# SHIYI CAO

shiyicao314@gmail.com ⋄ +41 764116188

## EDUCATION

**ETH Zurich**                                                                                     *Sept. 2020 - May. 2023*
Department of Computer Science
M.S. in Computer Science
Relevant Courses:

- Advanced Operating System, Advanced System Lab (6/6), Cloud Computing Architecture (5.75/6), Information Security Lab (5.75/6), Design of Parallel and High-Performance Computing (5.5/6)

**Shanghai Jiao Tong University**                                                                   *Sept. 2016 - 2020*
School of Electronic Information and Electrical Engineering
B.S. in Computer Science and Technology                                                             GPA: 88/100
Relevant Courses:

- Computer System Architecture (90), Computer Architecture Lab (94), Computing Theory (97), Mathematical Foundations of Computer Science (92), Operating Systems Lab (90), Cloud Computing (93)

## RESEARCH EXPERIENCE

**Generalized Pipeline Parallelism for Model Training**                                             *Current*
*Catalyst, CMU, Advisor: Zhihao Jia*

- Leading the end-to-end implementation for enabling generalized parallelism training strategies.

**Network-accelerated General-purpose RPC Framework**                                               *Aug. 2021 - Aug. 2022*
*SPCL Lab, ETH Zurich, Advisor: Salvatore Di Girolamo and Prof. Torsten Hoefler*

- Offloading data deserialization (ProtoBuf) to SmartNIC for efficient RPC framework.
- Designing the deserialization algorithm to enable parallel and streaming processing on the NIC.

**Deep Reinforcement Learning in Distributed Metadata Management**                                  *Sept. 2018 - Jan. 2019*
*Advanced Network Laboratory, SJTU, Advisor: Prof. Xiaofeng Gao*

- Introduced for the first time deep reinforcement learning in distributed metadata management.
- Proposed an adaptive fine-grained metadata management scheme AdaM, leveraging deep reinforcement learning.
- Conducted experiments on real-world data traces and compared AdaM with strong baselines.
- Demonstrated that AdaM can address the trade-off between load balance and locality preservation cost-effectively and is highly adaptive to time-varying access pattern.
- Published the paper in ICPP'19 (International Conference on Parallel Processing) as the *first author*.

## SELECTED PROJECTS

**Barrelfish OS Development**                                                                       *Mar. 2022 - Jun. 2022*
*Advanced Operating System Course by David Cock and Prof. Timothy Roscoe*

- Implemented our own memory management, paging, message passing, inter-core communication etc. on Barrelfish research operating system education version.

- Implemented and benchmarked the Network stack.

### Distributed DL Training on Bagua
*Oct. 2021 - Jan. 2022*

*DS3Lab, ETH, Advisor: Jiawei Jiang and Prof. Ce Zhang*

- Port, improve and benchmark existing distributed deep compression training algorithms to Bagua, a deep learning trainging acceleration framework for PyTorch.

### High-performance Image Compression Implementation
*Mar. 2021 - June. 2021*

*Advanced System Lab Course Project*

- Designed highly optimized implementations of the whole SPIHT image compression pipeline, leveraging techniques such as SIMD vectorization, memory rearrangement, and blocking.
- Our best optimized version achieves a runtime speedup of 100x and 200x for encoding and decoding respectively compared with the baseline implementation.

### High-performance Parallel Priority Queues
*Sept. 2020 - Dec. 2020*

*Design of Parallel and High-performance Computing Course Project*

- Implemented several lock-free parallel priority queues in C++, code optimized.
- Conduct in-depth analysis of a variety of parallel priority queues and developed a benchmarking framework to facilitate the experiments.

## TALKS & PRESENTATIONS

### Participant, International Conference on Parallel Processing
*Aug. 2019*

- Made oral presentation for the accepted paper Adam: An adaptive fine-grained scheme for distributed metadata management.

## PUBLICATIONS

1. **Shiyi Cao**, Yuanning Gao, Xiaofeng Gao, and Guihai Chen. Adam: An adaptive fine-grained scheme for distributed metadata management. In *International Conference on Parallel Processing (ICPP)*, 2019.

## AWARDS

- Academic Excellence Scholarship (Second Class), 2016-2017
- Academic Excellence Scholarship (Third Class), 2018-2019
- Meng Minwei International Exchange Fund (12000RMB), 2019

## SKILLS

| | |
|---|---|
| **English Proficiency** | GRE: 329 + 4.0 (V:160 Q:169 AW:4.0), TOEFL: 109 |
| **Programming** | C, C++, SSE/AVX, Python, PyTorch, Tensorflow |
| **Softwares** | Latex, Matlab, Unity3D |
| **GitHub** | https://github.com/caoshiyi |