2014 IEEE 22nd Annual International Symposium on Field-Programmable Custom Computing Machines

Accelerator of Stacked Convolutional Independent Subspace Analysis for Deep Learning-based Action Recognition

Lu He
Electrical & Computer Engineering
University of Massachusetts, Lowell
Lu_He@student.uml.edu

Yan Luo
Electrical & Computer Engineering
University of Massachusetts, Lowell
Yan_Luo@uml.edu

Yu Cao
Computer Science
University of Massachusetts, Lowell
ycao@cs.uml.edu

***Abstract*—Action recognition has been a research challenge in multimedia computing and machine vision. Recent advances in deep learning combined with stacked convolutional Independent Subspace Analysis (ISA) has achieved a better performance superior to all previously published results on several public available data sets. Unfortunately, one major issue in large-scale deployment of this new deep learning-based approach is the unacceptable latency of training with high-dimension data. In this paper, we propose a new hardware accelerator that can reduce the training time substantially for deep learning-based action recognition. Specifically, our proposed approach focuses on accelerating the convolutional stacked ISA algorithm, the core components of the deep learning-based action recognition algorithms. We design parallel pipelines, data parallelisms and look-up table to speed up the algorithm. With an embedded heterogeneous platform consisting of a general purpose processor and a FPGA, we are able to achieve up to 10X speedup for stacked ISA training compared to a software-only implementation.**

***Keywords*-Deep Learning, Independent Subspace Analysis, Accelerator**

I. Introduction

Deep Learning improves the effectiveness of accuracy of action recognition. However, it is impractical to employ the deep learning-based action recognition for real-world large-scale applications due to the slow learning procedure on such high-dimensional data.

In this paper, we focus on accelerating the stacked convolutional Independent Subspace Analysis (ISA) algorithm[1]. We design and implement an FPGA accelerator to speed up stacked convolutional independent subspace analysis for recognizing actions in multidimensional video data (RGB+Depth). Using an embedded heterogeneous platform with general purpose cores and FPGAs, we partition the workload such that time-consuming iterative weight estimation is carried out through parallel pipelines and with look-up tables. We model the system architecture to analyze the speedup potential, and our experiment results show up to 10X better performance is achieved on ISA training.

II. System Architecture

The detailed design of accelerator is shown in Figure 1. In the design, The general purpose microprocessor (Intel Atom) is responsible for overall system control and more importantly for workload partition and data pre-processing for the acceleration on FPGA. The FPGA is programmed to perform the time-consuming but simple functions.

In the FPGA design, we use pipelining, parallel data flows, and look-up tables to accelerate the algorithm. The accelerator reads initial weights and part of input video data to on-chip memory to take advantage of its high bandwidth. After transferring weights and input data, all the subsequent operations work on data in the on-chip memory. There are parallel five pipelines to accelerate the ISA training. Each pipeline starts from matrix multiplication and stops at the matrix multiplication for the next layer.

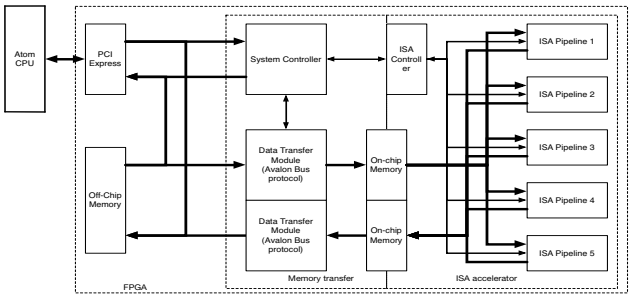

Figure 1. Accelerator Block Diagram

III. Performance Evaluation

Our design is implemented on a DE2i-150 development board, which consists of a dual-core Intel Atom Processor, Altera Cyclone IV FPGA and off-chip memory components. After synthesis, the accelerator engine on FPGA takes 21953 logic elements, 458984 bits memory and 362 embedded multipliers.

In the system, the frequency of PCI Express run at 100MHz and off-chip memory run at 150 MHz. The accelerator engine works at either 25 MHz or 50 MHz. The execution time of ISA training is improved by an order of magnitude (from 3.9 seconds to 0.395 second), compared to software only implementation.

References

[1] Le, Quoc V., et al. "Learning hierarchical invariant spatio-temporal features for action recognition with independent subspace analysis." Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on. IEEE, 2011.

978-1-4799-5111-6/14 $31.00 © 2014 IEEE
DOI 10.1109/FCCM.2014.37