Blind2Unblind: Self-Supervised Image Denoising with Visible Blind Spots

Zejin Wang1,2 Jiazheng Liu1,3 Guoqing Li1 Hua Han1,3,*

1National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences
2School of Artificial Intelligence, University of Chinese Academy of Sciences
3School 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-
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 Noise2Noise [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, Neighbor2Noise (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 \( y \) of the ground truth \( x \), Noise2Void aims to minimize the following empirical risk:

\[
\arg\min_{\theta} \mathbb{E}_y \| f_{\theta}(y_{RF(i)}) - y \|^2_2, \tag{1}
\]

where \( f_{\theta}(\cdot) \) denotes the denoising model with parameter \( \theta \), \( y_{RF(i)} \) is a patch around pixel \( i \), \( y_i \) 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:

\[
\arg\min_{\theta} \mathbb{E}_y \| h(\Omega_{\theta y}) - y \|^2_2 + \lambda \cdot \| f_{\theta}(y) - y \|^2_2, \tag{2}
\]

note that \( \Omega_{\theta y} \) denotes a noisy masked volume in which the contained blind spots are exactly at all positions of \( y \), \( h(\cdot) \) is a global-aware mask mapper that samples the denoised pixels where the blind spots are located, and \( \lambda \) is a constant. Using Eq. (2) as the objective function for training learns the identity. That is, \( f_{\theta}(y) \) 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:

\[
\arg\min_{\theta} \mathbb{E}_y \| h(\Omega_{\theta y}) - y \|_1 + \lambda \cdot \| f_{\theta}(y) - y \|_1, \tag{3}
\]

where \( \hat{f}_{\theta}(\cdot) \) means that they do not contribute to updating the gradient. That is, \( \hat{f}_{\theta}(y) \) can be considered as an ordinary constant. It only remains to design a new objective function and the derivative of \( f_{\theta}(\Omega_{\theta y}) \) should contain \( \hat{f}_{\theta}(y) \) 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:

\[
\arg\min_{\theta} \mathbb{E}_y \| h(\Omega_{\theta y}) - y \|_1 + \lambda \cdot \| \hat{f}_{\theta}(y) - y \|^2_2. \tag{4}
\]

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:

\[
\mathcal{T}(y) = \| h(\Omega_{\theta y}) - y \|_2 + \lambda \cdot \| \hat{f}_{\theta}(y) - y \|^2_2 = \| h(\Omega_{\theta y}) - y \|^2_2 + \lambda^2 \| \hat{f}_{\theta}(y) - y \|^2_2 + 2\lambda \| (h(\Omega_{\theta y}) - y)^T (\hat{f}_{\theta}(y) - y) \|_1. \tag{5}
\]

Let \( d_1 = h(\Omega_{\theta y}) - y, d_2 = \hat{f}_{\theta}(y) - y, \text{cond} = (h(\Omega_{\theta y}) - y)^T (\hat{f}_{\theta}(y) - y) \). Here, we divide the objective function \( \mathcal{T}(y) \) into two cases:

i) If \( \text{cond} \geq 0 \), we have

\[
\mathcal{T}(y) = \| h(\Omega_{\theta y}) \|_2 + \lambda \| \hat{f}_{\theta}(y) - (\lambda + 1) y \|^2_2, \tag{6}
\]

ii) Similarly, when \( \text{cond} < 0 \), it holds that

\[
\mathcal{T}(y) = \| h(\Omega_{\theta y}) - y \|_2 + (\lambda - 1) \| \hat{f}_{\theta}(y) - y \|^2_2. \tag{7}
\]

Note that when \( \text{cond} < 0 \), the learning objective of \( \hat{f}_{\theta}(y) \) becomes \( h(\Omega_{\theta y}) \), instead of \( y \) as Eq. (6). Considering the optimal denoiser \( f^*_{\theta} \) that is trained to convergence, then \( \hat{f}_{\theta}(y) \) should be approximate to \( h(f^*_{\theta}(\Omega_{\theta y})) \) in Eq. (7). However, the masking operation \( \Omega_{\theta y} \) causes information loss, resulting in low denoising upper bound for \( h(f^*_{\theta}(\Omega_{\theta y})) \).

Besides, \( f^*_{\theta}(y) - h(f^*_{\theta}(\Omega_{\theta y})) \) in Eq. (7) indicates that \( f^*_{\theta}(y) \) should approximate \( h(f^*_{\theta}(\Omega_{\theta y})) \), which in turn suppresses the performance of \( f_{\theta}(y) \). Combining the above analysis, the final re-visible loss can be formulated as:

\[
\arg\min_{\theta} \mathbb{E}_y \| h(\Omega_{\theta y}) \|_2 + \lambda \| \hat{f}_{\theta}(y) - (\lambda + 1) y \|^2_2. \tag{8}
\]

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

\[
\tilde{x} = h(f^*_{\theta}(\Omega_{\theta y})) + \frac{\lambda \hat{f}^*_{\theta}(y)}{\lambda + 1}. \tag{9}
\]

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

As described in Section 3.3, the blind part $|h(f_\theta(\Omega_y)) - y|$ is used as a transition to optimize the re-visible one $\lambda \cdot |f_\theta(y) - y|$. 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 $f_\theta(\Omega_y)$, forcing the blind part to be zero. Thus, based on Eq. (8), we have the following constrained optimization problem:

$$\min_{\theta} E_y \|h(f_\theta(\Omega_y)) + \lambda f_\theta(y) - (\lambda + 1)y\|_2^2,$$

subject to

$$\|h(f_\theta(\Omega_y)) - y\|_2^2 = 0.$$  \hspace{1cm} (10)

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

$$\min_{\theta} E_y \|h(f_\theta(\Omega_y)) + \lambda f_\theta(y) - (\lambda + 1)y\|_2^2 + \eta \cdot \|h(f_\theta(\Omega_y)) - y\|_2^2 = 0.$$  \hspace{1cm} (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 $\Omega_y$ 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 $y$ with width $W$ and height $H$, details of the global-aware mask mapper $h(f_\theta(\Omega_y))$ are described as follows:

1. First, the noisy image $y$ is divided into $[W/s] \times [H/s]$ 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 $\Omega^x_y$ consists of $[W/s] \times [H/s]$
blind spots, where \(i, j \in \{0, ..., s - 1\}\). These mask images are further stacked to form a masked volume \(\Omega_y\).

3. For positions on the denoised volumes \(f_\theta(\Omega_y)\) 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 \(h(f_\theta(\Omega_y))\) is obtained.

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

### 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:

\[
\mathcal{L} \rightarrow \mathcal{L}_{\text{rev}} + \eta \cdot \mathcal{L}_{\text{reg}} = \frac{1}{2} \| h(f_\theta(\Omega_y)) + \lambda f_\theta(y) - (\lambda + 1)y \|^2 + \eta \cdot \| h(f_\theta(\Omega_y)) - y \|^2.
\]

where \(f_\theta(\cdot)\) denotes an arbitrary denoising network with different structures, \(h(\cdot)\) is a mask mapper capable of global modeling, \(\eta\) is a fixed hyper-parameter that determines the initial contribution of the blind term and the training stability, and \(\lambda\) is a variable hyper-parameter that controls the intensity of visible parts when converting from blind to un-blind. The initial and final values of \(\lambda\) are \(\lambda_s\) and \(\lambda_f\), respectively. To prevent identity mapping, we disable the gradient updating of \(f_\theta(y)\) 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 over-fitting. 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 x 128 patches for training. In practice, masked pixels are a weighted average of pixels in a \(3 \times 3\) neighborhood, as advocated in [18].

**Datasets for Synthetic Denoising.** Following the setting in [15, 20], we select 44,328 images with sizes between 256 x 256 and 512 x 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 \(\sigma\) 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],...
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 $\sigma$ 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, Noiser2Noise, 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
the effects of over smoothing among neighboring pixels. Information contained within noisy images and mitigates Because the transition from blind to non-blind preserves all posed method outperforms the state-of-the-art (NBR2NBR) 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.

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.

<table>
<thead>
<tr>
<th>Noise Type</th>
<th>Method</th>
<th>KODAK</th>
<th>BSD300</th>
<th>Set14</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gaussian</td>
<td>Baseline, N2C [29]</td>
<td>32.430/0.884</td>
<td>31.05/0.879</td>
<td>31.400/0.869</td>
</tr>
<tr>
<td></td>
<td>Baseline, N2N [21]</td>
<td>32.410/0.884</td>
<td>31.04/0.878</td>
<td>31.370/0.868</td>
</tr>
<tr>
<td></td>
<td>CBM3D [9]</td>
<td>31.67/0.868</td>
<td>30.48/0.861</td>
<td>30.88/0.854</td>
</tr>
<tr>
<td></td>
<td>Self2Self [28]</td>
<td>31.28/0.864</td>
<td>29.86/0.849</td>
<td>30.08/0.839</td>
</tr>
<tr>
<td></td>
<td>N2V [18]</td>
<td>30.32/0.821</td>
<td>29.34/0.824</td>
<td>28.84/0.802</td>
</tr>
<tr>
<td></td>
<td>Laine19-mu [20]</td>
<td>30.62/0.840</td>
<td>28.62/0.803</td>
<td>29.93/0.830</td>
</tr>
<tr>
<td></td>
<td>Laine19-pme [20]</td>
<td>32.48/0.883</td>
<td>30.99/0.877</td>
<td>31.36/0.866</td>
</tr>
<tr>
<td></td>
<td>Noise2Noise [25]</td>
<td>30.70/0.845</td>
<td>29.32/0.833</td>
<td>29.64/0.852</td>
</tr>
<tr>
<td></td>
<td>DBSN [33]</td>
<td>31.64/0.856</td>
<td>29.80/0.839</td>
<td>30.63/0.846</td>
</tr>
<tr>
<td></td>
<td>R2R [26]</td>
<td>32.25/0.880</td>
<td>30.91/0.872</td>
<td>31.32/0.865</td>
</tr>
<tr>
<td></td>
<td>NBR2NBR [15]</td>
<td>32.08/0.879</td>
<td>30.79/0.873</td>
<td>30.90/0.864</td>
</tr>
<tr>
<td></td>
<td>Ours</td>
<td>32.77/0.880</td>
<td>30.87/0.872</td>
<td>31.27/0.864</td>
</tr>
<tr>
<td>Poisson</td>
<td>Baseline, N2C [29]</td>
<td>32.51/0.875</td>
<td>31.07/0.866</td>
<td>31.41/0.863</td>
</tr>
<tr>
<td></td>
<td>Baseline, N2N [21]</td>
<td>32.50/0.875</td>
<td>31.07/0.866</td>
<td>31.39/0.863</td>
</tr>
<tr>
<td></td>
<td>CBM3D [9]</td>
<td>32.02/0.860</td>
<td>30.56/0.847</td>
<td>30.94/0.849</td>
</tr>
<tr>
<td></td>
<td>Self2Self [28]</td>
<td>31.57/0.860</td>
<td>29.87/0.841</td>
<td>29.97/0.840</td>
</tr>
<tr>
<td></td>
<td>N2V [18]</td>
<td>30.44/0.806</td>
<td>29.31/0.801</td>
<td>29.01/0.792</td>
</tr>
<tr>
<td></td>
<td>Laine19-mu [20]</td>
<td>30.52/0.833</td>
<td>28.43/0.794</td>
<td>29.71/0.822</td>
</tr>
<tr>
<td></td>
<td>Laine19-pme [20]</td>
<td>32.48/0.870</td>
<td>30.95/0.861</td>
<td>31.21/0.855</td>
</tr>
<tr>
<td></td>
<td>DBSN [33]</td>
<td>30.58/0.826</td>
<td>28.34/0.788</td>
<td>29.49/0.814</td>
</tr>
<tr>
<td></td>
<td>R2R [26]</td>
<td>31.50/0.850</td>
<td>30.56/0.855</td>
<td>30.84/0.850</td>
</tr>
<tr>
<td></td>
<td>NBR2NBR [15]</td>
<td>31.20/0.870</td>
<td>30.73/0.861</td>
<td>31.05/0.858</td>
</tr>
<tr>
<td></td>
<td>Ours</td>
<td>32.32/0.872</td>
<td>30.86/0.861</td>
<td>31.14/0.857</td>
</tr>
</tbody>
</table>

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.

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

<table>
<thead>
<tr>
<th>Methods</th>
<th>Network</th>
<th>SIDD Benchmark</th>
<th>SIDD Validation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline, N2C [29]</td>
<td>U-Net [29]</td>
<td>50.60/0.991</td>
<td>51.19/0.991</td>
</tr>
<tr>
<td>Baseline, N2N [21]</td>
<td>U-Net [29]</td>
<td>50.62/0.991</td>
<td>51.21/0.991</td>
</tr>
<tr>
<td>BM3D [9]</td>
<td>U-Net [29]</td>
<td>48.69/0.986</td>
<td>48.92/0.986</td>
</tr>
<tr>
<td>N2V [18]</td>
<td>U-Net [29]</td>
<td>48.01/0.983</td>
<td>48.55/0.984</td>
</tr>
<tr>
<td>Laine19-mu (Gaussian) [20]</td>
<td>U-Net [29]</td>
<td>49.82/0.989</td>
<td>50.44/0.990</td>
</tr>
<tr>
<td>Laine19-pme (Gaussian) [20]</td>
<td>U-Net [29]</td>
<td>42.17/0.935</td>
<td>42.87/0.939</td>
</tr>
<tr>
<td>Laine19-mu (Poisson) [20]</td>
<td>U-Net [29]</td>
<td>50.28/0.989</td>
<td>50.89/0.990</td>
</tr>
<tr>
<td>Laine19-pme (Poisson) [20]</td>
<td>U-Net [29]</td>
<td>48.46/0.984</td>
<td>48.98/0.985</td>
</tr>
<tr>
<td>DBSN [33]</td>
<td>DBSN [33]</td>
<td>49.56/0.987</td>
<td>50.13/0.988</td>
</tr>
<tr>
<td>R2R [26]</td>
<td>U-Net [29]</td>
<td>46.70/0.978</td>
<td>47.20/0.980</td>
</tr>
<tr>
<td>NBR2NBR [15]</td>
<td>U-Net [29]</td>
<td>50.47/0.990</td>
<td>51.06/0.991</td>
</tr>
<tr>
<td>Ours</td>
<td>U-Net [29]</td>
<td>50.79/0.991</td>
<td>51.36/0.992</td>
</tr>
</tbody>
</table>

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.

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

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.

### 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 $f_0(y)$ in Eq. (7) is suppressed by the mask term $h(f_0(\Omega_y))$, 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 mapper, global-aware mask mapper, and re-visible loss. -/- means cannot converge.

**Ablation study on Regularization Term.** The hyper-parameter $\eta$ 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, XDB23032028), Program of Beijing Municipal Science & Technology Commission (Z201100008420004), and Science and Technology Innovation 2030 Major Projects of China (2021ZD0204503).
References


[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


