MINGDA WAN

Research Interest: Machine Learning
Homepage | dylan.r.mathison@gmail.com | Google Scholar

EDUCATION

Anhui University September 2021 – July 2025
Internet+ College *Anhui, China*
B.Eng. in Intelligent Science and Technology

- **Selected Coursework**: Deep Learning, Natural Language Processing, Pattern Recognition and Machine Learning, Curriculum Design of Machine Learning, Probability Theory and Mathematical Statistics, Linear Algebra, Applied Statistics, Advanced Mathematics, Discrete Mathematics, Algorithm Analysis and Design, Data Structure, Computer Organization and Architecture, Programming and Algorithm Comprehensive Practice.

PUBLICATIONS

Publications are listed in reverse chronological order, * denotes equal contribution or alphabetical order.
Accepted In Conferences:

1. Yingyu Liang*, Zhizhou Sha*, Zhenmei Shi*, Zhao Song*, **Mingda Wan***, Yufa Zhou*. "Unraveling the Smoothness Properties of Diffusion Models: A Gaussian Mixture Perspective" ***ICCV 2025***. Camera Ready Paper Preview

2. Bo Chen*, Chengyue Gong*, Xiaoyu Li*, Yingyu Liang*, Zhizhou Sha*, Zhenmei Shi*, Zhao Song*, **Mingda Wan*** and Xugang Ye*. "NRFlow: Towards Noise-Robust Generative Modeling via High-Order Flow Matching" ***UAI 2025***. Camera Ready Paper Preview

3. Zhihang Li*, Zhizhou Sha*, Zhao Song* and **Mingda Wan***. "Attention Scheme Inspired Softmax Regression" ***ICLR 2025 Workshop***. Camera Ready Paper

4. Zhihang Li*, Zhizhou Sha*, Zhao Song* and **Mingda Wan***. "An Improved Sample Complexity for Rank-1 Matrix Sensing" ***ICLR 2025 Workshop***. Camera Ready Paper

5. Bo Chen*, Chengyue Gong*, Xiaoyu Li*, Yingyu Liang*, Zhizhou Sha*, Zhenmei Shi*, Zhao Song* and **Mingda Wan***. "High-Order Matching for One-Step Shortcut Diffusion Models" ***ICLR 2025 Workshop***. Camera Ready Paper

Preprints:

1. Yingyu Liang*, Zhizhou Sha*, Zhenmei Shi*, Zhao Song* and **Mingda Wan***. "HOFAR: High-Order Augmentation of Flow Autoregressive Transformers" *arXiv preprint arXiv:2503.08032*. arXiv Link

2. Yang Cao*, Bo Chen*, Xiaoyu Li*, Yingyu Liang*, Zhizhou Sha*, Zhenmei Shi*, Zhao Song* and **Mingda Wan***. "Force Matching with Relativistic Constraints: A Physics-Inspired Approach to Stable and Efficient Generative Modeling" *arXiv preprint arXiv:2502.08150*. arXiv Link

3. Xiaoyu Li*, Yingyu Liang*, Zhenmei Shi*, Zhao Song* and **Mingda Wan***. "Theoretical Constraints on the Expressive Power of RoPE-based Tensor Attention Transformers" *arXiv preprint arXiv:2412.18040*. arXiv Link

RESEARCH EXPERIENCE

Dynamics and Efficiency Assessment of High-Order Flow Models Feb 2025 - May 2025
Instructed by Prof. Zhao Song at UC Berkeley, Prof. Yingyu Liang and Dr. Zhenmei Shi at UW-Madison

- Developed theoretical insights that enable high-order modeling without sacrificing computational efficiency.
- Performed empirical evaluations on the CIFAR-10 dataset to highlight improvement in image generation quality.
- Under review at NeurIPS 2025.

Improving Sampling Precision and Stability in Shortcut Diffusion Models Oct 2024 - Feb 2025

Instructed by Prof. Zhao Song at UC Berkeley, Prof. Yingyu Liang and Dr. Zhenmei Shi at UW-Madison

- Delivered smoother trajectory generation and improved distribution alignment on synthetic datasets.
- Provided theoretical guarantees of the superior approximation precision offered by high-order methods.
- Introduced principles from special relativity, providing a theoretical guarantee of sampling stability.
- UAI 2025 Poster, ICLR 2025 Workshop DeLTa Poster, Under review at NeurIPS 2025.

Exploring Expressiveness of RoPE-based Tensor Attention Transformers Aug 2024 - Dec 2024

Instructed by Prof. Zhao Song at UC Berkeley, Prof. Yingyu Liang and Dr. Zhenmei Shi at UW-Madison

- Proved RoPE-based tensor attention Transformer can be simulated by TC^0 circuit.
- Established hardness results that RoPE-based tensor attention Transformer cannot solve 2 key problems.
- Under review at NeurIPS 2025.

Analyzed Diffusion Models Lipschitz Continuity and Momentum Properties Apr 2024 - May 2024

Instructed by Prof. Zhao Song at UC Berkeley, Prof. Yingyu Liang and Dr. Zhenmei Shi at UW-Madison

- Established concrete error guarantees for various SDE and ODE diffusion solvers.
- Proved that for a k-mixture of Gaussian target distribution, the density of the entire diffusion process remains a k-mixture of Gaussians.
- ICCV 2025 Poster.

Self-Explainable Heterogeneous Graph Neural Network Mar 2024 - Aug 2024

Instructed by Prof. Zhenhua Huang at Anhui University

- Synthesized heterogeneous graph datasets for rigorous benchmarking of explanation methods.
- Embedded an explanation generator into the HGNN training pipeline, leveraging enhanced contrastive learning to boost prediction accuracy and explanation quality.
- WWW 2025 Oral. (Acknowledgements, initiative waived first author)

OCR Table Structure Recognition Project July 2023 - Sep 2023

Instructed by Prof. Zhoujun Li at Beijing University of Aeronautics and Astronautics

- Optimized models using quantization and channel pruning techniques, reducing model size to 35% of the original, enabling deployment on embedded business devices.
- Designed an end-to-end table structure recognition model based on a hybrid architecture of CNN and Transformers, achieving accurate detection of complex table lines and cell localization.

Theoretical Foundations of Attention-Inspired Softmax Regression Mar 2023 - May 2023

Instructed by Prof. Zhao Song at Adobe Research

- Offering theoretical insights into the training dynamics of attention in neural networks.
- Establishing a regularized softmax regression model with provable convergence guarantees.
- ICLR 2025 Workshop DeLTa Poster.

Efficient Algorithms and Analysis for Rank-1 Matrix Sensing Nov 2022 - Mar 2023

Instructed by Prof. Zhao Song at Adobe Research

- Introduced a novel sketching and analysis technique with provable guarantees, enhancing theoretical understanding of low-rank matrix recovery.
- Developed a new algorithm for rank-1 matrix sensing that improves sample complexity from $\widetilde{O}(\epsilon^{-2}dk^2)$ to $\widetilde{O}(\epsilon^{-2}(d+k^2))$ and reduces runtime from $\widetilde{O}(md^2k^2)$ to $\widetilde{O}(md^2k)$.
- ICLR 2025 Workshop DeLTa Poster.

SKILLS

Language ability: Mandarin (Native), English (Proficient), Cantonese (Conversational)

Academic service: Research Grants Council (Hong Kong), ICCV 2025, KDD 2024

Tools: Python, C, C#, C++, JavaScript, Swift, LaTeX, Bash Script, MATLAB, PyTorch, TensorFlow, JAX, Numpy, sklearn, Pandas, OpenCV, OpenMP, SQL, CUDA