

Bohan Zhou

Tel: +86 15797895657 | Email: zhoubh@stu.pku.edu.cn

EDUCATION

Peking University, School of Computer Science

Master's in Computer Applied Technology (GPA: 3.88 / 4.0)

Beijing, China

Sep. 2023 – Jul. 2026

Nankai University

Bachelor's in Intelligent Science and Technology (GPA: 3.95 / 4.0, Rank: 2/98)

Tianjin, China

Sep. 2019 – Jul. 2023

PUBLICATIONS

- [1] Zhou B, Li K, Jiang J, et al. Learning from visual observation via offline pretrained state-to-go transformer. *NeurIPS 2023*.
- [2] Zhou B, Zhan Y, Zhang Z, et al. MEgoHand: Multimodal Egocentric Hand-Object Interaction Motion Generation. *NeurIPS 2025*.
- [3] Zhou B, Zhang Z, Wang J, et al. NOLO: Navigate Only Look Once. *IROS 2025 Oral*.
- [4] Zhou B, Yuan H, Fu Y, et al. Learning diverse bimanual dexterous manipulation skills from human demonstrations. (*Under Review*).
- [5] Yuan H, Zhou B, Fu Y, et al. Cross-embodiment dexterous grasping with reinforcement learning. *ICLR 2025*.
- [6] Zheng S, Zhou B, Feng Y, et al. Unicode: Learning a unified codebook for multimodal large language models. *ECCV 2024*.
- [7] Luo H, Zhou B, Lu Z. Pre-trained Visual Dynamics Representations for Efficient Policy Learning. *ECCV 2024*.
- [8] Tan W, Zhang W, Xu X, Xia H, Ding Z, Li B, Zhou B, et al. Cradle: Empowering foundation agents towards general computer control. *ICML 2025*.
- [9] Hu Z, Yang Y, Zhai X, Yang D, Zhou B, et al. GFIE: A Dataset and Baseline for Gaze-Following from 2D to 3D in Indoor Environments. *CVPR 2023*.
- [10] Hu Z, Zhao K, Zhou B, et al. Gaze target estimation inspired by interactive attention. *TCSVT 2022*.
- [11] Yuan H, Bai Y, Fu Y, Zhou B, et al. Being-0: A Humanoid Robotic Agent with Vision-Language Models and Modular Skills. *CORL 2025 (Under Review)*.

INTERNSHIP EXPERIENCE

Being-0: A Humanoid Robotic Agent with Vision-Language Models and Modular Skills

Beijing Academy of Artificial Intelligence, Embodied AI

Oct. 2024 – Feb. 2025

- **A hierarchical robot agent framework** for efficient long-horizon humanoid robot control: **Top-level** black-box large multi-modal model (LMM, like GPT-4v) for task understanding + decomposition; **Mid-level** finetuned LMM for navigation planning; **Low-level** skill libraries for dexterous manipulation.
- **A VLM + rule based hybrid connector**: bridged high-level language plans with low-level motor skills, enabling coordinated locomotion, navigation and dexterous manipulation.

Cradle: Empowering Foundation Agents towards General Computer Control

Beijing Academy of Artificial Intelligence, Generalist Agents

Oct. 2023 – Feb. 2024

- **General computer control**: Pioneered a universal interface for agents to interact with any software using screenshots as input and keyboard/mouse actions as output, standardizing cross-environment interaction.
- **A LMM-powered Cradle framework**: integrated six core modules (Information Gathering, Self-Reflection, Task Inference, Skill Curation, Action Planning, and Memory) to automate task planning and execution via generating keyboard/mouse codes without built-in APIs.
- **Robust validation**: adaptable across **4 games** (Red Dead Redemption 2, Cities: Skylines, Stardew Valley and Dealer's Life 2), **5 software tools** (Chrome, Outlook, Feishu, Meitu and CapCut), and **OSWorld**.

RESEARCH EXPERIENCE

MEgoHand: Multimodal Egocentric Hand-Object Interaction Motion Generation

BeingBeyond, Embodied AI

Mar. 2025 – May. 2025

- **Large-Scale Dataset:** Curated **3.35M** hand-object interaction samples, **1.2K** objects, **24K** trajectories (10-15x prior datasets) via proposed inverse MANO retargeting and virtual RGB-D rendering.
- **Multimodal Perception & Spatial Understanding:** Based on **Eagle-2**, MEgoHand additionally incorporates **Unidepth-v2** for metric depth estimation.
- **Smooth Motion Generation:** **DiT**-based motion generator supervises future hand motion via Flow Matching. **Temporal orthogonal filtering** is designed to decode smooth pose sequence.

BiDexHD: Learning Diverse Bimanual Dexterous Manipulation Skills from Human Demonstrations

Beijing Academy of Artificial Intelligence, Embodied AI

Jun. 2024 – Sep. 2024

- **BiDexHD Framework:** Developed a unified framework for learning diverse bimanual skills from human demonstrations, unifying **task construction from HOI datasets** and **teacher-student policy learning** for scalable vision-based bimanual dexterous skills. To avoid task-specific reward engineering, a two-stage reward function is generally designed.
- **Zero-shot adaptability:** Achieved **74.59%** task fulfillment on trained tasks and **51.07%** on unseen tasks in the TACO benchmark (141 tasks), with **80.49%/65.99%** on ARCTIC.

NOLO: Navigate Only Look Once

Beijing Academy of Artificial Intelligence, Embodied AI

Feb. 2024 – May. 2024

- **Video Navigation:** Introduced a novel task requiring agents to finish image navigation using only a single **30-second** context video and real-time egocentric images.
- **NOLO Method:** Enhanced offline reinforcement learning by integrating **GMflow** via pseudo-action labeling and a **temporal coherence** loss.
- **Simulation and Real World Evaluation:** Demonstrated success in RoboTHOR and Habitat simulation and validated **real-world deployment on a Unitree Go2 robot** in a maze environment.

STG: Learning from Visual Observation via Offline Pretrained State-to-Go Transformer

Beijing Academy of Artificial Intelligence, Generalist Agents

Sep. 2022 – May. 2023

- **Two-stage framework to learn from pixels:** upstream offline pretrained State-to-Go Transformer on visual observations to guide **reward-free** online reinforcement learning downstream.
- **Joint Representation Learning:** Co-optimized a discriminator and temporal distance regressor in an adversarial manner to align latent embeddings temporally.

Human Intent Analysis in Indoor Environments for Service Robots

Tianjin Key Laboratory of Intelligent Robotics

Advisor: Prof. Jingtai Liu

Mar. 2021 – Aug. 2022

- **Human Intent Analysis Pipeline:** Involved semi-automatic dataset construction and gaze direction/target estimation for human intention understanding and forecasting, achieving **3rd prize**.
- **GFIE [CVPR 2023]:** Created multi-sensor gaze data collection system (Kinect + laser rangefinder) with novel algorithm for unbiased 2D/3D gaze target annotation via laser spot localization.
- **VSG-IA [TCSVT 2022]:** Proposed graph attention network for automatic gaze behavior detection and human-scene interaction analysis.

Awards & Honors

National Scholarship, Triple-A Outstanding Student Award

Peking University, China

Sep. 2025

Outstanding Graduates in Nankai University

Nankai University, China

Jun. 2023

Tianjin Municipal People's Government Scholarship (Top 2%)

Tianjin, China

Nov. 2021

2nd Prize, National College Student Mathematical Modeling Competition

Chinese Society for Industrial and Applied Mathematics

Oct. 2021

Honorable Mention, American College Mathematical Contest in Modeling

COMAP

May. 2021