

# Zichen "Charles" Zhang

☎ (+1) 651-600-5947 | ✉ charlesz@allenai.org | 🏠 zcczhang.github.io | 📺 zcczhang | 🌐 charleszzz | 🐦 @ZCCZHANG

## Education

**Macalester College, St. Paul, 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 agents, with a focus on developing elegant, generalizable, and effective approaches that can be applied across a wide range of tasks and scenarios. My ambition is to leverage large-scale world data to build a unified general-purpose agent capable of understanding and engaging in the multimodal and dynamic world. This includes interactions in 3D space (e.g. mobile manipulator), assistant capacities (e.g. grounded multimodal agent), and adaptive learning processes (e.g. lifelong autonomous learning).

## Research Experience

### Predoctoral Young Investigator (PYI)

Seattle, WA

*PRIOR team, Allen Institute for AI (AI2)*

Dec. 2022 - Present

- Co-led Unified-IO 2, a large multimodal generation model unifying over 200 understanding and generation tasks across vision, language, audio, and action, trained from scratch and open-sourced everything.
  - Main contributions in designing and augmenting the pretraining and instruction finetuning tasks and data for all modalities; investigating model architectures and training pipelines; extensive experiments and ablations to stabilize the training and improve the benchmark performance; large-scale evaluations for all modalities; conversion from tensorflow & jax/flax pipeline to torch for open-sourcing and embodiment evaluations.
- 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.
  - Proposed the idea of iterative decomposition using pretrained visual representations; designed and established the benefit of UVD as the “free-lunch” generalizations for IL and the optimally monotonic reward for RL; infrastructure, methods, models implementations; improved the data collection and online evaluation processes for real-world experiments.

### 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 novel multi-modal prompt-driven robotic manipulation tasks suite, VIMABench. Take part in a variety of foundation model implementations for sequence modeling, initial experiments, ablations, and baselines for VIMA.
- Participated in developing skill primitives in MineDojo; collaborated with the project for few-shot imitation learning (significantly sped up the training infrastructure; proposed the idea 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 is the near-irreversible (NI) states, we built the agent towards only request intervention (reset) when necessary, while nailing generalizations.
- Built mobile manipulation simulation benchmark STRETCH-P&P in AI2-THOR.
- Contributed to the modular and flexible learning 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

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, 2023

A large multimodal generation model unifying over 200 understanding and generation tasks across vision, language, audio, 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†, in submission, 2023, *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 training, task knowledge, and costs.

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

**Z. Zhang** and L. Weihs, in submission, *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†, 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

ICRA 2024, CVPR 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, [abh Gupta@cs.washington.edu](mailto:abh Gupta@cs.washington.edu)

**Jiasen Lu**, Research Scientist, Allen Institute for AI, [jiasenl@allenai.org](mailto:jiasenl@allenai.org)

**Linxi "Jim" Fan**, Senior Research Scientist, NVIDIA, [linxif@nvidia.com](mailto:linxif@nvidia.com)

**Luca Weihs**, Research Manager, Allen Institute for AI, [luca w@allenai.org](mailto:luca w@allenai.org)

**Shuhao Wang**, CTO, co-CEO, Thorough Future, [eric.wang@thorough.ai](mailto:eric.wang@thorough.ai)

**Yuke Zhu**, Assistant Professor, UT Austin, [yukez@cs.utexas.edu](mailto:yukez@cs.utexas.edu)

**Susan Fox**, DeWitt Wallace Professor and Department Chair, Macalester College, [fox@macalester.edu](mailto:fox@macalester.edu)