Xiao Xia

Email: xxia.200051@gmail.com | Tel: (858)539-3552 | Web: https://xiaoxia42.github.io/
LinkedIn: https://www.linkedin.com/in/xiao-xia-4b8a471a6/ | Github: https://github.com/xiaoxia42

Educations

University of California, San Diego Sept 2021 - March 2023
– M.S. in Computer Engineering, GPA 3.97/4.0

Fudan University Sept 2016 – Jun 2020
– B.Eng. in Microelectronics

Technical Skills

– **Programming Languages**: C, Python, C++, Java, Golang, JavaScript, TypeScript, CUDA
– **Tech Stacks**: React, SQL, MongoDB, TVM, Linux, Docker, Spring, MATLAB, Verilog, Mission Planner, PyTorch, PostgreSQL, OpenAPI, Kubernetes, Garden, LXD, AVX, SSE, MPI

Experience

Tesla Inc, Software Development Engineer Intern Jun 2022 – Aug 2022
Product: Tesla CSSP (Cloud Self-Service Portal)
– Independently developed and deployed the disk management system which will be used by **57,000** employees of Tesla.
– Developed seven front-end web modules and pages written in **React**, which include the modules of add/delete extra disks in the instance creation page and disk management pages.
– Implemented eight back-end APIs in **Golang**, which include publishing and receiving jobs from **RabbitMQ**, querying and persisting data into the **PostgreSQL** database, and managing virtual machines on the **LXD** cluster.
– Optimized the *instance_disk* **PostgreSQL** table with many duplicate recordings into *project_disk*, *instance_disk*, and *instance_node* tables to decouple the database operations and reduce the server load by about 5%.

Unity Technology, Software Development Engineer Intern May 2021 – Aug 2021
Product: Unity Distribution Portal: https://distribute.dashboard.unity.com/
– Developed front-end web modules written in **React** to help create the game distribution plan, which will benefit the game developers to publish their games more manageably to stores with about **one billion+** mobile gamers.
– Developed back-end **GraphQL APIs** in **Golang** to help game developers publish their games to the Samsung Galaxy Store and Huawei store and get the published result from the store using the callback APIs.
– Integrated the UDP website with the APIs provided by OneTrust Cookie Compliance.
– Upgraded the Support Form by integrating the Zendesk Support Form into the website.

Amazon Web Service, Software Development Engineer Intern Jun 2020 – May 2021
Product: MXNet: https://github.com/apache/incubator-mxnet TVM: https://github.com/apache/tvm
– Developed **Cython** and Ctype **FFI** (Foreign Function Interface) for **MXNet** Ops (Operators).
– Implemented dozens of Meta Ops, complemented their test cases with **TVM**, and developed several Meta type functions for inferring the type and shape of Ops.
– Developed Meta **IR Passes** to bring the Bring Your Own Codegen (**BYOC**) feature into Meta, making the compiler more scalable for self-defined languages.
– Implemented **BERT** Model Module and **ResNet** Model using the Ops provided by Meta and benchmarked these deep learning models to improve the functionality of Meta.

Projects

Parallel Computation, UC San Diego March 2022 – Jun 2022
– Implementing the **DGEMM** algorithm, memory packing, kernel size tuning, **vectorization**, and loop unrolling improved the Peak Gflops of 1024x1024 matrix multiply from 0.014 Gflops to 21.44 Gflops on AWS t2.micro instance.
– Using **CUDA** implemented the **CUTLASS** algorithm, memory coalescing, shared memory allocation, kernel size tuning, and pointer aliasing optimization increased the Peak Gflops of a 1024x1024 matrix multiply from 0.014 Gflops to 916.47 Gflops on an AWS p2.xlarge instance with a K80 GPU.
– Paralleled the computation of n=8000, i=8000 Aliev-Panfilov cardiac simulation using **MPI** and implemented loop fusion, separate initialization improved the computation performance from 11.5 to 2085 Gflops on SDSC's 384 core supercomputer.

Awards

– 2018 Intel Cup Undergraduate Electronic Design Contest - Embedded System Design Invitational Contest 3^{rd} place
– 2018 Fudan Ascendas Cup (1^{st} place)