

Wenyong Zhou

E-mail: wenyongz@connect.hku.hk

Education

The University of Hong Kong

Sep.2021 - Aug.2025

Ph.D. in Electrical and Electronic Engineering (*Postgraduate Scholarship*)

Supervisors: Prof. Ngai Wong / **Co-Supervisor:** Prof. Can Li

Research: Large Language Models, Implicit Neural Representations, In-Memory Computing

Thesis: Hardware-Aware Designs for Efficient Implicit Neural Representations

Northwestern University

Sep.2019 - Jun.2021

M.S. in Electrical and Computer Engineering

GPA: 4.0/4.0

Supervisor: Prof. Seda Memik

Thesis: Software-Hardware Co-Design on FPGA Using Differentiable Neural Architecture Search

Tianjin University

Sep.2015 - Jul.2019

Major: B.E. in Electronic Science and Technology / **Minor:** Law

GPA: 3.75/4.0

Thesis: Statistical Analysis of Voronoi Tessellation in 3D Space (Outstanding Thesis Award)

Experience

Zhicun Technology

Nov.2024 - Aug.2025

Computing Scientist

Beijing, China

JD Group

Jan.2024 - Jun.2024

Algorithm Intern

Beijing, China

ByteDance

Mar.2021 - Jun.2021

Algorithm Intern

Shanghai, China

Selected Publication

1. Wenyong Zhou, et al., "Towards RRAM-based Transformer-based Vision Models with Noise-aware Knowledge Distillation," *DATE*, Lyon, France, 2025.
2. Wenyong Zhou, et al., "Critical Weight Condensation for Robust Diffusion Models on Memristor-based CIM System," *IEEE Trans. Computer-Aided Design of Integrated Circuits and Systems*.
3. Wenyong Zhou, et al., "HaLoRA: Hardware-aware Low-Rank Adaptation for Large Language Models Based on Hybrid Compute-in-Memory Architecture," *IEEE Trans. Computer-Aided Design of Integrated Circuits and Systems*.
4. Wenyong Zhou, et al., "Binary Weight Multi-Bit Activation Quantization for Compute-in-Memory CNN Accelerators," *IEEE Trans. Computer-Aided Design of Integrated Circuits and Systems*.
5. Wenyong Zhou, et al., "RRAM-Based Isotropic CNNs with High Robustness and Resource Utilization Rate," *EDTM*, Hong Kong, China, 2025.
6. Wenyong Zhou, et al., "A Time- and Energy-Efficient CNN with Dense Connections on Memristor-Based Chips," *ASICON*, Nanjing, China, 2023.
7. Wenyong Zhou, et al., "A Hardware-Aware Neural Architecture Search Pareto Front Exploration for In-Memory Computing," *ICSICT*, Nangjing, China, 2022.

Awards and Scholarships

- HKU Postgraduate Scholarship (2021), HKU Best Teaching Assistant Award (2023), Tianjin University Outstanding Graduate (2019)