

YUHANG SONG

yuhangs@bu.edu | 857-600-1116 | <https://syh.one/about/>

EDUCATION

Boston University

Master of Science, Computer Science, GPA 3.96/4.0

Boston, MA

Sep 2022 - Present

South China University of Technology

Bachelor of Engineering, Software Engineering

Guangzhou, China

Sep 2017 - Jun 2021

EXPERIENCE

AI Framework Engineer

Intel Asia-Pacific Research & Development Ltd

Aug 2021 - Aug 2022

Shanghai, China

- Contributed to the oneDNN backend Graph Compiler, focusing on enhancing deep learning speeds, including fusion, loop rescheduling, graph inlining, and optimization using SIMD intrinsics.
- Implemented various elementwise and graph operators, including normalization and VNNI reordering. Developed graph patterns, such as MHA, to identify and optimize performance bottlenecks.

Software Engineer (Internship)

Tencent

Jun 2020 - Sep 2020

Guangzhou, China

- Developed a full-stack internal-use test case management system using Tornado and React.
- Created internal data visualization panels to display statistical data using Grafana.

PROJECT & RESEARCH

Accelerating Large-Scale GNN Training with Programmable SSDs

Sep 2023 - Present

- MS Thesis under Professor Vasiliki. Utilized Vitis HLS for FPGA onboard programming, enabling direct sampling without transferring data to host memory.
- Brings computation closer to SSD, conserving PCIe bandwidth for large-scale GNN sampling.
- Anticipate achieving higher performance compared to CPU sampling, with linear scalability.

Advanced Matrix Multiplication - Strassen Algorithm

Feb 2023 - May 2023

- Implemented and optimized the Strassen algorithm using OpenMP and SIMD intrinsic.
- Performance analysis and comparison of different matrix multiplication implementations.

Fine-Tuning Pretrained Vision Transformer Models

Feb 2023 - May 2023

- Fine-tuned models based on NAT, DiNAT, MaxViT, and DaViT, achieving image classification accuracy comparable to state-of-the-art models on the iNaturalist Dataset.
- Analyzed structural differences among various vision transformers.

Parallel Programming of "Group-By" Operation

Mar 2020 - Jun 2020

- Simulated "Group-By" operation of databases and calculated the group number of random numbers.
- Grouped data based on prefix and used OpenMP for parallel processing and MPI for cross-node parallel operation, distributing the computation task to each node and aggregating the results.

Design of Short Video Copyright Detection Algorithm

Sep 2019 - Nov 2019

- Designed an approximate detection method to search for short video clips in long videos.
- Used SIFT to extract feature points and FAISS vector database to build a long video feature library.
- Performed hierarchical matching and sliding window to achieve an accuracy of a 5-second range.

UWP-Based Property Management System

Jun 2019 - Sep 2019

- Created frontend of property management system using C++ and Universal Windows Platform.
- Developed backend with Python and Django, implementing RESTful APIs for communication.

HONORS & AWARDS

- **1st Scholarship of South China University of Technology**
- **2nd Prize of CCF National Olympiad in Informatics in Provinces**