Spectral Enhanced Rectangle Transformer for Hyperspectral Image Denoising

Miaoyu Li [1†], Ji Liu [2†], Ying Fu [1*], Yulun Zhang[3], Dejing Dou[4]
[1]Beijing Institute of Technology, [2]Baidu Inc., [3]ETH Zürich, [4]BCG X
miaoyli@bit.edu.cn, liuji04@baidu.com, fuying@bit.edu.cn,
yulun100@gmail.com, dejingdou@gmail.com

Abstract

Denoising is a crucial step for hyperspectral image (HSI) applications. Though witnessing the great power of deep learning, existing HSI denoising methods suffer from limitations in capturing the non-local self-similarity. Transformers have shown potential in capturing long-range dependencies, but few attempts have been made with specifically designed Transformer to model the spatial and spectral correlation in HSIs. In this paper, we address these issues by proposing a spectral enhanced rectangle Transformer, driving it to explore the non-local spatial similarity and global spectral low-rank property of HSIs. For the former, we exploit the rectangle self-attention horizontally and vertically to capture the non-local similarity in the spatial domain. For the latter, we design a spectral enhancement module that is capable of extracting global underlying low-rank property of spatial-spectral cubes to suppress noise, while enabling the interactions among non-overlapping spatial rectangles. Extensive experiments have been conducted on both synthetic noisy HSIs and real noisy HSIs, showing the effectiveness of our proposed method in terms of both objective metric and subjective visual quality. The code is available at https://github.com/MyuLi/SERT.

1. Introduction

With sufficient spectral information, hyperspectral images (HSIs) can provide more detailed characteristics to distinguish from different materials compared to RGB images. Thus, HSIs have been widely applied to face recognition [37, 38], vegetation detection [4], medical diagnosis [43], *etc*. With scanning designs [2] and massive wavebands, the photon numbers in individual bands are limited. HSI is easily degraded by various noise. Apart from poor visual effects, such undesired degradation also negatively affects the downstream applications. To obtain better visual effects and performance in HSI vision tasks, denoising is a fundamental step for HSI analysis and processing.

Similar to RGB images, HSIs have self-similarity in the spatial domain, suggesting that similar pixels can be grouped and denoised together. Moreover, since hyperspectral imaging systems are able to acquire images at a nominal spectral resolution, HSIs have inner correlations in the spectral domain. Thus, it is important to consider both spatial and spectral domains when designing denoising methods for HSI. Traditional model-based HSI denoising methods [10, 17, 21] employ handcrafted priors to explore the spatial and spectral correlations by iteratively solving the optimization problem. Among these works, total variation [20, 21, 52] prior, non-local similarity [19], low-rank [8, 9] property, and sparsity [42] regularization are frequently utilized. The performance of these methods relies on the accuracy of handcrafted priors. In practical HSI denoising, model-based methods are generally time-consuming and have limited generalization ability in diverse scenarios.

To obtain robust learning for noise removal, deep learning methods [7, 35, 41, 49] are applied to HSI denoising and achieve impressive restoration performance. However, most of these works utilize convolutional neural networks for feature extraction and depend on local filter response to separate noise and signal in a limited receptive field.

Recently, vision Transformers have emerged with competitive results in both high-level tasks [16, 39] and low-level tasks [1, 13, 50], showing the strong capability of modeling long-range dependencies in image regions. To diminish the unaffordable quadratically computation cost to image size, many works have investigated the efficient design of spatial attention [11, 46, 47]. Swin Transformer [28] splitted feature maps into shifted square windows. CSWin Transformer [15] developed a stripe window across the features maps to enlarge the attention area. As HSI usually has large feature maps, exploring the similarity beyond the noisy pixel can cause unnecessary calculation burden. Thus, how to efficiently model the non-local spatial similarity is still challenging for HSI denoising Transformer.

HSIs usually lie in a spectral low-rank subspace [9], which can maintain the distinguished information and sup-

[†]Equal Contribution, [*]Corresponding Author

press noise. This indicates that the non-local spatial similarity and low-rank spectral statistics should be jointly unitized for HSI denoising. However, existing HSI denoising methods [24, 45] mainly utilize the low-rank characteristics through matrix factorization, which is based on a single HSI and requires a long-time to solve. The global low-rank property in large datasets is hardly considered.

In this paper, we propose a **S**pectral **E**nhanced **R**ectangle **T**ransformerc (SERT) for HSI denoising. To reinforce model capacity with reasonable cost, we develop a multi-shape rectangle self-attention module to comprehensively explore the non-local spatial similarity. Besides, we aggregate the most informative spectral statistics to suppress noise in our spectral enhancement module, which projects the spatial-spectral cubes into low-rank vectors with the assistance of a global spectral memory unit. The spectral enhancement module also provides interactions between the non-overlapping spatial rectangles. With our proposed Transformer, the spatial non-local similarity and global spectral low-rank properly are jointly considered to benefit the denoising process. Experimental results show that our method significantly outperforms the state-of-the-art methods in both simulated data and real noisy HSIs.

Overall, our contributions can be summarized as follows:

- We propose a spectral enhanced rectangle Transformer for HSI denoising, which can well exploit both the non-local spatial similarity and global spectral low-rank property of noisy images.
- We present a multi-shape rectangle spatial self-attention module to effectively explore the comprehensive spatial self-similarity in HSI.
- A spectral enhancement module with memory blocks is employed to extract the informative low-rank vectors from HSI cube patches and suppress the noise.

2. Related Works

2.1. Hyperspectral Image Denoising

HSI denoising is a well-developed research area in computer vision [9, 19, 44] and remote sensing [34, 49]. Mainstream HSI denoising methods can be classified into model-based methods and deep learning methods.

Traditional model-based methods [10, 29, 29, 48, 54] illustrate noise removal as an iterative optimization problem with handcrafted priors. Adaptive spatial-spectral dictionary methods are proposed in [17]. Chang *et al*. [9] employed the hyper-Laplacian regularized unidirectional low-rank tensor recovery method to utilize the structure correlation in HSI. The spatial non-local similarity and global spectral low-rank property are integrated in [19] for denoising. Besides, other conventional spatial regularizers [29,52] and low-rank regularization [8] are also introduced to model the spatial and spectral properties of noisy HSI.

With great potential to automatically learn and represent features, deep learning methods [7,32,41,45] have been actively investigated for HSI denoising. Spectral-spatial features are exploited via residual convolutional network in HSID-CNN [49]. A deep spatial-spectral global reasoning network is proposed in [7] to consider both the local and global information for HSI denoising. Besides, a quasi-recurrent neural network was extended to HSI denoising task [32,41], showing the benefits of both convolutional and recurrent neural networks. Model-guided interpretable networks have also been actively explored in [3, 44]. Different from those convolution-based networks that have limited receptive field and fixed feature extraction paradigms, our proposed method utilizes a transformer to better model the inner similarity in spatial and spectral domains.

2.2. Vision Transformer

Transformer for RGB images. Transformers have been actively applied to vision tasks [16,18,39,47] due to its powerful ability in modeling long-range dependencies. Self-attention mechanism has been proven to be efficacious in previous works [23,40]. When applied to the spatial region, it is crucial for the Transformers to consider the trade-off between computation cost and model capacity. To cut down the quadratic computation growth to image size, Dosovitskiy *et al*. [16] first employed Transformer for image recognition with images spitted in small patches. Swin Transformer [28] was proposed with shifted window for self-attention in the spatial domain. To further enlarge the receptive field of self-attention, down-sampled attention was introduced in [13,39,47]. Without spatial information loss, Dong *et al*. [15] employed horizontal and vertical stripes to compute self-attention. However, for HSI denoising, the non-local spatial similarity is not efficiently explored as these Transformers conducted the spatial self-attention in limited windows or introduced unnecessary computation cost. Besides, the combined consideration of the spatial and spectral domains are rarely investigated.

Transformer for HSI. Recently, there is an emerging trend of using Transformer to HSI restoration [1, 36, 51] and HSI classification [22, 27]. An architecture search framework was proposed in [55] to find a suitable network consisting of spectral and spatial Transformer for HSI classification. A 3D quasi-recurrent and Transformer network was presented in [1] for hyperspectral image denoising, which combined the 3D quasi-recurrent layer with Swin blocks. Different from these works that tend to directly employ existing transformer blocks to another tasks, methods in [5, 6] solve the HSI reconstruction problem with task-oriented transformer block under the guidance of degradation mask. However, these works do not consider the similarity in both spatial and spectral domains. Here, we introduce our spectral enhanced rectangle Transformer to HSI denoising, exploring

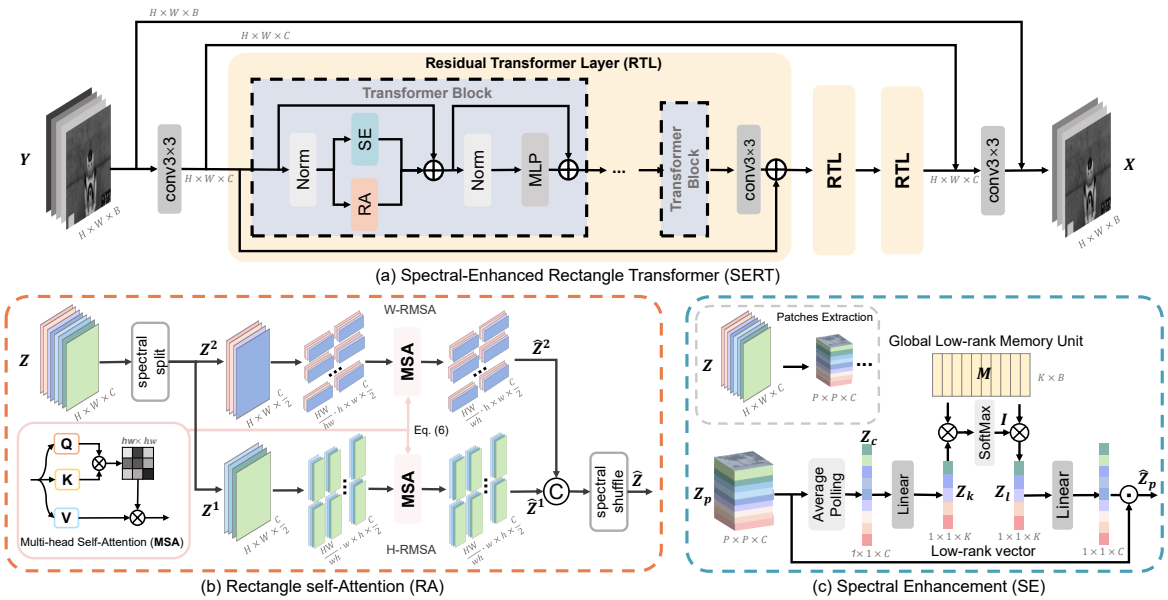


Figure 1. Overall framework of SERT. (a) SERT mainly includes two essential components, *i.e.*, SE for non-local spatial similarity and SE for global low-rank property. (b) spatial rectangle self-attention (RA) and (c) spectral enhancement (SE) module.

the most important two characteristics of HSI, including spatial non-local similarity and global low-rank properties.

3. Spectral Enhanced Rectangle Transformer

Assuming the degraded noisy HSI as $\boldsymbol{Y} \in \mathbb{R}^{H\times W\times B}$, where H, W, and B represent the height, width, and band of the HSI, the noise degradation can be formulated as

$$\boldsymbol{Y} = \boldsymbol{X} + \boldsymbol{n}, \tag{1}$$

where $\boldsymbol{X}\in\mathbb{R}^{H\times W\times B}$ is the desired clean HSI, and $\boldsymbol{n}\in\mathbb{R}^{H\times W\times B}$ denotes the addictive random noise. In realistic HSI degradation situations, HSIs are corrupted by various types of noise, *e.g.*, Gaussian noise, stripe noise, deadline noise, impulse noise, or a mixture of them.

In this section, we elaborately introduce our proposed spectral enhanced rectangle Transformer for HSI denoising. The overall architecture is shown in Figure 1. In our implementation, each Residual Transformer Layer (RTL) consists of 6 Transformer blocks. And the proposed Transformer Block mainly contains two essential components, *i.e.*, rectangle self-attention (RA) module and spectral enhancement (SE) module. Figure 1(b) and Figure 1(c) illustrate the detailed framework of RA module and SE module, respectively. The outputs of RA and SE are added together to achieve comprehensive feature embeddings for noise removal. Next, we discuss each module in detail.

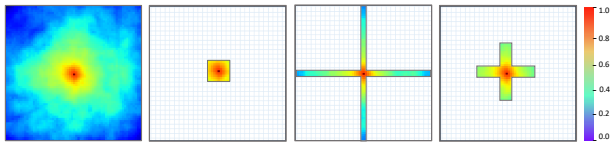
3.1. Spatial Rectangle Self-Attention

To remove noise from HSI, it is important to explore the similarity information in spatial domain [19], which implies that similar pixels can be aggregated together for denoising. Existing deep learning-based HSI denoising methods mainly utilize the convolutional layer to extract the local information with spatially invariant kernels, limiting the flexibility to model the non-local similarity.

For better model capacity, there are various attempts [28, 39, 50] that employ Transformer as an alternative solution to convolution neural network. The power of self-attention mechanism in modeling spatial information has also been proven in [13, 26]. Since the global self-attention in the spatial domain introduces high computational complexity, Swin Transformer [28] and CSWin Transformer [15] split the input feature into windows or stripes for attention operation. From the heatmap shown in Figure 2, we can observe that neighboring pixels are more similar to the center pixel than distant pixels. When conducting spatial self-attention, Swin (see Figure 2(b)) focuses on local information while CSwin (Figure 2 (c)) tends to utilize pixels which is less informative. Thus, how to effectively conduct the self-attention in the informative spatial regions to model non-local similarity is still challenging for HSI denoising.

Here, we propose a rectangle self-attention in the spatial domain, in which the feature maps are split into several non-overlapping rectangles. As shown in Figure 2, our rectangle Transformer focuses on the informative neighboring pixels and obtains more exhaustive information in non-local area. At different stages of the network, rectangles of different shapes are employed to explore better expression ability.

The details of our proposed RA module are shown in Figure 1(b). To obtain comprehensive features, the rectangle self-attention is conducted in vertically and horizontally



(a) similarity to central pixel (b) window self-attention (c) cross-shaped self-attention (d) rectangle self-attention (ours)

Figure 2. This similarity statistic is obtained via Realistic dataset [52]. As the distance becomes longer, the similarity decreases.

after the spectral split operation. Different from [15], we add a spectral shuffle [30] operation to exchange the information from two branches. Since rectangle self-attention in vertical and horizontal focuses on different regions and has different receptive fields, the shuffle operation also enlarges the respective field of the whole module.

Let $\mathrm{Z}\in\mathbb{R}^{H\times W\times C}$ denote the input features of RA module. The outputs of RA module is calculated via

$$\boldsymbol{Z}_1, \boldsymbol{Z}_2 = \mathrm{Split}(\boldsymbol{Z}), \tag{2}$$

$$\hat{\boldsymbol{Z}}^1 = \text{W-RMSA}(\boldsymbol{Z}^1), \hat{\boldsymbol{Z}}^2 = \text{H-RMSA}(\boldsymbol{Z}^2) \tag{3}$$

$$\hat{\boldsymbol{Z}} = \mathrm{Shuffle}([\hat{\boldsymbol{Z}}^1, \hat{\boldsymbol{Z}}^2]), \tag{4}$$

where W-RMSA denotes the horizontal rectangle multi-head self-attention, and H-RMSA denotes the vertical rectangle multi-head self-attention. $\boldsymbol{Z}$ is firstly divided into two parts in spectral domain, where $\boldsymbol{Z}^1\in\mathbb{R}^{H\times W\times \frac{C}{2}}$ and $\boldsymbol{Z}^2\in\mathbb{R}^{H\times W\times \frac{C}{2}}$. Then, $\boldsymbol{Z}^1$ and $\boldsymbol{Z}^2$ conduct the W-RMSA and H-RMSA separately.

Supposing the size of horizontal rectangle as $[h, w]$ and $h>w$, for W-RMSA, the input features $\boldsymbol{Z}^1$ is partitioned into non-overlapping rectangles as $\{\boldsymbol{Z}_1^1, \boldsymbol{Z}_2^1, ..., \boldsymbol{Z}_N^1\}$, in which $Z_i^1 \in \mathbb{R}^{h\times w\times \frac{C}{2}}$ and $N=\frac{W\times H}{h\times w}$. The output of each rectangle from W-RMSA is calculated as

$$\boldsymbol{Q}_i^1 = \boldsymbol{Z}_i^1\boldsymbol{W}_q^1, \quad \boldsymbol{K}_i^1 = \boldsymbol{Z}_i^1\boldsymbol{W}_k^1, \quad \boldsymbol{V}_i^1 = \boldsymbol{Z}_i^1\boldsymbol{W}_v^1 \tag{5}$$

$$\hat{\boldsymbol{Z}}_i^1 = \mathrm{SoftMax}(\boldsymbol{Q}_i^1{\boldsymbol{K}_i^1}^T/\sqrt{d} + \boldsymbol{P})\boldsymbol{V}_i^1, \tag{6}$$

where $\boldsymbol{W}_q^1, \boldsymbol{W}_k^1, \boldsymbol{W}_v^1\in\mathbb{R}^{\frac{C}{2}\times\frac{C}{2}}$ are the projection mappings of *query* $\boldsymbol{Q}_i^1\in\mathbb{R}^{h\times w\times\frac{C}{2}}$, *keys* $\boldsymbol{K}_i^1\in\mathbb{R}^{h\times w\times\frac{C}{2}}$, and *value* $\boldsymbol{V}_i^1\in\mathbb{R}^{h\times w\times\frac{C}{2}}$. $\boldsymbol{P}$ is the learnable parameter embedding the position and d is the feature dimension. Then the outputs of horizontal rectangle self-attention is aggregated by

$$\text{W-RMSA}(\boldsymbol{Z}^1) = \mathrm{Merge}(\hat{\boldsymbol{Z}}_1^1, \hat{\boldsymbol{Z}}_2^1, ..., \hat{\boldsymbol{Z}}_N^1). \tag{7}$$

For vertical rectangle self-attention H-RMSA, the size of the rectangle is $[w, h]$ while other operations are similar to W-RMSA. Moreover, at different layers of the network, rectangles in various shapes are employed to explore non-local similarity in different scales.

3.2. Spectral Enhancement

In traditional model-based HSI denoising methods, HSI is always represented by its extracted patches, and the low-rank property is widely explored in HSI denoising [9], compressive sensing [14], unmixing [24], implying that the low-dimensional spectral subspace is beneficial to HSI tasks. We also adopt the low-rank property to guide the HSI denoising process. However, without strong regularization like SVD decomposition [8], projecting the noisy HSI into a proper subspace is difficult. Thus, instead of introducing orthogonal linear projection as in [12] to HSI, we use the memory unit (MU) to store the low-rank statistics of HSI cubes. The network itself automatically learns how to represent the HSI cubes in subspace. The MU module can be denoted as a dictionary of global low-rank spectral vectors.

As shown in Figure 1(c), the features are firstly partitioned into several cube patches of size $P\times P\times C$ to explore the spectral-spatial correlation. In the implementation, P is set to the long side of the rectangle in RA module. Accordingly, the spectral enhancement block also provides information interactions between the inside rectangles. Moreover, shift operation [28] is employed in spatial domain to establish connections between adjacent cube patches.

The input of SE module is denoted as $\boldsymbol{Z}_p \in \mathbb{R}^{P\times P\times C}$. To obtain distinguished spectral information in a subspace, following [23] and [11], a squeeze operation is employed and aggregates the features across the cube patch $\boldsymbol{Z}_p$ to produce a projected spectral vector of size $1\times1\times K$. Specifically, a downsample operation is firstly conducted in the spatial domain to obtain aggregated spectral vector $\boldsymbol{Z}_c\in\mathbb{R}^{1\times1\times C}$. Then, it is projected to obtain $\boldsymbol{Z}_k \in \mathbb{R}^{1\times1\times K}$, which is in a subspace of rank K. The extraction is described as

$$\boldsymbol{Z}_c = \mathrm{AveragePool}(\boldsymbol{Z_p}), \tag{8}$$

$$\boldsymbol{Z}_k = \boldsymbol{Z}_c\boldsymbol{W}_k, \tag{9}$$

where $\boldsymbol{W}_k \in \mathbb{R}^{C\times K}$ is the projection mapping. Notably, instead of conducting a global aggregation on the whole image, we focus on the information inside the cube since neighboring pixels tend to share similar spectral statistics.

To explore the spatial-spectral correlation beyond the current HSI cube and enhance the expression ability of low-rank spectral vector, we introduce a memorizing unit (MU) to store the spectral information. The MU module maintains a global memory bank $\boldsymbol{M} \in \mathbb{R}^{K\times B}$, which is learned as parameters of the network. For spectral vector Z_k, we seek the most relevant spectral low-rank vectors in MU and use these vectors to assist in adjusting the projected vector $\boldsymbol{Z}_k$. The corresponding coefficients $\boldsymbol{I} \in \mathbb{R}^{1\times B}$ between $\boldsymbol{Z}_k$ and stored low-rank vectors $\boldsymbol{M}$ is extracted by

$$\boldsymbol{I} = \mathrm{Softmax}(\boldsymbol{Z}_k\boldsymbol{M}). \tag{10}$$

With coefficients matrix $\boldsymbol{I}$, the desired low-rank vector $\boldsymbol{Z}_l\in\mathbb{R}^{1\times1\times K}$ can be obtained from MU via

$$\boldsymbol{Z}_l = \boldsymbol{I}\boldsymbol{M}. \tag{11}$$

Method	10			30			50			70			10-70		
	PSNR	SSIM	SAM	PSNR	SSIM	SAM	PSNR	SSIM	SAM	PSNR	SSIM	SAM	PSNR	SSIM	SAM
Noisy	28.13	0.8792	18.72	18.59	0.5523	37.9	14.15	0.3476	49.01	11.23	0.2301	56.45	17.24	0.4782	41.94
BM4D [31]	40.78	0.9930	2.99	37.69	0.9872	5.02	34.96	0.9850	6.81	33.15	0.9554	8.40	36.62	0.9770	5.51
LLRT [9]	46.72	0.9983	1.60	41.12	0.9920	2.52	38.24	0.9830	3.47	36.23	0.9732	4.46	40.06	0.9860	3.24
NGMeet [19]	**47.90**	0.9988	1.39	42.44	0.9816	2.06	39.69	0.9658	2.49	38.05	0.9531	2.83	41.67	0.9937	2.19
HSID-CNN [49]	43.14	0.9918	2.12	40.30	0.9854	3.14	37.72	0.9746	4.27	34.95	0.9521	5.84	39.04	0.9776	3.71
GRNet [7]	45.25	0.9976	1.83	42.09	0.9957	2.18	40.25	0.9936	2.42	38.95	0.9914	2.63	41.44	0.9944	2.27
QRNN3D [41]	45.61	0.9977	1.80	42.18	0.9955	2.21	40.05	0.9929	2.63	38.09	0.9883	3.42	41.34	0.9938	2.42
T3SC [3]	45.81	0.9979	2.02	42.44	0.9957	2.44	40.39	0.9933	2.85	38.80	0.9904	3.26	41.64	0.9942	2.61
MAC-Net [45]	45.20	0.9974	1.87	42.10	0.9955	2.35	40.09	0.9931	2.79	38.64	0.9905	3.16	41.31	0.9941	2.52
SERT (Ours)	47.72	**0.9988**	**1.36**	**43.56**	**0.9969**	**1.77**	**41.33**	**0.9949**	**2.05**	**39.82**	**0.9929**	**2.30**	**42.82**	**0.9957**	**1.88**

Table 1. Averaged results of different methods under Gaussian noise levels on ICVL dataset. PSNR is in dB.

Method	Non-i.i.d Gaussian			Gaussian+Deadline			Gaussian+Impulse			Gaussian+Stripe			Gaussian+Mixture		
	PSNR	SSIM	SAM	PSNR	SSIM	SAM	PSNR	SSIM	SAM	PSNR	SSIM	SAM	PSNR	SSIM	SAM
Noisy	18.29	0.5116	46.20	17.50	0.4770	47.55	14.93	0.3758	46.98	17.51	0.4867	46.98	13.91	0.3396	51.53
BM4D [31]	36.18	0.9767	5.78	33.77	0.9615	6.85	29.79	0.8613	21.59	35.63	0.9730	6.26	28.01	0.8419	23.59
LLRT [9]	34.18	0.9618	4.88	32.98	0.9559	5.29	28.85	0.8819	18.17	34.27	0.9628	4.93	28.06	0.8697	19.37
NGMeet [19]	34.90	0.9745	5.37	33.41	0.9665	6.55	27.02	0.7884	31.20	34.88	0.9665	5.42	26.13	0.7796	31.89
HSID-CNN [49]	39.28	0.9819	3.80	38.33	0.9783	3.99	36.21	0.9663	5.48	38.09	0.9765	4.59	35.30	0.9588	6.29
GRNet [7]	35.19	0.9780	5.19	33.78	0.9744	5.42	32.78	0.9606	8.26	34.85	0.9772	5.41	30.91	0.9617	8.26
QRNN3D [41]	42.18	0.9950	2.84	41.69	0.9942	2.61	40.32	0.9914	4.31	41.68	0.9943	2.97	39.08	0.9892	4.80
T3SC [3]	41.95	0.9922	4.18	39.59	0.9924	4.86	37.85	0.9843	6.53	41.32	0.9937	3.27	35.53	0.9767	8.12
MAC-Net [45]	39.98	2.9662	4.55	36.68	0.9860	5.63	34.54	0.9553	10.20	39.03	0.9910	4.03	30.59	0.9300	14.51
SERT (Ours)	**44.20**	**0.9971**	**1.69**	**43.66**	**0.9969**	**1.99**	**42.67**	**0.9959**	**2.30**	**43.68**	**0.9969**	**1.97**	**40.00**	**0.9937**	**2.84**

Table 2. Averaged results of different methods under complex noise on ICVL dataset. PSNR is in dB.

Since $\boldsymbol{Z}_l$ represents the most informative spectral statistics of the noisy cube, to enhance the spatial-spectral correlation and suppress noise, we use the obtained low-rank vector as guidance to benefit the denoising process. The output of our spectral enhancement module is obtained by rescaling the input SHI cube $\boldsymbol{Z}_p$ with $\boldsymbol{Z}_l$ as

$$\hat{\boldsymbol{Z}}_p = \boldsymbol{Z}_p \cdot \boldsymbol{W}_c \boldsymbol{Z}_l, \tag{12}$$

where $\boldsymbol{W}_c \in \mathbb{R}^{C \times K}$ is the project mapping and $\cdot$ is the element-wise dot product.

4. Experiments

In this section, we first evaluate our method with synthetic experiments, including Gaussian noise cases and complex noise cases. Then we report results on real noisy datasets. Finally, we perform model analysis experiments to verify the effectiveness of the proposed model.

We compare several traditional model-based HSI denoising methods including the filter-based method (BM4D [31]), tensor-based method (LLRT [9]), and orthogonal basis-based method (NGMeet [19]). Five state-of-the-art deep learning-based methods, *i.e.*, HSID-CNN [48], GRNet [7], QRNN3D [41], T3SC [3], and MAC-Net [7] are also compared. Traditional methods are programmed in Matlab with Intel Core i9-10850K CPU. Our method as well as other deep networks is evaluated with an NVIDIA RTX 3090 GPU. Peak signal-to-noise ratio (PSNR), structural similarity index metric (SSIM) and spectral angle mapper (SAM) are used as the quantitative criteria.

4.1. Experiments on Synthetic Data

Datasets. Synthetic experiments are conducted on ICVL dataset, which has been widely used for simulated studies [3, 41]. ICVL contains 201 HSIs of size 1392×1300 with 31 bands from 400 nm to 700 nm. We use 100 HSIs for training, 5 HSIs for validating, and 50 HSIs used for testing. Following settings in [3] and [41], training images are cropped to size 64×64 at different scales. During the testing phase, HSIs are cropped to 512×512×31 to obtain an affordable computation cost for traditional methods.

Implementation Details. We use noise patterns in [41] to simulate the noisy HSIs. Specifically, the noise patterns are

- i.i.d Gaussian noise from level 10 to level 70.
- Complex noise cases. Five types of complex noise are included, *i.e.*, Non-i.i.d Gaussian noise, Gaussian + Stripe noise, Gaussian + Deadline noise, Gaussian + Impulse noise, and Mixture noise.

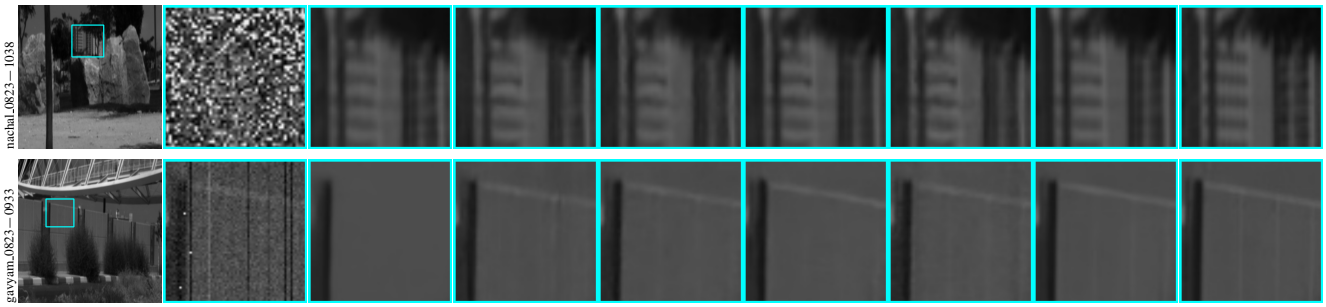


Figure 3. Visual comparison on ICVL. Images are from band 28. The top row exhibits the results under Gaussian noise with noise level 50 and the bottom row exhibits the results under deadline noise.

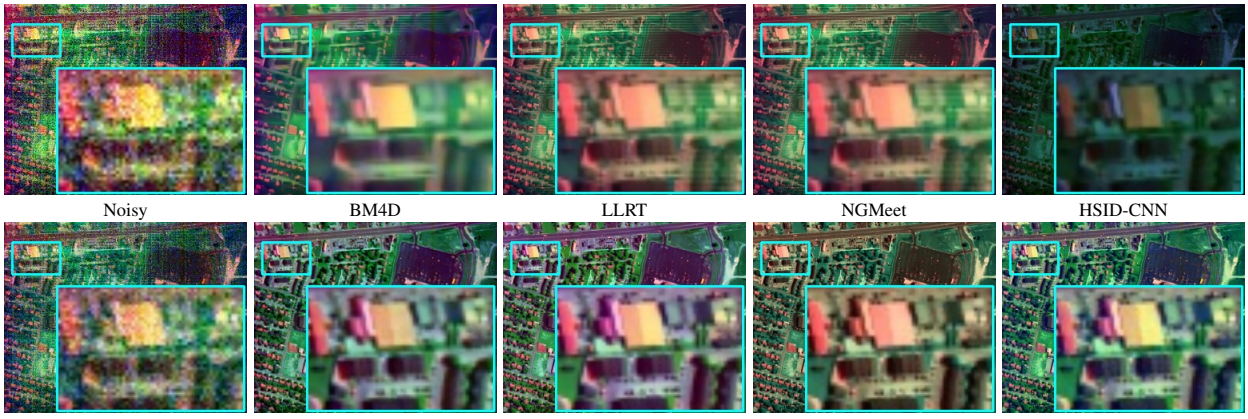


Figure 4. Visual quality comparison of real noisy HSI experiments on Urban dataset with bands 1, 108, 208.

For i.i.d Gaussian noise case, we train networks with random noise levels from 10 to 70 and test them under different levels of noise. For complex noise, networks are trained with a mixture of noise and tested under each case.

For our proposed model, the learning rate is set to $1e{-}4$ with Adam optimizer. After 50 epochs, the learning rate is divided by 10. The total epoch number is 80. we set the basic channel $C = 96$ and rank size $K = 12$. The size of the rectangle of each Transformer layer is set to $[16, 1]$, $[32, 2]$, and $[32, 4]$ respectively. For competing methods, we use the parameter settings in the referenced works and make a great effort to reproduce the best results.

Quantitative Comparison. We show the quantitative results of Gaussian noise experiments and complex noise experiments in Tables 1 and 2. Among these traditional methods, NGMeet performs well on Gaussian noise cases in Table 1 and surpasses the deep learning method HSID-CNN. However, results of NGMeet and other model-based methods under complex noise cases in Table 2 are much worse, showing the poor generalization ability of handcrafted priors. Our proposed method outperforms other deep learning methods by at least 0.9 dB for all noise cases. Notably, our method effectively recovers a more accurate image from the challenging complex noisy HSIs, demonstrating its impressive ability to handle various noise.

Visual Comparison. To further demonstrate the denoising performance of our method, we show the denoised results of different methods under random Gaussian noise and deadline noise in Figure 3. In the top row, QRNN3D and QRNN3D exhibit excessive smoothness for some more complex textures. Compared to NGMeet, our method has much fewer artifacts than other methods. In the bottom row, our method restores more texture details with less noise.

4.2. Experiments on Real Noisy Data

Datasets. Urban dataset and Realistic dataset from [53] are both adopted for our real data experiments.

Urban dataset contains a image of size 307×307 with 210 bands covering from 400 to 2500 nm. Since there is no clean HSI, we use APEX dataset [25] for pre-training, in which band-dependent noise levels from 0 to 55 are added to the clean HSIs. The settings are the same with [3].

For Realistic dataset [53], there are 59 noisy HSIs provided with paired clean HSIs. Each HSI contains 696×520 pixels in spatial resolution with 34 bands from 400 nm to 700 nm. We randomly select 44 HSIs from both indoor scenes and outdoor scenes. The left is used for testing.

Implementation Details. For Urban dataset experiment, networks are trained with their default parameter settings. The training epochs of our method is set to 100 epochs with

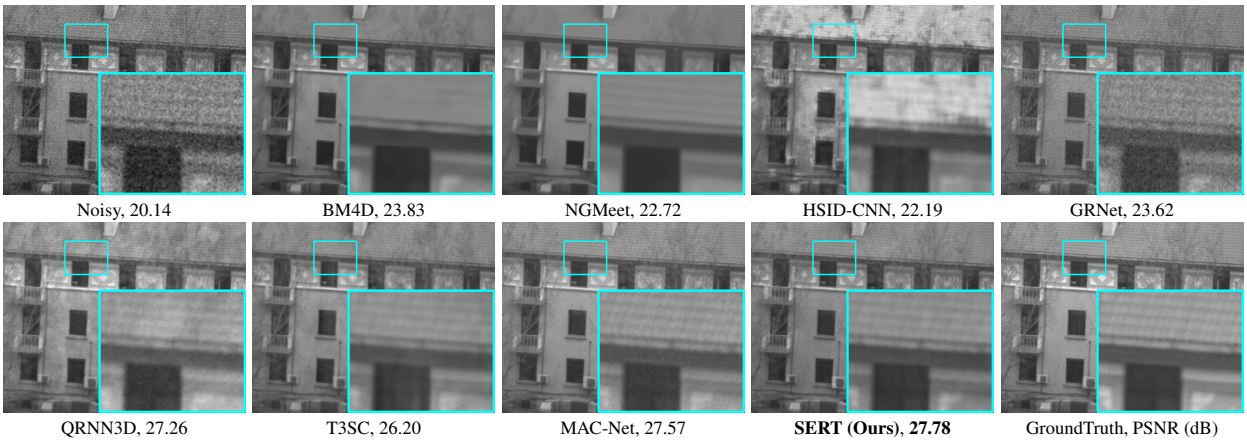


Figure 5. Visual comparison on Realistic dataset [52] of scene 5 with corresponding PSNR. The images are from band 12 on 550 nm.

Metric	Noisy	BM4D [31]	LLRT [9]	NGMeet [19]	HSID-CNN [48]	GRNet [7]	QRNN3D [41]	T3SC [3]	MAC-Net [7]	**SERT (Ours)**
PSNR	23.26	29.04	28.26	28.72	26.44	25.33	28.12	28.51	29.20	**29.68**
SSIM	0.7609	0.9471	0.9417	0.9511	0.8992	0.8381	0.9066	0.9323	0.9489	**0.9533**
SAM	17.329	3.087	3.960	2.735	5.242	9.737	5.590	4.408	4.099	**2.536**

Table 3. Average results of different methods on 15 real noisy HSIs. The PSNR is in dB, and best results are in bold.

Metric	SwinIR [26]	Restormer [50]	CSwin [15]	TRQ3D [33]	**SERT (Ours)**
GFLOPS	1473.0	3652.8	1129.5	2135.7	**1018.9**
Params (M)	2.98	90.94	58.53	**0.68**	1.91
PSNR (dB)	40.44	41.07	42.04	41.66	**42.82**
SSIM	0.9938	0.9945	0.9951	0.9947	**0.9957**
SAM	2.32	2.05	2.18	2.21	**1.88**

Table 4. Comparison with other Transformers under random Gaussian noise on ICVL dataset. SWinIR, Restormer, and CSWin are proposed for RBG image tasks. TRQ3D is for HSI denoising.

a learning rate $1e{-}4$. For the Realistic dataset [52], we crop overlapped $128{\times}128$ spatial regions with data augmentation to train deep networks. The data augmentation settings in [52] are also adopted. The training epoch is set to 1000.

Quantitative Comparison. Table 3 shows the averaged results of different methods on the Realistic dataset. Our proposed SERT significantly outperforms other HSI denoising methods by almost 0.5 dB, showing the effectiveness of our method in handling real noise.

Visual Comparison. We provide the denoising results of real noisy HSIs in Figures 4 and 5. Our method is superior to traditional denoising and deep learning methods in terms of both noise removal and detail retention. From Figure 4, we can observe that Urban image is corrupted by complex noise. The stripe noise has severely affected the visual effect of image. Denoised images obtained by other methods are either over-smoothed or still have obvious stripe noise. Our method provides a clean output image while preserving the textures and sharpness. For visual comparison of Realistic dataset in Figure 5, the competing methods generate incorrect texture and are less effective in noise removal. And our method achieves the most promising visual result.

4.3. Comparison with other Transformers

To show the effectiveness of our method in exploring spatial and spectral characteristics of HSIs, we evaluate our model with four Transformer methods in Table 4. Our model achieves the best results, implying that the proposed Transformer block is more suitable for HSI denoising.

Differences with existing RGB Transformers. Existing RGB Transformer methods consider the inner long-range dependency from the spatial dimension [13, 28] or spectral dimension [50]. Our Transformer explores the joint correlation. Besides, our Transformer block utilizes the non-local similarity and low-rank property, providing a better modeling capability to explore the rich information of HSI.

Differences with existing HSI Transformers. TRQ3D proposed a hybrid framework that employs both Swin Transformer and 3D quasi-recurrent network for HSI denoising [33]. With Transformer block adopted from RGB image tasks, the inner characteristic of HSI is hardly fully utilized in the proposed Transformer-based network.

4.4. Model Analysis

Model Complexity. In Table 5, we compare the average inference time, GFLOPs as well as denoising performance by different denoising methods on ICVL dataset and real noisy dataset [52]. Our method achieves the competing computation cost and inference time with better performance.

Component Analysis. The results of different component designs are given in Table 6(a). The first row presents Transformer with rectangle self-attention (RA) in spatial domain. Applying spectral enhancement (SE) to capture spatial-spectral information, it remarkably boosts the de-

Metric	Synthetic Noise (512×512×31)						Real Noise (512×512×34)					
	HSID-CNN	GRNet	T3SC	QRNN3D	MAC-Net	**SERT (Ours)**	HSID-CNN	GRNet	T3SC	QRNN3D	MAC-Net	**SERT (Ours)**
PSNR (dB)	39.04	41.44	41.34	41.64	41.31	42.82	26.44	25.33	28.13	28.51	29.20	29.68
Params (M)	0.40	44.39	0.83	0.83	0.43	1.91	0.40	44.40	0.83	0.83	0.43	1.91
GFLOPS	3249.7	610.7	-	2513.7	-	1018.9	3564.2	611.9	-	2756.9	-	1021.9
Time (s)	1.700	0.361	1.123	0.683	3.627	0.717	1.865	0.407	1.204	0.822	2.992	0.764

Table 5. Comparisons of PSNR, Params, FLOPS and inference time of different deep learning methods.

RA	SE	SS	MU	Params (M)	GFLOPS	PSNR (dB)	SAM
✓				1.75	973.5	42.06	2.32
✓	✓			1.88	1018.0	42.54	1.96
✓	✓	✓		1.88	1018.1	42.60	1.93
✓	✓	✓	✓	1.91	1018.9	42.82	1.88

(a) Break-down ablation studies to verify the effectiveness of modules.

Method	Params (M)	GFLOPS	PSNR (dB)	SAM
No SE	1.75	973.5	42.06	2.32
Global SE	1.91	1014.8	42.04	2.22
Local SE	1.84	993.8	42.60	1.93
Non-local SE	1.91	1018.9	42.82	1.88

(b) Ablation to the position of spectral enhancement (SE) module.

Table 6. Component analysis of various designs on ICVL dataset under random Gaussian noise.

noising performance by 0.42 dB improvement. The introduction of spectral shuffle (SS) also slightly improves the results, which validates the necessity of feature fusion. With memory unit (MU), the model gains 0.18 dB in PSNR, demonstrating the effectiveness of learning from a large-scale dataset to obtain representative low-rank vectors.

Position of SE Module. We further place our SE module at different positions to obtain the spatial-spectral correlation. The results are shown in Table 6(b). For global SE, the whole features of HSI is projected to one low-rank vector. Local SE stands for SE module that projected the feature inside a rectangle to one vector. Non-local SE, which is the employed design, projects several neighboring rectangles into one vector. Interestingly, global SE brings a slight decrease in performance, indicating extracting a low-rank vector from the entire HSI is inappropriate. As can be seen that non-local SE yields the best performance. We owe it to its ability to make interactions between spatial rectangles and aggregate information of neighboring similar pixels.

Visualization of Low-rank Vectors. To demonstrate the role of spectral enhancement module, we visualize several low-rank vectors obtained by SE module in Figure 6. The input cubes are severely influenced by noise and it is difficult to judge the similarities between cubes visually. However, low-rank vectors extracted from these noisy cube patches by SE module show clear similarities. Since the patch 7, 8 and 9 are all from the road area, their projected low-rank vectors are more similar to each other than to other vectors. This proves the ability of SE module to extract essential information from patches and suppress noise.

Parameter Analysis. We evaluate our proposed rectangle Transformer under different settings of rectangle size in Figure 7. We fix the width of rectangles and change their lengths for comparison. Since our method includes three layers of Transformer, we change the length in different layers. It can be observed that a rectangle with longer length may not bring better performance for HSI denoising, validating the essence of our proposed rectangle self-attention in modeling non-local similarity in the spatial domain.

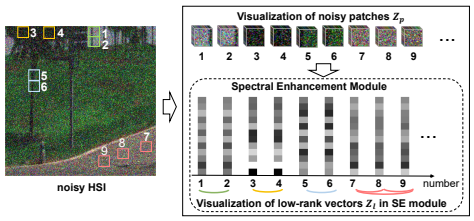


Figure 6. Visualization of low-rank vectors in SE module.

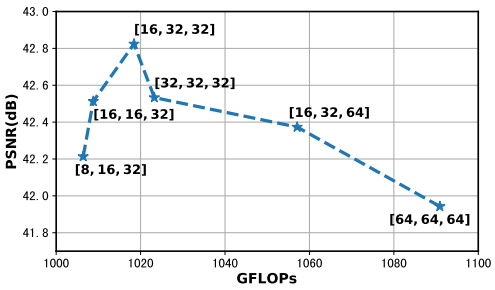


Figure 7. Different settings of rectangle's length at different layers. The widths is set to [1,2,4] for defaults.

5. Conclusion

In this paper, we present a spectral enhanced rectangle Transformer for HSI denoising, considering the spatial non-local similarity and spectral low-rank property of HSI. We exploit the non-local similarity via multi-shape rectangle self-attention in the spatial domain. Moreover, we integrate a spectral enhancement module with learnable memory units to explore the global spectral low-rank property of HSI. It introduces interactions across spatial rectangles while maintaining informative spectral characteristics. Extensive experiments demonstrate that our method significantly outperforms other competing methods with synthetic and real noisy HSIs. In the future, we plan to extend our method to cope with various HSI restoration tasks.

Acknowledgments This work was supported by National Key R&D Program of China (2022YFC3300704), and National Natural Science Foundation of China under Grants (62171038, 61827901, 62088101, and 62006023).

References

[1] Wele Gedara Chaminda Bandara and Vishal M Patel. Hypertransformer: A textural and spectral feature fusion transformer for pansharpening. In *CVPR*, pages 1767–1777, 2022. 1, 2

[2] Robert W Basedow, Dwayne C Carmer, and Mark E Anderson. Hydice system: Implementation and performance. In *Imaging Spectrometry*, volume 2480, pages 258–267. SPIE, 1995. 1

[3] Théo Bodrito, Alexandre Zouaoui, Jocelyn Chanussot, and Julien Mairal. A trainable spectral-spatial sparse coding model for hyperspectral image restoration. In *NeurIPS*, volume 34, pages 5430–5442, 2021. 2, 5, 6, 7

[4] Péter Burai, Balázs Deák, Orsolya Valkó, and Tamás Tomor. Classification of herbaceous vegetation using airborne hyperspectral imagery. *Remote Sensing*, 7(2):2046–2066, 2015. 1

[5] Yuanhao Cai, Jing Lin, Xiaowan Hu, Haoqian Wang, Xin Yuan, Yulun Zhang, Radu Timofte, and Luc Van Gool. Coarse-to-fine sparse transformer for hyperspectral image reconstruction. In *ECCV*, pages 686–704. Springer, 2022. 2

[6] 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 *CVPR*, pages 17502–17511, 2022. 2

[7] Xiangyong Cao, Xueyang Fu, Chen Xu, and Deyu Meng. Deep spatial-spectral global reasoning network for hyperspectral image denoising. *IEEE TGRS*, 2021. 1, 2, 5, 7

[8] Yi Chang, Luxin Yan, Xi-Le Zhao, Houzhang Fang, Zhijun Zhang, and Sheng Zhong. Weighted low-rank tensor recovery for hyperspectral image restoration. *IEEE TCYB*, 50(11):4558–4572, 2020. 1, 2, 4

[9] Yi Chang, Luxin Yan, and Sheng Zhong. Hyper-laplacian regularized unidirectional low-rank tensor recovery for multispectral image denoising. In *CVPR*, pages 4260–4268, 2017. 1, 2, 4, 5, 7

[10] Guangyi Chen and Shen-En Qian. Denoising of hyperspectral imagery using principal component analysis and wavelet shrinkage. *IEEE TGRS*, 49(3):973–980, 2010. 1, 2

[11] Xiangyu Chen, Xintao Wang, Jiantao Zhou, and Chao Dong. Activating more pixels in image super-resolution transformer. *arXiv preprint arXiv:2205.04437*, 2022. 1, 4

[12] Shen Cheng, Yuzhi Wang, Haibin Huang, Donghao Liu, Haoqiang Fan, and Shuaicheng Liu. Nbnet: Noise basis learning for image denoising with subspace projection. In *CVPR*, pages 4896–4906, 2021. 4

[13] Xiangxiang Chu, Zhi Tian, Yuqing Wang, Bo Zhang, Haibing Ren, Xiaolin Wei, Huaxia Xia, and Chunhua Shen. Twins: Revisiting the design of spatial attention in vision transformers. In *NeurIPS*, volume 34, pages 9355–9366, 2021. 1, 2, 3, 7

[14] Weisheng Dong, Guangming Shi, Xin Li, Yi Ma, and Feng Huang. Compressive sensing via nonlocal low-rank regularization. *IEEE TIP*, 23(8):3618–3632, 2014. 4

[15] Xiaoyi Dong, Jianmin Bao, Dongdong Chen, Weiming Zhang, Nenghai Yu, Lu Yuan, Dong Chen, and Baining Guo. Cswin transformer: A general vision transformer backbone with cross-shaped windows. In *CVPR*, pages 12124–12134, 2022. 1, 2, 3, 4, 7

[16] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, et al. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*, 2020. 1, 2

[17] Ying Fu, Antony Lam, Imari Sato, and Yoichi Sato. Adaptive spatial-spectral dictionary learning for hyperspectral image denoising. In *ICCV*, pages 343–351, 2015. 1, 2

[18] Ying Fu, Zichun Wang, Tao Zhang, and Jun Zhang. Low-light raw video denoising with a high-quality realistic motion dataset. *TMM*, 2022. 2

[19] Wei He, Quanming Yao, Chao Li, Naoto Yokoya, and Qibin Zhao. Non-local meets global: An integrated paradigm for hyperspectral denoising. In *CVPR*, pages 6868–6877, 2019. 1, 2, 3, 5, 7

[20] Wei He, Hongyan Zhang, Huanfeng Shen, and Liangpei Zhang. Hyperspectral image denoising using local low-rank matrix recovery and global spatial–spectral total variation. *IEEE J-STARS*, 11(3):713–729, 2018. 1

[21] Wei He, Hongyan Zhang, Liangpei Zhang, and Huanfeng Shen. Total-variation-regularized low-rank matrix factorization for hyperspectral image restoration. *IEEE TGRS*, 54(1):178–188, 2015. 1

[22] Danfeng Hong, Zhu Han, Jing Yao, Lianru Gao, Bing Zhang, Antonio Plaza, and Jocelyn Chanussot. Spectralformer: Rethinking hyperspectral image classification with transformers. *IEEE TGRS*, 60:1–15, 2021. 2

[23] Jie Hu, Li Shen, and Gang Sun. Squeeze-and-excitation networks. In *CVPR*, pages 7132–7141, 2018. 2, 4

[24] Jie Huang, Ting-Zhu Huang, Liang-Jian Deng, and Xi-Le Zhao. Joint-sparse-blocks and low-rank representation for hyperspectral unmixing. *IEEE TGRS*, 57(4):2419–2438, 2018. 2, 4

[25] 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. 6

[26] Jingyun Liang, Jiezhang Cao, Guolei Sun, Kai Zhang, Luc Van Gool, and Radu Timofte. Swinir: Image restoration using swin transformer. In *ICCV*, pages 1833–1844, 2021. 3, 7

[27] Bing Liu, Anzhu Yu, Kuiliang Gao, Xiong Tan, Yifan Sun, and Xuchu Yu. Dss-trm: deep spatial–spectral transformer for hyperspectral image classification. *Eur. J. Remote Sens*, 55(1):103–114, 2022. 2

[28] 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 *ICCV*, pages 10012–10022, 2021. 1, 2, 3, 4, 7

[29] Ting Lu, Shutao Li, Leyuan Fang, Yi Ma, and Jón Atli Benediktsson. Spectral–spatial adaptive sparse representation for hyperspectral image denoising. *IEEE TGRS*, 54(1):373–385, 2015. 2

[30] Ningning Ma, Xiangyu Zhang, Hai-Tao Zheng, and Jian Sun. Shufflenet v2: Practical guidelines for efficient cnn architecture design. In *ECCV*, pages 116–131, 2018. 4

[31] Matteo Maggioni, Vladimir Katkovnik, Karen Egiazarian, and Alessandro Foi. Nonlocal transform-domain filter for volumetric data denoising and reconstruction. *IEEE TIP*, 22(1):119–133, 2012. 5, 7

[32] Erting Pan, Yong Ma, Xiaoguang Mei, Fan Fan, Jun Huang, and Jiayi Ma. Sqad: Spatial-spectral quasi-attention recurrent network for hyperspectral image denoising. *IEEE TGRS*. 2

[33] Li Pang, Weizhen Gu, and Xiangyong Cao. Trq3dnet: A 3d quasi-recurrent and transformer based network for hyperspectral image denoising. *Remote Sensing*, 14(18):4598, 2022. 7

[34] Qian Shi, Xiaopei Tang, Taoru Yang, Rong Liu, and Liangpei Zhang. Hyperspectral image denoising using a 3-d attention denoising network. *IEEE TGRS*, 2021. 2

[35] Oleksii Sidorov and Jon Yngve Hardeberg. Deep hyperspectral prior: Single-image denoising, inpainting, super-resolution. In *CVPR*, pages 0–0, 2019. 1

[36] Xunyang Su, Jinjiang Li, and Zhen Hua. Transformer-based regression network for pansharpening remote sensing images. *IEEE TGRS*, 60:1–23, 2022. 2

[37] Muhammad Uzair, Arif Mahmood, and Ajmal Mian. Hyperspectral face recognition with spatiospectral information fusion and pls regression. *IEEE TIP*, 24(3):1127–1137, 2015. 1

[38] Muhammad Uzair, Arif Mahmood, and Ajmal S Mian. Hyperspectral face recognition using 3d-dct and partial least squares. In *BMVC*, volume 1, page 10, 2013. 1

[39] Wenhai Wang, Enze Xie, Xiang Li, Deng-Ping Fan, Kaitao Song, Ding Liang, Tong Lu, Ping Luo, and Ling Shao. Pyramid vision transformer: A versatile backbone for dense prediction without convolutions. In *ICCV*, pages 568–578, 2021. 1, 2, 3

[40] Xiaolong Wang, Ross Girshick, Abhinav Gupta, and Kaiming He. Non-local neural networks. In *CVPR*, pages 7794–7803, 2018. 2

[41] Kaixuan Wei, Ying Fu, and Hua Huang. 3-d quasi-recurrent neural network for hyperspectral image denoising. *IEEE TNNLS*, 32(1):363–375, 2020. 1, 2, 5, 7

[42] Wei Wei, Lei Zhang, Chunna Tian, Antonio Plaza, and Yanning Zhang. Structured sparse coding-based hyperspectral imagery denoising with intracluster filtering. *IEEE TGRS*, 55(12):6860–6876, 2017. 1

[43] Xueling Wei, Wei Li, Mengmeng Zhang, and Qingli Li. Medical hyperspectral image classification based on end-to-end fusion deep neural network. *IEEE T-IM*, 68(11):4481–4492, 2019. 1

[44] Fengchao Xiong, Jun Zhou, Shuyin Tao, Jianfeng Lu, Jiantao Zhou, and Yuntao Qian. Smds-net: Model guided spectral-spatial network for hyperspectral image denoising. *IEEE TIP*, 31:5469–5483, 2022. 2

[45] Fengchao Xiong, Jun Zhou, Qinling Zhao, Jianfeng Lu, and Yuntao Qian. Mac-net: Model-aided nonlocal neural network for hyperspectral image denoising. *IEEE TGRS*, 60:1–14, 2021. 2, 5

[46] Rui Yang, Hailong Ma, Jie Wu, Yansong Tang, Xuefeng Xiao, Min Zheng, and Xiu Li. Scalablevit: Rethinking the context-oriented generalization of vision transformer. *arXiv preprint arXiv:2203.10790*, 2022. 1

[47] Tian Ye, Mingchao Jiang, Yunchen Zhang, Liang Chen, Erkang Chen, Pen Chen, and Zhiyong Lu. Perceiving and modeling density is all you need for image dehazing. *arXiv preprint arXiv:2111.09733*, 2021. 1, 2

[48] Qiangqiang Yuan, Liangpei Zhang, and Huanfeng Shen. Hyperspectral image denoising employing a spectral–spatial adaptive total variation model. *IEEE TGRS*, 50(10):3660–3677, 2012. 2, 5, 7

[49] Qiangqiang Yuan, Qiang Zhang, Jie Li, Huanfeng Shen, and Liangpei Zhang. Hyperspectral image denoising employing a spatial–spectral deep residual convolutional neural network. *IEEE TGRS*, 57(2):1205–1218, 2018. 1, 2, 5

[50] 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 *CVPR*, pages 5728–5739, 2022. 1, 3, 7

[51] Feng Zhang, Kai Zhang, and Jiande Sun. Multiscale spatial–spectral interaction transformer for pan-sharpening. *Remote Sensing*, 14(7):1736, 2022. 2

[52] Hongyan Zhang, Lu Liu, Wei He, and Liangpei Zhang. Hyperspectral image denoising with total variation regularization and nonlocal low-rank tensor decomposition. *IEEE TGRS*, 58(5):3071–3084, 2019. 1, 2, 4, 7

[53] Tao Zhang, Ying Fu, and Cheng Li. Hyperspectral image denoising with realistic data. In *ICCV*, pages 2248–2257, 2021. 6

[54] Xiangtao Zheng, Yuan Yuan, and Xiaoqiang Lu. Hyperspectral image denoising by fusing the selected related bands. *IEEE TGRS*, 57(5):2596–2609, 2018. 2

[55] Zilong Zhong, Ying Li, Lingfei Ma, Jonathan Li, and Wei-Shi Zheng. Spectral–spatial transformer network for hyperspectral image classification: A factorized architecture search framework. *IEEE TGRS*, 60:1–15, 2021. 2