

MUYANG LI

✉ muyangli@mit.edu · 🌐 lmxxy · in Muyang Li · 🌐 lmxxy.me

🎓 EDUCATION

Massachusetts Institute of Technology

Ph.D. student at EECS, advised by [Prof. Song Han](#)

Sep. 2023 – Present

Cambridge, MA

Carnegie Mellon University

Master of Science in Robotics, advised by [Prof. Jun-Yan Zhu](#)

Aug. 2021 – May 2023

Pittsburgh, PA

Shanghai Jiao Tong University

Bachelor of Engineering in Computer Science

Sep. 2016 – Jun. 2020

Shanghai, China

- Member of [ACM Class](#), an elite CS program for the top 5% talented students.

🔍 RESEARCH INTERESTS

My research interest is in the intersection of machine learning, system, and computer graphics.

I am currently working on building efficient and hardware-friendly generative models with its applications in computer vision and graphics.

📄 PUBLICATIONS [GOOGLE SCHOLAR](#)

- [1] Haocheng Xi*, Shuo Yang*, Yilong Zhao, Chenfeng Xu, Muyang Li, Xiuyu Li, Yujun Lin, Han Cai, Jintao Zhang, Dacheng Li, Jianfei Chen, Ion Stoica, Kurt Keutzer, and Song Han, *Sparse VideoGen: Accelerating Video Diffusion Transformers with Spatial-Temporal Sparsity (ICML 2025)* [📄](#)
- [2] Enze Xie*, Junsong Chen*, Yuyang Zhao†, Jincheng Yu†, Ligeng Zhu†, Chengyue Wu, Yujun Lin, Zhekai Zhang, Muyang Li, Junyu Chen, Cai Han, Bingchen Liu, Daquan Zhou, and Song Han, *SANA 1.5: Efficient Scaling of Training-Time and Inference-Time Compute in Linear Diffusion Transformer (ICML 2025)* [📄](#)
- [3] Muyang Li*, Yujun Lin*, Zhekai Zhang*, Tianle Cai, Xiuyu Li, Junxian Guo, Enze Xie, Chenlin Meng, Jun-Yan Zhu, and Song Han, *SVDQuant: Absorbing Outliers by Low-Rank Components for 4-Bit Diffusion Models (ICLR 2025 Spotlight)* [📄](#)
- [4] Enze Xie*, Junsong Chen*, Junyu Chen, Han Cai, Haotian Tang, Yujun Lin, Zhekai Zhang, Muyang Li, Ligeng Zhu, Yao Lu, Song Han, *SANA: Efficient High-Resolution Image Synthesis with Linear Diffusion Transformers (ICLR 2025 Oral)* [📄](#)
- [5] Junyu Chen*, Han Cai*, Junsong Chen, Enze Xie, Shang Yang, Haotian Tang, Muyang Li, Yao Lu, and Song Han, *Deep Compression Autoencoder for Efficient High-Resolution Diffusion Models (ICLR 2025)* [📄](#)
- [6] Muyang Li*, Tianle Cai*, Jiaxin Cao, Qinsheng Zhang, and Han Cai, Junjie Bai, Yangqing Jia, Ming-Yu Liu, Kai Li, and Song Han, *DistriFusion: Distributed Parallel Inference for High-Resolution Diffusion Models (CVPR 2024 Highlight)* [📄](#)
- [7] Han Cai, Muyang Li, Zhuoyang Zhang, Qinsheng Zhang, Ming-Yu Liu, and Song Han, *Condition-Aware Neural Network for Controlled Image Generation (CVPR 2024)* [📄](#)
- [8] Muyang Li, Ji Lin, Chenlin Meng, Stefano Ermon, Song Han and Jun-Yan Zhu, *Efficient Spatially Sparse Inference for Conditional GANs and Diffusion Models (NeurIPS 2022 & T-PAMI 2023)* [📄](#)
- [9] Yihan Wang, Muyang Li, Han Cai, Wei-Ming Chen and Song Han, *Lite Pose: Efficient Architecture Design for 2D Human Pose Estimation (CVPR 2022)* [📄](#)
- [10] Muyang Li, Ji Lin, Yaoyao Ding, Zhijian Liu, Jun-Yan Zhu, and Song Han, *GAN Compression: Efficient Architectures for Interactive Conditional GANs (CVPR 2020 & T-PAMI 2021)* [📄](#)
- [11] Jindong Jiang*, Xiuyu Li*, Zhijian Liu, Muyang Li, Guo Chen, Zhiqi Li, De-An Huang, Guilin Liu, Zhiding Yu, Kurt Keutzer, Sungjin Ahn, Jan Kautz, Hongxu Yin, Yao Lu, Song Han, and Wonmin Byeon, *Token-Efficient Long Video Understanding for Multimodal LLMs* [📄](#)

👤 EXPERIENCES

NVIDIA <i>Summer Intern</i> Work with Prof. Song Han 4-bit quantization for diffusion models (SVDQuant, ICLR'25) and helped develop SANA (ICLR'25) .	May 2024 – Aug. 2024 Santa Clara, CA
NVIDIA <i>Summer Intern</i> Work with Prof. Song Han and Ming-Yu Liu Distributed diffusion models (DistriFusion, CVPR'24) and condition-aware networks (CAN, CVPR'24).	Jun. 2023 – Aug. 2023 Shanghai, China
CMU Generative Intelligence Lab <i>Master's Student</i> Advisor: Prof. Song Han and Prof. Jun-Yan Zhu Sparse image editing engine to accelerate GANs and diffusion models (SIGE, NeurIPS'22&T-PAMI'23).	Aug. 2021 – Present Pittsburgh, USA
OmniML Inc. <i>Summer Intern</i> Work with Prof. Song Han Efficient vision model deployment on edge devices (e.g., Jetson devices and mobiles).	May 2022 – Aug. 2022 San Jose, CA
Dawnlight Inc. <i>Data Scientist</i> Work with Prof. Song Han and Prof. Jia Li Efficient pose estimation (Lite Pose, CVPR'22).	Jul. 2020 – Jul. 2021 Shanghai, China
MIT HAN Lab <i>Research Assistant</i> Advisor: Prof. Song Han and Prof. Jun-Yan Zhu General conditional GANs' compression framework (GAN Compression, CVPR'20&T-PAMI'21).	Jul. 2019 – Jan. 2020 Cambridge, MA

💻 OPEN-SOURCED PROJECTS

🔗 mit-han-lab/nunchaku (2.2K Stars) <i>Python/CUDA</i> 4-bit diffusion model inference engine.	May 2024 – Present
🔗 mit-han-lab/ComfyUI-nunchaku (1.5K Stars) <i>Python/CUDA</i> ComfyUI plugin for Nunchaku.	Mar. 2025 – Present
🔗 mit-han-lab/deepcompressor <i>Python</i> Quantization library for LLMs and diffusion models.	May 2024 – Present
🔗 mit-han-lab/gan-compression (1.1K Stars) <i>Python</i> A general conditional GAN Compression framework.	Jul. 2019 – Apr. 2020
🔗 mit-han-lab/distrifuser (>600 Stars) <i>Python</i> A distributed framework to accelerate diffusion models with multiple GPUs.	Oct. 2023 – Feb. 2024
🔗 lmxyy/sige <i>Python/C++/CUDA/Metal</i> A sparse engine to accelerate image editing for GANs and diffusion models.	Jul. 2021 – Nov. 2022
🔗 mit-han-lab/litepose <i>Python</i> A light-weighted pose estimation model that could run on mobile devices.	Mar. 2021 – Jun. 2022

HONORS AND AWARDS

<i>Best Pitch Award</i> , Award on MARC 2025	Jan. 2025
<i>Seneff-Zue CS Fellowship (\$98K)</i>	Sep. 2023
<i>Gold Medal</i> , Award on CCPC2017 Harbin Regional, Ranked 10 th	Oct. 2017
<i>Gold Medal</i> , Award on ICPC2017 Qingdao Regional, Ranked 5 th	Nov. 2017
<i>3rd Runner-up</i> , Award on ICPC2017 Jakarta Regional	Nov. 2017
<i>1st Runner-up</i> , Award on Singing Competition of Zhiyuan College in SJTU	Dec. 2017
<i>Jin Long Yu Fellowship</i> , Award for top 1% students	Dec. 2017
<i>1st Runner up's Coach</i> , Award on ICPC 2018 Pathom Regional	Nov. 2018
<i>A-Class School-level Scholarship</i> , Award for top 1% students	Dec. 2018
<i>Zhiyuan Honorary Scholarship</i> (3 times), Award for top 5% students	2016, 2017, 2018
<i>Honorable Mention</i> , Award for 2019 American Interdisciplinary Contest in Modeling (ICM)	Jan. 2019

ACADEMIC SERVICES

- Conference Reviewer: ICML, ICLR, NeurIPS, ICCV, CVPR, SIGGRAPH Asia
- Journal Reviewer: T-PAMI, IJCV, TVCJ, TCSVT

TEACHING

SJTU ACM-ICPC Coach	Jun. 2018 – Apr. 2019
TA at SJTU Data Structure (CS147)	Mar. 2018 – May 2018

SKILLS

Programming Languages: C++/C/CUDA = Python > Java

Deep Learning Packages: PyTorch, TensorFlow, TVM, TensorRT

Languages: English - Proficient, Mandarin - Native speaker, Japanese - Amateur

Other: Pop Singing