

# Hao Luo

Ph.D. Student, School of Computer Science, Peking University

✉ [lh2000@pku.edu.cn](mailto:lh2000@pku.edu.cn)

🌐 <https://lighttet.github.io>

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

## EDUCATION

---

**Peking University, School of Computer Science** 2022 - 2027 expected

*Ph.D. Student in Computer Applied Technology; Research Focus: Embodied AI, Multimodal Foundation Models; GPA: 3.8/4.0; Advisor: Zongqing Lu*

**Peking University, School of Electronics Engineering and Computer Science** 2018 - 2022

*B.S. in Computer Science; GPA: 3.6/4.0*

## EXPERIENCES

---

**Beijing BeingBeyond Technology Co., Ltd.** 2025.4 - present

*Research Intern, Being Foundation Model Team*

- ♦ Developing the Being-H series of [embodied foundation models](#), progressively scaling embodied pretraining from thousands to tens of thousands of hours of human videos.
- ♦ Studying [scalable VLA](#) and [latent world-action models](#), both leading to papers.

**Beijing Academy of Artificial Intelligence (BAAI)** 2023.1 - 2025.3

*Research Intern, Multimodal Interaction Group*

- ♦ Studied [video pretraining](#) for policy learning and video understanding [VLM](#), both leading to papers.

## RECENT WORK

---

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

[\[Co-first author, first-listed\]](#) The earliest work to use large-scale human videos for VLA pre-training, treating human hands as a universal manipulator and serializing human motions through structured hand-motion tokenization, transferring human priors to downstream robotic dexterous manipulation.

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

[\[Co-first author, first-listed\]](#) Extends Being-H0 to **35,000+** hours of videos and builds a unified pretraining framework for human data and 30+ robot embodiments, enabling cross-embodiment policy in downstream tasks.

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

[\[Co-first author\]](#) Systematically studies action alignment, heterogeneous data mixing, and regularization in VLA scaling, clarifying how these factors affect scalability and robustness.

## Embodied AI / VLA

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

[CVPR 2026, First author] Developing JEPT, JALA scales VLA pretraining by aligning hidden states with inverse dynamics to learn unified latent actions from both labeled and unannotated videos, reducing data bottlenecks.

### 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†

## Video Representation for Policy Learning

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

[ICLR 2025, First author] Inspired by JEPA, JEPT learns predictive visual representations of task-relevant transitions through joint embedding prediction, enabling the earliest JEPA-style learning in policy learning.

### Pre-trained Visual Dynamics Representations for Efficient Policy Learning

[ECCV 2024, First author] PVDR learns abstract visual dynamics priors from unlabeled videos and adapts them for efficient policy learning in domain-gapped tasks.

### Reinforcement Learning Friendly Vision-Language Model for Minecraft

[ECCV 2024] Haobin Jiang\*, Junpeng Yue\*, **Hao Luo**, Ziluo Ding, Zongqing Lu†

## Vision-Language Models

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

[NeurIPS 2025, First author] Builds an 8.2M-sample egocentric video QA dataset and introduces semantic-aware visual token compression to improve VLMs in egocentric video understanding.

### 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†

## Reinforcement Learning

### A Survey on Transformers in Reinforcement Learning

[TMLR 2023] Wenzhe Li\*, **Hao Luo\***, Zichuan Lin\*, Chongjie Zhang†, Zongqing Lu†, Deheng Ye†

## HONORS

- ◆ Lingjun Pilot Scholarship, Peking University (PKU) 2025
- ◆ Merit Student, by Peking University (PKU) 2024-2025
- ◆ Academic Excellence Award, by Peking University (PKU) 2019-2021
- ◆ Social Work Award, by Peking University (PKU) 2018-2019

## SERVICES

- ◆ Reviewer: ICML, NeurIPS (before 2026), ICLR, CVPR, ICCV, ECCV, BMVC
- ◆ Teaching Assistant: Design and Analysis of Algorithms, PKU (Spring 2021, Spring 2022); Deep Reinforcement Learning, PKU (Spring 2024)