

林泽佳

✉ linzj39@mail2.sysu.edu.cn 🏠 zejia-lin.github.io ☎ 18022302700 🗨 Next-Ways

关键词: GPU, 编译优化, 推理优化, 混合精度

- 熟悉计算机系统和体系结构, 针对 GPU 的软硬件特性, 从编译/运行时寻找优化机会, 包括硬件资源共享、任务调度和混合精度计算。
- 独立完成和发表两篇 GPU 编译优化相关论文, 熟悉 CUDA, LLVM, PyTorch, vLLM, SGLang。

🎓 教育背景

中山大学, 国家超级计算广州中心, 广州 (985, 双一流) 2022 年 9 月 – 2027 年 6 月 (预计)
计算机科学与技术 博士研究生在读 导师: 卢宇彤、张献伟

西北工业大学, 西安 (985, 双一流) 2018 年 8 月 – 2022 年 6 月
软件工程 本科 GPA 排名: 22 / 259 优秀毕业生

📄 研究成果

- 『DAC'25, CCF-A』GoPTX: Fine-grained GPU Kernel Fusion by PTX-level Instruction Weaving
Kan Wu, Zejia Lin, Mengyue Xi, Zhongchun Zheng, Wenxuan Pan, Xianwei Zhang and Yutong Lu.
ACM/IEEE Design Automation Conference, 2025.
- 『LCTES'24, CCF-B』MixPert: Optimizing Mixed-Precision Floating-Point Emulation on GPU Integer Tensor Cores
Zejia Lin, Aoyuan Sun, Xianwei Zhang and Yutong Lu. *The 25th ACM SIGPLAN/SIGBED International Conference on Languages, Compilers, and Tools for Embedded Systems, 2024.*
- 『ICCD'23, CCF-B』KeSCo: Compiler-based Kernel Scheduling for Multi-task GPU Applications
Zejia Lin, Zewei Mo, Xuanteng Huang, Xianwei Zhang and Yutong Lu. *The IEEE 41th International Conference on Computer Design, 2023.*

👥 项目/实习/助教

通过 GPU 空间-时间编排提升大模型推理性能 2024 年 6 月 – 2025 年 4 月
在投论文 导师: 卢宇彤、张献伟

GPU 推理需要平衡延迟和吞吐, 我们发现 chunked prefill 不平衡地用吞吐换取延迟, 根本原因是小的 chunk size 带来严重的 GPU wave quantization, 并且冗余的 KVCache 访问进一步加剧计算资源利用率下降。我们设计了一种通过并发执行 prefill 和 decode 的方法, 通过动态调度和调整划分给 P/D 的 SM 比例, 在满足时延要求的前提下提升吞吐。实验表明与 SGLang 相比 P90 长尾下降 78.5 \times , 平均提升 1.09 \times 吞吐 (最大 1.21 \times)。

中山大学, 广州 2023 年 3 月 – 2023 年 7 月
编译原理课程助教 任课教师: 张献伟

基于 LLVM 实现了一个 C 语言子集的编译器, 并将任务拆解作为实验课作业。编写实验教程文档, 在实验课上现场教学。该编译器后续改进为 SYsU-Lang 实验框架, 集成 CMake 和测评机方便学生进行实验构建与自动化测试, 获得中国高校计算机教育大会 2024 年教学案例一等奖。

腾讯, 深圳 2021 年 7 月 – 2021 年 9 月
微信事业群/微信支付, 后台开发实习生 导师: 张璇

用 Django 搭建分布式的企业微信内部机器人，为开发人员自动回复与后台测试环境相关问题。接入腾讯内部 TAPD 和七彩石项目管理系统，完善问题流程追踪管理，自动更新缺陷库，用于持续训练轻量的 NLP 模型提升自动回复准确率。

⚙️ 专业技能

- 编程语言: C/C++, CUDA, Python
- 编译工具链: LLVM, CMake, Clang, GDB
- GPU 相关工具: NVIDIA nsys/ncu, CUTLASS, cuBLAS
- AI 生态: PyTorch, vLLM, SGLang
- 其它编程工具: Git, Docker, Vim, L^AT_EX
- CCF CSP 认证: Top 5.2%
- 英语六级: 610 / 710

★ 获奖情况

中山大学 一等奖学金	2024, 2023, 2022 年
中山大学计算机学院 腾讯奖学金	2023 年
西北工业大学 优秀毕业生	2022 年
西北工业大学 优秀本科毕业设计	2022 年
西北工业大学 一等奖学金	2021, 2020, 2019 年
美国大学生数学建模竞赛 M 奖 (6%)	2021 年 4 月
中国高校计算机大赛微信小程序应用开发赛 全国二等奖 (5%, 参赛队长)	2020 年 8 月

♡ 专业服务

- IEEE NAS'24, Sub-reviewer
- IEEE ICPADS'22, Sub-reviewer