Tianyu(Terry) 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
M.S., Computer Science **Sept. 2019 – Mar. 2021**

University of Science and Technology Beijing
B.E., Computer Science **Aug. 2015 – June 2019**

PROFESSIONAL EXPERIENCE

SambaNova Systems
Senior Software Enigneer **Apr. 2021 – Present**

- Designed and implemented highly-scalable infrastructure for large-scale data and model parallelism, supporting multi-dimensional data parallelism, fine-grained hierarchical data distribution management, and efficient cross-socket traffic planning.
- Designed and implemented compiler infrastructure for data parallelism on heterogeneous hardware, including bit-file packing and consistency checking support. US patent granted.
- Improved resource and performance modeling at compile time and extended compiler resource modeling for multiple architectures with varying compute resources and memory bandwidths.
- Brought up a config development toolkit, resulting in a 10X increase in efficiency and widely adopted by hundreds of applications.

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 leveraging pre-trained Re-ID models.

Tencent
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 face segmentation and alignment modules, which were subsequently adopted by a million-DAU mobile application.
- Proposed a state-of-the-art face reenactment model as part of a highly robust and efficient video generation system using generative adversarial networks.

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

- Proposed a method of boosting the accuracy of gait recognition by increasing the frame rate with generative adversarial networks. Published as *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 consisting of a high-level round-robin load balancer and a middle-level Raft load balancer. Implemented 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 in a single-node testing on an AWS EC2 instance.

Distributed Storage System

- Built a distributed storage system using Golang based on the Raft consensus algorithm.
- 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