Mingyuan Zhang

Boston, MA Homepage zhang.mingyua@northeastern.edu 781 219 1785 Google Scholar

EDUCATION

***Northeastern University* (NEU)** Boston, MA
• **Ph.D.** candidate in **Computer Engineering** Sep. 2023 – May 2028 (expected)
• Advisor: Prof. Yun Raymond Fu (Member of the Academy of Europe; Fellow of NAI, AAAS, AAAI, IEEE, AIMBE, OSA, SPIE, IAPR)
• Research Assistant at SmileLab • Focusing on computer vision, NLP and machine learning

***Northeastern University* (NEU)** Boston, MA
• **Master of Science** in **Electrical and Computer Engineering** Sep. 2021 – Dec. 2022
• Research Assistant at SmileLab
• **Concentration:** Computer Vision, Machine Learning and Algorithm

***Xidian University* (XDU)** Xi'an, China
• **Bachelor of Engineering** in **Telecommunications Engineering** Sep. 2017 – Jun. 2021
• Honor of Excellent Students Leader, 2020

RESEARCH INTEREST

Computer Vision | Efficiency | Visual Understanding | Embodied AI | Agents | Multimodal Learning | Generative Model

EXPERIENCE

Northeastern University | **SmileLab** | *Research Assistant* | Boston, MA *Sep. 2023 - Present*

Advisor: Prof. Yun Raymond Fu

- **Large Language Models and Multimodal Learning**
  - Developed efficient adaptation and integration methods for LLMs and VLMs, including fine-tuning, model merging, and latent MoE frameworks to improve performance with reduced computational overhead.
  - Built multimodal reasoning systems that enhance video-based MLLMs with perception, keyframe selection, and decision-making modules.
  - Investigated reliability challenges in large language and multimodal models, including hallucination in video LLMs and entropy-aware decoding strategies.
- **Pruning and Distillation**
  - Developed and surveyed model compression techniques for efficient foundation models, including sparse training, pruning, data pruning, and distillation methods that improve scalability.
- **Autonomous Driving and Robotics**
  - Developed and studied language and vision guided reasoning methods for autonomous systems, spanning visual evidence agents, VLA decision improvement, trajectory prediction, and language-driven motion understanding.
- **Generative Models**
  - Developed structure-guided generative models for visual synthesis and robust perception, spanning agent-centric visual generation and spatial-temporal view synthesis.

Amazon | **Industrial Robotics Group (IRG)** | *Applied Science Intern* | North Reading, MA *May. 2026 – Aug. 2026*

Mentor: Dr. Kai Gao, Dr. Jeroen van Baar

- **Long-horizon robot task with reasoning**
  - Developing world-model-based reasoning frameworks for long-horizon robotic manipulation, integrating visual evidence, future-state prediction, and VLA feedback to improve decision quality and task robustness.

Socure | **Computer Vision team** | *Research Intern* | Remote *Jun. 2025 – Aug. 2025*

Mentor: Dr. Weqiang Mu, Dr. Feng Xiao

- **Low-Level Vision**
  - Developed a computational framework for barcode reconstruction aimed at mitigating the effects of occlusion, blur, and partial data loss, thereby enhancing decoding robustness and accuracy.

AinnovationLabs | *Applied Scientist Intern* | Remote *Dec. 2022 – Dec. 2024*

Mentor: Prof. Yun Raymond Fu

- **Healthcare AI and App Development**
  - Invented **1 patent** for sleep signal processing that identifies sleep stages and detects sleep disorders from complex acoustic signals.
  - Developed a mobile app for recording sleep patterns and consulting sleep-related questions, enabling large-scale multimodal data collection for sleep research.
- **Computer Vision and Human-Centered Perception**
  - Developed a spatio-temporal 3D human pose estimation model with a self-collected dataset for accurate single-person motion tracking.
  - Built a pedestrian re-identification system capable of recognizing the same individual across diverse environments.

PUBLICATION

Under Review / Preprint

[**U-1**] ***Mingyuan Zhang***, Yue Bai, Huan Wang, Yizhou Wang, Qihua Dong, Yitian Zhang, and Yun Fu, "Boosting Large Language Models with Mask Fine-Tuning". [Paper] [Code]

[**U-2**] ***Mingyuan Zhang***, Yue Bai, Yifan Wang, Yiyang Huang, and Yun Fu, "Rethinking Fine-Tuning: Unlocking Hidden Capabilities in Vision-Language Models". [Paper]

[**U-3**] ***Mingyuan Zhang***, Jianglin Lu, Yiyang Huang, Yitian Zhang, Yizhou Wang and Yun Fu, "Selective Reduction in Foundation Models: A Survey".

[**U-4**] Yue Bai*, ***Mingyuan Zhang***, Huan Wang, Zhiqiang Tao, Kunpeng Li, and Yun Fu, "Random Sparse Networks Training with Sharpness-Aware Regularization".

[**U-5**] Yue Bai*, ***Mingyuan Zhang***, Handong Zhao, and Yun Fu, "Advancing Multi-Modal Foundation Models with Adapter Ensemble Strategies".

[**U-6**] Yi Xu, Ruining Yang, Yitian Zhang, Yizhou Wang, Jianglin Lu, ***Mingyuan Zhang***, Lili Su, and Yun Fu, "Trajectory Prediction Meets Large Language Models: A Survey". [Paper] [Code]

[**U-7**] Jianglin Lu, Hailing Wang, Xu Ma, Qihua Dong, ***Mingyuan Zhang***, Yizhou Wang, and Yun Fu, "MUSE: A Unified Multimodal Harness for MLLMs".

Conference

[**C-1**] Yizhou Wang, Lingzhi Zhang, Yue Bai, Mang Tik Chiu, Zhengmian Hu, ***Mingyuan Zhang***, Qihua Dong, Yu Yin, Sohrab Amirghodsi, and Yun Fu, "Cautious Next Token Prediction", ***ACL 2025***. [Paper] [Code]

[**C-2**] Yiyang Huang, Yitian Zhang, Yizhou Wang, ***Mingyuan Zhang***, Liang Shi, Huimin Zeng, and Yun Fu, "Distorted or Fabricated? A Survey on Hallucination in Video LLMs", ***ACL 2026***. [Paper] [Code]

Ongoing Work

[**O-1**] Xu Ma, Jianglin Lu, ***Mingyuan Zhang***, and Yun Fu, "Agentic Visual Generation in the Late 2020s".

[**O-2**] ***Mingyuan Zhang***, Yue Bai and Yun Fu, "SoupLM: Model Integration in Large Language and Multi-Modal Models".

ACADEMIC SERVICE

Reviewer:

- Conference: ACMMM 2024, AAAI 2025, UAI 2025, ACMMM 2025, NeurIPS 2025
- Journal: TKDD, TPAMI

Teaching:

- TA of EECE5639 Computer Vision, Spring 2026, NEU BOS