Super-Resolution of Medical Image Using Representation Learning

Xiong Yang†, Shu Zhan†*, Changsheng Hu†, Zhicheng Liang†, Dongdong Xie‡
†School of Computer and Information, Hefei University of Technology, Hefei, China 230009
‡Second Affiliated Hospital of Urology, Anhui Medical University, Hefei, China 230601
Email: *shu_zhan@hfut.edu.cn

***Abstract*—Super-resolution (SR) of single image is a meaningful challenge in medical images based diagnosis, while the image resolution is limited. Also, numerous deep neural networks based models were proposed and achieve excellent performance which is superior to the previous handcrafted methods. In this paper, we employ a deep convolutional neural networks for the super-resolution (SR) of single medical image, which learns the nonlinear mapping from the low-resolution space to high-resolution space directly. In addition, we use three sets imaging data (Mammary gland, Prostate tissue and Human brain) training deep network respectively. Firstly, we use Randomized Rectified Linear Unit (RReLU), which incorporates a nonzero slope for negative part to solve the problem of over compression. Secondly, for the purpose of enhancing the quality of reconstructed result and reducing the noise of over-fitting, Nesterov′s Accelerated Gradient (NAG) method on the SRCNN is used to accelerate the convergence of loss function and avoid the large oscillations. A comparative performance evaluation is carried out over a set of experiments using real imaging data to verify the validity of proposed algorithm.**

keywords: Image Super-Resolution; Deep Learning; Feature Map; Representation Learning; Convolutional Neural Networks

I. Introduction

With the application of computer vision technology, medical image has become an essential role in clinical application. In the scope of diagnosis using medical image, the pathological estimation of clinical expert and assistant computer will be influenced on account of the low quality image [1]–[3]. Nowadays, 3T MRI has been widely used in clinical treatment and 7T MRI is begin to be equipped in laboratory. Although the increased magnetic field strength can help increasing the signal-to-noise ratio(SNR), the image resolution is still limited. As the resolution of medical image is restricted by various economical, technological and physical condition, it is easily to cause misleading of diagnosis and misclassification of tissue [4], [5]. For example, calcification is the common symptoms of most breast cancer. As most calcification are too small to recognize and the intensity variation between the pathological tissues and healthy region is low, the diagnosis procedure become troublesome. Consequently, the task that improve the spatial resolution of medical image is become important.

In order to obtain the high resolution medical image, numerous image enhancement algorithm has been proposed. Image super-resolution(SR) [6], which is a embranchment of image enhancement, is an active and classic image processing task which accomplish the task by creating a visually pleasing and plausible high-resolution(HR) output image from the low-resolution(LR) input image. One of the earliest method about this problem was proposed by Tsai in 1980 [7]. Subsequently, many SR approaches had been presented in the past three decades [8], [9]. Although there were many possible solutions to generated an HR image from the LR image, We must use regularization of some forms, the prior information about HR image, to ensure the stability and uniqueness of the extension. For this purpose [10], the existing super resolution reconstruction approaches could be generally classified into three classed: interpolation based approaches, reconstruction (regularization) based approaches and learning (example) based approaches.

Previously, much attention had been paid on interpolation based approaches e.g. nearest-neighbor interpolation and linear interpolation. The interpolation-based methods [11], which were the primary image super-resolution approaches, would generally produce more favorable results by exploiting the priors of natural image. However, the simple interpolation approaches like Bicubic or Bilinear interpolation tended to produce over-smoothing images with jagged and choppy effects. Although these method were easy and fast to comprehend and operate, they were poor at processing complex and edges textures. Blurred textures and edges could be seen on interpolation based approaches frequently.

The reconstruction-based methods [12] performed a restriction that the down-sampled and smoothed edition of the evaluated HR image need be corresponding to its LR image. But the property of these reconstruction-based SR technologies depended mightily on the prior information of HR images and degraded rapidly as the available input images was less or the desired amplification element was large. Consequently, the result could be overly smooth and lacking essential high-frequency information [13].

Learning based method, which restore the high frequency information lacked by LR image through training database, had attracted huge interest in various field. Freeman et al. [14] first proposed a relation model between the local regions of images and scenes by using the Markov network. Yue et al. put forward a new approach to hallucinate landmark images via extracting local descriptors from its up-sampled edition and bundled the descriptors to retrieve correlated images from the web according to their spatial relationship [15]. Polatkan et al. developed a novel Bayesian nonparametric method for super-

978-1-5090-2860-3/16/$31.00 ©2016 IEEE

resolution which used a beta-Bernoulli procedure to study the recurring visual modes named dictionary factors from the data [16]. Yang provided a foundational sparse coding (ScSR) algorithm to reconstruct HR images [13].

In the past few years, the deep-learning based method have demonstrated excellent performance in computer vision which had built more powerful models and designed more effective strategies against overfitting. The neural networks have became more capable to fit training data, due to the increased complexity, new nonlinear activations, and sophisticated layer designs. Consequently, in the direction of super-resolution reconstruction, many researchers began to explore the connections between the SR and convolutional neural networks (CNN) [17].

Motivated by tremendous performance of deep learning method in computer vision, we employ a deep convolutional neural networks for super-resolution of single medical image in this paper, which learns the nonlinear mapping from low-resolution space to high-resolution space directly. Also, we used three sets imaging data(Mammary gland, the prostate gland and the human brain) to train deep network respectively. Firstly, we used Randomized Rectified Linear Unit (RReLU), which incorporate a nonzero slope for negative part to solve the problem of over compression. Secondly, for the purpose of enhancing the quality of reconstructed image and reduce the noise of overfitting, Nesterovs Accelerated Gradient (NAG) method on the SRCNN is applied to accelerate the convergence of loss function and avoid the large oscillations. We obtained obvious ascension both in subjective and objective comparison with the bicubic interpolation on the corresponding test sets.

The rest of this work is formulated as follows. Section II provides the detailed algorithms of modified deep convolutional networks. The experimental results and corresponding discussions are given in Section III. Finally, the conclusion is exhibited in Section IV.

II. Method

Generally, the main idea of single-image SR is: suppose X is a LR image, we need reconstruct the corresponding HR image Y. The task of image super-resolution could be formulated as:

$$Y = HX + V \tag{1}$$

where H is the degradation matrix which means the composite operator of down-sampling and blurring, X implies the original image, and V denotes the noise item. Therefore SR problem can be described in mathematical language: given a low resolution image X, find the optimal estimation $F(X)$ of high resolution image Y. The following we will particularly introduce how to obtain the optimal estimation $F(X)$ via modified deep convolution neural network.

Inspired by the literature [17], our modified convolution neural network also employed three layers structure whose function are correspondence to the patch extraction and representation, non-linear mapping, reconstruction respectively. Beyond all this, we must have a preprocessing layer/step for data preparation before the network training.

A. Preprocessing

In order to get the required LR images, we set the preprocessing steps which apply bicubic interpolation to upscale low-resolution images for obtaining desired size before the training part. In the test part, we also adopt the same method to generate the desired size of LR images from the original HR images. Actually, we can regard the bicubic interpolation as a convolutional operation, and it could be explained as a convolutional layer(see Fig.2). Nevertheless, in the process of fact, there is a fractional steps because the size of output result is larger than the input in this layer. In order to take full use of the prevalent well-optimized implementations Caffe [18], we exclude this 'layer' from our training network.

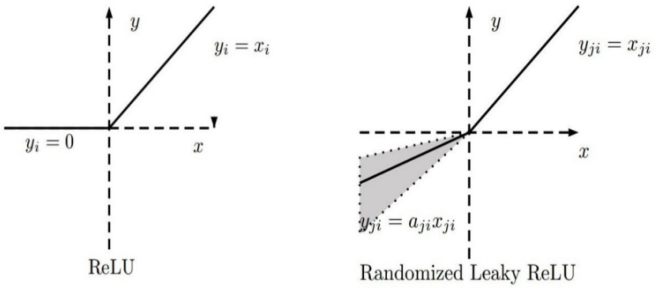

Fig. 1. For RReLU, a_{ji} is a random variable keeps sampling in a given range, and remains fixed in testing

B. Patch Extraction and Representation

Traditional neuron model is represented by the sigmoid activation function which evolve from the neuroscience. It will be excited when received signal exceeds a certain threshold. Also, it's appropriate that used to simulate the neurons received stimulation. However, there is a serious problem that it's easy to produce the saturation effect. Consequently, the function such as $abs(x), max(x, 0)$ came to the attention of the researchers. For the computers, it is easy to computer them by adjusting the sign bit. Besides, because the gradients of them is simple (-1, 1 or 0), No matter how many layer spread, the sum of gradients will remain in a relatively stable order of magnitude.

ReLU has the profound influence on neural network and first used in restricted Boltzmann machines [19], it can be defined as:

$$y = max(0, Wx + b) \tag{2}$$

Which is illustrated in Fig.1. However, when using ReLU, there is only one quadrant $(x > 0)$ of the information is retained and other areas of information has been compressed in different degrees. The amplitude is too big to recover. This kind of operation is unreasonable in applications.

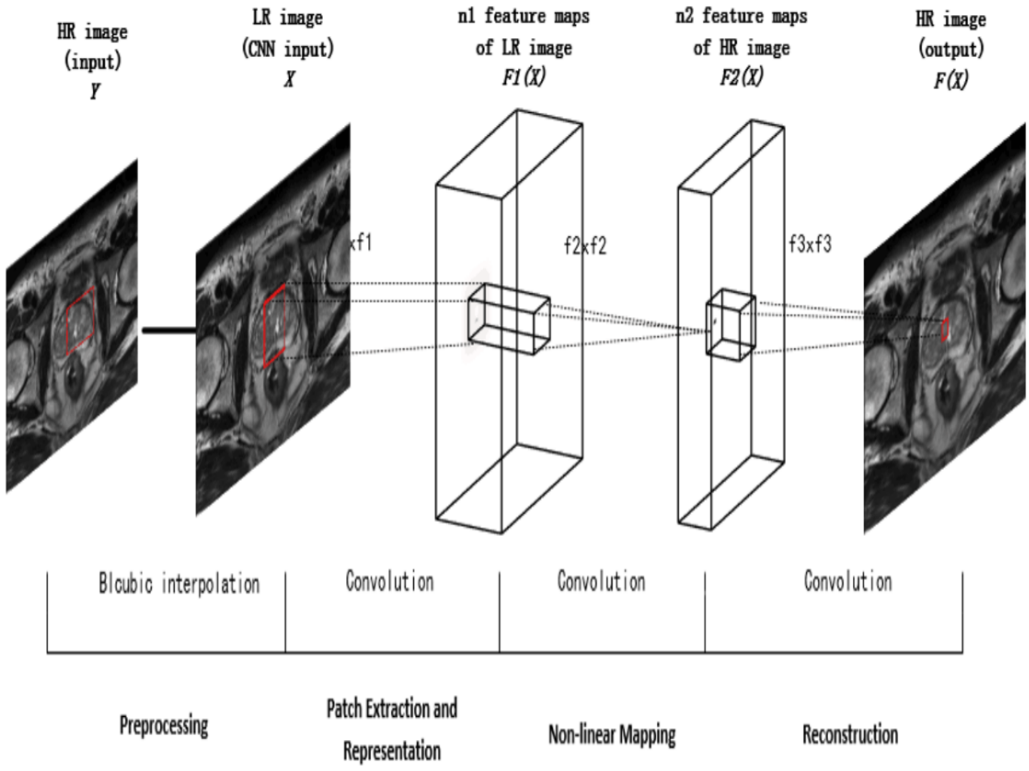

Fig. 2. the SRCNN structure chart.

For the purpose of avoiding this kind of defect, we apply the RReLU activation function in this paper which can be defined as:

$$y = \begin{cases} W_1 x + b_1, & \text{if } x > 0 \\ \frac{W_2 x + b_2}{a}, & \text{if } x < 0 \end{cases} \tag{3}$$

where a is random number sampled from a uniform distribution: $a \sim U(l, u)$. In order to be consistent with formula (2), we rewritten the formula (3) as:

$$y = max(0, W_1 x + b_1) + \frac{min(0, W_2 x + b_2)}{a} \tag{4}$$

which is illustrated in the right of Fig.1.

In the above section , we apply the RReLU to proceed the filter responses in the patch extraction and representation operation. The first layer of our method is expressed as an operation $F_1(X)$:

$$F_1(x) = max(0, W_1 * X + B_1) + \frac{min(0, W_1 * X + B_1)}{a} \tag{5}$$

where W_1 represents the filters and $*$ denotes the convolution operation. W_1 contains n_1 filters of size $c * f_1 * f_1$, c is the channel number of input image, f_1 implies the size of spatial filter and n_1 is the filters number. B_1 implies n_1-dimensional biases vector, whose element is related to a corresponding filter.

C. Non-linear Mapping

In the starting operation, we extracts a n_1-dimensional feature for each patch via a convolutional computation. The second layer map each n_1-dimensional vectors into n_2-dimensional. And the formulation of the second layer can be expressed as:

$$F_2(X) = max(0, W_2 * F_1(X) + B_2) + \frac{min(0, W_2 * F_1(x) + B_2)}{a} \tag{6}$$

where W_2 implies n_2 filters with size $n_1 \times f_2 \times f_2$, and B_2 is deviation. Each output n_2-dimensional vectors represents the high-resolution patch used for reconstruction in conceptually.

D. Reconstruction

In the paper [13], the final full image are usually produced by the predicted overlapping high-resolution patches via averaging. The averaging procedure could be seen as a pre-defined filter on a cluster of characteristic pattern [20]. Inspired by this, we can obtain the convolutional layer for producing the final HR image as:

$$F(X) = W_3 * F_2(X) + B_3 \tag{7}$$

where W_3 is a cluster of linear filters with size $n_2 \times f_3 \times f_3 \times c$ and B_3 is a c-demensional biases vector. An summarize of the network is exhibited in Fig.2.

E. Training

In order to learn the end-to-end mapping F, the Mean Squared Error (MSE) is used to estimate the network parameters $\Theta = \{W_1, W_2, W_3, B_1, B_2, B_3\}$ as the loss function in

this paper, and keep the loss function between the reconstructed results $F(X;\Theta)$ and corresponding ground truth HR images Y minimum. The MSE can be defined as:

$$L(\Theta) = \frac{1}{n}\sum_{i=1}^{n}||F(X_i,\Theta) - Y_i||^2 \tag{8}$$

As n represents training samples$'$ number.

Stochastic gradient descent(SGD) renews the weights W using the classical momentum(CM) method via a linear combination of previous weight update Δ_i and the negative gradient $\nabla L(W)$. The momentum μ $\in$[0,1] represents the weight of former update. The learning rate $\eta > 0$ implies the weight of negative gradient. Formally, we can use the following equations to estimate the renovate value Δ_{i+1}, the renewed weights W_{i+1} at iteration $i+1$, and the weight matrices as:

$$\Delta_{i+1} = \mu\Delta_i + \eta\nabla L(W_i^l) \tag{9}$$

$$W_{i+1}^l = W_i^l + \Delta_{i+1} \tag{10}$$

while $l \in \{1,2,3\}$ is the layers, i imply the iterations.

But in our study, We adopt the Nesterov$'$s accelerated gradient (NAG) [21] approach to updates the weights. And the weight renewal formulas is similar to the SGD updates shown above:

$$\Delta_{i+1} = \mu\Delta_i + \eta\nabla L(W_i^l + \mu\Delta_i) \tag{11}$$

$$W_{i+1}^l = W_i^l + \Delta_{i+1} \tag{12}$$

NAG first performs a partial update to W_i^l, computing $W_i^l + \mu\Delta_i$, which is similar to $W_(i+1)^l$, but missing the as yet unknown correction. This benign-looking difference seems to allow NAG to renew Δ in a faster way and obtain better response, which make NAG become more steady than CM in many cases, particularly for a higher μ. What$'$s more, when the optimization path taken by CM exhibits large oscillations along the high-curvature vertical direction, NAG is able to avoid these oscillations almost entirely. Which can confirm the intuition that it is much more effective than CM at decelerating over the course of multiple iterations. Consequently, the NAG is more tolerant of large values of μ compared to CM.

III. EXPERIMENTS

In this study, a lot of experimental results were exhibited to estimate the effectiveness of our modified method. In the experiments, the parameters of proposed method are set as follows: our experiments based on a three-layers SRCNN which the kernel size of the filter is $f_1 = 9$, $f_2 = 1$, $f_3 = 5$, respectively(9-1-5); We used three training sets (Mammary gland CT image, prostate gland MR image and brain tissue MR image) for deep network training by an upscaling factor 3; Then we set $n_1 = 64$, $n_2 = 32$ and μ=0.9 [22]; The initialized filter weights of each layer are randomly set as a Gaussian distribution with zero mean and standard deviation 0.001 (and biases is set as 0); The learning rate is fixed in this work with 10^{-4} for the first two layers and 10^{-5} for the last layer.

A. RReLU vs. ReLU

In this part, we compare the performance of two types of rectified activation functions (ReLU, RReLU) in SRCNN. The detail operation is that using the RReLU to replace the ReLU in the deep network and keeping the other parameters remain the same. And a is sampled from $a \sim U(3,8)$. In test time $a = \frac{11}{2}$, we use the same configuration in this paper. Then we using the peak signal noise ratio (PSNR) to evaluate and calculate the capacity(see Fig.3). And according to the shape of the curve, we also employ PSNR and structural similarity index measurement(SSIM) to compared their performance after 4×10^6 iterations (see Fig5).

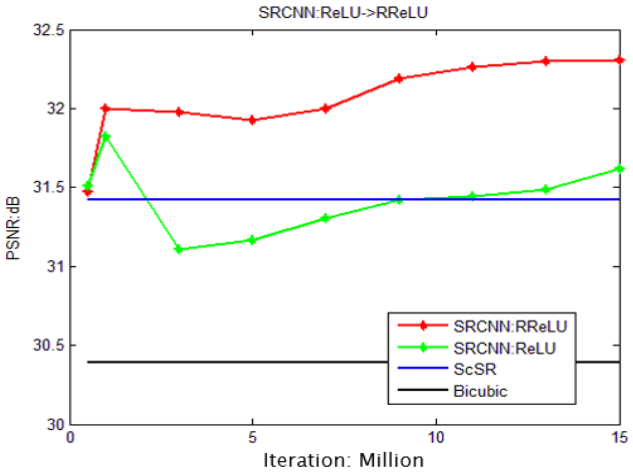

Fig. 3. The average PSNR comparison using different rectified linear unit. In the process of training, we find that the loss function to achieve the convergence after 1×10^6 iterations. So we start sampling from 1×10^6 iterations and the sampling interval is 2×10^6 (the same to Fig.4).

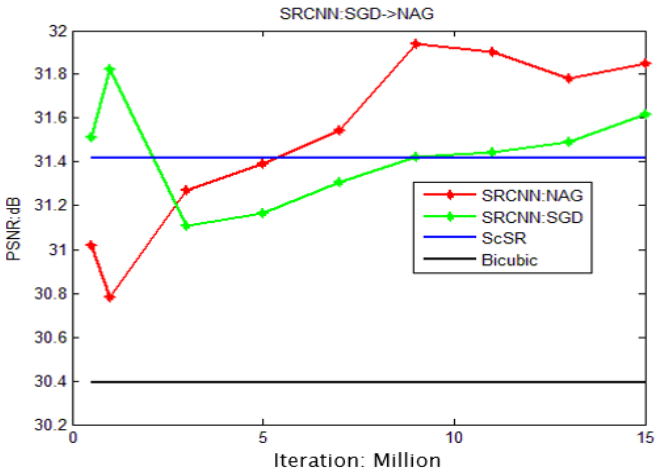

Fig. 4. The average PSNR comparison using different weight updating unit.

B. NAG vs. SGD

NAG is an "optimal" approach of convex optimization, obtaining a convergence degree of $O(\frac{1}{t^2})$ rather than $O(\frac{1}{t})$. In

fact, NAG is a valid approach to optimize the deep learning framework with decided types. From the weight value W where we figure the error gradient $\nabla L(W)$, we can distinguishes the approach from SGD. The detailed description is shown in Section II.E. In that part, we compared the influence for reconstruction quality on difference weight updating methods which show the performance in Fig.4.

(a)

Methods \ Images		1	2	3	4	5	6	7	8	Average
Bicubic	PSNR	25.70	25.51	28.20	27.91	27.49	31.15	30.43	30.27	28.33
	SSIM	0.6863	0.6902	0.7640	0.7533	0.7435	0.8063	0.8001	0.7985	0.7553
Ours	PSNR	26.27	26.10	28.86	28.60	28.16	31.83	31.12	31.00	**28.99**
	SSIM	0.7296	0.7333	0.7955	0.7867	0.7784	0.8330	0.8276	0.8271	**0.7889**

(b)

Methods \ Images		1	2	3	4	Average
Bicubic	PSNR	36.71	35.10	35.35	38.72	36.47
	SSIM	0.9430	0.9288	0.9912	0.9168	0.9450
Ours	PSNR	37.21	35.4	35.72	39.73	**36.98**
	SSIM	0.9557	0.9415	0.9948	0.9342	**0.9566**

(c)

Methods \ Images		1	2	3	4	5	6	Average
Bicubic	PSNR	46.41	44.83	44.11	44.09	43.85	34.39	42.95
	SSIM	0.9979	0.9974	0.9967	0.9966	0.9967	0.9892	0.9958
Ours	PSNR	49.13	47.97	47.44	47.29	46.99	37.92	**46.12**
	SSIM	0.9992	0.9990	0.9988	0.9987	0.9987	0.9951	**0.9983**

Fig. 5. The results of PSNR (dB), SSIM comparison on the prostate(a), breast(b), brain(c) image respectively.

C. Comprehensive Experiment Evaluation

In this part, the qualitative and quantitative results of proposed method is shown in Fig.5 and Fig.6. The related implementations using in this paper are publicly available codes offered by the authors, and down-sampling is processed on all images with same bicubic kernel. The training was run on the platform of GTX 960 GPU.

Fig.6 show the super-resolution results of different medical data. As can be observed, ours method obtained a satisfactory result than other methods without any visible artifacts in the image.

IV. Conclusion

This study provide a deep learning methods for super resolution which learns the nonlinear mapping from low-resolution space to high-resolution space directly. We use Randomized Rectified Linear Unit (RReLU) to solve the problem of over compression and apply Nesterovs Accelerated Gradient (NAG) method on the SRCNN to accelerate the convergence of loss function and avoid the large oscillations. The experimental results have exhibited the superior results in both subjective and objective evaluations compared with other methods. According to the experimental results, we firmly believe that we would obtain a nice result via enough training iterations (such as 8×10^8 [17]).

Acknowledgment

This paper was supported by National Natural Science Foundation of China Grant No: 61371156 and Anhui Province Scientific and Technological Research Programs Grant No: 1401B042019. The authors would like to thank the anonymous reviewers for their helpful and constructive comments and suggestions.

References

[1] X. Lu, Z. Huang, and Y. Yuan,"MR image super-resolution via manifold regularized sparse learning," *Neurocomputing*, vol.162, no.3, pp.94-104, 2015.

[2] A. Rueda, N. Malpica, and E. Romero, "Single-image super-resolution of brain MR images using overcomplete dictionaries," *Medical Image Analysis*, vol.17, no.1, pp.113-132, 2012.

[3] Y. Zhang, Z. Dong, P. Phillips, and et al, "Exponential Wavelet Iterative Shrinkage Thresholding Algorithm for compressed sensing magnetic resonance imaging," *Information Sciences*, vol.322, no.1, pp.115-132, 2015.

[4] R. Summers, "Evaluation of Computer-aided Detection Devices: Consensus Is Developing," *Academic Radiology*, vol.19, no.4, pp.377-379, 2012.

[5] F. Ciompi, C. Jacobs, E. Scholten, and et al, "Bag-of-Frequencies: A Descriptor of Pulmonary Nodules in Computed Tomography Images," *IEEE Transactions on Medical Imaging*, vol.34, no.4, pp.962-973, 2015.

[6] M. Irani, and S. Peleg, "Improving resolution by image registration," *Graphical Models and Image Processing*, vol.53, no.3, pp.231-239, 1991.

[7] R. Tsai, T. Huang, "Multiple frame image restoration and registration," *Advances in Computer Vision and Image Processing*, 1984.

[8] H. Greenspan, G. Oz, N. Kiryati, and S. Peled, "MRI inter-slice reconstruction using super-resolution," *Magnetic Resonance Imaging*, vol.20, no.5, pp.437-446, 2002.

[9] F. Calamante, J. Tournier, R. Smith, and A. Connelly, "A generalised framework for super-resolution track-weighted imaging," *NeuroImage*, vol.59, no.3, pp.2494-2503, 2012.

[10] S. Farsiu, M. Robinson, M. Elad, and P. Milanfar, "Fast and robust multiframe super resolution," *IEEE Trans. Image Process*, vol.13, no.10, pp.1327-1344, 2004.

[11] F. Zhou, W. Yang, and Q. Liao, "Interpolation-Based Image Super-Resolution Using Multisurface Fitting," *IEEE Transactions on Image Processing*, vol.21, no.7, pp.3312-3318, 2012.

[12] Q. Yan, Y. Xu, X. Yang, and T. Nguyen, "Single Image Superresolution Based on Gradient Profile Sharpness," *IEEE Transactions on Image Processing*, vol.24, no.10, pp.3187-3202, 2015.

[13] J. Yang, J. Wright, T. Huang, and Y. Ma: "Image super-resolution via sparse representation," *IEEE Trans. Image Processing*, vol.19, no.11, pp.2861-2873, 2010.

[14] W. Freeman, T. Jones, and E. Pasztor, "Example-based super resolution," *IEEE Comput. Graph.*, vol.22, no.2, pp.56-65, 2002.

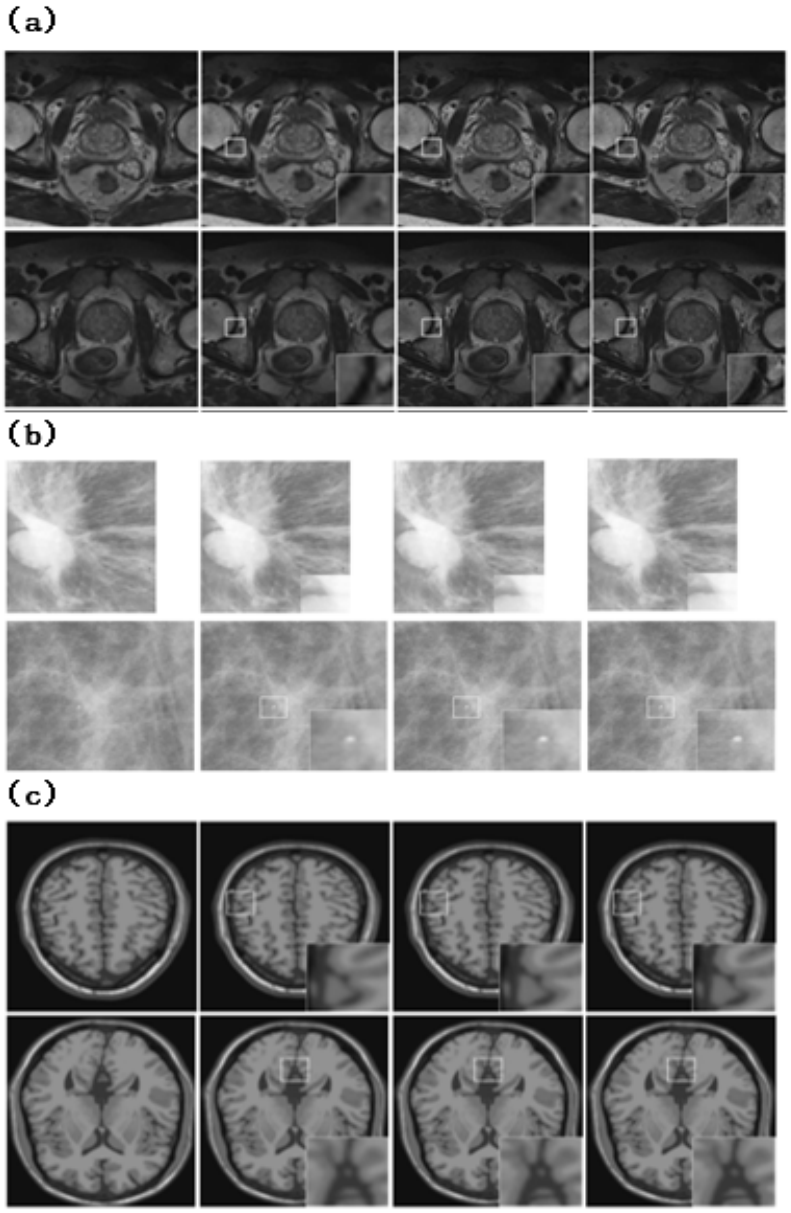

Fig. 6. Corresponding result of prostate(a), breast(b), brain(c) image respectively. The first, second, third and fourth columns show the LR image, the processing result of bicubic interpolation meyhod, the processing result of our method and the original image.

[15] H. Yue, X. Sun, J. Yang, and F. Wu: "Landmark Image Super-Resolution by Retrieving Web Images," *IEEE Transactions on Image Processing*, vol.22, no.12, pp.4865-4878, 2013.

[16] G. Polatkan, D. Blei, I. Daubechies, and L. Carin, "A Bayesian Non-parametric Approach to Image Super-Resolution," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol.37, no.2, pp.346-358, 2015.

[17] C. Dong, C.C. Loy, K. He, and X. Tang, "Image Super-Resolution Using Deep Convolutional Networks." *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol.38, no.2, pp.295-307, 2016.

[18] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, and et al, "Caffe: Convolutional Architecture for Fast Feature Embedding," *Eprint arXiv*, pp.675-678, 2014.

[19] V. Nair, and G.E. Hinton, "Rectified linear units improve restricted Boltzmann machines." *In ICML*, pp.807-814, 2010.

[20] J. Melendez, C. Snchez, G.B. Van, and et al, "Improving mass candidate detection in mammograms via feature maxima propagation and local feature selection," *Medical Physics*, vol.41, no.8, pp.081904-081904, 2014.

[21] J. Wright, Y. Ma, J. Mairal, and et al, "Sparse Representation for Computer Vision and Pattern Recognition," *Proceedings of the IEEE*, vol.98, no.6, pp.1031-1044, 2010.

[22] I. Sutskever, J. Martens, G. Dahl, and G. Hinton, "On the Importance of Initialization and Momentum in Deep Learning." *Proceedings of the 30th International Conference on Machine Learning, 2013*, 2013.