Image Super-resolution Reconstruction Algorithm Based on Convolutional Neural Network

He Jingxuan, Zhang Jian, Zhang Yonghui, Wang Rong
Hainan University, Haikou Hainan, China
180534@hainu.edu.cn

***Abstract*—Aiming at the problem that the existing SRCNN algorithm has too long training time, poor reconstruction performance and slow running speed, a new image super-resolution reconstruction algorithm based on convolutional neural network is proposed. The algorithm uses low-resolution images as the network input, the higher-order representation of the image is learned using the convolution operation, the high-resolution image is up-sampling by the deconvolution operation, and the residual structure is added to the network, so that the entire network can converge better. Experimental results on Set5, Set14, and BSD200 test sets show that compared with Bicubic, SRCNN and other methods, the method proposed in this paper is more efficient on super-resolution reconstruction of images and the convergence speed of the network is faster.**

Keywords—Image Processing, Convolutional neural network, Super-resolution reconstruction, Residual network, Deconvolution

I. INTRODUCTION

Image resolution is an important index to evaluate image quality, which largely determines the amount of information contained in the image.Therefore how to generate high-resolution images from low-resolution images is vital. Image super-resolution technology [1] is the technique of recovering high-resolution images or image sequences from a single frame or multi-frame of image. According to the existing literature, the technology of super-resolution based on learning method has become the mainstream method of single frame image super-resolution reconstruction as is superior to the interpolation method [2] and the reconstruction-based method [3] in both image quality and image type. Based on the learning method, the mapping relationship between low-resolution image and high-resolution image is obtained by studying the low-resolution-high-resolution sample library, and the results or high-resolution reconstruction are predicted , such as the neighbor embedding algorithm [4] proposed by Chang et al, the anchored neighborhood regression [5] and the adjusted anchored neighborhood regression [6]proposed by Timofte et al, the sparse coding algorithm [7] proposed by Yang et al, and the super-resolution (SRCNN) algorithm [8][9] for image using convolutional neural network proposed by Dong et al.

A new image super-resolution reconstruction algorithm based on convolutional Neural Network (CNN) is presented in this paper, which extracts rich image features by convolution, rebuilds high-resolution images by deconvolution, and adds residual structure to the network to accelerate the convergence of the whole network. The experimental results show that compared with the bicubic interpolation (Bicubic) [2] and SRCNN, the proposed algorithm is superior, as its better results of image super-resolution reconstruction, the computing speed and the convergence speed of the network are greatly improved.

II. RELATED THEORY

SRCNN, proposed by Dong et al, is an image super-resolution reconstruction algorithm based on CNN. Firstly, this algorithm uses the Bicubic method to enlarge the input low-resolution image to the same size as the high-resolution image to be output, then passes through three convolution layers, finally outputs the reconstructed high-resolution image. The SRCNN network structure of the SRCNN is shown in Figure 1.

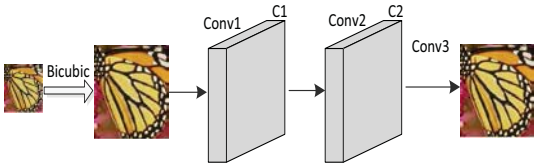

Fig. 1. SRCNN network structure

As shown in Figure 1, firstly, the enlarged image is obtained by the Bicubic method, the enlarged image is obtained by the Bicubic method, then the image features is extracted by the first layer convolution operation, after that the nonlinear mapping is realized by the second convolution layer. Finally, the image is reconstructed by the third convolution layer, and the output high-resolution image is obtained. Although SRCNN has solved the problem of using CNN for image super-resolution reconstruction, the algorithm needs to preprocess the image by using the Bicubic interpolation method, and the network training time is long, the learning ability is weak, the reconstruction performance is poor, and the operation speed is slow.

III. ALGORITHM IN THIS PAPER

In this paper, we use deconvolution to replace the traditional bicubic interpolation, and the residual structure is added to the network. The algorithm directly takes the unsampled low-resolution image as the network input, and fully extracts the original feature information of the image by using CNN. At the same time, the deconvolution layer is imported to sample the processed image, and the sampling factor is adjusted by adjusting the step size to realize the reconstruction of high-resolution image with different sizes. With the addition of residual network, the convergence speed of the whole network is accelerated, and the problem

978-1-5386-7861-9/18/$31.00 ©2018 IEEE

that the network is difficult to converge due to the deepening of the network layer is solved.

A. Network model structure

In this paper, the convolution neural network structure of image super-resolution reconstruction is constructed, which directly learning from low-resolution image to high-resolution image mapping. The network model consists of four convolution layers and one deconvolution layer, which mainly includes feature extraction, nonlinear mapping, and deconvolution reconstruction. The network framework algorithm in this paper is shown in Figure 2.

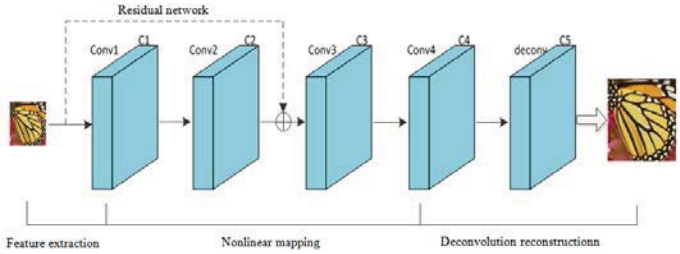

Fig. 2. Network framework of the algorithm

The first layer of convolution Conv1 acts as a feature extraction layer. The low-resolution image is directly put into the network, and the image is convoluted by 64 convolution kernels to extract the original feature information of the image, and then 64 feature images are generated to form the C1 layer. The size of convolution kernel of Conv1 layer is 5 × 5.

The nonlinear mapping layer is composed of Conv2, Conv3 and Conv4. The 64 feature images generated by Conv1 are put into the residual network module, and then they are outputted through the residual network, that is the C3 layer. Then 32 convolution kernels are used for convolution on feature images of the C3 layer to generate 32 feature images, forming the C4 layer. The size of the convolution kernels in Conv2, Conv3 and Conv4 layer are all 3×3.

The last layer of “deconv” is the deconvolution layer. The final process of low-resolution image reconstruction is realized by sampling on deconvolution. In the network model designed in this paper, the size of the deconvolution kernels are set to 9×9, and the sampling factor is adjusted by adjusting the step size to realize the reconstruction of different magnification dimensions of the low-resolution image.

B. Deconvolution

When the step size is greater than 1, the convolution is equivalent to the subsampling operation, and the deconvolution is equivalent to the upsampling operation. When the step size is equal to 1(s=1), the convolution is essentially the same as the deconvolution. The complexity of the network is directly related to the step size. The complexity of the network will be reduced S^2 when it is greater than 1. However, if the step size is too large, the quality of the image reconstruction will be seriously affected. The image can be reconstructed better and the higher resolution image can be obtained by increasing the size of the deconvolution dimension properly. The deconvolution operation process is shown in Figure 3. The procedure can be represented by Equation (1).

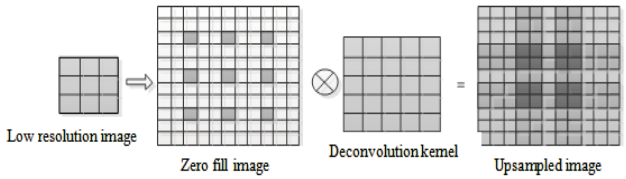

Fig. 3. Deconvolution operation process

$$F_1(Y) = \max(W_1 * Y + B_1, 0) + a_j \min(0, W_1 * Y + B_1) \quad (1)$$

As shown in Equation (1), Y is the input of low-resolution image, the parameter W_1 and B_1 are respectively the weight matrix and the offset quantity of the deconvolution kernel, and $F_1(Y)$ represent the upsampling image obtained by the deconvolution operation. The input image can be magnified by k times by setting the step size k (parameter "stride") in the deconvolution layer.

C. Residual network

Nonlinear mapping is the process of mapping feature images generated in feature extraction stage to other high dimensional feature images. The nonlinear mapping stage is the most important stage in the process of image super-resolution reconstruction, which has the most significant impact on performance. The number of network layers in the stage of deepening nonlinear mapping can achieve a better result of low-resolution image reconstruction, but with the increase of hidden layers, there will be some problems such as gradient explosion, gradient disappearance and network performance degradation [10]. Therefore, in this paper, the residual network is added to the nonlinear mapping stage to achieve a more efficient gradient descent, which speeds up the convergence process of the whole network and does not introduce any additional parameters or computational complexity. The residual network is composed by a series of residuals modules. The structure of a residual module in the residual network is shown in Fig.4. The network structure of residual can be described by Equation (2).

$$y = x + f(\theta, x) \quad (2)$$

As shown in Equation (2), x is the input of the residual network ,θ is a parameter of the residual network structure, and $f(\theta, x)$ is a residual mapping relationship learned by the network, y is the output of the residual network. The residual network in this paper contains only one residual module, and the residual module only contains the convolution layer, the activation function layer and the quick connection, and no batch unit is used.

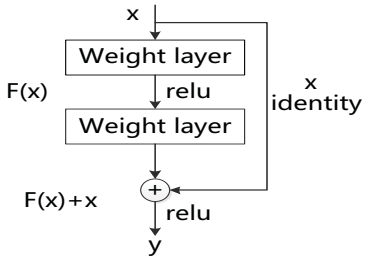

Fig. 4.Structure of a residual module

D. Activation function and loss function

The activation function used in this paper is a modified linear unit with parameters (PReLU) [11]. Compared with the ReLU [12]activation function, the PReLU activation function only adds a small amount of computation, but achieves higher accuracy and can avoid the ReLU activation function caused by the "dead features" phenomenon [13]. The PReLU activation function can play a regular effect to a certain extent.

In this paper, the loss function is represented by the mean square error(MSE) of real image and the network prediction image,and it can be described by Equation (3).

$$L(\Theta)=\frac{1}{N}\sum_{m=1}^{N}||F(Y_m;\Theta)-X_m||^2 \quad (3)$$

As shown in Equation (3), $\Theta=\{W_m,b_m\}$ is the network parameters representing of training learning, $F(Y_m;\Theta)$ and X_m are respectively the network prediction image and a set of real images, and the parameters Θ of each layer of the network are corrected by the back propagation though minimizing $L(\Theta)$, thereby the optimal mapping model can be obtained.

IV. Experiment and Result Analysis

A. Data sets and training samples

The research uses the Timofte dataset, a public dataset provided in reference [6], which contained 91 training images and two test data sets, Set5 and Set14.

In order to improve the validity of the training image data, 91 original training images are expanded by spinning and scaling. Finally 1,820 images are obtained as the training set. The images of Set5 are used as the verification set during training process.

B. Results and analysis

In order to verify the super-resolution reconstruction performance and the convergence rate of the network of the proposed method, 91 images were used in the experiment, 5 million iterations with 2 times magnification, 3 million iterations with 3 times magnification, 1 million iterations with 4 times magnification of training, respectively, and using the test set---Set5, Set14 both in the public Timofte dataset and the BSD200 test, a common test set in the field of image super-resolution reconstruction, and the test results are compared with Bicubic, RFL[14], SRCNN[15] and other methods. In this paper, root-mean-square-error (RMSE), peak signal-to-noise ratio (PSNR) and structural similarity index (SSIM) are used as quality evaluation criteria for generated high-resolution images.

Respectively, Table I and Table II shows that when the training set contains 91[6] images, the RMSE average and the PSNR average of all images in each test set after super - resolution reconstruction in the 2 times, 3 times and 4 times magnification of images.

Respectively, Table III shows that when the training set contains 91[6] images, the average value of SSIM of all images in each test set after super-resolution reconstruction in the 2 times, 3 times and 4 times magnification of images. Table IV shows that when the training set contains 91 images, the test time in each test set after of super-resolution reconstruction respectively in the 2 times, 3 times and 4 times magnification of images.

TABLE I. THE AVERAGE VALUE OF RMSE ON EACH TEST SET AFTER DIFFERENT TEST METHODS IN DIFFERENT MAGNIFICATION RATIOS

Dataset	Scale	Bicubic	RFL	FSRCNN-s	SRCNN	Algorithm in this paper
Set5	2	5.2910	3.7979	3.8022	3.8908	3.5689
Set14	2	7.8531	6.2164	6.2451	6.2956	6.0331
BSD200	2	8.3472	6.8792	6.8319	6.9110	6.5545
Set5	3	7.7097	6.0959	6.0123	6.0819	5.8350
Set14	3	10.7039	8.9958	8.9648	9.0373	8.9236
BSD200	3	11.0545	9.8637	9.7845	9.8410	9.6837
Set5	4	9.6725	7.9349	8.0268	7.9257	7.6303
Set14	4	12.7803	11.0800	11.2341	11.1183	10.9911
BSD200	4	12.8245	11.7095	11.7501	11.7636	11.5755

TABLE II. THE AVERAGE VALUE OF PSNR ON EACH TEST SET AFTER DIFFERENT TEST METHODS IN DIFFERENT MAGNIFICATION RATIOS

Dataset	Scale	Bicubic	RFL	FSRCNN-s	SRCNN	Algorithm in this paper
Set5	2	33.66	36.54	36.53	36.33	37.08
Set14	2	30.23	32.26	32.22	32.15	32.52
BSD200	2	29.70	31.38	31.44	31.34	31.80
Set5	3	30.39	32.43	32.55	32.45	32.81
Set14	3	27.54	29.05	29.08	29.01	29.12
BSD200	3	27.26	28.25	28.32	28.27	28.41
Set5	4	28.42	30.14	30.04	30.15	30.48
Set14	4	26.00	27.24	27.12	27.21	27.31
BSD200	4	25.97	26.76	26.73	26.72	26.86

TABLE III. THE AVERAGE VALUE OF SSIM ON EACH TEST SET AFTER DIFFERENT TEST METHODS IN DIFFERENT MAGNIFICATION RATIOS

Dataset	Scale	Bicubic	RFL	FSRCNN-s	SRCNN	Algorithm in this paper
Set5	2	0.9299	0.9530	0.9530	0.9521	0.9561
Set14	2	0.8687	0.9051	0.9050	0.9038	0.9079
BSD200	2	0.8625	0.9024	0.9026	0.9023	0.9047
Set5	3	0.8675	0.9049	0.9053	0.9035	0.9066
Set14	3	0.7736	0.8164	0.8165	0.8145	0.8184
BSD200	3	0.7638	0.8055	0.8058	0.8038	0.8064
Set5	4	0.8104	0.8499	0.8496	0.8532	0.8569
Set14	4	0.7019	0.7425	0.7421	0.7413	0.7464
BSD200	4	0.6949	0.7312	0.7311	0.7291	0.7349

TABLE IV. TEST TIME RECONSTRUCTED ON EACH TEST SET AFTER DIFFERENT TEST METHODS IN DIFFERENT MAGNIFICATION RATIO

Dataset	Scale	Bicubic	SRF	SRCNN	SRCNN-Ex	Algorithm in this paper
Set5	2	-	2.1	0.18	1.3	0.047
Set14	2	-	3.9	0.39	2.8	0.051
BSD200	2	-	3.1	0.23	1.7	0.048
Set5	3	-	1.7	0.18	1.3	0.063
Set14	3	-	2.5	0.39	2.8	0.066
BSD200	3	-	2.0	0.23	1.7	0.045
Set5	4	-	1.5	0.18	1.3	0.044
Set14	4	-	2.1	0.39	2.8	0.046
BSD200	4	-	1.7	0.23	1.7	0.044

It can be seen from Table I and Table II the value of RMSE proposed in this paper is significantly lower than that of other methods, while the value of PSNR is higher than that of other methods. It can be seen from Table III that the value of SSIM reconstructed by the algorithm in this paper is higher than that of other methods. It can be seen from Table IV that the test time required by the algorithm in this paper is significantly lower than other methods. The experimental results show that the proposed algorithm has achieved a higher reconstruction effect and the operation speed is greatly improved.

Respectively, Figure 6 and Figure 10 shows the reconstruction effects of Bicubic, SRCNN and the proposed algorithm "butterfly_GT" and "lenna" at the magnification ratio of 2 times and 4 times .

ORIGINAL / PSNR BICUBIC / 27.43 SRCNN / 32.20 ALGORITHM / 33.53

Fig. 5.Comparison of reconstruction effects of “butterfly_GT" at 2 times magnification

ORIGINAL / PSNR BICUBIC / 29.83 SRCNN / 31.20 ALGORITHM / 31.52

Figures 6.Comparison of reconstruction effects of "lenna" at 4 times magnification

V. SUMMARY

For the existing SRCNN algorithm network’s training time is too long, the reconstruction effect is inferior, and the computation speed is slow, a new image super-resolution reconstruction algorithm based on convolution neural network is proposed in this paper. The feature of the image is extracted in the low-resolution space, the computational complexity of the network is reduced, the final image reconstruction is realized by the deconvolution layer, and the residual structure is added in the network to speed up the network convergence. The experimental results show that, compared with Bicubic, SRCNN and other methods, the algorithm in this paper has achieved superior image reconstruction performance, greatly improved the computing speed, and achieved better results with fewer training iterations, thus improving the efficiency of the algorithm. The next step will be to explore other more effective network models to achieve better super-resolution reconstruction effects and meeting video real-time processing requirements.

ACKNOWLEDGMENT

This work was supported by Hainan Provincial Natural Science Foundation of China (Project Nos.618MS027) and Hainan University Education and Teaching Reform Research Project (Project Nos.hdjy1730). Zhang Jian is the corresponding author.

REFERENCES

[1] Heng S U. Survey of Super-resolution Image Reconstruction Methods[J]. Acta Automatica Sinica, 2013, 39(8):1202-1213.

[2] Aràndiga F. A nonlinear algorithm for monotone piecewise bicubic interpolation ☆[J]. Applied Mathematics & Computation, 2016, 272(P1):100-113.

[3] Li X, Orchard M T. New edge-directed interpolation.[J]. IEEE Trans Image Process, 2001, 10(10):1521-1527.

[4] Chang H, Yeung D Y, Xiong Y. Super-resolution through neighbor embedding[C]// IEEE Computer Society Conference on Computer Vision & Pattern Recognition. IEEE Computer Society, 2004:275-282.

[5] Timofte R, De V, Gool L V. Anchored Neighborhood Regression for Fast Example-Based Super-Resolution[C]// IEEE International Conference on Computer Vision. IEEE, 2014:1920-1927.

[6] Timofte R, Smet V D, Gool L V. A+: Adjusted Anchored Neighborhood Regression for Fast Super-Resolution[C]// Asian Conference on Computer Vision. Springer, Cham, 2014:111-126.

[7] Yang J, Wang Z, Lin Z, et al. Coupled Dictionary Training for Image Super-Resolution [J]. IEEE Transactions on Image Processing A Publication of the IEEE Signal Processing Society, 2012, 21(8):3467-3478.

[8] Dong C, Chen C L, He K, et al. Image Super-Resolution Using Deep Convolutional Networks [J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2016, 38(2):295-307.

[9] Dong C, Chen C L, Tang X. Accelerating the Super-Resolution Convolutional Neural Network[C]// European Conference on Computer Vision. Springer, Cham, 2016:391-407.

[10] He K, Zhang X, Ren S, et al. Deep Residual Learning for Image Recognition[C]// IEEE Conference on Computer Vision and Pattern Recognition. IEEE Computer Society, 2016:770-778.

[11] Xu C, Liu T, Tao D, et al. Local Rademacher Complexity for Multi-Label Learning [J]. IEEE Transactions on Image Processing, 2016, 25(3):1495-1507.

[12] Gong C, Liu T, Tao D, et al. Deformed Graph Laplacian for Semisupervised Learning.[J]. IEEE Transactions on Neural Networks & Learning Systems, 2015, 26(10):2261-2274.

[13] Fakhry A, Tao Z, Ji S. Residual Deconvolutional Networks for Brain Electron Microscopy Image Segmentation [J]. IEEE Transactions on Medical Imaging, 2017, 36(2):447-456.

[14] Schulter S, Leistner C, Bischof H. Fast and accurate image upscaling with super-resolution forests[C]// Computer Vision and Pattern Recognition. IEEE, 2015:3791-3799.

[15] Dong C, Chen C L, He K, et al. Learning a Deep Convolutional Network for Image Super-Resolution[C]// European Conference on Computer Vision. Springer, Cham, 2014:184-199.