remote sensing

MDPI

Article

IESRGAN: Enhanced U-Net Structured Generative Adversarial Network for Remote Sensing Image Super-Resolution Reconstruction

Xiaohan Yue [1], Danfeng Liu [1,*], Liguo Wang [1], Jón Atli Benediktsson [2], Linghong Meng [1] and Lei Deng [1]

[1] College of Information and Communication Engineering, Dalian Minzu University, Dalian 116600, China; 20211105324@dlnu.edu.cn (X.Y.); wangliguo@hrbeu.edu.cn (L.W.); 20211105267@dlnu.edu.cn (L.M.); 202211051004@dlnu.edu.cn (L.D.)

[2] Faculty of Electrical and Computer Engineering, University of Iceland, 107 Reykjavik, Iceland; benedikt@hi.is

* Correspondence: liudanfeng@dlnu.edu.cn

Abstract: With the continuous development of modern remote sensing satellite technology, high-resolution (HR) remote sensing image data have gradually become widely used. However, due to the vastness of areas that need to be monitored and the difficulty in obtaining HR images, most monitoring projects still rely on low-resolution (LR) data for the regions being monitored. The emergence of remote sensing image super-resolution (SR) reconstruction technology effectively compensates for the lack of original HR images. This paper proposes an Improved Enhanced Super-Resolution Generative Adversarial Network (IESRGAN) based on an enhanced U-Net structure for a 4× scale detail reconstruction of LR images using NaSC-TG2 remote sensing images. In this method, in-depth research has been performed and consequent improvements have been made to the generator and discriminator within the GAN network. Specifically, before introducing Residual-in-Residual Dense Blocks (RRDB), in the proposed method, input images are subjected to reflective padding to enhance edge information. Meanwhile, a U-Net structure is adopted for the discriminator, incorporating spectral normalization to focus on semantic and structural changes between real and fake images, thereby improving generated image quality and GAN performance. To evaluate the effectiveness and generalization ability of our proposed model, experiments were conducted on multiple real-world remote sensing image datasets. Experimental results demonstrate that IESRGAN exhibits strong generalization capabilities while delivering outstanding performance in terms of PSNR, SSIM, and LPIPS image evaluation metrics.

Keywords: super-resolution reconstruction; remote sensing images; generative adversarial networks

Citation: Yue, X.; Liu, D.; Wang, L.; Benediktsson, J.A.; Meng, L.; Deng, L. IESRGAN: Enhanced U-Net Structured Generative Adversarial Network for Remote Sensing Image Super-Resolution Reconstruction. *Remote Sens.* **2023**, *15*, 3490. https://doi.org/10.3390/rs15143490

Academic Editors: Xinghua Li, Benoit Vozel, Vladimir Lukin and Yakoub Bazi

Received: 31 May 2023
Revised: 29 June 2023
Accepted: 8 July 2023
Published: 11 July 2023

1. Introduction

Remote sensing technology can determine ground object targets and natural phenomena by collecting and analyzing electromagnetic waves [1]. Remote sensing also offers a repetitive and continuous perspective for observing Earth, making its value in monitoring short-term and long-term changes and the effects of human activities immeasurable [2]. Among other things, remote sensing images are a way to demonstrate the application of remote sensing data and image quality is directly related to the results of application analysis. Spatial resolution represents the smallest unit size or dimension that can be distinguished in remote sensing images and serves as an indicator of the image's ability to distinguish details of ground targets [3]. The higher the spatial resolution, the more information about ground objects is contained within remote sensing images, allowing for finer target identification. However, due to limitations such as under-sampling effects from imaging sensors and various degradation factors during image processing in transmission satellites, relying solely on hardware-level improvements for spatial resolution would result in high development costs and lengthy hardware iteration cycles. The image SR

technology provides a low-cost and effective way to obtain HR images by reconstructing HR images from relatively LR but easily available images [4].

Traditional SR reconstruction methods mainly include interpolation and prior-information-based reconstruction. Interpolation methods, such as bilinear interpolation [5], bicubic interpolation [6], and edge-guided interpolation [7], rely on neighboring pixels to estimate the current pixel value. Although interpolation methods have demonstrated good real-time performance, their results often have obvious edge effects and poor performance in detail recovery. Prior-information-based reconstruction methods use constraints, such as iterative back-projection [8], convex set projection [9], and maximum a posteriori probability method [10] to estimate the information points in the reconstructed image. However, these traditional methods are usually limited to specific application scenarios, with high computational complexity and limited generalization capabilities.

With the rapid development of deep learning, single-image SR methods based on deep learning outperform traditional single-image SR methods in the field of remote sensing SR and have broad application prospects. [11,12]. Dong et al. [13] first applied convolution neural network (CNN) technology to SR image reconstruction and proposed the SRCNN model, which performs non-linear mapping by extracting low-resolution image features to reconstruct images. Although SRCNN outperforms traditional methods in terms of performance, it is still limited by image region content, slow training convergence speed, and single-scale applicability. To address these issues, Shi et al. [14] proposed the ESPCN algorithm that uses sub-pixel convolution layers at the end of the network for up-sampling, preserving more low-resolution image texture regions, and increasing training speed. With the emergence of VGG networks [15], network model design tends towards deeper layers. However, deeper networks are prone to gradient vanishing problems. To solve this problem, He et al. [16] proposed a deep residual convolutional neural network (Residual Network, ResNet). More recently, Kim et al. [17] introduced ResNet and proposed the VDSR model, which uses a residual learning strategy to obtain high-frequency information residuals, thereby obtaining more image detail information. In addition, Li et al. [18] proposed a Multi-Scale Residual Network (MSRN) which uses Multi-Scale Residual Blocks (MSRB) combined with different scale convolution kernels for feature extraction and fusion. Lan et al. [19] pointed out that many CNN-based network models perform relatively poorly because they do not fully utilize low-level features. Therefore, they proposed the Cascading Residual Network (CRN) with multiple local shared groups and the Enhanced Residual Network (ERN) with a dual global path structure. Zhang et al. [20] introduced attention mechanisms into the SR field and proposed the Residual Channel Attention Network (RCAN), which adaptively adjusts each channel feature according to channel dependencies.

With continuous innovation and development in deep learning, Goodfellow [21] first proposed a revolutionary generative adversarial network (GAN). This method has achieved significant application results in many fields and laid a solid foundation for subsequent research. Ledig et al. [22] proposed the SRGAN model based on the GAN framework, using generators and discriminators for adversarial training. They found that mean square error loss leads to overly smooth reconstructed images and proposed perceptual loss to enhance the visual quality of reconstructed images. Wang et al. [23] further proposed the ESRGAN model, generating more realistic textures but still lacking high-frequency edge information in reconstructed images. In remote sensing applications, Rabbi et al. [24] targeted small object detection reconstruction performance in remote sensing images and proposed the EESRGAN algorithm using edge enhancement and different detector networks. Ma et al. [25] proposed a method based on Transferred Generative Adversarial Network that trains through transfer learning to improve remote sensing image reconstruction quality. Li et al. [26] proposed the SRAGAN algorithm using local and global attention mechanisms for different levels of feature extraction in remote sensing image ground scenes to reconstruct images. Salgueiro et al. [27] proposed the SEG-ESRGAN model, which combines semantic segmentation encoder–decoder architecture and uses multi-loss training methods. Zhu et al. [28] proposed an improved generative

adversarial network (an improved generative adversarial network via multi-scale residual blocks) that introduces multi-scale residual blocks in the generator network and uses attention mechanisms for multi-scale feature fusion. Zhao et al. [29] proposed the SA-GAN algorithm, which uses second-order channel attention mechanisms and region-level non-local modules in the generator network and employs region-aware loss to suppress artifact generation. Ali et al. [30] proposed an architecture for TESR (two-stage approach for enhancement and super-resolution) that exploits the power of visual deformers (ViT) and diffusion models (DM) to artificially improve the resolution of remotely sensed images.

Additionally, significant research has been conducted on resolution enhancement for other types of remote sensing images such as multisource image fusion [31,32] and hyperspectral imaging [33].

Although GAN has achieved remarkable success in fields such as image generation and style transfer, their training process still faces challenges, including mode collapse and gradient vanishing. Moreover, most current methods use pixel-level loss functions, such as mean squared error (MSE), which may lead to overly smooth reconstructed images lacking high-frequency details. Furthermore, remote sensing images exhibit more complex scenes and diverse target characteristics compared to ordinary images, necessitating consideration of real remote sensing dataset properties in reconstruction. Finally, while current super-resolution methods perform well on training data, they may lack generalization capabilities for unseen scenes and targets. Therefore, model design and training strategies should focus on enhancing robustness and generalization.

To address these issues, we propose IESRGAN: an improved GAN for remote sensing image super-resolution reconstruction based on an enhanced U-Net structure. The main adjustments and contributions include:

(1) Optimizing the generator network structure by adding reflection padding before the introduction of Residual-in-Residual Dense Blocks (RRDB), preventing image edge information loss and facilitating consistent feature map dimensions across RRDB layers to simplify skip connections and feature fusion processes.

(2) To improve performance further, we replace traditional discriminators with a U-Net-based discriminator and incorporate spectral normalization regularization. This allows for fusing image detail information at different resolution levels while enhancing the stability of the GAN discriminator.

(3) We demonstrate that our proposed IESRGAN exhibits strong generalization capabilities and performs well on real remote sensing images.

The rest of this paper is organized as follows. Section 2 details the structure of the IESRGAN; Section 3 verifies the effectiveness and generalization ability of IESRGAN by comparing it with other algorithms; Section 4 discusses the conclusions of IESRGAN in depth and points out future research directions.

2. Ideas and IESRGAN Methods

IESRGAN is composed of two main components: a generator and a discriminator. The overall workflow of IESRGAN is depicted in Figure 1. The generator is responsible for taking an input LR remote sensing image and reconstructing an HR image. It achieves this by utilizing operations such as convolution and up-sampling within its network structure. The generator network learns to map the LR image to an SR image with enhanced details and finer textures. Once the SR image is generated, it is passed through the U-Net-based discriminator. The discriminator's role is to compare the SR image with a real HR image and determine whether the SR image is realistic or not. The discriminator network is trained to identify flaws or discrepancies in the reconstructed images, enabling it to differentiate between real HR images and those generated by the generator. The generator and discriminator engage in continuous adversarial gameplay during training. The generator aims to produce SR images that are realistic enough to deceive the discriminator, while the discriminator strives to accurately identify the generated images. Through this adversarial process, both networks learn and improve their performance iteratively. As the training

progresses, the generator becomes more adept at generating high-quality and realistic HR images. Simultaneously, the discriminator becomes more discerning and capable of detecting flaws in the reconstructed images. This iterative training process leads to the generation of HR images with enhanced details and improved realism.

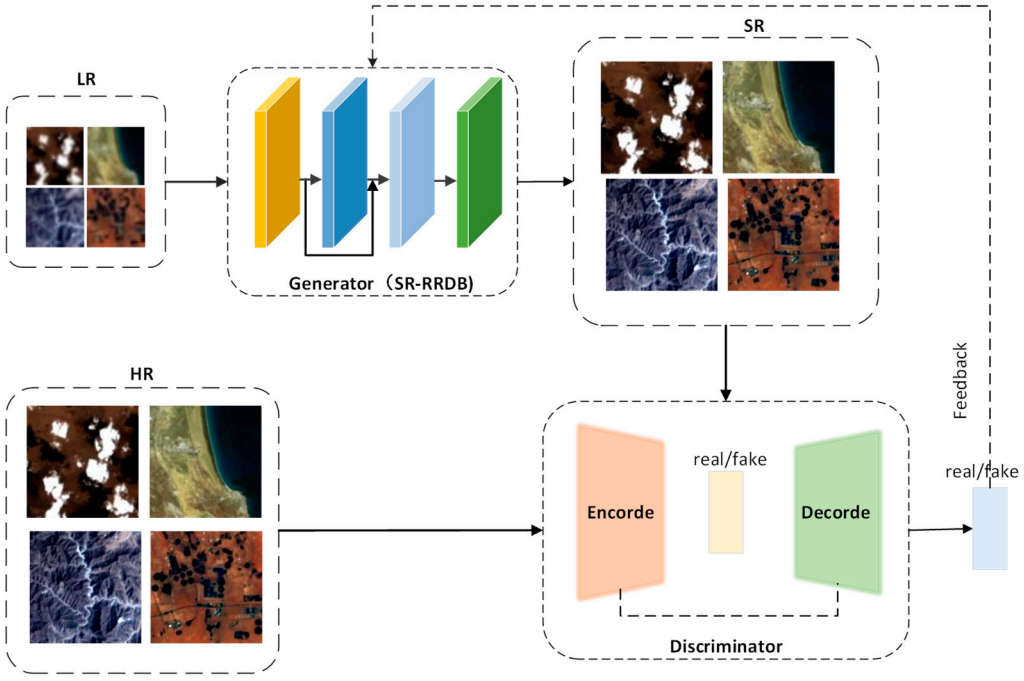


Figure 1. Network structure of IESRGAN.

2.1. Network Design of Generators—SR-RRDB

The generator network, depicted in Figure 2, is a CNN-based model. Initially, the input image undergoes a reflection padding layer, referred to as the ReflectionPad layer, which prevents edge information loss. Following this, RRDB are utilized to retain detail features while uncovering new ones. Notably, the generator comprises four primary modules.

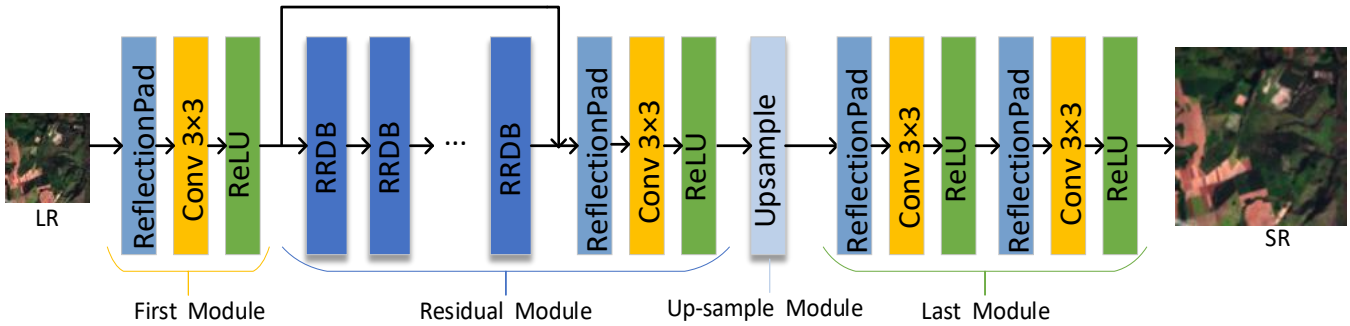


Figure 2. SR-RRDB: generator structure.

The first module is called the regular module, which consists of the ReflectionPad layer, Conv layer, and Rectified Linear Unit (ReLU) layer. The function of ReflectionPad is to perform reflection filling around the input image edges to extend edge information and avoid edge information loss and blurring; the Conv layer uses a 3×3 convolution kernel to perform convolution operation on the data in order to extract features; the ReLU layer performs a non-linear transformation to enhance the expressive power of the model. The ReLU layer has the advantages of simple computation, fast convergence, and no gradient disappearance problem.

The second module consists of 23 Residual-in-Residual Dense Block (RRDB) modules and a regular module with residual network connections. Among them, the RRDB combines the residual network structure and dense connectivity as shown in Figure 3. The residual network learns the residuals between the input and output, and most of the residuals can be 0 or smaller [34]. The dense connection is defined as $D = H([x_0, x_1, \ldots x_i])$, where $[x_0, x_1, \ldots x_i]$ denotes the network that combines $x_0, x_1, \ldots x_i$ layer-generated feature map connections as input [35]. Residual networks reuse features but are not good at mining new features while dense connections constantly explore new features but lead to higher redundancy [36]. RRDB combines the advantages of both network structures to make the model better adapted to complex data distributions and patterns, improving performance and accuracy.

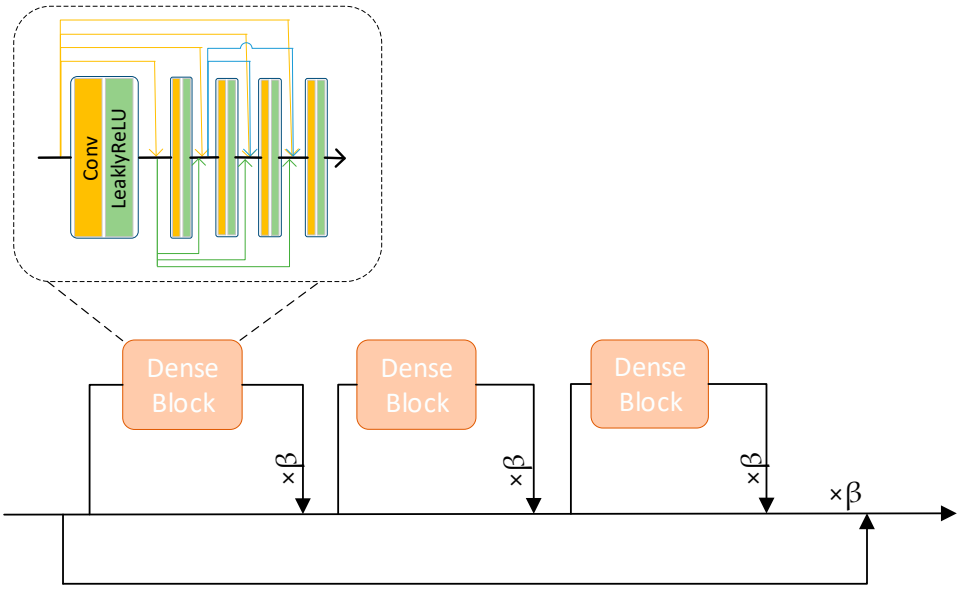


Figure 3. RRDB (Residual-in-Residual Dense Block).

The third module is up-sampling, which is used to increase the image size.

The last module consists of two regular modules where the convolution kernel is changed from 1×1 to 3×3 to enlarge the perceptual field and to learn features better. With the above generator network structure, called SR-RRDB, a high-resolution image corresponding to the input image is reconstructed.

2.2. Discriminator Network Design

In this study, instead of using the traditional discriminator structure, we chose a discriminator network based on the U-Net structure, as shown in Figure 4. This discriminator network structure consists of two main components: an encoder (down-sampling) and a decoder (up-sampling). The encoder is responsible for capturing the contextual information in the image, while the decoder is responsible for recovering the image details. To achieve information fusion, a jump connection is used between the two. As a result, this approach demonstrates its effectiveness in extracting multi-scale features from images with improved efficiency and accuracy.

It is worth noting that after entering the encoder from the initial convolution layer in this network structure, spectral normalization regularization is applied to stabilize the training of the discriminator network. Spectral normalization is a regularization method used in neural networks to prevent overfitting of neural networks by decomposing the weight matrix into eigenvalues and normalizing the result to limit the spectral norm of the weight

matrix. The specific algorithmic process is presented in Table 1. Spectral normalization [37] makes the spectral norm of weight matrix W satisfy the Lipschitz constraint $\sigma(W) = 1$:

$$\overline{W}_{\mathrm{SN}}(W) := W/\sigma(W) \tag{1}$$

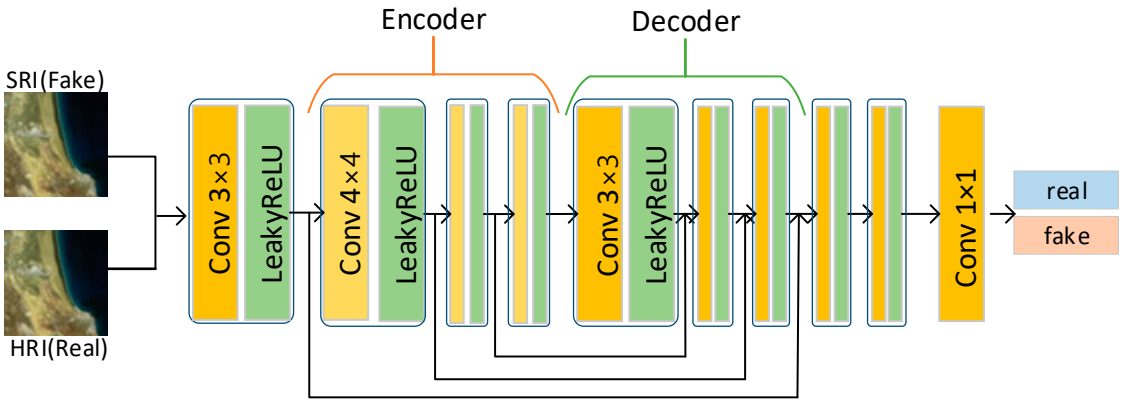


Figure 4. SR-RRDB: generator structure.

Table 1. Spectral normalization.

Spectral Normalization
· Initialize $\widetilde{u}_l \in \mathcal{R}^{d_l}$ for $l = 1, \ldots, L$ with a random vector (sampled from isotropic distribution)
· For each update and each layer l:
1. Apply power iteration method to an unnormalized weight W^l: $$\widetilde{v}_l \leftarrow \left(W^l\right)^{\mathrm{T}} \widetilde{u}_l / \|\left(W^l\right)^{\mathrm{T}} \widetilde{u}_l\|_2 \quad (2)$$ $$\widetilde{u}_l \leftarrow W^l \widetilde{v}_l / \|W^l \widetilde{v}_l\|_2 \quad (3)$$
2. Calculate $\overline{W}^l_{\mathrm{SN}}$ with the spectral norm: $$\overline{W}^l_{\mathrm{SN}}\left(W^l\right) = W^l/\sigma\left(W^l\right), \text{ where } \sigma\left(W^l\right) = \widetilde{u}_l^{\mathrm{T}} W^l \widetilde{v}_l \quad (4)$$
3. Update W^l on mini-batch dataset $\mathcal{D}_M$ with a learning rate α: $$W^l \leftarrow W^l - \alpha \nabla_{W^l} \updownarrow \left(\overline{W}^l_{\mathrm{SN}}\left(W^l\right), \mathcal{D}_M\right) \quad (5)$$

The use of a discriminator network based on the U-Net structure brings significant advantages. First, U-Net has jump connections, which fuse shallow features directly with deep features and alleviate the gradient disappearance problem. This allows the discriminator to learn semantic information at different scales and has a strong generalization capability. Secondly, since the U-Net structure fully considers the multi-scale information fusion, it can better capture the detail changes of small targets or local regions. This is important for generating high-quality images, especially in tasks that require the generation of fine structures and textures. Finally, U-Net restores features to the original input space step by step in the decoding stage by means of a deconvolution layer and continuously fuses shallow features. This allows the discriminator to take into account more contextual information, thus improving its ability to judge the quality of the generated images. Together, these advantages contribute to a significant improvement in GAN performance.

2.3. Loss Function

To enhance the robustness of the overall model, a fusion approach is employed in the loss function part. In the generator network, content loss, generation loss, and perceptual loss are included, where perceptual loss consists of content loss and generation loss. A binary cross entropy loss function (BCEWithLogitsLoss) is used in the discriminator network to counteract the loss.

The content loss is used to separately input the generated image and the target image into each convolutional layer in the VGG-19 network using the L1 norm and then calculate their differences in the feature space. The content loss formula is defined as:

$$\mathcal{L}_c = |G_l(\hat{y}) - G_l(y)|_1 \tag{6}$$

Here, $\hat{y}$ represents the generated image, y denotes the target image, $G_l(\cdot)$ signifies the feature map of layer l in the VGG-19 network, and $|\cdot|_1$ represents the L1 norm. The function of the content loss is to make the generated image closer to the pixel distribution of the target image, thus making the generated image more realistic. In the above formula, it is assumed that the feature map of a layer in the truncated VGG-19 network is represented as a three-dimensional tensor of $C_l \times H_l \times W_l$, where C_l indicates the number of channels, H_l indicates height, and W_l indicates width. Calculating generated image $\hat{y}$ at layer $l's$ feature map $G_l(\hat{y})$, its definition is as follows:

$$G_{l,i,j}(\hat{y}) = \frac{1}{C_l H_l W_l} \sum_{c=1}^{C_l} \sum_{h=1}^{H_l} \sum_{w=1}^{W_l} F_{l,c,h,w}(\hat{y}) \cdot \phi_{l,c,h,w}(i,j) \tag{7}$$

where $F_{l,c,h,w}(\hat{y})$ denotes the feature value of generated image $\hat{y}$ at layer l, channel c, row h, and column w; $\phi_{l,c,h,w}(i,j)$ represents the value of the convolution kernel at position (i,j) in layer l, channel c, row h, and column w of the VGG-19 network. $G_{l,i,j}(\hat{y})$ indicates the feature value of generated image $\hat{y}$ at layer l, row i, and column j.

In the generation loss, the discriminator is used to discriminate whether the SR-generated image is a "pseudo-image" or not, and then the discriminant result is obtained. Then, the BCEWithLogitsLoss is used to calculate adversarial loss, which is the difference between the probability of the generated image being discriminated as a real image and 1. The BCEWithLogitsLoss formula is expressed as:

$$\mathcal{L}_a = -\frac{1}{n} \sum_{i=1}^{n} [y_i \log \sigma(\hat{y}_i) + (1 - y_i) \log(1 - \sigma(\hat{y}_i))] \tag{8}$$

Here, n represents the number of samples, y_i denotes the label of the real image, $\hat{y}$ signifies the discriminant result of the discriminator on the generated image, and σ stands for sigmoid function.

The overall perceptual loss is defined as written in Equation (9):

$$\mathcal{L}_p = \mathcal{L}_c + \beta \, \mathcal{L}_a \tag{9}$$

The discrimination loss is calculated using the BCEWithLogitsLoss. First, the discriminant results are obtained by discriminating the SR-generated images and the real images separately. Next, the SR-generated image tensor is assumed to be 0, which means "false image", and the real image tensor is assumed to be 1, which means "true image". The formula expression is:

$$\mathcal{L}_a = \mathcal{L}_d^{sr} + \beta \, \mathcal{L}_a \tag{10}$$

where $\mathcal{L}_d^{sr}$ and $\mathcal{L}_d^{h}$ are, respectively, represented as:

$$\mathcal{L}_d^{sr} = -\frac{1}{n} \sum_{i=1}^{n} [{y_i}^{sr} \log \sigma(\hat{y}_i^{sr}) + (1 - {y_i}^{sr}) \log(1 - \sigma(\hat{y}_i^{sr}))] \tag{11}$$

In this equation, n indicates the number of samples; ${y_i}^{sr}$ is assumed to be a tensor with all zeros, which denotes the label of the fake image; ${\hat{y}_i}^{sr}$ is assumed to denote the discriminant result of the discriminator on the SR generated image, and σ signifies a sigmoid function.

$$\mathcal{L}_d^h = -\frac{1}{n}\sum_{i=1}^{n}\left[{y_i}^h \log\sigma\left(\hat{y}_i^h\right) + \left(1-{y_i}^h\right)\log\left(1-\sigma\left(\hat{y}_i^h\right)\right)\right] \tag{12}$$

Here, n indicates the number of samples; ${y_i}^h$ is assumed to be a tensor of all 1s, which denotes the label of the real image; ${\hat{y}_i}^h$ is assumed to denote the discriminant result of the real image, and σ signifies a sigmoid function.

The BCEWithLogitsLoss is advantageous in calculating the generative loss and the adversarial loss because it can not only measure the difference between the prediction result and the true result but also convert the prediction result into a probability value through the sigmoid function transformation, thus, reflecting the confidence level of the prediction result more accurately. In addition, BCEWithLogitsLoss can automatically handle the numerical stability problem and prevent numerical overflow or underflow in the calculation of the sigmoid function. In the adversarial training process, using BCEWithLogitsLoss can effectively evaluate the similarity between the generated image and the real image and provide better guidance for generator training.

3. Experiments

In this paper, we conduct model experiments with the following data and compare classical models in the super-resolution domain to verify the validity and generalization of the model.

3.1. Data Source

The remote sensing image data selected for this study include NaSC-TG2 [38], Satellite Images of Hurricane Damage [39], NWPU-RESISC45 [40], and UCMerced LandUse [41]. The NaSC-TG2 data originate from China's first space laboratory, Tiangong-2, which is equipped with a Wide-band Imaging Spectrometer (WIS) featuring 14 spectral channels covering visible light, near-infrared, short-wave infrared, and thermal infrared bands. The spatial resolution of these data at ground pixel distance is 100 m, 200 m, and 400 m. Satellite images of Hurricane Damage data are obtained from the Planet satellite constellation consisting of hundreds of Dove satellites (10 cm × 10 cm × 30 cm) that use optical systems and cameras to capture images in RGB and near-infrared bands with a ground pixel distance of 3~5 m. The NWPU-RESISC45 data come from Google Earth satellite images with spatial resolutions ranging from 0.2 m to 30 m, acquired through satellite imagery, aerial photography, and Geographic Information Systems (GIS). UCMerced LandUse data are sourced from the USGS National Map with a spatial resolution of 1 foot (0.3048 m). Table 2 summarizes the information on SR remote sensing image data used in this paper. Considering the spectral range differences across channels in these satellite image datasets, our experimental data only include RGB three-band images. The selection of these datasets will aid in further exploring remote sensing image processing techniques and provide theoretical support for enhancing practical applications.

Table 2. SR-RRDB: Generator structure.

Name	Size	Channel	Total Number	Spatial Resolution	Source
NaSC-TG2	128 × 128	3, 14	20,000	100 m, 200 m, 400 m	Tiangong-2
Satellite Image of Hurricane Damage	128 × 128	3	23,000	3~5 m	GeoEye1
NWPU-RESISC45	256 × 256	3	31,500	0.2~30 m	Google Earth
UCMerced LandUse	256 × 256	3	2100	0.3048 m	USGS National Map

In our experiments, we built a training set using 19,980 remote sensing images from the NaSC-TG2 dataset. Each HR image was down-sampled by a factor of four to obtain a low-resolution LR image. The HR images have a size of 128 × 128 pixels, and correspondingly, the LR images have a size of 32 × 32 pixels. Training with smaller-sized images allows the model to focus on rich local textures, structural features, and object information in remote sensing images. This approach helps capture important details and patterns necessary for accurate super-resolution reconstruction. Additionally, using smaller-sized images reduces computational complexity and memory consumption.

Figure 5 illustrates examples of the HR–LR pairs. To evaluate the generalization capability of our proposed model, we constructed four test sets by randomly selecting 120 images from the NaSC-TG2 dataset, 1000 images from the Satellite Image of Hurricane Damage dataset, 1890 images from the NWPU-RESISC45 dataset, and 420 images from the UCMerced LandUse dataset. These diverse datasets provide a representative sample of remote sensing images, enabling us to assess how well our model performs on different types of scenes and objects. Through this comprehensive evaluation, we aim to demonstrate the robustness and effectiveness of our model in handling a variety of remotely sensed image scenes.

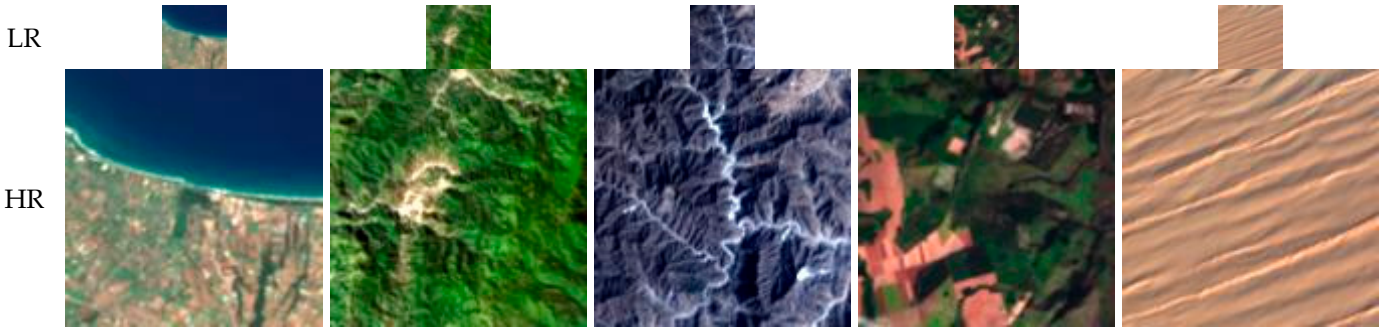


Figure 5. Examples of the HR–LR pair.

3.2. Experimental Environment and Parameter Settings

In this study, the experimental environment was set up on an Ubuntu operating system, equipped with a high-performance GeForce RTX 2080Ti GPU for efficient computation. The programming language utilized for code development is Python, while the Pytorch framework (available at https://pytorch.org/ (accessed on 1 July 2023)) was employed for effective algorithm modeling and implementation. The IESRGAN network architecture comprises two primary components: the generator network and the discriminator network. To conduct the experiments, a total of 19,800 HR remote sensing images from the NaSC-TG2 dataset were employed as the target images. As an initial step, a bicubic interpolation down-sampling technique was applied to generate a corresponding set of 19,800 LR remote sensing images required for input purposes. Subsequently, these LR images were fed into the SR-RRDB model, which consists of the generator network designed for training purposes. A comprehensive overview of the initial experimental details pertaining to the SR-RRDB model training can be found in Table 3.

Table 3. SR-RRDB: experimental details.

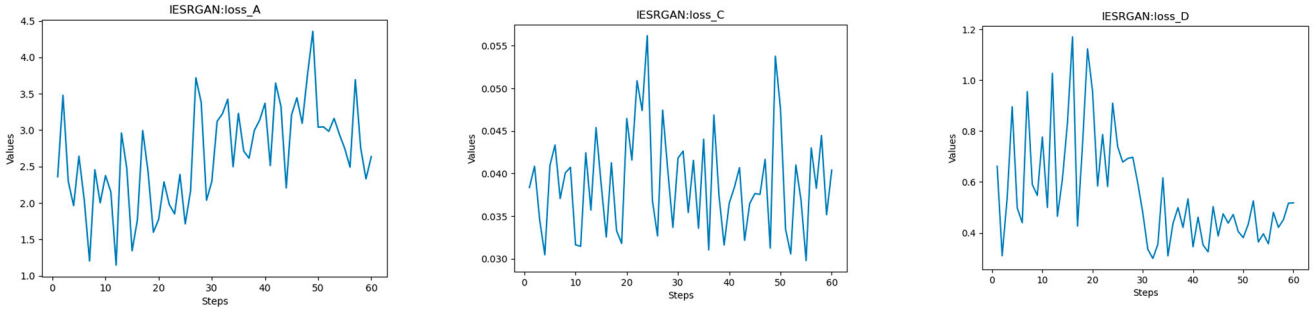
Parameter	Value
Scaling-factor	4
Batch size	16
Epochs	80
Learning rate	1×10^{-3}
Optimization method	Adam

The Cosine Annealing Learning Rate Schedule (CosineAnnealingLR) scheduler combined with the Adam optimizer was employed to effectively adjust learning rates during the training process. This method allows for the gradual reduction of learning rates, which in turn leads to enhanced convergence and ultimately improves the overall performance and generalization capability of the model. Upon completing this stage, the SR images generated by the well-trained SR-RRDB model were then introduced into a discriminator network that was designed based on the U-Net architecture. The purpose of this step was to efficiently discriminate between real HR images and those produced by the SR-RRDB model. Starting with the initialization of the SR-RRDB model, further experimental details pertaining to IESRGAN model training can be observed in Table 4. Notably, when the training reached its halfway point, there was an adjustment made wherein the learning rate was deliberately reduced to a half of its initial value. This strategic modification has been found to contribute significantly towards optimizing and refining both model performance and generalization effectiveness throughout the training process.

Table 4. Discriminator structure.

Parameter	Value
Scaling-factor	4
Batch size	16
Epochs	80
β	1×10^{-3}
Learning rate	1×10^{-4}
Optimization method	Adam

Figure 6 below shows the change curves of content loss, generation loss, and discriminative loss, respectively, throughout the training process.



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

Figure 6. Training loss variation curve. (**a**) Content loss value; (**b**) generation loss value; (**c**) discriminant loss value.

3.3. Experimental Evaluation Metrics

The Peak Signal-to-Noise Ratio (PSNR) [42] and Structural Similarity Index (SSIM) [43] have been used as standard evaluation metrics in image SR. Nevertheless, as revealed in some recent studies [44], super-resolved images may sometimes have high PSNR and SSIM scores with over-smoothed results but tend to lack realistic visual results. In this study, apart from the PSNR and SSIM, the learned perceptual image patch similarity (LPIPS) [45] is included in our experiments.

PSNR is used to evaluate pixel-wise differences between images. A higher PSNR value indicates a smaller difference between the processed image and the real image, implying better image quality. Its formula is:

$$PSNR = 10 \times \log_{10}\left(\frac{MAX^2}{MSE}\right) \tag{13}$$

In this formula, MAX represents the maximum pixel value, and MSE denotes the mean squared error between the reference image and the evaluated image. Its formula is given by:

$$MSE = \frac{1}{N}\sum_{i=1}^{n}(I_i - P_i)^2 \tag{14}$$

Here, N refers to the total number of pixels, while I_i and P_i represent the ith pixel values of the reference image and evaluated image, respectively.

SSIM takes into account factors such as the brightness, contrast, and structure of an image. Its formula is expressed as:

$$SSIM(X,Y) = \frac{(2u_X u_Y + C_1)(2\sigma_{XY} + C_2)}{\left(u_X^2 + u_Y^2 + C_1\right)\left(\delta_X^2 + \delta_Y^2 + C_2\right)} \tag{15}$$

The SSIM value ranges from [0,1] with higher values indicating better image quality.

LPIPS measures perceptual differences between two images, i.e., visual similarity between generated images and real images. A lower LPIPS score indicates a higher similarity between two images. Its formula is as follows:

$$d(x, x_0) = \sum_l \frac{1}{H_l W_l}\sum_{h,w} \parallel w_l \odot \left(\hat{y}_{hw}^l - \hat{y}_{0hw}^l\right) \parallel_2^2 \tag{16}$$

In the above equation, x and x_0 represent generated images and real images, respectively; $\hat{y}_{hw}^l$ denotes predicted feature maps for x at spatial position (h, w) and feature map l; $\hat{y}_{0hw}^l$ represents predicted feature maps for x_0 at the same spatial position and feature map. The weight matrix w_l is learned by the network to emphasize or de-emphasize certain features in an image.

3.4. Quantitative and Qualitative Comparison of Different Methods

In this section, an in-depth comparison is conducted between the proposed method and several classical single-image SR algorithms on four distinct test sets, focusing on their performance metrics. The SR algorithms under consideration encompass three CNN-based methods, specifically VDSR [17], SRResNet [22], and TESR [30], as well as two GAN-based methods, namely SRGAN [22] and ESRGAN [23]. Each of these methods has been meticulously optimized on the training set to guarantee the best possible performance and to ensure a fair comparison. To facilitate a more comprehensive comparison with both CNN-based and GAN-based algorithms, two networks are trained: SR-RRDB and IESRGAN. The proposed SR-RRDB is primarily a CNN-based algorithm that consists solely of the generator network. When trained exclusively with pixel loss, it can independently reconstruct HR images corresponding to LR ones. However, this approach may lack human perception since it relies solely on pixel loss for optimization. Therefore, a fair comparison between SR-RRDB and other CNN-based algorithms is made to evaluate their performance. On the other hand, the proposed IESRGAN is constructed upon a GAN network model, comprising both generator and discriminator networks. Its loss function incorporates perceptual loss through an innovative fusion method, which significantly enhances visual quality as perceived by the human eye. Thus, a fair comparison between IESRGAN and other GAN-based algorithms is conducted to assess their ability in delivering visually appealing results. In summary, this section aims to provide an extensive evaluation of the proposed method against traditional single-image SR algorithms in terms of performance metrics across four test sets. By comparing both CNN-based and GAN-based approaches using two different networks (SR-RRDB and IESRGAN), we strive to present a balanced analysis that highlights the strengths and limitations of each method while ensuring fairness in comparisons.

In this study, three metrics are employed to quantitatively evaluate the SR results, namely PSNR, SSIM, and LPIPS. The best results in each row are highlighted in red for easy comparison. As demonstrated in Table 5, the highest score in the PSNR metric is

achieved by the SR-RRDB method. Here it is noted that a higher PSNR value indicates a lower difference between the reconstructed image and the real image, ultimately resulting in superior image quality. As shown in Table 6, the highest score on the SSIM metric is also attained by the SR-RRDB method. A higher SSIM value suggests a greater similarity in brightness, contrast, and structure within a range of [0,1], indicating better preservation of these attributes during the super resolution process. Meanwhile, as displayed in Table 7, IESRGAN performs best on the LPIPS metric; a lower LPIPS value implies higher visual perceptual similarity between generated and real images. CNN-based SR methods offer advantages in terms of PSNR and SSIM due to their emphasis on preserving LR images' spatial structure. Consequently, super-resolution outcomes from CNN-based methods tend to lack realistic visual effects, leading to poor LPIPS performance. In contrast, GAN-based SR methods achieve better LPIPS performance while maintaining good PSNR and SSIM scores as they adopt adversarial loss and perceptual loss to encourage visually appealing results that closely resemble real images.

Table 5. Average PSNR/dB for different algorithms on the test sets selected by the NaSC-TG2, Satellite Image of Hurricane Damage, NWPU-RESISC45, and UCMerced LandUse. Red represents the best results.

Dataset	Bicubic	VDSR	SRResNet	SRGAN	ESRGAN	TESR	SR-RRDB (Proposed)	IESRGAN (Proposed)
1st test set	22.854	30.811	30.008	26.945	31.200	31.893	36.100	33.371
2nd test set	25.085	31.362	31.518	29.279	30.432	32.110	36.573	34.995
3rd test set	20.406	31.793	28.703	26.784	32.007	30.681	34.746	32.880
4th test set	19.259	29.123	28.238	26.468	30.401	31.004	34.315	31.952

Table 6. Average SSIM for different algorithms on the test sets selected by the NaSC-TG2, Satellite Image of Hurricane Damage, NWPU-RESISC45, and UCMerced LandUse. Red represents the best results.

Dataset	Bicubic	VDSR	SRResNet	SRGAN	ESRGAN	TESR	SR-RRDB (Proposed)	IESRGAN (Proposed)
1st test set	0.853	0.763	0.841	0.774	0.822	0.890	0.898	0.852
2nd test set	0.824	0.838	0.853	0.771	0.806	0.848	0.920	0.896
3rd test set	0.567	0.754	0.769	0.698	0.702	0.807	0.857	0.833
4th test set	0.757	0.727	0.777	0.717	0.772	0.850	0.859	0.832

Table 7. Average LPIPS of different algorithms on the test sets selected by the NaSC-TG2, Satellite Image of Hurricane Damage, NWPU-RESISC45, and UCMerced LandUse. Red represents the best results.

Dataset	Bicubic	VDSR	SRResNet	SRGAN	ESRGAN	TESR	SR-RRDB (Proposed)	IESRGAN (Proposed)
1st test set	0.289	0.286	0.252	0.131	0.186	0.205	0.198	0.091
2nd test set	0.305	0.373	0.320	0.174	0.224	0.394	0.257	0.134
3rd test set	0.453	0.298	0.402	0.244	0.256	0.249	0.326	0.202
4th test set	0.212	0.262	0.360	0.236	0.235	0.232	0.293	0.190

Figure 7 presents a comprehensive and intuitive comparison that enables a more profound comprehension of the quantitative results obtained in this study. Bicubic interpolation, as a traditional method, fails to generate any additional details or enhance image quality significantly. On the other hand, CNN-based super-resolution reconstruction algorithms, such as VDSR, SRResNet, and TESR, demonstrate relatively better performance in reconstructing some texture details by leveraging advanced learning techniques; however, they still suffer from contour blurring issues primarily due to the adoption of simplistic optimization strategies in their objective functions. In contrast, GAN-based super-resolution reconstruction algorithms like SRGAN and ESRGAN showcase notable

advantages in terms of visual effects and overall image enhancement. Nevertheless, these methods may inadvertently introduce artificial artifacts during the reconstruction process, which could potentially compromise the final output quality. The approach proposed here addresses these limitations by effectively recovering finer texture details compared to other SR methods available in the literature. Consequently, our method generates more realistic and visually appealing results that closely resemble natural images. This superior performance can be attributed to the innovative techniques employed in our algorithm design, which strike a delicate balance between optimizing visual quality and minimizing unwanted artifacts.

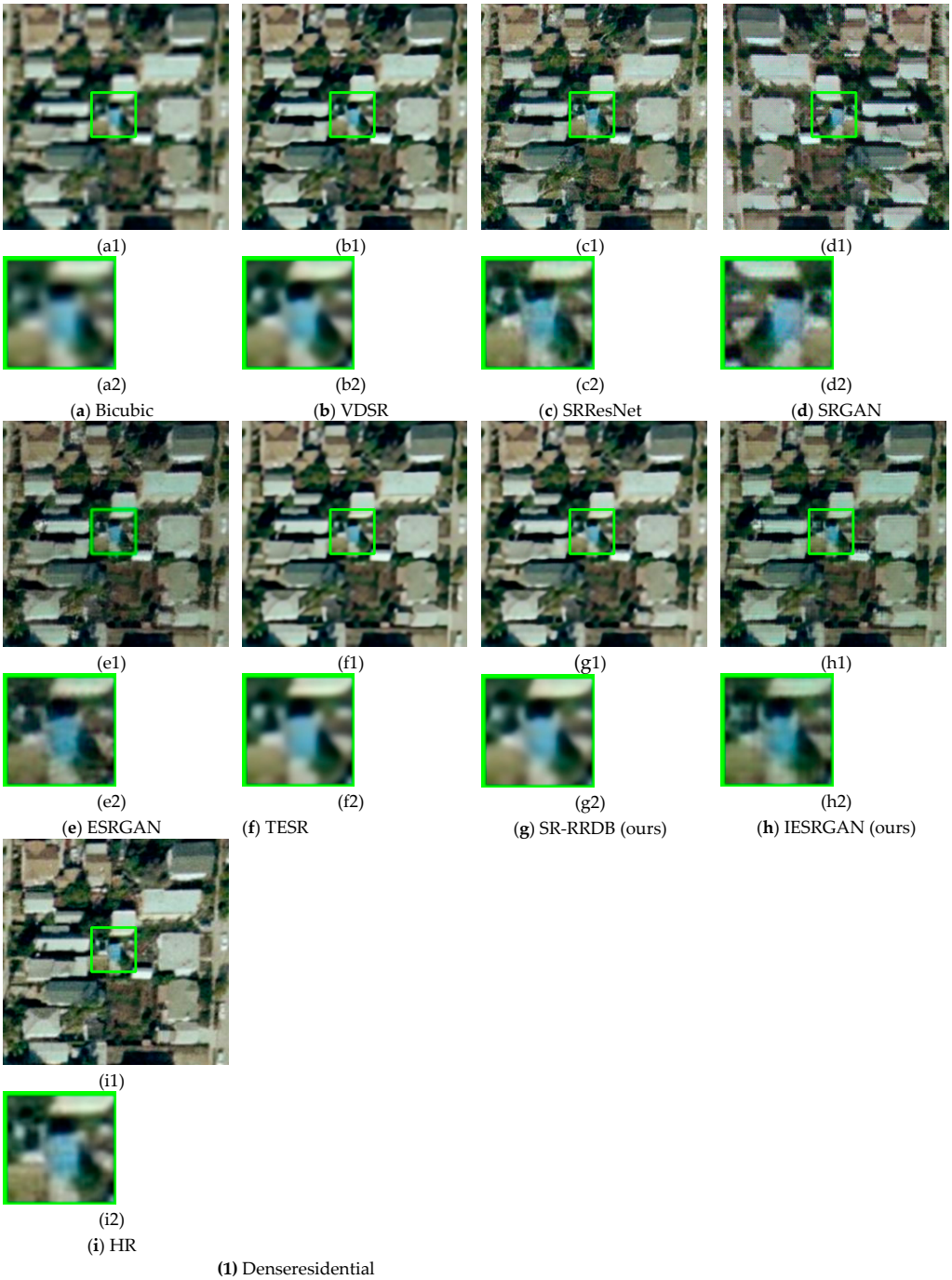


Figure 7. *Cont.*

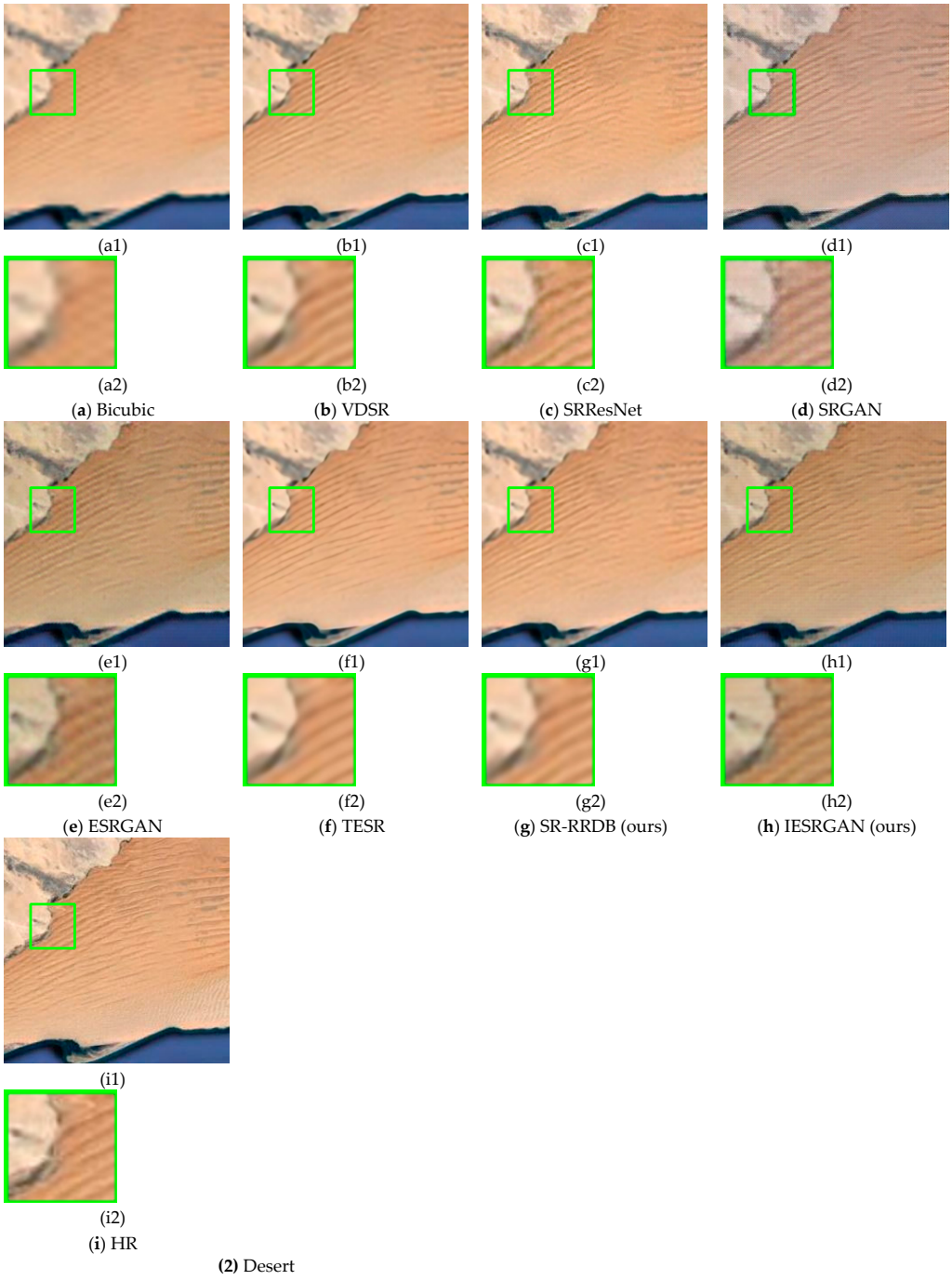


Figure 7. Visual comparison of our method with different SR methods on the test set. (1) Denseresidential; (2) Desert. (**a**) Bicubic result, (a2) is an enlarged view of the green area on (a1); (**b**) VDSR result, (b2) is an enlarged view of the green area on (b1); (**c**) SRResNet result, (c2) is an enlarged view of the green area on (c1); (**d**) SRGAN result, (d2) is an enlarged view of the green area on (d1); (**e**) ESRGAN result, (e2) is an enlarged view of the green area on (e1); (**f**) TESR result, (f2) is an enlarged view of the green area on (f1); (**g**) SR-RRDB (ours) result, (a2) is an enlarged view of the green area on (a1); (**h**) IESRGAN (ours) result (h2) is an enlarged view of the green area on (h1); (**i**) HR image, (i2) is an enlarged view of the green area on (i1).

3.5. Ablation Studies

In order to assess the effectiveness of the enhancements introduced by each component of our proposed method, a series of ablation experiments was performed. In these experiments, we gradually incorporated the RRDB strategy, Reflection Padding layer

(ReflectionPad), and U-Net structure into the baseline model. All models were trained using an identical configuration, and their performance was evaluated on a test set. The comparative data for various metrics are presented in Table 8, which clearly demonstrates an overall improvement in model performance throughout the refinement process.

Table 8. Ablation studies results. The higher the PSNR and SSIM scores, the better; the lower the LPIPS score, the better. Red represents the best results.

Test Dataset	Metric	Baseline	RRDB without ReflectionPad	RRDB with ReflectionPad	U-Net
1st test set	PSNR	30.008	32.024	32.623	33.279
	SSIM	0.841	0.801	0.858	0.864
	LPIPS	0.203	0.126	0.112	0.099
2nd test set	PSNR	30.047	32.205	32.967	34.371
	SSIM	0.807	0.866	0.869	0.852
	LPIPS	0.213	0.167	0.145	0.134
3rd test set	PSNR	29.703	31.731	32.378	32.878
	SSIM	0.769	0.726	0.899	0.823
	LPIPS	0.258	0.216	0.207	0.203
4th test set	PSNR	29.238	30.830	31.143	31.835
	SSIM	0.777	0.803	0.814	0.832
	LPIPS	0.218	0.198	0.203	0.192

Initially, increasing the number of RRDBs effectively contributes to enhancing image details and high-frequency information. This enhancement is achieved by mapping the image from an LR to an HR space through a deep network structure. Consequently, more image details are recovered, resulting in notable improvements in PSNR, SSIM, and LPIPS scores. Subsequently, adding a Reflection Padding layer on top of this foundation helps preserve edge information within the input image while reducing edge information loss. Edge information plays a critical role in generating HR images since it often contains high-frequency detail information that influences the level of detail present in the generated results. By introducing the Reflection Padding layer into our model, we achieve optimal SSIM values indicative of relatively ideal structural reconstruction effects. Lastly, incorporating a U-Net structure into the discriminator enables it to capture and integrate image features across multiple resolution levels more effectively. This enhanced capability assists in distinguishing generated images from real ones while simultaneously improving reconstructed image quality. In conjunction with our adopted fusion loss approach, this results in superior LPIPS values and improved perceptual quality for human observers. At the same time, both the PSNR and SSIM scores exhibit some degree of improvement as well—evidence that our model delivers higher-quality images. In summary, following these step-by-step enhancements to our initial design, our proposed method achieves significant improvements across all relevant metrics—thereby validating the effectiveness of each modification introduced.

4. Discussion

Remote sensing images have rich and complex scenes and different target features, and many existing algorithms have difficulty recovering these details accurately. To overcome this challenge, we propose an Enhanced U-Net Structured Generative Adversarial Network for Remote Sensing Image Super-Resolution Reconstruction (IESRGAN). IESRGAN consists of two parts; the first part is based on the RRDB module to improve the generator network to reconstruct the texture features of remote sensing images while preserving as many global details as possible. The second part is an improved discriminator network based

on the U-Net network, which has jump connections and can fuse shallow features with deep features directly. These are important for generating high-quality images, especially in tasks that require the generation of fine structures and textures. The results of our proposed IESRGAN model show good performance on NaSC-TG2, Satellite Image of Hurricane Damage, NWPU-RESISC45, and UCMerced LandUse datasets in terms of visual perception and quantitative measurements. In general, our proposed model outperforms other methods and provides a new approach for super-resolution reconstruction of remote sensing images.. There are several limitations of this work that need to be noted. First, the proposed algorithm is specifically designed for remotely sensed images and may not perform as well on other types of images. Second, we performed super-resolution reconstructions of remotely sensed images with a magnification factor of x4, which is not satisfactory for higher magnification factors such as x8.

5. Conclusions

Extensive experimental results show that the IESRGAN model performs well in quantitative evaluation metrics (such as PSNR, SSIM, and LPIPS) under different real remote sensing image datasets and thus has remarkable stability and generalization ability. The IESRGAN algorithm can provide a promising idea for the super-resolution reconstruction of remote sensing images, which can be applied to feature recognition classification, land detection, etc. There are several potential directions for future work in the proposed remote sensing super-resolution algorithm IESRGAN. A key area is the application of the algorithm to super-resolution reconstructions of remote sensing images at high magnifications (e.g., $\times 8$), aiming at better practical applications. In addition, the fusion of multi-source remote sensing image information can be explored to fully exploit the complementary information between different sources, thus improving the effectiveness of remote sensing image reconstruction. Finally, it would be beneficial to investigate the application of the algorithm in real-world processing tasks, such as land monitoring and object classification. By addressing these challenges, we will continue to advance the field of super-resolution reconstruction of remote sensing images and expand its applicability in various fields.

Author Contributions: Conceptualization, X.Y.; methodology, X.Y.; software, L.M.; validation, L.D.; formal analysis, X.Y.; investigation, L.M.; resources, X.Y.; data curation, L.D.; writing—original draft preparation, X.Y.; writing—review and editing, D.L. and J.A.B.; visualization, X.Y.; supervision, D.L. and J.A.B.; project administration, D.L.; funding acquisition, L.W. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by Leading Talents Project of the State Ethnic Affairs Commission and National Natural Science Foundation of China (No. 62071084).

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.

Acknowledgments: The authors sincerely thank the academic editors and reviewers for their useful comments and constructive suggestions.

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

Abbreviations

The following abbreviations are used in this manuscript:

HR	High resolution
LR	Low resolution
SR	Super resolution
CNN	Convolutional neural network
GAN	Generating adversarial network
RRDB	Residual-in-Residual Dense Block

ReflectionPad	Reflection padding
ReLU	Rectified linear unit
VGG	Very deep convolutional networks
CosineAnnealingLR	CosineAnnealingLR
CosineAnnealingLR	CosineAnnealingLR
PSNR	Peak Signal-to-Noise Ratio
SSIM	Structural similarity
LPIPS	Learned perceptual image patch similarity
VDSR	Very Deep Super-Resolution
SRResNet	Super-resolution residual network
SRGAN	Super-resolution generative adversarial network
ESRGAN	Enhanced super-resolution generative adversarial network

References

1. Wang, L.G.; Zhao, C.H. *Hyperspectral Image Processing Techniques*; National Defense Industry Press: Beijing, China, 2013.
2. Schowengerdt, R.A. *Remote Sensing Image Processing Models and Methods*; Electronic Industry Press: Beijing, China, 2010.
3. Hu, Y. Research on Super-Resolution Reconstruction Technology of Remote Sensing Images. Ph.D. Thesis, PLA Information Engineering University, Zhengzhou, China, 2004.
4. Jiang, K.; Wang, Z.; Yi, P.; Wang, G.; Lu, T.; Jiang, J. Edge-Enhanced GAN for Remote Sensing Image Superresolution. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 5799–5812. [CrossRef]
5. Blu, T.; Thevenaz, P.; Unser, M. Linear Interpolation Revitalized. *IEEE Trans. Image Process.* **2004**, *13*, 710–719. [CrossRef] [PubMed]
6. Zhang, X.-G. A New Kind of Super-Resolution Reconstruction Algorithm Based on the ICM and the Bicubic Interpolation. In Proceedings of the 2008 International Symposium on Intelligent Information Technology Application Workshops, Shanghai, China, 21–22 December 2008; pp. 817–820. [CrossRef]
7. Zhang, L.; Wu, X. An edge-guided image interpolation algorithm via directional filtering and data fusion. *IEEE Trans. Image Process.* **2006**, *15*, 2226–2238. [CrossRef]
8. Liang, X.; Gan, Z. Improved Non-local Iterative Back-Projection Method for Image Super-Resolution. In Proceedings of the 2011 Sixth International Conference on Image and Graphics, Hefei, China, 12–15 August 2011; pp. 176–181. [CrossRef]
9. Xi, H.; Xiao, C.; Bian, C. Edge Halo Reduction for Projections onto Convex Sets Super Resolution Image Reconstruction. In Proceedings of the 2012 International Conference on Digital Image Computing Techniques and Applications (DICTA), Fremantle, WA, Australia, 3–5 December 2012; pp. 1–7. [CrossRef]
10. Mofidi, M.; Hajghassem, H.; Afifi, A. Adaptive image super-resolution via controlled weighting coefficients of a maximum-a-posteriori estimator. *J. Electron. Imaging* **2018**, *27*, 043031. [CrossRef]
11. Fernandez-Beltran, R.; Latorre-Carmona, P.; Pla, F. Single-frame super-resolution in remote sensing: A practical overview. *Int. J. Remote Sens.* **2016**, *38*, 314–354. [CrossRef]
12. Lim, B.; Son, S.; Kim, H.; Nah, S.; Lee, K.M. Enhanced deep residual networks for single image super-resolution. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Honolulu, HI, USA, 21–26 July 2017; pp. 136–144. [CrossRef]
13. Dong, C.; Loy, C.C.; He, K.; Tang, X. Image Super-Resolution Using Deep Convolutional Networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2016**, *38*, 295–307. [CrossRef] [PubMed]
14. 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 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, June 26 to July 1 2016; pp. 1874–1883.
15. Tammina, S. Transfer learning using VGG-16 with deep convolutional neural network for classifying images. *Int. J. Sci. Res. Publ. (IJSRP)* **2019**, *9*, 143–150. [CrossRef]
16. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778. [CrossRef]
17. 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.
18. Li, J.; Fang, F.; Mei, K.; Zhang, G. Multi-scale residual network for image super-resolution. In Proceedings of the European Conference on Computer Vision, Munich, Germany, 8–14 September 2018; Springer: Munich, Germany, 2018; pp. 527–542. [CrossRef]
19. Lan, R.; Sun, L.; Liu, Z.; Lu, H.; Su, Z.; Pang, C.; Luo, X. Cascading and Enhanced Residual Networks for Accurate Single-Image Super-Resolution. *IEEE Trans. Cybern.* **2020**, *51*, 115–125. [CrossRef]
20. 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 (ECCV)*; Springer: Munich, Germany, 2018; pp. 294–310.
21. Goodfellow, I.; Pouget-Abadie, J.; Mirza, M. Generative Adversarial Networks. *Gener. Advers. Nets* **2014**, *2672*, 2680.

22. Ledig, C.; Theis, L.; Huszár, F.; Caballero, J.; Cunningham, A.; Acosta, A.; Aitken, A.P.; Tejani, A.; Totz, J.; Wang, Z.; et al. Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017.
23. Wang, X.; Yu, K.; Wu, S.; Gu, J.; Liu, Y.; Dong, C.; Qiao, Y.; Loy, C.C. ESRGAN: Enhanced Super-Resolution Generative Adversarial Networks. In Proceedings of the 15th European Conference on Computer Vision, ECCV 2018, Munich, Germany, 8–14 September 2018; pp. 63–79.
24. Rabbi, J.; Ray, N.; Schubert, M.; Chowdhury, S.; Chao, D. Small-Object Detection in Remote Sensing Images with End-to-End Edge-Enhanced GAN and Object Detector Network. *Remote Sens.* **2020**, *12*, 1432. [CrossRef]
25. Ma, W.; Pan, Z.; Guo, J.; Lei, B. Super-Resolution of Remote Sensing Images Based on Transferred Generative Adversarial Network. In Proceedings of the IGARSS 2018—2018 IEEE International Geoscience and Remote Sensing Symposium, Valencia, Spain, 22–27 July 2018; pp. 1148–1151. [CrossRef]
26. Li, Y.; Mavromatis, S.; Zhang, F.; Du, Z.; Sequeira, J.; Wang, Z.; Zhao, X.; Liu, R. Single-Image Super-Resolution for Remote Sensing Images Using a Deep Generative Adversarial Network with Local and Global Attention Mechanisms. *IEEE Trans. Geosci. Remote Sens.* **2021**, *60*, 1–24. [CrossRef]
27. Salgueiro, L.; Marcello, J.; Vilaplana, V. SEG-ESRGAN: A Multi-Task Network for Super-Resolution and Semantic Segmentation of Remote Sensing Images. *Remote Sens.* **2022**, *14*, 5862. [CrossRef]
28. Zhu, F.; Wang, C.; Zhu, B.; Sun, C.; Qi, C. An improved generative adversarial networks for remote sensing image super-resolution reconstruction via multi-scale residual block. *Egypt. J. Remote Sens. Space Sci.* **2023**, *26*, 151–160. [CrossRef]
29. Zhao, J.; Ma, Y.; Chen, F.; Shang, E.; Yao, W.; Zhang, S.; Yang, J. SA-GAN: A Second Order Attention Generator Adversarial Network with Region Aware Strategy for Real Satellite Images Super Resolution Reconstruction. *Remote Sens.* **2023**, *15*, 1391. [CrossRef]
30. Ali, A.M.; Benjdira, B.; Koubaa, A.; Boulila, W.; El-Shafai, W. TESR: Two-Stage Approach for Enhancement and Super-Resolution of Remote Sensing Images. *Remote Sens.* **2023**, *15*, 2346. [CrossRef]
31. Pan, Y.; Liu, D.; Wang, L.; Benediktsson, J.A.; Xing, S. A Pan-Sharpening Method with Beta-Divergence Non-Negative Matrix Factorization in Non-Subsampled Shear Transform Domain. *Remote Sens.* **2022**, *14*, 2921. [CrossRef]
32. Pan, Y.; Liu, D.; Wang, L.; Xing, S.; Benediktsson, J.A. A Multispectral and Panchromatic Images Fusion Method Based on Weighted Mean Curvature Filter Decomposition. *Appl. Sci.* **2022**, *12*, 8767. [CrossRef]
33. Yi, C.; Zhao, Y.-Q.; Chan, J.C.-W. Hyperspectral Image Super-Resolution Based on Spatial and Spectral Correlation Fusion. *IEEE Trans. Geosci. Remote Sens.* **2018**, *56*, 4165–4177. [CrossRef]
34. Feng, X.; Su, X.; Shen, J.; Jin, H. Single Space Object Image Denoising and Super-Resolution Reconstructing Using Deep Convolutional Networks. *Remote Sens.* **2019**, *11*, 1910. [CrossRef]
35. Huang, G.; Liu, Z.; Van Der Maaten, L.; Weinberger, K.Q. Densely connected convolutional networks. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 2261–2269. [CrossRef]
36. Chen, Y.; Li, J.; Xiao, H.; Jin, X.; Yan, S.; Feng, J. Dual path networks. *arXiv* **2017**, arXiv:1707.01629.
37. Miyato, T.; Kataoka, T.; Koyama, M.; Yoshida, Y. Spectral normalization for generative adversarial networks. *arXiv* **2018**, arXiv:1802.05957.
38. Zhou, Z.; Li, S.; Wu, W.; Guo, W.; Li, X.; Xia, G.; Zhao, Z. NaSC-TG2: Natural Scene Classification with Tiangong-2 Remotely Sensed Imagery. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2021**, *14*, 3228–3242. [CrossRef]
39. Cao, Q.D.; Choe, Y. Building Damage Annotation on Post-Hurricane Satellite Imagery Based on Convolutional Neural Networks. *arXiv* **2018**, arXiv:1807.01688. [CrossRef]
40. Cheng, G.; Han, J.; Lu, X. Remote Sensing Image Classification: Benchmark and State of the Art. *Proc. IEEE* **2017**, *105*, 1865–1883. [CrossRef]
41. Yang, Y.; Newsam, S. Bag-of-visual-words and spatial extensions for land-use classification. In Proceedings of the 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems, San Jose, CA, USA, 3–5 November 2010; pp. 270–279. [CrossRef]
42. Korhonen, J.; You, J. Peak Signal-to-Noise Ratio Revisited: Is simple beautiful? In Proceedings of the 2012 Fourth International Workshop on Quality of Multimedia Experience, Melbourne, Australia, 5–7 July 2012; pp. 37–38.
43. 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]
44. Blau, Y.; Mechrez, R.; Timofte, R.; Michaeli, T.; Zelnik-Manor, L. The 2018 PIRM challenge on perceptual image super-resolution. In Proceedings of the European Conference on Computer Vision (ECCV) Workshops; Springer: Berlin/Heidelberg, Germany; 2018. [CrossRef]
45. Zhang, R.; Isola, P.; Efros, A.A.; Shechtman, E.; Wang, O. The Unreasonable Effectiveness of Deep Features as a Perceptual Metric. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 586–595.