Zichen "Charles" Zhang

☐ (+1) 651-600-5947 | ✉ charlesz@allenai.org | 🏠 zcczhang.github.io | ⬛ zcczhang | 🐦 @ZCCZHANG | 🎓 Google Scholar

Education

Macalester College, St. Pual, MN   *Sep. 2019 - Dec. 2022*
GPA 3.95/4.0 **MAJOR GPA** 4.0/4.0   *Mathematics, C.S.*
- Charles J. Turck Presidential Honor Scholarship (Four-year scholarship); School Dean's List (2019 - 2022)

Research Interest

I am interested in both generalist and specialist embodied agents. My ambition is to My ambition is to build a **unified, general-purpose**, agent capable of understanding and engaging in the multimodal world. Such an agent should be capable of 3D interaction (e.g. mobile manipulation), assistant capacities (e.g. grounded multimodal agent), and adaptive learning (e.g. lifelong autonomous learning).

Research Experience

Predoctoral Young Investigator (PYI)   *Seattle, WA*
PRIOR team, Allen Institute for AI (AI2)   *Dec. 2022 - Present*
- Plan to integrate the large-scale, mobile manipulation benchmark CHORES at AI2 into the Open X-Embodiment dataset, aiming to assess how a diverse generalist model, potentially acting as an implicit or explicit world model, influences large-scale decision-making and emergent capabilities across various tasks and embodiment in sim and real.
- Part of a company-wide collaboration to create an open-source GPT4-V level generalist at AI2. Unlike the multi-modality generalist UNIFIED-IO 2, this project aims to maximize performance with a more narrow scope of vision and language capabilities.
- Co-led UNIFIED-IO 2, the first large multimodal generation model unifying over 200 understanding and generation tasks (and augmentations) across vision, language, audio, video, and action, trained from scratch.
- Led to extend the autonomous RL work with minimal human intervention with mobile manipulation robots in the real world.
- Led the project *Universal Visual Decomposer (UVD)*, with the collaboration of people in Upenn and UW.

Research   *Remote*
NVIDIA & Stanford Vision and Learning Lab (SVL) & UT Robot Perception and Learning Lab   *May. 2022 - Oct. 2022*
- Joined the research with people from NVIDIA and Stanford Vision & Learning (SVL), supervised by Linxi "Jim" Fan and Yuke Zhu.
- Developed and improved the multi-modal prompt-driven robotic manipulation tasks suite, VIMABench. Take part in a variety of foundation model implementations, initial experiments, and detection modules for object-centric manipulation for VIMA.
- Participated in developing skill primitives in MineDojo (precursor of Voyager); collaborated with the project for few-shot imitation learning using contrastive learning for analogy making.

Research (Intern)   *Seattle, WA*
PRIOR team, Allen Institute for AI (AI2)   *Sep. 2021 - Dec. 2022*
- Led the project for Autonomous RL (aka. reset-free RL) supervised by *Luca Weihs*. After identifying the main problem was near-irreversible (NI) states, we built the agent towards only request intervention (reset) when necessary, while nailing generalizations.
- Built mobile manipulation benchmark STRETCH-P&P in AI2-THOR; Contributed to the distributed RL framework *AllenAct*.

Summer Research   *St. Paul, MN*
Advised by Professor Lisa Naples, MACALESTER COLLEGE   *Jun. 2021 - Aug. 2021*
- After answering the Traveling Salesman Problem (TSP) in mathematics, developed and proved theorems and lemmas to extend the characterization of geometric measures that are carried by rectifiable curves in the dyadic cube system.

Research (Intern)   *Beijing, China*
R & D Department, Thorough Images (Now Thorough Future)   *Jan. 2021 - May 2021*
- Led the project of automated scoring systems for human epidermal growth factor receptor 2 (HER-2) after immunohistochemical (IHC) staining mentored by co-CEO & CTO *Shuhao Wang*.

Summer Research   *St. Paul, MN*
Advised by Prof. Esra Kadioglu Urtis, MACALESTER COLLEGE   *Jun. 2020 - Aug. 2020*
- Developed Q-learning-based and graph-based algorithms with simulations for UAVs coverage.

Publications / Preprints

* indicates equal contribution, † indicates equal advising

UNIFIED-IO 2: Scaling Autoregressive Multimodal Model with Vision, Language, Audio, and Action

J. Lu*, C. Clark*, S. Lee*, Z. Zhang* (Leading Authors), S. Khosla, R. Marten, D. Hoiem, A. Kembhavi, under review, 2023

A large multimodal generation model unifying over 200 understanding and generation tasks across vision, language, audio, video, and action. Mastering the dense and sparse, high-level and low-level understanding and generation tasks with emergent abilities.

Universal Visual Decomposer: Long-Horizon Manipulation Made Easy

Z. Zhang, Y. Li*, O. Bastani, A. Gupta, D. Jayaraman, Y. Ma†, and L. Weihs†, IEEE International Conference on Robotics and Automation (ICRA), 2024. Learning Effective Abstractions for Planning workshop (**Best Paper**) at CoRL 2023. Foundation Models for Decision Making (FMDM) (**oral, 6/112**) at NeurIPS 2023*

An off-the-shelf method immediately enables RL from vision without reward engineering and compositional generalizations in IL for long-horizon manipulation tasks in sim and real, without any extra data and training, task knowledge, and costs.

When Learning Is Out of Reach, Reset: Generalization in Autonomous Visuomotor Reinforcement Learning

*Z. Zhang and L. Weihs, under review. Out-of-Distribution Generalization in Robotics workshop at CoRL (**lightning talk**) 2023*

General and effective unsupervised irreversible transition measurements and a single-policy random-goal training framework, allowing agents to learn with much fewer resets and better generalize in positional, cosmetic, and structural variations in both (mobile, continuous) manipulation and navigation domains.

VIMA: General Robot Manipulation with Multimodal Prompts

Y. Jiang, A. Gupta, Z. Zhang*, G. Wang*, Y. Dou, Y. Chen, L. Fei-Fei, A. Anandkumar, Y. Zhu†, and L. Fan†, Proceedings of the Fortieth International Conference on Machine Learning (ICML) 2023. Foundation Models for Decision Making (FMDM) (**oral**) at NeurIPS 2022*

A transformer that ingests multimodal prompts and controls a robot arm for a wide range of manipulation tasks.

Automated Scoring System of HER2 in Pathological Images under the Microscope

Z. Zhang, L. Wang, and S. Wang, 18th European Congress on Digital Pathology (ECDP), 2022

Research paper for automatically recognizing and scoring HER-2 status under pathological images with interpretable procedures.

Characterization of Rectifiable Measures Carried by Lipschitz Curves

Z. Zhang, Y. Wu, and L. Naples, JMM Contributed Paper Session, AMS-PME Poster Session, 2022

Accepted for Joint Mathematics Meeting (JMM) 2022, AMS Contributed Paper Session on Functions of Complex Variables, Measure, and Integration Theory (**1** of **6** presenters), and America Mathematical Society-Pi Mu Epsilon (AMS-PME) Poster Session.

Service

CONFERENCE/WORKSHOP REVIEWER

CVPR 2024, ICRA 2024, NeurIPS 2023, CoRL 2023

TALKS

04/2023   "Autonomous Visuomotor Reinforcement Learning" at Ranjay Krishna group.

TEACHING ASSISTANT AT MACALESTER COLLEGE

Fall 2022	COMP 484: Intro to Artificial Intelligence
Spring 2022	MATH 378: Complex Analysis
Fall 2021	STAT/COMP 112: Intro to Data Science
Spring 2021	COMP 394 Topics Course: Reinforcement Learning design and write code implementations for homework and class materials
Fall 2020	COMP 128: Data Structure
Spring 2020	STAT/COMP 112: Intro to Data Science

References

Alphabetical order

Abhishek Gupta, *Assistant Professor, University of Washington, abhgupta@cs.washington.edu*
Jiasen Lu, *Research Scientist, Allen Institute for AI, jiasenl@allenai.org*
Linxi "Jim" Fan, *Senior Research Scientist, NVIDIA, linxif@nvidia.com*
Luca Weihs, *Research Manager, Allen Institute for AI, lucaw@allenai.org*
Shuhao Wang, *CTO, co-CEO, Thorough Future, eric.wang@thorough.ai*
Susan Fox, *DeWitt Wallace Professor and Department Chair, Macalester College, fox@macalester.edu*
Yuke Zhu, *Assistant Professor, UT Austin, yukez@cs.utexas.edu*