

FMCW 레이더 기반 FPGA 환경에서의 경량화된 CNN을 이용한 실시간 손동작 인식 시스템 구현

김태영, 이성주*

세종대학교 반도체시스템공학과, *세종대학교 AI융합전자공학과 및 지능형드론융합전공

taeyoung@itsoc.sejong.ac.kr, *seongjoo@sejong.ac.kr

Real-time Hand Gesture Recognition System using Lightweight CNN on FPGA based on FMCW Radar

Taeyoung Kim, Seongjoo Lee*

Dept. of Semiconductor Systems Engineering, Sejong Univ.,

*Dept. of Electrical Engineering and Dept. of AI Convergence Electronic Engineering for Intelligent Drone, Sejong Univ.

요약

최근 인공지능 기술 발전과 함께 데이터 처리 방식이 중앙집중형에서 에지 디바이스 중심의 분산형 방식으로 변화하고 있다. 에지 환경에서는 제한된 전력과 공간에서의 빠른 실시간 데이터 처리가 필수적이나, 기존 GPU는 크기와 전력 소모 문제로 활용하기 어렵다. 이에 높은 병렬 처리 성능과 저전력 특성을 가지는 FPGA가 효과적인 대안으로 주목받고 있다. 본 논문은 Frequency Modulated Continuous Wave (FMCW) 레이더를 이용한 비접촉식 손동작 인식 시스템을 Field Programmable Gate Array (FPGA) 환경에서 구현하였다. 레이더의 Intermediate Frequency 신호는 Short-Time Fourier Transform를 통해 micro-Doppler signature로 변환되고, 손동작의 유형에 따라 독특한 패턴을 나타낸다. 이를 효과적으로 분류하기 위해 Convolutional Neural Network (CNN) 기반의 경량화 모델을 설계하고, Xilinx의 FINN 프레임워크를 이용해 8-bit 양자화하여 하드웨어 친화적으로 최적화하였다. 실험 결과 FPGA 기반 시스템은 CPU 대비 약 7.8배 빠른 추론 속도와 약 7.70배 높은 처리량을 보였으며, 정확도는 미세한 감소(0.52%p)를 나타냈다. 결론적으로 FPGA 기반 CNN 모델이 에지 컴퓨팅 환경에서 우수한 실시간 성능을 제공할 수 있음을 확인하였다.

I. 서론

최근 인공지능 기술의 발전과 함께 데이터 처리 방식이 중앙집중형에서 edge device 중심의 분산형 처리로 변화하고 있다. 이러한 변화는 휴대용 기기나 IoT 센서 등 에지 환경에서 제한된 전력과 공간 내에서 빠른 실시간 처리가 필요하기 때문이다. 기존 GPU는 뛰어난 성능을 가지나, 크기와 전력 소모 문제로 소형 및 저전력 환경에서는 활용이 어렵다. 이에 높은 병렬 처리 성능과 설계 유연성을 가진 FPGA가 대안으로 부각되었으며, 최근 High-Level Synthesis (HLS)를 통해 개발 기간 단축도 가능해졌다. 손동작 인식 기술은 Human-Computer Interaction (HCI)의 성능 향상에 기여하는 중요한 분야로 [1], 특히 레이더 기반의 비접촉 방식이 조도 변화나 개인정보 문제에 강한 장점이 있어 최근 많은 주목을 받고 있다 [2]. 본 논문에서는 FMCW 레이더로부터 획득한 micro-Doppler signature를 FPGA 기반으로 실시간 분류하는 손동작 인식 시스템을 구현하였다. 이를 위해 경량화된 Convolutional Neural Network (CNN)를 설계하고, Xilinx의 FINN 프레임워크 [3]를 활용하여 8-bit 양자화하여 FPGA 환경에서의 효율적인 실시간 성능을 검증하였다.

II. 본론

본 연구에서는 FMCW 레이더를 활용한 손동작 인식 시스템을 구현하였다. FMCW 레이더는 송신된 연속적인 주파수 변조 신호가 물체에서 반사되어 돌아오는 시간 지연과 주파수 천이를 통해 거리와 속도를 측정하는 원리로 동작한다. 손과 같은 미세한 움직임은 Intermediate Frequency (IF) 신호를 Short-Time Fourier Transform (STFT)하여 얻은

micro-Doppler signature로 표현되며, 각 손동작에 따라 독특한 시간-주파수 패턴이 나타난다. 본 연구에서는 총 6가지의 손동작을 선정하고 CNN을 이용해 이를 실시간으로 분류하는 시스템을 구현하였다. Edge device 환경에서 효율적 실시간 추론을 위해 CNN은 convolutional layer, ReLU 활성화 함수, Max-pooling으로 구성된 4개의 convolutional block과 fully-connected layer로 설계하였다. 각 convolutional block은 공간적 특징을 효과적으로 추출하며 Max-pooling을 통해 연산량을 줄이는 방식으로 최적화하였다. 제안한 CNN의 구조는 표 1과 같다.

표 1. 제안된 CNN 모델 구조

Layer	Input size	output size	kernel size
Conv1	(3,64,64)	(16,64,64)	3x3
ReLU1	-	-	-
MP1	(16,64,64)	(16,32,32)	2x2
Conv2	(16,32,32)	(32,32,32)	3x3
ReLU2	-	-	-
MP2	(32,32,32)	(32,16,16)	2x2
Conv3	(32,16,16)	(64,16,16)	3x3
ReLU3	-	-	-
MP3	(64,16,16)	(64,8,8)	2x2
Conv3	(64,8,8)	(128,8,8)	3x3
ReLU3	-	-	-
MP3	(128,8,8)	(128,4,4)	2x2
FC	2048	6	

III. 실험

1. Dataset

* 교신저자

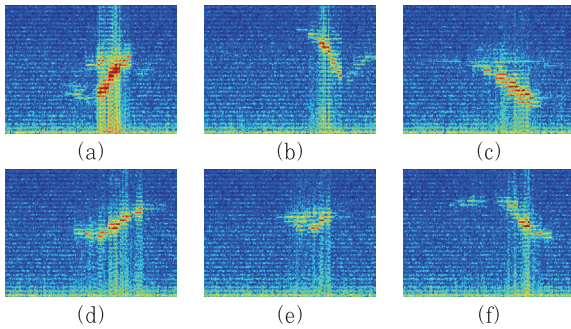


그림 1. 6가지 손동작에 대한 micro-Doppler signature.
(a) Takeoff, (b) Land, (c) Forward, (d) Return,
(e) Cancel, (f) Backward

학습 데이터의 다양성을 위해 총 8명의 참가자(남성 7명, 여성 1명, 나이 23~28세, 신장 165~183cm)가 참여하여 총 6가지 손동작을 측정하였다. 참가자별로 동작당 100회씩, 총 4,800장의 micro-Doppler signature 데이터를 수집하였으며, 측정된 데이터의 예시는 그림 1.과 같다. 이를 Train, validation, test 데이터셋으로 6:2:2 비율로 나누어 활용하였다.

2. 실험 환경

실험 환경은 소프트웨어와 하드웨어로 구성하였다. 소프트웨어 환경에서는 AMD Ryzen 5 5600X 프로세서 기반의 시스템에서 Python과 PyTorch로 모델의 학습과 추론을 진행하였다. 하드웨어 환경으로는 Xilinx사의 ZCU104 FPGA 보드를 이용하였다. FPGA 기반 구현을 위해 FINN 프레임워크 및 Vitis 2022.2 툴을 활용해 CNN 모델의 경량화 및 양자화를 수행하였으며, Xilinx PYNQ 플랫폼에서 FPGA 추론을 수행하였다.

IV. 결과

FMCW 레이더 기반 손동작 인식 시스템을 CNN 모델로 설계하고 FPGA(ZCU104) 환경에 최적화하여 성능 평가를 수행하였으며, 그 결과는 표 2.와 같다.

표 2. CPU 및 FPGA의 추론 성능

	Accuracy	Latency	Throughput
CPU	98.83%	2.247ms	445.03 images/sec
ZCU104	98.31%	0.286ms	3426.72 images/sec

정확도는 FPGA 기반 모델이 CPU 대비 0.52%p 낮았으나, FPGA 기반 시스템은 CPU 대비 약 7.86배 빠른 추론 속도와 약 7.70배 높은 처리량을 나타내었다. 이를 통해 FPGA의 병렬 처리 능력이 에지 환경에서 뛰어난 실시간 성능을 제공함을 확인하였다.

표 3. FPGA 자원 사용량

	Total Resources	Used Resources	Usage (%)
LUT	230400	1226674	54.98
FF	460800	144317	31.32
BRAM_36K	312	61	19.55
URAM	96	3	3.125
DSP	3	1728	100

리소스 사용 측면에서는 FPGA(ZCU104)의 LUT, FF, BRAM, URAM, DSP 리소스 사용량을 분석하였으며 그 결과는 표 3.과 같다. LUT는 54.98%, FF는 31.32%, BRAM_36K는 19.55%, URAM은 3.13% 사용되었

으며, DSP 자원은 100% 모두 활용되었다.

종합적으로 본 연구는 FPGA 기반의 경량화 CNN 모델이 에지 컴퓨팅 환경에서 높은 실시간 추론 성능과 효율성을 가짐을 입증하였다.

V. 결론

본 논문은 FMCW 레이더를 이용한 손동작 인식 시스템을 CNN을 활용하여 설계하고 FPGA를 통해 실시간 추론 환경을 구현하였다. 이를 위해 손동작 데이터를 IF신호로부터 micro-Doppler signature로 변환하여 CNN의 입력 데이터로 사용하였다. CNN 모델은 에지 환경의 실시간 추론 요구 조건을 만족하기 위해 구조가 경량화되었으며, Xilinx사의 FINN 프레임워크를 활용하여 8-bit 양자화를 통해 추가적인 모델 경량화와 FPGA 친화적 구현을 달성하였다.

실험 결과, 제안된 FPGA 기반 시스템은 CPU 기반 시스템과 유사한 정확도를 보이며, 약 7.86배 빠른 추론 속도와 약 7.70배 높은 처리량을 나타냈다. 리소스 활용 측면에서는 FPGA의 DSP 자원을 100% 모두 활용한 것으로 나타났으며, LUT, FF, BRAM 및 URAM 자원도 적정 수준의 사용률을 보였다. 그러나 DSP 자원의 높은 활용률로 인해 추가적인 연산 최적화나 복잡한 병렬 연산 구현에 있어 일부 제약이 발생할 수 있으므로, 향후 DSP 자원의 효율적 활용과 추가 최적화 기법을 통해 보다 높은 성능을 확보할 수 있을 것으로 기대된다.

결론적으로 본 연구는 FPGA를 활용한 CNN 기반의 손동작 인식 시스템이 에지 컴퓨팅 환경에서 높은 실시간 성능을 제공할 수 있음을 입증하였다. 앞으로의 연구에서는 FPGA 자원 활용 최적화와 CNN 구조 개선을 통해 더 높은 정확도와 효율성을 동시에 달성할 수 있는 방법을 모색할 예정이다.

ACKNOWLEDGMENT

본 연구는 과학기술정보통신부의 재원으로 한국연구재단, 무인이동체원천기술개발사업단의 지원을 받아 무인이동체원천기술개발사업을 통해 수행되었으며(No. 2023M3C1C1A01098414) 및 정부(교육부)의 재원으로 한국연구재단의 이공분야 대학중점연구소지원사업의 지원을 받아 수행된 연구임(No. 2020R1A6A1A03038540). 검증을 위한 EDA관련 툴은 IDEC의 지원을 받았음.

참 고 문 헌

- [1] Rakbin Song, Yuna Hong, & Noyoon Kwak (2023). User Interface Using Hand Gesture Recognition Based on MediaPipe Hands Model. Journal of Korea Multimedia Society, 26(2), 103-115.
- [2] Y. Dong and W. Qu, "Review of Research on Gesture Recognition Based on RADAR Technology," in Artificial Intelligence for Communications and Networks, vol. 356, S. Shi, L. Ye, and Y. Zhang, Eds., in Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering, vol. 356., Cham: Springer International Publishing, 2021, pp. 390 - 403.
- [3] Umuroglu, Y.; Fraser, N.J.; Gambardella, G.; Blott, M.; Leong, P.; Jahre, M.; Vissers, K. Finn: A framework for fast, scalable binarized neural network inference. In Proceedings of the 2017 ACM/SIGDA International Symposium on Field-Programmable Gate Arrays, Monterey, CA, USA, 22 - 24 February 2017; pp. 65 - 74.