LIU ZIHAN (Altair)

ilovehanhan1120@hotmail.com, +86 159 0215 7531

Personal Site: http://subjectnoi.github.io/about/

Education

2015.09-2019.06 Bachelor, Department of Computer Science, East China Normal University

- GPA: 3.81/4.00, Rank: 5/116, scholarship in 2018/2019. Awards in MCM, CCCC programming contest, etc.

2019.09-2022.03(Expected) Master, Department of Computer Science, Shanghai Jiao Tong University

- ReArch Lab, researching on computer system architecture, AI chip, compiler optimization, etc.

Job

2018.08-2019.01 Intern, **DBS(IBSO), SAP**

- Developing cloud and `S/4 HANA` application deployed on `Cloud Foundry` based on `Java`, `SAPUI5` with corresponding tools (`Spring`, `OData`, `MongoDB`), also responsible for unit and integrated test based on `karma`, `QUnit`, `Opa5`.

2019.01-2019.06 Intern, **GPU SM Arch, NVIDIA**

- Developing cycle level simulator with `C++`, `PTX`, `SASS` for functionality verification and performance modeling on next generation GPU circuit. I mainly worked on the logic of a series of new matrix instructions.

2020.06-2020.12 Intern, **LLVM CodeGen, Intel**

- Researching and developing on compiler back-end optimization (`LLVM`) for next generation Intel CPU. I mainly worked on the intrinsics related with new matrix instructions.

Project Experience

Compiling and optimizing tool chain for military intelligent chip.

2018.12-2020.07, SJTU, ICT-CAS, Logistic Department of Central Military Commission of P.R.C., project of National Defence Science and Technology Innovation Center.

- Implement a full-stack compiler tool chain for Cambricon MLU-100 in `C++`, `Python` with `ONNX` and `TVM` [1].
- Research and develop a series of hardware-aware graph level optimization procedures[2].

Profiling of Turing GPU and corresponding AI framework optimization.

2018.07-2019.06, ECNU, project of graduation design.

- Conduct the profiling from micro-benchmark to whole neural networks on new architecture GPUs, study the performance pattern and bottleneck.
- Adjust the `C++` and `PTX` code of Convolution computation with the feature of Tensor Core in `Tensor Flow` according to the profiling result to improve the throughput.

Simple compiler for C like programming language.

2018.01-2018.07, ECNU, project of Compiler Principle

- Implement the front-end and corresponding optimization procedures of a C-like language with `lex` and `yacc`, the generated IR is executed with an interpreter.

Skills

C/C++/CUDA/x86-64, ACM/ML/DL Algorithm and Data Structure, Linux, LaTeX, Python, Java, SQL/MongoDB, Unreal Engine 4, Verilog HDL.

Publications

1. Zihan Liu, Jingwen Leng, Minyi Guo et al. "Survey and Design of Paleozoic: a High-Performance Compiler Tool Chain for Deep Learning Inference Accelerator," In: *CCF Trans. HPC*, 2020.
2. Zihan Liu, Jingwen Leng, Minyi Guo et al. "DLFusion: An Auto-Tuning Compiler for Layer Fusion on Deep Neural Network Accelerator," In: *ISPA*, 2020.

Hobbies

Saxophone, PC/Console Game, Basketball