Haoyang Ma

Ph.D. Candidate in Computer Science and Engineering

(+852) 62200978
(+86) 17602273189
haoyanghkust@gmail.com
haoyang.ma@connect.ust.hk

EXPERIENCE

SiliconFlow (Formerly OneFlow), Beijing — *AI Framework Dev Intern*

Dec 2023 - Present

I am enhancing OneFlow's optimization process for stable diffusion models by introducing support for dynamic tensor shapes. This allows compiled models to be reused for receiving new inputs with varying shapes, resulting in savings in time and memory costs.

I am resolving operator alignment issues when transitioning PyTorch models to OneFlow. OneFlow is a user-friendly distributed deep learning framework that is compatible with PyTorch models. To seamlessly compile PyTorch models into low-level representations for distributed training and inference, it is necessary to align OneFlow operators with their corresponding counterparts in PyTorch. This ensures smooth compatibility between the two frameworks.

Unicloud Technology, Tianjin — *Big Data Dev Intern*

May 2019 - July 2019

I created Docker images for varying versions of a Kafka-related product that was under development at that time.

I assisted in performing regression testing by utilizing Docker containers.

INDUSTRIAL PROJECTS

OneFlow — *A Distributed Deep Learning Framework*

I refine next-generation OneFlow's high-level optimizations on deep-learning models by supporting dynamic tensor shapes. To achieve this goal, I analyze all special operators (e.g., unique, expand, and squeeze) that have no response to new tensor inputs and add such response as a temporary fix. Now I am reading literature, exploring more elegant approaches, and attempting to resolve this issue by helping OneFlow support an "Any" tensor shape type.

Onediff — *An out-of-the-box acceleration library for diffusion models*

I fix bugs inside torch2flow components to align PyTorch operators to OneFlow counterparts. My code patches enable Onediff users to accelerate PyTorch diffusion models with OneFlow more smoothly.

SKILLS

Language - C++, Python, C

Test - Fuzz Testing, Static Analysis

Compiler - LLVM, TVM

System - Linux, MLsys, Distributed System

AWARDS

Postgraduate studentship

Bronze prizes in ACMICPC Shenyang Division, ACMICPC Nanjing Division, and CCPC Jilin Division

LANGUAGES

Mandarin (native)

English (Full professional proficiency)

EDUCATION

The Hong Kong University of Science and Technology
PhD in Computer Science and Engineering
Sep 2021 - Present

Tianjin University
Bachelor in Engineering
Aug 2016 - Aug 2020

TVM — *An Open deep-learning compiler stack for CPU, GPU, and specialized accelerators*

I contribute code patches to resolve high-level optimization issues and model loading issues.

ACADEMIC PROJECTS

RecBi — *A Reinforcement-Learning-Based Bug Localizer for LLVM and GCC.*

RecBi can mutate C programs, collect code coverage, and analyze coverage differences before and after the mutation. The analysis feedbacks the A2C framework, a then-popular reinforcement-learning framework, to enhance the next-round mutation. This system can finally select several suspicious bug-triggered lines by analyzing the converging coverage difference and thus help localize the compiler bug.

HirGen — *A Generation-Based Fuzzer Deep Learning Compiler*

HirGen can generate diverse computational graphs from scratch, convert them into high-level IRs, and test deep learning compilers (e.g., TVM) with them. HirGen supports 58 operators in graph generation and has successfully detected 21 bugs in TVM.

PUBLICATIONS

Toward Understanding Solidity Compiler Bugs
Under Submission
Haoyang Ma, Wuqi Zhang, Qingchao Shen, Yongqiang Tian, Junjie Chen, Shing-Chi Cheung

A Survey of Modern Compiler Fuzzing
Preprint
Haoyang Ma

Fuzzing Deep Learning Compilers with HirGen
ISSTA' 23
Haoyang Ma, Qingchao Shen, Yongqiang Tian, Junjie Chen, Shing-Chi Cheung

A Comprehensive Study of Deep Learning Compiler Bugs
FSE' 21
Qingchao Shen, Haoyang Ma, Junjie Chen, Yongqiang Tian, Shing-Chi Cheung, Xiang Chen

Enhanced Compiler Bug Isolation via Memoized Search
ASE' 20
Junjie Chen, Haoyang Ma, Lingming Zhang