FPGA 기반 AI 가속에서 PYNQ의 효과적인 활용: Petalinux와의 비교

강유민[1], 민한율[1], 이채빈[1]
[1]서울과학기술대학교 전자IT미디어공학과 학부생
dbals4662@seoultech.ac.kr, croc0001@seoultech.ac.kr, leechaebin@seoultech.ac.kr

Effective Application of PYNQ for FPGA-Based AI Acceleration: A Comparative Research with Petalinux

Yu-min Kang[1], Han-yul Min[1], Chae-bin Lee[1]
[1]Dept. of Electronic and IT Media Engineering,
Seoul National University of Science and Technology

요 약

본 논문은 FPGA 기반의 Petalinux SDK와 PYNQ 프레임워크의 이미지 처리 속도를 비교한다. 연구에서는 YOLO v3 Tiny와 Darknet-19 알고리즘을 사용하여 FPGA에서 자체 제작한 CNN 가속기로 실험을 진행하였다. Petalinux SDK는 이미지 처리에 약 233.13ms가 소요된 반면, PYNQ 프레임워크는 약 2.55ms가 소요되어 더 빠른 속도를 보였다. 이를 통해 PYNQ의 잠재력과 활용 가능성을 강조하며, 추가 연구의 필요성을 제기한다.

1. 서론

최근 HW를 통한 AI 가속 구현 연구가 활발히 진행되며, 특히 Computer Vision 분야에서 HW를 이용한 이미지 처리 기술이 주목받고 있다.[1] HW 가속을 통한 연산 성능 향상은 SW만으로는 얻기 어려운 이점을 제공하며, 이에 따라 FPGA 등의 HW 플랫폼이 중요한 기술로 부상하고 있다.

본 논문은 Petalinux SDK와 PYNQ 프레임워크 두 가지 FPGA 기반 플랫폼을 비교하여 이미지 처리 시간을 측정하는 것을 목표로 한다. FPGA HW에는 자체 제작한 CNN 가속기를 사용하며, AI 모델은 YOLO v3 Tiny, 이미지 처리 알고리즘은 Darknet-19 알고리즘을 차용하였다.

Petalinux SDK는 Xilinx FPGA와 호환되는 임베디드 시스템 개발 도구이다. 이 도구를 사용하면 Xilinx HW에 맞는 임베디드 리눅스를 FPGA 보드 상에서 구동할 수 있다. Petalinux SDK는 복잡한 작업을 지원하는 다양한 유틸리티가 포함된 통합 환경으로 구성되어 있다.[2]

PYNQ 프레임워크는 ZYNQ 기반 FPGA HW를 프로그래밍하기 위한 고급 파이썬 인터페이스를 제공하는 Xilinx의 개발 플랫폼이다. PYNQ 라이브러리와 API, Overlay, Python 언어와 Jupyter-notebook을 통해 FPGA에 배치된 HW와 상호작용할 수 있다.[3]

본 논문은 위 두 개의 임베디드 플랫폼의 이미지 처리 속도를 분석하고자 한다.

2. 연구 환경

2.1 프로세서 사양

연구에 사용된 FPGA 보드의 프로세서 사양은 아래와 같다.

[표 1] 프로세서 사양

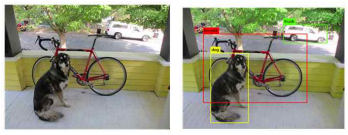
CPU	Cortex-A9 dual-core processor 650MHz
RAM	DDR3 512MB @ 1050Mbps

2.2 운영 체제

Petalinux SDK와 PYNQ 프레임워크 모두 BOOT 이미지 Ubuntu 18.04 LTS 환경에서 연구를 진행하였다.

2.3 입력 이미지 및 결과

[그림 1] 입력 및 출력 이미지



[표 2] 결과 데이터

물체	예측률
dog	100%
bicycle	99%
truck	92%

입력 이미지는 YOLO 모델의 대표 예제 이미지를 사용하였다.[4] 같은 YOLO v3 Tiny 모델을 이용하

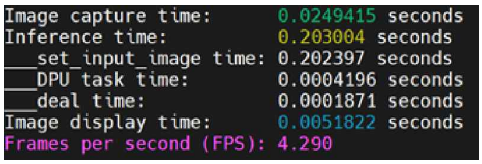
여 두 플랫폼 모두 같은 결과를 나타내었다.

3. 플랫폼 속도 비교

본 연구에서는 AI 가속을 위한 이미지 처리 속도 속도를 평가하였다.

3.1 Petalinux SDK

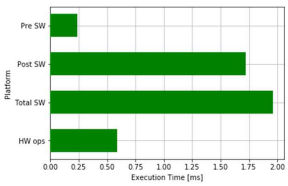
[그림 2] Petalinux SDK 이미지 처리 시간



Petalinux SDK를 사용한 실험 결과, 이미지 전처리 구간은 Image capture time과 set input image time을 합한 약 227.34ms가 소요되었으며, HW 연산 시간은 HW task time과 deal time을 합한 약 0.61ms이 소요되었다. 이미지 후처리 구간은 5.18ms가 소요되어 총 SW 처리 시간은 약 232.52ms가 소요 되었다. 한 장의 이미지를 처리하는 데에는 약 233.13ms가 소요되었다.

3.2 PYNQ 프레임워크

[그림 3] PYNQ 프레임워크 이미지 처리 시간



PYNQ 프레임워크를 사용한 실험 결과, SW 이미지 전처리 구간은 약 0.24ms의 처리 시간을 보였으며, HW 연산 과정에서 약 0.59ms의 처리 시간을 보였다. SW 후처리 구간은 약 1.72ms가 소요되어 총 SW 처리 시간은 약 1.96ms가 소요되었다. 한 장의 이미지를 처리하는 데에는 약 2.55ms가 소요되었다.

3.3 비교 요약 및 정리

플랫폼	처리 영역	이미지 처리 시간[ms]
Petalinux SDK	전체	233.13
	HW	0.61
	SW	232.52
PYNQ 프레임워크	전체	2.55
	HW	0.59
	SW	1.96

두 플랫폼의 속도 비교 결과, 두 플랫폼 모두 HW 기반 연산에서는 약 0.5 ~ 0.6 ms로 유사한 속도를 보였다. 그러나, SW 처리 시간에서 다음과 같은 차이가 나타났다. Petalinux SDK는 232.52ms, PYNQ 프레임워크는 1.96ms로, PYNQ가 Petalinux SDK와 비교해 약 99.16% 더 빠른 속도를 보였다.

4. 결론 및 고찰

4.1 분석 및 고찰

두 플랫폼 모두 하드웨어 연산 속도는 유사했지만, 속도 차이는 데이터 준비 단계에서 발생했다. Petalinux SDK는 입력 이미지 준비에 약 227.34ms가 소요된 반면, PYNQ 프레임워크는 이를 크게 단축하여 Petalinux 대비 약 99.16% 향상된 처리 속도를 보였다. Petalinux SDK는 임베디드 시스템 전반을 개발하는 것에 중점을 두어 이미지 처리와 같은 데이터 준비 단계에서 시간이 더 걸렸다. 반면, PYNQ 프레임워크는 FPGA 보드 상에서 직접적으로 필요한 동작만 처리하므로 데이터 준비와 처리 시간을 크게 단축시켰다.

4.2 결론

PYNQ 프레임워크는 이미지 처리와 같은 AI 가속 작업에서 Petalinux SDK 대비 빠른 속도를 보였으며, 특히 데이터 준비 시간 단축에 있어서 큰 장점을 가졌다. 이러한 결과는 PYNQ가 Xilinx FPGA 기반 AI 가속 응용 프로그램에서 더 나은 선택이 될 수 있음을 제안한다. 다만 본 연구는 Xilinx FPGA 플랫폼에 국한되어 있으므로, 향후 연구에서 다양한 AI 가속 SoC 하드웨어에 대해 다양한 플랫폼의 속도를 비교할 필요가 있다.

※ 본 논문은 과학기술정보통신부 대학디지털교육역량강화 사업의 지원을 통해 수행한 ICT멘토링 프로젝트 결과물입니다.

참고문헌

[1] Zhong, Juan, Zheng Liu, and Xi Chen. "Transformer-based models and hardware acceleration analysis in autonomous driving: A survey." arXiv preprint arXiv:2304.10891. 2023.

[2] 장영수. “딥 러닝 인식 성능을 위한 실시간 셀 셰이딩 알고리즘”. 국내석사학위논문 서울시립대학교 과학기술대학원, 서울, 2023.

[3] Allan, Douglas, et al. Software Defined Radio with Zynq Ultrascale+ RFSoC. No. 1st. Strathclyde Academic Media, 2023.

[4] Redmon, J. "You only look once: Unified, real-time object detection." Proceedings of the IEEE conference on computer vision and pattern. Las Vegas, NV, USA, 2016. pp. 779-788