H2TF for Hyperspectral Image Denoising: Where Hierarchical Nonlinear Transform Meets Hierarchical Matrix Factorization

Jia-Yi Li, Jin-Yu Xie, Yi-Si Luo, Xi-Le Zhao, and Jian-Li Wang

***Abstract*— Recently, tensor singular value decomposition (t-SVD) has emerged as a promising tool for hyperspectral image (HSI) processing. In the t-SVD, there are two key building blocks: 1) the low-rank enhanced transform and 2) the accompanying low-rank characterization of transformed frontal slices. Previous t-SVD methods mainly focus on the developments of 1), while neglecting the other important aspect, i.e., the exact characterization of transformed frontal slices. In this letter, we exploit the potentiality in both building blocks by leveraging the hierarchical nonlinear transform (HSI) and the hierarchical matrix factorization (HMF) to establish a new tensor factorization (termed as H2TF). Compared with shallow counter partners, e.g., low-rank matrix factorization (MF) or its convex surrogates, H2TF can better capture complex structures of transformed frontal slices due to its hierarchical modeling abilities. We then suggest the H2TF-based HSI denoising model and develop an alternating direction method of multipliers-based algorithm to address the resultant model. Extensive experiments validate the superiority of our method over state-of-the-art (SOTA) HSI denoising methods.**

***Index Terms*— Alternating direction method of multipliers (ADMM), hyperspectral denoising, tensor singular value decomposition (t-SVD).**

I. Introduction

HYPERSPECTRAL images (HSIs) inevitably contain mixed noise due to sensor failures or complex imaging conditions [1], [2], [3], which seriously affects subsequent applications. Traditional hand-crafted HSI denoising methods, e.g., low-rankness [4], total variation (TV) [5], sparse representations [6], and nonlocal self-similarity [7], use interpretable domain knowledge to design generalizable regularizations for HSI denoising. Their representation abilities may be inferior to data-driven methods using deep neural networks (DNNs) [8], [9], [10], which can learn representative denoising mappings via supervised learning with abundant training pairs. However, supervised deep learning methods mostly neglect the prior information of HSIs, which sometimes results in generalization issues over different HSIs and various types of noise.

More recently, tensor singular value decomposition (t-SVD) attracts much attention in HSI denoising [11], [12]. The t-SVD views HSI as an implicit low-rank tensor and exploits the low-rankness in the transformed domain. Under such a framework, there are naturally two key building blocks.

1) The selection of the low-rank enhanced transform. A suitable transform can obtain a lower rank transformed tensor and enhance the recovery quality [13], [14].
2) The characterization of low-rankness of transformed frontal slices. The implicit low-rankness of HSIs is exploited by the low-rank modeling of frontal slices in the transformed domain.

Classical t-SVD-based methods mainly focused on the first building blocks, i.e., the design of different transforms. For example, the discrete Fourier transform (DFT) [15] was first used in the t-SVD, and then the discrete cosine transform (DCT) [16] was used. Later methods exploited more representative and flexible transforms such as noninvertible transforms [17] and data-dependent transforms [18] to enhance the low-rankness of transformed frontal slices. These methods have achieved increasingly satisfactory results for HSI denoising [11], [12]. Nevertheless, these t-SVD methods pay less attention to the second building block, i.e., the exact characterization of transformed frontal slices. Specifically, they all use shallow representations such as low-rank matrix factorization (MF) [14], QR factorization [19], and nuclear norm [13], [17] to characterize the transformed frontal slices.

In this work, we exploit a more representative formulation to capture complex structures of transformed frontal slices. Specifically, we leverage the hierarchical matrix factorization (HMF), which tailors a hierarchical formulation of learnable matrices along with nonlinear layers to capture each frontal slice in the transformed domain. The hierarchical modeling ability of HMF makes it more representative to capture the complex structures of HSIs. Meanwhile, we leverage the hierarchical nonlinear transform (HNT) to enhance the low-rankness of transformed frontal slices. With the HNT and HMF, we develop a new tensor factorization method (termed as H2TF) under the t-SVD framework. Correspondingly, we develop the H2TF-based HSI denoising model. Attributed to the stronger representation abilities of HMF than shallow MF or its surrogates, our H2TF-based model can better capture fine details of the underlying clean HSI than the conventional

Manuscript received 20 April 2023; revised 2 July 2023; accepted 10 July 2023. Date of publication 13 July 2023; date of current version 25 July 2023. This work was supported in part by NSFC under Grant 12171072 and Grant 62131005, in part by the National Key Research and Development Program of China under Grant 2020YFA0714001, and in part by the Open Research Fund Program of Data Recovery Key Laboratory of Sichuan Province under Grant DRN2302. *(Jia-Yi Li and Jin-Yu Xie contributed equally to this work.) (Corresponding authors: Yi-Si Luo; Xi-Le Zhao.)*

Jia-Yi Li and Jin-Yu Xie are with the Yingcai Honors College, University of Electronic Science and Technology of China, Chengdu 610000, China (e-mail: lijiayi03531@gmail.com; xiejinyu554@gmail.com).

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

Xi-Le Zhao is with the School of Mathematical Sciences, University of Electronic Science and Technology of China, Chengdu 610000, China (e-mail: xlzhao122003@163.com).

Jian-Li Wang is with the School of Computing and Artificial Intelligence, Southwest Jiaotong University, Chengdu 610032, China (e-mail: wangjianli_123@163.com).

This article has supplementary downloadable material available at https://doi.org/10.1109/LGRS.2023.3294933, provided by the authors.

Digital Object Identifier 10.1109/LGRS.2023.3294933

t-SVD-based methods. Thus, our model is expected to deliver better HSI denoising results. Meanwhile, the parameters of H2TF can be inferred from the observed noisy HSI in an unsupervised manner. In summary, the contributions of this letter are given follows.

1) We propose a new tensor factorization, i.e., the H2TF, which leverages the expressive power of two key building blocks—the HNT and the HMF, to, respectively, enhance the low-rankness of transformed data and characterize complex structures of transformed frontal slices. By virtue of their hierarchical modeling abilities, H2TF can faithfully capture fine details of the clean HSI, and thus is beneficial for effectively removing heavy noise in the HSI.
2) We suggest an unsupervised H2TF-based HSI denoising model and develop an alternating direction method of multipliers (ADMM)-based algorithm. Extensive experiments on the simulated and real-world data validate the superiority of our method over the state-of-the-art (SOTA) HSI denoising methods, especially for details preserving and heavy noise removal.

II. Proposed H2TF

A. t-SVD Framework

We first introduce the general formulation of t-SVD. Suppose that the noisy HSI $\mathcal{Y} \in \mathbb{R}^{h\times w\times b}$ admits $\mathcal{Y} = \mathcal{X} + \mathcal{N}$, where $\mathcal{X}$ denotes the clean HSI and $\mathcal{N}$ denotes noise. To infer the underlying clean HSI $\mathcal{X}$ from the observed $\mathcal{Y}$, the t-SVD method generally formulates the following model:

$$\min_{\mathcal{Z},\theta} L(\mathcal{Y}, \mathcal{X}) + \sum_{k} \psi(\mathcal{Z}^{(k)}), \quad \text{where } \mathcal{X} = \phi_\theta(\mathcal{Z}). \tag{1}$$

Here, $L(\mathcal{Y}, \mathcal{X})$ denotes the fidelity term and $\psi(\mathcal{Z}^{(k)})$ represents the low-rank characterization of $\mathcal{Z}^{(k)}$ (which denotes the kth frontal (spatial) slice of $\mathcal{Z} \in \mathbb{R}^{h\times w\times b}$ [17]). $\phi_\theta(\cdot) : \mathbb{R}^{h\times w\times b} \to \mathbb{R}^{h\times w\times b}$ denotes a transform with learnable parameters θ, which transforms the low-rank representation $\mathcal{Z}$ into the original domain. Sometimes the transform $\phi_\theta(\cdot)$ may not be learnable (e.g., the fixed DFT [15]), and in those situations the optimization variable only includes $\mathcal{Z}$.

The philosophy of the t-SVD model (1) is to minimize the rank in the transformed domain, which can model the implicit low-rankness of HSI. There are naturally two key building blocks of the t-SVD-based methods, i.e., the selection of the transform $\phi_\theta(\cdot)$ and the exact low-rank characterization $\psi(\cdot)$ of the transformed frontal slice $\mathcal{Z}^{(k)}$. Most t-SVD-based methods focus on the design of different transforms $\phi_\theta(\cdot)$ (see examples in [14], [17], and [18]), but they pay less attention to the characterization of the transformed frontal slice. They mostly adopt shallow representations to characterize $\mathcal{Z}^{(k)}$, e.g., MF [14], [20], QR factorization [19], and nuclear norm [16], [17]. However, these shallow representations may not be expressive enough to capture fine details of HSIs. Therefore, more representative methods are desired to enhance the representation abilities of the model in the transformed domain.

B. HMF for Characterizing $\mathcal{Z}^{(k)}$

To cope with this challenge, we leverage the HMF to characterize $\mathcal{Z}^{(k)}$. The hierarchical modeling ability of HMF helps it more faithfully capture complex structures of the transformed frontal slice $\mathcal{Z}^{(k)}$ than shallow counter partners, e.g., SVD, MF, and QR factorization.

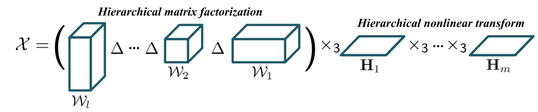

Fig. 1. General illustration of the H2TF representation of a tensor $\mathcal{X}$. The nonlinear layer $\sigma(\cdot)$ is omitted for space consideration.

The standard MF decomposes a low-rank matrix $\mathbf{Z} \in \mathbb{R}^{h\times w}$ into two factors as $\mathbf{Z} = \mathbf{W}_2\mathbf{W}_1$, where $\mathbf{W}_2 \in \mathbb{R}^{h\times r}$, $\mathbf{W}_1 \in \mathbb{R}^{r\times w}$, and r is the rank. To model the hierarchical structures of $\mathbf{Z}$, we extend the MF to the product of multiple matrix factors $\{\mathbf{W}_d\}_{d=1}^{l}$, i.e., $\mathbf{Z} = \mathbf{W}_l\mathbf{W}_{l-1}, \ldots, \mathbf{W}_1$, where $\mathbf{W}_d \in \mathbb{R}^{r_d\times r_{d-1}}$, $r_l = h$, and $r_0 = w$. It was shown in [21] that such a linear HMF induces an implicit low-rank regularization on $\mathbf{Z}$. Generally, the larger the l (i.e., adding depth to the HMF), the tendency of low-rankness goes stronger and oftentimes leads to better recovery performances. Thus, the HMF is suitable to play the role of low-rank regularization in the t-SVD.

Nevertheless, the linear HMF may not be sufficient to capture nonlinear interactions inside HSIs. It motivates us to use the nonlinear HMF [22], [23] to model the low-rank matrix $\mathbf{Z}$ via $\mathbf{Z} = \mathbf{W}_l\sigma(\mathbf{W}_{l-1}, \ldots, \sigma(\mathbf{W}_3\sigma(\mathbf{W}_2\mathbf{W}_1)))$, where $\sigma(\cdot)$ is a nonlinear scalar function. Classical HMF-based methods [21], [22] only use HMF to tackle the 2-D matrix. However, matrixing the HSI inevitably destroys its high-dimensional data structures. Therefore, we suggest tailoring b nonlinear HMFs to model the transformed tensor $\mathcal{Z}$ using each HMF to represent one of the frontal slices of $\mathcal{Z}$. Formally, we represent each frontal slice of $\mathcal{Z}$ by

$$\mathcal{Z}^{(k)} = \mathcal{W}_l^{(k)}\sigma\left(\mathcal{W}_{l-1}^{(k)}, \ldots, \sigma\left(\mathcal{W}_3^{(k)}\sigma\left(\mathcal{W}_2^{(k)}\mathcal{W}_1^{(k)}\right)\right)\right)$$
$$k = 1, 2, \ldots, b.$$

The above HMFs can be equivalently formulated as the tensor formulation $\mathcal{Z} = \mathcal{W}_l\Delta\sigma(\mathcal{W}_{l-1}\Delta, \ldots, \sigma(\mathcal{W}_3\Delta\sigma(\mathcal{W}_2\Delta\mathcal{W}_1)))$, where Δ is the tensor facewise product [24] and $\{\mathcal{W}_d \in \mathbb{R}^{r_d\times r_{d-1}\times b}\}_{d=1}^{l}$ are some factor tensors.

Compared with MF, QR factorization, and nuclear norm, the nonlinear HMF can better capture complex structures of HSIs due to its nonlinear hierarchical modeling abilities, which help capture fine details of HSI and remove heavy noise.

C. Proposed H2TF

Next, we introduce our H2TF. Recall that two key building blocks in the t-SVD are the selection of the transform $\phi_\theta(\cdot)$ and the characterization of the transformed frontal slice $\mathcal{Z}^{(k)}$. We suggest the HNT as the first building block $\phi_\theta(\cdot)$

$$\phi_\theta(\mathcal{Z}) := \sigma(\cdots\sigma(\mathcal{Z} \times_3 \mathbf{H}_1) \times_3 \cdots \times_3 \mathbf{H}_{m-1}) \times_3 \mathbf{H}_m$$

where $\sigma(\cdot)$ is a nonlinear scalar function, which is consistent with the nonlinear scalar function used in HMF, $\theta := \{\mathbf{H}_p \in \mathbb{R}^{b\times b}\}_{p=1}^{m}$ are the learnable parameters of HNT, and $\times_3$ is the mode-3 tensor–matrix product [25]. It was demonstrated [14] that the HNT can effectively enhance the low-rankness of transformed tensor, and thus obtain a better low-rank representation than shallow transforms (e.g., DFT [15] and DCT [16]), which benefits the implicit low-rank modeling.

Definition 1 (H2TF): Finally, we can define the following factorization modality of a certain low-rank tensor $\mathcal{X}$

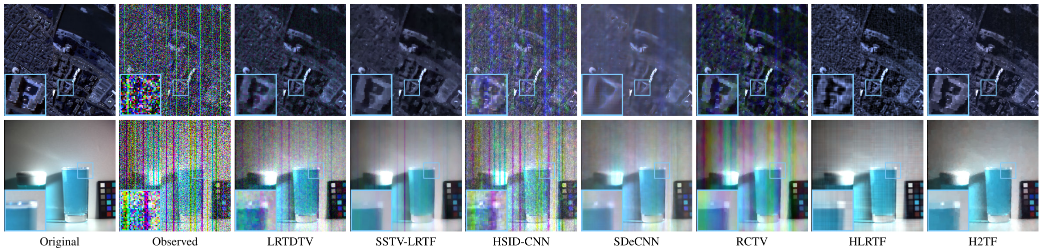

Fig. 2. Pseudocolor images of HSI denoising results by different methods on simulated data *PaviaC* Case 4 (first row) and *Cups* Case 5 (second row).

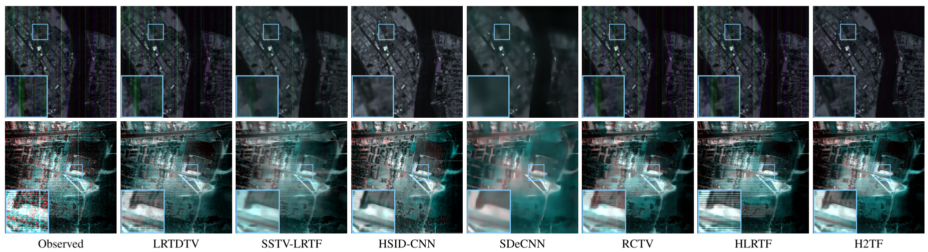

Fig. 3. Pseudocolor images of HSI denoising results by different methods on real-world data *Shanghai* (first row) and *Urban* (second row).

parameterized by $\{\mathcal{W}_d\}_{d=1}^{l}$ and $\{\mathbf{H}_p\}_{p=1}^{m}$

$$\mathcal{X} = \phi_\theta\big(\underbrace{\mathcal{W}_l \Delta \sigma(\mathcal{W}_{l-1}\Delta, \ldots, \sigma(\mathcal{W}_3 \Delta \sigma(\mathcal{W}_2 \Delta \mathcal{W}_1))}_{\text{Hierarchical matrix factorization}}\big)$$
$$\phi_\theta(\mathcal{Z}) := \underbrace{\sigma(\cdots\sigma(\mathcal{Z}\times_3 \mathbf{H}_1)\times_3 \cdots \times_3 \mathbf{H}_{m-1})\times_3 \mathbf{H}_m}_{\text{Hierarchical nonlinear transform}} \tag{2}$$

which we call the H2TF representation of $\mathcal{X}$.

A general illustration of H2TF is shown in Fig. 1. H2TF benefits from the HMF to exploit complex structures of transformed frontal slices and the HNT to enhance the low-rankness in the transformed domain. Therefore, H2TF can more faithfully capture fine details and rich textures of HSIs and remove heavy mixed noise. Now, we discuss the connections between H2TF and some popular matrix/tensor factorizations.

Remark 1: By changing the layer number of HMF (i.e., l) and the layer number of HNT (i.e., m), H2TF includes many matrix/tensor factorizations as special cases.

1) When $l = 2$, i.e., the HMF degenerates into the MF, our H2TF degenerates into the hierarchical low-rank tensor factorization [14].
2) When $m = 1$ and $\mathbf{H}_m$ is an identity matrix (i.e., the transform $\phi_\theta(\cdot)$ is an identical mapping), our H2TF degenerates into the plain HMFs [22], [23] applied on each frontal slice of the tensor separately. In the following, we interpret this case as "$m = 0$" since the transform is neglected.
3) When $l = 2$ and $m = 1$ with $\mathbf{H}_m$ being the fixed inverse DFT matrix, our H2TF degenerates into the classical low-tubal-rank tensor factorization [20], [26].

Moreover, H2TF can explicitly preserve the low-rankness of the tensor when omitting some nonlinearity, as stated below.

Lemma 1: Suppose that $\mathcal{X} = \phi(\mathcal{W}_l\Delta(\mathcal{W}_{l-1}\Delta, \ldots, \Delta\mathcal{W}_1)) \in \mathbb{R}^{h\times w\times b}$, where $\{\mathcal{W}_d \in \mathbb{R}^{r_d\times r_{d-1}\times b}\}_{d=1}^{l}$ ($r_l = h$ and $r_0 = w$) are factor tensors, $\phi(\mathcal{Z}) := \mathcal{Z}\times_3 \mathbf{F}^{-1}$ is the inverse DFT, and $\mathbf{F}^{-1}$ is the inverse DFT matrix (which is a special case of H2TF). Then we have $\text{rank}_t(\mathcal{X}) \le \min\{r_0, r_1, \ldots, r_l\}$, where $\text{rank}_t(\cdot)$ denotes the tensor tubal rank [13], [14], [15].

Lemma 1 indicates that H2TF can preserve the low-rankness in the linear special case, where the degree of low-rankness (the upper bound of tubal rank) is conditioned on the sizes of factor tensors. Therefore, we can readily control the degree of low-rankness by tuning the sizes of factor tensors in H2TF.

D. H2TF for HSI Denoising

H2TF is a potential tool for multidimensional data analysis and processing. We consider HSI denoising as a representative real-world application. By applying the H2TF representation (2) into (1), we can obtain the following HSI denoising model:

$$\min_{\{\mathcal{W}_d\}_{d=1}^{l}, \{\mathbf{H}_p\}_{p=1}^{m}} L(\mathcal{Y}, \mathcal{X}), \quad \text{where}$$
$$\mathcal{X} = \phi_\theta\big(\mathcal{W}_l\Delta\sigma(\mathcal{W}_{l-1}\Delta, \ldots, \sigma(\mathcal{W}_3\Delta\sigma(\mathcal{W}_2\Delta\mathcal{W}_1)))\big).$$

In the HSI denoising problem, we consider the fidelity term as $L(\mathcal{Y}, \mathcal{X}) = \|\mathcal{Y} - \mathcal{X} - \mathcal{S}\|_F^2 + \alpha_1\|\mathcal{S}\|_{\ell_1}$, where $\|\cdot\|_F^2$ denotes the Frobenius norm and we introduce $\mathcal{S} \in \mathbb{R}^{h\times w\times b}$ to represent sparse noise (often contains impulse noise and stripes). The ℓ_1-norm enforces the sparsity on $\mathcal{S}$ so that the sparse noise can be eliminated. Here, α_1 is a tradeoff parameter.

Meanwhile, our H2TF can be readily combined with other proven techniques to enhance the denoising abilities. Here, we consider the hybrid spatial–spectral TV (HSSTV) regularization [27] to further capture spatial and spatial–spectral local smoothness of HSIs. The HSSTV is formulated as $\|\mathcal{X}\|_{\text{HSSTV}} := \alpha_2\|\mathcal{X}\|_{\text{TV}} + \alpha_3\|\mathcal{X}\|_{\text{SSTV}}$, where $\|\mathcal{X}\|_{\text{TV}} := \|\nabla_x\mathcal{X}\|_{\ell_1} + \|\nabla_y\mathcal{X}\|_{\ell_1}$, $\|\mathcal{X}\|_{\text{SSTV}} := \|\nabla_x(\nabla_z\mathcal{X})\|_{\ell_1} + \|\nabla_y(\nabla_z\mathcal{X})\|_{\ell_1}$, and α_i ($i = 2, 3$) are tradeoff parameters. Here, the derivative operators are defined as $(\nabla_x\mathcal{X})_{(i,j,k)} := \mathcal{X}_{(i+1,j,k)} - \mathcal{X}_{(i,j,k)}$, $(\nabla_y\mathcal{X})_{(i,j,k)} := \mathcal{X}_{(i,j+1,k)} - \mathcal{X}_{(i,j,k)}$, and $(\nabla_z\mathcal{X})_{(i,j,k)} := \mathcal{X}_{(i,j,k+1)} - \mathcal{X}_{(i,j,k)}$, where $\mathcal{X}_{(i,j,k)}$ denotes the (i, j, k)th element of $\mathcal{X}$.

Based on the formulations of fidelity term and HSSTV, the proposed H2TF-based HSI denosing model is formulated as

$$\min_{\{\mathcal{W}_d\}_{d=1}^{l}, \{\mathbf{H}_p\}_{p=1}^{m}, \mathcal{S}} \|\mathcal{Y}-\mathcal{X}-\mathcal{S}\|_F^2 + \alpha_1\|\mathcal{S}\|_{\ell_1} + \|\mathcal{X}\|_{\text{HSSTV}}, \quad \text{where}$$
$$\mathcal{X} = \phi_\theta\big(\mathcal{W}_l\Delta\sigma(\mathcal{W}_{l-1}\Delta, \ldots, \sigma(\mathcal{W}_3\Delta\sigma(\mathcal{W}_2\Delta\mathcal{W}_1)))\big). \tag{3}$$

Compared with previous t-SVD-based HSI denoising methods [11], [12], H2TF has powerful representation abilities brought from the hierarchical structures and thus could better capture fine details of HSIs. Besides, the parameters of H2TF are unsupervisedly inferred from the noisy HSI by optimizing (3) without the requirement of training process.

E. ADMM-Based Algorithm

To tackle the problem (3), we develop an ADMM-based algorithm. By introducing auxiliary variables $\mathcal{V}_i$ ($i = 1, 2, 3, 4$), (3) can be equivalently formulated as

$$\begin{aligned}\min_{\substack{\{\mathcal{W}_d\}_{d=1}^{l}, \{\mathbf{H}_p\}_{p=1}^{m} \\ \mathcal{S}, \{\mathcal{V}_i\}_{i=1}^{4}}} & \|\mathcal{Y} - \mathcal{X} - \mathcal{S}\|_F^2 + \alpha_1 \|\mathcal{S}\|_{\ell_1} + \alpha_2 \|\mathcal{V}_1\|_{\ell_1} \\ & + \alpha_2 \|\mathcal{V}_2\|_{\ell_1} + \alpha_3 \|\mathcal{V}_3\|_{\ell_1} + \alpha_3 \|\mathcal{V}_4\|_{\ell_1} \\ \text{s.t. } & \mathcal{V}_1 = \nabla_x \mathcal{X}, \quad \mathcal{V}_2 = \nabla_y \mathcal{X}, \mathcal{V}_3 = \nabla_x(\nabla_z \mathcal{X}), \\ & \mathcal{V}_4 = \nabla_y(\nabla_z \mathcal{X})\end{aligned}$$

where $\mathcal{X} = \phi_\theta\big(\mathcal{W}_l \Delta \sigma(\mathcal{W}_{l-1} \Delta, \ldots, \sigma(\mathcal{W}_3 \Delta \sigma(\mathcal{W}_2 \Delta \mathcal{W}_1)))\big)$. The corresponding augmented Lagrangian function is

$$\begin{aligned}\mathcal{L}_\mu & \Big(\{\mathcal{W}_d\}_{d=1}^{l}, \big\{\mathbf{H}_p\big\}_{p=1}^{m}, \mathcal{S}, \{\mathcal{V}_i\}_{i=1}^{4}, \{\Lambda_i\}_{i=1}^{4}\Big) \\ &= \|\mathcal{Y} - \mathcal{X} - \mathcal{S}\|_F^2 + \alpha_1 \|\mathcal{S}\|_{\ell_1} + \alpha_2 \|\mathcal{V}_1\|_{\ell_1} \\ & \quad + \alpha_2 \|\mathcal{V}_2\|_{\ell_1} + \alpha_3 \|\mathcal{V}_3\|_{\ell_1} \\ & \quad + \alpha_3 \|\mathcal{V}_4\|_{\ell_1} + \frac{\mu}{2} \|\nabla_x \mathcal{X} - \mathcal{V}_1\|_F^2 + \frac{\mu}{2} \|\nabla_y \mathcal{X} - \mathcal{V}_2\|_F^2 \\ & \quad + \frac{\mu}{2} \|\nabla_x(\nabla_z \mathcal{X}) - \mathcal{V}_3\|_F^2 + \frac{\mu}{2} \|\nabla_y(\nabla_z \mathcal{X}) - \mathcal{V}_4\|_F^2 \\ & \quad + \langle \Lambda_1, \nabla_x \mathcal{X} - \mathcal{V}_1 \rangle + \langle \Lambda_2, \nabla_y \mathcal{X} - \mathcal{V}_2 \rangle \\ & \quad + \langle \Lambda_3, \nabla_x(\nabla_z \mathcal{X}) - \mathcal{V}_3 \rangle + \langle \Lambda_4, \nabla_y(\nabla_z \mathcal{X}) - \mathcal{V}_4 \rangle\end{aligned}$$

where μ is the penalty parameter, Λ_i ($i = 1, 2, 3, 4$) are multipliers, and $\mathcal{X}$ is defined as in (2). The joint minimization problem can be decomposed into easier subproblems, followed by the update of Lagrangian multipliers.

The $\mathcal{V}_i$ ($i = 1, 2, 3, 4$) subproblems are

$$\begin{cases} \min\limits_{\mathcal{V}_1} \dfrac{\mu}{2} \|\nabla_x \mathcal{X}^t + \dfrac{\Lambda_1^t}{\mu} - \mathcal{V}_1\|_F^2 + \alpha_2 \|\mathcal{V}_1\|_{\ell_1} \\ \min\limits_{\mathcal{V}_2} \dfrac{\mu}{2} \|\nabla_y \mathcal{X}^t + \dfrac{\Lambda_2^t}{\mu} - \mathcal{V}_2\|_F^2 + \alpha_2 \|\mathcal{V}_2\|_{\ell_1} \\ \min\limits_{\mathcal{V}_3} \dfrac{\mu}{2} \|\nabla_x(\nabla_z \mathcal{X}^t) + \dfrac{\Lambda_3^t}{\mu} - \mathcal{V}_3\|_F^2 + \alpha_3 \|\mathcal{V}_3\|_{\ell_1} \\ \min\limits_{\mathcal{V}_4} \dfrac{\mu}{2} \|\nabla_y(\nabla_z \mathcal{X}^t) + \dfrac{\Lambda_4^t}{\mu} - \mathcal{V}_4\|_F^2 + \alpha_3 \|\mathcal{V}_4\|_{\ell_1} \end{cases}$$

which can be exactly solved by $\mathcal{V}_1^{t+1} = \text{Soft}_{\alpha_2/\mu}(\nabla_x \mathcal{X}^t + (\Lambda_1^t/\mu))$, $\mathcal{V}_2^{t+1} = \text{Soft}_{\alpha_2/\mu}(\nabla_y \mathcal{X}^t + (\Lambda_2^t/\mu))$, $\mathcal{V}_3^{t+1} = \text{Soft}_{\alpha_3/\mu}(\nabla_x(\nabla_z \mathcal{X}^t) + (\Lambda_3^t/\mu))$, and $\mathcal{V}_4^{t+1} = \text{Soft}_{\alpha_3/\mu}(\nabla_y(\nabla_z \mathcal{X}^t) + (\Lambda_4^t/\mu))$, where $\big(\text{Soft}_v(\mathcal{X})\big)_{(i,j,k)} := \text{sign}(\mathcal{X}_{(i,j,k)}) \max\{|\mathcal{X}_{(i,j,k)}| - v, 0\}$.

The $\mathcal{S}$ subproblem is $\min_{\mathcal{S}} \|\mathcal{Y} - \mathcal{X}^t - \mathcal{S}\|_F^2 + \alpha_1 \|\mathcal{S}\|_{\ell_1}$, which can be exactly solved by $\mathcal{S}^{t+1} = \text{Soft}_{\alpha_1/2}(\mathcal{Y} - \mathcal{X}^t)$.

The $\mathcal{X}$ subproblem is

$$\begin{aligned}\min_{\mathcal{X}} & \|\mathcal{Y} - \mathcal{X} - \mathcal{S}^t\|_F^2 \\ & + \frac{\mu}{2} \Big(\big\|\nabla_x \mathcal{X} - \mathcal{D}_1^t\big\|_F^2 + \big\|\nabla_y \mathcal{X} - \mathcal{D}_2^t\big\|_F^2 \\ & \quad + \big\|\nabla_x(\nabla_z \mathcal{X}) - \mathcal{D}_3^t\big\|_F^2 + \big\|\nabla_y(\nabla_z \mathcal{X}) - \mathcal{D}_4^t\big\|_F^2 \Big)\end{aligned}$$

where $\mathcal{D}_i^t := \mathcal{V}_i^t - (\Lambda_i^t/\mu)$ ($i = 1, 2, 3, 4$) and $\mathcal{X}$ is parameterized by $\{\mathcal{W}_d\}_{d=1}^{l}$ and $\{\mathbf{H}_p\}_{p=1}^{m}$, as presented in (2). To tackle the nonlinear and nonconvex $\mathcal{X}$ subproblem, we apply the adaptive moment estimation (Adam) algorithm [28]. In each iteration of the ADMM-based algorithm, we use one step of the Adam to update $\{\mathcal{W}_d\}_{d=1}^{l}$ and $\{\mathbf{H}_p\}_{p=1}^{m}$.

Finally, the Lagrange multipliers are updated by $\Lambda_1^{t+1} = \Lambda_1^t + \mu(\nabla_x \mathcal{X}^t - \mathcal{V}_1^t)$, $\Lambda_2^{t+1} = \Lambda_2^t + \mu(\nabla_y \mathcal{X}^t - \mathcal{V}_2^t)$, $\Lambda_3^{t+1} = \Lambda_3^t + \mu(\nabla_x(\nabla_z \mathcal{X}^t) - \mathcal{V}_3^t)$, and $\Lambda_4^{t+1} = \Lambda_4^t + \mu(\nabla_y(\nabla_z \mathcal{X}^t) - \mathcal{V}_4^t)$.

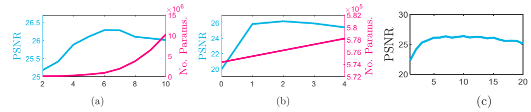

Fig. 4. Results on *Beads* Case 5 with (a) different layer numbers of HMF, (b) different layer numbers of HNT, and (c) different sizes of factor tensors.

We use the maximum iteration number (i.e., 500) as the stop criterion of the ADMM-based algorithm.

III. Experiments

A. Experimental Settings

We compare H2TF with the model-based methods LRTDTV [29], SSTV-LRTF [12], RCTV [5], and HLRTF [14] and deep learning methods HSID-CNN [10] and SDeCNN [9]. We use the pretrained models of HSID-CNN and SDeCNN provided by authors. All the hyperparameters of these methods are carefully adjusted to achieve the best results. We report the peak-signal-to-noise-ratio (PSNR) and structural similarity (SSIM).

We include four HSIs and three multispectral images (MSIs) as simulated datasets. The HSIs are *WDC* ($256 \times 256 \times 32$), *PaviaC* ($256 \times 256 \times 32$), *PaviaU* ($256 \times 256 \times 32$), and *Indian* ($145 \times 145 \times 32$). The MSIs are *Beads* ($256 \times 256 \times 31$), *Cloth* ($256 \times 256 \times 31$), and *Cups* ($256 \times 256 \times 31$) in the CAVE dataset [30]. The noise settings of simulated data are explained as below. **Case 1**: All the bands are added with Gaussian noise of standard deviation 0.2. **Case 2**: The Gaussian noise for Case 1 is kept. Besides, all the bands are added with impulse noise with sampling rate 0.1. **Case 3**: The same as Case 2 plus 50% of bands corrupted by deadlines. The number of deadlines for each chosen band is generated randomly from 6 to 10, and their spatial width is chosen randomly from 1 to 3. **Case 4**: The same as Case 2 plus 40% of bands corrupted by stripes. The number of stripes in each corrupted band is chosen randomly from 6 to 15. **Case 5**: The same as Case 2 plus both the deadlines in Case 3 and the stripes in Case 4. To test our method in real scenarios, we choose two real-world noisy HSIs *Shanghai* ($300 \times 300 \times 32$) and *Urban* ($307 \times 307 \times 32$) as real-world experimental datasets.

B. Experimental Results

1) Results: The quantitative results on simulated data are reported in Table I. Our H2TF obtains better quantitative results than other competitors with acceptable running time. H2TF outperforms other TV and tensor-factorization-based methods (LRTDTV, SSTV-LRTF, RCTV, and HLRTF), which shows the stronger representation abilities of H2TF than the existing shallow tensor factorizations, thanks to the hierarchical structures of H2TF. Some visual results on simulated and real data are shown in Figs. 2 and 3. We can observe that H2TF can more effectively remove heavy mixed noise. Also, H2TF preserves fine details of HSIs better than other methods. The superior performances of H2TF are mainly due to its hierarchical modeling abilities, which help better characterize fine details of HSI and robustly capture the underlying structures of HSI under heavy mixed noise.

2) Discussions: The HMF is an important building block in H2TF. We test the influence of the layer number of HMF (i.e., l); see Fig. 4(a). A suitable layer number of HMF (e.g., $l = 5$) can obtain both good performances and a lightweight model. The HNT is another important building

TABLE I

AVERAGE QUANTITATIVE DENOISING RESULTS AND AVERAGE RUNNING TIME (SECONDS) BY DIFFERENT METHODS

Dataset	Method	Case 1			Case 2			Case 3			Case 4			Case 5		
		PSNR	SSIM	Time	PSNR	SSIM	Time	PSNR	SSIM	Time	PSNR	SSIM	Time	PSNR	SSIM	Time
HSIs	LRTDTV	30.88	0.888	28.70	29.55	0.849	30.33	28.29	0.825	30.30	29.35	0.843	29.36	28.31	0.824	29.99
	SSTV-LRTF	30.77	0.887	17.96	30.35	0.879	18.15	28.32	0.839	16.59	29.51	0.859	18.16	27.42	0.812	17.60
WDC	HSID-CNN	29.61	0.863	153.93	22.89	0.691	154.29	21.98	0.661	154.83	22.16	0.669	155.65	21.22	0.635	155.66
PaviaC	SDeCNN	30.26	0.873	13.43	23.97	0.735	13.99	23.33	0.725	14.98	23.41	0.723	14.73	22.63	0.714	15.06
PaviaU	RCTV	29.53	0.853	**12.26**	29.05	0.839	**11.02**	26.75	0.782	**12.64**	28.47	0.826	**12.81**	26.34	0.772	**12.30**
Indian	HLRTF	30.12	0.868	15.40	29.65	0.855	15.37	29.58	0.853	14.67	29.15	0.846	15.20	29.03	0.841	14.23
	H2TF	**32.51**	**0.919**	15.66	**31.41**	**0.900**	14.31	**31.34**	**0.899**	14.80	**30.83**	**0.893**	14.34	**30.84**	**0.892**	14.44
	LRTDTV	28.85	0.889	35.95	27.05	0.838	35.69	26.31	0.828	34.94	26.83	0.830	36.21	26.13	0.819	34.38
MSIs	SSTV-LRTF	27.64	0.878	21.10	27.48	0.864	20.99	26.25	0.855	21.76	26.79	0.844	21.82	25.11	0.823	20.60
	HSID-CNN	25.86	0.827	185.50	21.22	0.660	185.42	20.97	0.645	185.82	20.68	0.646	185.94	20.34	0.626	186.03
Beads	SDeCNN	28.43	0.886	16.73	22.04	0.715	17.53	22.32	0.709	17.49	21.53	0.706	16.50	21.70	0.698	16.41
Cloth	RCTV	28.15	0.869	**13.26**	27.49	0.866	**14.18**	25.77	0.839	**12.95**	26.98	0.854	**13.44**	25.46	0.829	**12.71**
Cups	HLRTF	29.21	0.884	15.68	28.73	0.886	14.69	28.67	0.884	15.19	28.10	0.870	15.57	28.03	0.868	15.96
	H2TF	**31.51**	**0.940**	18.47	**29.46**	**0.906**	17.65	**29.47**	**0.901**	16.83	**29.22**	**0.908**	16.85	**29.03**	**0.896**	17.07

block. We change the layer number of HNT to test its influence; see Fig. 4(b). Also, a proper layer number of HNT (e.g., $m = 2$) can bring good performances. According to Lemma 1, the sizes of factor tensors in HMF, i.e., $\{r_d\}_{d=1}^4$, determine the degree of low-rankness. Hence, we test such connections by changing the sizes of factor tensors; see Fig. 4(c). (Here, r_0 and r_5 are fixed as the sizes of observed data and $\{r_d\}_{d=1}^4$ are selected in $\{(1, 2, 4, 8), (2, 4, 8, 16), (3, 6, 12, 24), \ldots, (20, 40, 80, 160)\}$.) When the sizes (rank) are too small, the model lacks representation abilities, and when the sizes (rank) are too large, the model overfits. Nevertheless, our method is quite robust with respect to $\{r_d\}_{d=1}^4$.

IV. CONCLUSION

We propose the H2TF for HSI denoising. Our H2TF leverages the HMF and HNT to compactly represent HSIs with powerful representation abilities, which can more faithfully capture fine details of HSIs than the classical tensor factorization methods. Comprehensive experiments validate the superiority of H2TF over SOTA methods, especially for HSI details preserving and heavy noise removal.

REFERENCES

[1] N. Liu, W. Li, R. Tao, and J. E. Fowler, "Wavelet-domain low-rank/group-sparse destriping for hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 12, pp. 10310–10321, Dec. 2019.

[2] W. He, H. Zhang, L. Zhang, W. Philips, and W. Liao, "Weighted sparse graph based dimensionality reduction for hyperspectral images," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 5, pp. 686–690, May 2016.

[3] L. Zhuang and M. K. Ng, "FastHyMix: Fast and parameter-free hyperspectral image mixed noise removal," *IEEE Trans. Neural Netw. Learn. Syst.*, early access, Sep. 29, 2021, doi: 10.1109/TNNLS.2021.3112577.

[4] F. Xu, Y. Chen, C. Peng, Y. Wang, X. Liu, and G. He, "Denoising of hyperspectral image using low-rank matrix factorization," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 7, pp. 1141–1145, Jul. 2017.

[5] J. Peng, H. Wang, X. Cao, X. Liu, X. Rui, and D. Meng, "Fast noise removal in hyperspectral images via representative coefficient total variation," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5546017, doi: 10.1109/TGRS.2022.3229012.

[6] L. Zhuang, L. Gao, B. Zhang, X. Fu, and J. M. Bioucas-Dias, "Hyperspectral image denoising and anomaly detection based on low-rank and sparse representations," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5500117.

[7] W. He et al., "Non-local meets global: An iterative paradigm for hyperspectral image restoration," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 4, pp. 2089–2107, Apr. 2022.

[8] Z. Wang, M. K. Ng, L. Zhuang, L. Gao, and B. Zhang, "Nonlocal self-similarity-based hyperspectral remote sensing image denoising with 3-D convolutional neural network," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5531617.

[9] A. Maffei, J. M. Haut, M. E. Paoletti, J. Plaza, L. Bruzzone, and A. Plaza, "A single model CNN for hyperspectral image denoising," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 4, pp. 2516–2529, Apr. 2020.

[10] Q. Yuan, Q. Zhang, J. Li, H. Shen, and L. Zhang, "Hyperspectral image denoising employing a spatial–spectral deep residual convolutional neural network," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 2, pp. 1205–1218, Feb. 2019.

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

[12] H. Fan, C. Li, Y. Guo, G. Kuang, and J. Ma, "Spatial–spectral total variation regularized low-rank tensor decomposition for hyperspectral image denoising," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 10, pp. 6196–6213, Oct. 2018.

[13] J.-L. Wang, T.-Z. Huang, X.-L. Zhao, T.-X. Jiang, and M. K. Ng, "Multi-dimensional visual data completion via low-rank tensor representation under coupled transform," *IEEE Trans. Image Process.*, vol. 30, pp. 3581–3596, 2021.

[14] Y. Luo, X. Zhao, D. Meng, and T. Jiang, "HLRTF: Hierarchical low-rank tensor factorization for inverse problems in multi-dimensional imaging," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 19281–19290.

[15] 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 Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 4, pp. 925–938, Apr. 2020.

[16] C. Lu, X. Peng, and Y. Wei, "Low-rank tensor completion with a new tensor nuclear norm induced by invertible linear transforms," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 5989–5997.

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

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

[19] Y. Zheng and A.-B. Xu, "Tensor completion via tensor QR decomposition and $L_{2,1}$-norm minimization," *Signal Process.*, vol. 189, Dec. 2021, Art. no. 108240.

[20] P. Zhou, C. Lu, Z. Lin, and C. Zhang, "Tensor factorization for low-rank tensor completion," *IEEE Trans. Image Process.*, vol. 27, no. 3, pp. 1152–1163, Mar. 2018.

[21] S. Arora, N. Cohen, W. Hu, and Y. Luo, "Implicit regularization in deep matrix factorization," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 32, 2019, pp. 7413–7424.

[22] Z. Li, T. Sun, H. Wang, and B. Wang, "Adaptive and implicit regularization for matrix completion," *SIAM J. Imag. Sci.*, vol. 15, no. 4, pp. 2000–2022, Dec. 2022.

[23] J. Fan and J. Cheng, "Matrix completion by deep matrix factorization," *Neural Netw.*, vol. 98, pp. 34–41, Feb. 2018.

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

[25] M. E. Kilmer, K. Braman, N. Hao, and R. C. Hoover, "Third-order tensors as operators on matrices: A theoretical and computational framework with applications in imaging," *SIAM J. Matrix Anal. Appl.*, vol. 34, no. 1, pp. 148–172, Jan. 2013.

[26] X.-Y. Liu, S. Aeron, V. Aggarwal, and X. Wang, "Low-tubal-rank tensor completion using alternating minimization," *IEEE Trans. Inf. Theory*, vol. 66, no. 3, pp. 1714–1737, Mar. 2020.

[27] S. Takeyama, S. Ono, and I. Kumazawa, "Mixed noise removal for hyperspectral images using hybrid spatio-spectral total variation," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2019, pp. 3128–3132.

[28] D. P. Kingma and J. L. Ba, "ADAM: A method for stochastic optimization," in *Proc. Int. Conf. Learn. Represent.*, 2014, pp. 1–15.

[29] Y. Wang, J. Peng, Q. Zhao, Y. Leung, X.-L. Zhao, and D. Meng, "Hyperspectral image restoration via total variation regularized low-rank tensor decomposition," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 4, pp. 1227–1243, Apr. 2018.

[30] F. Yasuma, T. Mitsunaga, D. Iso, and S. K. Nayar, "Generalized assorted pixel camera: Postcapture control of resolution, dynamic range, and spectrum," *IEEE Trans. Image Process.*, vol. 19, no. 9, pp. 2241–2253, Sep. 2010.