Wentao Zhou

wentao@gatech.edu | zwt.zip

Education

Georgia Institute of Technology — Aug. 2026–Present
Ph.D. student, Computer Science — Expected May 2029

University of Virginia — Aug. 2024–Apr. 2026
M.S., Computer Science

University of Michigan — Aug. 2022–Apr. 2024
B.S., Computer Science (dual degree)

Shanghai Jiao Tong University — Sep. 2020–Aug. 2024
B.S., Electrical and Computer Engineering (dual degree)

Publications

2026 **Wentao Zhou**, Weijie Gan, Jiayun Wang
Mutually Adversarial Self-Training with Evolving Data for Unified Multimodal Models
Manuscript in submission. [Paper]

Improves generation and understanding in a unified multimodal model through mutually adversarial self-training on evolving model-generated data.

2026 **Wentao Zhou**, Xuweiyi Chen, Charles Saunders, Aruni RoyChowdhury, Zezhou Cheng
Recurrent Point Transformer via Test-Time Training
Manuscript in submission. [Code]

Processes long 3D and 4D point-cloud sequences with compact fast-weight states, enabling online semantic segmentation with nearly constant memory.

IROS '26 **Wentao Zhou**, Xuweiyi Chen, Vignesh Rajagopal, Jeffrey Chen, Rohan Chandra, Zezhou Cheng
Stereo4DWalker: Learning 4D-aware Embodied Urban Navigation from Internet Stereo Videos
IEEE/RSJ International Conference on Intelligent Robots and Systems. [Paper]

Learns 4D-aware urban navigation from internet stereo videos, with diffusion and autoregressive policies evaluated for action generation.

CVPR '26 Xuweiyi Chen, **Wentao Zhou**, Zezhou Cheng
WildRayZer: Self-supervised Large View Synthesis in Dynamic Environments
IEEE/CVF Conference on Computer Vision and Pattern Recognition. ***Highlight.*** [Paper]

Separates static structure from transient motion to learn novel-view synthesis from dynamic scenes without supervision.

ICLR '26 Xuweiyi Chen, **Wentao Zhou**, Aruni RoyChowdhury, Zezhou Cheng
Point-MoE: Large-Scale Multi-Dataset Training with Mixture-of-Experts for 3D Semantic Segmentation
International Conference on Learning Representations. [Paper]

Enables joint training across indoor and outdoor 3D datasets with sparse experts, achieving 70.8 mean mIoU compared with 67.2 for PTv3-L.