# Haonan Chang

 Linkdein | 🌐 Home | G Scholar | ✉ chnme40cs@gmail.com | 📞 7348828797

## INTRODUCTION

I am a fifth-year Ph.D. student advised by Prof. Abdeslam Boularias at Rutgers University. My research focuses on developing and deploying foundation models, including **LLMs**, **VLMs**, and **Video Diffusion Models**, for robotic **manipulation** and **navigation systems**. I also specialize in scene understanding, 3D reconstruction, and SLAM.

## EDUCATION

| | | |
|---|---|---|
| 2020.9 - 2024.12 | Ph.D Candidate (CS) at **Rutgers University, New Brunswick** | (GPA: 3.9/4.0) |
| 2018.9 - 2020.5 | M.S. (Robotics and ME) at **University of Michigan, Ann Arbor** | (GPA: 4.0/4.0) |
| 2014.9 - 2018.5 | B.S. (ME and Math) at **Tsinghua University, Beijing** | (GPA: 3.7/4.0) |

## WORK EXPERIENCE

**Research Scientist (Intern) at ByteDance**                              June 2024 - Oct 2024

- Worked on deploying video diffusion models for long-horizon manipulation tasks.
- Contributed to a paper submission to CVPR2025.

**Research Scientist (Intern) at MERL**                              June 2023 - Dec 2023

- Developed a visual-force servoing system for robotic insertion tasks with novel digital parts under disturbances. Accuracy requirement for this task is under *1mm*.
- Published this work at IROS 2024 and secured a patent for the algorithm.

**Applied Scientist (Intern) at Amazon Lab126**                              June 2022 - Sept 2022

- Collaborated with the SLAM team on Amazon's Astro robot, adding a plane-based local registration module to enhance the SLAM system's accuracy by *15%*.

## SELECTED PUBLICATIONS

Here, we present a selection of publications. For a complete list, please visit my home page.

[1] Haonan Chang* et al. "ManipVDM: Hierarchical Long-Horizon Manipulation Policy via Video Diffusion Model". In: *The IEEE / CVF Computer Vision and Pattern Recognition Conference (**CVPR**) [In Submission]* (2025)
**Abstract:** We propose a new architecture that leverages the VideoDiffusion Model as a visual planner, combined with a multi-goal low-level policy, to enable long-horizon manipulation tasks.

[2] Haonan Chang* et al. "A3VLM: Actionable Articulation-Aware Vision Language Model". In: *Conference on Robot Learning (**CoRL**)* (2024)
**Abstract:** VLMs focus on understanding objects' affordances and articulation structures for manipulation. The object-centric representation of A3VLM allows it to integrate seamlessly with any manipulation hardware. [arXiv] [GitHub] [Video]

[3] Haonan Chang et al. "LGMCTS: Language-Guided Monte-Carlo Tree Search". In: *IEEE/RSJ International Conference on Intelligent Robots and Systems (**IROS**)* (2024)
**Abstract:** LLMs can generate semantically correct but non-executable plans. To address this, we combine

them with a Monte Carlo Tree Search planner. Algorithm achieves $S.O.T.A$ on multiple semantic rearrangement tasks. [arXiv] [GitHub] [Video]

[4] Haonan Chang, Abdeslam Boularias, and Siddarth Jain. "Insert-One: One-Shot Robust Visual-Force Servoing for Novel Object Insertion with 6-DoF Tracking". In: *IEEE/RSJ International Conference on Intelligent Robots and Systems (**IROS**)* (2024)
**Abstract:** A visual-force servoing-based system capable of inserting novel digital parts, even in the presence of disturbances. [Paper] [Video]

[5] Haonan Chang et al. "DAP: Diffusion-based Affordance Prediction for Multi-modality Storage". In: *IEEE/RSJ International Conference on Intelligent Robots and Systems (**IROS**)* (2024)
**Abstract:** Diffusion-based affordance prediction to identify storable regions among multiple candidates. [arXiv] [GitHub]

[6] Haonan Chang and et al. "Context-Aware Entity Grounding with Open-Vocabulary 3D Scene Graphs". In: *2023 Conference on Robot Learning (**CoRL**)* (2023)
**Abstract:** Constructing the scene as a scene graph with language features transforms the object-grounding problem into a graph search task. OVSG enables flexible language querying with contextual support. [arXiv] [GitHub] [Video]

[7] Haonan Chang et al. "Mono-STAR: Mono-camera Scene-level Tracking and Reconstruction". In: *2023 IEEE International Conference on Robotics and Automation (**ICRA**)* (2023)
**Abstract:** The first real-time 3D reconstruction system that simultaneously supports semantic fusion, fast motion tracking, non-rigid object deformation, and topological change under a unified framework. [arXiv] [GitHub] [Video]

[8] Haonan Chang and Abdeslam Boularias. "Scene-level Tracking and Reconstruction without Object Priors". In: *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (**IROS**)* (2022)
**Abstract:** The first real-time system capable of tracking and reconstructing, individually, every visible object in a given scene, without any form of prior on the scene. [arXiv] [GitHub] [Video]

## INVITED TALKS

| | |
|---|---|
| 2024 Aug | Invited by Prof. Hao Dong at Peking University |
| 2024 May | Invited by Prof. Lifeng Zhou at Drexel University |
| 2023 Dec | Invited by Dr. Peng Gao at ShangHai AI Lab |
| 2023 Oct | Invited by Prof. Hang Yang at Harvard University |

## SKILLS

I have extensive experience in developing deep learning and robotics systems, with hands-on expertise in deploying various algorithms across different robotic platforms in real-world environments.

| | |
|---|---|
| Code | Python, C++, CUDA, JavaScript, Pytorch, ROS. |
| Robot Platform | Kuka, Franka, UR5, Aloha, Fetch, Astros. |
| Robot Algorithm | Imitation Learning, Motion Planning, Task Planing, SLAM, Visual Servoing. |

## SERVICE

| | |
|---|---|
| Paper Reviewer | ICRA 2021, 2023 ISRR 2022, IEEE Transactions on Robotics, SIGGRAPH-Asia 2023, RSS 2023, ICRA 2024, CoRL 2024, ICRA2025 |