Junhan Zhu

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

Education

The Chinese University of Hong Kong (CUHK) — *Hong Kong SAR*
Incoming Ph.D. Student in Computer Science and Engineering — **Aug. 2027 – Expected 2031**
- **Advisor:** Prof. Weiyang Liu

Westlake University — *Hangzhou, China*
Bachelor of Engineering in Electronic and Information Engineering — **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) — *Singapore*
Exchange Student in School of Electrical and Electronic Engineering (EEE) — **Jan. 2026 – May 2026**

Experience

Duke Center for Computational Evolutionary Intelligence, Duke University — *Advisor*: *Prof. Yiran Chen*
Research Intern — **Incoming**
- **Research Focus:** Efficient generative models.

MMLab, Nanyang Technological University (NTU) — *Advisor*: *Prof. Ziwei Liu*
Research Intern — **Jan. 2026 - May 2026**
- **Research Focus:** Inference acceleration for native multimodal model.

ENCODE Lab, Westlake University — *Advisor*: *Prof. Huan Wang*
Research Intern — **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)