

PAPER

Nuclear Norm Minus Frobenius Norm Minimization with Rank Residual Constraint for Image Denoising

Hua HUANG[†], Yiwen SHAN^{††}, Chuan LI^{†††}, and Zhi WANG^{†††a)}, *Nonmembers*

SUMMARY Image denoising is an indispensable process of manifold high level tasks in image processing and computer vision. However, the traditional low-rank minimization-based methods suffer from a biased problem since only the noisy observation is used to estimate the underlying clean matrix. To overcome this issue, a new low-rank minimization-based method, called nuclear norm minus Frobenius norm rank residual minimization (NFRRM), is proposed for image denoising. The propose method transforms the ill-posed image denoising problem to rank residual minimization problems through excavating the nonlocal self-similarity prior. The proposed NFRRM model can perform an accurate estimation to the underlying clean matrix through treating each rank residual component flexibly. More importantly, the global optimum of the proposed NFRRM model can be obtained in closed-form. Extensive experiments demonstrate that the proposed NFRRM method outperforms many state-of-the-art image denoising methods.

key words: image denoising, nuclear norm minus Frobenius norm, rank residual, nonlocal self-similarity

1. Introduction

Image denoising, which aims to recover the original image from its noisy observation, is a problem of great challenge and popularity. It serves as an indispensable process of manifold high level tasks in image processing and computer vision [1]–[8]. Due to its importance and ill-posed nature, image denoising has drawn lots of research attentions in recent years. And a variety of state-of-the-art methods have been proposed based on low-rank minimization [9]–[15], sparse representation [16]–[18] and deep learning [19]–[23].

Benefited from the development of convex and nonconvex optimization, low-rank minimization has shown excellent performance in image denoising [24], [25]. Essentially, low-rank minimization is performed on each noise matrix that has distinctly low-rank properties. After denoising all of those matrices, the denoised image would be obtained. In practice, the noisy matrices can be constructed by gathering the similar structures in the observed image [26]. Mathematically, the low-rank minimization associated with the inputted noisy matrix, denoted as \mathbf{Y} , is

$$\min_{\mathbf{X}} \frac{1}{2\sigma_n^2} \|\mathbf{Y} - \mathbf{X}\|_F^2 + \lambda \cdot \text{rank}(\mathbf{X}), \quad (1)$$

where \mathbf{X} is the estimated matrix, $\|\cdot\|_F$ is the Frobenius norm, $\text{rank}(\cdot)$ is the rank function, σ_n is the standard deviation of noise, $\lambda > 0$ is the regularization parameter. Unfortunately, such a rank minimization problem (1) is NP-hard, which cannot be solved in polynomial time. Inspired by compressed sensing [27], [28], nuclear norm is used to replace the rank function, leading to the following nuclear norm minimization (NNM) problem:

$$\min_{\mathbf{X}} \frac{1}{2\sigma_n^2} \|\mathbf{Y} - \mathbf{X}\|_F^2 + \lambda \|\mathbf{X}\|_*, \quad (2)$$

where $\|\cdot\|_*$ is the nuclear norm. As pointed out by Fazel [29], the nuclear norm is the convex envelop, i.e., the “best” convex approximation, to the rank of matrix \mathbf{X} on the set $\{\mathbf{X} \in \mathbb{R}^{m \times n} : \|\mathbf{X}\|_2 \leq 1\}$, where $\|\mathbf{X}\|_2$ is the spectral norm. Moreover, Candès et al. [30] proved that the underlying matrix can be recovered with high accuracy and large probability. Solving the convex NNM problem has two advantages. First, a local optimum is indeed the global optimum. Second, there are a number of efficient algorithms to solve it in polynomial time, such as the interior point method [31], singular value thresholding [32], and accelerated proximal gradient with line search [33]. Albeit the theoretical guarantees and efficient solvers, NNM still has some limitations. As pointed out by [9], [34], the nuclear norm performs equal treatments on all singular values. This practice ignores the prior that singular values with different magnitude have different importance, and thus should be treated differently. Consequently, NNM would result in severe deviation between the recovered matrix and the optimum. To alleviate this problem, a number of nonconvex low-rank regularizers have been studied, including weighted norms [14], [15], truncated nuclear norm [35], [36], and capped nuclear norm [37], [38]. Those nonconvex regularizers give more flexible treatments on singular values. And their superiority over nuclear norm is validated in many studies. The weighted nuclear norm [14] sets the weights to be inversely proportional to the magnitude of singular values. Therefore, larger singular values, which quantify more information, are penalized less. The resultant weighted nuclear norm minimization (WNNM) method is applied to image denoising, achieving significant improvements over the original NNM. More importantly, the global optimum of WNNM model can be got in closed-form, which makes WNNM highly efficient.

Manuscript received December 11, 2023.

Manuscript revised February 26, 2024.

Manuscript publicized April 9, 2024.

[†]Information Construction and Management Center, Chongqing Vocational Institute of Engineering, Chongqing, China.

^{††}College of Computer and Information Science, Southwest University, Chongqing, China.

^{†††}Big Data and Intelligence Engineering School, Chongqing College of International Business and Economics, Chongqing, China.

a) E-mail: chiw@swu.edu.cn (Corresponding author)
DOI: 10.1587/transinf.2023EDP7265

However, Xie et al. [15] pointed out that weighted nuclear norm still tends to over-penalize the dominant singular values, which would result in WNNM obtaining biased solutions. On the same course, they proposed weighted Schatten p -norm minimization (WSNM), which achieved state-of-the-art denoising performance. However, solving WSNM is expensive since it no longer allows the closed-form solution and requires dedicated iterative solvers [39].

One common defect of the models aforementioned is that the low-rank matrix, i.e., \mathbf{X} in (1), is estimated only from its noisy version \mathbf{Y} . As the noise become stronger, it is more difficult to recognize the similar structures and gather into the noisy matrix. In other words, the low-rank nature of matrix \mathbf{Y} would be weakened. Consequently, the recovery would deviate from the optimum. To address this problem, dedicated constraints should be introduced. Rank residual constraint (RRC) [40] constructs a reference matrix for each noisy matrix. The reference matrices lead to a more accurate estimation to the underlying clean matrices. However, RRC model treats rank residual components equally, over-penalizing the dominant ones. To overcome this drawback, Zhang et al. [41] replaced the nuclear norm regularizer in RRC with weighted Schatten p -norm, and proposed the rank residual constraint with weighted Schatten p -norm (SRRC) model. SRRC model further improves the denoising performance over RRC. However, SRRC model requires iterative solvers [39], which might be expensive when recovering large images.

Considering the drawbacks of aforementioned models, in this paper we propose a new low-rank minimization model, called Nuclear norm minus Frobenius norm Rank Residual Minimization (NFRRM), and apply it to solve image denoising problem. The proposed model has two major advantages. First, it provides a more flexible treatment on different rank residual components, which indicates the underlying low-rank matrix can be estimated with more accuracy. Second, cheap closed-form solution is allowed [42], which means the proposed model can be solved with high efficiency. Through exploiting the framework of nonlocal self-similarity [26] prior, the proposed NFRRM model is applied to image denoising. Extensive experiments demonstrated that the proposed NFRRM method outperforms several state-of-the-art image denoising methods.

2. Related Works

2.1 Nonlocal Self-Similarity

As an ill-posed problem, image denoising has no unique solution. And the solutions vary discontinuously for small input changes. Therefore, prior knowledge should be introduced to regularize the solution space. The nonlocal self-similarity (NSS) [26] is one of the exemplary image prior. NSS describes the property that there spread many similar patterns across a natural image. Through gathering those similar structures, a flurry of low-rank matrices, called patch matrices, would be constructed. Low-rank minimization models

should be conducted on each patch matrix, estimating its underlying denoised version. Eventually, the denoised image would be obtained by concatenating all denoised matrices.

Figure 1 depicts the detailed procedure about generating a patch matrix from the observed image. Concretely, given a noisy image Y , a number of key patches are assigned across it. Those key patches are spaced an equal distance apart, and each spans $p \times p$ pixels. For each key patch, its k most similar neighbors (i.e., patches) are identified via k -Nearest Neighbors (k -NN) algorithm. The similar neighbors are found in a squared search region around the key patch, instead of the whole image. The distance between the key patch $\mathbf{K} \in \mathbb{R}^{p \times p}$ and its i th neighbor $\mathbf{P}_i \in \mathbb{R}^{p \times p}$ is measured by

$$d(\mathbf{K}, \mathbf{P}_i) = \|\mathbf{K} - \mathbf{P}_i\|_F, \quad (3)$$

where $i \in \{1, 2, \dots, S^2\}$, and S is the side length of the squared search region. After that, the chosen k most similar patches are stretched to column vectors, denoted as $\mathbf{y}_i \in \mathbb{R}^{p^2}$ for $i \in \{1, 2, \dots, k\}$. Those k vectors are stacked to form a patch matrix $\mathbf{Y} = \mathbf{C} + \mathbf{N} \in \mathbb{R}^{p^2 \times k}$, where \mathbf{C} , \mathbf{N} are the underlying clean matrix and the noise matrix, respectively. Note that the relationship between key patches and patch matrices is one-to-one. As patch matrix \mathbf{Y} groups similar structures, it has clear low-rank property. Therefore, low-rank minimization can be conducted on \mathbf{Y} . And the corresponding denoised version \mathbf{X} , which is desired to be as closed to the ground truth \mathbf{C} as possible, can be estimated. In summary, NSS prior bridges the gap between the low-rank minimization and image denoising problem, promoting the proposal of numerous low-rank based image denoising methods [13]–[15], [43], [44].

2.2 The Constraint of Rank Residual

As discussed in Introduction, defective results would be obtained by the traditional low-rank models since they estimate the underlying clean matrix only from its corresponding noisy observation. To address this issue, dedicated constraints should be introduced to regularize the solutions. In [40], Zha et al. proposed a novel low-rank minimization-based method with rank residual constraint (RRC) for image denoising and achieved the competitive results.

Mathematically, the rank residual, denoted as Γ , is defined as the difference between the the estimated matrix \mathbf{X} and the underlying ground truth matrix \mathbf{C} , that is $\Gamma \stackrel{\text{def}}{=} \mathbf{X} - \mathbf{C}$. However, the clean matrix \mathbf{C} is unavailable in image denoising problem. Therefore, accurate estimator should be introduced, for example, the nonlocal means [45]. Denote the estimated clean matrix as $\hat{\mathbf{C}}$. Then the rank residual can be modified as

$$\Gamma \stackrel{\text{def}}{=} \mathbf{X} - \hat{\mathbf{C}}. \quad (4)$$

Finally, a general form of low-rank minimization model with rank residual can be formulated as

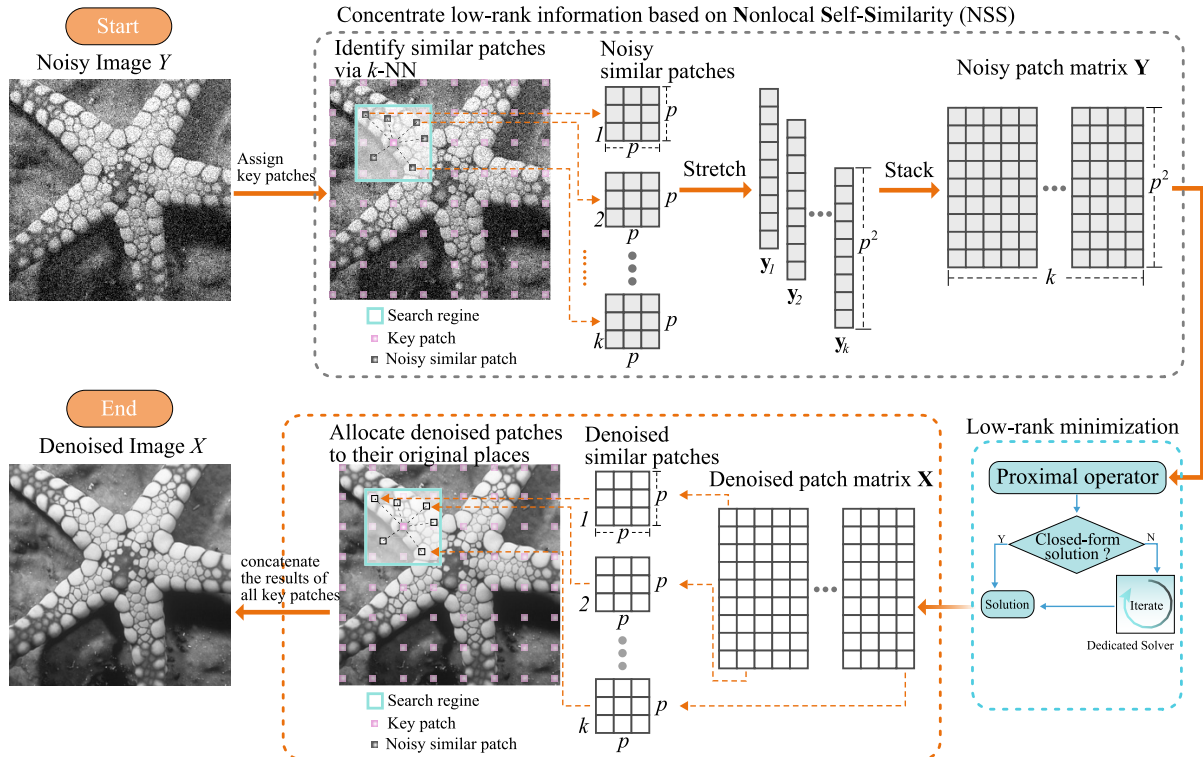


Fig. 1 Image denoising based on integrating NSS and low-rank minimization.

$$\min_{\mathbf{X}} \frac{1}{2} \|\mathbf{Y} - \mathbf{X}\|_F^2 + \lambda r(\mathbf{X} - \hat{\mathbf{C}}), \quad (5)$$

where $r(\cdot)$ denotes a low-rank regularizer that promotes the rank residual to be low-rank. It remains an open problem what regularizer fits for rank residual better [41].

2.3 Nuclear Norm Minus Frobenius Norm

Nuclear norm Minus Frobenius norm (NNFN) is a non-convex regularizer with multiple merits. Mathematically, NNFN of matrix $\mathbf{X} \in \mathbb{R}^{m \times n}$ is defined as

$$\begin{aligned} \|\mathbf{X}\|_{*-F} &= \|\mathbf{X}\|_* - \alpha \|\mathbf{X}\|_F \\ &= \sum_{i=1}^l \sigma_i(\mathbf{X}) - \alpha \left(\sum_{i=1}^l \sigma_i(\mathbf{X})^2 \right)^{\frac{1}{2}}, \end{aligned} \quad (6)$$

where $\alpha \geq 0$, $l = \min(m, n)$, and $\sigma_i(\mathbf{X})$ is the i th singular value of \mathbf{X} . As shown in Fig. 2, NNFN is adequate to treat singular values with much flexibility. More importantly, the optimization model with NNFN, which can be formulated as

$$\min_{\mathbf{X}} \frac{1}{2} \|\mathbf{Y} - \mathbf{X}\|_F^2 + \lambda \|\mathbf{X}\|_{*-F}, \quad (7)$$

allows the global optimum being obtained in closed-form [42], [46]. The NNFN-based low-rank models have achieved promising performance in matrix completion [46], recommendation system [46], and color image denoising [13].

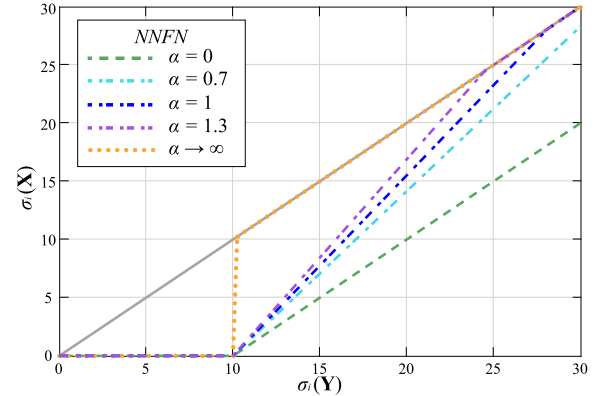


Fig. 2 The output pattern of the NNFN-based proximal operator with $\lambda = 10$. The horizontal axis represents the input, i.e., the i th singular value of the observed matrix \mathbf{Y} . The vertical axis represents the output, i.e., the i th singular value of the output matrix \mathbf{X} . Rational set of α will boost the flexibility of shrinkage. Concretely, when $\alpha = 0$, NNFN reduces to nuclear norm. Thus a constant shrinkage is imposed on $\sigma_i(\mathbf{X})$. As α becomes larger, the resultant NNFN shrinks less on the large singular values, or even preserves them. When $\alpha \rightarrow +\infty$, NNFN behaves like Capped nuclear norm, which only shrinks the small singular values.

3. The Proposed Model

3.1 Problem Formulation

Image denoising aims to recover the original clean image C from its noisy observation Y , which can be formulated as

$$Y = C + N, \quad (8)$$

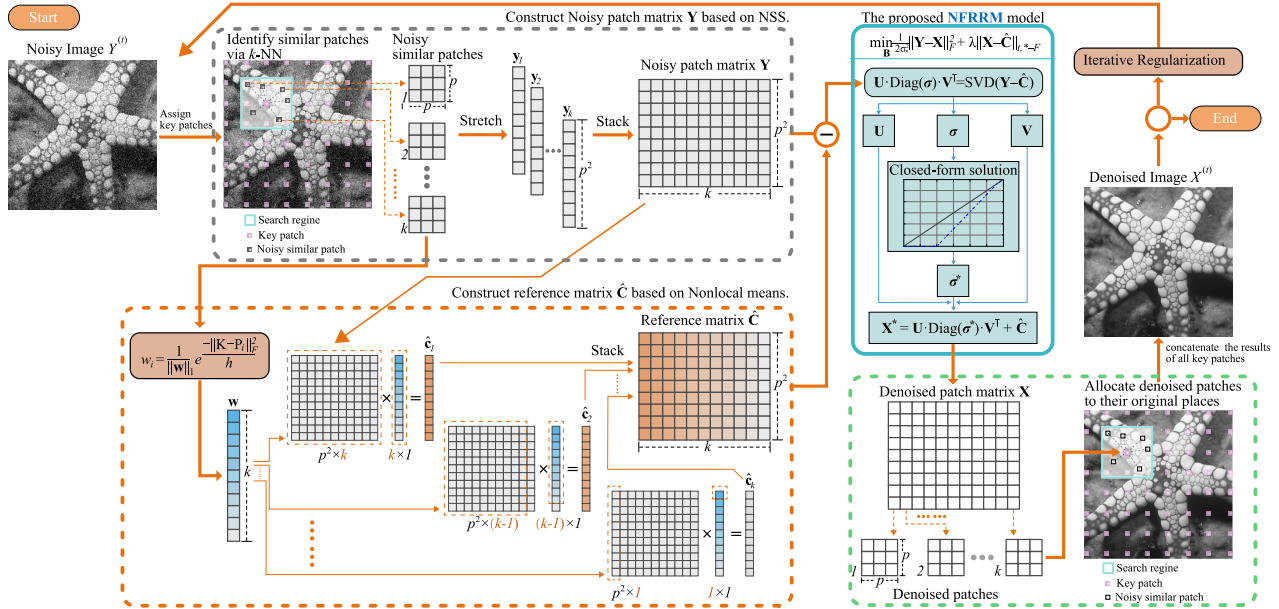


Fig. 3 The schematic framework of the proposed NFRRM method.

where $N \in \mathbb{R}^{m \times n}$ is the Gaussian white noise with each entry $N_{ij} \sim \mathcal{N}(0, \sigma_n^2)$. Problem (8) is severely ill-posed. Therefore, prior knowledge should be exploited to characterize the statistical features of the observed image Y . In this paper, NSS prior is introduced to concentrate the similar structures, form the low-rank matrices, and deliver them to the proposed model. As shown in Fig. 3, two matrices should be prepared for our model.

1) Patch matrix

Given the observed image $Y \in \mathbb{R}^{H \times W}$, we set M key patches across it. A key patch sizes $p \times p$ pixels. And the key patches are s pixels apart. In this paper, we set $s = \min(4, p - 1)$. Hence we have

$$M = \lceil (H - p)/s \rceil \times \lceil (W - p)/s \rceil, \quad (9)$$

where $\lceil \cdot \rceil$ is the ceil function. For each key patch, k most similar patches are identified via k -NN algorithm. The selected k similar patches are then scratched to column vectors, denoted as $\mathbf{y}_i \in \mathbb{R}^{p^2}$ for $i \in \{1, 2, \dots, k\}$. The vectors are stacked to form a patch matrix, denoted as $\mathbf{Y} \in \mathbb{R}^{p^2 \times k}$. The detailed procedure of generating the patch matrix can be found in Sect. 2.1.

2) Reference matrix

To calculate the rank residual in (4), a reference matrix should be constructed for each patch matrix. To recap, for each key patch, k most similar patches are extracted by k -NN. For each similar patch \mathbf{P}_i ($i \in \{1, 2, \dots, k\}$), its contribution, i.e., weight, to the reference matrix is

$$w_i = \frac{1}{\|\mathbf{w}\|_1} e^{-\frac{\|\mathbf{K} - \mathbf{P}_i\|_2}{h}}, \quad (10)$$

where $\mathbf{w} = [w_1, w_2, \dots, w_k]^T$, \mathbf{K} is the key patch, and h

is a constant. Intuitively, the weight w_i depends on the similarity between the key patch \mathbf{K} and the neighbor \mathbf{P}_i . And the similarity is modeled by a decreasing function of the Euclidean distance. After obtaining the weight vector \mathbf{w} , the reference matrix $\hat{\mathbf{C}}$ can be calculated. Concretely, the j th column of $\hat{\mathbf{C}}$ is given by ($j \in \{1, 2, \dots, k\}$)

$$\hat{\mathbf{C}}(1:p^2, j) = \mathbf{Y}(1:p^2, 1:k - j + 1) \cdot \mathbf{w}(1:k - j + 1), \quad (11)$$

where $\mathbf{Y} \in \mathbb{R}^{p^2 \times k}$ is the noisy patch matrix of key patch \mathbf{K} , the subscript in $\mathbf{w}(1:k - j + 1)$ means selecting the 1st through the p^2 th elements of \mathbf{w} . The mechanism in (11) stems from the nonlocal means [45]. And the constructed reference matrix $\hat{\mathbf{C}} \in \mathbb{R}^{p^2 \times k}$ is adequate to approximate the underlying clean patch matrix \mathbf{C} with sufficient accuracy. It is worth emphasizing that a key patch will correspond with only one patch matrix and one reference matrix.

To estimate the clean patch matrix \mathbf{X} with high accuracy and more efficiency, the Nuclear norm minus Frobenius norm Rank Residual Minimization (NFRRM) model is proposed. Mathematically, NFRRM model can be formulated as

$$\min_{\mathbf{X}} \frac{1}{2\sigma_n^2} \|\mathbf{Y} - \mathbf{X}\|_F^2 + \lambda \|\mathbf{X} - \hat{\mathbf{C}}\|_{s-F}, \quad (12)$$

where σ_n is the standard deviation of noise, λ is the regularization parameter, \mathbf{Y} and $\hat{\mathbf{C}}$ are the inputted patch matrix and reference matrix, respectively.

After the patch matrix $\mathbf{X} \in \mathbb{R}^{p^2 \times k}$ being estimated, it is decomposed back to k patches. Those denoised patches are then reverted back to their original places. Finally, the denoised image would be obtained by concatenating the results

all denoised patch matrices.

The above procedure is carried out several rounds in order to obtain better denoising performance. Denote the inputted image and outputted image at t -th iteration as $Y^{(t)}$ and $X^{(t)}$, respectively. To reduce the method noise, the following iterative regularization is adopted:

$$Y^{(t)} = X^{(t-1)} + \delta(Y - X^{(t-1)}), \quad (13)$$

where $t \in \mathbb{N}_+$, and $\delta \in [0, 1)$ controls the step-back in consecutive iterations.

3.2 Optimization

The optimization problem in (12) can be equivalently rewritten as follows:

$$\min_{\bar{\mathbf{X}}} \frac{1}{2} \|\bar{\mathbf{Y}} - \bar{\mathbf{X}}\|_F^2 + \lambda \sigma_n^2 \|\bar{\mathbf{X}}\|_{*-F}, \quad (14)$$

where $\bar{\mathbf{Y}} = \mathbf{Y} - \hat{\mathbf{C}}$, and $\bar{\mathbf{X}} = \mathbf{X} - \hat{\mathbf{C}}$. The following Theorem is proposed to show that the global optimum of problem (14) can be obtained in closed-form.

Theorem 1: Assume that $\bar{\mathbf{Y}}$ admits singular value decomposition (SVD) as $\mathbf{U}_{\bar{\mathbf{Y}}} \boldsymbol{\Sigma}_{\bar{\mathbf{Y}}} \mathbf{V}_{\bar{\mathbf{Y}}}^\top$, where $\boldsymbol{\Sigma}_{\bar{\mathbf{Y}}} = \text{Diag}(\boldsymbol{\sigma}(\bar{\mathbf{Y}}))$. Then the global optimum of problem (14) is

$$\mathbf{X}^* = \mathbf{U}_{\bar{\mathbf{Y}}} \text{Diag}(\boldsymbol{\rho}^*) \mathbf{V}_{\bar{\mathbf{Y}}}^\top + \hat{\mathbf{C}}, \quad (15)$$

with

$$\rho_i^* = \left(1 + \frac{\lambda \alpha \sigma_n^2}{\|\mathbf{s}\|_2}\right) \cdot s_i, \quad (16)$$

where $s_i = \max(\sigma_i(\bar{\mathbf{Y}}) - \lambda \sigma_n^2, 0)$.

Proof 1: Assume that $\bar{\mathbf{X}} \in \mathbb{R}^{p \times q}$ admits SVD as $\mathbf{U}_{\bar{\mathbf{X}}} \boldsymbol{\Sigma}_{\bar{\mathbf{X}}} \mathbf{V}_{\bar{\mathbf{X}}}^\top$, $l = \min(p, q)$. Denote $\lambda \cdot \sigma_n^2$ as Λ . Then the loss term in problem (14) can be rewritten as

$$\frac{1}{2} \|\bar{\mathbf{Y}} - \bar{\mathbf{X}}\|_F^2 = \frac{1}{2} (\|\bar{\mathbf{Y}}\|_F^2 - 2\langle \bar{\mathbf{Y}}, \bar{\mathbf{X}} \rangle + \|\bar{\mathbf{X}}\|_F^2). \quad (17)$$

According to the Von Neumann's trace inequality [47], we have

$$\langle \bar{\mathbf{Y}}, \bar{\mathbf{X}} \rangle \leq \text{Tr}(\boldsymbol{\sigma}(\bar{\mathbf{Y}})^\top \boldsymbol{\sigma}(\bar{\mathbf{X}})). \quad (18)$$

The equality occurs if and only if

$$\mathbf{U}_{\bar{\mathbf{Y}}} = \mathbf{U}_{\bar{\mathbf{X}}} \quad \text{and} \quad \mathbf{V}_{\bar{\mathbf{Y}}} = \mathbf{V}_{\bar{\mathbf{X}}}. \quad (19)$$

Then, for problem (14), we have

$$\min_{\bar{\mathbf{X}}} \frac{1}{2} \|\bar{\mathbf{Y}} - \bar{\mathbf{X}}\|_F^2 + \Lambda \|\bar{\mathbf{X}}\|_{*-F} \quad (20)$$

$$= \min_{\bar{\mathbf{X}}} \frac{1}{2} \|\bar{\mathbf{Y}}\|_F^2 - \text{Tr}(\boldsymbol{\sigma}(\bar{\mathbf{Y}})^\top \boldsymbol{\sigma}(\bar{\mathbf{X}})) + \frac{1}{2} \|\bar{\mathbf{X}}\|_F^2 + \Lambda \left(\sum_{i=1}^l \sigma_i(\bar{\mathbf{X}}) - \alpha \left(\sum_{i=1}^l \sigma_i(\bar{\mathbf{X}})^2 \right)^{\frac{1}{2}} \right) \quad (21)$$

$$= \min_{\bar{\mathbf{X}}} \sum_{i=1}^l \left(\frac{1}{2} \sigma_i(\bar{\mathbf{X}})^2 - \sigma_i(\bar{\mathbf{Y}}) \cdot \sigma_i(\bar{\mathbf{X}}) + \Lambda \cdot \sigma_i(\bar{\mathbf{X}}) \right) - \Lambda \alpha \left(\sum_{i=1}^l \sigma_i(\bar{\mathbf{X}})^2 \right)^{\frac{1}{2}} \quad (22)$$

Denote the objective function of (22) as $J(\boldsymbol{\sigma}(\bar{\mathbf{X}}))$. The minimum point of $J(\boldsymbol{\sigma}(\bar{\mathbf{X}}))$, denoted as $\boldsymbol{\rho}^* \in \mathbb{R}^l$, is given by the following first-order optimality condition:

$$\frac{\partial J}{\partial \boldsymbol{\sigma}(\bar{\mathbf{X}})} = \mathbf{0}, \quad (23)$$

where $\mathbf{0} \in \mathbb{R}^l$ is a zero vector. Formula (23) can be rewritten as

$$\left(1 - \frac{\Lambda \alpha}{\|\boldsymbol{\sigma}(\bar{\mathbf{X}})\|_2}\right) \boldsymbol{\sigma}(\bar{\mathbf{X}}) = \boldsymbol{\sigma}(\bar{\mathbf{Y}}) - \Lambda \mathbf{1}, \quad (24)$$

where $\mathbf{1} \in \mathbb{R}^l$ is a vector with all elements equal to 1. The solution of (24) is

$$\rho_i^* = \left(1 + \frac{\Lambda \alpha}{\|\mathbf{s}\|_2}\right) \cdot s_i, \quad (25)$$

where $s_i = \max(\sigma_i(\bar{\mathbf{Y}}) - \Lambda, 0)$. Therefore, the global optimum of $\bar{\mathbf{X}}$ is

$$\bar{\mathbf{X}}^* = \mathbf{U}_{\bar{\mathbf{Y}}} \text{Diag}(\boldsymbol{\rho}^*) \mathbf{V}_{\bar{\mathbf{Y}}}^\top. \quad (26)$$

Thus the global optimum of the original problem (12) is

$$\mathbf{X}^* = \bar{\mathbf{X}}^* + \hat{\mathbf{C}}. \quad (27)$$

Finally, the whole procedure of our image denoising method is summarized in Algorithm 1.

3.3 Complexity Analysis

We discuss the time complex of the steps in the inner *for*-loop of Algorithm 1. Step 6 costs $\mathcal{O}(S^2 \log S)$, where S is the side length of the squared search region. Step 7 costs

Algorithm 1: Image denoising via NFRRM.

Input: Noisy image Y ;

Output: Denoised image X ;

1 Initialize $X^0 = Y, T, s, k, \lambda, \alpha$;

2 **for** $i = 1 : T$ **do**

3 Iterative regularization $Y^{(t)} = X^{(t-1)} + \delta(Y - X^{(t-1)})$;

4 Assign M key patches in Y^t ;

5 **for** Each key patch \mathbf{K} **do**

6 k -NN: Extract k most similar neighbors of \mathbf{K} in a search region;

7 Construct the noisy patch matrix \mathbf{Y} ;

8 Construct the reference matrix $\hat{\mathbf{C}}$;

9 SVD: $\bar{\mathbf{Y}} = \mathbf{U} \text{Diag}(\boldsymbol{\sigma}(\bar{\mathbf{Y}})) \mathbf{V}^\top$;

10 Estimate the optimal \mathbf{X}^* by Theorem 1;

11 Decompose \mathbf{X}^* back to patches; revert them back to their original places;

12 **end**

13 **end**

Table 1 PSNR (dB) results of all competing methods.

Images	$\sigma_n = 20$								$\sigma_n = 30$							
	BM3D	GLIDE	OGLR	GSRC	RRC	SRRC	NSSRC	NFRRM	BM3D	GLIDE	OGLR	GSRC	RRC	SRRC	NSSRC	NFRRM
House	33.77	32.76	32.61	33.93	33.74	33.80	33.87	33.90	32.14	30.34	30.08	32.49	32.32	32.36	32.49	32.52
Straw	27.05	26.71	25.68	27.65	27.56	27.56	27.64	27.64	25.01	24.76	23.86	25.46	25.39	25.39	25.48	25.51
Lake	28.85	28.01	27.27	29.16	29.07	29.16	29.10	29.22	26.84	25.96	26.82	26.89	26.89	26.94	26.95	27.02
Starfish	29.56	29.03	27.53	30.09	30.01	30.05	30.07	30.17	27.54	26.92	25.32	28.02	27.98	27.99	28.01	28.04
Kodim	29.41	28.04	28.73	29.58	29.38	29.44	29.71	29.57	27.35	25.90	26.97	27.46	27.41	27.42	27.55	27.60
Monarch	30.43	29.51	28.38	30.99	31.02	31.03	30.88	31.09	28.40	27.50	26.42	28.80	28.80	28.81	28.78	28.90
Plants	32.68	31.29	30.37	32.84	32.82	32.80	32.86	32.88	30.63	28.95	28.26	30.87	30.90	30.79	30.79	31.05
Brodatz	26.64	25.53	26.56	27.25	27.20	27.26	27.39	27.33	24.45	23.35	24.41	25.15	25.05	25.13	25.22	
Parrots	32.30	31.13	30.65	32.28	32.39	32.45	32.46	32.59	30.30	29.47	28.47	30.26	30.52	30.60	30.78	
Boat	31.33	29.96	29.17	31.54	31.46	31.47	31.57	31.44	29.24	27.79	26.88	29.41	29.28	29.30	29.37	29.38
Average	30.20	29.20	28.70	30.53	30.46	30.50	30.56	30.58	28.19	27.09	26.59	28.48	28.45	28.47	28.51	28.60
Images	$\sigma_n = 40$								$\sigma_n = 50$							
	BM3D	GLIDE	OGLR	GSRC	RRC	SRRC	NSSRC	NFRRM	BM3D	GLIDE	OGLR	GSRC	RRC	SRRC	NSSRC	NFRRM
House	30.76	28.88	29.03	31.22	31.10	31.15	31.37	30.96	29.84	27.69	26.81	30.47	29.93	30.00	30.39	30.03
Straw	23.21	23.32	22.54	23.93	23.92	23.93	23.95	24.04	22.39	22.02	21.56	22.79	22.80	22.84	22.90	22.99
Lake	25.25	24.65	23.73	25.52	25.50	25.53	25.52	25.61	30.21	28.59	27.91	30.78	30.89	30.93	30.87	24.56
Starfish	25.89	25.39	23.81	26.53	26.47	26.48	26.54	26.55	24.89	24.10	22.93	25.36	25.36	25.36	25.44	25.40
Kodim	25.80	24.61	23.81	25.87	25.90	25.92	26.03	26.03	24.81	23.49	24.21	24.93	24.78	24.86	25.08	25.03
Monarch	26.78	26.13	25.41	27.35	27.34	27.35	27.32	27.43	25.72	25.07	23.66	26.25	26.22	26.23	26.28	26.30
Plants	29.00	27.79	26.30	29.42	29.51	29.36	29.38	29.60	27.93	26.61	26.62	28.25	28.32	28.21	28.27	28.49
Brodatz	22.20	21.85	23.25	23.41	23.49	23.55	23.66	23.68	21.81	20.66	22.25	22.24	22.20	22.26	22.48	22.53
Parrots	28.63	28.11	26.69	28.70	29.16	29.24	29.18	29.35	28.00	27.09	25.57	28.02	28.04	28.14	28.12	28.38
Boat	27.81	26.31	25.69	27.98	27.88	27.88	27.93	27.82	26.58	25.13	24.33	26.92	26.81	26.81	26.96	26.81
Average	26.53	25.70	25.03	26.99	27.03	27.04	27.09	27.11	25.63	24.55	23.99	25.97	25.90	25.91	26.04	26.05
Images	$\sigma_n = 75$								$\sigma_n = 100$							
	BM3D	GLIDE	OGLR	GSRC	RRC	SRRC	NSSRC	NFRRM	BM3D	GLIDE	OGLR	GSRC	RRC	SRRC	NSSRC	NFRRM
House	27.45	25.17	24.64	28.61	27.99	27.99	28.59	28.28	25.87	23.57	22.43	26.72	26.38	26.35	26.92	26.76
Straw	20.61	19.55	19.94	20.67	21.04	21.06	21.04	21.07	19.57	18.34	18.84	19.02	19.34	19.87	19.85	19.88
Lake	22.59	21.77	20.72	22.61	22.64	22.64	22.63	22.86	21.45	20.65	19.63	21.32	21.37	21.42	21.41	21.65
Starfish	22.93	22.04	21.05	23.32	23.37	23.38	23.41	23.55	21.85	20.92	19.76	21.86	21.98	22.05	22.06	22.26
Kodim	23.19	21.36	22.47	23.37	23.29	23.30	23.50	23.51	21.83	20.45	21.04	22.11	22.15	22.09	22.41	22.52
Monarch	23.72	22.71	21.73	24.35	24.24	24.23	24.28	24.50	22.31	20.80	19.21	22.87	22.76	22.76	22.82	23.11
Plants	26.09	24.82	23.66	26.18	26.40	26.28	26.21	26.61	24.70	23.74	22.30	24.18	24.91	24.93	24.92	25.24
Brodatz	19.94	18.43	20.29	20.21	20.10	20.18	20.28	20.37	18.70	16.98	18.77	19.07	19.14	19.12	19.25	19.34
Parrots	26.08	24.76	23.47	26.15	26.23	26.24	26.17	26.41	24.54	23.43	22.50	24.78	24.83	24.82	24.76	25.01
Boat	24.73	23.16	22.56	24.95	24.80	24.77	25.00	25.02	23.41	21.88	21.11	23.30	23.38	23.34	23.49	23.70
Average	23.73	22.38	22.05	24.04	24.01	24.01	24.11	24.22	22.42	21.08	20.56	22.52	22.62	22.67	22.79	22.95

$O(p^4k)$, where p is the side length of a patch and k is the number of most similar neighbors. Step 8 costs $O(p^2k^2)$ since there are $p^2(1 + 2 + \dots + k) = \frac{p^2}{2}k(k + 1)$ entries to be calculated. Step 9 (SVD) costs $O(p^2k^2)$. Step 10 and 11 cost $O(p^4k + p^2k^2)$ and $O(p^2k)$, respectively. Among them, the dominant cost lies in step 8. Therefore, the total time complexity of the proposed method is $O(p^2k^2 \cdot M \cdot T)$. And solving the proposed NFRRM model (i.e., step 9 and 10) costs $O(p^2k^2)$.

4. Experimental Results

To validate the effectiveness of the proposed NFRRM method, extensive experiments are implemented on image denoising. Seven state-of-the-art methods are chosen for comparison, including the block matching and 3d filtering (BM3D) [16], global image denoising (GLIDE) [48], optimal graph Laplacian regularization (OGLR) [49], group sparsity residual constraint with nonlocal priors (GSRC) [18], rank residual constraint (RRC) [40], SRRC [41], and Nonconvex Structural Sparsity Residual

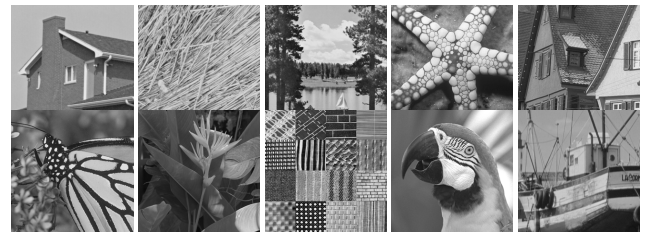


Fig. 4 Ten test images. The first row, from left-to-right: House, Straw, Lake, Starfish, and Kodim. The second row, from left-to-right: Monarch, Plants, Brodatz, Parrots, and Boat.

Constraint (NSSRC) [50]. The codes of all competing methods are obtained from their authors. The default parameters are kept. Both peak signal-to-noise ratio (PSNR) and structural similarity (SSIM) are considered to quantify the denoising performance. The higher PSNR and SSIM indicate the better quality of the denoised image. The competition between methods are conducted on ten widely used images, the thumbnails of which are shown in Fig. 4. The noisy observations are generated by zero-mean Gaussian with standard deviation $\sigma_n \in \{20, 30, 40, 50, 75, 100\}$.

Table 2 SSIM results of all competing methods.

Images	$\sigma_n = 20$								$\sigma_n = 30$							
	BM3D	GLIDE	OGLR	GSRC	RRC	SRRC	NSSRC	NFRRM	BM3D	GLIDE	OGLR	GSRC	RRC	SRRC	NSSRC	NFRRM
House	0.8726	0.8579	0.9156	0.8688	0.8665	0.8667	0.8680	0.8663	0.8480	0.8244	0.8832	0.8533	0.8529	0.8530	0.8525	0.8536
Straw	0.8962	0.8845	0.8545	0.9124	0.9120	0.9118	0.9120	0.9122	0.8320	0.8218	0.7780	0.8540	0.8562	0.8531	0.8547	0.8581
Lake	0.8798	0.8482	0.8685	0.8840	0.8809	0.8860	0.8845	0.8883	0.8286	0.7962	0.8268	0.8303	0.8323	0.8338	0.8345	0.8407
Starfish	0.8727	0.8541	0.8735	0.8770	0.8770	0.8761	0.8770	0.8777	0.8219	0.8045	0.8062	0.8327	0.8337	0.8333	0.8332	0.8351
Kodim	0.8645	0.7986	0.8498	0.8602	0.8594	0.8549	0.8694	0.8604	0.8029	0.7102	0.8040	0.7959	0.7935	0.7922	0.7973	0.8050
Monarch	0.9211	0.8986	0.9198	0.9246	0.9274	0.9270	0.9240	0.9279	0.8845	0.8585	0.8769	0.8939	0.8960	0.8957	0.8925	0.8978
Plants	0.8811	0.8492	0.8493	0.8835	0.8812	0.8807	0.8840	0.8818	0.8363	0.7938	0.7916	0.8453	0.8459	0.8412	0.8453	0.8498
Brodatz	0.9141	0.8778	0.8950	0.9235	0.9239	0.9223	0.9243	0.9225	0.8564	0.8059	0.8323	0.8715	0.8763	0.8699	0.8713	0.8762
Parrots	0.8985	0.8805	0.9120	0.8993	0.9007	0.8998	0.8997	0.9000	0.8705	0.8482	0.8677	0.8740	0.8765	0.8759	0.8718	0.8781
Boat	0.8897	0.8528	0.8773	0.8921	0.8893	0.8876	0.8927	0.8855	0.8392	0.7991	0.7971	0.8450	0.8427	0.8409	0.8437	0.8425
Average	0.8890	0.8602	0.8815	0.8925	0.8918	0.8913	0.8935	0.8923	0.8420	0.8063	0.8247	0.8496	0.8506	0.8489	0.8497	0.8537

Images	$\sigma_n = 40$								$\sigma_n = 50$							
	BM3D	GLIDE	OGLR	GSRC	RRC	SRRC	NSSRC	NFRRM	BM3D	GLIDE	OGLR	GSRC	RRC	SRRC	NSSRC	NFRRM
House	0.8282	0.7964	0.8582	0.8398	0.8395	0.8393	0.8415	0.8368	0.8150	0.7657	0.7819	0.8310	0.8248	0.8246	0.8300	0.8257
Straw	0.7459	0.7559	0.6967	0.7904	0.7983	0.7964	0.7900	0.8024	0.6810	0.6765	0.5957	0.7261	0.7461	0.7391	0.7339	0.7453
Lake	0.7809	0.7527	0.7381	0.7912	0.7921	0.7925	0.7928	0.8026	0.7440	0.7132	0.7183	0.7492	0.7571	0.7535	0.7581	0.7655
Starfish	0.7812	0.7550	0.7570	0.7957	0.7951	0.7950	0.7946	0.7974	0.7400	0.7035	0.7132	0.7540	0.7607	0.7599	0.7614	0.7611
Kodim	0.7243	0.6534	0.7411	0.7296	0.7219	0.7211	0.7284	0.7412	0.6762	0.6018	0.6981	0.6737	0.6777	0.6717	0.6861	0.6913
Monarch	0.8504	0.8255	0.8355	0.8620	0.8651	0.8645	0.8613	0.8686	0.8199	0.7941	0.7951	0.8308	0.8361	0.8352	0.8331	0.8422
Plants	0.7897	0.7534	0.7167	0.8138	0.8151	0.8061	0.8134	0.8196	0.7621	0.7193	0.7855	0.7814	0.7789	0.7707	0.7843	0.7877
Brodatz	0.7780	0.7337	0.7983	0.8165	0.8224	0.8173	0.8146	0.8248	0.7417	0.6772	0.7434	0.7401	0.7654	0.7577	0.7589	0.7709
Parrots	0.8413	0.8126	0.8294	0.8522	0.8555	0.8550	0.8502	0.8595	0.8298	0.7947	0.7790	0.8363	0.8372	0.8369	0.8327	0.8445
Boat	0.7994	0.7503	0.7611	0.8083	0.8051	0.8047	0.8053	0.8061	0.7584	0.7095	0.7237	0.7763	0.7738	0.7719	0.7765	0.7763
Average	0.7919	0.7589	0.7732	0.8099	0.8110	0.8092	0.8092	0.8159	0.7568	0.7156	0.7240	0.7699	0.7758	0.7721	0.7755	0.7811

Images	$\sigma_n = 75$								$\sigma_n = 100$							
	BM3D	GLIDE	OGLR	GSRC	RRC	SRRC	NSSRC	NFRRM	BM3D	GLIDE	OGLR	GSRC	RRC	SRRC	NSSRC	NFRRM
House	0.7601	0.7018	0.7339	0.8083	0.7952	0.7946	0.8068	0.8000	0.7203	0.6332	0.6330	0.7803	0.7655	0.7637	0.7810	0.7725
Straw	0.5344	0.4423	0.3696	0.5346	0.6227	0.6210	0.5885	0.5887	0.4115	0.3028	0.2214	0.3317	0.3689	0.5069	0.4672	0.4695
Lake	0.6675	0.6293	0.5962	0.6772	0.6822	0.6798	0.6767	0.6933	0.6070	0.5701	0.5263	0.6173	0.6233	0.6215	0.6218	0.6378
Starfish	0.6616	0.6119	0.6097	0.6712	0.6811	0.6797	0.6778	0.6868	0.5977	0.5461	0.5085	0.6006	0.6081	0.6132	0.6093	0.6208
Kodim	0.5864	0.4840	0.5967	0.6012	0.5932	0.5902	0.6025	0.6066	0.5171	0.4295	0.5352	0.5294	0.5287	0.5229	0.5412	0.5614
Monarch	0.7490	0.6978	0.7006	0.7779	0.7783	0.7778	0.7771	0.7869	0.6956	0.6311	0.5975	0.7302	0.7312	0.7300	0.7271	0.7409
Plants	0.6926	0.6479	0.5776	0.7188	0.7172	0.7110	0.7157	0.7270	0.6434	0.5945	0.5041	0.6642	0.6680	0.6683	0.6734	0.6805
Brodatz	0.6164	0.5162	0.6264	0.6055	0.6315	0.6151	0.6089	0.6215	0.5288	0.4032	0.4766	0.5189	0.5299	0.5236	0.5308	0.5551
Parrots	0.7750	0.7463	0.7089	0.8043	0.8028	0.8023	0.8004	0.8053	0.7330	0.6979	0.6798	0.7776	0.7729	0.7720	0.7712	0.7725
Boat	0.6876	0.6252	0.6183	0.7093	0.7003	0.6981	0.7093	0.7123	0.6348	0.5509	0.5557	0.6461	0.6428	0.6398	0.6521	0.6613
Average	0.6731	0.6103	0.6138	0.6908	0.7004	0.6970	0.6964	0.7028	0.6089	0.5359	0.5238	0.6196	0.6239	0.6362	0.6375	0.6472

Table 3 The settings of λ and α .

Noise	λ	α
$\sigma_n = 20$	0.95	2.400
$\sigma_n = 30$	0.93	2.125
$\sigma_n = 40$	0.92	2.110
$\sigma_n = 50$	0.93	1.855
$\sigma_n = 75$	0.93	1.000
$\sigma_n = 100$	0.94	0.900

For the proposed NFRRM method, parameters are set with respect to different noise levels. Concretely, the size of a patch $p \times p$ is set as 6×6 , 7×7 , 7×7 , 7×7 , 8×8 , 9×9 for $\sigma_n = 20, 30, 40, 50, 75, 100$, respectively. The number of similar neighbors k is set as 60, 60, 70, 80, 90, 100 for $\sigma_n = 20, 30, 40, 50, 75, 100$, respectively. The size of search region $S \times S = 25 \times 25$ for all noise levels. The iterative regularization parameter $\delta = 0.1$. The upper bound of iteration $T = 81$. And the settings of λ and α are listed in Table 3.

The PSNR results for all competing methods are presented in Table 1. The best results are highlighted in bold. The proposed NFRRM method achieves the highest PSNR in

42 out of 60 cases. Importantly, NFRRM outperforms other rank residual-based methods, i.e., RRC and SRRC, in all of 60 cases. The average improvements of NFRRM over RRC are 0.10 dB, 0.13 dB, 0.14 dB, 0.19 dB, 0.20 dB, 0.38 dB for $\sigma_n = 20, 30, 40, 50, 75, 100$, respectively. The improvement becomes more significant as the noise becomes larger. Moreover, NFRRM also outperforms the group sparse residual-based methods, i.e., GSRC and NSSRC.

The SSIM results are shown in Table 2. The proposed NFRRM achieves the best SSIM on 41 out of 60 cases. Concretely, NFRRM outperforms its counterparts, i.e., RRC and SRRC, in 53 out of 60 cases. The average improvements of NFRRM over SRRC are 1.6×10^{-3} , 2.6×10^{-3} , 5.8×10^{-3} , 6.3×10^{-3} , 1.02×10^{-2} , 1.83×10^{-2} for $\sigma_n = 20, 30, 40, 50, 75, 100$, respectively. When $\sigma_n = 20$, NFRRM fails to achieve the best average SSIM, narrowly losing to NSRRC. However, NFRRM achieves the highest average SSIM under all of the other noise levels. In summary, the proposed NFRRM method not only achieves the highest PSNR and SSIM in most cases, but also outperforms other state-of-the-art rank residual-based methods significantly.

The visual comparison between all competing meth-

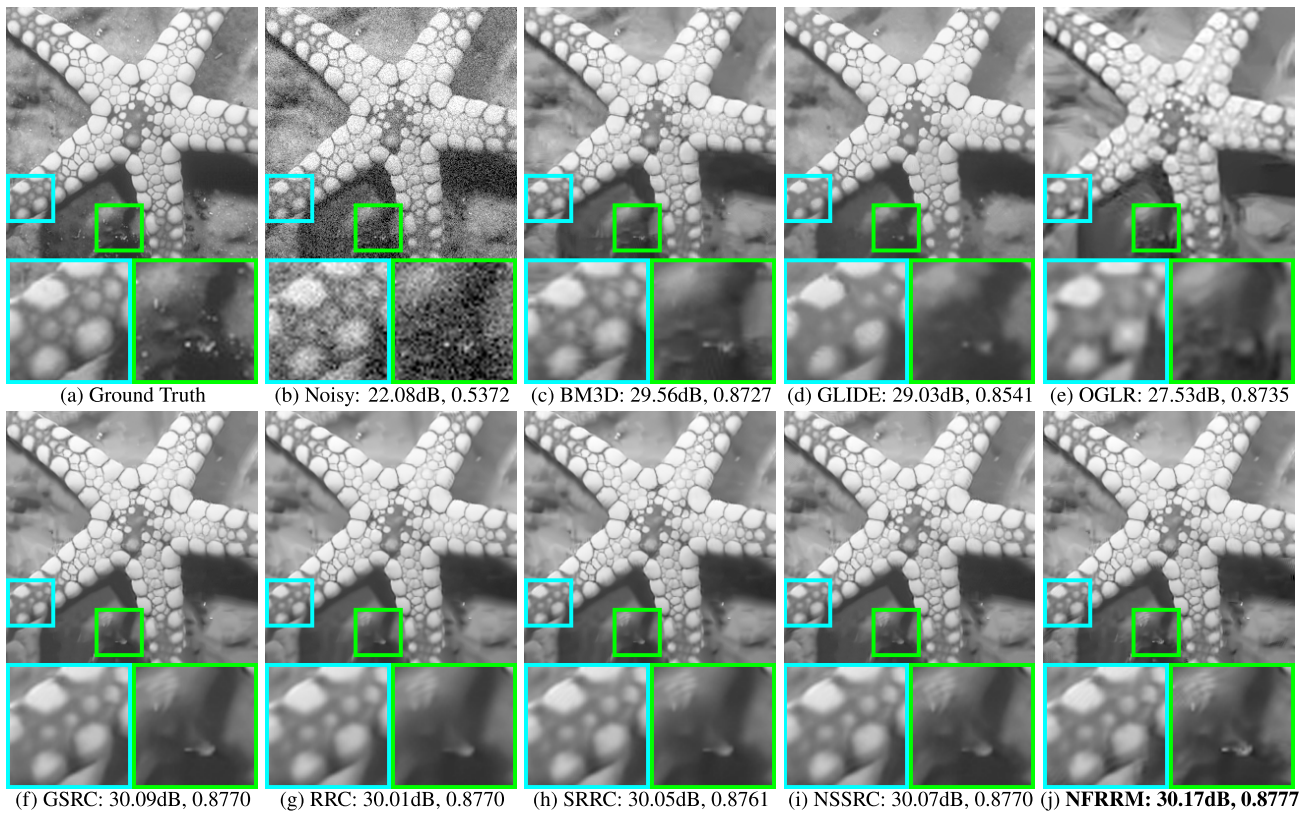


Fig. 5 Denoising results of *Starfish* with $\sigma_n = 20$. The subtitle format is “method: PSNR, SSIM”.

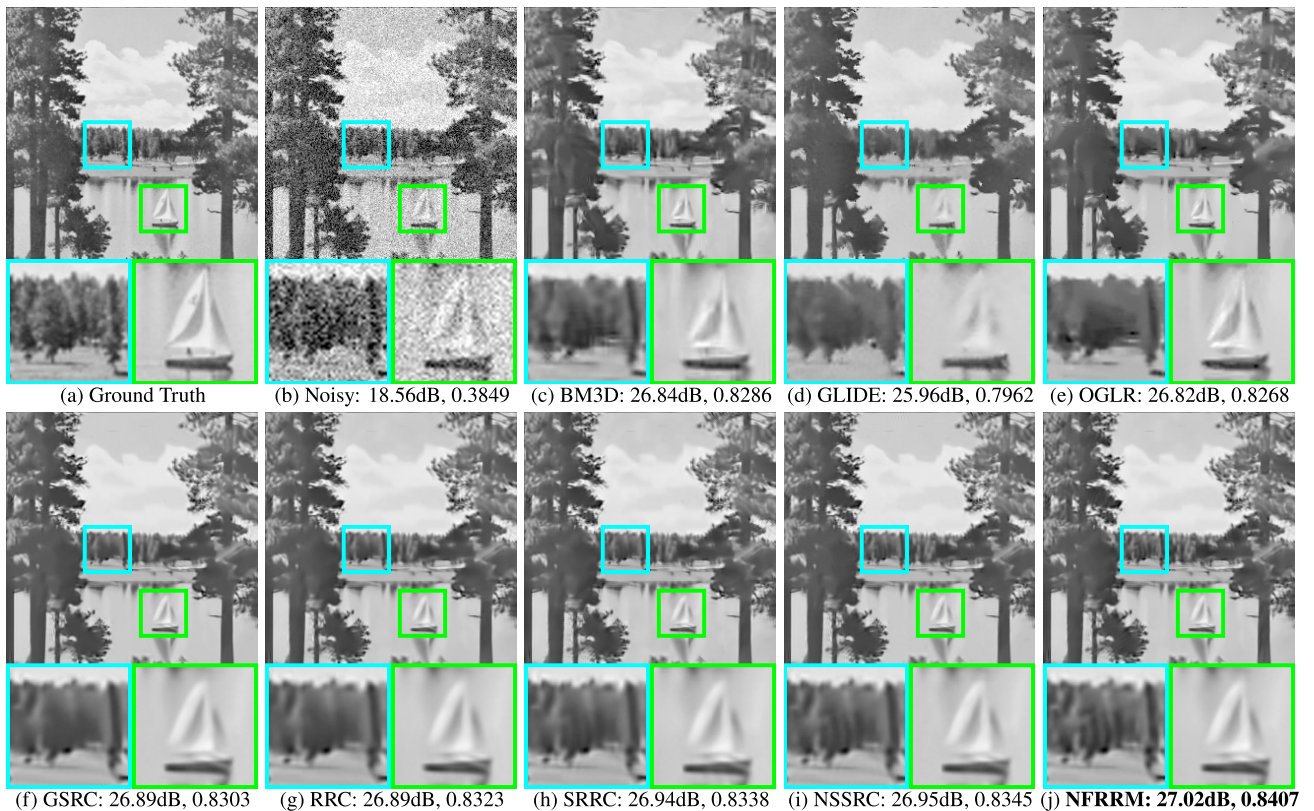


Fig. 6 Denoising results of *Lake* with $\sigma_n = 30$. The subtitle format is “method: PSNR, SSIM”.

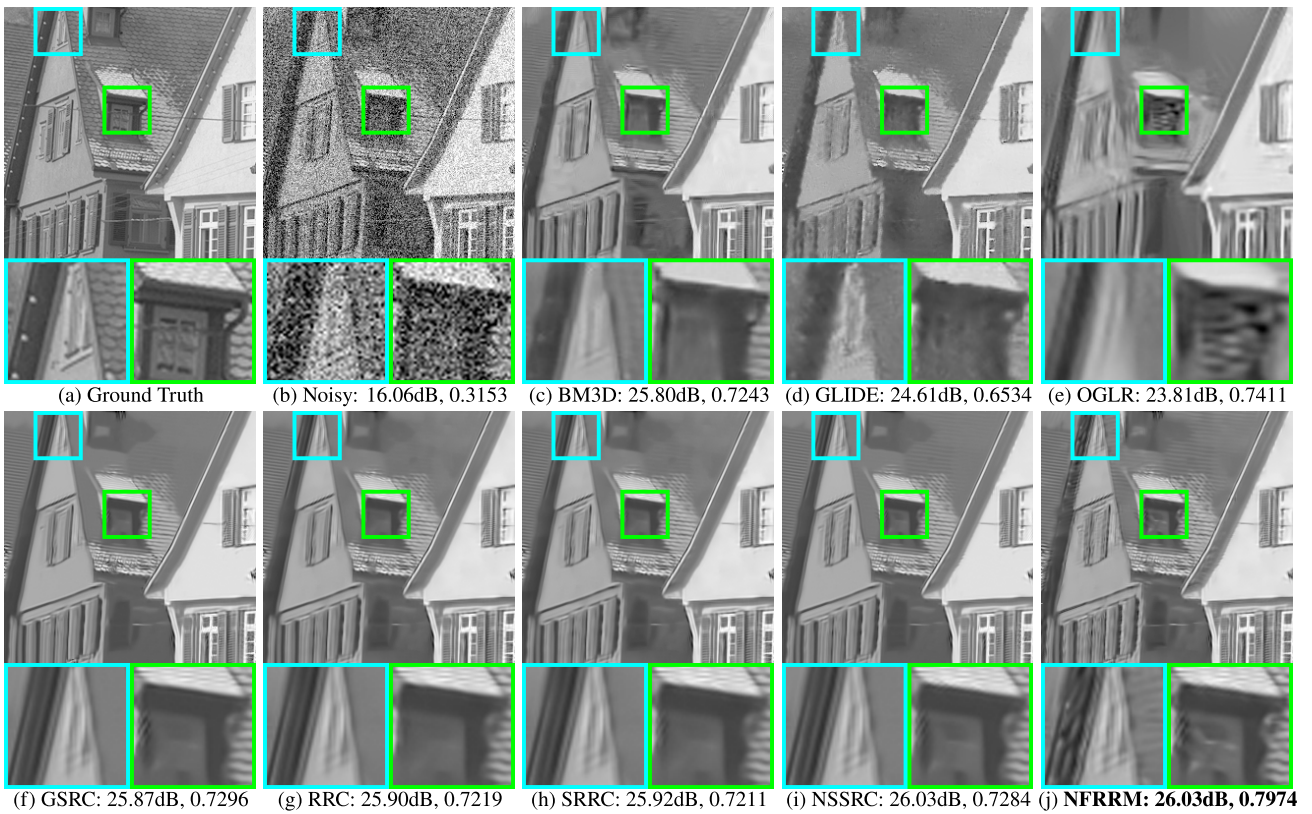


Fig. 7 Denoising results of *Kodim* with $\sigma_n = 40$. The subtitle format is “method: PSNR, SSIM”.

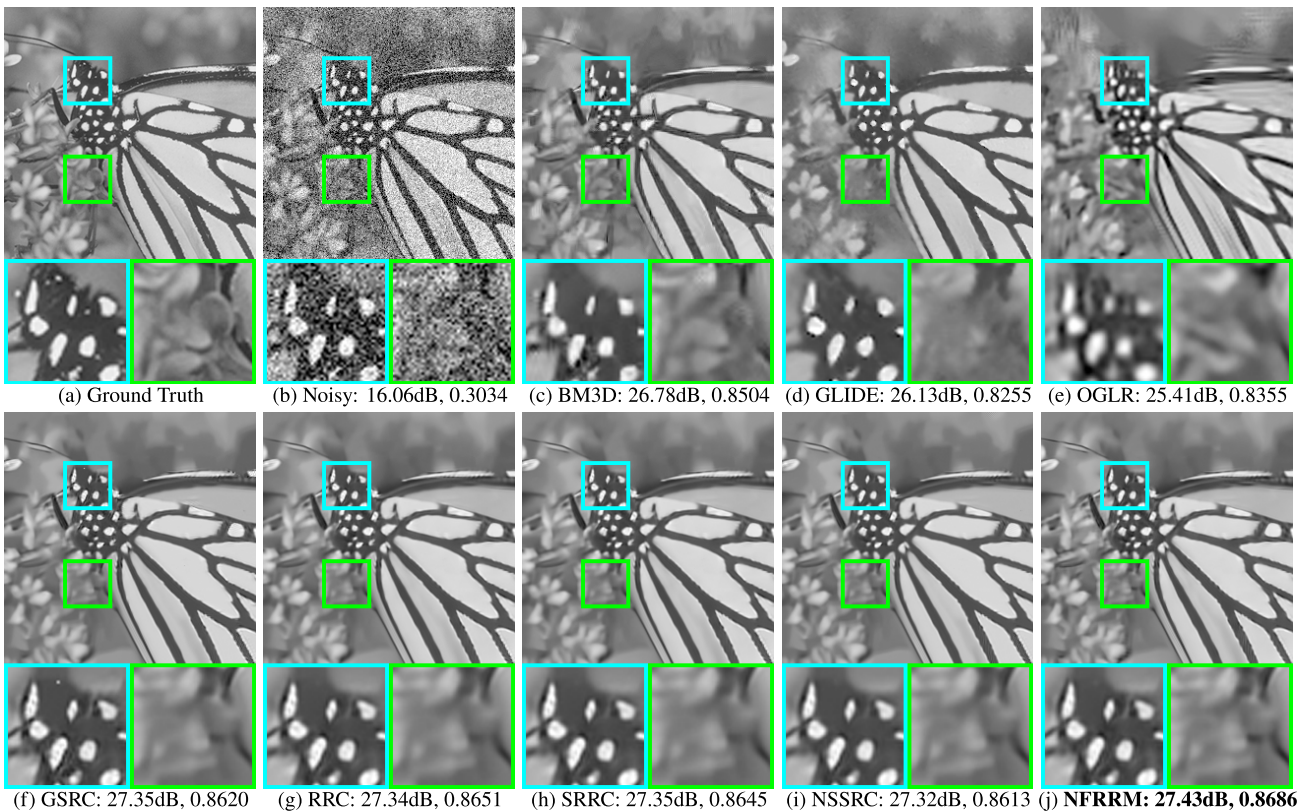


Fig. 8 Denoising results of *Monarch* with $\sigma_n = 40$. The subtitle format is “method: PSNR, SSIM”.

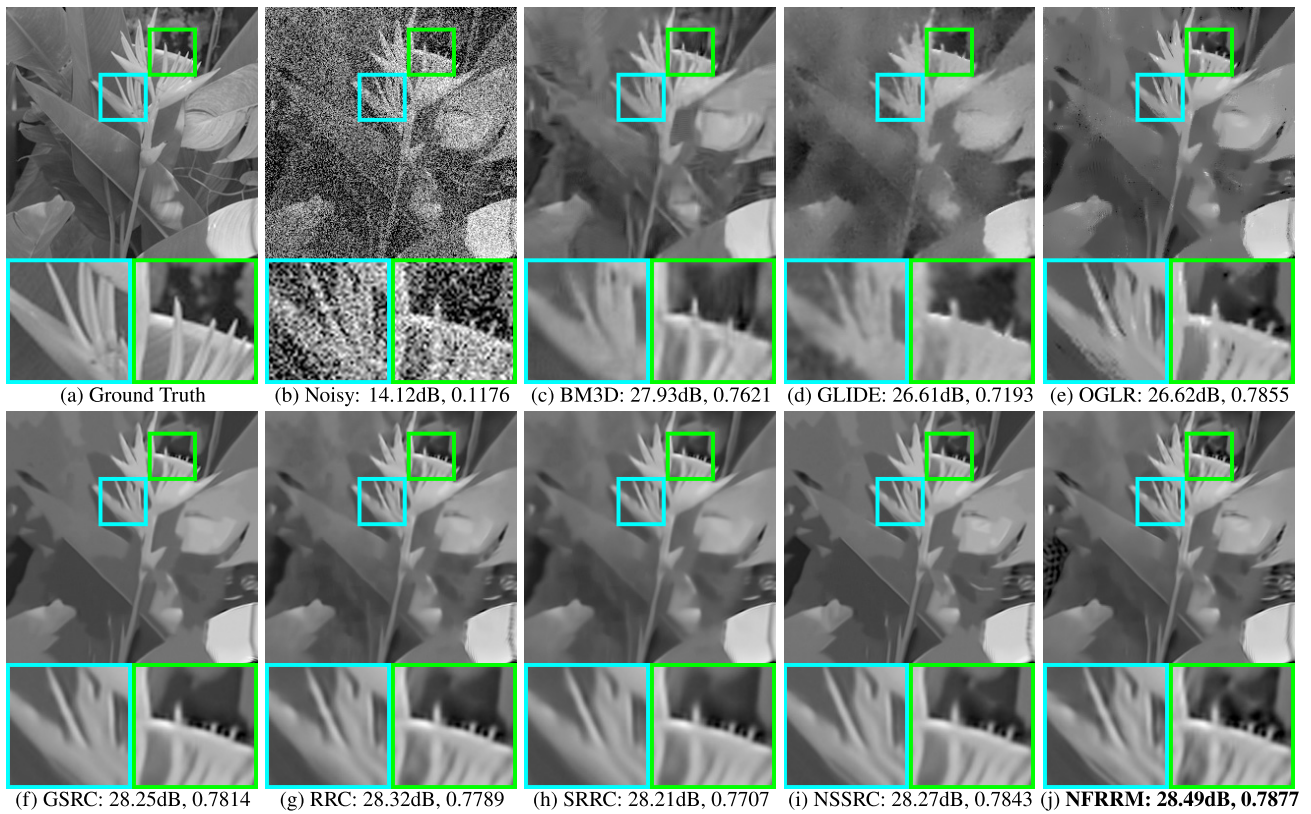


Fig. 9 Denoising results of *Plants* with $\sigma_n = 50$. The subtitle format is “method: PSNR, SSIM”.

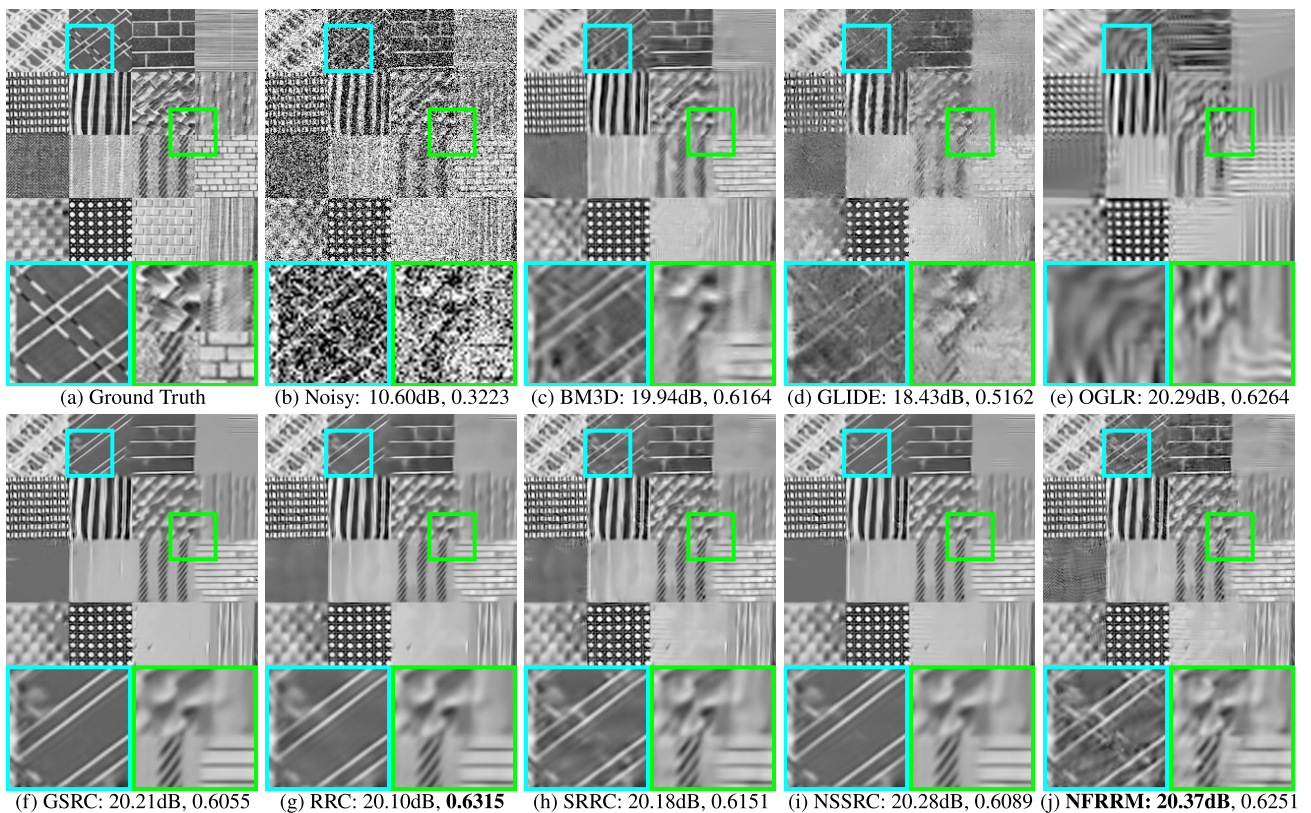


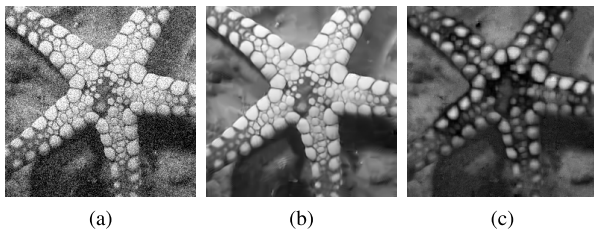
Fig. 10 Denoising results of *Brodatz* with $\sigma_n = 75$. The subtitle format is “method: PSNR, SSIM”.

Table 4 PSNR results of a BoostNet model (trained for $\sigma_n = 50$) and a NFRRM model ($\lambda = 0.93, \alpha = 1.855$).

#	$\sigma_n = 20$		$\sigma_n = 30$		$\sigma_n = 40$		$\sigma_n = 50$		$\sigma_n = 75$		$\sigma_n = 100$	
	BoostNet	NFRRM	BoostNet	NFRRM	BoostNet	NFRRM	BoostNet	NFRRM	BoostNet	NFRRM	BoostNet	NFRRM
1	31.32	33.91	31.19	32.51	30.93	31.28	30.44	30.03	18.46	27.85	14.13	25.84
2	22.70	27.64	22.92	25.51	23.13	24.04	23.15	22.99	17.33	21.17	13.77	20.03
3	24.67	29.23	24.77	27.03	24.80	25.65	24.57	24.56	18.47	22.71	14.35	21.31
4	25.46	30.16	25.48	28.04	25.41	26.56	25.15	25.40	18.11	23.42	14.15	21.90
5	25.33	29.57	25.35	27.61	25.35	26.07	25.27	25.03	18.11	23.32	14.08	22.05
6	27.27	31.10	27.21	28.91	27.00	27.48	26.64	26.30	18.35	24.30	14.16	22.55
7	29.02	32.91	28.93	31.06	28.65	29.68	28.19	28.49	18.90	26.44	14.45	24.84
8	21.51	27.33	21.59	25.22	21.71	23.68	21.76	22.53	17.23	20.34	13.67	19.10
9	28.85	32.60	28.55	30.79	28.07	29.44	27.44	28.38	18.46	26.23	14.22	24.52
10	27.49	31.48	27.54	29.41	27.49	27.95	27.16	26.81	18.13	24.72	14.00	23.01
Avg.	26.36	30.59	26.35	28.61	26.25	27.18	25.98	26.05	18.16	24.05	14.10	22.52

Table 5 SSIM results of a BoostNet model (trained for $\sigma_n = 50$) and a NFRRM model ($\lambda = 0.93, \alpha = 1.855$).

#	$\sigma_n = 20$		$\sigma_n = 30$		$\sigma_n = 40$		$\sigma_n = 50$		$\sigma_n = 75$		$\sigma_n = 100$	
	BoostNet	NFRRM	BoostNet	NFRRM	BoostNet	NFRRM	BoostNet	NFRRM	BoostNet	NFRRM	BoostNet	NFRRM
1	0.8354	0.8667	0.8354	0.8537	0.8352	0.8406	0.8356	0.8257	0.2056	0.7948	0.1018	0.7543
2	0.6771	0.9114	0.7010	0.8573	0.7280	0.7992	0.7509	0.7453	0.4447	0.6117	0.2573	0.5043
3	0.7401	0.8883	0.7473	0.8403	0.7559	0.8011	0.7637	0.7655	0.3506	0.6961	0.1908	0.6420
4	0.7367	0.8781	0.7418	0.8348	0.7492	0.7961	0.7597	0.7611	0.3535	0.6849	0.1888	0.6100
5	0.6557	0.8606	0.6612	0.8043	0.6664	0.7404	0.6803	0.6913	0.3351	0.6063	0.1787	0.5562
6	0.8567	0.9280	0.8563	0.8977	0.8537	0.8686	0.8502	0.8422	0.3560	0.7833	0.2079	0.7223
7	0.7786	0.8828	0.7796	0.8497	0.7788	0.8200	0.7835	0.7877	0.2224	0.7233	0.0954	0.6633
8	0.6966	0.9221	0.7055	0.8758	0.7213	0.8217	0.7461	0.7709	0.5475	0.6328	0.3680	0.5604
9	0.8469	0.9006	0.8454	0.8779	0.8431	0.8587	0.8435	0.8445	0.2516	0.8031	0.1193	0.7571
10	0.7798	0.8871	0.7852	0.8431	0.7909	0.8068	0.7973	0.7763	0.2817	0.7068	0.1416	0.6429
Avg.	0.7604	0.8926	0.7659	0.8535	0.7723	0.8153	0.7811	0.7811	0.3349	0.7043	0.1850	0.6413

**Fig. 11** The effects of λ . (a) $\lambda = 0.01$: PSNR = 18.56 dB, SSIM = 0.3929. (b) $\lambda = 1$: PSNR = 27.97 dB, SSIM = 0.8329. (c) $\lambda = 100$: PSNR = 11.43, SSIM = 0.5027.

ods are shown in Figs. 5–10. The proposed NFRRM removes the Gaussian noise completely in all of the chosen six cases. Meanwhile, the textures, edges, and image details are preserved well by NFRRM. Concretely, NFRRM presents a better recovery on textures of the sea bed in Fig. 5, the eave in Fig. 7, and the multiple kinds of surfaces in Fig. 10. In contrast, GSRC, RRC, and NSSRC over-smooth those textures, as shown in their highlighted windows. In the highlighted windows of Fig. 8, NFRRM reconstructs more structures on the mouthpart of the monarch and the flowers in the left-bottom corner. In contrast, the OGLR and GLIDE over-smooth the image and generate too much artifacts. In Fig. 6 and Fig. 9, NFRRM preserves the edges better. While BM3D, GLIDE, and OGLR failed to reconstruct those edges and details. In summary, the proposed NFRRM method presents strong denoising capability, recovering the images with promising visual qualities while achieving high PSNR and SSIM.

We also compare the proposed NFRRM model with a state-of-the-art deep learning model BoostNet [51]. For fairness, we use one NFRRM model ($\lambda = 0.93, \alpha = 1.855$) and one BoostNet model (trained for exactly $\sigma_n = 50$) to test all of noise levels $\sigma_n \in \{20, 30, 40, 50, 75, 100\}$. The PSNR and SSIM results are listed in Table 4 and Table 5, respectively. For $\sigma_n = 50$, NFRRM is narrowly inferior to BoostNet since it only obtains higher PSNR and SSIM in 4 images while BoostNet obtains 6. However, for $\sigma_n \in \{20, 30, 40, 75, 100\}$, the proposed NFRRM outperforms BoostNet significantly, achieving the higher PSNR and SSIM on all of the 50 test cases. Moreover, the denoising results of BoostNet (trained for $\sigma_n = 50$) become unreasonable when the tested noise levels become far from 50. As shown in Fig. 12, BoostNet (for $\sigma_n = 50$) over-smooths the images corrupted by $\sigma_n \in \{20, 30, 40\}$; while it remains too much noise in the images which are previously corrupted by $\sigma_n = \{75, 100\}$. On the contrary, the proposed NFRRM shows much more robustness since a same NFRRM model is adequate to obtain satisfactory denoising results in all of the six noise levels.

5. Sensitivity Analyses of Hyper-Parameters

In this section, we discuss the sensitivity of λ , α , and s . All of the tests are carried out on the image “Starfish” corrupted by noise $\sim \mathcal{N}(0, 30^2)$.

5.1 Regularization Parameter λ

Regularization parameter λ balances the influence of the in-

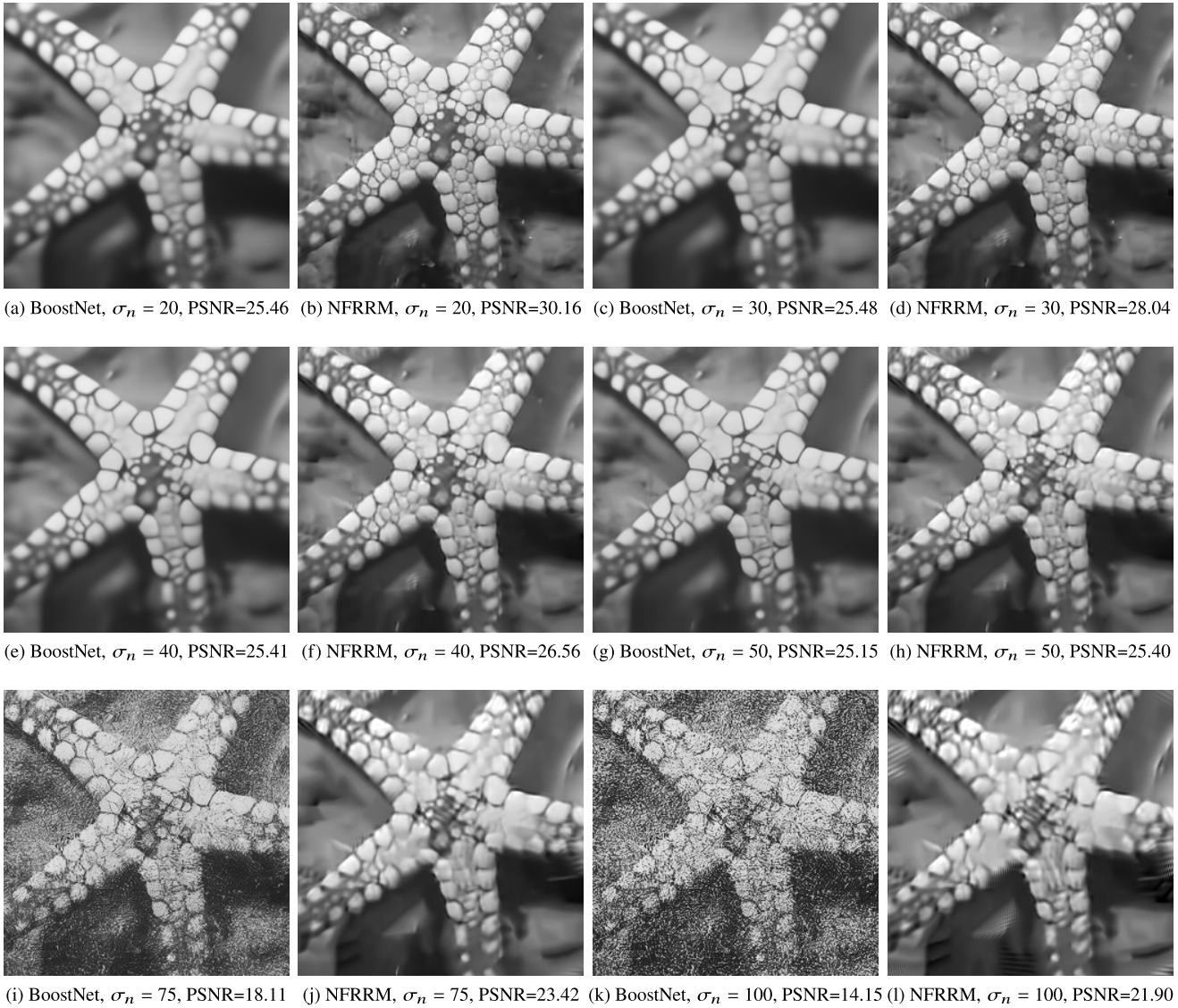


Fig. 12 Denoising results on “Starfish” of BoostNet model (trained for $\sigma_n = 50$) and NFRRM model ($\lambda = 0.93, \alpha = 1.855$).

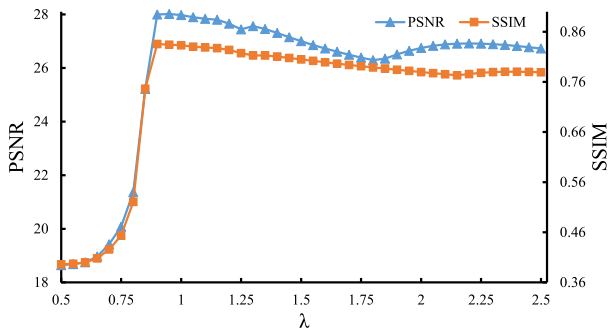


Fig. 13 The effects of λ on the denoising performance of NFRRM. The corrupted observation is generated by adding zero-mean Gaussian noise with standard deviation $\sigma_n = 30$ on image “Starfish”. Other parameters are all fixed.

delity term ($\|\cdot\|_F^2$) and the regularization term ($\|\cdot\|_{*-F}$). As the value of λ becomes too small, the NFRRM model would

mostly minimize the fidelity term, which makes the yielded image contain too much noise, as shown in Fig. 11 (a). As the value of λ becomes too large, the NFRRM model would mostly minimize the regularization term. Consequently, the yielded image would be over-smoothed and lose lots of details, as shown in Fig. 11 (c).

Figure 13 investigates the sensitivity of λ . As λ becomes larger, the denoising performance (PSNR and SSIM) improves sharply, and reaches the peak at $\lambda = 1$. Then the performance continues to decrease as $\lambda \rightarrow +\infty$. In this paper, the suitable values of λ are determined via experiments.

5.2 Parameter α

Parameter α controls the shrinkage on singular values. As it becomes too small ($\alpha \rightarrow 0$), the NNFN would behave like the nuclear norm, which is characterized by the green line in

Table 6 The best α for each image.

Group	Image	Plants	House	Parrots	Boat	Starfish	
Group1	$SSIM_0$	0.1176	0.1253	0.1351	0.1886	0.2322	Avg. ($SSIM_0$) = 0.1598
	Best α	0.95	1.55	1.20	2.50	2.00	Avg. (α) = 1.69 Std(α) = 0.4259
Group2	Image	Lake	Kodim	Monarch	Straw	Brodatz	
	$SSIM_0$	0.2365	0.2446	0.2451	0.3770	0.4834	Avg. ($SSIM_0$) = 0.3173
	Best α	2.60	1.50	6.45	2.80	4.00	Avg. (α) = 3.47 Std(α) = 1.688

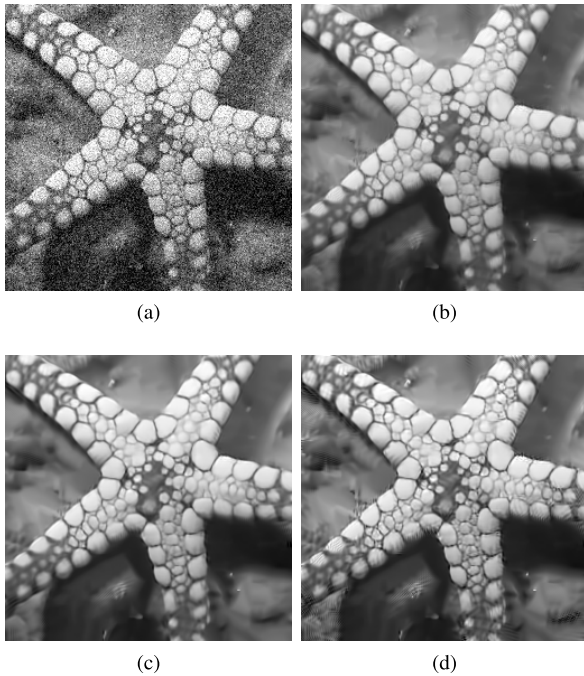


Fig. 14 The impact of α . (a) The corrupted observation ($\sigma_n = 20$), PSNR = 20.81 dB, SSIM = 0.4731. (b) $\alpha = 0$: PSNR = 27.63 dB, SSIM = 0.8292. (c) $\alpha = 2.4$: PSNR = 28.04 dB, SSIM = 0.8349. (d) $\alpha = 10$: PSNR = 27.38 dB, SSIM = 0.8258.

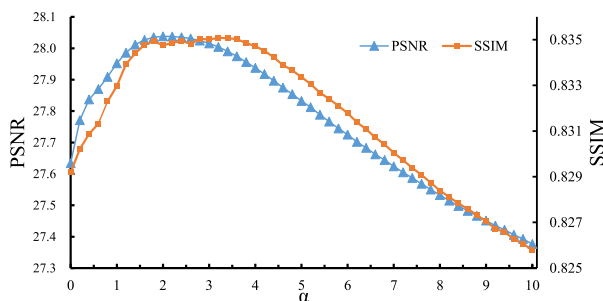


Fig. 15 PSNR and SSIM results on “Starfish” with different α .

Fig. 2. In that case, the leading singular values would be over-shrunk during the regularization process of NFRRM. Consequently, the output image would be over-smoothed, losing lots of details, as shown in Fig. 14 (b). As α becomes too large ($\alpha \rightarrow +\infty$), too many singular values would be zero-shrunk, which is characterized by the yellow line in Fig. 2. Consequently, the output image would contain noise and artifacts at the edges, as shown in Fig. 14 (d).

Figure 15 investigates the sensitivity of α . As can be seen, the best α dwells in the midst. In this paper, the suitable

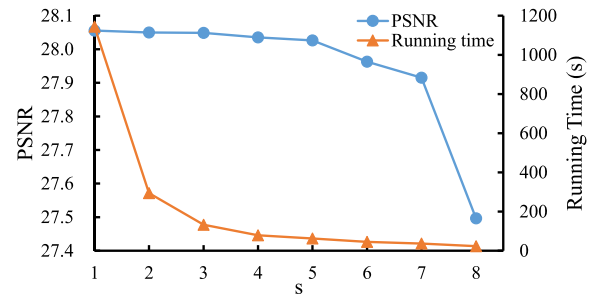


Fig. 16 Effects of s on denoising performance of NFRRM. Running time is in seconds.

α is also determined via experiments.

An empirical strategy is devised to choose α efficiently. If the corrupted image has a higher SSIM value (denoted as $SSIM_0$), a larger α is preferred, and vice versa. Considering the ten images corrupted by $\sigma_n = 50$. Their initial SSIM values, i.e., $SSIM_0$, are sorted in an ascending order. Then, those 10 images are divided into two groups according to their $SSIM_0$, as shown in Table 6. For each corrupted image, we search its “best α ”, i.e., an α that can make the NFRRM produce the highest SSIM on that image. And the “best α ” for each image are also listed in Table 6. As can be seen, the averages of “best α ” between two groups have a wide gap. And for the image in the second group, their “best α ” tend to be larger. Therefore, the aforementioned empirical strategy for choosing α can be obtained. This empirical strategy is efficient since it only uses the SSIM value of the corrupted image, which can be easily obtained.

5.3 The Interval of Key Patches s

Figure 16 investigates the sensitivity of s . As s becomes small ($8 \rightarrow 5$), the denoising performance would be improved significantly. That is attributed to that more key patches are generated, and hence the nonlocal self-similarity of the image is further exploited. However, as s becomes too small ($5 \rightarrow 1$), the running time would increase sharply, while the PSNR improves little. Therefore, the choice of s should be judicious in order to meet the balance between denoising performance and running time.

6. Conclusion

In this paper, a new image denoising method, called nuclear norm minus Frobenius norm rank residual minimization (NFRRM), was proposed. The proposed method converts the ill-posed image denoising problem to a nonconvex

optimization problem by exploiting the frameworks of image NSS prior. The sound NNFN regularizer was chosen to model the rank residual. With it, the proposed NFRMM model can treat different rank residual components with high flexibility. We derived that the global optimum of the nonconvex optimization problem can be easily obtained in closed-form. Extensive experimental results demonstrated that the proposed method outperforms several state-of-the-art image denoising methods.

Acknowledgements

This work was supported by the Science and Technology Research Program of Chongqing Municipal Education Commission under Grant KJQN202303413.

References

- [1] Z. Liu, D. Hu, Z. Wang, J. Gou, and T. Jia, "LatLRR for subspace clustering via reweighted Frobenius norm minimization," *Expert Syst. Appl.*, vol.224, Art. no.119977, Aug. 2023.
- [2] Y. Chang, L. Yan, T. Wu, and S. Zhong, "Remote sensing image stripe noise removal: From image decomposition perspective," *IEEE Trans. Geosci. Remote Sens.*, vol.54, no.12, pp.7018–7031, 2016.
- [3] G. Bi, G. Si, Y. Zhao, B. Qi, and H. Lv, "Haze removal for a single remote sensing image using low-rank and sparse prior," *IEEE Trans. Geosci. Remote Sensing*, vol.60, pp.1–13, Dec. 2022.
- [4] X. Ding, C. Shen, T. Zeng, and Y. Peng, "SAB Net: A semantic attention boosting framework for semantic segmentation," *IEEE Trans. Neural Netw. Learn. Syst.*, pp.1–13, 2022.
- [5] P.-H. Hsiao, F.-J. Chang, and Y.-Y. Lin, "Learning discriminatively reconstructed source data for object recognition with few examples," *IEEE Trans. Image Process.*, vol.25, no.8, pp.3518–3532, 2016.
- [6] X. Jiang and J. Lai, "Sparse and dense hybrid representation via dictionary decomposition for face recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol.37, no.5, pp.1067–1079, May 2015.
- [7] S. Liao, A.K. Jain, and S.Z. Li, "Partial face recognition: Alignment-free approach," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol.35, no.5, pp.1193–1205, May 2013.
- [8] S. Shen, "Accurate multiple view 3D reconstruction using patch-based stereo for large-scale scenes," *IEEE Trans. Image Process.*, vol.22, no.5, pp.1901–1914, May 2013.
- [9] Z. Wang, Y. Liu, X. Luo, J. Wang, C. Gao, D. Peng, and W. Chen, "Large-scale affine matrix rank minimization with a novel nonconvex regularizer," *IEEE Trans. Neural Netw. Learn. Syst.*, vol.33, no.9, pp.4661–4675, Sept. 2022.
- [10] Z. Wang, W. Wang, J. Wang, and S. Chen, "Fast and efficient algorithm for matrix completion via closed-form 2/3-thresholding operator," *Neurocomputing*, vol.330, pp.212–222, Feb. 2019.
- [11] Z. Wang, D. Hu, X. Luo, W. Wang, J. Wang, and W. Chen, "Performance guarantees of transformed Schatten-1 regularization for exact low-rank matrix recovery," *Int. J. Mach. Learn. Cybern.*, vol.12, pp.3379–3395, June 2021.
- [12] Z. Wang, C. Gao, X. Luo, M. Tang, J. Wang, and W. Chen, "Accelerated inexact matrix completion algorithm via closed-form q -thresholding ($q = 1/2, 2/3$) operator," *Int. J. Mach. Learn. Cybern.*, vol.11, pp.2327–2339, April 2020.
- [13] Y. Shan, D. Hu, Z. Wang, and T. Jia, "Multi-channel nuclear norm minus Frobenius norm minimization for color image denoising," *Signal Process.*, vol.207, Art. no.108959, June 2023.
- [14] S. Gu, Q. Xie, D. Meng, W. Zuo, X. Feng, and L. Zhang, "Weighted nuclear norm minimization and its applications to low level vision," *Int. J. Comput. Vis.*, vol.121, pp.183–208, July 2017.
- [15] 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, Aug. 2016.
- [16] 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, July 2007.
- [17] M. Elad and M. Aharon, "Image denoising via sparse and redundant representations over learned dictionaries," *IEEE Trans. Image Process.*, vol.15, no.12, pp.3736–3745, 2006.
- [18] Z. Zha, X. Yuan, B. Wen, J. Zhou, and C. Zhu, "Group sparsity residual constraint with non-local priors for image restoration," *IEEE Trans. Image Process.*, vol.29, pp.8960–8975, 2020.
- [19] 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, Feb. 2017.
- [20] 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, May 2018.
- [21] M. Zhao, G. Cao, X. Huang, and L. Yang, "Hybrid transformer-CNN for real image denoising," *IEEE Signal Process. Lett.*, vol.29, pp.1252–1256, May 2022.
- [22] F. Jia, L. Ma, Y. Yang, and T. Zeng, "Pixel-attention CNN with color correlation loss for color image denoising," *IEEE Signal Process. Lett.*, vol.28, pp.1600–1604, July 2021.
- [23] Y. Song, Y. Zhu, and X. Du, "Grouped multi-scale network for real-world image denoising," *IEEE Signal Process. Lett.*, vol.27, pp.2124–2128, Nov. 2020.
- [24] T. Liu, D. Hu, Z. Wang, J. Gou, and W. Chen, "Hyperspectral image denoising using nonconvex fraction function," *IEEE Geosci. Remote Sens. Lett.*, vol.20, pp.1–5, Art. no.5508105, Aug. 2023.
- [25] Y. Shi, T. Liu, D. Hu, C. Li, and Z. Wang, "Nonconvex regularization with multi-weighted strategy for real color image denoising," *Int. J. Intell. Syst.*, vol.2023, Art. no.8813500, Sept. 2023.
- [26] S. Wang, L. Zhang, and Y. Liang, "Nonlocal spectral prior model for low-level vision," *Asi. Conf. Comput. Vis.*, pp.231–244, Nov. 2012.
- [27] D.L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol.52, no.4, pp.1289–1306, 2006.
- [28] E.J. Candes and M.B. Wakin, "An introduction to compressive sampling," *IEEE Signal Process. Mag.*, vol.25, no.2, pp.21–30, 2008.
- [29] M. Fazel, *Matrix rank minimization with applications*, Ph.D. thesis, Stanford University, 2002.
- [30] E.J. Candès, X. Li, Y. Ma, and J. Wright, "Robust principal component analysis?," *J. ACM*, vol.58, no.3, Art. no.11, pp.1–37, 2011.
- [31] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, Cambridge, U.K., 2004.
- [32] J.-F. Cai, E.J. Candès, and Z. Shen, "A singular value thresholding algorithm for matrix completion," *SIAM J. Optim.*, vol.20, no.4, pp.1956–1982, March 2010.
- [33] K. Toh and S. Yun, "An accelerated proximal gradient algorithm for nuclear norm regularized linear least squares problems," *Pac. J. Optim.*, vol.6, no.615–640, p.15, 2010.
- [34] F. Nie, H. Huang, and C. Ding, "Low-rank matrix recovery via efficient Schatten p -norm minimization," *AAAI Conf. Artif. Intell.*, vol.26, no.1, pp.655–661, 2021.
- [35] Y. Hu, D. Zhang, J. Ye, X. Li, and X. He, "Fast and accurate matrix completion via truncated nuclear norm regularization," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol.35, no.9, pp.2117–2130, 2013.
- [36] T.-H. Oh, Y.-W. Tai, J.-C. Bazin, H. Kim, and I.S. Kweon, "Partial sum minimization of singular values in robust PCA: Algorithm and applications," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol.38, no.4, pp.744–758, April 2016.
- [37] Q. Sun, S. Xiang, and J. Ye, "Robust principal component analysis via capped norms," *19th ACM SIGKDD Int. Conf. Knowl. Discov. Data Mining*, New York, NY, USA, pp.311–319, Association for Computing Machinery, 2013.
- [38] F. Zhang, Z. Yang, Y. Chen, J. Yang, and G. Yang, "Matrix com-

- pletion via capped nuclear norm," *IET Image Process.*, vol.12, no.6, pp.959–966, 2018.
- [39] W. Zuo, D. Meng, L. Zhang, X. Feng, and D. Zhang, "A generalized iterated shrinkage algorithm for non-convex sparse coding," *IEEE Int. Conf. Comput. Vis.*, pp.217–224, 2013.
- [40] Z. Zha, X. Yuan, B. Wen, J. Zhou, J. Zhang, and C. Zhu, "From rank estimation to rank approximation: Rank residual constraint for image restoration," *IEEE Trans. Image Process.*, vol.29, pp.3254–3269, 2020.
- [41] T. Zhang, D. Wu, and X. Mo, "The rank residual constraint model with weighted Schatten p -norm minimization for image denoising," *Circuits Syst. Signal Process.*, vol.42, pp.4740–4758, 2023.
- [42] Y. Lou and M. Yan, "Fast L_1 - L_2 minimization via a proximal operator," *SIAM J. Sci. Comput.*, vol.74, no.2, pp.767–785, 2018.
- [43] J. Xu, L. Zhang, D. Zhang, and X. Feng, "Multi-channel weighted nuclear norm minimization for real color image denoising," *IEEE Int. Conf. Comput. Vis.*, pp.1105–1113, 2017.
- [44] X. Huang, B. Du, and W. Liu, "Multichannel color image denoising via weighted Schatten p -norm minimization," *Int. Joint Conf. Artif. Intell.*, pp.637–644, 2020.
- [45] A. Buades, B. Coll, and J.-M. Morel, "A non-local algorithm for image denoising," *IEEE Conf. Comput. Vis. Pattern Recognit.*, vol.2, pp.60–65, 2005.
- [46] Y. Wang, Q. Yao, and J. Kwok, "A scalable, adaptive and sound nonconvex regularizer for low-rank matrix learning," *Int. World Wide Web Conf.*, pp.1798–1808, 2021.
- [47] L. Mirsky, "A trace inequality of John von Neumann," *Monatshefte für mathematik*, vol.79, pp.303–306, 1975.
- [48] H. Talebi and P. Milanfar, "Global image denoising," *IEEE Trans. Image Process.*, vol.23, no.2, pp.755–768, 2014.
- [49] J. Pang and G. Cheung, "Graph Laplacian regularization for image denoising: Analysis in the continuous domain," *IEEE Trans. Image Process.*, vol.26, no.4, pp.1770–1785, 2017.
- [50] Z. Zha, X. Yuan, B. Wen, J. Zhang, and C. Zhu, "Nonconvex structural sparsity residual constraint for image restoration," *IEEE Trans. Cybern.*, vol.52, no.11, pp.12440–12453, 2022.
- [51] D.M. Vo, T.P. Le, D.M. Nguyen, and S.-W. Lee, "BoostNet: A boosted convolutional neural network for image blind denoising," *IEEE Access*, vol.9, pp.115145–115164, 2021.



Chuan Li received the M.S. degree in computer science and technology from Southwest University in 2013, Chongqing, China. He is currently a Ph.D. student in the College of Computer and Information Science, Southwest University, Chongqing, China. His current research interests include data mining, Education Informatization.



Zhi Wang received the B.S. and M.S. degrees in computer science and engineering, and the Ph.D. degree in statistics from the Southwest University, Chongqing, China, in 2005, 2009, and 2020, respectively. He is currently an Associate Professor with the College of Computer and Information Science, Southwest University, Chongqing, China. He has published more than 20 academic articles in reputable journals and conferences, including *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *IEEE Transactions on Neural Networks and Learning Systems*, *IEEE Geoscience and Remote Sensing Letters*, *Neurocomputing*, *Signal Processing*, *Expert Systems With Applications*, *International Journal of Machine Learning and Cybernetics*, *International Journal of Intelligent Systems*, *IJCNN*, *PRICAI*, *HPCC*, and so on. His research interests include low-rank matrix/tensor recovery, machine learning, and convex/nonconvex optimization. He is a member of the IEEE.



Hua Huang received the M.S. degree in computer science and engineering from the Chongqing University, Chongqing, China, in 2014. He is currently an Associate Professor with the Information Construction and Management Center, Chongqing Vocational Institute of Engineering, Chongqing, China.



Yiwen Shan is currently working toward the B.S. degree in the College of Computer and Information Science, Southwest University, Chongqing, China. His research focuses on machine learning.