Jinzhen Wang

+1-551-580-4404 jw447@njit.edu jinzhen-wang-54b6b0a8 jw447.github.io Jersey City, NJ

I'm a senior Ph.D. candidate in Computer Engineering, focusing on data compression and HPC data management. I'm a motivated and result-driven individual that can work effectively under limited supervision. I have experience working in a collaborative environment and am responsible for leading and managing research projects. I'm also proficient in writing and presentation at all levels.

Research interest

Data Compression, Scientific Data Management, Machine Learning, Scientific Visualization

Skills

Programming languages	Python, C, C++
Software & tools	ADIOS, HDF5, Foresight, Git, LaTeX, MarkDown
Communication	English (fluent), Mandarin (native)

Experience

Graduate research intern **Jan 2023 — present**
Los Alamos National Laboratory *Los Alamos, New Mexico, US*
Conduct a user study on the image quality of volume rendering. Correlate user feedback with image quality assessment.

Graduate research intern **Sep 2022 — Dec 2022**
Los Alamos National Laboratory *Los Alamos, New Mexico, US*
Explore the impact of lossy data compression on various volume rendering techniques as well as color modes. Design a user study on the users' feedback on the image quality.

Summer research intern **May 2022 — Aug 2022**
Los Alamos National Laboratory *Los Alamos, New Mexico, US*
Investigate the effect of lossy data compression on visualization for large-scale cosmology data.

Education

Ph.D. in Computer Engineering, *New Jersey Institute of Technology, NJ, US* Jan 2018 - Present

M.S. in Electrical Engineering, *New Jersey Institute of Technology, NJ, US* Sep 2015 - May 2017

B.S. in Internet of Things (Electrical Engineering), *Shandong University, China* Sep 2011 - May 2015

Research

Improving Progressive Retrieval for HPC Scientific Data using Deep Neural Network
Motivation — Progressive data retrieval framework has been developed to address the gap between the compute and I/O. However, the performance has been suffering from aggressive error control.

- We provide an in-depth investigation of the recently developed progressive retrieval framework;
- We propose two designs of prediction models (D-MGARD and E-MGARD) to estimate the amount of retrieved data size based on error bounds.
- We evaluate our proposed solutions and show that our solution is shown to read significantly less data (5% - 40% w. D-MGARD, 20% - 80% w. E-MGARD).

Jinzhen Wang

+1-551-580-4404 jw447@njit.edu jinzhen-wang-54b6b0a8 jw447.github.io Jersey City, NJ

Analyzing the Impact of Lossy Data Reduction on Volume Rendering of Cosmology Data
Motivation — HPC scientific users seek to reduce the amount of storage needed for application data while preserving enough quality for analysis and visualization. The key question is how much we can reduce the data before noticeable artifacts are introduced.

- Perform a thorough analysis of the effects of 3 cutting-edge error-bounded lossy compression techniques on large-scale scientific data;
- Investigate more than 20 image quality assessment metrics on quantifying visualization artifacts for volume rendering.

High-Ratio Lossy Compression: Exploring the Auto-encoder to Compress Scientific Data
Motivation — Auto-encoder has been widely used in image compression and dimension reduction and achieving superior performance, motivating the adoption of scientific data compression.

- Propose and design the first error-bounded auto-encoder compression framework that achieves superior reduction performance over traditional scientific lossy compression techniques.
- Optimize the auto-encoder compression framework via data pre-processing techniques and back-end lossless encoding techniques.
- Evaluate the compression framework over 20 large-scale scientific datasets (100+ GB) from more than 10 disciplines to show the generality.

Publications

- **Wang, J.**, Liang, X., Whitney, B., Chen, J., Gong, Q., He, X., Wan, L., Klasky, S., Podhorszki, N. Liu, Q. Improving Progressive Retrieval for HPC Scientific Data using Deep Neural Network. submitted to 2023 IEEE 39th international conference on data engineering (**major revision**).

- **Wang, J.**, Chen, Q., Liu, T., Liu, Q., He, X. zPerf: A Statistical Gray-box Approach to Performance Modeling and Extrapolation for Scientific Lossy Compression. IEEE Transactions on Computers (**major revision 2nd round**).

- Liu, T., **Wang, J.**, Liu, Q., Alibhai, S., Lu, T., & He, X. (2021). High-ratio lossy compression: Exploring the autoencoder to compress scientific data. IEEE Transactions on Big Data.

- Qin, Z., **Wang, J.**, Liu, Q., Chen, J., Pugmire, D., Podhorszki, N., & Klasky, S. (2020). Estimating Lossy Compressibility of Scientific Data Using Deep Neural Networks. IEEE Letters of the Computer Society, 3(1), 5-8.

- **Wang, J.**, Liu, T., Liu, Q., He, X., Luo, H., & He, W. (2019). Compression ratio modeling and estimation across error bounds for lossy compression. IEEE Transactions on Parallel and Distributed Systems, 31(7), 1621-1635.

- Liu, T., Alibhai, S., **Wang, J.**, Liu, Q., He, X., & Wu, C. (2019, August). Exploring transfer learning to reduce training overhead of hpc data in machine learning. In 2019 IEEE International Conference on Networking, Architecture and Storage (NAS) (pp. 1-7). IEEE.