

罗昊 Hao Luo

北京大学计算机学院博士研究生, 25 岁

✉ lh2000@pku.edu.cn

🌐 <https://lighttet.github.io>

📖 <https://scholar.google.com/citations?user=TwuNaTYAAAAJ>



教育背景

北京大学, 计算机学院

2022 - 2027 (预计)

博士研究生; 研究方向: 具身智能, 多模态基础模型; GPA: 3.8/4.0; 导师: 卢宗青 (长聘副教授)

北京大学, 信息科学技术学院

2018 - 2022

理学学士, 计算机科学与技术, GPA: 3.6/4.0

实习经历

北京智在无界科技有限公司 (BeingBeyond)

2025.4 - 至今

研究型实习生, Being 自研大模型团队

- ♦ 研发 Being-H 系列 具身基础模型, 逐步将数千到数万小时的人类视频数据用于具身预训练。
- ♦ 研究可规模化的视觉语言动作模型与基于隐空间的世界模型, 均有论文产出。

北京智源人工智能研究院 (BAAI)

2023.1 - 2025.3

研究型实习生, 多模态交互组

- ♦ 研究可用于策略学习的视频预训练方法, 以及专注视频理解的视觉语言模型, 均有论文产出。

近期工作

Being-H0: Vision-Language-Action Pretraining from Large-Scale Human Videos

【共一第一】最早将大规模人类视频用于 VLA 训练的工作, 通过将人手视为通用操作器, 通过结构化人手运动 tokenization 序列化人类操作进行物理意义的指令微调, 实现人类操作先验到下游机器灵巧操作的迁移。

Being-H0.5: Scaling Human-Centric Robot Learning for Cross-Embodiment Generalization

【共一第一】将 Being-H0 中人类视频预训练扩展至 35,000+ 小时视频数据, 构建横跨人类与 30 种本体机器人统一预训练框架, 实现下游跨本体的策略泛化。

Rethinking Visual-Language-Action Model Scaling: Alignment, Mixture, and Regularization

【共同一作】通过系统性对照 VLA 中的动作对齐方式、异构数据混合与正则化策略, 探究 VLA 模型的扩展性与稳健性中各因素的影响。

具身智能与 VLA

Joint-Aligned Latent Action: Towards Scalable VLA Pretraining in the Wild

【CVPR 2026 · 独立一作】前序工作 JEPT 的延伸, 通过将 VLA 表征与逆动力学表征联合对齐学习潜在动作空间, 使 VLA 能够利用 500 万条人类视频无标注数据进行可扩展预训练, 改善大规模机器人标注数据稀缺的瓶颈。

Spatial-Aware VLA Pretraining through Visual-Physical Alignment from Human Videos

【CVPR 2026】Yicheng Feng, Wanpeng Zhang, Ye Wang, **Hao Luo**, Haoqi Yuan, Sipeng Zheng, Zongqing Lu†

OpenT2M: No-frill Motion Generation with Open-source, Large-scale, High-quality Data

【CVPR 2026】Bin Cao, Sipeng Zheng, **Hao Luo**, Boyuan Li, Jing Liu, Zongqing Lu†

基于视频表征的策略学习

Learning Video-Conditioned Policy on Unlabelled Data with Joint Embedding Predictive Transformer

【ICLR 2025 · 独立一作】受 JEPa 启发, 通过联合嵌入预测与逆动力学同时建模视觉转移和动作, 使无标注视频也能提供有效监督, 解决视频条件策略学习中动作标注数据稀缺的问题。

Pre-trained Visual Dynamics Representations for Efficient Policy Learning

【ECCV 2024 · 独立一作】通过视频预测任务从无标注视频中预训练抽象视觉动力学表征, 并通过在线强化学习适配到目标环境, 解决视觉策略学习中的样本效率问题。

Reinforcement Learning Friendly Vision-Language Model for Minecraft

【ECCV 2024】Haobin Jiang*, Junpeng Yue*, **Hao Luo**, Ziluo Ding, Zongqing Lu†

视觉语言模型

OpenMMEgo: Enhancing Egocentric Understanding for LMMs with Open Weights and Data

【NeurIPS 2025 · 独立一作】构建 820 万条第一人称视频问答数据集, 并引入语义感知视觉 token 压缩与课程学习策略, 改善大语言模型在视角快速变化的自我中心视频理解中的不足。

Unified Multimodal Understanding via Byte-Pair Visual Encoding

【ICCV 2025 Highlight】Wanpeng Zhang, Yicheng Feng, **Hao Luo**, Yijiang Li, Zihao Yue, Sipeng Zheng, Zongqing Lu†

VideoOrion: Tokenizing Object Dynamics in Videos

【ICCV 2025】Yicheng Feng*, Yijiang Li*, Wanpeng Zhang, Sipeng Zheng, **Hao Luo**, Zihao Yue, Zongqing Lu†

强化学习

A Survey on Transformers in Reinforcement Learning

【TMLR 2023】Wenzhe Li*, **Hao Luo***, Zichuan Lin*, Chongjie Zhang†, Zongqing Lu†, Deheng Ye†

荣誉

- ◆ 灵均领航奖学金, 北京大学 2025
- ◆ 三好学生, 北京大学 2024-2025
- ◆ 学业优秀奖, 北京大学 2019-2021
- ◆ 社会工作奖, 北京大学 2018-2019

服务

- ◆ 审稿人: ICML、NeurIPS (2026 年前)、ICLR、CVPR、ICCV、ECCV、BMVC
- ◆ 助教: 北京大学《算法设计与分析》(2021 春、2022 春);《深度强化学习》(2024 春)