remote sensing

MDPI

Article

Super-Resolution Network for Remote Sensing Images via Preclassification and Deep–Shallow Features Fusion

Xiuchao Yue †, Xiaoxuan Chen †, Wanxu Zhang, Hang Ma, Lin Wang, Jiayang Zhang, Mengwei Wang and Bo Jiang *

School of Information Science and Technology, Northwest University, Xi'an 710127, China; yuexiuchao@stumail.nwu.edu.cn (X.Y.) chenxx@nwu.edu.cn (X.C.); wxzhang@nwu.edu.cn (W.Z.); mahang1@stumail.nwu.edu.cn (H.M.); wanglin@nwu.edu.cn (L.W.); zhangjiayang@stumail.nwu.edu.cn (J.Z.); 202021257@stumail.nwu.edu.cn (M.W.)

* Correspondence: jiangbo@nwu.edu.cn

† These authors contributed equally to this work.

Abstract: A novel super-resolution (SR) method is proposed in this paper to reconstruct high-resolution (HR) remote sensing images. Different scenes of remote sensing images have great disparities in structural complexity. Nevertheless, most existing SR methods ignore these differences, which increases the difficulty to train an SR network. Therefore, we first propose a preclassification strategy and adopt different SR networks to process the remote sensing images with different structural complexity. Furthermore, the main edge of low-resolution images are extracted as the shallow features and fused with the deep features extracted by the network to solve the blurry edge problem in remote sensing images. Finally, an edge loss function and a cycle consistent loss function are added to guide the training process to keep the edge details and main structures in a reconstructed image. A large number of comparative experiments on two typical remote sensing images datasets (WHURS and AID) illustrate that our approach achieves better performance than state-of-the-art approaches in both quantitative indicators and visual qualities. The peak signal-to-noise ratio (PSNR) value and the structural similarity (SSIM) value using the proposed method are improved by 0.5353 dB and 0.0262, respectively, over the average values of five typical deep learning methods on the ×4 AID testing set. Our method obtains satisfactory reconstructed images for the subsequent applications based on HR remote sensing images.

Keywords: remote sensing image; image super-resolution; convolutional neural network

check for updates

Citation: Yue, X.; Chen, X.; Zhang, W.; Ma, H.; Wang, L.; Zhang, J.; Wang, M.; Jiang, B. Super-Resolution Network for Remote Sensing Images via Preclassification and Deep–Shallow Features Fusion. *Remote Sens.* **2022**, *14*, 925. https://doi.org/10.3390/rs14040925

Academic Editors: Tao Lei, Tao Chen, Lefei Zhang and Asoke K. Nandi

Received: 27 December 2021
Accepted: 11 February 2022
Published: 14 February 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

1. Introduction

Because remote sensing images are obtained with long optical paths, one pixel in a remote sensing image generally corresponds to a size of several square meters on the ground. As a result, the remote sensing images generally are low-resolution (LR), which brings a lot of inconvenience to the later advanced processing, e.g., object detection [1,2] and semantic segmentation [3,4]. Therefore, it is significant to apply super-resolution (SR) methods to improve the resolutions of remote sensing images. SR is a technology to recover high-resolution (HR) images from its degraded low-resolution counterpart with only software algorithms instead of changing the hardware equipment. At present, the SR research is mainly for natural images and these SR methods are not appropriate for remote sensing images [5].

The particularity of the problem studied in this paper is reflected in three aspects: Firstly, unlike most of the images on the near ground side, the ground sizes of remote sensing images are very large [6]. However, the SR reconstruction has a strong demand on the correlation information between neighboring pixels. Therefore, the SR methods for natural images will not obtain satisfactory effect when they are directly applied to remote sensing images. Secondly, remote sensing images contain diverse scenes with great differences, such as urban buildings, forests, mountains, oceans, etc. Images of different scenes contain different details, so it is difficult to design an SR network that can acquire a satisfactory effect for all scenes [7]. Thirdly, the imaging optical paths are very long for remote sensing images. There are many degradation factors on the whole imaging link, such as noise, deformation, and movement, which generally weaken contours and edges in remote sensing images. At the same time, the lack of shallow features at the end of deep network also causes the edges of the reconstructed images to be blurry. Consequently, the quality of remote sensing images is generally not high, and it is challenging to design a suitable SR method for remote sensing images with a variety of scenarios.

In this paper, we research the aforementioned challenges and propose an SR method based on preclassification and deep–shallow features fusion, which can effectively reconstruct remote sensing images of different scenes and enhance the structure information in SR images. The main contributions of this work are as follows:

- We first introduce the preclassification strategy to the remote sensing image SR task. More specifically, we divide remote sensing images into three classes according to the structural complexity of scenes. Deep networks with different complexity are used for different classes of remote sensing images. The training difficulty is reduced with the declining number of training samples for each class. In this way, each network can learn the commonness of images in one class, improve the network's adaptability, and achieve good reconstruction effects for remote sensing images of different scenes and different complexity classes.
- We design a fusion network using the deep features and shallow features to deal with the problem of weak edge structure in remote sensing images. On the one hand, the multi-kernel residual attention (MKRA) modules are deployed to effectively extract the deep features of an LR image and learn the detail differences of images by using the global residual method. On the other hand, considering that the deep network lacks shallow features at its end, the shallow features of original data are integrated into the deep features at the end of the network. In fact, we take the main edge as the shallow features to solve the problem of weak edge structure of remote sensing images, which can well recover image edges and texture details.
- An edge loss and a cycle consistent loss are added to guide the training process. To avoid the trouble of weight hyperparameter, we adopt the charbonnier loss as the normal form of the loss function. The total loss function not only calculates the overall difference and edge difference between the HR image and the reconstructed SR image, but also calculates the difference between the LR image and the downsampled SR image, so as to better use the LR remote sensing image to guide the training process of the SR network.

The rest of this paper is organized as follows: we briefly review the related works on SR in Section 2. The proposed SR method is introduced in detailed in Section 3. The evaluation experiments of different methods are conducted in Section 4, which includes the quantitative and qualitative evaluations. Finally, Section 5 contains the conclusion.

2. Related Work

Since the learning-based methods are more advantageous in the field of image SR, many learning-based SR methods have been proposed in recent years [8]. These methods can fit the complex image degradation process and establish the mapping relationship between HR and LR images. The learning-based methods can be further divided into machine learning methods and deep learning methods. Sparsity [9] is a kind of typical prior information in machine learning, which is prevalently applied in sparse coding-based approaches. Yang et al. [10] proposed an SR method based on sparse representation prior. It trains the extracted features of LR and HR image patches to obtain the dictionaries, then the HR image patches can be obtained using HR image patch dictionary and sparse coefficients in corresponding LR image patches.

However, most machine learning methods use the low-level features of images for SR reconstruction, and the level of ability to represent these features greatly limits the reconstruction effect that is achievable. In addition, the long optical path and complex imaging environment of remote sensing imaging make the image degradation mechanism complex, so the related mapping is difficult to be effectively learned by traditional machine learning methods. Deep learning technology has been a research hotspot in image processing recently, such as classification [11], detection [12], semantic segmentation [13], and so on. Given the advent of the widespread popularity of deep learning, methods based on convolutional neural network (CNN) become the mainstream of SR tasks.

The basic principle of CNN-based SR reconstruction methods is to train a neural network using a dataset that includes both HR images and their corresponding LR counterparts. Then, the network takes new LR images as the input and outputs SR images. The seminal work based on CNN architecture is super-resolution convolutional neural network (SRCNN) [14], which first proposed a three-layer convolutional network for image SR. Later, Kim et al. [15] introduced a deeper network named very deep super-resolution (VDSR) with 20 layers. An efficient sub-pixel convolution layer was proposed in efficient sub-pixel convolutional neural network (ESPCN) [16] to upscale the final LR feature maps into the HR output. Because residual learning [11] can alleviate the training difficulty, super-resolution residual network (SRResNet) [17] took advantage of residual learning to construct a deeper network and achieved better performance. By removing unnecessary modules in conventional residual networks, Lim et al. [18] proposed enhanced deep super-resolution (EDSR) and multi-scale deep super-resolution (MDSR) by removing the batch normalization layer in SRResNet, which achieved significant improvement. Benefiting from the study of attention mechanism, pixel attention network (PAN) [19] constructed a pretty concise and effective network with a newly proposed pixel attention scheme.

The abovementioned methods have already achieved good SR effects for most natural images. However, it may have many challenges to apply these methods to remote sensing images directly [20]. There are many differences between remote sensing images and natural ground images. Since remote sensing images are characterized by diverse scenes, rich texture features, and fuzzy structure contours, the difficulty of SR reconstruction is increased, and the spatial resolution of remote sensing images is the main limiting factor for the subsequent advanced applications of remote sensing. Concerning the SR of remote sensing images, researchers have also proposed some SR methods for remote sensing images based on deep CNNs. Lei et al. [21] proposed an algorithm named local–global combined network (LGCNet) to learn multilevel representations of remote sensing images. Deep residual squeeze and excitation network (DRSEN) [22] proposed residual squeeze and excitation block (RSEB) to extract the features of remote sensing images and improve the upsampling module and the global residual pathway.

However, most methods mix all types of remote sensing images, ignore the structural complexity of different types of remote sensing images and the characteristics of weak edge structure in remote sensing images, blindly increase the network complexity to improve the SR effect, and increase the difficulty of training. Therefore, there is still a large space to study the SR of remote sensing images of multiple scenes, which is the content of this paper.

3. Proposed Method

In this section, we describe the overall architecture and specific details of our method, including the preclassification strategy, the network design, and the loss functions. To have a better understanding of our work, we first give a brief introduction to the method. The overall scheme is shown in Figure 1. According to the structural complexity of the input remote sensing images, different SR networks are designed to reconstruct the corresponding remote sensing images, reducing the training samples and difficulty of each network.

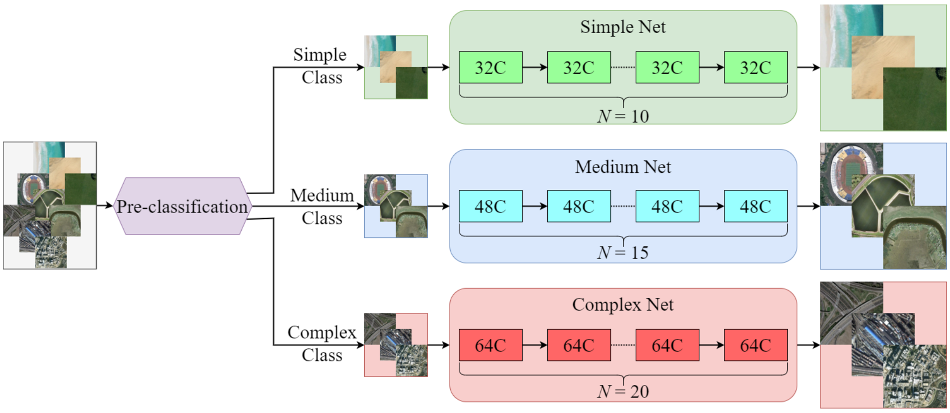


Figure 1. Overview of the proposed method. The input remote sensing LR images are first preclassified into three classes. Then, different SR nets are used to reconstruct SR images for each class.

3.1. Preclassification Strategy

Remote sensing images contain a variety of scene types, e.g., mountain, forest, city, ocean, desert, etc., whose scene structure complexity is very different. Existing SR networks process all kinds of remote sensing images without distinction, resulting in large training samples and difficulty to obtain excellent reconstruction results. In this paper, the idea of preclassification is innovatively highlighted. The datasets and networks are classified according to the different complexity of remote sensing images, to reduce the number of training samples and improve the effect of the SR network.

Since the complexity of remote sensing images is mainly reflected in image details such as edges, the average image gradients are used to measure the image complexity in this paper. With simplicity in mind, remote sensing images are divided into three classes according to their complexity, namely, the simple class, the medium class, and the complex class. Examples of each class and their gradient images are shown in Figure 2. Images in the simple class mostly contain monotonous and regular geographical areas, while images in the complex class contain a variety of ground objects. Remote sensing images of different complexity are designed to be processed by different networks, and the difference of networks is mainly reflected in the number of sub-modules, which does not increase the design complexity. Each network learns the commonality of the remote sensing images in each class, which gives the network good capability of reconstructing the congeneric images and reduces the overall training difficulty as well.

3.2. Deep–Shallow Features Fusion Network

Remote sensing images usually have the problem of weak edge details. When extracting image features, the deep network tends to weaken the shallow features, such as edges and contours, so we propose an SR network based on the fusion of deep and shallow features. Our network architecture is mainly divided into a deep feature branch and a shallow feature branch, as shown in Figure 3. The deep feature branch is used to extract deep features of remote sensing images, and the shallow feature branch integrates the shallow features of remote sensing images into the end of the network. The deep features and the shallow features are fused to generate the HR image.

Since the input image and the target image are highly correlated in the SR task, the global residual learning is adopted in the network, to reduce the complexity and learning difficulty of the network. To effectively extract the deep features of LR, we design the multi-kernel residual attention (MKRA) module. In general, when a network has deep layers, it can learn more complex representations, but it brings the increase of parameters number. Networks are built by controlling the number of MKRA and convolution filters to adapt to the different remote sensing images. Meanwhile, in view of the characteristics of weak edges in remote sensing images, we extract the main edge of LR images as shallow features to supervise the network to produce details. Specifically, the branch of shallow feature first smooths the image with L_0 gradient minimization [23] to reduce noise and secondary information, then extracts the edge and fuses it into the end reconstruction part to improve the edge reconstruction effects.

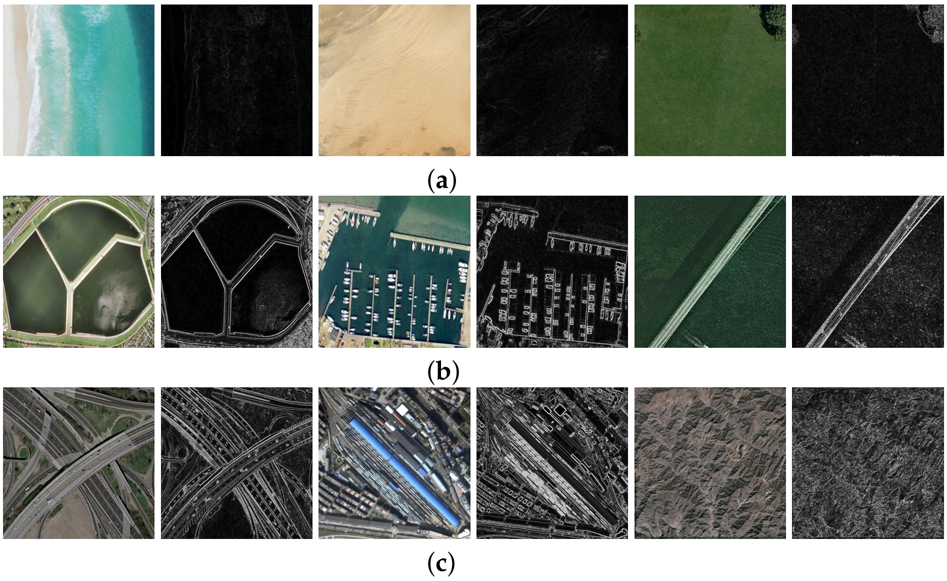


Figure 2. The examples of remote sensing images and their gradient images in three classes with different complexity. The images in columns 1, 3, and 5 are remote sensing images, and the images in columns 2, 4, and 6 are their corresponding gradient images. (**a**) Examples of the simple class; (**b**) Examples of the medium class; (**c**) Examples of the complex class.

In a word, the deep feature branch consists of convolution layers, multiple MKRA modules, and upsample parts. The shallow feature branch consists of an L_0 gradient minimization, a convolution layer, and an upsample part. Inspired by EDSR [18], sub-pixel convolution [16] is the upsampling part in the network.

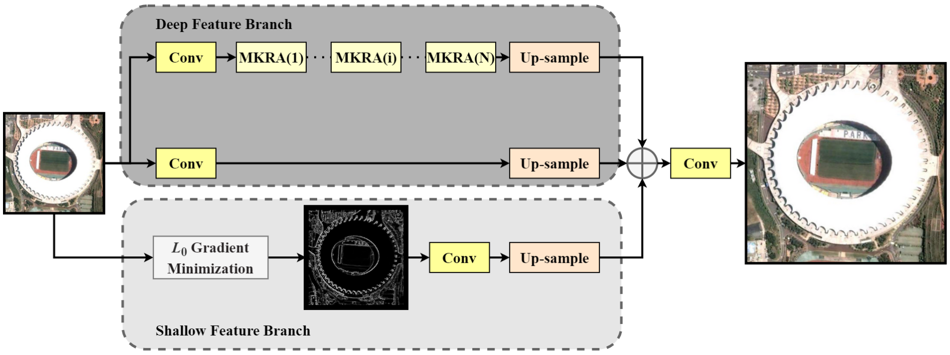


Figure 3. Network architecture of our method. The LR image is fused into an SR image by the deep feature branch and the shallow feature branch.

3.2.1. Multi-Kernel Residual Attention

For neural networks, higher-level feature extraction and data representation are all crucial [24]. Similarly, for the SR tasks of remote sensing images, stronger characterization ability is conducive to achieving better performance. The size of convolution kernel determines the way of feature extraction, so we adopt multi-kernel convolution for feature extraction to improve the richness of feature. Moreover, local residual connection and attention mechanism are adopted to further optimize the feature utilization capacity of the network. Each MKRA module shown in Figure 4 is composed of multi-kernel (MK) convolution sub-module, channel attention (CA) sub-module, and pixel attention (PA) sub-module.

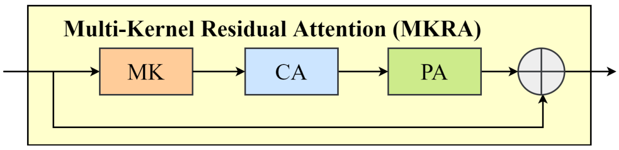


Figure 4. Structure of multi-kernel residual attention (MKRA).

Figure 5 demonstrates the detail of sub-module in MKRA. The feature maps are fused and activated by nonlinear function after four convolution kernels of different sizes (3 × 3, 1 × 3, 3 × 1, 1 × 1). To avoid size mismatches in the training process, the zero-padding approach is adopted to ensure the image size remains consistent during feature delivery. The local residual connection in the modules is beneficial to avoid the training instability and generalization ability caused by the deeper network.

Early CNN-based SR methods mainly focused on increasing the depth and width of the network, while features extracted from the network were treated equally in all channels and spatial regions. These methods lack the necessary flexibility for different feature mapping networks and waste computational resources in the task. The attention mechanism enables the network to pay more attention to the information features that are more useful to the target task, and suppress the useless features, so that the computing resources can be allocated more scientifically in the feature extraction process, to deepen the network effectively [25–27]. We use the cascade of channel attention and pixel attention to enhance the features and improve the learning ability of modules. Figure 5 and Table 1 report the MK, CA, PA design structure, and data flow in more detail.

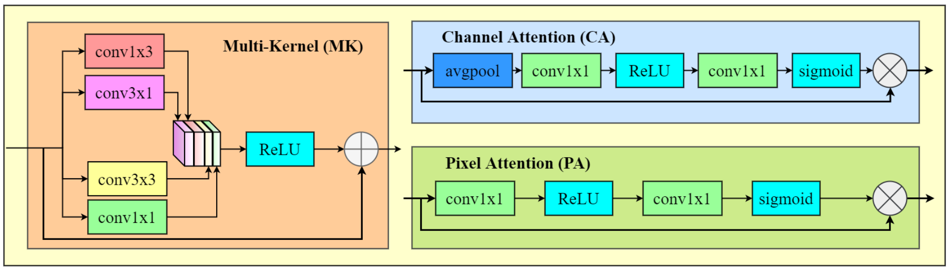


Figure 5. Structure of multi-kernel (MK), channel attention (CA), and pixel attention (PA) in MKRA.

Table 1. Network parameter settings of MKRA, where H and W denote the height and width of the feature map, and C denotes the channel.

Structure Component	Layer	Input	Output
MK module	conv1 × 3	$H \times W \times C$	$H \times W \times C/4$
	conv3 × 1	$H \times W \times C$	$H \times W \times C/4$
	conv3 × 3	$H \times W \times C$	$H \times W \times C/4$
	conv1 × 1	$H \times W \times C$	$H \times W \times C/4$
	ReLU	$H \times W \times C$	$H \times W \times C$
CA module	avgpool	$H \times W \times C$	$1 \times 1 \times C$
	conv1 × 1	$1 \times 1 \times C$	$1 \times 1 \times C/8$
	ReLU	$1 \times 1 \times C/8$	$1 \times 1 \times C/8$
	conv1 × 1	$1 \times 1 \times C/8$	$1 \times 1 \times C$
	sigmoid	$1 \times 1 \times C$	$1 \times 1 \times C$
	multiple	$H \times W \times C, 1 \times 1 \times C$	$H \times W \times C$
PA module	conv1 × 1	$H \times W \times C$	$H \times W \times C/8$
	ReLU	$H \times W \times C/8$	$H \times W \times C/8$
	conv1 × 1	$H \times W \times C/8$	$H \times W \times 1$
	sigmoid	$H \times W \times 1$	$H \times W \times 1$
	multiple	$H \times W \times C, H \times W \times 1$	$H \times W \times C$

3.2.2. Shallow Features Extraction

Deep CNNs with numerous convolution layers are hierarchical models and naturally give multilevel representations of input images, the lower layer representations focus on local details (e.g., edge and contours of an object) and the higher layer representations involve more global priority (e.g., environmental type). It also brings limitations that there are only high-dimensional deep features left, while the edge, texture, contour, and other shallow features of the image disappear at the end of the network.

Therefore, we extract the edge details of the original LR image and perform complementary fusion at the end of the network. However, for remote sensing images with complex scene structure, the main edge information is mixed with the secondary edge information, and the secondary information interferes with the neural network model to a certain extent. As a result, we use L_0 gradient minimization [23] to filter out the secondary edge information in the concrete implementation, and the gradient of the new image is more conducive to the recovery of main edge image information. The image differences with or without L_0 gradient minimization are shown in Figure 6. Note that images with L_0 gradient minimization maintain the main edge to the maximum extent and can effectively supplement the shallow feature deficiency caused by the deep network.

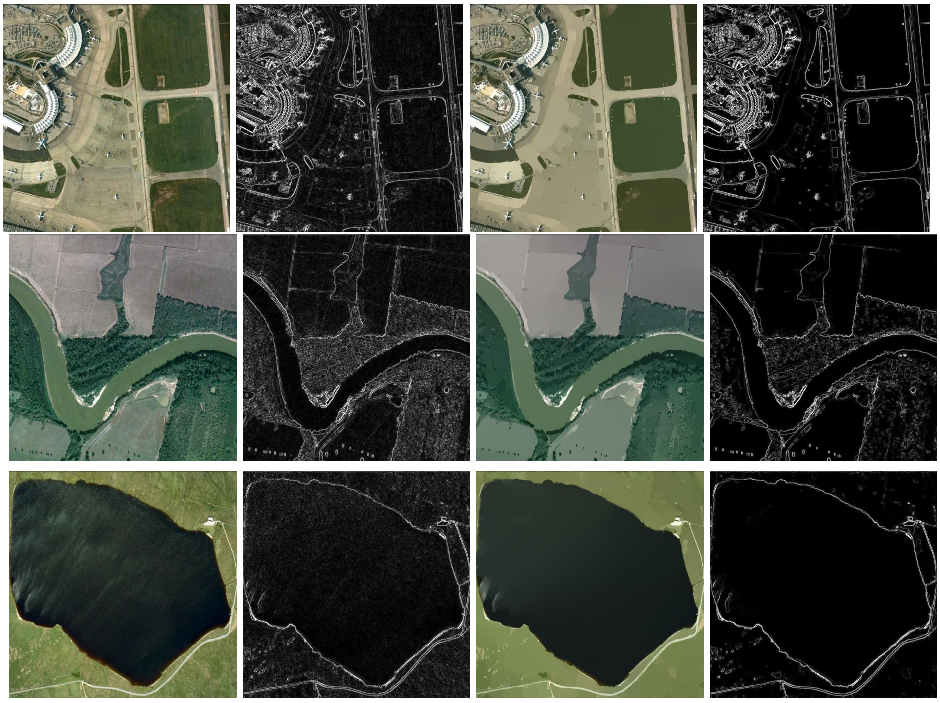


Figure 6. The original images and the gradient of the original images, and both of them after L_0 gradient minimization.

3.3. Loss Function

Our networks are trained by supervised learning, and the loss function is the ultimate goal of the network, which is very important to guide the training process of the network [28]. In light of the characteristics of remote sensing images, edge loss and cycle consistent loss [29] based on charbonnier loss [24] are added to make network convergence faster and easier. As usual, the loss function first calculates the holistic and detailed differences between HR image and SR image to guide the gradient optimization process of the network. The charbonnier loss calculates the overall difference between SR image and HR image:

$$L_{char} = \sqrt{||I_{HR} - I_{SR}||^2 + \epsilon^2} \tag{1}$$

where I_{HR} denotes the HR image and I_{SR} denotes the SR image, and the constant ϵ is empirically set to 10^{-3} for all the experiments.

To make full utilization of edge information, we apply edge loss to calculate the edge difference between HR images and SR images, to improve the effect of reconstruction of texture details such as edges. The calculation formula for edge loss is as follows:

$$L_{edge} = \sqrt{||\Delta(I_{HR}) - \Delta(I_{SR})||^2 + \epsilon^2} \tag{2}$$

where Δ denotes Laplacian operator.

In addition, SR is an inherently ill-posed problem where many more HR pixels need to be estimated under limited known LR pixels, which is where an LR may have multiple HR pairs. If we only focus on learning the mapping from LR images to HR images, the space of possible mapping functions may be very large, which makes training very difficult. The process from LR images to SR images is seen as positive and the process from SR images to LR images is seen as reversed, which can form a cycle. Obviously, downsampled I_{SR} should be consistent with I_{LR}, so we use cycle consistent loss to make better utilization of I_{LR} and narrow the range of SR solutions. In other words, we not only pay attention to the proximity of I_{HR} and I_{SR}, but also pay attention to the proximity of I_{LR} and I_{SR} after downsampling. The calculation formula for cycle consistent loss is as follows:

$$L_{cycle} = \sqrt{||I_{LR} - I_{SR} \downarrow ||^2 + \epsilon^2} \tag{3}$$

where $I_{SR} \downarrow$ represents the SR image downsampled by bicubic to the same resolution as I_{LR}.

In the end,the total loss function can be expressed as

$$L = L_{char} + \lambda_1 L_{edge} + \lambda_2 L_{cycle} \tag{4}$$

where λ_1 and λ_2 are used to adjust the weights of edge loss and cycle consistent loss.

As the same form of loss calculation is adopted, the calculation value is in the same order of magnitude, so we set λ_1 and λ_2 to 1, avoiding the difficulty of hyperparameters setting.

4. Experiment

In this section, we first introduce two remote sensing datasets and the implementation details of our SR networks. After that, we perform experiments to verify the effectiveness of preclassification strategy. Finally, we fully compare our method with various state-of-the-art methods, and display quantitative evaluation and visual comparison.

4.1. Dataset Settings

We choose two datasets with plentiful scenes to verify the robustness of our proposed method. There are some training images shown in Figure 7.

WHURS [30]: This is a classical remote sensing dataset, which consists of 1005 images in 19 classes of remote sensing images with different geographical topography, including airport, beach, bridge, commercial, etc. All images are in 600 × 600 pixels and the spatial resolution is up to 0.5 m/pixel. We randomly select 10 images from each class as the testing set, and the rest as the training set.

AID [31]: This is a large-scale aerial image dataset that collects sample images from Google Earth images. The AID dataset contains 10,000 images of 30 land-use scenes, including river, mountain, farmland, pond, and so on. All sample images of each category were carefully selected from different countries and regions of the world and extracted at different times and seasons under different imaging conditions, which increases the diversity in the classes of the data. We randomly select 20% of the total number as the testing set, and the remaining 80% as the training set.

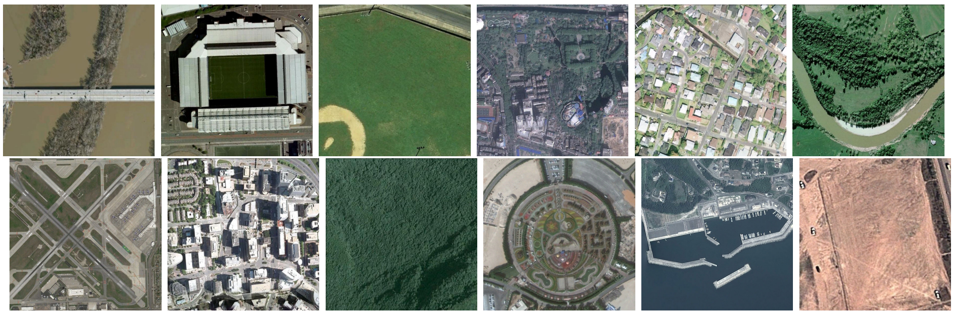


Figure 7. Examples of WHURS and AID datasets. The first line is the WHURS dataset, and the second line is the AID dataset.

4.2. Implementation Details

We design corresponding networks for remote sensing images of different complexity, and the main framework of these networks is similar, as shown in Figure 3. The three corresponding sub-networks (simple net, medium net, complex net) in Figure 1 are established by controlling the number of MKRA and the number of convolutional channels. To save memory and reduce computation, simple net has 10 MKRAs with 32 channels, medium net has 15 MKRAs with 48 channels, and complex net has 20 MKRAs with 64 channels.

Following the settings of EDSR [18], in each training batch, the input LR images are randomly cropped in a patch size of 48 × 48, and the corresponding input HR images with sizes of 96 × 96, 144 × 144, and 192 × 192 are cropped according to the upscaling factors ×2, ×3, and ×4, respectively. To produce the LR input frames, we downsample the HR frames through bicubic [32] interpolation. In addition, the training sets are also augmented via three image-processing methods: horizontal flipping, vertical flipping, and 90° rotation. More detailed parameter settings are indicated in Table 2. The proposed algorithm is implemented under the PyTorch [33] framework on a computer with an NVIDIA GTX 2080Ti GPU.

Table 2. Parameter settings during the training process.

Parameter	Setting
Batch size	8
Training epoch number	500
Optimization method	Adam [34], $\beta_1 = 0.9$, $\beta_2 = 0.999$, $\epsilon = 10^{-8}$
Learning rate (LR)	Initial $LR = 10^{-4}$, halved every 100 epochs

4.3. Preclassification Experiment

To verify the effectiveness of the preclassification strategy, we first use the classical SR network EDSR to conduct validation experiments on scales ×2, ×3, and ×4. As shown in Table 3, with the preclassification strategy, the remote sensing images with different complexity have been improved, especially the remote sensing images with higher complexity. The preclassification strategy has strong transferability and universal applicability, especially for various remote sensing images.

4.4. Quantitative and Qualitative Evaluation

In this section, the proposed method is evaluated with other methods quantitatively and qualitatively. To further verify the advancement and effectiveness of the proposed method, we compare our method with bicubic [32] and five other state-of-the-art methods: very deep super-resolution (VDSR) [15], enhanced deep super-resolution (EDSR) [18], pixel attention network (PAN) [19], local–global combined network (LGCNet) [21], and deep residual squeeze and excitation network (DRSEN) [22]. Bicubic interpolation is a representative interpolation algorithm. VDSR adopts residual learning to build a deep network. PAN builds a lightweight CNN with pixel attention for quick SR. EDSR is a representative version of deep network architectures with residual blocks. LGCNet and DRSEN are two SR methods for remote sensing images. To fairly compare the performance of the networks, the number of residual blocks for EDSR and the number of RSEB for DRSEN are set to 20, and both convolution filters are all set to 64. For a fair comparison, these methods are retrained under our training datasets.

Table 3. Preclassification verified on EDSR with WHURS dataset.

Scale	Preclassification	Metric	Simple Class	Medium Class	Complex Class
×2		PSNR SSIM	41.6271 0.9717	34.5859 0.9511	29.8556 0.9237
×2	√	PSNR SSIM	41.7225 0.9725	34.8476 0.9542	30.3133 0.9308
×3		PSNR SSIM	38.5069 0.9267	30.4516 0.8630	25.9097 0.8010
×3	√	PSNR SSIM	38.5845 0.9286	30.5792 0.8659	26.1368 0.8084
×4		PSNR SSIM	35.4589 0.8801	27.7834 0.7700	23.6102 0.6787
×4	√	PSNR SSIM	36.3193 0.8914	28.1385 0.7921	23.8651 0.7088

The model size is a critical issue in practical applications, especially in devices with low computing power. Furthermore, for the scale factor ×4, Figure 8 illustrates the comparison of the number of parameters between our SR network and other networks. Our simple net is close to the network with the minimum number of parameters, while the parameters number of complex net is less than EDSR and DRSEN. This provides an appropriate network for applications in different scenarios.

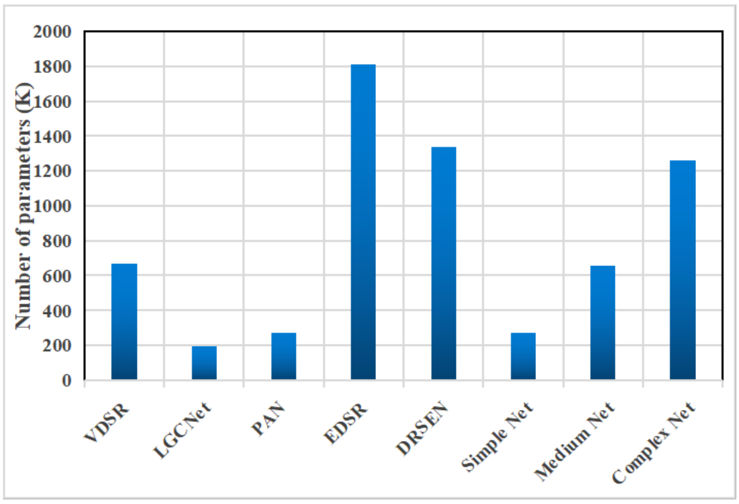


Figure 8. The number of network parameters (K) for scale ×4, simple net, medium net, and complex net are our SR nets in Figure 1.

4.4.1. Quantitative Evaluation

We adopt the peak signal-to-noise ratio (PSNR) [35] and structural similarity (SSIM) [36] as the objective evaluation indexes to measure the quality of the SR image reconstruction. The PSNR is one of the most widely used standards for evaluating image quality, and it compares the pixel differences between HR and SR images. Larger PSNR values indicate lower distortion and a better SR reconstruction effect. The SSIM is another widely used measurement index in SR image reconstruction, which is based on the luminance, contrast, and structure of HR image and LR image. If the SSIM value is closer to 1, then the similarity is greater between the SR image and the HR image, a.k.a., the higher the quality of the SR image.

It can be seen from the experimental results that the amount of training data of a single network is reduced to about 1/3 of the original dataset by the preclassification strategy proposed in this paper, but the reconstruction effect is greatly improved. Compared with other SOTA methods, our method achieves the best results in both PSNR and SSIM with different scale factors, and can adapt to the SR requirements of different scales. In addition, Table 4 implies the mean results of each method on WHURS and AID datasets with ×2, ×3, and ×4 scale factors, which reveals that our model outperforms other methods. On the WHURS testing set with scales ×2, ×3, and ×4, the PSNR and SSIM of our method reach 36.4095/0.9512, 31.8460/0.8701, and 29.4892/0.7976, respectively. It outperforms the second-best model, DRSEN, with PSNR gains of 0.1289 dB, 0.1954 dB, and 0.1428 dB, with SSIM gains of 0.0010, 0.0036, and 0.0104. The comparison results of PSNR and SSIM are more visually depicted in Figure 9. In particular, SSIM values under large scale (×4) are increased by 0.0104 and 0.0114 in WHURS and AID testing sets, respectively, equivalent to 1.32% and 1.49% higher than DRSEN (second place method), indicating that our method can effectively reconstruct the structural information of remote sensing images.

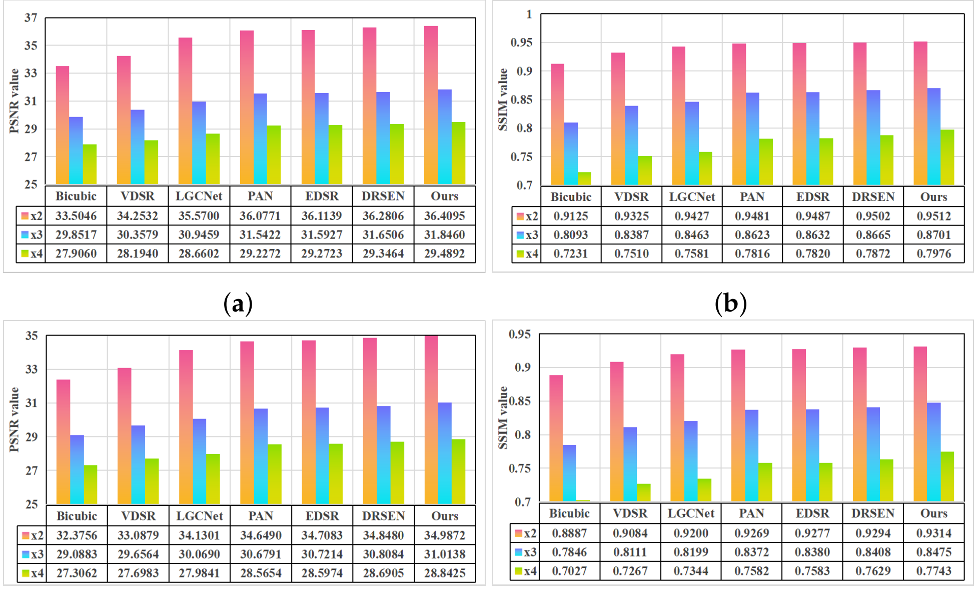
Table 4. Average PSNR and SSIM results of various SR methods. Bold index indicates the best performance.

Dataset	Scale	Metric	Bicubic	VDSR	LGCNet	PAN	EDSR	DRSEN	Ours
WHURS	×2	PSNR	33.5046	34.2532	35.5700	36.0771	36.1139	36.2806	**36.4095**
		SSIM	0.9125	0.9325	0.9427	0.9481	0.9487	0.9502	**0.9512**
	×3	PSNR	29.8517	30.3579	30.9459	31.5422	31.5927	31.6506	**31.8460**
		SSIM	0.8093	0.8387	0.8463	0.8623	0.8632	0.8665	**0.8701**
	×4	PSNR	27.9060	28.1940	28.6602	29.2272	29.2723	29.3464	**29.4892**
		SSIM	0.7231	0.7510	0.7581	0.7816	0.7820	0.7872	**0.7976**
AID	×2	PSNR	32.3756	33.0879	34.1301	34.6490	34.7083	34.8480	**34.9872**
		SSIM	0.8887	0.9084	0.9200	0.9269	0.9277	0.9294	**0.9314**
	×3	PSNR	29.0883	29.6564	30.0690	30.6791	30.7214	30.8084	**31.0138**
		SSIM	0.7846	0.8111	0.8199	0.8372	0.8380	0.8408	**0.8475**
	×4	PSNR	27.3062	27.6983	27.9841	28.5654	28.5974	28.6905	**28.8425**
		SSIM	0.7027	0.7267	0.7344	0.7582	0.7583	0.7629	**0.7743**

4.4.2. Qualitative Evaluation

To more fully illustrate the effectiveness of our method, the reconstruction results are examined qualitatively and some of the visual comparisons are demonstrated in Figure 10. It is noteworthy that our method achieves better results on the different scenes, reducing sawtooth and better reconstructing the structure and edge of the objects in the images. On the basis of numerical analysis in Table 5, we can find that large diversity exists among these remote sensing image classes, showing the authenticity and diversity of the test datasets. As can be seen from the visual results in Figure 10, compared with various typical deep learning SR methods, the proposed method in this paper has clearer details of reconstructed grain edges and richer details and textures. For a clearer comparison, a small patch marked by a red rectangle is enlarged and shown for each SR method.

The bicubic upsampling strategy results in loss of texture and blurry structure, which is more obvious for remote sensing images with weak edge details. VDSR and LGCNet take such bicubic upsampling results as network inputs, and then they produce erroneous structural and texture information and fail to recover more details, ultimately resulting in poor SR image quality. Other methods directly use LR as input, then achieve better results, but they do not take into account the characteristics of remote sensing images; the edges are still difficult to distinguish, as shown in Figure 10a. The results of our method have more clearly differentiated edges, which also can suppress noise and maintain the color consistency of local regions, as depicted in Figure 10c, while being closer to high-resolution images.



(**a**) (**b**)

(**c**) (**d**)

Figure 9. Average PSNR and SSIM results of various SR methods. (**a**) PSNR on WHURS; (**b**) SSIM on WHURS; (**c**) PSNR on AID; (**d**) SSIM on AID.

Table 5. PSNR and SSIM results of various SR methods on scale factor ×4 in Figure 10. Bold index indicates the best performance.

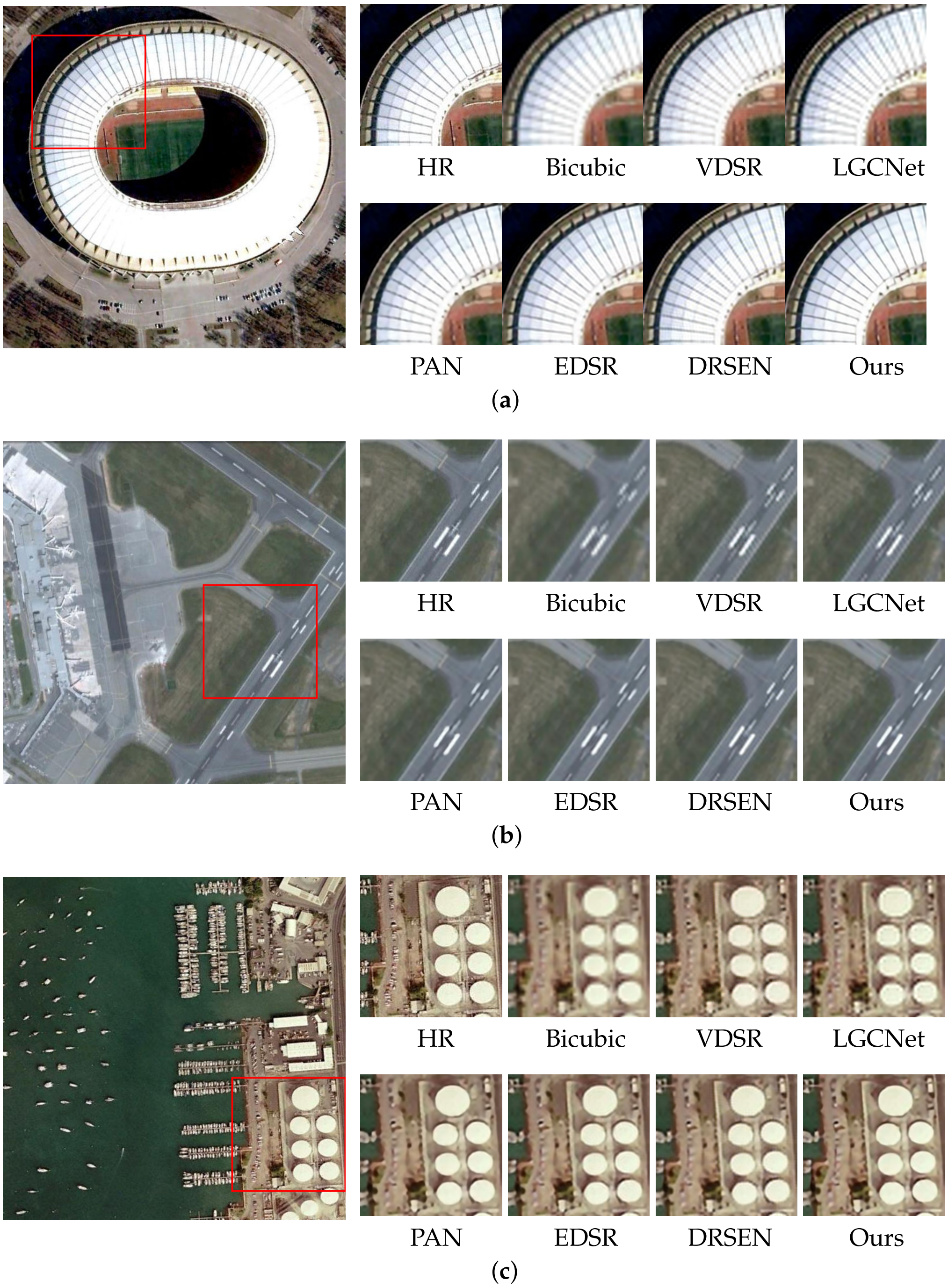
Image	Metric	Bicubic	VDSR	LGCNet	PAN	EDSR	DRSEN	Ours
(a) stadium	PSNR	24.4454	25.4497	25.3274	27.3530	27.3029	27.5224	**28.1065**
	SSIM	0.7624	0.7858	0.7923	0.8568	0.8513	0.8566	**0.8793**
(b) airport	PSNR	32.2469	32.9542	33.4316	34.5850	34.7783	34.9624	**35.2572**
	SSIM	0.8802	0.8890	0.8995	0.9173	0.9190	0.9222	**0.9279**
(c) port	PSNR	23.0638	23.8094	23.6353	24.4243	24.4345	24.5867	**24.7065**
	SSIM	0.7410	0.7685	0.7627	0.8044	0.8014	0.8068	**0.8187**
(d) river	PSNR	30.9059	31.3190	31.7969	32.1488	32.2085	32.2516	**32.3075**
	SSIM	0.8080	0.8236	0.8391	0.8460	0.8472	0.8487	**0.8519**

4.5. Discussion

According to the quantitative and qualitative evaluation in Section 4.4, the proposed method performs better than the other methods. Our method can be used as a reference solution for the problems such as the diversity of remote sensing images and the weakening of edge details. However, there are some limitations.

On the one hand, a mass of paired higher-quality images is necessary for deep learning, but it is often difficult to obtain pairs of high-resolution and degraded images. When training the network, we use bicubic method to produce the corresponding low-resolution images from high-resolution remote sensing images. However, in actual situations, the bicubic method does not fully represent the degradation process of remote sensing images. Our method may be inadequate for some extremely distorted images. The degradation process of remote sensing images needs further study in the future.

On the other hand, although some images at the intersection of intervals can be classified according to the proposed preclassification strategy, compared with the image inside the interval, this strategy is too simple and direct, so the classification based on fuzzy ideas can be explored in the future method.

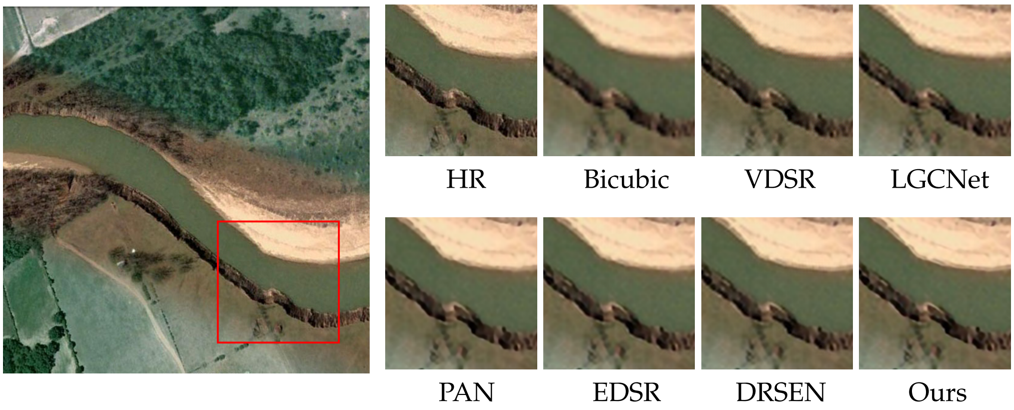


(**a**)

(**b**)

(**c**)

Figure 10. *Cont.*



(**d**)

Figure 10. Visual comparison of some representative SR methods and our model on ×4 factor: (**a**) stadium; (**b**) airport; (**c**) port; (**d**) river.

5. Conclusions

In view of the characteristics of remote sensing images, we propose an SR method for remote sensing images using preclassification strategy and deep–shallow features fusion. The preclassification strategy divides remote sensing images into three classes according to the structural complexity of scenes, and different networks are applied for each class. In this way, the training difficulty of each network is reduced, and each network can learn the commonness of same-class images. Moreover, considering the weak edge structure of remote sensing images, our networks are shallow features fused to deep features.We smooth the LR images by L_0 gradient minimization, and extract the main edge of the new LR images as the shallow features. The MKRA module is proposed to extract deep features, and the shallow features are integrated at the end of the deep network. Finally, the edge loss is added to improve the edge reconstruction effect and the cycle consistent loss is added to raise utilization of LR images. Numerous comparative experiments demonstrate that the SR method in this paper can enrich texture details of reconstructed images and provide better visual effects, and the PSNR and SSIM values of quantitative parameters are also generally improved. More advanced classification methods will be considered in the future to further reduce the number of training samples required, as well as the degradation process of different scenarios.

Author Contributions: Conceptualization, B.J.; methodology, X.C. and B.J.; software, X.Y. and X.C.; validation, X.C., B.J. and L.W.; data curation, X.Y.; writing—original draft preparation, X.Y., B.J. and J.Z.; writing—review and editing, X.Y., H.M., M.W. and W.Z.; supervision, W.Z., L.W. and X.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research is supported by the National Key Research and Development Program of China (No. 2019YFC1510503), the National Natural Science Foundation of China (Nos. 61801384, 61901371, 61901372, 61877050), the Key Research and Development Program of Shaanxi Province of China (Nos. 2020KW-010, 2021KW-05), the Natural Science Basic Research Program of Shaanxi Province of China (Nos. 2020JM-415, 2020JQ-600), and the Northwest University Paleontological Bioinformatics Innovation Team (No. 2019TD-012).

Data Availability Statement: The data of experimental images used to support the findings of this research are available from the corresponding author upon reasonable request.

Conflicts of Interest: All authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this paper:

SR	Super-resolution
LR	Low-resolution
HR	High-resolution
PSNR	Peak signal-to-noise ratio
SSIM	Structural similarity
MKRA	Multi-kernel residual attention
VDSR	Very deep super-resolution
LGCNet	Local–global combined network
EDSR	Enhanced deep super-resolution
PAN	Pixel attention network
DRSEN	Deep residual squeeze and excitation network

References

1. Chen, Z.; Zhang, T.; Ouyang, C. End-to-end airplane detection using transfer learning in remote sensing images. *Remote Sens.* **2018**, *10*, 139. [CrossRef]
2. Li, Y.; Chen, W.; Zhang, Y.; Tao, C.; Xiao, R.; Tan, Y. Accurate cloud detection in high-resolution remote sensing imagery by weakly supervised deep learning. *Remote Sens. Environ.* **2020**, *250*, 112045. [CrossRef]
3. Chen, K.; Fu, K.; Yan, M.; Gao, X.; Sun, X.; Wei, X. Semantic segmentation of aerial images with shuffling convolutional neural networks. *IEEE Geosci. Remote Sens. Lett.* **2018**, *15*, 173–177. [CrossRef]
4. Ran, S.; Gao, X.; Yang, Y.; Li, S.; Zhang, G.; Wang, P. Building multi-feature fusion refined network for building extraction from high-resolution remote sensing images. *Remote Sens.* **2021**, *13*, 2794. [CrossRef]
5. Ma, L.; Liu, Y.; Zhang, X.; Ye, Y.; Yin, G.; Johnson, B.A. Deep learning in remote sensing applications: A meta-analysis and review. *ISPRS J. Photogramm. Remote Sens.* **2019**, *152*, 166–177. [CrossRef]
6. Yuan, Q.; Shen, H.; Li, T.; Li, Z.; Li, S.; Jiang, Y.; Xu, H.; Tan, W.; Yang, Q.; Wang, J. Deep learning in environmental remote sensing: Achievements and challenges. *Remote Sens. Environ.* **2020**, *241*, 111716. [CrossRef]
7. Zhang, X.; Han, L.; Han, L.; Zhu, L. How well do deep learning-based methods for land cover classification and object detection perform on high resolution remote sensing imagery? *Remote Sens.* **2020**, *12*, 417. [CrossRef]
8. Yang, W.; Zhang, X.; Tian, Y.; Wang, W.; Xue, J.H.; Liao, Q. Deep learning for single image super-resolution: A brief review. *IEEE Trans. Multimed.* **2019**, *21*, 3106–3121. [CrossRef]
9. Dong, W.; Zhang, L.; Shi, G.; Wu, X. Image deblurring and super-resolution by adaptive sparse domain selection and adaptive regularization. *IEEE Trans. Image Process.* **2011**, *20*, 1838–1857. [CrossRef] [PubMed]
10. Yang, J.; Wright, J.; Huang, T.S.; Ma, Y. Image super-resolution via sparse representation. *IEEE Trans. Image Process.* **2010**, *19*, 2861–2873. [CrossRef] [PubMed]
11. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
12. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster R-CNN: Towards real-time object detection with region proposal networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2016**, *39*, 1137–1149. [CrossRef] [PubMed]
13. Long, J.; Shelhamer, E.; Darrell, T. Fully convolutional networks for semantic segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 3431–3440.
14. Dong, C.; Loy, C.C.; He, K.; Tang, X. Learning a deep convolutional network for image super-resolution. In Proceedings of the European Conference on Computer Vision, Zurich, Switzerland, 6–12 September 2014; Springer: Berlin/Heidelberg, Germany, 2014; pp. 184–199.
15. Kim, J.; Lee, J.K.; Lee, K.M. Accurate image super-resolution using very deep convolutional networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 1646–1654.
16. Shi, W.; Caballero, J.; Huszár, F.; Totz, J.; Aitken, A.P.; Bishop, R.; Rueckert, D.; Wang, Z. Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 1874–1883.
17. Ledig, C.; Theis, L.; Huszár, F.; Caballero, J.; Cunningham, A.; Acosta, A.; Aitken, A.; Tejani, A.; Totz, J.; Wang, Z.; et al. Photo-realistic single image super-resolution using a generative adversarial network. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 4681–4690.
18. Lim, B.; Son, S.; Kim, H.; Nah, S.; Mu Lee, K. Enhanced deep residual networks for single image super-resolution. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Honolulu, HI, USA, 21–26 July 2017; pp. 136–144.
19. Zhao, H.; Kong, X.; He, J.; Qiao, Y.; Dong, C. Efficient image super-resolution using pixel attention. In Proceedings of the European Conference on Computer Vision, Glasgow, UK, 23–28 August 2020; Springer: Berlin/Heidelberg, Germany, 2020; pp. 56–72.

20. Yang, D.; Li, Z.; Xia, Y.; Chen, Z. Remote sensing image super-resolution: Challenges and approaches. In Proceedings of the IEEE International Conference on Digital Signal Processing, Singapore, 21–24 July 2015; pp. 196–200.
21. Lei, S.; Shi, Z.; Zou, Z. Super-resolution for remote sensing images via local–global combined network. *IEEE Geosci. Remote Sens. Lett.* **2017**, *14*, 1243–1247. [CrossRef]
22. Gu, J.; Sun, X.; Zhang, Y.; Fu, K.; Wang, L. Deep residual squeeze and ecitation network for remote sensing image super-resolution. *Remote Sens.* **2019**, *11*, 1817. [CrossRef]
23. Xu, L.; Lu, C.; Xu, Y.; Jia, J. Image smoothing via L_0 gradient minimization. In Proceedings of the 2011 SIGGRAPH Asia Conference, Hong Kong, 12–15 December 2011; pp. 1–12.
24. Lai, W.S.; Huang, J.B.; Ahuja, N.; Yang, M.H. Fast and accurate image super-resolution with deep laplacian pyramid networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2018**, *41*, 2599–2613. [CrossRef] [PubMed]
25. Hu, J.; Shen, L.; Sun, G. Squeeze-and-excitation networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 7132–7141.
26. Cheng, X.; Li, X.; Yang, J.; Tai, Y. SESR: Single image super resolution with recursive squeeze and excitation networks. In Proceedings of the International Conference on Pattern Recognition, Beijing, China, 20–24 August 2018; pp. 147–152.
27. Zhang, Y.; Li, K.; Li, K.; Wang, L.; Zhong, B.; Fu, Y. Image super-resolution using very deep residual channel attention networks. In Proceedings of the European Conference on Computer Vision, Munich, Germany, 8–14 September 2018; pp. 286–301.
28. Zhao, H.; Gallo, O.; Frosio, I.; Kautz, J. Loss functions for image restoration with neural networks. *IEEE Trans. Comput. Imaging* **2016**, *3*, 47–57. [CrossRef]
29. Zhu, J.Y.; Park, T.; Isola, P.; Efros, A.A. Unpaired image-to-image translation using cycle-consistent adversarial networks. In Proceedings of the IEEE international Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 2223–2232.
30. Dai, D.; Yang, W. Satellite image classification via two-layer sparse coding with biased image representation. *IEEE Geosci. Remote Sens. Lett.* **2010**, *8*, 173–176. [CrossRef]
31. Xia, G.S.; Hu, J.; Hu, F.; Shi, B.; Bai, X.; Zhong, Y.; Zhang, L.; Lu, X. AID: A benchmark data set for performance evaluation of aerial scene classification. *IEEE Trans. Geosci. Remote Sens.* **2017**, *55*, 3965–3981. [CrossRef]
32. Keys, R. Cubic convolution interpolation for digital image processing. *IEEE Trans. Acoust. Speech Signal Process.* **1981**, *29*, 1153–1160. [CrossRef]
33. Paszke, A.; Gross, S.; Chintala, S.; Chanan, G.; Yang, E.; DeVito, Z.; Lin, Z.; Desmaison, A.; Antiga, L.; Lerer, A. *Automatic Differentiation in Pytorch*; 2017. Available online: https://openreview.net/forum?id=BJJsrmfCZ (accessed on 10 February 2022).
34. Kingma, D.P.; Ba, J. Adam: A Method for Stochastic Optimization. *arXiv* **2014**, arXiv:1412.6980.
35. Hore, A.; Ziou, D. Image quality metrics: PSNR vs. SSIM. In Proceedings of the International Conference on Pattern Recognition, Istanbul, Turkey, 23–26 August 2010; pp. 2366–2369.
36. Wang, Z.; Bovik, A.C.; Sheikh, H.R.; Simoncelli, E.P. Image quality assessment: From error visibility to structural similarity. *IEEE Trans. Image Process.* **2004**, *13*, 600–612. [CrossRef]