An Improved Combination of Image Denoisers Using Spatial Local Fusion Strategy

YIWEN LIU, SHAOPING XU, AND ZHENYU LIN

School of Information Engineering, Nanchang University, Nanchang 330031, China

Corresponding author: Shaoping Xu (xushaoping@ncu.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 61662044, Grant 51765042, and Grant 61163023; in part by the Jiangxi Provincial Natural Science Foundation under Grant 20171BAB202017; and in part by the Jiangxi Provincial Graduate Innovation Special Fund under Grant YC2018-S066.

ABSTRACT Image denoising is a well-researched problem in the image processing field. Numerous image denoising algorithms have been proposed in the past. Although researchers have continually focused on improving the denoising algorithm performance regarding denoising effect and outstanding results have been achieved, the improvement amplitude of a single denoising algorithm has decreased over time. Recently, the consensus neural network (CsNet), which combines multiple image denoisers to produce an overall better result compared to single algorithms, was proposed. However, the denoising process of CsNet is time-consuming owing to its MSE optimal weight setting and iterative boosting stages. Therefore, we propose an improved combination of nonlocally centralized sparse representation (NCSR) with a fast and flexible denoising convolutional neural network (FFDNet) using a spatial local fusion strategy (ICID). ICID uses a structural-based patch to decompose their denoised images into the strength, structure, and mean intensity components. Thereafter, an image patch is reconstructed and placed back into the fused image after fusing the three components separately. Experimental results verified that our algorithm is superior to CsNet, and it is faster. The combination of NCSR and FFDNet can harmonize the complementary denoising capabilities of different denoising algorithms. NCSR can preserve as many details as possible in natural images with numerous repeated structures, whereas FFDNet can achieve state-of-the-art results with a sufficiently large training set of images. Moreover, ICID uses the structural-based method, which considers more local details and preserves more textures, resulting in superior performance.

INDEX TERMS Image denoising, denoising performance, improved combination, spatial local fusion, image decomposition, image recomposition.

I. INTRODUCTION

Noise corruption in digital images is inevitable during image acquisition or transmission. Image denoising, which is an essential step in various image processing and analysis tasks, aims to estimate a high-quality image from its noisy observation while preserving the image edges, textures, and concrete details as much as possible. If image denoising is not processed effectively, the efficiency and performance of the various image processing algorithms will be affected. Although image denoising has been studied comprehensively and several successful image denoising algorithms have been proposed, researchers are continuously working to improve the performance of denoising algorithms.

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

Traditional methods for image denoising include total variation-based methods [1]–[3], low rankness-based methods [4]–[6], nonlocal means (NLM) methods [7]–[9], and sparse representation-based methods [10]–[12]. The majority of the leading denoising algorithms involve the intersection and improvement of the above algorithms, such as NLM, block-matching and 3D filtering (BM3D) [13], weighted nuclear norm minimization (WNNM) [4], learned simultaneous sparse coding (LSSC), and nonlocally centralized sparse representation (NCSR). BM3D searches for similar patches and groups them into 3D arrays to enhance the sparse representation, and then it conducts collaborative filtering on the patches. For its outstanding performance, it has been a benchmark algorithm. Mairal et al. [14] presented LSSC, which exploits the self-similarities in natural images combined with sparse coding to improve the dictionary performance. In
contrast to BM3D, LSSC avoids the pre-designed dictionary to achieve enhanced sparsity. Another classical dictionary learning method is K-SVD [15], which was proposed to reconstruct an underlying clean patch by sparse and redundant representations with an effective over-complete learned dictionary. The success of K-SVD in denoising inspired numerous subsequent works, such as multi-scale dictionary learning, double sparsity adaptive principal component analysis (PCA) dictionaries, and semi-coupled dictionary learning. Recently, Dong et al. [10] applied the concepts of sparse representation and nonlocal self-similarity, demonstrating a powerful image restoration capability. As NCSR consists of nonlocal patches with similar structures, it has exhibited outstanding performance for natural images with many repeated structures. Xu et al. [12] proposed a method known as patch group prior-based denoising (PGPD), which uses patch groups to learn from natural images to enable high-performance denoising. Xu et al. proposed Trilateral weight sparse coding (TWSC) [11], which utilizes a sparse coding strategy to handle realistic noise and image priors. These traditional model-based denoising methods are constantly improving, but the optimization amplitude is small. Moreover, the majority of the traditional denoising methods typically suffer from two major drawbacks regarding the denoising performance. Firstly, these methods generally involve a complex optimization problem during the testing stage, which makes the denoising process time consuming. Secondly, these models generally employ hand-crafted image priors and involve several manually selected parameters, thereby providing room to boost the denoising effect.

To overcome the above limitations, several learning-based methods have recently been proposed to learn the basic image prior and provide rapid inference from a training set of real-world image pairs. Learning-based methods provide abundant data to train a deep model, which result in rapid speeds and high efficiency. Thus, learning-based methods have achieved highly competitive results for image denoising, given a sufficiently large training set of images. One approach is to learn the stage-wise image priors, such as cascade of shrinkage fields [16] or trainable nonlinear reaction diffusion [17]. Another popular approach is plain discriminative learning, such as multilayer perception (MLP) [18] and convolutional neural network based methods [19], [20], among which the denoising CNN (DnCNN) [20] method has achieved highly competitive denoising performance. MLP [18] was proposed and began to utilize GPU for acceleration to boost the denoising effect. Xie et al. [21] presented stacked sparse denoising auto-encoders, which combined sparse coding and pre-trained deep networks to provide solutions to low-level vision tasks. However, deep neural networks could not effectively capture the inherent features of the original images. To address the above issues, numerous researchers have started to use deep CNNs for image denoising. Benefiting from the advances in deep CNNs, Mao et al. [22] proposed the residual encoder–decoder network (RED-Net), which applies skip connections to recover clean images and trains deep networks to boost the performance. Zhang et al. [23] proposed FFDNet, which is a fast and flexible CNN with an adjustable noise level map as the input. To provide an improved trade-off between accuracy and speed, Zhang et al. [24] introduced a seven-layer denoising network with dilated convolution [25] to expand the receptive field of the CNN. Nowadays, CNN-based methods exhibit superior performance, with rapid speeds and high visual quality in image processing tasks.

Image denoising has become an extensively studied problem in the image processing field. For example, some of the above methods have been demonstrated to achieve state-of-the-art image denoising performance regarding denoising effect. At present, researchers believe that the performance of existing algorithms has not yet reached their theoretical limits [26]. Improving the performance of the current state of the art is continually being pursued. Despite the fact that the existing single classical denoising algorithm still has room for improvement, its optimization amplitude is small. Choi et al. [27] proposed a framework called the consensus neural network (CsNet), which combines several denoisers to produce a better result. CsNet roughly consists of three major steps. Beginning with a group of image denoisers, CsNet uses a deep neural network to estimate the mean squared error (MSE). Then, the estimated MSE is used to solve a complex optimization problem and they obtain an initial improved denoised image. Finally, using a new deep neural network image booster, the combined denoising result is boosted further. CsNet exhibits better denoising effect compared with other state-of-the-art denoising algorithms; however, its denoising process is time-consuming, because of involving the MSE optimal weight setting and iterative boosting stages. Inspired by CsNet, we propose a spatial local fusion strategy to fuse denoised images generated by NCSR and FFDNet. Note that the spatial local fusion strategy actually provides a general solution to denoising applications that allows to fuse the combination of the preliminary denoised images generated by any two image denoisers fast. We analysis and find that the combination of NCSR and FFDNet obtains an overall superior result from extensive experiments. Generally, the denoising effect and computational complexity are two important indexes for evaluating the performance of denoising algorithms. In the process of algorithm research, researchers tend to improve denoising effect first, and then reduce the computational complexity after satisfying the improvement of denoising effect. In this work, our primary goal is to further boost the denoising effect, and then reduce the complexity. Extensive experiments demonstrated that the proposed improved combination of image denoisers (ICID) outperforms CsNet and the single state-of-the-art algorithms, including DnCNN, BM3D, and WNNM. Meanwhile, the speed of our strategy is faster than that of CsNet. This article makes the following contributions:

1) We introduce an improved denoiser that exploits the combination strategy to boost the denoising effect. We first obtain preliminary denoised images as image fusion sources by NCSR and FFDNet owing to their
complementarity and outstanding performance. Then, the preliminary denoised images are fused to obtain final result, which is superior to that by any single denoiser.

2) We exploit structural-based spatial local fusion to process the preliminary multiple fusion sources and expect more detailed information to be considered. Especially, we introduce graph-based visual saliency (GBVS) [28] to control the weight coefficient that reflects the mean pixel intensity according to the surrounding pixel intensities and the overall brightness.

The remainder of this article is organized as follows. Section II presents the NCSR model and FFDNet and underlines their contributions to our study. Section III outlines the generation of the fusion image source and proposed spatial local structural-based fusion. Section IV presents our experimental setting and the experimental results that are compared to other state-of-the-art methods. We discuss and evaluate the efficiency and effectiveness of the proposed method in Section V. The paper is concluded in Section VI.

II. RELATED WORK

A. NCSR FOR IMAGE RESTORATION

The core technique of the classical NCSR algorithm is sparse coding, which represents image patches as compact linear combinations of several atoms that are sparsely selected from a dictionary. This task aims to estimate other sparse coding coefficients of the tile given the dictionary. The use of sparsity-based regularization has attained satisfactory results.

Following the notation used in [10], for an image \( x \in \mathbb{R}^N \), let \( x_i = R_i x \) denote the extraction of an image patch of size \( \sqrt{n} \times \sqrt{n} \) at pixel \( i \), where \( R_i \) is the matrix that extracts patch \( x_i \) from \( x \) at pixel \( i \). Given a dictionary \( \Phi \in \mathbb{R}^{N \times M} \), with \( n \leq M \), each patch extracted from a given image can be sparsely represented as \( x_i \approx \Phi \alpha_{x,i} \) by solving a l1-minimization problem:

\[
\min_{\alpha_i} \sum_{i=1}^{P} \left( \| x_i - \Phi \alpha_i \|^2_2 + \lambda \| \alpha_i \|_1 \right),
\]

where \( P \) is the sum of the patches extracted from image \( x \), \( \alpha_i \) represents the sparse coding coefficients of \( x_i \) in contrast \( x_i \), and \( \lambda \) is the regularization parameter. Thus, the entire image \( x \) is represented by the set of sparse code \( \alpha_{x,i} \). The patches can be overlapped to obtain a redundant patch-based representation and to restrain boundary artifacts. The least-squares solution to reconstruct \( x \) from \( \alpha_{x,i} \) can be expressed as follows:

\[
x \approx \Phi \circ \alpha_x = \left( \sum_{i=1}^{N} R_i^T R_i \right)^{-1} \sum_{i=1}^{N} \left( R_i^T \Phi \alpha_{x,i} \right),
\]

where \( \alpha_x \) represents the set of all \( \alpha_{x,i} \). The above formulation indicates that the entire image \( x \) is reconstructed by averaging each reconstructed patch of \( x_i \). In general, the observed noisy image can be formulated by \( y = Hx + v \). Here, \( H \) denotes a degradation matrix and \( v \) is the noise. \( x \) is the original clean image and \( y \) is the observed noisy image. In this case, \( x \) is recovered from \( y \), which is sparsely coded with respect to \( \Phi \) initially, by solving the following minimization problem:

\[
\alpha_y = \arg \min_{\alpha} \left( \| x_i - \Phi \|^2_2 + \lambda \| \alpha_i \|_1 \right),
\]

Clearly, \( \alpha_y \) deviates from \( \alpha_x \), and the difference between \( \alpha_y \) and \( \alpha_x \) is defined as sparse coding noise (SCN): \( v_a = \alpha_y - \alpha_x \). For superior reconstruction of \( x \) from \( y \), the SCN \( v_a \) should be reduced as much as possible. Thereafter, the image \( x \) is estimated as \( \hat{x} = \Phi \circ \alpha_y \). Equation (3) can be rewritten as:

\[
\alpha_y = \arg \min_{\alpha} \left( \| x_i - \Phi \|^2_2 + \lambda \sum_{i=1}^{N} \| \alpha_i - \beta_i \|_p \right),
\]

The role of regularization parameter \( \lambda \) is to balance the fidelity term and the centralized sparsity term, so \( \lambda \) should be determined adaptively to improve performance. Here, \( \beta_i \) is an estimation of unknown sparse coding coefficient \( \alpha_i \). Only one regularization term \( \| \alpha_i - \beta_i \|_p \) exists in the above model. When \( p = 1 \), \( \beta_i \) is obtained by nonlocal regularization of the natural image. The regularization term is transformed into a nonlocal centralized sparsity term.

The NCSR model employs an adaptive sparse domain selection strategy. Eq. (4) can be solved iteratively. One essential part is the solving dictionary. The K-means clustering method and PCA dictionary are used to construct the dictionary \( \Phi \). Finally, the set of image patches extracted from a corrupted image is clustered into K clusters by K-means. For each cluster, a dictionary of PCA bases has to be learned. For a given patch, according to its distances to the means of the clusters, it is coded by a PCA sub-dictionary of the cluster. Another essential part of the NCSR algorithm is the estimation of \( \beta_i \). As natural images usually contain repetitive structures, that is, a significant quantity of nonlocal redundancies [29], \( \beta_i \) can be calculated as the weighted average of these sparse codes, which are associated with nonlocal patches similar to patch \( i \). Mathematically, let \( \Omega_i \) denote the set of patches similar to patch \( x_{i,q} \), following which \( \beta_i \) can be calculated as the weighted average of \( \alpha_{x,q} \) within set \( \Omega_i \):

\[
\beta_i = \sum_{q \in \Omega_i} w_{i,q} \alpha_{i,q},
\]

where \( w_{i,q} \) is the weight and \( \alpha_{i,q} \) is the sparsity of patch \( x_{i,q} \). The weight \( w_{i,q} \) can be calculated as inversely proportional to the distance between patches \( x_i \) and \( x_{i,q} \):

\[
w_{i,q} = \frac{1}{W} \exp\left(-\frac{\| x_i - x_{i,q} \|^2_2}{h}\right),
\]

where \( x_{i,q} = \Phi \alpha_{x_i} \), \( x_{i,q} = \Phi \alpha_{x_i,q} \). \( h \) is a predetermined scalar, and \( W \) is the normalization factor. Note that an effective estimation of \( \beta_i \) relies on the nonlocal redundancies of natural images.

In summary, NCSR uses sparsity-based regularization to achieve superior denoising performance, compared with classic regularization models that tend to over-smooth images.
owing to their piecewise constant assumption. Sparsity-based regularization uses the sparse representation and nonlocal self-similarity of the natural images to obtain outstanding performance. Consequently, we selected NCSR as a representative denoising algorithm in the denoiser combination for fusion sources. However, images with irregular textures would weaken this prior advantage, thereby leading to poor results. It is necessary for us to choose other algorithms that are complementary to NCSR, as one of denoiser in combination. Furthermore, the processing of NCSR is time-consuming, especially the process involving the training dictionary. In our implementation, we utilize our previous work, an improved NCSR model, called fast NCSR(FNCSR) \[30\] for fusion sources. However, images with irregular textures are generated by adding additive white Gaussian noise with a noise level \( \sigma \in [0, 75] \) to the clean patches \( x_i \) and constructing the corresponding noise level map \( M_i \). The noise level map \( M_i \) is uniform for each noisy patch \( y_i \). The output pixel is determined to its convolution layers. FFDNet outperforms state-of-the-art denoising models owing to its convolution layers.

The flexibility of the CNN-based denoiser can be improved by learning and analyzing how the model-based image denoising flexibly handles different levels of noise. In certain optimization algorithms, the solution for most model-based methods can be defined as

\[
\hat{x} = \mathcal{F}(y, \sigma, \lambda; \Theta). \tag{7}
\]

Based on the analysis of model-based methods, it is natural to exploit the CNN to learn an explicit mapping of Eq. (7), which puts the noise level and noise image in the input. Inspired by the patch-based denoising methods, which actually set \( \sigma \) for each patch, the noise level \( \sigma \) is stretched into a noise level map \( M \) to resolve the measurement mismatching problem (in which input \( y \) and \( \sigma \) have different measurements). In \( M \), all the elements are \( \sigma \); thus, Eq. (7) can be rewritten as

\[
\hat{x} = \mathcal{F}(y, M; \Theta). \tag{8}
\]

The training dataset is composed of input-output pairs \((y_i, M_i); x_i\)\text{init}\[N\], where the noisy patches \( y_i \) are generated by adding additive white Gaussian noise with a noise level \( \sigma \in [0, 75] \) to the clean patches \( x_i \) and constructing the corresponding noise level map \( M_i \). The noise level map \( M_i \) is uniform for each noisy patch \( y_i \). The output pixel is determined to the CNN to learn an explicit mapping of Eq. (7), which puts the noise level and noise image in the input.
by the local noisy input and local noise level, as FFDNet is a full CNN. In the case of using residual learning, the network outputs an estimation of the input noise $F(x_i)$. Thereafter, the denoised image is computed by subtracting the output noise from the noisy input $F(y, M; \Theta) - x_i$, where $\Theta$ is the collection of all learnable parameters. The ADAM algorithm [33] is applied to optimize FFDNet and then to minimize the following loss function, with its hyperparameters set to their default values:

$$L(\Theta) = \frac{1}{2N} \sum_{i=1}^{N} \| F(y, M; \Theta) - x_i \|. \tag{9}$$

Not only can FFDNet produce state-of-the-art results when the input noise level matches the ground-truth noise level, but also it can robustly control the tradeoff between noise reduction and detail preservation. Moreover, FFDNet exhibits faster speeds than competing methods such as BM3D and DnCNN. However, the performance of CNN-based methods tend to be limited for relying on the training data. Its training dataset may not be very consistent with the given noisy image. Its denoising results is not superior to NCSR for some images with regular structures. Therefore, the denoiser combination of FFDNet and NCSR can harmonize their complementary denoising capabilities and preserve more local details.

III. STRUCTURAL-BASED SPATIAL LOCAL FUSION

A. BASIC CONCEPT

In this section, we present the structural-based spatial local fusion that combines images denoised by NCSR and FFDNet. The majority of the existing pixel-based fusion algorithms [34]–[38], typically follow a weighted summation formulation:

$$\hat{X}(i) = \sum_{k=1}^{K} W_k(i) X_k(i), \tag{10}$$

where $K$ is the sum of input images source sequences; $W_k(i)$ and $X_k(i)$ represent the weight and intensity values at the $i$-th pixel in the $k$-th image source sequence, respectively; and $\hat{X}(i)$ denotes the $i$-th pixel in the fused image. In the transform domain, a straightforward extension of this approach is to replace $X_k(i)$ with transform coefficients. The weighting map $W_k$ often carries information regarding the structural preservation and perceptual importance of the $k$-th input image at the pixel level. By using specific models to quantify this information, existing pixel-based fusion methods vary mainly in the computation of $W_k$ and how it may fit into the space or scale based on the image content.

The aforementioned pixel-based fusion methods need to purposely take into account the noisy characteristics of $W_k$. The result by using pixel-wise fusion methods can lead to new distortion (i.e., artifacts) with inappropriate weight setting. Therefore, post-processing is necessary to produce a reasonable fused image, which is a major drawback of this type of method.

As illustrated in Fig. 2, we propose a structural-based spatial local fusion strategy that works on image patches instead of pixels. First, we assume that a set of denoisers can denoise noisy images and obtain $K$ preliminary denoised image fusion sources. In this study, we set $K = 2$. We use an improved denoiser combination of NCSR and FFDNet, owing to their complementarity and outstanding performance, and then gain two image fusion source sequences. In particular, for an input image $X$, we obtain a preliminary denoised image $X_1$ after processing by NCSR. Furthermore,
The approach contains two main parts: decomposing a given image into three components and multi-source fusion. Firstly, we obtain two image source sequences \( X_k = \{x_k|1 \leq k \leq 2\} \) for the fusion process. We let \( x_k \) be a set of co-located image patches in the source image sequence. For a given patch \( x_k \), we adopt the patch-based method to decompose it into three components: strength component, structure component, and mean intensity component:

\[
x_k = \| \bar{x}_k - \mu_s \| \frac{\| x_k - \mu_s \|}{\| x_k - \mu_s \|} + \mu_s
\]

In Eq. (11), \( \| \cdot \| \) represents the \( l_2 \) norm of a vector, \( \mu_s \) is the average value of the patch, and \( \bar{x}_k = x_k - \mu_s \) represents a mean-removed patch. Hence, the physical meanings of the scalar \( c = \| \bar{x}_k \| \), vector \( s = \frac{\bar{x}_k}{\| \bar{x}_k \|} \), and scalar \( l = \mu_s \) correspond to the strength component, structure component, and mean intensity component of \( x_k \), respectively. The concept of visual saliency is introduced into fusion processing. Next, we discuss how the three components are fused.

First, we process the strength component. The visibility of the local patch structure relies on the local contrast, which is straightforwardly related to the strength component. In general, higher contrast results in improved visibility. However, the higher contrast may also lead to unreality of the local structure. The following structure component is faced with the same problem. Here, the contrast of the fused patch relies on the highest contrast, because using GBVS [28] to control the weight of mean intensity in the fusion process can produce an overall superior result after the reconstructed image patches are placed back into the fused images. This also acts on following structure component. And this problem can be partially alleviated as discussed below.

\[
\hat{c} = \max(c_1, c_2) = \max(\| \hat{x}_1 \|, \| \hat{x}_2 \|)
\]

The patch structure \( s_k \) changes with the strength component. The structures of the fused image patch are expected to effectively represent the structures of all source image patches. A simple implementation of this relationship is computed by

\[
\hat{s} = \frac{\hat{s}}{\| \hat{s} \|}
\]

where \( S(\cdot) \) is a power-weighting function given by \( S(\bar{x}_k) = \| \bar{x}_k \|^p \), which determines the contribution of each source image patch in the structure of the fused image patch. \( \hat{s} \) can be expressed as \( \frac{S(\bar{x}_1)\hat{x}_1 + S(\bar{x}_2)\hat{x}_2}{S(\bar{x}_1) + S(\bar{x}_2)} \). Here, \( p \geq 0 \) is an exponent parameter that is determined adaptively based on the patch consistency measure [37]. The general formulation leads to a family of weighted functions with different meanings that can be realized by selecting different values of \( p \). A larger value of \( p \) results in a higher degree of attention to the patches that have relatively higher strength.

We introduce the GBVS [28] map into the mean intensity component \( l \). The HVS system is more sensitive to pixels

### Table 1. Proposed spatial local fusion strategy steps.

| Input: noisy image \( X \). |
|---|
| Preliminary Denoised Image Sources Generation: |
| (a) Denoise \( X \) with NCSR to obtain \( X_1 \). |
| (b) Denoise \( X \) with FFDNet to obtain \( X_2 \). |
| Structural Fusion: |
| for each patch position in above sources \( \{X_k\} = \{X_k|1 \leq k \leq 2\} \). |
| (c) Stack its channel pixels into a vector. |
| (d) Decompose the patches into \( \{c_1, s_1, l_1\} \) and \( \{c_2, s_2, l_2\} \). |
| (e) Fuse the components respectively based on Eq. (12), Eq. (13) and Eq. (14). |
| (f) Recompose the fused patch by Eq. (16). |
| (g) Obtain \( \hat{X} \) by accumulating and normalizing all the overlapping reconstructed patches. |

Output: \( \hat{X} \).
with higher luminance in the image, because these pixels are more stimulating to the human eye and easier to be observed. Specifically, we utilize the visual saliency metric as the weight coefficient of the mean intensity of each patch during fusion processing. Thus, the mean intensity component of the local patch can be expressed as:

$$
\hat{l}(i,j) = \sum_{k=1}^{K} W_l^k(i,j) \cdot l_k(i,j),
$$

(14)

where \((i,j)\) represents the location of image pixels, and \(W_l^k(i,j)\) is the control coefficient of the weight. The \(GBVS_k(i,j)\) is the GBVS map value of the \(k\)th image, which is used as the weight coefficient here, and its normalization is expressed as:

$$
W_l^k(i,j) = \frac{GBVS_k(i,j)}{\sum_{k=1}^{K} GBVS_k(i,j)},
$$

(15)

Once \(\hat{c}, \hat{s}\), and \(\hat{l}\) have been computed, we place them into the channels, reconstruct them, and define a new vector:

$$
\hat{x} = \hat{c} \cdot \hat{s} + \hat{l}.
$$

(16)

The proposed spatial local fusion strategy utilizes a moving window with a fixed stride to extract patches from a given denoised image. The pixels in the overlapping patches are averaged to produce the final output. The output patches by our proposed strategy contain abundant meaningful structural information in the same spatial location.

### IV. EXPERIMENTS

#### A. DATASETS AND EXPERIMENTAL SETUP

To assess the performance of the methods objectively, several existing state-of-the-art algorithms for image denoising, including BM3D [13], WNNM [4], NCSR [10], PGPD [12], TWSC [11], RED-Net [22], DnCNN [20], FFDNet [23], and CsNet [27], were selected for comparison with the proposed method. Denoising experiments were conducted on three datasets. The first dataset contained 10 images selected from commonly used image datasets in the literature for image processing, as illustrated in Fig. 3. To increase the testing difficulty, the second dataset was provided by the Berkeley segmentation dataset (BSD) [39], and we selected 50 representative images therefrom to generate the testing dataset, as illustrated in Fig. 4. It is suitable for testing the visual performance and robustness, for which high-quality natural and rich texture images were included. The third dataset was a real noise dataset, Darmstadt Noise Dataset (DND) [40].

To test the performance of the proposed methods, we used the visual effect and recognized the evaluating criteria [41]–[43] of the image quality, including the peak signal-to-noise ratio (PSNR) and structural similarity index (SSIM) [44] to verify the denoising effect on the first and second datasets. Through the integrated local natural image quality evaluator(ILNIQE) [45] and no-reference image quality metric for contrast distortion(NIQMC) [46], we evaluated the denoising effect of our method on real noise images. We used the same initial denoiser setting as the original authors. All of the experiments were performed in the MATLAB (R2017b) environment on a Lenovo desktop PC with a six-core Intel (R) Core (TM) i7-3770 CPU @3.40 GHz with 16 GB of RAM.

#### B. QUANTITATIVE EVALUATION METRICS

The PSNR, which is a quality measure commonly used in image processing, was applied to evaluate the visual quality of the denoised images. The larger the PSNR, the better the denoising effect in many cases. Table 2 presents the PSNR results of six images from the first testing dataset. From Table 2, it is clear that the higher PSNR values are obtained by our method, indicating that the effect of our algorithm on natural images is superior.
The results are listed in Table 3, indicating the average PSNR results of the different methods with various noise levels. The proposed method consistently outperformed other methods and achieved the highest average PSNR compared with other state-of-the-art methods, including the single NCSR and FFDNNet. Our method outperformed NCSR and FFDNNet by 0.2 to 0.6 dB in terms of PSNR. To evaluate the quality of the denoised images more reliably, the SSIM is also calculated. The value of SSIM being close to 1 indicates superior visual quality. SSIM can measure structural information effectively: a larger SSIM value indicates more visual comfortableness. Table 4 lists the average SSIM values on the first testing dataset. We utilize the visual saliency metric as the weight coefficient of the mean intensity of each patch to better reflect the characteristics of human visual attention to information; thus, our method yields the highest SSIM which is more suitable for human visual perception among the compared algorithms.

To further verify the robustness of the proposed approach, we conducted an experiment on the second test dataset. It can be observed from Table 5 that the proposed framework achieved a highly competitive performance compared with the other leading denoising methods. Our method produced very similar results to those of the experiment on the first test dataset and achieved the highest average PSNR and SSIM values among the compared algorithms for the six noise levels tested. For color image denoising, we also carried out experiment on color version of the BSD dataset. Actually, our fusion strategy can easily extend to denoising color image.
TABLE 5. Average PNSR (dB)/SSIM results of different methods on BSD with various noise levels.

| Method      | Noise Level | \( \sigma = 10 \) | \( \sigma = 20 \) | \( \sigma = 30 \) | \( \sigma = 40 \) | \( \sigma = 50 \) | \( \sigma = 60 \) |
|-------------|-------------|--------------------|--------------------|--------------------|--------------------|--------------------|--------------------|
| BM3D [13]   |             | 33.41 / 0.9230    | 29.56 / 0.8404    | 27.60 / 0.7829    | 26.26 / 0.7373    | 25.40 / 0.7021    | 24.73 / 0.6743    |
| FFDNet [23] |             | 33.61 / 0.9301    | 30.02 / 0.8604    | 28.16 / 0.8035    | 26.93 / 0.7569    | 26.04 / 0.7189    | 25.34 / 0.6875    |
| NCSR [10]   |             | 33.42 / 0.9243    | 29.58 / 0.8420    | 27.60 / 0.7860    | 26.27 / 0.7373    | 25.35 / 0.7032    | 24.59 / 0.6732    |
| DnCNN [20]  |             | 33.69 / 0.9320    | 30.06 / 0.8595    | 28.14 / 0.8062    | 26.87 / 0.7628    | 25.99 / 0.6948    | 25.23 / 0.6656    |
| WNNM [4]    |             | 33.53 / 0.9248    | 29.73 / 0.8449    | 27.81 / 0.7904    | 26.54 / 0.7460    | 25.63 / 0.7140    | 24.92 / 0.6844    |
| RED-Net [22]|             | 33.36 / 0.9229    | 29.51 / 0.8427    | 27.75 / 0.7851    | 26.59 / 0.7352    | 25.73 / 0.6964    | 25.07 / 0.6675    |
| PQPD [12]   |             | 33.30 / 0.9179    | 29.51 / 0.8315    | 27.57 / 0.7655    | 26.37 / 0.7170    | 25.46 / 0.6777    | 24.82 / 0.6507    |
| TWSC [11]   |             | 33.47 / 0.9266    | 29.69 / 0.8450    | 27.73 / 0.7773    | 26.46 / 0.7240    | 25.52 / 0.6819    | 24.79 / 0.6480    |
| CsNet [27]  |             | 33.59 / 0.9260    | 29.67 / 0.8474    | 27.90 / 0.7895    | 26.75 / 0.7440    | 25.91 / 0.7059    | 25.20 / 0.6731    |
| Proposed    |             | 34.72 / 0.9440    | 30.67 / 0.8726    | 28.79 / 0.8191    | 27.47 / 0.7697    | 26.56 / 0.7323    | 25.83 / 0.6998    |

TABLE 6. Average PNSR (dB) results of different methods on color version of the BSD with noise level \( \sigma = 30 \).

| Noise Level | BM3D [13] | FFDNet [23] | NCSR [10] | RED-Net [22] | CsNet [27] | Proposed |
|-------------|-----------|-------------|-----------|--------------|------------|----------|
| \( \sigma = 30 \) | 29.78 | 30.20 | 27.38 | 30.50 | 30.62 | 31.55 |

FIGURE 5. Denoising results of Lena with noise level \( \sigma = 40 \): (a) original; (b) noisy image; (c) BM3D [13] / 25.76dB; (d) NCSR [10] / 30.05dB; (e) TWSC [11] / 30.26dB; (f) PQPD [12] / 30.26dB; (g) WNNM [4] / 30.25dB; (h) RED-Net [22] / 30.34dB; (i) DnCNN [20] / 30.47dB; (j) FFDNet [23] / 30.80dB; (k) CsNet [27] / 28.04dB; (l) Proposed / 31.56dB.
FIGURE 6. Denoising results of House with noise level $\sigma = 30$: (a) original; (b) noisy image; (c) BM3D [13] / 29.36dB; (d) NCSR [10] / 32.12dB; (e) TWSC [11] / 32.32dB; (f) PGPD [12] / 32.39dB; (g) WNNM [4] / 32.45dB; (h) RED-Net [22] / 27.85dB; (i) DnCNN [20] / 32.19dB; (j) FFDNet [23] / 32.54dB; (k) CsNet [27] / 31.81dB; (l) Proposed / 33.09dB.

We set $C = 3$ (i.e., the number of channels is three) for color image. For example, Table 6 reports the denoising effect of different methods with noise level $\sigma = 30$ and demonstrates that our method is still effective in color images. In particular, as indicated in Tables 2 to 6, our method achieved superior performance in the PSNR and SSIM compared with CsNet (the setting details of CsNet are presented in [27], Section V-C).

TABLE 7. The average ILNIQE and NIQMC results on DND [40].

| Method  | ILNIQE | NIQMC |
|---------|--------|-------|
| Proposed | 86.0828 | 2.8866 |
| Brooks [47] | 94.4446 | 2.7756 |
| VDNet [48] | 92.2914 | 2.7836 |

We selected two no-reference (NR) image quality assessment (IQA) metrics to evaluate the performance on the real noise dataset. NIQMC [46] represents the contrast in the image. The higher the value, the higher the contrast. ILNIQE [45] is a general NR IQA metric that can evaluate many distortion types. The smaller the value, the better the image quality. We utilize the proposed spatial local fusion strategy to fuse the preliminary denoised images by variational denoising network (VDNet) [48] and Brooks et al. proposed [47] and obtain the final results. The result was compared to the two state-of-the-art real image denoisers, VDNet [48] and that proposed by Brooks et al. [47]. Table 7 presents the NIQMC and ILNIQE [45] values and shows the effectiveness of our method on real image noise.

It can be observed from our experimental data that the proposed approach outperformed the state-of-the-art denoising methods on the objective indicators. The proposed method exhibited stability when noise images from different datasets.
indicating the expansibility and practicality of our method. In the next section, we analyze the visual comparisons of denoised images using different denoising algorithms.

C. VISUAL COMPARISONS

To compare the visual quality of the proposed methods with other compared methods intuitively, we present the results of several test images with abundant texture information.

Fig. 5 presents a visual comparison of the denoising results of the Lena image with a noise level of $\sigma = 40$ when using different denoising methods. We selected the right eye part of the denoised image, which was magnified to illustrate the visual differences for a detailed comparison. NCSR could preserve the details of the eyelashes better, but it was not effective in maintaining the smooth area around the eye. FFDNet successfully removed noise without losing the detailed structures and textures, and consequently achieved a smoother denoised image. The performance of the proposed approach was superior to that of the nine above mentioned methods in terms of structural-based decomposition. The results obtained by our method were much closer to the original in terms of the eyelash details and eyeball positions. Moreover, the proposed approach achieved competitive results in preserving the tenable details of the textural regions of the hat. The visual quality of the edge and smooth region was also satisfactory. The proposed method was superior to CsNet and boosted the PSNR to 31.56 dB.

Figs. 6 and 7 present the visual results of the different methods on two images, namely House and Boat, respectively. As indicated in Fig. 6 (d), when the image House was denoised with NCSR, the chimney region was effectively preserved, whereas the other methods failed to achieve this. It demonstrates that NCSR performs better for images with
Figure 8. Denoising results of a color image with noise level $\sigma = 30$: (a) original; (b) noisy image; (c) BM3D [13] / 29.09dB; (d) NCSR [10] / 26.72dB; (e) RED-Net [22] / 29.77dB; (f) FFDNet [23] / 29.37dB; (g) CsNet [27] / 29.85dB; (h) Proposed / 30.88dB.

regular and repeated structures, because images with regular and repetitive structures meet the nonlocal similarity prior effectively. However, images with irregular textures would weaken the advantages of such a specific prior, thereby leading to poor results. It can be observed that NCSR, TWSC, and PGPD failed to preserve the details of a white line on the wall. FFDNet, which is a training-based method, produced better results on images with irregular textures. Our method completely exploited the combined advantages of NCSR and FFDNet, thereby obtaining combined results. Fig. 7 demonstrates that our method preserved more image details from the noisy image. NCSR and BM3D tended to produce over-smooth edges and textures. The tiny masts of the boat could barely be recognized in the image denoised by NCSR; however, FFDNet was capable of effectively preserving the sharp edges and details of the boat masts. It is clear that the proposed method was visually superior to the other methods. Our method not only yielded visually pleasant results with sharp details of the boat masts and edges, but also boosted the PSNR to 31.59 dB.

Moreover, Fig 8 presents visual comparison of the denoising results of color image with a noise level of $\sigma = 30$. It can be seen that textures of pattern tiles have been extensively lost in the denoised image by NCSR; moreover, FFDNet produce over-smoothed textures and edges. By comparison, the result obtained by our method preserves the information in the original image to the greatest extent. And position of pattern on paper and textures of pattern are the clearest and closest to the original image.

V. DISCUSSION

The denoising effect and computational complexity are two important indexes for evaluating the performance of denoising algorithms. Unfortunately, researchers usually achieve a superior denoising effect at the cost of computational complexity. The computation time of our proposed method consists of the running time of NCSR and FFDNet and the fusion time of the preliminary denoised images; therefore, it is clear that the computation time of our method is longer than that of the single denoiser. However, in contrast to deep learning-based methods (e.g., CsNet [27]), the fusion stage in the proposed method does not involve the process of training on data, which is time-consuming. Table 8 lists the fusion time of the proposed method for processing ten commonly used images with various noise levels. The noise levels ranged from 20 to 60 with steps of 20. It can be observed that the fusion process costs very little time. Therefore, the computational complexity is determined primarily by the two denoisers. A good denoiser should realize a trade-off between the PSNR and the running time [24]. Here, we check whether the contribution of the consumed fusion time to the increase in the PSNR is worthwhile. Our method can bring a gain of 0.2–0.7 dB by PSNR on different noise levels. Therefore, a compromise for improving denoising performance at the expense of additional increased a little running time is worthwhile.

In addition, our purpose is to introduce a novel spatial local fusion strategy for improving the denoising performance. We carried out extensive experiments for many combinations of several typical denoisers using spatial fusion strategy. When the number of denoisers reaches or exceeds three, there is a slight improvement in the denoising effect; however, the computational complexity increase more, thereby making the process not very efficient. From Table 9, it can be seen that the speed of the combination of FFDNet and DnCNN and the combination of BM3D and FFDNet is faster than that of combination of NCSR and FFDNet; however, their denoising effect is worse than that of the combination of NCSR and FFDNet. Thus, we select NCSR and FFDNet as initial denoisers. The combination of NCSR and FFDNet can explore both internal and external information of preliminary denoised images owing to their strong complementary nature. And using the spatial local fusion strategy can preserve...
TABLE 8. Fusion time(s) of the proposed method for processing ten commonly used images with various noise levels.

| Method Type | σ=20 | σ=40 | σ=60 |
|-------------|------|------|------|
| Barbara     | 1.31 | 1.32 | 1.34 |
| Boat        | 1.28 | 1.29 | 1.29 |
| Cameraman   | 0.29 | 0.29 | 0.29 |
| Couple      | 1.29 | 1.29 | 1.28 |
| Hill        | 1.29 | 1.29 | 1.28 |
| House       | 1.29 | 1.29 | 1.28 |
| Lena        | 1.29 | 1.29 | 1.28 |
| Man         | 1.32 | 1.33 | 1.33 |
| Monarch     | 0.30 | 0.29 | 0.29 |
| Peppers     | 0.29 | 0.29 | 0.29 |
| Average     | 0.89 | 0.90 | 0.89 |

TABLE 9. Combination types among different denoisers.

| Combination Types | σ=10 | σ=20 | σ=30 | σ=40 | σ=50 | σ=60 |
|-------------------|------|------|------|------|------|------|
| BM3D [13] and DnCNN [20] | 33.33 | 29.52 | 27.73 | 26.52 | 25.71 | 25.01 |
| BM3D [13] and NCSR [10] | 34.48 | 30.24 | 28.36 | 27.05 | 26.08 | 25.40 |
| BM3D [13] and FFDNet [23] | 33.28 | 29.49 | 27.72 | 26.54 | 25.71 | 25.05 |
| NCSR [10] and FFDNet [23] | 34.72 | 30.67 | 28.79 | 27.47 | 26.56 | 25.83 |
| NCSR [10] and DnCNN [20] | 34.68 | 30.30 | 28.60 | 27.46 | 26.52 | 25.80 |
| FFDNet [23] and WNNM [4] | 33.29 | 29.56 | 27.80 | 26.66 | 25.81 | 25.14 |
| BM3D [13] and WNNM [4] | 33.08 | 29.22 | 27.48 | 26.31 | 25.46 | 24.81 |
| NCSR [10] and WNNM [4] | 34.52 | 30.35 | 28.47 | 27.14 | 26.24 | 25.49 |
| FFDNet [23] and DnCNN [20] | 33.33 | 29.60 | 27.84 | 26.69 | 25.86 | 25.18 |
| DnCNN [20] and WNNM [4] | 33.34 | 29.58 | 27.81 | 26.64 | 25.80 | 25.11 |

more textures and details. Our approach yields the superior denoising effect at a faster speed. Our strategy can easily be extended to fusing the preliminary denoised images generated by any several image denoisers; thus, one can select other two complementary denoising algorithms and utilize our strategy to boost the denoising effect. Currently, researchers have been pursuing to improve the denoising performance continually. In future work, we can select more efficient complementary denoising algorithms to generate preliminary denoised images according to the demand by giving priority to the denoising effect over, computational complexity or the balance between the two goals. In this study, our primary goal is to further boost denoising effect.

Numerous image denoising algorithms have been proposed in the past few decades that have delivered outstanding denoising performance. It is worth noting that it has become more difficult to achieve even a small performance improvement over state-of-the-art algorithms. However, compared with FFDNet, our method could boost the PSNR by 0.49–1.11 dB, which is a noticeable improvement, as shown in Table 5. Thus, our method can apply to an abundant of real-world applications (e.g., machine vision, medical diagnoses, and remote sensing). Specifically, in digital medical diagnoses, the characteristics in denoised images may be ignored by a single denoising algorithm. In contrast, the denoiser combination of NCSR and FFDNet is complementary and can explore both internal and external information of preliminary denoised images. It can produce a better denoised image with more feature information and provide more accurate data for medical treatment, which is crucial for extracting features of lesion regions, and it can help doctors to accurately diagnose diseases. Thus, it has great clinical significance. In fact, our method is not limited to noise models, such as Gaussian noise (GN), and can be extended to other types of noise if it is allowed by the constituent denoising algorithms. Table 7 presents the results on real noise datasets, demonstrating the practicality and generality of our method. Thus, our method can also apply to real-world application scenarios.

VI. CONCLUSION

We proposed an improved combination of NCSR and FFDNet (i.e., ICID) using the spatial local fusion strategy. The proposed ICID involves two major steps. The input images are pre-processed by NCSR and FFDNet to generate the image fusion sources. The denoiser combination of NCSR and FFDNet is complementary and can explore both internal and external information of preliminary denoised images. Therefore, the preliminary denoised image sources are decomposed into three components (strength component, structure component, and mean intensity component), which contain an abundance of detail information and edge features. The reconstructed image patches are placed back into the fused images after fusing the three components separately. The results of extensive experiments confirmed the effectiveness of the spatial local fusion strategy and demonstrated that the proposed approach consistently produces fused images of superior quality, including sharp structure, according to both the quantitative evaluation metrics and visual comparisons. Compared with existing single state-of-the-art denoising algorithms, the spatial local fusion strategy exhibits superior performance and provides improved fusion, both quantificationally and visually. Moreover, compared to CsNet, which uses fusion strategy, the improvement of denoising effect of our approach is more remarkable and the speed of the fusion process is faster. Our strategy provides a general solution for denoising applications, which can combine complementary strengths of denoisers. In future research, we will explore more efficient selection of complementary denoising algorithms and then use our strategy to fuse their denoised images, resulting in preserving more local details of denoised images.
REFERENCES

[1] M. V. Afonso and J. M. R. Sanches, “A total variation recursive space-variant filter for image denoising,” Digit. Signal Process., vol. 40, pp. 101–116, May 2015.

[2] S. Oh, H. Woo, S. Yun, and M. Kang, “Non-convex hybrid total variation for image denoising,” J. Vis. Commun. Image Represent., vol. 24, no. 3, pp. 332–344, Apr. 2013.

[3] P. Rodriguez and B. Wohlberg, “Efficient minimization method for a generalized total variation function,” IEEE Trans. Image Process., vol. 18, no. 2, pp. 322–332, Feb. 2009.

[4] 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.

[5] Y. Xie, S. Gu, Y. Liu, W. Zuo, W. Zhang, and L. Zhang, “Weighted schatten p-norm minimization for image denoising and background subtraction,” IEEE Trans. Image Process., vol. 25, no. 10, pp. 4842–4857, Oct. 2016.

[6] J. Xu, L. Zhang, D. Zhang, and X. Feng, “Multi-channel weighted nuclear norm minimization for real color image denoising,” in Proc. IEEE Int. Conf. Comput. Vis., Oct. 2017, pp. 1096–1104.

[7] M. Mahmoudi and G. Sapiro, “Fast image and video denoising via nonlocal means of similar neighborhoods,” IEEE Signal Process. Lett., vol. 12, no. 12, pp. 839–842, Dec. 2005.

[8] Y. Wu, B. Tracey, P. Natarajan, and J. P. Noonan, “James–Stein type center pixel weights for non-local means image denoising,” IEEE Signal Process. Lett., vol. 20, no. 4, pp. 411–414, Apr. 2013.

[9] R. Vignesh, B. T. Oh, and C.-C.-J. Kuo, “Fast non-local means (NLM) computation with probabilistic early termination,” IEEE Signal Process. Lett., vol. 17, no. 3, pp. 277–280, Mar. 2010.

[10] W. Dong, L. Zhang, G. Shi, and X. Li, “Nonlocally centralized sparse representation for image restoration,” IEEE Trans. Image Process., vol. 22, no. 4, pp. 1620–1630, Apr. 2013.

[11] J. Xu, L. Zhang, and D. Zhang, “A trilateral weighted sparse coding scheme for real-world image denoising,” in Proc. Eur. Conf. Comput. Vis. (ECCV), 2018, pp. 20–36.

[12] J. Xu, L. Zhang, W. Zuo, D. Zhang, and X. Feng, “Patch group based nonlocal self-similarity prior learning for image denoising,” in Proc. IEEE Int. Conf. Comput. Vis. (ICCV), Dec. 2015, pp. 244–252.

[13] 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.

[14] J. Mairal, F. Bach, J. Ponce, G. Sapiro, and A. Zisserman, “Non-local self-similarity prior learning for image denoising,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., Jun. 2014, pp. 2862–2869.

[15] M. Aharon, M. Elad, and A. Bruckstein, “K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation,” IEEE Trans. Signal Process., vol. 54, no. 11, pp. 4311–4322, Nov. 2006.

[16] U. Schmidt and S. Roth, “Shrinkage fields for effective image restoration,” in Proc. IEEE 12th Int. Conf. Comput. Vis. (ICCV), Sep. 2009, pp. 54–62.

[17] M. Aharon, M. Elad, and A. Bruckstein, “K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation,” IEEE Trans. Signal Process., vol. 54, no. 11, pp. 4311–4322, Nov. 2006.

[18] V. Jagannath and S. Roth, “Shrink fields for effective image restoration,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., Jun. 2014, pp. 2774–2781.

[19] Y. Chen and T. Pock, “Trainable nonlinear reaction diffusion: A flexible framework for fast and effective image restoration,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 39, no. 6, pp. 1256–1272, Jun. 2017.

[20] H. C. Burger, C. J. Schuler, and S. Harmeling, “Image denoising: Can plain neural networks compete with BM3D?” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., Jun. 2012, pp. 2392–2399.

[21] V. Jain and S. Seung, “Natural image denoising with convolutional networks,” in Proc. Adv. Neural Inf. Process. Syst., 2009, pp. 769–776.

[22] K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang, “Beyond a Gaussian denoiser: Residual learning of deep CNN for image denoising,” IEEE Trans. Image Process., vol. 26, no. 7, pp. 3142–3155, Jul. 2017.

[23] J. Xie, L. Xu, and E. Chen, “Image denoising and inpainting with deep neural networks,” in Proc. Adv. Neural Inf. Process. Syst., 2012, pp. 341–349.

[24] X. Mao, C. Shen, and Y.-B. Yang, “Image restoration using very deep convolutional encoder-decoder networks with symmetric skip connections,” in Proc. 30th Int. Conf. Neural Inf. Process. Syst., 2016, pp. 2810–2818.

[25] 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.

[26] K. Zhang, W. Zuo, S. Gu, and L. Zhang, “Learning deep CNN denoiser prior for image restoration,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., Jul. 2017, pp. 3929–3938.

[27] Y. Liu: Improved Combination of Image Denoisers Using Spatial Local Fusion Strategy

[28] Z. Yue, H. Yong, Q. Zhao, D. Meng, and L. Zhang, “Variational denoising network: Toward blind noise modeling and removal,” in Proc. Adv. Neural Inf. Process. Syst., 2019, pp. 1690–1701.
YIWEN LIU is currently pursuing the bachelor’s degree with the School of Information Engineering, Nanchang University. Her current research interests include image processing and neural networks.

SHAOPING XU received the M.S. degree in computer application from the China University of Geosciences, Wuhan, China, in 2004, and the Ph.D. degree in mechatronics engineering from Nanchang University, Nanchang, China, in 2010. He is currently a Professor with the Department of Computer Science, School of Information Engineering, Nanchang University. He has published more than 30 research articles. His current research interests include digital image processing and analysis, computer graphics, virtual reality, and surgery simulation. He serves as a Reviewer for several journals, including the IEEE TRANSACTIONS ON INSTRUMENTATION AND MEASUREMENT.

ZHENYU LIN received the B.Tech. degree from the Jingdezhen Ceramic Institute, China, in July 2018. She is currently pursuing the M.S. degree with the School of Information Engineering, Nanchang University, China, under the supervision of Prof. S. Xu. Her research interests include digital image processing and machine learning.