Unsupervised Image Denoising with Frequency Domain Knowledge

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

Abstract

Supervised learning-based methods yield robust denoising results, yet they are inherently limited by the need for large-scale clean/noisy paired datasets. The use of unsupervised denoisers, on the other hand, necessitates a more detailed understanding of the underlying image statistics. In particular, it is well known that apparent differences between clean and noisy images are most prominent on high-frequency bands, justifying the use of low-pass filters as part of conventional image preprocessing steps. However, most learning-based denoising methods utilize only one-sided information from the spatial domain without considering frequency domain information. To address this limitation, in this study we propose a frequency-sensitive unsupervised denoising method. To this end, a generative adversarial network (GAN) is used as a base structure. Subsequently, we include spectral discriminator and frequency reconstruction loss to transfer frequency knowledge into the generator. Results using natural and synthetic datasets indicate that our unsupervised learning method augmented with frequency information achieves state-of-the-art denoising performance, suggesting that frequency domain information could be a viable factor in improving the overall performance of unsupervised learning-based methods.

1 Introduction

Based on clean and noisy image pairs, supervised learning-based image denoisers have shown impressive performance compared to prior-based approaches. A large number of high-quality image pairs play an important role in the performance of supervised learning-based methods. However, constructing large-scale paired datasets may be unavailable or

* The authors contributed equally.
Github: https://github.com/jdg900/UID-FDK

expensive in real-world situations. For this reason, image denoising methods that do not require clean and noisy image pairs have recently drawn attention.

A noisy image x is usually modeled as the sum of clean background y and noise n: $x = y + n$. Subsequently, noise corrupts the benign pixels, which makes it hard to distinguish the pixels of noise and content in the spatial domain. However, in the frequency domain, noise and content can be easily identified. As shown in Figure 1 (a), we observe that the noise lies in the high-frequency bands and semantic information lies in the low-frequency bands. Furthermore, in Figure 1 (b), we note that apparent differences between clean and noisy images are most prominent on high-frequency bands. It may indicate that the frequency domain provides useful evidence for noise removal. However, the recent learning-based denoisers overlook the frequency domain information and use only one-sided information from the spatial domain.

Motivated by these observations, we propose the unsupervised denoising method that reflects frequency domain information. Specifically, with a generative adversarial network as a base structure, we introduce the spectral discriminator and frequency reconstruction loss to transfer frequency knowledge to the generator. The spectral discriminator distinguishes the differences between denoised and clean images on high-frequency bands. By propagating this knowledge to the generator for noise removal, the generator considers the frequency domain and thus produces visually more plausible denoised images to fool the spectral discriminator. The frequency reconstruction loss, combined with the cycle consistency loss, improves the image quality and preserves the content of images while narrowing the gap between clean and denoised images in the frequency domain.

The main contributions of our method are summarized as follows: 1) We propose the GAN-based unsupervised image denoising method that preserves semantic information and produces a high-quality noise-free image. 2) To the best of our knowledge, it is the first approach to explore the potential of the frequency domain with Fourier transform in the field of noise removal tasks. The proposed spectral discriminator and frequency reconstruction loss make the generator concentrate on the noise and produce satisfying results. Denoised images recovered by our method are close to clean reference images in both spatial and frequency domain. 3) The proposed method outperforms existing unsupervised image denoisers by a considerable margin. Moreover, our performance is even comparable with supervised learning-based approaches trained with paired datasets.

2 Related Work

2.1 Image Denoising

Non-learning based image denoisers [2, 3, 9, 16, 22, 27, 34, 35, 36, 44] have tried to reconstruct clean images using pre-defined priors which model the distribution of noise. Specifically, a widely used prior in image denoising is non-local self-similarity prior [3, 9, 16, 35]. Assuming that similar patches exist in a single image, the methods based on non-local self-similarity [3, 9] remove the noise using these patches.

Recently, with the advent of deep neural networks, supervised learning-based image denoisers [23, 40, 41] show promising performance on a set of clean and noisy image pairs. However, it is challenging to construct clean and noisy image pairs in a real-world scenario. To address the above issues, denoisers that do not rely on clean and noisy image pairs have been proposed [7, 11, 20, 21, 30]. N2N [21] learns reconstruction using only noisy im-
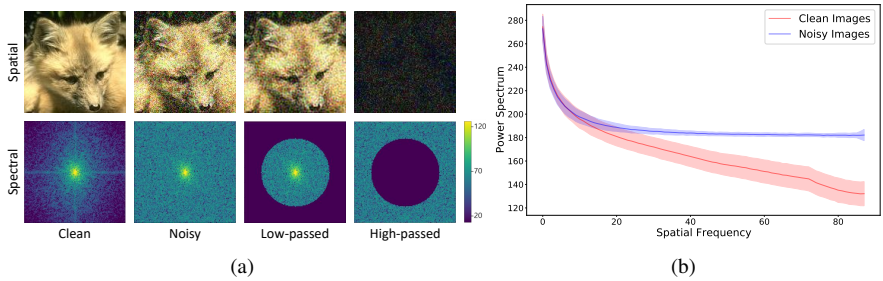
Figure 1: The spectrum analysis in the frequency domain. (a) Visualization of images in the spatial domain and corresponding spectrum maps in the frequency domain. (b) The statistics (mean and variance) after azimuthal integral over the power spectrum on clean and noisy images of *CBSD68*. We use AWGN with a noise level $\sigma = 50$ to yield noisy images.

age pairs without ground-truth clean images. N2V [20] estimates a corrupted pixel from its neighboring pixels based on a blind-spot mechanism. GCBD [7] generates the noisy images while modeling the real-world noise distribution through the GAN [15] and trains the denoiser with pseudo clean and noisy image pairs. LIR [11] trains an image denoiser by disentangling invariant representations from noisy images with an unpaired dataset.

2.2 Frequency Domain in CNNs

In traditional image processing, analyzing images in the frequency domain is known to be effective by transforming the image from the spatial domain to the frequency domain. Inspired by this idea, several works attempt to utilize the information from the frequency domain in deep neural networks. Xu *et al*. [37] accelerate the training of neural networks utilizing the discrete cosine transform. Dzanic *et al*. [13] observe that discrepancy exists between the images generated by the GAN [15] and the real images through the analysis of high-frequency Fourier modes. In addition, attempts to utilize the frequency domain information in the various fields, including image forensics [12, 14, 42], image generation [4, 8, 17], and domain adaptation [38, 39] are gradually increasing. However, image denoising methods combining the frequency domain analysis with DNN remain much less explored.

3 Method

In this section, we first introduce the spectral discriminator and frequency reconstruction loss that use information from the frequency domain. Then, we present an unsupervised framework for image denoising, integrating the proposed discriminator and loss with the GAN. The proposed framework is illustrated in Figure 2.

3.1 Frequency Domain Constraints

Spectral Discriminator The simple way for the generator to consider the frequency domain is that the discriminator transfers the frequency domain knowledge to the generator. To this end, we propose the spectral discriminator similar to that introduced by [8] to measure
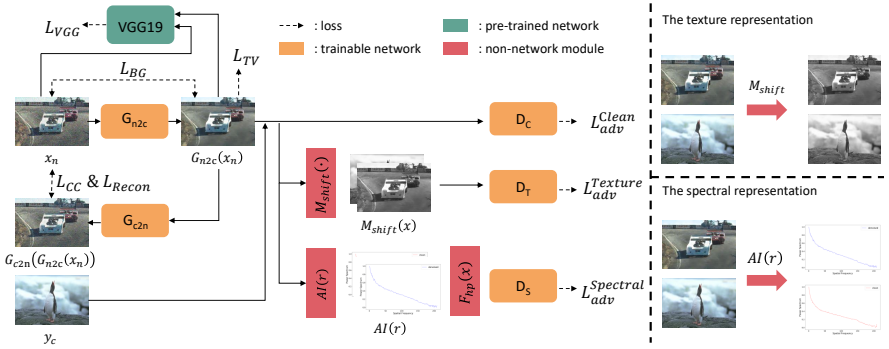
Figure 2: An overview of the proposed framework. Given an unpaired clean y_c and noisy image x_n, the generator G_{n2c} for image denoising takes the noisy image x_n as an input and learns the mapping for noise removal. Additional network G_{c2n} is used to impose the cycle consistency. Three discriminators D_C, D_T, and D_S try to distinguish the denoised image $G_{n2c}(x_n)$ from real clean image x_c in terms of both spatial domain and frequency domain. The whole framework is end-to-end trainable.

spectral realness. We compute the discrete Fourier transform on 2D image data $f(w,h)$ in size $W \times H$ to feed frequency representations to the discriminator.

$$F(k,l) = \sum_{w=0}^{W-1} \sum_{h=0}^{H-1} f(w,h)e^{-2\pi i\frac{kw}{W}}e^{-2\pi i\frac{lh}{H}} \tag{1}$$

for spectral coordinates $k = 0,...,W-1$ and $l = 0,...,H-1$.

Recent studies [8, 12] show that the 1D representation of the Fourier power spectrum is sufficient to highlight spectral differences. Following their works, we transform the result of Fourier transform to polar coordinate and compute azimuthal integration over θ.

$$F(r,\theta) = F(k,l) : r = \sqrt{k^2 + l^2}, \quad \theta = \arctan\frac{l}{k}, \quad AI(r) = \frac{1}{2\pi}\int_0^{2\pi}|F(r,\theta)|d\theta \tag{2}$$

where $AI(r)$ means the average intensity of the image signal about radial distance r.

We propose the spectral discriminator that allows the generator to focus on noise using high-frequency spectral information. To learn the differences on high-frequency bands, we pass the 1D spectral vector into the high-pass filter F_{hp} and input it to the spectral discriminator.

$$v_I = F_{hp}(AI(r)), \quad F_{hp}(x) = \begin{cases} x, & r > r_\tau, \\ 0, & otherwise \end{cases} \tag{3}$$

where r_τ is a threshold radius for high-pass filtering and v_I is a high-pass filtered 1D spectral vector of an input I.

Generally, the most distinct characteristics between clean and noisy images exist on high-frequency bands. Thus, if there is some remained noise on denoised images, the spectral discriminator easily distinguishes the difference between the clean and denoised images on high-frequency bands. By transferring this knowledge to the generator, the generator for noise removal learns to yield visually more plausible images to fool the spectral discriminator.

Frequency Reconstruction Loss Cai *et al.* [4] demonstrate the existence of a gap between the real and generated image in the frequency domain, which leads to artifacts in the spatial domain. Motivated by this observation, we propose to use frequency reconstruction loss with cycle consistency loss to ameliorate the quality of denoised images while reducing the gap. We aim that the frequency reconstruction loss which is complementary to cycle consistency loss enables the generator to consider the frequency domain. Furthermore, we expect that it can serve as an assistant in generating high-quality denoised images. To compute the frequency reconstruction loss, we map an input x_n and reconstructed image $G_{c2n}(G_{n2c}(x_n))$ to the frequency domain using Fourier transform. Then, we calculate the frequency reconstruction loss by measuring the difference between the two results of the Fourier transform and taking a logarithm to normalize it. Finally, we minimize the following objective:

$$L_{Freq} = log(1 + \frac{1}{WH} \sum_{k=0}^{W-1} \sum_{l=0}^{H-1} |F_{x_n}(k,l) - F_{G_{c2n}(G_{n2c}(x_n))}(k,l)|) \tag{4}$$

3.2 Unsupervised Framework for Image Denoising

Our goal is to learn a mapping from a noise domain X_N to a clean domain Y_C given unpaired training images $x_n \in X_N$ and $y_c \in Y_C$. To learn this mapping, we use the CycleGAN-like framework consisting of two generators, G_{n2c} and G_{c2n}, and three discriminators, D_C, D_T, and D_S. Given a noisy image x_n, the generator G_{n2c} learns to generate a denoised image $G_{n2c}(x_n)$. While distinguishing the denoised image $G_{n2c}(x_n)$ from the real clean image y_c, the discriminator D_C makes the generator produce the denoised images closer to the real clean domain Y_C. To stablize training, we use the Least Squares GAN (LSGAN) loss [24] for adversarial loss. The LSGAN loss for G_{n2c} and D_C is:

$$L_{adv}^{Clean} = E_{y_c \sim P_c}[(D_C(y_c))^2] + E_{x_n \sim P_n}[(1 - D_C(G_{n2c}(x_n)))^2] \tag{5}$$

where P_n and P_c are the data distributions of the domain X_N and domain Y_C, respectively.

As introduced in [31], we adopt the texture discriminator D_T in order to guide the generator to produce clean contour and preserve texture while removing the noise. Following the scheme of [31], a random color shift algorithm M_{shift} is applied to the denoised image $G_{n2c}(x_n)$. The texture loss for G_{n2c} and D_T is:

$$L_{adv}^{Texture} = E_{y_c \sim P_c}[(D_T(M_{shift}(y_c)))^2] + E_{x_n \sim P_n}[(1 - D_T(M_{shift}(G_{n2c}(x_n))))^2] \tag{6}$$

As discussed in Section 3.1, we use the spectral discriminator D_S to guide the generator to generate more realistic images by reducing the gap between the clean and denoised image in the frequency domain. The spectral loss for G_{n2c} and D_S is:

$$L_{adv}^{Spectral} = E_{y_c \sim P_c}[(D_S(v_{y_c}))^2] + E_{x_n \sim P_n}[(1 - D_S(v_{G_{n2c}(x_n)}))^2] \tag{7}$$

where v denotes the high-pass filtered 1D spectral vector in Eq. 3.

CycleGAN [43] imposes the two-sided cycle consistency constraint to learn the one-to-one mappings between two domains. On the other hand, we use only one-sided cycle consistency to maintain the content between noisy and denoised images. By incorporating a network G_{c2n}, we let $G_{c2n}(G_{n2c}(x_n))$ be identical to the noisy image x_n. The cycle consistency loss is expressed as:

$$L_{CC} = ||x_n - G_{c2n}(G_{n2c}(x_n))||_1 \tag{8}$$

where $||\cdot||_1$ is the L1 norm.

Furthermore, we add the reconstruction loss between the $G_{c2n}(G_{n2c}(x_n))$ and x_n to stabilize the training. We employ the negative SSIM loss [32] and combine it with the frequency reconstruction loss L_{Freq} in Eq. 4. The reconstruction loss is expressed as:

$$L_{Recon} = L_{Freq}(x_n, G_{c2n}(G_{n2c}(x_n))) + L_{SSIM}(x_n, G_{c2n}(G_{n2c}(x_n))) \qquad (9)$$

where $L_{SSIM}(a,b)$ denotes the negative SSIM loss, $-SSIM(a,b)$.

To impose the local smoothness and mitigate the artifacts in the restored image, we adopt the total variation loss [6]. The total variation loss is expressed as:

$$L_{TV} = \sum_{w,h}(||\nabla_w G_{n2c}(x_n)||_2 + ||\nabla_h G_{n2c}(x_n)||_2) \qquad (10)$$

where $||\cdot||_2$ denotes the L2 norm, ∇_w and ∇_h are the operations to compute the gradients in terms of horizontal and vertical directions, respectively.

Inspired by [11, 51], we use the perceptual loss [18] to ensure that extracted features from the noisy and denoised image are semantically invariant. This allows the image to keep its semantics even after the noise has been removed. The perceptual loss is expressed as:

$$L_{VGG} = ||\phi_l(x_n) - \phi_l(G_{n2c}(x_n))||_2 \qquad (11)$$

where $\phi_l(\cdot)$ denotes the pre-trained VGG-19 [29] on ImageNet [10], l denotes lth layer from VGG-19, and we use the *Conv5-4* layer of VGG-19 model in our experiments.

Moreover, we employ the background loss to preserve background consistency between the noisy and denoised image. The background loss constrains the L1 norm between blurred results of the noisy and denoised image. As a blur operator, we adopt a guided filter [13] that smooths the image while preserving the sharpness such as edges and details. The background loss is expressed as:

$$L_{BG} = ||GF(x_n) - GF(G_{n2c}(x_n))||_1 \qquad (12)$$

where $GF(\cdot)$ denotes the guided filter.

Our full objective for the two generators and the three discriminators is expressed as:

$$\min_{G_{n2c}, G_{c2n}} \max_{D_C, D_T, D_S} L_{adv}^{Clean} + L_{adv}^{Texture} + L_{adv}^{Spectral} + L_{CC} +$$
$$\lambda_{VGG}L_{VGG} + \lambda_{BG}L_{BG} + \lambda_{TV}L_{TV} + \lambda_{Recon}L_{Recon} \qquad (13)$$

We empirically define the weights in the full objective as: $\lambda_{VGG} = 2$, $\lambda_{BG} = 2$, $\lambda_{TV} = 0.2$, and $\lambda_{Recon} = 0.2$.

4 Experiment

In this section, we provide the implementation details of the proposed method. Then, we present extensive experiments on synthetic and real-world noisy images. Lastly, we conduct an ablation study to show the effectiveness of the proposed method. For synthetic noise, we use Additive White Gaussian Noise (AWGN) to synthesize the noisy images. We adopt the CBSD68 [25] for evaluation. For real noise, we use the Low-Dose Computed Tomography dataset [26] and real photographs SIDD [1] to demonstrate the generalization capacity of the proposed method. We employ PSNR and SSIM [32] to evaluate the results.
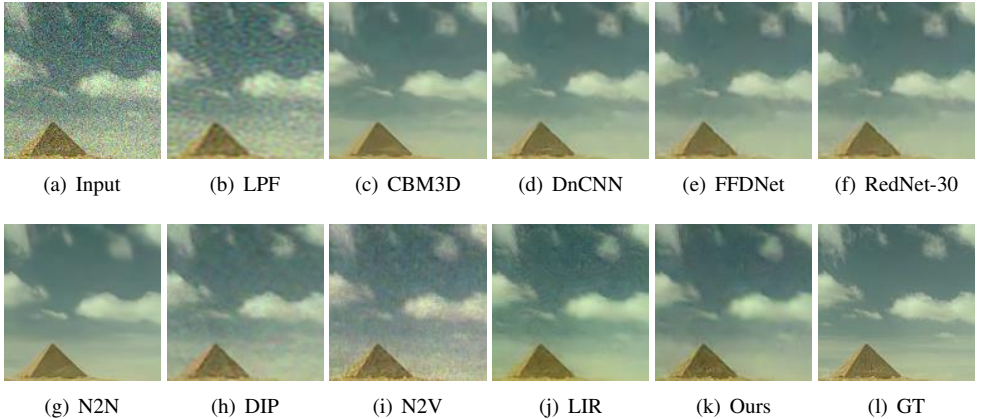
(a) Input   (b) LPF   (c) CBM3D   (d) DnCNN   (e) FFDNet   (f) RedNet-30

(g) N2N   (h) DIP   (i) N2V   (j) LIR   (k) Ours   (l) GT

Figure 3: Qualitative results of our method and other baselines on *CBSD68* corrupted by AWGN with a noise level $\sigma = 25$.

Methods	Traditional		Paired setting				Unpaired setting			
	LPF	CBM3D [8]	DnCNN [40]	FFDNet [41]	RedNet-30 [24]	N2N [21]	DIP [30]	N2V [20]	LIR [15]	Ours
Noise level	PSNR (dB)									
$\sigma = 15$	25.93	33.55	33.72	29.68	33.60	33.92	28.51	28.66	30.44	**32.21**
$\sigma = 25$	24.61	30.91	30.85	28.71	30.68	31.31	27.26	27.20	29.08	**29.37**
$\sigma = 50$	21.49	27.47	27.19	26.79	26.42	28.10	23.66	24.52	25.69	**26.03**
Noise level	SSIM									
$\sigma = 15$	0.7079	0.9619	0.9254	0.8616	0.9620	0.9301	0.8851	0.9024	0.9414	**0.9502**
$\sigma = 25$	0.6102	0.9331	0.8724	0.8254	0.9308	0.8857	0.8613	0.8684	0.9126	**0.9124**
$\sigma = 50$	0.4266	0.8722	0.7490	0.7463	0.8502	0.7973	0.7510	0.7927	0.8435	**0.8375**

Table 1: The average PSNR and SSIM results of our method and other baselines on *CBSD68* corrupted by AWGN with noise levels $\sigma = \{15, 25, 50\}$. Our results are marked in **bold**.

4.1 Implementation Details

We implement our method with Pytorch [28]. The generator and discriminator architectures are detailed in the supplementary material. We train our method up to 100 epochs on Nvidia TITAN RTX GPU and RTX A6000 in experiments. We adopt ADAM [19] for optimization. The initial learning rate is set to 0.0001, and we keep the same learning rate for the first 70 epochs and linearly decay the rate to zero over the last 30 epochs. We set the batch size to 16 in all experiments. We randomly crop 128×128 patches for synthetic noise removal and use input patches of size 256×256 for real-world noise removal. We randomly flip the images horizontally for data augmentation. For high-pass filter on spectral discriminator, r_τ is set to $\lfloor H/2\sqrt{2} \rfloor$ where H is the height of an image and $\lfloor \rfloor$ is a floor operator. Loss weights are described in Section 3.2. Our model is evaluated with three random seeds, and we report its average values for rigorous evaluation.

4.2 Synthetic Noise Removal

We train the model with DIV2K [25] that contains 800 images with 2K resolution. For the unpaired training, we randomly divide the dataset into two parts without intersection. To construct a noise set, we add the AWGN with noise levels $\sigma = \{15, 25, 50\}$ to images in one part using the other part as a clean set. For a fair comparison, we use only the noise set and their corresponding ground-truth when training other supervised learning-based methods.

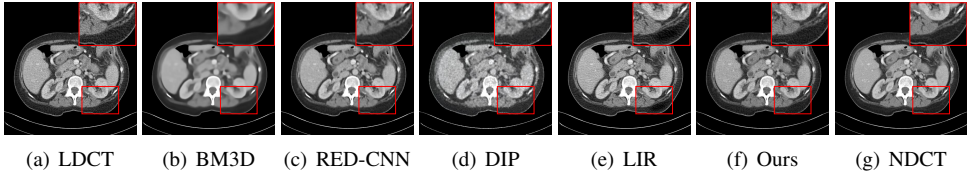| (a) LDCT | (b) BM3D | (c) RED-CNN | (d) DIP | (e) LIR | (f) Ours | (g) NDCT |

Figure 4: Qualitative results of our method and other baselines on *Mayo Clinic Low Dose CT dataset*. (a) Real low-dose. (b)-(f) Results of each methods. (g) Real normal-dose. As shown in the highlighted red box, the reconstructed image by our method has few noise and artifacts. The display window is $[160, 240]$ HU.

	Traditional	Paired setting	Unpaired setting		
Methods	BM3D [9]	RED-CNN [6]	DIP [30]	LIR [11]	Ours
PSNR (dB)	29.16	29.39	26.97	27.26	**30.11**
SSIM	0.8514	0.9078	0.8267	0.8452	**0.8728**

Table 2: The average PSNR and SSIM results of different methods on *Mayo Clinic Low Dose CT dataset*. Our results are marked in **bold**.

We select unsupervised methods, i.e. DIP [30], N2N [21], N2V [20], and LIR [11], and supervised methods, i.e. DnCNN [40], FFDNet [41], and RedNet-30 [23], to compare the performance. Traditional Low-Pass Filtering (LPF) and BM3D [9] are also evaluated. As shown in Figure 3, the unsupervised methods tend to shift the color and leave apparent visual artifacts in the sky. Especially, LIR removes the noise but fails to preserve the texture. With frequency domain information, our method successfully eliminates noise and preserves the texture. The classical LPF using Fourier transform alleviates the noise, but our framework that reflects not only the frequency domain knowledge but also spatial domain knowledge shows superior results. As shown in Table 1, our model outperforms other unsupervised methods, i.e. DIP, N2V, and LIR, by at least +0.29 dB in PSNR. Although our model is trained on unpaired images, it achieves superior performance in the SSIM than DnCNN and FFDNet trained on paired datasets. We conjecture that the reason for better noise removal is the use of the extra domain information that other previous methods do not consider.

4.3 Real-World Noise Removal

In this section, we evaluate the generalization ability of the proposed method on real-world noise, i.e. Low-Dose Computed Tomography (CT) and real photographs. For the comparison of the Low-Dose CT, we adopt BM3D [9], DIP [30], RED-CNN [6], and LIR [11] as baselines. For the comparison of the real photographs, BM3D [9], DIP [30], RedNet-30 [23], and LIR [11] are selected as baselines.

Denoising on Low-Dose CT Since Computed Tomography (CT) helps to diagnose abnormalities of organs, CT is widely used in medical analysis. Reducing the radiation dose in order to decrease health risks causes noise and artifacts in the reconstructed images. Like the real-world noise, the noise distributions of the reconstructed image are difficult to model analytically. Therefore, we adopt a CT dataset authorized by Mayo Clinic [26] to evaluate the generalization ability of our method on real-world noise. Mayo Clinic dataset consists of paired normal-dose and lose-dose CT images for each patient. The Normal-Dose CT
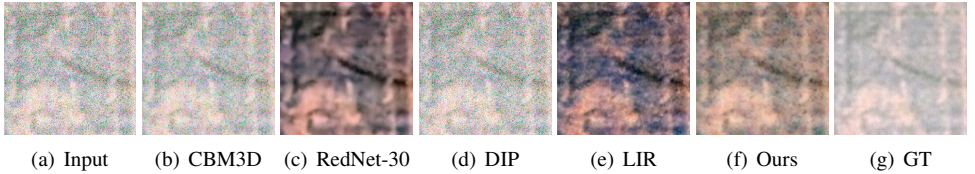
| (a) Input | (b) CBM3D | (c) RedNet-30 | (d) DIP | (e) LIR | (f) Ours | (g) GT |

Figure 5: Qualitative results of our method and other baselines on *SIDD*.

	Traditional	Paired setting	Unpaired setting		
Methods	CBM3D [■]	RedNet-30 [■]	DIP [■]	LIR [■]	Ours
PSNR (dB)	28.32	38.02	24.68	33.79	**34.30**
SSIM	0.6784	0.9619	0.5901	0.9466	**0.9334**

Table 3: The average PSNR and SSIM results of different methods on *SIDD*. Our results are marked in **bold**.

(NDCT) and the Low-Dose CT (LDCT) images correspond to clean and noisy images, respectively. For the training, we obtain 2,850 images in 512×512 resolution from 20 different patients. We construct 1,422 LDCT images from randomly selected 10 patients as a noise set and 1,428 NDCT images from the remaining patients as a clean set for unpaired training. For the test, we obtain 865 images from 5 different patients. As shown in Table 2, our method achieves the best and the second-best performance in PSNR and SSIM, respectively. Note that our model trained on the unpaired dataset outperforms the RED-CNN trained on the paired dataset in PSNR. It indicates that our method can be more practical in medical analysis where obtaining paired datasets is challenging. We also compare the qualitative results with other baselines. As shown in Figure 4, other methods tend to generate artifacts or lose details. On the other hand, our method shows a reasonable balance between noise removal and image quality. More qualitative results are provided in the supplementary material.

Denoising on Real Photographs To demonstrate the effectiveness of our method on real noisy photographs, we evaluate our method on SIDD [■] which is obtained from smartphone cameras. Because the images of the SIDD comprise various noise levels and brightness, this dataset is the best appropriate to validate the generalization capacity of the denoisers. The SIDD includes 320 pairs of noisy images and corresponding clean images with 4K or 5K resolutions for the training. For the unpaired training, we divide the dataset into 160 clean and 160 noisy images without intersection. The other training settings are the same as implementation details. For evaluation, we use 1280 cropped patches of size 256×256 in the SIDD validation set. As show in Figure 5, other baselines tend to leave the noise or fail to preserve the color of images. In contrast, our method removes the intense noise while keeping the color compared to other baselines. We also report the quantitative results in Table 3. More qualitative results are provided in the supplementary material.

4.4 Ablation Study

We conduct an ablation study to demonstrate the validity of our key components: the texture discriminator D_T, the spectral discriminator D_S, and the frequency reconstruction loss L_{Freq}. We employ an additional evaluation metric LFD [■] to measure the difference between denoised images and reference images in the frequency domain. The small LFD value indicates

D_S	D_T	L_{Freq}	PSNR (dB)	SSIM	LFD
✗	✗	✗	25.59	0.8290	6.5955
✓	✗	✗	25.79	0.8304	**6.5649**
✓	✓	✗	25.82	0.8334	6.5874
✓	✓	✓	**26.03**	**0.8375**	6.5795

Table 4: Ablation study. Quantitative results of our method with and without the texture discriminator D_T, spectral discriminator D_S, and frequency reconstruction loss L_{Freq} on CBSD68 corrupted by AWGN with a noise level $\sigma = 50$. we report the PSNR, SSIM (higher is better) and LFD (lower is better). The best results are marked in **bold**.

that the denoised images are close to the reference images. First, to verify the effectiveness of the D_S, we only add the D_S to the base structure. As shown in Table 4, when the D_S is integrated, both PSNR and SSIM increase by 0.2 dB and 0.0014, respectively. It demonstrates that the spectral discriminator leads the generator to remove high-frequency related noise effectively by transferring the difference between noisy and clean images on the high-frequency bands. Also, we see that the spectral discriminator makes the denoised images close to clean domain images in the frequency domain, resulting in the decrease of LFD. Next, to verify the effectiveness of the D_T, we integrate it with the D_S. Distinguishing the texture representations helps restore clean contours and fine details related to image quality, which improves the SSIM metric. A curious phenomenon is that the texture discriminator increases the LFD. We conjecture that the introduction of D_T causes a bias to the spatial domain in maintaining the balance between the spatial and frequency domains, thus increasing the distance in the frequency domain. Adding the L_{Freq} shows results validating our hypothesis that narrowing the gap in the frequency domain is crucial to generate the high-quality denoised image. In addition, through the decrease of LFD, the frequency reconstruction loss may help to maintain the balance between the spatial and frequency domain.

5 Conclusion

In this paper, we propose an unsupervised learning-based image denoiser that enables the image denoising without clean and noisy image pairs. To the best of our knowledge, it is the first approach that aims to recover a noise-free image from a corrupted image using frequency domain information. To this end, we introduce the spectral discriminator and frequency reconstruction loss that can propagate the frequency knowledge to the generator. By reflecting the information from the frequency domain, our method successfully focuses on high-frequency components to remove noise. Experiments on synthetic and real noise removal show that our method outperforms other unsupervised learning-based denoisers and generates more visually pleasing images with fewer artifacts. We believe that considering the frequency domain can be advantageous in other low-level vision tasks as well.

6 Acknowledgements

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] Michal Aharon, Michael Elad, and Alfred Bruckstein. K-svd: An algorithm for designing overcomplete dictionaries for sparse representation. *IEEE Transactions on signal processing*, 54(11):4311–4322, 2006.

[3] Antoni Buades, Bartomeu Coll, and J-M Morel. A non-local algorithm for image denoising. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, volume 2, pages 60–65. IEEE, 2005.

[4] Mu Cai, Hong Zhang, Huijuan Huang, Qichuan Geng, and Gao Huang. Frequency domain image translation: More photo-realistic, better identity-preserving. *arXiv preprint arXiv:2011.13611*, 2020.

[5] Antonin Chambolle. An algorithm for total variation minimization and applications. *Journal of Mathematical imaging and vision*, 20(1):89–97, 2004.

[6] Hu Chen, Yi Zhang, Mannudeep K Kalra, Feng Lin, Yang Chen, Peixi Liao, Jiliu Zhou, and Ge Wang. Low-dose ct with a residual encoder-decoder convolutional neural network. *IEEE transactions on medical imaging*, 36(12):2524–2535, 2017.

[7] Jingwen Chen, Jiawei Chen, Hongyang Chao, and Ming Yang. Image blind denoising with generative adversarial network based noise modeling. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3155–3164, 2018.

[8] Yuanqi Chen, Ge Li, Cece Jin, Shan Liu, and Thomas Li. Ssd-gan: Measuring the realness in the spatial and spectral domains. *arXiv preprint arXiv:2012.05535*, 2020.

[9] 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.

[10] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pages 248–255. Ieee, 2009.

[11] 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.

[12] Ricard Durall, Margret Keuper, and Janis Keuper. Watch your up-convolution: Cnn based generative deep neural networks are failing to reproduce spectral distributions. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7890–7899, 2020.

[13] Tarik Dzanic, Karan Shah, and Freddie Witherden. Fourier spectrum discrepancies in deep network generated images. *arXiv preprint arXiv:1911.06465*, 2019.

[14] Joel Frank, Thorsten Eisenhofer, Lea Schönherr, Asja Fischer, Dorothea Kolossa, and Thorsten Holz. Leveraging frequency analysis for deep fake image recognition. In *International Conference on Machine Learning*, pages 3247–3258. PMLR, 2020.

[15] Ian J Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial networks. *arXiv preprint arXiv:1406.2661*, 2014.

[16] Shuhang Gu, Lei Zhang, Wangmeng Zuo, and Xiangchu Feng. Weighted nuclear norm minimization with application to image denoising. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2862–2869, 2014.

[17] Liming Jiang, Bo Dai, Wayne Wu, and Chen Change Loy. Focal frequency loss for generative models. *arXiv preprint arXiv:2012.12821*, 2020.

[18] Justin Johnson, Alexandre Alahi, and Li Fei-Fei. Perceptual losses for real-time style transfer and super-resolution. In *European conference on computer vision*, pages 694–711. Springer, 2016.

[19] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[20] Alexander Krull, Tim-Oliver Buchholz, and Florian Jug. Noise2void-learning denoising from single noisy images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2129–2137, 2019.

[21] Jaakko Lehtinen, Jacob Munkberg, Jon Hasselgren, Samuli Laine, Tero Karras, Miika Aittala, and Timo Aila. Noise2noise: Learning image restoration without clean data. *arXiv preprint arXiv:1803.04189*, 2018.

[22] Julien Mairal, Francis Bach, Jean Ponce, Guillermo Sapiro, and Andrew Zisserman. Non-local sparse models for image restoration. In *2009 IEEE 12th international conference on computer vision*, pages 2272–2279. IEEE, 2009.

[23] 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.

[24] Xudong Mao, Qing Li, Haoran Xie, Raymond YK Lau, Zhen Wang, and Stephen Paul Smolley. Least squares generative adversarial networks. In *Proceedings of the IEEE international conference on computer vision*, pages 2794–2802, 2017.

[25] 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.

[26] 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.

[27] Stanley Osher, Martin Burger, Donald Goldfarb, Jinjun Xu, and Wotao Yin. An iterative regularization method for total variation-based image restoration. *Multiscale Modeling & Simulation*, 4(2):460–489, 2005.

[28] Adam Paszke, Sam Gross, Soumith Chintala, Gregory Chanan, Edward Yang, Zachary DeVito, Zeming Lin, Alban Desmaison, Luca Antiga, and Adam Lerer. Automatic differentiation in pytorch. 2017.

[29] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.

[30] 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.

[31] Xinrui Wang and Jinze Yu. Learning to cartoonize using white-box cartoon representations. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8090–8099, 2020.

[32] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4):600–612, 2004.

[33] Huikai Wu, Shuai Zheng, Junge Zhang, and Kaiqi Huang. Fast end-to-end trainable guided filter. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1838–1847, 2018.

[34] Jinjun Xu and Stanley Osher. Iterative regularization and nonlinear inverse scale space applied to wavelet-based denoising. *IEEE Transactions on Image Processing*, 16(2): 534–544, 2007.

[35] Jun Xu, Lei Zhang, David Zhang, and Xiangchu Feng. Multi-channel weighted nuclear norm minimization for real color image denoising. In *Proceedings of the IEEE international conference on computer vision*, pages 1096–1104, 2017.

[36] Jun Xu, Lei Zhang, and David Zhang. A trilateral weighted sparse coding scheme for real-world image denoising. In *Proceedings of the European conference on computer vision (ECCV)*, pages 20–36, 2018.

[37] Kai Xu, Minghai Qin, Fei Sun, Yuhao Wang, Yen-Kuang Chen, and Fengbo Ren. Learning in the frequency domain. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1740–1749, 2020.

[38] Yanchao Yang and Stefano Soatto. Fda: Fourier domain adaptation for semantic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4085–4095, 2020.

[39] Yanchao Yang, Dong Lao, Ganesh Sundaramoorthi, and Stefano Soatto. Phase consistent ecological domain adaptation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9011–9020, 2020.

[40] Kai Zhang, Wangmeng Zuo, Yunjin Chen, Deyu Meng, and Lei Zhang. Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising. *IEEE transactions on image processing*, 26(7):3142–3155, 2017.

[41] Kai Zhang, Wangmeng Zuo, and Lei Zhang. Ffdnet: Toward a fast and flexible solution for cnn-based image denoising. *IEEE Transactions on Image Processing*, 27(9):4608–4622, 2018.

[42] Xu Zhang, Svebor Karaman, and Shih-Fu Chang. Detecting and simulating artifacts in gan fake images. In *2019 IEEE International Workshop on Information Forensics and Security (WIFS)*, pages 1–6. IEEE, 2019.

[43] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proceedings of the IEEE international conference on computer vision*, pages 2223–2232, 2017.

[44] Daniel Zoran and Yair Weiss. From learning models of natural image patches to whole image restoration. In *2011 International Conference on Computer Vision*, pages 479–486. IEEE, 2011.