PAPER

Computed tomography super-resolution using deep convolutional neural network

View the article online for updates and enhancements.

Physics in Medicine & Biology

IPEM Institute of Physics and Engineering in Medicine

PAPER

Computed tomography super-resolution using deep convolutional neural network

Junyoung Park[1,2], Donghwi Hwang[1,2], Kyeong Yun Kim[1,2], Seung Kwan Kang[1,2], Yu Kyeong Kim[3] and Jae Sung Lee[1,2,4,5]

[1] Department of Biomedical Sciences, College of Medicine, Seoul National University, Seoul 03080, Republic of Korea
[2] Department of Nuclear Medicine, College of Medicine, Seoul National University, Seoul 03080, Republic of Korea
[3] Department of Nuclear Medicine, College of Medicine, Seoul National University Boramae Medical Center, Seoul 07061, Republic of Korea
[4] Institute of Radiation Medicine, Medical Research Center, Seoul National University College of Medicine, Seoul 03080, Republic of Korea
[5] Department of Nuclear Medicine, Seoul National University, College of Medicine, 103 Daehak-ro, Jongno-gu, Seoul 03080, Republic of Korea

E-mail: jaes@snu.ac.kr

Abstract

The objective of this study is to develop a convolutional neural network (CNN) for computed tomography (CT) image super-resolution. The network learns an end-to-end mapping between low (thick-slice thickness) and high (thin-slice thickness) resolution images using the modified U-Net. To verify the proposed method, we train and test the CNN using axially averaged data of existing thin-slice CT images as input and their middle slice as the label. Fifty-two CT studies are used as the CNN training set, and 13 CT studies are used as the test set. We perform five-fold cross-validation to confirm the performance consistency. Because all input and output images are used in two-dimensional slice format, the total number of slices for training the CNN is 7670. We assess the performance of the proposed method with respect to the resolution and contrast, as well as the noise properties. The CNN generates output images that are virtually equivalent to the ground truth. The most remarkable image-recovery improvement by the CNN is deblurring of boundaries of bone structures and air cavities. The CNN output yields an approximately 10% higher peak signal-to-noise ratio and lower normalized root mean square error than the input (thicker slices). The CNN output noise level is lower than the ground truth and equivalent to the iterative image reconstruction result. The proposed deep learning method is useful for both super-resolution and de-noising.

1. Introduction

The x-ray computed tomography (CT) system has been improved via several techniques, such as multi-slice detectors, spiral and fast gantry rotation, and fast computing via a graphics processing unit (GPU). Real-time reconstruction, a wider Z-coverage, helical scanning, and motion-dependent scanning, such as a cardiac CT, can be performed by those techniques. However, the immense amount of acquired data, which requires more computing power, is accompanied by a longer scan range and slower operating speed. Because of this limitation, which requires the high-throughput of acquired data, most commercial CT systems offer a preview reconstruction using small angle views with a thicker slice thickness (low resolution) during actual reconstruction, which is running in the background (Frey *et al* 2009). In this paper, we propose a method to enhance the resolution and contrast of CT images with thick slices.

In CT imaging, many technical parameters and components impact the image quality and radiation dose (McNitt-Gray 2003). The reconstructed slice thickness of an image relates to not only the Z-axis resolution, but also the noise, which is influenced by the x-ray tube current. Enhancing Z-axis resolution by reducing the slice thickness results in a tradeoff with increased noise and possibly an increased dose. To compensate for increased noise, we may increase the x-ray tube current to return the noise level to being equivalent to the original one

(McNitt-Gray 2003). The proposed approach enhances the CT image resolution similar to that of a thin slice without the increased cost in terms of dose and noise level.

Recently, convolutional neural networks (CNNs), a class of deep, feed-forward artificial neural networks, have outperformed conventional approaches in many computer vision applications, such as super-resolution (Dong *et al* 2014) and de-noising (Chen *et al* 2015) with a large dataset and high computing power GPU (Krizhevsky *et al* 2012). Super-resolution is a software technique of generating a high-resolution image from low-resolution images (Park *et al* 2003). Nowadays, many CNN-based super-resolution approaches have been proposed (Dong *et al* 2016, Kim *et al* 2016) in the field of computer vision. However, in biomedical image processing, most of those approaches have some drawbacks. First, the network needs to run for each patch, and there are lots of redundancy because of overlapping patches. This redundancy leads to slow computation. Second, there is a trade-off between localization accuracy and the use of context. More max-pooling layers, which reduce the localization accuracy, are required for larger patches. On the other hand, small patches allow the network to see the only little context. To address these problems, U-Net (Ronneberger *et al* 2015) provided an end-to-end mapping architecture consisting of a contracting path to capture the context and a symmetric expanding path that enables precise localization with a large receptive field.

The main purpose of the present study was to investigate whether the CNN is also potentially useful for CT image super-resolution. For this purpose, we designed a network that learns an end-to-end mapping between low (thick slice) and high (thin slice) resolution images using the modified U-Net shown in figure 1. To verify the proposed method, we trained and tested the CNN using axially averaged data of existing thin-slice CT images as input (thick slice) and their middle slice as the label (thin slice). Because we used average data as input, we expected an enhanced signal-to-noise ratio in the CNN output to be relative to the thin slice. Accordingly, we assessed the performance of the proposed method with respect to not only the resolution and contrast, but also the noise properties.

2. Materials and methods

2.1. Data set

Brain CT images from 65 clinical positron emission tomography (PET)/CT studies (26 males and 39 females, age $= 67.5 \pm 9.2$ years) with suspected Parkinson's disease were used. The PET/CT data were acquired using a Siemens Biograph mCT 40 scanner (Siemens Healthcare, Knoxville, TN). The retrospective use of the scan data and waiver of consent were approved by the Institutional Review Board of our institute. The head of each participant was positioned in a head holder attached to the patient bed, and the PET/CT scan followed the routine clinical protocol for brain studies. The CT images were reconstructed in a $512 \times 512 \times (60\text{–}80)$ matrix with voxel sizes of $0.59 \times 0.59 \times 3.0$ mm using multi slice spiral image reconstruction method named SureView by Siemens.

As a training dataset, five slices of a 3 mm slice thickness were averaged into one slice of a 15 mm slice thickness. Among those five slices, the middle slice was taken as the ground truth dataset, which represented the high-resolution image (figure 1). To generate as many data as possible, the averaged slices were generated using a simple moving average method.

A total of 52 patient data were used as the training set for the CNN; the other 13 patient data were employed as the test set. We performed five-fold cross-validation to confirm the performance consistency. For the CNN training and testing, the low-resolution images (thick slices) were used as the input, and the high-resolution images (thin slices) were used as the output, as previously mentioned. All input and output images were used in two-dimensional (2D) slice format. The total number of slices for training the CNN was 7670.

2.2. Network architecture

We implemented the proposed super-resolution method based on a deep CNN. The network directly learns an end-to-end mapping between thinner- and thicker-slice images (3.0 mm versus 15.0 mm). The proposed network architecture consists of a contracting path to capture the context and a symmetric expanding path that enables precise localization, as shown in figure 1. Each path is exploited by the five sequential layers. The contracting path follows the typical CNN architecture. Each layer of the path consists of repeated two 3×3 convolutions, each followed by a rectified linear unit (ReLU) and batch normalization (BN). Additionally, a 2×2 max pooling operation with stride 2 is followed for down-sampling. The number of feature maps is doubled in each down-sampling step, as shown in figure 2. Note that we draw only four feature maps at the first layer to enable the simplified drawing, and the actual number of feature maps is 32. The 32 feature maps become 512 as a result of the doubling per layer at the contracting path. Furthermore, mini-batches with the size of six images are used to achieve a relative speedy convergence (Ge *et al* 2015).

The expansion path also consists of repeated two 3×3 convolutions, each followed by the ReLU and BN, which is the same as in the contracting path (figure 3). However, 2×2 up-convolution is used for up-sampling

Figure 1. CNN architecture with data flow.

Figure 2. Detail of the contraction path.
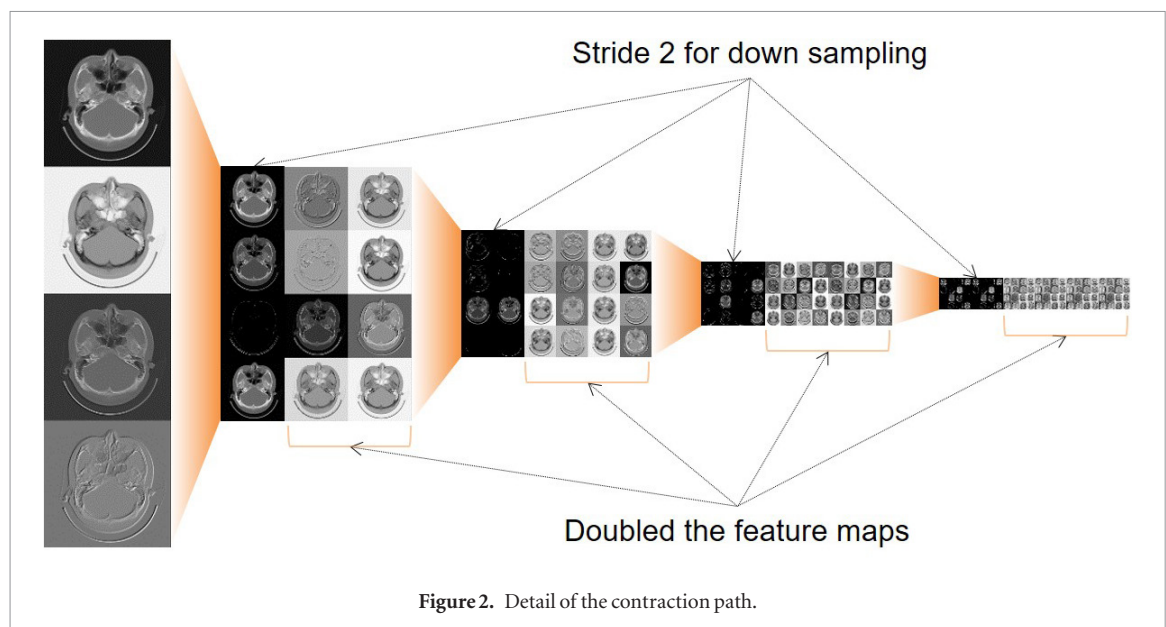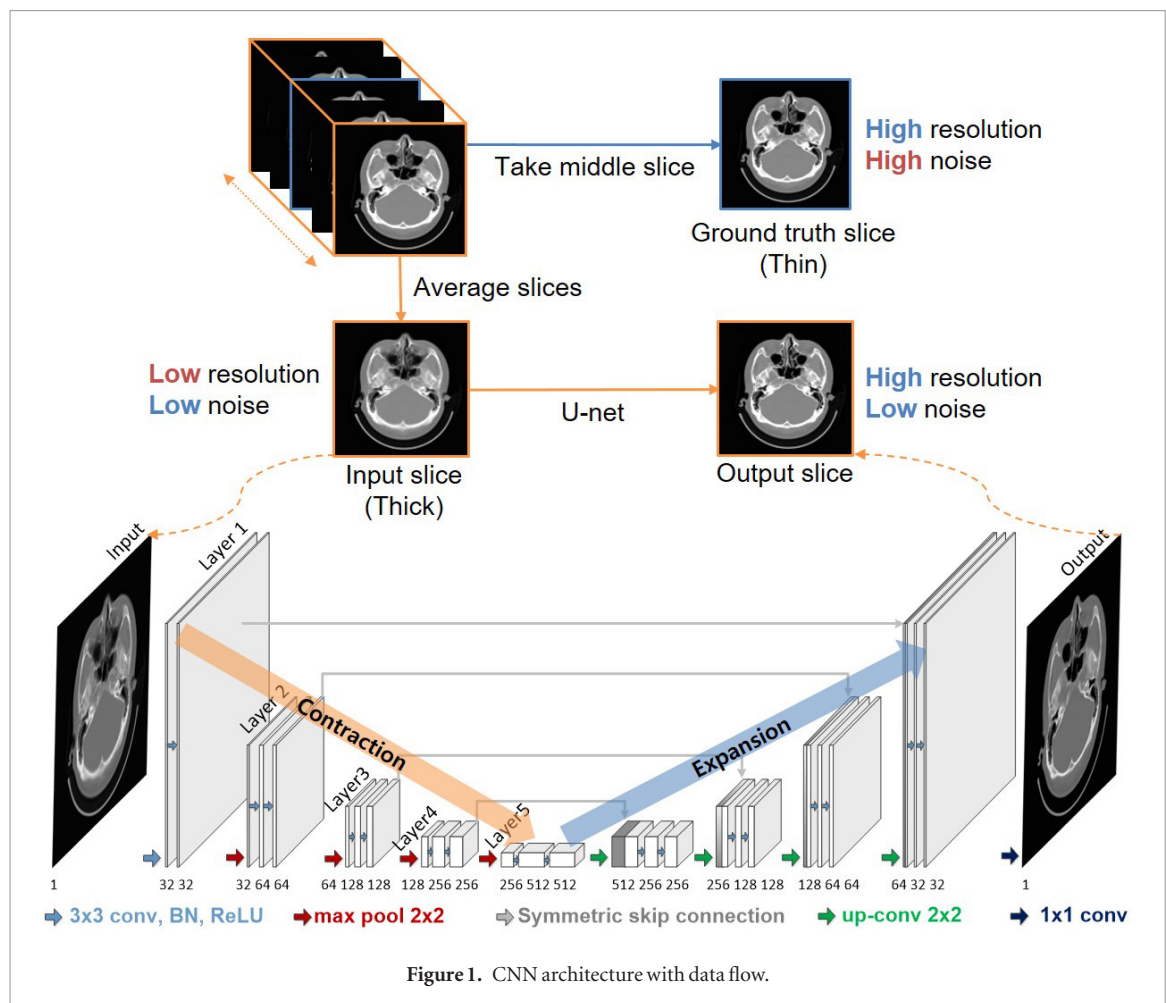
with stride 2. The 512 feature maps become 32 by reducing by half the number of feature maps in every layer. The expansive path is symmetric to the contracting path; thus, the U-shaped architecture is apparent. In this up-sampling part, we also have a large number of feature maps, which allow the network to propagate context information to higher resolution layers. The last 1×1 convolution layer performs the scaling.

We implemented the networks using the TensorFlow framework (version 1.3.0) (Abadi *et al* 2016). We employed symmetric skip connections (copy and concatenation), as shown in figure 1, between the convolutional and deconvolutional layers, thereby enabling the training to converge fast and attaining a high frequency of local features (Mao *et al* 2016).

Figure 3. Detail of the expansion path.
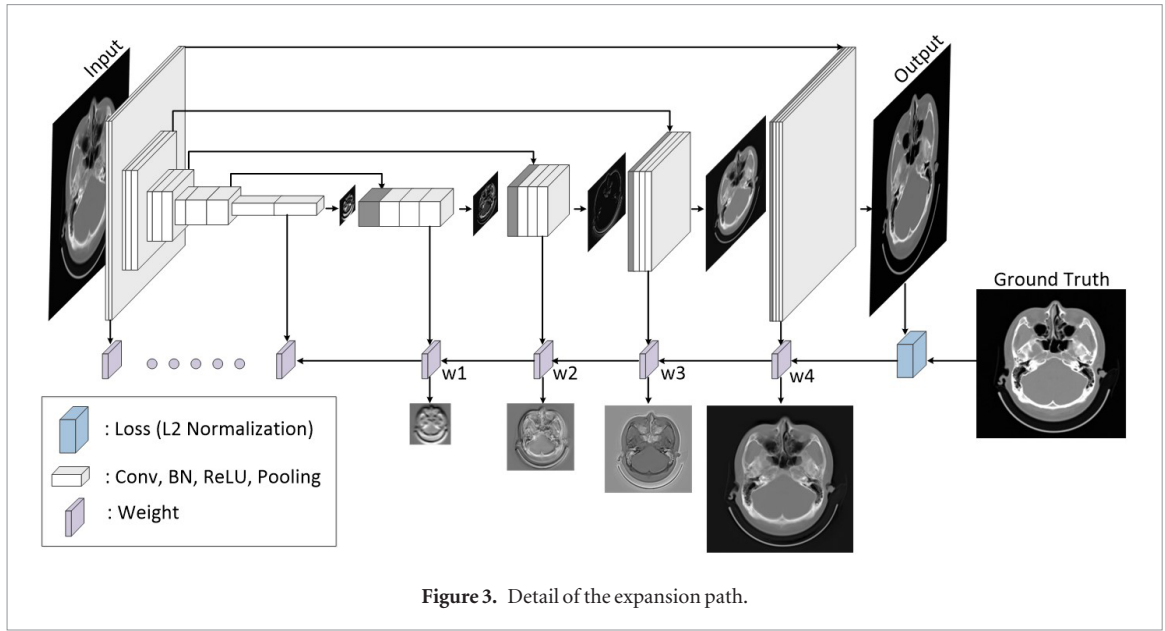
2.3. Data training and loss function

Learning the end-to-end mapping function F required the estimation of network variables, which are weight and bias (W). An initial weight is random value from a truncated normal distribution. We minimized loss function $L(v)$, which is the L^2 norm of the difference between reconstructed image $F(Y_i)$ from low-resolution image Y_i and high-resolution label image X_i. Activation function $A(x)$ was ReLU.

$$L(v) = \frac{1}{n}\sum_{i=0}^{n} \|F(Y_i; v) - X_i\|^2, \quad v = \{W_1, W_2 \ldots W_m\}$$

$$A(x) = \begin{cases} 0 \text{ for } & x < 0 \\ x \text{ for } & x \geqslant 0 \end{cases}$$

where n is the number of images in a mini-batch and m is the number of layers.

The weight factor (or kernel), W, for each layer was updated using the error back-propagation with Adam (adaptive moment estimation) optimizer, which is a stochastic optimization technique (Kingma *et al* 2014):

$$\Delta_{t+1} = \Delta_t - \frac{\eta}{\sqrt{\widehat{v_t}} + \epsilon}\widehat{m_t}, \quad \eta = \frac{\tau\sqrt{1 - \beta_2^t}}{1 - \beta_1^t}, \quad \widehat{m_t} = \frac{m_t}{1 - \beta_1^t}, \widehat{v_t} = \frac{v_t}{1 - \beta_2^t}.$$

Here, t is iteration number, Δ_t denotes the updated parameters, such as weight and bias at the tth iteration, and τ represents the learning rate. The first- and second-moment vector m_t and v_t are initialized to zero and β_1 and β_2 are respectively 0.9 and 0.999 with epsilon as 10^{-8}. The learning rate for determining to what extent the newly acquired information overrides the old information was initially 0.001 and exponentially decayed by half every two epochs. The number of epochs was 18. The training time was approximately 60 min/epoch when using one GTX 1080 TI GPU and i7-7700K CPU (3.40 GHz).

2.4. Image analysis

For the quantitative evaluation of the CNN output for the training set, 1515–1564 CT slices from 13 patient data were used for each five-fold cross-validation study among a total of 65 patient data sets. The image quality metrics, peak signal-to-noise ratio (PSNR) and normalized root mean square error (NRMSE) values were calculated.

$$PSNR = 10 \cdot \log_{10}\frac{MAX_I^2}{(\sum_{ij}^{PQ}[X(i,j) - Y(i,j)]^2)/N}$$

$$NRMSE = \frac{1}{MAX_I - MIN_I}\sqrt{\frac{\sum_{ij}^{PQ}[X(i,j) - Y(i,j)]^2}{N}}$$

where $X(i,j)$ is the reference high-resolution image with the matrix size of $P \times Q (=N)$ and $Y(i,j)$ is the output from the network. Additionally, MAX_I and MIN_I are the maximum and minimum intensities of the reference image.

For the comparison, 3D Richardson–Lucy deblurring (3D RL deblurring) algorithm (Al-Ameen *et al* 2012a, 2012b) was applied to the input images. The kernel size of the point spread function (PSF) was $3 \times 3 \times 3$ pixels
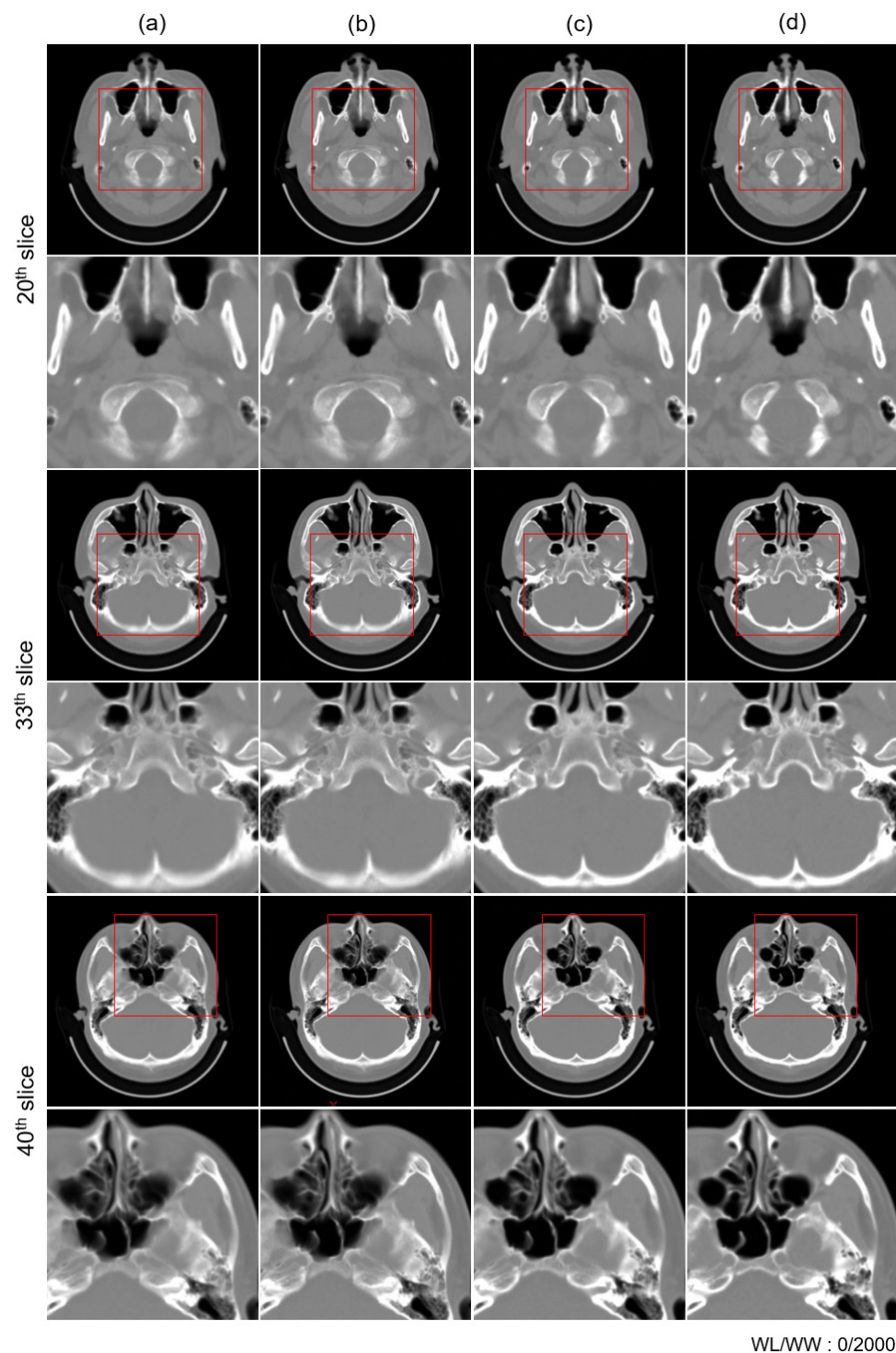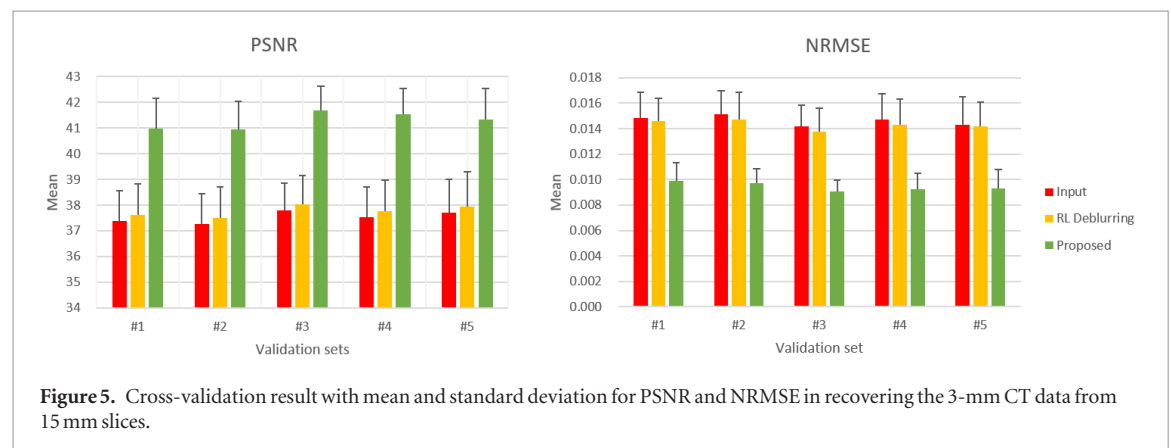
Figure 4. De-blurred CT images obtained with a deep CNN for recovering 3 mm slices from 15 mm ones. Multiple sections from (a) input image (15 mm slice), (b) input image focused by applying RL deblurring, (c) CNN output, and (d) ground truth (3 mm slice). Odd rows show CT slices obtained at different axial positions; even rows show magnified images from the red boxes on the odd rows.

($1.46 \times 1.46 \times 1.46$ mm) and sigma was one, which yielded highest PSRN and lowest NRMSE when we tested with the different kernel sizes from 0.5 pixels to 3.0 in the step of 0.5. Those parameters resulted in a better resolution than the input image with an approximately 0.5% lower PSNR; however, the larger kernel size and sigma yielded output that was too noisy.

The PSNR and NRMSE values of the original input, RL deblurred input, and network output were compared. To assess the noise characteristics of soft tissue, the region-of-interest (ROI)-based standard deviation was compared.

Table 1. PSNR and NRMSE (average across all axial positions) relative to the ground truth for 13 subjects employed in the test group of the third cross-validation.

Test group	Patient	Input		RL deblurring		Proposed method	
		PSNR	NRMSE	PSNR	NRMSE	PSNR	NRMSE
#3	1	36.34	0.017	36.52	0.017	40.78	0.010
	2	36.32	0.016	36.46	0.016	40.33	0.010
	3	36.73	0.016	36.91	0.016	40.82	0.010
	4	37.24	0.015	37.43	0.015	41.85	0.009
	5	39.35	0.012	39.64	0.011	42.79	0.008
	6	38.55	0.013	38.80	0.012	41.73	0.009
	7	37.68	0.014	37.96	0.013	41.28	0.009
	8	39.13	0.012	39.41	0.012	42.81	0.008
	9	37.10	0.015	37.35	0.014	41.07	0.010
	10	38.28	0.013	38.55	0.013	41.71	0.009
	11	37.43	0.015	37.66	0.014	40.90	0.010
	12	37.77	0.014	38.08	0.014	42.47	0.008
	13	39.36	0.012	39.67	0.012	43.36	0.008
	Mean	37.79	0.014	38.03	0.014	41.68	0.009
	SD	1.073	0.002	1.119	0.002	0.935	0.001

Figure 5. Cross-validation result with mean and standard deviation for PSNR and NRMSE in recovering the 3-mm CT data from 15 mm slices.

2.5. Additional verification for clinical practice

In clinical practice, it might be more interesting if one can recover a 1 mm slice from a 5 mm slice as the dose reduction would be quite significant when replacing a scanning protocol for a 1 mm slice with a 5 mm protocol. Thus, we also explored the possibility of applying the proposed super-resolution technique for recovering the 1 mm, high-resolution CT data from 5 mm slices. The CNN was also trained with 56 patient data set (9302 slices) and tested with 14 data set (2338 slices). The CT images were reconstructed in a $512 \times 512 \times (180–200)$ matrix with voxel sizes of $0.59 \times 0.59 \times 1.0$ mm. The additional verification was also performed through the ROI-based standard deviation comparison and the five-fold cross-validation as we have done in recovering the 3 mm CT data from 15 mm slices.

3. Results

3.1. Recovery from 3 mm slices from 15 mm ones

The CNN generated output images that were virtually equivalent to the ground truth, as observed in figure 4. From left to right, the (a) input image (15 mm slice), (b) input image focused by applying RL de-blurring, (c) CNN output, and (d) ground truth (3 mm slice) are presented. Three slices with complex anatomical structures were selected. Odd rows show CT slices obtained at different axial positions; even rows are magnified images from the red boxes on the odd rows. The most remarkable improvement in the image recovery by the CNN is the deblurring of boundaries of bone structures and air cavities. The better deblurring performance relative to the RL deblurring is distinct on the occipital bone around the cerebellum shown in the 33rd slice and the boundary of the maxillary sinus in the 40th slice. The cortical bone of the occipital condyle in the 20th slices and internal occipital crest and cerebellar fossa as well as external occipital protuberance shown in the 33th slices are better recognized in the ground truth image than input image. The medial/lateral pterygoid plate in the 20th slices, thin
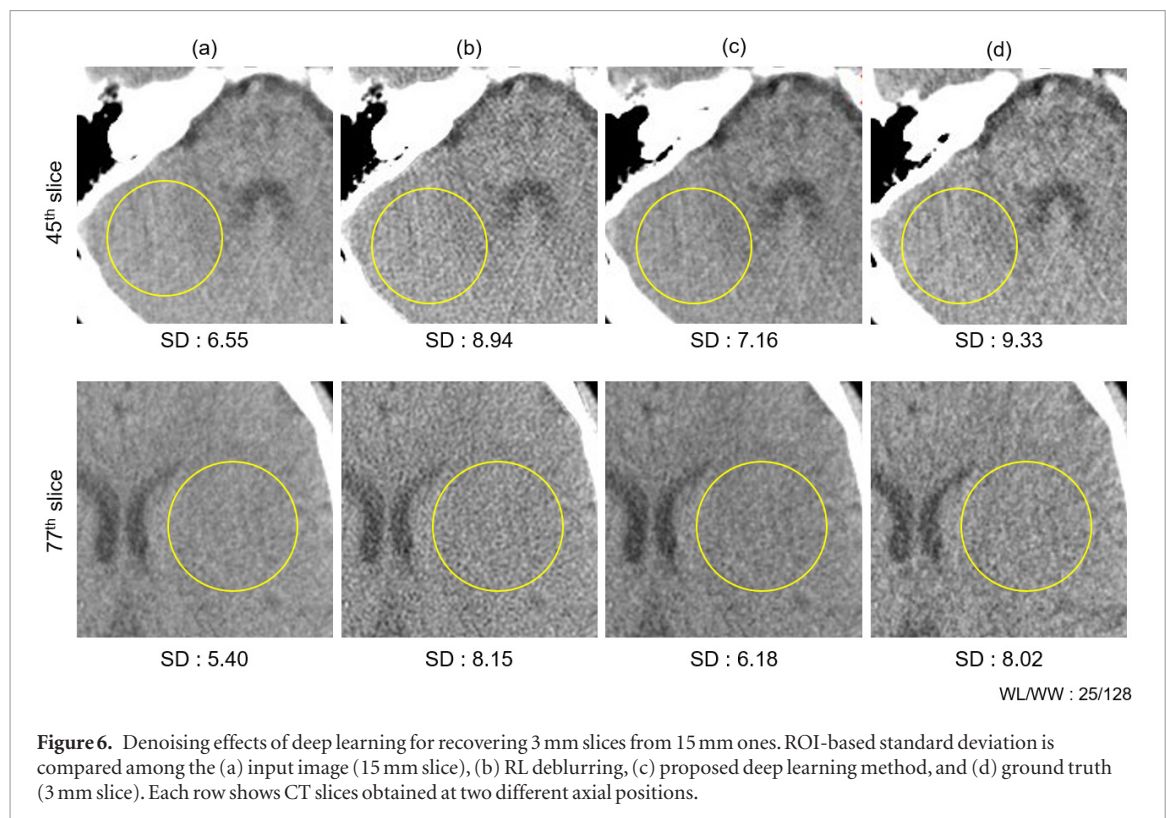
Figure 6. Denoising effects of deep learning for recovering 3 mm slices from 15 mm ones. ROI-based standard deviation is compared among the (a) input image (15 mm slice), (b) RL deblurring, (c) proposed deep learning method, and (d) ground truth (3 mm slice). Each row shows CT slices obtained at two different axial positions.

Table 2. PSNR and NRMSE (average across all axial positions) relative to the ground truth in recovering the 1 mm CT data from 5 mm slices.

Test group	Patient	Input		RL deblurring		Proposed method	
		PSNR	NRMSE	PSNR	NRMSE	PSNR	NRMSE
#3	1	35.38	0.019	35.50	0.018	39.40	0.012
	2	34.83	0.019	34.95	0.019	38.89	0.012
	3	38.13	0.014	38.17	0.014	41.33	0.010
	4	37.54	0.015	37.72	0.015	40.06	0.011
	5	35.96	0.017	36.07	0.017	39.99	0.010
	4	35.96	0.017	36.05	0.017	39.36	0.011
	7	36.11	0.018	36.27	0.017	40.06	0.011
	8	35.53	0.018	35.65	0.018	39.50	0.011
	9	36.83	0.017	36.88	0.016	40.32	0.011
	10	37.60	0.016	37.68	0.016	41.24	0.010
	11	36.21	0.017	36.34	0.017	39.90	0.011
	12	35.83	0.018	36.00	0.018	39.60	0.011
	12	36.10	0.018	36.27	0.018	39.96	0.011
	13	36.36	0.018	36.44	0.017	40.30	0.010
	Mean	36.31	0.017	36.43	0.017	39.99	0.011
	SD	0.920	0.001	0.905	0.001	0.675	0.001

septal structure of ethmoid sinus and sphenoid sinus are well identified in contrast of air cells in the 33th and 40th slices. The computation time of proposed CNN operation was only 30 ms per slice.

The quantitative analysis of PSNR and NRMSE confirmed these findings in the visual observations. Table 1 shows the PSNR and NRMSE (average across all axial positions) relative to the ground truth for 13 subjects employed (total 1478 slices) in the test group of the third cross-validation. The CNN output yielded approximately 10% higher PSNR and lower NRMSE values than the input (thicker slices).

Applying RL deblurring (figure 4(b)) yielded a slightly better PSNR than the input image with enhanced the image sharpness. These findings are consistent through all the cross-validations, as summarized in figure 5. In all the cross-validations, the CNN yielded a PSNR higher than 40 and an NRMSE smaller than 10%. The noise level of the CNN output was lower than the ground truth. Figure 6 shows the ROI-based standard deviation of a relatively uniform region of a representative case. On the contrary, the RL deblurring increased the noise level.
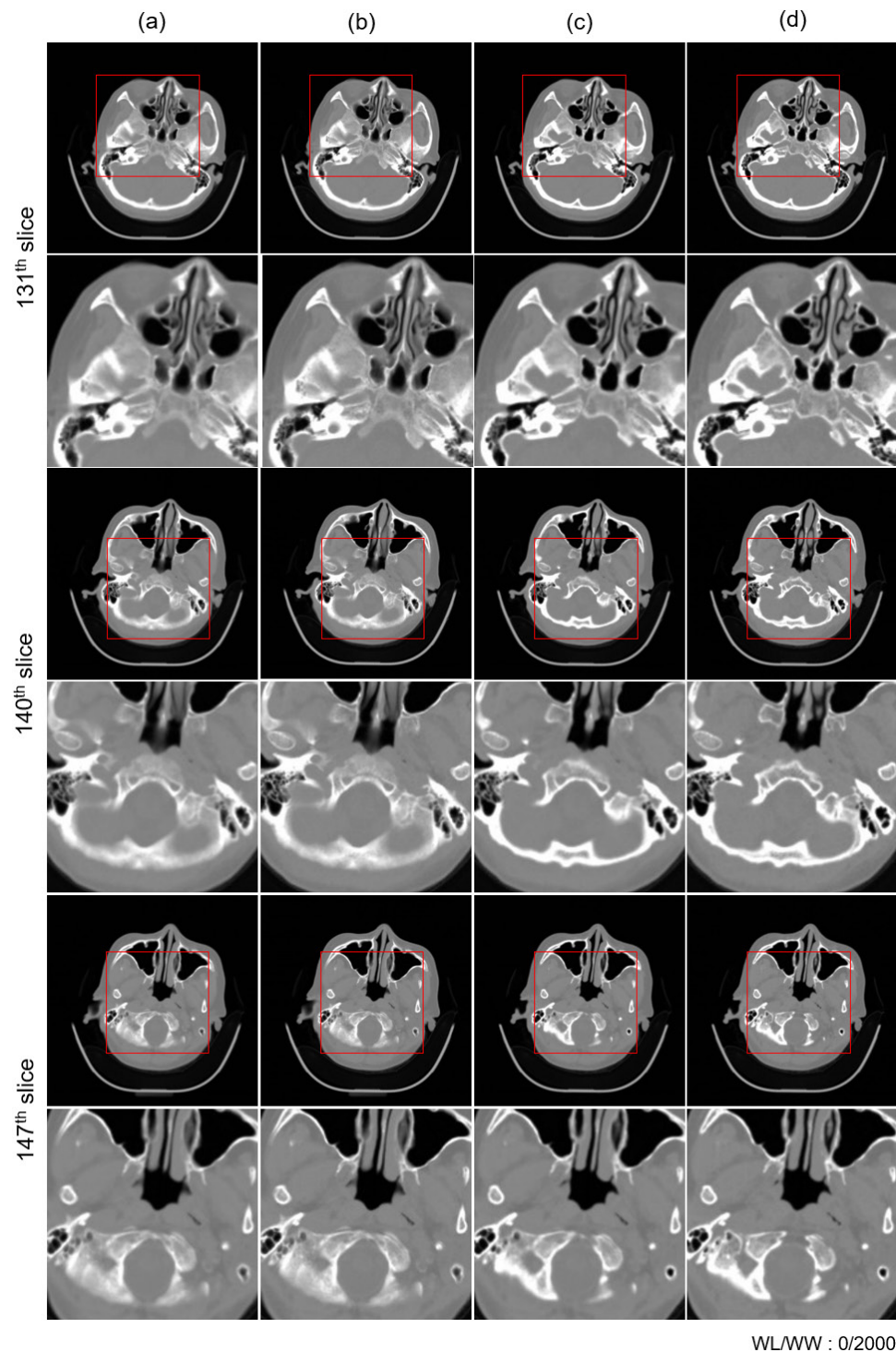
Figure 7. De-blurred CT images obtained with a deep CNN for recovering 1 mm slices from 5 mm ones. Multiple sections from (a) input image (5 mm slice), (b) input image focused by applying RL deblurring, (c) CNN output, and (d) ground truth (1 mm slice). Odd rows show CT slices obtained at different axial positions; even rows show magnified images from the red boxes on the odd rows.

3.2. Recovery from 1 mm slices from 5 mm ones

Figures 7–9 and table 2 shows the outcomes of additional verification for clinical practice in which we explored the feasibility of proposed method for recovering the 1 mm, high-resolution CT data from 5 mm slices. These results were all consistent with those shown previously in the recovery the 3 mm CT data from 15 mm slices. The additional verification also showed consistent performance through all the cross-validations, as summarized in figure 8. Figure 9 shows that the denoising effect exists consistently.

Figure 8. Cross-validation result with mean and standard deviation for PSNR and NRMSE in recovering the 1 mm CT data from 5 mm slices.
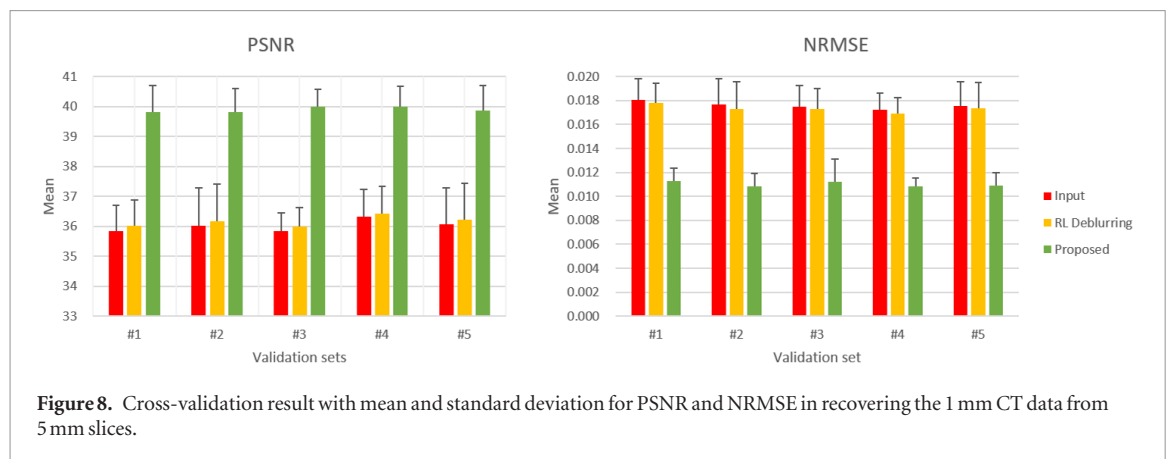
4. Discussion

In various biomedical imaging and engineering fields, machine learning has been widely used (Mohammadi *et al* 2016, Choi 2017, Yoo 2017). Recently, deep learning has gained considerable interest and is now actively investigated for medical imaging. In some physical artifact-reduction problems in x-ray CT and other medical image modalities, deep learning outperformed conventional mathematical and statistical approaches (Hong *et al* 2017, Hwang *et al* 2018, Kang *et al* 2018). Furthermore, deep learning has shown significant potential for computer-aided diagnosis and fully automatic organ segmentation (Cha *et al* 2016). In this study, we showed the potential of deep-learning-based super-resolution of x-ray CT images. A single image super-resolution technology based on CNN was applied to a thick CT slice having a higher SNR than a thin slice on account of the higher number of x-ray quanta per voxel. The CNN-generated output was almost indistinguishable from the ground truth thin slice both visually and numerically.

In this preliminary study, we did not systematically analyze the impact of super-resolution on lesion detectability because the CT images employed in the study were non-contrast-enhanced low-dose CT for PET attenuation correction and anatomical localization. Further investigations are thus warranted to determine whether this proposed approach will also be valid for diagnostic purposes. The feasibility of this approach for other medical imaging modalities, such as magnetic resonance imaging and PET, is also of interest. Although we cannot yet contend that the proposed method is clinically feasible on account of the above reasons, this method will nevertheless be useful at the minimum for previewing CT image reconstruction. Preview reconstruction (real-time reconstruction) employs fewer views owing to limited data throughput while the main reconstruction runs in the background. Because of insufficient data as well as the skipping of all post-processing steps, the output image of preview reconstruction is noisy and has poor spatial resolution. Therefore, we usually employ a thick slice thickness, of which the image resolution and contrast can be improved by the deep learning approach proposed herein. Because preview images were not saved in the Siemens mCT scanner used in this study, it was not possible to conduct training between the preview and main reconstruction images.

Thick CT slices may, at times, be preferable in reading CT images with a low soft-tissue contrast and high noise level. Moreover, under fewer view acquisitions with high-pitch helical scanning, a thick slice reduces the artifacts due to insufficient angular sampling and the chance of missing small lesions. The proposed method will be effective in these cases as well. However, we should note some limitations of the proposed method in these potential applications. Most of all, one potential issue of the proposed method is that the resolution recovery is done using 2D CNN. Due to the lack of the 3D information, a small lesion in any 3 mm slice would appear to be identical in the averaged, 15 mm slice. This creates an issue that the averaged information would not vary even if the actual location of the small lesion changed within the 15 mm slice. In addition, the CNN based super-resolution may also expect to reproduce small structure (common anatomical features among subjects) which is easily buried and invisible in low-resolution images. We, however, could not observe such cases in this study.

The spatial resolution of the CT image is mainly determined by the size of the detector elements, as defined by the detector spacing and detector pitch (Greenspan 2008). Although we can improve the spatial resolution by reducing the detector width and inter-detector distance, development of new detectors is time-intensive and incurs a high cost. In addition, smaller detector elements lead to an increase of noise. The image resolution can be improved by image processing methods, such as image sharpening by a high-pass filter (Dogra *et al* 2014), and image deblurring by a Laplacian filter and the Richardson–Lucy algorithm (Al-Ameen *et al* 2012a, 2012b). However, these kinds of filters and deblurring methods increase the noise level and are not as effective as the proposed deep learning approach, as shown in figures 6 and 9.
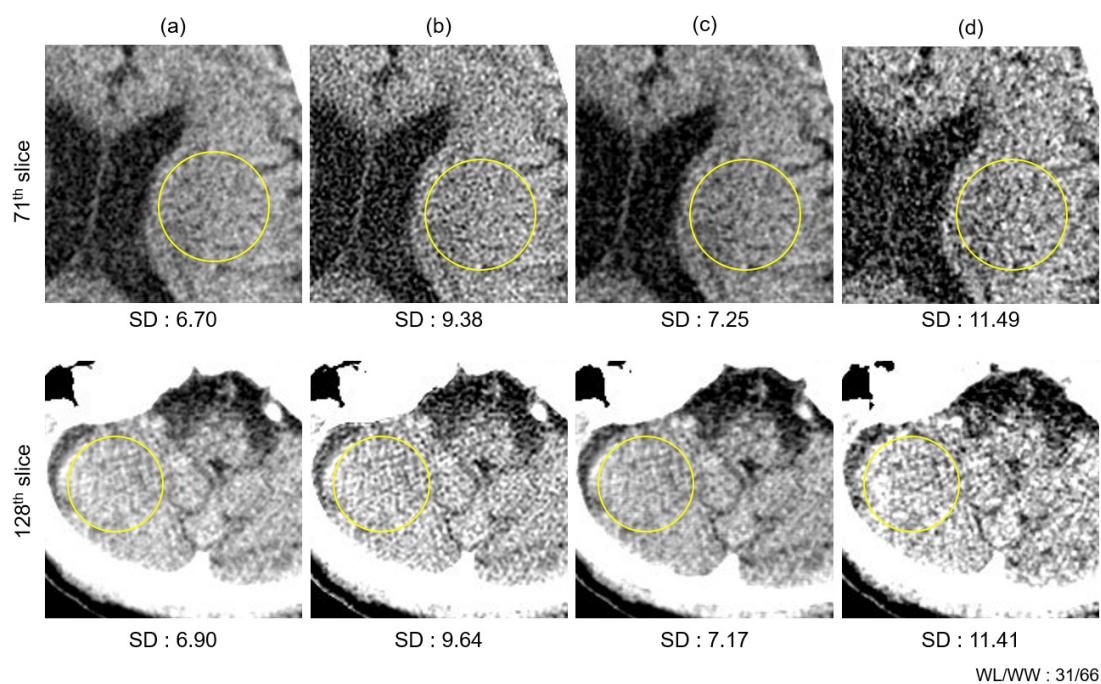
Figure 9. Denoising effects of deep learning for recovering 1 mm slices from 5 mm ones. ROI-based standard deviation is compared among the (a) input image (5 mm slice), (b) RL deblurring, (c) proposed deep learning method, and (d) ground truth (1 mm slice). Each row shows CT slices obtained at two different axial positions.
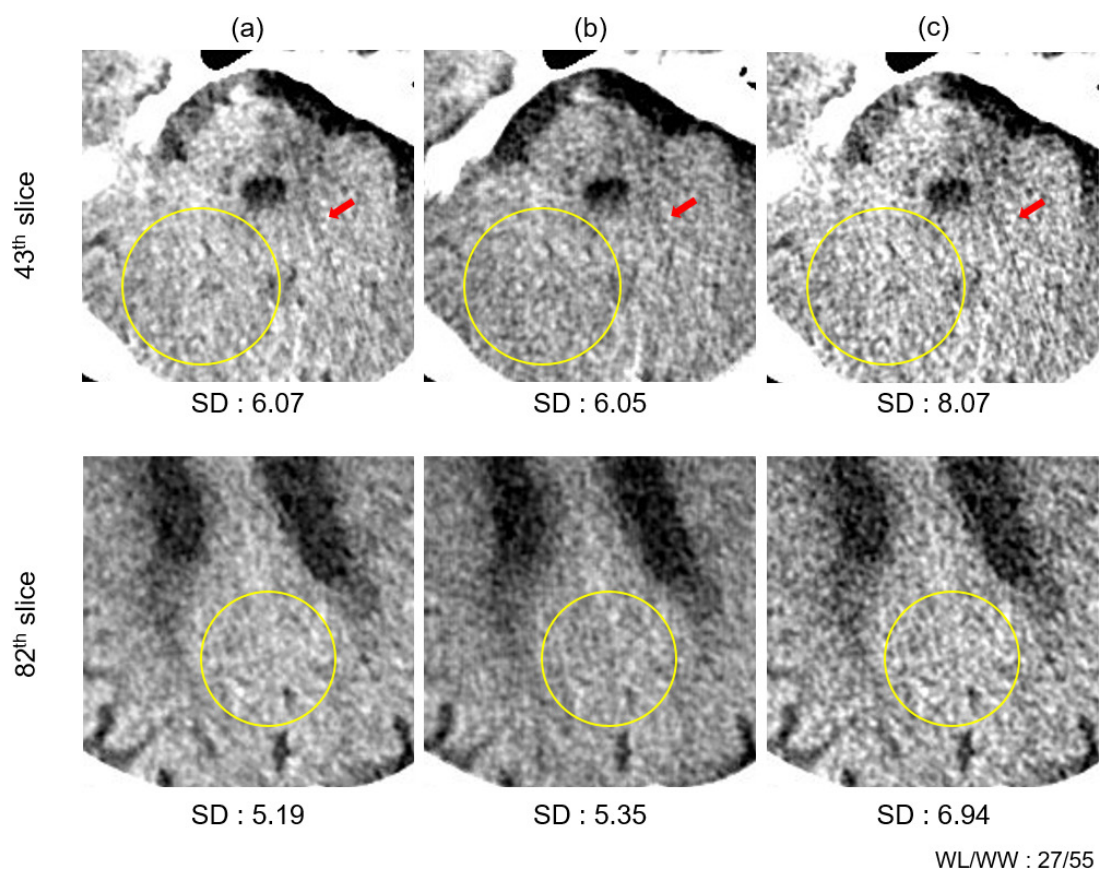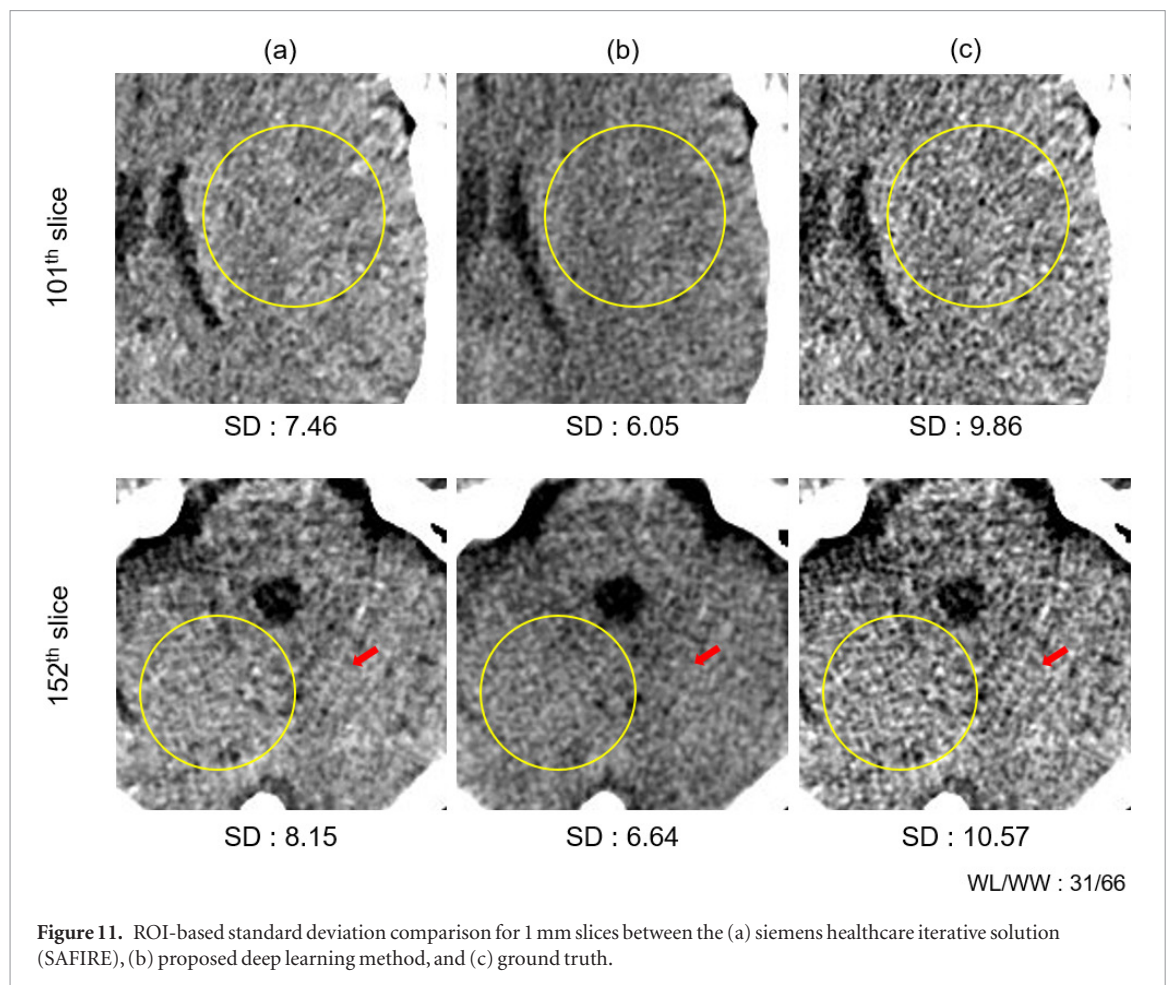

Figure 10. ROI-based standard deviation comparison for 3 mm slices between the (a) siemens healthcare iterative solution (SAFIRE), (b) proposed deep learning method, and (c) ground truth.

Figure 11. ROI-based standard deviation comparison for 1 mm slices between the (a) siemens healthcare iterative solution (SAFIRE), (b) proposed deep learning method, and (c) ground truth.

Some recent publications also used CNN-based super-resolution methods for improving mammography, angiography, CT, and MR images, etc (Oktay *et al* 2016, Yang *et al* 2016, Pham *et al* 2017a, 2017b, Umehara *et al* 2017a, 2017b, Zhang *et al* 2017). These related publications proposed methods for the recovery of original images from the under-sampled smaller-sized images with larger pixel size, and showed that the CNN-based super-resolution outperformed conventional interpolation method used in medical image magnification. However, in this study, we have focused on the image resolution and contrast recovery in the degraded images due to axial average required in real-time CT preview reconstruction. Our approach can be regarded as a kind of deconvolution operation for axial spatial resolution recovery. The results of these others' and our approaches indicate that we can use the CNN-based super-resolution techniques for improving the medical image quality and reducing scan time and processing burden in several different ways.

Noise reduction in a CT is most important for reducing the radiation dose and improving the diagnostic power. Statistical iterative image reconstruction and model-based image reconstruction are currently used algorithm-based approaches to reducing noise in CT images (Beister *et al* 2012). However, they require raw projection data and many computations, which result in a slow image reconstruction speed. To address this issue, an image-based nonlinear filter (Keselbrener *et al* 1992) and sinogram-based nonlinear smoothing (Li *et al* 2004) can also be applied. However, each of these methods produce low resolution and require projection data. In contrast, our proposed method applied to reconstructed image is effective in noise reduction without causing image blurring. As shown in figures 10 and 11, the proposed method can be also used as a noise-reduction solution. Compared to Siemens Healthcare iterative solution, named SAFIRE, our method yielded a similar noise level in the ROI-based standard deviation analysis. The streak artifact shown in the ground truth image (red arrow) was also reduced.

Some recent studies (i.e. Chen *et al* (2017) and Wolterink *et al* (2017)) showed that the CNN-based noise reduction is efficient in thin-slice CT images. In these publications, the CNNs learn the transition from the noisy image to denoised one or that from the noisy image to the residual between noisy and denoised images. Although these approaches look promising for reducing image noise or radiation dose in high-resolution CT scans, a concerns is raised in these direct denoising approaches: It is uncertain what the outcome of CNN will be if the noise level of an input image is different from the training set. On the contrary, our approach would be freer from this concern because the input to our CNN is axially averaged image with more consistent noise level than single thin-slice due to the saturated statistical fluctuation.

In this paper, we used several strategies to improve the CNN performance compared to the original U-Net. Mini-batches were used to accelerate CNN learning, although it requires more GPU memory. Moreover, there was a constraint in the number of feature maps and mini-batches because of the memory limitation. We used six mini-batches with half of the feature maps—specifically, 32 feature maps, compared to the originally designed 64 maps without mini-batches. However, using 12 mini-batches with five-fold fewer feature maps (16 of them) showed worse convergence and performance. Owing to the concern of overfitting, we once attempted the dropout and biased convolutional layers. However, BN (Ioffe *et al* 2015) was more effective than those strategies (dropout and biased convolution layers) for avoiding the local minima caused by improper initialization and achieving faster learning rates. We attempted to use batch re-normalization to reduce the dependency in multiple mini-batches (Ioffe 2017); however, it could not achieve a significant boost in training speed.

5. Summary and conclusion

In this study, we propose a deep-learning-based approach for CT image super-resolution. The CNN trained in this study yielded high-resolution images (thin slice) once the low-resolution image (thick slice) was given. We optimized the U-Net architecture and training strategies, including BN, learning rate decay, and feature-map and mini-batch size fine-tuning. The extraction and expansion paths of the network with a large receptive field effectively captured the high-resolution features as high-resolution features. Although this work mainly focused on resolution improvement, the *Z*-axis averaging plus super-resolution strategy was also useful for reducing noise.

Acknowledgments

ORCID iDs

Jae Sung Lee ⬤ https://orcid.org/0000-0001-7623-053X

References

Abadi M *et al* 2016 Tensorflow: large-scale machine learning on heterogeneous distributed systems (arXiv:1603.04467)

Al-Ameen Z, Sulong G, Gapar M and Johar M 2012a Reducing the Gaussian blur artifact from CT medical images by employing a combination of sharpening filters and iterative deblurring algorithms *J. Theor. Appl. Inf. Technol.* **46** 31–6

Al-Ameen Z, Sulong G, Johar M G M, Verma N, Kumar R, Dachyar M, Alkhawlani M, Mohsen A, Singh H and Singh S 2012b A comprehensive study on fast image deblurring techniques *Inj. J. Adv. Sci. Technol.* **44** 1–10

Beister M, Kolditz D and Kalender W A 2012 Iterative reconstruction methods in x-ray CT *Phys. Med.* **28** 94–108

Cha K H, Hadjiiski L, Samala R K, Chan H P, Caoili E M and Cohan R H 2016 Urinary bladder segmentation in CT urography using deep-learning convolutional neural network and level sets *Med. Phys.* **43** 1882–96

Chen H, Zhang Y, Zhang W, Liao P, Li K, Zhou J and Wang G 2017 Low-dose CT denoising with convolutional neural network *IEEE 14th Int. Symp. on Biomedical Imaging* pp 143–6

Chen Y, Yu W and Pock T 2015 On learning optimized reaction diffusion processes for effective image restoration *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition* pp 5261–9

Choi H 2017 Deep learning in nuclear medicine and molecular imaging: current perspectives and future directions *Nucl. Med. Mol. Imaging* **52** 109–18

Dogra A and Bhalla P 2014 Image sharpening by gaussian and butterworth high pass filter *Biomed. Pharmacol. J.* **7** 707–13

Dong C, Loy C C, He K and Tang X 2014 Learning a deep convolutional network for image super-resolution *European Conf. Computer Vision* pp 184–99

Dong C, Loy C C, He K and Tang X 2016 Image super-resolution using deep convolutional networks *IEEE. Trans. Pattern. Anal.* **38** 295–307

Frey S, Müller C, Strengert M and Ertl T 2009 Concurrent CT reconstruction and visual analysis using hybrid multi-resolution raycasting in a cluster environment *Adv. Vis. Comput.* **1** 357–66

Ge R, Huang F, Jin C and Yuan Y 2015 Escaping from saddle points—online stochastic gradient for tensor decomposition *Conf. Learning Theory* pp 797–842

Greenspan H 2008 Super-resolution in medical imaging *Comput. J.* **52** 43–63

Hong X, Zan Y, Weng F, Zhao Z and Huang Q 2017 Enhancing pet image quality via single image super resolution with deep residual learning *In the Proc. of The 14th Int. Meeting on Fully Three-Dimensional Image Reconstruction in Radiology and Nuclear Medicine* pp 673–6

Hwang D, Kim K Y, Kang S K, Seo S, Paeng J C, Lee D S and Lee J S 2018 Improving accuracy of simultaneously reconstructed activity and attenuation maps using deep learning *J. Nucl. Med.* (https://doi.org/10.2967/jnumed.117.202317)

Ioffe S 2017 Batch renormalization: towards reducing minibatch dependence in batch-normalized models (arXiv:1702.03275)

Ioffe S and Szegedy C 2015 Batch normalization: accelerating deep network training by reducing internal covariate shift *Int. Conf. Machine Learning* pp 448–56

Kang S K, Seo S, Shin S A, Byun M S, Lee D Y, Kim Y K, Lee D S and Lee J S 2018 Adaptive template generation for amyloid PET using a deep learning approach *Hum. Brain Mapp.* (https://doi.org/10.1002/hbm.24210)

Keselbrener L, Shimoni Y and Akselrod S 1992 Nonlinear filters applied on computerized axial tomography: theory and phantom images *Med. Phys.* **19** 1057–64

Kim J, Lee K J and Lee K M 2016 Accurate image super-resolution using very deep convolutional networks *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition* pp 1646–54

Kingma D P and Ba J 2014 Adam: a method for stochastic optimization (arXiv:1412.6980)

Krizhevsky A, Sutskever I and Hinton G E 2012 Imagenet classification with deep convolutional neural networks *Adv. Neural Inf. Process. Syst.* **1** 1097–105

Li T, Li X, Wang J, Wen J, Lu H, Hsieh J and Liang Z 2004 Nonlinear sinogram smoothing for low-dose x-ray CT *IEEE Trans. Nucl. Sci.* **51** 2505–13

Mao X, Shen C and Yang Y-B 2016 Image restoration using very deep convolutional encoder-decoder networks with symmetric skip connections *NIPS'16 Proc. 30th Int. Conf. Neural Information Processing Systems* pp 2802–10

McNitt-Gray M 2003 Tradeoffs in CT image quality and radiation dose *Med. Phys.* **30** 1410

Mohammadi M R, Khaleghi A, Nasrabadi A M, Rafieivand S, Begol M and Zarafshan H 2016 EEG classification of ADHD and normal children using non-linear features and neural network *Biomed. Eng. Lett.* **6** 66–73

Oktay O, Bai W, Lee M, Guerrero R, Kamnitsas K, Caballero J, de Marvao A, Cook S, O'Regan D and Rueckert D 2016 Multi-input cardiac image super-resolution using convolutional neural networks *Int. Conf. on Medical Image Computing and Computer-Assisted Intervention* pp 246–54

Park S C, Park M K and Kang M G 2003 Super-resolution image reconstruction: a technical overview *IEEE. Signal Proc. Med.* **20** 21–36

Pham C-H, Ducournau A, Fablet R and Rousseau F 2017a Brain MRI super-resolution using deep 3D convolutional networks *IEEE 14th Int. Symp. on Biomedical Imaging* pp 197–200

Pham C-H, Fablet R and Rousseau F 2017b Multi-scale brain MRI super-resolution using deep 3D convolutional networks (https://hal.univ-brest.fr/hal-01635455v1)

Ronneberger O, Fischer P and Brox T 2015 U-net: convolutional networks for biomedical image segmentation *Int. Conf. on Medical Image Computing and Computer-Assisted Intervention* pp 234–41

Umehara K, Ota J and Ishida T 2017a Application of super-resolution convolutional neural network for enhancing image resolution in chest CT *J. Digit. Imaging* (https://doi.org/10.1007/s10278-017-0033-z)

Umehara K, Ota J and Ishida T 2017b Super-resolution imaging of mammograms based on the super-resolution convolutional neural network *Open J. Med. Imaging* 180

Wolterink J M, Leiner T, Viergever M A and Išgum I 2017 Generative adversarial networks for noise reduction in low-dose CT *IEEE Trans. Med. Imaging* **36** 2536–45

Yang X, Zhant S, Hu C, Liang Z and Xie D 2016 Super-resolution of medical image using representation learning *8th Int. Conf. on Wireless Communications & Signal Processing (WCSP)* pp 1–6

Yoo Y 2017 On predicting epileptic seizures from intracranial electroencephalography *Biomed. Eng. Lett.* **7** 1–5

Zhang Y and An M 2017 Deep learning-and transfer learning-based super resolution reconstruction from single medical image *J. Healthc. Eng.* **2017** 5859727