UNIVERSITY OF GUELPH

College of Computational, Mathematical, and Physical Sciences

School of Computer Science

Computationally Speaking

Seminar Series

Title: **Understanding Human Activities from Video: Multimodal Learning with Limited Labels**

Speaker: **Dr. Yan Song**

Hosted by: Dr. Yan Yan

Where: In-person (REYN 1101) and Online (**Teams link**)

When: **Friday November 20, 2026, from 9:30am-10:30am**

Abstract: Understanding human activities from video has important applications in healthcare, assistive technologies, and smart environments. Despite significant advances in deep learning, challenges remain in integrating multiple data modalities, learning from limited labeled data, and detecting actions in continuous videos.

In this talk, I will present my research on human action recognition and temporal detection, highlighting my previous work on multimodal fusion, representation learning, and weakly and semi-supervised learning. I will also introduce my current ongoing research on generalized few-shot action recognition, particularly the challenge of recognizing rare but critical events while maintaining accurate recognition of common activities. Finally, I will discuss my long-term research vision for developing robust, annotation-efficient methods for human activity understanding, along with open research questions and opportunities for collaboration.

About the Speaker: **Dr. Yan Song** is an Assistant Professor in the Department of Computing Science at Thompson Rivers University (TRU), Canada, where she joined the faculty in 2024. She received her Ph.D. in Computer Science from the Chinese Academy of Sciences and her M.Sc. in Data Science from TRU. Before joining TRU, she was an Associate Professor at Nanjing University of Science and Technology, China. Her research focuses on multimedia analysis and machine learning, particularly human action recognition and temporal detection, multimodal learning, and learning with limited supervision. Her current research explores methods for understanding human activities from video with limited labeled data, with potential applications in intelligent environments.

