

RESEARCH INTERESTS

Machine Learning Systems, Graph Neural Networks, Distributed Deep Learning, High-Performance Computing, CXL Memory Systems

EDUCATION

08/24 – Present Texas State University

Ph.D. Student in Computer Science (Advisor: Dr. Chul-Ho Lee 🌐)

Selected Courses: High Performance Computing, Advanced Parallel Computing, Machine Learning and Pattern Recognition, Algorithms, Network Science

09/21 – 06/24 China University of Geosciences

M.Eng. in Electronic and Information (Advisor: Longsheng Wei)

Focus: Computer Vision

09/08 – 06/12 Hubei University of Technology

B.Eng. in Electrical Engineering and Automation

PROJECTS

05/26 – Present CXL Memory Pool Acceleration for cuGraph-based distributed GNN Training

- **Prototyped CXLGraph** to offload the graph store from GPU HBM to a CXL memory pool while comparing against cuGraph MGGraph-based sampling.
- **Explored CXL-backed feature storage** for read-only and read-write node features, including a look-ahead cache design for batched feature updates.
- **Investigated CXL-based communication paths** for PyTorch gradient all-reduce, NCCL collectives, and NVSHMEM-style GPU shared-memory communication.
- *Key technologies:* cuGraph, MGGraph, CXL memory, DAX, PyTorch Distributed, NCCL, NVSHMEM, RDMA, multi-GPU graph training.

08/25 – 04/26 Multi-GPU Temporal Graph Learning Research

Related manuscript under anonymous review.

- Conducted research on efficient multi-GPU training for temporal graph neural networks.

11/22 – 04/23 Building Low-Cost High-Performance GPU Computing Servers

- Assembled hardware components, including motherboard, PSU, RAM, GPU, and PCIe splitters, and built a custom **15-GPU server**.
- Installed Xubuntu and configured Docker, proxy rules, NAT, X11, and NFS.
- Wrote Dockerfiles to enable rapid deployment of frequently used environments.
- Performed performance tuning using htop, nvidia-smi, iotop, top, bcache, and swap.

06/22 – 08/22 EEG and EOG based Real-Time Driver Fatigue Detection [Code 📄]

- Developed module to convert MATLAB data into Python pickle format.
- Designed CNN-based model to fuse EEG and EOG features for real-time driver fatigue detection.

- 04/22 – 06/23 A Malfunction Detection System for Guangzhou Metro System**
- Developed OpenCV-based image preprocessing module for carriage segmentation and component-level region extraction.
 - Built component detection module based on YOLOv5 and text recognition module based on Fast R-CNN, and exposed the services via Flask API.
 - Created Docker Compose configuration to enable one-click deployment of all modules (Vue.JS frontend, Java backend, Python backend, and database).

EXPERIENCE

- 05/26 – 08/26 SK hynix America – GNN Advanced Memory System Research Intern**
- Researched CXL memory-pool acceleration for multi-GPU GNN training, including graph-store offloading, feature-store offloading, and CXL-based communication paths for PyTorch, NCCL, and NVSHMEM.
- 06/16 – 05/20 Wuhan Stack Tech Co., Ltd. – Founder, Developer**
- Developed websites, chrome extensions, and WeChat applets.
- 07/12 – 05/16 Ericsson – Radio Access Network Engineer**
- System integration and network optimization
 - Signal analysis and feature verification against 3GPP specs.
- 09/10 – 06/11 HGDOnline.net Coding Club – Web Developer**
- Participated in the redesign of the HBUT BBS.
 - Developed a student grade portal.

PUBLICATIONS

- Lei Zhao, Longsheng Wei, Zhi Ma, and Zhiheng Liu, “**A Pavement Crack Segmentation Method Based on Deformable Convolution and Enhanced Perceive Network**,” *Multimedia Tools and Applications*, 84(19):20995–21016 (2025).
- Zhi Ma, Lei Zhao, and Longsheng Wei, “**Single Human Parsing Based on Visual Attention and Feature Enhancement**,” *Journal of Advanced Computational Intelligence and Intelligent Informatics*, 27(4):561–566.
- Longsheng Wei, Lei Zhao, and Jian Peng, “**Reduced Reference Quality Assessment for Image Retargeting by Earth Mover’s Distance**,” *Applied Sciences*, 11(20):9776.

SKILLS

- **Deep Learning:** Pipeline optimization, PyTorch profiling, Distributed Data Parallel, DeepSpeed, 3D parallelism.
- **Programming:** Python (PyTorch, OpenCV, TraceViz), C/C++ (OpenMP), CUDA, Shell, Git.
- **Systems:** Linux, Docker, network analysis (Wireshark), Slurm, Aliyun ECS.

AWARDS

- **At China University of Geosciences (CUG):**
 - First Class Scholarship (Awarded twice), Second Class Scholarship (Awarded once)
 - Third Class Prize in the China Graduate Electronics Design Contest
- **At Hubei University of Technology (HBUT):**
 - Second Class Scholarship (Awarded once), Third Class Scholarship (Awarded twice)

– Outstanding Student

PATENTS

- A Human Body Posture Estimation Method and System Based on Shape Similarity. CN202210752411.0
- A Reference-Free Image Quality Evaluation Method Based on Graph Convolution and Multi-scale Features. CN202310801526.9
- A Twin Network Single Target Tracking Method and Device. CN202310897743.2
- Real-time Semantic Segmentation Method, Device, Electronic Equipment, and Storage Medium for Night Images. CN202310856554.0