Blind2Unblind: Self-Supervised Image Denoising with Visible Blind Spots

Zejin Wang[1,2] Jiazheng Liu[1,3] Guoqing Li[1] Hua Han[1,3,*]
[1]National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences
[2]School of Artificial Intelligence, University of Chinese Academy of Sciences
[3]School of Future Technology, University of Chinese Academy of Sciences

Abstract

Real noisy-clean pairs on a large scale are costly and difficult to obtain. Meanwhile, supervised denoisers trained on synthetic data perform poorly in practice. Self-supervised denoisers, which learn only from single noisy images, solve the data collection problem. However, self-supervised denoising methods, especially blindspot-driven ones, suffer sizable information loss during input or network design. The absence of valuable information dramatically reduces the upper bound of denoising performance. In this paper, we propose a simple yet efficient approach called Blind2Unblind to overcome the information loss in blindspot-driven denoising methods. First, we introduce a global-aware mask mapper that enables global perception and accelerates training. The mask mapper samples all pixels at blind spots on denoised volumes and maps them to the same channel, allowing the loss function to optimize all blind spots at once. Second, we propose a re-visible loss to train the denoising network and make blind spots visible. The denoiser can learn directly from raw noise images without losing information or being trapped in identity mapping. We also theoretically analyze the convergence of the re-visible loss. Extensive experiments on synthetic and real-world datasets demonstrate the superior performance of our approach compared to previous work. Code is available at https://github.com/demonsjin/Blind2Unblind.

1. Introduction

Image denoising, an essential task of low-level image processing, aims to remove noise and restore a clean image. In vision applications, the quality of denoising significantly affects the performance of downstream tasks, such as super-resolution [16], semantic segmentation [22], and object detection [31]. In addition, the denoiser can significantly improve the quality of images captured by mobile phones and other devices, reflecting a broad demand in imaging fields.

With the development of neural networks, learning-based denoisers [3, 6, 13, 29, 35, 36, 38, 39] have recently shown superior performance than traditional methods [5, 8, 9, 12]. However, supervised denoisers, *e.g.*, U-Net [29], DnCNN [38], FFDNet [39], RIDNet [3], SANet [6], rely on numerous noisy-clean pairs, which are costly and hard to collect. The performance of denoisers drops dramatically once processing unknown noise patterns. Lehtinen *et al*. [21] then propose to recover clean signals directly from corrupted image pairs. Using corrupted pairs reduces the difficulty of data collection but remains challenging for dynamic scenes with deformations and image quality variations.

To alleviate the above limitations, self-supervised denoising [4, 15, 18, 20, 26, 32, 33] that learns from a single noisy image has attracted much interest from researchers. Ulyanov *et al*. [32] learn deep prior only from a single noisy image. Namely, each degraded image has to be trained from scratch. Manual masking, *e.g.*, Noise2Self [4], Noise2Void [18], avoids custom denoising for each image. Since blind spots on the inputs occupy a large area, the receptive field of predicted pixels loses much valuable context, resulting in poor performance. Moreover, optimizing partial pixels in each iteration causes slow convergence. Laine *et al*. [20] design a blind spot network to bound the receptive field in four directions instead of manual masking. Masked convolution accelerates training and increases the receptive field to all areas except the blind spot. Similarly, the dilated blindspot network [33] sets blindspots on the receptive field without masking the inputs. Regardless of masked input or blind-spot networks, lower accuracy limits practical applications. Bayesian estimation [19, 20, 33] is used for explicit noise modeling as post-processing. However, noise modeling performs poorly on real-world data with complex patterns. Some works [25, 34] perform noise reduction for noisier-noise pairs even though the additional noise increases the information loss and requires that the extra noise has the same distribution as the original one. Subsequently, Pang *et al*. [26] develop a data augmentation technique with the known noise level to address the overfit-

*Corresponding author

ting caused by the absence of truth images. Recently, Huang *et al*. [15] propose to train the network with training pairs sub-sampled from the same noisy image. However, using sub-sampling pairs for supervision lead to over smoothing as neighboring pixels are approximated.

In this paper, we propose Blind2Unblind, a novel self-supervised denoising framework that overcomes the above limitations. Our framework consists of a global-aware mask mapper based on mask-driven sampling and a training strategy without blind spots based on re-visible loss. Specifically, we divide each noisy image into blocks and set specific pixels in each block as blind spots, so that we can obtain a global masked volume as input, which consists of a set of images with order masks. Then, the volume with global masks is fed into the network in the same batch. The global mapper samples denoised volumes at blind spots and projects them onto the same plane to generate denoised images. The operation speeds up training, enables global optimization, and allows the application of re-visible loss. However, masked images result in a sizable loss of valuable information, severely reducing the upper bound of denoising performance. Therefore, we consider learning from raw noisy images without masks and relief from identity mapping. Furthermore, the intermediate medium of gradient update must be introduced since raw noisy images cannot participate in backpropagation during training. We assume that masked images serve as a medium and propose a re-visible loss to enable the transition from blind-spot denoising to non-blind denoising. The proposed self-supervised denoising framework, which does not involve noise model prior or the removal of valuable information, shows surprising performance. Moreover, advanced models can be applied to our proposed method.

The contributions of our work are as follows:

1. We propose a novel self-supervised denoising framework that makes blind spots visible, without sub-sample, noise model priors and identity mapping.
2. We provide a theoretical analysis of re-visible loss and present its upper and lower bounds on convergence.
3. Our approach shows superior performance compared with state-of-the-art methods, especially on real-world datasets with complex noise patterns.

2. Related Work

2.1. Non-Learning Image Denoising

Non-learning denoising methods such as BM3D [9], CBM3D [8], WNNM [12], and NLM [5] usually iteratively perform custom denoising for each noisy image. However, since the denoising procedure is iteratively adjusted and the hyperparameters are limited, these methods suffer from long inference time and poor performance.

2.2. Supervised Image Denoising

Recently, many supervised approaches for image denoising have been developed. Zhang *et al*. [38], the pioneer in deep image denoising, first proposes DnCNN to process unknown noise levels using noisy-clean pairs as supervision. Then, U-Net [29] becomes a common denoiser based on multi-scale features. After that, more advanced denoisers [3, 6, 13, 35, 36, 39] are proposed to improve denoising performance under supervision. However, collecting noise-clean pairs remains challenging in practice since the cost of acquisition, deformation, and contrast variations. Moreover, supervision with a strong prior often performs poorly when noise patterns are unknown.

2.3. Self-Supervised Image Denoising

Noise2Noise [21] uses noisy-noisy pairs for training, which reduces the difficulty of data collection. Then, Noise2Self [4] and Noise2Void [18] propose masked schemes for denoising on individual noisy images. Since some areas on noisy images are blind, the accuracy of denoising is low. Laine19 [20], DBSN [33], and Probabilistic Noise2Void [19] transfer the masking procedure to the feature level for larger receptive fields, which reduces valuable information loss. Moreover, noise model priors are also introduced as post-processing to improve performance. Self2Self [28] trains a dropout denoiser on the pair generated by the Bernoulli sampler and averages the predictions of multiple instances. Noisy-As-Clean [34] and Noisier2Noise [25] introduce added noise to train the denoiser and require a known noise distribution, limiting their use in practice. Similarly, R2R [26] also corrupts noisy images with known noise levels to obtain training pairs. Recently, Neighbor2Neighbor (NBR2NBR) [15] obtains the noise pair for training by sub-sampling the noise image. However, approximating neighbor pixels leads to over smoothing, and sub-sampling destroys the structural continuity.

3. Theoretical Framework

3.1. Motivation

Classical blindspot denoising methods, *e.g*., masked inputs [4, 18, 19] and blindspot networks [20, 33], use an artificial masking scheme to form blind-noisy pairs. However, since the mask prior is suboptimal and lossy, their performance is severely limited. Intuitively, denoising a raw noisy image without blind spots can solve the performance degradation. Given a single raw noisy image, we assume that the model can perform denoising without losing valuable information. The only thing is to teach the model how to learn and what to learn. Since the model now has to learn denoising from the raw noisy image without blind spots, eliminating identity mapping is critical. We consider that using the masked input as a medium for updating the gradi-

ent prevents identity mapping. The challenge now remains to design a novel loss that associates blind spot denoising with raw noisy image denoising.

3.2. Noise2Void Revisit

Noise2Void [18] is a self-supervised denoising method using a training scheme that does not require noisy image pairs or clean target images. The approach only requires a single noisy image and then applies a blind-spot masking scheme to generate blind-noisy pairs for training. Given the noisy observation $\mathbf{y}$ of the ground truth $\mathbf{x}$, Noise2Void aims to minimize the following empirical risk:

$$\arg\min_{\theta} \mathbb{E}_{\mathbf{y}} \|f_\theta(\mathbf{y}_{\mathrm{RF}(i)}) - \mathbf{y}_i\|_2^2, \tag{1}$$

where $f_\theta(\cdot)$ denotes the denoising model with parameter θ, $\mathbf{y}_{\mathrm{RF}(i)}$ is a patch around pixel i, $\mathbf{y}_i$ means a single pixel i located at the patch center. This method assumes that noise is context-independent while signal relies on surrounding pixels. The blind spot architecture takes advantage of the above properties and thus relief itself from identity mapping.

3.3. Re-Visible without Identity Mapping

The blind spot scheme [4, 18] can use less information for its prediction. Therefore, its accuracy is lower than that of the normal network. The challenge here is how to convert the invisible blind spots into visible ones. In this way, we can use the structure of blind spots for self-supervised denoising and then use all the information to improve its performance. First, we see the loss in the multi-task form of denoising as follows:

$$\arg\min_{\theta} \mathbb{E}_{\mathbf{y}} \|h(f_\theta(\mathbf{\Omega}_{\mathbf{y}})) - \mathbf{y}\|_2^2 + \lambda \cdot \|f_\theta(\mathbf{y}) - \mathbf{y}\|_2^2, \tag{2}$$

note that $\mathbf{\Omega}_{\mathbf{y}}$ denotes a noisy masked volume in which the contained blind spots are exactly at all positions of $\mathbf{y}$, $h(\cdot)$ is a global-aware mask mapper that samples the denoised pixels where the blind spots are located, and λ is a constant. Using Eq. (2) as the objective function for training learns the identity. That is, $f_\theta(\mathbf{y})$ cannot explicitly participate in backpropagation. We hope this element can implicitly participate in the gradient update and realize the transition from blind to non-blind. We reformulate Eq. (2) in the 1-norm form:

$$\arg\min_{\theta} \mathbb{E}_{\mathbf{y}} \|h(f_\theta(\mathbf{\Omega}_{\mathbf{y}})) - \mathbf{y}\|_1 + \lambda \cdot \|\hat{f}_\theta(\mathbf{y}) - \mathbf{y}\|_1, \tag{3}$$

where $\hat{f}_\theta(\cdot)$ means that they do not contribute to updating the gradient. That is, $\hat{f}_\theta(\mathbf{y})$ can be considered as an ordinary constant. It only remains to design a new objective function and the derivative of $f_\theta(\mathbf{\Omega}_{\mathbf{y}})$ should contain $\hat{f}_\theta(\mathbf{y})$ to satisfy the non-blind requirement. We assume that the quadratic operation on Eq. (3) can achieve the implicit goal of re-visible and satisfy the average arithmetic form of the optimal solution for the denoising task, which then becomes:

$$\arg\min_{\theta} \mathbb{E}_{\mathbf{y}} \||h(f_\theta(\mathbf{\Omega}_{\mathbf{y}})) - \mathbf{y}| + \lambda \cdot |\hat{f}_\theta(\mathbf{y}) - \mathbf{y}|\|_2^2. \tag{4}$$

Direct application of Eq. (4) as the objective function is not appropriate, since the symbol of different terms reflects opposite optimization objectives. Specifically, we extend Eq. (4) as follows:

$$\begin{aligned} \mathcal{T}(\mathbf{y}) &= \||h(f_\theta(\mathbf{\Omega}_{\mathbf{y}})) - \mathbf{y}| + \lambda \cdot |\hat{f}_\theta(\mathbf{y}) - \mathbf{y}|\|_2^2 \\ &= \|h(f_\theta(\mathbf{\Omega}_{\mathbf{y}})) - \mathbf{y}\|_2^2 + \lambda^2 \|\hat{f}_\theta(\mathbf{y}) - \mathbf{y}\|_2^2 \\ &\quad + 2\lambda \|(h(f_\theta(\mathbf{\Omega}_{\mathbf{y}})) - \mathbf{y})^T (\hat{f}_\theta(\mathbf{y}) - \mathbf{y})\|_1. \end{aligned} \tag{5}$$

Let $d_1 = h(f_\theta(\mathbf{\Omega}_{\mathbf{y}})) - \mathbf{y}, d_2 = \hat{f}_\theta(\mathbf{y}) - \mathbf{y}, cond = (h(f_\theta(\mathbf{\Omega}_{\mathbf{y}})) - \mathbf{y})^T (\hat{f}_\theta(\mathbf{y}) - \mathbf{y})$. Here, we divide the objective function $\mathcal{T}(\mathbf{y})$ into two cases:
i) If $cond \geqslant 0$, we have

$$\mathcal{T}(\mathbf{y}) = \|h(f_\theta(\mathbf{\Omega}_{\mathbf{y}})) + \lambda \hat{f}_\theta(\mathbf{y}) - (\lambda + 1)\mathbf{y}\|_2^2; \tag{6}$$

ii) Similarly, when $cond < 0$, it holds that

$$\mathcal{T}(\mathbf{y}) = \|\hat{f}_\theta(\mathbf{y}) - h(f_\theta(\mathbf{\Omega}_{\mathbf{y}})) + (\lambda - 1)(\hat{f}_\theta(\mathbf{y}) - \mathbf{y})\|_2^2. \tag{7}$$

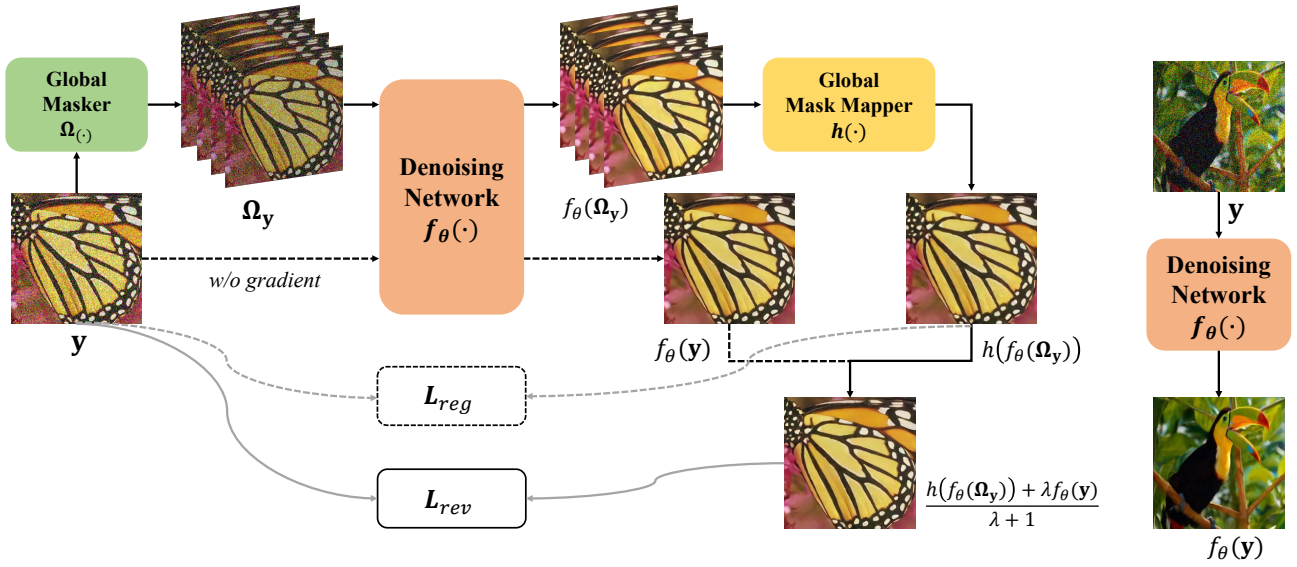
Note that when $cond < 0$, the learning objective of $\hat{f}_\theta(\mathbf{y})$ becomes $h(f_\theta(\mathbf{\Omega}_{\mathbf{y}}))$, instead of $\mathbf{y}$ as Eq. (6). Considering the optimal denoiser f_θ^* that is trained to convergence, then $\hat{f}_\theta^*(\mathbf{y})$ should be approximate to $h(f_\theta^*(\mathbf{\Omega}_{\mathbf{y}}))$ in Eq. (7). However, the masking operation $\mathbf{\Omega}_{(\cdot)}$ causes information loss, resulting in low denoising upper bound for $h(f_\theta(\mathbf{\Omega}_{\mathbf{y}}))$. Besides, $\hat{f}_\theta(\mathbf{y}) - h(f_\theta(\mathbf{\Omega}_{\mathbf{y}}))$ in Eq. (7) indicates that $\hat{f}_\theta(\mathbf{y})$ should approximate $h(f_\theta(\mathbf{\Omega}_{\mathbf{y}}))$, which in turn suppresses the performance of $\hat{f}_\theta(\mathbf{y})$. Combining the above analysis, the final re-visible loss can be formulated as:

$$\arg\min_{\theta} \mathbb{E}_{\mathbf{y}} \|h(f_\theta(\mathbf{\Omega}_{\mathbf{y}})) + \lambda \hat{f}_\theta(\mathbf{y}) - (\lambda + 1)\mathbf{y}\|_2^2. \tag{8}$$

When the denoiser converges to f_θ^*, the following holds with the optimal solution $\tilde{\mathbf{x}}$ of Eq. (8):

$$\tilde{\mathbf{x}} = \frac{h(f_\theta^*(\mathbf{\Omega}_{\mathbf{y}})) + \lambda \hat{f}_\theta^*(\mathbf{y})}{\lambda + 1}. \tag{9}$$

We assume that $h(f_\theta^*(\mathbf{\Omega}_{\mathbf{y}})) = \mathbf{x} + \varepsilon_1$ and $\hat{f}_\theta^*(\mathbf{y}) = \mathbf{x} + \varepsilon_2$. Empirically, $\|\varepsilon_1\|_1 > \|\varepsilon_2\|_1$. With Eqs. (4) and (8), the upper and lower bounds of $\tilde{\mathbf{x}}$ are respectively $\hat{f}_\theta^*(\mathbf{y})$ and $h(f_\theta^*(\mathbf{\Omega}_{\mathbf{y}}))$. Namely, $h(f_\theta^*(\mathbf{\Omega}_{\mathbf{y}})) \leq \tilde{\mathbf{x}} \leq \hat{f}_\theta^*(\mathbf{y})$. Therefore, we can consider the generation of denoised images $\tilde{\mathbf{x}}$ only from noisy original images $\mathbf{y}$ during inference. Given $\mathbf{y}$, it holds that $\lim_{\lambda \to +\infty} \tilde{\mathbf{x}} = \hat{f}_\theta^*(\mathbf{y})$.



(a) Training Phase (b) Inference Phase

Figure 1. Overview of our proposed Blind2Unblind framework. (a) Overall training process. The global masker $\mathbf{\Omega}_{(\cdot)}$ creates a masked volume by adding blind spots to a noisy image $\mathbf{y}$. Then, the global-aware mask mapper samples the denoised volume to obtain $h(f_\theta(\mathbf{\Omega}_\mathbf{y}))$. Meanwhile, the denoiser $f_\theta(\cdot)$ takes $\mathbf{y}$ as input and produces the denoised result $f_\theta(\mathbf{y})$. The re-visible loss realizes the transition from the blind to visible with the invisible term $h(f_\theta(\mathbf{\Omega}_\mathbf{y}))$ as a medium. Moreover, the regular term is used to stabilize the training phase. (b) Inference using the trained denoising model. The denoising network generates denoised images directly from noisy images $\mathbf{y}$ without additional operation.

3.4. Stabilize Transition Procedure

As described in Section 3.3, the blind part $|h(f_\theta(\mathbf{\Omega}_\mathbf{y})) - \mathbf{y}|$ is used as a transition to optimize the re-visible one $\lambda \cdot |\hat{f}_\theta(\mathbf{y}) - \mathbf{y}|$. Therefore, the cumulative error of the blind part can also affect the non-blind part. However, the pure re-visible loss lacks the separate constraint just for the transition term, which aggravates the instability during the training phase. We consider adding an extra constraint to correct $f_\theta(\mathbf{\Omega}_\mathbf{y})$, forcing the blind part to be zero. Thus, based on Eq. (8), we have the following constrained optimization problem:

$$\begin{aligned} &\min_\theta \mathbb{E}_\mathbf{y} \|h(f_\theta(\mathbf{\Omega}_\mathbf{y})) + \lambda \hat{f}_\theta(\mathbf{y}) - (\lambda+1)\mathbf{y}\|_2^2, \\ &\text{s.t.} \quad \|h(f_\theta(\mathbf{\Omega}_\mathbf{y})) - \mathbf{y}\|_2^2 = 0. \end{aligned} \tag{10}$$

We further reformulate it as the following optimization problem with a regularization term:

$$\begin{aligned} &\min_\theta \mathbb{E}_\mathbf{y} \|h(f_\theta(\mathbf{\Omega}_\mathbf{y})) + \lambda \hat{f}_\theta(\mathbf{y}) - (\lambda+1)\mathbf{y}\|_2^2 \\ &+ \eta \cdot \|h(f_\theta(\mathbf{\Omega}_\mathbf{y})) - \mathbf{y}\|_2^2 = 0. \end{aligned} \tag{11}$$

4. Main Method

Based on the theoretical analysis in Section 3, we propose Blind2Unblind, a novel deep self-supervised denoising framework learning from a single observation of noisy images. The framework consists of two main components: global-aware mask mapper and re-visible loss. The mask mapper generates masked volumes of noisy images as $\mathbf{\Omega}_\mathbf{y}$ and then samples denoised pixels in which there are blind spots to form the final denoised results. Moreover, we introduce a regularized re-visible loss to convert the blind spots into visible ones and overcome the challenge of having less information based on blind spots. An overview of our proposed Blind2Unblind framework can be found in Figure 1.

4.1. Global-Aware Mask Mapper

Manual masking methods [4, 18] hide part of pixels for training, and objective function focuses only on masked regions, leading to a decrease in accuracy and slow convergence. The mask mapper performs global denoising on blind spots. This mechanism constrains all pixels, promotes information exchange across whole masked regions, and increases noise reduction accuracy. In addition, our mask mapper also speeds up the training of manual masking schemes. The workflow of masked volumes generation and global mapping using mask mapper is shown in Figure 2. Given a noisy image $\mathbf{y}$ with width W and height H, details of the global-aware mask mapper $h\left(f_\theta\left(\mathbf{\Omega}_\mathbf{y}\right)\right)$ are described as follows:

1. First, the noisy image $\mathbf{y}$ is divided into $\lfloor W/s \rfloor \times \lfloor H/s \rfloor$ cells, each of size $s \times s$. For simplicity, we set $s = 2$.
2. The pixels in the i-th row and j-th column of each cell are masked and filled with black for illustration. Thus, the masked image $\mathbf{\Omega}_\mathbf{y}^{ij}$ consists of $\lfloor W/s \rfloor \times \lfloor H/s \rfloor$

blind spots, where $i, j \in \{0, ..., s-1\}$. These mask images are further stacked to form a masked volume $\mathbf{\Omega_y}$.

3. For positions on the denoised volumes $f_\theta(\mathbf{\Omega_y})$ corresponding to blind spots, the mask mapper $h(\cdot)$ samples them and assigns them to the same layer. In this way, a globally denoised image $h\left(f_\theta\left(\mathbf{\Omega_y}\right)\right)$ is obtained.

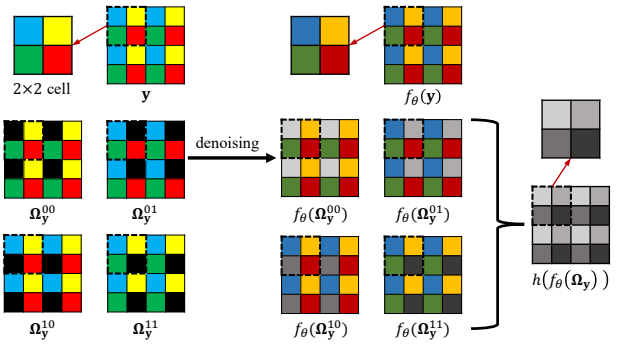


Figure 2. Details of a global-aware mask mapper. Taking a 2×2 cell $\mathbf{y}$ as an example, the global masker $\mathbf{\Omega}_{(\cdot)}$ hides four spots in $\mathbf{y}$ to create a global masked volumes $\mathbf{\Omega_y}$ consisting of four masked cells $\mathbf{\Omega}_\mathbf{y}^{ij}, i, j \in \{0, 1\}$. The mask mapper $h(\cdot)$ samples the denoised volumes $f_\theta(\mathbf{\Omega_y})$ in which there are blind spots. The final denoised cell $h(f_\theta(\mathbf{\Omega_y}))$ is formed by the mapper based on the sampled locations and pixel values.

4.2. Regularized Re-Visible Loss

Here we present how to use the Blind2Unblind framework for self-supervised denoising. Since blind spot denoising plays a mediating role, the training process should follow the transition from blind to non-blind. However, the pure re-visible loss uses only a single variable that can be backpropagated to optimize both the blind term and the visible term, resulting in unstable training. Therefore, a regular term is introduced to constrain the blind term and stabilize the training procedure. The re-visible loss with regularization is as follows:

$$\begin{aligned}\mathcal{L} &= \mathcal{L}_{rev} + \eta \cdot \mathcal{L}_{reg} \\ &= \|h(f_\theta(\mathbf{\Omega_y})) + \lambda \hat{f}_\theta(\mathbf{y}) - (\lambda+1)\mathbf{y}\|_2^2 \\ &+ \eta \cdot \|h(f_\theta(\mathbf{\Omega_y})) - \mathbf{y}\|_2^2.\end{aligned} \tag{12}$$

where $f_\theta(\cdot)$ denotes an arbitrary denoising network with different structures, $h(\cdot)$ is a mask mapper capable of global modeling, η is a fixed hyper-parameter that determines the initial contribution of the blind term and the training stability, and λ is a variable hyper-parameter that controls the intensity of visible parts when converting from blind to unblind. The initial and final values of λ are λ_s and λ_f, respectively. To prevent identity mapping, we disable the gradient updating of $\hat{f}_\theta(\mathbf{y})$ during training.

5. Experimental Results

5.1. Implementation Details

Training Details. We use the same modified U-Net [29] architecture as [15,20]. The batch size is 4. We use Adam [17] as our optimizer, with a weight decay of $1e^{-8}$ to avoid overfitting. The initial learning rate is 0.0003 for synthetic denoising in sRGB space and 0.0001 for real-world denoising, including raw-RGB space and fluorescence microscopy (FM). The learning rate decreases by half every 20 epochs, for 100 training epochs. As for the hyper-parameters included in the re-visible loss, we set $\eta = 1, \lambda_s = 2$ and $\lambda_f = 20$ empirically. We also randomly crop 128×128 patches for training. In practice, masked pixels are a weighted average of pixels in a 3×3 neighborhood, as advocated in [18]. All models are trained on a server using Python 3.8.5, Pytorch 1.7.1 [27], and Nvidia Tesla V100 GPUs.

Datasets for Synthetic Denoising. Following the setting in [15, 20], we select 44,328 images with sizes between 256×256 and 512×512 pixels from the ILSVRC2012 [10] validation set as the training set. To obtain reliable average PSNRs, we also repeat the test sets Kodak [11], BSD300 [24] and Set14 [37] by 10, 3 and 20 times, respectively. Thus, all methods are evaluated with 240, 300, and 280 individual synthetic noise images. Specifically, we consider four types of noise in sRGB space: (1) Gaussian noise with $\sigma = 25$, (2) Gaussian noise with $\sigma \in [5, 50]$, (3) Poisson noise with $\lambda = 30$, (4) Poisson noise with $\lambda \in [5, 50]$. Here, values of σ are given in 8-bit units. For grayscale image denoising, we use BSD400 [38] for training and three widely used datasets for testing, including Set12, BSD68 [30] and Urban100 [14].

Datasets for Real-World Denoising. Following the setting in [15], we take the SIDD [2] dataset collected by five smartphone cameras in 10 scenes under different lighting conditions for real-world denoising in raw-RGB space. We use only raw-RGB images in SIDD Medium Dataset for training and use SIDD Validation and Benchmark Datasets for validation and testing. As for real-world grayscale denoising on FM, we consider FMDD [40] dataset, which contains 12 datasets of images captured using either a confocal, two-photon, or widefield microscope. Each dataset contains 20 views with 50 noisy images per view. We select three datasets (Confocal Fish, Confocal Mice and Two-Photon Mice) and train with each dataset, using the 19th view for testing and the rest for training.

Details of Experiments. For fair comparison, we follow the experimental settings of NBR2NBR [15]. For the baseline, we consider two supervised denoising methods (N2C [29] and N2N [21]). We also compare the proposed Blind2Unblind with a traditional approach (BM3D [9]) and seven self-supervised denoising algorithms (Self2Self [28], Noise2Void (N2V) [18], Laine19 [20], Noisier2Noise [25],

Figure 3. Visual comparison of denoising sRGB images in the setting of $\sigma = 25$.

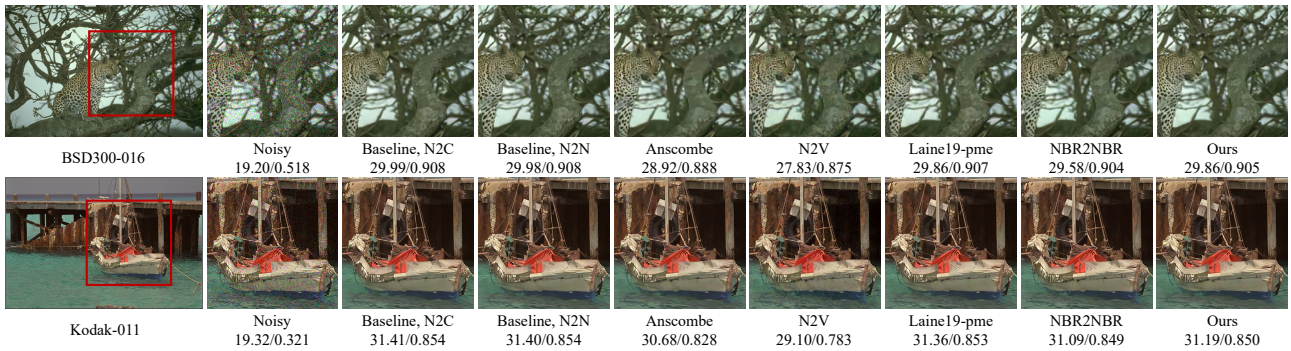


Figure 4. Visual comparison of denoising sRGB images in the setting of $\lambda = 30$.

DBSN [33], R2R [26] and NBR2NBR [15]).

In synthetic denoising, we use pre-trained models provided by [20] for N2C, N2N, and Laine19, retaining the same network architecture as [15, 20]. In addition, we use CBM3D [8], a multi-channel version of BM3D, to perform Gaussian denoising combined with the parameter σ estimated by [7]. For Poisson noise, Anscombe [23] is first run to convert Poisson noise into Gaussian distribution, and then BM3D is used for denoising. For Self2Self, N2V, Noisier2Noise, DBSN, R2R and NBR2NBR, we use the official implementation.

For real-world denoising in raw-RGB space, all methods use their official implementations and have been retrained on the SIDD Medium Dataset. Note that we split the single-channel raw image into four sub-images according to the Bayer pattern. For BM3D, we denoise the four sub-images individually and then integrate denoised sub-images into a whole image. For deep learning methods, we stack four sub-images to form a four-channel image for denoising and then integrate the denoised image into raw-RGB space. For real-world denoising on FM, we retrain all compared methods using official implementation.

5.2. Results for Synthetic Denoising

The quantitative comparison results of synthetic denoising for Gaussian and Poisson noise can be seen in Table 1. For Gaussian noise, whether the noise level is fixed or variable, our approach significantly outperforms the traditional denoising method BM3D and five self-supervised denoising methods, including Self2Self, N2V, Laine19-mu, DBSN, and NBR2NBR. R2R also performs worse than our method under variable Gaussian noise, despite a known noise prior. For fixed Poisson noise, our approach shows comparable performance to Laine19-pme, which is based on explicit noise modeling. However, explicit noise modeling means strong prior, leading to poor performance on real data. The following experiments on real-world datasets also illustrate this problem. When the Poisson noise level is variable, our method outperforms almost all methods to be compared, including Laine19-pme. Moreover, our approach outperforms the supervision baseline (N2C) with a gain of 0.13 dB and 0.08 dB on BSD300 and SET14, respectively. Namely, our method shows more superior performance for removing Poisson noise. In addition, whether the noise is fixed or variable, our method significantly outperforms the SOTA method NBR2NBR, with a maximum gain of 0.38 dB. Figures 3 and 4 illustrate the sRGB denoising results in the setting of $\sigma = 25$ and $\lambda = 30$, respectively. Compared with the recent best NBR2NBR, our method recovers more texture details. Table 2 conducts additional experiments on grayscale images. As the noise increases, our method gradually outperforms R2R, because re-corruption loses more information, offsetting the noise prior advantage. Moreover, our method infers directly, while R2R repeats 50 times.

5.3. Results for raw-RGB Denoising

In raw-RGB space, Table 3 shows the quality scores for quantitative comparisons on SIDD Benchmark and SIDD

Noise Type	Method	KODAK	BSD300	SET14
Gaussian $\sigma = 25$	Baseline, N2C [29]	32.43/0.884	31.05/0.879	31.40/0.869
	Baseline, N2N [21]	32.41/0.884	31.04/0.878	31.37/0.868
	CBM3D [8]	31.87/0.868	30.48/0.861	30.88/0.854
	Self2Self [28]	31.28/0.864	29.86/0.849	30.08/0.839
	N2V [18]	30.32/0.821	29.34/0.824	28.84/0.802
	Laine19-mu [20]	30.62/0.840	28.62/0.803	29.93/0.830
	Laine19-pme [20]	**32.40/0.883**	**30.99/0.877**	**31.36/0.866**
	Noisier2Noise [25]	30.70/0.845	29.32/0.833	29.64/0.832
	DBSN [33]	31.64/0.856	29.80/0.839	30.63/0.846
	R2R [26]	32.25/**0.880**	**30.91/0.872**	**31.32/0.865**
	NBR2NBR [15]	32.08/0.879	30.79/0.873	31.09/0.864
	Ours	32.27/0.880	30.87/0.872	31.27/0.864
Gaussian $\sigma \in [5, 50]$	Baseline, N2C [29]	32.51/0.875	31.07/0.866	31.41/0.863
	Baseline, N2N [21]	32.50/0.875	31.07/0.866	31.39/0.863
	CBM3D [8]	32.02/0.860	30.56/0.847	30.94/0.849
	Self2Self [28]	31.37/0.860	29.87/0.841	29.97/0.849
	N2V [18]	30.44/0.806	29.31/0.801	29.01/0.792
	Laine19-mu [20]	30.52/0.833	28.43/0.794	29.71/0.822
	Laine19-pme [20]	**32.40**/0.870	**30.95/0.861**	**31.21**/0.855
	DBSN [33]	30.38/0.826	28.34/0.788	29.49/0.814
	R2R [26]	31.50/0.850	30.56/0.855	30.84/0.850
	NBR2NBR [15]	32.10/0.870	30.73/**0.861**	31.05/**0.858**
	Ours	32.34/**0.872**	30.86/**0.861**	31.14/0.857
Poisson $\lambda = 30$	Baseline, N2C [29]	31.78/0.876	30.36/0.868	30.57/0.858
	Baseline, N2N [21]	31.77/0.876	30.35/0.868	30.56/0.857
	Anscombe [23]	30.53/0.856	29.18/0.842	29.44/0.837
	Self2Self [28]	30.31/0.857	28.93/0.840	28.84/0.839
	N2V [18]	28.90/0.788	28.46/0.798	27.73/0.774
	Laine19-mu [20]	30.19/0.833	28.25/0.794	29.35/0.820
	Laine19-pme [20]	**31.67/0.874**	**30.25/0.866**	**30.47/0.855**
	DBSN [33]	30.07/0.827	28.19/0.790	29.16/0.814
	R2R [26]	30.50/0.801	29.47/0.811	29.53/0.801
	NBR2NBR [15]	31.44/0.870	30.10/0.863	30.29/0.853
	Ours	31.64/0.871	**30.25**/0.862	30.46/0.852
Poisson $\lambda \in [5, 50]$	Baseline, N2C [29]	31.19/0.861	29.79/0.848	30.02/0.842
	Baseline, N2N [21]	31.18/0.861	29.78/0.848	30.02/0.842
	Anscombe [23]	29.40/0.836	28.22/0.815	28.51/0.817
	Self2Self [28]	29.06/0.834	28.15/0.817	28.83/0.841
	N2V [18]	28.78/0.758	27.92/0.766	27.43/0.745
	Laine19-mu [20]	29.76/0.820	27.89/0.778	28.94/0.808
	Laine19-pme [20]	30.88/0.850	29.57/0.841	28.65/0.785
	DBSN [33]	29.60/0.811	27.81/0.771	28.72/0.800
	R2R [26]	29.14/0.732	28.68/0.771	28.77/0.765
	NBR2NBR [15]	30.86/0.855	29.54/0.843	29.79/0.838
	Ours	**31.07/0.857**	**29.92/0.852**	**30.10/0.844**

Table 1. Quantitative denoising results on synthetic datasets in sRGB space. The highest PSNR(dB)/SSIM among unsupervised denoising methods is highlighted in **bold**, while the second is underlined.

Noise Type	Method	Network	BSD68	Set12	Urban100
Gaussian $\sigma = 15$	N2C [29]	U-Net [29]	31.58/0.889	32.60/0.899	31.95/0.910
	R2R [26]	U-Net [29]	**31.54/0.885**	**32.54/0.897**	**31.89/0.915**
	Ours	U-Net [29]	31.44/0.884	32.46/**0.897**	31.79/0.912
Gaussian $\sigma = 25$	N2C [29]	U-Net [29]	29.02/0.822	30.07/0.852	29.01/0.861
	R2R [26]	U-Net [29]	**28.99**/0.818	30.06/0.851	**29.17/0.865**
	Ours	U-Net [29]	**28.99/0.820**	**30.09/0.854**	29.05/0.864
Gaussian $\sigma = 50$	N2C [29]	U-Net [29]	26.08/0.715	26.88/0.777	25.35/0.766
	R2R [26]	U-Net [29]	26.02/0.705	26.86/0.771	25.48/0.768
	Ours	U-Net [29]	**26.09/0.715**	**26.91/0.776**	**25.57/0.868**

Table 2. Grayscale image denoising results.

Validation. Note that the online website [1] evaluates the quality scores for the SIDD Benchmark. In general, the proposed method outperforms the state-of-the-art (NBR2NBR) by 0.32 dB and 0.20 dB for the benchmark and validation. Because the transition from blind to non-blind preserves all information contained within noisy images and mitigates the effects of over smoothing among neighboring pixels.

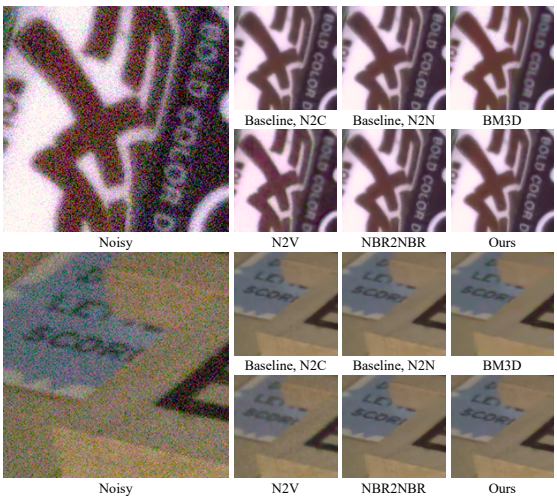


Figure 5. Visual comparison of denoising raw-RGB images in the challenging SIDD benchmark. The conversion from raw-RGB to sRGB is performed using the official ISP tool.

Methods	Network	SIDD Benchmark	SIDD Validation
Baseline, N2C [29]	U-Net [29]	50.60/0.991	51.19/0.991
Baseline, N2N [21]	U-Net [29]	50.62/0.991	51.21/0.991
BM3D [9]	-	48.60/0.986	48.92/0.986
N2V [18]	U-Net [29]	48.01/0.983	48.55/0.984
Laine19-mu (Gaussian) [20]	U-Net [29]	49.82/0.989	50.44/0.990
Laine19-pme (Gaussian) [20]	U-Net [29]	42.17/0.935	42.87/0.939
Laine19-mu (Poisson) [20]	U-Net [29]	50.28/0.989	50.89/0.990
Laine19-pme (Poisson) [20]	U-Net [29]	48.46/0.984	48.98/0.985
DBSN [33]	DBSN [33]	49.56/0.987	50.13/0.988
R2R [26]	U-Net [29]	46.70/0.978	47.20/0.980
NBR2NBR [15]	U-Net [29]	50.47/0.990	51.06/0.991
Ours	U-Net [29]	**50.79/0.991**	**51.36/0.992**

Table 3. Quantitative denoising results on SIDD benchmark and validation datasets in raw-RGB space.

Moreover, our method performs far better than R2R in the real world. Two primary reasons are lossy re-corruption and inaccurate noise priors in R2R. Interestingly, our approach outperforms the baseline (N2C) by 0.19 dB and 0.17 dB in the benchmark and validation, indicating improved generalization. The raw-RGB denoising performance in the real world demonstrates that the proposed approach is competitive in the presence of complex noise patterns. Figure 5 demonstrates our approach's superiority over alternative methods. Our method recovers more texture details and has a higher degree of continuity, whereas NBR2NBR exhibits obvious sawtooth diffusion defects due to the training scheme of approximating neighboring pixels.

5.4. Results for FM Denoising.

Table 4 shows the quantitative comparison of real-world fluorescence microscopy datasets. Our approach shows competitive performance against other methods. In particular, our method greatly outperforms self-supervised methods and slightly outperforms supervised methods (N2C and N2N) on both Confocal Mice and Two-Photon Mice. The results further confirm the superior performance of

our method on complex real-world noise patterns. Figure 6 shows a visual comparison of denoising FM images.

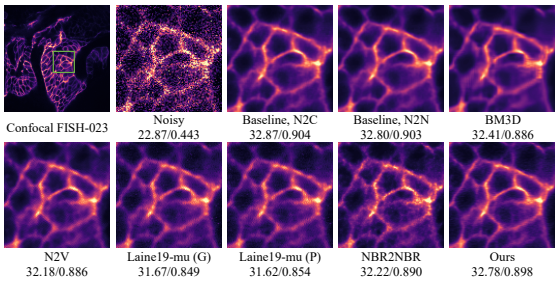


Figure 6. Visual comparison of denoising FM images.

Methods	Network	Confocal Fish	Confocal Mice	Two-Photon Mice
Baseline, N2C [29]	U-Net [29]	32.79/0.905	38.40/0.966	34.02/0.925
Baseline, N2N [21]	U-Net [29]	32.75/0.903	38.37/0.965	33.80/0.923
BM3D [9]	-	32.16/0.886	37.93/0.963	33.83/**0.924**
N2V [18]	U-Net [29]	32.08/0.886	37.49/0.960	33.38/0.916
Laine19-mu (G) [20]	U-Net [29]	31.62/0.849	37.54/0.959	32.91/0.903
Laine19-pme (G) [20]	U-Net [29]	23.30/0.527	31.64/0.881	25.87/0.418
Laine19-mu (P) [20]	U-Net [29]	31.59/0.854	37.30/0.956	33.09/0.907
Laine19-pme (P) [20]	U-Net [29]	25.16/0.597	37.82/0.959	31.80/0.820
NBR2NBR [15]	U-Net [29]	32.11/0.890	37.07/0.960	33.40/0.921
Ours	U-Net [29]	**32.74/0.897**	**38.44/0.964**	**34.03**/0.916

Table 4. Quantitative denoising results on Confocal Fish, Confocal Mice and Two-Photon Mice. G is for Gaussian and P is for Poisson.

5.5. Ablation Study

This section conducts ablation studies on the loss function, mask strategy, visible term, and regular term. Note that PSNR(dB)/SSIM is evaluated on Kodak.

Ablation study on Loss Function. Table 5 performs ablation experiments under two loss conditions. Since $\hat{f}_\theta(\mathbf{y})$ in Eq. (7) is suppressed by the mask term $h(f_\theta(\mathbf{\Omega}_\mathbf{y}))$, the performance of $\mathcal{L}_B$ is much lower than that of $\mathcal{L}_A$.

Ablation study on Mask Mapper. To evaluate our global-aware mask mapper (GM), we compare it with the random mask (RM) as advocated in Noise2Void [18]. Details are provided in the supplementary materials. Table 6 lists the performance of different mask strategies in self-supervised denoising. We can see that our GM considerably outperforms RM for all four noise patterns. Moreover, GM+V performs much better combined with re-visible loss, while RM+V is severely suppressed and cannot converge.

Ablation study on Visible Term. The hyper-parameter λ_f is positively correlated with the visibility of the non-blind term. Table 7 shows that the weight of the visible item is not the larger, the better: 1) When $\lambda_f = 20$, our approach achieves or approaches the best performance on four noise patterns. 2) When λ_f varies from 20 to 40, the performance on Poisson noise decreases. In contrast, the performance on Gaussian noise remains fixed. 3) When $\lambda_f > 40$, the performance is almost unchanged. In this paper, we set $\lambda_f = 20$.

Loss Type	$\sigma = 25$	$\sigma \in [5, 50]$	$\lambda = 30$	$\lambda \in [5, 50]$
$\mathcal{L}_A$	**32.27/0.880**	**32.34/0.872**	**31.64/0.871**	**31.07/0.857**
$\mathcal{L}_B$	31.37/0.873	30.85/0.851	30.95/0.863	30.01/0.840

Table 5. Ablation study on the loss function. $\mathcal{L}_A$ and $\mathcal{L}_B$ denote Eq. (6) and Eq. (7).

Noise Type	RM	GM	RM+V	GM+V
Gaussian $\sigma = 25$	30.32/0.831	30.56/0.839	- / -	**32.27/0.880**
Gaussian $\sigma \in [5, 50]$	30.23/0.822	30.46/0.831	- / -	**32.34/0.872**
Poisson $\lambda = 30$	29.89/0.821	30.11/0.830	- / -	**31.64/0.871**
Poisson $\lambda \in [5, 50]$	29.26/0.796	29.67/0.817	- / -	**31.07/0.857**

Table 6. Ablation study on mask mappers. RM, GM and V denote random mask mapper, global-aware mask mapper, and re-visible loss. - / - means cannot converge.

Noise Type	$\lambda_f = 2$	$\lambda_f = 20$	$\lambda_f = 40$	$\lambda_f = 100$
Gaussian $\sigma = 25$	32.11/0.879	**32.27/0.880**	**32.27**/0.879	32.26/0.879
Gaussian $\sigma \in [5, 50]$	32.00/0.871	32.34/0.872	32.34/0.872	**32.35/0.873**
Poisson $\lambda = 30$	31.47/0.869	**31.64/0.871**	31.52/0.869	31.51/0.869
Poisson $\lambda \in [5, 50]$	30.94/0.854	**31.07/0.857**	31.02/0.855	31.01/0.854

Table 7. Ablation study on visible term. Note that $\eta = 1, \lambda_s = 2$.

Noise Type	$\eta = 0$	$\eta = 1$	$\eta = 2$	$\eta = 3$
Gaussian $\sigma = 25$	32.19/0.877	32.27/0.880	**32.32/0.881**	32.21/0.878
Gaussian $\sigma \in [5, 50]$	32.21/0.870	**32.34/0.872**	32.31/**0.872**	32.23/0.871
Poisson $\lambda = 30$	31.54/0.869	**31.64/0.871**	31.54/0.869	31.52/0.869
Poisson $\lambda \in [5, 50]$	31.03/0.856	**31.07/0.857**	31.06/**0.857**	31.04/**0.857**

Table 8. Ablation study on regular term. Here, $\lambda_s = 2, \lambda_f = 20$.

Ablation study on Regularization Term. The hyper-parameter η controls the stability of re-visible loss. Table 8 shows that the denoising performance first increases and then decreases on four noise patterns. In particular, when $\eta = 1$, the performance is at or closed to the best, which proves that the regular term strengthens the training stability and improves the performance. In this paper, we set $\eta = 1$.

6. Conclusion

We propose Blind2Unblind, a novel self-supervised denoising framework, which achieves lossless denoising through the transition from blind to non-blind and enables blindspot schemes to use complete noisy images without information loss. The global-aware mask mapper samples the denoised volume at blind spots, and then re-visible loss realizes non-blind denoising under visible blind spots. Extensive experiments have shown the superiority of our approach against compared methods, especially for complex noise patterns.

Acknowledgments

This work is jointly supported by National Natural Science Foundation of China (32171461), the Strategic Priority Research Program of Chinese Academy of Science (XDA16021104, XDB32030208), Program of Beijing Municipal Science & Technology Commission (Z201100008420004), and Science and Technology Innovation 2030 Major Projects of China (2021ZD0204503).

References

[1] Abdelrahman Abdelhamed. Website. https://www.eecs.yorku.ca/~kamel/sidd/benchmark.php. Accessed: 2018-11-01. 7

[2] Abdelrahman Abdelhamed, Stephen Lin, and Michael S Brown. A high-quality denoising dataset for smartphone cameras. In *CVPR*, pages 1692–1700, 2018. 5

[3] Saeed Anwar and Nick Barnes. Real image denoising with feature attention. In *ICCV*, pages 3155–3164, 2019. 1, 2

[4] Joshua Batson and Loic Royer. Noise2self: Blind denoising by self-supervision. In *International Conference on Machine Learning*, pages 524–533, 2019. 1, 2, 3, 4

[5] Antoni Buades, Bartomeu Coll, and J-M Morel. A non-local algorithm for image denoising. In *CVPR*, volume 2, pages 60–65, 2005. 1, 2

[6] Meng Chang, Qi Li, Huajun Feng, and Zhihai Xu. Spatial-adaptive network for single image denoising. In *ECCV*, pages 171–187, 2020. 1, 2

[7] Guangyong Chen, Fengyuan Zhu, and Pheng Ann Heng. An efficient statistical method for image noise level estimation. In *ICCV*, pages 477–485, 2015. 6

[8] Kostadin Dabov, Alessandro Foi, Vladimir Katkovnik, and Karen Egiazarian. Color image denoising via sparse 3d collaborative filtering with grouping constraint in luminance-chrominance space. In *ICIP*, volume 1, pages I–313, 2007. 1, 2, 6, 7

[9] Kostadin Dabov, Alessandro Foi, Vladimir Katkovnik, and Karen Egiazarian. Image denoising by sparse 3-d transform-domain collaborative filtering. *TIP*, 16(8):2080–2095, 2007. 1, 2, 5, 7, 8

[10] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *CVPR*, pages 248–255, 2009. 5

[11] Rich Franzen. Kodak lossless true color image suite. *source: http://r0k. us/graphics/kodak*, 4(2), 1999. 5

[12] Shuhang Gu, Lei Zhang, Wangmeng Zuo, and Xiangchu Feng. Weighted nuclear norm minimization with application to image denoising. In *CVPR*, pages 2862–2869, 2014. 1, 2

[13] Shi Guo, Zifei Yan, Kai Zhang, Wangmeng Zuo, and Lei Zhang. Toward convolutional blind denoising of real photographs. In *CVPR*, pages 1712–1722, 2019. 1, 2

[14] Jia-Bin Huang, Abhishek Singh, and Narendra Ahuja. Single image super-resolution from transformed self-exemplars. In *CVPR*, pages 5197–5206, 2015. 5

[15] Tao Huang, Songjiang Li, Xu Jia, Huchuan Lu, and Jianzhuang Liu. Neighbor2neighbor: Self-supervised denoising from single noisy images. In *CVPR*, pages 14781–14790, 2021. 1, 2, 5, 6, 7, 8

[16] Yongqiang Huang, Zexin Lu, Zhimin Shao, Maosong Ran, Jiliu Zhou, Leyuan Fang, and Yi Zhang. Simultaneous denoising and super-resolution of optical coherence tomography images based on generative adversarial network. *Optics express*, 27(9):12289–12307, 2019. 1

[17] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In *ICLR*, 2015. 5

[18] Alexander Krull, Tim-Oliver Buchholz, and Florian Jug. Noise2void-learning denoising from single noisy images. In *CVPR*, pages 2129–2137, 2019. 1, 2, 3, 4, 5, 7, 8

[19] Alexander Krull, Tomáš Vičar, Mangal Prakash, Manan Lalit, and Florian Jug. Probabilistic noise2void: Unsupervised content-aware denoising. *Frontiers in Computer Science*, 2:5, 2020. 1, 2

[20] Samuli Laine, Tero Karras, Jaakko Lehtinen, and Timo Aila. High-quality self-supervised deep image denoising. *NIPS*, 32:6970–6980, 2019. 1, 2, 5, 6, 7, 8

[21] Jaakko Lehtinen, Jacob Munkberg, Jon Hasselgren, Samuli Laine, Tero Karras, Miika Aittala, and Timo Aila. Noise2noise: Learning image restoration without clean data. In *ICML*, 2018. 1, 2, 5, 7, 8

[22] Ding Liu, Bihan Wen, Jianbo Jiao, Xianming Liu, Zhangyang Wang, and Thomas S Huang. Connecting image denoising and high-level vision tasks via deep learning. *TIP*, 29:3695–3706, 2020. 1

[23] Markku Makitalo and Alessandro Foi. Optimal inversion of the anscombe transformation in low-count poisson image denoising. *TIP*, 20(1):99–109, 2010. 6, 7

[24] 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 *ICCV*, volume 2, pages 416–423, 2001. 5

[25] Nick Moran, Dan Schmidt, Yu Zhong, and Patrick Coady. Noisier2noise: Learning to denoise from unpaired noisy data. In *CVPR*, pages 12064–12072, 2020. 1, 2, 5, 7

[26] Tongyao Pang, Huan Zheng, Yuhui Quan, and Hui Ji. Recorrupted-to-recorrupted: Unsupervised deep learning for image denoising. In *CVPR*, pages 2043–2052, 2021. 1, 2, 6, 7

[27] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, et al. Pytorch: An imperative style, high-performance deep learning library. *NIPS*, 32:8026–8037, 2019. 5

[28] Yuhui Quan, Mingqin Chen, Tongyao Pang, and Hui Ji. Self2self with dropout: Learning self-supervised denoising from single image. In *CVPR*, pages 1890–1898, 2020. 2, 5, 7

[29] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241, 2015. 1, 2, 5, 7, 8

[30] Stefan Roth and Michael J Black. Fields of experts: A framework for learning image priors. In *CVPR*, pages 860–867. IEEE, 2005. 5

[31] B Shijila, Anju Jose Tom, and Sudhish N George. Simultaneous denoising and moving object detection using low rank approximation. *Future Generation Computer Systems*, 90:198–210, 2019. 1

[32] Dmitry Ulyanov, Andrea Vedaldi, and Victor Lempitsky. Deep image prior. In *CVPR*, pages 9446–9454, 2018. 1

[33] Xiaohe Wu, Ming Liu, Yue Cao, Dongwei Ren, and Wangmeng Zuo. Unpaired learning of deep image denoising. In *ECCV*, pages 352–368, 2020. 1, 2, 6, 7

[34] 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. *TIP*, 29:9316–9329, 2020. 1, 2

[35] Zongsheng Yue, Hongwei Yong, Qian Zhao, Deyu Meng, and Lei Zhang. Variational denoising network: Toward blind noise modeling and removal. *NIPS*, 32:1690–1701, 2019. 1, 2

[36] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, Ming-Hsuan Yang, and Ling Shao. Learning enriched features for real image restoration and enhancement. In *ECCV*, pages 492–511, 2020. 1, 2

[37] Roman Zeyde, Michael Elad, and Matan Protter. On single image scale-up using sparse-representations. In *International conference on curves and surfaces*, pages 711–730, 2010. 5

[38] Kai Zhang, Wangmeng Zuo, Yunjin Chen, Deyu Meng, and Lei Zhang. Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising. *TIP*, 26(7):3142–3155, 2017. 1, 2, 5

[39] Kai Zhang, Wangmeng Zuo, and Lei Zhang. Ffdnet: Toward a fast and flexible solution for cnn-based image denoising. *TIP*, 27(9):4608–4622, 2018. 1, 2

[40] Yide Zhang, Yinhao Zhu, Evan Nichols, Qingfei Wang, Siyuan Zhang, Cody Smith, and Scott Howard. A poisson-gaussian denoising dataset with real fluorescence microscopy images. In *CVPR*, 2019. 5

A. Training Framework for Blind2Unblind

The training framework for Blind2Unblind is shown in Algorithm 1.

Algorithm 1: Blind2Unblind

Input: A set of noisy images $Y = \{\mathbf{y}_i\}_{i=1}^{n}$;
Denoising network $f_\theta(\cdot)$;
Hyper-parameters η, λ;

while *not converged* **do**

Sample a noisy image $\mathbf{y}$;
Generate a global masker $\mathbf{\Omega}_{(\cdot)}$;
Derive a masked volume $\mathbf{\Omega}_\mathbf{y}$, where $\mathbf{\Omega}_\mathbf{y}$ is the network input, and $\mathbf{y}$ is the network target;
For the network input $\mathbf{\Omega}_\mathbf{y}$, derive the denoised volume $f_\theta(\mathbf{\Omega}_\mathbf{y})$;
Global mask mapper $h_{(\cdot)}$ samples the denoised volume $f_\theta(\mathbf{\Omega}_\mathbf{y})$ at blind spots, then obtain a blind denoised image $h(f_\theta(\mathbf{\Omega}_\mathbf{y}))$;
For the original noisy image $\mathbf{y}$, derive the visible denoised image $\hat{f}_\theta(\mathbf{y})$ without gradients;
Calculate re-visible loss $\mathcal{L}_{rev} = \|h(f_\theta(\mathbf{\Omega}_\mathbf{y})) + \lambda\hat{f}_\theta(\mathbf{y}) - (\lambda+1)\mathbf{y}\|_2^2$;
Calculate regularization $\mathcal{L}_{reg} = \|h(f_\theta(\mathbf{\Omega}_\mathbf{y})) - \mathbf{y}\|_2^2$;
Update network parameters θ by minimizing the regularized re-visible loss $\mathcal{L}_{rev} + \eta \cdot \mathcal{L}_{reg}$;

end

B. Details of Interpolation from Neighbors

Figure 7 shows the workflow of interpolation from neighbors. The workflow can be divided into the following three steps: 1) The mask is generated by random masking each 2×2 cells in image $\mathbf{y}$. The kernel convolves image $\mathbf{y}$ with stride 1 and padding 1 to produce $\mathbf{y}_c$. Then, $\mathbf{y}_m$ is obtained via Hadamard product $\mathbf{y}_c \circ mask$. 2) We perform Hadamard product $\mathbf{y} \circ (1 - mask)$ to gernerate $\mathbf{y}_{inv}$. 3) Sum by $\mathbf{y}_m$ and $\mathbf{y}_{inv}$, we finally obtain the masked image $\mathbf{\Omega}_\mathbf{y}$.

C. Details of Random Mask Strategy

The illustration of the random mask strategy is presented in Figure 8. The image y is divided into several blocks with 2x2 cells. A specific pixel in each cell is randomly set as a blind spot. Namely, there are four ways for random masking of 2x2 cells. After random masking, the masked image $\mathbf{\Omega}_\mathbf{y}$ is fed into the denoising network to generate the denoised image $f_\theta(\mathbf{\Omega}_\mathbf{y})$.

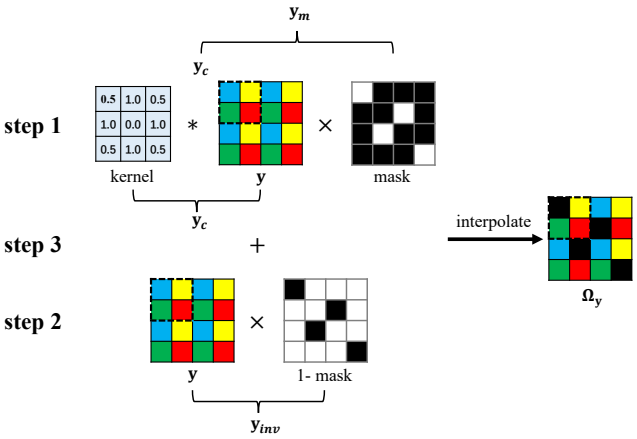


Figure 7. Details of interpolation from neighbors.

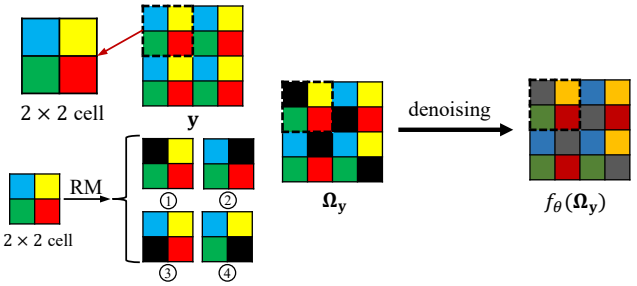


Figure 8. Details of random mask strategy.

D. More Experimental Results

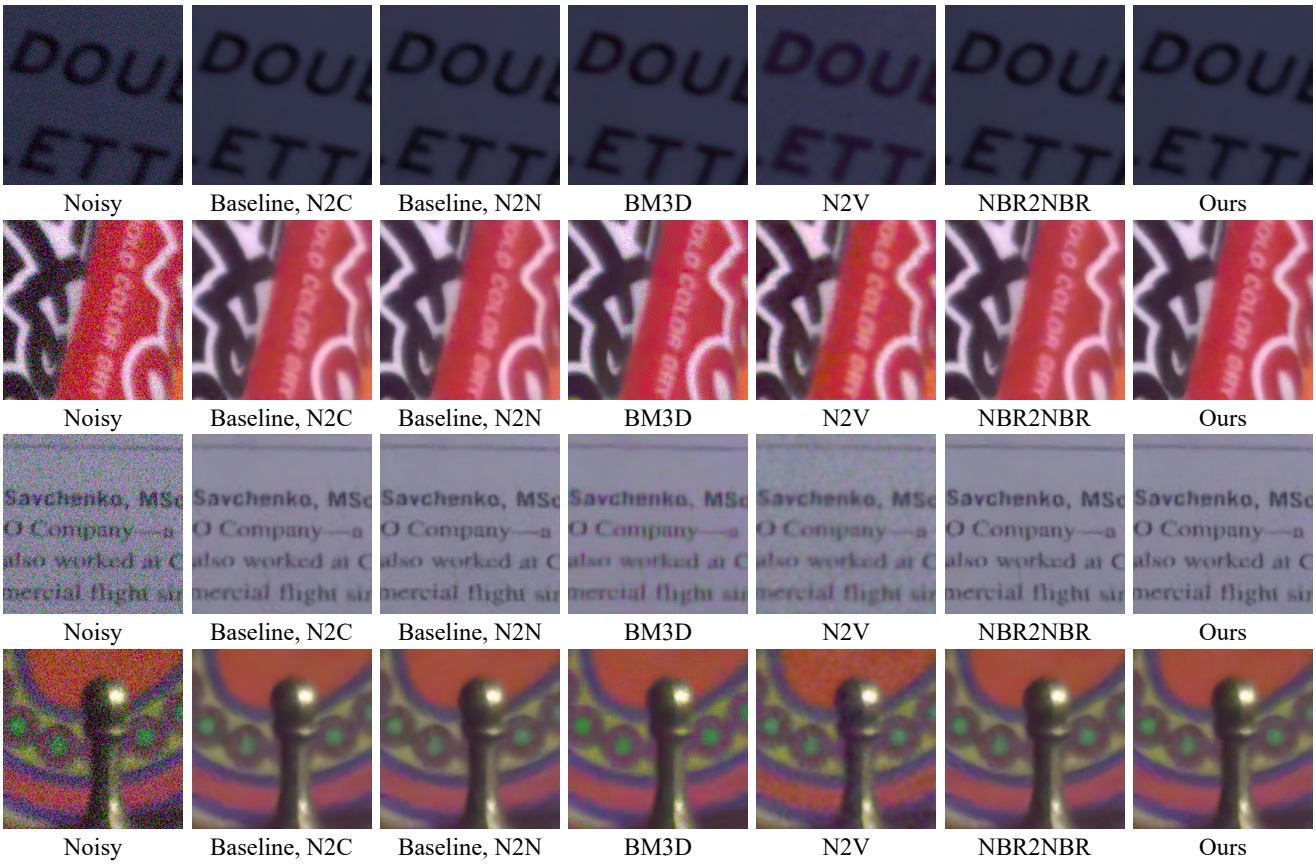
Figure 9 illustrates the steps of our proposed method while denoising sRGB images in the setting of $\sigma = 25$. Figure 10 shows the visual comparison of denoising raw-RGB images in the challenging SIDD benchmark.

Kodak-001	Noisy	Blind 26.90/0.759	Visible 29.58/0.860	Weighted 29.46/0.856

Kodak-006	Noisy	Blind 28.44/0.806	Visible 31.01/0.874	Weighted 30.92/0.871

Kodak-021	Noisy	Blind 28.68/0.839	Visible 31.42/0.891	Weighted 31.31/0.889

Figure 9. The steps of our proposed method while denoising sRGB images in the setting of $\sigma = 25$. Blind denotes $h(f_\theta(\mathbf{\Omega_y}))$, Visible denotes $\hat{f}_\theta(\mathbf{y})$, and Weighted denotes $\frac{h(f^*_\theta(\mathbf{\Omega_y}))+\lambda\hat{f}^*_\theta(\mathbf{y})}{\lambda+1}$.



Noisy	Baseline, N2C	Baseline, N2N	BM3D	N2V	NBR2NBR	Ours
Noisy	Baseline, N2C	Baseline, N2N	BM3D	N2V	NBR2NBR	Ours
Noisy	Baseline, N2C	Baseline, N2N	BM3D	N2V	NBR2NBR	Ours
Noisy	Baseline, N2C	Baseline, N2N	BM3D	N2V	NBR2NBR	Ours

Figure 10. Visual comparison of denoising raw-RGB images in the challenging SIDD benchmark.