

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) <i>Damiano Marsili, Aditya Mehta, Ryan Lin, Georgia Gkioxari</i>	CVPR 2026
No Labels, No Problem: Training Visual Reasoners with Multimodal Verifiers (PDF) <i>Damiano Marsili, Georgia Gkioxari</i>	ICLR 2026
Visual Agentic AI for Spatial Reasoning with a Dynamic API (PDF) <i>Damiano Marsili*, Rohun Agrawal*, Yisong Yue, Georgia Gkioxari</i>	CVPR 2025
<ul style="list-style-type: none">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.	
<ul style="list-style-type: none">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.	
<ul style="list-style-type: none">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 <i>Applied Science Intern</i>	Arlington, USA May 2023 — Sep 2023
<ul style="list-style-type: none">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) <i>Research Assistant</i>	Baltimore, USA Aug 2022 — May 2023
<ul style="list-style-type: none">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 <i>Research Intern</i>	Baltimore, USA Feb 2022 — May 2022
<ul style="list-style-type: none">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 <i>Teaching Assistant</i>	Johns Hopkins University Jan 2022 — May 2022
<ul style="list-style-type: none">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 <i>Mathematics & Computer Science Tutor</i>	Johns Hopkins University Sep 2021 — Jan 2022
<ul style="list-style-type: none">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.	