의학박사 학위논문

Improvement of Image quality in Low-Dose Computed Tomography using a Deep Learning-Based Denoising Algorithm

딥러닝을 이용한 저선량 전산화 단층촬영의 영상 화질 향상

2019 년 2 월

서울대학교 대학원

의과대학 임상의과학과

신 윤 주

A thesis of the Degree of Doctor of Philosophy

딥러닝을 이용한 저선량 전산화 단층촬영의 영상 화질 향상

Improvement of Image quality in Low-Dose Computed Tomography using a Deep Learning-Based Denoising Algorithm

February 2019

The Department of Clinical Medical Sciences,

Seoul National University

College of Medicine

Yoon Joo Shin

ABSTRACT

Introduction: To assess the image quality of low-dose (LD) computed tomography (CT) using a deep learning based denoising algorithm (DLA) compared with filtered back projection (FBP) and advanced modeled iterative reconstruction (ADMIRE).

Materials and Methods: A total of 100 patients who had undergone routine dose (RD) abdominal CT reconstructed with FBP were included to build the DLA training set. CT images at dose levels corresponding to 13%, 25%, and 50% of RD were simulated from RD CT images and reconstructed using FBP. We trained three DLAs using the simulated LD CT images with different dose levels as input data and the RD CT images as the ground truth (DLA-1, 2, 3 for 13%, 25%, and 50% dose levels, respectively). The American College of Radiology (ACR) CT phantom was used together with 18 patients who underwent abdominal LD CT to build a testing set. LD CT images of phantom and patients were reconstructed using FBP, ADMIRE, and processed using DLAs (LD-FBP, LD-ADMIRE, LD-DLA images). To compare the quality of reconstructed and

processed images, we measured noise power spectrum (NPS) and modulation transfer function (MTF) for various contrast objects in phantom images, and mean image noises in patient data. Statistical analysis was performed using paired t-tests and repeated measure analysis of variance with Bonferroni correction for pairwise comparisons. In addition, we evaluated the presence of additional artifacts in LD-DLA images.

Results: LD-DLAs achieved lower noise levels than LD-FBP and LD-ADMIRE in both phantom and patient studies (all $p < 0.001$), and LD-DLAs trained with lower radiation doses showed less image noise. There were no additional image artifacts in LD-DLA images. However, the MTFs of the LD-DLAs were significantly lower than those of LD-ADMIRE and LD-FBP (all $p < 0.001$) and decreased with decreasing training image dose, although the differences between the LD-DLAs and LD-FBP were minimal.

Conclusions: DLAs achieved less noise than FBP and ADMIRE in LD CT images, but did not maintain spatial resolution. Lower radiation doses in training images led to less noise.

Keywords: Computed tomography, Image denoising, Deep

learning, Iterative reconstruction

Student number: 2017-34436

CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ADMIRE: advanced modeled iterative reconstruction

CT: computed tomography

DLA: deep learning based denoising algorithm

FBP: filtered back projection

IR: iterative reconstruction

RD: routine dose

LD: low-dose

MSE: mean squared error

MTF: modulation transfer function

NPS: noise power spectrum

INTRODUCTION

Computed tomography (CT) is a widely used screening and diagnostic imaging tool due to its wide availability. However, CT is known to contribute significantly to the lifetime risk of radiation-related cancer, because of the increase in radiation exposure with increasing number of CT scans, particularly in patients who undergo repeated follow-up CT examinations or in younger patients [1, 2]. In this regard, various techniques have been developed to mitigate the radiation risk [3-6].

One of the common strategies used to reduce radiation dose is to adjust the current applied to the X-ray tube to reduce the number of X-ray photons emitted from the light source [7, 8]. However, this technique typically results in a reduction in image quality due to a low signal-to-noise ratio (SNR) [8]. Therefore, it is important to minimize noise and artifacts in low-dose (LD) CT, so as to preserve image quality while reducing radiation dose. The most successful recent attempt in this direction is the use of the iterative reconstruction (IR) technique in CT image reconstruction [4, 6, 8-11].

IR is a widely used method to improve the image quality of CT, but presents the disadvantage of requiring long computation times to execute the repetitive reconstruction process [9]. In addition, since the IR algorithm can treat only a few parameters, it is difficult to use the rich information provided by large-scale CT data. Thus, conventional methods do not completely solve the complex problem of image reconstruction in LD CT [9, 11, 12].

The use of deep learning techniques has become widespread in the medical imaging field due to recent advances in machine learning and improvements in hardware performance. In recent years, image denoising algorithms using artificial neural networks have been intensively researched and developed [13-15]. This noise reduction framework first learns the parameters of the neural network through supervised learning, using a large set of training data, and then uses the trained network to remove noise from the test data set. Recently, the possibility of improving the image quality of CT through deep learning has been proposed [12, 16]. According to these studies, deep learning algorithms trained through the

acquisition of routine dose (RD) CT and simulated LD CT images of the same patient were able to reconstruct CT images of quality comparable to conventional RD CT [12, 16].

Kang et al. proposed a new convolutional neural network (CNN) architecture optimized for CT denoising, which combines wavelet transformation with deep-processing [12]. The performance of this noise cancellation framework was evaluated using the dataset of the American Association of Physicists in Medicine (AAPM) Low Dose CT Grand Challenge [17] in 2016, and it was possible to reconstruct high-quality images even from quarter-dose CT images.

However, previous studies used only simulated LD CT images for testing, and fundamental studies on the applicability of these techniques in the clinical environment are lacking. Moreover, no study compared the image quality reconstructed by denoising technique through deep learning algorithms to that of the commercially available IR method. Another limit of previous studies is that the image quality was evaluated using a limited set of metrics to assess the similarity between LD and RD images, but lacked a quantitative evaluation of noise and

spatial resolution [12, 16, 18, 19]. Moreover, since the models developed in previous studies were trained only under fixed conditions, no insight was obtained about the dependence of performance on the training conditions.

Therefore, this study was undertaken to assess the image quality of LD CT using a deep learning-based denoising algorithm (DLA), compared with filtered back projection (FBP) and advanced modeled iterative reconstruction (ADMIRE), and to evaluate the changes in performance associated with various DLA training conditions.

MATERIALS AND METHODS

We conducted two separate phantom studies and one patient study to evaluate the performance of DLAs.

Patients included in the training and test sets

Our Institutional Review Board approved this retrospective study.

One-hundred patients who had undergone RD abdominal CT in our institution from August 2017 to January 2018 and had no metallic object in the scan range were included in the study to build the DLA training set. LD CT images at dose levels corresponding to 13%, 25%, and 50% of RD were simulated from RD CT images and reconstructed using FBP. This part of the study relied on ReconCT, a proprietary reconstruction software that allows simulating CT scans acquired with reduced radiation dose based on the raw data of original scans [20], and Siemens Healthcare (Erlangen, Germany) gave technical support in obtaining the simulated

scans.

To test the image quality improvement obtained using DLA in the clinical environment, we included eighteen patients who underwent low-dose (quarter dose) abdominal CT for clinical reasons from December 2017 to May 2018 as a test set.

Deep learning algorithm

A deep learning algorithm was implemented as a deep convolutional framelet-based denoising algorithm [19], which is an extended version of that developed in [12] and won the second place in the 2016 AAPM Low-Dose CT Grand Challenge. We used only a feed-forward network, except for the iterative process with a recursive neural network (RNN). Further implementation details and the network architecture can be found in [19]. All the model training and evaluation processes were carried out under the same computing environment: MATLAB (Version R2017a, The MathWorks Inc, Natick, MA) using two CUDA-enabled Nvidia Titan 12 GB graphic processing units (Nvidia Corporation, Santa Clara, Calif)

with CUDA 8.0/cuDNN 7.0.5 dependencies.

We trained the DLAs using the simulated LD CT images at each dose level as input data, and the RD FBP CT images as the ground truth. As a result, DLAs were developed under various training conditions, with three dose levels (DLA-1, 2, and 3 for the 13%, 25%, and 50% dose levels, respectively) and varying number of training cases (n = 1, 3, 5, 7, 10, 20, 50, and 100).

Scan protocol

All 100 patients included in the training set underwent routine dose abdominal CT with portal venous phase on a 128-channel dual energy scanner (SOMATOM Definition Edge, Siemens Healthcare, Forchheim, Germany) using the following parameters: 128*0.6 mm collimation, gantry rotation time of 0.5 seconds, reconstruction slice thickness of 4.0 mm, slice interval of 3.0 mm, tube potential of 100 kilovolt (peak) (kV[p]), variable milliamperage determined by x-, y-, and z-axis automated dose modulation (CARE Dose 4D, Siemens Healthcare) with a reference tube current-time of 210

effective mAs per patient. The timing of the portal venous phase scan was optimized using the bolus tracking technique.

For the test study, 18 patients underwent quarter dose abdominal CT with portal venous phase by changing the reference tube current-time to 40 effective mAs per patient under the same scan protocol as above. The phantom studies were also conducted under the same scanning protocol with fixed current-time values (50 mAs or 200 mAs to approximate 25% or 100% dose, respectively) and excluding dose modulation.

All LD CT images were initially reconstructed with the FBP and ADMIRE methods, and the LD-DLA images were obtained from LD-FBP images using DLAs.

Phantom studies

Two separate phantom studies were performed to assess the dependence of performance on the training set and to quantitatively evaluate image quality, respectively.

We performed CT scans of the anthropomorphic body

phantom (Kyoto Kagaku, Kyoto, Japan) and the American College of Radiology (ACR) CT accreditation phantom (model 464, Gammex-RMI, Middleton, WI).

The anthropomorphic phantom was used to get real paired body CT images at LD and RD without subjecting patients to additional radiation. We applied each of the trained DLAs with different numbers of training cases to the FBP images of the anthropomorphic phantom with 25% radiation dose. Then we obtained a mean squared error (MSE) between 25% dose phantom CT processed with DLA and RD phantom FBP images. The MSE is a commonly used metric to assess the error function in neural networks used for regression, and measures the distance between the predicted output and the true output [21]. Well-trained neural networks should have a very low MSE at the end of the training phase, and MSE is an adequate metric to compare the performance of differently trained DLAs.

To assess the objective image quality, noise power spectrum (NPS) and modulation transfer function (MTF) were calculated using the ACR phantom. To measure the NPS, we used the method proposed in a previous study [22]. Peak

frequency and area under the curve (AUC) of the NPS were calculated.

We used the circular edge method for MTF measurements, based on a study by Friedman et al. [22], since it may reflect the spatial resolution dependence of image contrast and noise level using a non-linear CT reconstruction algorithm, unlike conventional methods. The ACR phantom contains three inserts and one air cavity for Hounsfield unit (HU) accuracy measurement. We measured the MTF in the three inserts with different contrasts, which provided nominal HU values of 95, 955, and 120 kVp for polyethylene (Disc 1), bone (Disc 2) and acrylic (Disc 3), respectively. We obtained the average MTF value after 20 repeated CT scans of a phantom and measured the MTF-50 to simplify the MTF comparison. The MTF-50 is defined as the spatial frequency at which the MTF becomes one half of its zero- frequency value, in units of mm^{-1}.

The MSE of anthropomorphic body phantom images and the NPS and MTF of ACR phantom images were calculated with MATLAB.

Patient study

To objectively assess image quality in the clinical environment, a single radiologist (blinded, with 5 years of clinical experience in CT interpretation) measured the mean image noise of the LD-DLA, LD-FBP and LD-ADMIRE images of the 18 patients. Mean image noise is defined as the standard deviation of the CT attenuation value (HU) obtained by manually placing oval ROIs (100-150 mm^2) in homogeneous regions of the liver, subcutaneous fat, paraspinal muscles and abdominal aorta. Each ROI was placed at the exact same locations for the LD-DLA, LD-FBP, and LD-ADMIRE images.

All noise measurements for patient images were performed with a commercially available workstation using the RadiAnt DICOM viewer (Mexidant, Poznan, Poland).

Statistical analysis

All statistical analyses were performed with SPSS software (Version 22.0, Chicago, IL).

A paired t-test was performed to compare the MTF values

of phantom images, and repeated measure analysis of variance followed by pairwise comparisons was used to compare mean image noise levels of patient test images with different reconstruction methods and DLAs. After Bonferroni adjustment for the ten comparisons (all pairwise comparisons for FBP, ADMIRE, DLA-1, 2, and 3.), a p-value < 0.005 was considered significant. The modified Wald method was used to determine the confidence interval for the presence of image artifacts in 2,998 LD CT image slices of 18 patients.

RESULTS

The baseline patient characteristics and radiation dose information of the training and the test set are shown in Table 1. The age of the patients (mean ± standard deviation) was 63.5 ± 13.0 years and 35.0 ± 12.2 years in the training and test population, respectively. Because low-dose abdominal CT is performed in adults under 40 years of age in this institution, the test population was younger than the training population. Male subjects represented 56% (n = 56) and 66% (n = 12) of the two populations, respectively. The mean computed tomography dose index (CTDIvol), dose-length product (DLP), and peak mAs in the test population were approximately one fourth of those in the training population.

Phantom studies

Figure 1 shows the MSE measured using 25% dose CT in the anthropomorphic body phantom. As shown in the figure, the DLAs achieved the lowest MSE value with 10 training samples in DLA-1 and DLA-2, and 50 training samples in DLA-3.

Compared with the MSE value without training, the MSE value decreased by 32.59%, 31.52%, and 24.16% (DLA-1, DLA-2, and DLA-3, respectively) to its lowest value. After the lowest value was reached the MSE was considered to have reached a plateau, as it varied by 0.27%, 0.79%, and 0.88% in the three DLA models, respectively. The MSE of the DLA-3 CNN was higher than those of DLA-1 and DLA-2 regardless of the number of training cases. LD-DLA achieved lower noise level than LD-FBP and LD-ADMIRE but led to a loss of spatial resolution.

As shown in Figure 2 and Table 2, comparing LD-DLA, LD-ADMIRE, and LD-FBP, the peak frequency and AUC of the NPS curves in the LD-DLA images showed a lower value than the LD-FBP and LD-ADMIRE images, at all dose levels. In addition, the NPS curves were shifted towards lower spatial frequencies in LD-DLA images compared with the ADMIRE images.

Table 3 shows the mean MTF-50 value of each reconstructed CT image of three different disc objects of the phantom CT with 25% radiation dose level. The MTF-50 were

significantly lower in the LD-DLA images than in the LD-ADMIRE and LD-FBP images (all $p < 0.001$). Moreover, as the DLA training dose decreased, also the MTF decreased. Figure 3 shows the MTF for the DLA, ADMIRE, and FBP methods acquired at the 25% dose level for various contrast inserts.

Patient study

Mean image noise in all the LD-DLA images were significantly lower than in the LD-ADMIRE and LD-FBP images (all $p < 0.001$). Table 5 shows the mean image noise values according to the image reconstruction method. As the training radiation dose of DLA decreased, also the mean image noise significantly decreased ($p < 0.001$).

Figure 4 shows representative CT images of each reconstruction and processing method. The first column shows low-dose (25%) abdominal CT using five different methods and enlarged images are in the second column. The mean image noise of all LD-DLA images was lower than that of the LD-ADMIRE and LD-FBP images, and the DLA-1 image showed

the lowest mean image noise. As the training radiation dose of the DLA increased, the mean image noise of the processed CT images increased. Moreover, in a subjective analysis, DLA images showed no additional image artifacts in 2,998 image slices (0%, 95% confidence interval (CI), 0 – 4 slices)

DISCUSSION

We showed that a DLA can achieve lower image noise than ADMIRE, a state-of-the-art commercially available iterative reconstruction method. Moreover, to our knowledge, this is the first study evaluating the alternative approaches to CT denoising in a clinical setting, specifically comparing DLAs with the FBP and IR algorithms for real low-dose CT images of patients.

We sought to evaluate the feasibility of using a DLA for LD CT of both phantoms and patients. Because of the ethical difficulty of studying multiple doses in a single patient, quantitative image analysis was performed with phantoms. An anthropomorphic body phantom was used to validate the DLAs, and an ACR phantom was used to test its use. An additional test study with a small number of patients was conducted. As a result, DLAs achieved significant noise reduction in low-dose CT compared with previous reconstruction methods.

A significant strength of this study is that although the IR algorithm is vendor-specific and of limited applicability to

other CT systems, FBP-based DLAs can be used in a more general way because it allows image-based reconstruction without raw data in the FBP method [9]. In addition, a previous study showed another advantage of DLAs, i.e. the shorter computational time required compared with IR [12]. Although direct comparison was difficult in our experiment, because we could not test the ADMIRE method in the experiment, the DLAs might potentially be able to reduce the computing load also with respect to ADMIRE. Moreover, it is also notable that the study created the DLAs using only FBP images, a widely known, open and universal technology. Another strength of this study is that we evaluated DLA performance in quarter-dose CT, a dosage widely used in clinical low-dose CT [23, 24].

In the test study with eighteen patients, we could test the DLAs with low-dose CT in a real clinical environment. DLAs effectively reduced image noise in the test study and, moreover, did not show any additional image artifacts. We were initially concerned about the so-called "black-box" nature of deep learning techniques, a particularly significant problem in neural networks due to their complexity, and specifically we

were concerned that the processed image might show unknown or unexplained artifacts. But there were no image artifacts in about 3,000 images we tested, so that we can conclude that artifacts are highly unlikely under these test conditions. However, additional evaluation will be required under different conditions (e.g. images with beam hardening artifacts, images in elderly patients with motion artifacts, or images from other parts of the body) that we did not consider in either training or testing.

The resulting images appear over-smoothed, just like those produced by the first-generation IR algorithm [9]. The IR technique is currently used to lower the radiation dose by lowering image noise with better spatial resolution. However, from the radiologist's point of view, CT images reconstructed by IR have a plastic-like texture and feel unfamiliar or somewhat awkward to radiologists familiar with images reconstructed by FBP [9, 11]. Since we trained the DLAs with LD-FBP and RD-FBP images, we expected that they could reproduce the NPS of the FBP method and maintain the noise texture. However, compared to FBP, the NPS value decreased,

but the NPS curve was left-shifted in the DLA images. Moreover, DLAs appeared to give an image texture similar to that of ADMIRE. It is difficult to find a convincing explanation of these results, and further investigations are needed.

This study compared the image quality characteristics of LD-DLA CT with those of LD-FBP and LD-ADMIRE CT with quantitative assessment using NPS, MTF, and mean image noise. NPS and MTF measured in the phantom study showed less noise and less spatial resolution in the LD-DLA images. Regarding the dependence on training conditions, with decreasing DLA training radiation dose the NPS and MTF values of the LD-DLA images decreased. This tendency is difficult to explain, but it might be related to the intrinsic limitation of deep learning, which needs proper training intensity.

Since one of the most tangible costs of noise reduction is the loss of spatial resolution in CT image reconstruction, a reconstruction method with a better noise-resolution tradeoff is preferable. In this context, IR has been a popular method for CT reconstruction. Previous conventional MTF measurements

used highly dense materials and very low noise to improve measurement precision, and MTF was shown to be greater in IR than in FBP in high-contrast objects [25]. But in a real clinical image, the images could have lower contrast and higher noise than in an ideal environment. In this regard, Richard et al. showed how the MTF in IR could be different for varying contrast and noise levels [26]. The task-based MTF for model-based IR was shown to be greater than for FBP for high-dose and high-contrast objects, but for low-dose and low-contrast objects its performance decreased [26]. In this regard, we evaluated the MTF with three different contrast materials. Our results showed a little loss of spatial resolution by DLAs in all disc objects, but the difference was marginal compared to FBP. Although the MTF was increased as the training radiation dose increased, it may be difficult to go beyond the limits of FBP, as the DLA training was based on RD FBP images. But the actual performance difference is likely to be small between DLA and FBP, and further studies will be needed to explore this issue. If the DLAs were trained with CT images with improved spatial resolution, such as ADMIRE, as

the ground truth, the MTF could be further improved in DLA compared with FBP.

Regarding the dependence of the DLA performance on the training conditions, we calculated the MSE between the RD FBP and the LD DLA images. It is usually not appropriate to evaluate the performance with full-reference metrics such as mean square error and peak signal-to-noise ratio, since these measures refer only to the pixel-wise differences with the reference image, without considering factors that may influence the detectability, such as the NPS and MTF, which we have previously used to evaluate the model [27, 28]. However, by measuring the degree of proximity to the reference image we analyzed how the model converges as the number of training samples increases. Although the performance of deep learning algorithms is proportional to the amount of training data [29, 30], that of the model we propose quickly converges with 10-20 samples. Considering that there are 64 independent 32x32 patches per slice and a patient corresponds on average to 160 slices, a total of more than 100,000 data are used in the training set. We can conclude that a training dataset with a large number

of patches from one patient was sufficient to achieve satisfactory training in our experimental setting.

Our study has several limitations. First, we did not evaluate the task-based performance of the DLA. Despite the lower image noise in the objective physical measurements, the noise texture of the DLA image changed, which can affect clinical diagnostic performance such as lesion detection. Although this was a preliminary study aimed to investigate the performance of DLAs in image noise and spatial resolution, future studies are needed to confirm this exploratory result and determine the effects of DLAs on task-based performance. For DLAs to be used in a clinical environment, it will be necessary to demonstrate that they can improve image quality better and more effectively than the currently commercialized IR techniques. Second, the DLA used in this study was trained with simulated low-dose images. Therefore, we could not tell how the results would differ using actual low-dose images as training data. In addition, we did not explore the effects of image artifacts, because we included in the training set only qualified images without significant artifacts. The patients

included in the test were adults under the age of 40 who had quarter dose abdomen CT. Therefore, since the DLA was only tested on one kind of LD CT on a single body part, further studies using various LD CT protocols in different body parts may be required.

In conclusion, DLA could improve image quality in terms of noise reduction in CT. DLAs achieved less noise than FBP and ADMIRE in LD CT images, but did not maintain the spatial resolution. Regarding the dependence of performance on the training conditions, we found that lower radiation doses in the training images led to less noise.

Acknowledgement

We thank Seongyong Pak (Siemens Healthcare Ltd, Korea) for technical support on simulating low-dose CT images.

REFERENCES

1. Smith-Bindman R, Lipson J, Marcus R, et al. Radiation dose associated with common computed tomography examinations and the associated lifetime attributable risk of cancer. Archives of internal medicine. 2009;169:2078-86.

2. Brenner DJ, Hall EJ. Computed tomography—an increasing source of radiation exposure. New England Journal of Medicine. 2007;357:2277-84.

3. de González AB, Mahesh M, Kim K-P, et al. Projected cancer risks from computed tomographic scans performed in the United States in 2007. Archives of internal medicine. 2009;169:2071-7.

4. Hara AK, Paden RG, Silva AC, et al. Iterative reconstruction technique for reducing body radiation dose at CT: feasibility study. American Journal of Roentgenology. 2009;193:764-71.

5. Nakayama Y, Awai K, Funama Y, et al. Abdominal CT with low tube voltage: preliminary observations about radiation dose, contrast enhancement, image quality, and noise. Radiology. 2005;237:945-51.

6. Sagara Y, Hara AK, Pavlicek W, et al. Abdominal CT: comparison of low-dose CT with adaptive statistical iterative reconstruction and routine-dose CT with filtered back projection in 53 patients. American Journal of Roentgenology. 2010;195:713-9.

7. McCollough CH, Bruesewitz MR, Kofler Jr JM. CT dose reduction and dose management tools: overview of available options. Radiographics. 2006;26:503-12.

8. Yu L, Liu X, Leng S, et al. Radiation dose reduction in computed tomography: techniques and future perspective. Imaging in medicine. 2009;1:65.

9. Geyer LL, Schoepf UJ, Meinel FG, et al. State of the art: iterative CT reconstruction techniques. Radiology. 2015;276:339-57.

10. Holmquist F, Nyman U, Siemund R, et al. Impact of iterative reconstructions on image noise and low-contrast object detection in low kVp simulated abdominal CT: a phantom study. Acta Radiologica. 2016;57:1079-88.

11. Prakash P, Kalra MK, Kambadakone AK, et al. Reducing abdominal CT radiation dose with adaptive statistical iterative reconstruction technique. Investigative radiology. 2010;45:202-10.

12. Kang E, Min J, Ye JC. A deep convolutional neural network using directional wavelets for low- dose X- ray CT reconstruction. Medical Physics. 2017;44.

13. Jain V, Seung S. Natural image denoising with convolutional networks. Advances in Neural Information Processing Systems; 2009. p. 769-76.

14. Nasri M, Nezamabadi-pour H. Image denoising in the wavelet

domain using a new adaptive thresholding function. Neurocomputing. 2009;72:1012-25.

15. Xie J, Xu L, Chen E. Image denoising and inpainting with deep neural networks. Advances in Neural Information Processing Systems; 2012. p. 341-9.

16. Chen H, Zhang Y, Zhang W, et al. Low-dose CT via convolutional neural network. Biomedical optics express. 2017;8:679-94.

17. McCollough C. TU- FG- 207A- 04: Overview of the Low Dose CT Grand Challenge. Medical physics. 2016;43:3759-60.

18. Chen H, Zhang Y, Kalra MK, et al. Low-dose CT with a residual encoder-decoder convolutional neural network. IEEE transactions on medical imaging. 2017;36:2524-35.

19. Kang E, Chang W, Yoo J, et al. Deep convolutional framelet denosing for low-dose ct via wavelet residual network. IEEE transactions on medical imaging. 2018;37:1358-69.

20. Ellmann S, Kammerer F, Brand M, et al. A novel pairwise comparison-based method to determine radiation dose reduction potentials of iterative reconstruction algorithms, exemplified through circle of willis computed tomography angiography. Investigative radiology. 2016;51:331-9.

21. Silva LM, de Sá JM, Alexandre LA. Data classification with

multilayer perceptrons using a generalized error function. Neural Networks. 2008;21:1302-10.

22. Friedman SN, Fung GS, Siewerdsen JH, et al. A simple approach to measure computed tomography (CT) modulation transfer function (MTF) and noise- power spectrum (NPS) using the American College of Radiology (ACR) accreditation phantom. Medical physics. 2013;40:051907.

23. Kim K, Kim YH, Kim SY, et al. Low-dose abdominal CT for evaluating suspected appendicitis. New England Journal of Medicine. 2012;366:1596-605.

24. Kim SY, Lee KH, Kim K, et al. Acute appendicitis in young adults: low-versus standard-radiation-dose contrast-enhanced abdominal CT for diagnosis. Radiology. 2011;260:437-45.

25. Droege RT, Morin RL. A practical method to measure the MTF of CT scanners. Medical physics. 1982;9:758-60.

26. Richard S, Husarik DB, Yadava G, et al. Towards task- based assessment of CT performance: System and object MTF across different reconstruction algorithms. Medical physics. 2012;39:4115-22.

27. Chen H, Zhang Y, Zhang W, et al. Low-dose CT denoising with convolutional neural network. Biomedical Imaging (ISBI 2017), 2017 IEEE 14th International Symposium on: IEEE; 2017. p. 143-6.

28. Yang Q, Yan P, Zhang Y, et al. Low dose CT image denoising using a generative adversarial network with Wasserstein distance and perceptual loss. IEEE transactions on medical imaging. 2018.

29. Sajjadi M, Javanmardi M, Tasdizen T. Regularization with stochastic transformations and perturbations for deep semi-supervised learning. Advances in Neural Information Processing Systems; 2016. p. 1163-71.

30. Tarvainen A, Valpola H. Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results. Advances in neural information processing systems; 2017. p. 1195-204.

Table 1. Baseline characteristics of the study population

	Training population (n = 100)	Test population (n = 18)
Demographics		
Mean age (yr)	63.5 ± 13.0	35.0 ± 12.2
BMI	22.9 ± 3.55	23.0 ± 2.50
Radiation dose information		
CTDIvol (mGy)	6.50 ± 1.31	1.39 ± 0.31
DLP (mGycm)	312.8 ± 74.3	72.1 ± 20.4
Peak mAs	240.1 ± 58.6	52.8 ± 9.15

Note – BMI, body mass index; CTDIvol, computed tomography dose index volume; DLP, dose length product

Table 2. AUC and peak frequency of the noise power spectrum curve in each reconstruction method

	DLA−1	DLA−2	DLA−3	ADMIRE	FBP
AUC ($\times 10^4$, $HU^2\,mm^2$)	1.37	1.91	3.86	5.52	8.89
Peak frequency (mm^{-1})	0.05	0.07	0.13	0.18	0.22

Note− ADMIRE, advanced modeled iterative reconstruction; DLA, deep learning based denoising algorithm; FBP, filtered back projection; AUC, area under the curve.

Table 3. MTF-50s of the 25% ACR phantom CT according to the different discs and reconstruction methods

Reconstruction method	Disc 1 (Polyethylene)	Disc 2 (Bone)	Disc 3 (Acrylic)	*P value
DLA-1	0.410	0.445	0.415	
DLA-2	0.418	0.446	0.423	
DLA-3	0.434	0.447	0.435	< 0.001
FBP	0.454	0.448	0.452	
ADMIRE	0.563	0.499	0.551	

*P value for all pairwise comparison, Note- ADMIRE, advanced modeled iterative reconstruction; DLA, deep learning based denoising algorithm; FBP, filtered back projection; MTF, modulation transfer function.

Table 4. Mean image noise (HU) according to the image reconstruction method

Reconstruction method	DLA-1	DLA-2	DLA-3	ADMIRE	FBP	*P value
Measuring site and mean image noise (SD)						
Liver	6.29 (1.20)	8.54 (1.61)	11.42 (2.04)	16.16 (1.57)	23.0 (2.53)	<0.001
Subcutaneous fat	6.03 (1.23)	7.14 (1.48)	9.94 (1.53)	13.24 (1.27)	20.43 (2.47)	
Paraspinal muscle	5.34 (0.90)	6.58 (1.27)	9.86 (1.71)	14.46 (1.10)	21.98 (2.97)	
Abdominal aorta	6.31 (1.13)	7.74 (1.31)	11.06 (2.44)	15.48 (1.96)	21.33 (2.31)	

*P value for all pairwise comparison.

Note- ADMIRE, advanced modeled iterative reconstruction; DLA, deep learning based denoising algorithm; FBP, filtered back projection; MTF, modulation transfer function.

FIGURES

Figure 1. Mean squared errors of the denoised phantom images as a function of the number of training cases. Note – DLA-1, DLA-2, DLA-3, deep learning based denoising algorithm trained with 13%, 25%, and 50% doses.

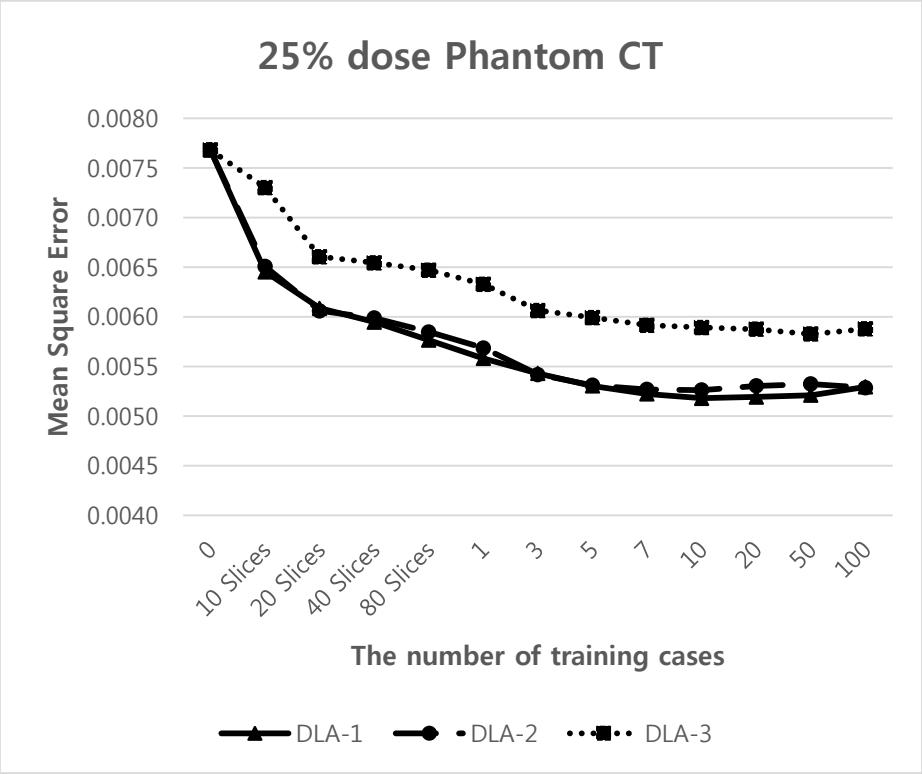


Figure 2. Noise power spectrum (NPS) of the 25% phantom CT according to the reconstruction method. The NPS curves are shifted towards lower spatial frequencies in the LD-DLA images, which were produced by a DLA trained on 50% radiation dose level (DLA-3). Note - DLA, deep learning based denoising algorithm

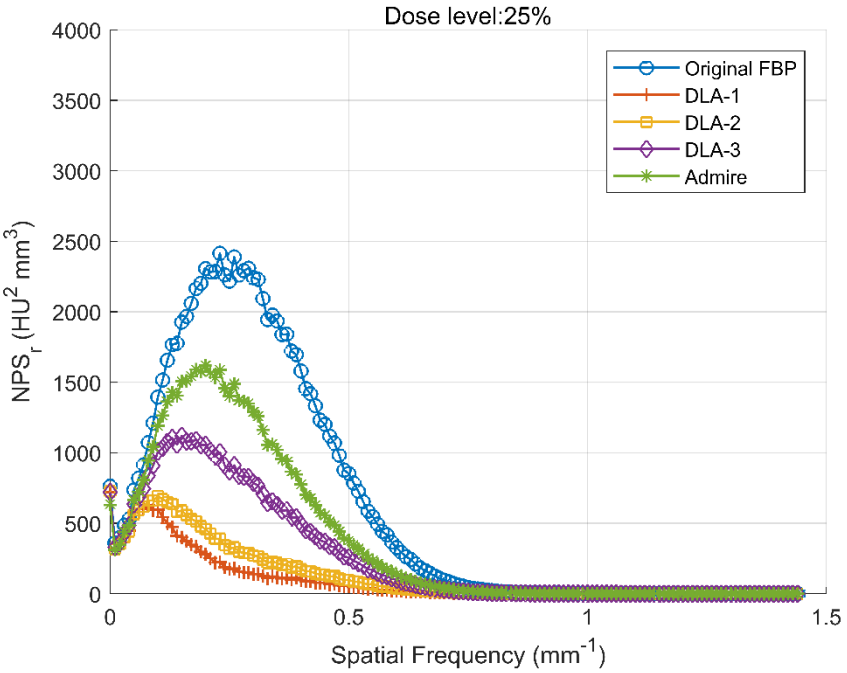
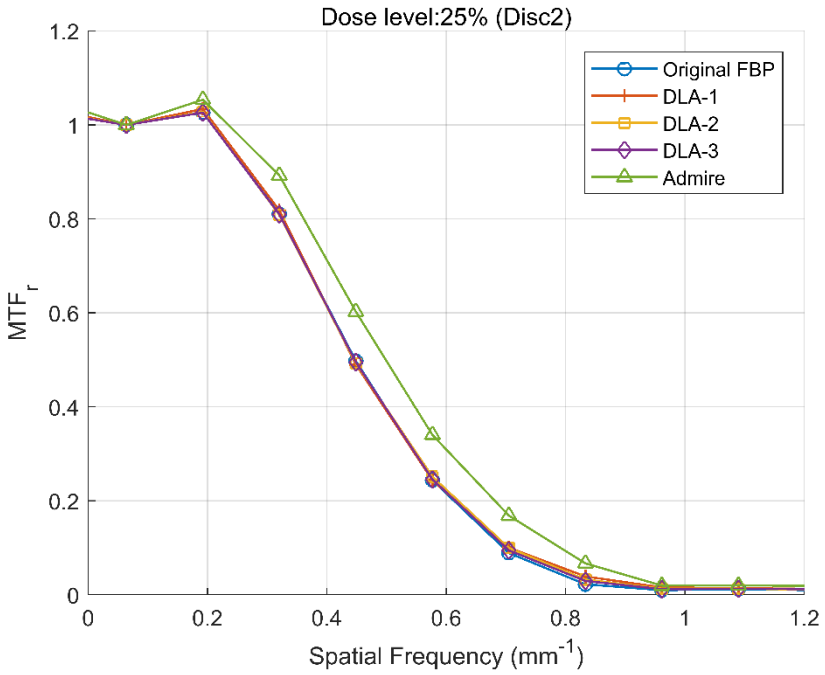
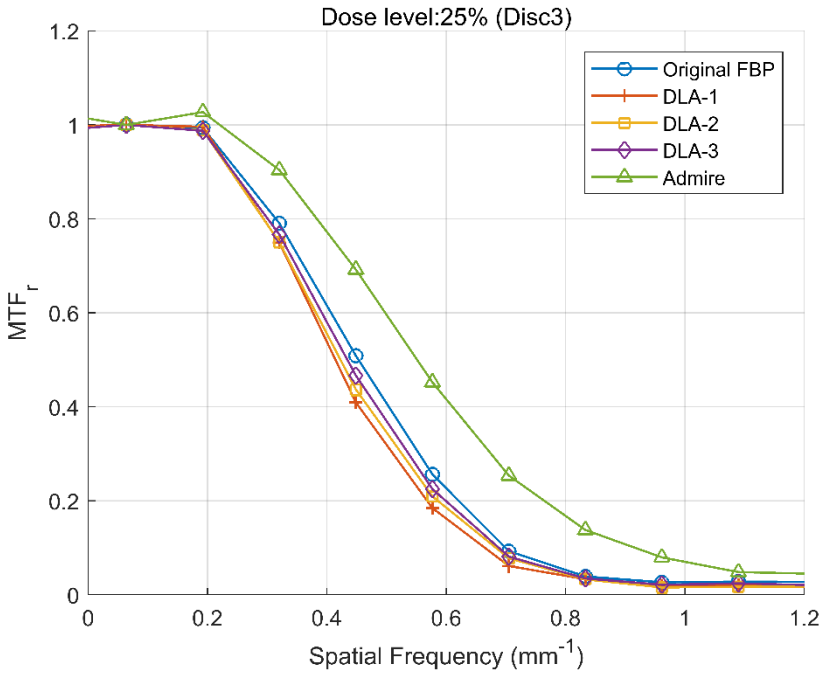


Figure 3. Comparison of MTF with five different CT reconstruction or processing methods in three different discs (A) polyethylene, (B) bone, (C) acrylic. Note – MTF, modulation transfer function.

(A)



(B)


Dose level:25% (Disc3)
Original FBP
DLA-1
DLA-2
DLA-3
Admire
MTF$_r$
Spatial Frequency (mm^{-1})
0
0.2
0.4
0.6
0.8
1
1.2

(C)

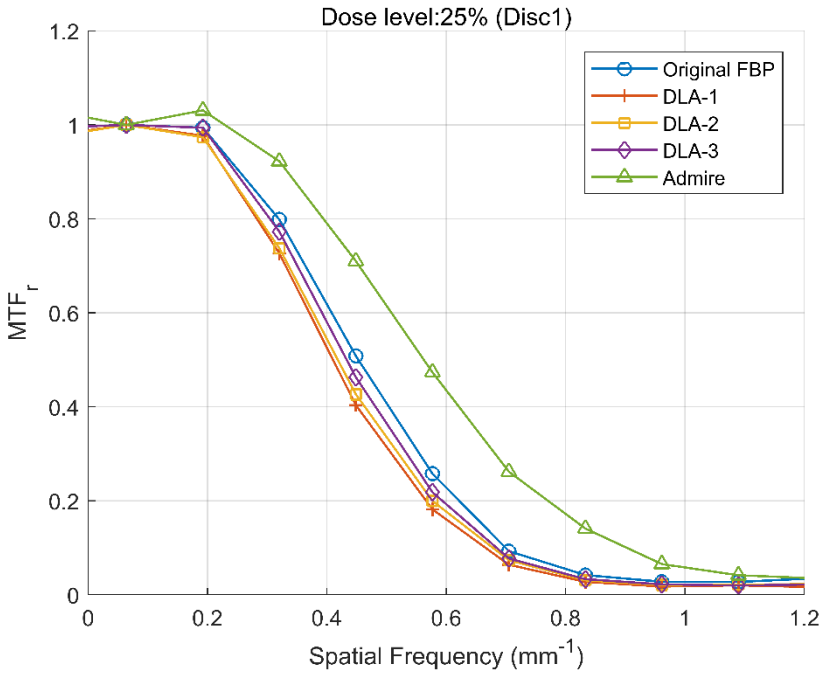
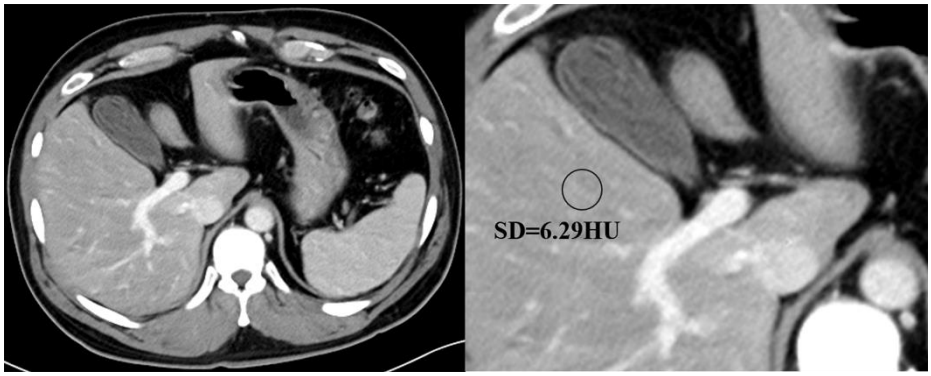
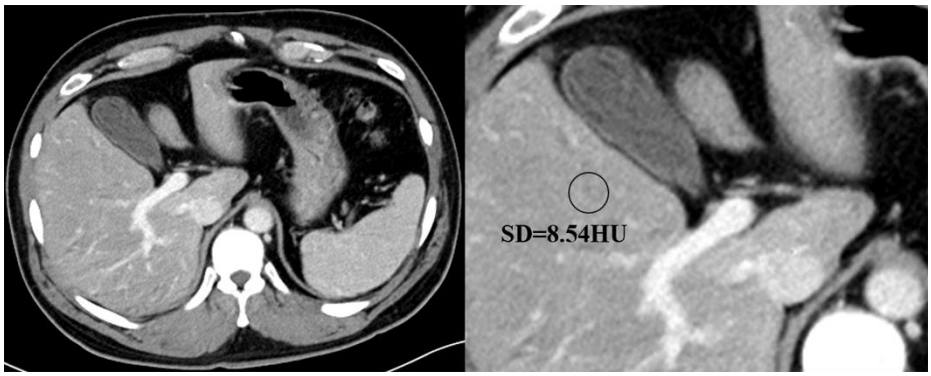

Dose level:25% (Disc1)
Original FBP
DLA-1
DLA-2
DLA-3
Admire
MTF_r
Spatial Frequency (mm^-1)
0
0.2
0.4
0.6
0.8
1
1.2

Figure 4. Low-dose abdominal CT images of the test set with different reconstruction methods. (A) LD-FBP, (B) LD-ADMIRE, (C) DLA-1, (D) DLA-2, (E) DLA-3.

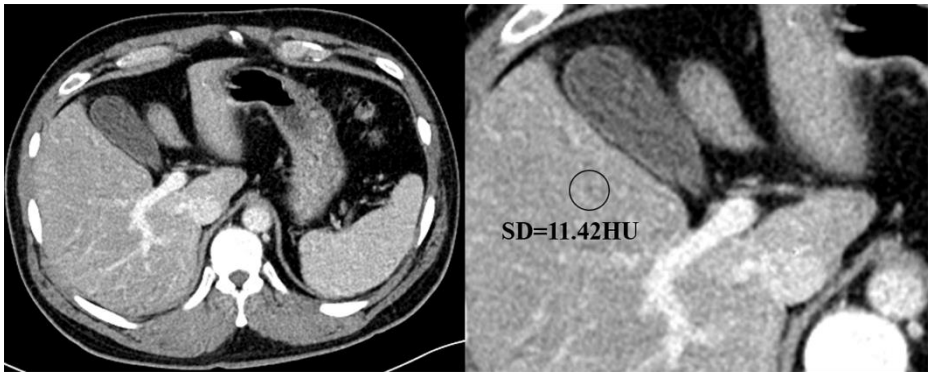
(A)



(B)

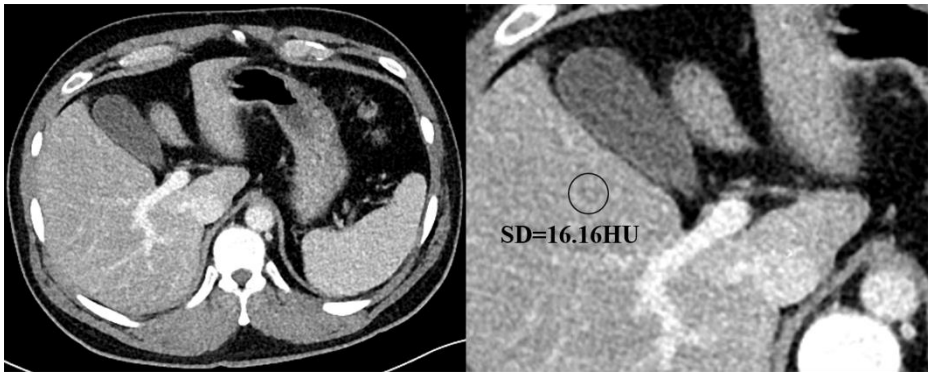


(C)



(D)

(E)



국문초록

서론: 저선량 전산화 단층촬영에서 FBP (filtered back projection) 및 ADMIRE (advanced modeled iterative reconstruction)와 비교하여 딥러닝 기반 알고리즘 (deep learning algorithm; DLA)을 이용하였을 때의 영상 화질 향상에 대한 연구이다.

방법: 이 후향적 연구는 기관 검토위원회의 승인을 받았다. FBP를 이용한 정상 선량 (routine dose, RD) 복부 CT를 시행한 총 100명의 환자를 대상으로 딥러닝 알고리즘의 훈련 세트를 만들었다. RD CT 영상으로부터 13 %, 25 %, 50 %의 선량 수준의 저선량 CT 영상을 시뮬레이션하고 FBP를 이용하여 재구성하였다. 우리는 시뮬레이션 된 저선량 CT 이미지를 입력 데이터로 사용하고 정상 선량 CT 이미지를 정답으로 하여 다양한 조건에서 DLA를 훈련시켰다. DLA의 유효성을 확인하기 위해 평균화 제곱 오류 (Mean squared error, MSE)를 의인화 팬텀을 사용하여 측정했다. 훈련과 검증을 거친 DLA를 시험하기 위해 팬텀을 이용한 연구와, 18명의 저선량 복부 CT를 시행한 환자를 대상으로 한 연구를 수행하였다. 각 연구에서 FBP, ADMIRE 및 DLA를 이용하여 팬텀 및 환자의 저선량 CT 영상을 재구성하였다. 각각의 방법으로 재구성된 영상에서 영상 품질을 테스트하고 비교하기 위해서, ACR 팬텀을 사용하여

잡음 전력 스펙트럼 (noise power spectrum, NPS)과 변조 전달 함수 (modulation transfer function, MTF)를 측정하고 환자 데이터를 사용하여 평균 영상 잡음(mean image noise)을 측정했다.

결과: LD-DLA는 팬텀 및 환자 연구 모두에서 LD-FBP 및 LD-ADMIRE보다 낮은 잡음 수준을 보였다. 팬텀 연구에서, LD-DLA의 NPS 곡선의 피크 값과 AUC는 LD-FBP 또는 LD-ADMIRE보다 낮았다. 환자 연구에서 LD-DLA 이미지는 LD-ADMIRE 이미지보다 유의하게 낮은 평균 이미지 노이즈를 보였고 (모두 $p<0.001$), 추가 인공물도 보이지 않았다.

결론: LD-DLA 영상은 LD-FBP 및 LD-ADMIRE 영상보다 잡음이 적음을 보여주었지만 공간 분해능은 개선시키지 못하였다. 그리고 더 적은 방사선량으로 촬영한 이미지로 훈련한 DLA일수록 잡음이 적게 나타났다.

주요어: 전산화 단층촬영, 이미지 노이즈 제거, 딥러닝, 반복적 재구성 영상

학번: 2017-34436