

경량화된 딥러닝 구조를 이용한 실시간 초고해상도 영상 생성 기술

*안세현, **강석주

서강대학교 전자공학과

*hh585@sogang.ac.kr, **sjkang@sogang.ac.kr

Deep Learning-based Real-Time Super-Resolution Architecture Design

*Saehyun Ahn, **Suk-Ju Kang

Department of Electronic Engineering, Sogang University

요 약

최근 딥러닝 기술은 여러 컴퓨터 비전 응용 분야에서 많이 쓰이고 있다. 물체 인식, 분류 및 영상 생성 등을 예로 들 수 있다. 특히 초고해상도 변환 문제에서 최근 딥러닝을 사용하면서 큰 성능 개선을 얻고 있다. Fast super-resolution convolutional neural network (FSRCNN)은 딥러닝 기반 초고해상도 알고리즘으로 잘 알려져 있으며, 여러 개의 convolutional layer로 추출한 저 해상도의 입력 특징을 활용하여 deconvolutional layer에서 초고해상도의 영상을 출력하는 알고리즘이다. 본 논문에서는 병렬 연산 효율성을 고려한 FPGA 기반 convolutional neural networks 가속기를 제안한다. 특히 deconvolutional layer를 convolutional layer로 변환하는 방법을 통해서 에너지 효율적인 가속기를 설계했다. 또한 제안한 방법은 FPGA 리소스를 고려하여 FSRCNN의 구조를 변형한 Optimal-FSRCNN을 제안한다. 사용하는 곱셈기의 개수를 FSRCNN 대비 2.4 배 압축하였고, 초고해상도 변환 성능을 평가하는 지표인 PSNR은 FSRCNN과 비슷한 성능을 내고 있다. 이를 통해서 FPGA에 최적화된 네트워크를 구현하여 FHD 입력 영상을 UHD 영상으로 출력하는 실시간 영상처리 기술을 개발했다.

1. 서론

디스플레이 산업의 지속적인 발전에 따라 초고해상도 디스플레이의 표준화가 이루어지고 있으며 이에 따라 기존 저해상도의 영상 콘텐츠를 효율적으로 고해상도의 디스플레이에 출력하는 연구가 활발히 이루어지고 있다. 최근 초고해상도 변환 알고리즘에 딥러닝 기술을 적용하여 높은 성능을 얻어 관련 연구가 늘어나고 있다. 이러한 높은 성능의 기술을 적용하기 위해 하드웨어에 대한 연구도 같이 진행되고 있다. Field programmable gate array (FPGA)는 높은 에너지 효율을 보여줄 뿐만 아니라 빠른 연산 속도를 구현할 수 있기 때문에 대안으로써 부각되고 있다. 우리는 딥러닝을 이용한 초고해상도 영상을 생성하는 알고리즘을 FPGA의 리소스를 고려하여

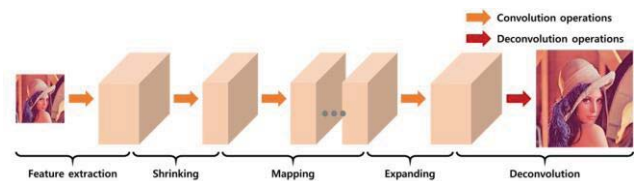


그림 1. FSRCNN 네트워크 구조

최적화하고 구현하였다.

2. 관련 연구

딥러닝을 이용한 초고해상도 알고리즘은 일반적으로 convolutional neural network (CNN) 구조이다. 그림 1과 같이 여러 개의 convolutional layer와 한 개의 deconvolutional layer로 구성되어 있는 fast super-resolution convolutional neural network (FSRCNN) [1]은 CNN 구조의 초고해상도 변환 알고리즘이다. FSRCNN에서 feature maps은 모두

저해상도이다. 반면 SRCNN [2]과 VDSR [3]는 고해상도의 영상을 입력으로 사용하기 때문에 FSRCNN 이 보다 연산 복잡도나 메모리 사용량에 대해서 하드웨어 설계에 유리하다.

Deconvolutional layer 를 하드웨어에 그대로 구현하는 경우 해당 layer 의 연산 특성상 출력 특징맵에서 겹치는 부분에 연산 합 문제가 발생한다. 이는 하드웨어로 구현할 경우 latency 와 에너지 효율 측면에서 치명적이다. 이러한 문제를 해결하기 위해 zero padded deconvolution 연산 방법 [4]-[5]과 deconvolutional layer 를 convolutional layer 로 변환하는 TDC 방법 [6]이 있다.

3. Optimal-FSRCNN 의 하드웨어 가속기 설계

우리는 주어진 하드웨어의 곱셈기 개수를 고려하여 FSRCNN 아키텍처 압축으로 연산 복잡도를 줄였다. 양자화 에러 문제를 해결하기 위해 두 가지 방법을 제안한다. 첫 번째, layer 개수를 가급적 적게 사용한다. 두 번째는 반올림 연산이다. 두 수를 곱셈 연산 후 양자화하는 과정으로 출력 영상의 색깔 틀어짐 문제를 해결할 수 있다. 다음으로 FPGA 의 곱셈기 개수를 고려하여 필터 개수를 결정하였다.

Optimal-FSRCNN 의 각 layer 의 구조가 다르다. 따라서 이러한 네트워크 구조를 하드웨어에 최적화하기 위해서는 multi-convolutional layer processor (CLP) 방법을 이용하여 하드웨어 자원을 최적화했다 [7]. 각 CLP 마다 독립적이고 병렬적으로 데이터를 처리하도록 설계하여 latency 를 최소화했다.

4. 실험 결과

Table 2 은 초고해상도 변환 성능을 평가하는 지표인 peak signal-to-noise ratio (PSNR)을 측정하여 FSRCNN 과 Optimal-FSRCNN 의 성능을 보여준다. FSRCNN 과 비교하여 Optimal-FSRCNN 의 PSNR 은 크게 떨어지지 않으며, 비슷한 성능을 유지한다. 반면 Table 1 처럼 곱셈기의 개수는 약 2.4 배 감소하였다. 연산량 대비 성능이 높아 효율적인 초고해상도 변환 알고리즘이다.

우리는 이러한 효율적인 알고리즘을 Xilinx Kintex7 UltraScale FPGA 에서 가속기를 설계하여 평가하였다. Vivado 2018.2 환경에서 설계하였으며, 구조 내의 parameter 는 16-bit 의 fixed-point 이다. FPGA 리소스는 Table 3 와 같이 사용하였으며, latency 는 78.15 μ s 이다.

5. 결론

이 논문에서, 우리는 경량화된 Optimal-FSRCNN 의 구조를 제안하였으며, 곱셈기의 개수를 2.4 배 적게 사용하면서 비슷한

Table 1. FSRCNN 과 Optimal-FSRCNN 의 곱셈 횟수

네트워크	FSRCNN	Optimal-FSRCNN
곱셈 횟수	10,116	4,261
압축률	1X	2.4X

Table 2. FSRCNN 과 Optimal-FSRCNN 의 PSNR 성능

Test Set	FSRCNN	Optimal-FSRCNN
Set5	36.94 dB	36.52 dB
Set14	32.51 dB	32.29 dB

Table 3. Optimal-FSRCNN 의 FPGA 리소스 사용량

BRAM18K	DSP48E	LUT	FFs
208	4,348	106,796	277,264

성능을 얻었다. 또한 FPGA 의 자원을 효율적으로 활용하여 latency 의 실시간 초고해상도 아키텍처를 제안하였다.

본 연구는 2020 년도 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원 (No. 2020M3H4A1A02084899), 과학기술정보통신부 및 정보통신기획평가원의 대학 ICT 연구센터지원사업의 연구결과 (IITP-2020-2018-0-01421) 및 문화체육관광부 및 한국콘텐츠진흥원의 연구개발지원사업으로 수행되었음 (과제번호: R2020040058)

참고문헌

- [1] Dong Chao, Chen Change Loy, and Xiaoou Tang, "Accelerating the super-resolution convolutional neural network," In *ECCV*, 2016.
- [2] Dong C., Loy C. C., He K., and Tang X., "Image super-resolution using deep convolutional networks," In *TPAMI*, pp.295-307, 2015.
- [3] Kim J., Kwon Lee J., and Mu Lee K., "Accurate image super-resolution using very deep convolutional networks," In *CVPR*, 2016.
- [4] Yazdanbakhsh, A., Samadi, K., Kim, N. S., and Esmaeilzadeh, H., "GANAX: A unified MIMD-SIMD acceleration for generative adversarial networks," In *ISCA*, 2018.
- [5] Song, M., Zhang, J., Chen, H., and Li, T., "Towards efficient microarchitectural design for accelerating unsupervised gan-based deep learning," In *HPCA*, 2018.
- [6] Chang Jung-Woo, and Suk-Ju Kang, "Optimizing FPGA-based convolutional neural networks accelerator for image super-resolution," In *ASP-DAC*, 2018.
- [7] Shen, Y., Ferdman, M., & Milder, P., "Overcoming resource underutilization in spatial CNN accelerators," In *FPL*, 2016.