Tianyu Sun

CONTACT INFORMATION

https://tianyu-sun.github.io mobile: +1 (858) 214-0007
https://www.linkedin.com/in/tianyu-sun e-mail: tterrysun@gmail.com

EDUCATION

University of California, San Diego, La Jolla, CA
M.S., Computer Science **Sept. 2019 – Mar. 2021**
University of Science and Technology Beijing, Beijing, China
B.E., Computer Science **Aug. 2015 – June 2019**

RELEVANT PROFESSIONAL EXPERIENCE

SambaNova Systems
Software Engineer **Apr. 2021 – Present**

- Co-designed and developed infra for large-scale DP and MP, supporting multi-dimensional DP, fine-grained data distribution management, and effcient cross-socket traffic planning.
- Designed and developed infra for DP on heterogeneous hardware architecture at compile time, including bit-file packing and consistency checking infra. US patent granted.
- Enhanced resource and performance modeling at compile time and extended compiler resource modeling for multiple architectures with different compute resources and memory bandwidth.
- Developed a config development toolkit, which boosted the efficiency of config development by 10X. The toolkit was integrated into the main workflow and adopted by hundreds of applications.
- Launched GUI for test coverage monitoring, making testing quality tracking much easier.

Aibee US
Research Intern **June 2020 – Sept. 2020**

- Designed and implemented a model that improves the vehicle Re-ID performance by considering pose. Increased performance from 85.4% to 97.3% on TPR@FPR=0.01. Converted the PyTorch model to a Caffe model and shipped it to intelligent parking lot production.
- Developed an internal tool for new car model discovery using Re-ID architecture.

Tencent
Research Intern @ Tencent AI Lab **Dec. 2018 – Aug. 2019**

- Participated in Virtual Host project, which aims at generating a virtual host for game streaming and weather broadcasting. Developed modules for face segmentation and alignment. Module adopted by a million-DAU mobile application.
- Worked on developing a robust and efficient system for generating realistic videos with generative adversarial networks. Proposed a stat-of-the-art face reenactment model.

Institute of Automation, Chinese Academy of Sciences
Research Intern @ National Laboratory of Pattern Recognition **June 2017 – Sept. 2018**

- Proposed a method of increasing the accuracy of gait recognition by heightening the frame rate with generative adversarial networks. The publication can be seen in *Frame-GAN*.

SELECTED PROJECTS

Lego-Serverless Distributed Platform

- Developed Lego-Serverless Platform, an event handling and function creation platform for modern serverless services.
- Designed a two-level load balancing mechanism, a high-level round-robin load balancer, and a middle-level Raft load balancer. Responsible for implementing data pipeline and high-level load balancing. Designed and developed data infrastructure based on Kafka and CouchDB.
- Lego-Serverless provides RESTful API for function and event CRUD. Additional management functions like user authentication and function authorization are supported too. The platform can handle 2,000 QPS based on single-node testing on AWS EC2 instance.

Distributed Storage System

- Built a distributed storage system based on the Raft consensus algorithm using Golang.
- Implemented leader election, file replication, and data persistence mechanisms. Designed RPC for communication between nodes. A crash recovery mechanism is implemented as well.

SKILLS

Frameworks and Tools
MLIR, PyTorch, Kafka, TensorFlow, OpenCV

Programming Languages
C++, Python, Golang, C, SQL, JAVA, Haskell