

# 가보웨이블릿 특징맵을 입력으로 한 CNN 기반 영상잡음제거기

권혁진, 조남익  
서울대학교

hjkwon@ispl.snu.ac.kr, nicho@snu.ac.kr

## Image Denoiser Based on Gabor Wavelets and Convolutional Neural Network

Hyuk Jin Kwon, Nam Ik Cho  
Seoul National University

### 요 약

최근 Convolutional Neural Network (CNN)에 영상이 아닌 비학습적 알고리즘으로부터 도출된 특징맵을 입력함으로써 영상처리 성능 및 계산자원 효율성 향상을 이룬 보고가 늘어나고 있다. 본 논문에서는 이러한 점을 바탕으로 가보웨이블릿 특징맵을 입력으로 하는 CNN 기반 영상잡음제거기를 제안하고 그 성능 및 특징을 고찰하였다. 즉 기존의 CNN에서는 일반적인 영상을 입력하는 반면에 본 논문에서는 영상으로부터 추출한 웨이블릿 계수들을 입력하였고, 이를 통하여 기존의 방법에 비하여 성능을 유지하면서 계산량을 줄일 수 있는 가능성을 확인하였다.

### 1. 서론

웨이블릿 변환은 주어진 영상에 대한 multiresolution analysis를 가능하게 하며 현재까지 다양한 영상처리에 도입되어왔다. 이 중 특히 영상잡음제거 분야의 경우 많은 영상잡음이 주파수 영역에서 특징적인 형태를 지니며 multiresolution analysis는 영상의 주파수 평면에서의 분석을 가능하게 하기 때문에 웨이블릿 변환의 이용이 활발하게 이루어지고 있다.

최근 그래픽 처리 하드웨어의 발전과 대용량 데이터베이스의 구축에 힘입어 학습 및 활용이 가능해진 심층신경망은 영상처리의 여러 분야에서 성과를 내고 있으며 영상잡음제거에도 활발하게 사용되고 있다. 이 중 영상잡음제거에 주로 활용되는 대표적인 심층신경망은 다층의 convolution 필터를 이용한 Convolutional Neural Network (CNN)이다.

영상잡음제거를 위한 위의 두가지 방식은 각각의 장단점을 지니고 있다. 첫번째 방법인 웨이블릿 기반의 방식은 영상 내의 특징 추출 시 추출대상의 물리적인 특징을 반영한 알고리즘의 설계가 가능하고 필요한 파라미터의 수를 조절할 수 있다는 장점이 있으나, 현재 성능측면에서 CNN을 이용한 방법에 뒤처진다는 단점이 있다. 두번째 방법인 CNN 기반의 방식의 경우 특징 추출을 위한 알고리즘의 설계를 요하지 않고 충분한 규모의 학습가능한 자료가 주어진 경우 높은 성능을 보이는 장점이 있으나, 이러한 성능을 확보하기 위해서는 많은 수의 학습가능한 파라미터와 학습을 위한 자료가 필요하다는 단점이 있다.

이에 따라 최근의 연구에서는 [1] 이러한 두가지 방식의 서로 상보적인 장단점을 이용해 웨이블릿 등 물리적인 의미를 가지고 설계된 특징을 CNN의 입력으로 하여 처리성능을 높이면서도 학습에 필요한 파라미터 수를 줄이는 접근이

이루어지고 있다. 이 중 특히 가보웨이블릿(Gabor Wavelet)은 질감을 주요한 특징으로 가지는 물체의 분석에 널리 사용되고 있고, 특히 나이추정과 같은 영상처리 분야에서 영상에서 추출된 가보웨이블릿 특징을 CNN의 입력으로 적용하였을 때 성능개선이 있음이 보고되고 있다. 본 논문에서는 이러한 사실들에 기반하여 가보웨이블릿 특징을 입력으로 한 CNN 기반 영상잡음제거기를 제안 및 구현하고 그 성능을 평가하였다.

### 2. 네트워크 구조

가보웨이블릿 특징 기반 영상잡음제거 CNN 네트워크는 [그림. 1]에서 제시된 바와 같이 특징맵을 계산하는 가보웨이블릿 뱅크 부분과 주어진 특징맵을 바탕으로 영상을 복원하는 CNN 부분으로 구성된다. 본 논문에서는 다음의 수식을 기반으로 가보 웨이블릿을 계산하였다.

$$G(x, y) = \exp\left(-\frac{X^2 + \gamma^2 Y^2}{2\sigma^2}\right) \times \cos\left(\frac{2\pi}{\lambda} X\right)$$

수식 1. 2차원 가보 웨이블릿

가보웨이블릿 뱅크 구성 시 웨이블릿의 이산시간에서의 크기는  $3 \times 3$ 으로 설정하였으며 잡음 검출을 위한 주파수평면에서의 해상도를 확보하기 위해 [수식 1]의  $\sigma$  값을  $[0, 10]$  구간에서 등간격으로 32개 추출하고, 시간평면에서의 해상도를 확보하기 위해 [수식 1]의  $\theta$ 를  $[0^\circ, 180^\circ]$  구간에서  $10^\circ$ 의 등간격으로 18개를 추출하여 총 576개의

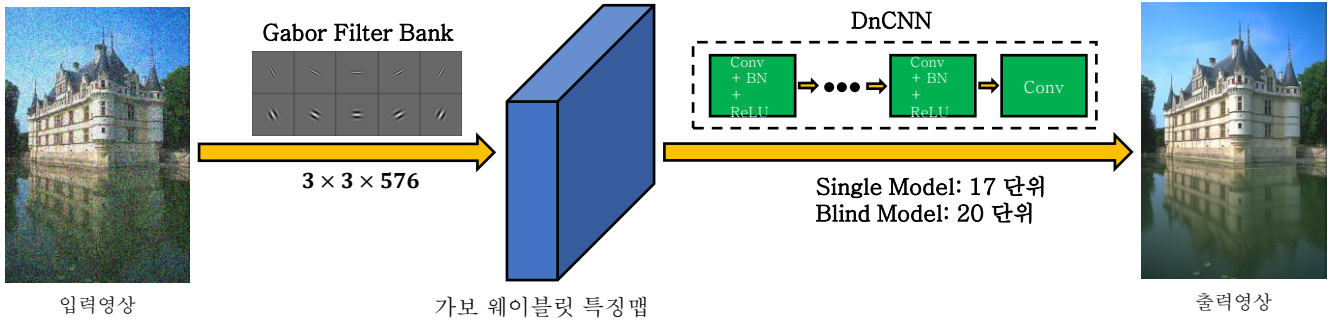


그림 1. 가보웨이블릿 특징맵을 입력으로 한 CNN 기반 영상잡음제거기

가보 웨이블릿 뱅크를 구성하였다. 그후 구성된 가보웨이블릿 뱅크를 입력영상에 적용하여 가보웨이블릿 특징맵을 계산하였고 이를 위해 입력영상을 흑백영상으로 변환하였다.

본 논문에서는 앞서 설명된 방법으로 계산된 입력영상의 가보웨이블릿 특징맵을 현재 널리 쓰이고 있는 CNN 기반 영상잡음제거기인 [2]의 첫 convolution 필터를 제외한 구조에 입력으로 주어 영상잡음제거기를 구성하였다. 제안한 영상잡음제거기의 CNN 부분의 구성을 위하여 본 논문에서는 [2]의 방식을 따라 convolution, batch-normalization, ReLU activation 의 세 종류 구성요소의 순차적인 처리를 한 단위로 정의하고 영상잡음제거기를 잡음의 표준편차를 알고 있다고 가정한 경우 (Single Model)와 모르고 있다고 가정한 경우 (Blind Model)의 2 가지로 구분한 뒤 Single Model 의 경우 17 단위, Blind Model 의 경우 20 단위를 가지도록 설계하였다.

### 3. 실험 방법 및 결과

본 논문에서는 [3]에서 제시된 방법과 동일하게 실험을 진행하고 잡음제거기의 성능을 PSNR 과 SSIM 을 기준으로 평가하였다. 구체적으로 설명하면, 주어진 400 장의 180 x 180 크기의 이미지를 상하 flip, 90° 회전, 90° 회전 후 상하 flip, 180° 회전, 180° 회전 후 상하 flip, 270° 회전, 270° 회전 후 상하 flip 을 통해 augmentation 한 후 Single Model 을 위해 경우 총 204,800 개의 40 x 40 크기의 patch 를, Blind Model 을 위해 총 387,200 개의 50 x 50 크기의 patch 를 만들었다.

본 논문에서는 이 patch 들을 이용해 영상잡음제거기의 CNN 부분을 학습하기 위해 Single Model 의 경우 표준편차가 각각 15, 25, 50 에 해당하는 random Gaussian noise 를 더한 3 가지 종류의 잡음영상을 생성하고 정답영상을 원본 patch 로 하는 3 쌍의 학습쌍을 구성하여 각 표준편차에 해당하는 3 개의 Single Model 을 학습하였고, Blind Model 의 경우 0 부터 55 사이의 값을 가지는 임의의 수를 추출한 뒤 이를 표준편차로 가지는 random Gaussian noise 를 각 patch 에 더하여 생성된 잡음영상과 원본 patch 를 정답으로 하는 학습자료를 구성하여 제시된 Blind Model 을 학습시켰다.

학습 시 설정된 batch 의 크기는 128 이며 초기화는 He initialization 을 이용하였고, 최적화는 Adam optimization 을 사용하였다. 학습속도는 1e-3 크기로 시작한 후 10 세대 마다 learning rate 를 1/10 만큼씩 감소시켰고, 전술 한 방법으로 총

30 세대를 학습하였다.

학습된 영상잡음제거기의 성능평가는 BSD6 데이터셋과 URBAN100 데이터셋에서 진행되었다. 영상잡음제거 성능평가를 위해 두 데이터셋의 원본 이미지에 표준편차를 각각 15, 25, 50 로 가지는 random Gaussian noise 를 더한 3 가지 종류의 잡음영상을 생성하고 학습된 4 가지 (표준편차 15, 25, 50 에서 학습된 3 개의 Single Model 과 1 개의 Blind Model)의 제시된 영상잡음제거기의 성능을 기존의 영상잡음제거 방법들과 BSD68 데이터셋의 경우 [표 1.]에서 PSNR 을 기준으로, [2]의 방법과 SSIM 을 기준으로 [표 3.]와 같이 비교하였고 URBAN100 데이터셋의 경우 [표 2.]에서 [2]의 방법과 PSNR 과 SSIM 을 기준으로 비교하였다.

실험 결과로부터 제시된 영상잡음제거기가 BSD68 데이터셋에서는 PSNR 을 기준으로 [2]와 비교할 때 Single Model 의 경우  $\sigma=15$  에서 0.12dB  $\sigma=25$  에서 0.07 dB,  $\sigma=50$  에서 0.11dB 만큼 성능이 떨어지나, [2]이 아닌 다른 방식의 영상잡음제거기들 보다는 뛰어난 성능을 보인 것을 확인하였고, Blind Model 의 경우  $\sigma=15$  에서는 0.02dB 만큼 성능이 높았고  $\sigma=25$  에서의 성능은 29.16 dB 로 동일 했으며  $\sigma=50$  에서는 0.02 dB 만큼 성능이 떨어지나 Single Model 과 마찬가지로 [2]가 아닌 다른 비교대상 보다는 좋은 성능을 보인 것을 확인할 수 있었다. SSIM 을 기준으로 판단하여도 Single Model 과 Blind Model 모두 제안된 방법과 [2]가  $\sigma=15, 25, 50$  에서 0.02 내의 차이의 값을 가져 서로 비슷한 성능을 보임을 확인하였다. URBAN100 데이터셋의 경우는  $\sigma=25, 50$  인 Single Model 의 경우를 제외한 모든 경우에서 제안된 방법이 PSNR 과 SSIM 측면에서 더 좋은 성능을 보였다.

### 4. 고찰

제안된 가보웨이블릿 특징맵을 입력으로 한 CNN 기반 영상잡음제거기는 현재 널리 쓰이는 CNN 기반 영상잡음제거기인 [2]와 비교 시 극선형태가 많은 BSD68 데이터셋에서는 잡음의 표준편차가 알려진 경우 모든 실험 구성에서 낮은 성능을, 잡음의 표준편차가 알려지지 않은 것을 가정한 실험구성에서는 모든 경우 0.02dB 내의 차이를 가지는 비슷한 성능을 보이고 상대적으로 직선형 물체가 많은 URBAN100 데이터셋에서는.  $\sigma=15$  인 Single Model 을 제외하고는 PSNR 과 SSIM 측면에서 좋은 성능을 보였다.

이러한 성능 향상은 가보 웨이블릿이 URBAN100 데이터셋에 주로 있는 직선형 물체의 특징맵을 생성하는데

표 1. BSD68 PSNR 비교

Methods	BM3D[4]	WNNM[5]	EPLL[6]	MLP[7]	CSF[8]	TNRD[3]	DnCNN[2] (Single)	DnCNN[2] (Blind)	Ours (Single)	Ours (Blind)
$\sigma = 15$	31.07	31.37	31.21	-	31.24	31.42	<b>31.73</b>	31.61	31.66	<b>31.63</b>
$\sigma = 25$	28.57	28.83	28.68	28.96	28.74	28.92	<b>29.23</b>	29.16	29.16	29.16
$\sigma = 50$	25.62	25.87	25.67	26.03	-	25.97	<b>26.23</b>	<b>26.23</b>	26.12	26.21

표 2. URBAN100 PSNR, SSIM 비교

Methods	PSNR				SSIM			
	DnCNN[2] (Single)	DnCNN[2] (Blind)	Ours (Single)	Ours (Blind)	DnCNN[2] (Single)	DnCNN[2] (Blind)	Ours (Single)	Ours (Blind)
$\sigma = 15$	<b>32.59</b>	32.10	32.58	<b>32.36</b>	<b>0.9275</b>	0.9150	0.9245	<b>0.9212</b>
$\sigma = 25$	29.83	29.63	<b>29.87</b>	<b>29.86</b>	0.8800	0.8735	<b>0.8810</b>	<b>0.8798</b>
$\sigma = 50$	26.14	26.16	<b>26.15</b>	<b>26.39</b>	0.7790	0.7806	<b>0.7797</b>	<b>0.7905</b>

표 3. BSD68 SSIM 비교

Methods	DnCNN[2] (Single)	DnCNN[2] (Blind)	Ours (Single)	Ours (Blind)
$\sigma = 15$	0.8826	<b>0.8860</b>	<b>0.8848</b>	0.8846
$\sigma = 25$	0.8190	0.8190	<b>0.8206</b>	<b>0.8190</b>
$\sigma = 50$	<b>0.7076</b>	<b>0.7014</b>	0.6925	0.7005

유리하기 때문에 분석된다. 또한 BSD68 데이터셋에서의 잡음 모델 별 성능 차이는 제안된 영상잡음제거기가 [2]와는 달리 압력으로부터 첫번째 convolution 필터를 학습이 불가능한 형태의 가보 웨이블릿으로 변경하면서 생긴 현상으로 파악된다. 즉, [2]의 경우 표준편차가 알려진 경우는 학습가능한 첫번째 convolution 필터가 주어진 표준편차를 가지는 잡음에만 적합하도록 일정 정도 과학습되어 동일한 표준편차를 가지는 잡음에 대해서는 유리한 결과를 가질 수 있었으나, 학습 시 잡음의 표준편차를 모르게 되는 경우 과학습을 통한 성능향상이 불가능 해지므로 첫번째 convolution 필터를 가보웨이블릿과 같이 미리 계산된 계수를 가진 학습이 불가능한 필터로 대체하여도 성능에 큰 차이가 발생하지 않고 가보 웨이블릿으로 특징맵을 추출하기에 유리한 직선형 물체들을 대상으로 한 경우는 성능향상이 이루어진 것으로 분석된다. 이러한 현상은 적절한 파라미터 값을 설정하여 가보웨이블릿으로 CNN의 convolution 필터들 중 일부를 변경할 수 있으며 이를 통해 현재 학습에 많은 수의 자료 및 계산능력이 필요한 CNN 기반 영상잡음제거기의 문제를 다소 완화할 수 있는 가능성이 있는 것으로 판단된다.

## 5. 결론 및 제언

본 논문에서는 가보웨이블릿 특징맵을 입력으로 하는 CNN 기반 영상잡음제거기를 제안하고 그 성능을 평가하였다. 제시된 영상잡음제거기는 영상잡음의 특성이 알려진 경우는 기존의 방식보다 낮은 성능을 보이나, 영상 잡음의 특성이 알려지지 않은 경우는 기존의 방식과 주목할 만한 차이가 발견되지 않음을 확인하였고 이로부터 CNN의 convolution 필터의 일정부분을 가보웨이블릿으로 변경함으로써 최종 성능에 영향 없이 CNN 기반 영상잡음제거기의 학습 시 필요한 자료 및 계산자원의 양을 다소 완화할 수도 있을 것이라는 가능성을 제시하였다.

## 감사의 글

This work was supported by the SNU-Hojeon Garment Smart Factory Research Center funded by Hojeon Ltd.

## 참고문헌

- [1] H.J. Kwon, H.I. Koo, J.W. Soh, and N.I. Cho "Age Estimation Using Trainable Gabor Wavelet Layers In A Convolutional Neural Network". in Proc. IEEE Int. Conf. Image Process., Sep. 2019, pp. 3626- 3630.
- [2] K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang, "Beyond a Gaussian Denoiser: Residual learning of deep CNN for image denoising," IEEE Trans. Image Process., vol. 26, no. 7, pp. 3142- 3155, Jul. 2017.
- [3] Y. Chen and T. Pock, "Trainable nonlinear reaction diffusion: A flexible framework for fast and effective image restoration," IEEE Trans. Pattern Anal. Mach. Intell., to be published, doi: 10.1109/TPAMI.2016.259.
- [4] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3-D transform-domain collaborative filtering," IEEE Trans. Image Process., vol. 16, no. 8, pp. 2080- 2095, Aug. 2007.
- [5] S. Gu, L. Zhang, W. Zuo, and X. Feng, "Weighted nuclear norm minimization with application to image denoising," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., Jun. 2014, pp. 2862- 2869.
- [6] D. Zoran and Y. Weiss, "From learning models of natural image patches to whole image restoration," in Proc. IEEE Int. Conf. Comput. Vis., Nov. 2011, pp. 479- 486.
- [7] H. C. Burger, C. J. Schuler, and S. Harmeling, "Image denoising: Can plain neural networks compete with BM3D?" in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., Jun. 2012, pp. 2392- 2399.

[8] U. Schmidt and S. Roth, "Shrinkage fields for effective image restoration," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., Jun. 2014, pp. 2774– 2781.