MRA-IDN: A Lightweight Super-Resolution Framework of Remote Sensing Images Based on Multiscale Residual Attention Fusion Mechanism

Wujian Ye, Bili Lin, Junming Lao, Yijun Liu, and Zhenyi Lin

***Abstract*—The emergence of deep-learning technology has significantly improved the performance of super-resolution algorithms for a single remote sensing image; however, the number of deep-learning model parameters is large, which limits its real-time deployment. In addition, the reconstructed image quality still needs improvement. To deal with the aforementioned problems, a lightweight multiscale residual attention information distillation network is proposed in this article. It achieves high-quality and fast super-resolution processing of remote sensing images. First, a high- and low-frequency (HF and LF) separation reconstruction strategy is adopted that enables the network to improve the reconstructed details of HF components while keeping the number of model parameters low. Second, a novel multiscale residual attention information distillation group is designed as the key component to further extract richer regional features with different perceptual fields and HF information while reducing the number of network parameters. This is achieved by combining a multiscale residual information distillation block that consists of multiple residual convolutional sub-blocks and an HF channel aware attention block. Last, the experimental results show that, compared with existing mainstream methods, such as the MHAN, the number of model parameters can be reduced by 75%, and the edge details of the reconstructed images are richer and more complete. The corresponding peak signal-to-noise ratio and SSIM can reach 31.59 dB and 0.824, respectively, under the condition of a $\times 4$ magnification factor, and 27.39 dB and 0.668, respectively, under the condition of a $\times 8$ magnification factor.**

***Index Terms*—Deep neural network (DNN), high- and low-frequency (LF) reconstruction, high-frequency (HF) attention, multiscale features, remote sensing super-resolution.**

I. Introduction

REMOTE sensing imaging is one of the key technologies for acquiring maps of the surface and atmospheric information in the field of earth and environmental sciences. It is extensively used in agriculture and forestry monitoring, urban mapping, military reconnaissance, etc. Spatial resolution is the minimum distance between two adjacent features that can be recognized on a remote sensing image. The higher the spatial resolution, the better its ability to recognize objects. The remote sensing images contain objective and accurate representation of target and background information, and consequently, their spatial resolution directly influences the accuracy and adequacy of the acquired information. Improvement of spatial resolution at the hardware-level requires increasing the number of optical sensors by adding a large amount of photoreceptors and thus augmenting the number of pixels; however, such methods have certain limitations. On the one hand, the size and weight of the optical system of remote sensing satellites need to be increased, which considerably raise the research and development cost. On the other hand, these methods cannot be used to improve the existing remote sensing satellites operating in orbit. Therefore, the enhancement of the spatial resolution of remote sensing images at the software level has become a key research direction for scholars.

Single image super-resolution (SISR) aims to restore a high-resolution (HR) image by adding high-frequency (HF) details to its low-resolution (LR) images [1], and it can be inexpensive and easy to implement. It also plays a vital role in security, biomedicine, remote sensing, and other fields, as well as in recovering precious historical data. In the past decade, researchers proposed all sort of SISR algorithms, based on interpolation, reconstruction, or learning techniques [2]. The images generated using Lanczos resampling [3] and bicubic interpolation [4] tend to exhibit blurred edge and texture details. Although the reconstruction algorithms proposed in [5] and [6] can effectively enhance the image generation quality, they require a significant amount of expertise and a quite complex a priori knowledge. So the generated images cannot be used in practical applications in the field of remote sensing.

Deep learning has achieved remarkable breakthroughs in the fields of image classification, object detection, and recognition, starting from the AlexNet [7], to the ResNet [8], [9], DenseNet [10], SENet [11], MGSNet [12], LSCNet [13], SLA-NET [14], and MIFNet [15]. It has also led to a revolution in image super-resolution techniques, which are also known as learning methods. These algorithms mainly use a deep neural network (DNN) to obtain the nonlinear relationship between the HR and

Manuscript received 27 November 2023; revised 17 January 2024 and 9 February 2024; accepted 16 March 2024. Date of publication 27 March 2024; date of current version 10 April 2024. This work was supported in part by the Key Area R&D Program of Guangdong Province under Grant 2018B030338001, in part by the Basic and Applied Basic Research Project of Guangzhou Basic Research Program under Grant 202201010595, in part by the Guangdong Education Department, and in part by the Guangdong University of Technology under Grant 220413548. *(Corresponding author: Yijun Liu.)*

Wujian Ye and Yijun Liu are with the School of Integrated Circuits, Guangdong University of Technology, Guangzhou 510006, China (e-mail: yewjian@126.com; yjliu@gdut.edu.cn).

Bili Lin, Junming Lao, and Zhenyi Lin are with the School of Information Engineering, Guangdong University of Technology, Guangzhou 510006, China (e-mail: 749923259@qq.com; 1135804550@qq.com; 1282962751@qq.com).

Digital Object Identifier 10.1109/JSTARS.2024.3381653

the corresponding LR images through continuous learning and training. This can result in the effective recovery of HF and low-frequency (LF) details of images when the scale factor is large. This is the mainstream direction of the current research. Deep learning also has a feature extraction function, which can deeply extract features that cannot be extracted by traditional methods. Moreover, the loss function can effectively supervise the learning of the model and improve the accuracy and stability of super-resolution. In 2014, a deep convolutional neural network was first proposed for super-resolution (named SRCNN) [16], which achieved remarkable results compared with traditional algorithms. Subsequently, other models, such as the VDSR [17], deep residual codec network [18], and DSRN [19], were proposed, all of which achieved superior performance. However, all the aforementioned algorithms are SISR models linked to the mean square error and all the metrics derived from it, like peak signal-to-noise ratio (PSNR). The images generated by these models usually have high PSNR but low perceptual quality.

To generate images that match the sensory requirements of the human eye, the SRGAN super-resolution model [20] was proposed in 2017. It was based on a generative adversarial network (GAN), which achieved better visual results than the convolutional neural network (CNN)-based models; however, as the network continued to become deeper, the batch normalization layer would affect the final imaging quality, resulting in the appearance of image artifacts. To overcome the limitations of SRGAN, ESRGAN [21], ESRGAN+ [22], and other GAN-based models have been proposed successively. These methods can generate images with more realistic details and more clear texture information, which are aligned closely with the objective evaluation of human eyes. In 2021, the VSR-transformer [23], [24], based on the transformer model, was proposed for video super-resolution reconstruction, and it resulted in high improvement in accuracy. Chen et al. [25] proposed a multiscale deformable transformer for single hyperspectral image super-resolution, in which the deformable convolution-based transformer module further extracts global spatial spectral information from the local multiscale feature species of the previous stage, and achieves good SR performance. However, the number of model parameters was larger than that of GAN and CNN. Recently, because the visual effects generated by the diffusion model are superior to other generative models [26], [27], image super-resolution methods based on the diffusion model have also been proposed by many researchers. Liu et al. [28] proposed a generative diffusion model with complementary details for remote sensing image super-resolution, which uses the diffusion model as a generative model and the LR image as the conditional information to guide the image generation. Han et al. [29] proposed a remote sensing image super-resolution algorithm based on an efficient hybrid conditional diffusion model, which makes full use of the conditional features to predict the noise data distribution to effectively recover HR images from noise. Although all of them achieved good visual results, but the model parameters are huge and resource consumption is extremely high.

In order to fully utilize remote sensing image data and achieve the goal of super-resolution algorithms to be widely applied to real-world scenarios, Wu et al. [30] proposed a scale-aware dynamic network for continuous-scale super-resolution of remote sensing images. Mishra and Hadar [31] proposed an unsupervised network that is trained directly based on LR images. In addition, blind super-resolution is introduced to super-resolve LR images with unknown degradation, and Xiao et al. [32] proposed a self-supervised degradation-guided adaptive network, which can adapt to a variety of degradation distributions. These methods expand the application scenarios while improving the image perception quality.

The existing super-resolution models, such as FSRCNN [33], SRGAN, ESRGAN, RCAN [34], second-order attention network (SAN) [35], and TTSR [36], can achieve better results compared with traditional algorithms. However, these super-resolution models have a very large number of parameters, and their complexity is hundreds of times higher than that of classic techniques, leading to high costs and unsatisfactory performance in the field of remote sensing. Networks with a large number of parameters limit their deployment on low-computing-capacity devices, resulting in insufficient memory and slow operation, which has a bad impact on real-time remote sensing image monitoring. Real-time remote sensing image monitoring technology utilizes HR remote sensing images to grasp the changes of urban indicators in real time, and then utilizes the effective information to help decision-makers make scientific planning and decision-making. Also, remote sensing imagery is quite complex, so its reconstruction process faces many challenges. First, remote sensing images have more complex backgrounds, textures, and HF details, and a wide variety of feature information compared with natural images. Second, they contain complex and rich content structures. Third, the distances of the scenes that they represent are very large, so the feature targets containing important information account for relatively small areas in the overall image. Last, they usually have low spatial resolution and low contrast.

To address the aforementioned problems, this article proposes a lightweight super-resolution scheme for processing remote sensing images called multiscale residual attention information distillation network (MRA-IDN). It accounts for the characteristics of remote sensing images just extensively addressed. The main contributions of this article are as follows.

1) To avoid the tendency of most existing models to learn LF information and lose some HF information, a lightweight super-resolution network model is designed by using an HF and LF reconstruction strategy. The LF information is reconstructed by bilinear interpolation. The HF information is obtained by our CNN network.
2) To reconstruct more HF information and reduce the number of parameters, a set of novel multiscale residual attention information distillation group (MRA-IDG) modules is designed as the backbone of the CNN network for learning HF information. A single MRA-IDG consists of a multiscale residual information distillation (MSR-ID) block and an HF channel aware (HFCA) attention block.

3) The MSR-ID modules are constructed and combined by using multiple residual convolutional subblocks (RCBs) with different sizes for solving the problems related to complex content structure and large differences for different scales of remote sensing images. The HFCA is designed for retaining the complex texture and HF details of image contents.
4) The experimental results show that, compared to the existing methods, such as the MHAN, the proposed MRA-IDN scheme significantly reduces the number of parameters by 75% because of the distillation group structure of MRA-IDG blocks. It can also enhance the edge details due to the high- and LF reconstruction strategy and the MRA-IDG attention module. Our MRA-IDN model shows better quality indexes values, clearer texture structure, and more accurate image details compared with other algorithms on the public remote sensing super-resolution datasets WHU-RS19 [37] and RSSCN7 [38].

The rest of this article is organized as follows. Section II presents our work, and describes the methods proposed by researchers in recent years as well as their advantages and disadvantages. Section III describes the design of the proposed method in detail. Section IV provides all experimental details to verify the performance of the algorithm and discusses the obtained results. Finally, Section V concludes this article.

II. Related Work

Recently, researchers have proposed many deep-learning-based super-resolution methods for remote sensing images, among which the CNN-based and GAN-based methods are widely used. Compared with the latter ones, the former neither require complex adversarial techniques during the training stage, nor iterative generation and optimization during the reasoning phase. The structure of the CNN model is easy to implement and adjust, and a higher amount of attention is paid to maintain image details and structural information. Moreover, digital images consisting of LF and HF components make the adoption of HF–LF reconstruction strategy more necessary. The attention mechanism can help the model to extract the desired features and thus improve the performance. Therefore, this article intends to investigate the CNN-based methods and utilize their advantages to design an improved lightweight super-resolution technique for generating high-quality remote sensing images.

A. CNN-Based Super-Resolution

The CNN commonly processes complex data, such as images and videos, by using convolutional operations to automatically extract features from data, and then uses fully connected layers for classification or regression tasks. In the field of image super-resolution, Dong et al. [16] first proposed a CNN-based super-resolution model, the SRCNN, which had a low number of layers. So it could not extract a sufficient amount of effective information [17]. Hara and Tanaka [39] proposed a super-deep image super-resolution model, known as VDSR, containing 20 convolutional layers. Shi et al. [40] presented an efficient subpixel CNN that utilizes a subpixel convolutional layer to boost the LR feature map to the final output. Dong et al. [41] also designed a fast super-resolution model (FSRCNN) trained directly on LR images. It uses a deconvolutional layer to upsample the feature maps, and its reconstruction speed was significantly improved compared to the existing models.

Improving the visual quality is important for the neural networks just cited because they usually yield low perceptual quality. Since the information features of an image cannot be adequately extracted by a single convolutional kernel, more researchers have proposed to use multiple convolutional kernels of different sizes to form different convolutional blocks to construct a multiscale information feature extraction module [42], [43], which improves the feature extraction capability of the network. Jiang et al. [44] proposed a new deep distillation recurrent network to improve the visual quality of reconstructed images. It uses a multiscale purification unit to compensate for the HF components lost during information propagation. Lu et al. [45] presented a multiscale residual neural network that can extract large-, medium-, and small-scale image features and fuse the multiscale information to generate images with high visual quality. Wang et al. [46] proposed an adaptive multiscale feature fusion network, which can adaptively learn multiscale features to improve the information usage efficiency. Zhang et al. [47] proposed a progressive residual deep neural network, which consists of a progressive residual structure that gradually learns different levels and different perception fields of the image feature maps to provide detailed features, and consequently generate more accurate edge and texture information.

B. Attention-Based Super-Resolution

Generally, the increasing number of CNN layers can improve the visual quality of the reconstructed images, but this may considerably raise the computational and memory resources. All channels are treated identically by the CNN kernels, and it is impossible to use LR images containing rich HF information. In image processing, the attention mechanism can be employed to solve this problem and it is divided into channel domain [11], [48] and spatial domain [49] attention. This technique can help a model focus on specific features or regions of the image, thereby improving its performance.

To solve the problem of equal interchannel weightage given to rich LF information contained in LR images, Zhang et al. [34] proposed a very deep residual network based on channel attention. Dai et al. [35] designed a deep SAN that uses second-order feature statistics for more powerful feature representation and correlation learning. Niu et al. [50] presented a holistic attention network that consists of a layer attention module and a channel space attention module, which can learn correlations between different layers. Huang et al. [51] implemented a deep dual-residual attention module, which enables the network to focus more on HF information regions to achieve global and local information fusion, and consequently generate detailed information that improves the visual quality of the reconstructed images.

C. Super-Resolution Based on High- and Low-Frequency Features

A digital image is composed of LF and HF components. The former represent the regions where the intensity value of pixels changes slowly in the image, and describes its main contour region. The latter correspond to the part of the image that shows quicker changes, i.e., the edge contour or noise, and thus the details in the image.

In previous articles, both LF and HF components were reconstructed using the same neural network, such as SRCNN, FSRCNN, VDSR, SRGAN, ESRGAN, RCAN, SAN, MHAN [52], EEGAN [53], and other models. However, this can cause the loss of HF components in LR images. Therefore, the current emphasis of research in super-resolution is to adopt different strategies for learning the HF and LF features of remote sensing images separately. Zhang et al. [54] proposed a residual-in-residual network that focuses on learning HF information. Tian et al. [55] proposed a lightweight enhanced CNN (LESRCNN) composed of an information extraction and enhancement block (IEEB), a reconstruction block (RB), and an information refinement block (IRB). The IEEB cannot only extract LF features but also remove redundant features. The RB converts LF features into HF features by fusing global and local features, and the IRB learns additional HF features to recover more HF detailed information. Peng et al. [56] presented a gated CNN, the PGCNN, which focuses on learning HF information by using migration learning and several residual blocks containing gated convolutional units with long hops to generate more texture and detailed information. Tian et al. [57] proposed an asymmetric CNN, known as ACNet, which consists of an asymmetric block (AB), a memory enhancement block (MEB), and an HF feature enhancement block (HFFEB). The MEB employs residual learning to fuse LF features from the AB and converts the obtained LF features into HF features. The HFFEB uses both LF and HF features to obtain more robust super-resolution features, thus mitigating the excessive feature enhancement problem.

III. Method

To achieve a reconstruction performance comparable to that of deep super-resolution networks, under the conditions of a smaller number of model parameters and a lower computational complexity, a novel lightweight super-resolution network framework, the MRA-IDN, is designed in this article. In this framework, the separate HF and LF reconstruction uses the traditional interpolation algorithm and the neural network to recover LF and HF components, respectively, a choice that significantly reduces the network depth. In the neural network, the MRA-IDG can extract information at multiple scales with a small number of parameters and improve the reconstruction performance with a slight increase in computational complexity. In the proposed design, one MRA-IDG consists of an MSR-ID block and an HFCA block. The details of the proposed MRA-IDN network are described as follows.

A. Network Structure of High- and Low-Frequency Separation Reconstruction

As there is different information in the LF and HF components of an image, the existing super-resolution models that use the same network to process both may be biased toward the reconstruction of the LF components, as they ignore the more detailed and complex HF components in the remote sensing images.

In recent years, some networks, such as PAN [58] and AAN [59], have demonstrated the superior reconstruction performance of HF and LF separation strategy. In these networks, the number of model parameters can be controlled to achieve better image reconstruction, even without deeper network layers. Therefore, in this article, we combine the multiscale attention mechanism and the HF–LF component separation reconstruction strategy to design an improved lightweight super-resolution framework. Its overall structure diagram is shown in Fig. 1.

In the figure, "MRA-IDG" is the multiscale residual attention information distillation group structure proposed in this article, "MSR-ID" is the multiscale residual information distillation block, and "RCB" and "HFCA" refer to the residual convolution sub-block and the HFCA attention block, respectively. "Upsampler" is the upsampling module, and "Conv 3×3" represents a convolution operation with a kernel of size 3×3. In this article, the bilinear interpolation method is used to reconstruct the LF components of the remote sensing images. The reconstruction of the HF components is more complicated and it is achieved by the neural network model. Finally, the LF and HF components of the three RGB channels are added together to obtain the final image.

As Fig. 1 shows, given an input LR image I^{LR}, the corresponding HR image is I^{HR}. The reconstruction of the LF component of the HR image can be expressed as follows:

$$\mathrm{LF} = f_{\mathrm{bilinear}}\left(I^{\mathrm{LR}}\right) \tag{1}$$

where f_{bilinear} is bilinear interpolation.

The reconstruction of the HF component of the HR image is carried out using the neural network. First, the primary features from I^{LR} are extracted using Conv 3×3 shown as follows:

$$x_0 = f_{\mathrm{ext}}\left(I^{\mathrm{LR}}\right) \tag{2}$$

where $f_{\mathrm{ext}}(*)$ denotes the convolution operation with a kernel of size 3×3, an input channel of 3, and an output channel of 64.

After the completion of the primary feature extraction, multiple MRA-IDG dense nonlinear mapping modules are stacked to compose a new powerful feature representation defined as follows:

$$x_g = f^{g}_{\mathrm{MRA-IDG}}\left(f^{g-1}_{\mathrm{MRA-IDG}} \cdots \left(f^{1}_{\mathrm{MRA-IDG}}\left(x_0\right)\right)\right) \tag{3}$$

where x_g represents the feature output after the $g\mathrm{th}$ MRA-IDG block

$$x_{\mathrm{feature}} = f_{\mathrm{fusion}}\left(x_g\right) + x_0. \tag{4}$$

Subsequently, information fusion is achieved by performing a convolution operation on the channel information extracted from each MRA-IDG block using a convolution kernel of size 3×3,

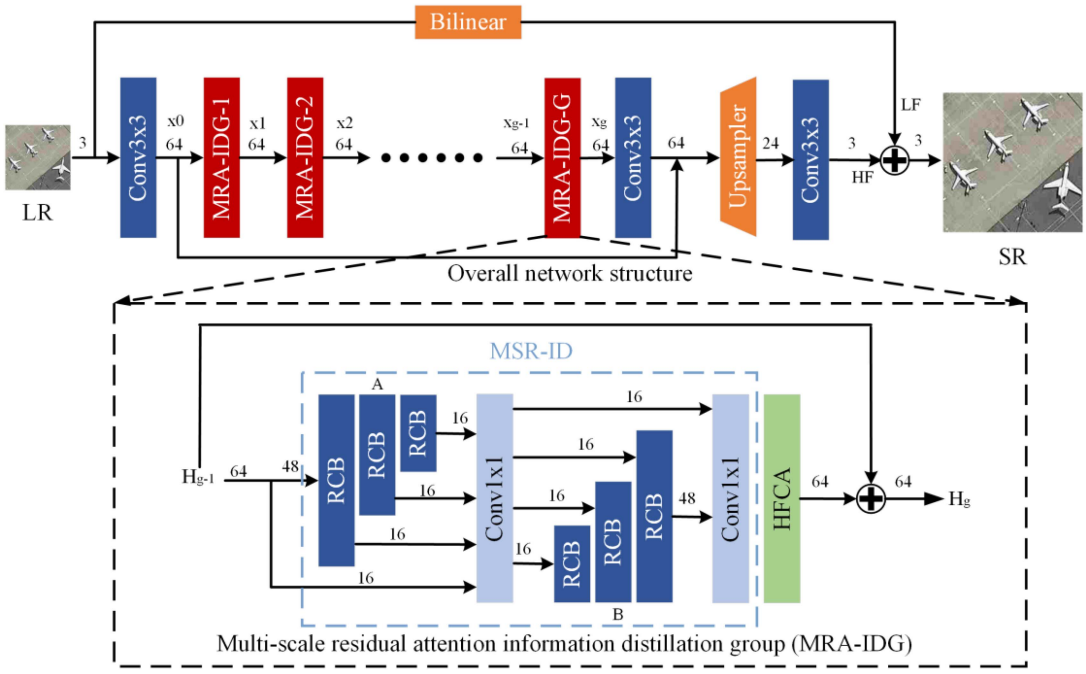


Fig. 1. Representation of the multiscale residual attention information distillation network (MRA-IDN).

as shown by the operator $f_{\text{fusion}}(*)$ in (4). Next, the features from deep and shallow layers are added using the information jump connection technique [18], [19], as shown in (4), which avoids loss of important details after multiple layers of convolution

$$\text{HF} = f_{\text{rec}}(f_{\text{up}}\ (x_{\text{feature}})). \tag{5}$$

The fused information x_{feature} is then integrated and amplified by the upsampling module, the operator $f_{\text{up}}(*)$ in (5). The amplified features are then fused and fed to a convolutional layer with a kernel of size 3×3, an operation represented by the function $f_{\text{rec}}(*)$ in (5).

At the end, the HF and LF components are simply added together. The super-resolved image can thus be expressed as the result of the MRA-IDN model shown as follows:

$$I^{\text{SR}} = H_{\text{MRA-IDN}}\ \left(I^{\text{LR}}\right) = \text{LF} + \text{HF} \tag{6}$$

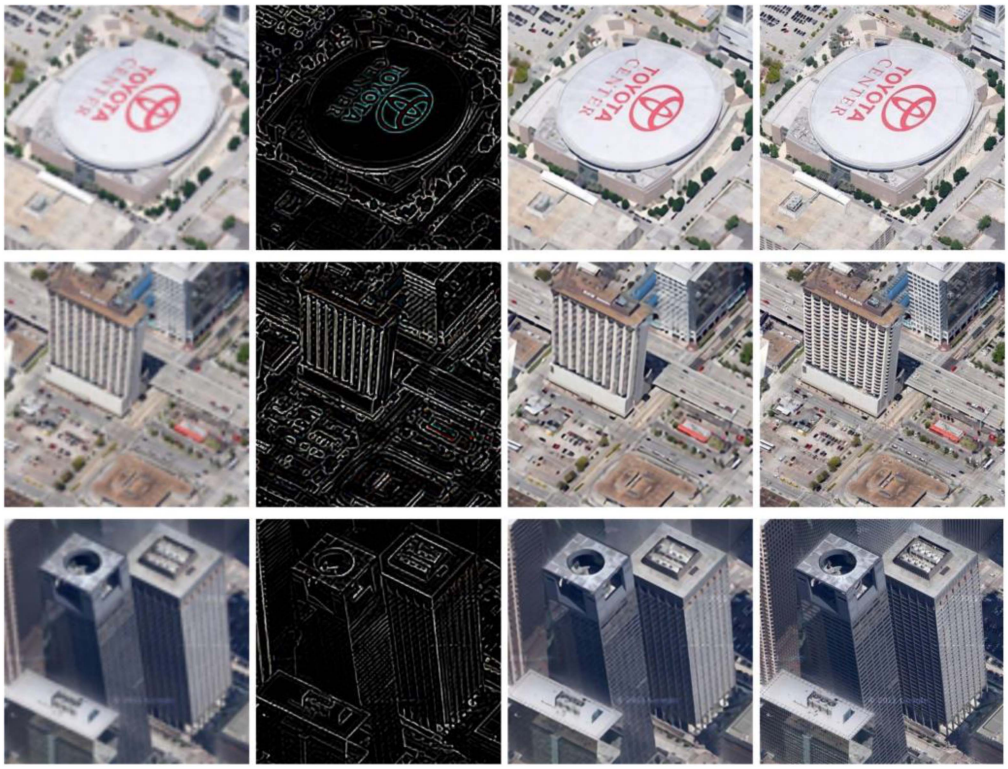
where $H_{\text{MRAIDN}}(*)$ represents the model proposed in this article.

Fig. 2 shows the reconstructed HF and LF components obtained by the MRA-IDN model as well as the final HR reconstructed images. Fig. 2(a) represents the LF components (for the three channels) of the HR image reconstructed from the LR image using the bilinear interpolation method. Fig. 2(b) represents the HF components (for the three channels) of the HR image reconstructed from the LR image by the neural network. The corresponding color channels of both components can be added one by one to obtain the final reconstructed HR image, as shown in Fig. 2(c).

B. Structure of MRA-IDG Block

The RCAN model for SISR is based on residual networks and it obtains a highly superior image super-resolution reconstruction performance by stacking more than 400 neural networks, in which the residual attention module plays an important role. However, the number of channels and layers of convolution in the residual attention module is very large, leading to a great amount of parameters and high model complexity. For example, the RACN network has a parameter count of 15592K. To address the aforementioned problems, this article uses multiple convolutional kernels in series to form a larger convolutional block, so that the convolutional block has the same feature extraction capability as a single multiple convolutional kernels, making it possible to reduce the number of parameters and the complexity of the network. Then, this article designs the novel MRA-IDG block, as shown in Fig. 3, which consists of MSR-ID and HFCA. MSR-ID increases the network nonlinearity by designing a symmetric structure while reducing the number of parameters and considerably improving the expression capability of the network. Meanwhile, the stacking of the RCBs in Fig. 4 with a convolutional kernel of size 3×3 is used to constitute a multiscale feature extraction network, thus enhancing its extraction capability.

The MRA-IDG structure is inspired by networks, such as the IMDN [60], the cross-SRN [61], and the DenseNet [10]. To reduce the computational complexity while enabling the network to have the multiscale feature extraction functionality, the MSR-ID adopts a channel splitting operation strategy, as shown more in detail in *Group* A in Fig. 3. First, the 64-dimensional (64-D) channel of the input feature H_{g-1} of the previous layer is split into four parts equally. Then, the channel features are



(a) LF (b) HF (c) SR (d) HR

Fig. 2. High- and low-frequency separation reconstruction. (a) LF represents the low-frequency component. (b) HF represents the high-frequency component. (c) SR represents the final reconstructed HR result obtained by summing the low-frequency and high-frequency components. (d) HR represents the HR image in the dataset.

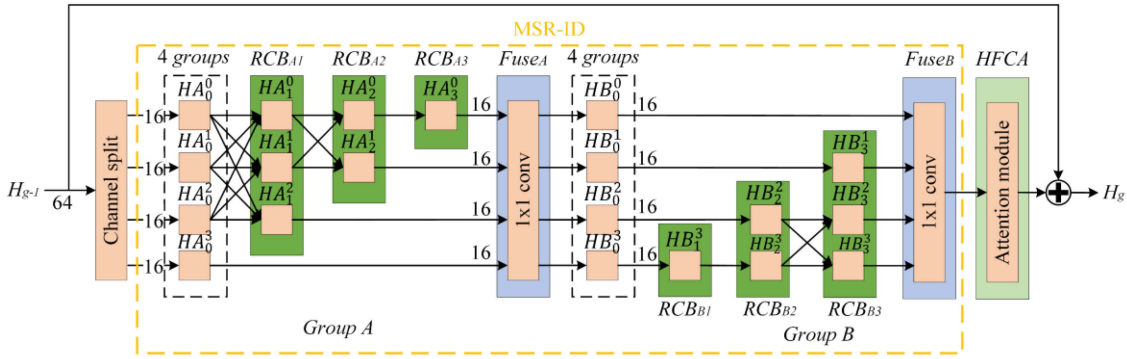


Fig. 3. Multiscale residual attention information distillation group (MRA-IDG).

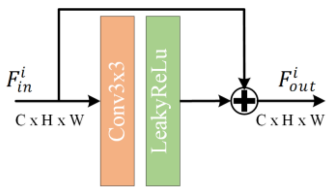


Fig. 4. Residual convolution sub-block (RCB).

successively split and input to the RCB block for convolution operation. The corresponding function mapping is composed of RCBs, whose structure is depicted in Fig. 4. It consists of a convolution module with a kernel of size 3×3 and a nonlinear activation function (the LeakyReLU). The main role of the mapping is to extract the multiscale feature information, which can be expressed as follows:

$$F_{\text{out}}^{(i)} = f\left(F_{\text{in}}^{(i)} \otimes k^{(i)} + b^{(i)}\right) + F_{\text{in}}^{(i)} \tag{7}$$

where f represents the nonlinear mapping, $F_{\text{in}}^{(i)}$ represents the feature input of the ith channel, $\otimes$ represents the convolution operation, $k^{(i)}$ represents the convolution kernel of the ith channel, and $b^{(i)}$ represents the bias of the ith channel.

Finally, the four 16-D channel features, HA_3^0, HA_2^1, HA_1^2, and HA_0^3 are fused using a convolutional kernel of size 1×1,

which can be expressed as follows:

$$HA = f_{\text{FuseA}}\left(\left[HA_3^0, HA_2^1, HA_1^2, HA_0^3\right]\right) \tag{8}$$

where the subscript i and superscript j of HA_i^j represent the ith RCB block and the jth group of information features, respectively.

To fully extract the features between each channel, after the RCB operation in group A, the HA 64-D channel features are split into four 16-D channel features. The split information features will be subjected to the Group B RCB operation which is centrosymmetric with the Group A RCB operation. Finally, the four 16-D channel features, HB_0^0, HB_3^1, HB_3^2, and HB_3^3 are fused using a convolutional kernel of size 1×1, as follows:

$$HB = f_{\text{FuseB}}\left(\left[HB_0^0, HB_3^1, HB_3^2, HB_3^3\right]\right) \tag{9}$$

where the subscript i and superscript j of HB_i^j represent the ith RCB block and the jth group of information features, respectively.

The above HB information features are then fed to the HFCA block to enable the neural network to concentrate more on the reconstruction of HF components. Finally, the information features given by the HFCA block are added one-by-one to the original feature inputs, respectively, thus forming the overall MRA-IDG block that can be expressed as follows:

$$H_g = f_{\text{HFCA}}\ (HB) + H_{g-1} \tag{10}$$

where H_{g-1} is the input feature.

C. HFCA Attention Mechanism

The channel attention introduced by the RCAN network can improve the visual quality of image super-resolution reconstruction; however, the complete information feature in each channel will undergo an average pooling during the process. Generally, the LF component occupies most of the energy of an image. In addition, the average pooling operation increases the feature channel weights of the LF component while suppressing some important channels with more discrete features. This does not satisfy the goal of HF detail reconstruction.

The standard deviation is a standardized metric of dispersion [62]. By pooling the standard deviation of the feature maps for each channel, a metric can be obtained about the variation of features in each channel. The standard deviation calculation of each channel will be updated during the training process, which makes the network can dynamically adjust the weight of each channel to adapt to the changes of different data distributions and tasks. In addition to the information provided by average pooling, standard deviation information can enable the network to fully explore the HF elements between channels. Therefore, this article includes the calculation of the standard deviation of the image channel features to the CA block to fully explore the relationship between feature channels. The HFCA module mines more HF information, allowing more HF details to be reconstructed, as shown in Fig. 5.

Fig. 5(b) shows the schematic diagram of the HFCA sub-block. First, the average pooling and global standard deviation calculation are performed for the feature maps of size $h \times w$ in c dimensions, and two sets of feature maps of size 1×1 in c dimensions can be obtained. Second, the two sets of feature maps are summed one by one by each channel to obtain a set of feature maps of size 1×1 in c dimensions. Third, the above feature maps are given to two fully connected layers for encoding, and a set of feature maps of size 1×1in c dimensions is also obtained. Fourth, after the application of the sigmoid activation function, the weights of size 1×1 in c dimensions are obtained, whose values are between 0 and 1. A value closer to 1 means that the channel weights are larger, and the channel features are more important. Last, the weights of the c dimensions of size 1×1 are multiplied with the corresponding original channel features to complete the assignment of channel attention weights.

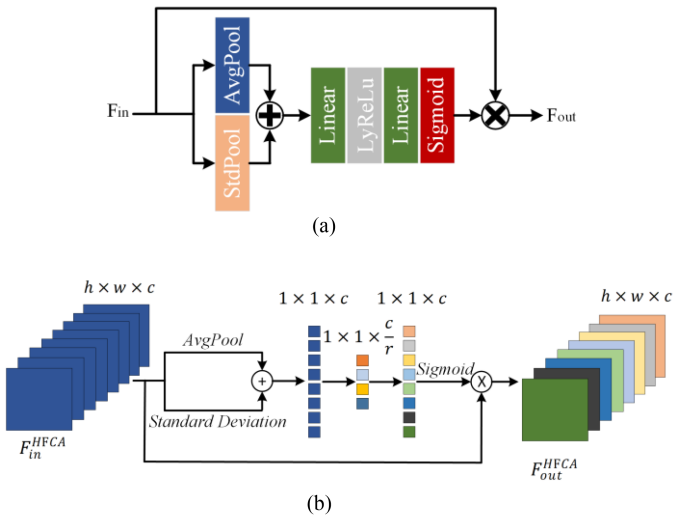


Fig. 5. High-frequency channel aware (HFCA) attention sub-block. (a) Structure diagram. (b) Schematic diagram.

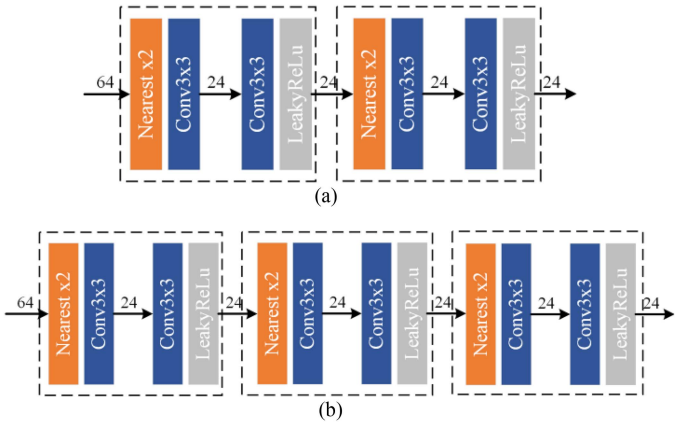


Fig. 6. Upsampling block. (a) $\times 4$ upsampling. (b) $\times 8$ upsampling.

D. Upsampling Block

The same upsampling module as that used in the AAN network is utilized in this article. As Fig. 6 shows, the upsampling blocks that achieve $\times 4$ and $\times 8$ upsampling consist of two and three sets of nearest-neighbor interpolation, respectively. Most

of the upsampling blocks of existing super-resolution networks utilize deconvolution. However, the deconvolution requires a large number of operations, which can further increase the network complexity and the number of parameters. Therefore, this article adopts the nearest-neighbor interpolation combined with a convolution scheme with a lower number of operations and number of parameters.

E. Loss Function

This article uses the mean absolute error or the $L1$ loss function as the optimization function of the network, and compares the performance of different models, such as SRCNN, FSRCNN, VDSR, RCAN, MHAN, etc. Given the training set $\{I_i^{\mathrm{LR}}, I_i^{\mathrm{HR}}\}_i^N$, where N is the number of training pairs of LR and HR images, the loss can be expressed as follows:

$$L\left(I^{\mathrm{SR}}, I^{\mathrm{HR}}\right) = \frac{1}{N}\sum_{i=1}^{N} \parallel H_{\mathrm{MRAIDN}}\left(I_i^{\mathrm{LR}}\right) - I_i^{\mathrm{HR}} \parallel_1 \quad (11)$$

where $\parallel * \parallel_1$ denotes the $L1$ criterion.

F. Evaluation Indicators

This article uses two objective evaluation indicators, PSNR and structural similarity (SSIM) [63], to quantitatively assess the performance of the super-resolution results. PSNR reflects the quality of the image reconstruction, which can be expressed as shown in (12). Generally, a PSNR value above 30 indicates excellent reconstruction performance, and below 20 is poor. SSIM reflects the similarity between the inferred image and the target image, as shown in (13), which is related to the intensity, brightness, and contrast of the image. In the super-resolution task, the larger the PSNR and SSIM values, the higher the similarity of the reconstructed image to the real image, and the better the image reconstruction performance

$$\mathrm{PSNR} = 10\log_{10}\frac{\left(2^N-1\right)^2}{\frac{1}{H\times W}\sum_{i=1}^{H}\sum_{j=1}^{W}\left[P\left(i,j\right)-G\left(i,j\right)\right]^2} \quad (12)$$

$$\mathrm{SSIM} = \left[s\left(P,G\right)\right]^{\alpha}\left[l\left(P,G\right)\right]^{\beta}\left[c\left(P,G\right)\right]^{\gamma} \quad (13)$$

where H and W depict the height and width of the image, respectively. $P(i,j)$ and $G(i,j)$ represent the pixel values in the inferred and target images, respectively. $s(P,G)$, $l(P,G)$, and $c(P,G)$ are the intensity, brightness, and contrast with respect to the inferred and target images, and α, β, and γ correspond to the weights of the significance of these three, respectively, with $\alpha=\beta=\gamma=1$.

IV. Experiments

A. Datasets

The AID dataset is a dataset of remote sensing images collected and produced by Xia et al. [64]. It contains 10 000 images belonging to 30 categories, including images of airports, farmlands, beaches, deserts, etc. The WHU-RS19 and RSSCN7 are datasets of remote sensing images produced by Dai and Wen [37] and Zou et al. [38], respectively. Further detailed information about them is provided in Table I. Fig. 7 shows some randomly extracted images from the three datasets.

To make an effective comparison with existing algorithms, the AID is used as the training set in this article, a choice consistent with MHAN. A total of 30 images are randomly and uniformly extracted from multiple categories in the RSSCN7 dataset as the validation set, which is named Test30 and partly are shown in Fig. 8. WHU-RS19 and RSSCN7 are used as the test sets to validate the performance of MRA-IDN.

B. Experimental Setting

To increase the data diversity, the AID dataset of size 600×600 is randomly cropped to obtain the training set of size 192×192. The generalization ability of the model is improved by using data enhancement strategies on the training set, such as random vertical flip and random horizontal flip, while the validation and test sets use the original unchanged data.

In this article, bicubic interpolation is used to degrade the images and obtain the LR and HR image pairs. The performance of the models is also tested under the magnification factors $\times 4$ and $\times 8$. To ensure a fair comparison, the same degradation model is used for all the considered models, and the same data enhancement strategy is used in the training process. The Adam optimizer [52] is employed to optimize the model for training, the momentum parameters are set to $\beta 1 = 0.9$, $\beta 2 = 0.99$, the batch size is set to 16, the initial learning rate is 0.0005, and a linear decay strategy is used during each round of training.

PyTorch 1.5.0 framework on the Ubuntu 18.04.5 Operating System is used to develop the models proposed in this article. The hardware platform consists of an NVIDIA RTX2080 graphics card that is used to accelerate the training of the models. All algorithms to be compared are tested on the same hardware and software platforms.

C. Ablation Study

Inspired by RCAN, IMDN, and PAN, the basic idea of the MRA-IDG block proposed in this article is derived from the residual channel attention block. Therefore, Fig. 1 is used as the overall framework of the network in the ablation experiments to test the effectiveness of the MRA-IDG blocks (12 layers) and the HFCA sub-block proposed in this article. When the number of MRA-IDG modules is 12, the model in this article is named MRA-IDN, and when the number is 24, the model is named MRA-IDN+.

Table II shows the PSNR metric, SSIM metric, and the number of model parameters for each module that reconstructs Test30 at a magnification factor of $\times 4$. It is known that when both the HFCA sub-block and the MRA-IDG block are simultaneously used in the model, the number of model parameters increases by a small amount. However, the PSNR and SSIM are at their highest values, and the reconstruction quality of the image is optimal at this time.

This article tests the effectiveness of each module through the PSNR trend curve for Test30 during the training process with a $\times 4$ magnification factor. It can be observed from both Fig. 9

TABLE I
DETAILS OF THE DATASETS

	Dataset	Number	Classes	Resolution	Spatial resolution
Train	AID	10000	30	600 × 600	Up to 0.5
Test	WHU-RS19	1005	19	600 × 600	Up to 0.5
	RSSCN7	2800	7	400 × 400	-

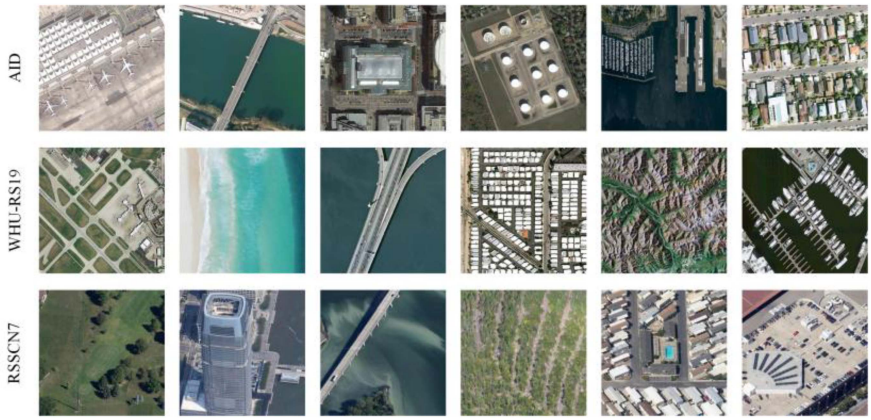


Fig. 7. Sample images from training and testing sets.

Fig. 8. Sample images of the Test30 dataset [40].

TABLE II
ABLATION STUDIES

Scale	HFCA	MRA-IDG	Parameters	Test30 PSNR/SSIM
× 4	×	×	962075	29.59/0.7715
	✓	×	962075	30.60/0.7717
	×	✓	**945131**	30.62/0.7717
	✓	✓	952091	**30.63/0.7732**

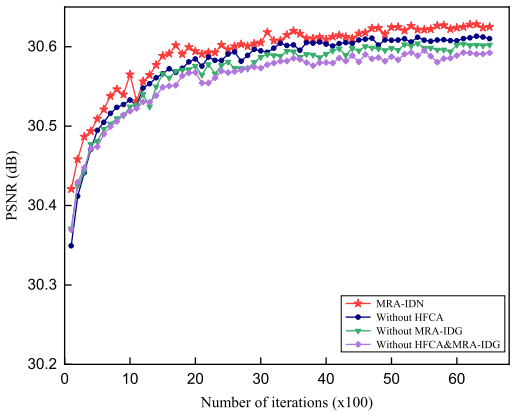


Fig. 9. Plot of PSNR variation in the validation set for the ablation experiment.

TABLE III
ABLATION STUDY OF THE ATTENTION MODULE

Module	Parameters	PSNR (dB)
MRA-IDG+CA	73604	30.61
MRA-IDG+HFCA	73604	**30.63**

PSNR values on the Test30 validation set at ×4 magnification factor and the number of parameters for individual MRA-IDG modules with the addition of different attention modules.

Fig. 10. Image details in the validation set obtained in the ablation experiment.

and Table II that both the HFCA sub-block as well as the MRA-IDG block enhance the reconstruction quality of the images and optimize the number of parameters. Fig. 10 shows in detail the effect of different blocks on the image reconstruction quality. The proposed model, MRA-IDN, which has the MRA-IDG and HFCA modules, reconstructs the images with sharpest and most accurate details. The comparison of WO-HFCA and W_HFCA cases, as depicted in Fig. 11, shows that the addition of HFCA attention block can help the network model reconstruct images with more accurate details.

As the HFCA block is inspired by the CA mechanism, ablation experiments are conducted to verify the effectiveness. Table III shows the PSNR values for their reconstructed validation set images and the number of parameters used by a single MRA-IDG block when the CA and HFCA module are added to it, respectively. It can be concluded from Table III that the comparison between the WO_HFCA image and the MRA-IDN image in Fig. 10, and the heat map of the attention mechanism in Fig. 11 that the reconstructed images have the best quality when the MRA-IDG block is added to the proposed HFCA sub-block.

Fig. 11 shows the feature map of the HFCA attention mechanism for reconstructing the images, where WO_HFCA and W_HFCA indicate the absence or presence of HFCA in the model, respectively. The HFCA sub-block can enable the network to pay more attention to HF information, such as lines, and suppress most of the responses of LF component pixels. This in turn makes the neural network focus more on the HF components that are difficult to reconstruct, while the restoration of most of the LF components is carried out by the traditional interpolation method.

In addition, this article tests the effectiveness of the upsampling module for HF networks. Ablation experiments are performed on the upsampling module based on the nearest interpolation in this article and subpixel convolution. It can be concluded from Table IV that the upsampling module in this article has lower parameters and higher PSNR values.

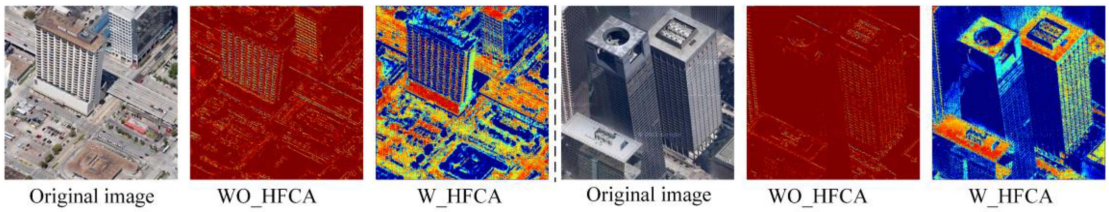


Fig. 11. Impact of HFCA attention module on the model.

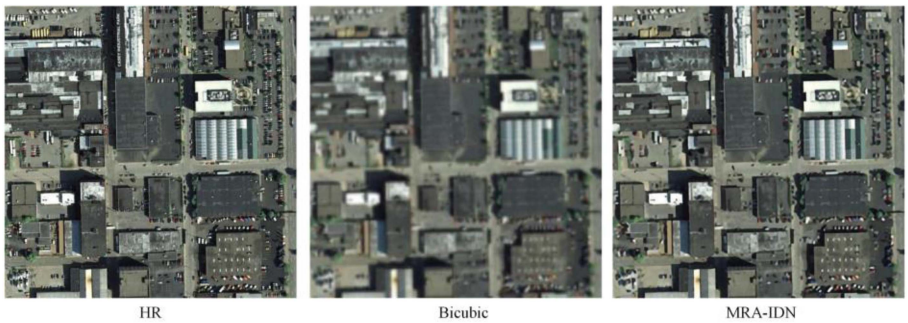


Fig. 12. Reconstructed images by different models under the condition of a $\times 4$ magnification factor.

TABLE IV
ABLATION STUDY OF THE UPSAMPLER MODULE

Module	Parameters (K)	PSNR (dB)
Pixel-shuffle layer	2027	30.66
Upsampler (MRA-IDN+)	1835	**30.67**

PSNR values on the Test30 validation set at ×4 magnification factor and the number of parameters for the network with addition of different upsampling modules.

TABLE V
ABLATION STUDY OF THE LOW-FREQUENCY NETWORK BASED ON INTERPOLATION

Interpolation	Parameters (K)	PSNR (dB)
Bicubic	1835	30.66
Bilinear (MRA-IDN+)	1835	**30.67**

PSNR values on the Test30 validation set at ×4 magnification factor and the number of parameters for the network with different interpolation.

Finally, this article tests the effectiveness of LF networks. Ablation experiments are performed on LF networks based on bilinear interpolation and bicubic interpolation. Other interpolation methods cannot be used due to different dimensions. It can be concluded from Table V that the bilinear interpolation in this article for the reconstruction of the LF component is better.

D. Comparison With Advanced Technologies

This article compares the proposed model with two kinds of existing advanced models using the WHU-RS19 and RSSCN7 datasets. The first one includes the generalized large-scale super-resolution reconstruction (GLSSR) algorithms, including RDN, D-DBPN [65], RCAN, SRFBN [66], SAN, and MHAN. The second one includes the lightweight remote sensing image super-resolution reconstruction (LRSISR) algorithms, including SR-CNN, VDSR, IMDN, PAN, LESRCNN, ACNet, AAN, FeNet [67], Omnisr [68], and HAUNet_S [69].

Fig. 12 shows the reconstruction performance of the bicubic interpolation method and our proposed model MRA-IDN on the LR remote sensing image under the condition of $\times 4$ magnification factor. It can be observed that the image reconstructed by the bicubic interpolation method is blurred and unclear, and lacks most of the HF components. Compared with the interpolation, the proposed algorithm based on deep learning can reconstruct and restore most of the HF components, and the reconstructed performance is closer to the original HR image.

To comprehensively analyze and compare the performance of the model proposed in this article, the computational resource consumption and reconstructed image quality metrics of each generic model under a $\times 4$ magnification factor are tested in this article. Table VI shows the actual running time, the occupied GPU memory, floating point operations (FLOPs), number of parameters, and PSNR values of the reconstructed images of Test30 under a $\times 4$ magnification factor. As Table VI shows, compared with the GLSSR algorithms, the FLOPs, the number of parameters, the occupied GPU memory, and the actual inference time of the proposed model are the lowest when the reconstruction results have the same PSNR values. On the other hand, the PSNR of the proposed model is higher when the FLOPs, the number of parameters, the occupied GPU memory, and the actual inference time are similar to those of the LRSISR algorithms.

The measured FLOPs, number of parameters, occupied GPU memory, and actual inference time for the MRA-IDN model are 2.21 G, 962 k, 749 MB, and 23.64 ms, respectively, under a $\times 8$ amplification factor, while the measured FLOPs, number of parameters, occupied GPU memory, and actual inference time for the MRA-IDN+ model are 3.29 G, 1845 k, 1491 MB, and

TABLE VI
DETAILS OF THE DATASETS

Type	Year	Method	FLOPs (G)	Parameters(K)	Memory (MB)	Time (ms)	PSNR (dB)
GLSSR	2018	RDN	79.69	16520	2345	31.01	30.54
	2018	D-DBPN	10234	10287	10145	63.41	30.65
	2018	RCAN	72.55	15592	2117	96.36	30.57
	2019	SRFBN	7715	3631	6289	55.38	30.64
	2019	SAN	12.73	11980	2465	90.45	30.68
	2021	MHAN	56.44	8931	1619	50.34	30.51
LRSISR	2014	SRCNN	**0.305**	**57**	791	**9.51**	30.14
	2016	VDSR	2.74	667	987	10.83	30.26
	2019	IMDN	3.14	715	836	13.68	30.60
	2020	PAN	2.20	272	868	17.30	30.57
	2020	LESRCNN	14.51	774	1160	17.42	30.55
	2021	ACNECT	22.03	1504	1346	22.45	30.57
	2021	AAN	5.44	1047	870	19.80	30.60
	2022	FeNet	4.78	366	822	22.40	30.50
	2023	Omnisr	4.10	805	1093	29.57	30.47
	2023	HAUNet_S	27.14	2019	2015	97.73	30.53
	2023	MRA-IDN	5.66	952	**729**	23.36	30.63
	2023	MRA-IDN+	10.00	1835	1321	35.58	**30.67**

Performance comparison of the reconstructed Test30 validation set for each model at a ×4 magnification factor.

TABLE VII
RECONSTRUCTION PERFORMANCE OF THE WHU-RS19 TEST SET FOR EACH MODEL AT A ×4 AMPLIFICATION FACTOR

WHU-RS19	RCAN	SRFBN	SAN	MHAN	FeNet	Omnisr	HAUNet_S	MRA-IDN+
	PSNR/SSIM	PSNR/SSIM	PSNR/SSIM	PSNR/SSIM	PSNR/SSIM	PSNR/SSIM	PSNR/SSIM	PSNR/SSIM
Airport	28.77/0.828	28.99/0.828	29.01/**0.831**	29.02/0.830	28.91/0.825	28.69/0.819	28.17/0.803	**29.06**/0.829
Beach	47.32/0.981	47.35/0.981	47.39/0.981	**47.42**/**0.981**	46.93/0.979	42.10/0.978	45.13/0.973	47.16/0.980
Bridge	35.25/0.916	35.58/0.918	35.54/0.919	35.56/**0.919**	35.05/0.914	35.05/0.914	34.60/0.908	**35.59**/0.918
Commercial	25.55/0.749	25.70/0.758	25.71/0.762	25.74/**0.762**	25.46/0.746	25.50/0.748	25.11/0.737	**25.78**/0.760
Desert	40.83/0.936	40.76/0.936	40.77/0.936	40.77/**0.937**	**40.88**/0.936	39.94/0.934	40.43/0.931	40.78/0.935
Farmland	37.55/0.889	37.67/0.891	**37.71**/**0.894**	37.70/0.894	37.47/0.889	37.31/0.888	36.65/0.874	37.69/0.892
FootballField	29.62/0.842	29.92/0.850	30.02/**0.856**	30.03/0.855	29.42/0.837	29.46/0.839	28.85/0.825	**30.05**/0.853
Forest	28.67/0.691	28.69/0.693	28.71/0.698	28.71/**0.698**	28.58/0.686	28.69/0.693	28.26/0.667	**28.74**/0.696
Industrial	28.01/0.799	28.26/0.808	28.29/0.813	28.30/**0.813**	27.88/0.796	27.91/0.797	27.29/0.782	**28.35**/0.811
Meadow	37.77/0.874	37.80/0.874	**37.82**/**0.875**	37.81/0.875	37.75/0.874	37.57/0.873	37.06/0.860	37.74/0.874
Mountain	25.43/0.631	25.45/0.633	25.50/0.637	25.51/0.637	25.39/0.627	25.45/0.632	25.52/0.634	**25.51**/**0.637**
Park	28.78/0.761	28.92/0.764	29.92/0.764	28.93/0.763	28.71/0.757	28.71/0.759	28.34/0.748	**28.96**/**0.766**
Parking	28.52/0.866	29.11/0.878	29.21/0.887	29.23/**0.887**	28.08/0.857	28.59/0.868	27.55/0.849	**29.50**/0.885
Pond	32.70/0.879	32.81/0.881	32.81/0.882	**32.84**/**0.882**	32.66/0.879	32.54/0.878	32.23/0.872	32.82/0.881
Port	28.19/0.856	28.49/0.863	28.60/**0.867**	28.61/0.866	27.96/0.853	28.13/0.855	27.62/0.847	**28.61**/0.864
RailwayStation	27.29/0.747	27.60/0.759	27.63/**0.764**	27.63/0.763	27.05/0.739	27.24/0.744	26.70/0.724	**27.66**/0.762
Residential	25.96/0.790	26.17/0.800	26.19/0.803	26.21/0.803	25.82/0.785	25.96/0.791	25.31/0.777	**26.35**/**0.804**
River	29.31/0.765	29.37/0.767	29.37/0.768	29.38/0.768	29.26/0.763	29.28/0.764	28.96/0.754	**29.39**/**0.768**
Viaduct	27.11/0.773	27.30/0.788	27.37/0.790	27.30/0.790	26.88/0.770	26.95/0.772	26.39/0.754	**27.44**/**0.791**
Average	31.32/0.822	31.35/0.824	31.38/0.827	31.40/**0.828**	31.01/0.816	30.76/0.818	30.49/0.806	**31.43**/0.827

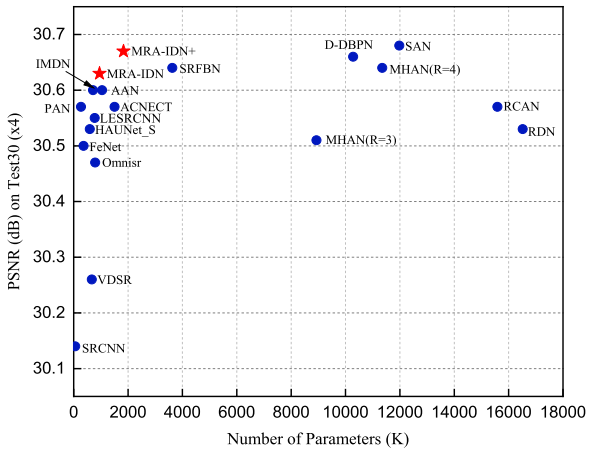
44.48 ms, respectively. Based on the values of the MRA-IDN and MRA-IDN+ models at ×4 and ×8 magnification factors, we can observe that the number of parameters and the actual inference time remain almost the same regardless of the magnification factor. Moreover, the number of parameters of this model as well as the time required for its computation are significantly lower than those of other large networks. Thus, our MRA-IDN is lightweight and has a stable performance, so it achieves a better tradeoff between inference time and model space.

The performance of the models is compared visually in Fig. 13(a) and (b), which show the 2-D plane plots between the model performance and the number of parameters, and between the model performance and the occupied GPU memory for the reconstruction of Test30 under a ×4 magnification factor for each model in Table VI, respectively. It can be observed that the algorithm proposed in this article has good performance in terms of both the number of parameters and the actual occupied GPU memory while achieving high PSNR and SSIM values.

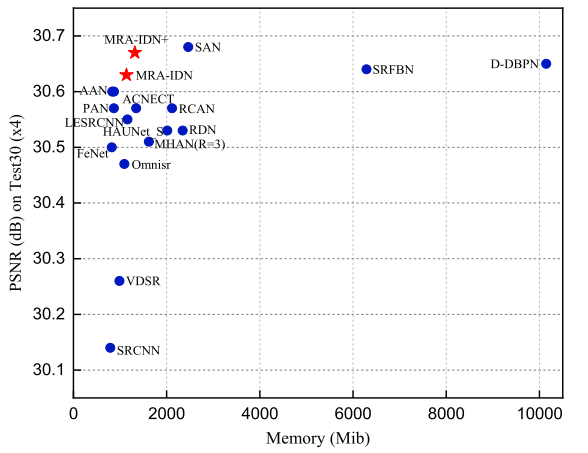
In order to carry out a more comprehensive comparison and analysis of the performance of the model proposed in this article, the reconstruction performance of various models is tested under the same experimental conditions, using the same training and test sets at different magnification factors, respectively. To ensure the accuracy and credibility of the data, the reconstruction

TABLE VIII
RECONSTRUCTION PERFORMANCE OF THE RSCCN7 TEST SET FOR EACH MODEL AT A $\times 4$ AMPLIFICATION FACTOR

RSSCN7	RCAN	SRFBN	SAN	MHAN	FeNet	Omnisr	HAUNet_S	MRA-IDN+
	PSNR/SSIM	PSNR/SSIM	PSNR/SSIM	PSNR/SSIM	PSNR/SSIM	PSNR/SSIM	PSNR/SSIM	PSNR/SSIM
aGrass	34.24/0.831	34.32/0.833	34.34/0.834	**34.34**/**0.834**	34.16/0.831	34.09/0.830	34.17/0.829	34.33/0.832
bField	33.32/0.757	33.36/0.758	33.38/0.759	**33.39**/0.759	33.26/0.757	33.19/0.755	33.21/0.754	33.36/**0.759**
cIndustry	26.14/0.732	26.36/0.743	26.42/**0.746**	26.43/0.745	26.10/0.731	26.09/0.732	26.15/0.731	**26.45**/0.745
dRiverLake	31.21/0.830	31.21/0.830	31.24/0.832	**31.24**/**0.832**	31.06/0.827	31.02/0.827	31.08/0.826	31.23/0.831
eForest	28.35/0.636	28.38/0.637	**28.41**/0.640	28.40/0.640	28.33/0.634	28.34/0.636	28.37/0.638	28.40/**0.640**
fResident	25.15/0.707	25.33/0.717	25.39/0.721	25.39/**0.721**	25.10/0.705	25.16/0.710	25.25/0.712	**25.40**/0.720
gParking	25.36/0.694	25.55/0.705	25.46/0.706	25.61/**0.709**	25.33/0.693	25.35/0.695	25.40/0.694	**25.61**/0.708
Average	29.25/0.747	29.22/0.745	29.26/0.749	**29.28**/0.750	29.05/0.740	29.03/0.741	29.09/0.741	29.25/**0.750**



(a)



(b)

Fig. 13. Comparison between the proposed model and existing algorithms for the reconstruction of the validation set Test30 at a $\times 4$ magnification factor. Reconstruction results, number of parameters, and the occupied GPU memory are considered. (a) Parameter comparison of different models. (b) Memory comparison of different models.

results data of other models in Tables VII–X are based on [52] except for the one of FeNet, Omnisr, HAUNet_S, and the proposed algorithm. The reconstruction results of FeNet, Omnisr, and HAUNet_S are included to allow a better comparison of the proposed model with the latest super-resolution models of remote sensing images.

Tables VII and VIII show the PSNR and SSIM corresponding to each category in the reconstructed WHU-RS19 and RSSCN7 for each model under a $\times 4$ magnification factor, where WHU-RS19 and RSSCN7 contain 19 and 7 categories, respectively. The last rows of the tables show the average values for each category.

As Tables VII and VIII show, the average reconstruction quality of this article's model in multiple categories is better in terms of PSNR or SSIM, and it reaches optimal values for complex scenes of Parking, Park, Railway Station, Residential, etc. This proves that the multiscale information extraction by the proposed MRA-IDG block can provide more detailed features in complex target scenes, thus improving the reconstruction quality of the neural network.

Fig. 14 compares the details of the reconstructed WHU-RS19 and RSSCN7 obtained using each of the algorithm of the aforementioned tables for a $\times 4$ magnification factor. It can be observed that the method proposed in this article provides the most complete and clearest image reconstruction details. This can be noticed from the better visual quality, clarity, and accuracy of the font lines in image c157, the building edges and window details in image buildings c165 and c160, and the color in image f015. The proposed MRA-IDN can achieve similar or better reconstruction quality than the existing methods, such as the MHAN, that belongs to the GLSSR methods, and the FeNet, that belongs to the LRSISR methods. However, Table VI and Fig. 14 show that the MRA-IDN model can have a lower model complexity, a smaller number of parameters, and a lower memory utilization while maintaining a high PSNR.

To carry out a more comprehensive evaluation of the performance of the proposed model with existing methods, this article also shows reconstruction results at a $\times 8$ magnification factor. Tables IX and X show the corresponding average PSNR and SSIM of each category of WHU-RS19 and RSSCN7 reconstructed by each model, respectively. The average reconstruction quality of the proposed model is the highest over multiple categories, and its reconstruction performance is better than those of other models of the same category under a $\times 8$ magnification factor. As the HF and LF components are separated, the neural network in MRA-IDN can focus on the reconstruction of HF components. Furthermore, the advantage of this structure is more prominent for reconstruction at a higher magnification and, therefore, the quality of the reconstructed images can be optimized.

Fig. 15 shows the reconstructed image details corresponding to the WHU-RS19 and RSSCN7 test sets, obtained by each model under the condition of a $\times 8$ magnification factor. It can be

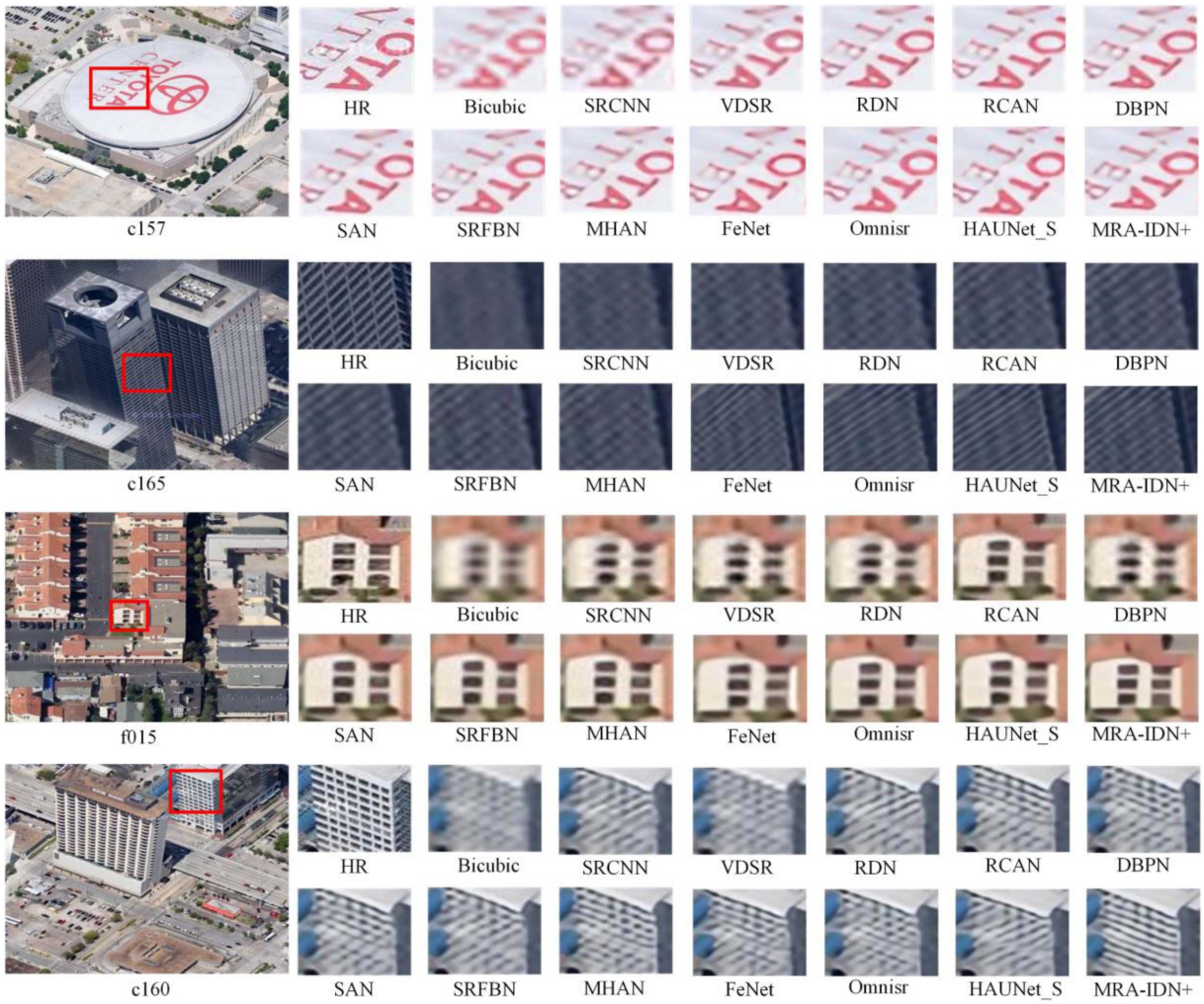


Fig. 14. Comparison of reconstructed image details between our model and existing algorithms at a $\times 4$ factor.

TABLE IX
PSNR AND SSIM METRICS OF THE RECONSTRUCTED WHU-RS19 TEST SET FOR EACH MODEL AT A $\times 8$ AMPLIFICATION FACTOR

WHU-RS19	RCAN	SRFBN	SAN	MHAN	FeNet	Omnisr	HAUNet_S	MRA-IDN+
	PSNR/SSIM	PSNR/SSIM	PSNR/SSIM	PSNR/SSIM	PSNR/SSIM	PSNR/SSIM	PSNR/SSIM	PSNR/SSIM
Airport	24.61/0.631	24.49/0.622	24.52/0.628	24.64/0.633	24.95/0.654	24.63/0.638	24.46/0.628	**25.05/0.658**
Beach	41.44/0.950	41.31/0.942	41.38/0.947	41.42/0.948	41.34/0.950	36.64/0.941	40.68/0.942	**41.67/0.951**
Bridge	30.43/0.837	30.30/0.833	30.41/0.837	30.51/0.840	30.55/0.840	30.52/0.840	30.49/0.840	**31.20/0.849**
Commercial	21.55/0.501	21.48/0.497	21.51/0.499	21.56/0.501	21.54/0.507	21.60/0.511	21.38/0.500	**21.87/0.530**
Desert	38.15/0.902	38.02/0.899	38.17/0.900	38.17/0.903	**38.29/0.903**	37.25/0.899	38.17/0.900	38.25/0.902
Farmland	33.17/0.791	33.07/0.787	33.11/0.790	33.20/0.793	33.22/0.794	33.08/0.793	32.96/0.783	**33.55/0.801**
FootballField	24.95/0.665	24.96/0.669	24.92/0.664	24.99/0.668	24.98/0.673	24.98/0.672	24.69/0.661	**25.63/0.701**
Forest	25.53/0.451	25.50/0.450	25.50/0.448	25.58/0.452	25.49/0.449	25.53/0.455	25.49/0.440	**25.61/0.459**
Industrial	23.23/0.559	23.15/0.553	23.19/0.557	23.20/0.557	23.33/0.572	23.29/0.570	23.01/0.556	**23.83/0.601**
Meadow	34.64/0.803	34.55/0.798	34.52/0.799	34.69/**0.806**	34.70/0.804	34.41/0.802	34.54/0.798	**34.75**/0.804
Mountain	22.67/0.392	22.61/0.388	22.64/0.389	22.73/0.394	22.66/0.392	22.69/0.398	**22.80**/0.397	22.76/**0.401**
Park	25.14/0.567	25.11/0.566	25.10/0.566	25.18/0.568	25.16/0.571	25.11/0.572	24.99/0.561	**25.38/0.582**
Parking	22.42/0.629	22.34/0.624	22.39/0.625	22.43/0.629	22.31/0.633	22.57/0.646	22.16/0.624	**23.40/0.690**
Pond	29.13/0.785	29.08/0.784	29.08/0.782	29.15/0.785	29.13/0.788	29.06/0.787	29.00/0.786	**29.41/0.794**
Port	23.59/0.700	23.50/0.695	23.51/0.692	23.63/0.702	23.57/0.706	23.72/0.711	23.41/0.703	**24.17/0.730**
RailwayStation	23.00/0.478	22.95/0.476	22.93/0.473	23.04/0.479	23.00/0.482	23.04/0.487	22.98/0.474	**23.38/0.507**
Residential	21.01/0.515	20.97/0.512	20.94/0.511	21.00/0.513	21.04/0.525	21.18/0.536	20.86/0.522	**21.60/0.569**
River	25.88/0.584	25.82/0.581	25.78/0.579	25.88/0.585	25.85/0.586	25.89/0.590	25.75/0.581	**26.05/0.595**
Viaduct	22.52/0.512	22.42/0.507	22.50/0.511	22.55/0.514	22.72/**0.650**	22.65/0.528	22.41/0.514	**23.18**/0.560
Average	27.00/0.644	26.93/0.641	26.95/0.642	27.03/0.642	27.01/0.650	26.70/0.651	26.83/0.642	**27.41/0.668**

TABLE X
PSNR AND SSIM METRICS OF THE RECONSTRUCTED RSCCN7 TEST SET FOR EACH MODEL AT A ×8 AMPLIFICATION FACTOR

RSSCN7	RCAN	SRFBN	SAN	MHAN	FeNet	Omnisr	HAUNet_S	MRA-IDN+
	PSNR/SSIM	PSNR/SSIM	PSNR/SSIM	PSNR/SSIM	PSNR/SSIM	PSNR/SSIM	PSNR/SSIM	PSNR/SSIM
aGrass	31.61/0.751	31.54/0.748	31.57/0.749	31.64/0.752	31.68/0.754	31.50/0.751	31.53/0.749	**31.83/0.756**
bField	31.40/0.692	33.30/0.688	31.35/0.690	31.38/0.691	31.51/0.696	31.34/0.693	32.66/0.752	**31.63/0.699**
cIndustry	22.63/0.522	22.54/0.516	22.55/0.517	22.66/0.524	22.75/0.535	22.70/0.533	22.19/0.504	**23.06/0.555**
dRiverLake	28.42/0.733	28.30/0.725	28.32/0.729	28.42/0.731	28.47/0.737	28.39/0.736	28.40/0.739	**28.67/0.743**
eForest	26.15/0.457	26.08/0.452	26.12/0.455	26.17/0.458	26.15/2.458	26.14/0.460	26.39/0.455	**26.21/0.463**
fResident	21.58/0.458	21.50/0.451	21.51/0.452	21.63/0.460	21.60/0.466	21.65/0.473	20.97/0.420	**21.87/0.489**
gParking	22.50/0.497	22.42/0.492	22.46/0.496	22.48/0.497	22.57/0.507	22.52/0.506	22.57/0.484	**22.78/0.522**
Average	26.32/0.587	26.24/0.581	26.26/0.584	26.34/0.588	26.39/0.593	26.32/0.593	26.39/0.586	**26.58/0.604**

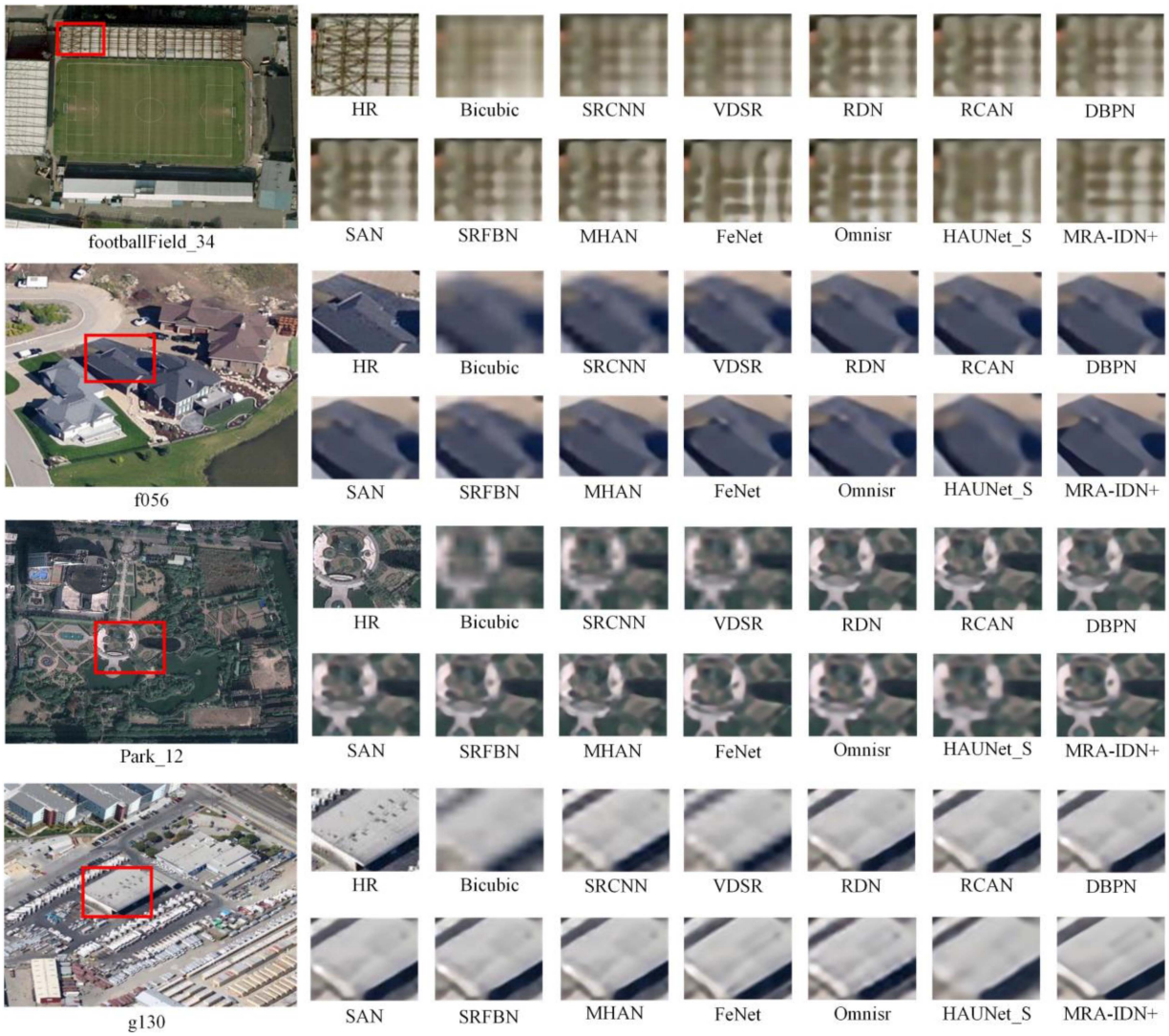


Fig. 15. Reconstructed image details obtained with our model and existing algorithms at a ×8 amplification factor.

observed that the details of the images reconstructed using the model proposed in this article are superior to those of the existing algorithms. In the house of f056, the image reconstructed by MRA-IDN has the highest definition, and the reconstructed line details of edges and borders are more accurate and clearer. For the steel bar details of footballField_34, the accuracy of the image reconstructed by the proposed algorithm is optimal. The g130 image shows that the reconstructed house is clearer and rounder with less noise.

Therefore, combining the visualization results of Figs. 13 and 14, for example, by image c157, c165, f015, c160, f056, and g130, it can be concluded that, on the RSSCN7 dataset, no matter it is a ×4 or ×8 magnification factor, the proposed model compares with the other models because of the effective

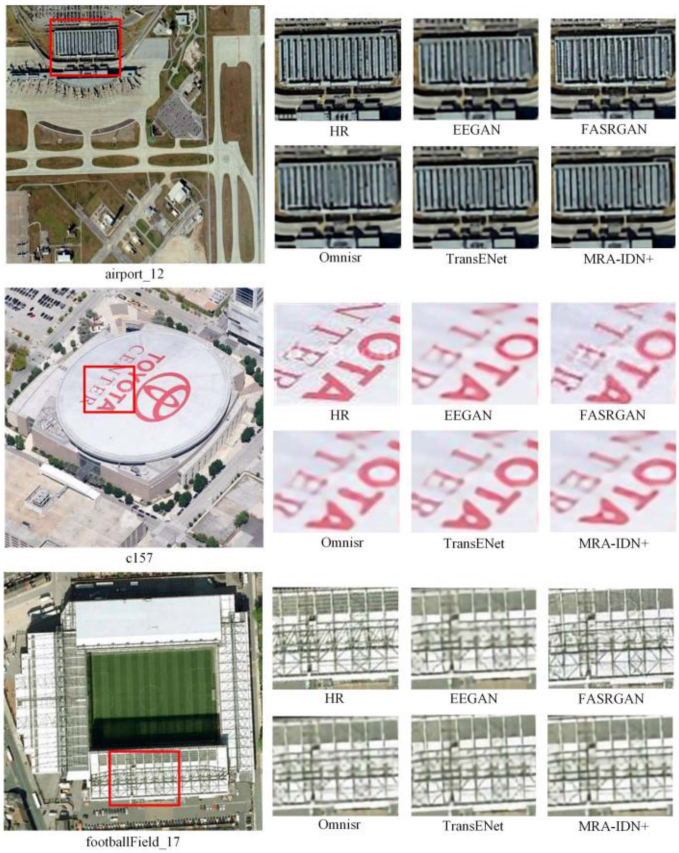


Fig. 16. Reconstructed image details obtained with our model and existing algorithms based on GAN and transformer at a $\times 4$ amplification factor.

extraction of features by the multiscale feature extraction module and the HF attention module, which makes the reconstructed concrete lines and structures clearer, the details of the edges of the windows and the walls more accurate and natural, and the details of text and objects on the roof of the building are richer and more obvious.

In order to compare the performance with more super-resolution algorithms, the proposed algorithm based on CNN compares the reconstruction effect with the latest GAN-based and transformer-based super-resolution algorithms, such as EEGAN, FASRGAN [70], Omnisr, and TransENet [71]. Since the number of network parameter, the occupied memory, and the running time of the CNN-based network are generally much less than those of the GAN and the transformer, only the reconstruction effect is compared here. Fig. 16 shows that, in image airport_12, the clarity of the lines reconstructed by the proposed algorithm are all better than other methods, except FASRGAN, but the choreography of the lines is closer to the original image compared to FASRGAN. The line choreography is uniform and more in line with the neatness of the building. In image c157, each letter is clearer than the other algorithms. In image footballField_17, the lines reconstructed by the proposed algorithm are better than the other algorithms, except for FASRGAN, but the color is closer to the original image than FASRGAN.

In this article, a lightweight super-resolution network is designed for remote sensing images that have complex and diverse feature information, but at the same time achieve super-resolution on real-world scenes. Table XII shows that the proposed algorithm achieves lightweight on remote sensing image super-resolution while also achieving the optimal performance, so only the comparison with LRSISR model is made on the real-world scene. It can be concluded from Table XI, the proposed algorithm achieves the highest PSNR and SSIM values on Set5 [72], Set14 [73], BSD100 [74], Urban [75], and Manga109 [76] datasets. Therefore, the proposed algorithm has wider scene applicability and better performance. Under a $\times 4$ amplification factor, HAUNet_S exceeds the maximum memory (20G) when testing Urban100 and Manga109 datasets, and the experimental environment does not support multicard testing, resulting in the failure to test these two datasets, so there is no information about the PSNR and SSIM values of these two datasets tested by HAUNet_S here.

Table XII summarizes the average PSNR and SSIM of each model corresponding to each category reconstruction for the WHU-RS19 and RSSCN7 test sets under different amplification factors. The algorithmic models in Table XII are categorized, which in turn makes it more convenient to compare the performance of the models with respect to different categories. All the GLSSR and LRSISR models are divided into four categories, based on whether the network structure uses the attention mechanism, and HF and LF features. These categories are as follows.

1) None. The network structure uses neither the attention mechanism nor HF and LF features.
2) Attention-based. The network structure uses only the attention mechanism.
3) Features-based. The network structure uses only the HF and LF features.
4) Mixing-based. The network structure uses the attention mechanisms as well as the HF and LF features.

Table XII shows that the remote sensing images reconstructed by our method have the optimal quality under multiple magnification factors. The attention mechanism, and HF and LF features are effective when applied to the network structure, which considerably improves the performance of the proposed model.

It can be observed from Table XII that under a $\times 8$ amplification factor, the MHAN method has the best performance out of the existing GLSSR algorithms, outperforming the SAN method with respect to the overall reconstruction performance. On the WHU-RS19 dataset, the PSNR only improves by 0.08 dB, but the SSIM does not improve. On the RSSCN7 dataset, the PSNR and SSIM only improve by 0.08 dB and 0.004, respectively. Compared with the MHAN algorithm, the proposed algorithm improves the PSNR and SSIM by 0.36 dB and 0.026, respectively, on the WHU-RS19 dataset, and by 0.23 dB and 0.016, respectively, on the RSSCN7 dataset. Moreover, compared with all the LRSISR algorithms in Table XII, the proposed algorithm improves the PSNR and SSIM by 0.42 dB and 0.019 on average, respectively, on the WHU-RS19 dataset, and by 0.22 dB and 0.013 on average, respectively, on the RSSCN7 dataset. The latest LRSISR algorithm, HAUNet_S, compared

TABLE XI
QUANTITATIVE COMPARISON FOR $\times 4$ AND $\times 8$ SR ON BENCHMARK DATASETS

Scale factor	Method	Set5	Set14	BSD100	Urban100	Manga109
		PSNR/SSIM	PSNR/SSIM	PSNR/SSIM	PSNR/SSIM	PSNR/SSIM
$\times$ 4	IMDN	31.48/0.885	28.19/0.772	27.38/0.729	25.66/0.772	29.12/0.886
	PAN	31.30/0.883	28.26/0.774	27.37/0.729	25.58/0.769	29.29/0.892
	LESRCNN	31.39/0.884	29.22/0.745	27.33/0.727	25.54/0.767	29.28/0.891
	ACNECT	31.43/0.884	28.23/0.773	27.34/0.727	25.58/0.768	29.32/0.891
	AAN	31.46/0.885	28.25/0.775	27.42/0.731	25.76/0.775	29.58/0.895
	FeNet	29.36/0.831	26.69/0.732	26.42/0.695	23.64/0.682	25.74/0.805
	Omnisr	31.48/0.884	28.18/0.772	27.38/0.729	25.57/0.767	29.30/0.891
	HAUNet_S	30.26/0.881	27.88/0.765	27.25/0.726	-	-
	MRA-IDN +	**31.65/0.888**	**28.34/0.776**	**27.48/0.733**	**25.94/0.781**	**29.64/0.896**
$\times$ 8	IMDN	26.49/0.753	24.68/0.627	24.67/0.589	22.20/0.600	23.62/0.741
	PAN	26.44/0.755	24.71/0.629	24.68/0.589	22.20/0.601	23.81/0.751
	LESRCNN	26.41/0.752	24.61/0.625	24.63/0.587	22.12/0.596	23.67/0.745
	ACNECT	26.32/0.750	24.59/0.626	24.63/0.587	22.15/0.598	23.69/0.746
	AAN	26.47/0.755	24.71/0.629	24.72/0.590	22.29/0.606	23.91/0.756
	FeNet	26.17/0.739	24.44/0.618	24.52/0.582	21.92/0.583	23.16/0.720
	Omnisr	26.22/0.741	24.46/0.620	24.53/0.584	21.99/0.586	23.30/0.725
	HAUNet_S	26.36/0.751	24.52/0.621	24.64/0.587	22.12/0.593	23.29/0.728
	MRA-IDN +	**26.64/0.764**	**24.77/0.631**	**24.74/0.592**	**22.35/0.611**	**23.91/0.757**

TABLE XII
PSNR AND SSIM METRICS OF THE RECONSTRUCTED WHU-RS19 AND RSCCN7 TEST SETS FOR EACH MODEL

Model	Type	Method	Scale	WHU-RS19	RSCCN7	Scale	WHU-RS19	RSCCN7
				PSNR/SSIM	PSNR/SSIM		PSNR/SSIM	PSNR/SSIM
GLSSR	None	RDN	$\times$ 4	31.20/0.825	29.10/0.741	$\times$ 8	26.95/0.643	26.27/0.584
	None	D-DBPN	$\times$ 4	31.36/0.826	29.23/0.747	$\times$ 8	26.97/0.642	26.30/0.585
	None	SRFBN	$\times$ 4	31.35/0.824	29.22/0.745	$\times$ 8	26.93/0.641	26.24/0.581
	Attention-based	RCAN	$\times$ 4	31.32/0.822	29.25/0.747	$\times$ 8	27.00/0.644	26.32/0.587
	Attention-based	SAN	$\times$ 4	31.38/0.827	29.26/0.749	$\times$ 8	26.95/0.642	26.26/0.584
	Attention-based	MHAN	$\times$ 4	31.40/**0.828**	**29.28/0.750**	$\times$ 8	27.03/0.642	26.34/0.588
LRSISR	None	SRCNN	$\times$ 4	30.06/0.783	28.41/0.710	$\times$ 8	26.33/0.617	25.86/0.565
	None	VDSR	$\times$ 4	30.74/0.807	28.85/0.731	$\times$ 8	26.74/0.635	26.15/0.580
	Attention-based	IMDN	$\times$ 4	31.27/0.821	29.16/0.744	$\times$ 8	27.28/0.660	26.51/0.599
	Attention-based	PAN	$\times$ 4	31.18/0.820	28.98/0.742	$\times$ 8	27.24/0.659	26.49/0.599
	Attention-based	AAN	$\times$ 4	31.29/0.823	29.18/0.745	$\times$ 8	27.30/0.661	26.53/0.600
	Attention-based	FeNet	$\times$ 4	31.01/0.816	29.05/0.740	$\times$ 8	27.01/0.650	26.39/0.593
	Attention-based	Omnisr	$\times$ 4	30.76/0.818	29.03/0.741	$\times$ 8	26.70/0.651	26.32/0.593
	Feature-based	ACNECT	$\times$ 4	31.19/0.819	29.12/0.742	$\times$ 8	27.18/0.656	26.45/0.596
	Feature-based	LESRCNN	$\times$ 4	31.15/0.818	29.11/0.741	$\times$ 8	27.13/0.656	26.46/0.596
	Mix-based	HAUNet_S	$\times$ 4	30.49/0.806	29.09/0.741	$\times$ 8	26.83/0.642	26.39/0.586
	Mix-based	**MRA-IDN**	$\times$ 4	31.35/0.824	29.21/0.746	$\times$ 8	**27.36/0.665**	**26.55/0.603**
	Mix-based	**MRA-IDN +**	$\times$ 4	**31.59**/0.827	29.25/**0.750**	$\times$ 8	**27.39/0.668**	**26.57/0.604**

with all the LRSISR algorithms, except the proposed algorithm, reduces the PSNR and SSIM by 0.16 dB and 0.007 on average, respectively, on the WHU-RS19 dataset, and the PSNR is higher by 0.04 dB on average, and SSIM is less by 0.005 on average, on the RSSCN7 dataset. The second-best-performing LRSISR algorithm, AAN, compared with all the LRSISR algorithms, except the proposed algorithm, improves the PSNR and SSIM by 0.36 dB and 0.013 on average, respectively, on the WHU-RS19 dataset, and by 0.19 dB and 0.010 on average, respectively, on the RSSCN7 dataset. It can be concluded that the proposed algorithm's PSNR and SSIM are improved by a large amount, and the performance, compared with other algorithms, is improved by a lot, while achieving the purpose of lightweight.

E. Discussion

The deep-learning-based super-resolution algorithm obtains the nonlinear relationship between HR and LR images through the continuous learning of the neural network, and achieves a

good quality of image reconstruction. The HF and LF details of the image can still be recovered well when the scaling factor is large. We propose an MSR-ID module, which fully extracts the features while reducing the number of parameters, and an HFCA module, which makes the complex texture and HF details fully preserved. Therefore, the MRA-IDN algorithm, compared with the GLSSR algorithm, achieves high visual reconstruction effect, while realizing lower FLOPs, number of parameters, occupied GPU memory, and actual inference time, such as the building lines reconstructed are clearer and the edges are more accurate. Compared with some evaluation metrics of the LRSISR algorithms, such as the number of parameters of FeNet and PAN, although the number of parameters is not the lowest, it has the highest PSNR value and better visual effect. The proposed algorithm will be continued to be improved to ensure high visual reconstruction effect while greatly reducing the number of network parameters. At $\times 4$ magnification factor, the proposed algorithm achieves optimal PSNR and SSIM values in most scenes and suboptimal values in the remaining few scenes. At $\times 8$ magnification factor, it achieves the optimal values in all scenes, except for the suboptimal values in the Desert scene. It can be seen that the proposed model can stably achieve good reconstruction visual effect in various remote sensing scenes, and will not reduce the reconstruction effect because of the scene change, which indicates that the model's performance is stable. It also reflects that the proposed algorithm can realize the highest visual reconstruction effect for the image super-resolution with high magnification factor in the majority of scenes, and can continue to improve the image super-resolution effect with low magnification factor in individual scenes.

V. Conclusion

This article presented a new, efficient, lightweight, and high-performance super-resolution reconstruction algorithm for remote sensing images. First, an MSR-ID consisting of multiple RCBs was designed. Second, a novel HFCA attention mechanism suitable for the super-resolution reconstruction task was proposed. These two mechanisms constituted an MRA-IDG with a small number of parameters, low complexity, and strong characterization ability. Last, a super-resolution scheme (MRA-IDN) combining frequency separation reconstruction, where the LF and HF components are restored by bilinear interpolation and a DNN, respectively, was proposed based on the backbone of a set of MRA-IDG modules. The MSR-ID block could extract multiscale residual features containing a higher amount of abundant information with different receptive fields under a lower number of network parameters. The HFCA attention block could guide the network to be biased toward the extraction and reconstruction of HF information. The HF and LF separation reconstruction scheme could improve the reconstruction performance of the HF component while keeping the number of model parameters low. The experimental results indicated that the proposed method, MRA-IDN, could reconstruct clearer images and more accurate HF detail information for complex remote sensing image scenes. Compared with the MHAN, the number of parameters of MRA-IDN could be reduced by 75%. It also showed advantages in terms of memory, running time, and GPU cost, and the subjective visual performance and objective quality of the reconstructed HR images were improved to some extent. However, due to the single degradation model, such as the interpolation used in the training process, the proposed scheme and other existing methods suffer from limitations for complex real-life scenes. Therefore, it is necessary to discuss and study different degradation models in the field of remote sensing to further optimize the proposed algorithm and obtain more complex and diversified remote sensing image data to train it, while realizing its application expansion in blind super-resolution. Moreover, the proposed algorithm is trained based on optical images, which is applicable to a limited number of remote sensing image types. So it can be subsequently trained based on more remote sensing image types to realize super-resolution for more image types, such as hyperspectral image super-resolution [77], [78], [79].

References

[1] H. Su, J. Zhou, and Z. Zhang, "A review of super-resolution image reconstruction methods," *Acta Automatica Sinica*, vol. 39, pp. 1202–1213, 2013.

[2] V. K. Ha, J. Ren, X. Xu, S. Zhao, G. Xie, and V. M. Varges, "Deep learning based single image super-resolution: A survey," in *Proc. 9th Int. Conf. Adv. Brain Inspired Cogn. Syst.*, 2018, pp. 106–119.

[3] C. E. Duchon, "Lanczos filtering in one and two dimensions," *J. Appl. Meteorol. Climatol.*, vol. 18, pp. 1016–1022, 1979.

[4] R. Keys, "Cubic convolution interpolation for digital image processing," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 29, no. 6, pp. 1153–1160, Dec. 1981, doi: 10.1109/TASSP.1981.1163711.

[5] H. Chang, D.-Y. Yeung, and Y. Xiong, "Super-resolution through neighbor embedding," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, 2004, pp. I–I.

[6] J. Sun, Z. Xu, and H. Y. Shum, "Image super-resolution using gradient profile prior," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2008, pp. 1–8.

[7] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Commun. Assoc. Comput. Mach.*, vol. 60, no. 6, pp. 84–90, 2017, doi: 10.1145/3065386.

[8] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 770–778.

[9] K. He, X. Zhang, S. Ren, and J. Sun, "Identity mappings in deep residual networks," in *Proc. Eur. Conf. Comput. Vis.*, 2016, pp. 630–645.

[10] G. Huang, Z. Liu, L. V. D. Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 4700–4708.

[11] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 7132–7141.

[12] J. Wang, W. Li, M. Zhang, R. Tao, and J. Chanussot, "Remote-sensing scene classification via multistage self-guided separation network," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, pp. 1–12, Jul. 2023, doi: 10.1109/TGRS.2023.3295797.

[13] J. Wang, W. Li, M. Zhang, and J. Chanussot, "Large kernel sparse convnet weighted by multi-frequency attention for remote sensing scene understanding," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, pp. 1–12, Nov. 2023, doi: 10.1109/TGRS.2023.3333401.

[14] M. Zhang, W. Li, X. Zhao, H. Liu, R. Tao, and Q. Du, "Morphological transformation and spatial-logical aggregation for tree species classification using hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, pp. 1–12, Jan. 2023, doi: 10.1109/TGRS.2022.3233847.

[15] J. Wang, W. Li, Y. Gao, M. Zhang, R. Tao, and Q. Du, "Hyperspectral and SAR image classification via multiscale interactive fusion network," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 34, no. 12, pp. 10823–10837, Dec. 2023, doi: 10.1109/TNNLS.2022.3171572.

[16] C. Dong, C. C. Loy, K. He, and X. Tang, "Image super-resolution using deep convolutional networks," *IEEE Trans. Pattern Anal. Mach.*

Intell., vol. 38, no. 2, pp. 295–307, Feb. 2016, doi: 10.1109/TPAMI.2015.2439281.
[17] J. Kim, J. K. Lee, and K. M. Lee, "Accurate image super-resolution using very deep convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 1646–1654.
[18] X. J. Mao, C. Shen, and Y. B. Yang, "Image restoration using convolutional auto-encoders with symmetric skip connections," 2016, *arXiv:1606.08921*.
[19] W. Han, S. Chang, D. Liu, M. Yu, M. Witbrock, and T. S. Huang, "Image super-resolution via dual-state recurrent networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 1654–1663.
[20] C. Ledig et al., "Photo-realistic single image super-resolution using a generative adversarial network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 4681–4690.
[21] X. Wang et al., "ESRGAN: Enhanced super-resolution generative adversarial networks," in *Proc. Eur. Conf. Comput. Vis.*, 2018, pp. 63–79.
[22] N. C. Rakotonirina and A. Rasoanaivo, "ESRGAN+: Further improving enhanced super-resolution generative adversarial network," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, 2020, pp. 3637–3641.
[23] A. Vaswani et al., "Attention is all you need," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, pp. 5998–6008.
[24] J. Cao, Y. Li, K. Zhang, and L. V. Gool, "Video super-resolution transformer," 2021, *arXiv:2106. 06847v1*.
[25] S. Chen, L. Zhang, and L. Zhang, "MSDformer: Multiscale deformable transformer for hyperspectral image super-resolution," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, pp. 1–14, Sep. 2023, doi: 10.1109/TGRS.2023.3315970.
[26] P. Dhariwal and A. Nichol, "Diffusion models beat GANs on image synthesis," *Adv. Neural Inf. Process. Syst.*, vol. 34, pp. 8780–8794, 2021.
[27] A. Q. Nichol and P. Dhariwal, " Improved denoising diffusion probabilistic models," in *Proc. PMLR Int. Conf. Mach. Learn.*, 2022, pp. 8162–8171.
[28] J. Liu, Z. Yuan, Z. Pan, Y. Fu, L. Liu, and B. Lu, "Diffusion model with detail complement for super-resolution of remote sensing," *Remote Sens.*, vol. 14, 2022, Art. no. 4834.
[29] L. Han et al., " Enhancing remote sensing image super-resolution with efficient hybrid conditional diffusion model," *Remote Sens.*, vol. 15, 2023, Art. no. 3452.
[30] H. Wu, N. Ni, and L. Zhang, "Learning dynamic scale awareness and global implicit functions for continuous-scale super-resolution of remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, pp. 1–15, Jan. 2023, doi: 10.1109/TGRS.2023.3240254.
[31] D. Mishra and O. Hadar, "Self-FuseNet: Data free unsupervised remote sensing image super-resolution," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 16, pp. 1710–1727, Jan. 2023, doi: 10.1109/JSTARS.2023.3239758.
[32] Y. Xiao et al., "From degrade to upgrade: Learning a self-supervised degradation guided adaptive network for blind remote sensing image super-resolution," *Inf. Fusion*, vol. 96, pp. 297–311, 2023.
[33] J. Kim, J. K. Lee, and K. M. Lee, "Deeply-recursive convolutional network for image super-resolution," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 1637–1645.
[34] Y. Zhang, K. Li, K. Li, L. Wang, B. Zhong, and Y. Fu, "Image super-resolution using very deep residual channel attention networks," in *Proc. Eur. Conf. Comput. Vis.*, 2018, pp. 286–301.
[35] T. Dai, J. Cai, Y. Zhang, S. T. Xia, and L. Zhang, "Second-order attention network for single image super-resolution," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 11065–11074.
[36] F. Yang, H. Yang, J. Fu, H. Lu, and B. Guo, "Learning texture transformer network for image super-resolution," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 5791–5800.
[37] D. Dai and Y. Wen, "Satellite image classification via two-layer sparse coding with biased image representation," *IEEE Geosci. Remote Sens. Lett.*, vol. 8, no. 1, pp. 173–176, Jan. 2011, doi: 10.1109/LGRS.2010.2055033.
[38] Q. Zou, L. 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, Nov. 2015, doi: 10.1109/LGRS.2015.2475299.
[39] M. Hara and K. Tanaka, "Accurate image super-resolution using very deep convolutional networks," *Public Health*, vol. 60, pp. 444–452, 2016.
[40] W. Shi et al., "Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 1874–1883.
[41] C. Dong, C. C. Loy, and X. Tang, "Accelerating the super-resolution convolutional neural network," in *Proc. Eur. Conf. Comput. Vis.*, 2016, pp. 391–407.
[42] L. C. Duta, L. Liu, F. Zhu, and L. Shao, "Pyramidal convolution: Rethinking convolutional neural networks for visual recognition," 2020, *arXiv:2006.11538*.
[43] C. Szegedy et al., "Going deeper with convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2015, pp. 1–9.
[44] K. Jiang, Z. Wang, P. Yi, J. Jiang, J. Xiao, and Y. Yao, "Deep distillation recursive network for remote sensing imagery super-resolution," *Remote Sens.*, vol. 10, no. 11, 2018, Art. no. 1700.
[45] T. Lu, J. Wang, Y. Zhang, Z. Wang, and J. Jiang, "Satellite image super-resolution via multi-scale residual deep neural network," *Remote Sens.*, vol. 11, no. 13, 2019, Art. no. 1588, doi: 10.3390/rs11131588.
[46] X. Wang, Y. Wu, Y. Ming, and H. Lv, "Remote sensing imagery super resolution based on adaptive multi-scale feature fusion network," *Sensors*, vol. 20, no. 4, 2020, Art. no. 1142, doi: 10.3390/s20041142.
[47] J. Zhang, S. Liu, Y. Peng, and J. Li, "Satellite image super-resolution based on progressive residual deep neural network," *J. Appl. Remote Sens.*, vol. 14, no. 3, 2020, Art. no. 032610, doi: 10.1117/1.jrs.14.032610.
[48] Q. Wang, B. Wu, P. Zhu, P. Li, W. Zuo, and Q. Hu, "ECA-Net: Efficient channel attention for deep convolutional neural networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 11534–11542.
[49] C. Dong, C. C. Loy, K. He, and X. Tang, "Learning a deep convolutional network for image super-resolution," in *Proc. Eur. Conf. Comput. Vis.*, 2014, pp. 184–199.
[50] B. Niu et al., "Single image super-resolution via a holistic attention network," in *Proc. Eur. Conf. Comput. Vis.*, 2020, pp. 191–207.
[51] B. Huang, B. He, L. Wu, and Z. Guo, "Deep residual dual-attention network for superresolution reconstruction of remote sensing images," *Remote Sens.*, vol. 13, no. 14, 2021, Art. no. 2784, doi: 10.3390/rs13142784.
[52] D. Zhang, J. Shao, X. 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, Jun. 2021, doi: 10.1109/TGRS. 2020.3009918.
[53] K. Jiang, Z. Wang, P. Yi, G. Wang, T. Lu, and J. Jiang, "Edge-enhanced GAN for remote sensing image superresolution," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 8, pp. 5799–5812, Aug. 2019, doi: 10.1109/TGRS.2019.2902431.
[54] Y. Zhang, Y. Tian, Y. Kong, B. Zhong, and Y. Fu, "Residual dense network for image super-resolution," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 2472–2481.
[55] C. Tian et al., "Lightweight image super-resolution with enhanced CNN," *Knowl. Syst.*, vol. 205, 2020, Art. no. 106235, doi: 10.1016/j.knosys.2020.106235.
[56] Y. Peng, X. Wang, J. Zhang, and S. Liu, "Pre-training of gated convolution neural network for remote sensing image super-resolution," *Int. Eng. Technol. Image Process.*, vol. 15, no. 5, pp. 1179–1188, 2021, doi: 10.1049/ipr2.12096.
[57] C. Tian, Y. Xu, W. Zuo, C.-W. Lin, and D. Zhang, "Asymmetric CNN for image superresolution," *IEEE Trans. Syst. Man Cybern. Syst.*, vol. 52, no. 6, pp. 3718–3730, Jun. 2022, doi: 10.1109/TSMC.2021.3069265.
[58] H. Zhao, X. Kong, J. He, Y. Qiao, and C. Dong, "Efficient image super-resolution using pixel attention," in *Proc. Eur. Conf. Comput. Vis.*, 2020, pp. 56–72.
[59] H. Chen, J. Gu, and Z. Zhang, "Attention in attention network for image super-resolution," 2021, *arXiv:2104.09497*.
[60] Z. Hui, X. Gao, Y. Yang, and X. Wang, "Lightweight image super-resolution with information multi-distillation network," in *Proc. Assoc. Comput. Mach. Int. Conf. Multimedia*, 2019, pp. 2024–2032.
[61] Y. Liu, Q. Jia, X. Fan, S. Wang, S. Ma, and W. Gao, "Cross-SRN: Structure-preserving super-resolution network with cross convolution," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 32, no. 8, pp. 4927–4939, Aug. 2022, doi: 10.1109/TCSVT.2021.3138431.
[62] C. Wu et al., "Full-reference image quality assessment via low-level and high-level feature fusion," *Int. J. Pattern Recognit. Artif. Intell.*, vol. 37, no. 11, 2023, Art. no. 2354016.
[63] R. Jing, F. Duan, F. Lu, M. Zhang, and W. Zhao, "Denoising diffusion probabilistic feature-based network for cloud removal in sentinel-2 imagery," *Remote Sens.*, vol. 15, no. 9, 2023, Art. no. 2217, doi: 10.3390/rs15092217.
[64] G.-S. Xia et al., "AID: A benchmark data set for performance evaluation of aerial scene classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 7, pp. 3965–3981, Jul. 2017, doi: 10.1109/TGRS.2017.2685945.
[65] M. Haris, G. Shakhnarovich, and N. Ukita, "Deep back-projection networks for super-resolution," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 1664–1673.

[66] Z. Li, J. Yang, Z. Liu, X. Yang, G. Jeon, and W. Wu, "Feedback network for image super-resolution," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 3867–3876.
[67] Z. Wang et al., "FeNet: Feature enhancement network for lightweight remote-sensing image super-resolution," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–12, Apr. 2023, doi: 10.1109/TGRS.2022.3168787.
[68] H. Wang, X. Chen, B. Ni, Y. Liu, and J. Liu, "Omni aggregation networks for lightweight image super-resolution," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2023, pp. 22378–22387.
[69] J. Wang, B. Wang, X. Wang, Y. Zhao, and T. Long, "Hybrid attention-based U-shaped network for remote sensing image super-resolution," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, pp. 1–15, Jun. 2023, doi: 10.1109/TGRS.2023.3283769.
[70] Y. Yan et al., "Fine-grained attention and feature-sharing generative adversarial networks for single image super-resolution," *IEEE Trans. Multimedia*, vol. 24, pp. 1473–1487, Mar. 2022, doi: 10.1109/TMM.2021.3065731.
[71] S. Lei, Z. Shi, and W. Mo, "Transformer-based multistage enhancement for remote sensing image super-resolution," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–11, Dec. 2023, doi: 10.1109/TGRS.2021.3136190.
[72] M. Bevilacqua et al., "Low-complexity single-image super-resolution based on nonnegative neighbor embedding," in *Proc. Brit. Mach. Vis. Conf.*, 2012, pp. 1–10.
[73] R. Zeyde, M. Elad, and M. Protter, "On single image scale-up using sparse-representations," in *Proc. Int. Conf. Curves Surfaces*, 2012, pp. 711–730.
[74] D. Martin, C. Fowlkes, D. Tal, and J. Malik, "A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics," in *Proc. 8th IEEE Int. Conf. Comput. Vis.*, 2001, pp. 416–423.
[75] J. Huang, A. Singh, and N. Ahuja, "Single image super-resolution from transformed self-exemplars," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2015, pp. 5197–5206.
[76] Y. Matsui et al., "Sketch-based manga retrieval using manga109 dataset," *Multimedia Tools Appl.*, vol. 76, pp. 21811–21838, 2017.
[77] Q. Li, M. Gong, Y. Yuan, and Q. Wang, "Symmetrical feature propagation network for hyperspectral image super-resolution," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–12, Sep. 2023, doi: 10.1109/TGRS.2022.3203749.
[78] Q. Li, M. Gong, Y. Yuan, and Q. Wang, "RGB-induced feature modulation network for hyperspectral image super-resolution," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, pp. 1–11, May 2023, doi: 10.1109/TGRS.2023.3277486.
[79] Q. Li, Y. Yuan, X. Jia, and Q. Wang, "Dual-stage approach toward hyperspectral image super-resolution," *IEEE Trans. Image Process.*, vol. 31, pp. 7252–7263, Nov. 2022, doi: 10.1109/TIP.2022.3221287.

Wujian Ye received the B.S. degree in computer science and technology from the School of Computers, Guangdong University of Technology, Guangzhou, China, in 2010, and the M.S. and Ph.D. degrees in computer science from the Dankook University, Yongin, South Korea, in 2012 and 2015, respectively.

Since 2016, he has been a Lecturer with the School of Integrated Circuits, Guangdong University of Technology. His research interests include deep learning and computer vision, machine-learning application, computer network and security analysis, and voice recognition.

Bili Lin received the B.S. degree in information engineering in 2022 from the School of Information Engineering, Guangdong University of Technology, Guangzhou, China, where she is currently working toward the master's degree in electronic information.

Her research interests include AI algorithm, deep learning, and super-resolution of remote sensing image.

Junming Lao received the B.S. degree in information engineering from the School of Information Engineering, Wuyi University, Jiangmen, China, in 2020, and the M.S. degree in information engineering from the School of Information Engineering, Guangdong University of Technology, Guangzhou, China, in 2023.

His research interests include AI algorithm, deep learning, and super-resolution of remote sensing image.

Yijun Liu received the B.S. degree from the Beijing Normal University, Beijing, China, in 1999, the M.Sc. degree from the Guangdong University of Technology, Guangzhou, China, in 2002, the M.Phil. degree, from University of Manchester, Manchester, U.K., in 2003, and the Ph.D. degree from the University of Manchester, Manchester, U.K., in 2005, all in computer science.

He is currently a Full Professor with the School of Integrated Circuits, Guangdong University of Technology, Guangzhou, China. His research interests include neuromorphic computing, deep learning, computer architecture, and GPS/Beidou navigation.

Zhenyi Lin received the B.S. degree in electronic information science and technology from the School of Electronic Information and Electrical Engineering, Huizhou University, Huizhou, China, in 2021. He is currently working toward the master's degree in electronic information with the Guangdong University of Technology, Guangzhou, China.

His research interests include deep learning and computer vision.