

깊은 합성곱 신경망 기반 초해상도 모델의 경량화 연구

김대희, 강주희, 유영준, 조영완, 이재구*

국민대학교 컴퓨터공학과

*jaekoo@kookmin.ac.kr

Lightweighting of Deep Convolutional Neural Network Based Super-Resolution Model

Daehee Kim, Juhee Kang, Youngjun Yoo, Youngwan Jo and Jaekoo Lee*
College of Computer Science, Kookmin University.

요약

초해상도는 저화질의 영상을 고화질의 영상으로 변환하는 과업으로, 무거운 연산을 활용하여 성능을 향상하는 연구에 비교적 집중되어 있다. 우리는 깊은 합성곱 신경망 기반 초해상도 모델을 다른 기술과 함께 사용하거나 모바일 환경에서 사용할 때, 모델의 경량화가 필요함을 인지하였다. 이에 따라 깊이별 분리 합성곱을 활용하여, 최신 초해상도 모델인 MSRN 을 경량화한 SMSRN 의 구조를 제안한다. SMSRN 의 매개변수는 MSRN 의 14.64% 수준으로 감량된 반면, 다양한 벤치마크 데이터 집합에 대한 정량적 실험을 진행한 결과로 PSNR 성능 지표는 98.53%를 유지함을 보였으며, 정성적 실험 결과로는 성능 저하를 확인하기 힘든 결과를 보였다. SMSRN 은 다양한 합성곱 필터 크기를 사용한 구조이므로, 다양한 구조의 초해상도 모델에도 적용하여 경량화할 수 있을 것으로 판단된다.

I. 서론

저화질의 영상을 고화질로 변환하는 초해상도(Super-resolution) 과업은 깊은 합성곱 신경망 (Deep Convolutional Neural Network)을 활용하여 급진적으로 발전하고 있다. 초해상도는 컴퓨터 비전의 필수적인 기술로서 광범위하게 연구되고 있는데, 무거운 연산을 활용하여 성능을 높이는 것에 비교적 집중되어 있다. 우리는 이전 연구에서 초해상도를 활용하여 깊은 신경망 기반 객체 인식 기술(Object Detection)의 근본적인 한계를 극복하고자 하였다[7]. 그러나 연구 과정에서 초해상도 모델을 객체 인식과 함께 사용하기 위해서는 실질적으로 모델의 경량화가 필요함을 인지하였다. 더불어 최근에는 고성능의 모바일 기기가 보편화되어 있으므로, 우리는 모바일 환경에서도 사용할 수 있도록 연산이 가벼우면서도 성능이 양호한 깊은 합성곱 신경망 기반 초해상도 모델 구조를 제안하고자 한다.

영상 분류 과업을 목적으로 구성된 MobileNet[1]의 구조에서 깊이별 분리 합성곱(Depth-wise Separable Convolution)이 제안된 이후, CNN 은 획기적인 모델 경량화가 가능하게 되었다. 우리는 이를 활용하여 깊은 합성곱 신경망 기반의 최신 초해상도 모델인 MSRN(Multi-Scale Residual Network) [2]을 경량화하였다. MSRN 에 깊이별 분리 합성곱을 적용하여 SMSRN(Separable Convolution Based Multi-Scale Residual Network) 모델을 [그림 2]와 같이 작성하였으며, 기존의 MSRN 처럼 다양한 합성곱 필터 크기를 사용하면서 성공적으로 경량화하였다.

[그림 1]은 Set5[3] 데이터 집합의 4 배 초해상도 결과로 계산한 PSNR(Peak Signal-to-Noise Ratio) 성능지표와 매개변수의 수를 기준으로 사용하여, SOTA(State Of The Art) 초해상도 모델들과 SMSRN 을 시각적으로 비교한 그래프이다. [그림 1]에서 볼 수 있듯이 SMSRN 은 매개변수의 수가 MSRN 의

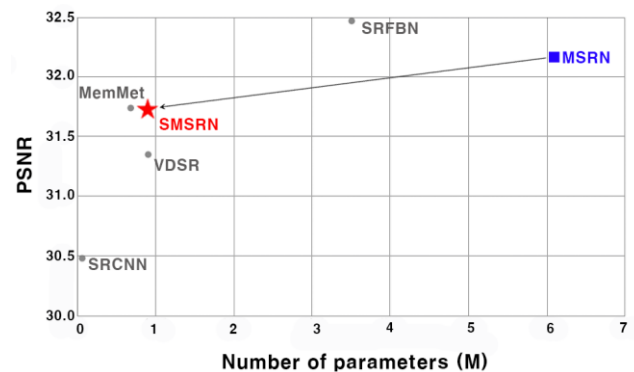


그림 1. SMSRN 과 SOTA 초해상도 모델 비교

14.64% 수준으로, 비약적인 감량에 성공하였다. 또한 다양한 벤치마크 데이터 집합을 기반으로 정량적인 실험을 진행하여 PSNR 성능지표를 비교하였을 때, SMSRN 은 평균적으로 MSRN 대비 98.53%의 성능을 유지하였다.

II. 본론

깊이별 분리 합성곱은 표준 합성곱 연산을 깊이별 합성곱(Depth-wise Convolution)과 점별 합성곱(Point-wise Convolution) 연산으로 분리한 후 연속하여 수행하는 방법이다. 깊이별 합성곱은 특징 맵(Feature Map)의 각 채널마다 독립적으로 합성곱 연산을 수행하는 것이며, 점별 합성곱은 1×1 필터 크기의 합성곱과 같은 과정이다. 이를 통하여 약간의 성능 저하는 있지만 매개변수의 수를 혁신적으로 줄일 수 있다.

MSRN 은 합성곱의 필터 크기를 3×3 과 5×5 로 2 가지를 사용하여 수용장(Receptive Field)의 크기를 다양하게 적용한 초해상도 모델이다. MSRN 의 전체적인 모델 구조는 [그림 2]에서 (a)와 같다. 우리는

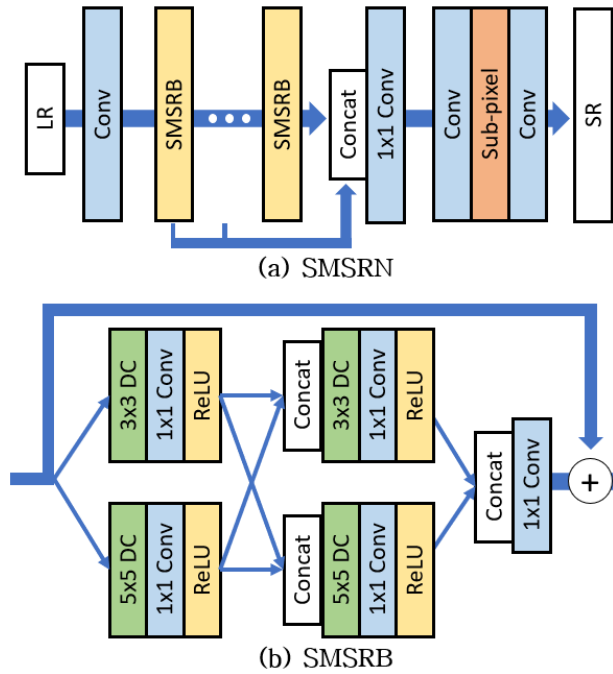


그림 2. SMSRN 모델 구조

[그림 2]의 (b)와 같이 표준 합성곱을 깊이별 합성곱과 점별 합성곱으로 분리, 대체하여 SMSRB(Seperable Multi-Scale Residual Block)를 구성하였다. (b)의 DC는 깊이별 합성곱을 의미한다.

III. 실험

SMSRN을 상세 학습 과정은 MSRN과 같으며, 정량적으로 비교하기 위해 Set5[3], Set14[4], B100[5], Urban100[6] 데이터 집합을 사용하여 실험하였다. [표 1]에서는 각 데이터 집합을 쌍입방 보간법(Bicubic Interpolation)으로 0.25배 다운샘플링(Downsampling)하고, 그 영상을 각각의 모델로 4배 초해상화한 결과를 PSNR과 SSIM(Structural Similarity) 평가 지표로 기록하였고, 매개변수의 수도 포함하였다. 4개의 데이터 집합에 대해 PSNR과 SSIM을 평균하였을 때, 각각 98.53%, 98.27%를 유지한 반면, 매개변수의 수는 MSRN의 14.63%까지 감소시킬 수 있었다. [그림 3]에서 MSRN과 SMSRN의 초해상화 결과를 확인할 수 있듯이 육안으로 확인하기 힘들 정도로 성능을 유지하였다.

IV. 결론

우리는 초해상도 모델을 모바일 환경에서 사용하거나, 다른 기술과 함께 사용할 때 경량화가 필요함을 인지하였다. 그에 따라 깊은 합성곱 신경망 기반 초해상도 모델인 MSRN에 깊이별 분리 합성곱을 적용하여 경량화하였으며, 그 결과로 SMSRN 모델을 제안하였다. SMSRN은 MSRN의 14.63% 수준의 매개변수로 구성되었으며, 이에 반해 PSNR 성능지표

Dataset	MSRN[2]	SMSRN
Set5[3]	32.16 / 0.8642	31.73 / 0.8558
Set14[4]	28.63 / 0.7460	28.30 / 0.7354
B100[5]	27.61 / 0.7139	27.39 / 0.7051
Urban100[6]	26.16 / 0.7679	25.45 / 0.7423
Average	28.64 / 0.7730	28.22 / 0.7597 (98.53%)/(98.27%)
Number of parameters	6,078,019	888,929 (14.63%)

표 1. SMSRN의 정량적 성능 비교 (PSNR / SSIM)

기준으로 98.53%를 유지하였다. SMSRN은 다양한 합성곱 필터 크기를 사용하는 모델임에도, 깊이별 분리 합성곱을 통해 성공적으로 경량화하였다. 따라서 여러 다양한 구조의 초해상도 모델에도 깊이별 분리 합성곱을 적용하여 경량화할 수 있을 것으로 판단된다.

ACKNOWLEDGMENT

이 성과는 2020년도 정부(미래창조과학부)의 재원으로 한국연구재단의 지원(No. NRF2018R1C1B5086441)과 과학기술정보통신부 및 정보통신기획평가원의 SW 중심대학지원사업(2016-0-00021)으로 수행된 연구임.

참고 문헌

- [1] Howard, Andrew G., et al. "Mobilenets: Efficient convolutional neural networks for mobile vision applications." arXiv preprint arXiv:1704.04861, 2017.
- [2] Li, Juncheng, et al. "Multi-scale residual network for image super-resolution." In ECCV, 2018.
- [3] Bevilacqua, Marco, et al. "Low-complexity single-image super-resolution based on nonnegative neighbor embedding.", 2012.
- [4] Zeyde, Roman, Michael Elad, and Matan Protter. "On single image scale-up using sparse-representations." In International conference on curves and surfaces, 2010.
- [5] Martin, David, et al. "A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics.", In ICCV, 2001.
- [6] Huang, Jia-Bin, et al. "Single image super-resolution from transformed self-exemplars." In CVPR, 2015.
- [7] 김대희, 김용욱, 유영준, 이재구. "얼굴 검출 성능 향상을 위한 깊은 신경망 기반 영상 화질 개선 연구" 한국통신학회 학술대회논문집, 42-43, 2020.

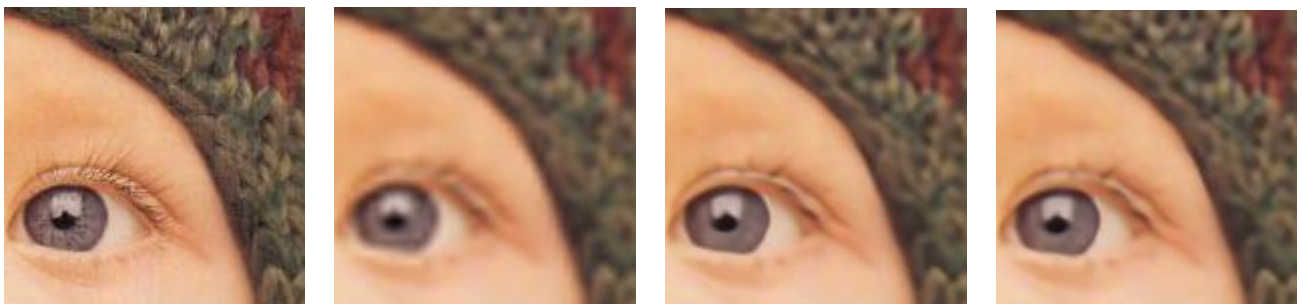


그림 3. Set5[3] 중 baby 영상을 4배 초해상화한 결과이다. 왼쪽부터 원본, 쌍입방 보간법, MSRN[2], SMSRN.