

Functional Neural Networks 기반의 자기 지도적 영상 잡음 제거

*장영일 **조남익

서울대학교

*jyicu@snu.ac.kr **nicho@snu.ac.kr

Functional Neural Networks for Self-supervised Image Denoising

*Jang, Yeong Il **Cho, Nam Ik

Seoul National University

요약

기존 합성곱 신경망 기반의 잡음 제거 네트워크들은 학습을 위한 noisy-clean 데이터 쌍을 필요로 한다. 하지만 실제 카메라 잡음의 경우, 잡음에 대한 깨끗한 원본 영상을 얻는 것은 불가능하거나 많은 비용이 소모된다. 따라서 이러한 방법을 해결하기 위하여 원본 영상 없이 잡음 영상만으로만 잡음 제거 네트워크를 학습하는 방법들이 제안되어왔다. 그 중 카메라 잡음 영상을 처리하기 위한 대표적인 방법으로 학습과 추론에서 비대칭적인 downsampling을 사용하는 AP-BSN이 제안되었다. 본 논문에서는 Functional neural network를 AP-BSN 알고리즘에 적용하여 다양한 downsampling ratio에 대응되는 하나의 네트워크를 학습하였다. 이를 통해 기존 hyperparameter로 사용되던 downsampling ratio에 대한 결과를 하나의 네트워크에서 분석 및 확인하였다. 또한 해당 파라미터를 조절함으로써 다양한 잡음 제거 후보들을 추출하고 사용자가 원하는 잡음 제거 정도를 조절할 수 있도록 하였다.

1. 서론

영상의 잡음 제거(image denoising)는 열화된 잡음 영상으로부터 깨끗한 원본 영상을 복원하는 기법이다. 최근 합성곱 신경망(Convolutional Neural Networks) 기반의 알고리즘들이 잡음 제거 분야에서 우수한 성능을 보여주고 있다. 이러한 합성곱 신경망 기반의 네트워크들은 대부분 부가적 백색잡음(Additive White Gaussian Noise)을 대상으로 하며 합성 잡음 영상을 입력으로하는 네트워크 출력과 깨끗한 원본 영상과의 차이를 손실함수로 하여 잡음 제거 네트워크를 학습시킨다.

깨끗한 원본 영상에서 잡음을 쉽게 생성할 수 있는 합성 잡음과 달리 실제 잡음의 경우 학습 데이터 쌍을 얻기 어렵다. 카메라 잡음은 AWGN보다 복잡한 확률 분포를 띠며 카메라 ISP를 거치면서 잡음 간에 채널 및 공간적 상관관계(correlation)가 발생하여 실제적인 잡음을 모델링하기 어렵다. 실제 카메라 이미지로부터 학습 데이터 셋을 만드는 방법도 있지만 잡음 영상에 대한 원본을 얻는 것은 대부분 불가능하거나 많은 촬영 및 후처리가 필요하다.

이에 따라 깨끗한 원본 이미지에 대한 필요성 없이 잡음 이미지만을 이용하여 잡음 제거를 수행하는 알고리즘들이 최근 다수 제안되었다. Noise2Void [1]와 Noise2Self [2]는 잡음제거 네트워크에서 입력 영상의 마스킹을 통해 receptive field에서 자기 자신을 제외하는 방법을 사용하였다. 이를 통해 자기 지도적 손실함수를 사용해도 네트워크가 항등 함수(identity)로 수렴하는 것을 막고 영상의 잡음을 제거하였다. Laine

et al. [3] 은 네트워크의 receptive field를 반 평면으로 제한한 뒤 평행 이동 후 합쳐 구조적으로 마스킹없이 blind spot network (BSN)를 구현하였으며 D-BSN [4]은 masked convolution과 dilated convolution을 이용하여 BSN을 구현하였다.

이러한 BSN 기반의 자기 지도적 잡음제거 알고리즘은 AWGN이나 Poisson 잡음을 효과적으로 제거하지만 실제 카메라에서 촬영된 잡음은 잘 제거하지 못한다. BSN 기반의 알고리즘들은 잡음의 평균이 0이고 픽셀 간 독립적이라는 가정 하에서 성립한다. 하지만 카메라 잡음의 경우 raw 도메인에서 생성된 잡음이 카메라 ISP를 통과하면서 채널 및 공간적으로 상관관계가 발생하게 된다. 따라서 BSN을 활용해도 해당 잡음에 대한 정보가 주변 픽셀에 남아있어 자기 지도적 손실함수를 사용하였을 때 자기 자신으로 수렴하여 잡음 영상을 출력한다.

이를 극복하기 위해 AP-BSN [5]은 실제 카메라 영상 잡음에서 공간적 correlation을 없애기 위해 Pixel Downshuffling (PD) [6]을 적용하였으며 학습과 테스트에서 비대칭적으로 PD stride를 사용해 sRGB에서 잡음 영상만으로 잡음 제거 네트워크를 학습시켰다.

본 논문에서는 AP-BSN에서의 학습과 테스트의 비대칭성에 주목하여 해당 PD stride를 네트워크 입력으로 같이 학습할 수 있는 방법을 제안한다. 본 논문에서는 AP-BSN구조에 functional neural network [7]를 적용하여 네트워크 파라미터가 PD stride에 대한 함수가 되도록 설계하였다. 이를 통해 각 stride에 대한 네트워크를 각각 학습시키지 않고서도 하나의 학습된 네트워크로 학습 PD ratio와 추론 PD ratio에

다른 성능향상을 분석하였으며 FNN을 적용하여도 AP-BSN의 성능을 유지함을 확인하였다. 또한 해당 파라미터를 조정함으로써 ground truth가 없는 실제 환경에서 사용자가 원하는 잡음 제거 정도를 조정할 수 있도록 하였다.

2. 이론적 배경

2.1. AP-BSN

AP-BSN [5]은 실제 카메라 잡음 영상만으로 잡음 제거 네트워크를 학습하였다. 해당 논문은 카메라 잡음에 대응하기 위하여 pixel-shuffle downsampling(PD) [6]를 사용한다. PD는 초해상도에서 많이 사용되는 pixel-shuffling [8]의 역연산으로 s 간격으로 이미지를 subsampling 하여 s^2 개의 subimage들을 생성한다. 이렇게 생성된 subimage들에서는 공간적 correlation이 존재하는 카메라 잡음을 픽셀 간 독립인 것처럼 다룰 수 있게 해준다.

AP-BSN은 학습과 추론에서 비 대칭적인 PD ratio를 사용하였다. PD ratio가 클수록 잡음의 공간적 상관관계가 줄어들어 BSN기반의 자기 지도적 잡음 제거 네트워크를 잘 학습시킬 수 있다. 반면 PD ratio가 작을수록 입력 영상의 구조가 잘 보존되며 네트워크가 더 많은 주변 픽셀을 보고 잡음 제거 영상을 추정할 수 있다. 따라서 학습 시에는 픽셀 간 독립 가정이 잘 성립하는 PD_5 로 학습하고 추론시에는 정보가 더 많은 PD_2 로 잡음 영상을 입력하여 카메라 잡음에 대한 잡음 제거 성능을 크게 향상시켰다.

2.2. Functional Neural Networks

Functional neural Networks (FNN) [7]는 하나의 네트워크로 noise level, quality factor와 같은 영상 열화(image degradation)에 관한 다양한 파라미터에 대응하기 위해 제안되었다. FNN은 네트워크의 학습 파라미터를 영상 열화 파라미터에 대한 함수로 표현한다.

$$G(H(x); \theta_a, \theta_b) = \frac{H(x) - H(x_a)}{H(x_b) - H(x_a)} (\theta_b - \theta_a) + \theta_a$$

이 때 x 는 영상 열화 파라미터이며 $H(x)$ 는 해당 열화와 관련된 임의의 함수이고 x_b 와 x_a 는 x 가 가질 수 있는 상한과 하한을 나타낸다. 그리고 θ_a 와 θ_b 는 학습되는 네트워크 파라미터이다. 위 방법을 통해 네트워크 전체의 학습 파라미터 숫자는 2배가 되지만 x_a 와 x_b 사이의 임의의 값에 대하여 네트워크를 최적화시킬 수 있으며 각 열화 수준마다 네트워크를 학습하는 대신 하나의 네트워크로 여러 열화 모델에 대응할 수 있게 해준다.

3. 제안하는 방법

본 논문에서는 AP-BSN의 PD ratio에 따른 trade-off를 분석하고자 PD ratio를 FNN의 파라미터로 하는 네트워크를 설계하였다. 네트워크 구조는 AP-BSN과 동일한 D-BSN 구조를 사용하였으며 PD ratio의

범위로 $x_a = 1$, $x_b = 5$ 를 사용하여 유의미한 구간에서만 downsampling이 일어나도록 하였다. 또한 $H(x)=x$ 를 사용하여도 수렴이 잘 되는 것을 확인하였다. 제안하는 네트워크를 학습시키기 위해서 각 batch마다 2에서 5 사이의 랜덤한 downsampling ratio d 를 샘플링한다. 이 때 최종적인 손실함수는 다음과 같다.

$$L = \|PD_d^{-1}(F(PD_d(x); G(d; \theta_a, \theta_b))) - x\|_1$$

학습된 네트워크를 추론할 때는 네트워크 파라미터가 a , 실제 PD ratio가 b 일 때

$$FNN_{a/b}(x) = PD_b^{-1}(F(PD_a(x); G(a; \theta_a, \theta_b)))$$

로 계산된다. 이를 통해 각 PD ratio마다 다른 네트워크를 학습시키지 않아도 하나의 네트워크로 임의의 a/b 에 대한 잡음제거 결과를 얻을 수 있다.

4. 실험 결과 및 분석

제안하는 네트워크는 Smartphone Image Denoising Dataset (SIDD) [9] medium set의 잡음 이미지만을 사용하여 네트워크를 학습하였다. 또한 90도 회전과 반전을 이용한 8개의 이미지를 랜덤하게 데이터 증강을 수행하였다. ADAM optimizer로 학습률 (learning rate) $1e-4$ 로 총 500,000회 학습을 진행하였다. 최종적으로는 Darmstadt Noise Dataset에서 온라인으로 성능평가를 수행하였다.

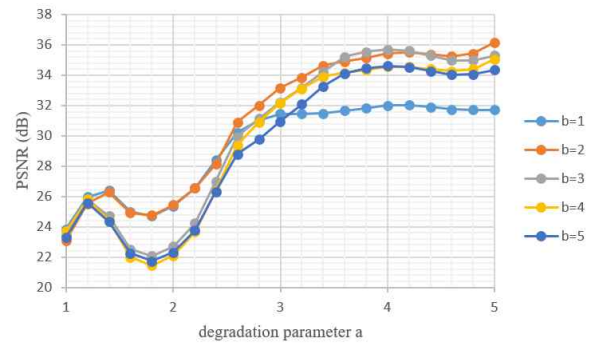


그림 1. FNN-BSN a/b에 따른 SIDD validation PSNR

그림 1은 네트워크를 결정하는 파라미터 a 와 PD ratio b 에 따라서 SIDD validation에서의 성능 변화를 측정한 그래프이다. 실험결과 전체적으로 a 가 커질수록 잡음 제거 성능이 증가하였다. 즉 실제 학습과 테스트의 입력 분포가 비슷한 것보다 픽셀 간 독립 조건이 학습에 더 큰 영향을 미치는 것을 알 수 있다. 입력의 PD ratio b 에 대해서는 $b=1$ 을 제외하고 b 가 작을수록 PSNR이 더 높았다. $b=1$ 인 경우는 네트워크 입력으로 downsampling을 수행하지 않고 원본 이미지를 그대로 넣어준 경우이다. $b=1$ 의 경우 학습 데이터로 사용하지 않고 잡음의 공간적 상관관계가 커 PSNR이 다른 downsampling ratio보다 낮게 측정되었다. 또한 FNN을 사용함으로써 AP-BSN과는 달리 학습 PD ratio를 float로

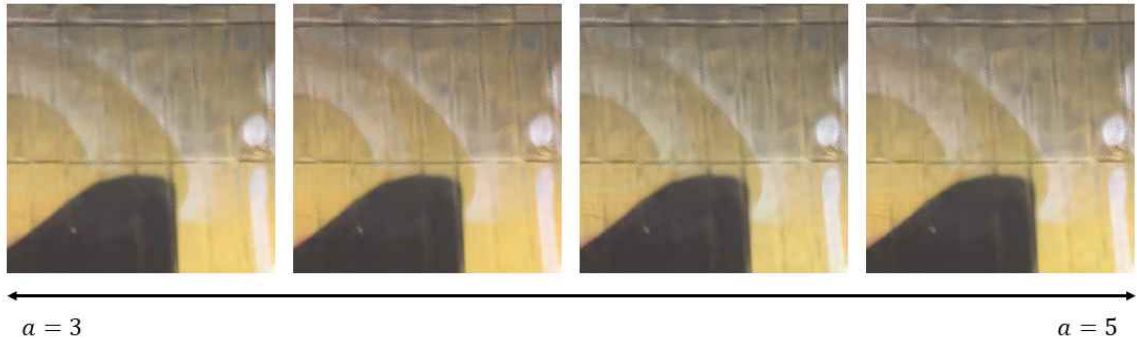


그림 2. 파라미터 α 에 따른 DND 영상의 잡음 제거 결과

조절할 수 있다. α 가 정수가 아닌 경우는 학습데이터에 없기 때문에 정수일 때 보다 성능이 일부 하락하지만 전체적으로 두 정수 α 의 함수 값에 대한 보간처럼 작동하는 것을 확인할 수 있다.

그림 2는 DND 이미지에 대하여 FNN-BSN 파라미터를 조절하며 출력한 결과이다. 이때 입력 이미지의 PD ratio b 는 2로 고정해주었다. 그림 2의 결과를 보면 모두 잡음이 어느 정도 제거된 상태에서 일부 색상 및 artifact 형태가 변화하는 것을 확인할 수 있다. 잡음 영상에 대한 원본이 없는 테스트 시에는 사용자가 잡음 제거 및 artifact를 조절할 수 있는 파라미터로 활용할 수 있을 것이라고 기대한다.

마지막으로 그림 1에서 가장 높은 성능을 보인 FNN-BSN5/2 모델에 대한 자기 지도적 잡음 제거 알고리즘 간 정량적 성능평가 결과는 표 1과 같다.

표 1. sRGB 도메인 DND 벤치마크에 대한 정량적 비교

	PSNR	SSIM
NAC [11]	36.20	0.925
CVF-SID(s^2) [12]	36.50	0.924
AP-BSN [5]	37.46	0.924
FNN-BSN 2/5	37.06	0.926

실험 결과 baseline인 AP-BSN보다 PSNR측면에서는 일부 하락하지만 SSIM은 소폭 상승하였으며 기존 성능을 유지하면서 파라미터를 조절하며 다양한 잡음 제거 후보들을 뽑을 수 있음을 확인하였다.

4. 결론

본 논문에서는 AP-BSN에 PD ratio를 입력으로 하는 functional neural network를 적용하여 학습시 사용한 downsampling ratio에 따라서 네트워크가 학습될 수 있도록 하였다. 이를 통해 성능에 큰 하락 없이 하나의 네트워크로 다양한 downsampling ratio에 대응하는 효과를 분석할 수 있었다. 하지만 제한하는 방법은 어떤 파라미터가 해당 이미지에 최선인지 결정하지 못한다는 한계가 있어 향후 각 이미지 별로 어떤 파라미터가 최선의 결과를 보이는 지 결정하는 연구가 필요하다.

감사의 글

이 논문은 2022년도 BK21 FOUR 정보기술 미래인재 교육연구단에 의하여 지원되었음. 본 연구는 산업통상자원부와 한국산업기술진흥원의 “지역혁신클러스터육성사업(R&D, P0002072)”으로 수행된 연구결과입니다. 이 논문은 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(2021R1A2C2C007220).

참고문헌

- [1] Krull, Alexander, Tim-Oliver Buchholz, and Florian Jug. "Noise2void-learning denoising from single noisy images." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019.
- [2] Batson, Joshua, and Loic Royer. "Noise2self: Blind denoising by self-supervision." International Conference on Machine Learning. PMLR, 2019.
- [3] Laine, Samuli, et al. "High-quality self-supervised deep image denoising." Advances in Neural Information Processing Systems 32 (2019): 6970-6980.
- [4] Wu, Xiaohe, et al. "Unpaired learning of deep image denoising." European conference on computer vision. Springer, Cham, 2020.
- [5] Lee, Wooseok, Sanghyun Son, and Kyoung Mu Lee. "AP-BSN: Self-Supervised Denoising for Real-World Images via Asymmetric PD and Blind-Spot Network." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022.
- [6] Zhou, Yuqian, et al. "When awgn-based denoiser meets real noises." Proceedings of the AAAI Conference on Artificial Intelligence. Vol. 34. No. 07. 2020.
- [7] Luo, Fangzhou, Xiaolin Wu, and Yanhui Guo. "Functional Neural Networks for Parametric Image Restoration Problems." Advances in Neural Information Processing Systems 34 (2021): 6762-6775.
- [8] Shi, Wenzhe, et al. "Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network." Proceedings of the IEEE conference on computer vision and pattern recognition. 2016.
- [9] Abdelhamed, Abdelrahman, Stephen Lin, and Michael S. Brown. "A high-quality denoising dataset for smartphone

- cameras." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018.
- [10] Plotz, Tobias, and Stefan Roth. "Benchmarking denoising algorithms with real photographs." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017.
- [11] Xu, Jun, et al. "Noisy-as-clean: Learning self-supervised denoising from corrupted image." IEEE Transactions on Image Processing 29 (2020): 9316-9329.
- [12] Neshatavar, Reyhaneh, et al. "CVF-SID: Cyclic multi-Variate Function for Self-Supervised Image Denoising by Disentangling Noise from Image." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022.