Zhanpeng Luo

ZhanpengLuo@pitt.edu | (+1) 412-224-8668
Github | Personal Website

EDUCATION

University of Pittsburgh — Pittsburgh, PA
Bachelor of Computer Sciene — Estimate to graduate at 26 Spring

PUBLICATIONS

Zhanpeng Luo, Haoxi Ran, Li Lu. Instant4D: 4D Gaussian Splatting in Minutes. *Accpted by* ***NeurIPS 25***, [Project Page] [Code] [NeurIPS 25]
Zhanpeng Luo, Ce Zhang … Guanya Shi, Katia Sycara, Yaqi Xie. pySpatial: Generating 3D Visual Programs for Zero-Shot Spatial Reasoning. *Under Review* [Paper]
Linna Wang, Leyi Zhao, **Zhanpeng Luo**, Xuan Wang, Ziliang Feng, and Li Lu "CARE-30: A Causally Driven Multi-Modal Model for Enhanced 30-Day ICU Readmission Predictions", **IEEE, BIBM**
Changjing Song, **Zhanpeng Luo**, Li Lu, and Qian Su " MLNet: Enhancing Joint Predictive Modeling of Chronic DiseasesUsing Deep Learning, **IEEE BIBM 23**
Lixian Zhu, **Zhanpeng Luo**, Xiaodong Yang, Bridgeless Power Factor Correction Circuit and Its Control Method. **China Patent** CN 119696357B

EXPERIENCE

Academic:
Poster Presentation: Robotics Institute Summer Scholar (RISS 2025), CMU
Conference Reviewer: ICLR 26
Dean List: 6 semester; GPA Major: 4.0 / 4
Internship:
25 Summer, RISS, Supervised by Katia Sycara
24 Summer, Research Internship, Supervised by Bernhard Kerbl
23 Summer, Research Internship, West China Hospital

ACTIVE RESEARCH INTEREST

Dynamic Reconstruction Generalized from 3D Pre-Training:
We have limited 4D data and relatively larger 3D datasets. It is important for us to find a way to train/fine-tune so that current LVSM or LRM can generalize to complex, diverse and dynamic scenarios. GEN3C shows it is possible and promising to learn from 3D data and generalize to dynamic scenarios. We can achieve this using visual SLAM to acquire point maps for static and dynamic scenarios and further designing a neural renderer for the point maps. However, this system with several modules has limitations: with fine-tuned data, we can't get better point map / camera model with gradient update; we cannot hallucinate the unseen region. Therefore, can we design a simpler, more scalable, and end-to-end solution? Here are some considerations I am studying:

- We need a simpler formulation for system, GEN3C only supports 14 frames in each chunk. As the sequence length goes longer, the memory overhead will be a burden. With the success of DUSt3R, LVSM and Rayzer, We start to believe that we are able to solve all 3D perception with a single transformer with suitable training strategy.
- We need to go a full latent way, *e.g.* LVSM. 1) We don't have a good enough dynamic modeling method for feedforward prediction. 2) As the sequence goes longer, the redundancy, memory computation overhead limits scalability; and 3) Modeling a scene or a motion sequence into a latent vector is compression, and compression is intelligence.
- We need to 1) utilize (*e.g.*weight, intermediate, supervision, self-supervision) 3D-Pretraining (*e.g.*Croco, DUSt3R, VGGT); 2) unite current dynamic dataset (*e.g.*Vitural KITTI2, DynamicReplica, Stereo4D); and 3) dynamic reconstruction method (*e.g.*Mega-SAM) to fine-tune a vision transformer so that it is able to reconstruct camera and do photo-realistic view synthesis.