깊은 잔차 U-Net 구조를 이용한 실제 카메라 잡음 영상 디노이징

*장영일 **조남익

서울대학교

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

Real-world noisy image denoising using deep residual U-Net structure

*Jang, Yeongil **Cho, Nam Ik

Seoul National University

요약

부가적 백색 잡음 모델(additive white Gaussian noise, AWGN)에서 학습된 깊은 신경망(deep neural networks)을 이용한 잡음 제거기는 제거하려는 잡음이 AWGN인 경우에는 뛰어난 성능을 보이지만 실제 카메라 잡음에 대해서 잡음 제거를 시도하였을 때는 성능이 크게 저하된다. 본 논문은 U-Net 구조의 깊은 인공신경망 모델에 residual block을 결합함으로서 실제 카메라 영상에서 기존 알고리즘보다 뛰어난 성능을 지니는 신경망을 제안한다. 제안한 방법을 통해 Darmstadt Noise Dataset에서 PSNR과 SSIM 모두 CBDNet 대비 향상됨을 확인하였다.

1. 서론

잡음 제거(denoising)는 잡음이 포함된 관측 영상으로부터 잡음이 없는 영상을 복원해 내는 것을 목표로 한다. 기존 연구들은 잡음을 모델링할 때 부가적 백색 잡음(additive white Gaussian noise, AWGN) 모델을 사용한다. 부가적 백색 잡음 모델은 다른 왜곡 없이 이미지에 독립적인 Gaussian 잡음이 더해졌다고 가정한다. 이는 다음과 같이 수식으로 나타낼 수 있다.

$$y = x + n$$

여기서 y는 관측값이고 x는 원본 이미지, n는 평균이 0이고 분산이 σ^2인 i.i.d Gaussian 잡음이다.

지난 수십 년간 많은 고전적인 방법들이 부가적 백색 잡음 모델을 사용하여 잡음 제거 연구를 진행하였다. 최근 다른 컴퓨터 비전 분야와 마찬가지로 잡음 제거 분야 역시 합성곱 신경망(convolutional neural networks, CNN)의 도입으로 그 성능이 크게 향상되었다. 대표적으로 DnCNN[1]은 부가적 백색 잡음 모델 하에서 배치 정규화(batch normalization)과 잔차 학습(residual learning)을 적용하여 고전적인 알고리즘보다 높은 잡음 제거 성능을 보였다.

이러한 학습 기반의 잡음 제거 모델은 잡음의 특성이 학습 시와 같은 AWGN일 때에는 뛰어난 성능을 보여주지만 실제 카메라에서 얻어지는 잡음에 대해서는 성능이 크게 저하된다. AWGN과 달리 실제 카메라 센서에서 포착되는 노이즈는 훨씬 더 복잡하며 카메라 내부 파이프라인을 거치면서 그 특성이 변화한다.

이를 극복하기 위하여 CBDNet[2] 은 raw 이미지에서의 잡음을 이분산성 가우시안 잡음(heteroscedastic Gaussian noise)으로 모델링하고 카메라 내부 파이프라인을 모델링하여 실제적인 합성 잡음을 생성하였다. 잡음의 표준편차를 예측하는 esimation subnetwork와 실제 잡음제거를 수행하는 denoiser subnetwork를 결합한 구조를 제안하였으며 합성 잡음과 실제 잡음 이미지를 같이 학습에 사용하여 실제 카메라 잡음에 대해서 뛰어난 성능을 보였다.

최근 SIDD[4] 등 추가적인 카메라 잡음 이미지 데이터셋이 공개되어 더 많은 실제 잡음 데이터를 확보 가능해졌다. 이에 따라 본 논문에서는 CBDNet을 기본 구조로 하여 많은 데이터를 효과적으로 학습할 수 있도록 U-Net[3] 구조에 더 깊고 많은 잔차 연결(residual connection)을 사용한 잡음 제거 네트워크 구조를 제안한다. 실험 결과를 Darmstadt Noise Dataset[5]에서 평가하여 성능 지표가 CBDNet보다 향상됨을 확인하였고 시각적으로도 더 좋은 결과를 얻었다.

2. 제안하는 구조

제안하는 구조는 CBDNet을 기본 구조로 하여 작업하였다. CBDNet과 마찬가지로 estimator network와 denoiser network로 네트워크를 분리하였다. 본 논문의 estimator는 CBDNet의 구조를 따라 64채널의 conv 레이어 5층으로 하였다. Estimator의 출력은 CBDNet과는 달리 각 채널간의 잡음 정보도 포함할 수 있도록 3채널을 사용하였다.

기존 CBDNet에서 denoiser부분은 기본 U-Net 구조를 사용하였으나 본 논문에서는 학습을 향상시키기 위하여 그림 1과 같이 더 깊고 residual block을 사용한 수정된 U-Net 스타일의 구조를 사용하였다. 본 네트워크의 기본 구조인 resblock은 배치 정규화 없이 [6]에서 제안한 conv - ReLU-conv 구조를 사용하였다.

Denoiser의 첫 레이어는 conv+ReLU로 32채널 피쳐맵을 생성한

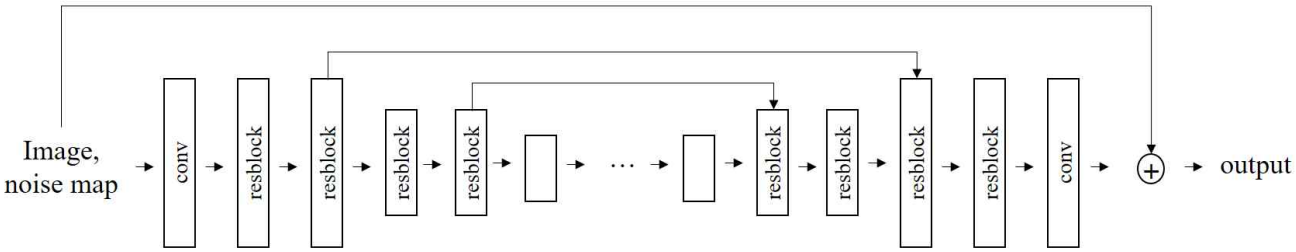


그림 1 제안하는 네트워크의 denoiser 구조

다. 이후 매 2 resblock 마다 간격이 2인 strided convolution을 사용하여 높이와 너비를 절반으로 줄이고 채널 수는 2배로 늘려주었다. 이를 총 4회 반복한 이후에는 다시 feature map의 크기를 키워주는 작업을 수행하였다. 이 때 [9]에서 제안한 subpixel shuffle을 사용하여 효과적으로 피쳐가 확장될 수 있게 하였다. 피쳐맵의 크기를 2배로 늘려준 이후 strided convolution 전의 피쳐맵과 채널 방향으로 결합시킨 후 1x1 convolution을 진행하여 채널 수를 조정하였다.

마지막 레이어에서는 활성 함수 없이 3채널 값을 출력하도록 하였고 [1, 2]와 마찬가지로 입력 영상의 잡음을 학습하는 잔차 학습(residual learning)을 사용하였다. 이 때 네트워크의 모든 합성곱의 필터 크기는 3x3을 사용하였으며 학습 시 feature map의 크기가 변하지 않도록 zero-padding을 사용하였다.

3. 학습 방법

학습을 위한 데이터 셋으로 SIDD medium dataset[4]의 sRGB 이미지를 사용하였다. SIDD mdeidum dataset은 320장의 실제 카메라 잡음 이미지와 이에 대응하는 잡음이 없는 이미지를 포함하고 있다. 각 이미지는 10개의 scene에 대해서 서로 다른 5개의 스마트폰 카메라와 다른 조명 및 ISO level에서 촬영하여 다양한 실제 잡음 이미지를 포함하고 있다.

제안하는 네트워크를 학습시키기 위하여 손실함수는 출력 이미지와 GT 이미지 간의 L1 loss를 사용하였다. 손실 함수를 최소화하기 위하여 $\beta_1 = 0.9$, $\beta_2 = 0.999$인 ADAM optimizer를 사용하였다. 이미지를 학습시키기 위하여 각 이미지에서 256x256 크기로 patch들을 추출하였으며 배치 크기(batch size)는 8로 학습을 진행하였다. 학습률(learning rate)의 경우 10^{-4}을 유지하여 500,000 회 학습한 후 추가로 10^{-5}으로 200,000회 학습하였다.

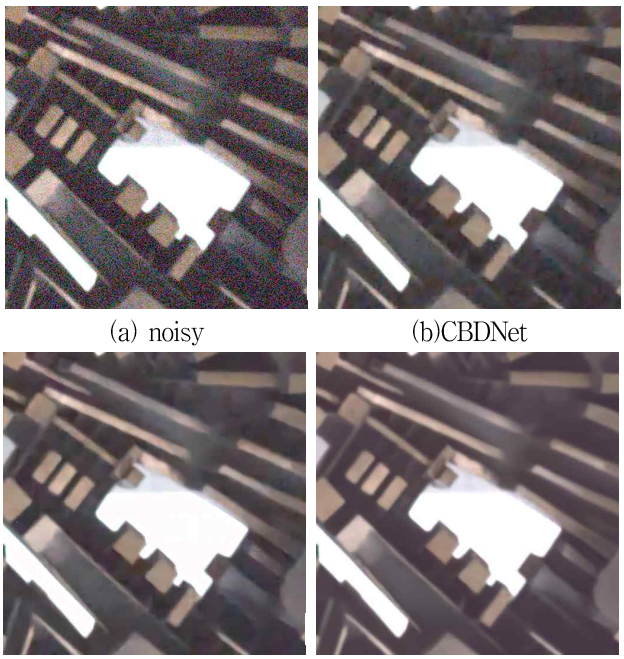
.

4. 실험 결과

실제 카메라 잡음에 대한 잡음 제거 성능을 평가하기 위하여 Darmstadt Noise Dataset를 사용하였다. DND는 4개의 다른 카메라와 다양한 ISO 레벨에서 촬영한 50장의 잡음 이미지로 구성되어 있다. DND는 사용자가 실험결과를 과적합 시킬 수 없도록 GT를 제공하지 않으며 온라인상으로 결과를 제출하여 평가하기 때문에 실제적 잡음 제거에서 신뢰성 있는 벤치마크로 사용되고 있다. 성능을 측정하기 위한 방법으로 기존 방법들과 PSNR(Peak Signal to Noise Ratio)과 SSIM(structural similarity)을 비교하였다. 실험 결과는 표 1과 같다.

표 1 DND 벤치마크 실험 결과

	PSNR	SSIM
DnCNN[1]	37.78	0.9308
CBDNet[2]	38.06	0.9421
N3Net[7]	38.32	0.9384
Path-restore[8]	39.00	0.9542
Ours	39.29	0.9519



(a) noisy (b)CBDNet

(c) Path-Restore (d) Ours

그림 2 DND 벤치마크에서의 정성적 실험 결과

실험 결과 기존 모델들보다 PSNR 측면에서 뛰어난 성능을 보였다. 특히 PSNR측면에서 CBDNet 보다 1.2dB, Path-restore보다 0.3dB 가량 향상된 성능을 보여주었다. 그림 2는 질적 평가를 위해 DND의 이미지를 나타내었다. CBDNet은 결과 이미지에 얼룩이 나타나는 반면 제안하는 네트워크는 CBDNet 대비 깨끗한 결과를 보여주었다.

5. 결론

본 논문에서는 CBDNet을 기반으로 SIDD 데이터 셋을 이용하여 더 깊고 연결이 많은 U-Net 구조의 네트워크를 제안하였다. 실제 잡음이 있는 이미지에 대하여 DND 벤치마크에서 PSNR과 SSIM 모두

향상된 결과를 얻었다. 하지만 제안하는 구조는 CBDNet과는 달리 잡음 수준에 대한 ground truth가 없어 estimator에 대한 손실 함수를 주기 어렵다. Estimator에 손실함수를 적용하여 잡음에 대한 정보를 학습할 수 있다면 추가적으로 더 성능을 향상시킬 수 있을 것으로 기대한다.

감사의 글

이 논문은 2019년도 BK21플러스 사업에 의하여 지원되었음

참고문헌

[1] Zhang, Kai, et al. "Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising." IEEE Transactions on Image Processing 26.7 (2017): 3142-3155.

[2] Guo, Shi, et al. "Toward convolutional blind denoising of real photographs." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2019.

[3] 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.

[4] 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.

[5] Plotz, Tobias, and Stefan Roth. "Benchmarking denoising algorithms with real photographs." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017.

[6] He, Kaiming, et al. "Deep residual learning for image recognition." Proceedings of the IEEE conference on computer vision and pattern recognition. 2016.

[7] Plötz, Tobias, and Stefan Roth. "Neural nearest neighbors networks." Advances in Neural Information Processing Systems. 2018.

[8] Yu, Ke, et al. "Path-Restore: Learning Network Path Selection for Image Restoration." arXiv preprint arXiv:1904.10343 (2019).

[9] 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.