Super-resolution reconstruction based on Gaussian transform and attention mechanism

Shuilong Zou[1], Mengmu Ruan[2], Xishun Zhu[1] and Wenfang Nie[1,3]

[1] Nanchang Normal College of Applied Technology, School of Electronic and Information Engineering, Nanchang, Jiangxi, China
[2] Nanchang Institute of Science & Technology, School of Wealth Management, Nanchang, Jiangxi, China
[3] Current Affiliation: School of Economics and Management, Jiangxi Manufacturing Polytechnic College, Nanchang, China

ABSTRACT

Image super-resolution reconstruction can reconstruct low resolution blurred images in the same scene into high-resolution images. Combined with multi-scale Gaussian difference transform, attention mechanism and feedback mechanism are introduced to construct a new super-resolution reconstruction network. Three improvements are made. Firstly, its multi-scale Gaussian difference transform can strengthen the details of low resolution blurred images. Secondly, it introduces the attention mechanism and increases the network depth to better express the high-frequency features. Finally, pixel loss function and texture loss function are used together, focusing on the learning of structure and texture respectively. The experimental results show that this method is superior to the existing methods in quantitative and qualitative indexes, and promotes the recovery of high-frequency detail information.

INTRODUCTION

Image super-resolution (SR) reconstruction is a basic task of image processing and is widely used in image compression (*Chen et al., 2021a*; *Shi, Li & Jiahuan, 2022*; *Chen et al., 2021b*), medical imaging (*Jiao et al., 2020*), and other fields. It is a research hotspot in the field of image processing. Current SR methods can be divided into two categories: reconstruction-based and learning-based methods.

Reconstruction-based methods can be divided into two categories: frequency domain methods and spatial domain methods. Image reconstruction based on the frequency domain (transform domain) is the indirect processing of images directly in the transform domain. The transformations used include Wavelet transform (*Jing & Wu, 2014*), Fourier transform (*Liu, 2018*), etc. (*Xiong et al., 2015*). These algorithms are simple in principle, have low computational complexity, and are easy to implement in hardware. Therefore, the frequency methods can only deal with global translational motion and cannot utilize the image a priori information.

The spatial domain methods mainly include iterative back-projection (IBP) (*Irani & Peleg, 1991*; *Song et al., 2010*; *Seema & Bailey, 2019*), projections onto convex sets (POCS) (*Patti & Altunbasak, 2001*; *Fan et al., 2017*; *Ma & Ren, 2020*), maximum a posteriori (MAP) probability estimation method (*Shen et al., 2007*; *Belekos, Galatsanos & Katsaggelos, 2010*; *Nascimento & Salles, 2020*) and other methods. These methods use the sub-pixel information existing between low-resolution (LR) images to provide additional information for reconstructing the images. The algorithms reconstruct well and are mainly used in scientific research, satellite remote sensing and other fields. However, these classical reconstruction methods require researchers to have a large amount of a priori knowledge and deep professional reserves. With the rise of deep learning, learning-based reconstruction methods have received wide attention because the technology does not require much a prior knowledge and the quality of reconstructed images is better than that of traditional algorithms.

Convolutional neural network (CNN) is one of the main algorithms for deep learning, and has excellent performance in areas such as image classification (*Wang, Wang & Wang, 2015*) and computer vision (*Gatys, Ecker & Bethge, 2015a*). Super-resolution convolutional neural network (SRCNN) was proposed by *Dong et al. (2014)* to apply CNN to super-resolution reconstruction. The proposed SRCNN is of milestone significance, but there are some shortcomings, such as over-reliance on contextual information of small image areas and slow convergence speed during training. In view of these shortcomings, *Dong, Loy & Tang (2016)* proposed the fast super-resolution convolutional neural network (FSRCNN). Increasing the number of network layers and using a smaller convolution kernel make the network deeper and learn more features. *Kim, Lee & Lee (2016)* proposed very deep convolutional networks for super-resolution (VDSR) models with increased number of network layers.

Deepening the network structure can bring more features, but it also tends to cause gradient disappearance or gradient explosion. ResNet (*He et al., 2016*) can effectively solve these problems and improve the expressiveness of the network. Subsequently, *Mao, Shen & Yang (2016)* applied it to the very deep residual encoder–decoder network (RED-Net). Unlike ResNet, DenseNet (*Huang et al., 2017*) connects each layer in series with other layers, which can better preserve the characteristic information of the original image. Even if it is transferred to the later layers, the image information is not easily lost, and the problem of gradient disappearance is well solved. Considering the advantages of DenseNet, *Tong et al. (2017)* proposed SRDenseNet. *Zhang et al. (2018b)* combined the residual block with the dense module to form the residual dense block (RDB). *Yu et al. (2018)* proposed the Wide Activation Image Super-Resolution (WASR) model. WASR consists of a convolution module, a residual module, and a pixel reorganization module. The pixel shuffle is realized by using the sub-pixel layer proposed by the efficient sub-pixel convolutional neural network (ESPCN) (*Shi et al., 2016*) for up sampling.

The images generated by these algorithms have a high peak signal-to-noise ratio (PSNR), but the perceptual quality is poor. To generate images for human eye perception, *Ledig et al. (2016)* proposed to realize image super-resolution by means of a Generative Adversarial Network (SRGAN). The main body of the network adopts a Generative Adversarial Network

(GAN) (*Goodfellow et al., 2014*). Although the PSNR is not the highest, the reconstructed image is more natural and clear, which is more in line with the visual effect of human eyes. *Wang et al. (2018b)* improved SRGAN and proposed Enhanced super-resolution Generative Adversarial Networks (ESRGAN). SRGAN uses activated features to calculate perceptual loss, while ESRGAN uses pre-activated features to calculate perceptual loss. *Wang, Yu & Dong (2018)* proposed SFTGAN and added a new spatial feature transform (SFT) to the model that combines effective a priori information with the neural network for end-to-end training. The reconstructed images appear visually more natural.

Most of the neural network-based SR methods do not fully utilize the information of the original LR image, resulting in unsatisfactory results. *Zhang et al. (2018a)* proposed the residual channel attention network (RCAN) by combining the channel attention mechanism with the residual block. *Zhang et al. (2018a)* also pointed out that there is a large amount of low-frequency information in low-resolution images, which can be transmitted directly to the last layer of the network through long-hop connections, allowing the network to focus on learning high-frequency information and reducing the learning burden of the network.

In addition to improving the network structure, different loss functions are used to generate different image qualities. Commonly used loss functions include pixel loss, content loss, confrontation loss (*Mao et al., 2017*), texture loss (*Gatys, Ecker & Bethge, 2015b*; *Gatys, Ecker & Bethge, 2016*), total variation loss (*Rudin, Osher & Fatemi, 1992*), and context loss (*Mechrez et al., 2018*). Since each loss function has its own emphasis, combining multiple loss functions to train the network, the image generated by the model will achieve good results in both objective evaluation and subjective visual effects. At present, most models adopt a joint training method with multiple loss functions, such as EnhanceNet proposed by *Sajjadi, Scholkopf & Hirsch (2017)*. The loss function consists of perceptual loss, confrontation loss, and texture loss, which can produce a realistic texture.

In order to generate realistic textures and natural details with high visual quality, this paper combines reconstruction methods and learning methods to build a deep learning network Garesat-Net based on Gaussian transform, residuals and channel attention. In the network, a new detail enhancement module based on Gaussian transform is constructed. The new module and the channel attention mechanism module are embedded into the residual structure to form a new residual block. In the network, the residual block is used to enhance the details of the image and send it to the subsequent layers for training. During the training process, the loss function comprises pixel loss and texture loss to limit the generation of super-resolution images. Through simulation experiments and real data testing, our algorithm can effectively utilize color image channel correlation and outperform existing algorithms in reconstruction.

The rest of the paper is organized as follows. The related work and evaluation metrics are given in the next section. Section 3 describes the workflow of the proposed model in detail. To evaluate the proposed algorithm, simulation and experimental results are given in Section 4. Finally, a brief conclusion is drawn in Section 5.

RELATED WORK AND EVALUATION INDEXES

Multi-scale Gaussian difference transform

Gaussian smoothing is a kind of digital image processing method, calculated from a two-dimensional normal distribution (Gaussian distribution) function. The Gaussian kernel used in the two-dimensional Gaussian distribution function is the product of two one-dimensional Gaussians x and y, and the standard deviation σ isusually the same in both dimensions:

$$G(x,y) = \frac{1}{2\pi\sigma^2} e^{\left(-\frac{x^2+y^2}{2\sigma^2}\right)}. \tag{1}$$

Gaussian smoothing is the convolution of an image with a two-dimensional Gaussian kernel of a certain size. Gaussian kernel is a discrete approximation of a continuous Gaussian function, which is usually obtained by discrete sampling and normalization of Gaussian surfaces. Here, normalization means that the sum of all elements of the convolution kernel is 1. The standard deviation σ determines the influence of the surrounding pixels on the current pixel. When σ increases, the influence of distant pixels on the central pixel is improved, with smoother filtering results. The Gaussian kernel is a discrete approximation of the continuous Gaussian. The more natural the window is, the better the approximation is. However, the Gaussian function is a bell-shaped curve. The farther away from the center, the smaller the value, which can be ignored if the distance is far enough. After the standard deviation is determined, the radius is 3σ, that is, the window size is $6\sigma \times 6\sigma$. Usually, the nearest odd number is taken.

The Gaussian multi-scale difference method (*Kim et al., 2015*) uses different Gaussian kernel functions to obtain different Gaussian blurred images, so as to extract different detailed images and fuse different detailed images into a whole. First, different Gaussian blurred images are obtained using different Gaussian kernel functions, and the image to be enhanced is I:

$$B_1(I) = G_1 * I, B_2(I) = G_2 * I, B_3(I) = G_3 * I \tag{2}$$

where G_k k $=1$ (, 2, 3) is the Gaussian kernel function. We extract fine detail D_1, intermediate detail D_2 and rough detail D_3, which are given by the following formula:

$$D_1(I) = I - B_1(I), D_2(I) = I - B_2(I), D_3(I) = I - B_3(I). \tag{3}$$

Merging the three-layer detail map to generate the overall detail and adding it to the original image:

$$I^* = I + (1 - \omega_1 \times \text{sgn}(D_1(I))) \times D_1(I) + \omega_2 \times D_2(I) + \omega_3 \times D_3(I). \tag{4}$$

where $\omega_1, \omega_2, \omega_3$ are the weight coefficients and sgn is a step function.

Gaussian multi-scale difference method is introduced in CNN to construct a Gaussian enhancement block, as shown in Fig. 1. The input multi-channel features are convoluted by a three-channel convolutional layer based on different Gaussian kernels, respectively, and three different Gaussian fuzzy feature maps are obtained. Three detail maps were obtained by making a circular difference with the original feature maps, and then fused
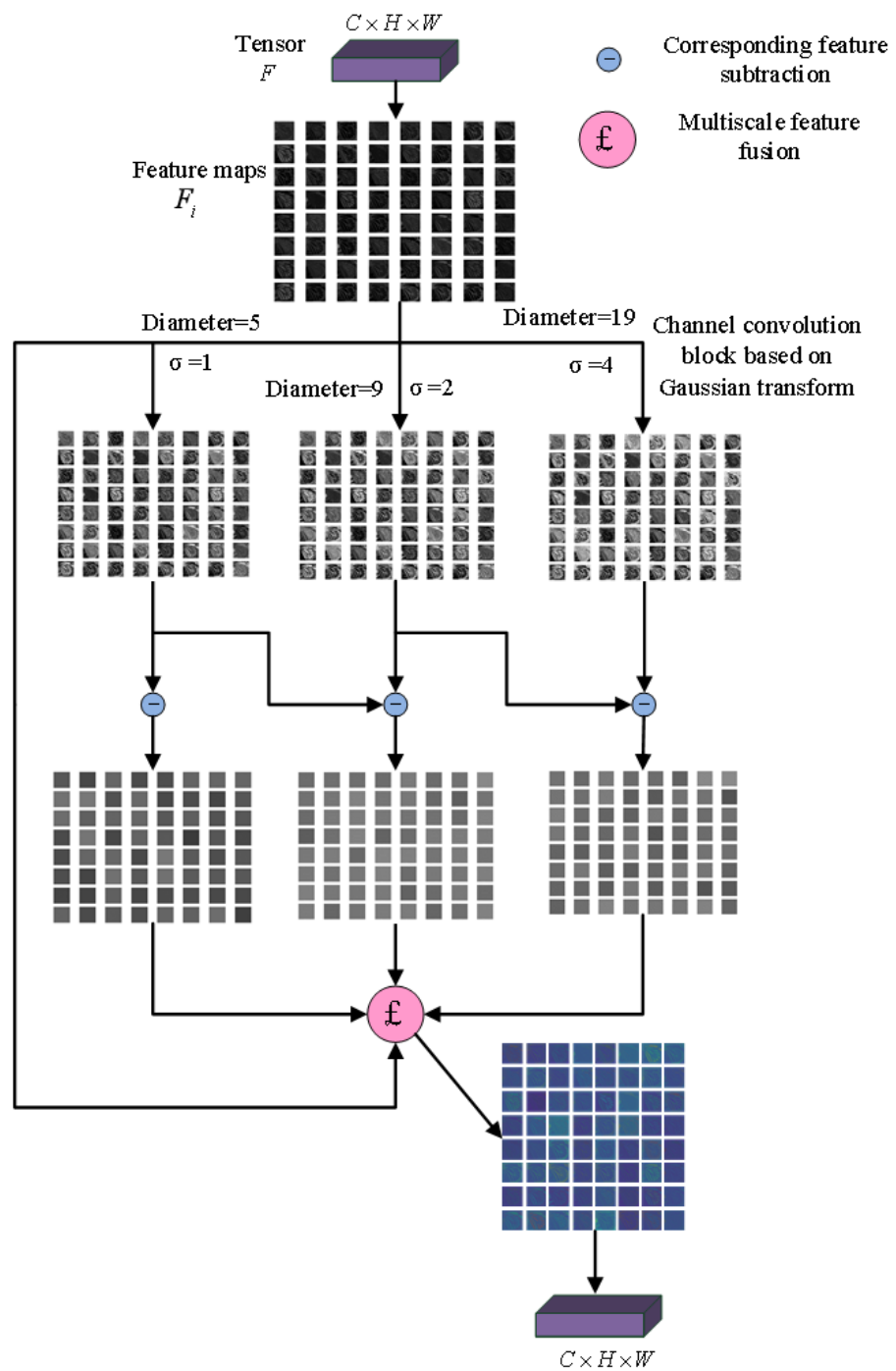
Figure 1 Gaussian enhancement block.

with the original feature map to obtain the enhanced feature map. The parameters of fusion $\omega_1, \omega_2, \omega_3$ were fixed at 0.5, 0.5 and 0.25.

The attention mechanism

Humans can naturally and effectively find significant regions in complex scenes. Inspired by this observation, an attention mechanism was introduced into computer vision to imitate this aspect of the human visual system. This attention mechanism can be viewed as a dynamic weight adjustment process based on the input image features, which is widely used in various types of deep learning tasks such as natural language processing, image recognition and speech recognition. It is one of the core technologies that deserve attention and deeper understanding in deep learning technology. The core goal of deep learning attention mechanism is to select more critical information for the current task from a large amount of information. Its representatives are channel attention and spatial attention.
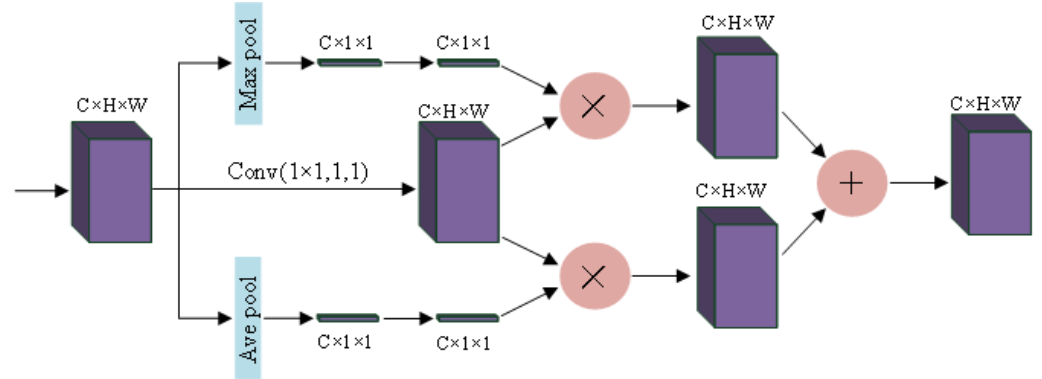
Channel attention block

Squeeze-and-Excitation Networks (SENet) (*Jie, Li & Gang, 2018*) considered the relationship between feature channels and an added attention mechanism to the feature channels. SENet automatically obtains the importance of each feature channel through learning and uses the obtained importance to improve features and suppress features that are not important for the current task. Using global maximum pooling and average pooling, the spatial dimension is compressed to obtain the channel weights of the different features, and this feature is applied to the original channel, and the features that need to be enhanced are augmented, and the two different enhanced features are fused and output. For more information, please see Fig. 2.

Spatial attention block

In the convolution layer, the receptive field is limited due to the size limitation of the convolution kernel. Wang et al. pointed out in *Wang et al. (2018a)* that nonlocal means that the receptive field can be very large, rather than the local field. If global information can be introduced into some layers, the problem of not seeing the global situation for local operations can be solved, bringing richer information to the later layers. A module called spatial transformer is proposed to transform the spatial information in the image so as to extract the key information. The essence of spatial attention is to locate the target and make some changes or obtain weights. As shown in Fig. 3, the target is located by compressing the channels with mean and maximum values to obtain the weights of different parts. Enhanced features are obtained by interacting with the original feature map, and the two enhanced feature maps are combined and convoluted and output.

Residual convolutional neural network

The residual convolutional neural network adds a jump connection branch (*He et al., 2016*) where the input feature map is added directly to the output and then activated. The residual network is a good solution to the degradation problem of deep neural networks (*Wang et al., 2019*). The error signal can be propagated directly to the lower layer without any intermediate weight matrix transformation, thus alleviating the gradient dispersion

Zou et al. (2023), *PeerJ Comput. Sci.*, DOI 10.7717/peerj-cs.1182

6/21

Figure 2 Channel attention block.
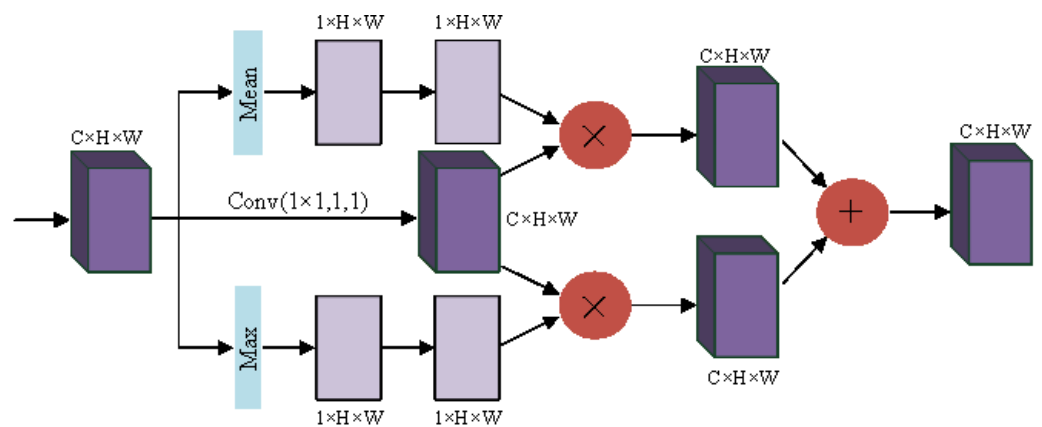
Figure 3 Spatial attention block.

problem and achieving better reconstruction. On the premise of the same number of layers, the convergence speed of the residual network is faster. In the super-resolution reconstruction proposed in this paper, the Gaussian transform enhancement module and the attention mechanism block are embedded into the residual module to construct a new residual block.

Evaluation indexes

The objective evaluation indexes of image hiding and recovery networks used in this paper include peak signal-to-noise ratio (PSNR) (*Welstead, 1999*), and structural similarity index measure (SSIM) (*Wang, Bovik & Sheikh, 2004*).

MATERIALS & METHODS

Garesat-Net

In this section, the super-resolution image reconstruction network, Garesat-Net, was proposed, and the pixel shuffle and the proposed ResBlock were used for this network.
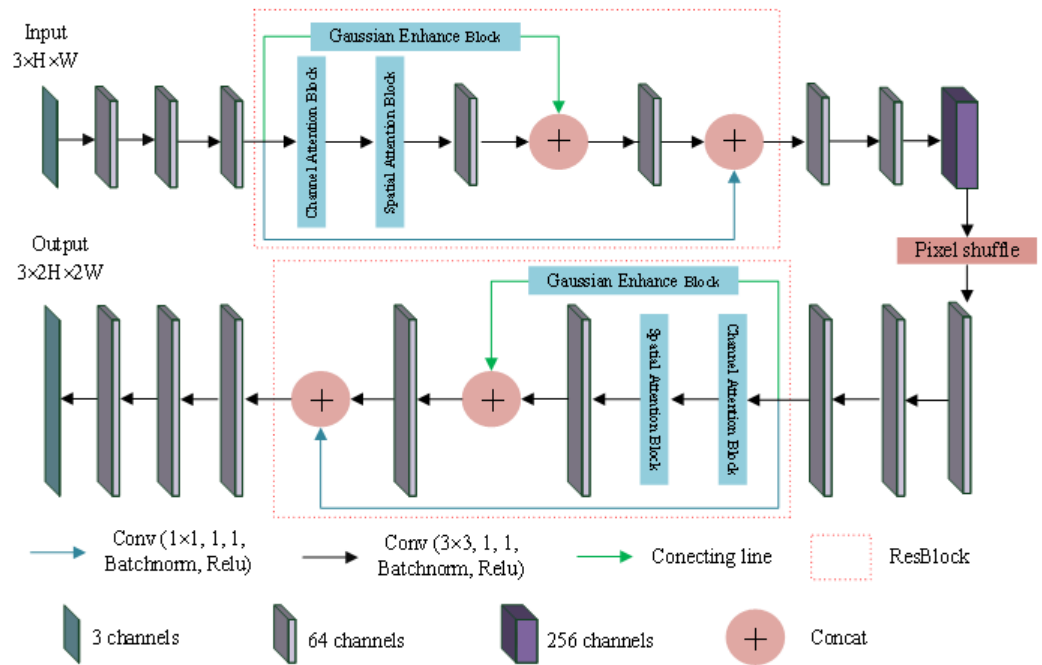
Figure 4 Garesat-Net. The struture of Garesat-Net.

The specific structure and parameter settings are shown in Fig. 4. The whole network consists of 11 convolutional layers, two ResBlocks and a pixel shuffle layer. Among them, 10 convolution layers have 64 convolution kernels, and only the sixth convolution layer has 256 convolution kernels, which is to prepare for the next pixel shuffle layer. The ResBlock has three branches. The first branch is the Gaussian enhancement block; the second branch is the channel attention module, as well as the spatial attention module and the convolutional layer, in that order. This output of the second branch is concatenated with the output of the first branch and fed to the next convolutional layer. Finally, the output is concatenated with the output of the third branch. The third branch has only one convolutional layer with a convolution kernel of 1×1. The input LR image passes through three convolutional layers, ResBlock, three convolutional layers, pixel shuffle layer, three convolutional layers, ResBlock and three convolutional layers in turn, and the output SR image is twice the length and width of the original image.

Loss function
During the training process, the loss function combines pixel loss and texture loss to limit the generation of super-resolution images.

Pixel loss
Most SR models currently adopt pixel loss. Pixel loss is generally divided into L_1 loss and L_2 loss, and their expressions are as follows:

$$L_1 = \frac{1}{CHW} \sum_{k=1}^{C} \sum_{i=1}^{H} \sum_{j=1}^{W} |\hat{y}_{i,j,k} - y_{i,j,k}| \tag{5}$$

$$L_2 = \frac{1}{CHW} \sum_{k=1}^{C} \sum_{i=1}^{H} \sum_{j=1}^{W} (\hat{y}_{i,j,k} - y_{i,j,k})^2 \tag{6}$$

where C is the number of channels of the image, generally 3; H is the height of the image; W is the width of the image; $\hat{y}_{i,j,k}$ is each pixel of the generated high-resolution image; $y_{i,j,k}$ is each pixel of the real image.

Texture loss

Gatys, Ecker & Bethge (2015b) and *Gatys, Ecker & Bethge (2016)* introduced texture loss to super-resolution reconstruction. The texture loss uses the Gram matrix. The Gram matrix expression is as follows:

$$G_{i,j}^l(I) = F_i^l(I) F_j^l(I) \tag{7}$$

where $F_i^l(I)$ is the feature map of the i-th channel on the *l*-th layer of Image *I*; $F_j^l(I)$ is the feature map of the *j* th channel on the *l*-th layer of Image *I*. The expression for texture loss is as follows:

$$L_{texture} = \frac{1}{C_l^2} \sqrt{\sum_{i=1}^{C_l} \sum_{j=1}^{C_l} (G_{i,j}^l(\hat{y}) - G_{i,j}^l(y))^2}. \tag{8}$$

The loss function is defined as:

$$
\begin{aligned}
L &= (L_2 + L_{texture})/2 \\
&= \left(\frac{1}{3HW} \sum_{k=1}^{3} \sum_{i=1}^{H} \sum_{j=1}^{W} (\hat{y}_{i,j,k} - y_{i,j,k})^2 + \frac{1}{3^2} \sqrt{\sum_{i=1}^{3} \sum_{j=1}^{3} (G_{i,j}(\hat{y}) - G_{i,j}(y))^2} \right)
\end{aligned}
\tag{9}
$$

where $\hat{y}_{i,j,k}, y_{i,j,k}$ represent each pixel of SR image and HR image, respectively, $\hat{y}, y$ represent SR image and HR image, respectively, and $G_{i,j}(\hat{y}) - G_{i,j}(y)$ represent the Gram matrix of layers i and j of SR image and HR image, respectively. In the training process, pixel loss was used to reconstruct the magnified image that was not fine enough, and texture loss was used to determine the detail part.

Optimization and parameter initialization settings

The proposed network was optimized using the Adam stochastic optimization method (*Kingma & Ba, 2014*). The convolution kernel was initialized with Gaussian distribution weights having zero expectation and variance of 0.02. In the experiment, the batch size was 4 and the initial learning rate was set to 0.005. The regularization coefficient in the loss function, the basic learning rate and the weight attenuation were set to 10e−4. After 400 epochs, the loss function became stable without further decline and stopped training.

RESULTS

In this section, we conducted several experiments to demonstrate the effectiveness of the proposed method.

Developmental environment and dataset

The proposed Garesat-Net model was trained and tested under a workstation operating system of Windows 10, accelerated with single GeForce RTX 3090Ti graphics adapters. The proposed model is based on Python (3.7) and Pytorch (1.1.0) to build a deep learning model.

As the training datasets, 200 images were used from the Berkeley Segmentation Dataset (BSD) (*Martin et al., 2001*). In the testing stage, we employed Set5 (*Bevilacqua et al., 2012*), Set14 (*Zeyde, Elad & Protter, 2010*), BSD100 (*Martin et al., 2001*), and Urban100 (*Huang, Singh & Ahuja, 2015*) datasets.

Comparative analysis of related algorithms

This experiment compares the method proposed in this paper with five existing deep learning-based SR algorithms with good performance in this field. The compared methods are RNAN (*Zhang, Li & Li, 2019*), LatticeNet (*Luo, Xie & Zhang, 2020*), SCGAN (*Liu & Chen, 2021*), FPMMLGPR (*Lu & Yu, 2022*), and AMG (*An & Wang, 2022*), where the bicubic interpolation method is used as the benchmark algorithm. This comparison verifies the effectiveness and generalization ability of the method proposed in this paper. The trained models published for these methods are used to produce all reconstruction results of the compared methods to ensure the fairness of the experiment. Table 1 shows the PSNR and SSIM indicators obtained for the above five mainstream deep learning-based ISR algorithms on the four test sets.

The proposed network reconstructs 2× super-resolution images and reconstructs 4x and 8x super-resolution images by reuse. The image quality gradually decreases as the number of uses increases. In the following illustration, we show the results of reconstructing 4× super-resolution images from LR images.

Figures 5, 6, 7 and 8 show the visual effects produced by the different 4× super-resolution reconstruction methods in BSD100, Set5, Set14 and Urban100, respectively. From a simple visual point of view, we cannot tell the difference. When magnifying a part of it, it can be seen that our proposed method is better in terms of texture features and closer to HR images. As can be seen from the spots on the anemone in Fig. 5, upon amplification, the first four and Bicubic are blurred at the edges and have differences in color. Only the AMG method and the proposed method have high quality in terms of reconstructed edges and colors. In Fig. 6, the blood filaments in the baboon's eyes have been blurred in the reconstruction of the previous methods, and our method only yields similar smoothed results.

In Fig. 7, the SR images generated by bicubic and the first three methods are not very natural, and the SR images generated by the latter two methods and our method are more natural. In Fig. 8, with a window in a high-rise building in the generated SR image, the image generated by our method is more natural and clear with better edge performance,

Table 1 Statistical analysis results of the seven methods on the four datasets of Set5, Set14, BSD100 and Urban100.

Methods	Evaluating indicator	Set5		Set14		BSD100		Urban100	
		2	4×	2×	4×	2×	4×	2×	4×
Bicubic	PSNR	32.87	28.45	30.25	25.38	29.56	24.80	26.68	22.89
	SSIM	0.9226	0.8777	0.8695	0.8245	0.8441	0.8111	0.8411	0.7905
RNAN	PSNR	36.49	30.89	33.88	28.34	32.29	27.64	32.69	27.66
	SSIM	0.9586	0.8856	0.9213	0.8476	0.9018	0.8398	0.9343	0.8746
LatticeNet	PSNR	36.93	31.11	33.88	28.49	32.30	27.45	32.82	27.94
	SSIM	0.9595	0.8853	0.9213	0.8473	0.9019	0.8367	0.9367	0.8665
SRGAN	PSNR	33.57	30.20	32.89	26.61	32.32	26.61	32.91	27.13
	SSIM	0.9543	0.8726	0.9232	0.7668	0.9021	0.7313	0.9366	0.7742
FPMMLGPR	PSNR	34.83	31.47	33.54	28.59	32.30	26.45	32.54	28.51
	SSIM	0.9527	0.8971	0.9367	0.8464	0.9034	0.8356	0.9367	0.8813
CMG	PSNR	37.14	32.45	33.88	29.69	32.35	**29.88**	32.79	**29.56**
	SSIM	0.9719	**0.9287**	0.9321	0.8888	0.9078	**0.8713**	0.9390	0.8985
Proposed	PSNR	**37.69**	**32.89**	**34.36**	**29.75**	**33.15**	29.54	**33.23**	29.46
	SSIM	**0.9775**	0.9246	**0.9413**	**0.8921**	**0.9162**	0.8691	**0.9411**	**0.9007**

Notes.

Bold indicates the best result in this column.

and for the one window in a tall building in Fig. 8, the details of the image reconstruction are improved using the texture loss function. Figure 9 shows the 2×, 4× and 8× SR images of the butterfly in Set5 reconstructed by the proposed method. Among them, the LR image used for reconstruction is a 32×32 image obtained by resizing the LR butterfly image in set5, and the SR images of 64×64, 128×128, and 256×256 are obtained by the proposed method. Their PSNR and SSIM are (38.22, 0.9777), (32.59, 0.9233), and (28.23, 0.8815), respectively. It can be seen that the image quality decreases rapidly as the number of reconstructions increases.

Table 1 shows the average performance of several methods on the four datasets. The proposed method performs best on 2xSR images and achieves the best effect on the four datasets. Our proposed method is not the best in terms of image evaluation indicators PSNR and SSIM, but it is the best in terms of detail and texture.

To further verify the method proposed in this paper, Table 2 gives the parameters and average testing time of the method proposed in this paper and other deep learning models.

As can be seen from Table 2, the number of parameters of the proposed model in this paper is only more than the number of RNAN parameters, and the average test speed is the fastest. The proposed method achieves a good balance between reconstruction performance and model complexity.

CONCLUSIONS

Garesat-Net SR was proposed to reconstruct realistic textures, edges and details by obtaining high visual quality. Gaussian enhancement block, spatial attention module and channel attention module were applied. The Gaussian enhancement block can enhance the edge

(a) Original

(b) Bicubic (26.89, 0.8351)

(c) RNAN (28.61, 0.8544)

(d) LatticeNet (28.87, 0.8656)

(e) GAN (28.82, 0.8715)

(f) APNet (29.11, 0.8758)

(g) AMG (29.65, 0.8832)

(h) Proposed (29.81, 0.8929)

Figure 5 **Compared in BSD100.** The visual effects produced by different methods of 4 × super-resolution reconstruction are compared in BSD100. Image credit: ©2001 IEEE. Reprinted, with permission, from *Martin et al. (2001)*.

Full-size ⬛ DOI: 10.7717/peerjcs.1182/fig-5
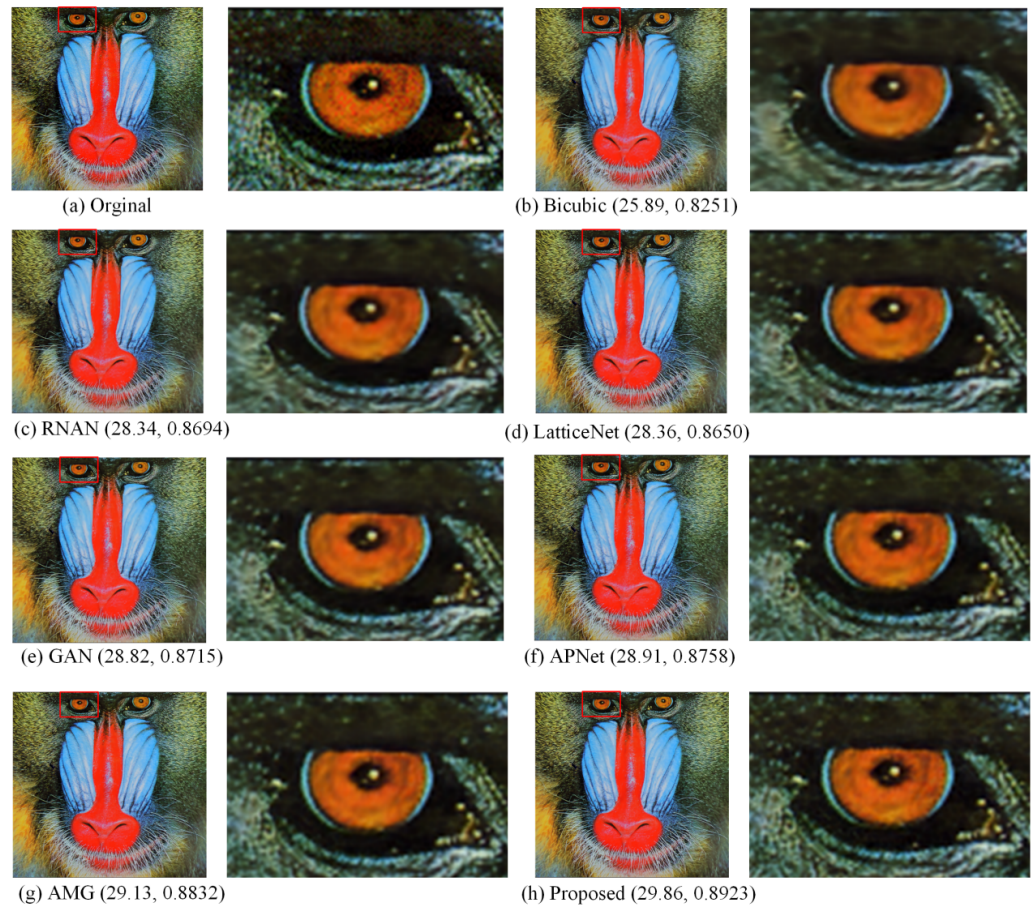
Figure 6 Compared in Set14. The visual effects produced by different methods of 4 × super-resolution reconstruction are compared in Set14. Image source credit: Reprinted by permission from SPRINGER NATURE LICENSE: Springer-Verlag Berlin, Curves and Surfaces (*Zeyde, Elad & Protter, 2010*). ©2012.

Full-size ⊡ DOI: 10.7717/peerjcs.1182/fig-6

information in the feature map, so that the final SR image is not blurred at the edges. During the learning process, the feature maps of LR images are obtained and enhanced by convolutional layer learning. Some feature maps contain only a small amount of information, while some learn most of the information of the LR image. Channel attention enhances the weights of the feature maps with more information, and the spatial attention module enhances the weights of the area containing important information in the feature map.

In the training process, the use of the pixel loss function ensures more similarity in color and structure. SR images score higher in PSNR and SSIM. The use of the texture loss function results in SR images with more detailed structure, more realistic texture and more natural details. Numerous experiments show that Garesat SR surpasses the existing models and achieves the most advanced performance. Finally, the proposed method and the five most advanced deep learning SR methods are tested on four datasets: Set 5, Set14, BSD100, and Urban100. Experimental results show that the PSNR and SSIM of the SR

Figure 7 Compared in Set5. The visual effects produced by different methods of 4 × super-resolution reconstruction are compared in Set5. Image source credit: *Bevilacqua et al. (2012)*. Dataset: https://people.rennes.inria.fr/Aline.Roumy/results/SR_BMVC12.html.

Full-size 🖼 DOI: 10.7717/peerjcs.1182/fig-7

(a) Original

(b) Bicubic (24.36, 0.8180)

(c) RNAN (25.68, 0.8625)

(d) LatticeNet (26.21, 0.8727)

(e) GAN (27.48, 0.8860)

(f) APNet (28.80, 0.8922)

(g) AMG (29.27, 0.9011)

(h) Proposed (29.55, 0.9032)

Figure 8 **Compared in Urban100.** The visual effects produced by different methods of 4 × super-resolution reconstruction are compared in Urban100. Image source credit: ©2015 IEEE. Reprinted, with permission, from *Huang, Singh & Ahuja (2015)*.

Full-size 🖼 DOI: 10.7717/peerjcs.1182/fig-8

(a) 2x (38.22, 0.9777)

(b) 4x (32.59, 0.9233)

(c) 8x (28.23, 0.8815)
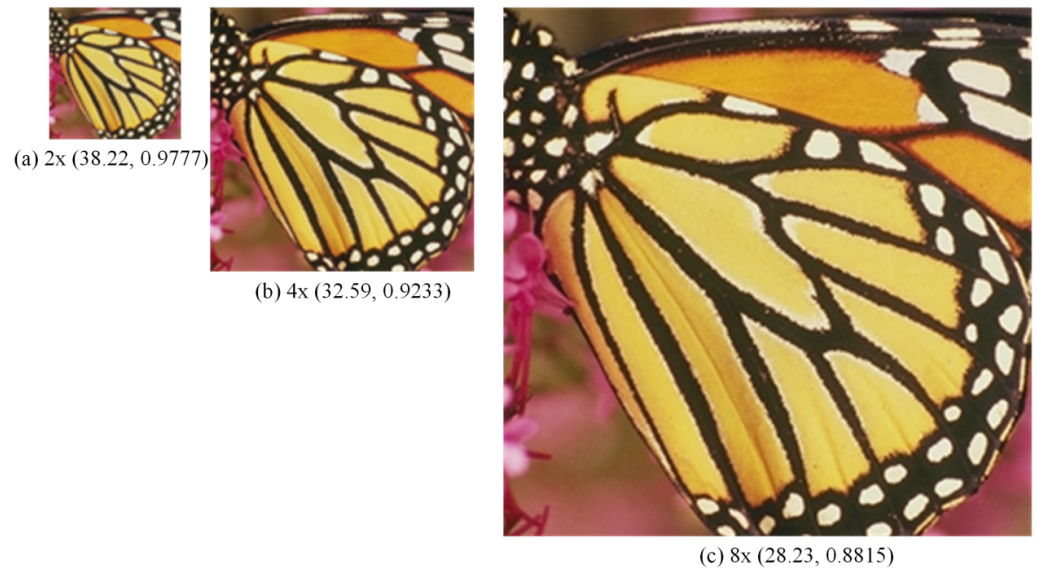
Figure 9 **SR images of butterfly.** The 2 ×, 4 × and 8 × SR images of butterfly in Set5 reconstructed by the proposed method. Image source credit: *Bevilacqua et al. (2012)*. Dataset: https://people.rennes.inria.fr/Aline.Roumy//results/SR_BMVC12.html.

Full-size 🖼 DOI: 10.7717/peerjcs.1182/fig-9

Table 2 **Deep learning model parameter comparison.**

	RNAN	LatticeNet	SRGAN	CMG	Proposed
Number of parameters	69087	909283	710343	1883045	793899
Average of test time	2.8	2.63	1.03	1.01	0.85

images obtained by the proposed method are not optimal during testing, but the obtained images are more natural and closer to HR images in terms of details and edges. After enlarging the details, we can also see that there are differences between SR images and HR images, which means that the SR algorithm needs further improvement, which will be our future work.

ADDITIONAL INFORMATION AND DECLARATIONS

Competing Interests

The authors declare there are no competing interests.

Author Contributions

- Shuilong Zou conceived and designed the experiments, authored or reviewed drafts of the article, and approved the final draft.
- Mengmu Ruan performed the experiments, analyzed the data, prepared figures and/or tables, and approved the final draft.
- Xishun Zhu performed the experiments, performed the computation work, prepared figures and/or tables, and approved the final draft.
- Wenfang Nie conceived and designed the experiments, authored or reviewed drafts of the article, and approved the final draft.

Data Availability

The following information was supplied regarding data availability:

The data is available from:

- In Proceedings British Machine Vision Conference 2012. Pages 135.1–135.10. http://dx.doi.org/10.5244/C.26.135

- Zeyde, R., Elad, M., Protter, M. (2012). On Single Image Scale-Up Using Sparse-Representations. In: , et al. Curves and Surfaces. Curves and Surfaces 2010. Lecture Notes in Computer Science, vol 6920. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-27413-8_47

- 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," Proceedings Eighth IEEE International Conference on Computer Vision. ICCV 2001, 2001, pp. 416-423 vol.2, doi: 10.1109/ICCV.2001.937655.

- J. -B. Huang, A. Singh and N. Ahuja, "Single image super-resolution from transformed self-exemplars," 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015, pp. 5197-5206, doi: 10.1109/CVPR.2015.7299156.

- Kaggle: Available at https://www.kaggle.com/datasets/msahebi/super-resolution; Nie, Wenfang (2022): BSD100, Set5, Set14, Urban100. figshare. Dataset. https://doi.org/10.6084/m9.figshare.21586188.v1

The code is available at Github: https://github.com/jbhuang0604/SelfExSR, DOI: 10.1109/CVPR.2015.7299156.

Supplemental Information

Supplemental information for this article can be found online at http://dx.doi.org/10.7717/peerj-cs.1182#supplemental-information.

REFERENCES

An F, Wang J. 2022. Image super-resolution reconstruction algorithm based on significant network connection-collaborative migration structure. *Digital Signal Processing* **127**:103566 DOI 10.1016/j.dsp.2022.103566.

Belekos S, Galatsanos N, Katsaggelos A. 2010. Maximum a posteriori video super-resolution using a new multichannel image prior. *IEEE Transactions on Image Processing* **19**:1451–1464 DOI 10.1109/TIP.2010.2042115.

Bevilacqua M, Roumy A, Guillemot C, Alberi Morel M-L. 2012. Low-complexity single image super-resolution based on nonnegative neighbor embedding. In: *Proceedings of the British Machine Vision Conference*. BMVA Press, 135 DOI 10.5244/C.26.135.

Chen Y, Liu L, Phonevilay V, Gu K, Xia R, Xie J, Zhang Q, Yang K. 2021a. Image super-resolution reconstruction based on feature map attention mechanism. *International Journal of Speech Technology* **51**(7):4367–4380.

Chen Y, Liu L, Tao J, Chen X, Xia R, Zhang Q, Xiong J, Yang K, Xie J. 2021b. The image annotation algorithm using convolutional features from intermediate layer of deep learning. *Multimedia Tools and Applications* **80**(3):4237–4261 DOI 10.1007/s11042-020-09887-2.

Dong C, Loy C, He K, Tang X. 2014. Learning a deep convolutional network for image super-resolution. In: *European conference on computer vision (ECCV)*. 184–199.

Dong C, Loy C, Tang X. 2016. Accelerating the super-resolution convollutional neural network. *Computer Vision* **25**:391–407.

Fan C, Wu C, Li G, Ma J. 2017. Projections onto convex sets super-resolution reconstruction based on point spread function estimation of low-resolution remote sensing images. *Sensors* **17**:362–361 DOI 10.3390/s17020362.

Gatys LA, Ecker AS, Bethge M. 2015a. A neural algorithm of artistic style. ArXiv preprint. arXiv:1508.06576.

Gatys L, Ecker A, Bethge M. 2015b. Texture synthesis using convolutional neural networks. ArXiv preprint. arXiv:1505.07376.

Gatys L, Ecker A, Bethge M. 2016. Image style transfer using convolutional neural networks. In: *Proceedings of the 2016 IEEE conference on computer vision and pattern recognition (CVPR)*. Piscataway: IEEE, 2414–2423.

Goodfellow I, Pouget-Abadie J, Mirza M, Xua B, Warde-Farleya D, Ozaira S, Courvillea A, Bengioa Y. 2014. Generative adversarial networks. *Advances in Neural Information Processing Systems* **3**:2672–2680.

He K, Zhang X, Ren S, Sun J. 2016. Deep residual learning for image recognition. In: *IEEE conference on computer vision and pattern recognition*. Piscataway: IEEE, 770–778.

Huang G, Liu Z, Laurens V, Weinberger KQ. 2017. Densely connected convolutional networks. *IEEE Computer Society* **2017**:2261–2269 DOI 10.1109/CVPR.2017.243.

Huang J, Singh A, Ahuja N. 2015. Single image super-resolution from transforme self-exemplars. In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Piscataway: IEEE, 5197–5206 DOI 10.1109/CVPR.2015.7299156.

Irani M, Peleg S. 1991. Improving resolution by image registration. *CVGIP: Graphical Models and Image Processing* **53**:231–239 DOI 10.1016/1049-9652(91)90045-L.

Jiao D, Wang D, Lv H, Peng Y. 2020. Super-resolution reconstruction on a digital elevation model based on a deep residual network. *Open Geosciences* **12**(1):1369–1382 DOI 10.1515/geo-2020-0207.

Jie H, Li S, Gang S. 2018. Squeeze-and-Excitation networks. In: *2018 IEEE/CVF conference on computer vision and pattern recognition*. Piscataway: IEEE, 7132–7141.

Jing W, Wu A. 2014. Super-resolution image reconstruction based on iteration wavelet transform. *Journal of Communication and Computer* **11(1)**:39–44.

Kim J, Lee J, Lee K. 2016. Accurate image super-resolution using very deep convolution networks. In: *Proceedings of the 2016 IEEE conference on computer vision and pattern reconition (CVPR)*. Piscataway: IEEE, 1646–1654.

Kim Y, Koh Y, Lee C, Kim S, Kim C-S. 2015. Dark image enhancement based onpairwise target contrast and multi-scale detail boosting. In: *IEEE international conference on image processing*. Piscataway: IEEE, 1404–1408.

Kingma D, Ba J. 2014. Adam: a method for stochastic optimization. Computer science. ArXiv preprint. arXiv:1412.6980.

Ledig C, Theis L, Huszar F, Caballero J, Cunningham A, Acosta A, Aitken A, Tejani A, Totz J, Wang Z, Shi W. 2016. Photo-realistic single image super-resolution using a generative adversarial network. In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 105–114 DOI 10.1109/CVPR.2017.19.

Liu B, Chen J. 2021. A super resolution algorithm based on attention mechanism and SR-GAN network. *IEEE Access* **9**:139138–139145 DOI 10.1109/ACCESS.2021.3100069.

Liu L. 2018. Infrared image reconstruction method based on Fourier transform. *Ship Science and Technology* **40(2)**:169–171.

Lu Z, Yu H. 2022. Single image super-resolution algorithm based on fixed-point multi-model gaussian process regression. *IOP Publishing Ltd* **2289**:012024.

Luo X, Xie Y, Zhang Y. 2020. Latticenet: towards lightweight image super-resolution with lattice block. In: *Computer vision—ECCV*. 272–289.

Ma Z, Ren G. 2020. Projection onto the convex sets model based on non-downsampling contourlet transform and high-frequency iteration. *Electronics Letters* **56**:1054–1056 DOI 10.1049/el.2020.0364.

Mao X, Li Q, Xie H, Lau RYK, Wang Z, Smolley SP. 2017. Least squares generative adversarial networks. In: *Proceedings of the 2017 IEEE international conference on computer vision (ICCV)*. Piscataway: IEEE, 2813–2821.

Mao X, Shen C, Yang Y. 2016. Image restoration using convolutional auto-encoders with symmetric skip connections. ArXiv preprint. arXiv:1606.08921v3.

Martin D, Fowlkes C, Tal D, Malik J. 2001. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In: *Proceedings Eighth IEEE International Conference on Computer Vision. ICCV 2001*. Piscataway: IEEE, 416–423 DOI 10.1109/ICCV.2001.937655.

Mechrez R, Talmi I, Shama F, Zelnik-Manor L. 2018. Maintaining natural image statistics with the contextual loss. In: *Proceedings of the 14th asian conference on computer vision (ACCV)*. 427–443.

Nascimento T, Salles E. 2020. Multi-Frame super-resolution combining demons registration and regularized bayesian reconstruction. *IEEE Signal Processing Letters* **27**:2009–2013 DOI 10.1109/LSP.2020.3033422.

Patti A, Altunbasak Y. 2001. Artifact reduction for set theoretic super resolution image reconstruction with edge adaptive constraints and higher-order interpolants. *IEEE Transactions on Image Processing* **10**:179–186 DOI 10.1109/83.892456.

Rudin L, Osher S, Fatemi E. 1992. Nonliner total variation based noise removal algorithms. *Physica D: Nonlinear Phenomena* **60(1–4)**:259–268 DOI 10.1016/0167-2789(92)90242-F.

Sajjadi M, Scholkopf B, Hirsch M. 2017. EnhanceNet: Single image super-resolution through automated texture synthesis. In: *Proceeding of the 2017 IEEE international conference on computer vision (ICCV)*. 4501–4510.

Seema R, Bailey K. 2019. Multi-frame image super-resolution by interpolation and iterative backward projection. In: *Proceedings of the 2nd international conference on signal processing and communication (ICSPC)*. 36–40.

Shen H, Zhang L, Huang B, Li P. 2007. A MAP approach for joint motion estimation, segmentation, and super resolution. *IEEE Transactions on Image Processing* **16**:479–490 DOI 10.1109/TIP.2006.888334.

Shi W, Caballero J, Huszár F, Totz J, Aitken AP, Bishop R, Rueckert D, Wang Z. 2016. Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. In: *2016 IEEE conference on computer vision and pattern recognition (CVPR)*. Piscataway: IEEE, 16541213 DOI 10.1109/CVPR.2016.207.

Shi X, Li Z, Jiahuan J. 2022. Single image super-resolution algorithm based on multi-scale recursive dense network. *Journal of Shanxi University* **43(4)**:719–726.

Song H, He X, Chen W, Sun Y. 2010. An improved iterative back-projection algorithm for video super-resolution reconstruction. In: *Proceedings of the symposium on photonics and optoelectronics*. DOI 10.1109/SOPO.2010.5504209.

Tong T, Li G, Liu X, Gao Q. 2017. Image super-resolution using dense skip connections. In: *IEEE international conference on computer vision*. Piscataway: IEEE, 4809–4817 DOI 10.1109/ICCV.2017.514.

Wang G, Zeng G, Cui X, Feng S. 2019. Dispersion analysis of the gradient weighted finite element method for acoustic problems in one, two, and three dimensions. *International Journal for Numerical Methods in Engineering* **120(4)**:473–497 DOI 10.1002/nme.6144.

Wang X, Girshick R, Gupta A, He K. 2018a. Non-local neural networks. In: *2018 IEEE/CVF conference on computer vision and pattern recognition*. 7794–7803 DOI 10.1109/CVPR.2018.00813.

Wang X, Yu K, Dong C. 2018. Recovering realistic texture in image super-resolution by deep spatial feature trasform. In: *Proceedings of the 2018 IEEE CVF conference on computer vision and pattern recognition (CVPR)*. Piscataway: IEEE, 606–615.

Wang X, Yu K, Wu S, Gu J, Liu Y, Dong C, Loy CC, Qiao Y, Tang X. 2018b. ESRGAN: Enhanced super-resolution generative adversarial networks. *Computer Vision* **11133**:63–79 DOI 10.1553/0x0038c0ea.

Wang Z, Bovik A, Sheikh H. 2004. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing* **13(4)**:24–33.

Wang Z, Wang X, Wang G. 2015. Learning fine-grained features via a CNN tree for large-scale classification. *Computer Science* **25**:1231–1240 DOI 10.1016/j.neucom.2017.09.061.

Welstead S. 1999. Fractal and wavelet image compression techniques. *SPIE Optical Engineering Press* 155–156.

Xiong S, Bin X, He Y, Zhang Z. 2015. Image reconstruction method based on CCD calibration in frequency domain. *Applied Optics* **54(14)**:4561–4565 DOI 10.1364/AO.54.004561.

Yu J, Fan Y, Yang J, Xu N, Wang Z, Wang X, Huang T. 2018. Wide activation for efficient and accurate image super-resolution. ArXiv preprint. arXiv:1808.08718.

Zeyde R, Elad M, Protter M. 2010. On single image scale-up using sparse representations. In: *International conference on curves and surfaces. Lecture Notes in Computer Science*, vol. 6920. Berlin, Heidelberg: Springer, 711–730 DOI 10.1007/978-3-642-27413-8_47.

Zhang Y, Li K, Li K. 2019. Residual non-local attention networks for image restoration. ArXiv preprint. arXiv:1903.10082.

Zhang Y, Li KP, Li K, Wang L, Zhong B, Fu Y. 2018a. Image super-resolution using very deep residual channel attention networks. In: *Proceedings of the 15th european conference on computer vision (ECCV)*. 286–301.

Zhang Y, Tian Y, Kong Y, Zhong B, Fu Y. 2018b. Residual dense network for image super-resolution. In: *Conference on computer vision and pattern recognition*. 2472–2481 DOI 10.1109/CVPR.2018.00262.