CoNoT: Coupled Nonlinear Transform-Based Low-Rank Tensor Representation for Multidimensional Image Completion

Jian-Li Wang, Ting-Zhu Huang, Xi-Le Zhao, Yi-Si Luo, and Tai-Xiang Jiang

***Abstract*— Recently, the transform-based tensor nuclear norm (TNN) methods have shown promising performance and drawn increasing attention in tensor completion (TC) problems. The main idea of these methods is to exploit the low-rank structure of frontal slices of the tensor under the transform. However, the transforms in TNN methods usually treat all modes equally and do not consider the different traits of different modes (i.e., spatial and spectral/temporal modes). To address this problem, we suggest a new low-rank tensor representation based on the coupled nonlinear transform (called CoNoT) for a better low-rank approximation. Concretely, spatial and spectral/temporal transforms in the CoNoT, respectively, exploit the different traits of different modes and are coupled together to boost the implicit low-rank structure. Here, we use the convolutional neural network (CNN) as the CoNoT, which can be learned solely from an observed multidimensional image in an unsupervised manner. Based on this low-rank tensor representation, we build a new multidimensional image completion model. Moreover, we also propose an enhanced version (called Ms-CoNoT) to further exploit the spatial multiscale nature of real-world data. Extensive experiments on real-world data substantiate the superiority of the proposed models against many state-of-the-art methods both qualitatively and quantitatively.**

***Index Terms*— Coupled nonlinear transform (CoNoT), low-rank tensor representation, tensor completion (TC), tensor nuclear norm (TNN).**

I. Introduction

BECAUSE of the flexible way of representing multidimensional images, tensor has drawn great interests in different fields, such as image denoising [1], video snapshot compressive imaging [2], hyperspectral image (HSI) restoration [3], [4], [5], and background model initialization [6], [7], [8], [9]. Unfortunately, the observed tensor is often incomplete due to the limitation of hardware, which severely hinders the subsequent tensor processing tasks, including but not limited to face recognition [10] and HSI classification [11]. Tensor completion (TC) aims to infer the missing information from its partial observation [12], [13], [14], which is a fundamental and valuable research topic. Generally speaking, without any assumptions, this is a classical ill-posed inverse problem. For solving it, the prior, which is also called regularization, needs to be adopted to constraint the solution space [15], [16], [17], [18], [19]. Good priors can yield the pleasing results.

Over the past several decades, the low-rank prior has proven to be an effective regularizer for constraining the solution space, based on the fact that multidimensional data admits many repetitive or redundant structures, i.e., it lives in a low-dimensional subspace. Thus, recovering a potential data from its partial observation usually can be modeled as the following low-rank TC (LRTC) problem:

$$\arg\min_{\mathcal{X}} \ \text{rank}(\mathcal{X}) \quad \text{s.t. } \mathcal{P}_{\Omega}(\mathcal{X}) = \mathcal{P}_{\Omega}(\mathcal{O})$$

where $\mathcal{X}$ is the underlying tensor, $\mathcal{O}$ is the observed tensor, Ω is the index set that corresponds to the observed entries, and $\mathcal{P}_{\Omega}(\cdot)$ is the projection operator that remains the elements of $\cdot$ in Ω, while other elements are filled with zeros. Unfortunately, the definition of tensor rank remains largely unexplored unlike the matrix case [20], [21]. So far, the research on tensor rank can be roughly divided into five mainstream works: CANDECOMP/PARAFAC rank [22], Tucker rank [23], [24], [25], [26], [27], tubal rank [28], [29], [30], tensor train rank [31], [32], and tensor ring rank [33], [34], [35], [36], [37], all of which provide a high-dimensional extension of the definition of matrix rank from a certain perspective.

Recently, the tubal rank-based methods, which are induced from the tensor singular value decomposition (t-SVD) [38], have shown promising performance and drawn increasing interests in TC problems [30]. The t-SVD was originally proposed by Braman [38], which is a direct third-order extension of the matrix singular value decomposition and decomposes a third-order tensor into two orthogonal tensors and one f-diagonal tensor based on a predefined tensor–tensor product (t-prod) [39]. The tubal rank of a tensor is defined as the number of nonzero tubes of the f-diagonal tensor in t-SVD manner. Since directly minimizing the tubal rank is a NP-hard problem, Zhang and Aeron [40] turned to minimize its convex

Manuscript received 7 October 2021; revised 18 June 2022; accepted 14 October 2022. This work was supported in part by the National Natural Science Foundation of China (NSFC) under Grant 12171072, Grant 61876203, and Grant 12001446; in part by the National Key Research and Development Program of China under Grant 2020YFA0714001; in part by the Key Project of Applied Basic Research in Sichuan Province under Grant 2020YJ0216; in part by the Applied Basic Research Project of Sichuan Province under Grant 2021YJ0107; in part by the Natural Science Foundation of Sichuan under Grant 2022NSFSC1798; and in part by the Fundamental Research Funds for the Central Universities under Grant JBK2202049 and Grant JBK2102001. *(Corresponding authors: Ting-Zhu Huang; Xi-Le Zhao.)*

Jian-Li Wang, Ting-Zhu Huang, and Xi-Le Zhao are with the Research Center for Image and Vision Computing, School of Mathematical Sciences, University of Electronic Science and Technology of China, Chengdu 611731, China (e-mail: wangjianli_123@163.com; tingzhuhuang@126.com; xlzhao122003@163.com).

Yi-Si Luo is with the School of Mathematics and Statistics, Xi'an Jiaotong University, Xi'an 710049, China (e-mail: yisiluo1221@foxmail.com).

Tai-Xiang Jiang is with the FinTech Innovation Center, School of Economic Information Engineering, Southwestern University of Finance and Economics, Chengdu 610074, China (e-mail: taixiangjiang@gmail.com).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TNNLS.2022.3217198.

Digital Object Identifier 10.1109/TNNLS.2022.3217198

relaxation, i.e., tensor nuclear norm (TNN), and built the corresponding TC model as follows:

$$\arg\min_{\mathcal{X}} \ \|\mathcal{X}\|_{\text{TNN}} \quad \text{s.t. } \mathcal{P}_\Omega(\mathcal{X}) = \mathcal{P}_\Omega(\mathcal{O})$$

where $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, $\|\mathcal{X}\|_{\text{TNN}} = \sum_{i=1}^{n_3} \|\bar{\mathcal{X}}^{(i)}\|_*$, $\bar{\mathcal{X}}^{(i)}$ is the ith frontal slice of $\bar{\mathcal{X}}$, $\bar{\mathcal{X}}$ is the result of performing a 1-D fast discrete Fourier transform (DFT) on $\mathcal{X}$ along its third mode (i.e., all the tubes of $\mathcal{X}$), which can be implemented by the MATLAB command fft, i.e., $\bar{\mathcal{X}} = \text{fft}(\mathcal{X}, [], 3)$, and $\|\cdot\|_*$ denotes the matrix nuclear norm (the sum of the singular values of the matrix). Due to the less effectiveness of the matrix norm in approximating the matrix rank, some nonconvex functions [41], [42] have been studied to better promote the low rankness, such as the logarithmic sum function [42].

It is worth noting that the main idea of TNN is to perform a 1-D fast DFT along the third mode of the third-order tensor and then minimize the nuclear norm of each frontal slice of the transformed tensor. In the literature, some researchers considered and studied other transforms to replace the DFT, for instance, the discrete cosine transform (DCT) [43], the unitary transform [44], [45], the invertible transform [39], [46], the framelet transform [47], the data-driven transform [48], [49], and the nonlinear transform [50]. These transform-based TNN methods for TC can be formulated as a unified model, i.e.,

$$\arg\min_{\mathcal{X}} \ \sum_{i=1}^{n_4} \|L(\mathcal{X})^{(i)}\|_* \quad \text{s.t. } \mathcal{P}_\Omega(\mathcal{X}) = \mathcal{P}_\Omega(\mathcal{O})$$

where $L(\mathcal{X})$ is obtained by applying the transform $L : \mathbb{R}^{1\times 1\times n_3} \to \mathbb{R}^{1\times 1\times n_4}$ to each tube of $\mathcal{X}$. If the transform is linear, $L(\mathcal{X})$ can be formulated as a mode-3 tensor-matrix product, i.e., $L(\mathcal{X}) = \mathcal{X} \times_3 \boldsymbol{L} = \texttt{Fold}_3(\boldsymbol{L}\texttt{Unfold}_3(\mathcal{X}))$, where $\boldsymbol{L} \in \mathbb{R}^{n_4 \times n_3}$ is the arbitrary matrix related to the linear transform L, and the definitions of Fold, Unfold are shown in Definition 1.

However, the above-mentioned methods usually consider the transform (correlation) of the one mode (i.e., spectral/temporal mode) and ignore the different traits of different modes (i.e., spatial and spectral/temporal modes). To address this problem, we suggest a new coupled nonlinear transform (CoNoT)-based low-rank tensor representation, in which spatial and spectral/temporal transforms in the CoNoT, respectively, exploit the different traits of different modes and are coupled together to boost the implicit low-rank structure, i.e., being not directly low rank in the original domain but low rank in the transformed domain with well-chosen transformation. Concretely, the proposed transform herein shares a similar topology structure with the popular convolutional neural network (CNN) with a nonlinear activation function, which can be learned solely from an observed multidimensional image in an unsupervised manner. The advantages of the proposed CoNoT are twofolds: 1) it admits a powerful expressive ability to exploit the implicit low-rank structure of spatial and spectral/temporal modes of the real-world data in a unified framework compared with previous transform-based TNN methods, which leads to a better low-rank approximation as by-product and 2) we consider the classical TNN framework, which enjoys an interpretable nature compared with deep learning (DL)-based methods [51], [52], [53], [54], [55].

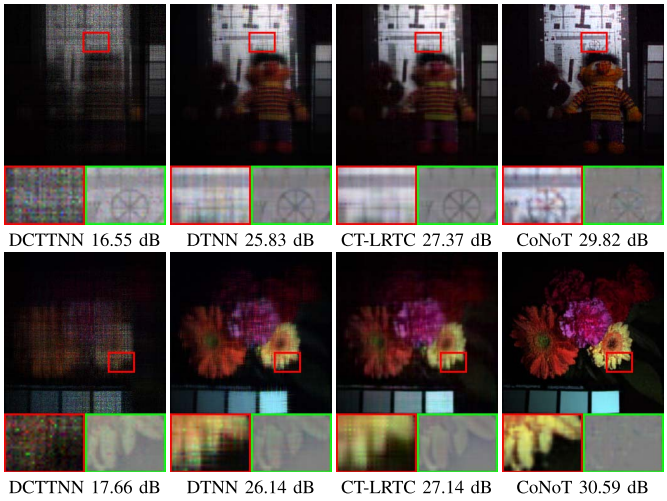


Fig. 1. Visualization of false-color restoration results (R:31, G:15, and B:5) by different methods on MSIs *Toy* (top) and *Flowers* (bottom) with SR = 2%.

Furthermore, we propose an enhanced version (called Ms-CoNoT) to fully exploit the spatial multiscale nature of real-world data, which enjoys a similar topology structure with the well-known U-Net [56], [57], [58].

Our contributions are summarized as follows.

1) We propose a CoNoT-based low-rank tensor representation, in which spatial and spectral/temporal transforms are coupled together to boost the implicit low-rank structure, leading to a better low-rank approximation as by-product. To fully exploit the spatial multiscale nature of real-world data, we further propose an enhanced version (called Ms-CoNoT).
2) We formulate two novel multidimensional image completion models based on the proposed CoNoT and Ms-CoNoT. To tackle them, we develop an effective gradient descent algorithm, which is specifically designed for optimizing deep neural networks.
3) Extensive experiments on real-world data substantiate the superiority of the proposed models against many state-of-the-art methods both qualitatively and quantitatively, especially in low sampling rates (SRs); see Fig. 1.

This article is organized as follows. Section II briefly reviews some related works. Section III gives the basic preliminaries about tensors. Section IV introduces the proposed CoNoT and Ms-CoNoT in detail for TC. Section V reports the experimental results and gives some discussions to demonstrate the effectiveness of the proposed models. Finally, this article is concluded in Section VI.

II. Related Work

In the literature, there were many transform-based TNN methods for LRTC. For instance, Madathil and George [43] used the real-valued DCT instead of DFT in TNN. The motivation is that the reflection boundary condition in DCT has advantages in preserving the head and tail frontal slices, and the real-valued operation replaces the complex operation in DFT, which greatly saves the computational cost. Furthermore, Song et al. [44] and Ng et al. [45] introduced the unitary transform in t-SVD to obtain a better low tubal-rank

approximation and proved that a tensor can be recovered accurately with overwhelming probability if its tubal rank is small enough, and its corrupted entries are fairly sparse. More generally, Lu et al. [39] and Kernfeld et al. [46] extend the so-called transform to any invertible linear transform in t-prod. Jiang et al. [47] broke the limitations of the invertibility and introduced the framelet transform. Some data-driven transforms were also proposed in [48] and [49]. Recently, Luo et al. [50] introduced the nonlinear transform into TNN, in which the nonlinear transform is implemented by a nonlinear multilayer neural network, which can be learned by using the observed tensor in a self-supervised manner. However, these methods only consider the transform of the one mode (i.e., spectral/temporal mode) and neglect the different traits of different modes. In contrast, we consider the transforms (correlations) of all modes (i.e., spatial and spectral modes) in a unified framework, which can exploit the different traits of different modes to boost the implicit low-rank structure.

Recently, the DL-based completion methods have been rapidly developed and can be divided into two groups, i.e., supervised DL-based methods and unsupervised DL-based methods. The supervised DL-based methods [52], [53] learn deep image priors (DIPs) from a large number of example images. Most of these methods are designed for particular tasks, e.g., inpainting (the tube-wise sampling) [52], [53], and their performance is essentially dependent upon the diversity and volume of training datasets. Therefore, the lack of generalization abilities hinders their direct application to the general TC problem for diverse samplings and data. To address this challenge, some unsupervised DL-based methods have been coming up like mushrooms these years. Ulyanov et al. [57] proposed an unsupervised image restoration framework, namely, DIP. The work in [57] showed that the proper neural network architectures, without any training samples, can "*encode*" the DIPs. Moreover, Sidorov and Yngve Hardeberg [59] extended the DIP idea to HSI processing (e.g., denoising, inpainting, and super-resolution). However, the DIP-based methods relatively lack interpretations.

In our work, we consider the classical TNN framework. The transform is a key module in the TNN to exploit the interactions of frontal slices. We employ the CoNoT to help obtain a better low-rank representation, which can boost the recovery performance. Meanwhile, this transform is learned by solely using the observed data in an unsupervised manner, which is suitable for diverse data and samplings in the general TC problem.

III. Preliminaries

In this section, we provide some required definitions following [39], [60], and [61]. Meanwhile, some basic notations used throughout this article are given in Table I.

Definition 1 (Mode-k Unfolding [61]): For an Nth-order tensor $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2, \ldots, \times n_N}$, its mode-$k$ unfolding $\boldsymbol{X}_{(k)}$ is an $n_k \times \prod_{i \neq k} n_i$ matrix, which satisfies that $\boldsymbol{X}_{(k)}(i_k, j)$ is mapped by $\mathcal{X}(i_1, i_2, \ldots, i_N)$, where $j = 1 + \sum_{s=1, s \neq k}^{N} (i_s - 1) J_s$ with $J_s = \sum_{m=1, m \neq k}^{s-1} n_m$. The corresponding operator and inverse operator are denoted as $\boldsymbol{X}_{(k)} = \texttt{Unfold}_k(\mathcal{X})$ and $\mathcal{X} = \texttt{Fold}_k(\boldsymbol{X}_{(k)})$, respectively.

Definition 2 (Mode-k Tensor-Matrix Product [61]): The mode-k tensor-matrix product of an Nth-order tensor $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2, \ldots, \times n_N}$ and a matrix $\boldsymbol{A} \in \mathbb{R}^{J \times n_k}$ is an $n_1 \times \cdots \times n_{k-1} \times J \times n_{k+1} \times \cdots \times n_N$ tensor, which is denoted by $\mathcal{X} \times_k \boldsymbol{A}$ and satisfied

$$(\mathcal{X} \times_k \boldsymbol{A})_{i_1, \ldots, i_{k-1}, j, i_{k+1}, \ldots, i_N} = \sum_{i_k=1}^{n_k} x_{i_1, i_2, \ldots, i_N} \cdot a_{j, i_k}.$$

Definition 3 (TNN [39]): Let $L : \mathbb{R}^{1 \times 1 \times n_3} \to \mathbb{R}^{1 \times 1 \times n_4}$ be any linear transform. Given $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, the TNN of $\mathcal{X}$ is defined as follows:

$$\|\mathcal{X}\|_{\text{TNN}} = \sum_{i=1}^{n_4} \|L(\mathcal{X})^{(i)}\|_*.$$

TABLE I
Notations

Notation	Description
x, $\boldsymbol{X}$, $\mathcal{X}$	Scalar, matrix, tensor.
$x_{i_1, i_2, \cdots, i_N}$ or $\mathcal{X}(i_1, i_2, \cdots, i_N)$	The $(i_1, i_2, \cdots, i_N)$-th element of $\mathcal{X}$.
$\mathcal{X}(i_1, i_2, :)$	The (i_1, i_2)-th tube of a third-order tensor $\mathcal{X}$.
$\mathcal{X}^{(i)}$ or $\boldsymbol{X}^i$	The i-th frontal slice of a third-order tensor $\mathcal{X}$.
$\|\mathcal{X}\|_F$	The Frobenius norm of $\mathcal{X}$, defined as $\|\mathcal{X}\|_F = \sqrt{\sum_{i_1, i_2, \cdots, i_N} x_{i_1, i_2, \cdots, i_N}^2}$.
$\texttt{Fold}_k(\boldsymbol{X})$	The mode-k folding of $\boldsymbol{X}$.
$\boldsymbol{X}_{(k)}$ or $\texttt{Unfold}_k(\mathcal{X})$	The mode-k unfolding of $\mathcal{X}$.
$\times_k$	The mode-k tensor-matrix product.
$\circledast$	The 2D spatial convolution operation.

IV. Proposed CoNoT for LRTC

In this section, we start by introducing the architecture of the CoNoT in Section IV-A. Based on CoNoT, we then present a novel multidimensional image completion model and build the corresponding algorithm to solve it in Sections IV-B and IV-C. Section IV-D gives an illustration to verify that the proposed CoNoT can help to obtain a better low-rank approximation. Finally, we further propose an enhanced version to fully exploit the spatial multiscale nature of real-world data in Section IV-E.

A. Architecture of the CoNoT

Classical methods induced by TNN rely on a transform along the third mode, e.g., DFT [38], DCT [43], and the data-driven transform [49]. However, these transforms usually are single mode and neglect the different traits of different modes. To address this problem, we propose a CoNoT-based low-rank tensor representation, in which spatial and spectral/temporal transforms in the CoNoT, respectively, exploit the different traits of different modes and are coupled together to boost the implicit low-rank structure. Here, we use the CNN with nonlinear activate function as the CoNoT (called $\mathcal{F}$), which can be learned solely from an observed multidimensional image in an unsupervised manner.

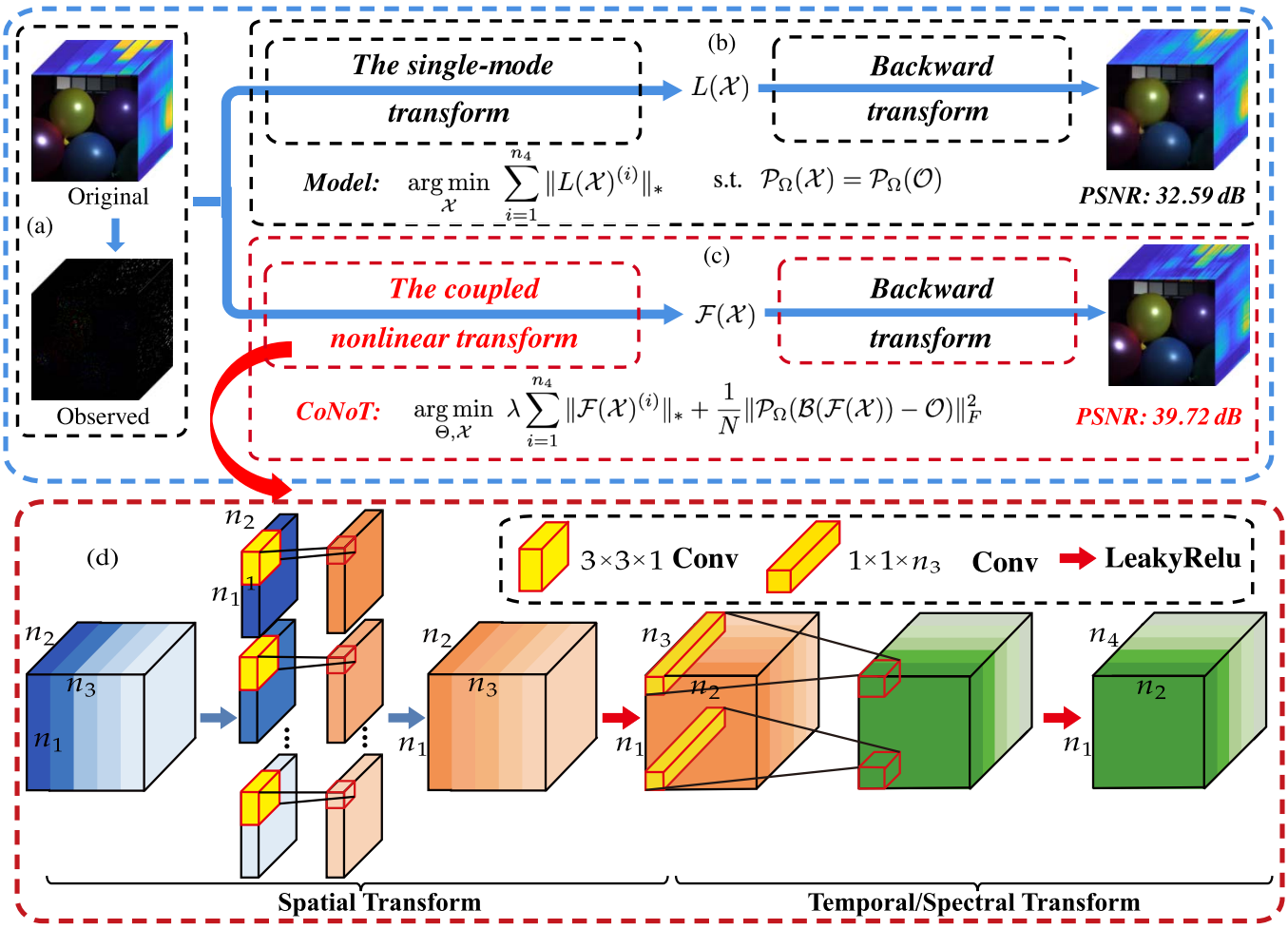


Fig. 2. Pipeline of the proposed CoNoT for multidimensional image completion. (a) Formation process of the observed tensor (here, the SR is 5%). (b) Demo of the single-mode transform-based TNN methods for TC. (c) Proposed CoNoT. (d) Convolutional implementation of the CoNoT.

The overall architecture of the CoNoT is presented in Fig. 2(d), which is mainly composed of two transforms, i.e., *spatial transform* (2-D spatial convolution layer + nonlinear transform layer) and *temporal/spectral transform* (1-D temporal/spectral convolution layer + nonlinear transform layer). Next, we will introduce them in detail.

1) Spatial Transform (2-D Spatial Convolution Layer + Nonlinear Transform Layer): To capture spatial correlation of the data, we first introduce the 2-D spatial convolution layer. Taking $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ for an example, the 2-D spatial convolution layer is implemented by convolving each of its frontal slice $\boldsymbol{X}^i$ with different filters of size $3 \times 3 \times 1$ (the number is n_3) to obtain a feature map of size $n_1 \times n_2 \times n_3$, which can capture the local spatial correlation of the data. In addition, we also perform a nonlinear mapping on the transformed data to exploit the nonlinear nature of real data. Mathematically, these two layers can be formulated as follows:

$$\mathcal{F}_1(\mathcal{X}) = \sigma\left\{\boldsymbol{W}_1^1 \circledast \boldsymbol{X}^1, \boldsymbol{W}_1^2 \circledast \boldsymbol{X}^2, \ldots, \boldsymbol{W}_1^{n_3} \circledast \boldsymbol{X}^{n_3}\right\} \quad (1)$$

where $\sigma(\cdot)$: $\mathbb{R} \to \mathbb{R}$ is the nonlinear activation function corresponds to the nonlinear transform layer, which handles each pixel individually, $\mathcal{W}_1$ denotes the set of learnable parameters in the 2-D spatial convolution layer and its size is $3 \times 3 \times n_3$, and $\circledast$ denotes the 2-D spatial convolution operation (zero padding $= 1$ and stride $= 1$) between the matrix and a 3×3 filter. More specifically, the (i, j, k)th element of $\mathcal{F}_1(\mathcal{X})$ is denoted as follows:

$$\sigma\left(\sum_{q=j-1}^{j+1}\sum_{p=i-1}^{i+1} \boldsymbol{W}_1^k(p-i+2, q-j+2)\boldsymbol{X}^k(p,q)\right).$$

In our implementation, we choose the leaky rectified linear unit (LeakyReLU) as nonlinear activation function, and the output of these two layers is $\mathcal{Y}$ of size $n_1 \times n_2 \times n_3$. Next, $\mathcal{Y}$ is treated as the input of the latter layer.

2) Temporal/Spectral Transform (1-D Temporal/Spectral Convolution Layer + Nonlinear Transform Layer): To capture interactions along the temporal/spectral mode, the 1-D temporal/spectral convolution layer is implemented by convolving each tube of $\mathcal{Y}$ with different filters of size $1 \times 1 \times n_3$ (the number is n_4) to obtain a feature map of size $n_1 \times n_2 \times n_4$, which can capture the global temporal/spectral correlation of the data. Similar to the previous one, we also add a nonlinear transform layer after the above convolution layer. Mathematically, this process can be written in the form of a tensor-matrix product, i.e.,

$$\mathcal{F}_2(\mathcal{Y}) = \sigma(\mathcal{Y} \times_3 \boldsymbol{W}_2) = \sigma(\texttt{Fold}_3(\boldsymbol{W}_2\texttt{Unfold}_3(\mathcal{Y}))) \quad (2)$$

where $\boldsymbol{W}_2$ denotes the set of learnable parameters in the temporal/spectral convolution layer, and its size is $n_4 \times n_3$

(n_4 is the number of filters of the temporal/spectral layer), and the output of these two layers is $\mathcal{Z}$ of size $n_1 \times n_2 \times n_4$. Here, a larger n_4 usually can bring redundancy of the transform to obtain a better low-rank representation [47].

By combining (1) and (2) cleverly, the CoNoT is expressed as follows:

$$\mathcal{F}(\mathcal{X}) = \mathcal{F}_2(\mathcal{F}_1(\mathcal{X})). \tag{3}$$

Similar to TNN family methods, we view $\mathcal{F}$ as the forward transform and introduce the corresponding backward transform, denoted by

$$\mathcal{B}(\mathcal{X}) = \mathcal{B}_2(\mathcal{B}_1(\mathcal{X})) \tag{4}$$

which has a similar structure with $\mathcal{F}$. For simplicity, we use $\Theta = \{\mathcal{W}_1, \boldsymbol{W}_2, \mathcal{W}_3, \boldsymbol{W}_4\}$ to denote the learnable parameters of $\mathcal{F}$ and $\mathcal{B}$. Here, $\mathcal{W}_1$ and $\boldsymbol{W}_2$ are learnable parameters of the convolution layers in $\mathcal{F}$; $\mathcal{W}_3$ and $\boldsymbol{W}_4$ are learnable parameters of the convolution layers in $\mathcal{B}$.

B. CoNoT for LRTC

We suggest a new low-rank tensor representation based on the above transform $\mathcal{F}$ and the corresponding backward transform $\mathcal{B}$ (called CoNoT) and formulate a novel multidimensional visual data completion model, in which the CoNoT can be implemented by CNN with nonlinear activation function, i.e.,

$$\arg\min_{\Theta,\mathcal{X}} \ \lambda \sum_{i=1}^{n_4} \|\mathcal{F}(\mathcal{X})^{(i)}\|_* + \frac{1}{N}\|\mathcal{P}_\Omega(\mathcal{B}(\mathcal{F}(\mathcal{X})) - \mathcal{O})\|_F^2 \tag{5}$$

where λ controls the strength of low rankness, $N = n_1 n_2 n_3$, $\mathcal{O} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ is the observed tensor, and $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ is the estimated tensor initialized by a single linear interpolation strategy [62], which is optimized along with the transform parameters Θ. $\mathcal{F} : \mathbb{R}^{n_1 \times n_2 \times n_3} \to \mathbb{R}^{n_1 \times n_2 \times n_4}$ and $\mathcal{B} : \mathbb{R}^{n_1 \times n_2 \times n_4} \to \mathbb{R}^{n_1 \times n_2 \times n_3}$ are the CoNoT and the corresponding backward transform defined by (3)–(4). $\Theta = \{\mathcal{W}_1, \boldsymbol{W}_2, \mathcal{W}_3, \boldsymbol{W}_4\}$ is the learnable parameters of $\mathcal{F}$ and $\mathcal{B}$. Our model only utilizes the observed data $\mathcal{O}$ without additional training data. Thus, the parameters of the CoNoTs $\mathcal{F}$ and $\mathcal{B}$ are inferred in an unsupervised manner.

After obtaining the optimal $\mathcal{F}$ and $\mathcal{B}$ by solving problem (5), it is expected to obtain the desired tensor via the learned transforms $\mathcal{F}$ and $\mathcal{B}$. Since we need to ensure that the elements in Ω are equal to the observed tensor, the underlying tensor is given by

$$\mathcal{P}_\Omega(\mathcal{O}) + \mathcal{P}_{\Omega^C}(\mathcal{B}(\mathcal{F}(\mathcal{X})))$$

where Ω^C is the complementary set of Ω.

Remark: We discuss the connection between the proposed work and the previous TNN-based methods. The transform-based TNN family methods can be roughly grouped into two categories: single-mode transform-based and all-mode transform-based methods. It is worth mentioning that most of the previous methods are single-mode transform-based methods, which only consider the transform of the one mode (i.e., spectral/temporal mode) of the data, such as DFT [38], DCT [43], the unitary transform [44], [45], the invertible linear transform [39], [46], the noninvertible linear transform [47], [48], [49], and the nonlinear transform [50]. Only [63] considers the transforms (correlations) of all modes (i.e., spatial and spectral modes) in a unified framework, in which the framelet transform is used to characterize the multiscale nature of data in spatial domain. On the contrary, the proposed CoNoT is not only an all-mode transform-based method, which can characterize the correlation of different modes in a unified framework, but also adaptively learns the corresponding transform from the data itself in an unsupervised manner, so that it can be more flexibly applied to various types of data.

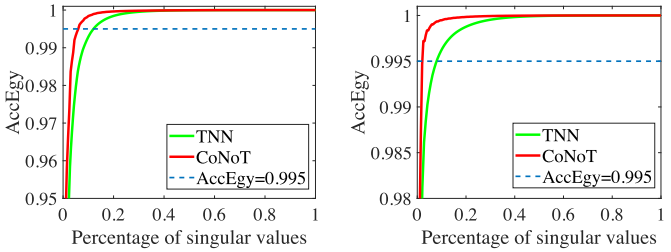


Fig. 3. AccEgy with respect to the percentage of singular values of the transformed frontal slices by using different transforms on MSI *Feathers* (left) and video *Foreman* (right).

C. Gradient Descent-Based Algorithm

The highly nonlinear structures of our optimization problem (5) prevent us from using general optimization algorithms designed for traditional models, e.g., half quadratic splitting (HQS) [64] and alternating direction method of multipliers (ADMM) [63]; we develop the gradient descent optimization to solve (5). Note that our algorithm is easy to implement and simple, which is specifically designed for optimizing deep neural networks.

For convenience, we rewrite the loss function as follows:

$$\mathcal{L} = \lambda \mathcal{L}_1 + \frac{1}{N}\mathcal{L}_2$$

where $\mathcal{L}_1 = \sum_i \|\mathcal{F}(\mathcal{X})^{(i)}\|_*$ is the regularization term, and $\mathcal{L}_2 = \|\mathcal{P}_\Omega(\mathcal{B}(\mathcal{F}(\mathcal{X})) - \mathcal{O})\|_F^2$ is the fidelity term.

We first introduce the subgradient of the matrix nuclear norm [65] used for the gradient descent-based algorithm.

Definition 4 (Subgradient of the Matrix Nuclear Norm [65]): For a matrix $\boldsymbol{X}$, the subgradient of its nuclear norm is

$$\frac{\partial \|\boldsymbol{X}\|_*}{\partial \boldsymbol{X}} = \left\{\hat{\boldsymbol{U}}_r \hat{\boldsymbol{V}}_r^\top + \boldsymbol{W} | \boldsymbol{U}^T \boldsymbol{W} = 0, \boldsymbol{W}\boldsymbol{V} = 0, \|\boldsymbol{W}\|_{\ell_1} \le 1\right\}$$

where $\boldsymbol{X} = \boldsymbol{U}\boldsymbol{S}\boldsymbol{V}^\top$ is the matrix singular value decomposition, $\hat{\boldsymbol{U}}_r, \hat{\boldsymbol{V}}_r$ are the first r columns of $\boldsymbol{U}, \boldsymbol{V}$, and r is the number of nonzero elements in $\boldsymbol{S}$.

In our algorithm, we use

$$\hat{\boldsymbol{U}}_r \hat{\boldsymbol{V}}_r^\top \in \frac{\partial \|\boldsymbol{X}\|_*}{\partial \boldsymbol{X}} \tag{6}$$

as the subgradient of the nuclear norm of the matrix $\boldsymbol{X}$ by setting $\boldsymbol{W}$ be a zero matrix. Now, we begin to establish our algorithm.

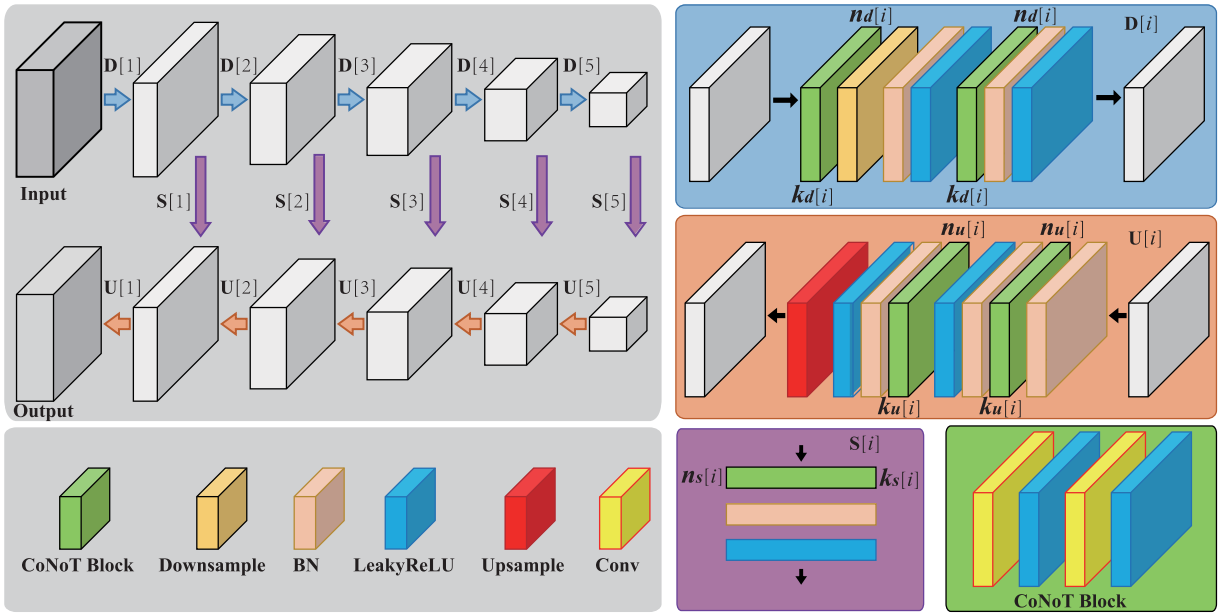


Fig. 4. Network architecture of the proposed Ms-CoNoT used in this article.

First, the gradient of $\mathcal{L}_1$ on the (u, v, w)th entry of $\mathcal{W}_1$ is expressed as follows:

$$\frac{\partial \mathcal{L}_1}{\partial (\mathcal{W}_1)_{u,v,w}} = \sum_i \frac{\partial \|\mathcal{L}(\mathcal{X})^{(i)}\|_*}{\partial (\mathcal{W}_1)_{u,v,w}} = \sum_i \sum_{s,t} \frac{\partial \|\mathcal{L}(\mathcal{X})^{(i)}\|_*}{\partial \left((\mathcal{L}(\mathcal{X}))^{(i)}\right)_{st}} \frac{\partial \left((\mathcal{L}(\mathcal{X}))^{(i)}\right)_{st}}{\partial (\mathcal{W}_1)_{u,v,w}}. \quad (7)$$

By combining (6) and (7) cleverly, the gradient of $\mathcal{L}_1$ on the (u, v, w)th entry of $\mathcal{W}_1$ is rewritten as follows:

$$\sum_i \sum_{s,t} \left(\hat{U}_r \hat{V}_r^\top\right)_{st} \frac{\partial \left((\mathcal{L}(\mathcal{X}))^{(i)}\right)_{st}}{\partial (\mathcal{W}_1)_{u,v,w}} \in \frac{\partial \mathcal{L}_1}{\partial (\mathcal{W}_1)_{u,v,w}}. \quad (8)$$

Similarly, the gradients of $\mathcal{L}_1$ on the (u, v)th entry of $\boldsymbol{W}_2$ and (u, v, w)th entry of $\mathcal{X}$ are

$$\begin{cases} \sum_i \sum_{s,t} \left(\hat{U}_r \hat{V}_r^\top\right)_{st} \frac{\partial \left((\mathcal{L}(\mathcal{X}))^{(i)}\right)_{st}}{\partial (\boldsymbol{W}_2)_{u,v}} \in \frac{\partial \mathcal{L}_1}{\partial (\boldsymbol{W}_2)_{u,v}} \\ \sum_i \sum_{s,t} \left(\hat{U}_r \hat{V}_r^\top\right)_{st} \frac{\partial \left((\mathcal{L}(\mathcal{X}))^{(i)}\right)_{st}}{\partial (\mathcal{X})_{u,v,w}} \in \frac{\partial \mathcal{L}_1}{\partial (\mathcal{X})_{u,v,w}}. \end{cases} \quad (9)$$

Second, the gradients of $\mathcal{L}_2$ on $\mathcal{W}_3, \boldsymbol{W}_4$, and $\mathcal{X}$ are as follows:

$$\begin{cases} \frac{\partial \mathcal{L}_2}{\partial (\mathcal{W}_3)_{u,v,w}} = 2 \sum_{l,s,t} \frac{\partial \mathcal{L}_2}{\partial \left(\mathcal{B}(\mathcal{F}(\mathcal{X}))\right)_{l,s,t}} \frac{\partial (\mathcal{B}(\mathcal{F}(\mathcal{X})))_{l,s,t}}{\partial (\mathcal{W}_3)_{u,v,w}} \\ \frac{\partial \mathcal{L}_2}{\partial (\boldsymbol{W}_4)_{u,v}} = 2 \sum_{l,s,t} \frac{\partial \mathcal{L}_2}{\partial \left(\mathcal{B}(\mathcal{F}(\mathcal{X}))\right)_{l,s,t}} \frac{\partial (\mathcal{B}(\mathcal{F}(\mathcal{X})))_{l,s,t}}{\partial (\boldsymbol{W}_4)_{u,v}} \\ \frac{\partial \mathcal{L}_2}{\partial (\mathcal{X})_{u,v,w}} = 2 \sum_{l,s,t} \frac{\partial \mathcal{L}_2}{\partial \left(\mathcal{B}(\mathcal{F}(\mathcal{X}))\right)_{l,s,t}} \frac{\partial (\mathcal{B}(\mathcal{F}(\mathcal{X})))_{l,s,t}}{\partial (\mathcal{X})_{u,v,w}}. \end{cases} \quad (10)$$

By combining (8)–(10), model (5) can be easily solved by the gradient descent-based algorithm. In this work, we adopt the adaptive moment estimation (ADAM) algorithm. In summary, the computational complexity of using the ADAM optimizer to solve the proposed CoNoT is $K^2 n_1 n_2 n_3 + K^2 n_1 n_2 n_4 + 2 n_1 n_2 n_3 n_4$ on a tensor of size $n_1 \times n_2 \times n_3$, where K is the size of the convolutional kernel. Since model (5) is nonconvex, the initialization of Θ and $\mathcal{X}$ is important. Here, we use the default normal distribution in PyTorch[1] to initialize the transform parameters and generate the initialization tensor $\mathcal{X}_0$ by a single linear interpolation strategy, which can be found in [62].

D. Enhanced Low-Rank Representation

The proposed CoNoT provides a better tensor low tubal-rank approximation as by-product compared with other transform-based TNN methods. We define the accumulation energy ratio (AccEgy) of top k singular values as $\sum_{i=1}^{k} \sigma_i^2 / \sum_j \sigma_j^2$, where σ_i is the ith singular value. From Fig. 3, we can observe that, after the CoNoT, the energy of singular values of the transformed tensor is more concentrated compared with TNN, which means it can get a more compact representation. More specifically, as indicated by the auxiliary dashed lines, the proportion of singular values required for the data guided by the CoNoT is smaller than that of TNN when the data can occupy 99.5% of the entire energy, i.e., the data have a better tensor low tubal-rank approximation and more flexible expression ability, which is also a major contribution and motivation of the proposed CoNoT.

E. Enhanced Version: Ms-CoNoT

Motivated by the fact that the real-world data have spatial multiscale nature, we propose an enhanced version with a U-Net structure (called Ms-CoNoT), which is derived from CoNoT, to further improve the capacity. As shown in Fig. 4, we plug-and-play the CoNoT module in a U-Net architecture to form our Ms-CoNoT.

The detailed network structure for Ms-CoNoT is shown in Fig. 4. Specifically, the proposed Ms-CoNoT network includes

[1]https://pytorch.org/docs/stable/nn.init.html

TABLE II
HYPERPARAMETER VALUES OF THE PROPOSED MS-CONOT

Hyperparameter	Value
$n_d = n_u$	$[128, 128, 128, 128, 128]$
$k_d = k_u$	$[3, 3, 3, 3, 3]$
n_s	$[4, 4, 4, 4, 4]$
k_s	$[1, 1, 1, 1, 1]$
Optimizer	ADAM
Number of iterations	6000
Learning rate	0.01
Weight decay	0
Momentum	0.9
LeakyReLU slope	0.2

five downsampling blocks $D[i]$, five upsampling blocks $U[i]$, and five skip-connection blocks $S[i]$ $(i = 1, 2, 3, 4, 5)$. Here, $n_d[i]$, $n_u[i]$, and $n_s[i]$ $(i = 1, 2, 3, 4, 5)$ denote the number of filters in the convolutional layer for the downsampling block $D[i]$, upsampling block $U[i]$, and the skip-connection block $S[i]$, respectively; $k_d[i]$, $k_u[i]$, and $k_s[i]$ $(i = 1, 2, 3, 4, 5)$ denote the corresponding sizes of the convolutional kernels. Moreover, we use max pooling as the downsampling operator and nearest upsampling as the upsampling operator. All hyperparameter values of the proposed Ms-CoNoT are provided in Table II for reproducibility.

V. EXPERIMENTAL RESULTS

We evaluate the effectiveness of the proposed CoNoT and Ms-CoNoT on diverse datasets and compare them with four state-of-the-art TC methods. First, some experimental settings are given in Section V-A. Then, we evaluate the performance of the proposed methods on diverse datasets, including multispectral images (MSIs), video dataset, and multitemporal remote sensing images in Sections V-B–V-D. Finally, Section V-E presents some necessary discussions about the proposed methods.

A. Experimental Settings

1) Preprocessing: We use three classical datasets to test the performance of the proposed methods, including Columbia multispectral database (CAVE)[2] dataset, video dataset,[3], and multitemporal remote sensing images *Morocco*.[4] For MSIs and video dataset, we test all methods for different SRs: 5%, 10%, 15%, 20%, and 25%. For multitemporal remote sensing images, we test all methods for a classical cloud/shadow removal application. Before the experiment, the pixel values of all datasets are normalized to [0, 1] band-by-band. All experiments are carried on the platform of Windows 10 with one Intel(R) Core(TM) i9-9900K CPU and NVIDIA RTX 2080Ti GPU with 32-GB RAM. Our methods are implemented on PyTorch 1.6.0 with CPU and GPU calculation. All the compared methods are implemented on MATLAB R2018b with CPU calculation.

[2]http://www.cs.columbia.edu/CAVE/databases/multispectral
[3]http://trace.eas.asu.edu/yuv/
[4]https://theia.cnes.fr/atdistrib/rocket/home

2) Baselines: To validate the performance of the proposed CoNoT and Ms-CoNoT, we compare them with four recent and state-of-the-art TC methods, including TRLRF [66], DCT-TNN [39], DTNN [49], and CT-LRTC [63]. The first one is a tensor network-based method, the second is a predefined transform-based TNN method, the third is a data-driven transform-based TNN method, and the fourth is a coupled linear transform-based TNN method. The codes of comparison methods are provided by the authors or downloaded from their homepages.

3) Parameter Setting: There is only one regularization parameter λ in (5) that needs to be manually set, which controls the strength of low rankness. We empirically select it from the candidate set $\{0.2 \times 10^{-p}\}$ $(p = 4, 5, 6, 7, 8)$ to ensure that our methods have relatively good performance in different data. For CoNoT and Ms-CoNoT, the learning rates are set to 0.001 and 0.01, respectively. We set $n_4 = 4n_3$ in all our experiments. To achieve the optimal performance of the comparison methods, all relevant parameters are manually adjusted by default or follow the rules in the authors' suggestions.

4) Evaluation Indices: For the MSIs dataset, we consider three numerical metrics to evaluate the reconstructed results of all methods, including the peak signal-to-noise ratio (PSNR) [67], the structural similarity (SSIM) [67], and the spectral angle mapper (SAM) [68]. The higher the PSNR value, the higher the SSIM value, and the lower the SAM value indicate better performance. For the video dataset, PSNR, SSIM, and the universal image quality index (UIQI) [69] are used to evaluate the performance of the different methods. Similar to PSNR and SSIM, a higher UIQI value indicates better results.

B. MSIs Dataset

In this section, we compare the proposed methods with other TC methods on the CAVE dataset to verify their effectiveness. The CAVE dataset contains 32 MSIs data and the original data of size $512 \times 512 \times 31$ with very complex structure and texture information, and we resize them to $256 \times 256 \times 31$ in our experiments.

Fig. 5 shows the PSNR values of the recovered results by different methods on the CAVE dataset with SR = 10%. We can see that the proposed methods (red columns) achieve the highest and second-highest values in all data, which validates the superior performance of the proposed methods. In what follows, we only show the numerical and visual results for four different scenarios (*Watercolors*, *Toy*, *Peppers*, and *Feathers*) in the CAVE dataset due to page limitations.

Table III reports the quantitative assessment results of different methods for the CAVE dataset under different SRs. The best and the second-best results for each quality index are highlighted by boldface and underline, respectively. In general, Ms-CoNoT and CoNoT produce the best and the second-best qualitative results at all SRs, respectively, which illustrates the advantage of the proposed methods in MSIs completion. TRLRF and CT-LRTC perform well when SR is low, while the DTNN performs well as SRs arise. The quantitative evaluation results of DTNN are better than DCTTNN, because the transform in DTNN is learned from data that has better and more flexible expression ability than predefined transforms.

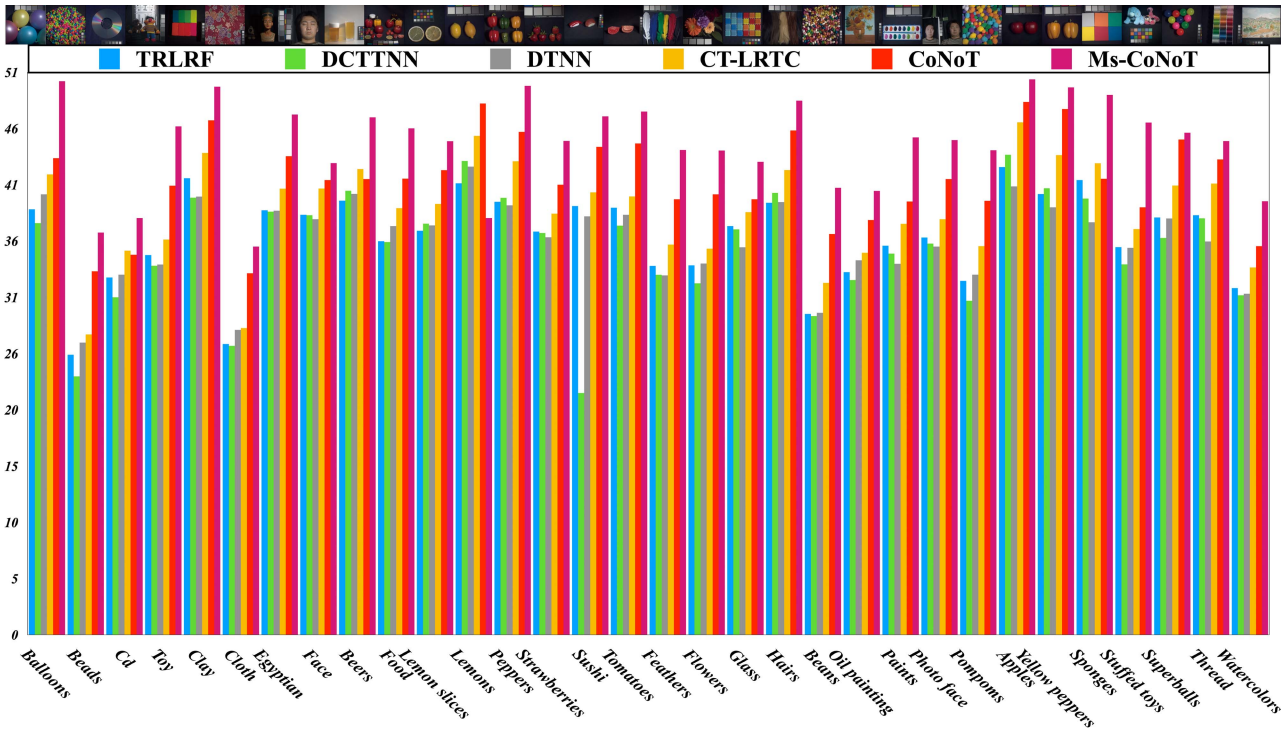


Fig. 5. PSNR values of the recovered results by different methods on the CAVE dataset with SR = 10%.

TABLE III

QUANTITATIVE ASSESSMENT RESULTS OF DIFFERENT METHODS FOR THE CAVE DATASET UNDER DIFFERENT SRS. THE **BEST** AND THE SECOND-BEST RESULTS FOR EACH QUALITY INDEX ARE HIGHLIGHTED BY **BOLDFACE** AND UNDERLINE, RESPECTIVELY

Dataset	SR	5%			10%			15%			20%			25%		
	Index	PSNR	SSIM	SAM	PSNR	SSIM	SAM	PSNR	SSIM	SAM	PSNR	SSIM	SAM	PSNR	SSIM	SAM
Watercolors	Observed	6.476	0.046	76.70	6.711	0.063	72.45	6.965	0.080	68.17	7.221	0.096	64.36	7.505	0.112	60.77
	TRLRF [66]	26.623	0.798	6.446	31.447	0.916	4.034	32.981	0.935	3.723	33.771	0.945	3.537	34.507	0.951	3.328
	DCTTNN [39]	26.002	0.785	8.389	30.794	0.914	4.756	34.318	0.957	3.295	37.133	0.976	2.509	39.460	0.985	2.024
	DTNN [49]	26.434	0.828	9.255	30.937	0.926	4.458	35.269	0.967	2.856	38.840	0.983	2.123	41.856	0.990	1.661
	CT-LRTC [63]	29.096	0.896	4.483	33.301	0.956	3.042	36.127	0.975	2.386	38.437	0.984	2.048	40.273	0.988	1.774
	CoNoT	31.488	0.919	4.326	35.239	0.961	3.322	38.862	0.980	2.573	41.509	0.988	**1.998**	45.061	0.995	**1.445**
	Ms-CoNoT	**33.873**	**0.961**	**2.427**	**39.307**	**0.985**	**2.421**	**43.237**	**0.994**	**1.641**	**45.129**	**0.995**	2.087	**46.360**	**0.996**	1.992
Toy	Observed	10.633	0.242	77.16	10.865	0.279	72.82	11.106	0.313	68.65	11.373	0.347	64.65	11.653	0.380	60.95
	TRLRF [66]	30.235	0.833	15.10	34.430	0.919	11.53	36.393	0.945	9.973	37.370	0.954	9.306	38.100	0.961	8.678
	DCTTNN [39]	29.273	0.860	16.94	33.456	0.934	10.97	36.584	0.963	8.037	39.062	0.978	4.124	41.479	0.986	4.998
	DTNN [49]	29.015	0.898	13.02	33.562	0.943	8.224	38.148	0.976	5.554	41.771	0.988	4.124	45.441	0.994	3.099
	CT-LRTC [63]	31.936	0.927	10.17	35.829	0.966	6.840	38.569	0.979	5.461	40.636	0.986	4.453	42.666	0.990	3.784
	CoNoT	34.653	0.937	12.53	40.716	0.985	5.334	43.504	0.990	4.688	46.359	0.993	3.795	48.087	0.996	2.767
	Ms-CoNoT	**36.994**	**0.968**	**7.611**	**45.552**	**0.996**	**2.725**	**46.097**	**0.997**	**2.401**	**50.434**	**0.997**	**2.402**	**52.276**	**0.998**	**2.118**
Peppers	Observed	12.937	0.176	79.04	13.168	0.216	74.78	13.419	0.254	70.38	13.690	0.290	66.31	13.953	0.323	62.51
	TRLRF [66]	34.829	0.920	5.341	39.958	0.964	3.600	42.309	0.977	2.904	43.911	0.983	2.548	44.903	0.986	2.355
	DCTTNN [39]	34.694	0.911	6.914	40.487	0.972	3.316	44.396	0.988	2.152	47.349	0.993	1.589	49.570	0.996	1.263
	DTNN [49]	33.327	0.887	17.85	38.754	0.945	11.89	43.450	0.987	2.096	47.676	0.994	1.426	50.714	0.997	1.081
	CT-LRTC [63]	38.786	0.973	3.422	43.493	0.990	2.050	46.691	0.995	1.503	49.086	0.996	1.230	50.746	0.997	1.051
	CoNoT	39.946	0.979	2.789	45.584	0.992	1.748	47.673	0.994	1.512	51.239	0.997	1.116	51.764	**0.998**	1.010
	Ms-CoNoT	**45.465**	**0.994**	**1.269**	**49.632**	**0.997**	**1.082**	**51.028**	**0.997**	**1.134**	**52.051**	**0.998**	**0.972**	**53.021**	**0.998**	**0.903**
Feathers	Observed	13.353	0.182	78.32	13.589	0.2216	73.90	13.589	0.222	73.90	14.106	0.295	65.52	14.379	0.330	61.69
	TRLRF [66]	29.473	0.812	11.63	33.444	0.888	8.889	35.271	0.917	7.697	36.200	0.930	7.144	36.782	0.937	6.793
	DCTTNN [39]	28.172	0.793	15.23	32.639	0.899	9.589	35.795	0.943	7.001	38.309	0.965	5.336	40.509	0.978	4.230
	DTNN [49]	28.460	0.853	17.37	32.576	0.923	7.649	36.783	0.965	5.018	40.343	0.982	3.462	43.325	0.990	2.622
	CT-LRTC [63]	31.670	0.909	8.786	35.375	0.954	5.976	38.085	0.972	4.671	40.081	0.981	3.850	41.829	0.986	3.263
	CoNoT	34.338	0.944	7.395	39.489	0.976	4.161	42.522	0.987	3.197	44.922	0.991	2.554	46.782	0.993	2.279
	Ms-CoNoT	**36.874**	**0.983**	**2.811**	**43.954**	**0.993**	**1.860**	**46.532**	**0.996**	**1.684**	**49.047**	**0.997**	**1.526**	**49.790**	**0.997**	**1.499**

Compared with the DTNN, in the proposed CoNoT, spatial and spectral/temporal transforms in the CoNoT, respectively, exploit the different traits of different modes and are coupled together to boost the implicit low-rank structure. As a

Fig. 6. Visualization of false-color restoration results (R:31, G:15, and B:5) by different methods on the CAVE dataset with SR = 5%. From top to bottom: *Watercolors*, *Toy*, *Peppers*, and *Feathers*. For better visualization, two demarcated areas (red and green rectangles) are magnified under each image.

result, the improvement of CoNoT compared with DTNN demonstrates the contribution of the CoNoT, which is a major contribution of the proposed CoNoT. For all SRs, the proposed Ms-CoNoT obtains the best quantitative evaluation results. Specifically, Ms-CoNoT is 2–3 dB higher than the second best (the proposed CoNoT) and 5–6 dB higher than the third best (CT-LRTC) in PSNR values on average, respectively. We attribute this to the fact that the U-Net structure can utilize the multiscale information of the data to better express the data.

Fig. 6 shows the false-color restoration results by different methods on the CAVE dataset with SR = 5%. For better visualization, two local areas are chosen and enlarged under each image. From Fig. 6, we can see that the result of DCTTNN contains some spatial blurring effects, and the color looks unnatural on *Toy* clothes and *Watercolor* trees. The DTNN and TRLRF have a better performance than DCTTNN visually; however, some blurring effects still exist, especially in the enlarged areas. The results of CT-LRTC have many spatial edges or textures in CAVE dataset compared with DTNN and TRLRF, but spectral differences can be found. On the contrary, CoNoT behaves well in spatial details, and the colors look more natural compared with the comparison methods. In addition, the spatial details and textures of Ms-CoNoT are clearer. The visual results again validate the superior performance of the proposed methods.

C. Video Dataset

In this section, we select three different video data to further verify the effectiveness of the proposed methods, i.e., *Akiyo*, *Carphone*, and *Foreman*. All data are of size $144 \times 176 \times 100$.

Quantitative assessment results of different methods for video dataset under different SRs are given in Table IV. We can observe that the proposed Ms-CoNoT and CoNoT achieve the best and the second-best results in most cases, respectively. Fig. 7 shows the PSNR values of each frame obtained by different methods under different SRs. From Fig. 7, one can see that the proposed methods, respectively, obtain the highest and second-highest PSNR values in almost all frames.

Fig. 8 shows the restoration results of different methods on video dataset with SR = 10%. For better visualization, two local areas are chosen and enlarged under each image. From Fig. 8, we can see that the result of DCTTNN contains some spatial blurring effects. DTNN and TRLRF have a better performance than DCTTNN visually; however, some blurring effects still exist. On the contrary, CoNoT behaves well in spatial details compared with the compared methods. In addition, the spatial details and textures of Ms-CoNoT are clearer.

D. Multitemporal Remote Sensing Images

In this section, we test all methods for a classical cloud/shadow removal application. The testing data are taken

TABLE IV

QUANTITATIVE ASSESSMENT RESULTS OF DIFFERENT METHODS FOR VIDEO DATASET UNDER DIFFERENT SRS. THE **BEST** AND THE SECOND-BEST RESULTS FOR EACH QUALITY INDEX ARE HIGHLIGHTED BY **BOLDFACE** AND UNDERLINE, RESPECTIVELY

Dataset	SR	5%			10%			15%			20%			25%		
	Index	PSNR	SSIM	UIQI	PSNR	SSIM	UIQI	PSNR	SSIM	UIQI	PSNR	SSIM	UIQI	PSNR	SSIM	UIQI
Akiyo	Observed	6.649	0.013	0.003	6.886	0.019	0.008	7.130	0.025	0.013	7.396	0.032	0.020	7.676	0.038	0.027
	TRLRF [66]	29.183	0.871	0.707	31.665	0.920	0.787	32.757	0.937	0.821	33.485	0.945	0.836	34.022	0.952	0.851
	DCTTNN [39]	30.039	0.909	0.802	32.612	0.948	0.880	34.597	0.966	0.916	36.305	0.976	0.939	37.719	0.983	0.954
	DTNN [49]	29.198	0.909	0.796	32.510	0.954	0.884	35.537	0.976	0.940	38.030	0.986	0.965	**39.957**	**0.990**	0.976
	CT-LRTC [63]	30.669	0.926	0.831	33.170	0.957	0.898	35.004	0.971	0.928	36.594	0.979	0.946	37.847	0.984	0.959
	CoNoT	32.112	**0.954**	**0.916**	34.529	0.972	**0.948**	36.710	0.982	0.968	38.311	0.986	0.975	39.730	**0.990**	**0.980**
	Ms-CoNoT	**32.656**	0.951	0.828	**35.460**	**0.977**	0.944	**37.197**	**0.984**	**0.973**	**38.954**	**0.989**	**0.980**	39.816	0.988	0.950
Carphone	Observed	5.911	0.011	0.003	6.148	0.016	0.008	6.393	0.022	0.013	6.661	0.028	0.020	6.940	0.033	0.027
	TRLRF [66]	25.929	0.727	0.526	28.136	0.805	0.611	28.928	0.833	0.645	29.468	0.849	0.667	29.949	0.863	0.686
	DCTTNN [39]	25.717	0.731	0.532	27.794	0.800	0.615	29.167	0.839	0.664	30.349	0.868	0.700	31.333	0.889	0.729
	DTNN [49]	27.401	0.839	0.663	29.856	0.894	0.742	31.543	0.919	0.782	32.957	0.935	0.811	34.134	0.947	0.832
	CT-LRTC [63]	27.705	0.826	0.635	29.658	0.874	0.705	30.882	0.899	0.745	31.901	0.917	0.774	32.785	0.930	0.800
	CoNoT	28.405	0.839	0.680	30.370	**0.907**	**0.778**	31.874	**0.931**	**0.816**	33.086	**0.945**	**0.840**	34.034	**0.954**	**0.860**
	Ms-CoNoT	**29.535**	**0.874**	**0.686**	**31.519**	0.901	0.757	**32.751**	0.922	0.778	**33.574**	0.930	0.790	**34.409**	0.939	0.807
Foreman	Observed	3.372	0.005	0.002	3.606	0.008	0.005	3.854	0.011	0.009	4.117	0.014	0.013	4.397	0.017	0.017
	TRLRF [66]	22.520	0.513	0.382	24.484	0.640	0.498	25.379	0.693	0.548	25.862	0.719	0.578	26.310	0.742	0.605
	DCTTNN [39]	22.280	0.510	0.369	24.353	0.622	0.500	25.848	0.696	0.579	27.153	0.753	0.639	28.369	0.798	0.687
	DTNN [49]	24.460	0.718	0.564	26.597	0.802	0.672	28.190	0.848	0.735	29.573	0.879	0.776	30.833	0.901	0.809
	CT-LRTC [63]	24.707	0.668	0.539	26.724	0.773	0.634	28.090	0.820	0.693	29.263	0.855	0.738	30.390	0.881	0.773
	CoNoT	25.330	0.744	0.603	27.348	0.829	**0.714**	29.035	**0.875**	**0.777**	30.444	**0.903**	**0.817**	31.414	**0.919**	**0.842**
	Ms-CoNoT	**27.027**	**0.774**	**0.605**	**29.271**	**0.837**	0.686	**30.211**	0.854	0.721	**31.373**	0.883	0.765	**31.735**	0.888	0.786

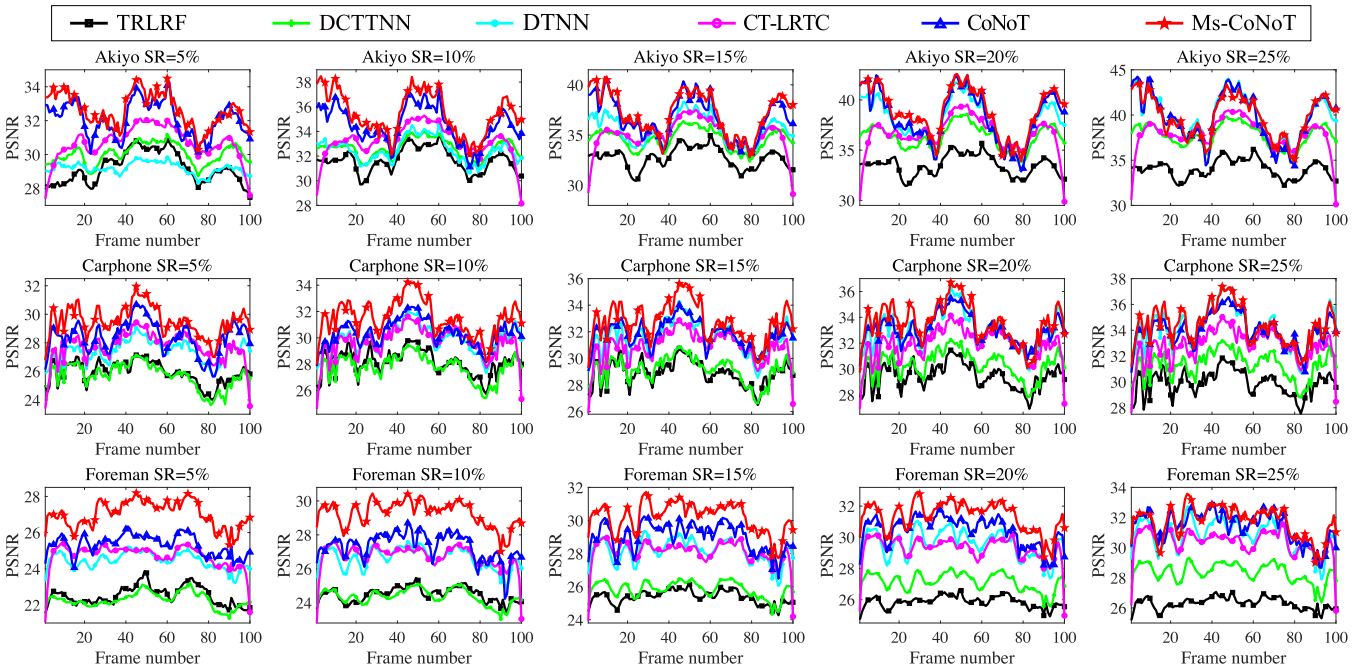


Fig. 7. PSNR values of each frames obtained by different methods under different SRs.

over *Morocco*, Africa, by Sentinel-2, which includes six different temporal data of size $400 \times 400 \times 4$ in our experiments. To simulate the real scene, we assume that all bands at the same time node are covered by the same type of clouds, and different time nodes are covered by different types of clouds.

Table V reports the quantitative assessment results of different methods on cloud/shadow removal application. The proposed methods, respectively, achieve the highest and second-highest PSNR, SSIM, and SAM values. Fig. 9 shows the false-color restoration results by different methods at two time nodes. For better visualization, one demarcated area (red rectangle) and the corresponding error (green rectangle) are magnified under each image. From Fig. 9, we can see that the result of TRLRF contains some spatial blurring effects. DTNN has a better performance than TRLRF visually; however, some blurring effects still exist. On the contrary, CoNoT behaves

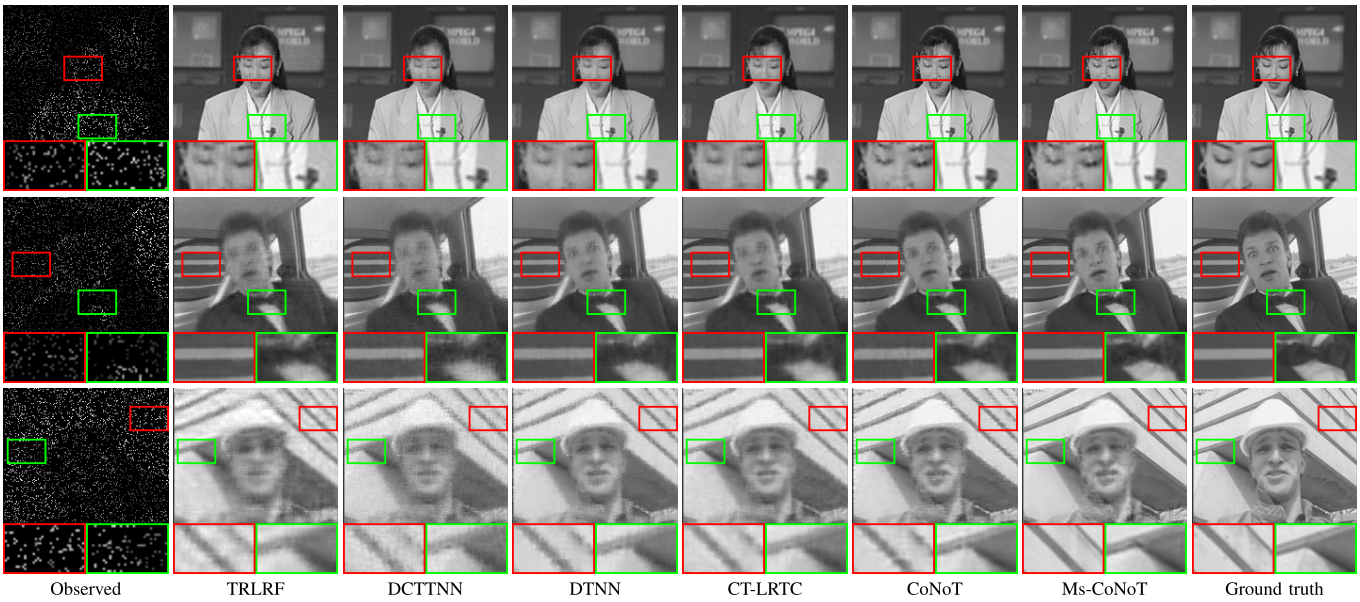


Fig. 8. Visualization of the restoration results by different methods on video dataset with SR = 10%. From top to bottom: the 85th frame of *Akiyo*, 55th frame of *Carphone*, and 60th frame of *Foreman*, respectively. For better visualization, two demarcated areas (red and green rectangles) are magnified under each image.

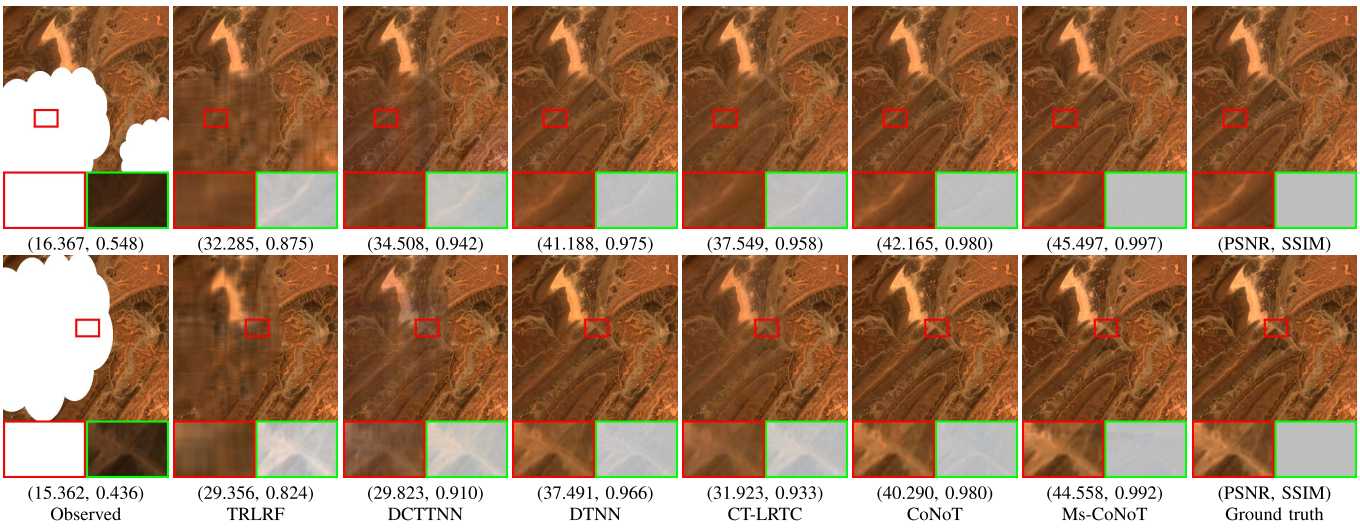


Fig. 9. Visualization of false-color restoration results (R:3, G:2, and B:1) by different methods on *Morocco*. From top to bottom: the results at the second-time node and the fourth-time node. For better visualization, one demarcated area (red rectangle) and the corresponding error (green rectangle) are magnified under each image.

well in spatial details or textures, and Ms-CoNoT's results are clearer.

E. Discussions

In this section, we present some necessary discussions about the proposed CoNoT and Ms-CoNoT.

1) Effectiveness of Spatial and Temporal/Spectral Transforms: The proposed CoNoT contains two transforms, which exploits the properties of spatial and spectral/temporal modes, respectively, and is internally connected and indeed complementary to each other. We illustrate this fact from two aspects. First, we show the AccEgy of different methods with respect to the percentage of singular values on MSI *Feathers* and video *Foreman* in Fig. 10. From Fig. 10, we can see that the energy of CoNoT is more concentrated compared with other methods, which means it can get a more compact representation, i.e., the data have a better tensor low tubal-rank approximation, and CoNoT has more flexible expression ability.

Then, we show the restoration results by different methods on MSI *Feathers* with SR = 5% and video *Foreman* with SR = 10% in Fig. 11. The result of CoNoT wo Tec. contains some spatial blurring effects on *Foreman*, and the color looks unnatural on *Feathers* compared with CoNoT, which implies the necessity of 1-D temporal/spectral convolution layer in the spectral (i.e., color) fidelity. The result of CoNoT wo Spc. loses some spatial details and textures compared with CoNoT, which implies the necessity of 2-D spatial convolution layer

TABLE V

QUANTITATIVE ASSESSMENT RESULTS OF DIFFERENT METHODS ON *Morocco*. THE **BEST** AND THE SECOND-BEST RESULTS FOR EACH QUALITY INDEX ARE HIGHLIGHTED BY **BOLDFACE** AND UNDERLINE, RESPECTIVELY

Dataset	Index	PSNR	SSIM	SAM
	Observed	17.017	0.567	39.36
	TRLRF [66]	34.672	0.898	3.137
	DCTTNN [39]	36.441	0.953	2.923
Decloud	DTNN [49]	44.219	0.983	1.146
	CT-LRTC [63]	40.267	0.969	2.040
	CoNoT	45.666	0.988	0.981
	Ms-CoNoT	**47.926**	**0.994**	**0.738**

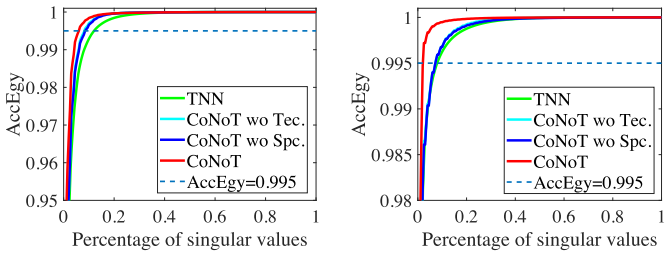


Fig. 10. AccEgy with respect to the percentage of singular values on MSI *Feathers* (left) and video *Foreman* (right). Here, CoNoT wo Tec. denotes the proposed CoNoT without 1-D temporal/spectral convolution layer and the corresponding nonlinear transform layer; CoNoT wo Spc. denotes the proposed CoNoT without 2-D spatial convolution layer and the corresponding nonlinear transform layer.

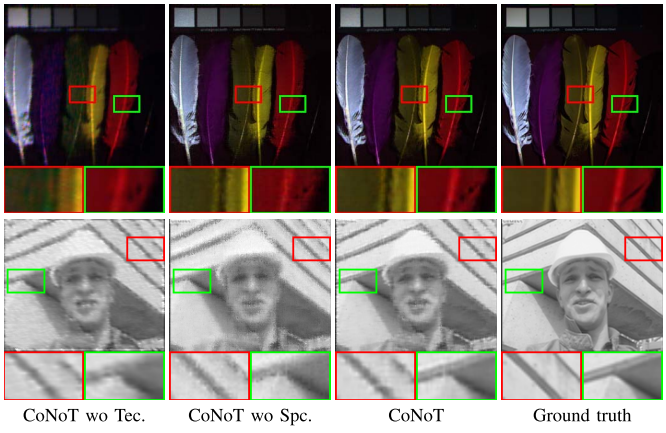


Fig. 11. Visualization of the restoration results by different methods on MSI *Feathers* (top) with SR = 5% and video *Foreman* (bottom) with SR = 10%.

in the spatial fidelity. On the contrary, CoNoT behaves well in spatial details, and the colors look natural compared with other methods, which once again visually shows that these two transforms are complementary to each other and validates our motivation.

2) Effectiveness of Nonlinearity: To verify the effectiveness of nonlinearity in the proposed CoNoT, we conduct the experiments to test the performance of different solvers, including the proposed CoNoT, CoNoT with different nonlinear activation functions, and CoNoT without nonlinear activation function. Table VI reports the quantitative assessment results of different methods on MSI *Balloons* with SR = 10%. From Table VI, we can observe that the methods with nonlinear activation functions can significantly improve performance in general. We attribute this to the powerful modeling capabilities of nonlinearity. Furthermore, CoNoT with LeakyReLU performs better than other nonlinear activation functions (i.e., ReLU, PReLU, ReLU6, and Sigmoid). Therefore, we choose LeakyReLU as the activation function in CoNoT in all our experiments.

TABLE VI

QUANTITATIVE ASSESSMENT RESULTS OF DIFFERENT METHODS ON MSI *Balloons* WITH SR = 10%. CONOT (LINEAR) INDICATES CONOT WITHOUT NONLINEAR ACTIVATION FUNCTION. CONOT (·) INDICATES CONOT WITH DIFFERENT NONLINEAR ACTIVATION FUNCTIONS

	Method	PSNR	SSIM	SAM	Time (s)
	CoNoT (Linear)	43.424	0.990	3.053	274
	CoNoT (ReLU)	42.354	0.986	3.591	266
Nonlinearity	CoNoT (LeakyReLU)	**44.220**	**0.991**	**2.392**	273
	CoNoT (PReLU)	43.996	**0.991**	2.653	221
	CoNoT (ReLU6)	42.871	0.986	3.562	221
	CoNoT (Sigmoid)	40.358	0.979	4.808	291

TABLE VII

QUANTITATIVE ASSESSMENT RESULTS BY THE PROPOSED CONOT WITH DIFFERENT n_4 VALUES ON MSIS WITH SR = 10%. THE **BEST** RESULTS FOR EACH QUALITY INDEX ARE HIGHLIGHTED BY **BOLDFACE**

Dataset	Index	$n_4 = 3n_3$	$n_4 = 4n_3$	$n_4 = 5n_3$	$n_4 = 6n_3$	$n_4 = 7n_3$
	PSNR	35.222	36.470	**37.023**	36.688	36.311
Watercolors	SSIM	0.964	0.974	**0.975**	0.974	0.973
	SAM	3.124	2.618	**2.433**	2.545	2.758
	PSNR	39.526	39.837	39.849	**40.415**	40.012
Feathers	SSIM	0.976	0.977	0.978	**0.982**	0.980
	SAM	4.250	4.111	4.077	**3.592**	4.128

3) Influence of n_4 Value: To discuss the influence of n_4 value on the performance of the proposed CoNoT, we report the numerical results with different n_4 values in Table VII, taking MSIs ***Watercolors*** and ***Feathers*** with SR = 10% as examples. From Table VII, we can observe that a moderate n_4 can achieve a satisfactory performance. However, the computational complexity of using the ADAM optimizer to solve the proposed CoNoT is $K^2n_1n_2n_3 + K^2n_1n_2n_4 + 2n_1n_2n_3n_4$ on a tensor of size $n_1 \times n_2 \times n_3$, where K is the size of the convolutional kernel. We can find that the computational complexity increases linearly with the increase of n_4. To balance the performance of the proposed CoNoT and the computational complexity, we set $n_4 = 4n_3$ in all our experiments.

4) Ablation Study: We replace 2-D + 1-D convolution layers with 3-D convolution layer in the proposed CoNoT block (named 3DCoNoT). Table VIII lists the quantitative indexes (PSNR/SSIM/SAM) and GPU times of the proposed CoNoT (i.e., 2-D + 1-D convolution layers in the proposed CoNoT block) and 3DCoNoT (i.e., 3-D convolution layer in the proposed CoNoT block) for CAVE dataset under different SRs. From Table VIII, we can observe that the proposed CoNoT consistently outperforms 3DCoNoT in terms of PSNR, SSIM, SAM values, and GPU time.

TABLE VIII

QUANTITATIVE ASSESSMENT RESULTS OF THE PROPOSED CONOT, 3DCONOT, AND DIP FOR CAVE DATASET UNDER DIFFERENT SRS. THE **BEST** AND THE SECOND-BEST RESULTS FOR EACH QUALITY INDEX ARE HIGHLIGHTED BY **BOLDFACE** AND UNDERLINE, RESPECTIVELY

Dataset	SR	10%			15%			20%			25%			Time (s)
	Index	PSNR	SSIM	SAM	PSNR	SSIM	SAM	PSNR	SSIM	SAM	PSNR	SSIM	SAM	
Peppers	Observed	13.168	0.216	74.78	13.419	0.254	70.38	13.690	0.290	66.31	13.953	0.323	62.51	–
	DIP [57]	47.888	0.996	1.223	48.982	0.996	1.196	49.284	0.996	1.228	50.548	0.997	1.084	1282
	3DCoNoT	42.831	0.990	1.977	43.070	0.990	2.140	46.466	0.995	1.466	50.862	**0.998**	1.059	828
	CoNoT	45.584	0.992	1.748	47.673	0.994	1.512	51.239	0.997	1.116	51.764	**0.998**	1.010	**262**
	Ms-CoNoT	**49.632**	**0.997**	**1.082**	**51.028**	**0.997**	**1.134**	**52.051**	**0.998**	**0.972**	**53.021**	**0.998**	**0.903**	846
Feathers	Observed	13.589	0.222	73.90	13.589	0.222	73.90	14.106	0.295	65.52	14.379	0.330	61.69	–
	DIP [57]	42.625	0.989	2.320	44.600	0.991	2.130	44.994	0.991	2.157	45.418	0.992	2.132	1273
	3DCoNoT	38.348	0.976	4.387	41.172	0.985	3.928	42.986	0.989	2.956	46.438	0.994	2.081	893
	CoNoT	39.489	0.976	4.161	42.522	0.987	3.197	44.922	0.991	2.554	46.782	0.993	2.279	**260**
	Ms-CoNoT	**43.954**	**0.993**	**1.860**	**46.532**	**0.996**	**1.684**	**49.047**	**0.997**	**1.526**	**49.790**	**0.997**	**1.499**	832

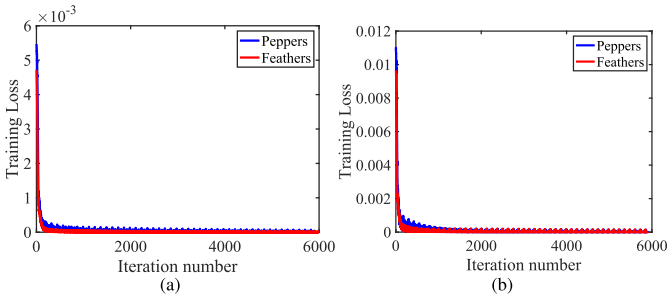


Fig. 12. Convergence analysis of the proposed ADAM algorithm. (a) Training loss versus iteration numbers with SR = 10%. (b) Training loss versus iteration numbers with SR = 20%.

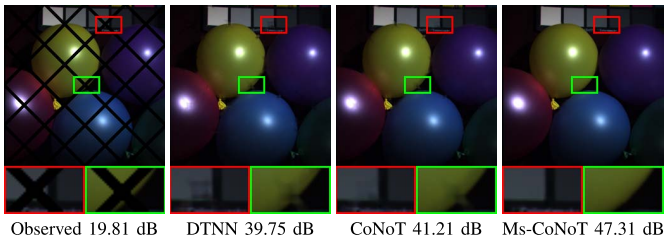


Fig. 13. Visualization of the restoration results by different methods on MSI *Balloons* with the structured missing.

5) Convergence Analysis: Since the proposed model is highly nonconvex and nonlinear, the theoretical convergence of the ADAM algorithm for the proposed model is still an open problem. Therefore, we provide the numerical convergence behavior of the proposed ADAM algorithm. Fig. 12 displays the curves of the training loss versus the number of iterations on MSIs ***Peppers*** and ***Feathers*** with SR = 10% and SR = 20%. From Fig. 12, we can observe that the training loss gradually decreases with slight fluctuations as the number of iterations increases, which empirically demonstrates the convergence behavior of the proposed ADAM algorithm.

6) Effectiveness of Multiscale Nature of Real-World Data: Here, we use a U-Net structure to faithfully capture the multiscale information of the data for better performance. To verify the effectiveness of the U-Net structure in Ms-CoNoT, in Fig. 13, we show the false-color restoration results by different methods on MSI *Balloons* with the structured missing. From Fig. 13, we can see that Ms-CoNoT outperforms CoNoT, where the spatial details and textures of the image are well protected. This is because the U-Net structure can further utilize the multiscale information of the data to better express the data, which also verifies our motivation.

7) Comparison With DIP: We compared the proposed methods with the unsupervised DL method [57]. The numerical results by our methods and DIP for CAVE dataset under different SRs are reported in Table VIII. From Table VIII, we can see that the proposed Ms-CoNoT consistently achieves best PSNR, SSIM, and SAM values compared with DIP, which indicates the advantage of the proposed method.

VI. CONCLUSION

In this work, we aim at the TC problems and propose a coupled nonlinear low-rank tensor representation called CoNoT. Based on this representation, we propose a novel multidimensional visual data completion model, which exhibits good representation ability. Furthermore, we also propose an enhanced version with the U-Net structure (called Ms-CoNoT), such that it can explore the spatial multiscale nature of real-world data. We directly use the gradient descent algorithm specifically designed for the deep neural network to tackle them. Extensive experimental results on real-world data demonstrate the effectiveness of the proposed methods for TC problems.

REFERENCES

[1] Y. Yang, Y. Feng, and J. A. K. Suykens, "Robust low-rank tensor recovery with regularized redescending M-estimator," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 27, no. 9, pp. 1933–1946, Sep. 2016.

[2] Y. Liu, X. Yuan, J. Suo, D. J. Brady, and Q. Dai, "Rank minimization for snapshot compressive imaging," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 41, no. 12, pp. 2990–3006, Dec. 2019.

[3] F. Fang, J. Li, Y. Yuan, T. Zeng, and G. Zhang, "Multilevel edge features guided network for image denoising," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 9, pp. 3956–3970, Sep. 2021.

[4] R. Dian, S. Li, and X. Kang, "Regularizing hyperspectral and multispectral image fusion by CNN denoiser," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 3, pp. 1124–1135, Mar. 2021.

[5] R. Dian, S. Li, and L. Fang, "Learning a low tensor-train rank representation for hyperspectral image super-resolution," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 30, no. 9, pp. 2672–2683, Sep. 2019.

[6] S. Javed, T. Bouwmans, and S. K. Jung, "SBMI-LTD: Stationary background model initialization based on low-rank tensor decomposition," in *Proc. Symp. Appl. Comput.*, Apr. 2017, pp. 195–200.

[7] I. Kajo, N. Kamel, Y. Ruichek, and A. Malik, "SVD-based tensor-completion technique for background initialization," *IEEE Trans. Image Process.*, vol. 27, no. 6, pp. 3114–3126, Jun. 2018.

[8] I. Kajo, N. Kamel, and Y. Ruichek, "Incremental tensor-based completion method for detection of stationary foreground objects," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 29, no. 5, pp. 1325–1338, May 2019.

[9] I. Kajo, N. Kamel, and Y. Ruichek, "Self-motion-assisted tensor completion method for background initialization in complex video sequences," *IEEE Trans. Image Process.*, vol. 29, pp. 1915–1928, 2019.

[10] T. N. Le, D. B. Giap, J.-W. Wang, and C.-C. Wang, "Tensor-compensated color face recognition," *IEEE Trans. Inf. Forensics Security*, vol. 16, pp. 3339–3354, 2021.

[11] S. Jia, K. Wu, J. Zhu, and X. Jia, "Spectral–spatial Gabor surface feature fusion approach for hyperspectral imagery classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 2, pp. 1142–1154, Feb. 2019.

[12] Q. Shi, Y.-M. Cheung, Q. Zhao, and H. Lu, "Feature extraction for incomplete data via low-rank tensor decomposition with feature regularization," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 30, no. 6, pp. 1803–1817, Jun. 2019.

[13] J. Xue, Y. Zhao, S. Huang, W. Liao, J. C.-W. Chan, and S. G. Kong, "Multilayer sparsity-based tensor decomposition for low-rank tensor completion," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 33, no. 11, pp. 1–15, Nov. 2021, doi: 10.1109/TNNLS.2021.3083931.

[14] A. Sobral and E.-H. Zahzah, "Matrix and tensor completion algorithms for background model initialization: A comparative evaluation," *Pattern Recognit. Lett.*, vol. 96, pp. 22–33, Sep. 2017.

[15] R. Dian, S. Li, L. Fang, T. Lu, and J. M. Bioucas-Dias, "Nonlocal sparse tensor factorization for semiblind hyperspectral and multispectral image fusion," *IEEE Trans. Cybern.*, vol. 50, no. 10, pp. 4469–4480, Oct. 2020.

[16] X. Xiao, Y. Chen, Y.-J. Gong, and Y. Zhou, "Prior knowledge regularized multiview self-representation and its applications," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 3, pp. 1325–1338, Mar. 2021.

[17] Y.-B. Zheng, T.-Z. Huang, X.-L. Zhao, and Q. Zhao, "Tensor completion via fully-connected tensor network decomposition with regularized factors," *J. Sci. Comput.*, vol. 92, no. 1, pp. 1–35, Jul. 2022, doi: 10.1007/s10915-022-01841-8.

[18] L.-B. Cui, Y.-D. Fan, Y.-S. Song, and S.-L. Wu, "The existence and uniqueness of solution for tensor complementarity problem and related systems," *J. Optim. Theory Appl.*, vol. 192, no. 1, pp. 321–334, Jan. 2022.

[19] L.-B. Cui, X.-Q. Zhang, and Y.-T. Zheng, "A preconditioner based on a splitting-type iteration method for solving complex symmetric indefinite linear systems," *Jpn. J. Ind. Appl. Math.*, vol. 38, no. 3, pp. 965–978, Sep. 2021.

[20] J. Yu, G. Zhou, C. Li, Q. Zhao, and S. Xie, "Low tensor-ring rank completion by parallel matrix factorization," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 7, pp. 3020–3033, Jul. 2021.

[21] J.-L. Wang, T.-Z. Huang, X.-L. Zhao, J. Huang, T.-H. Ma, and Y.-B. Zheng, "Reweighted block sparsity regularization for remote sensing images destriping," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 12, no. 12, pp. 4951–4963, Dec. 2019.

[22] C. Zeng, T.-X. Jiang, and M. K. Ng, "An approximation method of CP rank for third-order tensor completion," *Numerische Math.*, vol. 147, no. 3, pp. 727–757, Mar. 2021.

[23] X. Li, M. K. Ng, G. Cong, Y. Ye, and Q. Wu, "MR-NTD: Manifold regularization nonnegative tucker decomposition for tensor data dimension reduction and representation," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 28, no. 8, pp. 1787–1800, Aug. 2017.

[24] J. Liu, P. Musialski, P. Wonka, and J. Ye, "Tensor completion for estimating missing values in visual data," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 1, pp. 208–220, Jan. 2013.

[25] M. Al-Qizwini and H. Radha, "Fast smooth rank approximation for tensor completion," in *Proc. 48th Annu. Conf. Inf. Sci. Syst. (CISS)*, Mar. 2014, pp. 1–5.

[26] L.-T. Huang, H. C. So, Y. Chen, and W.-Q. Wang, "Truncated nuclear norm minimization for tensor completion," in *Proc. IEEE 8th Sensor Array Multichannel Signal Process. Workshop (SAM)*, Jun. 2014, pp. 417–420.

[27] Z.-F. Han, R. Feng, L.-T. Huang, Y. Xiao, C.-S. Leung, and H. C. So, "Tensor completion based on structural information," in *Proc. Int. Conf. Neural Inf. Process.*, 2014, pp. 479–486.

[28] X. Zhang, "A nonconvex relaxation approach to low-rank tensor completion," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 30, no. 6, pp. 1659–1671, Jun. 2019.

[29] C. Zhang, W. Hu, T. Jin, and Z. Mei, "Nonlocal image denoising via adaptive tensor nuclear norm minimization," *Neural Comput. Appl.*, vol. 29, no. 1, pp. 3–19, 2018.

[30] W. Hu, D. Tao, W. Zhang, Y. Xie, and Y. Yang, "The twist tensor nuclear norm for video completion," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 28, no. 12, pp. 2961–2973, Dec. 2017.

[31] C.-Y. Ko, K. Batselier, L. Daniel, W. Yu, and N. Wong, "Fast and accurate tensor completion with total variation regularized tensor trains," *IEEE Trans. Image Process.*, vol. 29, pp. 6918–6931, 2020.

[32] J. Yang, Y. Zhu, K. Li, J. Yang, and C. Hou, "Tensor completion from structurally-missing entries by low-TT-rankness and fiber-wise sparsity," *IEEE J. Sel. Topics Signal Process*, vol. 12, no. 6, pp. 1420–1434, Dec. 2018.

[33] H. Huang, Y. Liu, Z. Long, and C. Zhu, "Robust low-rank tensor ring completion," *IEEE Trans. Comput. Imag.*, vol. 6, pp. 1117–1126, 2020.

[34] X. P. Li and H. C. So, "Robust low-rank tensor completion based on tensor ring rank via $\ell_{p,\epsilon}$-norm," *IEEE Trans. Signal Process.*, vol. 69, pp. 3685–3698, 2021.

[35] W. He, N. Yokoya, L. Yuan, and Q. Zhao, "Remote sensing image reconstruction using tensor ring completion and total variation," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 11, pp. 8998–9009, Nov. 2019.

[36] X.-L. Zhao, J.-H. Yang, T.-H. Ma, T.-X. Jiang, M. K. Ng, and T.-Z. Huang, "Tensor completion via complementary global, local, and nonlocal priors," *IEEE Trans. Image Process.*, vol. 31, pp. 984–999, 2022.

[37] H. Zhang, X.-L. Zhao, T.-X. Jiang, M. K. Ng, and T.-Z. Huang, "Multiscale feature tensor train rank minimization for multidimensional image recovery," *IEEE Trans. Cybern.*, early access, Sep. 20, 2021, doi: 10.1109/TCYB.2021.3108847.

[38] K. Braman, "Third-order tensors as linear operators on a space of matrices," *Linear Algebra Appl.*, vol. 433, no. 7, pp. 1241–1253, 2010.

[39] C. Lu, X. Peng, and Y. Wei, "Low-rank tensor completion with a new tensor nuclear norm induced by invertible linear transforms," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 5996–6004.

[40] Z. Zhang and S. Aeron, "Exact tensor completion using t-SVD," *IEEE Trans. Signal Process.*, vol. 65, no. 6, pp. 1511–1526, Mar. 2017.

[41] H. Wang, F. Zhang, J. Wang, T. Huang, J. Huang, and X. Liu, "Generalized nonconvex approach for low-tubal-rank tensor recovery," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 33, no. 8, pp. 3305–3319, Aug. 2022, doi: 10.1109/TNNLS.2021.3051650.

[42] Y.-B. Zheng, T.-Z. Huang, X.-L. Zhao, T.-X. Jiang, T.-H. Ma, and T.-Y. Ji, "Mixed noise removal in hyperspectral image via low-fibered-rank regularization," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 1, pp. 734–749, Jan. 2020.

[43] M. Baburaj and S. N. George, "DCT based weighted adaptive multilinear data completion and denoising," *Neurocomputing*, vol. 318, pp. 120–136, Nov. 2018.

[44] G. Song, M. K. Ng, and X. Zhang, "Robust tensor completion using transformed tensor singular value decomposition," *Numer. Linear Algebra Appl.*, vol. 27, no. 3, p. e2299, 2020.

[45] M. K. Ng, X. Zhang, and X.-L. Zhao, "Patched-tube unitary transform for robust tensor completion," *Pattern Recognit.*, vol. 100, Apr. 2020, Art. no. 107181.

[46] E. Kernfeld, M. Kilmer, and S. Aeron, "Tensor–tensor products with invertible linear transforms," *Linear Algebra Appl.*, vol. 485, pp. 545–570, Nov. 2015.

[47] T.-X. Jiang, M. K. Ng, X.-L. Zhao, and T.-Z. Huang, "Framelet representation of tensor nuclear norm for third-order tensor completion," *IEEE Trans. Image Process.*, vol. 29, pp. 7233–7244, 2020.

[48] H. Kong, C. Lu, and Z. Lin, "Tensor Q-rank: New data dependent definition of tensor rank," *Mach. Learn.*, vol. 110, no. 7, pp. 1867–1900, Jul. 2021.

[49] T.-X. Jiang, X.-L. Zhao, H. Zhang, and M. K. Ng, "Dictionary learning with low-rank coding coefficients for tensor completion," *IEEE Trans. Neural Netw. Learn. Syst.*, early access, Aug. 31, 2021, doi: 10.1109/TNNLS.2021.3104837.

[50] Y.-S. Luo, X.-L. Zhao, T.-X. Jiang, Y. Chang, M. K. Ng, and C. Li, "Self-supervised nonlinear transform-based tensor nuclear norm for multi-dimensional image recovery," *IEEE Trans. Image Process.*, vol. 31, pp. 3793–3808, 2022.

[51] X.-Y. Liu and X. Wang, "Real-time indoor localization for smartphones using tensor-generative adversarial nets," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 8, pp. 3433–3443, Aug. 2021.
[52] H. Liu, Z. Wan, W. Huang, Y. Song, X. Han, and J. Liao, "PD-GAN: Probabilistic diverse GAN for image inpainting," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2021, pp. 9371–9381.
[53] Q. Zhang, Q. Yuan, C. Zeng, X. Li, and Y. Wei, "Missing data reconstruction in remote sensing image with a unified spatial–temporal-spectral deep convolutional neural network," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 8, pp. 4274–4288, Mar. 2018.
[54] H. Liu, Y. Li, M. Tsang, and Y. Liu, "CoSTCo: A neural tensor completion model for sparse tensors," in *Proc. 25th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Jul. 2019, pp. 324–334.
[55] J. Fan, C. Yang, and M. Udell, "Robust non-linear matrix factorization for dictionary learning, denoising, and clustering," *IEEE Trans. Signal Process.*, vol. 69, pp. 1755–1770, 2021.
[56] D. Hong et al., "Endmember-guided unmixing network (EGU-Net): A general deep learning framework for self-supervised hyperspectral unmixing," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 33, no. 11, pp. 6518–6531, Nov. 2021, doi: 10.1109/TNNLS.2021.3082289.
[57] V. Lempitsky, A. Vedaldi, and D. Ulyanov, "Deep image prior," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 9446–9454.
[58] Y.-C. Miao, X.-L. Zhao, X. Fu, J.-L. Wang, and Y.-B. Zheng, "Hyperspectral denoising using unsupervised disentangled spatiospectral deep priors," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–16, 2022.
[59] O. Sidorov and J. Y. Hardeberg, "Deep hyperspectral prior: Single-image denoising, inpainting, super-resolution," in *IEEE Int. Conf. Comput. Vis. Workshops*, Oct. 2019, pp. 3844–3851.
[60] Y. Zhang, X.-Y. Liu, B. Wu, and A. Walid, "Video synthesis via transform-based tensor neural network," in *Proc. 28th ACM Int. Conf. Multimedia*, Oct. 2020, pp. 2454–2462.
[61] T. G. Kolda and B. W. Bader, "Tensor decompositions and applications," *SIAM Rev.*, vol. 51, no. 3, pp. 455–500, 2009.
[62] N. Yair and T. Michaeli, "Multi-scale weighted nuclear norm image restoration," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 3165–3174.
[63] J.-L. Wang, T.-Z. Huang, X.-L. Zhao, T.-X. Jiang, and M. K. Ng, "Multi-dimensional visual data completion via low-rank tensor representation under coupled transform," *IEEE Trans. Image Process.*, vol. 30, pp. 3581–3596, 2021.
[64] B.-Z. Li, X.-L. Zhao, J.-L. Wang, Y. Chen, T.-X. Jiang, and J. Liu, "Tensor completion via collaborative sparse and low-rank transforms," *IEEE Trans. Comput. Imag.*, vol. 7, pp. 1289–1303, 2021.
[65] G. A. Watson, "Characterization of the subdifferential of some matrix norms," *Linear Algebra Appl.*, vol. 170, pp. 33–45, Jun. 1992.
[66] L. Yuan, C. Li, D. Mandic, J. Cao, and Q. Zhao, "Tensor ring decomposition with rank minimization on latent space: An efficient approach for tensor completion," in *Proc. AAAI Conf. Artif. Intell.*, vol. 33, 2019, pp. 9151–9158.
[67] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.
[68] R. H. Yuhas, A. F. H. Goetz, and J. W. Boardman, "Discrimination among semi-arid landscape endmembers using the spectral angle mapper (SAM) algorithm," in *Proc. Summaries 3rd Annu. JPL Airborne Geosci. Workshop*, 1992, pp. 147–149.
[69] Z. Wang and A. C. Bovik, "A universal image quality index," *IEEE Signal Process. Lett.*, vol. 9, no. 3, pp. 81–84, Aug. 2002.

Jian-Li Wang received the B.S. degree in mathematics and applied mathematics from Neijiang Normal University, Neijiang, China, in 2017. She is currently pursuing the Ph.D. degree with the School of Mathematical Sciences, University of Electronic Science and Technology of China, Chengdu, China.

Her research interests include high-dimensional data processing, tensor modeling and computing, computer vision, and deep learning. More information can be found in her homepage https://wangjianli123.github.io/homepage/.

Ting-Zhu Huang received the B.S., M.S., and Ph.D. degrees from the Department of Mathematics, Xi'an Jiaotong University, Xi'an, China, in 1986, 1992, and 2001, respectively, all in computational mathematics.

He is currently a Professor with the School of Mathematical Sciences, University of Electronic Science and Technology of China, Chengdu, China. His current research interests include scientific computation and applications, numerical algorithms for image processing, numerical linear algebra, preconditioning technologies, and matrix analysis with applications.

Dr. Huang is an Editor of the *Scientific World Journal*, *Advances in Numerical Analysis*, the *Journal of Applied Mathematics*, the *Journal of Pure and Applied Mathematics: Advances in Applied Mathematics*, and the *Journal of Electronic Science and Technology*, China.

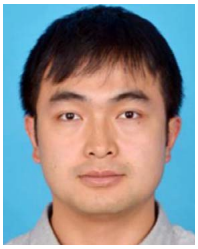
Xi-Le Zhao received the M.S. and Ph.D. degrees from the University of Electronic Science and Technology of China (UESTC), Chengdu, China, in 2009 and 2012, respectively.

He is currently a Professor with the School of Mathematical Sciences, UESTC. His research interests include model-driven and data-driven methods for image processing problems. His homepage is https://zhaoxile.github.io/.

Yi-Si Luo received the B.S. degree from the School of Mathematical Sciences, University of Electronic Science and Technology of China, Chengdu, China, in 2022. He is currently pursuing the M.S. degree with the School of Mathematics and Statistics, Xi'an Jiaotong University, Xi'an, China.

His research interests include model-based tensor modeling and unsupervised learning for low-level visual tasks, such as inpainting, denoising, and deraining.

Prof. Luo served as a regular Reviewer for IEEE Conference on Computer Vision and Pattern Recognition (CVPR) 2022, European Conference on Computer Vision (ECCV) 2022, and AAAI Conference on Artificial Intelligence (AAAI) 2023. His homepage is https://yisiluo.github.io/.



Tai-Xiang Jiang received the Ph.D. degree in mathematics from the University of Electronic Science and Technology of China (UESTC), Chengdu, China, in 2019.

From 2017 to 2018, he was a co-training Ph.D. student with the University of Lisbon, Lisbon, Portugal, supervised by Prof. Jose M. Bioucas-Dias. In 2019, he was a Research Assistant with Hong Kong Baptist University, Hong Kong, supported by Prof. Michael K. Ng. He is currently an Associated Professor with the School of Economic Information Engineering, Southwestern University of Finance and Economics, Chengdu. His research interests include sparse and low-rank modeling and tensor decomposition for multidimensional image processing, especially on the low-level inverse problems for multidimensional images. https://sites.google.com/view/taixiangjiang/