Accuracy and Energy Analysis of CNN Training Accelerator with Various Floating-Point Formats

Muhammad Junaid, Shoaib Sajid and HyungWon Kim

Department of Electronics Engineering, Chungbuk National University, Korea
Emails: {junaid, shoaib,hwkim} @cbnu.ac.kr

Introduction

CNN (Convolutional Neural network) is one of the most widely used technique for image recognition applications. With its rapid growth, the need for energy efficient acceleration has also increased significantly. In this paper, the relationship between energy and accuracy is evaluated using various Floating-point numbers

This analysis provides a baseline to develop the optimal CNN Training accelerator that represents the best trade-off between energy efficiency and accuracy of the network. The implemented architecture has shown in Fig.1 which includes both inference and training engines.

We implemented the proposed training accelerator optimized for MNIST dataset on FPGA and evaluated its training performance using comparison with a PC having GPU

Motivating Idea

It has been observed that by restricting the Exponent to 8-bits and decreasing the Mantissa Bits in Floating Point formats reduces the power consumption and resources utilization while the affect on Accuracy is minimal. The floating-point formats which we have evaluated are shown in figure 1.

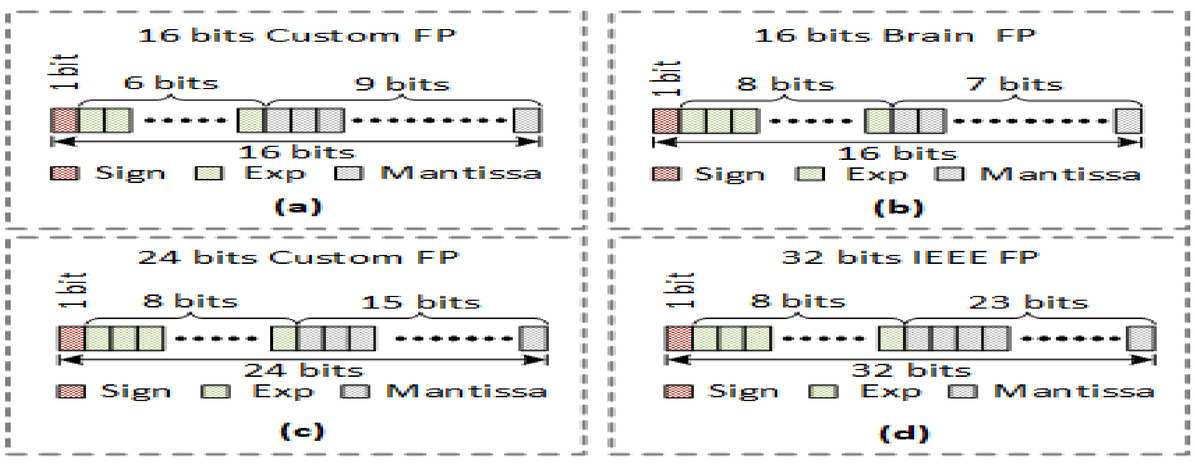

Fig. 1 Floating Point Formats. (a) 16-bits Custom FP. (b) 16-bits Brain FP. (c) 24-bits Custom FP (d) 32-bits IEEE FP

Proposed CNN Architecture

We propose a low power & energy efficient design for training on edge devices. Figure 2 represents our architecture used for both training as well as inference

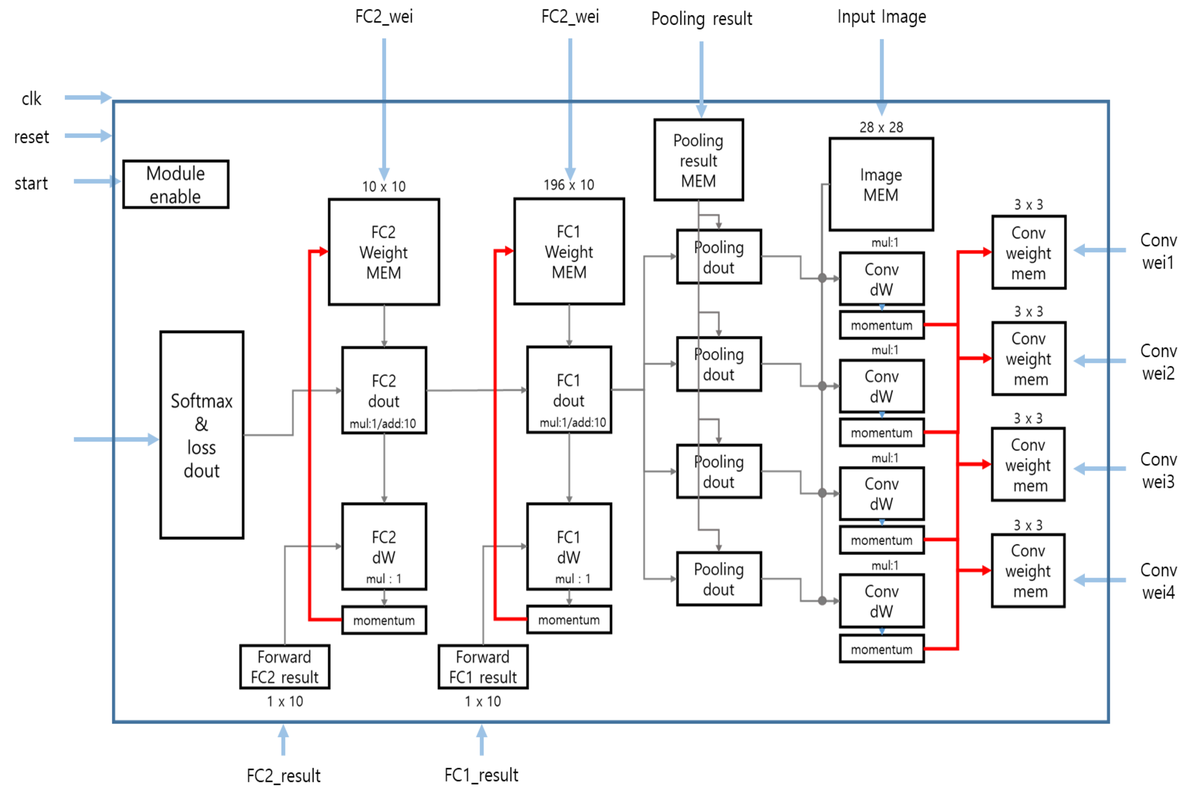

Fig.2 Block Diagram of implemented design

Implementation and Evaluation

We implemented the proposed CNN accelerator for multiple formats on Xilinx FPGA ZCU104 and Raspberry is being used to control the signals. Figure 3 is showing our implemented architecture on FPGA.

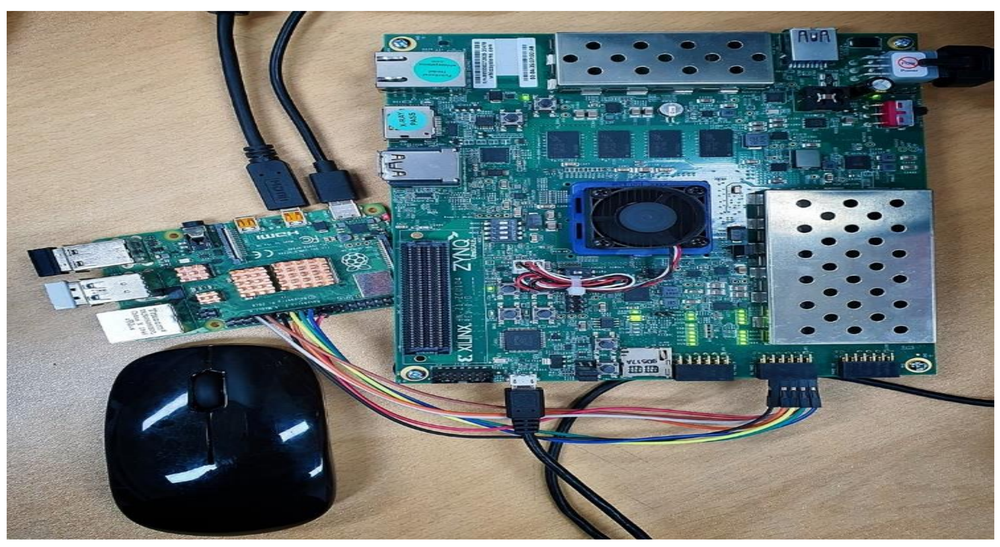

Fig. 3 Test Environment using Xilinx Ultrascale ZCU104 and Raspberry-Pi

Accuracy for multiple formats is shown in table 1, it can be observed that the optimum floating-point format with respect to accuracy and power consumption is 24-bits custom floating point.

Table 1 Accuracy comparison for different Floating-Point Formats

Different Floating Point Formats & their Accuracy					
No. of Bits	**Sign bits**	**Exponent bits**	**Mantissa bits**	**Training Accuracy (50000 images)**	**Test Accuracy (10000 images)**
16	X	Y	Z	9.8%	10%
32	1	8	23	95.4%	94.8%
16 - Bfloat	**1**	**8**	**7**	**91.1%**	**90.0%**
24 - Custom	**1**	**8**	**15**	**94.2**	**93.1%**

Power consumption comparison between different formats has been shown below in figure 4.

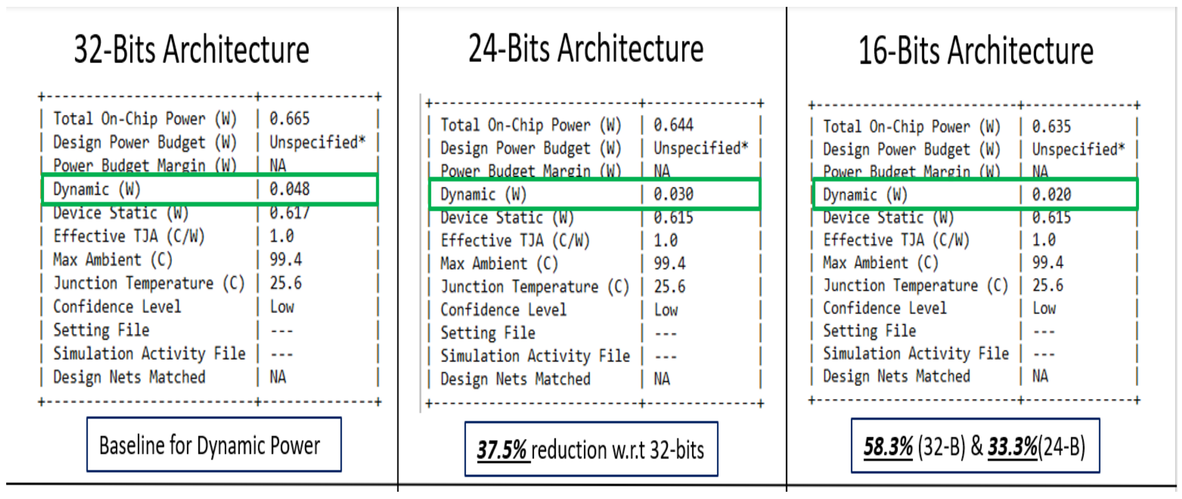

Fig.4 .Power Consumption Comparison.

References

1. **[Andrei Hagiescu., Martin Langhammer., Bogdan Pasca., Philip Colangelo., Jason Thong., Niayesh Ilkhani., BFLOAT MLP Training Accelerator for FPGAs, IEEE, Dec. 2019**
2. **Elsisi., M. Tran., M.-Q., Mahmoud.,Deep Learning-Based Industry 4.0 and Internet of Things towards Effective Energy Management for Smart Buildings. Sensors 2021**
3. **Lecun Y., Bottou, L., Bengio, Y., Haffner P., Gradient-Based Learning Applied to Document Recognition. Proc. IEEE 1998**