韩 虎生

1998.05.14 | 18801311826 | hanhusheng20z@ict.ac.cn
中国科学院大学 & 中科院计算所 | 在读工学博士 | 个人主页 | 谷歌学术

教育背景

2020.09 至今 | **中国科学院大学，中科院计算所** • 智能处理器研究中心 | 导师：郭崎、胡杏、陈云霁
计算机系统结构 • 直博 | 方向：安全高效的智能张量计算系统

2016.09 2020.06 | **清华大学** • 信息学院自动化系 | 导师：周杰
自动化 • 工学学士 | 学位论文：基于深度学习的相机阵列视觉定位

科研经历

研究方向 1: 安全高效的智能张量计算

- **安全高效的异构可信执行环境 (TEE) 设计** **2023.06-至今**
  项目: 面向异构可信计算，我们观察到异构设备保护粒度不一致导致了额外通信开销、及 CPU/NPU 小粒度的版本号/消息验证码产生了性能和存储开销。为此，我们基于张量的访存特性，提出虚拟支持 CPU 张量粒度版本号、延迟验证支持 NPU 张量粒度消息验证码、以及基于统一粒度的高效数据异构传输协议，消除片外访存和存储开销以及额外的通信开销，性能提升 4.0 倍，相比非安全执行开销仅为 2.1%。
  贡献/成果: 一作发表 CCF A 类论文 (***ASPLOS 2024***) 一篇，相关原型设计基于 Chipyard(Gemmini) 开发中。
- **全同态 (FHE) 加速器设计和开发** **2023.03-2024.02**
  项目: 面向场景动态且参数多样的全同态加密场景，我们设计了跨 CKKS 和 TFHE 方案的高效同态加速器，以统一的元操作为各种上层多项式算子提供高效支持和动态调度，消除了已有工作因固化的高层次操作模块化设计导致的调度低效和硬件利用率低下的问题，显著提升硬件利用率，加速 7 倍。
  贡献/成果: 二作发表 CCF A 类论文 (***DAC 2024***) 一篇，并在 FHE 基础算子库开发中进行架构设计和项目管理。
- **对抗补丁防御算法及系统设计** **2021.01-2023.07**
  项目: 为高效防御对抗补丁攻击，(1) 我们提出基于浅层显著特征的对抗补丁可证防御算法，并利用剪枝提升效率。（2）针对当前视频目标检测的对抗补丁防御任务，我们进一步基于帧间相似性提出关键帧计算复用框架，并基于多次推理中帧内的计算不变性提出高效数据缓存架构以消除重复计算，实现实时安全的视频目标检测系统。
  贡献/成果: 一作发表 CCF A 类论文两篇（***NeurIPS 2021*** 和 ***TCAD***），并基于 Vitis AI 在 FPGA 上实现原型系统。

研究方向 2: 处理器及加速器设计

- **通用处理器自动化设计** **2021.09-至今**
  项目: 通用处理器依赖大量人类专家的密集手工开发，难以满足当下多样多变的场景需求。（1）我参与设计并开发了” 启蒙” 一号，一个完全由输入-输出对驱动、借由二进制猜测图 (BSD) 和蒙特卡洛搜索自动设计的 32 位 RISC-V CPU 核，流片实机达到近似 Intel 80486SX CPU 的性能。（2）我们提出新的基于 LLM 的框架自动化设计框架，以微操作函数实现功能和性能优化的解耦，实现乱序 CPU 设计的高质量和高自动化。
  贡献/成果: (1) 合作发表 CCF A 类论文（***IJCAI 2024***）一篇，并参与模拟器编写、硬件验证以及 SPEC 适配。（2）合作投稿 CCF A 类论文（***ASPLOS 2025 under review***）一篇，参与思路讨论和论文撰写。
- **领域专用加速器设计（NeRF 及符号回归）** **2022.07-至今**
  项目: (1) 面向神经场景表示算法（NeRF），我们提出逐光线的硬件流水线和专用的片上访存系统，消除绝大多数的片外访存并解决了片上访存的板块冲突，相比 A100 实现 373.8 倍加速和 256.6 倍的能耗节省。(2) 面向边缘端深度符号回归，我们提出基于径向量基函数神经网络的近似算法，统一了规整的 NN 矩阵运算和不规整的 BFGS 优化的超越函数运算，相比于 GPU 实现 4.8 倍加速和 47.6 倍的能耗节省。
  贡献/成果: 合作发表/投稿 CCF A 类论文（***MICRO 2023***）和（***TCAD* 小修**），负责部分实验和论文撰写。

论文发表

已发表 6 篇 CCF A 类论文 (包括 ASPLOS, MICRO, DAC, TCAD, NeurIPS)，其中一作 3 篇，二作 1 篇。另有 2 篇在投。

研究方向 1: 安全高效的智能张量计算，包括异构可信执行环境的性能优化、同态加密加速器设计及对抗补丁防御等。

- **H. Han**, X. Zheng, Y. Wen, Y. Hao, E. Feng, L. Liang, et al.,“TensorTEE: Unifying Heterogeneous TEE Granularity for Efficient Secure Collaborative Tensor Computing,”(ASPLOS 2024, CCF A) (To appear)
- J. Mu, **H. Han**, S. Shi, J. Ye, Z. Liu, S. Liang, et al.,“Alchemist: A unified accelerator architecture for cross-scheme fully homomorphic encryption,”(DAC 2024, CCF A)
- **H. Han**, X. Hu, Y. Hao, K. Xu, P. Dang, Y. Wang, et al.,“Real-time robust video object detection system against physical-world adversarial attacks,”(TCAD 2023, CCF A)
- **H. Han**, K. Xu, X. Hu, X. Chen, L. Liang, Z. Du, et al.,“ScaleCert: Scalable certified defense against adversarial patches with sparse superficial layers,”(NeurIPS 2021, CCF A)

研究方向 2: 处理器及加速器设计，包括 CPU 自动化设计、领域专用加速器设计 (NeRF、符号计算等)。

- C. Li, P. Jin, T. Ma, **H. Han**, S. Cheng, D. Huang, ”AGON: Automated Design Framework for Customizing Out-of-Order Processors from ISA Documents” (ASPLOS 2025, CCF A, Under Review)
- S. Cheng, P. Jin, Q. Guo, Z. Du, R. Zhang, X. Hu, et al.,“Automated CPU Design by Learning from Input-Output Examples,”(IJCAI 2024, CCF A)
- X. Song, Y. Wen, X. Hu, T. Liu, H. Zhou, **H. Han**, et al.,“Cambricon-R: A fully fused accelerator for real-time learning of neural scene representation,”(MICRO 2023, CCF A)
- T. Ma, Y. Wen, X. Song, P. Jin, D. Huang, **H. Han**, et al. ”Harmonia: A Unified Architecture for Efficient Deep Symbolic Regression” (TCAD 2024, CCF A, Under Minor Revision Review)

© 专利

- 一种基于数据并行可信分布式神经网络加速器架构构建方法，发明人：胡杏、**韩虎生**、党朴成、宋新开，公开号：CN117195983A （第二发明人）
- 对抗补丁检测定位硬件架构及对抗补丁检测定位方法，发明人：胡杏、**韩虎生**、贺文凯、杜子东、郭崎，公开号：CN115422531A（第二发明人）
- 基于局部浅层重要神经元的对抗补丁检测定位方法及系统，发明人：胡杏、**韩虎生**、贺文凯、杜子东、郭崎，公开号：CN115422532A（第二发明人）
- 基于帧间相似性的对抗补丁检测定位方法及系统，发明人：胡杏、**韩虎生**、贺文凯、杜子东、郭崎，公开号：CN115422533A（第二发明人）
- 破除图像数据不可学习噪声的深度学习训练方法及系统，发明人：胡杏、党朴成、**韩虎生**、黄迪、张蕊、杜子东、郭崎、陈云霁，公开号：CN117475254A（第三发明人）

技能和语言

深度学习	熟练使用 PyTorch 框架；熟悉主流神经网络模型和算法如 LLM, CNN 等
系统安全	熟悉全同态、可信执行环境、多方安全计算、联邦学习等安全协议
体系架构	熟悉微架构知识、架构设计流程、架构模拟仿真工具等
编程语言	熟悉常用编程语言：Python, C/C++, Shell 等
工程技能	熟练使用 Linux 开发环境和工具链
语言	**英语** — 读写（优良），听说（日常交流）

荣誉奖项

- 2023-2024 年度，虑得投资博士生奖学金，中科院计算所
- 2023-2024 年度，优秀学生称号，处理器芯片全国重点实验室
- 2021-2022 年度，三好学生称号，中国科学院大学