Aarya Sumuk

asumuk@stanford.edu | +1-408-426-6961
aaryasum.github.io | Scholar

EDUCATION

Stanford University *June 2026*
Master of Science - *Robotics Track* GPA: 4.15/4.00
Selected Coursework: Robot Perception, Reinforcement Learning, Experimental Robotics

Indian Institute of Technology Madras *July 2024*
Bachelor of Technology in Mechanical Engineering GPA: 9.4/10.0
Selected Coursework: Mechatronics, Pattern Recognition and Machine Learning, Foundations of Data Science

RESEARCH INTERESTS

I am interested in grounded robot intelligence: how robotic policies learn spatial, semantic, and physical structure from perception, language, simulation, and real-world interaction. I am especially drawn to contact-rich manipulation as a testbed for generalization, failure recovery, and skill learning.

SELECTED RESEARCH EXPERIENCE

Bimanual Robotic Assembly with Contact-Rich Manipulation *Stanford, USA*
Interactive Perception and Robot Learning Lab (IPRL), *Prof. Jeannette Bohg* *May 2025 – Jan 2026*
Exploratory project investigating how hardware design, control strategies, and sensing modalities affect policy learning in real-world contact-rich manipulation tasks.

- Created the experimental substrate using dual Franka FR3 arms, Robosuite (MujoCo) environments, and Oculus teleoperation for real-robot data collection and insertion-style task evaluation.
- Studied bimanual workspace design and safety through task-coverage simulations, inclined mounting, potential-field collision avoidance, and exploratory OSCBF-style constraints for safe interaction.
- Evaluated BC, diffusion policy, and SERL-style online RL, finding that tight-tolerance insertion success relies heavily on force-aware interaction, exploration, and autonomous recovery strategies.

Visuo-Tactile World Models for Robot Policy Learning *Stanford, USA*
Assistive Robotics and Manipulation Lab (ARMLab), *Prof. Monroe Kennedy III* *Apr 2026 – Present*
Research project extending Toyota Research Institute's Interactive World Models (IWM) to study how tactile-conditioned latent dynamics affect policy performance in contact-rich tabletop manipulation tasks.

- Integrating high-resolution DenseTact sensor feedback into the IWM transformer-based architecture to investigate the role of tactile grounding in state-space representation and future-state prediction.
- Designing a multimodal data pipeline using a UMI-style gripper and iPhone RGB-D sensing to collect synchronized demonstrations that align tactile signals with visual and proprioceptive streams.
- Evaluating whether tactile rollouts enhance long-horizon prediction stability and offline policy evaluation for manipulation tasks involving complex object interaction, slip, and non-prehensile contact.

Tactile Failure Diagnosis and Recovery for Robotic Manipulation *Stanford, USA*
Independent Research Project, Manuscript in Preparation *Mar 2026 – Present*
Developing a tactile-aware framework to investigate how high-frequency physical signals can be mapped to semantic anomalies to condition adaptive recovery strategies in complex manipulation.

- Formulated a formal tactile failure taxonomy that maps high-dimensional physical signatures to latent anomalies, specifically characterizing micro-slip, misalignment, and stochastic surface irregularities.
- Developed robosuite simulation pipeline in Robosuite featuring automated failure injection and multimodal logging (RGB, tactile, proprio) to establish robust diagnostic and recovery benchmarks.
- Designing a diagnosis-conditioned recovery framework where latent failure representations guide residual corrective policies, enabling generalization across unmodeled contact dynamics and geometries.

Sample-Efficient RL via World Foundation Models for Synthetic Demonstrations *Stanford, USA*
CS234 Reinforcement Learning, *Prof. Emma Brunskill* *Feb 2026 – Mar 2026*
Investigated the utility of world foundation models as synthetic data generators to improve the sample efficiency of robotic policies in sparse-demonstration regimes.

- Fine-tuned a Cosmos Predict 2 world model on a limited set of ten expert demonstrations to generate high-fidelity synthetic trajectories and expand the effective training distribution.

- Developed a multimodal labeling pipeline using Pose Estimation Models (PEM) and Inverse Dynamics Models (IDM) to recover latent states and actions from generated video sequences.
- Evaluated Diffusion and residual RL policies trained on augmented datasets, finding that expert-only data outperformed expert+synthetic training under current video-to-action labeling quality.

Learning Actionable Affordances from Pairwise Human Preferences *Stanford, USA*
CS329H Machine Learning from Human Preferences, *Prof. Sanmi Koyejo* *Oct 2025 – Dec 2025*

Investigated human-aligned visual representations by learning robotic affordances from structured preference feedback to bridge the gap between semantic intent and physical grounding.

- Curated a specialized dataset of pairwise human graspability preferences over local image patches to provide a dense supervision signal for aligning visual features with functional utility.
- Optimized a Bradley–Terry preference model utilizing frozen self-supervised DINO embeddings to re-rank visual features based on human-perceived actionable regions for high-level manipulation.
- Generated dense affordance heatmaps that effectively suppress task-irrelevant visual saliency, highlighting graspable regions to improve the success rates of downstream motion planning policies.

Real-Time 6D Pose Estimation for Robotic Assembly *Odense, Denmark*
Elite Robotics Summer School, University of Southern Denmark *Aug 2025*

Developed high-speed spatial reasoning pipelines for complex industrial assembly tasks, focusing on the hybrid integration of classical geometric constraints and deep regression models.

- Designed robust 6D pose estimation pipelines integrating YOLO and Segment Anything Model (SAM) with classical PnP and ICP refinement for precise object localization in high-clutter scenes.
- Trained a ResNet-based regression architecture on a hybrid dataset of synthetic and real-world samples, utilizing composite translation and rotation losses to achieve sub-centimeter pose precision.
- Evaluated system performance using ADD/ADD-S metrics, achieving real-time estimation of 30+ brick poses with a 1.2 cm mean error, validated through closed-loop robotic grasping experiments.

Lightweight 3D Inpainting for Cultural Heritage Restoration *Stanford, USA*
CS231N Deep Learning for Computer Vision, *Prof. Fei-Fei Li* *Apr 2025 – Jun 2025*

Explored generative geometry completion using 3D diffusion models to restore the structural and semantic integrity of fragmented cultural heritage artifacts in volumetric space.

- Engineered a two-stage generative vision pipeline consisting of 2D U-Net mask prediction and 3D latent diffusion for high-fidelity geometric inpainting and reconstruction of voxelized data.
- Implemented a 3D diffusion model trained with composite BCE, L1, and perceptual losses to optimize the geometric consistency and structural detail of the recovered object surfaces.
- Quantified performance via Chamfer distance and F-score improvements, demonstrating superior reconstruction quality and semantic coherence compared to deterministic geometric baselines.

PUBLICATIONS

1. **Sumuk A.**, Khan Q. "A Sensing Device For Real-Time Road Condition Monitoring". In 2023 IEEE Asia Pacific Conference On Postgraduate Research In Microelectronics And Electronics, India, Nov 2023.
2. **Sumuk A.**, Martinez K. B., Rouhani H. "3D Modelling of Human Hand Using Instrumented Gloves". In 2023 Annual Alberta Biomedical Engineering (BME) Conference, Banff, Canada, Oct 2023.

TEACHING & MENTORSHIP

- **Principles of Robotic Autonomy I (Stanford):** Managed ROS2-based labs and provided technical mentorship for 200+ students on autonomous navigation, state estimation, and path planning.
- **Robot Dexterity (Stanford):** Supported technical modules on bimanual manipulation, impedance control, and tactile-driven feedback for students working with Franka FR3 hardware and MuJoCo simulators.

TECHNICAL SKILLS

- **Robotics & Control:** ROS2, MuJoCo, Robosuite, Franka FR3, Ufactory Xarm; experience training and deploying Imitation Learning (IL) and Reinforcement Learning (RL) policies on real hardware.
- **ML & Computer Vision:** PyTorch, OpenCV, YOLO, SAM, Dino, Diffusion Models, 3D Geometry Reconstruction, and Multimodal Representation Learning.
- **Programming & Tools:** Python (NumPy, PyTorch), C/C++, Bash; Git Version Control, Linux/Ubuntu environments, and high-density PCB design for custom tactile/force-torque sensor integration.