HIGH PERFORMANCE COMPUTING MODERNIZATION PROGRAM
RESEARCH PROJECT #: HPCMP-HIP-24-024

Benchmarking EfficientNet and EfficientDet on HPC, Edge and Neuromorphic Devices

About ERDC-ITL:

The U.S. Army Engineer Research and Development Center (ERDC) in Vicksburg, Mississippi, is the premier research and development laboratory complex for the Corps of Engineers. The ERDC DSRC operates a variety of large HPC systems to serve the HPC needs of engineers and scientists throughout the DoD.

RESEARCH LOCATION: Vicksburg, MS

PROJECT DESCRIPTION:

The ITL Supercomputing Research Center at ERDC has established an Edge Device Lab to bolster Edge Computing capabilities, particular in the domains of Edge AI, HPC Edge, and HPC Deployable. Notably, two cutting-edge neuromorphic devices have been added to the lab resources: the Intel Loihi-2 8-chip Kapoho Point and IBM NorthPole. The proposed project focuses on benchmarking EfficientNet and EfficientDet, models currently under training for the detection of invasive submersed aquatic vegetation. The objective is to compile a comprehensive performance matrix, encompassing various performance metrics, for each application's inference models running on HPC systems, Edge devices, and Neuromorphic chips.

ERDC-ITL Edge Computing Lab has benchmarked ESRGANs on traditional HPC, Edge HPC, and edge devices like Nvidia Jetson Xavier, Microsoft Azure Stack Edge Pro, etc. The proposed research extends this effort to benchmark EfficientNet and EfficientDet, the networks that are integral efforts in detecting underwater invasive plants within images captures by cameras mounted on ROVs. Expanded benchmarking research will not only include the data points from the two new devices but also encompasses the additional applications for Spike Neural Network (SNN). It's worth noting that the software stack for each of these two devices exhibits significant differences. Intel advises a training-centric approach for the SNN on the chip, while IBM's process involves neural network compilation and convergence for NorthPole. To explore their SDKs effectively, interns must grasp the hardware associated with these devices, which is vital for seamless integration of the respective SDKs. This comprehensive approach aims to evaluate and optimize the performance of various neural network models across diverse computing environments. Ultimately, these efforts will greatly enhance ERDC-ITL's capabilities.

The proposed project necessitates a diverse team, including interns specialized in Data Science, Computer Science, and Statistics, along with domain experts and hardware specialists. Each student intern will focus on one specific neuromorphic chip, becoming his/her expertise in the problem domain while collaborating with other team members. The challenges facing the interns encompass several aspects. 1) Acquiring the necessary science and engineering knowledge for the tasked application, 2) Utilizing DoD HPC machines for both the benchmarking and DoD application execution, 3) Handling substantial volumes of data in alignment with Intel's training-centric approach, and 4) Implementing quantization techniques to manage limited memory resources on the chip.

By the end of the internship program, the goal is for the interns to possess the skills and expertise to actively engage in the professional community of machine learning and AI researchers, especially in edge computing and neuromorphic computing.

ANTICIPATED START DATE:

May 2024 – Exact start dates will be determined at the time of selection and in coordination with the selected candidate.

QUALIFICATIONS:

Prospective candidates should have a background in Computer Science related skills, with basic programming and Linux environment experience. Some experience with Python, and machine learning frameworks such as TensorFlow or PyTorch would be valuable, as well as familiarity with SSH and command line interfaces.

ACADEMIC LEVEL:

Degree received within the last 60 months or currently pursuing:

- Bachelor's
- Master's
- Doctoral

DISCIPLINE NEEDED:

- Computer, Information, and Data Science
- Mathematics and Statistics
- Engineering
- Environmental and Marines Sciences
- Science & Engineering-related