ZISHAN SHAO

+1(336) 391-4963 ◇ Durham, NC
zishan.shao@duke.edu ◇ linkedin.com/in/zishan-shao/ ◇ zishan-shao.github.io

EDUCATION

M.S. in Statistical Science, Duke University Aug 2024 - May 2026
GPA: 3.8 Advisor: Dr. Yiran Chen

B.S. in Computer Science & Statistics, Wake Forest University Aug 2020 - May 2024
GPA: 3.97 Advisor: Dr. Aditya Devarakonda ***summa cum laude*** with honors in computer science

RELEVANT COURSEWORK

- **Computer Science:** Programming Languages; Operating Systems I & II; Data Structures & Algorithms I (CSC 201); Computer Systems I & II (CSC 250/251); Algorithm Design & Analysis (CSC 301); Computer Vision (CSC 391); High Performance Computing; **Grad** – Machine Learning & Deep Neural Nets (ECE 661)
- **Statistical Science:** Probability (STA 310); Statistical Inference (STA 311); Linear Models (STA 312); Multivariate Statistics (STA 362); Statistical Learning (STA 363); Time Series Forecasting (STA 368); Network Analysis (STA 352); **Grad** – Predictive Modeling & Statistical Learning (STA 521); Bayesian Statistical Modeling & Data Analysis (STA 602); Hierarchical Modeling (STA 610L); Applied Stochastic Processes (STA 621)
- **Mathematics:** Calculus I–III; Discrete Mathematics (MST 117); Linear Algebra I (MST 121); **Grad** – Real Analysis I (MATH 531); Numerical Linear Algebra, Optimization & Monte Carlo Simulation (MATH 561); Numerical Analysis (STA 612D)

SELECTED PROJECTS

FlashSVD: Memory-Efficient Inference & Training System for Low-Rank Models GitHub

- Developed the FlashSVD, a series of rank-aware streaming kernels (i.e. FlashSVDAttention, FlashSVDFFN V1 & V2) that support task-agnostic inference and training of SVD-Based transformer models.
- Reduce peak activation memory by 70.2% and transient memory by 75% on widely used encoder/decoder mdoels (i.e. BERT, GPT) with zero accuracy or latency penalty.

ECCD: Enhanced Cyclic Coordinate Descent for Generalized Linear Models GitHub

- Developed Enhanced Cyclic Coordinate Descent (ECCD), leveraging a novel Hessian-approximation to unroll vector recurrences into efficient batched operations and eliminate costly nonlinear gradient computations.
- Achieved up to 13× speedup over state-of-the-art solvers (i.e. `glmnet, BigLasso, ncvreg, ABESS, skglm`) on real-world and synthetic benchmarks with negligible relative loss in solution regardless of blocksize.

SADA: Stability-Guided Adaptive Diffusion Acceleration GitHub

- Introduced a training-free paradigm exploiting step and token-wise sparsity to accelerate diffusion sampling-achieving $\geq 1.8\times$ speedups on SD-2, SDXL, and Flux (DPM & DPM++ solvers) with LPIPS ≤ 0.10 and FID ≤ 4.5-and proposed a unified skipping method that outperforms existing training-free approaches.
- Demonstrated cross-modal generalization with approximate $1.81\times$ acceleration on MusicLDM and approximate $1.41\times$ on ControlNet-no fine-tuning required.

EXPERIENCE

Research Assistant, Center of Computational Evolutionary Intelligence (CEI), Duke University Fall 2024–Present

Research Assistant, Sparstitute, Wake Forest University Spring 2022–Present

Research Assistant, Intelligent Remote Sensing in Conservation & Discovery Group (IRSC), Wake Forest University Spring 2022–Present

PUBLICATIONS

- **Zishan Shao**, Hancheng Ye, Yixiao Wang, Ting Jiang, Qinsi Wang, Yiran Chen. *"FlashSVD++: Causality Enabled Inference & Memory Efficient Training System."* In submission to Machine Learning and Systems (**MLSys '26**).
- Ting Jiang*, Yixiao Wang*, **Zishan Shao***, Hancheng Ye, Mingyuan Ma, Yiran Chen. *"On Training-Free Acceleration of Generative Modeling."* In submission to International Conference on Learning Representations (**ICLR '26**).
- **Zishan Shao**, Yixiao Wang, Qinsi Wang, Ting Jiang, Zhixu Du, Hancheng Ye, Danyang Zhuo, Yiran Chen, Hai Li. *"FlashSVD: Memory Efficient Approach for SVD-Based Low Rank Model Inference."* In submission to Annual AAAI Conference on Artificial Intelligence (**AAAI '26**). [PDF]
- **Zishan Shao***, Yixiao Wang*, Ting Jiang, Aditya Devarakonda. *"Enhanced Cyclic Coordinate Descent Methods for Elastic Net Penalized Linear Models."* In submission to Neural Information Processing Systems (**NeurIPS '25**).
- Ting Jiang*, Hancheng Ye*, Yixiao Wang*, **Zishan Shao**, Jingwei Sun, Jingyang Zhang, Jianyi Zhang, Zekai Chen, Yiran Chen, Hai Li. *"SADA: Stability-guided Adaptive Diffusion Acceleration."* International Conference on Machine Learning (**ICML '25 Poster**). [PDF]
- **Zishan Shao**, Aditya Devarakonda. *"Scalable Dual Coordinate Descent for Kernel Methods."* International Conference on High Performance Computing in Asia-Pacific Region (**HPCAsia '25**, CORE B), *Outstanding Paper Award*. [PDF]
- Kangning Cui, **Zishan Shao**, Gregory Larsen, Victor Pauca, Sarra Alqahtani, David Segurado, João Pinheiro, Manqi Wang, David Lutz, Robert Plemmons, Miles Silman. *"PalmProbNet: A Probabilistic Approach to Understanding Palm Distributions in Ecuadorian Tropical Forest via Transfer Learning."* Proceedings of 2024 ACM-Southeast (**ACM-SE '24**). [PDF]

PROFESSIONAL SERVICE

- **Reviewer**, AAAI Conference on Artificial Intelligence 2025

MENTORING

- **Data+ Project Manager, Duke Rhodes iiD, Duke University (Durham, NC)** *Summer 2025*
  Advisor: Dr. Gregory Herschlag. Mentored a three-person team on textual analysis in agricultural research.
- **Peer Tutor, Center for Learning, Access, and Student Success (CLASS), Wake Forest University (Winston-Salem, NC)** *Oct. 2021–Dec. 2023*
  Provided on-site computer science tutoring and study support for undergraduates.

MEMBERSHIP & HONORS

- **Phi Beta Kappa**, Honorary Member
- **Upsilon Pi Epsilon**, Honorary Member
- **COMAP Interdisciplinary Contest in Modeling (ICM) 2022**, Meritorious Winner
- **Dean's List Scholar**, all semesters
- **Wake Forest Research Fellowship (2023)**
- **George Washington Greene Scholarship (2023)**, one of seven recipients
- **Upsilon Pi Epsilon Scholarship (2023)**