UNet-ESPC-Cascaded Super-Resolution Reconstruction in Spectral CT

Zhouxia Li, Xiaoni Wang, Lihui Wang, Wen Ji, Miaomiao Zhang, Yuemin Zhu, Feng Yang

UNet-ESPC-Cascaded Super-Resolution Reconstruction in Spectral CT

Zhouxia Li[1], Xiaoni Wang[1], Lihui Wang[3], Wen Ji[4], Yuemin Zhu[5],*, Feng Yang[2],*

[1]*School of Computer and Information Technology, Beijing Jiaotong University, No. 3 Shangyuancun, Beijing 100044 China*
[2]*National Library of Medicine, National Institute of Health, 8600 Rockvill Pike, Bethesda, MD 20894, USA*
[3]*Key Laboratory of Intelligent Medical Image Analysis and Precise Diagnosis of Guizhou Province, College of Computer Science and Technology, Guizhou University Guiyang 550025 China*
[4]*Beijing Key Laboratory of Mobile Computing and Pervasive Device, Institute of Computing Technology, Chinese Academy of Sciences, No. 6 Kexueyuannanlu, Beijing 100190 China*
[5]*Univ Lyon, INSA Lyon, CNRS, Inserm, CREATIS UMR 5220, U1206, F-69621, Lyon, France*
Email: feng.yang2@nih.gov, yue-min.zhu@creatis.insa-lyon.fr

Abstract—Spectral CT based on photon counting detectors is a promising imaging modality since it provides the possibility of both obtaining CT images from multi-energy bins with a single X-ray exposure and allowing low-dose imaging. However, the image quality such as spatial resolution reconstructed from multiple energy bins is degraded because of the use of narrow energy bins in spectral CT. We propose to use deep learning methods for super-resolution reconstruction of spectral CT images. To this end, we introduce an UNet-ESPC-cascaded model and perform a patch-based training to obtain the optimal parameters of the model. Experimental results on physical phantom datasets demonstrated that our deep learning based reconstruction method can reduce the F form error between the reconstructed super-resolution CT image and the ground truth, by 11.6% and 5.66% with respect to respectively bilinear-interpolation-based reconstruction and iterative back projection methods. Our method achieves best results with a patch size of 20 and a stride of 15.

Keywords—Spectral CT, Super Resolution, Deep Learning, UNet, ESPC

I. INTRODUCTION

Spectral computed tomography (CT) detects photons in all energy bins through using photon counting detectors (PCD), reconstructs images from multi-energy projection data [4], and distinguishes materials either directly from multi-energy attenuation coefficients or from multi-energy reconstructed spatial images [1, 2]. PCD-based spectral CT, simply named spectral in what follows, allows for low-dose imaging while achieving material decomposition. However, spectral CT suffers from degraded image quality such as low signal-to-noise ratio (SNR) and unclear images. Especially, when distinguishing materials of low atomic numbers, the materials tend to disappear in the image because spatial resolution is lacking. Super-resolution [3] can provide a solution to this problem and is also beneficial to the later use and processing of multi-energy CT images.

At present, dual spectral CT [1] is the main technology in medicine, and spectral CT has not been widely used. Since the lack of clinical spectral CT data, we choose physical phantoms. And the study of phantoms is conducive to grasp the characteristics of spectral CT images, and can provide theoretical support for the real spectral CT data.

So far, there are mainly three kinds of methods for super-resolution image reconstruction: interpolation-based methods [5, 6, 7], reconstruction-based methods [8, 9, 10] and learning-based methods [11, 12]. Interpolation-based methods fill in the missing pixels at high resolution by taking use of the pixel information at low resolution. Reconstruction-based method first determines a back projection filter and then applies a residual-driven approach to get high resolution images. Sparse learning methods and deep learning methods are both included in learning-based, which use data-driven to get the mapping between low and high resolution images.

Deep learning has gained popularity for super-resolution interpolation of natural images [13, 14, 15, 16]. Recently, it has also been applied to medical image reconstruction [17, 18, 19, 20]. However, it asks for a prior interpolation of samples for training, which means getting the same dimensions as the high resolution images, and thus results will be affected by different prior interpolations. In this paper, we propose an UNet-ESPC-based super-resolution model for spectral CT reconstruction, which does not need any prior operation. This work mainly contributes in two aspects. First, it is a new attempt to apply deep learning method to spectral CT image data. Second, Experimental results show that the Dual Network structure (i.e., the UNet-ESPC-cascaded network) is suitable for super resolution reconstruction in spectral CT.

The rest of the paper is organized as following. Section 2 describes the details of our method. Section 3 presents the experimental results, followed by conclusion and discussion in Section 4.

II. METHODS

A. Assumption of model

Assume that a low-resolution image $\mathcal{L}$ with a height H and a width W is interpolated with a scale factor r to the high resolution image $\mathcal{H}$. The low-resolution image $\mathcal{L}$ can be obtained by a filter $\mathcal{F}$ and a pool $\mathcal{M}$ from the high-resolution image $\mathcal{H}$, defined by Eq. (1).

$$\mathcal{L} = \mathcal{M}\big(\mathcal{F}(\mathcal{H})\big) + \varepsilon, \qquad (1)$$

where ε is noise.

B. Super resolution reconstruction

a) Interpolation-based methods[5, 6, 7]

In the present work, we use an interpolation factor $r=4$, so
$$\mathcal{L}(k, l) = \mathcal{H}(i, j), i = 4k - x, j = 4l - y, (x=0,1,2,3, y=0,1,2,3).$$

b) IBP(iterative back projection)-based reconstruction [8, 9, 10]

The principle of the reconstruction-based methods can be explained by the following algorithm. We designate the interpolated low resolution images as ILR and HR and HR' are both initialized as $\mathcal{L}$. GT represents ground truth. Since the spectral CT images collected from machine are high enough, we use it as the ground truth in this work and use Eq. (1) to get the low resolution images in the algorithm.

---

1. Get $\mathcal{L}$, $\mathcal{L} = \mathcal{M}\big(\mathcal{F}(\text{GT})\big)$;
2. ILR=$Interpolate(\mathcal{L})$ and HR= $\mathcal{L}$, $HR' = \mathcal{L}$;
3. Repeat:

   $HR' = Interpolate(\mathcal{M}\big(\mathcal{F}(HR')\big))$;

   error=HR-HR';

   $HR' = HR' + \sigma^{BP}$ (error);

   until error is lower than a given precision value.

In the above reconstruction algorithm, HR denotes original interpolated low resolution image. HR' is iterated for several times. When it is close to HR, HR' is close to GT. The algorithm is proved to be convergent[9]. σ^{BP}, which is related to back projection operator, is a super parameter. It is difficult to determine the IBP (iterative back projection) filter σ^{BP}, which projects error onto high resolution images. In the literature, empirical judgment is used.

c) Our Learning-based method: UNet-ESPC-Cascaded super-resolution reconstruction

UNet is the most widely used model in medical image segmentation [22, 23], with the input data first down-sampled and then transposed and convoluted to output data with the same size as the input. ESPC network is one of the most popular convolutional neural networks used for super-resolution reconstruction [21]; it does not need any prior operation on low-resolution images.

In the present study, we use the cascade of UNet and ESPC as the backbone of our model. The UNet is used as the embedding part, ESPC as the inference part, and the final sub-pixel convolution layer as the reconstruction part. We first filter the high-resolution data of spectral CT, and use the down-sampled images as the low-resolution images to be interpolated. Then, we split low-resolution images into patches, based on which, our UNet-ESPC-cascaded model is trained and tested. This is to avoid the overfitting problem caused by lack of data. The interpolated high resolution patches are integrated to get the final super-resolution images. We use a Gaussian filter of 5×5 and a down-sampling factor of 4 in the first step. The framework of our proposed method is depicted in Fig. 1. Figure 2 and Fig. 3 show the customized UNet and ESPC structures of the framework, respectively.
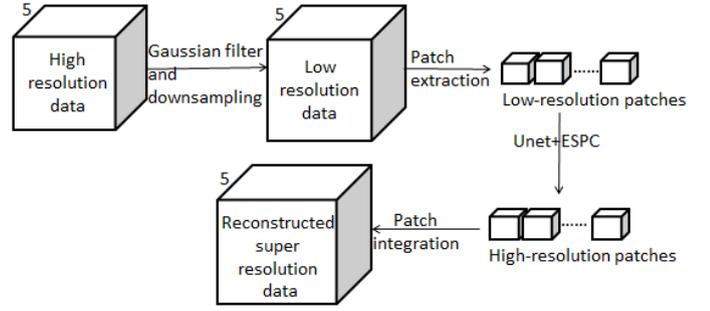

Fig. 1. Framework of our proposed UNet-ESPC cascaded model for spectral CT super-resolution reconstruction.
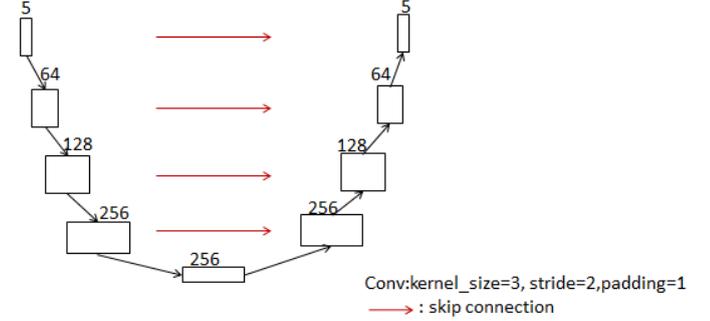

Fig. 2. UNet structure in our framework in Fig.1. 5,64,128,256 are channel numbers in each layer of the network.
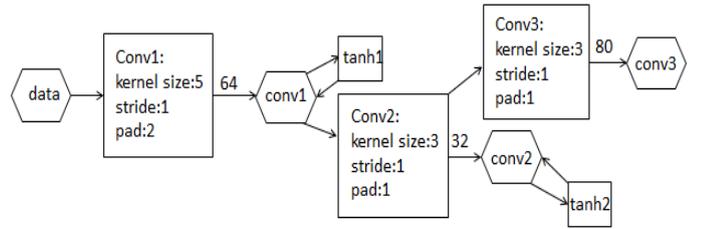

Fig. 3. ESPC structure in our framework in Fig.1.

Five energy bins in spectral CT are used as five channels. These five channels are 4 times interpolated from the low resolution images. Using our method, we obtain a data of 80 channels of which the first 16 channels will be integrated into one channel that corresponds to the interpolation results of the first energy bin [21], and the integrated 16 to 32 channels correspond to the interpolation results of the second energy bin, et cetera.

The objective function of this model is defined as follows:

$$min\ F = \sum_{e=1}^{5} \sum_{i=1}^{patch\_size} \sum_{j=1}^{patch\_size} |x_{eij} - \hat{x}_{eij}|, \tag{2}$$

where x_{eij} represents the attenuation coefficient of (i, j) in the e-th energy bin, and $\hat{x}_{eij}$ is the estimate of x_{eij}.

III. EXPERIMENTS AND RESULTS

A. Experimental data introduction

Two phantoms, Phantom1 and Phantom2, were prepared and then scanned by Philips Spectral CT under the current of 220 mA and the voltage of 120kvp at CREATIS laboratory of the National Institute of Applied Sciences in

Lyon (INSA Lyon), France. Phantom1 was filled with iodine (I) of different concentrations and Phantom2 with gadolinium (Gd) of different concentrations, as shown in Fig. 4. Each set of data is a four-dimensional data of $900 \times 900 \times 5 \times 9$, in which 900 is the image size whose pixels are material attenuation coefficients, 5 represents the number of energy bins, and 9 indicates the number of slices. The X-ray energy bins were set as: 30-51 kev, 51-62 kev, 62-72 kev, 72-81 kev and 81-150 kev. In our experiments, the original high resolution spectral CT images of the two phantoms are considered as ground truth. We use different methods to reconstruct images with the same resolution as the ground truth and then compare their performances.

Fig.4. Two phantoms filled by different materials with different concentrations. Left: Phantom 1 filled by I. Middle: Phantom 2 filled by Gd (unit: mg/cc, NA indicates no filled material). Right: Original image of the 5th slice of Phantom 1 for the 1st energy bin.

B. Evaluation metrics

To evaluate the performance of super-resolution reconstructions, we calculate the error between the reconstructed image and the original image, which is defined as F norm, a kind of matrix norm, of the difference between the original image and the reconstructed super-resolution image.

C. Reconstruction results based on traditional methods.

a) With interpolation-based method

Figure 5 shows the reconstruction results using a bilinear interpolation on each slice. Fig. 5(a) and (b) respectively indicates reconstructed super resolution image and the ground truth. It is observed blurred boundary and in the reconstructed image.
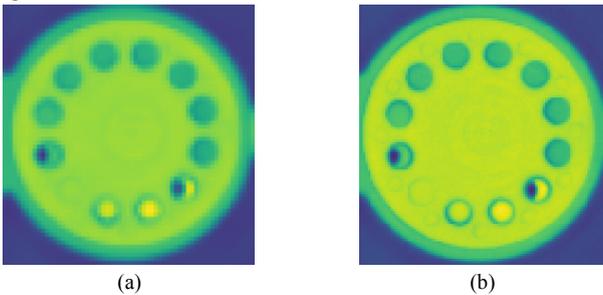

Fig. 5. Bilinear interpolation based reconstruction of the 6th slice of Phantom 2 for the 3rd energy bin.

b) With IBP

The IBP based reconstruction result of the 6th slice of Phantom2 for the 3rd energy bin is shown in Fig. 6(a). Serration and the blurred boundary are observed in the filling areas.

The F form errors are calculated for slice4-slice6 of Phantom2 for the 5 energy bins for both bilinear-interpolation-based and IBP-based reconstructions, as shown in Fig. 7. We see that the error of IBP-based reconstruction is obviously smaller than that of bilinear-interpolation-based reconstruction.

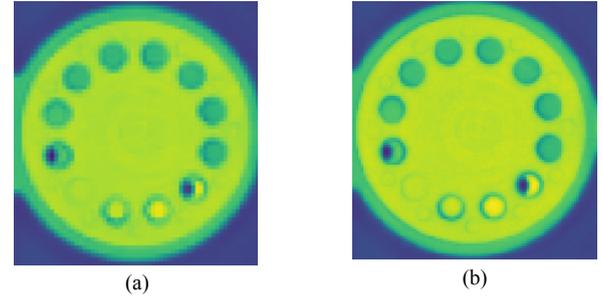(a)   (b)

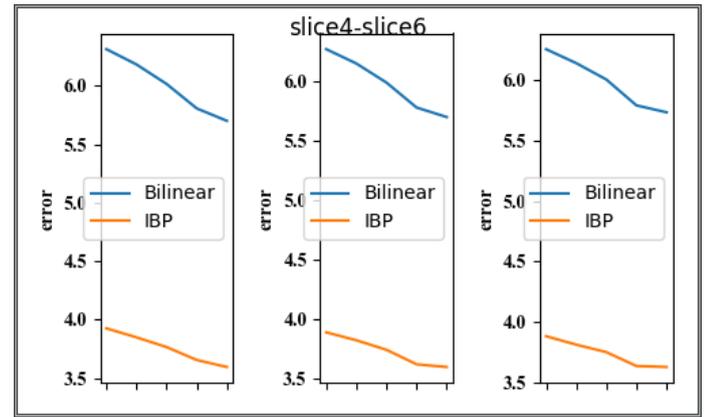Fig. 6. Iterative back projection results.

Fig. 7. Comparison of interpolation and iterative back projection error in slice4-slice6 of Phantom2. The x-axis indicates the energy bin number, and y-axis indicates the F form error.

D. Deep Learning method

For our UNet-ESPC-based method, we first decide the patch size and stride by experiments, and then use data augmentation to reduce the F form error.

a) Determine patch size

Extracted patches from the nine slices of Phantom1 are used as training set and those from Phantom2 as testing set. By setting the stride as 5, we compare three patch sizes of 15×15, 20×20 and 25×25, and the F norm error of the slice4-slice6 of Phantom 2 for the 5 energy bins is given in Fig. 8. We see that the lowest error is achieved with patch size 20×20 on the 4-6 slices. The third energy bin get the minimal error. Fig. 9 shows the qualitative results of the reconstructed images with different patch sizes.
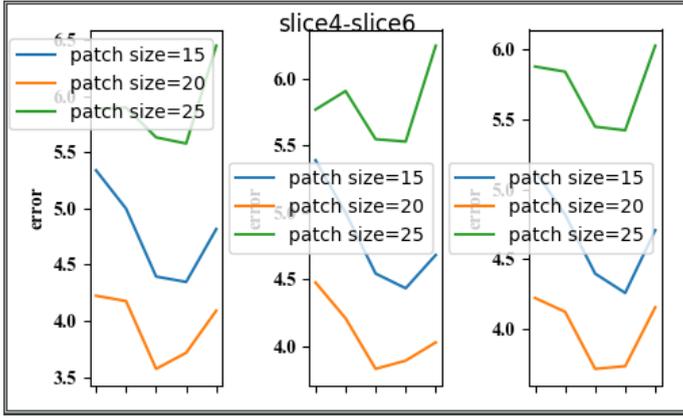
Fig. 8. Patch size comparison in slice4-slice6 of Phantom2. The x-axis indicates the energy bin number, and y-axis indicates the F form error
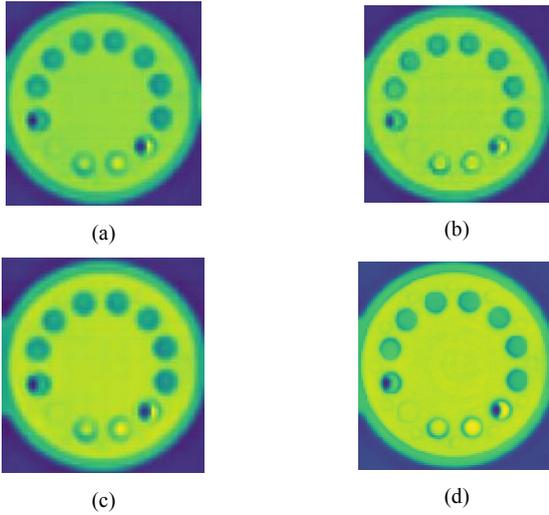

Fig. 9. Deep learning based super-resolution reconstruction of the sixth slice in of Phantom2 for the third energy bin. (a) Reconstruction result with patch size 15×15. (b) Reconstruction result with patch size 20×20. (c) Reconstruction result with patch size 25×25. (d) Ground truth.

b) Determine the stride

By fixing the patch size as 20×20, we compare the results with three different strides 5, 10 and 15. We use the extracted patches from 15 slices (9 slices of Phantom1 and slices 4 to 9 of Phantom2) for as training set. With the strides of 5, 10 and 15, so 26460, 6615 and 2940 patches are generated for each channel, respectively. Figure 10 shows the F form error with different strides in slice2 of Phantom2, having similar result with slice1 and slice3. We observe that the stride 15 achieves the smallest error.
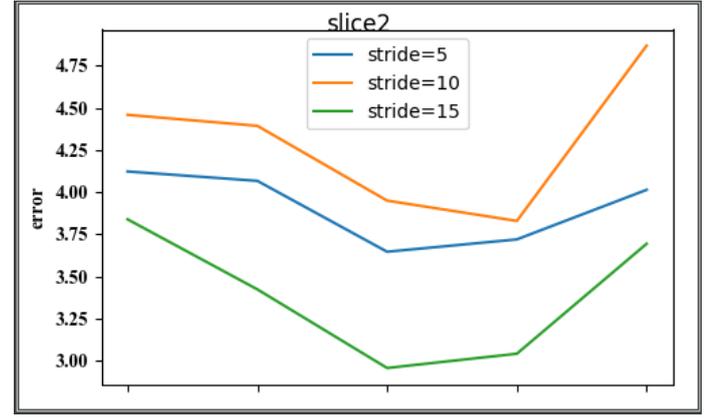

Fig. 10. Stride comparison on the slice2 of Phantom2 for the 5 energy bins. The x-axis indicates the energy bin number, and y-axis indicates the F form error.

c) Data augmentation

Since only 2940 patches can be obtained for training when the stride is 15, we perform a data augmentation by rotating patches of 90, 180 and 270 degrees respectively. As a result, 8610 patches are finally obtained for training. Fig. 11 plots the F form error of slice2 in the testing data for the five energy bins. Besides, the results of the slice1 and slice3 with augmentation are effective than those without augmentation. It is observed that the F form error of our model with data augmentation is reduced about 8.71% on average, compared to that without augmentation.
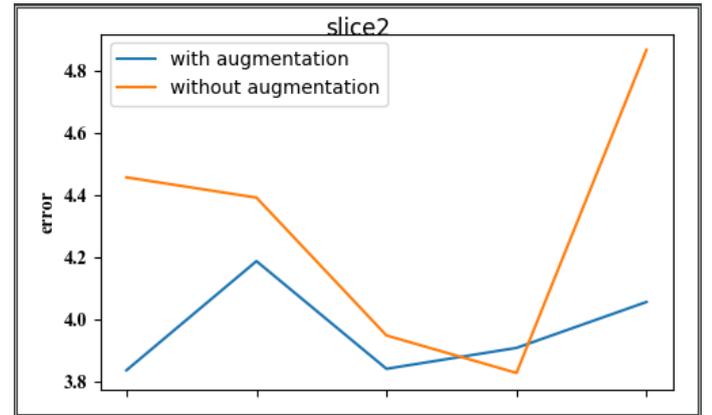

Fig. 11. Error comparison on slice2 of Phantom2 for the 5 energy bins between with and without data augmentation. The x-axis indicates the energy bin number, and y-axis indicates the F form error.
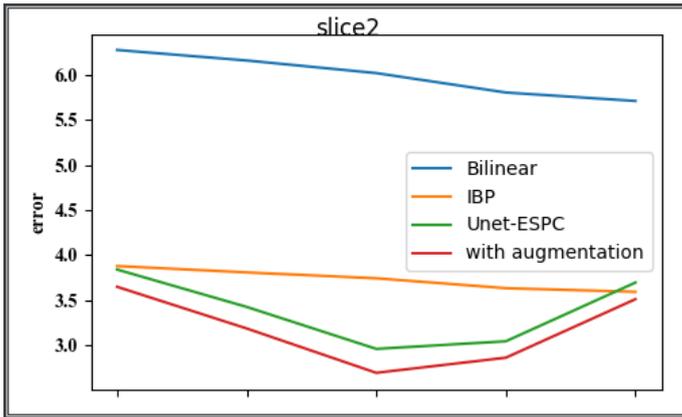
E. Comparison of all methods

Fig. 12. Error comparison on slice2 of Phantom2 for the 5 energy bins among all methods. The x-axis indicates the energy bin number, and y-axis indicates the F form error.

IV. DISCUSSION AND CONCLUSION

In this paper, we proposed a UNet-ESPC-cascaded model for super-resolution reconstruction in spectral CT. Our model does not ask for prior interpolation for training, and minimizes the serration and blurring phenomenon during reconstruction. Experimental results showed that our model can reduce the F form error by 11.6% and 5.66% compared to the bilinear-interpolation-based reconstruction and IBP-based reconstruction, respectively. In addition, since deep learning based reconstruction is a data-driven method, it does not ask for any prior operation such as the determination of filter in IBP-based method.

When fine-tuning the parameters of our model, we found that the first and ninth slices of Phantom2 have the largest error in the testing data. This is due to the fact that the filling areas of the 9th slice are too large to be reconstructed. We also found that the third energy bin contributes to the smallest error, while the first and the fifth energy bins contribute to the highest. This may be due to the fact that the photon numbers of the first bin is at the lowest level and the fifth energy bin at the highest level. Consequently, the image of gadolinium / iodine at the same concentration is dark (in the first energy bin) / bright (in the fifth energy bin), far from the third energy bin in the middle level.

REFERENCES

[1] Mccollough C H , Leng S , Yu L , et al. Dual- and Multi-Energy CT: Principles, Technical Approaches, and Clinical Applications[J]. Radiology, 2015, 276(3):637-653.

[2] Schirra, Carsten O., Brendel, Bernhard, Anastasio, Mark A., & Roessl, Ewald. . Spectral ct: a technology primer for contrast agent development. Contrast Media & Molecular Imaging, 9(1), 62-70.

[3] Zeng K, Ding S F.. Research progress of image super resolution reconstruction [J]. Computer Engineering and Applications, 053(016):29-35.

[4] Willemink, Martin J, Noël, Peter B. The evolution of image reconstruction for CT—from filtered back projection to artificial intelligence[J]. European Radiology.

[5] Li C J. Research on metal Artifact Removal Algorithm of CT Images based on Hybrid Interpolation [D]. Southern Medical University, 2009.

[6] Rajan D , Chaudhuri S . Generalized interpolation and its application in super-resolution imaging[J]. Image & Vision Computing, 2001, 19(13):p.957-969.

[7] Ding H Y, Zheng-Fu B. Super-resolution reconstruction of remote sensing image based on interpolation [J]. Microcomputer information, 2008, 24(21):292-293

[8] Irani M , Peleg S . Improving resolution by image registration[J]. CVGIP: Graphical Models and Image Processing, 1991, 53(3):231-239.

[9] Guo W W, Zhang P Z. Super-resolution image reconstruction based on iterative back projection [J]. Computer Science and Exploration, 2009(03):101-109.

[10] Tao Z Q; Li H L; Zhang H B. An iterative back-projection super-resolution reconstruction algorithm based on new edge-guided interpolation. Computer Engineering, 2016, 42.6: 255-260.

[11] Yang, Jianchao, et al. "Image super-resolution via sparse representation." IEEE transactions on image processing 19.11 (2010): 2861-2873.

[12] Zhang Y , Mou X , Wang G , et al. Tensor-based Dictionary Learning for Spectral CT Reconstruction[J]. IEEE Transactions on Medical Imaging, 2016:1-1.

[13] Dong, Chao, Loy, Chen Change, He, Kaiming, et al. Image Super-Resolution Using Deep Convolutional Networks[J]. IEEE Trans Pattern Anal Mach Intell, 38(2):295-307.

[14] Ledig C , Theis L , Huszar F , et al. Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network[C]// 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2017.

[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. 2016.

[16] LEDIG, Christian, etal. Photo-realistic single image super-resolution using a generative adversarial network. In: Proceedings of the IEEE conference on computer vision and pattern recognition.2017.p.4681-4690.

[17] Gao Y X. Research on super-resolution reconstruction algorithm of medical image based on convolutional neural network [D].

[18] Wu K Y. Research on image super resolution reconstruction algorithm based on deep learning [D].

[19] Xing X Y. Super-resolution reconstruction of medical image based on deep learning [D]. 2018.

[20] Wu X P. Super resolution reconstruction of medical images [D]. Central south university for nationalities, 2015.

[21] SHI, Wenzhe, et al. Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2016. p. 1874-1883.

[22] Li M D. Regional segmentation of pulmonary nodules based on convolutional neural network [D].

[23] Ronneberger, Olaf, Philipp Fischer, and Thomas Brox. "U-net: Convolutional networks for biomedical image segmentation." International Conference on Medical image computing and computer-assisted intervention. Springer, Cham, 2015.