

# Hongrui Wu

Homepage ◊ Google Scholar ◊ LinkedIn ◊ Email [wuhongrui2152131@gmail.com](mailto:wuhongrui2152131@gmail.com)

## EDUCATION

### Tongji University

Undergraduate in Computer Science and Technology (**Score:** 91.27/100)  
Undergraduate in Civil Engineering, AI in Construction (Transferred)

*Apr 2024 - Current  
Sep 2021 - Apr 2024*

## RESEARCH INTERESTS

My research focuses on enabling AI systems to build spatial intelligence by learning structured and physically grounded multi-modal representations of the world. I am particularly interested in generative models that can reconstruct and synthesize coherent 3D/4D scenes from visual observations, with the goal of producing representations that are directly usable for simulation, interaction, and downstream embodied tasks.

## PUBLICATIONS

- [1] **Hongrui Wu\***, S. Yoo\*, X. Zhang\*, C. Li, J. Xie, and Z. Tu, “Pixarmesh: Autoregressive mesh-native single-view scene reconstruction,” *In submission CVPR*, 2026.
- [2] **Hongrui Wu\***, Z. Gao\*, J. Cao, K. Yao, W. Shen, and Z. Wei, “Folk: Fast open-vocabulary 3d instance segmentation via label-guided knowledge distillation,” 2025. arXiv: [2510.08849 \[cs.CV\]](https://arxiv.org/abs/2510.08849).
- [3] J. Cao\*, **Hongrui Wu\***, Z. Feng, H. Bao, X. Zhou, and S. Peng, “Universe: Unleashing the scene prior of video diffusion models for robust radiance field reconstruction,” *ICCV*, 2025.

## EXPERIENCE

### MLPC lab in UC San Diego [1] Advisor: Prof. Zhuowen Tu

July 2025 - Present

- We develop the **first** scene-level 3D reconstruction framework that directly predicts **artist-created triangle meshes** from a single image, surpassing prior methods restricted to SDF-based geometry or object-level generation.
- We introduce a **unified autoregressive model** that jointly predicts instance poses, scales, and canonicalized geometry, enabling coherent multi-object reconstruction and spatially aligned scene assembly.
- Experiments on synthetic and real-world datasets show that our method achieves **state-of-the-art** scene reconstruction quality while producing lightweight, high-fidelity meshes suitable for downstream applications (**In Submission to CVPR 2026 [Paper]**).

### Tongji University [2] Advisor: Prof. Zhihua Wei and Prof. Wen Shen

Nov 2024 - Aug 2025

- FOLK is a fast open-vocabulary 3D instance segmentation framework that eliminates expensive 3D-to-2D projection by distilling open-vocabulary knowledge from 2D CLIP into a 3D student model.
- We introduce a novel teacher-student architecture where the teacher leverages multi-view CLIP embeddings with high visibility and viewpoint diversity, and the student learns a direct 3D embedding space via a label-guided distillation mechanism, ensuring semantic consistency and fast inference.
- FOLK achieves **state-of-the-art AP50 of 35.7 on the ScanNet200 dataset**, while being up to **6.0×-152× faster** than prior methods during inference. (**Under review [Paper]**)

### Zhejiang University [3] Advisor: Prof. Sida Peng and Prof. Xiaowei Zhou

Dec 2024 - Mar 2025

- UniVerse is a video generative model for robust 3D reconstruction from inconsistent multi-view images. Given unstructured multi-view inputs, we first estimate a camera trajectory, insert blank frames to form a pseudo video, and then restore the sequence and extract consistent frames for reconstruction.
- This work highlights the potential of decoupling robust reconstruction into **restoration and reconstruction**, instead of directly performing reconstruction on inconsistent inputs. (**Accepted by ICCV 2025. [Website] [Paper] [code]**)

## AWARDS

National Scholarship (Top 0.2% nationwide, highest scholarship in China).

TianXiang Scholarship (2/600)

Interdisciplinary Contest In Modeling hosted by COMAP Finalist Prize (Top 1.8% worldwide)

National Undergraduate Mathematics/Physics Competition: Third Prize/Second Prize

ASCE Concrete Canoe Competition (2nd Place in California Section)