Pattern Recognition 180 (2026) 114301

Contents lists available at ScienceDirect

Pattern Recognition

journal homepage: www.elsevier.com/locate/pr

Interpretable pan-sharpening via explicit spatial–spectral closed-loop priors

Chuang Liu [a], Zhiqi Zhang [b], Zhiwei Ye [b], Xiao Wu [c], Mi Wang [a,*], Jianhua Guo [d,e,*]

[a] State Key Laboratory of Information Engineering in Surveying, Mapping, and Remote Sensing, Wuhan University, Wuhan, 430079, China
[b] School of Computer Science, Hubei University of Technology, Wuhan, 430068, China
[c] School of Mathematical Sciences, University of Electronic Science and Technology of China, Chengdu, 611731, China
[d] International Research Center of Big Data for Sustainable Development Goals, Beijing, 100094, China
[e] Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing, 100094, China

ARTICLE INFO

Dataset link: https://github.com/JUSTM0VE0N/RSMIF_Project

Keywords:
Interpretable deep unfolding
Explicit prior modeling
Pan-sharpening
Multimodal image fusion
Cross-sensor generalization

ABSTRACT

Pan-sharpening is a multimodal image fusion problem that seeks to overcome the inherent spatial–spectral trade-off by fusing panchromatic and multispectral/hyperspectral observations under physical imaging constraints. Deep unfolding networks enhance interpretability by unrolling imaging models into learnable architectures; however, their priors are often encoded implicitly, obscuring the underlying physics and potentially causing unstable behavior manifested as sensitivity to sensor-dependent degradations and inconsistent spatial–spectral trade-offs across scenes and sensors. To mitigate this issue, we propose the CLPF framework, which formulates explicit spatial–spectral closed-loop priors and enforces clear bi-directional physical constraints. These constraints are expressed as jointly coupled constraints and are efficiently optimized via a variable-splitting strategy, yielding stable iterative updates. The derived network, UCLN, realizes this process through a closed-loop unfolding design. Extensive experiments on nine test sets from six satellite sensors demonstrate consistent improvements in fusion quality, efficiency, and cross-sensor generalization. The results validate the effectiveness of shifting from implicit to explicit prior modeling for high-fidelity and interpretable pan-sharpening.

1. Introduction

Pan-sharpening is a multimodal image fusion problem [1] that aims to recover high-resolution (HR) spectral images from panchromatic (PANI) and multispectral (MSI)/hyperspectral (HSI) observations under physical imaging constraints [2]. However, a critical gap persists in current methods: the pursuit of high-fidelity fusion often comes at the cost of physical interpretability [3]. This limitation undermines the reliability and trustworthiness of the results, thereby restricting their application in decision-critical remote sensing tasks that demand transparent and accountable outcomes. Such fused images are also valuable for downstream pattern recognition tasks, including object detection [4], change detection [5], and classification [6], where physical transparency and reliability are particularly important. For example, slight spectral bias can propagate to downstream indices and classifiers, causing systematic errors in vegetation or material mapping [7].

It is evident that the information contained in the input image pairs (denoted by $\mathbf{Y}$ and $\mathbf{Z}$) is significantly smaller than the desired image (denoted by $\mathbf{X}$) [8]. For instance, if $\mathbf{Z}$ is a PANI with a spatial scale of 400 × 400 and $\mathbf{Y}$ is an eight-band MSI with a size of 100 × 100 × 8, the total size of the input image pairs is 240,000 pixels, whereas the desired image (400 × 400 × 8) size is 1,280,000 pixels. Therefore, the fusion task can be considered as an ill-posed problem with an infinite number of solutions. To narrow the solution space, conventional methods typically estimate either a spectral degradation model between $\mathbf{Y}$ and $\mathbf{X}$ or a spatial degradation model between $\mathbf{Z}$ and $\mathbf{X}$, and then recover the fusion result by solving the corresponding inverse problem.

$$\begin{cases} \mathbf{Y} = \mathbf{X}\mathbf{R} + \mathbf{N}_1 \\ \mathbf{Z} = \mathbf{P}\mathbf{X} + \mathbf{N}_2, \end{cases} \tag{1}$$

Here, $\mathbf{X}$ is reshaped into a matrix by unfolding the spatial dimensions into the row dimension and the spectral bands into the column dimension, such that $\mathbf{R}$ and $\mathbf{P}$ act as spectral and spatial degradation operators, respectively. $\mathbf{N}_1$ and $\mathbf{N}_2$ denote the noise terms.

Following the observation model above, conventional pan-sharpening approaches are commonly grouped into three families [9]: (1) *Component Substitution (CS)* [10]: This type of method assumes that the PANI is a spectrally weighted sum of the HRMSI, and corrects the

---

* Corresponding authors.
E-mail addresses: chuang.liu@whu.edu.cn (C. Liu), zzq540@hbut.edu.cn (Z. Zhang), hgcsyzw@hbut.edu.cn (Z. Ye), wxwsx1997@gmail.com (X. Wu), wangmi@whu.edu.cn (M. Wang), guojh@aircas.ac.cn (J. Guo).

https://doi.org/10.1016/j.patcog.2026.114301
Received 19 January 2026; Received in revised form 21 April 2026; Accepted 16 June 2026

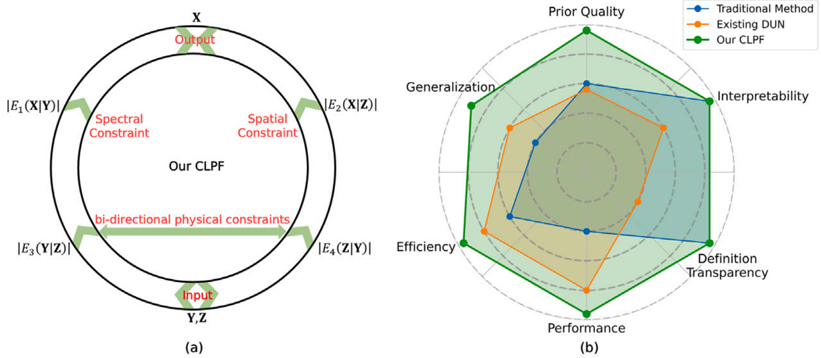

Fig. 1. Intuitive comparison with existing methods. (a) Schematic of spatial–spectral closed-loop priors. (b) Comparison with existing methods in terms of methodology.

spectral distortion by the degradation model; (2) *Multi-Resolution Analysis (MRA)* [11]: This type of method assumes that the high-frequency portion of PANI is consistent with the high-frequency portion of HRMSI, and the injection strategy, such as modulation transfer function (MTF) matching [12], is designed through the degradation model; and (3) *Variational Optimization (VO)* [13]: the fusion process is modeled as an inverse problem. The data fidelity term and regularization term are constructed through the degradation model, and the HRMSI is solved by optimizing the objective function. Evidently, these traditional methods possess clear physical interpretability and well-defined priors. However, their linear degradation models and hand-crafted priors limit their ability to represent complex nonlinear relationships in real-world scenes, resulting in a performance ceiling.

Recent advances in data-driven learning have shifted pan-sharpening from hand-crafted model design toward nonlinear representation learning with deep neural networks (DNNs) [14]. DNN-based methods [15] have shown impressive sharpening performance. Recent fusion models have further improved performance through stronger implicit spectral–spatial modeling, such as attention-based [16], Transformer-based [17], and Mamba-based [18] architectures. However, their black-box nature means that the underlying priors remain largely implicit, limiting physical transparency in high-reliability remote sensing applications. For this purpose, deep unfolding networks (DUNs) [19] have emerged, which embed physical prior information into DNNs to improve the interpretability of DNNs. Beyond in-domain benchmark accuracy, pan-sharpening models are expected to generalize across sensors and acquisition conditions while maintaining stable spatial–spectral consistency [20]. DUN-based methods [21] usually cast Eq. (1) into an optimization problem by minimizing a fidelity term together with a prior regularizer:

$$\hat{\mathbf{X}} = \arg\min_{\mathbf{X}} \{ F(\mathbf{X}, \mathbf{Y}, \mathbf{Z}) + |\Gamma(\mathbf{X})| \}, \tag{2}$$

where $F(\mathbf{X}, \mathbf{Y}, \mathbf{Z}) = \frac{\gamma_r}{2} \|\mathbf{X}\mathbf{R} - \mathbf{Y}\|_F^2 + \frac{\gamma_p}{2} \|\mathbf{P}\mathbf{X} - \mathbf{Z}\|_F^2$ denotes the joint fidelity term, with γ_r and γ_p balancing the contributions of each term. $|\Gamma(\mathbf{X})|$ denotes the prior term.

In general, $|\Gamma(\mathbf{X})|$ is often implemented implicitly, which means that existing DUNs are still translucent or black-box in nature. These methods fail to achieve accurate decoupling of physical prior knowledge, which prevents them from obtaining a stable solution space. To this end, we introduce a deep unfolding pan-sharpening framework (CLPF) that shifts the paradigm from implicit to explicit prior modeling for enhanced interpretability. By constructing explicit spatial–spectral closed-loop priors, our framework establishes bi-directional physical constraints driven by the imaging physics between different image pairs. As shown in Fig. 1(a), the four priors form a closed loop through jointly coupled constraints, enabling precise decoupling of physical knowledge and thereby ensuring stable, transparent, and high-fidelity solutions.

Further, the performance of these methods relies on stacking modules, resulting in a dramatic increase in sharpening computations and parameters. Building on the proposed CLPF, we implement a deep unfolding network termed UCLN, which realizes the framework with a compact closed-loop optimization design rather than heavy module stacking. As illustrated in Fig. 1(b), UCLN achieves a favorable trade-off among fusion quality, computational cost, and model size. The main contributions are summarized as follows:

- A novel framework denoted CLPF is derived for pan-sharpening. Moving beyond conventional implicit prior designs, this framework establishes four explicitly defined priors that form a closed-loop system through jointly coupled constraints modeling. This innovative design creates rigorous bi-directional physical constraints between input image pairs, significantly enhancing the stability of the solution space and providing a sound foundation for high-fidelity fusion with improved interpretability.
- We design UCLN, a lightweight and efficient network that accurately implements the CLPF. Unlike methods that loosely incorporate physical models, UCLN provides a precise computational realization of the CLPF optimization process. By incorporating a dedicated closed-loop prior optimization mechanism, the network simultaneously enforces spatial and spectral constraints without relying on heavy module stacking, achieving an optimal balance between performance, efficiency, and model complexity.
- Extensive experiments on nine test sets from six satellite sensors demonstrate the superiority of our method in MSI pan-sharpening, HSI pan-sharpening, and HSI super-resolution tasks. The results consistently show that UCLN improves both accuracy and efficiency, while maintaining robust cross-sensor generalization, indicating its robustness and applicability across diverse multimodal fusion settings.

2. Methodology

2.1. Model formulation

While offering a degree of interpretability over pure DNNs, existing DUN-based frameworks remain constrained by a core bottleneck: their reliance on implicit priors. This fundamentally hinders the decoupling of physical knowledge and fails to explore the inherent bi-directional constraints between PANI and MSI. To fill this gap, we introduce the CLPF, written as:

$$\hat{\mathbf{X}} = \arg\min_{\mathbf{X}} \left\{ F(\mathbf{X}, \mathbf{Y}, \mathbf{Z}) + \left| E_J(\mathbf{X}|\mathbf{Y}, \mathbf{Z}) \right| \right\}, \tag{3}$$

where $|E_J(\mathbf{X} \mid \mathbf{Y}, \mathbf{Z})|$ represents a joint prior acting on both spectral and spatial domains. We regularize this prior with an ℓ_1 penalty to encourage sparse, denoising-oriented representations. However, directly coupling the priors through E_J may mix spatial and spectral effects, reducing transparency. We therefore separate the joint term into two explicit components:

$$\hat{\mathbf{X}} = \arg\min_{\mathbf{X}} \{ F\left(\mathbf{X}, \mathbf{Y}, \mathbf{Z}\right) + \left|E_1\left(\mathbf{X}|\mathbf{Z}\right)\right| + \left|E_2\left(\mathbf{X}|\mathbf{Y}\right)\right| \}. \tag{4}$$

$|E_1(\mathbf{X}|\mathbf{Z})|$ denotes the spatial prior, which is a texture structure constraint guided by PANI $\mathbf{Z}$, while $|E_2(\mathbf{X}|\mathbf{Y})|$ denotes the spectral prior, which is a spectral information constraint guided by MSI $\mathbf{Y}$. Complementary information exists for $\mathbf{Z}$ and $\mathbf{Y}$ in both the spatial and spectral dimensions. Because neither $\mathbf{Y}$ nor $\mathbf{Z}$ alone contains sufficient information to uniquely recover $\mathbf{X}$, it is beneficial to explicitly model their mutual complementarity. For this reason, we construct explicit bidirectional mappings between $\mathbf{Z}$ and $\mathbf{Y}$. This approach formalizes their inherent complementarity into a closed-loop constraint, rigorously narrowing the solution space to achieve a solution that is both physically consistent and optimally balanced in spatial and spectral fidelity.

$$\hat{\mathbf{X}} = \arg\min_{\mathbf{X}} \left\{ F(\mathbf{X}, \mathbf{Y}, \mathbf{Z}) + \sum_{i=1}^{4} \lambda_i \| E_i(\cdot) \| \right\}, \tag{5}$$

where the terms $E_i(\cdot)$ are defined as:

$$\begin{aligned} E_1(\cdot) &:= E_1(\mathbf{X}|\mathbf{Z}), \\ E_2(\cdot) &:= E_2(\mathbf{X}|\mathbf{Y}), \\ E_3(\cdot) &:= E_3(\mathbf{Y}|\mathbf{Z}), \\ E_4(\cdot) &:= E_4(\mathbf{Z}|\mathbf{Y}). \end{aligned}$$

Here, λ_i controls the strength of each physical constraint.

2.1.1. *Spatial texture prior*

Spatial texture prior $E_1(\mathbf{X}|\mathbf{Z})$ primarily provides spatial texture during the fusion process. For this purpose, explicit relationship modeling can be performed by constraining the gradients of each band of $\mathbf{X}$ to be consistent with $\mathbf{Z}$.

$$\nabla_{e_1} \mathbf{X}^b = \alpha^b \nabla_{e_1} \mathbf{Z} + \beta^b, \quad \text{s.t. } b \in \{1, 2, \ldots, B\} \tag{6}$$

where b indexes spectral bands and B denotes the total number of bands. α and β are linear transformation coefficients; ∇_{e_1} denotes the spatial domain gradient, which can be obtained by local neighbor pixel differencing. Further, the spatial texture prior in CLPF can be expressed as:

$$\left|E_1\left(\mathbf{X}|\mathbf{Z}\right)\right| = \sum_{b \in B} \left(\nabla_{e_1} \mathbf{X}^b - \left(\alpha^b \nabla_{e_1} \mathbf{Z} + \beta^b \right) \right). \tag{7}$$

2.1.2. *Spectral information prior*

Spectral information prior $E_2(\mathbf{X}|\mathbf{Y})$ primarily provides spectral information during the fusion process. For this purpose, the explicit relationship can be modeled by constraining the spectral gradient of $\mathbf{X}$ and $\mathbf{Y}$. The spectral domain gradient can be obtained by calculating the difference between neighboring bands. As an example, the spectral gradient of $\mathbf{X}$ can be expressed as:

$$\begin{aligned} &\nabla_{e_2}^{\omega^b} \mathbf{X}^b = \omega_1^b \mathbf{X}^1 + \omega_2^b \mathbf{X}^2 + \cdots + \omega_B^b \mathbf{X}^B, \\ &\text{s.t. } \omega_i^b = \begin{cases} 1, & i = b \\ -1, & i = b + 1 \\ 0, & \text{otherwise} \end{cases} \end{aligned} \tag{8}$$

where ω_i^b denotes the linear weighting coefficients that encourage smooth spectral transitions across neighboring bands. Further, the spectral information prior in CLPF can be expressed as:

$$\left|E_2\left(\mathbf{X}|\mathbf{Y}\right)\right| = \sum_{b \in B} \left(\nabla_{e_2}^{\omega_i^b} \mathbf{X}^b - \left(\nabla_{e_2}^{\omega_i^b} \mathbf{Y}^b \right) \right). \tag{9}$$

2.1.3. *Bi-directional physical priors*

Bi-directional physical priors, $E_3(\mathbf{Y}|\mathbf{Z})$ and $E_4(\mathbf{Z}|\mathbf{Y})$, primarily provide complementary spatial–spectral information. In the fusion process, spatial detail and spectral distribution are coupled: enhanced spatial structures help refine spectral fidelity, while improved spectral consistency supports more reliable texture reconstruction. Motivated by this coupling, we further explicitly construct bi-directional physical constraints between $\mathbf{Z}$ and $\mathbf{Y}$. Specifically, the statistical relationship between $\mathbf{Y}$ and $\mathbf{Z}$ across resolutions is constructed through local mean alignment:

$$\left|E_3\left(\mathbf{Y}|\mathbf{Z}\right)\right| = \sum_{x,y,b} \left(\mathbb{E}\left[\mathbf{Y}^b_{(x,y)} \mid \mathbf{Z}_{\text{local}(x,y)} \right] - \mathbf{Y}^b_{(x,y)} \right)^2, \tag{10}$$

where $\mathbb{E}[\cdot]$ is computed as the local mean of $\mathbf{Z}$ within a 3×3 sliding window with unit stride; boundary pixels are handled by excluding padded values from the averaging denominator. In addition, the spectral response relationship between $\mathbf{Z}$ and $\mathbf{Y}$ can be constructed using the band linear mapping:

$$\left|E_4\left(\mathbf{Z}|\mathbf{Y}\right)\right| = \sum_{b} \left\| \rho_b \mathbf{Y}^b + \epsilon_b - \mathbf{Z} \right\|, \tag{11}$$

where ρ_b and ϵ_b are transformation coefficients. As illustrated in Fig. 2, the explicit closed-loop prior is formed by coupling the above prior. In the following, we concretize the way in which we optimize the closed-loop priors.

2.2. *Model optimization*

We solve CLPF using a variable-splitting scheme that combines gradient-based updates with a Split Bregman strategy. First, auxiliary variables are introduced to decouple the priors (non-smooth terms) in Eq. (5). Specifically, auxiliary variables $\mathbf{d}_i$, $i = 1, \ldots, 4$, can be defined as follows:

$$\begin{cases} \mathbf{d}_1 = E_1(\mathbf{X}|\mathbf{Z}) \\ \mathbf{d}_2 = E_2(\mathbf{X}|\mathbf{Y}) \\ \mathbf{d}_3 = E_3(\mathbf{Y}|\mathbf{Z}) \\ \mathbf{d}_4 = E_4(\mathbf{Z}|\mathbf{Y}) \end{cases} \tag{12}$$

By introducing the auxiliary variables, the original optimization can be equivalently rewritten as:

$$\min_{\mathbf{X}, \{\mathbf{d}_i\}_{i=1}^4} \; F(\mathbf{X}, \mathbf{Y}, \mathbf{Z}) + \sum_{i=1}^{4} \lambda_i \|\mathbf{d}_i\|_1 \quad \text{s.t. } \mathbf{d}_i = E_i(\cdot), \; i = 1, 2, 3, 4. \tag{13}$$

We construct the augmented Lagrangian as:

$$\mathcal{L} = F(\mathbf{X}, \mathbf{Y}, \mathbf{Z}) + \sum_{i=1}^{4} \left(\lambda_i \|\mathbf{d}_i\|_1 + \frac{\mu_i}{4} \|\mathbf{d}_i - E_i(\cdot) - \mathbf{b}_i\|_2^2 \right), \tag{14}$$

where μ_i are penalty coefficients. $\mathbf{b}_i$ are Bregman shift variables, which can be updated in iterations.

The constraints in Eq. (13) are decoupled through variable splitting. Subsequently, we minimize $\mathbf{X}$ and the auxiliary variable d_i in an iterative manner. Fixing d_i and b_i, the update of $\mathbf{X}$ is given by:

$$\mathbf{X}^{(k+1)} = \arg\min_{\mathbf{X}} \left\{ F(\mathbf{X}, \mathbf{Y}, \mathbf{Z}) + \sum_{i=1}^{4} \frac{\mu_i}{4} \|\mathbf{d}_i^{(k)} - E_i(\cdot) - \mathbf{b}_i^{(k)}\|_2^2 \right\} \tag{15}$$

where $k \in \{1, 2, \ldots, N\}$ indexes the unfolding stages. Similarly, $\mathbf{X}$ is fixed and d_i is updated by soft thresholding:

$$\mathbf{d}_i^{(k+1)} = \text{shrink}\left(E_i(\mathbf{X}^{(k+1)}) + \mathbf{b}_i^{(k)}, \frac{\lambda_i}{\mu_i} \right), \tag{16}$$

with the shrinkage operator:

$$\text{shrink}(z, \tau) = \text{sign}(z) \cdot \max(|z| - \tau, 0). \tag{17}$$

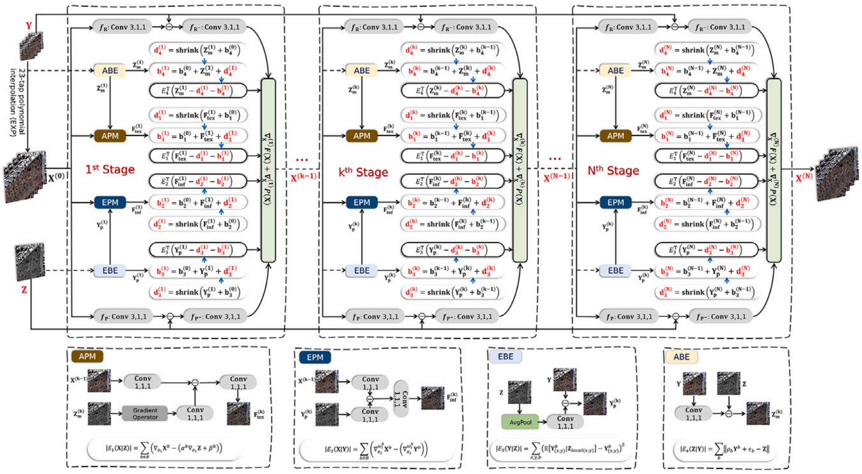

Fig. 2. Architecture of UCLN, obtained by unfolding the CLPF optimization. The network enforces explicit bi-directional physical constraints through the proposed spatial–spectral closed-loop priors.

The Bregman variables compensate the splitting residual and are updated as:

$$\mathbf{b}_i^{(k+1)} = \mathbf{b}_i^{(k)} + \left(E_i(\mathbf{X}^{(k+1)}) - \mathbf{d}_i^{(k+1)}\right). \tag{18}$$

From the above update process, it can be noticed that each subproblem only relies on the latest values of other variables during the iterations, instead of synchronous coupling. This manner allows avoiding joint optimization of high-dimensional variables, which yields convergent alternating updates for the convex subproblems and provides a stable iterative scheme that can be unfolded into a learnable network.

Further, the gradient descent algorithm is employed to solve for $\mathbf{X}$.

$$\mathbf{X}^{(k+1)} = \mathbf{X}^{(k)} - \eta\left(\nabla_{\mathbf{X}} F(\mathbf{X}^{(k)}) + \nabla_{\mathbf{X}} P(\mathbf{X}^{(k)})\right), \tag{19}$$

where η denotes the iteration step. $\nabla_{\mathbf{X}} F(\mathbf{X})$ and $\nabla_{\mathbf{X}} P(\mathbf{X})$ denote the fidelity term gradient and the prior term gradient, respectively. Of these, $\nabla_{\mathbf{X}} F(\mathbf{X})$ is computed as:

$$\nabla_{\mathbf{X}}^{(k+1)} F(\mathbf{X}) = \frac{\gamma_r}{2}(\mathbf{X}^{(k)\mathbf{R}} - \mathbf{Y})\mathbf{R}^{-} + \frac{\gamma_p}{2}\mathbf{P}^{-}(\mathbf{P}\mathbf{X}^{(k)} - \mathbf{Z}), \tag{20}$$

where $\mathbf{R}^{-}$ and $\mathbf{P}^{-}$ denote the pseudo-inverses of $\mathbf{R}$ and $\mathbf{P}$, respectively. The gradient contribution from the split prior terms is expressed as:

$$\begin{aligned}\nabla_{\mathbf{X}}^{(k+1)} P(\mathbf{X}) = &\frac{\mu_1}{4} E_1^{\top}(E_1(\mathbf{X}^{(k)}|\mathbf{Z}) - \mathbf{d}_1^{(k+1)} - \mathbf{b}_1^{(k+1)}) \\ &+ \frac{\mu_2}{4} E_2^{\top}(E_2(\mathbf{X}^{(k)}|\mathbf{Y}) - \mathbf{d}_2^{(k+1)} - \mathbf{b}_2^{(k+1)}) \\ &+ \frac{\mu_3}{4} E_3^{\top}(E_3(\mathbf{Y}^{(k)}|\mathbf{Z}) - \mathbf{d}_3^{(k+1)} - \mathbf{b}_3^{(k+1)}) \\ &+ \frac{\mu_4}{4} E_4^{\top}(E_4(\mathbf{Z}^{(k)}|\mathbf{Y}) - \mathbf{d}_4^{(k+1)} - \mathbf{b}_4^{(k+1)})\end{aligned} \tag{21}$$

where $E_i^{\top}$ represents the transpose version of E_i. In general, CLPF enables stable convergence to physically reasonable solutions by efficiently decoupling the fidelity and prior terms.

2.3. Model implementation

In line with the N iterations of the designed CLPF, we implement a deep unfolding network with N stages, titled UCLN. Rather than fixing the CLPF parameters by hand, UCLN learns the associated weights from data, enabling better adaptation of both fidelity and prior components to sensor-dependent degradations encountered in practice.

2.3.1. Estimation of fidelity terms $\nabla_{\mathbf{X}} F(\mathbf{X})$

As mentioned earlier, the weighting coefficients used to balance the different terms, such as λ_i, μ_i, η, γ_r, γ_p, and the degradation operators $\mathbf{R}$ and $\mathbf{P}$ are not subjected to manual fine-tuning. Instead, a data-driven manner of adaptive estimation is employed to fit the diverse data more exactly. For $\mathbf{R}$, $\mathbf{P}$, and their invertible versions, we estimate the degradation relations directly using either convolutions or transpose convolutions. Specifically, at each stage, we estimate it using a residual block whose structure consists of two convolution layers (kernel sizes and paddings set to 3, 1, and 1, respectively) followed by a ReLU activation. As a result, Eq. (20) can be remodeled as:

$$\begin{aligned}\nabla_{\mathbf{X}}^{(k)} F(\mathbf{X}) = &\frac{\gamma_r^{(k)}}{2} f_{\mathbf{R}^{-}}\left(f_{\mathbf{R}}(\mathbf{X}^{(k-1)}) - \mathbf{Y}\right) \\ &+ \frac{\gamma_p^{(k)}}{2} f_{\mathbf{P}^{-}}\left(f_{\mathbf{P}}(\mathbf{X}^{(k-1)}) - \mathbf{Z}\right).\end{aligned} \tag{22}$$

2.3.2. Estimation of prior terms $\nabla_{\mathbf{X}} P(\mathbf{X})$

The proposed closed-loop priors allow for the natural formation of bi-directional bridges, thereby establishing bi-directional physical constraints. We design a closed-loop prior optimization mechanism, which serves as the core building block of UCLN. By integrating our bespoke priors into a closed-loop prior optimization mechanism, we develop a spatial-to-spectral closed-loop prior module and a spectral-to-spatial closed-loop prior module, as shown in Fig. 2. In this case, the spatial-to-spectral closed-loop prior module consists of a spectral blind estimation module (EBE) and a spectral prior module (EPM), while the spectral-to-spatial closed-loop prior module consists of a spatial blind estimation block (ABE) and a spatial prior block (APM). APM, EPM, EBE and ABE are concrete implementations of $E_1(\mathbf{X}|\mathbf{Z})$, $E_2(\mathbf{X}|\mathbf{Y})$, $E_3(\mathbf{Y}|\mathbf{Z})$ and $E_4(\mathbf{Z}|\mathbf{Y})$, respectively.

More specifically, the four formulation-level priors are instantiated in UCLN in a one-to-one manner. The spatial-texture prior $E_1(\mathbf{X}|\mathbf{Z})$ is realized by APM. The spectral-information prior $E_2(\mathbf{X}|\mathbf{Y})$ is realized by EPM, where the spectral-gradient relation in Eqs. (8)–(9) is implemented through the spectral-gradient operator together with 1 × 1 convolutional projection. The local-mean alignment prior $E_3(\mathbf{Y}|\mathbf{Z})$ is realized by EBE, where the local mean in Eq. (10) is implemented by a 3 × 3 average-pooling operator with stride 1 and padding 1. The band-mapping prior $E_4(\mathbf{Z}|\mathbf{Y})$ is realized by ABE through 1 × 1 convolution-based channel transformation. In this way, the explicit physical relations defined in CLPF are preserved in UCLN through concrete module-level realizations rather than manually fixed coefficient sets. As a result, spectral closed-loop prior features and spatial closed-loop prior features are obtained by processing $\mathbf{X}$, $\mathbf{Y}$, and $\mathbf{Z}$ through two

Table 1
Quantitative comparisons on GF-2 test sets, with darker shades corresponding to better performance.

Method	Reduced-scale testing			Full-scale testing		
	SAM↓	ERGAS↓	Q2n↑	D_s ↓	QNR↑	HQNR↑
Traditional						
FS	1.681 ± 0.339	1.620 ± 0.353	0.890 ± 0.025	0.052 ± 0.017	0.925 ± 0.027	0.912 ± 0.021
BDSD-PC	1.724 ± 0.312	1.695 ± 0.390	0.885 ± 0.030	0.056 ± 0.020	0.932 ± 0.026	0.868 ± 0.037
DNN-based						
PanFormer	0.897 ± 0.165	0.844 ± 0.146	0.973 ± 0.009	0.042 ± 0.014	0.937 ± 0.021	0.919 ± 0.024
SSAFF	0.826 ± 0.143	0.813 ± 0.142	0.977 ± 0.010	0.038 ± 0.013	0.952 ± 0.018	0.936 ± 0.016
Pan-Mamba	0.755 ± 0.131	0.703 ± 0.094	0.981 ± 0.008	0.057 ± 0.012	0.933 ± 0.014	0.921 ± 0.015
PreMix	0.761 ± 0.142	0.690 ± 0.120	0.980 ± 0.007	0.031 ± 0.011	0.963 ± 0.015	0.951 ± 0.013
DUN-based						
GPPNN	1.344 ± 0.233	1.414 ± 0.251	0.937 ± 0.013	0.037 ± 0.013	0.953 ± 0.017	0.921 ± 0.020
GTP-PNet	1.385 ± 0.199	1.264 ± 0.151	0.939 ± 0.026	0.034 ± 0.013	0.952 ± 0.015	0.901 ± 0.026
LGTEUN	0.978 ± 0.121	0.827 ± 0.111	0.973 ± 0.012	0.035 ± 0.027	0.953 ± 0.017	0.935 ± 0.029
UTeRM	1.009 ± 0.185	0.925 ± 0.169	0.966 ± 0.011	0.038 ± 0.014	0.952 ± 0.019	0.934 ± 0.017
SSUN-Net	1.127 ± 0.189	0.891 ± 0.137	0.970 ± 0.011	0.024 ± 0.014	0.970 ± 0.013	0.951 ± 0.016
CLPF-based						
UCLN (Ours)	0.711 ± 0.134	0.656 ± 0.123	0.982 ± 0.007	0.025 ± 0.008	0.970 ± 0.014	0.957 ± 0.012

bi-directional closed-loop prior modules. The process can be formulated as:

$$\{\mathbf{Y}_{\mathrm{p}}^{(k)}, \mathbf{Z}_{\mathrm{m}}^{(k)}\} = \mathrm{EBE}(\mathbf{Z}, \mathbf{Y}), \mathrm{ABE}(\mathbf{Y}, \mathbf{Z}), \tag{23}$$

$$\{\mathbf{F}_{\mathrm{inf}}^{(k)}, \mathbf{F}_{\mathrm{tex}}^{(k)}\} = \mathrm{EPM}(\mathbf{X}^{(k)}, \mathbf{Y}_{\mathrm{p}}^{(k)}), \mathrm{APM}(\mathbf{X}^{(k)}, \mathbf{Z}_{\mathrm{m}}^{(k)}). \tag{24}$$

These module outputs are then used as the stage-wise realizations of the corresponding prior responses in Eq. (21), yielding the following implementation of $\nabla_{\mathbf{X}}^{(k)} P(\mathbf{X})$.

$$\begin{aligned} \nabla_{\mathbf{X}}^{(k)} P(\mathbf{X}) = & \frac{\mu_1}{4} E_1^{\top}(\mathbf{F}_{\mathrm{tex}}^{(k)} - \mathbf{d}_1^{(k)} - \mathbf{b}_1^{(k)}) \\ & + \frac{\mu_2}{4} E_2^{\top}(\mathbf{F}_{\mathrm{inf}}^{(k)} - \mathbf{d}_2^{(k)} - \mathbf{b}_2^{(k)}) \\ & + \frac{\mu_3}{4} E_3^{\top}(\mathbf{Y}_{\mathrm{p}}^{(k)} - \mathbf{d}_3^{(k)} - \mathbf{b}_3^{(k)}) \\ & + \frac{\mu_4}{4} E_4^{\top}(\mathbf{Z}_{\mathrm{m}}^{(k)} - \mathbf{d}_4^{(k)} - \mathbf{b}_4^{(k)}). \end{aligned} \tag{25}$$

3. Experimental setup

3.1. Training details

Since the core contribution is the spatial–spectral closed-loop priors, we adopt the commonly used ℓ_1 loss for training. It is noteworthy that we employ a consistent set of hyperparameters across all tasks, eliminating the need for task-specific adjustments. This approach demonstrates the method's strong generalization capability and practical applicability. We set the initial learning rate to 5×10^{-4}, train for 800 epochs with a batch size of 32, and adopt cosine annealing to decay the learning rate to 5×10^{-8}. The global random seed is fixed to 1024.

3.2. Dataset details

We evaluate the proposed method on six datasets, forming nine test sets from GF-2, QB, WV-III, WV-II [22], Botswana [23], and Pavia [24]. GF-2, QB, and WV-III are used for MSI pan-sharpening, while Botswana is used for HSI pan-sharpening and Pavia for HSI super-resolution. WV-II is further included to examine cross-sensor generalization. The adopted datasets are publicly available, and their data splits and preprocessing procedures follow the original sources.

3.3. Evaluation metrics

For simulated (reduced-scale) tests, we report four standard reduced-scale testing metrics, including PSNR [25], ERGAS [26], SAM, Q2n [27]. For real-data (full-scale) tests, we use the commonly adopted no-reference metrics, namely D_λ, D_λ^F, D_s, QNR and HQNR [28]. Given that QNR is calculated using D_λ and D_s, and HQNR is calculated using D_λ^F and D_s, D_λ and D_λ^F are not displayed in the quantitative experimental results.

3.4. Comparative methods

To benchmark fairly across settings, we compare representative baselines that are commonly adopted in the literature for each fusion task. All competitors are reported by family: (1) *MSI pan-sharpening.* Traditional: FS [12], BDSD-PC [11]. DNN-based: Panformer [29], SSAFF [30], Pan-Mamba [31], PreMix [32]. DUN-based: GPPNN [33], GTP-PNet [34], LGTEUN [19], UTeRM [35], SSUN-Net [36]. (2) *HSI pan-sharpening.* Traditional: GSA [37], CNMF [38], MTF-GLP [39]. DNN-based: PSRT [40], SFIGNet [41], MCIFNet [42]. DUN-based: DARN [43], Hyper-DSNet [23], DCINN [44]. (3) *HSI super-resolution.* Traditional: SFIM [45], CNMF, GFPCA [46]. DNN-based: QIS-GAN [47], SFIGNet [41], MCIFNet [42]. DUN-based: MDFL [48], 3DT-Net [49], DCINN [44]. For a fair comparison, all competing methods were retrained and evaluated under our unified open-source experimental framework, and the reproduced results were generally consistent with the performance reported in the original papers. The link to this framework is provided in the Data & Scripting Availability section. All experiments were executed on one compute node with two Intel Xeon Gold 6133 CPUs and two NVIDIA GeForce RTX 4090 GPUs.

4. Comparative study

4.1. Comparison experiments

4.1.1. Results on GF-2 test sets

Table 1 reports the quantitative results on the GF-2 test sets. Under reduced-scale testing, UCLN achieves the best scores across all metrics, improving SAM by 5.87% over Pan-Mamba and reducing ERGAS by 4.86% compared with PreMix, while also obtaining the highest Q2n. For full-scale testing, UCLN ranks first on QNR and HQNR and is only slightly behind SSUN-Net on D_s. Figs. 3 and 4 show qualitative results on simulated and real GF-2 data, respectively. Traditional methods suffer from blur and spectral inconsistency, and several learning-based baselines still miss fine structures. UCLN yields sharper textures and

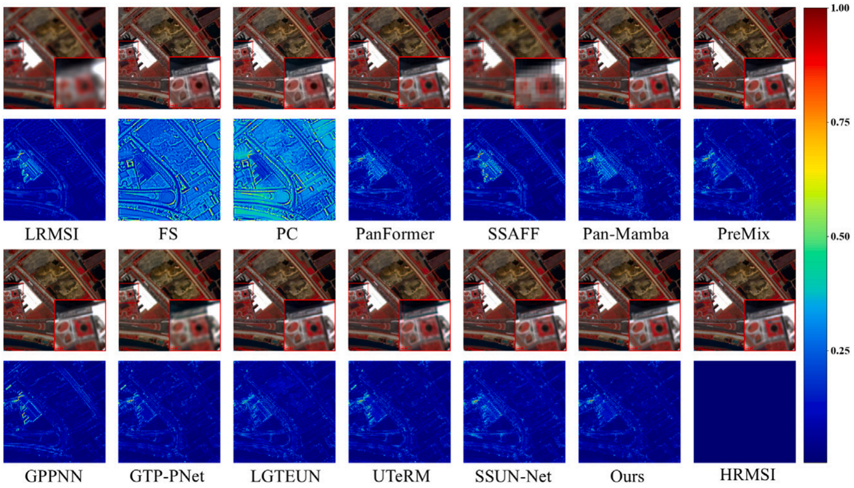

Fig. 3. The qualitative results (odd rows) and AEMs (even rows) on the GF-2 reduced-scale test set. For the AEMs, darker colors indicate smaller residual errors and thus better performance. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

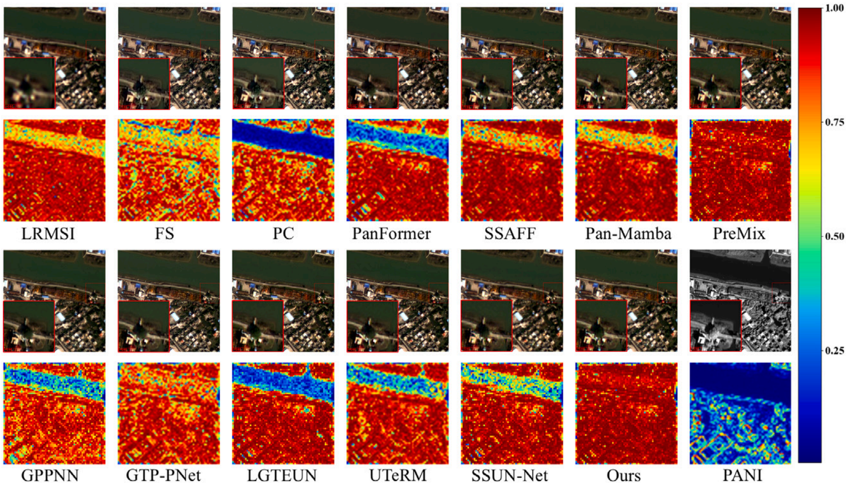

Fig. 4. The qualitative results (odd rows) and HQNR maps (even rows) on the GF-2 full-scale test set. For the HQNR maps, a redder overall appearance indicates better performance.

more coherent spectral distributions that better match the reference, as also reflected in the corresponding absolute error maps (AEMs) and HQNR maps.

4.1.2. Results on QB test sets

As shown in Table 2, the higher spatial resolution (0.61 m) and increased scene complexity degrade the performance of all methods. UCLN remains the top performer: on reduced-scale data it improves ERGAS, SAM, and Q2n by at least 3.22%, 5.40%, and 0.49%, respectively, and on full-scale real QB data it achieves the best non-reference metrics. Fig. 5 presents false-color composites on simulated QB data. Traditional methods (FS and BDSD-PC) exhibit clear structural and spectral distortions, and several learning-based baselines show detail loss or mild spectral bias. UCLN produces results and AEMs closest to the reference. On real QB data (Fig. 6), most methods show noticeable spectral distortion, whereas UCLN preserves a spectral distribution most consistent with LRMSI.

4.1.3. Results on WV-III test sets

Table 3 summarizes the quantitative results on the WV-III test sets. UCLN attains the best scores on all metrics; under reduced-scale testing it improves SAM, ERGAS, and Q2n by 1.60%, 2.95%, and 0.42% over PreMix, and under full-scale evaluation it ranks first on both QNR and HQNR. Figs. 7 and 8 report qualitative comparisons on simulated and real WV-III data. Traditional methods leave stronger residuals and more pronounced spatial–spectral distortions, and both Pan-Mamba and SSUN-Net degrade noticeably, particularly in vegetation regions. UCLN yields outputs closest to the reference, with the corresponding qualitative maps indicating reduced residual distortion and better no-reference quality.

4.2. Extensibility experiments

4.2.1. Extension on HSI pan-sharpening

The Botswana HSI contains 145 spectral bands, which makes spectrally faithful sharpening more challenging. Table 4 shows that UCLN improves PSNR, SAM and ERGAS by 1.1%, 4.1% and 8.7% over the second-best mamba-based MCIFNet, indicating improved reconstruction accuracy and spectral consistency. Fig. 9 further supports these results on representative scenes.

Table 2
Quantitative comparisons on QB test sets, with darker shades corresponding to better performance.

Method	Reduced-scale testing			Full-scale testing		
	SAM↓	ERGAS↓	Q2n↑	D_s ↓	QNR↑	HQNR↑
Traditional						
FS	7.866 ± 1.628	7.445 ± 0.551	0.834 ± 0.093	0.115 ± 0.023	0.855 ± 0.036	0.845 ± 0.030
BDSD-PC	8.181 ± 1.778	7.608 ± 0.574	0.829 ± 0.097	0.142 ± 0.031	0.836 ± 0.035	0.692 ± 0.046
DNN-based						
PanFormer	4.766 ± 0.826	4.124 ± 0.583	0.927 ± 0.091	0.047 ± 0.021	0.924 ± 0.035	0.895 ± 0.038
SSAFF	4.888 ± 0.888	4.615 ± 0.833	0.918 ± 0.091	0.051 ± 0.016	0.945 ± 0.016	0.893 ± 0.030
Pan-Mamba	5.046 ± 0.092	4.882 ± 0.093	0.909 ± 0.098	0.078 ± 0.032	0.853 ± 0.023	0.855 ± 0.029
PreMix	4.578 ± 0.741	3.820 ± 0.240	0.932 ± 0.092	0.071 ± 0.019	0.886 ± 0.022	0.883 ± 0.025
DUN-based						
GPPNN	5.860 ± 1.064	7.047 ± 0.643	0.872 ± 0.080	0.029 ± 0.020	0.942 ± 0.042	0.891 ± 0.036
GTP-PNet	6.635 ± 1.171	5.225 ± 0.371	0.891 ± 0.104	0.047 ± 0.021	0.920 ± 0.041	0.910 ± 0.033
LGTEUN	4.678 ± 0.765	3.840 ± 0.310	0.932 ± 0.090	0.069 ± 0.032	0.904 ± 0.032	0.857 ± 0.041
UTeRM	4.694 ± 0.777	3.819 ± 0.291	0.932 ± 0.089	0.066 ± 0.018	0.894 ± 0.032	0.899 ± 0.015
SSUN-Net	5.006 ± 0.824	4.043 ± 0.278	0.930 ± 0.086	0.065 ± 0.018	0.889 ± 0.026	0.892 ± 0.019
CLPF-based						
UCLN (Ours)	4.431 ± 0.743	3.613 ± 0.279	0.937 ± 0.087	0.022 ± 0.011	0.952 ± 0.034	0.937 ± 0.032

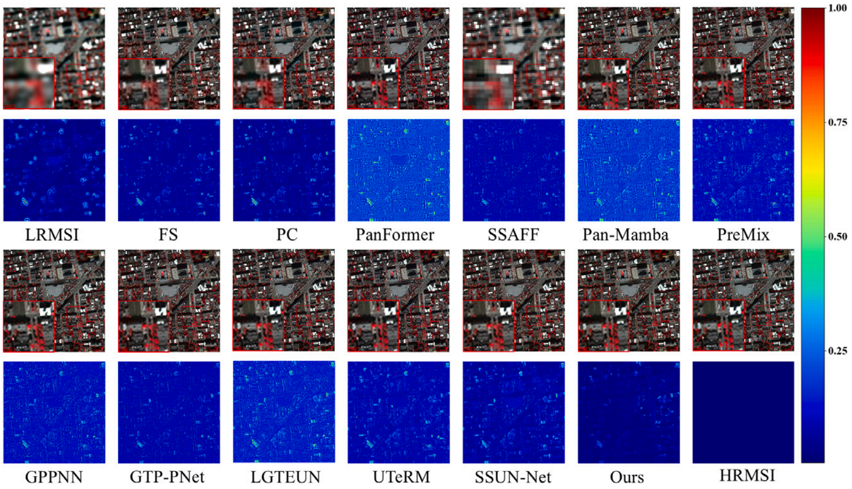

Fig. 5. The qualitative results (odd rows) and AEMs (even rows) on the QB reduced-scale test set. For the AEMs, darker colors indicate smaller residual errors and thus better performance. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

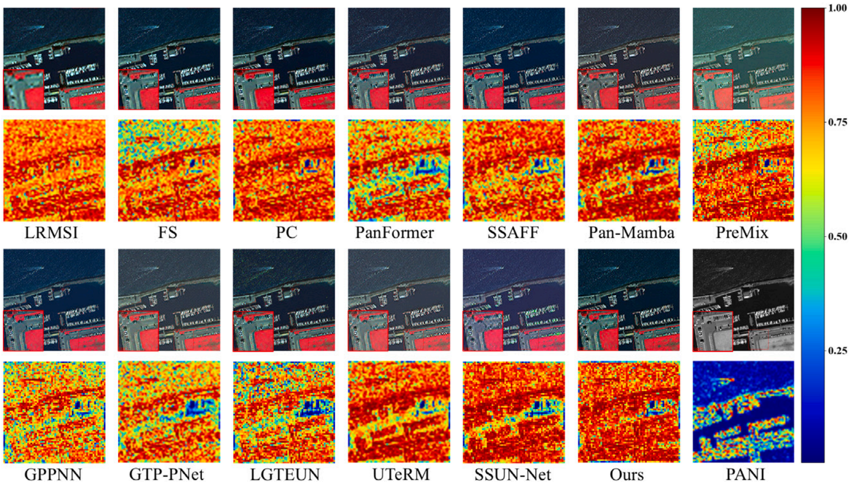

Fig. 6. The qualitative results (odd rows) and HQNR maps (even rows) on the QB full-scale test set. For the HQNR maps, a redder overall appearance indicates better performance.

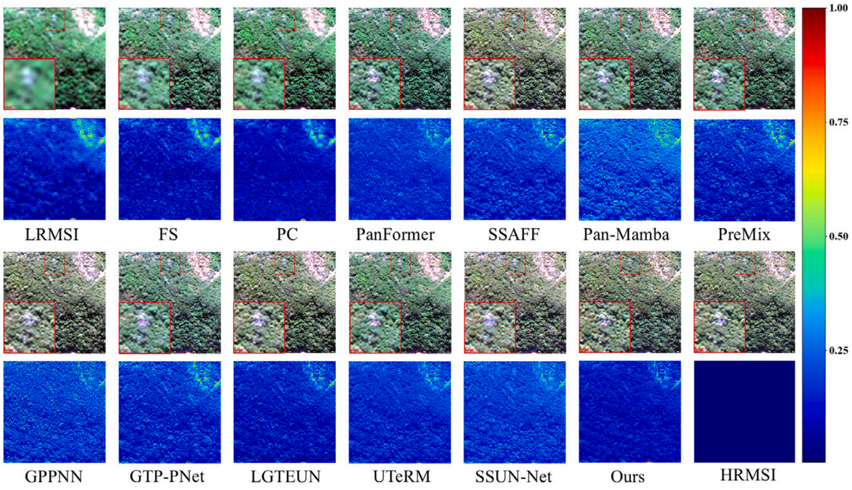

Fig. 7. The qualitative results (odd rows) and AEMs (even rows) on the WV-III reduced-scale test set. For the AEMs, darker colors indicate smaller residual errors and thus better performance. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

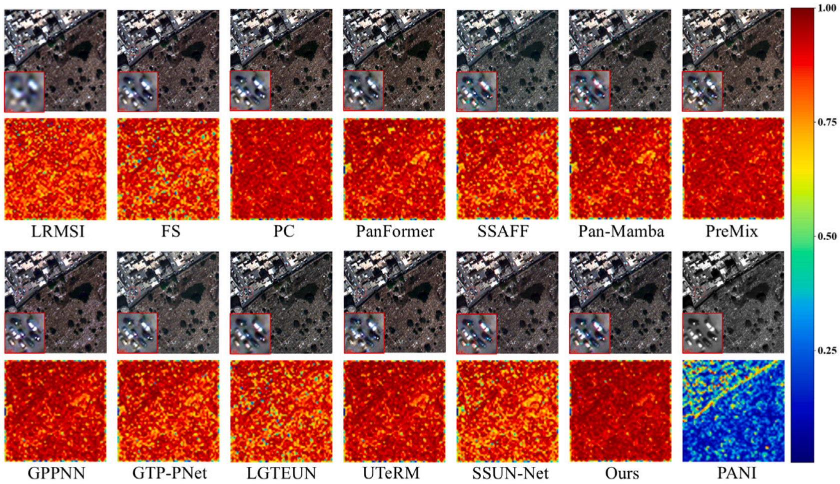

Fig. 8. The qualitative results (odd rows) and HQNR maps (even rows) on the WV-III full-scale test set. For the HQNR maps, a redder overall appearance indicates better performance.

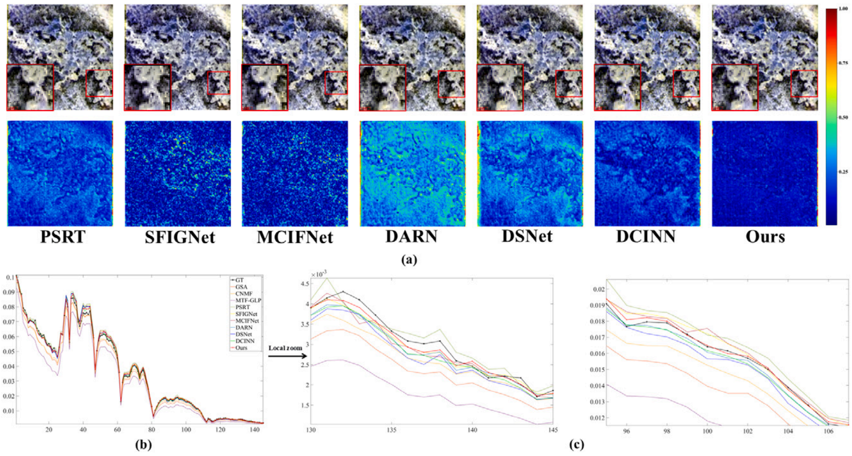

Fig. 9. Evaluation results on Botswana test set. (a) The qualitative results (Top) and AEMs (bottom) on the Botswana test set; (b) The comparisons of spectral vectors; (c) Zoomed-in view of subfigure (b). For the AEMs, darker colors indicate smaller residual errors and thus better performance. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 3
Quantitative comparisons on WV-III test sets, with darker shades corresponding to better performance.

Method	Reduced-scale testing			Full-scale testing		
	SAM↓	ERGAS↓	Q2n↑	D_s ↓	QNR↑	HQNR↑
Traditional						
FS	5.323 ± 1.611	4.645 ± 1.406	0.818 ± 0.099	0.085 ± 0.031	0.897 ± 0.043	0.897 ± 0.035
BDSD-PC	5.464 ± 1.671	4.650 ± 1.427	0.812 ± 0.104	0.091 ± 0.036	0.897 ± 0.043	0.853 ± 0.051
DNN-based						
PanFormer	3.278 ± 0.578	2.468 ± 0.556	0.901 ± 0.090	0.070 ± 0.022	0.911 ± 0.034	0.906 ± 0.023
SSAFF	3.208 ± 0.578	2.388 ± 0.533	0.902 ± 0.087	0.067 ± 0.040	0.902 ± 0.070	0.911 ± 0.057
Pan-Mamba	2.952 ± 0.552	2.237 ± 0.503	0.913 ± 0.075	0.031 ± 0.011	0.939 ± 0.024	0.952 ± 0.015
PreMix	2.950 ± 0.537	2.215 ± 0.490	0.915 ± 0.081	0.063 ± 0.020	0.924 ± 0.027	0.922 ± 0.022
DUN-based						
GPPNN	3.309 ± 0.626	2.483 ± 0.525	0.905 ± 0.082	0.062 ± 0.024	0.926 ± 0.034	0.899 ± 0.033
GTP-PNet	4.414 ± 0.664	3.040 ± 0.710	0.857 ± 0.109	0.076 ± 0.021	0.906 ± 0.030	0.898 ± 0.027
LGTEUN	3.152 ± 0.435	2.220 ± 0.471	0.904 ± 0.092	0.070 ± 0.010	0.912 ± 0.036	0.907 ± 0.029
UTeRM	3.691 ± 0.730	2.689 ± 0.596	0.873 ± 0.095	0.074 ± 0.024	0.911 ± 0.038	0.903 ± 0.030
SSUN-Net	3.232 ± 0.521	2.387 ± 0.561	0.902 ± 0.087	0.034 ± 0.014	0.947 ± 0.023	0.940 ± 0.020
CLPF-based						
UCLN (Ours)	2.902 ± 0.522	2.149 ± 0.466	0.918 ± 0.078	0.029 ± 0.010	0.948 ± 0.016	0.957 ± 0.012

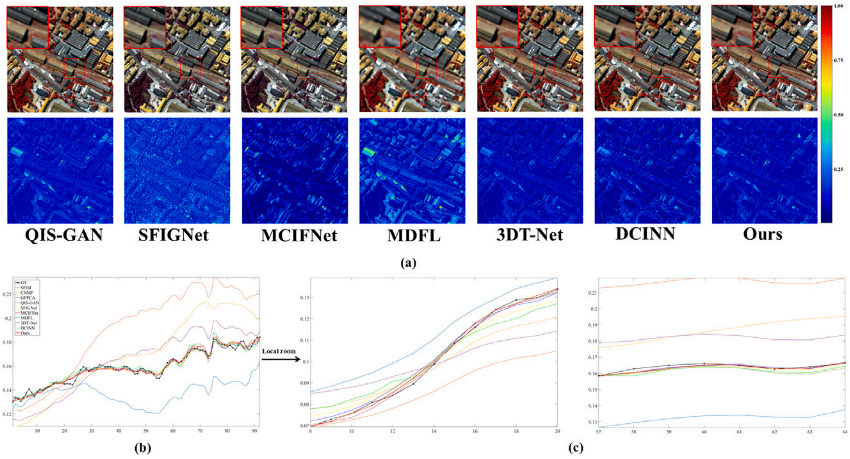

Fig. 10. Evaluation results on Pavia test set. (a) The qualitative results (Top) and AEMs (bottom) on the Pavia University test set; (b) The comparisons of spectral vectors; (c) Zoomed-in view of subfigure (b). For the AEMs, darker colors indicate smaller residual errors and thus better performance. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 4
Quantitative comparisons on HSI pan-sharpening task, with darker shades corresponding to better performance.

Method	Testing on Botswana		
	PSNR↑	SAM↓	ERGAS↓
Traditional			
GSA	34.887 ± 1.278	1.821 ± 0.222	2.189 ± 0.373
CNMF	31.954 ± 1.051	2.032 ± 0.244	2.881 ± 0.475
MTF-GLP	33.157 ± 1.265	1.928 ± 0.222	2.563 ± 0.352
DNN-based			
PSRT	37.554 ± 2.621	1.550 ± 0.199	1.814 ± 0.551
SFIGNet	37.502 ± 2.296	1.596 ± 0.184	1.857 ± 0.494
MCIFNet	38.127 ± 3.050	1.489 ± 0.205	1.763 ± 0.623
DUN-based			
DARN	37.538 ± 2.783	1.577 ± 0.220	1.821 ± 0.609
Hyper-DSNet	37.011 ± 2.485	1.646 ± 0.230	1.851 ± 0.502
DCINN	37.924 ± 2.356	1.591 ± 0.171	1.796 ± 0.484
CLPF-based			
UCLN (Ours)	38.545 ± 2.626	1.428 ± 0.173	1.610 ± 0.514

Table 5
Quantitative comparisons on HSI super-resolution task, with darker shades corresponding to better performance.

Method	Testing on pavia		
	PSNR↑	SAM↓	ERGAS↓
Traditional			
SFIM	26.268 ± 0.515	7.070 ± 0.665	8.428 ± 0.775
CNMF	30.659 ± 0.485	5.181 ± 0.182	5.463 ± 0.629
GFPCA	25.757 ± 0.575	7.605 ± 0.757	8.830 ± 0.788
DNN-based			
QIS-GAN	47.371 ± 1.334	1.563 ± 0.124	0.872 ± 0.161
SFIGNet	47.165 ± 1.186	1.622 ± 0.129	0.892 ± 0.170
MCIFNet	47.469 ± 2.026	1.499 ± 0.119	0.875 ± 0.216
DUN-based			
MDFL	28.060 ± 0.298	6.211 ± 0.678	6.837 ± 0.571
3DT-Net	47.210 ± 1.583	1.555 ± 0.109	0.896 ± 0.174
DCINN	44.612 ± 0.905	1.950 ± 0.173	1.100 ± 0.168
CLPF-based			
UCLN (Ours)	47.768 ± 1.614	1.477 ± 0.121	0.852 ± 0.168

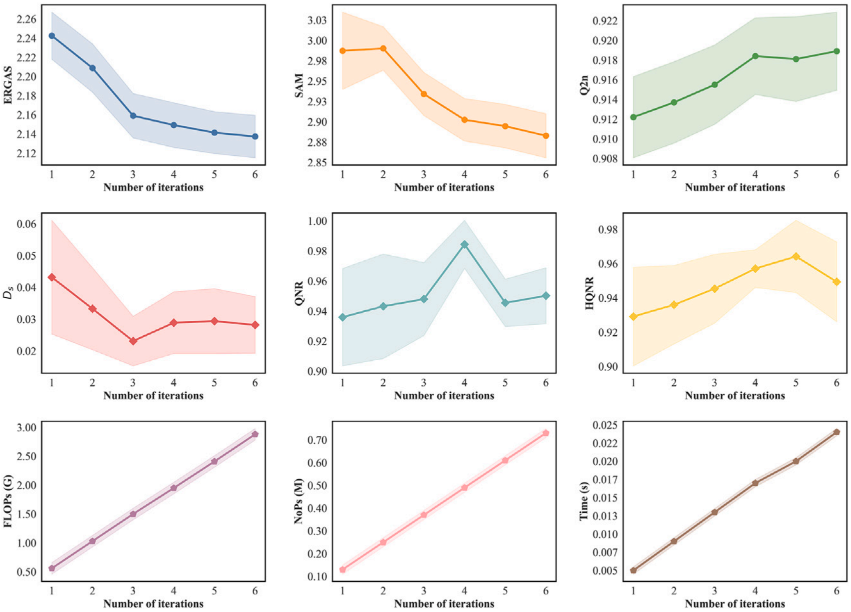

Fig. 11. Quantitative analysis on the number of iterations N.

4.2.2. Extension on HSI super-resolution

The Pavia HSI includes 92 spectral bands. Table 5 shows that UCLN improves PSNR and SAM by 0.6% and 1.5% over MCIFNet, and achieves a 2.3% improvement in ERGAS over QIS-GAN. As shown in Fig. 10, UCLN produces results closest to the reference, suggesting that the proposed closed-loop constraints generalize beyond MSI pan-sharpening to HSI super-resolution.

4.2.3. Comparisons of spectral vectors on two HSI datasets

We further compare spectral vectors at selected locations on Botswana (Fig. 9(b) and (c)) and Pavia (Fig. 10(b) and (c)). UCLN consistently yields curves that better follow the reference spectra and closely align with HRMSI, corroborating improved spectral preservation under multi-band fusion settings.

5. Ablation study and analysis

5.1. Ablation study

Since the core building of the proposed UCLN is spatial–spectral closed-loop priors, we perform ablation experiments on its variants on 8-band WV-III dataset.

5.1.1. Investigation of N

Increasing the iteration number N improves fusion quality but also increases computational cost and model size (see Fig. 11). Fig. 13(a) shows that larger N yields finer textures and more coherent spectral distributions, while the residual intensity in AEMs decreases accordingly. Overall, $N = 4$ achieves performance comparable to $N = 5$ and $N = 6$ with lower overhead; thus we set $N = 4$ in the following experiments.

5.1.2. Investigation of CLPF

We evaluate five UCLN variants: (I) using only $E_1(\mathbf{X}|\mathbf{Z})$, (II) using only $E_2(\mathbf{X}|\mathbf{Y})$, (III) using both $E_1(\mathbf{X}|\mathbf{Z})$ and $E_2(\mathbf{X}|\mathbf{Y})$, (IV) removing $E_3(\mathbf{Y}|\mathbf{Z})$, and (V) removing $E_4(\mathbf{Z}|\mathbf{Y})$. All variants are configured with the same number of parameters. Fig. 12 reports quantitative results and verifies the effectiveness of the explicit closed-loop design, while Fig. 13(b) shows that removing any closed-loop component degrades fidelity. These results indicate that the spatial, spectral, and bi-directional constraints contribute jointly to high-quality fusion.

5.1.3. Investigation of window size

We further study the window size used in the local-mean operator of Eq. (10). As shown in Fig. 14, the 3 × 3 setting yields the best overall performance under both reduced-scale and full-scale protocols. Enlarging the window to 5 × 5 or 7 × 7 consistently degrades performance. This suggests that a compact local window is more appropriate for Eq. (10), as it preserves locality more effectively, whereas larger windows introduce excessive averaging across heterogeneous regions and lead to a less favorable spatial–spectral trade-off.

5.1.4. Convergence and stability analysis

To further evaluate the stability of the proposed closed-loop unfolding framework, we examine the training process from both the optimization perspective and the unfolding perspective. Since the coefficients λ_i, μ_i, η, γ_r, and γ_p are estimated in a data-driven manner, we report both the training/validation loss curves and the validation metrics over training epochs in Fig. 15. The results show stable optimization behavior, with both losses decreasing steadily and the validation metrics evolving consistently toward better performance. Fig. 16 further presents the stage-wise behavior of the optimization-related coefficients in the unfolded updates. These results indicate that the learned coefficients remain well-behaved and that the reconstruction is progressively refined across unfolding stages. Therefore, the proposed closed-loop unfolding framework exhibits stable iterative behavior in practice.

5.2. Efficiency analysis

Fig. 17 provides a detailed performance-efficiency analysis of the proposed method. Beyond comparing UCLN with existing baselines, we additionally include five ablation variants (I-V) to reveal how different explicit prior components affect efficiency and fusion quality. The revised figure analyzes the trade-offs in terms of runtime, parameter count, and FLOPs, while also showing the corresponding performance changes. These results indicate that UCLN offers a favorable balance between fusion accuracy and computational cost, making it a practical solution for diverse fusion scenarios.

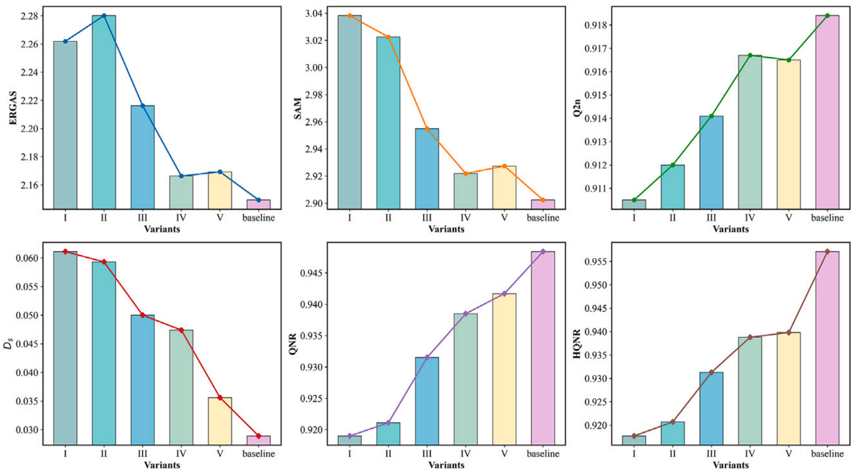

Fig. 12. Quantitative analysis on the CLPF architecture.

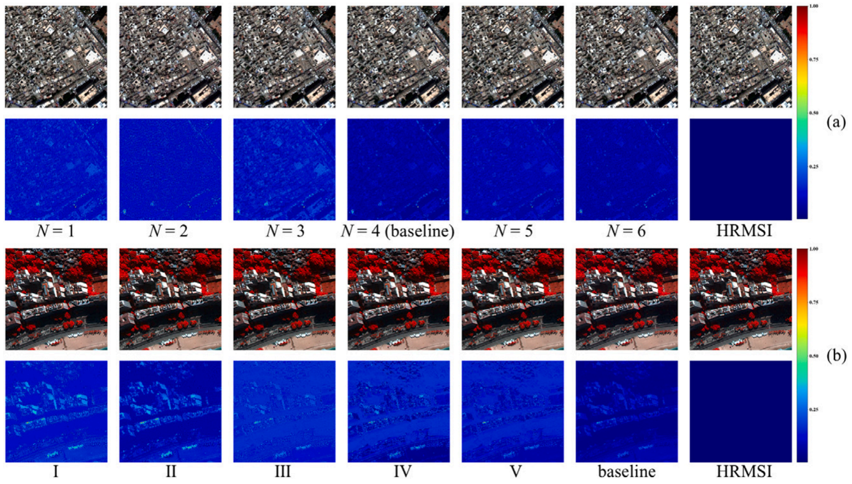

Fig. 13. Ablation study. (a) Visual comparisons (top) and corresponding AEMs (bottom) for different iteration numbers N. (b) Visual comparisons (top) and corresponding AEMs (bottom) for ablation variants of the CLPF components. For the AEMs, darker colors indicate smaller residual errors and thus better performance. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

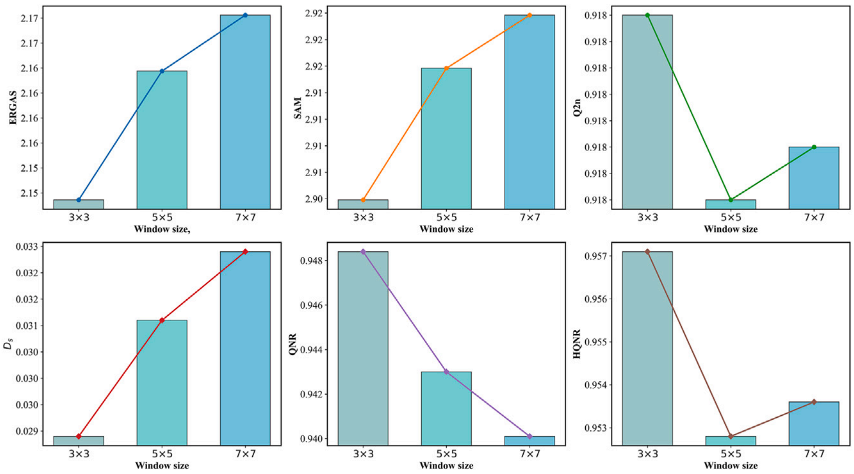

Fig. 14. Sensitivity analysis of the window size used in the local-mean operator of Eq. (10).

5.3. Cross-sensor generalizability analysis

To assess sensor transferability, we conducted a cross-sensor evaluation by training on WV-III and testing on WV-II imagery. The two sensors differ in spectral response functions and imaging characteristics, making this setting a realistic cross-sensor generalization scenario. As shown in Table 6, the proposed method consistently outperformed

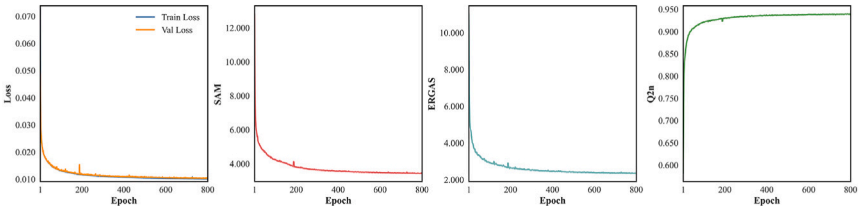

Fig. 15. Training and validation loss curves together with the validation metrics over training epochs.

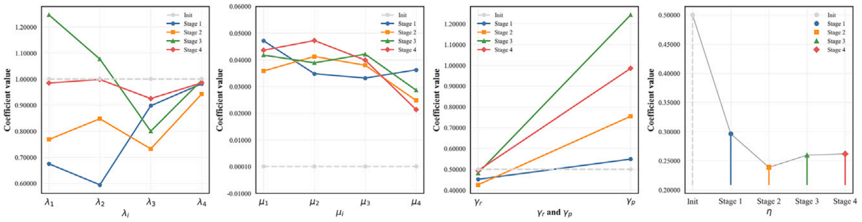

Fig. 16. Stage-wise analysis of the optimization-related coefficients λ_i, μ_i, η, γ_r, and γ_p in UCLN.

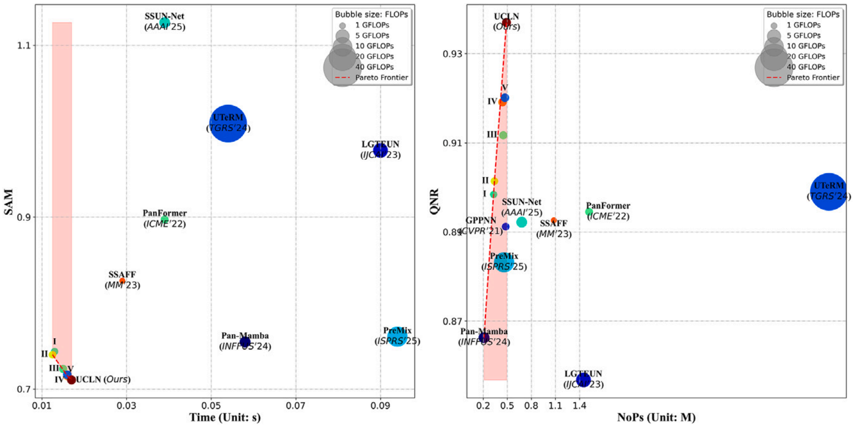

Fig. 17. Efficiency vs. reduced-scale performance (left) and Efficiency vs. full-scale performance (right) comparison.

representative DNN- and DUN-based baselines under this domain shift, indicating superior robustness to sensor-dependent discrepancies. We attribute this behavior to the explicit closed-loop spatial–spectral priors and bi-directional physical constraints, which reduce reliance on sensor-specific patterns and enhance generalization across platforms.

6. Discussion

6.1. Comparison with existing DUN-based methods

DUNs provide a principled bridge between physical imaging models and data-driven learning. However, many existing DUN-based approaches still encode priors implicitly in the spatial domain or spectral domain through transformer- or attention-based modules, which weakens physical traceability and interpretability. In addition, without stable bi-directional physical constraints across domains, it is difficult to maintain a well-balanced spatial–spectral trade-off when acquisition conditions vary. In contrast, the proposed CLPF explicitly formulates spatial, spectral, and cross-domain priors as a closed-loop formulation derived from the observation models. By decoupling these priors and enforcing them via bi-directional constraints, each unfolding update admits a clearer physical interpretation and helps stabilize the spatial–spectral trade-off without relying on heavy architectural stacking.

Table 6
Evaluation of Generalizability on WV-II Dataset, with darker shades corresponding to better performance.

Method	Generalizability testing		
	SAM↓	ERGAS↓	Q2n↑
Traditional			
FS	6.112 ± 0.757	4.562 ± 0.565	0.806 ± 0.077
BDSD-PC	6.248 ± 0.824	5.222 ± 0.856	0.789 ± 0.064
DNN-based			
PanFormer	7.175 ± 0.510	5.523 ± 0.296	0.817 ± 0.087
SSAFF	5.892 ± 0.495	5.412 ± 0.293	0.818 ± 0.077
Pan-Mamba	5.765 ± 0.451	5.033 ± 0.292	0.821 ± 0.088
PreMix	5.747 ± 0.433	5.015 ± 0.285	0.838 ± 0.076
DUN-based			
GPPNN	6.389 ± 0.644	4.829 ± 0.459	0.814 ± 0.075
GTP-PNet	6.411 ± 0.636	4.845 ± 0.463	0.813 ± 0.080
LGTEUN	5.857 ± 0.541	4.834 ± 0.397	0.819 ± 0.082
UTeRM	5.669 ± 0.557	4.793 ± 0.387	0.824 ± 0.081
SSUN-Net	5.635 ± 0.545	4.803 ± 0.386	0.841 ± 0.081
CLPF-based			
UCLN (Ours)	5.182 ± 0.476	3.859 ± 0.335	0.869 ± 0.082

6.2. Generalization across sensors and acquisition conditions

A key challenge in pan-sharpening is the variability across sensors and acquisition conditions, including differences in spatial resolution ratios, spectral response, MTF, and noise characteristics. Methods tuned to a single data distribution are therefore prone to performance degradation under sensor shift. The proposed explicit closed-loop priors mitigate this issue by constraining the fusion process with physically meaningful spatial–spectral relationships rather than sensor-specific patterns. As a result, the model exhibits more consistent behavior across sensors and preserves a more stable spatial–spectral trade-off under varying acquisition conditions.

6.3. *Scalability and extensibility of the framework*

Beyond benchmark-level evaluation, practical fusion models should remain efficient and extensible under larger-scale deployment and broader multimodal settings. The lightweight unfolding structure of UCLN supports scalable inference while preserving coherent spatial texture and spectral relationships. At the same time, the modular design of the closed-loop priors provides a flexible basis for extending the framework to more complex settings, such as multi-temporal fusion or scenarios with incomplete and degraded observations. This suggests that the proposed framework is not only effective for current pan-sharpening tasks, but also adaptable to broader multimodal fusion problems.

6.4. *Limitations*

Despite its strong performance and cross-sensor generalization, the proposed framework still has several limitations. First, it may face greater challenges under larger domain shifts, such as more heterogeneous multimodal fusion settings or severe misregistration and sensor-dependent degradations. Second, the number of unfolding iterations introduces a performance–efficiency trade-off, since increasing the iteration depth usually improves reconstruction fidelity but also raises computational cost.

7. Conclusion

This paper presents CLPF, an interpretable deep unfolding framework with explicit spatial–spectral closed-loop priors for pan-sharpening. The derived network, UCLN, preserves these priors through a closed-loop optimization design and achieves a favorable balance among fusion quality, efficiency, and cross-sensor generalization. Extensive experiments on MSI pan-sharpening, HSI pan-sharpening, and HSI super-resolution demonstrate that shifting from implicit to explicit prior modeling leads to more reliable and physically interpretable fusion. Despite these promising results, the current framework still has several limitations. In particular, its robustness under larger domain shifts and more heterogeneous multimodal settings remains to be further explored, and the unfolding depth introduces an inherent performance–efficiency trade-off. Future work will therefore focus on modality-aware extensions, uncertainty-aware constraints, and adaptive iteration mechanisms to further improve generalization and deployment efficiency. More broadly, beyond pan-sharpening itself, the proposed explicit-prior unfolding framework may offer a useful design paradigm for other multimodal image fusion and pattern recognition problems where physical interpretability, controllable optimization behavior, and cross-domain robustness are simultaneously desired. We hope that this explicit closed-loop modeling strategy can facilitate future research on more transparent and reliable learning-based fusion systems.

CRediT authorship contribution statement

Chuang Liu: Writing – original draft, Visualization, Validation, Methodology, Investigation, Formal analysis, Conceptualization. **Zhiqi Zhang:** Validation, Software. **Zhiwei Ye:** Validation, Software. **Xiao Wu:** Writing – review & editing, Validation, Investigation. **Mi Wang:** Resources, Project administration, Funding acquisition, Data curation, Conceptualization. **Jianhua Guo:** Writing – review & editing, Visualization, Formal analysis.

Declaration of competing interest

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

Acknowledgments

This research was supported in part by the National Science Fund for Distinguished Young Scholars under Grant 62425102.

Data availability

Code (including training and evaluation scripts, configuration files, and fixed random seeds) is available at: https://github.com/JUSTM0VE0N/RSMIF_Project. We will archive the exact code snapshot upon acceptance. All datasets used in this study are publicly available from their official providers; preprocessing and dataset split scripts are provided in the repository to reproduce the reported experiments. Pre-trained weights and inference demos are provided in the same repository.

References

[1] H. Ye, J. Chang, K. Wang, Z. Jia, W. Sun, Z. Li, A lightweight multilevel multiscale dual-path fusion network for remote sensing semantic segmentation, Pattern Recognit. 172 (2026) 112483, http://dx.doi.org/10.1016/j.patcog.2025.112483.

[2] L. Wu, Y. Yin, X. Jiang, T.C.E. Cheng, Pan-sharpening based on multi-objective decision for multi-band remote sensing images, Pattern Recognit. 118 (2021) 108022, http://dx.doi.org/10.1016/j.patcog.2021.108022.

[3] J. Li, K. Zheng, W. Liu, Z. Li, H. Yu, L. Ni, Model-guided coarse-to-fine fusion network for unsupervised hyperspectral image super-resolution, IEEE Geosci. Remote Sens. Lett. 20 (2023) 1–5, http://dx.doi.org/10.1109/LGRS.2023.3309854.

[4] L. Xiao, H. Li, S. Yao, C. Hou, EdgeFusionNet: Edge information-guided small object detection for remote sensing images, Pattern Recognit. 179 (2026) 113621, http://dx.doi.org/10.1016/j.patcog.2026.113621.

[5] Y. Sun, L. Lei, G. Kuang, Leveraging image transformation and optical flow for heterogeneous change detection under co-registration errors, Pattern Recognit. 177 (2026) 113275, http://dx.doi.org/10.1016/j.patcog.2026.113275.

[6] Z. Zhang, Z. Shao, J. Ma, J. Zhang, Y. Wang, Z. Liao, G. Cheng, J. Liu, M. Guo, L. Wu, Land-cover prior diffusion probabilistic model for remote sensing image super resolution, Pattern Recognit. 172 (2026) 112577, http://dx.doi.org/10.1016/j.patcog.2025.112577.

[7] J. Li, K. Zheng, L. Gao, Z. Han, Z. Li, J. Chanussot, Enhanced deep image prior for unsupervised hyperspectral image super-resolution, IEEE Trans. Geosci. Remote Sens. 63 (2025) 1–18, http://dx.doi.org/10.1109/TGRS.2025.3531646.

[8] H. Zheng, M. Gong, T. Liu, F. Jiang, T. Zhan, D. Lu, M. Zhang, HFA-Net: High frequency attention siamese network for building change detection in VHR remote sensing images, Pattern Recognit. 129 (2022) 108717, http://dx.doi.org/10.1016/j.patcog.2022.108717.

[9] C.-M. Lee, Y.-F. Lin, L.-W. Kang, C.-C. Hsu, Robust hyperspectral image pan-sharpening via sparse spatial-spectral representation, in: Proc. IEEE Int. Geosci. Remote Sens. Symp., IGARSS, 2025, pp. 2196–2201, http://dx.doi.org/10.1109/IGARSS55030.2025.11243541.

[10] J. Wang, Y. Lin, C. Chen, X. Huang, R. Zhang, Y. Wang, T. Lu, From forgotten to pan-sharpening, Pattern Recognit. 172 (2026) 112653, http://dx.doi.org/10.1016/j.patcog.2025.112653.

[11] G. Vivone, Robust band-dependent spatial-detail approaches for panchromatic sharpening, IEEE Trans. Geosci. Remote Sens. 57 (9) (2019) 6421–6433, http://dx.doi.org/10.1109/TGRS.2019.2906073.

[12] G. Vivone, R. Restaino, J. Chanussot, Full scale regression-based injection coefficients for panchromatic sharpening, IEEE Trans. Image Process. 27 (7) (2018) 3418–3431, http://dx.doi.org/10.1109/TIP.2018.2819501.

[13] H. Lu, Y. Yang, S. Huang, R. Liu, H. Guo, MSAN: Multiscale self-attention network for pansharpening, Pattern Recognit. 162 (2025) 111441, http://dx.doi.org/10.1016/j.patcog.2025.111441.

[14] X. Wei, M. Yuan, Adversarial pan-sharpening attacks for object detection in remote sensing, Pattern Recognit. 139 (2023) 109466, http://dx.doi.org/10.1016/j.patcog.2023.109466.

[15] J. Li, K. Zheng, L. Gao, L. Ni, M. Huang, J. Chanussot, Model-informed multistage unsupervised network for hyperspectral image super-resolution, IEEE Trans. Geosci. Remote Sens. 62 (2024) 1–17, http://dx.doi.org/10.1109/TGRS.2024.3391014.

[16] S. Liu, Y. Fan, Q. Hu, B. Li, Y. Zhang, S. Liu, SSMSFuse: A spectral and spatial multiscale coupling fusion model for hyperspectral and multispectral image, IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 18 (2025) 17310–17327, http://dx.doi.org/10.1109/JSTARS.2025.3586076.

[17] L. Chen, G. Vivone, J. Qin, J. Chanussot, X. Yang, Spectral–spatial transformer for hyperspectral image sharpening, IEEE Trans. Neural Netw. Learn. Syst. 35 (11) (2024) 16733–16747, http://dx.doi.org/10.1109/TNNLS.2023.3297319.

[18] M. Ma, M. Zhao, Y. Jiang, X. Li, W. Zhang, SSFMamba: Spatial-spectral fusion state space model for pansharpening, in: Proc. IEEE Int. Conf. Acoust., Speech, Signal Process., ICASSP 2025, 2025, pp. 1–5, http://dx.doi.org/10.1109/ICASSP49660.2025.10889877.

[19] M. Li, Y. Liu, T. Xiao, Y. Huang, G. Yang, Local-global transformer enhanced unfolding network for pan-sharpening, in: Proc. Int. Joint Conf. Artif. Intell., IJCAI, 2023, pp. 1071–1079.

[20] Z. Xue, G. Yang, X. Yu, A. Yu, Y. Guo, B. Liu, J. Zhou, Multimodal self-supervised learning for remote sensing data land cover classification, Pattern Recognit. 157 (2025) 110959, http://dx.doi.org/10.1016/j.patcog.2024.110959.

[21] L. Chen, G. Vivone, Z. Nie, J. Chanussot, X. Yang, Spatial data augmentation: Improving the generalization of neural networks for pansharpening, IEEE Trans. Geosci. Remote Sens. 61 (2023) 1–11, http://dx.doi.org/10.1109/TGRS.2023.3262262.

[22] L.-j. Deng, G. Vivone, M.E. Paoletti, G. Scarpa, J. He, Y. Zhang, J. Chanussot, A. Plaza, Machine learning in pansharpening: A benchmark, from shallow to deep networks, IEEE Geosci. Remote Sens. Mag. 10 (3) (2022) 279–315, http://dx.doi.org/10.1109/MGRS.2022.3187652.

[23] Y.-W. Zhuo, T.-J. Zhang, J.-F. Hu, H.-X. Dou, T.-Z. Huang, L.-J. Deng, A deep-shallow fusion network with multidetail extractor and spectral attention for hyperspectral pansharpening, IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 15 (2022) 7539–7555, http://dx.doi.org/10.1109/JSTARS.2022.3202866.

[24] Y. Xu, B. Du, L. Zhang, D. Cerra, M. Pato, E. Carmona, S. Prasad, N. Yokoya, R. Hänsch, B. Le Saux, Advanced multi-sensor optical remote sensing for urban land use and land cover classification: Outcome of the 2018 IEEE GRSS data fusion contest, IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 12 (6) (2019) 1709–1724, http://dx.doi.org/10.1109/JSTARS.2019.2911113.

[25] Y. Wang, H. Zheng, F. Li, Y. Lin, L. Fan, X. He, Y. Huang, X. Ding, Toward generalizable pansharpening: Conditional flow-based learning guided by implicit high-frequency priors, IEEE Trans. Geosci. Remote Sens. 63 (2025) 1–17, http://dx.doi.org/10.1109/TGRS.2025.3539013.

[26] Y. Cui, P. Liu, H. Zhang, Enpowering your pansharpening models with generalizability: Unified distribution is all you need, in: Proc. IEEE Int. Conf. Comput. Vis., ICCV, 2025, pp. 11850–11860.

[27] A. Garzelli, F. Nencini, Hypercomplex quality assessment of multi/hyperspectral images, IEEE Geosci. Remote Sens. Lett. 6 (4) (2009) 662–665, http://dx.doi.org/10.1109/LGRS.2009.2022650.

[28] A. Arienzo, G. Vivone, A. Garzelli, L. Alparone, J. Chanussot, Full-resolution quality assessment of pansharpening: Theoretical and hands-on approaches, IEEE Geosci. Remote Sens. Mag. 10 (3) (2022) 168–201, http://dx.doi.org/10.1109/MGRS.2022.3170092.

[29] H. Zhou, Q. Liu, Y. Wang, PanFormer: A transformer based model for pan-sharpening, in: Proc. IEEE Int. Conf. Multimedia Expo, ICME, 2022, pp. 1–6, http://dx.doi.org/10.1109/ICME52920.2022.9859770.

[30] Y. Yang, M. Li, S. Huang, H. Lu, W. Tu, W. Wan, Multi-scale spatial-spectral attention guided fusion network for pansharpening, in: Proc. 31st ACM Int. Conf. Multimedia (ACM MM), MM '23, Association for Computing Machinery, New York, NY, USA, 2023, pp. 3346–3354, http://dx.doi.org/10.1145/3581783.3613814.

[31] X. He, K. Cao, J. Zhang, K. Yan, Y. Wang, R. Li, C. Xie, D. Hong, M. Zhou, Pan-Mamba: Effective pan-sharpening with state space model, Inf. Fusion 115 (2025) 102779, http://dx.doi.org/10.1016/j.inffus.2024.102779.

[32] Y. Cui, P. Liu, Y. Ma, L. Chen, M. Xu, X. Guo, Pansharpening via predictive filtering with element-wise feature mixing, ISPRS J. Photogramm. Remote Sens. 219 (2025) 22–37, http://dx.doi.org/10.1016/j.isprsjprs.2024.10.029.

[33] S. Xu, J. Zhang, Z. Zhao, K. Sun, J. Liu, C. Zhang, Deep gradient projection networks for pan-sharpening, in: Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., CVPR, 2021, pp. 1366–1375, http://dx.doi.org/10.1109/CVPR46437.2021.00142.

[34] H. Zhang, J. Ma, GTP-PNet: A residual learning network based on gradient transformation prior for pansharpening, ISPRS J. Photogramm. Remote Sens. 172 (2021) 223–239, http://dx.doi.org/10.1016/j.isprsjprs.2020.12.014.

[35] T. Thanh Nhat Mai, E.Y. Lam, C. Lee, Deep unfolding tensor rank minimization with generalized detail injection for pansharpening, IEEE Trans. Geosci. Remote Sens. 62 (2024) 1–18, http://dx.doi.org/10.1109/TGRS.2024.3392215.

[36] S. Fang, H. Gan, SSUN-Net: Spatial-spectral prior-aware unfolding network for pan-sharpening, in: Proc. AAAI Conf. Artif. Intell., Vol. 39, 2025, pp. 32296–32305, http://dx.doi.org/10.1609/aaai.v39i3.32296.

[37] B. Aiazzi, S. Baronti, M. Selva, Improving component substitution pansharpening through multivariate regression of MS +pan data, IEEE Trans. Geosci. Remote Sens. 45 (10) (2007) 3230–3239, http://dx.doi.org/10.1109/TGRS.2007.901007.

[38] N. Yokoya, T. Yairi, A. Iwasaki, Coupled nonnegative matrix factorization unmixing for hyperspectral and multispectral data fusion, IEEE Trans. Geosci. Remote Sens. 50 (2) (2012) 528–537, http://dx.doi.org/10.1109/TGRS.2011.2161320.

[39] G. Vivone, L. Alparone, J. Chanussot, M. Dalla Mura, A. Garzelli, G.A. Licciardi, R. Restaino, L. Wald, A critical comparison among pansharpening algorithms, IEEE Trans. Geosci. Remote Sens. 53 (5) (2015) 2565–2586, http://dx.doi.org/10.1109/TGRS.2014.2361734.

[40] S.-Q. Deng, L.-J. Deng, X. Wu, R. Ran, D. Hong, G. Vivone, PSRT: Pyramid shuffle-and-reshuffle transformer for multispectral and hyperspectral image fusion, IEEE Trans. Geosci. Remote Sens. 61 (2023) 1–15, http://dx.doi.org/10.1109/TGRS.2023.3244750.

[41] C. Zhu, X. Song, Y. Li, S. Deng, T. Zhang, A spatial-frequency dual-domain implicit guidance method for hyperspectral and multispectral remote sensing image fusion based on Kolmogorov–Arnold Network, Inf. Fusion 123 (2025) 103261, http://dx.doi.org/10.1016/j.inffus.2025.103261.

[42] C. Zhu, S. Deng, X. Song, Y. Li, Q. Wang, Mamba collaborative implicit neural representation for hyperspectral and multispectral remote sensing image fusion, IEEE Trans. Geosci. Remote Sens. 63 (2025) 1–15, http://dx.doi.org/10.1109/TGRS.2025.3537638.

[43] Y. Zheng, J. Li, Y. Li, J. Guo, X. Wu, J. Chanussot, Hyperspectral pansharpening using deep prior and dual attention residual network, IEEE Trans. Geosci. Remote Sens. 58 (11) (2020) 8059–8076, http://dx.doi.org/10.1109/TGRS.2020.2986313.

[44] W. Wang, L.-J. Deng, R. Ran, G. Vivone, A general paradigm with detail-preserving conditional invertible network for image fusion, Int. J. Comput. Vis. 132 (2024) 1029–1054, http://dx.doi.org/10.1007/s11263-023-01924-5.

[45] J.G. Liu, Smoothing filter-based intensity modulation: A spectral preserve image fusion technique for improving spatial details, Int. J. Remote Sens. 21 (18) (2000) 3461–3472, http://dx.doi.org/10.1080/014311600750037499.

[46] W. Liao, X. Huang, F. Van Coillie, S. Gautama, A. Pižurica, W. Philips, H. Liu, T. Zhu, M. Shimoni, G. Moser, D. Tuia, Processing of multiresolution thermal hyperspectral and digital color data: Outcome of the 2014 IEEE GRSS data fusion contest, IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 8 (6) (2015) 2984–2996, http://dx.doi.org/10.1109/JSTARS.2015.2420582.

[47] C. Zhu, S. Deng, Y. Zhou, L.-J. Deng, Q. Wu, QIS-GAN: A lightweight adversarial network with quadtree implicit sampling for multispectral and hyperspectral image fusion, IEEE Trans. Geosci. Remote Sens. 61 (2023) 1–15, http://dx.doi.org/10.1109/TGRS.2023.3332176.

[48] Q. Li, Q. Wang, X. Li, Hyperspectral image super-resolution via multi-domain feature learning, in: Proc. IEEE Int. Geosci. Remote Sens. Symp., IGARSS, 2021, pp. 4135–4138, http://dx.doi.org/10.1109/IGARSS47720.2021.9554560.

[49] Q. Ma, J. Jiang, X. Liu, J. Ma, Learning a 3D-CNN and Transformer prior for hyperspectral image super-resolution, Inf. Fusion 100 (2023) 101907, http://dx.doi.org/10.1016/j.inffus.2023.101907.