Mu Cai

Contact Information

1210 W Dayton St
UW–Madison, CS Department
Madison, WI 53706 USA

Phone: (510) 280-4206
E-mail: mucai@cs.wisc.edu
WWW: https://pages.cs.wisc.edu/~mucai/

Research Interests

My recent research interest lies in the application and fundamental limitations of multimodal generative models. I am especially interested in video, visual prompting, and 3D understanding.

Education

University of Wisconsin–Madison, Madison, WI, USA

Ph.D. Student, Computer Sciences, 2020-2025 (Expected)

- Advisor: Prof. Yong Jae Lee
- GPA: 4.0/4.0

Xi'an Jiaotong University (XJTU), Xi'an, China

Bachelor in Electrical Engineering, July, 2020
Minor in Artificial Intelligence, July, 2020

- GPA: 4.05/4.3 (93.54/100, rank 1/350)
- Outstanding Student Pacesetter (highest honor, 10 out of ∼4,000 undergraduates)
- Advisor: Prof. Badong Chen

Publications & Manuscripts

Mu Cai, Jianwei Yang, Jianfeng Gao, Yong Jae Lee. Matryoshka Multimodal Models. *arXiv* 2024.

Jianrui Zhang*, **Mu Cai***, Yong Jae Lee. Vinoground: Scrutinizing LMMs over Dense Temporal Reasoning with Short Videos. *arXiv* 2024. (*equal contribution)

Thao Nguyen, Haotian Liu, Yuheng Li, **Mu Cai**, Utkarsh Ojha, Yong Jae Lee. Yo'LLaVA: Your Personalized Language and Vision Assistant. *NeurIPS* 2024.

Bocheng Zou*, **Mu Cai***, Jianrui Zhang, Yong Jae Lee. VGBench: Evaluating Large Language Models on Vector Graphics Understanding and Generation. *EMNLP main* 2024. (*equal contribution)

Yuheng Li, Haotian Liu, **Mu Cai**, Yijun Li, Eli Shechtman, Zhe Lin, Yong Jae Lee, and Krishna Kumar Singh. Removing Distributional Discrepancies in Captions Improves Image-Text Alignment. In *Proceedings of the European Conference on Computer Vision* (ECCV), 2024.

Jianrui Zhang*, **Mu Cai***, Tengyang Xie, Yong Jae Lee. CounterCurate: Enhancing Physical and Semantic Visio-Linguistic Compositional Reasoning via Counterfactual Examples. *ACL Findings*, 2024. (*equal contribution)

Mu Cai, Haotian Liu, Siva Karthik Mustikovela, Gregory P. Meyer, Yuning Chai, Dennis Park, Yong Jae Lee. Making Large Multimodal Models Understand Arbitrary Visual Prompts. *IEEE/CVF Conference on Computer Vision and Pattern Recognition* (CVPR), 2024.

Mu Cai, Chenxu Luo, Yong Jae Lee, Xiaodong Yang. Cross-Modal Self-Supervised Learning with Effective Contrastive Units for LiDAR Point Clouds. *IEEE/RSJ Intelligent Robots and Systems Conference* (IROS), 2024.

Xiang Li, Cristina Mata, Jongwoo Park, Kumara Kahatapitiya, Yoo Sung Jang, Jinghuan Shang,

Kanchana Ranasinghe, Ryan Burgert, **Mu Cai**, Yong Jae Lee, and Michael S. Ryoo. LLaRA: Supercharging Robot Learning Data for Vision-Language Policy. *arXiv* 2024.

Yuzhang Shang*, **Mu Cai***, Bingxin Xu, Yong Jae Lee†, Yan Yan†. LLaVA-PruMerge: Adaptive Token Reduction for Efficient Large Multimodal Models. *arXiv* 2024. (*equal contribution, †equal advisiing)

Mu Cai*, Zeyi Huang*, Yuheng Li, Haohan Wang, Yong Jae Lee. Leveraging Large Language Models for Scalable Vector Graphics-Driven Image Understanding. *arXiv* 2023. (*equal contribution)

Yuexiang Zhai, Shengbang Tong, Xiao Li, **Mu Cai**, Qing Qu, Yong Jae Lee, Yi Ma. Investigating the catastrophic forgetting in multimodal large language models. *Conference on Parsimony and Learning* (Proceedings Track, oral), 2023.

Zeyi Huang, Andy Zhou, Zijian Ling, **Mu Cai**, Haohan Wang, Yong Jae Lee. A Sentence Speaks a Thousand Images: Domain Generalization through Distilling CLIP with Language Guidance. In *IEEE/CVF International Conference on Computer Vision* (ICCV) , 2023.

Mu Cai, and Yixuan Li. Out-of-distribution Detection via Frequency-regularized Generative Models. In *IEEE/CVF Winter Conference on Applications of Computer Vision* (WACV), 2023, **Spotlight**

Haotian Liu, **Mu Cai**, and Yong Jae Lee. Masked Discrimination for Self-Supervised Learning on Point Clouds. In *Proceedings of the European Conference on Computer Vision* (ECCV), 2022.

Xuefeng Du, Zhaoning Wang, **Mu Cai**, and Yixuan Li. VOS: Learning What You Don't Know by Virtual Outlier Synthesis. In *International Conference on Learning Representations* (ICLR), 2022.

Mu Cai, Hong Zhang, Huijuan Huang, Qichuan Geng, Yixuan Li, Gao Huang. Frequency Domain Image Translation: More Photo-realistic, Better Identity-preserving. In *IEEE/CVF International Conference on Computer Vision* (ICCV), 2021.

Liting Sun*, **Mu Cai***, Wei Zhan, and Masayoshi Tomizuka. A Game-Theoretic Policy-Aware Interaction Strategy with Validation on Real Traffic Data. In *IEEE/RSJ Intelligent Robots and Systems Conference* (IROS), 2020. (*equal contribution)

Experience

Microsoft Research, Redmond, WA, USA

Research Intern **March, 2024 - Present**

Research project: Multimodal Video Models.
Supervised by Dr. Jianwei Yang and Prof Yong Jae Lee

Cruise AI Research, Sunnyvale, CA, USA

Research Intern **May, 2023 - December, 2023**

Research project: Making Large Multimodal Models Understand Arbitrary Visual Prompts. (CVPR2024)
Supervised by Prof Yong Jae Lee, Dr. Dennis Park, and Dr. Siva Karthik Mustikovela

QCraft Inc, San Jose, CA, USA

Research Intern **May, 2022 - March, 2023**

Research project: Cross-Modal Self-Supervised Learning with Effective Contrastive Units for LiDAR Point Clouds. (IROS2024)
Supervised by Dr. Xiaodong Yang

Kuaishou(Kwai) Technology, Beijing, China

Research Intern **June, 2020 - November, 2020**

Research project: Frequency Domain Image Translation: More Photo-realistic, Better Identity-preserving. (ICCV 2021)

Supervised by Prof. Gao Huang and Hong Zhang

SenseTime Research, Beijing, China

Research Intern **December, 2019 - June, 2020**

Supervised by Prof. Jifeng Dai

University of California, Berkeley, Berkeley, CA, USA

Visiting Student Researcher **July, 2019 - December, 2019**

Research project: A Game-Theoretic Policy-Aware Interaction Strategy with Validation on Real Traffic Data. (IROS 2020)

Supervised by Prof. Masayoshi Tomizuka

Professional Activities

- **Conference Reviewer:**
  - Computer Vision and Pattern Recognition (CVPR), 2023-24
  - Neural Information Processing Systems (NeurIPS), 2023-24
  - International Conference on Learning Representations (ICLR), 2023-25
  - International Conference on Machine Learning (ICML), 2024
  - European Conference on Computer Vision (ECCV), 2022, 2024
  - International Conference on Computer Vision (ICCV), 2023
  - Special Interest Group on Computer Graphics and Interactive Techniques (SIGGRAPH), 2024
  - IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 2024
  - AAAI Conference on Artificial Intelligence (AAAI), 2023-24
  - International Conference on 3D Vision (3DV), 2022
  - The International Joint Conference on Artificial Intelligence (IJCAI), 2023
  - IEEE/CVF Winter Conference on Applications of Computer Vision (WACV), 2023-25
- **Journal Reviewer:**
  - IEEE Transactions on Multimedia (TMM), 2022-24
  - IEEE Transactions on Neural Networks and Learning Systems (TNNLS), 2024
- **Teaching Assistant:**
  - CS639 Deep Learning for Computer Vision, UW Madison, Spring 2023
  - CS400 Programming III (Java), UW Madison, Fall 2020, Spring 2021, Fall 2021

Honors and Awards

- Qualcomm Innovation Fellowship Finalist, 2024
- UW-Madison, Computer Sciences Department Summer Research Award, 2021
- XJTU, First Prize, Huawei Large-scale Image Classification Competition, 2020
- XJTU, Outstanding Graduate, 2020
- XJTU, Thanksgiving Chinese Scholarship for Modern and Contemporary Scientists, 2019
- XJTU, Gold Prize, "Tengfei Cup" Student Technology Innovation Competition, 2018-2019

- XJTU, Ultra-High Voltage (UHV) Scholarship (Top 1%, highest honor in School of EE), 2018
- XJTU, First Prize, XJTU Mathematical Contest in Modeling, 2018
- XJTU, First Prize, The Chinese Mathematics Competitions, 2017

Skills

- Programming: Python, C/C++, Matlab, Java, SQL
- Misc: PyTorch, LaTeX, Markdown
- OS: Linux, macOS, Windows