Original Research Paper

Spectral recovery-guided hyperspectral super-resolution using transfer learning

SRTL for HSISR

Shaolei Zhang | Guangyuan Fu | Hongqiao Wang | Yuqing Zhao

Department of Information Engineering, Xi’an High-Tech Research Institution, Xi’an 710025, China

Correspondence
Shaolei Zhang, Xi’an High-tech Research Institution, Xi’an 710025, China.
Email: jianjianyuanqu11@163.com

Funding information
National Natural Science Foundation of China, Grant/Award Number: 61403397; Natural Science Basic Research Plan in Shaanxi Province of China, Grant/Award Number: 2020JM-358

Abstract
Single hyperspectral image (HSI) super-resolution (SR) has attracted researcher’s attention; however, most existing methods directly model the mapping between low- and high-resolution images from an external training dataset, which requires large memory and computing resources. Moreover, there are few such available training datasets in real cases, which prevent deep-learning-based methods from further improving performance. Here, a novel single HSI SR method based on transfer learning is proposed. The proposed method is composed of two stages: spectral down-sampled image SR reconstruction based on transfer learning and HSI reconstruction via spectral recovery module. Instead of directly applying the learned knowledge from the colour image domain to HSI SR, the spectral down-sampled image is fed into a spatial SR model to obtain a high-resolution image, which acts as a bridge between the colour image and HSI. The spectral recovery network is used to restore the HSI from the bridge image. In addition, pre-training and collaborative fine-tuning are proposed to promote the performance of SR and spectral recovery. Experiments on two public HSI datasets show that the proposed method achieves promising SR performance with a small paired HSI dataset.

1 | INTRODUCTION

Hyperspectral imaging technology captures spectral reflectance and spatial information of an imaging scene simultaneously, reflecting the objects’ nature comprehensively. The high resolution in the spectral domain makes objects with similar colour or textures distinguishable. Therefore, hyperspectral images (HSIs) have been widely applied in remote sensing and computer vision, such as land cover mapping [1], environmental protection [2], and anomaly detection [3]. However, the limitations in sensor technology make it essential that there be a trade-off between spatial, temporal, and spectral resolution to guarantee a sufficient signal-to-noise ratio. Compared with conventional images, HSIs have high spectral resolution but low spatial resolution, which is a disadvantage in their application. Therefore, many efforts have been made to obtain high-resolution HSIs.

Image super-resolution (SR) is a challenging computer-vision problem that aims to enhance the resolution of low-resolution images. Regarding HSI, most existing works concentrate on improving the spatial resolution, but spectral super-resolution (SSR), which reconstructs HSIs from colour images, has also attracted researcher’s attention recently. As to whether auxiliary images are needed or not, most HSI SR works are divided into fusion-based and single-image-based methods. Fusion-based methods adopted dictionary learning, matrix factorization, or tensor factorization with handcrafting prior to predicting high-resolution HSIs from low-resolution HSIs and corresponding auxiliary images. However, it is difficult to obtain auxiliary images in many scenes. Single-image-based image methods learn the mapping between low- and high-resolution images from an external image training dataset and directly predict the desired high-resolution HSI from the low-resolution one by the learned model. Early works adopted traditional methods, such as sparse coding and manifold learning, to learn the mapping, which have limited feature extraction and representation capability. Therefore, these methods achieve promising
performance only when super-resolving some images with simple structure.

With the success of deep learning in various computer-vision tasks [4], deep-learning-based methods have been proposed to solve the problem. The super-resolution convolutional neural network (SRCNN) [5] was the first method used to apply a deep network for colour image SR. Later, increasingly more complex networks were proposed to improve SR performance. Liebel et al. [6] directly applied the SRCNN to multi-spectral image SR in the band-by-band manner, which achieves good performance but exhibits severe spectral distortion. Therefore, several deep networks, such as SSIN [7] and DRAN [8], have been specially designed for HSI SR, and they exhibit state-of-the-art SR performance by integrating residual learning, dense connections, and attention mechanism. However, these complex networks have high temporal and spatial complexity, and demand a large training dataset to train the network well.

In addition, several attempts have been made to solve HSI SR by transfer learning [9–11]. In these works, it is assumed that the mapping between low- and high-resolution HSIs is the same as that between low- and high-resolution colour images. The mapping in the colour image domain learned by CNNs is transformed to the HSI domain in these works. However, the large gap between the source and target domains led to these works needing post-processing to explore spectral characteristics and prevent spectral distortion.

To overcome these problems, a novel HSI SR method based on transfer learning is proposed in the present paper. Specifically, the proposed method involves three steps, as shown in Figure 1c. First, low-resolution HSIs are spectrally down-sampled to low-resolution colour images by applying a spectral response function (SRF). Then, the spectrally down-sampled colour image is fed into the pre-trained spatial SR model to predict the high-resolution image. Finally, the desired high-resolution HSI is obtained from the high-resolution image by spectral recovery module. With the spectrally down-sampled image as a bridge, the proposed method can effectively utilize the knowledge from the colour image domain and reduce computational complexity compared with learning the mapping between low- and high-resolution HSIs directly. The proposed method consists of two parts: spatial SR module and spectral recovery module. The spectral recovery module is trained with low-resolution HSIs and spectrally down-sampled images, and the spatial SR model is trained with colour image training dataset. To avoid overfitting, the spatial SR super-resolution module and spectral recovery module are fine-tuned with available HSIs.

The main contributions of this paper are as follows:
1. A novel method is proposed for HSI SR based on transfer learning, in which the spatial module models the mapping between low- and high-resolution HSIs, and the spectral module models the mapping between spectrally down-sampled images and HSIs.
2. Spectral down-sampling and recovery is adopted to reduce computational complexity as well as the gap between source and target domains, facilitating the transfer of learned knowledge, which improves SR performance.
3. The proposed method is flexible and able to adopt any spectral recovery and SR model. Owing to the end-to-end manner in which it operates, it can utilize collaborative fine-tuning to further improve performance.

2 | PROPOSED METHOD

Directly predicting high-resolution HSIs from low-resolution ones has high computational complexity and demands sufficient samples for training network. Meanwhile, the gap between HSI and colour images makes knowledge transfer from the colour image domain to the HSI domain difficult. To solve these problems, a novel SR method based on transfer learning is proposed, the overall structure of which is shown in Figure 1c.

2.1 | Problem formulation

Here, low-resolution and high-resolution HSIs are denoted $X \in \mathbb{R}^{(m \times n \times B)}$ and $Y \in \mathbb{R}^{(M \times N \times B)}$, respectively, where
$m$, $n$, $M$, and $N$ denote the width and height of low-resolution and high-resolution images, respectively, and $B$ represents the number of bands. $X_C \in \mathbb{R}^{(M \times N \times C)}$ denotes spectrally down-sampled low-resolution images and $Y_C \in \mathbb{R}^{(M' \times N' \times C)}$ spectrally down-sampled high-resolution images. As shown in Figure 1a, deep-learning-based methods usually directly learn an end-to-end mapping $Y = F(X)$ from paired low- and high-resolution images, and the learned mapping $F$ predicts the desired high-resolution HSIs from low-resolution ones in the test stage.

As is well known, HSIs contain abundant spectral information. Specifically, the HSI for typical terrestrial applications usually have more than 30 bands, while those for airborne or spaceborne remote sensing applications have 200 or more bands. The SR network relies on its powerful feature-learning capability to achieve promising performance; for example, the number of intermediate features generated by convolution layers is approximately 20 times or more that of colour image channels. When using the same settings for super-resolving HSIs, the memory consumption of deep networks is huge, especially for satellite or airborne remote sensing HSIs. For example, there are $31 \times 16 = 496$ convolution features in SSIN [7] when super-resolving HSIs with 31 bands. At the same time, training such a network with insufficient samples is prone to overfitting. However, it is more difficult to acquire paired HSI data compared with colour image data.

To solve these problems, spectral recovery-guided HSI SR based on transfer learning is proposed. As shown in Figure 1c, the proposed method consists of two stages: (1) spatial SR based on transfer learning and (2) spectral reconstruction. Given paired HSIs, a SRF is first applied to the HSIs to synthesize colour images. Then, the spatial SR network takes synthesized colour images as input to predict high-resolution colour images, which is fed into the SSR network to obtain the desired high-resolution HSIs. The spectral recovery network is pre-trained with synthesized colour images and corresponding HSIs, and the spatial SR network is pre-trained with colour images to prevent overfitting.

### 2.2 SR using transfer learning

Several studies have shown that the increase in network depth and width is conducive to improving performance of many computer-vision tasks, but it means that more training data are demanded to fully train the network model. However, it is difficult to obtain enough labelled data in many tasks, for example, HSI SR. Transfer learning can alleviate the problem of lack of training data through cross-domain knowledge transfer. Recently, several attempts [9,10] were made to improve HSI SR by transferring knowledge from the colour image domain to the HSI domain. As shown in Figure 1b, in these works high-resolution HSIs are usually predicted through use of a pre-trained CNN in a band-by-band manner. To avoid the spectral distortion, post-processing, such as collaborative non-negative matrix factorization (CNMF), was utilized to exploit the spectral correlation between low-resolution and high-resolution HSIs.

To reduce the gap between the HSI and source domains, such as colour images or multi-spectral image, different from the literature [9,10] in which learned knowledge is transferred directly to the target domain, the proposed method utilizes a pre-trained CNN to super-resolve an intermediate image, namely, a spectrally down-sampled image. As shown in Figure 2, given HSI $X$, first a synthetic colour image is generated with the SRF, which is the input to predict a high-resolution image using a learned model. The spectral down-sampling process can be defined as

$$X_C = X \times \Phi_{srf},$$  \hspace{1cm} (1)

where $\Phi_{srf}$ denotes the SRF.

The release of benchmark datasets for colour SR allows design of more complex networks, resulting in deep networks having more powerful representation and generalization capabilities [9]. Existing super-resolution models, which usually learn the end-to-end mapping from low-resolution images to high-resolution images in a supervised training manner, are summarized in [12]. Common objective functions include L1 loss, L2 loss, and perceptual loss [13], among which L1 loss has better performance. In the present paper, the SAN [14] network is utilized to enhance the resolution of a spectrally down-sampled image $X_C$. As shown in Figure 3, the main part of the network is the region-level non-local module (RL-NL), which explores the abundant structure information and self-similarities. The SAN adopts pixel shuffling to enhance the resolution, and as it belongs to a post-upsampling framework, the computational burden is reduced. For discriminative feature-learning, second-order channel attention (SOCA) was proposed to learn feature interdependencies by second-order statistics. Given a colour image dataset $T_C = \{(X_C, Y_C), i = 1, 2, \ldots, N\}$, the deep network is trained as follows:

$$L_t (\Theta) = \frac{1}{N} \sum_{i=1}^{N} \|H_i - F_{SAN} (L_i)\|_1,$$  \hspace{1cm} (2)

where $\Theta$ denotes network parameters, $N$ the number of colour training samples, and $F_{SAN}(\#)$ the CNN.

The well-trained SAN can be used directly to enhance the resolution of spectrally down-sampled images. However, if some paired HSIs $(X, Y)$ exist, then the pre-trained model can be fine-tuned with the spectrally down-sampled image dataset $\{T_C = (X_C, Y_C), i = 1, 2, \ldots, Z\}$, which would provide better performance on synthetic colour images. The fine-tuning process retains the same setting with pre-training, and the fine-tuned network $F_{SR}$ is used to predict the high-resolution intermediate image:

$$\widetilde{Y}_C = F_{SR} (X_C),$$  \hspace{1cm} (3)

where $\widetilde{Y}_C$ denotes the predicted high-resolution image and $X_C$ the spectrally down-sampled low-resolution image.
The goal of the present work is to predict high-resolution HSIs from low-resolution HSIs. To reduce the gap between source and target domains, the original low-resolution HSI was firstly down-sampled along spectral dimensions. Therefore, only a high-resolution spectrally down-sampled image $\tilde{Y}$ was obtained by the SR network, which, acting as a bridge, the high-resolution HSI would be recovered by the SSR network.

Spectral recovery aims to recover a HSI from a RGB image. Early attempts concentrated on an appropriate prior with the sparse dictionary or radial basis network. With the development of CNNs, they have also been used to solve the problem. Galliani et al. [15] was the first to propose an end-to-end recovery of hyperspectral reflectance from RGB values using a CNN based on the Tiramisu network without the need for spectral response function. Since then, the NTIRE2018 and NTIRE 2020 spectral reconstruction challenges have been held, which offer large-scale, uniform benchmarks for the SSR problems. Many deep-learning-based methods achieve state-of-the-art performance with sufficient training samples. The adaptive weighted attention network (AWAN) [16] won first place in the ‘clean’ track of the NTIRE 2020 spectral reconstruction challenge. Therefore, the AWAN is used as the spectral recovery module in the present work.

Both parts of the proposed method adopt a deep network, meaning that the final learned mapping from low-resolution to high-resolution HSIs is end-to-end and we can train them jointly, which is another advantage compared with these works [9, 10].

The AWAN is shown in Figure 4. Similar to the SAN, the AWAN also includes shallow feature extraction and deep feature extraction. The attention mechanism was also introduced to capture the correlation between channels. Different from existing attention models [8, 14], adaptive weighted channel attention (AWCA) was proposed to learn the adaptive weight matrix. The patch-level second-order non-local module (PSNL) was proposed to explore the long-range interdependency, which
The framework of AWAN [16]

**Algorithm 1** Pre-training and collaborative fine-tuning for proposed model

**Input:** colour image \( T_p \), HSI \( T \), SRF \( F \)

**Output:** the well-trained network

1. Initialize learning rate, epochs for pre-training and fine-tuning, batch size etc.
2. Generate synthetic image \( T_c \) by applying SRF \( F \) on HSI \( T \) along spectral dimensions.
3. Pre-train spatial SR network with loss in Equation (2) on datasets \( T_p \) and \( T_c \).
4. Pre-train spectral recovery network with loss in Equation (5) on datasets \( T_s \) and \( T_s \) consists of \((X_C, X)\).
5. Fine-tune spatial SR network and spectral recovery network in collaborative way.

combines non-local operation and second-order statistics to learn discriminative representations. The object function of AWAN is defined as

\[
L_{\text{AWAN}} = \frac{1}{N} \sum_{p=1}^{N} \left( \| I_{\text{HSI}}^p - I_{\text{SR}}^p \|_1 \right)
+ \lambda \left( \Phi \| I_{\text{HSI}}^p - I_{\text{SR}}^p \|_1 \right),
\]

where \( N \) is the number of training samples.

Different from the AWAN training network with an external dataset, here the AWAN is trained with paired \((X_C, X)\), which are spectrally down-sampled images and HSIs, respectively. The \( X_i \) were generated using SRF according to Equation (1). In addition, L1 loss and spectral information divergence (SID) are used as object functions which jointly train the network in a supervised manner. The L1 loss is used to reduce the pixel loss of the reconstruction image and the reference image, and SID loss suppresses the spectral distortion between the reconstruction and reference images. The L1 and SID losses can be calculated as:

\[
L(\Phi) = L_1(\hat{X}_s, X) + L_{\text{SID}}(\hat{X}_s, X),
\]

\[
L_1(\hat{X}_s, X) = \frac{1}{N} \sum_{i=1}^{N} \| \hat{X}_i - X_i \|_1,
\]

\[
L_{\text{SID}}(\hat{X}_s, X) = \frac{1}{N} \sum_{i=1}^{N} \left( D(\hat{X}_i \| X) + D(X_i \| \hat{X}) \right),
\]

\[
D(\hat{X}_i \| X) = \hat{X}_i \log \left( \frac{\hat{X}_i + \varepsilon_2}{\hat{X}_i + \varepsilon_2} \right) + (X_i + \varepsilon_2),
\]

\[
D(X_i \| \hat{X}) = X_i \log \left( \frac{X_i + \varepsilon_2}{\hat{X} + \varepsilon_2} \right),
\]

where \( \Phi \) represents the network, \( N \) is the number of training samples, and \( \hat{X}_s, X_i \) are the predicted HSI and ground-truth HSI, respectively.
TABLE 1  Comparison of different methods on balloons from CAVE dataset

| Method | BICUBIC | SRCNN | TL-CNMF | SSIN | Proposed |
|--------|---------|-------|--------|------|----------|
| Scale factor = 3x | | | | | |
| RMSE   | 3.8351  | 3.1627 | 2.4497 | \textbf{1.8042} | 1.9115   |
| MPSNR  | 35.9812 | 38.1173| 40.4205| 41.0747| \textbf{41.0834}|
| MSSIM  | 0.9702  | 0.9843 | 0.9892 | \textbf{0.9928} | 0.9912   |
| SAM    | 1.8056  | 3.2383 | 1.4798 | \textbf{1.2721} | 1.3012   |
| Scale factor = 4x | | | | | |
| RMSE   | 4.5152  | 3.5627 | 2.8835 | 2.3304 | \textbf{2.2884}|
| MPSNR  | 35.0105 | 36.5173| 38.5632| 40.0802| \textbf{40.2045}|
| MSSIM  | 0.9662  | 0.9743 | 0.9848 | \textbf{0.9908} | 0.9892   |
| SAM    | 2.2719  | 3.2383 | 1.8162 | 1.5635 | \textbf{1.3566}|

TABLE 2  Comparison of different methods on imgb8 from Harvard dataset

| Method | BICUBIC | SRCNN | TL-CNMF | SSIN | Ours |
|--------|---------|-------|--------|------|------|
| Scale factor = 3x | | | | | |
| RMSE   | 7.1536  | 6.2918 | 7.0698 | 6.2927 | \textbf{5.5404}|
| MPSNR  | 38.8880 | 31.6707| 31.9739| 32.7622| \textbf{33.9173}|
| MSSIM  | 0.8392  | 0.8702 | 0.8438 | 0.8415 | \textbf{0.8943}|
| SAM    | 3.4994  | 4.4759 | 2.8256 | 2.8992 | \textbf{2.2603}|
| Scale factor = 4x | | | | | |
| RMSE   | 8.1431  | 8.2222 | 8.1095 | \textbf{7.1569} | 7.2018   |
| MPSNR  | 30.8179 | 29.4999| 30.7162| 31.2394| \textbf{31.7756}|
| MSSIM  | 0.7882  | 0.8014 | 0.8030 | 0.8024 | \textbf{0.8337}|
| SAM    | 4.0353  | 5.2108 | 3.3635 | 3.5773 | \textbf{2.5243}|

2.4  | Pre-training and collaborative fine-tuning

2.4.1  | Training process

The training for the proposed network includes two stages: pre-training and collaborative fine-tuning. For pre-training, the SAN network is trained on a color image benchmark dataset with the loss function shown in Equation (2), and the spectral recovery network is trained on the synthesized data (X̃, X). For collaborative fine-tuning, the method proposed in this paper is divided into two stages, including spatial SR and spectral recovery. The results of the spatial SR model can give feedback to the spectral recovery process, making transferred knowledge towards intermediate images, which in turn promotes and improves spectral recovery performance.

2.4.2  | Testing process

The spectrally down-sampled image is generated by applying the SRF to low-resolution HSIs, which are then passed sequentially through the spatial SR network and the SSR network to generate the high-resolution HSIs.

3  | EXPERIMENTS AND ANALYSIS

3.1  | Dataset and evaluation

3.1.1  | Dataset

The spatial SR network is pre-trained with the DIV2K [17] benchmark dataset, which contains 1000 images of different scenes and is divided into training, validation, and testing in the ratio {800, 100, 100}. The CAVE [18] dataset contains 32
images showing everyday objects with a spatial resolution of 512 × 512. Each image has 31 bands, covering the spectral range from 400 to 700 nm. The Harvard [19] dataset contain 50 images including indoor and outdoor scenes, representing both natural and artificial objects. Each image also has 31 bands, covering the spectral range from 420 to 720 nm. The spatial resolution is 1040 × 1392, and the upper left 1024 × 1024 pixel of each image was cropped in the experiments. The CAVE and Harvard datasets are common benchmark datasets for HSI SR. Three images were randomly selected from each dataset as test images, as shown in Figure 5, and the rest were used for fine-tuning and validation. The original HSIs in the datasets were used as ground truth, and the low-resolution HSIs are simulated by the Gaussian degradation model and bicubic down-sampling function. In the experiments, all of the images were normalized to [0, 1].

3.1.2 Evaluation metrics

Four common metrics, root-mean-square error (RMSE), spectral angle mapping (SAM), mean peak-signal-to-noise ratio (MPSNR), and mean structural similarity (MSSIM), were adopted to evaluate the reconstruction results. The larger the MPSNR, MSSIM, and SAM, the better the SR performance; the smaller the RMSE, the better the SR performance.

3.2 Implementation details

The proposed method is flexible and thus able to adopt any deep network models. In the work described in this paper, the SAN [14] was adopted for the first stage, and was pre-trained on the DIV2K dataset with the Adam optimizer. The
Adam parameters were $\beta_1 = 0.9$ and $\beta_2 = 0.99$ and the initial learning rate $1 \times 10^{-4}$. The AWAN [16] was used for the second stage, and was pre-trained on the CAVE and Harvard datasets (except for test images) with the Adam optimizer. The Adam parameters were $\beta_1 = 0.9$ and $\beta_2 = 0.99$, and the initial learning rate $1 \times 10^{-4}$. The collaborative fine-tuning also adopted the Adam optimizer with $\beta_1 = 0.9$ and $\beta_2 = 0.99$, but an initial learning rate of $5 \times 10^{-6}$. The epoch for fine-tuning was 20. All the networks in the present work were implemented based on the Pytorch framework.

### 3.3 Experimental results and analysis

To evaluate the performance of the proposed method, it was compared with several other methods.

**Bicubic:** The bicubic method is an interpolation-based SR method, which is often adopted as a baseline. The desired high-resolution HSI was obtained by applying bicubic interpolation on a low-resolution input image.

**SRCNN:** The SRCNN method is a classic deep-learning-based SR method for colour images [5]. The SRCNN was trained on the CAVE dataset and the desired high-resolution HSI was predicted from low-resolution input.

**TL-CNMF:** The TL-CNMF method is based on transfer learning [9]. The CNN was trained on the DIV2K dataset with the same settings used in [9]. The desired high-resolution HSI was predicted by the pre-trained CNN and CNMF.

**SSIN:** The SSIN method is a state-of-the-art deep-learning-based SR method for HSIs that was trained on the CAVE dataset with the same settings used in [7].

#### 3.3.1 Ablation experiment

The two stages comprising the proposed method are SR and spectral recovery. For ablation experiments, the SAN network in the proposed method was replaced by the CNN in the TL-CNMF method, and the others remained the same, and designated the CNN-AWAN. Compared with the proposed method, the performance of the CNN-AWAN decreased, but the decrease in the quality of image reconstruction was greater in terms of RMSE than in SAM. Specifically, RMSE increased by 7.5% and SAM increased by 54.4%. The CNN-AWAN exhibited better performance than the TL-CNMF, especially in SAM. Specifically, RMSE decreased by 3.9% and SAM decreased by 25.7%.

Tables 1 and 2 show the evaluation results of the proposed method and others on the ‘Balloons’ image from the CAVE dataset and on the ‘imgb8’ image from the Harvard dataset, respectively, under different scale factors. Tables 3 and 4 show the average quantitative results of different SR methods on test images under scale factors of 3X and 4X, respectively. The best results are highlighted in bold. Figures 6 and 7 show the pseudo-colour images of a patch from reconstruction of the ‘Balloons’ image and the error map between the reconstruction and ground-truth images at the 21st band. Figures 8 and 9 show the reconstruction results from the ‘imgb8’ image and the error map at the fifth band.

According to the results shown in the tables, it can be observed that the reconstruction results from the bicubic interpolation and SRCNN are significantly lower than those obtained using other methods. The SRCNN was originally designed for colour image SR and is only equipped with 64
FIGURE 9  Error maps of reconstructed image and ground-truth patches using different methods at fifth band: (a) bicubic, (b) SRCNN, (c) TL-CNMF, (d) SSIN, and (e) proposed.

FIGURE 10  Spectrum curves of randomly selected points from ‘imgb8’: (a) (500,850), (b) (750,600), (c) (550,720), and (d) (600,600).

Convolutional features and cannot extract enough spectral–spatial features from HSIs, which results in severe spectral distortion. The results prove that insufficient feature numbers may limit the SR performance. The TL-CNMF predicts the high-resolution HSI in a band-by-band manner, that is, 64 features are extracted from one band of HSI and the spectral correlation between low-resolution and high-resolution HSI is explored through the CNMF, which improves the SR performance by ensuring the same endmembers between reconstructed HSIs and low-resolution HSIs, especially in terms of SAM. The SSIN
achieves excellent performance in all metrics by extracting abundant features, but its memory consumption and computational time are the highest among all the methods. The base-feature size of group convolution was 16 in the experiment, meaning that 16 features were extracted from one band of each HSI. As mentioned earlier, these results prove that powerful feature extraction provides support for state-of-the-art performance. The proposed method has comparably the best SR reconstruction effects in all four metrics. By taking a spectrally downsampled image as a bridge, the spatial SR module takes full advantage of the knowledge from the colour image and reduces the need for the HSI training dataset. The results show that the spectral recovery module preserves useful spectral information of HSIs. Compared with the SSIN, the proposed method has lower memory requirements in the testing stage.

In terms of visualization results, the details in the high-resolution images reconstructed by the bicubic and SRCNN methods are blurred, e.g. the text in the 'balloons' image, as shown in Figure 6. There is no obvious difference in the difference maps between reconstructed and ground-truth images in Figure 8, which verifies the quantitative results shown in Table 2. Figure 10 shows the spectra of several points from ‘imgb8’, which indicate that the proposed method exhibits better performance in terms of spectral information preservation.

4 | CONCLUSIONS

Here, a novel general method based on transfer learning is proposed for HSI super-resolution (SR). Different from directly transferring the knowledge of colour images to HSI SR, the proposed method includes two stages: (1) spatial SR based on transfer learning, and (2) spectral recovery. Using a spectrally down-sampled image as a bridge, the proposed method predicts high-resolution images using the spatial SR network and then performs spectral recovery to generate the desired high-resolution HSI. On the one hand, spectral down-sampling reduces the gap between source and target domains, allowing learned knowledge to flow to the target domain better; on the other hand, it can reduce memory consumption and the computational complexity of the spatial SR network. In addition, the method proposed in this paper is flexible and able to integrate any existing SR model. Experimental results indicate that the proposed method achieves promising SR performance with a small amount of training data.

ACKNOWLEDGEMENTS

This work was jointly supported by the National Natural Science Foundation of China (Grant No. 61403397) and Natural Science Basic Research Plan in Shaanxi Province of China (Grant No 2020JM-358). The authors thank LetPub (www.letpub.com) for its linguistic assistance during the preparation of this manuscript.

REFERENCES

1. Vai, A., et al.: Deep learning for land use and land cover classification based on hyperspectral and multispectral earth observation data: A review. Remote Sens. 12(15), 2495 (2020)
2. Khan, M.J., et al.: Modern trends in hyperspectral image analysis: A review. IEEE Access 6, 14118–14129 (2018)
3. Zhao, C., Zhang, L.: Spectral-spatial stacked autoencoders based on low-rank and sparse matrix decomposition for hyperspectral anomaly detection. Infrared Phys. Technol. 92, 166–176 (2018)
4. Wang, N., et al.: Dynamic selection network for image inpainting. IEEE Trans. Image Process. 30, 1784–1798 (2021)
5. Dong, C., et al.: Learning a deep convolutional network for image super-resolution. Paper presented at the European Conference on Computer Vision (ECCV), pp. 184–199 (2014)
6. Liebel, L., Körner, M.: Single-image super resolution for multispectral remote sensing data using convolutional neural networks. ISPRS. XLIII-B3, 883–890 (2016)
7. Zheng, K., et al.: Separable-spectral convolution and inception network for hyperspectral image super-resolution. Int. J. Mach. Learn. Cybern. 10, 2593-2607 (2019). https://doi.org/10.1007/s13042-018-0991-4
8. Shi, Z., et al.: Deep residual attention network for spectral image super-resolution. In: European Conference on Computer Vision, pp. 214–229. Springer (2018)
9. Yuan, Y., et al.: Hyperspectral image super-resolution by transfer learning. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 10(5), 1963–1974 (2017). https://doi.org/10.1109/JSTARS.2017.2655112
10. He, Z., Liu, L.: Hyperspectral image super-resolution inspired by deep Laplacian pyramid network. Remote Sens. 10(12), 1939 (2018)
11. Li, X., et al.: Domain transfer learning for hyperspectral image super-resolution. Remote Sens. 11(6), 694 (2019).
12. Wang, Z., et al.: Deep learning for image super-resolution: A survey. IEEE Trans. Pattern Anal. Mach. Intell. 1-1 (2020). https://doi.org/10.1109/TPAMI.2020.2982166
13. Zhao, H., et al.: Loss functions for image restoration with neural networks. IEEE Trans. Comput. Imaging 3(1), 47–57 (2016). https://doi.org/10.1109/TCI.2016.2648465
14. Dai, T., et al.: Second-order attention network for single image super-resolution. In: 2019 Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 11065–11074 (2019)
15. Galliani, S., et al.: Learned spectral super-resolution. arXiv:1703.09470 (2017).
16. Li, J., et al.: Adaptive weighted attention network with camera spectral sensitivity prior for spectral reconstruction from RGB images. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, IEEE, pp. 462–463 (2020)
17. Agustsson, E., Tomico, R.: Nitre 2017 challenge on single image super-resolution: Dataset and study. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops, IEEE, pp. 126–135 (2017)
18. Yasuma, F., et al.: Generalized assorted pixel camera: Postcapture control of resolution, dynamic range, and spectrum. IEEE Trans. Image Process. 19(9), 2241–2253 (2010). https://doi.org/10.1109/TIP.2010.2046811
19. Chakrabarti, A., Zickler, T.: Statistics of real-world hyperspectral images. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, IEEE, pp. 193–200 (2011)

How to cite this article: Zhang, S., et al.: Spectral recovery-guided hyperspectral super-resolution using transfer learning. IET Image Process. 1–10 (2021). https://doi.org/10.1049/ipt2.12253