Zirui Wang

617-992-5314 | wzr@bu.edu | Github | Homepage

Education

Boston University — Boston, MA
Master of Science in Computer Science — *Sept. 2022 – Present*

Hangzhou Dianzi University — Hangzhou, China
Bachelor of Engineering in Computer Science and Technology — *Sept. 2018 – Jun. 2022*

Research Projects

ELF Compression Algorithm Acceleration — Jul. 2023 – Oct. 2023
Remote work with Prof.Cheng — *Boston, MA*

- Optimized ELF compression algorithm to increase the compression rate for 32-bit float floating-point numbers from around 1.2 to 1.25. Achieved parallel acceleration of ELF algorithm on SmartSSD, and used P2P transfer to significantly improve the IO performance, with the compression throughput reaching 1.3 GB/s in a single compute unit.

Stream Processing System with State Disaggregation — Feb. 2023 – May 2023
Boston University — *Boston, MA*

- Designed and implemented a streaming data processing system capable of automated task allocation, loading balancing, and state storage disaggregation.
- Implemented operators that handle the computation of stream data, including **stateless** operators such as Filter, KeyBy, Map, Union, and **stateful** operators such as Reduce, Count, and Sliding Window.
- Developed Task Manager that achieves state storage management, data I/O, and distribution. Implemented Control Plane to achieve load balancing, state routing, and other functions.
- Wrote test scripts in Java to test the latency of the system using local storage as well as remote state storage. Used **Prometheus** for real-time status monitoring of system latency.

Video Highlight Detection Based on Deep Learning Method — Sept. 2020 – Jul. 2021
Hangzhou Dianzi University — *Hangzhou, China*

- Used a hierarchical temporal context coding structure and proposed a low-rank decomposition-based video and audio fusion method to improve the detecting accuracy and speed. Successfully **exceeding the SOTA level** and improving the mAP value from 0.584 to 0.629. Paper accepted by **ICCV2021**.

Projects

Key-Value storage database engine — Mar. 2023 – Jun. 2023

- Based on **Bistcask**, developed a log-structure based KV storage database engine.
- Implemented basic CURD operations and support transactions.
- Optimized memory index (support ART, B+ Tree, B Tree), optimized file IO using MMap to speed up file reading, provided database state query to speed merge process.
- Completed support for HTTP, Redis data structures and the Redis protocol.

Sharded Key-Value storage system with fault-tolerant — May 2022 – Sept. 2022

- Based on the **Raft** algorithm, implemented leader election and log replication mechanisms.
- Developed **fault tolerance** mechanisms, including log compaction and snapshotting.
- Designed and implemented **sharding** mechanisms for distributing data across multiple servers.

Alibaba Tianchi Global Video Cloud Innovation Challenge — Mar. 2021 – Jul. 2021

- According to the competition problem, the Fast Instance Segmentation + Mask Refinement method is proposed to solve the problems of motion blur, frequent scene switching, and character edge refinement, making it possible to perform segmentation quickly and accurately.
- The competition ended up with a bronze prize (ranking 5/2904).

Publication

- ***ICCV '21*** *Temporal Cue Guided Video Highlight Detection with Low-Rank Audio-Visual Fusion.*
  International Conference on Computer Vision (**ICCV '21**).
  Qinghao Ye*, Xiyue Shen*, Yuan Gao*, **Zirui Wang***, Qi Bi, Ping Li, Guang Yang.

- ***VLDB '24*** *Everything You Always Wanted to Know About Storage Compressibility of Pre-Trained ML Models but Were Afraid to Ask.*
  50^{th} International Conference on Very Large Data Bases (**VLDB'24**), (*Under review*).
  Zhaoyuan Su, Ammar Ahmed, **Zirui Wang**, Ali Anwar, Yue Cheng.

Technical Skills

Language: Java, Go, Python, C/C++

Framework: PyTorch, Flink

Tools&Platforms: Git, Docker, AMD Vitis, Redis, Linux, SQL, Github, GitLab