

Han Zhang

Email: han.harry.zhang@gmail.com

EDUCATION

Shanghai Jiao Tong University (SJTU)

2020 - 2024 (*Expected*)

Bachelor of Engineering in Computer Science

Member of **ACM Class**, a selective CS program for the top 5% students.

Major GPA: 91.87/100 (5/35)

- Excellent professional and mathematical courses performance:

Calculus I,II: 95.5,99.75/100, Linear Algebra: 97/100, Probability and Statistics: 96/100, Data Structure: 97/100, Machine Learning: 94/100, Deep Learning: 97/100, Advanced Algorithm: 94/100

University of California, Los Angeles(UCLA)

2024 - 2029(*Expected*)

Computer Science PhD student, Advised by [Prof. Wei Wang](#)

PUBLICATION

GFS: Graph-based Feature Synthesis for Prediction over Relational Databases *Preprint*

Han Zhang, Gan Quan, David Wipf, Weinan Zhang

TaDA workshop at VLDB 2024

4DBInfer: A 4D Benchmarking Toolbox for Graph-Centric Predictive Modeling on Relational DBs *Preprint*

Minjie Wang, Quan Gan, David Wipf, Zhenkun Cai, Ning Li, Jianheng Tang, Yanlin Zhang, Zizhao Zhang, Zunyao Mao, Yakun Song, Yanbo Wang, Jiahang Li, **Han Zhang**, Guang Yang, Xiao Qin, Chuan Lei, Muhan Zhang, Weinan Zhang, Christos Faloutsos, Zheng Zhang

RESEARCH EXPERIENCE

Ma Lab, Carnegie Mellon University

Supervised by [Prof. Jian Ma](#)

Aug 2023 - Mar 2024

- Focus on computational genomics, particularly in unraveling gene expression complexities and transcriptional regulation mechanisms. Utilize Convolutional Neural Networks to distill long DNA sequences into dense vectors, and integrate transformer-based models with cell-specific data, including 2D Hi-C contact maps. Develop novel attention layer and transformer variants to incorporate 2D Hi-C data, aiming to predict gene expression. Key achievements include interpreting model weights, benchmarking gene expression outputs, and identifying biological patterns.

APEX Lab, Shanghai Jiao Tong University & AWS Shanghai AI Lab

Supervised by [Prof. Weinan Zhang](#) and [Dr. David Wipf](#)

Jul 2022 - Now

- (Jul 2022 - Dec 2022) Developed a feature space alignment method for tabular data using hard negative contrastive pretraining. This involved identifying hard negatives based on row similarities and employing contrastive learning with anchors, their corrupted views, and hard negatives to achieve a more uniform and well-aligned feature representation.
- (Dec 2022 - Now) Led the *GFS* project, innovating in deep learning over relational databases. Addressed the complex challenge of column predictions in target tables using surrounding relational data. Developed a novel framework that interprets databases as heterogeneous graphs, integrating single-table models for both embedding and prediction. This approach streamlined feature engineering, bypassing traditional labor-intensive processes. Demonstrated enhanced performance over conventional methods in experiments with four real-world datasets, showcasing *GFS*'s effectiveness and overcoming previous methods' limitations.

RESEARCH INTEREST

My primary research interests are centered around Artificial Intelligence and Machine Learning, particularly in the subfields of Data Mining, Graph Neural Networks, and LLM. Beyond these, I am enthusiastic about the interdisciplinary applications of AI, especially its intersection with biology. This includes but is not limited to genomics, protein science, and molecules, wherein AI algorithms have the potential to revolutionize our understanding and treatment of various biological problems.

OTHER EXPERIENCE

Teaching Assistant of Machine Learning

instructed by Prof. Weinan Zhang

Feb 2023 - Jun 2023

Assign five coding assignments for students about five different topics about Machine Learning, answer students' questions and correct their assignments.

Participants of Chinese Physics Olympiad(CPhO), Xuejun High School

Jul 2017 - Oct 2019

I Have won second prize twice in Zhejiang Province, and I have learned a lot of knowledge about physics. This experience greatly improved my learning ability and helped me get grades 100 and grade 95 in my Physics (1) and Physics (2) courses at SJTU

PROJECTS

Java-and-C-like Language Compiler (~10K lines in Java) [\[github\]](#)

The compiler can convert codes of program to an AST, then after semantic check to LLVM-like IR, and eventually to RISC-V assembly. Few optimizations were done to the Back-end, including mem2reg, graph-coloring.

Ray tracer(~4K lines in Rust) [\[github\]](#)

Construct a raytracer renderer in rust, you can set the position parameters of the objects and the camera parameters yourself to get the rendered real-world like image through the renderer.

RISC-V CPU of Tomasulo Architecture (~3K lines in Verilog) [\[github\]](#)

I-cache, and a 2-bit saturating counter branch predictor implemented to improve the performance. The CPU could run successfully on an FPGA board.

HONORS AND AWARDS

Ruiyuan-Hongshan Scholarship, top 5% students in ACM class

2023

Jin Long Yu Scholarship

2021

Zhiyuan Honorary Scholarship, top 5% students

2021, 2022, 2023

SKILLS

Programming Languages: C/C++ , Python , Java , Rust

Hardware: Verilog

Math: Calculus, Linear Algebra, Mathematical Logic, Measure Theory, Graph and Combinatorics

VOLUNTEER ACTIVITIES

- **Library Collection Relocation Volunteer.** Assisted in relocating library collections, including packing, moving, and organizing books and materials, ensuring their safe transfer and accessibility.
- **School Traffic Facilitation Volunteer.** Managed and directed traffic around the school, ensuring safe passage for students during peak hours and contributing to efficient traffic flow.

MISC

- 11th place in the *natural ecological* campus singing competition in Shanghai Jiao Tong University.
- I helped our team reach the semifinals in the basketball tournament as a starter in high school.