

李兴阳 Xingyang Li

邮箱: brucelee_sjtu@sjtu.edu.cn 主页: <https://acm.sjtu.edu.cn/xyli>

教育经历

上海交通大学 ACM 班

中国, 上海

工学学士

2022.9 - 2026.6(预计)

- GPA: 4.01/4.3, 排名: 3/30, 16 门满绩 (A+) 课程
- 核心课程分数: 计算机视觉: 100/100, 机器学习: 97/100, 计算机系统综合设计: 97/100, 信息论: 98/100, 高级编译器项目: 100/100, 线性代数: 100/100, 数理逻辑: 100/100, 编程综合实践: 100/100

研究兴趣

- 大模型系统:** 大模型推理硬件, 大模型新架构, 异构大模型推理系统
- 移动视觉计算:** 端侧视觉模型推理优化, 高效视觉模型训推算法

论文发表

(DAC'25) Harnessing Video Processing Insights for 3D-VGMs: A Comprehensive Attention-aware Way

Tianlang Zhao*, Jun Liu*, **Xingyang Li***, Li Ding, Jinhao Li, Shuaiheng Li, Jinbo Hu, Guohao Dai (*共同一作)

- 我们提出了一种相似性感知的混合精度注意力计算方法, 并采用协同设计的缓冲查找表架构加速混合精度计算, 相比领域内最优的加速器取得了无损的精度结果, 且推理速度提升 1.45 倍, 能效提升 1.63 倍。

科研经历

研究兴趣: 面向机器学习和视觉计算的高效算法和加速器

上海交通大学 DAI 实验室

中国, 上海

科研助理, 导师: 戴国浩 副教授

- 探索 3D 全注意力视频生成模型 (如 CogVideoX 和 Open-Sora-Plan) 的高效算法-硬件协同设计。(共同第一作者; 被 DAC 2025 接收)
- 作为核心作者, 提出了一种低成本的注意力推测技术, 允许在不损失精度的情况下进行性能感知的混合精度 3D 全注意力计算, 并协同设计硬件。

新兴并行计算研究中心

中国, 上海

科研助理, 导师: 冯宇 助理教授 和 冷静文 教授

- 探索高效且可扩展的高斯溅射系统, 主要目标为减少占据主要运行时间的细节层次 (LOD) 树遍历开销。(共同第一作者, 提交于 ICCAD 2025)
- 作为核心作者, 提出了一种内存友好的高效算法, 将树遍历的工作负载分配给不同的工作单元, 提升了高斯溅射模型的可拓展性, 解决了工作负载不平衡和内存不规则性的问题。

课程项目

Mx* 编译器/基于 LLVM 的即时编译器 (JIT)

上海交通大学 ACM 班编译器课程设计, 课程项目

该项目从 C++ 和类似 Java 的语言编译到 RV32I 汇编, 支持多种 LLVM 级和汇编级优化, 并支持 LLVM 的 JIT 编译。

基于 Verilog RTL 的 RISC-V CPU

上海交通大学 ACM 班计算机系统 (1), 课程项目

实现了基于 Tomasulo 算法的带有指令缓存和分支预测的乱序执行 RISC-V CPU, 可以成功在 FPGA 板上运行。

谷歌文件系统 (GFS)

上海交通大学 ACM 班计算机系统综合设计，课程项目
用 Golang 实现的 GFS，具有容错、快照的功能，并复现了论文中的基准测试。

STREAM: 一种实时且高效的领域适应模型

上海交通大学 ACM 班机器学习优秀论文，课程项目
该项目利用下载数据集的时间直接启动训练，并根据现有数据分布自适应调整训练权重，训练速度提升 25%。

获奖情况

- 2024 年华为奖学金（全校共 14 位获奖者）
- 2023 年龙湖优秀奖学金（致远学院共 10 位获奖者）
- 2022、2023、2024 年致远荣誉奖学金
- 2023、2024 年学业优秀奖学金（二等奖）
- 2024 年上海交通大学五·四优秀团员
- 全国第七届大学生艺术展演活动声乐一等奖（团体）
- 全国第七届大学生艺术展演活动上海赛区声乐一等奖（团体）

教学经历

机器学习 助教	上海交通大学，2024-25 春
数理逻辑 助教	上海交通大学，2023-24 春
程序设计 助教	上海交通大学，2023-24 秋

技术能力

语言: 汉语, 英语 (六级: 665/710, 托福: 115) .
编程语言: C/C++, CUDA, Go, Java, Python 和 Verilog.