

Zhihao Zhang

 Email |  Homepage |  LinkedIn |  Google Scholar

EDUCATION

University of North Carolina at Chapel Hill

Chapel Hill, NC, USA

- **Ph.D. in Computer Science**
Advisor: Prof. Xiaoming Liu

May. 2026 -

Michigan State University

East Lansing, MI, USA

- **Ph.D. in Computer Science**
Advisor: Prof. Xiaoming Liu

Aug. 2024 - May. 2026

Xi'an Jiaotong University

Xi'an, Shaanxi, China

- **M.S. in Computer Science and Technology** Ranking: 2/173
- **B.Eng. in Computer Science and Technology**
Honors Science Program in Qian Xuesen Honors College.

Sept. 2021 - Jun. 2024

Sept. 2017 - Jun. 2021

University of California, Berkeley

Berkeley, CA, USA

- Berkeley International Study Program

Aug. 2019 - Jan. 2020

RESEARCH INTERESTS

3D Vision: monocular 3D object detection, 3D shape understanding, multi-modal 3D perception

Spatial Intelligence: spatial reasoning and 3D grounding in vision-language models

World Models: self-supervised video pre-training via latent prediction, and predictive, action-conditioned modeling of 3D scenes

PUBLICATIONS & MANUSCRIPTS

(* means Equal Contribution)

- [1] **Zhihao Zhang**, Gengwei Zhang, Tianlong Chen, Xiaoming Liu. RefineAny3D: Depth Refinement as Semantic Alignment for Monocular 3D Detection. **NeurIPS 2026**.
- [2] **Zhihao Zhang**, Abhinav Kumar, Huaizhi Qu, Tianlong Chen, Xiaoming Liu. Rethinking Geometric Depth in Monocular 3D Object Detection: A Projection-Consistent Reformulation. **NeurIPS 2026**.
- [3] **Zhihao Zhang**, Abhinav Kumar, Girish Chandar Ganesan, Xiaoming Liu. Unleashing the Power of Chain-of-Prediction for Monocular 3D Object Detection. **CVPR 2026 Highlight**.
- [4] **Zhihao Zhang**, Abhinav Kumar, Xiaoming Liu. Towards Intrinsic-Aware Monocular 3D Object Detection. **CVPR 2026**.
- [5] **Zhihao Zhang***, Shengcao Cao*, Yu-Xiong Wang. TAMM: TriAdapter Multi-Modal Learning for 3D Shape Understanding. **CVPR 2024**.
- [6] **Zhihao Zhang***, Yiwei Chen*, Weizhan Zhang, Caixia Yan, Qinghua Zheng, Qi Wang, Wangdu Chen. Tile Classification Based Viewport Prediction with Multi-modal Fusion Transformer. **ACM-MM 2023**.

AWARDS & HONORS

SCHOLARSHIPS

- Engineering Co-op Scholarship, Michigan State University May. 2026
- **National Scholarship, Ministry of Education in China (Top 0.1%)** Nov. 2023

- Tencent Scholarship (Top **0.1%**) Nov. 2023
- Outstanding Freshman Scholarship, Xi'an Jiaotong University (Top 10%) Nov. 2021
- First Prize Scholarship, Xi'an Jiaotong University (Top 10%) Nov. 2022

HONORS

- **Outstanding Graduate Thesis**, Xi'an Jiaotong University (Top 1% of all graduates) Jun. 2024
- Outstanding Master Graduate, Xi'an Jiaotong University (Top 5% of all graduates) Jun. 2024
- Outstanding Bachelor Graduate, Xi'an Jiaotong University (Top 5% of all graduates) Jun. 2021

RESEARCH EXPERIENCE

Feed Quality, TikTok

San Jose, US

Research Intern. Mentor: Dr. [Jin Yang](#)

May 2026 – Aug. 2026

Research Project: Pre-training the next-generation video foundation model via auto-regressive modeling.

Computer Vision Lab, Michigan State University

East Lansing, US

Research Assistant. Advisor: Prof. [Xiaoming Liu](#)

Aug. 2024 - May. 2026

Research Project: Improve Monocular 3D Object Detection (Mono3D) from:

- *Sequential Reasoning*: Conventional Mono3D detectors overlook the inter-correlations of 3D attributes. Proposed **MonoCoP**, a novel detection framework that introduces a Chain-of-Prediction that progressively and conditionally predicts 3D attributes.
- *Intrinsic Awareness*: Demonstrated that 3D detectors generalize poorly to unseen camera intrinsics. Designed **MonoIA**, an intrinsic aware framework that transforms raw intrinsics into a semantic representation and enables robust detection across diverse camera settings.
- *Camera Height Robustness*: Explored ego height robustness in Mono3D models from single height data. Introduced **CHARM3R**, which generalizes to unseen ego heights through simple averaging of regressed depths and ground based object depths.

Computer Vision Group, University of Illinois at Urbana-Champaign

US

Research Assistant. Advisor: Prof. [Yuxiong Wang](#)

Jun. 2023 - Jan. 2024

- *3D Understanding*: Explored how to effectively leverage 2D foundation models to enhance 3D representation pre-training. Developed **TAMM**, a novel multimodal learning framework that consistently improves 3D representations across diverse 3D encoder architectures and downstream tasks.

ACADEMIC SERVICE

Conference Reviewer

- The International Conference on Learning Representations (ICLR 2024-2026)
- The International Conference on Computer Vision (ICCV 2025)
- Conference on Neural Information Processing Systems (NeurIPS 2025)
- ACM International Conference on Multimedia (ACM-MM 2023-2024)

LANGUAGE & SKILLS

Language	Mandarin (Native)
	English
Programming	Python, C/C++, Bash, LaTeX, MATLAB, HTML/CSS
Model Frameworks	LLMs (e.g., Llama, Qwen), MLLMs (e.g., LLaVA, Qwen-VL), VLMs (e.g., CLIP, BLIP) 2D/3D Detectors (e.g. DETR, RT-DETR)