

Xiaohan Zhang

xzhan1210@gmail.com | <https://keke-220.github.io> | (607)372-5856

RESEARCH SUMMARY

Robotics, Planning, Machine Learning, and Computer Vision

I develop learning and planning algorithms that enable autonomous agents (e.g., robots and vehicles) to perform service tasks in complex environments. Here are the topics (with references to representative first-author publications) I am interested in:

- Foundation models for long-horizon decision making [C1] [C4]
- Task and motion planning for mobile manipulation [C3] [C6]
- Language grounding with multi-modal perception [J1] [C7]

ACADEMIC BACKGROUND

State University of New York at Binghamton

Binghamton, NY

Ph.D. in Computer Science (Advisor: Shiqi Zhang)

Aug. 2020 – May 2024

M.S. in Computer Science

Aug. 2019 – May 2020

Thesis: Symbol Grounding for Task and Motion Planning in Robotics

University of Texas at Austin

Austin, TX

Visiting Ph.D. Student (Advisor: Peter Stone)

Jan. 2023 – Apr. 2023

Renmin University of China

Beijing, China

B.S. in Computer Science

Sep. 2015 – May 2019

INDUSTRIAL EXPERIENCES

Boston Dynamics AI Institute

Cambridge, MA

Applied Scientist

May 2024 – Present

Building the next-generation robotics foundation models.

Meta Platforms, Inc. – FAIR

Pittsburgh, PA

Research Scientist Intern (with Chris Paxton)

May 2023 – Dec. 2023

Vision-language models and 3D scene representations for task planning, embodied question answering, and primitive skill learning

- Co-lead the OpenEQA project [C1]: the first open-vocabulary benchmark dataset for embodied question answering which contains 1600+ high-quality and non-templated questions sourced from human annotators, spanning 180+ real-world environments, This work is covered by [Meta AI Blog](#).
- Trained a 13B Vision-Language Model (VLM) that processes multi-frame inputs enriched with 3D coordinate information, and outputs actions sequences for robot task planning
- Contributed to the SLAP project [C2]: combining natural language and point-cloud embeddings as tokens within a Transformer to learn robot low-level policies. SLAP performs 4x better than the state of the art in mobile manipulation tasks

Google, LLC – Google DeepMind

Mountain View, CA

Student Researcher (with Montserrat Gonzalez)

Mar. 2022 - Aug. 2022

High-level planning and reinforcement learning for mobile manipulation

- Worked on leveraging Neural Networks to optimize mobile manipulation actions for general-purpose home-assistive robots
- Trained a Fully Convolutional Network (FCN) with 30k+ point-cloud observations for predicting navigation goals in cluttered unseen environments. The model has been deployed on a real-time robot system for cleaning tasks in Google's offices, improving the task-completion rate from 19% to 64%

- Collaborated on building a table-wiping robot using reinforcement learning and trajectory optimization [C5]. This work is covered by [Google AI Blog](#)

SELECTED PUBLICATIONS See my [google scholar](#) for a full list.
 (*indicates equal contributions)

JOURNAL ARTICLES

- [J1] **Xiaohan Zhang**, Saeid Amiri, Jivko Sinapov, Jesse Thomason, Peter Stone, and Shiqi Zhang.
[Multimodal Embodied Attribute Learning by Robots for Object-Centric Action Policies.](#)
Autonomous Robots, 2023
- [J2] Yan Ding, **Xiaohan Zhang**, Saeid Amiri, Nieqing Cao, Hao Yang, Chad Esselink, and Shiqi Zhang.
[Integrating Action Knowledge and LLMs for Task Planning and Situation Handling in Open Worlds.](#)
Autonomous Robots, 2023
- [J3] Yan Ding, **Xiaohan Zhang**, Xingyue Zhan, and Shiqi Zhang.
[Learning to Ground Objects for Robot Task and Motion Planning.](#)
IEEE Robotics and Automation Letters (RA-L), 2022

CONFERENCE PAPERS

- [C1] Arjun Majumdar*, Anurag Ajay*, **Xiaohan Zhang***, Pranav Putta, Sriram Yenamandra, Mikael Henaff, Sneha Silwal, Paul Mcvay, Oleksandr Maksymets, Sergio Arnaud, Karmesh Yadav, Qiyang Li, Ben Newman, Mohit Sharma, Vincent Berges, Shiqi Zhang, Pulkit Agrawal, Yonatan Bisk, Dhruv Batra, Mrinal Kalakrishnan, Franziska Meier, Chris Paxton, Sasha Sax, and Aravind Rajeswaran.
[OpenEQA: Embodied Question Answering in the Era of Foundation Models.](#)
The IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2024
- [C2] Priyam Parashar, Vidhi Jain, **Xiaohan Zhang**, Jay Vakil, Sam Powers, Yonatan Bisk, and Chris Paxton.
[SLAP: Spatial-Language Attention Policies.](#)
Conference on Robot Learning (CoRL), 2023
- [C3] **Xiaohan Zhang**, Yifeng Zhu, Yan Ding, Yuqian Jiang, Yuke Zhu, Peter Stone, and Shiqi Zhang.
[Symbolic State Space Optimization for Long Horizon Mobile Manipulation Planning.](#)
IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 2023
- [C4] Yan Ding*, **Xiaohan Zhang***, Chris Paxton, and Shiqi Zhang.
[Task and Motion Planning with Large Language Models for Object Rearrangement.](#)
IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 2023
 (Extended Abstract) *RSS Workshop on Learning for Task and Motion Planning*, 2023
- [C5] Thomas Lew, Sumeet Singh, Mario Prats, Jeffrey Bingham, Jonathan Weisz, Benjie Holson, **Xiaohan Zhang**, Vikas Sindhvani, Yao Lu, Fei Xia, Peng Xu, Tingnan Zhang, Jie Tan, and Montserrat Gonzalez.
[Robotic Table Wiping via Reinforcement Learning and Whole-body Trajectory Optimization.](#)
IEEE International Conference on Robotics and Automation (ICRA), 2023

- [C6] **Xiaohan Zhang**, Yifeng Zhu, Yan Ding, Yuke Zhu, Peter Stone, and Shiqi Zhang. [Visually Grounded Task and Motion Planning for Mobile Manipulation](#). *IEEE International Conference on Robotics and Automation (ICRA)*, 2022
- [C7] **Xiaohan Zhang**, Jivko Sinapov, and Shiqi Zhang. [Planning Multimodal Exploratory Actions for Online Robot Attribute Learning](#). *Robotics: Science and Systems (RSS)*, 2021
(Extended Abstract) *RSS Workshop on Integrating Planning and Learning*, 2021
- [C8] Kishan Chandan, Jack Albertson, **Xiaohan Zhang**, Xiaoyang Zhang, Yao Liu, and Shiqi Zhang. [Learning to Guide Human Attention on Mobile Telepresence Robots with 360 Vision](#). *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2021
(Extended Abstract) *RSS Workshop on Closing the Academia to Real-World Gap in Service Robotics*, 2020
- [C9] Yan Ding, **Xiaohan Zhang**, Xingyue Zhan, and Shiqi Zhang. [Task-Motion Planning for Safe and Efficient Urban Driving](#). *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2020
(Extended Abstract) *RSS Workshop on Interaction and Decision-Making in Autonomous-Driving*, 2020

HONORS & AWARDS

- Outstanding Reviewer for IEEE Robotics and Automation Letters, 2024
- SUNY Binghamton Excellence in Research in Computer Science, 2024

ACADEMIC SERVICES

- Program committee member: AAAI (2023, 2024), IJCAI (2024)
- Reviewer: RA-L (2022, 2023), ICRA (2024), IROS (2021, 2022, 2023), CoRL (2024)

TECHNICAL SKILLS

- Programming Languages: Python (proficient), C/C++, PDDL
- Robotics Softwares and Simulators: ROS, Gazebo, Habitat, Isaac Sim
- Real Robot Hardware: UR5 (Universal Robots), HSR (Toyota), Stretch (Hello Robot), Spot (Boston Dynamics)
- Deep Learning Frameworks: PyTorch, Tensorflow