

翟明亮

✉ zmling22@gmail.com

☎ (+86) 177-0812-1099



🎓 教育背景

北京理工大学, 博士, 计算机科学与技术

2023 年 09 月 – 至今

北京理工大学, 硕士, 软件工程专业

2021 年 09 月 – 2023 年 7 月

四川师范大学, 学士, 计算机科学与技术专业

2014 年 09 月 – 2018 年 06 月

🔬 科研项目经历

动态三维长时序一致性保持的视频世界模型

2026 年 02 月 – 至今

基于 3D 高斯显式表示构建视频世界模型, 通过对场景进行连续空间建模, 实现高效渲染与动态表达。利用迭代式生成, 支持相机控制和显存优化, 结合长时序数据学习, 模型能够保持结构与运动的一致性, 提升复杂动态场景的重建与生成能力。可应用于具身机器人的 Imagined Planing 以实现最优规划。

具身问答模型学习多步推理能力

2025 年 03 月 – 2025 年 09 月

具身智能体在探索过程中由于缺乏显式推理能力, 往往难以有效收集关键证据, 从而影响最终决策与问答的准确性。为此, 我们提出一种面向具身问答任务的探索式推理学习方法, 使模型能够在交互过程中主动进行推理与信息整合。具体而言, 我们构建了一套自动化数据生成管线, 为探索轨迹引入细粒度的推理标注; 同时设计了「规划器-控制器-执行器」的闭环架构, 通过生成数据对控制器进行迭代训练, 使模型在探索过程中具备更强的推理能力, 并能够借助工具高效获取与利用信息, 显著提升复杂环境下的任务表现。

以记忆为中心的具身问答

2024 年 09 月 – 2025 年 02 月

低效的记忆构建和不加区分的检索使得模块化具身问答方法难以处理涉及多目标多区域的复杂任务。我们以记忆为系统核心设计了 MemoryEQA 方法, 构建多模态结构化记忆表示, 利用视点对比更新机制和模块感知的检索策略来为不通模块提供精准的推理证据。实现了在复杂任务上的高效探索和准确回答。

自动驾驶场景感知受限区域的世界知识增强推理

2024 年 03 月 – 2024 年 07 月

当自动驾驶车辆面对感知受限的区域 (动态或静态遮挡区域) 时, 模型很难有效地将感知能力与世界知识相结合进行推理。我们提出了一种即插即用的指令引导交互模块, 该模块弥合了模态差距, 显著缩短了输入序列长度, 使其能够有效地适应多视图视频输入。

Fast-StrucTexT: 一种用于文档理解的沙漏型 Transformer 架构 2022 年 07 月 – 2023 年 01 月

Transformer 架构依赖于与 Token 长度相关的二次复杂度的注意力机制, 在面对长上下文的场景时通常表现出较高的推理延迟。并且通用的高效 Transformer 模型难以直接应用于文档建模。所以我们提出了 Fast-StrucTexT, 利用模态引导的 Token Merge 机制来剪枝冗余 Token, 并结合上采样机制形成沙漏型 Transformer 架构, 此外, 为了利用多模态信息指导 Token 合并, 我们还提出了 Symmetry Cross Attention 机制。我们的方法能够在保持甚至提升推理精度的同时提升 1.9 倍模型吞吐量。

🔧 工程项目

四足机器人自动巡检系统

2025 年 07 月 – 2026 年 05 月

基于 Unitree Go2 四足机器人完成自定义巡检任务。集成 SLAM、导航、语义地图、多模态大模型和 Dashboard 服务等。核心功能是通过前端界面发布自然语言巡检指令, 自动规划路径完成巡检并生成巡检报告; 依赖功能是 object-goal 导航, 实现了室外区域的无图导航寻找目标。

📁 实习经历

算法实习生 长安汽车

2024 年 03 月 – 2024 年 7 月

开发基于世界知识的自动驾驶大模型，理解多视角图像，设计推理管线，输出当前条件下的最优驾驶策略。

算法实习生 百度公司

2022 年 06 月 – 2024 年 3 月

开发一种基于 StrucText 模型的小型化模型，服务于百度 OCR 开放平台。分析延迟瓶颈，针对性设计解决方案，QPS 提升 2 倍。

📄 学术成果

- **Mingliang Zhai**, Hansheng Liang, Xiaomeng Fan, Zhi Gao, Chuanhao Li, Che Sun, Xu Bin, Yuwei Wu, Yunde Jia. Multi-Step Reasoning for Embodied Question Answering via Tool Augmentation. *arXiv preprint arXiv:2510.20310*, 2025
- **Mingliang Zhai**, Zhi Gao, Che Sun, Yuwei Wu, Yunde Jia. Memory-Centric Embodied Question Answering. *arXiv preprint arXiv:2505.13948*, 2025
- **Mingliang Zhai**, Cheng Li, Zengyuan Guo, Ningrui Yang, Xiameng Qin, Sanyuan Zhao, Junyu Han, Ji Tao, Yuwei Wu, Yunde Jia. World Knowledge-Enhanced Reasoning Using Instruction-guided Interactor in Autonomous Driving. *The Association for the Advancement of Artificial Intelligence (AAAI)*, 2025
- **Mingliang Zhai**, Yulin Li, Xiameng Qin, Yi Chen, Qunyi Xie, Chengquan Zhang, Kun Yao, Yuwei Wu, Yunde Jia. Fast-StrucText: An Efficient Hourglass Transformer with Modality-guided Dynamic Token Merge for Document Understanding. *International Joint Conference on Artificial Intelligence (IJCAI)*, 2023

⚙️ IT 技能

- 编程语言: Python, ROS/ROS2, \LaTeX
- 深度学习平台: Pytorch, PaddlePaddle