(SD2020-332) F5-HD: Fast Flexible FPGA-based Framework for Refreshing Hyperdimensional Computing

Tech ID: 31915 / UC Case 2020-332-0

BACKGROUND

Hyperdimensional (HD) computing is a novel computational paradigm that emulates the brain functionality in performing cognitive tasks. The underlying computation of HD involves a substantial number of element-wise operations (e.g., addition and multiplications) on ultra-wise hypervectors, in the granularities of as small as a single bit, which can be effectively parallelized and pipelined. In addition, though different HD applications might vary in terms of number of input features and output classes (labels), they generally follow the same computation flow. Such characteristics of HD computing inimitably matches with the intrinsic capabilities of FPGAs, making these devices a unique solution for accelerating these applications.

TECHNOLOGY DESCRIPTION

Researchers from UC San Diego have developed F5-HD, a fast and flexible FPGA-based framework for refreshing the performance of HD computing. F5-HD eliminates the arduous task of handcrafted designing of hardware accelerators by automatically generating an FPGA implementation of HD accelerator leveraging a template of optimized processing elements, according to the applications specification and user's constraint. Comparison to benchmarks revealed that F5-HD provides 86.9× and 7.8× (11.9× and 1.7×) higher energy efficiency improvement and faster training (inference) as compared to an optimized implementation of HD on AMD R9 390 GPU, respectively.

APPLICATIONS

Using the invention architecture to design an accelerator for HD on FPGA.

ADVANTAGES

- F5-HD is a template-based framework that generates FPGA-based synthesizable architectures for accelerating HD computing. It automatically generates FPGA-based accelerator for classification problems.
- Leverages a novel hardware-friendly encoding approach that reduces the required Block RAM accesses, hence, enhances resource utilization.
- Provides the flexibility of customized accuracy by supporting different data-types (viz., fixed-point, binary, and power-of-two), and of customized power consumption bound by trading the parallelism.
- Enables simultaneous training and inference to refine the model without interrupting the system functionality

STATE OF DEVELOPMENT

It is currently a working prototype

INTELLECTUAL PROPERTY INFO

US patent rights are available. University is seeking licensees to commercially develop this technology.

PATENT STATUS

Patent Pending

CONTACT

Skip Cynar
scynar@ucsd.edu
tel: 858-822-2672.

OTHER INFORMATION

KEYWORDS

AI applications, Hyperdimensional (HD) computing, Brain-inspired Hyperdimensional Computing, Machine Learning; FPGA-based Acceleration, Automated Template-based Hardware Generation, hypervectors

CATEGORIZED AS

- **Computer**
  - Software
- **Engineering**
  - Engineering

RELATED CASES

2020-332-0

University of California, San Diego
Office of Innovation and Commercialization
9500 Gilman Drive, MC 0910, ,
La Jolla,CA 92093-0910

Tel: 858.534.5815
innovation@ucsd.edu
https://innovation.ucsd.edu
Fax: 858.534.7345