HSRMamba: Contextual Spatial-Spectral State Space Model for Single Hyperspectral Image Super-Resolution

Shi Chen[1] , **Lefei Zhang**[1*] and **Liangpei Zhang** [2]

[1]National Engineering Research Center for Multimedia Software, School of Computer Science, Wuhan University, Wuhan, 430072, P. R. China

[2]Aerospace Information Research Institute, Henan Academy of Sciences

{chenshi, zhanglefei, zlp62@whu.edu.cn}

Abstract

Mamba has demonstrated exceptional performance in visual tasks due to its powerful global modeling capabilities and linear computational complexity, offering considerable potential in hyperspectral image super-resolution (HSISR). However, in HSISR, Mamba faces challenges as transforming images into 1D sequences neglects the spatial-spectral structural relationships between locally adjacent pixels, and its performance is highly sensitive to input order, which affects the restoration of both spatial and spectral details. In this paper, we propose HSRMamba, a contextual spatial-spectral modeling state space model for HSISR, to address these issues both locally and globally. Specifically, a local spatial-spectral partitioning mechanism is designed to establish patch-wise causal relationships among adjacent pixels in 3D features, mitigating the local forgetting issue. Furthermore, a global spectral reordering strategy based on spectral similarity is employed to enhance the causal representation of similar pixels across both spatial and spectral dimensions. Finally, experimental results demonstrate our HSRMamba outperforms the state-of-the-art methods in quantitative quality and visual results. Code is available at: https://github.com/Tomchenshi/HSRMamba.

1 Introduction

Hyperspectral images (HSIs) typically comprise tens to hundreds of closely contiguous spectral bands spanning a broad spectral range, thereby capturing rich spectral and spatial information simultaneously [Xiao and Wei, 2023]. This capability enables a more accurate characterization of intrinsic spectral properties and subtle variations in materials, and consequently facilitates widespread applications in fields such as agriculture [Lu *et al.*, 2020], medical diagnosis [Lu and Fei, 2014], and remote sensing [Deng *et al.*, 2023]. However, achieving high spectral resolution often necessitates sacrifices in spatial resolution due to the inherent constraints of imaging hardware and procedures. Hyperspectral image
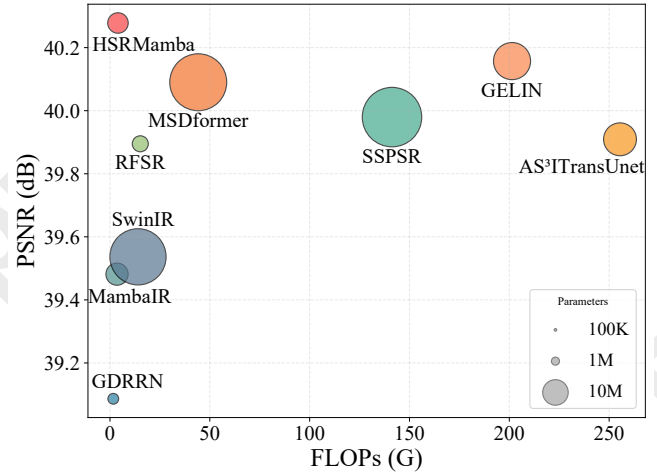
---
*Corresponding author

Figure 1: Comparison of trade-offs between model performance and effectiveness on Chikusei at scale factor ×4. Our method achieves superior performance with relatively low computational cost.

super-resolution (HSISR) addresses the challenge by converting low-resolution (LR) HSIs into the high-resolution (HR) HSIs, thereby enhancing spatial detail while preserving rich spectral information.

Generally, HSISR can be categorized into fusion-based approaches [Guo *et al.*, 2023; He *et al.*, 2023] and single hyperspectral image super-resolution (SHSR) [Zhang *et al.*, 2023], depending on whether auxiliary information is employed. Although fusion-based strategies can deliver superior results when aided by precisely registered additional images like multispectral image and panchromatic image, acquiring auxiliary information is often a significant challenge in real-world scenarios. Consequently, SHSR has attracted increasing interest. Over the past decade, deep learning SHSR methods have demonstrated remarkable advantages over traditional priors [Chen *et al.*, 2023a], owing to their potent capacity for modeling complex nonlinear relationships. Among these methods, Transformer-based networks [Liu *et al.*, 2022; Chen *et al.*, 2023b] have substantially enhanced HSISR performance by capturing long-range dependencies in spatial and spectral dimensions, thereby highlighting the critical importance of modeling long-range dependencies in HSISR.
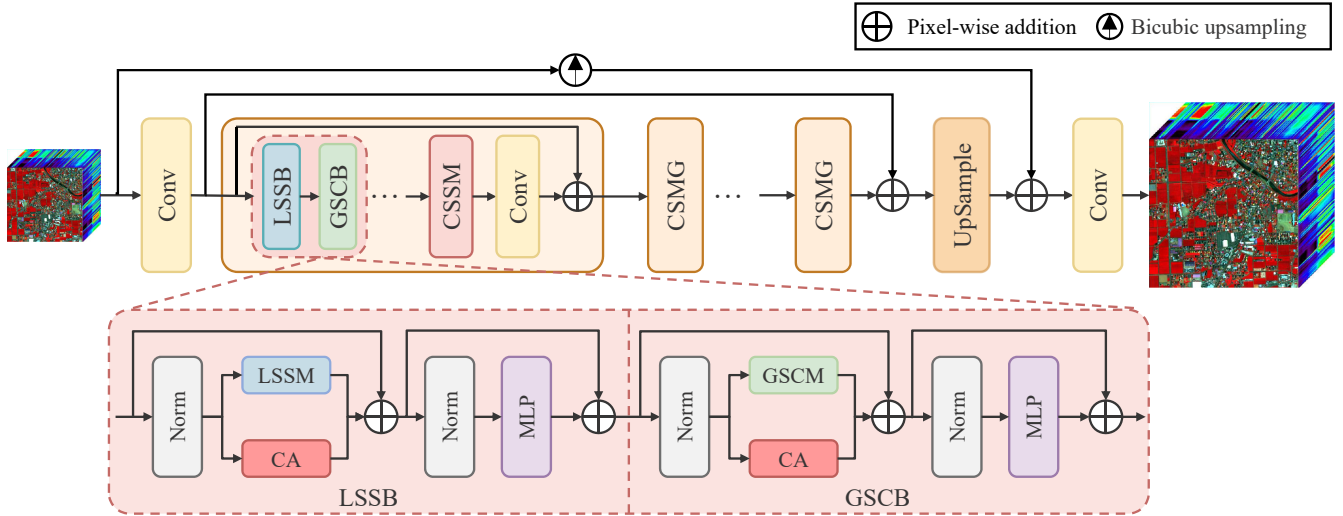
Figure 2: The overview of the proposed HSRMamba. HSRMamba primarily comprises multiple Contextual Spatial-Spectral Mamba Groups (CSMGs). Each CSMG consists of several consecutive Contextual Spatial-Spectral Mamba Blocks, which are further composed of a Local Spatial-Spectral Mamba Module (LSSM) and a Global Spectral Correlation Mamba Module (GSCM).

However, these models often face a trade-off between the global receptive field and computational efficiency [Li *et al.*, 2023; Zhang *et al.*, 2025; Dai *et al.*, 2024] due to the attention mechanism in Transformers, which severely limits further advancements in HSISR.

Recently, Mamba, an emerging state space model (SSM), has garnered significant attention for its ability to capture long-range dependencies with linear computational complexity. Despite its theoretical potential to address the challenge in attention mechanism, Mamba suffers from some issues for HSISR. Concretely, When transforming hyperspectral images into 1D sequences, the Mamba model overlooks the spatial-spectral structure among locally adjacent pixels, severely limiting its performance. Moreover, the output of Mamba is highly dependent on the input order, and a simple unfolding process disregards the modeling of spatial-spectral relationships between highly similar pixels, which is crucial for hyperspectral image restoration tasks.

To address these challenges, we propose HSRMamba, a state space model designed to efficiently capture both local and global long-range spatial-spectral dependencies. To the best of our knowledge, HSRMamba is the first attempt to employ Mamba tailored for SHSR. Specifically, we design a local spatial-spectral partitioning (LSSP) mechanism, which divides the 3D features into local spatial and spectral windows. By introducing the bidirectional SSM [Fu *et al.*, 2024] (BSSM), a local spaital-spectral Mamba module (LSSM) is constructed to establish causal relationships among neighboring pixels within these local 3-D windows, which enhances the network's local modeling capacity. Furthermore, we devise a global spectral reordering mechanism (GSRM), which rearranges the global spectra based on global spectral similarity. Subsequently we employ the Global Spectral Correlation Mamba module (GSCM) to strengthen causal modeling between highly similar pixels, facilitating detail reconstruction in hyperspectral images. This process further strengthens

causal relationships among pixels in the spatial and spectral dimensions that share high similarity. In summary, the main contributions of our work are as follows:

- We propose HSRMamba, the first SSM tailored for SHSR, which efficiently establishes local and global long-range spatial-spectral causal relationships.

- We devise a local spatial-spectral partitionin mechanism that captures patch-wise long-range spatial-spectral dependencies within a 3D window, thereby alleviating the inherent local pixel forgetting issue.

- We develop a global spectral-correlation Mamba module that globally extracts long-range spatial-spectral features by reordering spectra based on global spectral similarity, thus bolstering causal modeling among highly similar pixels.

- Extensive experiments on various datasets demonstrates the superiority and effectiveness of our proposed technique over the state-of-the-art methods.

2 Related Works

2.1 Single Hyperspectral Image Super-resolution

Without additional auxiliary data (e.g., panchromatic or multispectral images), SHSR exhibits broader applicability compared to fusion-based methods [Dian and Li, 2019; Vivone *et al.*, 2022; Dong *et al.*, 2023]. SHSR can be categorized into traditional methods [Dian *et al.*, 2017] based on hand-crafted priors and deep learning-based approaches [Wang *et al.*, 2023]. Over the past decade, numerous learning-based approaches, such as 3D convolution-based methods [Mei *et al.*, 2017; Li *et al.*, 2020; Fu *et al.*, 2021; Li *et al.*, 2021], group strategy-based methods [Li *et al.*, 2018; Jiang *et al.*, 2020; Li *et al.*, 2022; Wang *et al.*, 2022b; Wang *et al.*, 2022a], and Transformer-based methods [Wu *et al.*, 2023;

Chen *et al.*, 2024], have demonstrated significantly superior performance compared to traditional approaches.

In the above methods, convolution-based with 2D or 3D convolutions mainly focused on local spatial-spectral features, overlooking long-range spatial-spectral dependencies. Recently, Cai et al. [Cai *et al.*, 2022] proposed a spectral-wise multi-head self-attention for HSI reconstruction. Wang et al. [Wu *et al.*, 2023] introduced 3D-THSR, combining spectral self-attention with 3D convolutions to model spatial-spectral features in a global receptive field. Hu et al. [Hu *et al.*, 2024] transfered the HSI SR to the abundance domain with spectral-wise non-local attention to effectively incorporates valuable knowledge. Nevertheless, the computational complexity of Transformer networks grows quadratically with the input size, which significantly increases the demand for hardware resources when dealing with high-dimensional data such as hyperspectral images. While window-based self-attention mechanisms reduce computational costs by limiting the attention range, they fail to fully address the issue of high complexity. To overcome this limitation, we propose the Mamba network, a linear modeling framework designed to efficiently achieve hyperspectral image super-resolution reconstruction.

2.2 State Space Model

State Space Model [Gu *et al.*, 2022] are a mathematical framework designed to model temporal or sequential dependencies efficiently. More recently, Mamba, a SSMs-based model with linear computational complexity, has garnered significant attention for outperforming Transformers in natural language processing [Gu and Dao, 2023] and computer vision [Zhu *et al.*, 2024; Huang *et al.*, 2024] tasks. Subsequently, some Mamba networks designed for low-level vision tasks [Guo *et al.*, 2024; Qiao *et al.*, 2024; Xiao *et al.*, 2024; Zhi *et al.*, 2024; Fu *et al.*, 2024] have been proposed. Nonetheless, these methods are not suited for hyperspectral images. Firstly, they fail to consider the rich spectral information and the correlations between spatial and spectral dimensions inherent in hyperspectral data. Secondly, the Mamba network suffers from a local pixel forgetting issue, limiting its ability to model the structural information across spatial and spectral dimensions. Lastly, the performance of the Mamba network is highly dependent on the input sequence order, making it unable to fully exploit the relationships between highly similar pixels, which are crucial for image restoration.

3 Method

3.1 Preliminaries: State Space Models

SSMs provide a mathematical framework for modeling systems governed by latent states and their transitions over time. SSMs effectively model temporal dependencies in sequential data, and can be formulated as linear ordinary differential equation:

$$h'(t) = \mathbf{A}h(t) + \mathbf{B}x(t),$$
$$y(t) = \mathbf{C}h(t) + \mathbf{D}x(t), \qquad (1)$$

where $h(t) \in \mathbb{R}^N$ is the hidden state vector at time t, $\mathbf{A} \in \mathbb{R}^{N \times N}$, $\mathbf{B} \in \mathbb{R}^{N \times 1}$, $\mathbf{C} \in \mathbb{R}^{1 \times N}$ and $\mathbf{D} \in \mathbb{R}$ are the weight parameters, and N is the hidden state size.

Then, Eq.1 can be discretized using the zeroth-order hold (ZOH) rule, which transforms the continuous parameters $\mathbf{A}$, $\mathbf{B}$ to discrete parameters $\overline{\mathbf{A}}$, $\overline{\mathbf{B}}$ by the the timescale parameter Δ. It can be defined as:

$$\overline{\mathbf{A}} = \exp(\Delta \mathbf{A}),$$
$$\overline{\mathbf{B}} = (\Delta \mathbf{A})^{-1} (\exp(\Delta \mathbf{A}) - \mathbf{I}) \cdot \Delta \mathbf{B}. \qquad (2)$$

After discretization, the discrete version of Eq. 1 can be defined in the following RNN form:

$$h_k = \overline{\mathbf{A}} h_{k-1} + \overline{\mathbf{B}} x_k,$$
$$y_k = \mathbf{C} h_k + \mathbf{D} x_k. \qquad (3)$$

Furthermore, the SSM computation can be extended to a convolutional form as:

$$\overline{\mathbf{K}} \triangleq \left(\mathbf{C}\overline{\mathbf{B}}, \mathbf{C}\overline{\mathbf{A}}\overline{\mathbf{B}}, \dots, \mathbf{C}\overline{\mathbf{A}}^{L-1}\overline{\mathbf{B}} \right),$$
$$\mathbf{y} = \mathbf{x} * \overline{\mathbf{K}}, \qquad (4)$$

where L is the length of the input sequence and $*$ is the convolution operation, and $\overline{\mathbf{K}} \in \mathbb{R}^L$ respresents a structured convolutional kernel. Recently, leveraging a dynamic representation mechanism, Selective State Space Model (Mamba) enhances the long-range modeling capabilities of state space models with linear computational complexity.

3.2 Overview Architecture

As shown in Fig. 2, the overall architecture of HSRMamba consists of three primary components: a shallow feature extraction module, a deep feature extraction module, and an upsampling module. Assuming the input LR hyperspectral image is $I_{\text{LR}} \in \mathbb{R}^{H \times W \times B}$, where H, W, and B represent the spatial height, width, and the number of spectral bands, respectively. The SR hyperspectral image is denoted as $I_{\text{SR}} \in \mathbb{R}^{sH \times sW \times B}$, where s is the scaling factor for super-resolution reconstruction. The shallow features is extracted by using 3×3 convolutional layer. This process captures the structural information and adjusts the feature channel dimensions, which can be denoted as:

$$F_0 = f_s(I_{\text{LR}}), \qquad (5)$$

where $f_s(\cdot)$ represents 3×3 convolution, and F_0 denotes the extracted shallow features. Next, the shallow features are passed through the deep feature extraction module, consisting of multiple cascaded contextual spatial-spectral Mamba group (CSMG). The long-range spatial-spectral dependencies in deep features are extracted locally and globally. The process is defined as:

$$F_n = H_n(F_{n-1}), \quad n = 1, 2, \dots, N, \qquad (6)$$

where $H_n(\cdot)$ denotes the function of the n-th CSMG, and F_n represents the deep spatial-spectral features extracted by the n-th CSMG. Each CSMG consists of multiple consecutive contextual spatial-spectral Mamba modules (CSSM), each of which is composed of a local spatial-spectral Mamba block (LSSB) and a global spectral-correlation Mamba block (GSCB). Additionally, convolution is employed at the tail of each group to introduce inductive bias into the network.

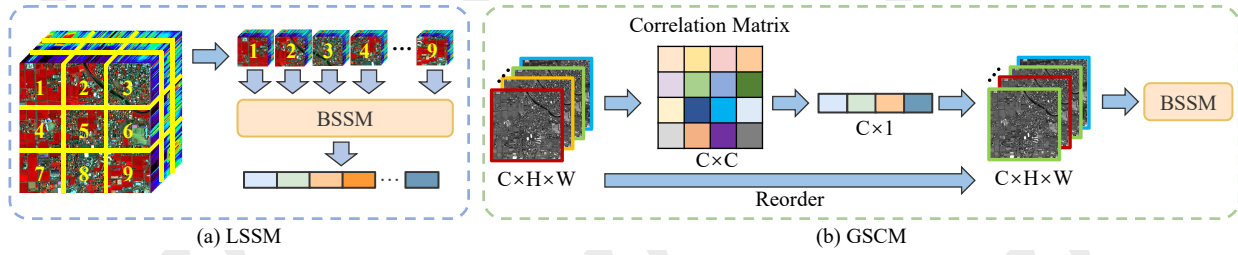(a) LSSM   (b) GSCM

Figure 3: The pipeline of the designed LSSM (left) and GSCM (right).

By leveraging long skip connections between shallow and deep features, the network focuses on learning high-frequency information beneficial for image reconstruction, improving both training efficiency and reconstruction performance. The merged deep features are then passed through a pixelshuffle layer to produce the upsampling deep features. Finally, SR HSI is obtained through long skip connections between the input features after bicubic upsampling and deep features.

3.3 Local Spaital-Spectral Mamba Module

As the performance of Mamba is heavily reliant on the order of the input, converting a hyperspectral image into a 1-D sequence for sequential scanning poses significant challenges in effectively establishing causal relationships between locally adjacent pixels across spatial and spectral dimensions. This inherent limitation greatly hinders the Mamba network's effectiveness in HSISR. Therefore, we design a local scanning mechanism capable of precisely capturing detailed connections between locally adjacent pixels, effectively addressing the limitations of Mamba in local causal modeling.

As shown in Fig. 2, LSSB consists of layer normalization, LSSM, channel atteniton (CA), and linear mapping layer. LSSM extract long-range spatial-spectral information from locally scanned regions. Additionally, CA mechanism is introduced to add inductive bias property into the network. Finally, the processed information is refined through linear layers, enhancing the module's capability for deep feature extraction. Assuming the input of the j-th local spatial-spectral Mamba module in the contextual spatial-spectral Mamba component is F_j, the above process can be expressed as:

$$\hat{F}_L^j = \text{LSSM}(\text{LN}(F_j)) + \text{CA}(\text{LN}(F_j)) + F_j,$$
$$F_L^j = \text{MLP}(\text{LN}(\hat{F}_L^j)) + \hat{F}_L^j, \tag{7}$$

where $\text{LN}(\cdot)$ denotes the layer normalization function, $\text{LSSM}(\cdot)$ represents the local spatial-spectral Mamba block, $\text{CA}(\cdot)$ denotes the channel attention mechanism, $\text{MLP}(\cdot)$ represents the linear layer function, and F_L^j is the final output of the module. LSSM consists of a LSSP and a BSSM [Fu *et al.*, 2024]. As illustrated in Fig. 3, the spatial-spectral partitioning mechanism divides the input features into $\frac{H}{h} \times \frac{W}{w} \times \frac{C}{c}$ local 3D feature blocks, each of size $h \times w \times c$, along the spatial and spectral dimensions. BSSM captures long-range spatial-spectral dependencies within each 3D local feature block. This process effectively addresses the local pixel forgetting issue inherent in traditional Mamba modules.
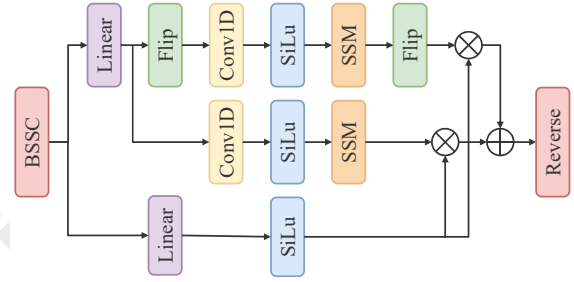

Figure 4: The pipeline of the proposed BSSM.

In this paper, we introduce a BSSM [Fu *et al.*, 2024] as the foundational unit for global spatial-spectral feature extraction. As shown in Fig. 4, the BSSM module takes the feature sequence from LSSP or GSRM as input of bidirectional branches and models comprehensive global spatial-spectral causal relationships from different directions via bidirectional spatial-spectral continuous scanning (BSSC). This design complementarily enhances contextual long-range dependency modeling, making it suited for hyperspectral images.

3.4 Global Spectral Correlation Mamba Module

In the Mamba network, the output of the current input relies solely on the preceding input in the data sequence. However, the recovery of fine details during the super-resolution process heavily depends on global similar pixels. To address this limitation, we propose a global spectral reordering mechanism and design the global spectral correlation Mamba.

As shown in Fig. 2, GSCB consists of layer normalization, the designed GSCM module, CA, and linear mapping layer. GSCM enhances glboal spatial-spectral causal modeling through the proposed global spectral reordering mechanism. Additionally, this module incorporates a channel attention mechanism to introduce inductive bias into the network. Finally, a linear layer is employed to enhance the representation capabilities of this paper.

Assuming the input to the j-th global spectral correlation Mamba module in the contextual spatial-spectral Mamba component is F_j, the above process can be expressed as:

$$\hat{F}_G^j = \text{GSCM}(\text{LN}(F_L)) + \text{CA}(\text{LN}(F_L)) + F_L,$$
$$F_G^j = \text{MLP}(\text{LN}(\hat{F}_G^j)) + \hat{F}_G^j, \tag{8}$$

where $\text{LN}(\cdot)$ denotes the layer normalization function, $\text{GSCM}(\cdot)$ represents the global spectral correlation Mamba

Method	Scale	Chikusei					Houston				
		PSNR↑	SSIM↑	SAM↓	CC↑	ERGAS↓	PSNR↑	SSIM↑	SAM↓	CC↑	ERGAS↓
GDRRN [Li *et al.*, 2018]	×4	39.0864	0.9265	3.0536	0.9421	5.7972	44.2964	0.9730	2.5347	0.9760	2.4700
SwinIR [Liang *et al.*, 2021]	×4	39.5366	0.9364	2.8327	0.9456	5.6280	46.0971	0.9808	1.9463	0.9864	2.0039
MambaIR [Guo *et al.*, 2024]	×4	39.4816	0.9353	2.9178	0.9456	5.6303	45.5947	0.9786	1.9953	0.9849	2.1239
SSPSR [Jiang *et al.*, 2020]	×4	39.9797	0.9393	2.4864	0.9528	5.1905	45.6017	0.9778	1.9650	0.9850	2.1380
RFSR [Wang *et al.*, 2022b]	×4	39.8950	0.9382	2.4656	0.9517	5.2334	45.8677	0.9792	1.8304	0.9858	2.0659
GELIN [Wang *et al.*, 2022a]	×4	40.1573	0.9410	2.4266	0.9543	5.0314	45.8715	0.9790	1.8759	0.9859	2.0778
AS³ITransUNet [Xu *et al.*, 2023]	×4	39.9093	0.9377	2.6056	0.9519	5.1900	45.8819	0.9792	1.8679	0.9862	2.0731
MSDformer [Chen *et al.*, 2023b]	×4	40.0902	0.9405	2.3981	0.9539	5.0818	46.2015	0.9807	1.7776	0.9870	1.9962
Ours	×4	**40.2781**	**0.9441**	**2.3160**	**0.9557**	**5.0131**	**46.9653**	**0.9838**	**1.6577**	**0.9891**	**1.8277**
GDRRN [Li *et al.*, 2018]	×8	34.7395	0.8199	5.0967	0.8381	9.6464	38.2592	0.9085	4.9045	0.9138	4.9135
SwinIR [Liang *et al.*, 2021]	×8	34.8785	0.8307	5.0413	0.8465	9.4743	39.4013	0.9194	4.0586	0.9370	4.3333
MambaIR [Guo *et al.*, 2024]	×8	35.1962	0.8365	4.7499	0.8556	9.0543	39.3049	0.9193	4.1935	0.9358	4.3648
SSPSR [Jiang *et al.*, 2020]	×8	35.1643	0.8299	4.6911	0.8560	9.0504	39.2844	0.9164	4.2673	0.9346	4.4212
RFSR [Wang *et al.*, 2022b]	×8	35.5049	0.8405	4.2785	0.8661	8.6338	39.4899	0.9211	3.8403	0.9379	4.2967
GELIN [Wang *et al.*, 2022a]	×8	35.6496	0.8464	4.1354	0.8707	8.4520	39.6387	0.9216	3.9231	0.9393	4.2453
AS³ITransUNet [Xu *et al.*, 2023]	×8	35.4999	0.8408	4.4746	0.8661	8.6793	39.8196	0.9254	3.9035	0.9422	4.1466
MSDformer [Chen *et al.*, 2023b]	×8	35.5914	0.8452	4.1381	0.8693	8.5203	39.7452	0.9227	3.6613	0.9417	4.2110
Ours	×8	**35.6812**	**0.8474**	**4.1148**	**0.8724**	**8.4508**	**39.9309**	**0.9265**	**3.5627**	**0.9443**	**4.1138**

Table 1: Quantitative performance on the Chikusei dataset and Houston dataset at different scale factors. Bold represents the best result and underline represents the second best.

block, $\mathrm{CA}(\cdot)$ denotes the channel attention mechanism, and $\mathrm{MLP}(\cdot)$ represents the linear layer function. $\hat{F}_G^j$ is the intermediate output of the j-th global spectral correlation Mamba module, and F_G^j is the final output of the module.

As illustrated in Fig. 3, the global spectral correlation Mamba consists of the GSRM and the BSSM. The global spectral reordering mechanism first computes the correlation coefficient matrix between spectral features. It then calculates the average of correlation coefficients among each spectrum as the global correlation value. Finally, the module reorders the spectral features based on their global correlation values, ensuring that pixels with high spectral correlation are closer in the spatial-spectral dimensions. This process significantly enhances the Mamba module's performance in extracting long-range spatial-spectral features.

3.5 Loss Function

The network is optimized using three losses: l_1 loss, spectral angle mapper (SAM) loss, and gradient loss in both spatial and spectral domains. The l_1 loss calculates the absolute pixel-wise difference between the reconstructed and original hyperspectral images, encouraging sharper and more detailed results compared to l_2 loss. The SAM loss ensures spectral consistency by considering the spectral characteristics of the data. Gradient loss enhances image sharpness by focusing on differences between adjacent pixels. The total loss is formulated as:

$$\mathcal{L}_{total}(\theta) = \mathcal{L}_1 + \lambda_s \mathcal{L}_{sam} + \lambda_g \mathcal{L}_{gra}, \quad (9)$$

where λ_s and λ_g are hyper-parameters that balance the losses, with $\lambda_s = 0.3$ and $\lambda_g = 0.1$ used empirically. The detailed function can be expressed as:

$$\mathcal{L}_1(\theta) = \frac{1}{N} \sum_{n=1}^{N} \|H_{hr}^n - H_{sr}^n\|_1,$$

$$\mathcal{L}_{sam}(\theta) = \frac{1}{N} \sum_{n=1}^{N} \frac{1}{\pi} \arccos\left(\frac{H_{hr}^n \cdot H_{sr}^n}{\|H_{hr}^n\|_2 \cdot \|H_{sr}^n\|_2}\right), \quad (10)$$

$$\mathcal{L}_{gra}(\theta) = \frac{1}{N} \sum_{n=1}^{N} \|M(H_{hr}^n) - M(H_{sr}^n)\|_1,$$

where N is the batch size, H_{hr}^n and H_{sr}^n represent the n-th HR and SR hyperspectral images, and $M(\cdot)$ denotes gradients in the horizontal, vertical, and spectral directions.

4 Experiments

4.1 Experimental Settings

Datesets

We conducted experiments on three hyperspectral image datasets: Chikusei [Yokoya and Iwasaki, 2016], Houston2018, and Pavia Center [Huang and Zhang, 2009]. The experiments on Pavia Center are presented in the supplementary materials. The Chikusei dataset, captured with the Headwall Hyperspec-VNIR-C sensor, includes 128 spectral bands over agricultural and urban areas in Japan, with a spatial resolution of 2517×2335 pixels. The Houston 2018 dataset is a part of the 2018 IEEE GRSS Data Fusion Contest, acquired by the ITRES CASI 1500 imager, covers the University of Houston and surrounding urban areas with 48 spectral bands and a resolution of 4172×1202 pixels.
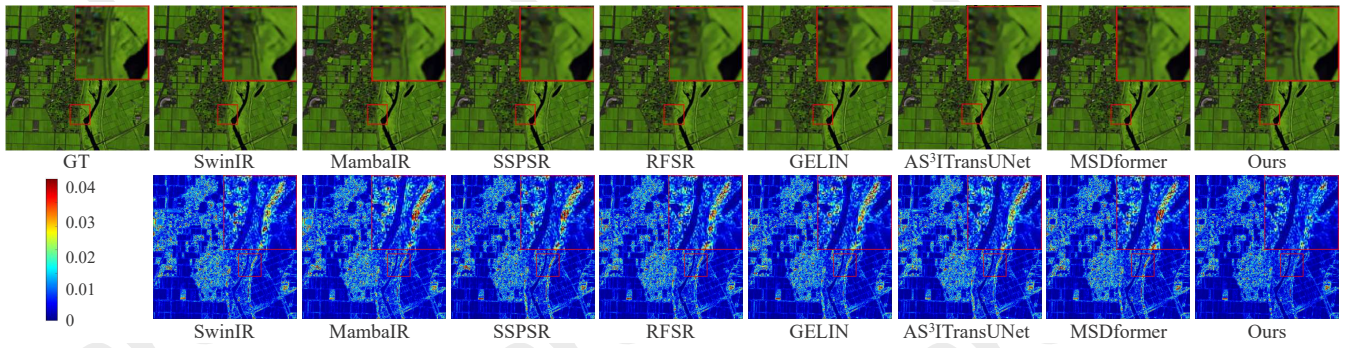
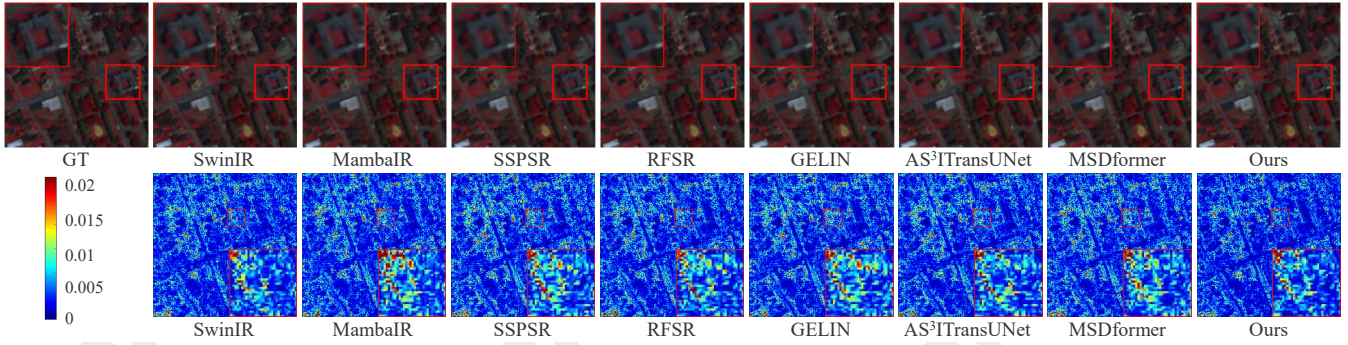Figure 5: Visual results on the Chikusei dataset with spectral bands 70-100-36 as R-G-B at scale factor ×4.

Figure 6: Visual results on the Houston dataset with spectral bands 26-20-10 as R-G-B at scale factor ×4.

Compared Methods and Metrics

We compare the proposed method with 8 deep learning approaches, including Transformer-based methods SwinIR [Liang *et al.*, 2021] for natural images, Mamba-based methods MambaIR [Guo *et al.*, 2024] for natural images, and hyperspectral image group-based methods such as GDRRN [Li *et al.*, 2018], SSPSR [Jiang *et al.*, 2020], RFSR [Wang *et al.*, 2022b], and GELIN [Wang *et al.*, 2022a]. We also include Transformer-based SHSR methods like AS³ITransUNet and MSDformer [Chen *et al.*, 2023b]. The performance of these methods is evaluated using six commonly used metrics in both spatial and spectral dimensions, including peak signal-to-noise ratio (PSNR), structure similarity (SSIM), spectral angle mapper (SAM), cross correlation (CC), root-mean-squared error (RMSE), and erreur relative global adimension-nellede synthese (ERGAS).

Implementation Details

The kernel size of the convolution is set to 3×3. We set the number of channels C to 64, the number of CSMG to 4, and the number of CSSM to 2. The initial learning rate is $1e^{-4}$, halving every 100 epochs until reaching 400 epochs. Following [Zhang *et al.*, 2018], the reduction ratio in channel attention (CA) is set to 16. During training, the Adam optimizer with Xavier initialization is used with a mini-batch size of 8. For image reconstruction, we use a progressive upsampling strategy via PixelShuffle [Shi *et al.*, 2016] to reduce parameters. The model is implemented in Pytorch and trained on NVIDIA RTX 4090 GPUs.

4.2 Comparison Results

Experiments on the Chikusei Dataset

For Chikusei dataset, 4 non-overlapping images with the size of $512 \times 512 \times 128$ are cropped from the top region. The remaining area is cropped into overlapping HR images for training (10% randomly selected for validation). The spatial size of the LR for training is 32×32, while the corresponding HR sizes at scale factors ×4, and ×8 are 128×128, and 256×256, respectively. All LR patches are generated by Bicubic downsampling at different scales.

Table 1 presents the quantitative results of our method and the compared approaches at different scale factors on the Chikusei dataset. The best results are highlighted in bold, while the second-best results are underlined. Our method surpasses SSPSR by 0.29 dB in PSNR and 0.32 in SAM at ×4, as group-based methods like SSPSR fail to leverage global spatial and spectral information effectively. Methods for natural images like SwinIR and MambaIR fail to fully utilize the rich spectral information and spatial-spectral correlations in hyperspectral images, resulting in poor spectral performance. Transformer-based SHSR methods like MSD-former and AS³ITransUNet achieve better SR performance than the above methods by modeling long-range dependencies. Notably, HSRMamba outperforms other methods for all metrics at scale factors ×4, and ×8, demonstrating the superiority and effectiveness of our method.

As shown in Fig. 5, group-based methods like GDRRN, often introduce blur details due to the limited spatial-spectral modeling. Transformer-based methods, such as MSDformer,
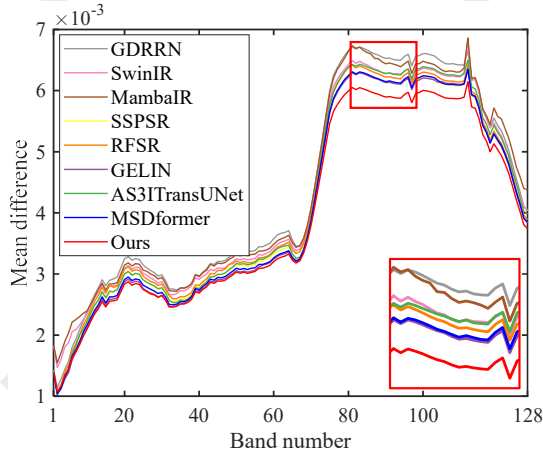
Figure 7: Mean spectral difference curves of different methods on the Chikusei dataset at scale factors ×4.

LSSP	GSRM	PSNR↑	SSIM↑	SAM↓
✗	✗	40.0714	0.9420	2.4078
✓	✗	40.1532	0.9427	2.3812
✗	✓	40.1849	0.9430	2.3662
✓	✓	**40.2781**	**0.9441**	**2.3160**

Table 2: Quantitative performance of different components evaluated on Chikusei dataset at the scale factor ×4.

Number (N)	Params.($\times 10^6$)	PSNR↑	SSIM↑	SAM↓
$N = 2$	1.0807	40.1797	0.9428	2.3479
$N = 4$	1.6073	**40.2781**	**0.9441**	**2.3160**
$N = 6$	2.1339	40.2312	0.9430	2.3523

Table 3: Quantitative comparisons of the number of Mamba groups over the Chikusei testing dataset at scale factor ×4.

consider global spatial and spectral information simultaneously, resulting in clearer boundaries and fewer artifacts compared to methods designed for natural images, such as MambaIR. HSRMamba reconstructs HR hyperspectral images with clearer and sharper details, which demonstrates that our HSRMamba can efficiently model long-range spatial-spectral dependencies. Additionally, mean error maps across all spectra in Fig. 5 show the reconstruction accuracy for individual pixels, with bluer regions indicating higher accuracy. And the mean spectral difference curves at ×4 in Fig. 7 evaluate the SR results from a spectral perspective. From the above visual results, it is evident that our method achieves results closer to the ground truth in both spatial and spectral dimensions compared to other methods.

Experiments on the Houston 2018 Dataset
Similar to Chikusei dataset, 8 images from the Houston2018 dataset with the size of $256 \times 256 \times 48$ are cropped from the top region for testing. The spatial resolution of LR and HR training patches is consistent with the Chikusei dataset.

The quantitative results of all methods on Houston dataset are shown in Table 1. HSRMamba outperforms the compared methods for all metrics at different scale factors. The visual results and mean error maps of all algorithms are presented in Fig. 6. We can also observe that HSRMamba provides more accurate results compared with other approaches.

4.3 Ablation Study
In this paper, we conduct the ablation experiments at factor scale ×4 on the Chikusei dataset. Additional ablation studies are listed in the supplementary materials.

Effects of LSSP and GSRM
LSSP effectively addresses the issue of local forgetting and GSRM relieves the sensitivity of the Mamba network to input order. As shown in Table 2, when LSSP is removed, PSNR of the network decreases by 0.09 dB, demonstrating the effectiveness of LSSP. The PSNR of model without GSRM decreases by 0.12 dB compared to the original method, demonstrating the effectiveness of GSRM. Finally, when both LSSP

and GSRM are removed, the performance drops significantly, further confirming the effectiveness of our method.

Effects of the Number of Groups
HSRMamba consists of multiple consecutive Mamba groups, and Table 3 shows the impact of the number of Mamba groups N. When $N = 2$, the performance is the weakest. As N increases to 4, the quantitative metrics improve. However, setting $N = 6$ leads to a decline in performance. This is mainly due to the increased network depth, causing the model to overfit. Therefore, considering both the experimental results and the model parameters, we set $N = 4$ in the paper.

Parameter and Complexity Analysis
To evaluate the computational efficiency of the proposed HSRMamba, we compare the model parameters, FLOPs, and PSNR results for different methods. As shown in Fig. 1, our method achieves better SR results with lower computational complexity and fewer parameters compared to other methods, demonstrating the effectiveness and efficiency of our approach. This indicates that our method strikes an excellent balance between model complexity and performance.

5 Conclusion
This paper presents HSRMamba, a contextual spatial-spectral relationship modeling algorithm designed for efficient HSISR. To address the issue of local pixel forgetting in hyperspectral images, we propose the LSSP to establish patch-wise long-range spatial-spectral correlations. Additionally, to overcome the challenge of insufficient causal modeling between highly similar pixels, we leverage the GSRM that rearranges the spectral dimension based on global spectral correlations. Using these algorithms, we construct the CSSM to efficiently capture long-range spatial-spectral dependencies in hyperspectral images. The CSSM module is composed of LSSM and GSCM, which enhance causal modeling by considering both local and global perspectives. Finally, extensive comparative experiments on different datasets and ablation studies validate the superiority and effectiveness of our proposed approach.

Acknowledgements

References

[Cai *et al.*, 2022] Yuanhao Cai, Jing Lin, Xiaowan Hu, Haoqian Wang, Xin Yuan, Yulun Zhang, Radu Timofte, and Luc Van Gool. Mask-guided spectral-wise transformer for efficient hyperspectral image reconstruction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 17481–17490, 2022.

[Chen *et al.*, 2023a] Chi Chen, Yongcheng Wang, Ning Zhang, Yuxi Zhang, and Zhikang Zhao. A review of hyperspectral image super-resolution based on deep learning. *Remote Sensing*, 15(11):2853, 2023.

[Chen *et al.*, 2023b] Shi Chen, Lefei Zhang, and Liangpei Zhang. Msdformer: Multiscale deformable transformer for hyperspectral image super-resolution. *IEEE Transactions on Geoscience and Remote Sensing*, 61:1–14, 2023.

[Chen *et al.*, 2024] Shi Chen, Lefei Zhang, and Liangpei Zhang. Cross-scope spatial-spectral information aggregation for hyperspectral image super-resolution. *IEEE Transactions on Image Processing*, 33:5878–5891, 2024.

[Dai *et al.*, 2024] Tao Dai, Jianping Wang, Hang Guo, Jinmin Li, Jinbao Wang, and Zexuan Zhu. Freqformer: Frequency-aware transformer for lightweight image super-resolution. In *Proceedings of the International Joint Conference on Artificial Intelligence*, pages 731–739, 2024.

[Deng *et al.*, 2023] Shangqi Deng, Liang-Jian Deng, Xiao Wu, Ran Ran, and Rui Wen. Bidirectional dilation transformer for multispectral and hyperspectral image fusion. In *Proceedings of the International Joint Conference on Artificial Intelligence*, pages 3633–3641, 2023.

[Dian and Li, 2019] Renwei Dian and Shutao Li. Hyperspectral image super-resolution via subspace-based low tensor multi-rank regularization. *IEEE Transactions on Image Processing*, 28(10):5135–5146, 2019.

[Dian *et al.*, 2017] Renwei Dian, Leyuan Fang, and Shutao Li. Hyperspectral image super-resolution via non-local sparse tensor factorization. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3862–3871, 2017.

[Dong *et al.*, 2023] Wenqian Dong, Jiahui Qu, Song Xiao, Tongzhen Zhang, Yunsong Li, and Xiuping Jia. Noise prior knowledge informed bayesian inference network for hyperspectral super-resolution. *IEEE Transactions on Image Processing*, 32:3121–3135, 2023.

[Fu *et al.*, 2021] Ying Fu, Zhiyuan Liang, and Shaodi You. Bidirectional 3d quasi-recurrent neural network for hyperspectral image super-resolution. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 14:2674–2688, 2021.

[Fu *et al.*, 2024] Guanyiman Fu, Fengchao Xiong, Jianfeng Lu, and Jun Zhou. Ssumamba: Spatial-spectral selective state space model for hyperspectral image denoising. *IEEE Transactions on Geoscience and Remote Sensing*, 2024.

[Gu and Dao, 2023] Albert Gu and Tri Dao. Mamba: Linear-time sequence modeling with selective state spaces. *arXiv preprint arXiv: 2312.00752*, 2023.

[Gu *et al.*, 2022] Albert Gu, Karan Goel, and Christopher Ré. Efficiently modeling long sequences with structured state spaces. In *Proceedings of the International Conference on Learning Representations*, 2022.

[Guo *et al.*, 2023] Wen-jin Guo, Weiying Xie, Kai Jiang, Yunsong Li, Jie Lei, and Leyuan Fang. Toward stable, interpretable, and lightweight hyperspectral super-resolution. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 22272–22281, 2023.

[Guo *et al.*, 2024] 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 *Proceedings of the European Conference on Computer Vision*, volume 15076, pages 222–241, 2024.

[He *et al.*, 2023] Jiang He, Qiangqiang Yuan, Jie Li, Yi Xiao, and Liangpei Zhang. A self-supervised remote sensing image fusion framework with dual-stage self-learning and spectral super-resolution injection. *ISPRS Journal of Photogrammetry and Remote Sensing*, 204:131–144, 2023.

[Hu *et al.*, 2024] 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*, 33:5260–5272, 2024.

[Huang and Zhang, 2009] 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.

[Huang *et al.*, 2024] Tao Huang, Xiaohuan Pei, Shan You, Fei Wang, Chen Qian, and Chang Xu. Localmamba: Visual state space model with windowed selective scan. *arXiv preprint arXiv:2403.09338*, 2024.

[Jiang *et al.*, 2020] Junjun Jiang, He Sun, Xianming Liu, and Jiayi Ma. Learning spatial-spectral prior for super-resolution of hyperspectral imagery. *IEEE Trans. Computational Imaging*, 6:1082–1096, 2020.

[Li *et al.*, 2018] Yong Li, Lei Zhang, Chen Ding, Wei Wei, and Yanning Zhang. Single hyperspectral image super-resolution with grouped deep recursive residual network. In *Proceedings of the International Conference on Multimedia Big Data*, pages 1–4, 2018.

[Li *et al.*, 2020] Qiang Li, Qi Wang, and Xuelong Li. Mixed 2d/3d convolutional network for hyperspectral image super-resolution. *Remote Sensing*, 12(10):1660, 2020.

[Li *et al.*, 2021] Qiang Li, Qi Wang, and Xuelong Li. Exploring the relationship between 2d/3d convolution for hyperspectral image super-resolution. *IEEE Transactions*

on Geoscience and Remote Sensing, 59(10):8693–8703, 2021.

[Li et al., 2022] Qiang Li, Yuan Yuan, Xiuping Jia, and Qi Wang. Dual-stage approach toward hyperspectral image super-resolution. IEEE Transactions on Image Processing, 31:7252–7263, 2022.

[Li et al., 2023] Guanxin Li, Jingang Shi, Yuan Zong, Fei Wang, Tian Wang, and Yihong Gong. Learning attention from attention: Efficient self-refinement transformer for face super-resolution. In Proceedings of the International Joint Conference on Artificial Intelligence, pages 1035–1043, 2023.

[Liang et al., 2021] 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 Workshops, pages 1833–1844, 2021.

[Liu et al., 2022] Yaoting Liu, Jianwen Hu, Xudong Kang, Jing Luo, and Shaosheng Fan. Interactformer: Interactive transformer and CNN for hyperspectral image super-resolution. IEEE Transactions on Geoscience and Remote Sensin, 60:1–15, 2022.

[Lu and Fei, 2014] Guolan Lu and Baowei Fei. Medical hyperspectral imaging: a review. Journal of biomedical optics, 19(1):010901–010901, 2014.

[Lu et al., 2020] 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.

[Mei et al., 2017] Shaohui Mei, Xin Yuan, Jingyu Ji, Yifan Zhang, Shuai Wan, and Qian Du. Hyperspectral image spatial super-resolution via 3d full convolutional neural network. Remote Sensing, 9(11):1139, 2017.

[Qiao et al., 2024] Junbo Qiao, Jincheng Liao, Wei Li, Yulun Zhang, Yong Guo, Yi Wen, Zhangxizi Qiu, Jiao Xie, Jie Hu, and Shaohui Lin. Hi-mamba: Hierarchical mamba for efficient image super-resolution. arXiv preprint arXiv:2410.10140, 2024.

[Shi et al., 2016] Wenzhe Shi, Jose Caballero, Ferenc Huszar, Johannes Totz, Andrew P. Aitken, Rob Bishop, Daniel Rueckert, and Zehan Wang. Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 1874–1883, 2016.

[Vivone et al., 2022] Gemine Vivone, Andrea Garzelli, Yang Xu, Wenzhi Liao, and Jocelyn Chanussot. Panchromatic and hyperspectral image fusion: Outcome of the 2022 whispers hyperspectral pansharpening challenge. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 16:166–179, 2022.

[Wang et al., 2022a] Xinya Wang, Qian Hu, Junjun Jiang, and Jiayi Ma. A group-based embedding learning and integration network for hyperspectral image super-resolution. IEEE Transactions on Geoscience and Remote Sensin, 60:1–16, 2022.

[Wang et al., 2022b] Xinya Wang, Jiayi Ma, and Junjun Jiang. Hyperspectral image super-resolution via recurrent feedback embedding and spatial-spectral consistency regularization. IEEE Transactions on Geoscience and Remote Sensing, 60:1–13, 2022.

[Wang et al., 2023] Xinya Wang, Qian Hu, Yingsong Cheng, and Jiayi Ma. Hyperspectral image super-resolution meets deep learning: A survey and perspective. IEEE/CAA Journal of Automatica Sinica, 10(8):1668–1691, 2023.

[Wu et al., 2023] Yiming Wu, Ronghui Cao, Yikun Hu, Jin Wang, and Kenli Li. Combining global receptive field and spatial spectral information for single-image hyperspectral super-resolution. Neurocomputing, 542:126277, 2023.

[Xiao and Wei, 2023] Jiahua Xiao and Xing Wei. Hyperspectral image denoising using uncertainty-aware adjustor. In Proceedings of the International Joint Conference on Artificial Intelligence, pages 1560–1568, 2023.

[Xiao et al., 2024] Yi Xiao, Qiangqiang Yuan, Kui Jiang, Yuzeng Chen, Qiang Zhang, and Chia-Wen Lin. Frequency-assisted mamba for remote sensing image super-resolution. IEEE Transactions on Multimedia, 27:1783–1796, 2024.

[Xu et al., 2023] Qin Xu, Shiji Liu, Jiahui Wang, Bo Jiang, and Jin Tang. As^{3}itransunet: Spatial-spectral interactive transformer u-net with alternating sampling for hyperspectral image super-resolution. IEEE Transactions on Geoscience and Remote Sensing, 61:1–13, 2023.

[Yokoya and Iwasaki, 2016] Naoto Yokoya and Akira Iwasaki. Airborne hyperspectral data over chikusei. Space Appl. Lab., Univ. Tokyo, Tokyo, Japan, Tech. Rep. SAL-2016-05-27, 2016.

[Zhang et al., 2018] Yulun Zhang, Kunpeng Li, Kai Li, Lichen Wang, Bineng Zhong, and Yun Fu. Image super-resolution using very deep residual channel attention networks. In Proceedings of the European Conference on Computer Vision, volume 11211, pages 294–310, 2018.

[Zhang et al., 2023] 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, 2023.

[Zhang et al., 2025] Xu Zhang, Jiaqi Ma, Guoli Wang, Qian Zhang, Huan Zhang, and Lefei Zhang. Perceive-ir: Learning to perceive degradation better for all-in-one image restoration. IEEE Transactions on Image Processing, 2025.

[Zhi et al., 2024] Ruicong Zhi, Xiaopei Fan, and Jingye Shi. Mambaformersr: A lightweight model for remote-sensing image super-resolution. IEEE Geoscience and Remote Sensing Letters, 21:1–5, 2024.

[Zhu et al., 2024] Lianghui Zhu, Bencheng Liao, Qian Zhang, Xinlong Wang, Wenyu Liu, and Xinggang Wang. Vision mamba: Efficient visual representation learning with bidirectional state space model. In Proceedings of the International Conference on Machine Learning, 2024.