Youliang HUANG

yliangh@bu.edu

PHO302, 8 St. Mary's Street
College of Engineering
Boston University
Boston, 02215, MA

https://youliangh.github.io

EDUCATION

Boston University — United States
Ph.D. in Computer Engineering — Sep. 2025 – Jul. 2030 (expected)

The Hong Kong University of Science and Technology (Guangzhou) — China
Master of Philosophy in Data Science and Analytics — Sep. 2023 – Jul. 2025

Guangzhou University — China
Bachelor of Engineering in Network Engineering — Sep. 2019 – Jun. 2023

SELECTED HONORS & AWARDS

Student Grant — **2026**
USENIX OSDI

Distinguished Computer Engineering Fellowship — **2025**
Boston University

Postgraduate Redbird Studentship — **2023, 2024**
The Hong Kong University of Science and Technology (Guangzhou)

The First-class Scholarship at Guangzhou University — **2020, 2021, 2022, 2023**
Guangzhou University, top 2.5% of all students

RESEARCH EXPERIENCE

Ph.D. Research, Boston University — **Sep. 2025 – Present**
Role: Ph.D. Student, **Supervisor:** Asst. Prof. Yigong Hu

Conducting full-time research on system vulnerability/reliability with bug detection and failure diagnosis. The current research is focusing on how to detect failures that might happen in complex cloud or distributed systems. The last project is about addressing performance issues caused by the contention of user-defined resources. The paper has been published on OSDI'26.

M.Phil. Research, The Hong Kong University of Science and Technology — **Sep. 2023 – Jul. 2025**
Role: M.Phil. Student, **Supervisor:** Asst. Prof. Zeyi Wen

Studying decomposition-based LLM compression and its application. Research involves repurposing LLM compression to fine-tuning, improving performance of LLM compression on downstream tasks, and improving inference efficiency by increasing more throughput and reducing memory footprint. Papers have been published on top-tier conferences.

DeepSE Lab, The Hong Kong University of Science and Technology — **Jan. 2023 – Sep. 2023**
Role: Research Assistant, **Supervisor:** Assoc. Prof. Sung Hun Kim

Studied cold-start problems in recommender systems with comprehensive benchmarks and in-depth analysis over simple rule-based data augmentation methods and contrastive learning methods, revealing combinations of simple data augmentation methods can outperform state-of-the-art contrastive learning in some scenarios. Published and presented two papers in TheWebConf.

Institute for Systems Rheology, Guangzhou University **Nov. 2021 – Aug. 2023**
Role: Research Assistant, **Supervisor:** Prof. Xue-Feng Yuan
Project: Ultra-large-scale Parallel Computing Test Verification

Profiled a computational fluid dynamics program called OpenFOAM where the goal is to generate billions of mesh on high-performing computing clusters in NSCC-TJ/GZ. Solved the out-of-memory problem incurred by inappropriate distributed communication. Contribute a solution to avoid memory spike during large-scale (over 1 billion) mesh generation in blockMesh and snappyHexMesh.

Project: Virtual Environment Terrain Generation in Digital Twin and Emulator

Focused on enhancing the controllability over terrains generated by generative adversarial networks. Proposed a novel GAN-based terrain modeling technique that disentangles embeddings in generator's latent space to achieve progressive modification on the resulting terrain during the generation. Obtained a software copyright and published a paper on *Computer & Graphics*.

TEACHING EXPERIENCES

Boston University Graduate Teaching Assistant
ENG EC327 Introduction to Software Engineering 2026 Fall

The Hong Kong University of Science and Technology (Guangzhou) Graduate Teaching Assistant
DSAA5003 Automatic Machine Learning 2024 Fall

PROFESSIONAL SERVICES

Reviewers: Eurographics'24, ICLR'25 '26 '27, AAAI'26 '27, NeurIPS'26, WWW'27

INVITED TALK

Seminar Talk **Harvard System Group**
Diagnosing Performance Issues in Application-Defined Resources Jul. 2026

SKILLS

Language: Mandarin Chinese (Native), English (IELTS: 8/8.5/6.5/6 Overall: 7.5)
Programming: Python (proficient), C/C++ (proficient), Java (good knowledge), CUDA (good knowledge)
Tools: PyTorch (proficient), GNU Toolchains (good knowledge), Git, Linux, LaTeX

PUBLICATIONS

[1]. Yigong Hu, **You-Liang Huang**, Haodong Zheng, Yicheng Liu, Dedong Xie, Baris Kasikci, "Diagnosing Performance Issues in Application-Defined Resources", *In 20th USENIX Symposium on Operating Systems Design and Implementation (OSDI 26)*: 647--664.

[2]. **You-Liang Huang**, Xinhao Huang, Chengxi Liao, and Zeyi Wen, "DeInfer: Efficient Parallel Inferencing for Decomposed Large Language Models", *The 63rd DAC, The Chips to Systems Conference (DAC)*, 2026.

[3]. Xinhao Huang, **You-Liang Huang**, Zeyi Wen, “Representation Drift Compensation: A Zero-Cost Enhancement for LLM Decomposition”, *The 43rd International Conference on Machine Learning (ICML)*, 2026.

[4]. Xinhao Huang, **You-Liang Huang**, Zeyi Wen, “SoLA: Leveraging Soft Activation Sparsity and Low-Rank Decomposition for Large Language Model Compression”, *Proceedings of the AAAI Conference on Artificial Intelligence. 2025, 39(16):* 17494-17502.

[5]. Minping Chen, **You-Liang Huang**, Zeyi Wen, “Towards Efficient Low-order Hybrid Optimizer for Language Model Fine-tuning”, *Proceedings of the AAAI Conference on Artificial Intelligence. 2025, 39(22)*: 23605-23613.

[6]. Peilin Zhou, Chao Liu, Jing Ren, Xinfeng Zhou, Yueqi XIE, Meng Cao, Zhongtao Rao, **You-Liang Huang**, Dading Chong, Junling Liu, Jae Boum KIM, Shoujin Wang, Raymond Chi-Wing Wong, Sunghun Kim, “When Large Vision Language Models Meet Multimodal Sequential Recommendation: An Empirical Study”, *Proceedings of the ACM on Web Conference 2025* (*WWW '25*): 275-292.

[7]. Peilin Zhou*, **You-Liang Huang***, Yueqi Xie, Jingqi Gao, Shoujin Wang, Jaeboum Kim and Sunghun Kim, “Is Contrastive Learning Necessary? A Study of Data Augmentation vs Contrastive Learning in Sequential Recommendation”, *Proceedings of the ACM on Web Conference 2024* (*WWW '24*): 3854–3863.

[8]. Peilin Zhou, Meng Cao, **You-Liang Huang**, et al. “Exploring Recommendation Capabilities of GPT-4V(ision): A Preliminary Case Study.” *arXiv preprint* arXiv:2311.04199 (2023).

[9]. **You-Liang Huang**, and Xue-Feng Yuan. "StyleTerrain: A novel disentangled generative model for controllable high-quality procedural terrain generation." *Computers & Graphics* 116 (2023): 373-382.

[10]. Wei Li, Lujun Li, Hao Gu, **You-Liang Huang**, Mark G. Lee, Shengjie Sun, Wei Xue, Yike Guo, “MoE-SVD: Structured Mixture-of-Experts LLMs Compression via Singular Value Decomposition”, *Proceedings of Machine Learning Research, 2025, volume 267, 35209-35230.*

[11]. Lujun Li, Cheng Lin, Dezhi Li, **You-Liang Huang**, Wei Li, Tianyu Wu, Jie Zou, Wei Xue, Sirui Han, Yike Guo, “Efficient Fine-Tuning of Large Models via Nested Low-Rank Adaptation”, Proceedings of the *IEEE/CVF International Conference on Computer Vision (ICCV), 2025, volume 267, 22252-22262.*