Hybrid-Scale Self-Similarity Exploitation for Remote Sensing Image Super-Resolution

Sen Lei, Zhenwei Shi*, *Member IEEE*

***Abstract*—Recently, deep convolutional neural networks (CNN) have made great progress in remote sensing image super-resolution. The CNN-based methods can learn powerful feature representation from plenty of low- and high-resolution counterparts. For remote sensing images, there are many similar ground targets recurred inside the image itself, both within the same scale and across different scales. In this paper, we argue that this internal recurrence can be used for learning stronger feature representation, and we propose a new hybrid-scale self-similarity exploitation network (HSENet) for remote sensing image super-resolution. Specifically, we introduce a single-scale self-similarity exploitation module (SSEM) to compute the feature correlation within the same scale image. Moreover, we design a cross-scale connection structure (CCS) to capture the recurrences across different scales. By combining SSEM and CCS, we further develop a hybrid-scale self-similarity exploitation module (HSEM) to construct the final HSENet, which simultaneously exploits single- and cross-scale similarities. Experimental results demonstrate that HSENet can obtain superior performance over several state-of-the-art methods. Besides, the effectiveness of our method is also verified by the assistance to the remote sensing scene classification task.**

***Index Terms*—Super-resolution, remote sensing images, deep convolutional neural networks, self-similarity**

I. Introduction

Super-resolution (SR) aims to recovery a high-resolution (HR) image from a given low-resolution (LR) image or a series of LR frames. SR technology is widely used in medical imaging [1, 2], video monitoring [3, 4] and remote sensing processing [5, 6]. In the remote sensing filed, high spatial resolution images often play a critical role on many applications such as object detection [7], change detection [8] and scene labeling [9], and thus the pursuit of HR images never ceases. Instead of developing physical imaging devices on board of remote sensing satellite, SR technology is an alternatively effective way to obtain HR remote sensing images [10–12].

Super-resolution from a single image is a very typical ill-posed problem, where image prior is often used to constrain the solution space of potential recovered HR results. In early time, some researchers introduced interpolation-based methods for single image super-resolution (SISR), such as bicubic interpolation and its improved algorithms [13, 14]. These methods are simply designed based on local image prior without any external information, suffering from the blurring of edges, contours, and other image details. After that, a series of learning-based SR algorithms were proposed, such as neighborhood embedding-based methods [15], sparse representation-based methods [16, 17], and local linear regression-based methods [18, 19]. Most of these methods assume that LR image patches and the corresponding HR ones are distributed on different sub-spaces and own a similar local manifold structure, where the learned dictionaries of HR and LR patches are used to perform image reconstruction in the test phase. However, these approaches are all designed based on low-level features, such as image edges, contours, and even image raw pixels, which limits their performances.

Deep convolutional neural networks (CNNs) are introduced in the natural image super-resolution community in recent years and have made great progress in both accuracy and visual performance. Specifically, CNNs can automatically learn high-level feature representations and further obtain superior performance over traditional approaches based on low-level features. SRCNN [20] is the first CNNs-based method specially designed for SR problem, which learns an end-to-end nonlinear mapping between low- and high-resolution images. SRCNN achieves the stat-of-the-arts results with a lightweight network of three convolutional layers. Since then, a large number of deep learning-based SR methods have been proposed in the last few years. Residual learning and residual blocks are incorporated into image super-resolution community to construct very deep SR networks [21–23]. Many researchers proposed recursive structures via reusing certain convolutional layers to increase the depth of SR network with fewer parameters [24, 25]. Furthermore, some works made full use of the features from each layer to obtain abundant feature expression through dense connections [26, 27].

Apart from exploiting the non-linear mapping of external examples, i.e., low- and high-resolution counterparts, some researchers also leverage image self-similarity prior to improve super-resolved results. The image self-similarity refers to the characteristic that similar patches redundantly recur within a single image, which is widely explored in some early works by combining with the classic example-based SR methods [28–30]. In recent, Shocher *et al.* [31] proposed a zero-shot super-resolution (ZSSR) network to only use the self-similarity information within the test LR image. However, for each new test LR image, additional training time would be required for

The work was supported by the National Key RD Program of China under the Grant 2019YFC1510905, the National Natural Science Foundation of China under the Grant 61671037 and the Beijing Natural Science Foundation under the Grant 4192034, and the Academic Excellence Foundation of BUAA for PhD Students under Grant 20200666 (*Corresponding author: Zhenwei Shi*).

Sen Lei and Zhenwei Shi (Corresponding author) are with Image Processing Center, School of Astronautics, Beihang University, Beijing 100191, China, and with Beijing Key Laboratory of Digital Media, Beihang University, Beijing 100191, China, and also with State Key Laboratory of Virtual Reality Technology and Systems, School of Astronautics, Beihang University, Beijing 100191, China. Sen Lei is also with Shenyuan Honors College, Beihang University, Beijing 100191, China. (e-mail: senlei@buaa.edu.cn; shizhenwei@buaa.edu.cn).

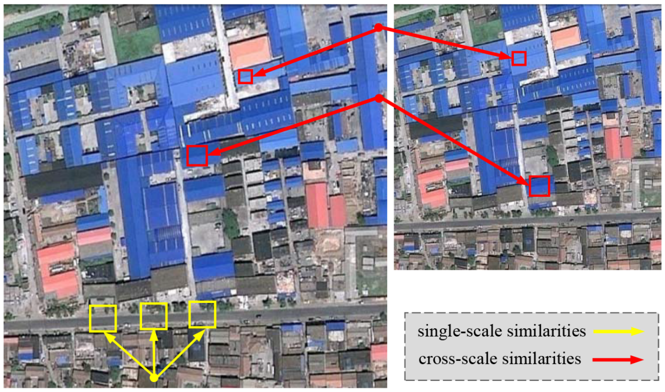

Fig. 1. An illustration of "recurrence of ground target" in remote sensing images with single- and cross-scale similarities. The left image shows the original image and the right one shows a down-scaled version. The similar road patches (marked with yellow boxes) are recurring in the same image scale and the buildings (marked with red boxes) are recurring across scales.

the ZSSR, thus it is not efficient for practical applications.

The remote sensing images also contain many self-similarities, i.e., internal recurrence of information. It often covers relatively large areas and similar ground targets tend to redundantly recur inside the images, both within the same scale and across different scales. Fig. 1 shows an example of "recurrence of ground target" in a typical remote sensing scene. In this example, similar road patches (marked with yellow boxes) are recurring in the same scale image while the roofs of buildings (marked with red boxes) are recurring across scales. These patches have similar appearances such as edges and textures, and this property can be incorporated in SR methods to boost super-resolved performance. In early time, Pan *et al.* [32] introduced the self-similarity prior of remote sensing images into a sparse representation framework. However, the sparse representation of SR is only based on low-level features and it is hard to make full use of the internal recurrences inside the whole remote sensing image. For this problem, a natural question arises that whether we can use the prevalent deep learning to exploit self-similarities of remote sensing images to obtain stronger feature representation.

In this paper, we propose a novel CNN-based super-resolution method to make full use of the internal recurrence of information in remote sensing images. We name our method as hybrid-scale self-similarity exploitation Network (HSENet). Specifically, we introduce a single-scale self-similarity exploitation module (SSEM) to learn the feature correlation within the same image scale, where a non-local operation is employed and its computed relevance is further taken as attentions to adaptively rescale the learned feature. Moreover, we propose a cross-scale connection structure (CCS) to capture the recurrences across different scales, in which an adjusted non-local block is designed to compute the relevance of two feature scales. By combining the SSEM and CCS, we further develop a hybrid-scale self-similarity exploitation module (HSEM) to construct the final HSENet, which simultaneously exploits single- and cross-scale similarities. Experimental results show that our method obtains superior super-resolved results in terms of both accuracy and visual performance.

The main contributions of this papers are summarized as follows:

- We propose a novel CNN-based method named hybrid-scale self-similarity exploitation network (HSENet) for remote sensing image super-resolution. Our proposed method learns single- and cross-scale internal recurrence of patterns in remote sensing images and obtain state-of-the-art SR performance on a public remote sensing dataset.
- We introduce a single-scale self-similarity exploitation module (SSEM) to learn the feature correlation within the same image scale and design a cross-scale connection structure (CCS) to capture the recurrences across different scales. Furthermore, by combining the SSEM and CCS, we develop a hybrid-scale self-similarity exploitation module (HSEM) to construct the final HSENet.

The rest parts of this paper are organized as follows. In Section II, we present related works of image self-similarity and image super-resolution. The framework of our proposed HSENet and details of this model are carefully described in Section III. In Section IV, we give a detailed description of our experimental dataset, ablation studies, experimental results and robust experiments. The final conclusions are drawn in Section V.

II. Related Works

A. Image Self-similarity

Local image patterns tend to redundantly recur in the image with similar contours and textures [28, 33]. The property that the internal data redundancy generally exists within a single image is regarded as *image self-similarity*, which is widely used in many low-level vision tasks including image denoise [34, 35], deblurring [36], super-resolution [29–31] and etc. In super-resolution, the pioneering work based on self-similarity prior was proposed in [28], where Glasner *et al.* proposed a unified framework combined exploited internal patches and example-based SR. Freedman *et al.* [29] followed local self-similarity assumption and extracted localized regional patches to reduce computation time. Furthermore, Yang *et al.* [30] proposed a very fast regression model based on in-place similarity with external- and self-examples. Recently, Shocher *et al.* [31] proposed a zero-shot super-resolution (ZSSR) network to perform unsupervised super-resolution using only the test LR image itself, where the self-similar patches of input LR image were fully exploited. Pan *et al.* [32] introduced structure self-similarity prior combined with sparse representation for remote sensing super-resolution problem. In this paper, for remote sensing image super-resolution, we aim to use the prevalent deep learning to leverage the self-similarity information of remote sensing images to to learn stronger feature representations.

B. CNN-based Image Super-resolution

Dong *et al.* [20] took the lead in applying deep learning to natural image super-resolution. They formulated image super-resolution as a regression task and built a three-layer convolutional neural network to directly learn the nonlinear mapping between low- and high-resolution images. After that, many researchers proposed deep CNN models to obtain more representative features. Kim $et\ al.$ [21] introduced a very deep convoluntinal (VDSR) with 20 layers, where image residuals are learned. Lim $et\ al.$ [22] proposed an enhanced deep super-resolution model (EDSR) based on improved residual blocks without batch noramlization layer. Some works leverage recurrent structures to reuse convoluntional layers in order to improve recovery performance with small model parameters. Kim $et\ al.$ [37] employed recursive blocks to enlarge the receptive field and introduced recursive supervision and skip connections to alleviate the training problem. Tai $et\ al.$ [24] proposed a recursive unit to learn the multi-layer expression of the current state as short-term memory, and by constructing several memory modules, the output was input into the gate unit as a long-term memory to solve the long-term dependence problem caused by the deepening of the network model. In early recent, several methods incorporated attention mechanism into CNN-based super-resolution model to readjust the importance on different feature. RCAN [23] incorporates residual channel attention mechanism to adaptively rescale feature. Dai $et\ al.$ [38] proposed a second-order channel attention module where second-order feature statistics were used to adaptively adjust the channel features, so as to learn more expressive features.

C. Remote Sensing Images Super-resolution

Methods for remote sensing images super-resolution can be broadly classified into two categories: sparse representation-based methods and deep learning-based methods. In early time, Pan $et\ al.$ [32] first introduced the sparse representation into remote sensing image super-resolution field and leveraged structure self-similarity prior to recovery remote sensing high-resolution images. Hou $et\ al.$ [10] developed a global joint dictionary model under global and local constraints to obtain better internal relationships between image patches. Shao $et\ al.$ [39] proposed a coupled sparse autoencoder to learn the mapping relationship between sparse representation coefficient of low-resolution images and high-resolution ones, in order to adopt remote sensing images of different spatial scale.

In recent years, deep learning is extensively used in the remote sensing super-resolution field. Lei $et\ al.$ [40] proposed a deep learning-based remote sensing super-resolution method combining local and global CNN features. Haut $et\ al.$ [41] introduced a deep compendium model (DCM) which integrates some components including residual unit, skip connection, and network-in-network structure. Pan $et\ al.$ [11] proposed residual dense back-projection blocks with up-projection and down-projection modules for remote sensing super-resolution. Moreover, many researchers address the remote sensing super-resolution problem from the point of wavelet analysis [42, 43]. Wang $ea\ al.$ [42] utilized several parallel shallow convolutional neural networks to learn different wavelet band information in different scales. Ma $et\ al.$ [43] transformed remote sensing images into wavelet domain, and proposed a recursive ResNet to learn LR-HR mapping on the wavelet domain. Zhang $et\ al.$ [44] introduced mixed high-order attention for feature extraction. Zhang $et\ al.$ [45] proposed a multiscale attention network (MSAN) to extract the multi-level features of remote sensing images and employed a scene-adaptive strategy to describe structural information on different scenes. Qin $et\ al.$ [46] introduced a deep gradient-aware network with image-specific enhancement (DGANet-ISE) and designed a gradient-aware loss to preserve the important gradient information of remote sensing images.

III. Methodology

In this section, we introduce the proposed hybrid-scale self-similarity exploitation network (HSENet) for remote sensing image super-resolution. We will first give a brief introduction to the overall framework of our method. The core of our method, including the single-Scale self-similarity exploitation module (SSEM) and the hybrid-scale self-similarity exploitation module (HSEM), are then discussed in Section III-B and Section III-C, respectively. The implementation details is provided in Section III-D.

A. Overall Framework

The overall framework of HSENet is illustrated in Fig. 2. Referring to some state-of-the-art methods [22, 23, 38], our method consists of three parts: shallow feature extraction part, deep feature extraction part and reconstruction part. The shallow feature extraction part is to extract the initial shallow feature of the LR inputs. We use only one convolution layer C_{SF} with kernel 3×3 to obtain the shallow feature F_0:

$$F_0 = C_{SF}(I_{LR}) \tag{1}$$

where I_{LR} and C_{SF} denote the LR input and the convolutional operation, respectively. F_0 is then taken as the input of the following deep feature extraction layers, and the extracted feature F_n can be computed as

$$F_n = BM_n(F_{n-1}) = BM_n(BM_{n-1}(...BM_0(F_0))...)) \tag{2}$$

where BM_n represents the n-th basic module, and F_{n-1} is the input of BM_n and F_n is the corresponding output. As Fig. 2 shows, the proposed hybrid-scale self-similarity exploitation module (HSEM) is a core part of the BM, which aims to learn more powerful feature representation by exploiting image self-similarity.

The final super-resolved output is further obtained through the reconstruction layers

$$I_{SR} = R(F_t + F_0) \tag{3}$$

where I_{SR} is the final super-resolved image and R denotes the reconstruction layers where residual learning is used to accelerate convergence speed. The main component of R is the up-sample layer shown in Fig. 2 which sub-pixel convolutions [47] are employed.

We train the above network by using a pixel-wise L1 loss function. Given super-resolved image I_{SR} and the corresponding high-resolution reference I_{HR}, the loss can be obtained as

$$L(\theta) = \frac{1}{N}\sum_{i=1}^{N} ||I_{SR}^{(i)} - I_{HR}^{(i)}||_1. \tag{4}$$

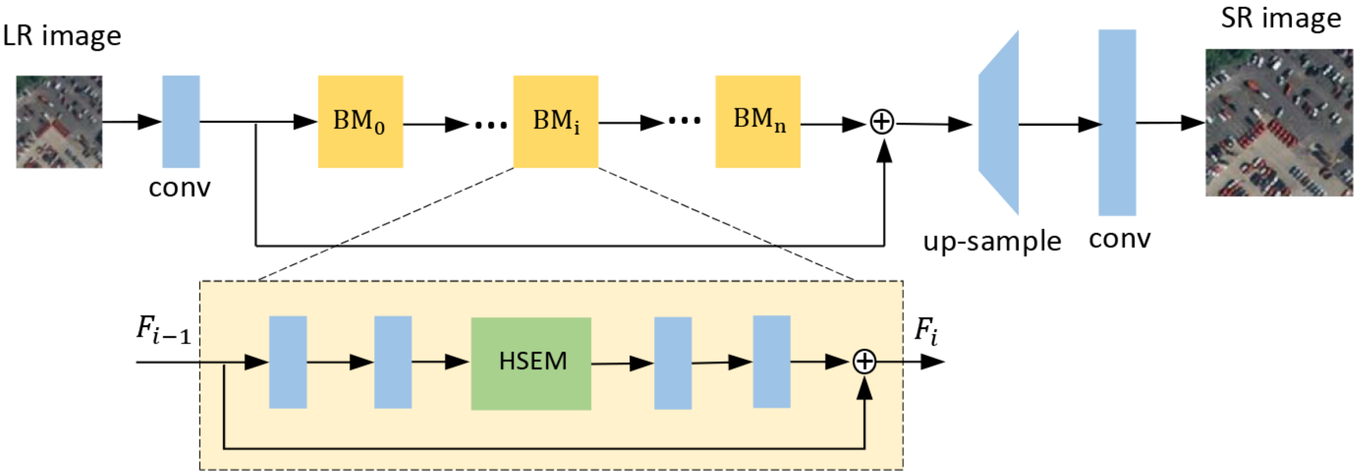

Fig. 2. The framework of our proposed model. The yellow block BM_i represents the basic module of our model, and HSEM denotes the hybrid-scale self-similarity exploitation module.

where N is the number of training images.

B. Single-Scale Self-similarity Exploitation Module

We first introduce single-scale self-similarity exploitation module (SSEM) to mine the feature correlation within the same scale of remote sensing images, and then present the hybrid-scale self-similarity exploitation module (HSEM) described in the next subsection is built upon the SSEM.

Traditional convolutional layers can only cover limited receptive fields and thus the relationships within local pixels would be explored. However, the non-local block (NLB) [48] can compute relevance among the whole input pixels and allow the network to concentrate more on informative areas. It can be regarded as one kind of self-attention models. Here, we incorporate the non-local operation into the SSEM to compute feature correlations, and the extracted self-similarity information is further taken as attentions for learning stronger feature representations.

As shown in Fig. 3 (a), we elaborately design the main branch and attention branch to perform single-scale feature representation. Inspired by some attention-based methods [49–51], we use the self-similarity information extracted by NLB to as attentions to better guide high-frequency feature extraction in the main branch. Specifically, in the main branch, two convolutional layers are utilized to extract higher-level features, and a non-local block (NLB) is employed in the attention branch to adaptively rescale the features upon the main branch with element-wise production.

Specifically, the non-local operation can be formulated as follows

$$y_i = \left(\sum_{\forall j} f(x_i, x_j) g(x_j) \right) / \sum_{\forall j} f(x_i, x_j) \tag{5}$$

where i is the index of the output position and j is the index that enumerates all positions. x and y denote the input and output of this operation, as shown in Fig. 3 (b). The pairwise function f can compute the correlation between x_i and all the x_j, and the function g extracts the feature representation of x_j. The single-scale self-similarity can be obtained by this pairwise operation and plays an important role in the attention branch.

We here use a embedding Gaussian function to learn the pairwise similarity:

$$f(x_i, x_j) = exp(\theta^T(x_i)\varphi(x_j)) \tag{6}$$

where $\theta(x_i) = W_\theta x_i$ and $\varphi(x_j) = W_\varphi x_j$ are the embeddings of x_i and x_j, respectively. Meanwhile, in order to reduce the computation cost of the non-local operation, the dimensions of x_i and x_j are controlled by a factor parameter r. Finally the output of the NLB is further obtained:

$$z_i = W_\phi y_i + x_i \tag{7}$$

where W_ϕ is a weight matrix and y_i can be rewritten as $softmax((W_\theta x_i)^T W_\varphi x_j)$.

The convolutional layer following the NLB is employed to transform the inputs to attention maps, which are then normalized by a sigmoid function. Furthermore, the output features of the main branch will multiply the attention maps, where the activation values of each spatial and channel position are rescaled.

C. Hybrid-Scale Self-Similarity Exploitation Module

In this subsection, we introduce the HSEM based on the aforementioned SSEM to simultaneously leverages single- and cross-scale similarity information of remote sensing images, which is illustrated in Fig 4 (a). Let us denote f_{in}^b as inputs of the HSEM which is also consider as the feature from a basic scale. In order to exploit the internal recurrence of information in different scales, the feature of a down-sampled scale f_{in}^d would be acquired by:

$$f_{in}^d = D_s(f_{in}^b) \tag{8}$$

where D_s represents the down-sample operation with scale factor s.

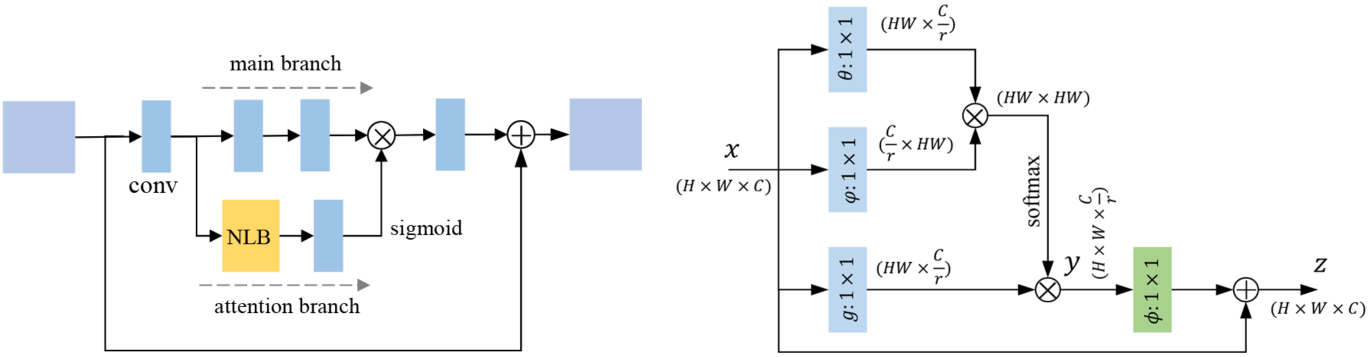

Fig. 3. Left: single-scale self-similarity exploitation module (SSEM). Right: the details of non-local block (NLB).

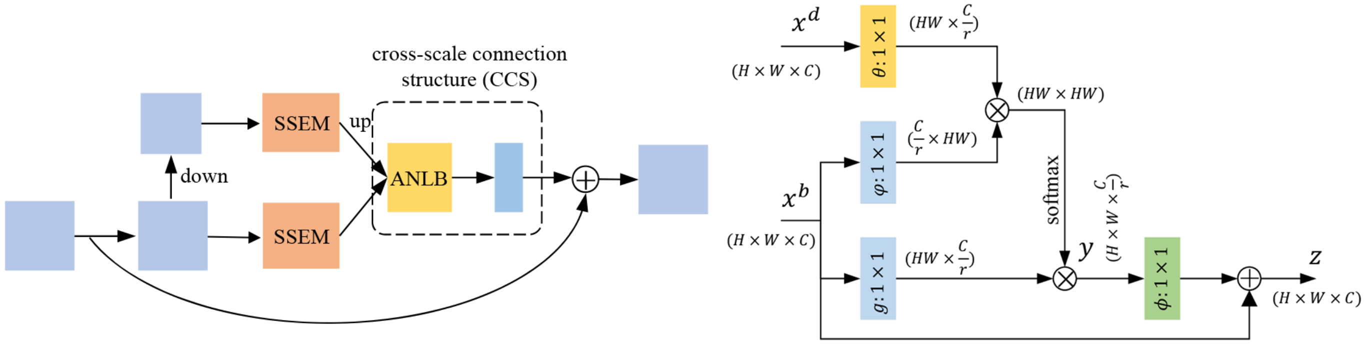

Fig. 4. Left: hybrid-scale self-similarity exploitation module (HSEM). Right: details of the adjusted non-local block (ANLB).

We then use two SSEMs to extract powerful feature representations with two different scales f_{in}^b and f_{in}^d respectively by leveraging the correlations within their whole scopes. The output of the down-sampled scale would be further up-sampled by a same scale factor s. Here x^b and x^d represent the output of the basic scale and the one of the down-sampled scale through the SSEMs, respectively, which are formulated as

$$\begin{aligned} x^b &= SSEM(f_{in}^b), \\ x^d &= U_s(SSEM(f_{in}^d)). \end{aligned} \tag{9}$$

where the U_s denotes an up-sample operation with scale factor s, and x^d has the same dimensions with x^b.

Moreover, we design a cross-scale connection structure (CCS) to exploit the similarity between x^b and x^d. An adjusted non-local block (ANLB) is the main component of this CCS, which is specially designed to leverage the relevance between two remote sensing image scales. As illustrated in Fig 4 (b), the main difference between ANLB and NLB is the input structure, and the followed self-similarity computation working flows are similar. Thus the y_i in formula (10) for ANLB would be rewritten as

$$y_i = \left(\sum_{\forall j} f(x_i^d, x_j^b) g(x_j^b)\right) / \sum_{\forall j} f(x_i^b, x_j^d) \tag{10}$$

where the $f(x_i^d, x_j^b)$ is computed as $exp(\theta^T(x_i^d)\varphi(x_j^b))$. It should be emphasized that x^b and x^d play different roles in the ANLB and x^d is only used in the computation of the pairwise function. In the cross-scale connection structure, the ANLB can fuse multiple scale features and leverage the similarity between them. One convolutional layer is then applied to further map the fusion features for output.

It should be noted that local skip connections are utilized both in the SSEM and the HSEM. These connections can be regarded as one kind of residual feature learning, and it allows us to form very deep networks with little training problems.

D. Implementation Details

In this paper, we focus on ×2, ×3 and ×4 scale factors and the up-sample blocks in the reconstruction part will be slightly adjusted according to the specific scale factors. In the training phase, 48×48 image patches will be randomly cropped from LR images and their ground-truth references will be extracted from HR ones corresponding to the scale factor. Furthermore, the training images are augmented by random rotation 90°, 180°, 270° and horizontally flipping. We finally set the number of BM to be ten and the parameter r and s are both set as 2. Besides, bicubic interpolation operation is used to performer the down-sampled and up-sampled in the HSEM.

We use Adam optimizer [52] to train our model with $\beta_1 = 0.9$, $\beta_2 = 0.99$, and $\epsilon = 10^{-8}$. The initial learn rate is set as 10^{-4} and the mini-batch size is 4. The overall training epochs are 500 and the learn rate decreases half at 400 epochs. Our proposed method is implemented by PyTorch [53] and all the experiments are run on a NIVIDIA GeForce GTX 1080Ti graphics card. Our codes will be publicly available at https://github.com/Shaosifan/HSENet.

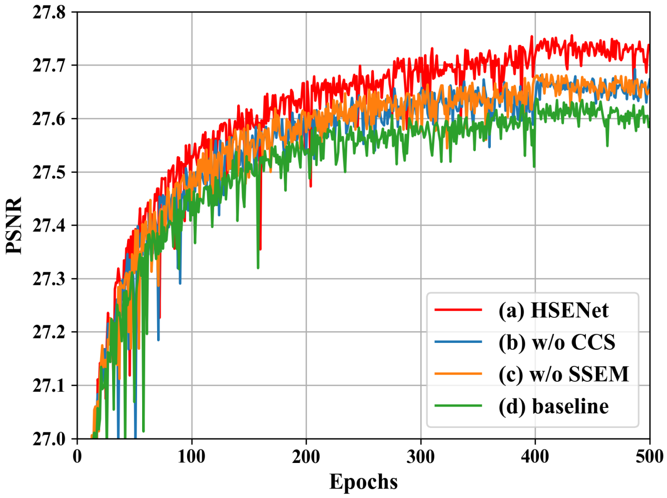

Fig. 5. PSNR comparison on the UCMerced test dataset on scale factor x4 during the overall training phase. (a) HSENet. (b) HSENet without cross-scale connection structure (CCS). (c) HSENet without single-scale self-similarity enhancement module (SSEM). (d) the baseline model.

IV. Experimental Results and Analyses

A. Experimental Data set and Settings

We select UCMecred Data Set (UCMerced) [54] to demonstrate the effectiveness of our proposed method. This data set has been extensively used in remote sensing super-resolution field for evaluation [40, 41, 46]. Specifically, UCMerced contains 21 classes in total which covers several remote sensing scenes, such as agricultural, airplane, baseball-diamond, beach and etc. There are 100 images for each class and all images are around in 256×256 pixels with a relatively high spatial resolution of 0.3 m/pixel. Following [41][46], the dataset is split into two balanced halves as training and test sets with 1050 samples each. In our experiments, LR images are downsampled from HR images by bicubic interpolation operation, and the corresponding HR ones are regarded as ground truth. All results are evaluated by peak signal-to-noise ratio (PSNR) and structural similarity index measure (SSIM) [55].

We further verify the robustness of HSENet on some real multi-spectral remote sensing data from GaoFen-1 and GaoFen-2 satellites. Three visible bands of these images are selected to generate RGB images and serve as LR inputs. Moreover, we conduct some experiments on NWPU-RESISC45 [56] to prove the assistance of our method to the remote sensing scene classification task.

B. Ablation Studies

In this subsection, we conduct some ablation studies to demonstrate the effectiveness of the main components of our proposed method, including single-scale self-similarity enhancement module (SSEM) and cross-scale connection structure (CCS). For the baseline model, we use 10 convolutional layers with a local skip connection to replace the basic module in our proposed method so that it has similar total parameters with other variants. Fig. 5 shows the PSNR comparisons on the UCMerced test dataset with scale factor $\times 4$, where the number of basic modules is set as 10. It verifies the effectiveness of the proposed components of SSEM and CCS. Meanwhile, our final HSENet obtains better super-resolved results with a faster convergence.

In Table I, we further explore the influence of the numbers of the basic module, where the baseline model and the proposed HSENet are compared. Table I shows the PSNR and SSIM evaluations on the UCMerced test dataset with scale factor $\times 4$. It can be seen that our method obtain the best performance when the number is 10 and has $+0.115dB$ higher than the corresponding baseline. Meanwhile, it should be noted that when the number of the basic module becomes larger the performance tends to degrade, which implies the occurrence of overfitting on UCMerced dataset.

C. Comparison with Other Methods

We further compare our method with some super-resolution methods including traditional bicubic interpolation, SC [57], SRCNN [58], FSRCNN [59], LGCNet [40], DCM [41], and DGANet-ISE [46] on UCMerced test data set. Bicubic interpolation is usually used as a weak baseline method in most SR literature. Among these methods, SC, SRCNN and FSRCNN are proposed in the natural image super-resolution field, and LGCNet, DCM and DGANet-ISE are recently proposed deep learning-based methods specifically designed for remote sensing super-resolution problem. The evaluation results of these comparisons on UCMerced test dataset are reported in some published works [40, 41, 46].

Table II lists the average quantitative evaluation results of different methods over all the UCMerced test data set for scale $\times 2$, $\times 3$ and $\times 4$. It can be seen that our method achieves the best performance in terms of PSNR on all scales. Specifically, the PSNR gains of our method over the second-best DGANet-ISE is 0.54 dB and 0.42 dB for scale 2 and scale 4, respectively. Moreover, the PSNR which our method obtains is 0.48 dB higher than DCM for the scale 3. In the case of SSIM metric, our method obtains the second performances which are 0.0017 and 0.0042 lower than DGANet-ISE for scale 2 and scale 4. Fig. 6 illustrates some reconstruction results of these methods. Compared with other methods, the high-resolution results recovered by our method have clearer edges and contours.

Moreover, Table III provides the detailed performances of different methods for scale factor $\times 3$ on all 21 classes [1] of UCMeced dataset. Since DGANet-ISE dose not report their performances on scale 3, the results of DGANet-ISE is not involved in Table III. From the results, it can be observed that our model obtains the best PSNR results in 13 UCMerced scene categories, and the second-best DCM obtains the best PSNR in other 8 categories. Comparing with DCM, our method is more effective in some scenes which contain rich edges and contours, such as 'Buildings', 'Denseresidential', 'Freeway', 'Storagetanks' and etc. Meanwhile, the proposed

[1] All these 21 classes: 1—Agricultural, 2—Airplane, 3—Baseballdiamond, 4—Beach, 5—Buildings, 6—Chaparral, 7—Denseresidential, 8—Forest, 9—Freeway, 10—Golfcourse, 11—Harbor, 12—Intersection, 13—Medium-residential, 14—Mobilehomepark, 15—Overpass, 16—Parkinglot, 17—River, 18—Runway, 19—Sparseresidential, 20—Storagetanks, 21—Tenniscourt.

TABLE I
PSNR(dB) AND SSIM RESULTS WITH DIFFERENT NUMBERS OF BASIC MODULE (BM) ON THE BASELINE METHOD AND THE PROPOSED METHOD FOR UCMERCED DATASET (X4 SCALE FACTOR).

Num. of BM	6	8	10	12	14
Baseline	27.568 / 0.7558	27.592 / 0.7572	27.619 / 0.7569	27.625 / 0.7578	27.615 / 0.7569
Proposed	27.668 / 0.7605	27.708 / 0.7611	**27.734 / 0.7623**	27.690 / 0.7611	27.687 / 0.7607

TABLE II
MEAN PSNR (dB) AND SSIM OVER THE UCMERCED TEST DATA SET

scale	Bicubic	SC[57]	SRCNN[58]	FSRCNN[59]	LGCNet[40]	DCM[41]	DGANet-ISE[46]	ours
	PSNR / SSIM	PSNR / SSIM	PSNR / SSIM	PSNR / SSIM	PSNR / SSIM	PSNR / SSIM	PSNR / SSIM	PSNR / SSIM
2	30.76 / 0.8789	32.77 / 0.9166	32.84 / 0.9152	33.18 / 0.9196	33.48 / 0.9235	33.65 / 0.9274	33.68 / **0.9344**	**34.22** / 0.9327
3	27.46 / 0.7631	28.26 / 0.7971	28.66 / 0.8038	29.09 / 0.8167	29.28 / 0.8238	29.52 / 0.8394	– / –	**30.00 / 0.8420**
4	25.65 / 0.6725	26.51 / 0.7152	26.78 / 0.7219	26.93 / 0.7267	27.02 / 0.7333	27.22 / 0.7528	27.31 / **0.7665**	**27.73** / 0.7623

TABLE III
MEAN PSNR (dB) OF EACH CLASS FOR UPSCALING FACTOR 3

class	Bicubic	SC [57]	SRCNN [58]	FSRCNN [59]	LGCNet [40]	DCM [41]	HSENet (ours)
1	26.86	27.23	27.47	27.61	27.66	**29.06**	27.64
2	26.71	27.67	28.24	28.98	29.12	**30.77**	30.09
3	33.33	34.06	34.33	34.64	34.72	33.76	**35.05**
4	36.14	36.87	37.00	37.21	37.37	36.38	**37.69**
5	25.09	26.11	26.84	27.50	27.81	28.51	**28.95**
6	25.21	25.82	26.11	26.21	26.39	**26.81**	26.70
7	25.76	26.75	27.41	28.02	28.25	28.79	**29.24**
8	27.53	28.09	28.24	28.35	28.44	28.16	**28.59**
9	27.36	28.28	28.69	29.27	29.52	30.45	**30.63**
10	35.21	35.92	36.15	36.43	36.51	34.43	**36.62**
11	21.25	22.11	22.82	23.29	23.63	**26.55**	24.88
12	26.48	27.20	27.67	28.06	28.29	**29.28**	29.21
13	25.68	26.54	27.06	27.58	27.76	27.21	**28.55**
14	22.25	23.25	23.89	24.34	24.59	**26.05**	25.70
15	24.59	25.30	25.65	26.53	26.58	27.77	**28.22**
16	21.75	22.59	23.11	23.34	23.69	**24.95**	24.66
17	28.12	28.71	28.89	29.07	29.12	28.89	**29.22**
18	29.30	30.25	30.61	31.01	31.15	**32.53**	31.15
19	28.34	29.33	29.40	30.23	30.53	29.81	**31.64**
20	29.97	30.86	31.33	31.92	32.17	29.02	**32.95**
21	29.75	30.62	30.98	31.34	31.58	30.76	**32.71**
AVG	27.46	28.23	28.66	29.09	29.28	29.52	**30.00**

method achieve higher PSNR of 0.48 dB than DCM for the overall evaluation. It also can been found that the PSNR results are very different for different scenes, in which the PNSR for "Beach" (class4) images is 37.69 dB but the PSNR for "Parking lot" (class16) images is only 24.66 dB. The reason for this phenomenon is that the image contents of different remote sensing scenes are widely varied, i.e, the scene of "Parking lot" owns more high-frequency information than the "Beach". The very smoothing scenes such as "Baseballdiamond" (class3), "Beach" and "Golfcourse" (class10) tend to have higher PSNR results, where little high-frequency information should be super-resolved.

D. Results on Real Remote Sensing Data

Previous experiments are conducted on the UCMerced data set, which contains high-resolution aerial remote sensing images with a spatial resolution of 0.3 m per pixel. In order to further verify the reconstruction ability of our method, we here use some real multi-spectral data from GaoFen-1 (GF-1) and GaoFen-2 (GF-2) satellites. The spatial resolutions of GF-1 and GF-2 are 8 m and 3.2 m per pixel respectively. Three visible bands of these images are selected to generate RGB images, which serve as LR inputs in this experiment. We use the pre-trained HSENet model on UCMerced data set to recovery high-frequency details given these LR inputs. As shown in Fig. 7 and Fig. 8, our HSENet can obtain promising results when dealing with real remote sensing data on some typical scenes including road, factories, paddy fields and buildings. Although the spatial resolutions of these inputs are different from LR images in the training data set, which are 0.9 m/pixel and 1.2 m/pixel for scale factor ×3 and ×4 respectively, our method still can improve the visual perception qualities of remote sensing images. It verifies the generalization of our HSENet.

E. Effects on Remote Sensing Scene Classification

Image super-resolution is often regarded as a pre-processing step for some high-level tasks such as image classification[60], small object detection [61] and etc. Specifically, when the inputs are LR images, SR methods can provide more image details and are beneficial to the downstream tasks. In order to further verify the effectiveness of the proposed HSENet, we conduct another experiment using a remote sensing images classification dataset - NWPU-RESISC45 (NWPU) [56]. NWPU contains 45 different scenes with 700 images per class, and the size of every image is 256×256. In this experiment, we randomly split NWPU dataset into two halves where one is for training and the rest for the test. ResNet-50 [62] is then re-trained on the training data set, whose weights are initialized with the model training on ImageNet. In the test phase, we take the original test data as HR images, and the corresponding LR ones are produced by down-sampling with scale factor ×4. The LR images are further super-resolved by many methods including bicubic interpolation, LGCNet[40], EDSR [22], RCAN [23] and our HSENet. The fine-tuned ResNet-50 is then used as a classifier for the super-resolved images, and Table IV lists the classification performance. The 'Ground Truth' in Table IV represents the original test HR images which are references for other methods. The ResNet-

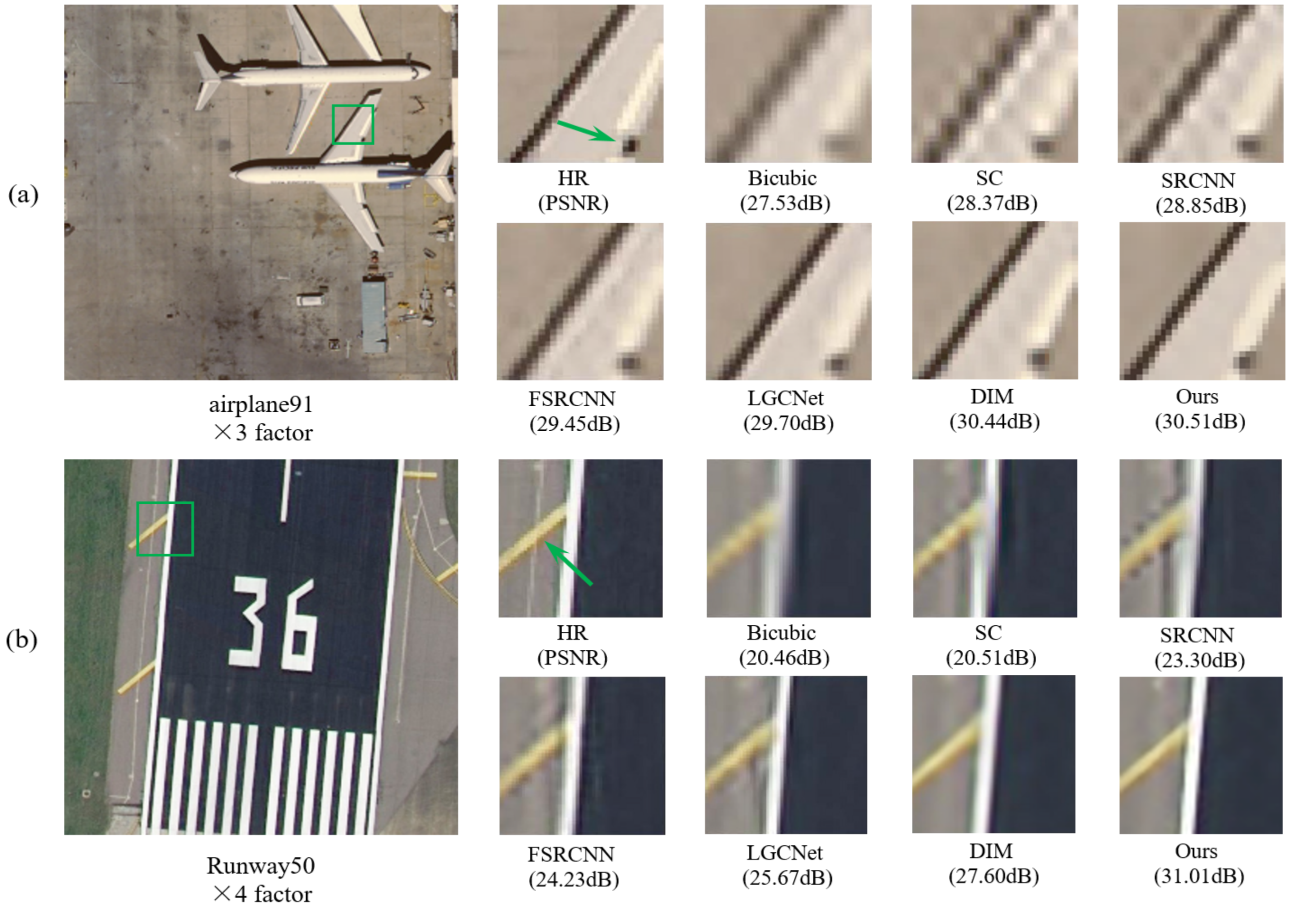

Fig. 6. Comparison of super-resolved outputs with different methods from UCMerced dataset: (a) *airplane91* with x3 scale factor; (b) *runway50* with x4 scale factor. (best view via zoom in)

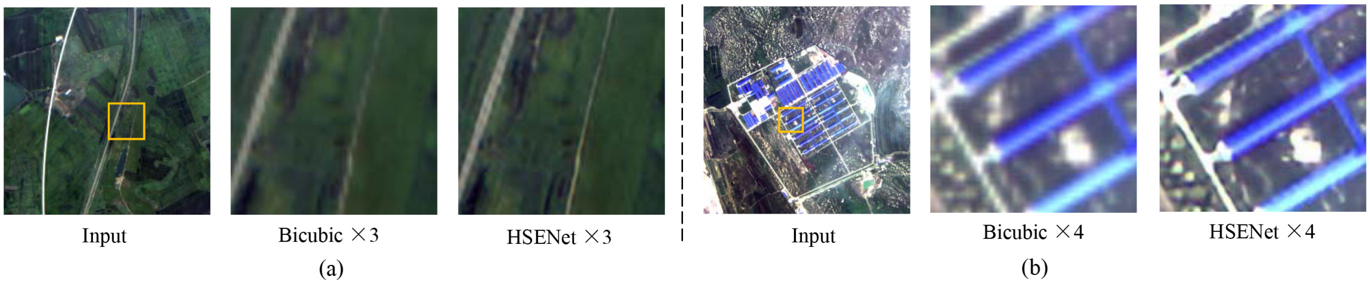

Fig. 7. Validation on real GaoFen-1 satellite data: (a) Road. (b) Factories.

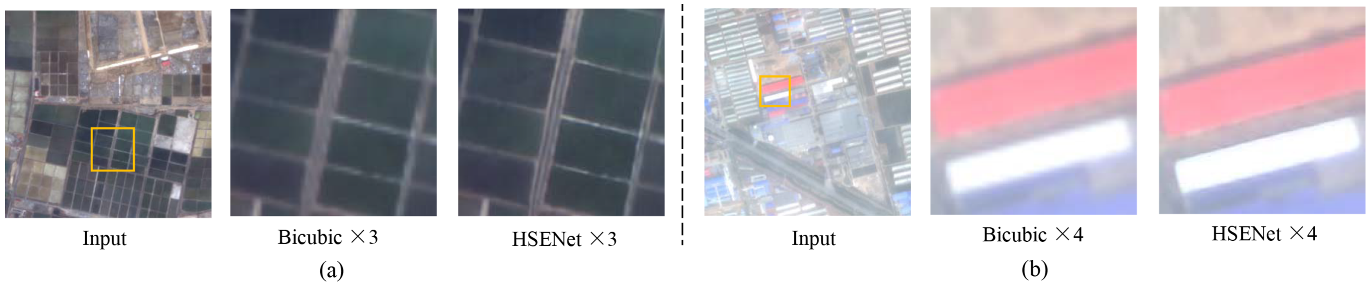

Fig. 8. Validation on real GaoFen-2 satellite data: (a) Paddy fields. (b) Buildings

TABLE IV
ACCURACY RATES (%) ON THE SUPER-RESOLVED IMAGES OF DIFFERENT METHODS GIVEN NWPU LR TEST IMAGES (SCALE FACTOR 4). THE CLASSIFIER IS A FINE-TUNED RESNET-50.

Methods	Top-1 Acc.	Top-5 Acc.
Ground Truth	94.77	99.54
Bicubic	74.04	91.82
LGCNet	78.75	94.91
EDSR	85.99	97.78
RCAN	85.83	97.67
ours	**86.22**	**97.80**

50 obtains the highest Top-1 and Top-5 accuracy on the super-resolved images of our method, which implies that our method can recovery more details of ground target and contributes to more accurate remote sensing scene classification than the other SR methods.

V. CONCLUSION

In this paper, we propose a novel hybrid-scale self-similarity exploitation network (HSENet) for remote sensing image super-resolution. The HSENet effectively leverages the internal recurrence of information both in single- and cross-scale within the images. We introduce a single-scale self-similarity exploitation module (SSEM) to mine the feature correlation within the same scale image. Meanwhile, we design a cross-scale connection structure (CCS) to capture the recurrences across different scales. By combining SSEM and CCS, we further develop a hybrid-scale self-similarity exploitation module (HSEM) to construct the final HSENet. The ablation studies demonstrate the effectiveness of the main components of the HSENet. Our method obtains better super-resolved results on UCMerced data set than several state-of-the-art approaches in terms of both accuracy and visual performance. Moreover, experiments on real-world satellite data (GF-1 and GF-2) verify the robustness of HSENet, and the experiments on NWPU data set show that the details of ground targets recovered by our method can contribute to more accurate classification when given low-resolution inputs.

REFERENCES

[1] H. Greenspan, "Super-resolution in medical imaging," *The computer journal*, vol. 52, no. 1, pp. 43–63, 2009.

[2] J. S. Isaac and R. Kulkarni, "Super resolution techniques for medical image processing," in *2015 International Conference on Technologies for Sustainable Development (ICTSD)*. IEEE, 2015, pp. 1–6.

[3] L. Zhang, H. Zhang, H. Shen, and P. Li, "A super-resolution reconstruction algorithm for surveillance images," *Signal Processing*, vol. 90, no. 3, pp. 848–859, 2010.

[4] P. Rasti, T. Uiboupin, S. Escalera, and G. Anbarjafari, "Convolutional neural network super resolution for face recognition in surveillance monitoring," in *International conference on articulated motion and deformable objects*. Springer, 2016, pp. 175–184.

[5] H. Ji, Z. Gao, T. Mei, and B. Ramesh, "Vehicle detection in remote sensing images leveraging on simultaneous super-resolution," *IEEE Geoscience and Remote Sensing Letters*, vol. 17, no. 4, pp. 676–680, 2019.

[6] J. Shermeyer and A. Van Etten, "The effects of super-resolution on object detection performance in satellite imagery," in *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, 2019, pp. 0–0.

[7] Z. Zou and Z. Shi, "Ship detection in spaceborne optical image with svd networks," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 10, pp. 5832–5845, 2016.

[8] H. Chen and Z. Shi, "A spatial-temporal attention-based method and a new dataset for remote sensing image change detection," *Remote Sensing*, vol. 12, no. 10, p. 1662, 2020.

[9] B. Pan, Z. Shi, X. Xu, T. Shi, N. Zhang, and X. Zhu, "Coinnet: Copy initialization network for multispectral imagery semantic segmentation," *IEEE Geoscience and Remote Sensing Letters*, vol. 16, no. 5, pp. 816–820, 2018.

[10] B. Hou, K. Zhou, and L. Jiao, "Adaptive super-resolution for remote sensing images based on sparse representation with global joint dictionary model," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 56, no. 4, pp. 2312–2327, 2017.

[11] Z. Pan, W. Ma, J. Guo, and B. Lei, "Super-resolution of single remote sensing image based on residual dense backprojection networks," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 10, pp. 7918–7933, 2019.

[12] S. Lei, Z. Shi, and Z. Zou, "Coupled adversarial training for remote sensing image super-resolution," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 58, no. 5, pp. 3633–3643, 2019.

[13] S. Thurnhofer and S. K. Mitra, "Edge-enhanced image zooming," *Optical Engineering*, vol. 35, no. 7, pp. 1862–1870, 1996.

[14] F. Fekri, R. M. Mersereau, and R. W. Schafer, "A generalized interpolative vq method for jointly optimal quantization and interpolation of images," in *Proceedings of the 1998 IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP'98 (Cat. No. 98CH36181)*, vol. 5. IEEE, 1998, pp. 2657–2660.

[15] H. Chang, D.-Y. Yeung, and Y. Xiong, "Super-resolution through neighbor embedding," in *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004. CVPR 2004.*, vol. 1. IEEE, 2004, pp. I–I.

[16] J. Yang, J. Wright, T. Huang, and Y. Ma, "Image super-resolution as sparse representation of raw image patches," in *2008 IEEE conference on computer vision and pattern recognition*. IEEE, 2008, pp. 1–8.

[17] J. Yang, Z. Wang, Z. Lin, S. Cohen, and T. Huang, "Coupled dictionary training for image super-resolution," *IEEE transactions on image processing*, vol. 21, no. 8, pp. 3467–3478, 2012.

[18] R. Timofte, V. De Smet, and L. Van Gool, "Anchored neighborhood regression for fast example-based super-

resolution," in *Proceedings of the IEEE international conference on computer vision*, 2013, pp. 1920–1927.

[19] ——, "A+: Adjusted anchored neighborhood regression for fast super-resolution," in *Asian conference on computer vision*. Springer, 2014, pp. 111–126.

[20] C. Dong, C. C. Loy, K. He, and X. Tang, "Image super-resolution using deep convolutional networks," *IEEE transactions on pattern analysis and machine intelligence*, vol. 38, no. 2, pp. 295–307, 2015.

[21] J. Kim, J. Kwon Lee, and K. Mu Lee, "Accurate image super-resolution using very deep convolutional networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 1646–1654.

[22] B. Lim, S. Son, H. Kim, S. Nah, and K. Mu Lee, "Enhanced deep residual networks for single image super-resolution," in *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, 2017, pp. 136–144.

[23] Y. Zhang, K. Li, K. Li, L. Wang, B. Zhong, and Y. Fu, "Image super-resolution using very deep residual channel attention networks," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 286–301.

[24] Y. Tai, J. Yang, and X. Liu, "Image super-resolution via deep recursive residual network," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 3147–3155.

[25] Y. Tai, J. Yang, X. Liu, and C. Xu, "Memnet: A persistent memory network for image restoration," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 4539–4547.

[26] Y. Zhang, Y. Tian, Y. Kong, B. Zhong, and Y. Fu, "Residual dense network for image super-resolution," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 2472–2481.

[27] M. Haris, G. Shakhnarovich, and N. Ukita, "Deep back-projection networks for super-resolution," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 1664–1673.

[28] D. Glasner, S. Bagon, and M. Irani, "Super-resolution from a single image," in *2009 IEEE 12th international conference on computer vision*. IEEE, 2009, pp. 349–356.

[29] G. Freedman and R. Fattal, "Image and video upscaling from local self-examples," *ACM Transactions on Graphics (TOG)*, vol. 30, no. 2, pp. 1–11, 2011.

[30] J. Yang, Z. Lin, and S. Cohen, "Fast image super-resolution based on in-place example regression," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2013, pp. 1059–1066.

[31] A. Shocher, N. Cohen, and M. Irani, ""zero-shot" super-resolution using deep internal learning," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 3118–3126.

[32] Z. Pan, J. Yu, H. Huang, S. Hu, A. Zhang, H. Ma, and W. Sun, "Super-resolution based on compressive sensing and structural self-similarity for remote sensing images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 51, no. 9, pp. 4864–4876, 2013.

[33] M. Zontak and M. Irani, "Internal statistics of a single natural image," in *CVPR 2011*. IEEE, 2011, pp. 977–984.

[34] A. Buades, B. Coll, and J.-M. Morel, "A non-local algorithm for image denoising," in *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, vol. 2. IEEE, 2005, pp. 60–65.

[35] J. Xu, L. Zhang, W. Zuo, D. Zhang, and X. Feng, "Patch group based nonlocal self-similarity prior learning for image denoising," in *Proceedings of the IEEE international conference on computer vision*, 2015, pp. 244–252.

[36] T. Michaeli and M. Irani, "Blind deblurring using internal patch recurrence," in *European conference on computer vision*. Springer, 2014, pp. 783–798.

[37] J. Kim, J. Kwon Lee, and K. Mu Lee, "Deeply-recursive convolutional network for image super-resolution," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 1637–1645.

[38] T. Dai, J. Cai, Y. Zhang, S.-T. Xia, and L. Zhang, "Second-order attention network for single image super-resolution," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2019, pp. 11 065–11 074.

[39] Z. Shao, L. Wang, Z. Wang, and J. Deng, "Remote sensing image super-resolution using sparse representation and coupled sparse autoencoder," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 12, no. 8, pp. 2663–2674, 2019.

[40] S. Lei, Z. Shi, and Z. Zou, "Super-resolution for remote sensing images via local–global combined network," *IEEE Geoscience and Remote Sensing Letters*, vol. 14, no. 8, pp. 1243–1247, 2017.

[41] J. M. Haut, M. E. Paoletti, R. Fernández-Beltran, J. Plaza, A. Plaza, and J. Li, "Remote sensing single-image super-resolution based on a deep compendium model," *IEEE Geoscience and Remote Sensing Letters*, vol. 16, no. 9, pp. 1432–1436, 2019.

[42] T. Wang, W. Sun, H. Qi, and P. Ren, "Aerial image super resolution via wavelet multiscale convolutional neural networks," *IEEE Geoscience and Remote Sensing Letters*, vol. 15, no. 5, pp. 769–773, 2018.

[43] W. Ma, Z. Pan, J. Guo, and B. Lei, "Achieving super-resolution remote sensing images via the wavelet transform combined with the recursive res-net," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 6, pp. 3512–3527, 2019.

[44] D. Zhang, J. Shao, X. Li, and H. T. Shen, "Remote sensing image super-resolution via mixed high-order attention network," *IEEE Transactions on Geoscience and Remote Sensing*, 2020.

[45] S. Zhang, Q. Yuan, J. Li, J. Sun, and X. Zhang, "Scene-adaptive remote sensing image super-resolution using a multiscale attention network," *IEEE Transactions on Geoscience and Remote Sensing*, 2020.

[46] M. Qin, S. Mavromatis, L. Hu, F. Zhang, R. Liu, J. Sequeira, and Z. Du, "Remote sensing single-image

resolution improvement using a deep gradient-aware network with image-specific enhancement," *Remote Sensing*, vol. 12, no. 5, p. 758, 2020.

[47] W. Shi, J. Caballero, F. Huszár, J. Totz, A. P. Aitken, R. Bishop, D. Rueckert, and Z. Wang, "Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 1874–1883.

[48] X. Wang, R. Girshick, A. Gupta, and K. He, "Non-local neural networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 7794–7803.

[49] F. Wang, M. Jiang, C. Qian, S. Yang, C. Li, H. Zhang, X. Wang, and X. Tang, "Residual attention network for image classification," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 3156–3164.

[50] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 7132–7141.

[51] Y. Zhang, K. Li, K. Li, B. Zhong, and Y. Fu, "Residual non-local attention networks for image restoration," *arXiv preprint arXiv:1903.10082*, 2019.

[52] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[53] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga *et al.*, "Pytorch: An imperative style, high-performance deep learning library," in *Advances in neural information processing systems*, 2019, pp. 8026–8037.

[54] Y. Yang and S. Newsam, "Bag-of-visual-words and spatial extensions for land-use classification," in *Proceedings of the 18th SIGSPATIAL international conference on advances in geographic information systems*, 2010, pp. 270–279.

[55] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE transactions on image processing*, vol. 13, no. 4, pp. 600–612, 2004.

[56] G. Cheng, J. Han, and X. Lu, "Remote sensing image scene classification: Benchmark and state of the art," *Proceedings of the IEEE*, vol. 105, no. 10, pp. 1865–1883, 2017.

[57] J. Yang, J. Wright, T. S. Huang, and Y. Ma, "Image super-resolution via sparse representation," *IEEE transactions on image processing*, vol. 19, no. 11, pp. 2861–2873, 2010.

[58] C. Dong, C. C. Loy, K. He, and X. Tang, "Image super-resolution using deep convolutional networks," *IEEE transactions on pattern analysis and machine intelligence*, vol. 38, no. 2, pp. 295–307, 2015.

[59] C. Dong, C. C. Loy, and X. Tang, "Accelerating the super-resolution convolutional neural network," in *European conference on computer vision*. Springer, 2016, pp. 391–407.

[60] D. Dai, Y. Wang, Y. Chen, and L. Van Gool, "Is image super-resolution helpful for other vision tasks?" in *2016 IEEE Winter Conference on Applications of Computer Vision (WACV)*. IEEE, 2016, pp. 1–9.

[61] J. Noh, W. Bae, W. Lee, J. Seo, and G. Kim, "Better to follow, follow to be better: towards precise supervision of feature super-resolution for small object detection," in *Proceedings of the IEEE International Conference on Computer Vision*, 2019, pp. 9725–9734.

[62] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 770–778.

Sen Lei received his B.S. degree from the Image Processing Center, School of Astronautics, Beihang University in 2015. He is currently working toward his doctorate degree in the Image Processing Center, School of Astronautics, Beihang University. His research interests include deep learning and image super-resolution.

Zhenwei Shi (M'13) received his Ph.D. degree in mathematics from Dalian University of Technology, Dalian, China, in 2005. He was a Postdoctoral Researcher in the Department of Automation, Tsinghua University, Beijing, China, from 2005 to 2007. He was Visiting Scholar in the Department of Electrical Engineering and Computer Science, Northwestern University, U.S.A., from 2013 to 2014. He is currently a professor and the dean of the Image Processing Center, School of Astronautics, Beihang University. His current research interests include remote sensing image processing and analysis, computer vision, pattern recognition, and machine learning.

Dr. Shi serves as an Associate Editor for the *Infrared Physics and Technology*. He has authored or co-authored over 100 scientific papers in refereed journals and proceedings, including the IEEE Transactions on Pattern Analysis and Machine Intelligence, the IEEE Transactions on Image Processing, the IEEE Transactions on Geoscience and Remote Sensing, the IEEE Geoscience and Remote Sensing Letters and the IEEE Conference on Computer Vision and Pattern Recognition. His personal website is http://levir.buaa.edu.cn/.