EAI Endorsed Transactions
on Cloud Systems
Research Article EAI.EU

Hardware Acceleration of Computer Vision and Deep Learning Algorithms on the Edge using OpenCL

B. Mishra[1]*, D. Chakraborty[1], S. Makkadayil[1], S. D. Patil[2] and B. Nallani[3]

[1]Intel Corporation, Bangalore, India
[2]Intel Corporation, Bangalore, India during the time of writing the paper
[3]Worked on the project at Intel Corporation, Bangalore, India

Abstract

Machine vision using CNN is a key application in Industrial automation environment, enabling real time as well as offline analytics. A lot of processing is required in real time, and in high speed environment variable latency of data transfer makes a cloud solution unreliable. There is a need for application specific hardware acceleration to process CNNs and traditional computer vision algorithms. Cost and time-to-market are critical factors in the fast moving Industrial automation segment which makes RTL based custom hardware accelerators infeasible. This work proposes a low-cost, scalable, compute-at-the-edge solution using FPGA and OpenCL. The paper proposes a methodology that can be used to accelerate traditional as well as machine learning based computer vision algorithms.

Keywords: CNN, OpenCL, Computer Vision, Machine Learning, Industrial Automation, FPGA, OCR, Hardware Acceleration.

Received on 08 September 2019, accepted on 02 November 2019, published on 05 November 2019

doi: 10.4108/_____

*Corresponding author. Email:bakshree.mishra@intel.com

1. Introduction

Computer vision and machine learning enable industrial environments to become more intelligent and enable more analytics in real time. The industrial environment is very fast moving, and the large number of cameras deployed generate a huge amount of data to be processed. This data enables online as well as offline analytics. Factors such as variable latency of data transfer and data privacy make a cloud solution for such analytics unfavourable. The high speed industrial environment thus calls for application-specific compute-at-the-edge hardware accelerators to process the sensor data using, for example, computer vision algorithms.

A custom hardware accelerator has challenges of its own, including cost of the hardware, as well as time-to-market for the acceleration solution [1] [2]. Field Programmable Gate Arrays (FPGAs) have proven to be reliable accelerators for rapidly changing industries. OpenCL, which is an open source high level synthesis (HLS) framework has further helped in reducing the time-to-market of FPGA solutions for target acceleration.

This paper addresses the critical factors mentioned above for acceleration of Computer Vision based applications, especially for industrial environments. In this paper, we propose a solution methodology for hardware acceleration of Convolutional Neural Networks (CNNs) based on a combination of a Cyclone V FPGA and an Intel Atom Processor. This methodology can be also implemented to accelerate traditional Computer Vision algorithms.

Convolutional neural networks are a class of machine learning algorithms which work on multiple layers of image convolutions. This can be thought of as a cascade of feature maps from low level features, e.g. directional edges, colours, to higher level features, e.g. complex curvatures or partial regions of objects. There can various types of layers used in CNN, in this work we deal with the following:

- Convolutional layers – An NxN convolution mask that operates on the images from the input or the previous layer. Each layer has many such feature masks.