AIMS Mathematics

https://www.aimspress.com/journal/Math

AIMS Mathematics, 10(6): 14140–14166
DOI: 10.3934/math.2025637
Received: 28 March 2025
Revised: 25 May 2025
Accepted: 04 June 2025
Published: 19 June 2025

Research article

Blind2Grad: Blind detail-preserving denoising via zero-shot with gradient regularized loss

Bolin Song[1], **Zhenhao Shuai**[2], **Yuanyuan Si**[1] **and Ke Li**[1,*]

[1] College of Artificial Intelligence, Dalian Maritime University, Dalian, 116026, China

[2] Department of Computer Science and Technology, Xiamen University, Xiamen, 361005, Fujian, China

* **Correspondence:** Email: ike@dlmu.edu.cn.

Abstract: Over the past few years, deep learning-based approaches have come to dominate the areas of image processing, with no exception to the field of image denoising. Recently, self-supervised denoising networks have garnered extensive attention in the zero-shot domain. Nevertheless, training the denoising network solely with zero-shot data will lead to significant detail loss, thereby influencing the denoising quality. In this work, we proposed a novel dual mask mechanism based on Euclidean distance selection as a global masker for the blind spot network, thereby enhancing image quality and mitigating image detail loss. Furthermore, we also constructed a recurrent framework, named Blind2Grad, which integrated the gradient-regularized loss and a uniform loss for training the denoiser and enabled the double-mask mechanism to be continuously updated. This framework was capable of directly acquiring information from the original noisy image, giving priority to the key information that needed to be retained without succumbing to equivalent mapping. Moreover, we conducted a thorough theoretical analysis of the convergence properties of our Blind2Grad loss from both probabilistic and game-theoretic viewpoints, demonstrating its consistency with supervised methods. Our Blind2Grad not only demonstrated outstanding performance on both synthetic and real-world datasets, but also exhibited significant efficacy in processing images with high noise levels.

Keywords: detail-preserving denoising; dual mask; gradient-regularized; convergence properties; self-supervised learning
Mathematics Subject Classification: 49N45, 68U10, 90C15, 94A08

1. Introduction

Image denoising, an essential task in low-level image processing, aims to recover valuable edges from noisy images [1–3]. Moreover, the quality of denoising methods significantly affects the performance of downstream tasks, such as classification, semantic segmentation, and target identification [4–6]. Thus, image denoising holds important practical value [7].

The field of image denoising has experienced notable advancements in the last few years [8], driven by the continuous development of advanced denoising techniques [9, 10]. Deep learning-based approaches provide target-driven yet powerful means for image denoising tasks comparing with traditional non-learning methods [11]. Specifically, these denoisers use a set of training samples to recover the clean counterpart $\mathbf{x}$ of the input noisy images $\mathbf{y}$ as well as learning the nonlinear mapping $f_\theta(\mathbf{y}) : \mathbf{y} \rightarrow \mathbf{x}$ with to-be-learned parameters [12]. These supervised denoising methods require large quantities of paired image (noisy/clean) data for training, presenting significant challenges in terms of data collection and annotations. [13, 14].

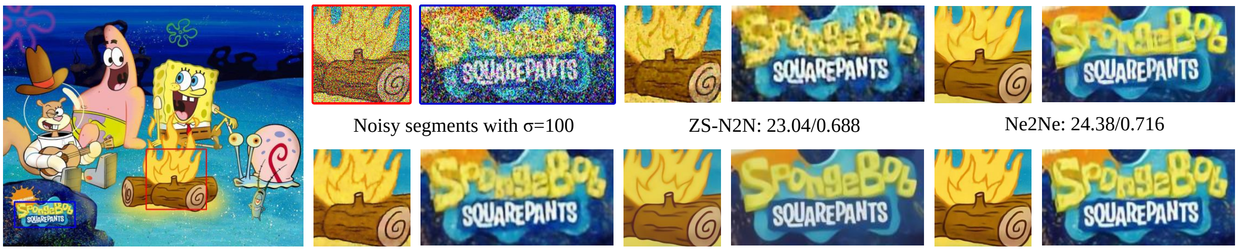
Currently, self-supervised denoisers without pre-collected paired data has become an enormous hotspot [15–17]. These methods follow two basic statistical assumptions: (1) the noise is independent and mean-zero; (2) the signals are not pixel-wise independent.

In blind spot denoising (BSD), the blind-spot-driven loss is to measure the blind-noisy difference, as shown in Eq (1.1)

$$\mathcal{L}_{d^c_{f_\theta}(\mathbf{y};\mathbf{\Omega}_m)} = \min_{\theta} \sum_{m=1}^{M} \|d^c_{f_\theta}(\mathbf{y};\mathbf{\Omega}_m)\|^2_{\mathbf{\Omega}_m}, \tag{1.1}$$

where $d^c_{f_\theta}(\mathbf{y};\mathbf{\Omega}_m) = f_\theta(\mathbf{\Omega}_m(\mathbf{y})) - \hat{\mathbf{y}}$ is the uniform loss of image content, $\|\cdot\|^2_{\mathbf{\Omega}_m} = \|(1-\mathbf{\Omega}_m)\odot\cdot\|^2_2$ is utilized to signify that the loss is calculated solely over the masked pixels, and $\hat{\mathbf{y}} = (\mathbf{I} - \mathbf{\Omega}_m)(\mathbf{y})$ with the identity operator $\mathbf{I}(\cdot)$; $\mathbf{\Omega}_m(\cdot)$ represents a masking operator where the masked pixel is set as zero. For self-supervised learning-based denoisers, f_θ indicates self-supervised blind-spot networks.

Blind-spot denoisers, employing random masking strategies, have acquired popularity in the field of image processing when clean image annotations or additional images are not required. Nevertheless, their performance is usually worse than the self-supervised methods with multiple images, let alone supervised denoisers [18]. Besides, the loss of the edge-aware information directly influences the quality of image denoising compared with those trivial contents of images, particularly for scenarios featuring high levels of diverse noises as shown in Figure 1.



Noisy segments with σ=100

ZS-N2N: 23.04/0.688

Ne2Ne: 24.38/0.716

S2S: 24.59/0.703

MPI: 23.43/0.697

Ours: 26.22/0.749

Figure 1. Superior performance of our Blind2Grad on removing Gaussian noise with σ=100, posted with peak signal-to-noise Ratio (PSNR) / structural similarity index measure (SSIM) values. It is worth noting that the image is referred to from the URL http://huaban.com/pins/606174904/.

In this paper, we present a novel self-supervised denoising framework, designated as Blind2Grad, to concurrently address the recovery challenges of multiple details and high-level noisy images. We introduce a dual mask mechanism that functions as a blind spot within the asymmetric networks, facilitating denoising effectively through zero-shot. To optimize the decoder of the asymmetric denoising networks, we adopt the threshold-based standout module to precisely retain the prior value of the gradient features, thereby maximizing the capacity of our proposed network to preserve edge information. Furthermore, we construct a gradient-regularized loss that incorporates the uniform loss to strike a balance of the information between the detailed and flat areas of the image. We utilize the potential attributes of Bayesian probability to demonstrate the relationship between the gradient-regularized loss and supervised loss, theoretically guaranteeing that Blind2Grad can maximize the retention of image edge information during denoising.

In summary, the main contributions of our work are as follows:

- We propose a novel idea for modifying the mask generation of BSD, which is not randomly sampled. Instead, the dual mask mechanism based on Euclidean distance is introduced herein, and the information of the flat area of the image is integrated to effectively preserve the crucial image texture in the denoising process, thereby enhancing the quality of the entire denoised image.
- We introduce a new framework called Blind2Grad (B2G), to enhance the accuracy of the double mask mechanism in capturing overlooked details within a rolling framework. In addition, the framework combines the gradient-regularized loss and the $\mathcal{T}$-Standout model (Threshold-based Standout), which can improve the preservation of detailed gradient information.
- Theoretically, we demonstrate that our loss function can converge to the supervised loss under certain probabilistic conditions. In combination with the global convergence property of the stable game, we can infer the convergence relationship between our proposed loss and the supervised loss. Our framework exhibits outstanding performance on both synthetic and real datasets.

The remainder of this paper is structured as follows: Section 2 undertakes a review of the related work. Section 3 puts forward our approach, which primarily encompasses the dual mask mechanism and the B2G framework. We conduct a theoretical study of the proposed loss as shown in Section 4. Section 5 showcases denoising experiments under diverse noise levels and compares the efficacy of proposed B2G with other typical self-supervised denoising methods. Finally, Section 6 draws conclusions and deliberates on future work.

2. Related work

2.1. Non-learning or learning with paired images

Denoising approaches initially originated from non-learning approaches, which are highly effective image processing techniques that can be divided as either filter-based [19] or model-based [20]. Model-based denoising approaches have surpassed filter-based approaches in terms of sophistication and effectiveness as the approaches have developed [21]. Nevertheless, these approaches might struggle to handle noise that does not conform to their handcrafted priors [22].

In the recent ten years, the advent of deep learning technology has led to a significant shift in the mainstream strategies of network-based denoising. For supervised denoising, convolutional neural networks (CNNs) trained with plenty of image pairs (noisy/clean) and have become the dominant technique [23–25]. However, supervised learning-based denoisers depend on a great number of noisy-clean image pairs for training.

2.2. *Self-supervised learning with multiple images*

One of the elegant methods known as Noise2Noise (N2N) emerged, using multiple noisy observations of identical scenes for training a deep denoising model [26]. By hypothesizing that the noise components are independent and of zero mean, minimizing the loss of the N2N method yields the same solution as the supervised training with ℓ_2-loss. Nevertheless, it is not always feasible to sample two independent noises for the same scene, which brings limitations to applications. Similarly, the Neighbor2Neighbor (Ne2Ne) method was proposed to acquire paired images via sub-sampling the noisy images for training [27]. Nevertheless, the sub-sampling strategy may compromise the integrity of image structures, resulting in an excessive level of smoothing. Noisy-As-Clean [28] and Noisier2Noise [29] introduce attached noise during the training of denoiser. Nevertheless, their practicability is limited due to the assumption of known noise distribution.

2.3. *Self-supervised learning with zero-shot*

On the one hand, as a pioneer in the field of self-supervised denoising approaches with zero-shot, Ulyanov et al., [30] put forward a deep image prior (DIP) approach for training the denoising network. On the other hand, existing blind-spot denoisers employ well-designed masking schemes wherein the neural network learns to fill in pixel gaps in the noisy images. For instance, Noise2Void (N2V) [31] and Noise2Self (N2S) [32] adopt the uniform pixel selection (UPS) strategy to constitute a particular mask, which is designed to prevent identity transformation. To improve its performance, this basic masking scheme is further refined by Self2Self (S2S) with the dropout skill [33]. Subsequently, Noise2Fast (N2F) [34] introduces a novel "checker-board" mask which is a down-sampling strategy that removes one half of all pixels in a checkerboard pattern. In addition, Zero-Shot Noise2Noise (ZS-N2N) [35] utilizes a lightweight network with regularized loss to achieve cost-effective noise reduction.

However, blind-spot denoisers may lose valuable structures (e.g., edges) while randomly masking pixels in single noisy image, thereby resulting in degraded noise canceling performance. The authors of Blind2Unblind (B2U) [36] and Masked Pretraining Inference (MPI) [37] have observed such weak points of blind-spot driven networks. Nevertheless, compared to other trivial points, the pixels that lie on valuable image structures should be directly emphasized while training blind-spot denoisers.

3. Proposed method

In this section, we first explain the motivation and strategy distribution of the dual mask mechanism. Besides, we establish a B2G framework based on gradient-regularized loss and the $\mathcal{T}$-standout model to further enhance the deployment of our proposed mask. Finally, we present the B2G training strategy and its computational complexity.

3.1. *Dual mask mechanism*

As shown on the first page, during the process of converting sensor analog electrical signals into output digital signals, noise reduction methods are employed for this purpose. However, the majority of the current zero-shot methods suffer from the problem of detail loss due to the randomness of the masks, which subsequently degrades the quality of denoised images. Therefore, it is essential to explore a new mask to complement the valuable details of the denoised image.

Given a clean image, we could initially observe the diverse manifestations of details and flatness via entropy, as depicted in Figure 2. If the random masking strategy (i.e., used in "S2S") is employed, it will be discovered that the outcomes presented in the penultimate column deviate significantly from the clean image in terms of detail restoration. This indicates that the random strategy is prone to losing valuable detail information, which prompts us to contemplate whether, instead of utilizing a random approach, we utilize conditional masking, that is, masking the flat and texture regions to varying degrees, thereby enabling the combination of more detail information with a certain degree of flat content for the prediction of the denoised image. Thus, a dual mask mechanism is selected. Specifically, a conditional masked convolution layer based on the random mask is introduced, as depicted in the blue box in Figure 3. The significance of the double mask mechanism is also affirmed in the results presented in the last column of Figure 2.

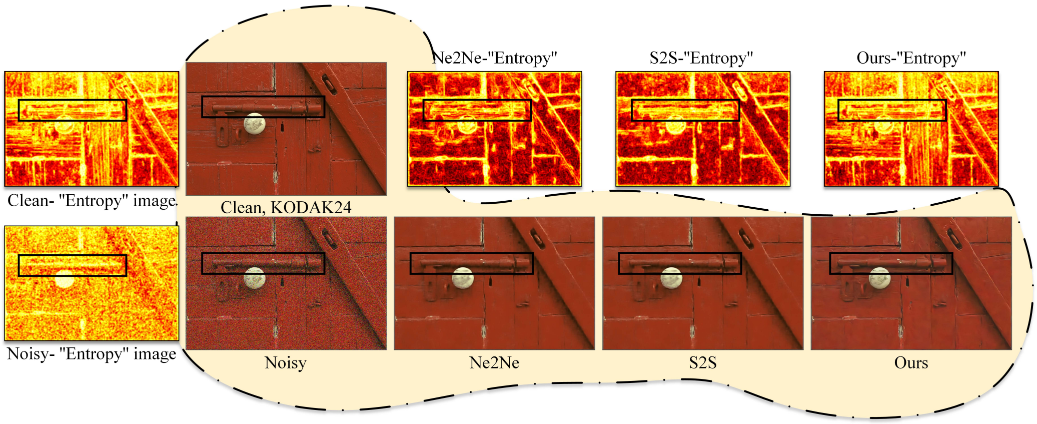


Figure 2. For the denoised image results (Gaussian noise with $\sigma = 50$), the entropy is closely associated with the used masking scheme. "S2S" represents the random masking strategy, while "Ne2Ne" denotes to the method without masking.

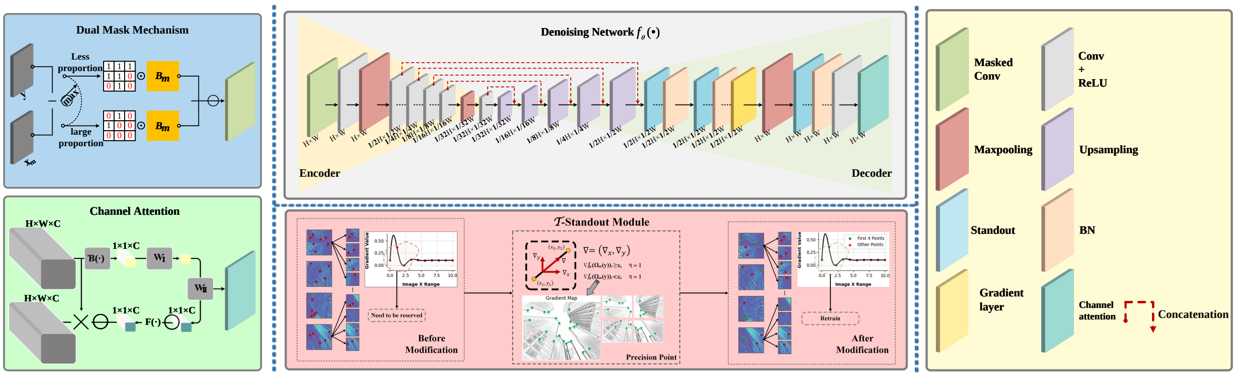


Figure 3. The architecture integrates four core components: (1) a dual mask mechanism based on Euclidean distance to to distinguish between flat and textured regions; (2) a denoising network $f_\theta(\cdot)$ composed of an encoder-decoder structure with convolutional, upsampling, and normalization layers; (3) a channel attention module that enhances feature representation by leveraging inter-channel dependencies; and (4) $\mathcal{T}$-Standout Module that identifies and emphasizes salient gradient-based regions for refined noise suppression.

For the dual mask mechanism, we have still used Bernoulli sampling for the first mask (preserving the content of the flat area), and select the conditional allocation based on Euclidean distance for the second mask, while maintaining the final mask rate at 25%. The specific dual mask mechanism is depicted on the left side of Figure 3. Given a medium clean image, denoted as $\mathbf{x}_m$, where the noise has been eliminated but the details are blurred, we compute the pixel difference between the noisy $\mathbf{y}$ and $\mathbf{x}_m$, as presented in formula (3.1):

$$\varepsilon_g = max\sqrt{(\mathbf{y}_i - \mathbf{x}_{m,i})^2}, \tag{3.1}$$

where variable "i" represents the pixel value, and the threshold ε_g is used to distinguish between the texture area and the flat area of the image.

Since noise has less impact on areas with substantial gradient changes as opposed to flat areas, the entropy graph in Figure 2 still indicates that there is still detailed information in the noisy image. Furthermore, the more distinct the detailed information of the denoised image itself is, the more prominent the performance becomes. We have utilized a binary indexed transform to obtain the indices $I_i := \{i_n\}_{n=1}^{N}$ of the details that need to be preserved, i.e., $i_n = 1$ if $\varepsilon_g \geq \mathcal{G}_{max}$; otherwise $i_n = 0$, with $\varepsilon_g < \mathcal{G}_{max}$.

To achieve the deployment of the dual mask mechanism, it is necessary to retain the pixel values encompassing detail information for variable $i_n = 1$, concurrently enhancing the masking ratio of variable $i_n = 0$. Subsequently, in the second conditional sampling based on Euclidean distance, we initially calculate the masking quantity corresponding to the Bernoulli sampling B_m as depicted in Eq (3.2).

$$\begin{aligned} IB_m &= I_i^{\mathcal{I}} \odot 5.4.2 + I_i^{\mathcal{II}} \odot B_m, \\ IB_m^c &= I_i^{\mathcal{I}^c} \odot B_m + I_i^{\mathcal{II}^c} \odot B_m. \end{aligned} \tag{3.2}$$

where $I_i^{\mathcal{I}}$ and $I_i^{\mathcal{II}}$ represent the first and second layers in the dual mask mechanism, with $I_i^{\mathcal{I}^c} = 1 - I_i^{\mathcal{I}}$. Our ultimate objective is to diminish the nonzero IB_m, denoted as IB_m^-; concurrently, augment the proportion of IB_m^c, designated as $IB_m^{c,+}$. Thus, the final double masking mechanism is expressed as:

$$\mathbf{\Omega}_m = IB_m^- + IB_m^c + IB_m^{c,+}. \tag{3.3}$$

What is requisite is that we propose a dual mask mechanism $\mathbf{\Omega}_m$ as a mask strategy for the blind spot denoising network, which can effectively retain the significant detail information of the denoised image to a certain extent. However, selecting the optimal value ε_g also demands the continuous enhancement of the clarity of $\mathbf{x}_m$. Therefore, in the subsequent section, we will introduce a novel loop framework B2G to generate high-quality $\mathbf{x}_m$, thereby obtaining the optimal ε_g for better discrimination of the texture and flatness of the image.

3.2. Blind2Grad framework

Since the excessive number of symbol definitions is to be presented in this chapter, a symbol summary table is provided initially (Table 1). Our proposed framework B2G is a recurrent strategy founded on the collaborative operation of gradient-regularized loss, $\mathcal{T}$-standout, and the channel attention mechanism, which is capable of further precisely guiding the generation of $\mathbf{\Omega}_m$ and preserving the significant details of the denoised image.

Table 1. Symbol summary table for our Blind2Grad.

Symbol	Definition	Example / Intuition
$\tilde{d}^{c}_{f_\theta}(\mathbf{y};\mathbf{\Omega}_m)$	Uniform loss (absolute pixel-wise difference)	$\lvert d^{c}_{f_\theta}(\mathbf{y};\mathbf{\Omega}_m)\rvert = \lvert f_\theta(\mathbf{\Omega}_m(\mathbf{y})) - (\mathbf{I}-\mathbf{\Omega}_m)(\mathbf{y})\rvert$ is the uniform loss of image content. E.g., appears at the top of Figure 2
$\tilde{d}^{g}_{f_\theta}(\mathbf{y};\mathbf{\Omega}_m)$	Gradient-regularized loss (absolute pixel-wise difference)	$\lvert d^{g}_{f_\theta}(\mathbf{y};\mathbf{\Omega}_m)\rvert = \lvert\nabla f_\theta(\mathbf{y}) - \nabla f_\theta(\mathbf{\Omega}_m(\mathbf{y}))\rvert$ is established to measure the gradient deviation between the whole denoised image and the recovered one from its masked instances. E.g., appears at the top of Figure 2
$d^{c_i}_{f_\theta}(\mathbf{y};\mathbf{\Omega}_m)$	The i-th elements of the uniform loss	$d^{c}_{f_\theta}(i)$
$d^{g_i}_{f_\theta}(\mathbf{y};\mathbf{\Omega}_m)$	The i-th elements of the gradient-regularized loss	$d^{g}_{f_\theta}(i)$
$\boldsymbol{\theta}^*,\gamma,\omega$	The corresponding network with $\boldsymbol{\theta}^*$	Used in: $f^*_{\boldsymbol{\theta}}(\cdot)$. E.g., appears in the middle of Figure "Denoising Network"
$f^*_{\boldsymbol{\theta}}$	Parameters for denoising/penalty/regularization network	E.g., CNN or U-Net
$\nabla\mathbf{x}^*$	Gradient of output with respect to input	$\nabla f^*_{\boldsymbol{\theta}}(\mathbf{y})$
ϵ_s	Threshold for neuron-level gradient selection	$\epsilon_s = 25$. E.g., appears in Eq (3.11)
$\mathcal{B}(\cdot)$	Global pooling operator	Average or max over channels. E.g., appears in Eq (3.12)
$f_\theta(\mathbf{\Omega}_m(\mathbf{y})(i,j)$	Value at position (i,j) of c-th output feature	A pixel value
$\mathcal{W}_{\mathrm{I}},\mathcal{W}_{\mathrm{II}}$	Convolution or downscaling weight sets	Used in network blocks. E.g., appears in the lower left corner of Figure "Channel Attention"

3.2.1. Gradient-regularized loss

To complement valuable structural details to the restored images, we believe that a blind spot denoiser should be learned by masking pairs not only in the color space as normal does, but also in the edge detail space of the image. In consequence, a loss of gradient is established to measure the gradient deviation between the whole denoised image (i.e., $f_{\boldsymbol{\theta}}(\mathbf{y})$) and the recovered one from its masked instances (i.e.,$f_{\boldsymbol{\theta}}(\mathbf{\Omega}_m(\mathbf{y}))$), that is,

$$d^{g}_{f_\theta}(\mathbf{y};\mathbf{\Omega}_m) = \nabla f_{\boldsymbol{\theta}}(\mathbf{y}) - \nabla f_{\boldsymbol{\theta}}(\mathbf{\Omega}_m(\mathbf{y})). \tag{3.4}$$

It is reasonable to minimize the loss of Eq (3.4) to expect that the structural information of the recovered image of $\mathbf{\Omega}_m(\mathbf{y})$ should not be deviated too much in comparison with the one with global details. In this way, image details that are lost due to the masking schemes can be properly supplemented.

Although compared to other pixels, image detail information is crucial for understanding what a graph is; it occupies a rather limited portion of the overall region, such that gradient minimization alone is not sufficient. Thus, we desire to propose a loss that considers both gradient and color differences, with a predetermined parameter ω as the balancing of the two terms. Besides, in order to avoid the case that $d^{c}_{f_\theta}(\mathbf{y};\mathbf{\Omega}_m) = -d^{g}_{f_\theta}(\mathbf{y};\mathbf{\Omega}_m) \neq 0$, the loss is proposed as:

$$\min_{\boldsymbol{\theta}} \sum_{m=1}^{M} \|\tilde{d}^{c}_{f_\theta}(\mathbf{y};\mathbf{\Omega}_m) + \omega\tilde{d}^{g}_{f_\theta}(\mathbf{y};\mathbf{\Omega}_m)\|^2_{\mathbf{\Omega}_m}, \tag{3.5}$$

where the i-th element of $\tilde{d}^{c}_{f_\theta}(\mathbf{y};\mathbf{\Omega}_m)/\tilde{d}^{g}_{f_\theta}(\mathbf{y};\mathbf{\Omega}_m)$ denotes the absolute value of the i-th element of $d^{c}_{f_\theta}(\mathbf{y};\mathbf{\Omega}_m)/d^{g}_{f_\theta}(\mathbf{y};\mathbf{\Omega}_m)$.

Specifically, we concretely formulate the Eq (3.5) as:

$$\min_{\boldsymbol{\theta}} \sum_{m=1}^{M} \sum\nolimits_i (|d^{c_i}_{f_\theta}(\mathbf{y};\mathbf{\Omega}_m)| + \omega|d^{g_i}_{f_\theta}(\mathbf{y};\mathbf{\Omega}_m)|)^2_{\mathbf{\Omega}_m}, \tag{3.6}$$

where the $d^{c_i}_{f_\theta}(\mathbf{y};\mathbf{\Omega}_m)$ and $d^{g_i}_{f_\theta}(\mathbf{y};\mathbf{\Omega}_m)$ are the i-th elements of $d^{c}_{f_\theta}(\mathbf{y};\mathbf{\Omega}_m)$ and $d^{g}_{f_\theta}(\mathbf{y};\mathbf{\Omega}_m)$, respectively.

There are theoretically two cases of $d^{c_i}_{f_\theta}(\mathbf{y}; \mathbf{\Omega}_m)$ in Eq (3.6), that is, $d^{c_i}_{f_\theta}(\mathbf{y}; \mathbf{\Omega}_m) \geq 0$ or $d^{c_i}_{f_\theta}(\mathbf{y}; \mathbf{\Omega}_m) < 0$. Nevertheless, it is an impossible situation of $d^{c_i}_{f_\theta}(\mathbf{y}; \mathbf{\Omega}_m) < 0$ since the complete original information will be preserved through $f_{\boldsymbol{\theta}}$. Meanwhile, the pixels of $\hat{\mathbf{y}}$ are less than that of $\mathbf{\Omega}_m(\mathbf{y})$.

With $d^{c_i}_{f_\theta}(\mathbf{y}; \mathbf{\Omega}_m) \geq 0$, there are two cases of $d^{g_i}_{f_\theta}(\mathbf{y}; \mathbf{\Omega}_m)$, i.e., $d^{g_i}_{f_\theta}(\mathbf{y}; \mathbf{\Omega}_m) \geq 0$ or $d^{g_i}_{f_\theta}(\mathbf{y}; \mathbf{\Omega}_m) < 0$. Similarly, due to the pixels contained in $\mathbf{\Omega}_m(\mathbf{y})$ being less than that of $\mathbf{y}$, the detail-preserving network will not cause the case of $d^{g_i}_{f_\theta}(\mathbf{y}; \mathbf{\Omega}_m) < 0$. The results confirm that only the case of $d^{c_i}_{f_\theta}(\mathbf{y}; \mathbf{\Omega}_m) \geq 0$ and $d^{g_i}_{f_\theta}(\mathbf{y}; \mathbf{\Omega}_m) \geq 0$ is a real possibility.

Therefore, we desire to propose our framework by minimizing a gradient-regularized loss, i.e.,

$$\min_{\boldsymbol{\theta}} \sum_{m=1}^{M} \|d^c_{f_\theta}(\mathbf{y}; \mathbf{\Omega}_m) + \omega d^g_{f_\theta}(\mathbf{y}; \mathbf{\Omega}_m)\|^2_{\mathbf{\Omega}_m}. \tag{3.7}$$

Hypothesize that $\boldsymbol{\theta}^*$ is an optimal solution of Eq (3.7), and the $f^*_{\boldsymbol{\theta}}$ is the corresponding denoising network with $\boldsymbol{\theta}^*$. Then, we denote an $\mathbf{x}^*$ as: $x^* = f^*_{\boldsymbol{\theta}}(\mathbf{\Omega}_m(\mathbf{y})) + \omega d^g_{f^*_\theta}(\mathbf{y}; \mathbf{\Omega}_m)$. If $\omega \to \infty$ exists, working out the Eq (3.7) will bring about $d^g_{f^*_\theta}(\mathbf{y}; \mathbf{\Omega}_m) \to 0$ and $f^*_{\boldsymbol{\theta}}(\mathbf{y_0}) = f^*_{\boldsymbol{\theta}}(\mathbf{\Omega}_m(\mathbf{y_0}))$ in turn. By denoting the unmasked pixels of $\Omega_m(\mathbf{y})$ as $\mathbf{y}_0$, that is, $\mathbf{y}_0 \in \mathbf{y} \cap \mathbf{\Omega}_m(\mathbf{y})$, we have $f^*_{\boldsymbol{\theta}}(\mathbf{y_0}) = f^*_{\boldsymbol{\theta}}(\mathbf{\Omega}_m(\mathbf{y_0}))$. Then, from the theorem of functional continuity, there holds $\lim_{\omega \to \infty} \mathbf{x}^* = f^*_{\boldsymbol{\theta}}(\mathbf{\Omega}_m(\mathbf{y})) = f^*_{\boldsymbol{\theta}}(\mathbf{y})$ with the given noisy $\mathbf{y}$. Thus, it is revealed that our gradient-regularized loss is in a position to preserve structural details while restoring the clean image.

It is indeed worthwhile to note that there are two potential conditions: when minimizing the gradient-regularized loss, the $d^{c_i}_{f_\theta}(\mathbf{y}; \mathbf{\Omega}_m) \geq 0$ and $d^{g_i}_{f_\theta}(\mathbf{y}; \mathbf{\Omega}_m) \geq 0$ should be covered. Furthermore, instead of adding those direct constraints, our objective is to minimize the gradient-regularized loss by restricting $d^c_{f_\theta}(\mathbf{y}; \mathbf{\Omega}_m) = 0$, forming the optimization as

$$\begin{aligned} &\min_{\boldsymbol{\theta}} \sum_{m=1}^{M} \|d^c_{f_\theta}(\mathbf{y}; \mathbf{\Omega}_m) + \omega d^g_{f_\theta}(\mathbf{y}; \mathbf{\Omega}_m)\|^2_{\mathbf{\Omega}_m}, \\ &\text{s.t. } d^c_{f_\theta}(\mathbf{y}; \mathbf{\Omega}_m) = 0. \end{aligned} \tag{3.8}$$

Our design accentuates the gradient-based loss term $d^g_{f_\theta}(\mathbf{y}; \mathbf{\Omega}_m)$, which directly focuses on image edge structures. The incorporation of the color consistency term $d^c_{f_\theta}(\mathbf{y}; \mathbf{\Omega}_m)$ serves a secondary function: it aids in maintaining the equilibrium between textured regions and flat areas within the reconstructed image. Therefore, we introduce the constraint $d^c_{f_\theta}(\mathbf{y}; \mathbf{\Omega}_m) = 0$ as a penalized regularization term with a penalty parameter γ. The term $d^c_{f_\theta}(\cdot)$ is designed primarily to supplement color consistency in flat regions and should not dominate the optimization. In addition, the final optimization model for learning our B2G framework is formulated as

$$\min_{\boldsymbol{\theta}} \sum_{m=1}^{M} (\|d^c_{f_\theta}(\mathbf{y}; \mathbf{\Omega}_m) + \omega d^g_{f_\theta}(\mathbf{y}; \mathbf{\Omega}_m)\|^2 + \gamma \|d^c_{f_\theta}(\mathbf{y}; \mathbf{\Omega}_m)\|^2). \tag{3.9}$$

We employ Eq (3.9) as the loss function of the entire B2G framework. The specific procedure can be observed as shown in Figure 4. Regarding $\mathbf{\Omega}_m$, we initially adopt the Bernoulli sampling approach as $\mathbf{\Omega}_1$, and subsequently obtain the corresponding $\mathbf{x}_1 = f_\theta(\mathbf{y})_1$ after the first network cycle. Hence, given a denoised image $\mathbf{y}$, when $m > 1$, our proposed B2G can be characterized as:

$$\mathbf{y} \to \{\mathbf{x}_m = f_\theta(\mathbf{y})_m \to \mathbf{\Omega}_m \to d^c_{f_\theta}(\mathbf{y}; \mathbf{\Omega}_m)/d^g_{f_\theta}(\mathbf{y}; \mathbf{\Omega}_m)\}_{m=1}^{m=M}. \tag{3.10}$$

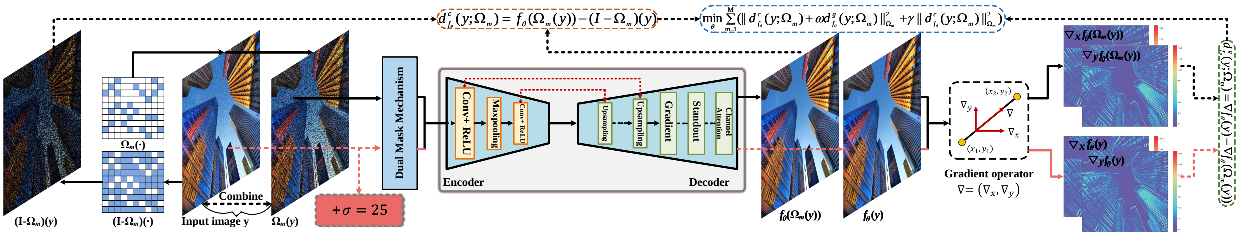

$d^c_{f_\theta}(y;\Omega_m) = f_\theta(\Omega_m(y)) - (I-\Omega_m)(y)$
$\min_\theta \sum_{m=1}^{M}(\| d^c_{f_\theta}(y;\Omega_m) + \omega d^g_{f_\theta}(y;\Omega_m)\|^2_{\Omega_m} + \gamma \| d^c_{f_\theta}(y;\Omega_m)\|^2_{\Omega_m})$
$\Omega_m(\cdot)$
$(I-\Omega_m)(y)$
$(I-\Omega_m)(\cdot)$
Combine
Input image y
$\Omega_m(y)$
$+\sigma = 25$
Dual Mask Mechanism
Conv+ ReLU
Maxpooling
Conv+ ReLU
Upsampling
Upsampling
Gradient
Standout
Channel Attention
Encoder
Decoder
$f_\theta(\Omega_m(y))$
$f_\theta(y)$
∇_y
(x_2, y_2)
∇
∇_x
(x_1, y_1)
Gradient operator
$\nabla = (\nabla_x, \nabla_y)$
$\nabla x f_\theta(\Omega_m(y))$
$\nabla y f_\theta(\Omega_m(y))$
$\nabla x f_\theta(y)$
$\nabla y f_\theta(y)$
$d^g_{f_\theta}(y;\Omega_m) = \nabla f_\theta(y) - \nabla f_\theta(\Omega_m(y))$

Figure 4. Architecture of our Blind2Grad framework. Specifically, the global mask $\mathbf{\Omega}_m$ is generated utilizing the dual mask mechanism to create a mask body with blind spots on the noisy image $\mathbf{y}$, Next, the gradient-regularized loss realizes the image valuable detail restoration with the gradient deviation term $d^g_{f_\theta}(\mathbf{y}; \mathbf{\Omega}_m)$ as a medium and the $d^c_{f_\theta}(\mathbf{y}; \mathbf{\Omega}_m)$ used to supplement content information about color content. By combining $\mathcal{T}$-standout module and channel attention, a denoising network $f_\theta(\cdot)$ with the optimal detail-preserving is obtained. Therefore, $\otimes$ denotes element-wise product, $\odot$ represents matrix dot product, and $|| \cdot ||$ represents the ℓ_2 norm of a vector throughout this paper.

3.2.2. $\mathcal{T}$-standout and channel attention

For the dropout, it merely randomly discards certain neurons, preventing them from participating in the calculation during the current training round. This method reduces overfitting while augmenting the quantity of training samples.

However, there exists a risk of losing detailed information in the process of random discarding. Hence, we have developed a new $\mathcal{T}$-standout strategy in the decoder, as depicted in the pink box diagram in Figure 3. Instead of randomly discarding neurons, the gradient threshold strategy is used to reselect the discarded neurons. Specifically, the discarded neurons can be represented as $\{f_\theta(\mathbf{\Omega}_m)_0, f_\theta(\mathbf{\Omega}_m)_1, \ldots, f_\theta(\mathbf{\Omega}_m)_i\}$, i.e., denoised image is randomly divided into any combination. We set the threshold ϵ_s, the weight $\{\eta_0, \eta_1, \ldots, \eta_j\}$, and $\{\eta_0, \eta_1, \ldots, \eta_j\} \subset \{0, 1\}$. such that

$$\begin{cases} \nabla f_\theta(\mathbf{\Omega}_m(\mathbf{y}))_i \geq \epsilon_s, & \eta_j = 1, \\ \nabla f_\theta(\mathbf{\Omega}_m(\mathbf{y}))_i < \epsilon_s, & \eta_j = 0, \end{cases} \tag{3.11}$$

exists herein. That is, when $\eta_j = 1$, the corresponding neuron $f_\theta(\mathbf{\Omega}_m(\mathbf{y}))_i$ cannot be discarded, thereby further increasing the quantity of training samples and replenishing the detailed information of the image once again.

On the other hand, to leverage the interdependence between feature channels, a channel attention mechanism is introduced, as shown on the left side of Figure 3. To capture the complete channel dependence from the collective information of the denoised image $f_\theta(\mathbf{\Omega}_m(\mathbf{y}))$, we opt to introduce gating with a sigmoid function mechanism:

$$\mathcal{S}^* = F(\mathcal{W}_{\text{II}}\mathcal{W}_{\text{I}}\mathcal{B}(f_\theta(\mathbf{\Omega}_m(\mathbf{y}))), \tag{3.12}$$

where $\mathcal{B}(\cdot) = 1/(H \times W) \sum_{i=1}^{H} \sum_{i=1}^{W} f_\theta(\mathbf{\Omega}_m(\mathbf{y})(i, j)$, $F(\cdot)$ represents the sigmoid gating function [38].

3.3. Network architecture and training scheme

Our B2G network employs an encoder-decoder architecture with a specialized $\mathcal{T}$-Standout Module for gradient-aware processing. Based on the architectural diagram as shown in Figure 3, we present a comprehensive description of each component:

The encoder model follows a progressive downsampling strategy to efficiently extract hierarchical features. (1) Input Processing: The network accepts a noisy input image of the size $H * W$; (2)Masked Convolution: Used for initial feature extraction with noise-aware processing; (3) Feature Extraction Path: The encoder consists of multiple stages with precise dimensional transformations; (4) Initial masked convolution at full resolution ($H*W$) maxpooling operation reducing dimensions to $H/2*W/2$.

Subsequent convolutional blocks with progressive downsampling through the following resolutions:

$$H/2 * W/2 \to H/4 * W/4 \to H/8 * W/8 \to H/16 * W/16 \to H/32 * W/32. \tag{3.13}$$

Each convolutional block incorporates convolution (Conv) + rectified linear unit (ReLU) operations as indicated in our architecture diagram.

Furthermore, the decoder implements a symmetric upsampling path with specialized components for feature refinement: Progressive Upsampling: Starting from the bottleneck features, the decoder gradually restores spatial dimensions:

$$H/32 * W/32 \to H/16 * W/16 \to H/8 * W/8 \to H/4 * W/4 \to H/2 * W/2. \tag{3.14}$$

Each upsampling block utilizes dedicated upsampling operations as shown in Figure 3. (1) Upsampling operations to increase spatial resolution Batch Normalization (BN) for training stability and improved convergence; (2) Gradient layer for edge-aware processing and structural preservation; (3) Channel attention mechanism to emphasize important feature channels concatenation operations (indicated by red dashed lines in our diagram) to merge features from different processing stages.

The decoder concludes with operations that restore the original $H * W$ resolution, producing the denoised output with preserved structural details. Besides, a key innovation in our architecture is the $\mathcal{T}$-Standout Module, which operates on intermediate features to enhance gradient preservation. The module incorporates a retrain component to further refine the processed features, ensuring optimal gradient preservation.

Moreover, we provide a quick summary of our proposed Blind2Grad training procedure in Algorithm 1. Specifically, we train the asymmetric denoising networks, with the gradient-regularized loss to achieve the goal of preserving detail while denoising.

The main computational complexity of our B2G arises from the dual mask machine ($\mathbf{\Omega}_m$) and $\mathcal{T}$-standout module. Specifically, for the $\mathcal{T}$-standout module, it is combined with standout, and its complexity $O(h(\Gamma((H \times (W-1)) + ((H-1) \times W))_g))$ is mainly related to the image resolution $H \times W$, where Γ and h represent the number of iterations and layers of the neural network, respectively. Additionally, for the dual mask $\mathbf{\Omega}_m(\cdot)$, its complexity mainly comes from Eq (3.1), and $O(h(\Gamma(H \times W)_{\varepsilon_g}))$ involves gathering the characteristics of Euclidean distance. Then, the whole complexity is obtained by

$$O(n) \sim O(h(\mathcal{M}(H-1, W-1; \Gamma)_g + \mathcal{M}(H, W; \Gamma)_{\varepsilon_g})), \tag{3.15}$$

where $\mathcal{M}(H-1, W-1; \Gamma)_g = \Gamma((H \times (W-1)) + ((H-1) \times W))_g$, $\mathcal{M}(H, W; \Gamma)_{\varepsilon_g} = \Gamma(H \times W)_{\varepsilon_g}$.

Algorithm 1 Training of Blind2Grad

```
Require: A noisy image $\mathbf{y}$;
    dual mask sampling $\mathbf{\Omega}_m(\cdot)$;
    denoising network $f_{\boldsymbol{\theta}}(\mathbf{y})$;
    gradient-regularized parameter $\omega$;
    penalty parameter $\gamma$.
Ensure: The denoising network $f^*_{\boldsymbol{\theta}}(\cdot)$ with the optimal $\boldsymbol{\theta}^*$.
while not converged do
    Sample a noisy image $\mathbf{y}$.
    Generate a dual-mask sampling $\mathbf{\Omega}_m(\cdot)$ with a determined rate.
    Derive a pair of sample images $(\mathbf{\Omega}_m(\mathbf{y}), \hat{\mathbf{y}})$.
    For input $\mathbf{\Omega}_m(\mathbf{y})$, compute the loss of color content difference $d^c_{f_\theta}(\mathbf{y}; \mathbf{\Omega}_m)$ using Eq (1.1).
    For the original noisy image $\mathbf{y}$, derive the gradient deviation $d^g_{f_\theta}(\mathbf{y}; \mathbf{\Omega}_m)$ by Eq (3.4);
    Combine $d^c_{f_\theta}(\mathbf{y}; \mathbf{\Omega}_m)$ and $d^g_{f_\theta}(\mathbf{y}; \mathbf{\Omega}_m)$ yielding the gradient-regularized loss in Eq (3.9);
    Using adaptive moment estimation (Adam) to minimize the objective function in Eq (3.9) to
      update the denoising network $f_{\boldsymbol{\theta}}$.
end while
```

4. Theoretical analysis

For being self-contained, we will provide necessary assumptions and propositions in Section 4.1 before presenting our main theoretical results.

4.1. Necessary preliminaries

For denoising with our proposed self-supervised blind-spot denoising networks, it is necessary to prevent the training from learning into the identity mapping. Nevertheless, it is necessary to assume that the denoising function $f_{\boldsymbol{\theta}}(\mathbf{y})_J$ should be $\mathcal{J}$-invariant [32], as claimed below.

Assumption 1 ($\mathcal{J}$-invariant)**.** *Our proposed denoising function $f_{\boldsymbol{\theta}}(\mathbf{y})_J$ is assumed to be $\mathcal{J}$-invariant, meaning that the value of $f_{\boldsymbol{\theta}}(\mathbf{y})_J$ does not depend on $\mathbf{y}_J$ for all $J \in \mathcal{J}$, where $\mathcal{J}$ is a partition of the dimensions $\{1, \ldots, m\}$; $f_{\boldsymbol{\theta}}(\mathbf{y})_J$ and $\mathbf{y}_J$ denote the values of $f_{\boldsymbol{\theta}}(\mathbf{y})$ and $\mathbf{y}$ on J, respectively [39].*

Definition 1. *We have $f_\theta(y) : \{i \in \mathbf{y} : i \in [0, 255]\}$ as the support set of $\mathbf{y}$. Hence, the state set encompassing the support of the $\mathbf{\Omega}_m(\mathbf{y})$ is defined as $f_\theta(\mathbf{\Omega}_m(\mathbf{y})) = \{\mathbf{\Omega}_m(\mathbf{y}) \in \sum_{m=1}^{m=M} \mathbf{\Omega}_m(\mathbf{y}) : f_\theta(\mathbf{\Omega}_m(\mathbf{y})) \subseteq f_\theta(\mathbf{y})\}$, along with $f_\theta(\mathbf{y}) = \sum_{m=1}^{m=M} f_\theta(\mathbf{\Omega}_m(\mathbf{y}))$ as the state set containing the support of the $\mathbf{\Omega}_m(\mathbf{y})$. At this juncture, a distance-like function $\mathbf{d}_{\mathbf{\Omega}_m} : \mathbf{\Omega}_m(\mathbf{y}) \rightarrow \mathcal{R}$ can be defined, which satisfies:*

$$\mathbf{d}_{\mathbf{\Omega}_m}(\mathbf{\Omega}_m(\mathbf{y})) = \sum_{i \in \mathbf{\Omega}_m(\mathbf{y})} \mathbf{x}^*_i log \frac{\mathbf{x}^*_i}{f_\theta(\mathbf{\Omega}_m(\mathbf{y}))_i}. \tag{4.1}$$

where $\mathbf{x}^*_i = f^*_\theta(\mathbf{y})$, and both $\mathbf{x}^*$ and $f_\theta(\mathbf{y})$ conform to the normal distribution of $\mathcal{N}(0, 1)$.

We present two necessary propositions for proving the main theorem as follows.

Proposition 1. *Supposing that there exists a constant $s \in [c, d]$ for the function $f(y_{j_x}, y_{j_y})$ with $y_{j_x} \in [a, b]$ and $y_{j_y} \in [c, d]$, then for every $\varepsilon > 0$, there has a $\delta > 0$, such that*

$$|f(y_{j_x}, y_{j_y}) - f(y_{j_x}, s)| < \varepsilon, \tag{4.2}$$

for all $f(y_{j_x}) \in [a, b]$ and $f(y_{j_y}) \in (s - \delta, s + \delta)$.

Then, $E(f(y_j)) = \int_a^b y_j f(y_j) dy_j$ is also continuous on $[a, b]$ or $[c, d]$.

Given the continuity of $E(f(y_j))$, it lays the conditional basis for the derivation of Proposition 1.

Proposition 2. *Supposing that $f(y_{j_x}, y_{j_y})$ and its derived function are continuous on $[a, b] \times [c, d]$, then with Proposition 1, we have that $I(y_j) = \int_a^b f(y_j) dy_j$ has a continuous derived function $I^{'}(y_j)$ on $[a, b]$ or $[c, d]$; meanwhile,*

$$\frac{d}{dy_j} \int_a^b y_j f(y_j) dy_j = \int_a^b \frac{d}{dy_j} y_j f(y_j) dy_j. \tag{4.3}$$

4.2. Main results

With the above given assumption and propositions, we will first provide a required corollary of the gradient operator we considered in our B2G network. For smooth reading, we leave all the detailed proofs in the Supplementary material.

Corollary 1. *Assume that the $f(\mathbf{\Omega}_m(\mathbf{y}))$ and $f(\mathbf{y})$ are unbiased estimators of $\hat{\mathbf{x}}$, that is, $E[f(\mathbf{\Omega}_m(\mathbf{y}))|\hat{\mathbf{x}}] = \hat{\mathbf{x}}$ and $E[f(\mathbf{y})|\hat{\mathbf{x}}] = \hat{\mathbf{x}}$, where $\hat{\mathbf{x}} = (\mathbf{I} - \mathbf{\Omega}_m)(\mathbf{x})$. By supposing that $f(\mathbf{\Omega}_m(\mathbf{y}))$ and $f(\mathbf{y})$ are continuous on closed rectangles, then we have*

$$E(\nabla f(\mathbf{\Omega}_m(\mathbf{y}))) = \nabla \hat{\mathbf{x}}, \quad E(\nabla f(\mathbf{y})) = \nabla \hat{\mathbf{x}}. \tag{4.4}$$

With the Corollary 1, we could derive the main theorem of our B2G.

Theorem 1. *Let $\mathcal{J}$ be a fixed partition of the dimensions of the input $\mathbf{y} \in \mathbb{R}^d$, and let $J \in \mathcal{J}$ denote any subset of dimensions. Consider a family of subspace projections $\mathbf{\Omega}_m : \mathbb{R}^d \to \mathbb{R}^{d'}$, and assume the following conditions:*

- *(i) The estimator $f(\mathbf{\Omega}_m(\mathbf{y}))_J$ and $f(\mathbf{y})_J$ as unbiased estimators of the true target $\hat{\mathbf{x}}_J$ on each subspace J, i.e.,*
$$E(f(\mathbf{\Omega}_m(\mathbf{y}))_J) = E(f(\mathbf{y})_J) = \hat{\mathbf{x}}_J.$$
- *(ii) The function $f(\cdot)$ is assumed to be* ***$\mathcal{J}$-invariant****, meaning $f(\mathbf{y})_J$ only depends on $\mathbf{y}_J$, i.e., for all $\mathbf{y}$, $\mathbf{y}^{'}$ such that $\mathbf{y}_J = \mathbf{y}^{'}_J$, we have $f(\mathbf{y})_J = f(\mathbf{y}')_J$.*
- *(iii) The function $f(\cdot)$ is also continuous on a compact subset of $\mathbb{R}^d$.*
Then, the following decomposition holds:

$$\begin{aligned} &E_y\|d^c_{f_\theta}(\mathbf{y}; \mathbf{\Omega}_m)_J + \omega d^g_{f_\theta}(\mathbf{y}; \mathbf{\Omega}_m)_J\|^2_{\mathbf{\Omega}_m} \\ &- E_y\|\gamma d^g_{f_\theta}(\mathbf{y}; \mathbf{\Omega}_m)_J\|^2_{\mathbf{\Omega}_m} \\ =&E_{x,y}\|f_\theta(\mathbf{\Omega}_m(\mathbf{y}))_J - \hat{\mathbf{x}}_J\|^2_{\mathbf{\Omega}_m} + E_{x,y}\|\hat{\mathbf{x}}_J - \hat{\mathbf{y}}_J\|^2_{\mathbf{\Omega}_m}. \end{aligned} \tag{4.5}$$

Proof. The proof is derived from the application of the expectation-variance decomposition to the denoiser.

First, we use the identity:

$$E_y\|f_\theta(\mathbf{\Omega}_m(\mathbf{y}))_J - \mathbf{y}_J\|^2_{\mathbf{\Omega}_m} = E_{x,y}\|f_\theta(\mathbf{\Omega}_m(\mathbf{y}))_J - \mathbf{x}_J\|^2_{\mathbf{\Omega}_m} + E_{x,y}\|\mathbf{x}_J - \mathbf{y}_J\|^2_{\mathbf{\Omega}_m} \tag{4.6}$$

where $f(\cdot)_J$ is the optimal denoiser under the invariance and unbiasedness assumptions. The cross-term vanishes in expectation due to assumption (i). Then, using the linearity of expectation and the independence of residual terms, the expression reduces to the desired decomposition, separating the estimation error from the self-supervised loss.

□

Theorem 2. *Let f_θ denote a denoising function, and suppose that the output $f_\theta(\mathbf{\Omega}_m(\mathbf{y}))$ lies in the interior of a probability simplex $D_{\mathbf{x}^*}$. Define a game $D_{\mathbf{x}^*}(x)$ over strategy space V_{f_θ} where each dimension i corresponds to a player with payoff given by the log-likelihood* $\log f_\theta(\mathbf{\Omega}_m(\mathbf{y}))_i$. *Assume:*

- *(i) $D_{\mathbf{x}^*}(x)$ is a strictly stable population game.*
- *(ii) There exists a unique Nash equilibrium $\mathbf{x}^\star \in D_{\mathbf{x}^*}(x)$ representing the optimal denoising result.*
- *(iii) The dynamics of V_{f_θ} follow a potential game with concave potential function $\mathbf{d}_{\mathbf{\Omega}_m}(\cdot)$.*
- *(iiii) The evolution of player strategies is governed by the replicator dynamic system V_{f_θ}:*

$$\dot{x}_i = x_i\left(f_\theta(\Omega_m(\mathbf{y}))_i - \mathbf{x}^\top f_\theta(\Omega_m(\mathbf{y}))\right).$$

Then the optimal denoising output $\mathbf{x}^\star$ is the unique global asymptotically stable equilibrium of the dynamics V_{f_θ}, and any initial distribution $\mathbf{x} \in D_{\mathbf{x}^*}$ converges to $\mathbf{x}^\star$ under the dynamics.

Proof. Let $D_{\mathbf{x}^\star}(\mathbf{x}) = \sum_{i\in\Omega_m(\mathbf{y})} x_i^\star \log\frac{x_i^\star}{x_i}$ denote the Kullback-Leibler divergence from $\mathbf{x}$ to the equilibrium $\mathbf{x}^\star$. This is a convex Lyapunov function for the replicator dynamics V_{f_θ}. By Jensen's inequality and the concavity of the payoff (log-likelihood), we have $D_{\mathbf{x}^\star}(\mathbf{x}) \geq 0$, and equality only if $\mathbf{x} = \mathbf{x}^\star$. Along the replicator dynamic, the time derivative satisfies:

$$\frac{d}{dt}D_{\mathbf{x}^\star}(\mathbf{x}) = (\mathbf{x}^\star - \mathbf{x})^\top f_\theta(\Omega_m(\mathbf{y})) \leq 0,$$

by strict stability of the game.

Thus, $D_{\mathbf{x}^\star}(\mathbf{x})$ decreases over time and converges to zero. Therefore, $\mathbf{x}^\star$ is globally asymptotically stable. □

As a consequence, the main results of the Theorems 1 and 2 claim that the gradient-regularized loss of our self-supervised denoiser has an equivalence relationship with the loss of a supervised denoising network. Besides, Figure 5 also reveals that our loss function possesses convergence consistency.

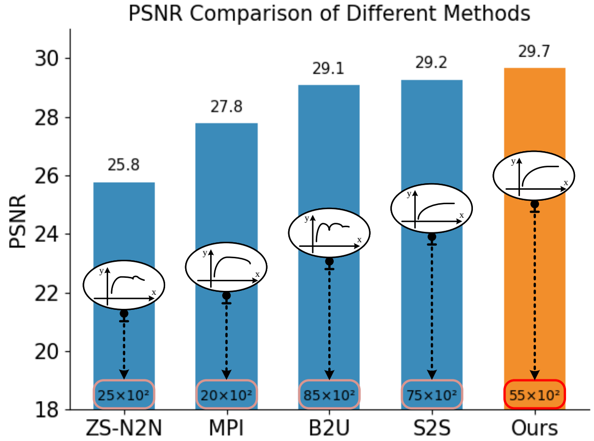


Figure 5. PSNR comparison and convergence stability of different methods. Bars represent the peak PSNR achieved, with the corresponding training iteration count shown below. The insets depict convergence trajectories during training; smoother curves show better stability.

5. Experimental results

In this section, we initially provide detailed descriptions of the experimental setups. Afterward, the method is evaluated by synthetic denoising and real-world denoising, with state-of-the-art methods including self-supervised denoisers with zero-shot or multiple images and a traditional training-free model. Furthermore, ablation studies are conducted to demonstrate the effectiveness of our proposed dual mask mechanism, gradient-regularized loss, $\mathcal{T} - standout$ model.

5.1. Experimental

5.1.1. Setup

We take advantage of the B2G architecture with the $1/2H \times 1/2W$ gradient convolution layer that is added third from the bottom of the network. Additionally, the kernel size for all the convolution layers is set to 3×3, and the corresponding stride and length of zero paddling are set as 2 throughout the whole experiment. The hyperparameter of each ReLU is set up to 0.1, and the probability of Bernoulli sampling is set up to 0.25. Specifically, our approach randomly selects 75% of pixel values to roughly estimate the remaining 25% in the whole noisy image. We utilize Adam optimizer with an initial learning rate of 2×10^{-5} for synthetic denoising experiments, and 10^{-5} for the real-world denoising experiments in different RGB spaces. Furthermore, we introduce a gradient convolutional layer of the size $1/2H \times 1/2W$ stacked with standout, calling it 40 times to get the final results during the training process. As for the gradient-regularized parameter ω and the penalty parameter γ used to balance the constraint of the $d^{c}_{f_\theta}(\mathbf{y}; \mathbf{\Omega})$, we set ω=0.8, γ=0.5 for the RGB synthetic denoising, and ω=0.5, γ=0.5 for the real-world denoising.

5.1.2. Datasets for denoising

We have evaluated our proposed B2G on the nine different datasets, including RGB synthetic images (i.e., Set9 [30], KODAK24 [40], Set14 [41], CBSD68 [42], BSD300 [42], Urban100 [43]), generated handwritten Chinese character images (H*à*nZi [32]), grey scale natural image

(BSD68 [42]), and physically captured 3D microscopy dataset (Planaria [44]). In addition, we validate our proposed B2G on images with different sizes and types, with different types and levels of noises, to demonstrate the effectiveness and robustness of the proposed framework. Moreover, the Planaria dataset measures the effectiveness of our approach from the perspective of biological cells. Furthermore, we consider four levels of noise for the synthetic image denoising: (1) Gaussian noise with σ=25, 50, 75, 100 for Set9, H$\grave{a}$nZi, and Urban100, (2) Gaussian noise with σ=25, 50 for KODAK24, CBSD68, Set14, and BSD300, (3) Gaussian noise with σ=25 and 50 for the BSD68 dataset, (4) Poisson noise with λ=30, 50 for the Planaria, and (5) the real-world denoising task the SIDD [45] dataset. Furthermore, we also conduct comparisons on the PolyU [46] and CC [47] datasets for the real-world denoising as detailed in the supplementary materials[*].

5.2. Synthetic Denoising

5.2.1. Gaussian Denoising

We also implement experiments for removing noise of zero-mean additive Gaussian. As indicated in Tables 2 and 3, our approach significantly outperforms the traditional denoising approach BM3D and all the benchmark self-supervised denoising approaches with zero-shot, including DIP, N2S, S2S, B2U, IDDP, ZS-N2N, and MPI. Although our proposed B2G can exhibit lower quantitative results than the approaches (such as Ne2Ne and N2N) on a few datasets, this can be mainly attributed to the fact that the Ne2Ne and N2N utilize more image content information through the use of image pairs, whereas our method only relies on single noisy images. Furthermore, while evaluating the overall quantitative effects, our method gains comparable performance.

Table 2. Quantitative comparisons of various approaches across different Gaussian noise levels on the Set9, Urban100, H$\grave{a}$nZi, and BSD68 datasets.

Datasets	σ	Traditional	Self-supervised learning with zero-shot								Multiple images learning	
		BM3D	DIP	N2S	S2S	B2U	IDDP	ZS-N2N	MPI	Ours	Ne2Ne	N2N
Set9	25	31.67/0.955	30.77/0.942	29.30/0.940	31.74/0.956	31.59/0.951	31.79/0.958	29.84/0.763	31.73/0.881	**32.01**/**0.965**	31.31/0.953	31.33/0.957
	50	28.95/0.922	28.33/0.910	27.25/0.904	29.25/0.928	29.07/0.918	29.24/0.925	25.76/0.614	27.77/0.754	**29.66**/**0.946**	29.22/0.927	28.94/0.929
	75	27.36/0.895	26.64/0.883	25.85/0.861	27.61/0.901	27.55/0.896	27.73/0.907	22.78/0.518	24.92/0.609	27.85/**0.918**	**27.86**/0.911	27.42/0.905
	100	26.04/0.868	25.41/0.858	24.67/0.848	26.27/0.877	26.23/0.874	26.41/0.885	20.46/0.446	22.74/0.482	**26.59**/**0.890**	26.55/0.881	26.45/0.876
Urban100	25	28.76/0.901	27.35/0.894	27.11/0.892	29.43/0.936	28.92/0.911	29.88/0.895	26.94/0.837	28.97/0.895	30.00/**0.952**	29.79/0.950	**30.66**/**0.952**
	50	25.48/0.876	24.88/0.856	24.72/0.853	26.22/0.884	26.12/0.835	26.57/0.895	22.86/0.698	23.51/0.711	**26.93**/**0.902**	26.07/0.891	25.09/0.869
	75	24.03/0.829	23.92/0.822	23.69/0.819	24.23/0.837	24.27/0.767	24.39/0.846	20.12/0.577	20.13/0.557	**24.69** /**0.848**	24.31/0.843	21.23/0.749
	100	22.36/0.759	22.29/0.757	22.18/0.745	22.57/0.768	22.97/0.711	22.80/0.783	18.12/0.478	17.75/0.444	**23.14**/**0.794**	23.04/0.790	20.32/0.662
H$\grave{a}$nZi	25	32.44/0.975	28.67/0.792	27.41/0.908	33.26/0.975	30.89/0.965	32.71/0.971	22.455/0.673	29.33/0.858	**35.28**/**0.984**	34.87/0.980	35.22/0.979
	50	27.78/0.943	20.91/0.598	23.82/0.861	28.45/0.968	29.89/0.942	29.90/0.970	18.99/0.625	21.03/0.674	**29.95**/**0.974**	29.91/0.970	22.14/0.659
	75	25.34/0.926	17.44/0.577	21.08/0.811	26.63/0.931	27.29/0.924	27.31/0.930	15.78/0.575	16.85/0.614	**27.54**/**0.939**	27.32/0.934	18.75/0.628
	100	23.86/0.899	15.02/0.497	18.77/0.783	25.02/0.920	25.38/0.913	25.41/0.925	13.26/0.515	14.08/0.560	**25.55**/**0.928**	25.37/0.923	15.60/0.573
BSD68	25	28.56/0.801	27.96/0.774	27.19/0.769	28.70/0.803	28.79/0.820	28.84/0.820	28.33/0.794	28.17/0.788	**28.89**/**0.823**	27.81/0.819	28.86/**0.823**
	50	25.62/0.687	25.04/0.645	24.53/0.642	25.92/0.699	25.90/0.697	26.07/0.709	25.58/0.681	25.44/0.667	**26.33**/**0.719**	25.68/0.703	25.77/0.700

[*] Detailed results are provided in the supplementary material.

Table 3. Quantitative comparison of PSNR/SSIM on the SIDD and DND datasets is conducted, wherein we present the official results as reported in the corresponding paper, which can also be cross-verified from benchmark websites. Hereinafter, red and blue colors are employed to denote the best and second-best outcomes among the self-supervised methods.

Datesets	σ	Traditional	Self-Supervised learning with zero-shot								Multiple images learning	
		BM3D	DIP	N2S	S2S	B2U	IDDP	ZS-N2N	MPI	Ours	Ne2Ne	N2N
KODAK24	25	31.87/0.868	27.20/0.751	30.96/0.843	31.28/0.864	32.27/0.880	32.09/0.871	29.17/0.773	30.28/0.846	32.46/**0.885**	32.08/0.879	**32.48**/0.884
	50	29.01/0.794	26.83/0.703	28.09/0.781	29.05/0.792	28.96/0.791	29.11/0.800	24.94/0.610	26.68/0.713	**29.36/0.813**	28.99/0.795	27.88/0.788
CBSD68	25	30.71/0.837	29.89/0.808	29.71/0.800	30.65/0.825	30.81/0.841	30.90/0.844	28.54/0.799	26.68/0.713	31.09/0.857	30.94/0.852	**31.22/0.858**
	50	27.60/0.779	25.94/0.769	26.58/0.774	27.56/0.778	27.64/0.780	27.67/0.782	24.39/0.644	26.46/0.699	**27.85/0.790**	27.69/0.782	26.51/0.775
Set14	25	30.88/0.854	27.16/0.802	29.80/0.815	30.08/0.839	31.27/0.864	30.99/0.860	28.26/0.770	30.28/0.846	**31.39/0.869**	31.09/0.864	31.37/0.868
	50	27.85/0.771	24.42/0.722	26.61/0.757	27.63/0.762	27.91/0.779	28.00/0.781	24.19/0.625	26.68/0.713	**28.36/0.796**	28.13/0.789	26.49/0.753
BSD300	25	30.48/0.861	26.38/0.708	29.67/0.826	29.89/0.849	30.87/0.872	30.92/0.874	28.53/0.798	30.79/0.874	**31.09/0.879**	30.79/0.873	31.07 / 0.878
	50	27.81/0.780	24.18/0.673	27.74/0.778	27.80/0.780	27.85/0.781	27.83/0.779	24.40/0.638	27.43/0.725	**27.98/0.784**	27.92/**0.784**	26.79/0.775

In Figure 6, we plotted the minimum-to-maximum range of the quantitative indicators for ZS-N2N, Ne2Ne, N2N, and our approach with various noise levels. From the results, we can observe that the approach maintains a stable state across different noise levels and different noisy images. Our approach shows commendably even under high noise levels, demonstrating the effectiveness of our proposed gradient-regularized loss.

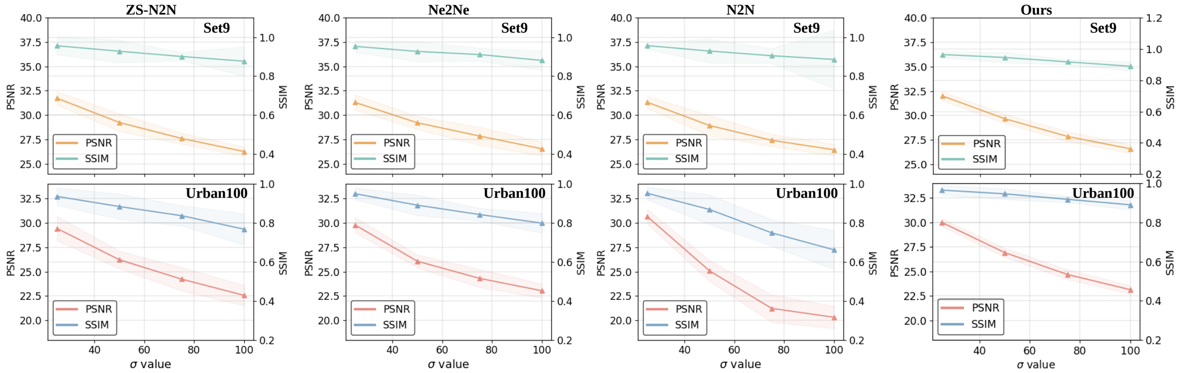
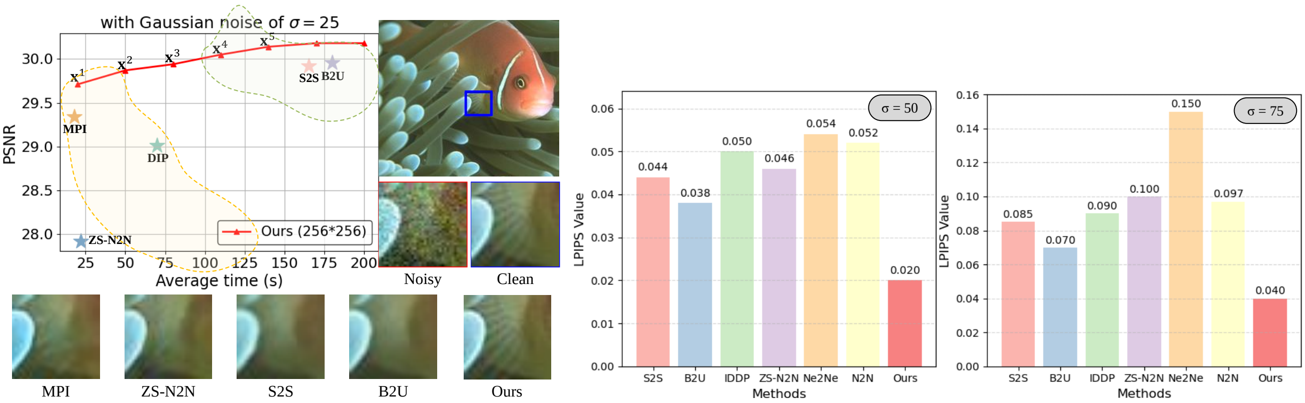


Figure 6. Performance of the proposed method compared with three methods on the Urban100 or Set9 dataset. The shaded region of the polyline is determined by the minimum and maximum values of PSNR/SSIM at each noise level point.

Due to the implementation of our B2G strategy, the average denoising time for a single image might surpass that of other state-of-the-art approaches. We provide a comparison in Figure 7 (a) on the time consumption with other self-supervised denoisers in zero-shot, i.e., DIP [30], S2S [32], B2U [36], ZS-N2N [35], and MPI [37]. In the curve graph provided in Figure 7 (a), existing methods can be roughly divided into two groups, namely, DIP, ZS-N2N, and MPI belonging to the first group, and S2S and B2U falling under the second group. The former approaches need a shorter time but yield relatively lower PSNR values, while the latter ones demand more time for achieving much better results. We also provide the PSNR values and corresponding time for each 3000 iteration in our B2G strategy. It is apparent that employing only 6000 iterations can yield PSNR values comparable to those obtained by other approaches, while also requiring almost the least computational time.

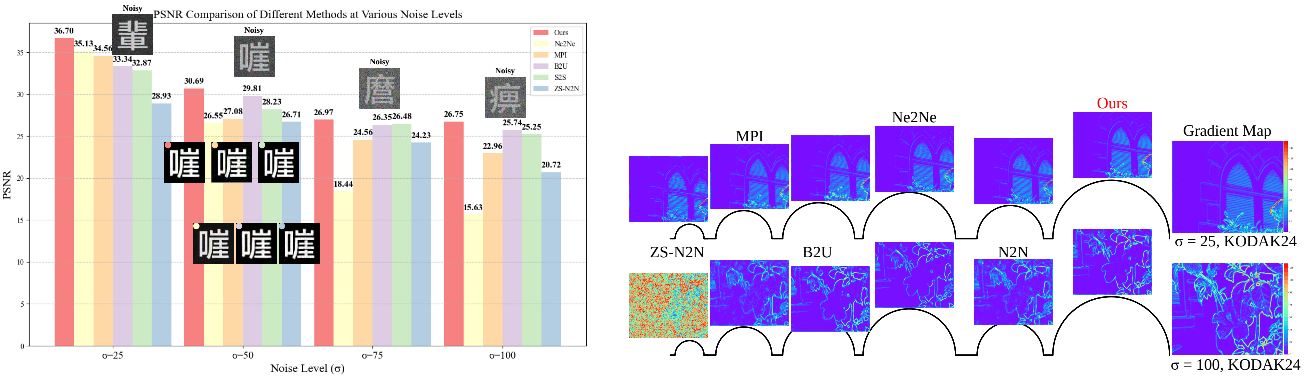
Additionally, in comparison to a specific existing method, we establish a fixed iteration to generate much better results within a comparable time (as circled by shadow areas). Thus, the time consumption problem associated with our proposed approach is no need to be concerned.



(a)

(b)

Figure 7. Given the image (with Gaussian noise, $\sigma = 25$), a comparison is made of the average time needed for denoising a single image with dimensions of 256×256 with other state-of-the-art approaches as shown in (a); (b) Comparison of diverse methods in terms of the Learned Perceptual Image Patch Similarity (LPIPS) indicator.



(a)

(b)

Figure 8. (a) PSNR comparison of different denoising methods across various noise levels ($\sigma = 25, 50, 75, 100$). For $\sigma = 50$, we provide visual examples of the noisy and denoised results to demonstrate the effectiveness of each method. The noisy input images are displayed at the top of each column, while the corresponding denoised outputs value are shown below; (b) Visual comparisons of the denoised by the compared models sRGB images with respect to image gradients that show edge retention even as noise is suppressed (Gaussian noise with σ=25/100). The "colorbar" located on the extreme right indicates the mapping between gradient values and corresponding colors.

Furthermore, Figures 8–10 also provide detailed comparisons of the denoising results of compared approaches at different noise levels. B2G exhibits outstanding capability of image detail restoration as shown in Figure 8 (b) surpassing the performance of other evaluated approaches under high noise conditions. The fluctuation of the curve in the Figure 8 (b) can further demonstrate that the method we proposed exhibits the greatest fluctuation (that is, it is most similar to the gradient map of the clean image). Compared to denoising methods with zero-shot, our B2G demonstrates a relatively stronger ability to restore lost details caused by the randomness of masks. In addition, we also introduce a novel indicator, LPIPS, to assess the similarity of human eye cognition, as shown in Figure 7 (b). As is evident from Figure 7 (b), regardless of whether the noise level is 50 or 75, our LPIPS indicator remains the lowest. From this viewpoint, it demonstrates that our proposed B2G approach holds an absolute superiority in human eye cognition.

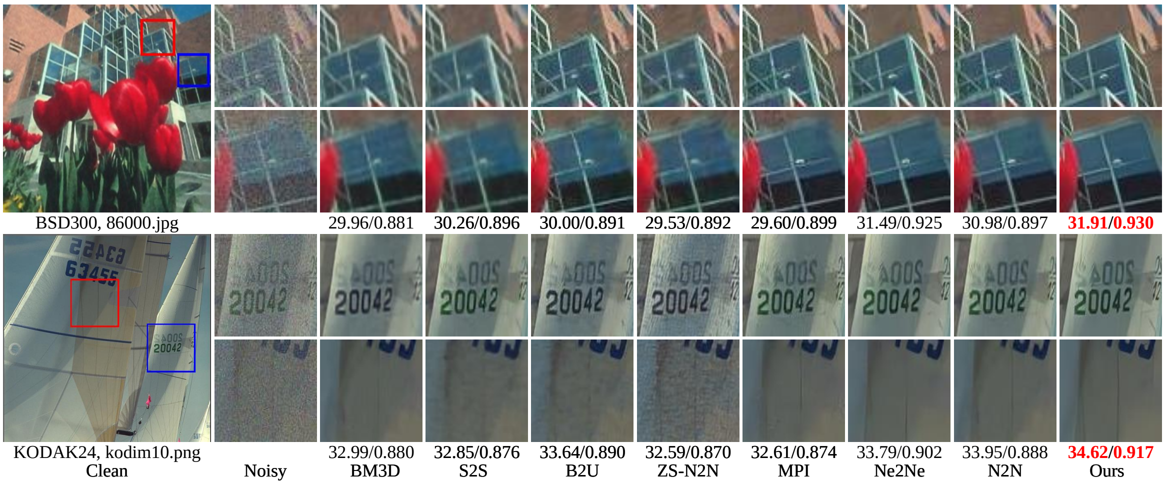


Figure 9. Visual comparisons of denoising results (PSNR/SSIM) produced by different approaches on provided images from BSD300 and KODAK24 corrupted by Gaussian noise (σ=25).

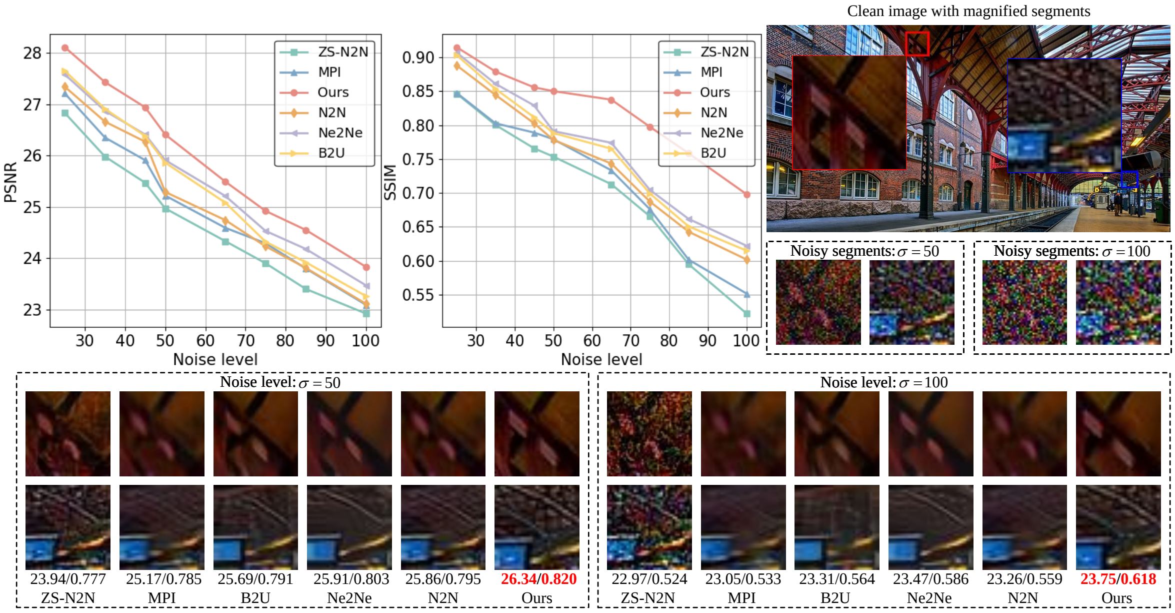


Figure 10. We selected an image from Urban100 (img054.png) and the visual and quantitative comparison results of sRGB cropped images with Gaussian noise (σ=50, 100) to facilitate comparison. The curves on the left display PSNR/SSIM for the Urban100 across various mask strategies under different Gaussian noise levels.

5.2.2. Poisson denoising

We then premeditate Poisson noise, which can be utilized to model photon noise in most imaging sensors. Similarly, the Poisson noise is also in accordance with the zero-mean assumption adopted in this paper. We added the Poisson noise with determined levels to the ground truth for the Planaria dataset, and the results are presented in Table 4. Although our SSIM metric is the same as the N2N, our PSNR is 0.29 dB higher. Furthermore, we achieved excellent visual results in terms of recovering details, as shown in Figure 11 for $\lambda = 30$ and $\lambda = 50$. This demonstrates the superiority of B2G for removing Poisson noise.

Table 4. Quantitative comparisons (PSNR/SSIM) of various approaches for different Poisson noise levels on the Planaria dataset.

λ	Traditional	Self-Supervised learning with zero-shot								Multiple images learning	
	BM3D	DIP	N2S	S2S	IDDP	B2U	ZS-N2N	MPI	Ours	Ne2Ne	N2N
30	34.64/0.924	34.22/0.918	34.59/0.922	34.98/0.924	34.77/0.920	35.07/0.925	34.48/0.915	34.86/0.921	**35.26/0.930**	35.01/0.923	34.97/**0.930**
50	33.98/0.908	33.47/0.899	33.83/0.905	34.25/0.908	34.01/0.900	34.28/0.910	33.66/0.900	33.91/0.905	**34.32/0.912**	34.29/0.911	34.16/0.909

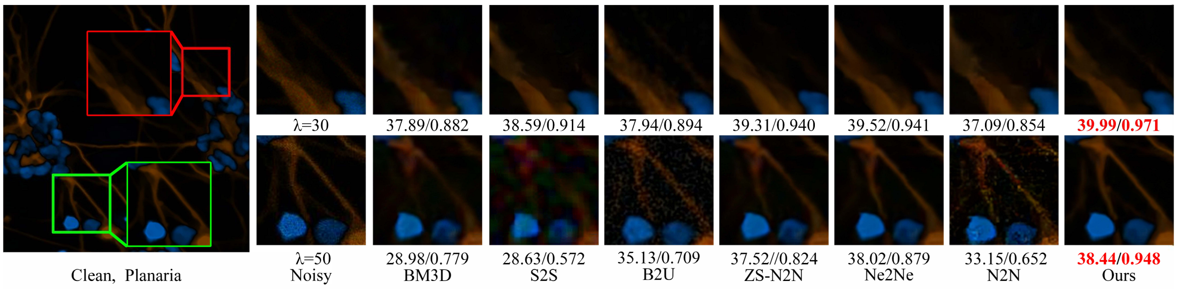


Figure 11. For denoising the provided image (Planaria, error.png), the visual and quantitative comparison results (PSNR/SSIM) of sRGB cropped images with Poisson noise ($\lambda = 30/50$) to facilitate comparison.

5.3. Real-world denoising

Confronting real-world denoising, we introduced a dilated convolution with the aim of disrupting the spatial correlation of real-world noise, and the network architecture is provided in the supplementary material. Table 5 indicates the quality scores for quantitative comparisons on the real-world dataset (SIDD [45]). Besides, we add comparisons with supervised approaches (named as plug-and-play (PnP) [25] and deep generalized unfolding network (DGUNet$^+$) [49]), where the denoising network parameters of DGUNet$^+$ in all stages vary and it is more appropriate for handling real-world noise. Although our approach is disadvantaged in comparison with DGUNet$^+$ on sRGB images, it achieves superior results compared to the state-of-the-art MASH (i.e., excels over other zero-shot approaches), both on sRGB and raw-RGB images. For the AP-BSN [50] and LG-BPN [51] approaches, we utilize these methods from [52] to directly denoise a single image. Table 5 comprehensively indicates the quality scores of quantitative comparisons on real-world denoising. In addition, the quality comparison provided in Figure 12 shows our advantage in effective denoising in real-world scenarios. Thus, our approach is adept at real-world denoising, offering a robust solution for image quality enhancement.

Table 5. Quantitative comparisons (PSNR/SSIM) on SIDD for real-world image denoising.

Category	Methods	raw-RGB		sRGB
		Benchmark	Validation	Validation
Non-learning	BM3D	48.60/0.986	48.92/0.986	34.21/0.821
Supervised learning	PnP	50.60/0.991	51.19/0.991	34.56/0.847
	DGUNet$^+$	50.78/0.992	51.23/0.992	**39.91/0.960**
Self-supervised with masked images	HQSSL	49.82/0.989	50.44/0.990	33.71/0.808
	S2S	46.48/0.973	46.54/0.975	30.84/0.726
	B2U	50.79/0.991	51.36/0.992	34.27/0.832
	AP-BSN	-	-	33.03/0.841
	ZS-N2N	49.20/0.983	49.22/0.983	25.61/0.559
	LG-BPN	-	-	34.74/0.843
	MPI	50.51/0.993	50.52/0.993	34.43/0.844
	MASH	52.15/0.992	52.17/0.993	35.06/0.851
	Ours	**52.36/0.994**	**52.37/0.995**	35.10/0.853
Self-supervised with entire images	Ne2Ne	50.47/0.990	51.06/0.991	34.33/0.838
	N2N	50.62/0.991	51.21/0.991	34.20/0.833

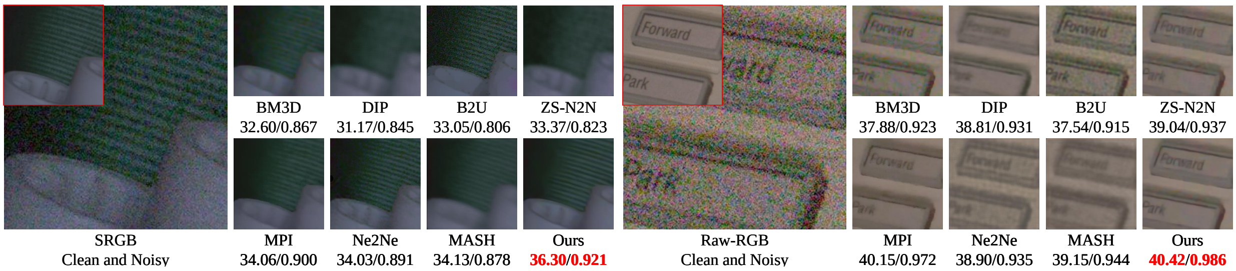


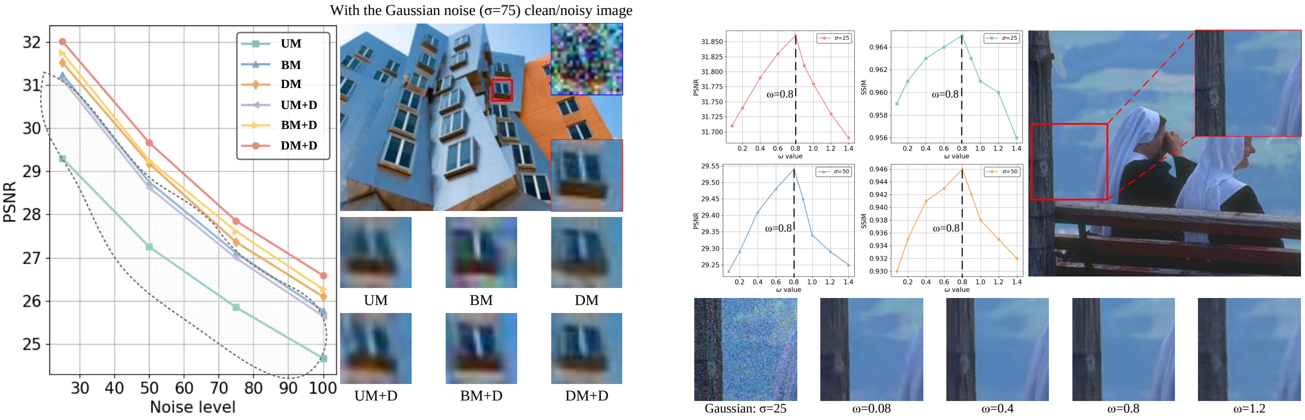
Figure 12. Visual comparisons (PSNR/SSIM) on SIDD the validation.

5.4. *Ablation studies*

In this section, we conduct comprehensive ablation studies on using dual mask mechanism with $\mathcal{T}$-standout, the penalty parameter γ, and the gradient-regularized parameter ω, on the Set9 dataset under the metrics of PSNR/SSIM.

5.4.1. Different mask strategies

To assess the efficacy of the dual mask mechanism (DM) under different noise levels, we contrast it with the uniform pixel selection method (UM) and Bernoulli sampling (BM) advocated in N2S [32] and S2S [33]. In addition, we validate the adaptability of the $\mathcal{T}$-standout module when applied to different masking mechanisms. Here, we explicitly define the D:= $\mathcal{T}$-standout module, which is used to enhance feature selection by preserving informative neurons. Figure 13 (a) depicts the performance of employing different mask strategies while maintaining detail loss. From the Figure 13 (a), it can be observed that our proposed dual mask mechanism "DM" significantly surpasses "BM" and "UM" in all four noise patterns. The "DM+D" (ours) exhibits superior performance when combined with $\mathcal{T}$-standout, and "UM+D" (i.e., the gray curve area in Figure 13 (a)) is severely restrained. These visual results on the right side of Figure 13(a) further illustrate the detailed preservation advantages brought by the "DM", especially when integrated with $\mathcal{T}$-Standout.



(a) **(b)**

Figure 13. (a) Ablation experiments on different mask strategies. The curves on the left display PSNR for Set9 across various mask strategies under different Gaussian noise levels. Besides, visual effects of denoised results for an example image (Urban100, img075.png) are also presented; (b) Effect of the gradient-regularized parameter ω. Upper left: given different values of the performance (PSNR/SSIM) of B2G varies over Gaussian noise levels (σ=25, 50) on the Set9. Below: visual effects of denoised results for an example image (BSD300, 126039.jpg) are obtained by varying ω.

5.4.2. Parameter ω

The gradient-regularized parameter ω is used to make up the common shortcoming of information loss caused by the masking schemes. As depicted in Figure 13 (b), with the increase of ω, the trend of PSNR shows an initial increase followed by the decrease after reaching the maximum value. Therefore, we set $\omega = 0.8$ as our experimental configuration in this research. In addition, it seems the gradient-regularized parameter ω is more sensitive than the adjustable parameter γ. Nevetheless, with respect to the two cases with Gaussian noise, for example, setting the ω within the interval of $[0.2, 1.0]$ is acceptable for superior performance than all the other BSD networks, which indicates the effectiveness of our gradient-regularized loss.

5.4.3. Parameter ϵ_s, ϵ_g

To assess the robustness and sensitivity of the dual mask mechanism and $\mathcal{T}$-standout module, an ablation study is carried out by modulating the threshold parameters ϵ_s and ϵ_g. These parameters respectively govern the selection of neurons with high detail pixel significance and high gradient significance. Multiple values of ϵ_s and ϵ_g are tested, where $\epsilon_s, \epsilon_g \in \{15, 25, 40, 55, 70, 85, 100\}$.

As depicted in Figure 14, the results clearly illustrate that when $\epsilon_s = 25$, the optimal denoising performance in terms of PSNR and SSIM is attained. This value effectively strikes a balance between preserving crucial detailed information and preventing redundancy. In addition, when $\epsilon_g = 40$, it yields the optimal performance by safeguarding the most informative gradient-aware features while avoiding noise amplification. These findings not only validate the robustness of our design decisions but also provide valuable guidance for the practical selection of threshold values.

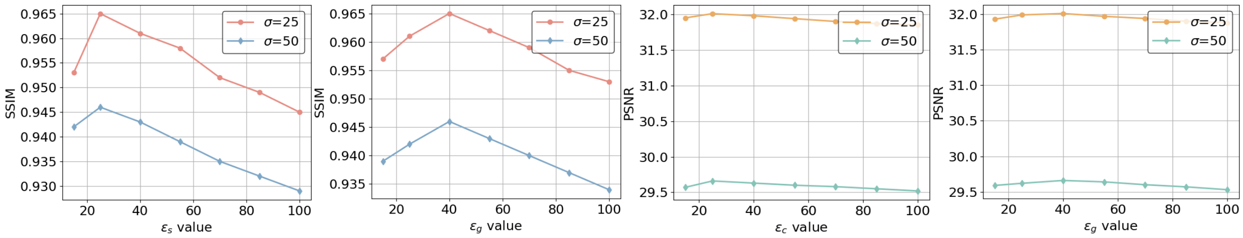


Figure 14. Sensitivity analysis on parameter ϵ_s and ϵ_g.

5.4.4. Parameter γ

The flexible penalty parameter γ is used to supplement background content information. In Figure 15, it is verified that the denoising performance exhibits a trend of initially increasing and subsequently decreasing with respect to different noise levels. It is noticed that when γ=0.5, PSNR reaches the optimal score, which more strongly confirms the role of the regularization term in constraining the overall loss function and enhancing the numerical efficacy. Thus, we set γ=0.5 throughout the experiments in this paper.

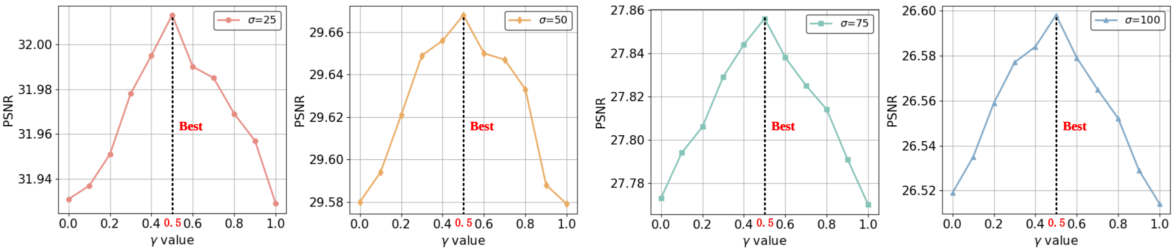


Figure 15. Ablation studies on the penalty parameter γ. The plots show that the PSNR performance of B2G as γ is varied for two different noise types on the Set9.

6. Conclusions

We present B2G, a novel self-supervised denoising framework designed to enhance edge restoration using dual mask denoisers. The framework employs a gradient-regularized loss function in conjunction with $\mathcal{T}$-standout to train the denoising network in a zero-shot manner. Simultaneously, the framework continuously updates the dual mask mechanism within an iterative loop, guaranteeing that valuable details are maintained throughout the denoising process. B2G employs a zero-shot way to offer a denoising framework that saves more training data. We validate the performance of our B2G across various noise-related tasks and demonstrate its robust capability in processing images with high noise levels. Therefore, our proposed B2G provides an efficient solution for image denoising, particularly in applications involving images with significant noise.

Author contributions

Bolin Song: Conceptualization, Methodology, Writing original draft, Resources; Zhenhao Shuai: Formal analysis, Supervision, Investigation; Yuanyuan Si: Data curation, Software, Project administration; Ke Li: Funding acquisition, Visualization, Writing review & editing.

Use of Generative-AI tools declaration

The authors declare that they have not used Artificial Intelligence (Al) tools in the creation of this article.

Acknowledgments

This research is supported in part by the National Natural Science Foundation of China (Nos. 62176036, 62102059), the Natural Science Foundation of Liaoning Province (No. 2023-MS-126), and the Fundamental Research Funds for the Central Universities (No. 3132023523).

Conflict of interest

All authors declare no conflicts of interest in this paper.

References

1. L. Lu, T. Deng, A method of self-supervised denoising and classification for sensor-based human activity recognition, *IEEE Sens. J.*, **23** (2023), 27997–28011. https://doi.org/10.1109/JSEN.2023.3323314
2. W. Dong, H. Wang, F. Wu, G. Shi, X. Li, Deep spatial-spectral representation learning for hyperspectral image denoising, *IEEE T. Comput. Imag.*, **5** (2019), 635–648. https://doi.org/10.1109/TCI.2019.2911881
3. L. Zhuang, M. K. Ng, L. Gao, Z. Wang, Eigen-CNN: Eigenimages plus eigennoise level maps guided network for hyperspectral image denoising, *IEEE T. Geosci. Remote*, **62** (2024), 5512018. https://doi.org/10.1109/TGRS.2024.3379199
4. Q. Jiang, H. Shi, S. Gao, J. Zhang, K. Yang, L. Sun, et al., Computational imaging for machine perception: Transferring semantic segmentation beyond aberrations, *IEEE T. Comput. Imag.*, **10** (2024), 535–548. https://doi.org/10.1109/TCI.2024.3380363
5. F. Xiao, R. Liu, Y. Zhu, H. Zhang, J. Zhang, S. Chen, A dense multicross self-attention and adaptive gated perceptual unit method for few-shot semantic segmentation, *IEEE T. Artif. Intell.*, **5** (2024), 2493–2504. https://doi.org/10.1109/TAI.2024.3369553
6. Y. Jin, S. Kwak, S. Ham, J. Kim, A fast and efficient numerical algorithm for image segmentation and denoising, *AIMS Mathematics*, **9** (2024), 5015–5027. https://doi.org/10.3934/math.2024243
7. J. Liao, C. He, J. Li, J. Sun, S. Zhang, X. Zhang, Classifier-guided neural blind deconvolution: A physics-informed denoising module for bearing fault diagnosis under noisy conditions, *Mech. Syst. Signal Pr.*, **222** (2025), 111750. https://doi.org/10.1016/j.ymssp.2024.111750

8. Y. Li, H. Chang, Y. Duan, CurvPnP: Plug-and-play blind image restoration with deep curvature denoiser, *Signal Process.*, **233** (2025), 109951. https://doi.org/10.1016/j.sigpro.2025.109951

9. A. C. Bovik, T. S. Huang, D. C. Munson, The effect of median filtering on edge estimation and detection, *IEEE T. Pattern Anal.*, **PAMI-9** (1987), 181–194. https://doi.org/10.1109/TPAMI.1987.4767894

10. L. K. Choi, J. You, A. C. Bovik, Referenceless prediction of perceptual fog density and perceptual image defogging, *IEEE T. Image Process.*, **24** (2015), 3888–3901. https://doi.org/10.1109/TIP.2015.2456502

11. H. Feng, L. Wang, Y. Wang, F. Han, H. Huang, Learnability enhancement for low-light raw image denoising: A data perspective, *IEEE T. Pattern Anal.*, **46** (2024), 370–387. https://doi.org/10.1109/TPAMI.2023.3301502

12. Y. Luo, B. You, G. Yue, J. Ling, Pseudo-supervised low-light image enhancement with mutual learning, *IEEE T. Circ. Syst. Vid.*, **34** (2024), 85–96. https://doi.org/10.1109/TCSVT.2023.3284856

13. Y. Zhao, Q. Zheng, P. Zhu, X. Zhang, M. Wenpeng, TUFusion: A transformer-based universal fusion algorithm for multimodal images, *IEEE T. Circ. Syst. Vid.*, **34** (2024), 1712–1725. https://doi.org/10.1109/TCSVT.2023.3296745

14. M. Kang, M. Jung, Nonconvex fractional order total variation based image denoising model under mixed stripe and Gaussian noise, *AIMS Mathematics*, **9** (2024), 21094–21124. https://doi.org/10.3934/math.20241025

15. J. Yin, X. Zhuang, W. Sui, Y. Sheng. J. Wang, R. Song, et al., A bearing signal adaptive denoising technique based on manifold learning and genetic algorithm, *IEEE Sens. J.*, **24** (2024), 20758–20768. https://doi.org/10.1109/JSEN.2024.3403845

16. X. Chen, L. Zhao, J. Xu, Z. Dai, L. Xu, N. Guo, et al., Feature-adaptive self-supervised leaning for microthrust measurement signals blind denoising, *IEEE Sens. J.*, **25** (2025), 1015–1028. https://doi.org/10.1109/JSEN.2024.3493104

17. S. Zhang, Y. Yang, Q. Qin, L. Feng, L. Jiao, Rapid blind denoising method for grating fringe images based on noise level estimation, *IEEE Sens. J.*, **21** (2021), 8150–8160. https://doi.org/10.1109/JSEN.2021.3050237

18. W. Wang, F. Wen, Z. Yan, P. Liu, Optimal transport for unsupervised denoising learning, *IEEE T. Pattern Anal.*, **45** (2023), 2104–2118. https://doi.org/10.1109/TPAMI.2022.3170155

19. Y. Hou, J. Xu, M. Liu, G. Liu, L. Liu, F. Zhu, et al., NLH: A blind pixel-level non-local method for real-world image denoising, *IEEE T. Image Process.*, **29** (2020), 5121–5135. https://doi.org/10.1109/TIP.2020.2980116

20. D. Gilton, G. Ongie, R. Willett, Deep equilibrium architectures for inverse problems in imaging, *IEEE T. Comput. Imag.*, **7** (2021), 1123–1133. https://doi.org/10.1109/TCI.2021.3118944

21. B. Jiang, Y. Lu, J. Wang, G. Lu, D. Zhang, Deep image denoising with adaptive priors, *IEEE T. Circ. Syst. Vid.*, **32** (2022), 5124–5136. https://doi.org/10.1109/TCSVT.2022.3149518

22. W. Wang, D. Yin, C. Fang, Q. Zhang, Q. Li, A novel multi-image stripe noise removal method based on wavelet and recurrent networks, *IEEE Sens. J.*, **24** (2024), 26058–26069. https://doi.org/10.1109/JSEN.2024.3421337

23. S. H. Chan, X. Wang, O. A. Elgendy, Plug-and-play ADMM for image restoration: Fixed-point convergence and applications, *IEEE T. Comput. Imag.*, **3** (2017), 84–98. https://doi.org/10.1109/TCI.2016.2629286

24. Y. Sun, B. Wohlberg, U. S. Kamilov, An online plug-and-play algorithm for regularized image reconstruction, *IEEE T. Comput. Imag.*, **5** (2019), 395–408. https://doi.org/10.1109/TCI.2019.2893568

25. K. Zhang, Y. Li, W. Zuo, L. Zhang, L. Van Gool, R. Timofte, Plug-and-Play Image Restoration With Deep Denoiser Prior, *IEEE T. Pattern Anal.*, **44** (2021), 6360–6376. https://doi.org/10.1109/TPAMI.2021.3088914

26. J. Lehtinen, J. Munkberg, J. Hasselgren, S. Laine, T. Karras, M. Aittala, et al., Noise2Noise: Learning image restoration without clean data, 2018, arXiv: 1803.04189 https://doi.org/10.48550/arXiv.1803.04189

27. T. Huang, S. Li, X. Jia, H. Lu, J. Liu, Neighbor2Neighbor: Self-supervised denoising from single noisy images, In: *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2021, 14776–14785. https://doi.org/10.1109/CVPR46437.2021.01454

28. J. Xu, Y. Huang, M. Cheng, L. Liu, F. Zhu, Z. Xu, et al., Noisy-as-clean: Learning self-supervised denoising from corrupted image, *IEEE T. Image Process.*, **29** (2020), 9316–9329. https://doi.org/10.1109/TIP.2020.3026622

29. N. Moran, D. Schmidt, Y. Zhong, P. Coady, Noisier2Noise: Learning to denoise from unpaired noisy data, In: *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020, 12061–12069. https://doi.org/10.1109/CVPR42600.2020.01208

30. V. Lempitsky, A. Vedaldi, D. Ulyanov, Deep image prior, In: *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018, 9446–9454. https://doi.org/10.1109/CVPR.2018.00984

31. A. Krull, T. Buchholz, F. Jug, Noise2Void-Learning denoising from single noisy images, In: *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019, 2124–2132. https://doi.org/10.1109/CVPR.2019.00223

32. J. Batson, L. Royer, Noise2Self: Blind denoising by self-supervision, 2019, arXiv: 1901.11365. https://doi.org/10.48550/arXiv.1901.11365

33. Y. Quan, M. Chen, T. Pang, H. Ji, Self2Self with dropout: Learning self-supervised denoising from single image, In: *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020, 1887–1895. https://doi.org/10.1109/CVPR42600.2020.00196

34. J. Lequyer, R. Philip, A. Sharma, W. Hsu, L. Pelletier, A fast blind zero-shot denoiser, *Nat. Mach. Intell.*, **4** (2022), 953–963. https://doi.org/10.1038/s42256-022-00547-8

35. Y. Mansour, R. Heckel, Zero-Shot Noise2Noise: Efficient image denoising without any data, In: *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2023, 14018–14027. https://doi.org/10.1109/CVPR52729.2023.01347

36. Z. Wang, J. Liu, G. Li, H. Han, Blind2Unblind: Self-supervised image denoising with visible blind spots, In: *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022, 2017–2026. https://doi.org/10.1109/CVPR52688.2022.00207

37. X. Ma, Z. Wei, Y. Jin, P. Ling, T. Liu, B. Wang, et al., Masked pre-training enables universal zero-shot denoiser, 2024, arXiv: 2401.14966. https://doi.org/10.48550/arXiv.2401.14966

38. O. Ghahabi, J. Hernando, Restricted Boltzmann machines for vector representation of speech in speaker recognition, *Comput. Speech Lang.*, **47** (2018), 16–29. https://doi.org/10.1016/j.csl.2017.06.007

39. Y. Xie, Z. Wang, S. Ji, Noise2Same: Optimizing a self-supervised bound for image denoising, 2020, arXiv: 2010.11971. https://doi.org/10.48550/arXiv.2010.11971

40. C. R. Hauf, J. S. Houchin, The FlashPix™ image file format, In: *Proc. IS&T 4th Color and Imaging Conf.*, **4** (1996), 234–234. https://doi.org/10.2352/CIC.1996.4.1.art00060

41. R. Zeyde, M. Elad, M. Protter, On single image scale-up using sparse-representations, In: *Curves and surfaces*, Berlin, Heidelberg: Springer, 2012, 711–730. https://doi.org/10.1007/978-3-642-27413-8_47

42. D. Martin, C. Fowlkes, D. Tal, J. 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*, 2001, 416–423. https://doi.org/10.1109/ICCV.2001.937655

43. J. Huang, A. Singh, N. Ahuja, Single image super-resolution from transformed self-exemplars, In: *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015, 5197–5206. https://doi.org/10.1109/cvpr.2015.7299156

44. M. Weigert, U. Schmidt, T. Boothe, A. Müller, A. Dibrov, A. Jain, et al., Content-aware image restoration: Pushing the limits of fluorescence microscopy, *Nat. Methods*, **15** (2018), 1090–1097. https://doi.org/10.1038/s41592-018-0216-7

45. A. Abdelhamed, S. Lin, M. S. Brown, A high-quality denoising dataset for smartphone cameras, In: *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018, 1692–1700. https://doi.org/10.1109/CVPR.2018.00182

46. J. Xu, H. Li, Z. Liang, D. Zhang, L. Zhang, Real-world Noisy Image Denoising: A New Benchmark, 2018, arXiv: 1804.02603. https://doi.org/10.48550/arXiv.1804.02603

47. S. Nam, Y. Hwang, Y. Matsushita, S. J. Kim, A holistic approach to cross-channel image noise modeling and its application to image denoising, In: *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, 1683–1691. https://doi.org/10.1109/CVPR.2016.186

48. K. Dabov, A. Foi, V. Katkovnik, K. Egiazarian, Image denoising by sparse 3-D transform-domain collaborative filtering, *IEEE T. Image Process.*, **16** (2007), 2080–2095. https://doi.org/10.1109/TIP.2007.901238

49. C. Mou, Q. Wang, J. Zhang, Deep generalized unfolding networks for image restoration, In: *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022, 17378–17389. https://doi.org/10.1109/CVPR52688.2022.01688

50. W. Lee, S. Son, K. M. Lee, AP-BSN: Self-supervised denoising for real-world images via asymmetric PD and blind-spot network, In: *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022, 17704–17713. https://doi.org/10.1109/CVPR52688.2022.01720

51. Z. Wang, Y. Fu, J. Liu, Y. Zhang, LG-BPN: Local and global blind-patch network for self-supervised real-world denoising, In: *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2023, 18156–18165. https://doi.org/10.1109/CVPR52729.2023.01741

52. H. Chihaoui, P. Favaro, Masked and shuffled blind spot denoising for real-world images, In: *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2024, 3025–3034. https://doi.org/10.1109/CVPR52733.2024.00292

AIMS Press

© 2025 the Author(s), licensee AIMS Press. This is an open access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0)