

잔차 블록 기반의 깊은 합성곱 신경망을 통한 단일 영상 초해상도 복원

김인구, *유송현, **정제창

한양대학교 전자컴퓨터통신공학과

ingu92@hanyang.ac.kr, *3069song@naver.com, **jjeong@hanyang.ac.kr

Image Super-Resolution Using Deep Convolutional Neural Networks Based on Residual Blocks

Kim, Ingu *Yu, Songhyun **Jeong, Jaechang

Department of Electronics and Computer Engineering, Hanyang University

요약

신경망은 깊어질수록 gradient vanishing/exploding과 같은 네트워크가 불안정해지는 문제가 발생 한다. 잔차 블록을 이용하여 이러한 문제를 해결 할 수 있다. 본 논문에서는 영상 인식 분야에서 훌륭한 성능을 보여준 잔차 블록 기반의 깊은 합성곱 신경망을 통한 단일 영상 초해상도 복원 기법을 제안 한다. 제안한 알고리즘은 EDSR에 사용된 잔차 블록을 다양한 크기의 합성곱 연산을 통해 영상의 특징들을 다르게 분석하도록 수정하고 VDSR과 비슷한 수준의 복잡도로 구성하여 향상된 성능을 얻었다. 실험 결과, VDSR에 비해 PSNR이 최대 0.1dB까지 증가했다.

1. 서론

단일 영상 초해상도 (SR: super-resolution)는 저해상도 (LR: low resolution) 입력 영상으로부터 고해상도 (HR: high resolution) 영상을 재구성한다. SR은 외부 정보의 사용여부에 따라서 크게 두 가지로 나눌 수 있다. Glasner 등 [6]은 영상의 자체 유사성을 사용하고 내부의 사전 (dictionary)을 통해 영상을 재구성한다. 외부의 데이터를 사용하는 대표적인 기술은 예제 기반 초해상도 (example-based SR) [5] 가 있다. 예제 기반 초해상도는 대량의 LR과 HR 패치쌍을 데이터베이스에 저장하고 저장된 정보를 참조하여 저해상도 패치에서 고해상도 패치로 맵핑을 하는 방식으로 외부 데이터를 사용하지 않는 방법들에 비해 성능향상을 보여주었다.

최근에는 합성곱 신경망 (CNN: convolutional neural network)을 SR에 적용하여 성능이 크게 향상되었다 [1], [3], [4]. SRCNN [1]은 SR에 합성곱 신경망을 적용한 대표적인 알고리즘으로 3개의 계층을 사용한 end-to-end 구조이다. SRCNN은 저해상도에서 비선형적으로 고해상도로 매핑을 하였고 이전의 SR 기법들에 비해 뛰어난 성능을 보여주었다. VDSR [3]은 SRCNN과 같은 end-to-end 방식에 20개의 합성곱 계층을 쌓고 영상의 잔차 성분만을 학습하기 위해 ResNet [2]에서 제안된 잔차 학습 (residual learning) 을 사용하여 깊은 신경망을 효율적으로

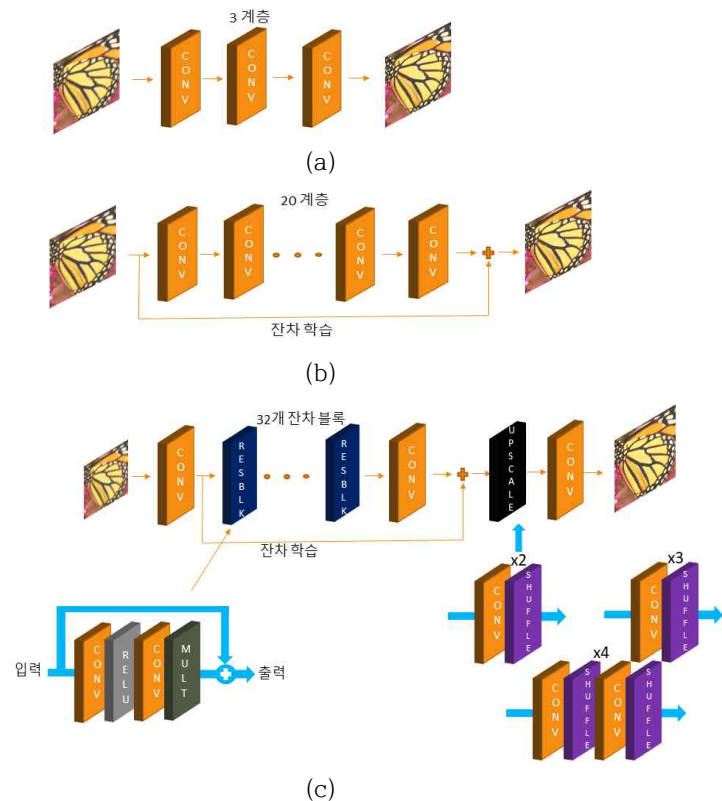


그림 1. SR 네트워크의 구조. (a) SRCNN (b) VDSR (c) EDSR.

로 학습했으며, EDSR [4] 은 잔차 블록의 구조를 SR에 맞게 수정하여 32개의 잔차 블록을 사용하고 전역 잔차 학습과 지역 잔차 학습을 사용하여 성능을 높였다. VDSR과 EDSR 모두 잔차 학습을 사용하였지만 각 네트워크 구조는 한계점이 있다. VDSR은 잔차 학습을 사용했지만 20개의 합성곱 계층을 쌓은 구조 (stacked convolution)로 역전파를 할 때 오차의 기울기가 잘 전달되지 않는다는 단점이 있다. EDSR은 잔차블록 내에서 지역 잔차 학습을 사용하여 local gradient의 흐름을 증가시켜 gradient vanishing/exploding 등의 문제점을 해결하였지만 영상을 분석할 때 단일 커널 사이즈의 사용으로 다양한 특징들을 분석하지 못하는 한계와 많은 파라미터 수 그리고 높은 계산복잡도로 인해 트레이닝 속도가 느리다는 단점이 있다.

본 논문에서는 파라미터 수를 VDSR과 비슷한 수준으로 유지하면서 EDSR에서 사용한 잔차 블록의 구조에 1×1 합성곱 계층을 추가하여 영상의 특징들을 다중 수용 필드로 분석할 수 있도록 하였다.

2. 관련 연구

CNN은 영상 인식과 분류에 많이 사용되는 알고리즘으로 영상에서 다양한 특징들을 추출한다. 최근 CNN을 활용하여 SR에서 훌륭한 성과들을 보여주고 있다. 이에 따라 본 절에서는 최근 딥 러닝 기반의 SR 방법들을 살펴본다.

SRCNN은 SR에 딥 러닝을 적용한 대표적인 알고리즘이다. 그림 1 (a)는 SRCNN의 구조를 나타낸다. 입력은 쌍입방 보간법 (bicubic interpolation)으로 확대된 영상을 사용하고 출력으로 HR 영상을 얻는다. SRCNN의 첫 번째 계층은 영상의 특징을 추출하고 (feature-extraction), 두 번째 계층에서 LR 영상에서 HR 영상으로 비선형 매핑을 하고 (non-linear mapping), 마지막 계층에서 디테일이 개선된 HR 영상을 재구성 한다 (reconstruction). VDSR은 20개의 깊은 합성곱 계층을 가지는 알고리즘이다. 그림 1 (b)는 VDSR의 구조를 나타낸다. 신경망의 입력은 SRCNN과 같은 구조로 보간된 영상을 사용한다. VDSR에서 입력인 LR 영상과 출력인 HR 영상은 많은 부분의 같은 정보를 공유하고 있기 때문에 네트워크를 통하여 HR과 LR 영상의 차이만을 학습하도록 모델링함으로써 학습 속도를 높였다. 그림 1 (a)와 (b)와 (c)에서 각 합성곱 계층 뒤에는 ReLU가 생략되었다. 그림 1 (c)는 EDSR의 구조를 나타낸다. EDSR은 ResNet에서 사용된 잔차블록을 SR에 적합하도록 BN(batch-normalization)을 제거하여 사용하였다. 그 결과 학습시 메모리 사용량을 절감하였고, 절감된 만큼 더 깊은 네트워크를 구성 하였다.

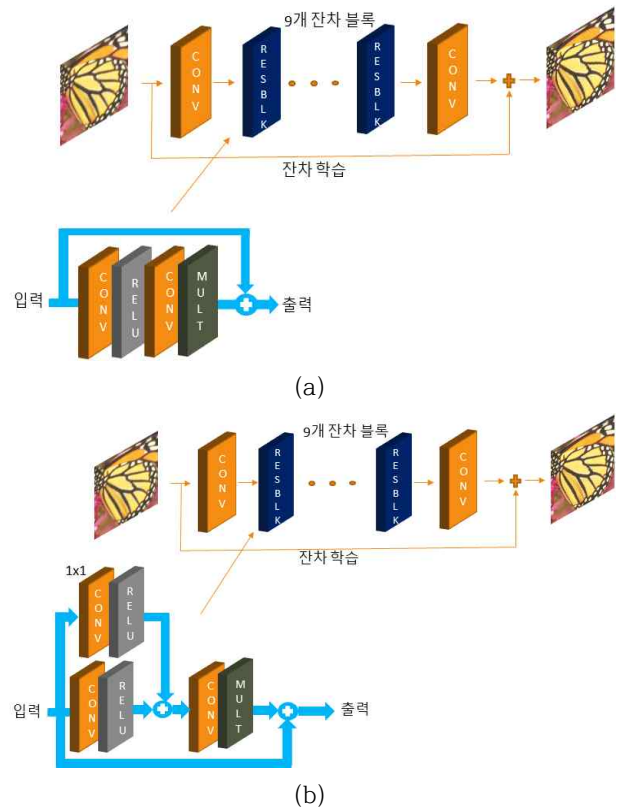


그림 2. 네트워크 구조 비교 (a) 비교를 위한 잔차 블록 기반 네트워크 (b) 제안하는 네트워크

3. 제안하는 방법

3.1 제안하는 네트워크 구조

그림 2 (a)는 EDSR의 잔차 블록구조를 사용한 네트워크 구조이고 VDSR과 동일한 파라미터를 가지도록 9개의 잔차 블록을 사용하였다. 그림 2 (b)는 제안하는 네트워크 구조를 나타낸 것으로 1×1 합성곱 계층과 ReLU를 추가하여 다중 수용 필드를 구성함으로써 입력 영상에 대해 다양한 특징들을 얻는 장점이 있다. 다중 수용 필드를 사용함으로써 VDSR에서 오차 기울기가 잘 전달되지 않는 한계점과 EDSR에서 단일 커널 크기의 사용과 많은 파라미터수를 가지는 한계점을 보완하여 학습시 높은 안정성과 유연한 네트워크를 구축하였다. 제안한 신경망 구조의 세부사항들은 표 1에 설명되어있다.

3.2 깊은 합성곱 신경망 학습

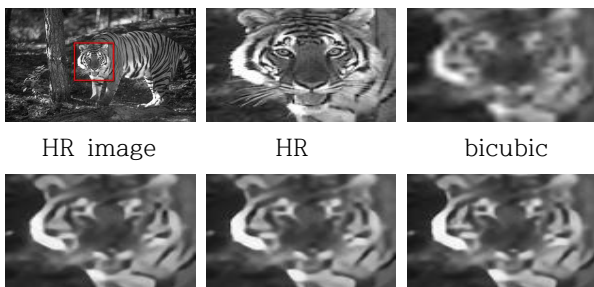
본 논문에서 제안한 네트워크를 학습시 사용한 손실 함수는 아래와 같다.

$$\text{Loss} = \frac{1}{N} \sum_{i=1}^N \| y_i - \hat{y} \| \quad (1)$$

	비교를 위한 네트워크	제안하는 네트워크
필터 크기	3×3	3×3 , 1×1
필터 개수	64	
배치 크기	64	
학습률	0.0005	
패치 크기	41×41	
손실 함수	L1 loss	

표 1. 제안한 신경망 구조의 세부사항

식(1)에서, N 은 학습에 사용한 배치의 크기를 의미하고 y_i 는 ground truth (HR)를 의미하고 \hat{y} 는 신경망을 통과한 출력 영상 (예측된 HR) 의미한다. 초해상도 영상 복원 과정에 사용한 손실함수는 L1 loss 함수를 사용하였다. MSE (Mean Squared Error) 또는 L2 loss 는 일반적으로 영상 복원과 성능 측정에 있어 흔히 사용하는 함수이다. 하지만 Zhao 등 [8]은 L2 loss가 다른 손실 함수에 비해 PSNR 측면에서 더 좋은 성능을 보장하지 않는 것을 보여주었다. 신경망을 학습할 때 Adam 최적화 알고리즘을 사용했고 학습률 (learning rate)은 0.0005로 설정하였다. 신경망 학습 시 입력데이터를 전체 영상이 아닌 일정한 크기인 패치 (Patch) 크기로 나누어 구성하였고 성능을 올리기 위해 영상의 반전/회전을 이용한 data augmentation을 사용해서 학습 데이터를 늘렸다.



VDSR 비교를 위한 네트워크 제안한 방법
(a)

그림3. 주관적 화질 비교. (a) “B100“, 108005, ×4

4. 실험결과

제안한 알고리즘은 291개의 영상 [9]으로 학습을 하였고 HR 영상에 쌍입방 다운샘플링을 사용하여 LR 영상을 얻고 쌍입방 보간법으로 영상의 크기를 재조정하여 보간된 LR 영상을 사용하였다.

최대 신호 대 잡음비 (PSNR: Peak Signal-to-noise ratio)는 영상의 화질 손실 정보를 평가할 때 사용된다. 실험 결과를 비교하기 위해 PSNR을 사용하고 실험 데이터셋으로는 SR 방법에 일반적으로 사용되는 Set5 [10], Set14 [11], B100 [12]을 사용하였고 제안하는 네트워크를 VDSR과 쌍입방 보간법 그리고 비교를 위한 네트워크와 비교하였다. 결과는 표 2에서 보여준다. 표 2에서는 PSNR 측면에서 성능을 비교하면 제안한 방법이 쌍입방 보간법과 VDSR 보다는 성능이 훨씬 좋고 비교를 위한 네트워크와 비교하면 부분적으로 성능이 더 좋은 것을 알 수 있다. 그림 3은 쌍입방 보간법과 VDSR 그리고 비교를 위한 네트워크를 주관적 화질측면에서 비교한 결과이다. VDSR은 눈 주위의 검은띠가 매끄럽게 이어지지 않지만 비교를 위한 네트워크는 VDSR보다 눈 주위의 검은띠가 부드럽게 연결되어 있다. 제안한 방법은 쌍입방 보간법과 VDSR 그리고 비교를 위한 네트워크보다 검은띠가 부드럽게 이어져 있고 흰색 부분이 조금 번져 있다.

5. 결론

기존의 VDSR은 SR에서 뛰어난 성능을 나타내지만 네트워크 구조의 한계를 가진다. 본 논문에서는 잔차 블록 기반의 영상 초해상도 복원 알고리즘을 제안 하였고, 제안된 알고리즘은 잔차 블록을 사용하여 기존의 VDSR보다 오차 기울기를 잘 전달하도록 개선하여 효율적인 신경망을 제안한다. Set5, Set14, B100의 데이터셋을 통해 자연 영상에 대해 실험을 하여, 제안하는 알고리즘이 기존 방법보다 성능이 개선됨을 알 수 있다.

방법	스케일	Set5	Set14	B100
bicubic	2	33.66	30.23	29.56
VDSR		37.53	33.04	31.90
비교를 위한 네트워크		37.63	33.11	31.95
제안한 방법		37.60	33.15	31.95
bicubic	3	30.38	27.54	27.20
VDSR		33.50	29.72	28.75
비교를 위한 네트워크		33.66	29.78	28.79
제안한 방법		33.62	29.79	28.79
bicubic	4	28.39	26.00	25.95
VDSR		31.19	27.88	27.19
제안한 방법1		31.23	27.93	27.23
제안한 방법		31.26	27.95	27.22

표 2. 평균 PSNR 비교 , 붉은색이 가장 좋은 성능

참고문헌

[1] C. Dong, C. C. Loy, K. He, and X. Tang. "Image super-resolution using deep convolutional networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 38, no. 2, Feb. 2016.

[2] K. He, X. Zhang, S. Ren, and J. Sun. "Deep residual learning for image recognition," *IEEE Conference on Computer Vision and Pattern Recognition*, 2016

[3] J. Kim, J. K. Lee and K. M. Lee. "Accurate Image Super-Resolution Using Very Deep Convolutional Networks," *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1646-1654, 2016.

[4] B. Lim, S. Son, H. Kim, S. Nah and K. M. Lee. "Enhanced Deep Residual Networks for Single Image Super-Resolution," *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1132-1140, 2017.

[5] W. T. Freeman, T. R. Jones, and E. C. Pasztor, "Example-based super resolution," *IEEE Comput.*

Graph. Appl. Vol. 22, no. 2, 2002.

[6] D. Glasner, S. Bagon, and M. Irani. "Super-resolution from a single image," In *IEEE International Conference on Computer Vision (ICCV)*, pages 349-356, 2009.

[7] K. Simonyan, A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image recognition," arXiv preprint, arXiv:1409.1556, 2014.

[8] H. Zhao, O. Gallo, J. Frosio, J. Kautz, "Loss functions for image restoration with neural networks," *IEEE Transactions on Computational Imaging*, vol. 3, no. 1, pp. 47-57, 2017.

[9] S. Schulter, C. Leistner, and H. Bischof. "Fast and accurate image upscaling with super-resolution forests," *IEEE Conference on Computer Vision and Pattern Recognition*, 2015.

[10] C. G. Marco Bevilacqua, Aline Roumy and M.-L. A. Morel, "Low-complexity single-image super-resolution based on nonnegative neighbor embedding," *In BMVC*, 2012.

[11] R. Zeyde, M. Elad, and M. Protter, "On single image scale-up using sparse-representations," *In Curves and Surfaces*, pages 711-730. Springer, 2012.

[12] D. Martin, C. Fowlkes, D. Tal, and J. Malik, "A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics," *In IEEE International Conference on Computer Vision (ICCV)*, 2001.