Summary

DianNao: A Small-Footprint High-Throughput Accelerator for Ubiquitous Machine-Learning [1]

Cui Weihao

SJTU

E-mail: reallygoodhao@126.com

[1] Tianshi Chen. DianNao: A Small-Footprint High-Throughput Accelerator for Ubiquitous Machine-Learning. ASPLOS, 2014.

Contents

Brief Introduction

In the paper, a novel machine-learning accelerator is designed for large-scale CNNs and DNNs, with a special emphasis on the impact of memory on accelerator design, performance and energy.

1   Main Contributions

- A synthesized (place & route) accelerator design for large-scale CNNs and DNNs, the state-of-the-art machine-learning algorithms.

- The accelerator achieves high throughput in a small area, power and energy footprint.

- The accelerator design focuses on memory behavior, and measurements are not cicumscribed to computational tasks, they factor in the performance and energy impact of memory transfer.

2   Composition of Accelerator

- Storage: an input buffer for input neurons (NBin), an out put buffer for output neurons (NBout), a third buffer for synaptic weights (SB).

- Computations: Neural Functional Unit.

- Control of Accelerator: Control Logic and layer code.

Details of the Accelerator's design

3 Neural Functional Unit (NFU)

The spirit of the NFU is to reflect the decomposition of a layer, which is included in CNNs and DNNs, into computational blocks.

- Arithmetic operators

  The computations of each layer type can be decomposed in either 2 or 3 stages. for classifer layers: multiplication

4 Storage: NBin, NBout and SB

5 Control and code