TIANCHEN ZHAO

Email: suozhang1998@gmail.com ⋄ Website: https://tianchen-zhao.info
Rohm Building 4-205, Tsinghua University
(+86) 17610272031

RESEARCH INTEREST

EfficientML and Software-Hardware Co-Optimization.

EDUCATION

Tsinghua University *2023/09 - Present*
Ph.D. in Electrical Engineering

Beihang University *2020/09 - 2023/06*
M.S. in Electrical Engineering

Beihang University *2016/09 - 2020/06*
B.S. in Electrical Engineering

EXPERIENCE

Infinigence, Research Intern *2023/06 - Now*
Conducted researches on efficient visual generation (diffusion) models.

- Faster evaluation of text-to-image generation (*FlashEval, CVPR24*)
- Mixed precision quantization for few-step diffusion models. (*MixDQ, ECCV24*).
- Quantization for diffusion transformer-based video and image generation. (*ViDiT-Q, Arxiv24*).

Novauto, Research Intern *2022/10 - 2023/03*
Conducted researches on efficient 3D perception:

- Transformer-basded backbone design for 3D point cloud (*CodedVTR, CVPR22*)
- Adaptive inference for 3D perception models (*Ada3D, ICCV23*).

SERVICES

Conferencqe Reviewer , for CVPR, ICCV, ECCV, NeurIPS, ICLR, ICML.
Outstanding Reviewer for NeurIPS2024.

PUBLICATIONS

"PARO: Hardware-Software Co-design with Pattern-aware Reorder-based Attention Quantization in Video Generation Models", Xinhao Yang*, **Tianchen Zhao***, Hongyi Wang, Wenheng Ma, Shulin Zeng, Zhenhua Zhu, Xuefei Ning, Huazhong Yang and Yu Wang, in Submission of DAC, 2025, *2024/10*

"ViDiT-Q: Efficient and Accurate Quantization of Diffusion Transformers for Image and Video Generation", **Tianchen Zhao**, Tongcheng Fang, Enshu Liu, Wan Rui, Widyadewi Soedarmadji, Shiyao Li, Zinan Lin, Guohao Dai, Shengen Yan, Huazhong Yang, Xuefei Ning+, Yu Wang+, in Submission of ICLR, 2025, [Project Page] *2024/10*

"MixDQ: Memory-Efficient Few-Step Text-to-Image Diffusion Models with Metric-Decoupled Mixed Precision Quantization", **Tianchen Zhao***, Xuefei Ning*+, Tongcheng Fang*, Enshu Liu,

Guyue Huang, Zinan Lin, Shengen Yan, Guohao Dai, and Yu Wang+, European Conference of Comput er Vision (ECCV), 2024, [Project Page] *2024/10*

"FlashEval: Towards Fast and Accurate Evaluation of Text-to-image Diffusion Generative Models", Lin Zhao*, **Tianchen Zhao***, Zinan Lin, Xuefei Ning+, Guohao Dai, Huazhong Yang, Yu Wang+, ,Computer Vision and Pattern Recognition (CVPR), 2024, [Project Page] *2024/06*

"DyPIM: Dynamic-inference-enabled Processing-In-Memory Accelerator", Tongxin Xie, **Tianchen Zhao**, Zhenhua Zhu+, Xuefei Ning, Bing Li, Guohao Dai, Huazhong Yang and Yu Wang+ , Design, Automation and Test in Europe Conference (DATE), 2024, *2024/03*

"Ada3D: Exploiting the Spatial Redundancy with Adaptive Inference for Efficient 3D Object Detection", **Tianchen Zhao**, Xuefei Ning, Ke Hong, Zhongyuan Qiu, Pu Lu, Linfeng Zhang, Yali Zhao, Lipu Zhou, Guohao Dai, Huazhong Yang and Yu Wang, International Conference of Computer Vision (ICCV), 2023, [Project Page] *2023/10*

"CodedVTR: Codebook-based Sparse Voxel Transformer with Geometric Guidance", **Tianchen Zhao**, Niansong Zhang, Xuefei Ning, He Wang, Li Yi, Yu Wang, Computer Vision and Pattern Recognition (CVPR), 2022, [Project Page] *2022/06*

"DSA: More Efficient Budgeted Pruning via Differentiable Sparsity Allocation", Xuefei Ning*, **Tianchen Zhao***, Wenshuo Li, Peng Lei, Yu Wang, Huazhong Yang, European Conference on Computer Vision (ECCV), 2020, *2020/10*