Artificial Intelligence Initiative (Aii)

https://ai.ucf.edu/

Invited Speaker

Dr. Wanhua Li, Harvard University



"Beyond 2D Pixels: Towards General -Purpose Intelligent Agents with Vision Foundation Models"

Thursday, June 26, 2025 • 11:30AM • R1 101 • Zoom

Abstract:
Vision Foundation Models (VFMs) have revolutionized computer vision, achieving remarkable generalization across diverse 2D image tasks. However, building general-purpose intelligent agents requires perception that goes beyond static 2D pixels—integrating language, 3D spatial reasoning, and temporal dynamics. In this talk, I will discuss how we can extend VFMs along these three critical dimensions. First, I will introduce Language-Guided VFMs, which leverage natural language as an interface to enhance visual reasoning. Next, I will present 3D-enabled VFMs, which bridge the gap between 2D vision models and the real-world 3D environments where intelligent agents operate. Finally, I will explore Dynamic-Aware VFMs, which incorporate temporal understanding for video and 4D scene reasoning. Throughout the talk, I will highlight key challenges, present novel approaches, and discuss the impact of these advancements on embodied AI, robotics, and multimodal intelligence. I will conclude by outlining future directions for enabling general-purpose vision agents that see, understand, and interact with the world in a human-like manner.

Bio:
Wanhua Li is a postdoc at Harvard University working with Prof. Hanspeter Pfister. He received his Ph.D. degree from Tsinghua University in 2022. His research focuses on computer vision, multimodal AI, and embodied AI. He has authored over 20 papers in top-tier conferences and journals in computer vision, machine learning, and artificial intelligence, such as NeurIPS, CVPR, ICCV, ECCV, ICLR, and TPAMI, with some of his work selected as CVPR Highlight and ICLR Oral presentations.