QITAO TAN

qitaotan@uga.edu Github Google Scholar

RESEARCH AREA

My research focuses on efficient training and deployment of large-scale deep learning models, with an emphasis on quantization, model compression, and hardware-software co-design for edge intelligence.

EDUCATION

University of Georgia, USA *August 2024 - Now*
Ph.D. in Computer Science, School of Computing
Advisor: Dr. Geng Yuan

Central China Normal University, China *August 2020 - June 2024*
Bachelor in Information Science, School of Information Management GPA: 3.67 / 4.0
Advisor: Dr. Guanghui Ye and Dr. Chuan Wu
Courses: Mathematics, Algo & Data Struct, Database, Game theory, Software Eng, etc.

SELECTED PUBLICATIONS & PREPRINTS

Qitao Tan, Xiaoying Song, Jin Lu, Guoming Li, Jun Liu, Lingzi Hong, Caiwen Ding, Jundong Li, Xiaoming Zhai, Shaoyi Huang, Wei Niu, Geng Yuan. Enabling end-to-end quantization-aware training for LLMs at the cost of inference, arXiv preprint, 2025. (Under review)

Qitao Tan, Jun Liu, Zheng Zhan, Caiwei Ding, Yanzhi Wang, Xiaolong Ma, Jaewoo Lee, Jin Lu, Geng Yuan. Harmony in divergence: Towards fast, accurate, and memory-efficient zeroth-order LLM fine-tuning[C]. Advances in Neural Information Processing Systems. (**NeurIPS 2025**)

Qitao Tan, Sung-En Chang, Rui Xia, Huidong Ji, Chence Yang, Ci Zhang, Jun Liu, Zheng Zhan, Zhenman Fang, Zhou Zou, Yanzhi Wang, Jin Lu, Geng Yuan. Perturbation-efficient zeroth-order optimization for hardware-friendly on-device training[C]. International Conference on Computer-Aided Design. (**ICCAD 2025**, Best Paper Nomination)

Ci Zhang, Chence Yang, **Qitao Tan**, Jun Liu, Ao Li, Yanzhi Wang, Jin Lu, Jinhui Wang, Geng Yuan. Towards memory-efficient and sustainable machine unlearning on edge using zeroth-order optimizer[C]. Proceedings of the Great Lakes Symposium on VLSI. (**GLSVLSI 2025**)

Sheng Li*, **Qitao Tan***, Yue Dai, Zhenglun Kong, Tianyu Wang, Jun Liu, Ao Li, Ninghao Liu, Yufei Ding, Xulong Tang, Geng Yuan. Mutual effort for efficiency: A similarity-based token pruning for vision transformers in self-supervised learning[C]. International Conference on Learning Representations. (**ICLR 2025**)

Guanghui Ye, Jinyu Wei, **Qitao Tan**, Chuan Wu, Xiaoying Song, Songye Li. Academic collaboration recommendation based on graph neural network and multi-attribute embedding[J]. **Journal of Information Science, 2024**.

Qitao Tan, Xiaoying Song, Guanghui Ye, Chuan Wu. An effective negative sampling approach for contrastive learning of sentence embedding[J]. **Machine Learning, 2023**.

HONORS AND AWARDS

IEEE/ACM William J. McCalla ICCAD Best Paper Nomination, 2025. (**Top 1.8%**)

Outstanding graduates Awards, Central China Normal University, 2024.

Excellent Student Awards, Central China Normal University, 2022-2023

MISCELLANEOUS

Coding: Python, Java, MySql; experiences with PyTorch, TensorFlow, Docker, Git, Bash, etc.

Service: Reviewer of NeurIPS, ICLR, AAAI, and IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems (TCAD).