HPDIFF: ZERO-SHOT BLIND HYPERSPECTRAL PANSHARPENING WITH DEGRADATION-GUIDED DIFFUSION SAMPLING

Anonymous authors
Paper under double-blind review

ABSTRACT

Hyperspectral pansharpening reconstructs a high-resolution hyperspectral image (HSI) from a low-resolution HSI (LRHS) and a high-resolution panchromatic (PAN) image. Existing methods trained with simulated degradations often rely on predefined spatial point spread functions (PSFs) and spectral response functions (SRFs). When these assumptions mismatch unknown sensor degradations, the reconstruction may fail to satisfy both observations. To address this issue, we propose HPDiff, a zero-shot blind hyperspectral pansharpening method with degradation-guided diffusion sampling. HPDiff first estimates the unknown PSF and SRF from a single test pair by matching spatially degraded PAN with spectrally projected LRHS in a shared low-resolution PAN domain. The estimated operators then define LRHS/PAN data consistency gradients to guide the reverse sampling of a pretrained low-rank Diffusion Sampling. During joint reconstruction, HPDiff further refines the degradation parameters around their initial estimates, improving observation adaptation while retaining physical plausibility. Observation Prior Guidance (OPG) integrates LRHS spectral information, PAN spatial structures, and lightweight task-information priors to regulate spectral, spatial, and edge constraints during reverse sampling. A Coupled Spatial-Spectral Low-Rank Decoder (CSLD) jointly estimates interacting spatial and spectral factors to map diffusion details into a full-band HSI. Experiments on multiple benchmarks demonstrate improved PSNR, SAM, and ERGAS, together with better preservation of spatial structures and spectral consistency. The code will be made publicly available upon acceptance.

1 INTRODUCTION

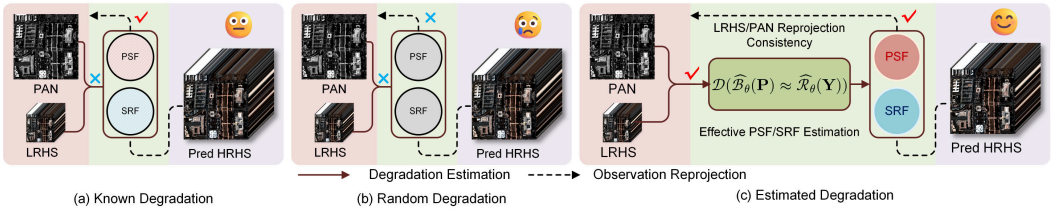


Figure 1: Comparison of three degradation settings in HSP: (a) accurate PSF/SRF operators provide reliable constraints but are generally unavailable in practice; (b) randomly specified operators cause degradation mismatch; and (c) the proposed blind estimator infers sample specific effective operators, enabling consistent reprojection of the reconstructed HRHS to the LRHS/PAN observations.

Hyperspectral images (HSIs) capture detailed spectral responses over contiguous narrow bands and support a wide range of remote sensing applications Shan et al. (2026). However, physical constraints in airborne and satellite sensors make it difficult to simultaneously achieve high spectral resolution, high spatial resolution, and wide coverage. HSIs therefore usually provide rich spectral information but coarse spatial sampling, whereas panchromatic (PAN) images record scene structures at a finer spatial resolution over a broad spectral range Ziaja et al. (2025). Hyperspectral

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.

pansharpening (HSP) aims to reconstruct a high spatial resolution HSI (HRHS) from a low spatial resolution HSI (LRHS) and a PAN image while preserving spectral fidelity Wang et al. (2026); Jiang & Chen (2025a).

The point spread function (PSF) describes spatial blur before LRHS downsampling, whereas the PAN spectral response function (SRF) projects HSI bands into a broadband PAN image Lin et al. (2024). Many recent HSP methods assume known or fixed degradation operators from the latent HRHS to the LRHS and PAN observations Lei et al. (2024). These operators are used to synthesize training pairs or impose observation fidelity during reconstruction Xu et al. (2026). In practice, however, the effective PSF and SRF vary with sensor characteristics, acquisition conditions, and registration accuracy, and are often unavailable with sufficient precision Hu et al. (2024). When the assumed operators deviate from the actual observation process, a reconstructed HRHS may contain plausible details but fail to map consistently back to the input LRHS/PAN. Such degradation mismatch weakens the observation constraint and may introduce spectral bias Liu et al. (2026). This problem is particularly severe in zero-shot HSP, where neither target references nor accurate sensor parameters are available to correct the mismatch Yang et al. (2025). As shown in Fig. 1, accurate degradation operators are difficult to obtain in practice, while HPDiff estimates effective PSF/SRF operators from the current observations to avoid degradation mismatch.

To address this problem, we propose HPDiff, a zero-shot blind HSP method based on degradation-guided diffusion sampling. To the best of our knowledge, HPDiff is the first HSP framework to explicitly estimate blind degradation parameters. It estimates effective PSF and SRF through low resolution PAN consistency and uses them as plug and play observation operators during reverse diffusion. Observation Prior Guidance adapts sampling to the spatial spectral content and task priors of the current observations, while the Coupled Spatial Spectral Low Rank Decoder combines diffusion details with LRHS spectral information to reconstruct the full band HRHS. During test time refinement, the degradation operators and decoder are jointly optimized under observation and physical constraints.

Our contributions are summarized as follows:

- We propose HPDiff, the first blind parameter estimation framework for HSP to our knowledge, enabling zero-shot reconstruction without paired training data, target HRHS references, or known degradation parameters.
- A physically constrained and plug-and-play blind PSF/SRF estimator is designed to estimate effective degradation operators through low resolution PAN consistency.
- Observation Prior Guidance is introduced to adapt reverse diffusion to the spectral information, spatial structures, and lightweight task priors of the current observations.
- A Coupled Spatial Spectral Low Rank Decoder is developed to couple diffusion details with LRHS spectral factors for full-band HRHS reconstruction.

2 Related Work

2.1 Hyperspectral Pansharpening

Conventional HSP methods follow component substitution, model based optimization, or multiresolution analysis Vivone et al. (2021; 2015); Aiazzi et al. (2006); He et al. (2014), while deep networks learn nonlinear LRHS/PAN-to-HRHS mappings from synthetic pairs Yang et al. (2017); Yuan et al. (2018); Deng et al. (2021); Guan & Lam (2022); Zini et al. (2024). Although recent architectures improve spatial detail recovery, their training pairs and observation constraints usually rely on predefined degradation operators Shang et al. (2024); Yang et al. (2026); Dong et al. (2022), which can limit transfer to sensors with unknown PSF/SRF responses.

2.2 Diffusion Models for Inverse Problems

Diffusion models provide strong generative priors Ho et al. (2020); Song et al. (2021), and posterior sampling adapts them to inverse problems by injecting observation fidelity during reverse sampling Chung et al. (2023); Rout et al. (2023); Li & Wang (2025). In HSI restoration and fusion, such priors support zero-shot detail recovery and subspace reconstruction Cao et al. (2024); Rui et al. (2024a);

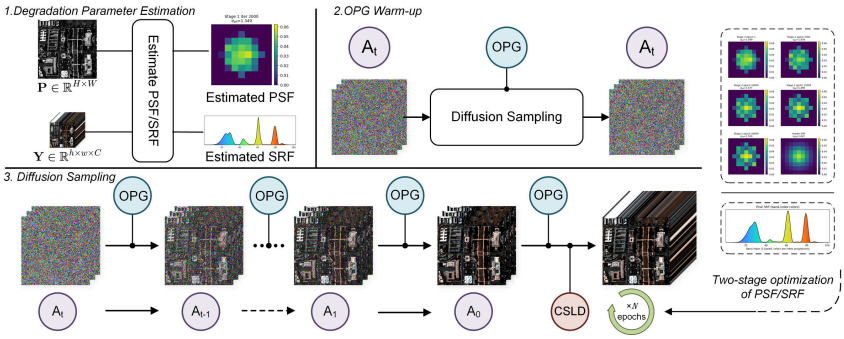


Figure 2: HPDiff estimates effective PSF/SRF operators from each LRHS/PAN pair, uses them with observation prior guidance to constrain low-dimensional reverse diffusion, and reconstructs the HRHS through a coupled low-rank decoder.

Zhu et al. (2025). However, the likelihood guidance still depends on the adopted forward operator, so inaccurate PSF/SRF assumptions can bias the reverse trajectory Liu et al. (2025); Jiang & Chen (2025b).

2.3 Zero-shot and Blind HSI Fusion

Zero-shot fusion avoids paired target-sensor training data through coupled factorization, low-rank modeling, or test-time adaptation Yokoya et al. (2012); Liu et al. (2021); Rui et al. (2024b). Recent diffusion based methods such as PLRDiff and DM-ZS further recover spatial details from pretrained priors while adapting to the current observations Xiao et al. (2025); Guarino et al. (2025). These methods improve generalization, but they typically retain fixed degradation forms. HPDiff instead estimates sample-specific PSF/SRF operators and embeds them into diffusion guidance and low-rank reconstruction.

3 Methodology

3.1 Problem Formulation

Given an LRHS $\mathbf{Y} \in \mathbb{R}^{h\times w\times C}$ and a PAN image $\mathbf{P} \in \mathbb{R}^{H\times W}$, HSP recovers an HRHS $\mathbf{X} \in \mathbb{R}^{H\times W\times C}$, where C is the number of spectral bands and $H/h = W/w$ is the resolution ratio. The observation process is modeled as

$$\mathbf{Y} = \mathcal{D}(\mathcal{B}(\mathbf{X})) + n_y, \qquad \mathbf{P} = \mathcal{R}(\mathbf{X}) + n_p, \tag{1}$$

where $\mathcal{B}(\cdot)$ is the band-independent blur induced by the PSF, $\mathcal{D}(\cdot)$ is downsampling, and $\mathcal{R}(\cdot)$ projects the spectral bands using the SRF. The terms n_y and n_p denote additive noise. In our blind setting, both $\mathcal{B}(\cdot)$ and $\mathcal{R}(\cdot)$ are unknown, and neither sensor calibration nor a target HRHS is available.

3.2 Overview of HPDiff

Fig. 2 summarizes the three stages of HPDiff. First, effective PSF and SRF operators are estimated from the current LRHS/PAN pair through low-resolution PAN consistency. Second, a pretrained RGB diffusion prior recovers low-dimensional spatial details, while the estimated operators and observation priors guide reverse sampling Bandara et al. (2025). Third, a coupled spatial-spectral low-rank decoder combines the recovered details with PAN structures and LRHS spectra to reconstruct the full-band HRHS.

3.3 Blind Spatial-Spectral Degradation Estimation

The unknown physical degradations cannot be uniquely identified from one observation pair. We therefore estimate sample-consistent effective operators $\widehat{\mathcal{B}}_\theta(\cdot)$ and $\widehat{\mathcal{R}}_\theta(\cdot)$. Since band-independent

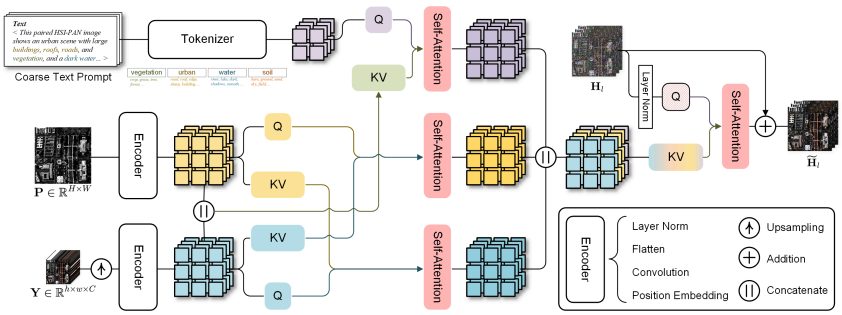


Figure 3: Observation Prior Guidance encodes LRHS, PAN, and an automatically generated task cue, performs task modulated observation interaction, and adaptively balances spectral, PAN, edge, and smoothness constraints during reverse diffusion.

spatial degradation commutes with linear spectral projection, the two observation paths should produce the same low-resolution PAN:

$$\mathcal{D}(\mathcal{B}(\mathbf{P})) \simeq \mathcal{R}_{\mathbf{r}}(\mathbf{Y}). \tag{2}$$

PSF parameterization. Let $\mathbf{U} = [u_{ij}] \in \mathbb{R}^{k_B \times k_B}$ denote the learnable parameters of a spatial kernel. The estimated kernel $\mathbf{b} = [b_{ij}]$ is parameterized using softmax:

$$b_{ij} = \frac{\exp(u_{ij})}{\sum_{m,n} \exp(u_{mn})}, \qquad \widehat{\mathcal{B}}_\theta(\mathbf{I}) = \mathbf{b} * \mathbf{I}, \tag{3}$$

where θ contains $\mathbf{U}$, and convolution is applied independently to each spectral band. This parameterization guarantees $b_{ij} \geq 0$ and $\sum_{i,j} b_{ij} = 1$, thereby enforcing non-negativity and energy preservation while retaining the flexibility to represent a non-predefined kernel shape.

To suppress noisy or spatially shifted kernels, smoothness and center priors are imposed:

$$\ell_{\mathcal{B}} = \eta_s \sum_{i,j} \left[(b_{i+1,j} - b_{ij})^2 + (b_{i,j+1} - b_{ij})^2 \right] + \eta_c \sum_{i,j} d_{ij}^2 b_{ij}. \tag{4}$$

The spatial differences are summed over all valid neighboring kernel positions, and d_{ij} denotes the normalized distance from position (i, j) to the kernel center.

SRF parameterization. Let $u_c \in [0, 1]$ denote the normalized spectral position of the cth hyperspectral band. The center μ and effective width ω of the PAN response window are parameterized as

$$\mu = \sigma(\alpha), \qquad \omega = \omega_{\min} + (1 - \omega_{\min})\sigma(\beta), \tag{5}$$

where $\sigma(\cdot)$ is the sigmoid function and $\omega_{\min} > 0$ prevents the response window from collapsing.

A continuous spectral window is then constructed as

$$g_c = \sigma\left(\frac{\omega/2 - |u_c - \mu|}{\tau} \right), \tag{6}$$

where τ controls the softness of the window boundaries. This window assigns larger responses to bands inside the learned support while smoothly attenuating bands outside it.

Let $a_c = \mathrm{softplus}(v_c)$ denote the nonnegative response amplitude of band c. The estimated PAN response and its spectral projection operator are defined as

$$r_{\theta,c} = \frac{a_c g_c}{\sum_{j=1}^{C} a_j g_j}, \qquad \widehat{\mathcal{R}}_{\mathbf{r}_\theta}(\mathbf{G}) = \sum_{c=1}^{C} r_{\theta,c} \mathbf{G}_c. \tag{7}$$

Here, θ additionally contains α, β, and $\{v_c\}_{c=1}^{C}$. The resulting response satisfies $r_{\theta,c} \geq 0$ and $\sum_{c=1}^{C} r_{\theta,c} = 1$. It jointly estimates the response center, effective width, and band amplitudes while maintaining contiguous spectral support.

The SRF regularizer is defined as

$$\ell_{\mathcal{R}} = \gamma_s \sum_{c=2}^{C-1} \left(r_{\theta,c+1} - 2r_{\theta,c} + r_{\theta,c-1}\right)^2 + \frac{\gamma_w}{C} \sum_{c=1}^{C} g_c. \tag{8}$$

The first term penalizes second-order spectral differences to suppress oscillatory responses, whereas the second encourages a compact response window. Its continuity is maintained by the soft-window parameterization in Eq. 6.

Collecting $\mathbf{U}$, α, β, and $\{v_c\}$ in θ, we optimize

$$\mathcal{L}_{\text{blind}}(\theta) = \frac{1}{N} \left\| \mathcal{D}\Big(\widehat{\mathcal{B}}_\theta(\mathbf{P})\Big) - \widehat{\mathcal{R}}_\theta(\mathbf{Y}) \right\|_F^2 + \lambda_{\mathcal{B}} \ell_{\mathcal{B}} + \lambda_{\mathcal{R}} \ell_{\mathcal{R}}, \tag{9}$$

where $N = hw$ and the first term implements Eq. 2. The solution $\hat{\theta}$ yields $\widehat{\mathcal{B}}_{\hat{\theta}}(\cdot)$ and $\widehat{\mathcal{R}}_{\hat{\theta}}(\cdot)$, interpreted as effective operators consistent with the current observations.

3.4 Diffusion-based Zero-shot Reconstruction

We employ a pretrained diffusion prior over low-dimensional spatial details. Let $\mathbf{x}_t$ be the noisy state at step t and $\mathbf{Z}$ the observation conditions constructed in the next subsection. The denoiser predicts the clean detail representation

$$\widehat{\mathbf{A}}_{0|t} = \frac{\mathbf{x}_t - \sqrt{1-\bar{\alpha}_t}\, \epsilon_\phi(\mathbf{x}_t, t; \mathbf{Z})}{\sqrt{\bar{\alpha}_t}}. \tag{10}$$

where $\bar{\alpha}_t$ denotes the cumulative noise schedule and ϵ_ϕ is the pretrained denoiser.

To evaluate physical constraints, this detail estimate is temporarily lifted to the HSI space. Let $\mathbf{Y}_{(3)}$ denote the spectral unfolding of LRHS, and let $\mathbf{S}$ contain its selected representative bands. The spectral lifting matrix and temporary HSI estimate are given by

$$\widehat{\mathbf{X}}_{t,(3)} = \mathbf{Y}_{(3)} \mathbf{S}^\top (\mathbf{S}\mathbf{S}^\top + \varepsilon \mathbf{I})^{-1} \widehat{\mathbf{A}}_{0|t,(3)} + \delta. \tag{11}$$

Here, δ is the interpolated residual of the low rank LRHS approximation.

Using the estimated degradations, we reproject it as

$$\widehat{\mathbf{Y}}_t = \mathcal{D}\Big(\widehat{\mathcal{B}}_{\hat{\theta}}(\widehat{\mathbf{X}}_t)\Big), \quad \widehat{\mathbf{P}}_t = \widehat{\mathcal{R}}_{\hat{\theta}}(\widehat{\mathbf{X}}_t), \tag{12}$$

and define the basic observation loss

$$\mathcal{L}_{\text{obs}}^t = \|\mathbf{Y} - \widehat{\mathbf{Y}}_t\|_F + \lambda_P \|\mathbf{P} - \widehat{\mathbf{P}}_t\|_F. \tag{13}$$

Treating $\mathcal{L}_{\text{obs}}^t$ as a negative log likelihood, its gradient approximates the likelihood score used to correct reverse diffusion sampling. The reverse posterior is first sampled and then corrected by the observation gradient:

$$\widetilde{\mathbf{x}}_{t-1} = \mu_t(\mathbf{x}_t, \widehat{\mathbf{A}}_{0|t}) + \sigma_t \mathbf{z}, \quad \mathbf{z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}), \quad \overline{\mathbf{x}}_{t-1} = \widetilde{\mathbf{x}}_{t-1} - s_t \nabla_{\mathbf{x}_t} \mathcal{L}_{\text{obs}}^t. \tag{14}$$

where μ_t and σ_t are the reverse posterior mean and standard deviation, and s_t controls the correction. Thus, $\widetilde{\mathbf{x}}_{t-1}$ denotes the unconstrained reverse sample, whereas $\overline{\mathbf{x}}_{t-1}$ is its observation corrected counterpart.

3.5 Observation Prior Guidance

As shown in Fig. 3, observation prior guidance (OPG) adapts the diffusion prior through condition injection and posterior correction. Given the current LRHS/PAN pair, a manually designed coarse-grained task cue q is used to summarize the observation setting. A lightweight rule parser converts q into nonnegative category weights $\{\pi_m(q)\}_{m=1}^{M_q}$ that sum to one. The three conditions are encoded as

$$\mathbf{Z}_Y = \Phi_Y(\mathcal{U}(\mathbf{Y})), \quad \mathbf{Z}_P = \Phi_P(\mathbf{P}), \quad \mathbf{Z}_q = \Phi_q\left(\sum_{m=1}^{M_q} \pi_m(q) \mathbf{U}_m\right), \tag{15}$$

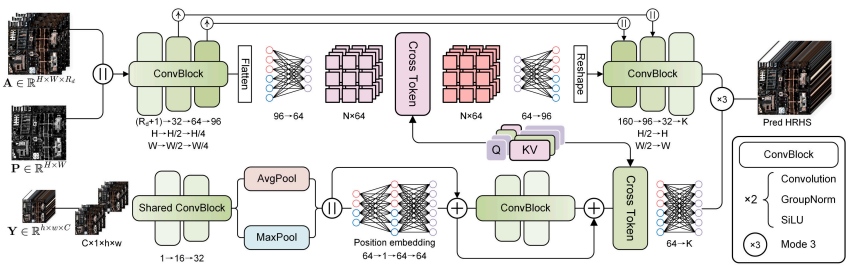


Figure 4: The coupled spatial-spectral low-rank decoder exchanges detail/PAN spatial tokens and LRHS spectral tokens through shared tokens, then reconstructs HRHS from the resulting spatial and spectral factors.

where $\mathcal{U}$ upsamples LRHS, Φ_Y and Φ_P are convolutional token encoders, and $\mathbf{U}_m$ is the task token bank for category m.

Let $\mathcal{A}(\mathbf{Q}, \mathbf{V})$ denote cross attention with queries $\mathbf{Q}$ and key/value tokens $\mathbf{V}$. The task tokens first aggregate evidence from LRHS and PAN, and their pooled representation generates modality gates $\mathbf{g}_i$. These gates regulate the bidirectional interaction between observation tokens:

$$\overline{\mathbf{Z}}_q = \mathbf{Z}_q + \kappa_q \mathcal{A}(\mathbf{Z}_q, [\mathbf{Z}_Y; \mathbf{Z}_P]), \quad \overline{\mathbf{Z}}_i = \mathbf{Z}_i + \kappa_i \mathbf{g}_i \odot \mathcal{A}(\mathbf{Z}_i, \mathbf{Z}_j), \tag{16}$$

where $(i, j) \in \{(Y, P), (P, Y)\}$ and κ_q, κ_i are learnable residual scales. The refined conditions are concatenated as $\mathbf{Z} = [\overline{\mathbf{Z}}_Y; \overline{\mathbf{Z}}_P; \overline{\mathbf{Z}}_q]$ and injected into the lth denoiser block:

$$\widetilde{\mathbf{H}}_l = \mathbf{H}_l + \mathcal{A}(\mathrm{Norm}(\mathbf{H}_l), \mathbf{Z}) \,. \tag{17}$$

Here, $\mathbf{H}_l$ and $\widetilde{\mathbf{H}}_l$ denote the original and condition enhanced features, respectively. Before sampling, the diffusion backbone is frozen, while the lightweight condition modules are adapted to the current observation pair for a few warm-up iterations.

To emphasize PAN-consistent spatial structures, OPG augments the basic observation loss with an edge consistency term:

$$\mathcal{L}^t_{\mathrm{opg}} = \mathcal{L}^t_{\mathrm{obs}} + \lambda_E \|\mathcal{S}(\mathbf{P}) - \mathcal{S}(\widehat{\mathbf{P}}_t)\|_1, \tag{18}$$

where $\mathcal{S}$ denotes the Sobel operator. This term encourages the sampled reconstruction to preserve high-frequency structures consistent with the PAN observation. Finally, OPG corrects the observation guided reverse sample by

$$\mathbf{x}_{t-1} = \overline{\mathbf{x}}_{t-1} - s_t \lambda_{\mathrm{opg}} \nabla_{\mathbf{x}_t} \mathcal{L}^t_{\mathrm{opg}}. \tag{19}$$

Thus, OPG acts through both conditional denoising and adaptive posterior correction.

3.6 Coupled Spatial–Spectral Low–Rank Decoder

As shown in Fig. 4, the coupled spatial–spectral low–rank decoder (CSLD) provides the final full-band reconstruction and participates in subsequent refinement sampling. Given the clean diffusion detail $\widehat{\mathbf{A}}_0$, its spatial and spectral branches form

$$\mathbf{Q}_s = E_s([\widehat{\mathbf{A}}_0; \mathbf{P}]), \qquad \mathbf{Q}_\lambda = E_\lambda(\mathbf{Y}), \tag{20}$$

where E_s extracts multiscale structures from the diffusion detail and PAN, while E_λ encodes LRHS bands and their spectral positions.

Learnable shared tokens $\mathbf{T}_0$ sequentially aggregate spatial and spectral information and propagate the coupled representation back to both branches:

$$\mathbf{T}_s = \mathcal{A}(\mathbf{T}_0, \mathbf{Q}_s), \quad \mathbf{T}_c = \mathcal{A}(\mathbf{T}_s, \mathbf{Q}_\lambda), \quad \widetilde{\mathbf{Q}}_s = \mathcal{A}(\mathbf{Q}_s, \mathbf{T}_c), \quad \widetilde{\mathbf{Q}}_\lambda = \mathcal{A}(\mathbf{Q}_\lambda, \mathbf{T}_c). \tag{21}$$

The decoded tokens yield a spatial factor $\mathbf{F}_s \in \mathbb{R}^{\rho \times H \times W}$ and a normalized spectral factor $\mathbf{F}_\lambda \in \mathbb{R}^{C \times \rho}$, where ρ denotes the decomposition rank. Their low-rank synthesis is

$$\widehat{\mathbf{X}}_{cij} = \sum_{k=1}^{\rho} (\mathbf{F}_\lambda)_{ck} (\mathbf{F}_s)_{kij}. \tag{22}$$

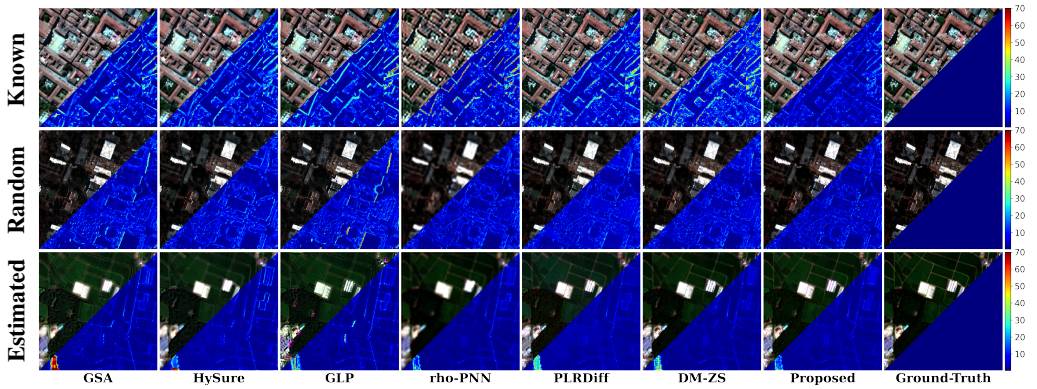


Figure 5: Qualitative comparison under three degradation settings: the rows correspond from top to bottom to Known degradation on Pavia Center, Random degradation on Houston, and Estimated degradation on Chikusei. In each result, the upper-left region shows the RGB pseudo-color image, while the lower-right region shows the SAM residual map in degrees. Darker residuals indicate better spectral fidelity.

Let ψ denote the decoder parameters. CSLD is optimized on the current observation pair through

$$\mathcal{L}_{\mathrm{CSLD}} = \|\mathbf{Y} - \mathcal{D}(\widehat{\mathcal{B}}_\theta(\widehat{\mathbf{X}}))\|_F + \lambda_P \|\mathbf{P} - \widehat{\mathcal{R}}_\theta(\widehat{\mathbf{X}})\|_F + \frac{\lambda_o}{\rho^2}\|\mathbf{F}_\lambda^\top \mathbf{F}_\lambda - \mathbf{I}\|_F^2. \tag{23}$$

The first two terms enforce LRHS/PAN consistency, while the last discourages redundant spectral factors. During refinement, the effective degradation operators are jointly fine-tuned around their initial estimates under the physical constraints described. All parameters are optimized only from the current LRHS/PAN pair, without target HRHS supervision.

4 Experiments

4.1 Experimental Settings

Reduced-resolution experiments use 8 Pavia Center samples and 10 samples each from Houston and Chikusei, all cropped without overlap at 256×256 pixels Vivone et al. (2023). Following the protocol from Wald et al. (1997), LRHS images are generated using a 9×9 Gaussian kernel with standard deviation 1.6986 and a downsampling factor of 4, while PAN images are obtained by averaging all spectral bands. The synthesis PSF and SRF are hidden under the blind setting. Full-resolution evaluation uses a 240×240 subimage from the real PRISMA FR1 scene without synthetic degradation Ziaja et al. (2023).

Six representative methods are selected for comparison. The conventional methods include GSA, HySure, and GLP. The deep learning method rho-PNN constructs its spatial degradation constraint using a fixed modulation transfer function kernel. We further compare two zero-shot diffusion methods, PLRDiff and DM-ZS, which perform reconstruction without paired HRHS training samples from the target scene.

All experiments are implemented in PyTorch on an RTX 4090 GPU with an Intel i9-14900KF CPU. Adam is used for blind PSF/SRF estimation, OPG warm-up, and CSLD optimization for 2000, 200, and 20000 iterations, with learning rates of 5×10^{-2}, 10^{-4}, and 10^{-3}, respectively. Reverse diffusion uses 200 steps, while the degradation operators are jointly refined with CSLD at a learning rate of 10^{-4}.

4.2 Quantitative Comparison

Tab. 1 evaluates the methods under known degradations. HPDiff achieves the best results on all 18 metrics across the three datasets. Compared with the strongest competitors, it improves PSNR by 1.383, 1.069, and 0.772 dB, while reducing SAM and ERGAS on Pavia Center, Houston, and

Table 1: Quantitative comparison under the known degradation setting on Pavia Center, Houston, and Chikusei. The best results are highlighted in bold.

Method	Pavia Center						Houston						Chikusei					
	PSNR↑	RMSE↓	SAM↓	ERGAS↓	SSIM↑	CC↑	PSNR↑	RMSE↓	SAM↓	ERGAS↓	SSIM↑	CC↑	PSNR↑	RMSE↓	SAM↓	ERGAS↓	SSIM↑	CC↑
GSA	26.341	5.053	10.997	11.029	0.793	0.971	32.619	2.425	8.165	7.909	0.871	0.974	21.477	9.219	11.672	35.735	0.612	0.978
HySure	25.211	5.872	12.278	12.664	0.741	0.958	31.337	2.850	7.616	8.319	0.865	0.972	24.555	6.155	8.706	24.653	0.717	0.987
GLP	24.950	6.266	11.119	12.385	0.787	0.971	32.857	2.428	7.327	7.679	0.870	0.972	21.864	11.525	5.403	16.419	0.681	0.988
rho-PNN	28.466	3.818	8.248	6.897	0.805	0.936	33.199	2.273	7.397	5.592	0.851	0.941	36.077	1.585	3.704	6.422	0.902	0.976
PLRDiff	33.531	2.132	8.388	4.315	0.916	0.978	37.851	1.331	6.241	3.560	0.928	0.979	38.955	1.136	3.476	4.771	0.937	0.988
DM-ZS	33.796	2.075	7.056	3.979	0.922	0.979	38.212	1.263	5.675	3.233	0.928	0.981	38.572	1.194	3.384	4.339	0.935	0.986
Proposed	**35.179**	**1.767**	**5.937**	**3.426**	**0.943**	**0.985**	**39.281**	**1.133**	**4.863**	**2.838**	**0.942**	**0.985**	**39.727**	**1.041**	**2.662**	**3.630**	**0.951**	**0.990**

Table 2: Quantitative comparison under blind degradations. * denotes randomly specified degradation operators, while † denotes the plug-and-play use of the degradation operators estimated by HPDiff. The best results are highlighted in bold.

Method	Pavia Center						Houston						Chikusei					
	PSNR↑	RMSE↓	SAM↓	ERGAS↓	SSIM↑	CC↑	PSNR↑	RMSE↓	SAM↓	ERGAS↓	SSIM↑	CC↑	PSNR↑	RMSE↓	SAM↓	ERGAS↓	SSIM↑	CC↑
rho-PNN*	28.289	3.896	8.284	7.345	0.783	0.929	32.283	2.490	6.252	6.080	0.826	0.927	35.437	1.691	5.036	7.786	0.853	0.973
PLRDiff*	32.169	2.507	8.085	4.785	0.897	0.971	35.458	1.771	6.945	4.428	0.892	0.966	34.383	1.930	6.133	10.590	0.813	0.967
DM-ZS*	31.432	2.726	8.729	5.207	0.880	0.966	34.933	1.898	7.553	4.791	0.875	0.959	33.823	2.064	6.404	10.443	0.802	0.962
Proposed*	32.360	2.430	8.770	4.745	0.903	0.973	35.884	1.671	6.837	4.375	0.906	0.971	35.754	1.644	4.733	8.231	0.862	0.976
rho-PNN†	28.552	3.785	6.918	7.041	0.772	0.930	32.984	2.323	5.480	5.665	0.831	0.936	36.573	1.492	**3.157**	6.451	0.907	0.979
PLRDiff†	33.128	2.229	8.164	4.365	0.916	0.976	37.385	1.393	6.285	3.691	0.923	0.977	37.481	1.347	3.565	5.426	0.920	0.983
DM-ZS†	32.973	2.272	6.986	4.246	0.911	0.976	37.199	1.407	5.928	3.622	0.917	0.976	37.084	1.414	3.625	5.562	0.918	0.981
Proposed†	**34.656**	**1.871**	**6.016**	**3.544**	**0.938**	**0.983**	**38.080**	**1.282**	**5.101**	**3.231**	**0.931**	**0.981**	**37.917**	**1.279**	3.708	**5.324**	**0.925**	**0.985**

Chikusei, respectively. These results demonstrate the effectiveness of its diffusion prior, observation guidance, and coupled low-rank decoding. Under randomly perturbed degradation operators, all methods suffer performance degradation, with HPDiff losing 2.819, 3.397, and 3.973 dB in PSNR on the three datasets. This confirms that inaccurate PSF/SRF assumptions adversely affect both spatial and spectral reconstruction. With the estimated operators denoted by †, 70 of 72 metric results improve over random degradation without changing the reconstruction architectures. HPDiff further gains 2.296, 2.196, and 2.163 dB in PSNR and achieves the best result on 17 of 18 metrics, except for SAM on Chikusei. These results demonstrate the effectiveness and plug-and-play generalization of the proposed degradation estimator.

4.3 Qualitative Comparison

The top row of Fig. 5 shows the results on Pavia Center under known degradation. Conventional methods retain the main structures but leave visible residuals, whereas learning-based methods recover sharper details. HPDiff produces clearer boundaries and lower SAM residuals, consistent with its quantitative advantage. The middle row presents the Houston results under randomly specified PSF/SRF operators. Degradation mismatch causes visible blurring and dense residuals for the competing methods. HPDiff better preserves building contours and reduces high-error regions, indicating improved robustness to inaccurate observation operators. The bottom row shows the Chikusei results using the estimated operators. The operators are applied to all methods without changing their reconstruction architectures. HPDiff achieves the closest appearance to the ground truth and fewer residuals, demonstrating the benefit of sample-specific degradation estimation and diffusion guidance.

4.4 Ablation Study

Tab. 3 verifies the effectiveness of OPG and CSLD across three metrics. OPG alone consistently improves PSNR and reduces RMSE and SAM on all datasets, although its isolated gains are modest. CSLD provides larger improvements, particularly on Pavia Center and Houston, demonstrating its effectiveness in reconstructing the full hyperspectral space from low-dimensional diffusion details. Combining both modules achieves the best results across all datasets and metrics, improving PSNR by 1.826, 1.383, and 0.865 dB while reducing SAM by 2.360°, 1.349°, and 0.893° over the

Table 3: Component ablation on Pavia Center, Houston, and Chikusei. The best results are highlighted in bold.

Case	OPG	CSLD	Pavia Center			Houston			Chikusei		
			PSNR↑	RMSE↓	SAM↓	PSNR↑	RMSE↓	SAM↓	PSNR↑	RMSE↓	SAM↓
1	×	×	33.353	2.175	8.297	37.898	1.324	6.212	38.862	1.152	3.555
2	✓	×	33.388	2.170	8.071	38.240	1.263	5.974	38.945	1.137	3.482
3	×	✓	34.527	1.905	6.261	39.123	1.157	5.008	38.980	1.131	2.850
4	✓	✓	**35.179**	**1.767**	**5.937**	**39.281**	**1.133**	**4.863**	**39.727**	**1.041**	**2.662**

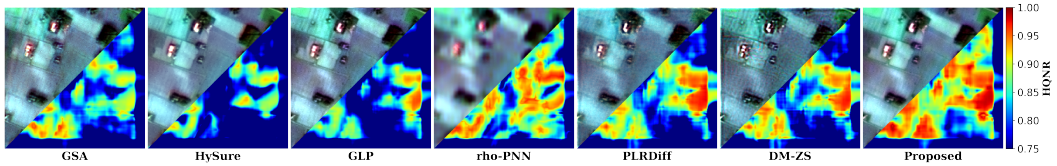


Figure 6: Visual comparison on the real PRISMA FR1 dataset. The upper-left region shows RGB pseudo-color image, while the lower-right region shows the HQNR maps, where warmer colors indicate higher reconstruction quality.

baseline. These consistent gains indicate that observation-guided diffusion and coupled low-rank reconstruction are complementary.

4.5 FR1 Real-World Dataset Evaluation.

Table 4: Full-resolution comparison on the real FR1 dataset.

Metric	GSA	HySure	GLP	rho-PNN*	PLRDiff*	DM-ZS*	Proposed†
HQNR↑	0.8206	0.7291	0.7563	0.7513	0.8424	0.8450	**0.8840**
$D_\lambda \downarrow$	0.0420	0.1546	**0.0363**	0.1427	0.0532	0.0414	0.0940
$D_S \downarrow$	0.1434	0.1376	0.2153	0.1236	0.1103	0.1223	**0.0243**

Fig. 6 presents the real FR1 results. GSA, HySure, and GLP produce smoother structures, while rho-PNN exhibits visible blurring. PLRDiff and DM-ZS recover sharper details but show uneven HQNR responses. The proposed method yields clearer boundaries and more continuous high-HQNR regions, indicating better spatial-spectral consistency on the real scene. Tab. 4 shows that HPDiff achieves the highest HQNR and lowest D_S on the real FR1 dataset. This indicates that the estimated PSF/SRF operators effectively adapt the observation constraints to unknown degradations, yielding the best overall balance between spatial and spectral consistency.

5 Conclusion

We presented HPDiff, a zero-shot blind HSP method that estimates sample-specific PSF/SRF operators and uses them to guide diffusion-based reconstruction. Experiments under reduced-resolution, blind-degradation, and real-world settings demonstrate consistent improvements in spatial detail recovery, spectral fidelity, and degradation adaptation. Future work will investigate cross-sensor registration for more robust practical deployment.

Statement

AI assisted with literature research and the writing process. Final analysis, interpretation, and editing were independently completed and verified by the authors.

REFERENCES

B. Aiazzi, L. Alparone, S. Baronti, A. Garzelli, and M. Selva. Mtf-tailored multiscale fusion of high-resolution ms and pan imagery. *Photogrammetric Engineering and Remote Sensing*, 72(5): 591–596, 2006. doi: 10.14358/PERS.72.5.591.

Wele Gedara Chaminda Bandara, Nithin Gopalakrishnan Nair, and Vishal Patel. Ddpm-cd: Denoising diffusion probabilistic models as feature extractors for remote sensing change detection. In *Proceedings of the Winter Conference on Applications of Computer Vision*, pp. 5250–5262, 2025.

Zihan Cao, Shiqi Cao, Liang-Jian Deng, Xiao Wu, Junming Hou, and Gemine Vivone. Diffusion model with disentangled modulations for sharpening multispectral and hyperspectral images. *Information Fusion*, 104:102158, 2024. doi: 10.1016/j.inffus.2023.102158.

Hyungjin Chung, Jeongsol Kim, Michael T. Mccann, Marc Klasky, and Jong Chul Ye. Diffusion posterior sampling for general noisy inverse problems. In *International Conference on Learning Representations*, 2023.

Liang-Jian Deng, Gemine Vivone, Cheng Jin, and Jocelyn Chanussot. Detail injection-based deep convolutional neural networks for pansharpening. *IEEE Transactions on Geoscience and Remote Sensing*, 59(8):6995–7010, 2021. doi: 10.1109/TGRS.2020.3031366.

Wenqian Dong, Tongzhen Zhang, Jiahui Qu, Song Xiao, Jie Liang, and Yunsong Li. Laplacian pyramid dense network for hyperspectral pansharpening. *IEEE Transactions on Geoscience and Remote Sensing*, 60:1–13, 2022. doi: 10.1109/TGRS.2021.3076768.

Peiyan Guan and Edmund Y. Lam. Multistage dual-attention guided fusion network for hyperspectral pansharpening. *IEEE Transactions on Geoscience and Remote Sensing*, 60:1–14, 2022. doi: 10.1109/TGRS.2021.3114552.

Giuseppe Guarino, Matteo Ciotola, Gemine Vivone, Giovanni Poggi, and Giuseppe Scarpa. Zero-shot hyperspectral pansharpening using hysteresis-based tuning for spectral quality control. *IEEE Transactions on Geoscience and Remote Sensing*, 63:1–19, 2025. doi: 10.1109/TGRS.2025.3583877.

Xiyan He, Laurent Condat, Jose M. Bioucas-Dias, Jocelyn Chanussot, and Junshi Xia. A new pansharpening method based on spatial and spectral sparsity priors. *IEEE Transactions on Image Processing*, 23(9):4160–4174, 2014. doi: 10.1109/TIP.2014.2333661.

Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. In *Advances in Neural Information Processing Systems 33*, pp. 6840–6851, 2020.

Baiyang Hu, Xiaodian Zhang, and Kun Gao. Unsupervised blind hyperspectral super-resolution for unregistered images. In *IGARSS 2024 - 2024 IEEE International Geoscience and Remote Sensing Symposium*, pp. 9349–9352, 2024. doi: 10.1109/IGARSS53475.2024.10642880.

Hongcheng Jiang and ZhiQiang Chen. Hyperspectral pansharpening with transformer-based spectral diffusion priors. In *2025 IEEE/CVF Winter Conference on Applications of Computer Vision Workshops (WACVW)*, pp. 544–553, 2025a. doi: 10.1109/WACVW65960.2025.00066.

Hongcheng Jiang and Zhiqiang Chen. Transformer-based diffusion and spectral priors model for hyperspectral pansharpening. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 18:18962–18977, 2025b. doi: 10.1109/JSTARS.2025.3590685.

Dajiang Lei, Genyuan Zhang, Junming Wang, Qun Liu, Weisheng Li, and Liping Zhang. Nonlocal deep unfolding pansharpening method based on degradation kernel estimation. *IEEE Transactions on Geoscience and Remote Sensing*, 62:1–14, 2024. doi: 10.1109/TGRS.2024.3454094.

Ji Li and Chao Wang. Efficient diffusion posterior sampling for noisy inverse problems. *SIAM Journal on Imaging Sciences*, 18(2):1468–1492, 2025. doi: 10.1137/24M1688321.

Hong Lin, Da Huang, Jun Li, Tong Zhou, Jian Long, and Yuanxi Peng. Semi-blind hyperspectral and multispectral image fusion via generalized inverse matrix optimization. *Infrared Physics and Technology*, 139:105316, 2024. doi: 10.1016/j.infrared.2024.105316.

Jianjun Liu, Lian Cheng, Mengjie Dong, Zebin Wu, and Liang Xiao. Causal degradation-guided network with spatial-frequency attention for blind hyperspectral image fusion. *IEEE Transactions on Geoscience and Remote Sensing*, 64:5517317–5517317, 2026. doi: 10.1109/TGRS.2026.3703367.

Na Liu, Lu Li, Wei Li, Ran Tao, James E. Fowler, and Jocelyn Chanussot. Hyperspectral restoration and fusion with multispectral imagery via low-rank tensor-approximation. *IEEE Transactions on Geoscience and Remote Sensing*, 59(9):7817–7830, 2021. doi: 10.1109/TGRS.2020.3049014.

Sen Liu, Jiahui Qu, Song Xiao, Wenqian Dong, and Yunsong Li. Learning generalization from various unaware degradations for blind hyperspectral image super-resolution via transparent diffusion model. *IEEE Transactions on Multimedia*, 27:9316–9329, 2025. doi: 10.1109/TMM.2025.3613121.

Litu Rout, Negin Raoof, Giannis Daras, Constantine Caramanis, Alex Dimakis, and Sanjay Shakkottai. Solving linear inverse problems provably via posterior sampling with latent diffusion models. In *Advances in Neural Information Processing Systems 36*, pp. 49960–49990, 2023. doi: 10.52202/075280-2174.

Xiangyu Rui, Xiangyong Cao, Yining Li, and Deyu Meng. Variational zero-shot multispectral pansharpening. *IEEE Transactions on Geoscience and Remote Sensing*, 62:1–16, 2024a. doi: 10.1109/TGRS.2024.3492059.

Xiangyu Rui, Xiangyong Cao, Li Pang, Zeyu Zhu, Zongsheng Yue, and Deyu Meng. Unsupervised hyperspectral pansharpening via low-rank diffusion model. *Information Fusion*, 107:102325, 2024b. doi: 10.1016/j.inffus.2024.102325.

Linsong Shan, Zecan Yang, Laurence T. Yang, Changlong Li, Honglu Zhao, and Xin Nie. Bayesian fully-connected tensor network for hyperspectral-multispectral image fusion. *IEEE Transactions on Image Processing*, 35:5910–5925, 2026. doi: 10.1109/TIP.2026.3696124.

Yanli Shang, Jianjun Liu, Jingyi Zhang, and Zebin Wu. Mft-gan: A multiscale feature-guided transformer network for unsupervised hyperspectral pansharpening. *IEEE Transactions on Geoscience and Remote Sensing*, 62:1–16, 2024. doi: 10.1109/TGRS.2024.3402058.

Yang Song, Jascha Sohl-Dickstein, Diederik P. Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-based generative modeling through stochastic differential equations. *International Conference on Learning Representations*, 2021.

Gemine Vivone, Luciano Alparone, Jocelyn Chanussot, Mauro Dalla Mura, Andrea Garzelli, Giorgio A. Licciardi, Rocco Restaino, and Lucien Wald. A critical comparison among pansharpening algorithms. *IEEE Transactions on Geoscience and Remote Sensing*, 53(5):2565–2586, 2015. doi: 10.1109/TGRS.2014.2361734.

Gemine Vivone, Mauro Dalla Mura, Andrea Garzelli, Rocco Restaino, Giuseppe Scarpa, Magnus O. Ulfarsson, Luciano Alparone, and Jocelyn Chanussot. A new benchmark based on recent advances in multispectral pansharpening: Revisiting pansharpening with classical and emerging pansharpening methods. *IEEE Geoscience and Remote Sensing Magazine*, 9(1):53–81, 2021. doi: 10.1109/MGRS.2020.3019315.

Gemine Vivone, Andrea Garzelli, Yang Xu, Wenzhi Liao, and Jocelyn Chanussot. Panchromatic and hyperspectral image fusion: Outcome of the 2022 whispers hyperspectral pansharpening challenge. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 16:166–179, 2023. doi: 10.1109/JSTARS.2022.3220974.

Lucien Wald, Thierry Ranchin, and Marc Mangolini. Fusion of satellite images of different spatial resolutions: Assessing the quality of resulting images. *Photogrammetric Engineering and Remote Sensing*, 63(6):691–699, 1997.

Xiaozheng Wang, Yong Yang, Shuying Huang, Nayu Liu, and Ziyang Liu. Umnet: Uncertainty-guided memory network for hyperspectral pansharpening. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, pp. 10199–10206, 2026. doi: 10.1609/aaai.v40i12.37988.

Jin-Liang Xiao, Ting-Zhu Huang, Liang-Jian Deng, Guang Lin, Zihan Cao, Chao Li, and Qibin Zhao. Hyperspectral pansharpening via diffusion models with iteratively zero-shot guidance. In *2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 12669–12678, 2025. doi: 10.1109/CVPR52734.2025.01182.

Mai Xu, Yongxuan Dou, Xin Deng, Xin Zou, and Zhenwei Shi. Blur-resistant hyperspectral image super-resolution via dual-degradation fusion model. *IEEE Transactions on Image Processing*, pp. 1–1, 2026. doi: 10.1109/TIP.2026.3714832.

Junfeng Yang, Xueyang Fu, Yuwen Hu, Yue Huang, Xinghao Ding, and John Paisley. Pannet: A deep network architecture for pan-sharpening. In *2017 IEEE International Conference on Computer Vision (ICCV)*, pp. 1753–1761, 2017. doi: 10.1109/ICCV.2017.193.

Pei Yang, Yong Ma, Xiaoguang Mei, Qihai Chen, Minghui Wu, and Jiayi Ma. Deep blind super-resolution for hyperspectral images. *Pattern Recognition*, 157:110916, 2025. doi: 10.1016/j.patcog.2024.110916.

Yong Yang, Xuan Liu, Shuying Huang, Xiaozheng Wang, and Jinlin Lu. 3sp-dunet: An interpretable deep-unfolded network with spatial-spectral sparse priors for hyperspectral pansharpening. *IEEE Transactions on Geoscience and Remote Sensing*, 64:1–16, 2026. doi: 10.1109/TGRS.2026.3674159.

Naoto Yokoya, Takehisa Yairi, and Akira Iwasaki. Coupled nonnegative matrix factorization unmixing for hyperspectral and multispectral data fusion. *IEEE Transactions on Geoscience and Remote Sensing*, 50(2):528–537, 2012. doi: 10.1109/TGRS.2011.2161320.

Qiangqiang Yuan, Yancong Wei, Xiangchao Meng, Huanfeng Shen, and Liangpei Zhang. A multiscale and multidepth convolutional neural network for remote sensing imagery pan-sharpening. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 11(3):978–989, 2018. doi: 10.1109/JSTARS.2018.2794888.

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, 2025. doi: 10.1109/CVPR52734.2025.01664.

Maciej Ziaja, Pawel Kowaleczko, Jakub Nalepa, Daniel Kostrzewa, Daniele Latini, Davide De Santis, Giorgia Salvucci, Ilaria Petracca, Valeria La Pegna, Fabio Del Frate, and Michal Kawulok. Hyperspectral image pansharpening: The prisma case study. In *2023 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, pp. 1633–1636, 2023. doi: 10.1109/IGARSS52108.2023.10282612.

Maciej Ziaja, Pawel Kowaleczko, Tomasz Tarasiewicz, Daniel Kostrzewa, and Michal Kawulok. Hyperspectral pansharpening enhanced with multi-image super-resolution for prisma data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 18:16562–16578, 2025. doi: 10.1109/JSTARS.2025.3581741.

Simone Zini, Mirko Paolo Barbato, Flavio Piccoli, and Paolo Napoletano. Deep learning hyperspectral pansharpening on large-scale prisma dataset. *Remote Sensing*, 16(12):2079, 2024. doi: 10.3390/rs16122079.