Department of Electrical Engineering
香港城市大學
City University of Hong Kong

IEEE Hong Kong Section 香港分會
CAS/COM Chapter

CityU Architecture Lab for Arithmetic and Security (CALAS) Seminar Series

Edge Intelligence : Hardware Challenges and Opportunities

Prof. Jose Nunez-Yanez
Linköping University, Sweden

Abstract:

In this talk we will initially discuss some basic edge computing concepts followed by the hardware challenges and opportunities of performing deep learning at the edge. We will then present the FADES (Fused Architecture for DEnse and Sparse tensor processing) heterogeneous architecture focusing on its application to Graph neural networks (GNN) acceleration. GNNs can deliver high accuracy when applied to non-Euclidean data in which data elements do not fit into a regular structure. They combine sparse and dense data characteristics and this, in turn, results in a combination of compute and bandwidth intensive requirements challenging to meet with general purpose hardware. FADES is a highly configurable architecture fully described with high-level synthesis integrated in TensorFlow Lite and Pytorch. It creates a dataflow of dataflows with multiple hardware treads and compute units that optimize data access and processing element utilization. This enables fine-grained stream hybrid processing of sparse and dense tensors suitable for multi-layer graph neural networks.

Biography:

Prof. Nunez-Yanez is a professor in hardware architectures for Machine Learning at Linköping University with over 20 years of experience in the design of high-performance embedded hardware. He holds a PhD in hardware-based parallel data compression from the University of Loughborough, UK, with three patents awarded on the topic of high-speed parallel data compression. Previously to joining Linköping University he was a reader (associate professor) at Bristol University, UK. He spent a few years working in industry at ST Micro (Milan), ARM (Cambridge) and Sensata Systems (Swindon) with Marie Curie and Royal Society fellowships. His main area of expertise is in the design of hardware architectures and heterogenous systems for signal processing and machine learning with a focus on run-time adaptation, high-performance via parallelism and energy-efficiency.

Date: Friday, 2nd Dec 2022, @ 4pm GMT+8, https://cityu.zoom.us/j/96742093029