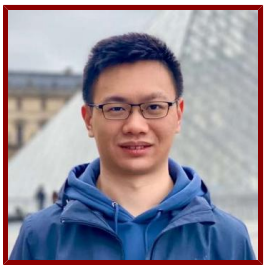
Stanford



Tao Sun

Ph.D. Student in Civil and Environmental Engineering, admitted Autumn 2024

Bio

BIO

Tao Sun is a PhD candidate in the Gradient Spaces Lab at Stanford University, working with Iro Armeni and Shuran Song. His research connects 3D vision, generative models, and robotics, with a focus on geometric representations and generative methods for scene understanding, pose estimation, shape assembly, and robot planning. He has been a research intern at NVIDIA Cosmos Lab, working on World-Action Models (WAMs) for robot planning and control. He is supported by the Stanford Graduate Fellowship and the Stanford Robotics Center Robotics Scholars Fellowship. His coauthored work includes Nothing Stands Still, recipient of the ISPRS Journal Best Paper Award for 2025, and Register Any Point, an ECCV 2026 Best Paper Award Candidate.

Previously, he worked on large language model post-training and reasoning at ByteDance Seed Lab and on scene understanding and multitask learning at ETH Zurich's Computer Vision Lab. He holds an M.Sc. in Computer Science from ETH Zurich and a B.Sc. from Tongji University.

HONORS AND AWARDS

- Stanford Robotics Center Robotics Scholars Fellowship, Stanford Robotics Center (May 2026)
- ISPRS Journal Best Paper Award for 2025, International Society for Photogrammetry and Remote Sensing (April 2026)
- ECCV 2026 Best Paper Award Candidate, European Computer Vision Association (ECVA) (Sept 2026)

LINKS

- Homepage: https://taosun.io
- Google Scholar: https://scholar.google.com/citations?user=x1O-cHYAAAAJ&hl=en

Research & Scholarship

CURRENT RESEARCH AND SCHOLARLY INTERESTS

My research asks how machines can understand the structure of the physical world and use that understanding to act. I work at the intersection of 3D computer vision, generative modeling, and robotics, developing geometric representations and learning methods that connect perception with spatial reasoning, planning, and control. My long-term goal is to enable robots to reason about unfamiliar scenes and objects, anticipate the consequences of their actions, and perform tasks that require geometric precision and adaptation.

One focus is 3D world understanding across changes in viewpoint, sensing modality, scene scale, and time. I study point cloud registration and pose estimation as foundations for relating observations and building consistent spatial representations. My work on Nothing Stands Still examines registration

under large geometric and temporal changes, including construction and renovation environments. Through Rectified Point Flow and Register Any Point, I investigate how flow-based generative formulations can support point cloud alignment and shape assembly, including settings with ambiguity and symmetry.

A second focus is generative geometry and spatial reasoning. I am interested in models that capture the structure and relationships of objects and scenes, rather than relying on superficial correlations. This includes representing multiple plausible geometric solutions and evaluating whether multimodal models genuinely use 3D information. My work on Real-3DQA examines shortcuts in spatial reasoning benchmarks and how to assess visually grounded understanding more reliably.

A third focus is connecting generative models to robot action. I study how short-horizon behaviors can be composed into coherent long-horizon plans, including energy-based methods that resolve inconsistencies between trajectory fragments. At NVIDIA Cosmos Lab, I work on World-Action Models (WAMs) for robot planning and control. My broader interests include how world models and vision-language-action models can incorporate geometric structure, predict action outcomes, and support decisions in changing physical environments.

My earlier work on large language model post-training and reasoning, including contributions to Seed1.5-Thinking, informs my interest in reinforcement learning and the relationship between training objectives, reasoning behavior, and generalization. Across these directions, I value rigorous evaluation, reusable benchmarks, and methods that connect advances in representation learning to measurable improvements in robot capabilities.