

Wenyan Cong

Ph.D. Student ◊ Department of Electrical and Computer Engineering ◊ University of Texas at Austin
wycong@utexas.edu ◊ <https://wenyancong.com/> ◊ Google Scholar

RESEARCH INTERESTS

Computer Vision:

- ▷ **Multimodal World Foundation Model, Generative AI, Neural Rendering, Computational Photography**

Embodied AI:

- ▷ **World-Action Models, Vision-Language-Action Models**

Machine Learning:

- ▷ **Vision-Language Foundation Models (Effectiveness and Efficiency), Reinforcement Learning**

EDUCATION

The University of Texas at Austin (UT Austin), TX, USA

Aug. 2022 - Oct. 2026 (Expected)

Ph.D., Electrical and Computer Engineering

- ▷ Advised by Prof. [Zhangyang \(Atlas\) Wang](#)

- ▷ **Honors: Outstanding Paper Honorable Mention, MLSys'25; Best Paper Award, CVPR'25 AI4CC Workshop;**

Shanghai Jiao Tong University (SJTU), Shanghai, China

Sept. 2019 - Mar. 2022

M.S., Computer Science and Engineering

- ▷ Advised by Prof. [Li Niu](#) and Prof. [Liqing Zhang](#)

- ▷ **Graduated with Highest Honors**

Shanghai Jiao Tong University (SJTU), Shanghai, China

Sept. 2015 - Jun. 2019

B.E., Computer Science and Engineering

PROFESSIONAL EXPERIENCE

NVIDIA Research, CA, USA

Feb 2026 – Present

Research Scientist Intern

Mentor: [Dr. Yan Wang](#)

- ▷ Interaction-aware driving world-action model (WAM) for safer ego-action generation (ongoing)
- ▷ RL objective design to improve VLA robustness in last-mile autonomous driving scenarios (in submission)
- ▷ OOD reasoning benchmark for driving VLAs released on nvidia/PhysicalAI-Autonomous-Vehicles

Adobe Research, CA, USA

May 2025 – Jan 2026

Research Scientist Intern

Mentor: [Dr. Yiqun Mei](#), [Dr. Miloš Hašan](#)

- ▷ Scaling Video Generation for Unified Lighting and View Control (in submission)

Meta FAIR, CA, USA

May 2024 – Dec 2024

Research Scientist Intern

Mentor: [Weiyao Wang](#), [Dr. Matt Feiszli](#)

- ▷ Efficient 3D reconstruction from videos (CVPR'25 AI4CC, 3DV'26)
- ▷ Generalizable 4D gaussian splatting

Snap Research, NY, USA

May 2023 – Dec 2023

Research Intern, Video Frame Interpolation

Mentor: [Dr. Jian Wang](#)

- ▷ High-quality video frame interpolation in pose domain.

AWARDS AND HONORS

Best Paper Award

CVPR AI4CC Workshop

2025

Outstanding Paper Honorable Mention

MLSys

2025

UT Austin Cockrell Engineering Fellowship

UT Austin

2022

Shanghai Outstanding Graduate

SJTU

2022

Academic Excellence Scholarship of Shanghai Jiao Tong University

SJTU

2021

National Scholarship for Graduate Students (Top 0.2% Nationwide)

SJTU

2020

Second Prize in China Post-Graduate Mathematical Contest in Modeling

SJTU

2019

Outstanding Graduate of Shanghai Jiao Tong University

SJTU

2019

PROFESSIONAL SERVICE

- ▷ **Conference Reviewer:** CVPR, ICCV, ECCV, ICML, Neurips, IJCAI, WACV, MM
- ▷ **Journal Reviewer:** Computer Graphics Forum

TEACHING & VOLUNTEER EXPERIENCES

- ▷ TA at ECE 379K - Introduction to Computer Vision (Spring 2024, Fall 2022), UT Austin
- ▷ TA at CS245 - Principles of Data Science (Spring 2021, Spring 2020), SJTU
- ▷ TA at CS7335 - Statistical Learning and Inference (Fall 2020), SJTU
- ▷ Volunteer teacher at Eryuan No.2 high school, Yunnan, China, Aug. 2017- Sept. 2017

PUBLICATIONS

Representative publications that I am a primary author.

- [I15] **Wenyan Cong**, Yiqing Liang, Yancheng Zhang, Ziyi Yang, Yan Wang, Boris Ivanovic, Marco Pavone, Chen Chen, Zhiwen Fan, Zhangyang Wang, "E3D-Bench: A Benchmark for End-to-End 3D Geometric Foundation Models" arXiv preprint arXiv:2506.01933. In submission. [[Paper](#); [Code](#)]
- [I14] Longfei Li, Zhiwen Fan, **Wenyan Cong**, Xinhang Liu, Yuyang Yin, Matt Foutter, Panwang Pan, Chenyu You, Yue Wang, Zhangyang Wang, Yao Zhao, Marco Pavone, Yunchao Wei, "Martian World Models: Controllable Video Synthesis with Physically Accurate 3D Reconstructions" in Conference on Neural Information Processing Systems (NeurIPS), 2025 [[Paper](#); [Code](#)]
- [I13] **Wenyan Cong***, Hanqing Zhu*, Peihao Wang, Bangya Liu, DeJia Xu, Kevin Wang, David Z. Pan, Yan Wang, Zhiwen Fan, Zhangyang Wang, "Can Test-Time Scaling Improve World Foundation Model?" In *Conference on Language Modeling (COLM)*, 2025. [[Paper](#); [Code](#)]
- [I12] **Wenyan Cong**, Hanqing Zhu, Kevin Wang, Jiahui Lei, Colton Stearns, Yuanhao Cai, Dilin Wang, Rakesh Ranjan, Matt Feiszli, Leonidas Guibas, Zhangyang Wang, Zhiwen Fan, Weiyao Wang, "VideoLifter: Lifting Videos to 3D with Fast Hierarchical Stereo Alignment" arXiv preprint arXiv:2501.01949. In *the 13th International Conference on 3D Vision (3DV)*, 2026. Preliminary version accepted by CVPR 2025 Workshop on AI for Content Creation (**Oral Presentation**; **Best Paper Award**). [[Paper](#); [Code](#)]
- [I11] Hanqing Zhu*, Zhenyu Zhang*, **Wenyan Cong**, Xi Liu, Sem Park, Vikas Chandra, Bo Long, David Z. Pan, Zhangyang Wang, Jinwon Lee, "APOLLO: SGD-like Memory, AdamW-level Performance." in *Conference on Machine Learning and Systems (MLSys)*, 2025 [[🏆 Outstanding Paper Honorable Mention](#); [Paper](#); [Code](#), [230+ stars](#); [Hacker News](#); [Huggingface Transformers](#); [LLaMA-Factory](#); [FluxML](#); [axolotl](#); [机器之心](#)]
- [I10] Hanqing Zhu, **Wenyan Cong**, Guojin Chen, Shupeng Ning, Ray Chen, Jiaqi Gu, and David Z. Pan, "PACE: Pacing Operator Learning to Accurate Optical Field Simulation for Complicated Photonic Devices," in *Conference on Neural Information Processing Systems (NeurIPS)*, 2024 [[Paper](#); [Code](#);]
- [I9] Zhiwen Fan*, Jian Zhang*, **Wenyan Cong**, Peihao Wang, Renjie Li, Kairun Wen, Shijie Zhou, Achuta Kadambi, Zhangyang Wang, Danfei Xu, Boris Ivanovic, Marco Pavone, Yue Wang, "Large Spatial Model: Real-time Unposed Images to Semantic 3D", in *Conference on Neural Information Processing Systems (NeurIPS)*, 2024 [[Paper](#); [Code](#);]
- [I8] Zhiwen Fan*, **Wenyan Cong***, Kairun Wen*, Kevin Wang, Jian Zhang, Xinghao Ding, Danfei Xu, Boris Ivanovic, Marco Pavone, Georgios Pavlakos, Zhangyang Wang, Yue Wang, "InstantSplat: Sparse-view SfM-free Gaussian Splatting in Seconds" arXiv preprint arXiv:2403.20309. In submission. [[Paper](#); [Code](#)]
- [I7] **Wenyan Cong***, Hanxue Liang*, Peihao Wang, Zhiwen Fan, Tianlong Chen, Mukund Varma, Yi Wang, and Zhangyang Wang, "Enhancing NeRF akin to Enhancing LLMs: Generalizable NeRF Transformer with Mixture-of-View-Experts" in *IEEE International Conference on Computer Vision (ICCV)*, 2023 [[Paper](#); [Code](#);]
- [I6] Li Niu, Junyan Cao, **Wenyan Cong**, and Liqing Zhang, "Deep Image Harmonization with Learnable Augmentation" in *IEEE International Conference on Computer Vision (ICCV)*, 2023 [[Paper](#); [Code](#);]
- [I5] Junyan Cao*, **Wenyan Cong***, Li Niu, Jianfu Zhang, and Liqing Zhang, "Deep Image Harmonization by Bridging the Reality Gap" in *The 33rd British Machine Vision Conference (BMVC)*, 2022 [[Paper](#);]
- [I4] **Wenyan Cong**, Xinhao Tao, Li Niu, Jing Liang, Xuesong Gao, Qihao Sun, and Liqing Zhang, "High-Resolution Image Harmonization via Collaborative Dual Transformations" in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022 [[Paper](#); [Code](#);]
- [I3] Li Niu, **Wenyan Cong**, Liu Liu, Yan Hong, Bo Zhang, Jing Liang, and Liqing Zhang, "Making Images Real Again: A Comprehensive Survey on Deep Image Composition" arXiv preprint arXiv:2106.14490 [[Paper](#);]
- [I2] **Wenyan Cong**, Li Niu, Jianfu Zhang, Jing Liang, and Liqing Zhang, "BargainNet: Background-Guided Domain Translation for Image Harmonization" in *International Conference on Multimedia and Expo (ICME)*, 2021. Oral [[Paper](#); [Code](#);]
- [I1] **Wenyan Cong**, Jianfu Zhang, Li Niu, Liu Liu, Zhixin Ling, Weiyuan Li, and Liqing Zhang, "DoveNet: Deep Image Harmonization via Domain Verification" in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020. [[Paper](#); [Code](#);]