Qitao Zhao

Webpage: qitaozhao.github.io
GitHub: github.com/QitaoZhao
Email: qitaoz@outlook.com (Preferred), qitaoz@andrew.cmu.edu
Location: Pittsburgh, PA, USA
Phone: (+1) 412-789-9102
Google Scholar: Profile

EDUCATION

Carnegie Mellon University, Pittsburgh, PA, USA — Aug, 2023 – Dec, 2024
Master of Science, Computer Vision — *GPA:* 4.11/4.3
Advisor: Shubham Tulsiani

Shandong University, Qingdao, Shandong, China — Sept, 2019 – June, 2023
Bachelor of Engineering, Electronic Science and Technology — *GPA:* 94.61/100

PUBLICATION & PREPRINT

[1] **DiffusionSfM: Predicting Structure and Motion via Ray Origin and Endpoint Diffusion**
Qitao Zhao, Amy Lin, Jeff Tan, Jason Y. Zhang, Deva Ramanan, Shubham Tulsiani.
In Submission, 2025. [Paper][Website]

[2] **Sparse-view Pose Estimation and Reconstruction via Analysis by Generative Synthesis**
Qitao Zhao, Shubham Tulsiani.
NeurIPS, 2024. [Paper][Website][Code]

[3] **A Single 2D Pose with Context is Worth Hundreds for 3D Human Pose Estimation**
Qitao Zhao, Ce Zheng, Mengyuan Liu, Chen Chen.
NeurIPS, 2023. [Paper][Website][Code]

[4] **PoseFormerV2: Exploring Frequency Domain for Efficient and Robust 3D Human Pose Estimation**
Qitao Zhao, Ce Zheng, Mengyuan Liu, Pichao Wang, Chen Chen.
CVPR, 2023. [Paper][Website][Code]

RESEARCH EXPERIENCE

My research focuses on inferring 3D structures from *sparse* 2D signals, such as predicting a 3D human pose from a few 2D poses or estimating 3D shapes from sparse-view images. I am passionate about recovering visually and physically plausible 3D structures from everyday observations (*e.g.,* photographs or video clips) or through our interactions with the world around us.

Carnegie Mellon University, Physical Perception Lab — Sept, 2023 – Present
Graduate Student Researcher — *Advisor:* Shubham Tulsiani
Topic: Sparse-view Camera Pose Estimation and Reconstruction

[1] **DiffusionSfM** (June, 2024 – Present)

- Proposed a diffusion-based framework to infer pixel-aligned camera ray origins and endpoints (*i.e.,* point clouds) from multiple input images, serving as a unified representation for scene geometry and camera pose
- Designed a *depth-mask conditioning* approach to deal with missing depth in ground truth data from real-world datasets when training diffusion-based models
- Introduced a *diffusion guidance* mechanism to guide the x_0-prediction from our model using mono-depth estimates from off-the-shelf models, which significantly improves the predicted geometry

[2] **SparseAGS** (Sept, 2023 – May, 2024)

- Introduced an analysis-by-generative-synthesis framework that jointly estimates 3D and camera viewpoints given sparse-view images, by integrating a 6-DoF novel-view

generative prior in an analysis-by-synthesis approach
- Proposed a *pose outlier identification and correction* approach by iteratively reconstructing 3D and investigating re-projection errors, which handles potential outlier(s) in the initial pose estimates from off-the-shelf methods
- Delivered an invited talk at the CMU VASC Seminar and presented a poster at the XRTC Symposium

University of Central Florida, CRCV — Apr, 2022 – Aug, 2023
Undergraduate Student Researcher — *Advisor:* Chen Chen
Topic: 3D Human Pose Estimation

[3] **Context-Aware PoseFormer** (Dec, 2022 – May, 2023)
- Proposed to leverage readily available visual representations (*i.e.,* multi-resolution feature maps) from 2D pose detectors to alleviate inherent ambiguities in single-view 2D-to-3D human pose estimation
- Developed a *Deformable Context Extraction* module to adaptively extract joint-centric contextual features from feature maps, accounting for the uncertainty in detected 2D joints within an end-to-end learnable framework

[4] **PoseFormerV2** (Apr, 2022 – Nov, 2022)
- Proposed to leverage a low-frequency representation for 2D joint sequences to enhance the efficiency on long sequences and robustness to keypoint noise of transformer-based methods
- Integrated the proposed frequency-domain representation into PoseFormer's temporal encoder, introducing novel designs such as *FreqMLP* to enhance time-frequency feature fusion
- Applied the proposed approach to other transformer-based SoTA methods

AWARDS

NeurIPS Travel Grant 2023
Outstanding Undergraduate Thesis Award 2023
Ranked 1st out of 7 finalists among 350+ students in the department
National Scholarship 2020
Awarded to the top 0.2% of students nationwide for academic excellence

SERVICE

Reviewer: NeurIPS 2024, ICLR 2024, ECCV 2024 Workshop

SKILLS

Programming: C, MATLAB, Python (NumPy, PyTorch, PyTorch3D)
Languages: English (fluent), Mandarin (native)
Test Scores: TOEFL (109; Speaking: 24), GRE (326+4.0)