IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 9 June 2024, accepted 28 July 2024, date of publication 6 August 2024, date of current version 19 August 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3439542

RESEARCH ARTICLE

Image Super-Resolution Reconstruction Based on Enhanced Attention Mechanism and Gradient Correlation Loss

YANLAN SHI[1], HUAWEI YI[1], XIN ZHANG[1], LU XU[2], AND JIE LAN[3]
[1]School of Electronics and Information Engineering, Liaoning University of Technology, Jinzhou 121001, China
[2]IT and Products Management Department, Agricultural Bank of China, Beijing 100005, China
[3]College of Science, Liaoning University of Technology, Jinzhou 121001, China

Corresponding author: Huawei Yi (yihuawei@126.com)

This work was supported in part by the National Natural Science Foundation for Youth Scientists of China under Grant 62203201, and in part by the Foundation Research Project of the Educational Department of Liaoning Province under Grant JYTMS20230860 and Grant LJKZZ20220085.

ABSTRACT In the field of super-resolution reconstruction, generative adversarial networks are able to generate textures that are more in line with the perception of the human eyes, but low-resolution images often encounter information loss and edge blurring problems in the process of reconstruction. In order to solve this problem, this article proposed an image super-resolution reconstruction model based on an enhanced attention mechanism and gradient correlation loss, which can better focus on important details in low-resolution images, thus improving the quality of reconstructed images. Firstly, an enhanced attention mechanism is proposed and incorporated into the generator model as a way to reduce the amount of information loss during image feature extraction and retain more image details. Furthermore, this paper proposed a gradient correlation loss function to maximize the correlation between the gradient of the generated image and the gradient of the original image. Thus, the generated image is more realistic and maintains a consistent edge structure. Finally, the experimental results on the standard dataset show that compared with other representative algorithms, the proposed algorithm has achieved some improvement in PSNR, SSIM, and LPIPS, which can verify the effectiveness of the algorithm.

INDEX TERMS Image reconstruction, super-resolution, enhanced attention mechanism, gradient correlation loss function.

I. INTRODUCTION

Super-resolution reconstruction of images is a technique for obtaining high-resolution images from single or multiple low-resolution images. The image super-resolution reconstruction technology is used to restore and reconstruct low-resolution images, which can effectively improve the details and quality of images. Super-resolution image reconstruction algorithms can be roughly divided into three categories, interpolation-based algorithms [1], reconstruction-based algorithms [2] and learning-based algorithms [3]. The first two categories belong to the traditional methods, which usually suffer from the drawbacks of overall blurring of the images and serious lack of details. In recent years, with the development of deep learning, learning-based super-resolution reconstruction technology has gradually become a hot topic. Among them, image super-resolution reconstruction methods based on convolutional neural network (CNN) and generative adversarial network (GAN) are widely used because the reconstruction performance of them is much better than the traditional algorithms.

In 2014, Dong et al. [4] proposed the super-resolution convolutional neural network (SRCNN), which used three convolutional layers for reconstruction and greatly improves the speed of reconstruction compared with traditional methods. In 2016, Kim et al. [5] proposed a recursive recurrent neural network (DRCN) that utilizes recurrent loops and jump connections to further improve the image quality compared

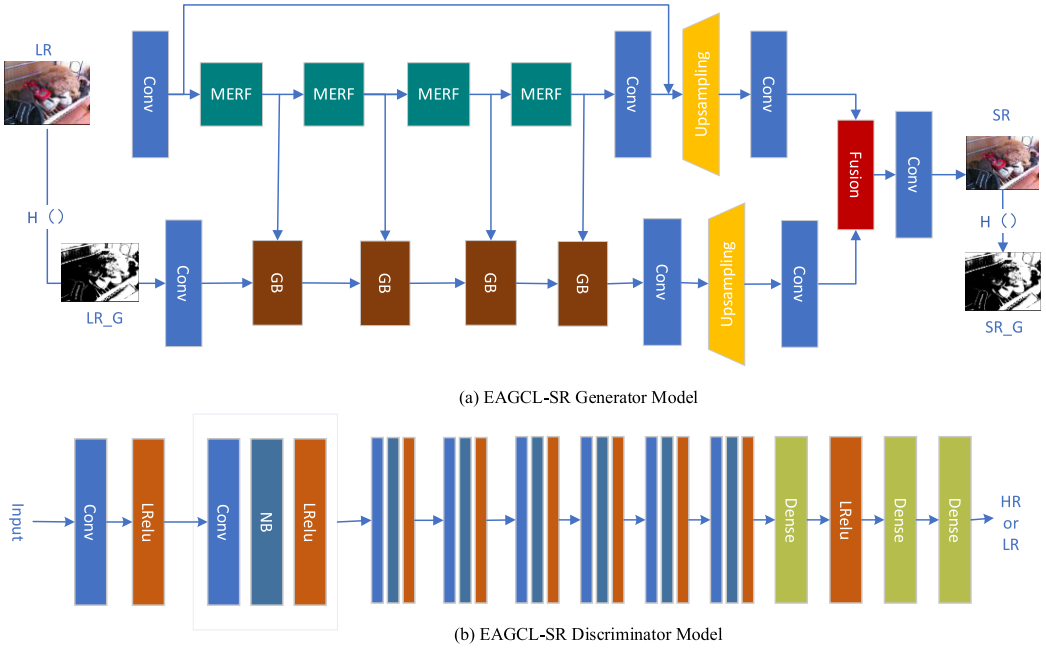
The associate editor coordinating the review of this manuscript and approving it for publication was Hengyong Yu.

with SRCNN. In 2017, Lim et al. [6] improved the residual network by removing the batch normalization layer (BN) from the residual blocks, thereby enhancing the generalization ability of the enhanced deep super-resolution network (EDSR). In 2018, Zhang et al. [7] introduced the channel attention mechanism into SR to construct residual channel attention networks (RCAN). RCAN is the first network that applies the attention mechanism to the SR problem, and the information learned by the network is more effective. In In 2020, Niu et al. [8] proposed holistic attention network (HAN) to address the problem of ignoring the correlation between different layers. This method introduces a hierarchical attention module to learn feature values through the interrelationships between multi-scale layers, and uses a channel-spatial attention module (CSAM) to learn the channel and spatial correlation of features at each layer.

In recent years, Generative Adversarial Networks (GAN) have been widely used in super-resolution reconstruction algorithms due to their ability to learn more meaningful loss functions through discriminators than those based on pixel differences. In 2014, The performance of the Generative Adversarial Network (GAN) model first proposed by Goodfellow et al. [9] in generating image data has greatly surprised researchers. Inspired by GAN [8], in 2017, Ledig et al. [10] applied it to the field of image super-resolution reconstruction and proposed the Super-Resolution Using a Generative Adversarial Network (SRGAN). The model combines perceptual loss and adversarial loss to recover the texture details of the image. Wang et al. [11] improved the network architecture, adversarial loss and perceived loss on the basis of SRGAN, and proposed an enhanced super-resolution generative Adversarial network (ESRGAN), which uses residual-dense blocks to replace residuals in the original generator and uses relative discriminator to further improve the quality of reconstructed images. In 2018, Luo et al. [12] proposed a new framework for Bi-GANs-ST super-resolution generative adversarial networks by introducing two complementary branches of generative adversarial networks. The framework uses a combination of pixel loss, perceptual loss, and adversarial loss for training, ultimately achieving a balance between objective image evaluation metrics and subjective perceptual visual effects. In 2019, Zhang et al. [13] proposed the Rank SRGAN model by incorporating content ranking into the SRGAN model framework, which uses content ranking loss to optimize the quality of generated images. In 2020, Prajapati et al. [14] used GAN for unsupervised learning of the SR algorithm and introduced a new objective learning function based on mean opinion score. Ma et al. [15] proposed a super-resolution generative adversarial network based on SPSR, which establishes the gradient feature mapping relationship between low-resolution and high-resolution images by adding a new gradient branch, introduces a gradient loss to better maintain the geometric structure of the reconstructed image, and incorporates a combination of minimum absolute value loss, perceptual loss, adversarial loss, these operations effectively preserve the overall edge structure features of the image. Rakotonirina et al. [16] proposed a super-resolution reconstruction method based on ESRGAN+, which improved the generative network of ESRGAN by adding residual skip connections of additional levels in its dense blocks, thereby enhancing the super-resolution image generation capability of the network. In 2021, Chen et al. [17] integrated the hierarchical feature extraction module into the SRGAN model framework and proposed the HSRGAN model, which extracts image features at multiple scales by hierarchically guiding the reconstruction, thus enhancing the visual fidelity of super-resolution reconstructed images. Zhang et al. [18] designed a more complex but practical degradation model for various degradation problems that could not cover real images. The model consists of random shuffling fuzzy, downsampling and noise degradation. It can help to significantly improve the practicability of deep super-resolvers, providing a powerful alternative solution for real SISR applications. In 2022, Liang et al. [19] proposed the LDL model for the problem of artifacts in images, which determines the regions, penalizes the image generation details, retains the useful textures, reduces the artifacts and makes the image more realistic. Li et al. [20] argued that the processing of images with a single loss would produce artifacts as well as part of the information would be too smooth, and therefore proposed a one-to-many supervised Beby-GAN. In 2023, Yoo et al. [21] combined CNN and Transformer and proposed a cross-scale marker attention module, allowing transform branches to efficiently exploit informative relationships between markers at different scales. In 2024, Lee et al. [22] performed meta-learning from the information contained in the distribution of the image, which greatly improved adaptation speed to new images as well as performance in kernel estimation and image fidelity. Although many scholars have achieved some results in the field of single-image super-resolution reconstruction, they often face the problems of information loss and blurred image edges in the reconstruction process. To address this problem,this paper proposes an image super-resolution reconstruction model based on enhanced attention and gradient correlation loss. The main contributions of this paper are itemized as follows:

An enhanced attention mechanism module is designed. The feature extraction capability of the generator is enhanced by reducing the number of channels, introducing stride connections and pooling layers to reduce the spatial dimensions of the network, as well as using convolutional groups to provide more variations and feature combinations.

A gradient correlation loss function is proposed. This loss function improves the visual effect of the reconstructed image by maximizing the correlation between the gradient of the generated image and the gradient of the original image, which makes the generated image more realistic and maintains a consistent edge structure.



(a) EAGCL-SR Generator Model

(b) EAGCL-SR Discriminator Model

FIGURE 1. Here is the EAGCL-SR model, where (a) is the generator model and (b) is the discriminator model.

II. MODEL

Based on the SPSR model proposed in literature [15], this section proposes a super-resolution reconstruction model based on enhanced attention mechanism and gradient correlation loss (EAGCL-SR),and the overall architecture is shown in Fig.1. Fig.1(a) shows the generator model of EAGCL-SR, which consists of two parts: one is the reconstruction parts of EAGCL-SR generator based on EA-RRDB, and the other is the reconstruction parts of EAGCL-SR generator based on gradient map. The latter adopts the gradient branch part mentioned in reference [15], and this section focuses on introducing the former. The features obtained from the two parts are fused by a fusion block, and then reconstructed by a convolutional layer to obtain the reconstructed image. Finally, the gradient map is obtained by gradient extraction. Fig.1(b) shows the discriminator model of EAGCL-SR,which adopts the relative discriminator design idea proposed in ESRGAN [11]. Conv denotes the regular convolutional layer, LRelu denotes the Leaky ReLU activation function, BN denotes the batch normalization layer, and Dense denotes the fully connected layer.

A. RECONSTRUCTION PART OF EAGCL-SR GENERATOR BASED ON EA-RRDB

In this section, the generator reconstruction model of EAGCL-SR based on EA-RRDB is given. Firstly, a multilevel residual dense connection module EA-RRDB based on enhanced attention mechanism is proposed and five EA-RRDBs are combined to obtain the MERF (Multi EA-RRDB Fusion) module. Fig.1(a) gives the framework of the reconstruction part of the EAGCL-SR generator based on EA-RRDB, which mainly performs reconstruction operations on low-resolution images (LR). Firstly, LR is input into the convolutional layer for shallow feature extraction. Then, the extracted features are passed to the MERF. Next, the features output from the MERF are passed to the next MERF while also passing them to the generator to reconstruct another part of the Gradient Block (GB), and so on. The gradient block (GB) can be any basic block which can extract higher level features, and the GB with 3×3 convolution kernel is used in this experiment. After the last MERF is executed, the features required for the reconstruction of this part of the generator are obtained through convolution and upsampling operations in turn.

1) ENHANCED ATTENTION MECHANISM

When extracting features from images, there is often a problem of losing details. To solve this problem, attention mechanism is integrated into the reconstruction model, which focuses on the details of the image and makes them less likely to be lost. However, ordinary attention mechanisms have the problem of focusing only on features in certain regions of the image and ignoring other key details, resulting in some important details or features being ignored or blurred, so that there is still the problem of detail loss.

In order to solve the above problems, this paper proposes an Enhanced Attention Block (EA), which aims to further solve the problem of detail loss in the process of image feature extraction. The module enables the EAGCL-SR to focus on feature-rich regions and extract more representative features,

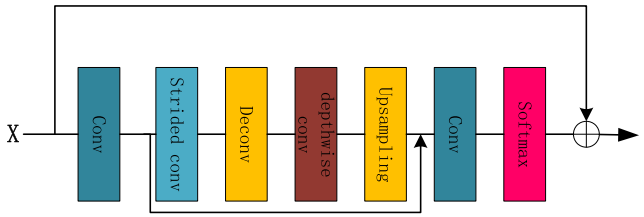


FIGURE 2. The structure of enhanced attention block module.

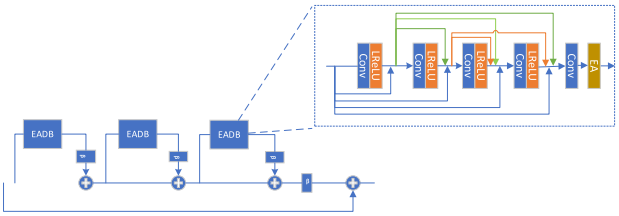


FIGURE 3. The structure of EA-RRDB module.

thus improving the reconstruction of the image by enhancing the detail information.

When designing the EA module, two aspects need to be considered. Firstly, the EA module needs to be inserted into multiple modules of the generator, so it should be designed as a lightweight module. Secondly, the EA module requires a large receptive field to better accomplish the task of image super-resolution reconstruction [23]. The EA module is designed as shown in Fig.2. Firstly, the feature x is input into a convolutional layer (Conv) which uses a 1×1 convolutional kernel to reduce the channel dimension. Then, a stride convolution (Strided conv) with a step size of 2 is used to expand the receptive field, and then the input features are amplified using a deconvolution (Deconv) to obtain richer high-frequency information. This combination of stride convolution and deconvolution can quickly reduce the spatial dimension of the network. Next, depthwise convolutions of 1×1, 3×3, and 5×5 are used to combine features in different ways to enhance the expressive power of the model. The spatial dimension is then recovered using an upsampling layer and the channel dimension is recovered using a 1×1 convolutional layer (Conv). Finally, the Softmax layer was used to obtain the deep features of the image, which were fused with the initial feature x to obtain the final feature. The module can effectively retain the detailed information of the image thus improving the effective transfer of features and the stability of network training.

2) EA-RRDB MODULE

In this section, the EA proposed above is fused with Residual-in-Residual Dense Block (RRDB) of the SPSR model proposed in [15] to obtain the Residual-Residual Dense Block (EA-RRDB) based on the Enhanced Attention Mechanism, which is shown in Fig.3. Each EA-RRDB module consists of three Enhanced Attention Mechanism-based Dense Connection Blocks (EADBs). Each EADB block consists of multiple densely connected residual blocks, and the input of each residual block includes the input and output of the previous level residual block. This densely connected approach helps to capture detailed information in the image and has more chances to pass gradients and can mitigate the problem of gradient vanishing. Based on this, the introduction of the enhanced attention mechanism allows the EADB to focus on important image regions and enhance feature extraction from these regions. This design enables the EA-RRDB module to extract effective features, which helps to improve image quality and detail retention during reconstruction.

B. LOSS FUNCTION

In the model of this paper, the classical loss function and gradient correlation loss function are used. The classical loss function includes pixel-based mean absolute error loss (MAE), perceptual loss, and adversarial loss. A comprehensive loss function is formed by a weighted summation of these four loss functions.

1) MAE LOSS FUNCTION

The MAE (Mean Absolute Error) loss function calculates the absolute value of the difference between the predicted value and the true value of each sample and then takes the average of the absolute differences of all samples as the loss. Specifically as shown in Equation (1):

$$l_{MAE} = \frac{1}{m}\sum_{i=1}^{m} |G(I_i^{LR}) - I_i^{HR}| \tag{1}$$

where l_{MAE} denotes the average absolute error loss function, m is the number of iterations, I_i^{HR} is the distribution of the ith real image and $G(I_i^{LR})$ is the distribution of the i-th high-resolution image generated by the generator.

2) PERCEPTUAL LOSS

The neural network is capable of extracting high-level features of images by training on large-scale datasets. Thus, the perceptual loss can calculate the difference between the two images through the pre-trained neural network, which is usually calculated by passing the input image and the target image through the pre-trained neural network separately to get their feature representations in the network. These feature representations are then used as the input of the loss function to calculate the Euclidean distance or Manhattan distance between them. Specifically, this can be expressed by the following Equation (2):

$$l_{Per} = \frac{1}{N}\sum_{i=1}^{N} (F_i(x) - F_i(y))^2 \tag{2}$$

where x is the input image, y is the target image, $F_i(x)$ and $F_i(y)$ denote their feature representations of i-th layer, respectively, in a pre-trained neural network and N denotes the number of feature layers. By minimizing the perceptual loss, the generator is forced to produce an image that is closer to the target image in terms of the feature space, which in turn improves the quality of the generated image.

3) ADVERSARIAL LOSS

A binary cross entropy loss function is used to measure the probability that an image generated by the generator is correctly discriminated as a real image or a fake image. This is shown in Equation (3):

$$lgan = -log\left(D\left(x\right)\right) - log\left(1\text{-}D\left(G\left(z\right)\right)\right) \tag{3}$$

where x denotes the real sample, $D\left(x\right)$ denotes the judgement result of the discriminator on the real sample, $G\left(z\right)$ denotes the fake sample generated by the generator, and $D\left(G\left(z\right)\right)$ denotes the judgement result of the discriminator on the fake sample. The goal of the discriminator is to minimize the adversarial loss function so that the judgement result for real samples is close to 1 and the judgement result for fake samples is close to 0.

4) GRADIENT CORRELATION LOSS FUNCTION

It has been shown that the use of classical loss functions in the training process can easily lead to the problem of over-smoothing of the reconstructed image, i.e., it is difficult to reconstruct the edges of low-resolution images with the trained model to achieve the desired results. The main reason of this problem is that the classical loss function (e.g. mean square error) focuses more on minimizing the global pixel-level differences in the optimization process while ignoring the importance of image details and edges. In this situation, the model tends to generate excessively smooth images that lack sharp edge features. To solve this problem, a gradient correlation loss function is proposed in this paper.

The concern of the proposed gradient correlation loss function is to ensure that the generated image is aligned with the original image in the gradient direction to maintain edge and texture consistency. The performance of the gradient correlation loss function is further enhanced by maximizing the correlation between the gradient of the generated image and the gradient of the original image. This loss function enables stronger constraints on the super-resolution model, which effectively maintains the structural information of the image and helps the generated high-resolution image to be more realistic and structurally consistent in terms of details and edges, thus improving the quality and visual effect of the reconstructed image. The specific calculation of the gradient correlation loss function is shown in Equation (4):

$$r_{LG} = \frac{cov\left(H\left(G\left(I^{LR}\right)\right), H(I^{HR})\right)}{\sqrt{\sigma(H\left(G\left(I^{LR}\right)\right))} * \sqrt{\sigma(H\left(I^{HR}\right))}} \tag{4}$$

In Equation (4), r_{LG} represents the correlation coefficient. The calculation result is that "1" indicates complete positive correlation (with the best loss function performance), -1 indicates negative correlation, and 0 indicates no correlation. $cov(\cdot)$ denotes the calculation of the gradient covariance of the generated image and the reconstructed image; $H\left(G\left(I^{LR}\right)\right)$ denotes the calculation of the gradient value of the reconstructed low-resolution image, and $H\left(I^{HR}\right)$ denotes the calculation of the gradient value of the high-resolution image. $\sigma(\cdot)$ denotes the calculation of the gradient covariance of the generated image and the reconstructed image.

The three classical loss functions proposed in Equations (1), (2) and (3) and the gradient correlation loss function proposed in Equation (4) are fused to obtain the final loss function as shown in Equation (5):

$$L_G = l_{Per} + \alpha l_{MAE} + \beta lgan + \eta r_{LG} \tag{5}$$

where $lgan$ denotes the adversarial loss, l_{Per} denotes the perceptual loss, and r_{LG} denotes the gradient correlation loss. α and β are the weights of the reconstructed image loss and η is the weight of the gradient correlation loss. Let α =0.01, $\beta = 0.005$, and $\eta = 0.005$.

TABLE 1. Commonly used super-resolution reconstruction datasets.

Dataset	Number	Scene content	Advantages	Disadvantages
Set5	5	Nature People	Small datasets Convenient for quick testing	Lack of diversity Small size
Set14	14	People Animals Landscape Nature	Medium data set with diverse scenario content	Dataset coverage is limited
BSDS 100	100	Cities Architecture Landscape nature	Scenario types are rich Diversity of degradation issues	Dataset is too complex

III. EXPERIMENT

A. EXPERIMENTAL ENVIRONMENT AND DATASET

The DIV2K [24] dataset is used during the training process, which includes 800 training images, 100 validation images and 100 test images. In order to avoid overfitting during the training process, data enhancement operations such as random rotation and horizontal flipping are performed on the training images as a way to increase the diversity of the data. In order to test the model effect, three standard benchmark datasets (Set5 [25], Set14 [26] and BSDS100 [27]) are used as the test sets, and the specific information is shown in TABLE 1:

B. EVALUATION INDICATORS

In this paper, PSNR, SSIM [28] and LPIPS [29] are used to evaluate the experimental results. As in Equation (6), PSNR can be evaluated by the grey level difference between the corresponding pixel points of two images. The higher the value of PSNR, the smaller the distortion.

$$P_{SNR}(X, Y) = 10 * lg \frac{255^2 * w * h * c}{\sum_{m=1}^{w} \sum_{n=1}^{h} \sum_{z=1}^{c} [X(m,n) - Y(m,n)]^2} \tag{6}$$

where X denotes the original high-resolution image; Y denotes the reconstructed image of generator; c denotes the

TABLE 2. Comparison of PSNR, SSIM, and LPIPS values for 4× reconstruction results of various algorithms.

Dataset	Metric	bicubic	SRGAN	ESRGAN	ESRGAN+	PDM-GAN	Beby-GAN	SPSR	Ours
Set5	PSNR	26.69	26.69	26.50	25.88	25.62	27.82	28.44	28.89
	SSIM	0.7736	0.7813	0.7565	0.7511	0.7304	0.8004	0.8241	0.8386
	LPIPS	0.3644	0.1305	0.1080	0.1178	0.1075	0.0875	0.0870	0.0802
Set14	PSNR	26.08	25.88	25.52	25.01	23.69	24.69	24.75	24.89
	SSIM	0.7467	0.7480	0.7175	0.7159	0.6716	0.7016	0.6960	0.7025
	LPIPS	0.3870	0.1421	0.1254	0.1362	0.1398	0.1094	0.1062	0.0972
BSDS100	PSNR	22.65	22.67	23.33	23.54	23.84	24.13	24.21	24.58
	SSIM	0.6014	0.6363	0.6133	0.6172	0.6235	0.6355	0.6554	0.6584
	LPIPS	0.4452	0.1636	0.1436	0.1434	0.1433	0.1274	0.1197	0.1125

number of channels of the image; w and h denote the width and height of the image, respectively; m denotes the m-th pixel on the width of the image; n denotes the n-th pixel on the height of the image; and z denotes the z-th channel of the three primary color channels.

As shown in Equation (7), SSIM evaluates the similarity of two images from brightness, contrast and structure. The SSIM value is close to 1, which indicates that the reconstructed image is closer to the structure of the original image and generates better results.

$$S_{SIM}(X, Y) = \frac{(2\mu_X\mu_Y + C_1)(2\sigma_{XY} + C_2)}{(\mu_X^2 + \mu_Y^2 + C_1)(\sigma_X^2 + \sigma_Y^2 + C_2)} \quad (7)$$

where μ_X denotes the mean value of X and μ_Y denotes the mean value of Y; μ_X^2 denotes the average value of X, μ_Y^2 denotes the average value of Y, and σ_{XY} denotes the covariance of X and Y; $C1$ and $C2$ are constants.

As in Equation (8), the LPIPS measures the difference between two images. LPIPS learns the reverse mapping of reconstructed images to real images, calculates the perceptual similarity between them, which can be used to evaluate the difference between two images. The lower the LPIPS value, the more similar the two images are, and vice versa. The lower the value of LPIPS, the more similar the two images are.

$$d(x, x_0) = \sum_l \frac{1}{H_l W_l} \sum_{h,\omega} ||\omega \odot (y^l_{h\omega} - y^l_{0h\omega})||_2^2 \quad (8)$$

where y^l, $y^l_0 \in R^{H_l \times W_l \times C_l}$ denotes that the inputs are sent to the neural network for feature extraction, and the outputs of each layer are normalized after activation, and ω denotes the layer of the network.

C. TRAINING DETAILS

In order to ensure the fairness of the experimental results, all the experiments in this paper use a 4-fold scale factor and are conducted in the same hardware environment. The hard ware device parameters used in this paper are: CPU: Intel(R) Xeon(R) CPU E5-2680 v4; RAM: 12G; number of cores: 28; GPU: 3080 Ti-12G. In the Linux operating system environment, we use PyTorch framework with version 1.13.1 to write code and utilize Cuda11.3 for accelerated learning.

In the training process, the parameter batch_size is set to 4 and the size of the cropped high-resolution image is set to 128 × 128. The training process is divided into two stages: first, the enhanced attention mechanism is incorporated into the RRDB module of the model to get a pre-trained model after training. Then, the obtained pre-trained model is used to initialize the generator and the generator is trained using a loss function. During the training process, the learning rate is set to 1e-4 and decayed to 0.5 times the original learning rate after every 5e4 iterations. The decay strategy of learning rate helps the model converge better during the training process. The above experimental setup ensures comparability and fairness of the experiments.

D. COMPARATIVE EXPERIMENTS

1) QUANTITATIVE COMPARISON

In the case of an amplification factor of 4, the model proposed in this paper (Ours) is compared with Bicubic, SRGAN [10], ESRGAN [11], ESRGAN+ [16], SPSR [15], Beby GAN [20] and PDM-GAN [30]. The experimental results are shown in TABLE 2. From TABLE 2, it can be seen that the PSNR values of our method on Set5, Set14, and BSDS100 datasets has been improved by 0.45dB, 0.14dB, and 0.37dB compared with SPSR, respectively. This indicates that our algorithm performs better in PSNR and the details of the generated image are clearer. On the Set5, Set14, and BSDS100 datasets, the SSIM values of our method are increased by 0.0145, 0.0065, and 0.0030 compared with SPSR, respectively, which indicates that our method performs better in maintaining image structural similarity. On the Set5, Set14, and BSDS100 datasets, the LPIPS values of our method are decreased by 0.0068, 0.0090, and 0.0072 compared with SPSR, respectively. In summary, the method proposed in this paper performs well in super-resolution image quality and is superior to other methods by comprehensively evaluating the PSNR, SSIM, and LPIPS indicators.

2) QUALITATIVE COMPARISON

In order to better highlightodel proposed in this paper we compared it with Bicubic, SRGAN, ESRGAN, ESRGAN+, SPSR, Beby GAN, and PDM-GAN. Figures 4-6 show some of the image reconstruction results.

From the visual perspective, Fig.4(a) shows a realistic image of the butterfly's back and wings, while Fig.4(b)

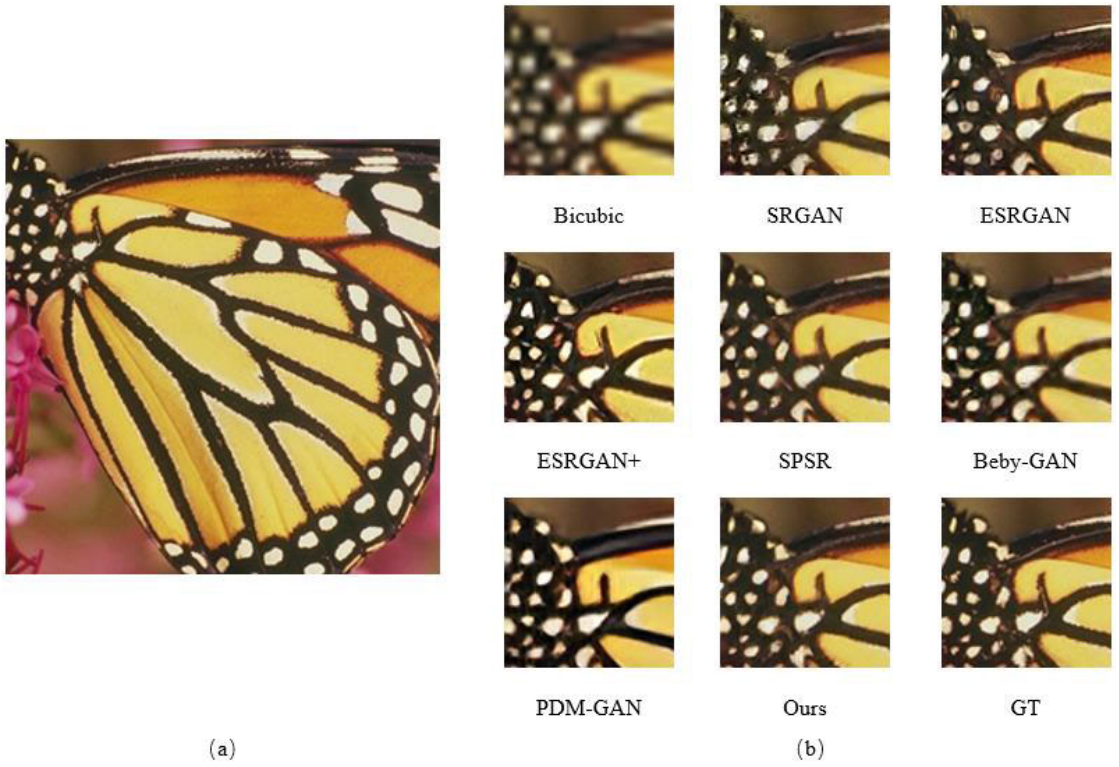


FIGURE 4. The reconstruction results of each algorithm at 4× scaling factor. Image "butterfly" from Set5. where (a) is the original image, and (b) is the local effect of the reconstructed image.

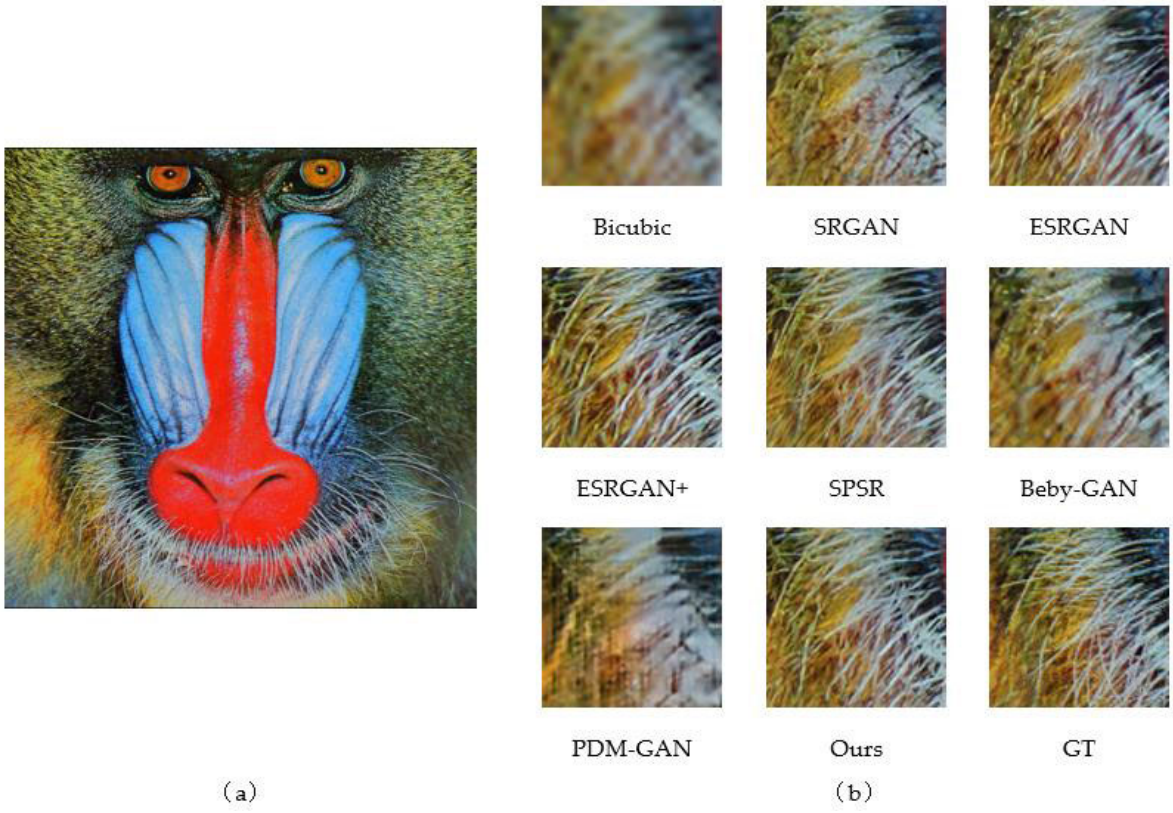


FIGURE 5. The reconstruction results of each algorithm at 4× scaling factor. Image "baboon" from Set14. where (a) is the original image, and (b) is the local effect of the reconstructed image.

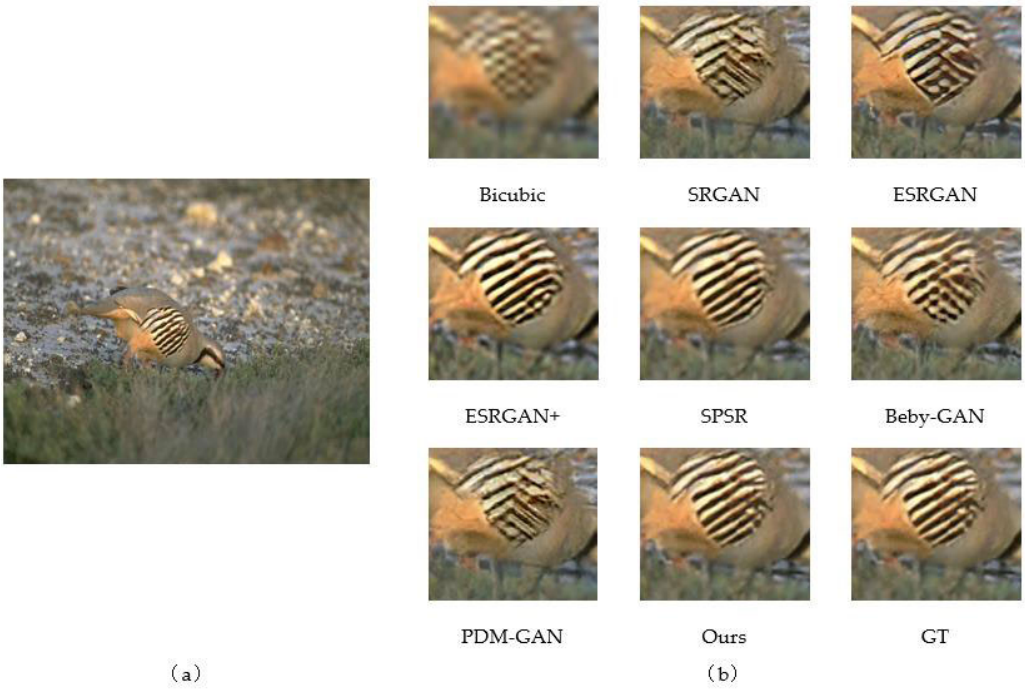


FIGURE 6. The reconstruction results of each algorithm at 4× scaling factor.Image "8023" from BSDS100. where (a) is the original image, and (b) is the local effect of the reconstructed image.

shows a local image of the wing root obtained using various reconstruction methods. As shown in Fig.4(b), the images generated by Bicubic are blurry and unclear, while the images generated by SRGAN, ESRGAN, and ESRGAN+ suffer from detail loss and severe sharpening. The reconstruction effect of SPSR has been improved to some extent, but the reconstruction effect in small areas is not good. The reconstructed image of Beby GAN has blurred wing lines and still suffers from detail loss. After analysis, it can be seen that the image reconstructed using the mode proposed in this article is closer to the real image (GT). Fig.5(a) shows a real image of Baboon, and Fig.5(b) shows the pattern of Baboon's left beard. By observing the texture of the beard, it can be seen that the image generated by Bicubic is blurry, and the reconstruction effects of SRGAN, Beby GAN, and PDM-GAN have severe detail loss and sharpening. The reconstruction effect of SPSR is relatively good, but there is still some detail loss compared with our proposed method. Therefore, the image reconstructed using our proposed method is closer to the real image (GT). Fig.6(a) is the actual picture of the bird, and Fig.6(b) is the wing pattern of the bird. By observing the wing details, it can be seen that the reconstructed images generated by Bicubic, SRGAN, ESRGAN, Beby-GAN and PDM-GAN models have serious blurring and sharpening problems. Relatively speaking, the reconstruction effect of the SPSR model is clear, but it is blurred near the first texture of the wing. Relatively speaking, the reconstruction effect of SPSR is relatively clear, but it can be seen to be quite blurry near the first texture of the wings. In contrast, our proposed method can reconstruct patterns that are closer to real images

FIGURE 7. Comparison of ablation experiments. Image "flowers" from set14.

(GT). In summary, compared with other comparison algorithms, the visual effect of reconstructing images using our proposed method can be closer to the details and textures of real images (GT).

E. ABLATION EXPERIMENTS

In order to verify the necessity of each part of the proposed model, the corresponding ablation experiments are conducted in this section. Given that the model proposed in this paper is based on the SPSR, two algorithms are designed for comparison. One algorithm (EAGCL-SR no L) did not use gradient correlation loss during training, but applied the EA network module in the network. Another algorithm (EAGCL-SR) is the complete model proposed in this article, which combines enhanced attention machines with gradient correlation loss. The experimental results are shown in Table 3.

From TABLE 3, it can be seen that compared with the SPSR, the performance of network with the EA module will have an improvement over the original network. In addition, it can be seen that EA performs well on PSNR. On this basis, the gradient correlation loss is added into the model, and the

TABLE 3. Comparison of different models under the ablation experiment.

Dataset	Metric	SPSR	EAGCL-SR no L	EAGCL-SR
Set5	PSNR	28.44	28.64	28.89
	SSIM	0.8241	0.8124	0.8386
	LPIPS	0.0870	0.0862	0.0802
Set14	PSNRS	24.75	24.72	24.89
	SIM	0.6960	0.6916	0.7025
	LPIPS	0.1062	0.1047	0.0972
BSDS 100	PSNR	24.21	24.43	24.58
	SSIM	0.6554	0.6347	0.6584
	LPIPS	0.1197	0.1152	0.1125

proposed EAGCL-SR model is obtained. The experimental results show that EAGCL-SR can effectively improve the quality of image reconstruction. Therefore, the effectiveness of the method proposed in this paper is verified. As shown in Fig.7, the reconstructed image obtained using the algorithm proposed in this paper has fewer blurry areas, resulting in a clearer pattern of the sepals.

IV. SUMMARY

For SR tasks with high visual quality requirements, this paper proposes an image super-resolution reconstruction model based on enhanced attention mechanism and gradient correlation loss. The purpose of this model is to solve the problems of information loss and edge blurring in image super-resolution reconstruction. The enhanced attention mechanisms is incorporated into the model to effectively focuses on important details in low resolution images, which can improve the quality of reconstructed images. At the same time, the gradient correlation loss function is used to make the generated image more realistic and maintain the consistency of the edge structure. The experimental results show that the proposed model achieves certain improvement in PSNR, SSIM and LPIPS, thus verifying the effectiveness of the model. In future work, more effective model architectures and training strategies will be explored to improve the quality of reconstruction results and reduce the computational cost.

REFERENCES

[1] S. Zhu, B. Zeng, and L. Zeng, "Image interpolation based on non-local geometric similarities," *IEEE Trans. Multimedia*, vol. 18, no. 9, pp. 1707–1719, Oct. 2016.
[2] V. Papyan and M. Elad, "Multi-scale patch-based image restoration," *IEEE Trans. Image Process.*, vol. 25, no. 1, pp. 249–261, Jan. 2016.
[3] C. Dong, C. Loy, and X. Tang, "Accelerating the super-resolution convolutional neural network," in *Computer Vision—ECCV 2016*. Cham, Switzerland: Springer, 2016, pp. 391–407.
[4] 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.
[5] J. Kim, J. Lee, and K. Lee, "Deeply-recursive convolutional network for image super-resolution," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 1–20.
[6] B. Lim, S. Son, and H. Kim, "Enhanced deep residual networks for single image super-resolution," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, 2017, pp. 136–144.
[7] Y. Zhang, Y. Tian, and Y. Kong, "Residual dense network for image super-resolution," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Oct. 2018, pp. 1–14.
[8] B. Niu, W. Wen, and W. Ren, "Single image super-resolution via a holistic attention network," in *Computer Vision—CCV 2020*. Cham, Switzerland: Springer, 2020, pp. 191–207.
[9] I. Goodfellow, J. Pouget-Abadie, and M. Mirza, "Generative adversarial networks," *Commun. ACM*, vol. 63, pp. 139–144, Aug. 2020.
[10] C. Ledig, L. Theis, F. Huszar, J. Caballero, A. Cunningham, and A. Acosta, "Photo-realistic single image super-resolution using a generative adversarial network," *IEEE Computer Society*, vol. 1, no. 1, pp. 1–15, Oct. 2016.
[11] 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.*, 2018, pp. 1–20.
[12] X. Luo, R. Chen, Y. Xie, Y. Qu, and C. Li, "Bi-GANs-ST for perceptual image super-resolution," in *Proc. ECCV*, 2019, pp. 1–18, doi: 10.1007/978-3-030-11021-5_2.
[13] X. Cecilia Zhang, Q. Chen, R. Ng, and V. Koltun, "Zoom to learn, learn to zoom," 2019, *arXiv:1905.05169*.
[14] K. Prajapati, V. Chudasama, H. Patel, K. Upla, R. Ramachandra, K. Raja, and C. Busch, "Unsupervised single image super-resolution network (USISResNet) for real-world data using generative adversarial network," in *Proc. IEEE/CVF Conf. Comput. Vis.*, Jun. 2020, pp. 1–26.
[15] C. Ma, Y. Rao, Y. Cheng, C. Chen, J. Lu, and J. Zhou, "Structure-preserving super resolution with gradient guidance," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 7766–7775.
[16] ICASSP N. C. Rakotonirina and A. Rasoanaivo, "ESRGAN+: Further improving enhanced super-resolution generative adversarial network," in *Proc. ICASSP*, 2020, p. 20.
[17] W. Chen, Y. Ma, and X. Liu, "Hierarchical generative adversarial networks for single image super-resolution," in *Proc. IEEE/CVF Winter Conf. Appl.*, Sep. 2021, pp. 1–18.
[18] K. Zhang, J. Liang, and L. Van Gool, "Designing a practical degradation model for deep blind image super-resolution," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, Oct. 2021, pp. 1–19.
[19] J. Liang, H. Zeng, and L. Zhang, "Supplementary material to 'details or artifacts: A locally discriminative learning approach to realistic image super-resolution," in *Proc. IEEE/CVF*, 2022, pp. 5657–5666.
[20] W. Li, K. Zhou, L. Qi, L. Lu, and J. Lu, "Best-buddy GANs for highly detailed image super-resolution," in *Proc. AAAI Conf. Artif. Intell.*, 2022, vol. 36, no. 2, pp. 1412–1420.
[21] J. Yoo, T. Kim, and S. Lee, "Rich CNN-transformer feature aggregation networks for super-resolution," 2022, *arXiv:2203.07682*.
[22] R. Lee, R. Li, and S. Venieris, "Meta-learned kernel for blind super-resolution kernel estimation," in *Proc. IEEE/CVF*, Jul. 2024, pp. 1–27.
[23] J. Liu, W. Zhang, and Y. Tang, "Residual feature aggregation network for image super-resolution," in *Proc. IEEE/CVF*, Oct. 2020, pp. 1–29.
[24] 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)*, Jul. 2017, pp. 1122–1131.
[25] M. Bevilacqua, A. Roumy, C. Guillemot, and M.-L. Alberi Morel, "Neighbor embedding based single-image super-resolution using semi-nonnegative matrix factorization," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Mar. 2012, pp. 1289–1292.
[26] Y. Yuan, S. Liu, J. Zhang, Y. Zhang, C. Dong, and L. Lin, "Unsupervised image super-resolution using cycle-in-cycle generative adversarial networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2018, pp. 814–81409.
[27] 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. Vision. ICCV*, vol. 2, Jul. 2001, pp. 416–423.
[28] 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.
[29] R. Zhang, P. Isola, A. A. Efros, E. Shechtman, and O. Wang, "The unreasonable effectiveness of deep features as a perceptual metric," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 586–595.
[30] Z. Luo, Y. Huang, and S. Li, "Learning the degradation distribution for blind image super-resolution," in *Proc. CVPR*, 2022.

YANLAN SHI received the B.S. degree in software engineering from Lanzhou City University, in 2022. She is currently pursuing the degree with the School of Electronic and Information Engineering, Liaoning University of Technology. Her main research interests include computer vision and image super-resolution reconstruction.

HUAWEI YI received the B.S. degree in computer science and technology from Liaoning University of Technology, in 2003, the M.Sc. degree in communication and information system from Lanzhou University of Technology, Lanzhou, China, in 2007, and the Ph.D. degree in computer science and technology from Yanshan University, Qinhuangdao, China, in 2017. She is currently an Associate Professor with the School of Electronics and Information Engineering, Liaoning University of Technology. Her main research interests include recommendation systems, trusted computing, and information security.

XIN ZHANG received the bachelor's degree in software engineering from Liaoning University of Technology, in 2023, where he is currently pursuing the degree with the School of Electronic and Information Engineering. His main research interests include computer vision and image super-resolution reconstruction.

LU XU received the bachelor's, master's, and Ph.D. degrees from Shenyang University of Technology. He is currently with the IT and Products Management Department, Agricultural Bank of China.

JIE LAN received the B.S. degree in applied mathematics from Jilin Agricultural University, Changchun, China, in 2005, and the M.S. degree in control theory and control engineering from Liaoning University of Technology, Jinzhou, China, in 2011. She is currently pursuing the Ph.D. degree in agricultural electrification and automation with Shenyang Agricultural University, Shenyang, China. She is a Lecturer with the College of Science, Liaoning University of Technology. Her research interests include adaptive fuzzy control, nonlinear control, neural network control, and multi-agent and swarm intelligence. She is a member of the Chinese Association of Automation.