

# OTLRM: Orthogonal Learning-based Low-Rank Metric for Multi-Dimensional Inverse Problems

Xiangming Wang<sup>1</sup>, Haijin Zeng<sup>2\*</sup>, Jiaoyang Chen<sup>1</sup>,  
Sheng Liu<sup>3</sup>, Yongyong Chen<sup>1\*</sup>, Guoqing Chao<sup>4</sup>

<sup>1</sup> Harbin Institute of Technology (Shenzhen)

<sup>2</sup> Gent University

<sup>3</sup> University of Electronic Science and Technology of China

<sup>4</sup> Harbin Institute of Technology (Weihai)

xmwang28@gmail.com

## Abstract

In real-world scenarios, complex data such as multispectral images and multi-frame videos inherently exhibit robust low-rank property. This property is vital for multi-dimensional inverse problems, such as tensor completion, spectral imaging reconstruction, and multispectral image denoising. Existing tensor singular value decomposition (t-SVD) definitions rely on hand-designed or pre-given transforms, which lack flexibility for defining tensor nuclear norm (TNN). The TNN-regularized optimization problem is solved by the singular value thresholding (SVT) operator, which leverages the t-SVD framework to obtain the low-rank tensor. However, it is quite complicated to introduce SVT into deep neural networks due to the numerical instability problem in solving the derivatives of the eigenvectors. In this paper, we introduce a novel data-driven generative low-rank t-SVD model based on the learnable orthogonal transform, which can be naturally solved under its representation. Prompted by the linear algebra theorem of the Householder transformation, our learnable orthogonal transform is achieved by constructing an endogenously orthogonal matrix adaptable to neural networks, optimizing it as arbitrary orthogonal matrices. Additionally, we propose a low-rank solver as a generalization of SVT, which utilizes an efficient representation of generative networks to obtain low-rank structures. Extensive experiments highlight its significant restoration enhancements.

**Code** — <https://github.com/xianggkl/OTLRM>

**Extended version** — <https://arxiv.org/abs/2412.11165>

## Introduction

Real-world multi-dimensional data, such as multispectral images (MSIs), videos, and Magnetic Resonance Imaging (MRI) data, are usually affected by unpredictable factors during capture and transmission. To reconstruct the original tensors, many tasks are extended under different observations, such as Tensor Completion (TC) (Qin et al. 2022; Mai, Lam, and Lee 2022), spectral imaging (Cai et al. 2022), and spectral denoising (Wang et al. 2022a).

Current approaches have achieved outstanding results benefiting from the low-rank nature of the original data

tensor. However, the tensor rank is still not well defined. Various definitions of the tensor rank are not deterministic and all have specific advantages and limitations. For example, the CANDECOMP PARAFAC (CP) rank (Carroll and Chang 1970) is defined as the number of rank-1 tensors obtained by the CP decomposition. Since computing the CP decomposition of a tensor is NP-hard and finding its accurate convex approximation is challenging, CP rank may not be appropriate to give solutions tailored to practical application areas. Additionally, Tucker rank (Tucker 1966; Sun, Vong, and Wang 2022) is based on the tensor unfolding scheme, which unfolds the tensor into matrices along different dimensions. As a result, it leads to a broken structure of the original tensor, and such an unfolding describes the correlation between only one mode of the tensor and all other modes (one mode versus the rest), which may bring undesirable results to the tensor reconstruction. Overall, how to define an appropriate tensor rank in different tensor decompositions is a problem worth discussing and analyzing.

With the tensor-tensor product (t-product) (Kernfeld, Kilmer, and Aeron 2015) gradually becoming a comprehensive approach for tensor multiplication, the tensor Singular Value Decomposition (t-SVD) (Zhang et al. 2014; Lu et al. 2019; Liu et al. 2024) which constructs the original tensor as the t-product of two orthogonal tensors and an  $f$ -diagonal tensor, has received widespread attention and research. Explicitly, the t-product is defined in a transform domain based on an arbitrary invertible linear transform. Given a tensor  $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$  and a transform matrix  $\mathbf{L} \in \mathbb{R}^{n_3 \times n_3}$ , the transformed tensor  $L(\mathcal{X})$  can be formulated as

$$L(\mathcal{X}) = \mathcal{X} \times_3 \mathbf{L}, \quad (1)$$

where  $\times_3$  denotes the mode-3 tensor product. Following the definition, t-SVD can be defined in the transform domain, which captures the low-rank structure of the tensor. *Based on some fixed linear invertible transforms such as Discrete Fourier Transform (DFT) (Zhang and Aeron 2016) and Discrete Cosine Transform (DCT) (Lu 2021), current transforms have theoretical guarantees, yet are fixed and not data-adaptive, which cannot fit different data instances well.*

Under various t-SVD definitions of the transform domain, there are many algorithms and models for solving the low-rank problem. Since directly minimizing the tensor tubal rank of a tensor is an NP-hard problem, the tensor nuclear

\*Corresponding author

norm (TNN) (Zhang et al. 2014) was proposed as its convex approximation, which is

$$\|\mathcal{X}\|_{\mathbf{L},*} := \sum_{i=1}^{n_3} \|L(\mathcal{X})(:, :, i)\|_*, \quad (2)$$

where  $\|\cdot\|_*$  is the matrix nuclear norm and  $\|\cdot\|_{\mathbf{L},*}$  is the tensor nuclear norm based on the transform  $\mathbf{L}$ . To solve the matrix nuclear norm, the singular value thresholding (SVT) operator was proposed (Cai, Candès, and Shen 2010), which can be formulated as:

$$\text{SVT}_\gamma(\mathbf{X}) = \mathbf{U}\mathbf{S}_\gamma\mathbf{V}^T, \quad (3)$$

where  $\mathbf{S}_\gamma = \max\{\mathbf{S} - \gamma, 0\}$ . With the assistance of SVT, recent models utilizing optimization algorithms (Wang et al. 2022a; Qin et al. 2022) and deep unfolding networks (Mai, Lam, and Lee 2022) have demonstrated promising outcomes. However, training neural networks based on SVD (Ionescu, Vantzos, and Sminchisescu 2015a; Wang et al. 2023) poses challenges due to its complexity and numerical instability when handling derivatives of eigenvectors. This can result in sub-optimal convergence during the training process. Despite the robust differentiability of eigenvectors, the partial derivatives may become large in cases where two eigenvalues are equal or nearly equal (Wang et al. 2022b).

To achieve a data-adaptive and theoretically invertible transform, fostering compatibility with deep networks, we introduce a novel approach: the learnable Orthogonal Transform-induced generative Low-Rank t-SVD Model (OTLRM). Compared with existing t-SVD methods, our OTLRM has two merits: (1) **Orthogonality and data adaptability**. We construct the learnable orthogonal transform  $\mathbf{L}$  by several learnable Householder transformations, possessing inherent orthogonality and adaptability. It enables the adjustment of the transform for each dataset while maintaining a theoretical guarantee for the requirement of “arbitrary invertible linear transform”. While current t-SVD methods (such as DTNN (Kong, Lu, and Lin 2021) and Tensor  $Q$ -rank (Jiang et al. 2023)) can only accommodate one or the other. (2) **Generative framework**. Under the t-SVD representation, OTLRM can directly generate the expected tensor with the guidance of the observations within the DNN optimization framework. While others decompose the target tensor by SVD and truncate the singular values. Especially, we introduce a dense rank estimation operator, which stores and enriches the rank information of each band in the transform domain. Primary contributions are outlined as follows:

- Prompted by the linear algebra theorem of the Householder transformation, we construct a learnable endogenously orthogonal transform into the neural network. Different from the predefined orthogonal transform, the proposed learnable endogenously orthogonal transform can be naturally embedded in the neural network, which has more flexible data-adaptive capability.
- With the endogenously orthogonal transform, we propose a low-rank solver as a generalization of SVT, which utilizes efficient t-SVD representation to obtain low-rank structures. In contrast to SVT-induced low-rank algorithms, OTLRM is solved by gradient descent-based algorithms within the DNN optimization framework.

- We conduct extensive experiments on several tasks: TC, snapshot compressive imaging, spectral denoising, under three types of datasets: MSIs, videos and MRI data. Abundant and superior experimental results validate the effectiveness of our method.

## Related Works

**Transform-based t-SVD.** Inspired by TNN (Zhang et al. 2014) in the Fourier domain, many t-SVD algorithms have been proposed. Specifically, Zhang et al. proposed the DFT-based t-SVD method (Zhang and Aeron 2016). Zhou et al. (Zhou and Cheung 2019) introduced the Bayesian version of tensor tubal rank to automatically determine the tensor rank. The weighted TNN (Mu et al. 2020) is also proposed to distinguish different tensor singular values under Fast Fourier Transform (FFT). Considering that t-product is defined on any invertible linear transforms, other transforms are explored. For example, Lu et al. (Lu, Peng, and Wei 2019) proposed the DCT-induced TNN with theoretical guarantees for exact recovery. While Song et al. (Song, Ng, and Zhang 2020) employed unitary transform matrices captured from the original datasets. However, fixed transforms can not be inherently suitable for current data instances. Therefore, Jiang et al. (Jiang et al. 2023) proposed a dictionary-based TNN (DTNN), which constructed a learnable data-adaptive dictionary as the transform. Additionally, Kong et al. (Kong, Lu, and Lin 2021) gave a definition of the tensor rank with a data-dependent transform, based on the learnable matrix  $\mathbf{Q}$ . Recently, Luo et al. (Luo et al. 2022) induced DNN as the transform and generated the transformed tensor by gradient descent algorithms. Due to the nonlinear transform used in the DNN, it does not fulfill the theoretical requirements for arbitrary invertible linear transforms.

**SVT in Neural Network.** Cai et al. proposed the SVT algorithm for matrix completion (Cai, Candès, and Shen 2010). To implement the SVT algorithm into the deep learning network, Ionescu et al. gave a sound mathematical apparatus and derived the gradient propagation formulas for SVD in deep networks (Ionescu, Vantzos, and Sminchisescu 2015a,b). However, although robustly calculating the derivatives of the SVD gradient propagation is direct, it becomes numerically unstable during the calculation of certain partial derivatives. Wang et al. introduced a Taylor expansion-based approach to compute the eigenvector gradients (Wang et al. 2022b) and Song et al. induced the orthogonality loss to improve the generalization abilities and training stability of the SVD (Song, Sebe, and Wang 2022). These approaches involve integrating the matrix-based SVT directly into deep neural networks, which still introduces complex SVD to solve low-rank problems.

## Notations and Preliminaries

### Notations

In this paper, scalars, vectors, matrices, and tensors are denoted respectively as lowercase letters, e.g.  $a$ , boldface lowercase letters, e.g.  $\mathbf{a}$ , boldface capital letters, e.g.  $\mathbf{A}$  and boldface Calligraphy letters, e.g.  $\mathcal{A}$ .  $\mathcal{A}^{(i)}$  represents the  $i$ -th frontal slice of the tensor  $\mathcal{A}$ .

## Orthogonal Transform

Orthogonal transform could maintain the original orthogonality between vectors without losing fine-level details. Based on the following linear algebra theorem, an orthogonal matrix  $\mathbf{L}$  is constructed, which satisfies  $\mathbf{L}^T \mathbf{L} = \mathbf{L} \mathbf{L}^T = \mathbf{I}$  and  $\mathbf{I}$  is the identity matrix. Proofs of Theorem 1 are presented in the supplementary material.

**Theorem 1.** (Uhlig 2001) *Every real orthogonal  $n \times n$  matrix  $\mathbf{L}$  is the product of at most  $n$  real orthogonal Householder transformations. And this also is true for complex unitary  $\mathbf{L}$  and complex Householders.*

Following the lines of Theorem 1, we can randomly generate a parameter matrix  $\mathbf{W} \in \mathbb{R}^{n \times n}$  containing  $n$  column vectors  $\mathbf{w}_i \in \mathbb{R}^{n \times 1}$ , and construct  $n$  orthogonal Householder transformations as below:

$$\mathbf{F}_i = \mathbf{I} - 2 \frac{\mathbf{w}_i \mathbf{w}_i^T}{\|\mathbf{w}_i\|^2}, \quad (4)$$

where  $1 \leq i \leq n$ . Then the orthogonal matrix  $\mathbf{L}$  is represented as the product of the  $n$  orthogonal Householder transformations  $\mathbf{F}_i$ :

$$\mathbf{L} = \mathbf{F}_1 \mathbf{F}_2 \cdots \mathbf{F}_n = \prod_{i=1}^n \left( \mathbf{I} - 2 \frac{\mathbf{w}_i \mathbf{w}_i^T}{\|\mathbf{w}_i\|^2} \right), \quad (5)$$

and it is worth noting that  $\mathbf{W}$  is the parameter matrix to be optimized and  $\mathbf{L}$  is the endogenous orthogonal matrix which is another form of  $\mathbf{W}$  based on the Householder transformation operations of Eq. (4) and Eq. (5).

In contrast to approaches that attain orthogonality by incorporating an orthogonality metric into the loss function, matrices formed through Householder transformations possess inherent orthogonality and learnability. This intrinsic property eliminates the need for external orthogonal constraints and ensures identical learnable parameters as a conventional transform matrix. Specifically, for optimizing the orthogonal transform, this chain multiplication enables the loss to be derived parallelly for each column vector  $\mathbf{w}_i$ . And all the optimization can be automatically solved by the built-in differentiation engine in PyTorch, facilitating straightforward integration into deep neural networks.

## $L$ -transformed Tensor Singular Value Decomposition

Distinguished from matrix SVD, t-SVD is based on the tensor-tensor product, which is defined directly in a transform domain with an arbitrary invertible linear transform  $\mathbf{L}$  (Kernfeld, Kilmer, and Aeron 2015).

**Definition 1. Mode-3 tensor-matrix product (Kolda and Bader 2009)** For any third-order tensor  $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$  with a matrix  $\mathbf{U} \in \mathbb{R}^{n \times n_3}$ , the mode-3 tensor-matrix product is defined as

$$\hat{\mathcal{A}} = \mathcal{A} \times_3 \mathbf{U} \Leftrightarrow \hat{\mathbf{A}}_{(3)} = \mathbf{U} \mathbf{A}_{(3)}, \quad (6)$$

where  $\hat{\mathcal{A}} \in \mathbb{R}^{n_1 \times n_2 \times n}$ ,  $\mathbf{A}_{(3)}$  and  $\hat{\mathbf{A}}_{(3)}$  are mode-3 matricization of  $\mathcal{A}$  and  $\hat{\mathcal{A}}$ , respectively.

**Definition 2. Tensor-Tensor face-wise product (Kernfeld, Kilmer, and Aeron 2015)** Given two tensors  $\mathcal{A} \in$

$\mathbb{R}^{n_1 \times \ell \times n_3}$  and  $\mathcal{B} \in \mathbb{R}^{\ell \times n_2 \times n_3}$ , the face-wise product of  $\mathcal{A}$  and  $\mathcal{B}$  is defined as

$$(\mathcal{A} \Delta \mathcal{B})^{(i)} = \mathcal{A}^{(i)} \mathcal{B}^{(i)}, \quad (7)$$

where  $\mathcal{A}^{(i)}$  is  $i$ -th frontal slice of  $\mathcal{A}$ .

**Definition 3. Tensor-tensor product in  $L$ -transform domain (Kernfeld, Kilmer, and Aeron 2015)** Define  $*_L$  as  $\mathbb{R}^{n_1 \times \ell \times n_3} \times \mathbb{R}^{\ell \times n_2 \times n_3} \rightarrow \mathbb{R}^{n_1 \times n_2 \times n_3}$ , we have the tensor-tensor product:

$$\mathcal{A} *_L \mathcal{B} = L^{-1}(L(\mathcal{A}) \Delta L(\mathcal{B})), \quad (8)$$

where  $L(\mathcal{A}) = \mathcal{A} \times_3 \mathbf{L}$  and  $L^{-1}(\cdot)$  is the inverse transform operator of  $L(\cdot)$ .

**Definition 4. Special tensors (Braman 2010; Kilmer et al. 2013)** **Identity tensor:** for an identity tensor  $\mathcal{I}$ , its every frontal slice is an identity matrix.  **$f$ -diagonal tensor:** for an  $f$ -diagonal tensor  $\mathcal{S}$ , its every frontal slice is a diagonal matrix. **Orthogonal tensor:** if the tensor  $\mathcal{U}$  satisfies that  $\mathcal{U}^T *_L \mathcal{U} = \mathcal{U} *_L \mathcal{U}^T = \mathcal{I}$ , it is called orthogonal tensor. **Semi-orthogonal tensor:** if the tensor  $\mathcal{U}$  satisfies  $\mathcal{U}^T *_L \mathcal{U} = \mathcal{I}$ , it is called semi-orthogonal tensor.

Subsequently, the definition of the t-SVD is expressed in terms of the  $L$ -transform and its inverse.

**Lemma 1. Tensor singular value decomposition (t-SVD) (Braman 2010; Kilmer and Martin 2011; Kilmer et al. 2013)** Given a tensor  $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ , the t-SVD can be formulated as

$$\mathcal{X} = \mathcal{U} *_L \mathcal{S} *_L \mathcal{V}^T, \quad (9)$$

where  $\mathcal{U} \in \mathbb{R}^{n_1 \times n_1 \times n_3}$ ,  $\mathcal{V} \in \mathbb{R}^{n_2 \times n_2 \times n_3}$  are orthogonal tensors and  $\mathcal{S} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$  is  $f$ -diagonal.

**Definition 5. Tensor tubal-rank (Zhang et al. 2014)** The tensor tubal-rank  $r$  of the target tensor  $\mathcal{X}$  is defined as the number of the non-zero singular tubes of the  $f$ -diagonal tensor  $\mathcal{S}$ , which is

$$\text{rank}(\mathcal{X}) = \#\{i, \mathcal{S}(i, i, :) \neq 0\}. \quad (10)$$

And an alternative definition is that the tensor tubal-rank of  $\mathcal{X}$  is the largest rank of every frontal slice of the  $L$ -transformed tensor  $L(\mathcal{X})$ .

**Remark 1. Skinny t-SVD (Kilmer et al. 2013; Zhang and Aeron 2016)** Given a tensor  $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$  which has tensor tubal-rank  $r$ , it's more efficient to compute the skinny t-SVD. And the decomposition can be reformulated as

$$\mathcal{X} = \mathcal{U} *_L \mathcal{S} *_L \mathcal{V}^T, \quad (11)$$

where  $\mathcal{U} \in \mathbb{R}^{n_1 \times r \times n_3}$ ,  $\mathcal{V} \in \mathbb{R}^{n_2 \times r \times n_3}$  are semi-orthogonal tensors,  $\mathcal{S} \in \mathbb{R}^{r \times r \times n_3}$  is  $f$ -diagonal and especially,  $\mathcal{U}^T *_L \mathcal{U} = \mathcal{I}$ ,  $\mathcal{V}^T *_L \mathcal{V} = \mathcal{I}$ .

**Lemma 2. Tensor Singular Value Thresholding (t-SVT) (Wang et al. 2022a)** Let  $\mathcal{X} = \mathcal{U} *_L \mathcal{S} *_L \mathcal{V}^T$  be the t-SVD for tensor  $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ . The t-SVT operator  $\delta$  is:

$$\delta_\gamma(\mathcal{X}) = \mathcal{U} *_L \mathcal{S}_\gamma *_L \mathcal{V}^T, \quad (12)$$

where  $\mathcal{S}_\gamma = L^{-1}(\max\{L(\mathcal{S}) - \gamma, 0\})$  and  $\gamma$  is the threshold value which controls the degree of the rank restriction.

## Learnable Endogenous Orthogonal Transform based Generative t-SVD Low-rank Model

Suppose that  $\mathcal{Y} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$  denotes the unknown pure tensor (real scene or video),  $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$  is the desired low-rank tensor and  $\mathbf{H}(\cdot)$  is defined as the capture operation that obtains the source data (hence,  $\mathbf{H}(\mathcal{Y})$  is the observed measurement). Based on the definition of transposition and orthogonality of the transform of the t-product, the desired low-rank tensor can be generated by the skinny low-rank t-SVD representation, which is  $\mathcal{X} = \mathcal{U} *_L \mathcal{S} *_L \mathcal{V}^T$ . With the endogenous orthogonal transform  $\mathbf{L} \in \mathbb{R}^{n_3 \times n_3}$  (and the parameter matrix to be optimized is  $\mathbf{W}$ ) which naturally satisfies  $\mathbf{L}^T \mathbf{L} = \mathbf{L} \mathbf{L}^T = \mathbf{I}$ , our generative t-SVD low-rank optimization model can be formulated as

$$\min_{\mathcal{U}, \mathcal{V}, \mathcal{S}, \mathbf{W}} \phi(\mathbf{H}(\mathcal{X}), \mathbf{H}(\mathcal{Y})), \text{ s.t. } \mathcal{X} = \mathcal{U} *_L \mathcal{S} *_L \mathcal{V}^T, \quad (13)$$

where  $\phi(\cdot)$  is the fidelity loss function which can be adjusted according to the target application,  $\mathcal{U} \in \mathbb{R}^{n_1 \times r \times n_3}$ ,  $\mathcal{V} \in \mathbb{R}^{n_2 \times r \times n_3}$  and  $\mathcal{S} \in \mathbb{R}^{r \times r \times n_3}$ . And explicitly,  $*_L$  denotes the orthogonal transform  $\mathbf{L}$ -based tensor-tensor product according to Eq. (5) and Eq. (8).

Since  $\mathcal{S}$  is an  $f$ -diagonal tensor, we construct the rank tensor by a transformed matrix  $\mathbf{S} \in \mathbb{R}^{n_3 \times r}$  instead, with the diagonalization  $Diag(\cdot)$ . It is reformulated as (please refer to the supplementary material for details):

$$\min_{\mathcal{U}, \mathcal{V}, \mathbf{S}, \mathbf{W}} \phi(\mathbf{H}(L^{-1}(L(\mathcal{U}) \Delta Diag(\mathbf{S}) \Delta L(\mathcal{V})^T)), \mathbf{H}(\mathcal{Y})). \quad (14)$$

### Dense Rank Estimation Operator

Traditionally, t-SVT is adopted to truncate and restrict the rank matrix by the soft thresholding operator. However, controlling the low-rank degree of the generated tensor  $\mathcal{X}$  by simply adjusting the rank  $r$  and the threshold value  $\gamma$  is coarse, which does not well exploit the data adaptability in DNNs. Inspired by the threshold contraction and to find the most suitable tensor tubal rank, we inject a DNN-induced dense rank estimation operator  $\rho(\cdot)$  into the optimal solution of the factor  $\mathcal{S}$ , which captures and enriches the rank information. Specifically,  $\rho(\cdot)$  can be viewed as a rank information extractor. In this paper, we use the fully connected layer for experiments. Given the rank matrix  $\mathbf{S}$  as input, the dense rank estimation operator can be formulated as:

$$\rho(\mathbf{S}) = LReLU \cdots (LReLU(\mathbf{S} \times_3 \mathbf{G}_1) \times_3 \cdots) \times_3 \mathbf{G}_k, \quad (15)$$

where  $k$  denotes the number of the layers,  $LReLU$  is the LeakyReLU (He et al. 2015) function and each  $\mathbf{G}$  is the learnable rank feature matrix. The optimization model is reformulated as

$$\min_{\mathcal{U}, \mathcal{V}, \mathbf{S}, \mathbf{W}} \phi(\mathbf{H}(L^{-1}(L(\mathcal{U}) \Delta Diag(\rho(\mathbf{S})) \Delta L(\mathcal{V})^T)), \mathbf{H}(\mathcal{Y})). \quad (16)$$

### Orthogonal Total Variation

To improve the generative capability and get a more suitable transform, we adopt the Orthogonal Total Variation (OTV)

constraint for  $\Theta = \{L(\mathcal{U}), L(\mathcal{V}), \theta_{L^{-1}}\}$ , where  $\theta_{L^{-1}}$  denotes the weight matrix of the  $L^{-1}(\cdot)$  module, aiming to enhance the local smoothness prior structure, which is

$$OTV(\Theta) = \|\nabla_x L(\mathcal{U})\|_{\ell_1} + \|\nabla_y L(\mathcal{V})^T\|_{\ell_1} + \|\nabla_x \theta_{L^{-1}}\|_{\ell_1}. \quad (17)$$

As a result, the final optimization model is

$$\min_{\mathcal{U}, \mathcal{V}, \mathbf{S}, \mathbf{W}} \phi(\mathbf{H}(L^{-1}(L(\mathcal{U}) \Delta Diag(\rho(\mathbf{S})) \Delta L(\mathcal{V})^T)), \mathbf{H}(\mathcal{Y})) + \lambda OTV(\Theta), \quad (18)$$

where  $\lambda$  is a trade-off parameter, and the complete algorithm is depicted in Algorithm 1.

---

Algorithm 1: The Proposed OTLRM Algorithm.

---

**Input:** The coarse estimated rank  $r$ , hyperparameter  $\lambda$ , and the maximum iteration  $t_{max}$ .

**Output:** The reconstructed tensor  $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ .

- 1: **Initialization:** The iteration  $t = 0$ .
  - 2: **while**  $t < t_{max}$  **do**
  - 3:   Compute  $\mathbf{L}$  via Eq. (5);
  - 4:   Compute  $\hat{\mathcal{U}}$  via  $\hat{\mathcal{U}} = L(\mathcal{U})$ ;
  - 5:   Compute  $\hat{\mathcal{V}}$  via  $\hat{\mathcal{V}} = L(\mathcal{V})$ ;
  - 6:   Compute  $\hat{\mathcal{S}}$  via  $\hat{\mathcal{S}} = Diag(\rho(\mathbf{S}))$ ;
  - 7:   Compute the loss via Eq. (18);
  - 8:   Perform gradient backpropagation;
  - 9: **end while**
  - 10: Get the final low-rank tensor  $\mathcal{X} = L^{-1}(\hat{\mathcal{U}} \Delta \hat{\mathcal{S}} \Delta \hat{\mathcal{V}}^T)$ .
- 

## Applications for The Proposed Model

### Tensor Completion

Given the unknown pure tensor  $\mathcal{Y} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ , tensor completion aims to recover underlying data from the observed entries  $\Omega = \{(i_1, i_2, i_3) | \zeta_{i_1, i_2, i_3} = 1\}$ , which follows the Bernoulli sampling scheme  $\Omega \sim \mathbf{Ber}(p)$ . And  $p$  is the probability of taking target Bernoulli variables with independent and identically distributed. Based on the above definition,  $\mathbf{H}(\cdot)$  can be specified as  $\mathbf{H}_\Omega(\cdot) : \mathbb{R}^{n_1 \times n_2 \times n_3} \rightarrow \mathbb{R}^{n_1 \times n_2 \times n_3}$ , which keeps the entries in  $\Omega$  fixed and sets the rest zero. The loss function of  $\phi(\cdot)$  can be formulated as

$$\phi(\mathcal{X}, \mathcal{Y}) = \|\mathbf{H}_\Omega(\mathcal{X}) - \mathbf{H}_\Omega(\mathcal{Y})\|_F^2. \quad (19)$$

### MSI Reconstruction in CASSI System

Coded aperture snapshot compressive imaging (CASSI), which aims at scanning scenes with spatial and spectral dimensions, has achieved impressive performance. With a coded aperture and a disperser, it encodes and shifts each band of the original scene  $\mathcal{Y} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$  with known mask  $\mathbf{M} \in \mathbb{R}^{n_1 \times n_2}$  and later blends all the bands to generate a 2-D measurement  $\mathbf{X} \in \mathbb{R}^{n_1 \times (n_2 + d \times (n_3 - 1))}$ , where  $d$  is the shift step. Considering the measurement noise  $\mathbf{N} \in \mathbb{R}^{n_1 \times (n_2 + d \times (n_3 - 1))}$  generated in the coding system, the whole process can be formulated as

$$\mathbf{X} = \sum_{k=1}^{n_3} \text{shift}(\mathcal{Y}(:, :, k) \odot \mathbf{M}) + \mathbf{N}. \quad (20)$$

For convenience, by the definition of the  $\mathbf{H}(\cdot)$ , the above operations can be simplified to the following formula:

$$\mathbf{X} = \mathbf{H}(\mathcal{Y}) + \mathbf{N}, \quad (21)$$

where for MSI reconstruction, the  $\mathbf{H}(\cdot)$  operator is defined as  $\mathbb{R}^{n_1 \times n_2 \times n_3} \rightarrow \mathbb{R}^{n_1 \times (n_2 + d \times (n_3 - 1))}$  in CASSI application. And the loss function of  $\phi(\cdot)$  can be formulated as

$$\phi(\mathbf{X}, \mathcal{Y}) = \|\mathbf{X} - \mathbf{H}(\mathcal{Y})\|_F^2. \quad (22)$$

### MSI Denoising

The purpose of MSI denoising is to recover clean MSIs from noise-contaminated observations. In that case, the  $\mathbf{H}(\cdot)$  operator can be defined as a noise-adding operation. The loss function of  $\phi(\cdot)$  can be formulated as

$$\phi(\mathcal{X}, \mathcal{Y}) = \|\mathcal{X} - \mathbf{H}(\mathcal{Y})\|_{\ell_1}. \quad (23)$$

### Computational Complexity Analysis

Suppose that the target tensor  $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$  and the orthogonal transform  $\mathbf{L} \in \mathbb{R}^{n_3 \times n_3}$ . For the process of the transform construction, the computational complexity is  $O((n_3)^4)$ . For the transformed tensors  $\hat{\mathcal{U}}, \hat{\mathcal{V}}$ , the computational complexity is  $O(n_1 r (n_3)^2) + O(r n_2 (n_3)^2)$ . For the  $\hat{\mathcal{S}}$ , given the number of DNN layers  $k$ , the computational complexity is  $O(n_3 r^2 k)$ . And for the product of the  $\hat{\mathcal{U}}, \hat{\mathcal{V}}, \hat{\mathcal{S}}$ , the computational complexity is  $O(n_3 (n_1 r^2 n_2))$ . Thus, the computational complexity of our method is  $O((n_3)^4) + O(n_1 r (n_3)^2) + O(r n_2 (n_3)^2) + O(n_3 r^2 k) + O(n_3 (n_1 r^2 n_2))$ . Due to the fact that  $r \ll \min\{n_1, n_2\}$ , it can be simplified as  $O((n_3)^4 + (n_1 + n_2)r(n_3)^2 + n_1 n_2 n_3 r^2)$ .

## Experiments

### Datasets and Settings

In this section, we introduce the datasets (videos, MSIs and MRI data) and the experiment settings used in three multi-dimensional inverse problems. Please refer to the supplementary material for extended experiments.

**Overall Settings:** All the experiments are implemented in PyTorch and conducted on one NVIDIA GeForce RTX 3090 GPU with 20GB RAM. Adam (Kingma and Ba 2015) is used to optimize Eq. (16) and Eq. (18). For the rank estimation module  $\rho(\cdot)$ , we used a two-layer DNN (composed of two linear transforms and a LeakyReLU (He et al. 2015) function) in all experiments (the  $k$  in Eq. (15) is 2). For the initialization of the learnable parameters in the model, we follow the typical kaiming initialization (He et al. 2015) and please refer to the supplementary material for the initialization ablation experiments. It is noteworthy that to enhance the model's data adaptability, we employ the endogenously orthogonal transforms  $\mathbf{L}_1$  and  $\mathbf{L}_2$  to operate on  $\mathcal{U}$  and  $\mathcal{V}$ , respectively, while  $\mathbf{L}_3$  serves as the inverse.

**Tensor Completion:** The evaluation encompasses four MSIs sourced from the CAVE database <sup>1</sup>, namely *Balloon*, *Beer*, *Pompom* and *Toy*, along with two videos obtained from

the NTT database <sup>2</sup>, labeled *Bird* and *Horse*. Each MSI is resized to dimensions of  $256 \times 256 \times 31$ , while the videos have a spatial size of  $288 \times 352$ . In our experiments, we utilize the initial 30 frames of each video. Regarding the MSI datasets, we employ sampling rates (SRs) of 0.05, 0.10, and 0.15, while for the video datasets, the sampling rates are set to 0.10, 0.15, and 0.20.

**MSI Reconstruction in CASSI System:** We select five datasets (*scene01-scene05*,  $256 \times 256 \times 31$ ) in KAIST <sup>3</sup> (Choi et al. 2017) for simulation. The shift step  $d$  is 2, which means that the measurement is of size  $256 \times 310$ .

**MSI Denoising:** We select three scenes (*scene01*, *scene02* and *scene10*) in KAIST (Choi et al. 2017) database (all of size  $256 \times 256 \times 31$ ) for testing. And the noise cases are set for Gaussian noise with standard deviations 0.2 and 0.3.

### Comparisons with State-of-the-Arts

In this section, we compare our method with state-of-the-arts. The **best** and **second-best** are highlighted in bold and underlined, respectively. **OTLRM\*** represents the model without OTV loss, and **OTLRM** denotes the whole model.

**Evaluation Metrics:** For numerical comparison, we use peak signal-to-noise ratio (PSNR) and structural similarity (SSIM) as metrics in all three problems. And feature similarity index measure (FSIM) is added for MSI denoising problem. The higher PSNR, SSIM, and FSIM the better.

**Tensor Completion:** Especially, we compare with other state-of-the-art methods, including TNN (Lu et al. 2019), TQRTNN (Wu et al. 2021), UTNN (Song, Ng, and Zhang 2020), DTNN (Jiang et al. 2023), LS2T2NN (Liu et al. 2023) and HLRTF (Luo et al. 2022).

Table 1 and 2 shows the numerical results with 1 to 4 dB improvement in PSNR, especially in low sampling rate. Compared to the traditional TNN-based methods, our model demonstrates excellent results of learnable orthogonal transforms and the generative t-SVD model, which maintains flexible data-adaption low-rank property. More results are provided in the supplementary material.

**MSI Reconstruction in CASSI System:** We compare our method with DeSCI(Liu et al. 2018),  $\lambda$ -Net(Miao et al. 2019), TSA-Net(Meng, Ma, and Yuan 2020), HDNet(Hu et al. 2022), DGSMP(Huang et al. 2021), ADMM-Net(Ma et al. 2019), GAP-Net(Meng, Yuan, and Jalali 2023), PnP-CASSI (Zheng et al. 2021), DIP-HSI(Meng et al. 2021) and HLRTF(Luo et al. 2022).

Table 3 exhibits clear advantages over self-supervised and zero-shot methods, even having a slight increase of at least 0.34 dB on average over supervised methods. More results are presented in the supplementary material.

**MSI Denoising:** For MSI denoising, we select: NonLRMA(Chen et al. 2017), TLRSSSTV(Zeng et al. 2020), LLxRGTV(Zeng and Xie 2021), 3DTNN(Zheng et al. 2019), LRTDCTV(Zeng et al. 2023), E3DTV(Peng et al. 2020), DDRM(Kawar et al. 2022), DDS2M(Miao et al. 2023) and HLRTF (Luo et al. 2022).

<sup>1</sup><https://www.cs.columbia.edu/CAVE/databases/multispectral/>

<sup>2</sup><http://www.brl.ntt.co.jp/people/akisato/saliency3.html>.

<sup>3</sup><https://vclab.kaist.ac.kr/siggraphasia2017p1/kaistdataset.html>

Method	Reference	SR=0.05		SR=0.10		SR=0.15		Time (s)	SR=0.05		SR=0.10		SR=0.15		Time (s)
		PSNR $\uparrow$	SSIM $\uparrow$	PSNR $\uparrow$	SSIM $\uparrow$	PSNR $\uparrow$	SSIM $\uparrow$		PSNR $\uparrow$	SSIM $\uparrow$	PSNR $\uparrow$	SSIM $\uparrow$	PSNR $\uparrow$	SSIM $\uparrow$	
Observed	None	13.53	0.12	13.76	0.15	14.01	0.18	None	11.66	0.05	11.89	0.08	12.14	0.11	None
TNN(Lu et al. 2019)	TPAMI	31.69	0.87	36.27	0.94	39.66	0.97	12	23.34	0.56	28.39	0.76	31.74	0.85	13
TQRTNN(Wu et al. 2021)	TCI	31.68	0.87	36.26	0.94	39.66	0.97	27	23.33	0.56	28.39	0.76	31.70	0.85	33
UTNN(Song, Ng, and Zhang 2020)	NLAA	33.78	0.92	39.28	0.97	43.20	0.99	234	25.32	0.66	31.34	0.86	35.62	0.93	150
DTNN(Jiang et al. 2023)	TNNLS	35.44	0.94	40.51	0.98	44.22	0.99	417	28.05	0.78	32.48	0.89	36.17	0.94	409
LS2T2NN(Liu et al. 2023)	TCSVT	38.35	0.97	42.99	0.99	45.58	0.99	54	31.75	0.87	36.57	0.95	39.23	0.97	220
HLRTF(Luo et al. 2022)	CVPR	37.25	0.96	42.43	0.98	45.65	0.99	22	30.04	0.82	37.65	0.96	40.92	0.98	21
<b>OTLRM*</b>	Ours	37.64	0.95	42.68	0.98	46.03	0.99	117	28.99	0.74	35.12	0.91	39.44	0.96	108
<b>OTLRM</b>	Ours	<b>40.20</b>	<b>0.98</b>	<b>44.34</b>	<b>0.99</b>	<b>46.92</b>	<b>0.99</b>	126	<b>35.42</b>	<b>0.93</b>	<b>39.41</b>	<b>0.97</b>	<b>42.20</b>	<b>0.98</b>	125
Observed	None	9.65	0.02	9.89	0.03	10.13	0.04	None	11.17	0.25	11.41	0.29	11.66	0.32	None
TNN(Lu et al. 2019)	TPAMI	32.41	0.88	37.81	0.96	41.35	0.98	13	27.38	0.82	31.45	0.90	34.42	0.94	13
TQRTNN(Wu et al. 2021)	TCI	32.46	0.88	37.57	0.95	41.28	0.98	33	27.12	0.81	31.43	0.90	34.37	0.94	33
UTNN(Song, Ng, and Zhang 2020)	NLAA	35.62	0.95	41.04	0.98	44.66	0.99	171	29.40	0.87	35.00	0.95	39.05	0.98	190
DTNN(Jiang et al. 2023)	TNNLS	35.95	0.95	41.29	0.97	45.05	0.99	415	30.34	0.90	35.38	0.96	39.33	0.97	411
LS2T2NN(Liu et al. 2023)	TCSVT	38.67	0.96	43.36	0.98	46.25	0.98	87	32.12	0.91	36.56	0.96	40.23	0.98	142
HLRTF(Luo et al. 2022)	CVPR	37.20	0.95	42.76	0.98	45.73	0.99	21	33.00	0.92	39.01	0.97	42.95	0.99	21
<b>OTLRM*</b>	Ours	39.16	0.97	43.53	0.99	46.92	0.99	111	32.63	0.91	38.41	0.97	42.46	0.98	113
<b>OTLRM</b>	Ours	<b>41.74</b>	<b>0.98</b>	<b>45.09</b>	<b>0.99</b>	<b>47.52</b>	<b>0.99</b>	118	<b>36.24</b>	<b>0.97</b>	<b>41.11</b>	<b>0.99</b>	<b>44.14</b>	<b>0.99</b>	128

Table 1: Evaluation on CAVE of tensor completion by different methods for MSIs under different SRs. Top Left: *Balloons*, Bottom Left: *Beer*; Top Right: *Pompom*, Bottom Right: *Toy*.

Method	Reference	SR=0.10		SR=0.15		SR=0.20		Time (s)	SR=0.10		SR=0.15		SR=0.20		Time (s)
		PSNR $\uparrow$	SSIM $\uparrow$	PSNR $\uparrow$	SSIM $\uparrow$	PSNR $\uparrow$	SSIM $\uparrow$		PSNR $\uparrow$	SSIM $\uparrow$	PSNR $\uparrow$	SSIM $\uparrow$	PSNR $\uparrow$	SSIM $\uparrow$	
Observed	None	7.74	0.02	7.98	0.03	8.25	0.04	None	6.53	0.01	6.78	0.02	7.04	0.02	None
TNN(Lu et al. 2019)	TPAMI	27.22	0.76	29.28	0.83	31.08	0.87	17	26.81	0.66	28.42	0.73	29.79	0.79	17
TQRTNN(Wu et al. 2021)	TCI	27.41	0.77	29.55	0.83	31.33	0.88	58	26.93	0.67	28.64	0.74	29.95	0.79	51
UTNN(Song, Ng, and Zhang 2020)	NLAA	27.83	0.79	30.11	0.85	31.95	0.89	156	27.27	0.68	29.00	0.76	30.30	0.81	146
DTNN(Jiang et al. 2023)	TNNLS	28.55	0.83	30.60	0.88	32.51	0.91	503	27.52	0.74	29.42	0.80	30.98	0.85	580
LS2T2NN(Liu et al. 2023)	TCSVT	30.40	0.85	33.38	0.91	35.55	0.94	63	29.48	0.76	31.44	0.83	33.45	0.88	83
HLRTF(Luo et al. 2022)	CVPR	26.25	0.70	28.50	0.78	31.27	0.86	22	25.73	0.60	26.37	0.63	28.61	0.76	23
<b>OTLRM</b>	Ours	<b>34.24</b>	<b>0.93</b>	<b>36.04</b>	<b>0.95</b>	<b>37.44</b>	<b>0.96</b>	124	<b>30.64</b>	<b>0.83</b>	<b>32.20</b>	<b>0.87</b>	<b>33.71</b>	<b>0.90</b>	125

Table 2: Evaluation on NTT of tensor completion by different methods for videos under different SRs. Left: *Bird*, Right: *Horse*.

Table 4 shows the performance with three metrics. More results are provided in the supplementary material. Compared with model-based methods, our OTLRM can deeply capture the low-rankness and help improve the abilities of MSI denoising within DNN. For diffusion-based methods and deep prior-induced PnP methods, our self-supervised method can also enhance the performance and achieve comparable results. Compared with the tensor-based method HLRTF, our method is more informative and smoother.

### Ablation Analysis

In this section, we do ablation experiments on the hyperparameter tensor rank  $r$ , the trade-off parameter  $\lambda$ , the number  $k$  of the DNN layers in the rank estimation module  $\rho(\cdot)$  and the visual results of the weight matrices in  $\rho(\cdot)$  respectively to demonstrate the advantages of our model. Additionally, in traditional skinny t-SVD algorithm,  $\mathcal{U}$  and  $\mathcal{V}$  are semi-orthogonal tensors. However, our model does not introduce semi-orthogonal constraints on these tensors. We explain this in the below ablation analysis. What's more, for the advantages of the orthogonal transform  $\mathbf{L}$ , we also compare our OTLRM with two typical methods, DTNN (Kong, Lu, and Lin 2021) and Tensor  $Q$ -rank (Jiang et al. 2023). And more ablation analysis and experiments are provided in the supplementary material. MSI *Balloons* (of size  $256 \times 256 \times 31$ ) in CAVE datasets is selected for TC.

**Effect of  $r$ :** The tensor rank  $r \in [1, \min\{n_1, n_2\}]$  which is

an integer characterizes the low-rankness of the generated tensor  $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ . In the experiment, we choose the rank  $r$  ranging from 10 to 100 at intervals of 10. From Figure 12 in the supplementary material, when the rank  $r$  is low, the performance is less effective due to the lost information of the original tensor. When the rank is too high, the result tensor is not well guaranteed to be low-rank, which leads to a sub-optimal performance. Thus, given  $n = \min\{n_1, n_2\}$ , the rank  $r$  can be set between  $[n/20, n/5]$ .

**Effect of  $\lambda$ :** The hyperparameter  $\lambda$  which controls the OTV loss is set using a strategy of sampling one point every order of magnitude from  $1e-1$  to  $1e-10$ . From Figure 13 in the supplementary material, we can find that when  $\lambda$  is large, the performance is decreased, probably because too strong the OTV loss aggravates the local similarities and loses the detailed features. Our method remains effective when  $\lambda$  falls between  $1e-7$  and  $1e-9$ . Thus, our model is easy to tune for the best  $\lambda$ . And to highlight our novelty in low-rank, more results are available in the supplementary material.

**Effect of  $k$ :** In all experiments, we simply use a two-layer DNN to represent the rank estimation module  $\rho(\cdot)$ . And we analyze the effect of  $\rho(\cdot)$  on the results by adjusting the number  $k$  of layers of the DNN. Here,  $k = 1$  indicates that only one LeakyReLU activation function is added. From Table 5, deeper DNN may have the potential for better results due to their better fitting ability.

**Effect of  $\rho(\cdot)$ :** For the rank information extractor  $\rho(\cdot)$ , we

Method	Category	Reference	scene01		scene02		scene03		scene04		scene05		Avg	
			PSNR $\uparrow$	SSIM $\uparrow$	PSNR $\uparrow$	SSIM $\uparrow$	PSNR $\uparrow$	SSIM $\uparrow$	PSNR $\uparrow$	SSIM $\uparrow$	PSNR $\uparrow$	SSIM $\uparrow$	PSNR $\uparrow$	SSIM $\uparrow$
DeSCI(Liu et al. 2018)	Model	TPAMI	28.38	0.80	26.00	0.70	23.11	0.73	28.26	0.86	25.41	0.78	26.23	0.77
$\lambda$ -Net(Miao et al. 2019)	CNN (*)	ICCV	30.10	0.85	28.49	0.81	27.73	0.87	37.01	0.93	26.19	0.82	29.90	0.86
TSA-Net(Meng, Ma, and Yuan 2020)	CNN (*)	ECCV	32.31	0.89	31.03	0.86	32.15	0.92	37.95	0.96	29.47	0.88	32.58	0.90
HDNet(Hu et al. 2022)	Transformer (*)	CVPR	<b>34.96</b>	<b>0.94</b>	<b>35.64</b>	<b>0.94</b>	<b>35.55</b>	<b>0.94</b>	41.64	<b>0.98</b>	32.56	<b>0.95</b>	<b>36.07</b>	<b>0.95</b>
DGSMP(Huang et al. 2021)	Unfolding (*)	CVPR	33.26	<u>0.92</u>	32.09	<u>0.90</u>	33.06	0.93	40.54	0.96	28.86	0.88	33.56	0.92
ADMM-Net(Ma et al. 2019)	Unfolding (*)	ICCV	34.03	<u>0.92</u>	<u>33.57</u>	<u>0.90</u>	34.82	0.93	39.46	<u>0.97</u>	31.83	0.92	34.74	<u>0.93</u>
GAP-Net(Meng, Yuan, and Jalali 2023)	Unfolding (*)	IJCV	33.63	0.91	<u>33.19</u>	<u>0.90</u>	33.96	0.93	39.14	<u>0.97</u>	31.44	0.92	34.27	<u>0.93</u>
PnP-CASSI(Zheng et al. 2021)	PnP (Zero-Shot)	PR	29.09	0.80	28.05	0.71	30.15	0.85	39.17	0.94	27.45	0.80	30.78	0.82
DIP-HSI(Meng et al. 2021)	PnP (Zero-Shot)	ICCV	31.32	0.86	25.89	0.70	29.91	0.84	38.69	0.93	27.45	0.80	30.65	0.82
HLRTF(Luo et al. 2022)	Tensor ( $\circ$ )	CVPR	34.56	0.91	33.37	0.87	<u>35.55</u>	<u>0.94</u>	<u>43.56</u>	<b>0.98</b>	<u>33.08</u>	<u>0.93</u>	36.02	<u>0.93</u>
<b>OTLRM</b>	Tensor ( $\circ$ )	Ours	<b>35.03</b>	0.91	32.90	0.82	<b>36.25</b>	<b>0.95</b>	<b>44.75</b>	<b>0.98</b>	<b>33.13</b>	0.92	<b>36.41</b>	0.92

Table 3: Evaluation on *KAIST* of **MSI Reconstruction in CASSI**. \* for supervised methods,  $\circ$  for self-supervised methods.

Method	Reference	PSNR $\uparrow$	SSIM $\uparrow$	FSIM $\uparrow$	Time (s)	Method	Reference	PSNR $\uparrow$	SSIM $\uparrow$	FSIM $\uparrow$	Time (s)
Noisy	None	16.18	0.12	0.401	None	Noisy	None	12.98	0.06	0.320	None
NonLRMA (Chen et al. 2017)	TGRS	21.26	0.41	0.803	<u>11</u>	NonLRMA (Chen et al. 2017)	TGRS	20.30	0.36	0.772	<u>11</u>
TLRLSSTV (Zeng et al. 2020)	TGRS	24.88	0.53	0.767	76	TLRLSSTV (Zeng et al. 2020)	TGRS	22.82	0.41	0.689	76
LLxRGTV (Zeng and Xie 2021)	SP	<u>31.15</u>	0.80	0.917	38	LLxRGTV (Zeng and Xie 2021)	SP	27.64	0.68	0.868	38
3DTNN (Zheng et al. 2019)	TGRS	28.04	0.78	0.881	20	3DTNN (Zheng et al. 2019)	TGRS	26.04	0.72	0.848	20
LRTDCTV (Zeng et al. 2023)	JSTAR	25.95	0.66	0.816	43	LRTDCTV (Zeng et al. 2023)	JSTAR	24.59	0.53	0.739	42
E3DTV (Peng et al. 2020)	TIP	30.34	<b>0.87</b>	<u>0.926</u>	<b>10</b>	E3DTV (Peng et al. 2020)	TIP	28.36	<b>0.82</b>	<u>0.900</u>	<b>9</b>
DDRM (Kawar et al. 2022)	NeurIPS	29.41	<b>0.87</b>	0.922	20	DDRM (Kawar et al. 2022)	NeurIPS	27.81	<u>0.79</u>	0.893	23
DDS2M (Miao et al. 2023)	ICCV	<b>32.80</b>	0.79	0.895	354	DDS2M (Miao et al. 2023)	ICCV	30.08	0.67	0.834	318
HLRTF (Luo et al. 2022)	CVPR	30.42	<u>0.83</u>	<b>0.940</b>	23	HLRTF (Luo et al. 2022)	CVPR	<u>30.34</u>	0.69	0.874	25
<b>OTLRM</b>	Ours	30.82	<u>0.72</u>	0.881	45	<b>OTLRM</b>	Ours	<b>32.56</b>	<b>0.82</b>	<b>0.902</b>	47

Table 4: Evaluation on *KAIST* of **MSI Denoising**. Left: case:  $\mathcal{N}(0, 0.2)$ -scene10, Right: case:  $\mathcal{N}(0, 0.3)$ -scene01.

$k$ -layers	SR=0.05		SR=0.10		SR=0.15	
	PSNR $\uparrow$	SSIM $\uparrow$	PSNR $\uparrow$	SSIM $\uparrow$	PSNR $\uparrow$	SSIM $\uparrow$
0	35.82	0.94	40.39	<u>0.98</u>	41.92	<u>0.98</u>
1	38.72	0.97	43.10	<b>0.99</b>	45.12	<b>0.99</b>
2	<b>40.20</b>	<b>0.98</b>	<u>44.34</u>	<b>0.99</b>	<b>46.92</b>	<b>0.99</b>
3	<u>40.15</u>	<b>0.98</b>	<b>44.56</b>	<b>0.99</b>	<u>46.85</u>	<b>0.99</b>

Table 5: Effect of  $k$  in the rank estimation module  $\rho(\cdot)$ .

adopted a two-layer deep neural network that contains one LeakyReLU layer sandwiched between two linear layers as a simple implementation. Visual results of the two weight matrices are shown in Figure 15 in the supplementary material. It can be seen that the weight matrices learn different and rich rank information for different dimensions and different positions of the rank matrix  $\mathbf{S}$  with low-rank property. **Analysis for The Semi-orthogonality of  $\mathcal{U}$  and  $\mathcal{V}$ :** The semi-orthogonality of the  $\mathcal{U}$  and  $\mathcal{V}$  is naturally and strictly guaranteed by the process of the SVD. However, our generative t-SVD model is solved based on the gradient descent-based algorithm, which means that we do not need the SVD. To verify the validity of the semi-orthogonality of  $\mathcal{U}$  and  $\mathcal{V}$ , we construct a semi-orthogonal loss term for ablation experiments, which is:

$$\Psi(\mathcal{U}, \mathcal{V}) = \beta(\|\mathcal{U}^T *_{L} \mathcal{U} - \mathcal{I}\|_F^2 + \|\mathcal{V}^T *_{L} \mathcal{V} - \mathcal{I}\|_F^2). \quad (24)$$

Figure 14 in the supplementary material shows the completion results under different penalty coefficient  $\beta$ . It can be observed that the results show a decreasing trend as  $\beta$  continues to rise. Therefore, we relax the semi-orthogonal constraints on  $\mathcal{U}$  and  $\mathcal{V}$  to enhance the performance.

**Comparisons on Orthogonal Transform  $\mathbf{L}$ :** Table 3 in the supplementary material shows the evaluation results of the above methods. **(1)** The transform of DTNN is a redundant dictionary with constraints on each column of the dictionary to have an F-norm of one. It may not technically be called orthogonal, which makes its inverse transform difficult to calculate. To obtain the transform  $\mathbf{L} \in \mathbb{R}^{n_3 \times d}$ , the computational complexity of DTNN is  $O((n_3)^3 d^2)$  and ours is  $O((n_3)^4)$  (running time **415s** vs **118s**). **(2)** Tensor  $Q$ -rank imposes constraints on the transform process with a “two-step” strategy. This involves first finding an appropriate transform based on certain selection criteria and then incorporating it into the reconstruction process. This “separate” strategy may not well capture the intrinsic characteristics of the data. In contrast, our learnable transform can be directly embedded and inherently achieves orthogonality.

## Conclusion

In this paper, we proposed a learnable orthogonal transform-induced generative low-rank framework based on t-SVD for multi-dimensional tensor recovery, possessing the orthogonal and learnable transform which enables flexible data-adaptive capability while maintains theoretical guarantee for “arbitrary invertible linear transform”. Constructed by a series of Householder transformation units, this transform can be learnable and seamlessly integrated into the neural network with endogenous orthogonality. Compared to traditional solutions of t-SVD, our generative t-SVD representation model can naturally maintain the low-rank structure and be solved by gradient descent-based algorithms. Comprehensive experiments verify the effectiveness of our method in MSIs and videos for three practical problems.

## Acknowledgments

This work was supported in part by the National Natural Science Foundation of China under Grants 62106063 and 62276079, by the Guangdong Natural Science Foundation under Grant 2022A1515010819, and by National Natural Science Foundation of China Joint Fund Project Key Support Project under U22B2049.

## References

- Braman, K. 2010. Third-order tensors as linear operators on a space of matrices. *Linear Algebra and its Applications*, 433(7): 1241–1253.
- Cai, J.-F.; Candès, E. J.; and Shen, Z. 2010. A Singular Value Thresholding Algorithm for Matrix Completion. *SIAM Journal on Optimization*, 20(4): 1956–1982.
- Cai, Y.; Lin, J.; Wang, H.; Yuan, X.; Ding, H.; Zhang, Y.; Timofte, R.; and Gool, L. V. 2022. Degradation-aware unfolding half-shuffle transformer for spectral compressive imaging. In *Proceedings of the Advances in Neural Information Processing Systems*, 37749–37761.
- Carroll, J. D.; and Chang, J.-J. 1970. Analysis of individual differences in multidimensional scaling via an N-way generalization of “Eckart-Young” decomposition. *Psychometrika*, 35(3): 283–319.
- Chen, Y.; Guo, Y.; Wang, Y.; Wang, D.; Peng, C.; and He, G. 2017. Denoising of hyperspectral images using nonconvex low rank matrix approximation. *IEEE Transactions on Geoscience and Remote Sensing*, 55(9): 5366–5380.
- Choi, I.; Kim, M.; Gutierrez, D.; Jeon, D.; and Nam, G. 2017. High-quality hyperspectral reconstruction using a spectral prior. Technical report.
- He, K.; Zhang, X.; Ren, S.; and Sun, J. 2015. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In *Proceedings of the IEEE International Conference on Computer Vision*, 1026–1034.
- Hu, X.; Cai, Y.; Lin, J.; Wang, H.; Yuan, X.; Zhang, Y.; Timofte, R.; and Van Gool, L. 2022. Hdnet: High-resolution dual-domain learning for spectral compressive imaging. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 17542–17551.
- Huang, T.; Dong, W.; Yuan, X.; Wu, J.; and Shi, G. 2021. Deep gaussian scale mixture prior for spectral compressive imaging. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 16216–16225.
- Ionescu, C.; Vantzos, O.; and Sminchisescu, C. 2015a. Matrix Backpropagation for Deep Networks With Structured Layers. In *Proceedings of the IEEE International Conference on Computer Vision*, 2965–2973.
- Ionescu, C.; Vantzos, O.; and Sminchisescu, C. 2015b. Training deep networks with structured layers by matrix backpropagation. *arXiv preprint arXiv:1509.07838*.
- Jiang, T.-X.; Zhao, X.-L.; Zhang, H.; and Ng, M. K. 2023. Dictionary Learning With Low-Rank Coding Coefficients for Tensor Completion. *IEEE Transactions on Neural Networks and Learning Systems*, 34(2): 932–946.
- Kawar, B.; Elad, M.; Ermon, S.; and Song, J. 2022. Denoising diffusion restoration models. In *Proceedings of the Advances in Neural Information Processing Systems*, 23593–23606.
- Kernfeld, E.; Kilmer, M.; and Aeron, S. 2015. Tensor–tensor products with invertible linear transforms. *Linear Algebra and its Applications*, 485: 545–570.
- Kilmer, M. E.; Braman, K.; Hao, N.; and Hoover, R. C. 2013. Third-order tensors as operators on matrices: A theoretical and computational framework with applications in imaging. *SIAM Journal on Matrix Analysis and Applications*, 34(1): 148–172.
- Kilmer, M. E.; and Martin, C. D. 2011. Factorization strategies for third-order tensors. *Linear Algebra and its Applications*, 435(3): 641–658.
- Kingma, D. P.; and Ba, J. L. 2015. Adam: A Method for Stochastic Optimization. In *Proceedings of The International Conference on Learning Representations*.
- Kolda, T. G.; and Bader, B. W. 2009. Tensor decompositions and applications. *SIAM Review*, 51(3): 455–500.
- Kong, H.; Lu, C.; and Lin, Z. 2021. Tensor Q-rank: new data dependent definition of tensor rank. *Machine Learning*, 110(7): 1867–1900.
- Liu, S.; Leng, J.; Zhao, X.-L.; Zeng, H.; Wang, Y.; and Yang, J.-H. 2023. Learnable Spatial-Spectral Transform-Based Tensor Nuclear Norm for Multi-Dimensional Visual Data Recovery. *IEEE Transactions on Circuits and Systems for Video Technology*.
- Liu, S.; Zhao, X.-L.; Leng, J.; Li, B.-Z.; Yang, J.-H.; and Chen, X. 2024. Revisiting High-Order Tensor Singular Value Decomposition From Basic Element Perspective. *IEEE Transactions on Signal Processing*, 72: 4589–4603.
- Liu, Y.; Yuan, X.; Suo, J.; Brady, D. J.; and Dai, Q. 2018. Rank minimization for snapshot compressive imaging. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 41(12): 2990–3006.
- Lu, C. 2021. Transforms Based Tensor Robust PCA: Corrupted Low-Rank Tensors Recovery via Convex Optimization. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 1145–1152.
- Lu, C.; Feng, J.; Chen, Y.; Liu, W.; Lin, Z.; and Yan, S. 2019. Tensor robust principal component analysis with a new tensor nuclear norm. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 42(4): 925–938.
- Lu, C.; Peng, X.; and Wei, Y. 2019. Low-rank tensor completion with a new tensor nuclear norm induced by invertible linear transforms. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 5996–6004.
- Luo, Y.; Zhao, X.-L.; Meng, D.; and Jiang, T.-X. 2022. HLRTF: Hierarchical Low-Rank Tensor Factorization for Inverse Problems in Multi-Dimensional Imaging. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 19303–19312.

- Ma, J.; Liu, X.-Y.; Shou, Z.; and Yuan, X. 2019. Deep tensor admn-net for snapshot compressive imaging. In *Proceedings of the IEEE International Conference on Computer Vision*.
- Mai, T. T. N.; Lam, E. Y.; and Lee, C. 2022. Deep Unrolled Low-Rank Tensor Completion for High Dynamic Range Imaging. *IEEE Transactions on Image Processing*, 31: 5774–5787.
- Meng, Z.; Ma, J.; and Yuan, X. 2020. End-to-end low cost compressive spectral imaging with spatial-spectral self-attention. In *Proceedings of the European Conference on Computer Vision*, 187–204. Springer.
- Meng, Z.; Yu, Z.; Xu, K.; and Yuan, X. 2021. Self-supervised neural networks for spectral snapshot compressive imaging. In *Proceedings of the IEEE International Conference on Computer Vision*, 2622–2631.
- Meng, Z.; Yuan, X.; and Jalali, S. 2023. Deep unfolding for snapshot compressive imaging. *International Journal of Computer Vision*, 131(11): 2933–2958.
- Miao, X.; Yuan, X.; Pu, Y.; and Athitsos, V. 2019. l-net: Reconstruct hyperspectral images from a snapshot measurement. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 4059–4069.
- Miao, Y.; Zhang, L.; Zhang, L.; and Tao, D. 2023. Dds2m: Self-supervised denoising diffusion spatio-spectral model for hyperspectral image restoration. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 12086–12096.
- Mu, Y.; Wang, P.; Lu, L.; Zhang, X.; and Qi, L. 2020. Weighted tensor nuclear norm minimization for tensor completion using tensor-SVD. *Pattern Recognition Letters*, 130: 4–11.
- Peng, J.; Xie, Q.; Zhao, Q.; Wang, Y.; Yee, L.; and Meng, D. 2020. Enhanced 3DTV regularization and its applications on HSI denoising and compressed sensing. *IEEE Transactions on Image Processing*, 29: 7889–7903.
- Qin, W.; Wang, H.; Zhang, F.; Wang, J.; Luo, X.; and Huang, T. 2022. Low-rank high-order tensor completion with applications in visual data. *IEEE Transactions on Image Processing*, 31: 2433–2448.
- Song, G.; Ng, M. K.; and Zhang, X. 2020. Robust tensor completion using transformed tensor singular value decomposition. *Numerical Linear Algebra with Applications*, 27(3): e2299.
- Song, Y.; Sebe, N.; and Wang, W. 2022. Improving covariance conditioning of the SVD meta-layer by orthogonality. In *Proceedings of the European Conference on Computer Vision*, 356–372. Springer.
- Sun, Y.; Vong, C. M.; and Wang, S. 2022. Fast AUC Maximization Learning Machine With Simultaneous Outlier Detection. *IEEE Transactions on Cybernetics*.
- Tucker, L. R. 1966. Some mathematical notes on three-mode factor analysis. *Psychometrika*, 31(3): 279–311.
- Uhlig, F. 2001. Constructive ways for generating (generalized) real orthogonal matrices as products of (generalized) symmetries. *Linear Algebra and its Applications*, 332: 459–467.
- Wang, A.; Li, C.; Bai, M.; Jin, Z.; Zhou, G.; and Zhao, Q. 2023. Transformed Low-Rank Parameterization Can Help Robust Generalization for Tensor Neural Networks. *arXiv preprint arXiv:2303.00196*.
- Wang, M.; Wang, Q.; Hong, D.; Roy, S. K.; and Chanutot, J. 2022a. Learning tensor low-rank representation for hyperspectral anomaly detection. *IEEE Transactions on Cybernetics*, 53(1): 679–691.
- Wang, W.; Dang, Z.; Hu, Y.; Fua, P.; and Salzmann, M. 2022b. Robust Differentiable SVD. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(9): 5472–5487.
- Wu, F.; Li, Y.; Li, C.; and Wu, Y. 2021. A Fast Tensor Completion Method Based on Tensor QR Decomposition and Tensor Nuclear Norm Minimization. *IEEE Transactions on Computational Imaging*, 7: 1267–1277.
- Zeng, H.; Huang, S.; Chen, Y.; Luong, H.; and Philips, W. 2023. All of Low-rank and Sparse: A Recast Total Variation Approach to Hyperspectral Denoising. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*.
- Zeng, H.; and Xie, X. 2021. Hyperspectral image denoising via global spatial-spectral total variation regularized non-convex local low-rank tensor approximation. *Signal Processing*, 178: 107805.
- Zeng, H.; Xie, X.; Cui, H.; Yin, H.; and Ning, J. 2020. Hyperspectral image restoration via global L 1-2 spatial-spectral total variation regularized local low-rank tensor recovery. *IEEE Transactions on Geoscience and Remote Sensing*, 59(4): 3309–3325.
- Zhang, Z.; and Aeron, S. 2016. Exact tensor completion using t-SVD. *IEEE Transactions on Signal Processing*, 65(6): 1511–1526.
- Zhang, Z.; Ely, G.; Aeron, S.; Hao, N.; and Kilmer, M. 2014. Novel methods for multilinear data completion and de-noising based on tensor-SVD. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 3842–3849.
- Zheng, S.; Liu, Y.; Meng, Z.; Qiao, M.; Tong, Z.; Yang, X.; Han, S.; and Yuan, X. 2021. Deep plug-and-play priors for spectral snapshot compressive imaging. *Photonics Research*.
- Zheng, Y.-B.; Huang, T.-Z.; Zhao, X.-L.; Jiang, T.-X.; Ma, T.-H.; and Ji, T.-Y. 2019. Mixed noise removal in hyperspectral image via low-fibered-rank regularization. *IEEE Transactions on Geoscience and Remote Sensing*, 58(1): 734–749.
- Zhou, Y.; and Cheung, Y.-M. 2019. Bayesian low-tubal-rank robust tensor factorization with multi-rank determination. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43(1): 62–76.