Residual CNN-based Image Super-Resolution for CT Slice Thickness Reduction using Paired CT Scans : Preliminary Clinical Validation

Woong Bae **Seungho Lee** **Gwangbeen Park** **Hyunho Park** **Kyuhwan Jung**
VUNO Inc.
Seoul, South Korea
{iorism, bleaf.lee, pgb, hyunho.park, khwan.jung}@vuno.co

Abstract

We propose a 2.5D image super resolution (SR) network based on fully residual convolutional neural networks(CNN) which reduce the effective slice thickness of CT scans. We demonstrate that the proposed network quantitatively outperforms the 2017 NTIRE winning method, when trained and tested with 100 pairs of chest CT scans acquired with different slice thickness (1 mm, 3 mm, 5 mm). Based on the knowledge of CT reconstruction algorithms, we also demonstrate the necessity of using real pairs of CT scans with different slice thickness rather than using simulated low-resolution data which are widely used in image super resolution studies. Furthermore, when the proposed SR method was applied, for CT images that are 3mm and 5mm slice thickness, we confirmed dramatic performance improvement in the CNN based lung nodule detection network.

1 Introduction

Most of deep learning based methods suffer from performance degradation when they are applied to the clinical environment due to the variability of image acquisition protocol such as slice thickness in the case of CT scans. Though it is known that thin slice thickness (<2.5mm) results in better performance both for the human reader and computer-aided detection(CAD) systems, thick slice thickness(>2.5mm) CT scans are widely used in clinical setting mainly because of the reading time efficiency and storage reduction. To overcome this limitation, in this paper, we propose a effective SR method for CT slice thickness reduction using CNN and demonstrate the quantitative performance for both Super Resolution and detection of lung nodules with real CT paired images.

On the other hand, in most SR papers, the low resolution (downsampled) image is generated by simulation using predefined simple kernel functions [3, 4]. However, low resolution CT images generated from high resolution using these simple kernels show significantly different characteristics and quality compared to the real low resolution CT scans. Therefore, in this paper, we show how important it is to use real paired data.

2 Methodology

As shown in Figure 1.(d), the SR for CT slice thickness aims to improve the resolution only in the depth direction, which can be interpreted as spatial information of the coronal and sagittal planes. Our main framework of our proposed network is shown in Figure 1.(a). Our proposed network is divided into preprocessing, non-linear mapping and reconstruction parts. The preprocessing and non-linear mapping parts consist of residual block (resblock) proposed in [4] and subpixel shuffling is used as an operator to create high resolution images. The preprocessing networks that handle

1st Conference on Medical Imaging with Deep Learning (MIDL 2018), Amsterdam, The Netherlands.

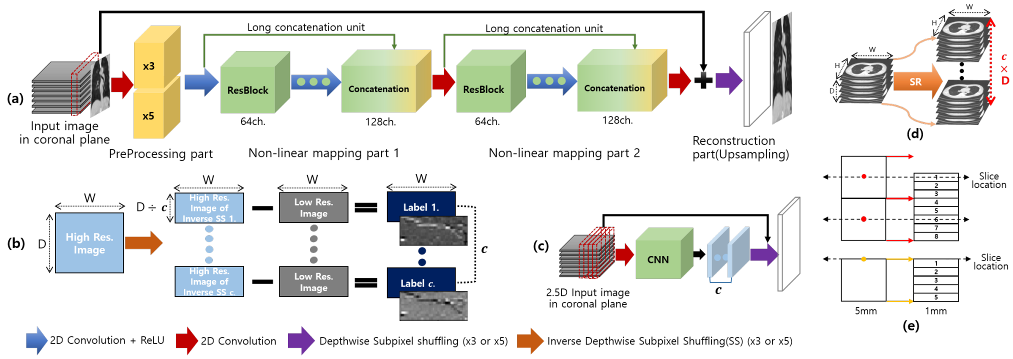

Figure 1: (a) Main framework of the proposed SR network for CT slice thickness. (b) Illustration of generating labels by residuals of inverse subpixel suffling. (c) The 2.5D SR network. (d) The problem of our super resolution for CT slice thickness. (e) Illustration of constructing paired data based on center (top) or boundary (bottom) of voxel.

variance from input images with different scales are composed of two resblocks with a 5x5 kernel, one for 3 mm and the other for 5 mm. The subpixel shuffling layer in our network is an operator that rearranges the elements of a $W \times (D \div c) \times c$ tensor to a tensor of shape $W \times D \times 1$ [2]. The weight parameters of all layers are shared except the preprocessing layer and the final convolution layer. Our main network has three characteristics. The first one is to train the difference between the label and the input image at low resolution in a coronal plane. The inverse function of subpixel shuffling is used to convert the label image into a low resolution image and obtain the difference from the input image. Thus, as shown in Figure 1.(b), the label image becomes a residual image, which makes optimization easier and thus improves network performance [3, 1]. The second one is the long concatenation unit in Figure 1.(a). It allows the creation of various combinations of features and reduces the effect of gradient vanishing. It can also be used as a cascade learning unit that trains by dividing non-linear mapping layers into two parts. Finally, no batch normalization layer is employed in the network. It can degrade the performance of SR, which is a regression task [4]. Formally, The proposed 2.5D input network can be expressed as $I_p^{SR} = f(I_{[p-S:p+S]}^{LR}; \theta)$ where p is the current position in the coronal plane and $S = \frac{T-1}{2}$ where T is the number of input frames. 2.5D networks allow the network to learn the spatial correlation between the coronal and sagittal planes with less computational burden.

Two useful training methods were applied. The first method is the pre-trained learning which use the pre-trained 3 mm network parameters when training the 5 mm network. The second method is the cascade learning method that gradually add new layers during training. To save training time, we used the cascade learning method based on the long concatenation units.

Pairing CT data should be done considering DICOM information of slice location, slice thickness and spacing between slice. As shown in Figure 1.(e), it is so important to generate pairs of low-resolution and high-resolution images based on the voxels' center positions. Because most CT images are reconstructed based on the voxels' center positions.

3 Materials and Results

For training and evaluation of performance of the SR network, we used CT data gathered from 100 patients. Each patient data is composed of paired CT data with the exactly same acquisition protocol except slice thickness (1, 3, 5 mm). The PSNR and SSIM were measured on the axial plane and the mean values of all slices were recorded.

The experimental results for the methods are shown in Table 1. The performance of all SR networks was better than the B-spline Bicubic, and our main network showed better performance than the winning network of NTIRE 2017 with the same number of parameters. This means that fully residual learning that uses the proposed inverse subpixel shuffling is effective. The performance of the 2.5D input network is better than the 2D input and increasing the information of the sagittal plane to 9

Table 1: The comparison results of networks according to input data shapes and learning methods. (**SD** : using Simulated low resolution data, **Bo** : using Boundary of voxel instead of center of voxel, **PT** : Pre-Trained learning, **CL** : Cascaded learning based on long concatenation units)

	SD	Bo	PT	CL	PSNR/SSIM(3mm)	PSNR/SSIM(5mm)	parameters(M)
B-Spline Bicubic					28.1971/0.7508	26.0959/0.6643	
Winner of NTIRE 2017						28.4296/0.7189	2.70
2D input						28.4869/0.7191	2.70
2.5D input (9 ch.)					30.7884/0.8067	28.9509/0.7335	2.70
	O				30.0018/0.7850	28.6133/0.7160	2.70
		O			28.8137/0.7618	26.6552/0.6678	2.70
			O		30.8579/0.8119	29.0490/0.7396	2.70
			O	O	**30.9009/0.8128**	**29.1192/0.7410**	2.70

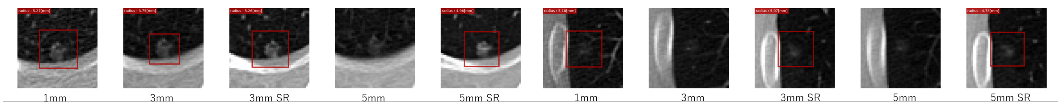

Figure 2: Performance comparison of nodule detection according to usage of proposed SR method

channels further increases the performance. To verify our arguments experimentally, simulated CT data of 3 mm or 5 mm was synthetically generated by simple summation of CT data with 1 mm slice thickness. Performance using simulation data is better than B-spline bicubic method, but lower than using real acquisition data. Pairs of low-resolution and high-resolution images generated based on the voxel boundary degrade performance significantly.

For the evaluation of performance of lung nodule detection networks which was trained in CT images with thin slice thickness (<2.5mm), we used 100 CT data which have 100 biopsy-proven lung nodules consisting of 46 GGN, 54 solid nodules, respectively.

As shown in Figure 2, in comparison with the nodule detection performance using B-Spline Bicubic as preprocessing method, we confirmed that the proposed SR had improved detection performance at 3mm and 5mm slice thickness. In case of 3mm slice thickness images, 2 solid nodules were missed (95.7% recall) while their corresponding super resolved 1mm images improved the recall to 97.8% (1 missed solid nodule). Recall of images with 5mm slice thickness was 89.1% and 85.2% while their super resolved 1mm images improved the recall to 100.0% and 96.3% for solid nodules and GGNs respectively.

4 Conclusion

Through experiments on architectural design and model training setups, we proposed an efficient 2.5D SR network for the reduction of CT slice thickness with real CT pairs. Furthermore, we demonstrated the proposed method have clinical impact by improving the performance of CAD systems. Further validation remains as future work on CT scans with various vendors and anatomical parts.

References

[1] Woong Bae, Jaejun Yoo, and Jong Chul Ye. Beyond deep residual learning for image restoration: Persistent homology-guided manifold simplification. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, pages 145–153, 2017.

[2] Jose Caballero, Christian Ledig, Andrew Aitken, Alejandro Acosta, Johannes Totz, Zehan Wang, and Wenzhe Shi. Real-time video super-resolution with spatio-temporal networks and motion compensation. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017.

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

[4] Bee Lim, Sanghyun Son, Heewon Kim, Seungjun Nah, and Kyoung Mu Lee. Enhanced deep residual networks for single image super-resolution. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, volume 1, page 3, 2017.