Shuo Zhang

✉ sz3177@columbia.edu 📞 (857) 350-2596 in shuo-zhang-leo ⓞ Inverse0

Education

Columbia University   *Sept 2023 – Dec 2024*
MS in Computer Science
- GPA: 3.95/4.00

Boston University   *Sept 2020 – May 2023*
BA in Computer Science and Economics
- GPA: 3.83/4.00; Honors: Magna Cum Laude

Publications

CAVE: Concurrency-Aware Graph Processing on SSDs   May 2024
Tarikul Islam Papon, Taishan Chen, *Shuo Zhang*, Manos Athanassoulis
[10.1145/3654928](https://doi.org/10.1145/3654928) ↗ (SIGMOD 2024)

Research Experience

Research Assistant   *New York, NY*
Columbia University Database Research Group   *May 2024 – present*
- Conducted research on data cleaning, data repairing, and Text-to-SQL conversion.
- Contributed to the development of Cocoon, an advanced data-cleaning tool utilizing large language models (LLMs) to automate and improve the data-cleaning process.
- Evaluated Cocoon's performance by comparing its accuracy, recall, and F1-score with existing data cleaning systems such as Holoclean and Raha on benchmark datasets, demonstrating superior performance across all metrics.

Research Assistant   *New York, NY*
Columbia University Internet Real-Time Lab   *Sept 2023 – present*
- Conducted research on Internet of Things (IoT) devices, systems, and applications.
- Enhanced IoT system manageability and streamlined regulatory compliance. Implementing a policy server to describe security, energy efficiency, and other constraints of systems through Python.
- Built a prototype for demonstration purposes, allowing users to construct an IoT model based on a relationship-centric Policy System.

Research Assistant   *Boston, MA*
Boston University Data-intensive Systems and Computing Lab   *June 2022 – May 2024*
- Conducted research on databases, data management, and storage systems.
- Contributed to the development of CAVE, an innovative graph processing system leveraging SSD parallelism to significantly enhance the efficiency of large-scale graph processing.
- Demonstrated the effectiveness of CAVE through rigorous testing on diverse real-world graph datasets across multiple SSD platforms, resulting in performance improvements up to an order of magnitude over existing systems such as Mosaic and GridGraph, and three orders of magnitude over GraphChi.
- Co-authored a research paper detailing the innovations and findings of the CAVE project, emphasizing its novel intra-subgraph and inter-subgraph parallelization approaches and their impact on computational efficiency in graph processing. Conducted experiments for CAVE and other baseline systems and authored the evaluation section for CAVE.

Industry Experience

Research Intern   *San Jose, CA*
VisionX LLC   *June 2022 – Aug 2022*

- ○ Developed an advanced object detection model utilizing PyTorch and a pre-trained YOLO framework, significantly enhancing the quality of a product designed for crop pest and disease identification. This model achieved a 98% accuracy rate and increased processing speed by 180%.
- ○ Optimized the model based on real-world product applications, employing fine-tuning techniques to improve performance on blurry and low-resolution images.
- ○ Constructed a Java-based graphical user interface (GUI) to enhance user experience.

Projects

Research Project on Video Enhancement

- ○ Developed a method that enhances both video resolution and smoothness by integrating advanced image super-resolution techniques with video interpolation strategies, significantly boosting video quality. The method achieved a Peak Signal-to-Noise Ratio (PSNR) of 34.72 on a subset of the Vimeo-90K benchmark.
- ○ Designed a novel frame selection algorithm that ensures a natural flow of video while preserving the original frame rate.

Database and Web application Project

- ○ Developed a shopping website, utilizing HTML for design and Flask for creating a RESTful API that interfaces with a MongoDB backend.
- ○ Implemented SQL to develop essential e-commerce functionalities such as shopping carts and order tracking systems, mirroring those found on established shopping platforms.

Research Project on Time Series Data Analysis   *2023*

- ○ Researched the potential of machine learning models for time series data forecasting.
- ○ Experimented with various state-of-the-art machine learning models, comparing their performance with traditional methods like autoregressive integrated moving average (ARIMA) and advanced techniques such as Long Short-Term Memory (LSTM) networks.
- ○ Evaluated on real-world GDP data and achieved an 85% improvement on Root Mean Squared Error (RMSE) compared with ARIMA prediction.

Research Project on Time Series Data Analysis   *2022*

- ○ Collaborated with Boston City Councilor Ruthzee to analyze small landlord data, identifying properties with the potential to join affordable housing programs.
- ○ Developed a strategy to integrate owner-occupied and small landlord properties, previously unenrolled, into affordable housing databases.
- ○ Cleaned and consolidated existing housing databases to identify eligible properties. Created models to cluster housing units and predict fair rental price ranges.

Technologies

Language: Chinese (Native)

Programming Languages: C/C++, HTML, Java, OCaml, Python, R, SQL

Technologies: MongoDB, MySQL, PostgreSQL, Stata