Xingyang Li

brucelee_sjtu@sjtu.edu.cn / https://acm.sjtu.edu.cn/~xyli

Education

Shanghai Jiao Tong University(SJTU) Shanghai, China

Bachelor of Computer Science *Sept. 2022 - Present*

- Member of ACM Honors Class, which is an elite CS program for top 5% talented students
- **GPA** (All): **4.01**/4.3, Ranking: **3**/30, 16 A+ courses
- Selected Core Courses: Computer Vision: 100/100, Machine Learning: 97/100, Comprehensive Design for Computer System: 97/100, Information Theory: 98/100, Advanced Compiler Project: 100/100, Linear Algebra: 100/100

Publications

Harnessing Video Processing Insights for 3D-VGMs: A Comprehensive Attention-aware Way

T. Zhao, J. Liu*, **X. Li***, L. Ding, J. Li, S. Li, J. Hu, G. Dai (*equal contributions)*

- Accepted by **DAC 2025**.
- We propose a novel similarity-aware mixed-precision attention computation method with co-designed buffered-LUT-based architecture, achieving lossless results with **1.45x speedup and 1.63x energy efficiency compared to SOTA accelerators.**.

Research Experience

Research interests: efficient algorithms and accelerators for machine learning and mobile visual computing

Design Automation Innovation & Domain-specific Artificial Intelligence Lab, SJTU Shanghai, China

Undergraduate researcher, advised by Prof. Guohao Dai

- Explored efficient algorithm-hardware co-design for 3D-full-attention video generation models like CogVideoX and Open-Sora-Plan. *(co-first author; accepted by DAC 2025)*
- I proposed a low-cost attention speculation technique on fine-grained and coarse-grained levels to allow performance-aware mixed-precision 3D full attention computation with no accuracy loss and co-designed hardware.

Emerging Parallel Computing Center, SJTU Shanghai, China

Advisors: Prof. Jingwen Leng and Prof. Yu Feng at EPCC Lab, SJTU

- Explored efficient scalable point-based neural rendering models like Gaussian Splatting by reducing the dominating level-of-detail(LOD) tree traversal overhead. *(first author, in submission)*
- I proposed a memory-friendly algorithm to distribute the workload of tree traversal to different working units, boosting scalable point-based neural rendering models by taming workload imbalance and memory irregularity.

Course Projects

Mx* Compiler with Just-in-time Compiler(JIT) for llvm

SJTU ACM Class Compiler Design and Implementation 2023 Assignment (CS2966 Course Project)

A compiler from a C++ & Java-like language to RV32I Assembly with multiple llvm-level and assembly-level optimizations. It also supports JIT compilation for llvm.

RISC-V CPU Implemented in Verilog RTL

SJTU ACM Class Computer Architecture 2023 Assignment (CS2951 Course Project)

A Tomasulo RISC-V CPU with I-Cache and branch predictor. It can be successfully implemented on the FPGA board.

An Implementation of the Google File System(GFS)

SJTU ACM Class Comprehensive Design for Computer System (CS2913 Course Project)

An implementation of GFS in Golang with fault-tolerance, snapshots, and reproduced benchmarks.

A Real-time Efficient Domain Adaptation Model

Outstanding Paper of SJTU ACM Class Machine Learning (CS3308 Course Project)

A domain-adaptation training framework leveraging the time of downloading dataset with 25% training speedup.

Honors & Awards

- 2024 Commercial Sponsorship Scholarship (**14 winners** each year in SJTU)
- 2023 Longfor Merit Scholarship (**Top 10** in Zhiyuan College, SJTU)
- 2022, 2023, 2024 Zhiyuan Honorary Scholarship (**Top 2%** in SJTU)
- 2023, 2024 Academic Excellence Scholarship (Ranked 4-th, 2-nd respectively in ACM Class of 2026)

Teaching Experience

Machine Learning
Teaching Assistant of Professor Weinan Zhang *Feb. 2025 - June. 2025*

Mathematical Logic
Teaching Assistant of Professor Qiang Yin and Professor Yijia Chen *Feb. 2024 - June. 2024*

Computer Programming
Teaching Assistant of Professor Huiyu Weng *Sept. 2023 - Jan. 2024*

Role as teaching assistant: Giving lectures and recitation classes, writing documents and sample solutions, grading homework, creating exam questions, etc.

Technical Skills

Languages: Mandarin (native), English (proficient, CET6: 665/710, TOEFL: 115).
Programming Languages: Proficient in C/C++, CUDA, Go, Java, Python, and Verilog.