ROBUST SELF-SUPERVISED IMAGE DENOISING WITH CYCLIC SHIFT AND NOISE-INTENSITY-AWARE UNCERTAINTY

Anonymous authors
Paper under double-blind review

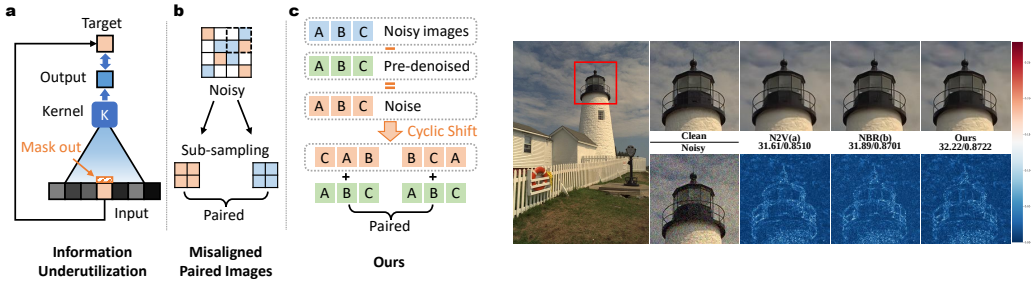
ABSTRACT

In self-supervised image denoising, it is challenging to construct paired noisy samples from a single noisy observation, and the quality of samples seriously influences the performance of the denoising model. Strategies for constructing pairs of samples for learning, such as blind-spot convolution and sub-sampling, are widely adopted in existing self-supervised denoising methods. However, these strategies suffer from the severe problems of information underutilization and pixel misalignment, which seriously hinder the further improvement of denoising performance. Furthermore, little attention has been paid to the sensitivity of denoising models to deal with unknown noise, which is of great significance in enhancing the practicality of denoising models. To overcome these challenges, we propose a very simple and effective method, called Cyclic Shift, to construct paired noisy images for self-supervised training. This new strategy solves the problems of information underutilization and pixel misalignment without additional computation, and it can be easily embedded into existing denoising methods and significantly boost their performance. In addition, we introduce the uncertainty-aware loss in training to enable the denoising network to perceive the noise intensity and have robust denoising performance. We theoretically explain the effectiveness of Cyclic Shift and analyze the ability of the uncertainty loss to endow the network with noise intensity perception. Extensive experimental results show that our approach achieves state-of-the-art self-supervised image denoising performance.

1 INTRODUCTION

Image denoising, which aims to recover a clean signal from noisy observations, is one of the essential tasks in image processing and low-level computer vision. Recently, with the development of neural networks, learning-based surpervised denoisers Anwar & Barnes (2019); Chang et al. (2020b); Ma et al. (2022) have achieved satisfactory performance. However, these methods depend heavily on noisy/clean or noisy/noisy paired images, and collecting such paired images is complex and expensive, most notably in dynamic scenes and medical imaging, resulting in the limited practice of supervised denoising methods.

To overcome this limitation, a series of un-/self-supervised denoising methods Laine et al. (2019); Cha et al. (2020); Xu et al. (2020); Pang et al. (2021); Huang et al. (2021); Wang et al. (2022) inspired by Noise2Noise Lehtinen et al. (2018) and Noise2Void Krull et al. (2019) have been proposed. These denoising methods learn from only single noisy images, which means that they have better practicability and broader application areas than supervised image denoising methods. However, this advantage requires constructing pairs of samples for self-supervised learning. The main idea of these denoising methods is to use the original noisy observations to get the paired noisy pixels/images, which have the same scene and independent and identically distributed (i.i.d.) noise. According to the strategies of constructing pairs of samples, they can be divided into two categories, blind-spot-based (Figure 1 (a)-a) and sub-sample-based methods (Figure 1 (a)-b). These blind-spot-based methods require masking out some pixels in the noisy images before feeding them to the network, causing severe information underutilization during training. In addition, sub-sample-based methods encounter the limitation of misaligned paired noisy images. These inherent limitations seriously hinder the further improvement of denoising performance.



(a) Comparison of the methods synthesizing paired noisy images from only single noisy images.

(b) Visual comparison of denoising sRGB images in the setting of Gaussian noise ($\sigma = 25$) on KODAK.

Figure 1: **(a) a**, represents the blind-spot-based methods. **b**, indicates the sub-sample-based methods, and **c**, denotes Cyclic Shift that can overcome the problems of information underutilization and pixel misalignment in **a** and **b**, respectively. A, B, and C denote different images. **(b)** Visual comparison of the denoising performance of three representative methods is shown in (a). For better viewing, we display the difference between the denoising results and Clean in colour.

To alleviate these problems mentioned above, in this paper, we propose a novel and easy-to-implement method, Cyclic Shift (Figure 1 (a)-c), for synthesizing paired noisy images from a group of different noisy observations. Specifically, we first obtain the group of pre-denoised images from a denoising network, which learns from only single noisy observations. Then we extract the group of noise from the corresponding noisy observations. After that, we can get two groups of noise with different orders by Cyclic Shift and then fuse them with the pre-denoised images to obtain paired noisy images with the same scene and i.i.d. noise. Finally, the constructed paired noisy images are used to train the denoising network.

Furthermore, most self-supervised image denoising methods pay much attention to improving denoising performance, with little attention to the robustness, resulting in these denoising models being sensitive to unseen noise. However, since there is no guarantee that the collected noisy observations are compliant in practice, it is essential for the denoising network to keep robust in facing complex scenarios. Therefore, we introduce the uncertainty-aware loss to enable the denoising network to perceive the noise intensity, resulting in robust denoising performance.

Overall, our contributions are summarized as follows:

- We propose a novel strategy to construct sample pairs for self-supervised learning called Cyclic Shift. It can well avoid the problems of insufficient information utilization and pixel misalignment existing in the popular sample pairs constructing strategies. As a result, it can extensively boost existing self-supervised denoising methods without any additional computational cost.
- We introduce the uncertainty-aware loss to endow the network with the ability to perceive the noise intensity to improve the robustness of the denoising network and obtain better denoising performance in the face of unseen noise. Besides, we theoretically illustrate the particularities of uncertainty modeling in self-supervised image denoising.
- By integrating Cyclic Shift and the uncertainty-aware loss, we obtain a robust and noise-intensity-aware self-supervised image denoising model (CSU), which achieves the state-of-the-art denoising performance on multiple datasets of sRGB, RAW, and grayscale domains.

2 Related Work

2.1 Image Denoising

2.1.1 Training with Clean Images

The first supervised image denoising approach using neural networks is DnCNN Zhang et al. (2017). Then, numerous advanced methods are proposed to further improve the denoising performance, such

as FFDNet Zhang et al. (2018), N3Net Plötz & Roth (2018), CBDNet Guo et al. (2019), RIDNet Anwar & Barnes (2019), NBNetCheng et al. (2021), and FADNet Ma et al. (2022). In practice, unpaired noisy/clean images are more accessible to obtain than collecting paired noisy/clean images used in supervised image denoising. Therefore, many denoising methodsChen et al. (2018); Hong et al. (2020); Jang et al. (2021); Wu et al. (2020); Lin et al. (2021) based on unpaired noisy/clean images are proposed. Besides, Lehtinen et al. Lehtinen et al. (2018) propose Noise2Noise, which can learn denoising from paired noisy images, and the denoising performance of this method is comparable to that of supervised denoising models. However, the difficulty and high cost of obtaining clean images make these methods challenging to apply in actual scenarios.

2.1.2 Training with Only Single Noisy Images

Inspired by Noise2Noise Lehtinen et al. (2018), Noise2Void Krull et al. (2019) proposes using the blind-spot convolution to learn denoising from only single noisy images. Subsequently, based on the blind-spot convolution, Noise2Self Batson & Royer (2019), AP-BSNLee et al. (2022), and Blind2Unblind Wang et al. (2022) are proposed, but they suffer a severe limitation of information underutilization. Besides, Noiser2Noise Moran et al. (2020), GAN2GAN Cha et al. (2020), R2R Pang et al. (2021), and IDR Zhang et al. (2022) employ prior knowledge to synthesize the paired noisy images from noisy observations to train the denoising network. In addition, Self2Self Quan et al. (2020) and NBR2NBR Huang et al. (2021) introduce two sampling approaches to generate paired noisy images. These methods do not require clean images but have varying degrees of information underutilization, pixel misaligned, and require prior knowledge about the noise.

2.2 Robustness of Denosing Model

In image denoising, it is common to use filtered datasets with similar distributions and intensities of noise and MSE/L1 loss for training, which results in the network performing well in denoising specific noise but is sensitive to unseen noise (*i.e.* poor robustness). Moreover, keeping the robustness of the model has received little attention in previous studies of image denoising but in other computer vision tasks Kendall & Gal (2017); Choi et al. (2019); Chang et al. (2020a); Upadhyay et al. (2021b); Sudarshan et al. (2021); Upadhyay et al. (2021a); Zhu et al. (2022); Jaskari et al. (2022), such as face recognition Chang et al. (2020a), image-to-image translation Upadhyay et al. (2021a), and depth completion Zhu et al. (2022). They demonstrate that model training with the uncertainty-aware loss can quantify the uncertainty in the prediction and exclude the interference of out-of-distribution pixels to make a sensible decision, and thus it has better robustness than standard networks. However, the role of this idea in image denoising has not been discussed.

3 Theoretical Contributions

Here, we introduce our twofold contributions. One is that We demonstrate the feasibility of training denoising networks with paired noisy images constructed by Cyclic Shift and indicate that the Cyclic Shift strategy can be applied to more realistic scenarios. The other is the noise-intensity-aware uncertainty estimation of the denoising network.

3.1 Cyclic Shift

In order to describe the theory more clearly, we assume that $Y = X + N$ is the noisy observation of $X \sim \mathcal{N}\left(0, \sigma_X^2\right)$, and $N \sim \mathcal{N}\left(0, \sigma_N^2\right)$ is the noise, which is consistent with the assumptions made in GAN2GAN (G2G) Cha et al. (2020). When given Y, the denoising result is the minimum MSE (MMSE) estimate of X,

$$f_{\mathrm{MMSE}}^*(Y) = \mathbb{E}(X \mid Y) = \frac{\sigma_X^2}{\sigma_X^2 + \sigma_N^2} Y. \tag{1}$$

Based on the above settings, we can get paired noisy images $Y_1 = X + N_1$ and $Y_2 = X + N_2$, in which N_1 and N_2 are two i.i.d. copies of the noise N. Therefore, for Noise2Noise (N2N), Eq.(1) is rewritten as

$$f_{\mathrm{N2N}}\left(Y_1\right) \triangleq \arg\min_f \mathbb{E}\left(Y_2 - f\left(Y_1\right)\right)^2 = \mathbb{E}\left(Y_2 \mid Y_1\right) = \mathbb{E}\left(X + N_2 \mid Y_1\right) \overset{(s)}{=} \mathbb{E}\left(X \mid Y_1\right) = \frac{\sigma_X^2}{\sigma_X^2 + \sigma_N^2} Y_1, \tag{2}$$

where (s) is inferred from Y_1 and N_2 independently of each other. However, it is challenging to get the pure and clean images in practice. Subsequently, G2G Cha et al. (2020) is proposed, which is the "Noisy" N2N. It has no X but the images X' with slight noise. Let $X' = X + N_0$, in which $N_0 \sim \mathcal{N}\left(0, \sigma_0^2\right)$, and the paired noisy observations are ${Y_1}' = X' + N_1$ and ${Y_2}' = X' + N_2$. Then, Eq. (2) can be further reformulated as:

$$f_{\text{G2G}}\left(Y_1', a\right) \triangleq \arg\min_f \mathbb{E}\left(Y_2' - f\left(Y_1'\right)\right)^2 = \mathbb{E}\left(X' \mid Y_1'\right) = \frac{\sigma_X^2(1+a)}{\sigma_X^2(1+a) + \sigma_N^2} Y_1', \tag{3}$$

where $a \triangleq \sigma_0^2/\sigma_X^2$ and $0 \leq a < 1$. It is obvious that Eq.(2) and Eq.(3) are the same when $a = \sigma_0^2 = 0$. Then, they proof that for a sufficiently large σ_0^2, $f_{\text{G2G}}(Y, a)$ gives a better estimate of X than X'. In other words, it is feasible to train the denoising network with Y_1 and Y_2.

We generalize G2G to the universal case by assuming that it is difficult to separate the noise N with the actual distribution from a single noisy image but can obtain an approximate noise N', which can be formulated as

$$Z = X' + N' = X + N_m + N - N_m, \tag{4}$$

where $X' = X + N_m$, $N' = N - N_m$, $N_m \sim \mathcal{N}\left(0, \sigma_m^2\right)$, and $N' \sim \mathcal{N}\left(0, \sigma_N^2 + \sigma_m^2\right)$. After that, we can get two i.i.d. noise sets, ${N_1}'$ and ${N_2}'$. The paired noisy images obtained by Cyclic Shift are $Z_1 = X' + {N_1}'$ and $Z_2 = X' + {N_2}'$. With these settings, the estimated value of the MMSE is

$$f_{\text{CS}}\left(Z_1, b\right) \triangleq \arg\min_f \mathbb{E}\left(Z_2 - f\left(Z_1\right)\right)^2 = \mathbb{E}\left(X' \mid Z_1\right) = \frac{\sigma_X^2(1+b)}{\sigma_X^2(1+2b) + \sigma_N^2} Z_1, \tag{5}$$

where $b \triangleq \sigma_m^2/\sigma_X^2$. We note that Eq.(5) is equal to Eq.(1) when $b = \sigma_m^2 = 0$, which is the same as the condition for the optimal value of G2G.

Theory 1. *It is feasible to train a denoising network using paired noisy images consisting of coarse-denoised images and i.i.d noise with an approximate distribution to the actual noise.*

Theory 1 motivates our approach, and the training process does not suffer severe information underutilization. From the above analysis, it can be inferred that the upper limit of our method is the result of the network training with the standard N2N datasets.

3.2 Noise-Intensity-Aware Uncertainty Estimation

Most previous studies in image denoising employ MSE/L1 loss to optimize the network. This leads to the denoiser that treats all pixel points equally and has satisfied performance in dealing with specific noise but has difficulty making the most informed decisions for unseen noise. Furthermore, modeling uncertainty can compensate for this deficiency very well. Uncertainty captures hard-to-perceive disturbances in the dataset Kendall & Gal (2017), as well as gives the network the ability to discriminate them.

Modeling the uncertainty in other computer vision tasks Kendall & Gal (2017); Chang et al. (2020a); Ning et al. (2021) is based on the following settings. Suppose that Q: $\{q_i\}_{i=1}^k$ and $g(S)(S$: $\{s_i\}_{i=1}^k)$ are the learning target and output of the network $g(\cdot)$, respectively. The detractors (*i.e.* uncertainty) in the dataset may be blurred, misaligned, slightly noisy, and so on, which can be denoted by an additive term U: $\{u_i\}_{i=1}^k$. The relationship between these three elements is usually represented as the following equation,

$$Q = g(S) + \epsilon U, \tag{6}$$

where $\epsilon \sim \mathcal{N}(0, I)$. Moreover, for a given input s_i and a corresponding target q_i, the Gaussian distribution is assumed for characterizing the likelihood function by

$$p(q_i \mid s_i) = \frac{1}{\sqrt{2\pi u_i^2}} \exp\left(-\frac{\|q_i - g(s_i)\|_2}{2u_i^2}\right). \tag{7}$$

For ease of calculation, the log-likelihood can be formulated as follows,

$$\ln p(q_i \mid s_i) = -\frac{\|q_i - g(s_i)\|_2}{2u_i^2} - \frac{1}{2}\ln u_i^2 - \frac{1}{2}\ln 2\pi. \tag{8}$$

Then, the likelihood maximization is reformulated as the minimization of the loss function,

$$\mathcal{L}_U = \frac{1}{K}\sum_{i=1}^{k}\frac{1}{2u_i^2}\left\|q_i - f\left(s_i\right)\right\|_2 + \frac{1}{2}\ln u_i^2. \tag{9}$$

where K is the number of samples in the training dataset. The network training with Eq.(9) is able to output both the denoising result $f(s_i)$ and the corresponding the uncertainty map u_i.

Different from the above settings, the uncertainty of the estimation in the image denoising task is closely related to the noise intensity. First, assuming that we have the paired noisy images $Z_1 = X' + N_1{}'$ and $Z_2 = X' + N_2{}'$. $g(Z_2)$ represents the denoised images learned by the network. Then, Eq.(6) can be formulated as

$$Z_1 = g(Z_2) + \epsilon U'. \tag{10}$$

Furthermore, we can infer that the uncertainty estimated by the network contains two components, the noise that needs to be removed and the unavoidable disturbances that exist in the dataset. Namely,

$$U' = N_1' + U, \tag{11}$$

Theory 2. *In image denoising, the uncertainty modeled and estimated has the property of noise intensity perception.*

We experimentally verify Theory 2 in Figure 2, which shows the uncertainty map corresponding to the noisy image used for testing. We train the uncertainty-based network using a dataset with Gaussian noise ($\sigma = 25$). During testing, we added two different intensities of Gaussian noise (*e.g.* $\sigma = 5$ (left) and 25 (right), first row) to the same image. The noisy images are fed into the network to obtain the corresponding uncertainty maps. It clearly shows that the network assigns higher uncertainty values to the pixels with high-intensity noise. Therefore, the uncertainty-based network can adaptively adjust the influence of pixels containing unseen noise on the denoising results according to the noise intensity, making the uncertainty-based denoising network more robust than the standard denoising network.

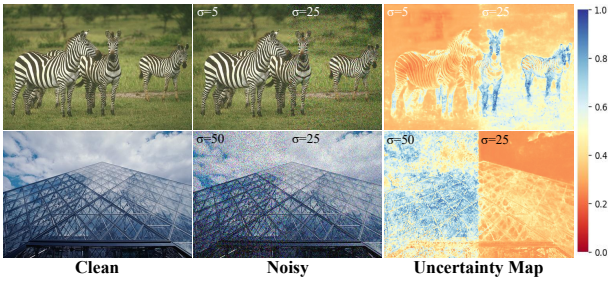


Figure 2: **Uncertainty maps visualization in BSD300**.

4 Implementation

Motivated and supported by the above theories, we propose a novel robust self-supervised image denoising method (CSU). An overview of the framework of CSU is shown in Figure 3. It consists of three steps: pre-denoising, constructing paired noisy images with Cyclic Shift, and uncertainty-aware denoising. Each of them will be described in detail below.

4.1 Step 1: Pre-denoising

First, assume that we have a group of noisy observations $Y = \{y_A, y_B, y_C, y_D\}$ from the whole noisy dataset, in which the different letters in the subscripts represent different images with i.i.d. noise. As shown in Figure 3 (Step 1), $f_{e\text{-}d}(\cdot)$ denotes the whole denoising network, which can be an arbitrary network that need to be optimized. The pre-denoised images getting from $f_{e\text{-}d}(\cdot)$ can be expressed as

$$X' = f_{e\text{-}d}(Y) = \{x'_A, x'_B, x'_C, x'_D\}. \tag{12}$$

4.2 Step 2: Constructing Paired Noisy Images with Cyclic Shift

After obtaining the pre-denoised image, we can get the noise N' by

$$N' = Y - X' = \{n'_A, n'_B, n'_C, n'_D\}, \tag{13}$$

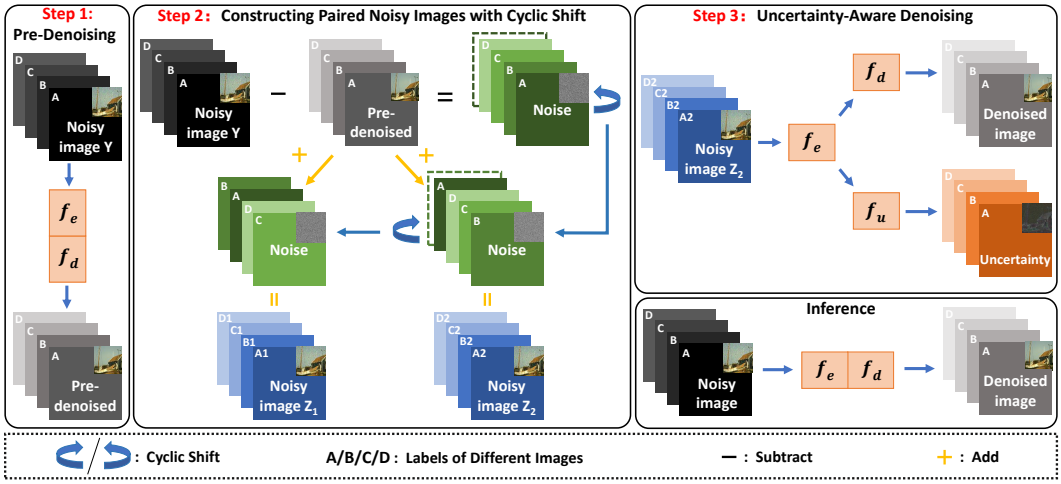


Figure 3: **The overall structure of CSU.** First, we get the pre-denoised images from the network $f_{e\text{-}d}(\cdot)$, which consists of two sub-networks, $f_e(\cdot)$ and $f_d(\cdot)$. Then, the pre-denoised images are subtracted from the noisy observations to obtain the noise. Subsequently, we obtain the paired noisy images by Cyclic Shift. Finally, we train the uncertainty-aware network $f_{e\text{-}d\text{-}u}(\cdot)$ to learn the denoising results and the uncertainty maps simultaneously. The small images in the upper right corner are a visual representation of different types of images.

where the noise patches are independent of each other. Subsequently, we fix the order within the group of X' and perform Cyclic Shift on N' twice. The detail of Cyclic Shift is to move the noise block n'_* that is in the first position in N' to the end. Then, we can get ${N_1}' = \{n'_B, n'_C, n'_D, n'_A\}$ and ${N_2}' = \{n'_C, n'_D, n'_A, n'_B\}$. The paired noisy images are

$$\begin{aligned} Z_1 &= X' + {N_1}', \\ Z_2 &= X' + {N_2}', \end{aligned} \tag{14}$$

in which

$$\begin{aligned} Z_1 &= \{z_{A1}, z_{B1}, z_{C1}, z_{D1}\}, \\ Z_2 &= \{z_{A2}, z_{B2}, z_{C2}, z_{D2}\}. \end{aligned} \tag{15}$$

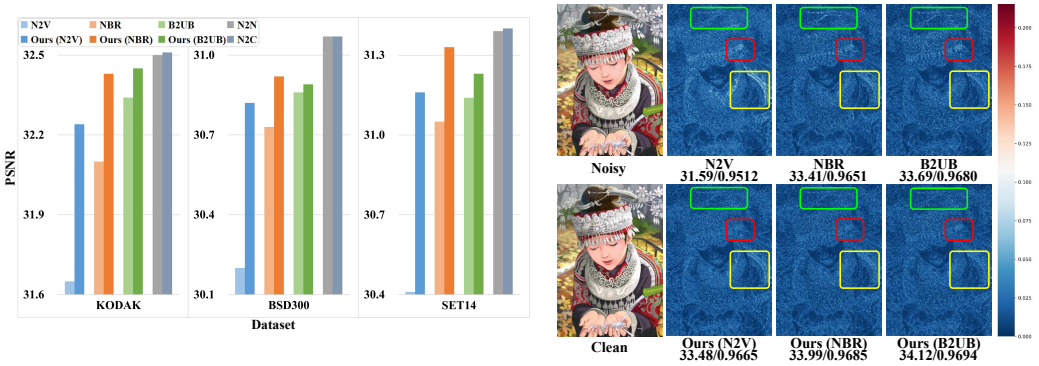
Noisy images with the same letter in the subscript form paired noisy images (*i.e.,* z_{A1} and z_{A2}) with the same pre-denoised image x'_* and i.i.d. noise.

4.3 Step 3: Uncertainty-Aware Denoising

We train the denoising network $f_{e\text{-}d\text{-}u}(\cdot)$ with Z_1 as the target and Z_2 as the input (Figure 3 (Step 3)). $f_e(\cdot)$ is used for extracting feature, and the two branches, $f_d(\cdot)$ and $f_u(\cdot)$ output the denoising results and the uncertainty maps, respectively. In addition, we define $\theta_i = \ln u_i^2$ in Eq.(9) as the output of $f_u(\cdot)$ to avoid the existence of zero and negative values that cause the network to crash due to the inability to calculate the logarithm with the following loss function,

$$\mathcal{L}_{NIAU} = \frac{1}{K}\sum_{i=1}^{k} \exp(-\theta_i) \left\| z_{1(i)} - f_{e\text{-}d}\left(z_{2(i)}\right) \right\|_2 + \lambda\theta_i, \tag{16}$$

where we set $\lambda = 2$ according to the Jeffreys prior Figueiredo (2001) to increase the sparsity of the uncertainty maps. In this way, $f_{e\text{-}d\text{-}u}(\cdot)$ can discriminate unseen noise or distractors presented in the noisy images, thus giving the model better robustness. After training, we use $f_{e\text{-}d}(\cdot)$ for testing (Figure 3 (Inference)).



(a) Effect of Cyclic Shift. (b) Visual comparison.

Figure 4: **(a) The ability of Cyclic Shift to significantly improve other denoising methods on three datasets with Gaussian noise** ($\sigma = [5, 50]$). Ours$(\cdot)$ indicates the improved results, *i.e.*, "Ours (N2V)" means that the N2V method uses the Cyclic Shift to construct sample pairs for training. Note that N2N and N2C are the upper limit of all self-supervised denoising methods. **(b) Visual comparison of the improved denoising methods by Cyclic Shift on SET14 with Gaussian noise** ($\sigma = [5, 50]$). For better viewing, we show the difference between the denoising results and Clean in colour.

5 Experiments

5.1 Experimental Settings

Datasets and Noise Type. For the experiments in sRGB space, ImageNet Deng et al. (2009) validation dataset is the training dataset. KODAK Rich Franzen (1999), BSD300 Martin et al. (2001), and Set14 Zeyde et al. (2010) are the testing datasets. In grayscale experiments, we use the noisy images from BSD400 Zhang et al. (2017) for training and Set12, BSD68 Roth & Black (2005) and Urban100 Huang et al. (2015) for testing. Moreover, two types of Gaussian noise ($\sigma = 25$ and $\sigma \in [0, 55]$) are considered in the above two experiments and the correlated noise is considered in grayscale experiments. Besides, we take the raw-RGB images in the real-world from SIDD Abdelhamed et al. (2018) Medium dataset for training and the SIDD validation dataset for testing.

Training Details. The modified U-Net Wang et al. (2022) architecture is used for the experiments on sRGB and raw-RGB datasets. We set batch size as 4 and use Adam Kingma & Ba (2014) as the optimizer with a weight decay of $1e^{-8}$ to avoid overfitting. The initial learning rate is $3e^{-4}$ and $1e^{-4}$ for sRGB and raw-RGB denoising, respectively, and they decrease by half every 20 epochs for 100 training epochs. When training the uncertainty-aware denoising network, we set the learning rate as $8e^{-5}$ for the sRGB experiments and $6e^{-5}$ for the raw-RGB experiments. In the experiments with grayscale images, we train the DnCNN Zhang et al. (2017) with an initial learning rate of $3e^{-4}$. The learning rate decreases by half every 30 epochs for 300 training epochs and remains constant after 120 epochs.

5.2 Cyclic Shift Improves SOTA Denoising Methods

Here, we demonstrate that Cyclic Shift can significantly improve the denoising performance of SOTA methods. We choose three representative SOTA denoising methods, N2V, NBR, and B2UB, where B2UB has achieved SOTA performance recently. Figure 4 (a) clearly shows that the Cyclic Shift significantly improves the denoising performance of the three methods. These results well demonstrate the effectiveness of Cyclic Shift. Note that Cyclic Shift does not requiring any additional computation when embedding into existing denoising methods. The visual comparison of the improved denoising methods with the original methods is shown in Figure 4 (b). It intuitively demonstrates that the Cyclic Shift can improve the denoising performance for most of the self-supervised denoising methods, especially the N2V-like models and the models that cannot synthesize aligned paired noisy images like NBR.

Table 1: Quantitative denoising results (PSNR(dB)/SSIM) of different methods on sRGB dataset. Bolded red and bolded black represents the highest and second highest results, respectively.

Noise Type	Dataset	Method										Upper Bound	
		CBM3D	S2S	N2V	Nr2N	DBSN	R2R	NBR	B2UB	Ours (NBR)	Ours (B2UB)	N2N	N2C
Gaussian $\sigma = 25$	KODAK	31.87/ 0.868	31.28/ 0.864	30.32/ 0.821	30.70/ 0.845	31.64/ 0.856	32.25/ 0.880	32.08/ 0.879	32.27/ 0.880	**32.28/ 0.882**	**32.35/ 0.883**	32.41/ 0.884	32.43/ 0.884
	BSD300	30.48/ 0.861	29.86/ 0.849	29.34/ 0.824	29.32/ 0.833	29.80/ 0.839	30.91/ 0.872	30.79/ 0.873	30.87/ 0.872	**30.93/ 0.877**	**30.90/ 0.875**	31.04/ 0.878	31.05/ 0.879
	SET14	30.88/ 0.854	30.08/ 0.839	28.84/ 0.802	29.64/ 0.832	30.63/ 0.846	**31.32/ 0.865**	31.09/ **0.864**	**31.27/ 0.864**	31.10/ **0.865**	31.07/ **0.864**	31.37/ 0.868	31.40/ 0.869
Gaussian $\sigma \in [5,50]$	KODAK	32.02/ 0.860	31.37/ 0.860	30.44/ 0.806	-	30.38/ 0.826	31.50/ 0.850	32.10/ 0.870	32.34/ 0.872	**32.43/ 0.874**	**32.45/ 0.875**	32.50/ 0.875	32.51/ 0.875
	BSD300	30.56/ 0.847	29.87/ 0.841	29.31/ 0.801	-	28.34/ 0.788	30.56/ 0.855	30.73/ 0.861	30.86/ 0.861	**30.92/ 0.864**	**30.89/ 0.863**	31.07/ 0.866	31.07/ 0.866
	SET14	30.94/ 0.849	29.97/ 0.849	29.01/ 0.792	-	29.49/ 0.814	30.84/ 0.850	31.05/ 0.858	31.14/ 0.857	**31.33/ 0.863**	**31.24/ 0.861**	31.39/ 0.863	31.41/ 0.863

Table 2: Experimental results (PSNR(dB)/SSIM) on three grayscale datasets. All methods in the table are based on DnCNN. The B2UB method was not included in the comparison because the official code runs poorly on grayscale images for denoising.

Noise Type	Dataset	Method				Upper Bound	
		N2V	NBR	Ours(N2V)	Ours(NBR)	N2N	N2C
Gaussian $\sigma = 25$	Urban100	26.82/0.793	**27.85**/0.815	27.67/**0.818**	**28.26/0.833**	28.86/0.854	28.91/0.859
	BSD68	27.27/0.746	**28.15**/0.774	27.94/**0.778**	**28.31/0.793**	28.63/0.804	28.68/0.809
	SET12	28.48/0.792	**29.16**/0.802	29.05/**0.808**	**29.38/0.815**	29.79/0.834	29.89/0.838
Gaussian $\sigma \in [5, 50]$	Urban100	26.12/0.732	26.80/0.746	**26.82/0.763**	**27.10/0.774**	27.22/0.777	27.74/0.794
	BSD68	27.25/0.722	**27.96**/0.743	27.78/**0.746**	**28.19/0.762**	28.40/0.766	28.72/0.781
	SET12	28.11/0.753	28.57/0.766	**28.65/0.774**	**28.89/0.788**	29.08/0.790	29.54/0.814

5.3 Comparison with State-of-the-art Methods

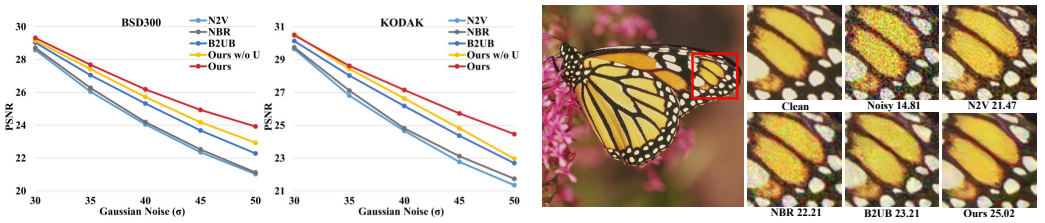
We compare our CSU against eight self-supervised denoising methods, Self2Self (S2S) Quan et al. (2020), Noise2Void (N2V) Krull et al. (2019), Laine19 (L19) Laine et al. (2019), Noisier2Noise (Nr2N) Moran et al. (2020), DBSN Wu et al. (2020), R2R Pang et al. (2021), NBR2NBR (NBR) Huang et al. (2021), and Blind2Unblind (B2UB) Wang et al. (2022), two upper bound methods, supervised denoising (N2C) and Noise2Noise (N2N) Lehtinen et al. (2018), and a traditional denoiser BM3D Dabov et al. (2007).

Results on sRGB Dataset. The quantitative comparison results are shown in Table 1. Obviously, our method significantly outperforms all other methods in the experiments with Gaussian noise $\sigma = 25$ and $\sigma \in [5, 50]$, and are comparable to the two upper bound methods. The reason for the improved denoising performance is that training the denoising network with pairs of noisy images constructed using cyclic shifting avoids the problems of information underutilization and pixel misalignment. We find that our method performs better based on NBR than based on B2UB on BSD300 and Set14. The reason for this phenomenon is that B2UB empirically fixes the weighted average hyperparameter during the training process, which leads to a bias in the noise learned by the network. This further leads to the paired noisy images synthesized by Cyclic Shift containing noise that deviates from the actual distribution. For NBR, the degradation of denoising performance is mainly caused by the misalignment of paired noisy images constructed by sub-sampling, which can be well addressed by our method.

Results on Grayscale Dataset. In the grayscale images denoising experiments, we retrain N2C, N2N, and two self-supervised methods, N2V and NBR, based on DnCNN Zhang et al. (2017). The quantitative denoising results are shown in Table 2. The results demonstrate that the proposed method achieves the best denoising performance. Moreover, we find that the denoising performance of ours CSU based on N2V is better than that of NBR in SSIM metric. We analyze the reasons for this phenomenon are twofold. 1) CSU compensates for this severe problem of missing information in the blind-spot-based methods. 2) There is a pixel bias in the paired noisy images obtained by sub-sampling in NBR, resulting in the over-smoothed results of the denoising network. In addition, to demonstrate that our method not only achieves excellent performance in removing Gaussian noise but also in handling noisy datasets with spatially correlated noise, we synthesize the noisy images

Table 3: Denoising results (PSNR(dB)/SSIM) on SIDD validation dataset in raw-RGB space.

Dataset	Method											Upper Bound	
	BM3D	N2V	L19-mu (Gaussian)	L19-pme (Gaussian)	L19-mu (Poisson)	L19-pme (Poisson)	DBSN	R2R	NBR	B2UB	**Ours (NBR)**	N2N	N2C
SIDD Validation	48.92/ 0.986	48.55/ 0.984	50.44/ 0.990	42.87/ 0.939	50.89/ 0.990	48.98/ 0.985	50.13/ 0.988	47.20/ 0.980	51.06/ **0.991**	**51.36/ 0.992**	**51.07/ 0.991**	51.21/ 0.991	51.19/ 0.991



(a) Comparison of the robustness. (b) Visual comparison.

Figure 5: **(a) Comparison of the robustness of different methods on KODAK and BSD300.** All methods are trained with noisy images containing Gaussian noise ($\sigma = 25$) and tested with noisy images containing different intensities of Gaussian noise. **(b) Visual comparison of the robustness of different denoising methods on SET14** (bottom values is PSNR). It shows the denoising results of these methods, trained with noisy images containing Gaussian noise ($\sigma = 25$) and tested with noisy images containing Gaussian noise ($\sigma = 50$).

using the way proposed by GAN2GAN Cha et al. (2020). The experimental results and details can be viewed in the Appendix.

Results on Raw-RGB Dataset. Table 3 indicates that our method achieves the second-best denoising performance. It is to be expected because the theoretical upper limit of our method is N2N, which is analyzed in the theoretical contributions section. We note that the denoising result of B2UB is higher than N2N and N2C, which is caused by interpolating the noisy image before sampling in B2UB. This approach is essentially a data enhancement technique and provides the network with more available information. Moreover, the SIDD dataset contains multiple noisy images of the same scene, which somewhat compensates for the lack of information during training.

5.4 Uncertainty Enhances Robustness of Denoising

We use the following strategy to investigate the robustness of different denoising models. The denoising network is trained using the noisy images with fixed Gaussian noise ($\sigma = 25$), and the denoising performance is quantified by providing images with different intensities of Gaussian noise. There are five methods for comparison N2V, NBR, B2UB, "Ours w/o U", and Ours (CSU). "U" denotes the uncertainty loss $\mathcal{L}_{NIAU}$ and "Ours w/o U" indicates that our method uses the MSE loss instead of $\mathcal{L}_{NIAU}$ when training the uncertainty-aware denoising network. As shown in Figure 5 (a), our method models and estimates the uncertainty in the dataset resulting in a smoother degradation trend for the denoised network than other methods, which means that the model can make informed choices in the face of unseen noise. Moreover, the visual comparison of different methods are shown in Figure 5 (b), which shows that our model incorporating uncertainty performs better than others in dealing with unseen noise.

6 Conclusion

We propose a novel strategy, Cyclic Shift, to construct sample pairs for self-supervised image denoising learning. It can extensively boost existing self-supervised denoising methods without additional computational cost and does not rely on complex network structures, making it more practical. In addition, we introduce the uncertainty-aware loss to improve the perception of noise intensity as well as the robustness of the denoising network. By integrating both them, our CSU method achieves the state-of-the-art denoising performance on multiple datasets.

REFERENCES

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*, pp. 1692–1700, 2018.

Saeed Anwar and Nick Barnes. Real image denoising with feature attention. In *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 3155–3164, 2019.

Joshua Batson and Loic Royer. Noise2self: Blind denoising by self-supervision. In *International Conference on Machine Learning*, pp. 524–533. PMLR, 2019.

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*, 2020.

Jie Chang, Zhonghao Lan, Changmao Cheng, and Yichen Wei. Data uncertainty learning in face recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 5710–5719, 2020a.

Meng Chang, Qi Li, Huajun Feng, and Zhihai Xu. Spatial-adaptive network for single image denoising. In *European Conference on Computer Vision*, pp. 171–187. Springer, 2020b.

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*, pp. 3155–3164, 2018.

Shen Cheng, Yuzhi Wang, Haibin Huang, Donghao Liu, Haoqiang Fan, and Shuaicheng Liu. Nbnet: Noise basis learning for image denoising with subspace projection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 4896–4906, 2021.

Jiwoong Choi, Dayoung Chun, Hyun Kim, and Hyuk-Jae Lee. Gaussian yolov3: An accurate and fast object detector using localization uncertainty for autonomous driving. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 502–511, 2019.

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.

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*, pp. 248–255. Ieee, 2009.

Mário Figueiredo. Adaptive sparseness using jeffreys prior. *Advances in neural information processing systems*, 14, 2001.

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*, pp. 1712–1722, 2019.

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*, volume 34, pp. 4140–4149, 2020.

Jia-Bin Huang, Abhishek Singh, and Narendra Ahuja. Single image super-resolution from transformed self-exemplars. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 5197–5206, 2015.

Tao Huang, Songjiang Li, Xu Jia, Huchuan Lu, and Jianzhuang Liu. Neighbor2neighbor: Self-supervised denoising from single noisy images. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 14781–14790, 2021.

Geonwoon Jang, Wooseok Lee, Sanghyun Son, and Kyoung Mu Lee. C2n: Practical generative noise modeling for real-world denoising. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 2350–2359, 2021.

Joel Jaskari, Jaakko Sahlsten, Theodoros Damoulas, Jeremias Knoblauch, Simo Särkkä, Leo Kärkkäinen, Kustaa Hietala, and Kimmo K. Kaski. Uncertainty-aware deep learning methods for robust diabetic retinopathy classification. *IEEE Access*, 10:76669–76681, 2022. doi: 10.1109/ACCESS.2022.3192024.

Alex Kendall and Yarin Gal. What uncertainties do we need in bayesian deep learning for computer vision? *Advances in neural information processing systems*, 30, 2017.

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

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*, pp. 2129–2137, 2019.

Samuli Laine, Tero Karras, Jaakko Lehtinen, and Timo Aila. High-quality self-supervised deep image denoising. *Advances in Neural Information Processing Systems*, 32, 2019.

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/CVF Conference on Computer Vision and Pattern Recognition*, pp. 17725–17734, 2022.

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.

Huangxing Lin, Yihong Zhuang, Yue Huang, Xinghao Ding, Xiaoqing Liu, and Yizhou Yu. Noise2grad: Extract image noise to denoise. In *IJCAI*, pp. 830–836, 2021.

Ruijun Ma, Shuyi Li, Bob Zhang, and Zhengming Li. Generative adaptive convolutions for real-world noisy image denoising. 2022.

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, pp. 416–423. IEEE, 2001.

Nick Moran, Dan Schmidt, Yu Zhong, and Patrick Coady. Noisier2noise: Learning to denoise from unpaired noisy data. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 12064–12072, 2020.

Qian Ning, Weisheng Dong, Xin Li, Jinjian Wu, and Guangming Shi. Uncertainty-driven loss for single image super-resolution. *Advances in Neural Information Processing Systems*, 34, 2021.

Tongyao Pang, Huan Zheng, Yuhui Quan, and Hui Ji. Recorrupted-to-recorrupted: unsupervised deep learning for image denoising. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 2043–2052, 2021.

Tobias Plötz and Stefan Roth. Neural nearest neighbors networks. *Advances in Neural information processing systems*, 31, 2018.

Yuhui Quan, Mingqin Chen, Tongyao Pang, and Hui Ji. Self2self with dropout: Learning self-supervised denoising from single image. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 1890–1898, 2020.

Rich Franzen. Kodak lossless true color image suite. `http://r0k.us/graphics/kodak`, 1999.

Stefan Roth and Michael J Black. Fields of experts: A framework for learning image priors. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, volume 2, pp. 860–867. IEEE, 2005.

Viswanath P Sudarshan, Uddeshya Upadhyay, Gary F Egan, Zhaolin Chen, and Suyash P Awate. Towards lower-dose pet using physics-based uncertainty-aware multimodal learning with robustness to out-of-distribution data. *Medical Image Analysis*, 73:102187, 2021.

Uddeshya Upadhyay, Yanbei Chen, and Zeynep Akata. Robustness via uncertainty-aware cycle consistency. *Advances in Neural Information Processing Systems*, 34:28261–28273, 2021a.

Uddeshya Upadhyay, Viswanath P Sudarshan, and Suyash P Awate. Uncertainty-aware gan with adaptive loss for robust mri image enhancement. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 3255–3264, 2021b.

Zejin Wang, Jiazheng Liu, Guoqing Li, and Hua Han. Blind2unblind: Self-supervised image denoising with visible blind spots. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 2027–2036, 2022.

Xiaohe Wu, Ming Liu, Yue Cao, Dongwei Ren, and Wangmeng Zuo. Unpaired learning of deep image denoising. In *European conference on computer vision*, pp. 352–368. Springer, 2020.

Jun Xu, Yuan Huang, Ming-Ming Cheng, Li Liu, Fan Zhu, Zhou Xu, and Ling Shao. Noisy-as-clean: Learning self-supervised denoising from corrupted image. *IEEE Transactions on Image Processing*, 29:9316–9329, 2020.

Roman Zeyde, Michael Elad, and Matan Protter. On single image scale-up using sparse-representations. In *International conference on curves and surfaces*, pp. 711–730. Springer, 2010.

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.

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.

Yi Zhang, Dasong Li, Ka Lung Law, Xiaogang Wang, Hongwei Qin, and Hongsheng Li. Idr: Self-supervised image denoising via iterative data refinement. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 2098–2107, 2022.

Yufan Zhu, Weisheng Dong, Leida Li, Jinjian Wu, Xin Li, and Guangming Shi. Robust depth completion with uncertainty-driven loss functions. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, pp. 3626–3634, 2022.

A EXPERIMENTAL RESULTS ON CORRELATED NOISE

For correlated noise, we generated the following noise for each ℓ-th pixel using the rules proposed by GAN2GAN Cha et al. (2020),

$$N_\ell = \eta M_\ell + (1-\eta)\left(\frac{1}{\sqrt{|\mathcal{NB}_\ell|}}\sum_{m\in\mathcal{NB}_\ell} M_m\right), \ell = 1, 2, \ldots \tag{17}$$

in which $\{M_\ell\}$ are white Gaussian $\mathcal{N}(0,\sigma^2)$, $\mathcal{NB}_\ell$ is the $k \times k$ neighborhood patch except for the pixel ℓ, and η is a mixture parameter. We set $\eta = 1/\sqrt{2}$ such that the marginal distribution of N_ℓ is also $\mathcal{N}(0,\sigma^2)$ and set $k = 16$. Note in this case, N_ℓ has a spatial correlation, and we tested with $\sigma = 15$. The results of six denoising methods are shown in Table 4. The experimental results illustrate that our method has better denoising performance in dealing with spatially correlated noise than other self-supervised denoising methods.

Table 4: Experimental results (PSNR(dB)/SSIM) on three grayscale datasets. All methods in the table are based on DnCNN. The B2UB method was not included in the comparison because the official code runs poorly on grayscale images for denoising. Bolded red and bolded black represents the highest and second highest results, respectively.

Noise Type	**Dataset**	**Method**				**Upper Bound**	
		N2V	NBR	**Ours(N2V)**	**Ours(NBR)**	N2N	N2C
Correlated $\sigma = 15$	Urban100	28.73/0.8708	**30.05**/0.8783	29.97/**0.8893**	**30.52/0.8909**	31.48/0.9103	31.42/0.9072
	BSD68	28.73/0.8410	**30.01**/0.8679	29.96/**0.8732**	**30.25/0.8776**	31.20/0.9032	31.23/0.9053
	SET12	29.98/0.8682	30.62/0.8783	**30.75/0.8857**	**30.82/0.8914**	31.90/0.9019	31.93/0.9044

B ABLATION STUDY

Table 5 shows the influence of our two modules, Cyclic Shift (C) and $\mathcal{L}_{NIAU}$ (U) in sRGB space. "Ours w/o C, U" means the denoising network for getting pre-denoised images. The data in Table 5 indicates that the improvement of the network denoising performance is mainly due to the paired noisy images with the same scene synthesized by Cyclic Shift, and $\mathcal{L}_{NIAU}$ has a slight improvement in the denoising performance.

Table 5: Ablation study on the two components Cycle Shift (C) and $\mathcal{L}_{NIAU}$ (U) in the sRGB space with Gaussian noise $\sigma = 25$. Bolded black means the best PSNR/SSIM.

Method	**KODAK**	**BSD300**	**SET14**
Ours w/o C, U	32.265/0.8796	30.809/0.8742	30.968/0.8635
Ours w/o U	32.348/0.8828	30.929/0.8760	31.099/0.8643
Ours	**32.350/0.8831**	**30.933/0.8767**	**31.102/0.8647**

C VISUAL COMPARISON

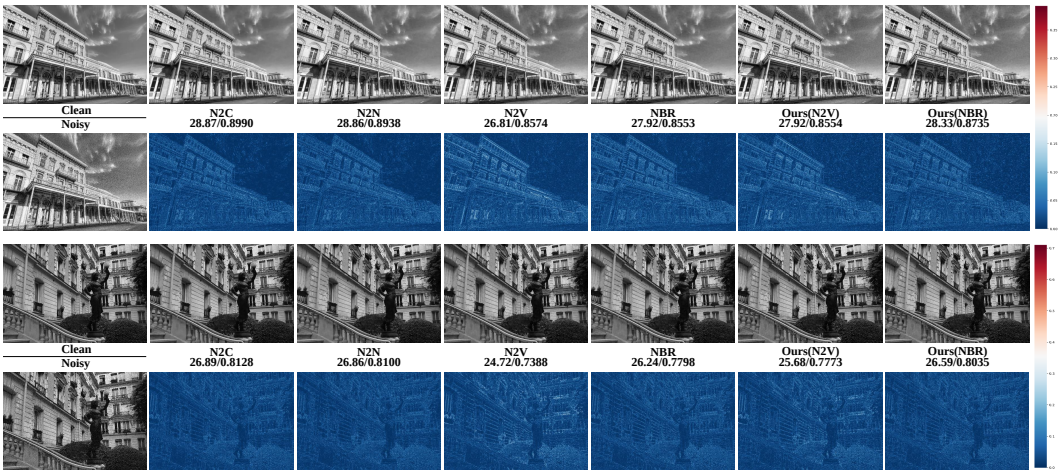


Figure 6: Visual comparison of denoising grayscale images in the setting of Gaussian noise ($\sigma = 25$) on Urban100. For better viewing, we show the difference between the denoising results and Clean in colour.

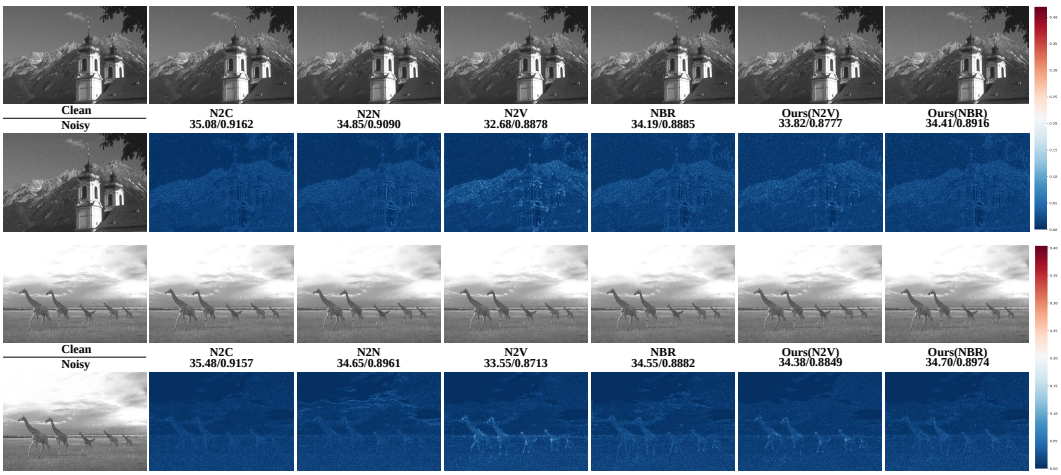


Figure 7: Visual comparison of denoising grayscale images in the setting of Gaussian noise ($\sigma \in [5, 50]$) on BSD68. For better viewing, we show the difference between the denoising results and Clean in colour.

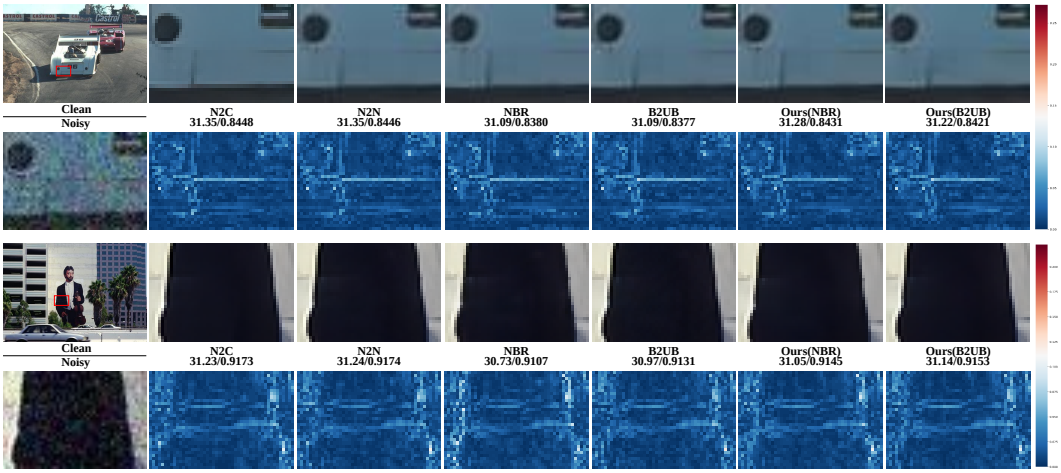


Figure 8: Visual comparison of denoising sRGB images in the setting of Gaussian noise ($\sigma = 25$) on BSD300. For better viewing, we show the difference between the denoising results and Clean in colour.

Figure 9: Visual comparison of denoising sRGB images in the setting of Gaussian noise ($\sigma \in [5, 50]$) on SET14. For better viewing, we show the difference between the denoising results and Clean in colour.