

金雨润

yurunjin@mail.ustc.edu.cn +86 190 5549 9020

在 LinS Lab 与 LightOrigins 实习，方向 Agentic for robotics。

探索从感知到行动的智能。我关注世界模型如何理解物理世界，以及机器人如何在其中学习与行动。

教育背景

中国科学技术大学

2023.09 至今

计算机科学与技术学院 · 计算机科学与技术专业 · 本科

研究兴趣

世界模型与机器人学习

- Action Condition 注入与具身世界模型：关注如何将机械臂运动学、末端执行器几何、历史观测、交互事件等 Action Condition，以图像域中间表示、latent conditioning 或 attention modulation 等方式注入生成式 World Model，从而提升未来状态预测、动作生成与闭环控制的一致性。
- WAM 与 AC-WM 范式互促：探索 WAM 与 AC-WM 两类世界模型范式在动作理解、环境预测和策略生成上的互补关系：一方面利用 action-centric 表征增强 World Model 的可控预测能力，另一方面将 World Model 的时空预测与状态转化能力反馈给动作生成、动作评估和策略优化。
- 物理知识增强的具身智能：后续希望研究如何让 World Model 更好理解接触、遮挡、几何约束、物体动力学、因果状态转化等物理知识，并将其转化为机器人长程规划、精细操作、跨场景泛化和真机部署稳定性的提升。

研究经历

STARRY: Spatial-Temporal Action-Centric World Modeling for Robotic Manipulation

2026.02 — 至今

(二作 · NeurIPS 在投) 中关村学院 具身智能研究组 (导师: 陈凯)

- Tech: Python, PyTorch, Wan 2.2, Motus, RoboTwin 2.0, Depth Anything V3, 方舟无限 R5, SAPIEN
- 论文创新及效果简介: STARRY 是一个时空 Action-Centric 的世界模型增强动作生成策略，联合去噪未来时空 latent 与 action 序列，并提出 Geometry-Aware Selective Attention Modulation (GASAM)，将预测深度与末端执行器几何转化为 token 对齐权重，对 action-attention 进行选择性地调制；在 RoboTwin 2.0 上 Clean / Randomized 平均成功率达 93.82% / 93.30%，真机平均成功率从 42.5% 提升至 70.8% (对标 $\pi_{0.5}$)。
- 个人贡献: 基于 Motus + Wan 2.2 为 STARRY 的 joint future-state 与 action-generation 设定完成模型实现，并搭建 RoboTwin 2.0 benchmark 实验的训练与评测 pipeline；在 方舟无限 R5 双臂 真机上完成真机验证，覆盖数据采集、处理、模型部署与真机条件评测。

EA-WM: Event-Aware Generative World Model with KVAFs (二作 · NeurIPS 在投)

2026.02 — 至今

中关村学院 具身智能研究组 (导师: 陈凯)

- Tech: Python, PyTorch, Wan 2.2, DiT, WorldArena
- 论文创新及效果简介: EA-WM 是基于 Wan 2.2 骨干的事件感知生成式世界模型，提出 Structured Kinematic-to-Visual Action Fields (KVAFs)，将机械臂运动学动作直接投影到目标相机视角，与 RGB 同域送入世界模型，既保留运动学拓扑又避免修改骨干；并引入 event-aware 双向融合模块捕捉物体状态变化与交互动力学；在 WorldArena 官方评测上取得 Overall 维度 SOTA，位列公开 Leaderboard 第 2 (清华 FIB-Lab 世界模型评测基准)。
- 个人贡献: 搭建 EA-WM 项目的原始训练 / 评测 pipeline，覆盖数据处理、训练脚本与 WorldArena 评测闭环；围绕 Action Condition 的图像域中间表示与注入方式协助开展 总计 50+ 次架构组合实验；持续在 WorldArena Benchmark 上做调参与消融，已取得榜一成绩。

PhysBrain — End-to-end Development & Evaluation of Embodied AI Foundation Model

2025.09 — 2025.12

中关村学院 具身智能研究组 (指导老师: 陈凯)

- Tech: Python, PyTorch, VLM/VLA, World Models, SimplerEnv, StarVLA
- 论文创新及效果简介: PhysBrain 是面向具身控制的多模态基础模型，针对大模型向机器人端到端控制迁移中的视角鸿沟 (egocentric $\rightarrow$ allocentric) 与具身监督信号缺失问题，将手物交互位姿估计、世界模型与 VLA 统一到同一训练框架，使 egocentric 数据可直接转化为机器人端到端控制的监督信号；在多个长程规划与接触式交互基准上展现出更强的物理先验与状态转化推理能力。
- 个人贡献: 深入调研手物交互位姿估计、多模态大模型与世界模型领域的前沿文献并输出技术路线综述；依托 HPC 集群完成 PhysBrain 微调模型在多个主流基准上的长程规划、状态转化推理与接触式交互能力评测。

中关村学院 具身智能研究组（导师：陈凯）

- Tech: Python, PyTorch, VLA, RoboTwin2, AliasBench, MemoryVLA, SimplierEnv, LIBERO, RoboCasa
- 论文创新及效果简介： IntentVLA 是面向观测混叠场景的历史条件化 VLA 框架，将近期视觉观测序列编码为短时意图表征，并作为 action chunk 生成条件，缓解逐帧条件化策略在部分可观测环境中的意图混淆、chunk 冲突和执行不稳定问题；同时提出 AliasBench，基于 RoboTwin2 构建 12 个混叠感知任务及配套训练和评估环境，用于隔离和量化短时观测混叠现象；在 AliasBench, SimplierEnv, LIBERO 和 RoboCasa 四个基准上，IntentVLA 的 rollout 稳定性和任务成功率均优于多个 VLA 基线。
- 个人贡献： 在 AliasBench 的 12 个任务上完成 MemoryVLA 的微调与效果测评，为论文提供了 memory-centric 类型基线的对比数据。

项目经历

星载多谱段宽视场大气瞬态发光事件监测相机设计 · 中国科学技术大学大创2025.09 — 2026.06

GCINE3243, IMX455, Vitis HLS, FPGA, ISP, 8K Imaging

- 面向 MERLIS 微纳卫星平台与 TLEs 广域遥感观测，参与多谱段宽视场相机载荷原理样机开发；基于 GCINE3243 和索尼 IMX455 平台完成图像采集、数据读出、图像处理和视频压缩链路验证
- 基于 Vitis HLS 与 Vitis Vision 设计 8K RAW 数据 ISP 处理流水线，样机已经实现 8K 图像拍摄，并为后续集成 ISP 标定、事件检测、目标识别和 ROI 提取等星上智能处理功能提供基础

自瞄系统研发与优化 · RoboMaster 竞赛2024.11 — 2025.08

C++, ROS2, OpenCV, ML

- 基于 ROS2 实现的高性能自动化目标识别与跟踪控制系统
- 使用先验的灯条几何属性识别装甲板，使用 CNN 识别装甲板数字
- 提取灯条关键点进行 PnP + 优化 Kalman 滤波来估计目标的 3D 姿态
- MPC 控制器实现对目标的预测性跟踪与弹道补偿
- 可对至少 5m 外的、转速至少为 10rad/s 的目标进行稳健识别，实体弹丸命中率超过 50%

自动兑矿系统研发 · RoboMaster 竞赛2024.11 — 2025.08

C++, Python, ROS2, OpenCV, DL

- 基于 ROS2 实现的自动化末端靶向对位与作业系统
- 七自由度机械臂的逆运动学求解与空间轨迹规划
- 基于双目/RGB-D 视觉的目标检测与 6D 位姿估计，复杂环境下的高鲁棒性靶点特征提取与空间定位
- 对于赛事中最高难度矿石兑换任务，成功率超过 80%

Mobile Using Agent 研发 · PineAI Agent Development Intern2025.05 — 2025.09

Python, LLM, ReAct, Android World, Vector DB

- 基于 ReAct 理念与 Android World 开源框架的大模型推理与 UI 交互闭环系统
- 设计并实现反思机制，提取任务执行过程中的成功动作序列并转化为结构化的标准操作步骤存入技能库
- 引入基于语义相似度的检索增强策略，使 Agent 在面对相似任务时能自主召回历史经验，绕过冗长的从零推理过程，提升了任务完成效率和成功率
- 利用 Android World 开源框架，使大模型能够精准理解 Android UI 无障碍模式视图树并输出确定性交互指令（点击、滑动、输入等），提升了复杂长步数任务的执行成功率

Agent Ability BenchMark · PineAI Agent Development Intern2025.05 — 2025.09

Python, Golang, LLM, Docker, Kafka, MySQL

- 基于 Kafka 异步会话管理与分布式 Docker 容器环境构建的交互式 Agent 能力评估系统
- 设计多维度动态博弈测试用例，通过配置 Mock Agent 模拟复杂用户意图与极端边缘场景，交互式的全面检验被测智能体的上下文理解、工具调用及纠错鲁棒性

获奖经历

RoboMaster 2025 联盟赛冠军 / 区域赛冠军 / 全国赛亚军2025.08

Robogame 2024 校一等奖2024.10

校园经历

主持人俱乐部 副社长 / Vice President, Hosts Club2025.09 — 至今

摄影协会 副社长 / Vice President, Photography Association2024.09 — 至今