

Junhan Zhu

✉ zhujunhan@westlake.edu.cn | 🌐 alrightlone.github.io |  LinkedIn |  GitHub

EDUCATION

Westlake University

Bachelor of Engineering in Electronic and Information Engineering

Hangzhou, China

Sept. 2023 – Present

- **GPA: 4.044/4.3 (Rank: 1/23)**
- **Early Admission:** Admitted by skipping the last year of high school due to academic excellence.

Nanyang Technological University (NTU)

Exchange Student in School of Electrical and Electronic Engineering (EEE)

Singapore

Jan. 2026 – May 2026

EXPERIENCE

Duke Center for Computational Evolutionary Intelligence, Duke University

Research Intern

Advisor: *Prof. Yiran Chen*

Incoming

- **Research Focus:** Efficient generative models.

MMLab, Nanyang Technological University (NTU)

Research Intern

Advisor: *Prof. Ziwei Liu*

Jan. 2026 - May 2026

- **Research Focus:** Inference acceleration for native multimodal model.

ENCODE Lab, Westlake University

Research Intern

Advisor: *Prof. Huan Wang*

Mar. 2025 - Dec. 2025

- **Research Focus:** Model compression and acceleration for diffusion models.

PUBLICATIONS

* Corresponding author. † Equal contribution (Co-first author).

[1] OBS-Diff: Accurate Pruning For Diffusion Models in One-Shot

Junhan Zhu, Hesong Wang, Mingluo Su, Zefang Wang, Huan Wang*

 arXiv |  Project |  Code

ICLR 2026 (Lead Author) | Jan. 2026

[2] Cross-Resolution Diffusion Models via Network Pruning

Jiaxuan Ren†, Junhan Zhu†, Huan Wang*

 arXiv |  Project

CVPR 2026 Findings (Co-first Author) | Feb. 2026

[3] LVOmniBench: Pioneering Long Audio-Video Understanding Evaluation for Omnimodal LLMs

Keda TAO, Yuhua Zheng, Jia Xu, Wenjie Du, Kele Shao, Hesong Wang, Xueyi Chen, Xin Jin, Junhan Zhu, Bohan Yu, Weiqiang Wang, Jian liu, Can Qin, Yulun Zhang, Ming-Hsuan Yang, Huan Wang*

 arXiv |  Project |  Code

Under Review | Mar. 2026

SELECTED PROJECTS

SparAlloc: Modular Framework for LLM Sparsity Allocation

Lead Developer |  Source Code

May 2025

- Built a modular Python framework that decouples sparsity allocation from importance metrics in LLM pruning, enabling flexible layer-wise compression strategies.

AWARDS

- **SenseTime Scholarship (Only 30 Recipients Nationwide)** June 2026
SenseTime | Highly prestigious national scholarship awarded to top undergraduate students in AI fields
- **Hongyi Scholarship** Dec. 2024
Westlake University | Awarded for meaningful social activities
- **Outstanding Bachelor's Student** Oct. 2025
Westlake University | Awarded for outstanding academic performance

- **Outstanding Undergraduate Scholarship** Oct. 2024 & Oct. 2025
Westlake University | *Awarded for top undergraduate students (¥2000/year)*
- **Innovation Award** Oct. 2024 & Oct. 2025
Westlake University | *Awarded for innovative research achievements*

ACADEMIC SERVICE

- **Journal Reviewer:** *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*
- **Conference Reviewer:** *NeurIPS 2026*

SKILLS

- **Technical Skills:** Python, C/C++, PyTorch, Hugging Face, Linux, Git, \LaTeX
- **Languages:** Chinese (Native), English (Fluent, IELTS 7.0)