

# Damiano Marsili

PhD student in Computing and Mathematical Sciences, Caltech, Pasadena, USA  
Email, Personal Website

## EDUCATION

---

**California Institute of Technology**, Pasadena, USA  
PhD Student, Computing and Mathematical Sciences

Sep 2023 — May 2028

**Johns Hopkins University**, Baltimore, USA  
BS Computer Science  
BA Mathematics

Aug 2020 — May 2023

## PUBLICATIONS

---

**Same or Not? Enhancing Visual Perception in Vision-Language Models** (PDF) CVPR 2026  
*Damiano Marsili, Aditya Mehta, Ryan Lin, Georgia Gkioxari*

- Created TWIN, a large-scale dataset of 561K VQA queries designed to improve fine-grained understanding in VLMs.
- Introduced FGVQA, a benchmark suite of 12,000 queries that repurposes retrieval datasets for fine-grained VQA.
- Demonstrated that post-training on TWIN improves fine-grained understanding in VLMs, measured by an improvement of up to 19.3% on FGVQA, without compromising performance on general VQA.

**No Labels, No Problem: Training Visual Reasoners with Multimodal Verifiers** (PDF) ICLR 2026  
*Damiano Marsili, Georgia Gkioxari*

- Developed VALOR, an annotation-free post-training framework that uses multimodal verifiers to jointly improve visual reasoning via reinforcement learning and visual grounding through automated hard-negative mining.

**Visual Agentic AI for Spatial Reasoning with a Dynamic API** (PDF) CVPR 2025  
*Damiano Marsili\*, Rohun Agrawal\*, Yisong Yue, Georgia Gkioxari*

- Designed a training-free agentic visual programming approach, VADAR, that dynamically generates new skills in Python and significantly outperforms previous visual programming methods on spatial reasoning in 3D.
- Introduced Omni3D-Bench, a benchmark for 3D understanding with complex queries involving multiple reasoning steps.

## WORK EXPERIENCE

---

**Amazon Robotics** Arlington, USA  
*Applied Science Intern* May 2023 — Sep 2023

- Engineered a large multimodal spatial reasoning dataset composed of over 300,000 grasp samples.
- Trained a Vision Language Model (VLM) to resolve spatial relationships for the task of targeted grasping.

**Applied Physics Laboratory (JHU APL)** Baltimore, USA  
*Research Assistant* Aug 2022 — May 2023

- Worked on self-supervised training methods to train robots for gesture recognition using both synthetic and real data. Project funded by Army Research Labs (ARL).
- Leveraged novel techniques in transfer learning to mitigate the synthetic-to-real gap for gesture recognition.

**Malone Center for Engineering in Healthcare, Johns Hopkins University** Baltimore, USA  
*Research Intern* Feb 2022 — May 2022

- Developed simulation environments used to train reinforcement learning agents for autonomous ventilators.
- Explored the impact of various medical insults on pulmonary compliance in the simulation environment.

## TEACHING EXPERIENCE

---

**Object Oriented Software Engineering** Johns Hopkins University  
*Teaching Assistant* Jan 2022 — May 2022

- Mentored a group of 7 students partaking in a semester-long software project.
- Arranged mock presentation sessions to provide feedback ahead of their final.

**Learning Den** Johns Hopkins University  
*Mathematics & Computer Science Tutor* Sep 2021 — Jan 2022

- Helped two students improve from a B- to an A-/A in Calculus II and Calculus III respectively.
- Constructed a tailored curriculum of practice sets to reinforce concepts the students found challenging.