Deep Rank-One Tensor Functional Factorization for Multi-Dimensional Data Recovery

Yanyi Li[1*], Xi Zhang[1*], Yisi Luo[1†], Deyu Meng[1,2,3]

[1]Xi'an Jiaotong University, Xi'an, China
[2]Pengcheng Laboratory, Shenzhen, China
[3]Macau University of Science and Technology, Macao, China
yyl0605@foxmail.com, xi.zhang@stu.xjtu.edu.cn, yisiluo1221@foxmail.com, dymeng@mail.xjtu.edu.cn

Abstract

Many real-world data are inherently multi-dimensional, e.g., color images, videos, and hyperspectral images. How to effectively and compactly represent these multi-dimensional data within a unified framework is an important pursuit. Previous methods focus on tensor factorizations, convolutional networks, or diffusion models for multi-dimensional data representation, which may not fully utilize inherent data structures and may lead to redundant parameters. In this work, we propose a Deep Rank-One Tensor Functional Factorization (DRO-TFF), which internally utilizes more comprehensive data priors facilitated by much fewer parameters. Concretely, our DRO-TFF consists of three organically integrated blocks: compact rank-one factorizations in the spatial domain, a deep transform to capture underlying low-dimensional structures, and smooth factors parameterized by implicit neural representations. Through a series of theoretical analysis, we show the rich data priors encoded in the DRO-TFF structure, e.g., Lipschitz smoothness and low-rankness. Extensive experiments on multi-dimensional data recovery problems, such as image and video inpainting, image denoising, and hyperspectral mixed noise removal, showcase the effectiveness of the proposed method.

1 Introduction

How to efficiently represent real-world multi-dimensional data, e.g., color images (Bengua et al. 2017; Liu et al. 2023), hyperspectral images (HSIs) (Hong et al. 2021), and videos (Hu et al. 2017), is receiving much attention in the AI community due to the wealthy information contained in multi-dimensional data structures and the promising applicability of these representation methods in low-level and high-level downstream tasks, e.g., inpainting (Liu et al. 2013), denoising (Xie et al. 2016), super-resolution (Kawulok et al. 2020), classification (Cheng et al. 2024), and detection (Guo et al. 2023). Within the fast developments of multi-dimensional data representation methods, current methodologies can be rationally categorized into three types: tensor decomposition-based methods (Vervliet, Debals, and De Lathauwer 2019), convolutional network-based methods (Wang et al. 2024b), and diffusion model-based methods (Özdenizci and Legenstein 2023), as introduced next.

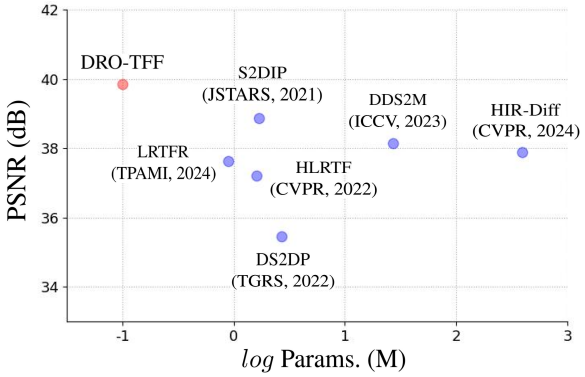

Figure 1: The average PSNR and number of parameters for hyperspectral image denoising across nine tests (see Table 2 for detailed results). The proposed DRO-TFF achieves better PSNR with fewer parameters.

For tensor decomposition-based methods, the most classical techniques are the CP decomposition and Tucker decomposition, which provide foundational approaches for representing multi-dimensional data (Kolda and Bader 2009). Another classical method is the tensor singular value decomposition (T-SVD) (Kilmer and Martin 2011), which emerges as a significant advancement that offers more flexible topologies for decomposing tensors. Some current methods have attempted to facilitate classical tensor decompositions with more advanced hierarchical and multi-scale structures. For example, the tensor network decompositions such as tensor train (TT) (Oseledets 2011), tensor ring (TR) (Yuan et al. 2019a), and fully connected tensor network (FCTN) (Zheng et al. 2021) decompositions are shown to be feasible in this field of multi-dimensional data representation. Deep neural networks have also been widely considered to improve tensor decompositions. For instance, Wang et al. (Wang et al. 2021, 2024a) utilize coupled transforms and convolutional neural networks to improve the performance of T-SVD. Luo et al. (Luo et al. 2024) introduce coordinate-based neural network in the Tucker decomposition to improve the flexibility of classical tensor decompositions. These tensor

*These authors contributed equally.
†Corresponding author.

decomposition-based methods provide foundations for representing complex multi-dimensional data structures.

In recent years, convolutional network-based methods have been extensively explored for multi-dimensional data representation. For example, the deep image prior (DIP) (Ulyanov, Vedaldi, and Lempitsky 2020) leverages the untrained convolutional network to capture image priors without the need for external training datasets. This innovative approach demonstrates significant improvements in image processing tasks by utilizing the implicit priors embedded within neural networks. Immediately after, the spatial-spectral constraint is introduced into the DIP (Luo et al. 2021) to enhance its capability for multi-dimensional data recovery. And the classical low-rank tensor decomposition is also considered to facilitate DIP (Miao et al. 2022) for multi-dimensional data. The elaborate local feature extraction abilities of the convolutional network have made it a popular backbone for many multi-dimensional data processing problems (Maffei et al. 2020; Yuan et al. 2019b).

More recently, diffusion model has emerged and become a powerful method for multi-dimensional data representation. For instance, the denoising diffusion restoration model (DDRM) (Kawar et al. 2022) is firstly proposed and attains good performance for color images, but it is not suitable for data-starved HSIs. Later, the unsupervised denoising diffusion spatial-spectral model (DDS2M) is proposed (Miao et al. 2023), which uses variational inference-based loss function and low-rank tensor decomposition for multi-dimensional HSI representation. Another recent effort (Pang et al. 2024) has shown the potential of pre-trained generative diffusion model for HSI restoration. Benefit from the powerful generative abilities of diffusion models, these methods attain superior performances than convolutional network-based methods for multi-dimensional data processing tasks.

Although the above methods have shown promising performances for multi-dimensional data representation, there is still room for improvements in terms of prior knowledge mining and efficiency. Specifically, current representation methods either fail to fully utilize comprehensive prior knowledge underlying multi-dimensional data, which leads to sub-optimal performances and redundant parameters, or have limited representation abilities due to the shallow and linear structures. In this work, we propose a Deep Rank-One Tensor Functional Factorization (DRO-TFF) for multi-dimensional data representation (see Fig. 2 for quick view). The proposed DRO-TFF utilizes more comprehensive prior knowledge of multi-dimensional data, obtaining a more compact tensor representation with powerful representation abilities. Specifically, DRO-TFF is a deep tensor decomposition-based model that has three organically integrated blocks, i.e., compact rank-one decompositions in the spatial domain, a deep transform for capturing low-dimensional structures along the spectral dimension, and smooth factors parameterized by implicit neural representations (INRs) (Sitzmann et al. 2020). The proposed DRO-TFF encodes low-rankness and implicit smoothness in the spatial and spectral dimensions, and has much fewer parameters as compared with convolutional and diffusion models. Therefore, DRO-TFF is expected to obtain good performances for multi-dimensional data processing tasks under favorable computational and space efficiency.

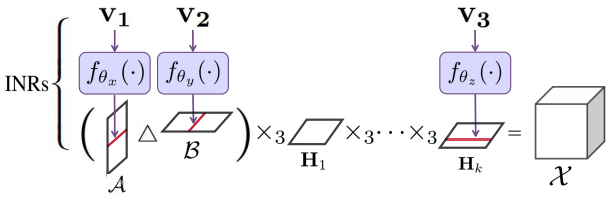

Figure 2: DRO-TFF contains three organically integrated blocks: compact rank-one factorizations, a mode-3 deep transform, and factor INRs.

The contributions of this paper are as follows:

- We propose a novel tensor representation termed DRO-TFF, which consists of three organically integrated blocks, i.e., compact rank-one factorizations in the spatial domain, a deep transform along the spectral domain, and spatial-spectral factors parameterized by implicit neural representations. The proposed DRO-TFF has a more compact structure with much fewer parameters and incorporates more comprehensive data priors.
- Theoretically, we reveal the potential representation abilities and low-rankness embedded in the rank-one formulation, and deduce the spatial-spectral smoothness brought by the functional factors parameterized by INRs. Hence, our method is expected to obtain good performances for data recovery with favorable efficiency.
- We apply our DRO-TFF to different multi-dimensional data recovery tasks including image and video inpainting, image denoising, and hyperspectral mixed noise removal. Extensive experiments demonstrate that DRO-TFF, which has a small number of parameters, has superior performances over state-of-the-art methods.

2 The Proposed DRO-TFF

2.1 Notations

In this paper, scalars, vectors, matrices, and tensors are defined by x, $\mathbf{x}$, $\mathbf{X}$, and $\mathcal{X}$. $\mathcal{X}^{(i)}$ denotes the i-th frontal slice of a tensor $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, which satisfies $\mathcal{X}^{(i)} \in \mathbb{R}^{n_1 \times n_2}$. The unfolding operator of a third-order tensor is denoted by $\texttt{unfold}(\cdot) : \mathbb{R}^{n_1 \times n_2 \times n_3} \to \mathbb{R}^{n_3 \times n_1 n_2}$, while $\texttt{fold}(\cdot) : \mathbb{R}^{n_3 \times n_1 n_2} \to \mathbb{R}^{n_1 \times n_2 \times n_3}$ denotes the folding operator. $\triangle$ denotes the face-wise product between two tensors, i.e., $\mathcal{C} = \mathcal{A} \triangle \mathcal{B} \Leftrightarrow \mathcal{C}^{(i)} = \mathcal{A}^{(i)}\mathcal{B}^{(i)}$. $\times_3$ denotes the mode-3 tensor-matrix product, i.e., $\mathcal{X} \times_3 \mathbf{H} = \texttt{fold}(\mathbf{H}\ \texttt{unfold}(\mathcal{X}))$. ∇_x, ∇_y and ∇_z denote the vertical, horizontal, and spectral derivative operators respectively, e.g., $\nabla_x(\mathcal{X}) = \mathcal{X}(i+1,:,:) - \mathcal{X}(i,:,:)$. Furthermore, we use $T(\cdot, i)$ to denote the unsqueeze operator in the i-th dimension, e.g., let $\mathbf{X} \in \mathbb{R}^{n_1 \times n_2}$ and we have $\mathcal{X} = T(\mathbf{X}, 2) \in \mathbb{R}^{n_1 \times 1 \times n_2}$.

2.2 DRO-TFF Representation

In this section, we give the detailed formulation of the deep rank-one tensor functional factorization. Specifically, our

DRO-TFF consists of three components: rank-one factorizations in the spatial domain, a deep transform along the third dimension, and factor INRs.

Spatial Rank-One Factorization To represent a spatial matrix, the low-rank matrix factorization has been widely considered, i.e., $\mathbf{X} = \mathbf{AB}$ where $\mathbf{A}, \mathbf{B}$ are low-rank factors. Real-world multi-dimensional data such as color images and videos are composed of multiple spatial matrices. Given a third-order tensor $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, it is natural to use n_3 low-rank matrix factorizations to separately represent each frontal slice. However, this leads to many learnable parameters when setting a suitable rank (often the rank is much larger than one). In this work, we propose to use rank-one factorizations with an additional transform matrix $\mathbf{H}$ to represent the third-order tensor. Specifically, instead of separately modeling each frontal slice, the rank-one factorizations have the following form:

$$\mathcal{X} = (\mathcal{A} \triangle \mathcal{B}) \times_3 \mathbf{H}, \tag{1}$$

where $\triangle$ denotes the face-wise product between two tensors. $\mathcal{A} \in \mathbb{R}^{n_1 \times 1 \times r}$ and $\mathcal{B} \in \mathbb{R}^{1 \times n_2 \times r}$ are rank-one factor tensors (i.e., their each frontal slice is a vector, and the resulting $\mathcal{A} \triangle \mathcal{B}$ is rank-one in each frontal slice), and $\mathbf{H} \in \mathbb{R}^{n_3 \times r}$ is a transform matrix. The advantages of the rank-one factorizations (1) are their lightweight structures, which are essential for processing large-scale multi-dimensional data. Also, due to the extra re-weighting matrix $\mathbf{H}$, the rank-one factorizations have enough expressive abilities for capturing spatial information of data, and such representation ability can be controlled by tuning the size r (see Theorem 2), which makes our model flexible and compact.

Mode-3 Deep Transform We propose to extend the mode-3 transform matrix $\mathbf{H}$ to a deep nonlinear transform to improve the expressiveness of the model. The proposed rank-one model under the deep transform has the following formulation:

$$\mathcal{X} = \psi(\cdots \psi((\mathcal{A} \triangle \mathcal{B}) \times_3 \mathbf{H}_1) \times_3 \cdots \times_3 \mathbf{H}_{k-1}) \times_3 \mathbf{H}_k, \tag{2}$$

where $\mathcal{A}, \mathcal{B}$ are rank-one factors and $\{\mathbf{H}_i\}_{i=1}^k$ are weight matrices of the deep transform with $\mathbf{H}_i \in \mathbb{R}^{r \times r}$ for $i = 1, 2, \cdots, k-1$ and $\mathbf{H}_k \in \mathbb{R}^{n_3 \times r}$. Here, $\psi(\cdot)$ is a nonlinear activation and is set as LeakyReLU in this work.

Factor Implicit Neural Representations To further introduce the smoothness into the model, we suggest factor implicit neural representations to encode the spatial-spectral smoothness that enhances the robustness of the model. Specifically, we use INRs (Sitzmann et al. 2020) to parameterize factor tensors/matrices $\mathcal{A}, \mathcal{B}, \mathbf{H}_k$ in the rank-one factorizations. Let $f_\theta(\cdot) : \mathbb{R}^n \to \mathbb{R}^{n \times r}$ be an INR with parameters θ (i.e., a coordinate-based neural network) that takes the coordinate vector $\mathbf{v} \in \mathbb{R}^n$ as the input, its formulation is a classical multilayer perceptron (MLP):

$$f_\theta(\mathbf{v}) := \mathbf{W}_d(\sigma(\mathbf{W}_{d-1} \cdots \sigma(\mathbf{W}_1 \mathbf{v}))), \tag{3}$$

where $\sigma(\cdot)$ is a nonlinear activation function. Specifically, for each element of the input coordinate vector $\mathbf{v} \in \mathbb{R}^n$, the INR $f_\theta(\cdot)$ maps it into an r-th dimensional vector, which results in a final output of size $n \times r$.

Remark 1 *Here, we emphasize two important techniques when using INRs to parameterize the rank-one model.*

- *First, the rank-one factorizations naturally allow us to use INRs to parameterize the factors. This is because the factors $\mathcal{A}, \mathcal{B}$ are actually matrices instead of tensors in the rank-one setting, which can be easily combined with the INR in* (3) *whose output is exactly a matrix.*
- *When using INR to parameterize the deep transform matrices $\{\mathbf{H}_i\}_{i=1}^k$, we only need to parameterize the last matrix $\mathbf{H}_k$ and let the remaining matrices $\{\mathbf{H}_i\}_{i=1}^{k-1}$ to be learnable themselves, because this is enough to encode the spectral smoothness (see Theorem 2) with little increase of parameter number.*

Based on the above analysis, we use three INRs to respectively parameterize factors $\mathcal{A}, \mathcal{B}$, and $\mathbf{H}_k$:

$$\begin{cases} \mathcal{A} = T(f_{\theta_x}(\mathbf{v}_1), 2), \ \mathbf{v}_1 \in \mathbb{R}^{n_1} \\ \mathcal{B} = T(f_{\theta_y}(\mathbf{v}_2), 1), \ \mathbf{v}_2 \in \mathbb{R}^{n_2} \\ \mathbf{H}_k = f_{\theta_z}(\mathbf{v}_3), \ \mathbf{v}_3 \in \mathbb{R}^{n_3}, \end{cases} \tag{4}$$

where $f_{\theta_x}(\mathbf{v}_1), f_{\theta_y}(\mathbf{v}_2), f_{\theta_z}(\mathbf{v}_3)$ are three INRs with input coordinate vectors $\mathbf{v}_1 = [1, \cdots, n_1]$, $\mathbf{v}_2 = [1, \cdots, n_2]$, and $\mathbf{v}_3 = [1, \cdots, n_3]$. $T(\cdot, i)$ denotes the unsqueeze operator from a matrix to a tensor. With the factor INRs, our model is no longer a discrete tensor decomposition but becomes a functional factorization, which implicitly encodes smoothness in the infinite dimensional domain (see Theorem 2).

By plugging the factor INR formulation (4) into the deep rank-one factorizations (2), we have the following explicit formulation of the proposed deep rank-one tensor functional factorization.

Definition 1 *Given a tensor $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, we call the following factorization as its DRO-TFF representation:*

$$\begin{aligned} \mathcal{X} = \psi(\cdots \psi((T(f_{\theta_x}(\mathbf{v}_1), 2) \triangle T(f_{\theta_y}(\mathbf{v}_2), 1)) \times_3 \mathbf{H}_1) \\ \times_3 \cdots \times_3 \mathbf{H}_{k-1}) \times_3 f_{\theta_z}(\mathbf{v}_3), \end{aligned} \tag{5}$$

where the learnable parameters are $\theta_x, \theta_y, \theta_z$ and $\{\mathbf{H}_i\}_{i=1}^{k-1}$.

2.3 Theoretical Analysis

Here, we give theoretical analysis to show that the proposed DRO-TFF encodes more comprehensive data priors within its structure, including its rank-r representation abilities, low-rankness, spatial smoothness, and spectral smoothness.

First, DRO-TFF has the representation abilities to capture rank-r information ($r > 1$) even with rank-one factorizations. This is revealed through the following lemma.

Lemma 1 *Let $\mathbf{X} \in \mathbb{R}^{n_1 \times n_2}$ and $rank(\mathbf{X}) = r$, then there exist r rank-one matrices $\{\mathbf{C}_i\}_{i=1}^r$ such that $\mathbf{X} = \sum_{i=1}^r \mathbf{C}_i$.*

Let us consider the basic formulation of our rank-one factorizations in (1), i.e., $\mathcal{X} = (\mathcal{A} \triangle \mathcal{B}) \times_3 \mathbf{H}$, where $\mathcal{A} \in \mathbb{R}^{n_1 \times 1 \times r}$ and $\mathcal{B} \in \mathbb{R}^{1 \times n_2 \times r}$ are rank-one factor tensors. Here, the mode-3 product $\times_3$ can be seen as the summation operator of r rank-one matrices in $\mathcal{A} \triangle \mathcal{B}$ (i.e., each frontal slice $\mathcal{A}^{(i)}\mathcal{B}^{(i)}$ is rank-one), and the only difference between Lemma 1 is that such summation has the weights $\mathbf{H}$. However, adding weights in the summation would not necessarily change the rank. Hence from Lemma 1 we see that the

proposed model has the potential to capture rank-r spatial information even with lightweight rank-one factorizations.

From another perspective, our basic model also preserves the low-rankness of the representation. This is because any matrix with rank larger than r can not be represented by the addition of r rank-one matrices.

Lemma 2 *Let $\mathbf{X} \in \mathbb{R}^{n_1 \times n_2}$ and $rank(\mathbf{X}) > r$. Then for r arbitrary rank-one matrices $\mathbf{C}_i \in \mathbb{R}^{n_1 \times n_2}$ $(i = 1, \cdots, r)$ we have $\mathbf{X} \neq \sum_{i=1}^{r} \mathbf{C}_i$.*

Proof Assume that there exist rank-one matrices $\mathbf{C}_i$ $(i = 1, \cdots, r)$ such that $\mathbf{X} = \sum_{i=1}^{r} \mathbf{C}_i$. Then $rank(\mathbf{X}) = rank(\sum_{i=1}^{r} \mathbf{C}_i) \leq \sum_{i=1}^{r} rank(\mathbf{C}_i) = r$, contradicting with the assumption $rank(\mathbf{X}) > r$. Therefore, the original statement is true. □

Theorem 1 *(Low-rankness) Let $\mathcal{X} = (\mathcal{A} \triangle \mathcal{B}) \times_3 \mathbf{H}$, where $\mathcal{A} \in \mathbb{R}^{n_1 \times 1 \times r}$, $\mathcal{B} \in \mathbb{R}^{1 \times n_2 \times r}$ are rank-one factors and $\mathbf{H} \in \mathbb{R}^{n_3 \times r}$. Then we have $rank(\mathcal{X}^{(i)}) \leq r$ for $i = 1, \cdots, n_3$.*

Proof Let $\mathcal{C} = \mathcal{A} \triangle \mathcal{B}$, then $\mathcal{C}^{(i)} = \mathcal{A}^{(i)} \mathcal{B}^{(i)}$. Since $\mathcal{A}^{(i)}$ and $\mathcal{B}^{(i)}$ are rank-one we have $rank(\mathcal{C}^{(i)}) = 1$. And the following equality holds: $\mathcal{X}^{(i)} = \sum_{j=1}^{r} \mathbf{H}(i,j) \times \mathcal{C}^{(j)}$, where $\mathbf{H}(i,j)$ is a scalar that doesn't affect the rank of $\mathcal{C}^{(j)}$. According to Lemma 2, we can get $rank(\mathcal{X}^{(i)}) \leq r$. □

Theorem 1 shows that the proposed rank-one factorizations preserve the low-rankness implicitly, i.e., the representation has spatial rank less than r. In practice, we introduce the mode-3 deep transform to enhance the representation abilities, which makes the tensor $\mathcal{X}$ not necessarily be rank-r. However, the rank-one factorizations still serve as low-rank regularization to constrain the solution space so that a good recovery result can be obtained.

Except for low-rankness, our DRO-TFF also internally preserves the smoothness of the representation along the spatial and spectral dimensions attributed to the functional representation brought by factor INRs. With both the low-rank and smooth constraints, our model is essentially suitable for data recovery tasks.

Theorem 2 *(Smoothness) Let $\mathcal{X}$ satisfy* (5)*. Suppose that the MLPs $f_{\theta_x}(\cdot), f_{\theta_y}(\cdot), f_{\theta_z}(\cdot)$ have the same activation function $\sigma(\cdot)$ and depth d. Besides, we assume that*

- *$\sigma(\cdot)$ is Lipschitz continuous with Lipschitz constant κ.*
- *The ℓ_1-norm of each weight matrix $\mathbf{W}_i$ (i.e., $\|\mathbf{W}_i\|_{\ell_1} := \sum_{j,k} |\mathbf{W}_i(j,k)|$) in the three MLPs and the ℓ_1-norm of each transform matrix $\mathbf{H}_i$ is bounded by $\eta > 0$.*

Then the resulting $\mathcal{X}$ in (5) *satisfies the following inequalities for all $(i,j,k)s$ $(i = 1, 2, \ldots, n_1,\ j = 1, 2, \ldots, n_2,\ k = 1, 2, \ldots, n_3)$:*

$$\begin{cases} |\mathcal{X}(i,j,k) - \mathcal{X}(i-1,j,k)| \leq \delta |\mathbf{v}_1^{(i)} - \mathbf{v}_1^{(i-1)}| \\ |\mathcal{X}(i,j,k) - \mathcal{X}(i,j-1,k)| \leq \delta |\mathbf{v}_2^{(j)} - \mathbf{v}_2^{(j-1)}| \\ |\mathcal{X}(i,j,k) - \mathcal{X}(i,j,k-1)| \leq \delta |\mathbf{v}_3^{(k)} - \mathbf{v}_3^{(k-1)}|, \end{cases} \tag{6}$$

where $\mathbf{v}_1^{(i)}$ denotes the i-th element of $\mathbf{v}_1$, $\delta = \eta^{3d+k-1} \kappa^{3d-3} \zeta^2$, and $\zeta = \max\{\|\mathbf{v}_1\|_\infty, \|\mathbf{v}_2\|_\infty, \|\mathbf{v}_3\|_\infty\}$.

From the theorem it can be seen that the proposed DRO-TFF implicitly preserves the local smoothness of the representation $\mathcal{X}$ attributed to the Lipschitz smoothness of the factor INRs. Hence, our model is expected to obtain robust performances for multi-dimensional data recovery tasks.

2.4 Applications of DRO-TFF

Image and Video Inpainting Image and video inpainting aims to recover the clean image from the one with missing pixels. By giving the observed incompleted data $\mathcal{O}$, the loss function of this task is defined as follows:

$$\min_{\Theta} \|\mathcal{P}_\Omega(\mathcal{O} - \mathcal{X})\|_F^2 + \gamma_1 \|\mathcal{X}\|_{\mathrm{TV}}, \tag{7}$$

where $\mathcal{X}$ admits (5). Here, $\Theta := \{\theta_x, \theta_y, \theta_z, \{\mathbf{H}_i\}_{i=1}^{k-1}\}$ are optimization parameters and $\mathcal{P}_\Omega(\cdot)$ is the projection operator that keeps the elements in the mask Ω unchanged while changing others to be zero. The total variation (TV) regularization $\|\mathcal{X}\|_{\mathrm{TV}} := \|\nabla_x \mathcal{X}\|_{l_1} + \|\nabla_y \mathcal{X}\|_{l_1}$ is used to enhance the stability, and γ_1 is a trade-off parameter.

Image Denoising Gaussian noise is a common degradation in multi-dimensional image. Here we consider the image denoising task for removing Gaussian noise, whose objective is as follows:

$$\min_{\Theta} \|\mathcal{O} - \mathcal{X}\|_F^2 + \gamma_1 \|\mathcal{X}\|_{\mathrm{TV}} + \gamma_2 \|\mathcal{X}\|_{\mathrm{SSTV}}, \tag{8}$$

where $\mathcal{O}$ is the noisy image, $\mathcal{X}$ admits (5), and $\Theta := \{\theta_x, \theta_y, \theta_z, \{\mathbf{H}_i\}_{i=1}^{k-1}\}$ are optimization parameters. Here, $\|\mathcal{X}\|_{\mathrm{SSTV}} := \|\nabla_x \nabla_z \mathcal{X}\|_{l_1} + \|\nabla_y \nabla_z \mathcal{X}\|_{l_1}$ is the spatial-spectral TV regularization (Aggarwal and Majumdar 2016) that enhances the spatial-spectral local smoothness of the recovered image. γ_2 is a trade-off parameter.

Hyperspectral Mixed Noise Removal The purpose of hyperspectral mixed noise removal is to obtain the clean HSI from its observation corrupted by mixed noise. Here, the noisy HSI contains mixed noise including Gaussian noise, sparse noise, and deadlines (missing columns). Due to the presence of sparse noise, we additionally consider a sparse term and change the formulation as

$$\min_{\Theta, \mathcal{S}} \|\mathcal{P}_\Omega(\mathcal{O} - \mathcal{X} - \mathcal{S})\|_F^2 + \gamma \|\mathcal{S}\|_{l_1} + \gamma_1 \|\mathcal{X}\|_{\mathrm{TV}} + \gamma_2 \|\mathcal{X}\|_{\mathrm{SSTV}}, \tag{9}$$

where $\mathcal{O}$ is the observed image corrupted by mixed noise, $\mathcal{X}$ admits (5), $\mathcal{S}$ is the underlying sparse noise to be estimated, and $\Theta := \{\theta_x, \theta_y, \theta_z, \{\mathbf{H}_i\}_{i=1}^{k-1}\}$ are optimization parameters of DRO-TFF. Here, we jointly optimize the parameters of DRO-TFF and the sparse noise $\mathcal{S}$ to remove mixed noise from the HSI. We can use the following alternating direction method of multipliers (ADMM) to address the optimization model (9). We remark that models (7) and (8) can be addressed in analogous ways.

Let $\mathcal{V}_1 = \nabla_x(\mathcal{X})$, $\mathcal{V}_2 = \nabla_y(\mathcal{X})$, $\mathcal{V}_3 = \nabla_x \nabla_z(\mathcal{X})$, $\mathcal{V}_4 = \nabla_y \nabla_z(\mathcal{X})$. We can write the augmented Lagrangian function of (9) by attaching Lagrangian multipliers Λ_i $(i = 1, 2, 3, 4)$ (see supplementary file for details).

Then the joint problem can be decomposed into several sub-problems. Specifically, the $\mathcal{V}$ sub-problem can be

solved by $\mathcal{V}_1^{t+1} = Soft_{\frac{\gamma_1}{\mu}}(\nabla_x(\mathcal{X}^t) + \frac{\Lambda_1^t}{\mu})$, $\mathcal{V}_2^{t+1} = Soft_{\frac{\gamma_1}{\mu}}(\nabla_y(\mathcal{X}^t) + \frac{\Lambda_2^t}{\mu})$, $\mathcal{V}_3^{t+1} = Soft_{\frac{\gamma_2}{\mu}}(\nabla_x\nabla_z(\mathcal{X}^t) + \frac{\Lambda_3^t}{\mu})$, and $\mathcal{V}_4^{t+1} = Soft_{\frac{\gamma_2}{\mu}}(\nabla_y\nabla_z(\mathcal{X}^t) + \frac{\Lambda_4^t}{\mu})$, where $Soft_\alpha(\cdot) = sgn(\cdot)\max\{|\cdot| - \alpha, 0\}$ is the soft thresholding operator. The $\mathcal{S}$ sub-problem can be solved by $\mathcal{S}^{t+1} = Soft_{2\gamma}(\mathcal{O} - \mathcal{X}^t)$. The $\mathcal{X}$ sub-problem is to optimize the parameters of the DRO-TFF, and can be formulated as

$$\begin{aligned}\min_{\Theta} \|\mathcal{P}_\Omega(\mathcal{O} - \mathcal{X} - \mathcal{S}^t)\|_F^2 + \frac{\mu}{2}(\|\nabla_x(\mathcal{X}) - \mathcal{D}_1^t\|_F^2 \\ + \|\nabla_y(\mathcal{X}) - \mathcal{D}_2^t\|_F^2 + \|\nabla_x\nabla_z(\mathcal{X}) - \mathcal{D}_3^t\|_F^2 \\ + \|\nabla_y\nabla_z(\mathcal{X}) - \mathcal{D}_4^t\|_F^2),\end{aligned} \tag{10}$$

where $\mathcal{D}_i^t = \mathcal{V}_i^t - \frac{\Lambda_i^t}{\mu}$ and $\mathcal{X}$ admits (5). We use the off-the-shell Adam optimizer to update Θ with one step in each iteration of the ADMM algorithm. The multipliers Λ can be updated by $\Lambda_1^{t+1} = \Lambda_1^t + \mu(\nabla_x(\mathcal{X}^t) - \mathcal{V}_1^t)$, $\Lambda_2^{t+1} = \Lambda_2^t + \mu(\nabla_y(\mathcal{X}^t) - \mathcal{V}_2^t)$, $\Lambda_3^{t+1} = \Lambda_3^t + \mu(\nabla_x\nabla_z(\mathcal{X}^t) - \mathcal{V}_3^t)$, and $\Lambda_4^{t+1} = \Lambda_4^t + \mu(\nabla_y\nabla_z(\mathcal{X}^t) - \mathcal{V}_4^t)$.

3 Experiments

3.1 Comparisons with State-of-the-Arts

In experiments, we perform a comparative analysis of all the introduced tasks. We use two commonly used evaluation metrics peak-signal-to-noise-ratio (PSNR) and structural similarity (SSIM).

Datasets and Compared Methods We use publicly available datasets. For image and video inpainting, we use three types of multi-dimensional data including color images (*House*, *Tree*, *Jelly Beans*)[1], natural HSIs (*Feathers*, *Flowers*, *Thread Spools*)[2], and videos (*News*, *Carphone*, *Hall Monitor*)[3] as testing data. We consider the sampling rates (SRs) $0.1, 0.2, 0.3$ to perform random missing. The compared methods include state-of-the-art tensor decomposition-based methods FTNN (Jiang et al. 2020), FCTN (Zheng et al. 2021), HLRTF (Luo et al. 2022), LRTFR (Luo et al. 2024), and diffusion model-based method DDS2M (Miao et al. 2023).

For image denoising, we adopt three HSIs (*Face*, *Fake and Real Peppers (Peppers)*, *Egyptian Statue*)[2]. We consider Gaussian noise with standard deviation (SD) $0.1, 0.2, 0.3$ to generate noisy images. The compared methods are convolutional network-based methods S2DIP (Luo et al. 2021), DS2DP (Miao et al. 2022), tensor decomposition-based methods HLRTF (Luo et al. 2022), LRTFR (Luo et al. 2024), and diffusion model-based methods DDS2M (Miao et al. 2023) and HIR-Diff (Pang et al. 2024).

For hyperspectral mixed noise removal, we adopt three natural HSIs (*Cloth*, *Chart and Stuffed Toy (Toy)*, *Fake and Real Food (Food)*)[2] and two remote sensing HSIs (*WDC Mall* and *Pavia University (Pavia)*)[4] as testing data. We consider three noisy cases. **Case 1** contains Gaussian noise with standard deviation 0.2 and deadlines in all spectral bands. **Case 2** contains Gaussian noise with standard deviation 0.2 and sparse noise with SR 0.1. **Case 3** contains the same noise as **Case 2** plus deadlines in all spectral bands. The compared methods are the same as image denoising.

[1]https://sipi.usc.edu/database/database.php

[2]https://www.cs.columbia.edu/CAVE/databases/multispectral/

[3]http://trace.eas.asu.edu/yuv/

[4]https://lesun.weebly.com/hyperspectral-data-set.html

SR	0.1		0.2		0.3		Params.	Time
Method	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM		
Color images *House*, *Tree*, *Jelly Beans* ($256 \times 256 \times 3$)								
Observed	4.796	0.014	5.310	0.028	5.884	0.044	-	-
FTNN	18.109	0.386	21.608	0.587	24.386	0.714	-	273s
FCTN	18.085	0.366	20.942	0.514	24.401	0.682	-	29s
HLRTF	21.664	0.586	25.231	0.747	27.700	0.833	1.0M	13s
DDS2M	23.386	0.693	26.217	0.744	27.149	0.743	27.3M	249s
LRTFR	23.519	0.655	25.948	0.757	26.942	0.798	0.5M	24s
DRO-TFF	**24.631**	**0.765**	**27.871**	**0.847**	**29.568**	**0.893**	0.2M	22s
HSIs *Feathers*, *Flowers*, *Thread Spools* ($512 \times 512 \times 31$)								
Observed	13.597	0.318	14.107	0.388	14.689	0.456	-	-
FTNN	35.973	0.974	40.425	0.992	43.416	0.996	-	275s
FCTN	36.560	0.951	38.376	0.966	39.490	0.973	-	752s
HLRTF	38.565	0.977	42.344	0.990	44.859	0.995	3.2M	47s
DDS2M	32.970	0.918	35.474	0.951	37.227	0.965	27.3M	718s
LRTFR	39.045	0.980	42.912	0.989	44.933	0.991	1.7M	73s
DRO-TFF	**41.338**	**0.992**	**43.986**	**0.995**	**45.942**	**0.997**	0.6M	126s
Videos *News*, *Carphone*, *Hall Monitor* ($144 \times 176 \times 100$)								
Observed	6.825	0.024	7.338	0.064	7.916	0.112	-	-
FTNN	22.336	0.844	25.699	0.930	28.144	0.961	-	268s
FCTN	29.328	0.967	30.614	0.976	31.361	0.980	-	138s
HLRTF	28.647	0.962	31.419	0.980	33.575	0.988	0.9M	19s
DDS2M	24.065	0.883	26.747	0.938	27.955	0.951	27.4M	542s
LRTFR	28.079	0.954	31.604	0.981	33.378	0.987	0.7M	29s
DRO-TFF	**29.970**	**0.973**	**32.590**	**0.985**	**34.153**	**0.989**	0.1M	23s

Table 1: The average quantitative results by different methods for image and video inpainting. The **best** and second-best values are highlighted.

Hyperparameter Settings For all tasks, the rank parameter r is set as $\text{int}(\beta \min(n_1, n_2))$, where n_i $(i = 1, 2)$ are the spatial sizes of the observed data, and β is a hyperparameter that is set to 0.7, 0.3, 0.3 for image and video inpainting, image denoising, and hyperspectral mixed noise removal. The depth of the mode-3 deep transform k is set to 3. The network depth of the factor INRs is set to 3. The sine activation $\sin(\omega_0 \cdot)$ is used as nonlinear activation because of its Lipschitz continuous property. For color image inpainting, ω_0 is set to 1, otherwise it is set to 2. The trade-off parameters γ_1, γ_2 in the loss function are set to 4×10^{-5} and 4×10^{-4} respectively. In the hyperspectral mixed noise removal model, we set $\gamma = 0.01$ and $\mu = 0.04$. For other hyperparameters, please refer to the source code.

Experimental Results The quantitative results of image and video inpainting, image denoising, and hyperspectral mixed noise removal are respectively shown in Tables 1-3. It is evident that the DRO-TFF consistently achieves superior performances across all tasks. This can be attributed to

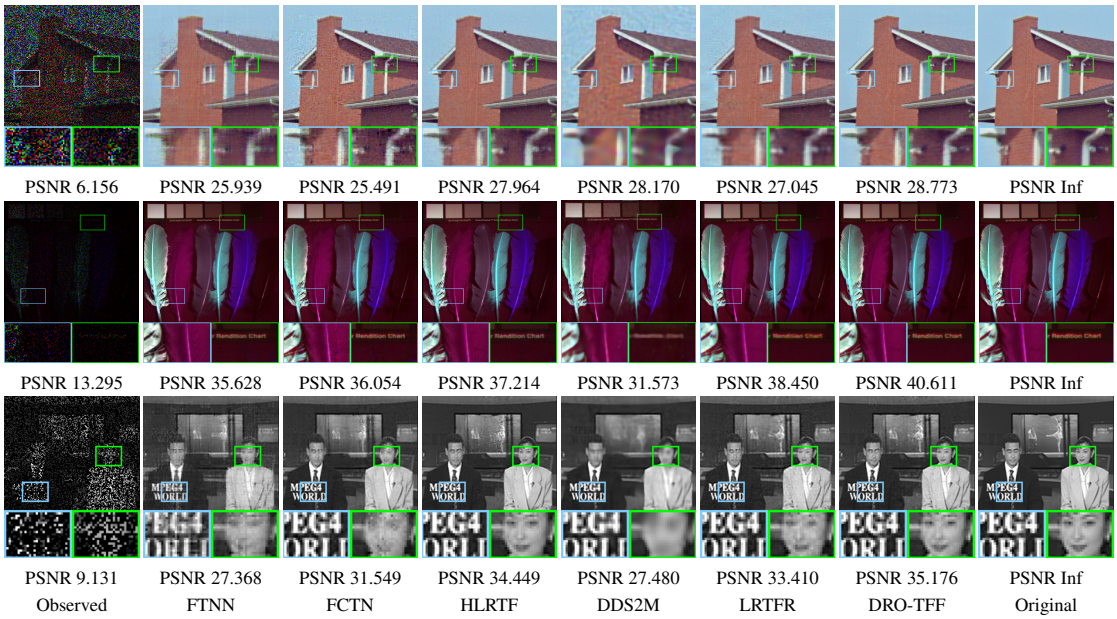

Figure 3: The results of image and video inpainting by different methods on color image *House* (SR=0.3), HSI *Feathers* (SR=0.1), and video *News* (SR=0.3).

SD	0.1		0.2		0.3		Params.	Time
Method	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM		
HSIs *Face*, *Peppers*, *Egyptian Statue* ($512 \times 512 \times 31$)								
Observed	20.000	0.165	13.980	0.058	10.457	0.030	-	-
S2DIP	41.616	0.977	38.516	0.964	36.487	0.940	1.7M	1016s
DS2DP	37.507	0.941	34.988	0.901	33.856	0.892	2.7M	2105s
HLRTF	39.821	0.963	36.609	0.916	35.170	0.894	1.6M	61s
DDS2M	40.323	0.959	37.922	0.934	36.179	0.901	27.2M	644s
LRTFR	40.059	0.972	37.040	0.950	35.805	0.925	0.9M	92s
HIR-Diff	39.188	0.950	37.766	0.943	36.706	0.932	391M	146s
DRO-TFF	**42.184**	**0.980**	**39.530**	**0.969**	**37.808**	**0.956**	0.1M	74s

Table 2: The average quantitative results by different methods for image denoising.

the comprehensive data prior knowledge utilization of DRO-TFF, e.g., spatial low-rankness and Lipschitz smoothness. Furthermore, in terms of parameter count, DRO-TFF has the lowest number of parameters among all methods. For convolutional network-based methods and diffusion model-based methods such as DDS2M and HIR-Diff, the number of parameters is much more than that of the proposed DRO-TFF. This indicates that DRO-TFF is highly efficient for representing multi-dimensional data, validating its favorable computational and space efficiency.

Some visual results are shown in Figs. 3-4. It can be observed that the DRO-TFF recovers image details better than other methods. For example, in Fig. 3 *house*, the pipes are more clearly restored using DRO-TFF, while the other methods suffer from different degrees of blurring (see zoom-in boxes). In addition, the results of DRO-TFF is generally free of noise and artifacts, which showcases the robustness of DRO-TFF for multi-dimensional data recovery.

3.2 Ablation Study

To test the significance of the three building blocks in DRO-TFF, we test different variants as follows.

- DRO-TFF without rank-one factorizations (dubbed D-TFF): To test the effectiveness of rank-one factorizations, we consider the rank-p factorizations, where p is set as 5.
- DRO-TFF without deep transform (dubbed RO-TFF): To verify the importance of the mode-3 deep transform, we remove the first $k-1$ linear layers $\mathbf{H}_1, \cdots, \mathbf{H}_{k-1}$ and the nonlinear LeakyReLU layers, and then apply the linear transform-based model for image recovery tasks.
- DRO-TFF without functional representation (dubbed DRO-TF): To validate the influence of factor INRs, we remove these INRs and directly optimize the learnable parameters $\mathcal{A}, \mathcal{B}, \{\mathbf{H}_i\}_{i=1}^k$ in the model.

We adopt natural HSIs *Oil Painting*, *Pompoms*, and *CD* in the CAVE dataset[2] to test different variants. We consider three tasks including inpainting (SR=0.1), denoising (SD=0.2), and mixed noise removal (Case 3). The results are shown in Table 4. It can be seen that DRO-TFF obtains better performances than other variants, which showcases the necessity of different building blocks. Especially, the rank-one factorizations preserve the underlying low-dimensional structures of multi-dimensional data with lightweight parameter number. The deep transform enhances the expressiveness of the model, and factor INRs preserve the spatial and spectral smoothness of the data. Fig. 5 shows some visual results. It can be observed that the deep transform is

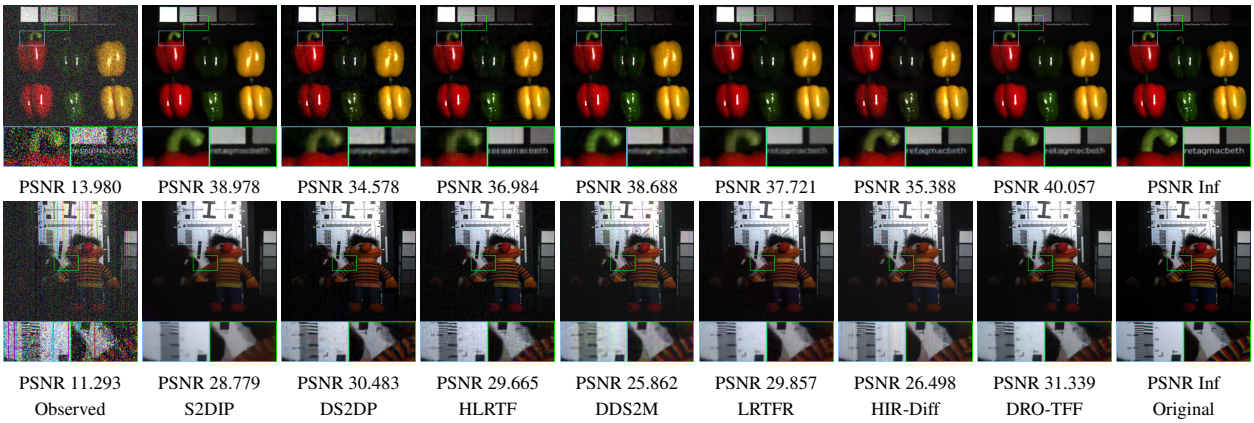

Figure 4: Top row: The results of image denoising by different methods on *Peppers* (SD=0.2). Bottom row: The results of hyperspectral mixed noise removal by different methods on *Toy* (Case 3).

Case	Case 1		Case 2		Case 3		Params.	Time
Method	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM		
HSIs *Cloth*, *Toy*, *Food* ($512 \times 512 \times 31$)								
Observed	13.853	0.152	11.346	0.109	11.373	0.100	-	-
S2DIP	31.364	0.892	29.784	0.785	28.301	0.732	1.7M	1001s
DS2DP	30.264	0.851	28.968	0.758	29.932	**0.827**	2.7M	2213s
HLRTF	31.679	0.880	28.823	0.746	28.817	0.744	1.6M	57s
DDS2M	30.494	0.894	24.777	0.643	25.439	0.637	27.2M	835s
LRTFR	30.631	0.854	28.373	0.718	28.289	0.717	0.9M	113s
HIR-Diff	31.149	0.907	24.994	0.631	25.997	0.638	391M	134s
DRO-TFF	**33.006**	**0.932**	**30.522**	**0.795**	**30.076**	0.789	0.1M	124s
HSIs *WDC Mall* ($256 \times 256 \times 191$), *Pavia* ($256 \times 256 \times 93$)								
Observed	15.359	0.200	12.581	0.135	12.603	0.128	-	-
S2DIP	28.940	0.809	28.967	0.873	26.905	0.763	1.8M	2290s
DS2DP	30.345	0.880	29.304	0.861	30.193	0.871	2.7M	1577s
HLRTF	31.347	0.884	29.185	0.861	28.997	0.853	0.4M	43s
DDS2M	30.320	0.888	24.788	0.835	25.847	0.819	27.5M	940s
LRTFR	28.685	0.812	27.269	0.794	27.259	0.791	0.4M	82s
HIR-Diff	31.436	**0.917**	24.911	0.845	26.374	0.853	391M	19s
DRO-TFF	**31.659**	0.912	**29.981**	**0.874**	**30.321**	**0.881**	0.1M	120s

Table 3: The average quantitative results by different methods for hyperspectral mixed noise removal.

Task	Inpainting		Denoising		Mixed noise removal	
Method	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
Observed	13.815	0.151	13.975	0.175	11.496	0.037
D-TFF	36.837	0.962	34.577	0.936	30.946	0.804
RO-TFF	30.279	0.813	31.524	0.858	31.159	0.806
DRO-TF	34.623	0.948	33.263	0.914	29.516	0.762
DRO-TFF	**37.601**	**0.971**	**35.389**	**0.951**	**32.160**	**0.816**

Table 4: The average quantitative results by different variants of the proposed method.

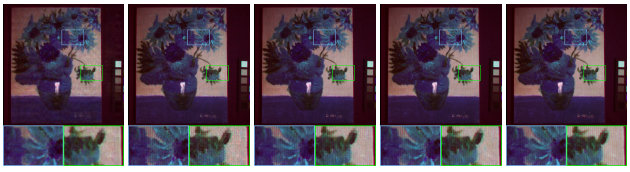

Figure 5: The results of image inpainting on *Oil Painting* (SR=0.1) by different variants of the proposed method.

crucial for capturing fine details of the image (RO-TFF v.s. DRO-TFF). The factor INRs are helpful to avoid artifacts (DRO-TF v.s. DRO-TFF), and the rank-one factorizations are more effective than the rank-5 factorizations (D-TFF v.s. DRO-TFF). Note that the rank-one factorizations also bring an overall lightweight structure. These results validate the importance of the three basic components, and they are organically integrated to obtain good performances.

4 Conclusions

In this paper, we propose a deep rank-one tensor functional factorization for multi-dimensional data representation. Our method contains three organically integrated components: the lightweight rank-one factorizations that preserve the low-rankness of the recovered data, the mode-3 deep transform that enhances the representation abilities of the model, and the factor INRs that capture the spatial-spectral smooth characteristics of data. Experiments on image/video inpainting, denoising, and hyperspectral mixed noise removal verify the superiority and efficiency of the proposed method as compared with state-of-the-art methods.

Acknowledgments

This research was supported by the National Key R&D Program of China (2022YFA1004100), the Major Key Project of PCL under Grant PCL2024A06, and the NSFC projects under contracts 124B2029, 12226004, and 62272375.

References

Aggarwal, H. K.; and Majumdar, A. 2016. Hyperspectral Image Denoising Using Spatio-Spectral Total Variation. *IEEE Geoscience and Remote Sensing Letters*, 13(3): 442–446.

Bengua, J. A.; Phien, H. N.; Tuan, H. D.; and Do, M. N. 2017. Efficient Tensor Completion for Color Image and Video Recovery: Low-Rank Tensor Train. *IEEE Transactions on Image Processing*, 26(5): 2466–2479.

Cheng, C.; Zhang, L.; Li, H.; Cui, W.; Gao, J.; and Cun, Y. 2024. A Deep High-Order Tensor Sparse Representation for Hyperspectral Image Classification. *IEEE Transactions on Geoscience and Remote Sensing*, 62: 1–16.

Guo, T.; He, L.; Luo, F.; Gong, X.; Li, Y.; and Zhang, L. 2023. Anomaly Detection of Hyperspectral Image With Hierarchical Antinoise Mutual-Incoherence- Induced Low-Rank Representation. *IEEE Transactions on Geoscience and Remote Sensing*, 61: 1–13.

Hong, D.; He, W.; Yokoya, N.; Yao, J.; Gao, L.; Zhang, L.; Chanussot, J.; and Zhu, X. 2021. Interpretable Hyperspectral Artificial Intelligence: When nonconvex modeling meets hyperspectral remote sensing. *IEEE Geoscience and Remote Sensing Magazine*, 9(2): 52–87.

Hu, W.; Tao, D.; Zhang, W.; Xie, Y.; and Yang, Y. 2017. The Twist Tensor Nuclear Norm for Video Completion. *IEEE Transactions on Neural Networks and Learning Systems*, 28(12): 2961–2973.

Jiang, T.-X.; Ng, M. K.; Zhao, X.-L.; and Huang, T.-Z. 2020. Framelet Representation of Tensor Nuclear Norm for Third-Order Tensor Completion. *IEEE Transactions on Image Processing*, 29: 7233–7244.

Kawar, B.; Elad, M.; Ermon, S.; and Song, J. 2022. Denoising Diffusion Restoration Models. In *Advances in Neural Information Processing Systems*, volume 35, 23593–23606.

Kawulok, M.; Benecki, P.; Piechaczek, S.; Hrynczenko, K.; Kostrzewa, D.; and Nalepa, J. 2020. Deep Learning for Multiple-Image Super-Resolution. *IEEE Geoscience and Remote Sensing Letters*, 17(6): 1062–1066.

Kilmer, M. E.; and Martin, C. D. 2011. Factorization strategies for third-order tensors. *Linear Algebra and its Applications*, 435(3): 641–658.

Kolda, T. G.; and Bader, B. W. 2009. Tensor Decompositions and Applications. *SIAM Review*, 51: 455–500.

Liu, J.; Liu, R. W.; Sun, J.; and Zeng, T. 2023. Rank-One Prior: Real-Time Scene Recovery. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(7): 8845–8860.

Liu, J.; Musialski, P.; Wonka, P.; and Ye, J. 2013. Tensor Completion for Estimating Missing Values in Visual Data. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(1): 208–220.

Luo, Y.; Zhao, X.; Li, Z.; Ng, M. K.; and Meng, D. 2024. Low-Rank Tensor Function Representation for Multi-Dimensional Data Recovery. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 46(5): 3351–3369.

Luo, Y.; Zhao, X.; Meng, D.; and Jiang, T. 2022. HLRTF: Hierarchical Low-Rank Tensor Factorization for Inverse Problems in Multi-Dimensional Imaging. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 19281–19290.

Luo, Y.-S.; Zhao, X.-L.; Jiang, T.-X.; Zheng, Y.-B.; and Chang, Y. 2021. Hyperspectral Mixed Noise Removal via Spatial-Spectral Constrained Unsupervised Deep Image Prior. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 14: 9435–9449.

Maffei, A.; Haut, J. M.; Paoletti, M. E.; Plaza, J.; Bruzzone, L.; and Plaza, A. 2020. A Single Model CNN for Hyperspectral Image Denoising. *IEEE Transactions on Geoscience and Remote Sensing*, 58(4): 2516–2529.

Miao, Y.; Zhang, L.; Zhang, L.; and Tao, D. 2023. DDS2M: Self-Supervised Denoising Diffusion Spatio-Spectral Model for Hyperspectral Image Restoration. In *IEEE/CVF International Conference on Computer Vision*, 12086–12096.

Miao, Y.-C.; Zhao, X.-L.; Fu, X.; Wang, J.-L.; and Zheng, Y.-B. 2022. Hyperspectral Denoising Using Unsupervised Disentangled Spatiospectral Deep Priors. *IEEE Transactions on Geoscience and Remote Sensing*, 60: 1–16.

Oseledets, I. V. 2011. Tensor-Train Decomposition. *SIAM Journal on Scientific Computing*, 33(5): 2295–2317.

Pang, L.; Rui, X.; Cui, L.; Wang, H.; Meng, D.; and Cao, X. 2024. HIR-Diff: Unsupervised Hyperspectral Image Restoration Via Improved Diffusion Models. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 3005–3014.

Sitzmann, V.; Martel, J.; Bergman, A.; Lindell, D.; and Wetzstein, G. 2020. Implicit Neural Representations with Periodic Activation Functions. In *Advances in Neural Information Processing Systems*, volume 33, 7462–7473.

Ulyanov, D.; Vedaldi, A.; and Lempitsky, V. 2020. Deep Image Prior. *International Journal of Computer Vision*, 128: 1867–1888.

Vervliet, N.; Debals, O.; and De Lathauwer, L. 2019. Exploiting Efficient Representations in Large-Scale Tensor Decompositions. *SIAM Journal on Scientific Computing*, 41(2): A789–A815.

Wang, J.-L.; Huang, T.-Z.; Zhao, X.-L.; Jiang, T.-X.; and Ng, M. K. 2021. Multi-Dimensional Visual Data Completion via Low-Rank Tensor Representation Under Coupled Transform. *IEEE Transactions on Image Processing*, 30: 3581–3596.

Wang, J.-L.; Huang, T.-Z.; Zhao, X.-L.; Luo, Y.-S.; and Jiang, T.-X. 2024a. CoNoT: Coupled Nonlinear Transform-Based Low-Rank Tensor Representation for Multidimensional Image Completion. *IEEE Transactions on Neural Networks and Learning Systems*, 35(7): 8969–8983.

Wang, Z.; Li, D.; Zhang, M.; Luo, H.; and Gong, M. 2024b. Enhancing Hyperspectral Images via Diffusion Model and Group-Autoencoder Super-resolution Network. *AAAI Conference on Artificial Intelligence*, 38(6): 5794–5804.

Xie, Q.; Zhao, Q.; Meng, D.; Xu, Z.; Gu, S.; Zuo, W.; and Zhang, L. 2016. Multispectral Images Denoising by Intrinsic Tensor Sparsity Regularization. In *IEEE Conference on Computer Vision and Pattern Recognition*, 1692–1700.

Yuan, L.; Li, C.; Mandic, D.; Cao, J.; and Zhao, Q. 2019a. Tensor Ring Decomposition with Rank Minimization on Latent Space: An Efficient Approach for Tensor Completion. *AAAI Conference on Artificial Intelligence*, 33(01): 9151–9158.

Yuan, Q.; Zhang, Q.; Li, J.; Shen, H.; and Zhang, L. 2019b. Hyperspectral Image Denoising Employing a Spatial–Spectral Deep Residual Convolutional Neural Network. *IEEE Transactions on Geoscience and Remote Sensing*, 57(2): 1205–1218.

Zheng, Y.-B.; Huang, T.; Zhao, X.; Zhao, Q.; and Jiang, T.-X. 2021. Fully-Connected Tensor Network Decomposition and Its Application to Higher-Order Tensor Completion. In *AAAI Conference on Artificial Intelligence*.

Özdenizci, O.; and Legenstein, R. 2023. Restoring Vision in Adverse Weather Conditions With Patch-Based Denoising Diffusion Models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(8): 10346–10357.