

합성곱 신경망 기반의 인공지능 FPGA 칩 구현

† 윤영

† 한국해양대학교 전자전기정보공학부 전파융합공학전공 교수

A Realization of CNN-based FPGA Chip for AI (Artificial Intelligence) Applications

† Young Yun

† Division of Electronics and Electrical Information Engineering, Korea Maritime and Ocean University, 727 Taejong-ro, Yeongdo-gu, Busan 49112, Korea

요 약 : 최근 인공지능 분야는 자율주행, 로봇 및 스마트 통신등 다양한 분야에 응용되고 있다. 현재의 인공지능 응용분야는 파이썬을 기반으로 한 tensor flow를 이용하는 소프트웨어 방식을 이용하고 있으며, 프로세서로는 PC의 그래픽 카드 내부에 존재하는 GPU (Graphics Processing Unit)를 이용하고 있다. 그러나 GPU 기반의 소프트웨어 방식은 하드웨어를 변경할 수 없다는 문제점을 가지고 있다. 이러한 문제점으로 인해 높은 수준의 판단이나 작업을 요구하는 경우에는 이에 적합한 높은 사양의 GPU가 필요하며, 이러한 경우에는 인공지능 작업을 처리하는 그래픽 카드로 교체해야 한다. 이러한 문제점을 해결하기 위해 본 연구에서는 HDL (Hardware Description Language)을 이용하여 반도체 내부의 회로를 변경할 수 있는 FPGA (Field Programmable Gate Array)를 기반으로 한 신경망 회로를 이용하여 합성곱 신경망 기반의 인공지능 시스템을 구현하고자 한다.

핵심용어 : 합성곱 신경망, 인공지능, FPGA (Field Programmable Gate Array)

Abstract : Recently, AI (Artificial Intelligence) has been applied to various technologies such as automatic driving, robot and smart communication. Currently, AI system is developed by software-based method using tensor flow, and GPU (Graphic Processing Unit) is employed for processing unit. However, if software-based method employing GPU is used for AI applications, there is a problem that we can not change the internal circuit of processing unit. In this method, if high-level jobs are required for AI system, we need high-performance GPU, therefore, we have to change GPU or graphic card to perform the jobs. In this work, we developed a CNN-based FPGA (Field Programmable Gate Array) chip to solve this problem.

Key words : CNN (Convolution Neural Network), AI (Artificial Intelligence), FPGA (Field Programmable Gate Array)

1. 서 론

최근 인공지능 분야는 자율주행, 로봇 및 스마트 통신등 다양한 분야에 응용되고 있다. 현재의 인공지능 응용분야는 파이썬을 기반으로 한 tensor flow를 이용하는 소프트웨어 방식을 이용하고 있으며, 프로세서로는 PC의 그래픽 카드 내부에 존재하는 GPU (Graphics Processing Unit)를 이용하고 있다. 그림 1은 최근에 GPU 기반의 소프트웨어 방식의 예를 보여준다. 이 방식은 C 언어를 이용하여 CPU (Central Processing Unit)을 이용하여 C 언의 코드를 실행하는 방법과 동일하다. 즉, PC에서 파이썬으로 이루어진 tensor flow 기반의 소프트웨어를 실행하면 GPU가 병렬처리를 이용하여 tensor flow 기반의 소프트웨어 코드를 실행하는 방법이다. 그러나, GPU 기반의 소프트웨어 방식은 하드웨어를 변경할 수 없다는 문제점을 가지고 있다. 이러한 문제점으로 인해 더욱 높은 수준의 판단이나 작업을 요구하는 경우에는 이에 적합한 높은 사양의 GPU가 필요하며, 이러한 경우에는 인공지능 작업을 처리하는

그래픽 카드로 교체해야 한다. 그 이유는 그래픽 카드 내부에 있는 GPU chip은 트랜지스터 소자 간 연결이 고정되어 있으므로, 소프트웨어적인 방법으로 외부에서 회로의 연결상태를 변경하는 것이 불가능하기 때문이다. 결국 GPU기반의 소프트웨어 방식을 사용하는 한, 높은 사양이 요구될 때마다 그 성능에 부응하는 GPU 혹은 그래픽카드로 교체해야 한다.

그리고, 또 하나의 큰 문제점은 GPU기반의 소프트웨어 방식을 사용하는 경우, 인공지능 시스템이 영상을 판단하여 인지하기 위해 GPU chip이 필요하며, 처리결과를 송신하기 위해서는 별도의 무선통신 모듈이 필요하다는 문제점이 있다. 예를 들어, 드론에 인공지능 기능을 부여하여 특정물체를 타겟으로 추적하는 경우, 드론에 GPU가 내장된 그래픽 카드 뿐 아니라 디지털 통신 모듈과 RF 모듈이 필요하게 된다. 그리고, 타겟을 인지한 드론이 타겟에 어떠한 행위를 하기 위해서는 MCU (Micro Control Unit)이 별도로 필요하게 된다. 따라서, GPU기반의 소프트웨어 방식을 이용하여 인공지능 드론을 구현하는 경우,드론의 사이즈는 매우 커지며 드론을 추진하기

위해 상당한 연료가 필요하게 된다. 드론뿐 아니라 로봇의 경우에도 마찬가지이다. 따라서, 인공지능 기능을 가진 드론이나 로봇등을 개발하기 위해서는 하나의 통신/제어/DSP/AI 기능등이 하나의 반도체에 집적된 on-chip 인공지능 시스템이 필요하다.

이러한 문제점을 해결하기 위해 본 연구에서는 HDL (Hardware Description Language)을 이용하여 반도체 내부의 회로를 변경할 수 있는 FPGA (Field Programmable Gate Array)를 기반으로 한 신경망 회로를 이용하여 합성곱 신경망 기반의 인공지능 시스템을 구현하고자 한다.

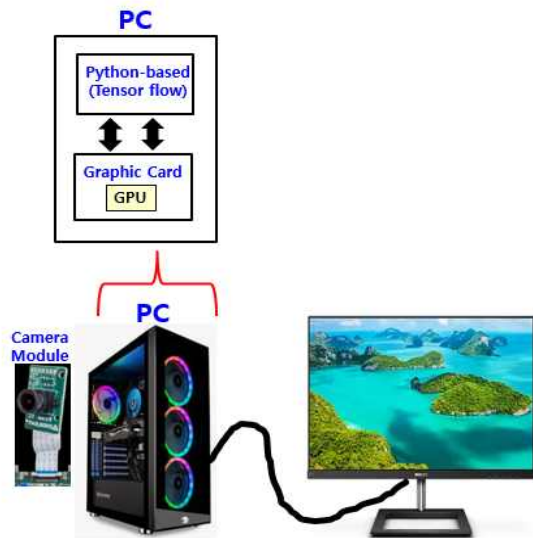


Fig. 1 AI system employing GPU-based software method

2. FPGA 기반 AI 시스템 구현

그림 2는 본 연구에서 제안하는 FPGA를 기반으로 한 신경망 회로를 이용하여 구현한 합성곱 신경망 기반의 인공지능 시스템이다. 구체적으로 본 논문에서는 자일링스 (Xilinx)에서 제공하는 DPU (Deep learning Processing Unit) [1]을 이용하여 합성곱 신경회로망을 FPGA chip 내부에 집적하였다. FPGA 방식을 이용하면 통신/제어/DSP/AI 기능등을 HDL 코딩을 이용하여 하나의 FPGA 내부에 집적할 수 있으며, 인공지능에 더욱 높은 수준의 판단이나 작업이 요구되는 경우에는 HDL 코드를 수정함으로써 DPU를 비롯한 하드웨어의 수정이 가능하다.

3. 결 론

본 논문에서는 DPU (Deep learning Processing Unit)을 이용하여 합성곱 신경회로망을 FPGA chip 내부에 집적하였다.

FPGA 방식을 이용하면 통신/제어/DSP/AI 기능등을 HDL 코딩을 이용하여 하나의 FPGA 내부에 집적할 수 있으며, 인공지능에 더욱 높은 수준의 판단이나 작업이 요구되는 경우에는 HDL 코드를 수정함으로써 DPU를 비롯한 하드웨어의 수정이 가능하다.

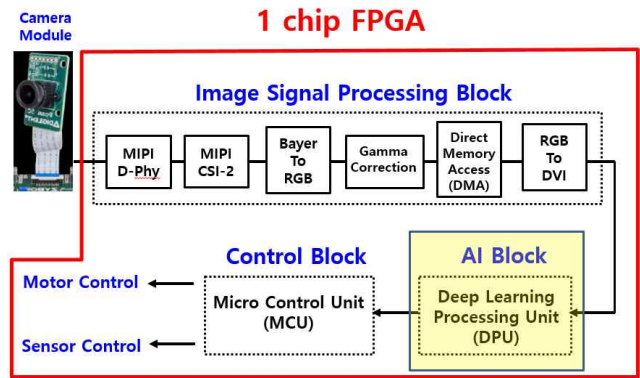


Fig. 2 AI system employing a CNN-based FPGA chip

참 고 문 헌

- [1] <https://www.xilinx.com/products/intellectual-property/dpu.html>