

Unsupervised Image Denoising with Frequency Domain Knowledge (Supplementary Material)

Nahyun Kim*
nhkim21@kaist.ac.kr

Donggon Jang*
jdg900@kaist.ac.kr

Sunhyeok Lee
sunhyeok.lee@kaist.ac.kr

Bomi Kim
by5747@kaist.ac.kr

Dae-Shik Kim
daeshik@kaist.ac.kr

Korea Advanced Institute of Science
and Technology (KAIST),
Daejeon, Korea

1 Overview

In this supplementary material, we describe the architecture details and show the additional experiments as follows:

- In Section 2, we describe the architectures of two generators, i.e. G_{n2c} and G_{c2n} , and three discriminators, i.e. D_C , D_T , and D_S , in our framework.
- In Section 3, we show the additional results on CBSD68 [9] corrupted by AWGN with a noise level $\sigma = 25$.
- In Section 4, we show the additional qualitative results on real-world noise, i.e. Low-Dose CT authorized by Mayo Clinic [10] and SIDD [11].
- In Section 5, we show the results of an additional ablation study to demonstrate the validity of the perceptual loss L_{VGG} , the cycle consistency loss L_{CC} , and the reconstruction loss L_{Recon} .
- In Section 6, we show the results on several noise types, such as structured noise and Poisson noise, to evaluate the generalization ability of our method.

2 The Details of Architectures

Generator G_{n2c} For the noise removal generator G_{n2c} , we adopt the network introduced by [10]. The main idea of this architecture is multiple cascading connections at global and local levels which help to propagate low-level information to later layers and remove noise. The details of G_{n2c} are illustrated in Figure 1 and 2.

Generator G_{c2n} For the generator G_{c2n} , we adopt the U-Net based network that is similar to the architecture introduced by [10]. The role of this network is to translate images from the noise domain to the clean domain. The details of G_{c2n} are illustrated in Figure 3 and 4.

Discriminators D_C and D_T For the discriminators D_C and D_T , we employ the 70×70 PatchGAN discriminator [10] which classifies whether 70×70 image patches are real or fake. The details of D_C and D_T are illustrated in Figure 5.

Discriminator D_S For the spectral discriminator D_S , we employ the single linear unit as the spectral discriminator. The D_S takes a high-pass filtered 1D spectral vector and aims to classify whether the spectral vector is real or fake.

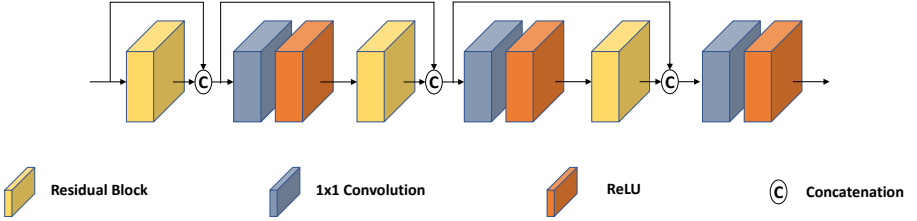


Figure 1: The architecture of Cascading Block used as the basic component in the G_{n2c} . We use the Residual Block proposed by [10] and the ReLU.

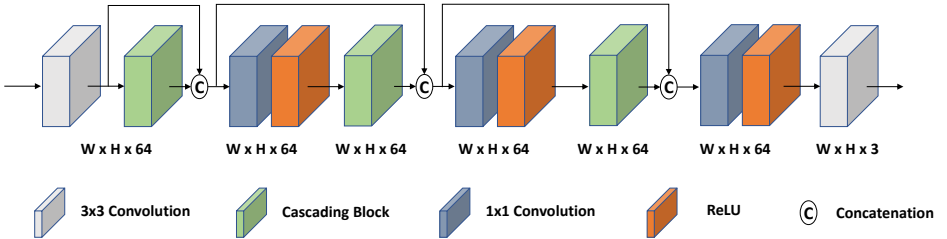


Figure 2: The architecture of generator G_{n2c} for noise removal. We use the convolution with kernel size=3, stride=1, and padding=1.

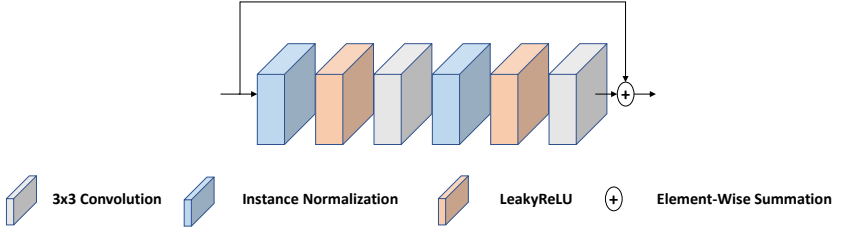


Figure 3: The architecture of Instance Residual Block used as the basic component in the G_{c2n} . We use the convolution with kernel size=3, stride=1, and padding=1 and the LeakyReLU with a slope of 0.2.

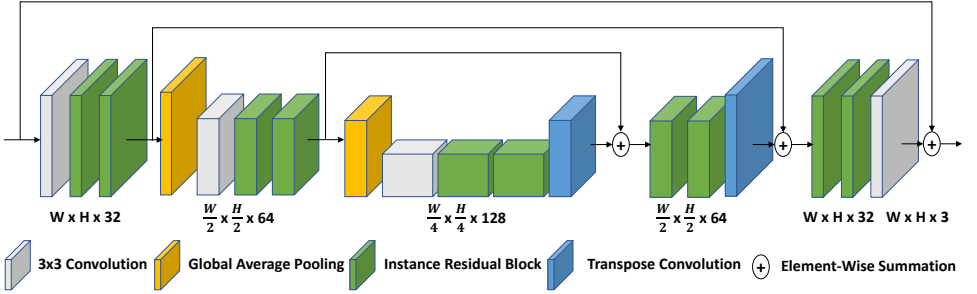


Figure 4: The architecture of generator G_{c2n} . We use the convolution with kernel size=3, stride=1, and padding=1 and transposed convolution with kernel size=3, stride=2, padding=1, and output padding=1.

3 Additional Results on AWGN

We additionally visualize the results for CBSD68 images corrupted by AWGN with a noise level $\sigma = 25$ and show the PSNR and SSIM in Figure 6 and 7. In Figure 6, our method outperforms other methods trained with unpaired dataset by at least +3.44dB and +0.08 in terms of PSNR and SSIM, respectively. LIR and N2V spoil the color and lights, but our method preserves both the color and lights and successfully removes the noise. We also show the challenging example that has repetitive high-frequency patterns hard to distinguish with noise in Figure 7. Our approach removes noise without artifact and also preserves the patterns of the zebra. Although our method is trained under unpaired settings, it shows comparable performance in PSNR and SSIM with the supervised models in Figure 7. Furthermore, compared to methods trained with unpaired dataset, our approach achieves the best performance in both PSNR and SSIM.

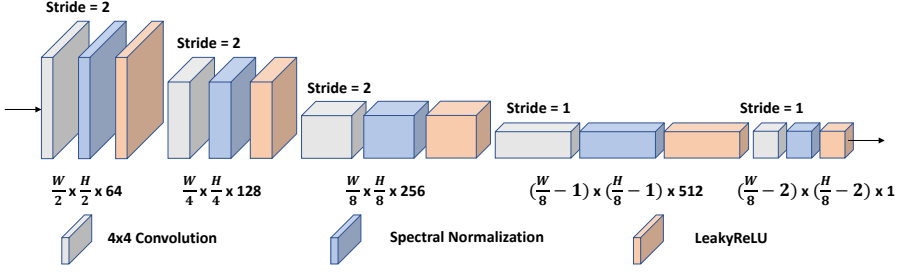


Figure 5: The architecture of discriminators D_C and D_T . We use the convolution with kernel size=4 and padding=1. Followed by the convolution, we use the spectral normalization [10] and the LeakyReLU with a slope of 0.2.

4 Additional Qualitative Results on Real-World Noise

4.1 Low-Dose CT

In this subsection, we show the additional qualitative results on Low-Dose CT dataset authorized by Mayo Clinic [10] in Figure 8. As shown in Figure 8, previous methods tend to lose details and generate blurred results. However, our method removes the noise, while preserving the details of organs. It shows that our method is also practical for medical image denoising.

4.2 Real Photographs

In this subsection, we visualize the additional qualitative results on SIDD [10] in Figure 9 and 10. As shown in Figure 9, previous methods tend to lose the texture and leave the noise. In contrast, our method removes the noise while preserving the texture compared to other baselines. In Figure 10, we observe that our method removes the intense noise while preserving the color of images compared to other baselines.

5 Additional Ablation Study

We conduct an additional ablation study to demonstrate the validity of the perceptual loss L_{VGG} , the cycle consistency loss L_{CC} , and the reconstruction loss L_{Recon} . First, to verify the effectiveness of the L_{VGG} , we only add the L_{VGG} . As shown in Table 1, when the L_{VGG} is used, both PSNR and SSIM increase by 0.07dB and 0.0068. It demonstrates that the perceptual loss L_{VGG} helps to improve the performance, preserving the semantics even after the noise has been removed. Next, to verify the contribution of L_{CC} , we integrate it with the L_{VGG} . We observe that the L_{CC} which enables the one-to-one mapping between noisy and denoised images improves the PSNR and SSIM by 0.08dB and 0.004. Finally, when we integrate the L_{Recon} with the L_{VGG} and the L_{CC} , both PSNR and SSIM increase by 0.15dB and 0.0063, thus showing the best results in PSNR and SSIM. Through this experiment, we validate that each of the losses contributes to the performance improvement.



Figure 6: Qualitative results of our method and other baselines on *CBSD68* corrupted by AWGN with a noise level $\sigma = 25$.

L_{VGG}	L_{CC}	L_{Recon}	PSNR (dB)	SSIM
✗	✗	✗	25.67	0.8204
✓	✗	✗	25.74	0.8272
✓	✓	✗	25.88	0.8312
✓	✓	✓	26.03	0.8375

Table 1: Ablation study. Quantitative results of our method with and without the perceptual loss L_{VGG} , the cycle consistency loss L_{CC} , and the reconstruction loss L_{Recon} on *CBSD68* corrupted by AWGN with a noise level $\sigma = 50$. We report the PSNR and SSIM (higher is better). The best results are marked in **bold**.

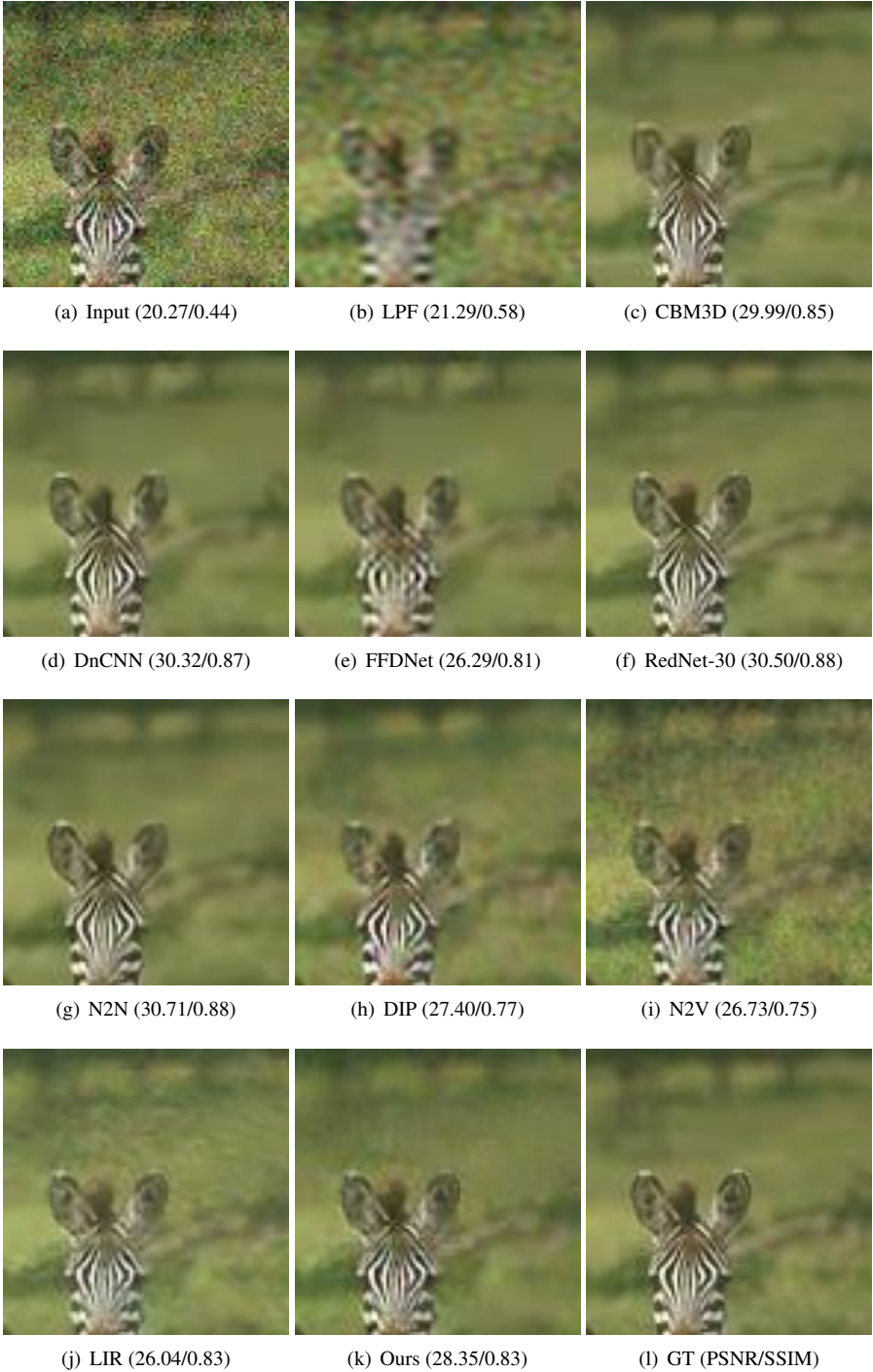


Figure 7: Qualitative results of our method and other baselines on *CBSD68* corrupted by AWGN with a noise level $\sigma = 25$.

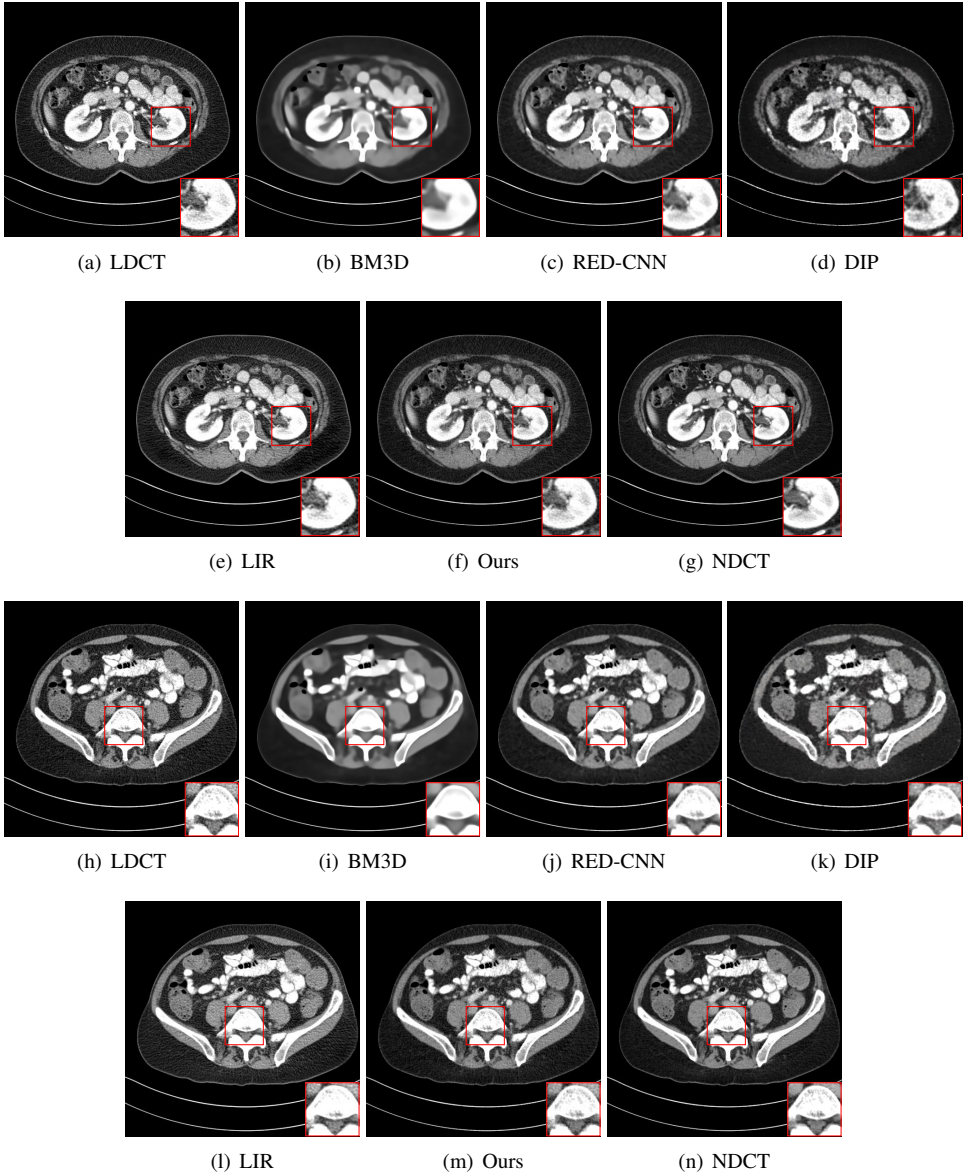


Figure 8: Qualitative results of our method and other baselines on *Mayo Clinic Low Dose CT dataset*. As shown in the highlighted red box, the reconstructed images by our method have few noise and preserve the details of organs. The display window is $[160, 240]$ HU.

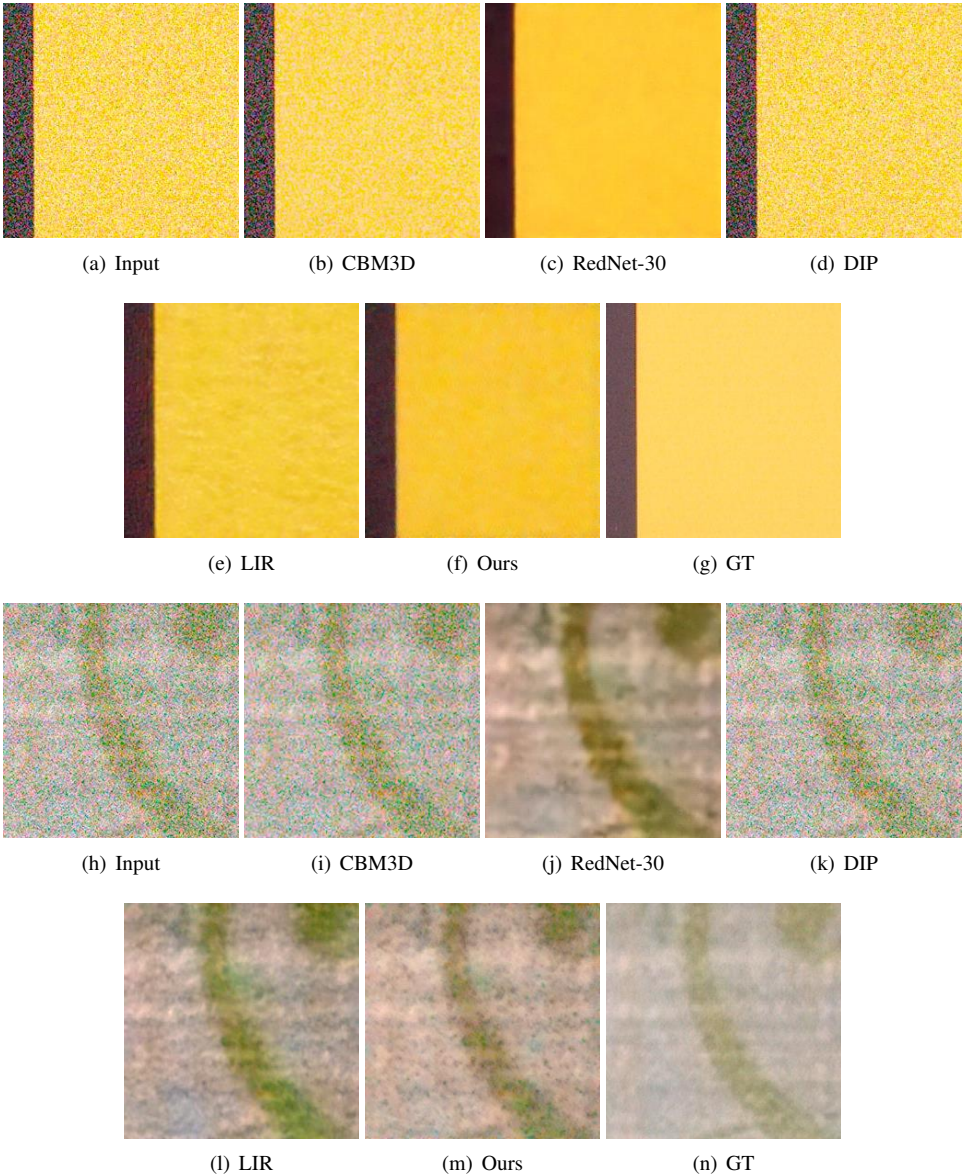


Figure 9: Qualitative results of our method and other baselines on real noisy data, *SIDD*.

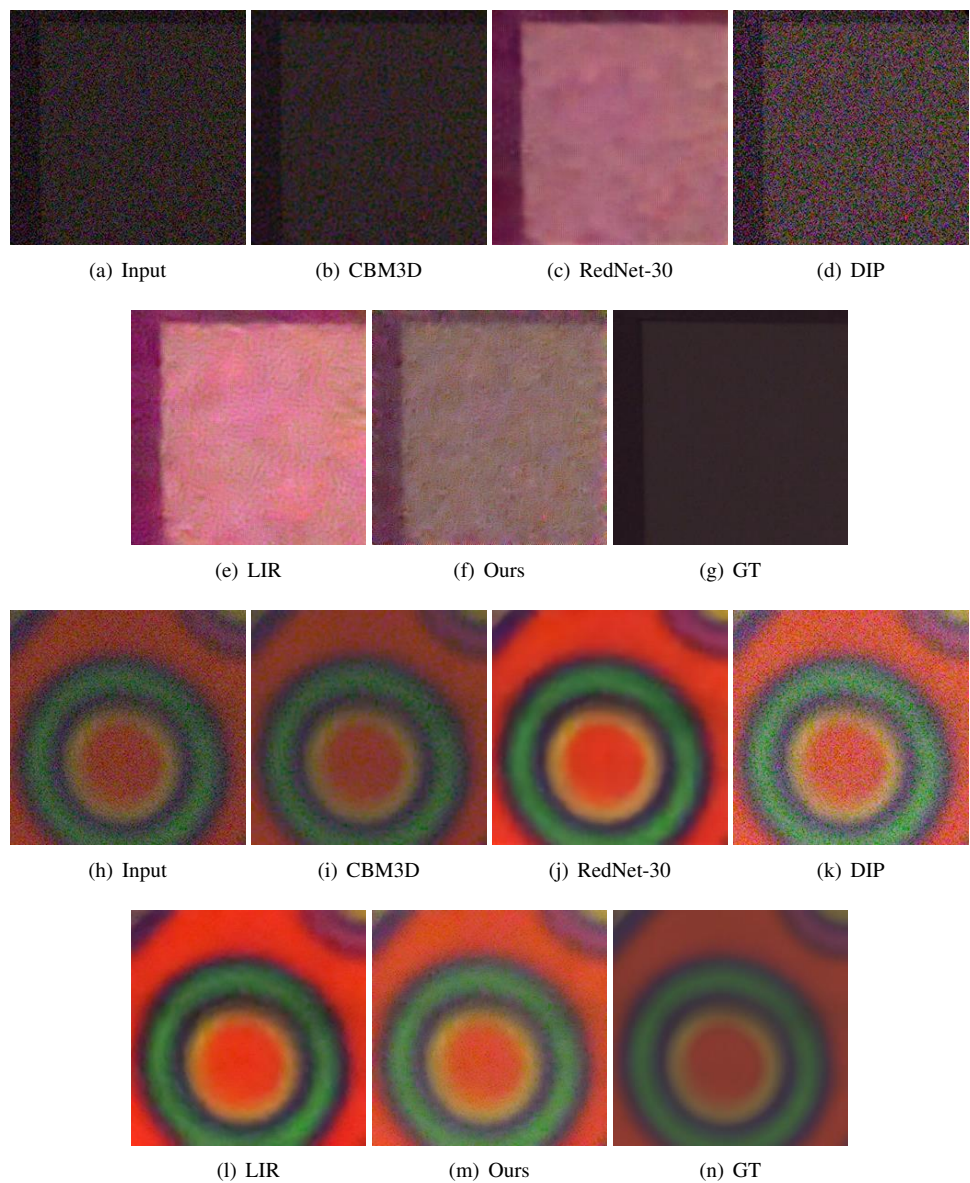


Figure 10: Qualitative results of our method and other baselines on real noisy data, *SIDD*.

6 Evaluation on Several Noise Types

6.1 Structured Noise

In this subsection, we show the results on structured noise. To generate the structured noise, we sample the pixel-wise i.i.d white noise, and convolve it with a 2D Gaussian filter whose a kernel size is 21×21 and σ is 3 pixel. For the train and evaluation, we follow the same setting as the setting for synthetic noise removal in the main paper. As shown in Figure 11, our method is able to remove complex noise compared to BM3D [9] and DIP [12]. Furthermore, while LIR [9] spoil the lights, our method successfully preserves both the color and lights. The quantitative results are summarized in Table 2. Our method outperforms the traditional and unsupervised methods, achieving the second-best performance in terms of PSNR and SSIM.

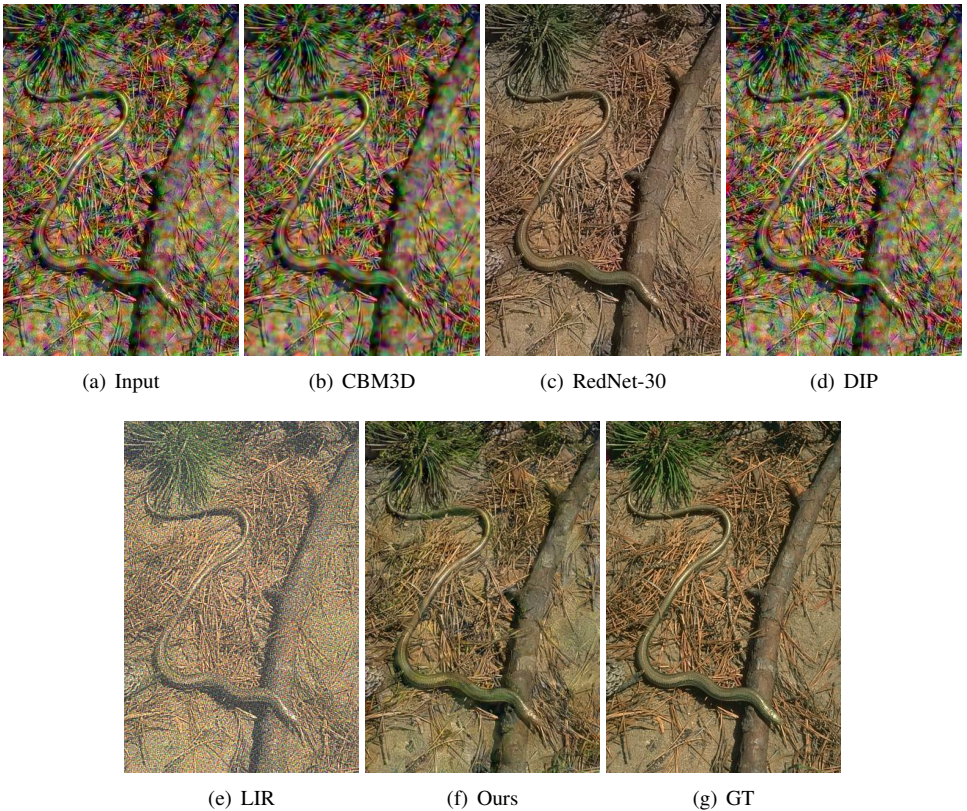


Figure 11: Qualitative results of our method and other baselines on *CBSD68* corrupted by structured noise.

6.2 Poisson Noise

In the comparisons of Poisson noisy images, we use Kodak24 as the test dataset. The images are corrupted by independent Poisson noise from Scikit-image library [13]. We train the

	Traditional	Paired setting	Unpaired setting		
Methods	CBM3D [8]	RedNet-30 [9]	DIP [10]	LIR [11]	Ours
PSNR (dB)	20.62	28.51	20.70	16.90	25.18
SSIM	0.5650	0.9588	0.7239	0.3738	0.9026

Table 2: The average PSNR and SSIM results of different methods on *CBSD68* corrupted by structured noise. Our results are marked in **bold**.

models following the settings in the main paper. The visualized results of Poisson noise removal are given in Figure 12 and 13. Our approach shows impressive noise removal results. While LIR and DIP fail to remove the Poisson noise, our method successfully eliminates the noise and preserves the colors. In Table 3, our method achieves the best performance in PSNR and the second-best performance in terms of SSIM even when it is trained under the unpaired dataset. It demonstrates that our method has robustness and generalization against various noise types. Note that we do not change any hyper-parameters when trained under several types of noise.

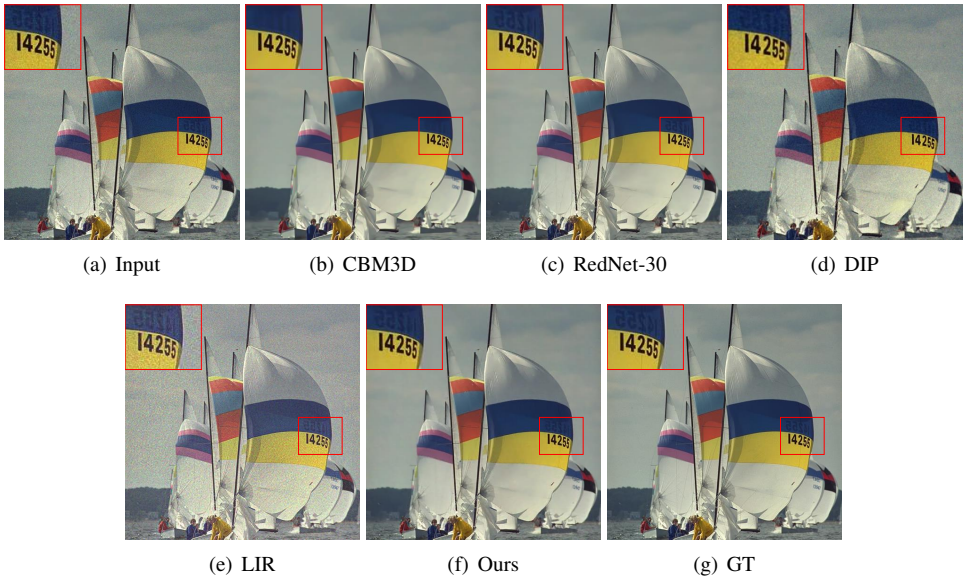


Figure 12: Qualitative results of our method and other baselines on *Kodak24* corrupted by Poisson noise.

	Traditional	Paired setting	Unpaired setting		
Methods	CBM3D [8]	RedNet-30 [9]	DIP [10]	LIR [11]	Ours
PSNR (dB)	32.36	29.59	29.59	26.20	34.93
SSIM	0.8694	0.9778	0.8774	0.7741	0.9691

Table 3: The average PSNR and SSIM results of different methods on *Kodak24* dataset corrupted by Poisson noise. Our results are marked in **bold**.

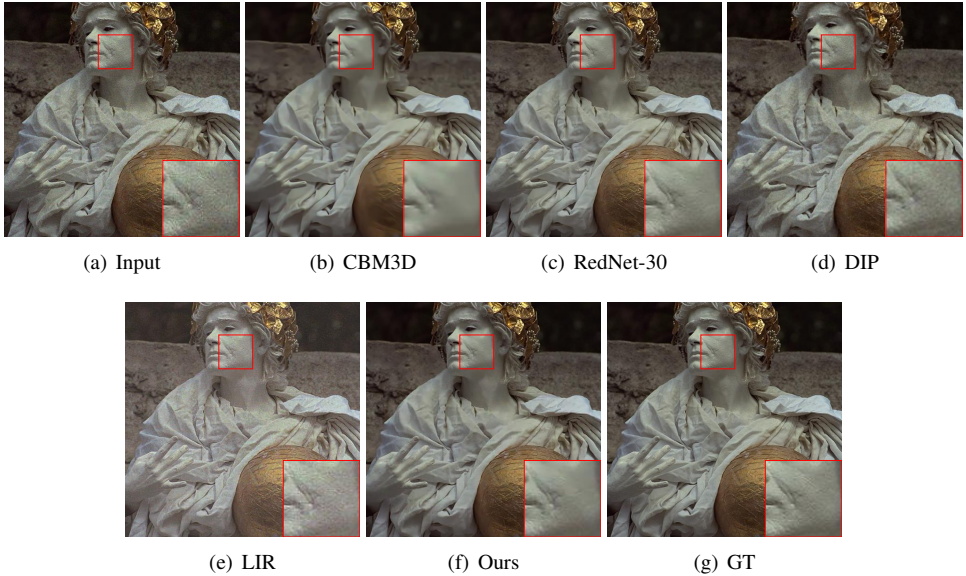


Figure 13: Qualitative results of our method and other baselines on *Kodak24* corrupted by Poisson noise.

References

- [1] Abdelrahman Abdelhamed, Stephen Lin, and Michael S Brown. A high-quality denoising dataset for smartphone cameras. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1692–1700, 2018.
- [2] Namhyuk Ahn, Byungkun Kang, and Kyung-Ah Sohn. Fast, accurate, and lightweight super-resolution with cascading residual network. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 252–268, 2018.
- [3] Kostadin Dabov, Alessandro Foi, Vladimir Katkovnik, and Karen Egiazarian. Image denoising by sparse 3-d transform-domain collaborative filtering. *IEEE Transactions on image processing*, 16(8):2080–2095, 2007.
- [4] Wenchao Du, Hu Chen, and Hongyu Yang. Learning invariant representation for unsupervised image restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 14483–14492, 2020.
- [5] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.
- [6] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros. Image-to-image translation with conditional adversarial networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1125–1134, 2017.
- [7] Yoonsik Kim, Jae Woong Soh, Gu Yong Park, and Nam Ik Cho. Transfer learning from synthetic to real-noise denoising with adaptive instance normalization. In *Pro-*

- ceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3482–3492, 2020.
- [8] Xiao-Jiao Mao, Chunhua Shen, and Yu-Bin Yang. Image restoration using very deep convolutional encoder-decoder networks with symmetric skip connections. *arXiv preprint arXiv:1603.09056*, 2016.
 - [9] David Martin, Charless Fowlkes, Doron Tal, and Jitendra Malik. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In *Proceedings Eighth IEEE International Conference on Computer Vision. ICCV 2001*, volume 2, pages 416–423. IEEE, 2001.
 - [10] Takeru Miyato, Toshiki Kataoka, Masanori Koyama, and Yuichi Yoshida. Spectral normalization for generative adversarial networks. *arXiv preprint arXiv:1802.05957*, 2018.
 - [11] Taylor R Moen, Baiyu Chen, David R Holmes III, Xinhui Duan, Zhicong Yu, Lifeng Yu, Shuai Leng, Joel G Fletcher, and Cynthia H McCollough. Low-dose ct image and projection dataset. *Medical physics*, 48(2):902–911, 2021.
 - [12] Dmitry Ulyanov, Andrea Vedaldi, and Victor Lempitsky. Deep image prior. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 9446–9454, 2018.
 - [13] Stefan Van der Walt, Johannes L Schönberger, Juan Nunez-Iglesias, François Boulogne, Joshua D Warner, Neil Yager, Emmanuelle Gouillart, and Tony Yu. scikit-image: image processing in python. *PeerJ*, 2:e453, 2014.