LEARNING HOW FACTORS INTERACT: CONDITIONED TENSOR PRODUCTS FOR UNSUPERVISED HYPERSPECTRAL SUPER-RESOLUTION

Anonymous authors
Paper under double-blind review

ABSTRACT

Hyperspectral image super-resolution (HSI-SR) aims to reconstruct a high-spatial-resolution hyperspectral image from a low-resolution hyperspectral observation and a registered high-resolution multispectral image (HR-MSI). Existing unsupervised factorized priors typically employ globally fixed factor interactions and discrete band-wise spectral parameterizations, limiting adaptation to spatially varying guidance and ordered spectral variation. To address these limitations, we propose CPF-RFJ, a unified test-time framework that jointly adapts factor interactions and spectral representations. We first propose the Conditioned Tensor Product (CTP), which allows auxiliary observations to select the multiplication law between latent factors, thereby shifting conditioning from feature representations to the algebraic interaction itself. Building upon CTP, the Conditioned Product Factorization (CPF) enables the local HR-MSI observation at each spatial location to govern spectral–spatial factor coupling. We further develop a Pole-Free Rational Fractional-Jacobi Kolmogorov–Arnold Network (PF-RFJ-KAN), which models the spectral factor as a coordinate-defined function through signed fractional spectral reparameterization, an adaptive Jacobi basis family, and a structurally pole-free rational representation. The complete CPF-RFJ model is optimized directly from the observations without external training data or target HR-HSI supervision. Extensive experiments demonstrate state-of-the-art spatial–spectral reconstruction performance against representative training-data-free methods.

1 INTRODUCTION

Hyperspectral images (HSIs) capture a scene across numerous contiguous spectral bands, providing fine-grained material signatures that are unavailable in standard color or multispectral imagery. Such high spectral fidelity, however, often comes at the expense of spatial resolution, acquisition throughput, and signal-to-noise ratio. Hyperspectral image super-resolution (HSI-SR) addresses this trade-off by fusing a low-spatial-resolution hyperspectral image (LR-HSI) with a registered high-spatial-resolution multispectral image (HR-MSI), aiming to recover an HR-HSI with both high spectral and spatial resolutions.

This fusion problem is severely ill-posed because the two observations undergo complementary information loss: spatial blur and downsampling discard fine-scale structures in the LR-HSI, whereas the spectral response function compresses many narrow hyperspectral bands into only a few multispectral channels. Model-based HSI-SR methods mitigate this ambiguity by imposing explicit spectral–spatial structural priors. Representative approaches exploit spectral subspaces and nonlocal low-rankness Yang et al. (2024), nonlinear tensor factorization Zou et al. (2025), recoverable block-term models Ding & Fu (2026), and joint low-rankness and smoothness regularization Zhang et al. (2026). Although these methods provide strong interpretability and effective structural priors, their factorization forms and interaction mechanisms are largely prescribed once the model is specified. Consequently, they do not explicitly adapt the interaction law to spatially varying auxiliary information provided by the registered HR-MSI.

Deep learning substantially enhances the representation capacity of HSI-SR through expressive spectral–spatial mappings and data-driven priors. Recent architectures increasingly combine phys-

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.

ical observation models with deep unfolding, attention, or multilinear tensor structures. For instance, Cao *et al.* Cao et al. (2025) develop a cross-domain-aware deep unfolding Transformer that couples scene-specific prior adaptation with spatial–spectral dependency modeling, while TPTransformer Xu et al. (2026a) introduces tensor–tensor products into Transformer interactions to capture higher-order spectral–spatial correlations. Fang *et al.* Fang et al. (2026) further integrate adaptive multi-scale wavelet priors with physical unfolding for content-aware reconstruction. Despite their strong representation capacity, such supervised models generally rely on external paired training data and priors tied to the training distribution, making their performance sensitive to variations in sensor responses, degradation processes, and target scenes.

Training-data-free and unsupervised test-time approaches avoid this distribution dependence by adapting their models directly to the observed image pair without external supervision. This paradigm has evolved from model-inspired autoencoders (MIAE) Liu et al. (2022) and coupled deep image priors (CS2DIPs) Fang et al. (2024) to enhanced deep representations (EDIP) Li et al. (2025), unsupervised high-order tensor unrolling networks Wang et al. (2026), and coupled diffusion posterior sampling Xu et al. (2026b). Despite substantial progress, two structural limitations remain. First, existing factorized test-time frameworks adapt latent spectral and spatial representations under a predefined and globally shared interaction rule, without explicitly conditioning factor interactions on the registered HR-MSI across spatial locations. Second, the spectral factor is commonly parameterized only at discrete band indices, leaving its ordered and potentially nonuniform variation along the spectral axis implicit. These limitations motivate a framework that jointly adapts factor interaction laws and spectral representations.

To overcome these limitations, we propose *CPF-RFJ*, a unified training-data-free framework that adapts factorized priors at both the interaction and spectral-representation levels. We first propose the *Conditioned Tensor Product* (CTP), a general formulation in which side information selects the multiplication law governing two tensor operands, thereby shifting conditioning from the operands to the algebraic interaction itself. Building upon CTP, the *Conditioned Product Factorization* (CPF) enables the registered HR-MSI at each spatial location to govern spectral–spatial factor coupling during reconstruction.

To model the spectral factor beyond discrete band indices, we develop a *Pole-Free Rational Fractional-Jacobi Kolmogorov–Arnold Network*. Rather than learning spectral values under a fixed coordinate system and basis, PF-RFJ-KAN represents the spectral factor as a coordinate-defined function by jointly adapting the fractional reparameterization of the spectral coordinate, the Jacobi basis family, and a structurally pole-free rational function class. The resulting coordinate function provides a structured residual correction to the deep-prior spectral factor, which is then coupled with the spatial factor through CPF. The complete CPF-RFJ framework is optimized directly from the observed LR-HSI and HR-MSI without external training data or target HR-HSI supervision.

The main contributions of this paper are summarized as follows:

- We propose the *Conditioned Tensor Product* (CTP), a general formulation in which side information selects the multiplication law between latent tensor factors, and establish its fundamental algebraic and stability properties.
- Building upon CTP, we propose the *Conditioned Product Factorization* (CPF), which combines conditioned tensor multiplication with block-trace contraction to enable spatially varying spectral–spatial factor interactions conditioned on local HR-MSI observations.
- We develop *PF-RFJ-KAN*, which represents the spectral factor as a coordinate-defined function by jointly adapting fractional spectral reparameterization, the Jacobi basis family, and a structurally pole-free rational function class, and incorporates it as a structured residual correction to the spectral factor.

2 Conditioned Tensor Product and Factorization

Conventional tensor factorizations employ a globally prescribed multiplication law to combine latent factors. As illustrated in Fig. 1, ordinary matrix multiplication and fixed transform-induced tensor products apply a globally shared interaction law, whereas the proposed CTP allows auxiliary information to select the multiplication law itself. We first introduce the *Conditioned Tensor Prod-*

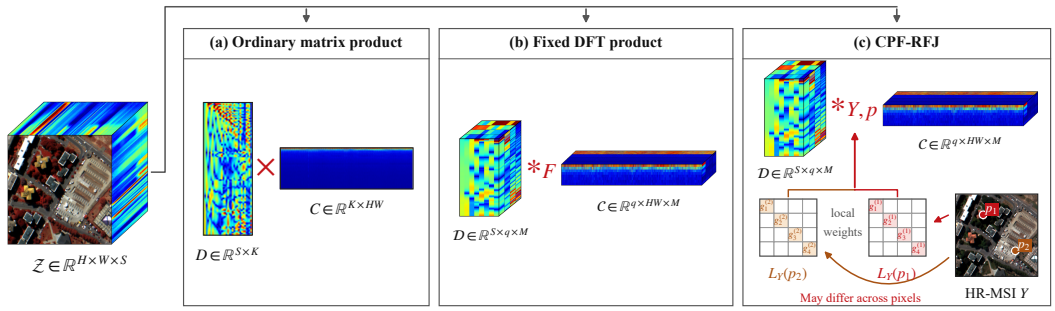


Figure 1: Comparison of different tensor multiplication schemes.

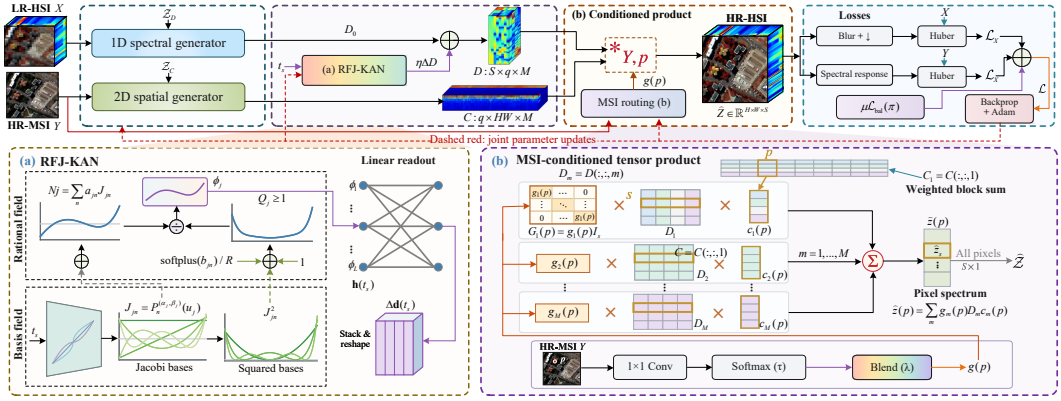


Figure 2: Overview of the proposed CPF-RFJ framework.

uct (CTP), then derive the *Conditioned Product Factorization* (CPF), and finally instantiate it for HSI-SR.

2.1 Conditioned Tensor Product

Let $\mathcal{A} \in \mathbb{F}^{n_1\times n_2\times M}$ and $\mathcal{B} \in \mathbb{F}^{n_2\times n_3\times M}$, where $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$, with frontal blocks $A^{(m)} = \mathcal{A}(:,:,m)$ and $B^{(m)} = \mathcal{B}(:,:,m)$ for $m = 1,\dots,M$. We denote facewise matrix multiplication by $\triangle$, i.e., $[\mathcal{A}\triangle\mathcal{B}]^{(m)} = A^{(m)}B^{(m)}$. For auxiliary conditioning information $\mathbf{c} \in \mathcal{C}$, let $L_{\mathbf{c}} = \Psi_\vartheta(\mathbf{c}) \in \mathrm{GL}(M, \mathbb{F})$ and define the block-mode transformation $\Phi_{\mathbf{c}}(\mathcal{A}) = \mathcal{A} \times_3 L_{\mathbf{c}}$.

Definition 2.1 (Conditioned Tensor Product)**.** For fixed $\mathbf{c}$, the *Conditioned Tensor Product* $\mathcal{A} \star_{\mathbf{c}} \mathcal{B}$ is the unique tensor satisfying

$$\Phi_{\mathbf{c}}(\mathcal{A} \star_{\mathbf{c}} \mathcal{B}) = \Phi_{\mathbf{c}}(\mathcal{A})\triangle\Phi_{\mathbf{c}}(\mathcal{B}). \tag{1}$$

Since $\Phi_{\mathbf{c}}$ is a linear bijection, $\mathcal{A}\star_{\mathbf{c}}\mathcal{B}$ is uniquely determined. Importantly, $\mathbf{c}$ is not an additional multiplicand; rather, it selects the block-mode coordinates and hence the multiplication law governing the interaction between the two operands.

Theorem 2.2 (Algebraic properties)**.** *For fixed* $\mathbf{c}$*,* $\star_{\mathbf{c}}$ *is bilinear and associative for all conformable tensors, with identity* $\mathcal{E}_{\mathbf{c}} = \Phi_{\mathbf{c}}^{-1}(\mathcal{I})$*, where* $\mathcal{I} = (I,\dots,I)$*. Moreover,* $\mathcal{A}$ *is* $\star_{\mathbf{c}}$*-invertible if and only if every frontal block of* $\Phi_{\mathbf{c}}(\mathcal{A})$ *is nonsingular.*

Associativity assumes that the same conditioning state is preserved throughout nested multiplications.

Theorem 2.3 (Stability)**.** *Under the block Frobenius norm* $\|\mathcal{A}\|_{\mathrm{bF}}^2 = \sum_{m=1}^{M}\|A^{(m)}\|_F^2$*,*

$$\|\mathcal{A}\star_{\mathbf{c}}\mathcal{B}\|_{\mathrm{bF}} \le \|L_{\mathbf{c}}^{-1}\|_2\|L_{\mathbf{c}}\|_2^2\|\mathcal{A}\|_{\mathrm{bF}}\|\mathcal{B}\|_{\mathrm{bF}}. \tag{2}$$

Explicit inverse constructions and proofs of Theorems 2.2 and 2.3 are provided in Appendix A.

Relation to fixed tensor products. CTP reduces to facewise multiplication when $L_{\mathbf{c}} = I_M$, and to a fixed transform-induced product when $L_{\mathbf{c}} \equiv L$ is independent of $\mathbf{c}$. A fixed discrete Fourier transform further recovers the classical t-product up to normalization Kilmer & Martin (2011); Kernfeld et al. (2015). Thus, CTP differs from fixed products by allowing side information to select the multiplication law itself.

2.2 Conditioned Product Factorization

For efficient factorization, we consider the positive-diagonal realization $L_{\mathbf{c}}^{+} = \mathrm{Diag}(\mathbf{g}(\mathbf{c}))$, where $\mathbf{g}(\mathbf{c}) \in \mathbb{R}_{++}^{M}$. Definition 2.1 then reduces to

$$[\mathcal{A} \star_{\mathbf{c}}^{+} \mathcal{B}]^{(m)} = g_m(\mathbf{c}) A^{(m)} B^{(m)}. \tag{3}$$

For complementary blocks $A^{(m)} \in \mathbb{F}^{r_a \times r_b}$ and $B^{(m)} \in \mathbb{F}^{r_b \times r_a}$, let $\mathrm{btr}(\mathcal{T}) = \sum_{m=1}^{M} \mathrm{tr}(T^{(m)})$ denote the block-trace contraction.

Definition 2.4 (Conditioned Product Factorization)**.** The *Conditioned Product Factorization* (CPF) is defined as

$$\mathrm{CPF}_{\mathbf{c}}(\mathcal{A}, \mathcal{B}) := \mathrm{btr}(\mathcal{A} \star_{\mathbf{c}}^{+} \mathcal{B}) = \sum_{m=1}^{M} g_m(\mathbf{c}) \, \mathrm{tr}\left(A^{(m)} B^{(m)}\right). \tag{4}$$

Thus, CTP governs how corresponding latent blocks interact, while block-trace contraction closes their complementary latent matrix dimensions.

Residual simplex parameterization. We parameterize the positive coefficients as $\boldsymbol{\pi}(\mathbf{c}) = \mathrm{softmax}(q_\vartheta(\mathbf{c})/\tau)$ and $\mathbf{g}(\mathbf{c}) = (1-\rho)\mathbf{1} + \rho M \boldsymbol{\pi}(\mathbf{c})$, where $\tau > 0$ and $0 < \rho < 1$. This gives $\mathbf{1}^\top \mathbf{g}(\mathbf{c}) = M$ and $1 - \rho < g_m(\mathbf{c}) < 1 - \rho + \rho M$. At the simplex barycenter $\boldsymbol{\pi}(\mathbf{c}) = \mathbf{1}/M$, we obtain $\mathbf{g}(\mathbf{c}) = \mathbf{1}$, exactly recovering facewise multiplication.

Proposition 2.5 (Simplex-induced product deviation)**.** *Let $\mathfrak{B}_{\mathbf{c}}$ denote the bilinear operator induced by Eq. equation 3. Then*

$$\|\mathfrak{B}_{\mathbf{c}} - \triangle\|_{\mathrm{bil}} = \|\mathbf{g}(\mathbf{c}) - \mathbf{1}\|_\infty = \rho M \left\| \boldsymbol{\pi}(\mathbf{c}) - \frac{\mathbf{1}}{M} \right\|_\infty, \tag{5}$$

where $\|\cdot\|_{\mathrm{bil}}$ denotes the induced bilinear operator norm defined in Appendix A.

Hence, displacement from the simplex barycenter directly quantifies the deviation of the conditioned multiplication law from facewise multiplication. The proof is provided in Appendix A.

2.3 CPF for Hyperspectral Super-Resolution

Let $\mathcal{X} \in \mathbb{R}^{S \times h \times w}$, $\mathcal{Y} \in \mathbb{R}^{s \times H \times W}$, and $\widehat{\mathcal{Z}} \in \mathbb{R}^{S \times H \times W}$ denote the LR-HSI, registered HR-MSI, and reconstructed HR-HSI, respectively. At each spatial location $p = (i, j)$, the local multispectral vector serves as the conditioning state $\mathbf{c}_p = \mathcal{Y}(:, i, j)$. A shared linear projection $q_\vartheta(\mathbf{c}_p) = W_g \mathbf{c}_p + b_g$ then produces the spatially adaptive coefficients $\mathbf{g}_p$ through the residual simplex parameterization.

We represent the latent scene using M pairs of spectral and spatial factors $\mathcal{G}^{(m)} \in \mathbb{R}^{r_a \times r_b \times S}$ and $\mathcal{H}^{(m)} \in \mathbb{R}^{r_b \times r_a \times H \times W}$. For spectral band k and spatial location $p = (i, j)$, let $G_k^{(m)} = \mathcal{G}^{(m)}(:, :, k)$ and $H_p^{(m)} = \mathcal{H}^{(m)}(:, :, i, j)$. Stacking these blocks over m forms $\mathcal{G}_k \in \mathbb{R}^{r_a \times r_b \times M}$ and $\mathcal{H}_p \in \mathbb{R}^{r_b \times r_a \times M}$. CPF reconstructs each spectral–spatial entry as

$$\widehat{\mathcal{Z}}_{k,p} = \mathrm{CPF}_{\mathbf{c}_p}(\mathcal{G}_k, \mathcal{H}_p) = \sum_{m=1}^{M} g_m(\mathbf{c}_p) \, \mathrm{tr}\left(G_k^{(m)} H_p^{(m)}\right). \tag{6}$$

The coefficients $\mathbf{g}_p$ are shared across all hyperspectral bands at spatial location p. Consequently, the HR-MSI does not simply reweight an existing output; it determines the local multiplication law under which the spectral and spatial latent factors interact inside CPF. The parameterization of these factors is introduced next.

3 PF-RFJ-ENHANCED DEEP-PRIOR FACTORIZATION

CPF specifies how spectral and spatial latent factors interact, but does not prescribe how these factors are parameterized. The complete CPF-RFJ pipeline is illustrated in Fig. 2. We generate the latent factors using coupled test-time deep priors and further introduce a *Pole-Free Rational Fractional-Jacobi Kolmogorov–Arnold Network* (PF-RFJ-KAN) to model the spectral factor as a continuous coordinate-defined function. The deep priors provide data-adaptive base factors, whereas PF-RFJ-KAN introduces a structured correction along the ordered spectral axis before factor coupling through CPF.

3.1 DEEP-PRIOR FACTOR GENERATION

We employ complementary one- and two-dimensional deep priors to generate the base spectral and spatial factors:

$$\left\{\mathcal{G}_{\mathrm{DIP}}^{(m)}\right\}_{m=1}^{M} = f_{\theta_s}(\mathcal{X}, \boldsymbol{\xi}_s), \qquad \left\{\mathcal{H}^{(m)}\right\}_{m=1}^{M} = f_{\theta_p}(\mathcal{Y}, \boldsymbol{\xi}_p), \tag{7}$$

where $\boldsymbol{\xi}_s$ and $\boldsymbol{\xi}_p$ are learnable latent inputs, $\mathcal{G}_{\mathrm{DIP}}^{(m)} \in \mathbb{R}^{r_a \times r_b \times S}$, and $\mathcal{H}^{(m)} \in \mathbb{R}^{r_b \times r_a \times H \times W}$. As illustrated in Fig. 2, the one-dimensional branch generates the base spectral factors from the LR-HSI, while the two-dimensional branch extracts high-resolution spatial factors from the HR-MSI. The spatial factors are further constrained to be positive and normalized at each spatial location; the explicit normalization is provided in Appendix C.

Although the one-dimensional deep prior provides an expressive base spectral representation, its output remains parameterized at discrete band indices. We therefore augment it with a coordinate-defined PF-RFJ correction that explicitly models functional variation along the spectral axis.

3.2 PF-RFJ-KAN SPECTRAL COORDINATE FUNCTION

KANs parameterize learnable univariate functions on network edges, with the original formulation using B-spline bases (Liu et al., 2025). Subsequent variants explore fractional Jacobi bases and rational function parameterizations (Aghaei et al., 2026; Aghaei, 2025). Motivated by these developments, PF-RFJ-KAN represents the spectral factor as a continuous function over normalized band coordinates, while jointly adapting the spectral coordinate, Jacobi basis family, and a structurally pole-free rational function class. Its construction is summarized in Fig. 2(a). Conceptually, each PF-RFJ atom takes the form

$$\phi_j(t) = R_{\alpha_j,\beta_j}\big(u_{\gamma_j}(t)\big), \tag{8}$$

where u_{γ_j} reparameterizes the spectral coordinate and R_{α_j,β_j} denotes a pole-free rational expansion in an adaptive Jacobi function space. PF-RFJ-KAN therefore jointly adapts the spectral coordinate parameterization, basis family, and rational function class.

For the k-th hyperspectral band, we define the normalized spectral coordinate $t_k = -1 + 2(k-1)/(S-1) \in [-1, 1]$. For atom j, this coordinate is reparameterized through the learnable signed fractional map

$$u_j(t) = \mathrm{sgn}(t)|t|^{\gamma_j}, \qquad \gamma_j > 0, \tag{9}$$

with $u_j(0) = 0$. For every admissible γ_j, u_j is a monotone bijection from $[-1, 1]$ onto itself. It therefore defines a learnable reparameterization of the spectral domain, allowing different atoms to represent the same ordered spectral axis under distinct coordinate geometries.

Adaptive Jacobi function space. Given $u_j(t)$, we employ the Jacobi polynomial family $\{P_n^{(\alpha_j,\beta_j)}\}_{n\geq 0}$ with $\alpha_j, \beta_j > -1$. Its associated weight function is $w_{\alpha_j,\beta_j}(u) = (1-u)^{\alpha_j}(1+u)^{\beta_j}$, under which

$$\int_{-1}^{1} P_n^{(\alpha_j,\beta_j)}(u) P_m^{(\alpha_j,\beta_j)}(u) w_{\alpha_j,\beta_j}(u)\, du = 0, \qquad n \neq m. \tag{10}$$

Hence, learning (α_j, β_j) adapts not merely the expansion coefficients but the weighted orthogonal function space itself. The three-term recurrence used for basis evaluation is provided in Appendix B.

Pole-free rational function class. For numerator and denominator degrees d_N and d_D, respectively, we define

$$N_j(u) = \sum_{n=0}^{d_N} c_{j,n} P_n^{(\alpha_j,\beta_j)}(u),$$
$$D_j(u) = 1 + \frac{1}{d_D} \sum_{n=1}^{d_D} \mathrm{Softplus}(d_{j,n}) \left[P_n^{(\alpha_j,\beta_j)}(u)\right]^2. \tag{11}$$

Let $\mathbf{p}_j(u) = [P_1^{(\alpha_j,\beta_j)}(u), \ldots, P_{d_D}^{(\alpha_j,\beta_j)}(u)]^\top$ and

$$\mathbf{\Lambda}_j = \frac{1}{d_D} \mathrm{Diag}\left(\mathrm{Softplus}(d_{j,1}), \ldots, \mathrm{Softplus}(d_{j,d_D})\right) \succeq 0.$$

Then $D_j(u) = 1 + \mathbf{p}_j(u)^\top \mathbf{\Lambda}_j \mathbf{p}_j(u) \geq 1$. Denominator positivity is therefore imposed structurally on the function class rather than through post-hoc clipping.

The resulting PF-RFJ atom is $\phi_j(t) = N_j(u_j(t))/D_j(u_j(t))$. Stacking J atoms as $\boldsymbol{\phi}(t) = [\phi_1(t), \ldots, \phi_J(t)]^\top$, the PF-RFJ-KAN output is

$$\boldsymbol{\delta}(t) = W_o \boldsymbol{\phi}(t) + \mathbf{b}_o \in \mathbb{R}^{M r_a r_b}. \tag{12}$$

Theorem 3.1 (Pole freedom and uniform boundedness)**.** *For atom j, define $B_{j,n} = \max_{u\in[-1,1]} |P_n^{(\alpha_j,\beta_j)}(u)|$ and $C_j = \sum_{n=0}^{d_N} |c_{j,n}| B_{j,n}$. Then, for every $t \in [-1, 1]$,*

$$D_j(u_j(t)) \geq 1, \qquad |\phi_j(t)| \leq C_j. \tag{13}$$

Consequently, every PF-RFJ atom is uniformly bounded over the modeled spectral interval and contains no real poles.

The proof of Theorem 3.1, together with fractional-coordinate regularity, parameter-gradient bounds, and approximation-density results, is provided in Appendix B.

3.3 PF-RFJ Spectral Factor Correction

Evaluating Eq. equation 12 at all normalized coordinates $\{t_k\}_{k=1}^S$ and reshaping the outputs yields $\{\Delta\mathcal{G}^{(m)}\}_{m=1}^M$, where

$$\Delta\mathcal{G}^{(m)} \in \mathbb{R}^{r_a \times r_b \times S}.$$

Rather than replacing the deep-prior spectral factor, PF-RFJ-KAN provides a structured residual correction:

$$\widetilde{\mathcal{G}}^{(m)} = \mathcal{G}_{\mathrm{DIP}}^{(m)} + \lambda \Delta\mathcal{G}^{(m)}, \qquad m = 1, \ldots, M, \tag{14}$$

where $\lambda > 0$ controls the correction scale. This residual blending step is illustrated in Fig. 2(a), where the PF-RFJ output is reshaped and added to the deep-prior spectral factor before CPF coupling.

For band k, let $\widetilde{G}_k^{(m)} = G_{\mathrm{DIP},k}^{(m)} + \lambda \Delta G_k^{(m)}$, and stack these blocks as $\widetilde{\mathcal{G}}_k$. Substituting the corrected spectral factor into CPF yields the complete HR-HSI reconstruction:

$$\widehat{\mathcal{Z}}_{k,p} = \mathrm{CPF}_{\mathbf{c}_p}\left(\widetilde{\mathcal{G}}_k, \mathcal{H}_p\right)$$
$$= \sum_{m=1}^{M} g_m(\mathbf{c}_p)\, \mathrm{tr}\left[\left(G_{\mathrm{DIP},k}^{(m)} + \lambda \Delta G_k^{(m)}\right) H_p^{(m)}\right]. \tag{15}$$

As illustrated in Fig. 2(b), the local HR-MSI observation produces the conditioning coefficients $\mathbf{g}_p$, which modulate the corresponding spectral–spatial block interactions before their aggregation into the reconstructed spectrum. The deep priors therefore establish data-adaptive base factors, PF-RFJ-KAN introduces coordinate-dependent spectral corrections, and CPF determines their observation-conditioned interaction law. A formal bound on the reconstruction perturbation induced by the PF-RFJ correction is provided in Appendix B.

Table 1: Quantitative comparison on three datasets under $\times 8$ and $\times 16$ spatial degradation. The best and second-best comparable results are highlighted in **red** and **blue**, respectively.

Method	PaviaC				PC				PU			
	PSNR↑	SSIM↑	SAM↓	ERGAS↓	PSNR↑	SSIM↑	SAM↓	ERGAS↓	PSNR↑	SSIM↑	SAM↓	ERGAS↓
$\times 8$												
MIAE_{22}	44.8876	**0.9869**	3.6439	0.7534	**47.1767**	**0.9977**	**1.8252**	**0.3106**	43.6221	0.9922	1.7920	0.5345
CS2DIPs_{24}	45.2687	0.9868	3.6585	0.7495	46.3360	0.9970	2.0510	0.3611	44.8158	0.9930	1.6304	0.4584
M2U-Net_{24}	40.5692	0.9798	4.1738	1.0019	40.3893	0.9901	2.6962	0.6765	41.3909	0.9890	1.9067	0.6572
EDIP-Net_{25}	41.1756	0.9811	4.1068	0.9555	42.4898	0.9944	2.3480	0.5204	41.3081	0.9887	1.9043	0.6564
SpectraLift_{25}	43.0082	0.9860	3.6290	0.8552	44.9947	0.9972	2.0230	0.3914	42.5493	0.9917	1.7472	0.5730
ARGS-Diff_{25}	44.4068	0.9852	4.1396	0.8045	44.4784	0.9955	2.6166	0.4414	43.5731	0.9910	1.8634	0.5297
CLoRF_{25}	44.0148	0.9846	3.9129	0.7974	44.6705	0.9959	2.4572	0.4148	44.1256	0.9916	1.7771	0.4982
CDPS_{26}	39.6074	0.9727	5.9322	1.2407	39.7363	0.9871	3.7697	0.7733	40.8845	0.9858	2.4065	0.7329
UCL_{26}	**45.4037**	**0.9869**	**3.5575**	**0.7347**	46.9320	0.9974	1.8827	0.3208	**45.8584**	**0.9943**	**1.4040**	**0.4155**
CPF-RFJ	**46.1971**	**0.9878**	**3.3989**	**0.6991**	**48.4836**	**0.9982**	**1.6337**	**0.2649**	**46.0958**	**0.9942**	**1.4153**	**0.3986**
$\times 16$												
MIAE_{22}	**44.8796**	**0.9868**	**3.7204**	**0.3821**	**46.2741**	**0.9974**	**2.0037**	**0.1734**	43.4073	**0.9922**	1.8805	0.2723
CS2DIPs_{24}	43.4632	0.9848	4.0512	0.4446	45.1068	0.9966	2.1448	0.2077	43.2510	0.9906	1.9817	0.3064
M2U-Net_{24}	40.5861	0.9786	4.4342	0.5052	41.1410	0.9913	2.7115	0.3116	40.9973	0.9879	2.0616	0.3454
EDIP-Net_{25}	40.8763	0.9796	4.3764	0.4945	41.8381	0.9932	2.6368	0.2757	40.8625	0.9871	2.0844	0.3453
SpectraLift_{25}	40.6931	0.9834	3.8881	0.5034	40.2180	0.9944	2.2712	0.3603	39.8119	0.9884	2.2476	0.4061
ARGS-Diff_{25}	43.2219	0.9832	4.8581	0.4478	44.3563	0.9956	2.6587	0.2196	42.7912	0.9903	2.0496	0.2919
CLoRF_{25}	43.6393	0.9839	4.0008	0.4130	44.6620	0.9959	2.4663	0.2090	**43.7426**	0.9910	**1.8580**	**0.2608**
CDPS_{26}	32.6072	0.9111	9.0010	1.1029	30.0204	0.9223	7.7931	1.0750	29.4108	0.8336	6.3607	1.2199
UCL_{26}	44.6746	0.9858	3.8006	0.3960	45.4191	0.9966	2.0192	0.1977	44.1578	0.9926	1.7559	0.2725
CPF-RFJ	**45.6395**	**0.9874**	**3.5929**	**0.3611**	**47.3300**	**0.9978**	**1.8385**	**0.1520**	**45.3455**	**0.9935**	**1.5520**	**0.2179**

3.4 Observation-Consistent Test-Time Optimization

The complete CPF-RFJ model is optimized directly from the observed LR-HSI and HR-MSI without external training data or target HR-HSI supervision. Let $\mathcal{P}_{\text{spe}}$ denote the spectral response operator and $\mathcal{D}_{\text{spa}}$ the spatial blur-and-downsampling operator. With Θ collecting all trainable variables, we solve

$$\min_{\Theta} \ \mathcal{L}_\delta \left(\mathcal{Y} - \mathcal{P}_{\text{spe}}(\widehat{\mathcal{Z}}_\Theta) \right) + \mathcal{L}_\delta \left(\mathcal{X} - \mathcal{D}_{\text{spa}}(\widehat{\mathcal{Z}}_\Theta) \right), \tag{16}$$

where $\mathcal{L}_\delta$ denotes the element-wise Huber loss with threshold $\delta = 0.1$. The two observation-consistency paths and the joint parameter updates are summarized in Fig. 2.

All model parameters and learnable latent inputs are jointly optimized using Adam, with latent-input perturbation, scheduled optimization, and clean-input checkpoint selection. Ground-truth HR-HSI is used only for evaluation. Further implementation details are provided in Appendix C.

4 Experiments

4.1 Comparison Methods and Evaluation Protocol

We evaluate CPF-RFJ on Pavia Centre (PaviaC), Pavia City Center (PC), and Pavia University (PU) under $\times 8$ and $\times 16$ spatial degradations. Nine representative training-data-free methods are compared: MIAE (Liu et al., 2022), M2U-Net (Li et al., 2024), EDIP-Net (Li et al., 2025), CS2DIPs (Fang et al., 2024), SpectraLift (Shah & Duarte, 2025), CLoRF (Wang et al., 2025), ARGS-Diff (Zhu et al., 2025), CDPS (Xu et al., 2026c), and UCL (Xiao et al., 2026). All methods use the same observations, and the reference HR-HSI is used only for evaluation. Full experimental details are provided in Appendix C.

Quantitative comparison. Table 1 reports the quantitative results under $\times 8$ and $\times 16$ spatial degradation. CPF-RFJ achieves the best performance in 22 out of the 24 dataset–metric–scale combinations and ranks second in the remaining two, demonstrating consistently strong reconstruction performance across different scenes and degradation levels. Under $\times 8$, CPF-RFJ achieves the best results on all four metrics for both PaviaC and PC. In particular, it improves the PSNR over the second-best method by 0.79 dB on PaviaC and 1.31 dB on PC. On PU, CPF-RFJ obtains the highest PSNR and the lowest ERGAS, while its SSIM and SAM are only marginally behind UCL (0.9942

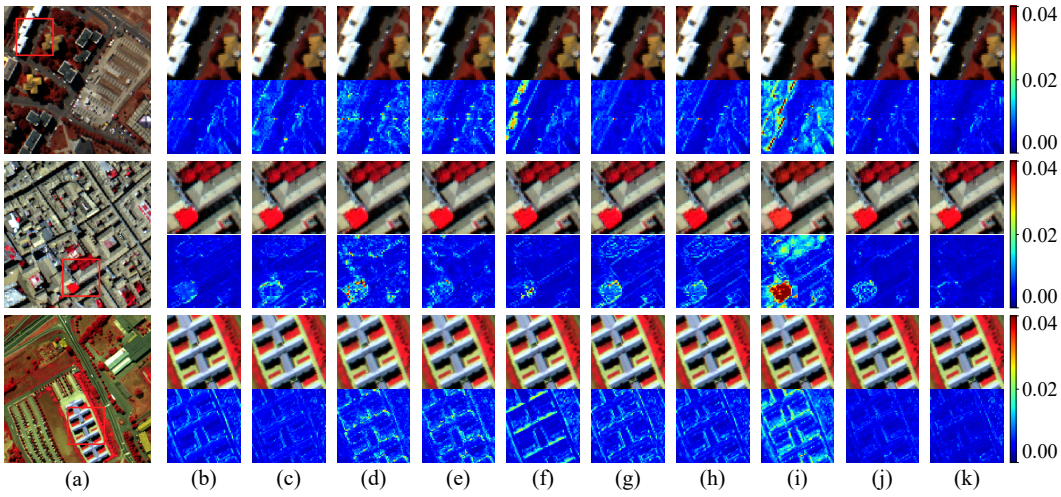


Figure 3: Qualitative comparison on three representative scenes under $\times 8$ spatial degradation. (a) Reference, (b) MIAE, (c) CS2DIPs, (d) M2U-Net, (e) EDIP-Net, (f) SpectraLift, (g) ARGS-Diff, (h) CLoRF, (i) CDPS, (j) UCL, and (k) CPF-RFJ. For each reconstruction method, the upper panel presents the reconstructed region and the lower panel shows the corresponding absolute error map.

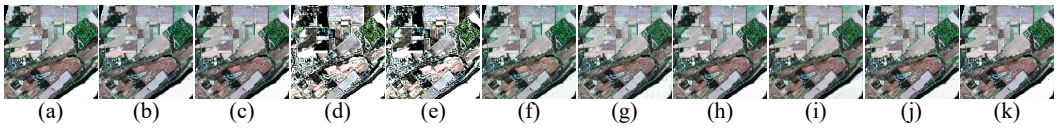


Figure 4: Qualitative comparison on the real Hyperion–Sentinel observation pair. (a) HR-MSI, (b) MIAE, (c) CS2DIPs, (d) M2U-Net, (e) EDIP-Net, (f) SpectraLift, (g) ARGS-Diff, (h) CLoRF, (i) CDPS, (j) UCL, and (k) CPF-RFJ.

vs. 0.9943 and 1.4153 vs. 1.4040, respectively). The advantage becomes more pronounced under the more challenging $\times 16$ setting, where CPF-RFJ ranks first on all twelve dataset–metric combinations. Compared with the second-best PSNR, CPF-RFJ gains 0.76 dB, 1.06 dB, and 1.19 dB on PaviaC, PC, and PU, respectively. Meanwhile, it consistently yields higher SSIM and lower SAM and ERGAS across all three datasets. These results indicate that CPF-RFJ not only provides high reconstruction accuracy under moderate degradation but also maintains a clear advantage as the spatial degradation becomes more severe.

Qualitative comparison. Fig. 3 presents qualitative comparisons on three representative scenes under $\times 8$ spatial degradation. The enlarged regions correspond to the red boxes in the reference images, and all error maps share the same color scale. Most competing methods recover the main structures but exhibit noticeable residuals around high-contrast boundaries, small objects, and elongated structures. In contrast, CPF-RFJ produces reconstructions closer to the reference, with sharper structural boundaries and fewer visible artifacts. Its error maps are predominantly dark blue, particularly around building edges and other high-frequency regions, demonstrating improved spatial detail preservation while maintaining overall reconstruction fidelity.

Real-data experiment. Fig. 4 presents the qualitative comparison on the real Hyperion–Sentinel observation pair. Since no ground-truth HR-HSI is available for this real scene, we focus on visual spatial fidelity and structural consistency. Several competing methods introduce noticeable local distortions, over-enhanced boundaries, or unnatural artifacts, particularly around buildings, roads, and vegetation regions. In contrast, CPF-RFJ produces more coherent spatial structures with sharper yet natural boundaries and fewer visible artifacts, demonstrating its applicability to real observations beyond simulated degradations. For additional quantitative assessment without ground-truth HR-

HSI, we report the MSI observation-consistency error in Appendix C.5, where CPF-RFJ achieves the lowest error among the compared methods.

5 Ablation and Analysis

Ablation on factor-interaction operators. We compare CPF with two fixed interaction rules, namely ordinary matrix multiplication and the classical t-product, while keeping the backbone, PF-RFJ correction, and optimization settings unchanged. As shown in Table 2, replacing matrix multiplication with the t-product does not yield a consistent gain, suggesting that a fixed transform-domain interaction alone is insufficient. In contrast, CPF achieves the highest PSNR under both $\times 8$ and $\times 16$, improving over the best fixed alternative by 0.17 dB and 0.01 dB, respectively. It also matches the best SSIM at $\times 8$ and achieves the highest SSIM at $\times 16$. These results support the advantage of adapting factor interactions to local multispectral observations rather than using a globally fixed multiplication rule.

Table 2: Ablation of factor-interaction operators on Pavia Centre.

Interaction	PaviaC $\times 8$		PaviaC $\times 16$	
formulation	PSNR↑	SSIM↑	PSNR↑	SSIM↑
Matrix product	46.0541	0.9877	45.6759	0.9873
t-product	46.0429	0.9878	45.4853	0.9872
CPF (ours)	**46.2250**	**0.9878**	**45.6851**	**0.9875**

Spatial and spectral error comparison. Fig. 5 evaluates different spectral-factor parameterizations from both spatial and spectral viewpoints. In Fig. 5(a), CPF-RFJ yields the lowest spatial reconstruction error on the selected ROI, with an MAE of 5.901×10^{-3}, improving upon B-spline KAN and Padé-rKAN by 6.2% and 10.3%, respectively. In Fig. 5(b), CPF-RFJ also shows the strongest spectral performance, achieving the smallest average excess error (0.96%) and the largest number of best-performing bands (55 out of 93). This demonstrates that CPF-RFJ improves reconstruction quality not only in local spatial details but also across a wide range of spectral bands.

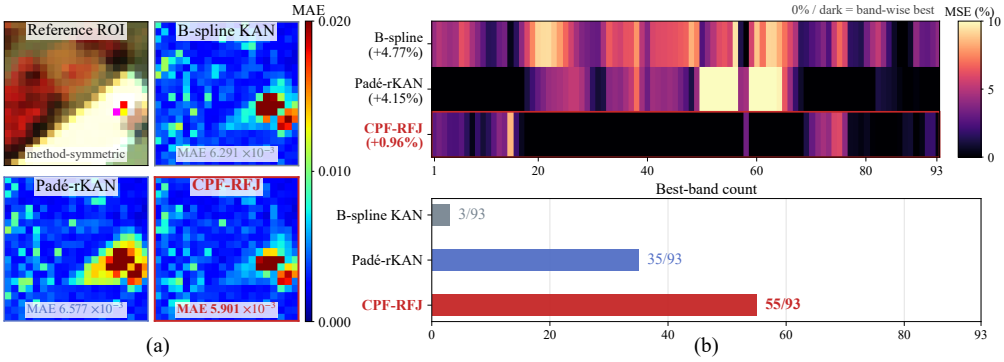


Figure 5: Comparison of different spectral-factor parameterizations in terms of spatial and spectral reconstruction errors. (a) Spatial error maps on a representative ROI. (b) Band-wise spectral error comparison, including the relative MSE heatmap and the number of best-performing bands.

6 Conclusion

We presented CPF-RFJ, an unsupervised test-time framework for hyperspectral image super-resolution that revisits both how latent factors interact and how spectral factors are represented. The proposed Conditioned Tensor Product allows side information to select the multiplication law between tensor factors, from which the Conditioned Product Factorization provides an efficient observation-dependent spectral–spatial factorization. We further introduced PF-RFJ-KAN, which represents the spectral factor through a jointly adapted coordinate parameterization, Jacobi function space, and structurally pole-free rational function class. Experiments under $\times 8$ and $\times 16$ spatial degradations demonstrate consistent improvements over training-data-free baselines, while targeted analyses validate both the conditioned factor interaction and the coordinate-function spectral representation. These results suggest that adapting not only latent representations but also their interaction law provides a promising direction for structured inverse problems beyond HSI-SR.

AI USE STATEMENT

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

REPRODUCIBILITY STATEMENT

We provide the formulation and theoretical properties of CTP and CPF in Section 2 and Appendix B, and the construction and additional analyses of PF-RFJ-KAN in Section 3 and Appendix C. Appendix D reports the implementation configuration, parameter initialization, computational statistics, hyperparameter-sensitivity analysis, spatial-factor normalization, optimization schedules, and checkpoint-selection procedure. The experimental section specifies the observation and evaluation protocol used for all compared methods. These details are provided to facilitate independent reproduction of the reported results.

REFERENCES

Alireza Afzal Aghaei. fkan: Fractional kolmogorov–arnold networks with trainable jacobi basis functions. *Neurocomputing*, 623:129414, 2025.

Alireza Afzal Aghaei, Mehdi Hosseinzadeh, and Kourosh Parand. rkan: Rational kolmogorov-arnold networks. *Neural Networks*, pp. 108888, 2026.

Xuheng Cao, Xuquan Wang, Xiong Dun, Yusheng Lian, Xinbin Cheng, and Xiaopeng Hao. Cross-domain-aware deep unfolding transformer for hyperspectral image super-resolution. *Pattern Recognition*, pp. 112374, 2025.

Meng Ding and Xiao Fu. Rethinking coupled tensor analysis for hyperspectral superresolution: Recoverable modeling under endmember variability. *SIAM Journal on Imaging Sciences*, 19(2): 752–781, 2026.

Yuan Fang, Yipeng Liu, Chong-Yung Chi, Zhen Long, and Ce Zhu. Cs2dips: Unsupervised hsi super-resolution using coupled spatial and spectral dips. *IEEE Transactions on Image Processing*, 33:3090–3101, 2024.

Yuan Fang, Yipeng Liu, Zhen Long, Chong-Yung Chi, and Ce Zhu. Content-adaptive unfolding wavelet transformer for hyperspectral image super-resolution. *IEEE Transactions on Image Processing*, 35:1923–1936, 2026. doi: 10.1109/TIP.2025.3636789.

Eric Kernfeld, Misha Kilmer, and Shuchin Aeron. Tensor–tensor products with invertible linear transforms. *Linear Algebra and its Applications*, 485:545–570, 2015. doi: 10.1016/j.laa.2015.07.021.

Misha E. Kilmer and Carla D. Martin. Factorization strategies for third-order tensors. *Linear Algebra and its Applications*, 435(3):641–658, 2011. doi: 10.1016/j.laa.2010.09.020.

Jiaxin Li, Ke Zheng, Lianru Gao, Li Ni, Min Huang, and Jocelyn Chanussot. Model-informed multistage unsupervised network for hyperspectral image super-resolution. *IEEE Transactions on Geoscience and Remote Sensing*, 62:1–17, 2024.

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.

Jianjun Liu, Zebin Wu, Liang Xiao, and Xiao-Jun Wu. Model inspired autoencoder for unsupervised hyperspectral image super-resolution. *IEEE Transactions on Geoscience and Remote Sensing*, 60: 1–12, 2022.

Ziming Liu, Yixuan Wang, Sachin Vaidya, Fabian Ruehle, James Halverson, Marin Soljacic, Thomas Hou, and Max Tegmark. Kan: Kolmogorov–arnold networks. In *International conference on learning representations*, volume 2025, pp. 70367–70413, 2025.

Ritik Shah and Marco F Duarte. Spectralift: Physics-guided spectral-inversion network for self-supervised hyperspectral image super-resolution. In *2025 15th Workshop on Hyperspectral Imaging and Signal Processing: Evolution in Remote Sensing (WHISPERS)*, pp. 1–5. IEEE, 2025.

He Wang, Yang Xu, Zhihui Wei, and Zebin Wu. Self-expressive high-order tensor unrolling network for unsupervised hyperspectral and multispectral image fusion. *IEEE Transactions on Image Processing*, 35:5714–5729, 2026. doi: 10.1109/TIP.2026.3695389.

Ting Wang, Zipei Yan, Jizhou Li, Xile Zhao, Chao Wang, and Michael Ng. Hyperspectral and multispectral image fusion with arbitrary resolution through self-supervised representations: T. wang et al. *International Journal of Computer Vision*, 133(11):7515–7535, 2025.

Jin-Liang Xiao, Ting-Zhu Huang, Liang-Jian Deng, Huidong Jiang, Qibin Zhao, and Gemine Vivone. Unsupervised coefficient learning framework for variational pansharpening. *Information Fusion*, 127:103790, 2026. ISSN 1566-2535. doi: https://doi.org/10.1016/j.inffus.2025.103790.

Honghui Xu, Chuangjie Fang, Yiqun Meng, Jiawei Jiang, Sixian Chan, Shiqing Zhang, and Jianwei Zheng. Tptransformer: Tensor-tensor product transformer for hyperspectral image super-resolution. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 1670–1679, 2026a.

Yang Xu, Jian Zhu, Danfeng Hong, Zhihui Wei, and Zebin Wu. Coupled diffusion posterior sampling for unsupervised hyperspectral and multispectral images fusion. *IEEE Transactions on Image Processing*, 35:69–84, 2026b. doi: 10.1109/TIP.2025.3647207.

Yang Xu, Jian Zhu, Danfeng Hong, Zhihui Wei, and Zebin Wu. Coupled diffusion posterior sampling for unsupervised hyperspectral and multispectral images fusion. *IEEE Transactions on Image Processing*, 35:69–84, 2026c.

Yiguo Yang, Dan Li, Yanyan Lv, Fanqiang Kong, and Qiang Wang. Multispectral and hyperspectral images fusion based on subspace representation and nonlocal low-rank regularization. *International Journal of Remote Sensing*, 45(9):2965–2984, 2024.

Jun Zhang, Chao Yi, Mingxi Ma, Mengling He, and Chao Wang. Low-rankness and smoothness meet subspace: A unified tensor regularization for hyperspectral image super-resolution. *Signal Processing*, pp. 110693, 2026.

Jian Zhu, He Wang, Yang Xu, Zebin Wu, and Zhihui Wei. Self-learning hyperspectral and multispectral image fusion via adaptive residual guided subspace diffusion model. In *2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 17862–17871. IEEE, 2025.

Qingke Zou, Jie Zhou, and Mingjie Luo. Hyperspectral super-resolution via nonlinear unmixing. *Information Fusion*, 123:103295, 2025.

A Additional Theory for Conditioned Tensor Products

This appendix provides the explicit inverse construction and proofs of the algebraic, stability, and product-deviation results stated in Section 2.

A.1 Algebraic Properties of CTP

For a fixed conditioning state $\mathbf{c}$, Definition 2.1 can equivalently be written as

$$\mathcal{A} \star_{\mathbf{c}} \mathcal{B} = \Phi_{\mathbf{c}}^{-1} \left(\Phi_{\mathbf{c}}(\mathcal{A}) \triangle \Phi_{\mathbf{c}}(\mathcal{B}) \right), \tag{17}$$

where $\Phi_{\mathbf{c}}$ is a linear bijection because $L_{\mathbf{c}} \in \mathrm{GL}(M, \mathbb{F})$.

Proof of Theorem 2.2. *Bilinearity.* For scalars $\alpha, \beta \in \mathbb{F}$,

$$\begin{aligned} &\Phi_{\mathbf{c}} \left((\alpha \mathcal{A}_1 + \beta \mathcal{A}_2) \star_{\mathbf{c}} \mathcal{B} \right) \\ &= \Phi_{\mathbf{c}}(\alpha \mathcal{A}_1 + \beta \mathcal{A}_2) \triangle \Phi_{\mathbf{c}}(\mathcal{B}) \\ &= \alpha \Phi_{\mathbf{c}}(\mathcal{A}_1) \triangle \Phi_{\mathbf{c}}(\mathcal{B}) + \beta \Phi_{\mathbf{c}}(\mathcal{A}_2) \triangle \Phi_{\mathbf{c}}(\mathcal{B}) \\ &= \Phi_{\mathbf{c}} \left(\alpha (\mathcal{A}_1 \star_{\mathbf{c}} \mathcal{B}) + \beta (\mathcal{A}_2 \star_{\mathbf{c}} \mathcal{B}) \right). \end{aligned} \tag{18}$$

Injectivity of $\Phi_{\mathbf{c}}$ establishes linearity in the first operand. Linearity in the second operand follows analogously.

Associativity. Using the same conditioning state $\mathbf{c}$ throughout,

$$\begin{aligned} &\Phi_{\mathbf{c}} \left((\mathcal{A} \star_{\mathbf{c}} \mathcal{B}) \star_{\mathbf{c}} \mathcal{D} \right) \\ &= \left[\Phi_{\mathbf{c}}(\mathcal{A}) \triangle \Phi_{\mathbf{c}}(\mathcal{B}) \right] \triangle \Phi_{\mathbf{c}}(\mathcal{D}) \\ &= \Phi_{\mathbf{c}}(\mathcal{A}) \triangle \left[\Phi_{\mathbf{c}}(\mathcal{B}) \triangle \Phi_{\mathbf{c}}(\mathcal{D}) \right] \\ &= \Phi_{\mathbf{c}} \left(\mathcal{A} \star_{\mathbf{c}} (\mathcal{B} \star_{\mathbf{c}} \mathcal{D}) \right), \end{aligned} \tag{19}$$

where the second equality follows from associativity of matrix multiplication in every frontal block. Since $\Phi_{\mathbf{c}}$ is injective,

$$(\mathcal{A} \star_{\mathbf{c}} \mathcal{B}) \star_{\mathbf{c}} \mathcal{D} = \mathcal{A} \star_{\mathbf{c}} (\mathcal{B} \star_{\mathbf{c}} \mathcal{D}).$$

Identity. Let $\mathcal{I} = (I, \ldots, I)$ denote the identity tensor under facewise multiplication, and define

$$\mathcal{E}_{\mathbf{c}} = \Phi_{\mathbf{c}}^{-1}(\mathcal{I}). \tag{20}$$

Then

$$\Phi_{\mathbf{c}}(\mathcal{A} \star_{\mathbf{c}} \mathcal{E}_{\mathbf{c}}) = \Phi_{\mathbf{c}}(\mathcal{A}) \triangle \mathcal{I} = \Phi_{\mathbf{c}}(\mathcal{A}),$$

which implies $\mathcal{A} \star_{\mathbf{c}} \mathcal{E}_{\mathbf{c}} = \mathcal{A}$. Similarly, $\mathcal{E}_{\mathbf{c}} \star_{\mathbf{c}} \mathcal{A} = \mathcal{A}$.

Inverse. Suppose every frontal block $[\Phi_{\mathbf{c}}(\mathcal{A})]^{(m)}$ is nonsingular. Define a tensor $\mathcal{Q}$ by

$$Q^{(m)} = \left([\Phi_{\mathbf{c}}(\mathcal{A})]^{(m)} \right)^{-1}, \qquad m = 1, \ldots, M, \tag{21}$$

and set

$$\mathcal{A}_{\star_{\mathbf{c}}}^{-1} = \Phi_{\mathbf{c}}^{-1}(\mathcal{Q}). \tag{22}$$

Then

$$\Phi_{\mathbf{c}} \left(\mathcal{A} \star_{\mathbf{c}} \mathcal{A}_{\star_{\mathbf{c}}}^{-1} \right) = \Phi_{\mathbf{c}}(\mathcal{A}) \triangle \mathcal{Q} = \mathcal{I}, \tag{23}$$

and therefore

$$\mathcal{A} \star_{\mathbf{c}} \mathcal{A}_{\star_{\mathbf{c}}}^{-1} = \mathcal{E}_{\mathbf{c}}.$$

The reverse multiplication gives the same result, so $\mathcal{A}_{\star_{\mathbf{c}}}^{-1}$ is a two-sided inverse.

Conversely, suppose a $\star_{\mathbf{c}}$-inverse exists. Applying $\Phi_{\mathbf{c}}$ to the inverse relation yields a facewise inverse of $\Phi_{\mathbf{c}}(\mathcal{A})$, which is possible only if each frontal block $[\Phi_{\mathbf{c}}(\mathcal{A})]^{(m)}$ is nonsingular. This proves the claim. □

Remark on conditioning states. The associativity result above concerns a single algebra $(\cdot, \star_{\mathbf{c}})$ with fixed $\mathbf{c}$. If the conditioning state is recomputed from intermediate operands, the successive products may belong to different members of the family $\{\star_{\mathbf{c}}\}$, and the preceding associativity argument no longer applies.

A.2 Stability of CTP

Proof of Theorem 2.3. Recall the block Frobenius norm

$$\|\mathcal{A}\|_{\mathrm{bF}}^2 = \sum_{m=1}^{M} \|A^{(m)}\|_F^2.$$

Because $\Phi_{\mathbf{c}}$ applies $L_{\mathbf{c}}$ along the block mode,

$$\|\Phi_{\mathbf{c}}(\mathcal{A})\|_{\mathrm{bF}} \le \|L_{\mathbf{c}}\|_2 \|\mathcal{A}\|_{\mathrm{bF}}. \tag{24}$$

Likewise,

$$\|\Phi_{\mathbf{c}}^{-1}(\mathcal{A})\|_{\mathrm{bF}} \le \|L_{\mathbf{c}}^{-1}\|_2 \|\mathcal{A}\|_{\mathrm{bF}}. \tag{25}$$

For facewise multiplication, Frobenius submultiplicativity gives

$$\|A^{(m)} B^{(m)}\|_F \le \|A^{(m)}\|_F \|B^{(m)}\|_F.$$

Consequently,

$$\begin{aligned} \|\mathcal{A} \triangle \mathcal{B}\|_{\mathrm{bF}}^2 &= \sum_{m=1}^{M} \|A^{(m)} B^{(m)}\|_F^2 \\ &\le \sum_{m=1}^{M} \|A^{(m)}\|_F^2 \|B^{(m)}\|_F^2 \\ &\le \left(\sum_{m=1}^{M} \|A^{(m)}\|_F^2 \right) \left(\sum_{m=1}^{M} \|B^{(m)}\|_F^2 \right), \end{aligned} \tag{26}$$

and hence

$$\|\mathcal{A} \triangle \mathcal{B}\|_{\mathrm{bF}} \le \|\mathcal{A}\|_{\mathrm{bF}} \|\mathcal{B}\|_{\mathrm{bF}}. \tag{27}$$

Using Eq. equation 17,

$$\begin{aligned} &\|\mathcal{A} \star_{\mathbf{c}} \mathcal{B}\|_{\mathrm{bF}} \\ &= \left\| \Phi_{\mathbf{c}}^{-1} \left(\Phi_{\mathbf{c}}(\mathcal{A}) \triangle \Phi_{\mathbf{c}}(\mathcal{B}) \right) \right\|_{\mathrm{bF}} \\ &\le \|L_{\mathbf{c}}^{-1}\|_2 \, \|\Phi_{\mathbf{c}}(\mathcal{A}) \triangle \Phi_{\mathbf{c}}(\mathcal{B})\|_{\mathrm{bF}} \\ &\le \|L_{\mathbf{c}}^{-1}\|_2 \|\Phi_{\mathbf{c}}(\mathcal{A})\|_{\mathrm{bF}} \|\Phi_{\mathbf{c}}(\mathcal{B})\|_{\mathrm{bF}} \\ &\le \|L_{\mathbf{c}}^{-1}\|_2 \|L_{\mathbf{c}}\|_2^2 \|\mathcal{A}\|_{\mathrm{bF}} \|\mathcal{B}\|_{\mathrm{bF}}. \end{aligned} \tag{28}$$

This proves Theorem 2.3. □

A.3 Positive-Diagonal CTP

For completeness, consider $L_{\mathbf{c}}^{+} = \mathrm{Diag}(\mathbf{g}(\mathbf{c}))$ with $\mathbf{g}(\mathbf{c}) \in \mathbb{R}_{++}^{M}$. Then

$$[\Phi_{\mathbf{c}}(\mathcal{A})]^{(m)} = g_m(\mathbf{c}) A^{(m)},$$

and

$$[\Phi_{\mathbf{c}}^{-1}(\mathcal{T})]^{(m)} = g_m(\mathbf{c})^{-1} T^{(m)}.$$

Therefore,

$$\begin{aligned} [\mathcal{A} \star_{\mathbf{c}}^{+} \mathcal{B}]^{(m)} &= g_m(\mathbf{c})^{-1} \left[g_m(\mathbf{c}) A^{(m)} \right] \left[g_m(\mathbf{c}) B^{(m)} \right] \\ &= g_m(\mathbf{c}) A^{(m)} B^{(m)}, \end{aligned} \tag{29}$$

which gives Eq. equation 3.

Under the residual simplex parameterization,

$$\mathbf{g}(\mathbf{c}) = (1-\rho)\mathbf{1} + \rho M \boldsymbol{\pi}(\mathbf{c}), \qquad \boldsymbol{\pi}(\mathbf{c}) = \mathrm{softmax}\left(\frac{q_\vartheta(\mathbf{c})}{\tau}\right),$$

each coefficient is strictly positive and satisfies

$$1-\rho < g_m(\mathbf{c}) < 1-\rho+\rho M.$$

Hence, $L_{\mathbf{c}}^{+}$ is nonsingular for every admissible conditioning state. Its spectral condition number satisfies

$$\kappa_2(L_{\mathbf{c}}^{+}) = \frac{\max_m g_m(\mathbf{c})}{\min_m g_m(\mathbf{c})} < \frac{1-\rho+\rho M}{1-\rho}. \tag{30}$$

A.4 Simplex-Induced Product Deviation

For a bilinear operator $\mathfrak{T}$ acting on two block tensors, define the induced bilinear operator norm as

$$\|\mathfrak{T}\|_{\mathrm{bil}} := \sup_{\|\mathcal{A}\|_{\mathrm{bF}}=\|\mathcal{B}\|_{\mathrm{bF}}=1} \|\mathfrak{T}(\mathcal{A},\mathcal{B})\|_{\mathrm{bF}}. \tag{31}$$

Proof of Proposition 2.5. Let

$$\Delta\mathbf{g}(\mathbf{c}) = \mathbf{g}(\mathbf{c}) - \mathbf{1}.$$

By Eq. equation 3, the m-th frontal block of $(\mathfrak{B}_{\mathbf{c}} - \triangle)(\mathcal{A},\mathcal{B})$ is

$$\Delta g_m(\mathbf{c}) A^{(m)} B^{(m)}.$$

Therefore,

$$\begin{aligned}
&\|(\mathfrak{B}_{\mathbf{c}} - \triangle)(\mathcal{A},\mathcal{B})\|_{\mathrm{bF}}^2 \\
&= \sum_{m=1}^{M} |\Delta g_m(\mathbf{c})|^2 \|A^{(m)}B^{(m)}\|_F^2 \\
&\le \|\Delta\mathbf{g}(\mathbf{c})\|_\infty^2 \sum_{m=1}^{M} \|A^{(m)}\|_F^2 \|B^{(m)}\|_F^2 \\
&\le \|\Delta\mathbf{g}(\mathbf{c})\|_\infty^2 \|\mathcal{A}\|_{\mathrm{bF}}^2 \|\mathcal{B}\|_{\mathrm{bF}}^2.
\end{aligned} \tag{32}$$

Thus,

$$\|\mathfrak{B}_{\mathbf{c}} - \triangle\|_{\mathrm{bil}} \le \|\mathbf{g}(\mathbf{c}) - \mathbf{1}\|_\infty. \tag{33}$$

To show that this bound is attained, choose

$$m^\star \in \arg\max_m |g_m(\mathbf{c}) - 1|.$$

Let u, v, and w be unit vectors of compatible dimensions and set

$$A^{(m^\star)} = uv^{\mathrm{H}}, \qquad B^{(m^\star)} = vw^{\mathrm{H}},$$

with all other frontal blocks equal to zero, where $^{\mathrm{H}}$ denotes transpose in the real case and conjugate transpose in the complex case. Then

$$\|\mathcal{A}\|_{\mathrm{bF}} = \|\mathcal{B}\|_{\mathrm{bF}} = 1$$

and

$$A^{(m^\star)}B^{(m^\star)} = uw^{\mathrm{H}}, \qquad \|uw^{\mathrm{H}}\|_F = 1.$$

Hence,

$$\begin{aligned}
&\|(\mathfrak{B}_{\mathbf{c}} - \triangle)(\mathcal{A},\mathcal{B})\|_{\mathrm{bF}} \\
&= |g_{m^\star}(\mathbf{c}) - 1| = \|\mathbf{g}(\mathbf{c}) - \mathbf{1}\|_\infty.
\end{aligned} \tag{34}$$

Together with Eq. equation 33, this yields

$$\|\mathfrak{B}_{\mathbf{c}} - \triangle\|_{\mathrm{bil}} = \|\mathbf{g}(\mathbf{c}) - \mathbf{1}\|_{\infty}. \tag{35}$$

Finally, the residual simplex parameterization gives

$$\begin{aligned}\mathbf{g}(\mathbf{c}) - \mathbf{1} &= (1-\rho)\mathbf{1} + \rho M \boldsymbol{\pi}(\mathbf{c}) - \mathbf{1} \\ &= \rho M \left(\boldsymbol{\pi}(\mathbf{c}) - \frac{\mathbf{1}}{M} \right).\end{aligned} \tag{36}$$

Therefore,

$$\|\mathbf{g}(\mathbf{c}) - \mathbf{1}\|_{\infty} = \rho M \left\| \boldsymbol{\pi}(\mathbf{c}) - \frac{\mathbf{1}}{M} \right\|_{\infty},$$

which proves Proposition 2.5. □

B Additional Theory for PF-RFJ-KAN

This appendix provides additional analytical results for the PF-RFJ-KAN introduced in Section 3, including the Jacobi recurrence, the proof of pole freedom, fractional-coordinate regularity, parameter-gradient bounds, approximation density, and a perturbation bound for the residual spectral correction.

B.1 Jacobi Polynomial Recurrence

For $\alpha, \beta > -1$, the first two Jacobi polynomials are

$$P_0^{(\alpha,\beta)}(u) = 1, \qquad P_1^{(\alpha,\beta)}(u) = \frac{1}{2}\left[(\alpha-\beta) + (\alpha+\beta+2)u\right]. \tag{37}$$

For $n \geq 2$, let $a = \alpha + \beta$ and define

$$\begin{aligned}A_n &= 2n(n+a)(2n+a-2), \\ B_n(u) &= (2n+a-1)\left[(2n+a)(2n+a-2)u + \alpha^2 - \beta^2\right], \\ C_n &= 2(n+\alpha-1)(n+\beta-1)(2n+a).\end{aligned} \tag{38}$$

The higher-order Jacobi polynomials are evaluated recursively as

$$P_n^{(\alpha,\beta)}(u) = \frac{B_n(u)P_{n-1}^{(\alpha,\beta)}(u) - C_n P_{n-2}^{(\alpha,\beta)}(u)}{A_n}. \tag{39}$$

This recurrence permits direct evaluation of the adaptive Jacobi basis without explicitly constructing high-order polynomial coefficients. In PF-RFJ-KAN, the parameters (α_j, β_j) are learned jointly with the remaining model parameters while satisfying $\alpha_j, \beta_j > -1$.

B.2 Proof of Pole Freedom and Uniform Boundedness

Proof of Theorem 3.1. For atom j, recall

$$D_j(u) = 1 + \frac{1}{d_D}\sum_{n=1}^{d_D} \mathrm{Softplus}(d_{j,n}) \left[P_n^{(\alpha_j,\beta_j)}(u)\right]^2. \tag{40}$$

Since $\mathrm{Softplus}(d_{j,n}) > 0$ and every squared basis value is nonnegative,

$$D_j(u) \geq 1, \qquad \forall u \in [-1,1]. \tag{41}$$

Moreover, because $\alpha_j, \beta_j > -1$, every Jacobi polynomial $P_n^{(\alpha_j,\beta_j)}$ is continuous on the compact interval $[-1,1]$. Therefore,

$$B_{j,n} = \max_{u\in[-1,1]} \left|P_n^{(\alpha_j,\beta_j)}(u)\right| < \infty.$$

For the numerator,

$$|N_j(u)| = \left| \sum_{n=0}^{d_N} c_{j,n} P_n^{(\alpha_j,\beta_j)}(u) \right| \leq \sum_{n=0}^{d_N} |c_{j,n}| \left| P_n^{(\alpha_j,\beta_j)}(u) \right| \leq \sum_{n=0}^{d_N} |c_{j,n}| B_{j,n} = C_j. \tag{42}$$

Because $D_j(u) \geq 1$,

$$|\phi_j(t)| = \frac{|N_j(u_j(t))|}{D_j(u_j(t))} \leq C_j, \qquad t \in [-1, 1]. \tag{43}$$

Hence, the denominator cannot vanish on the modeled spectral domain, and every PF-RFJ atom is uniformly bounded and free of real poles. □

B.3 Fractional-Coordinate Regularity

The signed fractional coordinate

$$u_\gamma(t) = \operatorname{sgn}(t)|t|^\gamma, \qquad \gamma > 0,$$

provides a controllable reparameterization of the normalized spectral axis. The following result characterizes its effect on the regularity of a PF-RFJ atom.

Theorem B.1 (Fractional-coordinate regularity)**.** *Let*

$$R_j(u) = \frac{N_j(u)}{D_j(u)}, \qquad \nu_j = \min\{\gamma_j, 1\},$$

and define

$$L_j = \max_{u \in [-1,1]} |R_j'(u)|.$$

Then, for any $s, t \in [-1, 1]$,

$$|\phi_j(t) - \phi_j(s)| \leq L_j \kappa_{\gamma_j} |t - s|^{\nu_j}, \tag{44}$$

where

$$\kappa_\gamma = \begin{cases} 2^{1-\gamma}, & 0 < \gamma \leq 1, \\ \gamma, & \gamma \geq 1. \end{cases} \tag{45}$$

Proof. Since $D_j(u) \geq 1$ on $[-1, 1]$, R_j is continuously differentiable on this compact interval and therefore $L_j < \infty$.

For $0 < \gamma \leq 1$, the signed fractional map satisfies

$$|u_\gamma(t) - u_\gamma(s)| \leq 2^{1-\gamma} |t - s|^\gamma. \tag{46}$$

When s and t have the same sign, this follows from the γ-Hölder continuity of x^γ on $[0, 1]$. When their signs differ, $|t - s| = |t| + |s|$, and concavity yields

$$|t|^\gamma + |s|^\gamma \leq 2^{1-\gamma} (|t| + |s|)^\gamma.$$

For $\gamma \geq 1$, u_γ has derivative magnitude $\gamma |t|^{\gamma-1} \leq \gamma$ away from the origin and is therefore γ-Lipschitz on $[-1, 1]$:

$$|u_\gamma(t) - u_\gamma(s)| \leq \gamma |t - s|. \tag{47}$$

Applying the mean-value bound for R_j gives

$$|\phi_j(t) - \phi_j(s)| = |R_j(u_j(t)) - R_j(u_j(s))| \leq L_j |u_j(t) - u_j(s)|,$$

which together with Eqs. equation 46 and equation 47 proves Eq. equation 44. □

The result shows that every admissible PF-RFJ atom remains continuous. For $\gamma_j > 1$, the fractional map becomes locally flat around the origin, whereas $0 < \gamma_j < 1$ permits a cusp-like local response when $R_j'(0) \neq 0$. Thus, different atoms can represent spectral variations with different local regularities.

B.4 Parameter-Gradient Bounds

The positive denominator also prevents parameter gradients from being amplified by near-zero denominator values.

Proposition B.2 (Parameter-gradient bounds)**.** *Under the notation of Theorem 3.1, the PF-RFJ atom satisfies*

$$\left|\frac{\partial \phi_j(t)}{\partial c_{j,n}}\right| \le B_{j,n}, \qquad 0 \le n \le d_N, \tag{48}$$

$$\left|\frac{\partial \phi_j(t)}{\partial d_{j,n}}\right| \le \frac{C_j B_{j,n}^2}{d_D}, \qquad 1 \le n \le d_D, \tag{49}$$

and

$$\left|\frac{\partial \phi_j(t)}{\partial \gamma_j}\right| \le \frac{L_j}{e\gamma_j}, \tag{50}$$

where $L_j = \max_{u\in[-1,1]} |R_j'(u)|$.

Proof. For the numerator coefficients,

$$\frac{\partial \phi_j(t)}{\partial c_{j,n}} = \frac{P_n^{(\alpha_j,\beta_j)}(u_j(t))}{D_j(u_j(t))}.$$

Since $D_j \ge 1$, Eq. equation 48 follows immediately.

For a raw denominator coefficient,

$$\frac{\partial \phi_j(t)}{\partial d_{j,n}} = -\frac{N_j(u_j(t))}{D_j(u_j(t))^2} \frac{\mathrm{sigmoid}(d_{j,n})}{d_D} \left[P_n^{(\alpha_j,\beta_j)}(u_j(t))\right]^2. \tag{51}$$

Using $|N_j| \le C_j$, $D_j \ge 1$, and $0 < \mathrm{sigmoid}(d_{j,n}) < 1$ gives Eq. equation 49.

Finally,

$$\frac{\partial \phi_j(t)}{\partial \gamma_j} = R_j'(u_j(t)) \frac{\partial u_j(t)}{\partial \gamma_j}.$$

For $t \ne 0$,

$$\left|\frac{\partial u_j(t)}{\partial \gamma_j}\right| = |t|^{\gamma_j} |\log|t||.$$

The scalar function $x^\gamma |\log x|$ on $x \in (0,1]$ attains its maximum at $x = e^{-1/\gamma}$, yielding

$$\max_{x\in(0,1]} x^\gamma |\log x| = \frac{1}{e\gamma}.$$

At $t = 0$, the derivative is defined by its limiting value 0. Therefore,

$$\left|\frac{\partial \phi_j(t)}{\partial \gamma_j}\right| \le L_j \frac{1}{e\gamma_j},$$

which proves Eq. equation 50. □

B.5 Approximation Density

The positivity constraint on the denominator does not prevent PF-RFJ from retaining a broad univariate approximation capability.

Theorem B.3 (Density in continuous univariate functions)**.** *Fix $d_D \ge 1$, and let $\mathcal{A}_{d_D}$ denote the union of admissible PF-RFJ atoms over numerator degrees $d_N \ge d_D$. Assume that the admissible parameter set contains $(\alpha, \beta, \gamma) = (0, 0, 1)$ and that the raw denominator parameters are unconstrained. Then $\mathcal{A}_{d_D}$ is dense in $C([-1,1])$ under the uniform norm.*

Proof. Let $f \in C([-1,1])$ and $\varepsilon > 0$. By the Weierstrass approximation theorem, there exists a polynomial p such that

$$\|f - p\|_\infty < \frac{\varepsilon}{2}. \tag{52}$$

Choose $(\alpha, \beta, \gamma) = (0, 0, 1)$. The Jacobi family then reduces to the Legendre polynomial family, which spans every polynomial of sufficiently low degree. Hence, for a sufficiently large d_N, the numerator can represent p exactly:

$$N(u) = p(u).$$

Because $\mathrm{Softplus} : \mathbb{R} \to (0, \infty)$ can approach zero arbitrarily closely, the raw denominator coefficients can be chosen such that

$$0 \le D(u) - 1 \le \eta$$

uniformly over $u \in [-1, 1]$, for any desired $\eta > 0$. Since $D(u) \ge 1$,

$$\begin{aligned} \left| \frac{p(u)}{D(u)} - p(u) \right| &= |p(u)| \frac{D(u) - 1}{D(u)} \\ &\le \|p\|_\infty \eta. \end{aligned} \tag{53}$$

Choosing η sufficiently small gives

$$\left\| \frac{p}{D} - p \right\|_\infty < \frac{\varepsilon}{2}.$$

Therefore,

$$\left\| f - \frac{p}{D} \right\|_\infty \le \|f - p\|_\infty + \left\| p - \frac{p}{D} \right\|_\infty < \varepsilon. \tag{54}$$

Thus, admissible PF-RFJ atoms are dense in $C([-1, 1])$. □

This is a representation result as the numerator degree increases; it does not imply that a finite-degree PF-RFJ parameterization is necessarily easier to optimize or superior to alternative parameterizations.

B.6 Bounded PF-RFJ Spectral Correction

Let $\widehat{\mathcal{Z}}^{(0)}_{k,p}$ denote the CPF reconstruction obtained from Eq. equation 15 with $\lambda = 0$.

Proposition B.4 (Bounded PF-RFJ correction)**.** *For every spectral band k and spatial location p,*

$$\left| \widehat{\mathcal{Z}}_{k,p} - \widehat{\mathcal{Z}}^{(0)}_{k,p} \right| \le |\lambda| \sum_{m=1}^{M} g_m(\mathbf{c}_p) \|\Delta G_k^{(m)}\|_F \|H_p^{(m)}\|_F. \tag{55}$$

Since $g_m(\mathbf{c}_p) > 0$ and $\sum_m g_m(\mathbf{c}_p) = M$,

$$\left| \widehat{\mathcal{Z}}_{k,p} - \widehat{\mathcal{Z}}^{(0)}_{k,p} \right| \le |\lambda| M \max_m \left(\|\Delta G_k^{(m)}\|_F \|H_p^{(m)}\|_F \right). \tag{56}$$

Proof. Subtracting the reconstruction with $\lambda = 0$ from Eq. equation 15 gives

$$\widehat{\mathcal{Z}}_{k,p} - \widehat{\mathcal{Z}}^{(0)}_{k,p} = \lambda \sum_{m=1}^{M} g_m(\mathbf{c}_p) \operatorname{tr}\left(\Delta G_k^{(m)} H_p^{(m)} \right).$$

Using the triangle inequality and

$$|\operatorname{tr}(AB)| \le \|A\|_F \|B\|_F$$

yields Eq. equation 55. The second inequality follows from positivity and the fixed-sum property $\sum_m g_m(\mathbf{c}_p) = M$. □

Thus, PF-RFJ acts as a controlled correction to the CPF spectral factor rather than as an unconstrained image-level bypass.

B.7 Huber Observation Loss

For a scalar residual e, the Huber penalty is

$$\rho_\delta(e) = \begin{cases} \frac{1}{2}e^2, & |e| \le \delta, \\ \delta\left(|e| - \frac{1}{2}\delta\right), & |e| > \delta. \end{cases} \tag{57}$$

For a residual tensor $\mathcal{R}$, we define

$$\mathcal{L}_\delta(\mathcal{R}) = \sum_i \rho_\delta(\mathcal{R}_i).$$

We use $\delta = 0.1$ in all experiments.

Accordingly, the observation-consistency objective in Eq. equation 16 jointly enforces consistency with the observed HR-MSI through the spectral response model and with the observed LR-HSI through the spatial degradation model.

B.8 Latent-Input Perturbation

Both deep-prior branches use learnable latent inputs. During iteration t, independent Gaussian perturbations are injected as

$$\widetilde{\boldsymbol{\xi}}_s^{(t)} = \boldsymbol{\xi}_s + \sigma_t \boldsymbol{\epsilon}_s^{(t)}, \qquad \widetilde{\boldsymbol{\xi}}_p^{(t)} = \boldsymbol{\xi}_p + \sigma_t \boldsymbol{\epsilon}_p^{(t)}, \tag{58}$$

where $\boldsymbol{\epsilon}_s^{(t)}$ and $\boldsymbol{\epsilon}_p^{(t)}$ are independently sampled standard Gaussian variables. The perturbation is used only during optimization; clean latent inputs are used for checkpoint evaluation.

B.9 Constant-Then-Cosine Scheduling

To reduce late-stage stochastic fitting, both the learning rate and the latent-input noise magnitude follow a constant-then-cosine schedule. For a generic quantity a_t with initial value a_0, terminal value $a_{\min}$, decay start T_a, and total number of iterations T, we use

$$a_t = \begin{cases} a_0, & t \le T_a, \\ a_{\min} + \dfrac{a_0 - a_{\min}}{2}\left[1 + \cos\left(\pi \dfrac{t - T_a}{T - T_a}\right)\right], & T_a < t \le T. \end{cases} \tag{59}$$

The same functional form is applied independently to the learning rate and perturbation magnitude, with their respective decay-start iterations and terminal values.

C Additional Experimental and Implementation Details

This appendix provides the implementation details omitted from the main text for compactness, including spatial-factor normalization, the observation loss, latent-input perturbation, optimization schedules, and checkpoint selection.

C.1 Datasets and Observation Generation

Datasets. We conduct simulated HSI–MSI fusion experiments on four remote-sensing hyperspectral datasets: PaviaC, Pavia City Center (PC), and Pavia University (PU).

1. *PaviaC.* The PaviaC scene was acquired by the ROSIS-03 sensor over Pavia, Italy. After removing noisy and water-absorption bands, we use a $192 \times 192 \times 102$ subcube for the experiments.
2. *Pavia City Center (PC).* The PC scene was also acquired by the ROSIS-03 sensor over Pavia, Italy. The original scene has a spatial size of 1096×715 pixels. We use a $200 \times 200 \times 80$ subcube containing representative roads, buildings, vegetation, and other urban structures. For the $\times 16$ setting, the spatial region is cropped to 192×192 so that its dimensions are divisible by the degradation factor.

3. *Pavia University (PU).* The PU scene was acquired by the ROSIS-03 sensor over the University of Pavia, Italy. The original scene has a spatial size of 610×340 pixels with a spatial resolution of approximately $1.3\,\mathrm{m}$. We use a $256 \times 256 \times 93$ subcube containing diverse urban structures, vegetation regions, and shadow areas.

Observation generation. All reference HSIs are converted to single precision and normalized by their global maximum. For each reference HR-HSI $\mathcal{Z}$, the LR-HSI $\mathcal{X}$ is generated by applying an 8×8 Gaussian blur with standard deviation 3, followed by spatial downsampling with scale $s \in \{8, 16\}$. The corresponding HR-MSI $\mathcal{Y}$ is generated using a four-channel spectral response matrix. For each dataset, the response matrix is truncated to match the number of hyperspectral bands and row-normalized before use. No additional observation noise is introduced.

Each scene–scale pair is optimized independently from its corresponding LR-HSI/HR-MSI observation pair. The reference HR-HSI is used only to generate the simulated observations and compute evaluation metrics, and is never used for optimization or checkpoint selection.

C.2 Evaluation Protocol

We evaluate reconstruction quality using PSNR, SSIM, SAM, and ERGAS. PSNR and SSIM are computed independently for each spectral band and then averaged over all bands, with the image dynamic range normalized to one. Reconstructed values are clipped to $[0, 1]$ for PSNR evaluation. SAM is reported in degrees and is computed over spatial locations where both the reference and reconstructed spectral vectors have ℓ_2 norms greater than 10^{-3}. ERGAS is computed using the corresponding spatial degradation factor, while bands whose reference mean is no greater than 10^{-4} are excluded.

All compared methods are evaluated using the same LR-HSI and HR-MSI observations. Methods assuming known degradations use identical spectral response matrices, spatial blur kernels, and scale factors, whereas blind methods retain their original degradation-estimation procedures. Published architectures and optimization settings are preserved whenever compatible with the experimental protocol.

C.3 Implementation Configuration

Table 3 summarizes the default configuration used for the main quantitative experiments. Each scene–scale setting is optimized independently from its corresponding LR-HSI/HR-MSI observation pair using the same configuration unless otherwise stated.

Parameter initialization. Both learnable latent inputs are initialized from $\mathcal{U}(0, 0.1)$. The weights and biases of the conditioning projection are initialized to zero, yielding uniform initial simplex probabilities. The routing strength is parameterized as $\rho = \mathrm{sigmoid}(r)$, where the learnable routing logit is initialized as $r = \mathrm{logit}(0.8)$.

For PF-RFJ-KAN, the Jacobi parameters are initialized as $\alpha_j = \beta_j = 0$, while the fractional exponent is initialized as $\gamma_j = 1$. In the implementation, α_j, β_j are constrained to $(-0.75, 1.5)$ and γ_j to $(0.25, 2.0)$. These ranges are implementation constraints used during optimization; the mathematical formulation in Section 3.2 only requires $\alpha_j, \beta_j > -1$ and $\gamma_j > 0$.

The numerator coefficients are initialized from a zero-mean Gaussian distribution with standard deviation $0.1/\sqrt{d_N + 1}$. The raw denominator coefficients are initialized to -4. The linear-readout weights are initialized from a zero-mean Gaussian distribution with standard deviation 0.01, with the corresponding bias initialized to zero.

Hardware and computational statistics. All reported runs were performed on NVIDIA GeForce RTX 3090 GPUs using Python 3.10.13, PyTorch 2.1.1+cu118, and CUDA 11.8. Across the six scene–scale settings reported in the main experiments, CPF-RFJ contains approximately 8.37–8.59 million trainable parameters. The recorded wall-clock time for a single test-time fit ranges from 13.9 to 18.7 minutes, with peak PyTorch-allocated GPU memory ranging from 2355 to 4060 MiB.

Table 3: Default configuration of CPF-RFJ used for the main quantitative experiments. Each scene–scale setting uses one independent self-supervised fit with random seed 766.

Parameter	Value
Number of CPF blocks M	6
Latent factor dimensions (r_a, r_b)	$(12, 12)$
PF-RFJ atom width J	32
Jacobi degrees (d_N, d_D)	$(4, 3)$
Gate temperature τ	0.20
Initial routing strength ρ_0	0.80
Spectral residual scale λ	0.20
Deep-prior channels (main, skip)	$(256, 8)$
Latent input channels (spatial, spectral)	$(10, 256)$
Huber threshold δ	0.10
Gate-balance regularization weight	0
Optimizer	Adam
Optimization iterations	7000
Initial/final learning rate	$10^{-3}/5 \times 10^{-5}$
Learning-rate decay start	3000
Initial/final input-noise std.	$1/30$ / 0.005
Input-noise decay start	2500
Clean-input evaluation interval	25 iterations
Checkpoint criterion	Minimum clean observation loss
Gradient clipping	Disabled
Random seed	766

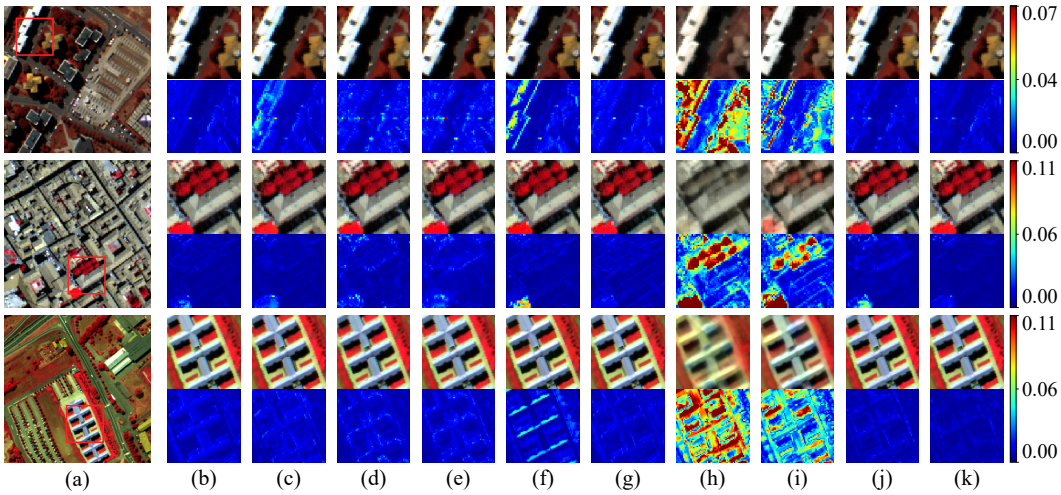


Figure 6: Qualitative comparison on three representative scenes under $\times 16$ spatial degradation. (a) Reference, (b) MIAE, (c) CS2DIPs, (d) M2U-Net, (e) EDIP-Net, (f) SpectraLift, (g) ARGS-Diff, (h) CLoRF, (i) CDPS, (j) UCL, and (k) CPF-RFJ. For each reconstruction method, the upper panel presents the reconstructed region and the lower panel shows the corresponding absolute error map.

C.4 Additional Qualitative Results under $\times 16$ Degradation

Figure 6 presents additional qualitative comparisons under the more challenging $\times 16$ spatial degradation. Compared with the $\times 8$ setting reported in the main paper, the substantially reduced spatial information in the LR-HSI increases the ambiguity of fine-detail recovery and makes the differences among competing methods more apparent.

Across the three scenes, most methods recover the dominant spatial layout, but noticeable errors remain around high-contrast boundaries, small structures, and elongated objects. In the first scene, several methods produce edge-aligned residuals around building contours and narrow structures,

Table 4: MSI observation-consistency error on the real Hyperion–Sentinel experiment. Lower is better.

Method	MSI_RE ↓
MIAE	0.0448
CS2DIPs	<u>0.0275</u>
M2U-Net	0.8034
EDIP-Net	0.8182
SpectraLift	0.0578
CSTF	0.1213
CLoRF	0.0349
LFCTNN	0.0883
UCL	0.1411
CPF-RFJ	**0.0253**

indicating incomplete recovery of local spatial transitions. Similar behavior can be observed in the second scene, where some competing methods introduce locally concentrated errors around the roof boundaries and adjacent structures. The third scene further highlights the difficulty of preserving repeated building patterns under severe downsampling, with several methods exhibiting strong residual responses along structural edges.

In contrast, CPF-RFJ consistently preserves clearer boundaries and more coherent local structures across all three scenes. Its error maps remain predominantly in the low-error range and contain fewer high-response regions around object boundaries and fine spatial details. This advantage is particularly evident in regions where spectral and spatial information must be jointly recovered from severely degraded observations. These results complement the quantitative improvements under $\times 16$ degradation and further support the effectiveness of observation-conditioned factor interactions in maintaining reconstruction fidelity under severe spatial information loss.

C.5 Detailed Results on the Real-World Experiment

We further provide detailed quantitative results for the real Hyperion–Sentinel experiment reported in the main paper. Since a ground-truth HR-HSI is unavailable for this real scene, conventional full-reference metrics such as PSNR, SSIM, SAM, and ERGAS cannot be directly computed. We therefore complement the qualitative comparison with an observation-consistency measure in the multispectral domain.

Specifically, the reconstructed HR-HSI $\widehat{\mathcal{Z}}$ is projected through the spectral response operator to obtain $\widehat{\mathcal{Y}} = \mathcal{P}_{\mathrm{spe}}(\widehat{\mathcal{Z}})$. We then compute the relative MSI observation error as

$$\mathrm{MSI_RE} = \frac{\left\|\widehat{\mathcal{Y}} - \mathcal{Y}\right\|_F}{\|\mathcal{Y}\|_F}, \tag{60}$$

where a lower value indicates better consistency between the reconstructed HSI and the observed HR-MSI.

As shown in Table 4, CPF-RFJ achieves the lowest MSI observation error of 0.0253, followed by CS2DIPs with 0.0275. This corresponds to an approximately 8.0% relative reduction over the second-best result. CLoRF and MIAE also maintain relatively low observation errors, whereas several competing methods exhibit substantially larger discrepancies after projection to the MSI domain.

Together with the qualitative results in Fig. 4, these results indicate that CPF-RFJ maintains strong agreement with the actual multispectral observation while preserving coherent spatial structures in the reconstructed HSI. Importantly, MSI_RE measures observation consistency rather than reconstruction accuracy with respect to an unavailable ground-truth HR-HSI; it should therefore be interpreted as complementary evidence for the real-world experiment rather than as a full-reference reconstruction metric.

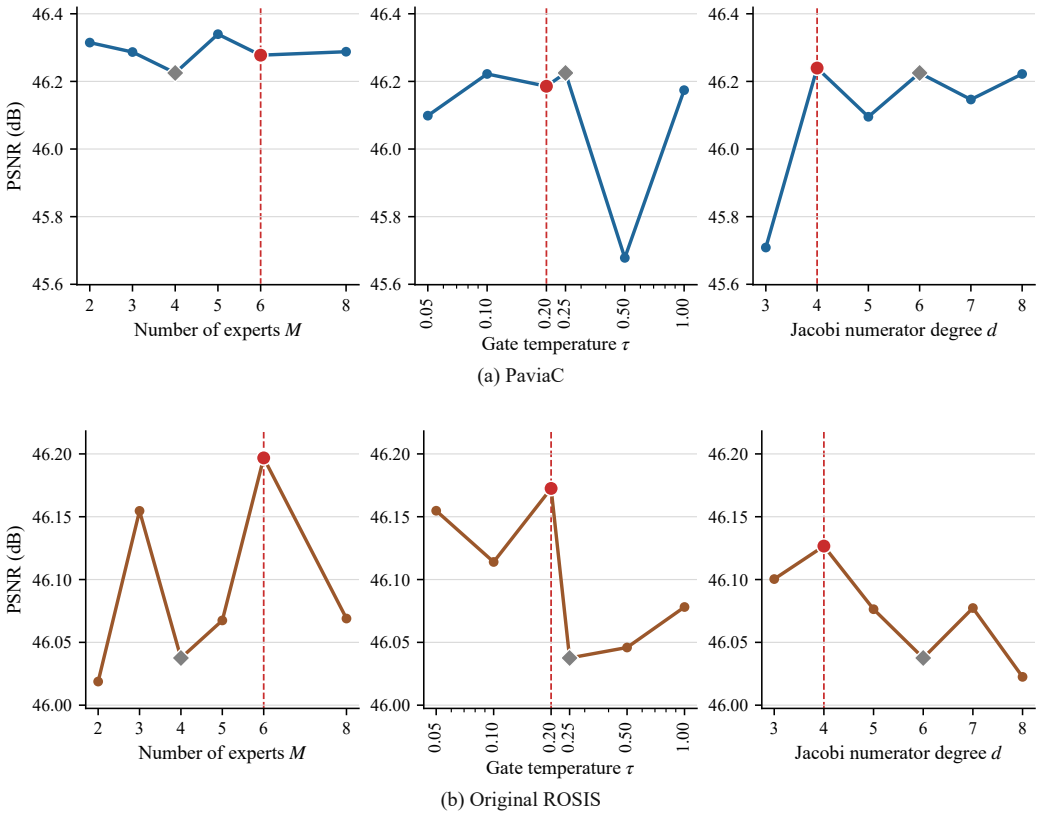


Figure 7: Sensitivity of CPF-RFJ to the number of CPF blocks M, gate temperature τ, and Jacobi numerator degree d_N on (a) PaviaC and (b) Original ROSIS. Red circles and dashed lines indicate the parameter values adopted in the main experiments.

C.6 Hyperparameter Sensitivity

We further analyze the sensitivity of CPF-RFJ to three representative hyperparameters: the number of CPF blocks M, the gate temperature τ, and the Jacobi numerator degree d_N. For each sweep, one parameter is varied while all remaining settings are kept fixed. We report the resulting PSNR on PaviaC and Original ROSIS in Fig. 7.

As shown in Fig. 7, the reconstruction performance remains relatively stable over a broad range of M and d_N, indicating limited sensitivity to the factorization size and polynomial degree. In comparison, the gate temperature τ has a more noticeable influence because it directly controls the concentration of the conditioning coefficients. Overall, the adopted configuration provides consistently competitive performance across the two scenes.

C.7 Spatial-Factor Normalization

Let $\overline{\mathcal{H}}^{(m)}$ denote the raw output of the two-dimensional spatial deep prior. At every high-resolution spatial location p, we enforce positivity and normalize jointly across the block and latent dimensions:

$$H^{(m)}_{b,a,p} = \frac{\mathrm{Softplus}\left(\overline{H}^{(m)}_{b,a,p}\right)}{\sum_{m'=1}^{M}\sum_{a'=1}^{r_a}\sum_{b'=1}^{r_b}\mathrm{Softplus}\left(\overline{H}^{(m')}_{b',a',p}\right)+\epsilon}. \tag{61}$$

The normalization yields nonnegative spatial factors with controlled aggregate magnitude at each spatial location, reducing scale ambiguity between the spectral and spatial factors.