

## Hyperspectral image super-resolution via joint network with spectral-spatial strategy

Yixin Dong, Bo Yang, Cong Liu, Zemin Geng & Taiping Wang

To cite this article: Yixin Dong, Bo Yang, Cong Liu, Zemin Geng & Taiping Wang (18 Jul 2025): Hyperspectral image super-resolution via joint network with spectral-spatial strategy, Geo-spatial Information Science, DOI: [10.1080/10095020.2025.2522909](https://doi.org/10.1080/10095020.2025.2522909)

To link to this article: <https://doi.org/10.1080/10095020.2025.2522909>



© 2025 Wuhan University. Published by Informa UK Limited, trading as Taylor & Francis Group.



Published online: 18 Jul 2025.



Submit your article to this journal 



View related articles 



View Crossmark data 

# Hyperspectral image super-resolution via joint network with spectral-spatial strategy

Yixin Dong <sup>a</sup>, Bo Yang  <sup>a</sup>, Cong Liu  <sup>b</sup>, Zemin Geng  <sup>a</sup> and Taiping Wang  <sup>a</sup>

<sup>a</sup>The State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan, China; <sup>b</sup>The School of Optical-Electrical and Computer Engineering, University of Shanghai for Science and Technology, Shanghai, China

## ABSTRACT

Hyperspectral image (HSI) super-resolution (SR) faces significant challenges due to the inherent difficulty in acquiring large-scale training data and the complex spectral-spatial relationships in HSIs that conventional deep-learning-based methods often fail to fully exploit. While existing approaches typically stack convolutional layers to increase network depth, they frequently overlook the structured continuity of spectral bands and non-local spatial similarities, resulting in limited performance and overfitting risks. To address these limitations, we propose SRLSGAT, a novel joint spectral-spatial network that combines a vertical-horizontal bi-directional LSTM (VH-BiLSTM) for modeling multi-directional spectral correlations and a multi-adjacent weight matrix graph attention network (MAW-GAT) for capturing non-local patch relationships. Besides, we design a spectral attention mechanism (SpeAM), which dynamically weights remote dependencies through bidirectional spectral sequence analysis, while the graph-based spatial processing adaptively learns patch similarities through optimized edge connections. Extensive experiments on three HSI datasets show that SRLSGAT has better performance relative to SOTA SR methods. The code is accessible at <https://github.com/Dongyx1128/SRLSGAT>.

## ARTICLE HISTORY

Received 4 March 2025

Accepted 11 June 2025

## KEYWORDS

Hyperspectral image; super-resolution; graph attention network; bi-directional LSTM; SpeAM

## 1. Introduction

Hyperspectral imaging technology captures the 2D geometric spatial and 1D spectral information of a target by detecting electromagnetic waves reflected from the object using hyperspectral sensors, thus obtaining the spatial and spectral data of ground targets. Hyperspectral images (HSI) contain hundreds of consecutive spectral bands captured by imaging systems. Compared to RGB remote sensing images (Akhtarmanesh et al. 2024; Felegari et al. 2023; Mahdipour et al. 2024; Marzvan et al. 2021; Mirhoseini Nejad, Abbasi-Moghadam, and Sharifi 2024), they provide richer spectral information. As a result, hyperspectral images have strong spectral diagnostic capability and greatly improve the ability to analyze things. They are widely used in many fields, such as land surface detection (Liu, Su, and Li 2016), surveillance videos (Rasti et al. 2016), cancer detection (Pike et al. 2016), and aerial field (Arun et al. 2019).

The spatial resolution of HSI is widely recognized as reflecting the degree of detail in the spatial information within an image, whereas spectral resolution plays a crucial role (Esmaeili et al. 2024) in characterizing the structural attributes of different features. However, the hardware limitations of hyperspectral imaging technologies require a trade-off between spectral

resolution and spatial resolution, where an increase in spectral resolution generally decreases spatial resolution. The purpose of the HSI SR technique is to enhance HSI spatial resolution via post-processing, without changing hardware performance (Sharifi and Safari 2025; Vafaeinejad et al. 2025). This technique has received wide attention from more and more researchers and has become a hot issue in computer vision.

Many SR methods have been introduced in recent years. Based on whether or not auxiliary information is utilized, HSI SR methods can be classified into two types: fusion-based HSI SR and single-image HSI SR (Yokoya, Grohnfeldt, and Chanussot 2017). Fusion-based HSI reconstruction methods (Yu et al. 2024) combine information from low-resolution hyperspectral images (LR-HSI) and high-resolution multispectral (HR-MSI) as prior data (Li et al. 2018). These methods primarily employ four approaches: dictionary learning (Han et al. 2020), sparse representation (Akhtar, Shafait, and Mian 2015; Dong et al. 2016; Yokoya, Yairi, and Iwasaki 2012), matrix/tensor decomposition (Bu et al. 2021; Dian and Li 2019; Dian, Fang, and Li 2017; Dian, Li, and Fang 2019), and deep-learning-based (Hong et al. 2022; Sun et al. 2021; Zhang et al. 2021) SR methods. However, in

**CONTACT** Bo Yang  XT000025@whu.edu.cn

© 2025 Wuhan University. Published by Informa UK Limited, trading as Taylor & Francis Group.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

practical scenarios, acquiring auxiliary images with a high degree of matching remains challenging.

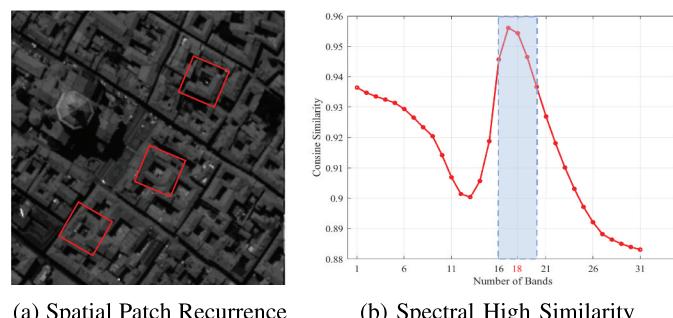
Single HSI SR achieves super-resolution without auxiliary data, enabling flexible image post-processing. This advantage makes it a research focus. Most current model-based methods focus on capturing spatial-spectral correlations using manual a prior strategies and matrix decomposition strategies. In the prior strategy, researchers usually use different priors (Li et al. 2025) applied to HSI, for example, non-local priors (Wang et al. 2017), sparse priors (Li et al. 2016), and low-rank priors (He et al. 2016). For matrix decomposition (Liu, Fan, and Zhang 2022; Zhao et al. 2011) strategies, researchers generally reformat the target 3D HR-HSI into a 2D spectral or spatial matrix, assuming that this matrix can be represented in a low-dimensional subspace via a spatial matrix. These algorithms focus on capturing the internal feature structure of the reconstructed HSI. However, these methods suffer from increased computational complexity and degraded performance. At present, learning-based methods show excellent performance in gray/RGB image SR, for example, the pioneering work SRCNN (Dong et al. 2016) proposed by Dong et al. learns a nonlinear mapping between LR and HR, demonstrating the advantages of convolutional neural networks (CNNs) over traditional methods. These residual block skip connection-based methods (GSCN (Liu and Lei 2021), FMNSR (Zhang et al. 2022), EDSR (Lim et al. 2017), LapSRN (Lai et al. 2017) and SRDenseNet (Tong et al. 2017)) have also demonstrated good performance by deepening the network hierarchy. Besides enhancing network depth, non-local attention methods have been extensively applied in many different computer vision tasks. Many SR networks, such as RCAN (Zhang et al. 2018), NLRN (Liu et al. 2018), and RNAN (Zhang et al. 2019), apply attention mechanisms to utilize the structural relevance of image features, which improves performance. However, compared to RGB SR, HSI SR has more demand. Li et al. (2022) proposed a dual-stage SR (DualSR) method, which mainly utilizes the correlation between spectral in adjacent ranges for

spectral angle constraints to achieve spectral consistency. But the method is simply achieves band-by-band SR and lacks to consider the correlation within the continuous spectral range. In summary, two problems exist with most of present approaches. Firstly, HSI contain hundreds of spectral bands, challenging the consideration of spectral correlation. Secondly, HSI covers a vast geographic area, but it has limited spatial resolution compared to the dimensions of satellite observations, making it more challenging to construct HSI SR models.

### 1.1. Motivation

In fact, HSI presents distinctive attributes in both spectral and spatial dimensions (Figure 1). In terms of spatial texture features, image patches are recurrent globally within non-local regions of the high-resolution image, and feature information can be shared by capturing similarity image patches. Within the range of spectral sequences, the bands of adjacent sequences can also explicitly assist in the reconstruction of the current band, as these bands also have high correlation. Adjacent bands within the same spectral range provide complementary information. This helps fill missing features during band reconstruction.

Therefore, the key problem is how to effectively exploit the recurrent and adjacent bands of image patches to enhance the SR performance of HSIs. Motivated by these discoveries, we propose an HSI SR algorithm (SRLSGAT) via joint network with spectral-spatial strategy to extensively explore the spectral and spatial dimensions. To obtain HR-HSI with the required spectral and spatial resolution, our method effectively models two key characteristics: (1) Spectral remote dependencies, we analyze multi-directional correlations across spectral bands. (2) Spatial non-local similarity, we capture global recurrence patterns of image patches. This approach fully utilizes both spectral sequences and spatial relationships. Specifically, we integrate key components. First, the vertical-horizontal bi-directional long short-term memory network (VH-BiLSTM) extracts depth



**Figure 1.** Illustration of the image patches recurrence in spatial non-local regions and high similarity in spectral adjacent bands. We show the similarity values between the 18th band and its adjacent bands. Here the similarity values are calculated using cosine similarity.

spectral-spatial aggregate feature. Second, the multi-adjacent weight matrix graph attention network (MAW-GAT) process spatial relationships. This combined approach effectively merges spectral and spatial information for enhanced reconstruction. Meanwhile, to enhance the continuity between bands and the non-local similarity of image patches, the spectral attention mechanism (SpeAM) is designed, which selectively extracts correlation information from multiple directions of spectral sequences. We introduce a multi-adjacent weight matrix construction method that enhances the global information interaction among image patches by using their inherent non-local self-similarity property. Extensive experiments show that our SRLSGAT method is effective in terms of performance improvement.

## 1.2. Contributions

To solve these problems, we propose an HSI SR algorithm (SRLSGAT) via joint network with spectral-spatial strategy. This paper's significant contributions are as follows:

- To address limited HSI training samples, we propose SRLSGAT, a parallel architecture network that integrates spectral sub-network (VH-BiLSTM) and spatial sub-networks (MAW-GAT) to achieve accurate LR-HSI to HR-HSI mapping.
- We design SpeAM to enhance reconstruction quality through bidirectional (vertical-horizontal) spectral sequence interactions, effectively constraining the reconstruction process.
- Our multi-weight adjacency matrix method improves non-local similarity modeling by adaptively weighting patch connections in the spatial sub-network.

The paper is organized as follows. Section II reviews the related literature of HSI SR. The details of the proposed SRLSGAT are presented in Section III. Experimental parameter discussion and results compared with SOTA methods are present in Section IV. Finally, Section V concludes the article.

## 2. Related work

In this section, we only review deep learning based SR methods, firstly, single HSI SR based methods are reviewed. Second, we review some Non-local neural network single image SR methods.

### 2.1. Single hyperspectral image super-resolution

Due to the rapid development of deep learning techniques (Safari et al. 2024), researchers have expanded SR methods for RGB images to SR methods for HSI. For example, Li et al. (2018) proposed the grouped deep recursive residual network (GDRRN) by embedding

grouped recursive modules into the global residual structure through 2D convolution. Jiang et al. (2020) proposed the grouped convolution progressive upsampling network (SSPSR), but due to the complexity of the hyperspectral image structure, they did not mine the correlation of the spectral, and the direct SR method using RGB images would have the problem of spectral distortion (Mei et al. 2017). Later, scholars have studied the SR of hyperspectral images by using 3D convolution. For example, Mei et al. (2017) proposed a 3D fully convolutional neural network (3D-FCNN) using a five-layer structure. Li, Wang, and Li (2020) designed a mixed 2D/3D convolutional network (MCNet). Later, Li, Wang, and Li (2021) proposed to explore 2D/3D relational networks (ERCSR) by sharing spatial domains. 3D convolution can extract both spectral and spatial features, retain more spectral related information, and improve network performance significantly (Farmonov et al. 2024). Li, Yuan, et al. (2022) proposed a dual-stage super-resolution (DualSR) method, in which the coarse-stage progressively generates super-resolution images band by band through an alternating spatial-spectral fusion mechanism. SFPN (Li, Gong, et al. 2022) proposes a two-stage learning approach for HSI SR, learning general spatial spectral priors and image-specific details, respectively. The rich HR content of HSI Spectral and RGB images is utilized to guide the model representation learning. IFMSR (Q. Li et al. 2023) also use RGB images to guide the learning of HSI SR, a multi-corresponding patch aggregation is designed to globally assemble this contextual information, which is beneficial for feature learning.

Recent progress in unsupervised HSI SR has introduced methods that eliminate reliance on prior sensor parameters. HyCoNet (Zheng et al. 2021) proposes an unsupervised HSI-MSI fusion framework using three coupled autoencoders and adaptive convolution to jointly learn spectral unmixing and sensor characteristics (PSF/SRF), enabling robust performance without prior knowledge. NonRegSRNet (Zheng et al. 2022) addresses nonrigid deformation issues by jointly performing registration and SR in a triplet CNN, adaptively estimating PSF/SRF parameters. EDIP-Net (Gao et al. 2023) enhances deep image prior (DIP)-based SR via a two-stage architecture: generating coarse spectral estimates through cross-scale learning and refining results with a dual U-Net incorporating degradation-aware fusion, effectively modeling hyperspectral priors in an unsupervised manner.

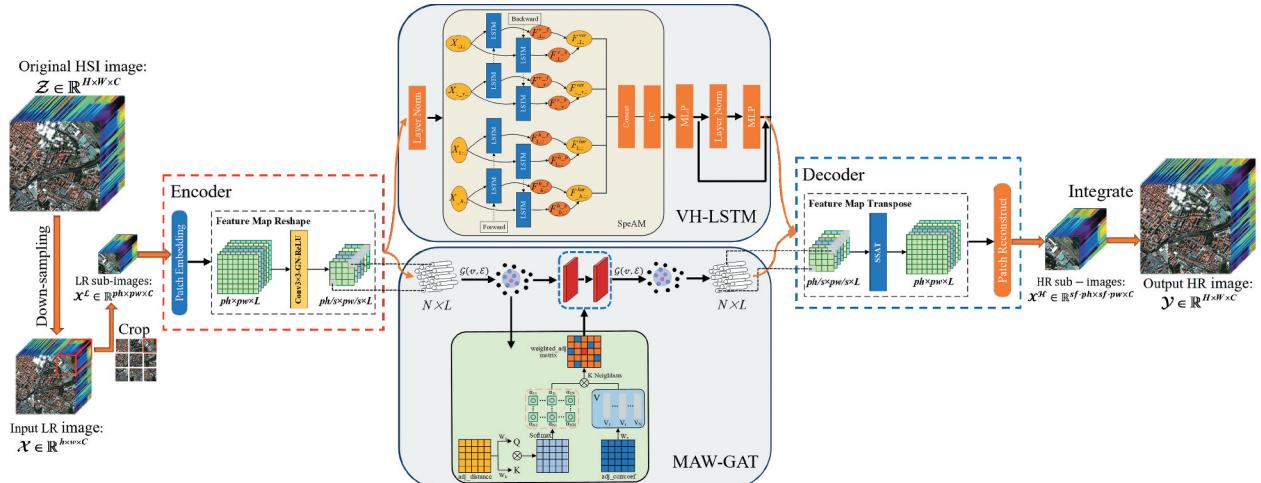
### 2.2. Non-local neural network single image super-resolution methods

SR is a low-level task which requires more contextual information to obtain more clarity. At present, some researchers design methods for global

information fusion using similarity image patches, all of them obtaining better results. Even though CNN-based methods have made a lot of progress, other methods based on deep learning have also tried to do SR on images. e.g. graph neural networks (GNN) (Gori, Monfardini, and Scarselli 2005), recurrent neural networks (RNN), and long short-term memory networks (LSTM) (Hochreiter and Schmidhuber 1997). Recently, graph-based SR algorithm methods have been proposed in the framework of using graph neural networks to capture the mapping relationship between LR and HR in images. For example, Zhou et al. (2020) used a new cross-scale internal graph neural network (IGNN) to explore the cross-scale patch recursion properties of natural images. Yang and Qi (2021) proposed the channel attention and spatial graph convolution network (CASGCN), which combines the channel attention mechanism and the spatial graph convolution network (GCN) to enhance feature extraction capability. Yan et al. (2021) proposed an SRGAT network based on a GAT that uses high-level information to facilitate low-level features via a feedback mechanism. Liu and Dong (2022) proposed a CNN-enhanced graph attention network (CEGATSR), which can fully capture different features by using a graph attention to capture different receptive fields by using relatively few layers.

### 3. The proposed method

This section outlines the fundamental architecture of the SRLSGAT network. The whole structure includes three components, starting with the Encoder-Decoder part of the network and including with the spectral and spatial branches arranged in parallel within the network.



**Figure 2.** Overall architecture of proposed SRLSGAT framework.

### 3.1. Network Architecture

As shown in Figure 2, the proposed SRLSGAT is designed to reconstruct high-resolution hyperspectral images HR HSI  $Y \in \mathbb{R}^{H \times W \times C}$  from low-resolution observations LR HSI  $X \in \mathbb{R}^{h \times w \times c}$  by jointly learning spectral and spatial features. The network consists of two parallel sub-networks: (1) a spectral feature extraction branch using a vertical-horizontal bi-directional long short-term memory network (VH-BiLSTM), which captures remote dependencies adjacent spectral bands in both vertical and horizontal directions. (2) A spatial feature extraction branch based on a multi-adjacent weight matrix graph attention network (MAW-GAT), which models non-local relationships among image patches via attention mechanisms on multiple weighted adjacency matrices. Additionally, a spectral attention mechanism (SpeAM) is incorporated to selectively emphasize informative spectral features, improving reconstruction under limited training data. (3) An Encoder-Decoder, a VH-BiLSTM branch for spectral feature extraction and a MAW-GAT branch for spatial feature extraction; the latter is the part that encodes and decodes the image information for processing. So we aim to predict the HR HSI from the input LR HSI using the proposed end-to-end SR algorithm SRLSGAT, i.e.

$$y = \text{Integrate}(f_{\text{Net}}(\text{Crop}(X))). \quad (1)$$

In addition,

$$\begin{aligned} \chi &= \text{Down Sampling}(Z) + \eta, \\ x^L &= \text{Crop}(X), \\ x^H &= f_{\text{Net}}(x^L), \\ Y &= \text{Integrate}(x^H), \end{aligned} \quad (2)$$

where  $Z \in \mathbb{R}^{H \times W \times C}$  is denoted as the original HSI,  $X \in \mathbb{R}^{h \times w \times C}$  is denoted as the input LR HSI after

downsampling from image  $Z$ , and  $Y \in \mathbb{R}^{H \times W \times C}$  is the output HSI learned by the SRLSGAT network from the LR HSI  $X$ . The LR sub-images set  $x^L \in \mathbb{R}^{ph \times pw \times C}$  is the sub-images cut from the input images  $X$ , and  $x^H \in \mathbb{R}^{sf \cdot ph \times sf \cdot pw \times C}$  is the HR sub-images learned by the network.  $H$  and  $W$  are the height and width of the spatial dimension,  $C$  is the number of spectral bands,  $ph$  and  $pw$  are the height and width of the sub-images respectively, and they satisfy  $H = sf \times h$  and  $W = sf \times w$ , where  $sf$  denotes the scale factor.  $\eta$  denotes the noise.  $f_{Net}(\cdot)$  denotes the proposed SRLSGAT network.  $Crop(\cdot)$  denotes the cut operation on the image,  $Integrate(\cdot)$  indicates a concatenation operation on the sub-images, and  $DownSampling(\cdot)$  denotes the downsampling operation.

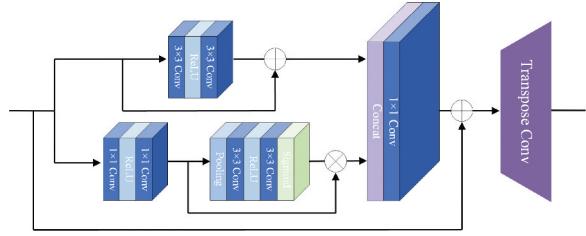
### 3.2. Encoder-Decoder

In the proposed HSI SR method (SRLSGAT) based on the spectral-spatial strategy with combined bi-directional LSTM and graph attention networks, we use an Encoder-Decoder component at the beginning and end of the network. This incorporation aims to input three-dimensional information from the HSI into traditional RNN and GAT networks during spectral and spatial feature extraction. The Encoder-Decoder structure facilitates necessary feature transformation and target reconstruction of sub-images from the image patch level perspective. This is achieved through the reconstruction network's pre-processing and post-processing stages within the Encoder-Decoder component.

1) *Encoder*: The  $Coder_E$  aims to extract shallow features and perform dimensional conversion on sub-images. For a given sub-images  $x^L$ , the initial step involves a patch embedding operation, which includes a stack of convolutional layers. Then, the feature are input into the feature map reshape module, where the extracted shallow features are handled by a convolutional layer with a step size of  $s$  and a convolutional kernel size of  $k$  to obtain a feature map  $H_E \in \mathbb{R}^{\frac{ph}{s} \times \frac{pw}{s} \times L}$ . Mathematically, the encoding process performed by the Encoder is represented as

$$H_E = Coder_E(x^L). \quad (3)$$

2) *Decoder*: The  $Coder_D$  aim to dimensional inverse conversion and target reconstruction of the feature maps after deep feature extraction. Specifically, the Encoder-Decoder perform the symmetric inverse conversion of the feature map's dimension using the feature map transpose module. The extracted deep features  $H_{spe-spa} \in \mathbb{R}^{\frac{ph}{s} \times \frac{pw}{s} \times L}$  from spectral-spatial sub-networks are subsequently directed into the spatial-spectral attention transposition module (SSAT), illustrated in Figure 3,



**Figure 3.** The spatial-spectral attention transposition module (SSAT).

which includes spatial attention, channel attention, and transpose convolution. After that, the target reconstruction is performed, and the  $x^H \in \mathbb{R}^{sf \cdot ph \times sf \cdot pw \times C}$  sub-images is obtained after the patch reconstruction layer. Mathematically, the process by the Decoder is represented as

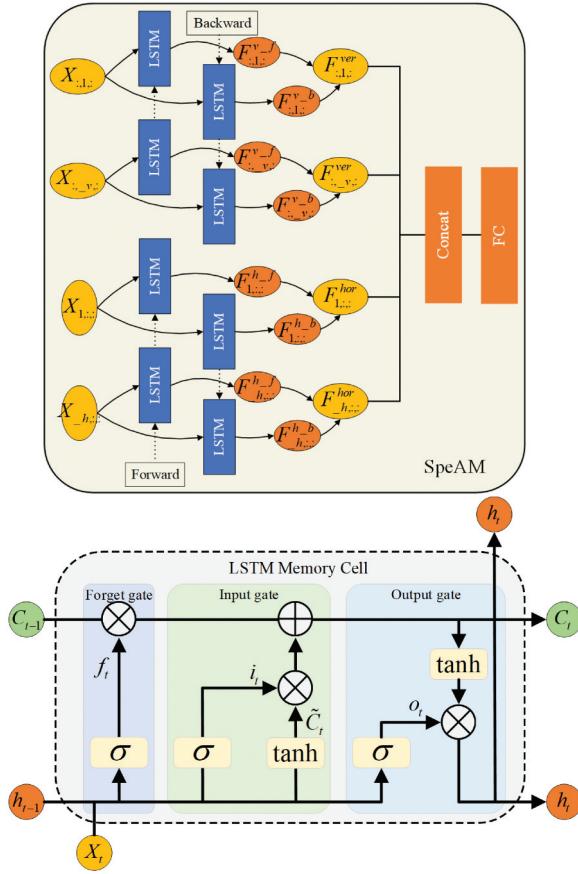
$$x^H = Coder_D(H_{spe-spa}). \quad (4)$$

### 3.3. Vertical-horizontal Bi-directional long short-term memory network (VH-BiLSTM)

HSI is known for its high spectral resolution precisely because HSI has a large number of spectral bands, so many recursive units are required to process the spectral information. However, most research has ignored the remote dependence among spectral bands. To effectively maintain the feature information in adjacent spectral bands, we propose a spectral attention mechanism (SpeAM) for spectral information extraction. This approach selectively models spectral remote dependencies by analyzing multi-directional correlations between spectral bands. The spectral sequences from different directions are adaptively weighted, effectively reducing spectral variations during reconstruction.

In the VH-BiLSTM of SRLSGAT framework, the structure of our proposed SpeAM is shown in Figure 4. First, the LayerNorm operation is used to preserve the feature relationships between spectral bands. For an input  $X \in \mathbb{R}^{\frac{ph}{s} \times \frac{pw}{s} \times L}$  is given to the SpeAM, we focus on using the correlation information of spectral sequences between bands in multiple directions. The SpeAM consists of two directions of BiLSTMs: a vertical spectral sequence BiLSTM and a horizontal spectral sequence BiLSTM. The structure of the LSTM cell is shown in Figure 4, and its feature propagation is defined as follows:

$$\begin{aligned} f_t &= \sigma(W_f \cdot [h_{t-1}, X_t] + b_f), \\ i_t &= \sigma(W_i \cdot [h_{t-1}, X_t] + b_i), \\ \tilde{C}_t &= \tanh(W_c \cdot [h_{t-1}, X_t] + b_c), \\ o_t &= \sigma(W_o \cdot [h_{t-1}, X_t] + b_o), \\ C_t &= f_t * C_{t-1} + i_t * \tilde{C}_t, \\ h_t &= o_t * \tanh(C_t). \end{aligned} \quad (5)$$



**Figure 4.** SpeAM and LSTM architecture.

where  $\sigma$  denotes the logistic sigmoid function and  $*$  represents the Hadamard product operation. The matrices  $W_f$ ,  $W_i$ ,  $W_c$ ,  $W_o$  are weight matrices. The vectors  $b_f$ ,  $b_i$ ,  $b_c$ ,  $b_o$  are bias vectors.

Specifically, in spectral sub-network VH-BiLSTM has an input  $X = [X_{:,1,:}, \dots, X_{:,-v,:}, \dots, X_{:,pw,:}]$ ,  $X_{:,-v,:} \in \mathbb{R}^{\frac{ph}{s} \times L}$  is regarded as the input set of vertical spectral sequences. Here,  $ph$  is the number of tokens in the vertical direction,  $pw$  is the number of vertical spectral sequences, and  $L$  is the number of hidden layer dimensions in the spectral band. All sequences  $X_{:,-v,:}$  are into the vertical BiLSTM:

$$\begin{aligned} F_{:,-v,:}^{ver} &= [LSTM_f(X_{:,-v,:}) || LSTM_b(X_{:,-v,:})] \\ &= [F_{:,-v,:}^{v,f} || F_{:,-v,:}^{v,b}] = BiLSTM(X_{:,-v,:}) \end{aligned} \quad (6)$$

Similarly, for input  $X = [X_{1,:,:}, \dots, X_{-h,:,:}, \dots, X_{ph,:,:}]$ ,  $X_{-h,:,:} \in \mathbb{R}^{\frac{ph}{s} \times L}$  is viewed as the input set of horizontal spectral sequences. All sequences  $X_{-h,:,:}$  are into the horizontal BiLSTM:

$$\begin{aligned} F_{-h,:,:}^{hor} &= [LSTM_f(X_{-h,:,:}) || LSTM_b(X_{-h,:,:})] \\ &= [F_{-h,:,:}^{h,f} || F_{-h,:,:}^{h,b}] = BiLSTM(X_{-h,:,:}) \end{aligned} \quad (7)$$

where the outputs of BiLSTM are  $F_{:,-v,:}^{ver} \in \mathbb{R}^{\frac{ph}{s} \times 2L}$  and  $F_{-h,:,:}^{hor} \in \mathbb{R}^{\frac{ph}{s} \times 2L}$ .  $LSTM_f(\cdot)$  and  $LSTM_b(\cdot)$  are the forward LSTM operation and the backward LSTM operation, respectively. Therefore, the SpeAM can utilize

more multiple directions spectral information extracted by the vertical and horizontal BiLSTM to enhance the continuity between spectral bands and reduce the spectral variations.

The feature map in the vertical and horizontal directions is then cumulated by channel dimension, respectively:

$$\begin{cases} H^{ver} = \sum_{v=1}^{pw} F_{:,-v,:}^{ver}; \\ H^{hor} = \sum_{h=1}^{ph} F_{-h,:,:}^{hor}, \end{cases} \quad (8)$$

Here, vertical and horizontal output  $H^{ver} \in \mathbb{R}^{\frac{ph}{s} \times \frac{pw}{s} \times 2L}$ ,  $H^{hor} \in \mathbb{R}^{\frac{ph}{s} \times \frac{pw}{s} \times 2L}$ .

We concatenate the output  $H^{ver}$  of the vertical spectral sequence with the output  $H^{hor}$  of the horizontal spectral sequence, followed by the full concatenation operation:

$$H_{spe}^{VH} = FC(H^{VH}) = FC([H^{ver} || H^{hor}]), \quad (9)$$

where the  $H^{VH} \in \mathbb{R}^{\frac{ph}{s} \times \frac{pw}{s} \times 4L}$ ,  $H_{spe}^{VH} \in \mathbb{R}^{\frac{ph}{s} \times \frac{pw}{s} \times L}$ , and  $FC(\cdot)$  denotes the fully connected operation.

Finally, the features are extracted by MLP, LayerNorm, and MLP layers to obtain the deep spectral feature map output  $H_{spe} \in \mathbb{R}^{\frac{ph}{s} \times \frac{pw}{s} \times L}$ :

$$H_{spe} = MLP(LayerNorm(MLP(H_{spe}^{VH}))). \quad (10)$$

### 3.4. Multi-adjacent weight matrix graph attention network (MAW-GAT)

HSIs not only have more spectral bands but also have different kinds of target features with similarity in the non-local, leading to recurrence on a global scale. Given the limited training samples of HSIs, existing CNN-based methods have difficulty capturing the global similarity structure of image patches, even within a large receptive field. So we design the spatial sub-network MAW-GAT, as shown in Figure 2, which aims to learn the global recurrence information of each patch pair feature space.

We construct a graph by utilizing image patches as node vectors. The importance of similarity between node vectors is computed to obtain the adjacent weight matrix among image patches, and through propagation and aggregation of similarity information among node vectors in the graph attention layer, achieve global information aggregation of image patches. Subsequently, the graph is symmetrically inverted, and node vectors are output as image non-local spatial features.

Specifically, we reshape feature map  $H_E \in \mathbb{R}^{\frac{ph}{s} \times \frac{pw}{s} \times L}$  to  $\mathbf{H} \in \mathbb{R}^{N \times L}$ , where  $N = \frac{ph}{s} \times \frac{pw}{s}$  and each row of size  $1 \times L$  corresponds to a node. Construct the nodes as a graph  $G(V, E)$ , where  $V$

denotes the vertex set,  $E$  denotes the edge set, and  $|V| = N$ ,  $E = V \times V$ . Liu et al. (Liu and Dong 2022) use the Euclidean distance between node vectors directly as a measure of the similarity weight between image patches. However, this method lacks scale normalization, leading to potential issues with data scale discrepancies in the vector space. Hence, we propose a multi-adjacent weight matrix method. In this method, the distance (Euclidean distance and Chebyshev distance) matrix and Pearson correlation coefficient matrix between each node vector are computed based on the graph  $G(V, E)$ , which is defined as Equation (11). Then, we perform a weight operation and select  $d$  nearest neighbors from the obtained adjacent matrix to obtain the multi-adjacent weight matrix.

$$\begin{aligned} adj\_euc, adj\_cheby &= Distance[G(V, E)], \\ adj\_corrcoef &= Corrcoef[G(V, E)]. \end{aligned} \quad (11)$$

Then initialize it using learnable nonlinear weights:

$$\begin{cases} Q = adj\_euc \odot W_q, \\ K = adj\_cheby \odot W_k, \\ V = adj\_corrcoef \odot W_v, \end{cases} \quad (12)$$

where  $W_q, W_k, W_v \in \mathbb{R}^{N \times N}$  are trainable weight parameters, the matrix  $Q$  works as the query distance target, matrix  $K$  serves as the transformation distance key for spatial inputs, distance correlation is established through a sequence key  $K$ , which informs the attention distribution  $\alpha$  based on the reference  $Q$ , and  $V$  is the correlation information value to be extracted.  $\odot$  denotes the dot product of two vectors, which represents the similarity between two node vectors. Then calculation of similarity using the scoring function:

$$\alpha_{ij} = \sum_{t=1}^N \frac{k_{it} \times q_{jt}}{\sqrt{N}}, \quad (13)$$

where  $k_i$  is the  $i$ -th row of  $k$ , and  $q_j$  is the  $j$ -th column of  $Q^T$ . We construct the similarity weight matrix  $\hat{\alpha}_{ij} \in (0, 1)$  based on the relevance of node vectors, which represents the importance of each position in the distance matrix. The normalization can be formulated as follows:

$$\hat{\alpha}_{ij} = \frac{\exp(\alpha_{ij})}{\sum_{j=1}^N \exp(\alpha_{ij})}, \quad (14)$$

We perform the dot product operation between the  $\hat{\alpha}_{ij}$  and the correlation coefficient matrix  $V$  to enhance the accuracy of the correlation information in the vector space and suppress the scale discrepancy problem. Finally, similarity selection operation is applied to the adjacent weight matrix  $A_{ij} \in \mathbb{R}^{N \times N}$ , and each node vector is selected with the  $d$ 's nearest neighbors

assigned value of 1, while all other values are set to 0, creating sparsity and enhancing the generalization ability of the network:

$$A_{ij} = \hat{\alpha}_{ij} \odot V, \quad (15)$$

$$A'_{ij} = \begin{cases} 1, & \text{if } A_{ij} \geq \varepsilon, \text{ and } j \text{ is the } d \text{ largest values in } A_i, \\ 0, & \text{otherwise} \end{cases} \quad (16)$$

Here,  $A_i$  is the  $i$ -th row of  $A_{ij}$ ,  $\varepsilon$  is a threshold value.  $A'_{ij}$  is the obtained multi-adjacent weight matrix.

Since the GAT layer can give more weight to features that are similar (Velickovic 2018), we use it to send the information in the node vectors. The GAT layer then aggregates the combined information from all nodes. The two inputs to the GAT layer are the feature vectors  $\mathbf{H} = \{\vec{h}_1, \dots, \vec{h}_N\} \in \mathbb{R}^{N \times L}$ ,  $\vec{h}_1 \in \mathbb{R}^L$  of the nodes in the graph and the constructed multi-adjacent weight matrix  $A'_{ij} \in \mathbb{R}^{N \times N}$ . The output of this layer is a new vector of node features  $\mathbf{H}' = \{\vec{h}_1', \dots, \vec{h}_N'\} \in \mathbb{R}^{N \times L'}, \vec{h}_1' \in \mathbb{R}^{L'}$ .

In order to extract deeper global similarity features, we add a learnable shared linear transform weight matrix  $\mathbf{W} \in L' \times L$  to each node, where we make  $L' = L$ , and then apply the shared self-attention mechanism  $a: \mathbb{R}^{L'} \times \mathbb{R}^L \rightarrow \mathbb{R}$  on the nodes to calculate the attention coefficients:

$$\begin{aligned} e_{xy} &= a(\mathbf{W}\vec{h}_x, \mathbf{W}\vec{h}_y) \\ &= \text{LeakyReLU}(\vec{\mathbf{a}}^T [\mathbf{W}\vec{h}_x \| \mathbf{W}\vec{h}_y]), \end{aligned} \quad (17)$$

The attention coefficient  $e_{xy}$  denotes the importance of node  $y$  to node  $x$ . We calculate only for  $y \in (0, N_x)$ , where  $N_x$  denotes the  $d$  first-order neighbor nodes of node  $x$  (including  $x$ ) in the graph. The attention mechanism  $a$  is a single-layer feedback neural network layer parameterized by  $\vec{\mathbf{a}}^T \in \mathbb{R}^{2C'}$ , after applying Leaky ReLU nonlinearity. To facilitate coefficient comparison across nodes, we normalize all  $y$ -selected coefficients using the softmax function:

$$\begin{aligned} \chi_{xy} &= \text{softmax}_y(e_{xy}) \\ &= \frac{\exp(\text{LeakyReLU}(\vec{\mathbf{a}}^T [\mathbf{W}\vec{h}_x \| \mathbf{W}\vec{h}_y]))}{\sum_{l \in N_x} \exp(\text{LeakyReLU}(\vec{\mathbf{a}}^T [\mathbf{W}\vec{h}_x \| \mathbf{W}\vec{h}_l]))}. \end{aligned} \quad (18)$$

The final output consists the updated HR features of each of the nodes generated by similar LR feature nodes:

$$\vec{h}_x' = \sigma \left( \sum_{y \in N_x} \chi_{xy} \mathbf{W}\vec{h}_y \right). \quad (19)$$

Using multi-head attention enhances the extraction of deep features, aligning with the number of GAT layers. The features from multiple attention mechanisms are cascaded after applying Equation (19), then the

features undergo averaging, and the ultimate HR features are obtained by applying non-linearity:

$$\vec{h}_x' = \sigma\left(\frac{1}{M} \sum_{m=1}^M \sum_{y \in N_x} \chi_{xy}^m \mathbf{W}^m \vec{h}_y\right), \quad (20)$$

where  $\chi_{xy}$ ,  $m$  and  $\mathbf{W}^m$  denote the normalized attention coefficients and the linear transformation weight matrix of the  $m$ th attention mechanism  $a^m$ .

### 3.5. Loss functions

To enhance the performance of the SR network, we propose a composite loss function that integrates two key components: the L1 norm loss and the spatial-spectral total variation (SSTV) loss. The L1 norm loss accelerates network convergence, while the SSTV loss simultaneously maintains high-quality spatial reconstruction and preserves inter-band spectral correlations. The final loss function is formulated as a weighted combination of these components:

$$\mathcal{L}(\Theta) = \mathcal{L}_1 + \alpha \mathcal{L}_{\text{SSTV}}, \quad (21)$$

In addition,  $\mathcal{L}_1$  and  $\mathcal{L}_{\text{SSTV}}$  are defined as

$$\mathcal{L}_1(\Theta) = \frac{1}{N} \sum_{n=1}^N \|Z^n - f_{\text{Net}}(X^n)\|_1, \quad (22)$$

$$\begin{aligned} \mathcal{L}_{\text{SSTV}}(\Theta) = & \frac{1}{N} \sum_{n=1}^N (\|\nabla_h Y^n\|_1 + \|\nabla_w Y^n\|_1 \\ & + \|\nabla_c Y^n\|_1), \end{aligned} \quad (23)$$

where,  $N$  represents the batch size of images used during training, and  $\Theta$  corresponds to the parameters set of the proposed SRLSGAT. The functions  $h$ ,  $w$  and  $c$  are utilized to calculate the horizontal, vertical, and spectral gradients of  $Y^n$  respectively.

## 4. Experiments

This section presents a comprehensive evaluation of the proposed network on multiple HSI datasets. First of all, we introduce the experimental datasets. Subsequently, the implementation details and evaluation metrics are outlined. Following this, parameter discussion and ablation studies are conducted to thoroughly examine the proposed network. Lastly, comparative results against several state-of-the-art methods are presented.

### 4.1. Datasets

In our work at Chikusei (Yokoya and Iwasaki 2016), Pavia Centre (Huang and Zhang 2009), and Washington DC, data sets are used to evaluate the SRLSGAT method. Since these datasets are acquired

using various hyperspectral imaging systems, they exhibit distinct attributes. This diversity enables independent training and testing for each dataset, a approach that differs significantly from the SR of natural images.

- (1) *Chikusei Dataset*: The remote sensing HSI dataset,<sup>1</sup> was captured using the Headwall Hyperspec-VNIR-C imaging sensor over agricultural and urban areas in Chikusei, Ibaraki, Japan. This dataset comprises 128 spectral bands covering a wavelength range of 363 nm to 1018 nm. The scene dimensions are  $2517 \times 2335$  pixels, with a ground sampling distance of 2.5 meters.
- (2) *Pavia Centre Dataset*: The Pavia Centre dataset,<sup>2</sup> a remote sensing HSI dataset, was captured by the ROSIS sensor during a flight campaign over northern Italy. The number of original spectral bands is 115, the dataset was preprocessed to remove water vapor absorption and noise bands, resulting in a final set of 102 bands. The image dimensions are  $1096 \times 715$  pixels image, with a geometric resolution of 1.3 meters.
- (3) *Washington DC Dataset*: The Washington DC Mall (WDC) dataset,<sup>3</sup> is an aerial HSI of the Washington Mall acquired by the Hydice sensor. It contains a total of 191 bands, covering the visible to near-infrared spectral (0. to 2.  $\mu\text{m}$ ), with a spatial resolution of  $1280 \times 307$  pixels. Feature categories include streets, grass, water, gravel paths, trees, shadows, and roofs.

### 4.2. Implementation Details

To augment the training samples, geometric transformations are implemented as part of the data augmentation pipeline. Specifically, each input image undergoes horizontal flipping followed by rotational transformations at  $90^\circ$ ,  $180^\circ$ , and  $270^\circ$  orientations. For the Chikusei dataset, edge regions with missing data are removed to maximize the retention of valid information. A subset of 31 spectral bands, specifically bands 50 to 80, is extracted to create a sub-image with dimensions of  $2304 \times 2048 \times 31$ . The upper section of the image is designated as the training set, with 10% of this portion reserved for validation, while the lower section is allocated as the test set. For the Pavia Centre dataset, the original image is utilized as the HR-HSI. A subset of 31 spectral bands, spanning bands 35 to 65, is selected to construct a sub-image with dimensions of  $1096 \times 715 \times 31$ . The left section of this sub-image is assigned as the training set, with 10% of the training data randomly allocated for validation, while the remaining portion

serves as the test set. The Washington DC dataset is processed by designating the original image as the HR-HSI. From this dataset, a subset of 31 contiguous spectral band between bands (81–111) is isolated to create a sub-image with size  $1280 \times 307 \times 31$ . The upper section of this sub-image is allocated as the training set, with 10% of these samples randomly partitioned for validation purposes. The remaining lower portion serves as the independent test set for performance evaluation. Three different HR-HSI patch types are created by cropping each HR-HSI into the three dimensions:  $32 \times 32$ ,  $64 \times 64$ , and  $128 \times 128$  with 50% overlap between adjacent patches (stride of 16, 32, and 6 pixels respectively). This HR-HSI patch is downsampled to generate LR-HSI patches with a resolution of  $16 \times 16$  for scaling factors of  $\times 2$ ,  $\times 4$ , and  $\times 8$ , respectively. There are  $512 \times 512$  and  $256 \times 256$  non-overlapping images in the test dataset. Our model is trained using the ADAM (Kingma and Ba 2014) optimizer with  $\beta_1 = 0.99$ ,  $\beta_2 = 0.999$ , and  $\epsilon = 10^{-8}$ . The network architecture employs 256 hidden units in all intermediate layers and processes inputs with a batch size of 16. According to the experimental findings, as shown in Figures 5a demonstrates that both training losses plateau after 70 epochs, confirming the optimal epoch setting. the model attain stable performance after 70 epochs. After 30 epochs, the learning rate dropped to 10 times from its initially setting of  $10^{-4}$ . Figure 5b demonstrates the variation of the loss value with different model parameter settings. The SRLSGAT network was implemented using PyTorch framework. All experiments were conducted on a workstation running Ubuntu 22.04.3 LTS, equipped with an Intel Core i9-13900KF CPU,

64GB of RAM, and an NVIDIA GeForce RTX 4090 GPU with 24GB of memory.

Five quantitative picture quality indices (PQIs) were used to evaluate the performance of SRLSGAT: peak signal-to-noise ratio (PSNR), structure similarity (SSIM) (Wang et al. 2004), spectral angle mapper (SAM) (Shuai et al. 2018), root mean squared error (RMSE), and universal image quality index (UIQI) (Z. Wang and Bovik 2002). For PSNR and SSIM, the mean values across all spectral bands were calculated. These five indices are widely recognized as standard metrics of assessing the reconstruction quality of the HSI.

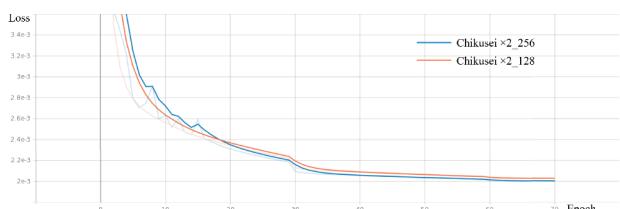
#### 4.3. Parameter discussion

To assess the influence of key parameters in the proposed SRLSGAT, we experiment on the Pavia Centre dataset with a  $sf$  of  $\times 4$  and 128 hidden layers to discuss the selection of important parameters. The key parameters include: (1) the number of convolutional layers  $PE$  in the patch embedding Encoder module, (2) the convolutional kernel size  $KS$  for feature map reshape, which corresponds to the step size  $S$ , (3) the number of adjacent nodes  $d$ , and (4) the number of multi-head attentions  $M$  in the GAT layer.

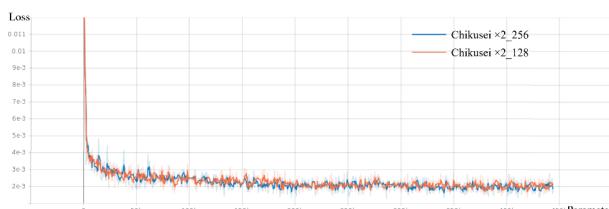
First, we discuss the choice of the number of convolutional layers  $PE$  in patch embedding. Testing values from 4 to 7 (Table 1) reveals that PSNR initially increases and then decreases with increases  $PE$ , peaking at  $PE = 6$ . Next, the selection of convolution kernel size  $KS$  and step size  $S$  is analyzed for the reshaping from pixel to image patch level. For  $KS/S$ , we choose  $3/1$ ,  $3/2$ ,  $5/2$ , and  $5/4$  for testing, as presented in Table 2, the experimental findings indicate that the network performs better when  $3/2$  is chosen. The primary factor lies in determining the optimal image patch size according to the size of the LR is necessary to support the quality of the network learning representation. Therefore, in the subsequent experiments, we set the  $KS/S$  value to  $3/2$ . For the adjacent nodes (including itself)  $d \in (4, 8)$  number, we set to  $d = 6$ , as shown in Table 3 experimental data. For the number of multi-head attentions  $M$  in the GAT layer, we test from 1 to 5, as shown in Table 4. Based on the comparative analysis, the highest PSNR value is achieved when the number of layers in  $M$  is assigned a value of 2.

#### 4.4. Ablation studies

The SRLSGAT architecture integrates five fundamental components: the spectral sub-network VH-BiLSTM, the spectral attention mechanism (SpeAM), the spatial sub-network MAW-GAT, the multi-adjacent weight matrix (QKV), and the spatial-



(a) The convergence process of the loss function with respect to the number of training epochs



(b) The variation of the loss value with different model parameter settings

**Figure 5.** Training and parameter analysis of the proposed SRLSGAT model.

**Table 1.** The study on the layers number of the patch embedding.

PE	4	5	6	7
PSNR ↑	33.7389	33.8103	33.8338	33.7732
SSIM ↑	0.8770	0.8785	0.8787	0.8776
SAM ↓	2.4884	2.4874	2.4867	2.4915
RMSE ↓	0.0232	0.0231	0.0230	0.0232
UIQI ↑	0.6375	0.6527	0.6527	0.6482

**Table 2.** Analysis of kernel size and stride in feature map reshape.

KS/S	3/1	3/2	5/2	5/4
PSNR ↑	33.7374	33.7705	33.7478	33.5748
SSIM ↑	0.8781	0.8786	0.8785	0.8721
SAM ↓	2.4893	2.4829	2.4853	2.4888
RMSE ↓	0.0232	0.0232	0.0232	0.0237
UIQI ↑	0.6244	0.6413	0.6300	0.6104

**Table 3.** The study on the number of the adjacent nodes.

d	4	5	6	7	8
PSNR ↑	33.7536	33.7720	33.7794	33.7593	33.7528
SSIM ↑	0.8779	0.8787	0.8784	0.8781	0.8779
SAM ↓	2.4891	2.4894	2.4823	2.4835	2.4826
RMSE ↓	0.0232	0.0232	0.0232	0.0232	0.0232
UIQI ↑	0.6400	0.6420	0.6437	0.6359	0.6323

**Table 4.** Impact of multi-head attention quantity in GAT.

M	1	2	3	4	5
PSNR ↑	33.7667	33.7792	33.7668	33.7704	33.7243
SSIM ↑	0.8787	0.8794	0.8783	0.8774	0.8773
SAM ↓	2.4865	2.4817	2.4859	2.4843	2.4864
RMSE ↓	0.0231	0.0231	0.0232	0.0232	0.0233
UIQI ↑	0.6389	0.6424	0.6438	0.6321	0.6179

spectral attention transpose module (SSAT). In this section, we evaluate the effects of individual components. This process is implemented by replacing different components with convolutional layers. The ablation study results are shown in [Table 5](#), these experimental were conducted on the Pavia Centre dataset, the  $sf$  was set to  $\times 4$ , and 128 hidden layers were used.

The images are input to parallel networks separately, enabling the modeling of feature map correlation in spectrally adjacent bands and spatial self-similarity features. To assess the efficacy of the spectral and spatial sub-networks, we substitute the sub-networks with convolution layers having an equivalent number of parameters, denoted as w/o VH-BiLSTM and w/o MAW-GAT, respectively. Additionally, in the spectral sub-network, we introduce SpeAM to capture remote dependencies of spectral sequences. To evaluate the effectiveness of SpeAM, we replace mechanism with a convolution layer having the same parameters

number, denoted as w/o SpeAM. In the spatial sub-network, to concentrate on capturing information related to the non-local self-similarity of image patches, the multi-adjacent weight matrix in the MAW-GAT is replaced by solely using Euclidean distance to image patch nodes, denoted as w/o QKV. In the Encoder-Decoder, a symmetric structure is employed to invert the feature map from the image patch level to the pixel level using the SSAT module. We further replace the spatial attention and channel attention with convolution layers, denoted as w/o SSAT. In addition, we introduce a convolution-only baseline model, which consists of a simple encoder-decoder architecture without any of the proposed modules. This baseline serves as a reference to evaluate the overall benefit brought by each component.

As shown in [Table 5](#), compared with the baseline, removing any single component results in varying

**Table 5.** Ablation studies quantitative comparisons among different components.

Model	PSNR↑	SSIM↑	SAM↓	RMSE↓	UIQI↑
w/o VH-BiLSTM	33.2762	0.8646	2.8332	0.0247	0.6223
w/o SpeAM	33.3843	0.8572	2.7689	0.0245	0.6287
w/o MAW-GAT	33.2912	0.8563	2.8214	0.0247	0.6255
w/o QKV	33.3755	0.8627	2.7744	0.0246	0.6341
w/o SSAT	33.3737	0.8644	2.7750	0.0246	0.6337
baseline	33.1524	0.8503	2.8502	0.0248	0.6186
Ours	33.4062	0.8731	2.7136	0.0242	0.6374

degrees of performance degradation. For instance, removing VH-BiLSTM leads to a PSNR drop from 33.4062 to 33.2762 and a UIQI drop from 0.6374 to 0.6223, highlighting the importance of remote spectral dependency modeling. Similarly, eliminating SpeAM results in a decline in SSIM from 0.8731 to 0.8572, demonstrating the effectiveness of the designed spectral attention mechanism in enhancing spectral fidelity. The MAW-GAT module also plays a critical role, as its removal causes a noticeable increase in SAM and a decrease in SSIM and UIQI, indicating that the global spatial context captured via the graph structure significantly contributes to spatial detail reconstruction. Most importantly, the baseline model, which lacks all five key components, performs the worst across all metrics PSNR (33.1524), SSIM (0.8503), SAM (2.8502), and UIQI (0.6186) underscoring the collective necessity of integrating both spectral and spatial priors into the reconstruction process. These results validate the design rationale of SRLSGAT and affirm the complementary benefits of its core modules.

To justify the necessity of the VH-BiLSTM module for spectral feature modeling, we further compare it with two alternative strategies: (1) a unidirectional LSTM (Uni-LSTM), which models spectral sequences in a single forward direction, and (2) a lightweight 1D self-attention module (1D-FAM), which lacks sequential recurrence and models pairwise spectral dependencies. As shown in Table 7, the unidirectional LSTM model achieves a PSNR of 33.2103, which is lower than the VH-BiLSTM variant (33.2762) and our full model (33.4062), indicating the limited capacity of one-directional encoding. Similarly, the attention-only model yields degraded SSIM and SAM values. These results demonstrate that the VH-BiLSTM design offers more robust spectral feature extraction by capturing richer inter-band dependencies.

#### 4.5. Comparisons with State-of-the-art methods

This section assesses the performance of the proposed SRLSGAT. Three widely used HSI datasets are utilized for evaluation. The assessment involves a comparison with seven existing SR methods, which are 3D-FCNN (Mei et al. 2017), GDRRN (Li et al. 2018), SSPSR (Jiang et al. 2020), MCNet (Li, Wang, and Li 2020), EUNet (Liu et al. 2023), MSDformer (Chen, Zhang, and Zhang 2023) and CLSCNet (Xu et al. 2024). All seven methods focus on super-resolution tasks with hyperspectral images. The proposed SRLSGAT is compared with the current SOTA methods. Detailed descriptions of these comparisons are provided below.

*1) Results on Chikusei Dataset:* Table 6 presents the comparative results of all methods, including five evaluation indices for different  $sf$  on the Chikusei dataset. The proposed SRLSGAT demonstrates superior performance across all evaluation indices compared to other methods. Specifically, the average PSNR values of SRLSGAT exceed those of the second-best method by 0.7020 dB ( $\times 2$ ), 0.5957 dB ( $\times 4$ ), and 0.7033 dB ( $\times 8$ ). Through comparison with existing SR methods, the proposed SRLSGAT exhibits significantly improved performance. The reason for this is that both SR methods fail to account for spectral band correlation and image patches recurrence. The proposed SRLSGAT also shows better performance than the seven HSI SR methods when compared to them. This performance may gain from SRLSGAT's enhanced feature extraction capability. Specifically, the spectral attention mechanism (SpeAM) improves feature discriminability, while non-local aggregation units capture remote dependencies. The 3D-FCNN performs the worst because of its shallow architecture (only three layers). In contrast, SSPSR benefits from deeper networks, yielding better results. MCNet and CLSCNet utilize 3D convolutions to capture spatial and spectral information, achieving strong reconstruction performance. However, their performance degrades with larger  $sf$ , unlike our algorithm, which effectively leverages spectral information across all scales.

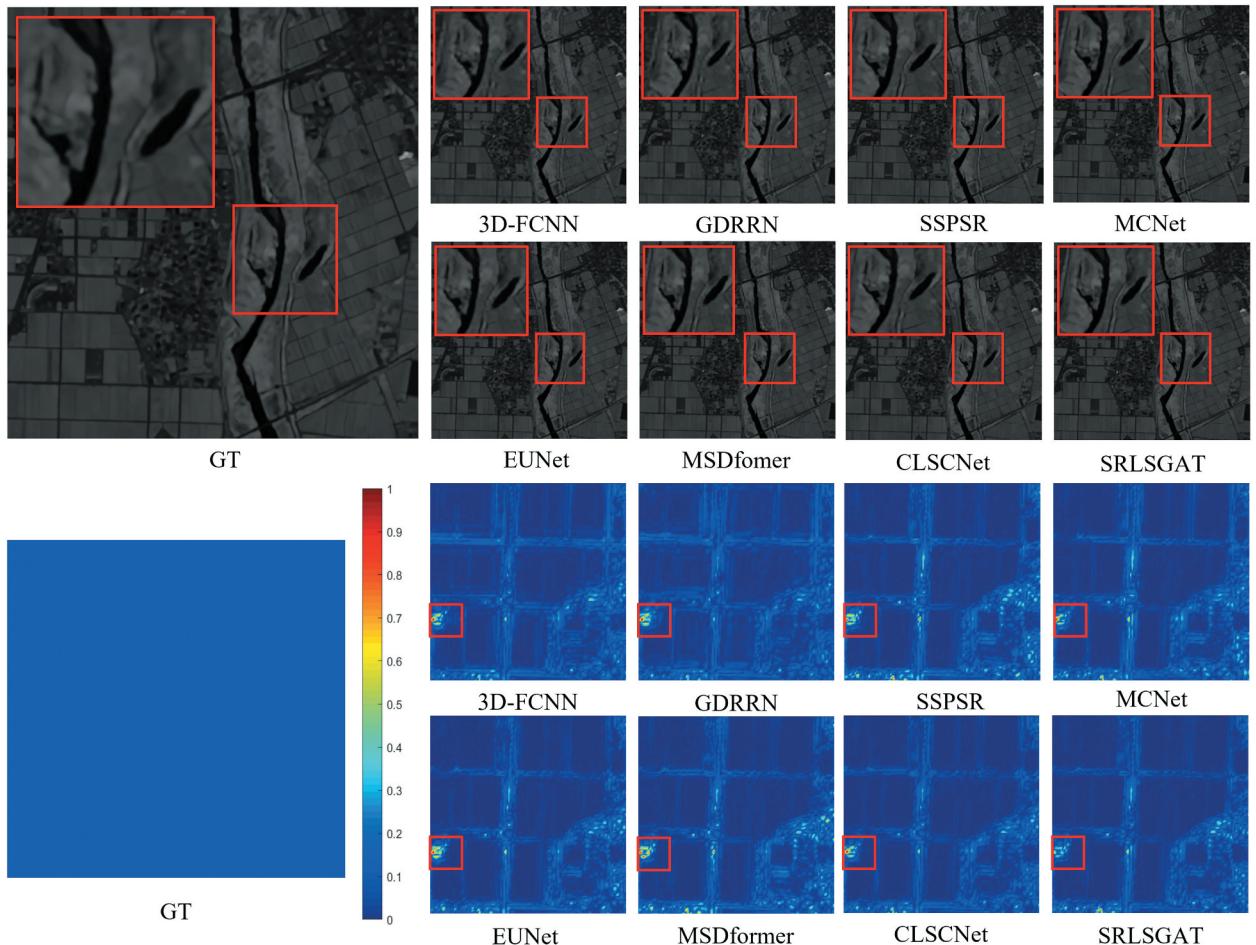
Figure 6 shows the results of visual quality comparisons for the Chikusei dataset for the 31st spectral with a  $sf$  of  $\times 4$ . The top row displays the reconstructed HSIs, while the bottom row shows their corresponding error maps. The proposed SRLSGAT generates sharper edges with minimal artifacts, as shown in the visual results. In the error maps, bluer regions indicate higher reconstruction accuracy, with all images normalized for comparison. In terms of reconstructing information between sequential spectral, SRLSGAT outperforms other methods in texture detail recovery, aligning with the quantitative results in Table 6.

PSNR and spectral curves were plotted to analyze spectral distortion in the reconstructed HSI, as illustrated in Figure 9a. The PSNR curves consistently show superior performance for our method. In the Reflectance curve plots, we randomly selected pixel positions (289, 340) to analyze the spectral distortion. While all methods produce spectral curves generally consistent with the original HR-HSI, the proposed SRLSGAT achieves the closest match. This confirms its superior ability to maintain spectral correlation during reconstruction.

*2) Results on Pavia Centre Dataset:* Table 8 presents the comparative results of all methods for the five indices across different  $sf$  on the Pavia Centre dataset.

**Table 6.** Qualitative computations of all methods on the Chikusei dataset across multiple sf. The red and blue annotations denote the top and second-best performances, respectively.

SF	Metrics	3D-FCNN	GDRRN	SSPSR	MCNet	EUNet	MSDformer	CLSCNet	SRLSGAT
x2	PSNR ↑	43.1938	43.3340	45.8551	44.5130	42.1486	45.5973	45.1618	46.5571
	SSIM ↑	0.9769	0.9777	0.9842	0.9783	0.9943	0.9837	0.9821	0.9910
	SAM ↓	1.3930	1.5410	0.8607	1.3124	0.7619	0.8586	1.0928	0.7589
	RMSE ↓	0.0093	0.0087	0.0073	0.0087	0.0048	0.0075	0.0079	0.0046
	UIQI ↑	0.9178	0.9005	0.9840	0.9392	0.9562	0.9473	0.9402	0.9878
	PSNR ↑	35.6606	35.0272	37.1766	36.1965	34.286	36.7182	36.7092	37.7723
	SSIM ↑	0.8781	0.8688	0.9067	0.8845	0.9634	0.9007	0.8897	0.9421
	SAM ↓	3.1375	3.8178	2.2235	3.1422	2.0537	2.2820	2.5575	1.9980
x	RMSE ↓	0.0205	0.0212	0.0178	0.0202	0.0119	0.0184	0.0186	0.0113
	UIQI ↑	0.6390	6.5961	0.9055	0.6858	0.7495	0.7170	0.7079	0.9182
	PSNR ↑	31.0946	30.7737	31.6472	31.6167	29.7302	31.656	30.9100	32.3597
	SSIM ↑	0.7340	0.7160	0.7620	0.7480	0.8900	0.7539	0.7248	0.8531
x8	SAM ↓	5.1712	6.1770	4.3424	4.9664	3.9812	4.3442	5.4301	3.8683
	RMSE ↓	0.0332	0.0332	0.0303	0.0311	0.0199	0.0307	0.0334	0.0188
x8	UIQI ↑	0.2758	0.2386	0.6936	0.3009	0.3509	0.3248	0.2525	0.7369



**Figure 6.** The SR results and corresponding error maps for Chikusei dataset at a sf of  $\times 4$ , focusing on the 31st spectral band. From left to right, the images represent the ground truth, followed by the result of 3D-FCNN, GDRRN, SSPSR, MCNet, EUNet, MSDformer, CLSCNet, and the proposed SRLSGAT method.

**Table 7.** Ablation studies on different spectral modeling approaches, including the proposed VH-BiLSTM, uni-LSTM, and 1D-FAM.

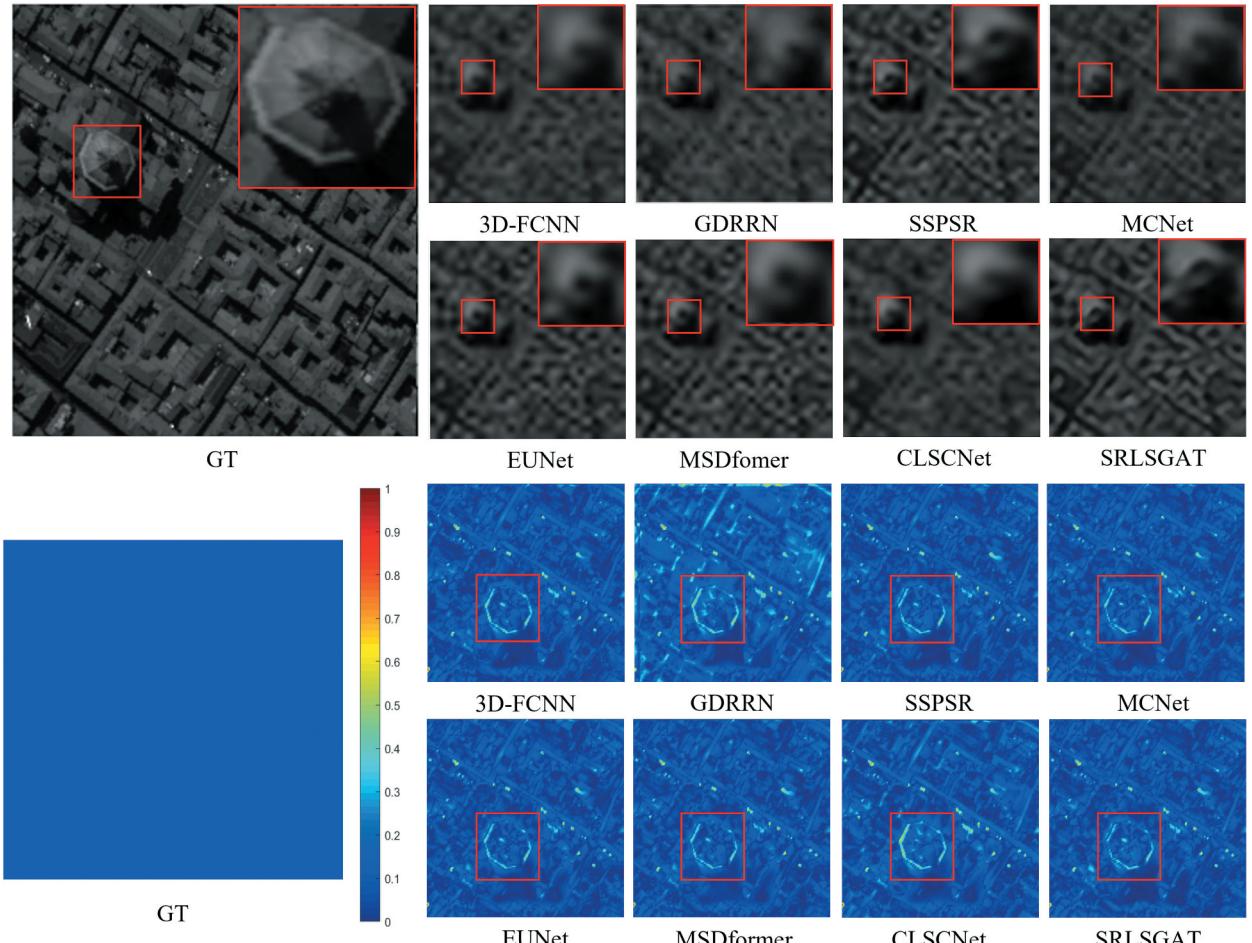
Model	PSNR↑	SSIM↑	SAM↓	RMSE↓	UIQI↑
w/o VH-BiLSTM	33.2762	0.8646	2.8332	0.0247	0.6223
Uni-LSTM	33.2103	0.8614	2.8647	0.0248	0.6201
1D-FAM	33.1858	0.8594	2.8732	0.0248	0.6189
Ours	33.4062	0.8731	2.7136	0.0242	0.6374

The proposed SRLSGAT outperforms other methods across all evaluation metrics. Compared to the second-best method, it achieves PSNR improvements of 0.1261 dB, 0.8571 dB, and 0.1636 dB for  $\times 2$ ,  $\times 4$ , and  $\times 8$ , respectively. Figure 7 shows that for the  $sf \times 8$ , the Pavia Centre dataset is in 31st spectral bands for comparison of the visual quality results. From this figure, we can see that the proposed SRLSGAT can reconstruct textural details such as building edges and

street trends very well, whereas the reconstructions of the compared methods are blurred. In Figure 9b, the PSNR curves demonstrate our performance in each spectral band, which is all better than the other methods. In the spectral curves, we randomly choose pixel points (189, 98) to display the spectral curves, the spectral curves of all methods generally align with those of the real topographic image, with the proposed SRLSGAT demonstrating the highest spectral fidelity.

**Table 8.** Qualitative computations of all methods on the Pavia Centre dataset across multiple  $sf$ . The red and blue annotations denote the top and second-best performances, respectively.

SF	Metrics	3D-FCNN	GDRRN	SSPSR	MCNet	EUNet	MSDformer	CLSCNet	SRLSGAT
$\times 2$	PSNR $\uparrow$	33.9334	33.2832	35.3763	34.5509	34.7863	35.8553	35.1508	35.9814
	SSIM $\uparrow$	0.9493	0.9435	0.9630	0.9560	0.9801	0.9630	0.9619	0.9644
	SAM $\downarrow$	2.1237	5.2872	2.0063	2.5115	1.8378	1.9969	2.2262	1.9908
	RMSE $\downarrow$	0.0210	0.0225	0.0178	0.0196	0.0183	0.0179	0.0183	0.0178
	UIQI $\uparrow$	0.9398	0.9327	0.9755	0.9481	0.9559	0.9604	0.9544	0.9790
	PSNR $\uparrow$	27.6996	26.6684	28.4201	27.7186	28.0471	28.1628	27.3117	29.2772
$\times 4$	SSIM $\uparrow$	0.7712	0.7220	0.8067	0.7713	0.8979	0.7922	0.7446	0.8088
	SAM $\downarrow$	3.1084	8.5396	2.8135	4.2911	2.6161	2.6200	4.0261	2.7688
	RMSE $\downarrow$	0.0418	0.0469	0.0384	0.0417	0.0398	0.0396	0.0436	0.0387
	UIQI $\uparrow$	0.7340	0.6762	0.8848	0.7346	0.7585	0.7573	0.6999	0.8873
	PSNR $\uparrow$	24.2821	22.8242	24.5491	24.4944	24.2442	24.4656	23.7146	24.7127
	SSIM $\uparrow$	0.4774	0.4131	0.5090	0.5029	0.7152	0.4913	0.4396	0.4999
$\times 8$	SAM $\downarrow$	4.1506	11.7500	3.6996	4.8353	3.1962	3.6231	6.6617	3.5250
	RMSE $\downarrow$	0.0616	0.0727	0.0598	0.0601	0.0619	0.0603	0.0657	0.0614
	UIQI $\uparrow$	0.3471	0.2698	0.6443	0.3874	0.3717	0.3858	0.2944	0.6542



**Figure 7.** The SR results and corresponding error maps for Pavia Centre dataset at a  $sf$  of  $\times 8$ , focusing on the 31st spectral band. From left to right, the images represent the ground truth, followed by the result of 3D-FCNN, GDRRN, SSPSR, MCNet, EUNet, MSDformer, CLSCNet and the proposed SRLSGAT method.

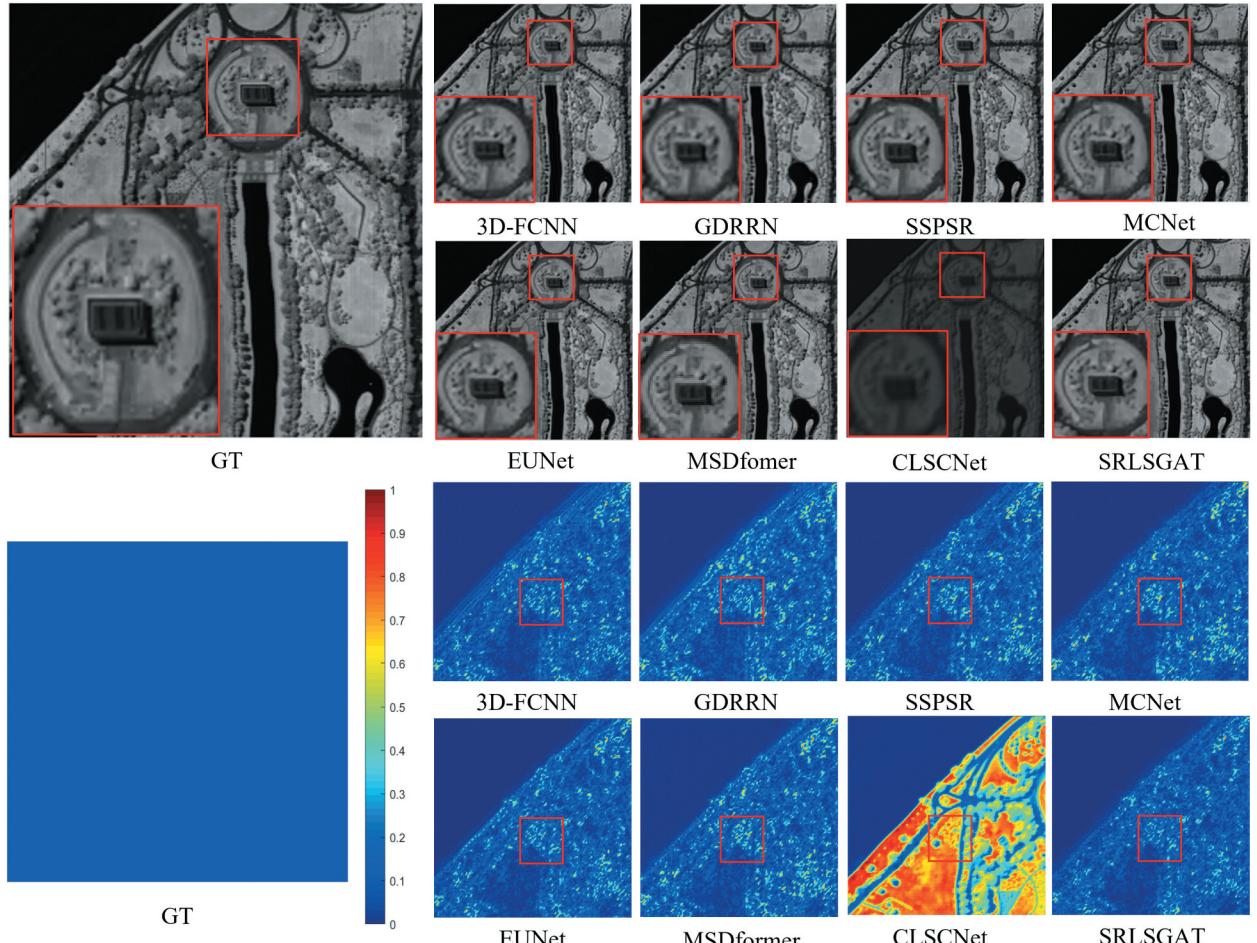
However, all methods exhibit significant errors on the real topographic image, which may be attributed to the limited training samples in the Pavia Centre dataset and the insufficient generalization ability of the trained models.

3) *Results on Washington DC Dataset:* Table 9 presents the average performance of all comparison methods on the WDC dataset for different  $sf$ . For  $sf$  of  $\times 2$ ,  $\times 4$  and  $\times 8$ , the proposed SRLSGAT

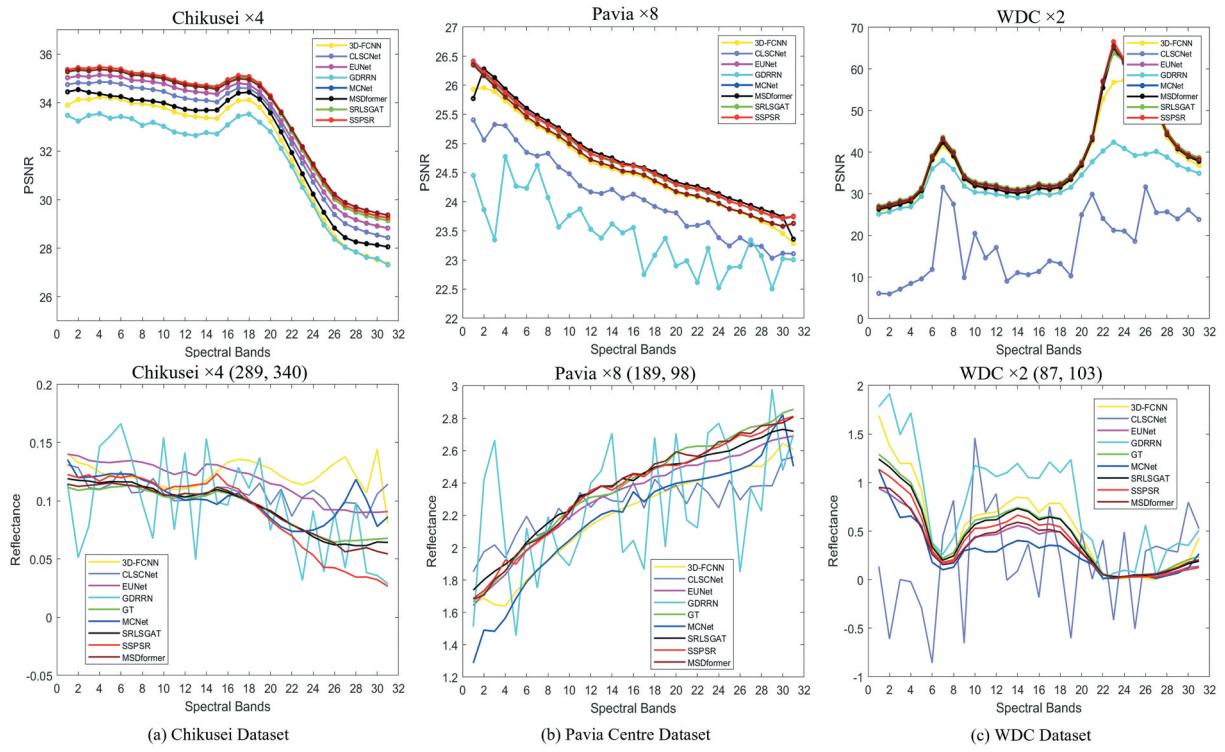
achieved average PSNR improvements of 0.1364 dB, 0.0894 dB, and 0.3997 dB, respectively, compared to the second-best method. Our SRLSGAT performs best in the index. For visual comparison, We selected the 4st spectral bands of test images in WDC dataset. Figure 8 presents the reconstructed HSI and their corresponding error maps obtained on all methods for  $sf \times 2$ . The proposed SRLSGAT excels in reconstructing fine structural details from

**Table 9.** Qualitative computations of all methods on the Washington DC dataset across multiple  $sf$ . The red and blue annotations denote the top and second-best performances, respectively.

SF	Metrics	3D-FCNN	GDRRN	SSPSR	MCNet	EUNet	MSDformer	CLSCNet	SRLSGAT
$\times 2$	PSNR $\uparrow$	37.9558	34.2548	40.1437	39.5587	26.4446	40.1574	38.5200	40.2948
	SSIM $\uparrow$	0.9384	0.8905	0.9500	0.9441	0.9860	0.9511	0.9431	0.9526
	SAM $\downarrow$	3.3276	5.9053	1.7643	2.8434	1.1991	1.6809	2.9203	1.6509
	RMSE $\downarrow$	0.0248	0.0288	0.0223	0.0238	0.0229	0.0215	0.0239	0.0223
	UIQI $\uparrow$	0.8011	9.6624	0.9319	0.8326	0.8470	0.8350	0.8059	0.9555
	PSNR $\uparrow$	31.7111	26.1858	33.0104	32.6834	20.7003	32.8204	31.7011	33.0998
	SSIM $\uparrow$	0.7785	0.6336	0.8057	0.7958	0.9433	0.7920	0.7582	0.8108
	SAM $\downarrow$	6.4466	11.4115	4.9144	5.4293	2.1366	4.8412	8.1270	4.7103
$\times 4$	RMSE $\downarrow$	0.0473	0.0593	0.0449	0.0461	0.0452	0.0460	0.0491	0.0438
	UIQI $\uparrow$	0.5423	0.3611	0.8287	0.5649	0.5629	0.5538	0.5039	0.8389
	PSNR $\uparrow$	28.0196	21.3299	29.3598	28.9051	17.8079	29.0585	27.5855	29.7595
	SSIM $\uparrow$	0.5777	0.4088	0.5978	0.5935	0.8792	0.5641	0.5055	0.6105
$\times 8$	SAM $\downarrow$	10.3972	22.1217	9.6524	9.8961	3.0354	9.7055	20.4793	9.5717
	RMSE $\downarrow$	0.0682	0.1059	0.0653	0.0674	0.0632	0.0677	0.0696	0.0625
	UIQI $\uparrow$	0.2187	0.1314	0.6561	0.2398	0.2419	0.2211	0.1976	0.6497



**Figure 8.** The SR results and corresponding error maps for Washington DC dataset at a  $sf$  of  $\times 2$ , focusing on the 4st spectral band. From left to right, the images represent the ground truth, followed by the result of 3D-FCNN, GDRRN, SSPSR, MCNet, EUNet, MSDformer, CLSCNet and the proposed SRLSGAT method.



**Figure 9.** PSNR curve and spectral curve on randomly selected pixel position (289,340), (189,98), (87,103) value of three datasets (Chikusei x4, Pavia Centre x8 and Washington DC x2) respectively.

the original images, as evidenced by the visual results. Additionally, the error maps indicate that SRLSGAT outperforms other methods in texture detail preservation. Specific features, such as the curved roof edges and grassy areas, are reconstructed with high accuracy.

Furthermore, spectral curves were generated for randomly selected pixel locations (87, 103) in the reconstructed images. These curves are displayed in Figure 9c, illustrated the reconstruction efficacy of the spectral information. The spectral curves generated by the proposed SRLSGAT closely match to those real ground truth images than those of other comparative methods, and the area magnification graphs confirm the experimental findings. This indicates that SRLSGAT better preserves the original image's spectral correlations. Additionally, the PSNR curve graph confirms our algorithm's superior performance across most scenarios.

#### 4.6. Complexity analysis of the model

This subsection, we evaluate the efficiency of the proposed SRLSGAT using three widely adopted metrics: Model Parameters, Floating-Point Operations (FLOPs), and Memory Access Cost (MAC). The experimental results obtained for the seven SOTA methods are showed in Table 10 to evaluate model on the Pavia Centre dataset with a *sf* ( $\times 8$ ) and 6 hidden layers.

First, we analyze the Model Parameters count. The SRLSGAT requires fewer parameters than SSPSR, MCNet, EUNet, MSDformer, and CLSCNet, but more than 3D-FCNN and GDRRN. This can be attributed to the fact that the last two comparison methods utilize fewer network layers than the proposed approach. Subsequently, FLOPs were analyzed to further evaluate computational efficiency. The proposed SRLSGAT performs better than SSPSR,

**Table 10.** Complexity comparison of different methods.

SF	Method	Para(K)	FLOP(M)	MAC(M)
	3D-FCNN	41.77	302.61	4700
	GDRRN	109.44	179.01	3508
	SSPSR	1825.64	7635.57	2270
	MCNet	2960.51	638656.30	7372
	EUNet	1547.79	3945.37	2135
	MSDformer	1458.47	1589.19	3234
	CLSCNet	2337.64	3673.20	2646
x8	Ours	1435.63	2151.30	1131

MCNet, EUNet, CLSCNet and is higher than the other three methods compared. It is probable that the proposed SRLSGAT transforms between the pixel level and image patch level, which significantly increases the computational operations. Finally, the MAC was analyzed to assess memory efficacy, we found that SRLSGAT has a greater memory advantage than all the other methods. The reason for this is that SRLSGAT uses the image patch level to capture spectral-spatial information, reducing memory costs.

## 5. Conclusions

In this study, a deep joint neural network based on Spectral-Spatial Strategy is introduced to solve the problem of single HSI SR. In particular, for the problem of small samples in HSI, we propose Vertical-Horizontal Bi-directional Long Short-Term Memory Network (VH-BiLSTM) and Multi-Adjacent Weight Matrix Graph Attention Network (MAW-GAT). By applying LSTM to capture remote dependencies between spectral and utilizing graph neural networks to learn the inherent relationships of small patch pairs, the parameters of the model can be greatly reduced, and it is possible to obtain stable training results under small data conditions. Experimental results show that the proposed SRLSGAT obtains comparative performance in comparison with the SOTA method.

Even though our model outperforms SOTA methods, three main limitations of the current approach should be noted: (1) The avoidance of cross-scale recursive iteration of image patches, implemented to maintain network simplicity, may limit performance on multi-scale features. Future work will investigate multi-scale strategy integration to address this constraint. (2) The model's performance remains dependent on training data quantity and quality. We propose to mitigate this through self-supervised learning techniques that leverage unlabeled data. (3) The limited sample size in HSI datasets frequently induces underfitting or overfitting. We will explore enhanced utilization of inherent image structural information to improve reconstruction robustness. Future work will investigate these aspects while maintaining the model's current strengths in spectral fidelity.

## Notes

1. <https://www.sal.t.u-tokyo.ac.jp/hyperdata/>
2. <http://www.ehu.eus/ccwintco/index.php>
3. <https://engineering.purdue.edu/biehl/MultiSpec/>

## Disclosure statement

No potential conflict of interest was reported by the author(s).

## Funding

This research was funded by the National Key Research and Development Program of China [grant number 2023YFD2201702].

## Notes on contributors

***Yixin Dong*** received the M.S. degree in Computer Science and Technology from University of Shanghai for Science and Technology, Shanghai, China in 2023, and is currently studying the Ph.D. degree in State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University. Her research interests focus on computer vision, image processing, and artificial intelligence.

***Bo Yang*** received the PhD degree in Photogrammetry and Remote Sensing from Wuhan University in 2014. He is an associate professor in the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, China. His research interests include satellite photogrammetry and remote sensing, computer vision, AI+RS, and 3D reconstruction.

***Cong Liu*** received the PhD degree in Computer Application Technology from East China Normal University in 2013. He is an associate professor at the School of Optical-Electrical and Computer Engineering, University of Shanghai for Science and Technology, China. His research interests include image processing, hyperspectral remote sensing image processing, machine learning, interpretable deep learning, intelligent optimization, and convex optimization.

***Zemin Geng*** He received the M.S. degree from Wuhan University in 2023 and is currently studying for the Ph.D. degree at Wuhan University State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing. His current research interests include satellite photogrammetry and remote sensing, satellite remote sensing image processing.

***Taiping Wang*** received the B.S. and M.S. degrees in Photogrammetry and Remote Sensing from Wuhan University in 2018 and 2021, respectively. He is currently a PhD student at State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University. His research interests include satellite photogrammetry and remote sensing, satellite remote sensing image processing.

## ORCID

Yixin Dong  <http://orcid.org/0009-0008-3976-7811>  
 Bo Yang  <http://orcid.org/0009-0000-4217-9174>  
 Cong Liu  <http://orcid.org/0000-0002-3034-4504>  
 Zemin Geng  <http://orcid.org/0009-0004-0090-7048>  
 Taiping Wang  <http://orcid.org/0009-0004-7888-3859>

## Data availability statement

The Chikusei, Pavia Centre and Washington DC dataset were downloaded from <https://www.sal.t.u-tokyo.ac.jp/hyperdata/>, <http://www.ehu.eus/ccwintco/index.php>, and <https://engineering.purdue.edu/biehl/MultiSpec/>, respectively. The code and experimental results used in this

study are available in the Github at <https://github.com/Dongyxl128/SRLSGAT>.

## References

- Akhtar, N., F. Shafait, and A. Mian. 2015. "Bayesian Sparse Representation for Hyperspectral Image Super Resolution." In *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Boston, MA, USA, 3631–3640. <https://doi.org/10.1109/CVPR.2015.7298986>.
- Akhtarmanesh, A., D. Abbasi-Moghadam, A. Sharifi, M. H. Yadkouri, A. Tariq, and L. Lu. 2024. "Road Extraction from Satellite Images Using Attention-Assisted Unet." *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 17:1126–1136. <https://doi.org/10.1109/JSTARS.2023.3336924>.
- Arun, P. V., I. Herrmann, K. M. Budhiraju, and A. Karnieli. 2019. "Convolutional Network Architectures for Super-Resolution/Sub-Pixel Mapping of Drone-Derived Images." *Pattern Recognit* 88:431–446. <https://doi.org/10.1016/j.patcog.2018.11.033>.
- Bu, Y., Y. Zhao, J. Xue, J. C.-W. Chan, S. G. Kong, C. Yi, J. Wen, and B. Wang. 2021. "Hyperspectral and Multispectral Image Fusion via Graph Laplacian-Guided Coupled Tensor Decomposition." *IEEE Transactions on Geoscience and Remote Sensing* 59 (1): 648–662. <https://doi.org/10.1109/TGRS.2020.2992788>.
- Chen, S., L. Zhang, and L. Zhang. 2023. "Msdformer: Multiscale Deformable Transformer for Hyperspectral Image Super-Resolution." *IEEE Transactions on Geoscience and Remote Sensing* 61:1–14. <https://doi.org/10.1109/TGRS.2023.3315970>.
- Dian, R., L. Fang, and S. Li. 2017. "Hyperspectral Image Super-Resolution via Non-Local Sparse Tensor Factorization." In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI, USA, 3862–3871.
- Dian, R., and S. Li. 2019. "Hyperspectral Image Super-Resolution via Subspace-Based Low Tensor Multi-Rank Regularization." *IEEE Transactions on Image Processing* 28 (10): 5135–5146. <https://doi.org/10.1109/TIP.2019.2916734>.
- Dian, R., S. Li, and L. Fang. 2019. "Learning a Low Tensor-Train Rank Representation for Hyperspectral Image Super-Resolution." *IEEE Transactions on Neural Networks and Learning Systems* 30 (9): 2672–2683. <https://doi.org/10.1109/TNNLS.2018.2885616>.
- Dong, C., C. C. Loy, K. He, and X. Tang. 2016. "Image Super-Resolution Using Deep Convolutional Networks." *IEEE Transactions on Pattern Analysis and Machine Intelligence* 38 (2): 295–307. <https://doi.org/10.1109/TPAMI.2015.2439281>.
- Dong, W., F. Fu, G. Shi, X. Cao, J. Wu, G. Li, and X. Li. 2016. "Hyperspectral Image Super-Resolution via Non-Negative Structured Sparse Representation." *IEEE Transactions on Image Processing* 25 (5): 2337–2352. <https://doi.org/10.1109/TIP.2016.2542360>.
- Esmaeili, M., D. Abbasi-Moghadam, A. Sharifi, A. Tariq, and Q. Li. 2024. "Resmorcnn Model: Hyperspectral Images Classification Using Residual-Injection Morphological Features and 3dcnn Layers." *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 17:219–243. <https://doi.org/10.1109/JSTARS.2023.3328389>.
- Farmonov, N., M. Esmaeili, D. Abbasi-Moghadam, A. Sharifi, K. Amankulova, and L. Mucsi. 2024. "Hypsldnet: 3-d-2-d Cnn Model and Spatial-Spectral Morphological Attention for Crop Classification with Desis and Lidar Data." *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 17: 11 969–11 996. <https://doi.org/10.1109/JSTARS.2024.3418854>.
- Felegari, S., A. Sharifi, M. Khosravi, and S. Sabanov. 2023. "Using Experimental Models and Multitemporal Landsat-9 Images for Cadmium Concentration Mapping." *IEEE Geoscience and Remote Sensing Letters* 20:1–4. <https://doi.org/10.1109/LGRS.2023.3291019>.
- Gao, L., J. Li, K. Zheng, and X. Jia. 2023. "Enhanced Autoencoders with Attention-Embedded Degradation Learning for Unsupervised Hyperspectral Image Super-Resolution." *IEEE Transactions on Geoscience and Remote Sensing* 61:1–17. <https://doi.org/10.1109/TGRS.2023.3267890>.
- Gori, M., G. Monfardini, and F. Scarselli. 2005. "A New Model for Learning in Graph Domains." *Proceedings. 2005 IEEE International Joint Conference on Neural Networks*, Montreal, Canada, 2, 729–73.
- Han, X., J. Yu, J. Xue, and W. Sun. 2020. "Hyperspectral and Multispectral Image Fusion Using Optimized Twin Dictionaries." *IEEE Transactions on Image Processing* 29:4709–4720. <https://doi.org/10.1109/TIP.2020.2968773>.
- He, S., H. Zhou, Y. Wang, W. Cao, and Z. Han. 2016. "Super-Resolution Reconstruction of Hyperspectral Images via Low Rank Tensor Modeling and Total Variation Regularization." *2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, Beijing, China, 6962–6965. <https://doi.org/10.1109/IGARSS.2016.7730816>.
- Hochreiter, S., and J. Schmidhuber. 1997. "Long Short-Term Memory." *Neural Computation* 9 (8): 1735–1780. <https://doi.org/10.1162/neco.1997.9.8.1735>.
- Hong, D., J. Yao, D. Meng, N. Yokoya, and J. Chanussot. "Decoupled-and-Coupled Networks: Self-Supervised Hyperspectral Image Super-Resolution with Subpixel Fusion" [Online]: <https://arxiv.org/abs/2205.03742.2022>
- Huang, X., and L. Zhang. 2009. "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. <https://doi.org/10.1080/01431160802559046>.
- Jiang, J., H. Sun, X. Liu, and J. Ma. 2020. "Learning Spatial-Spectral Prior for Super-Resolution of Hyperspectral Imagery." *IEEE Transactions on Computational Imaging* 6:1082–1096. <https://doi.org/10.1109/TCI.2020.2996075>.
- Kingma, D. P., and J. Ba. 2014. "Adam: A Method for Stochastic Optimization." *arXiv* 1412.6980. <https://arxiv.org/abs/1412.6980>.
- Lai, W.-S., J.-B. Huang, N. Ahuja, and M.-H. Yang. 2017. "Deep Laplacian Pyramid Networks for Fast and Accurate Super-Resolution." *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI, USA, 5835–5843.
- Li, J., Q. Yuan, H. Shen, X. Meng, and L. Zhang. 2016. "Hyperspectral Image Super-Resolution by Spectral Mixture Analysis and Spatial-Spectral Group Sparsity." *IEEE Geoscience and Remote Sensing Letters* 13 (9): 1250–1254. <https://doi.org/10.1109/LGRS.2016.2579661>.
- Li, J., K. Zheng, L. Gao, Z. Han, J. Li, Z. Chanussot, and J. Gao. 2025. "Enhanced Deep Image Prior for Unsupervised Hyperspectral Image Super-Resolution."

- IEEE Transactions on Geoscience and Remote Sensing* 63:1–18. <https://doi.org/10.1109/TGRS.2025.3577562>.
- Li, Q., M. Gong, Y. Yuan, and Q. Wang. 2022. “Symmetrical Feature Propagation Network for Hyperspectral Image Super-Resolution.” *IEEE Transactions on Geoscience and Remote Sensing* 60:1–12. <https://doi.org/10.1109/TGRS.2022.3203749>.
- Li, Q., M. Gong, Y. Yuan, and Q. Wang. 2023. “Rgb-Induced Feature Modulation Network for Hyperspectral Image Super-Resolution.” *IEEE Transactions on Geoscience and Remote Sensing* 61:1–11. <https://doi.org/10.1109/TGRS.2023.3277486>.
- Li, Q., Q. Wang, and X. Li. 2020. “Mixed 2d/3d Convolutional Network for Hyperspectral Image Super-Resolution.” *Remote Sensing* 12 (10): 1660. <https://doi.org/10.3390/rs12101660>.
- Li, Q., Q. Wang, and X. Li. 2021. “Exploring the Relationship Between 2d/3d Convolution for Hyperspectral Image Super-Resolution.” *IEEE Transactions on Geoscience and Remote Sensing* 59 (10): 8693–8703. <https://doi.org/10.1109/TGRS.2020.3047363>.
- Li, Q., Y. Yuan, X. Jia, and Q. Wang. 2022. “Dual-Stage Approach Toward Hyperspectral Image Super-Resolution.” *IEEE Transactions on Image Processing* 31:7252–7263. <https://doi.org/10.1109/TIP.2022.3221287>.
- Li, S., R. Dian, L. Fang, and J. M. Bioucas-Dias. 2018. “Fusing Hyperspectral and Multispectral Images via Coupled Sparse Tensor Factorization.” *IEEE Transactions on Image Processing* 27 (8): 4118–4130. <https://doi.org/10.1109/TIP.2018.2836307>.
- Li, Y., L. Zhang, C. Dingl, W. Wei, and Y. Zhang. 2018. “Single Hyperspectral Image Super-Resolution with Grouped Deep Recursive Residual Network.” *2018 IEEE Fourth International Conference on Multimedia Big Data (BigMM)*, Beijing, China, 1–4.
- Lim, B., S. Son, H. Kim, S. Nah, and K. M. Lee. 2017. “Enhanced Deep Residual Networks for Single Image Super-Resolution.” *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, Honolulu, HI, USA, 1132–1140.
- Liu, C., and Y. Dong. 2022. “Cnn-Enhanced Graph Attention Network for Hyperspectral Image Super-Resolution Using Non-Local Self-Similarity.” *International Journal of Remote Sensing* 43 (13): 4810–4835. <https://doi.org/10.1080/01431161.2022.2121188>.
- Liu, C., Z. Fan, and G. Zhang. 2022. “Gjtd-lr: A Trainable Grouped Joint Tensor Dictionary with Low-Rank Prior for Single Hyperspectral Image Super-Resolution.” *IEEE Transactions on Geoscience and Remote Sensing* 60:1–17. <https://doi.org/10.1109/TGRS.2022.3204049>.
- Liu, C., and P. Lei. 2021. “An Efficient Group Skip-Connecting Network for Image Super-Resolution.” *Knowledge-Based Systems* 222:107017. <https://doi.org/10.1016/j.knosys.2021.107017>.
- Liu, D., J. Li, Q. Yuan, L. Zheng, J. He, S. Zhao, and Y. Xiao. 2023. “An Efficient Unfolding Network with Disentangled Spatial-Spectral Representation for Hyperspectral Image Super-Resolution.” *Information Fusion* 94:92–111. <https://doi.org/10.1016/j.inffus.2023.01.018>.
- Liu, D., B. Wen, Y. Fan, C. C. Loy, and T. S. Huang. 2018. “Non-Local Recurrent Network for Image Restoration.” *Proceedings of the 32nd International Conference on Neural Information Processing Systems*, ser. NIPS’18. 1em plus 0.5em minus 0.4em Red Hook, NY, USA: 1680–1689. Curran Associates Inc.
- Liu, K., H. Su, and X. Li. 2016. “Estimating High-Resolution Urban Surface Temperature Using a Hyperspectral Thermal Mixing (Htm) Approach.” *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 9 (2): 804–815. <https://doi.org/10.1109/JSTARS.2015.2459375>.
- Mahdipour, H., A. Sharifi, M. Sookhak, and C. R. Medrano. 2024. “Ultrafusion: Optimal Fuzzy Fusion in Land-Cover Segmentation Using Multiple Panchromatic Satellite Images.” *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 17:5721–5733. <https://doi.org/10.1109/JSTARS.2024.3360648>.
- Marzvan, S., K. Moravej, S. Felegari, A. Sharifi, and M. S. Askari. 2021. “Risk Assessment of Alien Azolla Filiculoides Lam in Anzali Lagoon Using Remote Sensing Imagery.” *Journal of the Indian Society of Remote Sensing* 49 (8): 1801–1809. Aug. <https://doi.org/10.1007/s12524-021-01362-1>.
- Mei, S., X. Yuan, J. Ji, Y. Zhang, S. Wan, and Q. Du. 2017. “Hyperspectral Image Spatial Super-Resolution via 3d Full Convolutional Neural Network.” *Remote Sensing* 9 (11): 1139. <https://doi.org/10.3390/rs9111139>.
- Mirhoseini Nejad, S. M., D. Abbasi-Moghadam, and A. Sharifi. 2024. “ConvLstm-Vit: A Deep Neural Network for Crop Yield Prediction Using Earth Observations and Remotely Sensed Data.” *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 17: 17 489–17 502. <https://doi.org/10.1109/JSTARS.2024.3464411>.
- Pike, R., G. Lu, D. Wang, Z. G. Chen, and B. Fei. 2016. “A Minimum Spanning Forest-Based Method for Noninvasive Cancer Detection with Hyperspectral Imaging.” *IEEE Transactions on Biomedical Engineering* 63 (3): 653–663. <https://doi.org/10.1109/TBME.2015.2468578>.
- Rasti, P., T. Uiboupin, S. Escalera, and G. Anbarjafari. 2016. “Convolutional Neural Network Super Resolution for Face Recognition in Surveillance Monitoring.” In *Articulated Motion and Deformable Objects*, edited by F. J. Perales and J. Kittler, 175–184. Cham: Springer International Publishing.
- Safari, M. M., A. Sharifi, J. Mahmoodi, and D. Abbasi-Moghadam. 2024. “Mesoscale Eddy Detection and Classification from Sea Surface Temperature Maps with Deep Neural Networks.” *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 17: 10 279–10 290. <https://doi.org/10.1109/JSTARS.2024.3402823>.
- Sharifi, A., and M. M. Safari. 2025. “Enhancing the Spatial Resolution of Sentinel-2 Images Through Super-Resolution Using Transformer-Based Deep-Learning Models.” *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 18:4805–4820. <https://doi.org/10.1109/JSTARS.2025.3526260>.
- Shuai, Y., Y. Wang, Y. Peng, and Y. Xia. 2018. “Accurate Image Super-Resolution Using Cascaded Multi-Column Convolutional Neural Networks.” *2018 IEEE International Conference on Multimedia and Expo (ICME)*, San Diego, CA, USA, 1–6.
- Sun, W., K. Ren, X. Meng, C. Xiao, G. Yang, and J. Peng. 2021. “A Band Divide-and-Conquer Multispectral and Hyperspectral Image Fusion Method.” *IEEE Transactions on Geoscience & Remote Sensing* 60:1–13. <https://doi.org/10.1109/TGRS.2020.3046321>.

- Tong, T., G. Li, X. Liu, and Q. Gao. 2017. "Image Super-Resolution Using Dense Skip Connections." *2017 IEEE International Conference on Computer Vision (ICCV)*, Venice, Italy, 4809–4817.
- Vafaeinejad, A., N. Alimohammadi, A. Sharifi, and M. M. Safari. 2025. "Super-Resolution Ai-Based Approach for Extracting Agricultural Cadastral Maps: Form and Content Validation." *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 18:5204–5216. <https://doi.org/10.1109/JSTARS.2025.3530714>.
- Velickovic, P., G. Cucurull, A. Casanova, A. Romero, P. Lio, and Y. Bengio. 2018. "Graph Attention Networks." *International Conference on Learning Representations*, accepted as poster. [Online].<https://openreview.net/forum?id=rJXMpikCZ>.
- Wang, Y., X. Chen, Z. Han, and S. He. 2017. "Hyperspectral Image Super-Resolution via Nonlocal Low-Rank Tensor Approximation and Total Variation Regularization." *Remote Sensing* 9 (12): 1286. <https://doi.org/10.3390/rs9121286>.
- Wang, Z., and A. Bovik. 2002. "A Universal Image Quality Index." *IEEE Signal Processing Letters* 9 (3): 81–84. <https://doi.org/10.1109/97.995823>.
- Wang, Z., A. Bovik, H. Sheikh, and E. Simoncelli. 2004. "Image Quality Assessment: From Error Visibility to Structural Similarity." *IEEE Transactions on Image Processing* 13 (4): 600–612. <https://doi.org/10.1109/TIP.2003.819861>.
- Xu, Y., J. Hou, X. Zhu, C. Wang, H. Shi, J. Wang, Y. Li, and P. Ren. 2024. "Hyperspectral Image Super-Resolution with ConvLSTM Skip-Connections." *IEEE Transactions on Geoscience and Remote Sensing* 62:1–16. <https://doi.org/10.1109/TGRS.2024.3401843>.
- Yan, Y., W. Ren, X. Hu, K. Li, H. Shen, and X. Cao. 2021. "Srgat: Single Image Super-Resolution with Graph Attention Network." *IEEE Transactions on Image Processing* 30:4905–4918. <https://doi.org/10.1109/TIP.2021.3077135>.
- Yang, Y., and Y. Qi. 2021. "Image Super-Resolution via Channel Attention and Spatial Graph Convolutional Network." *Pattern Recognition* 112:107798. <https://doi.org/10.1016/j.patcog.2020.107798>.
- Yokoya, N., C. Grohnfeldt, and J. Chanussot. 2017. "Hyperspectral and Multispectral Data Fusion: A Comparative Review of the Recent Literature." *IEEE Geoscience and Remote Sensing Magazine* 5 (2): 29–56. <https://doi.org/10.1109/MGRS.2016.2637824>.
- Yokoya, N., and A. Iwasaki. 2016. "Airborne Hyperspectral Data Over Chikusei." *Space Application Laboratory, University of Tokyo, Japan, Tech. Rep. Sal-2016-05-27*. May. [Online] <http://park.itc.u-tokyo.ac.jp/sal/hyperdata/TechRepSAL20160527.pdf>.
- Yokoya, N., T. Yairi, and A. Iwasaki. 2012. "Coupled Nonnegative Matrix Factorization Unmixing for Hyperspectral and Multispectral Data Fusion." *IEEE Transactions on Geoscience and Remote Sensing* 50 (2): 528–537. <https://doi.org/10.1109/TGRS.2011.2161320>.
- Yu, H., Z. Ling, K. Zheng, L. Gao, J. Li, and J. Chanussot. 2024. "Unsupervised Hyperspectral and Multispectral Image Fusion with Deep Spectral-Spatial Collaborative Constraint." *IEEE Transactions on Geoscience and Remote Sensing* 62:1–14. <https://doi.org/10.1109/TGRS.2024.3509985>.
- Zhang, M., Q. Wu, J. Zhang, X. Gao, J. Guo, and D. Tao. 2022. "Fluid Micelle Network for Image Super-Resolution Reconstruction." *IEEE Transactions on Cybernetics* 53 (1): 1–14. <https://doi.org/10.1109/TCYB.2022.3163294>.
- Zhang, X., W. Huang, Q. Wang, and X. Li. 2021. "Ssr-Net: Spatial-Spectral Reconstruction Network for Hyperspectral and Multispectral Image Fusion." *IEEE Transactions on Geoscience and Remote Sensing* 59 (7): 5953–5965. <https://doi.org/10.1109/TGRS.2020.3018732>.
- Zhang, Y., K. Li, K. Li, L. Wang, B. Zhong, and Y. Fu. 2018. "Image Super-Resolution Using Very Deep Residual Channel Attention Networks." In *Computer Vision – ECCV 2018*, edited by V. Ferrari, M. Hebert, C. Sminchisescu, and Y. Weiss, 294–310. Cham: Springer International Publishing.
- Zhang, Y., K. Li, K. Li, B. Zhong, and Y. Fu. 2019. "Residual Non-Local Attention Networks for Image Restoration." *CoRR abs/1903 (10082)*. <http://arxiv.org/abs/1903.10082>.
- Zhao, Y., J. Yang, Q. Zhang, L. Song, Y. Cheng, and Q. Pan. 2011. "Hyperspectral Imagery Super-Resolution by Sparse Representation and Spectral Regularization." *EURASIP Journal on Advances in Signal Processing* 2011 (1): 1–10. <https://doi.org/10.1186/1687-6180-2011-87>.
- Zheng, K., L. Gao, D. Hong, B. Zhang, and J. Chanussot. 2022. "Nonregsrnet: A Nonrigid Registration Hyperspectral Super-Resolution Network." *IEEE Transactions on Geoscience and Remote Sensing* 60:1–16. <https://doi.org/10.1109/TGRS.2022.3225843>.
- Zheng, K., L. Gao, W. Liao, D. Hong, B. Zhang, X. Cui, and J. Chanussot. 2021. "Coupled Convolutional Neural Network with Adaptive Response Function Learning for Unsupervised Hyperspectral Super Resolution." *IEEE Transactions on Geoscience and Remote Sensing* 59 (3): 2487–2502. <https://doi.org/10.1109/TGRS.2020.3006534>.
- Zhou, S., J. Zhang, W. Zuo, and C. C. Loy. 2020. "Cross-Scale Internal Graph Neural Network for Image Super-Resolution." In *Advances in Neural Information Processing Systems*, 3499–3509. Vol. 33. The MIT Press. [https://proceedings.neurips.cc/paper\\_files/paper/2020/file/23ad3e314e2a2b43b4c720507cec0723-Paper.pdf](https://proceedings.neurips.cc/paper_files/paper/2020/file/23ad3e314e2a2b43b4c720507cec0723-Paper.pdf).