OOKAMI PROJECT APPLICATION

Date: Feb. 12, 2021

Project Title: The Large Scale Image Classification Benchmark

Usage:

- Testbed

Principal Investigator: Yuewei Lin

- University/Company/Institute:
  Brookhaven National Laboratory

- Mailing address including country:
  Building 725, POBox 5000.
  Upton NY, 11973
  United States.

- Phone number: 631-344-8704

- Email: ywlin@bnl.gov

Names & Email of initial project users:

- Yuewei Lin (ywlin@bnl.gov)

- Quinten De Man (quinten.deman@stonybrook.edu)

- Tony Curtis (anthony.curtis@stonybrook.edu)

- Gaurav Verma: (gaurav.verma@stonybrook.edu)

- Barbara Chapman (barbara.chapman@stonybrook.edu)

Usage Description:

Image classification tasks usually are treated as one of the base tasks in the computer vision field, as such, it is widely used as the baseline to evaluate and optimize the hardware, software platforms and their interactions. We will benchmark image classification datasets with variant scales both in single node

and multiple nodes modes, by using deep convolutional neural networks (CNN), which became a standard model for image based tasks.

Ookami, as a brand new supercomputer, we have limited knowledge about how it can efficiently collaborate with the popular (deep) machine learning platforms, such as Tensorflow and PyTorch. From our benchmark test, we will test different CNN models, with different model sizes, using different platforms. We expect to learn how the deep learning platforms based CNN models can collaborate with the A64fx architecture, as CNN models typically include a large number of highly parallelable operations. We also try to optimize the interactions between the A64fx architecture and platforms.

Computational Resources:

- Total node hours per year: 15,000

- Size (nodes) and duration (hours) for a typical batch job: 8 nodes for 72 hours per run.

- Disk space (home, project, scratch):

    - Home: 20 GB for debugging, visualization and other analyses.
    - Project: 10 TB for shared data, code, etc.
    - Scratch: 20 TB

Personnel Resources:

Basic training for Quinten De Man.

Required software:

- Tensorflow, PyTorch.

- TensorBoard for visualization.