

CNN 을 이용한 단일영상 고해상도 복원 및 수용영역 확장을 통한 성능 향상

박가람, *조남익

서울대학교 전기공학부 뉴미디어통신연구소(INMC)

saturnian77@gmail.com, *nicho@snu.ac.kr

Karam Park, *Nam Ik Cho

Seoul National University Electrical Computer Engineering, INMC

요 약

합성곱 신경망의 성능이 증가하면서 다양한 영상 처리 문제를 해결하기 위해 합성곱 신경망을 적용한 시도들이 증가하고 있다. 고해상도 복원 문제도 그 중 하나였으며, 보다 높은 성능을 얻기 위해 주로 신경망의 깊이를 깊게 하는 시도들이 있었다. 본 논문에서는 고해상도 복원 작업을 위한 합성곱 신경망의 성능 향상을 위해 깊이를 증가시키는 접근법이 아닌 수용영역을 확장시키는 접근법을 시도하였다. 논문에서 제시한 모델은 신경망 내부에 두 개의 브랜치를 두어, 하나의 브랜치는 Dilated Convolution 을 이용해 수용영역을 확장하는데 사용되며, 다른 하나는 이 브랜치를 통해 나온 feature 를 가공하는데 사용된다. 기본 모델은 EDSR 을 사용하였으며, 최종적으로 4.79M 의 파라미터로 평균 32.46dB 의 PSNR 을 보여주었다. 하지만 모델의 구조가 복잡하여 깊이를 늘이는 접근법을 적용하기 어렵다는 한계점이 있다.

1. 서론

컴퓨터 기술이 발전함에 따라 영상을 전자화하여 보관하는 경우가 증가하고 있다. 통신 기술이 보급되기 시작하던 초창기에는 송수신할 수 있는 데이터의 크기가 매우 작았기 때문에 매우 작은 크기와 낮은 화질을 가진 영상만을 컴퓨터에서 처리할 수 있었다. 하지만 통신 기술과 반도체 기술, 그리고 촬영 기술이 점차 발전함에 따라 컴퓨터가 다루는 영상의 크기와 화질 또한 점점 커지기 시작했다. 그럼에도 여전히 효율성 측면에서 데이터의 압축은 필수적이었고, 이러한 과정에서 손실된 영상의 정보를 채울 수 있는 방법에 대해 사람들은 많은 관심을 갖게 되었다.

AlexNet [1]의 성공적인 결과 이래로 합성곱 신경망의 성능이 크게 주목받으면서 컴퓨터를 통한 다양한 영상 및 신호 처리 문제 해결에 합성곱 신경망을 적용하는 시도들이 증가하고 있었다. 단일 영상을 이용한 저해상도에서의 초고해상도로의 복원 문제도 그 중 하나였으며, SRCNN [2]을 시작으로 합성곱 신경망을 이용해 초고해상도 복원 문제를 해결하려는 시도들 또한 많았다.

초고해상도 복원 문제 해결에 가장 큰 문제는 저해상도에서 고해상도로의 복원이 가지는 특성상 복원될 수 있는 결과가 one-to-many 라는 점이였으며, 복원 과정에서 생겨난 대부분의 아티팩트는 창틀의 휘어짐, 건물의 구조의 휘어짐 등 정보 부족이 원인인 복잡한 질감 복원의 실패에서 기인하였다.

이러한 성능 하락의 원인을 해결하기 위해 합성곱 신경망의 깊이를 늘림으로써 수용 영역을 확장하고, 이를 통해 특정 지점의 복원에 기여하는 정보량을 늘림으로써 성능 향상을 시도한 EDSR [3], 각 채널 별 feature 의 평균 값을 비교하고 평균 값이 큰 채널이 가지고 있는 정보가 유의미하다는 가정 하에 채널 별 Attention 을 평균값이 클수록 크게 주어 좀 더 중요한 정보에 가중치를 주는 방식으로 성능

향상을 시도한 CBAM [4], 순환 신경망 구조를 활용하여 매개 변수의 수를 크게 줄이고 저해상도 영상에서 고해상도 영상으로 혹은 고해상도 영상에서 저해상도 영상으로 매핑을 반복하며 얻은 두 종류의 영상들 간의 관계 속에서 최종 결과값을 결정하는 방식을 사용하여 성능을 극대화한 SRFBN [5] 등 다양한 기법들을 적용하여 신경망의 성능 향상을 시도해왔다.

본 논문에서는 Dilated convolution [6]을 이용하여 수용 영역을 확장하고 이것이 합성곱 신경망의 고해상도 복원 성능에 미치는 효과를 확인한다. 해당 모델은 수용 영역 확장을 위해 별도의 브랜치를 사용하였고, 이 브랜치에서 나온 특징값을 처리하는 브랜치를 별도로 사용하여 입력된 영상을 고해상도 영상으로 복원하도록 하였다. 이러한 방식을 통해 직접적으로 아티팩트가 생기는 영역이 얼마나 크게 줄어드는지 보여줌으로써 Dilated convolution 을 이용한 수용영역 확장이 고해상도 영상 복원에 미치는 효과를 확인하였다.

본 논문의 구성은 다음과 같다. 2 절에서는 본 논문에서 제안한 모델의 두 브랜치에 대해서 각각 설명한 뒤, 3 절에서는 실험 방법을 설명한다. 4 절에서는 본 논문에서 제안하는 모델의 성능을 분석하고, 5 절에서는 본 논문에 대한 결론을 맺는다.

2. 신경망 구조

본 논문에서 제안하는 신경망 모델은 저해상도 영상 I_{low} 를 입력으로 받아 고해상도 영상 I_{high} 로 복원한다. 신경망 모델의 구조는 그림 1 과 같이 크게 수용영역 확장을 위한 수용영역 확장 브랜치와 이 브랜치에서 추출한 특징을 가공하는 특징 추출 브랜치로 구성되어 있다. 전체적인 구성은 2 개의 합성곱 레이어 사이에 위치한 반복적인 합성곱 블록 구조와 첫 레이어 이후에서 이어지는 잔차 연결을 가진 EDSR 의 구조를 따랐으며, 세부적인 신경망의 구조는 그림 2 와 그림 3 과 같다.

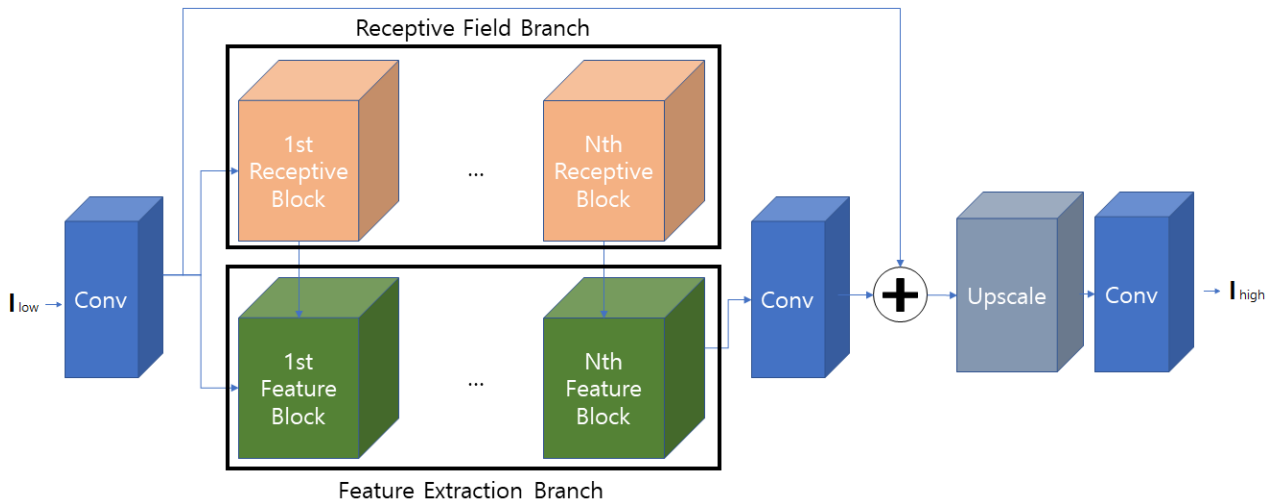


그림 1 제안한 신경망의 구조

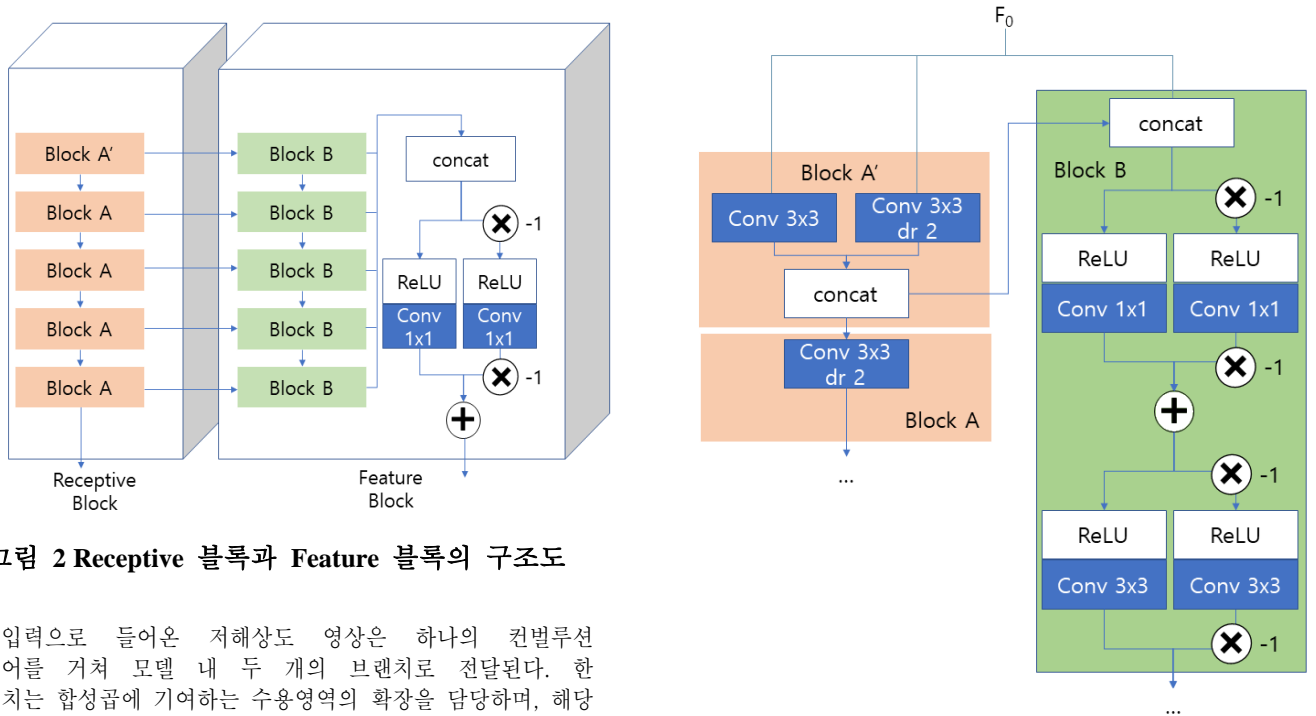


그림 2 Receptive 블록과 Feature 블록의 구조도

입력으로 들어온 저해상도 영상은 하나의 컨벌루션 레이어를 거쳐 모델 내 두 개의 브랜치로 전달된다. 한 브랜치는 합성곱에 기여하는 수용영역의 확장을 담당하며, 해당 브랜치를 구성하는 컨벌루션 레이어는 Dilated Convolution 을 사용하여 필요한 파라미터의 수를 줄였다. 이렇게 Dilated Convolution 을 거친 텐서는 수용영역 확장 브랜치 내부에서 Dilated Convolution 을 거듭해 수용영역의 크기를 확장한다.

동시에 Block A 에 속한 각각의 컨벌루션 레이어는 그림 2 와 3 에서 보이는 것과 같이 다른 하나의 브랜치인 특징 추출 브랜치의 Block B 로 텐서를 전달한다. 이 브랜치는 ReLU 와 컨벌루션 레이어로 구성되어 있으며, 커널의 크기는 1 과 3 을 사용한다. 커널의 크기가 1 인 컨벌루션 레이어는 거듭된 텐서의 접합으로 커진 채널의 수를 줄이는 역할을 한다. 텐서는 커널의 크기가 3 인 컨벌루션 레이어를 거칠 때마다 Dense Connection [7] 을 통해 연결되어 최종적으로 Feature 블록의 출력을 결정하는데 기여한다. 두 브랜치를 거쳐 다듬어진 특징값은 이후 Upscale 모듈을 통해 목표로 하는 해상도에 맞게 가공되어 고해상도 이미지로 출력된다.

그림 3 블록 A 와 블록 B 의 구조도

3. 실험 방법

해당 모델의 학습에는 DIV2K [8] 데이터셋을 사용하였으며, 데이터셋의 800 개의 영상에서 96 픽셀 크기의 정방형 패치들을 추출한 뒤 이를 학습에 사용하였다. 출력으로 하는 영상 크기는 2 배가 되도록 하였다. 그림 3 과 같이 Block A 에서 사용하는 Dilated Convolution 의 Dilation rate 는 2 를 적용하였으며, 두 브랜치에서는 5 개의 블록을 사용하였다. 학습에는 10^{-4} 의 Learning rate 를 사용하였다. 학습에 사용한 Loss 함수는 EDSR 과 마찬가지로 L1 loss 를 사용하여 학습을 진행하였다.

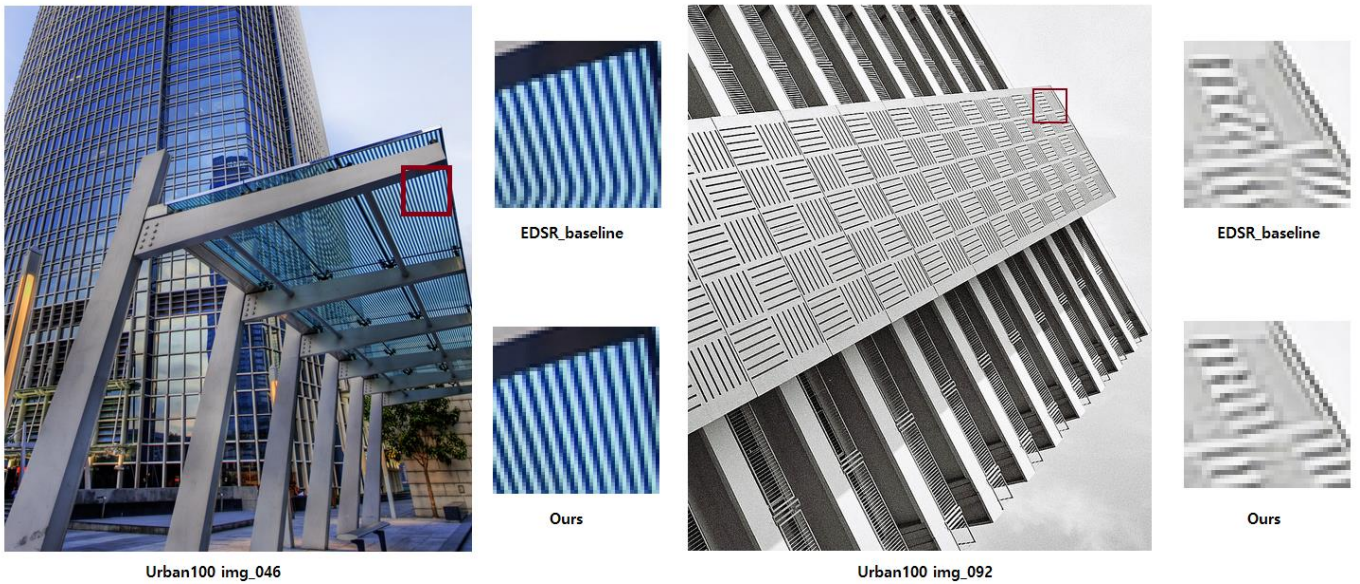


그림 4 EDSR baseline 과 비교한 결과

4. 실험 결과

표 1 Method 별 파라미터 수와 PSNR 비교

Method	Scale	Params	PSNR
LapSRN [9]	×2	0.81M	30.41
VDSR [10]	×2	0.67M	30.76
DRCN [11]	×2	1.77M	30.75
MDSR [12]	×2	6.92M	32.84
CARN [13]	×2	1.59M	31.92
EDSR_baseline	×2	1.37M	31.83
EDSR	×2	40.73M	32.93
Ours	×2	4.79M	32.46

학습을 마친 모델은 Urban100[14] 데이터셋을 사용하여 그 성능을 측정하고 다른 모델과 비교하였다. 서로 조건이 많이 다르지만 어느 정도 비교가 가능하도록 각 모델의 파라미터 수와 PSNR 을 측정하였다. 비교를 위한 여러 다른 모델들과 더불어 실험의 결과는 위의 표 1 에 정리하였다. 수용영역 확장을 위한 브랜치와 특징 추출을 위한 브랜치를 따로 두어 설계하면서 기본이 된 EDSR baseline 에 비해 파라미터의 수가 크게 증가하였다. 본래 설계 의도에 맞게 한 지점의 값을 결정하는데 기여하는 수용영역이 넓어지면서 아티팩트가 EDSR baseline 에 비해 줄어든 것을 그림 4 와 같이 확인할 수 있었다.

하지만 파라미터의 수가 크게 증가한 것에 비해 0.6dB 정도의 PSNR 증가는 큰 성능향상이라 보기 힘들다. 또한 모델을 구성하는 그림 2 의 두 블록을 증가시키는 방식으로 모델의 깊이를 늘려보았으나 유의미한 성능 향상을 확인할 수 없었으며, 이는 모델의 구조가 복잡하며 블록 내부에 많은 수의 컨벌루션 레이어가 존재해 그래디언트 전달이 불안정해졌고 이 때문에 학습이 원활히 진행되지 못한 것으로 보인다.

5. 결론

본 논문에서는 수용 영역 확장을 통해 고해상도 복원 작업을 수행하는 합성곱 신경망의 성능 향상을 시도하였으며, 결과적으로 제안한 모델을 통해 목표한 성능의 향상과 아티팩트 영역의 축소는 확인할 수 있었으나 기존에 존재하는 다른 모델들의 성능과 비교해 유의미한 정도의 성능 개선은 확인할 수 없었다. 향후 모델의 구조적 복잡함을 줄이고 고해상도 복원 작업에 높은 성능을 보여준 다른 모델의 기법들과 복합적으로 사용한다면 보다 효율적으로 파라미터를 사용하면서 성능 향상을 기대할 수 있을 것으로 보인다.

감사의 글

이 논문은 2019 년도 BK21 플러스 사업에 의하여 지원되었음.

참고문헌

- [1] A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. In NIPS, 2012.
- [2] Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang. Learning a deep convolutional network for image super-resolution. In Computer Vision – ECCV 2014 – 13th European Conference, Zurich, Switzerland, September 6–12, 2014, Proceedings, Part IV, pages 184–199, 2014.
- [3] Bee Lim, Sanghyun Son, Heewon Kim, Seungjun Nah, and Kyoung Mu Lee. Enhanced deep residual networks for single image super-resolution. In 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops, CVPR Workshops, Honolulu, HI, USA, July 21–

26, 2017, pages 1132– 1140, 2017.

[4] S. Woo, J. Park, J.-Y. Lee, and I. So Kweon. Cbam: Convolutional block attention module. In ECCV, 2018.

[5] Zhen Li, Jinglei Yang, Zheng Liu, Xiaomin Yang, Gwanggil Jeon, and Wei Wu. Feedback network for image super-resolution. In CVPR, 2019.

[6] F. Yu and V. Koltun. Multi-scale context aggregation by dilated convolutions. arXiv:1511.07122, 2015.

[7] G. Huang, Z. Liu, K. Q. Weinberger, and L. van der Maaten. Densely connected convolutional networks. arXiv preprint arXiv:1608.06993, 2016.

[8] R. Timofte, S. Gu, J. Wu, L. Van Gool, L. Zhang, M.-H. Yang, et al. Ntire 2018 challenge on single image super-resolution: Methods and results. In CVPRW, 2018.

[9] Wei-Sheng Lai, Jia-Bin Huang, Narendra Ahuja, and MingHsuan Yang. Deep laplacian pyramid networks for fast and accurate super-resolution. In 2017 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, Honolulu, HI, USA, July 21–26, 2017, pages 5835– 5843, 2017.

[10] Jiwon Kim, Jung Kwon Lee, and Kyoung Mu Lee. Accurate image super-resolution using very deep convolutional networks. In 2016 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2016, Las Vegas, NV, USA, June 27–30, 2016, pages 1646– 1654, 2016.

[11] Jiwon Kim, Jung Kwon Lee, and Kyoung Mu Lee. Deeplyrecursive convolutional network for image super-resolution. In 2016 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2016, Las Vegas, NV, USA, June 27–30, 2016, pages 1637– 1645, 2016.

[12] Bee Lim, Sanghyun Son, Heewon Kim, Seungjun Nah, and Kyoung Mu Lee. Enhanced deep residual networks for single image super-resolution. In 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops, CVPR Workshops, Honolulu, HI, USA, July 21–26, 2017, pages 1132– 1140, 2017.

[13] Namhyuk Ahn, Byungkon Kang, and Kyung-Ah Sohn. Fast, accurate, and lightweight super-resolution with cascading residual network. In Computer Vision – ECCV 2018 – 15th European Conference, Munich, Germany, September 8–14, 2018, Proceedings, Part X, pages 256– 272, 2018.

[14] J.-B. Huang, A. Singh, and N. Ahuja. Single image superresolution from transformed self-exemplars. In CVPR, 2015.