Deep Residual Attention Network
for Spectral Image Super-Resolution

Zhan Shi, Chang Chen, Zhiwei Xiong$^{(\boxtimes)}$, Dong Liu, Zheng-Jun Zha,
and Feng Wu

University of Science and Technology of China, Hefei, China
{zhanshi,changc}@mail.ustc.edu.cn,
{zwxiong,dongeliu,zhazj,fengwu}@ustc.edu.cn

Abstract. Spectral imaging sensors often suffer from low spatial resolution, as there exists an essential tradeoff between the spectral and spatial resolutions that can be simultaneously achieved, especially when the temporal resolution needs to be retained. In this paper, we propose a novel deep residual attention network for the spatial super-resolution (SR) of spectral images. The proposed method extends the classic residual network by (1) directly using the 3D low-resolution (LR) spectral image as input instead of upsampling the 2D bandwise images separately, and (2) integrating the channel attention mechanism into the residual network. These two operations fully exploit the correlations across both the spectral and spatial dimensions and greatly promote the performance of spectral image SR. In addition, for the scenario when stereo pairs of LR spectral and high-resolution (HR) RGB measurements are available, we design a fusion framework based on the proposed network. The spatial resolution of the spectral input is enhanced in one branch, while the spectral resolution of the RGB input is enhanced in the other. These two branches are then fused together through the attention mechanism again to reconstruct the final HR spectral image, which achieves further improvement compared to using the single LR spectral input. Experimental results demonstrate the superiority of the proposed method over plain residual networks, and our method is one of the winning solutions in the PIRM 2018 Spectral Super-resolution Challenge.

Keywords: Spectral image · Super-resolution · Channel attention

1 Introduction

Spectral imaging sensors aim to obtain the spectrum of the scene in terms of tens or hundreds of bandwise images, each corresponding to a certain narrow wavelength range. The spectral image obtained in this way plays an important role in many fields such as remote sensing [16], medical diagnosis [17], and agriculture [18], also including various computer vision tasks such as image segmentation [43], face recognition [34], and object tracking [44]. Therefore, the development

of spectral imaging systems has been an active research field in recent years [7,15,29,47,50,51].

However, to capture 3D spectral images with existing 2D sensors, trade-offs between spectral and spatial/temporal resolutions are inevitable [3,14]. Conventional imaging spectrometers usually operate in a scanning manner, which simply trades the temporal resolution for the spectral resolution [5], either scanning the space [3,35] or scanning the spectrum [14,37]. On the other hand, to enable spectral acquisition for dynamic scenes, snapshot spectral imagers relying on computational reconstruction have been developed in the last decade. The representative techniques along this line include computed tomographic imaging spectrometry (CTIS) [10], prism-mask spectral video imaging system (PMVIS) [7], and coded aperture snapshot spectral imager (CASSI) [47]. These snapshot spectral imagers support spectral video acquisition, yet at the cost of sacrificing the spatial resolution or reconstruction fidelity. To address this issue, dual-camera systems incorporating a snapshot spectral imager and an RGB/panchromatic camera with a beam splitter are proposed [30,48,49], which combine the high spectral resolution of the former and the high spatial resolution of the latter. Still, these systems are of high cost in terms of hardware implementation.

As an alternative solution, spatial super-resolution (SR) of spectral images has also attracted tremendous research efforts, where only a low-resolution (LR) spectral image is required as input and the high-resolution (HR) spectral image is obtained without the need of additional hardware. Similar to single-image SR, mainstream solutions for this ill-posed problem are example-based ones. As a representative, the sparse coding based approaches learn the relationship between LR spectral cubes and the corresponding HR counterparts as a dictionary from an external database [1,2,13], where the HR spectral images can be obtained through scanning-based spectrometers. With the success of deep learning in single-image SR [11,24], CNN-based methods rapidly emerge to directly learn an end-to-end mapping between LR and HR spectral images [27,28,31,54], which achieve promising performance while promoting the efficiency compared to sparse coding. Following this trend, we propose an advanced network architecture for spectral image SR, which integrates the powerful channel attention mechanism [20] with the classic residual blocks [19]. Consequently, the correlations across both the spectral and spatial dimensions are fully exploited and the performance of spectral image SR is significantly improved over plain residual networks.

In practical applications where spectral images are captured, it is often possible to capture an RGB/panchromatic image of the same scene at the same time with a much higher spatial resolution. Different from the above mentioned systems that using a beam splitter [30,48], the resulting spectral and RGB/panchromatic measurements are not aligned but rather in a stereo configuration, which thus relieves the difficulty of hardware implementation. This additional HR RGB/panchromatic image can be of great help for enhancing the LR spectral image, which is usually referred to as color-guided spectral image SR

[23,25,53] or pan-sharpening [26,36,38] and has been investigated extensively in the literature. Based on the proposed network, we then design a fusion framework for color-guided spectral image SR. The spatial resolution of the spectral input is enhanced in one branch, and the spectral resolution of the RGB input is enhanced in the other. These two branches are then fused together through the attention mechanism again to reconstruct the final HR spectral image, which achieves further improvement compared to using the single LR spectral input.

The main contributions of this paper can be summarized as follows:

– An advanced network architecture for spectral image SR, which integrates the powerful channel attention mechanism with the classic residual blocks.
– A fusion framework for color-guided spectral image SR based on the proposed network, which is designed for the scenario when stereo pairs of LR spectral and HR RGB measurements are available.
– State-of-the-art results on the above two tasks, and one of the winning solutions in the PIRM 2018 Spectral Image SR Challenge [40,41].

2   Related Work

Spectral Image SR. Different from reconstructing a spectral image from the corresponding RGB image [39,52] (sometimes also referred to as spectral SR), the spectral image SR here denotes to spatial resolution enhancement of a spectral image. Similar to single-image SR that enhances the spatial resolution of a 2D image, mainstream solutions for this ill-posed problem are example-based ones. As a representative, the sparse coding based approaches learn the relationship between LR spectral cubes and the corresponding HR counterparts as a dictionary from an external database [1,2,13], where the HR spectral images can be obtained through scanning-based spectrometers. With the success of deep learning in single-image SR [11,24], CNN-based methods rapidly emerge to directly learn an end-to-end mapping between LR and HR spectral images [27,28,31,54], which achieve promising results while promoting the efficiency compared to sparse coding. With even advanced network architectures such as the one proposed in this paper, the performance of spectral image SR is expected to be further improved.

Color-Guided Spectral Image SR. The approaches which fuse an LR spectral image with an HR RGB image based on matrix factorization have been actively investigated [23,25,53]. These approaches first unfold the spectral image as a matrix, and then decompose the matrix as spectral basis and corresponding coefficients. Spectral image SR then becomes the estimation of spectral basis and corresponding coefficients from the LR spectral and HR RGB measurements of the same scene. In addition to considering the spectral information, some approaches also use the spatial structures of spectral images for SR [12,42,46]. These matrix factorization based methods often start by unfolding the 3D data structures into matrices. However, operating with matrices makes it hard to fully exploit the inherent spatial-spectral correlations in spectral images. These

correlations can be better exploited using a CNN-based fusion framework, as demonstrated in this paper.

Attention Mechanism. Attention can be viewed as an adaptation mechanism which automatically allocates processing resources towards the most informative components of an input [45]. Following the successful application in neural machine translation [45], increasingly more methods have been proposed to apply attention into the deep network [4,6,22,32,55]. Inspired by [55], we integrate the channel attention mechanism [20] with the residual block [19], which promotes the reconstruction fidelity of spectral image SR.

3   Spectral Image SR Network

3.1   Baseline VDSR-3D Network

A 3D spectral image can be viewed as a stack of 2D bandwise grayscale images. Hence, classic single-image SR networks can be directly applied to the spectral image SR task. Take the well-known VDSR [24] network for example, the network can use each LR bandwise image as input and output the corresponding HR image, and the HR spectral image can be obtained by stacking these reconstructed HR bandwise images. However, this simple reuse of VDSR is not efficient since it ignores the correlations across the spectral dimension. A natural extension would be taking the 3D LR spectral image as input of the network, in which the spectral correlations are learned automatically. This modified VDSR network, denoted as VDSR-3D, is regarded as the baseline model in this paper.

Although VDSR-3D can achieve a better performance compared with its 2D version, as demonstrated later, there are still some problems in the VDSR architecture. First, it requires a bicubic-interpolated image as the input, which is a sub-optimal solution since a deconvolutional layer can easily learn a better operation than bicubic in a back-propagation manner. Moreover, according to the experimental results, it fails to achieve an improved performance when deepening the network structure, which is not favorable for pursuing more accurate solutions. To address these problems, we propose the deep residual attention network that greatly improve the capability of VDSR-3D.
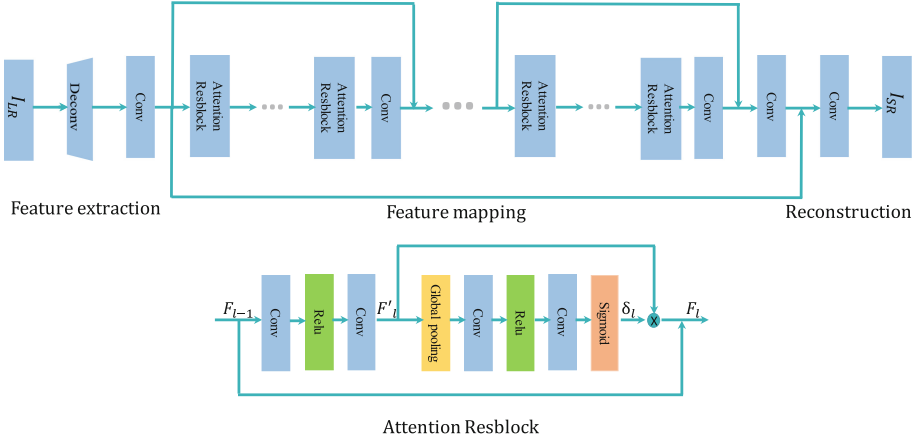
3.2   Deep Residual Attention Network

The architecture of our proposed deep residual attention network is shown in Fig. 1. The network mainly consists of three modules: feature extraction, feature mapping, and reconstruction. We use one deconvolutional layer and one convolutional layer to extract the feature as

$$F_0 = f_E(I_{LR}), \tag{1}$$

where f_E denotes the feature extraction module. Note that, we set the stride of deconvolution the same as the upsampling scale (i.e., $\times 3$ in our experiments) to

Fig. 1. Deep residual attention network for spectral image SR

convert an LR spectral input I_{LR} to the HR space for succeeding procedures. The obtained feature F_0 is then fed into the feature mapping module, where a number of residual blocks are adopted to implement this procedure.

Inspired by [55], we integrate the channel attention mechanism [20] with the residual block [19] to generate the attention resblock, which is also shown in Fig. 1. Let $F_l^{'}(l = 1, \ldots, L)$ denotes the intermediate feature in the lth attention resblock. The channel attention coefficient δ_l can be calculated as

$$\delta_l = Sig(C(P_G(F_l^{'}))),$$

$$P_G(F_{l,c}^{'}) = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} F_{l,c}^{'}(i,j), \tag{2}$$

where H and W denote the spatial resolution of the feature map, $F_{l,c}^{'}(i,j)$ denotes the value at position (i,j) in the cth channel of the intermediate feature in the lth attention resblock, $P_G(\cdot)$ denotes the global average pooling function, $C(\cdot)$ denotes the combining operation which includes two successive convolutional layers (with a 1×1 kernel size) and a ReLu activation function [33] in between, and $Sig(\cdot)$ denotes the sigmoid function for normalization.

Obtained the channel attention coefficient δ_l, the output of the lth attention resblock F_l can then be calculated as

$$F_l = F_{l-1} + \delta_l \cdot F_l^{'}. \tag{3}$$

With the proposed channel attention, the residual component in the attention resblock is adaptively weighted. Compare to the plain convolution in VDSR-3D, the attention resblock offers higher reconstruction fidelity with a much deepened network. Moreover, we adopt the skip connection in each stack of the attention

resblock to relieve the vanishing of gradient and ease the convergence of the network. The super-resolved spectral image I_{SR} is finally reconstructed as

$$I_{SR} = f_R(F_L), \tag{4}$$

where $f_R(\cdot)$ denotes the reconstruction module implemented using a single convolutional layer.
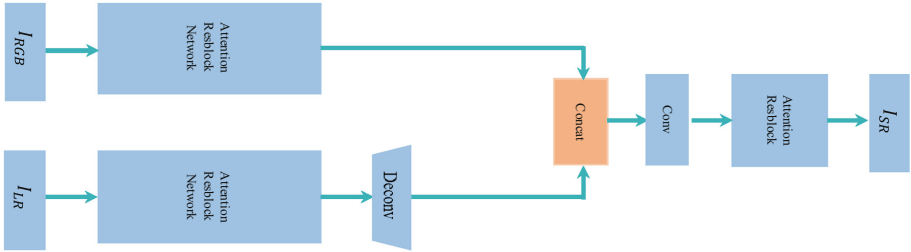
Fig. 2. Color-guided fusion framework for spectral image SR

3.3   Color-Guided Fusion Framework

In practical applications where spectral images are captured, it is often possible to capture an RGB image of the same scene at the same time with a much higher spatial resolution. However, the resulting spectral and RGB images are not aligned but rather in a stereo configuration. A registration step is needed to address the scene disparity and generate the aligned RGB image as an input for SR in addition to the spectral image. This aligned HR RGB image can provide more information in the spatial dimension to facilitate the spectral image SR.

To this end, we design a color-guided fusion framework with a two-branch architecture to implement the fusion between the spectral information in the spectral image and the spatial information in the RGB image. The fusion framework is shown in Fig. 2. The two branches share the same basic network structure, i.e., the deep residual attention network described in Sect. 3.2. Note that the upscaling operation is no longer needed in the RGB branch, while in the spectral branch it is moved to the tail of the network (due to the restriction of GPU memory in implementation). A concatenating operation is used to combine the features from these two branches. The reconstructed HR spectral image I_{SR} can be obtained as

$$I_{SR} = f_{FUS}(f_{SPE}(I_{LR}) \oplus f_{RGB}(I_{RGB})), \tag{5}$$

where $f_{RGB}(\cdot)$, $f_{SPE}(\cdot)$, and $f_{FUS}(\cdot)$ denote the RGB branch, the spectral branch, and the fusion part, respectively. The symbol $\oplus$ denotes the concatenating operation between features output from these two branches. Note that another attention resblock is used in the fusion part.

3.4   Implementation Details

Training Process. We train the network of Track 1 (spectral image SR) using LR spectral cubes with a size of $20 \times 20 \times 14$ and corresponding HR cubes with a size of $60 \times 60 \times 14$. We set the batch size as 64 and adopt the ADAM optimizer with $\beta_1 = 0.9$, $\beta_2 = 0.999$, and $\epsilon = 10^{-8}$. The initial leaning rate is set as 1×10^{-4} and then it is decreased to half every 1×10^5 iterations of back-propagation. We stop training when no notable decay of training loss is observed. Based on the default setting of Track 1, we add the aligned HR RGB patches as inputs to train the network of Track 2 (color-guided spectral image SR), which have the same spatial size with the HR spectral cubes. We implement the proposed networks based on the PyTorch framework and train them using a 1080Ti GPU. It takes about 12 h to train each network.

Testing Process. For Track 1, there is no difference between the training and testing processes. Yet for Track 2, we conduct a post-processing operation to further promote the performance. Since the aligned RGB images are obtained using the FlowNet 2.0 [21] according to the challenge configuration, black pixels are usually observed around the borders due to the imperfect registration. To address this issue, we first crop 12 pixels in the four-direction borders and then make them up using the network of Track 1. In other words, the center part of the reconstructed image is obtained using the color-guided fusion framework, while the boundary pixels are only reconstructed from the LR spectral image.

Loss Function. CNN-based methods for single-image SR usually adopt the mean square error (MSE) as the loss function during training [9, 11, 24], which has also been applied to spectral reconstruction tasks [52]. However, the luminance level in spectral images usually varies significantly among different bands, and the same deviation in the pixel value may have different influence to the bands with different luminance levels. It thus makes the MSE loss generate a bias towards the bands with high luminance levels, which is not desired because each band matters equally. Hence, recent methods adopt the mean relative absolute error (MRAE) as a substitute [39].

Let $I_{SR}(i)$ and $I_{GT}(i)$ denote the $i^{th}(i = 1, \dots, N)$ pixel of the reconstructed and groundtruth spectral images, respectively. The MRAE is formulated as

$$MRAE = \frac{1}{N} \sum_{i=1}^{N} \left(\mid I_{SR}^{(i)} - I_{GT}^{(i)} \mid / I_{GT}^{(i)} \right). \tag{6}$$

When there exists zero points in the groundtruth image, the MRAE loss will become infinite, making the network fail to converge. To address this issue, a small value of ϵ_1 is added to the denominator of MRAE. Considering the intensity range of the spectral data is from 0 to 65535, ϵ_1 is set to 1 according to the challenge configuration, deriving the modified MRAE as

$$MRAE' = \frac{1}{N} \sum_{i=1}^{N} \left(\mid O((I_{SR}^{(i)})) - I_{GT}^{(i)} \mid / (I_{GT}^{(i)} + \epsilon_1) \right), \tag{7}$$

where the operation $O(\cdot)$ rounds the small value less than 1 into 0, which further decreases the reconstruction error caused by the zero points.

To investigate the impact of different loss functions, we also implement the spectral information divergence (SID) function which is designed to evaluate the spectral similarity and discriminability [8]. The SID is formulated as

$$SID = \sum_{i=1}^{N} \left(D\left(I_{SR}^{(i)} \parallel I_{GT}^{(i)} \right) + D\left(I_{GT}^{(i)} \parallel I_{SR}^{(i)} \right) \right),$$

$$D\left(I_{SR}^{(i)} \parallel I_{GT}^{(i)} \right) = O((I_{SR}^{(i)})) \, log\left(\left(O((I_{SR}^{(i)})) + \epsilon_2 \right) / \left(I_{GT}^{(i)} + \epsilon_2 \right) \right), \quad (8)$$

$$D\left(I_{GT}^{(i)} \parallel I_{SR}^{(i)} \right) = I_{GT}^{(i)} \, log\left(\left(I_{GT}^{(i)} + \epsilon_2 \right) / \left(O((I_{SR}^{(i)})) + \epsilon_2 \right) \right),$$

where ϵ_2 is set to 1×10^{-3} according to the challenge configuration to avoid the infinite value caused by the zero points.

4   Experimental Results

4.1   Dataset and Evaluation

The experiments are conducted strictly following the instructions of the PIRM 2018 Spectral Image SR Challenge [40,41]. There are two tracks in this challenge. Track 1 requires to upscale an LR spectral image (with a size of $80 \times 160 \times 14$) by a factor of 3 in the spatial dimension. The dataset for this track consists of 240 spectral images captured by an IMEC 16-band snapshot spectral imager, which is split into 200 for training, 20 for validation, and 20 for testing. Track 2 requires to accomplish the same task with the help of corresponding HR RGB images (with a size of 240×480). The dataset contains 120 pairs of stereo images, with one view captured by the IMEC 16-band snapshot spectral imager and the other by an ordinary color camera. The images are split into 100 pairs for training, 10 for validation, and 10 for testing. Also, the dataset provides a registered version of the RGB images using the FlowNet 2.0 [21].

In the development phase, data augmentation is performed on the training images which are randomly rotated by 90°, 180°, 270°, and flipped horizontally. We reserve 10 pairs of images as our own validation set. Our reported results are all calculated on the official validation set. No additional preprocessing or dataset is needed for both tracks. The main ranking metrics of this challenge are MRAE, SID, and the mean opinion score (MOS). Note that the challenge utilizes the normalized SID as the final metric, which may be different from the one used in this paper.

4.2   Ablation Experiments

Loss Function. To investigate the impact of different loss functions, we conduct a series of experiments. The experimental results are shown in Table 1. We first

compare the results between using MRAE and SID as loss function alone. There seems to exist a trade-off between these two metrics and one increases while the other decreases. To prove this, we further design a combined loss function where the weight of SID is ten times that of MRAE, which gives a simple joint optimization according to cross-validation. The result shows that the SID can improve notably (i.e., 13%) by slightly sacrificing the performance in MRAE (i.e., 2%). We also use the SID as loss function to fine-tune the model pretrained by MRAE. During the training process, the SID turns to decrease but the MRAE increases at the same time, which again demonstrates the trade-off between SID and MRAE. Finally, it converges to a similar point to that achieved by using the SID as loss function alone. In the experiments below and in the challenge, we use the MRAE as loss function alone since it is the primary evaluation metric.

Table 1. Comparisons between different loss function settings. The MRAE and SID are adopted as the metrics

Loss function	MRAE	SID	Combined	Fine-tune
MRAE	0.1215	0.1761	0.1234	0.1707
SID	135	104	117	104
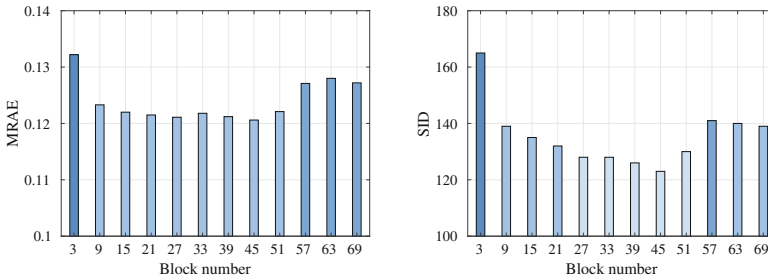
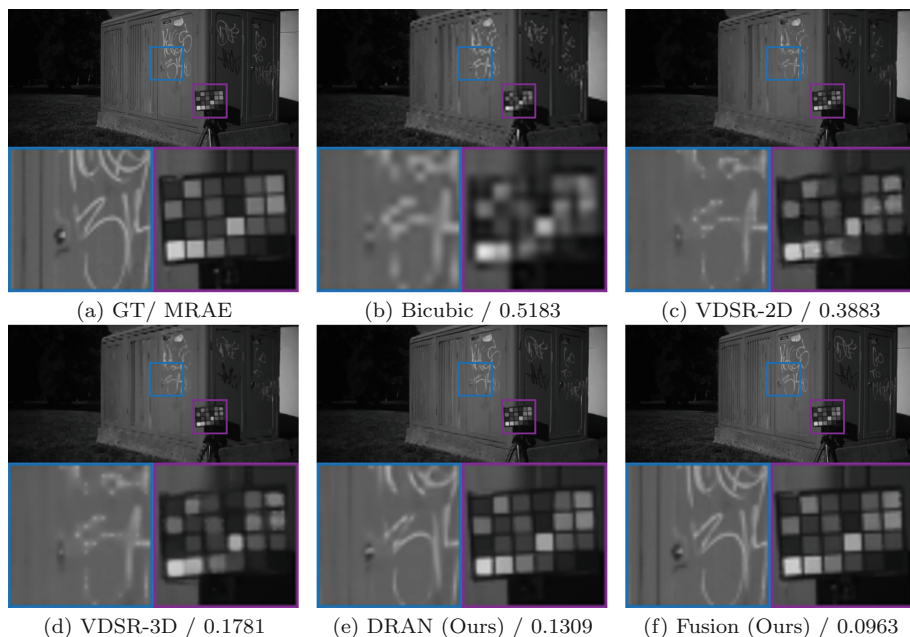**Fig. 3.** The influence of depth on the network performance

Table 2. Comparisons between our proposed methods and baseline models. The validation sets of the challenge are adopted for evaluation. Bold indicates the best performance and italic indicates the second best one

Dataset	Bicubic	VDSR-2D	VDSR-3D	DRAN	Fusion
Track1 (MRAE)	0.2406	0.1790	*0.1384*	**0.1215**	-
Track2 (MRAE)	0.4202	0.3044	0.1464	*0.1334*	**0.1173**
Running time (s)	-	0.08	0.09	1.11	2.50

Network Depth. The depth is an important factor to determine the basic capacity of the network. However, a deeper network does not always yield a better performance. As depicted in Fig. 3, a number of 45 attention resblocks seems to be an optimal depth under both SID and MRAE metrics. Note that when the block number goes up to 57, gradient exploding will occur in the training process. To overcome this, the initial learning rate is lowered down to 1/5 of the original, which could result in the performance decrease in the deeper structure. Also, we find that the initial value of the network can slightly affect the performance but the conclusion holds. The other settings of the network are kept the same to eliminate the influence of other hyper-parameters. In the experiments below and in the challenge, we use 15 attention resblocks in consideration of implementation speed.

4.3   Comparison with Baseline Models

To validate the effectiveness of the proposed deep residual attention network (DRAN) and the color-guided fusion framework (Fusion), we first compare their performance with the baseline models VDSR-2D and VDSR-3D (both with a network depth comparable to DRAN). All these models are trained using the whole training set of each track, and the official validation set is adopted for evaluation. The quantitative results are listed in Table 2. As can be seen, for both

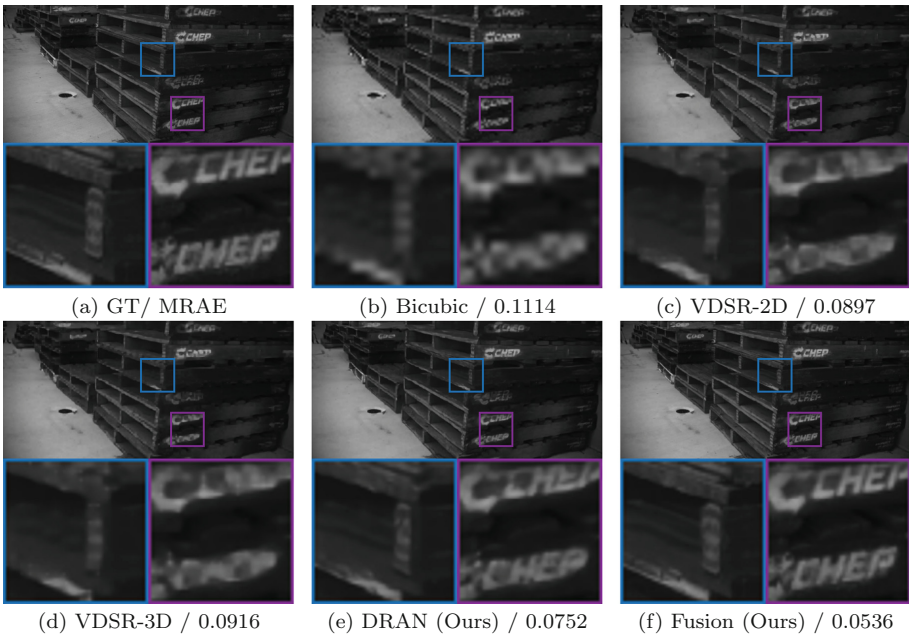| (a) GT/ MRAE	(b) Bicubic / 0.5183	(c) VDSR-2D / 0.3883
(d) VDSR-3D / 0.1781	(e) DRAN (Ours) / 0.1309	(f) Fusion (Ours) / 0.0963

Fig. 4. Visual comparison of a spectral image from the validation set of Track 2. All 14 bands are averaged for the evaluation of spatial information

tracks, VDSR-3D outperforms VDSR-2D by a large margin, which indicates the importance of exploiting correlations across different bands in spectral image SR. Meanwhile, the results of DRAN show that the architecture of attention resblocks has a distinct advantage over the plain residual networks. For Track 2, the fusion method achieves 12% decrease in MRAE compared with DRAN, which proves the usefulness of the HR RGB images as well as the fusion framework.

4.4   Ensemble Method and Running Time

For ensemble purpose, we flip and rotate the input image and treat it as another input similar to data augmentation. Then we apply an inverse transform to the corresponding output. Finally, we average the transformed output and the original output to generate the self-ensemble result. In this way, further improvement (3% decrease in MRAE) can be achieved.

We calculate the average running time using a 1080Ti GPU. The running time includes the process of ensembling. The fusion method cost nearly double time compared to a single DRAN since it needs the result of DRAN to recover the cropped borders. And the VDSR-based models are slightly faster than DRAN.

(a) GT/ MRAE   (b) Bicubic / 0.1114   (c) VDSR-2D / 0.0897

(d) VDSR-3D / 0.0916   (e) DRAN (Ours) / 0.0752   (f) Fusion (Ours) / 0.0536

Fig. 5. Visual comparison of a spectral image from the validation set of Track 2. All 14 bands are averaged for the evaluation of spatial information

4.5 Visual Quality Comparison

To evaluate the perceptual quality of spectral image SR, we show the visual comparison of two images in Track 2 (for which all methods can be compared together). Note that we average the spectral image across the spectral dimension for a better visual experience in the spatial dimension. As can be seen in Figs. 4 and 5, the edge regions of super-resolved images from DRAN and Fusion are notably shaper and clearer than the VDSR-based models and bicubic interpolation. Also, with the help of HR RGB images, the blurring artifacts are alleviated and more details of spectral images are recovered, if comparing the results of Fusion and DRAN.

To further visualize the spectral accuracy of the reconstructed HR spectral images, we show the error maps of three selected bands in the above two images in Figs. 6 and 7. As can be seen, on the one hand, the error of DRAN and Fusion are smaller than the VDSR-based models and bicubic interpolation, which again validates the effectiveness of the proposed method. On the other hand, the error
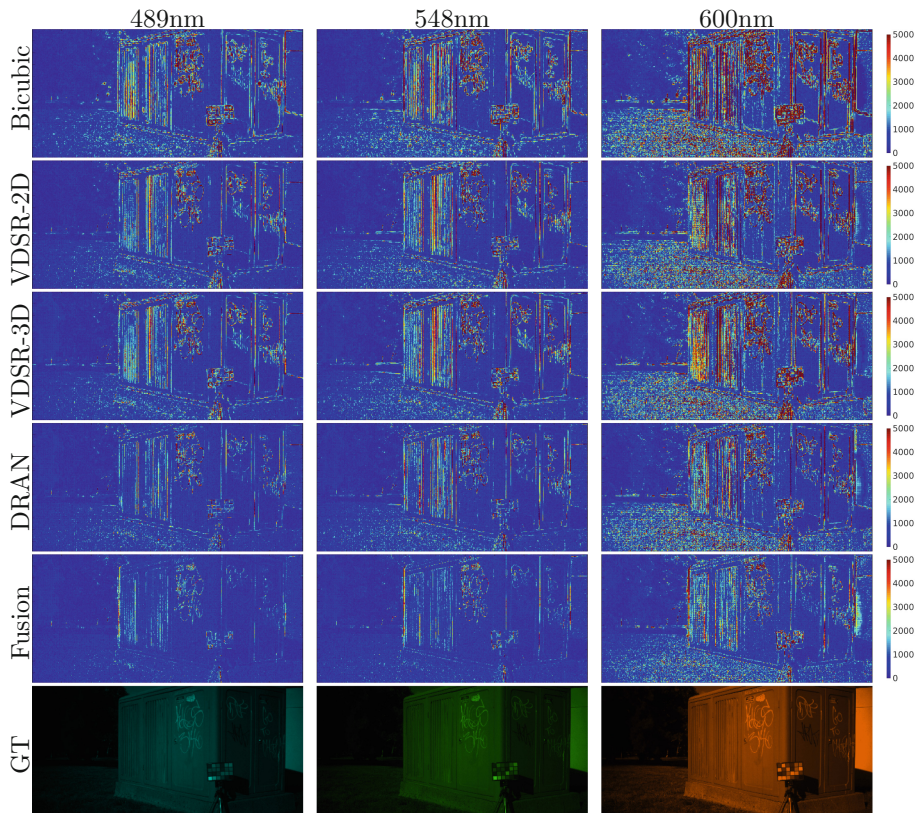
Fig. 6. Visual comparison of a spectral image from the validation set of Track 2. Three bands are selected for the evaluation of spectral accuracy
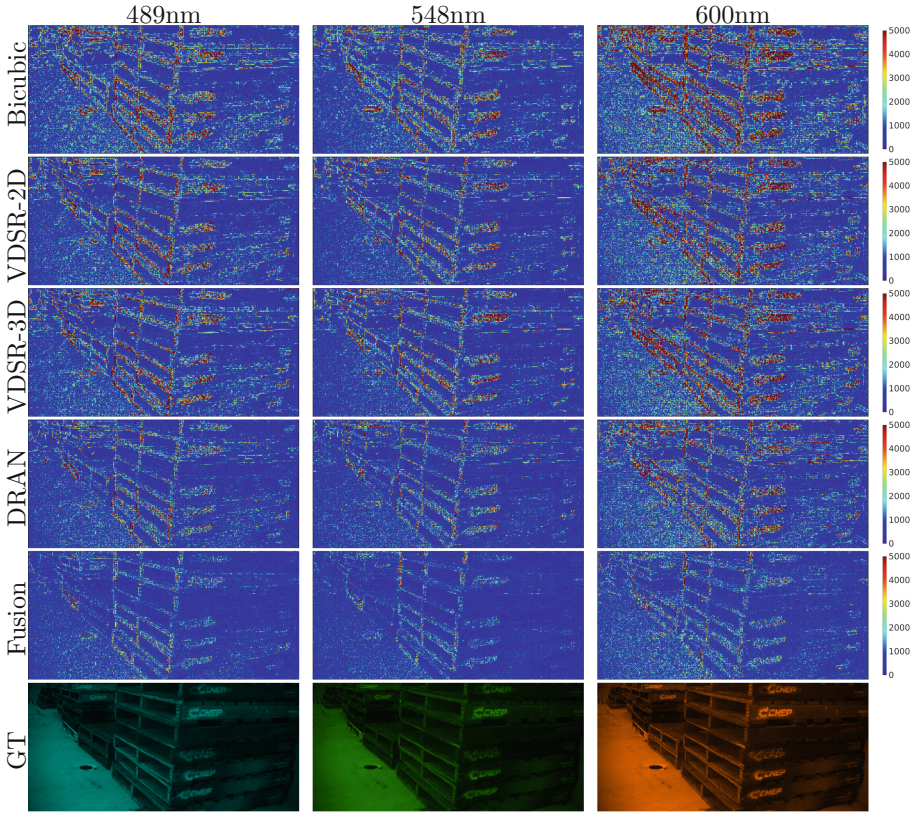
Fig. 7. Visual comparison of a spectral image from the validation set of Track 2. Three bands are selected for the evaluation of spectral accuracy

in the edge regions of Fusion is obviously smaller than that of DRAN, which demonstrate that the HR RGB image mainly contributes to the edge regions.

5   Conclusions

This paper presents a novel deep residual attention network for the spatial SR of spectral images. The proposed method integrates the channel attention mechanism into the residual network to fully exploit the correlations across both the spectral and spatial dimensions of spectral images, which greatly promotes the performance of spectral image SR. In addition, we design a fusion framework based on the proposed network when stereo pairs of LR spectral and HR RGB measurements are available, which achieves further improvement compared to using the single LR spectral input. Experimental results demonstrate the superiority of the proposed method over plain residual networks, and our method is one of the winning solutions in the PIRM 2018 Spectral Super-resolution Challenge.

References

1. Akhtar, N., Shafait, F., Mian, A.: Sparse spatio-spectral representation for hyper-spectral image super-resolution. In: Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T. (eds.) ECCV 2014. LNCS, vol. 8695, pp. 63–78. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-10584-0_5

2. Akhtar, N., Shafait, F., Mian, A.: Hierarchical beta process with gaussian process prior for hyperspectral image super resolution. In: Leibe, B., Matas, J., Sebe, N., Welling, M. (eds.) ECCV 2016. LNCS, vol. 9907, pp. 103–120. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-46487-9_7

3. Basedow, R.W., Carmer, D.C., Anderson, M.E.: HYDICE system: implementation and performance. In: Proceedings of SPIE (1995)

4. Bluche, T.: Joint line segmentation and transcription for end-to-end handwritten paragraph recognition. In: NIPS (2016)

5. Brady, D.J.: Optical Imaging and Spectroscopy. Wiley, Hoboken (2009)

6. Cao, C., et al.: Look and think twice: capturing top-down visual attention with feedback convolutional neural networks. In: ICCV (2015)

7. Cao, X., Du, H., Tong, X., Dai, Q., Lin, S.: A prism-mask system for multispectral video acquisition. IEEE Trans. Pattern Anal. Mach. Intell. **33**(12), 2423–35 (2011)

8. Chang, C.I.: Spectral information divergence for hyperspectral image analysis. In: IGARSS (1999)

9. Chen, C., Tian, X., Xiong, Z., Wu, F.: UDNET: up-down network for compact and efficient feature representation in image super-resolution. In: ICCVW (2017)

10. Descour, M., Dereniak, E.: Computed-tomography imaging spectrometer: experimental calibration and reconstruction results. Appl. Opt. **34**(22), 4817–4826 (1995)

11. Dong, C., Loy, C.C., He, K., Tang, X.: Image super-resolution using deep convolutional networks. IEEE Trans. Pattern Anal. Mach. Intell. **38**(2), 295–307 (2016)

12. Dong, W., et al.: Hyperspectral image super-resolution via non-negative structured sparse representation. IEEE Trans. Image Process. **25**(5), 2337–2352 (2016)

13. Fang, L., Zhuo, H., Li, S.: Super-resolution of hyperspectral image via superpixel-based sparse representation. Neurocomputing **273**, 171–177 (2018)

14. Gat, N.: Imaging spectroscopy using tunable filters: a review. In: Proceedings of SPIE (2000)

15. Goel, M., et al.: HyperCam: hyperspectral imaging for ubiquitous computing applications. In: UbiComp (2015)

16. Goetz, A.F., Vane, G., Solomon, J.E., Rock, B.N.: Imaging spectrometry for earth remote sensing. Science **228**(4704), 1147–1153 (1985)

17. Gowen, A., O'Donnell, C., Cullen, P., Downey, G., Frias, J.: Hyperspectral imaging–an emerging process analytical tool for food quality and safety control. Trends Food Sci. Technol. **18**(12), 590–598 (2007)

18. Haboudane, D., Miller, J.R., Pattey, E., Zarco-Tejada, P.J., Strachan, I.B.: Hyper-spectral vegetation indices and novel algorithms for predicting green LAI of crop canopies: modeling and validation in the context of precision agriculture. Remote Sens. Environ. **90**(3), 337–352 (2004)

19. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: CVPR (2016)

20. Hu, J., Shen, L., Sun, G.: Squeeze-and-excitation networks. In: CVPR (2018)
21. Ilg, E., Mayer, N., Saikia, T., Keuper, M., Dosovitskiy, A., Brox, T.: Flownet 2.0: evolution of optical flow estimation with deep networks. In: CVPR (2017)
22. Jaderberg, M., Simonyan, K., Zisserman, A., et al.: Spatial transformer networks. In: Advances in Neural Information Processing Systems (2015)
23. Kawakami, R., Matsushita, Y., Wright, J., Ben-Ezra, M., Tai, Y.W., Ikeuchi, K.: High-resolution hyperspectral imaging via matrix factorization. In: CVPR (2011)
24. Kim, J., Kwon Lee, J., Mu Lee, K.: Accurate image super-resolution using very deep convolutional networks. In: CVPR (2016)
25. Lanaras, C., Baltsavias, E., Schindler, K.: Hyperspectral super-resolution by coupled spectral unmixing. In: ICCV (2015)
26. Li, S., Yang, B.: A new pan-sharpening method using a compressed sensing technique. IEEE Trans. Geosci. Remote Sens. **49**(2), 738–746 (2011)
27. Li, Y., Hu, J., Zhao, X., Xie, W., Li, J.: Hyperspectral image super-resolution using deep convolutional neural network. Neurocomputing **266**, 29–41 (2017)
28. Liebel, L., Körner, M.: Single-image super resolution for multispectral remote sensing data using convolutional neural networks. Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci. **41**, 883–890 (2016)
29. Lin, X., Liu, Y., Wu, J., Dai, Q.: Spatial-spectral encoded compressive hyperspectral imaging. ACM Trans. Graph **33**(6), 233 (2014)
30. Ma, C., Cao, X., Tong, X., Dai, Q., Lin, S.: Acquisition of high spatial and spectral resolution video with a hybrid camera system. Int. J. Comput. Vision **110**(2), 141–155 (2014)
31. Mei, S., Yuan, X., Ji, J., Zhang, Y., Wan, S., Du, Q.: Hyperspectral image spatial super-resolution via 3D full convolutional neural network. Remote Sens. **9**(11), 1139 (2017)
32. Miech, A., Laptev, I., Sivic, J.: Learnable pooling with context gating for video classification. arXiv preprint arXiv:1706.06905 (2017)
33. Nair, V., Hinton, G.E.: Rectified linear units improve restricted Boltzmann machines. In: ICML (2010)
34. Pan, Z., Healey, G., Prasad, M., Tromberg, B.: Face recognition in hyperspectral images. IEEE Trans. Pattern Anal. Mach. Intell. **25**(12), 1552–1560 (2003)
35. Porter, W.M., Enmark, H.T.: A system overview of the airborne visible/infrared imaging spectrometer (AVIRIS). In: Proceedings of SPIE (1987)
36. Rahmani, S., Strait, M., Merkurjev, D., Moeller, M., Wittman, T.: An adaptive IHS pan-sharpening method. IEEE Geosci. Remote Sens. Lett. **7**(4), 746–750 (2010)
37. Schechner, Y., Nayar, S.: Generalized mosaicing: wide field of view multispectral imaging. IEEE Trans. Pattern Anal. Mach. Intell. **24**(10), 1334–1348 (2002)
38. Shah, V.P., Younan, N.H., King, R.L.: An efficient pan-sharpening method via a combined adaptive pca approach and contourlets. IEEE Trans. Geosci. Remote Sens. **46**(5), 1323–1335 (2008)
39. Shi, Z., Chen, C., Xiong, Z., Liu, D., Wu, F.: HSCNN+: Advanced CNN-based hyperspectral recovery from RGB images. In: CVPRW (2018)
40. Shoeiby, M., et al.: PIRM2018 challenge on spectral image super-resolution: methods and results. In: Leal-Taixé, L., Roth, S. (eds.) ECCV 2018 Workshops. LNCS, vol. 11133, pp. 356–371. Springer, Cham (2018)
41. Shoeiby, M., Robles-Kelly, A., Wei, R., Timofte, R.: PIRM2018 challenge on spectral image super-resolution: dataset and study. In: Leal-Taixé, L., Roth, S. (eds.) ECCV 2018 Workshops. LNCS, vol. 11133, pp. 276–287. Springer, Cham (2018)

42. Simões, M., Bioucas-Dias, J., Almeida, L.B., Chanussot, J.: A convex formulation for hyperspectral image superresolution via subspace-based regularization. IEEE Trans. Geosci. Remote Sens. **53**(6), 3373–3388 (2015)
43. Tarabalka, Y., Chanussot, J., Benediktsson, J.A.: Segmentation and classification of hyperspectral images using watershed transformation. Pattern Recognit. **43**(7), 2367–2379 (2010)
44. Van Nguyen, H., Banerjee, A., Chellappa, R.: Tracking via object reflectance using a hyperspectral video camera. In: CVPRW (2010)
45. Vaswani, A., et al.: Attention is all you need. In: NIPS (2017)
46. Veganzones, M.A., Simoes, M., Licciardi, G., Yokoya, N., Bioucas-Dias, J.M., Chanussot, J.: Hyperspectral super-resolution of locally low rank images from complementary multisource data. IEEE Trans. Image Process. **25**(1), 274–288 (2016)
47. Wagadarikar, A., John, R., Willett, R., Brady, D.: Single disperser design for coded aperture snapshot spectral imaging. Appl. Opt. **47**(10), B44–B51 (2008)
48. Wang, L., Xiong, Z., Gao, D., Shi, G., Zeng, W., Wu, F.: High-speed hyperspectral video acquisition with a dual-camera architecture. In: CVPR (2015)
49. Wang, L., Xiong, Z., Shi, G., Wu, F., Zeng, W.: Adaptive nonlocal sparse representation for dual-camera compressive hyperspectral imaging. IEEE Trans. Pattern Anal. Mach. Intell. **39**(10), 2104–2011 (2017)
50. Wang, L., Xiong, Z., Gao, D., Shi, G., Wu, F.: Dual-camera design for coded aperture snapshot spectral imaging. Appl. Opt. **54**(4), 848–858 (2015)
51. Wug Oh, S., Brown, M.S., Pollefeys, M., Joo Kim, S.: Do it yourself hyperspectral imaging with everyday digital cameras. In: CVPR (2016)
52. Xiong, Z., Shi, Z., Li, H., Wang, L., Liu, D., Wu, F.: HSCNN: CNN-based hyperspectral image recovery from spectrally undersampled projections. In: ICCVW (2017)
53. Yokoya, N., Yairi, T., Iwasaki, A.: Coupled nonnegative matrix factorization unmixing for hyperspectral and multispectral data fusion. IEEE Trans. Geosci. Remote Sens. **50**(2), 528–537 (2012)
54. Yuan, Y., Zheng, X., Lu, X.: Hyperspectral image superresolution by transfer learning. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. **10**(5), 1963–1974 (2017)
55. Zhang, Y., Li, K., Li, K., Wang, L., Zhong, B., Fu, Y.: Image super-resolution using very deep residual channel attention networks. In: Ferrari, V., Hebert, M., Sminchisescu, C., Weiss, Y. (eds.) ECCV 2018. LNCS, vol. 11211, pp. 294–310. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-01234-2_18