Jing Liu

liujing_95@outlook.com · https://www.jing-liu.com/ · https://github.com/liujingcs
https://scholar.google.com.au/citations?user=-lHaZH4AAAAJ&hl

Education

Monash University, Victoria, Australia 2021 – Present
Ph.D. candidate in Faculty of Information Technology
Supervisor: Asst. Prof. Bohan Zhuang, Prof. Jianfei Cai, and Prof. Chunhua Shen.

South China University of Technology, Guangzhou, Guangdong, China 2017 – 2020
Master in Software Engineering (SE)
Supervisor: Prof. Mingkui Tan and Prof. Qingyao Wu
GPA: 3.49/4.0

South China University of Technology, Guangzhou, Guangdong, China 2013 – 2017
Bachelor in Software Engineering (SE)
GPA: 3.73/4.0

Research Interests

Efficient training and inference for foundational models

Publications

(* indicates equal contributions)

MiniCache: KV Cache Compression in Depth Dimension for Large Language Models
Akide Liu, **Jing Liu**, Zizheng Pan, Yefei He, Gholamreza Haffari, Bohan Zhuang
Neural Information Processing Systems (NeurIPS) 2024.

ZipCache: Accurate and Efficient KV Cache Quantization with Salient Token Identification
Yefei He, Luoming Zhang, Weijia Wu, **Jing Liu**, Hong Zhou, Bohan Zhuang
Neural Information Processing Systems (NeurIPS) 2024.

QLLM: Accurate and Efficient Low-Bitwidth Quantization for Large Language Models
Jing Liu, Ruihao Gong, Xiuying Wei, Zhiwei Dong, Jianfei Cai, Bohan Zhuang
International Conference on Learning Representations (ICLR) 2024.

EfficientDM: Efficient Quantization-Aware Fine-Tuning of Low-Bit Diffusion Models
Yefei He, **Jing Liu**, Weijia Wu, Hong Zhou, Bohan Zhuang
International Conference on Learning Representations (ICLR) 2024. (**Spotlight, Top 5%**)

Stitched ViTs are Flexible Vision Backbones
Zizheng Pan, **Jing Liu**, Haoyu He, Jianfei Cai, Bohan Zhuang
European Conference on Computer Vision (ECCV), 2024.

Efficient Stitchable Task Adaptation
Haoyu He, Zizheng Pan, **Jing Liu**, Jianfei Cai, and Bohan Zhuang
Conference on Computer Vision and Pattern Recognition (CVPR) 2024.

TFMQ-DM: Temporal Feature Maintenance Quantization for Diffusion Models
Yushi Huang, Ruihao Gong, **Jing Liu**, Tianlong Chen, Xianglong Liu
Conference on Computer Vision and Pattern Recognition (CVPR) 2024. (**Spotlight, Top 11%**)

Pruning self-attentions into convolutional layers in single path

Haoyu He, **Jing Liu**, Zizheng Pan, Jianfei Cai, Jing Zhang, Dacheng Tao, Bohan Zhuang
IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI).

PTQD: Accurate Post-Training Quantization for Diffusion Models

Yefei He, Luping Liu, **Jing Liu**, Weijia Wu, Hong Zhou, Bohan Zhuang
Neural Information Processing Systems (NeurIPS), 2023.

BiViT: Extremely Compressed Binary Vision Transformers

Yefei He, Lou Zhenyu, Luoming Zhang, **Jing Liu**, Weijia Wu, Bohan Zhuang, Hong Zhou
International Conference on Computer Vision (ICCV) 2023.

Single-path Bit Sharing for Automatic Loss-aware Model Compression

Jing Liu, Bohan Zhuang, Peng Chen, Chunhua Shen, Jianfei Cai, Mingkui Tan
IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI).

A Survey on Efficient Training of Transformers

Bohan Zhuang, **Jing Liu**, Zizheng Pan, Haoyu He, Yuetian Weng, Chunhua Shen
International Joint Conference on Artificial Intelligence (IJCAI), 2023.

Dynamic Focus-aware Positional Queries for Semantic Segmentation

Haoyu He, Jianfei Cai, Zizheng Pan, **Jing Liu**, Jing Zhang, Dacheng Tao, Bohan Zhuang
Computer Vision and Pattern Recognition (CVPR), 2023.

EcoFormer: Energy-Saving Attention with Linear Complexity

Jing Liu, Zizheng Pan*, Haoyu He, Jianfei Cai, Bohan Zhuang
Neural Information Processing Systems (NeurIPS), 2022. (**Spotlight, Top 5%**)

Less is More: Pay Less Attention in Vision Transformers

Zizheng Pan, Bohan Zhuang, Haoyu He, **Jing Liu**, Jianfei Cai
AAAI Conference on Artificial Intelligence (AAAI), 2022.

Scalable visual transformers with hierarchical pooling

Zizheng Pan, Bohan Zhuang, **Jing Liu**, Haoyu He, Jianfei Cai
International Conference on Computer Vision (ICCV), 2021.

Discrimination-aware Network Pruning for Deep Model Compression

Jing Liu, Bohan Zhuang, Zhuangwei Zhuang, Yong Guo, Junzhou Huang, Jinhui Zhu, Mingkui Tan
IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI), 2021.

Effective Training of Convolutional Neural Networks with Low-bitwidth Weights and Activations

Bohan Zhuang*, Mingkui Tan*, **Jing Liu***, Lingqiao Liu, Ian Reid, Chunhua Shen
IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI), 2021.

Aqd: Towards accurate quantized object detection

Peng Chen*, **Jing Liu***, Bohan Zhuang, Mingkui Tan, Chunhua Shen
Conference on Computer Vision and Pattern Recognition (CVPR), 2021. (**Oral, Top 4%**)

Deep Transferring Quantization

Zheng Xie*, Zhiquan Wen*, **Jing Liu***, Zhiqiang Liu, Xixian Wu, Mingkui Tan
European Conference on Computer Vision (ECCV), 2020.

Generative Low-bitwidth Data Free Quantization

Shoukai Xu, Haokun Li, Bohan Zhuang, **Jing Liu**, Jiezhang Cao, Chuangrun Liang, Mingkui Tan
European Conference on Computer Vision (ECCV), 2020.

Discrimination-aware Channel Pruning for Deep Neural Networks

Zhuangwei Zhuang*, Mingkui Tan*, Bohan Zhuang*, **Jing Liu***, Yong Guo, Qingyao Wu, Junzhou Huang, Jinhui Zhu
Neural Information Processing Systems (NeurIPS), 2018.

Technical Report

MiniCache: KV Cache Compression in Depth Dimension for Large Language Models

Akide Liu, **Jing Liu**, Zizheng Pan, Yefei He, Gholamreza Haffari, Bohan Zhuang
Submitted to Neural Information Processing Systems (NeurIPS), 2024.

ZipCache: Accurate and Efficient KV Cache Quantization with Salient Token Identification

Yefei He, Luoming Zhang, Weijia Wu, **Jing Liu**, Hong Zhou, Bohan Zhuang
Submitted to Neural Information Processing Systems (NeurIPS), 2024.

Sharpness-aware Quantization for Deep Neural Networks

Jing Liu, Jianfei Cai, Bohan Zhuang
Submitted to Pattern Analysis and Machine Intelligence (TPAMI).

Professional Experiences

Journal Reviewer: TPAMI, IJCV, PR
Conference Program Committee: ICLR, NeurIPS, ICML, CVPR, ECCV, ICCV

Honors and Awards

Google Travel Grant	Mar. 2024
ICLR 2024 Financial Assistance Award	Mar. 2024
NeurIPS 2021 Outstanding Reviewer	Oct. 2021
Faculty of Information Technology Research Scholarship	Sept. 2020
The Second Prize Scholarship of SCUT	June 2020
The Third Prize Scholarship of SCUT	June 2019
The First Prize Scholarship of SCUT	June 2018