This ICCV paper is the Open Access version, provided by the Computer Vision Foundation.
Except for this watermark, it is identical to the accepted version;
the final published version of the proceedings is available on IEEE Xplore.

# MP-HSIR: A Multi-Prompt Framework for Universal Hyperspectral Image Restoration

Zhehui Wu[1] Yong Chen[2] Naoto Yokoya[3,4] Wei He[1,4*]
[1] Wuhan University [2] Jiangxi Normal University [3] The University of Tokyo
[4] RIKEN Center for Advanced Intelligence Project
{wuzhehui, weihe1990}@whu.edu.cn, chen_yong@jxnu.edu.cn, yokoya@k.u-tokyo.ac.jp

## Abstract

*Hyperspectral images (HSIs) often suffer from diverse and unknown degradations during imaging, leading to severe spectral and spatial distortions. Existing HSI restoration methods typically rely on specific degradation assumptions, limiting their effectiveness in complex scenarios. In this paper, we propose **MP-HSIR**, a novel multi-prompt framework that effectively integrates spectral, textual, and visual prompts to achieve universal HSI restoration across diverse degradation types and intensities. Specifically, we develop a prompt-guided spatial-spectral transformer, which incorporates spatial self-attention and a prompt-guided dual-branch spectral self-attention. Since degradations affect spectral features differently, we introduce spectral prompts in the local spectral branch to provide universal low-rank spectral patterns as prior knowledge for enhancing spectral reconstruction. Furthermore, the text-visual synergistic prompt fuses high-level semantic representations with fine-grained visual features to encode degradation information, thereby guiding the restoration process. Extensive experiments on 9 HSI restoration tasks, including all-in-one scenarios, generalization tests, and real-world cases, demonstrate that MP-HSIR not only consistently outperforms existing all-in-one methods but also surpasses state-of-the-art task-specific approaches across multiple tasks. The code and models are available at* https://github.com/ZhehuiWu/MP-HSIR.

## 1. Introduction

Hyperspectral images (HSIs) provide significantly higher spectral resolution than RGB images, as illustrated in Figure 2 (a), making them essential for urban planning [20, 60], agricultural production [40, 54], and environmental monitoring [2, 57]. However, due to various environmental factors and imaging limitations, HSIs are often affected by degradations during acquisition, compromising image quality and limiting subsequent analysis. Consequently, HSI restoration is a critical step in the image processing pipeline.

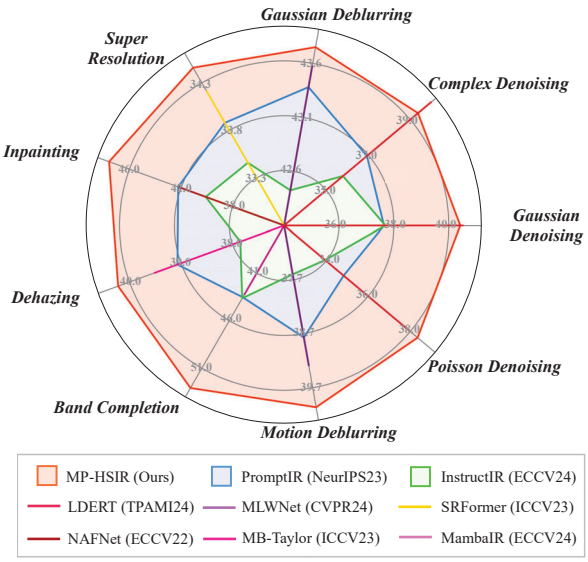
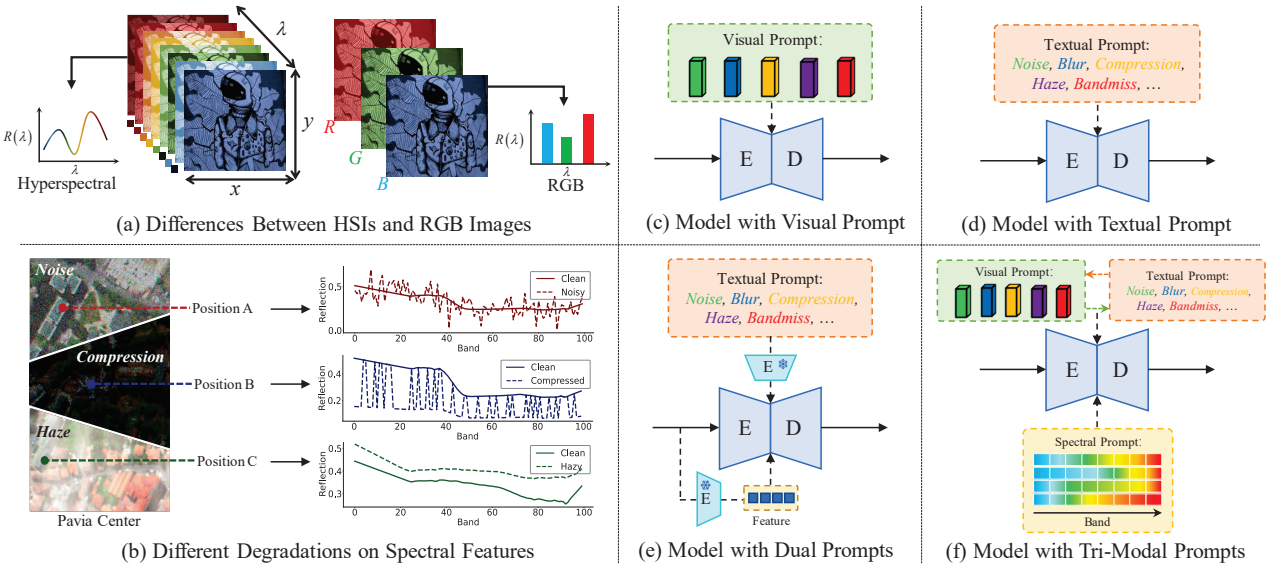



Figure 1. PSNR comparison with the state-of-the-art all-in-one and task-specific methods across 9 tasks. Average results across all test datasets. Best viewed in color.

Despite advancements in task-specific HSI restoration methods [9, 19, 33, 61, 67, 75], the diversity and complexity of degradations pose a significant challenge for a single model to handle multiple HSI restoration tasks. As illustrated in Figure 2 (b), different degradation types have distinct effects on the spectral features of HSIs. For instance, noise increases spectral fluctuations, compression reduces spectral reflectance in certain bands, and haze shifts the spectral curve globally. These diverse impacts hinder the generalization capability of task-specific methods across different scenarios. Therefore, a universal framework is required to address the unique distortions caused by different degradations.

In recent years, all-in-one image restoration methods have achieved significant progress in RGB images [24, 30, 63]. Among these, prompt learning-based approaches [1, 12, 17,

*Corresponding author




(a) Differences Between HSIs and RGB Images (c) Model with Visual Prompt (d) Model with Textual Prompt

(b) Different Degradations on Spectral Features (e) Model with Dual Prompts (f) Model with Tri-Modal Prompts

Figure 2. Overview of HSI characteristics and prompt-based all-in-one restoration methods. (a) Differences between HSIs and RGB images. (b) Impacts of different degradations on spectral features. (c) Visual prompt-based models. (d) Textual prompt-based models. (e) Dual-modality prompt-based models. (f) The proposed model integrating spectral, textual, and visual prompts.

34, 42, 50] have garnered attention for their flexibility in task adaptation and efficient degradation perception. Based on prompt modality, these methods can be categorized into visual, textual, and dual prompts. As shown in Figure 2 (c), visual prompt-based models [34, 50] guide restoration by introducing additional prompt components but suffer from limited interpretability. Figure 2 (d) illustrates textual prompt-based models [12, 17], which convert textual prompts into latent embeddings to assist restoration. Nevertheless, inconsistencies may arise when the semantic gap between text and images is too large. Figure 2 (e) depicts dual prompt-based models [1, 42], which leverage pre-trained vision-language models (VLMs) to extract deep representations from text instructions and degraded images, providing more precise degradation guidance. However, existing VLMs still exhibit limitations in modeling HSI degradation information. Specifically, **most all-in-one models fail to consider the unique spectral characteristics of HSIs, often resulting in spectral distortions in the restoration process**.

To address the challenges of all-in-one HSI restoration, we propose a universal multi-prompt framework, MP-HSIR, which integrates spectral, textual, and visual prompts, as illustrated in Figure 2 (f). In this framework, spectral prompts guide the spatial-spectral transformer for spectral recovery, while textual and visual prompts synergistically generate degradation-specific information to guide the global restoration. Specifically, the prompt-guided spatial-spectral transformer comprises two core components: spatial self-attention and dual-branch spectral self-attention. The spatial self-attention captures non-local spatial similarities, while the global spectral branch models long-range dependencies in the spectral domain. Meanwhile, the local spectral branch incorporates universal low-rank spectral patterns from the spectral prompts to enhance spectral feature reconstruction. Additionally, the text-visual synergistic prompt deeply fuses the semantic representations of textual prompts with the fine-grained features of visual prompts, enabling controlled restoration while embedding degradation information at the pixel level. As shown in Figure 1, our method outperforms existing all-in-one methods across 9 HSI restoration tasks and surpasses state-of-the-art task-specific methods in multiple tasks. Overall, our main contributions are as follows:

- We propose a novel multi-prompt framework for all-in-one HSI restoration, integrating spectral, textual, and visual prompts. The spectral prompt provides universal low-rank spectral patterns to enhance spectral reconstruction.
- We introduce a text-visual synergistic prompt that combines semantic representations from textual prompts with fine-grained features from visual prompts, enhancing the controllability, interpretability, and degradation adaptability of the HSI restoration process.
- Extensive experiments on 9 HSI restoration tasks and real-world scenarios demonstrate that MP-HSIR significantly outperforms compared all-in-one methods and surpasses state-of-the-art task-specific approaches in multiple tasks.

## 2. Related Work

### 2.1. Hyperspectral Image Restoration

Traditional HSI restoration methods mainly leverage prior knowledge, such as non-local similarity [44], low-rank struc-

ture [19, 37], and sparsity [66, 68], to constrain the solution space and employ optimization algorithms [10, 55] to solve the inverse problem. Deep learning-based HSI restoration methods directly learn the mapping between degraded and original images through deep neural networks, including convolutional neural networks (CNNs) [61, 72], transformers [31, 32, 71], and generative models [33]. Although these methods have demonstrated strong restoration capabilities, they are typically designed for specific types of degradation, such as noise removal [6, 27, 65], super-resolution [62, 75], and dehazing [43, 67], lacking universality and adaptability in practical applications.

Recent studies [26, 46, 48, 71] have explored using a single model for multiple HSI restoration tasks. DDS2M [46] designs a self-supervised diffusion model, while HIR-Diff [48] projects HSI into RGB space to leverage a pre-trained diffusion model for restoration. Nevertheless, these approaches frequently demand intricate hyperparameter tuning and exhibit slow inference speeds. Furthermore, they fail to adequately account for the distinct characteristics of different degradation types, limiting their effectiveness.

### 2.2. All-in-One Image Restoration

All-in-one image restoration [24, 42, 63] aims to establish a unified deep learning framework for restoring images with various degradations. However, developing an effective degradation-aware mechanism to distinguish different degradation remains challenging. AirNet [30] addresses this by employing contrastive learning to extract degradation representations, while HAIR [7] leverages a degradation-awareness classifier to capture global information and adaptively generate parameters via a hyper-selection network.

Recently, prompt learning [29, 41] has emerged as a promising paradigm for all-in-one image restoration due to its inherent flexibility and task-specific adaptability. PromptIR [50] integrates visual prompts for dynamic guidance, while PIP [34] refines them into degradation-aware and basic restoration prompts. To enable more controllable restoration, InstructIR [12] utilizes manually crafted instructions to guide the restoration process. Moreover, with the rise of VLMs, several studies [1, 42, 51] have explored the use of pre-trained VLMs to encode both textual and visual information for identifying degradation types. In the field of all-in-one HSI restoration, PromptHSI [28] introduces a text-to-feature modulation mechanism that applies textual prompts to composite degradation restoration in HSI, but it remains inherently limited to a single degradation mode. Overall, existing prompt-based all-in-one methods struggle to effectively characterize different degradation types in HSIs and lack guidance for spectral feature restoration.

## 3. Method

In this section, we present the multi-prompt framework for universal HSI restoration. First, we review the task of HSI restoration in 3.1. Then, we describe the overall architecture of the MP-HSIR framework in 3.2. Finally, we introduce the Spectral Self-Attention with Prompt Guidance module and the Text-Visual Synergistic Prompts module in 3.3 and 3.4.

### 3.1. Preliminary: Hyperspectral Image Restoration

HSI restoration aims to recover the clean image $\mathcal{X}$ from the degraded observation $\mathcal{Y}$ [14, 74], modeled as:

$$\mathcal{Y} = \mathcal{H}(\mathcal{X}) + \mathcal{N}, \tag{1}$$

where $\mathcal{H}$ represents multiplicative degradation operations and $\mathcal{N}$ denotes additive degradation operators. For the denoising task, $\mathcal{H}$ is an identity operation. For the deblurring task, $\mathcal{H}$ represents the blurring operation. For the inpainting task, $\mathcal{H}$ denotes a binary mask. In practice, the specific degradation in HSI often involves a combination of unknown multiplicative and additive operators. Without prior knowledge of the degradation type, selecting and deploying a specific restoration method requires maintaining multiple specialized models, inevitably leading to high storage costs and complex model management.

### 3.2. Overall Pipeline

Given the degraded observation $\mathcal{Y} \in \mathbb{R}^{H\times W\times B}$, where $H \times W$ represents the spatial resolution and $B$ denotes the number of spectral bands, MP-HSIR first applies a single-layer convolution to extract shallow features. As illustrated in Figure 3 (a), these features are processed through a three-level hierarchical encoder with downsampling operations. This process reduces spatial resolution while increasing channel capacity to obtain multi-scale deep features $F_l^e \in \mathbb{R}^{H\times W\times C}$, where $C$ is the number of channels. Each encoder-decoder pair consists of multiple Prompt-Guided Spatial-Spectral Transformer Blocks (PGSSTBs), with the number of blocks progressively increasing at deeper levels.

To integrate degradation information into the decoding process, we introduce the Text-Visual Synergistic Prompt (TVSP) module at each skip connection between the encoder and decoder. The TVSP module leverages textual and visual prompts to incorporate semantic guidance and fine-grained details, effectively capturing degradation characteristics for restoration. Finally, the restored HSI is generated via a $3 \times 3$ convolution layer followed by an image-level residual connection. The following sections detail the designs of the PGSSTB and TVSP modules.

### 3.3. Prompt-Guided Spatial-Spectral Transformer

The PGSSTB is the core component of MP-HSIR, integrating Spatial Self-Attention (SSA), Prompt-Guided Spectral

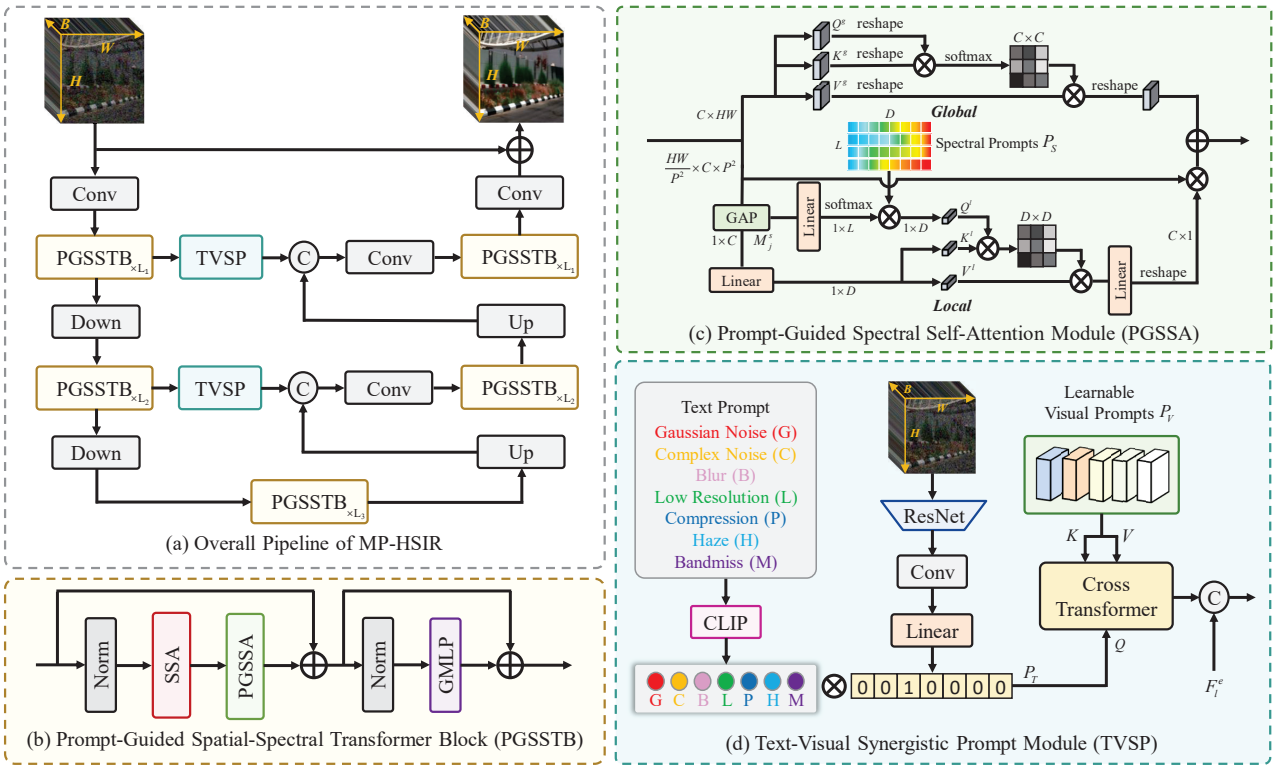



Figure 3. (a) The architecture of the proposed MP-HSIR. (b) Prompt-Guided Spatial-Spectral Transformer Block (PGSSTB). (c) Design of the Prompt-Guided Spectral Self-Attention (PGSSA) Module. (d) Text-Visual Synergistic Prompt (TVSP) Module.

Self-Attention (PGSSA), and a gated multi-layer perceptron (GMLP) [36]. To enhance training stability and convergence speed, layer normalization and skip connections [18] are incorporated, as shown in Figure 3 (b). SSA computes attention along the spatial dimension, capturing the non-local similarity in HSI. To improve computational efficiency while preserving the ability to model long-range dependencies, a sliding window [38] of size $P \times P$ is adopted.

While SSA captures spatial dependencies, spectral self-attention models long-range correlations across spectral bands, which is crucial for exploiting HSI's rich spectral information. To compute spectral similarity, the input features are transposed and reshaped into $F^{in} \in \mathbb{R}^{C \times HW}$ and processed through two complementary branches: global spectral self-attention and prompt-guided local spectral self-attention. The outputs from both branches are then fused and selectively refined using a GMLP. The following sections detail the two branches of PGSSA.

**Global Spectral Self-Attention.** Global spectral self-attention captures inter-channel dependencies across the entire HSI, as illustrated in the upper part of Figure 3 (c). Specifically, the input feature $F^{in}$ is first projected into the *query* $Q^g \in \mathbb{R}^{C \times HW}$, *key* $K^g \in \mathbb{R}^{C \times HW}$, and *value* $V^g \in \mathbb{R}^{C \times HW}$ using $1 \times 1$ point-wise convolutions $W^P$, followed by $3 \times 3$ depth-wise convolutions $W^D$:

$$[Q^g, K^g, V^g] = W^D W^P \left[F^{in}, F^{in}, F^{in}\right]. \tag{2}$$

To capture diverse spectral relationships across different bands, we further divide $Q^g$, $K^g$, and $V^g$ into multiple heads. Within each head, the global spectral attention map $A^g \in \mathbb{R}^{C \times C}$ is computed using the element-wise dot product between $Q^g$ and $K^g$, formulated as:

$$\begin{aligned} A^g &= \text{Softmax}\left(\frac{Q^g \cdot K^g}{\epsilon}\right), \\ \text{Attention}\left(Q^g, K^g, V^g\right) &= W^P\left(A^g V^g\right), \end{aligned} \tag{3}$$

where $\epsilon$ is a learnable scaling factor that normalizes the results of the dot product.

**Prompt-Guided Local Spectral Self-Attention.** Unlike global spectral self-attention, local spectral self-attention models spectral channel dependencies within localized spatial regions, as illustrated in the lower part of Figure 3 (c). Specifically, the input feature $F^{in}$ is divided into $\frac{HW}{P^2}$ non-overlapping local patches $\{M_1^{in}, M_2^{in}, \ldots, M_J^{in}\}$, where each patch $M_j^{in} \in \mathbb{R}^{C \times P^2}$ has the same spatial size as the spatial self-attention window. To extract representative spectral features $M_j^s \in \mathbb{R}^{1 \times C}$ from each patch, global average

pooling is applied along the spatial dimension:

$$M_j^s = \left( \text{AvgPool} \left( M_j^{in} \right) \right)^\top. \tag{4}$$

However, due to the susceptibility of spectral structures to various degradations, purely local modeling may lead to an overly dispersed attention distribution, preventing the model from focusing on the most informative spectral channels. To address this, a learnable spectral prompt, $P_S \in \mathbb{R}^{L \times D}$, is introduced to guide the attention mechanism toward essential spectral patterns, where $L$ denotes the number of universal low-rank spectral patterns in $P_S$, and $D$ represents the dimension. Specifically, $M_j^s$ undergoes a linear transformation followed by Softmax normalization to produce weights. These weights are then used to compute a weighted sum of $P_S$, which is subsequently projected to obtain the query $Q^l$. Meanwhile, $M_j^s$ is projected into the same low-dimensional space and further transformed to generate the key $K^l$ and value $V^l$, formulated as:

$$\begin{aligned} Q^l &= \text{Softmax} \left( M_j^s W_1^l \right) P_S W_3^l, \\ \left[ K^l, V^l \right] &= M_j^s W_2^l W_3^l, \end{aligned} \tag{5}$$

where the linear layer weights $W_1^l$, $W_2^l$, and $W_3^l$ have dimensions $C \times L$, $C \times D$, and $D \times D$, respectively.

The spectral prompt $P_S$, learned during training, provides universally applicable low-rank spectral patterns for local spectral self-attention, enabling any local spectral feature to be represented as a linear combination of these patterns. This enhances the performance of local spectral modeling and facilitates spectral information recovery, thereby improving the model's generalization across different datasets and degradation scenarios. Finally, the local spectral attention map $A^l \in \mathbb{R}^{D \times D}$ can be obtained by:

$$\begin{aligned} A^l &= \text{Softmax} \left( \frac{Q^l \cdot K^l}{\epsilon} \right), \\ \text{Attention} \left( Q^l, K^l, V^l \right) &= W^P \left( A^l V^l \right), \end{aligned} \tag{6}$$

where $W^P$ remaps the low-dimensional spectral features to the original dimension $M_j^{out} \in \mathbb{R}^{C \times 1}$, which is then used to weight $M_j^{in}$ to obtain the final output.

### 3.4. Text-Visual Synergistic Prompt

The spectral prompt exhibits strong robustness to HSIs with varying degradations but lacks explicit guidance on the degradation type. To address this limitation, we introduce the TVSP module, as illustrated in Figure 3 (d), which integrates textual prompts and learnable visual prompts to provide degradation-specific information for the restoration network.

First, we employ a degradation predictor $\Phi$ [11] to classify the degradation type of the input observation (details are provided in Section A of the supplementary material). Based on predefined textual descriptions $T_{text}$ corresponding to different degradation types, a frozen CLIP model [53] generates embeddings that serve as the textual prompt $P_T$. To bridge the domain gap between RGB and HSI, these embeddings are transformed into learnable parameters, enabling the network to adapt and refine the textual prompts during training for improved alignment with HSI characteristics. The final textual prompt is selected based on the classification result:

$$P_T = \Phi(\mathcal{X}) \cdot \text{Clip} \left( T_{text} \right). \tag{7}$$

However, the textual prompt $P_T$ provides only global degradation information and lacks pixel-level precision. To address this, we introduce a learnable visual prompt $P_V$ and fuse it with the textual prompt via a cross-attention. This allows $P_T$ to regulate global degradation while $P_V$ refines local features, enabling degradation information to be effectively integrated and adjusted, formally expressed as:

$$F_l^{out} = \text{Concat} \left( F_l^e, \text{Attention} \left( P_T, P_V \right) \right), \tag{8}$$

where the output $F_l^{out}$ is finally passed to the decoder via a skip connection. Compared to VLM-based dual prompts, TVSP is more flexible and adaptive, enabling prompt adjustments for different HSI domains.

## 4. Experiments

### 4.1. Experimental setup

**Datasets.** We employ 13 hyperspectral datasets for training and testing, comprising two natural scene HSI datasets and 11 remote sensing HSI datasets. The natural scene datasets include ICVL [3] and ARAD [4], while the remote sensing datasets consist of Xiong'an [69], WDC [78], Pavia City and Pavia University [22], Houston [64], Chikusei [70], Eagle [49], Berlin [47], APEX [23], Urban [5], and 10 EO-1 Hyperion scenes collected in this study. For the natural scene HSI datasets, 1,000 images are cropped into 64×64×31 patches for training, while an additional 100 non-overlapping images are reserved for testing. For the first eight remote sensing datasets, 80% of each image is cropped into 64×64×100 patches for training, with the remaining regions used for testing. The last three remote sensing datasets are designated for real-world experiments. Due to the substantial domain differences between natural scenes and remote sensing data, we train models separately on each dataset category. To ensure a comprehensive evaluation, multiple test datasets are randomly assigned to each task, ensuring that both natural scene and remote sensing HSI datasets are included in every task. Moreover, normalization is conducted on all datasets. A detailed description of the datasets is provided in Section B of the supplementary material.

**Settings.** To validate the effectiveness and generalization capability of the proposed model, we conducted evaluations

| Type | Method | *Gaussian Denoising (Sigma = 30, 50, 70)* | | | Method | *Complex Denoising (Case = 1, 2, 3, 4)* | | |
|---|---|---|---|---|---|---|---|---|
| | | ICVL [3] | ARAD [4] | Xiong'an [69] | | ICVL [3] | ARAD [4] | WDC [78] |
| | | PSNR / SSIM ↑ | PSNR / SSIM ↑ | PSNR / SSIM ↑ | | PSNR / SSIM ↑ | PSNR / SSIM ↑ | PSNR / SSIM ↑ |
| Task Specific | QRNN3D [61] | 39.99 / 0.947 | 39.18 / 0.932 | 36.06 / 0.829 | QRNN3D [61] | 41.60 / 0.966 | 41.11 / 0.960 | 30.82 / 0.868 |
| | SST [31] | 41.31 / 0.959 | 40.86 / 0.955 | 37.53 / 0.850 | SST [31] | 42.43 / 0.971 | 41.95 / 0.967 | 32.71 / 0.889 |
| | SERT [32] | 41.55 / 0.967 | 41.09 / 0.959 | 37.83 / 0.859 | SERT [32] | 43.31 / 0.976 | 42.87 / 0.973 | 33.31 / 0.903 |
| | LDERT [33] | 41.92 / 0.969 | 41.47 / 0.965 | 38.14 / 0.865 | LDERT [33] | 43.42 / 0.977 | 43.02 / 0.974 | 33.49 / 0.904 |
| All in One | AirNet [30] | 39.76 / 0.943 | 39.19 / 0.920 | 31.94 / 0.668 | AirNet [30] | 40.68 / 0.959 | 40.09 / 0.945 | 28.07 / 0.726 |
| | PromptIR [50] | 40.25 / 0.953 | 39.69 / 0.945 | 32.99 / 0.684 | PromptIR [50] | 41.29 / 0.965 | 40.71 / 0.954 | 28.83 / 0.735 |
| | PIP [34] | 40.87 / 0.958 | 40.25 / 0.952 | 32.64 / 0.674 | PIP [34] | 41.67 / 0.968 | 41.17 / 0.962 | 28.50 / 0.724 |
| | HAIR [7] | 40.51 / 0.956 | 39.91 / 0.949 | 32.54 / 0.679 | HAIR [7] | 40.63 / 0.958 | 40.15 / 0.947 | 28.18 / 0.729 |
| | InstructIR [12] | 41.02 / 0.960 | 40.32 / 0.954 | 31.74 / 0.666 | InstructIR [12] | 40.12 / 0.957 | 39.81 / 0.952 | 27.49 / 0.707 |
| | PromptHSI [28] | 40.65 / 0.960 | 40.04 / 0.956 | 38.07 / 0.881 | PromptHSI [28] | 39.14 / 0.955 | 38.75 / 0.936 | 33.77 / 0.912 |
| | MP-HSIR (Ours) | 41.62 / 0.964 | 41.09 / 0.960 | 38.81 / 0.897 | MP-HSIR (Ours) | 42.29 / 0.971 | 41.99 / 0.969 | 34.07 / 0.918 |
| **Type** | **Method** | ***Gaussian Deblurring (Radius = 9, 15, 21 & 7, 11, 15)*** | | | **Method** | ***Super-Resolution (Scale = 2, 4, 8)*** | | |
| | | ICVL [3] | PaviaC [22] | Eagle [49] | | ARAD [4] | PaviaU [22] | Houston [64] |
| | | PSNR / SSIM ↑ | PSNR / SSIM ↑ | PSNR / SSIM ↑ | | PSNR / SSIM ↑ | PSNR / SSIM ↑ | PSNR / SSIM ↑ |
| Task Specific | Stripformer [56] | 46.03 / 0.988 | 37.13 / 0.913 | 41.96 / 0.958 | SNLSR [21] | 36.05 / 0.898 | 30.55 / 0.730 | 31.70 / 0.787 |
| | FFTformer [25] | 46.65 / 0.988 | 37.96 / 0.921 | 42.76 / 0.962 | MAN [58] | 36.88 / 0.911 | 30.92 / 0.733 | 32.03 / 0.791 |
| | LoFormer [45] | 47.15 / 0.989 | 37.72 / 0.917 | 42.59 / 0.962 | ESSAformer [75] | 37.40 / 0.918 | 31.34 / 0.738 | 32.37 / 0.793 |
| | MLWNet [13] | 47.66 / 0.990 | 39.01 / 0.928 | 44.04 / 0.969 | SRFormer [77] | 38.02 / 0.922 | 31.80 / 0.751 | 32.80 / 0.803 |
| All in One | AirNet [30] | 47.21 / 0.989 | 37.76 / 0.918 | 42.42 / 0.963 | AirNet [30] | 36.73 / 0.910 | 30.93 / 0.735 | 32.00 / 0.791 |
| | PromptIR [50] | 47.67 / 0.990 | 38.72 / 0.927 | 43.74 / 0.968 | PromptIR [50] | 37.37 / 0.918 | 31.53 / 0.746 | 32.73 / 0.799 |
| | PIP [34] | 47.52 / 0.990 | 38.52 / 0.924 | 43.17 / 0.967 | PIP [34] | 38.36 / 0.926 | 31.77 / 0.749 | 32.89 / 0.805 |
| | HAIR [7] | 46.45 / 0.988 | 37.93 / 0.919 | 42.77 / 0.964 | HAIR [7] | 36.84 / 0.915 | 31.47 / 0.745 | 32.54 / 0.796 |
| | InstructIR [12] | 34.05 / 0.798 | 27.03 / 0.602 | 32.68 / 0.766 | InstructIR [12] | 36.51 / 0.913 | 31.34 / 0.741 | 32.52 / 0.793 |
| | PromptHSI [28] | 32.52 / 0.817 | 36.73 / 0.912 | 39.75 / 0.954 | PromptHSI [28] | 35.03 / 0.904 | 30.75 / 0.732 | 31.85 / 0.775 |
| | MP-HSIR (Ours) | 48.07 / 0.990 | 39.00 / 0.927 | 43.72 / 0.968 | MP-HSIR (Ours) | 38.25 / 0.924 | 31.97 / 0.760 | 33.06 / 0.810 |
| **Method** | ***Inpainting (Mask Rate = 0.7, 0.8, 0.9)*** | | **Method** | ***Dehazing (Omega = 0.5, 0.75, 1)*** | | **Method** | ***Completion (Rate = 0.1, 0.2 0.3)*** | |
| | ICVL [3] | Chikusei [70] | | PaviaU [22] | Eagle [49] | | ARAD [4] | Berlin [47] |
| | PSNR / SSIM ↑ | PSNR / SSIM ↑ | | PSNR / SSIM ↑ | PSNR / SSIM ↑ | | PSNR / SSIM ↑ | PSNR / SSIM ↑ |
| NAFNet [8] | 44.39 / 0.987 | 39.82 / 0.953 | SGNet [43] | 34.28 / 0.963 | 37.22 / 0.976 | NAFNet [8] | 47.04 / 0.995 | 38.35 / 0.907 |
| Restormer [73] | 45.79 / 0.990 | 36.33 / 0.899 | SCANet [16] | 36.59 / 0.978 | 39.64 / 0.985 | Restormer [73] | 48.34 / 0.995 | 35.07 / 0.606 |
| DDS2M [46] | 42.18 / 0.969 | 34.94 / 0.887 | MB-Taylor [52] | 37.99 / 0.983 | 41.03 / 0.991 | SwinIR [35] | 49.75 / 0.995 | 35.45 / 0.886 |
| HIR-Diff [48] | 38.91 / 0.949 | 37.65 / 0.916 | DCMPNet [76] | 37.20 / 0.985 | 40.24 / 0.990 | MambaIR [15] | 50.26 / 0.995 | 36.12 / 0.888 |
| AirNet [30] | 42.60 / 0.981 | 37.46 / 0.919 | AirNet [30] | 35.59 / 0.965 | 38.83 / 0.981 | AirNet [30] | 45.27 / 0.992 | 35.91 / 0.624 |
| PromptIR [50] | 46.38 / 0.990 | 38.07 / 0.930 | PromptIR [50] | 37.41 / 0.982 | 40.73 / 0.992 | PromptIR [50] | 46.60 / 0.994 | 40.45 / 0.652 |
| PIP [34] | 43.37 / 0.982 | 38.43 / 0.930 | PIP [34] | 37.62 / 0.982 | 40.74 / 0.990 | PIP [34] | 47.36 / 0.993 | 37.20 / 0.668 |
| HAIR [7] | 44.02 / 0.982 | 38.05 / 0.927 | HAIR [7] | 36.76 / 0.978 | 40.86 / 0.992 | HAIR [7] | 45.08 / 0.992 | 38.01 / 0.650 |
| InstructIR [12] | 44.07 / 0.986 | 36.11 / 0.907 | InstructIR [12] | 34.72 / 0.971 | 37.65 / 0.982 | InstructIR [12] | 51.31 / 0.997 | 35.97 / 0.580 |
| PromptHSI [28] | 41.48 / 0.972 | 37.33 / 0.946 | PromptHSI [28] | 36.77 / 0.974 | 39.78 / 0.984 | PromptHSI [28] | 47.34 / 0.994 | 43.36 / 0.972 |
| MP-HSIR (Ours) | 51.53 / 0.996 | 43.63 / 0.979 | MP-HSIR (Ours) | 39.59 / 0.986 | 42.41 / 0.995 | MP-HSIR (Ours) | 56.48 / 0.999 | 49.50 / 0.987 |

Table 1. **[All-in-one]** Quantitative comparison of all-in-one and state-of-the-art task-specific methods on 7 HSI restoration tasks. The best and second-best performances are highlighted in red and blue, respectively.

under three different settings. **(1) All-in-one:** A unified model was trained on seven HSI restoration tasks, including Gaussian denoising, complex denoising, Gaussian deblurring, super-resolution, inpainting, dehazing, and band completion. **(2) Generalization:** Few-shot and zero-shot learning were performed on unseen tasks. Motion deblurring was evaluated after fine-tuning, while Poisson denoising was assessed directly using the pre-trained model. **(3) Real-World:** For real-world denoising, following the settings of [31], the model was fine-tuned on the APEX dataset [23] and evaluated on the Urban dataset [5]. For real-world dehazing, the model was directly evaluated on the EO-1 Hyperion data. Detailed experimental settings for each task, including degradation synthesis models and intensity ranges, are provided in Section C of the supplementary material.

**Implementation Details.** Due to the difference in data volume, we employed distinct training strategies for natural scene and remote sensing HSI datasets. Both utilized the AdamW optimizer ($\beta_1 = 0.9, \beta_2 = 0.999$) with L1 loss and a batch size of 32. For natural HSIs, training was conducted for 100 epochs with an initial learning rate of $2 \times 10^{-4}$, which was gradually reduced to $1 \times 10^{-6}$ using cosine annealing [39]. In contrast, for remote sensing HSIs, the initial learning rate was set to $1 \times 10^{-4}$, and training was extended to 300 epochs. Given the higher spectral dimensionality of remote sensing HSIs, the width of the all-in-one model was increased by a factor of 1.5, while task-specific models were fine-tuned to achieve optimal performance.

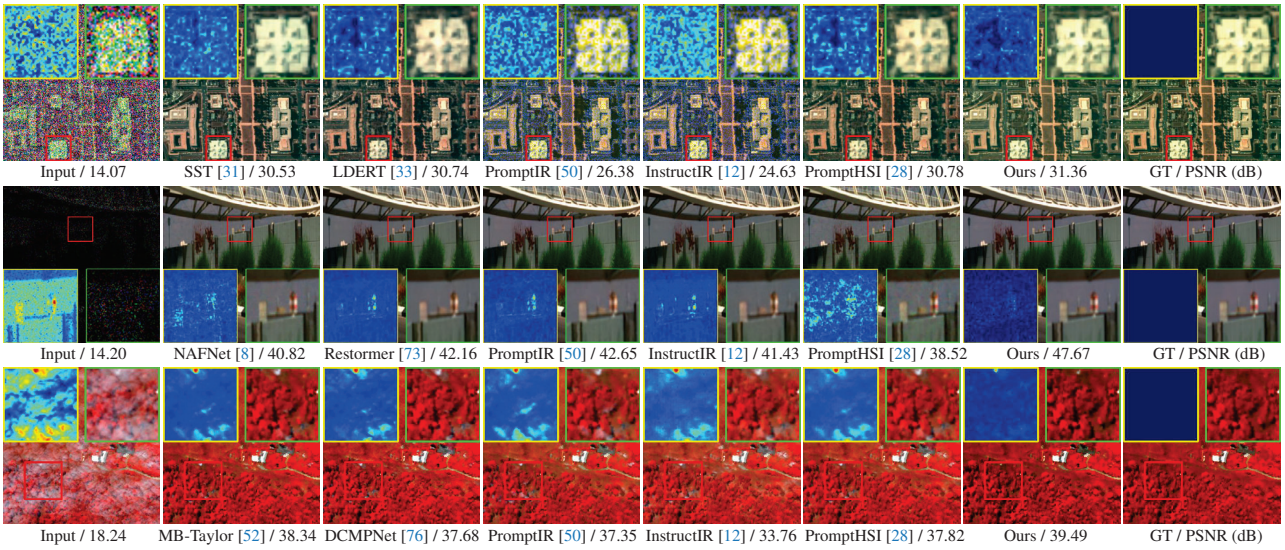



Figure 4. **[All-in-one]** Visual comparison results on ***complex denoising***, ***inpainting***, and ***dehazing***. The zoomed-in regions are highlighted with red boxes, the corresponding zoomed results are marked with green boxes, and the residual maps are outlined with yellow boxes.

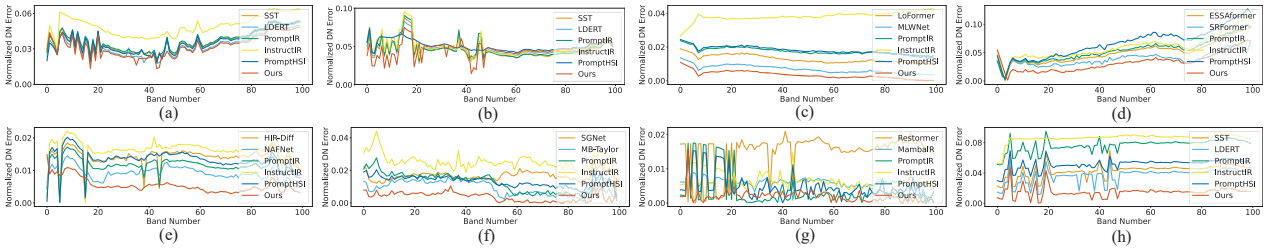



Figure 5. The spectral errors of the restored HSIs across multiple tasks. (a) ***Gaussian denoising***. (b) ***Complex denoising***. (c) ***Gaussian deblurring***. (d) ***Super-resolution***. (e) ***Inpainting***. (f) ***Dehazing***. (g) ***Band completion***. (h) ***Poisson denoising***. Zoom in for the best view.

| Type | Method | *Motion Deblurring* | | Methods | *Poisson Denoising* | |
|---|---|---|---|---|---|---|
| | | PSNR ↑ | SSIM ↑ | | PSNR ↑ | SSIM ↑ |
| Task Specific | Stripformer [56] | 37.91 | 0.948 | QRNN3D [61] | 35.98 | 0.910 |
| | FFTformer [25] | 38.42 | 0.950 | SST [31] | 37.51 | 0.914 |
| | LoFormer [45] | 38.94 | 0.953 | SERT [32] | 37.46 | 0.913 |
| | MLWNet [13] | 39.28 | 0.959 | LDERT [33] | 37.92 | 0.915 |
| All in One | AirNet [30] | 37.43 | 0.947 | AirNet [30] | 33.91 | 0.896 |
| | PromptIR [50] | 38.77 | 0.951 | PromptIR [50] | 34.83 | 0.902 |
| | PIP [34] | 38.55 | 0.950 | PIP [34] | 34.52 | 0.901 |
| | HAIR [7] | 37.58 | 0.948 | HAIR [7] | 34.38 | 0.901 |
| | InstructIR [12] | 36.58 | 0.942 | InstructIR [12] | 33.96 | 0.895 |
| | PromptHSI [28] | 36.79 | 0.945 | PromptHSI [28] | 37.25 | 0.912 |
| | MP-HSIR (Ours) | 40.06 | 0.965 | MP-HSIR (Ours) | 38.56 | 0.922 |

Table 2. **[Generalization]** The results of ***motion deblurring*** on the ICVL dataset [3] and ***Poisson denoising*** on the Houston dataset [64]. Task-specific methods are trained on the entire dataset, while all-in-one methods are fine-tuned using only 5% of the data.

**Evaluation Metrics and Comparisons Methods.** We employed PSNR and SSIM [59] as quantitative metrics for evaluation. The comparison methods included six all-in-one models, *i.e.*, AirNet [30], PromptIR [50], PIP [34], HAIR [7], InstructIR [12], and PromptHSI [28], along with 22 state-of-the-art task-specific models.

## 4.2. Comparison with state-of-the-art methods

**All-in-one.** Table 1 presents the quantitative performance of seven all-in-one methods and 22 task-specific approaches across seven HSI restoration tasks. Our method consistently outperforms all all-in-one methods and surpasses state-of-the-art task-specific approaches in multiple tasks, achieving significant PSNR improvements in inpainting, dehazing, and band completion. Furthermore, our method attains the best or near-best performance across all hyperspectral remote sensing datasets. Notably, InstructIR and PromptHSI exhibit anomalous results in the deblurring task due to their inability to effectively distinguish different types of HSI degradation. Visualization comparison of prompts is provided in Section D of the supplementary material.

Figure 4 presents visual comparisons for complex denoising, inpainting, and dehazing. The residual maps highlight the superiority of our method in structural recovery and fine-detail reconstruction. Figure 5 illustrates the normalized digital number (DN) error between the restored and ground truth images across spectral bands. Leveraging spectral prompts, our method achieves the most accurate spectral

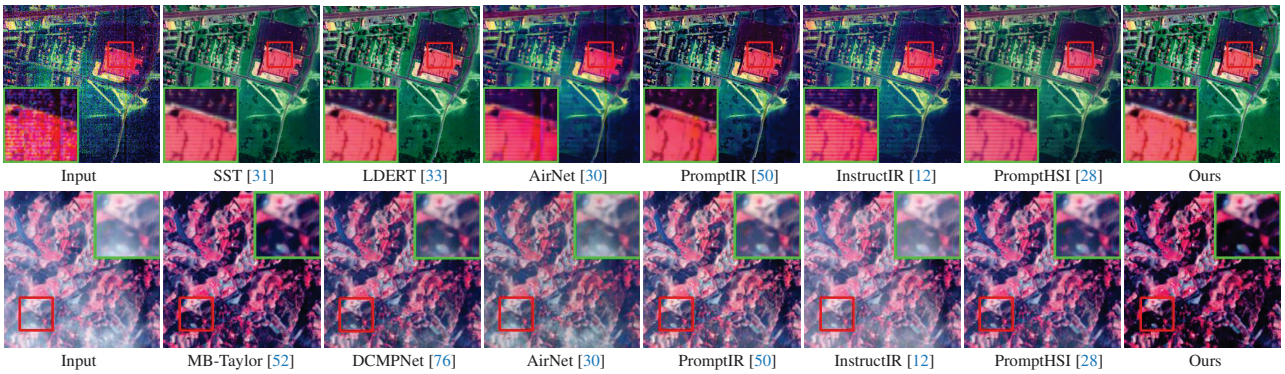



Figure 6. **[Real-world]** Visual comparison results on ***real denoising*** and ***real dehazing***. Zoom in for the best view.

| **Method** | PSNR ↑ | SSIM ↑ | Params (M) |
|---|---|---|---|
| Baseline (Only Spatial SA) | 39.24 | 0.963 | 20.93 |
| + Textual Prompt $P_T$ | 39.62 | 0.964 | 21.51 |
| + Visual Prompt $P_V$ | 39.57 | 0.964 | 23.68 |
| + Textual Prompt $P_T$ + Visual Prompt $P_V$ | 39.90 | 0.964 | 24.26 |
| + Global Spectral SA + $P_T$ + $P_T$ | 40.63 | 0.969 | 30.07 |
| + Local Spectral SA + $P_T$ + $P_V$ | 40.46 | 0.968 | 24.43 |
| + Local Spectral SA + $P_T$ + $P_V$ + Spectral Prompt $P_S$ | 41.05 | 0.971 | 25.10 |
| Full Model | **41.98** | **0.974** | 30.91 |

Table 3. Ablation study to verify the effectiveness of modules on Chikusei dataset in ***inpainting*** task with mask ratio = 0.9.

reconstruction across all tasks.

**Generalization.** Table 2 presents the experimental results for motion blur removal using few-shot learning and Poisson noise removal using zero-shot learning. The proposed method exhibits superior generalization ability compared to existing approaches under both learning paradigms.

**Real-world.** As shown in Figure 6, we present a comparative visualization of all-in-one and task-specific methods applied to real-world denoising and dehazing scenarios. The results highlight the proposed method's effectiveness in restoring texture details and spectral structure, while mitigating common artifacts like over-smoothing and spectral distortion.

### 4.3. Model Analysis

**Ablation Study.** To assess the effectiveness of each module in the proposed MP-HSIR framework, we conducted an ablation study, starting with a baseline model that includes only spatial self-attention (SA) and then progressively incorporating textual prompts $P_T$, learnable visual prompts $P_V$, global spectral self-attention, local spectral self-attention, and spectral prompts $P_S$. As shown in Table 3, both accuracy metrics for the inpainting task on the Chikusei dataset progressively improve with the addition of each module. Notably, the local spectral self-attention module guided by spectral prompts provides the most significant performance gain, attributed to the additional spectral guidance offered by spectral prompts. Results for other tasks are provided in Section E of the supplementary material.

**Visualization of Spectral Prompt.** To further investigate the role of spectral prompts, we visualize the activation distribution of universal low-rank spectral patterns under both noise and blur degradations in Figure 7. By selecting identical regions from both degraded images for comparison, we observe that despite different degradation types, the activation patterns remain highly consistent within the same regions. This indicates that spectral prompts effectively capture key spectral information in HSI and demonstrate strong robustness against different types of degradation.

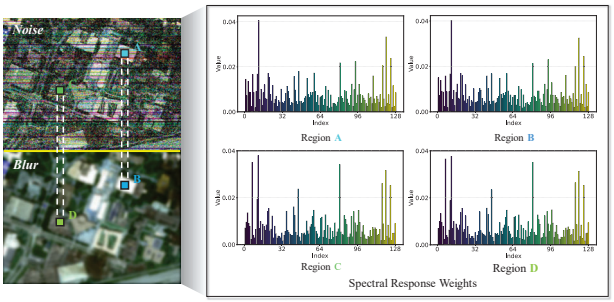



Figure 7. Visualization of the activation proportions for universal low-rank spectral patterns within the spectral prompt.

## 5. Conclusion

HSI restoration is challenging due to diverse and often unknown degradation types, while existing methods are typically limited to specific degradation scenarios. In this work, we propose MP-HSIR, a universal HSI restoration framework that integrates spectral, textual, and visual prompts. MP-HSIR introduces a novel prompt-guided spatial-spectral transformer, where spectral prompts provide universal low-rank spectral patterns to enhance spectral reconstruction. Furthermore, the text-visual synergistic prompts combine high-level semantic representations with fine-grained visual features to guide degradation modeling. Extensive experiments on 9 HSI restoration tasks—including all-in-one restoration, generalization testing, and real-world scenarios—demonstrate the superior robustness and adaptability of MP-HSIR under diverse degradation conditions.

## References


[1] Yuang Ai, Huaibo Huang, Xiaoqiang Zhou, Jiexiang Wang, and Ran He. Multimodal prompt perceiver: Empower adaptiveness generalizability and fidelity for all-in-one image restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 25432–25444, 2024. 1, 2, 3

[2] Margaret E Andrew and Susan L Ustin. The role of environmental context in mapping invasive plants with hyperspectral image data. *Remote Sensing of Environment*, 112(12):4301–4317, 2008. 1

[3] Boaz Arad and Ohad Ben-Shahar. Sparse recovery of hyperspectral signal from natural rgb images. In *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, the Netherlands, October 11–14, 2016, Proceedings, Part VII 14*, pages 19–34. Springer, 2016. 5, 6, 7

[4] Boaz Arad, Radu Timofte, Rony Yahel, Nimrod Morag, Amir Bernat, Yuanhao Cai, Jing Lin, Zudi Lin, Haoqian Wang, Yulun Zhang, et al. Ntire 2022 spectral recovery challenge and data set. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 863–881, 2022. 5, 6

[5] M. F. Baumgardner, L. L. Biehl, and D. A. Landgrebe. 220-band aviris hyperspectral image data set: June 12, 1992 indian pine test site 3, 2015. Available: https://purr.purdue.edu/publications/1947/1. 5, 6

[6] Théo Bodrito, Alexandre Zouaoui, Jocelyn Chanussot, and Julien Mairal. A trainable spectral-spatial sparse coding model for hyperspectral image restoration. *Advances in Neural Information Processing Systems*, 34:5430–5442, 2021. 3

[7] Jin Cao, Yi Cao, Li Pang, Deyu Meng, and Xiangyong Cao. Hair: Hypernetworks-based all-in-one image restoration. *arXiv preprint arXiv:2408.08091*, 2024. 3, 6, 7

[8] Liangyu Chen, Xiaojie Chu, Xiangyu Zhang, and Jian Sun. Simple baselines for image restoration. In *European Conference on Computer Vision*, pages 17–33. Springer, 2022. 6, 7

[9] Yong Chen, Wenzhen Lai, Wei He, Xi-Le Zhao, and Jinshan Zeng. Hyperspectral compressive snapshot reconstruction via coupled low-rank subspace representation and self-supervised deep network. *IEEE Transactions on Image Processing*, 33: 926–941, 2024. 1

[10] Yurong Chen, Hui Zhang, Yaonan Wang, Yimin Yang, and Jonathan Wu. Flex-dld: Deep low-rank decomposition model with flexible priors for hyperspectral image denoising and restoration. *IEEE Transactions on Image Processing*, 33: 1211–1226, 2024. 3

[11] Lu Chi, Borui Jiang, and Yadong Mu. Fast fourier convolution. *Advances in Neural Information Processing Systems*, 33:4479–4488, 2020. 5

[12] Marcos V Conde, Gregor Geigle, and Radu Timofte. Instructir: High-quality image restoration following human instructions. In *European Conference on Computer Vision*, pages 1–21. Springer, 2024. 1, 2, 3, 6, 7, 8

[13] Xin Gao, Tianheng Qiu, Xinyu Zhang, Hanlin Bai, Kang Liu, Xuan Huang, Hu Wei, Guoying Zhang, and Huaping Liu. Efficient multi-scale network with learnable discrete wavelet transform for blind motion deblurring. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2733–2742, 2024. 6, 7

[14] Pedram Ghamisi, Naoto Yokoya, Jun Li, Wenzhi Liao, Sicong Liu, Javier Plaza, Behnood Rasti, and Antonio Plaza. Advances in hyperspectral image and signal processing: A comprehensive overview of the state of the art. *IEEE Geoscience and Remote Sensing Magazine*, 5(4):37–78, 2017. 3

[15] Hang Guo, Jinmin Li, Tao Dai, Zhihao Ouyang, Xudong Ren, and Shu-Tao Xia. Mambair: A simple baseline for image restoration with state-space model. In *European Conference on Computer Vision*, pages 222–241. Springer, 2024. 6

[16] Yu Guo, Yuan Gao, Wen Liu, Yuxu Lu, Jingxiang Qu, Shengfeng He, and Wenqi Ren. Scanet: Self-paced semi-curricular attention network for non-homogeneous image dehazing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1885–1894, 2023. 6

[17] Yu Guo, Yuan Gao, Yuxu Lu, Huilin Zhu, Ryan Wen Liu, and Shengfeng He. Onerestore: A universal restoration framework for composite degradation. In *European Conference on Computer Vision*, pages 255–272. Springer, 2024. 1, 2

[18] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Identity mappings in deep residual networks. In *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part IV 14*, pages 630–645. Springer, 2016. 4

[19] Wei He, Quanming Yao, Chao Li, Naoto Yokoya, and Qibin Zhao. Non-local meets global: An integrated paradigm for hyperspectral denoising. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6868–6877, 2019. 1, 3

[20] Uta Heiden, Wieke Heldens, Sigrid Roessner, Karl Segl, Thomas Esch, and Andreas Mueller. Urban structure type characterization using hyperspectral remote sensing and height information. *Landscape and urban Planning*, 105 (4):361–375, 2012. 1

[21] Qian Hu, Xinya Wang, Junjun Jiang, Xiao-Ping Zhang, and Jiayi Ma. Exploring the spectral prior for hyperspectral image super-resolution. *IEEE Transactions on Image Processing*, 2024. 6

[22] Xin Huang and Liangpei Zhang. A comparative study of spatial approaches for urban mapping using hyperspectral rosis images over pavia city, northern italy. *International Journal of Remote Sensing*, 30(12):3205–3221, 2009. 5, 6

[23] Klaus I Itten, Francesco Dell'Endice, Andreas Hueni, Mathias Kneubühler, Daniel Schläpfer, Daniel Odermatt, Felix Seidel, Silvia Huber, Jürg Schopfer, Tobias Kellenberger, et al. Apex-the hyperspectral esa airborne prism experiment. *Sensors*, 8 (10):6235–6259, 2008. 5, 6

[24] Junjun Jiang, Zengyuan Zuo, Gang Wu, Kui Jiang, and Xianming Liu. A survey on all-in-one image restoration: Taxonomy, evaluation and future trends. *arXiv preprint arXiv:2410.15067*, 2024. 1, 3

[25] Lingshun Kong, Jiangxin Dong, Jianjun Ge, Mingqiang Li, and Jinshan Pan. Efficient frequency domain-based transformers for high-quality image deblurring. In *Proceedings of*

*the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5886–5895, 2023. 6, 7

[26] Yo-Yu Lai, Chia-Hsiang Lin, and Zi-Chao Leng. Hyperrestormer: A general hyperspectral image restoration transformer for remote sensing imaging. *arXiv preprint arXiv:2312.07016*, 2023. 3

[27] Zeqiang Lai, Chenggang Yan, and Ying Fu. Hybrid spectral denoising transformer with guided attention. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 13065–13075, 2023. 3

[28] Chia-Ming Lee, Ching-Heng Cheng, Yu-Fan Lin, Yi-Ching Cheng, Wo-Ting Liao, Chih-Chung Hsu, Fu-En Yang, and Yu-Chiang Frank Wang. Promptthsi: Universal hyperspectral image restoration framework for composite degradation. *arXiv preprint arXiv:2411.15922*, 2024. 3, 6, 7, 8

[29] Yiming Lei, Jingqi Li, Zilong Li, Yuan Cao, and Hongming Shan. Prompt learning in computer vision: a survey. *Frontiers of Information Technology & Electronic Engineering*, 25(1): 42–63, 2024. 3

[30] Boyun Li, Xiao Liu, Peng Hu, Zhongqin Wu, Jiancheng Lv, and Xi Peng. All-in-one image restoration for unknown corruption. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 17452–17462, 2022. 1, 3, 6, 7, 8

[31] Miaoyu Li, Ying Fu, and Yulun Zhang. Spatial-spectral transformer for hyperspectral image denoising. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 1368–1376, 2023. 3, 6, 7, 8

[32] Miaoyu Li, Ji Liu, Ying Fu, Yulun Zhang, and Dejing Dou. Spectral enhanced rectangle transformer for hyperspectral image denoising. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5805–5814, 2023. 3, 6, 7

[33] Miaoyu Li, Ying Fu, Tao Zhang, Ji Liu, Dejing Dou, Chenggang Yan, and Yulun Zhang. Latent diffusion enhanced rectangle transformer for hyperspectral image restoration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2024. 1, 3, 6, 7, 8

[34] Zilong Li, Yiming Lei, Chenglong Ma, Junping Zhang, and Hongming Shan. Prompt-in-prompt learning for universal image restoration. *arXiv preprint arXiv:2312.05038*, 2023. 2, 3, 6, 7

[35] Jingyun Liang, Jiezhang Cao, Guolei Sun, Kai Zhang, Luc Van Gool, and Radu Timofte. Swinir: Image restoration using swin transformer. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 1833–1844, 2021. 6

[36] Hanxiao Liu, Zihang Dai, David So, and Quoc V Le. Pay attention to mlps. *Advances in Neural Information Processing Systems*, 34:9204–9215, 2021. 4

[37] Yun-Yang Liu, Xi-Le Zhao, Yu-Bang Zheng, Tian-Hui Ma, and Hongyan Zhang. Hyperspectral image restoration by tensor fibered rank constrained optimization and plug-and-play regularization. *IEEE Transactions on Geoscience and Remote Sensing*, 60:1–17, 2021. 3

[38] Ze Liu, Yutong Lin, Yue Cao, Han Hu, Yixuan Wei, Zheng Zhang, Stephen Lin, and Baining Guo. Swin transformer: Hierarchical vision transformer using shifted windows. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 10012–10022, 2021. 4

[39] Ilya Loshchilov and Frank Hutter. Sgdr: Stochastic gradient descent with warm restarts. *arXiv preprint arXiv:1608.03983*, 2016. 6

[40] Bing Lu, Phuong D Dao, Jiangui Liu, Yuhong He, and Jiali Shang. Recent advances of hyperspectral imaging technology and applications in agriculture. *Remote Sensing*, 12(16):2659, 2020. 1

[41] Yuning Lu, Jianzhuang Liu, Yonggang Zhang, Yajing Liu, and Xinmei Tian. Prompt distribution learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5206–5215, 2022. 3

[42] Ziwei Luo, Fredrik K Gustafsson, Zheng Zhao, Jens Sjölund, and Thomas B Schön. Photo-realistic image restoration in the wild with controlled vision-language models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6641–6651, 2024. 2, 3

[43] Xiaofeng Ma, Qunming Wang, and Xiaohua Tong. A spectral grouping-based deep learning model for haze removal of hyperspectral images. *ISPRS Journal of Photogrammetry and Remote Sensing*, 188:177–189, 2022. 3, 6

[44] Matteo Maggioni, Vladimir Katkovnik, Karen Egiazarian, and Alessandro Foi. Nonlocal transform-domain filter for volumetric data denoising and reconstruction. *IEEE Transactions on Image Processing*, 22(1):119–133, 2012. 2

[45] Xintian Mao, Jiansheng Wang, Xingran Xie, Qingli Li, and Yan Wang. Loformer: Local frequency transformer for image deblurring. In *Proceedings of the 32nd ACM International Conference on Multimedia*, pages 10382–10391, 2024. 6, 7

[46] Yuchun Miao, Lefei Zhang, Liangpei Zhang, and Dacheng Tao. Dds2m: Self-supervised denoising diffusion spatio-spectral model for hyperspectral image restoration. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 12086–12096, 2023. 3, 6

[47] Akpona Okujeni, Sebastian van der Linden, and Patrick Hostert. Berlin-urban-gradient dataset 2009: An enmap preparatory flight campaign, 2016. Available: https://doi.org/10.2312/enmap.2016.0021. 5, 6

[48] Li Pang, Xiangyu Rui, Long Cui, Hongzhong Wang, Deyu Meng, and Xiangyong Cao. Hir-diff: Unsupervised hyperspectral image restoration via improved diffusion models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3005–3014, 2024. 3, 6

[49] Kabir Yunus Peerbhay, Onisimo Mutanga, and Riyad Ismail. Commercial tree species discrimination using airborne aisa eagle hyperspectral imagery and partial least squares discriminant analysis (pls-da) in kwazulu–natal, south africa. *ISPRS Journal of Photogrammetry and Remote Sensing*, 79:19–28, 2013. 5, 6

[50] Vaishnav Potlapalli, Syed Waqas Zamir, Salman H Khan, and Fahad Shahbaz Khan. Promptir: Prompting for all-in-one image restoration. *Advances in Neural Information Processing Systems*, 36:71275–71293, 2023. 2, 3, 6, 7, 8

[51] Chenyang Qi, Zhengzhong Tu, Keren Ye, Mauricio Delbracio, Peyman Milanfar, Qifeng Chen, and Hossein Talebi. Spire:

Semantic prompt-driven image restoration. In *European Conference on Computer Vision*, pages 446–464. Springer, 2024. 3

[52] Yuwei Qiu, Kaihao Zhang, Chenxi Wang, Wenhan Luo, Hongdong Li, and Zhi Jin. Mb-taylorformer: Multi-branch efficient transformer expanded by taylor formula for image dehazing. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 12802–12813, 2023. 6, 7, 8

[53] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, et al. Learning transferable visual models from natural language supervision. In *International Conference on Machine Learning*, pages 8748–8763. PmLR, 2021. 5

[54] Lankapalli Ravikanth, Digvir S Jayas, Noel DG White, Paul G Fields, and Da-Wen Sun. Extraction of spectral information from hyperspectral data and application of hyperspectral imaging for food and agricultural products. *Food and Bioprocess Technology*, 10:1–33, 2017. 1

[55] Dunbin Shen, Jianjun Liu, Zebin Wu, Jinlong Yang, and Liang Xiao. Admm-hfnet: A matrix decomposition-based deep approach for hyperspectral image fusion. *IEEE Transactions on Geoscience and Remote Sensing*, 60:1–17, 2021. 3

[56] Fu-Jen Tsai, Yan-Tsung Peng, Yen-Yu Lin, Chung-Chi Tsai, and Chia-Wen Lin. Stripformer: Strip transformer for fast image deblurring. In *European Conference on Computer Vision*, pages 146–162. Springer, 2022. 6, 7

[57] Yuting Wan, Xin Hu, Yanfei Zhong, Ailong Ma, Lifei Wei, and Liangpei Zhang. Tailings reservoir disaster and environmental monitoring using the uav-ground hyperspectral joint observation and processing: a case of study in xinjiang, the belt and road. In *IGARSS 2019-2019 IEEE International Geoscience and Remote Sensing Symposium*, pages 9713–9716. IEEE, 2019. 1

[58] Yan Wang, Yusen Li, Gang Wang, and Xiaoguang Liu. Multi-scale attention network for single image super-resolution. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5950–5960, 2024. 6

[59] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4):600–612, 2004. 7

[60] Cody Weber, Rahim Aguejdad, Xavier Briottet, J Avala, Sophie Fabre, Jean Demuynck, Emmanuel Zenou, Yannick Deville, Moussa Sofiane Karoui, Fatima Zohra Benhalouche, et al. Hyperspectral imagery for environmental urban planning. In *IGARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symposium*, pages 1628–1631. IEEE, 2018. 1

[61] Kaixuan Wei, Ying Fu, and Hua Huang. 3-d quasi-recurrent neural network for hyperspectral image denoising. *IEEE Transactions on Neural Networks and Learning Systems*, 32 (1):363–375, 2020. 1, 3, 6, 7

[62] Chanyue Wu, Dong Wang, Yunpeng Bai, Hanyu Mao, Ying Li, and Qiang Shen. Hsr-diff: Hyperspectral image super-resolution via conditional diffusion models. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 7083–7093, 2023. 3

[63] Gang Wu, Junjun Jiang, Kui Jiang, and Xianming Liu. Harmony in diversity: Improving all-in-one image restoration via multi-task collaboration. In *Proceedings of the 32nd ACM International Conference on Multimedia*, pages 6015–6023, 2024. 1, 3

[64] Hao Wu and Saurabh Prasad. Convolutional recurrent neural networks for hyperspectral data classification. *Remote Sensing*, 9(3):298, 2017. 5, 6, 7

[65] Jiahua Xiao, Yang Liu, and Xing Wei. Region-aware sequence-to-sequence learning for hyperspectral denoising. In *European Conference on Computer Vision*, pages 218–235. Springer, 2024. 3

[66] Ting Xie, Shutao Li, and Jibao Lai. Adaptive rank and structured sparsity corrections for hyperspectral image restoration. *IEEE Transactions on Cybernetics*, 52(9):8729–8740, 2021. 3

[67] Meng Xu, Yanxin Peng, Ying Zhang, Xiuping Jia, and Sen Jia. Aacnet: Asymmetric attention convolution network for hyperspectral image dehazing. *IEEE Transactions on Geoscience and Remote Sensing*, 61:1–14, 2023. 1, 3

[68] Wei Xu, Qing Zhu, Na Qi, and Dongpan Chen. Deep sparse representation based image restoration with denoising prior. *IEEE Transactions on Circuits and Systems for Video Technology*, 32(10):6530–6542, 2022. 3

[69] CEN Yi, Lifu Zhang, Xia Zhang, WANG Yueming, QI Wenchao, TANG Senlin, and Peng Zhang. Aerial hyperspectral remote sensing classification dataset of xiongan new area (matiwan village). *National Remote Sensing Bulletin*, 24(11): 1299–1306, 2020. 5, 6

[70] Naoto Yokoya and Akira Iwasaki. Airborne hyperspectral data over chikusei. *Space Appl. Lab., Univ. Tokyo, Tokyo, Japan, Tech. Rep. SAL-2016-05-27*, 5(5):5, 2016. 5, 6

[71] Dabing Yu, Qingwu Li, Xiaolin Wang, Zhiliang Zhang, Yixi Qian, and Chang Xu. Dstrans: Dual-stream transformer for hyperspectral image restoration. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 3739–3749, 2023. 3

[72] Qiangqiang Yuan, Qiang Zhang, Jie Li, Huanfeng Shen, and Liangpei Zhang. Hyperspectral image denoising employing a spatial–spectral deep residual convolutional neural network. *IEEE Transactions on Geoscience and Remote Sensing*, 57 (2):1205–1218, 2018. 3

[73] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, and Ming-Hsuan Yang. Restormer: Efficient transformer for high-resolution image restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5728–5739, 2022. 6, 7

[74] Hongyan Zhang, Wei He, Liangpei Zhang, Huanfeng Shen, and Qiangqiang Yuan. Hyperspectral image restoration using low-rank matrix recovery. *IEEE Transactions on Geoscience and Remote Sensing*, 52(8):4729–4743, 2013. 3

[75] Mingjin Zhang, Chi Zhang, Qiming Zhang, Jie Guo, Xinbo Gao, and Jing Zhang. Essaformer: Efficient transformer for hyperspectral image super-resolution. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 23073–23084, 2023. 1, 3, 6

[76] Yafei Zhang, Shen Zhou, and Huafeng Li. Depth information assisted collaborative mutual promotion network for single image dehazing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2846–2855, 2024. 6, 7, 8

[77] Yupeng Zhou, Zhen Li, Chun-Le Guo, Song Bai, Ming-Ming Cheng, and Qibin Hou. Srformer: Permuted self-attention for single image super-resolution. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 12780–12791, 2023. 6

[78] Feiyun Zhu. Hyperspectral unmixing: ground truth labeling, datasets, benchmark performances and survey. *arXiv preprint arXiv:1708.05125*, 2017. 5, 6