

특징 추출 블록과 이중 경로 합성곱 신경망을 이용한 영상 내 잡음 제거

김재엽[○], 이경준, 정제창

한양대학교 융합전자공학과

athurk94111@gmail.com, kjlee888@naver.com, jjeong@hanyang.ac.kr

요 약

본 논문에서는 합성곱 신경망(Convolutional Neural Network, CNN) 중 특징 추출 블록(Features Block)과 이중 경로 네트워크(Dual Network)를 이용한 영상 내 잡음 제거(Image Denoising) 신경망을 제안한다. 제안하는 네트워크는 입력 영상의 특징을 추출하는 특징 추출 블록과, 입력과 추출된 특징을 결합하여 이중 경로 네트워크를 거쳐 두 결과를 공유하는 구조를 가진다. 실험은 잡음 레벨(Noise Level)이 15, 25, 50인 세 가지 레벨의 가산 백색 가우스 잡음(Additive White Gaussian Noise, AWGN)이 추가된 영상을 네트워크에 입력으로 사용해 잡음을 제거하는 것으로 진행된다. 실험 결과 제안하는 네트워크는 객관적 평가인 PSNR(Peak Signal-to-Noise Ratio)에서 기존 알고리즘 및 네트워크에 비해 높은 결과 수치를 보였고, 주관적 평가인 네트워크 별 화질 비교에서도 기존 알고리즘 및 네트워크보다 영상 내 디테일을 선명하게 보존하는 결과를 보인다.

1. 서론

영상 내 잡음 제거는 영상 내에 존재하는 잡음(Noise)을 제거해 영상의 화질을 개선하는 기술을 말한다. 영상 내 잡음 제거는 기존의 필터 적용, 잡음 제거 알고리즘과 같은 기본적인 영상처리에서부터 최근에는 딥 러닝(Deep Learning)을 이용한 영상 내 잡음 제거까지 꾸준히 연구되고 있다. 또한 영상 내 잡음 제거 기법은 객체 검출 [1], 이미지 분류화 [2] 등 다양한 영상처리 분야에서 사전에 잡음을 제거하는 전처리 단계로 많이 사용되고 있다.

딥 러닝을 이용해 영상 내 잡음 제거 연구가 이루어지기 전에는 필터나 알고리즘을 통한 잡음 제거 연구가 이루어졌다. 잡음 제거 알고리즘의 대표적인 예로 BM3D[3]가 있다. BM3D는 영상 내에 설정된 참조 블록과 값이 유사한 블록을 찾아 블록을 3차원의 형태로 쌓고 임계값에 미치지 못하는 값을 제거한 후 역변환을 실행하여 잡음을 제거하는 알고리즘이다. 이와 같은 알고리즘에 이어 최근에는 딥 러닝을 이용한 영상 내 잡음 제거가 활발하게 이루어지고 있다. DnCNN[4]은 딥 러닝을 이용한 단일 네트워크로 합성곱, ReLU[5], 배치 정규화(Batch Normalization)[6]로 이루어진 17개의 계층으로 구성되어 있다. DnCNN은 딥 러닝을 이용하여 기존의 알고리즘을 이용한 잡음 제거 보다 우수한 잡음을 제거 성능을 보인다. FFDNet[7]은 잡음이 있는 영상과 잡음 레벨(Noise Level) 정보를 함께 입력으로 사용하며 DnCNN의 네트워크를 그대로 이용하였다.

본 논문에서는 앞서 언급한 DnCNN, FFDNet과 같이 단일 네트워크로 이루어진 합성곱 신경망과 달리 다양한 특징을 학습하기 위해 서로 다른 네트

워크의 결과를 공유하는 이중 경로 기반의 네트워크를 제안한다. 또한 특징 추출 블록을 이용하여 잡음이 있는 영상에 대해 특징을 추출한 후, 이를 잡음이 있는 영상과 결합해 이중 경로 네트워크의 입력으로 이용한다.

본 논문의 구성은 다음과 같다. 2장에서는 제안하는 네트워크에서 사용된 방법인 확장된 합성곱(Dilated Convolution) [8]과 잔차 학습(Residual Learning) [9]에 대하여 설명하도록 한다. 3장에서는 2장에서 설명한 방법을 기반으로 구성된 제안하는 네트워크의 전체적인 구조와 네트워크의 핵심적인 내용을 설명한다. 4장에서는 기존 네트워크와 제안한 네트워크의 실험 결과를 비교 분석하고 5장에서 결론을 맺는다.

2. 이론적 배경

2.1 확장된 합성곱(Dilated Convolution)

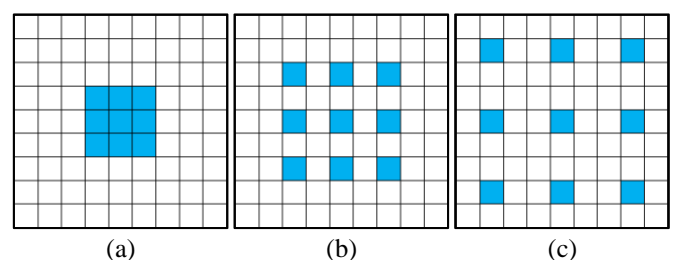


그림 1. 확장 계수에 따른 확장된 합성곱의 변화

학습 시 네트워크에서 필터의 크기가 3인 합성곱 연산은 그림 1(a)와 같이 3x3 필터로 영상 내에

서 연산이 이루어지며 이를 수용 영역(Receptive Field)이라고 한다. 확장된 합성곱은 확장 계수(Dilation Rate)를 증가시켜 넓은 범위의 특징을 학습할 수 있고 넓은 범위의 잡음 성질을 검출할 수 있다. 그림 1(b)와 그림 1(c)는 각각 확장 계수가 2와 3일 때의 수용영역이다. 확장 계수가 증가하면 넓은 범위의 수용 영역을 가진다.

2.2 잔차 학습(Residual Learning)

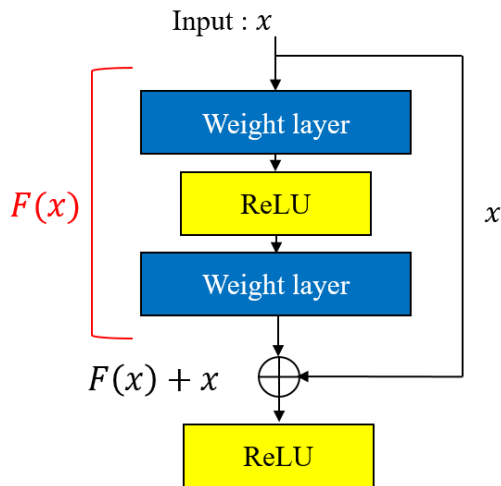


그림 2. 잔차 학습의 기본적인 구조

잔차 학습은 기존에 깊게 쌓인 네트워크의 출력 값에 입력 값을 더하는 스킵 연결(Skip Connection)을 추가하는 구조로 이루어진다. 네트워크가 깊어지면 학습의 성능이 저하되는 gradient vanishing과 exploding 현상이 발생한다. 이러한 문제를 해결하기 위해 그림 2의 입력 값 x 를 합성곱 신경망 $F(x)$ 에 더해주는 구조로 스킵 연결을 추가한다. 이 때 입력 값 x 가 더해진 합성곱 신경망 출력 값을 $H(x)$ 라고 한다.

$$H(x) = F(x) + x \quad (1)$$

$H(x)$ 는 식 (1)과 같이 정의할 수 있으며 합성곱 신경망 $F(x)$ 는 다음과 같다.

$$F(x) = H(x) - x \quad (2)$$

결국, 스킵 연결을 추가함에 따라 합성곱 신경망 $F(x)$ 는 식(2)와 같이 입력 값과 출력 값의 차이를 학습하는 구조로 구성된다.

결과적으로 잔차 학습은 단순히 출력 값에 입력 값을 더해주는 구조로 연산량이 증가하지 않고 입력 값과 출력 값의 차이에 대해서 학습한다. 역전파 과정에서 입력 값 x 가 더해진 출력 값 $H(x)$ 는 미분 시 최소 1 이상의 값으로 최소한의 기울기를 가진다. 때문에 깊게 쌓인 합성곱 신경망에서 안정적으로 학습하여 학습 성능이 저하되는 문제점을 해결할 수 있다.

3. 제안하는 네트워크

제안하는 네트워크에서는 특징 추출 블록과 이중 경로 네트워크를 이용하여 잡음을 제거하는 전체 네트워크에 대해 설명한다. 그림 3의 전체 네트워크 구조를 바탕으로 3.1은 네트워크의 전체적 구조에 대해 간략히 설명하고 3.2는 특징 추출 블록에 대해 설명하며 3.3은 이중 경로 네트워크를 설명한다.

3.1 네트워크 구조

제안하는 네트워크는 특징을 추출하는 특징 추출 블록과 이중 경로 네트워크를 통해 학습하는 두 단계로 나뉜다. 네트워크는 먼저 그림 3의 특징 추출 블록(Features Block)을 이용해 입력 영상의 특징을 추출한다. 다음으로 특징 추출 블록에 의해 추출된 특징 영상은 잡음이 포함된 영상과 연결(Concatenation)되어 이중 경로 네트워크의 입력으로 사용된다. 이중 경로 네트워크는 그림 3의 첫 번째 계층(Layer1)과 두 번째 계층(Layer2)과 같이 구성된다. 이중 경로 네트워크를 거친 후 첫 번째 계층(Layer1)과 두 번째 계층(Layer2) 결과에 네트워크의 입력영상을 각각 더해준 두 결과와 특징 추출 블록의 결과 영상을 연결한다. 연결한 세 가지 영상을 두 번의 3×3 합성곱, 배치 정규화, ReLU로 학습하고 전체 잔차 학습(Global Residual Learning)을 통해 최종적으로 잡음이 제거된 영상을 추출한다.

3.2 특징 추출 블록

제안하는 네트워크에서 특징 추출 블록은 그림 4(a)와 같은 잡음이 있는 입력 영상에서 영상의 특징을 추출한다. 그림 3의 특징 추출 블록(Feature Block)은 3×3 합성곱과 ReLU가 순서대로 네 차례 구성된 후 1개의 3×3 합성곱으로 구성된다. 그림 4(b)는 잡음이 있는 입력 영상이 특징 추출 블록을 거친 후 출력되는 결과 영상이다. 특징 추출 블록을 거쳐 특징이 추출된 영상은 잡음이 있는 입력 영상에 비해 평탄한 영역이 유사한 값으로 매핑되며 경계선을 보존하는 결과를 보인다. 특징 추출 블록은 영상의 전체적인 특징을 추출하여 입력 영상과 연결됨으로써 학습할 정보를 증가시킨다.

3.3 이중 경로 네트워크

Pan et al. [10]은 낮은 레벨 비전(Low-Level Vision)에서 디테일을 보완하기 위해 DualCNN을 제안하였다. 제안하는 네트워크는 DualCNN과 같이 영상 내 디테일을 보존하기 위해 이중 경로 네트워크를 사용한다. 두 개의 계층(Layer)을 각각 따로 학습한 후 학습된 결과를 공유하고 입력 영상을 연결하여 학습 결과를 도출하는 구조로 이루어진다. 첫 번째 계층(Layer1)은 기존 영상 내 잡음 제거 네트워크인

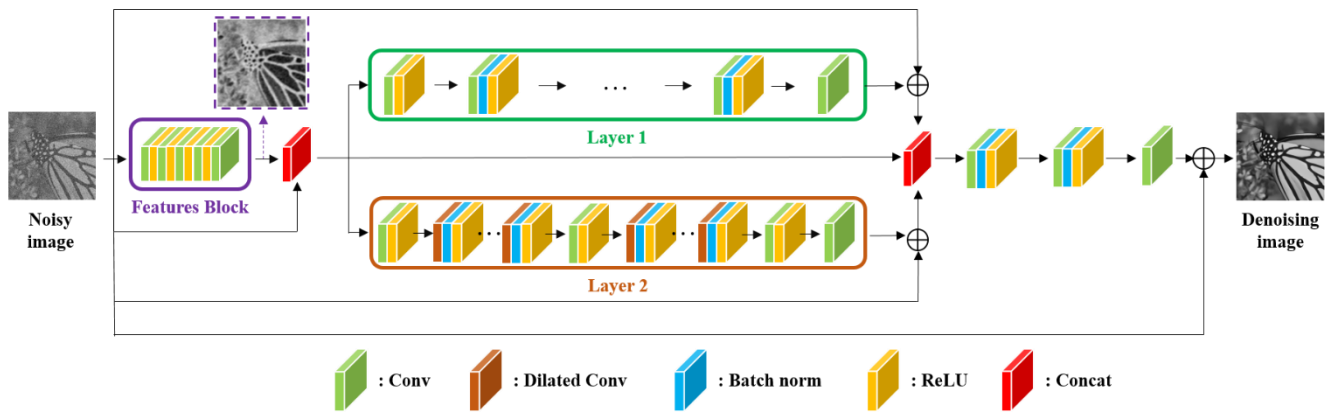


그림 3. 제안하는 특징 블록과 이중 경로 네트워크를 이용한 전체 네트워크 구조

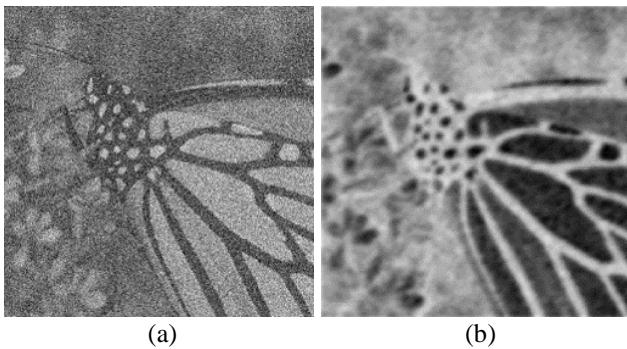


그림 4. 입력 영상과 특징 추출 블록의 결과 영상

17개의 계층으로 구성된 DnCNN을 사용하고, 두 번째 계층(Layer2)은 확장된 합성곱을 이용하여 넓은 수용 영역을 바탕으로 학습하도록 구성한다. 두 번째 계층(Layer2)은 3×3 합성곱과 ReLU의 결합구조와 확장된 합성곱과 배치 정규화 및 ReLU의 결합구조, 하나의 3×3 합성곱으로 총 세 가지 형식으로 구성한다. 이때 확장 계수가 2인 확장된 합성곱을 사용하고 첫 번째 계층(Layer1)과 같은 깊이의 네트워크를 구성하기 위해 17개의 계층으로 구성한다.

그림 5(a)는 DnCNN을 이용한 첫 번째 계층(Layer1)의 결과이며 전반적인 잡음과 영상 내 객체 위주로 특징을 추출한다. 반면, 그림 5(b)는 확장된 합성곱을 이용한 두 번째 계층(Layer2)의 결과이며 세세한 잡음과 영상 내에서 지역적인 특징을 추출한다. 이후 두 계층의 결과를 연결하여 공유함으로써 서로 다른 관점에서 잡음 정보와 특징을 파악한다. 결과적으로 동일한 입력 영상에 대해 서로 다른 결과를 추출하여 특징을 공유함으로써 다양한 특징을 학습할 수 있다.

4. 실험 결과 및 분석

제안하는 네트워크는 DIV2K 데이터 셋을 이용하여 학습을 진행하였으며 학습 시 96×96 크기의 패치(patch)로 영상을 분할하여 네트워크의 입력으로 사용하였다. 학습된 네트워크의 성능을 평가하는 Set12와 BSD68[11] 데이터 셋을 이용하여 객관적

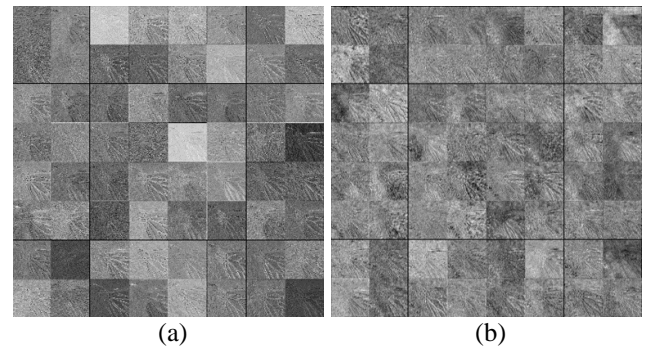


그림 5. 두 계층의 학습 결과 피쳐 맵(Feature map)

수치 및 주관적 화질을 비교 평가한다. 실험은 파이토치(PyTorch)를 이용하고 CPU는 i7-4790k, GPU는 RTX 2080Ti를 사용한다. 학습 시 배치의 크기는 16으로 설정하고 학습 횟수는 60 에폭(epoch)으로 학습한다. 학습률 (Learning Rate)은 0.001을 초기값으로 하고 30에폭 이후에는 0.0001로 학습률을 감소시킨다. 각 네트워크의 성능을 확인하기 위하여 객관적 지표인 PSNR을 사용하여 비교하고 주관적 비교는 확대한 영상을 바탕으로 화질과 디테일을 비교한다.

표 1은 입력 영상에 대한 4개의 기존 알고리즘 및 네트워크와 제안하는 네트워크의 PSNR 결과이다. 제안하는 네트워크는 잡음 레벨이 15, 25, 50 일 때 모두 기존 알고리즘 및 네트워크보다 향상된 PSNR 결과를 보인다. BSD68 데이터 셋을 기준으로 잡음 레벨 15일 때 가장 높은 수치를 보인 DnCNN보다 0.06이 향상되고 잡음 레벨 25일 때 가장 높은 수치를 보인 FFDNet보다 0.11이 향상되며 잡음 레벨 50일 때 가장 높은 수치를 보인 FFDNet보다 0.06이 향상된다. Set12 데이터 셋을 기준으로 잡음 레벨 15일 때 가장 높은 수치를 보인 DnCNN보다 0.17이 향상되고 잡음 레벨 25일 때 가장 높은 수치를 보인 FFDNet보다 0.18이 향상되며 잡음 레벨 50일 때 가장 높은 수치를 보인 WDnCNN[12]보다 0.13이 향상된다. 제안하는 네트워크를 이용한 잡음 제거 시 잡음 레벨이 25일 때 평가하는 모든 데이터 셋에서 PSNR이 기존 알고리즘 및 네트워크보다 크게 향상된다.

표 1. 기존 영상 잡음 제거 네트워크와 제안한 네트워크의 PSNR(dB) 비교

Methods	BSD68			Set12		
	Noise Level σ			Noise Level σ		
	15	25	50	15	25	50
Noisy	24.60	20.17	14.15	24.60	20.18	14.15
BM3D	31.13	27.80	25.68	32.40	30.00	26.74
DnCNN	<u>31.72</u>	29.20	26.22	<u>32.82</u>	30.38	27.14
FFDNet	31.63	<u>29.21</u>	<u>26.31</u>	32.76	<u>30.47</u>	27.32
WDnCNN	31.56	29.16	26.30	32.71	30.44	<u>27.37</u>
Proposed	31.78	29.32	26.37	32.98	30.65	27.50

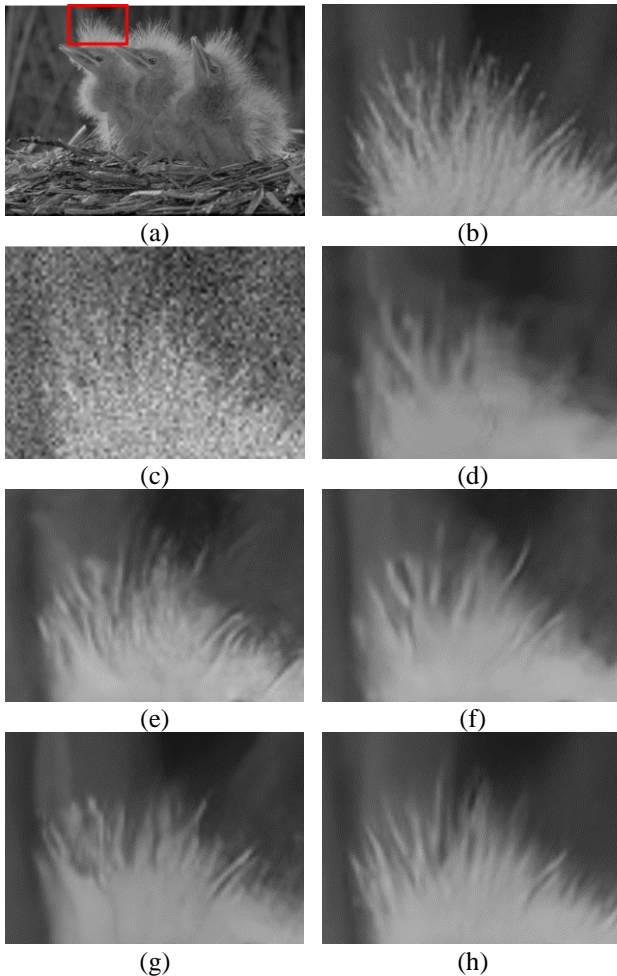


그림 6. BSD68 데이터 셋의 잡음이 제거된 확대 영상에 대한 주관적 화질 비교,
(a) 확대 영역 표시 영상 (b) 원본 영상 (c) 잡음 영상 ($\sigma = 25$) (d) BM3D (e) DnCNN (f) FFDNet
(g) WDnCNN (h) 제안하는 네트워크

그림 6은 BSD68 데이터셋 중 하나로 잡음이 제거된 영상을 주관적 비교를 위해 확대한 영상이다. 제안하는 네트워크인 그림 6(h)는 원본 영상의 확대한 부분인 그림 6(b)와 비교했을 때 다른 네트워크보다 동물 털의 질감을 잘 보존하며 복원하는 결과를 보인다. 또한 제안하는 네트워크는 기존 알고리즘 및 네트워크보다 잡음 제거 시 주변에 열화 현상이 존재하지 않아 영상을 선명하게 복원한다.

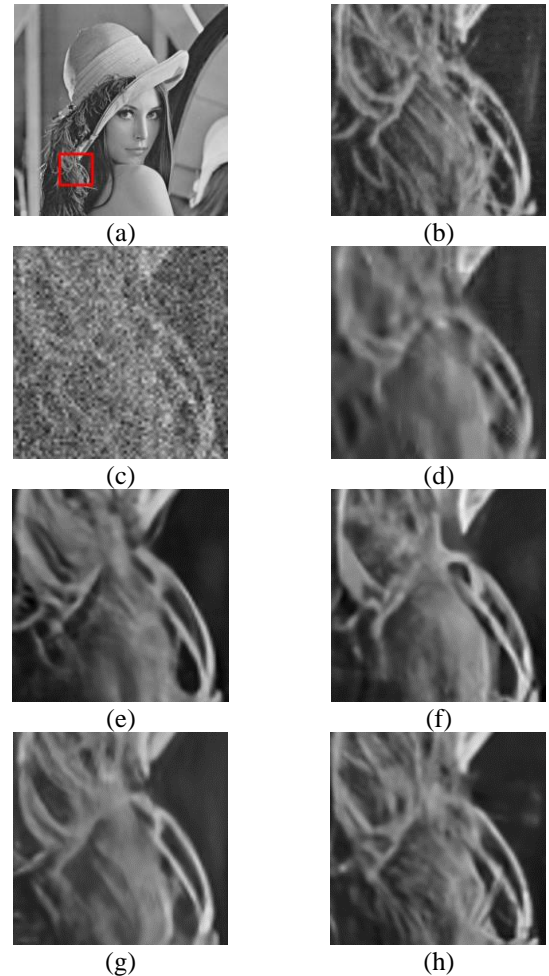


그림 7. Set12 데이터 셋의 잡음이 제거된 확대 영상에 대한 주관적 화질 비교,
(a) 확대 영역 표시 영상 (b) 원본 영상 (c) 잡음 영상 ($\sigma = 50$) (d) BM3D (e) DnCNN (f) FFDNet
(g) WDnCNN (h) 제안하는 네트워크

그림 7은 Set12 데이터 셋 중 Lena 영상의 잡음 제거 결과 영상의 주관적 비교를 위해 부분 확대한 영상이다. 확대한 부분에서 모자의 깃털 장식을 비교하였을 때 제안하는 네트워크인 그림 7(h)가 기존 알고리즘 및 네트워크에 비해 원본 영상인 그림 7(b)와 가장 유사하게 복원된다. 그림 7의 경우 잡음 레벨이 50으로 잡음이 심한 경우임에도 제안하는 네트워크는 영상 내 디테일을 원본과 유사하게 보존하는 잡음 제거 결과를 보인다.

5. 결론

본 논문에서는 특징 추출 블록과 이중 경로 네트워크를 이용하여 영상 내 잡음을 효율적으로 제거하는 기법을 제안한다. 특징 추출 블록을 통해 잡음이 있는 영상 내에서 영상의 특징을 추출하고 이중 경로 네트워크를 이용하여 각 경로의 서로 다른 결과를 공유함으로써 다양한 특징을 학습하게 된다. 그 결과 기존 알고리즘 및 네트워크와 비교하였을 때 객관적 지표인 PSNR에서 높은 수치를 보였으며 주관적 화질 비교에서도 동물의 털이나 깃털 장식과 같은 물체의 질감 및 디테일을 잘 보존하면서 잡음을 효과적으로 제거한다. 또한 기존 알고리즘 및 네트워크의 잡음 제거 결과와 비교하였을 때 제안하는 네트워크는 열화 현상이 적게 나타나 보다 영상을 선명하게 복원하는 결과를 보인다.

감사의 글

이 연구는 2020년도 산업통상자원부 및 한국산업기술평가관리원(KEIT) 연구비 지원에 의한 연구임('20013726').

참고문헌

- [1] X. Liu, Y. Lu, T. Wu, and T. Yuan, "An Improved Local Descriptor based Object Recognition in Cluttered 3D Point Clouds," *International Journal of Computers Communications & Control*, vol. 13, no. 2, pp. 221-234, Apr. 2018.
- [2] H. Hosseini, B. Xiao, and R. Poovendran, "Google's Cloud Vision API Is Not Robust To Noise," *International Conference on Machine Learning and Applications(ICMLA)*, pp. 101-105, Dec. 2017.
- [3] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3-D transform-domain collaborative filtering," *IEEE Transactions on Image Processing*, vol. 16, no. 8, pp. 2080-2095, Sep. 2007.
- [4] K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang, "Beyond a Gaussian Denoiser: Residual Learning of Deep CNN for Image Denoising," *IEEE Conference on Computer Vision and Pattern Recognition(CVPR)*, vol. 26, no. 7, pp. 3142-3155, Feb. 2017.
- [5] V. Nair and G. Hinton, "Rectified Linear Units Improve Restricted Boltzmann Machines," *International Conference on Machine Learning (ICML)*, pp. 807-812, Jun. 2010.
- [6] S. Ioffe and C. Szegedy "Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift," *Proceedings of the 32nd International Conference on Machine Learning Research(PMLR)* 37:448-456, Feb. 2015.
- [7] K. Zhang, W. Zuo, and L. Zhang, "FFDNet: Toward a Fast and Flexible Solution for CNN based Image Denoising," *IEEE Transactions on Image Processing*, vol. 27, no. 9, pp. 4608-4622, Sep. 2018.
- [8] F. Yu and V. Koltun, "Multi-Scale Context Aggregation by Dilated Convolutions," *International Conference on Learning Representations(ICLR)*, arXiv preprint arXiv:1511.07122, pp. 1-13, Nov. 2015.
- [9] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," *IEEE Conference on Computer Vision and Pattern Recognition(CVPR)*, pp. 770-778, Jun. 2016.
- [10] J. Pan et al., "Learning Dual Convolutional Neural Networks for Low-Level Vision," *IEEE Conference on Computer Vision and Pattern Recognition(CVPR)*, pp. 3070-3079, Jun. 2018.
- [11] S. Roth and M. J. Black, "Fields of experts," *International Journal of Computer Vision*, vol. 82, no. 2, pp. 205-229, Apr. 2009.
- [12] R. Zhao, K. M. Lam, and D. P. Lun, "Enhancement of a CNN-Based Denoiser Based on Spatial and Spectral Analysis," *2019 IEEE International Conference on Image Processing(ICIP)*, pp. 1124-1128, Sep. 2019.