Hyperspectral-Multispectral Image Fusion via Tensor Ring and Subspace Decompositions

Honghui Xu, Mengjie Qin, Shengyong Chen, *Senior Member, IEEE*, Yuhui Zheng, and Jianwei Zheng

***Abstract*—Fusion from a spatially low resolution hyperspectral image (LR-HSI) and a spectrally low resolution multispectral image (MSI) to produce a high spatial-spectral HSI (HR-HSI), known as Hyperspectral super-resolution, has risen to a preferred topic for reinforcing the spatial-spectral resolution of HSI in recent years. In this work, we propose a new model, namely low-rank tensor ring decomposition based on tensor nuclear norm (LRTRTNN), for HSI-MSI fusion. Specifically, for each spectrally subspace cube, similar patches are grouped to exploit both the global low-rank property of LR-HSI and the non-local similarity of HR-MSI. Afterwards, a joint optimization of all groups via the presented LRTRTNN approximation is implemented in a unified cost function. With the introduced tensor nuclear norm (TNN) constraint, all 3D tensor ring factors are no longer unfolded to suit the matrix nuclear norm used in conventional methods, and the internal tensor structure can be naturally retained. The Alternating Direction Method of Multipliers (ADMM) is introduced for coefficients update. Numerical and visual experiments on real data show that our LRTRTNN method outperforms most state-of-the-art algorithms in terms of fusing performance.**

***Index Terms*—Image fusion, Hyperspectral super-resolution, hyperspectral imaging, multispectral image, tenson ring, tensor nuclear norm, low-rank decomposition.**

I. Introduction

Hyperspectral imaging is a rising modality where a camera acquires images with tens or even hundreds of spectral bands in wide-range spectral coverage from a scene across. Since the materials usually have different reflectance for various wavelengths, hyperspectral images (HSIs) capacitate an explicit discernment of the materials present within the scene, which underlies a major amount of comprehensive applications on anomaly detection [1], [2], remote sensing [3]–[8], and classification [9], [10], etc. However, there is invariably a certain compromise between the spectral dimension and spatial dimension for existing cameras due to the quantity of the incident energy is restricted in the optical remote sensing systems [11]. Since high spectral dimension is required in HSIs, the spatial resolution is sacrificed due to the limitations of equipment. On the contrary, the acquired MSIs have a higher spatial dimension, but the spectral resolution is lower. Therefore, it is an increasingly promising and economical approach to generate the high spatial resolution HSI (HR-HSI) by fusing a high spatial resolution MSI (HR-MSI) with the corresponding low spatial resolution HSI (LR-HSI) of the same scene.

The recent HSI-MSI fusion methods can be principally categorized into four families: Pan-sharpening [12]–[15], deep learning [16]–[19], matrix factorization (MF) based approaches [20]–[25], and tensor representation based approaches [26]–[33]. The typical Pan-sharpening based Fusion methods attempt to fuse a high-resolution panchromatic (PAN) image with a low-resolution MSI for spatially resolution enhancement. According to the spectral coverage of HSIs, Chen *et al*. [14] separated the spectral channels into several groups, and combined each spectra of MSIs with the corresponding band of HSI following pan-sharpening scheme. Although this can achieve better fusion results and have low computational cost, the fused HSI often produces significant distortions once the spatial dimensions of MSI and HSI is quite different.

With the prosperous application of deep learning in object detection, pattern recognition, denoising, and inpainting problems, a large number of researchers have applied this technology into HSI-MSI fusion problem. Qu *et al*. [16] presented an unsupervised deep CNN model combining two encoders with a shared decoder for HSI-MSI fusion, which inherits an advantage of needless usage of the pretraining stage. Yuan *et al*. [17] offered a multi-scale and multi-depth CNN for pansharpening. Although deep CNN can achieve appealing fusion effect, it is extremely complex and cumbersome to evaluate the parameters in the network by using massive amounts of training samples. Furthermore, the generalization capability of those strategies is commonly restricted by the specific model structure and the priori distribution of training data.

The matrix factorization based HSI-MSI fusion can be further categorized into two branches: sparse representation and low-rank constraint. These approaches all unfold 3D HSI into 2D matrix, and some of them factor the unfolded matrix into a spectral basis matrix and a coefficient matrix for better revealing the intrinsic structure. Dong *et al*. [20] proposed a nonnegative over-complete dictionary learning method to estimate the spectral basis from LR-HSI and used the sparse property to learn the coefficient matrix. Simoes *et al*. [25] applied the vertex component analysis (VCA) to HSI to extract

This work was supported in part by the National Key R&D Program of China under Grant 2018YFE0126100, the National Natural Science Foundation of China under Grant 62020106004, and Grant 61602413, Zhejiang Provincial Natural Science Foundation of China under Grant LY19F030016, Open Research Projects of Zhejiang Lab under Grant 2019KD0AD01/007, and the Scientific Research Fund of the National Health Commission of China under Grant WKJ-ZJ-2102. (Corresponding author: Jianwei Zheng.)

H. Xu, M. Qin, and J. Zheng are with the College of Computer Science and Technology, Zhejiang University of Technology, Hangzhou 310023, China (e-mail: xhh@zjut.edu.cn; qmj@zjut.edu.cn; zjw@zjut.edu.cn).

Shengyong Chen is with the College of Computer Sciences and Engineering, Tianjin University of Technology, Tianjin 300384, China (e-mail: sy@ieee.org).

Y. Zheng is with School of Computer and Software, Nanjing University of Information Science and Technology, Nanjing 210044, China, and also with the Engineering Research Center of Digital Forensics, Ministry of Education. (Email: zhengyh@vip.126.com).

the spectral dictionary and the spatial total variation constraint was also applied to obtain the smooth coefficients. As for the low-rank based approaches, Veganzones *et al*. [23] learned the spectral base of HSI by dividing the whole image into patches, which makes full use of the low-rank prior of HSI. Wei *et al*. [24] first described the multi-band fusion as an ill-posed inverse problem, then took consideration on both the low-rank property of linear spectral mixing model and the non-negativity resulting from inherent physical properties of the abundance to solve the fusion problem. In practical applications, although the matricization operation involved in these methods would not change any values in the original 3D data, it indeed breaks the spatial-spectral structure information of HSIs and suffers from performance degradation.

In the last few years, tensor factorization has emerged rapidly in the field of multi-frame image denoising, completion, compressive sensing, and classification [10], [34]. Nowadays, tensor factorization has become another hot scheme to handle the HSI-MSI fusion problem. Dian *et al*. [11] divided the HSI into smaller full-band cubes and clustered these patches following the guided information from the HR-MSI. Since that cubes in the same group share the same dictionaries, they made full use of non-local self-similarities in HSIs. Then, Li *et al*. [26] factorized the HSI as a sparse core tensor and three dictionaries of different modes via proximal altercating optimization, which was called coupled sparse Tucker decomposition (CSTF) algorithm. In [31], Kanatsoulis *et al*. used CANDECOMP/PARAFAC (CP) decomposition to divide HR-HSI into three parts and then updated each part iteratively. Chang *et al*. [29] also decomposed the HSI into a core tensor multiplied by three matrices along with different channels, whose advantage lies in the assignment of different weights to the singular values (SV) of the core factor. In [32], the authors formed 4D tensors from grouped clusters for holding the intrinsic spatial-spectral structure. On that basis, the tensor-train (TT) rank regularization was imposed on the 4-D tensors. Subsequently, a new tensor nuclear norm regularizer was further imposed to measure the correlation of non-local similar image patches [28]. He *et al*. [27] applied the tensor ring (TR) decomposition to get a core tensor with high spectral resolution from HSI and two other factors with high spatial resolution from MSI. Meanwhile, they penalized matrix nuclear norm constraint on the spectral factor along mode-2 for exploiting the globally spectral low-rank property. Among all these tensor decompositions, tensor networks based completion methods such as tensor train [35], tensor tree [36], and tensor ring [27], [37]–[40] decompositions achieve superior performance than the rest ones since they can capture more inherent correlations. Numerically, for any Nth-order tensor, the space complexity of CP, TT, and TR decompositions are linear with $O(N)$, while the case of Tucker decomposition grows exponentially with N. Besides, although CP decomposition follows the same space complexity as TT and TR decompositions, it is not easy to search for the best latent tensors. Since TT constitutes the rank vector of factors by a well-designed matricization scheme, it can better reveal the global correlation of tensor elements. However, TT needs rank-1 condition on the border factors and larger ranks for the relatively middle factors, which ends up with restricted representation ability and flexibility. Comparatively, TR can be regarded as a linear combination of TT, which provides a more powerful and generalized expression ability to alleviate the shortcomings of TT. That is to say, TR not only has all the dominant properties of TT, but also relaxes the constraints of TT ranks. Moreover, since tensor ring has the rotation invariance property of latent factors, its model constraints are more flexible than TT. In practice, TR can enhance the compression ability and also improve the interpretability of latent factors. The minimization of TR rank has become a hot research topic on the purpose of correctly discovering the latent space of tensor ring factors. He *et al*. [27] introduced the matrix nuclear norm to the third core tensor along mode-2. Moreover, Yuan *et al*. [37] imposed low-rankness on each of the TR factors. However, although acceptable performance can be achieved from these algorithms, they all unfolded the tensor core to mode-n matrix which may result in loss of optimality in the representation.

In this study, to boost the HSI-MSI fusion performance, a new tensor ring and subspace decomposition method is proposed. Specifically, the spectral subspace from LR-HSI is estimated by singular value decomposition, in which the full band cubes with a given size can be partitioned and gathered. For each separated cube, similar patches can be collected to form a tensor group centered at the referred cube. On that basis, all groups for different cubes would be jointly optimized using the proposed LRTRTNN constraint. The main difference of our method from several state-of-the-arts is threefold. First, we handle the fusion problem in a patch-wise way for more detailed information collection, instead of the treatment in [26] and [27] directly on the whole image. Second, we collect similar patches for each cube to take advantage of the stricter low-rank property, rather than a simple clustering of the divided cubes in [28]. More importantly, for better TR representation capability, we impose tensor nuclear norm on the 3D ring factors for capturing the internal structure of the tensors, rather than the matrix nuclear norm used in [37] and [27]. Generally speaking, our LRTRTNN method takes all the existing issues into account and the primary contributions are highlighted as follows.

1) A general framework combing both global spectral low-rank and non-local spatial self-similarity is presented for HSI-MSI fusion problem. Specifically, the singular values decomposition of LR-HSI is used to learn the spectral subspace at first. Thereafter, the non-local similar patches are grouped for revelation of better spatially low-rank prior. On that basis, our proposed tensor ring decomposition under the tensor nuclear norm constraint can be naturally adopted.
2) A new and compact tensor ring decomposition model are proposed to describe the intrinsic spatial, spectral, and non-local similarly simultaneously, in which the tensor nuclear norm is imposed for all TR factors. The popular alternating direction method of multiplier method (ADMM) algorithm is exploited for model optimization.
3) The proposed method, namely Low-rank tensor ring de-

composition based on tensor nuclear norm (LRTRTNN), is applied to HSI-MSI fusion tasks. Numerical experiments verify that the proposed method outperforms most state-of-the-art approaches in terms of fusing performance.

II. NOTATIONS AND PRELIMINARIES

In this section, we first briefly present the basic principles of the fusion problem, then introduce the specific definitions and properties of tensor ring decomposition and tensor nuclear norm.

A. PROBLEM FORMULATION

In this work, all of the LR-HSI, HR-MSI, and HR-HSI are denoted as 3D tensors. We seek to restructure the target HR-HSI $\boldsymbol{\mathcal{Z}} \in \mathbb{R}^{W\times H\times S}$ by fusing the acquired LR-HSI $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{w\times h\times S}$ and the corresponding HR-MSI $\boldsymbol{\mathcal{Y}} \in \mathbb{R}^{W\times H\times s}$ of the same scene. Here W and H denote the spatial dimension of HR-HSI and HR-MSI, w and h represent the spatial size of LR-HSI. S and s are the band sizes of HR-HSI and HR-MSI, respectively.

Note that $W > w$ and $H > h$ always hold, and LR-HSI $\boldsymbol{\mathcal{X}}$ could be downsampled from $\boldsymbol{\mathcal{Z}}$ with regarding to the spatial dimension, in which the relationship can be written as

$$\boldsymbol{X}_{(3)} = \boldsymbol{Z}_{(3)}\boldsymbol{B}\boldsymbol{S} \tag{1}$$

where $\boldsymbol{Z}_{(3)} \in \mathbb{R}^{S\times HW}$ and $\boldsymbol{X}_{(3)} \in \mathbb{R}^{S\times hw}$ are the third mode unfolded matrices of $\boldsymbol{\mathcal{Z}}$ and $\boldsymbol{\mathcal{X}}$, respectively. The spatial downsampling matrix $\boldsymbol{S} \in \mathbb{R}^{WH\times wh}$ selects the corresponding spectral pixels. The block circulant matrix $\boldsymbol{B} \in \mathbb{R}^{HW\times HW}$ that is diagonalized by fast Fourier transform (FFT) represents the convolution blur, and denoted by

$$\boldsymbol{B} = \boldsymbol{F}\boldsymbol{K}\boldsymbol{F}^{-1} \tag{2}$$

where the diagonal matrix $\boldsymbol{K}$ holds eigenvalues of $\boldsymbol{B}$, $\boldsymbol{F}$ and $\boldsymbol{F}^{-1}$ respectively denote the fast Fourier transform and its inverse operation.

We can identify the same scene through multispectral imaging sensors to obtain HR-MSI $\boldsymbol{\mathcal{Y}}$, and model the HR-MSI as the downsampled version from $\boldsymbol{\mathcal{Z}}$ with regarding to the spectral dimension.

$$\boldsymbol{Y}_{(3)} = \boldsymbol{R}\boldsymbol{Z}_{(3)} \tag{3}$$

where $\boldsymbol{Y}_{(3)} \in \mathbb{R}^{s\times HW}$ is the folding matrix of $\boldsymbol{\mathcal{Y}}$ along the third dimension. $\boldsymbol{R} \in \mathbb{R}^{s\times S}$ denotes the spectral response matrix of MSI sensor.

B. Tensor-Ring Decomposition

The TR decomposition can rewrite a higher-order tensor $\boldsymbol{\mathcal{T}} \in \mathbb{R}^{I_1\times I_2\times\cdots\times I_N}$ as the circular multilinear products across a series of low-order latent core factors $\boldsymbol{\mathcal{G}} := \left\{\boldsymbol{\mathcal{G}}^{(1)}, \ldots, \boldsymbol{\mathcal{G}}^{(N)}\right\}$, in which $\boldsymbol{\mathcal{G}}^{(n)} \in \mathbb{R}^{R_n\times I_n\times R_{n+1}}$ with $n = 1, 2, \cdots, N$. The syntax $\{R_1, R_2, \ldots, R_N\}$ is the TR-rank controlling the model complexity after TR decomposition. Note the trace operation is adopted in TR decomposition and all the core factors are set to be 3-order. Specifically, the elementwise relationship of $\boldsymbol{\mathcal{T}}$ with the core factor $\boldsymbol{\mathcal{G}}$ can be written by

$$\boldsymbol{\mathcal{T}}\left(i_1, i_2, \cdots, i_N\right) = \mathrm{Tr}\left(\mathbf{G}^{(1)}\left(i_1\right)\mathbf{G}^{(2)}\left(i_2\right)\cdots\mathbf{G}^{(N)}\left(i_N\right)\right) \tag{4}$$

where $\mathrm{Tr}(\cdot)$ is the matrix trace operation and $\mathbf{G}^{(n)}\left(i_n\right)$ denotes the i_n-th lateral slice of $\boldsymbol{\mathcal{G}}^{(n)}$. According to the characteristics of the trace operation, the product of these slices is a square matrix. Benefiting from this operation, tensor ring decomposition relaxes the rank constraint on the first and last core of TT to $R_1 = R_{N+1}$, rather than the original condition $R_1 = R_{N+1} = 1$. A brief illustration of TR decomposition is provided in Fig. 1. Next, we tend to present two necessary characteristics of this operation.

Definition 1: (Tensor Multilinear Product [40]) Let $\boldsymbol{\mathcal{G}}^{(n)}$ and $\boldsymbol{\mathcal{G}}^{(n+1)}$ denote two adjacent cores of tensor ring decomposition. The tensor multilinear product of the two nearby cores is denoted by

$$\boldsymbol{\mathcal{G}}^{(n,n+1)}\left(\left(j_l-1\right)I_n + i_k\right) = \boldsymbol{\mathcal{G}}^{(n)}\left(i_k\right)\boldsymbol{\mathcal{G}}^{(n+1)}\left(j_l\right) \tag{5}$$

for $i_k = 1, \cdots, I_n$, and $j_l = 1, \cdots, I_{n+1}$. From this definition, we can obtain this product of all tensor cores $[\boldsymbol{\mathcal{G}}] = \prod_{k=1}^{N}\boldsymbol{\mathcal{G}}^{(k)} = \boldsymbol{\mathcal{G}}^{(1,2,\cdots,N)} = \left\{\boldsymbol{\mathcal{G}}^{(1)}, \boldsymbol{\mathcal{G}}^{(2)}, \cdots, \boldsymbol{\mathcal{G}}^{(N)}\right\} \in \mathbb{R}^{R_1\times I_1 I_2\cdots I_N\times R_1}$. Therefore, the TR decomposition (Eq. (4)) can be reexpressed as $\boldsymbol{\mathcal{T}} = \Phi[\boldsymbol{\mathcal{G}}]$, in which Φ represents the tensor ring transformation operator converting the 3D tensor $[\boldsymbol{\mathcal{G}}]$ to tensor $\boldsymbol{\mathcal{T}}$, i.e., $\Phi : \mathbb{R}^{R_1\times I_1 I_2\cdots I_N\times R_1} \rightarrow \mathbb{R}^{I_1\times I_2\times\cdots\times I_N}$.

Proposition 1: (Circular dimensional permutation invariance [40] [41]) For the tensor $\boldsymbol{\mathcal{T}} \in \mathbb{R}^{I_1\times I_2\times\cdots\times I_N}$, if $\overleftarrow{\boldsymbol{\mathcal{T}}_n} \in \mathbb{R}^{I_{n+1}\times\cdots\times I_N\times I_1\times\cdots\times I_n}$ is represented as circularly moving the dimensions of $\boldsymbol{\mathcal{T}}$ by n, then we get

$$\overleftarrow{\boldsymbol{\mathcal{T}}_{\boldsymbol{n}}} = \Phi\left(\left\{\boldsymbol{\mathcal{G}}^{(n+1)}, \ldots, \boldsymbol{\mathcal{G}}^{(N)}, \boldsymbol{\mathcal{G}}^{(1)}, \ldots, \boldsymbol{\mathcal{G}}^{(n)}\right\}\right) \tag{6}$$

Proposition 1 shows that TR decomposition holds the capacity to reveal the correlation of elements in multiple dimensions, therefore we could readily move the middle factors of TR to the marginal position and analyze each factor separately.

C. t-SVD and TNN

In this subsection, we further introduce some definitions about TNN and tensor singular value decomposition (t-SVD). Note that tensors used in this subsection are all 3-order tensors.

Definition 2. (t-product [42]) For any two acquired tensors $\boldsymbol{\mathcal{A}} \in \mathbb{R}^{W\times H\times S}$ and $\boldsymbol{\mathcal{B}} \in \mathbb{R}^{H\times l\times S}$, the t-product $\boldsymbol{\mathcal{C}}$ of $\boldsymbol{\mathcal{A}}$ and $\boldsymbol{\mathcal{B}}$ is denoted as a tensor with size $W \times l \times S$, i.e.,

$$\boldsymbol{\mathcal{C}} = \boldsymbol{\mathcal{A}} * \boldsymbol{\mathcal{B}} = \mathrm{fold}(\mathrm{bcirc}(\boldsymbol{\mathcal{A}}) \cdot \mathrm{unfold}(\boldsymbol{\mathcal{B}})) \tag{7}$$

where $\cdot$ is the matrix product, bcirc($\boldsymbol{\mathcal{A}}$) is a block circulant matrix whose form can be written as

$$\mathrm{bcirc}(\boldsymbol{\mathcal{A}}) = \begin{bmatrix} \boldsymbol{A}^{(1)} & \boldsymbol{A}^{(S)} & \cdots & \boldsymbol{A}^{(2)} \\ \boldsymbol{A}^{(2)} & \boldsymbol{A}^{(1)} & \cdots & \boldsymbol{A}^{(3)} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{A}^{(S)} & \boldsymbol{A}^{(S-1)} & \cdots & \boldsymbol{A}^{(1)} \end{bmatrix}, \tag{8}$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JSTARS.2021.3108233, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing

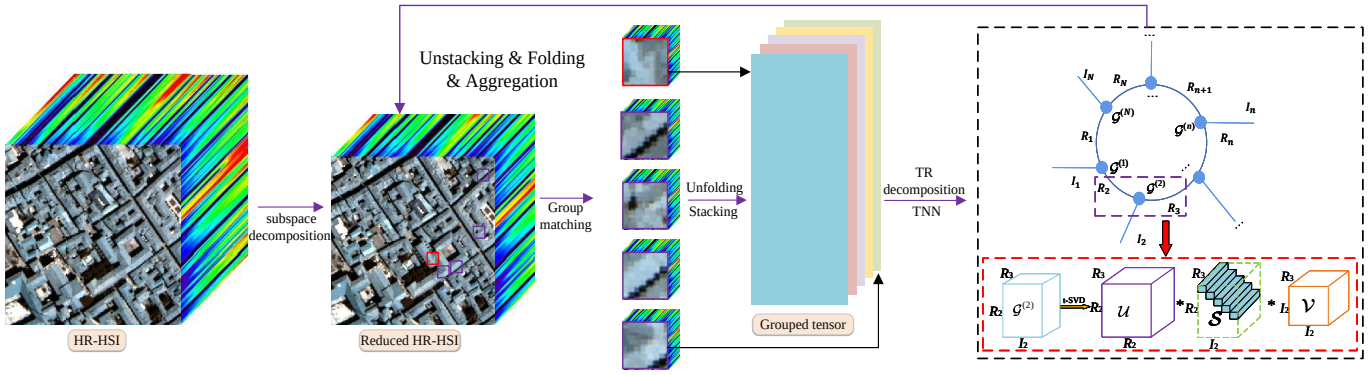


Fig. 1. Illustration of the proposed LRTRTNN method for the HR-HSI. The red box represents the TNN constraint on some TR factor.

and the unfold operator as well as its inverse operator are as follows.

$$\text{unfold}(\boldsymbol{\mathcal{A}}) = \begin{bmatrix} \boldsymbol{A}^{(1)} \\ \boldsymbol{A}^{(2)} \\ \vdots \\ \boldsymbol{A}^{(S)} \end{bmatrix}, \text{fold}(\text{unfold}(\boldsymbol{\mathcal{A}})) = \boldsymbol{\mathcal{A}}. \quad (9)$$

Indeed, t-product is homologous to the matrix product with an added operation of circulant convolution, therefore Eq. (7) can also be written in the Fourier domain as

$$\boldsymbol{\mathcal{C}} = \boldsymbol{\mathcal{A}} * \boldsymbol{\mathcal{B}} \Longleftrightarrow \bar{\boldsymbol{C}} = \bar{\boldsymbol{A}}\bar{\boldsymbol{B}}, \quad (10)$$

where the terms $\bar{\boldsymbol{A}}$, $\bar{\boldsymbol{B}}$, and $\bar{\boldsymbol{C}}$ are all block diagonal matrices whose frontal slices are diagonal blocks, e.g., $\bar{\boldsymbol{A}}$ is written as:

$$\bar{\boldsymbol{A}} = \text{bdiag}(\bar{\boldsymbol{\mathcal{A}}}) = \begin{bmatrix} \bar{\boldsymbol{A}}^{(1)} & & & \\ & \bar{\boldsymbol{A}}^{(2)} & & \\ & & \ddots & \\ & & & \bar{\boldsymbol{A}}^{(S)} \end{bmatrix}, \quad (11)$$

where $\bar{\boldsymbol{\mathcal{A}}}$ represents the result of Discrete Fourier transformation (DFT) of $\boldsymbol{\mathcal{A}}$, bdiag$(\cdot)$ turns tensor $\bar{\boldsymbol{\mathcal{A}}}$ into block diagonal matrix $\bar{\boldsymbol{A}}$.

Definition 3. (t-SVD [42]) Any given tensor $\boldsymbol{\mathcal{A}} \in \mathbb{R}^{W\times H\times S}$ can be factorized as

$$\boldsymbol{\mathcal{A}} = \boldsymbol{\mathcal{U}} * \boldsymbol{\mathcal{S}} * \boldsymbol{\mathcal{V}}^T \quad (12)$$

where $\boldsymbol{\mathcal{U}} \in \mathbb{R}^{W\times W\times S}$, $\boldsymbol{\mathcal{V}} \in \mathbb{R}^{H\times H\times S}$ are orthonormal tensors and $\boldsymbol{\mathcal{S}} \in \mathbb{R}^{W\times H\times S}$ is an f-diagonal tensor whose frontal slices are diagonal matrices. The red box in Fig. 1 shows an example. Through Eq. (10), t-SVD can also be written as

$$\bar{\boldsymbol{A}}^{(i)} = \bar{\boldsymbol{U}}^{(i)}\bar{\boldsymbol{S}}^{(i)}(\bar{\boldsymbol{V}}^{(i)})^T, \quad i = 1, 2, \ldots, S. \quad (13)$$

Definition 4.(tensor tubal rank [43])The tensor tubal rank of $\boldsymbol{\mathcal{A}} \in \mathbb{R}^{W\times H\times S}$, which is denoted as $rank_t(\boldsymbol{\mathcal{A}})$, is the amount of nonzero singular tubes of $\boldsymbol{\mathcal{S}}$,

$$rank_t(\boldsymbol{\mathcal{A}}) = \#\left\{i, \boldsymbol{\mathcal{S}}(i,i,:) \neq 0\right\}, \quad (14)$$

where $\boldsymbol{\mathcal{S}}$ is from the t-SVD of $\boldsymbol{\mathcal{A}}$.

Definition 5.(tensor nuclear norm [43]) For any tensor $\boldsymbol{\mathcal{A}} \in \mathbb{R}^{W\times H\times S}$, TNN is denoted as the summation of SVs of each frontal slice of $\bar{\boldsymbol{\mathcal{A}}}$

$$\|\boldsymbol{\mathcal{A}}\|_* = \frac{1}{S}\sum_{i=1}^{S}\left\|\bar{\boldsymbol{A}}^{(i)}\right\|_*. \quad (15)$$

III. PROPOSED METHOD

In this section, we mainly introduce our unified HSI-MSI fusion model which integrates the new tensor ring decomposition based on tensor nuclear norm, global spectrum low-rank, and non-local similarity priors in the same framework. We first utilize the globally spectral low-rank prior of LR-HSI to obtain the low dimensional subspace. Secondly, we group similar image patches together from the HR-MSI to take full advantage of non-local similarities. Finally, the new tensor ring decomposition based on TNN is applied to explore the low-rank prior. The whole process is shown in Fig. 1, which sequentially consists of a subspace decomposition operation, a group matching operation, and the tensor ring strategies based on the tensor nuclear norm.

A. Globally low-rank property

Due to the fact that HSI often has a close correlation among different bands, the globally spectral low-rank property is a dominant prior for HSI super-resolution. Particularly, the pixels located at the same spatial position but with different spectral bands often live in a low dimensional subspace whose ambient size is much less than that of the original one. Therefore, we use the subspace decomposition technology for HR-HSI to reveal the spectral low-rank property, i.e.,

$$\boldsymbol{\mathcal{Z}} = \boldsymbol{\mathcal{A}} \times_3 \boldsymbol{D} \quad (16)$$

where $\boldsymbol{D} \in \mathbb{R}^{S\times L}$, $L << S$ is the orthogonal basis matrix which discovers the common subspace of different spectra and the tensor $\boldsymbol{\mathcal{A}} \in \mathbb{R}^{W\times H\times L}$ is the reduced image, where $\boldsymbol{\mathcal{A}}(i,j,:)$ denotes the coefficient of the spectral pixel $\boldsymbol{\mathcal{Z}}(i,j,:)$. Note that $\boldsymbol{\mathcal{A}} \in \mathbb{R}^{W\times H\times L}$ used here is different from the $\boldsymbol{\mathcal{A}} \in \mathbb{R}^{W\times H\times S}$ used in Subsection II.C. For convenience, we use symbol $\boldsymbol{\mathcal{A}}$ to represent the reduced image in the rest of the paper. Eq. (16) can also be rewritten as

$$\boldsymbol{Z}_{(3)} = \boldsymbol{D}\boldsymbol{A}_{(3)} \quad (17)$$

where $\boldsymbol{A}_{(3)} \in \mathbb{R}^{L\times WH}$ is the unfolded matrix of $\boldsymbol{\mathcal{A}}$ along the third dimension. In this way, the fusion problem can be converted into an learning problem of subspace $\boldsymbol{D}$ and coefficients $\boldsymbol{\mathcal{A}}$ from the given $\boldsymbol{\mathcal{X}}$ and $\boldsymbol{\mathcal{Y}}$. LR-HSI and HR-HSI can be regarded as living in the same spectral subspace, because LR-HSI retains most of the spectral information from HR-HSI. Therefore, we apply singular value decomposition on LR-HSI to obtain the low-dimensional subspace, i.e.,

$$\boldsymbol{X}_{(3)} = \boldsymbol{U}_1\boldsymbol{\Sigma}_1\boldsymbol{V}_1^T \tag{18}$$

where $\boldsymbol{U}_1$ and $\boldsymbol{V}_1$ are column-orthogonal matrices, and $\boldsymbol{\Sigma}_1$ is the diagonal matrix. We can obtain the low-dimensional approximation of $\boldsymbol{X}_{(3)}$ by reserving r largest singular values from $\boldsymbol{\Sigma}_1$, i.e.,

$$\overline{\boldsymbol{X}}_{(3)} = \overline{\boldsymbol{U}}_1\overline{\boldsymbol{\Sigma}}_1\overline{\boldsymbol{V}}_1^T \tag{19}$$

where $\overline{\boldsymbol{U}}_1 = \boldsymbol{U}_1(:, 1:r)$ and $\overline{\boldsymbol{V}}_1 = \boldsymbol{V}_1(:, 1:r)$. Hence, the learned subspace $\boldsymbol{D}$ can be acquired by

$$\boldsymbol{D} = \overline{\boldsymbol{U}}_1 = \boldsymbol{U}_1(:, 1:r) \tag{20}$$

When $\boldsymbol{D}$ is known, the coefficients $\boldsymbol{A}_{(3)}$ can be learned by tackling the following cost function:

$$\min_{\boldsymbol{A}_{(3)}} \left\|\boldsymbol{X}_{(3)} - \boldsymbol{D}\boldsymbol{A}_{(3)}\boldsymbol{B}\boldsymbol{S}\right\|_F^2 + \left\|\boldsymbol{Y}_{(3)} - \boldsymbol{R}\boldsymbol{D}\boldsymbol{A}_{(3)}\right\|_F^2 + \lambda R(\boldsymbol{\mathcal{A}}) \tag{21}$$

where $R(\boldsymbol{\mathcal{A}})$ represents the regularization term on the coefficient $\boldsymbol{\mathcal{A}}$, λ is the tradeoff parameter. The first two are data fidelity terms. As we know, without the regularization term, Eq. (21) may be an ill-posed inverse problem, therefore it is crucial to employ various prior knowledge to shrink the solution space in order to alleviate the indeterminacy of the HSI-MSI fusion task.

B. Low-Rank Tensor Ring Approximation Based on TNN and Non-Local Similarity Prior

The success of the HSI-MSI fusion strongly depends upon whether we select proper prior constraint. From the perspective of data format, we can divide the existing HSI-MSI fusion methods into 1) sparse representation methods based on 1-D vectors, 2) low-rank matrix restoration methods based on 2-D matrices, and 3) low-rank tensor methods based on 3-D tensors. Nowadays, the tensor based HSI-MSI fusion has become a hot topic due to its intrinsically preservation of the original structure. Researchers have investigated a great amount of tensor decomposition algorithms to solve this problem. For instance, Dian *et al*. imposed the tensor-train (TT) rank [32] and subsequently used low tensor multi-rank prior on tensors [28]. Chang *et al*. [29] combined the weighted low-rank tensor prior with the data fidelity term, and optimized the cost function by merits of higher order SVD (HOSVD). On the purpose of substantially preserving the intrinsic correlation of 3D image, we also adopt the tensor form for the third term in Eq. (21).

The discussion in Sec. I shows that tensor ring decomposition owns a more powerful capability to capture the inherent correlation in tensors. The superiority of TR can also be interpreted with the capacity of ensuring more abundant representation. Taking a 4D data $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{I_1\times I_2\times I_3\times I_4}$ as example, Tucker can discover the low-rank prior of one mode against the remains, including matrices of $\boldsymbol{X}[I_1, I_2I_3I_4]$, $\boldsymbol{X}[I_2, I_3I_4I_1]$, $\boldsymbol{X}[I_3, I_4I_1I_2]$, and $\boldsymbol{X}[I_4, I_1I_2I_3]$. TT can factorize $\boldsymbol{\mathcal{X}}$ into a sequence of 3D factors with fixed order, which stresses the low-rank prior of $\boldsymbol{X}[I_1, I_2I_3I_4]$, $\boldsymbol{X}[I_1I_2, I_3I_4]$, and $\boldsymbol{X}[I_1I_2I_3, I_4]$. Benefiting from the trace operation, besides the shown matrices in Tucker and TT, TR is capable of capturing more low-rank prior for matrices such as $\boldsymbol{X}[I_2I_3, I_4I_1]$ and $\boldsymbol{X}[I_4I_1, I_2I_3]$. Note that in most real-world problems, it is empirically verified that low-rank approximation usually reaches state-of-the-art performance when the tensors are converted to more size-balanced matrices [40].

As a result, we select tensor ring decomposition for the subsequent operations. That is, the regularization term $R(\boldsymbol{\mathcal{A}})$ in Eq. (21) can be modeled as $\left(\|\boldsymbol{\mathcal{A}} - \Phi[\boldsymbol{\mathcal{G}}]\|_F^2\right)$. By finely tuning an appropriate tradeoff parameter, the original fusion problem can be formulated as follows.

$$\begin{aligned}\min_{\boldsymbol{A}_{(3)}} \left\|\boldsymbol{X}_{(3)} - \boldsymbol{D}\boldsymbol{A}_{(3)}\boldsymbol{B}\boldsymbol{S}\right\|_F^2 + \left\|\boldsymbol{Y}_{(3)} - \boldsymbol{R}\boldsymbol{D}\boldsymbol{A}_{(3)}\right\|_F^2 + \\ \lambda\left(\|\boldsymbol{\mathcal{A}} - \Phi[\boldsymbol{\mathcal{G}}]\|_F^2\right)\end{aligned} \tag{22}$$

The most original intuition in Eq. (22) is to minimize the TR rank finding the low-rank structure of the data and approximating the recovered tensor. Note that Ref. [35] and [39] have imposed matrix nuclear norm on the TR factors. Although acceptable restoration performance can be achieved , they all unfold the tensor core $\boldsymbol{\mathcal{G}}$ to mode-n matrix, which may result in certain loss of optimality in the representation. In order to maintain the original high-order structural correlation, we instead impose the tensor nuclear norm on each of the TR factors. Fig. 1 shows the tensor-ring decomposition based on tensor nuclear norm. A TR factor, e.g., $\boldsymbol{\mathcal{G}}^{(2)} \in \mathbb{R}^{R_2\times I_2\times R_3}$, is penalized by the tensor nuclear norm regularization. It is worth noting that Fig. 1 only shows the process of imposing the tensor nuclear norm on one TR factor. In fact, we have carried out tensor nuclear norm constraint on all TR factors. Based on this setting, Eq. (22) can be rewritten as follows.

$$\begin{aligned}\min_{\boldsymbol{A}_{(3)}} \left\|\boldsymbol{X}_{(3)} - \boldsymbol{D}\boldsymbol{A}_{(3)}\boldsymbol{B}\boldsymbol{S}\right\|_F^2 + \left\|\boldsymbol{Y}_{(3)} - \boldsymbol{R}\boldsymbol{D}\boldsymbol{A}_{(3)}\right\|_F^2 + \\ \left(\lambda\|\boldsymbol{\mathcal{A}} - \Phi[\boldsymbol{\mathcal{G}}]\|_F^2 + \eta\sum_{n=1}^{N}\left\|\boldsymbol{\mathcal{G}}^{(n)}\right\|_{TNN}\right)\end{aligned} \tag{23}$$

where η is a tradeoff parameter, $\|\cdot\|_{TNN}$ is the TNN regularization.

Spatially non-local self-similarity is also a mighty property which can utilize not only the neighborhood entries but also the far away entries with similar textures or appearances. Therefore, we impose this prior on our model to capture the intrinsic sparsity configurations. However, the HR-HSI is unknown, we cannot directly group its patches. To solve this problem, we group similar patches of HR-HSI based on the learned spatial locations from the HR-MSI. For more details, we divide the reduced image $\boldsymbol{\mathcal{A}}$ into overlapping 3D patches to capture the local information of HR-MSI. For each patch $\boldsymbol{\mathcal{B}}$, we search the M (including $\boldsymbol{\mathcal{B}}$ itself) most similar 3D patches, measured by the Euclidean distance, in

a relatively larger window. Then, by stacking the similar patches in the i-th exemplar patch together, a third-order tensor $\boldsymbol{\mathcal{A}}_i \in \mathbb{R}^{P^2 \times S \times M}$ can be formed, where P denotes the patch size. Mathematically, we represent $\boldsymbol{\mathcal{A}}_i = \mathcal{R}_i \boldsymbol{\mathcal{A}}$ as the extracted 3D tensor for each exemplar cubic at position i. Hence, the cost function of LRTRTNN can further be regarded as

$$\min_{\boldsymbol{A}_{(3)}} \left\| \boldsymbol{X}_{(3)} - \boldsymbol{D}\boldsymbol{A}_{(3)}\boldsymbol{B}\boldsymbol{S} \right\|_F^2 + \left\| \boldsymbol{Y}_{(3)} - \boldsymbol{R}\boldsymbol{D}\boldsymbol{A}_{(3)} \right\|_F^2 + \sum_i \left(\lambda \left\| \mathcal{R}_i \boldsymbol{\mathcal{A}} - \Phi[\boldsymbol{\mathcal{G}}_i] \right\|_F^2 + \eta \sum_{n=1}^{N} \left\| \boldsymbol{\mathcal{G}}_i^{(n)} \right\|_{TNN} \right) \tag{24}$$

In order to directly compare the practical performance of applying different low-rank constraint on TR factors, we respectively exhibit the experimental results of imposing matrix nuclear norm, tensor nuclear norm, and the ones without any regularization in Fig. 2. It can be seen from this figure that imposing nuclear norm constraint on TR factors can greatly improve the fusion effect. Moreover, the PSNR values of all bands from tensor nuclear norm are all higher than those from matrix nuclear norm, which demonstrates that tensor nuclear norm constraint can faithfully retain the intrinsic information of the tensor. Therefore, we choose to impose tensor nuclear norm constraint on the model for better performance.

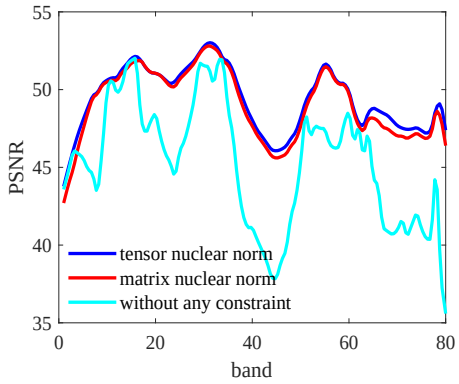


Fig. 2. PSNR values along with different spectral bands on Pavia city center dataset. The blue, red, and cyan lines are regarding to tensor nuclear norm, matrix nuclear norm, and the one without any constraint, respectively.

Compared with conventional MSI-HSI fusion model, the proposed LRTRTNN has two evident advantages. On the one hand, it learns the low dimensional subspace from HSI to probe into the global spectral low-rank prior and extracts non-local similar 3D patches from the reduced image to make full use of spatial self-similarity. On the other hand, the format of tensor ring approximation based on TNN can more faithfully capture the structural correlation of the learned tensor.

C. Optimization Procedure

Although the cost function (24) is an unconstrained problem, it cannot be directly optimized due to the hybrid usage of extraction operator $\mathcal{R}_i$, the tensor ring decomposition $\Phi[\boldsymbol{\mathcal{G}}_i]$, and the TNN constraint. We introduce the well-known ADMM algorithm [44] to solve this model. By adopting the variable $\boldsymbol{\mathcal{J}} = \boldsymbol{\mathcal{A}}$, $\{\boldsymbol{\mathcal{M}}_i^n\}_{n=1}^N = \{\boldsymbol{\mathcal{G}}_i^n\}_{n=1}^N$, the following augmented Lagrangian function shall be acquired:

$$\begin{aligned} &\mathcal{L}(\boldsymbol{\mathcal{A}}, \boldsymbol{\mathcal{J}}, \boldsymbol{\mathcal{P}}_1, \boldsymbol{\mathcal{P}}_2, \boldsymbol{\mathcal{M}}, \boldsymbol{\mathcal{G}}) = \\ &\left\| \boldsymbol{X}_{(3)} - \boldsymbol{D}\boldsymbol{A}_{(3)}\boldsymbol{B}\boldsymbol{S} \right\|_F^2 + \left\| \boldsymbol{Y}_{(3)} - \boldsymbol{R}\boldsymbol{D}\boldsymbol{A}_{(3)} \right\|_F^2 \\ &+ \tfrac{\mu_1}{2} \left\| \boldsymbol{\mathcal{J}} - \boldsymbol{\mathcal{A}} + \tfrac{\boldsymbol{\mathcal{P}}_1}{\mu_1} \right\|_F^2 + \lambda \sum_i \left(\left\| \mathcal{R}_i \boldsymbol{\mathcal{J}} - \Phi[\boldsymbol{\mathcal{G}}_i] \right\|_F^2 \right) + \\ &\sum_i \sum_{n=1}^{N} \left(\eta \left\| \boldsymbol{\mathcal{M}}_i^{(n)} \right\|_{TNN} + \tfrac{\mu_2}{2} \left\| \boldsymbol{\mathcal{M}}_i^{(n)} - \boldsymbol{\mathcal{G}}_i^{(n)} + \tfrac{\boldsymbol{\mathcal{P}}_{2(i)}^{(n)}}{\mu_2} \right\|_F^2 \right) \\ &s.t. \boldsymbol{\mathcal{J}} = \boldsymbol{\mathcal{A}}, \{\boldsymbol{\mathcal{M}}_i^n\}_{n=1}^N = \{\boldsymbol{\mathcal{G}}_i^n\}_{n=1}^N \end{aligned} \tag{25}$$

where $\boldsymbol{\mathcal{P}}_1$ and $\boldsymbol{\mathcal{P}}_2$ are the Lagrangian multipliers, μ_1 and μ_2 are penalty parameters. On that basis, we can separate the minimization of problem (25) into several solvable subproblems:

1) For $\boldsymbol{A}_{(3)}$-subproblem:

$$\begin{aligned} \boldsymbol{A}_{(3)} &\in \underset{\boldsymbol{A}_{(3)}}{\operatorname{argmin}} \mathcal{L}(\boldsymbol{\mathcal{A}}, \boldsymbol{\mathcal{J}}, \boldsymbol{\mathcal{P}}_1) \\ &= \underset{\boldsymbol{A}_{(3)}}{\operatorname{argmin}} \left\| \boldsymbol{X}_{(3)} - \boldsymbol{D}\boldsymbol{A}_{(3)}\boldsymbol{B}\boldsymbol{S} \right\|_F^2 \\ &\quad + \left\| \boldsymbol{Y}_{(3)} - \boldsymbol{R}\boldsymbol{D}\boldsymbol{A}_{(3)} \right\|_F^2 + \tfrac{\mu_1}{2} \left\| \boldsymbol{J}_{(3)} - \boldsymbol{A}_{(3)} + \tfrac{\boldsymbol{P}_{1(3)}}{\mu_1} \right\|_F^2 \end{aligned} \tag{26}$$

where $\boldsymbol{P}_{1(3)}$ and $\boldsymbol{J}_{(3)}$ are the unfolded matrices of $\boldsymbol{\mathcal{P}}_1$ and $\boldsymbol{\mathcal{J}}$ along the third dimension, respectively. Due to the strongly convexity of Eq. (26), the solution of $\boldsymbol{A}_{(3)}$ has a unique solution. Specifically, we can obtain a Sylvester-type equation by forcing the derivative of (26) on $\boldsymbol{A}_{(3)}$ to be zero.

$$\boldsymbol{H}_1 \boldsymbol{A}_{(3)} + \boldsymbol{A}_{(3)} \boldsymbol{H}_2 = \boldsymbol{H}_3 \tag{27}$$

where $\boldsymbol{H}_1 = \left[(\boldsymbol{R}\boldsymbol{D})^T \boldsymbol{R}\boldsymbol{D} + \mu_1 \boldsymbol{I}_L \right]$ is positive definite whose eigenvalues are all positive values with $\boldsymbol{I}_L \in \mathbb{R}^{L \times L}$ being identity matrix, $\boldsymbol{H}_2 = (\boldsymbol{B}\boldsymbol{S})(\boldsymbol{B}\boldsymbol{S})^T$ is semipositive, and $\boldsymbol{H}_3 = (\boldsymbol{R}\boldsymbol{D})^T \boldsymbol{Y}_{(3)} + \frac{\mu_1}{2} \left(\boldsymbol{J}_{(3)} + \frac{\boldsymbol{P}_{1(3)}}{\mu_1} \right) + \boldsymbol{D}^T \boldsymbol{X}_{(3)} (\boldsymbol{B}\boldsymbol{S})^T$. By employing the properties of $\boldsymbol{B}$ and $\boldsymbol{S}$, we can solve Eq. (27) in a closed-from solution as described in [28], [45]. The practical procedure is provided in Algorithm 1.

Algorithm 1: The solution of (27) for $\boldsymbol{A}_{(3)}$

Input: $\boldsymbol{H}_1$, $\boldsymbol{H}_2$, $\boldsymbol{H}_3$, $\boldsymbol{B}$, $\boldsymbol{S}$.
Output: $\boldsymbol{A}_{(3)}$.

```
Eigen-decomposition of B: B = F K F^{-1};
K̃ = K (1_d ⊗ 1_{wh});
Eigen-decomposition of H_1: H_1 = Q_1 Λ Q_1^{-1};
H̃_3 = Q_1^{-1} H_3 F;
for l= 1 to L do
    o_l = λ_l^{-1} (H̃_3)_l − λ_l^{-1} (H̃_3)_l K̃ (λ_l d I_n + Σ_{t=1}^{d} K_t^2) K̃^T;
end for;
Set A_(3) = Q_1 O F^T;
```

In Algorithm 1, $\mathbf{1}_d \in \mathbb{R}^d$ represent a vector with its elements all ones, and d denotes the spatial degradation factor. The diagonal matrix $\boldsymbol{K} \in \mathbb{R}^{WH \times WH}$ decomposed from the blur

matrix $\boldsymbol{B}$ is written as

$$\boldsymbol{K} = \begin{bmatrix} \boldsymbol{K}_1 & 0 & \cdots & 0 \\ 0 & \boldsymbol{K}_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \boldsymbol{K}_d \end{bmatrix} \tag{28}$$

We can acquire the unitary matrix $\boldsymbol{Q}_1$ and the diagonal matrix $\boldsymbol{\Lambda}$ by the eigen-decomposition of $\boldsymbol{H}_1$, i.e., $\boldsymbol{H}_1 = \boldsymbol{Q}_1\boldsymbol{\Lambda}\boldsymbol{Q}_1^{-1}$, where the columns of $\boldsymbol{Q}_1$ hold eigenvectors of $\boldsymbol{H}_1$ and $\boldsymbol{\Lambda}$ can be written as

$$\boldsymbol{\Lambda} = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_L \end{bmatrix} \tag{29}$$

$\boldsymbol{O}$ can be computed as $\boldsymbol{O} = \boldsymbol{Q}^{-1}\boldsymbol{A}_{(3)}\boldsymbol{F} = \left[\boldsymbol{o}_1^T, \boldsymbol{o}_2^T, \ldots, \boldsymbol{o}_L^T\right]^T$, from which we can know that $\boldsymbol{o}_l$ is the l-th row of $\boldsymbol{O}$.

2) For $\boldsymbol{\mathcal{G}}_i^{(n)}$-subproblem: According to the work in [37], assuming the size of each group is $\mathcal{R}_i\boldsymbol{\mathcal{J}} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$ and for the n-th core tensor $\boldsymbol{\mathcal{G}}^{(n)}$, we have $\mathcal{R}_i\boldsymbol{J}_{<n>} = \boldsymbol{G}_{i,(2)}^{(n)}\boldsymbol{G}_{i,<2>}^{(\neq n)}$, where $\mathcal{R}_i\boldsymbol{J}_{<n>} \in \mathbb{R}^{I_n \times \prod_{i=1, i\neq n}^{N} I_i}$ denotes mode-n matricization of $\mathcal{R}_i\boldsymbol{\mathcal{J}}$, $\boldsymbol{G}_{<2>}^{(\neq n)} \in \mathbb{R}^{\prod_{i=1, i\neq n}^{N} I_i \times R_n R_{n+1}}$ is a mode-2 unfolded matrix generated by merging all the factor matrices except the n-th core, and $\boldsymbol{G}_{(2)}^{(n)} \in \mathbb{R}^{I_n \times R_n R_{n+1}}$ is a n-th unfolded matrix. Because each of the core tensors is independent, we can optimize them independently. By using Eq. (25), the subproblem w.r.t. $\boldsymbol{\mathcal{G}}_i^{(n)}$ can be formed as

$$\begin{aligned} \min_{\boldsymbol{\mathcal{G}}_i^{(n)}} \left(\lambda \left\| \mathcal{R}_i\boldsymbol{J}_{<n>} - \left(\boldsymbol{G}_{i,(2)}^{(n)}\right)\left(\boldsymbol{G}_{i,<2>}^{(\neq n)}\right)^T \right\|_{\mathrm{F}}^2 \right) + \\ \left(\frac{\mu_2}{2} \left\| \boldsymbol{\mathcal{M}}_i^{(n)} - \boldsymbol{\mathcal{G}}_i^{(n)} + \frac{\boldsymbol{\mathcal{P}}_{2(i)}^{(n)}}{\mu_2} \right\|_F^2 \right) \end{aligned} \tag{30}$$

Eq. (30) is a least square problem, which has a closed-form solution as follows:

$$\boldsymbol{G}_{i,(2)}^{(n)} = \begin{array}{l} \left(\left(\mu_2\boldsymbol{M}_{i,(2)}^{(n)} + \boldsymbol{P}_{2[i,(2)]}^{(n)} \right) + \lambda\mathcal{R}_i\ \boldsymbol{J}_{<n>}\boldsymbol{G}_{i,<2>}^{(\neq n)} \right) \\ \left(\lambda\left(\boldsymbol{G}_{i,<2>}^{(\neq n)}\right)^T \left(\boldsymbol{G}_{i,<2>}^{(\neq n)}\right) + \mu_2\boldsymbol{I} \right)^{-1} \end{array} \tag{31}$$

where $\boldsymbol{I}$ is an identity matrix. Since the solution form of Eq. (31) is a matrix, a folding operation is performed to convert $\boldsymbol{G}_{i,(2)}^{(n)}$ into a tensor form $\boldsymbol{\mathcal{G}}_i^{(n)}$.

3) For $\boldsymbol{\mathcal{J}}$-subproblem: The subproblem of $\boldsymbol{\mathcal{J}}$ is

$$\begin{aligned} \boldsymbol{\mathcal{J}} \in \operatorname*{argmin}_{\boldsymbol{\mathcal{J}}} \boldsymbol{\mathcal{L}}(\boldsymbol{\mathcal{A}}, \boldsymbol{\mathcal{J}}, \boldsymbol{\mathcal{P}}_1, \boldsymbol{\mathcal{G}}) = \lambda \sum_i \|\mathcal{R}_i\boldsymbol{\mathcal{J}} - \Phi[\boldsymbol{\mathcal{G}}_i]\|_F^2 + \\ \frac{\mu_1}{2} \left\| \boldsymbol{\mathcal{J}} - \boldsymbol{\mathcal{A}} + \frac{\boldsymbol{\mathcal{P}}_1}{\mu_1} \right\|_F^2 \end{aligned} \tag{32}$$

The quadratic optimization problem (32) has an analytical solution, provided by

$$\boldsymbol{\mathcal{J}} = (\mu_1\boldsymbol{I} + \lambda\sum_i \mathcal{R}_i^T\mathcal{R}_i)^{-1}(\lambda\sum_i \mathcal{R}_i^T\Phi[\boldsymbol{\mathcal{G}}_i] + \mu_1\boldsymbol{\mathcal{A}} - \boldsymbol{\mathcal{P}}_1) \tag{33}$$

where $\lambda\sum_i \mathcal{R}_i^T\mathcal{R}_i$ denotes the number of overlapping cubics that cover the pixel location, and $\lambda\sum_i \mathcal{R}_i^T\Phi[\boldsymbol{\mathcal{G}}_i]$ means the sum value of all overlapping reconstruction cubics that cover the pixel location.

4) For $\boldsymbol{\mathcal{M}}_i$-subproblem: By keeping other variables fixed, we can obtain $\boldsymbol{\mathcal{M}}_i$ by

$$\min_{\boldsymbol{\mathcal{M}}_i} \sum_{n=1}^{N} \left(\eta \left\| \boldsymbol{\mathcal{M}}_i^{(n)} \right\|_{TNN} + \frac{\mu_2}{2} \left\| \boldsymbol{\mathcal{M}}_i^{(n)} - \boldsymbol{\mathcal{G}}_i^{(n)} + \frac{\boldsymbol{\mathcal{P}}_{2(i)}^{(n)}}{\mu_2} \right\|_F^2 \right) \tag{34}$$

Eq. (34) also has a closed-form solution by adopting the tensor singular value thresholding (t-SVT) operator [43]

$$\boldsymbol{\mathcal{M}}_i^{(n)} = \mathrm{SVT}_{\frac{1}{\eta}}(\boldsymbol{\mathcal{G}}_i^{(n)} - \frac{\boldsymbol{\mathcal{P}}_{2(i)}^{(n)}}{\mu_2}) \tag{35}$$

The t-SVT operator is given as:

$$\mathrm{SVT}_\tau(\boldsymbol{\mathcal{L}}) = \boldsymbol{\mathcal{U}} * \boldsymbol{\mathcal{S}}_\tau * \boldsymbol{\mathcal{V}}^T \tag{36}$$

where $\boldsymbol{\mathcal{S}}_\tau = \mathrm{ifft}\left((\overline{\boldsymbol{\mathcal{S}}} - \tau)_+, [], 3\right)$ with t_+ representing the positive part of t, i.e., $t_+ = \max(t, 0)$. In other words, t-SVT simply adopts a soft-thresholding rule to the SVs of $\overline{\boldsymbol{\mathcal{S}}}$ that computed from the frontal slices of $\boldsymbol{\mathcal{G}}_i^{(n)} - \frac{\boldsymbol{\mathcal{P}}_{2(i)}^{(n)}}{\mu_2}$.

5) Update Lagrangian multipliers $\boldsymbol{\mathcal{P}}_1$ and $\boldsymbol{\mathcal{P}}_2$: According to ADMM, the multipliers can be updated as:

$$\begin{aligned} \boldsymbol{\mathcal{P}}_1 &= \boldsymbol{\mathcal{P}}_1 + \mu_1(\boldsymbol{\mathcal{J}} - \boldsymbol{\mathcal{A}}) \\ \boldsymbol{\mathcal{P}}_{2(i)}^n &= \boldsymbol{\mathcal{P}}_{2(i)}^n + \mu_2(\boldsymbol{\mathcal{M}}_i^n - \boldsymbol{\mathcal{G}}_i^n) \end{aligned} \tag{37}$$

The whole procedure to solve the proposed LRTRTNN method is given in Algorithm 2.

D. Computational Complexity

For the acquired LR-HSI $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{w \times h \times S}$ and HR-MSI $\boldsymbol{\mathcal{Y}} \in \mathbb{R}^{W \times H \times s}$, the per-iteration cost of each group mainly lies in the subproblems of updating $\boldsymbol{\mathcal{M}}^n$ and $\boldsymbol{\mathcal{G}}^n$. For simplicity, assuming the size of each group is $\mathcal{R}_i\boldsymbol{\mathcal{J}} \in \mathbb{R}^{I \times I \times I}$ and the TR rank is set as $R_1 = R_2 = R_3 = R$, the cost of updating $\boldsymbol{\mathcal{G}}^n$ in each iteration is $O(R^6 + I^2R^4 + I^3R^2)$. The computation of $\boldsymbol{\mathcal{M}}^n$ is dominated by the FFT transformation and $\lceil \frac{R+1}{2} \rceil$ SVDs of $R \times I$ matrices, which costs $O(R^2 I log R + I^2R^2)$ and is evidently less than that for $\boldsymbol{\mathcal{G}}^n$. Unfortunately, for images with relatively larger spatial size, the computational cost of the algorithm is very high. For computation alleviation, our algorithm can be performed in a parallel way for different groups. Moreover, subspace decomposition shrinks the spectral dimensionality of HSI, which could further reduce the computational cost.

TABLE I
SETUP OF THE PARAMETERS IN THE EXPERIMENTS

	NSSR	CSTF	UTV	STEREO	SC-LL1	LTMR	LRTRTNN
Sub-dimension L	/	/	/	/	/	10	10
Group size M	/	/	/	/	/	100	150
Patch size P	/	/	/	/	/	7	6
Step size s	/	/	/	/	/	4	3
Tradeoff parameter	[0.015,10^{-4}]	10^{-5}	[10^{-6}, 10^{-8}, 10^{-7}, 10^{3}]	0.8	[0.005, 10^{-5}, 10^{-4}]	[0.03, 10^{-3}]	[1, 0.007]
Dictionary atoms	80	[300, 300, 12]	[300, 300, 12]	/	/	/	/

Algorithm 2: ADMM for solving (24)

Input: HR-MSI $\boldsymbol{\mathcal{Y}}$; LR-HSI $\boldsymbol{\mathcal{X}}$; tradeoff parameter λ and η; Randomly sampled $\boldsymbol{\mathcal{G}}^n$ by distribution $N \sim (0, 1)$; patch size P; step size s; TR Rank R; maximum iteration T.

Output: HR-HSI $\boldsymbol{\mathcal{Z}}$

Initialization:
$\lambda = 1, \eta = 0.7, R = 11, T = 30, \mu_1 = \mu_2 = 10^{-3}, P, s$, iter=0;
learn the orthogonal basis matrix $\boldsymbol{D}$ from $\boldsymbol{X}_{(3)}$ via SVD;
while *not converged* **do**
  iter=iter+1;
  Update $\boldsymbol{A}_{(3)}$ via (26);
  For $n = 1, \cdots, N$, Update $\boldsymbol{\mathcal{G}}^n$ via (31);
  Update $\boldsymbol{\mathcal{J}}$ via (33);
  For $n = 1, \cdots, N$, Update $\boldsymbol{\mathcal{M}}^n$ via (35);
  Update Lagrangian multipliers $\boldsymbol{\mathcal{P}}_1$ and $\boldsymbol{\mathcal{P}}_2$ via (37);
end
$\boldsymbol{\mathcal{Z}} = \boldsymbol{\mathcal{A}} \times_3 \boldsymbol{D}$.

IV. Experiments

We conduct the comparisons between LRTRTNN with several state-of-the-art HSI-MSI fusion methods, including NSSR (nonnegative structured sparse representation) [20], CSTF (coupled sparse tensor factorization) [26], LTMR (subspace based low tensor multi-rank regularization) [28], UTV(Unidirectional Total Variation) [30], STEREO(Super-resolution Tensor reconstruction) [31], and SC-LL1 (structured coupled LL1 decomposition) [4]. Among the compared approaches, NSSR follows the matrix factorization mode, while CSTF, LTMR, UTV, STEREO, and SC-LL1 are all tensor factorization based methods. All the undetermined parameters for the compared methods are optimally tuned to report their best results. Unless otherwise specified, the used parameter setting is given in Table I. The experiments are run using MATLAB R2020a in Windows 10 and on hardware platform with AMD R5-3600 and 16-GB memory.

A. Description of Data Sets

We employ four data sets to verify the performance of LRTRTNN. The first dataset is Cuprite Mine in Nevada, which was obtained by the NASAs Airborne Visible and Infrared Imaging Spectrometer (AVIRIS) [46]. The original size of the image is 256 × 256 × 224, and it covers the wavelength from 400nm to 2500nm with 10nm interval. For reasons connected with the downsampling process, the used HSI is resized to be with 200 × 200 × 103 tensor.

The second one is University of Pavia image obtained by the reflective optics system imaging spectrometer (ROSIS) optical airborne sensor over the area of the University of Pavia, Italy [47]. Since some spectral bands of this dataset are heavily contaminated by noise, we firstly remove a few bands of this data. The finally used HSI has been resized to 256 × 256 × 93.

The third dataset is Pavia city center dataset collected by the reflective optics system imaging spectrometer (ROSIS-03)[1]. The HSI consists of 610 × 340 pixels and 93 spectral bands and we resize this HSI to 200 × 200 × 80 in our experiments.

The fourth one is Washington DC Mall dataset collected by the Hyperspectral Digital Imagery Collection Experiment (HYDICE) [48][2]. The original image contains 1208 × 307 pixels and 191 spectral bands. We use a subimage of size 256 × 256 ×103 in the experiment.

To build the whole process of MSI-HSI fusion, we first apply IKONOS-like spectral reflectance response filter [49] to obtain the four-band HR-MSIs of the size 200 × 200 × 4, 256 × 256 × 4, 200 × 200 × 4, and 256 × 256 × 4, respectively. Afterwards, a 7 × 7 Gaussian blur with standard deviation 2 is filtered to each band of HR-HSI, followed by a downsampling operator mapping every four pixels to two spatial modes for all data sets.

B. Quantitative Metrics

Five quantitative metrics, including mean peak signal-to-noise ratio (MPSNR), mean structural similarity (MSSIM) [50], spectral angle mapper (SAM), the relative dimensionless global error in synthesis (ERGAS), and degree of distortion (DD) are utilized to measure the comprehensive quality of the generated results. They can be respectively defined as follows:

$$\mathrm{MPSNR} = \frac{1}{S}\sum_{i}^{S} 10 \times \lg \frac{\|\boldsymbol{A}_i\|_\infty^2}{\frac{1}{WH}\left\|\widehat{\boldsymbol{A}}_i - \boldsymbol{A}_i\right\|_F^2}, \tag{38}$$

$$\mathrm{MSSIM} = \frac{1}{S}\sum_{i}^{S} \frac{(2\mu_{\boldsymbol{A}_i}\mu_{\widehat{\boldsymbol{A}}_i} + C_1)(2\sigma_{\boldsymbol{A}_i\widehat{\boldsymbol{A}}_i} + C_2)}{(\mu_{\boldsymbol{A}_i}^2 + \mu_{\widehat{\boldsymbol{A}}_i}^2 + C_1)(\sigma_{\boldsymbol{A}_i}^2 + \sigma_{\widehat{\boldsymbol{A}}_i}^2 + C_2)}, \tag{39}$$

[1] http://www.ehu.eus/ccwintco/index.php
[2] https://engineering.purdue.edu/ biehl/MultiSpec/hyperspectral.html

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JSTARS.2021.3108233, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing

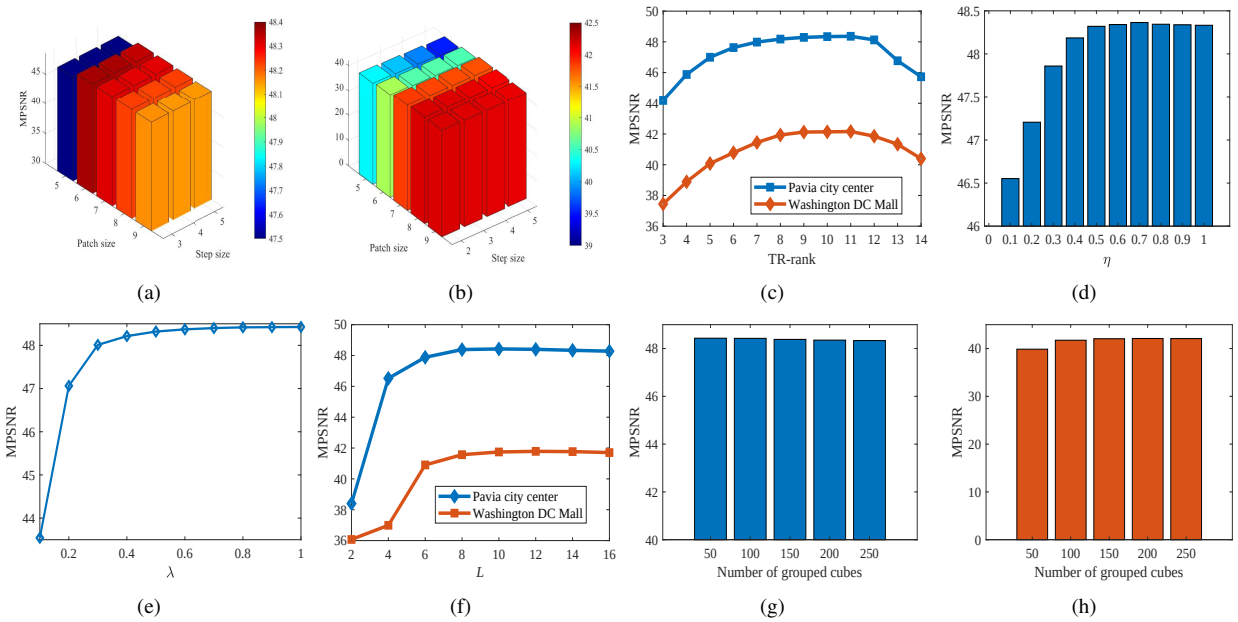


Fig. 3. PSNR results against different selections of (a) patch size P and step size s on Pavia city center, (b) patch size P and step size s on Washington DC Mall, (c) tensor ring rank R, (d) tradeoff parameter η, (e) tradeoff parameter λ, (f) subspace dimension L, (g) group size on Pavia city center, and (h) group size on Washington DC Mall.

$$\mathrm{SAM}(\boldsymbol{\mathcal{A}}, \widehat{\boldsymbol{\mathcal{A}}}) = \frac{1}{WH}\sum_{i=1}^{WH} \arccos \frac{\widehat{\boldsymbol{a}}_i^T \mathbf{a}_i}{\|\widehat{\mathbf{a}}_i\|_2 \|\mathbf{a}_i\|_2} \tag{40}$$

$$\mathrm{ERGAS}(\boldsymbol{\mathcal{A}}, \widehat{\boldsymbol{\mathcal{A}}}) = \frac{100}{d}\sqrt{\frac{1}{S}\sum_{i=1}^{S}\frac{\mathrm{MSE}\left(\boldsymbol{A}_i, \widehat{\boldsymbol{A}_i}\right)}{\mu^2_{\left(\widehat{\boldsymbol{A}_i}\right)}}}, \tag{41}$$

$$\mathrm{DD}(\boldsymbol{\mathcal{A}}, \widehat{\boldsymbol{\mathcal{A}}}) = \frac{1}{WHS}\|\mathrm{vec}(\boldsymbol{\mathcal{A}}) - \mathrm{vec}(\widehat{\boldsymbol{\mathcal{A}}})\|_1 \tag{42}$$

where $\boldsymbol{A}_i$ and $\widehat{\boldsymbol{A}_i}$ denote the i-th spectra of $\boldsymbol{\mathcal{A}_i} \in \mathbb{R}^{W\times H\times S}$ and the generated result $\widehat{\boldsymbol{\mathcal{A}}} \in \mathbb{R}^{W\times H\times S}$, respectively. C_1 and C_2 are constants, $\sigma_{\boldsymbol{A}_i\widehat{\boldsymbol{A}}_i}$ denotes the covariance matrix between $\boldsymbol{A}_i$ and $\widehat{\boldsymbol{A}_i}$, $\mu_{\boldsymbol{A}_i}$ and $\mu_{\widehat{\boldsymbol{A}_i}}$ represent the average values, $\sigma_{\boldsymbol{A}_i}$ and $\sigma_{\widehat{\boldsymbol{A}_i}}$ are the standard deviation of $\boldsymbol{A}_i$ and $\widehat{\boldsymbol{A}_i}$, respectively. d represents the spatial downsampling factor (DF), $\mathrm{MSE}(\cdot)$ is abbreviated for the mean square error, $\mu_{(\cdot)}$ is the symbol to compute the mean value, and $\mathbf{a}_i$ and $\hat{\mathbf{a}}_i$ are the ground truth pixel and the estimated one, respectively.

C. Parameters Analysis

In our method, several key parameters, i.e., patch size P, step size s, tensor ring rank R, tradeoff parameter η and λ, subspace dimension L, and the number of grouped cubes M, need to be pre-determined in practical usage. Table I lists the key parameter values in the experiment.

1) The influence of patch size P and step size s: Fig. 3(a) and Fig. 3(b) report the quantitative metrics (MPSNR) against different patch sizes and step sizes on datasets Pavia city center and Washington DC Mall, respectively. The detailed performance difference is compared through different depths of the presented colors. It is well known that different patch sizes and step sizes perform inconsistently to different datasets. Given a fixed step size, our first observation from Fig. 3(a) and Fig. 3(b) is that the performance fluctuates slightly against varying patch sizes. Specifically, the best fusion results lie in $P = 6$ and $P = 8$ on Pavia city center and Washington DC Mall, respectively. Nevertheless, while $P = 6$ is clearly the best setting for Pavia city center, the result on Washington DC Mall from $P = 9$ is very close to that from $P = 8$. By taking the running efficiency into account, $P = 8$ is decided to be our final setting since it achieves well-balanced compromise between patch size and patch number, which further makes it run faster than $P = 9$. As for the step size, we can see that it has lower impact on the fusion performance. In practice, a smaller step size would lead to much more patches, which further results in better performance due to deeper utilization of spatial information. However, while the performance gain is trivial, the running efficiency would evidently cut down. For a well balance, we set $s = 3$ for Pavia city center and $s = 4$ for Washington DC Mall. Since Cuprite Mine (Resp. University of Pavia) shares the spatial dimension with Pavia city center (Resp. Washington DC Mall), we simply set their P, s pair as [6,3] and [8,4], respectively. Note again that the best pair values are inconsistent concerning different datasets. For practical usage, we suggest a preliminary attempt on pairs [6,3] and [8,4]. Then the values around the winner can further be tried for a better visual assessment.

2) Discussion of TR Rank R: R is an important parameter since it determines how low rank the recovered tensors would be. While a loose approximation using higher rank values is relatively easier to determine, it would result in accuracy loss and incremental computation due to involving of redundant information. Conversely, when a much lower rank value is

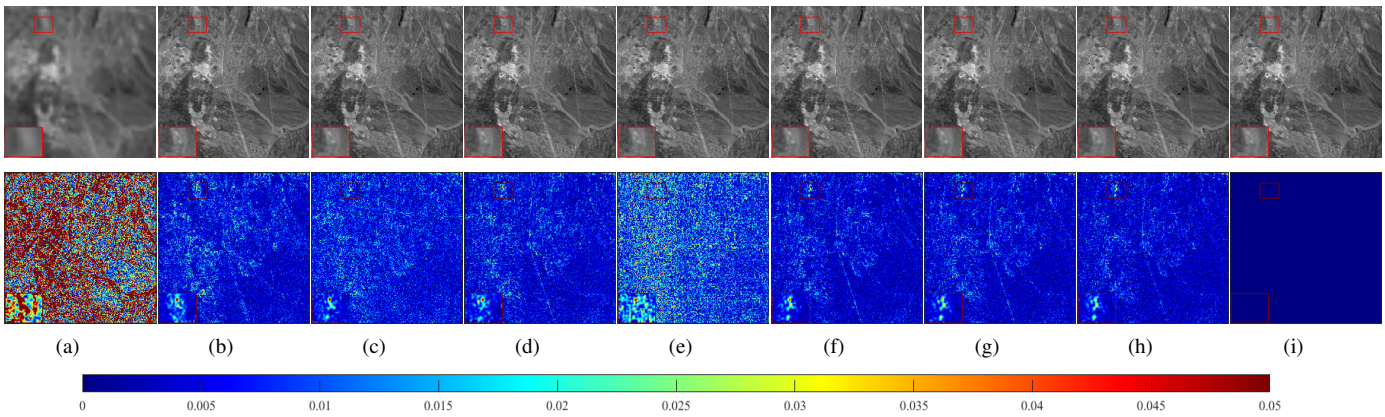


Fig. 4. The upper row gives the fused images by the competing methods for Cuprite Mine at band 20. The lower row gives the error images of the test methods for Cuprite Mine at band 20.(a) LR-HSI. (b) NSSR. (c) CSTF. (d) UTV. (e) STEREO. (f) SCL11. (g) LTMR. (h) LRTRTNN. (i) Ground truth.

given, the performance degradation would be also remarkable due to loss of the intrinsic information. For simplicity, we assume the TR ranks for different factors are equal, i.e., $R_1 = R_2 = \cdots = R_N$. To seek a relatively tighter representation, in Fig. 3(c), we illustrate the variations of MPSNR values against with different settings of TR ranks on Pavia city center and Washington DC Mall. From this figure, one can observe that the experimental results have a sharp rise as parameter R varies from 2 to 11. However, when the peak value is surpassed, the performance declines rapidly as R grows further. The clear trend on all datasets would evidently relieve the tedious searching process of parameter R. In our experiments, we set R as 11 due to its stability.

3) The influence of tradeoff parameters η and λ: η and λ directly affect the solution of (35) and (31), respectively. Given the searching span $\eta \in [0.1, 1) * 10^{-3}$, Fig. 3(d) shows the generated results of LRTRTNN on Pavia city center dataset. Note that when $\eta = 0$, the TNN penalty is eliminated from the cost function, which turns into an ill-posed inverse problem and the solutions are infinite. Conversely, when one set $\eta = \infty$, then the fidelity term would contribute little to the final solution. From Fig. 3(d), one can check that the generated results are comparatively stable in the given span. Although there would be certain fluctuations in efficacy and efficiency in different scenarios or on various target data, $\eta = 0.7 * 10^{-3}$ is suggested as the default value due to the appealing performance with this setting in all our experiments. The parameter λ controls the relative weight of tensor ring decomposition. From Fig. 3(e), it can be observed that the values of MPSNR get higher with the increasing λ on Pavia city center dataset. However, when λ continues to increase, the growth rate of MPSNR slows down, and the best results are achieved when $\lambda = 1$. Based on these results, the empirical setting $\lambda = 1$ is used in all experiments.

4) Analysis on different choices of atoms L: L controls the reduced spatial dimension. In Fig. 3(f), we plot the MPSNR curves against different L on Pavia city center and Washington DC Mall. As can be seen, the MPSNR values have a rising trend as L changes from 2 to 10. When L is greater than 10, the performance tends to be stable and the value variation is trivial. This phenomenon is reasonable since that a smaller dimension is insufficient to support the subspace range. At least 10 atoms should be retained to preserve the spectral information for the used data.

5) Analysis of group size M: The results of MPSNR values with M being searched in $\{50, 100, 150, 200, 250\}$ are shown in Fig. 3(g) and (h) on Pavia city center and Washington DC Mall, respectively. In both of these two subfigures, the performance is relatively stable along with different settings of parameter M. Thus, the patch number is empirically set to 50 and 150 for Pavia city center (Cuprite Mine) and Washington DC Mall (Pavia University), respectively.

D. Synthetic Data Experiments

In this subsection, we exhibit the generated results of all the competing methods on the four selected datasets, i.e., Cuprite Mine, Pavia University, Pavia city center, and Washington DC Mall. The values of MPSNR, MSSIM, SAM, ERGAS, and DD of the recovered HSIs for these datasets are given in Table II to Table V, where the highest values are purposely shown in bold face.

TABLE II
QUANTITATIVE METRICS (PSNR, SSIM, SAM, ERGAS, DD) OF ALL COMPARED METHODS ON CUPRITE MINE

Image	Method	MPSNR	MSSIM	SAM	ERGAS	DD
Cuprite Mine	NSSR	37.9809	0.9647	1.6736	0.9734	2.5756
	CSTF	39.5205	0.9751	1.5867	0.8049	2.1809
	UTV	41.3791	0.9799	1.4537	0.7658	1.8645
	STEREO	37.274	0.9513	2.3719	1.2419	3.0987
	SC-LL1	42.6106	0.9859	1.2207	0.6465	1.5449
	LTMR	42.7091	0.9866	1.2021	0.6416	1.5351
	LRTRTNN	**43.5868**	**0.9891**	**1.0908**	**0.5664**	**1.3787**

For the Cuprite Mine dataset, LRTRTNN generates the best results compared with other competing approaches in terms of all quantitative metrics, which verifies the superiority of our joint consideration of tensor ring decomposition and TNN prior. Specific to MPSNR, the improvement of LRTRTNN over NSSR, CSTF, UTV, STEREO, SC-LL1 and LTMR,

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JSTARS.2021.3108233, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing

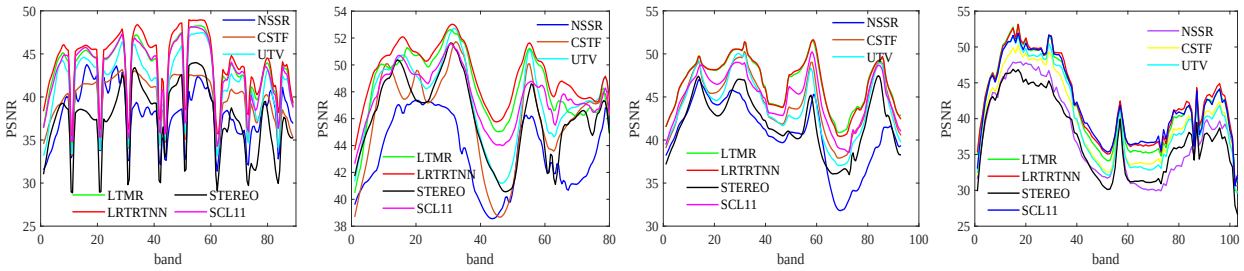


Fig. 5. PSNR values against different spectral bands on four datasets. (a) Cuprite Mine, (b) Pavia city center, (c) Pavia University, and (d) Washington DC Mall.

are 5.6059dB, 4.0663dB, 2.2077dB, 6.3128dB, 0.9762dB and 0.8777dB, respectively. Besides, on the purpose of visually comparing the performance of LRTRTNN with the others, some reconstructed images and the corresponding error maps are shown in Fig. 4. Moreover, for better visual perception, the representative areas of these images are magnified. Note that the error map reflects the difference between the fusion result and the ground truth, in which deeper blue denotes less residuals. At first blush, all the competitors can roughly learn the spatial details from HR-MSI, little difference can be found with naked eyes. However, from the error maps, especially in the magnified regions, one can easily find that our proposed method tends to generate much bluer and smoother results, which demonstrates that the fused HR-HSIs by NSSR, CSTF, UTV, STEREO, SC-LL1, and LTMR are with more flaws and scattered points. To exhibit the efficacy of our method in the spectral dimension, Fig. 5(a) further shows the band-wise PSNR values. We can see that LRTRTNN is consistently superior to other methods in all bands. This sub-figure together with the previous results demonstrates the superior performance of our hybrid usage of non-local similarity, global correlation, and TNN constrained tensor ring decomposition.

TABLE III
QUANTITATIVE METRICS (PSNR, SSIM, SAM, ERGAS, DD) OF ALL COMPARED METHODS ON PAVIA CITY CENTER

Image	Method	MPSNR	MSSIM	SAM	ERGAS	DD
Pavia city center	NSSR	43.1814	0.9921	2.5299	1.0353	1.2726
	CSTF	45.5026	0.9943	2.5448	0.8662	1.0488
	UTV	46.6649	0.9959	2.223	0.7017	0.8934
	STEREO	45.6477	0.9948	2.544	0.7666	1.024
	SC-LL1	46.9932	0.9967	2.0266	0.6462	0.8086
	LTMR	47.5827	0.9968	1.9134	0.6439	0.7553
	LRTRTNN	**48.3645**	**0.9974**	**1.7581**	**0.5491**	**0.6844**

As for the Pavia city center dataset, we can observe from Table III that LRTRTNN again ranks at the first place consistently for all quantitative metrics. Taking metric ERGAS as example, the result of SC-LL1 that ranks the second place lags behind ours by 0.0971. This improvement is quite remarkable since all the competing methods have been reported as achieving the state-of-the-art results in recent years. Taking the 50th band as example, we show the fused images and error maps in Fig. 6 for visual comparison. Similarly, we magnify a representative region of each image (or map) for easier observation. From the spatial perspective, the generated results are all of sharp edges and clear surfaces. However, part of methods still generates certain artifacts and whose results suffer from the loss of texture details. For instance, in Fig. 6 (b), the house roof in the magnified region is slightly blurrier, while the ones from other methods are closer to the ground truth. More evidently, from the error maps, we can easily observe the difference. The maps generated by LRTRTNN, SC-LL1 and LTMR have fewer residuals than those produced by NSSR, CSTF, STEREO, and UTV. Among these three, the map from LRTRTNN is significantly bluer. Fig. 5(b) further shows the PSNR values along with different spectral bands. Again, the curve of our method is clearly higher than other competing methods, which together with previous analysis illustrates the advantages of our method both spatially and spectrally. Fig. 7 further shows the behavior of spectral consistency on representative pixels. It is evident that the profiles generated from CSTF, UTV, and STEREO deviate from the original ones on the 40th channel to the 60th channel, while NSSR cannot perfectly capture the spectral characteristics of 35th to 45th channels. SCL11, LTMR, and LRTRTNN can capture the main spectral consistency. From a deeper insight, the curves of our method are much closer to the original one, which again demonstrates the superiority of tensor ring decomposition based on non-local similarity and tensor nuclear norm constraints.

TABLE IV
QUANTITATIVE METRICS (PSNR, SSIM, SAM, ERGAS, DD) OF ALL COMPARED METHODS ON PAVIA UNIVERSITY

Image	Method	MPSNR	MSSIM	SAM	ERGAS	DD
Pavia University	NSSR	40.5277	0.9767	2.4363	0.7384	1.6721
	CSTF	44.0264	0.9896	1.8907	0.5074	1.1301
	UTV	43.5311	0.9841	2.1483	1.0667	1.2813
	STEREO	41.7181	0.9791	2.5621	0.9409	1.005
	SC-LL1	44.505	0.99172	1.7261	**0.6462**	**0.8086**
	LTMR	45.9688	**0.9932**	1.4591	0.7905	0.8325
	LRTRTNN	**45.9746**	0.9931	**1.4513**	0.7928	0.8307

The overall quantitative results and the PSNR values along

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JSTARS.2021.3108233, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing

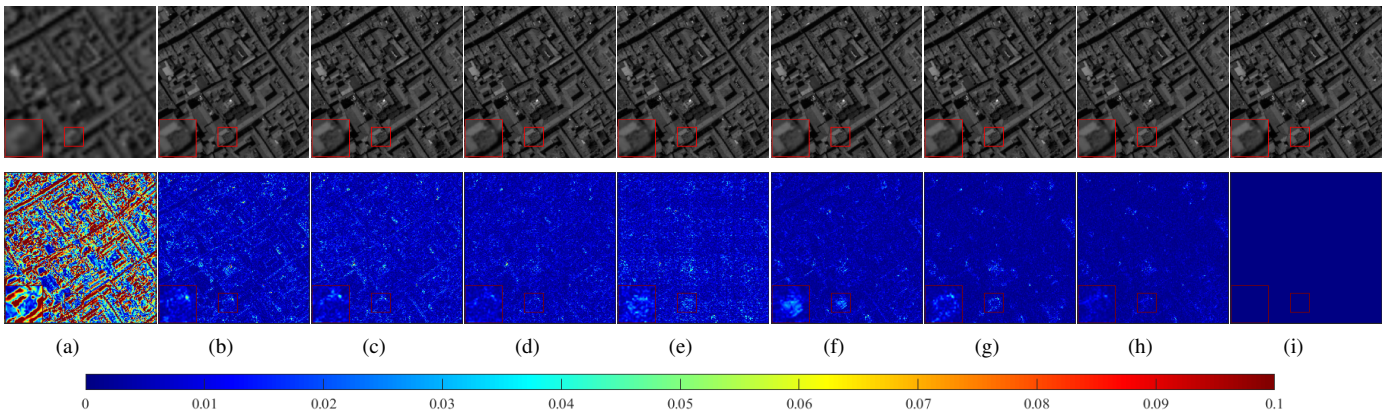


Fig. 6. The first row gives the fused images by all competing methods on the 50th band of Pavia city center. The second row gives the error images of all competitors on the 50th band of Pavia city center. (a) LR-HSI. (b) NSSR. (c) CSTF. (d) UTV. (e) STEREO. (f) SCL11. (g) LTMR. (h) LRTRTNN. (i) Ground truth.

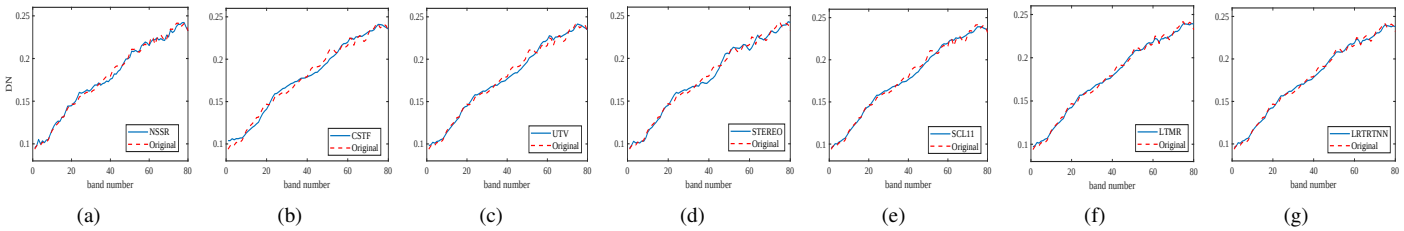


Fig. 7. Spectrum of pixel (20,20) in Pavia city center. (a) NSSR. (b) CSTF. (c) UTV. (d) STEREO. (e) SCL11. (f) LTMR. (g) LRTRTNN.

TABLE V
QUANTITATIVE METRICS (PSNR, SSIM, SAM, ERGAS, DD) OF ALL COMPARED METHODS ON WASHINGTON DC MALL

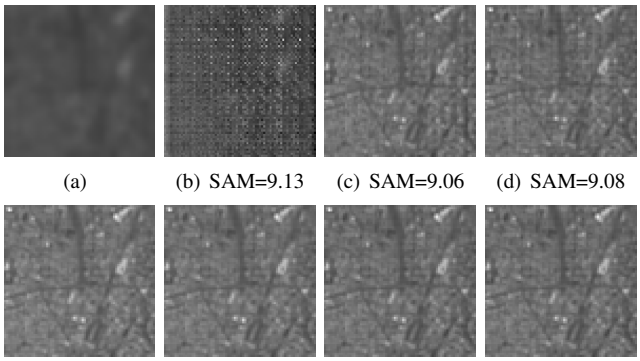
Image	Method	MPSNR	MSSIM	SAM	ERGAS	DD
Washington DC Mall	NSSR	37.7007	0.9747	3.6326	1.7767	2.7466
	CSTF	39.8669	0.9821	3.8171	1.4413	2.2982
	UTV	39.9599	0.9806	3.8765	1.4713	2.3601
	STEREO	37.3332	0.9718	5.2696	1.9013	3.1418
	SC-LL1	41.7423	0.98881	2.8578	1.1347	1.7204
	LTMR	41.4686	0.9872	2.9819	1.1963	1.8349
	LRTRTNN	**42.1144**	**0.9894**	**2.6210**	**1.0914**	**1.6188**

the spectral dimension for the Pavia University dataset are given in Table IV and Fig. 5(c), respectively. In Table IV, the generated values of our LRTRTNN are on par with LTMR. Among the five metrics, our MSSIM is inferior to that from LTMR, and our ERGAS and DD values are inferior to those from SC-LL1. However, our MPSNR and SAM values are cutting-edge. In Fig. 5(c), from the 45th to 55th bands, the result from our method is ahead of that from LTMR. Conversely, the curve of LTMR is slightly higher than ours in the 65-75 bands. We consider that LRTRTNN, LTMR, and SC-LL1 share the first place on this dataset since the numerical difference is marginal. Lastly, for the Washington DC Mall dataset, all numerical values of LRTRTNN are again much higher than others from Table V. In Fig. 5(d), the PSNR values along with all bands also confirm this result. This shows that our LRTRTNN can not only preserve the spatial structure but also maintain the spectral information. Taking SAM as example, our average results on four datasets are ahead of LTMR, SC-LL1, STEREO, UTV, CSTF, and NSSR by 0.159, 0.227, 1.457, 0.695, 0.702, 0.838, respectively. In general, all the experimental results verify the superiority of our method over other state-of-the-art HSI-MSI fusion methods, which demonstrates the appealing property by combining the global low-rank spectral structure and the spatially non-local self-similarity in the HSI-MSI fusion problem.

E. Semi-real experiment

In order to further evaluate our method in more practical scenarios, we apply our method to semi-real HSI and MSI fusion, in which the spatial downsampling operator is given while the spectral degradation function is unknown [51]. The HSI and MSI datasets were initially published in [25] and collected by the Hyperion and the Advanced Land Imager instruments on board the Earth Observing-1 Mission satellite. We treat the HSI which contains 72×72 pixels and 128 bands as the reference image and the MSI whose size is $72\times72\times9$ as the spectrally low resolution image. To obtain the LR-HSI, the reference image is downsampled at a ratio of 3. The spectral response matrix $\boldsymbol{R}$ is learned from the observed images via the approach presented in [25]. A visual assessment and the SAM results for all competing methods are given in Fig. 8. From the figure, our first observation is that NSSR fails in this job. Compared with the original LR-HSI, the other methods all significantly improve the spatial resolution. By digging into the details, there are more artifacts and ghost lines in the fused images from CSTF and UTV. The visual advantage

of our method over STEREO, SCL11, and LTMR cannot be easily judged by human eyes. However, the numerical results provide a lower SAM value for our LRTRTNN, which verifies its excellent performance over the other algorithms.



(a) (b) SAM=9.13 (c) SAM=9.06 (d) SAM=9.08

(e) SAM=10.05 (f) SAM=8.97 (g) SAM=8.26 (h) SAM=8.05

Fig. 8. Visual and numerical comparison of all competing methods on semi-real scenario. (a) LR-HSI. (b) NSSR. (c) CSTF. (d) UTV. (e) STEREO. (f) SCL11. (g) LTMR. (h) LRTRTNN.

F. Analysis and Discussion

1) **Convergence behavior:** Taking Pavia city center and Washington DC Mall as example, Fig. 9 depicts the relative error of the successively restored image $\boldsymbol{\mathcal{A}}^{t+1}$ and $\boldsymbol{\mathcal{A}}^{t}$, i.e., $\|\boldsymbol{\mathcal{A}}^{t+1}-\boldsymbol{\mathcal{A}}^{t}\|_F/\|\boldsymbol{\mathcal{A}}^{t}\|_F$, against the increasing iterations. From the curves, we can observe that the relative error drops sharply at the initial phase. Then, it keeps decreasing with a mild trend until the relative changes approach to zero. This empirically demonstrates the strong convergence behavior of LRTRTNN. Since the tolerant threshold would be reached with large probability when the number of iterations is 25. It is suggested as the default setting in all experiments.

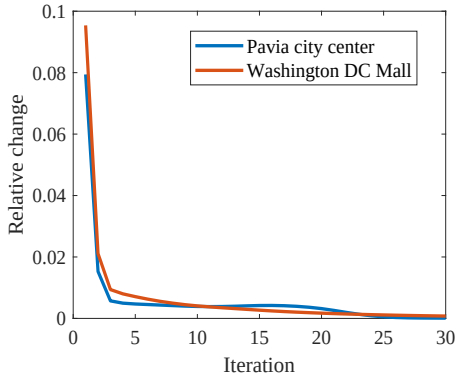


Fig. 9. Relative error of $\|\boldsymbol{\mathcal{A}}^{t+1}-\boldsymbol{\mathcal{A}}^{t}\|_F/\|\boldsymbol{\mathcal{A}}^{t}\|_F$ versus increasing iterations on Pavia city center and Washington DC Mall.

2) **Running efficiency:** To compare the computational cost of all approaches, Fig. 10 reports the elapsed time (in seconds) on Pavia city center dataset. Unfortunately, as can be seen from the figure, our method runs much slower than the others. The computational bottleneck lies in the sequentially searching of non-local patches and the joint optimization for all grouped cubes. Some more efficient optimization schemes or the relief of neighborhood searching are worthy of further in-depth study.

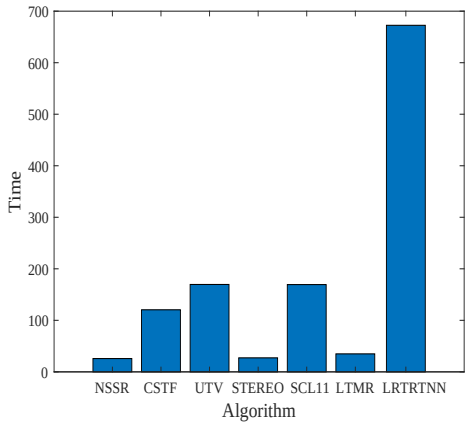


Fig. 10. The elapsed time (in seconds) of all compared approaches on Pavia city center dataset.

3) **The roles of TR and TNN:** To demonstrate the effectiveness of our TNN constrained TR decomposition, an ablation experiment is designed measuring the role of different components. The quantitative results on Pavia city center are shown in Table VI. Taking the final version of LRTRTNN as the baseline, the competing models include the one with matrix nuclear norm, the one without any regularization, and the one without the TR decomposition (setting both $\lambda = \eta = 0$). Thanks to the spectrally subspace decomposition operation, the model could still work without any regularization or any decomposition. However, the performance of this bald model deviates far away from the acceptable results. The introduction of TR evidently contributes to better efficacy. Moreover, a rank minimization constraint, whatever with a matrix or a tensor manner, would further enhance the performance. Between these two, TNN behaves consistently better than the counterpart in terms of all quantitative metrics, which supports the motivation of our usage of the TNN constrained TR model.

TABLE VI
QUANTITATIVE RESULTS OF ABLATION EXPERIMENTS ON PAVIA CITY CENTER

	PSNR	SSIM	SAM	ERGAS	DD
LRLRTNN	48.3645	0.9974	1.7581	0.5491	0.6844
With matrix NN	47.8754	0.990	1.8481	0.5968	0.7317
Without constraint	41.9079	0.9883	4.0787	1.3024	1.7146
Without TR	37.2683	0.96561	6.3497	2.2778	2.9547

V. CONCLUSION

We centre around the problem of recovering the HR-HSI by fusing the LR-HSI with HR-MSI in the same scene. In particular, a newly proposed HSI-MSI fusion method, namely Low-rank TR decomposition based on TNN (LRTRTNN), is

presented that involves three main merits, including spectral subspace learning, similar patches grouping, and the penalization of TNN prior on all tensor ring factors, which jointly considers the inherent correlation along with the spatial, spectral, and non-local modes. For the solving of the complex cost function, ADMM provides a theoretically solid and computationally efficient solution for a high-quality HSI with super-resolution both in spatial and spectral dimension. Extensively experiments results demonstrate that LRTRTNN can resoundingly enlarge the spatial resolution for the LR-HSIs with spectral dimension holding the same as the original ones, which exceeds most state-of-the-art approaches both visually and numerically.

In the near future, we aim to extend the method in three directions. First, for efficacy improvement, since the intensities of the HSI are naturally non-negative, we consider to impose non-negative constraint on the tensor factorization. Second, for lighter model selection, we attempt to adaptively estimate the intrinsic low-rank prior and incorporate it into model optimization. Finally, for efficiency promotion, the faster implementation of each subproblem, the easier search of similar patches, and different clustering schemes are also worthy of in-depth study.

REFERENCES

[1] L. Li, W. Li, Q. Du, and R. Tao, "Low-rank and sparse decomposition with mixture of gaussian for hyperspectral anomaly detection," *IEEE Transactions on Cybernetics*, pp. 1–10, 2020. doi:10.1109/TCYB.2020.2968750.

[2] J. Lei, G. Yang, W. Xie, Y. Li, and X. Jia, "A low-complexity hyperspectral anomaly detection algorithm and its fpga implementation," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 14, pp. 907–921, 2020.

[3] J. Zheng, J. Jiang, H. Xu, Z. Liu, and F. Gao, "Manifold-based nonlocal second-order regularization for hyperspectral image inpainting," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 14, pp. 224–236, 2021.

[4] M. Ding, X. Fu, T.-Z. Huang, J. Wang, and X.-L. Zhao, "Hyperspectral super-resolution via interpretable block-term tensor modeling," *IEEE Journal of Selected Topics in Signal Processing*, vol. 15, no. 3, pp. 641–656, 2021.

[5] J. Lin, T.-Z. Huang, X.-L. Zhao, T.-X. Jiang, and L. Zhuang, "A tensor subspace representation-based method for hyperspectral image denoising," *IEEE Transactions on Geoscience and Remote Sensing*, pp. 1–19, 2020. doi:10.1109/TGRS.2020.3032168.

[6] P. Ghamisi, N. Yokoya, J. Li, W. Liao, S. Liu, J. Plaza, B. Rasti, and A. Plaza, "Advances in hyperspectral image and signal processing: A comprehensive overview of the state of the art," *IEEE Geoscience and Remote Sensing Magazine*, vol. 5, no. 4, pp. 37–78, 2017.

[7] Y. Chang, L. Yan, and S. Zhong, "Hyper-laplacian regularized unidirectional low-rank tensor recovery for multispectral image denoising," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 4260–4268, 2017.

[8] 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 Transactions on Geoscience and Remote Sensing*, vol. 58, no. 1, pp. 734–749, 2019.

[9] N. He, M. E. Paoletti, J. M. Haut, L. Fang, S. Li, A. Plaza, and J. Plaza, "Feature extraction with multiscale covariance maps for hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 2, pp. 755–769, 2018.

[10] J. Zheng, Y. Feng, C. Bai, and J. Zhang, "Hyperspectral image classification using mixed convolutions and covariance pooling," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 59, no. 1, pp. 522–534, 2020.

[11] R. Dian, L. Fang, and S. Li, "Hyperspectral image super-resolution via non-local sparse tensor factorization," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 5344–5353, 2017.

[12] G. Vivone, L. Alparone, J. Chanussot, M. Dalla Mura, A. Garzelli, G. A. Licciardi, R. Restaino, and L. Wald, "A critical comparison among pansharpening algorithms," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, no. 5, pp. 2565–2586, 2014.

[13] X. He, L. Condat, J. M. Bioucas-Dias, J. Chanussot, and J. Xia, "A new pansharpening method based on spatial and spectral sparsity priors," *IEEE Transactions on Image Processing*, vol. 23, no. 9, pp. 4160–4174, 2014.

[14] Z. Chen, H. Pu, B. Wang, and G.-M. Jiang, "Fusion of hyperspectral and multispectral images: A novel framework based on generalization of pan-sharpening methods," *IEEE Geoscience and Remote Sensing Letters*, vol. 11, no. 8, pp. 1418–1422, 2014.

[15] J. Ma, W. Yu, C. Chen, P. Liang, X. Guo, and J. Jiang, "Pan-gan: An unsupervised pan-sharpening method for remote sensing image fusion," *Information Fusion*, vol. 62, pp. 110–120, 2020.

[16] Y. Qu, H. Qi, and C. Kwan, "Unsupervised sparse dirichlet-net for hyperspectral image super-resolution," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 2511–2520, 2018.

[17] Q. Yuan, Y. Wei, X. Meng, H. Shen, and L. Zhang, "A multiscale and multidepth convolutional neural network for remote sensing imagery pan-sharpening," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 11, no. 3, pp. 978–989, 2018.

[18] Q. Xie, M. Zhou, Q. Zhao, D. Meng, W. Zuo, and Z. Xu, "Multispectral and hyperspectral image fusion by ms/hs fusion net," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 1585–1594, 2019.

[19] R. Dian, S. Li, and X. Kang, "Regularizing hyperspectral and multispectral image fusion by cnn denoiser," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 32, no. 3, pp. 1124–1135, 2021.

[20] W. Dong, F. Fu, G. Shi, X. Cao, J. Wu, G. Li, and X. Li, "Hyperspectral image super-resolution via non-negative structured sparse representation," *IEEE Transactions on Image Processing*, vol. 25, no. 5, pp. 2337–2352, 2016.

[21] R. Wu, W.-K. Ma, X. Fu, and Q. Li, "Hyperspectral super-resolution via global–local low-rank matrix estimation," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 58, no. 10, pp. 7125–7140, 2020.

[22] R. Dian, S. Li, L. Fang, and Q. Wei, "Multispectral and hyperspectral image fusion with spatial-spectral sparse representation," *Information Fusion*, vol. 49, pp. 262–270, 2019.

[23] M. A. Veganzones, M. Simoes, G. Licciardi, N. Yokoya, J. M. Bioucas-Dias, and J. Chanussot, "Hyperspectral super-resolution of locally low rank images from complementary multisource data," *IEEE Transactions on Image Processing*, vol. 25, no. 1, pp. 274–288, 2015.

[24] Q. Wei, J. Bioucas-Dias, N. Dobigeon, J.-Y. Tourneret, M. Chen, and S. Godsill, "Multiband image fusion based on spectral unmixing," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 12, pp. 7236–7249, 2016.

[25] M. Simoes, J. Bioucas Dias, L. B. Almeida, and J. Chanussot, "A convex formulation for hyperspectral image superresolution via subspace-based regularization," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, no. 6, pp. 3373–3388, 2015.

[26] S. Li, R. Dian, L. Fang, and J. M. Bioucas-Dias, "Fusing hyperspectral and multispectral images via coupled sparse tensor factorization," *IEEE Transactions on Image Processing*, vol. 27, no. 8, pp. 4118–4130, 2018.

[27] W. He, Y. Chen, N. Yokoya, C. Li, and Q. Zhao, "Hyperspectral super-resolution via coupled tensor ring factorization," *arXiv preprint arXiv:2001.01547*, 2020.

[28] R. Dian and S. Li, "Hyperspectral image super-resolution via subspace-based low tensor multi-rank regularization," *IEEE Transactions on Image Processing*, vol. 28, no. 10, pp. 5135–5146, 2019.

[29] Y. Chang, L. Yan, X.-L. Zhao, H. Fang, Z. Zhang, and S. Zhong, "Weighted low-rank tensor recovery for hyperspectral image restoration," *IEEE transactions on cybernetics*, vol. 50, no. 11, pp. 4558–4572, 2020.

[30] T. Xu, T.-Z. Huang, L.-J. Deng, X.-L. Zhao, and J. Huang, "Hyperspectral image superresolution using unidirectional total variation with tucker decomposition," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 13, pp. 4381–4398, 2020.

[31] C. I. Kanatsoulis, X. Fu, N. D. Sidiropoulos, and W.-K. Ma, "Hyperspectral super-resolution: A coupled tensor factorization approach," *IEEE Transactions on Signal Processing*, vol. 66, no. 24, pp. 6503–6517, 2018.

[32] R. Dian, S. Li, and L. Fang, "Learning a low tensor-train rank representation for hyperspectral image super-resolution," *IEEE transactions on neural networks and learning systems*, vol. 30, no. 9, pp. 2672–2683, 2019.

[33] Y. Xu, Z. Wu, J. Chanussot, and Z. Wei, "Nonlocal patch tensor sparse representation for hyperspectral image super-resolution," *IEEE Transactions on Image Processing*, vol. 28, no. 6, pp. 3034–3047, 2019.
[34] H. Xu, J. Zheng, X. Yao, Y. Feng, and S. Chen, "Fast tensor nuclear norm for structured low-rank visual inpainting," *IEEE Transactions on Circuits and Systems for Video Technology*, pp. 1–14, 2021. doi: 10.1109/TCSVT.2021.3067022.
[35] J.-H. Yang, X.-L. Zhao, T.-Y. Ji, T.-H. Ma, and T.-Z. Huang, "Low-rank tensor train for tensor robust principal component analysis," *Applied Mathematics and Computation*, vol. 367, p. 124783, 2020.
[36] Y. Liu, Z. Long, and C. Zhu, "Image completion using low tensor tree rank and total variation minimization," *IEEE Transactions on Multimedia*, vol. 21, no. 2, pp. 338–350, 2018.
[37] 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 *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, pp. 9151–9158, 2019.
[38] A. Ahad, Z. Long, C. Zhu, and Y. Liu, "Hierarchical tensor ring completion," *arXiv preprint arXiv:2004.11720*, 2020.
[39] Y. Chen, W. He, N. Yokoya, T.-Z. Huang, and X.-L. Zhao, "Nonlocal tensor-ring decomposition for hyperspectral image denoising," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 58, no. 2, pp. 1348–1362, 2019.
[40] W. He, N. Yokoya, L. Yuan, and Q. Zhao, "Remote sensing image reconstruction using tensor ring completion and total variation," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 11, pp. 8998–9009, 2019.
[41] Q. Zhao, G. Zhou, S. Xie, L. Zhang, and A. Cichocki, "Tensor ring decomposition," *arXiv preprint arXiv:1606.05535*, 2016.
[42] Z. Zhang and S. Aeron, "Exact tensor completion using t-svd," *IEEE Transactions on Signal Processing*, vol. 65, no. 6, pp. 1511–1526, 2016.
[43] C. Lu, J. Feng, Y. Chen, W. Liu, Z. Lin, and S. Yan, "Tensor robust principal component analysis with a new tensor nuclear norm," *IEEE transactions on pattern analysis and machine intelligence*, vol. 42, no. 4, pp. 925–938, 2019.
[44] J. Zheng, P. Yang, X. Yang, and S. Chen, "Truncated low-rank and total p variation constrained color image completion and its moreau approximation algorithm," *IEEE Transactions on Image Processing*, vol. 29, pp. 7861–7874, 2020.
[45] Q. Wei, N. Dobigeon, and J. Tourneret, "Fast fusion of multi-band images based on solving a sylvester equation," *IEEE transactions on image processing : a publication of the IEEE Signal Processing Society*, vol. 24, no. 11, pp. 4109–21, 2015.
[46] G. Vane, R. O. Green, T. G. Chrien, H. T. Enmark, E. G. Hansen, and W. M. Porter, "The airborne visible/infrared imaging spectrometer (aviris)," *Remote sensing of environment*, vol. 44, no. 2-3, pp. 127–143, 1993.
[47] F. Dell'Acqua, P. Gamba, A. Ferrari, J. A. Palmason, J. A. Benediktsson, and K. Árnason, "Exploiting spectral and spatial information in hyperspectral urban data with high resolution," *IEEE Geoscience and Remote Sensing Letters*, vol. 1, no. 4, pp. 322–326, 2004.
[48] R. W. Basedow, D. C. Carmer, and M. E. Anderson, "Hydice system: Implementation and performance," in *Imaging Spectrometry*, vol. 2480, pp. 258–267, International Society for Optics and Photonics, 1995.
[49] Q. Wei, J. Bioucas-Dias, N. Dobigeon, and J.-Y. Tourneret, "Hyperspectral and multispectral image fusion based on a sparse representation," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, no. 7, pp. 3658–3668, 2015.
[50] 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 Transactions on Image Processing*, vol. 30, pp. 3581–3596, 2021.
[51] R. A. Borsoi, T. Imbiriba, and J. C. M. Bermudez, "Super-resolution for hyperspectral and multispectral image fusion accounting for seasonal spectral variability," *IEEE Transactions on Image Processing*, vol. 29, pp. 116–127, 2020.

Honghui Xu received the B.E. degree from Zhejiang University of Technology, Hangzhou, China in 2018. He is currently pursuing the Ph.D. degree in the School of Computer Science and Engineering, Zhejiang University of Technology, Hangzhou, China. His current research interests include image processing and optimization algorithm.

Mengjie Qin received the B.Eng. degree in information management and information system from Guangdong Ocean University, Zhanjiang, China, in 2017. She is currently working toward the Ph.D. degree with the College of Computer Science and Engineering, Zhejiang University of Technology, Hangzhou, China. Her current research interests include image and video enhancement, pattern recognition, and machine learning.

Shengyong Chen (Senior Member, IEEE) received the Ph.D. degree in computer vision from the City University of Hong Kong, Hong Kong, in 2003. He worked with the University of Hamburg, from 2006 to 2007. He is currently a Professor with the Tianjin University of Technology and the Zhejiang University of Technology, China. He has published over 100 scientific articles in international journals. His research interests include computer vision, robotics, and image analysis. He is also a Senior Member of CCF and a Fellow of IET. He has received the fellowship from the Alexander von Humboldt Foundation of Germany. He has also received the National Outstanding Youth Foundation Award of China, in 2013.

Yuhui Zheng (M'15) was born in Ruicheng, Shanxi, China, in 1982. He received the B.S. degree in chemistry and the Ph.D. degree in computer science from the Nanjing University of Science and Technology, Nanjing, China, in 2004 and 2009, respectively. From 2014 to 2015, he was a Visiting Scholar with the Digital Media Laboratory, School of Electronic and Electrical Engineering, Sungkyunkwan University, Suwon, South Korea. He is currently a Professor with the School of Computer and Software, Nanjing University of Information Science and Technology, Nanjing. His current research interests include inverse problems in image processing, scientific computing, data mining, and pattern recognition.

Jianwei Zheng received the B.E. degree in electronic and computer engineering and the Ph.D. degree in control theory and control engineering from the Zhejiang University of Technology, Hangzhou, China, in 2005 and 2010, respectively. He is an Associate Professor with the College of Computer Science and Technology, Zhejiang University of Technology. His research interests include machine learning and data analysis. He has published more than 70 academic papers in reputable journals and conferences, including IEEE Trans. Image Processing, IEEE Trans. Neural Networks and Learning Systems, IEEE Trans. Geoscience and Remote Sensing, IEEE Trans. Industrial Informatics, Pattern Recognition, and so on.