Beyond Low-rankness: Guaranteed Matrix Recovery via Modified Nuclear Norm

Jiangjun Peng[1,2], **Yisi Luo**[3], **Xiangyong Cao**[4*], **Shuang Xu**[1,2], **Deyu Meng**[3*]
[1] School of Mathematics and Statistics, Northwestern Polytechnical University, Xi'an 710129, China
[2] Shenzhen Research Institute of Northwestern Polytechnical University, Shenzhen 518057, China
[3] School of Mathematics and Statistics and Ministry of Education Key Lab of Intelligent Networks and Network Security, Xi'an Jiaotong University, Xi'an 710049, China
[4] School of Computer Science and Technology, Xi'an Jiaotong University, Xi'an 710049, China
{pengjj, xs}@nwpu.edu.cn, yisiluo1221@foxmail.com, {caoxiangyong,dymeng}@mail.xjtu.edu.cn

Abstract

The nuclear norm (NN) has been widely explored in matrix recovery problems, such as Robust PCA and matrix completion, leveraging the inherent global low-rank structure of the data. In this study, we introduce a new modified nuclear norm (MNN) framework, where the MNN family norms are defined by adopting suitable transformations and performing the NN on the transformed matrix. The MNN framework offers two main advantages: (1) it jointly captures both local information and global low-rankness without requiring trade-off parameter tuning; (2) under mild assumptions on the transformation, we provide theoretical recovery guarantees for both Robust PCA and MC tasks—an achievement not shared by existing methods that combine local and global information. Thanks to its general and flexible design, MNN can accommodate various proven transformations, enabling a unified and effective approach to structured low-rank recovery. Extensive experiments demonstrate the effectiveness of our method. Code and supplementary material are available at https://github.com/andrew-pengjj/modified_nuclear_norm.

1 Introduction

The nuclear norm (NN), serving as a convex relaxation of the matrix rank, is widely applied to various matrix optimization problems due to its ability to effectively preserve the global low-rank property of data while possessing many desirable theoretical properties [Candes and Recht, 2012; Candès *et al.*, 2011; De Silva and Lim, 2008; Recht *et al.*, 2010; Gu *et al.*, 2014]. The data property is typically encoded by a regularizer term $\mathcal{R}(\cdot)$ and incorporated into the following energy function (1):

$$\min_{\mathbf{X}} E(\mathbf{X}, \mathbf{Y}) := L(\mathbf{X}, \mathbf{Y}) + \lambda \mathcal{R}(\mathbf{X}), \tag{1}$$

where $\mathbf{X}$ and $\mathbf{Y}$ represent the data to be recovered and the observed data, respectively. The function $L(\mathbf{X}, \mathbf{Y})$ is the loss function, which can be in the form of ℓ_1 and ℓ_2 norm.

*Corresponding authors

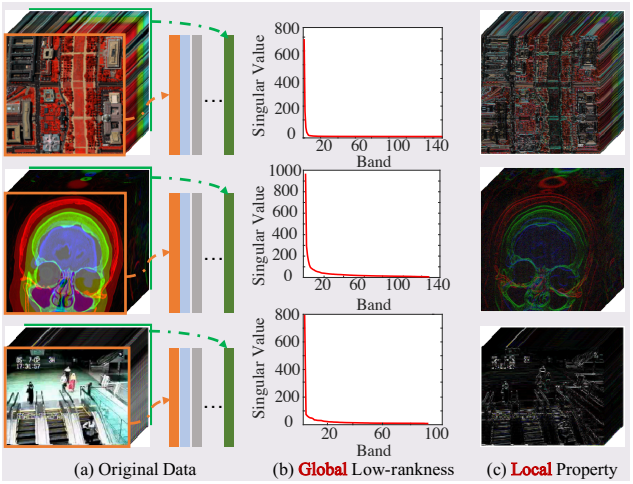


Figure 1: Demonstrations of the global low-rank property and local prior of data. (a) From top to bottom, showcasing the hyperspectral data, CT images and surveillance video data, as well as a matrix expanded along the third dimension; (b) Singular value curve of the mode-3 unfolding matrices of the data in (a); (c) Difference maps are obtained by applying difference operators. Difference maps have sparsity, which means that the original data has local smoothness.

However, in many cases, alongside the global low-rankness of data, local prior information also plays a crucial role, especially in data related to computer vision tasks such as hyperspectral image [Wang *et al.*, 2017], multispectral image [Xie *et al.*, 2017], CT images [Wang *et al.*, 2023], and surveillance videos [Peng *et al.*, 2022b], as shown in Fig. 1. Efficiently integrating global and local information to obtain a model with theoretical guarantees is an important issue for restoration tasks. Previous methods of integrating global and local information typically involve constructing multiple regularizer terms and combining them additively [Wang *et al.*, 2017; He *et al.*, 2015; Xue *et al.*, 2021; He *et al.*, 2017; Shang *et al.*, 2023; Chen *et al.*, 2023]. Taking global low-rankness and local smoothness properties as an example, the traditional fused regularization method is given by:

$$\mathcal{R}(\mathbf{X}) := \|\mathbf{X}\|_* + \beta \|\mathbf{X}\|_{\text{TV}}, \tag{2}$$

where $\|\mathbf{X}\|_{\text{TV}}$ is the total variation (TV) regularizer to encode the sparsity of difference map of the original data [Rudin *et*

al., 1992], and β is the trade-off parameter need to be fine-tuned for each data. The TV regularizer often fails to capture the local smoothness of the data adequately. Therefore, in many cases, additional regularizers need to be incorporated in the manner (2). This inevitably introduces the problem of selecting more balanced parameters. Besides, except for NN, many regularizers don't hold good theoretical properties.

To tackle the above two drawbacks, we introduce a new modified nuclear norm (MNN) that can capture both the local information and global low-rankness of data. Specifically, the MNN is defined by adopting a suitable transformation and then performing the NN on the transformed matrix, i.e.,

$$\|\mathbf{X}\|_{\text{MNN}} := \|\mathcal{D}(\mathbf{X})\|_*. \tag{3}$$

Compared to NN, MNN (3) can concurrently capture local information and global low-rankness of data. Specifically, the transformation $\mathcal{D}(\cdot)$ exploits useful local correlation information of the matrix and performing the NN on the transformed matrix could finely utilize such local correlation based on the compatibility of norms.

Specifically, the proposed MNN framework offers two key advantages over existing methods that integrate global and local information. First, MNN eliminates the need for parameter tuning required in additive manner (2). Second, under mild transformation assumptions, MNN provides exact recovery guarantees for two typical matrix applications: Robust PCA and Matrix Completion (MC). Building on theoretical tools related to nuclear norms [Candes and Recht, 2012; Candès *et al.*, 2011; Chen, 2015; Candes and Plan, 2010; Negahban and Wainwright, 2012; Shahid *et al.*, 2015], the MNN framework leverages these tools to offer theoretical guarantee. It incorporates various classical transformations, such as the difference operator [Peng *et al.*, 2022b; Wang *et al.*, 2023; Liu *et al.*, 2023], the Sobel operator [Kanopoulos *et al.*, 1988], and the Laplacian operators [Wang, 2007], enabling the integration of diverse local information. In summary, the contributions of this paper are:

Modeling: We propose the MNN framework, which can simultaneously exploit global and local information of data by performing the NN on a transformed matrix that encodes local correlations through a suitable transformation operator without trade-off parameters.

Theory: Under mild conditions, we prove the exact recoverability theory of MNN on two types of problems: Robust PCA and MC. Although the theoretical bounds are not improved compared to NN, from the perspective of embedding local smoothness, MNN provides a unified recoverability theory framework for integrating low-rank and multi-layer local information priors. Extensive experiments validate the efficacy of the MNN framework.

Operator: Classical operators are introduced into the MNN framework to characterize local smoothness. This paper demonstrates that for low-rank image data, the difference, Sobel, and Laplacian operators can be directly embedded into MNN to improve model's performance. In particular, Laplacian operators can provide richer local information than the widely used first-order differences operator.

2 Modified Nuclear Norm

In this part, we will give the forms and theoretical results about the MNN-induced Robust PCA and MC models.

2.1 Motivations

The previous global and local information fusion model [Wang *et al.*, 2017; He *et al.*, 2015; Xue *et al.*, 2021; He *et al.*, 2017; Peng *et al.*, 2022a; Peng *et al.*, 2020] was characterized by constructing multiple regularizers and summing them up, as follows:

$$\sum_{i=1}^{n} \tau_i \mathcal{R}_i(\mathbf{X}), \tag{4}$$

where $\mathcal{R}_i(\mathbf{X})(i = 1, \cdots, n)$ are regularizers, such as the nuclear norm (NN) and total variation (TV) regularizer in the form (2), and τ_i are trade-off parameters, manner (4) is intuitive and effective. However, it suffers from two key issues: the challenge of parameter selection among multiple regularizers and the lack of theoretical recoverability. To address these limitations, inspired by the norm compatibility and exact recovery guarantees of the nuclear norm, we propose the MNN defined in Eq. (3). As demonstrated in Remark 1, Eq. (3) effectively models both global and local information, eliminating the need for hyperparameter tuning.

Remark 1. *MNN (3) can simultaneously encode global and local information based on two aspects.*

***1) MNN could capture global low-rankness**. We assume that $\mathcal{D}(\cdot)$ can be reformulated as a full-rank matrix-induced linear operator, i.e., there exists a full-rank matrix $\mathbf{A}$ such that*

$$rank(\mathbf{AX}) = min(rank(\mathbf{A}), rank(\mathbf{X})) = rank(\mathbf{X}). \tag{5}$$

Based on the above derivation, it can be inferred that the low-rankness of $\mathcal{D}(\mathbf{X})$ stems from the low-rankness of $\mathbf{X}$, i.e., minimizing the rank of $\mathcal{D}(\mathbf{X})$ is equivalent to minimizing the rank of $\mathbf{X}$. Hence, MNN can characterize the global low-rankness of the data.

***2) MNN could capture local information**. According to the norm compatibility theorem, we have*

$$\|\mathcal{D}(\mathbf{X})\|_F \leq \|\mathcal{D}(\mathbf{X})\|_* \leq \|\mathcal{D}(\mathbf{X})\|_1, \tag{6}$$

holds for all transformation operators. Thus, it can be inferred that minimizing $\|\mathcal{D}(\mathbf{X})\|_*$ can simultaneously minimize both $\|\mathcal{D}(\mathbf{X})\|_F$ and $\|\mathcal{D}(\mathbf{X})\|_1$. If an appropriate transformation operator $\mathcal{D}(\cdot)$ is chosen so that $\mathcal{D}(\mathbf{X})$ can capture useful local information of data (e.g., edge structures), then MNN can effectively utilize such local information based on Eq. (6). For example, when $\mathcal{D}(\cdot)$ is chosen as the first-order difference operator $\nabla(\cdot)$, $\|\nabla(\mathbf{X})\|_F$ and $\|\nabla(\mathbf{X})\|_1$ correspond to isotropic [Rudin et al.*, 1992] and anisotropic [He* et al.*, 2017] TV regularizer, respectively. Therefore, our MNN can then characterize the local smoothness of data* [1].

[1] If $\mathcal{D}(\cdot)$ is set as identity mapping, $\|\mathcal{D}(\mathbf{X})\|_*$ degenerates to $\|\mathbf{X}\|_*$. Since $\|\mathbf{X}\|_F$ and $\|\mathbf{X}\|_1$ cannot encode local structural information of the data, $\|\mathbf{X}\|_*$ can only characterize the global low rankness.

Indeed, several works such as Correlated Total Variation (CTV [Wang *et al.*, 2023; Peng *et al.*, 2022b; Liu *et al.*, 2023]) can be seen as a special case of our MNN by setting the transformation operator as the first-order difference.

Remark 2. *Unlike the first-order difference in CTV, which only captures adjacent similarities, the proposed MNN framework accommodates operators with larger receptive fields, enabling more effective modeling of local smoothness across broader regions. What's more, MNN also removes the need for tuning the parameters between multiple regularization terms. While combining several regularizers improves interpretability, it often requires careful hyperparameter tuning—poor choices can yield suboptimal results.*

2.2 Theoretical Guarantees

Next, we establish the exact recovery guarantees of MNN-induced RPCA and MC models.

Models

Robust PCA [Candès *et al.*, 2011] is a problem aiming at accurately separating a low-rank matrix $\mathbf{X}_0 \in \mathbb{R}^{n_1 \times n_2}$ and a sparse matrix $\mathbf{S}_0$ from the observed data matrix $\mathbf{M} = \mathbf{X}_0 + \mathbf{S}_0$. By using MNN and ℓ_1 norms to respectively encode the global low-rankness and local priors of low-rank matrices, as well as the sparsity of sparse matrices, we can derive the following MNN-RPCA model:

$$\min_{\mathbf{X},\mathbf{S}} \|\mathcal{D}(\mathbf{X})\|_* + \lambda\|\mathbf{S}\|_1, s.t.\ \mathbf{M} = \mathbf{X} + \mathbf{S}. \tag{7}$$

MC [Candes and Recht, 2012] is a problem aiming to accurately infer the clean data $\mathbf{X}_0$ from a limited set of observed data $\mathbf{M}$ (the support set denoted as Ω). Using MNN, we can obtain the following MNN-MC model:

$$\min_{\mathbf{X}} \|\mathcal{D}(\mathbf{X})\|_*, s.t.\ \mathcal{P}_\Omega(\mathbf{M}) = \mathcal{P}_\Omega(\mathbf{X}), \tag{8}$$

where $\mathcal{P}_\Omega(\cdot)$ is a mapping operator. If $(i,j) \in \Omega$, then $\mathcal{P}_\Omega(\mathbf{X}_{ij}) = \mathbf{X}_{ij}$; otherwise, it is set to 0.

In our MNN-based model, the single regularization term simultaneously captures both global and local information, which avoids the need for carefully tuning the trade-off parameter τ_i required by additive model-based methods (4).

Assumptions

Before giving the theorems, we need three mild assumptions.

Assumption 1 (Incoherence Condition)**.** *For the low-rank matrix $\mathbf{X}_0 \in \mathbb{R}^{n_1 \times n_2}$ with rank r, it follows the incoherence condition with parameter μ, i.e.,*

$$\begin{aligned} \max_k \|\mathbf{U}^* e_k\| \le \frac{\mu r}{n_1}, \max_k \|\mathbf{V}^* e_k\| \le \frac{\mu r}{n_2}, \\ \|\mathbf{U}\mathbf{V}^*\|_\infty \le \sqrt{\frac{\mu r}{n_1 n_2}}, \end{aligned} \tag{9}$$

where $\mathbf{U} \in \mathbb{R}^{n_1 \times r}$ and $\mathbf{V} \in \mathbb{R}^{n_2 \times r}$ are obtained from the singular vector decomposition of $\mathbf{X}_0$ and e_k is the unit orthogonal vector.

Incoherence condition is a widely used assumption for low-rank recovery problems [Candès *et al.*, 2011; Chen, 2015; Candes and Romberg, 2007] to control the dispersion of the elements of the low-rank matrix. If the data is a rank-r identity matrix, since the observed values are likely to be zero, it is difficult to infer the non-zero elements, so we need this condition to keep the data away from the identity matrix.

The random distribution assumption for $\mathbf{S}_0$ and the normalization assumption of $\mathcal{D}(\cdot)$ are as follows:

Assumption 2 (Random Distribution)**.** *For the sparse term $\mathbf{S}_0$, its support Ω is chosen uniformly among all sets of cardinality m, and the signs of supports are random, i.e.*

$$\begin{aligned} \mathbb{P}[(\mathbf{S}_0)_{i,j} > 0 | (i,j) \in \Omega] = \mathbb{P}[(\mathbf{S}_0)_{i,j} \le 0 | (i,j) \in \Omega] \\ = 0.5. \end{aligned} \tag{10}$$

Assumption 3 (Normalization)**.** *The Frobenius norm of the linear transformation $\mathcal{D}(\cdot)$ in Eq. (6) is one.*

Assumption 3 ensures that the elements of the transformed data $\mathcal{D}(\mathbf{X})$ remain bounded, allowing key inequalities in the dual verification to hold. If the Frobenius norm of $\mathcal{D}(\cdot)$ is not one, normalization can be used to satisfy Assumption 3.

Main Results

Based on Assumptions 1-3, we can derive that:

Theorem 1 (MNN-RPCA Theorem)**.** *Suppose that $\mathcal{D}(\mathbf{X}_0) \in \mathbb{R}^{n_1 \times n_2}$, $\mathbf{S}_0$ and $\mathcal{D}(\cdot)$ obey Assumptions 1-3, respectively. Without loss of generality, suppose $n_1 \ge n_2$. Then, there is a numerical constant $c > 0$ such that with probability at least $1 - cn_1^{-10}$(over the choice of support of $\mathbf{S}_0$), the MNN-RPCA model (7) with $\lambda = 1/(\sqrt{n_1})$ is exact, i.e., the solution $(\hat{\mathbf{X}}, \hat{\mathbf{S}}) = (\mathbf{X}_0, \mathbf{S}_0)$, provided that*

$$\textit{rank}(\mathbf{X}_0) \le \rho_r n_2 \mu^{-1} (\log n_1)^{-2}, \textit{ and } m \le \rho_s n_1 n_2, \tag{11}$$

where ρ_r and ρ_s are some positive numerical constants, and m is the number of the support set of $\mathbf{S}_0$.

Theorem 2 (MNN-MC Theorem)**.** *Suppose that $\mathcal{D}(\mathbf{X}_0) \in \mathbb{R}^{n_1 \times n_2}$ and $\mathcal{D}(\cdot)$ obey Assumptions 1 and 3, $\Omega \sim Ber(p)$ and m is the number of Ω, where $Ber(p)$ represents the Bernoulli distribution with p. Without loss of generality, suppose $n_1 \ge n_2$. Then, there exist universal constants $c_0, c_1 > 0$ such that $\mathbf{X}_0$ is the unique solution to MNN-MC model (8) with probability at least $1 - c_1 n_1^{-3} \log n_1$, provided that*

$$m \ge c_0 \mu r n_1^{5/4} \log(n_1). \tag{12}$$

Theorems 1 and 2 establish the exact recoverable theory for the fusion-induced model combining low-rank and general local smoothness priors, which was unavailable in previous studies. Proofs are provided in the supplementary materials.

2.3 Optimizations

The exact recoverable theory, as outlined in Theorems 1 and 2, asserts that the optimal solution $(\hat{\mathbf{X}}, \hat{\mathbf{S}})$ of the model accurately reflects the true value $(\mathbf{X}_0, \mathbf{S}_0)$. Thus, the following corollary can be inferred with high probability.

Corollary. *Suppose $\mathbf{X}_0$ and $\mathbf{S}_0$ satisfy Assumptions 1 and 2, and transformation operator $\mathcal{D}(\cdot)$ satisfy Assumption 3. Denote the objective functions of the RPCA and MC models as*

$$\begin{aligned} \mathcal{J}_1^{\mathcal{D}}(\mathbf{X}) :=& \|\mathcal{D}(\mathbf{X})\|_* + \lambda\|\mathbf{M} - \mathbf{X}\|_1, \\ \mathcal{J}_2^{\mathcal{D}}(\mathbf{X}) :=& \|\mathcal{D}(\mathbf{X})\|_* + \mu\|\mathcal{P}_\Omega(\mathbf{M} - \mathbf{X})\|_F^2, \end{aligned} \tag{13}$$

respectively, where $\lambda = 1/\sqrt{n_1}$ *and* $\mu = (\sqrt{n_1} + \sqrt{n_2})\sqrt{p}\sigma$ *according to [Candes and Plan, 2010], and* n_1, n_2, σ, p *are the sizes of matrix, noise standard variance, and missing ratio. Then, for any* $\mathbf{X}$*, we have:*

$$\mathcal{J}_1^{\mathcal{D}}(\mathbf{X}) \geq \mathcal{J}_1^{\mathcal{D}}(\mathbf{X}_0), \mathcal{J}_2^{\mathcal{D}}(\mathbf{X}) \geq \mathcal{J}_2^{\mathcal{D}}(\mathbf{X}_0). \tag{14}$$

Then, according to Corollary 2.3, we can directly perform a simple gradient descent on the objective function (14) to obtain the final solution. More details can be seen in the supplementary materials.

2.4 Some Suitable Transformation Operators for MNN

In this section, we shift our attention to another important aspect, namely, the selection of the transformation operator $\mathcal{D}(\cdot)$. The local smoothness of images is typically achieved through convolution operators. The most common one is the first-order difference operator $\nabla(\cdot)$ =[-1,1]/[-1;1]. Based on the first-order difference operator [Rudin *et al.*, 1992] and its fast Fourier transform-based solving algorithm [Chambolle, 2004; Huang *et al.*, 2008], various types of total variation regularization have been developed in recent decades [Wang *et al.*, 2017; He *et al.*, 2015; Peng *et al.*, 2020]. However, the difference operators may not be able to characterize local information of data with multiple directions and higher-order derivative information. Hence, we consider several other transformations by taking advantage of multiple directions and higher-order derivatives.

Specifically, we examine four common convolution kernels: the first-order central difference operator, the Sobel operator, and two types of Laplacian operators. Their forms and edge extraction effects on the "Barbara" image are shown in Figure 2. It can be observed that compared to the central difference operator, the latter three operators capture more abundant local information about the image.

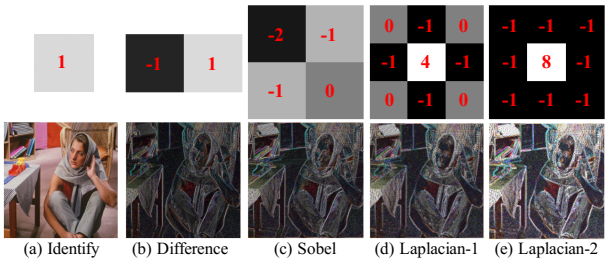


Figure 2: Demonstrates the forms of several typical convolution kernels and its edge extraction effect after acting on the image.

Next, we give an intuitive mathematical explanation of the superior local correlation excavation abilities of the Sobel and Laplacian operators by Remark 3.

Remark 3. *For an image, it can be modeled as a bivariate function* $z = f(x, y)$*, where* x, y *represent the coordinates, and* z *represents the grayscale value. The first-order difference of* $f(x, y)$ *is:*

$$\begin{aligned} \frac{\partial f}{\partial x} &= f(x+1, y) - f(x, y), \\ \frac{\partial f}{\partial y} &= f(x, y+1) - f(x, y). \end{aligned} \tag{15}$$

The second-order difference of $f(x, y)$ *can be represented in the* x *and* y *directions as follows:*

$$\begin{aligned} \frac{\partial^2 f}{\partial x^2} &= f(x+1, y) - 2f(x, y) + f(x-1, y), \\ \frac{\partial^2 f}{\partial y^2} &= f(x, y+1) - 2f(x, y) + f(x, y-1). \end{aligned} \tag{16}$$

The first-order difference operator $\frac{\partial f}{\partial x}/\frac{\partial f}{\partial y}$ *corresponds to the convolution kernel [-1,1]/[-1;1]. The Sobel operator* $\frac{\partial f}{\partial x} + \frac{\partial f}{\partial y}$ *and Laplacian operator* $\frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2}$ *have effects equivalent to [-2,-1;-1,0] and [0,-1,0;-1,4,-1;0,-1,0], respectively. The first derivative detects edges, while the second derivative detects the rate of change of edges. Derivatives in multiple directions can comprehensively characterize edge information. Therefore, the Sobel operator and the two types of Laplacian operators provided in Figure 2 can exploit richer local priors.*

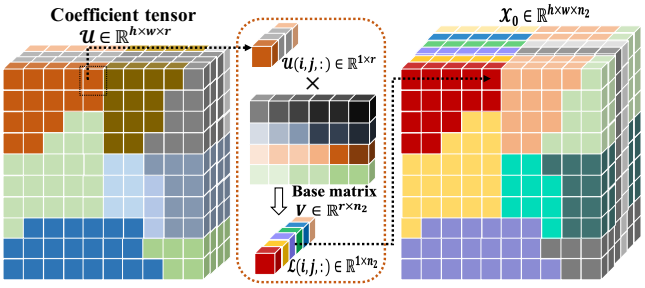


Figure 3: The simulated data generated mechanism of joint low rank and local smoothness data $\mathbf{X}_0$.

3 Simulated Experiment

The numerical experiments are conducted to validate Theorems 1 and 2. As described in Section 2.4, we use difference, Sobel, and Laplacian operators in MNN, denoted as MNN-Diff, MNN-Sobel, MNN-L1, and MNN-L2 in Figure 2. Note that MNN-Diff reduces to the CTV regularizer. Following Corollary 2.3, we set $\lambda = 1/\sqrt{\max\{n_1, n_2\}}$ for the TRPCA task and $\mu = (\sqrt{n_1} + \sqrt{n_2})\sqrt{p}\sigma$, $\sigma = 1e^{-4}$ for the MC noiseless task. Further performance gains can be achieved by fine-tuning λ and μ. All simulations are run on a PC with an Intel Core i5-10600KF CPU (4.10 GHz), 32 GB RAM, and a GeForce RTX 3080 GPU.

3.1 Data Generation

We create two factor matrices $\mathbf{U} \in \mathbb{R}^{n_1\times r}$ and $\mathbf{V} \in \mathbb{R}^{r\times n_2}$, yielding $\mathbf{X}_0 = \mathbf{U}\mathbf{V}^T$ as a low-rank matrix, where $n_1 = hw$. To introduce image-like properties, each column of $\mathbf{U}$ (i.e., $\mathbf{U}(:, i)$) is reshaped into an $h \times w$ matrix, thus we can get the tensor $\mathcal{U} \in \mathbb{R}^{h\times w\times r}$ satisfies $\mathbf{U} = \text{unfold}_3(\mathcal{U})$ where $\text{unfold}_3(\cdot)$ is the unfold operator to unfold the tensor to matrix along the third-mode. Each slice $\mathcal{U}(:, :, i)$ is randomly divided into c regions, with elements within each region being consistent and drawn from $\mathcal{N}(0, 1)$ distribution. The elements of $\mathbf{V}$ are selected from $\mathcal{N}(0, 1)$ distribution. The entire generation process is shown in Figure 3. Besides, the support set Ω is

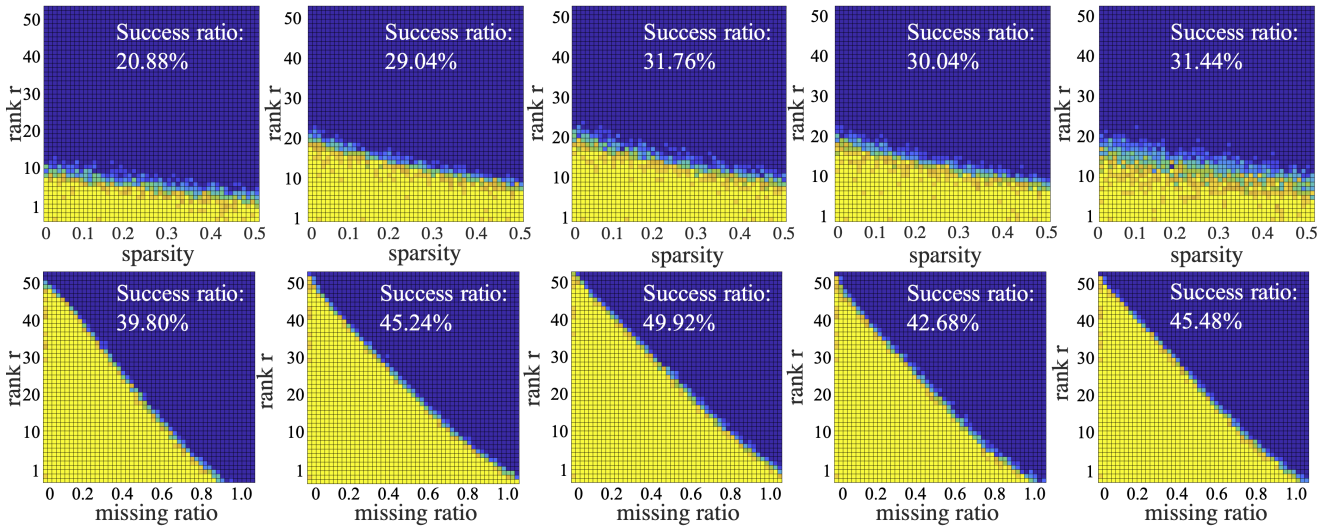
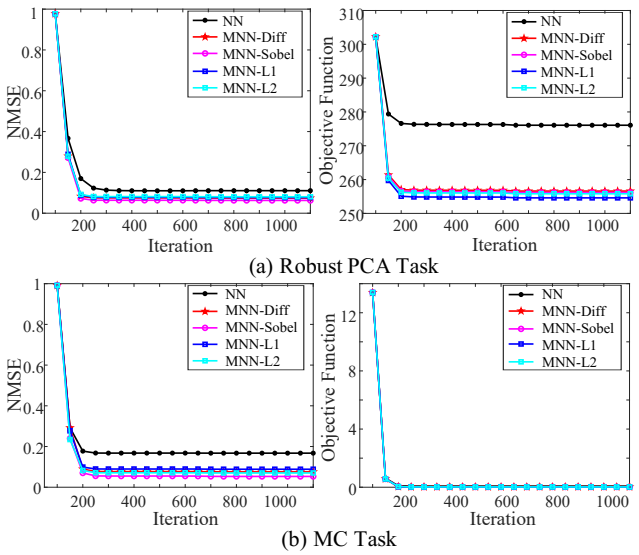


Figure 4: Phase transition diagrams of MNN variants for Robust PCA (top row) and MC (bottom row). Blue (0%) and yellow (100%) indicate success ratios. From left to right: NN, MNN-Diff, MNN-Sobel, MNN-L1, MNN-L2.



(a) Robust PCA Task

(b) MC Task

Figure 5: Objective function and relative recovery error over iterations: (a) RPCA with $r = 10, \rho_s = 0.1$; (b) MC with $r = 10$, sampling ratio 0.2.

chosen randomly. For RPCA task, $\mathbf{S}_0$ is set as $\mathbf{S}_0 = \mathcal{P}_\Omega(\mathbf{E})$ and the observed matrix $\mathbf{M}$ is set as $\mathbf{M} = \mathbf{X}_0 + \mathbf{S}_0$, where $\mathbf{E}$ is a matrix with independent Bernoulli ± 1 entries. For MC task, $\mathbf{M}$ is set as $\mathbf{M} = \mathcal{P}_\Omega(\mathbf{X}_0)$.

3.2 Experiment Settings and Result Analysis

Experiment Settings. In all experiments, we set $h = w = 50$, $n1 = hw = 2500$, and $n2 = 100$. We evaluate how the rank r, sparsity ρ_s (RPCA), and missing ratio ρ (MC) affect performance by varying ρ_s in $(0.01, 0.5)$ (step 0.01), ρ in $(0.01, 0.99)$ (step 0.02), and r in $(1, 50)$ (step 1).

Phase Transition Diagram. For each (r, ρ_s) and (r, ρ), we run 10 trials and consider recovery successful if $\|\hat{\mathbf{X}} - \mathbf{X}_0\|_F / \|\mathbf{X}_0\|_F \leq 0.05$. Figure 4 shows the phase diagrams with recovery rates. Compared to NN, MNN variants (Diff, Sobel, L1, L2) significantly expand the recovery region by leveraging local structures. Notably, MNN-Sobel, MNN-L1, and MNN-L2 outperform MNN-Diff, highlighting the advantage of high-order operators as noted in Remark 3.

Type	Methods
Non-convex Decomposition	KBR, LRTDTV,
Multi-regularization Fusion	LRTV, LRTDTV,
Matrix-based Methods	NN, LRTV, CTV
Tensor-based Methods	TCTV,SNN,TNN, Qrank, Framelet

Table 1: Comparison of methods for fusing global and local priors.

Empirical analysis of convergence. According to Corollary 2.3, the objectives of models (7) and (8) reach their minima at $\mathbf{X}_0$ and $\mathbf{S}_0$ under Assumptions 1–3. Even when the assumptions are only partially met, the objective values still decrease, and NMSE sequence $\|\mathbf{X}_k - \mathbf{X}_0\|_F / \|\mathbf{X}_0\|_F$ steadily declines. Figure 5 shows that under a learning rate of $1e^{-4}$, NN and all MNN variants converge stably.

4 Real Application

We then assess the restoration performance on real datasets.

Datasets. We selected four commonly used low-rank image datasets, which are HSI data used in [Wang *et al.*, 2023], MSI [2], color video sequences [3], and MRI and CT images [4]. Among them, hyperspectral images, multispectral images, color video sequences, and MRI and CT images contain 5, 11, 10, and 4 images, respectively. We adopt the same noise addition mechanism and sampling scheme as in the simulation experiments section.

[2]https://www.cs.columbia.edu/CAVE/databases/multispectral/
[3]http://trace.eas.asu.edu/yuv/
[4]https://www.cancerimagingarchive.net/

SR	Metrics	NN	SNN	KBR	TNN	LRTV	LRTD TV	CTV	TCTV	MNN-Diff	MNN-Sobel	MNN-L1	MNN-L2
Hyperspectral Image Denoising: Five Datasets													
0.1	PSNR↑	46.16	29.58	35.97	45.43	32.57	31.21	48.53	47.30	47.57	**49.52**	49.33	47.99
	SSIM↑	0.998	0.897	0.974	0.989	0.928	0.894	0.998	0.992	0.997	**0.999**	**0.999**	0.998
0.3	PSNR↑	39.93	23.21	33.31	40.34	30.26	29.30	45.69	43.55	44.99	**47.08**	46.70	45.52
	SSIM↑	0.992	0.608	0.959	0.980	0.883	0.845	0.997	0.988	0.997	**0.999**	**0.998**	0.997
	SSIM↑	0.690	0.138	0.519	0.051	0.644	0.642	0.962	0.366	0.781	**0.991**	0.910	**0.991**
Average Time/s		8.62	116.8	58.39	136.4	33.79	102.9	96.63	277.1	77.61	75.04	74.12	77.69
Multispectral Image Denoising: Eleven Datasets													
0.1	PSNR↑	29.85	33.39	35.76	41.70	33.56	36.32	40.64	**44.64**	40.15	42.14	42.92	40.74
	SSIM↑	0.968	0.967	0.965	0.993	0.957	0.977	0.995	**0.996**	0.995	**0.996**	**0.996**	0.995
0.3	PSNR↑	24.96	31.14	34.32	39.14	31.90	34.48	38.57	**42.95**	38.34	40.47	41.39	39.43
	SSIM↑	0.920	0.945	0.956	0.988	0.943	0.965	0.992	**0.995**	0.992	0.994	**0.995**	0.993
Average Time/s		17.01	294.5	106.8	275.6	52.91	220.9	233.5	513.9	68.62	70.24	67.24	67.35
RGB Video Denoising: Ten Datasets													
0.1	PSNR↑	32.26	27.44	26.48	38.01	29.97	23.51	35.73	38.69	37.14	38.61	**39.77**	37.62
	SSIM↑	0.958	0.774	0.827	0.980	0.892	0.809	0.976	0.983	0.980	0.983	**0.984**	0.981
0.3	PSNR↑	30.16	10.58	25.89	35.47	28.52	22.43	33.91	36.63	34.90	36.57	**37.45**	35.84
	SSIM↑	0.933	0.097	0.808	0.963	0.870	0.777	0.963	0.967	0.970	0.977	**0.979**	0.975
Average Time/s		28.17	104.1	72.61	142.9	52.67	119.9	129.2	293.1	157.6	159.8	157.8	159.9

Table 2: Restoration comparison of all competing methods under different salt and pepper noise noise variance (SR). The best and second results are highlighted in bold italics and underlined.

SR	Metrics	NN	SNN	KBR	TNN	Framlet	Qrank	CTV	TCTV	MNN-Diff	MNN-Sobel	MNN-L1	MNN-L2
Hyperspectral Image Completion: Five Datasets													
5%	PSNR↑	21.85	20.20	30.28	27.41	29.26	28.38	35.29	29.48	29.29	**36.03**	32.04	32.57
	SSIM↑	0.625	0.428	0.871	0.771	0.825	0.803	0.972	0.862	0.913	**0.981**	0.957	0.961
10%	PSNR↑	33.98	21.91	42.60	31.01	32.28	28.06	41.90	33.54	37.10	**43.09**	39.34	40.46
	SSIM↑	0.909	0.528	0.989	0.876	0.884	0.861	0.992	0.935	0.982	**0.995**	0.990	0.992
Average Time/s		11.30	35.23	258.6	143.3	222.3	112.1	213.2	474.5	41.30	42.98	41.97	41.85
Multispectral Image Completion: Eleven Datasets													
5%	PSNR↑	16.71	27.56	26.49	32.58	31.61	30.23	35.71	36.12	30.64	**36.98**	36.10	36.05
	SSIM↑	0.583	0.820	0.888	0.920	0.912	0.889	0.963	0.973	0.921	**0.978**	0.975	0.975
10%	PSNR↑	22.41	30.84	35.45	36.29	35.48	34.72	39.89	40.09	35.42	**41.11**	40.32	40.05
	SSIM↑	0.793	0.892	0.980	0.963	0.958	0.952	0.983	0.988	0.971	**0.992**	0.991	0.991
Average Time/s		15.4	77.73	431.5	275.4	512.2	332.1	303.1	800.6	31.70	32.36	31.99	31.49
RGB Video Completion: Ten Datasets													
5%	PSNR↑	23.01	17.72	26.90	28.14	26.53	25.35	29.43	28.70	25.76	**29.53**	27.95	28.10
	SSIM↑	0.605	0.484	0.802	0.801	0.790	0.751	0.890	0.856	0.826	**0.909**	0.892	0.894
10%	PSNR↑	29.06	20.48	32.81	30.54	31.55	29.36	**33.18**	31.89	29.32	32.75	31.61	31.72
	SSIM↑	0.846	0.589	0.917	0.861	0.893	0.879	0.937	0.910	0.898	**0.945**	0.935	0.936
Average Time/s		20.51	58.22	401.8	214.8	449.4	168.6	426.2	685.1	67.44	68.16	64.92	64.83
MRI-CT Completion: Four Datasets													
5%	PSNR↑	14.14	17.40	21.58	20.65	19.82	19.94	23.70	24.80	20.48	24.45	24.87	**24.94**
	SSIM↑	0.204	0.448	0.532	0.368	0.506	0.380	0.688	0.647	0.625	0.801	0.810	**0.813**
10%	PSNR↑	18.36	19.89	25.68	22.37	22.71	23.94	25.48	26.73	22.75	26.34	**27.04**	26.98
	SSIM↑	0.369	0.535	0.704	0.457	0.621	0.699	0.761	0.721	0.737	0.858	0.869	**0.870**
Average Time/s		20.07	69.32	484.0	316.8	537.8	235.2	424.7	1009	60.81	62.13	60.04	60.01

Table 3: Restoration comparison of all competing methods under different sampling ratio (SR). The best and second results are highlighted in bold italics and underlined.

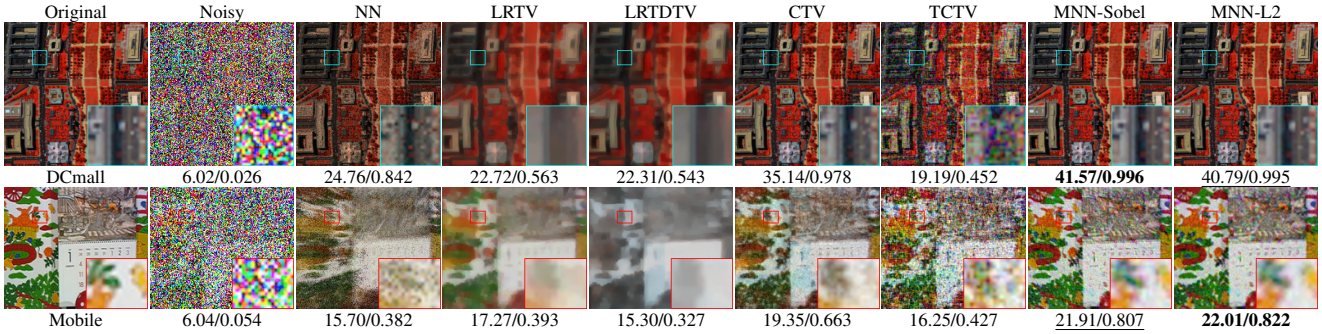


Figure 6: Recovered pseudo-colored images under 70% sparse noise. The PSNR and SSIM values are placed below the images, and the best and second-best results are bolded and underlined respectively.

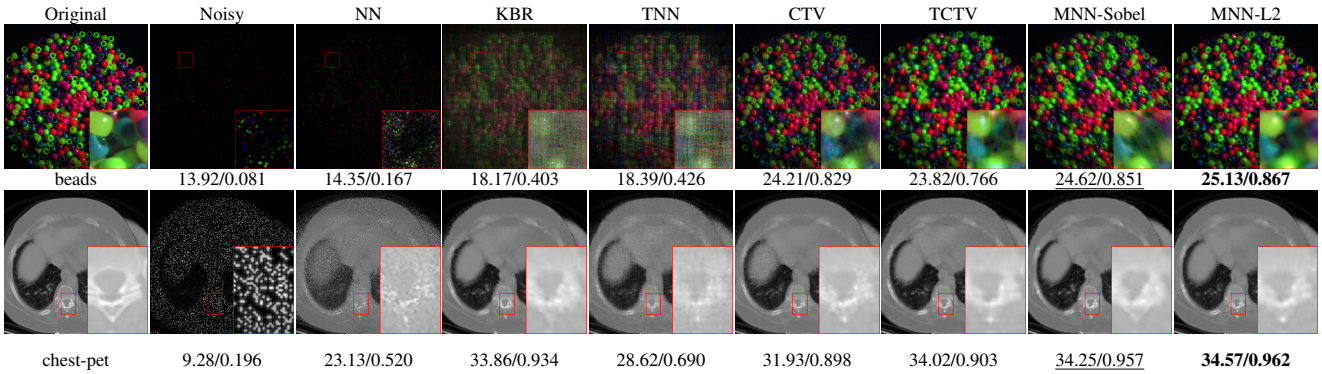


Figure 7: Restored pseudo-colored images under sampling rates of 0.02 (first row) for beans data (R-G-B:23-13-4) and 0.2 (second row) for chest-pet dataset at band 80.

Comparison methods. To validate the effectiveness of MNN in fusing global and local priors, we selected methods such as LRTV [He *et al.*, 2015], LRTDTV [Wang *et al.*, 2017], CTV [Peng *et al.*, 2022b], Qrank [], KBR [Xie *et al.*, 2017], TCTV [Wang *et al.*, 2023], SNN [Liu *et al.*, 2012], Qrank [Kong *et al.*, 2021], TNN [Lu *et al.*, 2019] and Framlet [Jiang *et al.*, 2020]. It is worth noting that CTV can actually be regarded as a special case of MNN since it utilizes multiple differential operators from different methods and employs the ADMM [Boyd *et al.*, 2011] algorithm to solve the model. To facilitate readers' understanding of these comparison methods, the description of these methods are placed in Table 1.

4.1 Robust Principal Component Analysis Tasks

We evaluate denoising performance on HSI, MSI, and RGB video data with salt-and-pepper noise levels ranging from 10% to 70%, using PSNR and SSIM as metrics. Table 2 summarizes the average results. The following observations can be observed: 1) In most cases, the MNN-Sobel model achieves the best performance; 2) Models such as MNN-Sobel, MNN-L1, and MNN-L2 generally outperform MNN-Diff, partially supporting Remark 3, which suggests that expanding the receptive field of convolutional operators enhances performance; 3) In some scenarios, MNN-L1 and MNN-L2 outperform MNN-Sobel, indicating that different transformation operators are required for different types of data and noise. If suitable transformation operators can be adaptively selected, further improvements in restoration performance are achievable.

Figure 6 shows restored pseudo-color images for two datasets, where MNN-Sobel and MNN-L2 visibly enhance noise reduction and color fidelity. More results are provided in the supplementary materials.

4.2 Completion Tasks

We conduct completion experiments on four datasets using random sampling. Table 3 reports the average restoration metrics, leading to the same three conclusions as in the denoising results (Table 2). Notably, despite being matrix-based, MNN consistently outperforms tensor-based methods like TNN, TCTV, and KBR, demonstrating its ability to effectively capture both global and local features.

Figure 7 shows visual results on two datasets, where MNN-based models excel in structure preservation, color fidelity, and overall restoration. Additional results are provided in the supplementary materials.

5 Conclusion

This paper proposes a modified nuclear norm (MNN) regularizer that integrates global low-rankness and local priors via transformation operators. Under mild and general theoretical conditions, MNN accommodates a broad range of transformations, enabling flexible and effective fusion of local and global information within a unified framework.

Acknowledgments

This work was in part by the National Key R& D Program of China (2020YFA0713900), the National Natural Science Foundation of China under Grant 12401674, 62272375 and 12201497, Tianyuan Fund for Mathematics of the National Natural Science Foundation of China (Grant No. 12426105), the Major Key Project of Pengcheng Laboratory under Grant PCL2024A06, the Guangdong Basic and Applied Basic Research Foundation under Grant 2025A1515011453, the Fundamental Research Funds for the Central Universities under Grant D5000240095, and the Key Research, Development Plan of Shaanxi Province, China (No. 2024GX-YBXM-145).

References

[Boyd *et al.*, 2011] Stephen Boyd, Neal Parikh, Eric Chu, Borja Peleato, Jonathan Eckstein, et al. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends® in Machine learning*, 3(1):1–122, 2011.

[Candes and Plan, 2010] Emmanuel J Candes and Yaniv Plan. Matrix completion with noise. *Proceedings of the IEEE*, 98(6):925–936, 2010.

[Candes and Recht, 2012] Emmanuel Candes and Benjamin Recht. Exact matrix completion via convex optimization. *Communications of the ACM*, 55(6):111–119, 2012.

[Candes and Romberg, 2007] Emmanuel Candes and Justin Romberg. Sparsity and incoherence in compressive sampling. *Inverse problems*, 23(3):969, 2007.

[Candès *et al.*, 2011] Emmanuel J Candès, Xiaodong Li, Yi Ma, and John Wright. Robust principal component analysis? *Journal of the ACM (JACM)*, 58(3):1–37, 2011.

[Chambolle, 2004] Antonin Chambolle. An algorithm for total variation minimization and applications. *Journal of Mathematical imaging and vision*, 20:89–97, 2004.

[Chen *et al.*, 2023] Yang Chen, Wenfei Cao, Li Pang, Jiangjun Peng, and Xiangyong Cao. Hyperspectral image denoising via texture-preserved total variation regularizer. *IEEE Transactions on Geoscience and Remote Sensing*, 2023.

[Chen, 2015] Yudong Chen. Incoherence-optimal matrix completion. *IEEE Transactions on Information Theory*, 61(5):2909–2923, 2015.

[De Silva and Lim, 2008] Vin De Silva and Lek-Heng Lim. Tensor rank and the ill-posedness of the best low-rank approximation problem. *SIAM Journal on Matrix Analysis and Applications*, 30(3):1084–1127, 2008.

[Gu *et al.*, 2014] Shuhang Gu, Lei Zhang, Wangmeng Zuo, and Xiangchu Feng. Weighted nuclear norm minimization with application to image denoising. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2862–2869, 2014.

[He *et al.*, 2015] Wei He, Hongyan Zhang, Liangpei Zhang, and Huanfeng Shen. Total-variation-regularized low-rank matrix factorization for hyperspectral image restoration. *IEEE transactions on geoscience and remote sensing*, 54(1):178–188, 2015.

[He *et al.*, 2017] Wei He, Hongyan Zhang, and Liangpei Zhang. Total variation regularized reweighted sparse nonnegative matrix factorization for hyperspectral unmixing. *IEEE Transactions on Geoscience and Remote Sensing*, 55(7):3909–3921, 2017.

[Huang *et al.*, 2008] Yumei Huang, Michael K Ng, and You-Wei Wen. A fast total variation minimization method for image restoration. *Multiscale Modeling & Simulation*, 7(2):774–795, 2008.

[Jiang *et al.*, 2020] Tai-Xiang Jiang, Michael K Ng, Xi-Le Zhao, and Ting-Zhu Huang. Framelet representation of tensor nuclear norm for third-order tensor completion. *IEEE Transactions on Image Processing*, 29:7233–7244, 2020.

[Kanopoulos *et al.*, 1988] Nick Kanopoulos, Nagesh Vasanthavada, and Robert L Baker. Design of an image edge detection filter using the sobel operator. *IEEE Journal of solid-state circuits*, 23(2):358–367, 1988.

[Kong *et al.*, 2021] Hao Kong, Canyi Lu, and Zhouchen Lin. Tensor q-rank: new data dependent definition of tensor rank. *Machine Learning*, 110(7):1867–1900, 2021.

[Liu *et al.*, 2012] Ji Liu, Przemyslaw Musialski, Peter Wonka, and Jieping Ye. Tensor completion for estimating missing values in visual data. *IEEE transactions on pattern analysis and machine intelligence*, 35(1):208–220, 2012.

[Liu *et al.*, 2023] Xinling Liu, Jingyao Hou, Jiangjun Peng, Hailin Wang, Deyu Meng, and Jianjun Wang. Tensor compressive sensing fused low-rankness and local-smoothness. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, pages 8879–8887, 2023.

[Lu *et al.*, 2019] Canyi Lu, Jiashi Feng, Yudong Chen, Wei Liu, Zhouchen Lin, and Shuicheng Yan. Tensor robust principal component analysis with a new tensor nuclear norm. *IEEE transactions on pattern analysis and machine intelligence*, 42(4):925–938, 2019.

[Negahban and Wainwright, 2012] Sahand Negahban and Martin J Wainwright. Restricted strong convexity and weighted matrix completion: Optimal bounds with noise. *The Journal of Machine Learning Research*, 13:1665–1697, 2012.

[Peng *et al.*, 2020] Jiangjun Peng, Qi Xie, Qian Zhao, Yao Wang, Leung Yee, and Deyu Meng. Enhanced 3dtv regularization and its applications on hsi denoising and compressed sensing. *IEEE Transactions on Image Processing*, 29:7889–7903, 2020.

[Peng *et al.*, 2022a] Jiangjun Peng, Hailin Wang, Xiangyong Cao, Xinling Liu, Xiangyu Rui, and Deyu Meng. Fast noise removal in hyperspectral images via representative coefficient total variation. *IEEE Transactions on Geoscience and Remote Sensing*, 60:1–17, 2022.

[Peng *et al.*, 2022b] Jiangjun Peng, Yao Wang, Hongying Zhang, Jianjun Wang, and Deyu Meng. Exact decomposition of joint low rankness and local smoothness plus sparse matrices. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(5):5766–5781, 2022.

[Recht *et al.*, 2010] Benjamin Recht, Maryam Fazel, and Pablo A Parrilo. Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization. *SIAM review*, 52(3):471–501, 2010.

[Rudin *et al.*, 1992] Leonid I Rudin, Stanley Osher, and Emad Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D: nonlinear phenomena*, 60(1-4):259–268, 1992.

[Shahid *et al.*, 2015] Nauman Shahid, Vassilis Kalofolias, Xavier Bresson, Michael Bronstein, and Pierre Vandergheynst. Robust principal component analysis on graphs. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 2812–2820, 2015.

[Shang *et al.*, 2023] Wenting Shang, Jiangjun Peng, Zebin Wu, Yang Xu, Mohamad Jouni, Mauro Dalla Mura, and Zhihui Wei. Hyperspectral anomaly detection via sparsity of core tensor under gradient domain. *IEEE Transactions on Geoscience and Remote Sensing*, 2023.

[Wang *et al.*, 2017] Yao Wang, Jiangjun Peng, Qian Zhao, Yee Leung, Xi-Le Zhao, and Deyu Meng. Hyperspectral image restoration via total variation regularized low-rank tensor decomposition. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 11(4):1227–1243, 2017.

[Wang *et al.*, 2023] Hailin Wang, Jiangjun Peng, Wenjin Qin, Jianjun Wang, and Deyu Meng. Guaranteed tensor recovery fused low-rankness and smoothness. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2023.

[Wang, 2007] Xin Wang. Laplacian operator-based edge detectors. *IEEE transactions on pattern analysis and machine intelligence*, 29(5):886–890, 2007.

[Xie *et al.*, 2017] Qi Xie, Qian Zhao, Deyu Meng, and Zongben Xu. Kronecker-basis-representation based tensor sparsity and its applications to tensor recovery. *IEEE transactions on pattern analysis and machine intelligence*, 40(8):1888–1902, 2017.

[Xue *et al.*, 2021] Jize Xue, Yong-Qiang Zhao, Yuanyang Bu, Wenzhi Liao, Jonathan Cheung-Wai Chan, and Wilfried Philips. Spatial-spectral structured sparse low-rank representation for hyperspectral image super-resolution. *IEEE Transactions on Image Processing*, 30:3084–3097, 2021.