TSINGHUA SCIENCE AND TECHNOLOGY
ISSN 1007-0214 05/14 pp475–485
DOI: 10.26599/TST.2022.9010018
Volume 28, Number 3, June 2023

RFCNet: Remote Sensing Image Super-Resolution Using Residual Feature Calibration Network

Yuan Xue, Liangliang Li, Zheyuan Wang, Chenchen Jiang, Minqin Liu, Jiawen Wang, Kaipeng Sun, and Hongbing Ma*

Abstract: In the field of single remote sensing image Super-Resolution (SR), deep Convolutional Neural Networks (CNNs) have achieved top performance. To further enhance convolutional module performance in processing remote sensing images, we construct an efficient residual feature calibration block to generate expressive features. After harvesting residual features, we first divide them into two parts along the channel dimension. One part flows to the Self-Calibrated Convolution (SCC) to be further refined, and the other part is rescaled by the proposed Two-Path Channel Attention (TPCA) mechanism. SCC corrects local features according to their expressions under the deep receptive field, so that the features can be refined without increasing the number of calculations. The proposed TPCA uses the means and variances of feature maps to obtain accurate channel attention vectors. Moreover, a region-level nonlocal operation is introduced to capture long-distance spatial contextual information by exploring pixel dependencies at the region level. Extensive experiments demonstrate that the proposed residual feature calibration network is superior to other SR methods in terms of quantitative metrics and visual quality.

Key words: Convolutional Neural Network (CNN); remote sensing image; Super-Resolution (SR); attention mechanism

1 Introduction

Single-image Super-Resolution (SR) is a popular issue aiming to restore High-Resolution (HR) images from the corresponding Low-Resolution (LR) images. As we all know, images with higher spatial resolution can provide more detailed information, so SR can serve various remote sensing tasks, e.g., environmental monitoring[1], resource exploration[2], and target detection[3]. SR is a serious ill-posed problem, because LR images can be mapped to multiple HR images. To generate accurate HR images, several SR methods have been proposed, including interpolation[4], reconstruction[5], and learning-based methods[6–12]. Due to the lack of consideration of the edge characteristics of images, interpolation-based (e.g., nearest, bilinear, and bicubic) methods fail to restore HR images with high quality. Flexible reconstruction-based methods work by exploiting powerful image priors, such as nonlocal similarity and sparsity priors. Although such reconstruction-based methods can restore accurate HR images, they are time-consuming and suffer severe performance degradation in the absence of prior information. Deep learning development has greatly promoted progress in various fields. Due to powerful nonlinear fitting

• Yuan Xue, Zheyuan Wang, Chenchen Jiang, and Minqin Liu are with the College of Information Science and Engineering, Xinjiang University, Urumqi 830046, China. E-mail: 18309886088@163.com; wzy_696@163.com; 482049605@qq.com; liumqxz@163.com.
• Liangliang Li and Hongbing Ma are with the Department of Electronic Engineering, Tsinghua University, Beijing 100084, China. E-mail: leeliangliang@163.com; hbma@tsinghua.edu.cn.
• Jiawen Wang and Kaipeng Sun are with Shanghai Institute of Satellite Engineering, Shanghai 201109, China. E-mail: 313474093@qq.com; sunkaipeng@509.sast.casc.
* To whom correspondence should be addressed.
Manuscript received: 2022-01-13; revised: 2022-04-05; accepted: 2022-05-24

ability and end-to-end training paradigm, Convolutional Neural Networks (CNNs) can handle SR tasks and achieve remarkable performance. Although several methods[8, 13, 14] have achieved prominent performance in remote sensing SR tasks, some limitations still exist: (1) Most CNNs enlarge receptive fields by deepening networks. Accordingly, model complexity and memory consumption also increase, hindering the application in practice needs; (2) Most existing CNN-based SR methods ignore the calibration of features that is not conducive to improving the quality of restored remote sensing images.

To address these issues, we propose an efficient Residual Feature Calibration Network (RFCNet) for accurate remote sensing image SR. Figure 1 shows the structure of the proposed network, where multiple Nonlocal Share-source Frameworks (NSFs) are grouped together to extract deep features. Specifically, one Region-Level NonLocal (RLNL) operation is first used to capture long-distance spatial contextual information, and nonlocal features with large receptive fields can be obtained at extremely small costs. Generated nonlocal features with rich information are then transmitted to latter Residual Feature Calibration Blocks (RFCBs) by Share-Source Skip Connections (S3Cs). To make RFCBs obtain valuable information from nonlocal features, a learnable weight vector is used to multiply nonlocal features element-wise. The weight vector can adaptively scale nonlocal features at the corresponding dimension before it is passed forward. As for RFCBs, after obtaining residual features, we split them into two parts along the channel axis, one part is calibrated pixel-wise by Self-Calibrated Convolution (SCC), and the other part is rescaled channel-wise by the proposed Two-Path Channel Attention (TPCA). Then, the two parts of features are concatenated together to add to the features on the identity branch.

The contributions can be summarized as follows:

(1) A notable RFCNet is proposed for efficient remote sensing image SR, where the proposed NSF skillfully combines RLNL, S3C, and RFCB. Features with large receptive fields extracted through nonlocal operation are transmitted to latter RFCBs by S3Cs. Extensive experiments on several remote sensing datasets show that the proposed RFCNet can achieve better SR performance than other CNN-based SR methods in terms of quantity and visual quality.

(2) An effective RFCB is proposed to improve residual module performance. SCC is used to calibrate one part of residual features, and TPCA is employed to rescale the other parts of features. The proposed RFCB greatly increases the model fitting ability.

(3) A flexible TPCA is proposed to obtain an accurate channel attention descriptor. Here, the two branches are used to calculate the means and variances of feature maps, thereby greatly increasing the ability to characterize features of complex remote sensing images.

In this paper, Section 2 introduces some representative works from two aspects: remote sensing and general SR. Section 3 first introduces the basic network architecture of our method, and then describes NSF and RFCB in detail, and RFCB further includes SCC and TPCA. Section 4 first introduces the used datasets and implementation details of methods. Next, the model analysis part shows the effectiveness of our proposed sub-modules, and finally we compare the performance indicators and visualization results of our method with other excellent methods. In conclusion part, we generally describe this article.

2 Related Work

SR, as a classical low-level task, plays an important role in the computer vision community. In this section, we focus on recent CNNs for remote sensing image and general image SR.

2.1 CNNs for remote sensing image SR

Remote sensing images with high spatial resolution are conducive to other remote sensing tasks[1–3, 13]. Compared with improving hardware, the SR technique

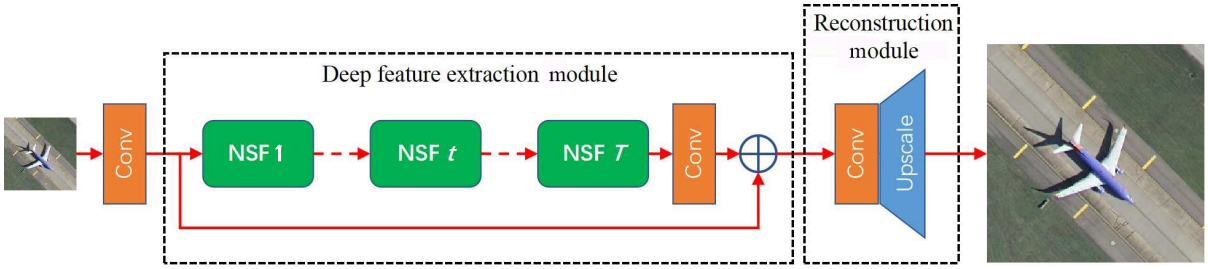

Fig. 1 Architecture of the proposed RFCNet, where ⊕ denotes element-wise addition.

is a cheaper way, which improves image quality without changing the original hardware.

Recently, numerous CNN-based architectures have been proposed to obtain high-quantity remote sensing images. Affected by the success of CNNs in natural scenes, a Local-Global Combined Network (LGCNet)[14] is proposed to learn multilevel representations. Pan et al.[8] constructed the residual dense back-projection network and explored the SR of remote sensing images with median- and large-scale factors. To further explain restoration, Zhang et al.[15] first introduced a mixed high-order attention module to low-level vision tasks. Sustika et al.[16] used a residual dense network as a generative adversarial network generator and achieved high performance. Zhang et al.[17] proposed a novel Cycle Convolutional Neural Network (Cycle-CNN), which can be trained with unpaired data. To release additional power when dealing with large-scale SR tasks, Dong et al.[18] developed a dense sampling SR network to explore the large-scale SR reconstruction of remote sensing images.

2.2 CNNs for general image SR

Deep learning development has greatly promoted progress in various fields. CNNs can handle SR tasks and achieve remarkable performance because of their powerful nonlinear fitting ability and end-to-end training paradigm. Super-Resolution Convolutional Neural Network (SRCNN)[10] is the pioneering work for image SR, which outperforms traditional methods by a large margin. A huge room for improvement exists, as SRCNN only has three convolutional layers. Later, numerous CNN-based SR methods have been proposed to learn the mapping function from LR input to the corresponding HR output. To accelerate the reasoning speed, Shi et al.[11] used a sub-pixel convolution layer to upscale the final LR feature maps. The network can extract features from LR images directly and solve the time cost of processing images at HR. Later, Kim et al.[12] designed a deep VDSR with 20 layers, which eased training difficulty by using residual learning. Lim et al.[19] thought that the batch normalization layer is not conducive to the SR task. They stacked a long string of modified residual blocks and achieved state-of-the-art performance. Zhang et al.[20] incorporated a dense block into the residual block to make full use of the hierarchical features of LR images. Zhang et al.[21] later introduced the channel attention mechanism to adaptively rescale features by exploring the correlation among channels. Qiu et al.[22] further proposed a second-order attention network to adaptively rescale features by considering feature statistics higher than first-order RCAN. Liu et al.[23] proposed a Residual Feature Aggregation Network (RFANet) to produce representative features to fully use hierarchical features on residual branches.

3 Methodology

3.1 Basic network architecture for image SR

As illustrated in Fig. 1, the proposed RFCNet comprises three modules: Shallow feature extraction Module (SM), Deep feature extraction Module (DM), and Reconstruction Module (RM). For SM, a 3×3 convolution is used to obtain coarse features F_0. Given input I_{LR}, the coarse features can be obtained by

$$F_0 = H_{SM}(I_{LR}) \tag{1}$$

where H_{SM} denotes the function of SM. Then, the extracted coarse features are fed to DM, which further contains a series of stacked NSFs. The output of the t-th NSF can be obtained by

$$F_t = H_{NSF}^t\left(H_{NSF}^{t-1}\left(\cdots H_{NSF}^1(F_0)\cdots\right)\right),\ t \in 1,\ldots,T \tag{2}$$

where H_{NSF}^t represents the function of the t-th NSF. F_t denotes the output features of the t-th NSF. Finally, a 3×3 convolution is used to refine the features generated by the last NSF. Global residual learning is also used to facilitate information flow. This process can be formulated as

$$F_{DF} = F_0 + H_{conv}^{3\times3}(F_t) \tag{3}$$

where $H_{conv}^{3\times3}$ represents a 3×3 convolutional layer, and F_{DF} represents deep features. Then, RM is used to generate the SR image. Here, RM contains an upscale layer and a 3×3 convolutional layer. For the upscale layer, ESPCN[11] is chosen to enlarge the spatial size and achieve an improved trade-off between computational burden and performance, the super-resolved image with the target spatial resolution is as follows:

$$I_{SR} = H_R(F_{DF}) = H_{RFCNet}\left(H_{conv}^{3\times3}(F_{DF})\right) \tag{4}$$

where $H_R(\cdot)$ and $H_{RFCNet}(\cdot)$ denote RM and the proposed RFCNet, respectively.

Following previous studies[21, 22], RFCNet is then optimized with L_1 loss function. Given a training dataset with N LR image patches (I_{LR}) and the corresponding HR image patches (I_{HR}), the loss function can be formulated as

$$L(\theta) = \frac{1}{N}\sum_{i=1}^{N}\left|\left|H_{RFCNet}\left(I_{LR}^i\right) - I_{HR}^i\right|\right|_1 \tag{5}$$

where θ denotes the parameter set of the proposed RFCNet.

3.2 NSF

The proposed NSF (Fig. 2) comprises one RLNL module, and several S3Cs and RFCBs. Dai et al.[24] verified that performing RLNL operations is better than conducting global-level nonlocal operations. Here, we first utilize the RLNL[24] operation to capture long-distance dependencies by the divided image patches,

$$Y_0 = H_{RLNL}(X) \tag{6}$$

where $H_{RLNL}(\cdot)$ denotes the function of RLNL. X and Y_0 stand for the input of NSF and output features of RLNL, respectively. To make full use of nonlocal features with long-distance dependencies, we pass them to latter RFCBs through S3C. The output of the k-th RFCB can be obtained by

$$Y_k = H^k_{RFCB}(Y_{k-1}) + V_{S3C} \times Y_0, k \in 1, \ldots, K \tag{7}$$

where $H^k_{RFCB}(\cdot)$ stands for the function of the k-th RFCB. V_{S3C} denotes a learnable vector used to scale nonlocal features. Then, the output of NSF is obtained by

$$F_t = H^{3\times3}_{conv}(Y_K) + X \tag{8}$$

where $H^{3\times3}_{conv}$ stands for a 3×3 convolutional layer.

3.3 RFCB

As illustrated in Fig. 3, assuming the input of RFCB is Y_k, the input features are first processed by the convolutional block function, i.e., Conv $3 \times 3 \rightarrow$ ReLu $\rightarrow$ Conv 3×3. Then, the obtained features are split into x_1 and x_2 along the channel axis. For the two branches, SCC and TPCA are used to refine and rescale the two features, respectively.

For SCC, the features are first downscaled to obtain a huge receptive field by the *AdaptiveMaxPool2D* function with kernel size 7 and stride 3. After adjusting

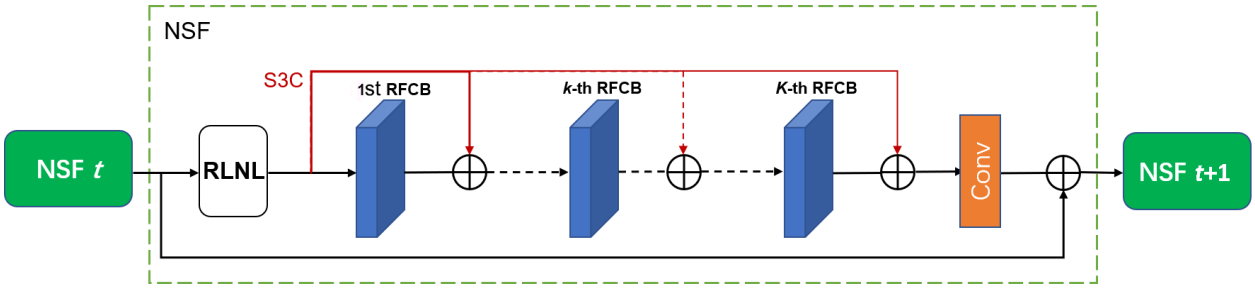

Fig. 2 Architecture of NSF.

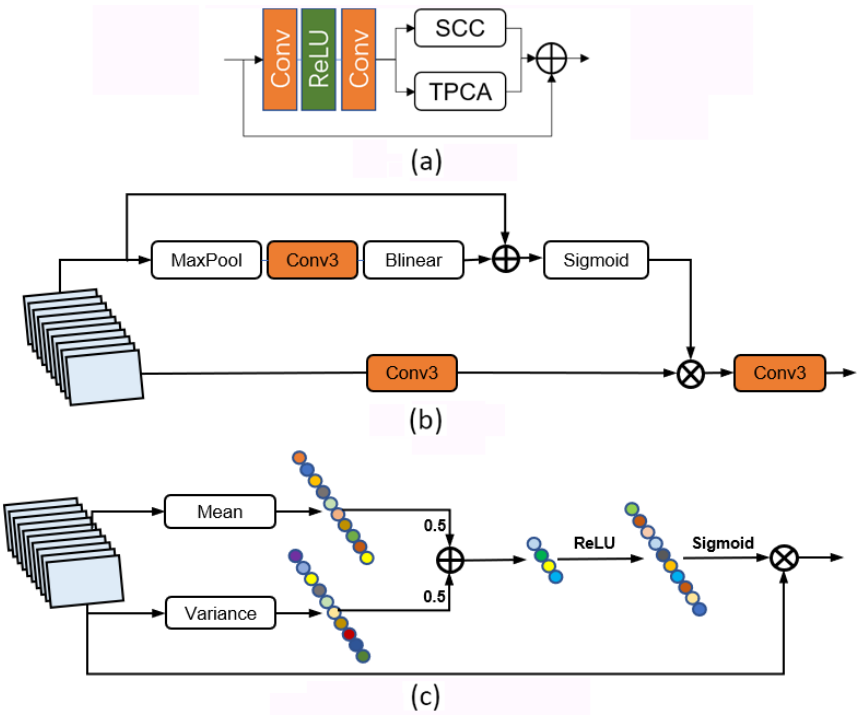

Fig. 3 Architecture of RFCB, which includes SCC and TPCA mechanisms. (a) RFCB, (b) SCC, and (c) TPCA mechanism.

by a 3×3 convolution layer, *bilinear* interpolation function is exploited to upscale the features to input size. Finally, a *sigmoid* function is adopted to obtain the final spatial statistics s_p. This process can be described as

$$w = \delta(H_{\uparrow}(H_{conv}^{3\times3}(H_{\downarrow}(x_1)))) \tag{9}$$

$$x_1^{out} = H_{conv}^{3\times3}(w \times H_{conv}^{3\times3}(x_1)) \tag{10}$$

where w denotes attention matrix, which is utilized to scale the refined feature x_1. $\delta(\cdot)$ denotes the $sigmoid$ function. $H_{\uparrow}(\cdot)$ denotes the upsampling operation through $bilinear$ interpolation, and $H_{\downarrow}(\cdot)$ denotes the downsampling operation using the $AdaptiveMaxPool2D$ function.

The structure of a remote sensing image is complex, and obtaining the channel descriptor only through global average pooling is inaccurate. We consider that the variance can reflect the information richness of a feature map, and add a branch to obtain the channel descriptive vector according to the variance. Here, we have

$$s_{avg} = H_{mean}(x_2^c), \ \ c \in 1, \ldots, 32 \tag{11}$$

$$s_{var} = H_{mean}(x_2^c - s_{avg}), \ \ c \in 1, \ldots, 32 \tag{12}$$

where s_{avg} and s_{var} stand for the channel descriptors obtained by average and variance, respectively. $H_{mean}(\cdot)$ indicates the operation of calculating the mean value along the channel axis. To make full use of the two descriptors, we first use two convolutions to reduce the dimension of the two descriptors and fuse them in an additive way. Then, a 1×1 convolutional layer is used to increase the dimension of the fused vectors. Finally, a $sigmoid$ function is used to generate the final channel descriptor. Here, we have

$$s_{aggregation} = H_{conv1}^{f=4}(s_{avg}) + H_{conv1}^{f=4}(s_{var}) \tag{13}$$

$$s = Sigmoid(H_{conv1}^{f=32}(ReLU(s_{aggregation}))) \tag{14}$$

where $H_{conv1}^{f=4}$ and $H_{conv1}^{f=32}$ denote the 1×1 convolution with 4 and 32 kernels, respectively. The final channel descriptor s can be flexibly used as the expression of a feature map at the spatial level. We use it to rescale split features x_2,

$$x_2^{out} = x_2 \times s \tag{15}$$

After obtaining the calibrated x_1^{out} and rescaled x_2^{out}, we concatenate the two features on the corresponding branches to obtain the output of RFCB,

$$Y_{k+1} = Concat(x_1^{out}, \ x_2^{out}) + Y_k \tag{16}$$

4 Experiment

4.1 Datasets

The WHU-RS19[25] dataset is a remote sensing image dataset containing 19 types of scene images with a size of 600×600 pixel, about 50 images in each type, with a total of 1005 images. Considering that the changes in lighting-, scale-, resolution-, and viewpoint- dependent appearances in some categories are more complex than those in other datasets, images from the WHU-RS19 dataset are selected to make the training dataset. Specifically, 45 images are randomly selected from each category, a total of 855 images, as the training dataset. Meanwhile, one image from each category is selected as the validation dataset named RS19-V, and the remaining images are used as the first test dataset named RS19-T. The RSSCN7[26] dataset contains seven typical scene categories. The pixel size of each image in the dataset is 400×400 pixel. Images with a size of 400×400 pixel come from different seasons and are sampled on the basis of four different scales. Ten images are chosen from each category as the second test dataset named RSSCN7-T to prove the power of the proposed RFCNet method.

4.2 Implementation details

To obtain LR training images, HR images are downscaled with scaling factors (×2, ×3, and ×4) through $bicubic$ interpolation. The training images are randomly rotated by 90°, 180°, and 270°, and images are flipped horizontally to make full use of the training data. Moreover, 10 images are randomly selected from each category of the remaining images, and then 300 images are obtained as the first test dataset named AID-T. In each training batch, the input of RFCNet is 16 LR RGB-channel patches with a size of 48× 48 pixel. The learning rate is set to 1×10^{-4} and halved at every 100 epochs. Specifically, the number of NSF is set to 10, and one nonlocal module and eight RFCB exist in each NSF. The network has about 160 convolutional layers with a parameter of 8.9×10^6. When evaluating the SR results, the RGB images are first transformed into YCbCr space and then the average PSNR and structural similarity are calculated only on the Y channel of the YCbCr space. The pixel range of input images is scaled to [0,1] to ensure numeric stability. The model in this study is optimized by an ADAM optimizer with $\beta_1 = 0.9$ and $\beta_2 = 0.999$. All experiments are implemented using the PyTorch[27] framework.

4.3 Model analysis

As displyed in Fig. 1, RLNL, S3C, and RFCBs are combined into a powerful NSF. To prove the effectiveness of some key modules, we compare NSF with its variants, and test them on the RS19-V dataset.

Here, the long and short skip connections have been proven to be efficient in a previous study[20]; hence they are used for all cases. From Table 1, we can clearly see that the base module using only residual blocks achieves a result of 29.01 dB, and the addition of S3C, RLNL, RFCB, and TPCA improves the performance by 0.06 dB, 0.01 dB, 0.06 dB, and 0.03 dB, respectively. These modules are thus beneficial to the reconstruction of remote sensing images. When SCC or TPCA is used alone, the other branch in RFCB does not perform any processing (Fig. 3a). Compared with the use of a single module, the combined version of SCC and TPCA achieves better performance. Specifically, RFCB is regarded as a whole in the experiments. Moreover, RLNL based on S3C has achieved a leap in performance, because nonlocal features are important for the feature refinement of the latter modules. If RLNL is not combined, then S3C only transmits input features, which have little performance gain. Finally, the ingenious combination achieves 29.19 dB.

4.4 Comparisons with state-of-the-art methods

The proposed RFCNet is compared with several state-of-the-art SR methods, including SRCNN[10], FSRCNN[28], VDSR[12], IDN[29], RCAN[21], RFANet[23], and RFDN[30] to test the effectiveness of the proposed RFCNet. Here, RFDN is a lightweight and efficient network.The purpose of comparison with RFDN is to show the advantages of our method and the gap between a lightweight method and our method. Table 2 shows all the quantitative comparisons for ×2, ×3, and ×4 SR. The RFCNet method achieves the best performance among the previously proposed methods with ×2 and ×3 magnifications, even better than the top RCAN and RFANet. This fully demonstrates the high performance of the network. On scale ×4, the RFCNet method obtaines a better PSNR value than other methods in the RS19 test dataset. However, RCAN achieves better performance in the RSSCN7 test dataset, likely because the residual in residual structure makes RCAN to be designed deeper than usual. Accordingly, the parameters of RCAN (about 15.6×10^6) are almost twice those of the proposed RFCNet (8×10^6) (see Table 3). Figure 4 shows visual comparisons on scales ×3 and ×4. The details are richer than those of SRCNN, FSRCNN, and VDSR methods.

Table 1 Effects of different modules on PSNR. We report the best Peak Signal-to-Noise Ratio (PSNR) values on RS19-V (×2) in 5×10^4 iterations. (dB)

Module	PSNR
Base	29.01
RLNL	29.07
S3C	29.02
SCC	29.07
TPCA	29.04
SCC+TPCA	29.10
RLNL+S3C	29.11
RLNL+SCC+TPCA	29.14
S3C+SCC+TPCA	29.09
RLNL+S3C+SCC+TPCA	29.19

Table 2 Quantitative comparison with several state-of-the-art methods based on ×2, ×3, and ×4 SR with bicubic degradation model. The best results are in red. The second bests result are in green.

Method	Scale	PSNR (dB)/SSIM	
		RS19-T	RSSCN7-T
Bicubic[10]	×2	34.70/0.9574	31.72/0.9170
SRCNN[10]		36.43/0.9714	32.82/0.9343
FSRCNN[28]		36.54/0.9723	32.90/0.9351
VDSR[12]		37.04/0.9746	33.22/0.9382
IDN[29]		37.32/0.9759	33.46/0.9407
RCAN[21]		37.39/0.9763	33.51/0.9411
RFANet[23]		37.43/0.9764	33.53/0.9414
RFDN[30]		37.26/0.9756	33.40/0.9403
RFCNet (ours)		37.47/0.9768	33.59/0.9423
Bicubic[10]	×3	30.89/0.9013	29.29/0.8587
SRCNN[10]		31.96/0.9239	29.98/0.8799
FSRCNN[28]		32.11/0.9258	30.06/0.8156
VDSR[12]		32.40/0.9297	30.25/0.8852
IDN[29]		32.75/0.9347	30.51/0.8905
RCAN[21]		32.84/0.9363	30.59/0.8921
RFANet[23]		32.83/0.9356	30.53/0.8911
RFDN[30]		32.73/0.9345	30.50/0.8904
RFCNet (ours)		32.94/0.9370	30.61/0.8924
Bicubic[10]	×4	28.91/0.8478	28.05/0.8128
SRCNN[10]		29.66/0.8731	28.58/0.8340
FSRCNN[28]		29.80/0.8772	28.67/0.8375
VDSR[12]		30.12/0.8836	28.83/0.8425
IDN[29]		30.44/0.8914	29.05/0.8488
RCAN[21]		30.56/ 0.8952	29.18/0.8524
RFANet[23]		30.48/0.8921	29.12/0.8502
RFDN[30]		30.39/0.8903	29.05/0.8485
RFCNet (ours)		30.62/0.8950	29.15/0.8516

4.5 Comparisons with remote sensing SR methods

The aforementioned experiments show the effectiveness of the proposed RFCNet. We now compare the proposed RFCNet with LGCNet[14] and MHAN[15] to further demonstrate the performance of the proposed network, which is specially designed for remote sensing SR.

Table 3 Model parameter and FLOP of some compared methods (e.g., SRCNN, FSRCNN, VDSR, LGCNet, IDN, RCAN, RFANet, RFDN, MHAN, and RFCNet). The FLOPs of models with ×3 enlargement are calculated corresponding to a 1280 × 720 pixel HR image.

Method	SRCNN	FSRCNN	VDSR	LGCNet	IDN	RCAN	RFANet	RFDN	MHAN	RFCNet
Number of parameters	2.00×10^4	2.50×10^4	6.82×10^5	1.93×10^5	5.91×10^5	1.56×10^7	1.11×10^7	5.73×10^5	1.14×10^7	8.90×10^6
FLOPS	1.85×10^{10}	1.36×10^{10}	7.00×10^{10}	1.78×10^{11}	1.06×10^{11}	1.59×10^{12}	1.03×10^{12}	5.60×10^{10}	1.20×10^{12}	8.45×10^{11}

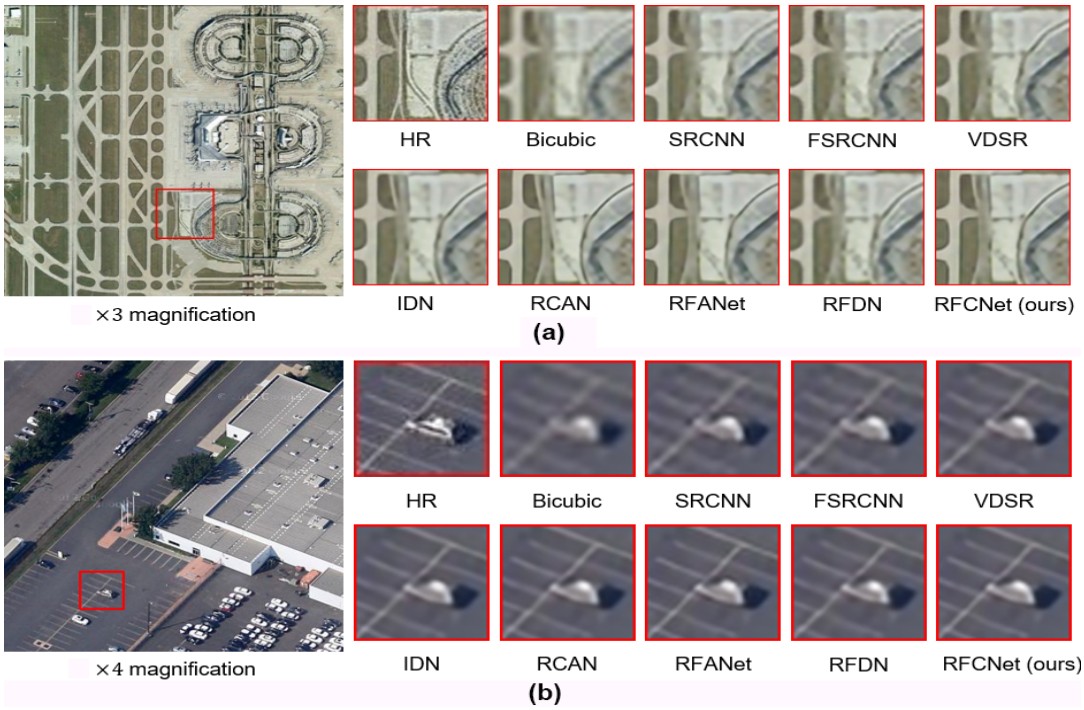

Fig. 4 Visual comparison of some representative SR methods and our model on ×3 (a) and ×4 (b) factors.

As presented in Table 4, it is easy to find that the performance of the proposed RFCNet far exceeds that of LGCNet. Improved results are also achieved for the recently proposed MHAN. All images in all categories of RSSCN7 are tested on scale factor ×2, ×3, and ×4 to prove the generalization ability of the proposed network (Table 5). Our RFCNet achieves the best results on all classes among SRCNN, FSRCNN, VDSR, and RFANet. In addition, two zoomed-in remote sensing images are displayed, where four local details of several methods are given. The RFCNet recover many details and textures from LR images. As illustrated in Fig. 5, the original LR images are seriously blurred. The LGCNet and traditional bicubic only restore some edge textures, and the restored images are still blurred. The proposed RFCNet and the recently proposed MHAN can eliminate artificial artifacts well. In addition, from the enlarged letters and other details, the reconstructed image contours are better than MHAN.

Table 4 Comparison of different remote sensing SR methods. The PSNR and SSIM values are evaluated on RS19-T and RSSCN7-T dataset with scale factors ×2, ×3, and ×4. The best results are in red. The second best results are in green.

Method	Scale	PSNR (dB)/SSIM	
		RS19-T	RSSCN7-T
Bicubic[10]	×2	34.70/0.9574	31.72/0.9170
LGCNet[14]		36.94/0.9741	33.18/0.9379
MHAN[15]		37.32/0.9761	33.49/0.9410
RFCNet (ours)		37.47/0.9768	33.59/0.9423
Bicubic[10]	×3	30.89/0.9013	29.29/0.8587
LGCNet[14]		32.39/0.9299	30.27/0.8858
MHAN[15]		32.74/0.9364	30.57/0.8923
RFCNet (ours)		32.94/0.9370	30.61/0.8924
Bicubic[10]	×4	28.91/0.8478	28.05/0.8128
LGCNet[14]		30.01/0.8820	28.79/0.8416
MHAN[15]		30.55/0.8942	29.15/0.8513
RFCNet (ours)		30.62/0.8950	29.15/0.8516

5 Conclusion

In this study, the RFCNet is proposed for effective single remote sensing image SR. Our RFCNet comprises several NSFs, where the long and short skip connections

Table 5 Mean PSNR (dB)/SSIM values of each class of RSSCN7 for Bicubic, SRCNN, FSRCNN, VDSR, RFANet, and RFCNet. The results are evaluated on ×2, ×3, and ×4 enlargements.

Class	Scale	Bicubic	SRCNN	FSRCNN	VDSR	RFANet	RFCNet
Grass	×2	37.55/0.9664	38.74/0.9741	38.82/0.9744	39.05/0.9753	39.22/0.9760	39.26/0.9762
Field		35.57/0.9231	36.41/0.9320	36.44/0.9321	36.59/0.9325	36.83/0.9345	36.87/0.9353
Industry		28.78/0.9137	30.31/0.9379	30.45/0.9393	31.04/0.9455	31.54/0.9503	31.64/0.9511
RiverLake		33.89/0.9524	35.03/0.9633	35.10/0.9636	35.39/0.9653	35.62/0.9666	35.65/0.9669
Forest		31.69/0.9060	32.65/0.9272	32.69/0.9274	32.76/0.9284	32.88/0.9302	32.90/0.9307
Resident		28.13/0.9093	29.56/0.9342	29.70/0.9357	30.17/0.9410	30.55/0.9449	30.62/0.9456
Parking		27.87/0.8943	29.03/0.9194	29.14/0.9209	29.61/0.9283	30.07/0.9348	30.15/0.9361
Average		31.93/0.9236	33.10/0.9411	33.19/0.9419	33.52/0.9452	33.82/0.9482	33.87/0.9487
Grass	×3	34.70/0.9363	35.35/0.9448	35.43/0.9452	35.63/0.9466	35.82/0.9482	35.88/0.9486
Field		33.48/0.8918	33.97/0.8997	34.02/0.9002	34.18/0.9017	34.37/0.9036	34.42/0.9042
Industry		25.96/0.8363	27.03/0.8703	27.16/0.8732	27.47/0.8799	27.99/0.8912	28.15/0.8940
RiverLake		31.39/0.9154	32.12/0.9286	32.20/0.9294	32.37/0.9315	32.61/0.9343	32.67/0.9355
Forest		29.13/0.8277	29.61/0.8511	29.64/0.8517	29.68/0.8530	29.78/0.8557	29.81/0.8567
Resident		25.35/0.8263	26.30/0.8621	26.44/0.8651	26.69/0.8710	27.06/0.8800	27.17/0.8826
Parking		25.46/0.8148	26.23/0.8461	26.32/0.8487	26.55/0.8553	26.95/0.8676	27.07/0.8706
Average		29.35/0.8641	30.09/0.8861	30.17/0.8876	30.37/0.8913	30.65/0.8972	30.74/0.8987
Grass	×4	33.29/0.9139	33.77/0.9215	33.88/0.9232	34.07/0.9252	34.24/0.9272	34.30/0.9279
Field		32.49/0.8723	32.86/0.8787	32.93/0.8804	33.13/0.8830	33.32/0.8855	33.39/0.8864
Industry		24.45/0.7702	25.27/0.8061	25.41/0.8118	25.69/0.8212	26.17/0.8361	26.29/0.8391
RiverLake		30.06/0.8859	30.64/0.8993	30.73/0.9014	30.91/0.9043	31.14/0.9082	31.16/0.9093
Forest		27.88/0.7686	28.22/0.7896	28.26/0.7924	28.29/0.7937	28.38/0.7973	28.40/0.7993
Resident		23.78/0.7494	24.50/0.7888	24.64/0.7947	24.85/0.8028	25.19/0.8157	25.25/0.8181
Parking		24.18/0.7514	24.77/0.7826	24.87/0.7879	25.08/0.7964	25.42/0.8103	25.48/0.8132
Average		28.02/0.8160	28.57/0.8381	28.68/0.8417	28.86/0.8467	29.12/0.8543	29.18/0.8563

Fig. 5 Visual comparison of remote sensing SR methods and the proposed model on ×4 factor. Two red patches of each image are enlarged for better comparison.

facilitate information flow. In NSF, RLNL is used to generate nonlocal features with large receptive fields. After being scaled by a learnable vector, nonlocal features are transmitted to latter RFCBs through S3Cs. Hence, each RFCB can obtain useful information from source features. As for RFCB, we first divide the features into two parts along the channel dimension. One part flows to SCC to be further refined, and the other part is rescaled by the proposed TPCA mechanism. SCC corrects the local features according to their expressions under the deep receptive field, so that the features can be refined without increasing the number of calculations. The proposed TPCA uses the means and variances of feature maps to obtain accurate channel attention vectors. Extensive experiments on the model demonstrate the effectiveness of the proposed RFCNet method for SR.

Acknowledgment

This work was supported by Shanghai Aerospace Science and Technology Innovation Fund (No. SAST2019-048) and the Cross-Media Intelligent Technology Project of Beijing National Research Center for Information Science and Technology (BNRist) (No. BNR2019TD01022).

References

[1] D. Moroni, G. Pieri, O. Salvetti, and M. Tampucci, Proactive marine information system for environmental monitoring, in *Proc. OCEANS 2015 -Genova*, Genova, Italy, 2015, pp. 1–5.

[2] R. Booysen, R. Gloaguen, S. Lorenz, R. Zimmermann, L. Andreani, and P. A. M. Nex, The potential of multi-sensor remote sensing mineral exploration: Examples from southern Africa, in *Proc. 2019 IEEE Int. Geoscience and Remote Sensing Symp.*, Yokohama, Japan, 2019, pp. 6027–6030.

[3] X. L. Wang, Y. Ban, H. M. Guo, and L. Hong, Deep learning model for target detection in remote sensing images fusing multilevel features, in *Proc. 2019 IEEE Int. Geoscience and Remote Sensing Symp.*, Yokohama, Japan, 2019, pp. 250–253.

[4] L. Zhang and X. L. Wu, An edge-guided image interpolation algorithm via directional filtering and data fusion, *IEEE Trans. Image Proc.*, vol. 15, no. 8, pp. 2226–2238, 2006.

[5] K. B. Zhang, X. B. Gao, D. C. Tao, and X. L. Li, Single image super-resolution with non-local means and steering kernel regression, *IEEE Trans. Image Proc.*, vol. 21, no. 11, pp. 4544–4556, 2012.

[6] R. Timofte, V. De, and L. Van Gool, Anchored neighborhood regression for fast example-based super-resolution, in *Proc. 2013 IEEE Int. Conf. on Computer Vision*, Sydney, Australia, 2013, pp. 1920–1927.

[7] T. Peleg and M. Elad, A statistical prediction model based on sparse representations for single image super-resolution, *IEEE Trans. Image Proc.*, vol. 23, no. 6, pp. 2569–2582, 2014.

[8] Z. X. Pan, W. Ma, J. Y. Guo, and B. Lei, Super-resolution of single remote sensing image based on residual dense backprojection networks, *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 10, pp. 7918–7933, 2019.

[9] Z. Wang, L. Li, Y. Xue, C. Jiang, J. Wang, K. Sun, and H. Ma, FeNet: Feature enhancement network for lightweight remote-sensing image super-resolution, *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–12, 2022.

[10] C. Dong, C. C. Loy, K. M. He, and X. O. Tang, Image super-resolution using deep convolutional networks, *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 2, pp. 295–307, 2016.

[11] W. Z. Shi, J. Caballero, F. Huszár, J. Totz, A. P. Aitken, R. Bishop, D. Rueckert, and Z. H. Wang, Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network, in *Proc. 2016 IEEE Conf. on Computer Vision and Pattern Recognition* (*CVPR*), Las Vegas, NV, USA, 2016, pp. 1874–1883.

[12] J. Kim, J. K. Lee, and K. M. Lee, Accurate image super-resolution using very deep convolutional networks, in *Proc. 2016 IEEE Conf. on Computer Vision and Pattern Recognition* (*CVPR*), Las Vegas, NV, USA, 2016, pp. 1646–1654.

[13] D. D. Guo, S. Y. Zhu, and J. A. Wei, Research on vehicle identification based on high resolution satellite remote sensing image, in *Proc. 2019 Int. Conf. on Intelligent Transportation, Big Data & Smart City* (*ICITBS*), Changsha, China, 2019, pp. 62–65.

[14] S. Lei, Z. W. Shi, and Z. X. Zou, Super-resolution for remote sensing images via local-global combined network, *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 8, pp. 1243–1247, 2017.

[15] D. Y. Zhang, J. Shao, X. Y. Li, and H. T. Shen, Remote sensing image super-resolution via mixed high-order attention network, *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 6, pp. 5183–5196, 2021.

[16] R. Sustika, A. B. Suksmono, D. Danudirdjo, and K. Wikantika, Generative adversarial network with residual dense generator for remote sensing image super resolution, in *Proc. 2020 Int. Conf. on Radar, Antenna, Microwave, Electronics, and Telecommunications* (*ICRAMET*), Tangerang, Indonesia, 2020, pp. 34–39.

[17] H. P. Zhang, P. R. Wang, and Z. G. Jiang, Nonpairwise-trained cycle convolutional neural network for single remote sensing image super-resolution, *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 5, pp. 4250–4261, 2021.

[18] X. Y. Dong, X. Sun, X. P. Jia, Z. H. Xi, L. R. Gao, and B. Zhang, Remote sensing image super-resolution using novel dense-sampling networks, *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 2, pp. 1618–1633, 2021.

[19] B. Lim, S. Son, H. Kim, S. Nah, and K. M. Lee, Enhanced deep residual networks for single image super-resolution, in

Proc. 2017 IEEE Conf. on Computer Vision and Pattern Recognition Workshops (*CVPRW*), Honolulu, HI, USA, 2017, pp. 1132–1140.

[20] Y. L. Zhang, Y. P. Tian, Y. Kong, B. N. Zhong, and Y. Fu, Residual dense network for image super-resolution, in *Proc. 2018 IEEE/CVF Conf. on Computer Vision and Pattern Recognition*, Salt Lake City, UT, USA, 2018, pp. 2472–2481.

[21] Y. L. Zhang, K. P. Li, K. Li, L. C. Wang, B. N. Zhong, and Y. Fu, Image super-resolution using very deep residual channel attention networks, in *Proc. 15th European Conf. on Computer Vision*, Munich, Germany, 2018, pp. 294–310.

[22] Y. Qiu, R.Wang, D.Tao, and J.Cheng, Embedded block residual network: A recursive restoration model for single-image super-resolution, in *Proceedings of Conference on Computer Vision and Pattern Recognition* (*CVPR*), Long Beach, CA, USA, 2019, pp. 4180–4189.

[23] J. Liu, W. J. Zhang, Y. T. Tang, J. Tang, and G. S. Wu, Residual feature aggregation network for image super-resolution, in *Proc. 2020 IEEE/CVF Conf. on Computer Vision and Pattern Recognition* (*CVPR*), Seattle, WA, USA, 2020, pp. 2356–2365.

[24] T. Dai, J. R. Cai, Y. B. Zhang, S. T. Xia, and L. Zhang, Second-order attention network for single image super-resolution, in *Proc. 2019 IEEE/CVF Conf. on Computer Vision and Pattern Recognition* (*CVPR*), Long Beach, CA, USA, 2019, pp. 11057–11066.

[25] D. X. Dai and W. Yang, Satellite image classification via two-layer sparse coding with biased image representation, *IEEE Geosci. Remote Sens. Lett.*, vol. 8, no. 1, pp. 173–176, 2011.

[26] Q. Zou, L. H. Ni, T. Zhang, and Q. Wang, Deep learning based feature selection for remote sensing scene classification, *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 11, pp. 2321–2325, 2015.

[27] A. Paszke, S. Gross, S. Chintala, G. Chanan, E. Yang, Z. DeVito, Z. M. Lin, A. Desmaison, L. Antiga, and A. Lerer, Automatic differentiation in PyTorch, presented at the 31st Conf. on Neural Information Processing Systems, Long Beach, CA, USA, 2017.

[28] C. Dong, C. C. Loy, and X. O. Tang, Accelerating the super-resolution convolutional neural network, in *Proc. 14th European Conf. on Computer Vision*, Amsterdam, The Netherlands, 2016, pp. 391–407.

[29] Z. Hui, X. M. Wang, and X. B. Gao, Fast and accurate single image super-resolution via information distillation network, in *Proc. 2018 IEEE/CVF Conf. on Computer Vision and Pattern Recognition*, Salt Lake City, UT, USA, 2018, pp. 723–731.

[30] J. Liu, J. Tang, and G. S. Wu, Residual feature distillation network for lightweight image super-resolution, in *Proc. European Conf. on Computer Vision*, Glasgow, UK, 2020, pp. 41–55.

Yuan Xue received the BEng degree in safety engineering from Shenyang Aerospace University, Shenyang, China in 2015. He is currently a master student in computer science and technology at Xinjiang University, Urumqi, China. His research interests include deep learning and remote sensing image super-resolution.

Liangliang Li received the PhD degree in communication and information systems from Jilin University, Changchun, China in 2019. He is currently a postdoctoral researcher at the Department of Electronic Engineering, Tsinghua University, Beijing, China. His research interests include image processing and pattern recognition.

Zheyuan Wang received the BEng degree in computer science and technology from Henan University of Technology, Zhengzhou, China in 2019. He is currently a master student in computer technology at Xinjiang University, Urumqi, China. His research interests include deep learning and image super-resolution.

Chenchen Jiang received the BEng degree in computer science and technology from Tianjin University of Technology, Tianjin, China in 2018. She is currently a master student in computer technology at Xinjiang University, Urumqi, China. Her research interests include deep learning and sematic segmentation.

Minqin Liu received the BEng degree in electronic information science and technology from Shandong University of Technology, Shandong, China in 2017. He is a master student at Xinjiang University, Urumqi, China. His research interests include machine learning and image processing.

Jiawen Wang received the MEng degree in navigation guidance and control from Nanjing University of Aeronautics and Astronautics, Nanjing, China in 2012. She is currently an engineer at Shanghai Institute of Satellite Engineering. Her research interest covers the overall design of optical remote sensing satellite and hyperspectral detection technology.

Kaipeng Sun received the BEng degree in aircraft design & engineering and the PhD degree in dynamics & control from Nanjing University of Aeronautics and Astronautics, Nanjing, China in 2009 and 2015, respectively. He is currently a senior engineer at Shanghai Institute of Satellite Engineering. His research interests include the overall design of optical remote sensing satellites and hyperspectral detection technology.

Hongbing Ma received the BEng degree from Hebei Normal University, Shijiazhuang, China in 1985; and the MEng and PhD degrees from Peking University, Beijing, China in 1996 and 1999, respectively. He is currently a professor at the Department of Electronic Engineering, Tsinghua University, Beijing, China. His research interests include image processing and pattern recognition.