

Pouyan Navard

"Turning pixels into 3D intelligence."

📍 Columbus, OH 📩 bnd.pouyan@gmail.com 🔍 bnavard.github.io

Skills

Languages: Python, C++

Deep Learning: Hydra, Pytorch Lightning, Diffusers, HuggingFace, Wan, CogVideoX, MMSeg

Architecture: CogVideoX, Wan, Vision Transformer, Wav2Lip, Mixture of Experts (MOE), Diffusion Models

Software Engineering: MLOps, Object Oriented Programming, Linux, Bash, Git, Docker

Professional Experience

Computer Vision Engineer

Path Robotics Inc.

Columbus, Ohio

Nov 2024 – Present

- **Weld Anything Model:** Architecting an end-to-end deep learning system for generalizable robotic welding.
 - Defined the problem formulation and developed proof-of-concepts to validate feasibility.
 - Built a self-supervised multimodal backbone enabling cross-geometry generalization in welding scenarios.
 - Fine-tuned and deployed the model in production via a real-time ROS-based inference service, supporting continuous robotic adaptation.
- **Machine Learning Infrastructure:** Built a flexible MLOps framework to streamline deep learning project development and deployment.
 - Reduced model setup time by **60–80%** through configuration-driven initialization templates.
 - Improved experiment **reproducibility**, version control, and standardization across research teams.
 - Enabled **push-button scaling**, automated training tracking, and continuous integration (CI/CD) for production workflows.
- **3D Neural Asset GenAI:** Led the development of a diffusion-based 3D asset generation pipeline with controllable lighting for high-fidelity robotic simulation data.
 - Designed a lighting-aware generation strategy to enhance visual realism and geometric consistency.
 - Reduced real-world data collection time by **86.7%**, enabling large-scale pretraining.
 - Advanced research toward scalable, simulation-driven learning for robotic perception and control.

Research Assistant

Photogrammetric Computer Vision Lab

Columbus, Ohio

Feb 2021 – Aug 2025

- **Visual Instruction Tuning:** Developed a multimodal instruction-tuning framework aligning large vision-language models for complex visual reasoning tasks.
 - Curated and generated high-quality image–text instruction datasets via self-instruct pipelines and domain-specific caption mining.
 - Designed a fine-tuning strategy to progressively ground linguistic concepts in visual context.
 - Achieved **81.3%** recall overlap with ground-truth tokens, improving multimodal reasoning and response coherence on visual QA benchmarks.
- **Self-Supervised Multimodal Learning:** Developed a video–text fusion framework for learning joint representations in self-supervised settings.
 - Implemented dynamic token selection to identify redundant video tokens and enrich them with semantically aligned text features.
 - Improved cross-modal reasoning and representation alignment through fine-grained video–text fusion.
 - Enhanced retrieval and captioning performance via more coherent video–language representations.

Selected Publications

Large Language-and-Vision Assistant for Lunar Exploration (LLaVA-LE) June 2025
Pouyan Navard, Gokce Inal, Alper Yilmaz
Arxiv (under review)

Developed a multimodal conversational AI that interprets lunar imagery and answers open-ended questions to support lunar mapping and geo related exploration.

KnobGen: Controlling the Sophistication of Artwork in Sketch-based Diffusion Models June 2025
Pouyan Navard, Amin Monsefi, Harry Chao, Alper Yilmaz, Rajiv Ramanath
CVPR Workshops, 2025

Proposed a dual-path diffusion framework with coarse- and fine-grained controllers unified via a “**knob**” **inference mechanism** for adaptive sketch-to-image synthesis. Achieved robust controllability and fidelity on MultiGen-20M, advancing human-controllable generative modeling.

SegFormer3D: An Efficient Transformer for 3D Medical Image Segmentation June 2024
Shehan Perera, **Pouyan Navard**, Alper Yilmaz
CVPR Workshops, 2024

Co-developed a memory-efficient Vision Transformer achieving **33× fewer parameters** and **13× lower compute cost** while maintaining SOTA accuracy on Synapse, BRaTs, and ACDC—making high-quality 3D segmentation feasible on standard GPUs.

A Probabilistic Drift Correction Module for Visual-Inertial SLAMs June 2024
Pouyan Navard, Alper Yilmaz
ISPRS Proceedings, 2024

Designed a plug-and-play probabilistic drift correction module reducing motion drift by up to **10×** in long-term SLAM trajectories, improving robustness in autonomous navigation and AR/VR tracking.

Distribution-Aware Learning for Sparsely Labeled and Imbalanced Spatiotemporal Data 2025
Pouyan Navard
Doctoral Dissertation, The Ohio State University

Introduced a probabilistic video classification framework modeling labels as Gaussian distributions, enabling uncertainty-aware prediction under sparse supervision using a self-supervised video transformer.

Awards and Honors

Robert E. Altenhofen Memorial Scholarship Award  ISPRS, 2022
◦ National-level scholarship recognizing excellence in photogrammetry and remote sensing research.

Education

The Ohio State University Feb 2021 – Aug 2025
PhD in Electrical and Computer Engineering
◦ GPA: 3.9/4.0
◦ Dissertation: *Distribution-Aware Learning for Sparse and Imbalanced Visual Data*
◦ Research focus on 3D perception, diffusion-based generative modeling, and multimodal learning under supervision of Prof. Alper Yilmaz.

Isfahan University Sep 2015 – Sep 2019
BSc in Electrical and Computer Engineering
◦ GPA: 3.6/4.0
◦ Concentration in signal processing and computer vision.