Hyperspectral Pansharpening via Diffusion Models with Iteratively Zero-Shot Guidance

Jin-Liang Xiao[1,3], Ting-Zhu Huang[1]*, Liang-Jian Deng[1], Guang Lin[2,3], Zihan Cao[1], Chao Li[3], Qibin Zhao[3]*
[1] University of Electronic Science and Technology of China, Chengdu, China
[2] Tokyo University of Agriculture and Technology, Tokyo, Japan
[3] RIKEN AIP, Tokyo, Japan

Abstract

Hyperspectral pansharpening refers to fusing a panchromatic image (PAN) and a low-resolution hyperspectral image (LR-HSI) to obtain a high-resolution hyperspectral image (HR-HSI). Recently, guiding pre-trained diffusion models (DMs) has demonstrated significant potential in this area, leveraging their powerful representational abilities while avoiding complex training processes. However, these DMs are often trained on RGB images, not well-suited for pansharpening tasks, limited in adapting to the hyperspectral images. In this work, we propose a novel guided diffusion scheme with zero-shot guidance and neural spatial-spectral decomposition (NSSD) to iteratively generate the RGB detail image and map the RGB detail image to target HR-HSI. Specifically, zero-shot guidance employs an auxiliary neural network that trained only with a PAN and LR-HSI to guide pre-trained DMs in generating the RGB detail image, informed by specific prior knowledge. Then, NSSD establishes a spectral mapping from the generated RGB detail image to the final HR-HSI. Extensive experiments are conducted on Pavia, Washington DC, Chikusei, and FR1 datasets to demonstrate that the proposed method significantly enhances the performance of DMs for hyperspectral pansharpening tasks, outperforming existing methods across multiple metrics and achieving improvements in visualization results. The code is available at https://github.com/Jin-liangXiao/DM-zs.

1. Introduction

Hyperspectral images (HSIs), rich in spatial and spectral information, have been extensively used in a variety of applications [22, 46, 48]. In practice, however, the spatial and spectral information in HSIs are often divided across different image types: the panchromatic image (PANs) with high spatial resolution and low-resolution hyperspectral images (LR-HSIs) with abundant spectral information, due to hardware constraints. This separation highlights the need for hyperspectral pansharpening, which aims to fuse these image types to generate high-resolution hyperspectral images (HR-HSIs) [26].

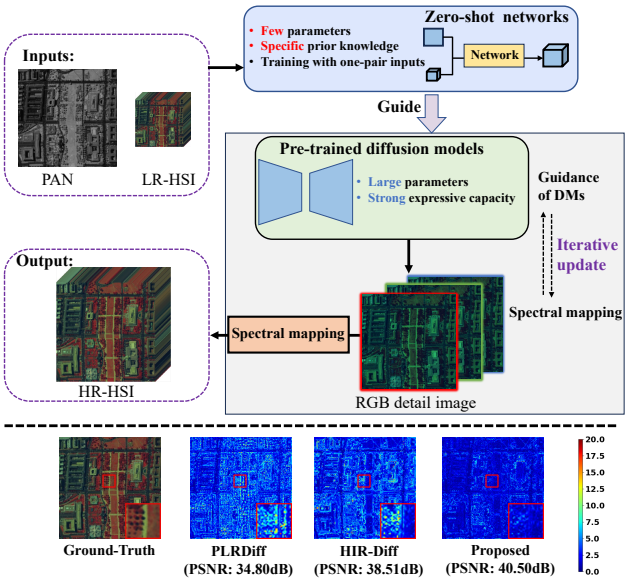

Figure 1. The upper row: illustration of the proposed method that includes iterative update of **zero-shot guidance** for pre-trained diffusion models and **spectral mapping** from RGB detail image to target HR-HSI. The bottom row: error map comparison between state-of-the-art pre-trained diffusion based hyperspectral pansharpening models (i.e., PLRDiff [30], HIR-Diff [29], and the proposed method).

Methods for hyperspectral pansharpening generally fall into two categories: single-image approaches [10, 14, 16, 41, 47] and deep learning (DL) methods [4, 12, 13, 13, 15, 17, 18]. Single-image approaches require only a single image pair instead of an entire dataset for training, which can save significant training costs [36]. While single-image approaches are effective, DL methods can provide stronger

*Corresponding author

representations due to the powerful feature-extraction ability of deep neural networks. However, DL methods come with the trade-off of requiring intensive computational resources for training [7, 18, 37] and consistency in the dataset [40]. Specifically, if there is a distribution shift [8] between the training data and testing data, the performance can degrade significantly, which leads to poor generalization. This issue is particularly serious in hyperspectral pansharpening, where hyperspectral data are both more varying and limited than data in general computer vision tasks.

Foundation models, particularly diffusion models (DMs) [21, 34], have gained wide attention in computer vision for their capability of generating high-quality images. In the context of hyperspectral pansharpening, DMs have shown improved generalization due to flexibility of the guidance techniques. Recent studies [29, 30] have demonstrated the potential of DMs, extending pre-trained DMs to hyperspectral pansharpening through specialized guidance. These methods utilize the degenerate process of pansharpening to guide the DM. Due to the powerful representational capabilities, pre-trained DM as a foundation model can be quickly transferred to hyperspectral pansharpening tasks, which can reduce a significant amount of training costs. However, since existing methods lack the consideration of specific prior knowledge of hyperspectral data, the guidance of DMs still needs improvement. The extension from pre-trained DMs to the hyperspectral pansharpening task is non-trivial: *We need effective guidance to adapt the pre-trained DMs originally designed with only RGB channels, to manage the complex spectral fidelity while preserving high-resolution spatial details*. This adaptation becomes even more challenging when we only have a single (PAN, LR-HSI) image pair. These challenges naturally give rise to two important questions: **1)** How can we shift the data distribution of pre-trained DMs to HSIs for hyperspectral pansharpening with limited data through effective guidance? **2)** How to accurately transform the RGB information of the generated images to HR-HSIs?

To address the aforementioned challenges, we propose a novel guided diffusion scheme with zero-shot guidance and neural spatial-spectral decomposition (NSSD) to iteratively generate an RGB spatial detail image and map the detail image to the target HR-HSI. As shown in the green box of Figure 1, we utilize the pre-trained DMs as the backbone, with all their parameters being frozen. Then, we introduce the specific prior knowledge of testing data by a trainable zero-shot network to guide the DMs to generate the detail image enriched with expected features, as shown in the blue box. Unlike conventional deep neural networks that require extensive training data and complex training processes, the proposed zero-shot network is a lightweight model with few parameters, can be trained naively on a single testing image pair, i.e., a LR-HSI and PAN. This network encodes specific prior knowledge into pre-trained DMs and guides the DMs to generate detail image more effectively (see Figure 3). Additionally, motivated by the tensor decomposition in single-image approaches [44], we recognize that spatial-spectral correlation is critical for the quality of final data and introduce a neural spatial-spectral decomposition (NSSD) to capture the spatial-spectral correlation and map from the RGB detail image to the final HR-HSI. The NSSD and the guidance of pre-trained DMs are updated iteratively, as displayed in the gray box of Figure 1, making NSSD adaptive to the results of guided DMs. Compared with existing pre-trained DMs methods, the proposed method achieves the state-of-the-art result (refer to the bottom row of Figure 1).

The main contributions are summarized as follows:

- We propose a zero-shot network to efficiently guide the pre-trained DM to generate a detail image enriched with specific prior knowledge.
- We introduce a neural spatial-spectral decomposition (NSSD) to iteratively construct the target HR-HSI from the generated detail image.
- Experiments on the Pavia, Washington DC, Chikusei, and FR1 datasets demonstrate that the proposed method achieves the state-of-the-art performance.

2. Preliminaries & Related Works

In this section, we provide definitions, notations, degenerate process of pansharpening, a brief introduction to score-based DMs with guidance, and related works.

2.1. Notations

Throughout the paper, we denote tensors $\mathcal{X} \in \mathbb{R}^{H\times W\times S}$, $\mathcal{Y} \in \mathbb{R}^{h\times w\times S}$, $\mathbf{P} \in \mathbb{R}^{H\times W}$ as HR-HSI, LR-HSI, PAN, respectively. For tensor $\mathcal{X}$, $\mathbf{X}_{(3)} \in \mathbb{R}^{S\times HW}$ denotes its mode-3 unfolding [24]. Additionally, $\mathcal{X} \times_3 \mathbf{R}$ indicates the mode-3 multiplication between tensor $\mathcal{X}$ and matrix $\mathbf{R}$. $\mathbf{x} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ means the variable $\mathbf{x}$ satisfies Gaussian distribution with the mean of $\mathbf{0}$ and variance of matrix $\mathbf{I}$.

2.2. Degenerate Process of Pansharpening

For the hyperspectral pansharpening task, inputs PAN and LR-HSI are mainly modeled as HR-HSI after the spectral and spatial degradation, respectively [45]. The degenerate process of the hyperspectral pansharpening problem can be mathematically represented as follows:

$$\begin{aligned} \mathcal{Y} &= \tau(\mathcal{X}), \\ \mathbf{P} &= \mathcal{X} \times_3 \mathbf{R}, \end{aligned} \tag{1}$$

where τ indicates blurring and down-sampling operations, and $\mathbf{R} \in \mathbb{R}^{1\times S}$ denotes the spectral response function [43]. The goal of the task is to obtain the HR-HSI $\mathcal{X}$ from the LR-HSI $\mathcal{Y}$ and PAN $\mathbf{P}$.

2.3. Score-based Diffusion Models with Guidance

Diffusion models (DMs) [21, 33] have proven to be a potent class of generative models capable of generating high-quality images. Among them, score-based DMs are a powerful family that models the forward process by stochastic differential equations (SDEs) and the reverse process by corresponding reverse-time SDEs [34]. Given a input $\mathbf{x}$, we have

$$\mathrm{d}\mathbf{x} = f(\mathbf{x}, t)\mathrm{d}t + g(t)\mathrm{d}\mathbf{w}, \tag{2}$$

$$\mathrm{d}\mathbf{x} = [f(\mathbf{x}, t) - g^2(t)\nabla_{\mathbf{x}} \log p_t(\mathbf{x})] + g(t)\mathrm{d}\bar{\mathbf{w}}, \tag{3}$$

where t represents timestep, $f(\cdot, t)$ denotes a vector valued function called drift coefficient and $g(t)$ is a scalar valued function called diffusion term. $\mathbf{w}$ and $\bar{\mathbf{w}}$ represent a Wiener process and a reverse-time Wiener process, respectively. Eq. (2), corresponding to the forward SDE, indicates adding noise from $t = 0$ to T, which diffuses the image to pure Gaussian noise. In the reverse process, the corresponding reverse-time SDEs, i.e., Eq. (3), we utilize a learned score network $s_\theta(\mathbf{x}_t, t) := \nabla_{\mathbf{x}} \log p_t(\mathbf{x})$ to generate denoised sample $\mathbf{x}_0$ from a simple Gaussian noise $\mathbf{x}_T \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$. To more effectively enable DMs to generate the desired images, Dhariwal *et al.* and Pang *et al.* [9, 29] proposed the guided DM, which is a valid path for general specific tasks, leveraging the powerful generative capabilities of DMs without the need for complex retraining. The guidance of pre-trained DMs works by introducing gradient in reverse process, which can be represented as

$$\hat{s}_\theta(\mathbf{x}_t, t) = s_\theta(\mathbf{x}_t, t) + s(t) \cdot \nabla_{\mathbf{x}_t} \log p(y|\mathbf{x}_t), \tag{4}$$

where $s(t)$ is the guidance strength in t step, and $p(y|\mathbf{x}_t)$ is the relationship between estimated $\mathbf{x}_t$ and condition y, which is crucial for the guidance of DMs on specific tasks.

2.4. Related Works

Leveraging the powerful generative capabilities of DMs, numerous studies have been conducted on hyperspectral pansharpening [5, 25, 39, 42, 50]. However, the performance of these models often depends on complex and time-consuming training processes and substantial training data. Recently, efforts have also been made to bridge the connection between inverse problems and pre-trained DMs. Kawar *et al.* [23] employed a pre-trained DM to solve linear inverse problems for RGB images. Since pre-trained DMs are primarily trained on RGB data, there remains a gap between the generated outputs and the target hyperspectral images. Miao *et al.* [27] further extended this approach to hyperspectral image restoration.

For the hyperspectral pansharpening task, a nonlinear inverse problem, different loss functions are proposed to guide DMs. Rui *et al.* [30] proposed a pre-trained diffusion framework that guides the diffusion process to generate an RGB detail image $\mathcal{A}$. This image is then transformed into the final pansharpened hyperspectral image by applying a fixed transformation matrix along the third dimension, as $\mathcal{X} = \mathcal{A} \times_3 \mathbf{E}$, where $\mathcal{X} \in \mathbb{R}^{H\times W\times S}$ is the target HR-HSI image, $\mathcal{A} \in \mathbb{R}^{H\times W\times b}$ is the image generated by a pre-trained DM, and $\mathbf{E} \in \mathbb{R}^{S\times b}$ is the fixed transform matrix. Pang *et al.* [29] explore a new choice for estimating the transform matrix $\mathbf{E}$ and apply total variation (TV) prior to diffusion guided function. However, this issue remains only partially resolved, as their fixed transformation matrix $\mathbf{E}$ in the overall algorithm causes the final result to depend heavily on the initial estimate of $\mathbf{E}$. In contrast, our method introduces an auxiliary neural network to provide the specialized guidance for generating an RGB detail image and iteratively employs a neural spatial-spectral decomposition to adaptively construct the mapping from the RGB detail image to HR-HSI.

3. Methodology

In this section, we elaborate on the proposed guided diffusion shcheme. As illustrated in Figure 2, the scheme primarily consists of two components: diffusion with zero-shot guidance and neural spatial-spectral decomposition (NSSD). The zero-shot guidance operates on the pre-trained DM to generate an RGB detail image, enabling better preservation of spatial details and adaptation to specific prior knowledge. NSSD establishes the mapping from the detail image to the HR-HSI.

3.1. Diffusion with Zero-Shot Guidance

Unlike the existing DM-based hyperspectral pansharpening methods [29, 30], we aim to guide the reverse process of DMs for generating the RGB detail image $\mathcal{A}$ by introducing the specific prior knowledge of testing data. To this end, we design a novel diffusion scheme with zero-shot guidance, where the reverse process of $\mathcal{A}$ satisfies SDE as follows:

$$\begin{aligned}\mathrm{d}\mathcal{A}_t = {} & g(t)\mathrm{d}\overline{w} + f(\mathcal{A}, t)\mathrm{d}t \\ & - g^2(t)[s_\theta(x_t, t) + \nabla_{\mathcal{A}_t} \log\ p_t(\tilde{\mathcal{Y}}|\mathcal{A}_t)]\mathrm{d}t,\end{aligned} \tag{5}$$

where $p_t(\tilde{\mathcal{Y}}|\mathcal{A}_t)$ is the pansharpening guidance estimated from the condition, defined as

$$\tilde{\mathcal{Y}} = (\mathcal{Y}, \mathbf{P}, \mathcal{M}, f_\psi(\mathcal{Y}, \mathbf{P})), \tag{6}$$

where $\mathcal{Y}$ and $\mathbf{P}$ are input image pair LR-HSI and PAN, $\mathcal{M}$ is a mapping from RGB detail image to HR-HSI (deferred in next subsection), and $f_\psi(\mathcal{Y}, \mathbf{P})$ is an auxiliary neural network that is trained with $\mathcal{Y}$ and $\mathbf{P}$. Following [9], we have

$$p_{\theta,\psi}(\mathcal{A}_{t-1}|\mathcal{A}_t, \tilde{\mathcal{Y}}) \propto p_\theta(\mathcal{A}_{t-1}|\mathcal{A}_t)\ p_\psi(\tilde{\mathcal{Y}}|\mathcal{A}_t). \tag{7}$$

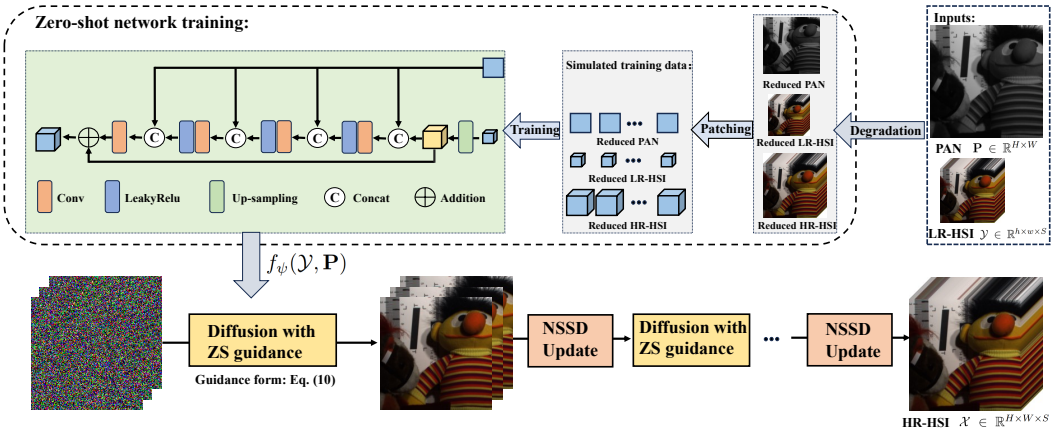

Figure 2. Illustration of the whole process of the proposed hyperspectral pansharpening method that includes diffusion with zero-shot (ZS) guidance and neural spatial-spectral decomposition (NSSD). The part in the dashed box is the proposed zero-shot guided function that is trained with a single testing image pair. NSSD builds the spectral mapping from the detail image generated by DM to HR-HSI. The guidance of the DM and NSSD are iteratively updated.

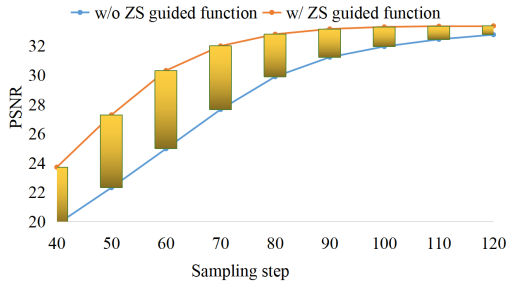

Figure 3. The analysis of the proposed zero-shot guidance. The DMs with the proposed guidance can achieve better performance with fewer sampling steps. (The length of the yellow box means the distance of PSNR values between the DM with and without ZS guidance)

Based on [6], we define the variable $\hat{\mathcal{A}}_0 := \mathbb{E}[\mathcal{A}_0|\mathcal{A}_t] = \mathbb{E}_{\mathcal{A}_0 \sim p(\mathcal{A}_0|\mathcal{A}_t)}[\mathcal{A}_0]$, which means

$$\hat{\mathcal{A}}_0 \simeq \frac{1}{\sqrt{\bar{\alpha}_t}}(\mathcal{A}_t + (1-\bar{\alpha}_t)s_\theta(\mathcal{A}_t, t)), \tag{8}$$

$$p_\psi(\tilde{\mathcal{Y}}|\mathcal{A}_t) \simeq p_\psi(\tilde{\mathcal{Y}}|\hat{\mathcal{A}}_0), \tag{9}$$

where $\bar{\alpha}_t := \prod_{s=1}^{t} \alpha_s$ is the noise scale of DMs. We propose the zero-shot guidance as follows,

$$\begin{aligned}\log\ p_\psi(\tilde{\mathcal{Y}}|\hat{\mathcal{A}}_0) = &\left\|\tau(\mathcal{M}(\hat{\mathcal{A}}_0)) - \mathcal{Y}\right\|_F^2 \\ &+ \left\|\mathcal{M}(\hat{\mathcal{A}}_0) \times_3 \mathbf{R} - \mathbf{P}\right\|_F^2 + \lambda_1 \left\|\mathcal{M}(\hat{\mathcal{A}}_0) - f_\psi(\mathcal{Y}, \mathbf{P})\right\|_F^2,\end{aligned} \tag{10}$$

where λ_1 is a weighting scale. Finally, the reverse process is illustrated as:

$$\begin{aligned}\mathcal{A}_{t-1} = &\frac{1}{\sqrt{\alpha_t}}(\mathcal{A}_t - \frac{\beta_t}{\sqrt{1-\bar{\alpha}_t}} s_\theta(\mathcal{A}_t, t)) \\ &- \lambda \cdot \frac{\beta_t}{\sqrt{1-\bar{\alpha}_t}\sqrt{\alpha_t}} \nabla_{\mathcal{A}_t} log\ p_\psi(\tilde{\mathcal{Y}}|\hat{\mathcal{A}}_0),\end{aligned} \tag{11}$$

where β_t, as a hyperparameter, is generally set to linearly interpolate from 0.0001 to 0.02 [28] and $\alpha_t := 1 - \beta_t$. The weighting scale λ can balance the effect of guidance.

The rest of this part is how to train the zero-shot network $f_\psi(\mathcal{Y}, \mathbf{P})$. The network structure and training process are shown within the dotted box in Figure 2, which has been verified its effectiveness in [11]. Based on the "invariance among scales" assumption, we train the network in the reduced-resolution scale. In specific, we use the input LR-HSI as the training ground-truth. Training PAN and training LR-HSI are simulated by PAN and LR-HSI after spatial and spectral degenerate processes, respectively. This approach allows for efficient training of the zero-shot network using only a single testing input pair. Figure 3 shows the change curve of PSNR values with the increase of iteration number. One can observe that the DM guided by zero-shot guidance can achieve better performance with fewer sampling steps. Besides, this single-image training property provides specific prior knowledge to the pre-trained DM.

3.2. Neural Spatial-Spectral Decomposition

This subsection describes neural spatial-spectral decomposition (NSSD), i.e., $\mathcal{M}$ in Eq. (10). In hyperspectral pansharpening, beyond the fidelity term in (1), the spatial-spectral correlation exhibits the low-rank properties characteristic of hyperspectral data [44]. The mode-3 decom-

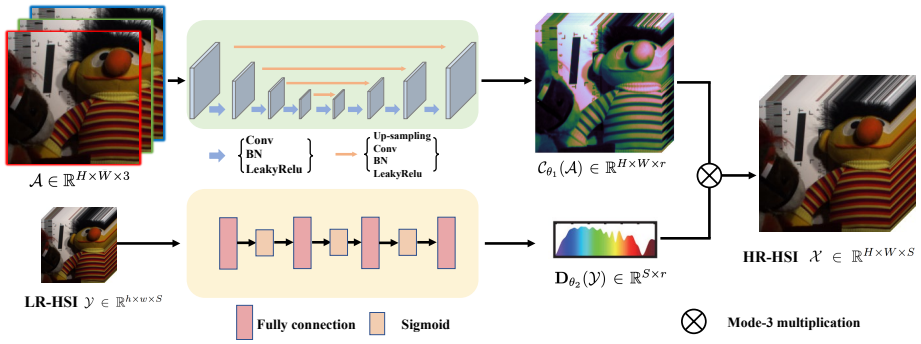

Figure 4. The structure of neural spatial-spectral decomposition (NSSD) consists of a U-net structure with skip connection and the fully connection structure to extract spatial and spectral information, respectively.

position format, i.e., $\mathcal{X} = \mathcal{A} \times_3 \mathbf{E}$, has shown to be effective [19, 20, 31], which satisfies the following inequation:

$$\text{rank}(\mathbf{X}_{(3)}) \leq \min\{\text{rank}(\mathbf{A}_{(3)}), \text{rank}(\mathbf{E})\} \leq b, \tag{12}$$

where $\mathcal{A} \in \mathbb{R}^{H\times W\times b}$, and $\mathbf{E} \in \mathbb{R}^{S\times b}$. The low value of the size of $\mathbf{E}$ limits the expressive ability. Since pre-trained DMs are primarily trained on RGB images, the size of $\mathcal{A}$ is typically set to $H \times W \times 3$. Additionally, the larger value of r in (12) inevitably increases the time required for the reverse process in the DM.

To mitigate the drawbacks, we propose a NSSD to capture the spatial-spectral correlation, as follows:

$$\mathcal{X} = \mathcal{M}(\mathcal{A}) := \mathcal{C}_{\theta_1}(\mathcal{A}) \times_3 \mathbf{D}_{\theta_2}(\mathcal{Y}), \tag{13}$$

where $\mathcal{C}_{\theta_1}(\mathcal{A}) \in \mathbb{R}^{H\times W\times r}$ and $\mathbf{D}_{\theta_2}(\mathcal{Y}) \in \mathbb{R}^{S\times r}$ with hyperparameter r are produced by two neural network with the input $\mathcal{A} \in \mathbb{R}^{H\times W\times 3}$ and $\mathcal{Y} \in \mathbb{R}^{h\times w\times S}$, respectively, as illustrated in Figure 4. Specifically, $\mathcal{C}_{\theta_1}(\cdot)$ is the spatial-factor network that extracts the spatial information from the RGB detail image. The structure of $\mathcal{C}_{\theta_1}(\cdot)$ is the U-net type with skip connection, whose detail-extraction capability has been verified [35]. The spectral-factor network, $\mathbf{D}_{\theta_2}(\cdot)$, consists of the fully connection structure and sigmoid to efficiently absorb spectral information from LR-HSI. NSSD does not require pre-training. It produces the final result through iterative steps using the following loss:

$$\begin{aligned} Loss = &\|\tau(\mathcal{C}_{\theta_1} \times_3 \mathbf{D}_{\theta_2}) - \mathcal{Y}\|_F^2 \\ &+\lambda_2 \|(\mathcal{C}_{\theta_1} \times_3 \mathbf{D}_{\theta_2}) \times_3 \mathbf{R} - \mathbf{P}\|_F^2, \end{aligned} \tag{14}$$

where λ_2 is a balanced parameter. NSSD models the relationship between the generated detail image and the target HR-HSI. Compared with the classic format $\mathcal{X} = \mathcal{A} \times_3 \mathbf{E}$, the proposed form $\mathcal{C}_{\theta_1}(\mathcal{A}) \times_3 \mathbf{D}_{\theta_2}(\mathcal{Y})$ offers a more flexible representation for HR-HSI, particularly for data with higher rank. Furthermore, the NSSD evolves in the guidance of pre-trained DMs, as shown in Eq. (10). The guidance of the DMs and the update of NSSD are mutually adaptive. Thus, NSSD provides a more flexible and DM-compatible representation of the spatial-spectral correlation in HR-HSI.

Algorithm 1: The Whole Solving Process

Input: $\mathcal{A}_T$ sampled from $\mathcal{N}(0, I)$, DM s_θ, noise scales $\{\bar{\alpha}_t\}_{t=1}^T$, f_ψ, LR-HSI $\mathcal{Y}$, PAN $\mathbf{P}$
Output: HR-HSI $\mathcal{X}$

while *not converged* **do**
  for $t = T, T-1, \cdots 1$ **do**
    step 1: estimate $\hat{\mathcal{A}}_0$ by Eq. (8)
    step 2: compute $\log\ p_\psi(\tilde{\mathcal{Y}}|\mathcal{A}_t)$ by Eq. (10)
    step 3: sample $\mathcal{A}_{t-1}$ by Eq. (11)
  end
  NSSD Update: update $\mathcal{M}$ by minimizing (14)
end
$\mathcal{X} = \mathcal{C}_{\theta_1}(\mathcal{A}) \times_3 \mathbf{D}_{\theta_2}(\mathcal{Y})$
return $\mathcal{X}$

Whole Solving Algorithm: Since the guidance of pre-trained DM is related to the mapping $\mathcal{M}$, the generated detail image is adaptive to the mapping $\mathcal{M}$. To enhance the flexibility of mapping and the suitability of diffusion guidance, we alternatively conduct guidance of pre-trained DM and NSSD update as shown in Figure 2. During each guidance step, the parameters of the NSSD are held fixed. The whole solving algorithm is summarized in Algorithm 1.

4. Experiments

To demonstrate the effectiveness of the proposed method, we compare it with corresponding state-of-the-art methods, including GSA [1], CNMF [47], HySure [32], ZSL [11], PLRDiff [30], and HIR-Diff [29]. It is worth to be remarked that HIR-Diff [29] is a general inverse problem framework. We rewrite the fidelity for the hyperspectral pansharpening task. The parameters of all methods are adjusted according to the authors' recommendation. The metrics peak

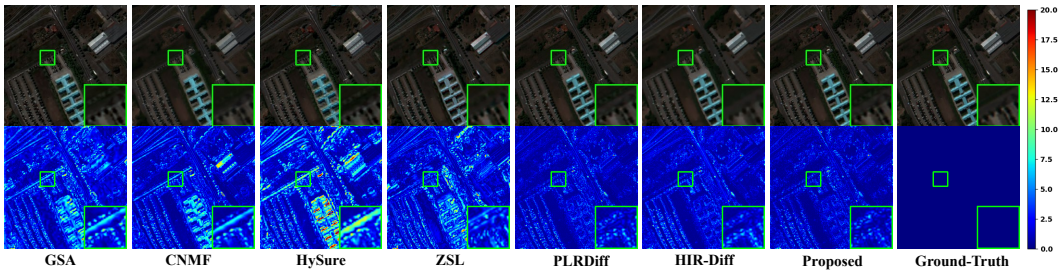

Figure 5. Visual comparisons on Pavia dataset. The first row shows the predicted HR-HSIs of different methods, and the second row displays the error maps between predicted HR-HSIs and ground-truth.

signal-to-noise ratio (PSNR), the structural similarity index (SSIM) [38], the relative dimensionless global error in synthesis (ERGAS) [2], the spectral angle mapper (SAM) [49], and root mean square error (RMSE) are utilized for assessment. QNR [3] is employed for evaluating real-world data. We assess the results produced by these methods on Pavia[1], Washington DC (WDC)[2], Chikusei[3], and FR1 datasets[4]. The simulated LR-HSI is gained by applying a 5×5 Gaussian blurring kernel and the downsampling factor 4. The simulated PAN image is obtained from ground-truth by multiplying a spectral mean response vector. The downsampling factor of FR1 dataset is 6. The detailed information about these datasets is listed as follows,

- **Pavia dataset:** Pavia dataset are originally obtained with the size of $610 \times 340 \times 115$. We choose 93 bands by eliminating the bands associated with water vapor absorption and crop the up-left cube with size $256 \times 256 \times 93$.
- **WDC dataset:** WDC dataset is obtained from the Hyperspectral Digital Imagery Collection Experiment (HYDICE) sensor. We eliminate some unusable bands and choose the area with size $256 \times 256 \times 191$.
- **Chikusei dataset:** The Headwall Hyperspec-VNIR-C imaging sensor captures the Chikusei dataset with 128 bands and the spectral range is from 363 to 1018 nm. We crop a subimage with size $256 \times 256 \times 128$.
- **FR1 real dataset:** The FR1 dataset is a real-world data distributed for the PRISMA contest, for pansharpening at the full spatial resolution. We crop a subimage with PAN size 240×240.

4.1. Implementation Details

All experiments are conducted on the same PC with Pytorch 2.0 and MATLAB (R2022a) on a computer of 64Gb RAM, Intel(R) Core(TM) i9-10900KF CPU @3.70GHz, and NVIDIA GeForce RTX 4070. The update of NSSD is optimized by Adam optimizer with 0.001 learning rate (LR). The balanced parameter of guidance λ_1 is 0.1, and $\lambda_2 = 1$. The diffusion reverse process step number is set to 100. The number of the whole iterations is 20000. The training of the zero-shot neural network is with Adam optimizer with 2000 iterations, and the LR is 0.001 decayed by $gamma = 0.8$ each 1000 iterations. The rank of NSSD for Pavia, WDC, and Chikusei datasets are set to 20, 40, and 40, respectively. For FR1 real dataset, $\lambda_1 = \lambda_2 = 10$, and the number of the whole iterations is 10000.

[1] https://rslab.ut.ac.ir/data
[2] https://engineering.purdue.edu/biehl/MultiSpec/hyperspectral.html
[3] https://naotoyokoya.com/Download.html
[4] https://github.com/liangjiandeng/HyperPanCollection

Table 1. Quantitative results for Pavia dataset. (Bold: best; Underline: second best)

Method	PSNR↑	SSIM↑	ERGAS↓	SAM↓	RMSE↓
GSA [1]07'TGRS	28.304	0.812	5.819	7.753	10.368
CNMF [47] 11'TGRS	28.896	0.859	5.495	5.949	9.558
HySure [32] 14'TGRS	25.847	0.758	7.861	10.059	14.065
ZSL [11]23'TPAMI	32.060	0.905	3.549	5.318	6.819
PLRDiff [30] 24'IF	<u>33.060</u>	<u>0.905</u>	<u>3.291</u>	5.334	<u>6.001</u>
HIR-Diff [29]24'CVPR	31.974	0.875	3.618	<u>4.705</u>	6.742
Proposed	**34.227**	**0.926**	**2.863**	**4.463**	**5.284**
Ideal value	**+∞**	**1**	**0**	**0**	**0**

4.2. Results on Pavia Dataset

In this part, we evaluate the effectiveness of the proposed method on the Pavia dataset. Figure 5 shows the visual comparison of these methods in the benchmark. Diffusion-based methods, i.e., PLRDiff [30], HIR-Diff [29], and the proposed method, achieve excellent performance for detail extraction. The residual images, i.e., the second row of Figure 5 obviously verify that our method can preserve more details compared to others. The result produced by HySure [32] leads to spatial distortions. Table 1 provides the quantitative results for compared approaches. For all metrics, the proposed method achieves the best performance.

4.3. Results on WDC Dataset

For the WDC dataset, there exist more challenges to preserve the spectral information due to its more spectral

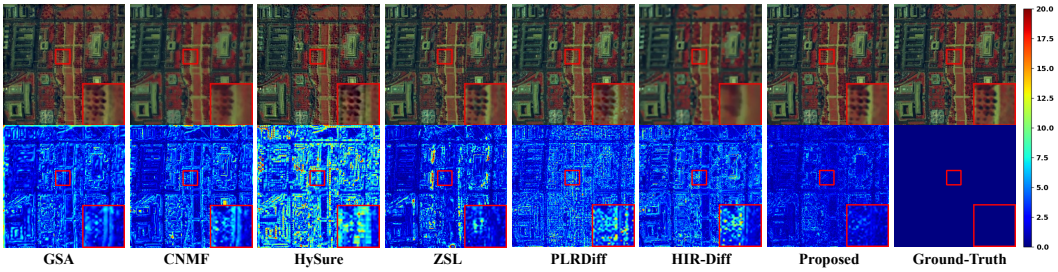

Figure 6. Visual comparisons on WDC dataset. The first row shows the predicted HR-HSIs of different methods, and the second row displays the error maps between predicted HR-HSIs and ground-truth.

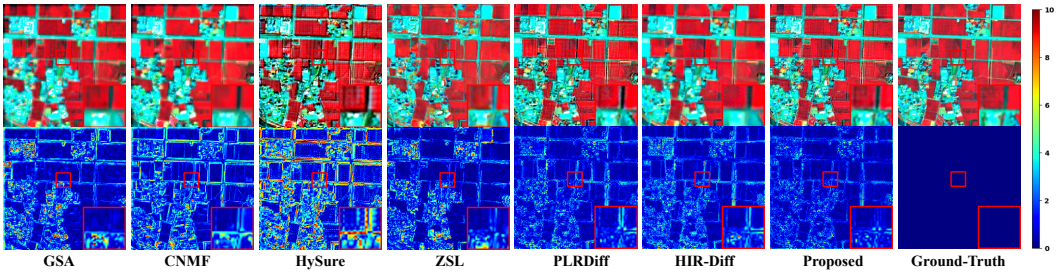

Figure 7. Visual comparisons on Chikusei dataset. The first row shows the predicted HR-HSIs of different methods, and the second row displays the error maps between predicted HR-HSIs and ground-truth.

Table 2. Quantitative results for WDC dataset. (Bold: best; Underline: second best)

Method	PSNR↑	SSIM↑	ERGAS↓	SAM↓	RMSE↓
GSA [1]07'TGRS	39.629	0.910	5.924	7.079	7.665
CNMF [47] 11'TGRS	39.703	0.897	**5.864**	6.448	7.665
HySure [32] 14'TGRS	35.071	0.798	9.926	11.898	12.928
ZSL [11]23'TPAMI	39.685	0.916	6.335	7.825	7.301
PLRDiff [30] 24'IF	34.799	0.833	18.202	9.544	8.667
HIR-Diff [29]24'CVPR	38.506	0.844	7.102	6.686	8.079
Proposed	**40.503**	**0.943**	12.254	**5.047**	**4.631**
Ideal value	**+∞**	**1**	**0**	**0**	**0**

Table 3. Quantitative results for Chikusei dataset. (Bold: best; Underline: second best)

Method	PSNR↑	SSIM↑	ERGAS↓	SAM↓	RMSE↓
GSA [1]07'TGRS	32.459	0.887	10.888	6.762	6.662
CNMF [47] 11'TGRS	33.578	0.878	8.592	4.893	5.905
HySure [32] 14'TGRS	30.270	0.752	14.159	9.883	8.417
ZSL [11]23'TPAMI	35.258	0.917	8.367	5.685	4.632
PLRDiff [30] 24'IF	37.176	0.917	6.994	4.838	3.651
HIR-Diff [29]24'CVPR	36.389	0.905	6.333	**3.894**	4.247
Proposed	**38.450**	**0.939**	**5.945**	4.197	**3.192**
Ideal value	**+∞**	**1**	**0**	**0**	**0**

bands. As shown in Table 2, we provide the visual results of different methods. Our approach also performs best on this dataset. Compared with traditional method CNMF [47], diffusion-based methods PLRDiff [30] and HIR-Diff [29] lead to spectral distortions. As demonstrated in Figure 6, PLRDiff [30] also preserves excellent spatial detail. The proposed method has the advantages of both spatial detail preservation and spectral information extraction.

4.4. Results on Chikusei Dataset

Figure 7 evaluates the effectiveness of the proposed method on the Chikusei dataset. Our method achieves the best result due to the spatial detail preservation. For the detail edge of the image, the proposed method can preserve the spatial detail well. As shown in Table 3, the proposed method demonstrates improvements in PSNR by 1.274dB/2.061dB, respectively, compared with diffusion-based methods PLRDiff [30] and HIR-Diff [29].

4.5. Ablation Study

To demonstrate the effectiveness of the proposed zero-shot guidance and NSSD, we conduct the ablation study on the Pavia dataset. Table 4 shows the results produced by Four models. For the model I, we use the fixed transformation in [30] and regular fidelity to substitute the proposed zero-shot guidance and NSSD, respectively. As shown in Table 4, the results produced by the model with both zero-shot guidance and NSSD are best. Besides, we design more detailed experiments to explore more properties of the two parts.

Zero-shot Guidance: The zero-shot guidance achieves

Table 4. Ablation study about the proposed zero-shot (ZS) guidance and neural spatial-spectral decoder (NSSD) on Pavia dataset.

Ablation	ZS	NSSD	PSNR	SSIM	ERGAS	SAM	RMSE
Model I	✗	✗	33.060	0.905	3.291	5.334	6.001
Model II	✓	✗	33.267	0.912	3.219	5.285	5.877
Model III	✗	✓	33.972	0.921	2.946	4.605	5.460
Proposed	✓	✓	**34.179**	**0.925**	**2.877**	**4.533**	**5.332**

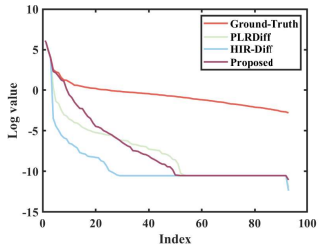

Figure 8. Singular value distributions of results produced by PLRDiff [30], HIR-Diff [29], and the proposed method.

Table 5. Performance of proposed zero-shot guidance compared with conventional regularization term.

Guidance	PSNR	SSIM	ERGAS	SAM	RMSE
Fidelity	32.470	0.896	3.455	**5.114**	6.395
TV	32.770	0.902	3.356	5.156	6.204
Proposed	**33.374**	**0.914**	**3.174**	5.150	**5.793**

excellent guided performance. To demonstrate the effectiveness, we compare it with the conventional total variation regularized term, which is widely used and effective for image processing tasks. As shown in Table 5, the regularized term is helpful for the guidance of DMs. The proposed zero-shot guidance offers more specific prior knowledge for the pre-trained DM to achieve the best performance.

NSSD: NSSD describes the spatial-spectral correlation with the low-rank property. Figure 8 shows the singular value distribution of target HR-HSI produced by different methods. We compare the proposed method with NSSD with two fixed Diffusion transformation estimation methods in PLRDiff [30] and HIR-Diff [29]. One can observe that the proposed NSSD provides a flexible transformation and performs the best approximation with the pre-trained DM guidance. Our method can achieve the closest distance between the singular values of ground-truth.

4.6. Guided Network

In this part, we analyze the influence of zero-shot network structure for guided performance. Table 6 shows the final results with U-Net with skip-connections [35], DSNet [51], and CONet [11]. U-Net contains multi-scale information, DSNet is a structure used in the supervised method, and CONet is a light convolution-based cascade structure.

Table 6. Performance of zero-shot guided network with different network structures.

Guided Structure	PSNR	SSIM	ERGAS	SAM	RMSE
U-Net	34.015	0.921	2.953	4.511	5.426
DSNet	33.775	0.917	3.033	4.677	5.609
CONet	**34.227**	**0.926**	**2.863**	**4.463**	**5.284**

The proposed method with the above network structures can achieve excellent results. The convolution-based cascade structure achieves the best performance.

4.7. More Discussions

Real-data Experiment: Table 7 shows the evaluation on the FR1 real-world dataset, which demonstrates the superior performance of the proposed method on real data.

Table 7. Real data experiment on FR1 dataset.

Method	GSA	CNMF	HySure	ZSL	PLRDiff	HIR-Diff	Proposed
QNR↑	0.590	0.679	0.668	0.802	0.630	0.625	**0.832**

More Experimental Comparisons: We provide further discussion on running time, downsampling factor, and kernel size, which are shown in the supplementary material.

Limitations: While the computational cost of DMs with the proposed zero-shot guidance is comparable to that of previous DM-based methods, such as PLRDiff [30], NSSD requires more time to iteratively construct the spectral mapping from the RGB detail image to the HR-HSI, as detailed in the supplementary material. We will explore the use of pre-trained networks to address this limitation in the future.

5. Conclusion

In this paper, we introduced a novel guided diffusion scheme for hyperspectral pansharpening, combining the iterative update of zero-shot guidance and NSSD. The zero-shot guidance is conducted with an auxiliary convolutional network that can be trained just on one-pair inputs, while NSSD acts as a flexible transformation mapping the generated images by pre-trained DMs to the hyperspectral images. Our method demonstrated significant effectiveness on benchmark datasets, including Pavia, WDC, and Chikusei. Through extensive ablation studies, we analyzed the contributions of the guidance and NSSD and evaluated the impact by different zero-shot network architectures.

Acknowledgement This work was supported in part by the NSFC (12171072, 12271083, 62273077), Natural Science Foundation of Sichuan Province (2024NSFSC0038), and National Key Research and Development Program of China (2020YFA0714001).

References

[1] B. Aiazzi, S. Baronti, and M. Selva. Improving component substitution pansharpening through multivariate regression of MS + Pan data. *IEEE Trans. Geosci. Remote Sens.*, 45 (10):3230–3239, 2007. 5, 6, 7

[2] L. Alparone, L. Wald, J. Chanussot, C. Thomas, P. Gamba, and L. M. Bruce. Comparison of pansharpening algorithms: Outcome of the 2006 GRS-S data-fusion contest. *IEEE Trans. Geosci. Remote Sens.*, 45(10):3012–3021, 2007. 6

[3] L. Alparone, B. Aiazzi, S. Baronti, A. Garzelli, F. Nencini, and M. Selva. Multispectral and panchromatic data fusion assessment without reference. *Photogramm. Eng. Remote Sens.*, 74(2):193–200, 2008. 6

[4] W. G. C. Bandara, J. M. J. Valanarasu, and V. M. Patel. Hyperspectral pansharpening based on improved deep image prior and residual reconstruction. *EEE Trans. Geosci. Remote Sens.*, 60:1–16, 2021. 1

[5] Z. Cao, S. Cao, L.-J. Deng, X. Wu, J. Hou, and G. Vivone. Diffusion model with disentangled modulations for sharpening multispectral and hyperspectral images. *Inf. Fusion*, 104: 102158, 2024. 3

[6] H. Chung, J. Kim, M. T. Mccann, M. L. Klasky, and J. C. Ye. Diffusion posterior sampling for general noisy inverse problems. *arXiv preprint arXiv:2209.14687*, 2022. 4

[7] M. Ciotola, G. Guarino, G. Vivone, G. Poggi, J. Chanussot, A. Plaza, and G. Scarpa. Hyperspectral pansharpening: Critical review, tools and future perspectives. *arXiv preprint arXiv:2407.01355*, 2024. 2

[8] L.-J. Deng, G. Vivone, M. E. Paoletti, G. Scarpa, J. He, Y. Zhang, J. Chanussot, and A. Plaza. Machine learning in pansharpening: A benchmark, from shallow to deep networks. *IEEE Geosci. Remote Sens. Mag.*, 10(3):279–315, 2022. 2

[9] P. Dhariwal and A. Nichol. Diffusion models beat gans on image synthesis. *Adv. Neural Inf. Process. Syst. (NeurIPS)*, 34:8780–8794, 2021. 3

[10] R. Dian, L. Fang, and S. Li. Hyperspectral image super-resolution via non-local sparse tensor factorization. In *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, pages 5344–5353, 2017. 1

[11] R. Dian, A. Guo, and S. Li. Zero-shot hyperspectral sharpening. *IEEE Trans. Pattern Anal. Mach. Intell.*, 45(10):12650–12666, 2023. 4, 5, 6, 7, 8

[12] W. Dong, T. Zhang, J. Qu, S. Xiao, J. Liang, and Y. Li. Laplacian pyramid dense network for hyperspectral pansharpening. *IEEE Trans. Geosci. Remote Sens.*, 60:1–13, 2021. 1

[13] Y. Duan, X. Wu, H. Deng, and L.-J. Deng. Content-adaptive non-local convolution for remote sensing pansharpening. In *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recog.*, pages 27738–27747, 2024. 1

[14] X. Fu, Z. Lin, Y. Huang, and X. Ding. A variational pan-sharpening with local gradient constraints. In *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recog.*, pages 10265–10274, 2019. 1

[15] J. Gao, J. Li, X. Su, M. Jiang, and Q. Yuan. Deep image interpolation: A unified unsupervised framework for pansharpening. In *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recog.*, pages 609–618, 2022. 1

[16] M. Gong, H. Zhang, H. Wang, J. Chen, J. Huang, X. Tian, and J. Ma. Cross-scale domain adaptation with comprehensive information for pansharpening. In *Int. Joint Conf. Artif. Intell*, pages 785–793, 2024. 1

[17] L. He, H. Ye, D. Xi, J. Li, A. Plaza, and M. Zhang. Tree-structured neural network for hyperspectral pansharpening. *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.*, 17: 2516–2530, 2023. 1

[18] L. He, Z. Fang, J. Li, J. Chanussot, and A. Plaza. Two spectral-spatial implicit neural representations for arbitrary-resolution hyperspectral pansharpening. *IEEE Trans. Geosci. Remote Sens.*, 2024. 1, 2

[19] W. He, Q. Yao, C. Li, N. Yokoya, and Q. Zhao. Non-local meets global: An integrated paradigm for hyperspectral denoising. In *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recog.*, pages 6868–6877, 2019. 5

[20] W. He, Q. Yao, C. Li, N. Yokoya, Q. Zhao, H. Zhang, and L. Zhang. Non-local meets global: An iterative paradigm for hyperspectral image restoration. *IEEE Trans. Pattern Anal. Mach. Intell.*, 44(4):2089–2107, 2020. 5

[21] Jo. Ho, A. Jain, and P. Abbeel. Denoising diffusion probabilistic models. In *Proc. Adv. Neural Inf. Process. Syst. (NeurIPS)*, pages 6840–6851, 2020. 2, 3

[22] S. Huang, M. Ren, Y. Yang, X. Wang, and Y. Wei. MFTN: A multi-scale feature transfer network based on IMatchformer for hyperspectral image super-resolution. In *Forty-first Int. Conf. Mach. Learn.*, 2024. 1

[23] B. Kawar, M. Elad, S. Ermon, and J. Song. Denoising diffusion restoration models. *Proc. Adv. Neural Inf. Process. Syst.*, 35:23593–23606, 2022. 3

[24] Tamara G Kolda and Brett W Bader. Tensor decompositions and applications. *SIAM Rev.*, 51(3):455–500, 2009. 2

[25] S. Li, S. Li, and L. Zhang. Hyperspectral and panchromatic images fusion based on the dual conditional diffusion models. *IEEE Trans. Geosci. Remote Sens.*, 2023. 3

[26] L. Loncan, L. B. De Almeida, J. M. Bioucas-Dias, X. Briottet, J. Chanussot, N. Dobigeon, S. Fabre, W. Liao, G. A. Licciardi, M. Simoes, et al. Hyperspectral pansharpening: A review. *IEEE Geosci. Remote Sens. Mag.*, 3(3):27–46, 2015. 1

[27] Y. Miao, L. Zhang, L. Zhang, and D. Tao. DDS2M: Self-supervised denoising diffusion spatio-spectral model for hyperspectral image restoration. In *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, pages 12086–12096, 2023. 3

[28] A. Quinn Nichol, P. Dhariwal, A. Ramesh, P. Shyam, P. Mishkin, B. Mcgrew, I. Sutskever, and M. Chen. GLIDE: Towards photorealistic image generation and editing with text-guided diffusion models. In *Int. Conf. Mach. Learn.*, pages 16784–16804. PMLR, 2022. 4

[29] L. Pang, X. Rui, L. Cui, H. Wang, D. Meng, and X. Cao. HIR-Diff: Unsupervised hyperspectral image restoration via improved diffusion models. In *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recog.*, pages 3005–3014, 2024. 1, 2, 3, 5, 6, 7, 8

[30] X. Rui, X. Cao, L. Pang, Z. Zhu, Z. Yue, and D. Meng. Unsupervised hyperspectral pansharpening via low-rank diffusion model. *Inf. Fusion*, 107:102325, 2024. 1, 2, 3, 5, 6, 7, 8

[31] K. Shi, J. Peng, J. Gao, Y. Luo, and S. Xu. Hyperspectral image denoising via double subspace deep prior. *IEEE Trans. Geosci. Remote Sens.*, 62:1–15, 2024. 5

[32] M. Simoes, J. Bioucas-Dias, L. B. Almeida, and J. Chanussot. A convex formulation for hyperspectral image superresolution via subspace-based regularization. *IEEE Trans. Geosci. Remote Sens.*, 53(6):3373–3388, 2014. 5, 6, 7

[33] Y. Song and S. Ermon. Generative modeling by estimating gradients of the data distribution. In *Proc. Adv. Neural Inf. Process. Syst. (NeurIPS)*, pages 11895–11907, 2019. 3

[34] Y. Song, J. Sohl-Dickstein, D. P. Kingma, A. Kumar, S. Ermon, and B. Poole. Score-based generative modeling through stochastic differential equations. In *Int. Conf. Learn. Representations*, 2021. 2, 3

[35] D. Ulyanov, A. Vedaldi, and V. Lempitsky. Deep image prior. In *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, pages 9446–9454, 2018. 5, 8

[36] G. Vivone, M. Dalla Mura, A. Garzelli, R. Restaino, G. Scarpa, M. O. Ulfarsson, L. Alparone, and J. Chanussot. A new benchmark based on recent advances in multispectral pansharpening: Revisiting pansharpening with classical and emerging pansharpening methods. *IEEE Geosci. Remote Sens. Mag.*, 9(1):53–81, 2020. 1

[37] W. Wang, L.-J. Deng, R. Ran, and G. Vivone. A general paradigm with detail-preserving conditional invertible network for image fusion. *Int. J. Comput. Vis.*, 132(4):1029–1054, 2024. 2

[38] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE Trans. Image Process.*, 13(4):600–612, 2004. 6

[39] C. Wu, D. Wang, Y. Bai, H. Mao, Y. Li, and Q. Shen. HSR-Diff: Hyperspectral image super-resolution via conditional diffusion models. In *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, pages 7083–7093, 2023. 3

[40] J.-L. Xiao, T.-Z. Huang, L.-J. Deng, Z.-C. Wu, X. Wu, and G. Vivone. Variational pansharpening based on coefficient estimation with nonlocal regression. *IEEE Trans. Geosci. Remote. Sens.*, 61:1–15, 2023. 2

[41] W. Xie, J. Lei, Y. Cui, Y. Li, and Q. Du. Hyperspectral pansharpening with deep priors. *IEEE Trans. Neural Netw. Learn. Syst.*, 31(5):1529–1543, 2019. 1

[42] Y. Xing, L. Qu, S. Zhang, K. Zhang, Y. Zhang, and L. Bruzzone. Crossdiff: Exploring self-supervised representation of pansharpening via cross-predictive diffusion model. *IEEE Trans. Image Process.*, 33:5496–5509, 2024. 3

[43] T. Xu, T.-Z. Huang, L.-J. Deng, and N. Yokoya. An iterative regularization method based on tensor subspace representation for hyperspectral image super-resolution. *IEEE Trans. Geosci. Remote Sens.*, 60:1–16, 2022. 2

[44] T. Xu, T.-Z. Huang, L.-J. Deng, J.-L. Xiao, C. Broni-Bediako, J. Xia, and N. Yokoya. A coupled tensor double-factor method for hyperspectral and multispectral image fusion. *IEEE Trans. Geosci. Remote Sens.*, 62:1–17, 2024. 2, 4

[45] J. Xue, Y.-Q. Zhao, Y. Bu, W. Liao, J. C.-W. Chan, and W. Philips. Spatial-spectral structured sparse low-rank representation for hyperspectral image super-resolution. *IEEE Trans. Image Process.*, 30:3084–3097, 2021. 2

[46] J. Yao, D. Hong, J. Chanussot, D. Meng, X. Zhu, and Z. Xu. Cross-attention in coupled unmixing nets for unsupervised hyperspectral super-resolution. In *Proc. Eur. Conf. Comput. Vis.*, pages 208–224. Springer, 2020. 1

[47] N. Yokoya, T. Yairi, and A. Iwasaki. Coupled nonnegative matrix factorization unmixing for hyperspectral and multispectral data fusion. *IEEE Trans. Geosci. Remote Sens.*, 50 (2):528–537, 2011. 1, 5, 6, 7

[48] Y. Yu, E. Pan, X. Wang, Y. Wu, X. Mei, and J. Ma. Unmixing before fusion: A generalized paradigm for multi-source-based hyperspectral image synthesis. In *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recog.*, pages 9297–9306, 2024. 1

[49] R. H. Yuhas, A. F. H. Goetz, and J. W. Boardman. Discrimination among semi-arid landscape endmembers using the spectral angle mapper (SAM) algorithm. In *Proc. Summaries 3rd Annu. JPL Airborne Geosci. Workshop*, pages 147–149, 1992. 6

[50] Y. Zhong, X. Wu, L.-J. Deng, and Z. Cao. SSDiff: Spatial-spectral integrated diffusion model for remote sensing pansharpening. *arXiv preprint arXiv:2404.11537*, 2024. 3

[51] Y.-W. Zhuo, T.-J. Zhang, J.-F. Hu, H.-X. Dou, T.-Z. Huang, and L.-J. Deng. A deep-shallow fusion network with multi-detail extractor and spectral attention for hyperspectral pansharpening. *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.*, 15:7539–7555, 2022. 8