The Egyptian Journal of Remote Sensing and Space Sciences 26 (2023) 151–160

HOSTED BY

ELSEVIER

Contents lists available at ScienceDirect

The Egyptian Journal of Remote Sensing and Space Sciences

journal homepage: www.sciencedirect.com

Research Paper

An improved generative adversarial networks for remote sensing image super-resolution reconstruction via multi-scale residual block

Fuzhen Zhu [a,*], Chen Wang [a], Bing Zhu [b], Ce Sun [a], Chengxiao Qi [a]

[a] *Key laboratory of Remote Sensing Image Processing, Electronic Engineering College, Heilongjiang University, Harbin 150080, PR China*
[b] *Institute of Image Information Technology and Engineering, Harbin Institute of Technology, Heilongjiang, Harbin 150001, PR China*

ARTICLE INFO

Article history:
Received 12 August 2022
Revised 11 November 2022
Accepted 26 December 2022
Available online 9 January 2023

Keywords:
Super-resolution reconstruction
GAN
RaGAN
Multi-scale residual block

ABSTRACT

Existing image super-resolution algorithms still suffer from the problems of not extracting rich image features and losing realistic high-frequency details. In order to solve these problems, this paper proposes an improved generative adversarial network algorithm for super-resolution reconstruction of remote sensing images by multi-scale residual blocks. The original generative adversarial network (GAN) structure is improved and multi-scale residual blocks are introduced in the generator to fuse features at different scales. After extracting the parallel information of multi-scale features, information is exchanged between multi-resolution information streams to obtain contextual information through spatial and channel attention mechanisms, and multi-scale features are fused according to the attention mechanism. In the discriminator, the concept of relative average GAN (RaGAN) is introduced, and the loss function of the network is redesigned so that the discriminator can predict relative probabilities instead of absolute probabilities thus enabling clear learning of edge and texture details. Experimental results show that the proposed method in this paper significantly outperforms state-of-the-art (SOTA) methods in terms of both subjective and objective metrics.In three test datasets, compared with SOTA methods, the Peak Signal to Noise Ratio(PSNR) is improved by a maximum of 1.18 dB, 0.84 dB and 1.29 dB respectively, and the Structural Similarity Index (SSIM) is improved by 0.0264, 0.0077 and 0.0109 respectively in scale of 2, 3 and 4 times images super-resolution.The model proposed in this paper effectively improved the super-resolution re-construction results of remote sensing images.

1. Introduction

Super-resolution reconstruction (SRR) is a method of generating high-resolution images (HRI) from a series of low-resolution images (LRI) using some algorithm without changing the imaging hardware conditions (Glasner et al., 2009; Zhou et al., 2016). Image super-resolution reconstruction technology has been widely used in military and civilian fields such as military target strikes (Guo et al., 2018), battlefield environment monitoring (Zhang et al., 2020), target identification and localization (Peng et al., 2016), ultra-high definition television, emergency event monitoring, and medical image diagnosis (Wang and Ying, 2021).

In 1984, Huang and Tsai introduced the concept of image SRR for the first time and realized image SRR on frequency domain. Since then, SRR techniques have been developed rapidly. At present, image SRR methods mainly include reconstruction-based super-resolution methods and learning-based super-resolution methods. The reconstruction-based super-resolution method first obtains the registration relationship between LRI and HRI, obtains the contribution of the gray value of each low-resolution pixel to the high-resolution pixel, and derives a set of equations connecting the high-resolution pixel vector and the low-resolution pixel vector. Finally, the resultant image of SSR is obtained by solving the equations. There are many typical reconstruction-based methods such as nonuniform interpolation (Rojo-Álvarez et al., 2007), iterative inverse projection (Wunsch and Hirzinger, 1996), convex set projection (Nedić, 2010), Bayesian analysis (Berger et al., 1994), adaptive filtering (Diniz, 1997), etc. In contrast, learning-based super-resolution reconstruction techniques use the relational relationship between LRI and HRI to reconstruct super-resolution images, which can maximize the use of prior knowledge, generate

Abbreviations: SRR, super-resolution reconstrution; GAN, generative adversarial network; RaGAN, relative average generative adversarial network; PSNR, Peak Signal to Noise Ratio; SSIM, Structural Similarity Index.

[*] Corresponding author.

E-mail addresses: zhufuzhen@hlju.edu.cn (F. Zhu), 2211737@s.hlju.edu.cn (C. Wang), zhubing@hit.edu.cn (B. Zhu), 2201661@s.hlju.edu.cn (C. Sun), 2191314@s.hlju.edu.cn (C. Qi).

https://doi.org/10.1016/j.ejrs.2022.12.008
1110-9823/© 2022 National Authority of Remote Sensing & Space Science. Published by Elsevier B.V.

new high-frequency detail information, and obtain better results with the same number of LRI samples. Typical learning-based SRR methods are as follows: Markov network learning (Diniz, 1997), image pyramid learning (Zeng et al., 2019), streaming learning (Gomes et al., 2019) sparse coding (Olshausen and Field, 2004), and deep learning, etc. In recent years, the application of deep learning techniques in the field of single-frame image SRR has achieved better results. Dong et al. (2015) first proposed the SRCNN model for image SRR, which extracts the features of the original input image through multiple convolutional layers and learns the feature mapping relationship between LRI and HRI to obtain better SRR results than traditional super-resolution algorithms. To further solve the problems of insufficient detail information in the results, large computational effort, and slow reconstruction, Kim et al. (2016a) proposed the FSRCNN algorithm, which removes the upsampling layer and nonlinear mapping layer, and concatenates multiple small convolutional kernels, which can reduce the computation and extract features with different perceptual field sizes, and is suitable for small-size LRI processing. Kim et al. (2016b) further improved FSRCNN by using the teacher network (Lee et al., 2020) to extract intermediate features of images by secondary sampling of HR images passed to the student network for training, which substantially improved the performance of FSRCNN. Shi et al. (2016) proposed the ESPN algorithm to achieve direct LR feature extraction, which solved the problem of dramatically increasing the number of model parameters. Zhang et al. (2018) combined the channel attention mechanism with residuals to construct a deeper network to weaken a large amount of low-frequency information in LR images. Woo et al. (2018) fused channel and spatial attention modules to form the CBAM model (Ahn et al., 2018) shifted the center of the network to images with more feature information regions, focusing on extracting features with important information. Huang et al. (2018) proposed a densely connected network. Using jump-length (short) connections between layers to fully fuse feature information from different layers can alleviate gradient disappearance, reduce the number of parameters, enhance network stability, and further improve the image reconstruction. Chen et al. (2020) proposed AdderNet to extract high and low frequency information respectively by using additive operations to solve the problem of increased running memory and computation.

Ledig et al. (2017) proposed a super-resolution image reconstruction algorithm based on Generative Adversarial Networks (GAN), which got good SRR and higher operation speed. In the SRGAN (Ledig et al., 2017) algorithm, the generator network and discriminator network are used for adversarial training of the SRR network, with the generator being used to generate HR images and the discriminator being used to discriminate newly constructed HR images from the original HR images, as well as the inverse optimization of the generator and discriminator networks. At the same time, the traditional MSE loss function was replaced by perceptual loss (Yang et al., 2018; Li et al., 2020; Ouyang et al., 2019) to enhance the recovery details and ensure the high fidelity and quality of the reconstructed images. Wang et al. (2018) proposed an ESRGAN model based on SRGAN, which removes the BN layer to improve the perceptual loss and uses pre-activated features to improve the visual quality and obtain more natural textures. Ma et al. (2020) proposed a GAN-based gradient-guided SR network to avoid structural distortion of reconstructed images.

Above SRR models have relatively good peak signal-to-noise ratio (PSNR) and structural similarity index (SSIM). However, in contrast to these approaches that tend to build deeper and more complex networks, we make full use of the feature information of different resolution images and propose a multi-scale residual block to improve the GAN framework for remote sensing image SRR. The overall digram of our improved GAN is shown in Fig. 1. The main contributions of this paper are summarized as follows:

(1) A multi-scale residual blocks are introduced as the core blocks of the generator. Multi-scale parallel feature information can be shared among each other while selective kernel networks are employed to dynamically combine parallel branches to preserve the original feature information of each spatial resolution image.
(2) A dual attention unit (DAU) is used to capture semantic information in both spatial and channel dimensions. The channel attention mechanism and spatial attention mechanism are used and integrated into various existing deep learning networks as plug-and-play modules, reducing the parameters and computation of the network.
(3) The standard discriminator is replaced by a relativistic discriminator, and its loss function is designed by introducing RaGAN loss, which can make the SRR result images more realistic texture details.

2. Materials and Methods

2.1. Algorithm of this paper

In generator, a multi-scale residual block (Li et al., 2018; Gao et al., 2019; Qin et al., 2020; Liu et al., 2020) is introduced in this paper. Inside it, information fully interacts between the multi-scale feature maps, and a selective kernel feature fusion (SKFF) (Chen et al., 2022) mechanism is used to exchange information for each set of information streams.Instead of simply concatenating or accumulating features, this fusion approach selects useful kernel sets from each branch representative in the multi-scale residual block and dynamically combines parallel branches using SKFF, preserving the original information at each spatial resolution. The information captures contextual information in both spatial and channel dimensions through a dual attention unit (DAU) (Zamir et al., 2020; Li et al., 2020). At the end of the cascaded multiscale residual blocks, the convolutional layer acts as a bottleneck layer to chieve dimensionality reduction and fusion of features.

The standard discriminator replaced by the relativistic discriminator is a binary classification process used to predict the probability of true or false input images. And the relativistic discriminator predicts the probability that the image is more true than the false image generated by the generator, resulting in a reconstructed image with clearer and more realistic details.

2.2. Generator

In this paper, a generator based on multiscale residual blocks is designed as shown in Fig. 2. Low-resolution images I^{LR} is obtained by bicubic interpolation, using C as the image channel, here $C = 3$ represents the RGB channel and W and H are the width and height of the image respectively. The input dimension I^{LR} is $W \times H \times C$, and the dimensions of I^{HR}, I^{SR} are $rW \times rH \times C$, where r represents the upsampling factor. The final goal of the generator is to learn the mapping relationship from the low-resolution image I^{LR} to the super-resolution image I^{SR}.

2.2.1. Multi-scale residual block

The multiscale residual block is the core of the generator, and its structure is shown in Fig. 3. First, the features are downsampled at different scales, and three streams of information with different resolutions represent different levels of features. In this paper, multiscale residual blocks are introduced into the GAN network. Since the multi-scale feature maps exchange multi-resolution traf-

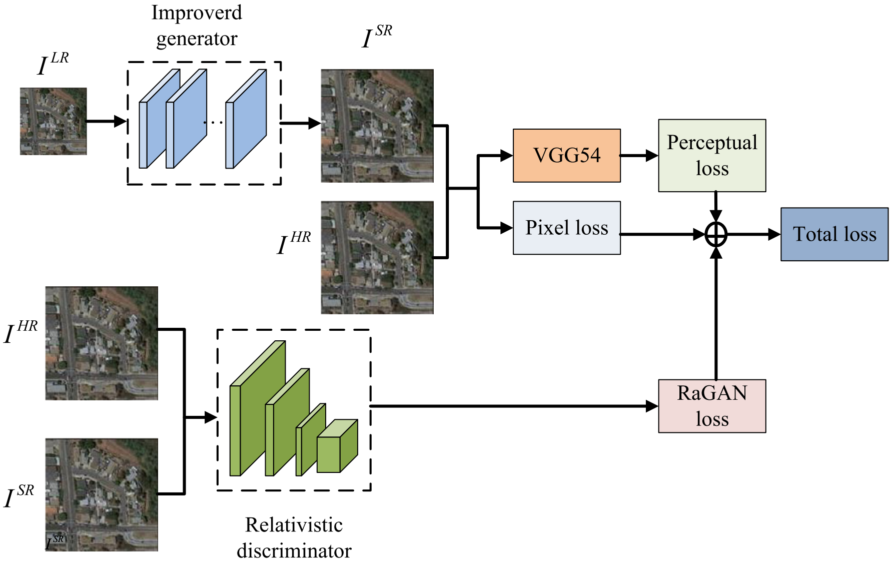


Fig. 1. An overall diagram of our improved GAN for remote sensing image SRR.

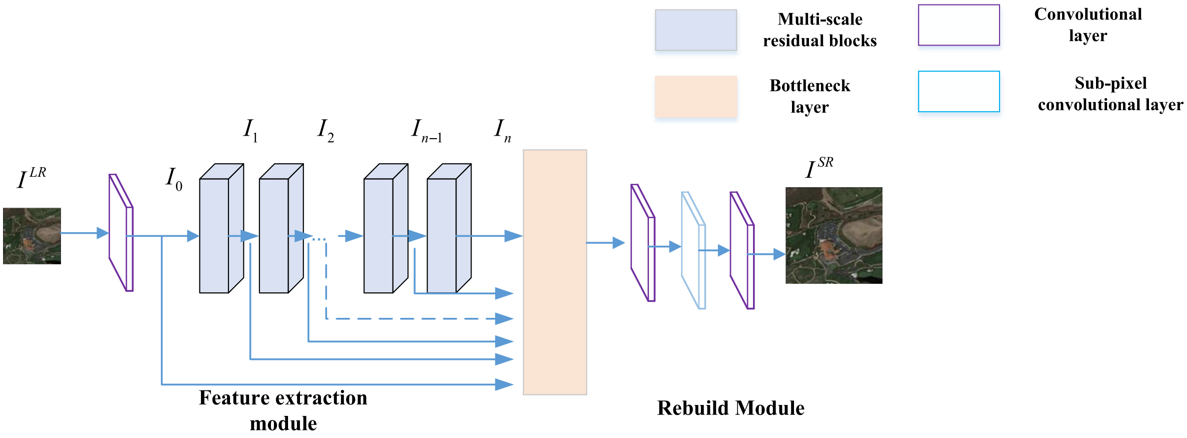


Fig. 2. The overall pipeline of our generator.

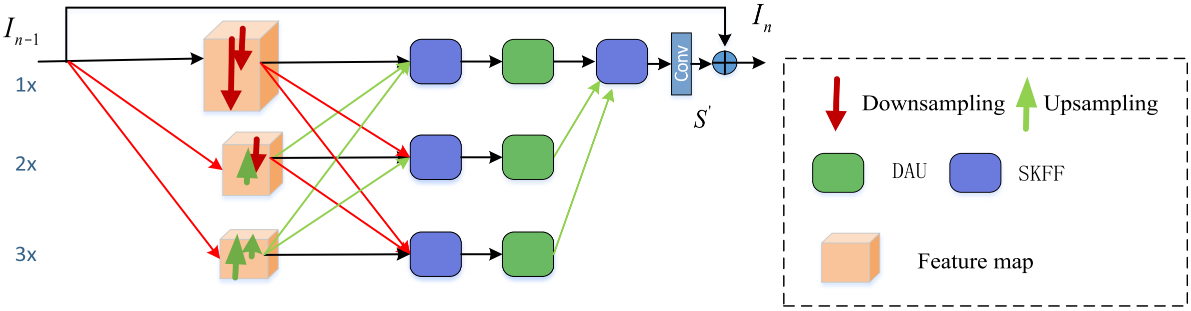


Fig. 3. The structure of the multi-scale residual block.

fic information in a fully interactive manner, the multi-scale residual blocks are able to aggregate the feature attention mechanisms from multiple branches via SKFF. Meanwhile, contextual information is captured by the DAU unit in both spatial and channel dimensions. The multiscale residual block contains three components: SKFF, DAU, and local residual learning (Shi et al., 2018; Hou et al., 2020; Liu and Lee, 2019), each of which is described in detail below.

- SKFF mechanism

The automatic selection operation of selective kernel (SK) convolution between multiple nuclei of different sizes enables neurons to adjust their receptive fields adaptively. As shown in Fig. 4, the SKFF module dynamically adjusts the receptive field by fusion and weight selection of multi-resolution streams. The fusion operation produces descriptions of global features by combining information

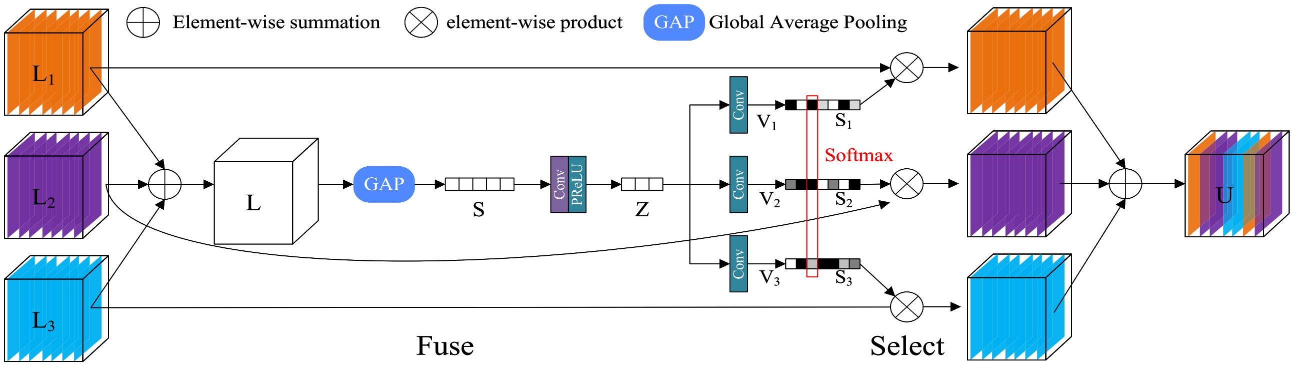


Fig. 4. Schematic for SKFF.

from multi-resolution streams, and the selection operation uses these descriptors to recalibrate the feature maps of different streams before aggregation.

Firstly, information from multiple branches is aggregated through fusion operations to obtain a global representation of the selection weights, which enable neurons to adjust the size of their receptive fields adaptively. The result of fusing branches by element summation is represented as $L = L_1 + L_2 + L_3$, then global average pooling (GAP) is applied to compute the channel statistic $s \in R^{1\times1\times C}$ in the spatial dimension of $L \in R^{H\times W\times C}$. In the experiments, the channel descending convolutional layer is applied to generate a compact feature representation $z \in R^{1\times1\times r}$, where $r = \frac{C}{8}$. Finally, the feature vector z corresponds to three resolution streams respectively through three parallel channel convolution layers. At the same time, three feature descriptors v_1, v_2 and v_3 are provided, with dimensions $U = s1 \cdot L_1 + s2 \cdot L_2 + s3 \cdot L_3$.

- DAU mechanism

In order to effectively capture the global dependencies of features, inspired by low-level vision methods based on attention mechanisms, DAU is introduced in this paper to extract features in the information stream, as shown in Fig. 5. The DAU mechanism allows only informative features to pass while suppressing less useful features. Channel attention and spatial attention mechanisms are applied to achieve feature recalibration.

The squeeze and excitation operations make channel attention branching using inter-channel relations of the convolutional feature map. Given a feature map $M \in R^{H\times W\times C}$, the squeeze operation applies GAP across spatial dimensions to encode global context, producing a feature descriptor $d \in R^{1\times1\times C}$. The excitation operator passes d through two convolution layers and a sigmoid function and generates an activation feature map $\hat{d} \in R^{1\times1\times C}$. Finally, the output of the CA branch is obtained by rescaling with the activated feature map $\hat{d}$.

The goal of the spatial attention (SA) branch is to generate a spatial attention map using the spatial interdependence of convolutional features to recalibrate the input features. The SA branch independently performs global average pooling and maximum pooling of features in the channel dimension and concatenates the outputs to form a feature map to generate a spatial attention map $f \in R^{H\times W\times 2}$. The map f is activated by convolution and sigmoid to obtain a spatial attention map $\hat{f} \in R^{H\times W\times 4}$, which will be used to rescale M.

- Local residual learning

To solve problems of gradient disappearance and gradient explosion caused by deeper network layers, and to alleviate the difficulty of training deeper networks, residual learning is chosen for each multi-scale residual block, which is achieved by jump joining and addition operation. It can be described by the following Eq. 1.

$$I_n = S' + I_{n-1} \tag{1}$$

where I_n and I_{n-1} represent the input and output of the multi-scale residual blocks respectively, and $S' + I_{n-1}$ are accumulated operations employing jump connections, residual learning reduces the computational complexity greatly and improves the stability of the network.

2.2.2. Hierarchical feature fusion structure

The features of the input image can be reconstructed by image SRR transfer to a deeper network. Since some features fade away as the network goes deeper, a simple layered feature fusion structure

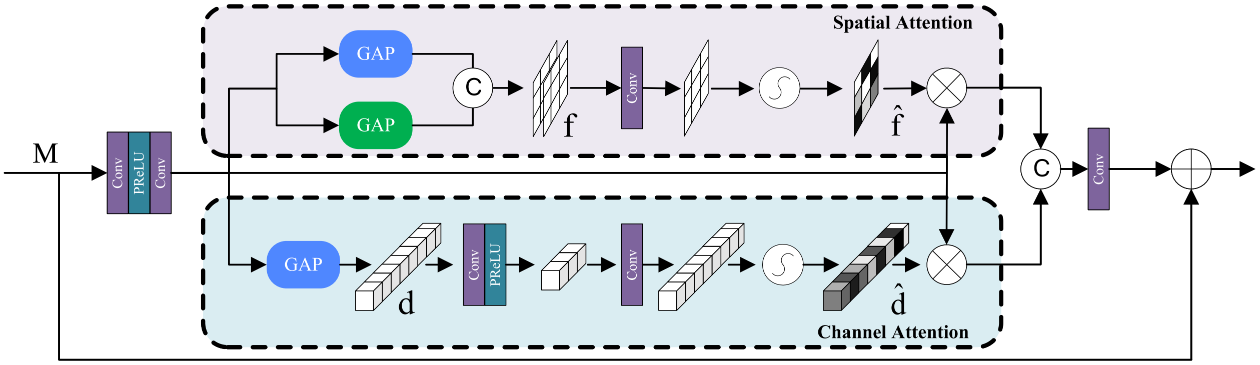


Fig. 5. Schematic for DAU.

with a 1Œ1 kernel convolutional layer as the bottleneck layer is used to solve the problem that hopping connections between different layers does not fully utilize the information redundancy generated by the input image features. The output of the layered feature fusion structure is shown in Eq. 2.

$$F_{LR} = W * [I_0, I_1, I_2 \ldots, I_N] + b \tag{2}$$

where I_0 denotes the output of the first convolution layer, I_i is the output of the multiscale residual block i^{th}, and $[I_0, I_1, I_2 \ldots, I_N]$ denotes the concatenation operation.

2.3. Relativistic discriminator

In this paper, the improvement of the discriminator based on the relativistic discriminator is shown in Fig. 6. Unlike the standard discriminator in SRGAN, the original discriminator is defined as $D(x) = \mathrm{sigmoid}(C(x))$, which predicts the probability that the generated image is true or false, while, in the relative probability that the HRI is more realistic than the generated image, and it is redefined as $D(x) = \mathrm{sigmoid}(C(x_f) - C(x_r))$. As shown in Fig. 6, x_r and x_f represent the real image and the fake image generated by the generator respectively, $\sigma$is the sigmoid activation function, $C(x)$ is the untransformed discriminator output, E represents the averaged data over each mini-batch, and D_{Ra} represents the relativistic discriminator that predicts a relative probability. By this relative calculation of the generated image and the real image, it can make the relative discriminator more global and the SRR result of the model reconstruction is closer to the real image.

2.4. Loss function

The loss function is crucial for deep learning network training and SRR results. By adding perceptual loss and improved adversarial loss to the loss function, the final loss function can be expressed as:

$$L = L_p + \lambda_1 L_1 + \lambda_2 L_a \tag{3}$$

where λ_1 and λ_2 are the regular factors that adjust the weights of loss term L_1 and L_a.

In this paper, we use L_1 loss as the pixel loss to calculate the errors between SR image and HR image in the corresponding pixel position, which can avoid the abnormal values and make the trained model converge quickly and more robust. Its corresponding loss function is defined as Eq. 4:

$$L_1 = \frac{1}{WHC}\sum_{i}^{W}\sum_{j}^{H}\sum_{k}^{C}\left\|I^h_{i,j,k} - I^g_{i,j,k}\right\|_1 \tag{4}$$

To improve the quality of remotely sensed SRR images and to obtain more high frequency features, perceptual loss is added to the loss function and the discriminator uses VGG54 to extract features from SRR images and the original HRI. The perceptual loss L_p is defined as Eq. 5.

$$L_p = \frac{1}{W_i H_i C_i}\|D_i(G(x)) - D_i(y)\|_2^2 \tag{5}$$

where, D_i is the $i-th$ layer of the discriminator network, whose corresponding number of channels is C_i, and the length and width of the layer i feature map are H_i and W_i.

Based on the above description of the relativistic discriminator, the discriminator loss is then defined as the Eq. 6:

$$L_D = -E_{x_r}\left[\log\left(1 - D_{Ra}(x_r, x_f)\right)\right] - E_{x_f}\left[\log\left(D_{Ra}(x_f, x_r)\right)\right] \tag{6}$$

The adversarial loss of the generator is in a symmetrical form:

$$L_a = -E_{x_r}\left[\log\left(1 - D_{Ra}(x_r, x_f)\right)\right] - E_{x_f}\left[\log\left(D_{Ra}(x_f, x_r)\right)\right] \tag{7}$$

where E_{x_r} denotes the averaging operation of all fake image data in a minimal batch, $x_f = G(x_i)$ and x_i represent the LRI. The adversarial loss of the generator includes x_r and x_f.

3. Experimental result and data analysis

3.1. Datasets and Metrics

In the SRR experiments, it is necessary to select some high-resolution images as the target for network training and learning. Our method focuses on the super-resolution reconstruction of visible remote sensing images, and the visible remote sensing image datasets are mainly used for the network training. So in our study, RGB images with high spatial resolution and sufficient feature details were selected from three publicly available datasets: NWPU VHR-10 (spatial resolution is 0.5 m~2 m), LEVIR (spatial resolution is 0.2 m~1.0 m) and RSOD (spatial resolution is 0.5 m~2 m). The total number of used HRIs is 23728. The experimental results were evaluated with PSNR and SSIM metrics.

3.2. Experimental Environment and Image preprocessing

The experiments were conducted on the hardware environment of Intel Core Xeon E3 3.30 GHz, Ubuntu18.0 operation, NVIDIA GTX 2080Ti GPU, and the PyTorch was chosen as the deep learning framework to train the deep learning network. The mapping relationships of the deep learning network require a large number of samples to learn and few public remote sensing datasets are available. To ensure a sufficient number of training samples, the original images are cross-cropped at intervals of 30 pixels, the size of training image is reduced to 256Œ256, and the number of training samples for the network is expanded to 53388. Then the dataset is divided into 1/2 training set, 1/6 validation set and 1/3 test set (Euijeong et al., 2021). They are divided into three groups for testing, corresponding to the experiment Dataset1, Dataset2, and Dataset3 respectively. The specific preprocessing of the dataset is shown in Fig. 7.

3.3. Experiments and Implementation

The parameter initialization method using normalizing the data to a Gaussian distribution does not solve the problem of gradient disappearance with the increase of net layers number (Es-SAFI and HARCHLI, 2016).In order to avoid the variance of activation values decreasing layer by layer, this paper uses Xavier's weight initialization method so that the gradient, input values and activation values are similar on all layers. The initialization is automatically determined according to the number of input and output

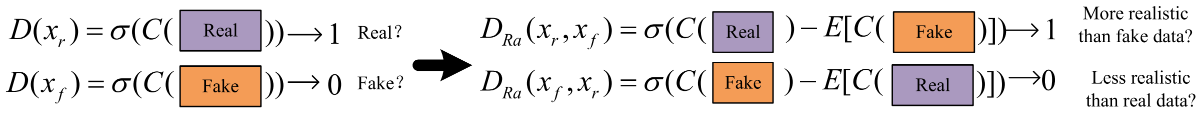

$D(x_r) = \sigma(C($ Real $)) \longrightarrow 1$
Real?
$D(x_f) = \sigma(C($ Fake $)) \longrightarrow 0$
Fake?
$D_{Ra}(x_r, x_f) = \sigma(C($ Real $) - E[C($ Fake $)]) \longrightarrow 1$
More realistic than fake data?
$D_{Ra}(x_f, x_r) = \sigma(C($ Fake $) - E[C($ Real $)]) \longrightarrow 0$
Less realistic than real data?

Fig. 6. Difference between the standard discriminator and relativistic discriminator.

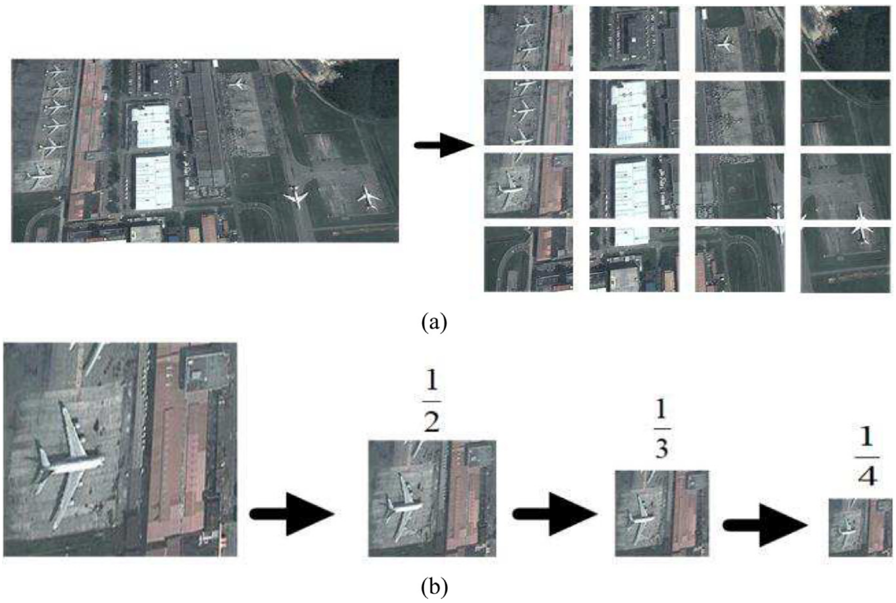


Fig. 7. Pre-processing of training sample images. (a) Cropping process, and (b) Downsampling process.

neurons, defining the input dimension of the layer is fan_{in}, the output dimension is fan_{out}, and the weight matrix W_{ij} can be sampled from a normal distribution with the following standard deviation:

$$\sigma = \sqrt{2/(fan_{\text{in}} + fan_{out})} \tag{8}$$

$$W_{ij} \sim N\left(0, \sqrt{\frac{2}{fan_{in} + fan_{out}}}\right) \tag{9}$$

More detailed parameters are indicated in Table 1. The initial learning rate is 0.1, and after every 20 rounds of training, the learning rate decreases by a factor of 10, and there is no weight decay. At the same time, there is no weight decay. Because the initial learning rate is larger, the gradient is cropped after each round of training. If the current value of the gradient g exceeds a given threshold, the gradient is reasonably reassigned to a smaller value g', calculated as Eq. 10:

$$g' = \min\left(\frac{\theta}{\|g\|}, 1\right) \times g \tag{10}$$

Taking the Œ2 SRR experiment as an example, parameter settings of the GAN training process is shown in Table 1. The HRI is reconstructed with the L_1 loss constraints generator to ensure the effectiveness of the adversarial loss, and the regular coefficients of the adversarial loss are continuously adjusted to finally obtain the best performance. When the regular factor is set $\lambda_2 = 0.005$, the loss gradually becomes a dynamic equilibrium state during the training process, and the total loss stabilizes and tends to converge under the current weight factor setting. The network training process lasted 16 h, and the training was stopped when the overall loss converged and the PSNR and SSIM of the tested images no longer increased. The loss profiles after training are shown in Fig. 8, which are: the L_1 pixel loss for training the generator network; the perceptual loss for reconstructing the HRI closer to human eye perception with the rule factor $\lambda_1 = 0.5$ set; the RaGAN loss representing the discriminator loss under the fixed-weight generator network condition; and the total loss. It can be seen from the loss curve that the total loss gradually converged to a stable.

Table 1
Parameter settings of our Œ2 SRR GAN training process.

Parameter	Setting
Bath size	4
Optimization method	Adam, $\beta_1 = 0.9$, $\beta_2 = 0.999, \sigma = 10^{-8}$
Training epoch number	60
Regular factors of adjusting loss	$\lambda_1 = 5, \lambda_2 = 0.05$

3.4. Analysis of experimental results

3.4.1. Comparisons with State-of-the-Arts

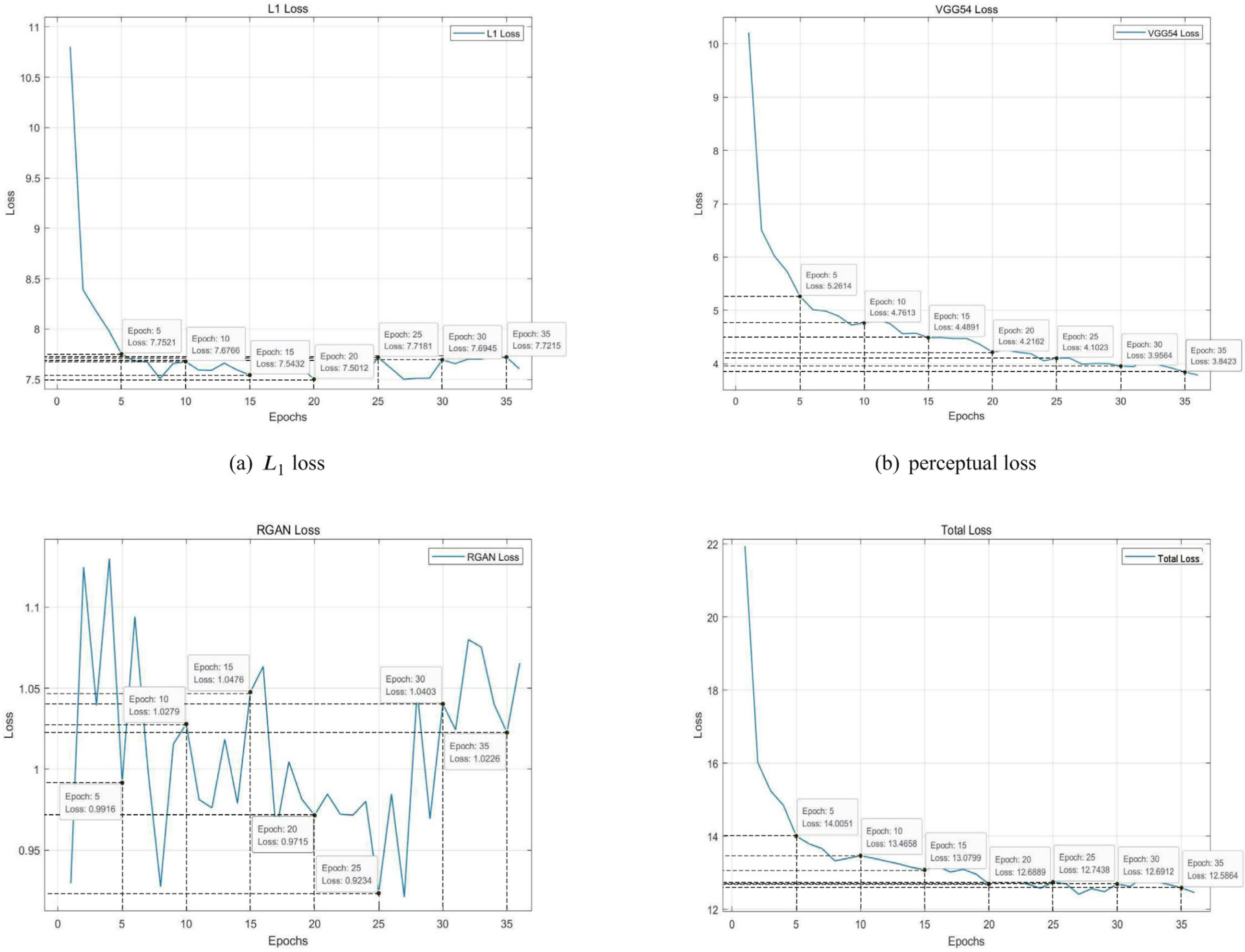
To verify the advancedness of the model proposed in this paper, the super-resolution results of Œ2, Œ3 and Œ4 are respectively compared with the bicubic method and other four SOTA SRR methods, i.e. very deep super-resolution (VDSR), enhanced deep super-resolution (EDSR), residual channel attention networks (RCAN), super-resolution generative adversarial networks (SRGAN). The subjective visual comparison results of the former experiments are shown in Fig. 9–11.

From above comparison results of remote sensing image SRR in Fig. 9 to Fig. 11, the results of Bicubic, VDSR, EDSR and RCAN often show over-smoothing and lack of some details. While our method introduces GAN loss, so the SRR results are richer in detailed information, the Œ2 super-resolution results perform well, the artifacts can be observed in the higher magnification experiments and jagged irregularities appear in the boundary structure. The boundary of the structure SRR image proposed in this paper is more regular and the line details are sharper and clear. Therefore, the visual results of the reconstruction method proposed in this paper are closer to the human eye perception.

3.4.2. Quantitative Evaluation and Qualitative Evaluation

In this section, the proposed method is evaluated with other methods qualitatively and quantitatively. The results of Œ2, Œ3 and Œ4 SRR were evaluated by PSNR and SSIM, respectively, and the results are presented in Tables 2–4.

As can be seen from the qualitative evaluation table above, our method is able to adapt to SRR requirements at different scales and both obtain optimal PSNR and SSIM results compared to other SOTA methods. On the Dataset3 test set with scale Œ2, the PSNR



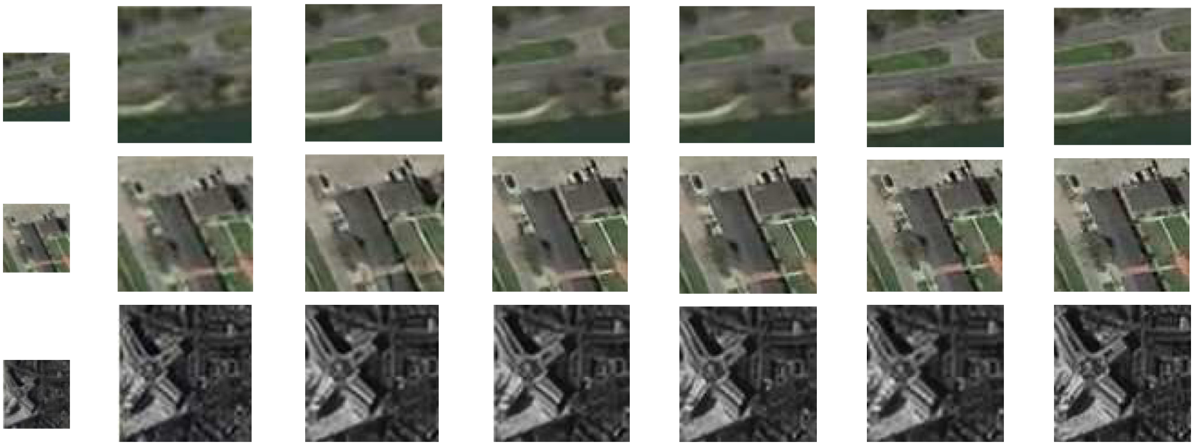
(a) L_1 loss

(b) perceptual loss

(c) RaGAN loss

(d) total loss

Fig. 8. Loss curve of our net training.



(a) LRI (b) Bicubic (c) VDSR (d) EDSR (e) RCAN (f) SRGAN (g) Ours

Fig. 9. Comparison results of Œ2 SRR. Images of the first row are from the Dataset1, images of the second row are from the Dataset2, and images of the third row are from the Dataset3.

and SSIM of our method reach 34.28/0.8887, which outperforms the second-best SRGAN model with PSNR gains of 1.18 dB and SSIM gains of 0.0264. On the Dataset2 testing set with scales Œ3 and Œ4 the PSNR and SSIM of our method reach 28.94 dB/0.7517 and 27.93 dB/0.7466. It outperforms the second-best model, with PSNR gains of 0.84 dB and 1.29 dB, but the value of SSIM is not

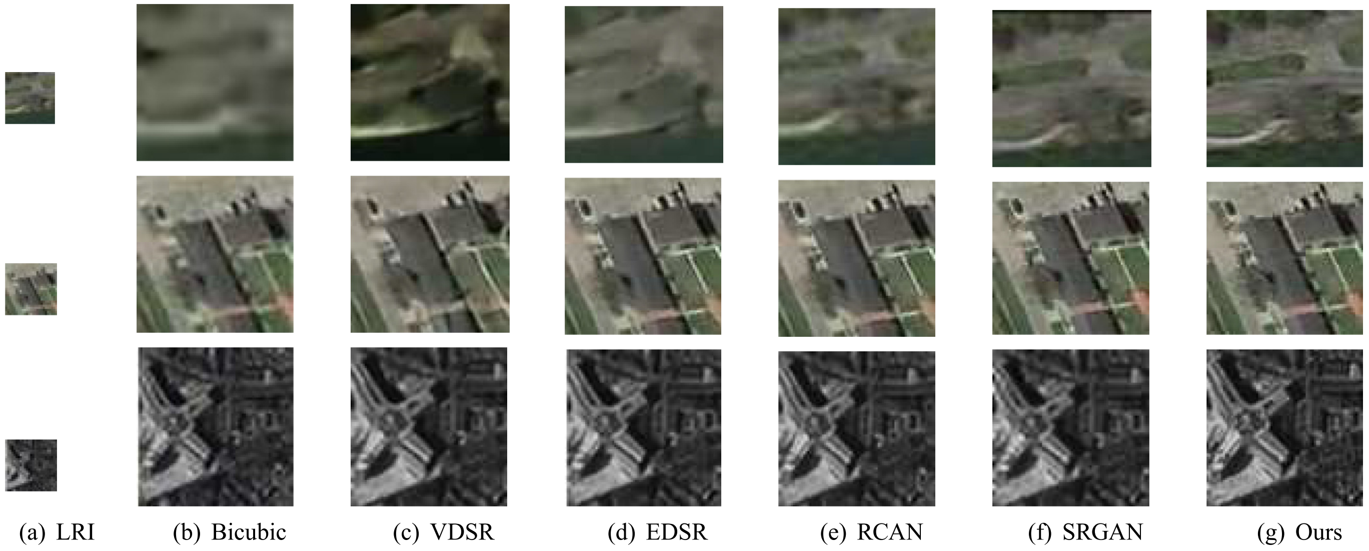

(a) LRI (b) Bicubic (c) VDSR (d) EDSR (e) RCAN (f) SRGAN (g) Ours

Fig. 10. Comparison results of Œ3 SRR. Images of the first row are from the Dataset1, images of the second row are from the Dataset2, and images of the third row are from the Dataset3.

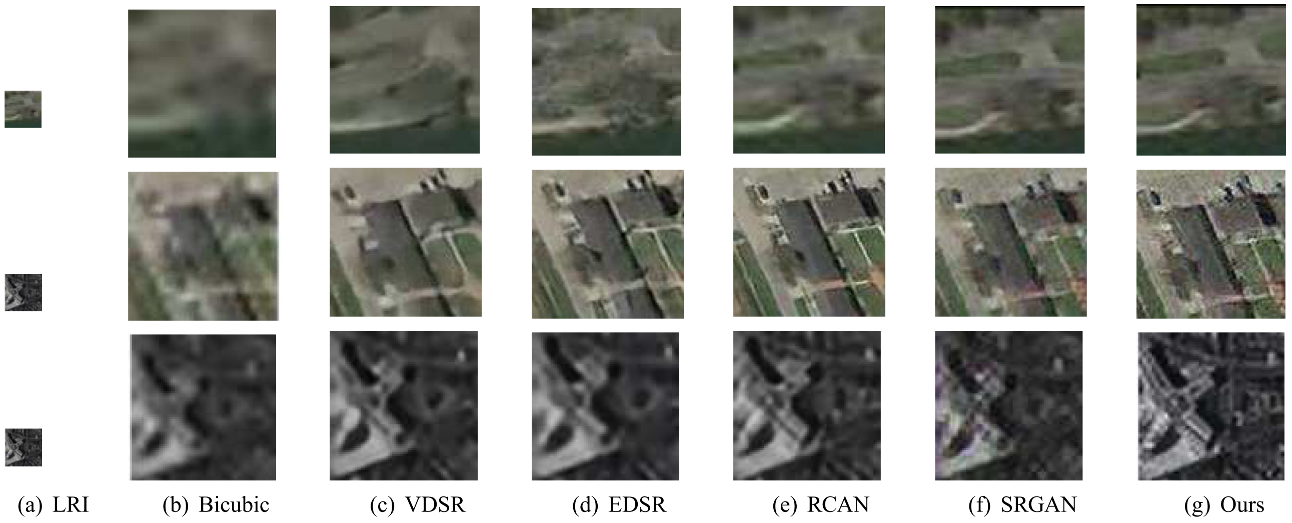

(a) LRI (b) Bicubic (c) VDSR (d) EDSR (e) RCAN (f) SRGAN (g) Ours

Fig. 11. Comparison results of Œ4 SRR. Images of the first row are from the Dataset1, images of the second row are from the Dataset2, and images of the third row are from the Dataset3.

Table 2
Comparison results of PSNR (dB) and SSIM for Œ2 remote sensing images SRR.

Method	Scale	Dataset1		Dataset2		Dataset3	
		PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
Bicubic	Œ2	26.95	0.7288	25.94	0.6347	27.84	0.7032
VDSR	Œ2	30.86	0.8456	29.44	0.7899	30.98	0.8432
EDSR	Œ2	31.32	0.8566	30.90	0.8002	31.54	0.8511
RCAN	Œ2	32.86	0.8922	31.21	0.8111	32.03	0.8597
SRGAN	Œ2	31.44	0.8777	31.89	0.8437	33.10	0.8623
Ours	Œ2	**32.90**	**0.9002**	**32.76**	**0.8599**	**34.28**	**0.8887**

as good as RCAN in the Œ4 SSR experiment. In the three sets of test data, compared with SOTA, the PSNR values of Œ2, Œ3 and Œ4 scales are improved by a maximum of 1.18 dB, 0.84 dB and 1.29 dB, respectively, and the SSIM values are improved by 0.0264, 0.0077 and 0.0109, which indicates that the structural information of remote sensing images can be effectively reconstructed with our method.

3.4.3. Ablation experiment and analysis

To investigate the effects of multiscale residual blocks and relativistic GAN loss functions on the performance of the proposed models, ablation experiments were conducted. In the ablation experiments, the multiscale residual blocks and the relativistic GAN loss function are retained and eliminated, and four network models are generated after swapping groups, and the pairs are

Table 3
Comparison results of PSNR (dB) and SSIM for Œ3 remote sensing images SRR.

Method	Scale	Dataset1		Dataset2		Dataset3	
		PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
Bicubic	Œ3	23.18	0.6738	22.36	0.5838	24.16	0.6493
VDSR	Œ3	28.03	0.8113	26.62	0.6872	28.34	0.8091
EDSR	Œ3	29.11	0.8297	27.83	0.7099	28.98	0.8174
RCAN	Œ3	29.56	0.8256	28.10	0.7482	29.64	0.8256
SRGAN	Œ3	29.11	0.8224	27.95	0.7183	29.71	0.8278
Ours	Œ3	**29.95**	**0.8310**	**28.94**	**0.7517**	**29.74**	**0.8355**

Table 4
Comparison results of PSNR (dB) and SSIM for Œ4 remote sensing images SRR.

Method	Scale	Dataset1		Dataset2		Dataset3	
		PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
Bicubic	Œ4	20.85	0.5518	20.12	0.5162	21.13	0.5131
VDSR	Œ4	26.69	0.7456	25.73	0.7107	25.81	0.8172
EDSR	Œ4	26.99	0.8013	26.16	0.7433	24.44	0.8279
RCAN	Œ4	27.03	0.8125	26.03	**0.7488**	26.21	0.8279
SRGAN	Œ4	27.83	0.8125	26.64	0.7463	25.34	**0.8290**
Ours	Œ4	**28.10**	**0.8234**	**27.93**	0.7466	**26.59**	0.8254

Table 5
The ablation experiments with four models.

	Multi-scale residual blocks	Relativistic average GAN loss	PSNR (dB)	SSIM
Model 1	X	X	26.54	0.3636
Model 2	✓	X	28.73	0.4030
Model 3	X	✓	28.43	0.3967
Model 4	✓	✓	29.85	0.4258

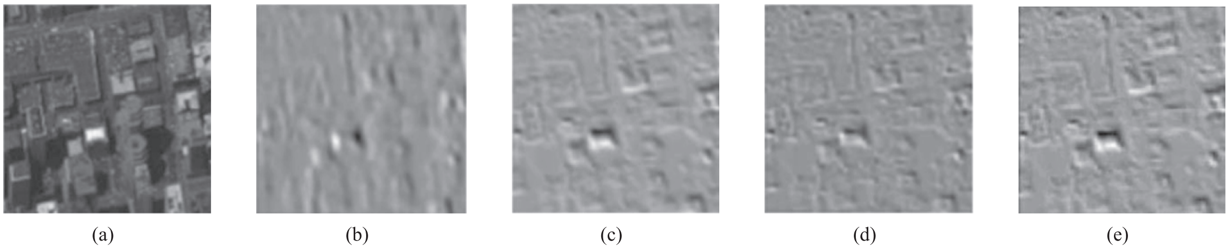


Fig. 12. The feature maps of four models ablation experiments. (a) Original image; (b) feature maps of Model 1; (c) feature maps of Model 2; (d) feature maps of Model 3; (e) feature maps of Model 4.

compared under the same conditions. Among them, Model 1 is a generator containing a cascade with four ordinary residual blocks and generating the objective function with pixel loss; Model 2 is a generator containing a cascade consisting of the same number of multiscale residual blocks as proposed in this paper; Model 3 is a loss function proposed in this paper added to Model 1; and Model 4 is the model proposed in this paper. Using PSNR and SSIM as evaluation criteria, the detailed information and quantitative results of each model are shown in the Table 5. Using the VGG54 pre-trained model, the features of the same layers are extracted from the reconstruction results of the above four cases, and their corresponding feature maps are shown in Fig. 12.

From above ablation experimental results, it can be seen that the model containing RaGAN loss has better boundary information. The model containing multi-scale residual blocks has better constraint on pixel information and richer content information. The ablation experiments were performed and the results were quantified in SRR of remote sensing images with scale factor Œ4. The test results of the above four models in the same three datasets are shown in Table 5. It can be seen that the network model that retains both the multiscale residual block and RaGAN loss function has better SRR results than the other three network models.

4. Conclusions

In this paper, we improve the SRGAN network in the generator and discriminator, respectively. A multi-scale residual block network is introduced into the generator, which includes two main modules, SKFF and DAU mechanism. The SKFF mechanism can exchange knowledge on information streams of different resolutions, and the DAU is used to capture contextual information in spatial and channel dimensions. The discriminator is improved with the idea of RaGAN so that it can predict relativistic probabilities instead of absolute probabilities, enabling the network to learn more realistic texture details. Experimental results show that the method in this paper significantly outperforms the SOTA method in both subjective and objective metrics. In the test experiments of three datasets, the PSNR improves 1.18 dB, 0.84 dB and 1.29 dB over SOTA at Œ2, Œ3 and Œ4 scales, respectively; the SSIM improves 0.0264, 0.0077 and 0.0109, respectively. the results of the ablation experiments also show that the method proposed in this paper can effectively improve the SRR results of remote sensing images.In the future, we will expand our research to SRR of multispectral images or infrared images and focus more attention on the complexity and computations of network models.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

This work is supported in part by the Fundamental Research Funds for the Heilongjiang Provincial Universities (Heilongjiang University Special Funds Project), inpart by the National Natural Science Foundation China (61601174)in part by the Postdoctoral Research Foundation of Heilongjiang Province (LBH-Q17150)in part by the Science and Technology Innovative Research Team in Higher Educational Institutions of Heilongjiang Province (No. 2012TD007).

References

Ahn, N., Kang, B., Sohn, K.A., 2018. Fast, accurate, and lightweight super-resolution with cascading residual network, in: Proceedings of the European conference on computer vision (ECCV), pp. 252–268.

Berger, J.O., Moreno, E., Pericchi, L.R., et al., 1994. An overview of robust bayesian analysis. Test 3, 5–124.

Chen, F., Wei, J., Xue, B., et al., 2022. Feature fusion and kernel selective in inception-v4 network. Applied Soft Computing 119, 108582.

Chen, H., Wang, Y., Xu, C., et al., 2020. Addernet: Do we really need multiplications in deep learning?, in: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp. 1468–1477.

Diniz, P.S. et al., 1997. Adaptive filtering, volume 4. Springer.

Dong, C., Loy, C.C., He, K., et al., 2015. Image super-resolution using deep convolutional networks. IEEE transactions on pattern analysis and machine intelligence 38, 295–307.

Es-SAFI, Abdelatif et al., 2016. HARCHLI: Adaptation of multilayer perceptron neural network to unsupervised clustering using a developed version of k-means algorithm. WSEAS Transactions on Computers 15, 103–116.

Euijeong, S., Seokjung, K., Seok, C., et al., 2021. Srps–deep-learning-based photometric stereo using superresolution images. Journal of Computational Design and Engineering 4.

Gao, S.H., Cheng, M.M., Zhao, K., et al., 2019. Res2net: A new multi-scale backbone architecture. IEEE transactions on pattern analysis and machine intelligence 43, 652–662.

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

Gomes, H.M., Read, J., Bifet, A., et al., 2019. Machine learning for streaming data: state of the art, challenges, and opportunities. ACM SIGKDD Explorations Newsletter 21, 6–22.

Guo, R., Shi, X.P., Jia, D.K., 2018. Learning a deep convolutional network for image super-resolution reconstruction. Journal of Engineering of Heilongjiang University.

Hou, J., Si, Y., Yu, X., 2020. A novel and effective image super-resolution reconstruction technique via fast global and local residual learning model. Applied Sciences 10, 1856.

Huang, G., Liu, S., Van der Maaten, L., et al., 2018. Condensenet: An efficient densenet using learned group convolutions, in: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 2752–2761.

Kim, J., Lee, J.K., Lee, K.M., 2016a. Accurate image super-resolution using very deep convolutional networks, in: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 1646–1654.

Kim, J., Lee, J.K., Lee, K.M., 2016b. Deeply-recursive convolutional network for image super-resolution, in: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 1637–1645.

Ledig, C., Theis, L., Huszár, F., et al., 2017. Photo-realistic single image super-resolution using a generative adversarial network, in: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 4681–4690.

Lee, W., Lee, J., Kim, D., et al., 2020. Learning with privileged information for efficient image super-resolution. In: European Conference on Computer Vision. Springer, pp. 465–482.

Li, F., Bai, H., Zhao, Y., 2020. Learning a deep dual attention network for video super-resolution. IEEE transactions on image processing 29, 4474–4488.

Li, J., Fang, F., Mei, K., et al., 2018. Multi-scale residual network for image super-resolution, in: Proceedings of the European conference on computer vision (ECCV), pp. 517–532.

Li, M., Hsu, W., Xie, X., et al., 2020. Sacnn: Self-attention convolutional neural network for low-dose ct denoising with self-supervised perceptual loss network. IEEE transactions on medical imaging 39, 2289–2301.

Liu, H., Cao, F., Wen, C., et al., 2020. Lightweight multi-scale residual networks with attention for image super-resolution. Knowledge-Based Systems 203, 106103.

Liu, W., Lee, J., 2019. An efficient residual learning neural network for hyperspectral image superresolution. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 12, 1240–1253.

Ma, C., Rao, Y., Cheng, Y., et al., 2020. Structure-preserving super resolution with gradient guidance, in: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp. 7769–7778.

Nedić, A., 2010. Random projection algorithms for convex set intersection problems, in: 49th IEEE Conference on Decision and Control (CDC), IEEE. pp. 7655–7660.

Olshausen, B.A., Field, D.J., 2004. Sparse coding of sensory inputs. Current opinion in neurobiology 14, 481–487.

Ouyang, J., Chen, K.T., Gong, E., et al., 2019. Ultra-low-dose pet reconstruction using generative adversarial network with feature matching and task-specific perceptual loss. Medical physics 46, 3555–3564.

Peng, X., Yongping, L.I., Zhang, X., 2016. Binocular stereo matching algorithm based on deep learning.

Qin, J., Huang, Y., Wen, W., 2020. Multi-scale feature fusion residual network for single image super-resolution. Neurocomputing 379, 334–342.

Rojo-Álvarez, J.L., Figuera-Pozuelo, C., Martínez-Cruz, C.E., et al., 2007. Nonuniform interpolation of noisy signals using support vector machines. IEEE Transactions on Signal Processing 55, 4116–4126.

Shi, J., Liu, Q., Wang, C., et al., 2018. Super-resolution reconstruction of mr image with a novel residual learning network algorithm. Physics in Medicine & Biology 63, 085011.

Shi, W., Caballero, J., Huszár, F., et al., 2016. 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, pp. 1874–1883.

Wang, X., Yu, K., Wu, S., et al., 2018. Esrgan: Enhanced super-resolution generative adversarial networks, in: Proceedings of the European conference on computer vision (ECCV) workshops, pp. 0–0.

Wang, Y., Ying, X., 2021. Symmetric parallax attention for stereo image super-resolution, in: Computer Vision and Pattern Recognition.

Woo, S., Park, J., Lee, J.Y., et al., 2018. Cbam: Convolutional block attention module, in: Proceedings of the European conference on computer vision (ECCV), pp. 3–19.

Wunsch, P., Hirzinger, G., 1996. Registration of cad-models to images by iterative inverse perspective matching, in: Proceedings of 13th International Conference on Pattern Recognition, IEEE. pp. 78–83.

Yang, Q., Yan, P., Zhang, Y., et al., 2018. Low-dose ct image denoising using a generative adversarial network with wasserstein distance and perceptual loss. IEEE transactions on medical imaging 37, 1348–1357.

Zamir, S.W., Arora, A., Khan, S., et al., 2020. Learning enriched features for real image restoration and enhancement. In: European Conference on Computer Vision. Springer, pp. 492–511.

Zeng, Y., Fu, J., Chao, H., et al., 2019. Learning pyramid-context encoder network for high-quality image inpainting, pp. 1486–1494.

Zhang, T., Gu, Y., Huang, X., 2020. Stereo endoscopic image super-resolution using disparity-constrained parallel attention.

Zhang, Y., Li, K., Li, K., et al., 2018. Image super-resolution using very deep residual channel attention networks, in: Proceedings of the European conference on computer vision (ECCV), pp. 286–301.

Zhou, L.Y., Cai-Xia, S.U., Cao, Y.F., 2016. Image super-resolution via sparse representation. Computer Engineering and Design.