

3DeepRep: 3D deep low-rank tensor representation for hyperspectral image inpainting[☆]

Yunshan Li^{a, b, 1} , Wenwu Gong^{b, 1}, Qianqian Wang^b, Chao Wang^{b, *}, Lili Yang^{a, b, **}

^a Shenzhen Key Laboratory of Safety and Security for Next Generation of Industrial Internet, Southern University of Science and Technology, Shenzhen, 518055, Guangdong, China

^b Department of Statistics and Data Science, Southern University of Science and Technology, Shenzhen, 518055, Guangdong, China

ARTICLE INFO

Communicated by N. Zeng

Keywords:

Deep transform
3-directional tensor nuclear norm (3DTNN)
Hyperspectral image inpainting
Low-rank tensor representation
Self-supervised learning

ABSTRACT

Hyperspectral image (HSI) inpainting is a fundamental challenge in remote sensing. However, existing transform-based low-rank methods face significant limitations: they often rely on unstable generator-based latent characterization and suffer from directional inflexibility by applying transformations solely along the spectral dimension. To address these issues, we propose 3-directional Deep Low-rank Representation (3DeepRep), a fully learning-guided model that bridges low-rank tensor modeling with deep representation learning. 3DeepRep introduces three key innovations: (1) Generator-Free Optimization: We treat latent tensors as explicit learnable variables under nuclear norm constraints, eliminating the parameter redundancy and optimization instability associated with generator networks; (2) Multi-directional Deep Transform: We extend the coupled transform block to all three tensor modes, enabling the comprehensive exploitation of spatial-spectral correlations; and (3) Learnable Aggregation: A data-driven module adaptively fuses directional reconstructions, eliminates the need for manual hyperparameter tuning. The entire framework is unified within a fully gradient-driven self-supervised optimization scheme. Extensive experiments on three benchmark datasets validate the effectiveness of our approach. 3DeepRep consistently outperforms state-of-the-art methods, achieving remarkable PSNR gains of up to 7.553 dB in complex degradation scenarios.

1. Introduction

Hyperspectral images (HSIs) play a pivotal role in characterizing material properties and environmental dynamics by capturing rich spectral information across hundreds of contiguous bands. Such fine-grained spectral fidelity has advanced diverse domains ranging from urban planning [1,2] and agriculture [3] to industrial intelligence [4,5]. However, in practical acquisition scenarios, HSIs are inevitably degraded by sensor limitations, atmospheric interference (e.g., clouds, haze), and transmission errors. These degradations disrupt the spatial-spectral continuity, severely affect downstream tasks such as target detection and classification [6–8]. Consequently, recovering the underlying high-quality data

from corrupted observations is a crucial task for reliable hyperspectral analysis [9].

Fundamentally, an HSI is modeled as a third-order tensor $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, where the spatial dimensions and the spectral dimension are interacted. The key to successful restoration lies in exploiting the high redundancy inherent among spectral and spatial dimensions, the strong global correlation across spectral bands and the non-local self-similarity within spatial domains [10,11]. This structural redundancy provides the mathematical insight for recovering missing information from partial observations, motivating the extensive use of low-rank representation in HSI inpainting tasks.

[☆] This research was funded by the SUSTech Presidential Postdoctoral Fellowship, the China Postdoctoral Science Foundation (Grant No. 2025M773057), Shenzhen Science and Technology Program (Grant No. ZDSYS20210623092007023), the Guangdong Province Universities and Colleges Key Areas of Special Projects (Grant No. 2021222012), and the Shenzhen Key Laboratory of Safety and Security for Next Generation of Industrial Internet, Southern University of Science and Technology.

^{*} Corresponding author.

^{**} Corresponding author at: Shenzhen Key Laboratory of Safety and Security for Next Generation of Industrial Internet, Southern University of Science and Technology, Shenzhen, 518055, Guangdong, China.

Email addresses: 12331105@mail.sustech.edu.cn (Y. Li), gongww@sustech.edu.cn (W. Gong), 12331103@mail.sustech.edu.cn (Q. Wang), wangc6@sustech.edu.cn (C. Wang), yangll@sustech.edu.cn (L. Yang).

¹ Co-first authors.

<https://doi.org/10.1016/j.neucom.2026.133160>

Received 29 December 2025; Received in revised form 10 February 2026; Accepted 24 February 2026

Available online 25 February 2026

0925-2312/© 2026 Elsevier B.V. All rights are reserved, including those for text and data mining, AI training, and similar technologies.

Early approaches extended matrix rank concepts to tensors. Typical examples include CP [12], Tucker decompositions [13,14], and tensor network representations [15–18]. However, these methods struggle to balance computational complexity with flexible dependency modeling. To address this, the tensor tubal rank was introduced based on the tensor singular value decomposition (t-SVD) [19]. This approach demonstrates superior capability in capturing spatial–spectral correlations [20,21]. Since directly minimizing the tubal rank is NP-hard, the Tensor Nuclear Norm (TNN) is adopted as a convex surrogate [22]. TNN is defined as the sum of the singular values of frontal slices in a transform domain. A standard TNN-based restoration model is formulated as:

$$\min_{\hat{\mathcal{X}}} \sum_k \|\hat{\mathcal{X}}^{(k)}\|_* + \lambda \mathcal{L}(\mathcal{X}, \mathcal{O}) \quad \text{s.t. } \hat{\mathcal{X}} = f(\mathcal{X}), \quad (1)$$

where $f(\cdot)$ represents a unitary transform (traditionally the DFT). This transform maps the data into a domain where the low-rank property is more compact.

Although traditional TNN models offer theoretical interpretability, their efficacy is inherently constrained by the linearity of fixed transforms (e.g., DFT [23], DCT [24], or Framelets [25]). Such linear mappings often fail to adequately characterize the complex, non-linear distributions intrinsic to real-world HSI data [26,27]. Conversely, deep learning approaches excel at feature extraction but lack explicit structural constraints. To bridge this gap, the current research has shifted towards integrating low-rank tensor modeling with deep representation learning. This hybrid framework aims to combine the mathematical reliability of low-rank priors with the non-linear representational power of neural networks [28]. Instead of relying on fixed bases, recent state-of-the-art methods replace the linear transform $f(\cdot)$ in Eq. (1) with learnable, non-linear deep networks. For instance, S2TNN [29] introduced deep fully connected networks to learn spectral transformations, while CoNoT [30] further advanced this by coupling spatial convolutional networks with spectral transforms to capture intricate spatial-spectral dependencies. By embedding deep networks into the iterative low-rank optimization framework, these methods achieve a Deep Transform, enabling the model to adaptively learn a low-rank latent tensor representation. This evolution marks a transition from handcrafted linear assumptions to data-driven non-linear representations, setting the stage for high-fidelity HSI inpainting.

1.1. Motivation

Despite the promising integration of deep learning with low-rank tensor representation, existing approaches suffer from high-fidelity HSI restoration. We summarize these challenges into three aspects:

Inefficiency and Instability in Latent Representation. Most current deep tensor methods rely on networks to generate low-rank latent tensors. As noted in recent studies, this dependency introduces substantial parameter redundancy and cause the sensitivity of the optimization process to the network’s inductive biases and random initialization. In self-supervised settings, this often results in optimization instability, making it difficult to converge to a reliable low-rank representation.

Restricted Directional Modeling. The majority of existing models apply transformations exclusively along a single tensor mode (typically the spectral dimension). This directional inflexibility neglects the intrinsic low-rank structures inherent in the spatial modes. Although some approaches like 3DTNN [31] and 3DSTNN [32] attempt to extend transforms to multiple modes, they typically employ predefined bases (lacking data adaptivity). Crucially, they fail to implement a fully coupled transform mechanism (such as the spatial-spectral interaction in [30]) across all dimensions, thereby failing to exploit the joint spatial-spectral correlations.

Inflexible Fusion Strategies. Since multi-directional transformations yield distinct reconstruction candidates, effective fusion is important. However, existing methods rely on heuristic strategies, such as minimizing a weighted sum of nuclear norms. These global weighting schemes

lack the adaptivity required to handle local variations in degradation intensity. A fixed set of global weights cannot distinguish valid structural information, often leading to suboptimal fusion where high-quality local details are influenced by lower-quality candidates.

Overall, existing learning-guided low-rank tensor representation methods suffer from two key limitations: (1) they are *ineffective and unstable* to capture coupled spatial-spectral correlations across all tensor modes, and (2) they lack an *adaptive* learning mechanism to balance directional contributions during reconstruction. These limitations motivate a multi-directional deep low-rank tensor representation equipped with a learnable aggregation module, which enables comprehensive hyperspectral correlation modeling, improved directional flexibility, and more robust recovery under diverse degradation patterns.

1.2. Contribution

To address these issues, we propose 3-directional deep low-rank tensor representation (**3DeepRep**), a fully learning-guided method to bridge the gap between low-rank tensor modeling and deep representation learning. As illustrated in Fig. 2, our proposal initializes three explicit latent tensors $\{\hat{\mathcal{X}}_i\}_{i=1}^3$, maps them to the pixel space via multi-directional deep transforms, and fuses the resulting image candidates $\{\mathcal{X}_i\}_{i=1}^3$ through a learnable aggregation module to produce the final recovered image. Fig. 1 demonstrates the superiority of 3DeepRep over single-mode baselines (e.g., CoNoT [30]). The main contributions are summarized as follows:

- **Multi-directional deep representation with generator-free optimization.** We extend the deep spatial-spectral transform to all three tensor modes, enabling a comprehensive exploitation of low-rank structures and correlations across different dimensions. Crucially, we introduce a generator-free strategy that treats latent tensors $\{\hat{\mathcal{X}}_i\}_{i=1}^3$ as explicit learnable variables. This design reduces parameter redundancy compared to generator-based schemes while avoiding the optimization instability.
- **Adaptive fusion via learnable aggregation module.** We propose a deep aggregation module to replace the heuristic weighting

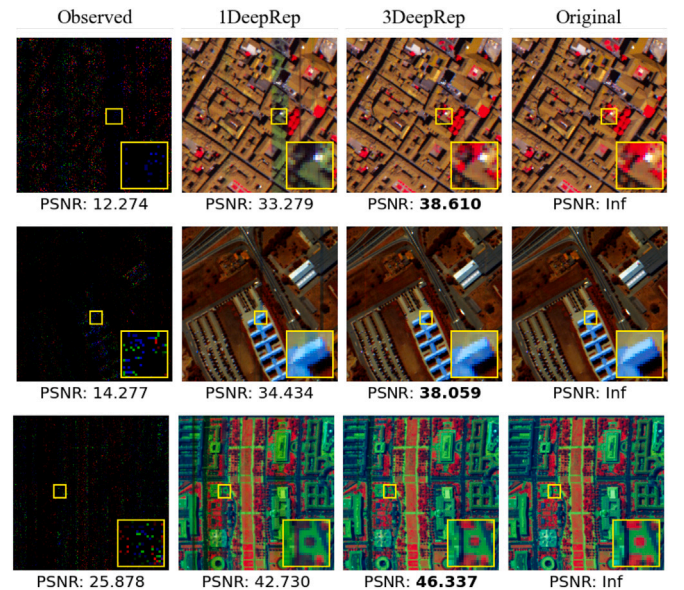


Fig. 1. Qualitative and quantitative comparison of HSI inpainting methods: including 1DeepRep and the proposed 3DeepRep. All the HSI images are false-color images composed of bands (R: 70, G: 40, and B: 10). From top to bottom, we report results on *Pavia Centre*, *Pavia University*, and *Washington DC Mall*, respectively. The missing-data setting follows a mixed corruption scenario that combines point missing, stripe missing, and deadline missing (see Section 5.2.1).

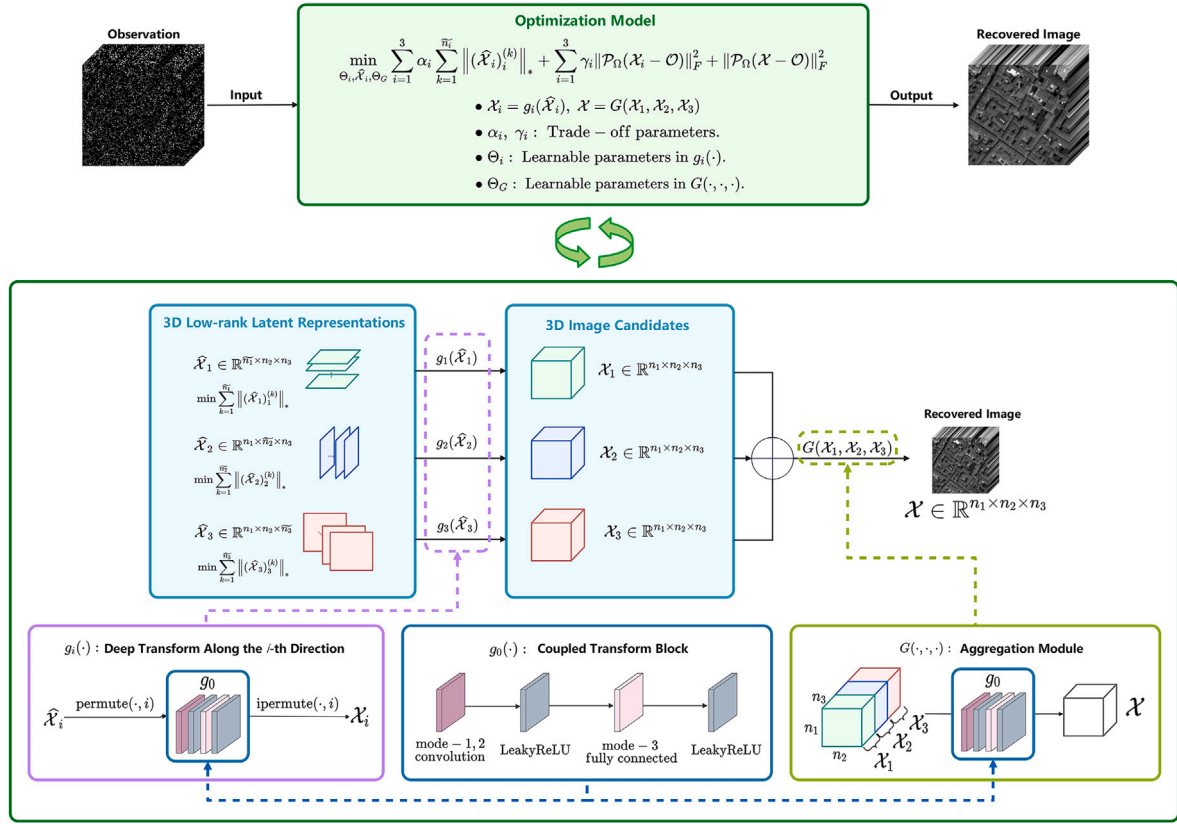


Fig. 2. Flowchart of the proposed 3DeepRep framework for HSI inpainting. The process initiates with three explicit Latent Variables ($\{\hat{\mathcal{X}}_i\}_{i=1}^3$), corresponding to the three tensor directions. Low-rank structures are enforced on these variables by minimizing the nuclear norms of their respective frontal slices. Subsequently, these latent representations are mapped to the pixel space via multi-directional deep transforms (implemented by Coupled Transform Blocks), yielding three Image Candidates ($\{\mathcal{X}_i\}_{i=1}^3$). Finally, these directional intermediates are adaptively fused by a learnable Aggregation Module to produce the Recovered Image ($\mathcal{X}$).

schemes used in traditional multi-directional methods. By modeling the non-linear spatial-spectral interactions among directional reconstructions, this data-driven fusion enhances recovery quality and eliminates the dependence on handcrafted hyperparameters for balancing directional contributions.

- **Unified gradient-driven self-supervised learning.** We develop a fully gradient-driven optimization strategy that enables the direct backpropagation of transformation weights, aggregation parameters, and latent variables in a self-supervised manner.

The remainder of this paper is organized as follows. Section 2 reviews related work. Section 3 presents some preliminaries. Section 4 proposes the 3DeepRep method step by step. Section 5 presents the experiment results and a discussion of the model. Section 6 concludes this paper.

2. Related work

2.1. Deep low-rank latent tensor-based methods

Deep low-rank latent tensor-based methods aim to bridge the gap between traditional low-rank modeling and deep representation learning. These approaches typically recover the target tensor by mapping a low-rank latent tensor representation to the image space via a deep neural network. Formally, these methods can be formulated as:

$$\min_{\mathcal{X}} L(\mathcal{X}, \mathcal{O}) \quad \text{s.t. } \mathcal{X} = g(\hat{\mathcal{X}}), \quad \hat{\mathcal{X}} \text{ is low-rank,}$$

where $g(\cdot)$ denotes a deep neural network parameterized by θ , mapping the low-rank latent tensor $\hat{\mathcal{X}}$ to the image space.

A core divergence in existing literature lies in the generation mechanism of the low-rank latent tensor $\hat{\mathcal{X}}$. Early approaches treated $\hat{\mathcal{X}}$ as a set

of *explicit learnable variables* directly optimized under factorization constraints. For instance, Bacca et al. [33] imposed a Tucker decomposition on the latent variables, HLRTF [34] and D-FCTN [35] utilized fully connected tensor network factorizations. While parameter-efficient, these methods are often constrained by the multilinear decompositions, limiting their adaptability to real-world data distributions.

To enhance representational capacity, recent trends have shifted towards *implicit generation* strategies (i.e., generator-based methods), where $\hat{\mathcal{X}}$ is generated from random noise by a neural network. S2NTNN [29] and CoNoT [30] employ fully connected networks or coupled non-linear transforms to model the latent space. Although these generator-based designs introduce powerful inductive biases, they suffer from significant parameter redundancy and optimization instability due to the “black-box” mapping from noise to the latent space.

Regardless of the generation strategy, a critical challenge in most aforementioned methods is *directional inflexibility*. Previous methods apply the deep transform solely along the spectral mode (Mode-3) [29,30]. This mode-specific strategy fails to fully exploit the spatial-spectral correlations inherent in HSIs, limiting the model’s ability to capture complex structures across different dimensions. Consequently, there is need for a multi-directional design that extends deep transforms to all tensor modes to achieve comprehensive low-rank representation.

2.2. 3D low-rank tensor-based methods

To address the directional inflexibility issue, 3D low-rank tensor-based methods extend low-rank tensor representation to all tensor modes:

$$\min_{\mathcal{X}} L(\mathcal{X}, \mathcal{O}) \quad \text{s.t. } \mathcal{X} = g_i(\hat{\mathcal{X}}_i), \quad \hat{\mathcal{X}}_i \text{ is low-rank, } i = 1, 2, 3,$$

Table 1
Notations.

Notation	Explanation
$x, \mathbf{x}, \mathbf{X}, \mathcal{X}$	Scalar, vector, matrix, tensor.
$\mathcal{X}(l, j, s)$	The (l, j, s) -th element of $\mathcal{X}$.
$\hat{\mathcal{X}}_i$	Latent low-rank tensor under the i -th direction, $i \in \{1, 2, 3\}$.
$\mathcal{X}_i$	Intermediate directional reconstruction from the i -th direction.
$(\hat{\mathcal{X}}_1)_1^{(k)}$	Mode-1 frontal slice [36] of $\hat{\mathcal{X}}_1$: $\hat{\mathcal{X}}_1(k, :, :)$.
$(\hat{\mathcal{X}}_2)_2^{(k)}$	Mode-2 frontal slice [36] of $\hat{\mathcal{X}}_2$: $\hat{\mathcal{X}}_2(:, k, :)$.
$(\hat{\mathcal{X}}_3)_3^{(k)}$	Mode-3 frontal slice [36] of $\hat{\mathcal{X}}_3$: $\hat{\mathcal{X}}_3(:, :, k)$.
$\ \mathbf{X}\ _*$	The matrix nuclear norm of $\mathbf{X}$.
$\ \mathcal{X}\ _F$	Frobenius norm: $\sqrt{\sum_{i,j,s} \mathcal{X}_{i,j,s}^2}$.
$\times_i$	Mode- i tensor-matrix product.
$\otimes$	2D convolution operation.
$w_1(\cdot)$	2D convolutional transform.
$w_2(\cdot)$	Fully-connected (FC) transform.
$g_i(\cdot)$	Directional deep transform.
$G(\cdot, \cdot, \cdot)$	Learnable aggregation module.

where $g_i(\cdot)$ denotes the neural network (i.e., deep transform) along the i -th mode, mapping the corresponding low-rank latent tensor $\hat{\mathcal{X}}_i$ into the image domain. These approaches treat the recovered image as a fusion of reconstructions from multiple views, formulated as $\mathcal{X} = \text{Fusion}(\{g_i(\hat{\mathcal{X}}_i)\}_{i=1}^3)$.

Traditional multi-directional methods, such as 3DTNN [36] and WSTNN [31], rely on a *global weighted average* strategy to fuse directional reconstructions. The fusion weights are typically manually tuned hyperparameters that remain fixed for the entire dataset. This “sum-of-weighted-parts” assumption is inherently heuristic, as it ignores the local variation of information reliability and cannot adaptively distinguish between valid structures and artifacts in different spatial regions. Although recent works like JRTQN [37] have introduced data-dependent linear mappings, they fundamentally rely on the Alternating Direction Method of Multipliers (ADMM) for optimization. The reliance on ADMM imposes a challenge to end-to-end learning: the complex variable splitting hinders the integration of deep non-linear aggregation modules. To overcome these limitations, a shift towards fully gradient-driven frameworks with *learnable aggregation modules* is essential, enabling the pixel-wise, adaptive fusion of multi-directional information.

3. Preliminaries

The notations used in this paper are presented in Table 1. In addition, we introduce the tensor mode- i permutation operator and concatenation function as follows.

Definition 1 (Tensor Mode- i Permutation [36]). Let $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ be a third-order tensor. For any $i \in \{1, 2, 3\}$, we define the mode- i permutation of a tensor $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ as

$$\text{permute}(\mathcal{X}, i) \in \mathbb{R}^{p_1 \times p_2 \times p_3},$$

where (p_1, p_2, p_3) is a permutation of (n_1, n_2, n_3) satisfying $p_3 = n_i$, and the relative order of the remaining two indices is preserved. The element-wise correspondence between $\mathcal{X}$ and $\text{permute}(\mathcal{X}, i)$ is expressed as

$$\begin{aligned} \mathcal{X}(l, j, s) &= \text{permute}(\mathcal{X}, 1)(j, s, l) \\ &= \text{permute}(\mathcal{X}, 2)(l, s, j) \\ &= \text{permute}(\mathcal{X}, 3)(l, j, s), \\ l &\in [1, n_1], j \in [1, n_2], s \in [1, n_3]. \end{aligned} \quad (2)$$

Note that $\text{permute}(\mathcal{X}, 3) = \mathcal{X}$. The inverse operation of $\text{permute}(\cdot, i)$ is denoted by $\text{ipermute}(\cdot, i)$, $i \in \{1, 2, 3\}$.

Definition 2 (Concatenation Function). Let $\mathcal{X}_1, \mathcal{X}_2, \mathcal{X}_3 \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ be three tensors of identical dimensions. Their concatenation along the

third mode is defined as

$$\text{concat}(\mathcal{X}_1, \mathcal{X}_2, \mathcal{X}_3)(l, j, s) = \begin{cases} \mathcal{X}_1(l, j, s), & 1 \leq s \leq n_3, \\ \mathcal{X}_2(l, j, s - n_3), & n_3 < s \leq 2n_3, \\ \mathcal{X}_3(l, j, s - 2n_3), & 2n_3 < s \leq 3n_3, \end{cases} \quad (3)$$

where $\text{concat}(\mathcal{X}_1, \mathcal{X}_2, \mathcal{X}_3) \in \mathbb{R}^{n_1 \times n_2 \times 3n_3}$.

Coupled Transform Block. Following the design in [30], the *Coupled Transform Block* (CTB) module effectively exploits dependencies among neighboring tubes. This module is designed to reconstruct the HSI data from the latent space, formulated as follows:

$$g_0(\cdot) = w_2(w_1(\cdot)), \quad (4)$$

where $w_1(\cdot)$ denotes a slice-wise convolutional transform applied to the frontal slices of latent tensor, and $w_2(\cdot)$ denotes a fully connected transform (implemented via 1×1 convolution) operating along the tensor tubes. We specifically utilize this module to reconstruct the hyperspectral image from the low-rank latent representation, ensuring the preservation of intrinsic spatial-spectral correlations. The learnable weights of this coupled reconstruction transform are denoted by Θ .

4. Methodology

In this section, we propose 3DeepRep, a fully learning-guided tensor recovery framework designed for high-quality HSI inpainting. Unlike traditional approaches that rely on latent tensor characterization or complex ADMM solvers, 3DeepRep bridges the gap between low-rank tensor modeling and deep representation learning within a unified, differentiable optimization problem. As illustrated in Fig. 2, the proposed framework consists of three integral components: (1) A *Generator-Free Latent Representation* strategy that treats latent tensors as explicit learnable variables subject to nuclear norm constraints, thereby ensuring parameter efficiency and optimization stability; (2) A *Multi-directional Deep Transform* that extends the CTB to all tensor modes, enabling the comprehensive exploitation of spatial-spectral correlations across different dimensions; and (3) A *Learnable Aggregation Module* that adaptively fuses directional reconstructions via a data-driven strategy, effectively eliminating the need for manual parameter tuning. Crucially, we introduce a *Fully Gradient-driven Optimization* strategy, allowing for the end-to-end joint optimization of network parameters and latent representations using standard backpropagation.

4.1. Generator-free latent representation

Existing deep tensor representation methods (e.g., DTR [38]) typically rely on generator networks to map random noise into latent tensors. This dependency introduces a substantial number of redundant parameters and renders the optimization process sensitive to the specific inductive biases and random initializations of the generator, often leading to instability in self-supervised settings.

To address these limitations, we propose a *Generator-Free* strategy that offers a direct and parameter-efficient optimization path. Instead of generating latent tensor via neural network, we treat the multi-directional latent tensors $\{\hat{\mathcal{X}}_i\}_{i=1}^3$ as explicit learnable variables. These variables are initialized directly and optimized jointly with the network parameters via backpropagation. To ensure the latent representations capture the underlying low-rank structures, we impose a nuclear norm constraint on their frontal slices:

$$\mathcal{R}(\hat{\mathcal{X}}_i) = \sum_{k=1}^{\tilde{n}_i} \left\| (\hat{\mathcal{X}}_i)_i^{(k)} \right\|_* \quad (5)$$

Remark 1 (Merits of Generator-Free Design). By decoupling the latent representation from a generator network, this design yields two

significant advantages: (1) **Parameter Efficiency and Regularization.** By eliminating generator networks, we substantially reduce the number of learnable network parameters compared to generator-based schemes (e.g., DTR [38]). This reduction minimizes model complexity and effectively mitigates the risk of overfitting in self-supervised settings (see Table 5). (2) **Optimization Stability.** Directly optimizing the latent tensors avoids the sensitivity to inductive biases and random initializations inherent in deep generator networks. This leads to more robust convergence and consistent reconstruction performance (demonstrated in Table 5).

4.2. Multi-directional deep transform

We employ the *Coupled Transform Block* (CTB) as the fundamental decoding module to map the latent tensors back to the original pixel space. To elucidate the reconstruction process, we first consider the transformation along the spectral dimension (Mode-3). Let $\hat{\mathcal{X}}_3 \in \mathbb{R}^{n_1 \times n_2 \times \tilde{n}_3}$ denote the latent tensor corresponding to the 3rd directional representation. The reconstruction begins with the spatial transform $w_1(\cdot)$, which applies a slice-wise 2D convolution to each frontal slice $(\hat{\mathcal{X}}_3)_3^{(k)}$ to recover local spatial information. This is mathematically expressed as:

$$w_1(\hat{\mathcal{X}}_3) = \sigma \left(\left\{ \mathbf{W}_1^k \otimes (\hat{\mathcal{X}}_3)_3^{(k)} \right\}_{k=1}^{\tilde{n}_3} \right),$$

where $\mathbf{W}_1^k$ represents the convolution kernel for the k -th slice, and $\otimes$ denotes the 2D convolution operator. Specifically, we utilize a 3×3 kernel with a padding and stride of 1, followed by a LeakyReLU activation function [39].

Subsequently, the spectral transform $w_2(\cdot)$ captures the inter-slice correlations by performing a mode-3 tensor-matrix product with a learnable projection matrix $\mathbf{W}_2 \in \mathbb{R}^{n_3 \times \tilde{n}_3}$. Let $\hat{\mathcal{Y}} = w_1(\hat{\mathcal{X}}_3)$ denote the spatially-refined intermediate tensor; the spectral transform yields:

$$w_2(\hat{\mathcal{Y}}) = \sigma \left(\hat{\mathcal{Y}} \times_3 \mathbf{W}_2 \right).$$

By combining these operations with the permutation operators defined in Definition 1, the reconstructed tensor $\mathcal{X}_3$ is generated from $\hat{\mathcal{X}}_3$ as follows:

$$\begin{aligned} \mathcal{X}_3 &= \text{ipermute} \left(w_2 \left(w_1(\hat{\mathcal{X}}_3) \right), 3 \right) \\ &= \text{ipermute} \left(g_0 \left(\text{permute}(\hat{\mathcal{X}}_3, 3) \right), 3 \right), \end{aligned} \quad (6)$$

where $g_0(\cdot)$ represents the CTB operation.

To fully exploit the low-rank structures embedded across all tensor modes and capture comprehensive multi-dimensional correlations, we extend the CTB from a single-mode operation to a multi-directional framework. Motivated by the multidimensional transformations in TNN methods [36], we generalize the CTB module $g_0(\cdot)$ to the i -th mode ($i = 1, 2, 3$) via:

$$\mathcal{X}_i = g_i(\hat{\mathcal{X}}_i) = \text{ipermute} \left(g_0 \left(\text{permute}(\hat{\mathcal{X}}_i, i) \right), i \right), i \in \{1, 2, 3\}. \quad (7)$$

which takes the latent tensor $\hat{\mathcal{X}}_i$ as input and outputs the intermediate image candidate $\mathcal{X}_i$.

Remark 2 (Mechanism of Directional Transforms). As formulated in Eq. (7), the operator $g_i(\cdot)$ applies the slice-wise convolutional transform $w_1(\cdot)$ to the slices orthogonal to mode- i to capture local intra-slice dependencies, while utilizing the fully-connected transform $w_2(\cdot)$ along mode- i to model global inter-slice (tube-wise) correlations. By generalizing the CTB to all three modes, this hierarchical design ensures that the model fully exploits the low-rank structures inherent in the data (see Table 6). This multi-directional strategy enables the nonlinear transforms to flexibly integrate strong local spatial correlations with global spectral dependencies, thereby achieving a more compact and robust low-rank representation.

4.3. Learnable aggregation module

Existing multi-directional low-rank tensor methods (e.g., WSTNN [36]) address inter-modal correlations by minimizing a weighted sum of nuclear norms. However, such heuristic strategies rely on handcrafted weights to balance mode-wise contributions. This dependence necessitates extensive parameter tuning and limits generalization, as optimal weights are highly sensitive to degradation patterns.

To overcome this limitation and adaptively integrate the complementary information from the three directional transformations, we propose a learnable *Aggregation Module*. Instead of a simple or heuristically weighted averaging, this module fuses the directional reconstructions $\{\mathcal{X}_i\}_{i=1}^3$ through a non-linear data-driven approach. Specifically, the final output $\mathcal{X}$ is generated by concatenating the intermediate tensors and passing them through a deep coupled transform:

$$\mathcal{X} = G(\mathcal{X}_1, \mathcal{X}_2, \mathcal{X}_3) = g_3 \left(\text{concat}(\mathcal{X}_1, \mathcal{X}_2, \mathcal{X}_3) \right), \quad (8)$$

where concat is defined in Definition 2. The aggregation function g_3 shares the architecture of the directional transform g_i (Eq. 7) and is parameterized by Θ_G . Specifically, the aggregation module $G(\cdot)$ processes the input image candidates through the following three steps:

1. **Channel Expansion:** The three directional reconstructions $\{\mathcal{X}_i\}_{i=1}^3 \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ are concatenated along the spectral mode (mode-3). This yields an intermediate tensor $\mathcal{X}_{\text{cat}} \in \mathbb{R}^{n_1 \times n_2 \times 3n_3}$, preserving all structural information from the directional candidates.
2. **Spatial Refinement (w_1):** A slice-wise 2D convolution (3×3 kernel, stride 1, padding 1) is applied to each of the $3n_3$ frontal slices of $\mathcal{X}_{\text{cat}}$. This operation refines local spatial features while maintaining the channel dimension:

$$\mathcal{X}_{\text{spatial}} = \text{LeakyReLU}(w_1(\mathcal{X}_{\text{cat}})) \in \mathbb{R}^{n_1 \times n_2 \times 3n_3}.$$

3. **Spectral Fusion and Dimension Reduction (w_2):** To fuse the expanded spectral information and recover the original dimensions, we apply a mode-3 fully connected transform $w_2(\cdot)$. This is implemented as a 1×1 convolution along the tubes, mapping the channel dimension from $3n_3$ to n_3 :

$$\mathcal{X} = \text{LeakyReLU}(\mathcal{X}_{\text{spatial}} \times_3 \mathbf{W}_{\text{agg}}),$$

where $\mathbf{W}_{\text{agg}} \in \mathbb{R}^{3n_3 \times n_3}$ is the learnable weight matrix responsible for adaptively weighting and merging the directional components.

We explicitly select spectral direction for aggregation to preserve spatial consistency. Aggregating along this dimension ensures that the fusion remains pixel-wise aligned, thereby maintaining spatial coherence (See Table 5).

Remark 3 (Superiority over Handcrafted Fusion). Unlike traditional ADMM-based frameworks [32,36,40] that rely on manually tuned weights to balance directional tensors, our proposal employs a fully learnable aggregation strategy. By replacing linear weighting with a nonlinear mapping, the proposed module can adaptively weigh the directional contributions based on local data characteristics. This data-driven approach not only eliminates hyperparameter tuning but also significantly enhances recovery fidelity by automatically distinguishing valid structural information from artifacts, as evidenced by the consistent superiority of 3DeepRep over 3DTNN and 3DSTNN in the quantitative experiments (see the quantitative results in Tables 2–4).

4.4. 3DeepRep for HSI inpainting

We formulate HSI inpainting as a unified self-supervised optimization problem, in which network parameters and latent tensors are jointly

Table 2

Quantitative comparison on the **Pavia Centre** dataset across four degradation cases. Note that 3DeepRep achieves a significant improvement of over 7 dB in Case 3, demonstrating the robustness of the multi-directional strategy against structural missing.

Case	MR	Metric	Observed	TNN	3DTNN	t-CTV	S2NTNN	CoNoT	3DSTNN	HIR-Diff	3DeepRep
Case 1	0.99	PSNR	12.013	12.013	17.552	26.352	26.001	<u>29.523</u>	25.136	18.870	30.799 $\pm$ 0.369*
		SSIM	0.004	0.004	0.243	0.752	0.755	<u>0.885</u>	0.679	0.513	0.915 $\pm$ 0.001*
		SAM	2.207	2.207	0.178	0.128	0.090	<u>0.075</u>	0.107	0.452	0.056 $\pm$ 0.001*
	0.95	PSNR	12.192	28.879	37.296	33.786	37.932	<u>40.456</u>	35.499	26.821	43.214 $\pm$ 0.275*
		SSIM	0.018	0.860	0.981	0.950	0.981	<u>0.989</u>	0.970	0.867	0.994 $\pm$ 0.001*
		SAM	1.381	0.126	0.034	0.074	0.033	<u>0.026</u>	0.042	0.162	0.019 $\pm$ 0.001*
	0.90	PSNR	12.426	32.637	45.952	38.717	46.416	<u>48.980</u>	41.617	28.128	51.660 $\pm$ 0.215*
		SSIM	0.035	0.930	0.997	0.981	0.997	<u>0.998</u>	0.992	0.887	0.999 $\pm$ 0.001
		SAM	1.255	0.097	0.013	0.047	0.014	<u>0.012</u>	0.025	0.128	0.009 $\pm$ 0.001*
	0.85	PSNR	12.676	35.520	52.911	42.773	52.127	<u>54.990</u>	44.655	28.480	58.676 $\pm$ 0.281*
		SSIM	0.053	0.957	<u>0.999</u>	0.991	<u>0.999</u>	<u>0.999</u>	0.996	0.891	1.000 $\pm$ 0.001*
		SAM	1.177	0.078	0.007	0.033	0.009	<u>0.006</u>	0.020	0.119	0.003 $\pm$ 0.001*
Case 2	0.50	PSNR	14.940	14.940	26.319	<u>28.560</u>	23.178	<u>25.963</u>	21.798	28.428	30.237 $\pm$ 0.176*
		SSIM	0.216	0.216	0.772	<u>0.857</u>	0.588	0.766	0.586	0.835	0.905 $\pm$ 0.001*
		SAM	1.571	1.157	<u>0.042</u>	0.047	0.067	0.069	0.043	0.075	0.036 $\pm$ 0.001*
	0.30	PSNR	17.164	17.164	30.642	<u>32.187</u>	29.789	31.216	28.306	31.031	36.871 $\pm$ 0.331*
		SSIM	0.381	0.381	0.913	<u>0.939</u>	0.897	0.926	0.863	0.908	0.977 $\pm$ 0.000*
		SAM	0.942	0.607	0.022	0.025	0.032	<u>0.021</u>	0.024	0.044	0.019 $\pm$ 0.001
	0.10	PSNR	22.137	22.137	39.751	39.270	39.610	<u>43.017</u>	38.135	34.780	47.702 $\pm$ 0.289*
		SSIM	0.703	0.703	0.989	0.988	0.989	<u>0.995</u>	0.983	0.956	0.998 $\pm$ 0.001*
		SAM	0.298	0.235	<u>0.005</u>	0.007	0.012	0.004	0.006	0.038	0.011 $\pm$ 0.001*
Case 3		PSNR	18.415	27.614	39.722	35.799	39.754	<u>43.499</u>	39.291	29.682	51.164 $\pm$ 0.213*
		SSIM	0.648	0.889	0.988	0.959	0.990	<u>0.995</u>	0.988	0.901	0.999 $\pm$ 0.001*
		SAM	0.470	0.139	0.024	0.064	0.016	<u>0.012</u>	0.022	0.107	0.007 $\pm$ 0.001*
Case 4		PSNR	12.274	20.251	30.421	31.890	32.644	<u>33.279</u>	31.552	26.880	39.648 $\pm$ 0.109
		SSIM	0.024	0.583	0.934	0.930	0.951	<u>0.958</u>	0.939	0.851	0.987 $\pm$ 0.001*
		SAM	1.512	0.333	0.085	0.078	<u>0.042</u>	<u>0.042</u>	0.050	0.147	0.021 $\pm$ 0.000*

Note: The **best** and **second-best** values are highlighted in bold and underline, respectively. The reported mean $\pm$ standard deviation values for 3DeepRep are obtained from 10 repeated experiments. * indicates statistically significant improvement over the 1D method CoNoT ($p < 0.05$), as determined by paired t-tests based on 10 repeated experiments of 3DeepRep and CoNoT.

Table 3

Quantitative results on the **Pavia University** dataset across four degradation cases. 3DeepRep consistently outperforms baselines, demonstrating robustness to different spatial resolutions.

Case	MR	Metric	Observed	TNN	3DTNN	t-CTV	S2NTNN	CoNoT	3DSTNN	HIR-Diff	3DeepRep
Case 1	0.99	PSNR	14.010	14.010	18.738	27.384	27.220	<u>30.774</u>	25.900	18.938	30.991 $\pm$ 0.281
		SSIM	0.012	0.012	0.418	0.739	0.745	0.877	0.681	0.433	<u>0.868</u> $\pm$ 0.001*
		SAM	2.056	2.056	0.271	0.150	0.122	<u>0.093</u>	0.146	0.460	0.092 $\pm$ 0.001
	0.95	PSNR	14.189	28.565	36.651	33.778	36.697	<u>38.725</u>	35.290	22.721	40.413 $\pm$ 0.340*
		SSIM	0.035	0.766	0.963	0.920	0.956	<u>0.968</u>	0.948	0.643	0.971 $\pm$ 0.001*
		SAM	1.364	0.162	0.054	0.090	0.058	<u>0.052</u>	0.063	0.281	0.049 $\pm$ 0.001*
	0.90	PSNR	14.425	32.629	42.285	37.965	42.194	<u>43.167</u>	40.000	29.193	44.170 $\pm$ 0.237*
		SSIM	0.059	0.883	0.983	0.960	0.977	<u>0.981</u>	0.975	0.850	0.980 $\pm$ 0.000
		SAM	1.255	0.116	0.038	0.064	0.045	<u>0.041</u>	0.047	0.138	0.042 $\pm$ 0.001
	0.85	PSNR	14.674	35.376	<u>44.933</u>	40.892	44.654	44.887	41.755	29.869	45.392 $\pm$ 0.273*
		SSIM	0.081	0.925	0.987	0.974	0.982	<u>0.985</u>	0.981	0.863	0.983 $\pm$ 0.001*
		SAM	1.177	0.093	0.032	0.051	0.039	<u>0.036</u>	0.042	0.126	0.039 $\pm$ 0.001
Case 2	0.50	PSNR	16.969	16.969	27.869	<u>30.472</u>	24.803	27.907	23.438	30.452	33.082 $\pm$ 0.253*
		SSIM	0.252	0.252	0.819	<u>0.877</u>	0.681	0.819	0.629	0.860	0.925 $\pm$ 0.001*
		SAM	1.571	1.155	0.072	<u>0.062</u>	0.090	0.066	0.066	0.084	0.055 $\pm$ 0.001*
	0.30	PSNR	19.154	19.154	32.520	<u>34.074</u>	31.647	33.515	29.992	32.607	39.111 $\pm$ 0.203*
		SSIM	0.393	0.393	0.932	<u>0.945</u>	0.910	0.942	0.885	0.903	0.975 $\pm$ 0.001*
		SAM	0.933	0.608	0.032	<u>0.031</u>	0.044	0.027	0.032	0.075	0.036 $\pm$ 0.000*
	0.10	PSNR	19.154	24.339	41.314	40.480	40.525	<u>42.402</u>	39.980	34.962	44.975 $\pm$ 0.244*
		SSIM	0.393	0.707	0.990	0.987	0.985	<u>0.991</u>	0.985	0.928	0.993 $\pm$ 0.000*
		SAM	0.933	0.172	<u>0.008</u>	0.009	0.024	0.007	0.008	0.070	0.017 $\pm$ 0.001*
Case 3		PSNR	20.779	25.773	37.672	33.391	<u>39.818</u>	39.790	37.328	29.860	42.883 $\pm$ 0.302*
		SSIM	0.679	0.831	0.971	0.940	0.978	<u>0.979</u>	0.971	0.866	0.990 $\pm$ 0.001*
		SAM	0.411	0.192	0.039	0.066	0.031	<u>0.028</u>	0.037	0.130	0.022 $\pm$ 0.001*
Case 4		PSNR	14.277	21.620	33.355	30.665	32.999	<u>34.434</u>	31.809	26.536	38.060 $\pm$ 0.288*
		SSIM	0.043	0.542	0.921	0.875	0.914	<u>0.934</u>	0.908	0.810	0.959 $\pm$ 0.000*
		SAM	1.494	0.402	0.075	0.122	0.079	<u>0.074</u>	0.099	0.208	0.054 $\pm$ 0.001*

Note: The **best** and **second-best** values are highlighted in bold and underline, respectively. The reported mean $\pm$ standard deviation values for 3DeepRep are obtained from 10 repeated experiments. * indicates statistically significant improvement over the 1D method CoNoT ($p < 0.05$), as determined by paired t-tests based on 10 repeated experiments of 3DeepRep and CoNoT.

Table 4
Quantitative results on the Washington DC Mall dataset across four degradation cases.

Case	MR	Metric	Observed	TNN	3DTNN	t-CTV	S2NTNN	CoNoT	3DSTNN	HIR-Diff	3DeepRep
Case 1	0.99	PSNR	25.615	25.615	31.825	33.404	40.718	<u>43.150</u>	38.161	31.019	43.304 ± 0.188
		SSIM	0.266	0.266	0.604	0.776	0.908	<u>0.940</u>	0.874	0.660	0.949 ± 0.001*
		SAM	1.725	1.725	0.295	0.175	0.105	<u>0.090</u>	0.126	0.515	0.067 ± 0.001*
	0.95	PSNR	25.794	33.314	48.903	38.670	50.505	<u>53.151</u>	45.809	35.608	54.399 ± 0.215*
		SSIM	0.289	0.811	0.986	0.934	0.990	<u>0.994</u>	0.984	0.836	0.995 ± 0.000*
		SAM	1.355	0.155	0.046	0.090	0.032	<u>0.029</u>	0.044	0.277	0.025 ± 0.001*
	0.90	PSNR	25.615	36.826	56.240	42.567	54.997	<u>57.666</u>	48.526	36.681	59.522 ± 0.152*
		SSIM	0.241	0.902	<u>0.998</u>	0.972	<u>0.998</u>	<u>0.998</u>	0.993	0.858	0.999 ± 0.000*
		SAM	1.725	0.106	0.020	0.058	<u>0.018</u>	0.019	0.031	0.242	0.015 ± 0.000*
	0.85	PSNR	26.276	39.481	<u>60.024</u>	45.403	57.920	59.377	50.761	37.779	61.296 ± 0.251*
		SSIM	0.340	0.941	0.999	0.984	0.998	0.999	0.995	0.888	0.999 ± 0.000
		SAM	1.177	0.081	0.013	0.043	0.014	0.017	0.023	0.196	0.013 ± 0.000*
Case 2	0.50	PSNR	28.629	28.629	37.956	36.657	35.057	35.650	33.214	<u>38.755</u>	38.862 ± 0.283*
		SSIM	0.512	0.512	0.855	<u>0.883</u>	0.762	0.793	0.733	0.888	0.880 ± 0.000*
		SAM	1.571	1.148	0.089	0.068	0.138	0.089	0.105	0.091	<u>0.083</u> ± 0.001*
	0.30	PSNR	30.676	30.676	<u>41.415</u>	40.448	39.123	40.559	37.983	40.391	41.762 ± 0.253*
		SSIM	0.634	0.634	0.931	0.946	0.900	0.920	0.886	0.927	<u>0.941</u> ± 0.000*
		SAM	0.933	0.601	0.045	0.034	0.063	<u>0.040</u>	0.046	0.075	<u>0.040</u> ± 0.001
	0.10	PSNR	35.198	35.198	47.103	46.431	44.160	<u>47.379</u>	45.354	42.199	49.647 ± 0.201*
		SSIM	0.850	0.850	0.983	<u>0.987</u>	0.974	0.984	0.978	0.956	0.991 ± 0.001*
		SAM	0.307	0.215	0.012	0.010	0.026	<u>0.011</u>	0.013	0.064	0.017± 0.001*
Case 3		PSNR	31.983	29.252	<u>47.561</u>	36.035	46.106	46.120	42.462	38.174	47.908 ± 0.141*
		SSIM	0.748	0.773	0.962	0.913	<u>0.967</u>	0.959	0.946	0.879	0.980 ± 0.000*
		SAM	0.514	0.407	<u>0.047</u>	0.103	0.054	0.057	0.085	0.144	0.033 ± 0.001*
Case 4		PSNR	25.878	27.151	<u>42.863</u>	34.989	41.147	42.730	38.713	36.309	46.369 ± 0.170*
		SSIM	0.297	0.621	0.944	0.884	0.937	<u>0.949</u>	0.916	0.857	0.977 ± 0.001*
		SAM	1.497	0.504	<u>0.070</u>	0.121	0.092	0.078	0.115	0.199	0.036 ± 0.001*

Note: The **best** and second-best values are highlighted in bold and underline, respectively. The reported mean ± standard deviation values for 3DeepRep are obtained from 10 repeated experiments. * indicates statistically significant improvement over the 1D method CoNoT ($p < 0.05$), as determined by paired t-tests based on 10 repeated experiments of 3DeepRep and CoNoT.

optimized to minimize a composite objective comprising local directional constraints and a global aggregation loss. For each directional component $i \in \{1, 2, 3\}$, we define a specific sub-problem that simultaneously enforces the low-rank structure on the latent tensor $\hat{\mathcal{X}}_i$ and ensures consistency with the observed entries in $\mathcal{O}$. The loss function for the i -th direction is formulated as:

$$\mathcal{L}_i = \sum_{k=1}^{\tilde{n}_i} \left\| (\hat{\mathcal{X}}_i^{(k)})_i \right\|_* + \lambda_i \left\| \mathcal{P}_\Omega \left(g_i(\hat{\mathcal{X}}_i) - \mathcal{O} \right) \right\|_F^2, \quad (9)$$

where $\mathcal{P}_\Omega(\cdot)$ denotes the projection operator that keeps entries on the observed set Ω unchanged and sets others to zero. The hyperparameter λ_i governs the trade-off between the low-rank regularization strength and the data fidelity.

The overall optimization objective of 3DeepRep combines these directional sub-losses with a global constraint on the final fused output. The complete end-to-end optimization problem is expressed as:

$$\min_{\{\Theta_i, \hat{\mathcal{X}}_i\}_{i=1}^3, \Theta_G} \sum_{i=1}^3 \alpha_i \mathcal{L}_i + \left\| \mathcal{P}_\Omega \left(G(g_1(\hat{\mathcal{X}}_1), g_2(\hat{\mathcal{X}}_2), g_3(\hat{\mathcal{X}}_3)) - \mathcal{O} \right) \right\|_F^2, \quad (10)$$

where $\{\alpha_i\}$ are weighting coefficients that balance the contribution of each directional representation, and $G(\cdot)$ is the learnable aggregation module (as defined in Eq. (8)).

4.5. Fully gradient-driven optimization

The proposed model introduces complex nonlinear dependencies among variables, rendering traditional splitting methods like ADMM inefficient. Instead, we adopt a *fully gradient-driven* strategy. Let $\hat{\mathcal{X}}_i$ denote the latent tensor which serves as the low-rank representation. The CTB module g_i maps this latent representation to the image space, yielding

the directional reconstruction $\mathcal{X}_i = g_i(\hat{\mathcal{X}}_i)$. Finally, the aggregation module G fuses these intermediate results to produce the final output $\mathcal{X}$. We compute subgradients for the entire objective $\mathcal{L} = \sum \alpha_i \mathcal{L}_i + \mathcal{L}_{agg}$ via backpropagation.

4.5.1. Gradients w.r.t. network parameters

The gradients for the transformation weights Θ_i and aggregation weights Θ_G are derived via the chain rule:

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial \Theta_i} &= 2\alpha_i \lambda_i \sum_{l,j,s} [P_\Omega(\mathcal{X}_i - \mathcal{O})]_{l,j,s} \frac{\partial (\mathcal{X}_i)_{l,j,s}}{\partial \Theta_i} \\ &\quad + 2 \sum_{l,j,s} [P_\Omega(\mathcal{X} - \mathcal{O})]_{l,j,s} \frac{\partial \mathcal{X}_{l,j,s}}{\partial \Theta_i}. \end{aligned} \quad (11)$$

$$\frac{\partial \mathcal{L}}{\partial \Theta_G} = 2 \sum_{l,j,s} [P_\Omega(\mathcal{X} - \mathcal{O})]_{l,j,s} \frac{\partial \mathcal{X}_{l,j,s}}{\partial \Theta_G}. \quad (12)$$

4.5.2. Gradients w.r.t. latent tensors

Computing gradients for $\hat{\mathcal{X}}_i$ involves the subgradient of the nuclear norm [41]. Let $\mathbf{M} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^\top$ be the SVD of a matrix $\mathbf{M}$; its subgradient is given by $\partial \|\mathbf{M}\|_* = \mathbf{U}_r \mathbf{V}_r^\top$, where subscript r denotes truncation to the rank r .

The total gradient of $\mathcal{L}$ with respect to the element $(\hat{\mathcal{X}}_i)_{l,j,s}$ is composed of three terms: the nuclear norm regularization gradient (d_i^1), the directional reconstruction gradient (d_i^2), and the aggregation gradient (d_i^3):

$$\frac{\partial \mathcal{L}}{\partial (\hat{\mathcal{X}}_i)_{l,j,s}} = \alpha_i d_i^1 + 2\alpha_i \lambda_i d_i^2 + 2d_i^3. \quad (13)$$

Regularization Term (d_i^1): This term arises from $\sum \|(\hat{\mathcal{X}}_i)^{(k)}\|_*$. Depending on the mode i , the singular vectors correspond to different unfolding directions:

$$d_i^1 = \begin{cases} (\tilde{\mathbf{U}}_l^1 \tilde{\mathbf{V}}_l^{1\top})_{j,s}, & \text{if } i = 1, \\ (\tilde{\mathbf{U}}_j^2 \tilde{\mathbf{V}}_j^{2\top})_{l,s}, & \text{if } i = 2, \\ (\tilde{\mathbf{U}}_s^3 \tilde{\mathbf{V}}_s^{3\top})_{l,j}, & \text{if } i = 3. \end{cases} \quad (14)$$

where $\tilde{\mathbf{U}}$ and $\tilde{\mathbf{V}}$ are the truncated singular vectors of the corresponding frontal slices.

Fidelity Terms (d_i^2 and d_i^3): These terms propagate the reconstruction errors back to the latent space:

$$d_i^2 = \sum_{l^*, j^*, s^*} [\mathcal{P}_\Omega(\mathcal{X}_i - \mathcal{O})]_{l^*, j^*, s^*} \frac{\partial(\mathcal{X}_i)_{l^*, j^*, s^*}}{\partial(\hat{\mathcal{X}}_i)_{l,j,s}}, \quad (15)$$

$$d_i^3 = \sum_{l^*, j^*, s^*} [\mathcal{P}_\Omega(\mathcal{X} - \mathcal{O})]_{l^*, j^*, s^*} \frac{\partial \mathcal{X}_{l^*, j^*, s^*}}{\partial(\hat{\mathcal{X}}_i)_{l,j,s}}. \quad (16)$$

This derivation confirms that the proposed framework allows for the efficient computation of exact gradients for all components using standard automatic differentiation frameworks, enabling robust end-to-end training with optimizers such as Adam [42].

5. Experiments

To comprehensively evaluate the efficacy and robustness of the proposed 3DeepRep, we conduct extensive experiments on both simulated and real-world degraded Hyperspectral Images (HSIs). All pixel intensities of the HSIs are normalized to the range [0, 1] prior to processing. We benchmark our method against representative state-of-the-art approaches from two distinct perspectives: *model-guided* and *learning-guided* methods. Regarding the computational environment, all deep learning-based models are trained on an NVIDIA A100 GPU. Traditional optimization-based experiments are implemented in MATLAB (R2023b) on a workstation equipped with an Intel Core i9-10900K 3.70GHz processor and 128-GB RAM.

5.1. Baseline methods

We select seven representative methods for comparison, categorized by their modeling structure and underlying optimization path.

5.1.1. Model-guided methods

Methods in this category rely on explicit optimization formulations regularized by pre-defined transforms. While theoretically grounded, their reconstruction quality is often constrained by the fixed transformations and sensitivity to initialization.

- **TNN** [43]: The classical Tensor Nuclear Norm model based on the Fourier transform. It serves as a representative baseline for 1D transform-based low-rank regularization frameworks.
- **3DTNN (WSTNN)**² [31]: A 3D extension of TNN that utilizes predefined multi-directional transforms to capture spatial-spectral correlations simultaneously.
- **t-CTV**³ [40]: A unified optimization model integrating low-rank properties with 3D Total Variation transform. Its performance heavily depends on handcrafted priors and the careful tuning of hyperparameter.

5.1.2. Learning-guided methods

This category employs deep neural networks to parameterize the transformation, optimizing the tensor recovery via data-driven strategies. These approaches aim to address the limitations of predefined transformations found in model-guided methods:

- **S2NTNN**⁴ [29]: A self-supervised framework that replaces the fixed Fourier transform in TNN with a fully connected learnable transform, enabling deep representation learning.
- **CoNoT** [30]: A coupled nonlinear 1D transform network designed to jointly learn spatial-spectral representations, enhancing the low-rank tensor representation.
- **3DSTNN** [32]: A 3D extension of S2NTNN that employs learnable transforms across all modes. However, it retains an explicit solver-based estimation step (similar to 3DTNN), rendering the solution sensitive to initialization.
- **HIR-Diff**⁵ [44]: A state-of-the-art generative framework combining diffusion models with low-rank tensor decomposition for high-fidelity HSI recovery.

5.1.3. Our proposed model

3DeepRep is a fully learning-guided 3D tensor representation designed to bridge the gap between low-rank modeling and deep representation learning. Unlike the baselines, it achieves two key innovations:

- Fully learnable representation:** Both the network parameters and the latent tensor representations are jointly optimized, moving beyond fixed generators or predefined transformations. It adopts a unified, end-to-end optimization strategy without relying on explicit splitting solvers.
- Robustness and stability:** The model exhibits significant robustness to initialization, ensuring stable convergence and consistent high-performance recovery across diverse and complex degradation patterns.

5.1.4. Implementation details

Model Hyperparameter Settings. The proposed 3deepRep model involves three groups of hyperparameters: latent-space dimensions $\{\tilde{n}_i\}_{i=1}^3$, directional weights $\{\alpha_i\}_{i=1}^3$, and regularization coefficients $\{\lambda_i\}_{i=1}^3$. First, we set $\tilde{n}_i = n_i$ ($i = 1, 2, 3$) to align the latent-space dimensionality with the input tensor, preserving full spatial-spectral resolution. Second, to reduce the search space, we parameterize the directional weights as $(\alpha_1, \alpha_2, \alpha_3) = \beta \frac{(1, 1, \theta)}{2 + \theta}$. Here, β controls the global strength of low-rank regularization, while θ adjusts the relative emphasis between spatial modes (modes 1 & 2) and the spectral mode (mode 3). Third, for parameter efficiency, we couple the weights by constraining $\alpha_i \lambda_i = \gamma$ for all i , implying $\lambda_i = \gamma / \alpha_i$. This simplifies the tuning process to a single scalar γ , which governs the data fidelity of the intermediate reconstructions $\mathcal{X}_i$. Consequently, the optimization objective in Eq. (10) is reformulated as:

$$\mathcal{L} = \beta \sum_{i=1}^3 \tilde{\alpha}_i \sum_{k=1}^{\tilde{n}_i} \left\| (\hat{\mathcal{X}}_i)^{(k)} \right\|_* + \gamma \sum_{i=1}^3 \left\| \mathcal{P}_\Omega(\mathcal{X}_i - \mathcal{O}) \right\|_F^2 + \left\| \mathcal{P}_\Omega(\mathcal{X} - \mathcal{O}) \right\|_F^2, \quad (17)$$

where $(\tilde{\alpha}_1, \tilde{\alpha}_2, \tilde{\alpha}_3) = \frac{(1, 1, \theta)}{2 + \theta}$. For fair comparison, all hyperparameters of the baseline methods are tuned according to the authors' official recommendations.

Initialization. The learnable network parameters $\{\Theta_i\}_{i=1}^3$ and Θ_G are initialized using PyTorch's default normal distribution.⁶ For the explicit

² <https://yubangzheng.github.io/>.

³ <https://github.com/wanghailin97>.

⁴ <https://github.com/YisiLuo/S2NTNN>.

⁵ <https://github.com/LiPang/HIRDiff>.

⁶ <https://pytorch.org/docs/stable/nn.init.html>.

latent tensors $\{\hat{\mathcal{X}}_i\}_{i=1}^3$, we employ a ‘warm start’ strategy: they are initialized using the result of the conventional TNN method [43]. This provides a reasonable initial estimate of the image structure, facilitating faster convergence and stability compared to random initialization.

Optimization Settings. All models were optimized using the Adam optimizer with a learning rate of 0.01, while all other Adam parameters were set to their default values in PyTorch. The number of training iterations was determined based on a convergence criterion, where training was terminated once the loss remained stable without further noticeable decrease over the latter half of the optimization process. Random seeds were not explicitly fixed, unless otherwise stated. Our code is available at: <https://github.com/Kendradra/3DeepRep>.

5.2. HSI inpainting on simulated data

5.2.1. Experimental settings

Datasets. We evaluate our proposal on three benchmark hyperspectral datasets widely used in simulation experiments: (1) **Pavia Centre**⁷: A scene acquired over Pavia, Italy, with a size of $200 \times 200 \times 80$. (2) **Pavia University**⁸: A dataset capturing the university area, with dimensions $256 \times 256 \times 103$. (3) **Washington DC Mall**⁹: An urban scene dataset with dimensions $256 \times 256 \times 191$.

Quantitative Indices. To assess the reconstruction quality, we employ three standard metrics: Peak Signal-to-Noise Ratio (PSNR), Structural Similarity (SSIM), and Spectral Angle Mapper (SAM). PSNR and SSIM measure spatial fidelity and structural consistency, respectively, while SAM evaluates the spectral fidelity by calculating the angle between reconstructed and ground-truth spectra. Higher PSNR/SSIM values and lower SAM values indicate superior inpainting performance.

Degradation Scenarios. To test the robustness of the proposed method, we compare it against baselines under four distinct degradation scenarios:

- **Case 1 (Random Pixel Missing):** We simulate random data loss by subsampling entries from the HSI tensor. The observed ratios (sampling rates) are set to 0.15, 0.10, 0.05, and 0.01 (corresponding to missing rates of 0.85, 0.90, 0.95, and 0.99, respectively).
- **Case 2 (Stripe Slice Missing):** This case simulates structural loss by removing complete vertical stripe slices from the tensor. We test missing rates of 0.10, 0.30, and 0.50 to evaluate the recovery of spatial continuity.
- **Case 3 (Consecutive Deadline Missing):** We simulate deadline noise (dead columns) across four groups of consecutive bands, with each group covering 1/4 of the total spectral bands. The missing rates (defined as the ratio of total deadline width to image width) for the four patterns are set to 0.1, 0.2, 0.3, and 0.4, respectively. For each pattern, the number of deadlines is randomly selected from $\{6, 7, \dots, 10\}$. A visual illustration of these patterns is provided in Fig. 3.
- **Case 4 (Mixture Missing):** We design a challenging hybrid scenario to evaluate the model’s capability in handling complex, coupled degradations. By integrating random missing (Case 1, MR=0.90), stripe missing (Case 2, MR=0.10), and deadline patterns (Case 3), this case creates a highly difficult HSI task. This scenario provides a critical benchmark to highlight 3DeepRep’s superiority performance in disentangling mixed noise patterns compared to selected baselines.

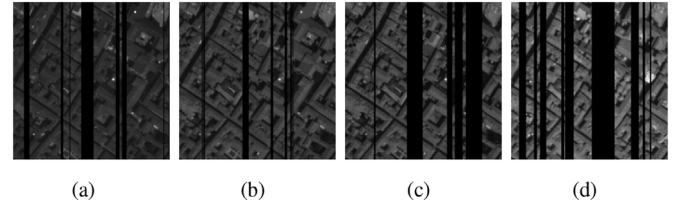


Fig. 3. Visualization of the four deadline patterns applied to the Pavia Centre dataset. (a) Pattern for bands 1–20 (width ratio: 0.2). (b) Pattern for bands 21–40 (width ratio: 0.1). (c) Pattern for bands 41–60 (width ratio: 0.3). (d) Pattern for bands 61–80 (width ratio: 0.4).

5.2.2. Pavia centre dataset results

Table 2 reports the quantitative comparison on the Pavia Centre dataset. The proposed 3DeepRep achieves state-of-the-art performance, outperforming all seven baselines across almost all degradation scenarios. We highlight two critical observations that directly validate our core contributions:

Validation of Multi-directional Deep Transform: In the deadline missing scenario (Case 3), 3DeepRep surpasses the second-best method (CoNoT) by a remarkable margin of 7.553 dB. This result provides strong empirical evidence for the effectiveness of the *Multi-directional Deep Transform*. In this scenario, the spectral continuity is strictly severed by dead columns, rendering spectral (Mode-3) smoothness priors largely ineffective. However, 3DeepRep treats the latent tensor as a unified 3D volume. By permuting the tensor, our model effectively shifts the optimization focus to minimize the nuclear norms of Mode-1 and Mode-2 unfoldings. Consequently, the model reconstructs the missing spectral bands by propagating valid spatial geometric structures from uncorrupted bands. This confirms that the multi-directional design successfully exploits complementary structural priors: when one dimension is occluded, the low-rank constraints from orthogonal dimensions guide the recovery.

Validation of Learnable Aggregation Module: The advantage of 3DeepRep is further verified in the most extreme mixture missing scenario (Case 4). As visualized in Fig. 4, 3DeepRep restores the most accurate color fidelity and distinct spatial details (e.g., the preservation of red dots within the zoomed-in yellow square). The error maps show the lowest residual noise (closest to pure dark blue). This success is attributed to the *Learnable Aggregation Module*, which functions as a semantic gating mechanism. Unlike heuristic 3D methods (3DTNN, 3DSTNN) that use fixed weights, our module dynamically suppresses directional reconstructions exhibiting high error residuals (e.g., rejecting Mode-2 artifacts in stripe-corrupted regions) while amplifying the contribution of valid modes. This mechanism explains why 3DeepRep maintains structural fidelity where baseline methods suffer from significant spectral distortion and residual stripe noise.

5.2.3. Pavia university dataset results

Table 3 shows the quantitative superiority of 3DeepRep, which maintains the highest PSNR and SSIM scores across all degradation scenarios. The consistent performance observed across datasets with varying spatial resolutions (Pavia Centre vs. Pavia University) underscores the robustness of our **Generator-Free** optimization strategy. Unlike generator-based methods that may overfit to specific spatial structural priors or struggle with initialization, our explicit latent variable approach adapts flexibly to different spatial resolutions.

Fig. 5 presents the visual comparison. 3DeepRep produces a restored image closely aligned with the ground truth and sharp structural boundaries. Crucially, it effectively eliminates both stripe and deadline noise, whereas comparative methods (e.g., tCTV and 3DSTNN) often leave residual artifacts in the texture-rich regions. This visual fidelity verifies the model’s ability to balance local smoothness with global structure preservation.

⁷ https://www.ehu.es/ccwintco/index.php/Hyperspectral_Remote_Sensing_Scenes.

⁸ https://www.ehu.es/ccwintco/index.php/Hyperspectral_Remote_Sensing_Scenes.

⁹ <http://lesun.weebly.com/hyperspectral-data-set.html>.

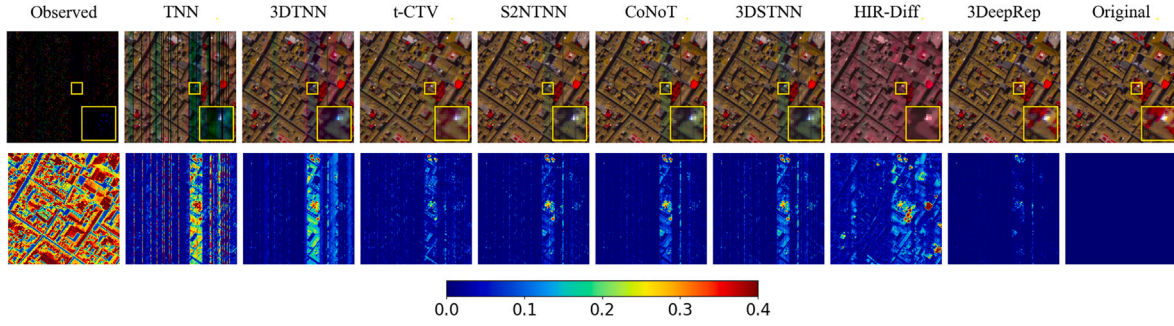


Fig. 4. Visual comparison on the **Pavia Centre** dataset under Case 4 (Mixture Missing). **Top rows:** False-color images (R:70, G:40, B:10) and zoomed-in details. **Bottom rows:** Corresponding absolute error maps of Band 58. Note that 3DeepRep effectively suppresses the stripe noise and spectral distortions that are prevalent in other methods (e.g., 3DTNN, t-CTV and 3DSTNN), validating the effectiveness of the Learnable Aggregation Module in decoupling mixed degradations.

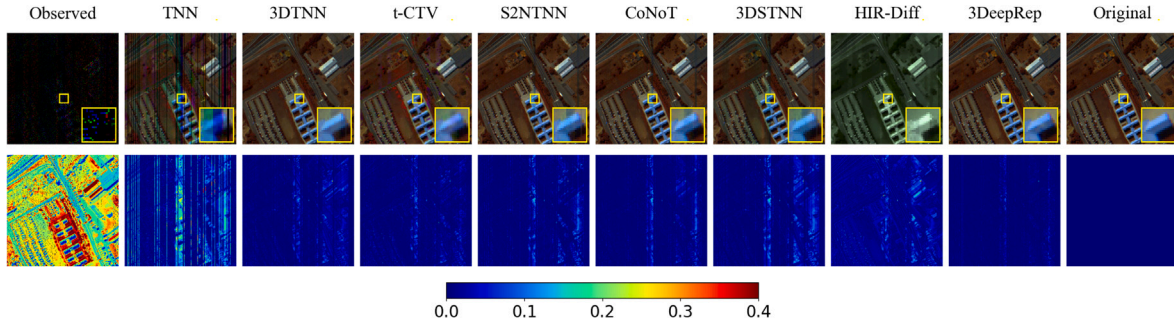


Fig. 5. Visual comparison on the **Pavia University** dataset under Case 4 (Mixture Missing). **Top rows:** False-color images and zoomed-in patches. **Bottom rows:** Absolute error maps of Band 75. 3DeepRep effectively removes complex mixed noise patterns, preserving clear boundaries.

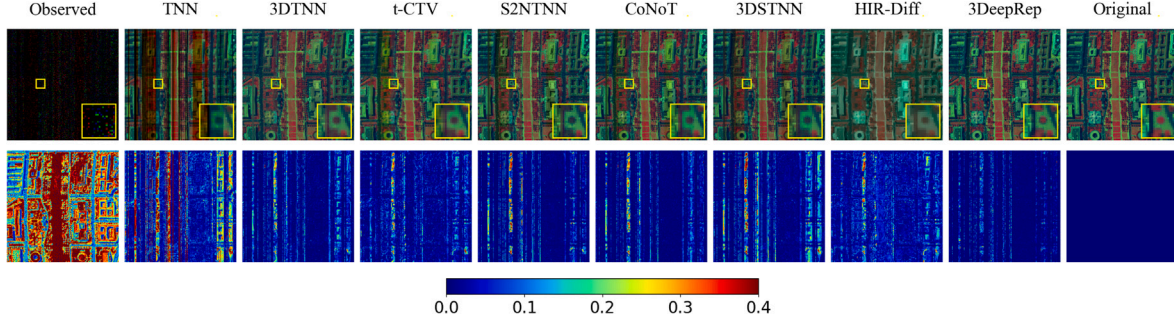


Fig. 6. Visual comparison on the **Washington DC Mall** dataset under Case 4 (Mixture Missing). **Top rows:** False-color images and zoomed-in details. **Bottom rows:** Absolute error maps of Band 60. 3DeepRep shows superior preservation of spatial textures and minimal residual error compared to other methods.

5.2.4. Washington DC mall dataset results

Fig. 6 visualizes the restoration results on the Washington DC Mall dataset under the severe Case 4 scenario. Among all baselines, 3DeepRep reconstructs the sharpest spatial details, with an error map that is dark blue, indicating the highest fidelity to the ground truth. It is worth noting that: 1D-based methods (e.g., CoNoT) exhibit improved relative performance here compared to their results on the Pavia datasets. We attribute this to the DC Mall's significantly higher spectral resolution (191 bands), which offers abundant redundancy along the spectral dimension that 1D methods can exploit. Even under these conditions, 3DeepRep still achieves the highest PSNR across almost all test cases (as detailed in Table 4). This validates the scalability of our framework, proving that 3DeepRep can effectively leverage deep spatial-spectral priors to outperform spectral-only methods, maintaining consistent superiority even when handling HSIs with substantially more bands or different spectral sampling rates.

5.3. Real-world HSI inpainting

To evaluate the generalization capability of 3DeepRep in real-world scenarios, we conduct experiments on two real-world hyperspectral datasets: MODIS Level-1B¹⁰ and Landsat 7 ETM+. ¹¹ For **MODIS**, the cropped MODIS data volume has a size of $256 \times 256 \times 10$ and consists of emissive bands. Unlike the simulated experiments, this dataset exhibits *intrinsic sensor degradation*: Band 27 suffers from severe stripe noise caused by structured pixel missing, while the remaining nine bands are fully observed. The degradation pattern is visualized in the leftmost column of Fig. 7. For **Landsat 7 ETM+**, the cropped Landsat 7 data volume has a size of $256 \times 256 \times 7$. This dataset also suffers from *intrinsic sensor*

¹⁰ <https://ladsweb.modaps.eosdis.nasa.gov/missions-and-measurements/products/MOD021KM>.

¹¹ <https://earthexplorer.usgs.gov/>.

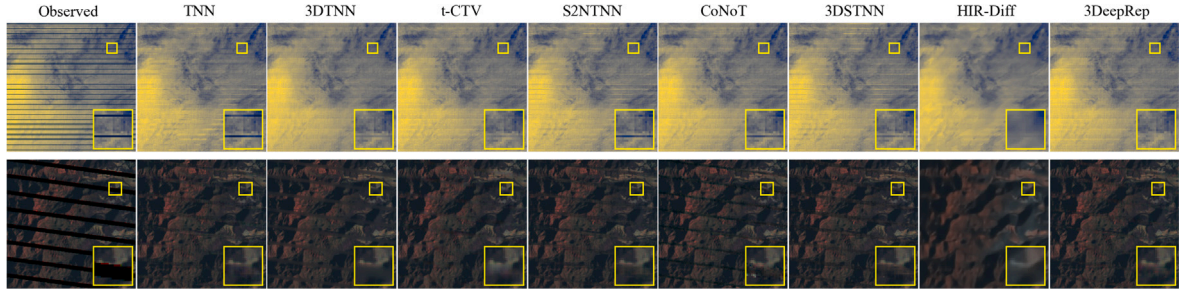


Fig. 7. Real-world hyperspectral inpainting results. **Row 1:** Band 27 of a MODIS Level-1B dataset (Antarctic Peninsula, Feb. 26, 2011). **Row 2:** Landsat 7 ETM+ dataset (western Venezuela, Jan. 18, 2024) shown in false color. Both images suffer from intrinsic sensor stripe noise. **3DeepRep** (last column) effectively removes stripe artifacts without over-smoothing, outperforming both 1D-based and conventional 3D-based methods.

degradation, where the same oblique stripe-like missing patterns appear simultaneously across all spectral bands, as shown in Fig. 7.

We compare 3DeepRep with seven baselines on the two datasets. The inpainting results are presented in Fig. 7 (We only display Band 27 for MODIS, as it is the sole corrupted band). It is observed that the baselines exhibit varying degrees of failure.

- **1D-based methods (TNN, S2NTNN, CoNoT):** These methods fail to remove the degradation, retaining obvious stripes. This indicates that relying solely on spectral constraints is insufficient to restore large-scale spatial structural defects.
- **Conventional 3D methods (3DTNN, t-CTV, 3DSTNN):** While these methods utilize spatial priors, they still leave slight residual artifacts, (for MODIS, particularly visible in the lower-middle part of the zoomed box). This suggests that their handcrafted weighting schemes are not flexible enough to separate complex stripe noise.
- **Generative method (HIR-Diff):** Although it removes the stripes, the resulting image suffers from excessive smoothing, leading to a loss of fine details.
- **3DeepRep (Ours):** Our method achieves the superior visual quality. It effectively suppresses the sensor-induced stripe artifacts while preserving the sharp edges and natural texture. This success is attributed to the *Multi-directional Deep Transform*, which allows the model to reconstruct the corrupted stripes through 3-directional correlations. Furthermore, the *Learnable Aggregation Module* automatically suppresses the artifacts generated by the corrupted direction, ensuring a clean and high-fidelity output.

5.4. Ablation study

In this section, we conduct a comprehensive ablation study to validate the effectiveness of the proposed 3DeepRep model. We evaluate the contributions of the core components, including the generator-free latent representation strategy, the multi-directional deep transform, and the learnable aggregation module. All experiments reported in this section are performed on the Pavia Centre dataset using the mixed missing mechanism (Case 4). The quantitative comparisons are presented in Tables 5 and 6.

5.4.1. Learnable latent tensors vs. generated latent tensors

A core contribution of 3DeepRep is the **generator-free optimization strategy**. Unlike previous deep tensor methods (e.g., DTR [38]) that rely on generator networks to map random noise into latent tensors, we treat the multi-directional latent tensors $\{\hat{\mathcal{X}}_i\}_{i=1}^3$ as explicit learnable variables. To validate the claims of parameter efficiency and optimization stability presented in our contributions, we compare our approach with a *Generator-based Variant*. In this variant, the explicit latent tensor $\hat{\mathcal{X}}_i$ is replaced by the output of a neural network $f_i(\mathcal{X}^0)$, where $\mathcal{X}^0$ denotes the input random noise and $f_i(\cdot)$ shares the same architecture as the decoder $g_i(\cdot)$. Consequently, the optimization problem for this

Table 5

Quantitative ablation study on the Pavia Centre Dataset (Case 4).

Module	Setting	PSNR	SSIM	SAM	Network Params
Latent Strategy	Generated	36.823	0.980	0.025	202.8 k
	Learnable (Ours)	39.514	0.987	0.021	112.1 k
Aggregation Direction	Mode-1	33.912	0.952	0.054	216.1 k
	Mode-2	22.623	0.684	0.241	216.1 k
	Mode-3 (Ours)	39.514	0.987	0.021	112.1 k
Initialization	Random $\mathcal{U}(0, 1)$	39.078	0.986	0.022	112.1 k
	Observed (Zero-filled)	39.302	0.987	0.022	112.1 k
	TNN [43] Result	39.514	0.987	0.021	112.1 k

Note: The best results are marked in **bold**. **Network Params** = network parameters.

variant is formulated as:

$$\min_{\Theta^{f,g}, \Theta_G} \sum_{i=1}^3 \alpha_i \mathcal{L}_i(f_i(\mathcal{X}^0)) + \left\| P_{\Omega} \left(G(\{g_i(f_i(\mathcal{X}^0))\}_{i=1}^3) - \mathcal{O} \right) \right\|_F^2,$$

where $\Theta^{f,g}$ denotes the parameters for both the generator f_i and the decoder g_i . The comparison results in the first block of Table 5 reveal two key insights: (1) the generator-free design significantly reduces the number of learnable network parameters (from 202.8 k to 112.1 k). By eliminating the redundant generator parameters, we minimize model complexity and mitigate the risk of overfitting. (2) 3DeepRep achieves superior PSNR (39.514 dB vs. 36.823 dB). This confirms that directly optimizing the latent variables avoids the sensitivity to inductive biases inherent in deep generators.

5.4.2. Effectiveness of multi-directional transforms

To verify the contribution of the **multi-directional deep representation**, we evaluate the impact of extending the deep transform CTB to all tensor modes. We compare the full 3DeepRep model against 1D configurations where CTB is applied only along a single direction ($i = 1, 2$, or 3). The results in the first block of Table 6 demonstrate that the multi-directional strategy (3D) consistently outperforms all 1D-based methods. This verifies that 3DeepRep effectively exploits the comprehensive spatial-spectral correlations across different dimensions, rather than relying on a single mode. Furthermore, among the single-mode variants, the spectral transform (Mode-3) yields the highest accuracy, confirming its critical role in capturing global spectral dependencies.

5.4.3. Analysis of aggregation strategy

Our proposed **learnable aggregation module** is designed to replace heuristic weighting schemes with a data-driven fusion strategy. While the module is learnable, the direction of aggregation is critical for preserving spatial consistency. We compare aggregating along Mode-1,

Table 6

Quantitative ablation study with computational cost comparison on the Pavia Centre Dataset (Case 4).

Module	Setting	PSNR	SSIM	SAM	Network Params	GPU Mem	Iters	Time (s)
Transform Direction	Mode-1 Only	30.838	0.925	0.077	41.8 k	0.30 GB	10 k	278
	Mode-2 Only	21.907	0.647	0.266	41.8 k	0.30 GB	10 k	273
	Mode-3 Only	33.939	0.961	0.036	7.1 k	0.30 GB	10 k	493
	3D (Ours)	39.514	0.987	0.021	112.1 k	1.01 GB	20 k	1065
Latent Dimension (k)	k = 0.5	39.079	0.987	0.022	66.7 k	0.85 GB	10 k	533
	k = 1.0 (Default)	39.514	0.987	0.021	112.1 k	1.01 GB	20 k	1065
	k = 2.0	38.913	0.986	0.022	202.8 k	1.31 GB	40 k	2140
	k = 3.0	38.601	0.986	0.022	293.5 k	1.62 GB	50 k	2975
	k = 4.0	38.31	0.984	0.022	384.2 k	1.96 GB	60 k	3600
Compactness	CoNoT [30]	33.279	0.958	0.042	54.8 k	0.66 GB	10 k	633
	3DeepRep (k = 0.3)	38.657	0.985	0.022	48.6 k	0.79 GB	7 k	357

Note: The best results are marked in **bold**. **Network Params** = network parameters; **GPU Mem** = GPU memory; **Iters** = number of iterations, determined when the loss shows no clear decreasing trend over the latter half of training.

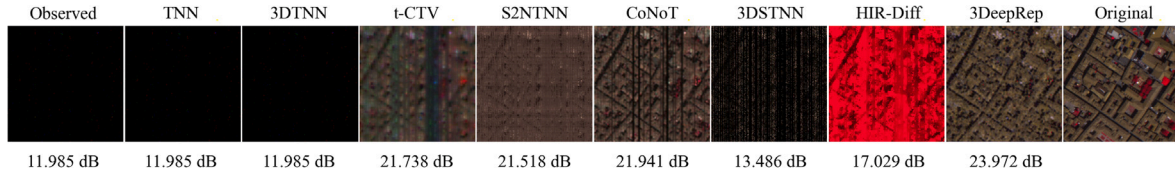


Fig. 8. Robustness analysis under a compound extreme degradation scenario. The input is corrupted by three simultaneous factors: 99% random missing pixels (Case 1), 50% stripe noise (Case 2), and deadline patterns (Case 3). Although such severe information loss causes significant performance degradation across all methods, 3DeepRep (Ours) effectively preserves structural integrity and achieves the highest PSNR, demonstrating superior resilience compared to state-of-the-art baselines.

Mode-2, and Mode-3. As defined in Remark 3, Mode-3 aggregation concatenates tensors along the spectral dimension, enriching spectral depth while allowing the subsequent transform to refine local spatial details. The results in the second block of Table 5 confirm that Mode-3 aggregation significantly outperforms other directions (e.g., +5.6 dB over Mode-1). This validates that our aggregation module not only adaptively fuses information but is more effective when maintaining pixel-wise spatial alignment.

5.4.4. Robustness to initialization

A key advantage of our Generator-Free framework is its robustness. Since we treat latent tensors as explicit variables optimized via backpropagation, the process should be less sensitive to initial states compared to generative models that require careful tuning of noise inputs. We test three initialization strategies for $\{\hat{\mathcal{X}}_i\}_{i=1}^3$: (a) Random uniform distribution $U(0, 1)$; (b) Zero-filled observed image; and (c) TNN-based initialization [43]. The third block of Table 5 shows that 3DeepRep yields comparable high-performance results across all strategies (PSNR fluctuation < 0.5 dB). This empirical evidence supports our claim of optimization stability, demonstrating that the gradient-driven updates can effectively converge regardless of the starting point.

5.5. Discussion

5.5.1. Hyperparameter analysis

We conduct a sensitivity analysis of the key hyperparameters in 3DeepRep using the Pavia Centre dataset under the mixed missing mechanism (Case 4). The parameters analyzed include the latent space dimension ratio k , the spatial-spectral balancing weight θ , the low-rank regularization weight β , and the intermediate fidelity weight γ .

Latent Space Dimensions (k): We define the dimensions of the latent tensors as $\tilde{n}_i = kn_i$ and evaluate $k \in \{0.5, 1, 2, 3, 4\}$. As detailed in Table 6, the optimal performance is achieved at $k = 1$. Setting $k = 0.5$ results in information loss, leading to inferior reconstruction. Conversely, increasing k beyond 1 does not yield performance gains. This behavior suggests

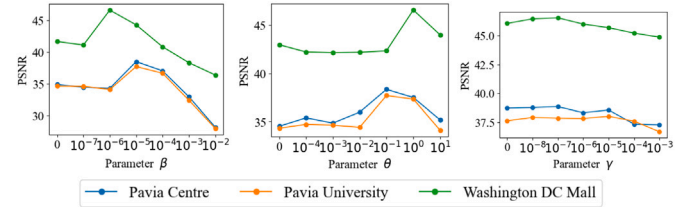


Fig. 9. Sensitivity analysis of hyperparameters (β, θ, γ). All experiments were conducted under Case 4. The curves demonstrate that the model is sensitive to β and θ , while remaining relatively robust to variations in γ .

that $k = 1$ shows the optimal balance; higher dimensions increase parameter complexity and may relax the low-rank constraints necessary for effective regularization, while $k = 1$ sufficiently captures the intrinsic data structure without overfitting.

Regularization Weights (θ, β, γ): The sensitivity of the model to these weights is illustrated in Fig. 9.

- **Spatial-spectral Balance (θ):** This parameter balances the low-rank constraints between spatial modes ($i = 1, 2$) and the spectral mode ($i = 3$). The model exhibits sensitivity to θ , with optimal values of 10^{-1} for Pavia Centre/University and 10^0 for Washington DC Mall.
- **Low-rank Regularization (β):** β controls the strength of the nuclear norm minimization. The distinct peaks in the PSNR curves confirm the critical role of low-rank regularization, with optimal performance observed at $\beta = 10^{-5}$ for the Pavia datasets and 10^{-6} for Washington DC Mall.
- **Intermediate Fidelity (γ):** This parameter governs the consistency of the directional reconstructions $\{\hat{\mathcal{X}}_i\}_{i=1}^3$ with the observed data. As shown in Fig. 9, the model is relatively robust to variations in γ . This indicates that while intermediate fidelity is useful, the final aggregation module compensates for minor deviations in the directional branches. The optimal γ values are 10^{-7} , 10^{-8} , and 10^{-7} for the three respective datasets.

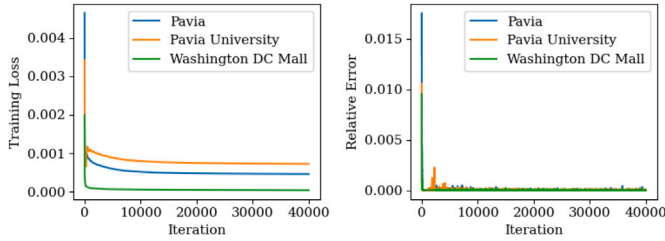


Fig. 10. Convergence analysis of the proposed algorithm. Experiments on the three datasets were all conducted under Case 4. The relative error is computed as $\text{RelErr} = \|\mathcal{X}^t - \mathcal{X}^{t-1}\|_F^2 / \|\mathcal{X}^{t-1}\|_F^2$, where $\mathcal{X}^t$ represents the recovered HSI at iteration t .

5.5.2. Complexity analysis and efficiency trade-offs

To evaluate the computational cost of 3DeepRep, we conduct a quantitative analysis of its **computational efficiency and resource scalability**. Table 6 details the evolution of GPU memory consumption and runtime as the transform modes extend from 1D to 3D. The results exhibit a near-linear scaling behavior for both memory usage and execution time relative to the number of modes. Specifically, the full three-mode configuration completes the reconstruction process in approximately 18 minutes (1065 s). Given the substantial performance gains, this computational cost remains competitive for HSI inpainting tasks, confirming that the integration of multi-directional modeling does not impose a heavy burden in real-world applications.

We further investigate the influence of the latent dimension k on the trade-off between reconstruction fidelity and computational efficiency. As evidenced in Table 6, increasing k beyond 1.0 yields small returns in performance while significantly escalating memory overhead and latency. This trend suggests that a latent dimension of $k = 1.0$ offers sufficient representational capacity, rendering higher-dimensional expansions redundant. Besides, we provide an insight when comparing the model's compactness against the state-of-the-art 1D transform-based method, CoNoT [30]. By constraining the latent dimension to $k = 0.3$, our compressed variant achieves a smaller parameter footprint (48.6 k vs. 54.8 k) and reduces runtime by nearly 50% (357 s vs. 633 s) compared to CoNoT. Despite this reduced complexity, 3DeepRep delivers a marked improvement in reconstruction accuracy (38.657 dB vs. 33.279 dB). These empirical results strongly substantiate that the superiority of 3DeepRep stems from the effectiveness of its multi-directional low-rank characterization rather than mere parameter scaling, ensuring robustness and efficiency.

Moreover, considering practical scenarios where the number of spectral bands increases substantially, the computational burden can be effectively mitigated by reducing the latent dimension k , while incurring only marginal performance degradation.

5.5.3. Convergence of gradient-driven optimization

Finally, we validate the convergence of our fully gradient-driven optimization strategy. Unlike traditional ADMM solvers that require complex variable splitting, our method enables end-to-end joint optimization. Fig. 10 plots the training loss and relative error curves. The rapid and stable decline of the loss confirms that our formulation allows for the efficient computation of exact gradients for all components (latent variables and network weights) using standard backpropagation.

5.6. Robustness assessment and operational boundaries

To evaluate the robustness and delineate the operational boundaries of 3DeepRep, we simulate extreme degradation scenarios that limits HSI inpainting. Fig. 8 visualizes the reconstruction results under a compound extreme setting, which simultaneously imposes 99% random missing pixels, 50% stripe corruption, and deadline patterns.

While 3DeepRep demonstrates superior resilience compared to competing methods, performance degradation is inevitably observed under such severe conditions. We analyze these limitations from two critical perspectives:

Sensitivity to Extreme Sparsity. The fundamental issue of low-rank tensor completion is that sufficient observed entries must exist to uniquely characterize the underlying low-rank structure. When data sparsity reaches pathological levels (e.g., the compound setting in Fig. 8), the remaining information give an extremely ill-posed inverse problem. In such scenario, fine-grained textures are smoothed out, and only coarse structural skeletons are preserved. However, as evidenced in Fig. 8, 3DeepRep maintains a significantly higher PSNR and better visual coherence than other baselines. This suggests that our multi-directional deep prior provides robust regularization even when observational data is critically insufficient.

Dependency on Inter-mode Correlations. 3DeepRep relies on the synergy between spatial and spectral correlations. A specific failure mode arises in full-band structural corruption, where missing patterns (e.g., thick stripes or dead lines) span identical spatial locations across all spectral bands. In such degenerate cases, spectral-based regularization becomes ineffective, as the spectral correlation is rendered uninformative for imputation. As observed in the real-world example in Fig. 7, slight residual artifacts persist in the reconstruction because standard spectral low-rank priors cannot effectively suppress artifacts that are spectrally consistent (i.e., present in every band). Nevertheless, unlike 1D transform-based methods (e.g., S2NTNN and CoNoT) that rely heavily on spectral transforms, 3DeepRep's 3D design enables the exploitation of spatial correlations along the remaining two modes. This allows it to partially mitigate full-band artifacts, yielding visually superior results, although complete artifact removal remains theoretically challenging without external priors.

In summary, 3DeepRep is most effective for scenarios where (i) the missing ratio permits the basic structural semantics, and (ii) corruptions do not completely sever correlations across all three tensor modes simultaneously. Within these boundaries, the proposed multi-directional deep representation offers a state-of-the-art solution; beyond them, it exhibits a robust reconstruction compared with existing baselines.

6. Conclusion

This paper proposes 3DeepRep, a 3-directional deep low-rank tensor representation model for HSI inpainting. 3DeepRep learns multi-directional deep transforms along all tensor modes and introduces a learnable aggregation module to adaptively fuse directional reconstructions. The proposed model is trained in a fully gradient-driven, self-supervised manner by jointly optimizing the transform parameters and low-rank latent tensors, thereby eliminating handcrafted tuning while improving parameter efficiency. Extensive experiments on three benchmark datasets demonstrate that 3DeepRep consistently outperforms state-of-the-art model-guided and learning-guided methods, particularly in challenging scenarios involving complex, mixed degradations. Ablation studies further confirm the effectiveness of direct optimization of latent tensors, multi-directional modeling, and adaptive aggregation.

3DeepRep provides a compact, stable, and self-supervised framework that effectively bridges low-rank tensor representation and self-supervised learning, offering a promising solution for HSI recovery. Future work will focus on reducing the computational overhead associated with nuclear norm minimization by integrating more efficient low-rank tensor decomposition strategies or implicit regularization techniques, thereby further improving scalability to large-scale hyperspectral datasets.

CRedit authorship contribution statement

Yunshan Li: Writing – original draft, Validation, Methodology, Conceptualization. **Wenwu Gong:** Writing – review & editing, Methodology, Conceptualization. **Qianqian Wang:** Software,

Resources. **Chao Wang:** Supervision, Resources. **Lili Yang:** Supervision, Resources.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

The authors would like to thank Xi-le Zhao for his valuable guidance and insightful suggestions throughout the development of this work. The authors also appreciate the contributions of researchers who generously shared their code, which facilitated comparison and reproducibility in our experimental evaluations.

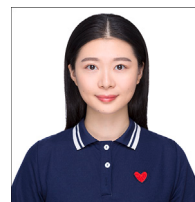
Data availability

The experimental data are available at the link provided in the footnotes.

References

- [1] M.J. Khan, H.S. Khan, A. Yousaf, K. Khurshid, A. Abbas, Modern trends in hyperspectral image analysis: a review, *IEEE Access* 6 (2018) 14118–14129.
- [2] R. Booyens, S. Lorenz, S.T. Thiele, W.C. Fuchsloch, T. Marais, P.A.M. Nex, R. Gloaguen, Accurate hyperspectral imaging of mineralised outcrops: an example from lithium-bearing pegmatites at uis, Namibia, *Remote Sens. Environ.* 269 (2022) 112790.
- [3] B.G. Ram, P. Oduor, C. Igathinathane, K. Howatt, X. Sun, A systematic review of hyperspectral imaging in precision agriculture: analysis of its current state and future prospects, *Comput. Electron. Agric.* 222 (2024) 109037.
- [4] H. Chen, P. Wu, W. Wen, N. Zeng, Dla-net: a dynamically learnable attention network for intelligent surface visual inspection of aero-engine blades, *IEEE Trans. Instrum. Meas.* 74 (2025) 1–14.
- [5] P. Wu, W. Wen, H. Li, N. Zeng, Ai-driven Automation of aviation equipment inspection: insights from a complex adaptive systems perspective, *The Innov.* 7 (1) (2026) 101084.
- [6] X. Sun, Y. Qu, L. Gao, X. Sun, H. Qi, B. Zhang, T. Shen, Target detection through tree-structured encoding for hyperspectral images, *IEEE Trans. Geosci. Remote Sens.* 59 (2021) 4233–4249.
- [7] D.P. Yadav, D. Kumar, A.S. Jalal, B. Sharma, Hyperspectral image denoising through hybrid spectral transformer network, *Adv. Space Res.* 76 (11) (2025) 6673–6693.
- [8] M. Ahmad, S. Distefano, A.M. Khan, M. Mazzara, C. Li, H. Li, J. Aryal, Y. Ding, G. Vivone, D. Hong, A comprehensive survey for hyperspectral image classification: the evolution from conventional to transformers and mamba models, *Neurocomputing* (2025) 130428.
- [9] A. Bhargava, A. Sachdeva, K. Sharma, M.H. Alsharif, P. Uthansakul, M. Uthansakul, Hyperspectral imaging and its applications: a review, *Heliyon* 10 (12) (2024) e33208.
- [10] T.G. Kolda, B.W. Bader, Tensor decompositions and applications, *SIAM Rev.* 51 (2009) 455–500.
- [11] J. Liang, Y. Gao, Lightweight memory-driven self-attention for hyperspectral image classification with cnn-transformer cross-feature fusion, *Neurocomputing* (2025) 130998.
- [12] C.-H. Lin, Y. Liu, C.-Y. Chi, C.-C. Hsu, H. Ren, T.Q. Quek, Hyperspectral tensor completion using low-rank modeling and convex functional analysis, *IEEE Trans. Neural Netw. Learn. Syst.* 35 (8) (2023) 10736–10750.
- [13] W. Gong, Z. Huang, L. Yang, Accurate regularized tucker decomposition for image restoration, *Appl. Math. Model.* 123 (11) (2023) 75–86, <https://doi.org/10.1016/j.apm.2023.06.031>
- [14] J. Cai, J. Jiang, Y. Wang, J. Zheng, H. Xu, Image inpainting via smooth tucker decomposition and low-rank hankel constraint, *Int. J. Comput. Appl.* 45 (6) (2023) 421–432.
- [15] Q. Zhao, G. Zhou, S. Xie, L. Zhang, A. Cichocki, Tensor ring decomposition, *arXiv arxiv:1606.05535*, 2016.
- [16] Y.-B. Zheng, T. Huang, X. Zhao, Q. Zhao, T.-X. Jiang, Fully-connected tensor network decomposition and its application to higher-order tensor completion, in: *AAAI Conference on Artificial Intelligence*, 2021.
- [17] Y.-B. Zheng, T.-Z. Huang, X.-L. Zhao, Q. Zhao, Tensor completion via fully-connected tensor network decomposition with regularized factors, *J. Sci. Comput.* 92 (1) (2022) 8.
- [18] Z. Tu, J. Lu, H. Zhu, W. Hu, Q. Jiang, M.K. Ng, Fully-connected tensor network decomposition and group sparsity for multitemporal images cloud removal, *Inverse Probl. Imaging* 19 (1) (2025) 59–86.
- [19] X.-Y. Liu, S. Aeron, V. Aggarwal, X. Wang, Low-tubal-rank tensor completion using alternating minimization, *IEEE Trans. Inf. Theory* 66 (2016) 1714–1737.
- [20] R. Dian, S. Li, Hyperspectral image super-resolution via subspace-based low tensor multi-rank regularization, *IEEE Trans. Image Process.* 28 (2019) 5135–5146.
- [21] X. Xiao, Y. Chen, Y. jiao Gong, Y. Zhou, Low-rank preserving t-linear projection for robust image feature extraction, *IEEE Trans. Image Process.* 30 (2020) 108–120.
- [22] Z. Zhang, G. Ely, S. Aeron, N. Hao, M.E. Kilmer, Novel methods for multilinear data completion and de-noising based on tensor-SVD, in: *2014 IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 3842–3849.
- [23] W.-H. Xu, X. Zhao, M.K.P. Ng, A fast algorithm for cosine transform based tensor singular value decomposition, *arXiv arxiv:1902.03070*, 2019.
- [24] T.-X. Jiang, M.K.P. Ng, X. Zhao, T. Huang, Framelet representation of tensor nuclear norm for third-order tensor completion, *IEEE Trans. Image Process.* 29 (2019) 7233–7244.
- [25] G. Song, M.K. Ng, X. Zhang, Robust tensor completion using transformed tensor singular value decomposition, *Numer. Linear Algebra Appl.* 27 (3) (2020) e2299.
- [26] H. Kong, C. Lu, Z. Lin, Tensor q-rank: new data dependent definition of tensor rank, *Mach. Learn.* 110 (2019) 1867–1900.
- [27] T.-X. Jiang, X. Zhao, H. Zhang, M.K.P. Ng, Dictionary learning with Low-Rank coding coefficients for tensor completion, *IEEE Trans. Neural Netw. Learn. Syst.* 34 (2020) 932–946.
- [28] B.-Z. Li, X.-L. Zhao, T.-Y. Ji, X.-J. Zhang, T.-Z. Huang, Nonlinear transform induced tensor nuclear norm for tensor completion, *J. Sci. Comput.* 92 (3) (2022) 83.
- [29] Y.-S. Luo, X.-L. Zhao, T.-X. Jiang, Y. Chang, M.K. Ng, C. Li, Self-supervised nonlinear transform-based tensor nuclear norm for multi-dimensional image recovery, *IEEE Trans. Image Process.* 31 (2022) 3793–3808.
- [30] J.-L. Wang, T.-Z. Huang, X.-L. Zhao, Y.-S. Luo, T.-X. Jiang, CoNoT: coupled nonlinear transform-based low-rank tensor representation for multidimensional image completion, *IEEE Trans. Neural Netw. Learn. Syst.* 35 (7) (2022) 8969–8983.
- [31] Y.-B. Zheng, T. Huang, X. Zhao, T.-X. Jiang, T.-Y. Ji, T.-H. Ma, Tensor n-tubal rank and its convex relaxation for low-rank tensor recovery, *Inf. Sci.* 532 (2018) 170–189.
- [32] G. Li, Z. Tu, J. Lu, C. Wang, L. Shen, Multi-dimensional image recovery via self-supervised nonlinear transform based a three-directional tensor nuclear norm, *Numer. Math.-Theory Methods Appl.* 17 (3) (2024) 727–750.
- [33] J. Bacca, Y. Fonseca, H. Arguello, Compressive spectral image reconstruction using deep prior and low-rank tensor representation, *Appl. Opt.* 60 14 (2021) 4197–4207.
- [34] Y. Luo, X. Zhao, D. Meng, T.-X. Jiang, HLRTF: hierarchical low-rank tensor factorization for inverse problems in multi-dimensional imaging, in: *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022, pp. 19281–19290.
- [35] R. Su, X.-L. Zhao, W.-H. Wu, S. Liu, J. He, Deep fully-connected tensor network decomposition for multi-dimensional signal recovery, *Signal Process.* 233 (2025) 109903.
- [36] Y.-B. Zheng, T.-Z. Huang, X.-L. Zhao, T.-X. Jiang, T.-H. Ma, T.-Y. Ji, Mixed noise removal in hyperspectral image via low-fibered-rank regularization, *IEEE Trans. Geosci. Remote Sens.* 58 (1) (2019) 734–749.
- [37] X. Cheng, W. Kong, X. Luo, W. Qin, F. Zhang, J. Wang, Tensor completion via joint reweighted tensor q-nuclear norm for visual data recovery, *Signal Process.* 219 (2024) 109407.
- [38] T.-W. Zhou, X.-L. Zhao, J.-L. Wang, Y.-S. Luo, M. Wang, X.-X. Bai, H. Yan, DTR: a unified deep tensor representation framework for multimedia data recovery, *IEEE Trans. Multimedia* (2025) 1–11.
- [39] A.L. Maas, A. Hannun y, A. Ng y, et al., Rectifier nonlinearities improve neural network acoustic models, in: *Proc. Icml*, vol. 30, Atlanta, 2013, pp. 3, GA.
- [40] H. Wang, J. Peng, W. Qin, J. Wang, D. Meng, Guaranteed tensor recovery fused low-rankness and smoothness, *IEEE Trans. Pattern Anal. Mach. Intell.* 45 (2023) 10990–11007.
- [41] J.-F. Cai, E.J. Candès, Z. Shen, A singular value thresholding algorithm for matrix completion, *SIAM J. Optim.* 20 (4) (2010) 1956–1982.
- [42] K.D.B.J. Adam, et al., A method for stochastic optimization, *arXiv preprint arXiv:1412.6980*, 2014, vol. 1412, no. 6.
- [43] Z. Zhang, S. Aeron, Exact tensor completion using t-SVD, *IEEE Trans. Signal Process.* 65 (2015) 1511–1526.
- [44] L. Pang, X. Rui, L. Cui, H. Wang, D. Meng, X. Cao, Hir-diff: unsupervised hyperspectral image restoration via improved diffusion models, in: *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2024, pp. 3005–3014.

Author biography



Yunshan Li received her M.S. degree in Mathematics from Southern University of Science and Technology, Shenzhen, China, in 2022. She is currently pursuing a Ph.D. degree in Mathematics with the Department of Statistics and Data Science at Southern University of Science and Technology, Shenzhen, China. Her research interests include low-rank tensor modeling and medical image analysis.



Wenwu Gong is currently a Postdoctoral at the Southern University of Science and Technology (SUSTech), where he also received his Ph.D. in Mathematics in 2024. His research interests include low-rank tensor representation, interpretable artificial intelligence, and their applications in computational imaging and data science.



Qianqian Wang received her B.S. degree in Information and Data Science from Henan University, Kaifeng, China, in 2018, and her M.S. degree in Mathematics from Southern University of Science and Technology, Shenzhen, China, in 2022. She is currently pursuing a Ph.D. degree in Mathematics with the Department of Statistics and Data Science at Southern University of Science and Technology, Shenzhen, China. Her research interests include causal machine learning theory, debiasing in recommendation systems, and out-of-distribution generalization.



Chao Wang received the Ph.D. degree in mathematics from The Chinese University of Hong Kong, Hong Kong, in 2017. He is currently an Assistant Professor with the Southern University of Science and Technology (SUSTech), Shenzhen, China. His research interests include scientific computing, compressed sensing, interdisciplinary mathematical modeling, convex and nonconvex optimization, medical imaging, machine learning, and numerical linear algebra.



Lili Yang is a Professor in the Department of Statistics and Data Science at Southern University of Science and Technology. Her main research areas include big data analytics, data imputation, deep learning, risk analysis, and modeling. She has published over 150 papers in journals such as Nature sub-journals, European Journal of Operational Research (EJOR), Information Systems Research (ISR), Environmental Research Letters, and Applied Energy, among others. Previously, she served as a professor and university council member at Loughborough University, UK. She holds both a Master's and a Ph.D. in Computer Science from institutions in the United Kingdom. Dr. Yang is a Fellow of the British Computer Society and a Chartered IT Professional. She has led and completed more than 20 major research projects funded by organizations such as the European Commission (FP7), the UK Engineering and Physical Sciences Research Council (EPSRC), and the UK Ministry of Defence (MoD). Her research results have been successfully applied in practice. Dr. Yang was nominated for national-level awards, including the Times Higher Education (THE) Awards for International Impact and the British IT Industry Awards for Innovation and Entrepreneurship. Since returning to China, she has led or participated in 15 projects funded by the National Natural Science Foundation of China (NSFC), the Ministry of Science and Technology, Guangdong Province, and Shenzhen Municipality. Her research findings have been applied in the Shenzhen Civic Center and earned the 2021 Science and Technology Progress Award, jointly awarded by the Ministry of Science and Technology and the Chinese Society for Measurement.