International Journal of Remote Sensing

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/tres20

# CNN-Enhanced graph attention network for hyperspectral image super-resolution using non-local self-similarity

Cong Liu & Yaxin Dong

**To cite this article:** Cong Liu & Yaxin 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, DOI: 10.1080/01431161.2022.2121188

**To link to this article:** https://doi.org/10.1080/01431161.2022.2121188

Published online: 14 Sep 2022.

INTERNATIONAL JOURNAL OF REMOTE SENSING
2022, VOL. 43, NO. 13, 4810–4835
https://doi.org/10.1080/01431161.2022.2121188

Taylor & Francis
Taylor & Francis Group



# CNN-Enhanced graph attention network for hyperspectral image super-resolution using non-local self-similarity

Cong Liu and Yaxin Dong

School of Optical-Electrical and Computer Engineering, University of Shanghai for Science and Technology, Shanghai, China

**ABSTRACT**

The small-sample problem that widely existed in the hyperspectral image (HSI) super-resolution task will lead to insufficient feature extraction in network training. Therefore, it is necessary to design an effective network to extract the feature of HSIs fully. In addition, existing HSI super-resolution (SR) networks usually capture multiple receptive fields by staking massive convolutions, which will inevitably produce many parameters. In this paper, we propose a novel HSI SR network based on the convolution neural network enhanced graph attention network (CEGATSR), which can fully capture different features by using a graph attention block (GAB) and a depthwise separable convolution block (DSCB). Moreover, the graph attention block can also capture different receptive fields by using relatively few layers. Specifically, we first divide the whole spectral bands into several groups and extract the features separately for each group to reduce the parameters. Second, we design a parallel feature extraction unit to extract non-local and local features by combining the graph attention block (GAB) and the depthwise separable convolution block (DSCB). The graph attention block makes full use of the non-local self-similarity strategy not only to self-learn the effective information but also to capture the multiple receptive fields by using relatively few parameters. The depthwise separable convolution block is designed to extract the local feature information with few parameters. Third, we design a spatial-channel attention block (SCAB) to capture the global spatial-spectral features and to distinguish the importance of different channels. A large number of experiments on three hyperspectral datasets show that the proposed CEGATSR performs better than the state-of-the-art SR methods. The source code is available at [Online]. Available: https://github.com/Dongyx1128/CEGATSR.

**ARTICLE HISTORY**
Received 22 March 2022
Accepted 28 August 2022

**KEYWORDS**
Hyperspectral image(HSI); graph attention network(GAT); super-resolution(SR); group spectral

## 1. Introduction

The hyperspectral imaging technique has recently been an active research topic because of its rich spectral information obtained by hundreds of narrow spectral bands simultaneously for the same scene. Thanks to the rich spectral and detailed information availability, HSIs have shown a solid spectral diagnostic ability and can distinguish substances

**CONTACT** Cong Liu congl2014@usst.edu.cn School of Optical-Electrical and Computer Engineering, University of Shanghai for Science and Technology, Shanghai 200093, China

similar to humans. Consequently, the HSI has been extensively exploited in many applications, such as land cover detection (Liu, Su and Li 2016), surveillance security (Rasti et al. 2016), medical diagnosis (Pike et al. 2016), aerospace field (Arun et al. 2019) and many other fields. As everyone knows, the spatial and spectral resolution reflect the degree of the details of the spatial and spectral information, respectively, in an HSI. However, due to the hardware limitation, there is a tradeoff between the spectral and spatial resolution in acquiring an HSI. In this way, researchers usually maintain the high spectral resolution while sacrificing the spatial resolution to get a low-spatial-resolution HSI. As a result, the spatial resolution of an HSI is often poorer than that of natural or multispectral images (Qu, Qi and Kwan 2018). It is necessary to find or design an effective and economical way to enhance the spatial resolution. The HSI super-resolution is such a way that gets a high spatial-resolution HSI from its one or more degraded counterparts by using some image post-processing techniques.

In general, HSI SR is a very challenging ill-posed inverse problem because the mapping between low-resolution (LR) and high-resolution (HR) pairs has multiple solutions. To solve this inverse problem, many HSI SR methods have been proposed. According to whether the auxiliary images are utilized, the existing HSI SR methods can be roughly divided into two categories, i.e. the fusion-based HSI SR and the single HSI SR (Yokoya, Grohnfeldt and Chanussot 2017). The former usually employs the matrix or tensor factorization (Dian, Li and Fang 2019; Sun et al. 2021; Zhang et al. 2021), sparse representation (Han et al. 2020) and recently advanced deep learning (Zhu et al. 2021) to design SR models, which has played an important role in recent years and achieved considerable performance (Akhtar, Shafait and Mian 2015; Dian, Fang and Li 2017; Dong et al. 2016; Yokoya, Yairi and Iwasaki 2012). However, obtaining auxiliary images with a high matching degree is difficult in practical applications.

The single HSI SR well avoids the shortcoming contained in the fusion-based HSI SR. It is a signal postprocessing technique that usually gets the desired HR-HSI by only one corresponding LR-HSI without requiring any auxiliary images. Early SR techniques are usually designed for grey/RGB images, containing two categories, i.e. model- and learning-based methods. The model-based methods focus on the utilization of handcraft priors, such as non-local prior, sparse prior, low-rank prior (Dong et al. 2011; Ren et al. 2019), and autoregressive model (Hung and Siu 2012) to capture the inner structure of the reconstructed image. This strategy has shown good performance in restoring precise details. However, it will lead to a large time complexity and performance degradation (He et al. 2016; Huang, Yu and Sun 2014; Irmak, Akar and Yuksel 2018; Wang et al. 2017) with the increase of the scaling factor. The learning-based methods can be further divided into neighbourhood embedding methods, sparse coding methods, and convolution neural network (CNN) based methods. Among these methods, CNN-based methods (Dong et al. 2016; Lim et al. 2017; Zhang and Li 2018) with their strong representation ability, have already been demonstrated to be a very feasible way for the grey/RGB image SR, such as SRCNN (Dong et al. 2016), RCAN (Zhang and Li 2018), EDSR (Lim et al. 2017) and so on.

It is difficult to directly borrow these natural image SR methods into the HSI SR task because they do not consider the high correlation among the spectral bands and lead to spectral distortion (Mei et al. 2017). Hence, recent researches pay more attention to the inherent spectral correlation preservation. In model-based methods, the spectral correlation can be captured by using the spectral matrix factorization or directly low-rank prior

for the spectral direction. As for the matrix factorization strategy, researchers usually reorder the desired 3D HR-HSI into a spectral matrix and assume that the matrix can be mapped into a low-dimensional subspace by using a spectral dictionary and a set of spectral coefficients. As for the low-rank prior strategy, researchers usually directly apply the low-rank prior into the desired HR-HSI. CNN-based methods apply the spectral difference learning (Hu et al. 2020; Hu, Zhao and Li 2019) and the 3D convolution (Mei et al. 2017) to enable this capability. For the past few years, many CNN-based methods for HSI SR have been proposed (Hu et al. 2020; Hu, Zhao and Li 2019; Jiang et al. 2020; Li, Wang and Li 2020, 2021; Liu, Li and Yuan 2021; Wang, Li and Li 2021) and have achieved satisfactory results relatively in both visually and quantitatively.

Although CNN-based methods have achieved significant improvement in the field of HSI SR, it is still a challenging task and has some space for further improvement. First, many HSIs are obtained by using different sensors or cameras, so the available training samples are relatively few, which will lead to insufficient feature extraction in the network, making it difficult to guarantee the training process (Fu, Liang and You 2021; Jiang et al. 2020). Second, it is known that deepening or widening the network by stacking massive convolutions can enable diversified receptive fields with more contextual and details information. However, it also increases the model complexity because a larger number of parameters in the network are needed. Third, some CNN-based methods usually assume that all the spectral bands are equally important. Precisely speaking, they exploit the regular convolutions such as 3D convolution or the pseudo 3D (2D + 1D) convolution to extract the features, which will lose the discriminating differences of spectral bands since they assign the same weight to different spectral bands.

Recently, the graph neural network (Velickovic et al. 2018) is an emerging technique, widely used in many image processing tasks, such as denosing (Valsesia, Fracastoro and Magli 2020) and SR (Yan et al. 2021). Unlike the regular convolution, which captures the adjacent relationship using a normal convolution kernel, the graph-based convolution can capture the long-term relationship feature. That is, the regular convolution can capture the only single receptive field (we call it the local feature). On the contrary, the graph-based convolution can capture multiple different receptive fields (we call it the non-local feature). By analysing the characteristic of the graph-based convolution, we find that taking the graph-based convolution into single HSI SR task can solve the first and the second limitations introduced above. For the first limitation, an HSI contains many similar regions, and it does not mean that all regions lose vital information in the degradation process. Hence, all the similar regions can learn the information from each other by using the graph-based convolution, which can alleviate the problem of the insufficient feature extraction. For the second limitation, the size of the receptive field in the graph-based convolution is dynamic, which can capture multi-level features by using only fewer layers. For the third limitation, the widely used channel attention mechanism can be fed into our network to capture the high-frequency information better.

Based on the above analyses, in this paper, we propose a CNN-enhanced graph attention HSI super-resolution network (CEGATSR) to extract more powerful features for reconstructing the high spatial HSIs better. To reduce the computational cost caused by massive spectral bands in an HSI, we first divide the whole spectral bands into several groups. Second, we design a parallel feature extraction unit to better capture the effective feature information by combining two complementary feature extraction units, i.e. the

non-local feature extraction unit and the local feature extraction unit. In the non-local feature extraction unit, we apply the graph attention block (GAB) to not only capture the non-local feature information but also self-learn the feature information from similar regions. In the local feature extraction unit, we apply the depthwise separable convolution block (DSCB) with relatively few parameters to replace the regular convolution to capture the local feature information. Thirdly, these groups are merged to form the whole spectral bands. We design a global feature mapping spatial-channel attention block (SCAB) composed of the spatial block and the channel attention block to further capture the global and high-frequency information. Experiments on public data sets verify the effectiveness the proposed CEGATSR. To sum up, our main contributions are as follows:

- Aiming at the problem of insufficient feature extraction caused by inadequate training samples of the HSI SR network, we propose a novel CEGATSR network for the single HSI SR.
- For capturing more features from the HSI, we design a parallel feature extraction unit. In this unit, both the non-local and local features as complementary features are extracted by using a graph attention block (GAB) and a depthwise separable convolution block, respectively.
- To further capture the global feature and high-frequency information, we design a spatial-channel attention block (SCAB) by combining a spatial block and a channel attention mechanism, which can re-weight the importance of different channels.
- We evaluate the proposed CEGATSR on three widely used HSI datasets, and the experimental results show that the proposed CEGATSR outperforms the most advanced methods.

The rest of the paper is organized as follows: We first review the related literature in Section **2**. The details of the proposed CEGATSR are presented in Section **3**. Experimental results compared with existing methods are evaluated in Section **4**. Finally, Section **5** concludes the paper.

## 2. Related work

This section briefly introduces existing deep learning-based single SR methods, including deep learning-based single RGB image SR methods and deep learning-based single HSI SR methods.

### *2.1. Deep learning-based single RGB image SR methods*

In recent years, CNN-based single RGB image SR methods have emerged one after another. SRCNN (Dong et al. 2016) is a groundbreaking work in the SR task proposed by Dong et al. and achieves better performance than traditional methods by learning an end-to-end nonlinear feature mapping between LR and HR image pairs. Soon afterwards, they further proposed FSRCNN (Dong, Loy and Tang 2016) to reduce the number of parameters by performing the upsampling procedure at the end of the network. Both two approaches are only performed on the shallow network. For extracting more useful information, many methods are designed by using a deeper network. Kim et al. proposed

VDSR (Kim, Lee and Lee 2016), an extremely deep network using 20 convolution layers. In addition, LapSRN (Lai et al. 2017) introduced the Laplace pyramid structure to SR coarse-to-fine way, achieving more efficient performance with fewer operations. Ledig et al. introduced ResNet (He et al. 2016) to build a deeper network called SR with deep ResNet (SRRes-Net) (Ledig et al. 2017). To obtain better performance, Lim et al. designed EDSR (Lim et al. 2017), which is a very deep and wide network by using a modified residual block and makes significant improvements. Lately, many studies have applied the attention mechanism to distinguish the importance of different features and got good performance. Zhang et al. (Zhang and Li 2018) presented a very deep network (RCAN) that combines the advantages of the residual block and the attention mechanism. It is the first time the channel attention mechanism has been applied to the image SR task. Dai et al. (Dai et al. 2019) noticed that many methods neglect the feature correlation of intermediate layers and then develop a second-order attention network (SAN) to extract more helpful feature information. CNN-based methods usually capture features in a local way because of the fixed receptive field of the convolutional operation, which fails to capture global self-similarity properties effectively in an image (Yang and Qi 2021). Therefore, many recent studies have taken the graph neural networks (GNNs) into the natural image super-resolution. Zhou et al. designed a single image super-resolution method, which explores the natural image's cross-scale patch recurrence property using a novel cross-scale internal graph neural network (IGNN) (Zhou et al. 2020). Yang et al. (Yang and Qi 2021) proposed channel attention and spatial graph convolution network (CASGCN), combining the channel attention mechanism and spatial graph convolution network (GCN) to enhance the ability to extract features. Yan et al. proposed an SRGAT (Yan et al. 2021) network based on the graph attention network (GAT), using high-level information to promote low-level features through a feedback mechanism. However, it is very difficult to directly copy these methods into the field of the single HSI SR because the band-wise reconstruction will lose the high correlation across the spectral bands and lead to spectral distortion (Mei et al. 2017) in the super-resolved HSI (Wang et al. 2017).

### 2.2. Deep- learning-based single HSI SR methods

Deep-learning-based single HSI SR methods have recently attracted increasing attention because of the powerful representation ability of the convolution neural network. As introduced above, the spectral correlation preservation is the main theme in this field. Li and Hu proposed three HSI SR networks named SCT_SDCNN (Li et al. 2017), SEC_SDCNN (Hu, Li and Xie 2017) and IFN (Hu et al. 2020), all of which apply the spectral difference learning to maintain the correlation of different spectral bands. In (Hu, Zhao and Li 2019), Hu applied an intra-fusion strategy to capture the correlation between the selected and unselected bands. To fully exploit the rich spectral information, other studies use 3D convolution to extract the spatial and spectral information simultaneously. For example, Mei et al. (Mei et al. 2017) proposed a 3D full convolution neural network (3D-FCNN) with a five-layer structure. However, the network does not fill all convolution during reconstruction, resulting in the estimated size of the HSI being smaller than the size of the input image, so it is not suitable for image reconstruction. Moreover, the number of parameters in the network with 3D convolution is significantly larger than that with 2D convolution, making it difficult to design deeper networks with

3D convolution. Hence, many researchers apply the separable 3D convolution replacing the regular 3D convolution to reduce the parameters. Li and Wang proposed ERCSR (Li, Wang and Li 2021), SFCSR (Wang, Li and Li 2021), SSRNet (Wang, Li and Li 2020) and MCNet (Li, Wang and Li 2020), all of which apply the separable 3D convolution to reduce the computational cost. Besides, other studies group the whole spectral bands into several sub-bands to reduce the parameters. Li et al.(Li et al. 2018) proposed the grouped depth recursive residual networks (GDRRN) by embedding the grouped recursive module into the global residual structure. Although they can contribute to the spatial resolution, they do not consider the spectral correlation, resulting in the spectral distortion. Later, both SSPSR (Jiang et al. 2020) and RFSR (Wang, Ma and Jiang 2021) applied spectral band grouping strategies to reduce the network parameters to enable a deep network effectively.

## 3. Methods

### 3.1. Model formulation

From a generative perspective, the observed LR-HSI view $I_{LR} \in \mathbb{R}^{h\times w\times C}$ is generated by the degradation model

$$I_{LR} = fold(B \times unfold(I_{HR}) + \eta), \tag{1}$$

where $I_{HR} \in \mathbb{R}^{H\times W\times C}$ denotes the corresponding HR-HSI. $h$, $w$ and $C$ represent the height, width and spectral band number of $I_{LR}$ respectively. $h$ and $w$ represent the height and width of $I_{HR}$ respectively. They satisfy with $H = sf \times h$ and $W = sf \times w$ and $sf$ denotes the scaling factor. *unfold* is an operator that transforms a 3D cube with the size of $H \times W \times C$ into a 2D matrix with the size of $HW \times C$. *fold* is its inverse operator that transforms a 2D matrix with a size of $HW \times C$ into a 3D cube with the size of $H \times W \times C$. $B \in \mathbb{R}^{hw\times HW}$ represents the degraded matrix, which is often assumed to be composed of a downsampling operator and a blurring filter. $\eta \in \mathbb{R}^{h\times w\times C}$ is the noises contained in $I_{LR}$. Our goal is to get a corresponding HR-HSI view $I_{SR} \in \mathbb{R}^{H\times W\times C}$ from the observed $I_{LR}$ by using the proposed CEGATSR, which can be formulated as

$$I_{SR} = H_{CEGATSR}(I_{LR}), \tag{2}$$

where $H_{CEGATSR}(\cdot)$ represents the proposed CEGATSR network. In the following, we denote $Conv_{(e,f)}$ to represent a convolution layer, where $e$ and $f$ represent the size of a filter and the number of the filters, respectively.

### 3.2. Network architecture

The overall structure of the proposed CEGATSR is shown in Figure 1. Similar to some existing HSI SR networks, CEGATSR employs the global residual learning to increase the convergence speed and learn the low-frequency information. Consequently, the main path of CEGATSR is used to recover the residual image (high-frequency information) between the LR-HSIs and their corresponding HR-HSIs. For the global residual learning, we employ the double cubic interpolation to interpolate the input LR-HSI to the desired size of the output HR-HSI band by band, which is formulated as

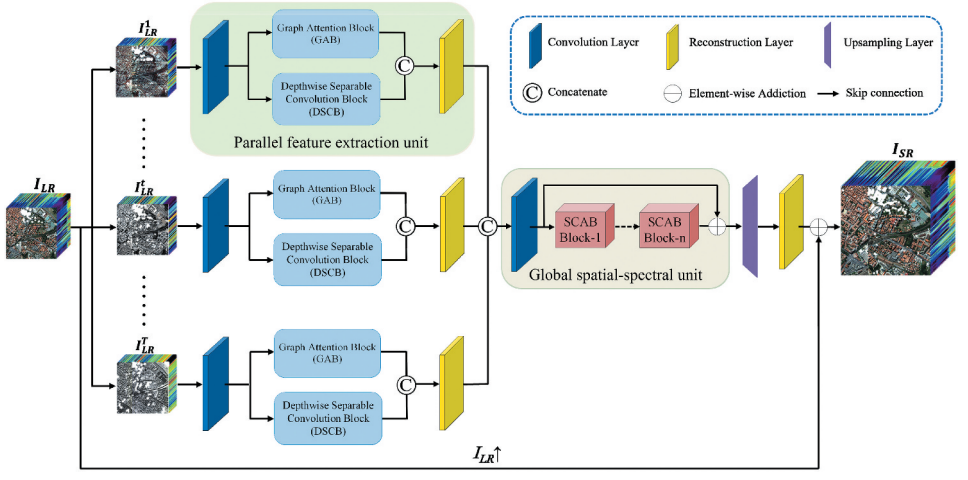


**Figure 1.** Overall architecture of the proposed CEGATSR model.

$$I_{LR}\uparrow = H_{cubic}(I_{LR}), \tag{3}$$

where $H_{cubic}$ denotes the double cubic interpolation operator and $I_{LR}\uparrow \in \mathbb{R}^{H\times W\times C}$ denotes the interpolated LR-HSI.

The main path consists of the following five parts, i.e. (1) the grouped parallel module for reducing the computational cost, (2) the graph attention block (GAB) for capturing the non-local similarity information, (3) the depthwise separable convolution block (DSCB) for capturing the local feature information, (4) the spatial-channel attention block (SCAB) to capture the global spatial-spectral feature and capture the importance of different channels, and (5) the upsampling layer and reconstruction layer to get the desired HR-HSI.

### 3.2.1. Grouped parallel module

Since an HSI often contains massive spectral bands, directly extracting the features will inevitably increase the computational cost. Recent studies tend to apply the grouped strategy to overcome this problem (Wang, Ma and Jiang 2021). Here, our proposed network also borrows this strategy. Specifically, the whole spectral bands are split into several overlapping groups, and the feature extraction is performed for each group. Assuming that we split $C$ spectral bands into $T$ overlapping groups and each group has $g$ spectral bands with $C < T \times g$, the $T$th group band ($t \in \{1., T\}$) can be represented as $I_{LR}^{t} \in R^{h\times w\times g}$.

### 3.2.2. Parallel feature extraction unit

The crucial issue of a convolution network is how to extract the useful features. Existing networks usually stack multiple convolution kernels to capture different feature maps. However, they will lead to large parameters. As we all know, regular convolution can capture the local feature. On the contrary, the graph attention block can capture the non-local feature according to the similarity of image patches. Therefore, here, we combine the two types of convolutions to design a parallel feature extraction unit to better extract different types of features with few layers. It has been confirmed that the depthwise separable

convolution significantly reduces the computational cost without reducing the accuracy. Therefore, we apply the depthwise separable convolution block (DSCB) to replace the regular convolution to reduce the number of parameters. We apply the $t$th group feature map $I_{LR}^t$ to expound the details. We first apply a convolution layer $Conv_{(3,L)}$ to expand the channel number of $I_{LR}^t$ from $g$ to $L$ to capture more information, which can be modelled as

$$G^t = Conv_{(3,L)}(I_{LR}^t), \tag{4}$$

where $G^t \in \mathbb{R}^{h\times w\times L}$ denotes the expanded feature map. And then, it is fed into GAB and DSCB in parallel, which can be formulated as

$$G_{GAB}^t = H_{GAB}(G^t), \tag{5}$$

and

$$G_{DSCB}^t = H_{DSCB}(G^t), \tag{6}$$

respectively, where $H_{GAB}$ and $H_{DSCB}$ represent the GAB and DSCB operators, respectively. More details of them will be introduced in Section 3.3 and Section 3.4 respectively. After obtaining the local and non-local features, we concatenate them and use a convolution $Conv_{(3,g)}$ to reduce the channel number from $L$ to $g$, which is modelled as

$$G_{pre}^t = Conv_{(3,g)}([G_{GAB}^t, G_{DSCB}^t]), \tag{7}$$

where $G_{pre}^t \in R^{h\times w\times g}$ denotes the reduced feature and $[\cdot]$ represents the concatenation operator. After performing all the groups, we merge them into a pre-HSI, $F_{pre} = [G_{pre}^1 \cdots G_{pre}^T] \in \mathbb{R}^{h\times w\times C}$.

### 3.2.3. Global spatial-spectral unit

Both GAB and DSCB focus on extracting the spatial and spectral features for each group without considering the spatial and spectral features for the whole spectral bands. In addition, recent studies have confirmed that the high-frequency information in the HSIs is the fundamental information in HSI SR. Consequently, in this subsection, we propose a spatial-channel attention block (SCAB) (introduced in section 3.5) to capture the global spatial and spectral features and the high-frequency information. Moreover, we stack multiple SCABs to extract the deeper feature information. Besides, we apply a short skip mechanism to propagate the low-level feature from the former layers to the latter layers. Before the series of operations, we apply a convolution $Conv_{(3,L)}$ to expand the channel number, which is shown as

$$EF = Conv_{(3,L)}(F_{pre}), \tag{8}$$

where $EF \in R^{h\times w\times L}$ represents the expanded feature map. The SCAB operator is performed as

$$F_{SCAB,R} = H_{SCAB,R}(\cdots H_{SCAB,r}(\cdots H_{SCAB,1}(EF)\cdots)\cdots) + EF, \tag{9}$$

where $H_{SCAB,r}$ represents the $r$th SCAB operator and $r \in \{1., R\}$.

### 3.2.4. Upsampling layer and reconstruction layer

Finally, we perform the upsampling and reconstruction layers to obtain the desired HR-HSI. The former uses a convolution layer and a sub-pixel convolution layer sequence to

upsample the channel number to the desired size. The latter also utilizes a convolution layer $Conv_{(3,C)}$ to reduce the channel number to the desired spectral number. And then, combining the interpolated LR-HSI $I_{LR}\uparrow$, we can get the final HR-HSI. This operator can be formulated as

$$I_{SR} = Conv_{(3,C)}(F_{up}(H_{global})) + I_{LR}\uparrow, \tag{10}$$

where $F_{up}(\cdot)$ represents the upsampling layer.

### 3.3. Graph attention block (GAB)

As illustrated above, GAB can capture the non-local feature according to the similarity of image patches. The structure of GAB is shown in Figure 2. We use $F_{in} \in \mathbb{R}^{h\times w\times L}$, which is a simple substitution of $G^t$ ($t \in \{1., T\}$), to represent the input feature of GAB. As shown in Figure 2, GAB is roughly divided into three steps. First, $F_{in}$ is fed into a convolution layer with a stride of $s$ and a kernel size of $3 \times 3$, to get the patch feature $P \in \mathbb{R}^{\frac{h}{s}\times\frac{w}{s}\times L}$, in other words, each position of $P$ corresponds to a patch of $F_{in}$. Second, the patch feature $P$ is reshaped into a feature matrix $H$ with size of $N \times L$, where $N = \frac{h}{s} \times \frac{w}{s}$. According to the feature matrix $H$, a specific graph is created. Each row of $H$ can be regarded as a node of this graph and the similarity between each pair of rows can be regarded as an edge of the graph. we apply Euclidean distance to calculate the similarity between each pair of nodes. After calculating the similarity of all pair of nodes, we can get the adjacency matrix $adj \in \mathbb{R}^{N\times N}$. In order to reduce the memory and computational overhead caused by getting edges, we select the nearest $k$ neighbouring nodes for each node to construct the adjacency matrix, so it is a sparse matrix with only $k$ values of 1 (the rest are 0) per row. The graph can be represented as $\mathcal{G}(H, adj)$, where $H$ and $adj$ can also be regarded as the vertex set and the edge set, respectively. For matrixing description, we reshape $H$ into $\mathbf{H}_l = \left\{\vec{h}_1, \cdots \vec{h}_N\right\} \in \mathbb{R}^{L\times N}$ by the transpose operator, where $\vec{h}_i \in \mathbb{R}^L$ of the nodes in the graph. Since the similar nodes should have more positive correlation with each other, the GAT layer (Velickovic et al. 2018) assign more weights to more similar nodes. By using GAT layer, a set of new nodes $\mathbf{H}_l^{'} = \left\{\vec{h}_1^{'}, \cdots \vec{h}_N^{'}\right\} \in \mathbb{R}^{L'\times N}$, where $\vec{h}_i^{'} \in \mathbb{R}^{L'}$, can be gotten. Many references set $L' = L$ to guarantee that the input dimension is equal to the output dimension. And third, we transform $\mathbf{H}_l^{'}$ to obtain a 3D feature map with size of $\frac{H}{s} \times \frac{W}{s} \times L$ and then apply a transposed convolution layer to change the 3D feature map to obtain the output feature of GAB $F_{out} \in \mathbb{R}^{h\times w\times L}$.

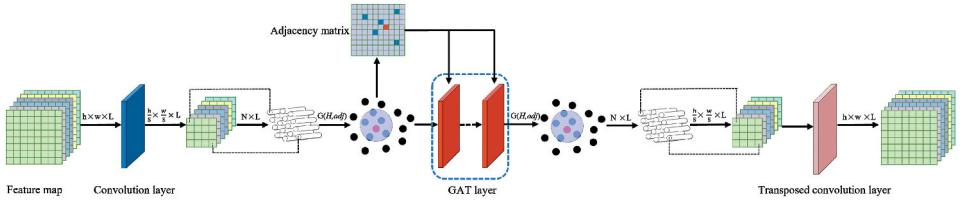


**Figure 2.** The structure of the graph attention block (GAB).

Next, we use the $i$th node $\vec{h}_i$ and its nearest $k = 5$ neighbouring nodes as an example to describe the procedure of the GAT layer, which is shown in Figure 3(b) and the output feature is $\vec{h}_i'$. The details are shown in Figure 3(a). First, a learnable shared linear transformation weight matrix $\mathbf{W} \in {}^{L' \times L}$ is added for each node to generate more deeper features and then apply the shared self-attention mechanism $a : \mathbb{R}^{L'} \times \mathbb{R}^{L} \rightarrow \mathbb{R}$ to calculate the influence of $\vec{h}_j$ to $\vec{h}_i$.

$$e_{ij} = a(\mathbf{W}\vec{h}_i, \mathbf{W}\vec{h}_j) = \text{LeakyReLU}(\vec{\mathbf{a}}^{\mathrm{T}}\left[\mathbf{W}\vec{h}_i \parallel \mathbf{W}\vec{h}_j\right]), \tag{11}$$

where $a$ is a single-layer feedback neural network layer as shown in Figure 3(a), which is parameterized by $\vec{\mathbf{a}} \in \mathbb{R}^{2L'}$ and then uses LeakyReLU nonlinearity. The attention coefficient $e_{ij}$ represents the importance of the node $\vec{h}_j$ to the node $\vec{h}_i$. We only calculate $e_{ij}$ for nodes $j \in \mathcal{N}_i$, where $\mathcal{N}_i$ represents $k$ first-order neighbour nodes of $\vec{h}_i$ (including $i$) in the graph. In order to compare the coefficients between different nodes, we use the softmax function to normalize all the coefficients of $j$ selection:

$$\alpha_{ij} = \text{softmax}_j(e_{ij}) = \frac{\exp(\text{LeakyReLU}(\vec{\mathbf{a}}^{\mathrm{T}}\left[\mathbf{W}\vec{h}_i \parallel \mathbf{W}\vec{h}_j\right]))}{\sum_{l \in \mathcal{N}_i} \exp(\text{LeakyReLU}(\vec{\mathbf{a}}^{\mathrm{T}}\left[\mathbf{W}\vec{h}_i \parallel \mathbf{W}\vec{h}_l\right]))}. \tag{12}$$

The output feature $\vec{h}_i'$ is gotten by aggregating all the neighbouring features.

$$\vec{h}_i' = \sigma(\sum_{j \in \mathcal{N}_i} \alpha_{ij}\mathbf{W}\vec{h}_j). \tag{13}$$

The multi-head attention mechanism is an extension for better extracting deep features, and it is shown in Figure 3(b). Assuming that we have $M$ independent attention

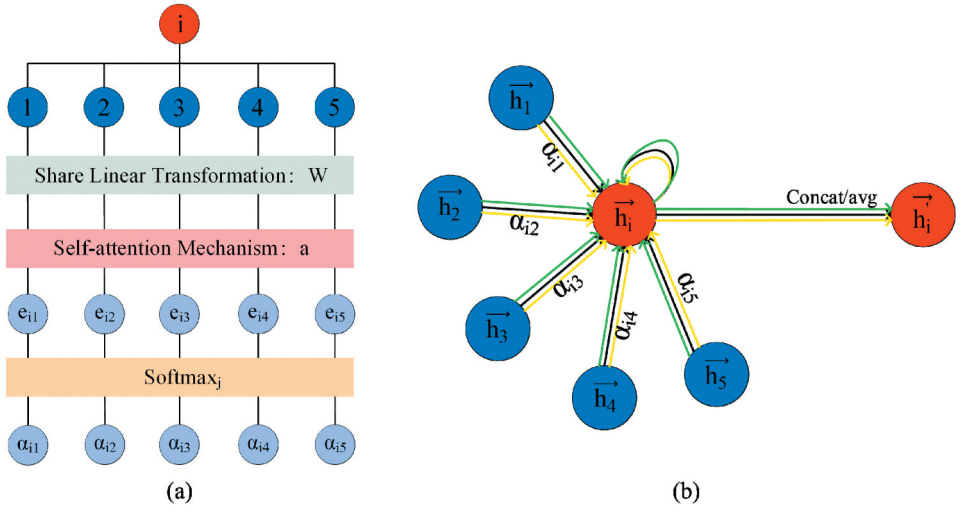


**Figure 3.** (a) the procedure of the GAT Layer, (b) Multi-head GAT layer. The red circle $i$ and blue circles represent the $i$th node and its nearest 5 neighboring nodes, respectively. Different arrow colors denote independent attention computations, represented in the (b) as three heads of attention.

mechanisms for each graph and each attention mechanism can get an updated feature for the *i*th feature by using Equation (13). The updated feature maps obtained by *M* independent attention mechanisms are first concatenated and then averaged to obtain the final updated features,

$$\vec{h}_i^{\,'} = \sigma\left(\frac{1}{M}\sum_{m=1}^{M}\sum_{j\in\mathcal{N}_i}\alpha_{ij}^{m}\mathbf{W}^{m}\vec{h}_j\right), \tag{14}$$

where $\alpha_{ij}^{m}$ and $\mathbf{W}^{m}$ are the normalized attention coefficient computed and the linear transformation's weight matrix of the *m*th attention mechanism($a^k$). More details are shown in (Velickovic et al. 2018).

### 3.4. Depthwise separable convolution block (DSCB)

In general, the regular convolution can restore the local feature information such as edge and texture details. However, since the spectral band number of the HSI itself is high, it will lead to a large number of parameters in the network. Hence, in this paper, we apply the depthwise separable convolution replacing the regular convolution to reduce the computation cost. Depthwise separable convolution was proposed by Andrew et al. (Howard et al. 2017), which greatly improves the calculation speed than traditional convolution neural networks and has only a little performance loss. Many researchers have taken the depthwise separable convolution into different image processing tasks recently (Chollet 2017). This fact shows the superiority of the depthwise separable convolution.

However, directly applying the depthwise separable convolution into our model will lead to the spectral distortion because of the high spectral bands in HSIs, so we design a new depthwise separable convolution block (DSCB) by combining the spectral convolution and spatial convolution to better extract the local feature. The structure of the depthwise separable convolution is shown in Figure 4. As illustrated above, the input of the block is $G^t \in \mathbb{R}^{h\times w\times L}$. We first apply a convolution layer with $1\times 1$ to capture the correlation of the feature maps to get the fused feature map $FG^t \in \mathbb{R}^{h\times w\times L}$. Then, the fused feature map is fed into the spatial convolution. Each feature channel of $FG^t$ is convoluted using a convolution kernel $Conv_{(3,1)}$. For example, the *i*th feature channel of $FG^t$ is convoluted by using

$$\widetilde{FG_i^t} = Conv_{(3,1)}(FG_i^t). \tag{15}$$

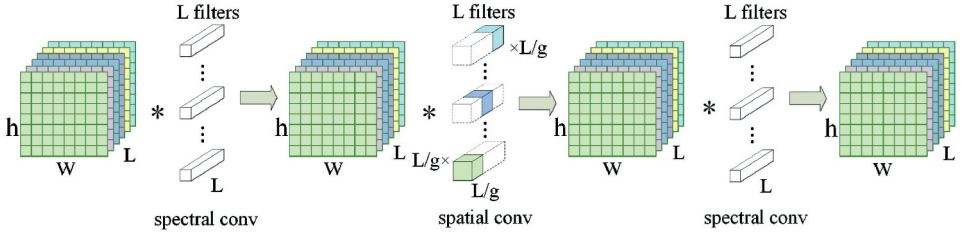


Figure 4. Depthwise separable convolution block (DSCB).

Next, the newly feature channels are concatenated to form the new feature map $\widetilde{FG^t} = [\widetilde{FG_1^t}, ., \widetilde{FG_L^t}]$. However, the correlation between feature maps can not be captured. To address this problem, we reuse a convolution layer with $1 \times 1$ to capture the correlation among feature maps to get the output feature map $G_{DSCB}^t$.

$$G_{DSCB}^t = Conv_{(1,L)}(\widetilde{FG^t}). \tag{16}$$

### 3.5. Spatial-Channel attention block (SCAB)

This subsection will introduce the details of SCAB, which is used to extract the global spatial and spectral features of the whole spectral bands and capture the importance of different features by using the channel attention mechanism (Hu et al. 2020). We propose a spatial-channel attention block (SCAB) to capture the global spatial and spectral features and the high-frequency information. The structure of SCAB is shown in Figure 5, which contains a two-branch parallel structure. The upper branch is used to capture the spatial feature maps, and the lower branch is used to capture the spectral band correlation and the attention mechanism block. For the $r$th SCAB block $H_{SCAB,r}$, its input and output feature maps are $FS_{r-1}$ and $FS_r$ respectively. Noticed that $FS_0$ and $EF$ (obtained in Equation (8)) are the same. For the upper branch, we apply two serial convolutions and a skip connection to capture the global spatial feature, which is formulated as

$$F_{upper} = Conv_{(3,L)}(Relu(Conv_{(3,L)}(FS_{r-1}))) + FS_{r-1}. \tag{17}$$

For the lower branch, we first apply two convolutions with $1 \times 1$ to capture the spectral correlation and then apply the channel attention mechanism block to distinguish the importance of different feature channels. For the channel attention mechanism, we use an average pooling layer and two fully connected layers with a simple gating mechanism to learn the weight vector, which implies the importance of different feature maps. Next, the weight vector is used to reweight the feature maps. After this procedure, we can get the feature map of the lower branch $F_{lower}$.

After that, the two feature maps $F_{upper}$ and $F_{lower}$ are concatenated. And, we apply a convolution with $1 \times 1$ to reduce the concatenated feature maps to $L$. Combining the

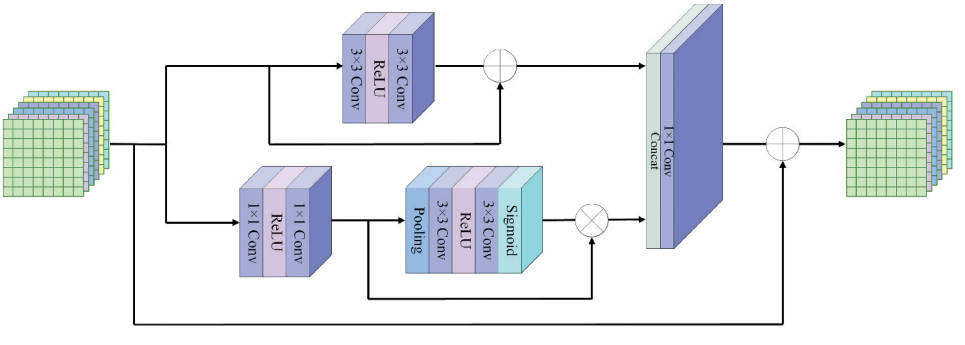


Figure 5. Spatial-Channel attention block (SCAB).

input feature maps with a long skip connection, we can get the output feature of SCAB $F_{SCAB,r}$.

$$FS_r = Conv_{(1,L)}([F_{upper}, F_{lower}]) + FS_{r-1}. \tag{18}$$

### 3.6. Loss function

To improve the performance of the SR network, we combine two loss functions, including L1 norm loss and spatial-spectral total variation (SSTV) loss, to design the whole loss function. L1 norm can make the network more convergent. At the same time, SSTV loss can guarantee not only the performance of the spatial reconstruction but also the correlation among spectral bands. We use the weighted sum of the two-loss functions as the final loss function of the model:

$$\mathcal{L}(\Theta) = \mathcal{L}_{\infty} + a\mathcal{L}_{\mathcal{SSTV}}, \tag{19}$$

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

$$\mathcal{L}_{\infty}(\Theta) = \frac{1}{N}\sum_{n=1}^{N} \left\| I_{HR}^{n} - f_{Net}\left(I_{LR}^{n}\right) \right\|_1, \tag{20}$$

and

$$\mathcal{L}_{\mathcal{SSTV}}(\Theta) = \frac{1}{N}\sum_{n=1}^{N}(\left\| \ {}_h I_{SR}^{n} \right\|_1 + \left\| \ {}_w I_{SR}^{n} \right\|_1 + \left\| \ {}_c I_{SR}^{n} \right\|_1), \tag{21}$$

respectively, where $N$ denotes the batch size of images in training, and $\Theta$ denotes the parameter set of the proposed CEGATSR. $_h$, $_w$ and $_c$ are the functions to compute the horizontal, vertical and spectral gradient of $I_{SR}^{n}$ respectively.

## 4. Experiments and results

To evaluate the performance of the proposed CEGATSR, we conduct extensive experiments on three public HSI datasets, including two hyperspectral remotely sensed images and one natural HSI dataset. First of all, we introduce the experimental datasets and evaluation metrics. Then, the implementation details and the parameter discussion are listed. After that, we apply the ablation experiment to analyse the proposed network. Finally, the comparison results with several state-of-the-art methods are displayed.

### 4.1. Experimental data sets

#### 4.1.1. Chikusei dataset

This is a remotely sensed HSI dataset(https://www.sal.t.u-tokyo.ac.jp/hyperdata/) taken by Headwall Hyperspec-VNIR-C imaging sensor over agricultural and urban areas in Chikusei, Ibaraki, Japan. It consists of 2517 × 2335 pixels, and each pixel contains 128 spectral bands in the spectral range from 363 nm to 1018 nm.

#### 4.1.2. Pavia Centre dataset

This is also a remotely sensed HSI dataset (http://www.ehu.eus/ccwintco/index.php) acquired by the ROSIS sensor during a flight campaign over Pavia, northern Italy. This image consists of 1096 × 715 pixels, and each pixel has 115 spectral bands.

#### 4.1.3. Cave dataset

This is a natural HSI dataset (http://www.cs.columbia.edu/CAVE/databases/multispectral/) gathered by a cooled CCD camera and widely used in many multi-spectral image SR tasks. It consists of 32 scenes of a wide variety of real-world materials and objects whose spatial resolution is 512 × 512 pixels, including 31 spectral bands ranging from 400 nm to 700 nm at 10 nm steps.

### 4.2. Implementation details

The training and test HR-HSIs are obtained by the following operations. For the Chikusei dataset, we discard the missing edge to retain more valid information and select 31 spectral bands between spectral bands 50 to 80 to get a sub-image of size 2304 × 2048 × 31. We apply the top region of the image as the training set (10% of the training data is used as the verification set) and the rest region as the test set. For Pavia Centre dataset, we apply the original image as the HR-HSI and select 31 spectral bands between bands 35 to 65 to get a sub-image of size 1096 × 715 × 31. We extract the left part of the sub-image as the training set (10% randomly selected as the validation set) and the rest region as the test set. For the Cave dataset, we randomly select 20 HSIs as the training set (10% randomly selected as the validation set), and the remaining 12 HSIs are taken as the test set. Each HR-HSI is cropped to three types of HR-HSI patches with 32 × 32, 64 × 64 and 128 × 128. To get the LR-HSI, three types of HR-HSI patches are down-sampled to get the LR-HSI patches with 16 × 16 for scaling factors × 2, × 4 and × 8, respectively. The test dataset contains non-overlap images with 128 × 128. Since different HSIs are usually gathered by different hyperspectral cameras, so each HSI is trained and tested separately, which is different from the natural image. For this reason, the training images are relatively few, which is unsuitable for training the deep learning method, so we expand the training dataset through the data enhancement. Each patch is flipped and rotated horizontally (90°, 180° and 270°). In the spectral grouping, the number of overlapping bands is 1, and the edge bands are divided by $ceil(\cdot)$. We use ADAM (Kingma and Ba 2014) optimizer to train our model with $\beta_1 = 0.99$, $\beta_2 = 0.999$ and $\in = 10^{-8}$. The experimental results show that the model achieves stable performance at 60 epochs. The learning rate is initially set to $10^{-4}$ and then declined to 10 times after 30 epochs. The proposed CEGATSR is performed using PyTorch on Ubuntu 18.04.3 LTS with i9-9900KF CPU, 32GB RAM, and NVIDIA GeForce RTX 2080Ti GPU, 11GB.

We apply five kinds of quantity picture quality indices to evaluate CEGATSR qualitatively, including peak signal-to-noise ratio (PSNR), structure similarity (SSIM) (Wang et al. 2004), spectral angle mapper (SAM), root mean squared error (RMSE) and the relative dimensionless global error in synthesis (ERGAS). For PSNR and SSIM, we use their average values of all spectral bands. PSNR, SSIM, and RMSE are commonly used quantitative image restoration quality indices, and the other two indices are widely used in HSI fusion tasks.

### 4.3. Parameter discussion

To evaluate the effectiveness and sensitivity of the critical parameters in the proposed CEGATSR, we do some experiments to discuss their selection. There are the number of bands in each group $g$, the number of adjacent nodes $k$, the number of multiple heads in the GAT layer $M$ and the number of global SCAB blocks $R$. We use the Pavia Centre dataset as the training set, and the scaling factor is $\times$ 4. The results are shown in Figure 6.

First, we discuss the selection of $g$, an important parameter in the grouped strategy. We test different $g$ from 3 to 8, as shown by the red curve in Figure 6. We can see that the PSNR values are increasing at first and then decreasing as the increasing of $g$. The network can achieve the best performance when the parameter is set to 4.

Second, we discuss the selection of $k$ and $M$, two important parameters in GAB. For $k$, we test it from 3 to 12, as shown by the green curve in Figure 6. As seen from this curve, the experimental results show that the network's performance is better when $k \in [5, 9]$. So we set the number of adjacent nodes (including itself) to $k = 5$. For $M$, we test it from 1 to 8, as shown by the blue curve in Figure 6. It can be seen from the curve that the PSNR value can get the best when $M$ is set to 2 layers.

Third, we discuss the selection of $R$ and test it from 1 to 8, as shown by the pink curve in Figure 6. As can be seen from this curve, the PSNR value can achieve the best performance when $R = 6$.

### 4.4. Ablation study

The proposed CEGATSR contains four core components, i.e. spectral group (SG), graph attention (GA), separate convolution (SC) and channel attention (CA). In this section, we investigate the influence of different combinations by removing them. Table 1 shows the ablation study about these combinations on the Pavia Centre dataset for the scaling factor undefined. The spectral group strategy is used to divide the whole spectral bands into several groups, which can not only exploit the correlation among the neighbouring spectral bands but also reduce the parameters of the model. To verify the effectiveness of the spectral group, we remove the group

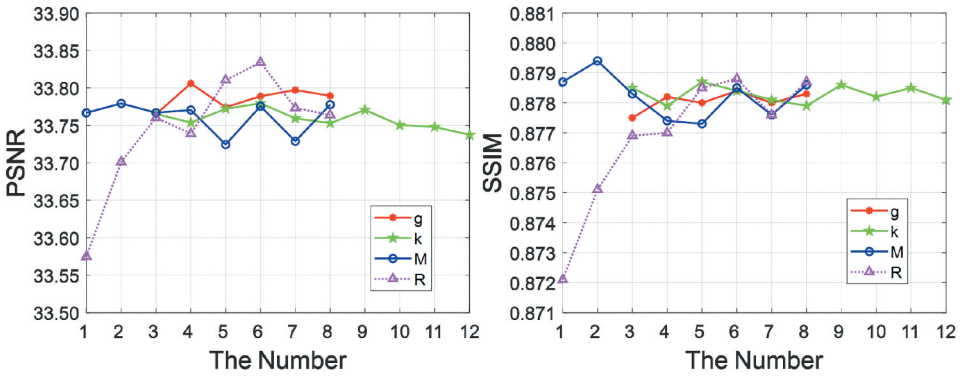


Figure 6. Parameters discussion.

**Table 1.** Ablation study. Qualitative computations obtained by using different combination on the Pavia Centre dataset for scaling factor. $\times 4$

| Models | PSNR ↑ | SSIM ↑ | SAM ↓ | RMSE ↓ | ERGAS ↓ |
|---|---|---|---|---|---|
| w/o SG | 33.5974 | 0.8733 | 2.5376 | 0.0236 | 5.2815 |
| w/o GA | 33.7033 | 0.8767 | 2.4661 | 0.0233 | 5.2211 |
| w/o SC | 33.1709 | 0.8628 | 2.6615 | 0.0247 | 5.5585 |
| w/o CA | 33.7619 | 0.8778 | 2.4834 | 0.0232 | 5.1875 |
| CEGATSR | 33.7831 | 0.8787 | 2.4740 | 0.0231 | 5.1751 |

module and reconstruct the HR-HSI band by band, which is represented as w/o SG. The GAB module is used to extract the non-local features. We replace it with two regular convolutions with $3 \times 3$ to test its effects, which are represented as w/o GA. To reduce the parameters of the network, we apply the separate convolution (SC) to extract the local feature. For testing this module, we replace it with two regular convolutions with $3 \times 3$, which is represented as w/o SC. In SCAB, we plug the channel attention mechanism into SCAB to distinguish the importance of different feature maps. We remove the channel attention mechanism to investigate its effectiveness, which is represented as w/o CA. Compared to the proposed CEGATSR with four combinations in Table 1, we can conclude that all the components are indispensable in CEGATSR.

**Table 2.** Quantitative comparison results of different SR methods on Chikusei dataset for different scaling factors. The red and blue indicate the best and second-best performances, respectively.

| SF | Method | PSNR ↑ | SSIM ↑ | SAM ↓ | RMSE ↓ | ERGAS ↓ |
|---|---|---|---|---|---|---|
| | SRCNN | 44.5358 | 0.9792 | 1.9890 | 0.0069 | 3.0357 |
| | VDSR | 44.6151 | 0.9798 | 1.7157 | 0.0066 | 3.2274 |
| | EDSR | 46.7891 | 0.9864 | 1.1532 | 0.0056 | 2.3112 |
| | RCAN | 47.1644 | 0.9875 | 1.2562 | 0.0053 | 2.2218 |
| | 3D-FCNN | 45.8663 | 0.9837 | 1.4478 | 0.0061 | 2.5788 |
| | GDRRN | 46.4444 | 0.9859 | 1.3703 | 0.0056 | 2.4313 |
| | SSPSR | 47.2311 | 0.9875 | 1.1019 | 0.0053 | 2.2044 |
| | MCNet | 47.4402 | 0.9879 | 1.1421 | 0.0052 | 2.1595 |
| x2 | CEGATSR | 47.5610 | 0.9883 | 1.0322 | 0.0051 | 2.1270 |
| | SRCNN | 38.4292 | 0.9095 | 3.5694 | 0.0142 | 6.2116 |
| | VDSR | 38.5280 | 0.9119 | 3.2138 | 0.0138 | 6.3195 |
| | EDSR | 39.5978 | 0.9270 | 2.6559 | 0.0126 | 5.3807 |
| | RCAN | 39.6842 | 0.9254 | 2.9587 | 0.0126 | 5.3142 |
| | 3D-FCNN | 38.9418 | 0.9168 | 3.1056 | 0.0135 | 5.8053 |
| | GDRRN | 39.5493 | 0.9263 | 2.8644 | 0.0126 | 5.4102 |
| | SSPSR | 40.0981 | 0.9325 | 2.3772 | 0.0121 | 5.0732 |
| | MCNet | 39.8811 | 0.9289 | 2.8033 | 0.0124 | 5.1932 |
| x4 | CEGATSR | 40.3875 | 0.9395 | 2.2665 | 0.0117 | 4.9025 |
| | SRCNN | 35.0406 | 0.8174 | 5.3457 | 0.0210 | 9.2205 |
| | VDSR | 35.0463 | 0.8180 | 5.1443 | 0. 0209 | 9.2813 |
| | EDSR | 35.2694 | 0.8227 | 5.1362 | 0.0204 | 9.0130 |
| | RCAN | 35.1122 | 0.8204 | 5.4764 | 0.0208 | 9.1662 |
| | 3D-FCNN | 35.2028 | 0.8196 | 5.0974 | 0.0207 | 9.0421 |
| | GDRRN | 35.2159 | 0.8234 | 5.1073 | 0.0205 | 9.0715 |
| | SSPSR | 35.4435 | 0.8280 | 4.8208 | 0.0201 | 8.7980 |
| | MCNet | 35.1757 | 0.8181 | 5.2134 | 0.0207 | 9.1226 |
| x8 | CEGATSR | 35.6894 | 0.8367 | 4.4190 | 0.0196 | 8.5199 |

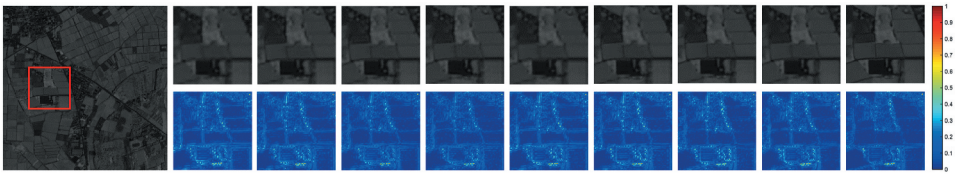

**Figure 7.** Reconstructed HSIs and their corresponding absolute error images of Chikusei dataset at the 31st bands for the scaling factors 4 by different compared methods. (1) SRCNN. (2) VDSR. (3) EDSR. (4) SCAN. (5) 3D-FCNN. (6) GDRRN. (7)SSPSR. (8)mcnet. (9)CEGATSR.

### *4.5. Comparison with state-of-the-art methods*

In this subsection, we evaluate the performance of the proposed CEGATSR on the test set for different scaling factors and compare it with eight existing SR methods, including four natural image SR methods, i.e. SRCNN (Dong et al. 2016), VDSR (Kim, Lee and Lee 2016), EDSR (Lim et al. 2017), RCAN (Zhang and Li 2018) and four single HSI SR methods, i.e. 3D-FCNN (Mei et al. 2017), GDRRN (Li et al. 2018), SSPSR (Jiang et al. 2020), MCNet (Li, Wang and Li 2020). For a fair and convincing comparison, we slightly adjust the parameters of

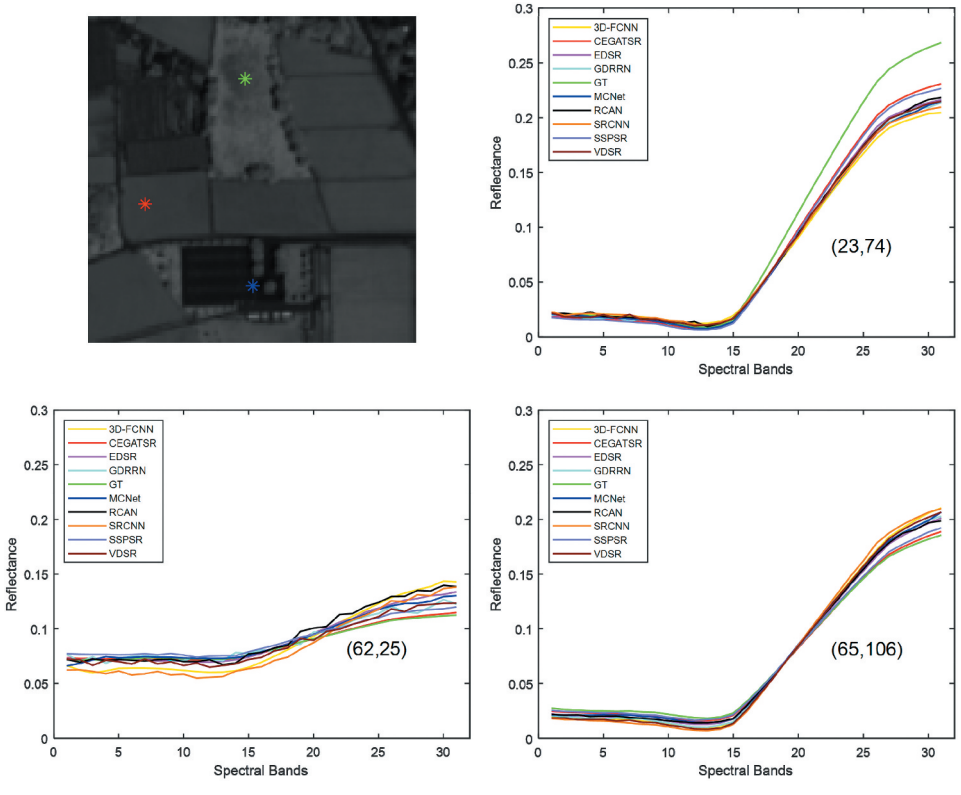


**Figure 8.** The spectral curves of three randomly selected positions ((23,74), (62,25), (65,106)) in the reconstructed Chikusei for the scaling factor undefined by different methods. The top right, bottom left and bottom right corresponds to the red, green and blue points respectively.

**Table 3.** Quantitative comparison results of different SR methods on Pavia Centre dataset for different scaling factors. The red and blue indicate the best and second-best performances, respectively.

| SF | Method | PSNR↑ | SSIM↑ | SAM↓ | RMSE↓ | ERGAS↓ |
|---|---|---|---|---|---|---|
| | SRCNN | 37.1826 | 0.9538 | 4.0224 | 0.0153 | 3.4437 |
| | VDSR | 37.8721 | 0.9583 | 3.1101 | 0.0143 | 3.2046 |
| | EDSR | 38.9764 | 0.9667 | 3.1638 | 0.0125 | 2.8044 |
| | RCAN | 38.9052 | 0.9670 | 3.2582 | 0.0126 | 2.8239 |
| | 3D-FCNN | 38.0512 | 0.9608 | 3.3600 | 0.0139 | 3.1872 |
| | GDRRN | 38.2656 | 0.9632 | 3.6129 | 0.0135 | 3.0510 |
| | SSPSR | 38.8462 | 0.9669 | 2.5448 | 0.0128 | 2.8491 |
| | MCNet | 38.9045 | 0.9675 | 2.3556 | 0.0127 | 2.8561 |
| x2 | CEGATSR | 39.4427 | 0.9705 | 2.2316 | 0.0119 | 2.6697 |
| | SRCNN | 32.4285 | 0.8469 | 4.0752 | 0.0265 | 6.0096 |
| | VDSR | 32.7387 | 0.8513 | 3.8803 | 0.0259 | 5.8392 |
| | EDSR | 33.1776 | 0.8625 | 3.8431 | 0.0245 | 5.5226 |
| | RCAN | 32.8388 | 0.8543 | 4.4538 | 0.0253 | 5.7319 |
| | 3D-FCNN | 32.4855 | 0.8467 | 3.7921 | 0.0263 | 6.0006 |
| | GDRRN | 32.9504 | 0.8588 | 3.9978 | 0.0251 | 5.6755 |
| | SSPSR | 33.2005 | 0.8635 | 3.0937 | 0.0246 | 5.5296 |
| | MCNet | 33.0445 | 0.8595 | 3.2660 | 0.0250 | 5.6315 |
| x4 | CEGATSR | 33.2975 | 0.8650 | 2.7119 | 0.0244 | 5.4716 |
| | SRCNN | 28.8311 | 0.6882 | 4.5369 | 0.0394 | 9.1522 |
| | VDSR | 28.8769 | 0.6827 | 3.8638 | 0.0396 | 9.1533 |
| | EDSR | 28.6959 | 0.6677 | 8.4210 | 0.0399 | 9.2585 |
| | RCAN | 28.1315 | 0.6669 | 9.1294 | 0.0421 | 9.8170 |
| | 3D-FCNN | 28.9528 | 0.6928 | 4.8761 | 0.0389 | 9.0463 |
| | GDRRN | 28.7122 | 0.6825 | 6.4792 | 0.0399 | 9.2904 |
| | SSPSR | 29.0846 | 0.6967 | 4.6124 | 0.0384 | 8.9011 |
| | MCNet | 28.9974 | 0.6929 | 5.9700 | 0.0387 | 9.0671 |
| x8 | CEGATSR | 29.1910 | 0.6997 | 3.8387 | 0.0381 | 8.8203 |

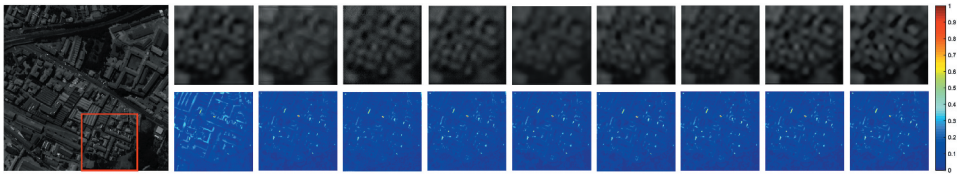

**Figure 9.** Reconstructed HSIs and their corresponding absolute error images of Pavia Centre dataset at the 31st bands for the scaling factors 4 by different compared methods. (1) SRCNN. (2) VDSR. (3) EDSR. (4) SCAN. (5) 3D-FCNN. (6) GDRRN. (7)SSPSR. (8)mcnet. (9)CEGATSR.

these methods and train them over our experimental dataset to get the best performance. The details and comparison results are described below.

### *4.5.1. Results on Chikusei dataset*

Table 2 shows the comparison results of five indices obtained by all compared methods on the Chikusei dataset for different scaling factors. We can easily observe that the proposed CEGATSR is superior to compared methods in all indices. The average PSNR value of the proposed CEGATSR is 0.1203 dB (undefined), 0.2894 dB (undefined) and 0.2459 dB (undefined) higher than the second-best method. For comparing with the four natural image SR methods, we can see that the proposed CEGATSR shows the best performance than them. The reason is four natural SR methods do not consider the correlation among spectral bands. For comparing with the four HSI SR methods, the

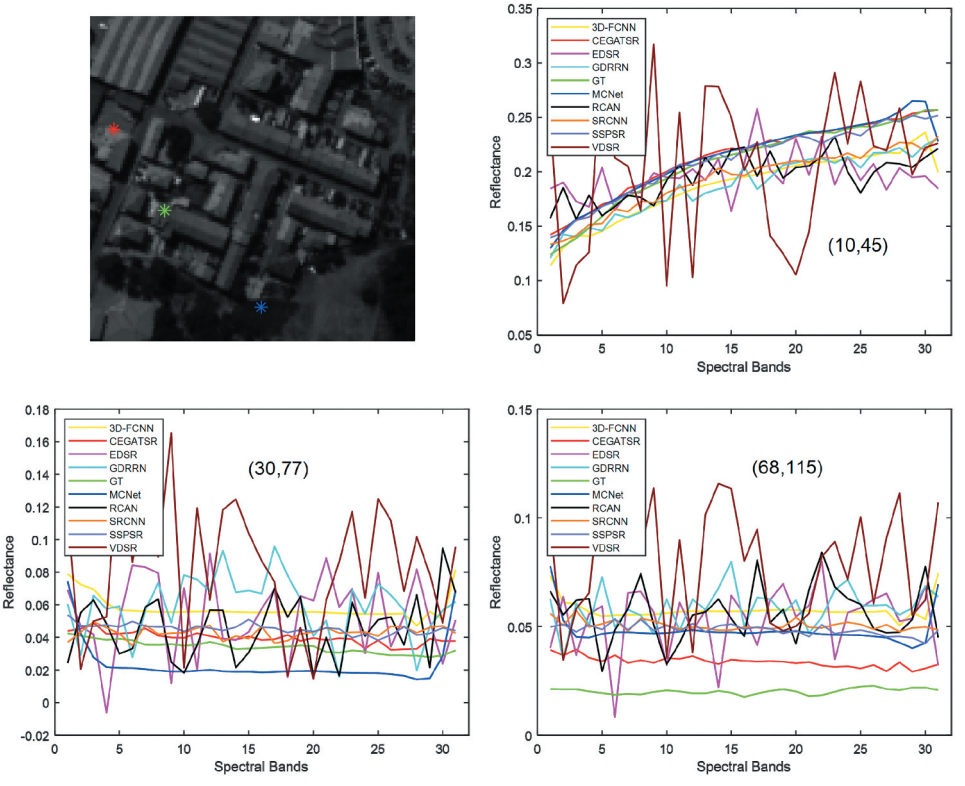


**Figure 10.** The spectral curves of three randomly selected positions ((10, 45), (30, 77), (68, 115)) in the reconstructed Pavia Centre dataset for the scaling factor undefined by different methods. The top right, bottom left, and bottom right correspond to the red, green and blue points, respectively.

proposed CEGATSR also shows the best performance than them. The reason might be that, the proposed CEGATSR can capture more useful feature information by using the non-local feature extraction unit and the local feature extraction unit.

Figure 7 shows the visual quality comparison results of the Chikusei dataset at the 31st spectral band for the scaling factor undefined. The reconstructed HSIs and their corresponding absolute error maps are listed in the first and the second rows, respectively. From the enlarged image in the red box region, we can see that the proposed CEGATSR produces clearer edges without obvious artefacts. As for the absolute error maps, the bluer the error map colour is, the closer the reconstruction effect is to the real image (the image has been normalized). It can be seen that CEGATSR is superior to the compared methods in restoring texture details, which is consistent with the analysis in Table 2. We also draw the spectral curve to visualize the spectral distortion of the reconstructed HSIs as shown in Figure 8. We randomly select three pixel positions (23,74), (62,25) and (65,106) to analyse the spectral distortion. Although the spectral curves of all methods are basically consistent with the real spectral curves of the original HR-HSI, the spectral curve obtained by the proposed CEGATSR is the closest one. This also proves that the proposed CEGATSR has better reconstruction performance in spectral correlation preservation.

#### 4.5.2. Results on Pavia Centre dataset

Table 3 shows the comparison results of five indices obtained by all compared methods on the Pavia Centre dataset for different scaling factors. We can observe that the proposed CEGATSR is better than compared methods in all indices. Compared with the second-best method, the PSNR values of the proposed CEGATSR are higher by 0.5382 dB, 0.0970 dB and 0.1064 dB for undefined, undefined and undefined, respectively.

Figure 9 shows the visual quality comparison results of the Pavia Centre dataset at the 31st spectral band for the scaling factor undefined. From this figure, we can see that the proposed CEGATSR can well reconstruct the texture details, such as the edge of the building and the trend of the street, while the reconstructed results of the compared methods are blurred. We also randomly select three points ((10, 45), (30, 77) and (68, 115)) from the red box region to show their spectral curves as shown in Figure 10. We can see that the spectral curves of all methods are consistent with that of the real terrain image, and the proposed CEGATSR achieves the best spectral fidelity. All the methods have a significant error with the real terrain image, which may be due to the limited training samples of the Pavia Centre data set, and the trained model does not have a good generalization ability.

#### 4.5.3. Results on Cave dataset

Table 4 reports the average performance over the Cave dataset obtained by all compared methods for different scaling factors. We can easily observe that the average PSNR values of the proposed CEGATSR are higher by 0.1305 dB(undefined), 0.2775 dB(undefined), 0.2173 dB(undefined) than that of the second-best methods for scaling factors undefined, undefined and undefined, respectively. The proposed CEGATSR performs the best performance in all indices.

We select a test image in the CAVE dataset for visual comparison. Figure 11 shows the reconstructed images and their corresponding error maps obtained by using various methods on *chart_and_stuffed_toy* at the 31st spectral band for the scaling factor undefined. From the reconstructed image, we can see that the proposed CEGATSR can reconstruct the detailed structure of the original image. It can also be seen from these error maps that the proposed CEGATSR achieves the best fidelity in terms of texture details. For example, the hair edges and the facial features of a doll are well reconstructed. In addition, we also randomly selected three positions (32,60), (72,19) and (84,125) of the reconstructed image and plotted their spectral curves shown as Figure 12 to show the spectral information. In most cases, the spectral curve of the proposed CEGATS is always closer to the spectral curve of the real ground image than that of other compared methods, which shows that our proposed CEGATS better retains the spectral correlation of the original HSI.

### 4.6. Analysis of the model complexity

In this subsection, we use three widely used indicators, the model parameters, the floating-point operations (FLOPs) and the memory access cost (MAC), to evaluate the efficiency of the proposed CEGATSR. These experiments are still performed on the Pavia Centre dataset for scaling factor undefined, and the experiment results obtained by using different methods are provided in Table 5.

**Table 4.** Quantitative comparison results of different SR methods on Cave dataset for different scaling factors. The red and blue indicate the best and second-best performances, respectively.

| SF | Method | PSNR↑ | SSIM↑ | SAM↓ | RMSE↓ | ERGAS↓ |
|---|---|---|---|---|---|---|
| | SRCNN | 43.4312 | 0.9872 | 3.1358 | 0.0106 | 3.8644 |
| | VDSR | 44.5017 | 0.9804 | 2.6498 | 0.0098 | 3.5711 |
| | EDSR | 44.7531 | 0.9824 | 2.6044 | 0.0894 | 3.4207 |
| | RCAN | 44.1311 | 0.9807 | 2.9954 | 0.0098 | 3.5526 |
| | 3D-FCNN | 43.9723 | 0.9813 | 2.7080 | 0.0101 | 3.1488 |
| | GDRRN | 44.2439 | 0.9811 | 2.3840 | 0.0098 | 3.1723 |
| | SSPSR | 45.1344 | 0.9849 | 2.1695 | 0.0088 | 2.8316 |
| | MCNet | 45.7919 | 0.9857 | 2.0575 | 0.0086 | 2.6795 |
| x2 | CEGATSR | 45.9224 | 0.9880 | 2.1403 | 0.0073 | 2.7226 |
| | SRCNN | 38.1583 | 0.9262 | 4.6237 | 0.0203 | 6.6231 |
| | VDSR | 39.2040 | 0.9313 | 4.0546 | 0.0189 | 6.0766 |
| | EDSR | 39.4937 | 0.9366 | 3.8830 | 0.0184 | 5.8954 |
| | RCAN | 39.3103 | 0.9340 | 4.4421 | 0.0184 | 5.9484 |
| | 3D-FCNN | 38.5321 | 0.9324 | 4.1252 | 0.0198 | 5.7299 |
| | GDRRN | 38.5782 | 0.9297 | 3.7352 | 0.0197 | 5.7729 |
| | SSPSR | 40.1508 | 0.9553 | 3.1209 | 0.0138 | 3.3384 |
| | MCNet | 40.2219 | 0.9452 | 3.1300 | 0.0175 | 4.9312 |
| x4 | CEGATSR | 40.4994 | 0.9590 | 3.0090 | 0.0168 | 4.7654 |
| | SRCNN | 34.1090 | 0.8469 | 6.1897 | 0.0320 | 9.8286 |
| | VDSR | 35.1556 | 0.8533 | 5.5035 | 0.0298 | 9.1628 |
| | EDSR | 34.2196 | 0.8369 | 7.0051 | 0.0317 | 9.8291 |
| | RCAN | 33.8253 | 0.8253 | 9.1824 | 0.0348 | 10.9709 |
| | 3D-FCNN | 34.7838 | 0.8584 | 6.4452 | 0.0303 | 9.0984 |
| | GDRRN | 33.6954 | 0.8413 | 6.9812 | 0.0325 | 10.2729 |
| | SSPSR | 35.8747 | 0.8772 | 5.1294 | 0.0275 | 8.3042 |
| | MCNet | 35.3834 | 0.8650 | 5.4852 | 0.0293 | 8.6461 |
| x8 | CEGATSR | 36.0920 | 0.8770 | 4.9347 | 0.0273 | 8.1876 |

First, we analyse the number of parameters. CEGATSR has fewer parameters than EDSR, RCAN, SSPSR and MCNet, while it has more parameters than SRCNN, VDSR, 3D-FCNN and GDRRN. The reason is that the last four compared methods contain fewer network layers than CEGATSR. Second, we analyse the floating-point operations (FLOPs). The proposed CEGATSR is only better than SSPSR and MCNet, while it is larger than the rest six compared methods. It is probably that the graph attention layer of the proposed CEGATSR is operated at the pixel level, which needs a large number of flops. Third, we analyse the memory access cost (MAC). CEGATSR has more memory advantage than 3D-FCNN, GDRRN and MCNet. The reason is that CEGATSR uses an improved depthwise separable convolution to capture the spatial-spectral information, reducing the memory overhead.

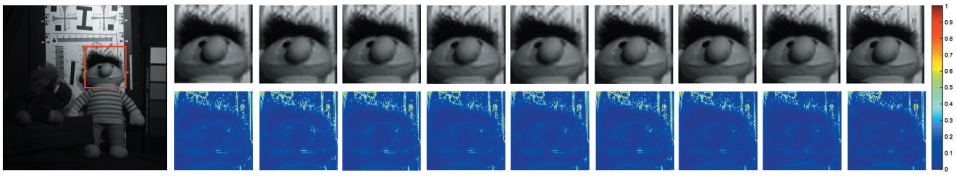

**Figure 11.** Reconstructed HSIs and their corresponding absolute error images of *chart_and_stuffed_toy* at the 31st bands for the scaling factors 4 by different compared methods. (1) SRCNN. (2) VDSR. (3) EDSR. (4) SCAN. (5) 3D-FCNN. (6) GDRRN. (7)SSPSR. (8)mcnet. (9)CEGATSR.

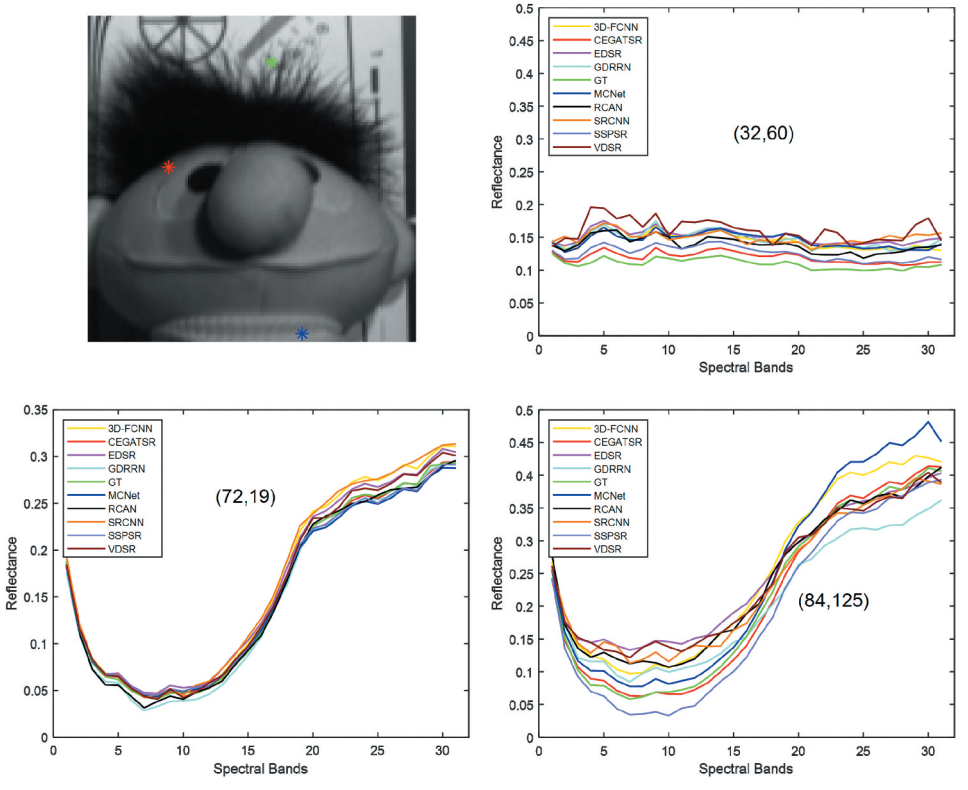


**Figure 12.** The spectral curves of three randomly selected positions ((32, 60), (72, 19), (84, 125) ✶ in the reconstructed *chart_and_stuffed_toy* for the scaling factor undefined by different methods.

**Table 5.** Complexity comparison of different methods on the Pavia Centre dataset. The network uses 64 filters for scaling factor ×8.

| Scaling factor | Method | Params(K) | FLOPs(M) | MAC(M) |
|---|---|---|---|---|
| x8 | SRCNN | 187.55 | 48.01 | 1033 |
| | VDSR | 699.26 | 179.01 | 857 |
| | EDSR | 2874.24 | 1703.26 | 1414 |
| | RCAN | 15,772.35 | 4980.57 | 1160 |
| | 3D-FCNN | 41.77 | 302.61 | 4700 |
| | GDRRN | 109.44 | 179.01 | 3508 |
| | SSPSR | 1825.64 | 7635.57 | 2270 |
| | MCNet | 2960.51 | 638,656.27 | 7372 |
| | CEGATSR | 1113.59 | 5501.02 | 2979 |

## 5. Conclusion

In this paper, a novel convolution neural network named CEGATSR has been proposed for HSI super-resolution. To make full of the inner structure of the insufficient training samples in the HSI SR task, we develop a parallel feature extraction unit by combining a non-local feature extraction unit and a local feature extraction unit. The non-local feature extraction unit employs the graph attention block (GAB) to explore non-local features and self-learn the similar structure in an HSI. The local feature extraction unit

applies the depthwise separable convolution block (DSCB) to extract the local texture details. Moreover, in order to future utilize the global feature information, we have presented a spatial-channel attention block (SCAB) by combining a two-branch parallel structure and a channel attention module. Extensive experiments on public hyperspectral image datasets SR show the effectiveness of our CEGATSR in terms of quantitative and visual results.

In the future, we plan to extend our model in two aspects. Firstly, in the graph attention block (GAB), the information aggregation in the cross-scale direction of image patches is not fully utilized, so this can effectively use the cross-scale information between different scaling factors of image patches to improve the structure of the network. Second, the paper uses channel attention to capture the robust correlation information between spectral bands. In the future, it is necessary to focus on the features between bands to mine spectral information fully.

## Acknowledgment

The authors would like to appreciate all anonymous reviewers for their insightful comments and constructive suggestions to polish this paper in high quality. This research was supported by the National Natural Science Foundation of China (No.61703278).

## Disclosure statement

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

## Funding

The work was supported by the National Natural Science Foundation of China [No.61703278]

## 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, 3631–3640.

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 Recognition* 88: 431–446. doi:10.1016/j.patcog.2018.11.033.

Chollet, F. 2017. "Xception: Deep Learning with Depthwise Separable Convolutions." In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR),* Honolulu, HI, 1800–1807 .

Dai, T., J. Cai, Y. Zhang, S. Xia, and L. Zhang. 2019. "Second-Order Attention Network for Single Image Super-Resolution." In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR),* Long Beach, CA, , 11057–11066.

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, 3862–3871.

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. doi:10.1109/TNNLS.2018.2885616.

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. doi:10.1109/TIP.2016.2542360.

Dong, W., Z. Lei, G. Shi, and X. Wu. 2011. "Image Deblurring and Super-Resolution by Adaptive Sparse Domain Selection and Adaptive Regularization." *IEEE Transactions on Image Processing* 20 (7): 1838–1857. doi:10.1109/TIP.2011.2108306.

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. doi:10.1109/TPAMI.2015.2439281.

Dong, C., C. C. Loy, and X. Tang. 2016. "Accelerating the Super-Resolution Convolutional Neural Network." In *2016 European Conference on Computer Vision (ECCV),* Amsterdam, The Netherlands, 391–407.

Fu, Y., Z. Liang, and S. You. 2021. "Bidirectional 3D Quasi-Recurrent Neural Network for Hyperspectral Image Super-Resolution." *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 14: 2674–2688. doi:10.1109/JSTARS.2021.3057936.

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. doi:10.1109/TIP.2020.2968773.

He, K., X. Zhang, S. Ren, and J. Sun. 2016. "Deep Residual Learning for Image Recognition." In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR),* Las Vegas, NV, 770–778.

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." In *2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS),* Beijing, China, 6962–6965.

Howard, A. G., M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam. 2017. "MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications." *CoRr,* abs/1704.04861 .

Huang, H., J. Yu, and W. Sun. 2014. "Super-Resolution Mapping via Multi-Dictionary Based Sparse Representation." In *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP),* Florence, Italy, 3523–3527.

Hu, J., X. Jia, Y. Li, G. He, and M. Zhao. 2020. "Hyperspectral Image Super-Resolution via Intrafusion Network." *IEEE Transactions on Geoscience and Remote Sensing* 58 (10): 7459–7471. doi:10.1109/TGRS.2020.2982940.

Hu, J., Y. Li, and W. Xie. 2017. "Hyperspectral Image Super-Resolution by Spectral Difference Learning and Spatial Error Correction." *IEEE Geoscience and Remote Sensing Letters* 14 (10): 1825–1829. doi:10.1109/LGRS.2017.2737637.

Hung, K. and W. Siu. 2012. "Robust Soft-Decision Interpolation Using Weighted Least Squares." *IEEE Transactions on Image Processing* 21 (3): 1061–1069. doi:10.1109/TIP.2011.2168416.

Hu, J., L. Shen, S. Albanie, G. Sun, and E. Wu. 2020. "Squeeze-And-Excitation Networks." *IEEE Transactions on Pattern Analysis and Machine Intelligence* 42 (8): 2011–2023. doi:10.1109/TPAMI.2019.2913372.

Hu, J., M. Zhao, and Y. Li. 2019. "Hyperspectral Image Super-Resolution by Deep Spatial-Spectral Exploitation." *Remote Sensing* 11 (10): 1229. doi:10.3390/rs11101229.

Irmak, H., G. B. Akar, and S. E. Yuksel. 2018. "A MAP-Based Approach for Hyperspectral Imagery Super-Resolution." *IEEE Transactions on Image Processing* 27 (6): 2942–2951. doi:10.1109/TIP.2018.2814210.

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. doi:10.1109/TCI.2020.2996075.

Kim, J., J. K. Lee, and K. M. Lee. 2016. "Accurate Image Super-Resolution Using Very Deep Convolutional Networks." In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR),* Las Vegas, NV, 1646–1654.

Kingma, D. and J. Ba. 2014. "Adam: A Method for Stochastic Optimization." In *2014 International Conference on Learning Representations*, San Diego, CA, 1–15.

Lai, W., J. Huang, N. Ahuja, and M. Yang. 2017. "Deep Laplacian Pyramid Networks for Fast and Accurate Super-Resolution." In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR),* Honolulu, HI, 5835–5843.

Ledig, C., L. Theis, F. Huszár, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, et al. 2017. "Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network." In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR),* Honolulu, HI, 105–114.

Li, Y., J. Hu, X. Zhao, W. Xie, and J. Li. 2017. "Hyperspectral Image Super-Resolution Using Deep Convolutional Neural Network." *Neurocomputing* 266 (29): 29–41. doi:10.1016/j.neucom.2017.05.024.

Lim, B., S. Son, H. Kim, S. Nah, and K. M. Lee. 2017. "Enhanced Deep Residual Networks for Single Image Super-Resolution." In *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW),* Honolulu, HI, 1132–1140.

Liu, D., J. Li, and Q. Yuan. 2021. "A Spectral Grouping and Attention-Driven Residual Dense Network for Hyperspectral Image Super-Resolution." *IEEE Transactions on Geoscience and Remote Sensing* 59 (9): 7711–7725. doi:10.1109/TGRS.2021.3049875.

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. doi:10.1109/JSTARS.2015.2459375.

Li, Q., Q. Wang, and X. Li. 2020. "Mixed 2D/3D Convolutional Network for Hyperspectral Image Super-Resolution." *Remote Sensing* 12 (10): 1660. doi: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. doi:10.1109/TGRS.2020.3047363.

Li, Y., L. Zhang, C. Ding, W. Wei, and Y. Zhang. 2018. "Single Hyperspectral Image Super-Resolution with Grouped Deep Recursive Residual Network." In *2018 IEEE Fourth International Conference on Multimedia Big Data (BigMM),* Xi'an, China, 1–4.

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. doi:10.3390/rs9111139.

Pike, R., G. Lu, D. Wang, Z. 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. doi:10.1109/TBME.2015.2468578.

Qu, Y., H. Qi, and C. Kwan. 2018. "Unsupervised Sparse Dirichlet-Net for Hyperspectral Image Super-Resolution." In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition,* Salt Lake City, UT, 2511–2520.

Rasti, P., T. Uiboupin, S. Escalera, and G. Anbarjafari. 2016. "Convolutional Neural Network Super Resolution for Face Recognition in Surveillance Monitoring." *Articulated Motion and Deformable Objects* 9756: 175–184.

Ren, C., X. He, Y. Pu, and T. Q. Nguyen. 2019. "Enhanced Non-Local Total Variation Model and Multi-Directional Feature Prediction Prior for Single Image Super Resolution." *IEEE Transactions on Image Processing* 28 (8): 3778–3793. doi:10.1109/TIP.2019.2902794.

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 and Remote Sensing* 60: 1–13.

Valsesia, D., G. Fracastoro, and E. Magli. 2020. "Deep Graph-Convolutional Image Denoising." *IEEE Transactions on Image Processing* 29: 8226–8237. doi:10.1109/TIP.2020.3013166.

Velickovic, P., G. Cucurull, A. Casanova, A. Romero, P. Liò, and Y. Bengio. 2018. "Graph Attention Networks." In *2018 International Conference on Learning Representations,* Vancouver, BC, 1–12.

Wang, Z., A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli. 2004. "Image Quality Assessment: From Error Visibility to Structural Similarity." *IEEE Transactions on Image Processing* 13 (4): 600–612.

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.

Wang, Q., Q. Li, and X. Li. 2020. "Spatial-Spectral Residual Network for Hyperspectral Image Super-Resolution." *CoRR* abs/2001.04609.

Wang, Q., Q. Li, and X. Li. 2021. "Hyperspectral Image Super-Resolution Using Spectrum and Feature Context." *IEEE Transactions on Industrial Electronics* 68 (11): 11276–11285. doi:10.1109/TIE.2020.3038096.

Wang, X., J. Ma, and J. Jiang. 2021. "Hyperspectral Image Super-Resolution via Recurrent Feedback Embedding and Spatial-Spectral Consistency Regularization." *IEEE Transactions on Geoscience and Remote Sensing* 60: 1–13.

Yang, Y. and Y. Qi. 2021. "Image Super-Resolution via Channel Attention and Spatial Graph Convolutional Network." *Pattern Recognition* 112: 107798. doi:10.1016/j.patcog.2020.107798.

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. doi:10.1109/TIP.2021.3077135.

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. doi:10.1109/MGRS.2016.2637824.

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. doi:10.1109/TGRS.2011.2161320.

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. doi:10.1109/TGRS.2020.3018732.

Zhang, Y. and K. Li. 2018. "Image Super-Resolution Using Very Deep Residual Channel Attention Networks." In *2018 European Conference on Computer Vision (ECCV),* Munich, Germany, 294–310.

Zhou, S., J. Zhang, W. Zuo, and C. C. Loy. 2020. "Cross-Scale Internal Graph Neural Network for Image Super-Resolution." *Advances in Neural Information Processing Systems* 33: 3499–3509.

Zhu, Z., J. Hou, J. Chen, H. Zeng, and J. Zhou. 2021. "Hyperspectral Image Super-Resolution via Deep Progressive Zero-Centric Residual Learning." *IEEE Transactions on Image Processing* 30: 1423–1438. doi:10.1109/TIP.2020.3044214.