校级优秀毕业生 (专业前 10%) | 2018.09 – 2022.06 |

实习经历

- | | |
|--|-------------------|
| 北京智源人工智能研究院 , 具身多模态大模型研究中心, 研究实习生 | 北京, 中国 |
| <ul style="list-style-type: none">参与构建开源项目 RoboBrain 1.0 与 2.0, 针对性提升 LLM 与 VLM 的空间理解、时序建模、长链推理等能力。参与提出端到端 VLA 机器人移动操作方案 AC-DiT, 在多个 Benchmark 上达到 SOTA。负责松灵 Mobile ALOHA 双臂机器人与睿尔曼复合移动升降单臂机器人的硬件调试、真机数据采集、真机任务训练与测试, 熟悉多种移动操纵机器人的 I/O 接口与程序架构。基于多种机器人模拟器构建移动操作数据集与 Benchmark, 并设计支持 VR 与空间鼠标的遥操作数据采集方案。主导构建了基于不同模拟器的机器人数字资产, 包括 100+ 物体与 10+ 场景。 | 2024.08 – 2025.08 |
| 至简动力 (初创企业), 研究实习生 | 北京, 中国 |
| <ul style="list-style-type: none">面向真实移动操纵场景, 推进视觉-语言-动作模型 (VLA) 在机器人系统中的落地, 覆盖从任务理解、动作生成到执行反馈的端到端流程。围绕移动操纵任务, 系统性优化机器人全身控制与移动-操纵耦合建模, 提升复杂运动状态下动作生成的稳定性与一致性。引入 World Model 相关建模思想, 强化模型对长程任务与动态环境的建模能力, 在保证推理效率的同时有效缓解长时序执行中的误差累积问题。融合多模态感知信息 (视觉、空间结构、本体状态等) 进行感知强化, 提升机器人在复杂环境与移动状态下的三维感知与交互鲁棒性。熟练掌握星海图 Galaxea R1 Lite 与 Franka Emika Panda 机器人的部署与操作, 能够独立完成从真机数据采集、清洗、模型训练到测试验证的完整工作流程。熟悉基于 3D 打印 的机器人零部件设计与制作流程, 支持快速结构迭代与真实系统调试。 | 2025.09 – 至今 |

研究经历

AC-DiT: Adaptive Coordination Diffusion Transformer for Mobile Manipulation (Co-first author) **NeurIPS 2025**

关联方向：Embodied AI (Mobile Manipulation), VLA

作为论文的核心作者之一，提出并实现端到端移动操纵的视觉-语言-动作模型框架，设计两阶段动作生成机制（粗预测 + 扩散细化），并主导核心模型实现及其在仿真与真实机器人系统中的系统化评测，在多个 Benchmark 上与真机实验中取得显著优于现有方法的性能表现。

RoboMIND: Benchmark on Multi-embodiment Intelligence Normative Data for Robot Manipulation **RSS 2025**

关联方向：Embodied AI

参与提出 RoboMIND 多形态机器人遥操作数据集，涵盖 10.7 万演示轨迹、479 个任务、96 类物体及 4 种机器人形态，并包含失败案例与数字孪生环境。实验验证其显著提升 VLA 模型成功率与泛化能力，是当前规模最大、质量最高的同类数据集之一。

Chain of Thought Prompt Tuning in Vision Language Models

关联方向：VLM

提出思维链提示微调方法，将 CoT 引入视觉语言模型，联合利用视觉与文本嵌入信息，显著提升图像分类任务的泛化与迁移能力，并在图文检索与视觉问答任务中展现更强的推理性能，实现该方法在视觉任务中的首次成功应用。

Implicit Neural Image Field for Biological Microscopy Image Compression

关联方向：AI for Science

提出基于隐式神经表示 (INR) 的自适应压缩流程，支持任意形状图像与像素级解压，实现可控高压缩比（最高 512 倍），并在多种真实显微图像中验证其有效性，在显著减轻存储与共享压力的同时保留关键分析信息。

**Nature
Computational
Science**

开源项目

CAMEL: Communicative Agents for "Mind" Exploration of Large Language Model Society **GitHub · 15k ★**

<https://github.com/camel-ai/camel>

项目定位：多智能体 (Agent Society) 研究与应用框架，强调 Role-Playing、工具集成与任务自动化

基于 CAMEL 框架开展具身场景下的 Agent 任务实现与系统化评测：设计面向具身任务的 Agent Tools (工具调用) 与 Prompt 方案，支持 **Zero-shot** (仅基于任务文本描述) 与 **One-shot** (单条示教：视频 + 文本, Demonstration-conditioned / In-context imitation) 两种任务执行设置。在 RoboTwin Benchmark 上，所构建的 Agent 系统在未见任务的 Zero-shot 条件下展现出显著优于现有代表性 VLA 模型的任务完成能力，并在 One-shot 示教条件下进一步提升成功率，验证了 Agent-centric 方法在具身任务泛化与快速适应方面的有效性。

AC-DiT: Adaptive Coordination Diffusion Transformer

GitHub · 29 ★

<https://github.com/PKU-HMI-Lab/AC-DiT>

项目定位：NeurIPS 2025 论文官方代码与可复现训练/评测框架

负责核心模块实现与工程化，包括多模态输入处理、两阶段动作生成（粗预测 + 扩散细化）以及仿真/真机评测脚本与配置。

<https://github.com/leofan90/Awesome-World-Models>

项目定位：World Model 方向的研究型综述与资源索引（Survey & Knowledge Base）

参与维护并持续扩展 World Model 相关研究资源，系统性整理涵盖视频预测、动作条件生成 (action-conditioned dynamics)、具身智能与强化学习等方向的代表性工作。该项目为 World Model 与 Embodied AI 社区提供结构化研究入口，被广泛用于论文调研、方法对比与研究方向梳理，具有较高的学术参考价值。

专业技能

- **研究与系统落地能力：**能够将方法型研究工作完整落地至真实机器人系统，覆盖数据采集、数据清洗、模型训练、评测与部署的闭环流程。
- **长程任务建模与系统鲁棒性：**具备处理长时序执行、误差累积与动态不确定性的经验，能够在模型与系统层面对鲁棒性与推理效率进行针对性设计与优化。
- **跨平台迁移与复现：**能够在不同机器人平台与仿真环境中快速复现并扩展研究方法，支持从仿真到真实系统的迁移与对比评测。
- **工程与研究工具链：**熟练使用 PyTorch、ROS/ROS2 及相关仿真工具，具备复杂研究代码库的工程化实现、复现与系统化评测能力。

专利与软件著作权

参与发明专利

- 一种特定目标的多摄像机联动实时分析方法
- 一种地面摄像机固定监控和天空移动无人机联动分析方法

参与软件著作权

- 天地一体化视频大数据离线检索与轨迹回放系统
- 天地一体化智能人车目标的实时跟踪与离线跟踪方法
- 天地一体化智能监控平台

荣誉奖项

校级三好学生、校级优秀团员、学科竞赛奖学金特等奖、学习优秀奖学金一等奖、创新创业奖学金一等奖、学生工作奖学金二等奖

其他

学生工作经历：曾担任北京大学视频与视觉技术国家工程研究中心班长、团支书；曾担任北京航空航天大学党委宣传部学生记者。

社团活动经历：北京航空航天大学电子信息工程学院羽毛球队队长、华为花粉俱乐部部长、电子科协教学部部长。

个人兴趣爱好：羽毛球、摄影。