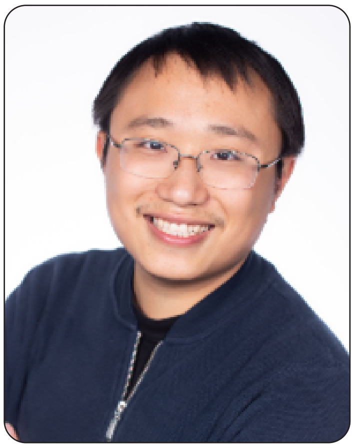
27th International Conference on
Pattern Recognition
December 01-05, 2024, Kolkata, India



Xiaolong Wang

Affiliation: Assistant Professor, Electrical and Computer Engineering,

University of California San Diego

Google Scholar: https://scholar.google.com/citations?user=Y8O9N_0AAAAJ&hl=en

DBLP: https://dblp.org/pid/91/952-4.html

Biography:

Xiaolong Wang is an Assistant Professor in the ECE department at the University of California, San Diego, and a Visiting Professor at NVIDIA Research. He received his Ph.D. in Robotics at Carnegie Mellon University. His postdoctoral training was at the University of California, Berkeley. His research focuses on the intersection between computer vision and robotics. His specific interest lies in learning visual representations from videos and physical robotic interaction data. These comprehensive representations are utilized to facilitate the learning of human-like robot skills, with the goal of generalizing the robot to interact effectively with a wide range of objects and environments in the real physical world. He is the recipient of the J. K. Aggarwal Prize, NSF CAREER Award, Intel Rising Star Faculty Award, and Research Awards from Sony, Amazon, Adobe, and CISCO.

Title: From Perception to Embodied AI: Modeling Humans for Humanoid Robots

Abstract:

The vast development in visual perception has enabled significant advancement and countless applications in robotic systems. Entering the LLM era, the connection between language and vision has enabled the robot to not only perceive but also reason and plan its interaction with the physical world. Among all the robot platforms, the humanoid robot provides a general-purpose platform to conduct diverse tasks we do in our daily lives. In this talk, I will present a 2-level learning framework designed to equip humanoid robots with robust mobility and manipulation skills, enabling them to generalize across diverse tasks, objects, and environments. The first level focuses on training Vision-Language-Action (VLA) models with human video data for both navigation and manipulation. These models can predict “mid-level” actions which predict precise movements or trajectories for the human body and hands, conditioned on language instructions. The second level involves developing low-level robot manipulation skills through teleoperation, and low-level humanoid whole-body control skills via motion imitation and Sim2Real. By combining human VLA with low-level robot skills, this framework offers a scalable pathway toward realizing general-purpose humanoid robots.

Organizing / Technical Partners