中山大學 SUN YAT-SEN UNIVERSITY

席梦悦

17358300383 ximy@mail2.sysu.edu.cn
mengyue414141 https://xi-meng.github.io/

基本信息

- 席梦悦，硕士。目前就读于中山大学，导师张献伟。2022-2023年在北京喜得网络科技有限公司（Cider）实习，2023年于重庆大学毕业。目前研究集中在**GPU软硬件设计优化、编译技术**等领域，包括**GPU系统，缓存管理**，编译优化，内核融合。研究成果发表于ASPDAC会议，CGO会议在投。参与导师的国家自然科学基金面上项目、联合实验室项目等。此外，担任NAS 2024协审审稿人。
- 编程语言：C++/C，Python， CUDA C/C++，HIP，Verilog，pytorch，paddle；掌握技术：CUDA/ROCm生态相关技术栈，LLVM框架，机器学习基本模型，Stable Diffusion，llama.cpp等等。

教育经历

中山大学 - 计算机科学与技术 硕士 2023.09 - 2026.06
专业绩点：94.92/100.00；专业排名：**2**/78（2.56%）；年级排名：**2**/427（1%）

重庆大学 - 计算机科学与技术（卓越） 本科 2019.09 - 2023.06
专业绩点：3.81/4.00；专业排名：**16**/219（7.31%），年级排名：**16**/295（5.42%）

项目经历

基于编译的GPU软件层资源管理 - 国家自然科学基金面上项目 2023.12 - 2024.09

- 提出了一种面向多层级缓存的自动化旁路管理技术。通过离线分析评估GPU全局内存访问指令在多层缓存中的旁路亲和度，以及指令间的协作与竞争关系。构建指令交互图进行分组，设计启发式算法为每条指令生成最优的缓存管理策略。与默认缓存策略相比，提升系统性能1.15倍。
- 提出了一种更细粒度的GPU内核融合技术。通过基本块划分解决负载均衡问题，利用基本块融合实现更精细的融合粒度，使用指令交织技术增加指令级平行，同时解决同步导致的死锁问题。与CUDA流机制相比，平均性能提升了11.2%。

软硬件协同的全智慧数据要素治理平台 - “恒超联算公共资源交易应用”联合实验室 2024.04 - 2024.08

- 项目核心成员，设计海关报关单校对技术、数据要素校验、多源数据汇总分析方案；使用开源大模型技术、ocr识别技术以及RAG技术，构建后端接口和vue2完成前端设计初步项目原型，并完成在多个机器A100、RTX3090、RTX 4090等的测评和分析。

基于MLIR循环交换的可重构计算架构软件流水优化 - 重庆大学计算机学院优秀毕业设计 2023.01 - 2023.06

- 基于可重构计算架构（CGRA）提出了一种基于多级中间表示（MLIR）的软件流水编译优化方法。使用多面体模型实现循环交换，并设计了最内层循环交换算法以探索多种循环结构。建立基于数据流图的执行时间评估模型，作为性能评估标准以识别循环结构的最优解。实验结果表明，最内层循环交换算法的解空间从超指数级优化至线性增长，数据重用提升了1.77倍，性能在最佳情况下提升了1.14至1.16倍。

Cider 跨境电商独角兽 - 推荐工程实习生 2022.09 - 2023.06

主要负责推荐算法场景下的工程部署和落地。

- 推荐模型全栈式自动化流程部署，基于airflow搭建每日自动化训练流程：读取用户今日新增数据，传输数据，模型增量训练，模型参数更新，更新模型部署接口。
- 以图搜图TOB应用部署，基于flask和milvus构建部署图片搜索库后端接口，采用多线程并发及k8s部署，完成全量图片嵌入向量训练、每日增量部署、搜索相似图片返回接口功能。

成果产出

Mpache: Interaction Aware Multi-level Cache Bypassing on GPUs - 已录用ASP-DAC，一作作者 2024.07

GoPTX: Fine-grained GPU Kernel Fusion by PTX-level Instruction Weaving - 在投CGO，三作作者 2024.09

“编译原理”及“编译器构造实验”课程改革 - CCF-计算机教育大会优秀教学案例一等奖 2024.08

- 基于友好开发体验的LLVM编译实践教学，使用cmake、docker基于vscode编辑器构建实验生成框架，主要负责语法分析实验框架构建，使用flex和bison（或者anltr）完成语法分析、类型检查、语法分析树的构建和转化。

荣誉奖项

计算机教育大会优秀教学案例一等奖（2024）、中山大学一等奖学金（2024）
国家奖学金、国家励志奖学金、重庆大学优异生、优秀学生干部、优秀毕业生、优秀学生（2019-2023）

其他

NAS 2024 协审审稿人，2019级计科（卓越）01班班长，高性能计算与人工智能协同创新国际论坛2024志愿者，英语辩论队队员；英语：517（六级），606（四级）