IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received November 4, 2020, accepted November 17, 2020, date of publication November 25, 2020,
date of current version December 10, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3040424

Image Super-Resolution Reconstruction Based on a Generative Adversarial Network

YUN WU, LIN LAN, HUIYUN LONG, GUANGQIAN KONG, XUN DUAN, AND CHANGZHUAN XU

School of Computer Science and Technology, Guizhou University, Guiyang 550025, China

Corresponding author: Lin Lan (l_lanlin@163.com)

This work was supported in part by the National Natural Science Foundation of China under Grant [2018]61741124, and in part by the Science Planning Project of Guizhou Province, Guizhou Science and Technology Cooperation Platform Talent [2018] under Grant 5781.

ABSTRACT In the field of computer vision, super-resolution reconstruction techniques based on deep learning have undergone considerable advancement; however, certain limitations remain, such as insufficient feature extraction and blurred image generation. To address these problems, we propose an image super-resolution reconstruction model based on a generative adversarial network. First, we employ a dual network structure in the generator network to solve the problem of insufficient feature extraction. The dual network structure is divided into an upsample subnetwork and a refinement subnetwork, which upsample and optimize a low-resolution image, respectively. In a scene with large upscaling factors, this structure can reduce the negative effect of noise and enhance the utilization of high-frequency details, thereby generating high-quality reconstruction results. Second, to generate sharper super-resolution images, we use the perceptual loss, which exhibits a fast convergence and excellent visual effect, to guide the generator network training. We apply the ResNeXt-50-32 $\times$ 4d network, which has few parameters and a large depth, to calculate the loss to obtain a reconstructed super-resolution image that is highly realistic. Finally, we introduce the Wasserstein distance into the discriminator network to enhance the discrimination ability and stability of the model. Specifically, this distance is employed to eliminate the activation function in the last layer of the network and avoid the use of the logarithm in calculating the loss function. Extensive experiments on the DIV2K, Set5, Set14, and BSD100 datasets demonstrate the effectiveness of the proposed model.

INDEX TERMS Deep learning, dual network structure, generative adversarial network, perceptual loss, super-resolution.

I. INTRODUCTION

With the development of information technology and progress in the digital age, the amount of all types of information is rapidly increasing. As the main carrier of information dissemination, images are widely used in various fields. However, due to the limitations of hardware and cost, directly obtaining high-resolution (HR) images is highly challenging, and low-resolution (LR) images are inadequate in the context of specific scenes. Therefore, increasing the image resolution and enhancing the image quality have emerged as critical problems in recent years, having notable research significance and application value.

In the field of image processing, image super-resolution (SR) reconstruction is a fundamental and critical issue, as a key step in many image related applications, including object tracking [1], object detection [2], semantic segmentation [3], image annotation [4], image inpainting [5], and image classification [6]. In these applications, a higher resolution of the image corresponds to more satisfactory results.

In general, image super-resolution reconstruction algorithms can be divided into interpolation based, reconstruction based and learning based SR reconstruction. Among these algorithms, the interpolation based image SR reconstruction algorithm involves a simple calculation and it easy to understand; however, the reconstructed image is usually blurry and involves missing details (especially high-frequency details) and ringing effects. The reconstruction based image SR reconstruction algorithm considers the degradation of the image and combines the prior knowledge to achieve satisfactory results. However, under high upscaling factors or when the number of input images is small, the reconstructed

The associate editor coordinating the review of this manuscript and approving it for publication was Shiqi Wang.

image quality is low, and the image is excessively smooth. The learning based image SR reconstruction algorithm fully considers the mapping relationship among images, solves the problems of missing details and blurring of the reconstructed image to a large extent, and achieves enhanced results.

In recent years, image SR reconstruction based on deep learning [7] has achieved remarkable results; nevertheless, certain challenges remain. In particular, the LR image is inherently blurry and damaged, and the image lacks sufficient information regarding the details, which hinder the network learning process. Second, the model feature extraction is insufficient, and most of the spatial information is lost, owing to which, the feature map is excessively rough to suitably describe the image; thus, the generated image is blurry.

To solve the above mentioned problems, we propose a unified end to end convolutional neural network model based on a generative adversarial network (GAN) to enhance the image SR reconstruction process. We divide the network structure into generator and discriminator networks. In the generator network, a dual network structure is implemented, composed of an upsample subnetwork and a refinement subnetwork. The upsampling subnetwork is used to sample the input image to a finer scale to enable the network to extract more detailed information, reduce distortion and enhance the image quality. The refinement subnetwork is used to recover certain missing details and spatial information in the upsampled image to enhance the learning ability of the network and generate clear SR images. In the discriminator network, the Wasserstein distance in the WGAN is incorporated [8], the sigmoid activation function in the last layer of the network is eliminated, and the loss function does not consider the logarithm value, to optimize the network and discriminating ability of the network.

The remaining paper is organized as follows. Section II provides certain background knowledge by describing the related works. Section III describes the proposed image super-resolution reconstruction model and the design of the dual network structure and loss function. Section IV presents the comparison and analysis of the experimental results. Finally, section V describes the concluding remarks and scope for future work.

II. RELATED WORK

A. IMAGE SUPER-RESOLUTION RECONSTRUCTION

Image SR reconstruction refers to the process of recovering the corresponding HR image from a given LR image by using relevant knowledge in the fields of digital image processing and computer vision, by employing specific algorithms and processing techniques [9].

The amount of information contained in an image depends primarily on the image resolution. A higher image resolution corresponds to a clearer edge of the object in the image, and thus, a larger amount of detailed information contained in the image. According to the considered number of LR images, image SR algorithms can be divided into two categories: single image (SISR) and multiple image (MISR). The SISR algorithm reconstructs the corresponding HR image from a given LR image. The MISR algorithm reconstructs the HR image based on a series of LR images, through a specific technique. The LR image is obtained by degrading the HR image. During the image degradation, the high-frequency information of the HR images is often lost. Moreover, because an HR image corresponds to countless LR images, the image SR problem is a typical ill posed problem.

B. TRADITIONAL METHOD

Based on different classification criteria, the image SR reconstruction technology can be divided into different categories. In terms of the number of input LR images, the technique can be divided into SISR and MISR (video) reconstruction. In terms of the transform space, the technique can be divided into frequency and spatial domain SR reconstruction. In terms of the reconstruction algorithms, the technique can be divided into interpolation, reconstruction and learning based SR reconstruction.

The interpolation based image SR reconstruction algorithm regards each image as a point on the image plane. Thus, the SR image is estimated by fitting the unknown pixel information on the plane with the known pixel information, usually using a predefined transformation function or interpolation kernel. The common interpolation based image SR reconstruction algorithms can be categorized as nearest neighbor interpolation [10], bilinear interpolation [11] or bicubic interpolation [12].

The reconstruction based image SR reconstruction algorithm first considers the degradation model of the image. Specifically, the LR image is assumed to be obtained after the HR image has undergone the appropriate motion transformation, blur and noise related processes. Therefore, the algorithm builds a model for the image degradation process and inversely solves the corresponding HR image based on the input LR image [13]. The common reconstruction based image SR reconstruction algorithms include the iterative back projection (IBP) [14], projections onto convex sets (POCS) [15], and maximum a posterior (MAP) techniques.

The learning based image SR reconstruction algorithm uses a considerable amount of training data to learn the correspondence between the LR and HR images and performs the prediction based on the learned mapping relationship to realize the image SR reconstruction. The common learning based image SR reconstruction algorithms include manifold learning, sparse coding and deep learning methods.

C. DEEP LEARNING METHOD

Owing to the promising potential of deep learning in the field of image processing, image SR algorithms based on deep learning have been widely applied. In 2014, Dong *et al.* [16] proposed a CNN based image SR reconstruction algorithm, namely, the super-resolution convolutional neural network (SRCNN) technique. Subsequently, Dong *et al.* [17] improved the SRCNN model and proposed the FSRCNN

model to have a higher image reconstruction rate. Inspired by the VGG [18] and ResNet [19] models, in 2016, Kim *et al.* [20] proposed a very deep image super-resolution (VDSR) model based on the VGG network, which outperformed the SRCNN on the SISR. Subsequently, Kim *et al.* [21] proposed an RCNN based image SR reconstruction algorithm, deeply-recursive convolutional network (DRCN), which outperformed the VDSR. In 2017, Tai *et al.* [22] proposed the deep recursive residual network (DRRN) model to improve the image SR effect. Moreover, in 2017, Lai *et al.* [23] proposed the Laplacian super-resolution network (LapSRN) model with a multicascade structure, which could gradually enlarge the LR image to obtain the required HR image. In 2017, Lim *et al.* [24] proposed the enhanced deep residual network (EDSR) model to remove the redundant modules of SRResNet [26], thereby expanding the model and enhancing the quality of the generated image.

D. GENERATIVE ADVERSARIAL NETWORK (GAN)

In 2014, Goodfellow *et al.* [25] first proposed the generative adversarial network (GAN), which marked a significant advancement in unsupervised learning. The GAN is trained to generate model G through an adversarial process. The generator network G and discriminator network D are trained simultaneously. The two networks compete with each other and undergo alternate optimization during the training process. The generator G trains the input data to generate the corresponding samples to deceive discriminator D, and the discriminator D undergoes continuous training to distinguish between the ground truth and fake data generated by the generator. The objective function of the original GAN can be defined as in equation (1):

$$\min_G \max_D V(D,G) = E_{x\sim p_{data}(x)}\left[\log D(x)\right] + E_{z\sim p_z(z)}\left[\log\left(1 - D\left(G\left(z\right)\right)\right)\right] \quad (1)$$

Here, x represents the ground truth sample, and z represents the random noise variable. G and D are trained to minimize and maximize the probability of the objective function, respectively.

Inspired by the GAN, Ledig *et al.* [26] proposed a GAN based image SR algorithm, SRGAN, in 2017. The method incorporated the GAN model and perceptual loss [27], owing to which, the reconstructed SR image exhibited an excellent visual effect and was more realistic.

III. METHOD

This section describes the overall design of the model. First, we introduce the design of the generator and discriminator networks and present the corresponding network structure and architecture diagrams. Subsequently, we describe the design of the various model components and loss function. Finally, we present the overall objective function of the model.

A. NETWORK STRUCTURE

In this work, we divide the network structure into a generator network and discriminator network. In the generator network, a dual network structure is implemented to solve problems such as insufficient feature extraction and the generation of blurry images. The generator network consists of an upsample subnetwork and refinement subnetwork, which are used jointly to solve the problems of SR and optimization, to ensure that the network can extract a larger amount of image feature information. Moreover, the perceptual loss is used to ensure that the network can recover more detailed information, enhance the learning ability of the network, and generate clear SR images. In the discriminator network, the Wasserstein distance is incorporated in the WGAN to optimize the network to enhance the discrimination ability of the network and model stability.

1) GENERATOR NETWORK

The generator network has a dual network structure, including an upsample subnetwork and refinement subnetwork. The overall network corresponds to a deep CNN architecture. The dual network structure is employed because the LR images often lose a large amount of high-frequency details during the upsampling process, and thus, a refinement subnetwork is introduced as an additional component. We input the upsampling results of the LR image to the refinement subnetwork and fully extract the features of the image to recover the lost high-frequency details. The generator network architecture is presented in Table 1.

a: UPSAMPLING SUBNETWORK

We input the LR image into the upsample subnetwork and perform convolution operations on the image. Specifically, using the concept of the residual network, we design 8 residual blocks (res_blocks×8) to increase the depth of the network and enhance the model performance to ensure that the network can upsample the image to a finer scale, thereby obtaining more detailed information. During the image upsampling process, the network performs the processing in two sequentially convolutional layers (i.e., the convolutional and deconvolution layers) [28]. Each deconvolution layer is composed of the learned kernels, which upsample the LR image by 2 times and finally output the ×4 SR image. The structure of the upsampling subnetwork is illustrated in Fig. 1.

b: REFINEMENT SUBNETWORK

As shown in Fig. 2, to reconstruct SR images with a higher quality, we design a refinement subnetwork that can fully extract the features of the image and recover the lost high-frequency detail information. The input of the refinement subnetwork is the output of the upsampling subnetwork. In this network, we add the components of the residual block (res_blocks×16) to ensure that the network can recover more detailed information and the spatial information lost in the

TABLE 1. Architecture of the generator network.

	Upsample Subnetwork							Refinement Subnetwork					
Layer	Conv	res_blocks×8	Deconv	Conv	Deconv	Conv	Conv	Conv	res_blocks×16	Conv	Conv	Conv	Conv
Channel	64	64	128	128	256	256	3	64	64	128	256	512	3
Stride	1	1	2	1	2	1	1	1	1	1	1	1	1
Kernel	3	3	3	3	3	3	1	3	3	3	3	3	3
AF	LReLU	LReLU	LReLU	LReLU	LReLU	LReLU	LReLU	LReLU	LReLU	LReLU	LReLU	LReLU	Tanh

*Channel: Number of channels produced by each layer; Stride: Stride of the convolution; Kernel: Size of the convolving kernel; AF: Activation function; Conv: Convolutional layer; res_blocks×8: 8 residual blocks; Deconv: Deconvolution layer; res_blocks×16: 16 residual blocks; LReLU: LeakyReLU.

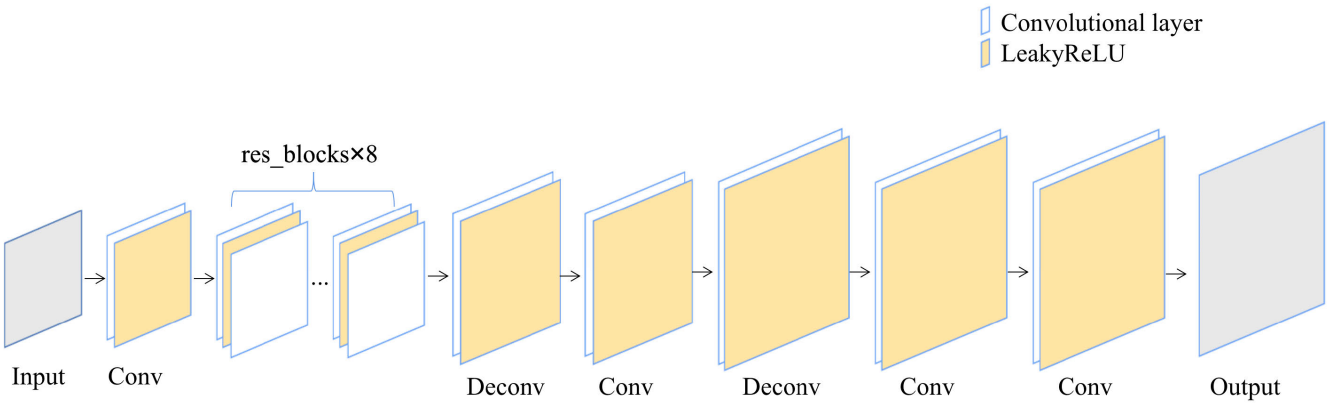

FIGURE 1. Structure of the upsampling subnetwork. 'Input', 'Conv', res_blocks×8', 'Deconv', and 'Output' represent the LR image, convolutional layer, 8 residual blocks, deconvolution layer, and output of the upsampling subnetwork, respectively.

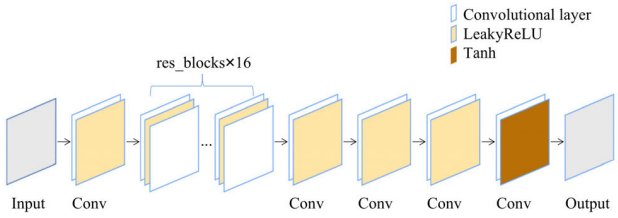

FIGURE 2. Structure of the refinement subnetwork. 'Input', 'Conv', 'res_blocks×16', and 'Output' represent the output result of the upsampling subnetwork, convolutional layer, 16 residual blocks, and reconstructed SR image, respectively.

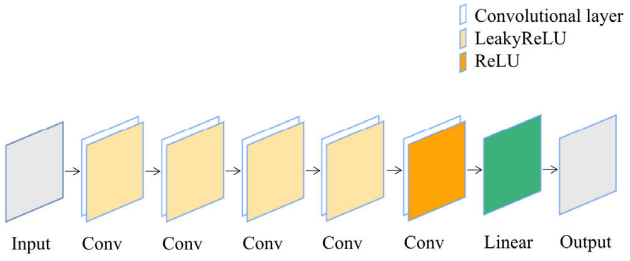

FIGURE 3. Structure of the discriminator network. 'Input', 'Conv', 'Linear', and 'Output' represents the reconstructed SR image, convolutional layer, fully connected layer, and probability that the reconstructed SR image is the real HR image, respectively.

upsampling operation. In this manner, the learning ability of the network is enhanced, and a clear SR image is generated.

In the generator network, data preprocessing is performed on all the input images, and the images are normalized to $[-1, 1]$. To prevent the problem of the vanishing gradient in the network, we apply the LeakyReLU activation function after each convolutional layer, except for the last layer of the refinement subnetwork, to ensure that the neurons with negative outputs are not lost. In the last layer of the refinement subnetwork, the Tanh activation function is applied. Moreover, because the input and output data distributions of the image SR reconstruction are nearly identical and not independently distributed, batch normalization (BN) [29] is not applied in the entire network structure.

2) DISCRIMINATOR NETWORK

The discriminator network adopts the deep CNN architecture. The network structure is shown in Fig. 3, and the architecture is described in Table 2.

TABLE 2. Architecture of the discriminator network.

Discriminator Network						
Layer	Conv	Conv	Conv	Conv	Conv	Linear
Channel	64	128	256	512	512	1
Stride	2	2	2	2	1	-
Kernel	3	3	3	3	3	-
AF	LReLU	LReLU	LReLU	LReLU	ReLU	-

*Channel: Number of channels produced by each layer; Stride: Stride of the convolution; Kernel: Size of the convolving kernel; AF: Activation function; Conv: Convolutional layer; Linear: Fully connected layer; LReLU: LeakyReLU.

As shown in Fig. 3, to avoid excessive downsampling of the image, which may lead to a loss of the high-frequency information of the image, we adopt a relatively simple design for the discriminator network convolution structure and do not incorporate the pooling operations and BN. Although the GAN can generate clearer and more realistic samples than those generated by other models, it exhibits limitations such as unstable training and the occurrence of the vanishing gradient and mode collapse. To address these problems, we introduce the Wasserstein distance in the WGAN into the discriminator network such that the last layer of the network does not use any activation function, to enhance the discrimination ability of the network and model stability. In addition, the input of the discriminator network is an SR image, and the image is trained by the network to return the probability of the image being a real image.

B. LOSS FUNCTION

In the designed GAN, we adopt the pixelwise loss and perceptual loss in the generator network to jointly optimize the network and introduce the Wasserstein loss in the discriminator network to optimize the network. Therefore, the generator network can generate a high-quality SR image, and the discriminator network can distinguish the true sample from the false sample more accurately.

1) PIXELWISE LOSS

In the generator network, the input of the network is an LR image rather than random noise. To ensure that the output of the generator network fits the distribution of the ground truth, the pixelwise loss is used to detect the deviation between the predicted value and ground truth of the model. The calculation is performed using equation (2):

$$Loss_{MSE} = \frac{1}{N}\sum_{i=1}^{N}\left(\| G_1\left(I_i^{LR}\right) - I_i^{HR} \|^2 + \| G_2\left(G_1\left(I_i^{LR}\right)\right) - I_i^{HR} \|^2\right) \quad (2)$$

Here, $Loss_{MSE}$ represents the mean square error loss function; I^{LR} and I^{HR} represent the pixel value of the input LR image and HR image, respectively; G_1 and G_2 represent the upsample and refinement subnetworks, respectively.

2) PERCEPTUAL LOSS

The pixelwise loss operation may result in the loss of the high-frequency information in the image, resulting in insufficient image details and unclear outlines. To address these problems, we design and add the perceptual loss function to the generator network to guide the network training in combination with the pixelwise loss. In this work, we use a convolutional neural network to extract the features of the output image and ground truth and calculate the sum of the Euclidean distance point by point on the feature map of the output image and ground truth. The perceptual loss function can be expressed as in equation (3):

$$Loss_p = \frac{1}{N}\sum_{i=1}^{N}\left(\varphi\left(G_2\left(G_1\left(I_i^{LR}\right)\right)\right) - \varphi\left(I_i^{HR}\right)\right)^2 \quad (3)$$

Here, $Loss_p$ represents the perceptual loss function, and φ represents the neural network. We input the output of the refinement subnetwork and real HR image to φ, extract the respective image features and calculate the sum of the Euclidean distances. Therefore, the network can obtain more high-frequency information during the training process and guide the generator network to generate clearer SR images. To calculate the perceptual loss, as the neural network of φ, we choose the ResNeXt-50-32 $\times$ 4d network, which is pretrained on ImageNet. Compared with VGG19, this network involves fewer network parameters and a smaller amount of calculation, exhibits a larger depth, and achieves a higher accuracy.

In summary, the loss of the generator network consists of two parts, that is, the pixelwise and perceptual losses. The generator network loss function $Loss_G$ can be defined as in equation (4):

$$Loss_G = Loss_{MSE} + Loss_p \quad (4)$$

3) DISCRIMINATOR LOSS

The GAN involves problems such as difficult convergence and easy collapse of the model. Consequently, we introduce the Wasserstein distance in the WGAN [8], in which the loss function does not consider the logarithm. Finally, we design the corresponding loss function according to the network structure. The discriminator loss function is as shown in equation (5):

$$Loss_D = \frac{1}{N}\sum_{i=1}^{N}\left[D\left(G_2\left(G_1\left(I_i^{LR}\right)\right)\right) - D\left(I_i^{HR}\right)\right] \quad (5)$$

Here, $Loss_D$ represents the discriminator loss function, and D represents the discriminator network. We input the output of the generator network and real HR image to the network, calculate the distance related difference between the inputs,

and adopt the average value to find the minimum cost of the generated sample and real sample.

C. OBJECTIVE FUNCTION

According to the designed loss functions, we modify equation (1) to obtain the final objective function of the GAN, as shown in equation (6):

$$\begin{aligned}\min_G \max_D V(D,G) = \frac{1}{N}\sum_{i=1}^{N}\Big[&\| G_1\left(I_i^{LR}\right) - I_i^{HR} \|^2 \\ &+ \| G_2\left(G_1\left(I_i^{LR}\right)\right) - I_i^{HR} \|^2 \\ &+ \left(\varphi\left(G_2\left(G_1\left(I_i^{LR}\right)\right)\right) - \varphi\left(I_i^{HR}\right)\right)^2 \\ &+ \left[D\left(G_2\left(G_1\left(I_i^{LR}\right)\right)\right) - D\left(I_i^{HR}\right)\right]\Big]\end{aligned} \tag{6}$$

IV. EXPERIMENTS

A. DATASETS

We perform experiments on four widely used benchmark datasets, DIV2K, Set5, Set14 and BSD100 (the test set of BSD300). The DIV2K dataset consists of 800 training images and 100 validation images; the BSD100 dataset consists of 100 images; and the Set5 and Set14 datasets consist of 5 and 14 images, respectively.

In this work, to avoid the overfitting of the model, the training data are preprocessed to increase the diversity of the data and enhance the overall generalization ability of the model. The data preprocessing involves the following data augmentation methods that are implemented randomly: random scaling, random rotation, and random flip.

B. TRAINING DETAILS AND PARAMETERS

The training and verification of the experiment are based on the Kaggle platform. The GPU model is NVIDIA TESLA P100, the memory is 16 GB, and the programming framework is PyTorch, based on Python.

The model is trained from scratch, and the iteration ratio of the discriminator D to generator G is 5:1. During training, we set $\alpha = 0.2$ for the activation function, LeakyReLU [28], with the batch_size $= 8$. In the generator network, we use the Adam optimizer with the momentum terms $\beta_1 = 0.5$ and $\beta_2 = 0.9$, and set the initial learning rate as 3e-4. In the discriminator network, we use the RMSprop optimization method and set the initial learning rate as 5e-5.

In the generator network, the weights in each layer are initialized using a zero mean Gaussian distribution with a standard deviation of 0.02, and the biases are initialized as 0. In the discriminator network, the weights in the fully connected layers are initialized using a zero mean Gaussian distribution with a standard deviation of 0.1, and the biases are initialized as 0.

To train the generator and discriminator networks, we preprocess the training image. First, we crop the HR image to the corresponding size. Second, we use the bicubic method with a factor of 4 to downsample the HR image to generate the corresponding LR image. Finally, we normalize the image data to $[-1, 1]$.

C. EVALUATION INDEX

In this work, we calculate the peak signal to noise ratio (PSNR) [30] and structural similarity (SSIM) [31] of the reconstructed image as the evaluation indexes.

The PSNR is an engineering term for the ratio between the maximum possible power of a signal and the power of the corrupting noise that affects the fidelity of the signal representation. This term is used to measure the pixel difference between the processed image and corresponding image and is the most widely used image quality evaluation index in the field of SR. The PSNR considers the MSE of the two images to calculate the similarity, and its unit is decibel (dB). A higher PSNR corresponds to a smaller distortion of the SR image and a more desirable effect. The PSNR and MSE can be defined as in equations (7) and (8), respectively:

$$PSNR = 10\log_{10}\left(\frac{(2^n-1)^2}{MSE}\right) \tag{7}$$

$$MSE = \frac{1}{H\times W}\sum_{i=1}^{H}\sum_{j=1}^{W}[X(i,j) - Y(i,j)]^2 \tag{8}$$

Here, n is the number of bits of the image pixel; H and W are the height and width of the image, respectively; i and j represent the position of the pixel; X and Y denote the original HR image and reconstructed SR image, respectively, and have equal sizes.

The SSIM is used to measure the similarity among image structures. The SSIM index defines the structural information considering that the image composition is independent of the brightness and contrast, thereby reflecting the properties of the object structure in the scene, and it models the distortion as a combination of three different factors: brightness, contrast, and structure. Moreover, the mean and standard deviation are considered as an estimate of the brightness and contrast, respectively, and the covariance is considered as a measure of the structural similarity. The index is usually used to evaluate the quality of image denoising results. The range of the evaluation value is [0, 1]. A value closer to 1 indicates that the structures of two images are more similar. The SSIM is defined as in formula (9):

$$SSIM(x,y) = \frac{\left(2\mu_x\mu_y + C_1\right)\left(2\sigma_{xy} + C_2\right)}{\left(\mu_x^2 + \mu_y^2 + C_1\right)\left(\sigma_x^2 + \sigma_y^2 + C_2\right)} \tag{9}$$

Here, x and y denote the original HR image and the reconstructed SR image respectively; μ_x and μ_y are the average values of x and y, respectively, expressed as the estimated value of brightness; σ_x and σ_y are the standard deviations of x and y, respectively; σ_{xy} is the covariance of x and y; C_1 and C_2 are constants, used to ensure the stability of the SSIM.

D. DETAILS OF THE PROPOSED ALGORITHM

Algorithm 1 Proposed Image Super-Resolution Reconstruction Algorithm

Input: LR image (the shape is [n, c, h, w], where n is the batch size, c is the number of channels, and h and w are the height and width of the feature map, respectively), HR image.
Output: SR image.
Initialization: Generator network G, discriminator network D, ResNeXt-50-32 × 4d network φ. I^{LR} and I^{HR} represent the pixel value of the input LR image and HR image, respectively. The discriminator involves k iterations (5, in this work).
For number of training epochs **do**
  For k iterations **do**
    Step 1: Input I^{LR}, train the upsample subnetwork, and obtain the output $G_1(I^{LR})$.
    Step 2: Input $G_1(I^{LR})$, train the refinement subnetwork, and obtain the output $G_2(G_1(I^{LR}))$.
    Step 3: Input $G_2(G_1(I^{LR}))$ and I^{HR}, train D, and obtain the output $D(G_2(G_1(I^{LR})))$ and $D(I^{HR})$.
    Step 4: Calculate the $Loss_D$ according to equation (5).
    Step 5: Update the weights of D.
  end for
  Step 6: Input $G_2(G_1(I^{LR}))$ and I^{HR}, call φ, and obtain the output $\varphi(G_2(G_1(I^{LR})))$ and $\varphi(I^{HR})$.
  Step 7: Calculate $Loss_{MSE}$ according to equation (2).
  Step 8: Calculate $Loss_p$ according to equation (3).
  Step 9: Calculate $Loss_G$ according to equation (4).
  Step 10: Update weights of G (i.e., the upsample and refinement subnetworks).
end for

E. RESULTS AND ANALYSIS

During training, we first crop the DIV2K dataset to a fixed size through random cropping and perform the corresponding data preprocessing as the HR real result. Subsequently, we downsample the image (downsample factor of 4) into the corresponding LR image through the bicubic method. Finally, we input the LR image into GAN to start training. The model generates the corresponding SR image, and we compare the SR image with the HR image.

During testing, we use three datasets: Set5, Set14, and BSD100. We input the three datasets to the trained model, in turn, to generate the corresponding SR image and calculate the average PSNR and SSIM values of the obtained images. We compare the experimental results with those of other classical SR models.

Ablation Study (Dual Network Structure): In this work, we employed a dual network structure to jointly solve the problems of super-resolution and optimization. To demonstrate the effectiveness of the dual network structure, experiments were conducted using the structure. In the ablation study, we conducted experiments to examine the effect of removing the refinement subnetwork. As shown in Table 3, when the refinement subnetwork was removed, the overall performance of the proposed model degraded. According to the experimental results, on the Set5, Set14 and BSD100 datasets, the average PSNR index of the dual network structure is larger than that of the model without the refinement subnetwork by 0.79, 0.78, and 0.25 dB, respectively, and the average SSIM index is larger by 0.0081, 0.0068, and 0.0013, respectively. In this manner, the ablation study demonstrates the effectiveness of the dual network structure and indicates that the added refinement subnetwork can enhance the model performance and optimize the quality of the generated image.

TABLE 3. Ablation study (dual network structure).

Datasets	Index	No Refinement Subnetwork	Dual Network Structure
Set5	PSNR	30.78	31.57
	SSIM	0.8784	0.8865
Set14	PSNR	27.58	28.36
	SSIM	0.7659	0.7727
BSD100	PSNR	27.24	27.49
	SSIM	0.7457	0.7470

*No Refinement Subnetwork: We remove the refinement subnetwork in the proposed model and use only the upsampling subnetwork for training. Dual Network Structure: We use both the upsample and refinement subnetworks for the training.

Table 4 presents the comparison of the average PSNR and SSIM values of different SR reconstruction algorithms, and the boldfaced data corresponds to the optimal result. The proposed model achieves the optimal experimental results when the image magnification is 4. As shown in Table 4, on the Set5, Set14 and BSD100 datasets, the PSNR value of the proposed model is higher than that of the bicubic method by 3.15, 2.36, and 1.53 dB, respectively, SRCNN model by 1.09, 0.86, and 0.59 dB, respectively, and LapSRN model by 0.03, 0.17, and 0.17 dB, respectively. The corresponding SSIM value is higher than that of the bicubic method

TABLE 4. Results of SSIM and PSNR (Unit: dB) (×4).

Datasets	Index	SR Algorithm						
		Bicubic	SRCNN [16]	VDSR [20]	DRCN [21]	LapSRN [23]	SRGAN [26]	Proposed
Set5	PSNR	28.42	30.48	31.35	31.53	31.54	29.40	**31.57**
	SSIM	0.8104	0.8628	0.8838	0.8854	0.8850	0.8472	**0.8865**
Set14	PSNR	26.00	27.50	28.01	28.02	28.19	26.02	**28.36**
	SSIM	0.7027	0.7513	0.7674	0.7670	0.7720	0.7397	**0.7727**
BSD100	PSNR	25.96	26.90	27.29	27.23	27.32	25.16	**27.49**
	SSIM	0.6675	0.7101	0.7251	0.7233	0.7280	0.6688	**0.7470**

FIGURE 4. Comparison of the magnified reconstruction results of different SR methods: Super-resolution results of 'ppt3' (Set14) and 'baboon' (Set14) with a scale factor of ×4. Top: The texts obtained using the proposed method are sharp; the character edges corresponding to the other methods are blurry. Bottom: The proposed method recovers sharp results, while the other models yield blurry results.

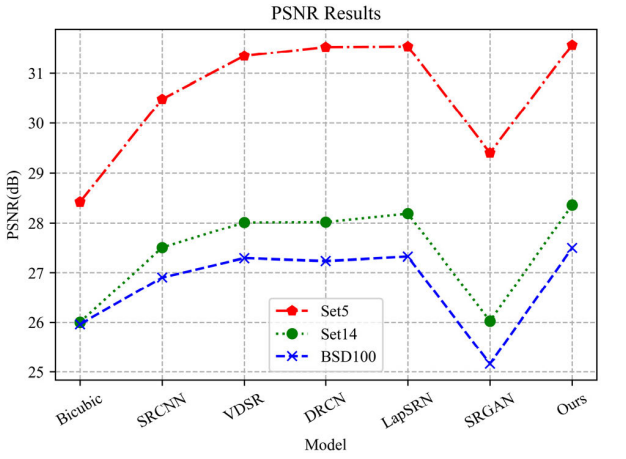

FIGURE 5. Comparison of the PSNR (dB) values. The abscissa and ordinate indicate the different SR models and PSNR evaluation results for the different models, respectively.

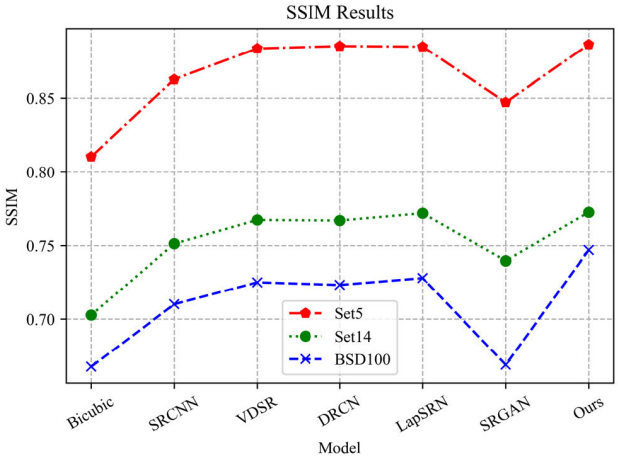

FIGURE 6. Comparison of the SSIM values. The abscissa and ordinate indicate the different SR models and SSIM evaluation results for the different models, respectively.

by 0.0761, 0.07, and 0.0795, respectively, SRCNN model by 0.0237, 0.0214, and 0.0369, respectively, and LapSRN model by 0.0015, 0.0007, and 0.019, respectively. Moreover, the PSNR and SSIM values for the proposed model are more suitable than those of the other models. The results show that the proposed dual network structure can extract more image features than other models, and the perceptual loss function promotes the network to recover the detailed information, thereby generating clear HR images.

FIGURE 7. Partial super-resolution reconstruction results of the Set5 and Set14 datasets. From left to right, the images correspond to those obtained using the bicubic model, SRCNN model, and proposed model, and the original HR image. The proposed method yields sharp results with higher quality and richer texture details, while the images obtained using the other models are blurry.

As shown in Fig. 4, different SR methods are used to reconstruct part of the images in the Set14 dataset. We perform the same operation on the image reconstructed by different SR methods; specifically, we crop the same area from the reconstructed image and enlarge the cropped part for comparison. As shown in Fig. 4, the SR image reconstructed by the bicubic method is extremely blurry after being enlarged, with critical image distortion, and a large amount of the detailed information is missing. The SR image reconstructed using the SRCNN method is better than that of the bicubic method, although several shortcomings remain. In comparison, the SR image reconstructed by the proposed model is the most similar to the original HR image. The image is relatively clear and has excellent visual effects; moreover, the reconstruction results demonstrate more realistic texture details. These aspects demonstrate the feasibility of the proposed model.

Both the proposed model and SRGAN model use a generative adversarial network structure. However, in contrast to the SRGAN, the proposed model does not employ batch normalization [32]. In particular, when BN is implemented deep in the network and trained under the GAN framework, the image may produce artifacts, thereby limiting the generalization ability of the model. Removing the BN can not only improve the generalization ability of the model but also alleviate the computational complexity and memory usage. In addition, the proposed model uses the Wasserstein distance in the WGAN to optimize the objective function. In particular, the last layer of the discriminator network does not use any activation function, and the loss function does not consider the logarithm. As shown in Table 4, the average PSNR and SSIM values of the proposed model are higher than those of the SRGAN model by 2.17, 2.34, and 2.33 dB and 0.0393, 0.033, and 0.0782, respectively. These values demonstrate the effectiveness of the implementation of the Wasserstein distance, which can effectively improve the discrimination ability of the network, enhance the stability of the model, and help generate high-quality images.

Figs. 5 and 6 show the comparison of the average PSNR and SSIM values of different models on the Set5, Set14, and BSD100 datasets. As shown in Figs. 5 and 6, the proposed model outperforms the other models in terms of the PSNR and SSIM values, along with the experimental results. These findings demonstrate the effectiveness of the proposed model.

Fig. 7 shows the SR reconstruction results of certain images in the Set5 and Set14 datasets. From left to right, the images correspond to those obtained using the bicubic model, SRCNN model, and proposed model, and the original HR image. As shown in Figs. 4 and 7, in both the cases of comparing the reconstructed image directly or cropping a part of the area for partial enlargement and comparison, the reconstructed image of the proposed model is clearer than that obtained using the other methods. Moreover, the reconstructed image based on the bicubic method is blurry both globally and locally, considerable noise is present in the image, and the distortion is critical. The SRCNN method yields a reconstructed image with a higher quality. The global image is clearer, although the image is slightly blurry and excessively smooth compared to the original image. Moreover, the image appears blurry after partial enlargement and lacks the high-frequency detail information. Compared with the images reconstructed using the bicubic and SRCNN methods, the image reconstructed using the proposed model exhibits a higher quality, richer texture details, better color matching with the original image, and better visual effects. These findings demonstrate the effectiveness of the proposed model, which can recover more high-frequency details and generate clear SR images.

V. CONCLUSION

In this paper, to address the problems of insufficient feature extraction and blurred generated images of SR reconstruction, we propose an image SR reconstruction model based on the GAN. In the generator network, a dual network structure is adopted to solves the problems of simultaneous super-resolution and optimization, thereby enabling the network to fully extract the image features and recover more missing details, enhance the learning ability of the network, and optimize the quality of the generated image. The design of the perceptual loss guides the network to generate the SR image with more detailed information and clear outlines. In the discriminator network, we use the Wasserstein distance to optimize the network, thereby enhancing the stability and discrimination ability of the network.

The results of extensive experiments conducted on the DIV2K, Set5, Set14, and BSD100 datasets demonstrate that when the reconstructed image has large upscaling factors ($4\times$), the proposed model can recover more high-frequency details of the image, thereby making the reconstructed image clearer and more realistic. Moreover, the proposed model outperforms other state of the art methods in terms of the PSNR and SSIM. Overall, the experimental results indicate that the proposed approach is accurate and effective.

Currently, our model is only applicable to the SISR problem. In future work, we aim to extend the model to the MISR problem to ensure that the model can achieve excellent results in multi-image and video reconstruction as well. In addition, semantic segmentation technology based on deep learning has been fully developed. In future research, the semantic segmentation technology can be combined with the image super-resolution reconstruction technology to further enhance the model performance.

REFERENCES

[1] Y. Chen, J. Wang, S. Liu, X. Chen, J. Xiong, J. Xie, and K. Yang, "Multi-scale fast correlation filtering tracking algorithm based on a feature fusion model," *Concurrency Comput., Pract. Exper.*, Oct. 2019, Art. no. e5533, doi: 10.1002/cpe.5533.

[2] Y. Chen, J. Tao, Q. Zhang, K. Yang, X. Chen, J. Xiong, R. Xia, and J. Xie, "Saliency detection via the improved hierarchical principal component analysis method," *Wireless Commun. Mobile Comput.*, vol. 2020, May 2020, Art. no. 8822777, doi: 10.1155/2020/8822777.

[3] Y. Chen, J. Tao, L. Liu, J. Xiong, R. Xia, J. Xie, Q. Zhang, and K. Yang, "Research of improving semantic image segmentation based on a feature fusion model," *J. Ambient Intell. Humanized Comput.*, May 2020, doi: 10.1007/s12652-020-02066-z.

[4] Y. Chen, L. Liu, J. Tao, X. Chen, R. Xia, Q. Zhang, J. Xiong, K. Yang, and J. Xie, "The image annotation algorithm using convolutional features from intermediate layer of deep learning," *Multimedia Tools Appl.*, pp. 1–25, Sep. 2020, doi: 10.1007/s11042-020-09887-2.
[5] Y. Chen, L. Liu, J. Tao, R. Xia, Q. Zhang, K. Yang, J. Xiong, and X. Chen, "The improved image inpainting algorithm via encoder and similarity constraint," *Vis. Comput.*, pp. 1–5, Jul. 2020, doi: 10.1007/s00371-020-01932-3.
[6] D. Lu and Q. Weng, "A survey of image classification methods and techniques for improving classification performance," *Int. J. Remote Sens.*, vol. 28, no. 5, pp. 823–870, Mar. 2007.
[7] A. Shrestha and A. Mahmood, "Review of deep learning algorithms and architectures," *IEEE Access*, vol. 7, pp. 53040–53065, 2019.
[8] M. Arjovsky, S. Chintala, and L. Bottou, "Wasserstein GAN," 2017, *arXiv:1701.07875*. [Online]. Available: http://arxiv.org/abs/1701.07875
[9] X. Sun, L. Xiao-Guang, L. Jia-Feng, and Z. Li, "Review on deep learning based image super-resolution restoration algorithms," *Acta Automatica Sinica*, vol. 43, no. 5, pp. 697–709, 2017.
[10] R. R. Schultz and R. L. Stevenson, "A Bayesian approach to image expansion for improved definition," *IEEE Trans. Image Process.*, vol. 3, no. 3, pp. 233–242, May 1994.
[11] H. Hou and H. Andrews, "Cubic splines for image interpolation and digital filtering," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-26, no. 6, pp. 508–517, Dec. 1978.
[12] X. Li and M. T. Orchard, "New edge-directed interpolation," *IEEE Trans. Image Process.*, vol. 10, no. 10, pp. 1521–1527, Oct. 2001.
[13] K. Zhang, W. Zuo, S. Gu, and L. Zhang, "Learning deep CNN denoiser prior for image restoration," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 2808–2817.
[14] S. Singh, M. K. Kalra, J. Hsieh, P. E. Licato, S. Do, H. H. Pien, and M. A. Blake, "Abdominal CT: Comparison of adaptive statistical iterative and filtered back projection reconstruction techniques," *Radiology*, vol. 257, no. 2, pp. 373–383, Nov. 2010.
[15] M. Jiang and Z. Zhang, "Review on POCS algorithms for image reconstruction," *Computerized Tomogr. Theory Appl.*, vol. 12, no. 1, pp. 51–55, 2003.
[16] C. Dong, C. C. Loy, K. He, and X. Tang, "Learning a deep convolutional network for image super-resolution," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2014, pp. 184–199.
[17] C. Dong, C. C. Loy, and X. Tang, "Accelerating the super-resolution convolutional neural network," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2016, pp. 391–407.
[18] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*. [Online]. Available: http://arxiv.org/abs/1409.1556
[19] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.
[20] J. Kim, J. K. Lee, and K. M. Lee, "Accurate image super-resolution using very deep convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 1646–1654.
[21] J. Kim, J. K. Lee, and K. M. Lee, "Deeply-recursive convolutional network for image super-resolution," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 1637–1645.
[22] Y. Tai, J. Yang, and X. Liu, "Image super-resolution via deep recursive residual network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 2790–2798.
[23] W.-S. Lai, J.-B. Huang, N. Ahuja, and M.-H. Yang, "Deep Laplacian pyramid networks for fast and accurate super-resolution," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 5835–5843.
[24] B. Lim, S. Son, H. Kim, S. Nah, and K. M. Lee, "Enhanced deep residual networks for single image super-resolution," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jul. 2017, pp. 1132–1140.
[25] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Proc. Adv. Neural Inf. Process. Syst.*, 2014, pp. 2672–2680.
[26] C. Ledig, L. Theis, F. Huszar, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, and W. Shi, "Photo-realistic single image super-resolution using a generative adversarial network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 105–114.
[27] J. Johnson, A. Alahi, and L. Fei-Fei, "Perceptual losses for real-time style transfer and super-resolution," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2016, pp. 694–711.
[28] A. Radford, L. Metz, and S. Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks," 2015, *arXiv:1511.06434*. [Online]. Available: http://arxiv.org/abs/1511.06434
[29] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," 2015, *arXiv:1502.03167*. [Online]. Available: http://arxiv.org/abs/1502.03167
[30] A. Hore and D. Ziou, "Image quality metrics: PSNR vs. SSIM," in *Proc. 20th Int. Conf. Pattern Recognit.*, Aug. 2010, pp. 2366–2369.
[31] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.
[32] X. Wang, K. Yu, S. Wu, J. Gu, Y. Liu, C. Dong, Y. Qiao, and C. C. Loy, "ESRGAN: Enhanced super-resolution generative adversarial networks," in *Proc. Eur. Conf. Comput. Vis. Workshops (ECCVW)*, 2018, pp. 63–79.

YUN WU was born in Jingdezhen, China, in 1973. He received the B.E. degree in computer science from the Nanjing University of Aeronautics and Astronautics (NUAA), Nanjing, China, in 1999, and the M.E. and Ph.D. degrees in computer science from Guizhou University, Guiyang, China, in 2006 and 2009, respectively.

He is currently an Associate Professor with Guizhou University, the Director of the Institute of Computer Simulation, Guizhou University, an Evaluation Expert of Guizhou University, and the Academic Leader of the Cloud Computing and Information Research Team. He presided over and participated in a number of national, provincial, and ministerial scientific research projects, published more than 20 academic articles, obtained three invention patents, and ten software copyrights. He has served as an Expert for the China Institute of Communications-Cloud Computing and Big Data Application Committee, a member of the Expert Group of the Guizhou Big Data Academy, a member of the China Computer Federation, and a member of the Chinese Information Processing Society of China. His research interests include artificial intelligence, big data technology, computer vision, data mining, and deep learning.

LIN LAN received the B.E. degree in Internet of Things engineering from the School of Information and Communication Engineering, Dalian Minzu University, Dalian, in 2018. He is currently pursuing the M.E. degree in computer technology with the School of Computer Science and Technology, Guizhou University, Guizhou, China. His research interests include deep learning, computer vision, image processing, and image super resolution.

HUIYUN LONG received the Ph.D. degree in computer science from Guizhou University, Guiyang, China, in 2009. She is currently an Associate Professor and the Associate Dean of the School of Computer Science and Technology, Guizhou University. Her main research interests include process algebra, formalization of Web services, and deep learning.

GUANGQIAN KONG received the B.S. degree from Central South University, Hunan, China, in 1996, and the M.S. and Ph.D. degrees from Guizhou University, Guizhou, China, in 2002 and 2009, respectively. He is currently an Associate Professor, a Graduate Supervisor, and a member of the China Computer Federation. His research interests include computer networks, big data, deep learning, and their applications.

XUN DUAN received the B.S. degree from the Chongqing University of Posts and Telecommunications, in 1995, the Ph.D. degree in computer science from Guizhou University, in 2007. He has been an Associate Professor with the School of Computer Science and Technology, Guizhou University, since 2007. His current interests include big data, deep learning, and computer vision.

CHANGZHUAN XU received the B.S. degree in mathematics from the School of Information and Computer Science, Xiamen University of Technology, Xiamen, China, in 2018. She is currently pursuing the M.E. degree in computer technology with the School of Computer Science and Technology, Guizhou University, Guizhou, China. Her research interests include deep learning, computer vision, image processing, and image semantic segmentation.