Research of Single Image Super-Resolution Reconstruction with Sawtooth Dilated Residual Convolution*

LI Lan, LIN Guoliang, MA Shaobin

(*School of Digital Media, Lanzhou University of Arts and Science, Lanzhou Gansu 730000, China*)

Abstract: In order to solve the problems of the limited receptive field, low-resolution, high complexity and loss of edge information in the super-resolution reconstruction method of residual learning, adilated residual convolution neural network is proposed. Firstly, we design the sawtooth dilated convolution based on the ResNet network to expand the receptive field of the network and eliminate the "zero filling" of the network, the image features are transferred to the deeper network by adding the jump connection. Secondly, the residual image with the same size as the original image is obtained through the last convolution layer. Finally, the input LR image and the residual image are linearly fused to output the final super-resolution image. The experimental data on set 5 and set 14 shows that compared with the existing algorithms, the algorithm of this paper has better reconstruction effect and better learning performance.

Key words: residual network; dilated convolution; deep learning; image super-resolution reconstruction

DOI: 10.13568/j.cnki.651094.651316.2020.07.30.0002

CLC number: TP391 **Document Code:** A **Article ID:** 2096-7675(2021)02-0174-17

Citation format: LI L, LIN G L, MA S B. Research of single image super-resolution reconstruction with sawtooth dilated residual convolution[J]. Journal of Xinjiang University(Natural Science Edition in Chinese and English), 2021, 38(2): 174-190.

0 Introduction

The idea of super-resolution image reconstruction (SRIR or SR) is to use a group of low quality and low-resolution images (LR) to obtain single or multi frame high quality and high-resolution images through computer technology and image processing technology[1,2].SR is an important research direction of digital image processing, and has a wide range of applications in the field of computer vision, such as intelligent transportation, safety monitoring, image generation and medical imaging[3].At present, SR methods are mainly divided into 3 categories: difference-based method[4], reconstruction-based method[5] and learning-based method[6,7].The difference-based method is to take the image as a point, and use the prior knowledge to fit the unknown information on the plane by a predefined transformation function or interpolation, so as to calculate the high-resolution image. The main disadvantage of this method is that it is easy to appear the phenomenon of ladder sawtooth and edge blurring. The reconstruction-based method is one of the widely studied. This kind of method mainly uses the under-sampling technology to fuse the multi frame information of low pixel accuracy on one or more low-resolution image information , and reconstruct the image with higher resolution. This method has high reconstruction accuracy, but it can only make use of the relationship between high-resolution and low-resolution images, so it is difficult to build a mathematical model, and the texture of the reconstructed image is not clear.

In recent years, the deep learning-based method has become a research hotspot. It mainly uses the internal similarity of the same image and a large number of training sample data to find the mapping relationship between high and low-resolution image pair, and complete the transformation of high-resolution image features, so as to realize the SR process. This method

* **Received Date:** 2020-07-30

Foundation Item: Higher Education Teaching Achievement Cultivation Project of Gansu Province in 2020（2020-204）; Innovation and Entrepreneurship Education Project of Gansu Province in 2019; Research Project of Gansu University(2018A-138); Industry Support and Guidance Project of Gansu Provincial Department of Education（2019C-09）.

Biography: LI Lan (1978-), female, master's degree, associate professor, research on deep learning, intelligent information processing, E-mail: 148439473@qq.com.

requires very high centralized characteristics of modeling data. Dong et al[8] firstly used the deep learning method. A total of three-layer convolution neural network (CNN) was set up in the network to realize super-resolution reconstruction, and achieved good results. Since then, it has opened the upsurge of deep learning to realize SR. In the process of training, with the increase of the number of network layers, there will be problems such as too many hyper parameters, gradient diffusion/explosion and so on. The reconstructed image is usually too smooth, losing high-frequency details, and the image quality still needs to be improved.

Kim et al[9] proposed a deep residual network (VDSR) model, which uses residual learning to accelerate the convergence speed of the network. It is proved that this method can improve the performance of super-resolution, but it will increase the computational complexity and the gradient disappearance.Yang et al[10] used the sparsity of the image to constrain the sparse representation under the dictionary corresponding to the high and low-resolution images to realize image super-resolution reconstruction. The reconstruction effect of this method is good, but the disadvantage is that dictionary training takes a long time and there will be noise at the edge of the image. Tai et al[11] proposed DRRN model, which uses recursive network module with weight sharing to increase the network depth to 52 layers and reduce the parameters of the model. However, each recursive unit is not optimized enough and the reconstruction effect is not obvious.Yu et al[12] and Chen et al[13] proposed the dilated convolution. By changing the size of convolution kernel without adding parameters in the network, we can obtain larger receptive field and obtain more original image information, and which achieves good results in reconstruction. However, this method will appear the phenomenon of ”gridding”, and more image information will be lost after convolution.

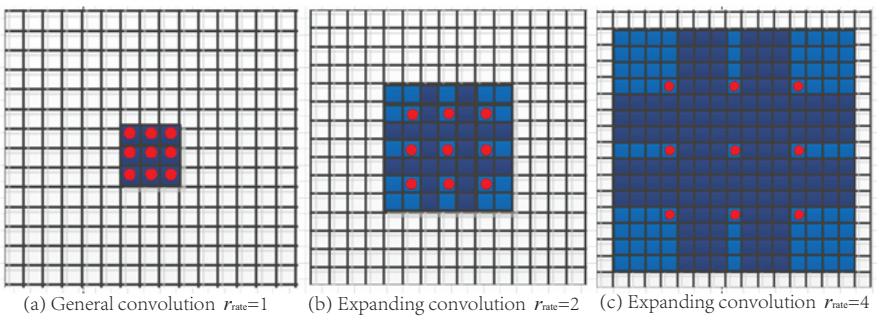
In this paper, we improve the above methods and propose an image super-resolution reconstruction method combining residual network and sawtooth dilated convolution. The model uses the dilated convolution network to extract image features, then uses residual network combined with sawtooth hole convolution to carry out image nonlinear mapping, and then applies convolution network to obtain residual image with the same size as the input image. The final super-resolution image is obtained by linear fusion of low-resolution image and residual image. Adaptive moment estimation (Adam) is used to speed up the convergence of the network. Dilated convolution is used to enlarge the receptive field of image features and recover the texture information of the image with high quality, which improves the visual effect of the reconstructed image.

1 Related work

1.1 Dilated convolution

Dilated convolution is a data sampling method on the feature map. The network expansion coefficient is increased by adding 0 pixel value between each pixel of ordinary convolution kernel, it can effectively increase the receptive field without increasing the model parameters or calculation. Dilated convolution can be applied to image global information or voice text which needs long sequence information dependence[10].

For a 3 ×3 convolution network, its expansion rate and receptive field are shown in Fig 1.



(a) General convolution r_{rate}=1 (b) Expanding convolution r_{rate}=2 (c) Expanding convolution r_{rate}=4

Fig 1 3 × 3 pixels convolution with different dilation rate

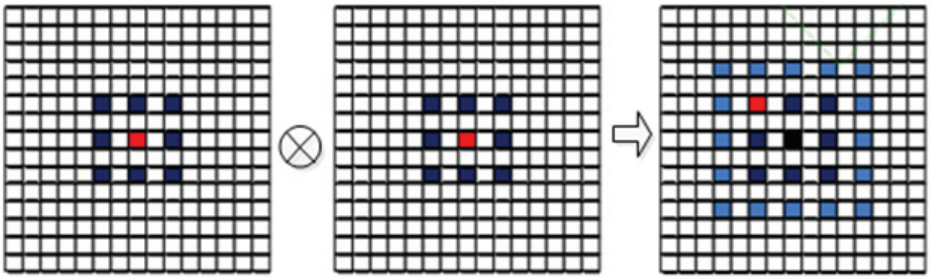
Fig 1(a) is a general convolution (hole convolution with expansion rate = 1), and the data represented is the convolution of 3 ×3 pixels in the original figure, and the receptive field is 3 ×3, equivalent to no expansion, r_{rate}= 1; Fig 1(b) is the dilated convolution with expansion rate r_{rate}= 2, receptive field is 7 ×7, dilation = 2; Fig 1 (c) is dilated convolution with expansion

rate r_{rate}= 4, receptive field is 15 ×15, dilation = 4. The improvement of dilated convolution to ordinary convolution is to obtain larger receptive field. The calculation of receptive field is expressed by formula (1).

$$v = ((k_{\text{size}} - 1)(r_{\text{rate}} - 1) + k_{\text{size}})^2 \tag{1}$$

where k_{size} represents the size of convolution kernel, r_{rate} represents the expansion rate of dilated convolution, and v represents the size of receptive field.

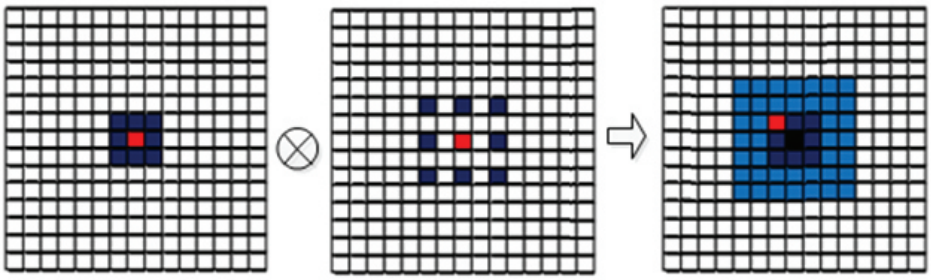
The internal feature information of the image can not only be retained, but also the loss of resolution caused by pooling can be avoided using the dilated convolution. However, it still has some defects: the adjacent pixels are convoluted from independent subsets in the same dilated convolution which result in a certain layer, and the "grid" effect is caused due to the lack of mutual dependence and the information discontinuity. The specific convolution is shown in Fig 2.



(a) Receptive field of r_{rate}=2 (b) Receptive field of r_{rate}=2 (c) Receptive field after convolution

Fig 2 Hole convolution with the same expansion rate

In Fig 2 (a), the convolution layer with expansion rate r_{rate}= 2, and the receptive field is 5 × 5; Fig 2(b) shows the second convolution layer, the receptive field is 5×5. Fig 2(c) is the result of one convolution for the dilated convolution with expansion rate r_{rate}=2, and the receptive field is 9. If all the dilated convolutions use the same expansion rate, the calculation method is similar to the chessboard format, and there is no dependence between large-scale data. With the further increase of the depth of the network, the important information will be lost. To solve this problem, this paper proposes a sawtooth mixed dilated convolution, as shown in Fig 3.



(a) Receptive field of r_{rate}=1 (b) Receptive field of r_{rate}=2 (c) Receptive field after convolution

Fig 3 Convolution results with different expansion rates

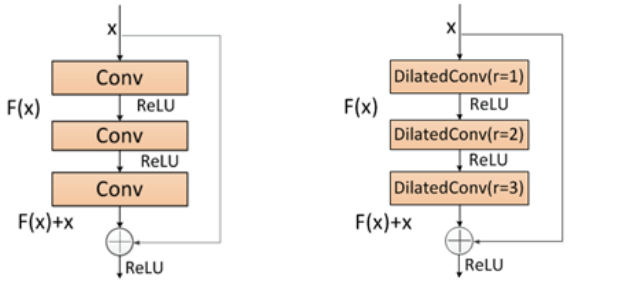
As can be seen from Fig 3, the receptive field is increased meanwhile has no blind area covering the whole area based on the sawtooth dilated convolution, so it can effectively extract features and improve the accuracy of reconstruction. Because of 0 pixel value is not involved in convolution operation, and the complexity remains unchanged.

1.2 Residual network (ResNet)

For CNN, simply increasing the depth will lead to gradient dispersion or gradient explosion. He et al[14] proposed the residual learning network ResNet, which directly connects shallow network and deep network by adding jump connection (identity mapping). In ResNet, there is a batch normalization(BN) after each convolution layer. It is pointed out in reference [15] that BN layer can improve the generalization ability of the network and accelerate the convergence process of training. However, the spatial information of the image is destroyed and the training parameters are increased in a certain extent, which

leads to poor network performance. Therefore, the ResNet is improved in this paper. The convolution layer in residual unit adopts dilated convolution, and BN layer is removed, which is helpful to obtain better image super-resolution reconstruction results.

If the same expansion rate is used for convolution, the convolution kernel will produce "grid" phenomenon. At the same time, it will carry redundant information and cause unnecessary memory occupation if the convolution kernel volume is too large. Therefore, the sawtooth dilated residual unit is designed in this paper, and the feature map with the same structure as the standard convolution is obtained by cyclic operation, which can increase the receptive field of the network while retaining the detailed information of the image, so it can better fit the target boundary and improve the image reconstruction quality. Therefore, the sawtooth dilated residual unit is designed in this paper, and the feature map with the same structure as the standard convolution is obtained by cyclic operation, which can increase the receptive field of the network while retaining the detailed information of the image, so it can better fit the target boundary and improve the image reconstruction quality. The structure of serrated cavity residual element and ResNet residual element is shown in Fig 4.



(a) Original residual block (b) the hole convolution residual block

Fig 4 Residual network unit

2 Network and implementation process

The idea of this algorithm is to obtain more image information using smaller convolution kernel, not increasing the number of layers and complexity of the network, can speed up the convergence speed of the network, at the same time to avoid the "blind area" phenomenon because of holes. Firstly, the input image of the residual network is trained by dilated convolution with different expansion rates, and then the super-resolution image is reconstructed by linear addition of the input low-resolution image and the output residual image from the residual network.

2.1 Network structure

With the increasing depth of the network, we find that the deeper CNN network layer, the better the performance. The convergence speed of the network will be affected with increasing the number of layers when the network reaches a certain depth, and the receptive field will be decreased, which can cause the reduction of context information and poor reconstruction effect. He Kaiming et al[14] proposed the deep residual network in 2015. In image classification of the ImageNet, the accuracy of network classification can be improved by increasing network depth, and the problems of gradient disappearance and network performance degradation can be solved by residual learning.

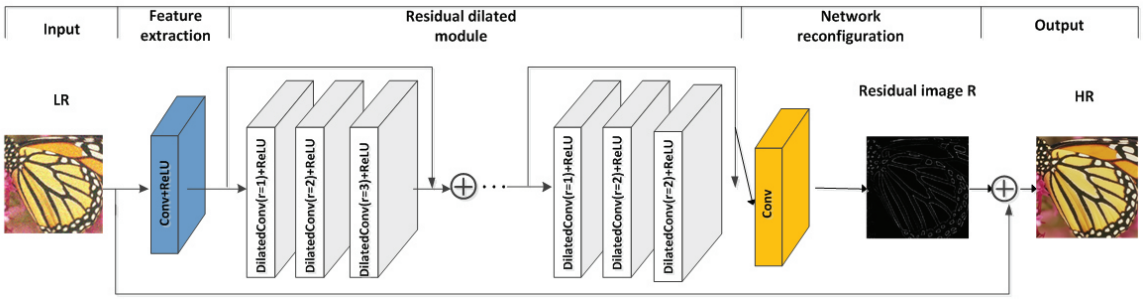


Fig 5 Network structure of single image super-resolution reconstruction method

Based on reference [13], a residual network of sawtooth dilated convolution is constructed in this paper. The overall structure of the network is shown in Fig 5, including 20 convolution layers. According to the function, it is divided into 5 parts: low-resolution image input part, convolution feature extraction part, the followed by which is 6 dilated residual blocks, each residual block contains 3 residual units, and the next is the residual image feature layer. The feature map size of each output convolution layer can be unchanged using the dilated residuals.

2.2 Image super-resolution reconstruction process

The network input is formed by a single color channel of the image data, and is preprocessed by Bicubic difference.

The dilated residual sawtooth convolution neural network super-resolution reconstruction proposed in this paper mainly includes the following three steps:

(1) Feature extraction layer (Conv + ReLu layer). Except that the size of convolution kernel in the last layer is 1 pixel, the convolution feature size of the other middle layers is $3 \times 3 \times 64$ pixels. In the feature extraction layer, the input low-resolution image X is convoluted with a convolution kernel of $3 \times 3 \times 1$ pixel size to obtain n_1 feature maps (here $n_1 = 224$), where 1 represents the number of channels. The feature extraction layer contains a convolution layer and an activation function, and each neuron obtained from this layer is transferred to the residual unit module. The expression of feature extraction process is as follows:

$$B_0 = f(W_1 * X + b_1) \tag{2}$$

where * is the convolution operation, W_1 is the convolution kernel of the first convolution layer; b_1 is the offset term, which is consistent with that of the convolution kernel in dimension; B_0 is the input feature of the first residual block extracted from the input x; f is the ReLU activation function.

In this paper, the application of ReLU in the network can accelerate the training speed and shorten the convergence time of the model, and at the same time, it can restrain the phenomenon of gradient disappearance in a certain extent. Its performance is better than the traditional activation function Sigmoid[16] (gradient explosion and gradient loss are caused by gradient reverse transmission in deep convolution network), the expression is as follows:

$$f(x_i) = \begin{cases} x_i, & x_i > 0 \\ 0, & x_i < 0 \end{cases} \tag{3}$$

where x_i is the input of the ReLU function and $f(x)_i$ is the output of the ReLU function.

(2) Nonlinear mapping. For low-resolution image of the input, the sawtooth residual network proposed is used for image training. The structure of the residual module is composed of 6 dilated residual units, each of which is composed of 3 dilated convolution layers and 3 nonlinear activation function ReLU. Each residual cell has two parts: jump connection and identity mapping. In this way, the residual information can be retained, and the image features can be transferred backward by jumping connection, which helps to maintain the diversity of features.

The convolution kernel of 3×3 pixels is used, 3 convolution layers with 1, 2, 3 are connected in series to form a residual block. In order to keep the size of convolution kernel unchanged, the output of 13×13 feature map is guaranteed at the end of network calculation. After the dilated residual block, the receptive field is expanded and the more original image information is obtained. The expression of each residual cell is as follows:

$$H_m = G_m(H_{m-1}) = F(H_{m-1}, W_m) + H_{m-1} \tag{4}$$

where H_{m-1} and H_m is the input and output of the m-th residual unit respectively, F is the residual mapping learned, that is:

$$F(H_{m-1}, W_m) = W_m^2 * f(W_m^i * H_{m-1}) \tag{5}$$

where, W_m^i, i= 1, 2, 3 is the weight of the i convolution layer learned, which is a simplified formula with omitting the bias term, f is ReLU function and $*$ is convolution operation.

(3) Image reconstruction and output. Firstly, the output feature map of the residual network is taken as the input of the last convolution layer, which is convoluted with the 3×3 pixels convolution kernel to generate the residual image of the same size as the input image, and then is linearly superimposed with the input interpolation image to output

Fig 6 Image reconstruction phase

the final super-resolution image. The structure of image reconstruction phase is shown in Fig 6.

2.3 Loss function

The sawtooth dilated residual is applied in the network to make the input image equal output image in size, and the problem of network convergence is solved. x represents the input image after the double cubic difference, $f_{Res}(x)$ represents the residual image of the input image after being passed through the whole neural network, y represents the original high-resolution image, and y^* represents the super-resolution image predicted by the network, that is:

$$y^* = x + f_{\text{Res}}(x)$$

In this paper, the Mean Square Error (MSE) function is used as the loss function of the whole network. The minimum value is achieved and the optimal solution is obtained by calculating the MSE of the image generated and the original high-resolution image. The calculation formula of loss function is:

$$Loss(\Theta) = \frac{1}{n}\sum_{i=1}^{n} \|y_i - f(x_i;\Theta)\|^2 \tag{6}$$

where n is the number of samples, y_i is the high-resolution image, $f(x_i;\Theta)$ is the prediction output image of the network, $\Theta = \{w_1, w_2, \cdots, b_1, b_2, \cdots\}$.

3 Experiments and analysis

3.1 Experiment environment

The experiment environment of the image super-resolution reconstruction algorithm is shown in Tab 1.

Tab 1 Experiment environment

programming environment	server hardware configuration	server software environment
programming language: MatlabR2016a	CPU: Intel(R) Core(TM)i7-8700k 3.7 GHz Video card：11 G GTX1080 GPU: NVIDIA Titan CUDA RAM : 16 G	system: Windows 7 Operating System
deep learning framework: Caffe[17]		

3.2 Experiment setup and evaluation index

Because the network is relatively deep, the algorithm needs to use larger training sets to get better training results. 291 images which are the same as the reference [13] are selected as the training sets, and these images are respectively from 91 images proposed in reference [18] and 200 from BSD (Berkeley Segmentation Dataset)[13]. In order to make full use of the depth image, the dataset images were flipped horizontally, vertically and horizontally vertically, and scaled according to the coefficients of 0.9 and 0.8. Then the images were saved and a total of 5 820 images were generated, and the image size was no more than 512×512.

In the process of training images using this network, the training images are transformed into YCbCr space[19]. Compared with change of color (CBCR channel information), human beings are more sensitive to the change of brightness (Y channel), so the network designed in this paper only deals with y channel. The image convolution kernel size is 3×3, and the number of characteristic image channels is 64. The optimization algorithm used in training is Adam[20]. Compared with SGD (stochastic gradient descent, SGD), Adam optimization method is more flexible and adaptive, which can control the learning rate of each iteration within a certain range, making the parameter learning more stable. The initial learning rate of the network is set to 10^{-4}, the momentum parameter is 0.9, and the mini-batch training mode is adopted, and the batch size is 64, which is reduced to half of the original value every 100 000 iterations.

In the experiment, set 5 and set 14 are used as test sets, the original high-resolution image is X, the magnification is set s=2,3,4, and the preprocessed image is used for the input of the network.

Evaluation method of image super-resolution reconstruction is to verify whether the image super-resolution reconstruction method meets the expectation. At present, there are two methods to evaluate the quality of reconstructed images: subjective evaluation and objective evaluation.

Subjective evaluation mainly refers to judging the sensory difference of the reconstructed image through human eye and prior knowledge, such as the similarity between the texture, color and other features of the image and the original high-definition image. Subjective evaluation mainly depends on the artificial aesthetic standards, and there are some errors in image quality judgment due to various factors.

Objective evaluation is to quantitatively analyze image resolution and evaluate image quality using specific indicators. The common objective evaluation index includes peak signal to noise ratio (PSNR) and structural similarity index (SSIM). PSNR is to measure whether there is distortion in the reconstructed image by calculating the ratio of the maximum value of pixels to the power of additive noise. The larger the value is, the closer the performance and authenticity of the reconstructed image is to the original high-resolution image. This calculation method is directly related to the mean square error (MSE) of the image. The calculation formula is as follows (7).

$$MSE = \frac{1}{HW}\sum_{i=0}^{H-1}\sum_{j=0}^{W-1}[I^{SR}(i,j) - I(i,j)]^2 \tag{7}$$

$$PSNR = 10\lg(L^2/MSE) \tag{8}$$

where H, W is the size of the image, MSE is the mean square error, I^{SR} is the reconstructed image, I is the original image, L is usually taken as 255. The larger the PSNR value is, the smaller the distortion of the output image is, and the closer it is to the original image.

SSIM is a comprehensive measure of the similarity in structure, brightness and contrast between two or more images. The closer the value is to 1, the better the output image quality is. The specific formula is shown in formula (9).

$$S_{SSIM}(I, I^{SR}) = \frac{(2u_I u_{I^{SR}} + C_1)(2\sigma_{II^{SR}} + C_2)}{(u_I^2 + u_{I^{SR}}^2 + C_1)(\sigma_I^2 + \sigma_{I^{SR}}^2 + C_2)} \tag{9}$$

Where u_I and $u_{I^{SR}}$ represents the mean value of the original image I and the reconstructed image I^{SR}; σ_I and $\sigma_{I^{SR}}$ represents the variance of the original image I and the reconstructed image I^{SR}; $\sigma_{II^{SR}}$ represents the covariance of the two; C_1 and C_2 are constants.

3.3 Experiment analysis

The algorithm in this paper is compared with the representative methods in the field of single image super-resolution reconstruction, such as Bicubic algorithm, FSRCNN algorithm and VSDR algorithm. The subjective effect comparison is shown in Fig 7~Fig 9, and the red box is the interesting region.

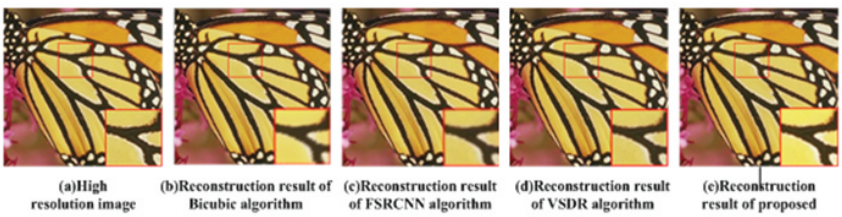


Fig 7 Reconstruction results of different algorithms on butterfly

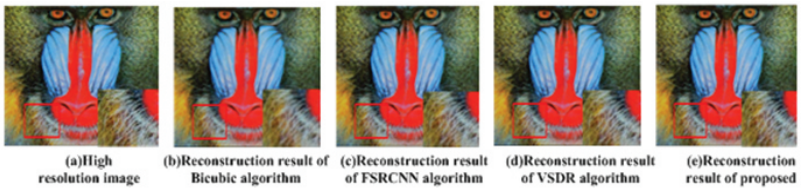


Fig 8 Reconstruction results of different algorithms on baboon

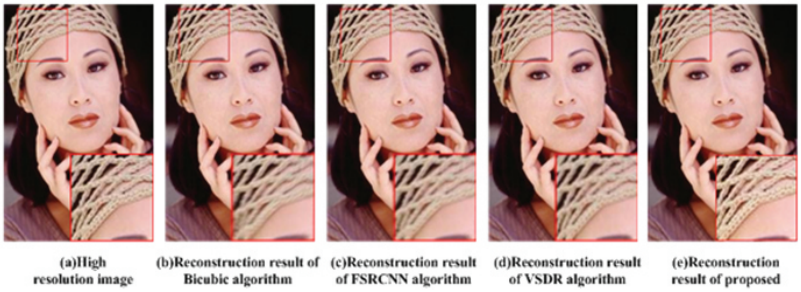


Fig 9 Reconstruction results of different algorithms on woman

It can be compared from Fig 7~Fig 9 and find that the local receptive field of the double Bicbuic algorithm is small, and the available regional features are single, and the high-resolution reconstructed image still has the disadvantage of edge definition, and the image reconstruction effect is poor. The processing effect of FSRCNN and VSDR algorithms on baboon's hair and mouth is worse than that of the original high-resolution image.

The algorithm of this paper has a good effect in distinguishing image edge and improving texture details, and has a good improvement on the clarity on head ornament texture details of the woman.

Tab 2 PSNR and SSIM results of different methods on set 5 dataset (magnification = 2,3,4)

set5	multiple	Bicubic algorithm		FSRCNN algorithm		VSDR algorithm		proposed algorithm	
		PSNR/dB	SSIM	PSNR/dB	SSIM	PSNR/dB	SSIM	PSNR/dB	SSIM
baby	2	38.32	0.961 4	39.13	0.963 1	38.91	0.913 8	38.42	0.968 9
	3	34.95	0.921 8	35.34	0.922 8	34.96	0.923 1	36.13	0.932 4
	4	33.12	0.891 1	33.02	0.886	33.27	0.891 7	34.13	0.890 2
bird	2	40.71	0.968 4	41.17	0.991 7	41.82	0.984 3	41.27	0.997 6
	3	35.21	0.949 2	36.29	0.952 3	36.31	0.948 3	35.69	0.951 4
	4	32.13	0.912 3	33.12	0.913 2	33.37	0.897 5	33.61	0.917 2
butterfly	2	32.27	0.951 4	32.79	0.967 5	32.82	0.958 9	34.03	0.978 4
	3	27.61	0.891 5	28.32	0.908 5	28.29	0.915 6	28.41	0.931 8
	4	25.11	0.841	26.05	0.857 3	25.98	0.852 3	26.03	0.862 5
head	2	35.64	0.890 1	36.14	0.885 1	36.09	0.876 4	36.27	0.894 2
	3	33.52	0.824	33.73	0.834 9	33.64	0.838 9	34.12	0.834 1
	4	32.16	0.779 2	32.25	0.784 1	32.31	0.792 3	32.23	0.788 1
woman	2	36.02	0.959 3	33.41	0.959 8	33.29	0.952 6	34.59	0.962 8
	3	31.14	0.923 1	32.2	0.938 1	32.31	0.937 6	32.57	0.939 8
	4	28.03	0.872 1	29.03	0.876 5	28.86	0.863 8	29.79	0.895 8
average	2	36.59	0.946 1	36.53	0.953 4	36.59	0.937 2	36.92	0.960 4
	3	32.49	0.901 9	33.18	0.911 3	33.10	0.912 7	33.38	0.917 9
	4	30.11	0.859 1	30.69	0.863 4	30.76	0.859 5	31.16	0.870 8

PSNR and SSIM is used to evaluate the proposed algorithm objectively, and is compared with Bicubic algorithm, FSRCNN algorithm and VSDR algorithm respectively. The scale factor of input image is magnified by 2, 3, 4 times in set 5 and set 14 test sets. After experiment comparison of different methods, the comparison results of reconstructed data are listed in Tab 2.

It can be seen from the table 2, compared with Bicubic algorithm, FSRCNN algorithm and VSDR algorithm, the average PSNR of proposed algorithm is increased by 0.33 dB, 0.338 dB and 0.324 dB respectively when the expansion factor is 2, 0.282 dB, 0.208 dB and 0.898 dB is increased when the expansion factor is 3, and 0.4 dB, 0.464 dB, 1.048 dB is increased when the expansion factor is 4. SSIM is increased by 0.041 3, 0.009 6 and 0.023 2 when the expansion factor was 2, 0.016 0, 0.006 6 and 0.005 2 when the expansion factor was 3, 0.011 6, 0.007 3 and 0.011 2 when the expansion factor is 4.

Tab 3 PSNR and SSIM results of different methods on set14 dataset (magnification = 2,3,4)

Image	multiple	Bicubic interpolation		SRCNN algorithm		FSRCNN algorithm		The algorithm of this paper	
		PSNR/dB	SSIM	PSNR/dB	SSIM	PSNR/dB	SSIM	PSNR/dB	SSIM
baboon	2	25.58	0.806 5	25.67	0.838 6	25.79	0.886 9	25.93	0.962 4
	3	23.21	0.683 2	23.42	0.725 5	24.23	0.728 3	24.58	0.752 9
	4	22.46	0.624 1	22.54	0.630 8	22.69	0.658 7	22.73	0.698 4
barbara	2	30.62	0.923 2	30.71	0.937 3	30.89	0.945 7	30.94	0.969 8
	3	26.27	0.788 2	26.63	0.858 9	26.91	0.838 6	27.89	0.870 8
	4	25.85	0.637 8	25.92	0.648 3	26.05	0.651 3	26.17	0.693 7
bridge	2	25.96	0.727 5	25.99	0.741 8	26.08	0.752 1	26.78	0.761 3
	3	25.39	0.796 1	26.88	0.832 4	27.13	0.809 1	27.94	0.850 6
	4	24.02	0.802 4	25.84	0.827 1	25.93	0.836	26.13	0.841 2
coast guard	2	30.6	0.798 5	30.53	0.800 4	30.67	0.812 9	30.91	0.842 7
	3	27.01	0.698 2	27.74	0.831 2	27.89	0.827 6	28.81	0.816 2
	4	25.29	0.593	25.36	0.613 8	25.72	0.628 1	26.24	0.629 3
comic	2	26.14	0.678 5	26.53	0.681 2	26.67	0.691 2	26.93	0.701 7
	3	23.09	0.792 4	25.81	0.891	26.32	0.912 5	28.84	0.908 3
	4	22.83	0.738 4	22.89	0.745 6	22.94	0.746 9	24.05	0.751 2
face	2	33.59	0.845 7	33.68	0.857 6	33.75	0.891 3	33.81	0.898 6
	3	32.76	0.819 4	34.23	0.871 2	34.73	0.858 2	26.12	0.878 2
	4	31.26	0.765 2	31.36	0.773 2	31.84	0.780 5	33.94	0.789 3
flowers	2	29.46	0.913 4	29.57	0.921 5	29.68	0.930 7	29.83	0.931 2
	3	28.12	0.878 1	29.05	0.918 9	30.87	0.896 7	34.29	0.919 3
	4	26.34	0.753 1	26.47	0.765 2	26.58	0.780 3	28.13	0.790 4
average	2	28.85	0.813 3	28.95	0.825 4	29.07	0.844 4	29.30	0.866 8
	3	26.55	0.779 3	27.68	0.847 0	28.297 1	0.838 7	28.352 8	0.856 6
	4	25.4357	0.702	25.768 5	0.714 8	25.964 2	0.725 9	26.77	0.741 9

It can be seen from the table 3 compared with Bicubic algorithm, FSRCNN algorithm and VSDR algorithm, the average PSNR of proposed algorithm is increased by 0.45 dB, 0.35 dB and 0.23 dB respectively when the expansion factor is 2, 1.8 dB, 0.67 dB and 0.05 dB is increased when the expansion factor is 3, and 1.3 dB, 1.0 dB and 0.8dB is increased when the expansion factor is 4. SSIM is increased by 0.05, 0.04 and 0.02 when the expansion factor was 2, 0.8, 0.0096 and 0.001 7 when the expansion factor was 0.039 9, 0.027 and 0.015 when the expansion factor is 4. From the quantitative data analysis, we can see that the sawtooth dilated residual convolution can enhance the texture information and reconstruction effect of the reconstructed image to a certain extent, reduce the operation time and improve the image reconstruction efficiency.

4 Conclusion

In this paper, an improved image super-resolution reconstruction method is proposed based on VSDR algorithm. Firstly, the low-resolution image is interpolated processing into the network, and then the image features are extracted by one-time convolution. Secondly, the nonlinear mapping of image features is realized by using 6 continuous sawtooth dilated residual convolutions, and can effectively avoid the “grid” phenomenon which is caused by the same expansion rate. The receptive field is expanded without changing the image size, and get the output residual image, and the problems of network gradient disappearance and gradient explosion is solved. Finally, the original low-resolution image and the residual image are linearly added to output the final high-resolution image, which improves the quality and efficiency of reconstruction. The experiment results show that: compared with the other three methods, the proposed algorithm achieves better results in PSNR and SSIM for a single image. Compared with the original color high-resolution image, there is still a gap in texture and clarity. In the next research work, the sawtooth dilated convolution is applied to a better deep learning framework, and more expansion rate combinations are tried to achieve image super-resolution reconstruction and better use of image features.

基于锯齿空洞残差卷积的单幅图像超分辨率重建研究*

李 岚，蔺国梁，马少斌
(兰州文理学院 数字媒体学院，甘肃 兰州 730000)

摘 要：针对残差学习的超分辨率重建方法中存在感受野受限、分辨率低、复杂性较高、边缘信息丢失等问题，提出一种锯齿空洞残差卷积的神经网络. 首先，基于ResNet网络设计了锯齿空洞卷积，扩大网络的感受野，消除网络的“网格化”，并增加跳跃连接，将图像特征传递到更深的网路；然后，通过最后一个卷积层得到与原始图像大小相等的残差图像；最后，将输入LR图像与残差图像进行线性融合输出最终的超分辨率图像. 在set5和set14数据集上的实验数据表明：与现有算法相比，本文算法具有更好的重建效果，学习性能有较大提高.

关键词：残差网络；锯齿空洞卷积；深度学习；图像超分辨率重建
DOI：10.13568/j.cnki.651094.651316. 2020.07.30.0002
中图分类号：TP391 **文献标识码**：A **文章编号**：2096-7675(2021)02-0174-17
引文格式：李岚, 蔺国梁, 马少斌. 基于锯齿空洞残差卷积的单幅图像超分辨率重建研究[J]. 新疆大学学报(自然科学版)(中英文), 2021, 38(2): 174-190.

0 引 言

超分辨率图像重建（Super Resolution Image Reconstruction, SRIR 或SR）的思想是利用一组低质量、低分辨率图像（Low Resolution, LR）经过计算机技术和图像处理技术产生单帧或多帧高质量、高分辨率图像[1,2]. SR是数字图像处理的一个重要研究方向，在计算机视觉领域有广泛的应用，如智能交通、安全监控、图像生成和医学成像等[3]. 目前SR方法主要分为3类：基于差值的方法[4]、基于重建的方法[5]和基于学习的方法[6,7]. 基于差值的方法是将图像看作一个点，利用先验知识由一个预定义的变换函数或插值来拟合平面上未知的信息，从而计算出高分辨率图像，该方法的主要缺点是很容易出现阶梯锯齿状现象及边缘模糊现象. 基于重建的方法是目前被广泛研究的方法之一，这类方法主要是应用下采样技术使得一个或者多个低分辨率图像信息进行融合低像素精度的多帧信息，重建出更高分辨率的图像. 这种方法重构精度高，但只能够利用高、低分辨率图像之间的关系，数学模型构建比较困难，导致重建图像纹理不清晰.

基于深度学习的方法近年来成为研究的热点，主要是利用相同图像的内部相似性及大量的训练样本数据，从中求出高、低分辨率图像对之间的映射关系，完成高分辨率图像特征的转化，从而实现SR的过程，这种方法对建模数据的集中特性要求非常高. Dong等[8]首次采用深度学习方法，网络总共设置了一个3层的卷积神经网络(Convolution Neural Network,CNN)实现超分辨率重建，并取得了良好的效果，从此开启了深度学习实现SR的热潮. 这种方法在训练的过程中随着网络层数的增加，会出现超参数过多、梯度弥散/爆炸等问题，重建后的图像通常过于平滑，丢失了高频细节信息，图象质量仍有待提升.

Kim等[9]提出深度残差网络（VDSR）模型，利用残差学习加快网络收敛速度，证明了该方法对超分辨率的性能提升，但有增加计算的复杂度和梯度消失问题. Yang等[10]利用图像的稀疏性，约束高低分辨率图像对应的字典下的稀疏表示实现图像超分辨率重建，重建效果较好，其缺点是字典训练需要花费较长时间，并且在图像边缘会出现噪声. Tai等[11]提出DRRN模型，运用递归使用权重共享的网络模块，增加网络深度至52层，减少了模型的参数，但是每个递归单元优化不够，重建效果不够明显. Yu 等[12]和Chen等[13]提出了空洞卷积（Dilated Convolution），通过在网络中不增加参数而改变卷积核大小的情况下获得更大的感受野，获取更多的原始图像信息，在重建中取得较好的效果，但该方法会出现“网格化”现象，卷积后导致更多图像信息丢失.

本文对以上方法进行了改进，提出了一种残差网络与锯齿空洞卷积相结合的图像超分辨率重建方法. 该模型利用空洞卷积网络提取图像特征，然后运用残差网络结合锯齿空洞卷积进行图像非线性映射，再应用卷积网

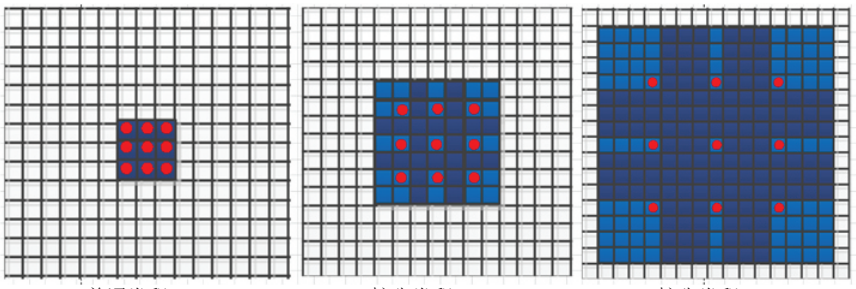
* **收稿日期**：2020-07-30
基金项目：2020年甘肃省高等教育教学成果培育项目；2019年甘肃省创新创业项目；2018年甘肃高等学校科研项目；2019年甘肃省教育厅产业支撑引导项目.
作者简介：李岚（1978-），女，硕士，副教授，从事深度学习、智能信息处理领域的研究，E-mail：148439473@qq.com.

络得到与输入图像大小相等的残差图像，最后将预处理的低分辨率图像与残差图像线性融合得到最终的超分辨率图像. 网络训练使用自适应时刻估计方法（Adaptive Moment Estimation，Adam）加速网络收敛，通过空洞卷积操作扩大了图像特征感受野的同时高质量地恢复了图像的纹理信息，提升了重建图像的视觉效果.

1 相关工作

1.1 空洞卷积

空洞卷积（Dilated Convolution）也称扩张卷积，是一种在特征图上进行数据采样的方式. 通过在普通卷积核的每个像素之间补充0像素值，用来增加网络的扩张系数，可以在卷积核有效增大感受野的同时不增加模型参数或者计算量. 在图像全局信息或者语音文本需要较长的sequence信息依赖的问题中，都能较好地应用空洞卷积[10]. 对于一个3*3的卷积网络，其扩张率和感受野如图1所示.



（a）普通卷积$r_{\text{rate}}=1$　（b）扩张卷积$r_{\text{rate}}=2$　（c）扩张卷积$r_{\text{rate}}=4$

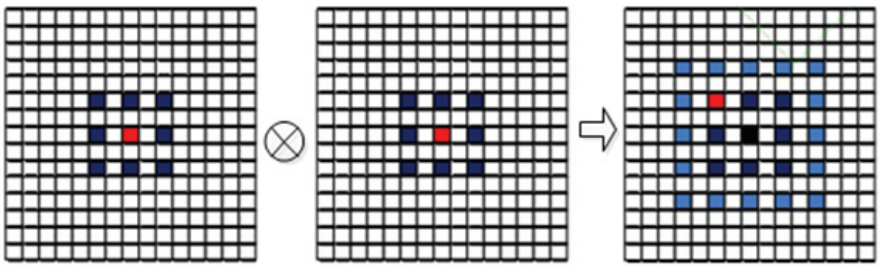
图 1　不同扩张率的3×3 pixels 卷积

图1（a）是普通的卷积（即扩张率r_{rate}=1的空洞卷积），代表的数据是原图3*3元素的卷积，感受野是3*3，相当于不扩充，r_{rate}=1；图1（b）是扩张率r_{rate}=2的空洞卷积，感受野是7*7，dilation=2；图1（c）是采用扩张率r_{rate}=4的空洞卷积，感受野是15*15，dilation =4. 空洞卷积对普通卷积的改进就是为了获得更大的感受野，感受野的大小计算用公式（1）表示.

$$v=((k_{\text{size}}-1)(r_{\text{rate}}-1)+k_{\text{size}})^2 \tag{1}$$

其中：k_{size}表示卷积核的大小，r_{rate}表示空洞卷积的扩张率，v表示感受野的大小.

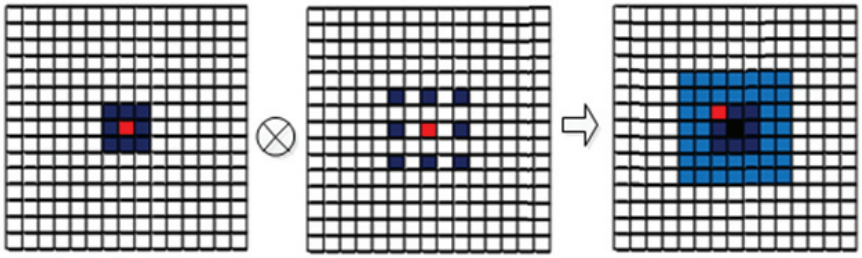
空洞卷积不仅可以保留图像内部的特征信息，还能避免pooling造成的分辨率丢失，但其仍存在缺陷：相同的空洞卷积得到某一层的结果中，邻近的像素是从相互独立的子集中卷积得到的，相互之间缺少依赖，造成信息不连续，会出现“网格”效应，具体卷积如图2所示.



（a）r_{rate}=2 的感受野　（b）r_{rate}=2 的感受野　（c）卷积后的感受野

图 2　相同扩张率的空洞卷积

图2（a）对应扩张率$r_{\text{rate}}=2$的卷积层，感受野为5×5；图2（b）表示第二层卷积核，感受野为5×5 ，图2（c）为扩张率$r_{\text{rate}}=2$的空洞卷积进行一次卷积的结果，感受野为9. 如果所有空洞卷积都使用相同的扩张率，则计算方式类似于棋盘格，大范围的数据相互之间没有依赖关系，随着深度网络的层数进一步增加，导致重要信息丢失. 针对这种问题，本文提出一种锯齿混合空洞卷积，如图3所示.



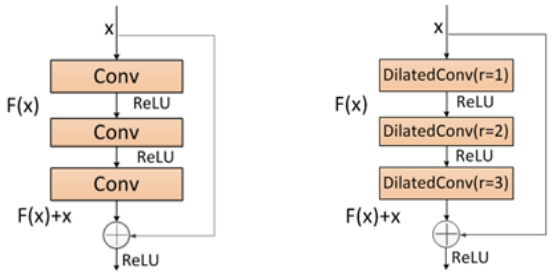
（a）r_{rate} =1 的感受野　（b）r_{rate} =2 的感受野　（c）卷积后的感受野

图 3　不同扩张率的卷积结果

从图3可以看出，锯齿空洞卷积在增大感受野的同时能保证无盲区覆盖整个区域，从而能够有效提取特征，提高重建的准确率. 由于插入0像素值不参与卷积运算，参与运算的仍是相同尺寸的卷积核，复杂度不变.

1.2　残差网络（ResNet）

对于CNN，如果简单地增加深度，会导致梯度弥散或梯度爆炸. He等[14]提出了残差学习网络ResNet，通过增加跳跃连接（恒等映射）来直接连接浅层网络与深层网络. 在ResNet中每个卷积层后面都有一个批归一化（Batch Normalization），文献[15]中指出，BN层可以将特征归一化处理，提高网络的泛化能力，加速训练的收敛过程，但在一定程度上破坏了图像的空间信息，增加网络的训练参数，导致网络性能变差. 所以，本文对ResNet进行了改进，残差单元中卷积层采用空洞卷积，并移除了BN层，有助于获得更好的图像超分辨率重建结果.



（a）原始残差块　（b）本文空洞卷积残差块

图 4　残差网络单元

如果采用同一个扩张率进行卷积后，卷积核会产生“网格”现象，同时卷积核体积过大会携带冗余信息，造成不必要的内存占用. 因此，本文设计了锯齿空洞残差单元，通过循环运算获得与标准卷积相同结构的特征图，在保留图像细节信息的同时可以增大网络的感受野，进而更好地拟合目标边界，提高图像重建质量. 本文中的锯齿空洞残差单元与ResNet的残差单元结构如图4所示.

2　本文网络及实现过程

本文算法的思想是为了获取更多的图像信息，使用较小的卷积核，既不增加网络的层数和复杂度，又能够加快网络的收敛速度，同时避免了因为空洞出现的“盲区”现象. 首先在残差网络中采用不同扩张率的空洞卷积对网络的输入图像进行训练，输出残差图像，然后将输入的低分辨率图像与残差网络输出的残差图像相加重建超分辨率图像.

2.1　网络结构

随着网络深度的不断增加，人们发现深度CNN网络并不是网络层数越深性能就越好，当达到一定的深度后继续增加层数反而会影响网络的收敛速度，感受野会减小，导致上下文信息减少，重建效果差. 何凯明等[14]2015年提出深度残差网络（Deep Residual Network,ResNet），在ImageNet图像分类中通过增加网络深度可以提高网络分类的准确性，通过残差学习解决梯度消失和网络性能退化问题.

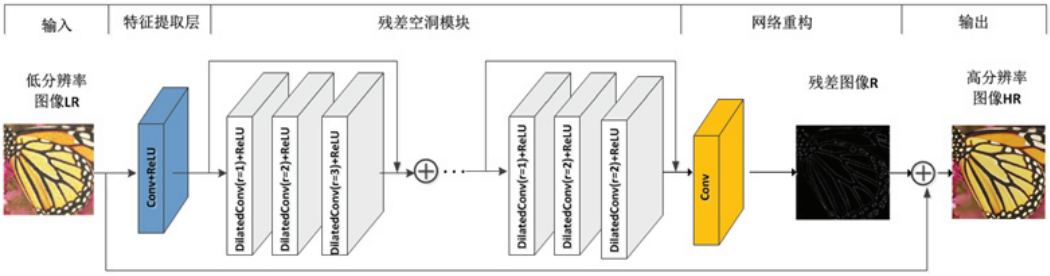


图 5　本文单幅图像超分辨率重建方法网络结构

本文基于文献[15]构建了一个锯齿洞卷积的残差网络，网络整体结构如图5所示，共包含20个卷积层，按照功能划分为5个部分，分别是低分辨率图像输入部分，卷积特征提取部分，紧接着是6个空洞残差块，每个残差块包含3个残差单元，然后是残差图像特征层，通过空洞残差使每个卷积层输出的特征图保持尺寸不变.

2.2 图像超分辨率重建过程

网络输入由图像单一颜色通道形成的数据经双三次差值预处理.

本文提出的空洞残差卷积神经网络超分辨率重建主要包括以下3个步骤：

（1）特征提取层（Conv+ReLU层）

除了最后一层卷积核大小为1个像素外，其余中间各层卷积特征尺寸都为3×3×64个像素点. 在特征提取层中将输入的低分辨率图像X与3×3×1个像素大小的卷积核进行卷积操作，得到n_1个特征图（这里n_1=224），这里的1表示通道数. 特征提取层包含一个卷积层和一个激活函数，并将该层获得的每个神经元传递到残差单元模块. 特征提取过程的表达式为：

$$B_0 = f(W_1 * X + b_1) \tag{2}$$

其中：*为卷积操作，W_1为第一个卷层的卷积核；b_1为第一个卷积层的偏置项，其维度与卷积核的维度保持一致；B_0为从输入X中提取的第一个残差块的输入特征；f为ReLU激活函数.

本文网络中应用ReLU可以加快模型训练速度和缩短模型收敛时间，同时一定程度上抑制梯度消失现象，性能优于传统的激活函数Sigmoid[16]（在深度卷积网络中梯度反向传递时导致梯度爆炸和梯度消失），表达式为：

$$f(x_i) = \begin{cases} x_i, & x_i > 0 \\ 0, & x_i < 0 \end{cases} \tag{3}$$

其中：x_i为ReLU函数的输入，$f(x_i)$为ReLU函数的输出.

（2）非线性映射

对输入的低分辨率图像，采用提出的锯齿残差网络进行图像训练. 该残差模块结构由6个空洞残差单元组成，每个空洞残差单元由3个空洞卷积层和3个非线性激活函数ReLU组成. 每个残差单元设计有两部分：跳跃连接和恒等映射. 这样既可以保留残差信息，又能通过跳跃连接将图像特征向后传递，有助于保持特征的多样性.

采用3×3 像素的卷积核，扩张率的选择采用1、2、3的3个卷积层进行串联成为一个残差块，为了不改变卷积核的大小，在网络计算结束时保证输出为13×13的特征图. 经过空洞残差块之后特征图像素点的感受野得到了扩张，获取更多的原始图像信息，每个残差单元的表达式为：

$$H_m = G_m(H_{m-1}) = F(H_{m-1}, W_m) + H_{m-1} \tag{4}$$

式中：H_{m-1}和H_m分别为第m个残差单元的输入与输出，F为学习到的残差映射，即：

$$F(H_{m-1}, W_m) = W_m^2 * f(W_m^i * H_{m-1}) \tag{5}$$

其中：W_m^i，i=1，2，3为学习到的第i个卷积层权重. 为简化公式，省略了偏置项，f为ReLU函数，$*$为卷积操作.

由于在深度卷积网络中设置了锯齿空洞卷积，经过网络训练，减少了“网格化”现象导致的图像信息不连续的问题.

（3）图像重构与输出

首先将残差网络输出的特征图作为最后一个卷积层的输入，与3×3的卷积核进行卷积生成与输入图像大小相等的残差图像，然后与输入的插值图像进行线性叠加，输出最终的超分辨率图像. 图像重构阶段的结构如图6所示.

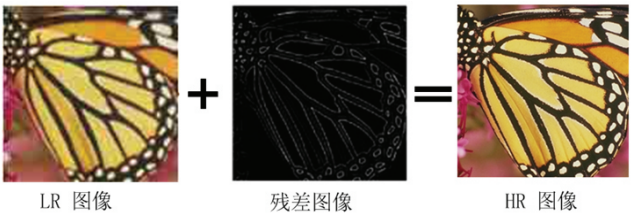


图 6 图像的重构阶段

2.3　损失函数

网络模型应用了空洞残差，使得输入图像与输出图像大小相等，同时解决网络难以收敛的问题. 以x表示双三次差值后的输入图像，$f_{\mathrm{Res}}(x)$表示输入图像x经过整个神经网络以后输出的残差图像，y表示原始高分辨率图像，y^*表示网络预测输出的超分辨率图像，即

$$y^* = x + f_{\mathrm{Res}}(x)$$

本文网络中应用均方误差（Mean Squared Error，MSE）函数作为整个网络的损失函数，通过计算生成的图像和原始高分辨率图像的均方误差达到最小值，得到最优解. 损失函数计算公式为

$$\mathrm{Loss}(\Theta) = \frac{1}{n}\sum_{i=1}^{n} \|y_i - f(x_i;\Theta)\|^2 \tag{6}$$

其中：n为样本数，y_i为高分辨率图像，$f(x_i;\Theta)$为网络的预测输出图像，$\Theta = \{w_1, w_2, \cdots, b_1, b_2, \cdots\}$.

3　实验与分析

3.1　实验环境

本文图像超分辨率重建算法的实验环境如表1所示.

表 1　本文实验环境

编程环境	服务器硬件配置	服务器软件环境
编程语言：MatlabR2016a	CPU：Intel(R) Core(TM)i7-8700k 3.7 GHz	系统：Windows 7操作系统
深度学习框架：Caffe[17]	显卡：11 G GTX1080	
	GPU：NVIDIA Titan CUDA	
	RAM ：16 G	

3.2　实验设置与评价指标

由于网络相对较深，算法要使用更大的训练集才能训练出更好的结果，分别选取与文献[13]相同的291张图片作为训练集，这些图片分别来自文献[18]提出的91张图片和BSD（Berkeley Segmentation Dataset）训练集中的200张图片[11]，共291张图像. 为了充分利用深度图像，对数据集图像进行水平翻转、垂直翻转和水平垂直翻转，并按照0.9、0.8的系数缩放，然后保存图片，共生成5 820张图像，图像尺寸大小不超过512×512.

在应用本文网络对图像进行训练的过程中，将训练图像转换到YCbCr空间[19]. 相对于颜色（CbCr通道信息）变化，人类对亮度（Y通道）的变化更敏感，所以本文设计的网络中只对Y通道做处理. 图像卷积核大小都为3×3，特征图像通道数量为64. 训练所采用的优化算法是Adam[20]与SGD（Stochastic Gradient Descent, SGD）相比较，Adam优化方法更灵活，具有自适应性，可以将每次迭代的学习率控制在一定范围内，使得参数学习比较稳定. 设置网络的初始学习率为10^{-4}，动量参数设置为0.9，采用mini-batch训练方式，batch-size为64，每10万次迭代降低为原来的一半.

实验中用Set5和Set14作为测试集，原始高分辨率图像为X，放大倍数取s=2,3,4，预处理后的图像作为网络的输入.

为了验证图像超分辨率重建方法是否符合预期，需要对重建后的高分辨率图像进行评价. 目前，重建图像的质量评价方法分为两种：主观评价和客观评价. 主观评价主要是指通过人眼观测和先验知识判断重建图像的感官差异，如图像的纹理、颜色等特征与原始高清图像的近似程度. 主观评价主要依靠人为审美标准，受多方面因素影响，对图像质量判断存在一定的误差. 客观评价是对图像分辨率进行定量分析，通过具体的指标评价图像质量. 常见的客观评价指标有峰值信噪比（Peak Signal to Noise Ratio，PSNR）和结构相似比（Structural Similarity Index，SSIM）. PSNR是通过计算图像内像素最大值与加性噪声功率的比值来衡量重建图像是否存在失真问题，其数值越大，说明重建图像的性能和真实程度越接近原始高分辨率图像，其计算方法与图像的均方误差（MSE）有直接关系，计算公式如下式（7）所示.

$$MSE = \frac{1}{HW}\sum_{i=0}^{H-1}\sum_{j=0}^{W-1}[I^{SR}(i,j) - I(i,j)]^2 \tag{7}$$

$$PSNR=10\lg(L^2/MSE) \tag{8}$$

其中：H、W为图像的尺寸，MSE为均方根误差，I^{SR}为重建图像，I为原始图像，L的值通常取255. PSNR的值越大，表示输出图像失真越小，与原始图像越接近.

SSIM是综合衡量两幅或者多幅图像之间结构度、亮度和对比度的相似程度，其值与1进行比较，越接近于1，表示输出图像质量越好，其具体公式如式（9）所示.

$$S_{\mathrm{SSIM}}(I,I^{SR})=\frac{(2u_Iu_{I^{SR}}+C_1)(2\sigma_{II^{SR}}+C_2)}{(u_I^2+u_{I^{SR}}^2+C_1)(\sigma_I^2+\sigma_{I^{SR}}^2+C_2)} \tag{9}$$

其中：u_I和$u_{I^{SR}}$表示原始图像I和重建图像I^{SR}的均值；σ_I和$\sigma_{I^{SR}}$表示原始图像I和重建图像I^{SR}的方差；$\sigma_{II^{SR}}$表示两者的协方差；C_1和C_2是常数.

3.3 实验分析

首先将本文算法分别与Bicubic算法、FSRCNN算法和VSDR算法等单幅图像超分辨率重建领域中具有代表性的方法进行对比实验，主观效果对比如图7~9所示，其中红色框为感兴趣的区域.

图 7 不同算法在Butterfly上的重建结果

图 8 不同算法在Baboon上的重建结果

图 9 不同算法在Woman上的重建结果

比较图7~9中可以看出双Bicbuic算法局部感受野较小，可利用的区域图像特征单一，高分辨率重建的图像仍然存在边缘清晰度差的缺点，图像重建效果差. FSRCNN和VSDR算法对图像Baboon的毛发和嘴周围处理效果比原始高分辨率图像差.

本文算法在区分图像边缘和改善纹理细节中有较好的效果，对Woman的重建中头部饰品纹理细节在清晰度上有较好的提升. 采用PSNR和SSIM两种指标对本文算法进行客观评价，分别与Bicubic算法、FSRCNN算法和VSDR算法进行比较，在Set5、Set14测试集上将输入图像的尺度因子放大2，3，4倍，经过不同方法的实验对比，重建的数据对比结果如表2所列.

从表2中可以看出，本文算法相较于Bicubic算法、FSRCNN和VSDR算法在扩大因子为2时，PSNR均值上分别提升了0.33、0.338和0.324 dB；扩大因子为3时提升了0.282、0.208、0.898；扩大因子为4时提升了0.4、0.464、1.048. 在扩大因子为2时，SSIM平均提升了0.041 3、0.009 6和0.023 2；扩大因子为3时提升了0.016 0、0.006 6、0.005 2；扩大因子为4时提升了0.011 6、0.007 3、0.011 2.

从表3中可以看出，本文算法相较于Bicubic算法、FSRCNN和VSDR算法PSNR值在扩大因子为2时，均值上分别提升了0.45 dB、0.35 dB和0.23 dB；扩大因子为3时提升了1.8 dB、0.67 dB和0.05 dB；扩大因子为4时提升了13 dB、10 dB和0.8 dB. SSIM值在扩大因子为2时，平均提升了0.05、0.4 和 0.02，扩大因子为3时提升了0.8、0.009 6、0.001 7；扩大因子为4时提升了0.399、0.27、0.015. 由定量数据分析可知，本文算法引入锯齿空洞残差卷积在一定程度上增强了重建图像的纹理信息和重建效果，缩减了运算时间，提高了图像重建效率.

表 2　不同方法在set5数据集上的PSNR和SSIM结果（放大倍数=2,3,4）

set5	倍数	Bicubic算法		FSRCNN算法		VSDR算法		本文算法	
		PSNR/dB	SSIM	PSNR/dB	SSIM	PSNR/dB	SSIM	PSNR/dB	SSIM
baby	2	38.32	0.961 4	39.13	0.963 1	38.91	0.913 8	38.42	0.968 9
	3	34.95	0.921 8	35.34	0.922 8	34.96	0.923 1	36.13	0.932 4
	4	33.12	0.891 1	33.02	0.886	33.27	0.891 7	34.13	0.890 2
bird	2	40.71	0.968 4	41.17	0.991 7	41.82	0.984 3	41.27	0.997 6
	3	35.21	0.949 2	36.29	0.952 3	36.31	0.948 3	35.69	0.951 4
	4	32.13	0.912 3	33.12	0.913 2	33.37	0.897 5	33.61	0.917 2
butterfly	2	32.27	0.951 4	32.79	0.967 5	32.82	0.958 9	34.03	0.978 4
	3	27.61	0.891 5	28.32	0.908 5	28.29	0.915 6	28.41	0.931 8
	4	25.11	0.841	26.05	0.857 3	25.98	0.852 3	26.03	0.862 5
head	2	35.64	0.890 1	36.14	0.885 1	36.09	0.876 4	36.27	0.894 2
	3	33.52	0.824	33.73	0.834 9	33.64	0.838 9	34.12	0.834 1
	4	32.16	0.779 2	32.25	0.784 1	32.31	0.792 3	32.23	0.788 1
woman	2	36.02	0.959 3	33.41	0.959 8	33.29	0.952 6	34.59	0.962 8
	3	31.14	0.923 1	32.2	0.938 1	32.31	0.937 6	32.57	0.939 8
	4	28.03	0.872 1	29.03	0.876 5	28.86	0.863 8	29.79	0.895 8
average	2	36.59	0.946 1	36.53	0.953 4	36.59	0.937 2	36.92	0.960 4
	3	32.49	0.901 9	33.18	0.911 3	33.10	0.912 7	33.38	0.917 9
	4	30.11	0.859 1	30.69	0.863 4	30.76	0.859 5	31.16	0.870 8

表 3　不同方法在set14数据集上的PSNR和SSIM结果（放大倍数=2,3,4）

图像	倍数	双三次插值法		SRCNN算法		FSRCNN算法		本文算法	
		PSNR/dB	SSIM	PSNR/dB	SSIM	PSNR/dB	SSIM	PSNR/dB	SSIM
baboon	2	25.58	0.806 5	25.67	0.838 6	25.79	0.886 9	25.93	0.962 4
	3	23.21	0.683 2	23.42	0.725 5	24.23	0.728 3	24.58	0.752 9
	4	22.46	0.624 1	22.54	0.630 8	22.69	0.658 7	22.73	0.698 4
barbara	2	30.62	0.923 2	30.71	0.937 3	30.89	0.945 7	30.94	0.969 8
	3	26.27	0.788 2	26.63	0.858 9	26.91	0.838 6	27.89	0.870 8
	4	25.85	0.637 8	25.92	0.648 3	26.05	0.651 3	26.17	0.693 7
bridge	2	25.96	0.727 5	25.99	0.741 8	26.08	0.752 1	26.78	0.761 3
	3	25.39	0.796 1	26.88	0.832 4	27.13	0.809 1	27.94	0.850 6
	4	24.02	0.802 4	25.84	0.827 1	25.93	0.836	26.13	0.841 2
coastguard	2	30.6	0.798 5	30.53	0.800 4	30.67	0.812 9	30.91	0.842 7
	3	27.01	0.698 2	27.74	0.831 2	27.89	0.827 6	28.81	0.816 2
	4	25.29	0.593	25.36	0.613 8	25.72	0.628 1	26.24	0.629 3
comic	2	26.14	0.678 5	26.53	0.681 2	26.67	0.691 2	26.93	0.701 7
	3	23.09	0.792 4	25.81	0.891	26.32	0.912 5	28.84	0.908 3
	4	22.83	0.738 4	22.89	0.745 6	22.94	0.746 9	24.05	0.7512
face	2	33.59	0.845 7	33.68	0.857 6	33.75	0.891 3	33.81	0.898 6
	3	32.76	0.819 4	34.23	0.8712	34.73	0.858 2	26.12	0.878 2
	4	31.26	0.765 2	31.36	0.773 2	31.84	0.780 5	33.94	0.789 3
flowers	2	29.46	0.913 4	29.57	0.921 5	29.68	0.930 7	29.83	0.9312
	3	28.12	0.878 1	29.05	0.918 9	30.87	0.896 7	34.29	0.919 3
	4	26.34	0.753 1	26.47	0.765 2	26.58	0.780 3	28.13	0.790 4
average	2	28.85	0.813 3	28.95	0.825 4	29.07	0.844 4	29.30	0.866 8
	3	26.55	0.779 3	27.68	0.847 0	28.297 1	0.838 7	28.352 8	0.856 6
	4	25.435 7	0.702	25.768 5	0.714 8	25.964 2	0.725 9	26.77	0.741 9

4 结 论

本文提出了一种基于VSDR算法改进的锯齿空洞残差卷积的图像超分辨率重建方法. 该方法首先将低分辨率图像插值预处理后输入网络，通过一次卷积提取图像特征，然后应用6个连续的锯齿空洞残差卷积实现图像特征的非线性映射，有效避免了相同扩张率产生的“网格化”现象而出现的图像盲区，在扩大感受野的同时不改变图像尺寸，得到输出的残差图像，解决了网络梯度消失和梯度爆炸问题. 最后将原始的低分辨率图像与残差图像进行线性相加，输出最终的高分辨率图像，提高了重建质量和效率. 实验结果表明：与其他3种方法相比，对于单幅图像本文算法在峰值信噪比和结构相似比上达到了更优的效果. 与原始彩色高分辨率图像比较，在纹理和清晰度上还有差距，下一步的研究工作中，将锯齿空洞卷积应用到更优的深度学习框架中，尝试更多的扩张率组合，实现图像超分辨率重建，更好地利用图像特征.

参考文献：

[1] 苏衡, 周杰, 张志浩. 图像超分辨率重建方法综述[J]. 自动化学报, 2013, 39(8)：1202-1213.

[2] 杨东旭, 赖惠成, 班俊硕, 等. 基于改进DCNN结合迁移学习的图像分类方法[J]. 新疆大学学报(自然科学版), 2018, 35(2): 195-202.

[3] THÉVENAZ P, Blu T, UNSER M. Handbook of Medical Imaging Processing and Analysis[M]. [S. l.]: Academic Press, 2000.

[4] 王知人, 谷昊晟, 任福全, 等. 基于深度卷积残差学习的图像超分辨[J]. 郑州大学学报(理学版), 2020, 53(3): 42-48.

[5] 雷为民, 王玉楠, 李锦环. 基于FSRCNN的图像超分辨率重建算法优化研究[J]. 传感器与微系统, 2020, 39(2): 54-57.

[6] 周雷, 阿里甫·库尔班, 吕情深, 等. 新疆大学软件学院基于R-FCN的中国手指语识别[J]. 新疆大学学报（自然科学版）（中英文）, 2020, 37(2): 170-176.

[7] 张圣祥, 郑力新, 朱建清, 等. 采用深度学习的快速超分辨率图像重建方法[J]. 华侨大学学报（自然科学版）, 2019, 40(2): 245-250.

[8] DONG C, LOY C C, HE K, et al. Image super-resolution using deep convolutional networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2016, 38(2): 295-307.

[9] KIM J, KWON L J, MU L. Accurate image super-resolution using very deep convolutional networks[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE, 2016: 1646-1654.

[10] YANG J C, WRIGHT J, HUANG T S, et al. Image super-resolution via sparse representation [J]. IEEE Transactions on Image Processing, 2010, 19(11): 2861-2873.

[11] TAI Y, YANG J, LIU X. Image super-resolution via deep recursive residual network[C]// IEEE Computer Vision and Pattern Recognition (CVPR 2017). IEEE, 2017. DOI: 10. 1109.

[12] YU F, KOLTUN V. Multi-scale context aggregation by dilated convolutions[J/OL]. arXiv: 1511. 07122, (2016-04-30) [2017-05-16]. https: //arxiv. org /abs /1511. 07122.

[13] CHEN L C, PAPANDREOU G, KOKKINOS I, et al. Deep Lab: semantic image segmentation with deep convolutional nets, atrousconvolution, and fully connected CRFs[J/OL]. IEEE Transactions on Pattern Analysis &Machine Intelligence, 2017[2017-05-26]. http: // ieeexplore. Ieee. org /document /7913730 /.

[14] HE K, ZHANG X, R EN S, et al. Deep residual learning for image recognition [C]// Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 770-778．[DOI: 10．1109 / CVPR. 2016. 90]

[15] YANG Z, ZHANG K, LIANG Y, et al. Single image super-resolution with a parameter economic residual-like convolutional neural network[C]//International Conference on Multimedia Modeling. Springer, Cham, 2017: 353-364.

[16] SHI W, CABALLERO J, HUSZAR F, et al. Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network[C]//Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition．Piscataway, NJ: IEEE, 2016: 1874 -1883.

[17] JIA Y Q, SHELHAMER E, DONAHUE J, et al. Caffe: convolutional architecture for fast feature embedding[C/OL]. [2018-10-22]. https// arxiv. org / pdf / 1408. 5093. pdf.

[18] DONG C, CHEN C L, HE K, et al. Learning a deep convolutional network for image super-resolution[C]//European Conference on Computer Vision. Cham: Springer International Publishing, 2014: 184-199.

[19] 李岚, 张云, 马少斌. 改进的梯度与肤色融合均值移动粒子滤波人脸跟踪[J]. 延边大学学报(自然科学版), 2018, 44(2): 139-142.

[20] K D, BA J. Adam: A method for stochastic optimization[C]//The International Conference on Learning Representations, San Diego, USA, 2015.

责任编辑：闫新云