Hao Shi (石昊)

2002-05 | M.S. in Embodied AI, Tsinghua University
shihao1895@gmail.com | 13966968553 (Phone, WeChat)
https://shihao1895.github.io

Education

Tsinghua University 2023.09 - 2026.06
Artificial Intelligence, Master | Department of Automation Beijing

- Research Interests: Embodied AI, Robot Learning, VLA, Visual Perception
- Advisors: Gao Huang (黄高), Xiangyu Zhang (张祥雨)
- GPA: 3.8 / 4.0
- Honors: THU Comprehensive Excellence 1st Prize Scholarship (Top 10%, awarded twice).

Tianjin University 985 2019.09 - 2023.06
Computer Science, Bachelor | College of Intelligence and Computing Tianjin

- GPA: 91.2 / 100, 3.81 / 4.0 | Rank: Top 6%
- Academic advisor: Di Lin (林迪)
- Honors: Outstanding Bachelor's Thesis Award, Excellent Graduate Award, Distinguished Student.

Research

1. [Robot Learning] MemoryVLA: Perceptual-Cognitive Memory in Vision-Language-Action Models for Robotic Manipulation, 2025

- ***Hao Shi****, Bin Xie, Yingfei Liu, Lin Sun, Fengrong Liu, Tiancai Wang, Erjin Zhou, Haoqiang Fan, Xiangyu Zhang, Gao Huang*✉
- MemoryVLA is a Cognition-Memory-Action framework for robotic manipulation inspired by human memory systems. It builds a hippocampal-like perceptual-cognitive memory to capture the temporal dependencies essential for current decision-making, enabling long-horizon, temporally aware action generation. [Paper]
- Invited talk @ 具身智能之心 & 3D视觉工坊

2. [Embodied Perception] DenseGrounding: Improving Dense Language-Vision Semantics for Ego-centric 3D Visual Grounding, ICLR 2025 & CVPR 2024 Workshop, Oral

- *Henry Zheng*,* ***Hao Shi*(equal contrib.)****, Qihang Peng, Yong Xien Chng, Rui Huang, Yepeng Weng, Zhongchao Shi, Gao Huang*✉
- DenseGrounding is a framework for embodied 3D visual grounding. It tackles the loss of fine-grained visual semantics from sparse fusion of point clouds with multi-view images, as well as the limited textual semantic context from arbitrary instructions, enabling more accurate and context-aware grounding. [Paper]
- **1st Place & Innovation Award** in CVPR 2024 Autonomous Grand Challenge (1/154).
- Oral Presentation at CVPR 2024 Workshop on Foundation Models for Autonomous Systems.

3. [Robot Learning] SpatialActor: Exploring Disentangled Spatial Representations for Robust Robotic Manipulation, 2025

- ***Hao Shi****, Bin Xie, Yingfei Liu, Yang Yue, Tiancai Wang, Haoqiang Fan, Xiangyu Zhang, Gao Huang*✉
- SpatialActor is a disentangled framework for robust robotic manipulation. It decouples perception into complementary high-level geometry from fine-grained but noisy raw depth and coarse but robust depth expert priors, along with low-level spatial cues and appearance semantics. [Paper]

4. [Visual Perception] Open Compound Domain Adaptation with Object Style Compensation for Semantic Segmentation, NeurIPS 2023

- *Tingliang Feng*,* ***Hao Shi*(equal contrib.)****, Xueyang Liu, Wei Feng, Liang Wan, Yanlin Zhou, Di Lin*✉
- We propose object style compensation for open compound domain adaptation. It builds an object-level discrepancy memory bank to capture fine-grained source-target domain gaps and compensates target features to align with source distribution, enabling cross-domain segmentation. [Paper]

5. [Robot Learning] GeoVLA: Enpowering 3D Representations in Vision-Language-Action Models, 2025

- *Lin Sun*, Bin Xie*, Yingfei Liu,* ***Hao Shi****, Tiancai Wang, Jiale Cao*✉
- GeoVLA is a framework that bridges 2D semantics and 3D geometry for VLA. By encoding geometric embeddings with a dual-stream design and leveraging a Mixture-of-Experts 3D-Aware Action Expert, it achieves robustness across diverse camera views, object heights, and sizes. [Paper]

6. [Embodied Perception] Grounding Beyond Detection: Enhancing Contextual Understanding in Embodied 3 D Grounding, 2025

- *Yani Zhang*, Dongming Wu*,* ***Hao Shi****, Yingfei Liu, Tiancai Wang, Haoqiang Fan, Xingping Dong✉*
- DEGround transfers general proposals from detection into grounding with shared queries, and mitigates vision–language misalignment through region activation and query-wise modulation, achieving 1st place on EmbodiedScan leaderboard. [Paper]

Honors & Awards

1. 1st Place & Innovation Award in CVPR 2024 Autonomous Grand Challenge, Embodied 3D Grounding Track. *IEEE CS & CVPR, 2024.* (1 / 154, $ 9000)
2. Philobiblion Scholarship, 1st Prize. *THU & 杨绛-钱钟书, 2024.* (Top 10%, ￥10000)
3. CXMT Scholarship, 1st Prize. *THU & CXMT Inc, 2023.* (Top 10%, ￥10000)
4. Outstanding Bachelor's Thesis Award. *TJU, 2023.*
5. Excellent Graduate Award. *TJU, 2023.*
6. Huawei Intelligent Base Scholarship, Ministry of Education-Huawei Intelligent Base Future Stars. *Ministry of Education & Huawei Inc, 2021.*
7. 1st Prize in Mathematical Contest in Modeling. *COMAP, 2021.*
8. 1st Prize in Tianjin Undergraduate Mathematical Contest in Modelling. *CSIAM, 2021.*

Internship

Intern @ MEGVII & Dexmal 2024.08 - Present

- Department: Embodied Fundation Research Group.
- Mentors: Tiancai Wang (汪天才), Yingfei Liu (刘迎飞).

Service

- Reviewer / PC Member: ICLR 2026, AAAI 2026, ICLR 2025, ICCV 2025
- Invited Talk:
  - Talk about MemoryVLA, *3D视觉工坊, Online, 2025.*
  - Talk about MemoryVLA, *具身智能之心, Online, 2025.*
  - Talk about DenseGrounding, *CVPR 2024 Workshop on Foundation Models for Autonomous Systems, Seattle, 2024.*
  - Talk about DenseGrounding, *Technical Seminar on End-to-End Embodied Agent, Shanghai, 2024.*

Skills

- Programming: Python, PyTorch, C++, etc.
- Robots: Franka, WidowX, etc.

Ph.D. Plan

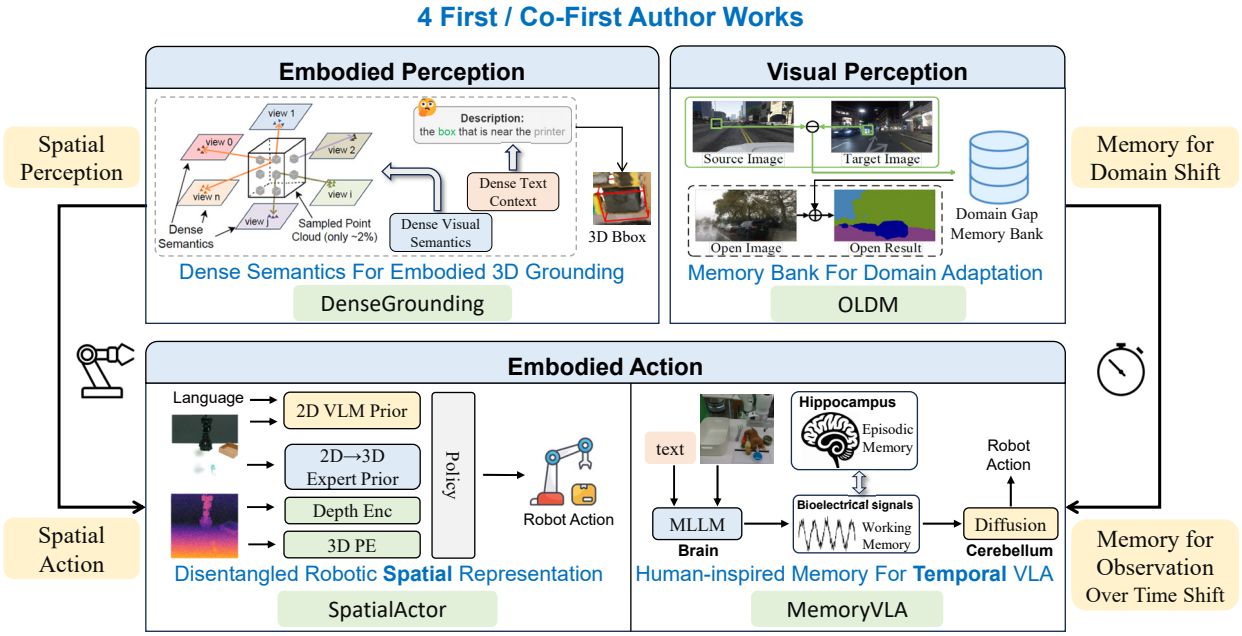
I am dedicated to building **VLA foundation models** for general robotic systems, toward embodied AGI, and hope to explore:

1. Perception & Cognition: Enhance spatial understanding and embodied reasoning by equipping VLMs with physical commonsense, affordance grounding, task planning, progress estimation, and trajectory prediction.
2. Memory: Develop human-like hierarchical memory systems for long-term context and reflective reasoning.
3. Thinking: Advance reasoning paradigms such as CoT, cross-domain co-training, and RL for VLA.
4. Action: Bridge perception and decision space, novel action expert to handle multimodal behaviors.
5. Exploration: Reinforcement pre-training to better leverage suboptimal data, RL post-training to explore OOD corner cases, test-time computation, and closed-loop feedback mechanism for autonomous intelligence.
6. Data: Leverage cross-domain data (web, human video, simulation) to enhance open-world generalization.
7. System: Balance reproducibility and sim-real consistency in evaluation, deploy VLA to varied embodiments.

I am very eager for the opportunity to pursue a Ph.D. to explore these challenges, and I hope to seek a faculty position after completing my doctoral studies.

Appendix

What Have I Done



What Will I Do