

# 독립운용이 가능한 임베디드 인공지능 프로세서 설계

조권능 · 최도영 · 정영우 · 이승은\*

서울과학기술대학교

## Design of Stand-alone AI Processor for Embedded System

Kwon Neung Cho · Do Young Choi · Young Woo Jeong · Seung Eun Lee\*

Seoul National University of Science and Technology

E-mail : seung.lee@seoultech.ac.kr

### 요 약

모바일 산업의 발달과 인공지능 기술에 대한 관심이 높아지면서 임베디드 시스템에 적용 가능한 인공지능 프로세서에 대한 연구가 활발히 진행되고 있다. 임베디드 시스템에서 인공지능을 구현하는 경우 제한된 자원과 소비 전력을 고려한 설계가 필수적이며, 낮은 연산 성능을 보완할 수 있는 전용 가속기를 포함하는 것이 효율적이다. 본 연구는 독립 운용이 가능한 임베디드 인공지능 프로세서를 제안한다. 제안하는 인공지능 프로세서는 거리연산 기반의 경량 인공지능 알고리즘이 적용된 하드웨어 가속기를 포함하며, 프로그래밍 가능한 범용 프로세서와 함께 운용되어 다양한 임베디드 시스템에 적용 가능하다. 인공지능 프로세서는 Verilog HDL을 사용하여 설계되었으며 Field Programmable Gate Array (FPGA)를 통해 기능을 검증하였다.

### ABSTRACT

With the development of the mobile industry and growing interest in artificial intelligence (AI) technology, a lot of research for AI processors which applicable to embedded systems is under study. When implementing AI to embedded systems, the design should be considered the restriction of resource and power consumption. Moreover, it is efficient to include a dedicated hardware accelerator in order to complement the low computational performance of the embedded system. In this paper, we propose an stand-alone embedded AI processor. The proposed AI processor includes a hardware accelerator that is dedicated to the distance-based AI algorithm and a general-purpose MCU that supports flexible programmability for application to various embedded systems. The AI processor was designed with Verilog HDL and verified by implementing on Field Programmable Gate Array (FPGA).

### 키워드

Embedded AI processor, Embedded system, Hardware accelerator, Verilog HDL, FPGA

### 1. 서 론

반도체 기술과 컴퓨터 시스템이 발전함에 따라 인공지능 기술과 이를 이용한 응용 시스템에 대한 연구가 활발히 진행되고 있다 [1,2]. 인공지능 응용 시스템의 구현은 많은 연산량과 컴퓨팅 자원을 요

구하므로, 고성능의 서버와 데이터센터를 이용한 구현이 주를 이루었다 [3]. 그러나 최근 웨어러블 디바이스, 사물인터넷 (Internet of Things) 등 모바일 산업의 발달과 함께 서버와 데이터센터를 거치지 않고 단일 디바이스에서 인공지능을 수행하는 엣지 컴퓨팅에 대한 관심이 높아지면서 임베디드, 모바일 디바이스 등에 인공지능 알고리즘을 적용하기 위한 경량화 연구가 진행되었다 [3,4]. 임베디

\* corresponding author

드 시스템 환경에서 응용 시스템을 구현하는 경우 제한된 컴퓨팅 자원과 적은 소비전력을 고려한 설계가 필수적이다 [5]. 따라서 딥러닝과 같은 인공지능 알고리즘을 적용하는 것은 무리가 있으며, 임베디드 시스템의 설계 환경을 고려하여 경량화 된 알고리즘을 적용하는 것이 필요하다. 또한 임베디드 시스템의 낮은 연산 성능으로는 인공지능 알고리즘을 수행하는 것에 어려움이 있으므로 인공지능 연산을 위한 전용 하드웨어 가속기를 포함하는 것이 효율적이다 [6].

본 논문에서는 독립운용이 가능한 임베디드 인공지능 프로세서를 제안한다. 제안하는 프로세서는 하드웨어 인공지능 가속기를 포함하며 프로그램 가능한 ARM의 Cortex-M0와 함께 운용되어 다양한 임베디드 시스템에 적용 가능하다. 설계한 인공지능 가속기는 Cortex-M0와 시스템 버스를 통해 인터페이스를 수행하고 사용자는 프로그래밍을 통해 해당 가속기를 제어해 인공지능 기능을 구현할 수 있다. 인공지능 가속기는 거리연산 기반의 학습 및 인식 알고리즘을 전용으로 수행하며, 인식 과정에서 거리연산을 병렬처리 하여 연산을 가속하였다.

한편, UART (Universal Asynchronous Receiver Transmitters), SPI (Serial Peripheral Interfaces), I2C (Inter-Integrated Circuits) 통신 모듈을 설계하여, 프로세서가 PC, 센서 등 외부 시스템과 통신을 수행할 수 있도록 했으며 영상 기반의 인공지능 기능 구현을 위해 외부 카메라 모듈과 인터페이스 할 수 있는 카메라 인터페이스 모듈을 설계하였다. 제안하는 인공지능 프로세서는 Verilog HDL을 사용하여 설계되었으며 Field Programmable Gate Array (FPGA)를 통해 기능을 검증하였다.

## II. 임베디드 인공지능 프로세서 구조

제안하는 임베디드 인공지능 프로세서의 구조는 그림 1과 같다. 프로세서는 두 종류의 시스템 버스 AHB (Advanced High Performance Bus), APB (Advanced Peripheral Bus)를 사용한다. Cortex-M0는 AHB의 마스터로서 AHB 프로토콜에 따라 연결된 주변기기에 접근한다. AHB에는 프로그램 코드 또는 데이터 저장을 위한 내부 메모리가 연결되어 있으며 외부 메모리 인터페이스, 인공지능 가속기 및 GPIO (General Purpose Input Output)가 연결되어 있다. AHB to APB bridge는 AHB 프로토콜을 APB 프로토콜로 변환하여 Cortex-M0가 APB에 연결된 주변기기에 접근할 수 있도록 한다. APB에는 UART, SPI, I2C, Camera Interface와 같이 외부 통신을 위한 모듈이 연결되어 있으며 시간에 따른 제어를 위한 Timer, Watchdog Timer 모듈이 연결되어 있다.

인공지능 가속기는 거리기반의 알고리즘을 수행하여 학습 및 인식을 통한 인공지능 기능을 제공

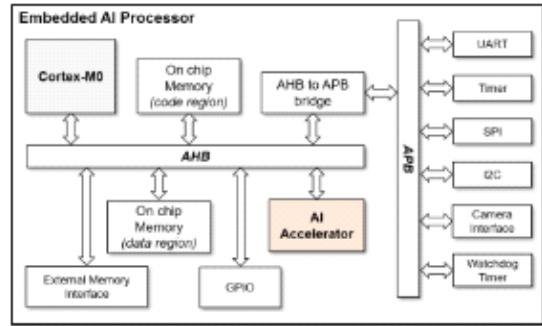


그림 1. 인공지능 프로세서 구조

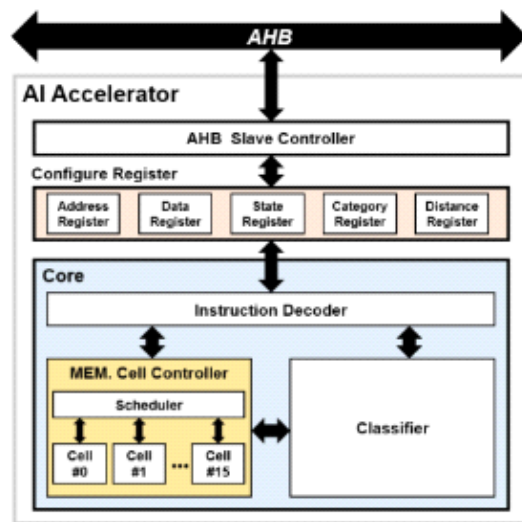


그림 2. 인공지능 가속기 구조

한다. 그림 2는 제안하는 인공지능 가속기의 구조를 나타낸다. 인공지능 가속기의 코어는 Instruction Decoder, Memory Cell Controller, Classifier로 구성되고 Memory Cell Controller는 Scheduler 모듈과 다수의 Cell 모듈을 포함한다. Instruction Decoder는 레지스터 설정을 통해 입력 받은 명령에 따라 학습 및 인식을 수행하기 위한 컨트롤 신호를 생성한다. Scheduler는 발생한 컨트롤 신호에 따라 Cell 모듈을 설정하며 학습 및 인식과정을 제어한다. Cell 모듈 내부에는 학습을 위한 메모리가 존재하고, 하나의 Cell은 하나의 학습 데이터와 학습 데이터에 대한 카테고리 값을 저장할 수 있다. 인식이 수행될 때 Cell 모듈은 학습된 데이터와 입력 받은 인식 데이터와의 거리를 계산하여 출력한다. Classifier 모듈은 출력된 거리 값 중 가장 작은 값을 출력한 Cell의 카테고리를 인식 결과로 출력한다.

인공지능 가속기는 AHB를 통해 범용 프로세서와 인터페이스하기 위해 AHB Slave Controller를

포함한다. 범용 프로세서는 AHB Slave Controller를 통해 인공지능 가속기의 레지스터를 설정하여 인공지능 가속기를 제어할 수 있다. 인공지능 가속기는 Address, Data, State, Category, Distance 레지스터를 포함한다. 사용자는 Address, Data 레지스터를 설정하여 인공지능 가속기에 명령을 전달할 수 있고 State 레지스터 값을 읽어 동작 상태를 확인할 수 있다. 인식을 수행한 후 인공지능 가속기가 인식 결과로 출력한 카테고리 값과 거리 계산 값은 각각 Category, Distance 레지스터에 저장되어 범용 프로세서가 인식 결과를 확인 할 수 있도록 한다.

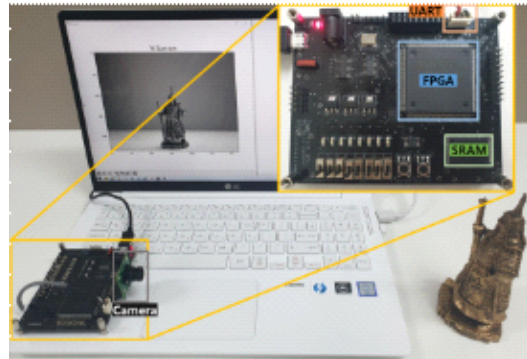


그림 3. 인공지능 프로세서 기능 검증 환경

### III. 구현 및 검증

설계한 임베디드 인공지능 프로세서는 Verilog HDL을 사용하여 설계되었으며 FPGA에 구현되었다. 그림 3은 인공지능 프로세서의 기능을 검증하기 위한 테스트 보드 및 실험 환경이다. 테스트 보드는 FPGA, 외부 SRAM, PC 통신을 위한 UART 포트 및 외부 카메라 인터페이스 포트를 포함한다. 구현된 인공지능 프로세서는 총 16개의 학습 데이터를 학습 할 수 있으며, 학습 및 인식 데이터는 1024-byte 크기를 가진다.

우선, 외부 카메라는 통해 피사체를 grayscale 이미지로 촬영한 뒤 외부 SRAM에 저장한다. 이후 저장된 이미지는 학습 데이터의 크기와 같은 크기로 resize 된다. 전처리된 데이터는 입력받은 카테고리 값과 함께 인공지능 프로세서에 학습된다. 이와 같은 학습 과정을 반복하여 총 4개의 사물에 대한 학습을 수행하였다. 인식 과정도 학습 과정과 같이 카메라를 통해 피사체를 촬영한 후 전처리 과정을 거치며, 인식 명령을 통해 전처리된 데이터를 인공지능 가속기로 입력하여 인식을 수행한다. 인식 결과와 인식한 사물에 대한 원본 이미지는 UART를 통해 PC로 전송되어 출력된다. PC에서 출력된 이미지를 확인하여 카메라 인터페이스 모듈의 동작을 검증하였고, 인식 결과를 확인하여 인공지능 가속기에서 학습 및 인식이 정상적으로 수행되었음을 확인하였다.

### IV. 결론

본 논문에서는 독립 운용이 가능한 임베디드 인공지능 프로세서를 제안한다. 제안하는 인공지능 프로세서는 거리연산 기반의 경량 인공지능 알고리즘이 적용된 하드웨어 가속기를 포함하며, 범용 프로세서와 함께 운용되어 프로그래밍을 통해 다양한 응용 시스템을 구현할 수 있다. 외부 시스템과의 인터페이스를 위한 통신 모듈을 설계하였으며 외부 카메라 모듈로부터 이미지를 입력 받아 영상 기반의 인공지능 기능을 구현 하였다. 설계한

인공지능 프로세서는 Verilog HDL을 사용하여 설계되었으며 FPGA에 구현되었다. 기능 검증을 위해 실험환경을 구축하였고, 영상 기반의 사물인식을 수행하고 결과를 출력하여 기능을 검증하였다.

### References

- [1] Z. Gao and C. Chen, "AI Deep Learning with Multiple Labels for Sentiment Classification of Tweets," in *Proceeding of the 2019 IEEE International Symposium on Circuits and Systems*, Sapporo, pp. 1-5, 2019.
- [2] A. P. James, "Towards Strong AI with Analog Neural Chips," in *Proceeding of the 2020 IEEE International Symposium on Circuits and Systems*, Seville, pp. 1-5, 2020.
- [3] T. Jia, Y. Ju, R. Joseph and J. Gu, "NCPU: An Embedded Neural CPU Architecture on Resource-Constrained Low Power Devices for Real-time End-to-End Performance," in *Proceeding of the 2020 IEEE/ACM International Symposium on Microarchitecture*, Athens, pp. 1097-1109, 2020.
- [4] D. H. Hwang, Y. H. Yoon, C. Y. Han, and S. E. Lee, "Performance Analyzer for Embedded AI Processor", *The Journal of Internet Computing and Services*, Vol. 21, No. 5, pp. 149-157, Oct. 2020.
- [5] G. B. Hwang, K. N. Cho, C. Y. Han, H. W. Oh, Y. H. Yoon, S. E. Lee, "Lossless Decompression Accelerator for Embedded Processor with GUI" *Micromachines* Vol. 12, No. 2, pp. 145, Jan. 2021.
- [6] Y. H. Yoon, D. H. Hwang, J. H. Yang, and S. E. Lee, "Intellino: Processor for Embedded Artificial Intelligence", *Electronics*, Vol. 9, No. 7, pp. 1169, July. 2020.