applied sciences

MDPI

Article

A Dual-Branch Self-Boosting Network Based on Noise2Noise for Unsupervised Image Denoising

Yuhang Geng, Shaoping Xu *, Minghai Xiong, Qiyu Chen and Changfei Zhou

School of Mathematics and Computer Sciences, Nanchang University, Nanchang 330031, China; 6109121105@email.ncu.edu.cn (Y.G.); 419100220136@email.ncu.edu.cn (M.X.); 419100230159@email.ncu.edu.cn (Q.C.); 409100220054@email.ncu.edu.cn (C.Z.)
* Correspondence: xushaoping@ncu.edu.cn

Abstract: While unsupervised denoising models have shown progress in recent years, their noise reduction capabilities still lag behind those of supervised denoising models. This limitation can be attributed to the lack of effective constraints during training, which only utilizes noisy images and hinders further performance improvements In this work, we propose a novel dual-branch self-boosting network called DBSNet, which offers a straightforward and effective approach to image denoising. By leveraging task-dependent features, we exploit the intrinsic relationships between the two branches to enhance the effectiveness of our proposed model. Initially, we extend the classic Noise2Noise (N2N) architecture by adding a new branch for noise component prediction to the existing single-branch network designed for content prediction. This expansion creates a dual-branch structure, enabling us to simultaneously decompose a given noisy image into its content (clean) and noise components. This enhancement allows us to establish stronger constraint conditions and construct more powerful loss functions to guide the training process. Furthermore, we replace the UNet structure in the N2N network with the proven DnCNN (Denoising Convolutional Neural Network) sequential network architecture, which enhances the nonlinear mapping capabilities of the DBSNet. This modification enables our dual-branch network to effectively map a noisy image to its content (clean) and noise components simultaneously. To further improve the stability and effectiveness of training, and consequently enhance the denoising performance, we introduce a feedback mechanism where the network's outputs, i.e., content and noise components, are fed back into the dual-branch network. This results in an enhanced loss function that ensures our model possesses excellent decomposition ability and further boosts the denoising performance. Extensive experiments conducted on both synthetic and real-world images demonstrate that the proposed DBSNet outperforms the unsupervised N2N denoising model as well as mainstream supervised models trained with supervised methods. Moreover, the evaluation results on real-world noisy images highlight the desirable generalization ability of DBSNet for practical denoising applications.

Keywords: image denoising; unsupervised training; training constraints; dual-branch network; feedback mechanism; denoising performance

Citation: Geng, Y.; Xu, S.; Xiong, M.; Chen, Q.; Zhou, C. A Dual-Branch Self-Boosting Network Based on Noise2Noise for Unsupervised Image Denoising. *Appl. Sci.* **2024**, *14*, 4735. https://doi.org/10.3390/app14114735

Academic Editor: Andrea Prati

Received: 1 May 2024
Revised: 24 May 2024
Accepted: 28 May 2024
Published: 30 May 2024

1. Introduction

Image denoising, a fundamental task in computer vision [1–4], involves the restoration of clean images from noisy ones, thereby enhancing their quality. The effectiveness of denoising directly impacts various downstream tasks in computer vision applications, including super resolution [5–7], semantic segmentation [8–10], and object detection [11–13]. Moreover, denoising techniques play a crucial role in improving the image quality captured by diverse devices like mobile phones, reflecting the widespread demand in imaging domains. In the context of image denoising, the objective is to restore a 2D image $x \in \mathcal{I} = \mathbb{R}^m$

from a noisy observation $y \in \mathcal{I}$ corrupted by random noise n, where both x and y belong to the image space $\mathcal{I}$. In general, the noisy image y can be modeled as

$$y = x + n \tag{1}$$

It is common to assume that the entries of the noise vector n are mutually independent. Many image denoising methods rely on this assumption [14,15]. Moreover, these methods assume that the image exhibits some statistically meaningful structure that can be exploited to remove the noise [14,16]. For example, there is an expectation that certain structures of the image are repeated elsewhere in the image. Over the past decades, a multitude of image denoising techniques has emerged, leading to significant advancements in the performance of state-of-the-art methods. Generally, these approaches can be categorized into two main groups: model-based methods and deep-learning-based denoisers. Model-based techniques, exemplified by seminal works like BM3D [14], NLM [17], NCSR [16], and WNNM [18], excel in addressing denoising challenges across varying noise levels through model optimization procedures. Nevertheless, these approaches often incur significant computational overhead due to their intricate algorithmic intricacies. Moreover, model-based methodologies typically rely on handcrafted priors such as nonlocal self-similarity [17] and sparsity [16], which may lack the robustness needed to effectively capture intricate image structures.

Improved by their robust learning and representation capabilities, convolutional neural networks (CNNs) demonstrate superior performance in tasks related to image denoising. Over the past decade, methods based on deep neural networks (DNNs) have garnered remarkable success. One notable example is the DnCNN (Denoising Convolutional Neural Network) [15], which leverages residual learning and batch normalization to enhance both training efficiency and denoising efficacy. However, conventional CNN architectures may encounter challenges in handling objects across various locations and scales, particularly for larger targets or those with non-fixed positions within the image. To overcome the inherent limitations of CNN structures, the Transformer architecture has been introduced into low-level visual image processing. Proposed in 2022, Restormer [19] stands out as an efficient Transformer-based method capable of handling high-resolution images. It accomplishes this by modeling global context through self-attention mechanisms across channels, thereby reducing computational overhead. More recently, Guo [20] proposed a pixel-wise crossmodal image denoising method (CMID) based on deep reinforcement learning, aiming to effectively address noise across different modalities. However, the performance of networks trained on synthetic data may significantly degrade when tested on real-world data due to domain gaps between synthetic and real datasets. Although existing supervised learning models demonstrate outstanding denoising performance, they often suffer from data bias issues, necessitating a substantial number of noisy–clean image pairs to enhance their generalization ability. However, acquiring such clear images poses a significant challenge.

To mitigate this issue, researchers have proposed a series of unsupervised and self-supervised methods that obviate the need for clean images to train denoising networks. For instance, Lehtinen et al. [21] introduced a statistical point estimation procedure and devised the Noise2Noise (N2N) denoising model, which advocates training models solely with corrupted image pairs for different noise distributions. This approach notably streamlines the training process by relaxing the demand for clean images. However, its performance may degrade when handling images divergent from the training dataset, and its generalization capability remains to be improved. Subsequently, Ulyanov et al. proposed an unsupervised method named Deep Image Prior (DIP) [22], which requires only a noisy image and a suitable network structure (e.g., UNet) to accomplish the denoising task, showcasing remarkable flexibility. DIP combines optimization iterations with the prior modeling prowess of deep neural networks (DNNs). As it solely relies on the noisy image, DIP exhibits desirable practical applicability. Nonetheless, the denoising performance of DIP and its variants still lags behind supervised or unsupervised learning, particularly for

synthetic images. In contrast, Noise2Void (N2V) [23] and Noise2Self (N2S) [24] employ a blind spot strategy by randomly discarding some input pixels and utilizing their neighboring pixels to predict them. This approach helps prevent overfitting during DNN training for mapping noisy images to themselves. However, the large area occupied by blind spots on the inputs leads to valuable context loss in the receptive field of predicted pixels, resulting in subpar performance. Additionally, optimizing partial pixels in each iteration slows down convergence. Recently, Zhang proposed a multi-mask strategy, Multi-Masked BSNs (MM-BSN), for self-supervised sRGB image denoising [4]. A multi-mask strategy can effectively break down large noise structures that were previously challenging to handle with single-center masked models. Han proposed SS-Attention, a novel self-similarity-based self-attention module that integrates SS-Attention into the blind spot network (SS-BSN), enabling training in a fully self-supervised manner [25]. Nonetheless, the performance of these denoising models still leaves room for improvement. In 2021, Huang et al. introduced the Neighbor2Neighbor (Ne2Ne) denoising model, which employs a random neighbor sub-sampling strategy to generate noisy image pairs for training [26]. Ne2Ne further reduces the acquisition difficulty of training images in the unsupervised model. However, the sampling operation reduces image resolution and may degrade denoising effectiveness. Moreover, Ne2Ne is unsuitable for addressing spatially correlated noise and extremely dark images. In summary, while unsupervised denoising models can be trained without clear images, they are generally less effective than supervised models.

In this work, our main objective is to enhance the performance of unsupervised denoising models. To achieve this, we propose a novel approach called the dual-branch self-boosting network (DSBNet) based on the Noise2Noise (N2N) framework. DSBNet incorporates both residual mapping and nonresidual mapping in a dual learning manner, enabling accurate estimation of the latent clean image, i.e., denoised image. Unlike N2N, which generates a single latent clean image, DSBNet employs two separate branches within a single network framework. One branch is responsible for generating the noise component prediction, while the other branch predicts the content (clean) component. These two branches have their own strengths and complement each other. During unsupervised training, we leveraged multiple loss terms to facilitate mutual enhancement between the branches, eliminating the need for reference images. The training process involved generating noise and content prediction results using the dual-branch network. Then, while keeping the network parameters fixed, we fed these results back into the network to obtain new noise and content prediction results. By providing preliminary separation of image content or noise as input to the network, we expected the processed output to yield more accurate estimates, thereby improving the denoising performance. To guide the network training, we constructed a powerful loss function based on the network outputs which updates the parameter values. The inter-branch loss plays a crucial role in promoting the performance of both branches during unsupervised learning. During the inference stage, we only utilized the content prediction branch for image denoising. We conducted extensive evaluations on four public datasets (Set12, BSD68, SIDD, and DND) to validate the effectiveness of DSBNet. Additionally, ablation studies were performed to analyze the key modules and loss function, demonstrating the superiority of our method. DSBNet achieved significant performance improvements compared to previous unsupervised methods like N2N, surpassing even strongly supervised approaches when applied to synthetic and real-world noisy images, showcasing its excellent generalization capability. The code for DSBNet can be accessed at the following GitHub repository: https://github.com/gengyuhang/DBSNet (accessed on 16 April 2024). The main contributions of our work can be summarized as follows:

(1) Denoising, a foundational task in image processing, can be approached through two distinct perspectives: directly predicting the image content or predicting the noise component of the image. To enhance the effectiveness of denoising, we propose a novel dual-branch network architecture. This architecture enables simultaneous prediction of both the noise component and the clean image within a unified frame-

work. By incorporating two parallel and complementary generators, our methodology effectively suppresses noise and extracts intricate image details, resulting in the reconstruction of a clean image. This unified framework leverages the benefits of residual learning to tackle challenges associated with deep architectures and vanishing gradients while also taking advantage of the comprehensive mapping capabilities offered by nonresidual learning;

(2) In the proposed DSBNet, when a noisy image is inputted, it undergoes a decomposition process within the dual-branch network architecture, resulting in the extraction of a clean image and a noise component. To further refine the network's performance, we introduced a feedback mechanism where a single re-input of either the noise component or the image content component exclusively activates one branch and deactivates the other. Leveraging this mechanism, we reintroduced the clean image and the noise component into the network, incorporating two additional loss terms alongside the classic N2N loss function. These additional loss terms act as robust constraints during unsupervised training, enhancing stability and reducing the complexity of the training process. With the support of the dual-branch architecture, we established stronger constraints to effectively train the network model, resulting in the generation of more accurate image content;

(3) The proposed DSBNet was assessed on three publicly available test datasets, showcasing outstanding performance in terms of PSNR and SSIM metrics. Moreover, the subjective evaluation revealed richer texture details in the denoised images. Through rigorous experimental validation, our DBDNet was shown to exhibit superior denoising prowess compared to alternative methods across both grayscale and color images. It is worth mentioning that the test images encompassed a combination of synthesized noise and real-world noise present in various scenes.

The rest of this paper is organized as follows. Section 2 reviews the existing related image denoising methods with particular emphasis on highlighting the Noise2Noise approach, and Section 3 presents the proposed dual-branch image denoising model. Section 4 provides details from the experimental evaluations of the proposed method. Section 5 concludes this paper.

2. Related Work

2.1. Noise2Noise

Lehtinen [21] introduced Noise2Noise, a model for unsupervised noise reduction, enabling image restoration without dependency on a clear image as the training target. This innovative approach solely relies on paired independent noisy images originating from the same ground truth. Noise2Noise operates under the assumption of accessing a collection of noisy image pairs $y_1 = x + n_1, y_2 = x + n_2$, where n_1, n_2 denote independent noise vectors. The optimization of $\boldsymbol{\theta}$ in Noise2Noise aims to minimize the following loss:

$$\frac{1}{n}\sum_{i=1}^{n}\left\|f_{\boldsymbol{\theta}}\left(y_1^i\right)-y_2^i\right\|_2^2 \tag{2}$$

where $f_{\boldsymbol{\theta}}$ denotes the denoising network parameterized by $\boldsymbol{\theta}$. Considering Equation (2), if the dataset is sufficiently large, we have

$$\begin{aligned}\arg\min_{\theta}&\mathbb{E}\left\|f_\theta\left(x_i+n_2^i\right)-\left(x_i+n_1^i\right)\right\|_2^2\\ &=\arg\min_{\theta}\mathbb{E}\|f_\theta(y_i)-x_i\|_2^2-2\mathbb{E}\left[{n_1^i}^T f_\theta(y_i)\right].\end{aligned} \tag{3}$$

Because n_1^i and n_2^i are two independent noises, so $\mathbb{E}\left[n_1^i \mid x_i + n_2^i\right] = \mathbb{E}\left[n_1^i \mid x_i\right] = 0$. Therefore, we have

$$\begin{aligned}\mathbb{E}\left[n_1^{iT} f_\theta(y_i)\right] &= \mathbb{E}\left[\mathbb{E}\left[n_1^i \mid f_\theta(y_i)\right]^T f_\theta(y_i)\right] \\ &= \mathbb{E}\left[\mathbb{E}\left[n_1^i \mid x_i + n_2^i\right]^T f_\theta(y_i)\right] \\ &= 0.\end{aligned} \tag{4}$$

In expectation of such noisy instances, and assuming zero mean noise, training a network in a supervised manner to map a noisy image to another noisy image is equivalent to mapping it to a clean image, i.e.,

$$\arg\min_{\theta} \mathbb{E}\left[\|f_\theta(y_1) - x\|_2^2\right] = \arg\min_{\theta} \mathbb{E}\left[\|f_\theta(y_1) - y_2\|_2^2\right]. \tag{5}$$

The N2N denoising model, trained on pairs of noisy images without access to ground truth clean images, utilizes an unsupervised learning approach that offers certain advantages. In theory, N2N training can achieve the same performance as Noise2Clean (N2C) training given an infinitely large dataset. However, in practice, due to limited training set size, N2N falls short of N2C.

While N2N [21] demonstrates great flexibility in image denoising, its results may not always be optimal. One key limitation of the Noise2Noise (N2N) model is its restricted generalization ability when confronted with noise characteristics that differ significantly from the training data. Furthermore, in order to achieve optimal denoising performance, the N2N technique may require the selection of different loss functions tailored to specific application scenarios. Additionally, the N2N model assumes that the noise statistics follow an independent and identically distributed pattern. However, real-world noise often exhibits more complex, spatially correlated characteristics. Consequently, the denoising performance of N2N inevitably degrades when dealing with this type of spatially correlated noise. It is also worth noting that the performance of the N2N model is influenced by the quality and capacity of its underlying backbone network architecture. In the current implementation, a UNet architecture is utilized as the backbone network, as illustrated in Figure 1. However, the use of a more advanced network structure has the potential to enhance the denoising performance of the N2N model further. The limitations of the N2N denoising model mentioned above can be attributed to its overreliance on a simplistic loss function definition, which only establishes constraint between the network output and input. This limited constraint fails to ensure the stability and generalization capability of the model during training. In order to address these issues, more advanced loss items need to be employed that capture the complex relationships and structural information within the clean images. By incorporating such loss items, the N2N model can better guide the denoising process and improve its stability and generalization ability.

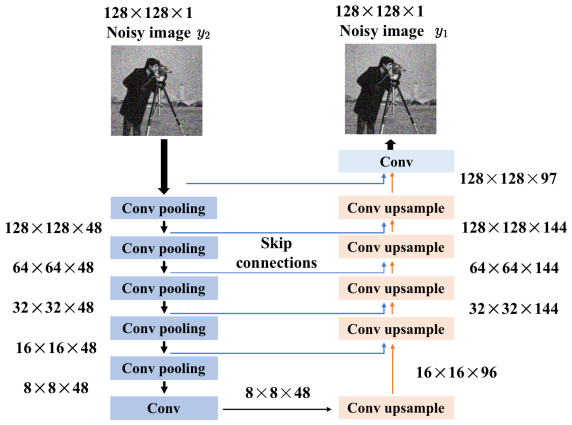

Figure 1. The UNet architecture used in unsupervised N2N. During the training phase, both the input and target images of the network are noisy images depicting the same scene.

2.2. Nonresidual- and Residual-Learning-Based Denoising Models

To provide a more comprehensive description of our work, we have reformulated Equation (1) for clarity as follows:

$$I_{\text{noisy}} = I_{\text{content}} + I_{\text{noise}} \tag{6}$$

where I_{noisy}, I_{content}, and I_{noise} represent the noisy image to be denoised, the clean image, and the noise, respectively. Based on the diverse mapping objectives, current deep-learning-based denoising models can be categorized into two methodologies: nonresidual (content) learning and residual (noise) learning. Both methodologies possess distinct advantages while being capable of predicting the latent clean image. Nonresidual-learning-based denoising models, such as N2N (Noise2Noise) [21] and FFDNet (Fast and Flexible Denoising Network) [27], typically employ neural networks denoted as f_{content}, which aim to transform a given noisy image I_{noisy} into an approximation $\hat{I}_{\text{content}} = f\text{content}(I_{\text{noisy}})$ of the clean image I_{content}. Residual-learning-based denoising models, such as the DnCNN (Denoising Convolutional Neural Network) [15] and Restormer [19], first employ neural networks designed to learn residual mappings. The final clean image is obtained by subtracting the predicted noise from the noisy image, formulated as $I_{\text{content}} = I_{\text{noisy}} - \hat{I}_{\text{noise}} = I_{\text{noisy}} - f_{\text{noise}}(I_{\text{noisy}})$. Generally speaking, the advantage of nonresidual learning lies in its simplicity and straightforwardness as the model is relatively lightweight and easy to train and understand. However, without residual connections, deeper networks may face issues such as vanishing or exploding gradients, leading to training difficulties. On the other hand, residual learning addresses the vanishing or exploding gradient problem more effectively by introducing residual connections, ensuring more stable training. Additionally, residual learning can learn deeper relationships between input and output, thereby improving the model's representational and generalization capabilities. Note that most of existing denoising models predict a single task while neglecting the potential of dual learning with two tasks, i.e., estimating residuals and estimating latent clean images [28,29].

3. Methodology

In this section, we propose a novel denoiser based on a dual-branch neural network named DSBNet. Considering the gap between predicting residuals and directly learning mapping from noisy images to clean data, the proposed DSBNet is designed to investigate the correlations between these two tasks. To begin, we present the fundamental principles underlying our proposed solution. Following that, we introduce the network architecture employed in our approach. Lastly, we delve into the explanation of the loss function used for optimization.

3.1. Basic Idea

As mentioned above, it is evident that Noise2Noise exhibits its greatest advantage in not requiring clean reference images for training. This characteristic simplifies both the model design and implementation, making it highly accessible and easy to deploy. The model's ability to learn directly from pairs of noisy images enables effective noise suppression while preserving important details, enhancing its robustness in real-world scenarios. Additionally, its simplicity and straightforward implementation make it a practical and efficient solution for denoising tasks. However, despite these advantages, it is important to note that Noise2Noise also has some limitations.

One of the main limitations of Noise2Noise is its focus on predicting image content while neglecting noise prediction. Incorporating noise prediction within the framework of the N2N model can lead to performance enhancement. Firstly, by leveraging the noise component, finer details in the image can be better predicted by comprehending noise characteristics. Noise often correlates with high-frequency content and fine textures in an image. Explicitly modeling and predicting noise enables the model to more effectively capture and reconstruct these subtle details, ensuring effective noise removal while preserving

underlying content, thereby enhancing image quality. Secondly, the introduction of an additional branch imposes stronger constraints during training, enhancing stability and efficacy. The noise prediction branch serves as an auxiliary task, introducing additional supervision signals to guide learning. This supplementary information aids in regularizing the model, promoting the acquisition of more robust representations. Consequently, network training becomes more stable, with the noise prediction branch contributing to a stronger regularization effect and mitigating overfitting, rendering the network more efficient for practical applications.

In conclusion, the incorporation of the dual-branch self-boosting technique in N2N allows for the exploitation of content learning and noise learning, leading to improved denoising performance. This new approach not only enhances the stability and efficiency of generating clean images but also results in images with richer details. As a result, the utilization of the dual-branch self-boosting model can significantly enhance the effectiveness of denoising compared to the original N2N model.

3.2. Dual-Branch Constraints

Given the simplistic constraints utilized in training the N2N model, we propose a novel approach to leverage the dual-branch network to simultaneously generate image content and noise components. Subsequently, we reintroduce the generated image content and noise components back into the dual-branch network, thereby imposing stronger constraints on the denoising process. As shown in Figure 2, we utilize a dual-branch network denoted by f_θ to decompose a given noisy image into its content and noise components, i.e., $(I_{content}, I_{noise})$. Based on this, we denote the noisy image I_{noisy} as $I_{noisy} = g(s_1 I_{content}, s_2 I_{noise})$, where $s_1 = s_2 = 1$. Here, g serves as an image additive blending function. Following the decomposition of the provided noisy image into its constituent clean image and noise components, we can then input both the content and noise components back into the identical network, utilizing the exact same set of parameters θ as in the initial image decomposition process. Ideally, since the predicted content image should not contain any noise, it can be modeled as $I_{content} = g(s_1 I_{content}, s_2 I_{noise})$, where $s1 = 1$ and $s2 = 0$. Similarly, the noise component should not contain any content; it can be modeled as $I_{noise} = g(s_1 I_{content}, s_2 I_{noise})$, where $s1 = 0$ and $s2 = 1$. Thus, these two conditions can be leveraged to formulate additional constraints for training the dual-branch network, thereby augmenting the efficacy of the N2N framework. The network learns to effectively separate the content and noise components and generate denoised images by leveraging the shared parameters θ. This approach capitalizes on the fact that the network can benefit from a nonresidual and residual learning process, where the same parameters are utilized for both the initial decomposition and the subsequent reintroduction of the components. Overall, dual-branch constraints promote a more robust and accurate denoising process, enabling the network to effectively exploit the relationships between the content and noise components, ultimately producing high-quality denoised images.

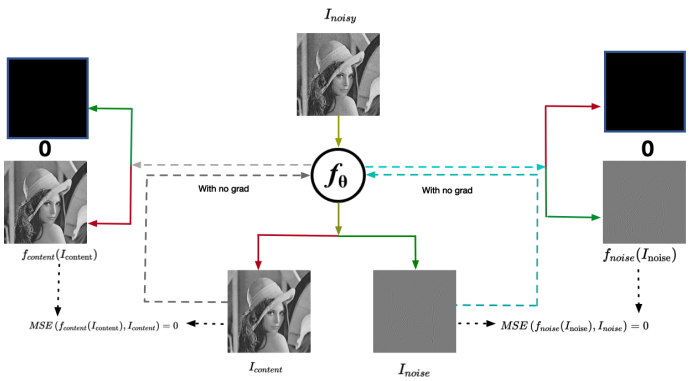

Figure 2. The additional constraints established on the basis of the dual-branch architecture. Once the neural network f_θ decomposes the input noisy image I_{noisy} into its content and noise components

($I_{content}$,I_{noise}), these decomposed parts can be fed back into the network as inputs again. The resulting decomposed image components will be exclusive; when the image content is input into the network f_θ, it should generate content $f_{content}(I_{content})$ on the content branch consistent with the input $I_{content}$ while remaining completely unresponsive on the noise branch $f_{noise}(I_{content}) == \mathbf{0}$. Conversely, when the decomposed noise part I_{noise} is fed into the dual-branch network, the response obtained on the noise branch $f_{noise}(I_{noise})$ should be identical to the input noise component I_{noise} while remaining completely unresponsive on the image content branch $f_{content}(I_{noise}) == \mathbf{0}$. Based on these constraints, network training will become more stable and capable of generalization.

3.3. Network Architecture

This work introduces a novel approach to image denoising by utilizing a dual-branch network architecture. The schematic representation of the entire network is depicted in Figure 3. The model consists of two main components: the clean image generator and the noise generator. To reconstruct the clean output image $I_{content}$ from a given noisy image, the clean generator employs the DnCNN architecture [15] without skip connections and batch normalization layers (BN). The DnCNN architecture has been widely used for image denoising tasks and has demonstrated excellent performance. By omitting skip connections and batch normalization layers, we aimed to simplify the network architecture while maintaining its effectiveness in denoising. The noise generator, on the other hand, is responsible for generating the noise component of the image. It comprises three types of layers, as depicted in Figure 3. The initial layer utilizes 64 filters of size $3 \times 3 \times c$ to generate 64 feature maps. Rectified linear units (ReLU) are then applied for nonlinearity. Here, 'c' represents the number of image channels, where 'c = 1' for grayscale images, and 'c = 3' for color images. For layers 2 to $(D-1)$, we continued to employ 64 filters of size $3 \times 3 \times 64$. Additionally, batch normalization was incorporated between the convolution and ReLU layers in order to accelerate the training process and enhance the network's generalization ability. Finally, the last layer of the noise generator utilizes 'c' filters of size $3 \times 3 \times 64$ to reconstruct the denoised output image. By combining the clean image generator and the noise generator within the dual-branch network, we aimed to exploit the complementary information captured by these two components. The clean image generator focuses on recovering the underlying content, while the noise generator aims to estimate and remove the noise component. This joint learning process enables the network to effectively denoise the input images and produce high-quality output results.

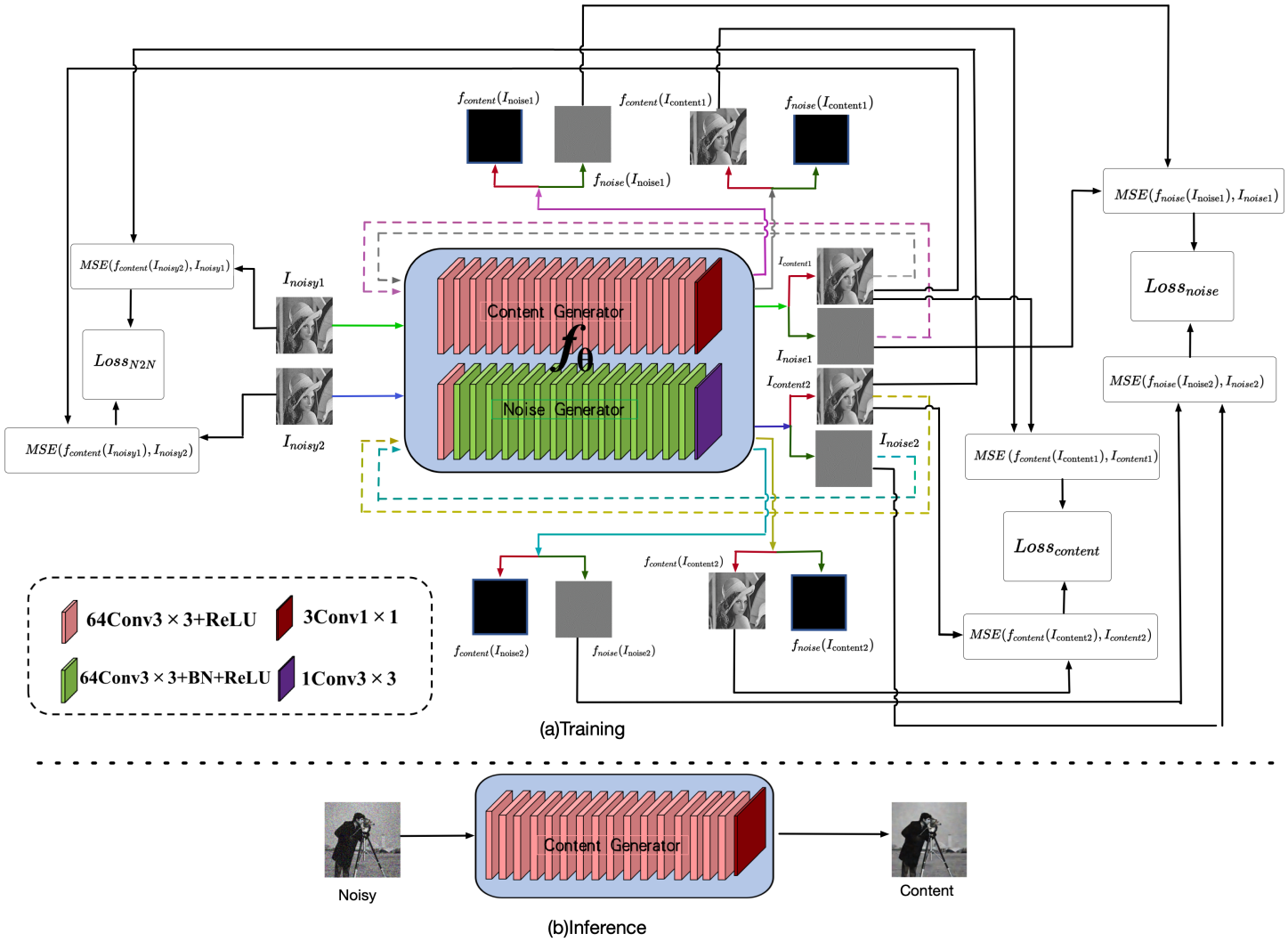

Figure 3. Overview of our proposed DBSNet framework. (**a**) The given noisy image during training passes through the content generator and the noise generator. The decomposed image content and noise component will then be fed back to the DBSNet to establish more powerful constraints for training. (**b**) When inferencing, the given noisy image only passes through the content generator to obtain the denoised image.

3.4. Loss Function

It is well known that choosing an appropriate loss function plays a key role in training a DNN. In this subsection, we devise three mean squared error (MSE) losses to formulate a compound loss function aimed at addressing the corresponding optimization problem. The MSE metric computes the squared difference between pixel values of images, thereby promoting smoothness and contrast in the resulting image output. Notably, in the augmentation of the existing MSE loss function within the N2N model, we introduced two additional loss items for image content and noise, thus crafting a more potent composite loss function. The standard MSE loss definition is as follows:

$$MSE(I_1, I_2) = \frac{1}{M \times N} \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} [I_1(i,j) - I_2(i,j)]^2 \tag{7}$$

where $I_1(i,j), I_2(i,j)$ denote the input images, and M and N represent the size of input images. In this work, there are three loss functions: $Loss_{N2N}, Loss_{content}, Loss_{noise}$. When we feed the noisy image I_{noisy1} into the network, we can obtain the clean image $I_{content1}$ and the noise I_{noise1}, and, when we feed the noisy image I_{noisy2} into the network, we can obtain the clean image $I_{content2}$ and the noise I_{noise2}.

N2N loss: The N2N loss, which is similar to the loss function in the original N2N, evaluates the difference between the generated content $I_{content1}$ and another noisy picture I_{noisy2}. The expression is as follows:

$$\begin{aligned} Loss_{N2N} = {} & MSE\left(f_{content}\left(I_{noisy1}\right), I_{noisy2}\right) \\ & + MSE\left(f_{content}\left(I_{noisy2}\right), I_{noisy1}\right) \end{aligned} \tag{8}$$

Content loss: When our network takes a denoised image $I_{content1}$ as an input, we expect the model to predict the same image as output without any noise terms. Similarly, when $I_{content2}$ is the input, we expect the output image to be free of noise. Therefore, the loss function that constrains the content branch of the image decomposition is defined as follows:

$$\begin{aligned} Loss_{content} = {} & MSE\left(f_{content}(I_{\text{content1}}), I_{content1}\right) \\ & + MSE(f_{content}(I_{\text{content2}}), I_{content2}) \end{aligned} \tag{9}$$

Noise loss: When we take noise I_{noise1} as an input, we expect that the output is noise without content. Similarly, when I_{noise2} is the input, we expect the output image to be free of content. Therefore, the loss function that constrains the noise branch of the image decomposition is defined as follows:

$$\begin{aligned} Loss_{noise} = {} & MSE\left(f_{noise}(I_{\text{noise1}}), I_{noise1}\right) \\ & + MSE(f_{noise}(I_{\text{noise2}}), I_{noise2}) \end{aligned} \tag{10}$$

In contrast to the loss functions employed by the N2N denoising model, the enhanced loss function utilized in this paper is capable of constraining the network to generate ideal content and noise components, thereby achieving complete separation between image content and noise.

4. Experiments

4.1. Datasets and Experiment Setup

In this section , we present extensive experiments conducted on diverse benchmark datasets to validate the advantage of the proposed DBSNet for both synthetic and real-world noise removal scenarios. Our method is compared with several state-of-the-art techniques, including BM3D [14], N2N [21], DnCNN [15], IRCNN [30], FFDNet [27], VDNet [31], SwinIR [32], Restormer [19], and FEUNet [33]. The benchmark datasets utilized in the experiments include Set12 [15], BSD68 [34], SIDD [35], and DND [36]. Set12 is widely acknowledged as a benchmark database in the literature, while BSD68 comprises a random

selection of 68 images converted to grayscale from the BSD500 database [37]. These images are characterized by rich content and texture details, rendering them suitable for evaluating the robustness of competing methods to image content. SIDD is a representative real-world dataset containing well-aligned noise–clean image pairs for training purposes. The DND benchmark consists of 50 noisy images captured with consumer-grade cameras of various sensor sizes. Each image is cropped into 20 patches of size 512×512, resulting in a total of $50 \times 20 = 1000$ samples for evaluation. Compared to the SIDD dataset, images in the DND dataset are captured under normal lighting conditions, thus exhibiting weaker noise. Moreover, to quantify the performance of different methods, we adopted PSNR (peak signal-to-noise ratio) and the SSIM (structural similarity index measure) as evaluation metrics. PSNR evaluates the intensity similarity between undistorted and generated images, with higher PSNR values indicating better denoising ability. The SSIM is an image quality evaluation metric closely related to human perception of image quality, where higher SSIM values imply more satisfactory synthesis performance visually. Our experiments were implemented in PyTorch 1.7.1, developed by Facebook's artificial intelligence research team, [38] and conducted on a single NVIDIA RTX 4090 GPU, a powerful graphics card manufactured by NVIDIA Corporation in Santa Clara, CA, United States. The neural network was initialized using the Xavier initialization method. The bias parameters were set to a constant value of zero. The Adam optimizer was employed with a learning rate of 0.0001 and a weight decay parameter of 1×10^{-9} .

4.2. Ablation Experiment

As mentioned earlier, the N2N model exhibits remarkable capability in recovering a clean image from a noisy input, although its overall performance on the entire image still falls short of that achieved by supervised state-of-the-art counterparts. This discrepancy arises from the absence of constraints imposed on the N2N model. To investigate the effects of different loss items, we conducted several experiments aimed at elucidating the impact of the loss item employed by our unsupervised approach. The experiments were conducted on the Set12 database with a noise level of $\sigma = 25$. We assessed various combinations of loss terms, namely, the N2N loss term $\mathcal{L}_{N2N}$, the content loss term $\mathcal{L}_C$, and the noise loss term $\mathcal{L}_N$. Table 1 presents the results of each loss function during training. As shown in Table 1, while incorporating the loss term $\mathcal{L}_{N2N}$ is crucial for training our model, adding either $\mathcal{L}_C$ or $\mathcal{L}_N$ alone can enhance the denoising effect. Furthermore, the simultaneous inclusion of both $\mathcal{L}_C$ and $\mathcal{L}_N$ leads to a significant performance improvement and enhances the stability of the learning process. The loss item specifically tailored for one branch will, in turn, enhance the performance of the other branch. The above experiment shows the self-boosting characteristics of the dual-branch structure.

Table 1. Ablation study with different combinations of loss items on the Set12 dataset (σ = 25).

$\mathcal{L}_{N2N}$	$\mathcal{L}_{Content}$	$\mathcal{L}_{Noise}$	PSNR
√			31.05
√	√		31.83
√		√	31.76
√	√	√	32.21

In the original N2N framework, the clean image generator employed was UNet. However, it has been observed that UNet is susceptible to rotation and scale variations in the input image, leading to unstable performance at different angles or scales. To overcome this limitation and enhance robustness, additional data augmentation or preprocessing steps are typically necessary. In our method, we utilize DnCNN as the clean image generator while excluding the skip connection and batch normalization layers (BN). A comparison between our approach and N2N was conducted on the Set12 dataset, incorporating various levels of noise. The results, presented in Table 2, clearly demonstrate the superior performance of our clean image generator using DnCNN.

Table 2. PSNR(dB) results using different backbone networks on the Set12 dataset with noise level σ at 15, 25, and 50.

	$\sigma = 15$	$\sigma = 25$	$\sigma = 50$
UNet	32.19	31.32	27.64
Ours	34.65	32.21	28.91

The proposed approach utilizes a dual-branch network to generate both the clean image (content) and noise simultaneously. In order to showcase the efficacy of our dual-branch network, we conducted a comparative analysis with methods that directly predict noise and directly predict clean images. The outcomes, as presented in Table 3, highlight the effectiveness of the dual-branch network. The highest PSNR values are highlighted in bold. Notably, the PSNR result of the dual-branch network method exhibits a significant improvement compared to the PSNR result achieved using a single-branch network. This observation underscores the capability of our dual-branch strategy to facilitate productive collaboration between the content and noise branches, thereby yielding superior denoising performance.

Table 3. PSNR(dB) results of networks using different branches on the Set12 dataset with noise level σ at 15, 25, and 50.

	σ = 15	σ = 25	σ = 50
Dual branch	**34.65**	**32.21**	**28.91**
Directly predict content	33.67	30.29	25.71
Directly predict noise	33.08	30.69	27.53

4.3. Overall Performance Comparison

To thoroughly evaluate the denoising effectiveness of DBSNet, we conducted comparative experiments on the grayscale image datasets Set12 and BSD68. We introduced Gaussian noise to the images at noise levels of 15, 25, and 50. In order to analyze the denoising performance of each method in more detail, we present the specific PSNR results for each image at noise levels 15, 25, and 50 in Tables 4–6. Additionally, Table 7 showcases the average PSNR results obtained from the BSD68 dataset. The highest PSNR values are highlighted in bold. By examining Tables 4–7, it becomes evident that our proposed unsupervised method, DBSNet, outperforms other unsupervised methods and surpasses the majority of state-of-the-art supervised methods in terms of PSNR.

Table 4. PSNR results of various competing methods on individual images from Set12 dataset corrupted with Gaussian noise ($\sigma = 15$).

Methods	BM3D	DnCNN	FFDNet	Restormer	IRCNN	SwinIR	FEUNet	N2N	Ours
Cameraman	31.90	32.14	32.42	33.02	32.53	33.04	32.48	33.72	**34.39**
House	34.92	34.96	35.01	**36.18**	34.88	35.27	36.18	35.85	36.15
Peppers	32.71	33.09	33.10	33.63	33.21	33.66	33.35	34.45	**34.67**
Starfish	31.12	31.92	32.02	32.58	31.96	32.57	32.05	33.69	**34.35**
Monarch	31.95	33.08	32.77	33.73	32.98	33.68	32.88	34.58	**35.17**
Airplane	31.10	31.54	31.58	32.07	31.66	32.08	31.69	33.57	**34.10**
Eagle	31.31	31.64	31.77	32.14	31.88	32.16	31.96	33.20	**33.49**
Lena	34.23	34.52	34.63	34.99	34.50	34.97	34.73	35.10	**36.06**
Barbara	33.04	32.03	32.50	33.56	32.41	**33.98**	32.53	32.19	33.91
Boat	32.09	31.93	32.35	32.83	32.36	32.81	32.41	33.26	**34.14**
Man	31.90	31.98	32.40	32.63	32.36	32.67	32.43	33.33	**34.68**
Couple	32.04	32.38	32.45	32.81	32.37	32.83	32.54	33.47	**34.28**
Avg.	32.36	32.67	32.75	33.35	32.76	33.36	33.86	33.86	**34.61**

Table 5. PSNR results of various competing methods on individual images from Set12 dataset corrupted with Gaussian noise ($\sigma = 25$).

Methods	BM3D	DnCNN	FFDNet	Restormer	IRCNN	SwinIR	FEUNet	N2N	Ours
Cameraman	29.38	30.03	30.06	30.72	30.12	30.69	30.15	31.44	**31.46**
House	32.86	33.04	33.27	34.18	33.02	33.91	33.37	34.08	**34.21**
Peppers	30.20	30.73	30.79	31.31	30.81	31.28	31.01	31.84	**32.33**
Starfish	28.52	29.24	29.33	30.01	29.21	29.96	29.47	30.69	**31.52**
Monarch	29.30	30.37	30.14	31.02	30.20	30.95	30.30	32.02	**32.57**
Airplane	28.42	29.06	29.05	29.47	29.05	29.44	29.18	31.22	**31.45**
Eagle	28.94	29.35	29.43	29.75	29.47	29.75	29.64	30.94	**30.95**
Lena	32.07	32.40	32.59	33.00	32.40	32.98	32.73	32.97	**33.90**
Barbara	30.66	29.67	29.98	31.39	29.93	**31.69**	30.29	30.89	31.27
Boat	29.87	30.19	30.23	30.73	30.17	30.64	30.31	31.18	**31.89**
Man	29.59	30.06	30.10	30.33	30.02	30.32	30.12	31.90	**32.19**
Couple	29.70	30.05	30.18	30.60	30.05	30.57	30.25	31.28	**31.74**
Avg.	29.96	30.35	30.43	31.04	30.37	31.01	30.57	31.70	**32.12**

Table 6. PSNR results of various competing methods on individual images from Set12 dataset corrupted with Gaussian noise ($\sigma = 50$).

Methods	BM3D	DnCNN	FFDNet	Restormer	IRCNN	SwinIR	FEUNet	N2N	Ours
Cameraman	26.19	27.26	27.03	**27.88**	27.16	27.79	27.88	27.34	27.00
House	29.57	29.91	30.43	**31.56**	29.91	31.11	31.56	30.73	30.98
Peppers	26.67	27.35	27.43	28.04	27.33	27.91	28.04	27.71	**28.23**
Starfish	24.90	25.60	25.77	26.66	25.48	26.55	26.66	25.87	**27.16**
Monarch	25.71	26.84	26.88	27.39	26.66	27.31	27.39	27.01	**28.55**
Airplane	25.17	25.82	25.90	26.17	25.78	26.14	26.17	26.01	**27.40**
Eagle	25.87	26.48	26.58	**26.92**	26.48	26.91	26.92	26.67	26.65
Lena	29.02	29.34	29.68	30.19	29.36	30.11	30.19	29.81	**30.49**
Barbara	27.21	26.32	26.48	28.31	26.17	**28.41**	28.31	27.09	27.29
Boat	26.73	27.18	27.32	27.83	27.17	27.70	27.83	27.47	**28.36**
Man	26.79	27.17	27.30	27.51	27.14	27.45	27.51	27.28	**28.60**
Couple	26.47	26.87	27.07	27.61	26.86	27.53	27.61	27.24	**28.02**
Avg.	26.70	27.18	27.32	28.01	27.12	27.91	28.01	27.52	**28.37**

Table 7. The average PSNR results of various competing methods on images from BSD68 datasets corrupted with Gaussian noise (σ = 15, σ = 25, σ = 50).

Methods	BM3D	DnCNN	FFDNet	Restormer	IRCNN	SwinIR	FEUNet	N2N	Ours
$\sigma = 15$	31.15	31.62	31.65	31.88	31.65	31.91	31.69	32.86	**33.92**
$\sigma = 25$	28.57	29.14	29.17	29.41	29.13	29.41	29.28	30.55	**31.22**
$\sigma = 50$	25.58	26.18	26.24	26.49	26.14	26.47	26.39	26.89	**27.78**

To evaluate the performance of our proposed method in real-world denoising scenarios, we conducted experiments on the SIDD and DND datasets. We compared our denoising results with the results obtained with BDNet [39], DnCNN [15], R2R [40], RIDNet [41], N2V [24], AP-BSN [1], and MM-BSN [4] methods. The real-world image denoising performance of different methods on the SIDD and DND datasets is presented in Tables 8 and 9. The highest PSNR values are highlighted in bold. It is evident from the results that our proposed method exhibits significant superiority over the unsupervised methods R2R and N2V in terms of PSNR and the SSIM, surpassing the performance of the majority of supervised methods.

Table 8. Quantitative comparison of real-world denoising on images in SIDD benchmark dataset. We compared DBSNet with other denoising methods in terms of PSNR and SSIM.

Method	DnCNN	N2V	CBDNet	R2R	RIDNet	AP-BSN	MM-BSN	Ours
PSNR	23.66	27.68	33.28	34.78	**38.70**	35.97	37.37	38.49
SSIM	0.583	0.668	0.868	0.844	0.950	0.925	0.936	**0.963**

Table 9. Quantitative comparison of real-world denoising on RGB images in DND benchmark dataset. We compared DBSNet with other denoising methods in terms of PSNR and SSIM.

Method	DnCNN	N2V	CBDNet	R2R	RIDNet	AP-BSN	MM-BSN	Ours
PSNR	32.43	35.61	38.05	38.41	39.25	38.13	38.82	**39.27**
SSIM	0.790	0.824	0.942	0.949	0.952	0.937	0.940	**0.959**

4.4. Visual Comparison

Due to its rich texture details, we chose the Lena image to facilitate a comprehensive visual comparison. The outcomes obtained from various competing denoising methods are presented in Figure 4. For closer scrutiny, we zoomed in on specific regions of the image. It is notable that, in regions exhibiting intricate texture details, methods such as FFDNet, DnCNN, and BM3D showed signs of excessive smoothing, leading to the loss of image details. Particularly in Lena's eye region, our method distinctly outlined her lower eyelashes, a feature smoothed out by alternative methods. Our proposed approach adeptly mitigates visual artifacts while retaining sharp edges and intricate details. The comparative analysis with competing methods underscores the superior visual fidelity achieved by our proposed methodology.

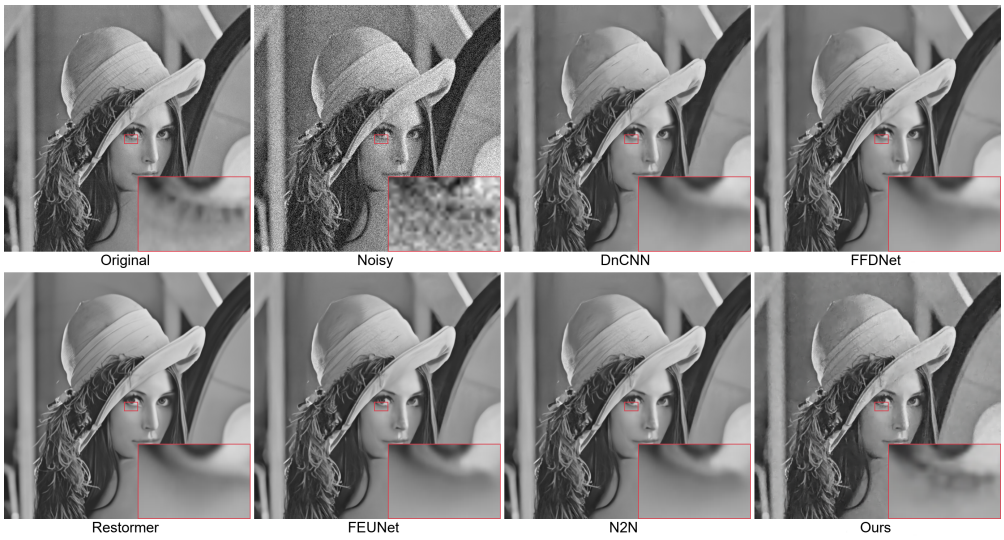

Figure 4. Visual denoising results of various methods on Lena image.

The distribution of actual noise in real-world applications is often unknown, leading to a significant degradation in the performance of state-of-the-art denoising methods when applied to real-world noisy images. To further illustrate the effectiveness of our proposed method, experiments were conducted on real-world noisy images, and the denoised results were compared with those obtained using FFDNet [27], SwinIR [32], and Restormer [19]. An image of a staircase alongside its denoised versions obtained through different methods is showcased in Figure 5. As depicted in Figure 5, most methods struggled to accurately restore details present in the noisy image, such as the cracks on the wall, whereas our method effectively restored these details. In summary, our approach consistently achieves superior visual fidelity, whether encountering synthetic or real noise scenarios.

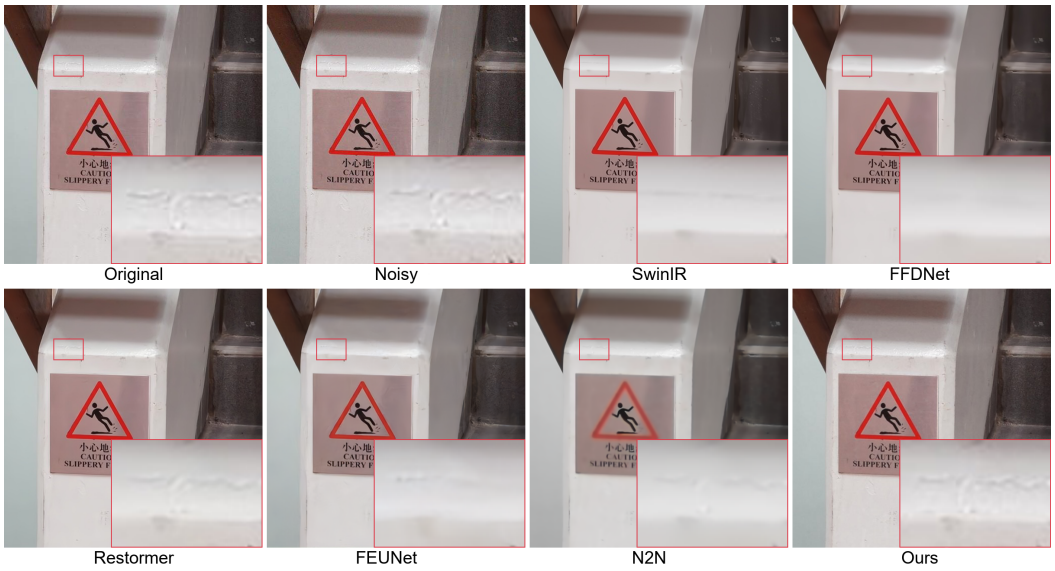

Figure 5. Visual denoising results of various methods on staircase image.

4.5. Computation Time

To assess the computational time of the model, comparative experiments were conducted on the Set12 dataset. The Set12 dataset comprises a total of 12 images, including seven images of size 256 × 256 pixels and five images of size 512 × 512 pixels. As shown in Table 10, the running time of our model was 0.236 ms, 6 times faster than Restormer, 20 times faster than AP-BSN, and 50 times faster than N2N. As a result, our method is more advantageous than before when dealing with images.

Table 10. Computation time (ms) of various competing methods on Set12 dataset images.

Method	DnCNN	CBDNet	Restormer	FEUNet	AP-BSN	MM-BSN	N2N	Ours
Time	0.014	0.092	1.684	0.032	5.024	0.224	12.107	0.236

5. Conclusions

In this work, we introduce a novel DBSNet, a dual-branch self-boosting network designed for precise, robust, and flexible image denoising without reliance on annotated images. Through meticulous analysis, we unveil a symbiotic relationship between content prediction (nonresidual mapping) and noise prediction (residual mapping) training strategies. To effectively harness the predictive power of both tasks, we propose a dual-branch network that concurrently trains on these two objectives and seamlessly integrates them within a unified architecture. To enhance network stability and denoising performance, we introduce a pioneering feedback mechanism. By incorporating this mechanism, we can introduce additional loss terms to refine the network's ability to predict both content and noise accurately. Extensive experiments validate the efficacy of our proposed dual-branch approach, which continually reinforces each branch during unsupervised training on unlabeled images. These findings demonstrate that DBSNet possesses the flexibility of unsupervised denoising networks and the efficacy of supervised denoising models. Our approach offers a solution or method to effectively tackle challenges in real-world image denoising, including limited training data, complex noise modeling, and the need for robust generalization.

Author Contributions: S.X. and Y.G. contributed to the conception of the study, Y.G. wrote the main manuscript text, M.X. and Q.C. contributed significantly to the analysis and manuscript preparation, and C.Z. conducted the experiments. All authors reviewed the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Natural Science Foundation of China, grant number 62162043.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available on request from the corresponding author. The data are not publicly available due to privacy.

Acknowledgments: The authors would like to extend their gratitude to the authors who generously provided the source code for their contribution to this research. Their code has been instrumental in enhancing the quality and reproducibility of our work.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Lee, W.; Son, S.; Lee, K.M. Ap-bsn: Self-supervised denoising for real-world images via asymmetric pd and blind-spot network. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New OR, LA, USA, 18–24 June 2022; pp. 17725–17734.
2. Li, J.; Zhang, Z.; Liu, X.; Feng, C.; Wang, X.; Lei, L.; Zuo, W. Spatially adaptive self-supervised learning for real-world image denoising. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Vancouver, BC, Canada, 17–24 June 2023; pp. 9914–9924.
3. Tian, C.; Zheng, M.; Zuo, W.; Zhang, S.; Zhang, Y.; Lin, C.W. A cross Transformer for image denoising. *Inf. Fusion* **2024**, *102*, 102043. [CrossRef]
4. Zhang, D.; Zhou, F.; Jiang, Y.; Fu, Z. Mm-bsn: Self-supervised image denoising for real-world with multi-mask based on blind-spot network. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Vancouver, BC, Canada, 17–24 June 2023; pp. 4188–4197.
5. Al-Mekhlafi, H.; Liu, S. Single image super-resolution: A comprehensive review and recent insight. *Front. Comput. Sci.* **2024**, *18*, 181702. [CrossRef]
6. Yue, Z.; Wang, J.; Loy, C.C. Resshift: Efficient diffusion model for image super-resolution by residual shifting. In *Advances in Neural Information Processing Systems*; MIT Press: Cambridge, MA, USA, 2024; Volume 36.
7. Zhang, W.; Zhao, W.; Li, J.; Zhuang, P.; Sun, H.; Xu, Y.; Li, C. CVANet: Cascaded visual attention network for single image super-resolution. *Neural Netw.* **2024**, *170*, 622–634. [CrossRef] [PubMed]
8. Jin, Z.; Hu, X.; Zhu, L.; Song, L.; Yuan, L.; Yu, L. IDRNet: Intervention-driven relation network for semantic segmentation. In *Advances in Neural Information Processing Systems*; MIT Press: Cambridge, MA, USA, 2024; Volume 36.
9. Nguyen, Q.; Vu, T.; Tran, A.; Nguyen, K. Dataset diffusion: Diffusion-based synthetic data generation for pixel-level semantic segmentation. In *Advances in Neural Information Processing Systems*; MIT Press: Cambridge, MA, USA, 2024; Volume 36.
10. Zhang, F.; Zhou, T.; Li, B.; He, H.; Ma, C.; Zhang, T.; Yao, J.; Zhang, Y.; Wang, Y. Uncovering prototypical knowledge for weakly open-vocabulary semantic segmentation. In *Advances in Neural Information Processing Systems*; MIT Press: Cambridge, MA, USA, 2024; Volume 36.
11. Kaur, J.; Singh, W. A systematic review of object detection from images using deep learning. *Multimed. Tools Appl.* **2024**, *83*, 12253–12338. [CrossRef]
12. Xie, C.; Zhang, Z.; Wu, Y.; Zhu, F.; Zhao, R.; Liang, S. Described Object Detection: Liberating Object Detection with Flexible Expressions. In *Advances in Neural Information Processing Systems*; MIT Press: Cambridge, MA, USA, 2024; Volume 36.
13. Deepshikha, K.; Yelleni, S.H.; Srijith, P.K.; Mohan, C.K. Monte Carlo DropBlock for modeling uncertainty in object detection. *Pattern Recognit.* **2024**, *146*, 110003.
14. Dabov, K.; Foi, A.; Katkovnik, V.; Egiazarian, K. Image denoising by sparse 3-D transform-domain collaborative filtering. *IEEE Trans. Image Process.* **2007**, *16*, 2080–2095. [CrossRef] [PubMed]
15. Zhang, K.; Zuo, W.; Chen, Y.; Meng, D.; Zhang, L. Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising. *IEEE Trans. Image Process.* **2017**, *26*, 3142–3155. [CrossRef] [PubMed]
16. Dong, W.; Zhang, L.; Shi, G.; Li, X. Nonlocally centralized sparse representation for image restoration. *IEEE Trans. Image Process.* **2012**, *22*, 1620–1630. [CrossRef]
17. Buades, A.; Coll, B.; Morel, J.M. A non-local algorithm for image denoising. In Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), San Diego, CA, USA, 20–25 June 2005; Volume 2, pp. 60–65.
18. Gu, S.; Zhang, L.; Zuo, W.; Feng, X. Weighted nuclear norm minimization with application to image denoising. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014; pp. 2862–2869.

19. Zamir, S.W.; Arora, A.; Khan, S.; Hayat, M.; Khan, F.S.; Yang, M.H. Restormer: Efficient transformer for high-resolution image restoration. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–24 June 2022; pp. 5728–5739.
20. Guo, Y.; Gao, Y.; Hu, B.; Qian, X.; Liang, D. CMID: Crossmodal Image Denoising via Pixel-Wise Deep Reinforcement Learning. *Sensors* **2023**, *24*, 42. [CrossRef] [PubMed]
21. Lehtinen, J.; Munkberg, J.; Hasselgren, J.; Laine, S.; Karras, T.; Aittala, M.; Aila, T. Noise2Noise: Learning image restoration without clean data. *arXiv* **2018**, arXiv:1803.04189.
22. Ulyanov, D.; Vedaldi, A.; Lempitsky, V. Deep image prior. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 9446–9454.
23. Batson, J.; Royer, L. Noise2self: Blind denoising by self-supervision. In Proceedings of the International Conference on Machine Learning. PMLR, Long Beach, CA, USA, 9–15 June 2019; pp. 524–533.
24. Krull, A.; Buchholz, T.O.; Jug, F. Noise2void-learning denoising from single noisy images. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 2129–2137.
25. Han, Y.J.; Yu, H.J. SS-BSN: Attentive blind-spot network for self-supervised denoising with nonlocal self-similarity. *arXiv* **2023**, arXiv:2305.09890.
26. Huang, T.; Li, S.; Jia, X.; Lu, H.; Liu, J. Neighbor2neighbor: Self-supervised denoising from single noisy images. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 20–25 June 2021; pp. 14781–14790.
27. Zhang, K.; Zuo, W.; Zhang, L. FFDNet: Toward a fast and flexible solution for CNN-based image denoising. *IEEE Trans. Image Process.* **2018**, *27*, 4608–4622. [CrossRef] [PubMed]
28. Lu, W.; Onofrey, J.A.; Lu, Y.; Shi, L.; Ma, T.; Liu, Y.; Liu, C. An investigation of quantitative accuracy for deep learning based denoising in oncological PET. *Phys. Med. Biol.* **2019**, *64*, 165019. [CrossRef] [PubMed]
29. Pinzón-Arenas, J.O.; Jiménez-Moreno, R.; Pachón-Suescún, C.G. ResSeg: Residual encoder-decoder convolutional neural network for food segmentation. *Int. J. Electr. Comput. Eng.* **2020**, *10*, 1017–1026. [CrossRef]
30. Zhang, K.; Zuo, W.; Gu, S.; Zhang, L. Learning deep CNN denoiser prior for image restoration. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 3929–3938.
31. Yue, Z.; Yong, H.; Zhao, Q.; Meng, D.; Zhang, L. Variational denoising network: Toward blind noise modeling and removal. *Adv. Neural Inf. Process. Syst.* **2019**, *32*, 1690–1701.
32. Liang, J.; Cao, J.; Sun, G.; Zhang, K.; Van Gool, L.; Timofte, R. Swinir: Image restoration using swin transformer. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, BC, Canada, 10–17 October 2021; pp. 1833–1844.
33. Wu, W.; Lv, G.; Liao, S.; Zhang, Y. FEUNet: A flexible and effective U-shaped network for image denoising. *Signal Image Video Process.* **2023**, *17*, 2545–2553. [CrossRef]
34. Zhang, Y.; Li, K.; Li, K.; Sun, G.; Kong, Y.; Fu, Y. Accurate and fast image denoising via attention guided scaling. *IEEE Trans. Image Process.* **2021**, *30*, 6255–6265. [CrossRef] [PubMed]
35. Abdelhamed, A.; Lin, S.; Brown, M.S. A high-quality denoising dataset for smartphone cameras. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 1692–1700.
36. Plotz, T.; Roth, S. Benchmarking denoising algorithms with real photographs. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 1586–1595.
37. Arbelaez, P.; Maire, M.; Fowlkes, C.; Malik, J. Contour detection and hierarchical image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2010**, *33*, 898–916. [CrossRef] [PubMed]
38. Quan, Y.; Chen, M.; Pang, T.; Ji, H. Self2self with dropout: Learning self-supervised denoising from single image. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 1890–1898.
39. Guo, S.; Yan, Z.; Zhang, K.; Zuo, W.; Zhang, L. Toward convolutional blind denoising of real photographs. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 1712–1722.
40. Pang, T.; Zheng, H.; Quan, Y.; Ji, H. Recorrupted-to-recorrupted: Unsupervised deep learning for image denoising. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 19–25 June 2021; pp. 2043–2052.
41. Anwar, S.; Barnes, N. Real image denoising with feature attention. In Proceedings of the IEEE/CVF international Conference on Computer Vision, Seoul, Republic of Korea, 27 October–2 November 2019; pp. 3155–3164.