

EDUCATION

Tianjin University

Tianjin, China

▪ B.S. in Data Science and Big Data Technology, School of Mathematics

Aug. 2021–Jul. 2025 (Expected)

- **GPA:** 3.52/4.0, **Rank:** 7/32
- **Main Courses:** Mathematic Analysis (98), Theory of Probability (92), Data Visualization (100), Deep Learning (93), Machine Learning (95), Algorithm Design & Analysis (95), Data Science (94), Combinatorial Mathematics (93)

RESEARCH INTEREST

My primary research goal is to develop reliable and efficient machine learning models/algorithms to address real-world challenges. With this vision, my work focuses on the Generalization of Multimodal Learning and the Alignment of MLLMs. Currently, my research interests include:

Multimodal Learning

▪ *Multimodal Fusion, Imbalanced Multimodal Learning, Explainable Machine Learning*

Alignment of Foundation Models

▪ *Inference Time Alignment of VLM/LLMs, Steering FMs towards Human Preference, Ethical Standards*

Trustworthy AI

▪ *AI Safety, Fairness, Uncertainty, etc.*

PUBLICATIONS

(* denotes equal contribution)

[P1] **ETA: Evaluating Then Aligning Safety of Vision Language Models at Inference Time**

Yi Ding, Bolian li, Ruqi Zhang (Advisor)

Preprint (Under Review 2024)  

[C1] **Predictive Dynamic Fusion**

Bing Cao (Advisor), Yinan Xia*, Yi Ding*, Changqing Zhang, Qinghua Hu (Advisor)

International Conference on Machine Learning (ICML 2024) 

[C2] **Test-Time Dynamic Image Fusion**

Bing Cao (Advisor), Yinan Xia*, Yi Ding*, Changqing Zhang, Qinghua Hu (Advisor)

Neural Information Processing Systems (NeurIPS 2024) 

RESEARCH EXPERIENCE

RZ-Lab, Purdue University

USA

▪ *Research Intern, Advised by Prof. Ruqi Zhang*

May 2024–Present

- Proposed a two-phase plug-and-play alignment framework (ETA) featuring a multimodal evaluator and bi-level alignment, provided a new perspective of safety challenges in vision-language models caused by continuous visual token embeddings. ETA ensures responses are both safe and useful, improving harmlessness and helpfulness without additional training or data while maintaining the VLM's general performance. (*In Submission*)

MLDM Lab, Tianjin University

Tianjin, China

▪ *Research Intern, Advised by Prof. Bing Cao and Prof. Qinghua Hu*

Sep. 2023–Present

- Revealed that the key to dynamic fusion lies in the correlation between the weights and the loss, providing theoretical foundations for multimodal decision-level fusion from the perspective of the generalization error bound. Based on this insight, proposed the *Predictive Dynamic Fusion* (PDF) algorithmic framework offers trustworthy priors for decision-level fusion in multimodal systems or multi-agent settings, thereby achieving better generalization. (*Accepted to ICML 2024*)
- Proved theoretically that dynamic image fusion outperforms static image fusion and introduced Relative Dominability, providing a formal framework to enhance the interpretability of complex network architectures. This theoretical proof supports dynamically fusing advantageous regions and adjusting fusion weights at test time, leading to significant improvements in image quality across various baselines. (*Accepted to NeurIPS 2024*)

UNITES Lab, UNC Chapel Hill

North Carolina, USA

▪ *Research Intern*, Advised by **Prof. Tianlong Chen**

Feb. 2024-May 2024

- Developed a neural network ensemble framework for time series stock prediction, integrating multiple models (*1D-CNN, GRU, etc.*) to improve accuracy. Additionally, proposed an efficient fine-tuning system for time series foundation models (EFT-TSFM) to enhance parameter and memory efficiency. Findings were summarized in the NN in Finance technical report.

SKILLS

Languages: Chinese Mandarin (Native), English (TOEFL 102(22))

Research Abilities: Proficient in coding: Python, \LaTeX , MATLAB; Enjoys mathematical derivations; Solid foundation in mathematics and statistics.

SERVICE

Conference Reviewer

- ICLR 2025