Zhenghong Zhou

Rochester, US 14623 • (+1) 585-509-2735• zhouzhenghong1999@gmail.com • Website • Google Scholar • Github

EDUCATION BACKGROUND

University of Rochester, Computer science — Sep. 2024–Now
Ph.D. in department of Computer science, Advisor: Jiebo Luo
Focus: Computer vision, multi-modal model, 3D/4D generation

Huazhong University of Science and Technology, Information and Communication Engineering — Sep. 2021–Jun. 2024
M.S. in school of EIC, Advisor: Xinggang Wang, Wenyu Liu

Huazhong University of Science and Technology, Biomedical Engineering — Sep. 2017–Jun. 2021
B. E. in school of Engineering Sciences, GPA: **90.4/100,** rank: 2/28

PUBLICATIONS

Arxiv (submitted to top AI conference), 2024, link

- **Zhenghong Zhou***, Jie An*, Jiebo Luo. Latent-Reframe: Enabling Camera Control for Video Diffusion Model without Training.
- This project proposes a method for enabling camera control in video diffusion (like T2V) models without requiring finetuning. It reframes the latent code of frames to match the input camera trajectory using time-aware point clouds. Latent code inpainting and harmonization refine the model's latent space, ensuring high-quality video generation. Experiments show that Latent-Reframe achieves comparable or superior camera control and video quality to training-based methods.

Arxiv (prepare to submit), 2024, link

- **Zhenghong Zhou***, Huangxuan Zhao*, Jiemin Fang, et al. TiAVox: Time-aware Attenuation Voxels for Sparse-view 4D DSA Reconstruction.
- This project proposes a Time-aware Attenuation Voxel (TiAVox) approach for sparse-view 4D DSA reconstruction, which paves the way for high-quality 4D imaging. TiAVox can reconstruct 4D/3D/2D result from the real clinical scene of radiocontrast agent flow. TiAVox does not require 3D supervision, and only needs 30 or less views to reconstruct.

Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), 2023, link

- Kan Wu*, Houwen Peng*, **Zhenghong Zhou***, et al. TinyCLIP: CLIP Distillation via Affinity Mimicking and Weight Inheritance.
- This project proposes a novel cross-modal distillation method for large-scale language-image pre-trained models, introduces two core techniques: affinity mimicking and weight inheritance, reduces the size of the pre-trained CLIP ViT-B/32 by 50% while maintaining comparable zero-shot performance, demonstrates the good transferability of TinyCLIP.

Cell Reports Medicine (IF=11.7), 2022, link

- Huangxuan Zhao*, **Zhenghong Zhou***, Feihong Wu*, et al. Self-supervised learning enables excellent 3D digital subtraction angiography reconstruction from ultra-sparse 2D projection views: A multicenter study.
- This project proposes a self-supervised learning method to realize 3D-DSA reconstruction using ultra-sparse 2D projections. Two radiologists scored the reconstructed images from internal and external datasets using eight projections and identified all 82 lesions with high confidence. The radiation dosages are approximately 1/16.7 compared with the gold standard method.

*: equal contribution.

EXPERIENCE

- **Microsoft Research Asia**, Beijing China. — Dec. 2022 – Jun. 2023
  Research intern, mentored by Dr. Houwen Peng, working on visual-language models

ONGOING PROJECT

Controllable Training-free Scene Generation, 2024, at University of Rochester

- Main contributor, supervised by Prof. Jiebo Luo.
- Recent advancements in dense 3D/4D reconstruction techniques, exemplified by DUSt3R and Gaussian splatting, have been impressive. Meanwhile, video diffusion models can generate realistic videos from text or image prompts. We aim to integrate reconstruction and generation models for controllable, training-free scene generation.
- This project is aimed at ICCV 2025.

AWARDS

- Merit Graduate Student award, HUST — Oct. 2023 and Oct. 2023
- Outstanding Graduate award, HUST — Jun. 2021