영상 분할 지도를 활용한 영상 잡음 제거

*양혜윤 **장영일 ***소재웅 ****조남익

서울대학교 전기정보공학부 뉴미디어통신연구소

*liz1y98@ispl.snu.ac.kr **jyicu@ispl.snu.ac.kr

***soh90815@ispl.snu.ac.kr ****nicho@snu.ac.kr

Image Denoising Using Image Segmentation Map

*Yang, Haeyoon **Jang, Yeong Il ***Soh, Jae Woong ****Cho, Nam Ik

Department of ECE, INMC, Seoul National University

요약

영상 잡음 제거는 잡음으로 저하된 영상으로부터 잡음 없는 영상을 복원하는 기술이다. 최근 영상 처리에 딥러닝을 사용한 학습 기반 방법 중 저수준 컴퓨터 비전 분야에 고수준 영상 정보를 활용하는 접근이 있었다. 본 논문에서는 고수준 영상 정보인 영상 분할 지도를 활용하여 영상 속 가산 백색 잡음 제거 연구를 진행하였다. 잔차 연결을 활용한 구조의 인공신경망 모델에 잡음 영상, 잡음 수준 지도, 영상 분할 지도를 입력으로 넣어 고수준 영상 정보를 활용할 수 있게 하였다. 본 논문에서 제안한 인공신경망을 Outdoor Scene Dataset과 CBSD68 Dataset에 대해 확인해본 결과, PSNR과 인지적인 측면에서 DnCNN과 FFDNet보다 성능이 향상되는 것을 확인하였다.

1. 서론

영상 잡음 제거(image denoising)는 영상 복원 문제들 중 하나로서, 잡음으로 인해 화질이 저하된 영상에서 잡음 없는 영상을 복원하는 기술이다. 잡음이 있는 영상과 잡음이 없는 원본 영상과의 관계는 다음 수식으로 표현할 수 있다.

$$y = x + n$$

여기서, y는 잡음이 있는 관측 영상이고, x는 잡음이 없는 원본 영상, n은 잡음을 의미한다. 본 논문에서는 잡음을 가산 백색 잡음 (additive white Gaussian noise)으로 가정하며, n은 픽셀당 평균이 0이고 공분산이 σ^2I인 Gaussian 잡음이다.

과거에는 영상 잡음 제거에 주로 Bayesian 모델을 이용한 모델 기반 방법을 사용하였다[1, 2, 3, 4, 5]. 최근에는 합성곱 신경망(convolutional neural network)의 도입과 발달로 합성곱 신경망을 사용한 딥러닝 학습 기반 방법을 주로 사용한다[6, 7, 8, 9]. 합성곱 신경망 기반 방법의 예로는, DnCNN[10]과 FFDNet[11]이 있다. DnCNN은 잔차 학습(residual learning)과 배치 정규화(batch normalization)를 사용해 잡음 제거 성능을 높였고, FFDNet은 잡음 수준 지도(noise level map)를 사용해 하나의 모델로 다양한 수준의 잡음을 제거할 수 있었다.

더 나아가, 영상 잡음 제거와 영상 초해상도 (image super-resolution)과 같은 저수준(low-level) 컴퓨터 비전 분야에 영상 분류(image classification)와 영상 분할(image segmentation)과 같은 고수준(high-level) 컴퓨터 비전 정보를 적용하는 연구도 진행되었다. 예를 들어, SFT-GAN[12]에서는 영상 초해상도 복원에 이미지의 분할 지도(segmentation map)를 사용하여 성능을 높였고, [13]에서는 영상 복원 네트워크와 영상 분할 네트워크를 결합해 두 네트워크의 상호작용을 통해 영상 복원 성능을 높일 수 있었다.

위와 같이 영상 복원에 영상 분할 정보를 사용한다면, 하늘과 같이 스무딩(smoothing)이 많이 필요한 부분과 동물과 풀과 같이 디테일한 질감이 살아야 하는 부분을 나눠서 처리할 수 있다. 이에 따라 본 논문에서는 영상 잡음 제거에 영상 분할 지도를 활용한다. 영상 분할 정보를 cascade로 활용하거나 중간 레이어에 추가하는 기존 방법[12, 13]과는 다르게 영상 분할 지도를 입력으로 사용하였다. 입력으로 사용한 영상 분할 지도는 SFT-GAN에서 사용한 Outdoor Scene Dataset[15]을 활용한 8 채널의 원핫 행렬(one-hot matrix)을 사용하였다. 또한, FFDNet에서 제안한 잡음 수준 지도를 사용하여 잡음 수준마다 다른 모델을 만든 것이 아닌 하나의 모델로 σ = [0, 75] 에 대해 잡음 제거가 가능하게 하였다. 잡음 영상과 잡음 수준 지도, 영상 분할 지도를 합쳐 잔차 연결(residual connection)을 사용한 네트워크의 입력으로 넣어주었고 잡음이 제거된 색깔 영상을 출력으로 받았다.

Outdoor Scene Dataset[15]과 CBSD68[16]에 대해 실험을 진행한 결과, 제안하는 방법이 DnCNN과 FFDNet보다 성능이 향상되는 것을 확인할 수 있었다. 또한, 다양한 영상 분할 지도를 가지고 실험하여 영상 분할 지도가 네트워크에 주는 영향과 효과를 확인하였다.

2. 본론

2.1. 네트워크 구조

제안하는 네트워크의 구조는 잔차 연결을 활용한 EDSR 구조를 기반으로, 전역 잔차 연결(global residual connection)과 지역 잔차 연결(local residual connection)을 모두 사용하였다. 제안하는 네트워크의 구조는 아래 그림 1과 같다.

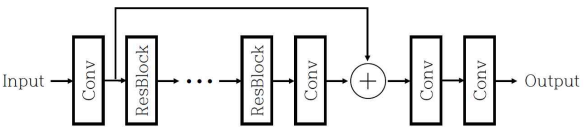


그림 1. 제안하는 네트워크 구조

네트워크의 입력은 잡음이 있는 관측 영상과 잡음 수준 지도, 영상 분할 지도로 이루었다. 관측 영상으로는 색깔 영상을 사용하여 RGB 3 채널로, 잡음 수준 지도는 픽셀당 가산 백색 잡음 σ의 정보를 담은 1 채널로 구성하였다. 영상 분할 지도는 배경(background), 하늘(sky), 건물 (building), 식물(plant), 동물(animal), 풀(grass), 산(mountain), 물 (water)로 총 8가지 항목에 대해 영상 픽셀별로 원핫 인코딩(one-hot encoding)을 하여 8 채널로 구성하였다. 네트워크의 입력으로는 이 세 가지를 모아 12 채널로 구성하여 사용하였다.

처음 12 채널의 입력이 네트워크에 들어가면 96 채널의 Conv 레이어를 통과하고 12 개의 residual block (ResBlock)을 통과한다. ResBlock은 EDSR의 잔차 블록(residual block) 구조를 차용해 96 채널의 Conv-ReLU-Conv 구조로 구성하였고 지역 잔차 연결을 사용하였다. 이후 96 채널의 Conv 레이어를 통과한 후 전역 잔차 연결로 앞선 피처맵(feature map)과 더해주었고, 2개의 96 채널 Conv 레이어를 통과한 후 3 채널 (RGB)의 영상을 출력하였다. 네트워크의 합성곱의 필터 크기는 3×3을 사용하였고, 영상의 크기를 보존하기 위해 매 레이어마다 제로 패딩(zero-padding)하였다.

2.2. 학습 방법

학습 데이터셋으로는 영상 분할 지도가 제공되어있는 Outdoor Scene Dataset[15]의 트레이닝셋 중 9600장을 학습에 사용하였고 300장을 validation에 사용하였다. 영상 패치의 크기는 90x90으로 설정하여 영상마다 90x90 크기의 패치를 추출하였고, 영상을 돌리거나 뒤집는 데이터 증강(data augmentation) 방법을 사용하였다. 배치 크기는 32로 설정하였고, 손실함수는 L1 loss를, 옵티마이저는 Adam optimizer를 사용하였다. 네트워크의 3 채널 출력 영상과 잡음이 없는 원본 이미지 사이 L1 loss를 계산하고 Adam optimizer를 사용해 네트워크를 업데이트하였다. 학습률(learning rate)은 2×10^{-4}로 시작하여 90,000회 학습할 때마다 0.5배씩 줄여주었고, 총 420,000회 학습하였다.

3. 실험 결과

실험 데이터셋으로는 [15]의 테스트셋 300장과 CBSD68[16] 데이터 68장을 사용하였다. 성능 비교는 DnCNN과 FFDNet에 대해 진행하였고, 잡음 σ가 50인 경우에 대해 실험하였다. 세 모델의 성능을 측정하고 비교하는 방법으로는 PSNR을 사용하였고, 이 결과를 아래 표 1에 정리하였다. 제안한 네트워크를 [15]의 데이터셋에 대해 실험을 할 때는 제공된 영상 분할 지도를 사용하였고, [16]의 데이터셋에 대해 실험을 할 때는 영상 분할 지도의 전체 픽셀을 배경이라는 항목으로 설정하여 사용하였다.

표 1. σ=50에서 PSNR 비교

Model ＼ Data	Outdoor Scene	CBSD68
DnCNN	27.51	27.95
FFDNet	27.63	27.97
Ours	**27.97**	**28.16**

그림 2는 출력 영상의 질적 비교를 위하여, [15]와 [16] 각각 하나의 영상에 대해 원본 영상과 잡음 영상, 제안한 네트워크를 통과한 영상, DnCNN을 통과한 영상, FFDNet을 통과한 영상을 나타낸 것이다.

표 1과 그림 2를 통해서 제안한 네트워크가 기존 방법인 DnCNN과 FFDNet보다 성능이 좋은 것을 확인할 수 있다. 표 1을 보면 PSNR은 Outdoor Scene Dataset[15]에 대해 DnCNN, FFDNet 대비 약 0.46 dB, 0.34 dB씩, CBSD68 Dataset[16]에 대해 약 0.2 dB가 향상되었다. [15]에 대해 실험한 경우에는 영상에 알맞은 영상 분할 지도를 사용하였기 때문에 PSNR이 더 많이 향상되었다고 볼 수 있다.

그림 2(a)의 영상들을 비교해보면, 제안한 네트워크의 출력 영상이 DnCNN, FFDNet 출력 영상 대비 뿌옇거나 뭉개지는 느낌이 적었다. 건물의 경우 창문과 벽이 더 깨끗하고 명확하게 표현되었고, 물과 하늘, 나무에서는 디테일함과 질감이 더 잘 표현되는 것을 확인할 수 있다. [15]의 다른 출력 영상들을 확인한 결과, 제안한 모델이 건물 등에서 더 깨끗한 가장자리와 직선을 표현하였고 하늘과 물, 풀 등의 질감을 잘 표현한 것을 볼 수 있었다. 그림 2(b)의 영상들을 보면, 영상 분할 정보를 활용하지 않았기 때문에 [15]의 결과에 비해 영역별로 디테일한 질감을 표현하지는 못했지만 DnCNN, FFDNet 결과 대비 더 깨끗하게 잡음이 제거되는 것을 확인할 수 있었다.

다음으로는, 영상 분할 지도의 영향과 효과를 확인해보았다. 먼저 영상 분할 지도가 출력 영상에 어떠한 영향을 주는지 확인하기 위해서 네트워크 입력의 영상 분할 지도를 180도 돌려서 넣어주었다. σ= 50에 대하여 올바른 영상 분할 지도를 넣어준 입력과 잘못된 영상 분할 지도를 넣어준 입력의 출력 영상을 비교해보았고 이를 그림 3에 나타내었다. 그림 3에서 하늘과 물의 영상 분할 지도가 바뀌어서 들어갔기 때문에 하늘 쪽은 잡음이 제대로 제거가 되지 않았고 물 쪽은 뿌옇게 표현되는 것을 볼 수 있다. 게다가, 건물이 있는 부분에서도 가장자리나 직선이 명확하게 표현되지 않았고 뭉개져 있는 것을 확인할 수 있다.

또한, 영상 분할 지도가 효과가 있는지를 확인하기 위해서 영상에 한 항목의 영상 분할 지도씩을 주어서 결과를 확인해보았다. 예를 들어, 동물이 크게 있는 영상에 입력 영상 분할 지도를 배경으로만, 하늘로만, 동물로만 등 8개 항목에 대해서 주었을 때, 영상 분할 지도 전체가 동물인 경우에서 가장 PSNR이 높게 측정되었다. 대부분 들판으로 이루어진 영상에서는 입력 영상 분할 지도 전체가 풀인 경우에 가장 PSNR이 높았다.

이렇게 영상 분할 지도가 있는 데이터셋 [15]과 없는 데이터셋 [16]에 대해 실험을 한 결과, 네트워크 자체도 학습이 잘 되어 잡음 제거가 잘 되는 것과 영상 분할 지도가 영상의 질감을 표현하는 데 있어서 큰 역할은 한다는 것을 확인하였다. 또한, 영상과 다른 영상 분할 지도를 사용해 실험한 결과, 입력으로 들어간 영상 분할 지도가 학습에 사용되었고 영향과 효과를 준다는 것을 확인하였다.

4. 결론

본 논문에서는 영상 분할 지도를 활용한 영상 잡음 제거 연구를 진행하였다. 제안한 모델은 잔차 연결을 활용한 네트워크에 잡음 수준 지도와 영상 분할 지도를 추가적으로 입력에 넣어준 모델이었다. 이전의 딥러닝 학습 기반 방법들과 비교하였을 때, PSNR과 질적 성능 모두 향상됨을 확인할 수 있었다. 영상 분할 지도가 제대로 활용되지 못했을 때도 성능이 향상되었지만 제대로 활용되었을 때는 영상의 디테일이 더 향상되는 것을 확인하였다. 또한, 맞지 않는 영상 분할 지도를 입력으로 넣어준 경우와 비교하여 영상 분할 지도의 효과와 영향을 알아보았다.

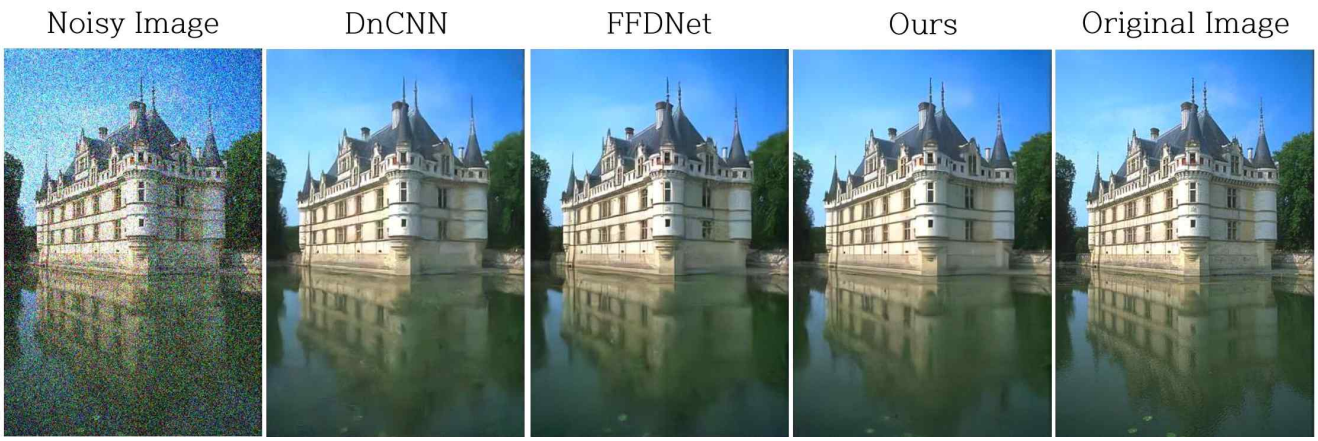
본 논문의 연구에서는 영상 분할 지도가 주어진 데이터셋에 대해 학습과 실험을 진행하였지만, 실제로는 영상 분할 지도가 주어지지 않은 영상이 훨씬 많이 존재한다. 따라서, 영상 분할 지도를 추출하는 네트워크를 사용하거나 이를 같이 학습하는 방향으로 추가 연구를 진행할 수 있을 거라 기대한다.

감사의 글

이 논문은 삼성전자의 지원과 2021년도 과학기술정보통신부 및 정보통신기획평가원의 대학ICT연구센터지원사업의 연구결과로 수행되었음 (IITP-2020-2016 -0-00288).

참고문헌

[1] Dong, Weisheng, et al. "Sparsity-based image denoising via dictionary learning and structural clustering." CVPR 2011. IEEE, 2011.
[2] Mairal, Julien, Michael Elad, and Guillermo Sapiro. "Sparse representation for color image restoration." IEEE Transactions on image processing 17.1 (2007): 53-69.
[3] Malfait, Maurits, and Dirk Roose. "Wavelet-based image denoising using a Markov random field a priori model." IEEE Transactions on image processing 6.4 (1997): 549-565.
[4] Buades, Antoni, Bartomeu Coll, and J-M. Morel. "A non-local algorithm for image denoising." 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05). Vol. 2. IEEE, 2005.
[5] Dabov, Kostadin, et al. "BM3D image denoising with shape-adaptive principal component analysis." SPARS'09 -Signal Processing with Adaptive Sparse Structured Representations. 2009.
[6] Burger, Harold C., Christian J. Schuler, and Stefan Harmeling. "Image denoising: Can plain neural networks compete with BM3D?." 2012 IEEE conference on computer vision and pattern recognition. IEEE, 2012.
[7] Chen, Yunjin, and Thomas Pock. "Trainable nonlinear reaction diffusion: A flexible framework for fast and effective image restoration." IEEE transactions on pattern analysis and machine intelligence 39.6 (2016): 1256-1272.
[8] Zhang, Yulun, et al. "Residual dense network for image super-resolution." Proceedings of the IEEE conference on computer vision and pattern recognition. 2018.
[9] He, Kaiming, et al. "Deep residual learning for image recognition." Proceedings of the IEEE conference on computer vision and pattern recognition. 2016.
[10] Zhang, Kai, et al. "Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising." IEEE transactions on image processing 26.7 (2017): 3142-3155.
[11] Zhang, Kai, Wangmeng Zuo, and Lei Zhang. "FFDNet: Toward a fast and flexible solution for CNN-based image denoising." IEEE Transactions on Image Processing 27.9 (2018): 4608-4622.
[12] Wang, Xintao, et al. "Recovering realistic texture in image super-resolution by deep spatial feature transform." Proceedings of the IEEE conference on computer vision and pattern recognition. 2018.
[13] Liu, Ding, et al. "When image denoising meets high-level vision tasks: A deep learning approach." arXiv preprint arXiv:1706.04284 (2017).
[14] Lim, Bee, et al. "Enhanced deep residual networks for single image super-resolution." Proceedings of the IEEE conference on computer vision and pattern recognition workshops. 2017.
[15] Zhang, Yanfu, Li Ding, and Gaurav Sharma. "Hazerd: an outdoor scene dataset and benchmark for single image dehazing." 2017 IEEE international conference on image processing (ICIP). IEEE, 2017.
[16] Martin, David, et al. "A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics." Proceedings Eighth IEEE International Conference on Computer Vision. ICCV 2001. Vol. 2. IEEE, 2001.



(a) Outdoor Scene Dataset

(b) CBSD68 Dataset

그림 2. Outdoor Scene Dataset과 CBSD68 Dataset에 대한 영상 비교

Original segmentation map　　Wrong segmentation map

그림 3. 영상 분할 지도의 영향 비교