Xiandong Zhao

Institute of Computing Technology Chinese Academy of Sciences
University of Chinese Academy of Sciences
Beijing, China, (86) 18811176157
GitHub: https://github.com/hustzxd
zhaoxiandong@ict.ac.cn

EDUCATIONS

Ph.D. of Computer Architecture, Expected in Jun.2022,
Institute of Computing Technology Chinese Academy of Sciences(ICT)
University of Chinese Academy of Sciences(UCAS)

Bachelor of Computer Secience and Technology, Jun.2017,
Huazhong University of Science and Technology(HUST)

EXPERIENCE

Research Intern in Jeejio AI LAB

- Low-precision quantization for specialized hardware. [1][p1]   2018-2019
- Network pruning for bit-serial accelerators. [2][p1]   2018-2019
- Pattern pruning with ADMM for specialized accelerators. [p1]   2020

2013 - 2017 in HUST

- Five-stage pipeline CPU implementation   2016

PROJECTS

[p1] A PyTorch Framework for Efficient Pruning and Quantization for specialized NPU. Algorithm: LLSQ, LSQ, BitPruner, ADMM NPU Pruner, ADMM Level Pruner, Sparsity and Quantization (SQ).

[p2] The PyTorch implementation of Learned Step size Quantization (LSQ) in ICLR2020.

PUBLICATION

[1] **Xiandong Zhao**, Ying Wang, Xuyi Cai, Cheng Liu, Lei Zhang. "Linear Symmetric Quantization of Neural Networks for Low-precision Integer Hardware." In International Conference on Learning Representations (ICLR 2020).

[2] **Xiandong Zhao**, Ying Wang, Cheng Liu, Cong Shi, Kaijie Tu, Lei Zhang. "BitPruner: Network Pruning for Bit-serial Accelerators." In Design Automation Conference (DAC 2020)

[3] Yintao He, Ying Wang, **Xiandong Zhao**, Huawei Li and Xiaowei Li. "Towards State-Aware Computation in ReRAM Neural Networks." In Design Automation Conference (DAC 2020)

[4] Songyun Qu, Ying Wang, Bing Li, **Xiandong Zhao** and Lei Zhang. "RaQu: An automatic high-utilization CNN quantization and mapping framework for general-purpose RRAM Accelerator." In Design Automation Conference (DAC 2020)