Runsheng Bai

brs21@mails.tsinghua.edu.cn / +86 13520785205

EDUCATION

School of Software, Tsinghua University (THU), Beijing *Sep 2021 – Jul 2025 (Expected)*

Major in Software Engineering

Overall GPA: 3.95/4.0; **Major GPA:** 3.95/4.0

GPA Ranking: 2nd among 79 students

Core courses: *Programming Methodology (A), Data Structure (A), Introduction to Algorithms (A-), Operating System (A, scored 96 in the final exam, 1/112), Advanced topic in Linear Algebra (A, top 5 in the final exam), Probability and Statistics (A), Introduction to Complex Analysis (A+), Ordinary Differential Equations (A+).*

Honors and Awards:

Software Creative Innovation Competition at Tsinghua University *2023*

Dong Fang Electronic Corporation Scholarship *2023*

Scholarship For Freshmen year(1 out of 82) *2022*

PUBLICATION

- Kaichao You, **Runsheng Bai**, Meng Cao, Jianmin Wang, Mingsheng Long. depyf: Open the Opaque Box of PyTorch Compiler for Machine Learning Researchers. *Journal of Machine Learning Research.*

RESEARCH EXPERIENCE

SKM: Quantization using scaled kmeans with mixed precision | UTA | *Research Assistant* *March 2024 – Present*

Advisor: Associate Professor Qiang Liu, Department of Computer Science, UT Austin

- Proposed a novel quantization technique called SKM based on kmeans clustering.
- Incorporated: 1. a trainable scaling vector on undifferencial grouping operator; 2. greedy algorithm to solve the optimal bit allocation accross rows, into the sensitivity-based kmeans quantization proposed by SqueeseLLM.
- Developed the code individually, and conducted various experiment to test the effectiveness of our proposed method. For instance, we reduced the layer quantization error by an average of 75.6% and decreased the perlexity of quantized 3-bit Llama-7B on wikitext2 from 6.33 to 6.19.
- Expected to publish the paper as the first author.

Project Depyf: A Tool for Understanding torch.compile() | THU | *Research Assistant* *Oct 2023 – March 2024*

Advisor: Associate Professor Mingsheng Long, School of Software, THU

- Developed depyf, a useful tool to open the opaque box of torch.compile() by decompiling the python bytecode and hooking functions.
- Assisted in designing the code logic to handle the reachability issues that had been caused by JUMP ABSOLUTE in Python bytecode, and helped deal with the with ... except ... code block.
- Conducted various experiments, including verifying the correctness of the decompiler and assessing the ability of depyf to open the models compiled by torch (Models are from Hugging face, Torchbench and Timm).
- Achieved a 100% success rate on the tested 85 python decompilation cases and 140 deep learning models, while the best among the others achieved 78 out of 85 and 27 out of 140, which demonstrated its potential.
- Published the paper to JMLR as the second author. Github link: https://github.com/thuml/depyf.git

Low-Light Image Reconstruction from Event Cameras | THU | *Participant* *Oct 2022 – May 2023*

Advisor: Associate Professor Yue Gao, School of Software, THU

- Proposed using event camera data which shows light changes over time to reconstruct images in low-light conditions.
- Individually built a reconstruction network by combining U-Net and Vision Transformer ideas via Pytorch, extracting event information with convolution and merging it into the image for subsequent reconstruction.
- Transformed 63.82% of the unrecognizable images in the given dataset into relatively recognizable ones, and achieved an "A" grade in the project assessment ; gained valuable skills in machine learning through this project.

SKILLS

- **Proficient:** C/C++, Python, Pytorch, Git, Latex.
- **Mastered:** Java, JavaScript, html/css.
- **Known:** Matlab, R, Win32 assembly.