In Favour of Posting

THE CHINESE UNIVERSITY OF HONG KONG
Institute of Network Coding
and
Department of Information Engineering
Seminar

Block Minifloat Arithmetic for Deep Learning Inference and Training

by

Prof. Philip Leong

University of Sydney, Australia

Date : 24 March 2023 (Friday)
Time : 2:00pm – 3:00pm
Venue : Rm833, Ho Sin Hang Engineering Building, CUHK

Abstract

In this talk we present Block Minifloat (BM) arithmetic, a parameterised minifloat format which is optimised for low-precision deep learning applications. While standard floating-point representations have two degrees of freedom, the exponent and mantissa, BM exposes an additional exponent bias allowing the range of a block to be controlled. Results for inference, training and transfer learning using 4-8 bit precision which achieve similar accuracy to floating point will be presented.

Biography

Philip Leong received the B.Sc., B.E. and Ph.D. degrees from the University of Sydney. In 1993 he was a consultant to ST Microelectronics in Milan, Italy working on advanced flash memory-based integrated circuit design. From 1997-2009 he was with the Chinese University of Hong Kong. He is currently Professor of Computer Systems in the School of Electrical and Information Engineering at the University of Sydney, Visiting Professor at Imperial College, Chief Technology Advisor to ClusterTech and Chief Technology Officer at CruxML.

**** ALL ARE WELCOME ****

Host: Prof. Raymond Yeung (Tel: 3943-8375, Email: whyeung@ie.cuhk.edu.hk)
Enquiries: Department of Information Engineering, CUHK (Tel.: 3943-8385)