

Unsupervised Image Denoising in Real-World Scenarios via Self-Collaboration Parallel Generative Adversarial Branches

Xin Lin

Sichuan University

linxin@stu.scu.edu.cn

Chao Ren*

Sichuan University

chaoren@scu.edu.cn

Xiao Liu

Sichuan University

liux@stu.scu.edu.cn

Jie Huang

Sichuan University

huangjiechn@stu.scu.edu.cn

Yinjie Lei

Sichuan University

yinjie@scu.edu.cn

Abstract

Deep learning methods have shown remarkable performance in image denoising, particularly when trained on large-scale paired datasets. However, acquiring such paired datasets for real-world scenarios poses a significant challenge. Although unsupervised approaches based on generative adversarial networks (GANs) offer a promising solution for denoising without paired datasets, they are difficult in surpassing the performance limitations of conventional GAN-based unsupervised frameworks without significantly modifying existing structures or increasing the computational complexity of denoisers. To address this problem, we propose a self-collaboration (SC) strategy for multiple denoisers. This strategy can achieve significant performance improvement without increasing the inference complexity of the GAN-based denoising framework. Its basic idea is to iteratively replace the previous less powerful denoiser in the filter-guided noise extraction module with the current powerful denoiser. This process generates better synthetic clean-noisy image pairs, leading to a more powerful denoiser for the next iteration. In addition, we propose a baseline method that includes parallel generative adversarial branches with complementary “self-synthesis” and “unpaired-synthesis” constraints. This baseline ensures the stability and effectiveness of the training network. The experimental results demonstrate the superiority of our method over state-of-the-art unsupervised methods.

<https://github.com/linxin0/SCPGabNet>

1. Introduction

Image denoising aims to recover noise-free images from noisy observations by reducing the potential noise. Al-

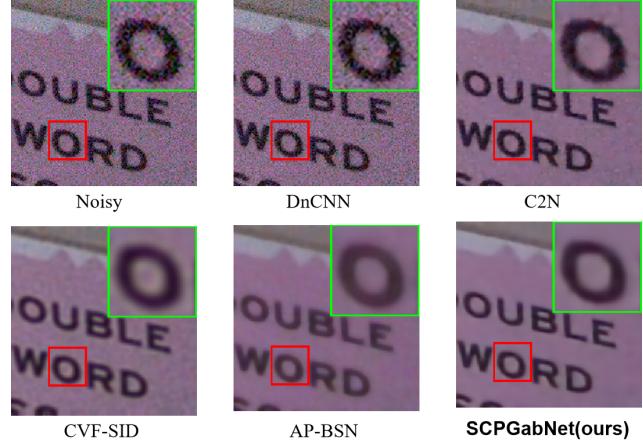


Figure 1. A real noisy image from the SSID Validation dataset. Our SCPGabNet achieves better results compared to other denoising methods.

though it is one of the oldest and most classical tasks in low-level computer vision, its fundamental nature continues drawing much interest. In general, existing image denoising algorithms can be divided into three groups: filtering-based [11, 9, 5, 14, 35], model-based methods [16, 31, 7, 19, 48, 41] and learning-based [46, 28, 36, 30, 42, 44, 29, 8, 2, 20] methods.

The additive white Gaussian noise (AWGN) assumption is widely used in image denoising. However, it is complex and challenging to adaptively achieve denoising based on filtering-based or model-based methods for high performance. In contrast, learning-based methods have demonstrated their superiority in image denoising. However, these methods [47, 36, 20, 39, 37, 45] are data-driven and typically require pairs of clean-noisy datasets to train their models. These noise samples are usually obtained through a predefined AWGN formulation that assumes the noise is

*Corresponding Author

signal-independent.

On the contrary, the real-world noise is more complex and much different from the ideal AWGN assumption. Using the AWGN model directly for the real scenes leads to poor performance. Therefore, numerous methods [4, 44, 2, 8, 38, 22, 37] have been proposed to capture paired clean-noisy image datasets from real scenes to promote the training of deep networks. However, these paired image-based methods focus on enhancing performance by improving network structures, and acquiring well-aligned pairs of clean-noisy images is time-consuming and laborious.

To solve the aforementioned problems, unsupervised denoising-based methods [10, 21, 15, 43] have emerged. The existing approaches are typically based on generative adversarial network (GAN) frameworks, which mainly focus on generating higher quality pseudo-noisy images. GAN2GAN [6] identified one of the key limitations of unsupervised denoising frameworks is the gap between the real and synthetic images, and proposed a novel approach using multiple generators and discriminators to generate images that closely conform to the real noise distribution. However, the performance of existing unsupervised denoising frameworks remains unsatisfactory due to the difficulty of adversarially training. Moreover, after training the model, the existing frameworks cannot further maximize the denoising potential without significantly changing its structure or increasing the inference complexity (e.g., using certain self-ensemble strategy) for denoisers. To address the previous limitations, we innovatively propose an unsupervised real-world denoising network called Self Collaboration Parallel Generative Adversarial Branches (SCPGabNet). The self-collaboration (SC) strategy, which provides the framework a powerful self-boosting capability. This enables the denoisers obtained from the conventional GAN framework to continuously evolve themselves and significantly improve their performance. The major contributions of our method are as follows:

- We design a novel filter-guided synthetic noisy image generator with the noise extraction (NE) module to synthesize high-quality clean-noisy image pairs, which serve as the foundation for implementing the SC strategy.
- We propose an effective parallel generative adversarial branches framework with complementary “self-synthesis” and “unpaired-synthesis” constraints as our powerful baseline.
- We are the first to propose the SC strategy, which significantly enhances the performance of the GAN-based denoising framework without increasing its inference complexity. Experimental results demonstrate the superiority of our SCPGabNet over state-of-the-art unsu-

pervised image denoising methods with large margins on the SIDD and DND benchmarking datasets.

2. Related Work

2.1. Denoising for Synthetic Noisy Image

Image denoising techniques for synthetic noisy images, such as AWGN, can be classified into classical non-deep-learning-based, and deep-learning-based denoising approaches. Classical non-deep-learning-based approaches include filtering-based [11, 9, 5, 14, 35], model-based [16, 31, 7, 19, 48, 41], and traditional learning-based methods. However, the filtering-based methods require manually designed filters, and model-based methods require several preset hyperparameters and entail a significant computational burden. These limitations restrict their flexibility and efficacy. To address these challenges, DnCNN [46] was proposed and achieved significant performance improvements over traditional methods. Subsequently, deep network-based methods [46, 49, 39] became the mainstream in AWGN image denoising. For instance, several learning-based methods with advanced network designs have been proposed after DnCNN, including FFDNet [47], N^3 Net [36], and MemNet [39]. However, as pointed out by Guo et al. [20], AWGN-based training methods suffer in real-world denoising due to the domain differences between real and synthetic noise.

2.2. Denoising for Real-World Image

Real paired-dataset based supervised denoising methods. These approaches typically involve designing high-performance denoising networks that are trained in a supervised manner using pairs of clean-noisy image datasets captured from real scenes. Anwar et al. [2] proposed a method that combines synthetic and real images during training to enhance the generality of the denoising model. Cheng et al. [13] generated a set of image basis vectors from the noisy input images and reconstructed them from the subspace formed by these basis vectors to obtain image-denoising results. Ren et al. [37] proposed a novel model-based denoising method that informs the design of the network for both synthetic and real denoising. Recently, Ren et al. [38] proposed a novel depth-unfolding network based on a latent space blind model via self-correction alternative optimization.

Synthetic paired-dataset based two-step denoising methods. While several real-world noisy datasets have been introduced, the limited number of training datasets motivates some approaches to divide the denoising task into two steps based on noise modeling: 1) synthesizing paired datasets by modeling noise or incorporating a priori noise information, and 2) training the denoiser. By leveraging prior knowledge of the image signal processing (ISP) pipeline, CBDNet [20]

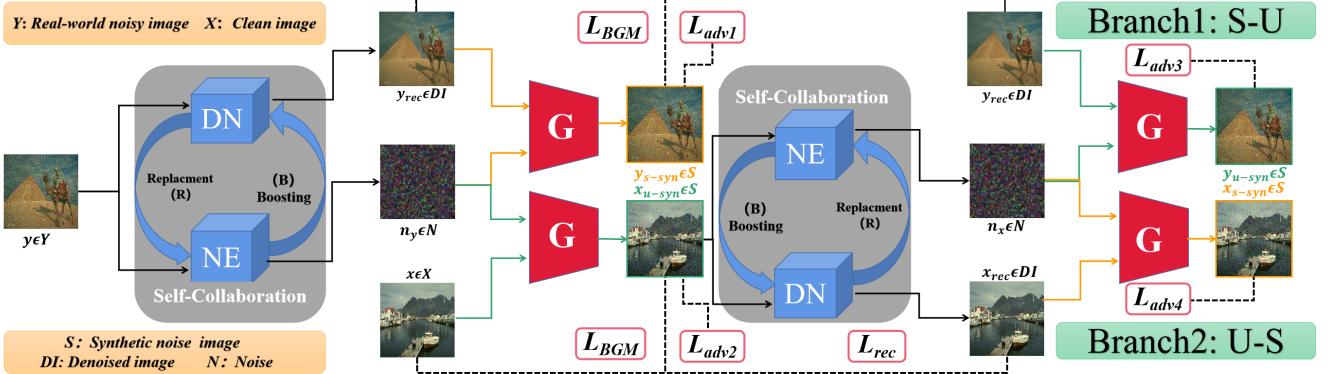


Figure 2. The architecture of our proposed SCPGabNet framework consists of two branches: Branch1: “**S**elf-synthesis - **U**npaired-synthesis” (Left: obtain **S**elf-synthesis image y_{s-syn} from y_{rec} and n_y ; Right: obtain **U**npaired-synthesis image y_{u-syn} from y_{rec} and n_x) and Branch2: “**U**npaired-synthesis - **S**elf-synthesis” (Left: obtain **U**npaired-synthesis image x_{u-syn} from x and n_y ; Right: obtain **S**elf-synthesis image x_{s-syn} from n_x and x_{rec}). Each branch contains a self-collaborative operation that involves a noise extraction module **NE** and a denoiser **DN**. This process is essentially an **R** (replacement)-**B** (boosting) iteration, where the **NE** module extracts noise from the noisy image, and the **DN** removes the noise from the noisy image to generate a clean image.

emulated the pipeline using a gamma correction and demosaicking process. Then, the synthesized anisotropic Gaussian noise is transformed into a realistic noise signal that is used to generate training pairs for supervised learning. Liu et al. [29] employed a priori information on image degradation to synthesize realistic noisy images and achieve good denoising performance. Zhou et al. [18] generated synthetic Gaussian-distributed noisy images and trained a Gaussian denoiser on a paired dataset of such images. For testing, they applied a conventional noise distribution conversion method that approximates real noise as Gaussian through pixel shuffling. Jang et al. [23] proposed a Clean-to-Noisy generator network based on a GAN that learns the features of real-world noises, capable of accurately representing various noise types. The generated pairs are then used to train a denoiser.

Image Denoising without Paired Dataset. Obtaining paired datasets in practice is extremely challenging, which has led to the emergence of simple unsupervised methods. These methods do not require prior knowledge of the noise model but instead combine image synthesis and denoising within a single framework. Chen et al. [10] firstly proposed a noise generator that can create pseudo-noisy images to train a denoiser. Cha et al. [6] proposed GAN2GAN, which uses a multi-generator/discriminator structure to better extract noisy information and generate pseudo images that better match the real noise distribution. Hong et al. [21] proposed UIDNet, which employs a sharpening processing mechanism to achieve noise separation and better train unpaired denoising models. In addition, several methods have been proposed to train models using only noisy images. Ne-

shatavar et al. [32] developed a self-supervised network that can handle real-world signal-dependent noise with greater adaptability to realistic noise. LEE et al. [27] proposed an unpaired learning approach that combines cyclic adversarial learning and self-supervised residual learning.

3. Proposed Method

In this section, we introduce a promising denoising network called SCPGabNet. The details are illustrated in the following subsections.

3.1. Parallel Generative Adversarial Branches for Unsupervised Denoising (PGabNet)

To apply the SC strategy, we first propose a high-performance baseline in this subsection, named PGabNet, which ensures the stability and effectiveness of the training process.

3.1.1 Filter-Guided Synthetic Noisy Image Generator

While supervised denoising frameworks generally outperform unsupervised ones (e.g., [17] 40.30dB, [37] 39.35dB vs. [32] 34.71dB, [27] 34.90dB on the SIDD Benchmark), they require a large number of paired datasets. Conversely, unsupervised denoising frameworks are more robust to situations where training images are inadequate or unpaired, compared to supervised methods. According to [6, 23], the reason for this performance gap is the difference between the synthetic noisy images and the real ones. The more realistic the training images are, the better the denoiser’s per-

formance will be. If the quality of the synthetic image is good enough (infinitely close to the real image), then the performance of the denoiser will not differ between supervised and unsupervised frameworks. Therefore, improving the quality of the synthetic image is a general method to enhance the performance of unsupervised denoising framework.

As depicted in Figure 3, synthesizing real noisy images with high quality can be challenging due to the influence of different image contents. To mitigate this difficulty, we propose a novel filter-guided synthetic noisy image generator for our GAN model, which better captures noise information. Since it is difficult to learn the noise distribution directly, instead of directly inputting a noisy image and a clean image into G , we use the NE module to obtain the noise. Specifically, a learnable convolutional block called DN_0 is used for denoising, and the noise is then obtained by subtracting the denoised image from the original noisy image. This approach reduces the learning difficulty of the generator and encourages it to focus more on synthesizing high-quality noisy images, thereby improving the overall denoising performance.

3.1.2 Parallel Generative Adversarial Branches

The overall framework is a GAN structure that utilizes unpaired clean and noisy images. We employ a Resnet with 6 residual blocks as the generator, a PatchGAN as the discriminator [51] and a DeamNet [37] as the denoiser. The specific structures of these variants are shown in detail in the [Supplementary Material](#).

There are two cases in learning synthetic pseudo noisy images in PGabNet: (1) when the clean and noisy images are different; (2) when the clean and noisy images are the same. Unpaired synthesis is a common approach in many computer vision works, where one degraded image guides the generation of a synthetic image from another clean image (e.g., [15, 23]). The goal is to learn the noise distribution properties from the input noise of one noisy image to guide the generation of a similar pseudo-signal-dependent noisy image from another clean image. This method imposes unpaired constraints on the generator, which captures more prior information and improves the quality of synthetic noisy images. A robust generator should learn the real noise distribution properties of different inputs. To balance the noise extracted from “same image” and “different images”, we propose self-synthesis. These two complementary constraints can improve the adversarial performance of the generator-discriminator and produce synthetic noisy images that are more consistent with the true noise distribution.

As illustrated in Figure 2, the PGabNet comprises two branches, with each branch implementing the complemen-

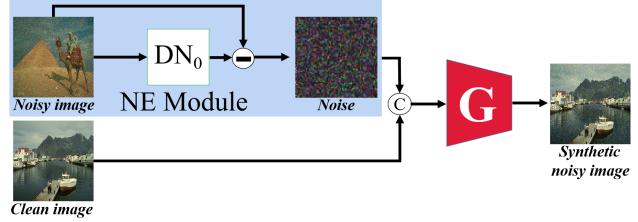


Figure 3. The architecture of the filter-guided synthetic noisy image generator is designed to extract real noise through the noise extraction (NE) module and project the noise distribution directly onto the clean image. This approach reduces the difficulty for the generator to synthesize better noisy images.

tary “self-synthesis” and “unpaired synthesis” constraints, respectively. Branch 1 utilizes “self-synthesis - unpaired synthesis” architecture, while branch 2 employs the “unpaired synthesis - self-synthesis” architecture. Specifically, branch 1 generates the self-synthesis noisy image y_{s-syn} and the unpaired synthesis noisy image y_{u-syn} , while branch 2 generates the unpaired synthesis noisy image x_{u-syn} and the self-synthesis noisy image x_{s-syn} . These images are then fed as inputs to the discriminator, along with the real noisy image y . The “self-synthesis” constraint and the “unpaired synthesis” constraint are strongly complementary within each branch of PGabNet. Additionally, the “self-synthesis” constraint and the “unpaired synthesis” constraint between these two branches are also complementary.

3.1.3 GAN-based Noise Synthesize and Loss function

As depicted in Figure 2, x and y represent the clean and real-world noisy images. Correspondingly, X and Y denote the clean and real-world noisy domains. The generator G aims to perform domain transformation by learning the image distribution in an unsupervised GAN framework. Simultaneously, the discriminator D (Due to space limitation, it is not given in this figure. Please see [Supplementary Material](#) for details.) evaluates whether the output of the generator belongs to the same domain as the target image. The generator and discriminator are trained in an adversarial manner to accomplish the domain transformation. Many studies [23, 20, 50] have demonstrated that signal-dependent noise can be modeled by additivity. The G in Figure 3, we first extract the noise n_y from the noisy image using the NE module, and then input both the noise and clean image into G to synthesize a pseudo noisy image:

$$x_{u-syn} = G(x, NE(y)) \quad (1)$$

We simultaneously train a discriminator D to distinguish whether a given noisy image is synthesized by our generator G or sampled from a real-world dataset. Here, y and x_{u-syn} denote the real-world image and synthetic image,

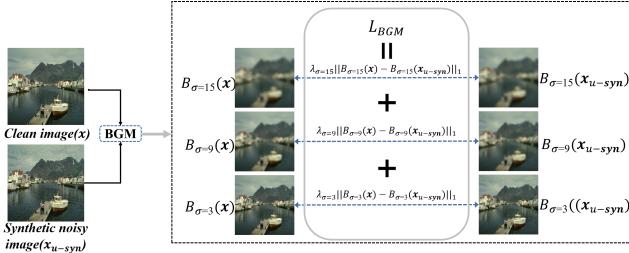


Figure 4. Example of BGM loss.

respectively. To prevent model degradation during training and improve the representation capability of the network, we use the least squares loss as the adversarial loss for L_{adv2} . The mathematical expression is as follows:

$$L_{adv2}(D, G) = -E_{y \sim Y} [\|D(y) - 1\|_2^2] - E_{x_{u-syn} \sim S} [\|D(x_{u-syn}) - 0\|_2^2] \quad (2)$$

That means for the generated image x_{u-syn} , its adversarial loss L_{adv2} is constrained between y and x_{u-syn} . The other three diversial losses can be constructed similarly by constraining the current generated image and y . Then, the total loss can be given by:

$$L_{GAN} = L_{adv1} + L_{adv2} + L_{adv3} + L_{adv4} \quad (3)$$

Inspired by [24], we further apply a Background Guidance Module (BGM) to provide additional reliable supervision. The BGM maintains the consistency of the background between the synthetic noisy image and the clean image, constraining their low-frequency contents to be similar. Specifically, we illustrate this approach using L_{BGM} in branch 2. Low-frequency contents are extracted by using several low-pass filters and constrained to be similar to each other through the L1 norm loss :

$$L_{BGM} = E_{x_{u-syn} \sim S, x \sim X} \left[\sum_{\sigma=3,9,15} \lambda_\sigma \|B_\sigma(x) - B_\sigma(x_{u-syn})\|_1 \right] \quad (4)$$

where $B_\sigma(\bullet)$ denotes the Gaussian filter operator with blurring kernel size σ , and λ_σ denotes the weight for the level σ . An example of the BGM loss is shown in Figure 4. We empirically set σ -s to 3, 9 and 15, and λ -s to 0.01, 0.1, and 1, respectively.

In our image denoising framework, we utilize pseudo-paired samples denoted by x_i and $(x_{rec})_i$. The denoiser is trained by optimizing the following loss functions:

$$L_{DN}(\Theta) = \frac{1}{2m} \sum_{i=1}^m [\|(x_{rec})_i - x_i\|_1 + \lambda_{SSIM} L_{SSIM}((x_{rec})_i, x_i)] \quad (5)$$

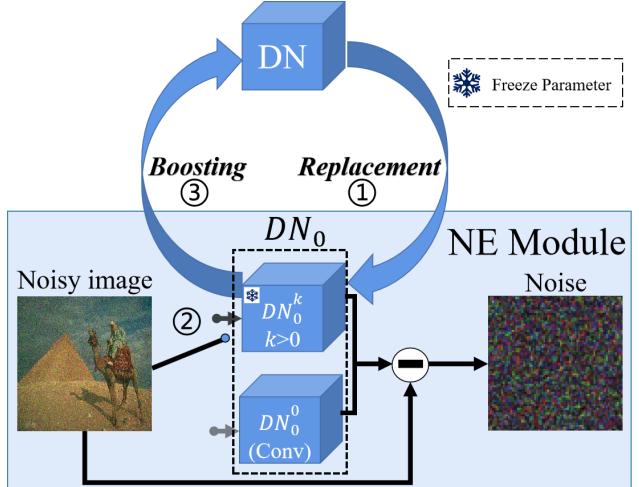


Figure 5. Visual illustration of the SC strategy. In NE, k denotes the number of iterations. It is a learnable convolutional block when $k=0$. However, when $k>0$, the iterative collaboration of DN_0 and the denoiser DN is conducted. Specifically, the weaker denoiser DN_0 is successively replaced by the current more powerful denoiser DN, which enables the synthesis of better clean-noisy image pairs. This iterative process results in a series of increasingly powerful denoisers. ①: the current denoiser DN is applied to replace the previous weaker denoiser DN_0 . This replacement allows for the creation of a new and more powerful denoiser DN. ②: the updated NE is obtained to generate better clean-noisy image pairs that are used to train the denoiser DN. ③: the denoiser DN is trained using the updated clean-noisy image pairs, which further boosts its performance.

where m denotes the total number of the sample pairs, $(x_{rec})_i$ is the clean image estimated by the denoising network, $L_{SSIM}(\bullet)$ represents the structural information used by SSIM loss to constrain the image, and λ_{SSIM} is the weight for L_{SSIM} . In conclusion, the total loss function is:

$$L = \min_G \max_D L_{GAN} + \lambda_{BGM} L_{BGM} + L_{DN}(\Theta) \quad (6)$$

where λ_{BGM} denotes the hyperparameter of background consistency loss.

3.2. Proposed SC based PGabNet (SCPGabNet)

In the previous subsection, we introduced our powerful baseline, PGabNet. However, as PGabNet is still a conventional GAN-based unsupervised framework, it is highly challenging to further improve its performance without significantly altering the network architecture or increasing inference complexity. To address this problem, we propose SC-based PGabNet (SCPGabNet).

3.2.1 Self-Collaboration Strategy

Our SC strategy is an innovative approach that enables a denoiser trained in conventional structures to self-correct and improve its performance without requiring modifications to the denoiser’s structure or increased complexity. By incorporating the SC strategy into PGabNet, we aim to surpass the performance limit of conventional GAN-based unsupervised frameworks.

The SC strategy is illustrated in Figure 5, and it functions more like an outer loop than an inner loop. The structure consists of a noise extraction module (NE) and a denoiser (DN). As described in the previous sub-sections, the NE module extracts noise from a noisy image and guides the generator to produce high-quality synthetic noisy images. To train the PGabNet, we initially use a simple and learnable linear convolutional layer as DN_0^0 in the NE module. Then, the denoiser is iteratively replaced and boosted, with k denoting the number of iterations. During each iteration, the current more powerful denoiser DN replaces the previous weaker denoiser DN_0 in the NE module, significantly improving the performance of the updated denoiser. By using the SC strategy for the denoiser inside and outside the NE module, better DN_0 can be obtained, which can extract more accurate noise. This results in the production of more realistic synthetic noisy pairs, and iteratively improves the performance of the updated DN with higher-quality synthetic samples. Excitingly, we observe a significant improvement in the performance of DN by using the SC strategy than the original one without SC.

During the SC stage, the loss functions of G and D are the same as before. The loss function of DN:

$$\begin{aligned} L_{DN-SC} = & L_{DN} + \frac{1}{2m} \left[\sum_{i=1}^m \| (x_{rec})_i - DN_0(x_{u-syn}) \|_1 \right. \\ & + \| (y_{rec})_i - DN_0(y_i) \|_1 \\ & + \lambda_{SSIM} L_{SSIM}((x_{rec})_i, DN_0(x_{u-syn})) \\ & \left. + \lambda_{SSIM} L_{SSIM}((y_{rec})_i, DN_0(y_i)) \right] \end{aligned} \quad (7)$$

The basic idea of the SC is to use the result of the previous stage as feedback information to guide and improve the subsequent stages. The ‘feedback information’ can either be a structure or an output image. This approach has significant potential in low-level vision tasks.

3.2.2 Analysis of SC for PGabNet

To develop a more powerful denoising framework, we integrate the SC strategy with our previous PGabNet, which we now refer to as SCPGabNet. At the start of each iteration, the new NE can more accurately capture noise in the noisy

image by replacing DN_0 in the NE module with a more powerful DN. This reduces the impact of the image content on the synthetic noisy image generation process. As illustrated in Figure 2, with more precise noise extracted from the noisy image y , our denoiser can achieve better results in both self-synthesis in branch 1 and unpaired synthesis in branch 2, resulting in a higher quality synthetic noisy image. Similarly, more accurate noise extracted from the synthetic noisy image x_{u-syn} can improve unpaired synthesis in branch 1 and self-synthesis in branch 2, thus promoting complementary constraints between the two branches and making the modules in the network more interconnected. As a result, our SC strategy creates a self-boosting framework that enables better denoiser training and performance.

The implementation of SC strategy in SCPGabNet involves several steps: First, after the original framework (PGabNet) has reached convergence, we replace DN_0 in the NE module with the latest DN and fix its parameters to generate better pseudo-noisy images. Next, we retrain G, D, and DN until convergence is achieved. Finally, we repeat the previous processes until the performance of DN no longer improves.

4. Experiments

In this section, we first describe the experimental settings, including the datasets and training details. After that, to evaluate the effectiveness, the proposed method is compared with some representative supervised/unsupervised methods. Finally, we analyze the proposed method in-depth.

4.1. Experimental Setting

Training and Testing Data. To train and test the model, we first equally divide the SIDD Medium training set (consisting of 320 pairs of noise images and corresponding clean images captured by multiple smartphones) into separate noisy and clean image parts. Then, we use 160 clean images from the first part and 160 noisy images from the second part to construct an unpaired dataset of real images for training the algorithm presented in this paper. We evaluate the denoising performance on three widely used real-world noisy datasets SIDD[1] Validation, SIDD[1] Benchmark, and the DND[40] Benchmark. Note that the denoised images of SIDD Benchmark and DND Benchmark can be uploaded to the SIDD and DND websites to obtain the PSNR and SSIM results.

Implementation Details. To optimize the proposed network, we adopt the Adam optimizer algorithm with $\beta_1=0.9$, $\beta_2=0.999$, and the initial learning rate is set to 10^{-4} . The mini-batch size is set to 10, the used framework is PyTorch, and the used GPU is GeForce RTX 3090. For noise learning, the background consistency loss hyperparameter

Table 1. Denoising results of several competitive methods on SIDD Validation, SIDD Benchmark, and DND Benchmark

	Methods	Conference/Journal	SIDD Validation		SIDD Benchmark		DND Benchmark	
			PSNR↑	SSIM↑	PSNR↑	SSIM↑	PSNR↑	SSIM↑
Non-learning	BM3D [14] WNNM [19]	TIP 2007 CVPR 2014	31.75 —	0.7061 —	25.65 25.78	0.6850 0.8090	34.51 34.67	0.8510 0.8650
Real pairs(Supervised)	TNRD [12] DnCNN [46] FFDNet [47] RIDNet [3] AINDNet [25] InvDN [29] DeamNet [37] ScaoedNet [38]	TPAMI 2016 TIP2017 TIP 2018 CVPR 2019 CVPR 2020 CVPR 2021 CVPR 2021 NeurIPS 2022	26.99 26.20 26.21 38.76 38.96 38.30 39.40 39.52	0.7440 0.4414 0.6052 0.9132 0.9123 0.9064 0.9169 0.9187	24.73 28.46 29.30 37.87 38.84 39.28 39.35 39.48	0.6430 0.7840 0.6940 0.9430 0.9510 0.9550 0.9550 0.9570	33.65 32.43 34.40 39.25 39.34 39.57 39.63 40.17	0.8310 0.7900 0.8470 0.9530 0.9520 0.9520 0.9531 0.9597
Synthetic pairs(two pipeline)	DnCNN [46] CBDNet [20] PD+ [18] C2N+DnCNN [23] C2N+DIDN [23]	TIP 2017 CVPR 2019 AAAI 2020 ICCV 2021 ICCV 2021	— 30.83 34.03 — —	— 0.7541 0.8810 — —	23.66 33.28 34.00 33.76 35.02	0.5830 0.8680 0.8980 0.9010 0.9320	32.43 38.06 38.40 36.08 36.12	0.7900 0.9420 0.9450 0.9030 0.8820
Unsupervised	N2V [26] GCBD [10] UIDNet [21] R2R [34] CVF-SID (S^2) [32] AP-BSN[27] PGabNet(baseline) SCPGabNet(ours)	CVPR 2019 CVPR 2018 AAAI 2020 CVPR 2021 CVPR 2022 CVPR 2022 — —	29.35 — — 35.04 — 34.46 34.66 36.53	0.6510 — — 0.8440 — 0.8501 0.8517 0.8860	27.68 32.48 34.78 34.71 34.71 34.90 34.67 36.53	0.6680 0.8970 0.8980 0.9170 0.9170 0.9000 0.8950 0.9250	— — — 36.20 36.50 37.46 36.87 38.11	— 35.58 — 0.9220 0.9240 0.9240 0.9267 0.9393

λ_{BGM} in the loss function of Eq. 4 is set to 6. For the denoising network, the hyperparameter λ_{SSIM} of the SSIM constraint term in Eq. 5 is set to 1.

4.2. Real-World Image Denoising Analysis

In this subsection, we evaluate the denoising performance of our method on real images from the SIDD Validation, SIDD Benchmark, and DND Benchmark. We compare our method with a range of traditional-based methods, representative supervised methods based on paired images, and latest unsupervised methods based on unpaired images from recent years. We utilize evaluation metrics (PSNR and SSIM) to assess the effectiveness of each method.

Table 1 presents the quantitative evaluation of various methods on the SIDD Validation, SIDD Benchmark, and DND Benchmark in the sRGB space. It is observed that SCPGabNet achieves the best performance among the unsupervised approaches. Specifically, compared with the latest unpaired methods AP-BSN and CVF-SID presented at CVPR 2022, SCPGabNet provides a PSNR gain of 1.63 dB and 1.82 dB, and an SSIM gain of 0.25 and 0.008 on the SIDD benchmark. Moreover, SCPGabNet outperforms CBDNet, which is a classic supervised real denoising network. In terms of the two-stage synthetic denoising method, SCPGabNet outperforms two-stage pipeline methods with self-ensemble (e.g., C2N+DnCNN*, C2N+DIDN*) on both SIDD and DND datasets. Although PD+[18] performs slightly better than SCPGabNet on the DND Benchmark, SCPGabNet’s performance is significantly superior to PD+ on both SIDD Benchmark and SIDD validation by more than 2dB. SCPGabNet is trained only on the SIDD training set and still exhibits a greater advantage on other bench-

marks, indicating its greater generalizability than two-step methods. Although our method’s denoising performance is not as good as some of the latest supervised methods based on real image pairs, such as ScaoedNet[38], the majority of these methods depend on a large number of paired images. Consequently, they may not be effectively applied to real-world image-denoising tasks when training images are insufficient or cannot be paired. In contrast, our SCPGabNet can achieve image denoising without using paired images, making it flexible enough to handle a wide range of real-world denoising scenarios. Figure 6 depicts denoised examples on the SIDD Validation dataset. Our method outperforms comparable unsupervised and two pipeline methods in noise removal and detail preservation.

4.3. Ablation Study

In this subsection, we demonstrate the importance of the PGabNet structure and BGMloss in Table 2. V1 is the GAN-based unsupervised denoising network with only unpaired synthesis. V2 is the V1 + BGMloss, V3 is the V2 + NE module. V4 is the branch U-S, i.e., SGabNet and V5 is our baseline (PGabNet). Due to the limited space, the specific structure of these variants are provided in the Supplementary Material.

By comparing V1 with V2, we found that adding the BGMloss to the GAN-based unsupervised denoising network can lead to a more stable training process. Furthermore, adding the NE module to V2 to obtain V3, the PSNR increase is about 0.20dB, verifying the effectiveness of the NE module and the improvement in the quality of synthetic images. Comparing V3 with V4 (SGabNet), we observed that the denoiser experienced a more significant improve-

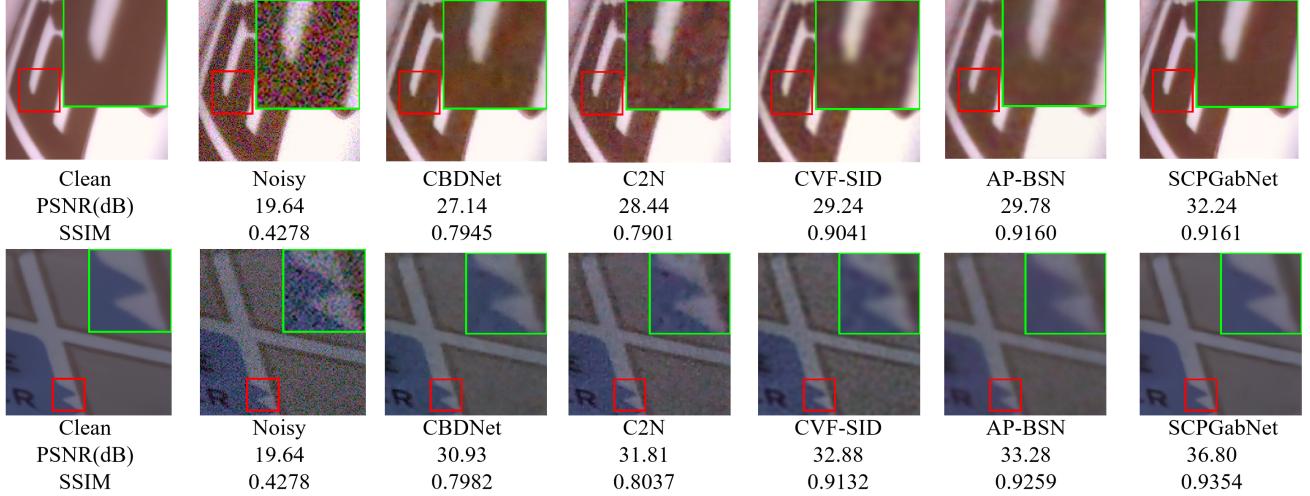


Figure 6. Visual comparison of our method against other competing methods. The quantitative PSNR(dB)/SSIM results are listed as well.

Table 2. **V1:** (U) Conventional GAN-based unsupervised denoising network only with unpaired synthesis; **V2:** V1 + BGMloss; **V3:** V1 + BGMloss + NE module; **V4:** (S) SGabNet(V1 + BGMloss + NE module + self-synthesis) **V5:** (P) PGabNet(our baseline).

Methods	V1	V2	V3	V4	V5(ours)
U	✓	✓	✓	✓	✓
BGMloss		✓	✓	✓	✓
NE module			✓	✓	✓
S				✓	✓
P					✓
PSNR(dB)	33.14	33.26	33.45	34.27	34.67

ment after adding the “self-synthesis” constraint. Specifically, the PSNR improved by 0.82dB on the SIDD Benchmark, demonstrating that “self-synthesis + unpaired synthesis” can better train the network and enhance the denoiser’s performance. Finally, compared PGabNet with SGabNet, we observed a significant performance improvement in PGabNet with PSNR gains of 0.40dB on the SIDD Benchmark. This is because PGabNet enhances PGabNet with SGabNet, leading to the production of better synthetic noisy images that are more consistent with the real image distribution and improving the denoiser’s performance.

4.4. Effectiveness of SC strategy

In this subsection, we apply the SC strategy to our baseline method (PGabNet) and obtain SCPGabNet. The details of this process are as follows: In the first stage, the batch size is 8 and the patchsize is 112. In the second stage, we increase the patchsize to 128 and reduce the batchsize to 4 for fine-tuning. The learning rate is keep at 10^{-4} for both stages. As shown in Figure 7, we display the performance of the SC on the SIDD Benchmark. In the initial few iter-

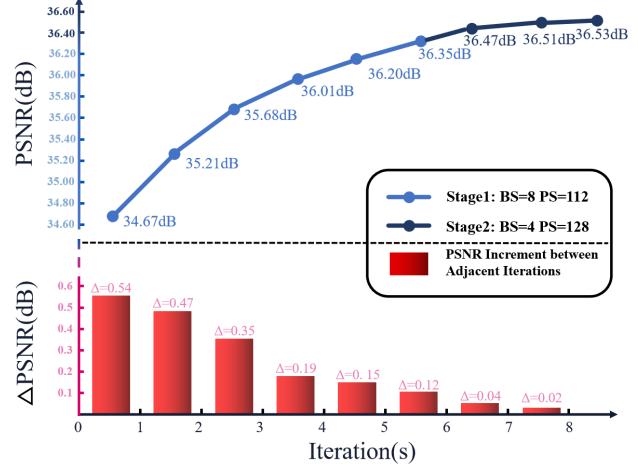


Figure 7. The entire process involves eight iterations, consisting of two stages. In the first stage: the batchsize=8 and the patchsize=112. In the second stage, fine-tuning is performed with a batch size of 4 and a patch size of 128. The upper part of the figure shows the results of SCPGabNet on the SIDD Benchmark after each iteration, while the lower part displays the improvement achieved in each iteration.

tions, we observe a notable improvement in denoising performance, especially in the first iteration where the denoiser improves by more than 0.5dB. Based on the PSNR increments observed between adjacent iterations, the Δ PSNR decreases as the iterations progress, with the last iteration resulting in only a 0.02dB improvement. By comparison with PGabNet, with a PSNR of 36.53dB, significantly improved by 1.86dB on the SIDD Benchmark. This improvement highlights the effectiveness of our proposed approach in achieving state-of-the-art performance in image denoising.

4.5. Analysis on Transferability

To evaluate the generality and transferability of our approach, we apply the proposed SC strategy to several classical and the latest denoising networks, including DnCNN [46], UNet [33], DeamNet [37] and DBSNL [43], which serves as the denoiser for the AP-BSN [27]. The framework still adopts the PGabNet proposed in section 3 with different denoiser networks.

Table 3 illustrates that our proposed SC strategy is effective to these networks. For instance, after applying the SC strategy to DnCNN, the PSNR/SSIM improvements on SIDD Validation are 0.96dB/0.0085. For UNet and DBSNL, the gains are 1.58dB/0.0373 and 1.52dB/0.0035 respectively. Consequently, the results demonstrate that our SC strategy has strong transferability and can be potentially applied to other unsupervised denoising methods.

Table 3. The effects of SC for different denoisers.

Networks	Self-Collaboration	SIDD Validation		SIDD Benchmark		DND Benchmark	
		PSNR↑	SSIM↑	PSNR↑	SSIM↑	PSNR↑	SSIM↑
DnCNN [46]	No	30.60	0.8553	30.56	0.8150	29.79	0.7876
	Yes	31.56	0.8638	31.52	0.8260	30.77	0.7892
UNet [33]	No	34.46	0.8417	34.63	0.8960	35.81	0.9140
	Yes	36.04	0.8790	36.00	0.9180	38.04	0.9401
DBSNL [43]	No	34.38	0.8757	34.29	0.9100	36.51	0.9257
	Yes	35.90	0.8792	35.86	0.9160	37.72	0.9295
Deamnet [37]	No	34.66	0.8517	34.67	0.8950	36.87	0.9267
	Yes	36.53	0.8860	36.53	0.9250	38.11	0.9393

5. Conclusion

In this paper, we first introduce a parallel generative adversarial branches for unsupervised real-world image denoising as our baseline. Furthermore, we innovatively propose an SC strategy that can provide the denoiser a self-boosting capacity and significantly improve denoising performance. Our experimental results demonstrate that the proposed method achieves state-of-the-art performance. In addition, we validate the transferability of the SC strategy on various denoisers and anticipate its potential applicability to many low-level computer vision tasks.

6. Limitations and Future Work

Our current focus is on unsupervised tasks, with supervised tasks slated for future exploration. We specifically investigate the denoising task and, at the same time, consider the potential to delve into restoration tasks. In the SC strategy, each iteration necessitates manually selecting the optimal model iteration or retraining within the phase based on metrics, which is labour-intensive. The SC strategy can be simplified in the future by using automated iteration. The overall framework is based on CNNs, and better performance can be obtained in the future with transformers.

Acknowledgement This work was supported by the National Natural Science Foundation of China under Grant 62171304, and in part by 62276176.

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 (CVPR)*, pages 1692–1700, 2018. 6
- [2] Saeed Anwar and Nick Barnes. Real image denoising with feature attention. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 3155–3164, 2019. 1, 2
- [3] S. Anwar and N. Barnes. Real image denoising with feature attention. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3155–3164, 2019. 7
- [4] Tim Brooks, Ben Mildenhall, Tianfan Xue, Jiawen Chen, Dillon Sharlet, and Jonathan T Barron. Unprocessing images for learned raw denoising. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11036–11045, 2019. 2
- [5] Antoni Buades, Bartomeu Coll, and Jean-Michel Morel 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 vol. 2, 2005. 1, 2
- [6] Sungmin Cha, Taeeon Park, Byeongjoon Kim, Jongduk Baek, and Taesup Moon. Gan2gan: Generative noise learning for blind denoising with single noisy images. In *International Conference on Learning Representations (ICLR)*, pages 1712–1722, 2019. 2, 3
- [7] Antonin Chambolle. An algorithm for total variation minimization and applications. *Mathematical Imaging and Vision*, 20(1):89–97, 2004. 1, 2
- [8] Meng Chang, Qi Li, Huajun Feng, and Zhihai Xu. Spatial-adaptive network for single image denoising. In *European Conference on Computer Vision*, pages 171–187. Springer, 2020. 1, 2
- [9] Jingdong Chen, J. Benesty, Yiteng Huang, and S. Doclo. New insights into the noise reduction wiener filter. *IEEE Transactions on Audio, Speech, and Language Processing*, 14(4):1218–1234, 2006. 1, 2
- [10] J. Chen, J. Chen, H. Chao, and M. Yang. Image blind denoising with generative adversarial network based noise modeling. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3155–3164, 2018. 2, 3, 7
- [11] Tao Chen, Kai-Kuang Ma, and Li-Hui Chen. Tri-state median filter for image denoising. *IEEE Transactions on Image Processing*, 8(12):1834–1838, 1999. 1, 2
- [12] Yunjin Chen and Thomas Pock. Trainable nonlinear reaction diffusion: A flexible framework for fast and effective image restoration. *IEEE transactions on pattern analysis and machine intelligence*, 39(6):1256–1272, 2016. 7

- [13] Shen Cheng, Yuzhi Wang, Haibin Huang, Donghao Liu, Haoqiang Fan, and Shuaicheng Liu. Nbnnet: Noise basis learning for image denoising with subspace projection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4896–4906, 2021. 2
- [14] 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. 1, 2, 7
- [15] Wenchao Du, Hu Chen, and Hongyu Yang. Learning invariant representation for unsupervised image restoration. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 14483–14492, 2020. 2, 4
- [16] Michael Elad and Michal Aharon. Image denoising via sparse and redundant representations over learned dictionaries. *IEEE Transactions on Image Processing*, 15(12):3736–3745, 2006. 1, 2
- [17] Chen L et al. Simple baselines for image restoration. In *ECCV*, 2022. 3
- [18] Yuqian Zhou et al. When awgn-based denoiser meets real noises. In *AAAI*, 2020. 3, 7
- [19] Shuhang Gu, Lei Zhang, Wangmeng Zuo, and Xiangchu Feng. Weighted nuclear norm minimization with application to image denoising. In *2014 IEEE Conference on Computer Vision and Pattern Recognition*, pages 2862–2869, 2014. 1, 2, 7
- [20] Shi Guo, Zifei Yan, Kai Zhang, Wangmeng Zuo, and Lei Zhang. Toward convolutional blind denoising of real photographs. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1712–1722, 2019. 1, 2, 4, 7
- [21] Zhiwei Hong, Xiaocheng Fan, Tao Jiang, and Jianxing Feng. End-to-end unpaired image denoising with conditional adversarial networks. In *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*, pages 4140–4149, 2020. 2, 3, 7
- [22] Jie Huang, Zhibo Zhao, Chao Ren, Qizhi Teng, and Xiaohai He. Aprior-guideddeepnetworkforrealimagedenoisingandits applications. *Knowledge-Based Systems*, 255:109776, 2022. 2
- [23] Geonwoon Jang, Wooseok Lee, Sanghyun Son, and Kyounghu Lee. C2n: Practical generative noise modeling for real-world denoising. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, pages 2350–2359, 2021. 3, 4, 7
- [24] Xin Jin, Zhibo Chen, Jianxin Lin, Zhikai Chen, and Wei Zhou. Unsupervised single image deraining with self-supervised constraints. In *IEEE International Conference on Image Processing (ICIP)*, pages 2761–2765, 2019. 5
- [25] 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 *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2020. 7
- [26] Alexander Krull, Tim-Oliver Buchholz, and Florian Jug. Noise2void-learning denoising from single noisy images. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2129–2137, 2019. 7
- [27] Wooseok Lee, Sanghyun Son, and Kyoung Mu Lee. Ap-bsn: Self-supervised denoising for real-world images via asymmetric pd and blind-spot network. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 17725–17734, 2022. 3, 7, 9
- [28] Ding Liu, Bihan Wen, Yuchen Fan, Chen Change Loy, and Thomas S Huang. Non-local recurrent network for image restoration. *Advances in Neural Information Processing Systems*, 2018:1673–1682, 2018. 1
- [29] Yang Liu, Zhenyue Qin, Saeed Anwar, Pan Ji, Dongwoo Kim, Sabrina Caldwell, and Tom Gedeon. Invertible denoising network: A light solution for real noise removal. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 13365–13374, 2021. 1, 3, 7
- [30] Jan Lukas, Jessica Fridrich, and Miroslav Goljan. Digital camera identification from sensor pattern noise. *IEEE Transactions on Information Forensics and Security*, 1(2):205–214, 2006. 1
- [31] Julien Mairal, Michael Elad, and Guillermo Sapiro. Sparse representation for color image restoration. *IEEE Transactions on Image Processing*, 17(1):53–69, 2008. 1, 2
- [32] Reyhaneh Neshatavar, Mohsen Yavartanoo, Sanghyun Son, and Kyoung Mu Lee. Cvf-sid: Cyclic multi-variate function for self-supervised image denoising by disentangling noise from mage. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 17583–17591, 2022. 3, 7
- [33] Ronneberger O, Fischer P, and Brox T. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241, 2015. 9
- [34] Tongyao Pang, Huan Zheng, Yuhui Quan, and Hui Ji. Recorrupted-to-recorrupted: Unsupervised deep learning for image denoising. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2043–2052, 2021. 7
- [35] Shibin Parameswaran, Enming Luo, and Truong Q Nguyen. Patch matching for image denoising using neighborhood-based collaborative filtering. *IEEE Transactions on Circuits and Systems for Video Technology*, 28(2):392–401, 2016. 1, 2
- [36] Tobias Plötz and Stefan Roth. Neural nearest neighbors networks. *Advances in Neural Information Processing Systems*, 31:1087–1098, 2018. 1, 2
- [37] Chao Ren, Xiaohai He, Chuncheng Wang, and Zhibo Zhao. Adaptive consistency prior based deep network for image denoising. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 8596–8606, June 2021. 1, 2, 3, 4, 7, 9
- [38] Chao Ren, Yizhong Pan, and Jie Huang. Enhanced latent space blind model for real image denoising via alternative optimization. In *36th Conference on Neural Information Processing Systems*, pages 10581–10590, 2022. 2, 7
- [39] Ying Tai, Jian Yang, Xiaoming Liu, and Chunyan Xu. Memnet: A persistent memory network for image restoration. In

- Proceedings of the IEEE international conference on computer vision*, pages 4539–4547, 2017. 1, 2
- [40] TobiasPlotz and StefanRoth. Benchmarking denoising algorithms with real photographs. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1586–1595, 2017. 6
- [41] Chuncheng Wang, Chao Ren, Xiaohai He, and Linbo Qing. Deep recursive network for image denoising with global non-linear smoothness constraint prior. *Neurocomputing*, 426:147–161, 2021. 1, 2
- [42] Kaixuan Wei, Ying Fu, Jiaolong Yang, and Hua Huang. A physics-based noise formation model for extreme low-light raw denoising. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2758–2767, 2020. 1
- [43] Xiaohe Wu, Ming Liu, and Yue Cao. Unpaired learning of deep image denoising. In *European conference on computer vision*, pages 352–368, 2020. 2, 9
- [44] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, Ming-Hsuan Yang, and Ling Shao. Cycleisp: Real image restoration via improved data synthesis. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2693–2702, 2020. 1, 2
- [45] Sen Zhai, Chao Ren, Zhengyong Wang, Xiaohai He, and Linbo Qing. An effective deep network using target vector update modules for image restoration. *Pattern Recognition*, 122:108333, 2022. 1
- [46] 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. 1, 2, 7, 9
- [47] 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. 1, 2, 7
- [48] Xiaoqin Zhang, Jingjing Zheng, Di Wang, and Li Zhao. Exemplar-based denoising: A unified low-rank recovery framework. *IEEE Transactions on Circuits and Systems for Video Technology*, 30(8):2538–2549, 2019. 1, 2
- [49] Yulun Zhang, Yapeng Tian, Yu Kong, Bineng Zhong, and Yun Fu. Residual dense network for image restoration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43(7):2480–2495, 2020. 2
- [50] Rui Zhao, Daniel PK Lun, and Kin-Man Lam. Ntgan: Learning blind image denoising without clean reference. In *BMVC*, 2020. 4
- [51] JunYan Zhu, Taesung Park, Phillip Isola, and Alexei A. Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *IEEE International Conference on Computer Vision (ICCV)*, pages 2223–2232, 2017. 4