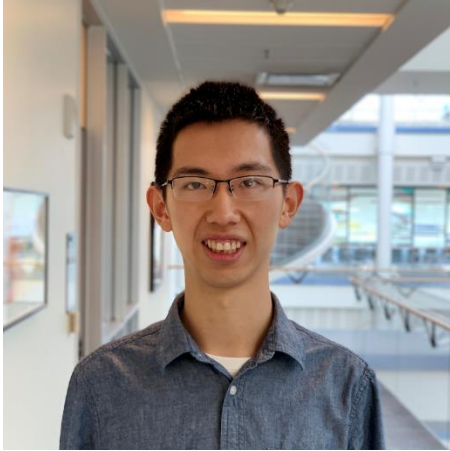




DISSERTATION DEFENSE



Shengyi Qian

Learning to Interact with the
3D World

Friday, July 19, 2024

1:00pm – 3:00pm

3725 Beyster

Virtual – [Zoom](#)

ABSTRACT: Enabling machines to perceive, understand, and interact with the 3D world remains a fundamental challenge in Computer Vision and Robotics. This dissertation explores novel approaches to teach machines to understand and interact with arbitrary objects in diverse scenes. The development of such a system has the potential to greatly enhance the ability of AI agents to navigate and operate in both the physical world and its digital twin.

The dissertation is structured into two primary sections. The first part focuses on the methodology of training deep networks using a vast array of unstructured Internet videos. To construct this methodology from the ground up, the research delves into the curation of diverse video data and introduces a novel approach for predicting 3D object interactions from a single image.

The second section extends the approach in two key directions. Firstly, it investigates the distillation of comprehensive world knowledge from Vision Language Models to further enhance the system's understanding of object affordances and interactions. Secondly, it employs the developed system as a visual pretraining mechanism for robotics, aiming to improve the performance and generalization of manipulation policies.

The key contributions of this dissertation lie in the development of techniques spanning from passive 3D perception to active object manipulation. By leveraging the vast knowledge embedded in Vision Language Models and applying the system to robotic pretraining, this research takes significant steps towards endowing machines with the ability to intelligently interact with the 3D world. The proposed methodologies, spanning data curation, interaction prediction, knowledge distillation, and policy learning, collectively advance the state-of-the-art in machine perception and interaction, paving the way for more capable and adaptable AI agents.

CHAIR: Prof. David Fouhey and Prof. Joyce Chai