Wavelet 기반의 영상 디테일 향상 잡음 제거 네트워크

*정군, 위승우, 정제창
한양대학교 융합전자공학과
*gooni0906@gmail.com, slike0910@hanyang.ac.kr, jjeong@hanyang.ac.kr

WDENet: Wavelet-based Detail Enhanced Image Denoising Network

*Jun Zheng, Seungwoo Wee, and Jechang Jeong
Department of Electronic Engineering, Hanyang University

요 약

최근 딥 러닝 기법의 하나인 합성곱 신경망(Convolutional Neural Network, CNN)은 영상 잡음(Noise) 제거 분야에서 전통적인 기법보다 좋은 성능을 나타내고 있지만 학습하는 과정에서 영상 내 디테일한 부분이 손실될 수 있다. 본 논문에서는 웨이블릿 변환(Wavelet Transform)을 기반으로 영상 내 디테일 정보도 같이 학습하여 영상 디테일을 향상하는 잡음 제거 합성곱 신경망 네트워크를 제안한다. 제안하는 네트워크는 디테일 향상 서브 네트워크(Detail Enhancement Subnetwork)와 영상 잡음 추출 서브 네트워크(Noise Extraction Subnetwork)를 이용하게 된다. 실험을 통해 제안하는 방법은 기존 알고리듬보다 디테일 손실 문제를 효과적으로 해결할 수 있었고 객관적 품질 평가인 PSNR(Peak Signal-to-Noise Ratio)와 주관적 품질 비교에서 모두 우수한 결과가 나온 것을 확인하였다.

1. 서론

현재 카메라 성능이 점점 발전해왔지만 얻은 디지털 영상 내에는 코딩, 전송, 수집, 처리하는 과정 또는 야간이나 악천후 등의 환경적 요인으로 인해 영상 내 잡음이 생성되기도 한다. 이러한 영상 내에 존재하는 잡음(Noise)을 제거해 영상 화질을 개선하는 과정을 영상 잡음 제거(Image Denoising)라고 한다.

잡음으로는 가산성 백색 가우시안 잡음(Additive White Gaussian Noise, AWGN)이 많이 알려져 있고, 모든 주파수 대역에 존재하는 잡음을 의미하며 가우시안 형태의 확률 변수를 가지고 있기 때문에 자연계에서 흔히 볼 수 있는 잡음과 유사한 효과를 얻을 수 있다. 따라서 AWGN으로 필터링 기반의 방법부터 딥 러닝 기반의 방법까지 다양한 영상 잡음 제거 기술이 연구되고 있다.

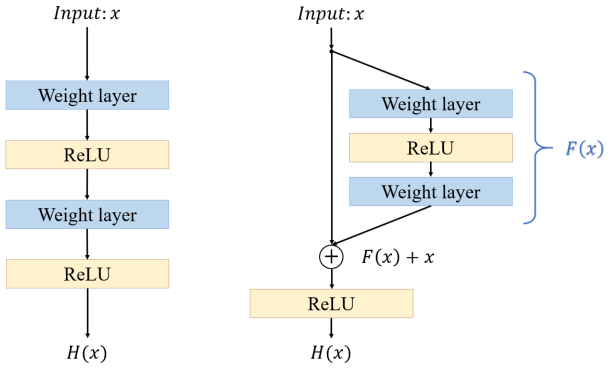
전통적인 잡음 제거 기술 중에서 잘 알려진 알고리듬으로는 비 지역적 평균 필터(Non-local Mean Filter) [1]를 사용하는 BM3D [2]가 있다. 이는 잡음이 있는 영상에서 블록을 추출하여 전체 영역에서 비슷한 블록들을 찾아 3차원 그룹으로 구성하고, 그룹에 협업 필터링(Collaborative Filtering)을 실행한 다음 얻어진 추정치를 원래 위치에 반환하는 알고리듬이다. 최근에는 이런 필터링 기반의 잡음 제거 기술에 이어 딥 러닝 기반의 알고리듬들이 높은 성능을 보이면서 이를 이용한 잡음 제거 기술이 많이 개발되고 있다. 즉, 네트워크로 얻은 결과와 원본의 차이를 최소화하게 학습하여 잡음을 제거하는 방법이다. 대표적인 것이 잔차 학습(Residual Learning) [3], 합성곱, 배치 정규화(Batch Normalization) [4]와 ReLU를 이용하여 AWGN을 제거하는 신경망인 DnCNN [5]이 있다. 하지만 이러한 방법들은 잡음을 제거하는 과정에서 영상 내 디테일이 손실되는 문제점이 있다.

웨이블릿 변환은 주파수 영역에서 잡음과 원본 영상을 분리하여 디테일 정보를 학습하는 데에 효과적이고 U-Net [6] 구조는 경계선 검출(Edge Detection)에 예민하다. 따라서 본 논문에서는 기존 딥 러닝 기반의 단일 CNN 모델과 달리 웨이블릿 변환과 U-Net을 기반으로 영상 내 디테일 정보도 같이 학습하여 영상 디테일을 향상시키는 잡음 제거 합성곱 신경망 네트워크를 통해 영상 잡음 제거 과정에서 디테일한 부분이 손실되는 문제를 개선한다. 제안하는 네트워크는 디테일 향상 서브 네트워크와 영상 잡음 추출 서브 네트워크를 이용한다. 실험은 잡음 레벨(Noise Level)이 각각 15, 25, 50인 AWGN이 추가된 영상을 사용하여 잡음 제거가 진행되고 객관적 품질 평가는 PSNR을 사용했다.

본 논문의 구성은 다음과 같다. 2장에서는 제안한 방법에서 사용되는 이론적 지식인 잔차 학습과 2D 영상신호에 대한 이산 웨이블릿 변환(Discrete Wavelet Transform, DWT)을 살펴본 후, 3장에서는 제안하는 방법의 전체적인 구조에 대해 설명한다. 4장에서는 실험 결과를 통해 기존 방법과 제안하는 방법을 비교하여 성능을 확인하고 마지막으로 5장에서는 결론을 맺는다.

2. 이론적 배경

2.1 잔차 학습



(a) CNN (b) Residual Learning Block

그림 1. 기존 합성곱 신경망 구조와 잔차 학습

CNN을 통해 영상의 특징을 추출하는 과정에서 네트워크의 층(Layer)을 너무 깊게 쌓으면 파라미터 수의 증가에 의해 에러가 커지므로 과적합(Overfitting)의 문제와 달리 기울기가 소실 및 증폭(Gradient Vanishing/Exploding) 현상이 발생함으로써 성능이 저하되는 문제점이 있다.

이를 해결하기 위해 입력을 그대로 출력에 더하는 숏컷 연결(Shortcut Connection)이 추가되는 구조를 가진 잔차 학습[3]이 제안되었다. 즉 잔차 학습은 그림 1 (a)처럼 기존 CNN의 출력 값 $H(x)$를 학습하는 대신 그림 1 (b)에서 입력 값 x와 출력 값 $H(x)$의 차이 값 $F(x)$를 학습하는 구조로 구성된다. 따라서 CNN의 출력 값 $H(x)$를 다음과 같이 재정의한다.

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

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

이렇게 되면 위의 숏컷 연결을 합성곱에 사용하여 역전파(Back Propagation)과정에서 x를 미분하면 적어도 1이상의 값으로 최소한의 기울기를 만들어서 네트워크가 깊더라도 기존 방법보다 안정적인 학습이 가능하고 층에 대해서 잔차를 학습한다는 구조로 성능 저하되는 문제를 해결할 수 있다.

2.2 2D 영상신호에 대한 이산 웨이블릿 변환

DWT는 피처 맵(Feature Map)의 주파수와 위치 정보를 파악할 수 있는 특성이 있으므로 영상 내 디테일한 부분을 강조하는 데 도움이 된다. 따라서 본 논문에서는 2D 이산 웨이블릿 변환을 사용하게 되고 그중 계산량이 적은 Haar 웨이블릿을 선택하였다. 그리고 DWT는 가역성을 가지고 있기 때문에 영상 내의 모든 정보를 다운 샘플링하고 역 웨이블릿 변환(Inverse Wavelet Transform, IWT)로 에지(Edge) 성분에 손실 없이 복원할 수 있는 특성을 가지고 있다.

그림 2처럼 DWT는 영상의 크기가 가로, 세로 각 $\frac{1}{2}$씩 줄어든 입력 영상이 저역 필터(Low-pass Filter) f_{LL}와 고역 필터(High-pass Filter) f_{LH}, f_{HL}, f_{HH} 총 4개의 필터를 거쳐 4개의 서브 밴드(Sub Band) 영상 x_{LL} (평균), x_{LH} (수평), x_{HL} (수직), x_{HH} (대각선)로 나누어지고 4개의 서브 밴드 영상에 IWT을 적용하게 되면 원본 영상으로 복원이 가능하다. Haar 웨이블릿의 경우 4개의 필터는 아래와 같이 정의된다.

$$f_{LL} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}, \; f_{LH} = \begin{bmatrix} -1 & -1 \\ 1 & 1 \end{bmatrix},$$

$$f_{HL} = \begin{bmatrix} -1 & 1 \\ -1 & 1 \end{bmatrix}, \; f_{HH} = \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} \quad (3)$$

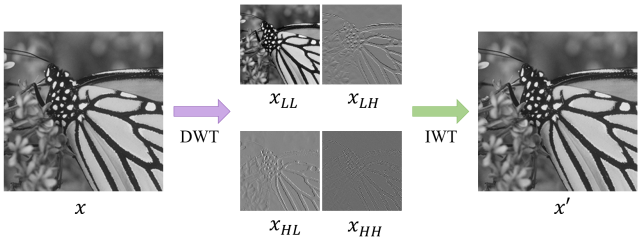


그림 2. 2D 영상의 DWT와 IWT의 변환 과정 (Haar 웨이블릿)

3. 제안하는 방법

3장에서는 제안하는 WDENet 모델의 전체적인 구조와 두 서브 네트워크를 사용하는 이유에 대해 설명한다.

3.1 WDENet 모델 구조

그림 3은 제안하는 WDENet의 전체 네트워크 구조이다. 제안하는 WDENet 모델은 그림 3과 같이 디테일을 향상하는 서브 네트워크와 영상 잡음을 추출하는 서브 네트워크로 나눌 수 있다. WDENet 구조에서는 먼저 입력 영상인 잡음 영상으로부터 두 개의 서브 네트워크를 거친다. 디테일 향상 서브 네트워크는 영상 내 디테일을 강조시키는 역할로 디테일 맵(Detail Map)을 추출하고, 영상 잡음 추출 서브 네트워크는 DnCNN 모델과 비슷한 구조로 잡음을 추출하여 디테일이 손실된 잡음 제거 영상 피처 맵을 추출한다. 다음으로 디테일 정보 특징과 잡음 정보를 동시 파악하기 위해 서로 다른 관점에서 추출한 두 결과를 연결(Concatenation)한다. 마지막으로 세 번의 깊이(Depth)가 128인 3×3 합성곱, 배치 정규화와 ReLU의 결합 구조와 한 번의 깊이가 128인 3×3 합성곱에 잔차 학습을 추가해 최종적으로 디테일이 향상된 잡음 제거 영상을 얻는다.

3.2 영상 디테일 향상 서브 네트워크

제안하는 WDENet 모델에서 영상 디테일 향상 서브 네트워크는 딥 러닝 기반의 영상 잡음 제거 과정에서 디테일한 부분이 손실되는 문제를 개선하기 위해 사용되고 웨이블릿 변환과 U-Net 구조를 사용한다

웨이블릿 변환은 주파수 영역에서 잡음과 원본 영상을 분리하여 디테일 정보를 학습하는 데에 효과적이고 DWT는 가역성을 가지고 있기 때문에 영상 내의 모든 정보를 손실 없이 복원할 수 있는 특성이 있다. U-Net 구조는 생체의학 분야에서 영상 분할(Image Segmentation)을 목적으로 제안된 구조로 경계선 검출에 예민하다. 이러한 장점 때문에 기존보다 우수한 디테일

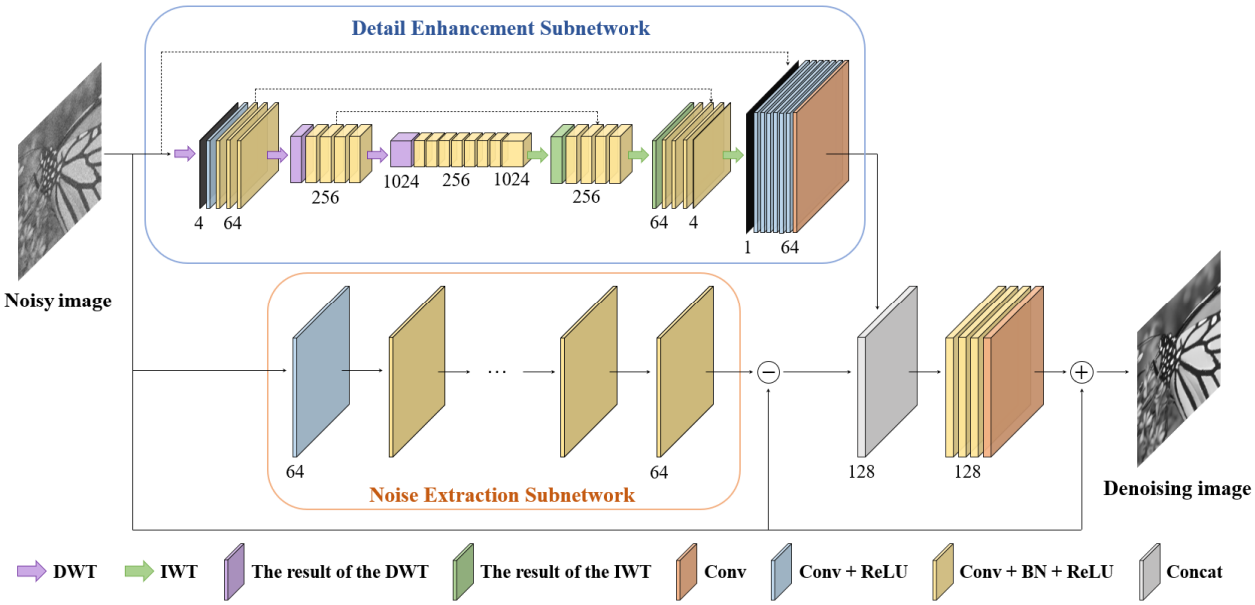


그림 3. 제안하는 WDENet 모델의 전체 네트워크 구조

추출을 위해 이 두 기법을 결합하여 WDENet에 사용한다. 즉, 그림 3처럼 U-Net 구조에서 DWT로 다운 샘플링(Down-sampling)하여 크기가 절반으로 줄어든 결과를 연결하고 IWT로 업 샘플링(Up-sampling)하여 크기가 두 배씩 늘어나게 된다. WDENet의 U-Net 구조는 6개의 합성곱 블록으로 이루어진다. 각 블록에서는 4번의 3×3 합성곱 계층을 사용하고 첫 번째 블록의 깊이는 64이고, 두 번째와 세 번째 블록의 깊이는 256으로 설정하고 나머지 블록의 깊이는 대응되게 설정해 준다. 추가로 네 번째 블록의 마지막 합성곱 깊이는 업 샘플링을 위해 1024로 설정해 준다. 그리고 학습을 돕기 위해 블록마다 배치 정규화와 ReLU를 추가한다. 마지막으로는 U-Net 구조로 얻어진 1채널 디테일 정보 영상으로부터 6번의 깊이가 64인 3×3 합성곱과 ReLU의 결합 구조와 한 번의 깊이가 64인 3×3 합성곱 계층을 통해 디테일 맵이 추출된다.

3.3 영상 잡음 추출 서브 네트워크

영상 잡음 추출 서브 네트워크는 DnCNN [5]모델과 유사하게 17번의 깊이가 64인 3×3 합성곱 계층으로 구성되고 첫 번째 계층은 합성곱과 ReLU의 결합 구조이고 나머지 계층은 합성곱, 배치 정규화와 ReLU의 결합 구조를 사용하여 영상 디테일 향상 서브 네트워크에서 추출한 결과와 같은 깊이를 얻게 된다. 여기서는 잡음과 영상 내 객체 위주의 피처 맵을 추출한다.

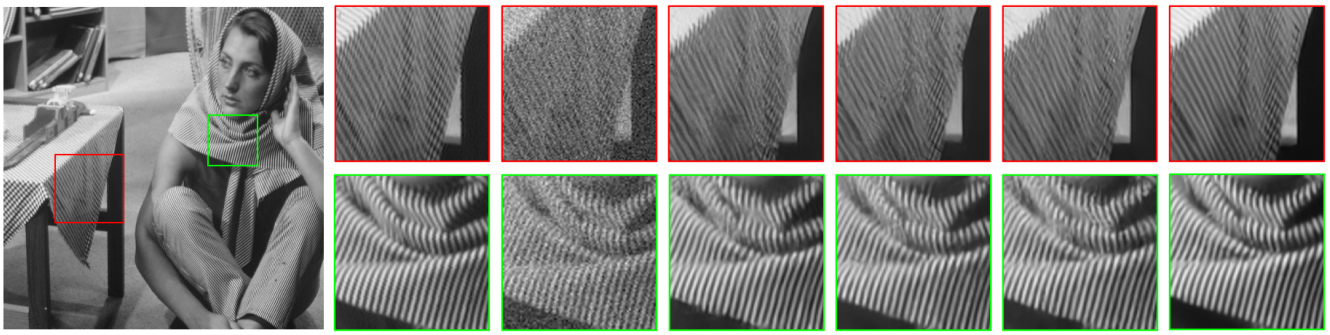
4. 실험 결과 및 분석

본 논문의 실험은 Python 3.7기반의 Pytorch를 사용하고 CPU는 i7-4790, GPU는 GeForce GTX 1080 Ti를 사용하여 성능을 평가했다. 학습에 필요한 데이터 셋(Dataset)은 DIV2K을 사용하고, 패치 크기(Patch Size)는 96×96 으로 설정하여 사용했다. 배치 크기(Batch Size)는 16, 에폭(Epoch)은 55로 설정했다. Loss함수로는 제곱 오차 함수(Mean Squared Error, MSE)를 이용하였고, 옵티마이저(Optimizer)로는 아담 옵티마이저(Adaptive Moment Estimation-optimizer, Adam-optimizer)를 사용하였다. 옵티마이저에 사용하는 학습률(Learning Rate)은 0.0001로 시작하고 30에폭 이후에는 0.00001로 감소한 학습률을 사용하여 학습시킨다. 그리고 학습된 모델의 성능을 평가하기 위해 테스트 셋은 Set12와 BSD68 데이터 셋을 이용했다. 실험 결과는 객관적인 수치인 PSNR결과와 주관적인 품질과 디테일을 관찰하여 기존 방법과 비교한다.

표 1. 다양한 잡음 레벨에 따라 기존 방법과 제안한 방법의 PSNR(dB) 결과 비교
(기존 방법 중 가장 좋은 결과를 초록색으로, 제안한 방법은 빨간색으로, 두 결과의 차이를 파란색으로 강조하여 표시한다.)

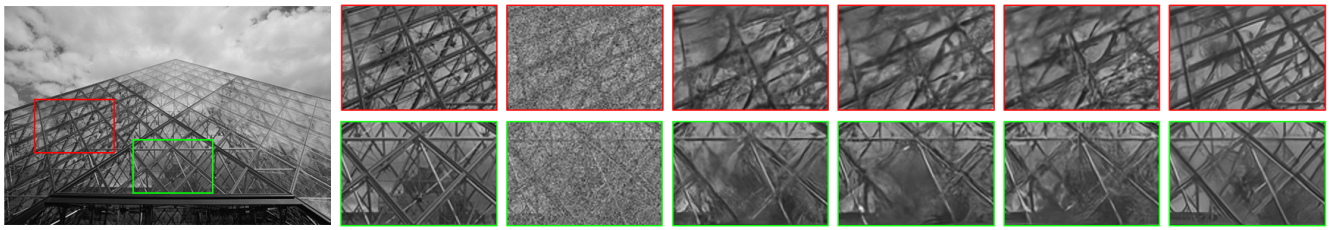
Methods	Set12			BSD68		
	Noise Level σ			Noise Level σ		
	15	25	50	15	25	50
Noisy	24.60	20.17	14.15	24.61	20.17	14.15
BM3D [2]	32.40	30.00	26.74	31.10	28.19	25.65
IRCNN [7]	32.73	30.36	27.08	31.63	29.14	26.18
DnCNN [5]	32.82	30.38	27.12	31.72	29.21	26.23
FFDNet [8]	32.73	30.42	27.29	31.63	29.19	26.29
Proposed	33.04	30.71	27.63	31.81	29.37	26.41
Difference	+0.22	+0.29	+0.34	+0.09	+0.16	+0.12

표 1은 잡음 레벨이 15, 25, 50인 경우 기존 방법과 제안한 방법의 잡음 제거 성능을 객관적 수치인 PSNR 결과로 보여준다. 제안하는 방법은 데이터 셋, 잡음 레벨과 관계없이 모든 상황에서 기존 알고리듬 및 네트워크에 비해 높은 수치를 보여주고 있다. Set12 데이터 셋의 경우 제안한 WDENet은 15의 잡음 레벨에서 가장 좋은 결과를 얻은 기존 방법보다 0.22만큼 향상하고, 25의 잡음 레벨에서 가장 좋은 결과를 얻은 기존 방법보다 0.29만큼 향상하며 50의 잡음 레벨에서 가장 좋은 결과를 얻은 기존 방법보다 0.34만큼 더 높은 수치를 보였다. BSD68 데이터 셋의



(a) Ground truth (b) Noisy (c) IRCNN (d) DnCNN (e) FFDnet (f) Proposed

그림 4. 잡음이 제거된 결과 화질 비교 (Set12, $\sigma = 25$)



(a) Ground truth (b) Noisy (c) IRCNN (d) DnCNN (e) FFDnet (f) Proposed

그림 5. 잡음이 제거된 결과 화질 비교 (BSD68, $\sigma = 50$)

경우 15의 잡음 레벨에서 가장 좋은 결과를 얻은 기존 방법보다 0.09만큼 향상하고, 25의 잡음 레벨에서 가장 좋은 결과를 얻은 기존 방법보다 0.16만큼 향상하며 50의 잡음 레벨에서 가장 좋은 결과를 얻은 기존 방법보다 0.12만큼 더 높은 수치를 보였다. 위의 결과를 통해 제안하는 방법은 낮은 잡음 레벨보다는 높은 잡음 레벨에 대해 더 나은 성능을 발휘하는 것을 알 수 있다.

그림 4는 Set12 데이터 셋의 Barbara영상으로 기존 잡음 제거 방법과 제안한 방법을 비교하기 위한 결과 영상이다. 영상 중 목도리 부분 및 테이블 보 부분을 비교하였을 때 제안하는 WDENet으로 형성한 그림 4 (f)는 기존 알고리듬에 비해 목도리의 구겨짐과 테이블 보의 라인 패턴이 원본 영상인 그림 4 (a)에 가장 가깝게 표현되었고 영상의 디테일 정보가 향상되었다.

그림 5는 BSD68 데이터 셋으로 기존 잡음 제거 방법과 제안한 방법을 비교하기 위한 결과 영상이다. 영상 내 복잡한 경계선들이 모여 있는 부분을 비교해 봤을 때 잡음 레벨이 50임에도 불구하고 제안하는 WDENet으로 얻은 그림 5 (f)는 기존 알고리듬에 비해 디테일 정보의 보존이 눈에 띄게 향상되었고, 블러(Blur) 현상이 적어 영상을 디테일하게 복원할 수 있었다.

5. 결론

본 논문에서는 웨이블릿 U-Net 구조를 기반으로 영상 잡음 제거 과정에서 손실되는 디테일 정보를 보완하는 합성곱 신경망 네트워크를 제안한다. 기존의 단일 CNN 모델과 달리 제안하는 WDENet은 입력 영상인 잡음 영상을 디테일 향상 서브 네트워크와 잡음 추출 서브 네트워크를 통해 다른 관점에서 얻은 두 결과를 연결하여 같이 학습하는 구조로 잡음과 함께 디테일 정보가 손실되는 문제를 개선할 수 있다. 실험 결과에서 볼 수 있듯이 WDENet은 기존 잡음 제거 알고리듬 및 네트워크보다 객관적 품질 평가 수치인 PSNR 결과와 주관적 품질 비교에서 모두 뛰어난 결과가 나타났고 디테일 향상에 좋은 성능을 보여주었다.

참고문헌

[1] Buades, B. Coll, and J.M. Morel, "A non-local algorithm for image denoising," *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, Vol. 2, pp. 60-65, 2005.

[2] 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, 2007.

[3] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 770-778, 2016.

[4] S. Ioffe and C. Szegedy, "Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift," *International Conference on Machine Learning*, pp. 448-456, 2015.

[5] K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang, "Beyond a Gaussian Denoiser: Residual Learning of Deep CNN for Image Denoising," *IEEE Transactions on Image Processing*, Vol. 26, No. 7, pp. 3142-3155, 2017.

[6] O. Ronneberger, P. Fischer, and T. Brox, "Unet: Convolutional Networks for Biomedical Image Segmentation," *Proceeding of Medical Image Computer-Asisted Intervention*, pp. 234-241, 2015.

[7] K. Zhang, W. Zuo, S. Gu, and L. Zhang, "Learning Deep CNN Denoiser Prior for Image Restoration," *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3929-3938, 2017.

[8] 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, 2018.