A DEEP CONVOLUTIONAL NETWORK FOR MEDICAL IMAGE SUPER-RESOLUTION

Yunxing Gao
School of Information Science and Engineering, University of Jinan
Shandong Provincial Key Laboratory of Network Based Intelligent Computing
Jinan 250022, China

Hengjian Li
School of Information Science and Engineering, University of Jinan
Shandong Provincial Key Laboratory of Network Based Intelligent Computing
Jinan 250022, China

Jiwen Dong*
School of Information Science and Engineering, University of Jinan
Shandong Provincial Key Laboratory of Network Based Intelligent Computing
Jinan 250022, China
ise_dongjw@ujn.edu.cn

Guang Feng
School of Information Science and Engineering, University of Jinan
Shandong Provincial Key Laboratory of Network Based Intelligent Computing
Jinan 250022, China

***Abstract*—This paper presents a novel deep network model specifically for medical image super-resolution reconstruction. We take full advantage of the fact that all medical images basically have distinct repetitive structure and large black border without any texture information. Compared with the Super-Resolution Convolution Neural Network (SRCNN) structure, firstly, we add a convolution layer to carry out secondary feature extraction which make the feature more representative. Then we add overlapping pooling layers in order to highlight the important features and fine processing them. At last, in order to use local features and global features to complete the reconstruction together, we add a link layer, establish a connection between the second pooling layer and the reconstruction layer. The experimental results show that average PSNR gains achieved by our algorithm are higher than the original SRCNN. The reconstructed CT images can clearly provide important reference for clinicians to make the correct treatment decisions, and also have important guiding significance for the difficulty and risk assessment of surgical feasibility.**

Keywords—convolutional neural network; medical image; super-resolution; deep learning

I. Introduction

Medical image as an important basis for medical diagnosis require high clarity, but due to hardware devices and existing imaging technology limitations, doctors often do not get the ideal high-definition images. The application of image super-resolution technology in medical images enables doctors to obtain high-quality, high-resolution medical images like CT or MRI images. High-definition medical images can provide more accurate clues for medical diagnosis .With these high-definition medical images, doctors can more accurately identify the lesion location. So to improve the resolution of medical images has great practical significance.

Image super-resolution [1-2] technology is to obtain a single high-resolution image from one or more low-resolution images. In recent years, researchers have made many achievements in the study of super-resolution algorithms for natural images. Some successful algorithms are proposed by them. Such as Freeman et al. [3] proposed an algorithm based on case learning in 2002. This algorithm uses confidence propagation to solve the Markov random field to learn the mapping between low resolution and high resolution. And in 2004, Chang et al. [4] proposed a super-resolution reconstruction algorithm based on neighborhood embedding using local linear embedding. This algorithm maps the spatial local geometry of low-resolution image blocks into high-resolution spaces and reconstructs the image with a linear combination of neighborhoods. In 2010, Yang et al. [5] proposed a new algorithm for super-resolution reconstruction based on sparse expression. Dong et al. [6] achieved super-resolution reconstruction by convolution neural network (SRCNN).

Super-resolution reconstruction algorithms for medical image are not as popular as natural images but researchers also have made good progress. Between 2001 and 2009 [7-9], the super-resolution algorithm began to be applied to medical imaging equipment. D. Wallach et al. [10] proposed a super-resolution algorithm for four-dimensional positron emission tomography (4D-PET) in 2008. They use multiple lung slice images and the Maximum a-posteriori (MAP) algorithm to estimate the high-resolution image, and then they use the gradient descent algorithm to find the optimal solution. In 2013, Wang Y H et al. [11] used the double-dictionary algorithm based on sparse representation to reconstruct MRI images. First, they train the corresponding high and low resolution dictionaries, and then use the calculated input image sparse representation to reconstruct high-resolution images. Finally, they try to use the global optimization to enhance the recovery effect. In 2010, P. Akhtar and F. Azhar [12] combined

*Corresponding author. Email: ise_dongjw@ujn.edu.cn

with cubic interpolation algorithm and 2-D interpolation algorithm, proposed a single image super-resolution algorithm based on interpolation algorithm for biomedical imaging. Dong W et al. [13] proposed a super-resolution reconstruction algorithm for medical images based on adaptive sparse domain selection and adaptive regularization in 2011. Those above algorithms have made some progress already but their performance are not ideal. This is mainly due to the fact that these algorithms are not considered the distinct repetitive structure of medical images and large black border without any texture information when they reconstruct high-resolution medical images.

In our study, we take full advantage of the fact that all medical images basically have distinct repetitive structure to design a new network structure, and successfully applied the convolution neural network to the medical images super-resolution reconstruction. We are inspired by the Dong convolution neural network image super - resolution algorithm and use the neural network model on Imagenet contest for reference. After through our algorithm the image clarity has improved, the details of the texture is more abundant. In short, the contributions of this paper include:

- We add a convolution layer to carry out secondary feature extraction.
- We add overlapping pooling layers in order to highlight the important features and fine processing them. And reduce the dimension of feature maps, reduce the computational complexity of the algorithm.
- We add a link layer, establish a connection between the second convolution layer and the reconstruction layer, in order to use local features and global features to complete the reconstruction together.

In the following, we will first review the detail flow of the SRCNN algorithm in Sec.2. And then we introduced our medical image super-resolution model in detail in Sec.3. Extensive experimental results are reported in Sec.4, and conclusions are drawn in Sec.5.

II. SUPER-RESOLUTION CONVOLUTION NEURAL NETWORK

Learning-based image super-resolution method [14-15] is a relatively popular research field in recent years. This type of method uses the image database or the image itself to learn the association between high and low resolution images. And use it as a priori constraint to generate high-resolution images. The image super-resolution based on deep learning is one of the branches, mainly divided into the following steps: (1) Create an external image dataset for training. (2) Build a network model. (3) Train the dataset, estimating and optimizing the network parameters to obtain the feature expression and the prior knowledge of the input data. (4) Input the low-resolution image to the trained network model, and get the output of the high-resolution image.

Influenced by the great success of deep learning method is applied to the field of computer vision, Dong et al. [6]apply the depth learning method to image super-resolution reconstruction and propose a single image super-resolution reconstruction algorithm based on convolution neural network named SRCNN. This algorithm avoids the artificially designed feature extraction method, achieves the end-to-end learning and improves the image reconstruction precision. The SRCNN consists of three convolutional layers. They are patch extraction and representation layer, non-linear mapping layer, reconstruction layer. An overview of the network is depicted in Fig. 1.

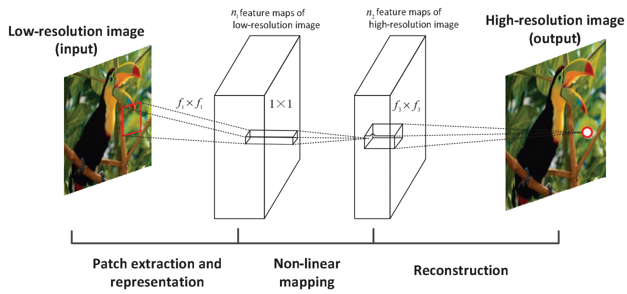

Fig.1 Given a low-resolution image Y, the first convolutional layer of the SRCNN extracts a set of feature maps. The second layer maps these feature maps nonlinearly to high-resolution patch representations. The last layer combines the predictions within a spatial neighbourhood to produce the final high-resolution image F(Y).

The first is the **patch extraction and representation layer. In** this layer the algorithm extracts (overlap-ping) patches from the low-resolution image **Y** and represents each patch as a high-dimensional vector. These vectors comprise a set of feature maps, of which the number equals to the dimensionality of the vectors. The output of this layer is all the features of the input image which is expressed as $F_1(Y)$. Formally, the operation of first layer is expressed as (1):

$$F_1(Y) = \max(0, W_1 * Y + B_1) \qquad (1)$$

Here Y is the input image, $*$ stands for convolution operation, W_1 and B_1 represent the filters and biases respectively. Here W_1 is of a size $c \times f_1 \times f_1 \times n_1$, where c is the number of channels in the input image, f_1 is the spatial size of a filter, and n_1 is the number of filters. The output is composed of n_1 feature maps. B_1 is an n_1 -dimensional vector, whose each element is associated with a filter. The feature maps obtained by convolution is processed by the activation function ReLU.

The first layer is followed by a non-linear mapping layer. In this layer the algorithm maps each of these n_1 -dimensional vectors into an n_2 -dimensional one. Each mapped vector represents a high-resolution pixel block, and these vectors form another feature map set $F_2(Y)$. The operation of the second layer is shown in (2):

$$F_2(Y) = \max(0, W_2 * F_1(Y) + B_2) \qquad (2)$$

Here W_2 is of a size $n_1 \times f_2 \times f_2 \times n_2$. If the second convolution layer contains n_2 convolution kernels, after the convolution operation the algorithm will generate n_2 -dimensional feature map. The output of each vector of n_2 -

dimension represents a high-resolution pixel block which will be applied to the reconstruction process.

The last layer is reconstruction, in this layer the algorithm aggregates the above high-resolution patch-wise representations to generate the final high-resolution image. This image $F(Y)$ is expected to be similar to the ground truth X. The operation of the last layer is shown in (3):

$$F(Y) = W_3 * F_2(Y) + B_3 \qquad (3)$$

Here W_3 is of size $n_2 \times f_3 \times f_3 \times c$, and B_3 is a c-dimensional vector.

The parameters of the SRCNN can be expressed as $\Theta = \{W_1, W_2, W_3, B_1, B_2, B_3\}$. The process of training the entire network is to estimate and optimize the parameters. This is achieved through minimizing the loss between the reconstructed images $F(Y;\Theta)$ and the corresponding ground truth high-resolution images X. Given a set of high-resolution image $\{X_i\}$ and their corresponding low-resolution images $\{Y_i\}$, the algorithm use Mean Squared Error (MSE) as the loss function:

$$L(\Theta) = \frac{1}{n}\sum_{i=1}^{n} \| F(Y_i;\Theta) - X_i \|^2 \qquad (4)$$

Here n is the number of training samples. The loss is minimized using stochastic gradient descent with the standard backpropagation.

SRCNN is a universal model, but its performance in medical images is not ideal. We specially design an algorithm for medical image super-resolution. Our algorithm takes full advantage of the fact that all medical images basically have distinct repetitive structure and large black border without any texture information. The specific improvements and algorithm flows and detailed network configuration parameters is shown in Sec.3.

III. THE IMPROVED SRCNN

VGGNet [16] is a deep convolution neural network developed by Oxford University Computer Vision Group and researchers in Google DeepMind. VGGNet explores the relationship between the depth of the convolution neural network and its performance. VGGNet has significantly reduced the error rate compared to the state-of-the-art network structure. This move fully illustrates the performance of deep network is better than shallow network. And in the end of each convolution VGGNet has a pooling which is used to reduce the size of images. Inspired by those idea our network is one more convolutional and two pooling layers than SRCNN. These moves can make our network reach the depth of 6 layers and reduce the size of feature maps. The overview of our network is shown in the Fig.2.

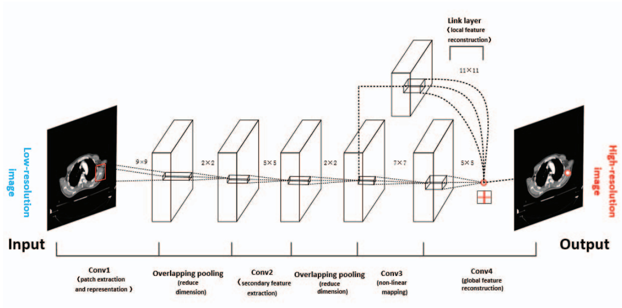

Fig.2 Our network structure.

A. *Convolutional layer*

We add a new convolution layer after the original first patch extraction and representation layer. This operation perform secondary feature extraction from the output of original first layer. This secondary feature extraction works directly on the feature map rather than the original image. The purpose of this move is to make more representative of the extracted features and deepen the depth of the network at the same time. The operation of the new convolution layer is shown in (5):

$$F_2{}'(Y) = \max(0, W_2{}' * F_1(Y) + B_2{}') \qquad (5)$$

Here $W_2{}'$ is of a size 32*5*5*32, $B_2{}'$ is 32-dimensional.

Visualization of learned feature maps after the feature extraction layer obtained using two layer convolution and single layer convolution are shown in the Fig.3.

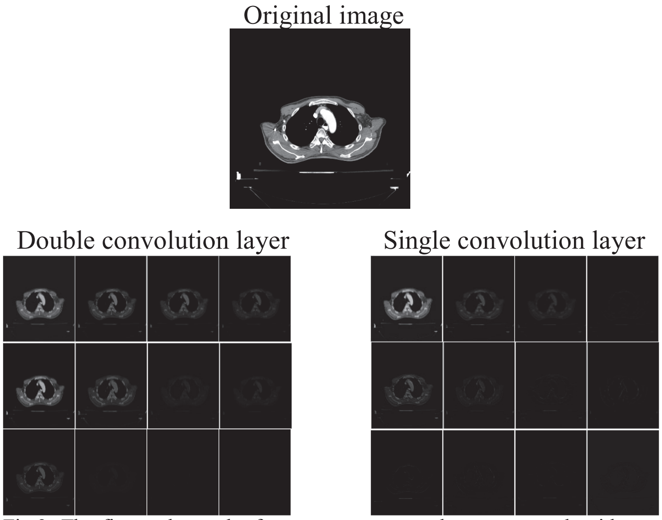

Fig.3. The figure shows the feature map contrast between our algorithm and the SRCNN.

B. *Overlapping pooling*

SRCNN doesn't have pooling layer. We add two pooling layers after the original first convolution layer and the new convolution layer. We need to be very careful and fine when dealing with medical images because of the special use of medical images. In this paper, we use the overlapping pooling [17]. It can effectively avoid missing any useful features.

Pooling layers in CNNs summarize the outputs of neighboring groups of neurons in the same kernel map. Traditionally, the neighborhoods summarized by adjacent pooling units do not overlap. To be more precise, a pooling layer can be thought of as consisting of a grid of pooling units spaced s pixels apart, each summarizing a neighborhood of size $z \times z$ centered at the location of the pooling unit. If we set $s = z$, we obtain traditional local pooling as commonly employed in CNNs. If we set $s < z$, we obtain overlapping pooling. This is what we use throughout our network, with $s = 1$ and $z = 2$.

After adding the pooling layer and the new convolution layer, our network has reached the depth of 6 layers, deep network structure is more conducive to learning more accurate expression of image data. The pooling layer can also effectively reduce the number of parameters, help to simplify the network, improve the efficiency of parameter training. We generally observe during training that models with overlapping pooling find it slightly more difficult to overfit. For the pooling layer, the number of input feature maps does not change, but the size of each output feature map is reduced. The essence of this process is down sampling, the operation of overlapping layers are shown in (6):

$$x_j^l = f(\beta_j^l down(x_j^{l-1}) + b_j^l) \qquad (6)$$

Here x_j^l and x_j^{l-1} represent the j-th feature map of the current layer and the previous layer respectively. down(•) represent the down sampling. f(•) represent the activate function. β_j^l is multiplicative paranoid and b_j^l is additive biases vector.

C. *Link Layer*

In this paper, we add a new connection [18] (we call it link layer) between the secondary pooling layer and the last reconstruction layer because all medical images basically have distinct repetitive structure. After this operation, we can use the local features obtained by the twice convolutions and twice pooling operation to reconstruct an image in the link layer. And then we use the global feature obtained by Conv3 (non-linear mapping layer) to reconstruct another image in the Conv4 (reconstruction layer). In the end we add the results of the two reconstruction images to get the final super-resolution results. The combination of global and local features can effectively let us get a clearer high-resolution image. The special use of medical images makes us dare not lose a little bit of information in them so we use local feature to compensate for the loss of information in the training process.

According to experience and constant experiment, we adjust the learning rate of the algorithm to 10-4. So that our algorithm can quickly converge. Detailed network parameters are shown in TABLE Ⅰ. The algorithm flow is shown in Fig.4.

TABLE I. DETAILED NETWORK PARAMETERS

Network Configuration	
Conv1	9*9*32 and Stride of 1 Followed by ReLU
Overlapping pooling	2*2 stride of 1
Conv2	32*5*5*32 and Stride of 1 Followed by ReLU
Overlapping pooling	2*2 stride of 1
Link layer	32*11*11*1 Stride of 1
Conv3	32*7*7*32 and Stride of 1 Followed by ReLU
Conv4	32*5*5*1 Stride of 1

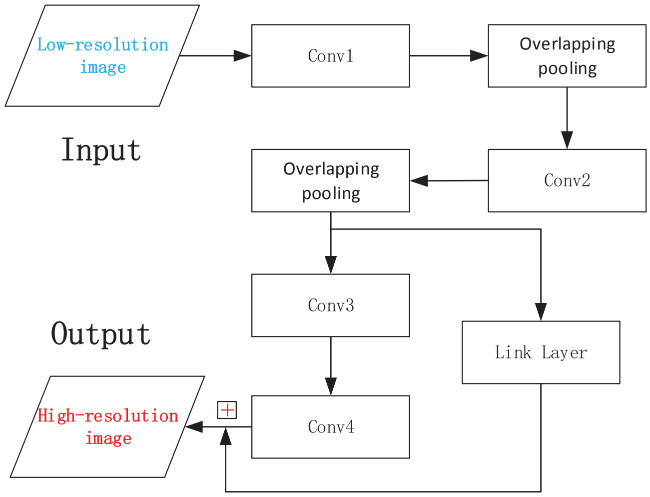

Fig.4.The algorithm flow.

IV. EXPERIMENT

In this paper, we use the CT images provided by Shandong Provincial Tumor Hospital as the experimental dataset. We have 376 CT images of the esophagus, pelvic and nasal. We randomly selected 155 as training set, 35 as test sets and 20 as validation set.

In the choice of contrast algorithm, we first chose the super-resolution algorithm with universality such as: Bicubic, SC (sparse coding) [5], SCN (sparse coding based network) [19], SRCNN [6]. Then we chose several super-resolution algorithms specifically design for medical images such as: SPSR (statistical prediction model) [20], ASDS (adaptive sparse domain selection) [13].

We choose two CT images of the pelvis, two CT images of the esophagus and two CT images of the nasal in our contract experiments. As shown in TABLE Ⅱ, TABLEⅢand TABLE Ⅳ, the proposed algorithm yields the highest average PSNR in all scale 2, scale 3 and scale 4 experiments. Specifically, as shown in those Tables, the average gains achieved by improved algorithm are 2.1 dB, 0.6dB, and 1dB, higher than the original SRCNN. And compare with algorithm SPSR which is specifically for medical images, our algorithm are 0.59dB and 1.64dB higher than SPSR.

TABLE II. THE RESULT OF PSNR (dB) WITH AN UPSCALING FACTOR 2 ON VERIFICATION SET.

Images	Scale	Bicubic	SC [5]	SCN [19]	SRCNN [6]	SPSR [20]	ASDS [13]	Our
		PSNR	PSNR	PSNR	PSNR	PSNR	PSNR	PSNR
pelvic1	2	42.8	47.2	46.8	46.4	48.0	45.2	48.4
pelvic2	2	41.4	46.2	45.3	45.4	46.2	44.2	46.9
esophagus3	2	41.0	45.7	44.5	44.4	46.9	43.5	47.0
esophagus4	2	39.4	44.6	43.4	43.1	44.7	42.1	44.5
nasal5	2	42.7	47.6	46.9	46.6	47.3	45.7	47.5
nasal6	2	41.4	46.0	45.2	44.8	46.7	44.9	46.9
Average	2	41.5	46.2	45.3	45.1	46.6	44.3	47.2

TABLE III. THE RESULT OF PSNR (dB) WITH AN UPSCALING FACTOR 3 ON VERIFICATION SET.

Images	Scale	Bicubic	SC [5]	SCN [19]	SRCNN [6]	SPSR [20]	ASDS [13]	Our
		PSNR	PSNR	PSNR	PSNR	PSNR	PSNR	PSNR
pelvic1	3	37.9	40.0	41.5	42.6	41.5	42.8	43.3
pelvic2	3	36.5	38.8	40.3	41.3	40.2	41.2	41.7
esophagus3	3	35.1	37.0	38.5	39.5	38.8	39.1	40.4
esophagus4	3	34.2	36.5	37.5	38.5	37.5	38.1	38.9
nasal5	3	35.1	39.5	41.1	42.3	41.0	41.5	42.8
nasal6	3	36.3	38.7	40.3	41.4	40.3	41.3	42.1
Average	3	35.8	38.4	39.9	40.9	39.9	40.7	41.5

TABLE IV. THE RESULT OF PSNR (dB) WITH AN UPSCALING FACTOR 4 ON VERIFICATION SET.

Images	Scale	Bicubic	SC [5]	SCN [19]	SRCNN [6]	SPSR [20]	ASDS [13]	Our
		PSNR	PSNR	PSNR	PSNR	PSNR	PSNR	PSNR
pelvic1	4	34.8	36.0	38.1	38.8	-	38.0	39.4
pelvic2	4	33.7	34.9	36.9	37.2	-	36.6	37.7
esophagus3	4	32.2	33.4	35.3	35.8	-	34.4	36.7
esophagus4	4	31.2	32.6	34.3	34.6	-	34.0	34.9
nasal5	4	35.1	36.1	37.6	38.1	-	37.5	39.1
nasal6	4	33.3	34.9	36.6	36.8	-	36.5	38.0
Average	4	33.4	34.7	36.5	36.9	-	36.2	37.9

Part of the reason for the improvement is the twice feature extraction. With the secondary feature extraction layer we can get more representative features in all the acquired features. With two over pooling layers we can effectively avoid missing any useful features. After these moves our network has reached the depth of 6 layers, deep network structure is more conducive to learning more accurate expression of image data. Another part of the reason is the new connection. With the link layer we can combine the global features and local features to reconstruct the final high-resolution image. The global features and local features complement each other so that we can get a clearer high-resolution image than SRCNN.

The Fig.5, Fig.6 and Fig.7 corresponding to the pelvic, esophagus, nasal CT images. These images show the detail of every algorithm and our algorithm. We can see obviously the subjective visual effect is improved.

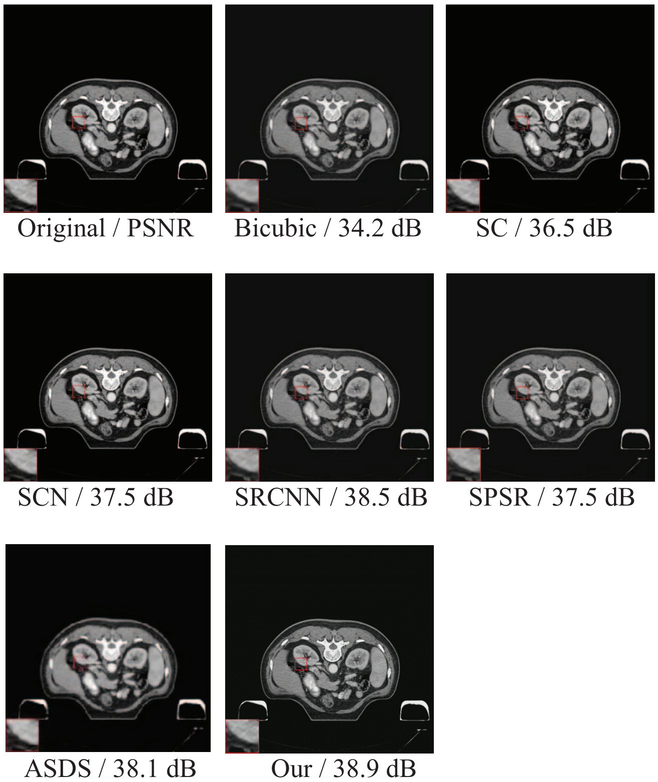

Fig.5. "Pelvic" CT image from validation set with an upscaling factor 3.

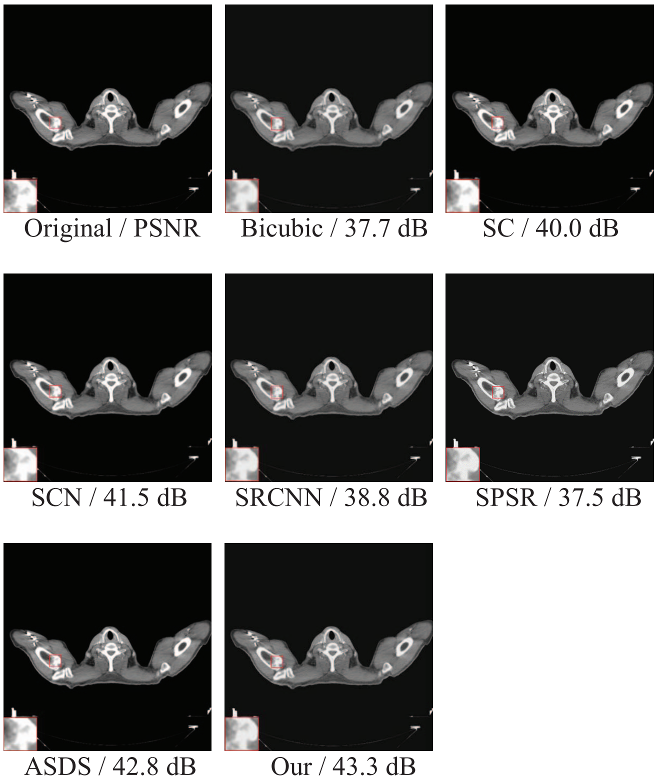

Fig.6. "Esophagus" CT image from validation set with an upscaling factor 3.

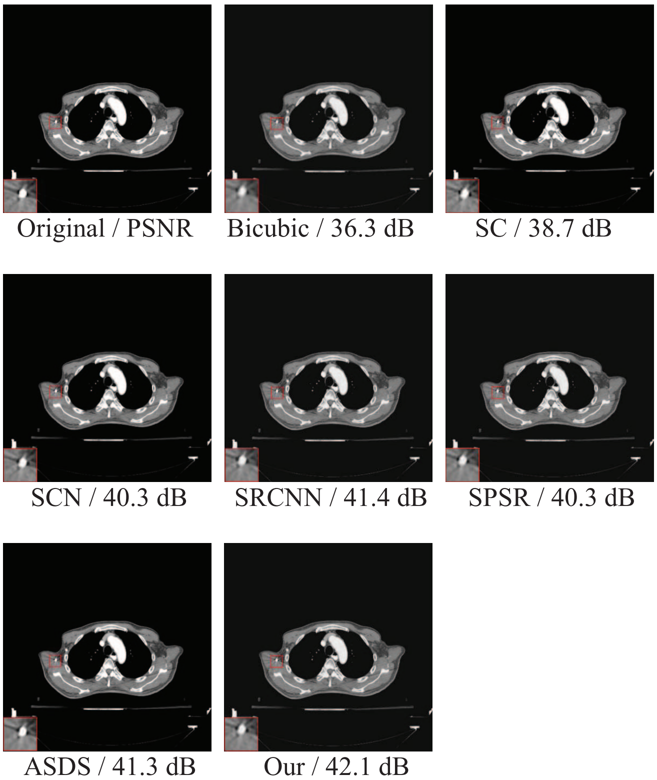

Fig.7. “Nasal” CT image from validation set with an upscaling factor 3.

V. CONCLUSION

In this paper, we presented a novel deep network model specifically for medical image super-resolution reconstruction. We fully consider the characteristics of medical image structure repetition and black border. Based on SRCNN model, we add a convolution layer to carry out secondary feature extraction to improve the feature performance, and overlapping pooling layers is adopted to highlight the important features. Furthermore, we establish a link layer between the second convolution layer and the reconstruction layer, which make local features and global features to complete the reconstruction together. The experimental results show that average PSNR gains achieved by our algorithm are 2.1 dB, 0.6dB, and 1dB, higher than the original SRCNN. And the super-resolution results in medical images achieved by our model not only better than SRCNN but also better than other contrast algorithms. The image reconstructed by our algorithm obtain a higher objective evaluation index PSNR and a better subjective visual effect than before.

REFERENCES

[1] Tsai, "Multiframe image restoration and registration," Adv. Comput. Vis. Image Process, vol. 1.2, pp.317-339, 1984.

[2] Goodman J W. “Introduction to Fourier optics,” Roberts and Company Publishers, 2005.

[3] Freeman, William T., Thouis R. Jones, and Egon C. Pasztor, "Example-based super-resolution," IEEE Computer graphics and Applications, vol.22.2, pp. 56-65, 2002.

[4] Chang, Hong, Dit-Yan Yeung, and Yimin Xiong, "Super-resolution through neighbor embedding," Computer Vision and Pattern Recognition, 2004. [Proceedings of the 2004 IEEE Computer Society Conference on. Vol. 1. IEEE, 2004].

[5] Yang, J., Wright, J., Huang, T. S., and Ma, Y., “Image super-resolution via sparse representation,” IEEE transactions on image processing, vol.19.11, pp. 2861-2873, 2010.

[6] Dong, C., Loy, C. C., He, K., and Tang, X., “ Image super-resolution using deep convolutional networks,” IEEE transactions on pattern analysis and machine intelligence, vol.38.2, pp.295-307, 2016.

[7] Peeters, Ronald R., et al., "The use of super‐resolution techniques to reduce slice thickness in functional MRI," International Journal of Imaging Systems and Technology, vol.14.3, pp.131-138, 2004.

[8] Kennedy, J. A., Israel, O., Frenkel, A., Bar-Shalom, R., and Azhari H., “Super-resolution in PET imaging,” IEEE transactions on medical imaging, vol.25.2, pp.137-147, 2006.

[9] Greenspan H.,"Super-resolution in medical imaging," The Computer Journal, vol.52.1, pp.43-63, 2008.

[10] Wallach, D., et al., "Super-resolution in 4D positron emission tomography," Nuclear Science Symposium Conference Record, NSS'08. IEEE.IEEE, pp. 4285-4287, October 2008.

[11] Wang, X., D. Chen, and Q. Ran, "Image super-resolution reconstruction with content based dual-dictionary learning and sparse representation," Chinese Journal of Scientific Instrument, vol.34.8, pp.1690-1695, 2013.

[12] Akhtar, Pervez, and Faisal Azhar, "A single image interpolation scheme for enhanced super resolution in bio-medical imaging," Bioinformatics and Biomedical Engineering [4th International Conference on. IEEE pp.1-5], 2010.

[13] Dong, W., Zhang, L., Shi, G., and Wu X., “Image deblurring and super-resolution by adaptive sparse domain selection and adaptive regularization,” IEEE Transactions on Image Processing, vol.20.7, pp.1838-1857, 2011.

[14] Kim, Jiwon, Jung Kwon Lee, and Kyoung Mu Lee, "Accurate image super-resolution using very deep convolutional networks," Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp.1646-1654, 2016.

[15] Kim, Jiwon, Jung Kwon Lee, and Kyoung Mu Lee, "Deeply-recursive convolutional network for image super-resolution," Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1637-1645, 2016.

[16] Simonyan Karen, and Andrew Zisserman, "Very deep convolutional networks for large-scale image recognition." arXiv preprint arXiv.1409.1556, 2014.

[17] Krizhevsky Alex, Ilya Sutskever, and Geoffrey E. Hinton, "Imagenet classification with deep convolutional neural networks," Advances in neural information processing systems, pp.1097-1105, 2012.

[18] Sun, Yi, Xiaogang Wang, and Xiaoou Tang, "Deep learning face representation from predicting 10,000 classes," Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp.1891-1898, 2014.

[19] Wang, Z., Liu, D., Yang, J., Han, W., and Huang, T., “Deep networks for image super-resolution with sparse prior,” In Proceedings of the IEEE International Conference on Computer Vision, pp. 370-378, 2015.

[20] Peleg, Tomer, and Michael Elad, "A statistical prediction model based on sparse representations for single image super-resolution," IEEE transactions on image processing, vol.23.6, pp.2569-2582, 2014.