

Received January 17, 2020, accepted February 4, 2020, date of publication February 18, 2020, date of current version February 27, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2974759

Image Super-Resolution Reconstruction Using Generative Adversarial Networks Based on Wide-Channel Activation

XUDONG SUN¹, ZHENXI ZHAO^{1,2,3,4}, SONG ZHANG^{1,2,3,4}, JINTAO LIU^{1,2,3,4,5},
XINTING YANG^{1,2,3,4}, AND CHAO ZHOU^{1,2,3,4}

¹School of Mechatronics Engineering, East China Jiaotong University, Nanchang 330013, China

²National Engineering Research Center for Information Technology in Agriculture, Beijing 100097, China

³Beijing Research Center for Information Technology in Agriculture, Beijing 100097, China

⁴National Engineering Laboratory for Agri-Product Quality Traceability, Beijing 100097, China

⁵School of Engineering, University of Almería, 04120 Almería, Spain

Corresponding author: Chao Zhou (supperchao@hotmail.com)


This work was supported in part by the National Key Technology Research and Development Program of China under Grant 2019YFD0901004, in part by the National Natural Science Foundation of China under Grant 31960497, in part by the Youth Research Fund of Beijing Academy of Agricultural and Forestry Sciences under Grant QNJJ202014, and in part by the Beijing Natural Science Foundation under Grant 4184089.

ABSTRACT In recent years, residual learning has shown excellent performance on convolutional neural network (CNN)-based single-image super-resolution (SISR) tasks. However, CNN-based SISR approaches have focused mainly on the design of deep architectures, and the rectified linear units (ReLUs) used in these networks hinder shallow-to-deep information transfer. As a result, these methods are unable to utilize some shallow information, and improving model performance is difficult. To solve the above issues, this paper proposes an image SR reconstruction method based on a generative adversarial network with a residual dense architecture. First, before ReLU activation, the number of feature channels is expanded by a factor of 6~9 using a 1×1 convolutional layer, which improves the utilization of shallow information. Next, the original discriminator is replaced with a relativistic average discriminator, thereby improving the authenticity of the discriminative network. Finally, preactivation features are used to improve the perceptual loss, thus providing stronger monitoring for brightness consistency and texture restoration. Experimental results show that the proposed algorithm improves the utilization of shallow information in a deep network. Structural similarity (SSIM) index evaluations show that the overall utilization of shallow information is increased by 105.52%. In addition, the average runtime is 0.42 sec/frame, nearly 3.6 times faster than those of traditional methods. Moreover, the recovered images have an average natural image quality evaluator value of 3.4 and high perceptual quality, showing that the proposed method is suitable for image reconstruction applications in fields such as agriculture and medicine.

INDEX TERMS Super-resolution, residual block, relativistic average discriminator, generative adversarial network, perceived quality.

I. INTRODUCTION

Image super-resolution (SR), in which algorithms are used to reconstruct an image from low resolution (LR) to high resolution (HR), is an important class of image processing techniques. In addition to improving image perceptual quality, image SR techniques are in high demand for applications such as agricultural imaging.

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

SR improvements can help enhance the performance of various computer vision tasks [1]–[5]. As an essential artificial intelligence tool, deep learning has gradually been introduced into the field of image SR reconstruction [6]–[9]. From the early convolutional neural network (CNN)-based approaches (such as SRCNN [10]) to the more recent generative adversarial network (GAN)-based approaches (such as SRGAN [11]), continuous advancements have been achieved in research on image SR reconstruction [10]–[14].

At present, deep-learning-based SR models such as SRCNN [10], EDSR [15], VDSR [13], and RCAN [16] have achieved significant improvements in terms of the peak signal-to-noise ratio (PSNR) and structural similarity (SSIM) index. However, these models typically use the minimum absolute deviation (l_1) or the minimum squared error (MSE or l_2) as the loss function; as a result, the generated images are too smooth, and they easily lose details and texture information [11]. As an alternative to the above models, a GAN consists of two different subnetworks, namely, a generative network and a discriminative network, rather than simply a single network, and GANs adopt an adversarial training method [17], [18]. Compared with other generative models, such as Boltzmann machines and generative stochastic networks (GSNs), GANs do not need complex Markov chains and can produce more transparent and realistic samples.

The purpose of using GANs is to restore image texture information and improve the overall quality of images. The SRGAN model added an adversarial loss function to the original l_2 loss function [11]. Compared with the outputs of the SRCNN, FSRCNN [19], DRCN [17] and DRRN [20], the images reconstructed with SRGAN are more realistic. However, the PSNR and SSIM scores of SRCNN tend to be relatively low, and it can easily produce visual artifacts. In EnhanceNet [21], perceptual and texture matching losses were introduced into the traditional MSE loss function, and the perceptual quality of SISR was improved via joint countermeasure training; however, this model produces high-frequency noise, and the reconstructed SR images lack high-frequency details. Under the original SR framework, single-image SR with feature discrimination (SRFeat) was proposed based on two kinds of discriminators: one acting in the image domain and one acting in the feature domain. The generator includes long-range jump connections to improve the mobility of remote information [22]. However, the batch normalization (BN) layer uses only a small batch of data rather than the entire training set to calculate the mean and variance; this is equivalent to introducing noise during the gradient calculations and is not suitable for noise-sensitive GANs. Based on the SRGAN model, the enhanced SR GAN (ESRGAN) model was introduced by using nested residual-in-residual dense blocks (RRDBs) to improve the network structure. The BN layer was removed, and the deep network was trained using residual scaling and small initial network parameters. A deep network improves image quality; however, as the network depth increases, shallow information is less fully utilized, ultimately resulting in the lack of further image enhancement. Moreover, ESRGAN focuses primarily on image enhancement while ignoring image authenticity; thus, it can easily generate erroneous texture information [23].

While deep network structures can improve image quality, as the number of layers increases, it becomes increasingly difficult for the model to converge, leading to an unstable model effect, inconsistent brightness of the reconstructed images, and insufficient use of shallow information.

Although BN layers can accelerate the network convergence speed and enable a high learning rate to be used, the introduction of noise often causes the reconstructed images to contain artifacts.

To address the above issues, this paper proposes an image SR reconstruction algorithm based on variant residual dense blocks (VRDBs) for GANs. This approach resolves the underutilization of shallow information by enhancing the channel information characteristics. Better brightness monitoring and the improved stability of the model structure help not only to produce higher-quality images but also to avoid unrealistic textures.

II. RELATED WORK

A. GENERATIVE ADVERSARIAL NETWORKS

Although CNN-based SISR models have achieved breakthroughs in both accuracy and speed, the images they reconstruct lack fine textural details for images with large sampling factors [24]–[26]. These SISR models are driven by the objective loss function. Current mainstream algorithms, such as SRCNN, FSRCNN [19], DRCN, and DRRN, focus primarily on the MSE loss. Thus, while the resulting recovered images have a high PSNR, they usually lack high-frequency details, and their visual effect is usually unsatisfactory. In contrast, a GAN model consists of a generator and a discriminator and adopts a pairwise adversarial approach. The generator, which is responsible for creating the SR images, attempts to trick the discriminator into failing to distinguish between a real HR image and an artificial reconstructed SR image. Using this approach leads to the generation of SR images with better perceptual qualities; however, the PSNR value corresponding to the reconstructed SR image is usually lower than that of the original HR image, indicating that the popular PSNR evaluation metric cannot robustly evaluate the perceptual differences between SR and HR images [11], [23]. To solve the above problem, the SRGAN model was proposed, which changes the loss function by replacing the traditional MSE loss with a perceptual loss and a content loss while simultaneously introducing a GAN to convert the content loss in the traditional pixel space into a similarity of adversarial properties; finally, SRGAN introduces a deep residual network to extract detailed textures from images [11]. In addition to a pixel-level MSE loss, EnhanceNet uses two other loss terms: first, a perceptual loss function is defined as l_1 , which is on the intermediate feature representation of the pretraining network [27]; second, a texture-matching loss function is used to match LR images with HR images. A texture-matching loss function is also quantized into l_1 between the Gram matrix calculated from the depth feature. The entire network architecture is trained in the reverse direction, where the goal of the generative network is to gain the ability to deceive the discriminator. Based on the SRGAN model, the ESRGAN model [23] offers three improvements: (1) the basic network unit is changed from a residual unit to an RRDB; (2) the GAN is replaced with an improved version, namely, a relativistic average

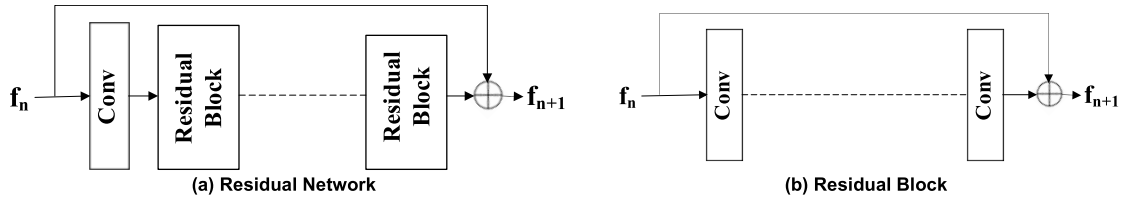


FIGURE 1. Residual network structure: (a) diagram of the overall residual network; (b) a local residual block.

GAN (RaGAN); and (3) experiments have demonstrated that using the VGG feature before activation can provide stronger brightness monitoring for the reconstructed image, and thus, the reconstructed image is darker after activation. Using VGG features before activation also helps produce sharper edges and more luxurious textures. This is because the use of dense features before activation provides stronger supervision than the use of sparse features after activation.

B. DEEP CONVOLUTIONAL NEURAL NETWORK

The design of neural network structures has become an essential part of deep learning. Using a deep network model helps to output high-quality images, but as the number of network layers increases, network convergence becomes more difficult to achieve, and exploding and vanishing gradient problems are more likely to occur. To address these challenges, researchers have modified the network connections and developed residual learning strategies; the overall residual network and the residual block structure are shown in Figure 1. Residual networks can be classified mainly into networks focusing on global and local residual learning. Global residual learning aims to restore the high-frequency information lost during the conversion from LR to HR images. Finally, the LR image and the learned residual information are combined to form the reconstructed HR image [28], [29].

Local residual learning, which is similar to residual learning in ResNet, is used to alleviate the degradation caused by increasing the depth and learning ability of the network [30], [31]. As the number of network layers increases, the features in each convolutional layer will correspond to receptive fields of different sizes. However, the Laplacian pyramid SR network (LapSRN) [14], EDSR, and MemNet [32] models fail to fully use the information from each convolutional layer. Although the proposed gating unit in the memory block controls short-term memory [23], a local convolutional layer cannot directly access subsequent layers, and thus, it is difficult to make full use of the information from all the internal layers of a network. Consequently, researchers often add complex network connections (such as dense connections) to residual blocks.

A densely connected layer uses the feature mappings from all the previous layers as its input; then, its own feature mapping is used as the input to all the subsequent layers, leading to $l \cdot (l - 1) / 2$ dense blocks (where l is the total number of layers with dense connections). Dense connections not only help

to alleviate the vanishing gradient problem, enhance signal propagation and promote feature reuse but also significantly reduce the number of parameters by enabling the use of low growth rates and the squeezing of channels after joining. Tong *et al.* [33] used a dense block structure to construct a 69-layer SRDenseNet model by inserting dense connections between different dense blocks. In addition, Zhang *et al.* [29] proposed a CNN-based residual structure with wide-channel activation (WDSR_A and WDSR_B). On the premise of maintaining the original calculation cost, the number of feature channels before activation was increased, and the feature channel was extended before the rectified linear unit (ReLU) allowed more information to pass through while simultaneously maintaining the nonlinearity of extremely deep neural networks. Consequently, low-level SR from the shallow layer can be more easily propagated to the final layer, which can produce a better image reconstruction effect. This method was promoted in [34], [35].

C. LOSS FUNCTION

Early SR models mainly used either the minimum absolute deviation l_1 (used by models such as LapSRN, EDSR, and MemNet) or the MSE (used by models such as SRCNN, VDSR, and the dense-deep backprojection network (D-DBPN) [36]) as the loss function for SR image reconstruction. The formulas are as follows:

$$l_1 = \frac{1}{n \times m} \sum_{i=1}^{i=n} \sum_{j=1}^{j=m} |Y_{i,j} - f(x_{i,j})|, \quad (1)$$

$$l_2 = \frac{1}{n \times m} \sum_{i=1}^{i=n} \sum_{j=1}^{j=m} (Y_{i,j} - f(x_{i,j}))^2, \quad (2)$$

where $Y_{i,j}$ is a pixel value in the real image, $f(x_{i,j})$ is the pixel value at the corresponding position in the reconstructed image, and n and m represent the image dimensions.

However, the use of these loss functions results in insufficient image texture information remaining after the reconstruction, and the image is too smooth and does not match human visual perception. In an adversarial neural network, both an adversarial loss and a content loss can be applied to make the image look more natural; the content loss ensures that the reconstructed image has characteristics similar to those of the original LR image [11]. Adversarial loss is similar to the loss used in traditional GAN applications: the main innovation lies in the content loss. The content loss can be defined as a Euclidean distance loss based on the differences between the reconstructed HR image and the original image.

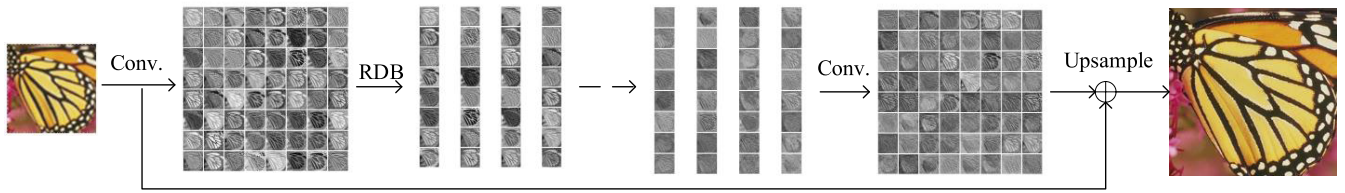


FIGURE 2. Overall network architecture of the proposed algorithm.

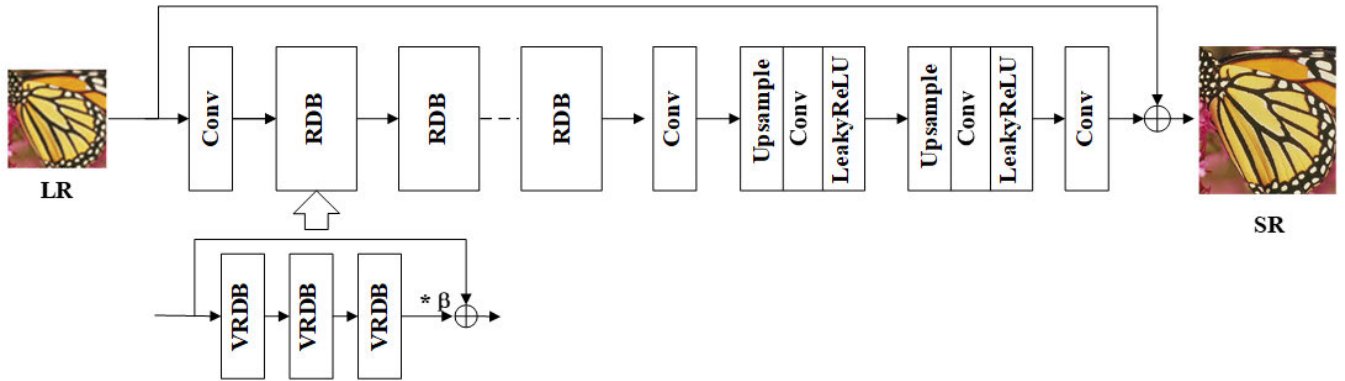


FIGURE 3. Network diagram of the algorithm proposed in this paper.

The adversarial loss formula is as follows:

$$l_{Gen}^{SR} = \sum_{n=1}^{n=N} -\log D_{\theta D}(G_{\theta G}(I^{LR})), \quad (3)$$

where $D_{\theta D}(G_{\theta G}(I^{LR}))$ denotes the probability that the reconstructed image is a real, accurate image.

The content loss function consists of l_2 and the VGG loss after activation. The calculation formula for the VGG loss function is as follows:

$$l_{VGG/i,j}^{SR} = \frac{1}{W_{i,j} \times H_{i,j}} \sum_{x=1}^{x=W_{i,j}} \sum_{y=1}^{y=H_{i,j}} (\phi_{i,j}(I^{HR})_{x,y} - \phi_{i,j}(G_{\theta G}(I^{HR}))_{x,y})^2, \quad (4)$$

where $\phi_{i,j}$ indicates that the j -th convolution is activated in VGG19 and that the feature map is obtained before the i -th max-pooling layer and $W_{i,j}$ and $H_{i,j}$ are the dimensions of the feature map.

III. PRINCIPLE OF THE PROPOSED MODEL

To avoid the insufficient use of shallow information, SR models such as EDSR, SRDenseNet, the residual deconvolutional network (RDN) [30], and D-DBPN use jump connections and other connection methods. By contrast, rather than using various jump connections, this paper considers that the use of nonlinear ReLUs hinders the flow of information from shallow to deep layers [37]. To address this problem, a 1×1 convolutional layer is inserted before activating the ReLU to expand the number of feature channels by a factor of 6-9. This approach maintains a high degree of nonlinearity in the deep neural network while improving the utilization of shallow information, endowing the neural network with

a better effect when predicting dense pixels. Based on the above considerations, this paper uses 1×1 convolution to expand the feature combination before activating the ReLU in each VRDB, allowing more information to pass through and producing both a better texture effect and an increased PSNR value. In addition, to obtain more detailed texture information, residual scaling and the leaky ReLU unsaturated excitation function are used in the model network. The BN layer is removed, and a 5×5 convolution kernel is used instead of the original 3×3 convolution kernel to increase the size of the receptive field. In the design of the loss function, this paper uses preactivation features instead of the perceptual loss function L_{percep} after activation. The discriminative power of the discriminator is improved by using the relativistic average discriminator D_{Ra} instead of the standard discriminator [23]. The overall network structure diagram of the proposed algorithm is shown in Figure 2.

A. MODEL NETWORK STRUCTURE

The proposed SR image restoration method is based primarily on the existing GAN and SRDenseNet methods. The overall network structure is illustrated in Figure 3. First, a convolutional layer is used to learn the characteristics of the lower layer; then, upsampling is performed in two steps, each time by a factor of 2, to learn the upsampling filter parameters. Finally, a convolutional layer produces an HR output image. The three main contributions of this paper are as follows:

(1). The BN layer is removed, and each residual block is multiplied by a scaling factor β to remove noise and strengthen the stability of the network model, resulting in a deep network model that is more suitable for training.

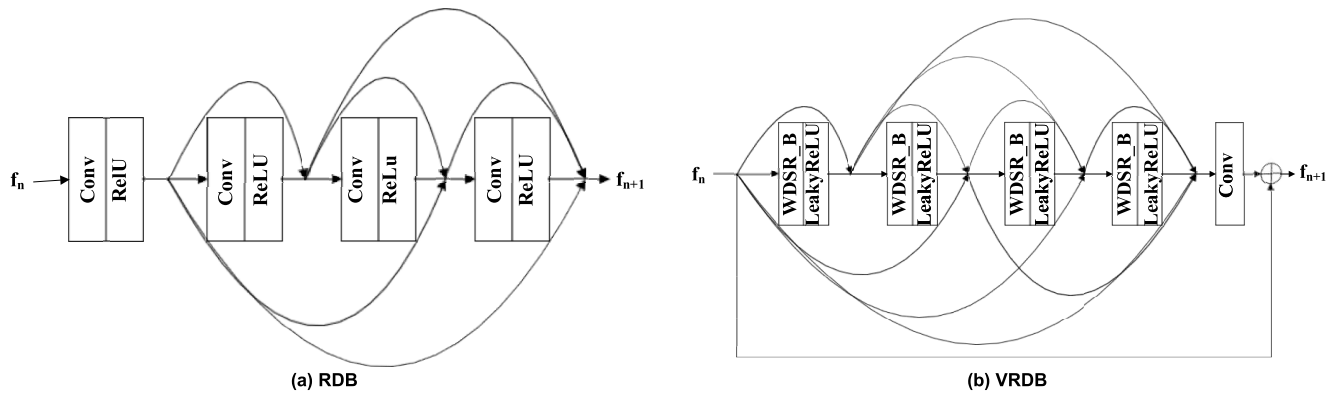


FIGURE 4. Residual network structure comparison: (a) structural diagram of an RDB; (b) VRDB diagram.

(2). Every group of three VRDBs is treated as a complete RDB. A deep network with 22 weighted layers is used, and the final number of neurons is 64. A 3×3 convolution kernel is used to extract image features. Each VRDB module is composed of 4 WDSR_B [37] and leaky ReLU layers with dense jump connections and a convolutional layer to extract local image features.

(3). The 128-layer VGG ImageNet Large Scale Visual Recognition Challenge (ILSVRC) network structure is used for the discriminative network. The VGG feature extractor is a filter with a 3×3 convolution kernel, a step size of 1, and a fill mode of the same size. Thus, the overall model structure can be expressed as follows:

$$\begin{cases} F_1(Y) = f(W_1 * I^{LR} + b_1) \\ F_2(Y) = G_1(F_1(Y)) \\ F_3(Y) = G_2(F_2(Y)) \\ \vdots \\ F_{23}(Y) = G_{22}(F_{22}(Y)) \\ F_{24}(Y) = f(W_{23} * F_{23}(Y) + b_{23}) \\ Y^{HR1} = P(W_{24} * F_{24}(Y) + b_{24}) \\ Y^{HR2} = f(W_{25} * Y^{HR1} + b_{25}) \end{cases} \quad (5)$$

where I^{LR} is the input LR image, $W_j(j = 1, 2, \dots, 25)$ is the convolution filter, $*$ represents a convolution operation, and $f(x)$ is the excitation function. $G_i(F(Y))$ is a nonlinear feature extraction function:

$$\begin{aligned} G_i(F(Y)) &= \beta \times D(D(F_i(Y))), \\ D(F_i(Y)) &= f_{i+1}, i = 1, 2, \dots, 22; \beta \in (0, 1). \end{aligned} \quad (6)$$

$D(x)$ uses a stable connection to output f_{i+1} for an input $F_i(Y)$, $P(x)$ denotes the upsampling of the extracted image features to obtain an HR image, and Y^{HR2} is the final SR image.

B. LOCAL RESIDUAL NETWORK MODEL

Increasing the network depth is beneficial for obtaining more useful texture information. However, vanishing and exploding gradients represent obstacles when training deep

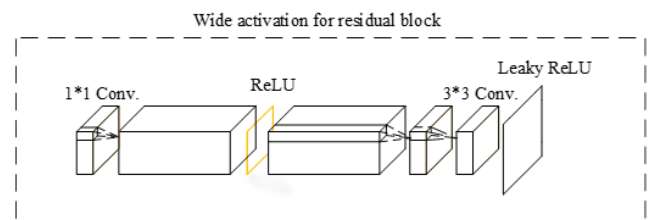


FIGURE 5. Network structure diagram of a WDRB.

networks, making it difficult for such a network to converge. To increase the network convergence speed, a combination of local residual learning with a multipath mode and multiweight recursive learning is used. The specific connection mode and structure are shown in Figure 4, where Figure 4(a) shows the RDBs used in SRDenseNet and Figure 4(b) shows the VRDB module proposed in this paper.

The VRDB module consists of four WDSR_B and leaky ReLU layers with dense jump connections and one convolutional layer. Together, one WDSR_B layer and one leaky ReLU layer form a wide-channel activation residual block (WDRB). The network structure used in this paper is depicted in Figure 5. The WDSR_B design of the linear low-rank convolution stack also increases the computational overhead of activation. In the WDSR_B design, all the activated ReLUs are applied only between two extended sets of features (features with an increased number of channels). Shi and Chu [37] also showed the benefits of such wide-channel activation that are achieved through more extensive activation and linear low-rank convolution. This model improves the accuracy of SR image reconstruction without requiring additional parameters or calculations. A comparison of the results before and after wide-channel activation is provided in Figure 6.

First, the input f_n yields more abundant texture information when processed with the wide-channel activation layer (WDSR_B). Then, the image feature extraction process is completed by the leaky ReLU layer. The specific network structure is shown in Figure 4(b). Next, the input passes sequentially through three identical pairs of WDSR_B and leaky ReLU layers through jump connections to further

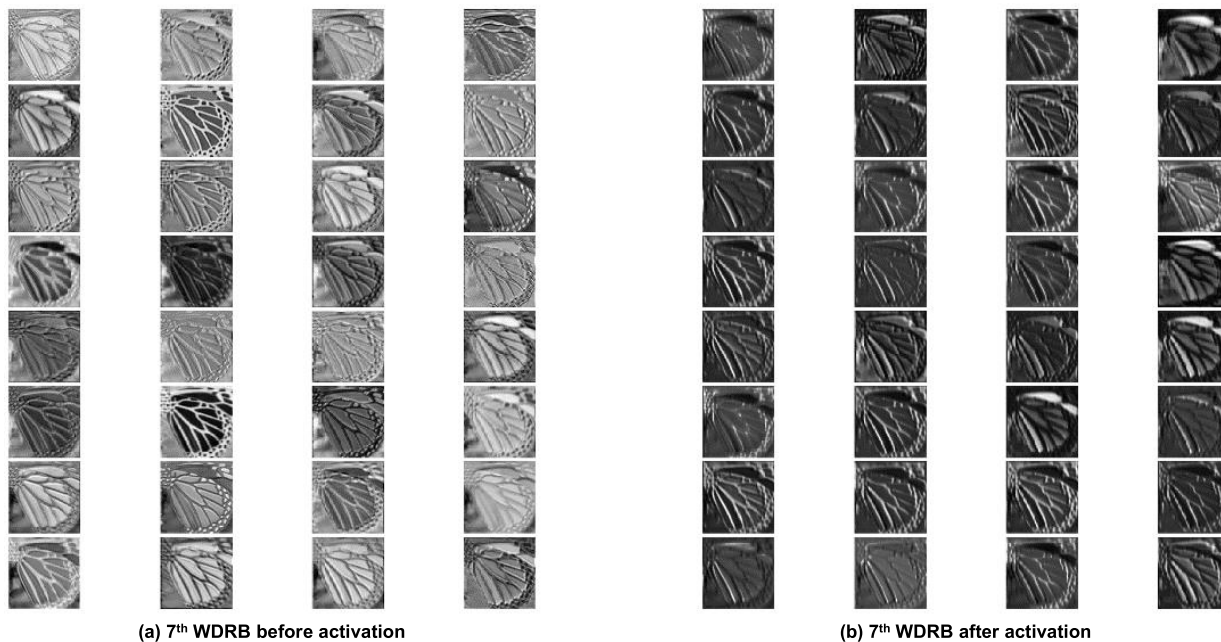


FIGURE 6. Comparison of the full-channel 7th WDRB before and after activation: (a) 7th WDRB before activation; (b) 7th WDRB after activation.

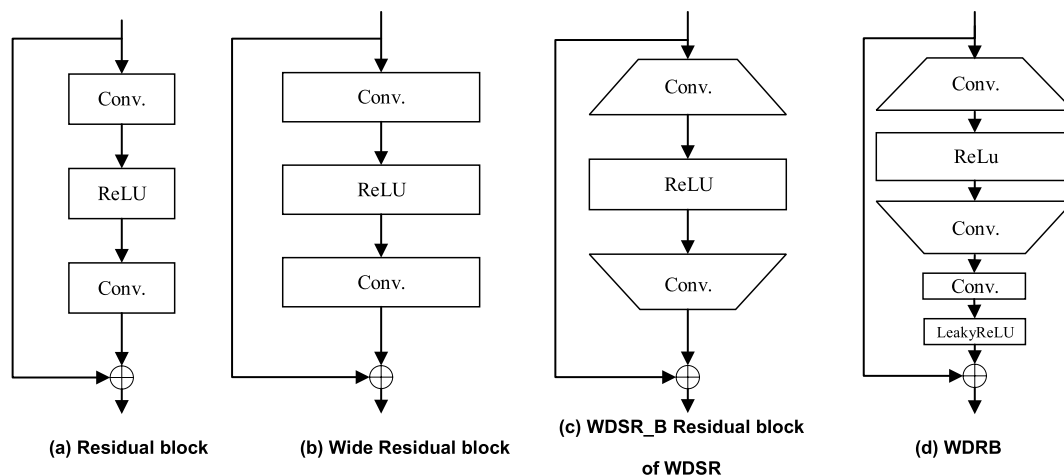


FIGURE 7. Comparison among residual block structures: (a) conventional residual block; (b) wide-channel residual structure; (c) wide-channel activated residual block; and (d) the WDRB proposed in this paper.

extract deeper image features. Finally, after a 3×3 convolutional layer, the high-frequency image features are learned and added to f_n to obtain f_{n+1} . In this local residual network structure, the local residual units are stacked sequentially, and different residual units have different inputs. Subsequently, a multipath jump connection structure is used such that all the remaining units share the same input. Compared with the recursive mode, this multipath mode is more conducive to learning and is not easily susceptible to overfitting.

C. COMPARATIVE DESIGN OF RESIDUAL MODULES

ReLU is commonly used, unsaturated activation functions that make the output of a network sparse, thereby reducing the number of calculations while retaining the information

pertaining to the main features. However, it is easy to lose other small texture information. Nevertheless, the loss of fine texture information should be avoided as much as possible. This paper considers the use of a 1×1 convolution kernel to expand the number of feature channels before activation, which is helpful for alleviating the sparsity of the network output after activation [38]. To this end, this paper improves upon and develops a WDRB that is suitable for GAN structures (as shown in Figure 7(d)). The proposed image SR reconstruction method based on a GAN generates more image feature information than does the SR model. Therefore, this paper adds a convolutional layer and a leaky ReLU layer on the basis of the original WDSR_B to better retain and extract effective image feature information.

Different from Figure 7(a), (b), and (c), the residual blocks are directly connected in series. In this paper, four WDSR_B layers and a convolution layer are connected as a VRDB by using dense jump connections. The effective use of shallow information to speed up network calculations is also conducive to generating higher-quality reconstructed images.

The wider residual block shown in Figure 7(b) allows more influx of shallow information [39] than does the structure shown in Figure 7(a) [15] but also adds additional calculations. In contrast, the version in Figure 7(c) increases the number of input channels to the active layer without increasing the parameters [37]. The number of parameters required for the residual block in Figure 7(a) is $W = 2 \times W_1 \times W_2 \times k^2 = 2 \times W_1^2 \times k^2$, which is equivalent to $r = 1$. In the residual block calculations illustrated in Figure 7(c), the number of parameters required is $W = 2 \times \hat{W}_1 \times \hat{W}_2 \times k^2 = 2 \times \hat{W}_1^2 \times k^2$, where k represents the convolution kernel size. To ensure that the number of parameters does not increase, $\hat{W}_1 = 1/\sqrt{r}W_1$, and the corresponding $W_2 = r \times 1/\sqrt{r} = \sqrt{r}W_1$. The proposed WDRB has the advantages of the block illustrated in Figure 7(b) but requires only the number of calculations of that shown in Figure 7(a). W_1 or \hat{W}_1 and W_2 or \hat{W}_2 replace the weights required before and after convolution, respectively [29].

D. PERCEPTUAL LOSS

To improve the overall perceptual quality of the final SR images, this paper uses a relativistic average discriminator, denoted by D_{Ra} , instead of the standard discriminator. The standard discriminator used in SRGAN can be expressed as $D(x) = \sigma(C(x))$, where σ is the sigmoid function and $C(x)$ is the nontransformed discriminator output. D_{Ra} can be expressed as $D_{Ra}(x_r, x_f) = \sigma(C(x_r) - IE_{x_f}[C(x)])$, where IE_{x_f} represents the operation of averaging all the fake data in the mini-batch. The discriminator loss can be defined [23] as follows:

$$L_D^{Ra} = -E_{x_r}[\log(D_{Ra}(x_r, x_f))] - E_{x_f}[\log(1 - D_{Ra}(x_f, x_r))]. \quad (7)$$

The adversarial loss function is expressed as follows:

$$L_G^{Ra} = -E_{x_r}[\log(1 - D_{Ra}(x_r, x_f))] - E_{x_f}[\log(D_{Ra}(x_f, x_r))]. \quad (8)$$

In Formula (8), $x_f = G(x_i)$, where x_i represents the input LR image. The adversarial loss of the generative network includes both x_f and x_r . Therefore, the generator benefits from the gradient of both the data generated during training and the actual data, while in SRGAN, only the generated data exert an effect. This discriminator modification helps the network learn to generate sharper edges and more delicate textures.

Regarding the design of the loss function, the perceptual loss function L_{percep} based on the preactivation constraint features is used instead of the loss function after activation, as in SRGAN. Thus, in combination with Formula (8),

the loss function of the generator is obtained as follows:

$$L_G = L_{percep} + \lambda L_G^{Ra} + \eta l_1, \quad (9)$$

where $l_1 = IE_{x_a} \| G(x_i) - y \|_1$ is the minimum absolute deviation (l_1) loss between the estimated recovered image $G(x_i)$ and the real image y and λ and η are coefficients that balance the different loss terms [23].

E. EVALUATION INDICATORS

To better evaluate the performances of various image SR algorithms, the PSNR, SSIM, and natural image quality evaluator (NIQE) [40] can be used as image quality evaluation metrics. The PSNR is a full-reference image quality evaluation indicator. Let MSE denote the mean square error between the current image X and the reference image Y ; let H and W denote the image height and width, respectively; and let n denote the number of bits per pixel (generally 8), meaning that the number of possible gray levels of a pixel is 256. The PSNR is expressed in units of dB. The larger the PSNR is, the smaller the distortion. The MSE and PSNR are calculated as follows:

$$MSE = \frac{1}{H \times W} \sum_{i=1}^{i=H} \sum_{j=1}^{j=W} (X(i, j) - Y(i, j))^2, \\ PSNR = 10 \log_{10} \left(\frac{(2^n - 1)^2}{MSE} \right). \quad (10)$$

The SSIM is another full-reference image quality evaluation metric that measures image similarity from three perspectives: brightness, contrast, and structure. SSIM values closer to 1 indicate greater similarity between the original (x) and reconstructed image blocks (\hat{x}) and represent a better reconstruction effect. The formula for calculating the SSIM is as follows:

$$SSIM(x, \hat{x}) = \frac{(2u_x u_{\hat{x}} + C_1)(2\sigma_{x\hat{x}} + C_2)}{(u_x^2 + u_{\hat{x}}^2 + C_1)(\sigma_x^2 + \sigma_{\hat{x}}^2 + C_2)}. \quad (11)$$

In Formula (11), u_x and $u_{\hat{x}}$ represent the means of image blocks x and \hat{x} , respectively; σ_x , $\sigma_{\hat{x}}$ and $\sigma_{x\hat{x}}$ represent the variances of image blocks x and \hat{x} and their covariance, respectively; and C_1 and C_2 are constants.

The PSNR and SSIM are the most commonly used objective indicators for image evaluation. However, they are based on the errors between corresponding pixels; that is, they evaluate image quality based on error sensitivity. Since these measures do not consider the visual characteristics of the human eye, the evaluation results are often inconsistent with human subjective perception. Therefore, this study uses the NIQE, an objective evaluation measure for image quality that more closely reflects the subjective evaluation of the human eye.

The formula for calculating the NIQE is as follows:

$$D(v_1, v_2, \sum_1, \sum_2) = ((v_1 - v_2)^T \left(\frac{\sum_1 + \sum_2}{2} \right)^{-1} (v_1 - v_2)) \frac{1}{2}. \quad (12)$$

TABLE 1. List of public SR image datasets.

Dataset	Amount	Avg. Resolution	Avg. Pixels	Format	Category Keywords
Set14	14	492×446	230,203	PNG	Humans, animals, insects, flowers, vegetables, comics, slides, etc.
BSD100	100	435×367	154,401	PNG	Animals, buildings, food, landscape, people, plants, etc.
Urban100	100	984×797	774,314	PNG	Architecture, cities, structure, urban environments, etc.
BSD200	200	433×370	154,401	PNG	Animals, building, food, landscapes, people, plants, etc.

Here, the means and variance matrices of the multivariate Gaussian (MVG) models for the natural and distorted images are v_1, v_2, Σ_1 and Σ_2 . The larger the values of these parameters are, the better the image quality. The MVG models are calculated as follows:

$$f_x(x_1, x_2, \dots, x_k) = \frac{\exp[-\frac{1}{2}(x - v)^T \Sigma^{-1}(x - v)]}{(2\pi)^{k/2} |\Sigma|^{1/2}}, \quad (13)$$

where $x_1, x_2, \dots, x_k (k = 1, 2, \dots, n, n \in \mathbb{N}^*)$ denote the extracted image features, v is the mean, and Σ is the variance matrix. Here, v and Σ can be obtained through maximum likelihood estimation.

IV. RESULTS AND DISCUSSION

A. MODEL TRAINING PARAMETERS

In this study, the DIV2K dataset was used as the training set [41]. A comparative analysis was then performed using four standard datasets: Set14 [42], BSD100 [43], Urban100 [44], and BSD200 [43]. The characteristics of each dataset and the differences among them are shown in Table 1. The bicubic degradation model was used in the experiments. The PSNR, SSIM, and NIQE were used to evaluate the SR images based on the Y channel in the YCbCr color space. Among them, the NIQE is a no-reference quality indicator that is more consistent with subjective quality evaluations based on the human visual system. To improve the network convergence speed and visual effect, each batch of images was cropped to 128×128 subpixel blocks, and the training data were composed of the cropped HR pixel blocks. The training dataset contained a total of 32,208 images. The Adam optimizer was used to train the model with settings of $\beta_1 = 0.9$ and $\beta_2 = 0.99$. The learning rate was initially set to 10^{-4} , and attenuation by a factor of two was performed for every 2×10^5 small batches. The generator was trained using the loss function in Formula (9) with $\lambda = 5 \times 10^{-3}$ and $\eta = 1 \times 10^{-2}$. The generative and discriminative networks

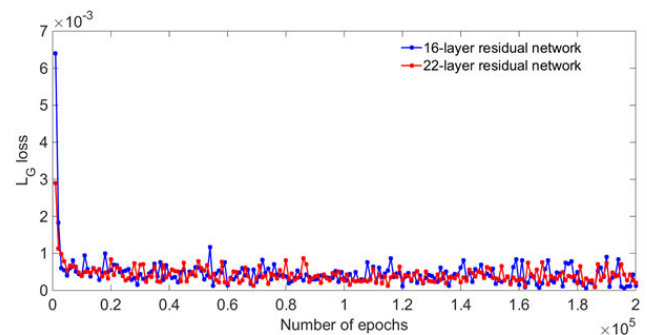


FIGURE 8. L_G loss versus the number of iterations with different numbers of network layers.

were alternately updated until the relativistic average discriminator could continuously discriminate between the real and reconstructed images with a probability of 0.5, at which time the model was deemed to have converged. Generators with two different structures were tested: one contained 16 residual blocks with a capacity similar to that of SRGAN, and the other was a deeper model with 22 RDBs. The proposed model was implemented using the PyTorch framework on an NVIDIA 1060 GPU.

B. RESIDUAL LAYER OPTIMIZATION

Network depth is an essential factor that affects the quality of SR images [45], [46]. Increasing the network depth enables more features to be extracted. However, the deeper the network is, the worse its convergence behavior becomes, making it more susceptible to overfitting [13].

In this study, the size of the dataset was increased, and a local residual network structure was used to accelerate network convergence while effectively avoiding overfitting. At a magnification of 4, network models with 16 and 22 residual layers were trained separately. To ensure complete network convergence, 200,000 iterations were performed. Figure 8 shows the relationship between the L_G loss for the 16-layer

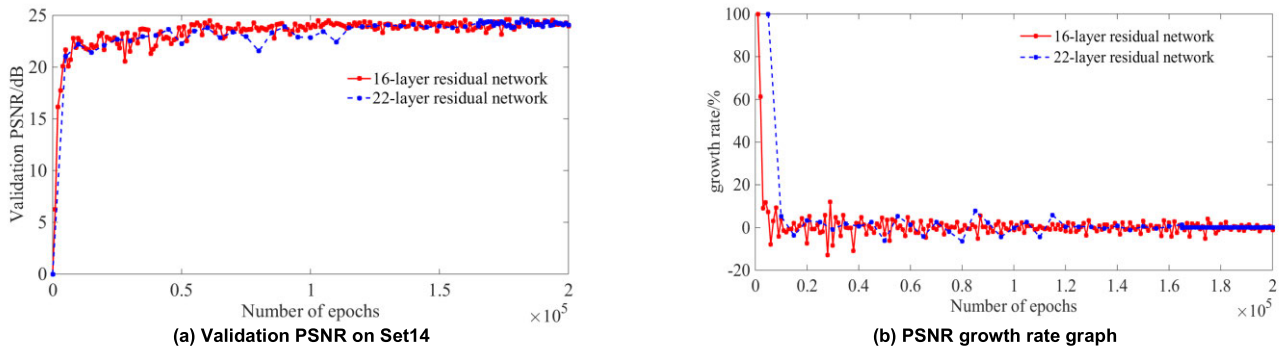


FIGURE 9. Changes in the PSNR with the number of iterations: (a) average PSNR value on Set14 versus the number of iterations; (b) rate of change in the PSNR.

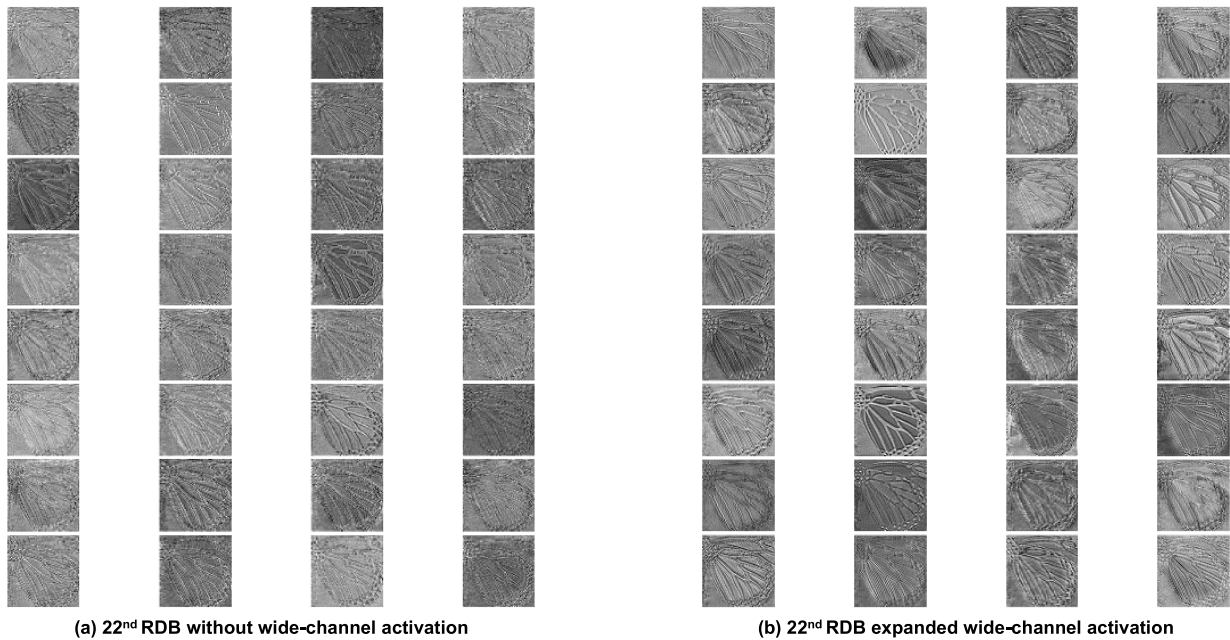


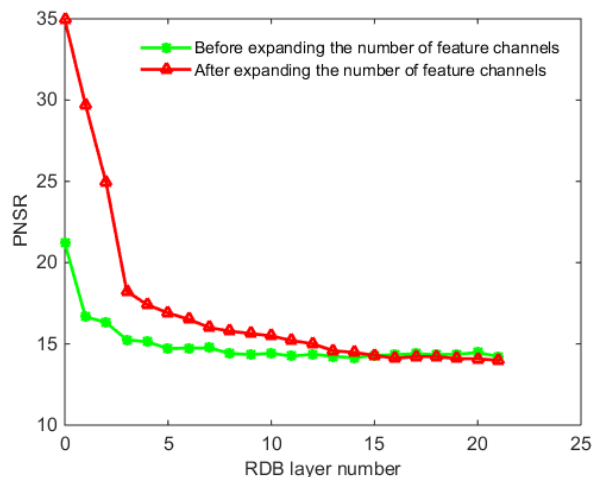
FIGURE 10. Comparison of RDBs with and without wide-channel activation: (a) feature map of the 22ndRDB layer without wide-channel activation; (b) feature map of the 22ndRDB layer with wide-channel activation.

and 22-layer networks and the number of iterations. This figure shows that the final LG_loss and PSNR tend to become stable and approximately parallel, indicating that the network has fully converged and that no overfitting phenomenon has occurred. To further verify the validity of the model, Set14 was selected as the test set, and the average PSNR was used as the evaluation index. Figure 9(a) shows that the average PSNR in the test data set (Set14) changes with the number of iterations. Figure 9(b) presents a graph showing the PSNR growth rate curve, demonstrating that the average PSNR of the 22-layer network is slightly larger than that of the 16-layer network. The exact values are shown in Table 4. The convergence of the 16-layer residual network is significantly faster than that of the 22-layer network. The 16-layer residual network begins to converge after 70,000 iterations, whereas the 22-layer residual network begins to converge only after 115,000 iterations. These results show that when

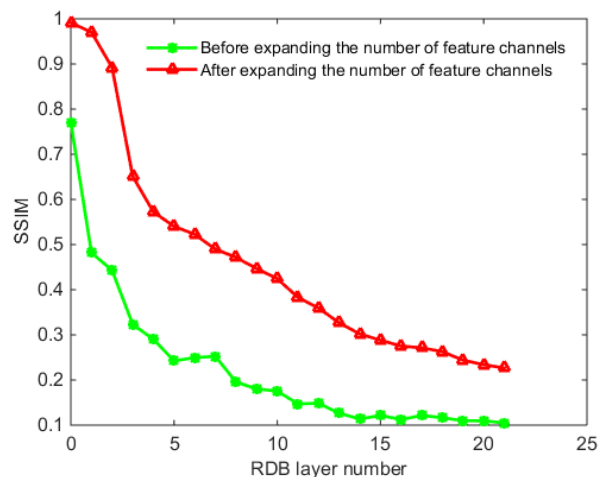
a model with high convergence speed and only moderate accuracy is required, the 16-layer residual network should be selected. When the perceptual quality of SR images is of greater concern, the 22-layer residual network should be adopted, as it yields better average PSNR and SSIM values and better image quality. This paper focuses on the perceptual quality and visual effect of SR images; therefore, the 22-layer network model is selected for subsequent experiments.

C. ACTIVATION COMPARISON WITH DIFFERENT NUMBERS OF FEATURED CHANNELS

To verify whether using 1×1 convolution to expand the number of feature channels can improve the utilization of shallow information, this paper selected the feature visualization of the RDB layer. The butterfly image in the Set5 [44] dataset was used as the input image. The LR images are shown in Figure 2. Figure 10 shows a comparison between



(a) PSNR curve for the 22-layer RDB network with and without wide-channel activation



(b) SSIM curve for the 22-layer RDB network with and without wide-channel activation

FIGURE 11. PSNR and SSIM curves for the 22-layer RDB network with and without wide-channel activation: (a) PSNR curve for the 22-layer RDB network with and without wide-channel activation; (b) SSIM curve for the 22-layer RDB network with and without wide-channel activation.

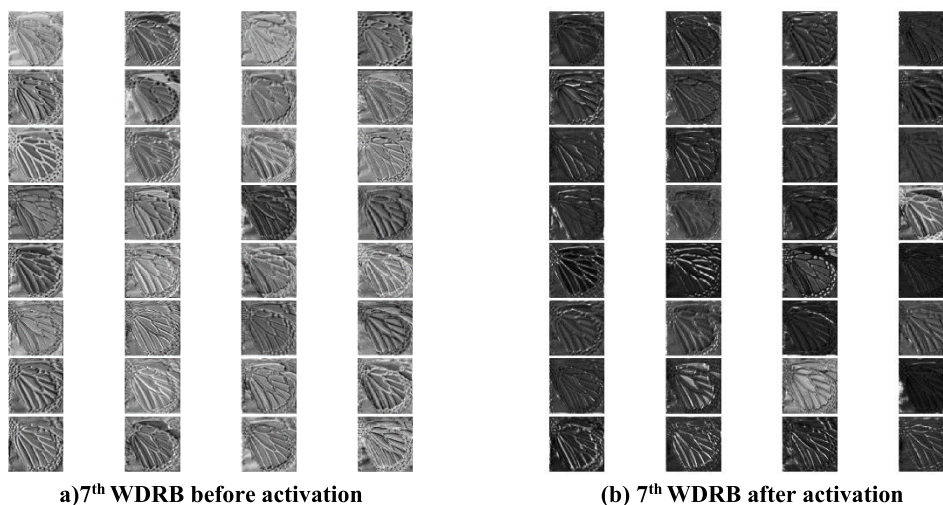


FIGURE 12. Compare the 7th WDRB without wide-channel activation before and after: (a) 7th WDRB before activation; (b) 7th WDRB after activation.

the RDBs with and without the 1×1 convolutional expansion of the number of feature channels. Figure 11 presents the further verification of the results. As the number of network layers increases, the utilization of shallow information is significantly reduced, and thus, fewer features are extracted. To test this hypothesis, this article compares the feature maps of the 7th WDRB with and without wide-channel activation (see Figure 6 and Figure 12). A comparison of Figure 6(b) and Figure 12(b) clearly shows that more feature information is retained by using a WDRB. The experiments show that expanding feature channels before ReLU activation is helpful for alleviating network sparsity after activation (as shown in Figure 6 and Figure 12).

Figure 10 shows that the feature information of the 22nd layer is quite different after wide-channel activation.

There is a serious loss of RDB feature information without wide-channel activation, whereas the RDB layer with wide-channel activation retains distinct feature information. The shallow information of the network is lost in the deeper layers, resulting in a lower utilization rate; thus, a worse performance is achieved. By contrast, the activation of a wide feature channel enables better retention of the original shallow feature information and improves the utilization rate.

In Figure 11, the experimental results show that as the number of network layers increases, either the PSNR or the SSIM in each RDB layer decreases compared to that in the first layer. After activation, the PSNR or SSIM significantly increases, and the overall perceived quality is better, which indicates that the RDB feature map activated with wide channels includes more shallow information than that

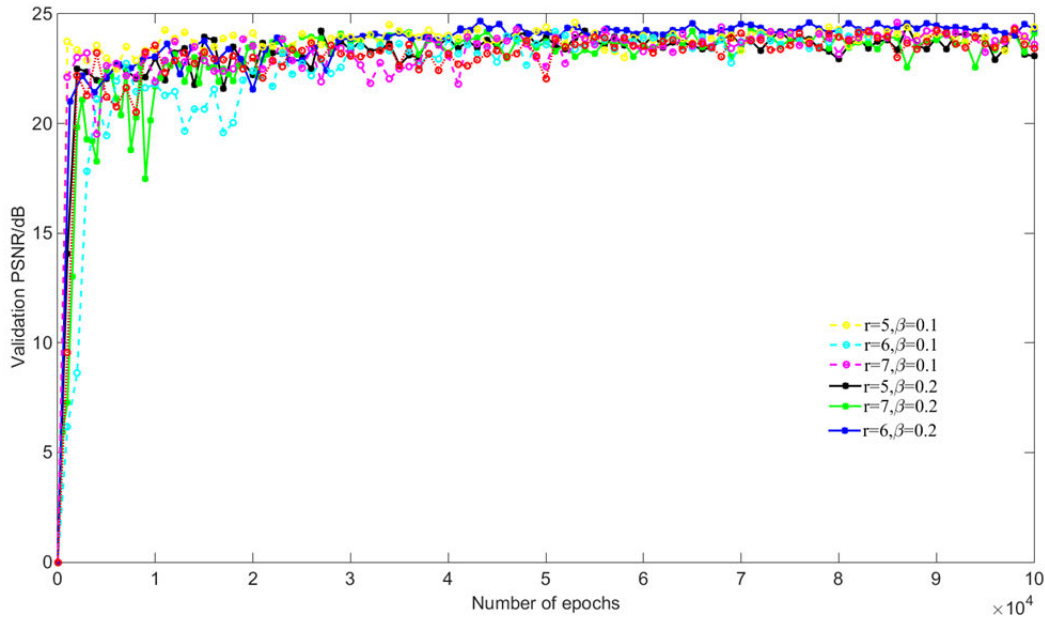


FIGURE 13. Variation in the training process under different feature expansion parameters and residual scaling parameters.

without wide-channel activation. The average PSNR after activation is 17.27, which is 2.30 higher than that before activation, and the overall utilization of shallow information increases by approximately 15.34%. According to the SSIM evaluation, the average SSIM values after and before activation are 0.46 and 0.22, respectively. Therefore, the utilization of shallow information overall improves by approximately 105.52%; that is, after wide-channel activation, the utilization of feature information is significantly improved.

D. ABLATION ANALYSIS

An ablation analysis was performed for the residual scaling parameter ($\beta = 0.1, 0.2, 1$) and the feature expansion parameter ($r = 1, 5, 6, 7$). First, the existence of residual scaling and feature expansion channels are investigated. If feature expansion and residual scaling are not present, they are both set to 1; otherwise, the default values are used for feature expansion and residual scaling (6 and 0.2, respectively). Set14 is used as the test dataset, and the PSNR is used as the image quality evaluation index. To further demonstrate the effectiveness of feature expansion and residual scaling on the model performance, the residual scaling parameters are set to 0.1 and 0.2, and the feature expansion parameters are set to 5, 6, and 7. The specific experimental results are shown in Figure 13.

Figure 13 demonstrates that when the number of iterations is 30,000, various structures may seem to converge. To better reflect the influence of each parameter structure on the model, the average PSNR value of the RGB channel during training epochs 30,000-100,000 was selected as the evaluation index. The specific results are shown in Figure 14.

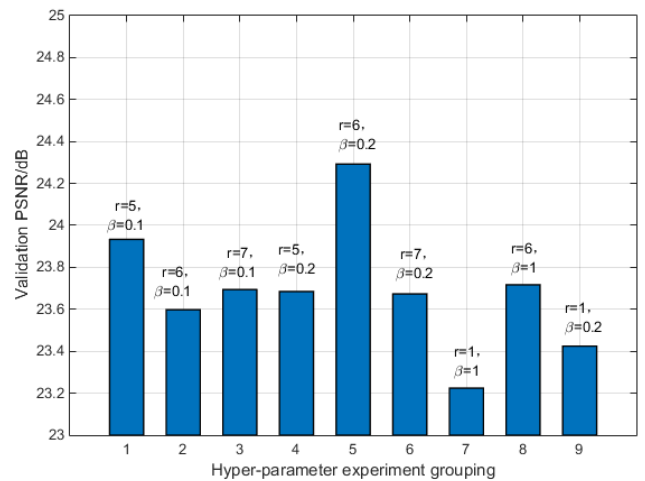


FIGURE 14. Comparison of the average PSNR values of RGB channels under different parameter structures.

Figure 14 shows that when the feature expansion parameter is 6 and the residual scaling parameter is 0.2 ($r = 6, \beta = 0.2$), the average PSNR value is the highest. In contrast, the average PSNR value is the lowest when feature expansion and residual scaling are not used ($r = 1, \beta = 1$ in Figure 14). In addition, from the comparison of $r = 6, \beta = 1$ and $r = 1, \beta = 1$, the use of feature expansion significantly improves the model performance, while the simple use of residual scaling does not significantly improve the network performance ($r = 1, \beta = 1$ and $r = 1, \beta = 0.2$ in Figure 14). Therefore, it can be concluded that when either feature expansion or residual scaling is employed alone, the former has a greater impact on model performance

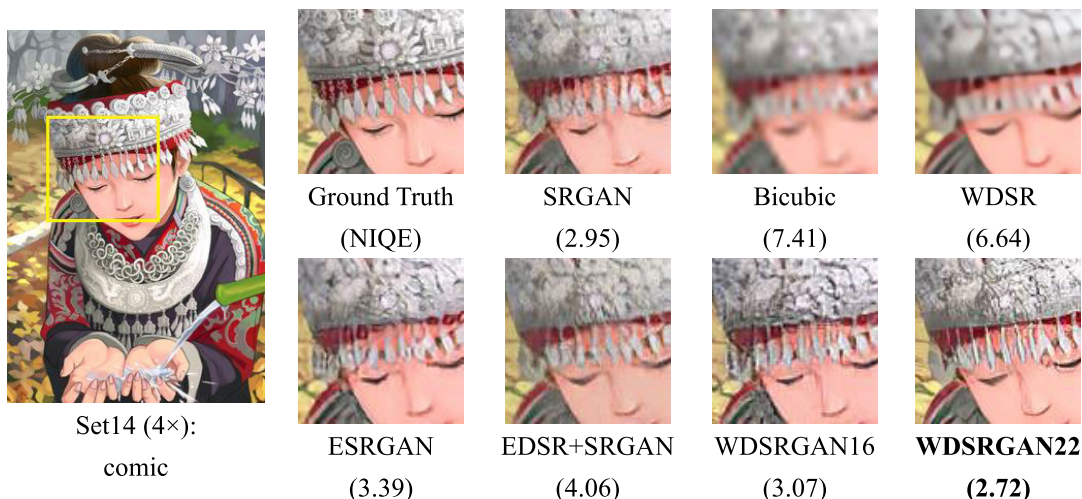


FIGURE 15. Comparison among the textural details of the different algorithms on a comic image; the best results are highlighted.



FIGURE 16. Comparison of the image authenticity among the different algorithms for image 138032; the best results are highlighted.

than the latter. When the feature expansion parameter is 5, a residual scaling parameter of 0.1 performs best. When using the feature expansion parameter is 6, a residual scaling parameter of 0.2 performs best. Hence, residual scaling has little effect on the performance of the model with a feature expansion parameter of 7.

E. QUALITATIVE ANALYSIS

When evaluating the image perception quality, we should not only consider the high-frequency texture information and visual effect of the image but also emphasize the image restoration accuracy instead of blindly pursuing smooth but unrealistic images. Therefore, the restoration of image textures, the generation of erroneous texture information, and the existence of other undesirable (noisy or unrealistic) information (such as artifacts, partial image details, and color distortion) are all critical considerations when evaluating image quality. To test the effectiveness of the proposed algorithm,

the SRGAN [11], bicubic WDSR [37], EDSR+SRGAN, and ESRGAN [23] models were compared with the two proposed structural models.

WDSR is the winning model from NTIRE 2018; it uses a wide-channel activation method similar to that proposed in this study. The model in [39] uses a wide-channel residual block similar to DRRN. ESRGAN ranked first in the PIRM2018-SR Challenge competition and obtained a higher perceived quality index than both EDSR and DRRN [47]. The proposed models, called WDSRGAN16 and WDSRGAN22, and the other five models were tested on the Set14, BSDS200, and Urban100 datasets (with a magnification of 4). The experimental results show that the proposed algorithm recovers more image texture details and that its visual quality is the best (Figure 15). On “image 138032”, the ESRGAN model restores the detail of the rope poorly and produces the wrong texture information (Figure 16). On “img_011”, the EDSR+SRGAN model suffers from

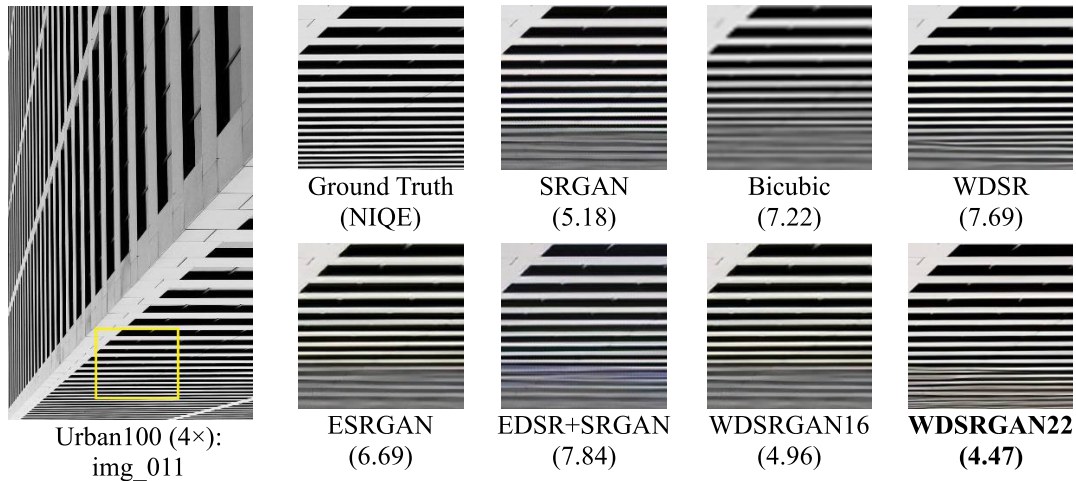


FIGURE 17. Comparison of the color and brightness monitoring performance among the different algorithms on *img_011*; the best results are highlighted.

color distortions at the edge of the tall building, which is blue (Figure 17).

Comparing the various algorithms on the comic image shown in Figure 15, the lower the NIQE value is, the better the perceptual quality of the image. Among the tested algorithms, the bicubic algorithm has the worst image restoration quality, characterized by a high degree of ambiguity. A large amount of high-frequency information is not recovered because simple linear interpolation is insufficient for the highly complex task of image reconstruction. WDSR is used as the loss function l_2 . For this kind of SR image, this loss function restores only the primary outline information and the image texture; hence, the

SR image feature information extracted by WDSR is insufficient. EDSR+SRGAN and SRGAN preserve more details than do the above algorithms; however, their results often contain artifacts and abnormal variations in image color. This is because the GAN is sensitive to the noise introduced by the BN layer. ESRGAN performs effectively in terms of monitoring the image color and brightness and recovering texture information; however, compared to the real image, the restoration quality of the content is insufficient. WDSRGAN22 shows superior performance compared to all the above algorithms, yielding an NIQE value of 2.72, which is 19.76% lower than that of ESRGAN, indicating higher-quality detail restoration compared with ESRGAN. This is because the WDSRGAN22 network structure uses a 1×1 convolutional layer to expand the number of feature channels by a factor of 6–9, thus improving the utilization of shallow information and making the network more conducive to the transmission of information. In addition, the proposed residual network structure differs from the previous residual structure: three VRDBs are used to form a wide local residual network block, and residual scaling is used, where each VRDB structure contains dense jump connections. The improvement in utilizing shallow information contributes to

the transmission of more information throughout the network layers, thereby improving the quality of the restored image. The map with the best perceptual quality perception is highlighted with bold font in each figure.

Figure 16 shows a comparison of the image authenticity test results among the various algorithms. The SR images reconstructed with WDSR and the bicubic algorithm are very fuzzy, whereas SRGAN and EDSR+SRGAN successfully restore the texture information of part of the rope image but also introduce noticeable artifacts. Because of the excessive pursuit by the ESRGAN model of an aesthetically pleasing visual effect, the contents of the images restored by ESRGAN do not match the original image information; in fact, for images with rich texture information, ESRGAN often produces texture information that fails to match the original image. In contrast, the results of the proposed network structure are more realistic with a minimum NIQE value of 4.02.

In the comparison among the color and brightness monitoring effects of the different algorithms presented in Figure 17 (*img_011*), SRGAN, WDSR, ESRGAN, and WDSRGAN16 show little differences in brightness monitoring performance. However, the NIQE value of WDSRGAN22 is 4.47; its generated image is sharper than those generated by the other algorithms, and its brightness is stronger than that of the SR images generated by the other algorithms. The SR image reconstructed by the EDSR+SRGAN algorithm exhibits an undesirable bluish phenomenon, but it achieves the highest NIQE value.

In addition to the qualitative and quantitative analyses performed on several commonly used image datasets with high variability, a real underwater fish dataset was also used to verify the reliability of the proposed algorithm. The same six algorithms, i.e., the bicubic, SRGAN, ESRGAN, WDSR, EDSR+SRGAN, and WDSRGAN algorithms, were tested and compared on this dataset in Figure 18, and the quality

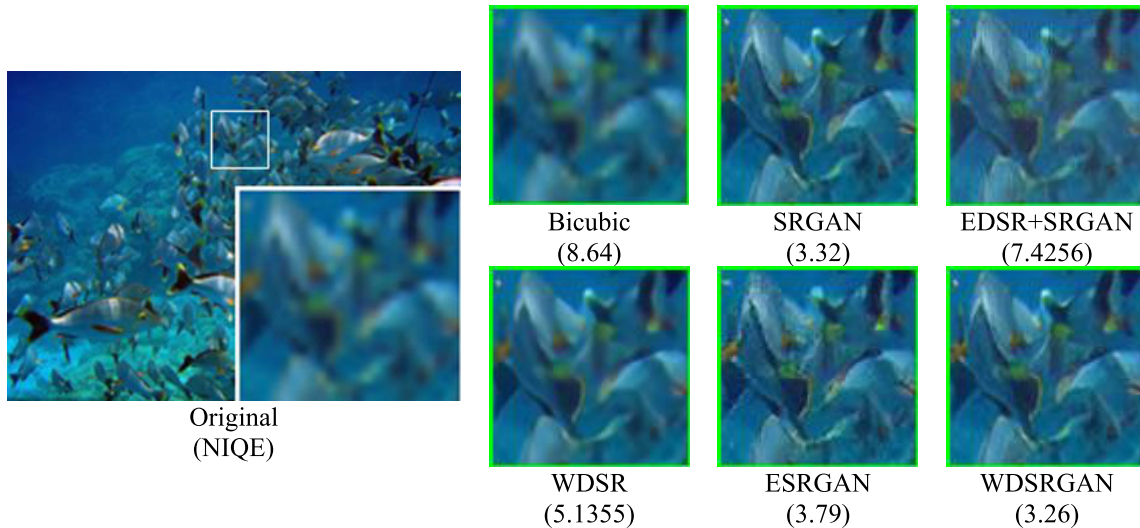


FIGURE 18. Comparison among the enhancement and reconstruction results for a real underwater image of fish using the different algorithms; the best results are highlighted.

TABLE 2. Average NIQE results on different datasets.

Dataset	Bicubic	EDSR+SRGAN	SRGAN	WDSR	ESRGAN	WDSRGAN 16	WDSRGAN 22
Set14	7.83	4.87	3.82	6.71	3.76	3.63	3.52
Urban100	7.32	5.04	3.69	6.79	4.09	3.52	3.44
BSD100	7.68	5.21	3.43	6.89	3.52	3.46	3.26
BSD200	7.66	5.47	3.69	6.99	3.66	3.61	3.41

of the SR images was evaluated using the NIQE. Again, the bicubic algorithm achieves the worst reconstruction effect relative to the original image with an NIQE of 8.64. The reconstruction effect of the proposed algorithm is the best with an NIQE value of 3.26. The effect of the SRGAN is suboptimal, although it can sharpen the fuzzy details of the original image. These experimental results show that the proposed algorithm is similarly suitable for enhancing real underwater fish images, and it achieves the best performance among the tested algorithms.

F. QUANTITATIVE ANALYSIS

This paper reports the results of testing the bicubic, SRGAN, ESRGAN, WDSR, EDSR+SRGAN, 16-layer WDSRGAN, and 22-layer WDSRGAN algorithms on four different datasets: Set14, BSDS100, BSDS200 and Urban100 (at a magnification of 4). The quality of the SR images was evaluated using the NIQE, and the results are shown in Table 2. Considering that the human eye is sensitive to brightness contrasts and to better show the differences among the algorithms, experimental calculations were performed on the Y channel of the YCbCr color space to calculate the PSNR and SSIM for the Y channel only. The results of these calculations are shown in Table 3 and Table 4.

As seen in Table 2, the NIQE value of WDSRGAN16 is only slightly higher than that of WDSRGAN22, which reports the lowest NIQE among all the algorithms considered for the comparison with an average value of 3.4 (the lower the NIQE value is, the better the image quality). The average NIQE value of WDSRGAN22 is 0.3 lower than that of ESRGAN, which achieves the highest perceptual quality score. The average NIQE value of WDSR is 6.84, indicating that the resulting SR images do not match human perception; they are too smooth and lack textural details. The SR images produced through bicubic interpolation are the most ambiguous with an average NIQE value of 7.62, which is 4.22 higher than that of WDSRGAN22.

From the Y-channel PSNR and SSIM results presented in Table 3 and Table 4, the PSNR and SSIM values of the ESRGAN and WDSRGAN16 models are similar, but the

WDSRGAN22 model performs better. This shows that strengthening the use of shallow information can help improve network model performance. On the Set14 and BSDS100 datasets, the average PSNR of WDSRGAN22 is 0.3 dB higher than that of ESRGAN, and the average SSIM is 0.45 higher. Most notably, the average PSNR of WDSRGAN22 increases by 0.45 dB on the Urban100 dataset. Thus, the proposed algorithm exhibits better performance, and is

TABLE 3. Average Y-channel PSNR results on different datasets.

Dataset	Bicubic	SRGAN	ESDR+ SRGAN	ESRGAN	WDSR	WDSRGAN16	WDSRGAN22
Set14	26.10	26.65	25.40	26.29	28.14	26.36	26.54
BSDS100	25.96	24.88	24.96	25.33	27.21	25.32	25.63
BSDS200	26.42	25.11	25.38	25.91	27.64	25.76	26.16
Urban100	23.14	24.37	23.45	24.37	25.58	23.73	24.82

TABLE 4. Average Y-channel SSIM results on different datasets.

Dataset	Bicubic	SRGAN	ESDR+ SRGAN	ESRGAN	WDSR	WDSRGAN16	WDSRGAN22
Set14	0.70	0.72	0.71	0.69	0.76	0.70	0.77
BSDS100	0.66	0.66	0.66	0.65	0.73	0.65	0.66
BSDS200	0.69	0.69	0.69	0.68	0.74	0.68	0.68
Urban100	0.65	0.72	0.70	0.73	0.76	0.70	0.73

not only applicable for the SR processing of a wide variety of images but also achieves higher image quality. Furthermore, according to the results, the SRGAN model yields a lower overall score than does WDSRGAN22 and tends to produce unpleasant visual artifacts. The WDSR algorithm, although it achieves a higher score and obtains a sharper picture than the bicubic algorithm, does not meet the requirements of human visual perception.

G. ALGORITHM EXECUTION SPEED COMPARISON

An increase in network depth is beneficial for improving image quality after reconstruction, but the resulting increase in network complexity also reduces the execution speed and increases the difficulty of network convergence. Therefore, this paper presents the results of an experiment conducted to compare the execution speeds of EDSR+SRGAN, SRGAN, and WDSRGAN22, all of which are similar in structure. The proposed reconstruction algorithm makes full use of dense jump connections and residual learning, which significantly reduces the computational complexity and improves the running time. Although the WDSRGAN22 model has approximately 85,470,908 parameters, its running time is 0.42 sec/frame, which is 3.6 times faster than that of SRGAN (1.411 sec/frame), even though SRGAN has the same number of parameters as WDSRGAN16. EDSR+SRGAN has a runtime similar to WDSRGAN22, but its PSNR value is 1.1 dB higher than that of WDSRGAN 22.

V. CONCLUSION

The WDSRGAN algorithm proposed in this paper achieves higher perceptual image quality and a higher PSNR value than previous algorithms. Based on the network structure of the original SRGAN model, dense jump connections

are adopted in each VRDB module. Every set of three VRDBs forms an RDB module, increasing the depth of the network while maintaining a runtime speed of 0.42 sec/frame—3.6 times faster than that of SRGAN. In addition, 1×1 convolution is used to combine the extended features in VRDB. Based on the PSNR and SSIM evaluations of each layer of feature images, the overall utilization of shallow information increases by 15.34% and 105.52%, respectively. An ablation experiment shows that a model network width expansion setting of $r = 6$ and a residual scaling setting of $\beta = 0.2$ are optimal. In tests on four standard datasets, the average NIQE value of the reconstructed images is 3.4, and the perceptual quality is the highest among all tested algorithms. In qualitative image evaluations, the image texture restoration effect, the generation of erroneous texture information, and the presence of noisy or unrealistic image information are essential considerations.

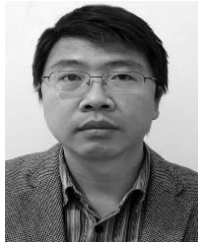
The method proposed in this paper, which combines a GAN with wide-channel activation for feature extraction, offers beneficial improvements compared with existing algorithms and has practical value for specific applications. Future work can focus on further optimizing the network model while ensuring the image quality to make the model lighter and enable its applications in other fields, such as medical imaging, hyperspectral imaging, and the SR reconstruction of terahertz images.

REFERENCES

- [1] P. Yi, Z. Wang, K. Jiang, Z. Shao, and J. Ma, "Multi-temporal ultra dense memory network for video super-resolution," *IEEE Trans. Circuits Syst. Video Technol.*, to be published.
- [2] L. Zhou, Z. Wang, Y. Luo, and Z. Xiong, "Separability and compactness network for image recognition and superresolution," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 30, no. 11, pp. 3275–3286, Nov. 2019.

- [3] K. Jiang, Z. Wang, P. Yi, G. Wang, T. Lu, and J. Jiang, "Edge-enhanced GAN for remote sensing image super-resolution," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 8, pp. 5799–5812, Aug. 2019.
- [4] Z. Shao, L. Wang, Z. Wang, and J. Deng, "Remote sensing image super-resolution using sparse representation and coupled sparse autoencoder," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 12, no. 8, pp. 2663–2674, Aug. 2019.
- [5] Z. Wang, P. Yi, K. Jiang, J. Jiang, Z. Han, T. Lu, and J. Ma, "Multi-memory convolutional neural network for video super-resolution," *IEEE Trans. Image Process.*, vol. 28, no. 5, pp. 2530–2544, May 2019.
- [6] Y. Sun, W. Zhang, H. Gu, C. Liu, S. Hong, W. Xu, J. Yang, and G. Gui, "Convolutional neural network based models for improving super-resolution imaging," *IEEE Access*, vol. 7, pp. 43042–43051, 2019.
- [7] Y. Wang, L. Wang, H. Wang, and P. Li, "End-to-end image super-resolution via deep and shallow convolutional networks," *IEEE Access*, vol. 7, pp. 31959–31970, 2019.
- [8] Y. Pang, J. Cao, J. Wang, and J. Han, "JCS-net: Joint classification and super-resolution network for small-scale pedestrian detection in surveillance images," *IEEE Trans. Inf. Forensics Secur.*, vol. 14, no. 12, pp. 3322–3331, Dec. 2019.
- [9] J. Shi, Q. Liu, C. Wang, Q. Zhang, S. Ying, and H. Xu, "Super-resolution reconstruction of MR image with a novel residual learning network algorithm," *Phys. Med. Biol.*, vol. 63, no. 8, Mar. 2018, Art. no. 085011.
- [10] C. Dong, C. C. Loy, K. He, and X. Tang, "Image super-resolution using deep convolutional networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 2, pp. 295–307, Feb. 2016.
- [11] 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. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Honolulu, HI, USA, Jul. 2017, pp. 4681–4690.
- [12] 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)*, Zürich, Switzerland, Sep. 2014, pp. 184–199.
- [13] 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)*, Las Vegas, NV, USA, Jun. 2016, pp. 1646–1654.
- [14] 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)*, Honolulu, HI, USA, Jul. 2017, pp. 624–632.
- [15] 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)*, Honolulu, HI, USA, Jul. 2017, pp. 136–144.
- [16] Y. Zhang, K. Li, K. Li, L. Wang, B. Zhong, and Y. Fu, "Image super-resolution using very deep residual channel attention networks," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, Munich, Germany, Sep. 2018, pp. 286–301.
- [17] 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)*, Las Vegas, NV, USA, Jun. 2016, pp. 1637–1645.
- [18] A. Kadurin, S. Nikolenko, K. Khrabrov, A. Aliper, and A. Zhavoronkov, "DruGAN: An advanced generative adversarial autoencoder model for de novo generation of new molecules with desired molecular properties in silico," *Mol. Pharmaceutics*, vol. 14, no. 9, pp. 3098–3104, Sep. 2017.
- [19] C. C. Dong, C. C. Loy and X. Tang, "Accelerating the super-resolution convolutional neural network," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, Amsterdam, The Netherlands, Oct. 2016, pp. 391–407.
- [20] Y. Tai, J. Yang, and X. Liu, "Image super-resolution via deep recursive residual network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Honolulu, HI, USA, Jul. 2017, pp. 3147–3155.
- [21] M. S. M. Sajjadi, B. Scholkopf, and M. Hirsch, "EnhanceNet: Single image super-resolution through automated texture synthesis," in *Proc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Venice, Italy, Oct. 2017, pp. 4491–4500.
- [22] S. Park, H. Son, S. Cho, K. Hong, and S. Lee, "SRFeat: Single image super-resolution with feature discrimination," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, Munich, Germany, Sep. 2018, pp. 439–455.
- [23] 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," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, Munich, Germany, Sep. 2018, pp. 63–79.
- [24] C. Zhou, X. Yang, B. Zhang, K. Lin, D. Xu, Q. Guo, and C. Sun, "An adaptive image enhancement method for a recirculating aquaculture system," *Sci. Rep.*, vol. 7, no. 1, Dec. 2017, Art. no. 6243.
- [25] C. Zhou, D. Xu, L. Chen, S. Zhang, C. Sun, X. Yang, and Y. Wang, "Evaluation of fish feeding intensity in aquaculture using a convolutional neural network and machine vision," *Aquaculture*, vol. 507, pp. 457–465, May 2019.
- [26] C. Zhou, K. Lin, D. Xu, L. Chen, Q. Guo, C. Sun, and X. Yang, "Near infrared computer vision and neuro-fuzzy model-based feeding decision system for fish in aquaculture," *Comput. Electron. Agricult.*, vol. 146, pp. 114–124, Mar. 2018.
- [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)*, Amsterdam, The Netherlands Oct. 2016, pp. 694–711.
- [28] W. Han, S. Chang, D. Liu, M. Yu, M. Witbrock, and T. S. Huang, "Image super-resolution via dual-state recurrent networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Salt Lake City, UT, USA, Jun. 2018, pp. 1654–1663.
- [29] D. Guo, Y. Niu, and P. Xie, "Speedy and accurate image super-resolution via deeply recursive CNN with skip connection and network in network," *IET Image Process.*, vol. 13, no. 7, pp. 1201–1209, May 2019.
- [30] Y. Zhang, Y. Tian, Y. Kong, B. Zhong, and Y. Fu, "Residual dense network for image super-resolution," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Salt Lake City, UT, USA, Jun. 2018, pp. 2472–2481.
- [31] N. Ahn, B. Kang, and K.-A. Sohn, "Fast, Accurate, and lightweight super-resolution with cascading residual network," 2018, *arXiv:1803.08664*. [Online]. Available: <http://arxiv.org/abs/1803.08664>
- [32] Y. Tai, J. Yang, X. Liu, and C. Xu, "MemNet: A persistent memory network for image restoration," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Venice, Italy, Oct. 2017, pp. 4539–4547.
- [33] T. Tong, G. Li, X. Liu, and Q. Gao, "Image super-resolution using dense skip connections," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Venice, Italy, Oct. 2017, pp. 4799–4807.
- [34] H. Chen, X. He, C. An, and T. Q. Nguyen, "Deep AC residual network based joint blocking and color bleeding artifacts reduction for 4:2:0 JPEG-compressed images," *IEEE Signal Process. Lett.*, vol. 26, no. 1, pp. 79–83, Jan. 2019.
- [35] Y. FanJ Yu and T. S. Huang, "Wide-activated deep residual networks based restoration for BPG-compressed Images.," in *Proc. conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Salt Lake City, UT, USA, Jun. 2018, pp. 2621–2624.
- [36] M. Haris, G. Shakhnarovich, and N. Ukita, "Deep AC networks for super-resolution," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Salt Lake City, UT, USA, Jun. 2018, pp. 1664–1673.
- [37] J. Yu, Y. Fan, J. Yang, N. Xu, Z. Wang, X. Wang, and T. Huang, "Wide activation for efficient and accurate image super-resolution," 2018, *arXiv:1808.08718*. [Online]. Available: <http://arxiv.org/abs/1808.08718>
- [38] S. Shi and X. Chu, "Speeding up convolutional neural networks by exploiting the sparsity of rectifier units," 2017, *arXiv:1704.07724*. [Online]. Available: <http://arxiv.org/abs/1704.07724>
- [39] J. Shi, Z. Li, S. Ying, C. Wang, Q. Liu, Q. Zhang, and P. Yan, "MR image super-resolution via wide residual networks with fixed skip connection," *IEEE J. Biomed. Health Informat.*, vol. 23, no. 3, pp. 1129–1140, May 2019.
- [40] A. Mittal, R. Soundararajan, and A. C. Bovik, "Making a 'completely blind' image quality analyzer," *IEEE Signal Process. Lett.*, vol. 20, no. 3, pp. 209–212, Mar. 2013.
- [41] E. Agustsson and R. Timofte, "NTIRE 2017 challenge on single image super-resolution: Dataset and study," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Honolulu, HI, USA, Jul. 2017, pp. 126–135.
- [42] R. Zeyde, M. Elad, and M. Protter, "On single image scale-up using sparse-representations," in *Proc. Int. Conf. Curves Surf.*, Avignon, France, Jun. 2010, pp. 711–730.
- [43] D. Martin, C. Fowlkes, D. Tal, and J. Malik, "A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics," in *Proc. 8th IEEE Int. Conf. Comput. Vis. (ICCV)*, Nov. 2002, pp. 416–423.
- [44] J.-B. Huang, A. Singh, and N. Ahuja, "Single image super-resolution from transformed self-exemplars," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Boston, MA, USA, Jun. 2015, pp. 5197–5206.

- [45] B. Du, Z. Wang, L. Zhang, L. Zhang, and D. Tao, "Robust and discriminative labeling for multi-label active learning based on maximum correntropy criterion," *IEEE Trans. Image Process.*, vol. 26, no. 4, pp. 1694–1707, Apr. 2017.
- [46] P. Liu, G. Wang, H. Qi, C. Zhang, H. Zheng, and Z. Yu, "Underwater image enhancement with a deep residual framework," *IEEE Access*, vol. 7, pp. 94614–94629, 2019.
- [47] S. Anwar, S. Khan, and N. Barnes, "A deep journey into super-resolution: A survey," 2019, *arXiv:1904.07523*. [Online]. Available: <https://arxiv.org/abs/1904.07523>



XUDONG SUN received the B.S. and M.S. degrees from China Agriculture University, Beijing, China, in 2002 and 2005, respectively, and the Ph.D. degree from Donghua University, Shanghai, China, in 2017. Since 2016, he has been an Associate Professor of nondestructive technology applications in food and agriculture with East China Jiaotong University. His research interests include machine vision, superresolution, and nondestructive technology.



ZHENXI ZHAO received the B.S. degree in mechanical engineering and automation from Pingxiang University, China, in 2018. He is currently pursuing the M.S. degree with East China Jiaotong University. His research interests include superresolution, computer vision, and deep learning.



SONG ZHANG received the B.S. degree in electrical engineering and automation from the Tianjin University of Science and Technology, in 2017, where he is currently pursuing the M.S. degree with the School of Electronic Information and Automation. His current research interests include image processing, computer vision, and deep learning.



JINTAO LIU received the B.S. and M.S. degrees from Qingdao Agricultural University, Qingdao, China, in 2014 and 2016, respectively. In 2018, he applied for the Ph.D. program in computer science at the University of Almería, Spain. He is currently conducting research with the Beijing Research Center for Information Technology in Agriculture, China. His current research interests include fish behavior, computer vision, and machine learning.



XINTING YANG received the B.S. degree from Qingdao Agricultural University, Qingdao, China, in 1997, the M.S. degree from Zhejiang University, in 2000, and the Ph.D. degree from the Research Center for Eco-Environmental Sciences, Chinese Academy of Sciences, China, in 2008. In July 2000, he worked with the Beijing Agricultural Information Technology Research Center and successively served as a Research Intern, an Assistant Researcher, an Associate Researcher, and a Researcher. Since 2012, he has been the Deputy Director of the Center. His research interests include machine vision, agricultural image, and machine learning.



CHAO ZHOU received the B.S. degree in automation engineering from Qingdao University, Qingdao, China, in 2007, the M.S. degree in control theory and control engineering from the Taiyuan University of Technology, China, in 2010, and the Ph.D. degree in control science and engineering from the Beijing Institute of Technology, Beijing, China, in 2018. In 2010, he joined the Department of Intelligent Systems, China National Engineering Research Center for Information Technology in Agriculture (NERCITA), where he is currently a Senior Engineer. His research interests include machine vision, intelligent control, and the application of machine learning techniques to fish farming.

...