IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 13 August 2024, accepted 24 August 2024, date of publication 28 August 2024, date of current version 10 September 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3450842

RESEARCH ARTICLE

Single Image Denoising via a New Lightweight Learning-Based Model

SADJAD REZVANI[1], FATEMEH SOLEYMANI SIAHKAR[2], YASIN REZVANI[1], ABDORREZA ALAVI GHARAHBAGH[3], (Member, IEEE), AND VAHID ABOLGHASEMI[4], (Senior Member, IEEE)

[1]Faculty of Computer Engineering, Shahrood University of Technology, Shahrud 3614915374, Iran
[2]Faculty of Electrical Engineering, Shahid Beheshti University, Tehran 1983969411, Iran
[3]Faculty of Electrical and Computer Engineering, Islamic Azad University, Shahrud 1598995712, Iran
[4]School of Computer Science and Electronic Engineering, University of Essex, CO4 3SQ Colchester, U.K.

Corresponding author: Vahid Abolghasemi (v.abolghasemi@essex.ac.uk)

ABSTRACT Restoring a high-quality image from a noisy version poses a significant challenge in computer vision, particularly in today's context where high-resolution and large-sized images are prevalent. As such, fast and efficient techniques are required to address noise reduction in such images effectively. Deep CNN-based image-denoising algorithms have gained popularity due to the rapid growth of deep learning and convolutional neural networks (CNNs). However, many existing deep learning models require paired clean/noisy images for training, limiting their utility in real-world denoising scenarios. In this paper, we propose a fast residual denoising framework (FRDF) designed based on zero-shot learning to address this issue. The FRDF first employs a novel downsampling technique to generate six different images from the noisy input, which are then fed into a lightweight residual network with 23K parameters. The network effectively utilizes a hybrid loss function, including residual, regularization, and guidance losses, to produce high-quality denoised images. Our innovative downsampling approach incorporates zero-shot learning principles, enabling our framework to generalize to unseen noise types and adapt to diverse noise conditions without needing labelled data. Extensive experiments conducted on synthetic and real images confirm the superiority of our proposed approach over existing dataset-free methods. Extensive experiments conducted on synthetic and real images show that our method achieves up to 2 dB improvements in PSNR on the McMaster and Kodak24 datasets. This renders our approach applicable in scenarios with limited data availability and computational resources.

INDEX TERMS Deep learning, image denoising, self-supervision, downsampling, zero-shot learning.

I. INTRODUCTION

Image denoising is a critical area of research in low-level vision and image processing, to recover high-quality images from their noisy pairs [1]. This task presents a significant challenge due to the inherent difficulty in distinguishing fine textures and details from the noise. Noise interference during image acquisition and transmission is an unavoidable factor that can severely impact the visual quality of images, underscoring the importance of noise removal for various image processing tasks [2], [3], [4].

The associate editor coordinating the review of this manuscript and approving it for publication was Yizhang Jiang.

Noise in different images can originate from several sources, including sensor limitations, environmental conditions, and transmission artifacts [5]. Common types of noise include Gaussian noise, which is additive and follows a Gaussian distribution, and Poisson noise, which arises from photon counting processes and is commonly encountered in low-light imaging conditions [6]. The presence of noise can degrade image quality, reduce the peak signal-to-noise ratio (PSNR), and hinder the performance of subsequent tasks such as object detection, segmentation, and recognition [7], [8], [9].

Denoising methods aim to obtain a clean image (p) by effectively eliminating the noise component (n) from its noisy observation (y), as represented by $p = y - n$.

Traditional denoising techniques include spatial filters such as Gaussian smoothing, Median filtering and Wiener filter, which operate on pixel intensities or local neighbourhoods to reduce noise while preserving image details [10], [11]. However, traditional methods, such as BM3D [12], may not always be effective in handling complex noise patterns or preserving fine details in high-resolution images [13].

Recently, deep learning-based methods have been demonstrated to be powerful tools for image denoising, leveraging the capacity of CNNs to learn hierarchical representations directly from data [14]. Deep CNN-based denoising algorithms have demonstrated superior performance compared to traditional methods, particularly in scenarios with complex noise characteristics and limited availability of labelled data [15], [16]. By training on large datasets of noisy-clean image pairs, deep learning models can effectively learn to distinguish between signal and noise, enabling them to generalize well to unseen data and noise types [17], [18].

Despite the success of deep learning approaches, challenges remain in adapting these models to real-world denoising scenarios such as medical imaging [19], surveillance footage [20], or low-light photography [21]. Many existing deep learning models require paired clean and noisy images for training, which may be impractical or costly to obtain in certain applications such as medical imaging [22], [23], remote sensing [24], or historical document restoration [25]. Moreover, deep learning models trained on specific noise distributions may struggle to generalize to unseen noise types or adapt to varying noise conditions encountered in practice [26], [27].

In light of these challenges, there is a growing demand for self-supervised learning methods that operate without needing a clean image, meaning they can generate a denoised image solely based on a set of noisy images. For instance, in the Neighbor2Neighbor [27] method, multiple samples are generated from a single noisy image, and during training, these images are used as input and output for the network. In self-supervised models, the denoising performance is lower compared to models trained on clean/noisy image pairs. However, self-supervised models are more suited for use in real-world image-denoising applications because they don't need to prepare a large number of image pairs. Recently, there has been a lot of attention given to a specific type of self-supervised learning model called the zero-shot method. These models do not require a specific dataset during training and only use one noisy image for denoising. However, many of these models face challenges, including high computational costs [28], limited noise types [29], and low accuracy [30]. Our goal in this research is to develop a model that achieves good accuracy with fewer parameters and does not rely on a specific type or noise level. To do this, we generated multiple image samples from a noisy image at different stages without using a clean image at any stage. In summary, the proposed method utilizes a lightweight model with only about 23,000 parameters. Therefore, compared to all current models, the proposed method has a low execution time and can be easily run on a CPU. Additionally, due to the use of many down-sampled images from the noisy image and their application during the calculation of network hybrid loss, the denoised output image also has relatively good quality.

Our three key contributions are as follows:

1) We introduce a new downsampling strategy that generates multiple sub-sampled noisy images from a single noisy input. This approach allows our framework to leverage zero-shot learning principles, enabling it to generalize to unseen noise types and adapt to diverse noise conditions without the need for labelled data.
2) We design an effective hybrid loss function that includes residual loss, regularization loss, and guidance loss. This innovative loss function helps the network to better capture and reduce noise, while preserving essential image details, thereby enhancing the overall denoising performance.
3) We conduct extensive experiments on both synthetic and real image denoising tasks using various noise types. Our results with three diverse well-known datasets illustrate superior performance compared to state-of-the-art.

II. RELATED WORK

Image-denoising models encompass three distinct categories. The subsequent discussion aims to provide a comprehensive overview of each category, with a specific focus on single image-based denoising methods.

A. DENOISING MODELS TRAINED ON CLEAN/NOISY IMAGE PAIRS

The prevailing dominance of CNNs in contemporary image processing tasks is undeniable. This popularity is underscored by their ability to achieve state-of-the-art performance, where networks are trained to map noisy images to clean ones in an end-to-end manner. In real-world applications, when the noise level in test images varies substantially from the training noise level, the denoising effectiveness of these models declines. This limitation is attributable to the nature of Deep Neural Networks (DNNs), which heavily rely on the training data for generalization. To address these issues, Zhang et al. introduced the Fast and Flexible Denoising Network (FFDNet) [31]. FFDNet distinguishes itself by incorporating an adjustable noise level map as the model input. FFDNet is notable for incorporating an adjustable noise level map as part of its input, enabling it to manage a wide array of noise levels effectively. This feature allows FFDNet to adapt to spatially varying noise, making it capable of handling changeable or non-uniform noise maps.

The Dual-branch Residual Attention Network (DRANet) [32], addresses challenges faced by traditional deep convolutional neural networks (CNNs) when dealing with spatially variant noise. Unlike previous approaches that increase network depth, DRANet improves performance by expanding network width and incorporating attention-guided feature learning [33], [34]. The model features two parallel branches

with Residual Attention Blocks (RABs) and Hybrid Dilated Residual Attention Blocks (HDRABs) designed to capture complementary features and filter out unimportant ones.

Residual Wavelet-Conditioned Convolutional Autoencoder (Res-WCAE) [35] with Kullback-Leibler divergence regularization has been proposed specifically for fingerprint image denoising. Res-WCAE integrates two encoders—an image encoder and a wavelet encoder—and a single decoder, utilizing residual connections to preserve spatial details. The wavelet encoder enhances the model by processing both approximation and detail subimages in the wavelet-transform domain.

B. DENOISING MODELS TRAINED ON NOISY IMAGES

While deep neural network denoising techniques, trained on sets of clean and noisy images, have demonstrated exceptional performance in numerous denoising applications, the acquisition of clean ground truth images is often impractical or unattainable due to cost constraints in real-world scenarios. In response to this challenge, researchers have begun exploring methods to address image denoising without relying on clean data. Consequently, several DNN image denoising approaches have been introduced, focusing on training with pairs of noisy images or multiple noisy images.

The Noise2Noise [36] method, introduced by Lehtinen et al. in 2018, demonstrates impressive performance by utilizing two noisy images of a static scene for training, without requiring corresponding ground truth images. It is interesting to note that with zero-mean noise, training a network can map a noisy image to another one of the same scene with a performance similar to the ground truth. Although acquiring pairs of noisy images of identical scenes could be difficult in real-world situations, Noise2Noise has inspired further investigation into self-supervised methods.

Context-aware denoiser [37] leverages a dual-branch structure, incorporating global and local feature extraction through Context-aware Denoise Transformer (CADT) units. Additionally, a Secondary Noise Extractor (SNE) block is introduced for secondary global noise extraction, enabling two-stage denoising.

LG-BPN [38] employs a self-supervised training approach, enabling it to learn from unlabeled data without the need for paired noisy-clean image pairs. By leveraging self-supervised training, LG-BPN can effectively exploit the spatial correlation statistics of real-world noise and model long-range dependencies within the image.

C. DENOISING MODELS BASED ON VISUAL TRANSFORMERS

In recent years, there has been notable progress in the development of image recovery techniques using visual transformers. Image denoising models based on Vision Transformers (ViT) exhibit the ability to capture extensive dependencies among image pixels through global self-attention, leading to remarkable performance improvements. Transformers rely solely on the attention mechanism, eliminating the need for convolution operations. Moreover, in comparison to numerous deep convolutional neural network denoising models, transform-based models demand less training time while delivering competitive and promising performances.

DenSformer [39] integrates both Transformer and convolutional layers to capture local and global features, significantly improving denoising performance. It comprises a preprocessing module, a local-global feature extraction module with Sformer blocks, and a reconstruction module.

The Lightweight Image Denoising Transformer (LIDFormer) [40] incorporates Triple Multi-Dconv Head Transposed Attention (TMDTA) and Discrete Wavelet Transform (DWT). LIDFormer reduces computational complexity by transforming the input image into a low-frequency space using DWT. This transformation maintains performance while minimizing the computational burden.

TransCT-net [41] introduced transformers for high-frequency and low-frequency inference but still relied on convolutional operations. Addressing this gap, they propose a convolution-free Transformer Encoder-decoder Dilation network(TED-net) that leverages Token-to-Token (T2T) vision transformers for LDCT denoising.

D. DENOISING MODELS TRAINED ON SINGLE NOISY IMAGE

As a notable advancement in denoising models trained on noisy images, there has been a growing interest in models relying solely on a single noisy image in recent years [42]. The constraint of deep neural networks requiring a sufficient amount of sample data for effective training poses a limitation. However, zero-shot models, which do not necessitate the preparation of training image pairs, have emerged as a promising solution. This type of denoising model is particularly well-suited for practical denoising applications in real-world scenarios.

Ulyanov et al. introduced the deep image prior (DIP) [30], a denoising model utilizing untrained convolutional neural networks for image restoration. DIP involves adapting a generative neural network to map a random input to a specified degraded image, aiming to denoise the image through training with early stopping. Despite its straightforward approach, the performance of DIP is often unsatisfactory and can be sensitive to the iteration number, making it challenging to determine the optimal value for effective denoising.

Noise2Fast [43] is an efficient method for referenceless denoising in single images. The acceleration of this method is achieved through the training process involving four downsampling images. However, Noise2Fast drops pixel values during downsampling, leading to a degradation in the quality of the produced images.

Traditional non-learning-based methods like BM3D [12] and WNNM [44] perform well with Gaussian and Poisson

noise, respectively, and rely on the input of the noise level for optimal functionality.

In summary, image denoising models fall into four main categories, each with its strengths and limitations. Models trained on pairs of clean and noisy images, like FFDNet, perform well but struggle with generalization when noise levels in test images differ significantly from those in the training set. Models trained on noisy images, such as Noise2Noise, bypass the need for clean data but may encounter difficulties in obtaining pairs of noisy images for training. Transformer-based models, exemplified by Uformer, offer promising performance but may face challenges with deep architectures and computational costs. Models trained on a single noisy image, like DIP and Noise2Fast, address practical denoising scenarios but may suffer from performance issues and sensitivity to hyperparameters. Our approach aims to mitigate these challenges and enhance the effectiveness of single-noisy-image denoising.

III. METHODOLOGY

In the realm of image denoising, various approaches are utilized to mitigate the diverse effects of noise in images. These methods can be categorized into supervised, self-supervised, and zero-shot denoising approaches.

Supervised methods typically employ neural networks represented as N_θ, which map a noisy image y to an estimate $N_\theta(y)$ of the clean image p. These supervised denoising methods are conventionally trained on pairs of clean images and their corresponding noisy measurements $y = p+n$ where n represents the noise.

Self-supervised methods involve training neural networks on various noisy observations of the same clean image. In Noise2Noise [36], a network is trained to effectively map noisy images to one another. Acquiring sets of noisy images depicting identical static scenes can pose difficulties. For instance, the subject being photographed may exhibit movement or lighting conditions could undergo rapid changes. Neighbor2Neighbor [27] is a self-supervised approach that builds upon Noise2Noise by enabling training solely with individual noisy images. This is achieved by extracting subsamples from a noisy image to generate pairs of noisy images and training on many images.

Zero-shot methods, often referred to as dataset-free approaches, differ fundamentally from self-supervised and supervised methods by not requiring a large-scale dataset for training. Instead, they operate on a single image, meaning that the training and inference phases occur simultaneously. By integrating the training and inference processes, zero-shot methods can quickly adapt to new images without the need for retraining on a separate dataset, making them highly efficient and versatile for real-world applications. ZS-N2N is a zero-shot denoising network that produces a pair of noise maps [45] from a noisy image and employs these maps for denoising purposes.

Our research builds upon the concepts introduced by ZS-N2N [45] and Neighbor2Neighbor [27] by introducing a novel approach that allows training using only a single noisy image.

There are discernible differences between the pixel characteristics of a clean, natural image and those of random noise. By intentionally manipulating the pixel lattice and modifying the relationships between neighbouring pixels, we effectively denoise images in a self-supervised manner. Specifically, the neighbouring pixels in a clean image typically display strong correlation and similarity in their values, whereas the pixels representing noise lack organization and operate independently [45]. By generating various downsampled versions of a single noisy image where noise lacks correlation across different positions within subsets, we utilize these downsampled images as both input and output for training a neural network. The rationale behind utilizing a downsampled pair of noisy images lies in the inherent characteristics of clean and noisy pixels. In clean images, neighbouring pixels exhibit strong correlation and similarity in values, whereas noise pixels lack structure and operate independently. Consequently, the downsampled pair retains similar signal characteristics but independent noise patterns. This allows the pair to approximate two noisy observations of the same scene, with one serving as the input and the other as the target for denoising.

From a theoretical perspective, using noisy downsampled images as both input and output can achieve the same performance as a supervised approach where the input is noisy and the output is clean. Assume D_1 and D_2 are two downsampled images derived from a noisy image y, and p is a clean image. In our approach, the network $N(\theta)$ learns to map D_1 to D_2, while in the supervised approach, the network learns to map y to p. By proving Equation (1), our approach can reach the same performance as supervised methods if the dataset size is infinitely large. In practical scenarios, zero-shot approaches fall slightly short of supervised methods.

$$\arg\min_\theta \mathbb{E}\left[\|\mathcal{N}_\theta(D_1)-p\|_2^2\right] = \arg\min_\theta \mathbb{E}\left[\|\mathcal{N}_\theta(D_1)-D_2\|_2^2\right] \quad (1)$$

Proof of Equation (1) is provided in the supplementary material.

Our Fast Residual Denoising Framework (FRDF) relies on three main elements:

1) Down-sampling technique: this step generates subsample noisy images that are used to train our network.
2) Fast residual network: this network is designed to learn and extract noise from a single noisy input image.
3) Filter operator: utilising a Gaussian filter, this step is designed to guide the model to capture noise information and to retain low-frequency background consistency, especially in images with complex scenes.

A. FRAMEWORK OVERVIEW

In this section, we introduce FRDF, which is based on our zero-shot approach. Denoising models, leveraging deep neural networks, have become instrumental in achieving high

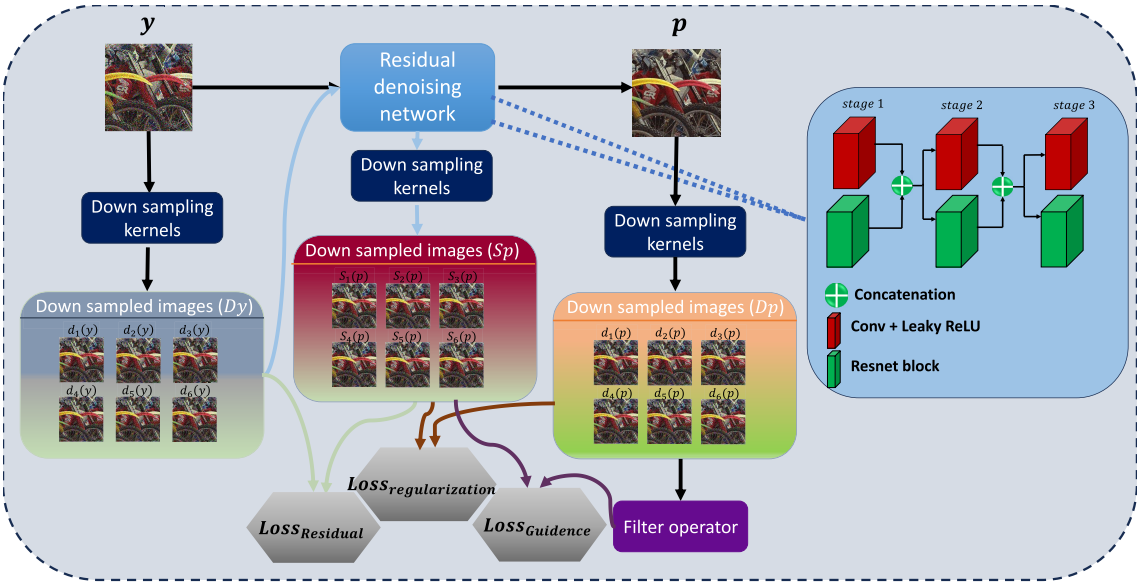

FIGURE 1. Architecture of the proposed lightweight zero-shot network.

accuracy across various applications. However, the training demands associated with these models, including the need for extensive datasets and significant computational resources, pose challenges. The denoising process in such models is notably time-consuming. In practical scenarios such as autonomous vehicles [46], surveillance systems [20] and live video streaming [47] where computational resources are often limited, there is a growing demand for real-time denoising methods compatible with both CPU and GPU platforms. Our proposed network distinguishes itself with a modest architecture, comprising around 23000 parameters. This lightweight design contrasts with deep networks, which typically boast millions of parameters. This approach takes a single noisy image as input and produces a denoised version of that input.

Fig.1 illustrates the workflow for the training process. Initially, down-sampled images (D_y) are generated from a single noisy image (y) using a down-sampling kernel (details in Section III-B). These down-sampled images are then fed into the residual denoising network. The output of this stage (S_p) is used to calculate our hybrid loss (details in Section III-C). Specifically, the residual loss is calculated using D_y and S_p, while the regularization loss is calculated using the generated down-sampled images from the network output (D_p) and S_p. Additionally, the guidance loss is calculated using S_p and the low-pass downsamplers (D_f), which are generated from D_p (details in Section III-C). The training process continues until the network converges, typically requiring 1.5K iterations. After training finishes, the network extracts noise from the input image. The denoised image can then be obtained by subtracting the output of the network (N_θ) from the input image y:

$$p = y - N_\theta(y) \tag{2}$$

Illustrated in Fig. 1, our residual network consists of three stages, each incorporating a ResNet block alongside a convolutional layer. The fundamental building block of our lightweight residual block comprises two convolution layers, conv1 and conv2, employing leaky ReLU activation functions with a negative slope of 0.2. This design choice allows for the passage of some negative values, enhancing the network's adaptability. The output of the second convolutional layer is combined with the residual connection. The network's operation begins with a single noisy image fed into the initial part. As the process unfolds, each subsequent section receives the concatenated output of the preceding part, strengthening the network's ability to capture fine textures.

To facilitate visual understanding of the network output, we present some exemplary outputs of the network in Fig. 2. As can be observed, Fig. 2(a) shows the noisy image (y) with Gaussian noise at a level of 75 and a PSNR of 12.28 dB. Fig. 2 (b) displays the output of the network before the start of training. As evident from Fig. 2 (c) and (d), the network extracts noise during training. Finally, the denoised image is obtained by subtracting the noise extracted from the network (i.e., Fig. 2 (d)) from the noisy image (Fig. 2 (a)), resulting in a PSNR of 19.45 dB.

B. DOWNSAMPLING TECHNIQUE

Our model needs to learn the denoising process directly from noisy images without explicit supervision from clean reference images. Nearby pixels in a clean image are highly correlated, reflecting the underlying structure of the scene, while noise patterns are unstructured and operate independently [45]. Therefore, we innovatively create three downsampler pairs $((d_1, d_2),(d_3, d_4),(d_5, d_6))$ to generate downsampled images that retain similar signal content but

FIGURE 2. Visual comparison of the proposed denoising process. (a) Noisy image. (b),(c) and (d) are network outputs in different epochs. (e) is the denoised image which is obtained by subtracting the network's output from the noisy image.

exhibit independent noise patterns. Our network N_θ, where θ represents network weights, is trained to learn mapping:

$$\begin{aligned} N_\theta(d_1) &\rightarrow d_2 \\ N_\theta(d_3) &\rightarrow d_5 \\ N_\theta(d_5) &\rightarrow d_6 \end{aligned} \tag{3}$$

Downsampling enables our model to exploit characteristics of noise present in the image. By generating downsampled versions of the noisy input, the model can focus on learning denoising patterns.

The key idea behind using downsampling to achieve independent noise patterns lies in the fact that when an image is downsampled, the spatial resolution is reduced, causing a redistribution of pixel values. Noise, being unstructured and independent, does not retain its spatial correlation after downsampling. Consequently, the noise patterns in downsampled versions of the image are less correlated with each other compared to the original noisy image.

To perform downsampling, initially, 6 fixed kernels with different dimensions are created. These kernels (k) are then convolved with the image y ($d_i = k_i \circledast y$) where $\circledast$ denotes the convolution operator. This operation is applied channel-wise and is performed on all channels of the image ($h \times w \times c$), where h is the height, w is the width, and c is the number of channels. Fig. 3 illustrates how the downsampling is performed for d_1 and d_2.

Table 1 provides detailed information about each pair of downsamplers. It is evident that in producing images for each pair, we utilized kernels where one generates diagonal pixels while the other generates non-diagonal pixels. This approach to kernel selection results in the production of non-overlapping downsamplers in each pair. This strategy ensures that the similarity of the downsampled images is maintained and effectively spreads out the noise across different pixels in the downsampled images, reducing the likelihood that noise in one downsampled image will correlate with noise in another. Each downsampled image thus represents a different, independently sampled subset of the original noisy data.

Furthermore, by ensuring that these kernels do not overlap in the regions they sample from the original image, we enhance the independence of the noise patterns. This means that the noise observed in one downsampled image will not influence the noise in another downsampled image, making the noise patterns truly independent.

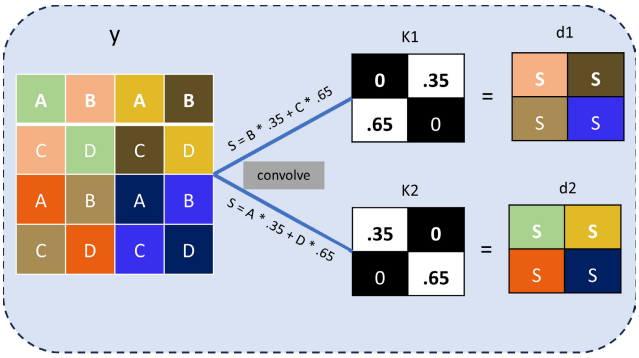

FIGURE 3. Visual exemplification of the different steps involved in generating two downsampled images (d_1, d_2). Here, y represents the input noisy image, and $K1$ and $K2$ are our kernels. By applying kernels with a stride of 2, as shown in Table 1, each [2, 2] pixel block convolves with the kernels. Notably, the values 0.25 and 0.65 are from our first kernel in the proposed method. As a result, our downsampled images have dimensions of $h/2 \times w/2 \times c$.

TABLE 1. The specifications of six distinct down samplers (denoted as d1 to d6). Each down sampler is characterized by its kernel configuration, responsiveness to dilation parameters, and suitability for a stride of 2.

Downsampler	Kernel	Dilation=1	Stride=2
d1	$\begin{bmatrix} 0 & 0.35 \\ 0.65 & 0 \end{bmatrix}$	×	✓
d2	$\begin{bmatrix} 0.35 & 0 \\ 0 & 0.65 \end{bmatrix}$		
d3	$\begin{bmatrix} 0 & 0.50 \\ 0.50 & 0 \end{bmatrix}$	✓	✓
d4	$\begin{bmatrix} 0.50 & 0 \\ 0 & 0.50 \end{bmatrix}$		
d5	$\begin{bmatrix} 0 & 0.5 & 0 \\ 0.5 & 0 & 0 \end{bmatrix}$	×	✓
d6	$\begin{bmatrix} 0.5 & 0 & 0 \\ 0 & 0.5 & 0 \end{bmatrix}$		

Algorithm 1 The Proposed FRDF Method

Input: A single noisy image y;
Denoising network N_θ;
while *not converged* **do**
  Generate down-samplers D:
  $D_y = \{(d_1(y), d_2(y)), (d_3(y), d_4(y)), (d_5(y), d_6(y))\}$;
  For the network input subset of D, ranging from 1 to 6: $N_\theta(d_i(y))$;
  Derive a denoised result of the network from down-sampler D:
  $\mathrm{Sp}_i = d_i(y) - N_\theta(d_i(y))$;
  Calculate $\mathrm{loss}_{\mathrm{res}}$:
  $\mathrm{loss}_{\mathrm{res}} = \frac{1}{6}\left(\sum_i |d(y)_i - \mathrm{Sp}_{i+1}|^2 + |d(y)_{i+1} - \mathrm{Sp}_i|^2\right)$;
  Derive a denoised result of the network from y:
  $p = y - N_\theta(y)$;
  Generate down-samplers according to the denoised input image p:
  $D_p = \{(d_1(p), d_2(p)), (d_3(p), d_4(p)), (d_5(p), d_6(p))\}$;
  Calculate $\mathrm{loss}_{\mathrm{reg}}$: $\mathrm{loss}_{\mathrm{reg}} = \frac{1}{6}\sum_{i=1}^{6} |d(p)_i - \mathrm{Sp}_i|^2$;
  Generate low-pass down-samplers by Gaussian filters GF:
  $D_f = GF_{\sigma=3,9,15}(D_p)$;
  Calculate $\mathrm{loss}_{\mathrm{guid}}$:
  $\mathrm{loss}_{\mathrm{guid}} = \frac{1}{6}\sum_{i=1}^{6} |d(f)_i - \mathrm{Sp}_i|^2$;
  Update the denoising network N_θ by minimizing the objective $\mathrm{loss}_{\mathrm{res}} + \mathrm{loss}_{\mathrm{reg}} + \mathrm{loss}_{\mathrm{guid}}$;

Our downsampling approach is crucial for our denoising method because it allows the network to learn to distinguish between the structured, correlated signal of the clean image and the unstructured, independent noise. By training on these downsampled images, the network can effectively learn to predict and remove noise, leveraging the fact that the noise patterns are independent and thus easier to identify and eliminate.

C. LOSS FUNCTION

In the domain of deep learning, the choice of loss function plays a critical role in guiding the optimization process and ultimately shaping the performance of the model. The loss function serves as a measure of dissimilarity between the predicted denoised image of the model and downsampled images. Through the minimization of this discrepancy during training, the model endeavours to converge towards accurate predictions of noise patterns and, consequently, produce a denoised image.

Inspired by recent advancements in self-supervised learning [45], [48], [49], we explore the efficacy of symmetric loss functions in the context of image denoising. Symmetric loss functions, as the name suggests, exhibit a balanced behaviour wherein the penalty incurred for predicting one value when the ground truth is another is symmetrically equivalent to predicting the second value when the ground truth is the first. Based on Equation (3), in our case, these values correspond to our three downsampler subsets.

The procedure of network optimization is shown in algorithm 1. First, we produce denoised downsampled images:

$$D_y = \{(d_1(y), d_2(y)), (d_3(y), d_4(y)), (d_5(y), d_6(y))\} \tag{4}$$

$$N_\theta(d_i(y)\ i = 1, 2, .., 6 \tag{5}$$

$$\mathrm{Sp}_i = d_i(y) - N_\theta(d_i(y)) \tag{6}$$

where D_y represents our six downsampled noisy images, and N_θ represents our network. θ denotes the network weights. Sp_i is the denoised version of the downsampled images. As shown in Fig. 2, the network N_θ extracts noise during training, and the denoised downsampled images (Sp_i) are obtained by subtracting the noise extracted from the noisy image (D_y).

For training the network, we employ two types of losses: residual loss and regularization loss. The residual loss, serving as our symmetric loss, is computed by evaluating the mean squared error (MSE) between each subset of downsampled noisy images (D_y) and their corresponding denoised counterparts (Sp_i):

$$\mathrm{loss}_{\mathrm{res}} = \frac{1}{6}\left(\sum_i |d(y)_i - \mathrm{Sp}_{i+1}|^2 + |d(y)_{i+1} - \mathrm{Sp}_i|^2\right) \quad i = 1, 3, 5 \tag{7}$$

We also add regularization loss which serves as a means to prevent overfitting and promote generalization. In this function, unlike the residual loss, we downsample the output of the network (D_p) (Equation (2)):

First, we generate 6 downsamples of the noisy input image, and then the regularization loss is calculated with the help of the denoised versions of the downsampled images (Sp_i):

$$p = y - N_\theta(y) \tag{8}$$

$$D_p = \{(d_1(p), d_2(p)), (d_3(p), d_4(p)), (d_5(p), d_6(p))\} \tag{9}$$

$$\mathrm{loss}_{\mathrm{reg}} = \frac{1}{6}\sum_{i=1}^{6} |d(p)_i - \mathrm{Sp}_i|^2 \tag{10}$$

Additionally, we incorporate a guidance loss to further reduce high-frequency noise by preserving essential low-frequency content. To do this, we first feed the D_p image set into the Gaussian low-pass filter. Then, the guidance loss is calculated using the sp_i images and the output of the filters D_f:

$$D_f = GF_{\sigma=3,9,15}(D_p) \tag{11}$$

$$\mathrm{loss}_{\mathrm{guid}} = \frac{1}{6}\sum_{i=1}^{6} |d(f)_i - \mathrm{Sp}_i|^2 \tag{12}$$

Finally, the total loss function can be obtained through summation of three functions described above:

$$\text{loss}_{\text{final}} = \beta_1 \text{loss}_{\text{res}} + \beta_2 \text{loss}_{\text{reg}} + \beta_3 \text{loss}_{\text{guid}} \tag{13}$$

In this study, we employ the gradient descent (GD) method to minimize $loss_{final}$, resulting in the optimization of the network parameters θ. In Equation (13), β denotes the scalar regularization parameter, which has been defined empirically through our experiments.

D. FILTER OPERATOR

In images with high levels of noise, the model may mistakenly interpret background variations as noise, leading to suboptimal denoising performance. To address this issue, we employ Gaussian low-pass filters. These filters serve a dual purpose: they preserve essential low-frequency content while effectively reducing high-frequency noise, resulting in smoother images that better represent the underlying scene [50]. By integrating these filters into our denoising process and utilizing the L2 norm for loss calculation, the model is directed to prioritize the extraction of crucial image features over the noise components. This ensures that the denoising process is guided to focus on retaining the structural integrity and essential details of the image, thus improving overall denoising performance.

In the filter operator phase, we apply Gaussian filtering to the downsampled images generated from the network output (p) of a single noisy image (y). The formula for the Gaussian filter expression is as follows:

$$f(x, y) = \frac{1}{2\pi^2} e^{\frac{x^2+y^2}{2\sigma^2}} \tag{14}$$

In the formula, $f(x, y)$ is the value of the Gaussian function at the coordinate (x, y). σ represents the standard deviation of the Gaussian distribution. Empirically, we use three different kernel sizes: 3, 9, and 15 (see Fig. 4). The values of these kernels are obtained using Gaussian filters and then applied to the image via 2D convolution.

IV. EXPERIMENTS

In this section, we initially outline the experimental settings. Subsequently, for assessing efficacy, the proposed approach undergoes a comparison with state-of-the-art denoising methods. Additionally, ablation studies are performed to scrutinize the effectiveness of the proposed method.

A. EXPERIMENTAL SETTINGS

We conduct a comprehensive comparison involving five state-of-the-art models, encompassing a variety of methodologies, including supervised, self-supervised, and zero-shot techniques. This comparative analysis aims to evaluate the proposed method against established denoising approaches.

With regard to supervised methods, we compare the proposed method with a modified version of UNet as the current state-of-the-art denoising algorithm. UNet, renowned for its exceptional performance, has become the standard choice

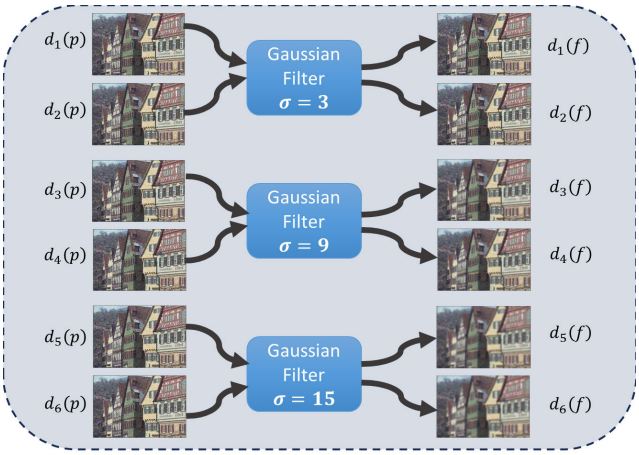

FIGURE 4. The process of generating filtered images with different kernel sizes (3, 9, and 15) from the noisy image $d_i(p)$, where $i = 1, 2, \ldots, 6$, to the filtered images $d_i(f)$.

in recent denoising studies [27], [29], [51], showcasing its effectiveness in capturing complex image features and preserving important details during the denoising process. We also compare the proposed method with Noise2Void [29], Neighbour2Neighbour [27], Noise2same [52] and Noisy-As-Clean [53] as self-supervised methods. These models exhibit proficiency in managing unknown noise distributions, enabling their evaluation on real-world datasets.

In terms of zero-shot methods, we compare our approach to deep learning algorithms such as DIP [30], Self2Self [28] and Self2Self+ [54]. Additionally, we also compare our method to the classical algorithm BM3D [12].

To thoroughly assess the effectiveness of the proposed method, we explore its performance under different noise conditions. Specifically, we consider Gaussian and Poisson noise, characterized by noise levels σ and λ, respectively. The fixed noise levels chosen for our experimentation are $\sigma = 10, 25$ and 50 for Gaussian noise, and $\lambda = 10, 25$ and 50 for Poisson noise. The σ values representing Gaussian noise align with pixel values within the range of $[0, 255]$, whereas the λ values associated with Poisson noise correspond to values within the interval $[0, 1]$.

TABLE 2. Summary of datasets used in our work.

Dataset	No. of images	Size of images	Real noise
Kodak24	24	768×512	$\times$
MacMaster18	18	500×500	$\times$
SSID	300	256×256	✓
CC	15	512×512	✓
FMDD	240	512×512	✓

As shown in Table 2, a detailed summary of the dataset characteristics, including the number of images, image sizes, and the presence of real noise, is provided. We employed both real-world and synthetic datasets to provide a comprehensive evaluation of the proposed method. The models are rigorously

TABLE 3. Average PSNR (dB) of the denoised images from the Top seven Models on Kodak24 summarizes the performance of various denoising methods, including Unet, Noise2Void, Neighbour2Neighbour, DIP, Self2Self, Noise2Same, Noise-As-Clean, BM3D, and our proposed method. The evaluation is based on different levels of Gaussian and Poisson noise.

Method	Gaussian noise			Poisson noise		
	$\sigma = 10$	$\sigma = 25$	$\sigma = 50$	$\lambda = 50$	$\lambda = 25$	$\lambda = 10$
Unet	33.5	28.3	25.67	29.45	27.3	26
Noise2Void	29.84	26.18	23.75	27.84	25.62	23.77
Neighbour2Neighbour	32.2	27.9	24.9	29.4	26.95	25.36
DIP	32.15	27.3	24.7	27.56	25.82	23.75
Self2Self	29.64	28.5	26.37	28.9	28.38	27.38
Noise2Same	32.15	29.96	**26.94**	30.45	28.98	**27.53**
Noise-As-Clean	33.35	29.47	26.43	30.08	27.21	25.14
BM3D	32.72	28.5	23.84	28.31	26.54	24.18
FRDF(ours)	**33.89**	**30.02**	26.27	**30.74**	**29.1**	27.2

TABLE 4. Average PSNR (dB) of the denoised images from the Top nine Models on McMaster18 dataset.

Method	Gaussian noise			Poisson noise		
	$\sigma = 10$	$\sigma = 25$	$\sigma = 50$	$\lambda = 50$	$\lambda = 25$	$\lambda = 10$
Unet	32.86	28.4	**25.96**	29.85	28.28	26.25
Noise2Void	30.5	26.47	23.8	28	25.78	23.45
Neighbour2Neighbour	33	28.14	25.18	29.7	27.6	25.68
DIP	33.72	27.25	22.9	28.5	27.46	24.9
Self2Self	30.9	29.1	25.13	30.11	**29.43**	**27.75**
Noise2Same	34.03	29.32	26.44	30.84	28.11	27.68
Noise-As-Clean	33.87	28.98	26.23	30.36	27.5	27.44
BM3D	33.48	28.51	23.5	27.34	24.75	22
FRDF(ours)	**34.67**	**29.55**	25.21	**31.15**	28.6	27.43

tested on two synthetic datasets: Kodak24 and McMaster18 [55]. Additionally, to simulate real-world scenarios, the models are tested on the SSID dataset [56], CC dataset [57] and FMDD [58], enhancing the robustness of the evaluation process.

B. COMPARISON WITH STATE-OF-THE-ART MODELS

1) QUANTITATIVE EVALUATION

In Tables 3 and 4, we present the denoising effectiveness of various methods. In Table 3, we test the methods on the Kodik24 dataset, while in Table 4, we test them on the MacMaster18 dataset. We trained Unet, Noise2Void, and Neighbour2Neighbour on 700 images using two different approaches. First, we trained them with an unknown noise level, and second, trained on that exact noise level between [10, 50]. Finally, we averaged the results from both approaches and recorded them in two tables. Based on the information from Tables 3 and 4, our approach consistently performs well for both Gaussian and Poisson noise, particularly at low noise levels. However, when considering Gaussian noise with a level of ($\sigma = 50$), Noise2same method achieves better results than our approach in both datasets.

As seen from the results of Tables 3 and 4, other zero-shot methods work well for specific noise levels and types. For example, BM3D exhibited lower scores in the Poisson noise distribution. In contrast to alternative zero-shot techniques, our method stands out as the sole dataset-independent denoising algorithm capable of delivering effective performance across diverse noise distributions and levels.

To assess the performance of our model on authentic noisy images, we performed testing on three different datasets, i.e., SSID, CC and Fluorescence Microscopy Denoising dataset (FMDD). SSID comprises images taken by various smartphone cameras, showcasing diverse lighting conditions and noise patterns. The CC contained 15 real noisy images of different ISO, i.e., 1,600, 3,200 and 6,400. Additionally, the FMDD contains real-world noisy fluorescence microscopy images with various noise levels, obtained using commercial two-photon, confocal, and widefield microscopes. The dataset includes raw images with high noise levels and images with lower noise levels created through image averaging. Particularly, images are averaged within sequences in the same field of view (FOV) of 50 images. By averaging S (where $S = 2, 4, 8, 16$) raw images, lower noise levels are achieved. Ground truth images are effectively obtained by image averaging, and noisy images are created at five different noise levels: raw (no averaging), and averaged with $S = 2, 4, 8$, and 16.

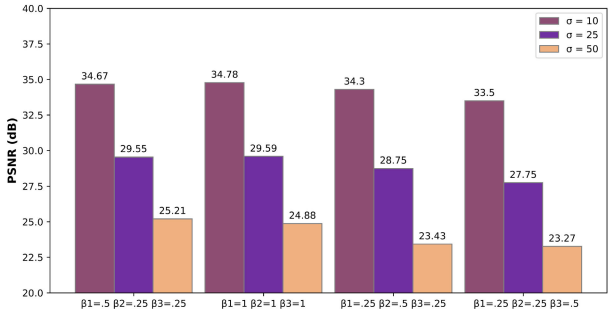

FIGURE 5. Quantitative comparison of β for our hybrid loss. The evaluation was conducted on MacMaster 18 dataset.

Firstly, we performed testing on a set of 300 images sourced from the SIDD dataset. Our result achieves the highest performance with a PSNR score of 34.43 dB, closely followed by DIP with 34.35 dB. However, BM3D shows a sharp drop in performance when compared to synthetic noise as the comparison is shown in Table 5.

For the CC dataset, our model achieves a PSNR of 35.74 dB, slightly lower than the highest score of 36.38 dB achieved by the S2S+ model. However, our model outperforms BM3D, which has a PSNR of 35.19 dB, and is close to S2S with a PSNR of 36.1 dB, indicating competitive performance on real noisy images across different ISO levels.

In the FMDD dataset, our model shows a strong performance with a PSNR of 35.89 dB. Although S2S+ achieves the highest PSNR of 36.14 dB, our model surpasses DIP, which has a PSNR of 33.18 dB, and BM3D, which scores 32.16 dB. This demonstrates the robustness of our approach in handling real noisy fluorescence microscopy images, maintaining high performance across diverse biological samples and microscopy techniques.

Moreover, it is noteworthy that the parameter number of S2S+ is around 1.2 million, while our model has significantly fewer parameters, approximately 23K, which is about 1/50 of the parameters of S2S+. This highlights the efficiency of our model in achieving comparable or superior performance with a substantially reduced computational complexity. According to the quantitative comparison in Fig. 5, to determine the weight of different loss in Equation (13), we set $\beta_1 = 0.5$, $\beta_2 = 0.25$, $\beta_3 = 0.25$.

TABLE 5. Average PSNR (dB) of the denoised images from the zero-shot models on real-world noise.

Dataset	FRDF(ours)	DIP	S2S	BM3D	S2S+
SIDD	**34.43**	34.35	33.85	28.05	34.11
CC	35.74	35.69	36.1	35.19	**36.38**
FMDD	35.89	33.18	35.72	32.16	**36.14**

2) QUALITATIVE EVALUATION

Fig. 6, 7 shows the denoising results of our method and four various models with 3 different noise levels 10, 25 and 50. We can find that the other zero-shot methods like BM3D generated blurred results while, our method has good robustness to both high-level and low-level noise to recover the actual texture and structures. Self-supervised method N2N does not perform well and exhibits a more than 2 dB drop in performance compared to our approach.

Our approach and Self2Self demonstrate comparable scores, slightly surpassing other baseline methods. Despite the similarity in performance metrics, a visual inspection of the denoised images reveals distinctions: Our method yields visually sharper images and retains slightly more details, whereas Self2Self produces relatively low-quality images. This distinction is particularly evident in images containing fine details, such as MRI images. In Fig. 8, showcasing a prostate image from the fastMRI [59] dataset, it is evident that our method successfully preserves all pixel values during downsampling. Moreover, our method exhibits superior image quality compared to the Self2Self method.

TABLE 6. FRDF's performance under different settings.

Setting number	Residual block	Parameters	PSNR (dB)
1	✓	264	27.697
	×	84	25.081
2	✓	6885	29.176
	×	747	28.165
3	✓	23109	29.584
	×	5955	29.151
4	✓	39333	29.589
	×	11163	29.408

C. ABLATION STUDY

In this section, to further evaluate the effectiveness of our zero-shot method FRDF, we conducted ablation studies using the McMaster18 dataset contaminated with Gaussian noise of $\sigma = 50$.

To illustrate the impact of network architecture on denoising performance, we specifically analyse two critical factors: the number of network stages and the presence of residual blocks. The quantitative comparison results are reported in Table 6. Our investigation aims to provide insights into achieving optimal image quality while striking a balance between performance and complexity.

Each stage in our network consists of a convolutional layer alongside a residual block (as depicted in Fig. 1). We evaluate the PSNR under various configurations to understand how these architectural choices affect denoising results. Including a residual block tends to yield higher PSNR values. These residual connections facilitate learning identity-like mappings, leading to improved image quality. For instance, consider the network with just one stage: with a residual block (264 parameters), the PSNR is 27.697 dB,

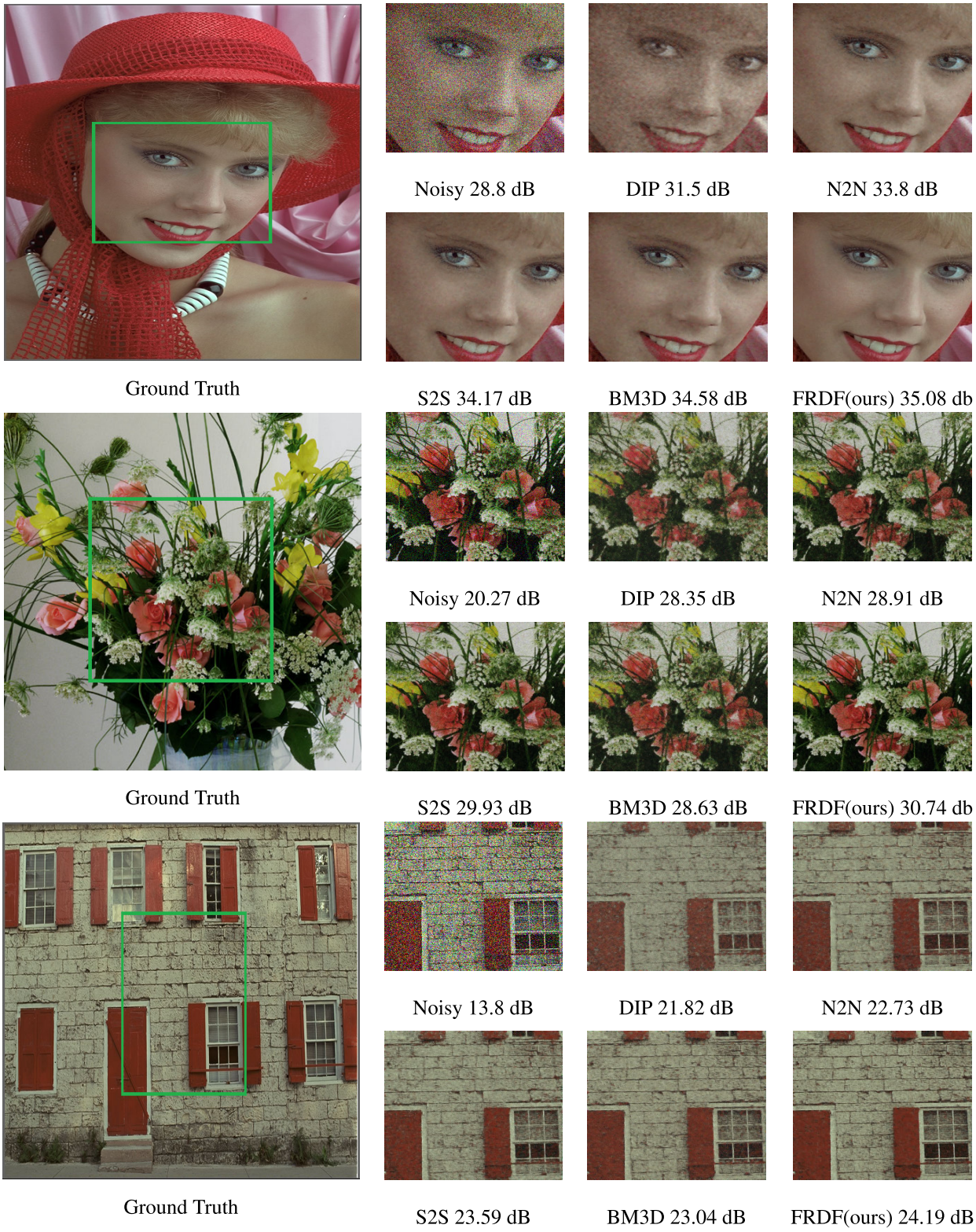

FIGURE 6. **Qualitative comparison of Gaussian denoising for different methods along with the corresponding PSNR. Upper row: $\sigma = 10$, middle row: $\sigma = 25$, lower row $\sigma = 50$.**

whereas without the block (84 parameters), the PSNR drops to 25.081 dB. Similar trends are observed across different stages.

As we increase the number of network parts, the PSNR generally improves. However, increasing the number of parameters (as seen in the network with four parts) does

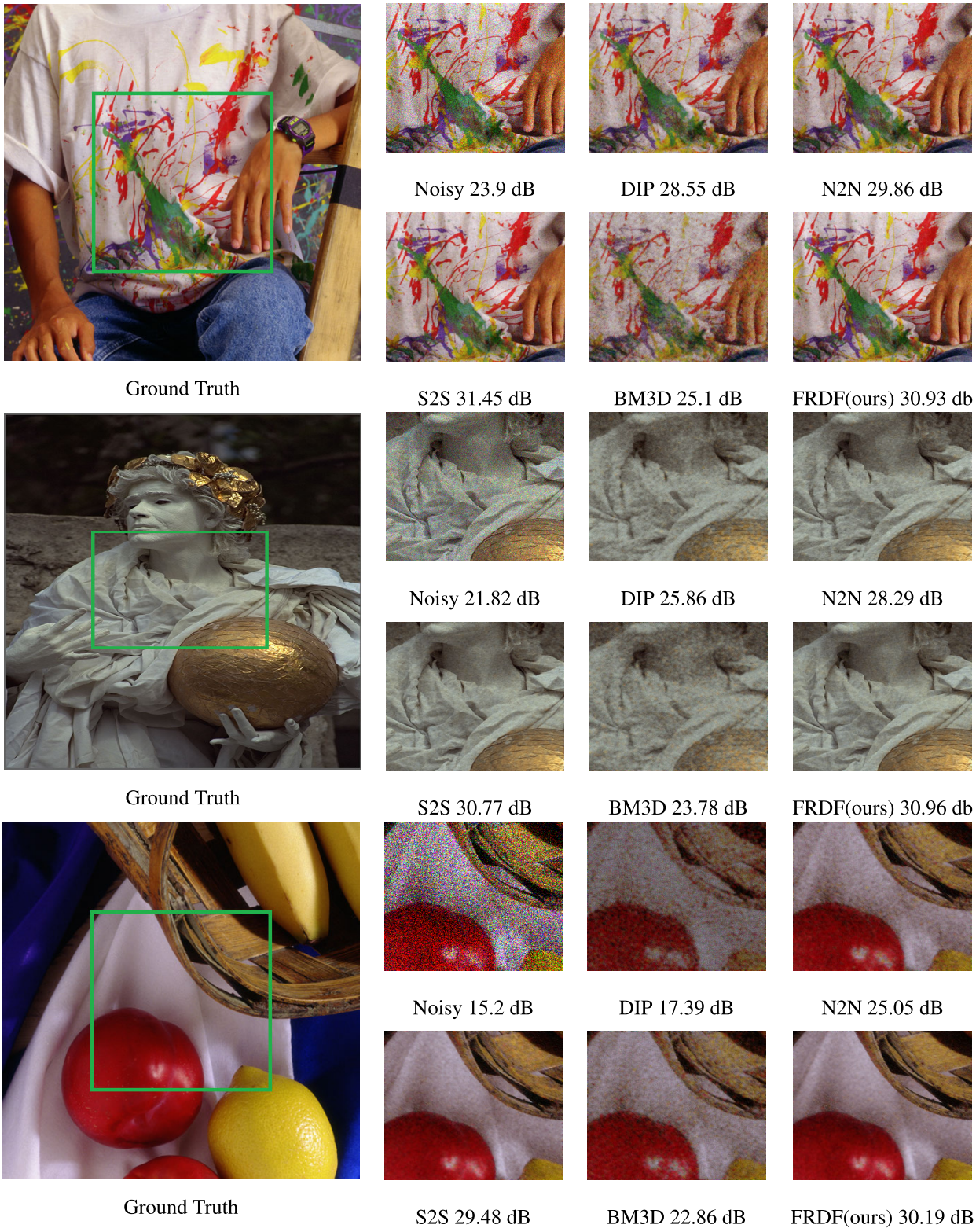

FIGURE 7. Qualitative comparison of Poisson denoising for different methods along with the corresponding PSNR. Upper row: $\lambda = 50$, middle row: $\lambda = 25$, lower row $\lambda = 10$.

not result in a substantial PSNR gain. Specifically, the PSNR for four network parts with a residual block (39,333 parameters) is only slightly higher than that for three parts (29.589 dB vs. 29.584 dB). Our findings emphasize the importance of thoughtful network design. Optimal denoising results can be achieved by strategically choosing the right

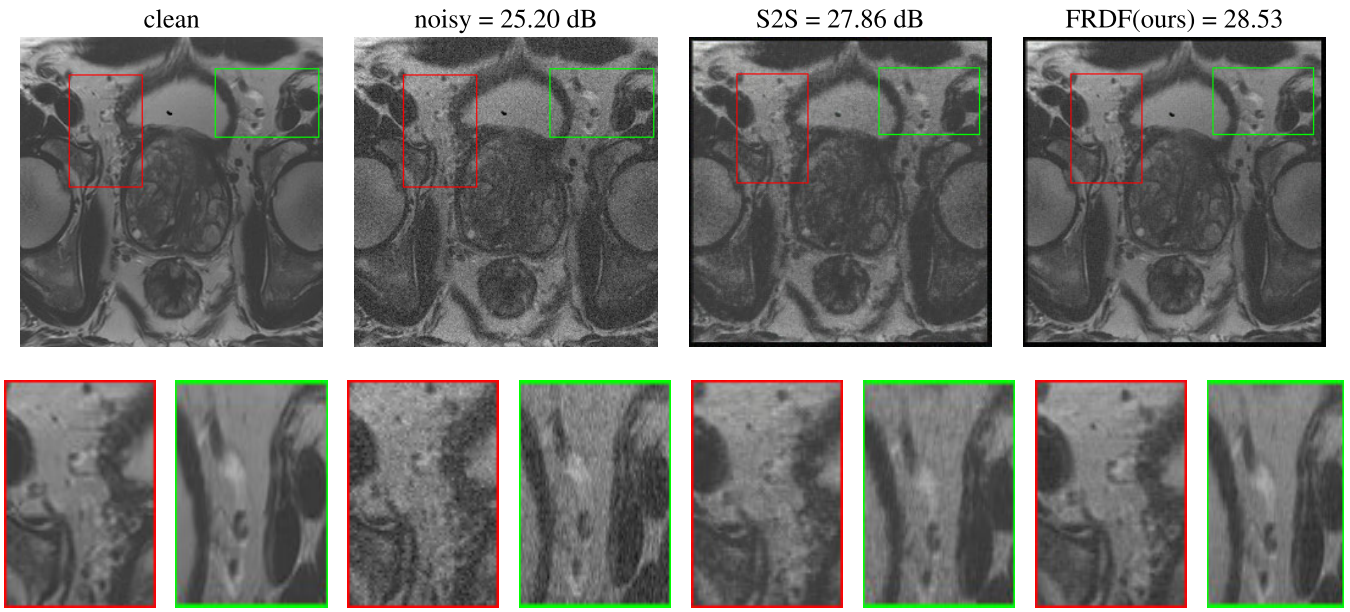

FIGURE 8. Visual comparison between FRDF and self2self for denoising Gaussian noise on a prostate image. Red and green regions are examples of enlarged regions for clearer inspection.

TABLE 7. Ablation study on the impact of different downsampling strategies using various kernel configurations. Each row represents the performance (PSNR) achieved using different combinations of downsampling kernels.

Number of kernels	Kernels								PSNR (dB)
#	$\begin{bmatrix}0.35 & 0\\ 0 & 0.65\end{bmatrix}$	$\begin{bmatrix}0 & 0.35\\ 0.65 & 0\end{bmatrix}$	$\begin{bmatrix}0.5 & 0\\ 0 & 0.5\end{bmatrix}$	$\begin{bmatrix}0 & 0.5\\ 0.5 & 0\end{bmatrix}$	$\begin{bmatrix}0 & 0.5 & 0\\ 0.5 & 0 & 0\end{bmatrix}$	$\begin{bmatrix}0.5 & 0 & 0\\ 0 & 0.5 & 0\end{bmatrix}$	$\begin{bmatrix}0.25 & 0\\ 0 & 0.75\end{bmatrix}$	$\begin{bmatrix}0 & 0.25\\ 0.75 & 0\end{bmatrix}$	#
2	✓	✓	×	×	×	×	×	×	24.53
2	×	×	✓	✓	×	×	×	×	24.63
2	×	×	×	×	✓	✓	×	×	24.58
2	×	×	×	×	×	×	✓	✓	23.7
4	✓	✓	✓	✓	×	×	×	×	25.88
4	✓	✓	×	×	×	×	✓	✓	25.63
4	×	×	✓	✓	×	×	✓	✓	25.78
4	✓	✓	×	×	✓	✓	×	×	25.84
6	✓	✓	✓	✓	✓	✓	×	×	25.96
6	×	×	✓	✓	✓	✓	✓	✓	25.81
6	✓	✓	✓	✓	×	×	✓	✓	25.27
8	✓	✓	✓	✓	✓	✓	✓	✓	25.91

combination of network stages and residual blocks while avoiding unnecessary parameter inflation.

We conducted an ablation study to evaluate the impact of different downsampling strategies on the performance of our method. Specifically, we analyzed how varying the number and combination of downsampling kernels affects the PSNR of the denoised images.

Table 7 presents the results of this ablation study. Each row in the table corresponds to a different set of downsampling kernels used during training, with the PSNR values indicating the denoising performance of the network. The columns under ''kernels'' display the specific kernel configurations used for downsampling. From the table, we observe that using different pairs of kernels shows varying levels of performance. As the number of downsampling kernels increases, the PSNR generally improves. For example, using four kernels instead of two leads to better performance (25.88 dB vs. 24.53 dB). This indicates that having more diverse downsampling kernels helps in better capturing the noise patterns and enhancing the denoising process. Furthermore, utilizing all eight downsampling kernels results in a PSNR of 25.91 dB, which is close to the highest performance observed in our experiments. This demonstrates that employing a comprehensive set of downsampling strategies can significantly enhance the network's ability to denoise images effectively.

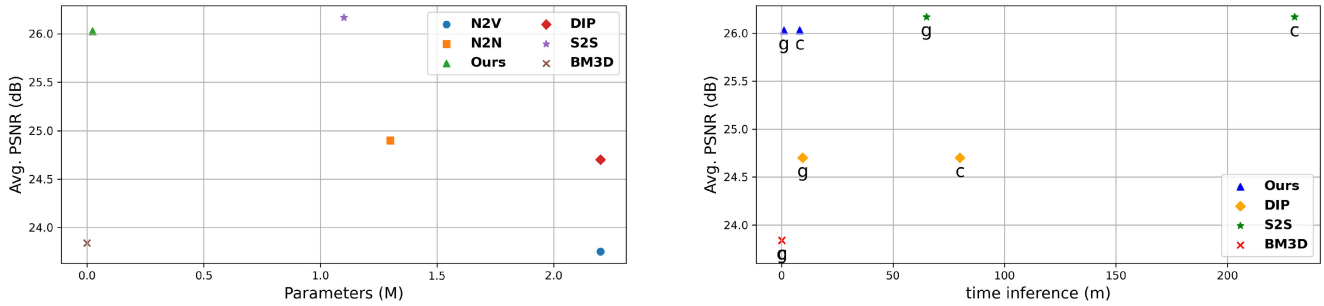

FIGURE 9. Parameters and inference time study on image denoising using a noise level of $\sigma = 50$ from the MacMaster test set. (a) Parameter size comparison and (b) Inference time comparison.

Furthermore, utilizing 6 downsampling kernels (row 9 of Table 7) results in a PSNR of 25.96 dB, which is the highest performance observed in our experiments. This demonstrates that employing selected downsampling kernels can enhance the network's ability to denoise images effectively.

To investigate computational complexity, we examine both denoising processing time and memory requirements, as indicated by the number of network parameters. Denoising time is assessed on both CPU and GPU platforms, with the GPU tested on an NVIDIA GeForce RTX 3060 laptop and the CPU on an Intel Core i7 11730 H. Fig. 9 illustrates the size of trainable parameters and the time required for denoising a single image from the MacMaster dataset.

In Fig. 9(a), despite our model's superior performance, its parameter size is smaller compared to most models, except for BM3D. The S2S model, while achieving slightly better results than our approach, has a parameter count 60 times larger. Moving on to Fig. 9(b), the runtime evaluation demonstrates that our proposed model achieves competitive speeds with outstanding performance, especially when deployed on a CPU. Notably, compared with S2S, our method exhibits significant speed improvements. On the CPU, our approach takes 8.2 minutes to denoise a single image with size 256 $\times$ 256, whereas S2S requires approximately 230 minutes. Only BM3D denoises faster than our approach, attributed to BM3D's non-utilization of deep networks and reliance on traditional methods.

V. CONCLUSION

We introduced an innovative zero-shot image denoising algorithm that operates without the need for training examples or information about the noise model or level. Our approach utilizes a lightweight three-stage network with just 23K parameters, facilitating efficient denoising within a relatively short time, even without GPU acceleration. The method demonstrates strong performance on both simulated noise and actual camera noise in real-world scenarios. Compared to existing dataset-free methods, our approach strikes a favourable balance between denoising quality and computational efficiency.

Our method employing an innovative downsampling technique to create pairs of downsampled images, allowing our network to learn effective denoising mappings without the need for clean reference images. Additionally, we introduce a hybrid loss function that combines residual loss, regularization loss, and guidance loss, which enhances the model's ability to differentiate between noise and important image features, improving the overall denoising performance.

While our method shows promising results, there are still some limitations. One area for improvement is the reduction of network parameters and the acceleration of the denoising process. This could be achieved by incorporating a variety of filters, such as bilateral filters, which might enhance the denoising performance while reducing the computational load. Additionally, further research could explore the optimization of the loss function to improve the model's ability to distinguish between noise and essential image features. By addressing these limitations, future work can enhance the robustness and efficiency of zero-shot image denoising methods.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT

The source code implementing the FRDF model and other relevant code utilized in this research is openly accessible and available at [https://github.com/sadjadrz/Zero-shot-image-denoising-based-on-downsampling].

REFERENCES

[1] M. Mafi, H. Martin, M. Cabrerizo, J. Andrian, A. Barreto, and M. Adjouadi, "A comprehensive survey on impulse and Gaussian denoising filters for digital images," *Signal Process.*, vol. 157, pp. 236–260, Apr. 2019.

[2] F. Liu, L. Jiao, and X. Tang, "Task-oriented GAN for PolSAR image classification and clustering," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 30, no. 9, pp. 2707–2719, Sep. 2019.

[3] J. Liu, Y. Wang, Y. Li, J. Fu, J. Li, and H. Lu, "Collaborative deconvolutional neural networks for joint depth estimation and semantic segmentation," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 11, pp. 5655–5666, Nov. 2018.

[4] T. Wu, B. Li, Y. Luo, Y. Wang, C. Xiao, T. Liu, J. Yang, W. An, and Y. Guo, "MTU-Net: Multilevel TransUNet for space-based infrared tiny ship detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5601015.

[5] Z. Han, H. Shangguan, X. Zhang, X. Cui, and Y. Wang, "A coarse-to-fine multi-scale feature hybrid low-dose CT denoising network," *Signal Process., Image Commun.*, vol. 118, Oct. 2023, Art. no. 117009.

[6] K. Sun, T.-H. Tran, R. Krawtschenko, and S. Simon, "Multi-frame super-resolution reconstruction based on mixed Poisson–Gaussian noise," *Signal Process., Image Commun.*, vol. 82, Mar. 2020, Art. no. 115736.

[7] N. Zhu and Z. Li, "Blind image splicing detection via noise level function," *Signal Process., Image Commun.*, vol. 68, pp. 181–192, Oct. 2018.

[8] X. Zhang, Y. Ning, X. Li, and C. Zhang, "Anti-noise FCM image segmentation method based on quadratic polynomial," *Signal Process.*, vol. 178, Jan. 2021, Art. no. 107767.

[9] A. Fateh, M. Fateh, and V. Abolghasemi, "Enhancing optical character recognition: Efficient techniques for document layout analysis and text line detection," *Eng. Rep.*, Dec. 2023, Art. no. e12832, doi: 10.1002/eng2.12832.

[10] H. Lv, P. Shan, H. Shi, and L. Zhao, "An adaptive bilateral filtering method based on improved convolution kernel used for infrared image enhancement," *Signal, Image Video Process.*, vol. 16, no. 8, pp. 2231–2237, Nov. 2022.

[11] X. Dong, J. Zhao, M. Sun, and X. Zhang, "Median value filtering method for non-circular signal DOA estimation with nested arrays in the presence of impulsive noise scenarios," in *Proc. 14th Int. Conf. Signal Process. Syst. (ICSPS)*, Nov. 2022, pp. 285–290.

[12] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3-D transform-domain collaborative filtering," *IEEE Trans. Image Process.*, vol. 16, no. 8, pp. 2080–2095, Aug. 2007.

[13] F. Zhang, N. Cai, J. Wu, G. Cen, H. Wang, and X. Chen, "Image denoising method based on a deep convolution neural network," *IET Image Process.*, vol. 12, no. 4, pp. 485–493, Apr. 2018.

[14] R. S. Thakur, R. N. Yadav, and L. Gupta, "State-of-art analysis of image denoising methods using convolutional neural networks," *IET Image Process.*, vol. 13, no. 13, pp. 2367–2380, Nov. 2019.

[15] S. Gai and Z. Bao, "New image denoising algorithm via improved deep convolutional neural network with perceptive loss," *Exp. Syst. Appl.*, vol. 138, Dec. 2019, Art. no. 112815.

[16] X. Jia, D. Meng, X. Zhang, and X. Feng, "PDNet: Progressive denoising network via stochastic supervision on reaction-diffusion–advection equation," *Inf. Sci.*, vol. 610, pp. 345–358, Sep. 2022.

[17] X. Jia, X. Feng, and S. Liu, "Dual non-autonomous deep convolutional neural network for image denoising," *Inf. Sci.*, vol. 572, pp. 263–276, Sep. 2021.

[18] C. Wang, Y. Huang, C. Ci, H. Chen, H. Wu, and Y. Zhao, "EIDNet: Extragradient-based iterative denoising network for image compressive sensing reconstruction," *Exp. Syst. Appl.*, vol. 250, Sep. 2024, Art. no. 123829.

[19] X. Li, Y. Li, Y. Zhou, J. Wu, Z. Zhao, J. Fan, F. Deng, Z. Wu, G. Xiao, J. He, and Y. Zhang, "Real-time denoising enables high-sensitivity fluorescence time-lapse imaging beyond the shot-noise limit," *Nature Biotechnol.*, vol. 41, no. 2, pp. 282–292, Feb. 2023.

[20] S. Roka and M. Diwakar, "Deep stacked denoising autoencoder for unsupervised anomaly detection in video surveillance," *J. Electron. Imag.*, vol. 32, no. 3, Jun. 2023, Art. no. 033015.

[21] T. Yu, S. Wang, W. Chen, F. R. Yu, V. C. M. Leung, and Z. Tian, "Joint self-supervised enhancement and denoising of low-light images," *IEEE Trans. Emerg. Topics Comput. Intell.*, vol. 8, no. 2, pp. 1800–1813, Apr. 2024.

[22] A. A. Hendriksen, D. M. Pelt, and K. J. Batenburg, "Noise2Inverse: Self-supervised deep convolutional denoising for tomography," *IEEE Trans. Comput. Imag.*, vol. 6, pp. 1320–1335, 2020.

[23] V. Sterzentsenko, L. Saroglou, A. Chatzitofis, S. Thermos, N. Zioulis, A. Doumanoglou, D. Zarpalas, and P. Daras, "Self-supervised deep depth denoising," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 1242–1251.

[24] L. Cheng and P. Chen, "Remote sensing image denoising based on Gaussian curvature and shearlet transform," *IEEE Access*, vol. 11, pp. 97716–97725, 2023.

[25] H. Neji, M. Ben Halima, J. Nogueras-Iso, T. M. Hamdani, J. Lacasta, H. Chabchoub, and A. M. Alimi, "Doc-attentive-GAN: Attentive GAN for historical document denoising," *Multimedia Tools Appl.*, vol. 83, no. 18, pp. 55509–55525, Nov. 2023.

[26] W. Lee, S. Son, and K. M. Lee, "AP-BSN: Self-supervised denoising for real-world images via asymmetric PD and blind-spot network," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 17704–17713.

[27] T. Huang, S. Li, X. Jia, H. Lu, and J. Liu, "Neighbor2Neighbor: Self-supervised denoising from single noisy images," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 14776–14785.

[28] Y. Quan, M. Chen, T. Pang, and H. Ji, "Self2Self with dropout: Learning self-supervised denoising from single image," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 1887–1895.

[29] A. Krull, T.-O. Buchholz, and F. Jug, "Noise2Void—Learning denoising from single noisy images," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 2124–2132.

[30] V. Lempitsky, A. Vedaldi, and D. Ulyanov, "Deep image prior," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 9446–9454.

[31] K. Zhang, W. Zuo, and L. Zhang, "FFDNet: Toward a fast and flexible solution for CNN-based image denoising," *IEEE Trans. Image Process.*, vol. 27, no. 9, pp. 4608–4622, Sep. 2018.

[32] W. Wu, S. Liu, Y. Xia, and Y. Zhang, "Dual residual attention network for image denoising," *Pattern Recognit.*, vol. 149, May 2024, Art. no. 110291.

[33] S. Rezvani, M. Fateh, and H. Khosravi, "ABANet: Attention boundary-aware network for image segmentation," *Exp. Syst.*, vol. 41, no. 9, Sep. 2024, Art. no. e13625.

[34] A. Fateh, R. T. Birgani, M. Fateh, and V. Abolghasemi, "Advancing multilingual handwritten numeral recognition with attention-driven transfer learning," *IEEE Access*, vol. 12, pp. 41381–41395, 2024.

[35] Y. Liang and W. Liang, "ResWCAE: Biometric pattern image denoising using residual wavelet-conditioned autoencoder," 2023, *arXiv:2307.12255*.

[36] J. Lehtinen, J. Munkberg, J. Hasselgren, S. Laine, T. Karras, M. Aittala, and T. Aila, "Noise2Noise: Learning image restoration without clean data," 2018, *arXiv:1803.04189*.

[37] D. Zhang and F. Zhou, "Self-supervised image denoising for real-world images with context-aware transformer," *IEEE Access*, vol. 11, pp. 14340–14349, 2023.

[38] Z. Wang, Y. Fu, J. Liu, and Y. Zhang, "LG-BPN: Local and global blind-patch network for self-supervised real-world denoising," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2023, pp. 18156–18165.

[39] C. Yao, S. Jin, M. Liu, and X. Ban, "Dense residual transformer for image denoising," *Electronics*, vol. 11, no. 3, p. 418, Jan. 2022.

[40] Y. Zhou, J. Lin, F. Ye, Y. Qu, and Y. Xie, "Efficient lightweight image denoising with triple attention transformer," in *Proc. AAAI Conf. Artif. Intell.*, 2024, vol. 38, no. 7, pp. 7704–7712.

[41] D. Wang, Z. Wu, and H. Yu, "Ted-Net: Convolution-free T2T vision transformer-based encoder–decoder dilation network for low-dose CT denoising," in *Proc. 12th Int. Workshop Mach. Learn. Med. Imag.* Strasbourg, France: Springer, Sep. 2021, pp. 416–425.

[42] M. A. Nazari Siahsar, S. Gholtashi, V. Abolghasemi, and Y. Chen, "Simultaneous denoising and interpolation of 2D seismic data using data-driven non-negative dictionary learning," *Signal Process.*, vol. 141, pp. 309–321, Dec. 2017.

[43] J. Lequyer, R. Philip, A. Sharma, W.-H. Hsu, and L. Pelletier, "A fast blind zero-shot denoiser," *Nature Mach. Intell.*, vol. 4, no. 11, pp. 953–963, Oct. 2022.

[44] S. Gu, L. Zhang, W. Zuo, and X. Feng, "Weighted nuclear norm minimization with application to image denoising," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 2862–2869.

[45] Y. Mansour and R. Heckel, "Zero-shot Noise2Noise: Efficient image denoising without any data," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2023, pp. 14018–14027.

[46] J. Rock, M. Toth, E. Messner, P. Meissner, and F. Pernkopf, "Complex signal denoising and interference mitigation for automotive radar using convolutional neural networks," in *Proc. 22th Int. Conf. Inf. Fusion (FUSION)*, Jul. 2019, pp. 1–8.

[47] C. Qi, J. Chen, X. Yang, and Q. Chen, "Real-time streaming video denoising with bidirectional buffers," in *Proc. 30th ACM Int. Conf. Multimedia*, Oct. 2022, pp. 2758–2766.

[48] X. Chen and K. He, "Exploring simple Siamese representation learning," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 15745–15753.

[49] X. Lin, C. Ren, X. Liu, J. Huang, and Y. Lei, "Unsupervised image denoising in real-world scenarios via self-collaboration parallel generative adversarial branches," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, Oct. 2023, pp. 12642–12652.
[50] H. Qu, K. Liu, and L. Zhang, "Research on improved black widow algorithm for medical image denoising," *Sci. Rep.*, vol. 14, no. 1, p. 2514, Jan. 2024.
[51] J. Deng and C. Hu, "A new multi-scale CNN with pixel-wise attention for image denoising," *Signal, Image Video Process.*, vol. 18, no. 3, pp. 2733–2741, Apr. 2024.
[52] Y. Xie, Z. Wang, and S. Ji, "Noise2Same: Optimizing a self-supervised bound for image denoising," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 33, 2020, pp. 20320–20330.
[53] J. Xu, Y. Huang, M.-M. Cheng, L. Liu, F. Zhu, Z. Xu, and L. Shao, "Noisy-as-clean: Learning self-supervised denoising from corrupted image," *IEEE Trans. Image Process.*, vol. 29, pp. 9316–9329, 2020.
[54] J. Ko and S. Lee, "Self2Self+: Single-image denoising with self-supervised learning and image quality assessment loss," 2023, *arXiv:2307.10695*.
[55] S. M. Kasar and S. D. Ruikar, "Image demosaicking by nonlocal adaptive thresholding," in *Proc. Int. Conf. Signal Process., Image Process. Pattern Recognit.*, Feb. 2013, pp. 34–38.
[56] A. Abdelhamed, S. Lin, and M. S. Brown, "A high-quality denoising dataset for smartphone cameras," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 1692–1700.
[57] S. Nam, Y. Hwang, Y. Matsushita, and S. J. Kim, "A holistic approach to cross-channel image noise modeling and its application to image denoising," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 1683–1691.
[58] Y. Zhang, Y. Zhu, E. Nichols, Q. Wang, S. Zhang, C. Smith, and S. Howard, "A Poisson–Gaussian denoising dataset with real fluorescence microscopy images," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 11702–11710.
[59] J. Zbontar, F. Knoll, A. Sriram, T. Murrell, Z. Huang, M. J. Muckley, A. Defazio, R. Stern, P. Johnson, and M. Bruno, "FastMRI: An open dataset and benchmarks for accelerated MRI," 2018, *arXiv:1811.08839*.

SADJAD REZVANI received the master's degree from the Shahrood University of Technology, with a thesis on masked face recognition using deep learning techniques. At Semnan Technology Park, he involved in diverse industry projects, as a Computer Vision Engineer. His tasks included developing face recognition software, license plate recognition systems, and a salt crack sorting machine. He enjoys finding innovative solutions to practical challenges in computer vision and research in this field. He is deeply involved in researching artificial intelligence, particularly focusing on machine learning.

FATEMEH SOLEYMANI SIAHKAR received the bachelor's degree in electrical engineering from Shahid Chamran University, Ahvaz, Iran, and the M.S. degree in electrical engineering from Shahid Beheshti University, Tehran, Iran. She is highly interested in AI and computer vision fields. She was involved in several projects as a Machine Learning Engineer, such as monitoring urban traffic, forecasting, and analyzing stock markets.

YASIN REZVANI is currently pursuing the bachelor's degree in computer science with the Shahrood University of Technology. With a strong passion for artificial intelligence, he specializes in leveraging deep learning techniques for computer vision tasks. His research interests include medical image analysis, where he aims to develop innovative solutions to improve diagnostic accuracy and efficiency. He is dedicated to advancing the capabilities of machine learning in healthcare, contributing to a future where technology plays a pivotal role in medical advancements.

ABDORREZA ALAVI GHARAHBAGH (Member, IEEE) received the bachelor's degree in electrical engineering from the Ferdowsi University of Mashhad, Iran, and the M.S.C. degree in communication engineering from Urmia University, Iran. He is currently a Staff Member with Islamic Azad University, Shahrood Branch. His programming skills include C/C++, Python, and MATLAB (professional level). As an Research and Development Consultant, he involved in various areas, including computer vision (recognition, detection, and classification problems), signal and image processing, data analysis and pattern recognition, modeling and optimization (particularly genetic algorithms), renewable energy, machine learning, natural language processing, stochastic processes and mathematical programming, data mining, big data, deep learning, and parallel processing.

VAHID ABOLGHASEMI (Senior Member, IEEE) received the Ph.D. degree in signal processing from the University of Surrey, Guildford, U.K., in 2011. He is currently an Associate Professor with the School of Computer Science and Electronic Engineering, University of Essex, Colchester, U.K. His research interests include signal and image processing, compressive sensing, and machine learning. His expertise extends to cutting-edge technologies, including smart and adaptive low-power sensing and communication, wireless image transmission, compressed and lightweight neural networks, and artificial intelligence.