

생성적 적대 신경망을 이용한 영상 잡음 제거

*박구용, 김윤식, 조남익

*서울대학교 전기정보공학부, 뉴미디어통신공동연구소

*benkay@ispl.snu.ac.kr, terryoo@ispl.snu.ac.kr, nicho@snu.ac.kr

Image denoising using Generative Adversarial Network

*Gu Yong Park, Yoonsik Kim, Nam Ik cho

INMC, Department of Electrical and Computer Engineering,
Seoul National University

요 약

영상 잡음 제거 알고리즘은 잡음으로 오염된 영상으로부터 잡음이 제거된 깨끗한 영상을 추정하여 복원하는 연구이다. 기존의 모델 기반 방법의 영상 잡음 제거 알고리즘은 영상을 복원하는 과정에서 최적화 문제를 풀어야 한다는 단점과 매개변수를 직접 선택을 해주어야 한다는 단점을 가진다. 본 논문에서는 딥러닝을 이용한 학습 기반 방법의 영상 잡음 제거 연구를 소개한다. 먼저, 신경망의 구축을 위하여 신경망의 구성 요소는 Instance Normalization 과 컨볼루션 신경망을 이용한 모델을 제안하였고, 여러 연구 분야에서 좋은 성능을 보이는 U-Net 구조를 전체적인 구조로 차용하였다. 신경망의 학습을 위하여 DnCNN 에서 제안한 잡음을 학습하는 잔여 학습 기법을 채택하였고, 기존의 영상 잡음 제거 알고리즘의 단점인 결과 영상이 흐릿해지는 현상을 보완하기 위하여 생성적 적대 신경망 학습 방법을 적용하였다. 본 논문에서 제안한 신경망을 이용한 잡음 제거 영상의 결과가 기존의 연구 방법들 보다 인지적인 측면에서 좋은 결과를 보임을 확인하였다.

1. 서론

초소형 카메라 하드웨어의 발달과 스마트폰의 높은 보급률 덕분에 누구나 쉽게 영상을 획득할 수 있게 되었다. 하지만 카메라 하드웨어의 한계로 인하여 shot noise, read noise 등 잡음(noise)이 생기게 되어 잡음이 없는 깨끗한 영상을 얻는 것은 불가능하다. 잡음 제거 알고리즘(image denoising)은 잡음으로 오염된 영상으로부터 잡음이 없는 깨끗한 영상을 복원하는 것을 목적으로 하며, 수식으로 표현하면 다음과 같다. 잡음으로 오염된 영상을 y , 잡음이 제거된 깨끗한 영상을 x , 잡음을 n 이라고 하면 영상 잡음 제거 알고리즘을 $y = x + n$ 라는 수식으로 모델링할 수 있다. 카메라를 통해 획득한 잡음으로 오염된 영상 y 로부터 잡음이 없는 깨끗한 영상 x 를 추정하는 것은 ill-posed inverse problem 이다.

영상 잡음 제거 알고리즘은 카메라의 발전과 함께 수십년 전부터 연구되어 온 분야이다. 영상 잡음 제거 알고리즘은 방법론에 따라서 모델 기반 방법과 학습 기반 방법으로 나뉜다. 먼저 모델 기반 방법은 Bayesian 모델을 이용해서 표현된 최적화 문제에서 해답을 찾아내는 방법으로 영상 잡음 제거 알고리즘을 접근한다. 베이지안 모델을 이용한 방법에서는 사후 확률(posterior)을 최대로 하는 x 를 찾는 것이 원하는 해답을 얻는 것이다. 모델 기반 방법의 영상 잡음 제거 알고리즘은 잡음 제거에 좋은 성능을 보여주지만, 테스트 단계에서 복잡한 최적화 문제를 풀어야 한다는 단점을 가진다. 또한 최적화 문제들이 대부분 non-convex 문제라서, 매개변수를 수동적으로 선택해야 한다는 단점을 가진다[1,2,3,4,5].

학습 기반 방법은 잡음으로 오염된 영상 y 로부터 잡음이 없는 깨끗한 영상 x 로의 mapping 함수를 학습하는 방법이다. 잡음으로 오염된 정보를 정확한 정보로 추론하는 mapping 함수를 찾는 것은 어려운 일이기 때문에 연구가 활발히 진행되지 않았다. 최근 들어 컨볼루션 신경망(Convolutional Neural Network)을 이용한 딥러닝(deep learning) 학습 기반 방법이 많은 컴퓨터 비전 분야에서 좋은 성능을 내고있고, 영상 잡음 제거 분야에서 또한 효과적임이 증명되었다[6,7,8].

본 논문에서는 딥러닝을 이용한 학습 기반 방법을 이용한 영상 잡음 제거 연구를 진행하였다. 논문의 실험에서 사용한 신경망의 구조는 최근에 많은 분야에서 좋은 성능을 나타내고 있는 U-Net 구조[9]를 기초로 하였고, 신경망 모델의 구성요소는 Instance Normalization[10]과 컨볼루션 신경망을 이용하여 구성하였다. 신경망의 학습 기법은 기존의 영상 잡음 제거 알고리즘인 DnCNN[8]에서 사용한 잔여 학습 방법(residual learning)과 생성적 적대 신경망(generative adversarial network)[11] 학습 기법을 이용하였다. 학습 데이터셋은 DIV2K 데이터[12]를 이용하였고, RGB 색공간으로 표현된 학습 데이터 세트를 만들어 실험을 수행하였다.

2. 본론

2.1. 학습 기반 방법의 영상 잡음 제거

영상 잡음 제거 알고리즘 분야에서 딥러닝을 이용한 학습

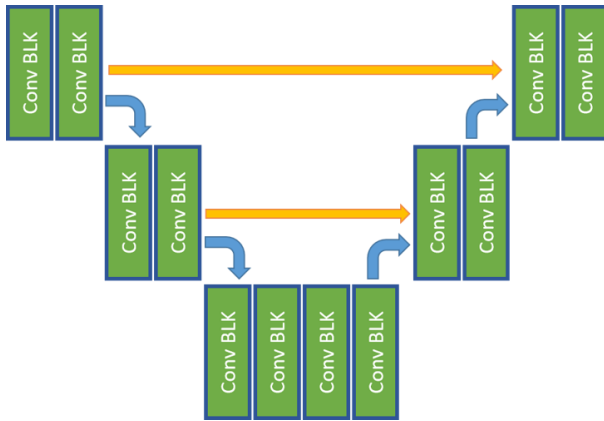


Fig 1. 실험에서 사용된 U-Net 구조

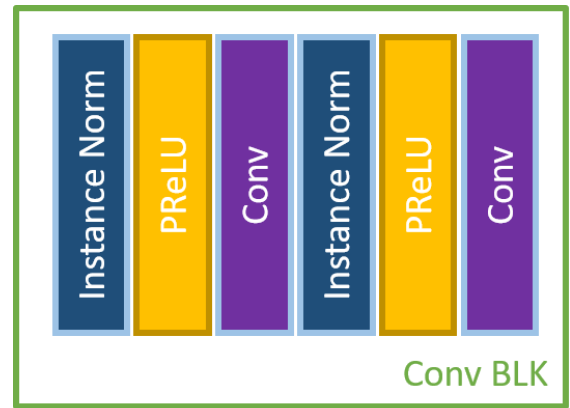


Fig 2. U-Net 구조의 구성요소 Conv BLK 의 구조

방법이 기존의 모델 기반 방법보다 뛰어난 성능을 보여주고 있다[7,8]. 기존의 모델 기반 방법들은 영상의 사전 확률 정보(prior knowledge)를 이용하여 잡음이 없는 깨끗한 영상을 복원하였다. 또한 인간의 지식을 기반으로 한 handcrafted prior 를 이용하여 특징(feature)를 추출하였고, 이를 이용하여 연구를 진행하였다[1,2,3,4,5]. 이에 반해 딥러닝을 이용한 학습 기반 방법은 특징을 직접 추출하고 이를 사용하기 때문에, 인간의 지식을 기반으로 하여 추출한 특징에 비해 유연한 정보를 이용할 수 있다는 장점을 가진다. 또한 딥러닝을 이용한 학습 방법은 기존의 모델 기반 방법과는 달리 여러 장의 영상을 동시에 사용할 수 있기 때문에 더욱 많은 정보를 이용할 수 있다는 장점을 가진다.

본 논문에서는 기존의 모델 기반 방법보다 좋은 성능을 보여주는 딥러닝을 이용한 학습 기반 방법을 채택하였다. 영상 잡음 제거 연구에서 좋은 성능을 보이는 DnCNN[8]에서 영상을 이용하여 학습을 진행하는 것보다 잡음을 이용하여 학습을 진행하는 잔여 학습이 좋은 성능을 보인다는 것을 보여주었다. 본 논문에서도 잔여 학습 방법을 채택하였고, 학습 데이터로는 영상 복구(image restoration) 알고리즘 분야에서 널리 쓰이고 있는 DIV2K 데이터[12]를 사용하였다.

2.1. 학습 기반 방법의 영상 잡음 제거

생성적 적대 신경망은 한 개의 신경망을 학습하기 위하여 또 다른 하나의 신경망을 이용하여 학습을 진행하는 신경망으로 기존의 컨볼루션 신경망과는 다른 방법으로 학습을 진행한다. generator 라고 불리는 신경망을 학습하기 위하여 discriminator 라고 불리는 신경망을 도입하였다. Generator 는 discriminator 를 속이는 것에 목표를 두고 학습을 행하고, discriminator 는 generator 가 만든 데이터와 원본 데이터를 구별하는 것에 목표를 두고 학습을 진행한다. 두 개의 신경망이 경쟁하며 학습을 진행하게 되고, 학습이 성공적으로 끝나게 되면 generator 는 원본과 비슷한 데이터를 생성하게 된다. 생성적 적대 신경망의 학습 기법을 이용하면 복잡한 분포를 가진 데이터에 대해서도 신경망의 학습을 진행할 수 있다[11]. 이런 강력한 장점을 바탕으로 생성적 적대 신경망은 여러 연구 분야에서 활발히 적용되고 있다[13,14,15,16,17].

하지만 생성적 적대 신경망의 학습 방법은 기존의 컨볼루션 신경망의 학습 방법에 비해 수렴이 굉장히 어렵다는 단점을 가진다. 두 개의 신경망을 경쟁시키는 학습 방법이기 때문에 학습과정이 매우 불안정하여 수렴을 시키기 어렵다. 이와 같은 단점을 보완하기 위하여 DCGAN[13], WGAN[18], WGAN-GP[19], RaGAN[20] 등과 같은 여러가지 학습 방법 등이 소개되었다.

본 논문에서는 기존의 영상 잡음 제거 알고리즘의 단점으로 꼽히는 결과 영상이 흐릿해지는 현상을 완화하기 위하여 생성적 적대 신경망 학습 방법 적용하였다. WGAN-GP[19] 학습 방법을 채택하여 실험을 진행하였다. 실험에 사용된 discriminator 의 구조는 Fig 3 과 같다.

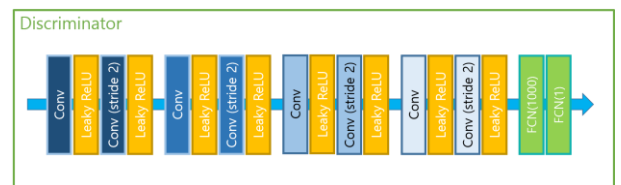


Fig 3. Discriminator 의 구조

3. 실험

3.1. 실험 설정

본 논문의 실험에서 사용한 신경망의 구조는 U-Net 구조 [9]이고 Fig 1 과 같다. 기존의 U-Net 신경망 구조의 구성요소는 컨볼루션 신경망과 ReLU 활성화 함수로 구성된다. 본 논문에서는 기존의 U-Net 구조를 전체적인 신경망의 구조로 차용하였지만 신경망의 구성요소(Conv BLK)는 Fig 2 와 같고 Instance Normalization[10]과 Parametric ReLU 활성화 함수(PReLU)를 이용하여 구성하였다.

DnCNN[8]에서 사용한 잡음을 학습하는 잔여 학습을 하기 위하여 신경망의 입력 부분과 출력 부분에 skip-connection 을 이용하였다. Skip-connection 은 신경망의 학습속도를 빠르게 해주며, 신경망의 깊이를 더욱 깊게 만들 수 있게 해주어 신경

망의 표현능력을 향상시켜준다[21].

실험에서 사용한 학습 데이터는 DIV2K 데이터로 총 800 장의 학습 데이터와 100 장의 검증 데이터로 이루어져 있다. 잡음을 추가하기 위하여 가우시안 잡음(Gaussian noise)을 추가 해주었고, blind 환경의 실험을 진행하기 위하여 잡음의 범위는 [5,35]로 설정하였다. 실험에 사용한 손실함수(loss function)는 식 1 과 식 2 의 합과 같다. 식 1 에서 x_i 는 잡음이 없는 깨끗한 영상이고 y_i 는 깨끗한 영상에 잡음을 추가하여 만든 잡음 영상이다.

$$\mathcal{L}(\theta) = \frac{1}{2N} \sum_{i=1}^{N1} \|F(y_i; \theta) - x_i\|_2^2 \quad \text{식 1}$$

생성적 적대 신경망의 학습 방법을 적용하기 위하여 WGAN-GP[19] 학습 방법을 채택하였고, 실험에 사용된 손실함수는 식 2 와 같다.

$$\mathcal{L}_{GAN} = E_{x \sim p_d} [D(x)] - E_{x \sim p_r} [D(x)] + \lambda E_{x \sim p_r} [(\|\nabla_x D(x)\|_2 - 1)^2] \quad \text{식 2}$$

RMSprop 알고리즘[22]을 이용하여 손실함수의 최적화를 진행하였고, 학습률(learning rate)은 10^{-4} 로 실험을 진행하였다.

3.2. 실험 결과 및 분석

영상 잡음 제거 연구에서는 실험간 결과 비교를 위하여 최대 신호 대 잡음비(PSNR)를 측정 기준으로 하여 결과 비교를 진행한다. 기존의 모델 기반 방법인 CBM3D[5]와 학습기반 방법인 DnCNN[8]와 FFDNET[23]을 이용하여 결과 비교를 진행하였고, 실험 결과는 표 1 과 같다.

Table 1. 잡음 수준 25 에서의 실험 결과

신경망 구조	최대 신호 대 잡음비
CBM3D [5]	30.71 dB
C-DnCNN [8]	31.23 dB
FFDNET [23]	31.21 dB
Proposed network	29.32 dB

실험에 사용된 Instance Normalization 과 컨볼루션 신경망을 모델의 구성요소로 하는 U-Net 구조는 29.32 dB 의 결과를 보여주었다. 이는 모델 기반 방법의 CBM3D 의 결과 30.71 dB 와 학습 기반 방법인 DnCNN 의 결과 31.23 dB 과 FFDNET 의 결과 31.21 dB 보다 낮은 결과를 나타낸다. 이러한 결과를 나타내는 원인으로는 2 가지를 꼽을 수 있다. 먼저, 제안한 모델은 가우시안 잡음 수준이 [5,35] 수준으로 설정하여 blind 환경의 실험을 진행하였다. 반면, 실험 결과 비교에 사용된 알고리즘들은 테스트에 사용한 가우시안 잡음과 일치하는 신경망 모델을 사용하였다. CBM3D 의 경우 모델 기반 방법이 기 때문에 직접적으로 가우시안 잡음 정보를 주었고, DnCNN 과 FFDNET 의 경우, 가우시안 잡음 수준 25 로 학습된 모델을 이용하였다.

두 번째로는 제안한 모델은 생성적 적대 신경망의 학습 기법을 선택하여 학습을 진행하였다. 생성적 적대 신경망 학습의 경우 인지적 측면(perceptual quality)에서 좋은 결과를 나타내지만, 최대 신호 대 잡음비 측면에서 좋지 않은 결과를 나타낸다는 것이 알려져 있다. 실험 결과 비교에 사용된 학습 기반 방법의 알고리즘인 DnCNN 과 FFDNET 은 기본적인 신경망 학

습을 채택한 모델이다. 이러한 이유로 인해 실험 결과 비교에 사용된 알고리즘들이 제안한 신경망보다 최대 신호 대 잡음비 측면에서 더 좋은 성능을 보이고 있다.

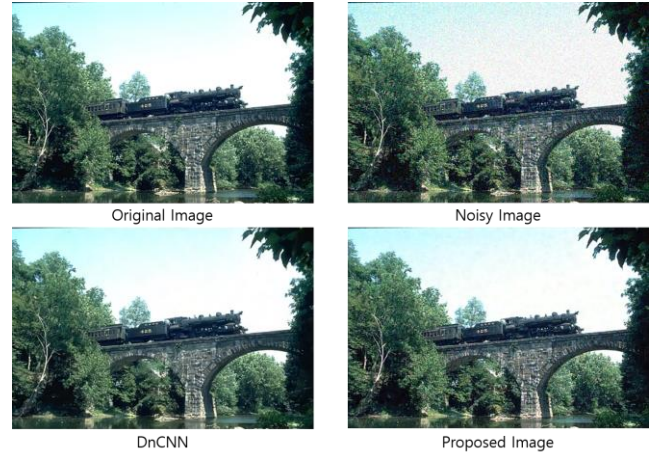


Fig 4. 잡음 수준 25 에서의 실험 결과 영상

Fig 4 를 보면 기본적인 신경망 학습을 채택한 모델인 DnCNN 의 결과 영상이 제안한 모델의 결과 영상보다 선명한 영상임을 알 수 있다. 또한, 기존의 단점으로 꼽히던 결과 영상이 흐릿해지는 현상 또한 발생했음을 알 수 있다. 제안한 모델의 결과영상의 경우 DnCNN 보다 선명하진 않지만 흐릿한 영역이 많이 줄어들었음을 알 수 있다. 특히, 왼쪽에 위치한 나무영역과 가운데에 위치한 다리 영역과 같이 high frequency 성분이 있는 영역에서 뚜렷한 차이를 보인다. 생성적 적대 신경망 학습 기법을 통해서 인지적인 측면에서 좋은 결과를 보임을 확인할 수 있었다. 반면 하늘과 같이 smooth 한 영역에서는 잡음이 완전히 제거가 되지 않았음을 알 수 있다. 이는 blind 환경 학습으로 인해 잡음을 제대로 예측을 못 하여 잡음 제거가 제대로 이루어지지 않음을 알 수 있다. 이 때문에 최대 신호 및 잡음비에서 좋지 않은 결과를 나타냄을 알 수 있었다.

4. 결론 및 향후 연구 방향

본 논문에서는 생성적 적대 신경망을 이용한 영상 잡음 제거 연구를 진행하였다. 컴퓨터 비전의 많은 분야에서 좋은 성능을 보여주고 있는 U-Net 구조의 신경망 모델을 차용하고, 신경망의 구성요소를 Instance Normalization 과 컨볼루션 신경망을 이용하여 구성하였다. 기존의 학습 기반 방법의 영상 잡음 제거 알고리즘은 DnCNN[8]에서 사용한 잡음을 학습하는 잔여 학습 방법과 생성적 적대 신경망 학습 방법을 이용하여 실험을 진행하였다. 실험 결과 생성적 적대 신경망 학습 기법을 통하여 영상 잡음 제거 알고리즘의 단점으로 꼽히던 결과영상의 흐릿해지는 현상을 완화할 수 있었다. 하지만 smooth 한 영역에서의 잡음 예측이 제대로 되지 않아서 잡음 제거가 완벽하게 이루어지지 않음을 알 수 있다. 이를 해결하기 위하여 신경망 모델의 구조를 바꾸는 연구, 잡음 예측 신경망을 추가하는 연구, 생성적 적대 신경망의 학습 방법을 효과적으로 하기위한 방법 등을 향후 연구 방향으로 설정할 것이다.

감사의 말

본 연구는 삼성전자의 지원으로 이루어진 것입니다

참고 문헌

- [1] W. Dong, X. Li, L. Zhang, and G. Shi, "Sparsity-based image denoising via dictionary learning and structural clustering," in Proc. of the IEEE CVPR, 2011, pp. 457–464.
- [2] J. Mairal, M. Elad, and G. Sapiro, "Sparse representation for color image restoration," IEEE Transactions on image processing, vol. 17, no. 1, pp.53–69, 2008.
- [3] M Malfait, and D Roose. "Wavelet-based image denoising using a Markov random field a priori model." IEEE Transactions on image processing 6.4 (1997): 549–565.
- [4] A Buades, B Coll, and J–M. Morel. "A non-local algorithm for image denoising." Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on. Vol. 2. IEEE, 2005.
- [5] K Dabov, A Foi, V Katkovnik, and K Egiazarian. "BM3D image denoising with shape-adaptive principal component analysis." SPARS'09–Signal Processing with Adaptive Sparse Structured Representations. 2009.
- [6] HC. Burger, C. Schuler, and S. Harmeling. "Image denoising: Can plain neural networks compete with BM3D?." Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on. IEEE, 2012.
- [7] Y. Chen, and T. Pock. "Trainable nonlinear reaction diffusion: A flexible framework for fast and effective image restoration." IEEE transactions on pattern analysis and machine intelligence 39.6 (2017): 1256–1272.
- [8] K. Zhang, W. Zuo, Y. Chen, and D. Meng. "Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising." IEEE Transactions on Image Processing 26.7 (2017): 3142–3155.
- [9] Ronneberger, Olaf, Philipp Fischer, and Thomas Brox. "U-net: Convolutional networks for biomedical image segmentation." International Conference on Medical image computing and computer-assisted intervention. Springer, Cham, 2015.
- [10] Ulyanov, Dmitry, Andrea Vedaldi, and Victor Lempitsky. "Instance normalization: The missing ingredient for fast stylization." arXiv preprint arXiv:1607.08022 (2016).
- [11] I. Goodfellow, J. Pouget–Abadie, and M. Mirza. "Generative adversarial nets." Advances in neural information processing systems. 2014.
- [12] <https://data.vision.ee.ethz.ch/cvl/DIV2K/>
- [13] A. Radford, L. Metz, and S. Chintala. "Unsupervised representation learning with deep convolutional generative adversarial networks." arXiv preprint arXiv:1511.06434 (2015).
- [14] M. Mirza, and S. Osindero. "Conditional generative adversarial nets." arXiv preprint arXiv:1411.1784 (2014).
- [15] C. Ledig, L. Theis, F. Huszár, and J. Caballero. "Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network." CVPR. Vol. 2. No. 3. 2017.
- [16] MSM. Sajjadi, B. Schölkopf, and M. Hirsch. "Enhancenet: Single image super-resolution through automated texture synthesis." Computer Vision (ICCV), 2017 IEEE International Conference on. IEEE, 2017.
- [17] X. Wang, K. Yu, C. Dong, and CC. Loy. "Recovering realistic texture in image super-resolution by deep spatial feature transform." arXiv preprint arXiv:1804.02815 (2018).
- [18] M. Arjovsky, S. Chintala, and L. Bottou. "Wasserstein gan." arXiv preprint arXiv:1701.07875 (2017).
- [19] I. Gulrajani, F. Ahmed, and M. Arjovsky. "Improved training of wasserstein gans." Advances in Neural Information Processing Systems. 2017.
- [20] Jolicœur–Martineau, Alexia. "The relativistic discriminator: a key element missing from standard GAN." arXiv preprint arXiv:1807.00734 (2018).
- [21] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In CVPR, 2016.
- [22] T. Tieleman and G. Hinton. Lecture 6.5|RmsProp: Divide the gradient by a running average of its recent magnitude. COURSERA: Neural Networks for Machine Learning, 2012.
- [23] 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.