TENSOR HOURGLASS DECOMPOSITION FOR HYPERSPECTRAL IMAGE SUPER-RESOLUTION

Anonymous authors
Paper under double-blind review

ABSTRACT

Hyperspectral image super-resolution (HSISR) aims to reconstruct a high-resolution hyperspectral image from a low-resolution hyperspectral image and a high-resolution multispectral image. A central challenge is to recover fine spatial details from the multispectral observation while maintaining consistency with the spectral information provided by the hyperspectral observation. To address this challenge, we propose a tensor hourglass decomposition (THD) model that separately characterizes heterogeneous spatial and spectral structures while preserving their high-order interactions. THD organizes the spatial and spectral representations into two closed peripheral cycles and couples their spoke modes through a shared high-order waist core, thereby enabling flexible many-to-many interactions between spatial structures and spectral signatures. By exploiting the multilinearity of THD, the spatial degradation and spectral response operators are directly absorbed into the corresponding physical-mode-carrying factors, yielding a structured coupled observation model for HSISR. Extensive experiments on benchmark datasets demonstrate that the proposed method achieves competitive or superior spatial–spectral reconstruction performance.

1 INTRODUCTION

Hyperspectral images (HSIs) record dense reflectance signatures across contiguous spectral bands Hong et al. (2026); Xu et al. (2026c), providing rich spectral information for environmental monitoring Wilson (2026) and target detection He et al. (2024). However, in optical imaging systems, the trade-off among spatial resolution, spectral resolution, and signal-to-noise ratio typically limits its spatial detail. In contrast, multispectral sensors capture the same scene at substantially higher spatial resolution but with only a few broad bands. Hyperspectral image super-resolution aims to reconstruct a high-resolution HSI (HR-HSI) by combining the spectral fidelity of a low-resolution HSI (LR-HSI) with the spatial detail of a high-resolution multispectral image (HR-MSI).

Existing HSISR methods can be broadly categorized into data-driven and model-driven approaches. Deep neural networks have demonstrated strong reconstruction performance owing to their nonlinear representation capacity Xu et al. (2025a; 2026a;b); Liu et al. (2026); Wang et al. (2026). However, supervised deep learning methods generally rely on simulated training pairs and may generalize poorly when the sensor characteristics Yang et al. (2026); Huang et al. (2026); Miao et al. (2026), spatial point-spread function (PSF), spectral response, or noise distribution at test time differ from those used during training. In contrast, model-driven methods explicitly incorporate spatial and spectral degradation processes, offering clear physical interpretability and independence from external paired training data.

Low-rank tensor modeling has become an important model-driven paradigm for HSISR because it preserves the multidimensional structure of hyperspectral data while exploiting their intrinsic redundancy. Beyond classical CP and Tucker models Li et al. (2018), tensor-ring (TR) Xu et al. (2021) decomposition arranges third-order cores in a cyclic network, offering flexible representation but primarily propagating information through neighboring connections. Fully connected tensor network (FCTN) Zheng et al. (2022); Su et al. (2025) strengthens global interaction by connecting every pair of modes, at the cost of rapidly increasing connectivity and model complexity. Tensor-wheel (TW) decomposition Wu et al. (2026) instead introduces a central core to coordinate peripheral factors, yielding a more compact global interaction mechanism. Despite their different topologies,

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

these methods generally organize all tensor modes within a single homogeneous network, without explicitly separating group-specific structures from inter-group transformations.

Such a unified treatment is not fully aligned with the physical structure of an HR-HSI. The height and width modes jointly encode spatial geometry, local textures, and nonlocal similarities, whereas the spectral mode is primarily characterized by material-dependent signatures, smooth inter-band variations, and a compact latent subspace. These heterogeneous properties suggest that spatial and spectral information should be represented by group-specific structures rather than being forced into an identical tensor-network organization. Meanwhile, the two groups should remain jointly coordinated so that spatial details can be transferred without compromising spectral fidelity. Therefore, an effective representation should simultaneously preserve group-specific dependencies and model their higher-order interaction.

Motivated by this observation, we propose Tensor Hourglass Decomposition (THD), a partition-aware tensor-network architecture for modeling heterogeneous mode interactions. Unlike TW, which organizes all physical modes within a single peripheral network coupled to a central core, THD partitions heterogeneous modes into group-specific peripheral cycles and coordinates them through a shared high-order waist core. The peripheral cycles preserve within-group dependencies, while the waist core captures higher-order cross-group interactions, thereby separating group-specific representation from inter-group coordination and better matching the heterogeneous spatial–spectral structure of an HR-HSI.

For HSI–MSI fusion, we specialize THD to the heterogeneous spatial and spectral structures of an HR-HSI. The spatial branch jointly characterizes the height and width modes, while the spectral branch captures the latent dependencies among spectral bands. Their group-specific representations are coordinated by the shared waist core and jointly estimated under the spatial and spectral degradation models of the LR-HSI and HR-MSI. Consequently, the proposed framework integrates spatial structure preservation, spectral fidelity, and cross-group interaction within a structured and physically interpretable reconstruction model.

Our main contributions are as follows.

- We propose Tensor Hourglass Decomposition (THD), a partition-aware tensor network that organizes heterogeneous modes into group-specific peripheral structures coupled by a shared high-order waist core. We derive its formulation, grouped-unfolding form, and bond-dimension-induced rank bounds.
- We develop a physically grounded THD model for HSI–MSI fusion, in which spatial and spectral structures are separately characterized and jointly coupled under the spatial and spectral degradation processes.
- Extensive experiments on simulated benchmarks and real data demonstrate the effectiveness of THD through comparative, ablation, robustness, convergence, and computational-efficiency analyses.

2 Notations and Problem Formulation

Notations Scalars, vectors, matrices, and tensors are denoted by lowercase, bold lowercase, bold uppercase, and calligraphic uppercase letters, respectively. For $\mathcal{X} \in \mathbb{R}^{I_1 \times \cdots \times I_N}$, $\mathcal{X}(i_1, \ldots, i_N)$ and $\boldsymbol{X}_{(n)}$ denote its $(i_1, \ldots, i_N)$th entry and mode-n unfolding, respectively. We use $\|\cdot\|_F$ and $\otimes$ for the Frobenius norm and Kronecker product.

Problem Formulation Let $\mathcal{X} \in \mathbb{R}^{H \times W \times S}$ denote the latent HR-HSI. The available observations are an LR-HSI $\mathcal{Y} \in \mathbb{R}^{h \times w \times S}$ and an HR-MSI $\mathcal{Z} \in \mathbb{R}^{H \times W \times s}$, where $h < H$, $w < W$, and $s < S$. The two observations provide complementary measurements of the target image: $\mathcal{Y}$ retains dense spectral information at reduced spatial resolution, whereas $\mathcal{Z}$ preserves spatial details over a limited number of spectral bands.

Under the commonly used separable spatial-degradation model, the observation processes are expressed as

$$\mathcal{Y} = \mathcal{X} \times_1 \boldsymbol{P}_1 \times_2 \boldsymbol{P}_2 + \mathcal{N}_Y, \qquad \mathcal{Z} = \mathcal{X} \times_3 \boldsymbol{P}_3 + \mathcal{N}_Z. \tag{1}$$

where $\boldsymbol{P}_1 \in \mathbb{R}^{h\times H}$ and $\boldsymbol{P}_2 \in \mathbb{R}^{w\times W}$ denote the spatial degradation operators accounting for blur and downsampling, and $\boldsymbol{P}_3 \in \mathbb{R}^{s\times S}$ is the spectral response matrix. The tensors $\mathcal{N}_Y \in \mathbb{R}^{h\times w\times S}$ and $\mathcal{N}_Z \in \mathbb{R}^{H\times W\times s}$ denote noise and modeling errors.

Accordingly, recovering the latent HR-HSI can be generally formulated as

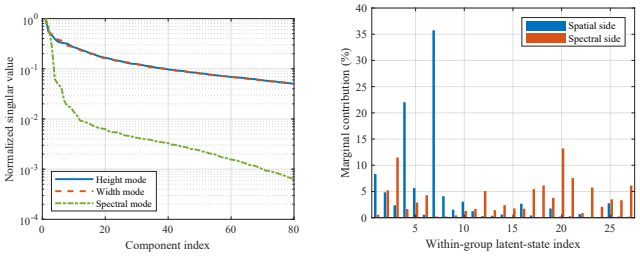
$$\min_{\mathcal{X}} \quad \frac{1}{2}\left\|\mathcal{Y} - \mathcal{X} \times_1 \boldsymbol{P}_1 \times_2 \boldsymbol{P}_2\right\|_F^2 + \frac{\lambda}{2}\left\|\mathcal{Z} - \mathcal{X} \times_3 \boldsymbol{P}_3\right\|_F^2, \tag{2}$$

where $\lambda > 0$ balances the contributions of the LR-HSI and HR-MSI observations.

3 Tensor Hourglass Decomposition

3.1 Motivation and General Topology

A natural way to solve the above HSISR problem is to parameterize the latent HR-HSI using a low-rank tensor-network decomposition. The representation capability of a tensor-network decomposition is largely determined by its topology. However, most existing models organize all physical modes within a unified topology Zheng et al. (2021); Wu et al. (2026), despite their heterogeneous low-rank characteristics and latent dependencies. As shown in Fig. 1, the height and width modes exhibit similar singular-value decay, whereas the spectral mode decays faster and presents distinct latent contributions. This motivates jointly modeling structurally related modes while separating heterogeneous groups. Accordingly, consider an Nth-order tensor $\mathcal{X} \in \mathbb{R}^{I_1\times\cdots\times I_N}$ whose physical modes are organized into two ordered groups:



(a) (b)

Figure 1: Mode-wise heterogeneity of a hyperspectral tensor: rank decay per mode and marginal latent contributions.

$$\boldsymbol{\pi}_L = (\pi_1, \ldots, \pi_K), \qquad \boldsymbol{\pi}_R = (\rho_1, \ldots, \rho_J), \tag{3}$$

where $K + J = N$, and $(\pi_1, \ldots, \pi_K, \rho_1, \ldots, \rho_J)$ is a permutation of $(1, \ldots, N)$. Thus, each physical mode belongs to exactly one group.

3.2 Construction and Definition of THD

Following the above mode grouping, we construct one cyclic peripheral representation for each mode group. Let $M_L \geq K$ and $M_R \geq J$ denote the lengths of the left and right peripheral cycles, respectively. For each peripheral position, its dimension and corresponding index are defined as

$$(\widetilde{I}_k^L, \widetilde{i}_k^L) = \begin{cases} (I_{\pi_k}, i_{\pi_k}), & k \leq K, \\ (1,1), & \text{otherwise}, \end{cases} \qquad (\widetilde{I}_j^R, \widetilde{i}_j^R) = \begin{cases} (I_{\rho_j}, i_{\rho_j}), & j \leq J, \\ (1,1), & \text{otherwise}. \end{cases} \tag{4}$$

The first K and J positions carry the original physical modes in $\boldsymbol{\pi}_L$ and $\boldsymbol{\pi}_R$, respectively, whereas the remaining positions are virtual singleton positions. Although their position indices are fixed to one, their cyclic and spoke dimensions remain active in the latent contractions.

The left and right peripheral factors are specified as

$$\mathcal{G}_k \in \mathbb{R}^{R_k\times \widetilde{I}_k^L \times L_k \times R_{k+1}}, \quad k = 1, \ldots, M_L, \qquad \mathcal{F}_j \in \mathbb{R}^{\widehat{R}_j \times \widetilde{I}_j^R \times U_j \times \widehat{R}_{j+1}}, \quad j = 1, \ldots, M_R. \tag{5}$$

with cyclic boundary conditions

$$R_{M_L+1} = R_1, \qquad \widehat{R}_{M_R+1} = \widehat{R}_1. \tag{6}$$

Each factor contains one position index, two cyclic indices connecting adjacent factors within the same mode group, and one spoke index connected to the shared waist core. Hence, $\{\mathcal{G}_k\}_{k=1}^{M_L}$ and $\{\mathcal{F}_j\}_{j=1}^{M_R}$ form two closed peripheral cycles for modeling group-specific dependencies.

To coordinate the latent representations of the two mode groups, all spoke indices are jointly connected to a shared high-order waist core. As illustrated in Fig. 2, rank-one coupling yields a substantially larger error, whereas diagonal, low-rank, and full cores provide markedly improved approximations. The full core attains the lowest error in the selected setting, supporting flexible cross-group coupling.

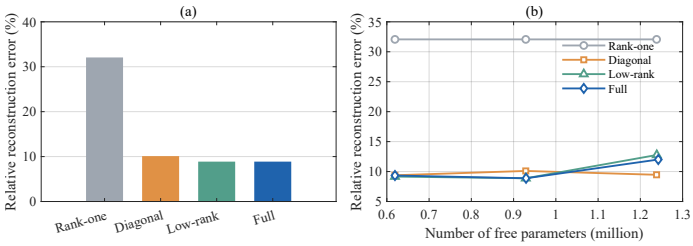


Figure 2: Effect of waist-core parameterization. (a) Relative errors of independently optimized core variants under the selected configuration. (b) Error–parameter trade-offs under varying latent dimensions.

Accordingly, the shared waist core is defined as

$$\mathcal{C} \in \mathbb{R}^{L_1 \times \cdots \times L_{M_L} \times U_1 \times \cdots \times U_{M_R}}. \tag{7}$$

Its first M_L modes are connected to the left peripheral factors, whereas the remaining M_R modes are connected to the right peripheral factors. As illustrated in Fig. 3, the two peripheral cycles model intra-group dependencies, while $\mathcal{C}$ mediates cross-group interactions, resulting in an hourglass-shaped tensor-network topology. Unlike TT and TR, which organize physical modes within a single chain or cycle, and TW, which couples a single peripheral network to a central core, THD separates group-specific peripheral representations from cross-group coordination.

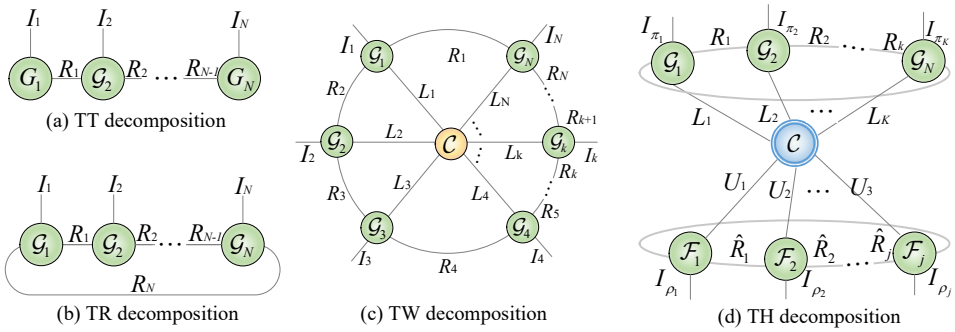


Figure 3: Topological comparison of representative tensor-network decompositions: (a) tensor train; (b) tensor ring; (c) tensor wheel; and (d) the proposed tensor hourglass decomposition.

Based on the above construction, Tensor Hourglass Decomposition is formally defined as follows.

Definition 1 (Tensor Hourglass Decomposition) *Let $\mathcal{X} \in \mathbb{R}^{I_1 \times \cdots \times I_N}$ be an Nth-order tensor whose physical modes are partitioned into two ordered groups $\boldsymbol{\pi}_L = (\pi_1, \ldots, \pi_K)$ and $\boldsymbol{\pi}_R = (\rho_1, \ldots, \rho_J)$, where $K + J = N$ and $(\pi_1, \ldots, \pi_K, \rho_1, \ldots, \rho_J)$ is a permutation of $(1, \ldots, N)$.*

Given peripheral-cycle lengths $M_L \geq K$ and $M_R \geq J$, $\mathcal{X}$ is said to admit a Tensor Hourglass Decomposition if

$$\mathcal{X}(i_1, \ldots, i_N) = \sum_{\boldsymbol{r},\boldsymbol{q},\boldsymbol{l},\boldsymbol{u}} \mathcal{C}(\boldsymbol{l}, \boldsymbol{u}) \prod_{k=1}^{M_L} \mathcal{G}_k(r_k, \widetilde{i}_k^{L}, l_k, r_{k+1}) \prod_{j=1}^{M_R} \mathcal{F}_j(q_j, \widetilde{i}_j^{R}, u_j, q_{j+1}). \tag{8}$$

where $r_{M_L+1} = r_1$, $q_{M_R+1} = q_1$, and all latent indices run over their corresponding dimensions.

For compactness, the representation in equation 8 is denoted by

$$\mathcal{X} = \mathrm{THD}^{(M_L, M_R)}_{\boldsymbol{\pi}_L | \boldsymbol{\pi}_R} \left(\{\mathcal{G}_k\}_{k=1}^{M_L}; \mathcal{C}; \{\mathcal{F}_j\}_{j=1}^{M_R} \right), \tag{9}$$

where the peripheral factors and the shared waist core are collectively referred to as the THD factors.

Remark 1 (Peripheral-Cycle Realizations) *For fixed ordered mode groups, (M_L, M_R) specifies the lengths of the two peripheral cycles. The choice $(M_L, M_R) = (K, J)$ gives the minimal realization, whereas $M_L = M_R$ defines a balanced one; in particular, $M_L = M_R = N$ is order matched. A length-one cycle reduces to a self-contracted factor. Different realizations preserve the physical tensor order and dimensions but induce different latent parameterizations and capacities.*

Definition 2 (THD-Rank) *For a fixed peripheral-cycle realization (M_L, M_R), a tuple*

$$\boldsymbol{r}_{\mathrm{THD}}^{(M_L,M_R)} = \big(R_1, \ldots, R_{M_L}; \widehat{R}_1, \ldots, \widehat{R}_{M_R}; L_1, \ldots, L_{M_L}; U_1, \ldots, U_{M_R}\big) \tag{10}$$

is called an admissible THD-rank of $\mathcal{X}$ if $\mathcal{X}$ admits a THD representation with these cyclic and spoke dimensions.

As in other tensor-network models, the THD-rank specifies the internal bond dimensions of a particular realization and is not necessarily unique.

3.3 Algebraic Properties of THD

Proofs of all results in this subsection are deferred to Appendix.

To characterize the interaction between the two peripheral cycles, let $I_L = \prod_{k=1}^{K} I_{\pi_k}$ and $I_R = \prod_{j=1}^{J} I_{\rho_j}$ denote the aggregate physical-mode dimensions, and let $L = \prod_{k=1}^{M_L} L_k$ and $U = \prod_{j=1}^{M_R} U_j$ denote the aggregate spoke dimensions. The grouped unfolding associated with $(\boldsymbol{\pi}_L, \boldsymbol{\pi}_R)$ is

$$\boldsymbol{X}_{\langle L|R\rangle} = \mathrm{reshape}\big(\mathrm{permute}(\mathcal{X}, [\boldsymbol{\pi}_L, \boldsymbol{\pi}_R]), I_L, I_R\big). \tag{11}$$

With $r_{M_L+1} = r_1$ and $q_{M_R+1} = q_1$, the peripheral contractions are

$$\begin{aligned} \mathcal{G}^{(L)}\big(i_{\pi_1}, \ldots, i_{\pi_K}, l_1, \ldots, l_{M_L}\big) &= \sum_{r_1,\ldots,r_{M_L}} \prod_{k=1}^{M_L} \mathcal{G}_k\big(r_k, \widetilde{i}_k^{L}, l_k, r_{k+1}\big), \\ \mathcal{F}^{(R)}\big(i_{\rho_1}, \ldots, i_{\rho_J}, u_1, \ldots, u_{M_R}\big) &= \sum_{q_1,\ldots,q_{M_R}} \prod_{j=1}^{M_R} \mathcal{F}_j\big(q_j, \widetilde{i}_j^{R}, u_j, q_{j+1}\big). \end{aligned} \tag{12}$$

The corresponding peripheral and core matrices are

$$\begin{aligned} &\boldsymbol{G} = \mathrm{reshape}\left(\mathcal{G}^{(L)}, I_L, L\right) \in \mathbb{R}^{I_L \times L}, \boldsymbol{F} \quad = \mathrm{reshape}\left(\mathcal{F}^{(R)}, I_R, U\right) \in \mathbb{R}^{I_R \times U}, \\ &\boldsymbol{C} = \mathrm{reshape}\left(\mathcal{C}, L, U\right) \in \mathbb{R}^{L \times U}. \end{aligned} \tag{13}$$

The above grouped contractions lead to the following compact matrix representation of THD.

Theorem 1 (Grouped-Unfolding Form of THD) *If $\mathcal{X}$ admits the THD representation in equation 8, then its grouped unfolding satisfies*

$$\boldsymbol{X}_{\langle L|R\rangle} = \boldsymbol{G}\boldsymbol{C}\boldsymbol{F}^{\top}. \tag{14}$$

Theorem 1 expresses THD as $\boldsymbol{X}_{\langle L|R\rangle} = \boldsymbol{G}\boldsymbol{C}\boldsymbol{F}^{\top}$, where $\boldsymbol{G}$ and $\boldsymbol{F}$ are generated by two disjoint group-specific peripheral cycles, while $\boldsymbol{C}$ coordinates their latent representations. This explicitly separates within-group dependency modeling from cross-group coupling. Although $\boldsymbol{C}$ can be absorbed into either group matrix, the resulting matrix need not retain the same peripheral-cycle parameterization. Thus, THD is not an unconstrained three-factor factorization, but a structured group-wise representation coupled through the shared waist core.

We next isolate each peripheral factor for block-wise optimization.

Proposition 1 (Single-Factor Separation of THD) *When all THD factors except one peripheral factor are fixed, the represented tensor is linear with respect to the mode-2 unfolding of that factor:*

$$\begin{cases} \boldsymbol{X}_{(\pi_k)} = \boldsymbol{G}_{k(2)}\boldsymbol{M}_{\neq\mathcal{G}_k}, & 1 \le k \le K, \\ \mathrm{vec}(\mathcal{X})^{\top} = \boldsymbol{G}_{k(2)}\boldsymbol{M}_{\neq\mathcal{G}_k}, & K < k \le M_L, \end{cases} \quad \begin{cases} \boldsymbol{X}_{(\rho_j)} = \boldsymbol{F}_{j(2)}\boldsymbol{M}_{\neq\mathcal{F}_j}, & 1 \le j \le J, \\ \mathrm{vec}(\mathcal{X})^{\top} = \boldsymbol{F}_{j(2)}\boldsymbol{M}_{\neq\mathcal{F}_j}, & J < j \le M_R. \end{cases} \tag{15}$$

Here, $\boldsymbol{G}_{k(2)}$ and $\boldsymbol{F}_{j(2)}$ denote the mode-2 unfoldings of $\mathcal{G}_k$ and $\mathcal{F}_j$, respectively. The corresponding environment matrices satisfy

$$\begin{aligned} \boldsymbol{M}_{\neq \mathcal{G}_k} &\in \mathbb{R}^{R_k L_k R_{k+1} \times \left((\prod_{n=1}^{N} I_n)/\widetilde{I}_k^L\right)}, \quad k = 1, \ldots, M_L, \\ \boldsymbol{M}_{\neq \mathcal{F}_j} &\in \mathbb{R}^{\widehat{R}_j U_j \widehat{R}_{j+1} \times \left((\prod_{n=1}^{N} I_n)/\widetilde{I}_j^R\right)}, \quad j = 1, \ldots, M_R. \end{aligned} \tag{16}$$

Any fixed column permutation induced by the adopted unfolding or vectorization convention is absorbed into the corresponding environment matrix.

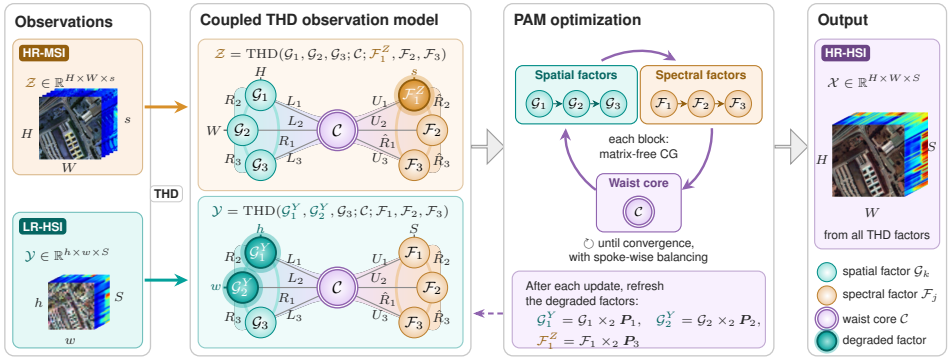


Figure 4: Overview of the proposed THD-based HSISR framework.

Together, the factor-wise linear forms and grouped-unfolding representation yield the following rank bounds.

Proposition 2 (Grouped- and Mode-Wise Rank Bounds) *Let $\mathcal{X}$ admit the THD representation in equation 8. Then, its grouped unfolding satisfies*

$$\operatorname{rank}\left(\boldsymbol{X}_{\langle L|R\rangle}\right) \le \min\{\operatorname{rank}(\boldsymbol{G}), \operatorname{rank}(\boldsymbol{C}), \operatorname{rank}(\boldsymbol{F})\} \le \min\{I_L, I_R, L, U\}. \tag{17}$$

Moreover, the mode-wise unfoldings associated with the original physical modes satisfy

$$\begin{aligned} \operatorname{rank}\left(\boldsymbol{X}_{(\pi_k)}\right) &\le \min\{I_{\pi_k}, R_k L_k R_{k+1}\}, \quad k = 1, \ldots, K, \\ \operatorname{rank}\left(\boldsymbol{X}_{(\rho_j)}\right) &\le \min\left\{I_{\rho_j}, \widehat{R}_j U_j \widehat{R}_{j+1}\right\}, \quad j = 1, \ldots, J. \end{aligned} \tag{18}$$

Together, Theorem 1 and Propositions 1–2 provide the algebraic basis for the fusion model: the grouped form separates group-specific representations and cross-group coupling, the factor-wise linearity enables the block-wise matrix-free optimization in Section 4, and the rank bounds characterize the mode-wise and cross-group capacities controlled by the THD ranks.

4 THD-Based HSI–MSI Fusion Model

Motivated by the ability of THD to separately characterize heterogeneous spatial and spectral structures while preserving their high-order interactions, we specialize THD to the latent HR-HSI $\mathcal{X} \in \mathbb{R}^{H \times W \times S}$. The spatial modes are assigned to the left group, whereas the spectral mode is assigned to the right group, i.e.,

$$\boldsymbol{\pi}_L = (1, 2), \qquad \boldsymbol{\pi}_R = (3), \qquad K = 2, \qquad J = 1. \tag{19}$$

We adopt the order-matched balanced realization $M_L = M_R = N = 3$. Accordingly, the effective position dimensions of the spatial and spectral peripheral cycles are

$$\left(\widetilde{I}_1^L, \widetilde{I}_2^L, \widetilde{I}_3^L\right) = (H, W, 1), \qquad \left(\widetilde{I}_1^R, \widetilde{I}_2^R, \widetilde{I}_3^R\right) = (S, 1, 1). \tag{20}$$

Under this realization, the latent HR-HSI is represented as

$$\mathcal{X} = \mathrm{THD}^{(3,3)}_{(1,2)|(3)}\left(\{\mathcal{G}_k\}_{k=1}^3; \mathcal{C}; \{\mathcal{F}_j\}_{j=1}^3\right). \tag{21}$$

The spatial and spectral groups are represented by $\{\mathcal{G}_k\}_{k=1}^3$ and $\{\mathcal{F}_j\}_{j=1}^3$, respectively, and are coupled through a shared sixth-order waist core $\mathcal{C}$. Their factor dimensions are

$$\begin{aligned} \mathcal{G}_1 &\in \mathbb{R}^{R_1\times H\times L_1\times R_2}, \quad \mathcal{G}_2 \in \mathbb{R}^{R_2\times W\times L_2\times R_3}, \quad \mathcal{G}_3 \in \mathbb{R}^{R_3\times 1\times L_3\times R_1}, \\ \mathcal{F}_1 &\in \mathbb{R}^{\widehat{R}_1\times S\times U_1\times \widehat{R}_2}, \quad \mathcal{F}_2 \in \mathbb{R}^{\widehat{R}_2\times 1\times U_2\times \widehat{R}_3}, \quad \mathcal{F}_3 \in \mathbb{R}^{\widehat{R}_3\times 1\times U_3\times \widehat{R}_1}. \end{aligned} \tag{22}$$

Here, $\mathcal{G}_1$, $\mathcal{G}_2$, and $\mathcal{F}_1$ carry the height, width, and spectral modes, respectively, whereas $\mathcal{G}_3$, $\mathcal{F}_2$, and $\mathcal{F}_3$ are virtual singleton positions. These positions introduce no additional physical dimensions but close the two three-factor peripheral cycles and remain active in the latent contractions.

The shared waist core is given by

$$\mathcal{C} \in \mathbb{R}^{L_1\times L_2\times L_3\times U_1\times U_2\times U_3}. \tag{23}$$

By jointly coupling the six spoke indices, $\mathcal{C}$ coordinates selective many-to-many interactions between the spatial and spectral peripheral representations.

Under the observation model in equation 1, the spatial degradations act on the physical-position modes of $\mathcal{G}_1$ and $\mathcal{G}_2$, while the spectral response acts on that of $\mathcal{F}_1$. Accordingly, we define

$$\mathcal{G}_1^Y = \mathcal{G}_1 \times_2 \boldsymbol{P}_1, \qquad \mathcal{G}_2^Y = \mathcal{G}_2 \times_2 \boldsymbol{P}_2, \qquad \mathcal{F}_1^Z = \mathcal{F}_1 \times_2 \boldsymbol{P}_3, \tag{24}$$

where $\mathcal{G}_1^Y \in \mathbb{R}^{R_1\times h\times L_1\times R_2}$, $\mathcal{G}_2^Y \in \mathbb{R}^{R_2\times w\times L_2\times R_3}$, and $\mathcal{F}_1^Z \in \mathbb{R}^{\widehat{R}_1\times s\times U_1\times \widehat{R}_2}$.

Owing to the multilinearity of THD with respect to its peripheral factors, the spatial and spectral degradation operators can be absorbed into the corresponding factors:

$$\begin{aligned} \mathcal{X} \times_1 \boldsymbol{P}_1 \times_2 \boldsymbol{P}_2 &= \mathrm{THD}_{(1,2)|(3)}^{(3,3)} \left(\mathcal{G}_1^Y, \mathcal{G}_2^Y, \mathcal{G}_3; \mathcal{C}; \mathcal{F}_1, \mathcal{F}_2, \mathcal{F}_3\right), \\ \mathcal{X} \times_3 \boldsymbol{P}_3 &= \mathrm{THD}_{(1,2)|(3)}^{(3,3)} \left(\mathcal{G}_1, \mathcal{G}_2, \mathcal{G}_3; \mathcal{C}; \mathcal{F}_1^Z, \mathcal{F}_2, \mathcal{F}_3\right). \end{aligned} \tag{25}$$

Accordingly, HSISR is formulated as the joint estimation of the six peripheral factors and the shared waist core:

$$\begin{aligned} \min_{\substack{\mathcal{G}_1,\mathcal{G}_2,\mathcal{G}_3,\\ \mathcal{F}_1,\mathcal{F}_2,\mathcal{F}_3,\mathcal{C}}} & \frac{1}{2}\left\| \mathcal{Y} - \mathrm{THD}_{(1,2)|(3)}^{(3,3)}\left(\mathcal{G}_1^Y,\mathcal{G}_2^Y,\mathcal{G}_3;\mathcal{C};\mathcal{F}_1,\mathcal{F}_2,\mathcal{F}_3\right)\right\|_F^2 \\ & + \frac{\lambda}{2}\left\| \mathcal{Z} - \mathrm{THD}_{(1,2)|(3)}^{(3,3)}\left(\mathcal{G}_1,\mathcal{G}_2,\mathcal{G}_3;\mathcal{C};\mathcal{F}_1^Z,\mathcal{F}_2,\mathcal{F}_3\right)\right\|_F^2. \end{aligned} \tag{26}$$

As illustrated in Fig. 4, the spatial and spectral degradation operators are absorbed into the corresponding THD factors. At each outer iteration, the peripheral factors are updated sequentially as $\mathcal{G}_1 \to \mathcal{G}_2 \to \mathcal{G}_3 \to \mathcal{F}_1 \to \mathcal{F}_2 \to \mathcal{F}_3$, followed by the shared waist core $\mathcal{C}$. Each block subproblem is solved by CG using matrix-free normal operators, with the detailed updates provided in Appendix.

Computational Complexity Analysis We analyze the dominant per-iteration cost of the matrix-free PAM solver for equation 26. Let $d_k^G = R_k L_k R_{k+1}$, $d_j^F = \widehat{R}_j U_j \widehat{R}_{j+1}$, $L = L_1L_2L_3$, $U = U_1U_2U_3$, and $\Omega = hwS + HWs$, where $k, j = 1, 2, 3$, $R_4 = R_1$, and $\widehat{R}_4 = \widehat{R}_1$. The dominant cost of one outer iteration is

$$\begin{aligned} \mathcal{T}_{\mathrm{THD}} = \mathcal{O}\Bigg(& \Omega\left[\sum_{k=1}^{3} N_{\mathcal{G}_k} d_k^G + \sum_{j=1}^{3} N_{\mathcal{F}_j} d_j^F\right] + hw\,U(L+S) \\ & + HW\,U(L+s) + (hw+HW)L^2 + (S+s)U^2 + N_{\mathcal{C}} LU(L+U)\Bigg), \end{aligned} \tag{27}$$

where $N_{\mathcal{G}_k}$, $N_{\mathcal{F}_j}$, and $N_{\mathcal{C}}$ denote the corresponding CG iteration numbers. The matrix-free implementation avoids forming large environment and Kronecker-product coefficient matrices. A detailed derivation is provided in Appendix.

Table 1: Quantitative results on PU, PC, and WDC. Best and second-best results are highlighted in **red** and **blue**, respectively.

Method	Pavia University					Pavia City Center					Washington DC Mall				
	PSNR↑	SSIM↑	SAM↓	ERGAS↓	DD↓	PSNR↑	SSIM↑	SAM↓	ERGAS↓	DD↓	PSNR↑	SSIM↑	SAM↓	ERGAS↓	DD↓
20-dB SNR															
CSTF	29.4022	0.8391	12.7213	4.7804	6.9396	30.8972	0.7893	8.4307	4.3806	5.8261	28.1827	0.8531	11.4790	4.8591	8.0172
LFCTNN	30.6909	0.8619	11.1675	4.3524	6.1157	32.1298	0.8217	7.0683	3.8827	5.0469	**29.5687**	**0.8802**	9.7879	**4.2399**	**6.8599**
GTNN	30.8906	0.8693	10.8997	4.1857	5.9318	32.0510	0.8219	7.2275	3.8498	5.0735	29.2789	0.8745	9.9343	4.3263	7.0571
CTDF	**31.2498**	0.8823	10.2565	**3.8196**	**5.6038**	31.6139	0.8065	7.5147	3.9727	5.3611	28.2974	0.8501	11.4909	4.7387	8.0762
GNLTN	30.7796	0.8633	11.0221	4.3143	6.0631	32.1799	0.8230	6.9416	3.8736	5.0199	29.1303	0.8708	10.2014	4.4009	7.2175
MHTNN	30.8102	0.8639	10.9533	4.3021	6.0422	32.1897	0.8230	6.9344	3.8705	5.0136	29.1179	0.8705	10.1806	4.4073	7.2259
EDIP-Net	29.5495	0.8665	10.5597	4.4827	6.5453	**32.8047**	**0.8925**	**5.3853**	**3.3433**	**4.2707**	27.9329	0.8682	**8.5881**	4.8726	7.7988
EU2ADL	31.1880	**0.8884**	**7.4850**	3.9215	5.8165	24.4350	0.5757	8.9118	8.9176	10.3801	20.2037	0.3699	12.2199	11.3017	19.3440
THD	**33.3185**	**0.9356**	**6.8992**	**3.1226**	**4.2225**	**34.2468**	**0.8931**	**4.8427**	**2.8862**	**3.8758**	**30.7153**	**0.9114**	**7.9639**	**3.4714**	**6.1013**
15-dB SNR															
CSTF	24.1998	0.6392	19.8803	8.8293	12.6449	25.5450	0.5645	15.1910	8.0870	10.8440	23.3517	0.6891	18.0520	8.5431	13.9211
LFCTNN	23.6372	0.6001	21.0644	10.4565	14.1842	26.2871	0.5935	13.7928	7.5142	10.0844	24.4803	0.7289	16.1135	7.7204	12.3877
GTNN	23.9682	0.6152	20.5037	9.9770	13.5836	26.3585	0.5948	13.8509	7.4107	9.9928	24.5281	0.7284	16.0272	7.7030	12.2914
CTDF	24.4933	0.6508	19.3605	8.2918	12.1350	24.9227	0.5261	16.3574	8.5573	11.6507	23.0493	0.6702	18.7164	8.6808	14.5807
GNLTN	23.6551	0.6015	20.8607	10.3907	14.0742	26.2457	0.5914	13.8136	7.5412	10.1281	24.4054	0.7247	16.2003	7.7735	12.4887
MHTNN	23.6499	0.6014	20.8561	10.3955	14.0800	26.2513	0.5920	13.7986	7.5367	10.1213	24.4020	0.7243	16.2001	7.7784	12.4938
EDIP-Net	24.2984	0.6688	18.0631	8.2019	11.9480	**29.0561**	**0.7581**	**8.9056**	**5.0885**	**6.8871**	**25.2781**	**0.7810**	**12.3154**	**6.4248**	**10.6912**
EU2ADL	**28.2747**	**0.8020**	**10.5801**	**5.3944**	**7.8505**	23.6593	0.4992	12.1592	9.7601	11.9951	19.6546	0.3320	15.4702	12.1046	20.7751
THD	**30.3470**	**0.8733**	**8.8003**	**4.1514**	**6.1353**	**30.9320**	**0.8055**	**6.3094**	**4.2043**	**5.6497**	**26.7365**	**0.8355**	**9.5184**	**5.4379**	**9.1541**

5 EXPERIMENTS

For evaluation, we compare THD with six model-driven methods, CSTF Li et al. (2018), CTDF Xu et al. (2024b), LFCTNN Xu et al. (2024a), GTNN Dian et al. (2025), GNLTN Xu et al. (2025b), and MHTNN Tu et al. (2026), and two deep methods, EU2ADL Gao et al. (2023) and EDIP-Net Li et al. (2025). We report PSNR, SSIM, SAM, ERGAS, and DD, with higher PSNR/SSIM and lower SAM/ERGAS/DD preferred. Experimental and implementation details are provided in Appendix.

Quantitative Comparison. Table 1 reports the quantitative results at 20- and 15-dB SNRs. THD achieves the best performance across all datasets and evaluation metrics, demonstrating consistent advantages in both spatial reconstruction and spectral preservation. At 20 dB, THD improves PSNR over the second-best methods by 2.07, 1.44, and 1.15 dB on PU, PC, and WDC, respectively. Under the more challenging 15-dB setting, the corresponding PSNR gains are 2.07, 1.88, and 1.46 dB. Compared with EU2ADL on PU and EDIP-Net on PC and WDC, THD further reduces SAM by 16.82%, 29.15%, and 22.71%, respectively, indicating stronger spectral consistency under severe noise. When the SNR decreases from 20 to 15 dB, its PSNR drops by only 2.97, 3.31, and 3.98 dB on the three datasets, confirming its stable noise robustness.

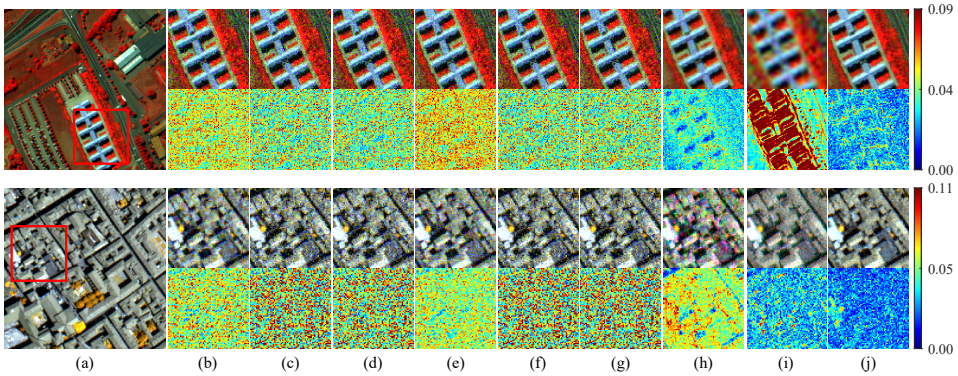


Figure 5: Visual comparison on the original ROSIS (top) and Pavia (bottom) scenes at $\times 4$ downsampling and 15-dB SNR: (a) Reference; (b) CSTF; (c) LFCTNN; (d) GTNN; (e) CTDF; (f) GNLTN; (g) MHTNN; (h) EDIP-Net; (i) EU2ADL; and (j) THD. For (b)–(j), the upper and lower panels show pseudo-color results and absolute-error maps, respectively.

Visual Comparison. Fig. 5 presents visual comparisons on the original ROSIS and Pavia scenes under $\times 4$ downsampling and 15-dB SNR. On ROSIS, CSTF, LFCTNN, GTNN, GNLTN, and MHTNN preserve the main roof structures but leave noticeable noise around parallel edges, while CTDF suppresses noise at the cost of smoothed local details. EDIP-Net introduces residual artifacts, whereas EU2ADL suffers from blurring and color distortion. Similar trends are observed on Pavia, where competing methods exhibit scattered errors, spectral-color deviations, or loss of fine structures. In contrast, THD reconstructs sharper boundaries and pseudo-colors closer to the references, with broader low-error regions in the absolute-error maps.

Real-data experiment. Fig. 6 presents pseudo-color results on a real HSI–MSI pair without reference HR-HSI. The LR-HSI preserves the overall spectral appearance but lacks fine spatial details. Although competing methods recover the main structures, some results remain oversmoothed or show local contrast and color deviations around field boundaries, vegetation, and narrow structures. In contrast, THD produces clearer and more continuous boundaries, preserves richer internal textures, and maintains visual consistency with the LR-HSI, providing qualitative evidence of spatial-detail transfer under real cross-sensor observations.

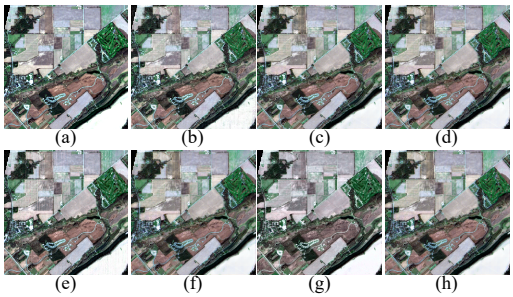


Figure 6: Real-data pseudo-color results (bands 10/9/1): (a) HR-MSI; (b)–(h) CSTF, LFCTNN, GTNN, CTDF, GNLTN, MHTNN, and THD.

5.1 Discussion

Topology Effectiveness Analysis. Fig. 7 compares THD with TT, TR, and TW. All models use the same fidelity objective, optimization protocol, and comparable parameter budgets, with detailed ranks, parameter counts, runtime, and peak memory reported in the Appendix. Fig. 7(a) shows that THD achieves lower reconstruction errors across most of the 93 spectral bands. In Fig. 7(b), the image is divided into 8×8 spatial blocks, where THD attains the minimum error in 94% of all spatial-block–band combinations, confirming its local advantage. Since TW already employs central-core coupling, these results support the benefit of group-specific peripheral partitioning beyond central-core coupling alone.

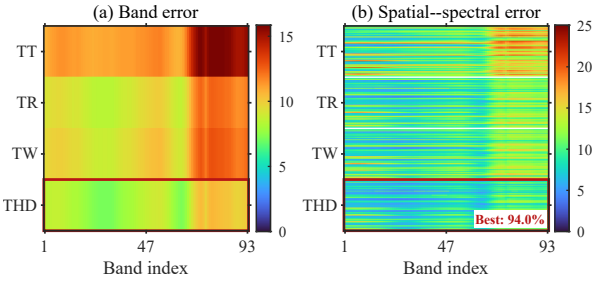


Figure 7: Topology comparison on the PU dataset under $\times4$ downsampling and 15-dB SNR: (a) bandwise RMSE; (b) spatial–spectral blockwise RMSE.

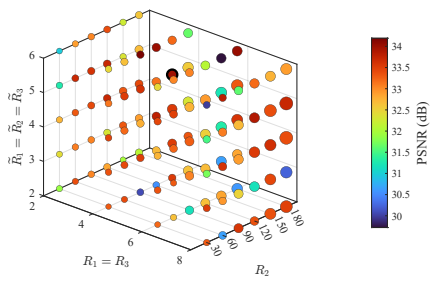


Figure 8: Rank sensitivity of THD. Sphere color and size denote PSNR and runtime, respectively.

Rank Sensitivity. Fig. 8 examines the sensitivity of THD to its cyclic ranks, where sphere color and size denote PSNR and runtime, respectively. We set $R_1 = R_3$ and $\widehat{R}_1 = \widehat{R}_2 = \widehat{R}_3$, while fixing the spoke ranks $[L_1, L_2, L_3, U_1, U_2, U_3]$ as detailed in Appendix. Small ranks limit representation capacity, whereas large ranks increase runtime without consistent gains. The rank configuration $[R_1, R_2, R_3, \widehat{R}_1, \widehat{R}_2, \widehat{R}_3] = [4, 150, 4, 5, 5, 5]$ is selected once on a representative setting for favorable accuracy and runtime, and is then shared across all datasets and noise levels.

6 Conclusion

In this work, we proposed Tensor Hourglass Decomposition (THD) for hyperspectral image super-resolution. THD organizes spatial and spectral representations into two peripheral cycles and couples their spoke modes through a shared high-order waist core, enabling group-specific modeling and flexible cross-group interactions. By exploiting the multilinearity of THD, the spatial degradation and spectral response operators are absorbed into the corresponding physical-mode-carrying factors, yielding a structured coupled observation model for the LR-HSI and HR-MSI. The resulting model is optimized by a matrix-free PAM scheme with spoke-wise balancing. Experiments on simulated and real datasets demonstrate accurate spatial–spectral reconstruction, strong robustness to additive noise, and competitive computational efficiency.

AI USE STATEMENT

We acknowledge the assistance of OpenAI's GPT-5.6 for grammatical correction, language polishing, and presentation refinement. The research design, mathematical derivations, experimental results, and scientific conclusions were developed, verified, and finalized by the authors.

REPRODUCIBILITY STATEMENT

We provide the formulation and algebraic properties of THD in Section 3 and Appendix A, detailed optimization derivations and the matrix-free implementation in Appendix B, computational and storage complexity analyses in Appendix C, and dataset preparation, observation generation, baseline settings, evaluation protocols, ablations, robustness, convergence, and parameter-sensitivity analyses in Appendix D. All compared methods are evaluated using identical observations under a unified experimental protocol. These details are provided to facilitate independent reproduction of the reported results.

REFERENCES

Fabio Dell'Acqua, Paolo Gamba, Alessio Ferrari, Jon Aevar Palmason, Jón Atli Benediktsson, and Kolbeinn Árnason. Exploiting spectral and spatial information in hyperspectral urban data with high resolution. *IEEE Geoscience and Remote Sensing Letters*, 1(4):322–326, 2004.

Renwei Dian, Yuanye Liu, and Shutao Li. Hyperspectral image fusion via a novel generalized tensor nuclear norm regularization. *IEEE Transactions on Neural Networks and Learning Systems*, 36 (4):7437–7448, 2025. doi: 10.1109/TNNLS.2024.3385473.

Lianru Gao, Jiaxin Li, Ke Zheng, and Xiuping Jia. Enhanced autoencoders with attention-embedded degradation learning for unsupervised hyperspectral image super-resolution. *IEEE Transactions on Geoscience and Remote Sensing*, 61:1–17, 2023.

Chengxun He, Yang Xu, Zebin Wu, and Zhihui Wei. Connecting low-level and high-level visions: A joint optimization for hyperspectral image super-resolution and target detection. *IEEE Transactions on Geoscience and Remote Sensing*, 62:1–16, 2024. doi: 10.1109/TGRS.2024.3385448.

Danfeng Hong, Chenyu Li, Naoto Yokoya, Bing Zhang, Xiuping Jia, Antonio Plaza, Paolo Gamba, Jon Atli Benediktsson, and Jocelyn Chanussot. Hyperspectral imaging. *Nature Reviews Methods Primers*, 6(1):19, 2026.

Kewen Huang, Yiping Xu, Yichun Dong, Ke Li, Hu Wang, and Yan Tian. Dbnsr: Hyperspectral image super-resolution based on dual-branch network. *IEEE Transactions on Geoscience and Remote Sensing*, 64:1–18, 2026. doi: 10.1109/TGRS.2026.3678075.

Jiaxin Li, Ke Zheng, Lianru Gao, Zhu Han, Zhi Li, and Jocelyn Chanussot. Enhanced deep image prior for unsupervised hyperspectral image super-resolution. *IEEE Transactions on Geoscience and Remote Sensing*, 63:1–18, 2025.

Shutao Li, Renwei Dian, Leyuan Fang, and José M Bioucas-Dias. Fusing hyperspectral and multispectral images via coupled sparse tensor factorization. *IEEE Transactions on Image Processing*, 27(8):4118–4130, 2018.

Pan Liu, Yuanyang Bu, Jingxiang Yang, and Yongqiang Zhao. Dynamic deep subspace residual regularizer for hyperspectral image super-resolution. *Information Fusion*, pp. 104632, 2026.

Guohua Miao, Zhihua Xie, and Fan Yang. Wavelet-enhanced spatial–spectral prior injection mamba network for hyperspectral image super-resolution. *IEEE Transactions on Geoscience and Remote Sensing*, 64:1–17, 2026. doi: 10.1109/TGRS.2025.3650662.

Ruoyang Su, Xi-Le Zhao, Wei-Hao Wu, Sheng Liu, and Junhua He. Deep fully-connected tensor network decomposition for multi-dimensional signal recovery. *Signal Processing*, 233:109903, 2025. ISSN 0165-1684. doi: https://doi.org/10.1016/j.sigpro.2025.109903. URL https://www.sciencedirect.com/science/article/pii/S0165168425000180.

Zhihui Tu, Yisha Gao, Wenyu Hu, Yong Chen, Honghui Xu, and Jian Lu. Hierarchical progressive transform with learnable prior for hyperspectral image super-resolution. *IEEE Transactions on Geoscience and Remote Sensing*, 64:5520419–5520419, 2026. doi: 10.1109/TGRS.2026.3703738.

Nan Wang, Aiping Mo, Renwei Dian, Jujue Wei, and Shutao Li. An unsupervised deep unfolded network for mosaic-based hyperspectral fusion imaging. *IEEE Transactions on Geoscience and Remote Sensing*, pp. 1–1, 2026. doi: 10.1109/TGRS.2026.3714859.

Maxwell Z Wilson. Environmental monitoring from the air with hyperspectral reporters: Environmental biotechnology. *Nature Biotechnology*, 44(2):194–195, 2026.

Zhong-Cheng Wu, Liang-Jian Deng, Ting-Zhu Huang, Hong-Xia Dou, Gemine Vivone, and Yu Liu. Tensor wheel decomposition: Theory and application to tensor completion. *IEEE Transactions on Image Processing*, 2026.

Honghui Xu, Mengjie Qin, Shengyong Chen, Yuhui Zheng, and Jianwei Zheng. Hyperspectral-multispectral image fusion via tensor ring and subspace decompositions. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 14:8823–8837, 2021.

Honghui Xu, Chuangjie Fang, Yilin Ge, Yubin Gu, and Jianwei Zheng. Cascade-transform-based tensor nuclear norm for hyperspectral image super-resolution. *IEEE Transactions on Geoscience and Remote Sensing*, 62(5540616):1–16, 2024a. doi: 10.1109/TGRS.2024.3508456.

Honghui Xu, Chuangjie Fang, Yibin Wang, Jie Wu, and Jianwei Zheng. Laboring on less labors: Rpca paradigm for pan-sharpening. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 11393–11402, 2025a.

Honghui Xu, Yueqian Quan, Mengjie Qin, Yibin Wang, Chuangjie Fang, Yan Li, and Jianwei Zheng. Nonlinear learnable triple-domain transform tensor nuclear norm for hyperspectral image super-resolution. *IEEE Transactions on Geoscience and Remote Sensing*, 63:1–17, 2025b. doi: 10.1109/TGRS.2025.3568627.

Honghui Xu, Junwei Zhu, Yubin Gu, Yueqian Quan, Chuangjie Fang, Hong Qiu, Jianwei Zheng, et al. Trt: Harnessing tensor ring transformer for hyperspectral image super-resolution. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, pp. 11232–11240, 2026a.

Junnan Xu, Honghui Xu, Wenbin Lu, Junwei Zhu, Junjie Zheng, Jiawei Jiang, and Jianwei Zheng. Semantic-continuous mamba for space-spectral fusion. *Information Fusion*, pp. 104616, 2026b.

Ting Xu, Ting-Zhu Huang, Liang-Jian Deng, Jin-Liang Xiao, Clifford Broni-Bediako, Junshi Xia, and Naoto Yokoya. A coupled tensor double-factor method for hyperspectral and multispectral image fusion. *IEEE Transactions on Geoscience and Remote Sensing*, 62:1–17, 2024b. doi: 10.1109/TGRS.2024.3389016.

Yinghao Xu, Xuanwen Tao, Zihao Ni, Zonghua Liu, Yimin Fu, and Peng Ren. Material proportion guided hyperspectral image super-resolution via unmixing. *ISPRS Journal of Photogrammetry and Remote Sensing*, 235:584–598, 2026c.

Yong Yang, Aoqi Zhao, Shuying Huang, Weiguo Wan, and Xiaozheng Wang. Ehen: Eigendecomposition-based hyperchannel enhancement network for hyperspectral image super-resolution. *Pattern Recognition*, pp. 114111, 2026.

Yu-Bang Zheng, Ting-Zhu Huang, Xi-Le Zhao, Qibin Zhao, and Tai-Xiang Jiang. Fully-connected tensor network decomposition and its application to higher-order tensor completion. In *Proceedings of the AAAI conference on artificial intelligence*, volume 35, pp. 11071–11078, 2021.

Yu-Bang Zheng, Ting-Zhu Huang, Xi-Le Zhao, and Qibin Zhao. Tensor completion via fully-connected tensor network decomposition with regularized factors. *Journal of Scientific Computing*, 92(1):8, 2022.

APPENDIX

Appendix A Proofs of the Algebraic Properties of THD p. 13

- Graph representation and topological properties
- Grouped- and mode-wise unfolding properties
- Physical-mode operator absorption
- Exact representability
- General cut-set rank bounds
- Waist-core coupling rank
- Spoke-wise gauge invariance
- Tucker-type degeneration

Appendix B Detailed Optimization Derivations p. 20

- Unified linear-operator formulation
- Peripheral-factor and waist-core updates
- Initialization and spoke-wise balancing
- Strong convexity and unique solvability
- Matrix-free conjugate-gradient implementation
- Sufficient decrease and stopping criterion

Appendix C Computational and Storage Complexity p. 28

- Peripheral-factor update costs
- Per-iteration computational complexity
- Waist-core update cost
- Parameter count and working memory

Appendix D Additional Experimental Details p. 29

- Dataset preparation and observation generation
- Peripheral-realization ablation
- Noise robustness and computational efficiency
- Compared methods and evaluation metrics
- Complete quantitative and visual comparisons
- Convergence behavior and parameter sensitivity

A Proofs of the Algebraic Properties of THD

Throughout this appendix, all tensor unfoldings, vectorizations, and reshaping operations follow the index ordering adopted in the main text. For a positive integer m, let $[m] = \{1, \dots, m\}$.

A.1 Graph Representation and Topological Properties

The algebraic structure of THD can be equivalently described by a labeled tensor-network multigraph. This representation is useful for interpreting the information paths induced by the two peripheral cycles and the shared waist core.

Definition 3 (THD contraction graph) *For a fixed realization (M_L, M_R), the THD contraction graph is the labeled multigraph*

$$\mathscr{G}_{\mathrm{THD}} = (\mathscr{V}, \mathscr{E}), \tag{A.1}$$

with vertex set

$$\mathscr{V} = \{\mathcal{C}\} \cup \{\mathcal{G}_k\}_{k=1}^{M_L} \cup \{\mathcal{F}_j\}_{j=1}^{M_R}. \tag{A.2}$$

The edge set consists of the left cyclic bonds, the right cyclic bonds, and the spoke bonds:

$$\mathscr{E}_L^{\mathrm{cyc}} = \left\{(\mathcal{G}_k, \mathcal{G}_{k+1}; R_{k+1})\right\}_{k=1}^{M_L}, \tag{A.3}$$

$$\mathscr{E}_R^{\mathrm{cyc}} = \left\{(\mathcal{F}_j, \mathcal{F}_{j+1}; \widehat{R}_{j+1})\right\}_{j=1}^{M_R}, \tag{A.4}$$

$$\mathscr{E}^{\mathrm{spoke}} = \left\{(\mathcal{G}_k, \mathcal{C}; L_k)\right\}_{k=1}^{M_L} \cup \left\{(\mathcal{F}_j, \mathcal{C}; U_j)\right\}_{j=1}^{M_R}, \tag{A.5}$$

where $\mathcal{G}_{M_L+1} = \mathcal{G}_1$ and $\mathcal{F}_{M_R+1} = \mathcal{F}_1$. The value after the semicolon denotes the dimension of the corresponding contracted index. Self-loops and parallel edges are retained when a peripheral cycle has length one or two.

Proposition 3 (Linear connectivity and graph diameter) *The THD contraction graph contains*

$$|\mathscr{V}| = M_L + M_R + 1 \tag{A.6}$$

vertices and

$$|\mathscr{E}| = 2(M_L + M_R) \tag{A.7}$$

latent bonds when edges are counted with multiplicity. Moreover, after self-loops and parallel edges are ignored for the purpose of shortest-path distance,

$$\operatorname{diam}(\mathscr{G}_{\mathrm{THD}}) = 2. \tag{A.8}$$

Proof 1 *There are M_L left peripheral vertices, M_R right peripheral vertices, and one waist-core vertex. Each peripheral cycle contributes one cyclic bond per peripheral factor, giving $M_L + M_R$ cyclic bonds. Every peripheral factor is additionally connected to the waist core through one spoke bond, giving another $M_L + M_R$ bonds.*

Every peripheral vertex is adjacent to $\mathcal{C}$. Consequently, any two vertices are connected by a path of length at most two. A left peripheral vertex and a right peripheral vertex are not directly connected, so their shortest path has length two through $\mathcal{C}$. Hence the diameter equals two.

The proposition shows that THD retains a number of latent bonds that grows linearly with the peripheral-cycle lengths, while any pair of peripheral factors can communicate through the shared core in at most two graph hops. The graph result is purely topological; the actual representation capacity is controlled by the associated bond dimensions.

A.2 A General Cut-Set Rank Bound

The rank bounds in the main text can be interpreted as special cases of a general cut-set property of the THD network. Let $\mathcal{S} \subseteq [N]$ be a subset of physical modes and let $\boldsymbol{X}_{\langle \mathcal{S} | \mathcal{S}^c \rangle}$ denote the unfolding whose row modes are indexed by $\mathcal{S}$. A vertex bipartition $\mathscr{P} = (\mathscr{V}_{\mathcal{S}}, \mathscr{V}_{\mathcal{S}^c})$ is called compatible with $\mathcal{S}$ if every peripheral factor carrying a mode in $\mathcal{S}$ belongs to $\mathscr{V}_{\mathcal{S}}$ and every factor carrying a mode in $\mathcal{S}^c$ belongs to $\mathscr{V}_{\mathcal{S}^c}$. Virtual singleton factors and the waist core may be assigned to either side. Let $\delta(\mathscr{P})$ be the set of latent bonds crossing the partition and let d_e be the dimension of a cut bond e.

Theorem 2 (THD cut-set rank bound) *For every compatible vertex bipartition $\mathscr{P}$,*

$$\operatorname{rank}\left(\boldsymbol{X}_{\langle \mathcal{S}|\mathcal{S}^c\rangle}\right) \leq \prod_{e\in\delta(\mathscr{P})} d_e. \tag{A.9}$$

Consequently,

$$\operatorname{rank}\left(\boldsymbol{X}_{\langle \mathcal{S}|\mathcal{S}^c\rangle}\right) \leq \min_{\mathscr{P} \text{ compatible}} \prod_{e\in\delta(\mathscr{P})} d_e. \tag{A.10}$$

Proof 2 *Fix a compatible partition and collect all cut indices into a single aggregate index γ of dimension*

$$D_{\mathscr{P}} = \prod_{e\in\delta(\mathscr{P})} d_e. \tag{A.11}$$

Contract all factors on the $\mathcal{V}_{\mathcal{S}}$ side while leaving the physical indices in $\mathcal{S}$ and the aggregate cut index γ free. This produces a matrix $\boldsymbol{A}_{\mathscr{P}}$ with $D_{\mathscr{P}}$ columns. Contracting the other side analogously produces $\boldsymbol{B}_{\mathscr{P}}$ with the same number of columns. Up to a fixed row or column permutation induced by the unfolding convention,

$$\boldsymbol{X}_{\langle \mathcal{S}|\mathcal{S}^c\rangle} = \boldsymbol{A}_{\mathscr{P}}\boldsymbol{B}_{\mathscr{P}}^{\top}. \tag{A.12}$$

Therefore its matrix rank is at most $D_{\mathscr{P}}$. Minimizing over all compatible partitions gives equation A.10.

Corollary 1 (Grouped and single-mode cuts) *For the grouped unfolding associated with $(\boldsymbol{\pi}_L, \boldsymbol{\pi}_R)$,*

$$\operatorname{rank}\left(\boldsymbol{X}_{\langle L|R\rangle}\right) \leq \min\{L, U\}. \tag{A.13}$$

For a nondegenerate left physical factor $\mathcal{G}_k$ and a nondegenerate right physical factor $\mathcal{F}_j$,

$$\operatorname{rank}\left(\boldsymbol{X}_{(\pi_k)}\right) \leq R_k L_k R_{k+1}, \tag{A.14}$$

$$\operatorname{rank}\left(\boldsymbol{X}_{(\rho_j)}\right) \leq \widehat{R}_j U_j \widehat{R}_{j+1}. \tag{A.15}$$

Proof 3 *For the grouped unfolding, placing the waist core on the right side cuts all left spoke bonds and gives the bound L; placing it on the left side cuts all right spoke bonds and gives the bound U. For a single physical mode, isolate its peripheral factor from the remaining graph. The three cut bonds are its two cyclic bonds and its spoke bond. The length-one-cycle case is covered directly by the factor separation proof below, because a self-contracted cyclic bond does not define an ordinary graph cut.*

A.3 Proof of Theorem 1

Proof 4 *For notational convenience, define the grouped physical index tuples*

$$\boldsymbol{i}_L = (i_{\pi_1}, \ldots, i_{\pi_K}), \qquad \boldsymbol{i}_R = (i_{\rho_1}, \ldots, i_{\rho_J}), \tag{A.16}$$

and the grouped spoke-index tuples

$$\boldsymbol{l} = (l_1, \ldots, l_{M_L}), \qquad \boldsymbol{u} = (u_1, \ldots, u_{M_R}). \tag{A.17}$$

Substituting the definitions of the left and right peripheral contractions in equation 12 into the THD representation in equation 8 gives

$$\mathcal{X}(i_1, \ldots, i_N) = \sum_{\boldsymbol{l}} \sum_{\boldsymbol{u}} \mathcal{G}^{(L)}(\boldsymbol{i}_L, \boldsymbol{l})\, \mathcal{C}(\boldsymbol{l}, \boldsymbol{u})\, \mathcal{F}^{(R)}(\boldsymbol{i}_R, \boldsymbol{u}). \tag{A.18}$$

Let

$$\alpha = \operatorname{ind}_L(\boldsymbol{i}_L) \in [I_L], \qquad \beta = \operatorname{ind}_R(\boldsymbol{i}_R) \in [I_R] \tag{A.19}$$

denote the linear indices induced by the grouped physical modes, and let

$$\xi = \operatorname{ind}_{\mathcal{L}}(\boldsymbol{l}) \in [L], \qquad \eta = \operatorname{ind}_{\mathcal{U}}(\boldsymbol{u}) \in [U] \tag{A.20}$$

denote the linear indices induced by the spoke modes. Here,

$$I_L = \prod_{k=1}^{K} I_{\pi_k}, \qquad I_R = \prod_{j=1}^{J} I_{\rho_j}, \tag{A.21}$$

and

$$L = \prod_{k=1}^{M_L} L_k, \qquad U = \prod_{j=1}^{M_R} U_j. \tag{A.22}$$

According to the reshape operations in equation 13, the corresponding matrix entries satisfy

$$\begin{aligned} \boldsymbol{G}(\alpha, \xi) &= \mathcal{G}^{(L)}\left(\boldsymbol{i}_L, \boldsymbol{l}\right), \\ \boldsymbol{C}(\xi, \eta) &= \mathcal{C}\left(\boldsymbol{l}, \boldsymbol{u}\right), \\ \boldsymbol{F}(\beta, \eta) &= \mathcal{F}^{(R)}\left(\boldsymbol{i}_R, \boldsymbol{u}\right). \end{aligned} \tag{A.23}$$

Meanwhile, the grouped unfolding in equation 11 gives

$$\boldsymbol{X}_{\langle L|R\rangle}(\alpha, \beta) = \mathcal{X}(i_1, \ldots, i_N). \tag{A.24}$$

Therefore, equation A.18 can be written entrywise as

$$\begin{aligned} \boldsymbol{X}_{\langle L|R\rangle}(\alpha, \beta) &= \sum_{\xi=1}^{L} \sum_{\eta=1}^{U} \boldsymbol{G}(\alpha, \xi) \boldsymbol{C}(\xi, \eta) \boldsymbol{F}(\beta, \eta) \\ &= \left(\boldsymbol{G}\boldsymbol{C}\boldsymbol{F}^{\top}\right)(\alpha, \beta). \end{aligned} \tag{A.25}$$

Since this equality holds for every $\alpha \in [I_L]$ and $\beta \in [I_R]$, we obtain

$$\boldsymbol{X}_{\langle L|R\rangle} = \boldsymbol{G}\boldsymbol{C}\boldsymbol{F}^{\top}, \tag{A.26}$$

which proves equation 14.

A.4 Proof of Proposition 1

Proof 5 *We first consider a left peripheral factor $\mathcal{G}_k$ and fix all the remaining THD factors. Define the latent index tuple associated with $\mathcal{G}_k$ as*

$$\boldsymbol{a}_k^G = \left(r_k, l_k, r_{k+1}\right), \tag{A.27}$$

whose aggregate dimension is

$$D_k^G = R_k L_k R_{k+1}. \tag{A.28}$$

Let

$$a_k^G = \mathrm{ind}_{G_k}\left(\boldsymbol{a}_k^G\right) \in [D_k^G] \tag{A.29}$$

denote the corresponding linear latent index.

By the adopted mode-2 unfolding convention,

$$\boldsymbol{G}_{k(2)} \in \mathbb{R}^{\widetilde{I}_k^L \times D_k^G} \tag{A.30}$$

and its entries satisfy

$$\boldsymbol{G}_{k(2)}\left(\widetilde{i}_k^L, a_k^G\right) = \mathcal{G}_k\left(r_k, \widetilde{i}_k^L, l_k, r_{k+1}\right), \tag{A.31}$$

up to the fixed ordering used to combine (r_k, l_k, r_{k+1}). The derivation below is written for cycles of length greater than one. For a length-one cycle, the cyclic equality constraint is absorbed into the environment contraction; the same factor-wise linear form follows, with inactive off-diagonal latent pairs contributing zero.

We now collect all contractions not belonging to $\mathcal{G}_k$ into an environment tensor. For fixed $\boldsymbol{a}_k^G$ and fixed original physical indices other than the physical index carried by $\mathcal{G}_k$, define

$$\begin{aligned}
&\mathcal{E}_{\neq \mathcal{G}_k}\left(\boldsymbol{a}_k^G; \boldsymbol{i}_{\neq k}^G\right) \\
&= \sum_{\substack{\{r_t\}_{t\notin\{k,k+1\}} \\ \{q_j\}_{j=1}^{M_R} \\ \{l_t\}_{t\neq k} \\ \{u_j\}_{j=1}^{M_R}}} \mathcal{C}\left(l_1, \ldots, l_{M_L}, u_1, \ldots, u_{M_R}\right) \\
&\quad \times \prod_{\substack{t=1 \\ t\neq k}}^{M_L} \mathcal{G}_t\left(r_t, \widetilde{i}_t^L, l_t, r_{t+1}\right) \prod_{j=1}^{M_R} \mathcal{F}_j\left(q_j, \widetilde{i}_j^R, u_j, q_{j+1}\right),
\end{aligned} \tag{A.32}$$

where r_k, r_{k+1}, and l_k remain free and are contained in $\boldsymbol{a}_k^G$. The tuple $\boldsymbol{i}_{\neq k}^G$ contains all original physical indices except the one carried by $\mathcal{G}_k$ when $k \leq K$. For $k > K$, it contains all original physical indices because the corresponding position is virtual.

With the environment tensor in equation A.32, the THD contraction can be rewritten as

$$\begin{aligned}
\mathcal{X}(i_1, \ldots, i_N) = \sum_{r_k=1}^{R_k} \sum_{l_k=1}^{L_k} \sum_{r_{k+1}=1}^{R_{k+1}} &\mathcal{G}_k\left(r_k, \widetilde{i}_k^L, l_k, r_{k+1}\right) \\
&\times \mathcal{E}_{\neq \mathcal{G}_k}\left(r_k, l_k, r_{k+1}; \boldsymbol{i}_{\neq k}^G\right).
\end{aligned} \tag{A.33}$$

Reshaping $\mathcal{E}_{\neq \mathcal{G}_k}$ consistently gives the environment matrix

$$\boldsymbol{M}_{\neq \mathcal{G}_k} \in \mathbb{R}^{D_k^G \times \left(\left(\prod_{n=1}^N I_n\right)/\widetilde{I}_k^L\right)}. \tag{A.34}$$

For $k = 1, \ldots, K$, the factor carries the original physical mode π_k, so

$$\widetilde{I}_k^L = I_{\pi_k}, \qquad \widetilde{i}_k^L = i_{\pi_k}. \tag{A.35}$$

Unfolding equation A.33 with i_{π_k} as the row index therefore yields

$$\boldsymbol{X}_{(\pi_k)} = \boldsymbol{G}_{k(2)} \boldsymbol{M}_{\neq \mathcal{G}_k}, \qquad k = 1, \ldots, K, \tag{A.36}$$

up to a fixed column permutation determined by the unfolding convention.

For $k = K+1, \ldots, M_L$, the corresponding position is virtual, and hence

$$\widetilde{I}_k^L = 1, \qquad \widetilde{i}_k^L = 1. \tag{A.37}$$

In this case, $\boldsymbol{G}_{k(2)} \in \mathbb{R}^{1 \times D_k^G}$, while all original physical indices remain on the environment side. Consequently, equation A.33 becomes

$$\mathrm{vec}(\mathcal{X})^\top = \boldsymbol{G}_{k(2)} \boldsymbol{M}_{\neq \mathcal{G}_k}, \qquad k = K+1, \ldots, M_L, \tag{A.38}$$

up to a fixed column permutation.

We next consider a right peripheral factor $\mathcal{F}_j$. Define

$$\boldsymbol{a}_j^F = (q_j, u_j, q_{j+1}) \tag{A.39}$$

with aggregate dimension

$$D_j^F = \widehat{R}_j U_j \widehat{R}_{j+1}. \tag{A.40}$$

The mode-2 unfolding satisfies

$$\boldsymbol{F}_{j(2)} \in \mathbb{R}^{\widetilde{I}_j^R \times D_j^F}. \tag{A.41}$$

Collecting all contractions except $\mathcal{F}_j$ into an environment tensor and reshaping it consistently gives

$$\boldsymbol{M}_{\neq \mathcal{F}_j} \in \mathbb{R}^{D_j^F \times \left(\left(\prod_{n=1}^N I_n\right)/\widetilde{I}_j^R\right)}. \tag{A.42}$$

Therefore, for an original physical-mode factor $j = 1, \ldots, J$,

$$\boldsymbol{X}_{(\rho_j)} = \boldsymbol{F}_{j(2)} \boldsymbol{M}_{\neq \mathcal{F}_j}, \tag{A.43}$$

whereas for a virtual factor $j = J+1, \ldots, M_R$,

$$\mathrm{vec}(\mathcal{X})^\top = \boldsymbol{F}_{j(2)} \boldsymbol{M}_{\neq \mathcal{F}_j}. \tag{A.44}$$

All fixed column permutations induced by the adopted unfolding or vectorization convention can be absorbed into the corresponding environment matrices. This proves equation 15–equation 16.

A.5 Proof of Proposition 2

Proof 6 *By Theorem 1, the grouped unfolding of $\mathcal{X}$ satisfies*

$$\boldsymbol{X}_{\langle L|R\rangle} = \boldsymbol{G}\boldsymbol{C}\boldsymbol{F}^{\top}. \tag{A.45}$$

Using the standard rank inequality for a product of matrices gives

$$\begin{aligned}\operatorname{rank}\left(\boldsymbol{X}_{\langle L|R\rangle}\right) \leq \min\big\{ &\operatorname{rank}(\boldsymbol{G}), \operatorname{rank}(\boldsymbol{C}), \\ &\operatorname{rank}(\boldsymbol{F}^{\top})\big\}.\end{aligned} \tag{A.46}$$

Since

$$\operatorname{rank}(\boldsymbol{F}^{\top}) = \operatorname{rank}(\boldsymbol{F}), \tag{A.47}$$

we obtain

$$\operatorname{rank}\left(\boldsymbol{X}_{\langle L|R\rangle}\right) \leq \min\left\{\operatorname{rank}(\boldsymbol{G}), \operatorname{rank}(\boldsymbol{C}), \operatorname{rank}(\boldsymbol{F})\right\}. \tag{A.48}$$

From equation 13,

$$\boldsymbol{G} \in \mathbb{R}^{I_L \times L}, \qquad \boldsymbol{C} \in \mathbb{R}^{L \times U}, \qquad \boldsymbol{F} \in \mathbb{R}^{I_R \times U}. \tag{A.49}$$

Therefore,

$$\begin{aligned}\operatorname{rank}(\boldsymbol{G}) &\leq \min\{I_L, L\}, \\ \operatorname{rank}(\boldsymbol{C}) &\leq \min\{L, U\}, \\ \operatorname{rank}(\boldsymbol{F}) &\leq \min\{I_R, U\}.\end{aligned} \tag{A.50}$$

Combining these inequalities with equation A.48 yields

$$\operatorname{rank}\left(\boldsymbol{X}_{\langle L|R\rangle}\right) \leq \min\left\{I_L, I_R, L, U\right\}, \tag{A.51}$$

which proves equation 17.

We next consider an original physical mode π_k in the left mode group. By Proposition 1, there exists a fixed permutation matrix $\boldsymbol{P}_k^G$ such that

$$\boldsymbol{X}_{(\pi_k)} = \boldsymbol{G}_{k(2)}\boldsymbol{M}_{\neq \mathcal{G}_k}\boldsymbol{P}_k^G. \tag{A.52}$$

The permutation matrix does not change matrix rank. Therefore,

$$\begin{aligned}\operatorname{rank}\left(\boldsymbol{X}_{(\pi_k)}\right) &\leq \operatorname{rank}\left(\boldsymbol{G}_{k(2)}\right) \\ &\leq \min\left\{I_{\pi_k}, R_k L_k R_{k+1}\right\},\end{aligned} \tag{A.53}$$

because

$$\boldsymbol{G}_{k(2)} \in \mathbb{R}^{I_{\pi_k} \times R_k L_k R_{k+1}}. \tag{A.54}$$

This proves the first mode-wise rank bound in Proposition 2.

Similarly, for an original physical mode ρ_j in the right mode group, there exists a fixed permutation matrix $\boldsymbol{P}_j^F$ such that

$$\boldsymbol{X}_{(\rho_j)} = \boldsymbol{F}_{j(2)}\boldsymbol{M}_{\neq \mathcal{F}_j}\boldsymbol{P}_j^F. \tag{A.55}$$

Since

$$\boldsymbol{F}_{j(2)} \in \mathbb{R}^{I_{\rho_j} \times \widehat{R}_j U_j \widehat{R}_{j+1}}, \tag{A.56}$$

it follows that

$$\operatorname{rank}\left(\boldsymbol{X}_{(\rho_j)}\right) \leq \min\left\{I_{\rho_j}, \widehat{R}_j U_j \widehat{R}_{j+1}\right\}. \tag{A.57}$$

This proves the second mode-wise rank bound and completes the proof.

A.6 WAIST-CORE COUPLING RANK

The grouped matrix form also provides an explicit algebraic interpretation of the shared waist core. Because the factors of a tensor-network decomposition are generally nonunique, the following quantity is defined for a fixed THD representation rather than as an intrinsic tensor rank.

Definition 4 (Waist-core coupling rank) *For a fixed THD representation with*

$$\boldsymbol{C} = \mathrm{reshape}(\mathcal{C}, L, U), \tag{A.58}$$

the waist-core coupling rank is

$$r_{\mathrm{w}} = \mathrm{rank}(\boldsymbol{C}). \tag{A.59}$$

Proposition 4 (Separable cross-group expansion) *Let r_{w} be defined by equation A.59. Then the grouped unfolding admits the expansion*

$$\boldsymbol{X}_{\langle L|R\rangle} = \sum_{s=1}^{r_{\mathrm{w}}} \sigma_s (\boldsymbol{G}\boldsymbol{a}_s)(\boldsymbol{F}\boldsymbol{b}_s)^\top, \tag{A.60}$$

where

$$\boldsymbol{C} = \sum_{s=1}^{r_{\mathrm{w}}} \sigma_s \boldsymbol{a}_s \boldsymbol{b}_s^\top \tag{A.61}$$

is a compact singular-value decomposition. If $\boldsymbol{G}$ and $\boldsymbol{F}$ both have full column rank, then

$$\mathrm{rank}\left(\boldsymbol{X}_{\langle L|R\rangle}\right) = r_{\mathrm{w}}. \tag{A.62}$$

Proof 7 *Substituting the compact singular-value decomposition of $\boldsymbol{C}$ into $\boldsymbol{X}_{\langle L|R\rangle} = \boldsymbol{G}\boldsymbol{C}\boldsymbol{F}^\top$ gives equation A.60. Hence the waist core combines the left and right group representations through r_{w} separable coupling channels.*

The product rank inequality gives $\mathrm{rank}(\boldsymbol{X}_{\langle L|R\rangle}) \le r_{\mathrm{w}}$. If $\boldsymbol{G}$ and $\boldsymbol{F}$ have full column rank, they admit left inverses $\boldsymbol{G}^\dagger$ and $\boldsymbol{F}^\dagger$. Therefore,

$$\boldsymbol{C} = \boldsymbol{G}^\dagger \boldsymbol{X}_{\langle L|R\rangle} (\boldsymbol{F}^\dagger)^\top, \tag{A.63}$$

which implies $r_{\mathrm{w}} \le \mathrm{rank}(\boldsymbol{X}_{\langle L|R\rangle})$. Combining the two inequalities proves equation A.62.

A rank-one waist core permits only one separable left–right interaction channel, whereas a higher-rank core supports multiple coordinated channels. This explains why a full waist-core parameterization can be substantially more expressive than rank-one or strongly constrained alternatives, while a low-rank core provides an intermediate capacity.

A.7 ABSORPTION OF PHYSICAL-MODE LINEAR OPERATORS

The physical degradation model used for HSI–MSI fusion is a special case of the following multilinearity property.

Proposition 5 (Physical-mode operator absorption) *Suppose that $\mathcal{X}$ admits the THD representation in equation 8. For a left physical mode $n = \pi_k$ and a matrix $\boldsymbol{P} \in \mathbb{R}^{J_n \times I_n}$,*

$$\begin{aligned}\mathcal{X} \times_n \boldsymbol{P} = \mathrm{THD}^{(M_L,M_R)}_{\boldsymbol{\pi}_L|\boldsymbol{\pi}_R}\big(&\mathcal{G}_1, \ldots, \mathcal{G}_k \times_2 \boldsymbol{P}, \\ &\ldots, \mathcal{G}_{M_L}; \mathcal{C}; \{\mathcal{F}_j\}_{j=1}^{M_R}\big).\end{aligned} \tag{A.64}$$

For a right physical mode $n = \rho_j$,

$$\begin{aligned}\mathcal{X} \times_n \boldsymbol{P} = \mathrm{THD}^{(M_L,M_R)}_{\boldsymbol{\pi}_L|\boldsymbol{\pi}_R}\big(&\{\mathcal{G}_k\}_{k=1}^{M_L}; \mathcal{C}; \\ &\mathcal{F}_1, \ldots, \mathcal{F}_j \times_2 \boldsymbol{P}, \ldots, \mathcal{F}_{M_R}\big).\end{aligned} \tag{A.65}$$

Operators acting on distinct physical modes can be absorbed simultaneously into their corresponding peripheral factors.

Proof 8 *Consider $n = \pi_k$. For a new mode index $a \in [J_n]$,*

$$(\mathcal{X} \times_n \boldsymbol{P})(i_1, \ldots, a, \ldots, i_N) = \sum_{i_n=1}^{I_n} \boldsymbol{P}(a, i_n)\mathcal{X}(i_1, \ldots, i_n, \ldots, i_N). \tag{A.66}$$

Substituting the THD element-wise contraction and exchanging finite summations, the only term depending on i_n is $\mathcal{G}_k(r_k, i_n, l_k, r_{k+1})$. Its contraction with $\boldsymbol{P}$ is

$$\sum_{i_n=1}^{I_n} \boldsymbol{P}(a, i_n)\mathcal{G}_k(r_k, i_n, l_k, r_{k+1}) = (\mathcal{G}_k \times_2 \boldsymbol{P})(r_k, a, l_k, r_{k+1}). \tag{A.67}$$

All other factors are unchanged, proving equation A.64. The right-mode case is identical. Because mode products on distinct physical modes act on distinct peripheral factors, the argument can be applied repeatedly.

Corollary 2 (Fusion degradation identities) *For the $(3, 3)$ THD realization used in the main text,*

$$\mathcal{X} \times_1 \boldsymbol{P}_1 \times_2 \boldsymbol{P}_2 = \mathrm{THD}^{(3,3)}_{(1,2)|(3)}\big(\mathcal{G}_1 \times_2 \boldsymbol{P}_1, \mathcal{G}_2 \times_2 \boldsymbol{P}_2, \mathcal{G}_3; \mathcal{C}; \mathcal{F}_1, \mathcal{F}_2, \mathcal{F}_3\big), \tag{A.68}$$

$$\mathcal{X} \times_3 \boldsymbol{P}_3 = \mathrm{THD}^{(3,3)}_{(1,2)|(3)}\big(\mathcal{G}_1, \mathcal{G}_2, \mathcal{G}_3; \mathcal{C}; \mathcal{F}_1 \times_2 \boldsymbol{P}_3, \mathcal{F}_2, \mathcal{F}_3\big). \tag{A.69}$$

A.8 Spoke-Wise Gauge Invariance

THD factors are not unique because an invertible scale can be transferred between a peripheral spoke and the corresponding waist-core mode. The balancing operation used in the algorithm is a gauge fixing of this invariance.

Proposition 6 (Diagonal spoke-gauge invariance) *For each left spoke, let $\boldsymbol{D}_k \in \mathbb{R}^{L_k \times L_k}$ be an invertible diagonal matrix, and for each right spoke let $\boldsymbol{E}_j \in \mathbb{R}^{U_j \times U_j}$ be an invertible diagonal matrix. Define*

$$\widetilde{\mathcal{G}}_k = \mathcal{G}_k \times_3 \boldsymbol{D}_k^{-1}, \qquad \widetilde{\mathcal{F}}_j = \mathcal{F}_j \times_3 \boldsymbol{E}_j^{-1}, \tag{A.70}$$

$$\widetilde{\mathcal{C}} = \mathcal{C} \times_1 \boldsymbol{D}_1 \cdots \times_{M_L} \boldsymbol{D}_{M_L} \times_{M_L+1} \boldsymbol{E}_1 \cdots \times_{M_L+M_R} \boldsymbol{E}_{M_R}. \tag{A.71}$$

Then

$$\mathrm{THD}^{(M_L,M_R)}_{\boldsymbol{\pi}_L|\boldsymbol{\pi}_R}\big(\{\widetilde{\mathcal{G}}_k\}_{k=1}^{M_L}; \widetilde{\mathcal{C}}; \{\widetilde{\mathcal{F}}_j\}_{j=1}^{M_R}\big) = \mathrm{THD}^{(M_L,M_R)}_{\boldsymbol{\pi}_L|\boldsymbol{\pi}_R}\big(\{\mathcal{G}_k\}_{k=1}^{M_L}; \mathcal{C}; \{\mathcal{F}_j\}_{j=1}^{M_R}\big). \tag{A.72}$$

Proof 9 *Because the gauge matrices are diagonal, a fixed left spoke index l_k contributes the factor $\boldsymbol{D}_k^{-1}(l_k, l_k)$ through $\widetilde{\mathcal{G}}_k$ and the reciprocal factor $\boldsymbol{D}_k(l_k, l_k)$ through $\widetilde{\mathcal{C}}$. Their product equals one. The same cancellation holds for every right spoke index and every $\boldsymbol{E}_j$. Thus each summand in the element-wise THD contraction is unchanged, proving equation A.72.*

The spoke-wise normalization in equation B.48 and equation B.51 is obtained by taking $\boldsymbol{D}_k$ and $\boldsymbol{E}_j$ to be the diagonal matrices of the measured spoke-slice norms.

A.9 Exact Representability and a Tucker-Type Special Case

Proposition 7 (Exact representability) *Every finite-dimensional tensor $\mathcal{X} \in \mathbb{R}^{I_1 \times \cdots \times I_N}$ admits an exact THD representation for any ordered two-group partition $(\boldsymbol{\pi}_L, \boldsymbol{\pi}_R)$.*

Proof 10 *Choose the minimal realization $M_L = K$ and $M_R = J$, set all cyclic dimensions to one, and set*

$$L_k = I_{\pi_k}, \qquad U_j = I_{\rho_j}. \tag{A.73}$$

Define the peripheral factors by Kronecker deltas:

$$\mathcal{G}_k(1, i, l, 1) = \delta_{i,l}, \qquad \mathcal{F}_j(1, i, u, 1) = \delta_{i,u}. \tag{A.74}$$

Define the waist core as the mode-permuted entries of the original tensor:

$$\mathcal{C}(l_1, \ldots, l_K, u_1, \ldots, u_J) = \mathcal{X}(i_1, \ldots, i_N), \tag{A.75}$$

where $i_{\pi_k} = l_k$ and $i_{\rho_j} = u_j$. In the THD contraction, the delta factors force every spoke index to equal its corresponding physical index. The remaining core entry is exactly $\mathcal{X}(i_1, \ldots, i_N)$.

Corollary 3 (Tucker-type degeneration) *In a minimal THD realization, if all cyclic dimensions equal one, THD reduces to a Tucker-type multilinear model under the mode ordering $(\boldsymbol{\pi}_L, \boldsymbol{\pi}_R)$.*

Proof 11 *With unit cyclic dimensions, define factor matrices*

$$\boldsymbol{A}_{\pi_k}(i, l) = \mathcal{G}_k(1, i, l, 1), \qquad \boldsymbol{A}_{\rho_j}(i, u) = \mathcal{F}_j(1, i, u, 1). \tag{A.76}$$

The THD element-wise expression becomes

$$\begin{aligned} \mathcal{X}(i_1, \ldots, i_N) = \sum_{l_1, \ldots, l_K} \sum_{u_1, \ldots, u_J} & \mathcal{C}(l_1, \ldots, l_K, u_1, \ldots, u_J) \\ & \times \prod_{k=1}^{K} \boldsymbol{A}_{\pi_k}(i_{\pi_k}, l_k) \prod_{j=1}^{J} \boldsymbol{A}_{\rho_j}(i_{\rho_j}, u_j), \end{aligned} \tag{A.77}$$

which is precisely a Tucker multilinear product after the specified mode permutation.

B Detailed Optimization Derivations

This appendix presents the detailed optimization procedure for the THD-based HSI–MSI fusion model in equation 26. Since the six peripheral factors and the shared waist core are multilinearly coupled, the overall problem is jointly nonconvex. Nevertheless, when the remaining variable blocks are fixed, the objective is quadratic with respect to each individual block.

We therefore adopt a proximal alternating minimization scheme and update the seven variable blocks in the Gauss–Seidel order

$$\mathcal{G}_1 \to \mathcal{G}_2 \to \mathcal{G}_3 \to \mathcal{F}_1 \to \mathcal{F}_2 \to \mathcal{F}_3 \to \mathcal{C}. \tag{B.1}$$

The same proximal parameter $\rho > 0$ is used for the six peripheral factors, whereas $\rho_C > 0$ is used for the shared waist core. The parameter $\lambda > 0$ weights the HR-MSI fidelity term, consistently with equation 26.

Unless otherwise specified, each environment matrix below is constructed using the most recently available Gauss–Seidel iterates of all THD components except the factor currently being updated. To simplify notation, iteration superscripts on the environment matrices are omitted.

For later use, the mode-wise observation unfoldings satisfy

$$\begin{aligned} \boldsymbol{Y}_{(1)} &\in \mathbb{R}^{h \times wS}, \quad \boldsymbol{Z}_{(1)} \in \mathbb{R}^{H \times Ws}, \\ \boldsymbol{Y}_{(2)} &\in \mathbb{R}^{w \times hS}, \quad \boldsymbol{Z}_{(2)} \in \mathbb{R}^{W \times Hs}, \\ \boldsymbol{Y}_{(3)} &\in \mathbb{R}^{S \times hw}, \quad \boldsymbol{Z}_{(3)} \in \mathbb{R}^{s \times HW}. \end{aligned} \tag{B.2}$$

For the virtual singleton factors, define

$$\boldsymbol{y}^\top = \mathrm{vec}(\mathcal{Y})^\top \in \mathbb{R}^{1 \times hwS}, \qquad \boldsymbol{z}^\top = \mathrm{vec}(\mathcal{Z})^\top \in \mathbb{R}^{1 \times HWs}. \tag{B.3}$$

The unfolding and vectorization orders are chosen consistently with Proposition 1. Any fixed column permutation induced by these conventions is absorbed into the corresponding environment matrix.

B.1 Unified Block Form and Unique Solvability

All seven updates share the same operator structure. Let $\mathcal{V}$ denote one peripheral factor or the waist core, and keep all other blocks fixed at their current Gauss–Seidel values. The two observation models then define linear maps

$$\mathscr{A}_Y^{\mathcal{V}} \quad \text{and} \quad \mathscr{A}_Z^{\mathcal{V}} \tag{B.4}$$

from the space of $\mathcal{V}$ to the corresponding observation spaces. The generic proximal block problem is

$$\begin{aligned}\mathcal{V}^+ = \arg\min_{\mathcal{V}} \quad & \frac{1}{2}\left\|\mathscr{A}_Y^{\mathcal{V}}(\mathcal{V}) - \mathcal{Y}\right\|_F^2 + \frac{\lambda}{2}\left\|\mathscr{A}_Z^{\mathcal{V}}(\mathcal{V}) - \mathcal{Z}\right\|_F^2 \\ & + \frac{\rho_{\mathcal{V}}}{2}\left\|\mathcal{V} - \mathcal{V}^{\text{old}}\right\|_F^2,\end{aligned} \tag{B.5}$$

where $\rho_{\mathcal{V}} = \rho$ for a peripheral factor and $\rho_{\mathcal{V}} = \rho_C$ for the waist core. The symbols $\mathcal{Y}$ and $\mathcal{Z}$ in equation B.5 also represent their consistent unfoldings when the block variable is matricized.

Proposition 8 (Strong convexity and unique block solution) *The normal operator associated with equation B.5 is*

$$\mathscr{H}_{\mathcal{V}} = (\mathscr{A}_Y^{\mathcal{V}})^* \mathscr{A}_Y^{\mathcal{V}} + \lambda(\mathscr{A}_Z^{\mathcal{V}})^* \mathscr{A}_Z^{\mathcal{V}} + \rho_{\mathcal{V}} \mathscr{I}. \tag{B.6}$$

It is self-adjoint and satisfies

$$\langle \mathcal{V}, \mathscr{H}_{\mathcal{V}}(\mathcal{V})\rangle_F \geq \rho_{\mathcal{V}} \|\mathcal{V}\|_F^2. \tag{B.7}$$

Therefore every proximal block subproblem is $\rho_{\mathcal{V}}$-strongly convex and has a unique minimizer. Its normal equation is suitable for conjugate gradient under the Frobenius inner product.

Proof 12 *For any block perturbation $\mathcal{V}$,*

$$\begin{aligned}\langle \mathcal{V}, \mathscr{H}_{\mathcal{V}}(\mathcal{V})\rangle_F = \ & \left\|\mathscr{A}_Y^{\mathcal{V}}(\mathcal{V})\right\|_F^2 + \lambda\left\|\mathscr{A}_Z^{\mathcal{V}}(\mathcal{V})\right\|_F^2 \\ & + \rho_{\mathcal{V}}\|\mathcal{V}\|_F^2 \geq \rho_{\mathcal{V}}\|\mathcal{V}\|_F^2.\end{aligned} \tag{B.8}$$

The adjoint-composition terms and the identity are self-adjoint. Since $\rho_{\mathcal{V}} > 0$, the operator is strictly positive definite. This is equivalent to strong convexity of the quadratic objective and implies existence and uniqueness of its minimizer.

The explicit updates below are obtained by instantiating $\mathscr{A}_Y^{\mathcal{V}}$ and $\mathscr{A}_Z^{\mathcal{V}}$ for each THD block.

B.2 Updates of the Spatial Peripheral Factors

B.2.1 Update of $\mathcal{G}_1$

The factor $\mathcal{G}_1$ carries the height mode. According to Proposition 1, the LR-HSI and HR-MSI reconstructions involving $\mathcal{G}_1$ can be written as

$$\boldsymbol{P}_1 \boldsymbol{G}_{1(2)} \boldsymbol{M}_{\neq \mathcal{G}_1}^Y \qquad \text{and} \qquad \boldsymbol{G}_{1(2)} \boldsymbol{M}_{\neq \mathcal{G}_1}^Z, \tag{B.9}$$

where

$$\boldsymbol{M}_{\neq \mathcal{G}_1}^Y \in \mathbb{R}^{R_1 L_1 R_2 \times wS}, \qquad \boldsymbol{M}_{\neq \mathcal{G}_1}^Z \in \mathbb{R}^{R_1 L_1 R_2 \times Ws}. \tag{B.10}$$

The LR-HSI environment contracts $(\mathcal{G}_2^Y)^{(t)}$, $\mathcal{G}_3^{(t)}$, $\mathcal{C}^{(t)}$, and $\{\mathcal{F}_j^{(t)}\}_{j=1}^3$. The HR-MSI environment contracts $\mathcal{G}_2^{(t)}$, $\mathcal{G}_3^{(t)}$, $\mathcal{C}^{(t)}$, $(\mathcal{F}_1^Z)^{(t)}$, $\mathcal{F}_2^{(t)}$, and $\mathcal{F}_3^{(t)}$.

The update of $\mathcal{G}_1$ is equivalently written in terms of its mode-2 unfolding as

$$\begin{aligned}\boldsymbol{G}_{1(2)}^{(t+1)} = \arg\min_{\boldsymbol{G}_{1(2)}} \quad & \frac{1}{2}\left\|\boldsymbol{P}_1 \boldsymbol{G}_{1(2)} \boldsymbol{M}_{\neq \mathcal{G}_1}^Y - \boldsymbol{Y}_{(1)}\right\|_F^2 \\ & + \frac{\lambda}{2}\left\|\boldsymbol{G}_{1(2)} \boldsymbol{M}_{\neq \mathcal{G}_1}^Z - \boldsymbol{Z}_{(1)}\right\|_F^2 \\ & + \frac{\rho}{2}\left\|\boldsymbol{G}_{1(2)} - \boldsymbol{G}_{1(2)}^{(t)}\right\|_F^2.\end{aligned} \tag{B.11}$$

Setting the derivative with respect to $\boldsymbol{G}_{1(2)}$ to zero gives

$$\begin{aligned}&\boldsymbol{P}_1^\top \boldsymbol{P}_1 \boldsymbol{G}_{1(2)}^{(t+1)} \boldsymbol{M}_{\neq \mathcal{G}_1}^Y \left(\boldsymbol{M}_{\neq \mathcal{G}_1}^Y\right)^\top \\ &\quad + \lambda \boldsymbol{G}_{1(2)}^{(t+1)} \boldsymbol{M}_{\neq \mathcal{G}_1}^Z \left(\boldsymbol{M}_{\neq \mathcal{G}_1}^Z\right)^\top + \rho \boldsymbol{G}_{1(2)}^{(t+1)} \\ &= \boldsymbol{P}_1^\top \boldsymbol{Y}_{(1)} \left(\boldsymbol{M}_{\neq \mathcal{G}_1}^Y\right)^\top + \lambda \boldsymbol{Z}_{(1)} \left(\boldsymbol{M}_{\neq \mathcal{G}_1}^Z\right)^\top + \rho \boldsymbol{G}_{1(2)}^{(t)}. \end{aligned} \tag{B.12}$$

After solving equation B.12, the tensor factor and its spatially degraded counterpart are recovered as

$$\mathcal{G}_1^{(t+1)} = \operatorname{fold}_2\left(\boldsymbol{G}_{1(2)}^{(t+1)}\right), \qquad (\mathcal{G}_1^Y)^{(t+1)} = \mathcal{G}_1^{(t+1)} \times_2 \boldsymbol{P}_1. \tag{B.13}$$

B.2.2 Update of $\mathcal{G}_2$

The factor $\mathcal{G}_2$ carries the width mode. Using the latest $\mathcal{G}_1^{(t+1)}$, the two factor-isolated observation forms are

$$\boldsymbol{P}_2 \boldsymbol{G}_{2(2)} \boldsymbol{M}_{\neq \mathcal{G}_2}^Y \qquad \text{and} \qquad \boldsymbol{G}_{2(2)} \boldsymbol{M}_{\neq \mathcal{G}_2}^Z, \tag{B.14}$$

where

$$\boldsymbol{M}_{\neq \mathcal{G}_2}^Y \in \mathbb{R}^{R_2 L_2 R_3 \times hS}, \qquad \boldsymbol{M}_{\neq \mathcal{G}_2}^Z \in \mathbb{R}^{R_2 L_2 R_3 \times Hs}. \tag{B.15}$$

The LR-HSI environment uses $(\mathcal{G}_1^Y)^{(t+1)}$, whereas the HR-MSI environment uses $\mathcal{G}_1^{(t+1)}$. Both environments further include $\mathcal{G}_3^{(t)}$, $\mathcal{C}^{(t)}$, and the current spectral factors.

The corresponding subproblem is

$$\begin{aligned}\boldsymbol{G}_{2(2)}^{(t+1)} = \arg\min_{\boldsymbol{G}_{2(2)}} \quad & \frac{1}{2} \left\| \boldsymbol{P}_2 \boldsymbol{G}_{2(2)} \boldsymbol{M}_{\neq \mathcal{G}_2}^Y - \boldsymbol{Y}_{(2)} \right\|_F^2 \\ & + \frac{\lambda}{2} \left\| \boldsymbol{G}_{2(2)} \boldsymbol{M}_{\neq \mathcal{G}_2}^Z - \boldsymbol{Z}_{(2)} \right\|_F^2 \\ & + \frac{\rho}{2} \left\| \boldsymbol{G}_{2(2)} - \boldsymbol{G}_{2(2)}^{(t)} \right\|_F^2. \end{aligned} \tag{B.16}$$

Its normal equation is

$$\begin{aligned}&\boldsymbol{P}_2^\top \boldsymbol{P}_2 \boldsymbol{G}_{2(2)}^{(t+1)} \boldsymbol{M}_{\neq \mathcal{G}_2}^Y \left(\boldsymbol{M}_{\neq \mathcal{G}_2}^Y\right)^\top \\ &\quad + \lambda \boldsymbol{G}_{2(2)}^{(t+1)} \boldsymbol{M}_{\neq \mathcal{G}_2}^Z \left(\boldsymbol{M}_{\neq \mathcal{G}_2}^Z\right)^\top + \rho \boldsymbol{G}_{2(2)}^{(t+1)} \\ &= \boldsymbol{P}_2^\top \boldsymbol{Y}_{(2)} \left(\boldsymbol{M}_{\neq \mathcal{G}_2}^Y\right)^\top + \lambda \boldsymbol{Z}_{(2)} \left(\boldsymbol{M}_{\neq \mathcal{G}_2}^Z\right)^\top + \rho \boldsymbol{G}_{2(2)}^{(t)}. \end{aligned} \tag{B.17}$$

The updated factors are

$$\mathcal{G}_2^{(t+1)} = \operatorname{fold}_2\left(\boldsymbol{G}_{2(2)}^{(t+1)}\right), \qquad (\mathcal{G}_2^Y)^{(t+1)} = \mathcal{G}_2^{(t+1)} \times_2 \boldsymbol{P}_2. \tag{B.18}$$

B.2.3 Update of $\mathcal{G}_3$

The factor $\mathcal{G}_3$ corresponds to a virtual singleton position. Its mode-2 unfolding is therefore a row vector, and the factor-isolated observation unfoldings reduce to $\boldsymbol{y}^\top$ and $\boldsymbol{z}^\top$ in equation B.3.

The two environment matrices satisfy

$$\boldsymbol{M}_{\neq \mathcal{G}_3}^Y \in \mathbb{R}^{R_3 L_3 R_1 \times hwS}, \qquad \boldsymbol{M}_{\neq \mathcal{G}_3}^Z \in \mathbb{R}^{R_3 L_3 R_1 \times HWs}. \tag{B.19}$$

The LR-HSI environment uses $(\mathcal{G}_1^Y)^{(t+1)}$ and $(\mathcal{G}_2^Y)^{(t+1)}$, whereas the HR-MSI environment uses $\mathcal{G}_1^{(t+1)}$ and $\mathcal{G}_2^{(t+1)}$. Both environments further include the current core and spectral factors.

The update of $\mathcal{G}_3$ is

$$\begin{aligned}\boldsymbol{G}_{3(2)}^{(t+1)} = \arg\min_{\boldsymbol{G}_{3(2)}} \quad & \frac{1}{2} \left\| \boldsymbol{G}_{3(2)} \boldsymbol{M}_{\neq \mathcal{G}_3}^Y - \boldsymbol{y}^\top \right\|_F^2 \\ & + \frac{\lambda}{2} \left\| \boldsymbol{G}_{3(2)} \boldsymbol{M}_{\neq \mathcal{G}_3}^Z - \boldsymbol{z}^\top \right\|_F^2 \\ & + \frac{\rho}{2} \left\| \boldsymbol{G}_{3(2)} - \boldsymbol{G}_{3(2)}^{(t)} \right\|_F^2. \end{aligned} \tag{B.20}$$

Its normal equation is

$$\begin{aligned}&\boldsymbol{G}_{3(2)}^{(t+1)}\boldsymbol{M}_{\neq\mathcal{G}_3}^{Y}\left(\boldsymbol{M}_{\neq\mathcal{G}_3}^{Y}\right)^{\top}\\&\quad+\lambda\boldsymbol{G}_{3(2)}^{(t+1)}\boldsymbol{M}_{\neq\mathcal{G}_3}^{Z}\left(\boldsymbol{M}_{\neq\mathcal{G}_3}^{Z}\right)^{\top}+\rho\boldsymbol{G}_{3(2)}^{(t+1)}\\&=\boldsymbol{y}^{\top}\left(\boldsymbol{M}_{\neq\mathcal{G}_3}^{Y}\right)^{\top}+\lambda\boldsymbol{z}^{\top}\left(\boldsymbol{M}_{\neq\mathcal{G}_3}^{Z}\right)^{\top}+\rho\boldsymbol{G}_{3(2)}^{(t)}.\end{aligned}\tag{B.21}$$

The updated tensor factor is recovered as

$$\mathcal{G}_3^{(t+1)}=\mathrm{fold}_2\left(\boldsymbol{G}_{3(2)}^{(t+1)}\right).\tag{B.22}$$

B.3 Updates of the Spectral Peripheral Factors

B.3.1 Update of $\mathcal{F}_1$

The factor $\mathcal{F}_1$ carries the spectral mode. After updating the three spatial factors, its environment matrices satisfy

$$\boldsymbol{M}_{\neq\mathcal{F}_1}^{Y}\in\mathbb{R}^{\widehat{R}_1U_1\widehat{R}_2\times hw},\qquad\boldsymbol{M}_{\neq\mathcal{F}_1}^{Z}\in\mathbb{R}^{\widehat{R}_1U_1\widehat{R}_2\times HW}.\tag{B.23}$$

The LR-HSI environment contracts the updated degraded spatial factors, whereas the HR-MSI environment contracts the updated full-resolution spatial factors. Both environments further include $\mathcal{C}^{(t)}$, $\mathcal{F}_2^{(t)}$, and $\mathcal{F}_3^{(t)}$.

The update of $\mathcal{F}_1$ is

$$\begin{aligned}\boldsymbol{F}_{1(2)}^{(t+1)}=\arg\min_{\boldsymbol{F}_{1(2)}}\quad&\frac{1}{2}\left\|\boldsymbol{F}_{1(2)}\boldsymbol{M}_{\neq\mathcal{F}_1}^{Y}-\boldsymbol{Y}_{(3)}\right\|_F^2\\&+\frac{\lambda}{2}\left\|\boldsymbol{P}_3\boldsymbol{F}_{1(2)}\boldsymbol{M}_{\neq\mathcal{F}_1}^{Z}-\boldsymbol{Z}_{(3)}\right\|_F^2\\&+\frac{\rho}{2}\left\|\boldsymbol{F}_{1(2)}-\boldsymbol{F}_{1(2)}^{(t)}\right\|_F^2.\end{aligned}\tag{B.24}$$

The corresponding normal equation is

$$\begin{aligned}&\boldsymbol{F}_{1(2)}^{(t+1)}\boldsymbol{M}_{\neq\mathcal{F}_1}^{Y}\left(\boldsymbol{M}_{\neq\mathcal{F}_1}^{Y}\right)^{\top}\\&\quad+\lambda\boldsymbol{P}_3^{\top}\boldsymbol{P}_3\boldsymbol{F}_{1(2)}^{(t+1)}\boldsymbol{M}_{\neq\mathcal{F}_1}^{Z}\left(\boldsymbol{M}_{\neq\mathcal{F}_1}^{Z}\right)^{\top}+\rho\boldsymbol{F}_{1(2)}^{(t+1)}\\&=\boldsymbol{Y}_{(3)}\left(\boldsymbol{M}_{\neq\mathcal{F}_1}^{Y}\right)^{\top}+\lambda\boldsymbol{P}_3^{\top}\boldsymbol{Z}_{(3)}\left(\boldsymbol{M}_{\neq\mathcal{F}_1}^{Z}\right)^{\top}+\rho\boldsymbol{F}_{1(2)}^{(t)}.\end{aligned}\tag{B.25}$$

The updated factor and its spectrally degraded counterpart are

$$\mathcal{F}_1^{(t+1)}=\mathrm{fold}_2\left(\boldsymbol{F}_{1(2)}^{(t+1)}\right),\qquad(\mathcal{F}_1^{Z})^{(t+1)}=\mathcal{F}_1^{(t+1)}\times_2\boldsymbol{P}_3.\tag{B.26}$$

B.3.2 Update of $\mathcal{F}_2$

The factor $\mathcal{F}_2$ corresponds to a virtual singleton position. Its mode-2 unfolding is a row vector, and the corresponding observation unfoldings are $\boldsymbol{y}^{\top}$ and $\boldsymbol{z}^{\top}$.

Its environment matrices satisfy

$$\boldsymbol{M}_{\neq\mathcal{F}_2}^{Y}\in\mathbb{R}^{\widehat{R}_2U_2\widehat{R}_3\times hwS},\qquad\boldsymbol{M}_{\neq\mathcal{F}_2}^{Z}\in\mathbb{R}^{\widehat{R}_2U_2\widehat{R}_3\times HWs}.\tag{B.27}$$

The LR-HSI environment uses $\mathcal{F}_1^{(t+1)}$, while the HR-MSI environment uses $(\mathcal{F}_1^{Z})^{(t+1)}$. Both environments further include $\mathcal{F}_3^{(t)}$, the updated spatial factors, and $\mathcal{C}^{(t)}$.

The subproblem is

$$\begin{aligned}\boldsymbol{F}_{2(2)}^{(t+1)}=\arg\min_{\boldsymbol{F}_{2(2)}}\quad&\frac{1}{2}\left\|\boldsymbol{F}_{2(2)}\boldsymbol{M}_{\neq\mathcal{F}_2}^{Y}-\boldsymbol{y}^{\top}\right\|_F^2\\&+\frac{\lambda}{2}\left\|\boldsymbol{F}_{2(2)}\boldsymbol{M}_{\neq\mathcal{F}_2}^{Z}-\boldsymbol{z}^{\top}\right\|_F^2\\&+\frac{\rho}{2}\left\|\boldsymbol{F}_{2(2)}-\boldsymbol{F}_{2(2)}^{(t)}\right\|_F^2.\end{aligned}\tag{B.28}$$

Its normal equation is

$$\begin{aligned}&\boldsymbol{F}_{2(2)}^{(t+1)}\boldsymbol{M}_{\neq\mathcal{F}_2}^{Y}\left(\boldsymbol{M}_{\neq\mathcal{F}_2}^{Y}\right)^{\top}\\&\quad+\lambda\boldsymbol{F}_{2(2)}^{(t+1)}\boldsymbol{M}_{\neq\mathcal{F}_2}^{Z}\left(\boldsymbol{M}_{\neq\mathcal{F}_2}^{Z}\right)^{\top}+\rho\boldsymbol{F}_{2(2)}^{(t+1)}\\&=\boldsymbol{y}^{\top}\left(\boldsymbol{M}_{\neq\mathcal{F}_2}^{Y}\right)^{\top}+\lambda\boldsymbol{z}^{\top}\left(\boldsymbol{M}_{\neq\mathcal{F}_2}^{Z}\right)^{\top}+\rho\boldsymbol{F}_{2(2)}^{(t)}.\end{aligned}\tag{B.29}$$

The updated tensor factor is

$$\mathcal{F}_2^{(t+1)}=\mathrm{fold}_2\left(\boldsymbol{F}_{2(2)}^{(t+1)}\right).\tag{B.30}$$

B.3.3 Update of $\mathcal{F}_3$

The factor $\mathcal{F}_3$ is also associated with a virtual singleton position. Its environment matrices are

$$\boldsymbol{M}_{\neq\mathcal{F}_3}^{Y}\in\mathbb{R}^{\widehat{R}_3U_3\widehat{R}_1\times hwS},\qquad\boldsymbol{M}_{\neq\mathcal{F}_3}^{Z}\in\mathbb{R}^{\widehat{R}_3U_3\widehat{R}_1\times HWs}.\tag{B.31}$$

The LR-HSI environment uses the latest $\mathcal{F}_1^{(t+1)}$ and $\mathcal{F}_2^{(t+1)}$, whereas the HR-MSI environment uses $(\mathcal{F}_1^{Z})^{(t+1)}$ and $\mathcal{F}_2^{(t+1)}$. Both environments further include the updated spatial factors and $\mathcal{C}^{(t)}$.

The corresponding subproblem is

$$\begin{aligned}\boldsymbol{F}_{3(2)}^{(t+1)}=\arg\min_{\boldsymbol{F}_{3(2)}}\quad&\frac{1}{2}\left\|\boldsymbol{F}_{3(2)}\boldsymbol{M}_{\neq\mathcal{F}_3}^{Y}-\boldsymbol{y}^{\top}\right\|_F^2\\&+\frac{\lambda}{2}\left\|\boldsymbol{F}_{3(2)}\boldsymbol{M}_{\neq\mathcal{F}_3}^{Z}-\boldsymbol{z}^{\top}\right\|_F^2\\&+\frac{\rho}{2}\left\|\boldsymbol{F}_{3(2)}-\boldsymbol{F}_{3(2)}^{(t)}\right\|_F^2.\end{aligned}\tag{B.32}$$

Its normal equation is

$$\begin{aligned}&\boldsymbol{F}_{3(2)}^{(t+1)}\boldsymbol{M}_{\neq\mathcal{F}_3}^{Y}\left(\boldsymbol{M}_{\neq\mathcal{F}_3}^{Y}\right)^{\top}\\&\quad+\lambda\boldsymbol{F}_{3(2)}^{(t+1)}\boldsymbol{M}_{\neq\mathcal{F}_3}^{Z}\left(\boldsymbol{M}_{\neq\mathcal{F}_3}^{Z}\right)^{\top}+\rho\boldsymbol{F}_{3(2)}^{(t+1)}\\&=\boldsymbol{y}^{\top}\left(\boldsymbol{M}_{\neq\mathcal{F}_3}^{Y}\right)^{\top}+\lambda\boldsymbol{z}^{\top}\left(\boldsymbol{M}_{\neq\mathcal{F}_3}^{Z}\right)^{\top}+\rho\boldsymbol{F}_{3(2)}^{(t)}.\end{aligned}\tag{B.33}$$

The updated factor is recovered as

$$\mathcal{F}_3^{(t+1)}=\mathrm{fold}_2\left(\boldsymbol{F}_{3(2)}^{(t+1)}\right).\tag{B.34}$$

B.4 Update of the Shared Waist Core

Let

$$L=L_1L_2L_3,\qquad U=U_1U_2U_3,\tag{B.35}$$

and define the matricized waist core as

$$\boldsymbol{C}=\mathrm{reshape}\left(\mathcal{C},L,U\right)\in\mathbb{R}^{L\times U}.\tag{B.36}$$

For the core update, define

$$\boldsymbol{Y}=\boldsymbol{Y}_{(3)}^{\top}\in\mathbb{R}^{hw\times S},\qquad\boldsymbol{Z}=\boldsymbol{Z}_{(3)}^{\top}\in\mathbb{R}^{HW\times s}.\tag{B.37}$$

After updating all six peripheral factors, contract the latest spatial and spectral cycles to obtain

$$\begin{aligned}&\boldsymbol{G}^{(t+1)}\in\mathbb{R}^{HW\times L},\quad\boldsymbol{G}^{Y,(t+1)}\in\mathbb{R}^{hw\times L},\\&\boldsymbol{F}^{(t+1)}\in\mathbb{R}^{S\times U},\qquad\boldsymbol{F}^{Z,(t+1)}\in\mathbb{R}^{s\times U}.\end{aligned}\tag{B.38}$$

Here, $\boldsymbol{G}^{Y,(t+1)}$ is formed from $(\mathcal{G}_1^{Y})^{(t+1)}$, $(\mathcal{G}_2^{Y})^{(t+1)}$, and $\mathcal{G}_3^{(t+1)}$, whereas $\boldsymbol{F}^{Z,(t+1)}$ is formed from $(\mathcal{F}_1^{Z})^{(t+1)}$, $\mathcal{F}_2^{(t+1)}$, and $\mathcal{F}_3^{(t+1)}$.

Algorithm 1 Matrix-Free PAM for THD-Based HSI–MSI Fusion

Require: $\mathcal{Y}$, $\mathcal{Z}$, $\boldsymbol{P}_1$, $\boldsymbol{P}_2$, $\boldsymbol{P}_3$, $\boldsymbol{r}_{\mathrm{THD}}^{(3,3)}$, λ, ρ, ρ_C, ϵ_{tol}, ϵ_0, and $T_{\max}$.
Ensure: Reconstructed HR-HSI $\widehat{\mathcal{X}}$.
1: Initialize the six peripheral factors and normalize their spoke slices.
2: Compute the initial degraded factors and construct $\boldsymbol{G}^{(0)}$, $\boldsymbol{G}^{Y,(0)}$, $\boldsymbol{F}^{(0)}$, and $\boldsymbol{F}^{Z,(0)}$; initialize $\mathcal{C}^{(0)}$ by matrix-free CG.
3: Reconstruct and project $\widehat{\mathcal{X}}^{(0)}$ onto $[0,1]$.
4: **for** $t = 0, \ldots, T_{\max} - 1$ **do**
5: Sequentially update $\mathcal{G}_1^{(t+1)}$, $\mathcal{G}_2^{(t+1)}$, and $\mathcal{G}_3^{(t+1)}$ by matrix-free CG; refresh $(\mathcal{G}_1^{Y})^{(t+1)}$ and $(\mathcal{G}_2^{Y})^{(t+1)}$.
6: Sequentially update $\mathcal{F}_1^{(t+1)}$, $\mathcal{F}_2^{(t+1)}$, and $\mathcal{F}_3^{(t+1)}$ by matrix-free CG; refresh $(\mathcal{F}_1^{Z})^{(t+1)}$.
7: Construct $\boldsymbol{G}^{(t+1)}$, $\boldsymbol{G}^{Y,(t+1)}$, $\boldsymbol{F}^{(t+1)}$, and $\boldsymbol{F}^{Z,(t+1)}$, and update $\mathcal{C}^{(t+1)}$ by matrix-free CG.
8: Perform spoke-wise balancing and absorb the removed scales into $\mathcal{C}^{(t+1)}$; recompute the degraded factors and peripheral matrices.
9: Reconstruct $\widehat{\mathcal{X}}^{(t+1)}$ and project it onto $[0,1]$.
10: **if** $\delta^{(t+1)} < \epsilon_{\mathrm{tol}}$ **then**
11: **break**
12: **end if**
13: **end for**
14: **return** the latest $\widehat{\mathcal{X}}$.

Under the column-major spatial vectorization convention, these matrices equivalently satisfy

$$\boldsymbol{G}^{Y,(t+1)} = (\boldsymbol{P}_2 \otimes \boldsymbol{P}_1)\boldsymbol{G}^{(t+1)}, \qquad \boldsymbol{F}^{Z,(t+1)} = \boldsymbol{P}_3 \boldsymbol{F}^{(t+1)}. \tag{B.39}$$

By Theorem 1, the observation models with respect to $\boldsymbol{C}$ are

$$\boldsymbol{Y} \approx \boldsymbol{G}^{Y,(t+1)} \boldsymbol{C} \boldsymbol{F}^{(t+1)\top}, \qquad \boldsymbol{Z} \approx \boldsymbol{G}^{(t+1)} \boldsymbol{C} \boldsymbol{F}^{Z,(t+1)\top}. \tag{B.40}$$

The core subproblem is therefore

$$\begin{aligned} \boldsymbol{C}^{(t+1)} = \arg\min_{\boldsymbol{C}} \quad & \frac{1}{2} \left\| \boldsymbol{Y} - \boldsymbol{G}^{Y,(t+1)} \boldsymbol{C} \boldsymbol{F}^{(t+1)\top} \right\|_F^2 \\ & + \frac{\lambda}{2} \left\| \boldsymbol{Z} - \boldsymbol{G}^{(t+1)} \boldsymbol{C} \boldsymbol{F}^{Z,(t+1)\top} \right\|_F^2 \\ & + \frac{\rho_C}{2} \left\| \boldsymbol{C} - \boldsymbol{C}^{(t)} \right\|_F^2 . \end{aligned} \tag{B.41}$$

Setting the derivative to zero gives

$$\begin{aligned} & \left(\boldsymbol{G}^{Y,(t+1)} \right)^\top \boldsymbol{G}^{Y,(t+1)} \boldsymbol{C}^{(t+1)} \boldsymbol{F}^{(t+1)\top} \boldsymbol{F}^{(t+1)} \\ & \quad + \lambda \boldsymbol{G}^{(t+1)\top} \boldsymbol{G}^{(t+1)} \boldsymbol{C}^{(t+1)} \left(\boldsymbol{F}^{Z,(t+1)} \right)^\top \boldsymbol{F}^{Z,(t+1)} \\ & \quad + \rho_C \boldsymbol{C}^{(t+1)} \\ & = \left(\boldsymbol{G}^{Y,(t+1)} \right)^\top \boldsymbol{Y} \boldsymbol{F}^{(t+1)} \\ & \quad + \lambda \boldsymbol{G}^{(t+1)\top} \boldsymbol{Z} \boldsymbol{F}^{Z,(t+1)} + \rho_C \boldsymbol{C}^{(t)}. \end{aligned} \tag{B.42}$$

After solving this equation, the sixth-order waist core is recovered by

$$\mathcal{C}^{(t+1)} = \mathrm{reshape}\left(\boldsymbol{C}^{(t+1)}, L_1, L_2, L_3, U_1, U_2, U_3 \right). \tag{B.43}$$

B.5 Matrix-Free Conjugate-Gradient Solution

The environment matrices introduced above are used only to derive the block-wise least-squares problems and normal equations. They are not explicitly assembled in the implementation.

For each peripheral-factor update, the left-hand side of the corresponding normal equation is evaluated by contracting a trial factor with the remaining THD components, applying the appropriate

spatial or spectral degradation, and back-projecting the resulting residual through the adjoint contractions. Therefore, neither the environment Gram matrices nor any Kronecker-product coefficient matrices need to be explicitly formed.

The core normal equation is treated similarly. Its left-hand side is evaluated through successive matrix multiplications by $\boldsymbol{G}^Y$, $\boldsymbol{G}$, $\boldsymbol{F}$, and $\boldsymbol{F}^Z$ and their transposes. Conjugate gradient is applied directly to the matrix unknown under the Frobenius inner product

$$\langle \boldsymbol{A}, \boldsymbol{B} \rangle_F = \operatorname{tr}\left(\boldsymbol{A}^\top \boldsymbol{B}\right), \tag{B.44}$$

without explicitly vectorizing the unknown matrices.

Proposition 8 shows that all block-wise normal operators are self-adjoint and positive definite. The current estimate of each factor or core is used as the initial point of the corresponding CG iteration. In the reported implementation, CG is terminated when the relative residual is below 10^{-6} or after 10 inner iterations, whichever occurs first. The complete implementation settings are summarized in Table 2.

B.6 Initialization and Spoke-Wise Balancing

The six peripheral factors are randomly initialized and normalized along their spoke modes. After constructing the initial peripheral matrices $\boldsymbol{G}^{(0)}$, $\boldsymbol{G}^{Y,(0)}$, $\boldsymbol{F}^{(0)}$, and $\boldsymbol{F}^{Z,(0)}$, the initial core is obtained by solving

$$\begin{aligned} \boldsymbol{C}^{(0)} = \arg\min_{\boldsymbol{C}} \quad & \frac{1}{2}\left\| \boldsymbol{Y} - \boldsymbol{G}^{Y,(0)} \boldsymbol{C} \boldsymbol{F}^{(0)\top} \right\|_F^2 \\ & + \frac{\lambda}{2}\left\| \boldsymbol{Z} - \boldsymbol{G}^{(0)} \boldsymbol{C} \boldsymbol{F}^{Z,(0)\top} \right\|_F^2 + \frac{\rho_C}{2}\left\|\boldsymbol{C}\right\|_F^2 . \end{aligned} \tag{B.45}$$

This problem has the same normal-equation structure as equation B.42, with the proximal center set to zero.

To alleviate factor–core scale imbalance, spoke-wise balancing is performed after every complete Gauss–Seidel sweep. Let $\epsilon_0 > 0$ be a small numerical safeguard. For $i = 1, 2, 3$ and $\ell = 1, \ldots, L_i$, define

$$a_{i,\ell} = \begin{cases} \left\| \mathcal{G}_i(:, :, \ell, :) \right\|_F, & \left\| \mathcal{G}_i(:, :, \ell, :) \right\|_F > \epsilon_0, \\ 1, & \text{otherwise.} \end{cases} \tag{B.46}$$

Let

$$\boldsymbol{a}_i = \left(a_{i,1}, \ldots, a_{i,L_i}\right)^\top . \tag{B.47}$$

The corresponding spatial factor and core mode are rescaled according to

$$\mathcal{G}_i(:, :, \ell, :) \leftarrow \frac{\mathcal{G}_i(:, :, \ell, :)}{a_{i,\ell}}, \qquad \mathcal{C} \leftarrow \mathcal{C} \times_i \operatorname{Diag}\left(\boldsymbol{a}_i\right). \tag{B.48}$$

Similarly, for $j = 1, 2, 3$ and $u = 1, \ldots, U_j$, define

$$b_{j,u} = \begin{cases} \left\| \mathcal{F}_j(:, :, u, :) \right\|_F, & \left\| \mathcal{F}_j(:, :, u, :) \right\|_F > \epsilon_0, \\ 1, & \text{otherwise.} \end{cases} \tag{B.49}$$

Let

$$\boldsymbol{b}_j = \left(b_{j,1}, \ldots, b_{j,U_j}\right)^\top . \tag{B.50}$$

The corresponding spectral factor and core mode are rescaled as

$$\mathcal{F}_j(:, :, u, :) \leftarrow \frac{\mathcal{F}_j(:, :, u, :)}{b_{j,u}}, \qquad \mathcal{C} \leftarrow \mathcal{C} \times_{3+j} \operatorname{Diag}\left(\boldsymbol{b}_j\right). \tag{B.51}$$

Each removed spoke scale is absorbed into the corresponding core mode. By Proposition 6, these transformations leave the THD reconstruction and both data fidelity terms unchanged while improving numerical conditioning.

B.7 Sufficient Decrease of an Exact PAM Sweep

Let $\boldsymbol{\Theta}$ collect the six peripheral factors and the waist core, and define the unregularized fusion objective

$$\Phi(\boldsymbol{\Theta}) = \frac{1}{2}\left\|\mathcal{Y} - \widehat{\mathcal{Y}}(\boldsymbol{\Theta})\right\|_F^2 + \frac{\lambda}{2}\left\|\mathcal{Z} - \widehat{\mathcal{Z}}(\boldsymbol{\Theta})\right\|_F^2, \tag{B.52}$$

where $\widehat{\mathcal{Y}}$ and $\widehat{\mathcal{Z}}$ are the two THD observation reconstructions. Let $\overline{\boldsymbol{\Theta}}^{(t+1)}$ denote the factors obtained after an exact seven-block Gauss–Seidel sweep and before spoke-wise balancing.

Proposition 9 (Sufficient decrease) *For exact minimization of every proximal block subproblem,*

$$\begin{aligned}\Phi(\boldsymbol{\Theta}^{(t)}) - \Phi(\overline{\boldsymbol{\Theta}}^{(t+1)}) \\ \geq \frac{\rho}{2}\sum_{k=1}^{3}\left\|\overline{\mathcal{G}}_k^{(t+1)} - \mathcal{G}_k^{(t)}\right\|_F^2 + \frac{\rho}{2}\sum_{j=1}^{3}\left\|\overline{\mathcal{F}}_j^{(t+1)} - \mathcal{F}_j^{(t)}\right\|_F^2 \\ + \frac{\rho_C}{2}\left\|\overline{\mathcal{C}}^{(t+1)} - \mathcal{C}^{(t)}\right\|_F^2.\end{aligned} \tag{B.53}$$

If $\boldsymbol{\Theta}^{(t+1)}$ is obtained from $\overline{\boldsymbol{\Theta}}^{(t+1)}$ by the balancing operation, then

$$\Phi(\boldsymbol{\Theta}^{(t+1)}) = \Phi(\overline{\boldsymbol{\Theta}}^{(t+1)}), \tag{B.54}$$

and therefore $\{\Phi(\boldsymbol{\Theta}^{(t)})\}$ is nonincreasing. Moreover, the sum of the squared exact pre-balancing block displacements is finite.

Proof 13 *Consider one block $\mathcal{V}$. Exact minimization of its proximal subproblem implies that the objective value at the new block is no larger than the value obtained by substituting the old block. After canceling the zero proximal term at the old block,*

$$\Phi(\boldsymbol{\Theta}_{\text{before}}) - \Phi(\boldsymbol{\Theta}_{\text{after}}) \geq \frac{\rho_{\mathcal{V}}}{2}\left\|\mathcal{V}^{\text{new}} - \mathcal{V}^{\text{old}}\right\|_F^2. \tag{B.55}$$

Summing this inequality over the seven Gauss–Seidel updates cancels all intermediate objective values and gives equation B.53. The balancing step preserves the represented tensor by Proposition 6, and therefore preserves Φ. Since $\Phi \geq 0$, summing equation B.53 over the outer iterations proves finiteness of the squared pre-balancing block displacements.

The proposition is stated for exact block solves. The same monotonicity conclusion holds for inexact CG updates whenever each inner solve achieves the corresponding sufficient objective decrease. It does not by itself remove the intrinsic gauge nonuniqueness of the THD factors; the balancing step is used to select a numerically stable representative.

B.8 Reconstruction and Stopping Criterion

After updating and balancing all THD factors, the latest HR-HSI estimate is reconstructed as

$$\widehat{\mathcal{X}}^{(t+1)} = \text{reshape}\left(\boldsymbol{G}^{(t+1)}\boldsymbol{C}^{(t+1)}\boldsymbol{F}^{(t+1)\top}, H, W, S\right). \tag{B.56}$$

For output evaluation and stopping-condition computation, the reconstructed HR-HSI is projected element-wise onto the valid intensity interval:

$$\widehat{\mathcal{X}}^{(t+1)} \leftarrow \min\left\{\max\left\{\widehat{\mathcal{X}}^{(t+1)}, 0\right\}, 1\right\}. \tag{B.57}$$

This projection is applied only to the reconstructed output and does not modify the THD factors during their block-wise optimization.

The iterations terminate when

$$\frac{\left\|\widehat{\mathcal{X}}^{(t+1)} - \widehat{\mathcal{X}}^{(t)}\right\|_F}{\max\left\{\left\|\widehat{\mathcal{X}}^{(t)}\right\|_F, \epsilon_0\right\}} < \epsilon_{\text{tol}}, \tag{B.58}$$

or when the prescribed maximum number of outer iterations is reached, where $\epsilon_{\text{tol}} > 0$ denotes the convergence tolerance.

C Computational and Storage Complexity

This section derives the dominant cost reported in the main text for one outer iteration of the matrix-free PAM solver. Define

$$d_k^G = R_k L_k R_{k+1}, \qquad d_j^F = \widehat{R}_j U_j \widehat{R}_{j+1}, \tag{C.1}$$

for $k, j = 1, 2, 3$, with cyclic conventions $R_4 = R_1$ and $\widehat{R}_4 = \widehat{R}_1$. Also define

$$L = L_1 L_2 L_3, \qquad U = U_1 U_2 U_3, \qquad \Omega = hwS + HWs. \tag{C.2}$$

Let N_{G_k}, N_{F_j}, and N_C denote the numbers of CG iterations used for the corresponding block updates.

C.1 Peripheral-Factor Updates

For a trial spatial factor $\mathcal{G}_k$, one matrix-free normal-operator evaluation consists of a forward THD contraction into the LR-HSI and HR-MSI observation spaces, followed by the corresponding adjoint contractions. The active latent dimension per physical-mode row is d_k^G. A conservative dominant bound for one such forward–adjoint pair is therefore

$$\mathcal{O}(\Omega d_k^G). \tag{C.3}$$

The same argument gives $\mathcal{O}(\Omega d_j^F)$ for a spectral factor. Hence the six peripheral updates require

$$\mathcal{O}\left(\Omega\left[\sum_{k=1}^{3} N_{G_k} d_k^G + \sum_{j=1}^{3} N_{F_j} d_j^F\right]\right). \tag{C.4}$$

This estimate includes the leading contractions with the two observations and suppresses lower-order index permutations, small diagonal scalings, and reshaping operations.

C.2 Waist-Core Update

The right-hand side of equation B.42 is formed without Kronecker products. For the LR-HSI term, first computing $\boldsymbol{Y}\boldsymbol{F}$ and then left-multiplying by $(\boldsymbol{G}^Y)^\top$ costs

$$\mathcal{O}(hwSU + hwLU) = \mathcal{O}\big(hwU(S + L)\big). \tag{C.5}$$

The HR-MSI term analogously costs

$$\mathcal{O}\big(HWU(s + L)\big). \tag{C.6}$$

The spatial Gram matrices $(\boldsymbol{G}^Y)^\top \boldsymbol{G}^Y$ and $\boldsymbol{G}^\top \boldsymbol{G}$ require

$$\mathcal{O}\big((hw + HW)L^2\big), \tag{C.7}$$

whereas the spectral Gram matrices $\boldsymbol{F}^\top \boldsymbol{F}$ and $(\boldsymbol{F}^Z)^\top \boldsymbol{F}^Z$ require

$$\mathcal{O}\big((S + s)U^2\big). \tag{C.8}$$

For a trial core matrix $\boldsymbol{Q} \in \mathbb{R}^{L \times U}$, each normal-operator evaluation has the form

$$\mathscr{H}_C(\boldsymbol{Q}) = \boldsymbol{A}_Y \boldsymbol{Q} \boldsymbol{B}_Y + \lambda \boldsymbol{A}_Z \boldsymbol{Q} \boldsymbol{B}_Z + \rho_C \boldsymbol{Q}, \tag{C.9}$$

where the two $\boldsymbol{A}$ matrices are $L \times L$ and the two $\boldsymbol{B}$ matrices are $U \times U$. Successive left and right multiplications cost

$$\mathcal{O}(L^2 U + L U^2) = \mathcal{O}\big(LU(L + U)\big). \tag{C.10}$$

Thus the N_C core CG iterations require

$$\mathcal{O}\big(N_C LU(L + U)\big). \tag{C.11}$$

Combining equation C.4 with the core costs gives the dominant per-iteration complexity

$$\begin{aligned} T_{\mathrm{THD}} = \mathcal{O}\Bigg(\Omega\left[\sum_{k=1}^{3} N_{G_k} d_k^G + \sum_{j=1}^{3} N_{F_j} d_j^F\right] + hwU(L + S) \\ + HWU(L + s) + (hw + HW)L^2 \\ + (S + s)U^2 + N_C LU(L + U)\Bigg), \end{aligned} \tag{C.12}$$

which is the expression stated in the main paper.

C.3 Parameter Count and Working Memory

For a general realization, the number of free scalar parameters is

$$P_{\mathrm{THD}} = \sum_{k=1}^{M_L} R_k \widetilde{I}_k^L L_k R_{k+1} + \sum_{j=1}^{M_R} \widehat{R}_j \widetilde{I}_j^R U_j \widehat{R}_{j+1} + LU. \tag{C.13}$$

For the $(3, 3)$ HSI–MSI realization, this becomes

$$\begin{aligned} P_{\mathrm{THD}}^{(3,3)} = {} & HR_1L_1R_2 + WR_2L_2R_3 + R_3L_3R_1 \\ & + S\widehat{R}_1U_1\widehat{R}_2 + \widehat{R}_2U_2\widehat{R}_3 + \widehat{R}_3U_3\widehat{R}_1 + LU. \end{aligned} \tag{C.14}$$

The singleton factors contain no additional physical-mode size but retain cyclic and spoke parameters.

Besides the observations and model parameters, the dominant stored intermediates are the four peripheral matrices $\boldsymbol{G}$, $\boldsymbol{G}^Y$, $\boldsymbol{F}$, and $\boldsymbol{F}^Z$, the waist core, and a constant number of CG work vectors. The resulting working-memory bound is

$$\mathcal{O}\left(\Omega + (HW + hw)L + (S + s)U + LU + P_{\mathrm{THD}}\right). \tag{C.15}$$

In particular, the implementation does not store the explicit environment matrices or the Kronecker-product normal coefficients. Their dimensions can be much larger than the active factors, which is the principal memory advantage of the matrix-free formulation.

D Additional Experimental Details

D.1 Datasets and Observation Generation

Datasets. We conduct simulated HSI–MSI fusion experiments on three widely used remote-sensing hyperspectral datasets. The detailed information and cropping settings are described as follows.

1. *Pavia University (PU):* The PU dataset was acquired by the Reflective Optics System Imaging Spectrometer (ROSIS-03) over the University of Pavia, Italy. The original image has a spatial size of 610×340 pixels and a spatial resolution of approximately 1.3 m. After removing noisy and water-absorption bands, we select a $256 \times 256 \times 93$ sub-cube containing diverse urban structures, vegetation regions, and shadow areas as the ground-truth HR-HSI.
2. *Pavia City Center (PC):* The PC dataset was also collected by the ROSIS-03 sensor over the city center of Pavia, Italy. It covers a larger urban scene with an original spatial size of 1096×715 pixels and a spatial resolution of approximately 1.3 m. After discarding severely corrupted bands, a $200 \times 200 \times 80$ region containing roads, buildings, vegetation, and other fine spatial structures is cropped as the reference HR-HSI.
3. *Washington DC Mall (WDC):* The WDC dataset was captured by the Hyperspectral Digital Imagery Collection Experiment (HYDICE) airborne sensor over the National Mall in Washington, DC. The original scene contains 210 spectral bands and has a spatial size of 1280×307 pixels. After removing noisy and water-vapor-absorption bands, we extract a $256 \times 256 \times 90$ sub-cube as the ground-truth HR-HSI. This region contains complex urban materials, vegetation, roads, and building boundaries, providing a challenging test for both spatial reconstruction and spectral preservation.

These three datasets exhibit different spatial layouts, material compositions, texture densities, and spectral distributions, enabling a comprehensive evaluation of the generalization and robustness of the compared fusion methods.

Observation generation. Following the uniform-sampling protocol commonly used in model-based HSISR evaluation, the LR-HSI is generated by applying a 7×7 Gaussian blur with standard deviation 2 to the reference HR-HSI, followed by spatial downsampling by a factor of four along both spatial dimensions. The HR-MSI is synthesized using an IKONOS-like spectral response function,

Table 2: Implementation and observation settings of THD.

Setting	Value
THD parameters	
Cyclic ranks $[R_1, R_2, R_3, \widehat{R}_1, \widehat{R}_2, \widehat{R}_3]$	$[4, 150, 4, 5, 5, 5]$
Spoke ranks $[L_1, L_2, L_3, U_1, U_2, U_3]$	$[3, 3, 3, 3, 3, 3]$
Observation weight λ	1
Proximal coefficients (ρ, ρ_C)	$(0.1, 0.1)$
Maximum outer iterations	30
CG iterations per block	10
CG tolerance	10^{-6}
Outer stopping tolerance	10^{-5}
Spoke-wise balancing	Enabled
Observation model	
Spatial downsampling factor	$\times 4$
Blur kernel	7×7 Gaussian
Gaussian standard deviation	2
Spectral response	IKONOS-like response
LR-HSI/HR-MSI SNR	$20/20$ and $15/15$ dB
Random seed	10

Table 3: Quantitative comparison on PU, PC, and WDC under the $\times 8$ spatial downsampling factor at SNRs of 20 and 15 dB. The best and second-best results are highlighted in red and blue, respectively.

Method	Pavia University					Pavia City Center					Washington DC Mall				
	PSNR↑	SSIM↑	SAM↓	ERGAS↓	DD↓	PSNR↑	SSIM↑	SAM↓	ERGAS↓	DD↓	PSNR↑	SSIM↑	SAM↓	ERGAS↓	DD↓
×8, SNR = 20 dB for both LR-HSI and HR-MSI															
CSTF	29.7956	0.8464	12.2907	2.3847	6.7695	31.0828	0.7963	8.1290	2.1802	5.7382	28.4116	0.8628	11.0432	2.3456	7.8321
LFCTNN	29.0636	0.8052	13.8291	2.8538	7.8601	31.8393	0.8120	7.3246	2.0459	5.2683	29.2206	0.8747	10.2429	2.2060	7.1543
GTNN	29.8449	0.8346	12.5726	2.5008	6.9788	31.6433	0.8100	7.6059	2.0486	5.3542	28.6840	0.8652	10.4713	2.2983	7.5452
CTDF	31.4353	0.8870	10.0209	1.8846	5.4751	31.7088	0.8148	7.3688	1.9753	5.2752	27.3450	0.8329	12.5779	2.6013	8.9732
GNLTN	29.2910	0.8137	13.3768	2.7389	7.5924	31.6522	0.8088	7.4201	2.0858	5.3668	27.9807	0.8522	11.3825	2.4874	8.2687
MHTNN	29.2950	0.8138	13.3708	2.7362	7.5873	31.6274	0.8090	7.4161	2.0925	5.3748	27.9831	0.8523	11.3777	2.4877	8.2655
EDIP-Net	27.5549	0.7970	12.6814	2.8718	8.3447	29.9065	0.8248	8.6406	2.3405	6.1425	26.2429	0.8216	11.3626	2.9196	9.5425
EU2ADL	25.9630	0.7162	11.1150	3.9682	11.3079	33.3811	0.8960	4.8459	1.5623	4.1963	17.9748	0.2077	14.8135	7.3086	25.3316
THD (Ours)	33.9809	0.9368	6.9596	1.3784	4.0525	33.9659	0.8874	4.9915	1.5054	4.0118	30.3088	0.9187	7.8919	1.8446	6.2931
×8, SNR = 15 dB for both LR-HSI and HR-MSI															
CSTF	23.6756	0.6116	20.8585	5.0671	13.9399	26.1068	0.5924	14.1679	3.8773	10.3111	23.8091	0.7099	17.3365	4.0731	13.3930
LFCTNN	22.2595	0.5335	24.2860	6.5573	17.6170	26.2192	0.5902	13.9287	3.8547	10.2963	24.3058	0.7218	16.3873	3.9404	12.7296
GTNN	22.5434	0.5475	23.4793	6.2015	16.7785	26.2926	0.5929	13.8331	3.8066	10.1726	24.2232	0.7176	16.3975	3.9383	12.7866
CTDF	25.9991	0.7176	16.4796	3.5058	10.2078	26.0457	0.5830	14.1914	3.8020	10.2113	22.0780	0.6309	20.6199	4.8285	16.1672
GNLTN	23.0457	0.5726	22.3449	5.7988	15.6307	26.1620	0.5879	13.9800	3.8737	10.3576	23.8424	0.7072	17.0704	4.0997	13.3892
MHTNN	23.0438	0.5725	22.3523	5.7997	15.6358	26.1552	0.5879	13.9803	3.8762	10.3636	23.8388	0.7069	17.0622	4.1022	13.3944
EDIP-Net	23.7743	0.6523	19.3650	4.4216	12.6813	24.5846	0.5891	16.4106	4.3395	11.4355	22.3388	0.6657	18.3282	4.5948	14.8372
EU2ADL	25.7892	0.7294	13.0294	3.5629	10.4021	28.2911	0.6945	9.6202	2.9683	7.8530	17.6351	0.1944	17.6315	7.6234	26.4444
THD (Ours)	28.8564	0.8221	11.4767	2.5148	7.3692	30.0906	0.7663	7.4954	2.3150	6.3598	26.9224	0.8281	11.0362	2.6587	9.2664

resulting in four multispectral bands. Zero-mean Gaussian noise is independently added to the LR-HSI and HR-MSI. For a clean observation $\mathcal{A}$ and noise $\mathcal{N}$, the noise magnitude is scaled to satisfy

$$\mathrm{SNR} = 10 \log_{10} \frac{\|\mathcal{A}\|_F^2}{\|\mathcal{N}\|_F^2}. \tag{D.1}$$

We evaluate two conditions in which both observations have an SNR of 20 or 15 dB. Every method receives exactly the same LR-HSI and HR-MSI for each dataset and noise level.

D.2 Real-Data Setting

For the real-world HSISR experiment, we use a registered Hyperion HSI and Sentinel-2 MSI pair covering the same geographical region. The Hyperion sensor aboard the Earth Observing-1 satellite acquires 220 spectral bands over approximately 400–2500 nm at a spatial resolution of 30 m. After removing noisy and severely corrupted bands, 89 bands are retained, yielding an LR-HSI crop of size $180 \times 180 \times 89$.

The Sentinel-2 MSI contains 13 spectral bands with different spatial resolutions. We select the four 10-m bands centered at approximately 490, 560, 665, and 842 nm, corresponding to the blue, green, red, and near-infrared channels, respectively. A spatially aligned region of size $540 \times 540 \times 4$ is extracted as the HR-MSI. Therefore, the two observations cover the same scene with a spatial-resolution ratio of three.

Because no reference HR-HSI is available for this real-data pair, the spatial blur kernel and spectral response matrix are estimated from the observations using the method in Dell'Acqua et al. (2004). The estimated operators are then used in the THD observation model without introducing synthetic noise or simulated downsampling. Reconstruction quality is evaluated primarily through visual spatial-detail preservation, spectral-color consistency, and consistency with the observed LR-HSI and HR-MSI.

D.3 Compared Methods and Evaluation Protocol

We compare THD with six model-driven methods, namely CSTF Li et al. (2018), CTDF Xu et al. (2024b), LFCTNN Xu et al. (2024a), GTNN Dian et al. (2025), GNLTN Xu et al. (2025b), and MHTNN Tu et al. (2026), together with two deep learning methods, EU2ADL Gao et al. (2023) and EDIP-Net Li et al. (2025). These baselines cover coupled tensor factorization, tensor-network decomposition, transform-based tensor nuclear norms, and data-driven fusion frameworks.

We use the official implementations and follow the parameter settings recommended by the original authors whenever available. Unspecified parameters are selected within the reported ranges and kept consistent under the same experimental condition. For the deep learning methods, the released training and inference protocols are followed. All methods are evaluated using identical LR-HSI and HR-MSI observations generated with the same spatial degradation, spectral response, and noise level. Since the original studies adopt different crops, degradation operators, noise levels, and preprocessing protocols, their published results are not directly comparable; we therefore rerun all available baselines on identical observations under the unified protocol.

All model-driven methods are evaluated in MATLAB R2025 on a workstation equipped with an AMD R5 CPU and 32 GB RAM. The deep-learning baselines are executed using PyTorch 1.12 on an NVIDIA RTX 3090 GPU. Runtime comparisons include only methods evaluated in the same MATLAB environment.

D.4 Evaluation Metrics

Five full-reference metrics are employed for quantitative evaluation: peak signal-to-noise ratio (PSNR), structural similarity index (SSIM), spectral angle mapper (SAM), relative dimensionless global error in synthesis (ERGAS), and distortion degree (DD). PSNR measures the average radiometric reconstruction accuracy, whereas SSIM evaluates the structural similarity between the reconstructed and reference images. SAM quantifies the angular discrepancy between the reconstructed and reference spectral vectors and is therefore particularly sensitive to spectral distortion. ERGAS measures the normalized global reconstruction error, while DD characterizes the overall distortion of the fused image relative to the reference. Higher PSNR and SSIM values indicate better reconstruction quality, whereas lower SAM, ERGAS, and DD values are preferred. In all tables, the best and second-best results are highlighted in red and blue, respectively.

D.5 Effect of Peripheral-Cycle Realizations

To evaluate the effect of virtual singleton positions and peripheral-cycle balance, we compare four THD realizations on the Pavia-80 crop under $\times 4$ spatial downsampling, with both the LR-HSI and HR-MSI corrupted at 15-dB SNR. All realizations use the same degradation operators, optimization protocol, and stopping criterion.

As shown in Table 4, the order-matched balanced realization $(3, 3)$ achieves the best reconstruction accuracy across all four metrics. Compared with the minimal realization $(2, 1)$, it improves PSNR by 1.85 dB and SSIM by 0.0634, while reducing SAM and ERGAS by 22.80% and 23.32%, respectively. These gains indicate that the additional virtual positions can enhance latent spatial–spectral coordination when they are introduced on both peripheral sides.

Table 4: Comparison of different peripheral-cycle realizations on Pavia-80 under $\times 4$ downsampling and 15-dB SNR.

Realization	PSNR↑	SSIM↑	SAM↓	ERGAS↓	Time↓
$(2,1)$	<u>28.24</u>	<u>0.7942</u>	<u>12.37</u>	<u>5.604</u>	**29.7**
$(2,2)$	26.86	0.7460	14.66	6.452	<u>35.4</u>
$(3,1)$	27.43	0.7688	13.64	6.026	164.9
$(3,3)$	**30.09**	**0.8576**	**9.55**	**4.297**	162.5

Simply increasing one peripheral-cycle length does not guarantee improved performance. Both $(2,2)$ and $(3,1)$ perform worse than the minimal $(2,1)$ realization. In particular, $(3,1)$ requires a runtime comparable to that of $(3,3)$ but produces substantially lower reconstruction accuracy. This comparison suggests that the advantage of $(3,3)$ does not arise merely from a longer optimization time; rather, the balanced extension of both peripheral cycles provides a more effective latent parameterization for coordinating the heterogeneous spatial and spectral groups.

The improved accuracy is accompanied by additional computational cost. The runtime of $(3,3)$ is approximately 5.47 times that of $(2,1)$. Therefore, $(2,1)$ offers a more efficient alternative when runtime is the primary concern, whereas $(3,3)$ is adopted in the main experiments because it provides the most favorable reconstruction accuracy.

D.6 Quantitative Analysis under $\times 8$ Downsampling

Table 3 presents the quantitative results under the more challenging $\times 8$ spatial downsampling setting. Among the 30 dataset–metric–noise combinations, THD ranks first in 28 cases and second in the remaining two, showing a consistent advantage across different scenes and noise levels.

Results at SNR = 20 dB. On PU, THD achieves a PSNR of 33.9809 dB, exceeding the second-best CTDF by 2.5456 dB. It also improves SSIM by 0.0498 and reduces SAM, ERGAS, and DD by 30.55%, 26.86%, and 25.98%, respectively. These substantial improvements indicate that THD can recover fine spatial structures while maintaining the spectral characteristics of the original HSI under severe spatial degradation.

On PC, THD obtains the highest PSNR, ERGAS, and DD results. Compared with EU2ADL, the strongest competing method on this dataset, THD improves PSNR by 0.5848 dB and reduces ERGAS and DD by 3.64% and 4.40%, respectively. EU2ADL achieves slightly better SSIM and SAM, with margins of only 0.0086 and 0.1456, respectively. This comparison suggests that the deep model is competitive in local structural and angular spectral measurements, whereas THD provides better overall reconstruction accuracy and lower global distortion.

On WDC, THD outperforms the second-best LFCTNN by 1.0882 dB in PSNR and 0.0440 in SSIM. Moreover, it reduces SAM, ERGAS, and DD by 22.95%, 16.38%, and 12.04%, respectively. The consistent gains on WDC are notable because this scene contains complex urban boundaries and heterogeneous materials, which place greater demands on spatial–spectral coordination.

Results at SNR = 15 dB. Under the stronger noise condition, THD achieves the best result for every metric on all three datasets. On PU, it exceeds the second-best PSNR by 2.8573 dB and improves SSIM by 0.0927. The corresponding reductions in SAM, ERGAS, and DD are 11.92%, 28.27%, and 27.81%. On PC, THD outperforms EU2ADL by 1.7995 dB in PSNR and reduces SAM, ERGAS, and DD by 22.09%, 22.01%, and 19.01%, respectively. On WDC, the PSNR advantage over LFCTNN reaches 2.6166 dB, while SAM, ERGAS, and DD are reduced by 32.65%, 32.49%, and 27.21%, respectively.

Overall, the superiority of THD becomes more evident at 15 dB, particularly on WDC. This behavior indicates that separately modeling the heterogeneous spatial and spectral structures prevents noise from being indiscriminately propagated across different modes. Meanwhile, the shared waist core coordinates the two mode groups, allowing spatial details to be recovered under the constraint of the spectral information provided by the LR-HSI. Consequently, THD maintains both reconstruction accuracy and spectral fidelity under simultaneous severe spatial degradation and observation noise.

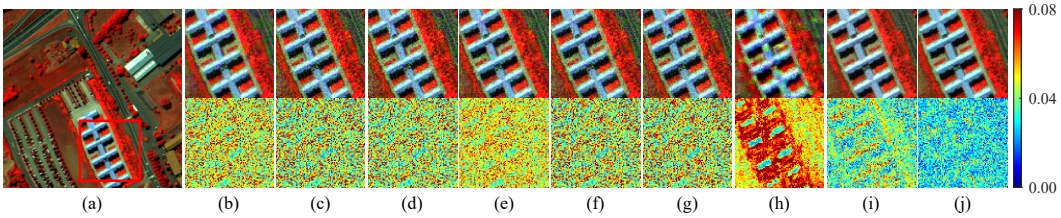


Figure 9: Visual comparison on the original ROSIS scene at $\times 8$ downsampling and 15-dB SNR: (a) Reference; (b) CSTF; (c) LFCTNN; (d) GTNN; (e) CTDF; (f) GNLTN; (g) MHTNN; (h) EDIP-Net; (i) EU2ADL; and (j) THD. For (b)–(j), the top and bottom panels show pseudo-color reconstructions and absolute-error maps, respectively.

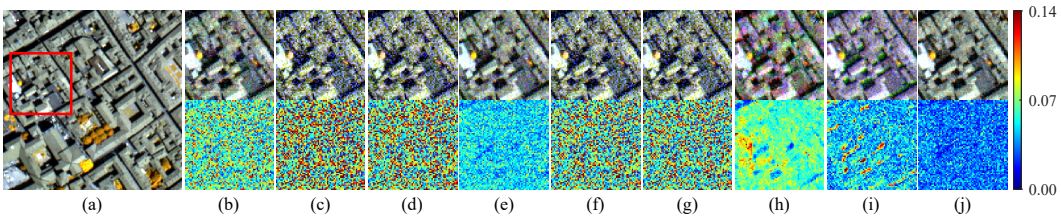


Figure 10: Visual comparison on the Pavia scene at $\times 8$ downsampling and 15-dB SNR: (a) Reference; (b) CSTF; (c) LFCTNN; (d) GTNN; (e) CTDF; (f) GNLTN; (g) MHTNN; (h) EDIP-Net; (i) EU2ADL; and (j) THD. For (b)–(j), the top and bottom panels show pseudo-color reconstructions and absolute-error maps, respectively.

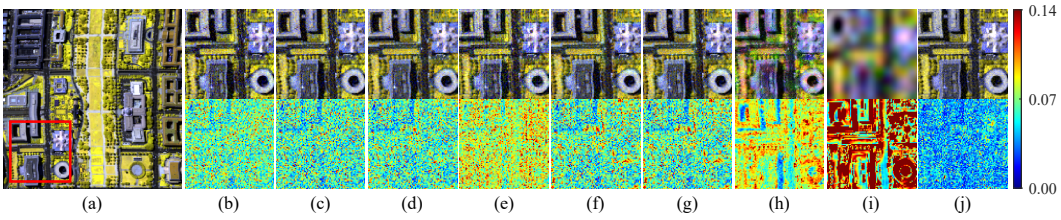


Figure 11: Visual comparison on the WDC scene at $\times 8$ downsampling and 15-dB SNR: (a) Reference; (b) CSTF; (c) LFCTNN; (d) GTNN; (e) CTDF; (f) GNLTN; (g) MHTNN; (h) EDIP-Net; (i) EU2ADL; and (j) THD. For (b)–(j), the top and bottom panels show pseudo-color reconstructions and absolute-error maps, respectively.

D.7 Visual Comparison under $\times 8$ Downsampling

Figures 9, 10, and 11 present the visual comparisons on the original ROSIS, Pavia, and WDC scenes under $\times 8$ spatial downsampling and an SNR of 15 dB. Under this severe degradation, the model-driven methods generally recover the main scene structures, but their enlarged results still contain visible residual noise, oversmoothing, or inaccurate edge reconstruction. The two deep learning methods exhibit larger performance variations across scenes, indicating limited robustness to the strong noise and spatial degradation considered here.

On the original ROSIS scene, CSTF, LFCTNN, GTNN, GNLTN, and MHTNN retain the overall arrangement of the parallel roof structures, but their reconstructions contain noticeable granular noise and local color deviations. CTDF preserves the primary geometry more effectively, although some fine boundaries remain smoothed. EDIP-Net introduces pronounced artifacts around the roof edges, which are clearly visible as concentrated high-error responses, whereas EU2ADL produces relatively blurred textures. In contrast, THD reconstructs sharper parallel lines and more faithful local colors. Its absolute-error map is dominated by lower-error blue regions, with substantially fewer responses around structural boundaries.

Table 5: PSNR decrease (dB) from 20 to 15 dB. Lower is better.

Method	PU	PC	WDC
CSTF	5.2025	5.3522	4.8310
LFCTNN	7.0537	5.8427	5.0883
GTNN	6.9224	5.6925	4.7508
CTDF	6.7565	6.6912	5.2481
GNLTN	7.1245	5.9342	4.7249
MHTNN	7.1603	5.9384	4.7159
THD	**2.9715**	**3.3148**	**3.9787**

On the Pavia scene, the errors of CSTF, LFCTNN, GTNN, GNLTN, and MHTNN are mainly distributed around buildings, roads, and small high-frequency structures. Although CTDF suppresses part of the pixel-level error, its reconstruction tends to smooth local textures and narrow boundaries. EDIP-Net exhibits evident spectral-color distortion and scattered artifacts, while EU2ADL loses part of the fine structural information because of excessive smoothing. THD better preserves building contours, road boundaries, and local material appearances, while producing a more uniform low-error map than the competing methods.

The WDC scene contains more complex roads, building boundaries, and heterogeneous materials. Most model-driven competitors recover the global spatial layout but leave visible noise and reconstruction errors around narrow roads, circular structures, and building edges. The degradation is more apparent for the deep learning methods: EDIP-Net shows strong local color inconsistency, whereas EU2ADL generates severe blurring and loses several fine structures. By comparison, THD preserves clearer geometric boundaries and a pseudo-color appearance closer to the reference image. Its error map contains fewer high-error responses, particularly around the road network and high-contrast structures.

Overall, the visual results agree with the quantitative comparisons. THD consistently suppresses observation noise while retaining fine spatial structures and reliable spectral appearances. This advantage supports the effectiveness of using group-specific peripheral structures to characterize heterogeneous spatial and spectral dependencies, together with a shared waist core to coordinate their information interaction.

D.8 Noise-Robustness Analysis

Table 5 reports the PSNR decrease when the SNR is reduced from 20 to 15 dB. To isolate the behavior of optimization-based models under an identical protocol, this comparison focuses on the model-driven methods. THD exhibits the smallest degradation on all three datasets.

The robustness gain is consistent across scenes rather than being specific to one spectral distribution. Compared with conventional global low-rank constraints, the group-specific peripheral subnetworks avoid forcing spatial and spectral modes into an identical structure. The shared waist core then transfers information between the two groups without discarding their distinct dependencies, limiting noise propagation during coupled reconstruction.

D.9 Computational-Efficiency Analysis

Table 6 reports the runtime averaged over the 15- and 20-dB experiments for each dataset. Only methods whose runtimes were recorded under the same experimental environment are included; the deep-learning methods are omitted from this table because directly comparing runtimes obtained from different implementations or hardware would be misleading.

THD has an average runtime close to GTNN and CSTF, but substantially outperforms both methods in all reconstruction metrics. Although CTDF is faster, its reduced runtime is accompanied by consistently lower accuracy. In contrast, LFCTNN, GNLTN, and MHTNN require considerably more computation while remaining inferior to THD. These results indicate that the shared waist-core mechanism achieves a favorable balance between expressive cross-group interaction and optimization cost.

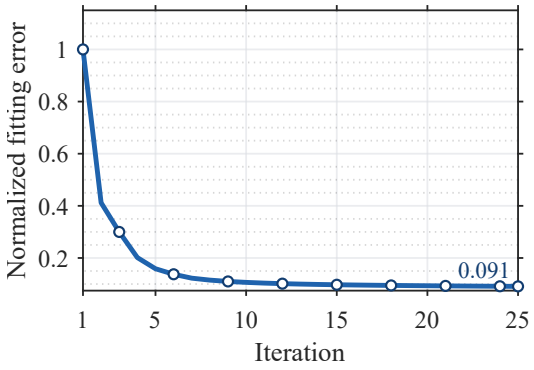


Figure 12: Convergence curve of the proposed algorithm in terms of the normalized fitting error.

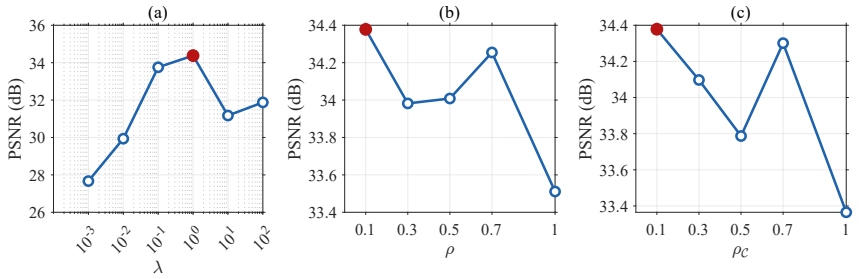


Figure 13: Sensitivity of THD to λ, ρ, and $\rho_{\mathcal{C}}$ in terms of PSNR: (a) λ; (b) ρ; and (c) $\rho_{\mathcal{C}}$. Red markers denote the adopted settings.

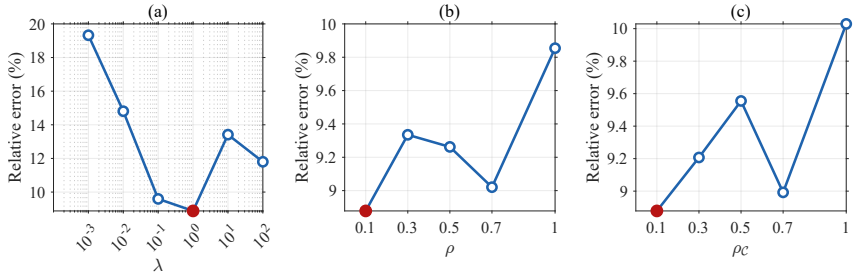


Figure 14: Sensitivity of THD to λ, ρ, and $\rho_{\mathcal{C}}$ in terms of relative reconstruction error: (a) λ; (b) ρ; and (c) $\rho_{\mathcal{C}}$. Red markers denote the adopted settings.

D.10 Convergence Analysis

Fig. 12 illustrates the empirical convergence behavior of the proposed algorithm. The normalized fitting error drops from approximately 1.0 to below 0.2 within five iterations, then decreases gradually and stabilizes after about ten iterations. At the 25th iteration, it reaches 0.091 without noticeable oscillation, indicating stable empirical convergence under the adopted truncated-CG updates.

D.11 Sensitivity to λ, ρ, and $\rho_{\mathcal{C}}$

Figs. 13 and 14 investigate the sensitivity of THD to the observation-balancing parameter λ and the proximal coefficients ρ and $\rho_{\mathcal{C}}$ for the peripheral factors and shared waist core, respectively. The red markers indicate the parameter values adopted in all experiments.

As shown in Fig. 13(a), increasing λ from 10^{-3} to 1 substantially improves PSNR, while a further increase leads to noticeable degradation. The corresponding relative errors in Fig. 14(a) exhibit the opposite trend and attain the minimum at $\lambda = 1$. A very small λ provides insufficient spatial guidance from the HR-MSI, whereas an excessively large value overemphasizes the multispectral

Table 6: Average runtime (seconds) under the two noise levels.

Method	PU	PC	WDC	Average
CSTF	62.39	51.43	53.01	55.61
LFCTNN	86.96	118.42	122.84	109.41
GTNN	48.32	69.57	71.61	63.17
CTDF	19.13	30.11	34.42	27.89
GNLTN	144.75	195.95	192.72	177.81
MHTNN	189.98	242.34	246.34	226.22
THD	45.57	79.88	74.77	66.74

fidelity term and weakens the spectral constraint imposed by the LR-HSI. Therefore, $\lambda = 1$ provides the most favorable balance between spatial-detail injection and spectral preservation.

Figs. 13(b)–(c) and 14(b)–(c) show that THD remains relatively stable for moderate values of ρ and $\rho_{\mathcal{C}}$. Both parameters achieve the highest PSNR and lowest relative error at 0.1, whereas setting them to 1 causes a clear performance drop. This indicates that mild proximal regularization stabilizes the block updates without excessively restricting their adaptation, while large coefficients may overconstrain the factor and core updates. Accordingly, we set $\lambda = 1$, $\rho = 0.1$, and $\rho_{\mathcal{C}} = 0.1$ throughout the experiments.