

Beier Zhu

 beier.zhu@ntu.edu.sg  GoogleScholar  beierzhu.github.io

Abbreviations

NTU	Nanyang Technological University	MM	ACM International Conference on Multimedia
THU	Tsinghua University	ICCV	IEEE International Conference on Computer Vision
USTC	University of Science and Technology of China	AAAI	AAAI Conference on Artificial Intelligence
CVPR	Conference on Computer Vision and Pattern Recognition	TIP	IEEE Transactions on Image Processing
NeurIPS	Conference on Neural Information Processing Systems	TSG	IEEE Transactions on Smart Grid
ICLR	International Conference on Learning Representations	EE	Electrical Engineering

Research Interests **Robust and fair learning with provable guarantees**, including imbalanced learning, group robustness, OOD generalization, and fast diffusion solvers.
Multimodal foundation models (VLMs, MLLMs, diffusion models) with a focus on robust adaptation, faithful reasoning, and controllable generation.

Working Experience Research Scientist, NTU, Singapore Mar 2025 – Present
Researcher, SenseTime, Beijing Jul 2019 – Nov 2020

Education Ph.D. in Computer Science, NTU, Singapore Jan 2021 – Jan 2025
supervisor: Prof. Hanwang Zhang
M.S. in EE, THU, Beijing. Sep 2016 – Jul 2019
B.E. in EE, THU, Beijing. Sep 2012 – Jul 2016

Selected Publications Listed in chronological order. *: equal contribution. †: corresponding author.

First-Authored Publications

1. Reducing class-wise performance disparity via margin regularization.
B. Zhu, K. Zhao, J. Cui, Q. Sun, Y. Zhou, X. Yang, H. Zhang
2. Parallel diffusion solver via residual dirichlet policy optimization.
R. Wang, Z. Li*, B. Zhu*, L. Yuan, H. Zhang, X. Yang, X. Chang, C. Zhang*
3. Adaptive stochastic coefficients for accelerating diffusion sampling.
R. Wang, B. Zhu*, J. Li, L. Yuan, C. Zhang*
4. Distilling parallel gradients for fast ODE solvers of diffusion models.
B. Zhu, R. Wang*, T. Zhao, H. Zhang, C. Zhang*
5. Project-probe-aggregate: efficient fine-tuning for group robustness.
B. Zhu, J. Cui, H. Zhang, C. Zhang
6. Robust fine-tuning of zero-shot models via variance reduction.
B. Zhu, J. Cui, H. Zhang
7. Enhancing zero-shot vision models by label-free prompt distribution learning and bias correcting.
X. Zhu, B. Zhu*, Y. Tan, S. Wang, Y. Hao, H. Zhang*
8. Selective vision-Language subspace projection for few-shot CLIP.
X. Zhu, B. Zhu*, Y. Tan, S. Wang, Y. Hao, H. Zhang*
9. Generalized logit adjustment: calibrating fine-tuned models by removing label bias in foundation models.
B. Zhu, K. Tang, Q. Sun and H. Zhang
10. Prompt-aligned gradient for prompt tuning.
B. Zhu, Y. Niu, Y. Han, Y. Wu and H. Zhang
11. Debiased fine-tuning for vision-language models by prompt regularization.
B. Zhu, Y. Niu, S. Lee, M. Hur, H. Zhang
12. Cross-domain empirical risk minimization for unbiased long-tailed classification.
B. Zhu, Y. Niu, X. Hua and H. Zhang
13. Polygonal region detection.
Beier Zhu, Rui Zhang
14. Structure-coherent deep feature learning for robust face alignment.
C. Lin, B. Zhu*, Q. Wang, R. Liao, C. Qian, J. Lu and J. Zhou*

Corresponding-Authorized Publications

- TIP 2026** 15. Hybrid granularity distribution estimation for few-shot learning: statistics transfer from categories and instances.
S. Wang, T. Qi, X. Zhu, Y. Hao, †B. Zhu, H. Zhang, M. Wang
- ICLR 2026** 16. Real-time motion-controllable autoregressive video diffusion.
K. Zhao, J. Shi, †B. Zhu, J. Zhou, X. Shen, Y. Zhou, Q. Sun, H. Zhang
- ICLR 2026** 17. Look carefully: adaptive visual reinforcements in multimodal large language models for hallucination mitigation.
X. Zhu, K. Zhao, L. Yi, S. Wang, Z. Wang, †B. Zhu, H. Zhang, X. He
- AAAI 2026** 18. DEPO: dual-efficiency preference optimization for LLM agents.
S. Chen, M. Zhao, L. Xu, Y. Zhao, †B. Zhu, H. Zhang, S. Zhao, C. Lu
- MM 2025
Oral** 19. Benchmarking and bridging emotion conflicts for multimodal emotion reasoning.
Z. Han, †B. Zhu, Y. Xu, P. Song, X. Yang
- ICCV 2025** 20. Unsupervised visual chain-of-thought reasoning via preference optimization.
K. Zhao, †B. Zhu, Q. Sun, H. Zhang
- TSG 2019** 21. Fault location for radial distribution network via topology and reclosure-generating traveling waves.
S. Shi, †B. Zhu, A. Lei and X. Dong
- TSG 2018** 22. Fault classification for transmission lines based on group sparse representation.
S. Shi, †B. Zhu, S. Mirsaeidi and X. Dong

Others

- ICLR 2026** 23. CoDi: subject-consistent and pose-diverse text-to-image generation.
Z. Gao, B. Zhu, L. Yao, J. Yang, Y. Tai
- ICLR 2026** 24. GuardAlign: robust safety alignment in multimodal large language models.
X. Zhu, B. Zhu, J. Fang, S. Wang, Y. Zhang, X. Wang, X. He
- ICLR 2026** 25. PMI: flow-based inversion correction via proximal operator.
C. Wang, B. Zhu, C. Zhang
- ICLR 2026** 26. Streaming drag-oriented interactive video manipulation: drag anything, anytime!
J. Zhou, Y. Zhou, K. Zhao, Q. Xu, B. Zhu, R. Hong, H. Zhang
- AAAI 2026** 27. Hierarchical semantic alignment for image clustering.
X. Zhu, B. Zhu, Y. Li, J. Fang, S. Wang, K. Zhao, H. Zhang
- Preprint 2025** 28. Generative distribution distillation.
J. Cui, B. Zhu, Q. Xu, X. Xu, P. Chen, X. Qi, B. Yu, H. Zhang, R. Hong
- Preprint 2025** 29. Generalized kullback-leibler divergence loss.
J. Cui, B. Zhu, Q. Xu, Z. Tian, X. Qi, B. Yu, H. Zhang, R. Hong
- NeurIPS 2025
Spotlight** 30. Enhancing CLIP robustness via cross-modality alignment.
X. Zhu, B. Zhu, S. Wang, K. Zhao, H. Zhang
- ICCV 2025** 31. Dynamic multimodal prototype learning in vision-language models.
X. Zhu, S. Wang, B. Zhu, M. Li, Y. Li, J. Fang, Z. Wang, D. Wang, H. Zhang
- CVPR 2025** 32. Devils in middle layers of large vision-language models: interpreting, detecting and mitigating object hallucinations via attention lens.
Z. Jiang, J. Chen, B. Zhu, T. Luo, Y. Shen, X. Yang
- CVPR 2025** 33. Stylestudio: text-driven style transfer with selective control of style elements.
M. Lei, X. Song, B. Zhu, H. Wang, C. Zhang
- CVPR 2024** 34. Classes are not equal: an empirical study on image recognition fairness.
J. Cui, B. Zhu, X. Wen, X. Qi, B. Yu, H. Zhang
- AAAI 2023
Oral** 35. Leveraging modality-specific representations for audio-visual speech recognition via reinforcement learning.
C. Chen, Y. Hu, Q. Zhang, H. Zou, B. Zhu, E. Chng

- Mentoring and Supervision** **Xingyu Zhu**, P.h.D at USTC
Supervisor for papers accepted at MM 2024, NeurIPS 2024/2025, ICCV 2025, AAAI 2026, and ICLR 2026.
- Kesen Zhao**, Ph.D. at NTU
Supervisor for paper accepted at ICCV 2025 and ICLR 2026.
- Ruoyu Wang**, Ph.D. at Westlake University
Supervisor for papers accepted at ICCV 2025 and NeurIPS 2025.
- Zhiyuan Han**, Ph.D. at USTC

Supervisor for paper accepted at MM 2025.

Sirui Chen, Ph.D. at Tongji University

Supervisor for paper accepted at AAAI 2026.

Zhanxin Gao, Ph.D. at Nanjing University

Supervisor for paper accepted at ICLR 2026.

Service	<i>Journal Reviewer:</i> TPAMI, TIP, TMM, TCSV, TCE, IJCV, Information Fusion <i>Conference Reviewer:</i> CVPR, ICCV, AAAI, MM, UAI, AISTATS, NeurIPS (Top Reviewer 2025), ICLR, ICML.	
Awards	<i>Honorable Mention</i> , Nanyang Speech Forum, Singapore. <i>AISG PhD Fellowship</i> , National Research Foundation, Singapore. <i>First Price Scholarship</i> , THU, China. <i>Scholarship</i> , China Scholarship Council, China.	2023 2021 2018 2014
Internship	Research Intern, SenseTime, Beijing Research Intern, Tencent YouTu Lab, Shanghai Research Intern, MeiTuan, Beijing	Mar 2018 – Oct 2018 Jul 2018 – Sep 2018 Aug 2017 – Dec 2017

[CV compiled on 2026-01-31]