



## DISSERTATION DEFENSE



### Yunseok Jang

#### Enhancing Video Understanding Through Deep Generative Models and Task Comprehension

Tuesday, December 17, 2024

4:00pm – 6:00pm

Virtual – [Zoom](#) Passcode: defense

**ABSTRACT:** This thesis explores video understanding through deep learning, focusing on realistic video generation and semantic interpretation of real-world videos. The research addresses challenges in conditional Generative Adversarial Networks (cGANs) and enhances video-based task understanding.

The work begins by tackling the mode-collapse issue in cGANs through a diversity-sensitive term that ensures varied outputs from inputs. Building on this foundation, the research introduces the RiCS 2D representation to better encode volumetric information, effectively bridging the gap between 3D volume comprehension and 2D image generation. This method improves the realism of synthetic videos by precisely representing self-occlusion in 2D camera space. The diversity-sensitive approach is further extended to develop the Learning to Learn from Diverse Attacks (L2L-DA) project, advancing adversarial attack and defense mechanisms.

In the realm of video understanding, the research focuses on semantic interpretation of instructional videos, particularly in delineating task sequences. The MSG<sup>2</sup> work develops methods for generating subtask graphs from these videos, leveraging techniques from Reinforcement Learning. Inspired by these results but seeking to eliminate manual annotation requirements, we introduce MONDAY, a framework that automatically extracts and comprehends mobile operating system navigation procedures from online videos, demonstrating how deep learning approaches can be made more scalable and practical.

This thesis contributes to video understanding by developing integrated deep learning solutions that enhance both video generation and understanding capabilities.

**CHAIR:** Prof. Honglak Lee