

효율적인 DCNN 연산을 위한 FPGA 기반 TDC 가속기

장혜림*, 문병인*,**

*경북대학교 일반대학원 전자전기공학부

**경북대학교 IT대학 전자공학부

kitty0421@knu.ac.kr, bihmoon@knu.ac.kr

An Efficient FPGA Based TDC Accelerator for Deconvolutional Neural Networks

Hyerim Jang*, Byungin Moon*,**

*School of Electronic and Electrical Engineering,

Kyungpook National University

**School of Electronics Engineering, Kyungpook National University

요약

딥러닝 알고리즘 중 DCNN(DeConvolutional Neural Network)은 이미지 업스케일링과 생성·복원 등 다양한 분야에서 뛰어난 성능을 보여주고 있다. DCNN은 많은 양의 데이터를 병렬로 처리할 수 있기 때문에 하드웨어로 설계하는 것이 유용하다. 최근 DCNN의 하드웨어 구조 연구에서는 overlapping sum 문제를 해결하기 위해 deconvolution 필터를 convolution 필터로 변환하는 TDC(Transforming the Deconvolutional layer into the Convolutional layer) 알고리즘이 제안되었다. 하지만 TDC를 CPU(Central Processing Unit)로 수행하기 때문에 연산의 최적화가 어려우며, 외부 메모리를 사용하기에 추가적인 전력이 소모된다. 이에 본 논문에서는 저전력으로 구동할 수 있는 FPGA 기반 TDC 하드웨어 구조를 제안한다. 제안하는 하드웨어 구조는 자원 사용량이 적어 저전력으로 구동 가능할 뿐만 아니라, 병렬 처리 구조로 설계되어 빠른 연산 처리 속도를 보인다.

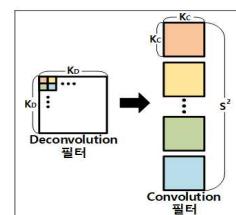
1. 서론

딥러닝 알고리즘 중 CNN(Convolutional Neural Network)은 영상처리 및 컴퓨터 비전 분야에 뛰어난 성능을 보이며, CNN 연산을 효율적으로 처리하기 위해 하드웨어 구조에 대한 연구가 활발히 진행되고 있다[1, 2]. CNN을 응용해 이미지를 업스케일링하는 DCNN(DeConvolutional Neural Network)은 GAN(Generative Adversarial Network), SR(Super Resolution) 등에 다양하게 응용되고 있다[3, 4].

최근 연구에서는 DCNN을 하드웨어로 구현 시 발생하는 overlapping sum 문제를 해결하기 위해 deconvolution 필터를 convolution 필터로 변환하는 TDC(Transforming the Deconvolutional layer into the Convolutional layer) 알고리즘이 제안되었다[3]. 그러나 제안된 DCNN 하드웨어 구조에서는 TDC를 CPU(Central Processing Unit)로 수행하기 때문에 연산의 최적화가 어렵다. 이에 본 논문에서는 TDC 알고리즘을 on-chip 상에서 효율적으로 수행하는 하드웨어 구조를 제안한다.

2. 본론

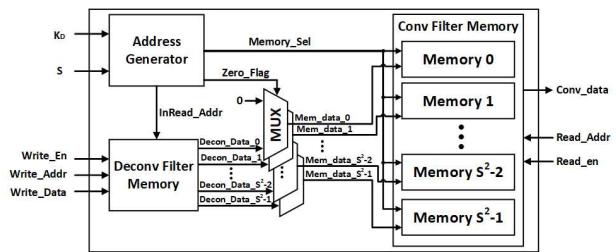
TDC는 그림 1과 같이 크기가 K_D^2 인 deconvolution 필터를 크기가 S (stride)의 제곱 개의 크기가 K_C^2 인 convolution 필터로 변환하는 알고리즘이다[3]. 변환된 convolution 필터들은 서로 의존성이 없으므로, TDC 연산은 병렬로 수행될 수 있다.



(그림 1) TDC 알고리즘

본 논문에서 제안하는 TDC 하드웨어 구조는 그림 2와 같으며, 주소 생성기(Address Generator), 입력 필터 메모리(DeConv Filter Memory), 출력 필터

메모리(Conv Filter Memory)로 구성된다. 주소 생성기는 K_D 와 S 에 대한 파라미터 값을 입력받으며, 내부적으로 생성된 픽셀의 좌표를 통해 convolution 필터의 가중치에 대한 주소를 계산한다. 입력 필터 메모리는 deconvolution 필터를 저장하며, 계산된 convolution 필터의 주소에 해당하는 deconvolution 필터의 주소와 각각 매핑한다. 출력 필터 메모리는 매핑된 S^2 개의 convolution 필터를 해당하는 주소의 메모리에 저장한다.



(그림 2) 제안하는 TDC 하드웨어 구조

3. 실험 및 결과

본 논문에서 제안된 TDC 가속기는 TDC에 사용되는 K_D 와 S 값을 각각 최대 9, 4까지 유동적으로 처리할 수 있도록 설계되었다. 설계된 하드웨어 구조의 자원 사용량 검증을 위해 Xilinx 사의 Vivado 를 사용하여 xc7z020clg484-1 chip을 기반으로 합성 하였으며, 합성 결과는 표 1과 같다. 또한, 제안된 하드웨어 구조의 성능 검증을 위해 CPU와의 TDC 수행 시간을 비교하였으며, 비교 결과는 표 2와 같다. CPU는 Python으로 구현하여 TDC를 수행하는데 소요된 시간으로 도출하였으며, 제안된 구조는 clock period와 (cycle+1)의 곱을 통해 도출하였다. 제안한 하드웨어 구조는 자원 사용량이 적어 저전력으로 구동 가능할 뿐만 아니라 병렬 처리 구조로 설계되어 연산에 소요되는 시간이 매우 적은 것을 확인할 수 있다.

<표 1> Vivado 합성 결과

Estimation		
Resource	LUT	59
	FF	9
Power	Static Power	0.104 W
	Dynamic Power	0.003 W
	Total On-Chip Power	0.108 W
Timing	Clock Period	5.000 ns
	Arrival Time	6.822 ns
	Required Time	7.270 ns
	Slack	0.448 ns

<표 2> CPU와 제안한 하드웨어 간 속도 비교

TDC에 소요되는 시간	
CPU(Python)	850 us (3.4 GHz 기준)
제안한	최소 10 ns ($S=0$ 일 때)
하드웨어	최대 130 ns ($K_D=9, S=1$ 일 때)

4. 결론

본 논문에서는 FPGA 기반 TDC 가속기를 제안하였다. 제안된 하드웨어 구조는 저면적으로 구현되며, TDC 연산을 병렬로 처리하기 때문에 속도 또한 크게 줄일 수 있다. 제안된 구조를 DCNN에 적용할 경우 DCNN의 속도를 보다 향상시킬 수 있으며, 저전력 및 실시간성이 요구되는 DCNN 응용 시스템에 활용 가능할 것이다.

ACKNOWLEDGMENT

본 연구는 IDEC에서 EDA Tool을 지원받아 수행하였습니다.

이 논문은 2021년도 정부(산업통상자원부)의 재원으로 한국연구재단-설종아동 등 신원확인을 위한 복합인지기술개발사업의 지원을 받아 수행된 연구임 (2018M3E3A1057248).

참고문헌

- [1] Y. Ma, Y. Cao, S. Vrudhula and J.-S. Seo, "Optimizing the convolution operation to accelerate deep neural networks on FPGA," IEEE Trans. Very Large Scale Integr. (VLSI) Syst., vol. 26, no. 7, pp. 1354 - 1367, July 2018.
- [2] Y.-H. Chen, T. Krishna, J. S. Emer and V. Sze, "Eyeriss: An energy-efficient reconfigurable accelerator for deep convolutional neural networks," IEEE J. Solid-State Circuits, vol. 52, no. 1, pp. 127 - 138, Jan. 2017.
- [3] J.-W. Chang, K.-W. Kang and S.-J. Kang, "An energy-efficient FPGA-based deconvolutional neural networks accelerator for single image super-resolution," IEEE Trans. Circuits Syst. Video Technol., vol. 30, no. 1, pp. 281 - 295, Jan. 2020.
- [4] A. Radford, L. Metz, S. Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks," CoRR, arXiv:1511.06434, pp. 1-16, 2015.