International Journal of Innovative Computing, Information and Control
Volume **19**, Number **2**, April **2023**

ICIC International ©2023 ISSN 1349-4198
pp. **385–401**

GENERATIVE ADVERSARIAL NETWORK WITH NEW BATCH NORMALIZATION AND FEATURE EXTRACTION BLOCK FOR IMAGE SUPER-RESOLUTION RECONSTRUCTION

Liquan Zhao[1,*], Jingjing Wu[1] and Yanfei Jia[2]

[1]Key Laboratory of Modern Power System Simulation and Control
and Renewable Energy Technology, Ministry of Education
Northeast Electric Power University
No. 169, Changchun Road, Jilin 132012, P. R. China
wu_jingjing2022@163.com
[*]Corresponding author: zhaoliquan@neepu.edu.cn

[2]College of Electric and Information Engineering
Beihua University
No. 3999, Beijing Road, Jilin 132013, P. R. China
jia_yanfei@163.com

Received August 2022; revised November 2022

Abstract. *To improve the quality of the recovered image by the generative adversarial network, an improved generative adversarial network is proposed. Firstly, it designs a new batch normalization block to avoid gradient explosion and disappearance. The traditional batch normalization will reduce the standard deviation of feature pixels, which causes degradation of reconstructed image quality. To solve the problem, an adaptive standard deviation of feature pixels modulator is designed to amplify the deviation of feature pixels and is introduced to traditional batch normalization to construct new batch normalization. Secondly, to extract more useful features, a new block is designed. The proposed block consists of two branches with different network depths. It fuses the different extracted features from the two branches to obtain more useful features. Besides, the proposed batch normalization block also is introduced into the new block. Thirdly, the new block is used to construct a dense network with skip connection characteristics for extracting features. Besides, the new block is also used alone at the end of the feature extraction network to fuse different features. Compared with EnhanceNet, SRGAN, ESRGAN, R-SRGAN, SAM+VAM on Set5, Set14, BSD200, and Urban100 datasets, our proposed method still has the greatest average PSNR, and SSIM for recovered images from the images downsampled the super-resolution images using a bicubic kernel with a scaling factor of ×2, ×3 and ×4, respectively. The recovered image by our proposed method is closer to the ground-truth image than other methods.*

Keywords: Image super-resolution reconstruction, Generative adversarial network, Feature extraction block, Batch normalization

1. **Introduction.** With the rapid development of computer technology, computer image processing technology has emerged and is more and more widely used. It can help various industries to collect and analyze all kinds of images and data, including image segmentation [1], image denoising [2], non-photorealistic rendering of images [3,4], and image super-resolution [5]. Image super-resolution reconstruction is to recover a high-resolution image from a low-resolution image. It can make blurred images clear to reduce the impacts of acquisition equipment or environment on the image. Image super-resolution reconstruction has important applications in many fields. In the medical field, super-resolution

DOI: 10.24507/ijicic.19.02.385

reconstruction can improve the quality of medical images to improve the performance of automated analysis of medical images and assist doctors in better diagnosing diseases [6]. In the field of the infrared image field, the low-resolution infrared image is very blurred, and the temperature measurement is not accurate. It is unable to directly meet the functional needs of household appliances, especially air-conditioning products which need to obtain the object temperature, location, and orientation in the target area. The super-resolution reconstruction can improve the accuracy of infrared temperature telemetry while increasing the resolution to meet the needs of practical scene applications [7].

Image super-resolution reconstruction methods can be divided into three classes: image super-resolution reconstruction based on interpolation methods, image super-resolution reconstruction based on reconstruction methods, and image super-resolution reconstruction based on learning methods. Recently, deep learning has been introduced into the field of image super-resolution reconstruction [8-10]. For image super-resolution reconstruction based on interpolation methods, the pixel information of the image is not fully utilized during the interpolation, so it cannot recover more detailed information. For image super-resolution reconstruction based on reconstruction methods, it supposes that the low-resolution image is obtained through a certain degradation process, and recovers the high-resolution image completely based on this assumption. The assumed prior information is inaccurate in real application, so the reconstruction method is not widely used in practice. The image super-resolution reconstruction based on deep learning uses many high-resolution images to build a training dataset to train the reconstruction network. It introduced prior information obtained from the training process to the reconstruction network. Therefore, it performs better than the image super-resolution reconstruction methods based on interpolation and reconstruction methods.

The image super-resolution reconstruction methods based on normal deep learning can recover higher quality images with the deep convolutional neural network. However, minimizing the Euclidean distance between predicted and ground truth pixels will encourage the deep convolutional neural network to produce blurry results. The generative adversarial network (GAN) is one of the special deep learning networks. It also has been used in image super-resolution reconstruction to solve this problem. The generative adversarial networks consist of two networks: generative network and adversarial network. The generative network is used to recover the super-resolution image. The adversarial network is used to determine whether the image is generated or the original super-resolution image. The optimal generative network is used to recover the high-resolution image from the low-resolution image by gaming the generative and adversarial networks. Enhanced super-resolution generative adversarial network (ESRGAN) [11] is one of the image super-resolution reconstruction methods based on a generative adversarial network. To improve the quality of recovered images, we design a new generative network based on ESRGAN to extract more effective features and recover higher resolution images.

The main contributions of this paper are summarized as the following.

1) We design a new batch normalization block to avoid gradient explosion and gradient disappearance in the generative network. Although the traditional batch normalization block can solve the problem, it will reduce the standard deviation of feature pixels, which causes degradation of reconstructed image quality. Therefore, we design an adaptive standard deviation of feature pixels modulator to amplify the deviation of feature pixels. The new batch normalization block consists of two branches. One branch is the traditional batch normalization, and the other is the designed adaptive standard deviation of feature pixels modulator. In the end, we introduce the new batch normalization block to the feature extraction network of the generative network.

2) We design a new feature extraction block and use it as a sub-block to construct a dense network with skip connection characteristics to extract features. The proposed block consists of two branches with different network depths and architectures. It fuses the different extracted features from the two branches to obtain more effective features. The designed dense network with skip connection characteristics is used as part of the feature extraction network. Besides, the new feature extraction block is also used alone at the end of the feature extraction network to fuse different features.

3) We also use the designed dense network as a sub-network to construct a network with residual architecture. The constructed residual network is also part of the feature extraction network. The whole of our proposed feature extraction network consists of three parts: the original residual-in-residual dense block of ESRGAN with our proposed batch normalization block, our designed dense network with skip connection characteristic, and our designed residual network based on the designed dense network.

This section introduces the background of image super-resolution reconstruction and our contribution. In Section 2, we describe related work on image super-resolution reconstruction. In Section 3, we present the detail of our proposed method. In Section 4, we show and compare our simulation results. In Section 5, the conclusion is given.

2. **Related Work.**

The super-resolution reconstruction method based on a convolutional neural network (CNN) has been widely used in recent years. Zhang et al. proposed a deep learning method with convolutional neural networks (CNNs) using skip connections with layer groups for super-resolution image reconstruction [8]. In the proposed method, entire CNN layers for residual data processing are divided into several layer groups. Skip connections with different multiplication factors are applied from input data to these layer groups. With the proposed method, the processed data in hidden layer units tend to be distributed wider. Consequently, the feature information from input data is transmitted to the output more robustly. Hu et al. proposed a method of image super-resolution reconstruction based on a hybrid deep convolutional network [9]. In 2020, Zhang et al. proposed an image super-resolution reconstruction algorithm combined with a multi-residual network and multi-feature SCSR (MRMFSCSR) [10]. In the proposed method, according to improving the very deep network super resolution (VDSR) deep network and introducing the feature fusion idea, the multi-residual network structure (MR) was designed.

With further research on neural networks, generative adversarial network (GAN) was first published in 2014 by Goodfellow et al. [12]. The generative adversarial network improves the performance of the model through continuous confrontation learning of the generator and discriminator. Therefore, GAN is widely used in different fields. In 2017, Ledig et al. proposed super-resolution generative adversarial networks (SRGAN) [13], using GAN for the first time in the field of image super-resolution. This method improved the visual effect of the generated image by improving texture details. In 2018, Wang et al. proposed enhanced super-resolution generative adversarial networks (ESRGAN) [11], which further improved the structure of SRGAN. To improve the generated network, a five-layer convolutional dense block and residual block were used to replace the original basic block to deepen the network depth. This paper proposed a lighter algorithm [14] based on SRGAN. The algorithm applied the feedback structure to the image generator to process the feedback information and enhance the high-frequency information of the image. Sun et al. proposed an image SR reconstruction method based on a generative adversarial network with a dense residual architecture [15]. It improved the utilization of shallow information in a deep network. Nan et al. [16] proposed a single image super-resolution reconstruction model called the Res_WGAN based on ResNeXt. The generator was constructed by the ResNeXt network, which reduced the computational complexity of the

model generator to 1/8 that of the SRGAN. This paper proposed a novel dense generative adversarial network for real aerial imagery super-resolution reconstruction (NDSRGAN) [17], and it produced image datasets with paired high and low-resolution real aerial remote sensing images. Dou et al. [18] proposed to improve the two-dimensional convolution layer in the SRGAN network structure into a three-dimensional convolution layer, use an attention mechanism to process the multiple features from the three-dimensional convolution layer, and improve the output of the model by improving the generator loss function, which can efficiently carry out the super-resolution reconstruction task. Zhang et al. [19] proposed a method based on the residual dense connection network (RDN), which integrated multiple residual dense blocks and can effectively extract feature information. Shao et al. [20] proposed a multi-scale generative adversarial network model based on a cross-layer attention transfer mechanism. The model utilized the cross-layer attention transfer module to make the high-level feature map guide the filling of the low-level feature map and ensure the visual effect of reconstruction. Hsu and Lin [21] proposed to add one more shortcut between two dense blocks, as well as add a shortcut between two convolution layers inside a dense block. This simple strategy of adding more shortcuts in the proposed network enabled a faster learning process as the gradient information can be back-propagated more easily. Based on the improved ESRGAN, the dual reconstruction was proposed to learn different aspects of the super-resolved image for judiciously enhancing the quality of the reconstructed image.

RFB-ESRGAN [22] proposed to use a multi-scale network structure in the generation network of ESRGAN to reconstruct finer details and textures of super-resolution images. Multi-scale network structure could propose features of different scales from the previous feature graph and make full use of features to achieve more image details. Rakotonirina and Rasoanaivo [23] designed a novel basic block network structure to replace the original basic block network structure of ESRGAN and introduced noise input into the generation network to make use of random changes to obtain a more realistic texture of the generated image. Residual super-resolution generative adversarial networks (R-SRGAN) [24] were used to build the model and realize image super-resolution. By adding residual blocks between adjacent convolutional layers of the GAN generator, more detailed information is retained. At the same time, the Wasserstein distance was used as a loss function to enhance the training effect and achieve image super-resolution. Wang [25] proposed an image super-resolution algorithm (SR) based on U-Net GAN [26]. As the basic block, this method introduced the residual-in-residual self-calibrated convolution with pixel attention block (RRSCPA) [27]. From [28] to heuristic, they designed the discriminator into a U-shaped structure, which could provide per-pixel feedback to the generator and promote the generator to generate a more realistic high-resolution image.

Although the recovered images by the above super-resolution reconstruction methods based on the generative adversarial network are relatively clear, there are some losses of detail features, which affect recovered image quality. The local blurring of the recovered image will affect the image segmentation and object detection based on artificial intelligence, human perception of image and diagnosis of disease, etc. To recover more detailed information from low-resolution reconstruction and improve the quality of the recovered image, we propose a new generative adversarial network for super-resolution image reconstruction.

3. **Method.**

To further improve the super-resolution image reconstruction effect, we propose a super-resolution image reconstruction algorithm based on a generative adversarial network, which can obtain more realistic and natural textures and details. The reconstruction network consists of two parts: the generative network, and the adversarial network.

The generative network is used to recover a high-resolution image from a low-resolution image. The adversarial network is used to determine whether the input image is the recovered image or the original real high-resolution image. The adversarial network and the generative network compete with each other to get the optimal generative network. Finally, the generative network recovers a high-resolution image close to the real high-resolution image from a low-high-resolution image. Based on ESRGAN, we design an improved generative network. The proposed generative network is shown in Figure 1. Compared with ESRGAN, we improve the Trunk Block, design a new Trunk-GM module, GM module and increase the short skip operation to fuse the shallow and deep feature information of the network and avoid information loss. The short skip operation is marked in red in Figure 1. In the following, we will introduce each module of the proposed generative network in turn.

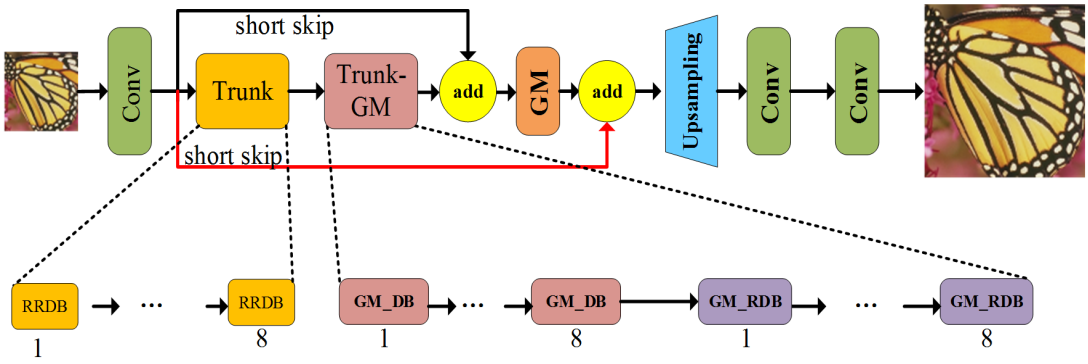

FIGURE 1. Proposed generative network

3.1. **Proposed batch normalization.** Generative adversarial network suffers from certain problems such as gradient explosion and gradient disappearance during training. To solve the problem, we propose to introduce batch normalization into the generative network. The traditional batch normalization will reduce the standard deviation of feature pixels. The standard deviation of feature pixels reflects the amount of variation of pixel values. If the standard deviation of feature pixels decreases, the feature information of the image edge will be lost, which causes degradation of reconstructed image quality. To solve the problem, we design an adaptive standard deviation of feature pixels modulator to amplify the deviation of feature pixels. We introduce the designed adaptive standard deviation modulator to the traditional batch normalization to construct the new batch normalization. The proposed batch normalization is shown in Figure 2. The BN block is the normal batch normalization. The Module in Figure 2 is the module where the proposed batch normalization will be added. The x is the input feature that is also the input feature of the original Module. The std(), log(), $f()$ and exp() are the standard deviation of input feature, logarithmic function, linear function, and an exponential function, respectively.

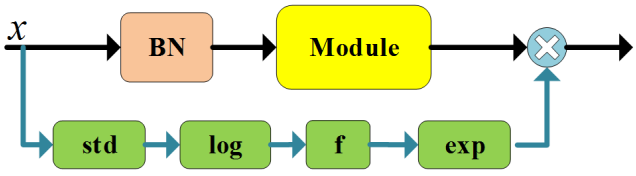

FIGURE 2. Proposed batch normalization

The designed adaptive standard deviation modulator is expressed as

$$O = y \times \exp(f(\log(\delta(x)))) = y \times \exp(w \times (\log(\delta(x))) + b) \tag{1}$$

where O is the output feature of the whole block in Figure 2, y is the output feature of Module, x is the input feature, and $f()$ is a linear function that consists of weight w and bias b. They can be updated by backpropagation during the training process to get the optimal weight and bias. $\delta(x)$ is used to compute the standard deviation of x. The $\exp(w \times (\log(\delta(x))) + b)$ is used to amplify the standard deviation to reduce the effect of normal batch normalization on standard deviation. In the training process, we can obtain the optimal weight and bias to modify the standard deviation to improve the quality of the recovered image.

3.2. **Proposed feature extraction block (GM block).** To obtain more effective feature information to make the recovered image contain more detailed information and improve the quality of the recovered image, we propose a new feature extraction block that is called GM block. The proposed feature extraction block is shown in Figure 3. We also use our proposed BN block to avoid gradient explosion and gradient disappearance. The proposed block contains two branches. The upper branch is our proposed BN block to reduce the effect of the normal BN block on the standard deviation of feature pixels. The lower branch is used to extract the feature. We use double branches architecture to design feature extraction blocks. The architecture can better utilize the correlation information between channels to improve the feature extraction capability of the block.

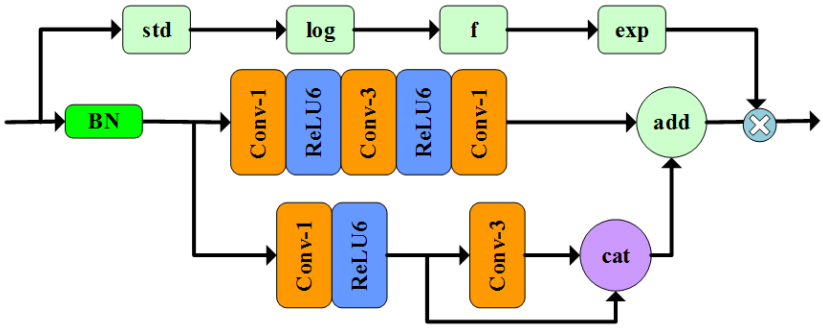

FIGURE 3. Proposed feature extraction block

The features are divided into two branches after the BN layer. The first branch successively consists of a 1×1 convolution, a ReLU6 activation function, a 3×3 convolution, a ReLU6 activation function, and a 1×1 convolution. The first 1×1 convolution is mainly used for halving the number of channels to reduce the computation. The 3×3 convolution is mainly used to extract features without changing the number of channels. The last 1×1 convolution is used to expand the number of channels to recover the number of compressed channels. It makes the number of feature channels in the feature extraction process remain constant. The second branch successively consists of a 1×1 convolution, a ReLU6 activation function, and a residual network composed of a 3×3 convolution. The first 1×1 convolution is also used to reduce the number of channels. To obtain feature information of the shallow layer, the 3×3 convolution with residual network architecture is used to extract feature information. The output feature of 3×3 convolution and output feature of the short skip is concatenated in channel dimension by cat operation. The two branches contain different feature information and have the same number of channels. Therefore, the add operation is used to fuse the feature information obtained from two branches to extract more effective feature information.

3.3. **Improved feature extraction network.** Our proposed feature extraction network mainly contains three parts: improved Trunk module, Trunk_GM module, and GM module. The original feature extraction network in ESRGAN only includes the Trunk module. The Trunk module contains sixteen RRDB blocks, and each RRDB block contains three RDB blocks. In our proposed feature extraction network, we reduce sixteen RRDB blocks to eight RRDB blocks and introduce eight GM_DB blocks and eight GM_RDB blocks.

To avoid gradient explosion and gradient disappearance, which affect the recovered image quality, we introduce our proposed batch normalization into the RDB and RRDB, respectively. We also use the Leaky-ReLU function instead of ReLU function that is used in the RDB blocks to avoid the dying ReLU problem. The improved RDB and RRDB blocks are shown in Figure 4. The improved RDB block is then used to construct the RRDB block, and eight improved RRDB blocks are used to build the Trunk module shown in Figure 1.

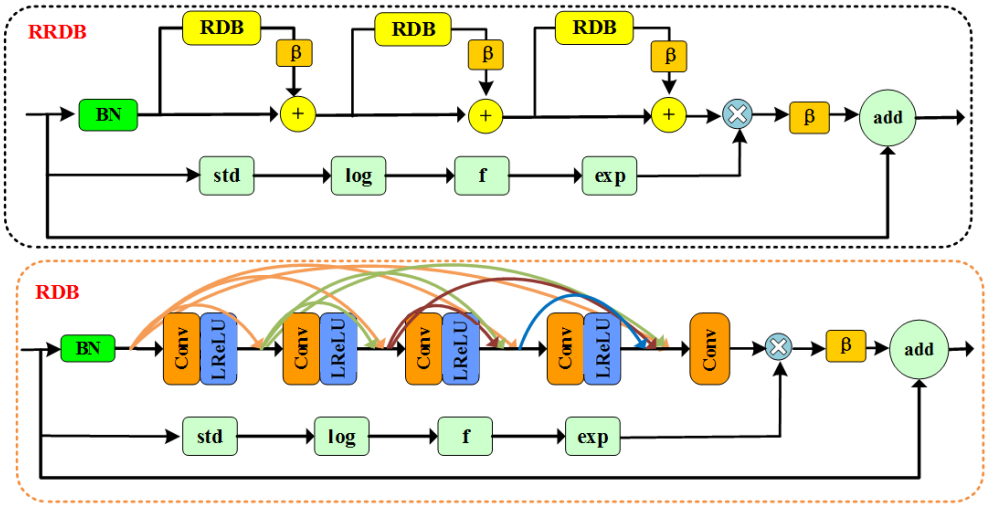

FIGURE 4. Improved RDB and RRDB blocks

The Trunk module is constructed by eight RRDB blocks with residual architecture, so it has better global feature extraction ability. To improve the local feature extraction ability, we use our proposed GM block and new BN block to construct a dense network with short skip connection characteristics. The proposed dense network is named GM_DB which is shown in Figure 5. It consists of a proposed BN block, four GM blocks, four Leaky-ReLU layers, and a 3×3 convolution layer. The GM_DB feeds each layer to all subsequent layers, which is useful for extracting more effective local features. To further achieve the extraction and fusion of global and local features, we use the GM_DB block as a sub-block to construct GM_RDB with residual architecture. GM_RDB is shown in Figure 5. It consists of three GM_DB blocks and an improved BN block. It has a residual structure, so it is useful to extract more effective global features. We used three different blocks that are RRDB, GM_DB, and GM_RDB to construct the feature extraction network. The RRDB and GM_RDB are mainly used to extract global features. The GM_DB is mainly used to extract local features. Therefore, the feature extraction network can extract more effective local and global features, which is useful to improve the quality of the recovered image by the generative network. Besides, we also introduce the GM block at the end of the Trunk module, and the Trunk_GM module is shown in Figure 1 to realize the global fusion of features.

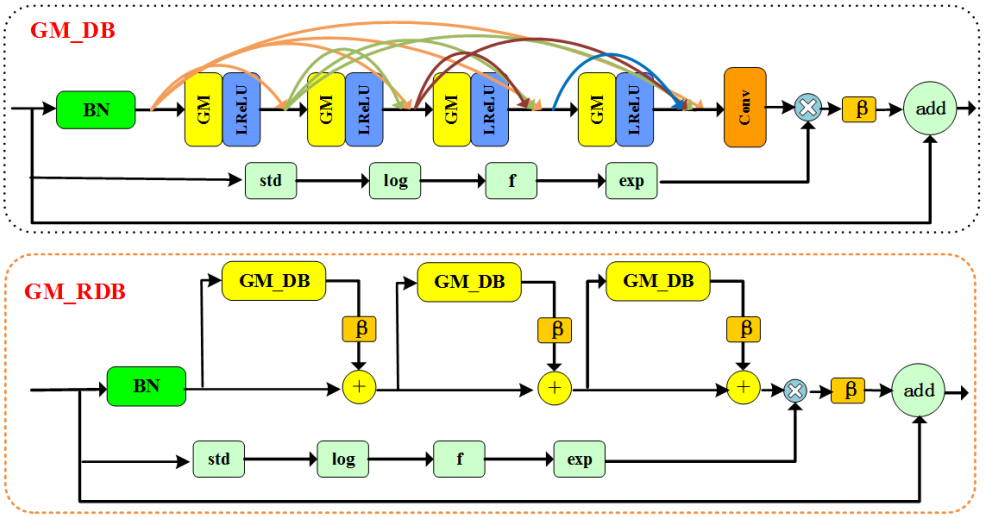

FIGURE 5. Proposed GM_DB and GM_RDB blocks

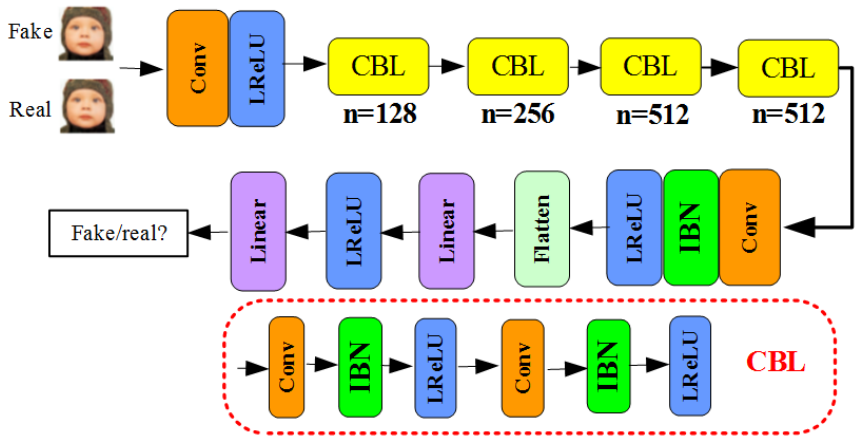

FIGURE 6. Improved adversarial network

3.4. **Improved adversarial network.** The adversarial network is used to determine whether the input image is the recovered image or the original high-resolution image. It improves image recovery in generative networks through the gaming of generative and adversarial networks. To improve the discriminatory ability of the adversarial network, we proposed an improved adversarial network based on ESRGAN. The improved adversarial network is shown in Figure 6. The improved adversarial network consists of a 3×3 convolution layer with 64 channels, an activation function, four CBL modules with the channel number of 128, 256, 512, and 512, a 4×4 convolution layer with 512 channels, an improved BN layer, an activation function layer, a flatten layer that is used to flatten the feature map, a linear layer, an activation layer, and a linear layer. The last linear layer is mainly used to discriminate whether the image is the original high-resolution image or the generated high-resolution image. The CBL module is used to extract features. The CBL module consists of a 4×4 convolutional layer, a modified BN layer, an activation function layer, a 3×3 convolutional layer, a modified BN layer, and a Leaky-ReLU activation function layer. The stride is 2 and the padding is 1 for the 4×4 convolution used in the CBL module. To reduce the effect of normal batch normalization on image quality, we first introduce two of our proposed BN (IBN) blocks into the CBL block of the adversarial network in ESRGAN. The proposed IBN block is shown in Figure 2. The traditional batch

normalization will reduce the standard deviation of feature pixels, which causes the loss of the image edge feature information and further affects the quality of recovered images. Therefore, we design an adaptive standard deviation of feature pixels modulator to amplify the deviation of feature pixels and introduce it to the traditional batch normalization to construct the IBN block. Besides, we also introduce the IBN behind the last convolution in the adversarial network to further improve the performance of discrimination ability. The improved adversarial network can extract more edge feature information by introducing an IBN block and obtain more accurate discrimination results, which is useful to optimize the generative network to improve the quality of recovered images.

3.5. **Loss function.** We directly use the loss function of ESRGAN as the loss function of our proposed method. The loss function consists of three parts: perceptual loss, pixel loss, and adversarial loss. It can be expressed as

$$L_G = L_P + \alpha_1 L_C + \alpha_2 L_A \tag{2}$$

where L_P is perceptual loss function. L_C is pixel loss function. L_A is the adversarial loss function, and α_1 and α_2 are weighting factors of the loss function. The perceptual loss function is expressed as follows:

$$L_P = E\left\{\|\varphi[G(x)] - \varphi(y)\|_2^2\right\} \tag{3}$$

where x is the lower resolution image; $G(x)$ is the high-resolution image generated by the generative network; y is the super-resolution image corresponding to the lower resolution image x. $\varphi()$ is the output features of the VGG19 network. The pixel loss function is expressed as follows:

$$L_C = E[\|G(x) - y\|_1] \tag{4}$$

where $\|\ \|_1$ is the L1 distance. $G(x)$ and y have the same meaning as (3). The adversarial loss function is expressed as follows:

$$L_G = E\left[(C(G(x)) - E[C(y)] - 1)^2\right] + E\left[(C(y) - E[C(G(x))] + 1)^2\right] \tag{5}$$

where $C()$ is the output of the adversarial network.

4. **Simulation and Discussion.** In this section, we use the DIV2K as a training dataset and test the performance of different methods on four different datasets Set5, Set14, BSD200, and Urban100. We compare our method with five image super-resolution reconstruction methods: EnhanceNet [29], SRGAN [13], ESRGAN [11], R-SRGAN [24], and SAM+VAM [21] method. We use two evaluation indexes to quantitatively evaluate each method. They are PSNR (peak signal to noise ratio) and SSIM (structural similarity). The operation system is Ubuntu18.04, and GPU is NVIDIA GTX1080Ti. The size of the input image is 128×128. For a fair comparison, we use the same Adam optimizer and parameters with $\beta_1 = 0.9$ and $\beta_2 = 0.99$ that are used in ESRGAN [11,21] to optimize the loss function of the network. The learning rate is 2×10^{-4}. The batch size is 8, and other parameters are set to zero to initialize the parameters. $\alpha_1 = 0.01$, $\alpha_2 = 5 \times 10^{-3}$. When the number of feature maps increases (too many residual blocks) beyond a certain level, the training process will be numerically unstable. To solve this problem, we introduce residual scaling $\beta = 0.2$, which largely stabilizes the training process without losing feature information.

4.1. **Datasets and metrics.** The DIV2K dataset contains 800 high-quality images with 2K resolution. The dataset contains people, natural landscapes, human landscapes, etc. The Set5 dataset, Set14 dataset, BSD200 dataset and Urban100 dataset consist of real images of natural landscapes, people, animals, buildings, etc. We rotate the image

randomly to expand dataset and crop all images to size 128×128. We use MATLAB software to downsample the images in test datasets at $\times 2$, $\times 3$, and $\times 4$, respectively. The downsampled images are used to test the performance of different methods.

We use two evaluation indexes: PSNR and SSIM to quantitatively compare our method with other methods. The PSNR can be expressed as

$$PSNR = 10 \log_{10} \left(\frac{MAX_I}{RMSE} \right) \tag{6}$$

where MAX_I is the maximum value of image pixel coloration, and $RMSE$ is the root mean square error. The RMSE can be expressed as

$$RMSE = \sqrt{\frac{1}{mn} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} \|I_{pre}(i,j) - I_{gt}(i,j)\|^2} \tag{7}$$

where m and n denote the size of the image. I_{pre} and I_{gt} represent the original super-resolution image and the generated high-resolution image, respectively.

The SSIM is expressed as follows:

$$SSIM(x,y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)} \tag{8}$$

where μ_x and μ_y are the mean values of x and y, σ_x^2 and σ_y^2 are the variances of x and y; and σ_{xy} is the covariance of x and y. The more SSIM closes to 1, the higher quality of the recovered image is.

4.2. **Simulation on four test datasets.** We randomly select four super-resolution images from the Set5 dataset, Set14 dataset, BSD200 dataset, and Urban100 dataset. We obtain the low-resolution images that are used as test images by downsampling the four super-resolution images using a bicubic kernel with a scaling factor of $\times 4$. We use EnhanceNet, SRGAN, ESRGAN, R-SRGAN, SAM+VAM, and our proposed method to recover high-resolution images from low-resolution images. The images in Figure 7 to Figure 10 are selected from the Set5 dataset, Set14 dataset, BSD200 dataset and Urban100 dataset, respectively. From Figure 7 to Figure 10, figure (a) is the original high-resolution image, figure (b) is the low-resolution image obtained by downsampling with a scaling factor of $\times 4$, and figure (c) is the local enlargement image from the original high-resolution images. The images in figure (d) to figure (i) are the local enlargement images of recovered images by SRGAN, EnhanceNet, ESRGAN, R-SRGAN, SAM+VAM, and our proposed method, respectively.

In Figure 7, the image recovered by SRGAN has a wide range of distortion and missing pixel blocks. The recovered images by EnhanceNet, ESRGAN, R-SRGAN, and SAM+VAM also have some blur. Compared with other methods, the eyelashes and freckles in the recovered image by our proposed method are clearer. In Figure 8, the image recovered by EnhanceNet is the most blurred. The image recovered by SRGAN contains more pixel blocks, and the edges of the image are not smooth. Compared with other methods, the leaves and flower hearts of the image recovered by our method are clearer in visual perception, and the recovered image is more lost to the original image. In Figure 9, compared with other methods, the mouth is clearer and the edges of the branches are smoother for the recovered image by our proposed method. In Figure 10, the wall lines of recovered images by SRGAN and R-SRGAN are curved and not smooth, and the glasses of the windows are not clear and have large pseudo-shadows. The wall lines are smoother, and the tree branches in glass are more clear for the image recovered by our method than

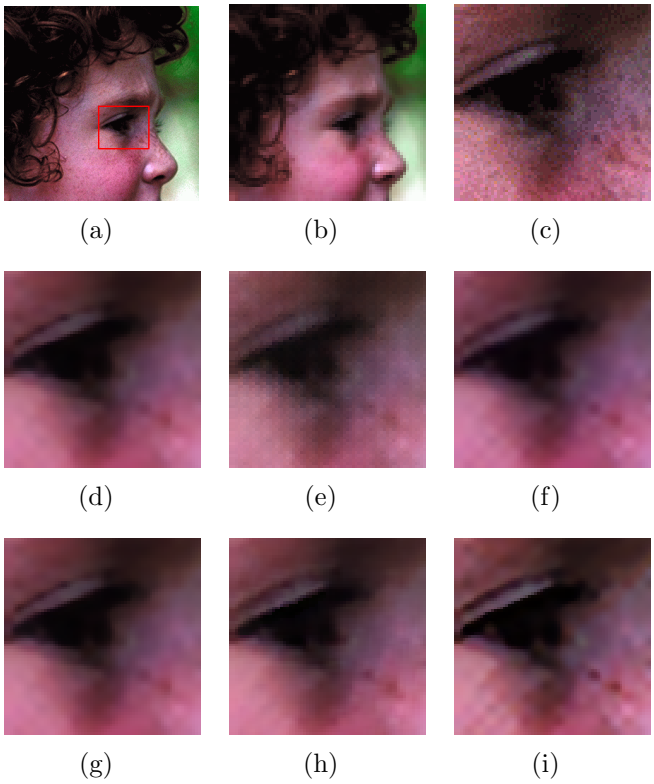

FIGURE 7. Image selected from Set5 dataset, downsampling image, local enlargement images of the original image, and recovered images by different methods

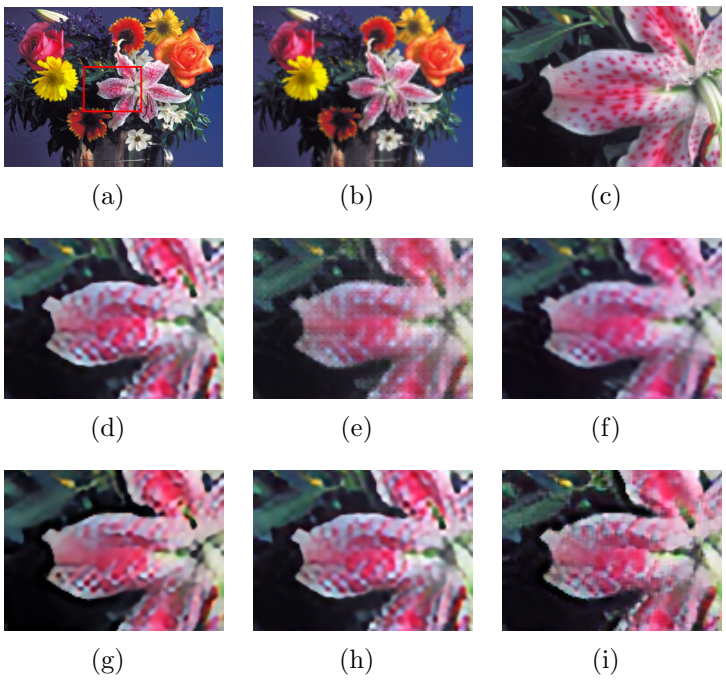

FIGURE 8. Image selected from Set14 dataset, downsampling image, local enlargement images of the original image, and recovered images by different methods

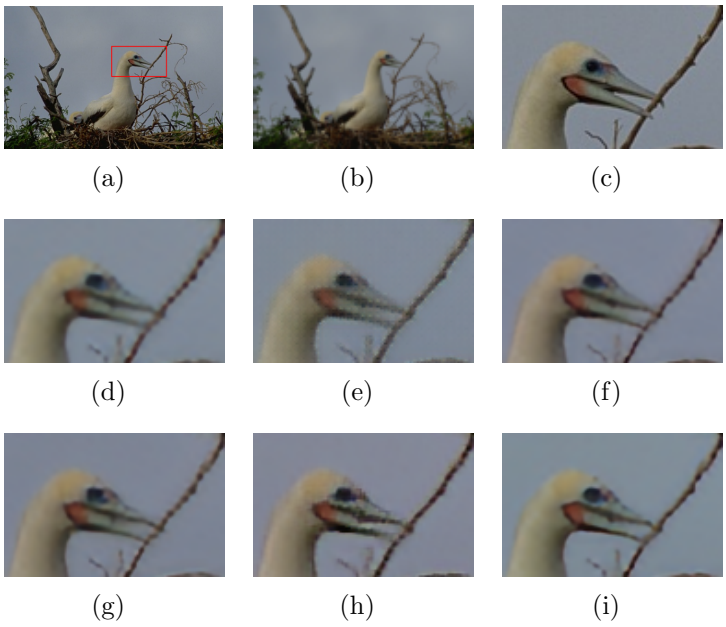

FIGURE 9. Image selected from BSD200 dataset, downsampling image, local enlargement images of the original image, and recovered images by different methods

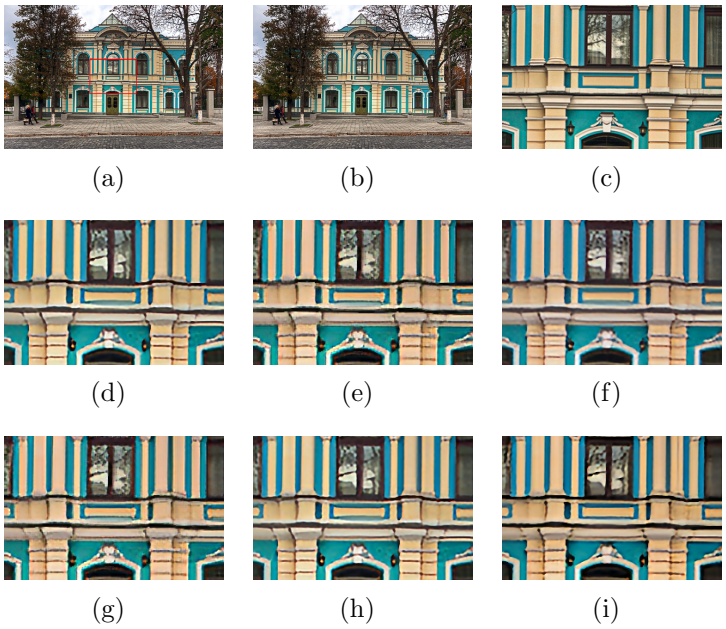

FIGURE 10. Image selected from Urban100 dataset, downsampling image, local enlargement images of the original image, and recovered images by different methods

other methods. From Figure 7 to Figure 10, the images recovered by our methods contain more detailed information and are closer to the ground-truth image.

We also use all images in different datasets to test the performance of different methods. Firstly, we downsample the original images using a bicubic kernel with a scaling factor of $\times 2$. The results are shown in Table 1. For the Set5 dataset, the SAM+VAM has

the greatest PSNR (34.87 dB), followed by our proposed method (34.72 dB). The R-SRGAN has the greatest SSIM (0.9401), followed by the SAM+VAM (0.9397) and our proposed method (0.9396). For the Set14 dataset, our proposed method has the greatest PSNR (32.12 dB) and SSIM (0.9013), followed by the R-SRGAN method. For the BSD200 dataset, the EnhanceNet has the greatest PSNR (32.17 dB), followed by R-SRGN (32.02 dB) and our method (31.96 dB). The SRGAN has the greatest SSIM (0.8986), followed by the R-SRGAN (0.8944) and our proposed method (0.8895). For the Urban100 dataset, our proposed method has the greatest PSNR (30.92 dB) and SSIM (0.8975), followed by the EnhanceNet method and R-SRGAN. Although our proposed method has not had the greatest PSNR and SSIM on the Set5 dataset and the BSD200 dataset, it has the greatest PSNR and SSIM on the Set14 dataset and Urban100 dataset and the greatest average PSNR and SSIM.

Table 1. Comparison of recovered performance on four datasets with a downsampling scaling factor of $\times 2$

	PSNR/SSIM				
	Set5	Set14	BSD200	Urban100	Average
SRGAN	33.33/0.9365	31.44/0.9004	31.33/**0.8986**	29.26/0.8904	31.34/0.9065
EnhanceNet	33.99/0.9354	31.88/0.8963	**32.17**/0.8870	30.21/0.8968	32.06/0.9039
ESRGAN	34.23/0.9383	31.53/0.8927	31.23/0.8891	29.88/0.8924	31.72/0.9031
R-SRGAN	34.21/**0.9401**	32.01/0.8965	32.02/0.8944	30.02/0.8954	32.07/0.9066
SAM+VAM	**34.87**/0.9397	31.86/0.8942	31.04/0.8767	29.39/0.8955	31.79/0.9015
Our method	34.72/0.9396	**32.12/0.9013**	31.96/0.8895	**30.92/0.8975**	**32.43/0.9070**

Secondly, we downsample the original images using a bicubic kernel with a scaling factor of $\times 3$. The results are shown in Table 2. For the four datasets, our proposed method has the greatest PSNR and SSIM. Compared with ESRGAN, the values of PSNR of our proposed method on Set5 dataset, Set14 dataset, BSD200 dataset and Urban100 dataset are increased by 1.51 dB (4.86%), 0.75 dB (2.64%), 1.20 dB (4.27%) and 1.09 dB (4.03%), respectively. Compared with ESRGAN, the values of SSIM of our proposed method on the Set5 dataset, Set14 dataset, BSD200 dataset, and Urban100 dataset increased by 1.08%, 2.48%, 4.14%, and 11.88%, respectively. Compared with R-SRGAN, the values of PSNR of our proposed method on Set5 dataset, Set14 dataset, BSD200 dataset and Urban100 dataset are increased by 1.05 dB (3.33%), 0.06 dB (0.21%), 1.10 dB (3.90%) and 0.82 dB (3.00%), respectively. Compared with R-SRGAN, the values of SSIM of our proposed method on the Set5 dataset, Set14 dataset, BSD200 dataset, and Urban100 dataset increased by 0.69%, 2.11%, 4.14%, and 11.14%, respectively. Compared with SAM+VAM, the values of PSNR of our proposed method on Set5 dataset, Set14 dataset,

Table 2. Comparison of recovered performance on four datasets with a downsampling scaling factor of $\times 3$

	PSNR/SSIM				
	Set5	Set14	BSD200	Urban100	Average
SRGAN	30.46/0.8861	28.28/0.8006	27.95/0.7841	26.52/0.7046	28.30/0.7939
EnhanceNet	30.96/0.8865	28.86/0.8106	28.81/0.7741	27.37/0.7209	29.00/0.7980
ESRGAN	31.08/0.8907	28.45/0.8072	28.08/0.7843	27.05/0.7247	28.67/0.8017
R-SRGAN	31.54/0.8941	29.14/0.8101	28.18/0.7843	27.32/0.7257	29.05/0.8036
SAM+VAM	32.01/0.8920	28.74/0.8041	28.11/0.7825	27.02/0.7256	28.97/0.8011
Our method	**32.59/0.9003**	**29.20/0.8272**	**29.28/0.8168**	**28.14/0.8109**	**29.80/0.8388**

BSD200 dataset and Urban100 dataset are increased by 0.58 dB (1.87%), 0.46 dB (1.60%), 1.17 dB (4.16%) and 1.12 dB (4.15%), respectively. Compared with SAM+VAM, the values of SSIM of our proposed method on the Set5 dataset, Set14 dataset, BSD200 dataset, and Urban100 dataset are increased by 0.93%, 2.87%, and 4.38%, and 11.75%, respectively.

Thirdly, we downsample the original images using a bicubic kernel with a scaling factor of $\times 4$. The results are shown in Table 3. For the four datasets, our proposed method still has the greatest PSNR and SSIM. Compared with ESRGAN, the values of PSNR of our proposed method on Set5 dataset, Set14 dataset, BSD200 dataset and Urban100 dataset are increased by 1.81 dB (6.26%), 2.02 dB (7.55%), 1.88 dB (7.11%) and 2.25 dB (9.27%), respectively. Compared with ESRGAN, the values of SSIM of our proposed method on the Set5 dataset, Set14 dataset, BSD200 dataset, and Urban100 dataset are increased by 4.42%, 9.52%, 6.51%, and 13.47%, respectively. Compared with R-SRGAN, the values of PSNR of our proposed method on Set5 dataset, Set14 dataset, BSD200 dataset and Urban100 dataset are increased by 0.68 dB (2.26%), 1.43 dB (5.23%), 0.48 dB (7.11%) and 1.27 dB (5.03%), respectively. Compared with R-SRGAN, the values of SSIM of our proposed method on the Set5 dataset, Set14 dataset, BSD200 dataset, and Urban100 dataset are increased by 5.87%, 7.98%, 5.96%, and 11.79%, respectively. Compared with SAM+VAM, the values of PSNR of our proposed method on Set5 dataset, Set14 dataset, BSD200 dataset and Urban100 dataset are increased by 0.85 dB (2.85%), 1.34 dB (4.88%), 0.82 dB (2.98%) and 1.37 dB (5.45%), respectively. Compared with SAM+VAM, the values of SSIM of our proposed method on the Set5 dataset, Set14 dataset, BSD200 dataset, and Urban100 dataset are increased by 4.28%, 8.67%, 6.36%, and 11.51%, respectively.

Table 3. Comparison of recovered performance on four datasets with a downsampling scaling factor of $\times 4$

	PSNR/SSIM				
	Set5	Set14	BSD200	Urban100	Average
SRGAN	27.65/0.8220	26.39/0.7302	26.18/0.7053	24.01/0.6476	26.06/0.7263
EnhanceNet	28.46/0.8378	26.04/0.7045	27.02/0.7213	24.66/0.6683	26.55/0.7330
ESRGAN	28.91/0.8568	26.76/0.7468	26.43/0.7280	24.26/0.6700	26.59/0.7504
R-SRGAN	30.04/0.8451	27.35/0.7575	27.83/0.7318	25.24/0.6801	27.62/0.7536
SAM+VAM	29.87/0.8580	27.44/0.7536	27.49/0.7290	25.14/0.6818	27.49/0.7556
Our method	**30.72/0.8947**	**28.78/0.8179**	**28.31/0.7754**	**26.51/0.7603**	**28.58/0.8121**

Based on the above analysis, our proposed method has the greatest PSNR and SSIM on the Set14 dataset and Urban100 dataset in Table 1 and has the greatest PSNR and SSIM on all test datasets in Table 2 and Table 3. On the whole, our proposed method has better performance in recovering images than others, especially for recovering images from lower-resolution images. Compared with ESRGAN, the proposed generative network uses more types of modules to extract features. The extracted features contain more abundant information. The different lengths of short skip connections are used to fuse different depth features by contacting the extracted features of the bottom and top layers to extract more effective features. Besides, the residual and dense connections in ESRGAN are also still used, which strengthens the transmission of feature information. Therefore, compared with ESRGAN, the proposed network can extract more effective features, which can improve the quality of recovered images.

5. **Conclusions.** This paper proposed a super-resolution image reconstruction based on an improved generative adversarial network. It designs a new batch normalization block to avoid gradient explosion and gradient disappearance in the generative network. The new block consists of the normal batch normalization block and an adaptive standard deviation of the feature pixels modulator. The adaptive standard deviation modulator is used to reduce the effect of normal batch normalization on recovered image quality. It also introduces the new batch normalization block to adversarial networks and feature extraction blocks that are used to construct feature extraction networks in the generative network. Besides, it designs a new block (named GM block) with double branches and residual structure to extract and fuse more effective features. It uses the GM block to construct a dense network with skip connection characteristics as one part of the feature extraction network. The new dense network is also used as a sub-block to construct a complex network with a residual structure that is used as one part of the feature extraction network. Finally, it also introduces the GM block into the end of the feature extraction network to fuse more effective features. Four datasets are used as test datasets.

Compared with ESRGAN, the average PSNR is increased by 0.71 dB (2.24%), 1.13 dB (3.94%), and 1.99 dB (7.48%) for downsampling with scaling factors of $\times 2$, $\times 3$, and $\times 4$, respectively. The average SSIM is increased by 0.43%, 4.63%, and 8.22% for downsampling with scaling factors of $\times 2$, $\times 3$, and $\times 4$. Compared with SAM+VAM, the average PSNR is increased by 0.64 dB (2.01%), 0.83 dB (2.87%), and 1.09 dB (3.97%) for downsampling with scaling factors of $\times 2$, $\times 3$, and $\times 4$, respectively. The average SSIM is increased by 0.61%, 4.71%, and 7.48% for downsampling with scaling factors of $\times 2$, $\times 3$, and $\times 4$. Compared with R-SRGAN, the average PSNR is increased by 0.36 dB (1.12%), 0.75 dB (2.58%), and 0.96 dB (3.48%) for downsampling with scaling factors of $\times 2$, $\times 3$, and $\times 4$, respectively. The average SSIM is increased by 4.38%, 3.48%, and 7.76% for downsampling with scaling factors of $\times 2$, $\times 3$, and $\times 4$. On the whole, the quality of the recovered image by our proposed method is higher than other methods.

Although the recovered images by our proposed network are more close to the original high-resolution images, the network structure is relatively complex, which requires hardware with larger computing power. Therefore, we will research how to reduce the network complexity and improve the speed of image reconstruction while ensuring the quality of the recovered images in our future work.

Acknowledgment. This research has been funded by the National Natural Science Foundation of China (61271115), Scientific and Technological Developing Scheme of Jilin Province (YDZJ202101ZYTS172), Research Foundation of Education Bureau of Jilin Province (JJKH20210042KJ, JJKH20210095KJ) and Doctoral Scientific Research Foundation of Beihua University (20171424).

REFERENCES

[1] V. Hemamalini, R. Subramanian, N. Subramanian, M. Sambath, D. Thiyagarajan, B. K. Singh and A. Raghuvanshi, Food quality inspection and grading using efficient image segmentation and machine learning-based system, *Journal of Food Quality*, DOI: 10.1155/2022/5262294, 2022.

[2] J. Zhang, H. Zhang, J. Zhang, X. Peng and X. Shi, Sparse reconstruction method based on starlet transform for high noise astronomical image denoising, *International Journal of Innovative Computing, Information and Control*, vol.16, no.5, pp.1639-1654, DOI: 10.24507/ijicic.16.05.1639, 2020.

[3] T. Hiraoka, Generation of colorful stripe patchwork images by filtering based on ratio between RGB and inverse filter, *ICIC Express Letters*, vol.15, no.11, pp.1197-1201, 2021.

[4] T. Hiraoka, Generation of cell-like images with variable pattern size from RGB-D images, *ICIC Express Letters*, vol.16, no.7, pp.723-729, 2022.

[5] X. Xu and Q. Huang, Depth map restoration algorithm based on improved super-resolution and FMM by using weight function, *International Journal of Innovative Computing, Information and Control*, vol.18, no.2, pp.577-590, DOI: 10.24507/ijicic.18.02.577, 2022.
[6] W. Ahmad, H. Ali, Z. Shah and S. Azmat, A new generative adversarial network for medical images super resolution, *Scientific Reports*, vol.12, 9533, DOI: 10.1038/s41598-022-13658-4, 2022.
[7] S. Zhu, C. Zhou and Y. Wang, Super resolution reconstruction method for infrared images based on pseudo transferred features, *Displays*, vol.74, 102187, DOI: 10.1016/j.displa.2022.102187, 2022.
[8] Z. Zhang, Z. Xiong, B. Zhang, Y. Yang and E. Fu, Detection for small target ship in remote sensing image based on super resolution reconstruction technology, *Journal of Northeast Electric Power University*, vol.42, no.4, pp.33-40, 2022.
[9] X. Hu, H. Guo and R. Zhu, Image super-resolution reconstruction based on hybrid deep convolutional network, *Journal of Computer Applications*, vol.40, no.7, pp.2069-2076, DOI: 10.11772/j.issn.1001-9081.2019122149, 2020.
[10] J. Zhang, M. Shao, L. Yu and Y. Li, Image super-resolution reconstruction based on sparse representation and deep learning, *Signal Processing: Image Communication*, vol.87, 115925, DOI: 10.1013/j.image.2020.115925, 2020.
[11] X. Wang, K. Yu, S. Wu, J. Gu, Y. Liu, C. Dong, C. C. Loy, Y. Qiao and X. Tang, ESRGAN: Enhanced super-resolution generative adversarial networks, *Proc. of the European Conference on Computer Vision (ECCV) Workshops*, 2018.
[12] I. J. Goodfellow, J. P. Abadie, M. Mirza, B. Xu, D. W. Farley, S. Ozair, A. Courville and Y. Bengio, Generative adversarial nets, *Proc. of the 27th International Conference on Neural Information Processing Systems*, vol.2, pp.2672-2680, 2014.
[13] C. Ledig, L. Theis, J. Caballero et al., Photo-realistic single image super-resolution using a generative adversarial network, *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp.105-114, DOI: 10.1109/CVPR.2017.19, 2017.
[14] Y. Hu, M. Jing, Y. Jiao and K. Sun, Images super-resolution using improved generative adversarial networks, *2021 3rd International Conference on Intelligent Control, Measurement and Signal Processing and Intelligent Oil Field (ICMSP)*, pp.254-258, DOI: 10.1109/ICMSP53480.2021.9513225, 2021.
[15] X. Sun, Z. Zhao, Z. Song, J. Liu, X. Yang and C. Zhou, Image super-resolution reconstruction using generative adversarial networks based on wide-channel activation, *IEEE Access*, vol.8, pp.33838-33854, DOI: 10.1109/ACCESS.2020.2974759, 2020.
[16] F. Nan, Q. Zeng, Y. Xing and Y. Qian, Single image super-resolution reconstruction based on the ResNext network, *Multimedia Tools and Applications*, vol.79, no.45, pp.34459-34470, 2020.
[17] M. Guo, Z. Zhang, H. Liu and Y. Huang, NDSRGAN: A novel dense generative adversarial network for real aerial imagery super-resolution reconstruction, *Remote Sensing*, vol.14, no.7, 1574, DOI: 10.3390/rs14071574m, 2022.
[18] X. Dou, C. Li, Q. Shi and M. Liu, Super-resolution for hyperspectral remote sensing images based on the 3D attention-SRGAN network, *Remote Sensing*, vol.12, no.7, 1204, DOI: 10.3390/rs12071204, 2020.
[19] Y. Zhang, Y. Tian, Y. Kong, B. Zhong and Y. Fu, Residual dense network for image restoration, *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol.43, no.7, pp.2480-2495, 2021.
[20] M. Shao, W. Zhang, W. Zuo and D. Meng, Multiscale generative adversarial inpainting network based on cross-layer attention transfer mechanism, *Knowledge Based Systems*, vol.196, 105778, DOI: 10.1016/j.knosys.2020.105778, 2020.
[21] C.-C. Hsu and C.-H. Lin, Dual reconstruction with densely connected residual network for single image super-resolution, *2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW)*, pp.3643-3650, DOI: 10.1109/ICCVW.2019.00449, 2019.
[22] T. Shang, Q. Dai, S. Zhu, T. Yang and Y. Guo, Perceptual extreme super resolution network with receptive field block, *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp.1778-1787, DOI: 10.1109/CVPRW50498.2020.00228, 2020.
[23] N. C. Rakotonirina and A. Rasoanaivo, ESRGAN+: Further improving enhanced super-resolution generative adversarial network, *2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp.3637-3641, DOI: 10.1109/ICASSP40776.2020.9054071, 2020.
[24] X. Xue, X. Zhang, H. Li and W. Wang, Research on GAN-based image super-resolution method, *2020 IEEE International Conference on Artificial Intelligence and Computer Applications (ICAICA)*, pp.602-605, DOI: 10.1109/ICAICA50127.2020.9182617, 2020.

[25] Y. Wang, Single image super-resolution with U-Net generative adversarial networks, *2021 IEEE the 4th Advanced Information Management, Communicates, Electronic and Automation Control Conference (IMCEC)*, pp.1835-1840, DOI: 10.1109/IMCEC51613.2021.9482317, 2021.
[26] E. Schönfeld, B. Schiele and K. Anna, A U-Net based discriminator for generative adversarial networks, *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp.8204-8213, DOI: 10.1109/CVPR42600.2020.00823, 2020.
[27] H. Zhao, X. Kong, J. He, Y. Qiao and C. Dong, Efficient image super-resolution using pixel attention, in *Computer Vision – ECCV 2020 Workshops. ECCV 2020. Lecture Notes in Computer Science*, A. Bartoli and A. Fusiello (eds.), Cham, Springer, 2020.
[28] Y. Jo, S. Yang and S. J. Kim, Investigating loss functions for extreme super-resolution, *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp.1705-1712, DOI: 10.1109/CVPRW50498.2020.00220, 2020.
[29] M. S. M. Sajjadi, B. Schölkopf and M. Hirsch, EnhanceNet: Single image super-resolution through automated texture synthesis, *2017 IEEE International Conference on Computer Vision (ICCV)*, pp.4501-4510, DOI: 10.1109/ICCV.2017.481, 2017.

Author Biography

Liquan Zhao received the B.Eng. degree from Harbin University of Science and Technology, China, 2005; the Ph.D. degree in Communication and Information System, from Harbin Engineering University, China, 2009.

Prof. Zhao is currently a full-time professor at the College of Electric Engineering, Northeast Electric Power University, China. His research interests include artificial intelligence and object detection. He has published over 40 papers in journals and conferences.

Jingjing Wu received the B.Eng. degree in Electronic Information Science and Technology from the Shanxi Agricultural University, Shanxi, China, in 2019. She is studying for a master's degree in Northeast Electric Power University, Jilin, China. Her main research interests include deep learning and generative adversarial networks.

Yanfei Jia received the B.Eng. degree from Harbin University of Science and Technology, China, 2005; the Ph.D. degree in Communication and Information System, from Harbin Engineering University, China, 2018.

Dr. Jia is currently a full-time associate professor at the College of Electric and Information Engineering, Beihua University, China. Her research interests include artificial intelligence, wireless sensor network, and blind source separation. She has published over 10 papers in journals and conferences.