Korean Journal of Radiology

KJR

Low-Dose Abdominal CT Using a Deep Learning-Based Denoising Algorithm: A Comparison with CT Reconstructed with Filtered Back Projection or Iterative Reconstruction Algorithm

Yoon Joo Shin, MD[1], Won Chang, MD[2], Jong Chul Ye, PhD[3], Eunhee Kang, PhD[3], Dong Yul Oh, BS[4], Yoon Jin Lee, MD[2], Ji Hoon Park, MD[2], Young Hoon Kim, MD[2]

[1]Department of Radiology, Konkuk University Medical Center, Seoul, Korea; [2]Department of Radiology, Seoul National University Bundang Hospital, Seongnam, Korea; [3]Bio Imaging and Signal Processing Lab, Department of Bio and Brain Engineering, KAIST, Daejeon, Korea; [4]Interdisciplinary Program in Bioengineering, Seoul National University, Seoul, Korea

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

Materials and Methods: One hundred routine-dose (RD) abdominal CT studies reconstructed using FBP were used to train the DLA. Simulated CT images were made at dose levels of 13%, 25%, and 50% of the RD (DLA-1, -2, and -3) and reconstructed using FBP. We trained DLAs using the simulated CT images as input data and the RD CT images as ground truth. To test the DLA, the American College of Radiology CT phantom was used together with 18 patients who underwent abdominal LD CT. LD CT images of the phantom and patients were processed using FBP, ADMIRE, and DLAs (LD-FBP, LD-ADMIRE, and LD-DLA images, respectively). To compare the image quality, we measured the noise power spectrum and modulation transfer function (MTF) of phantom images. For patient data, we measured the mean image noise and performed qualitative image analysis. We evaluated the presence of additional artifacts in the LD-DLA images.

Results: LD-DLAs achieved lower noise levels than LD-FBP and LD-ADMIRE for both phantom and patient data (all $p < 0.001$). LD-DLAs trained with a lower radiation dose showed less image noise. However, the MTFs of the LD-DLAs were lower than those of LD-ADMIRE and LD-FBP (all $p < 0.001$) and decreased with decreasing training image dose. In the qualitative image analysis, the overall image quality of LD-DLAs was best for DLA-3 (50% simulated radiation dose) and not significantly different from LD-ADMIRE. There were no additional artifacts in LD-DLA images.

Conclusion: DLAs achieved less noise than FBP and ADMIRE in LD CT images, but did not maintain spatial resolution. The DLA trained with 50% simulated radiation dose showed the best overall image quality.

Keywords: *Deep learning; Denoising; Iterative reconstruction; CT; Phantoms; Radiation dose*

INTRODUCTION

Computed tomography (CT) is a widely used imaging modality because of its wide availability. However, concerns regarding radiation exposure persist as CT itself emits ionizing radiation and the number of CT scans performed increases (1, 2). Since reducing the radiation dose inevitably affects the performance of CT, many attempts have been made to find the best compromise between radiation risk and diagnostic performance (3-6). In this context, one of the most successful attempts is the use of the iterative reconstruction (IR) technique in CT image reconstruction (4, 6-10). However, commercially available IR algorithms are vendor-specific and only partly disclosed by manufacturers. Furthermore, it can present the disadvantage of requiring long computation times to execute repetitive reconstruction process (7).

Deep learning techniques have become widespread in the medical imaging field due to recent advances in machine learning and improvements in hardware performance. Image denoising algorithms using artificial neural networks have been intensively researched and developed (11-13). Recently, the possibility of using deep learning to improve the image quality of CT has been proposed (14, 15). According to these studies, deep learning algorithms trained using the acquisition of routine-dose (RD) CT and simulated low-dose (LD) CT images of the same patient provide similar quality to conventional RD CT (14, 15). Kang et al. (14) proposed a new convolutional neural network architecture optimized for CT denoising, which combines wavelet transformation with deep-processing. The performance of this denoising framework was evaluated using the dataset of the American Association of Physicists in Medicine Low Dose CT Grand Challenge (16) in 2016 and they were able to reconstruct high-quality images even from quarter-dose CT images.

However, previous studies have several limitations. First, they used only simulated LD CT images for testing and fundamental studies on the applicability of these techniques in the clinical environment are lacking. Second, image quality was evaluated using a limited set of metrics and a quantitative evaluation of noise and spatial resolution was not performed (14, 15, 17, 18). Third, they did not compare the quality of CT images processed using deep learning algorithms to that of commercially available IR methods. Finally, since the previously developed models were trained only under fixed conditions, no insight was obtained regarding the dependence of performance on training conditions.

Therefore, this study tried to assess the image quality of LD CT using a deep learning-based denoising algorithm (DLA) and compare it with filtered back projection (FBP) and advanced modeled iterative reconstruction (ADMIRE) methods, as well as 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. First, we collected CT images from 100 patients and built a training set. The DLAs was trained with simulated LD CT from 100 patients as input and the RD CT of those patients as the ground truth under various training conditions. After training, a phantom study was performed to validate the DLAs. Finally, various DLAs were tested in another phantom study using the noise power spectrum (NPS) and modulation transfer function (MTF), as well as patient data from another 18 patients (Fig. 1).

Patients Included in the Training and Test Sets

Our Institutional Review Board approved this retrospective study and waived the requirement for obtaining informed consent. One-hundred patients who had undergone RD abdominal CT in our institution from August 2017 to January 2018 and had no metallic objects in the scan range were included to build the training set for DLA. 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 (Siemens Healthineers, Forchheim, Germany), a proprietary reconstruction software that allows simulating CT scans acquired with reduced radiation dose based on the raw data of original scans (19). Siemens Healthineers provided technical support in obtaining the simulated scans.

To test the image quality improvement obtained using DLA in the clinical environment, we recruited another 18 patients who underwent LD (quarter dose) abdominal CT for clinical reasons from December 2017 to May 2018 as a test set. Therefore, we included 100 patients in the training datasets and 18 patients in the test datasets, regardless of their disease or chief complaint. Most patients were outpatients receiving routine cancer surveillance or who had nonspecific symptoms such as abdominal pain. In our
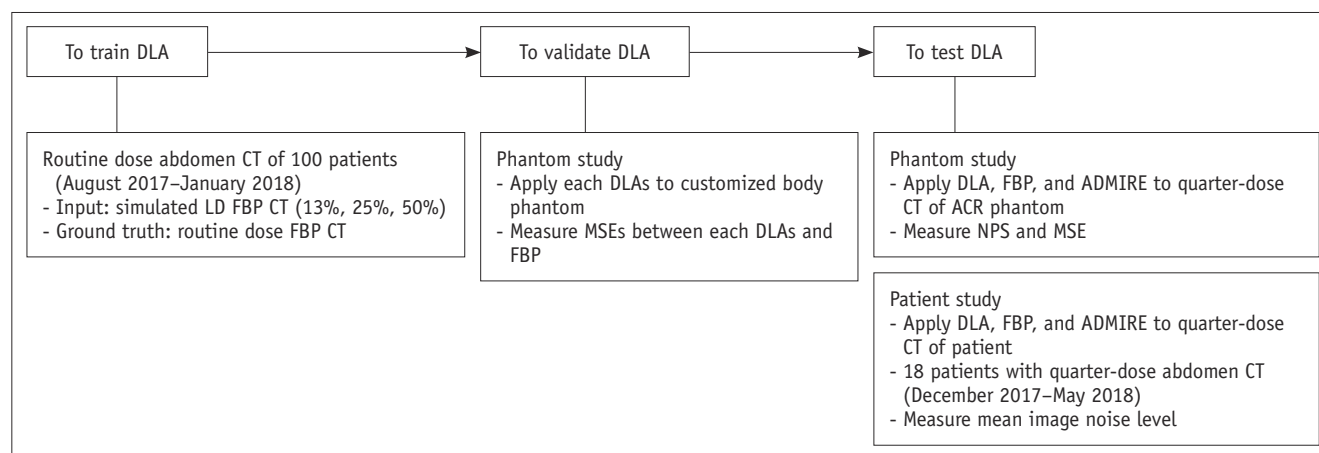
Fig. 1. Schematic diagram showing study population in phantom and patient studies. ACR = American College of Radiology, ADMIRE = advanced modeled iterative reconstruction, DLA = deep learning-based denoising algorithm, FBP = filtered back projection, LD = low-dose, MSE = mean squared error, NPS = noise power spectrum

institution, we used quarter dose abdominal CT to reduce the radiation risk in young patients (under 45 years of age) with abdominal symptoms, thereby we could use these LD CT to test the DLA.

Deep Learning Algorithm

The deep learning algorithm that was implemented was a deep convolutional framelet-based denoising algorithm (18), which was an advanced version of that proposed by Kang et al. (14) in 2016. We only used a feed-forward network, except for the iterative process, which used a recursive neural network. Further details can be found in the original article by Kang et al. (18). All the model training and evaluation processes were carried out under the same computing environment, MATLAB (Version R2017a, The MathWorks, Inc., Natick, MA, USA) using two compute unified device architecture (CUDA)-enabled Nvidia Titan 12 GB graphic processing units (Nvidia Corporation, Santa Clara, CA, USA) with CUDA 8.0/cuDNN 7.0.5 dependencies.

We trained the DLAs using the simulated LD CT images at each dose level as the input data and the RD CT images reconstructed with FBP 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).

CT Scanning Protocol

All patients included in the training set underwent RD abdominal CT with portal venous phase on a 128-channel scanner (SOMATOM Definition Edge, Siemens Healthineers)

using the following parameters: 128 x 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 kVp, variable milliamperage determined by x-, y-, and z-axis-automated dose modulation (CARE Dose 4D, Siemens Healthineers) with a reference tube current-time of 210 effective mAs per patient. The timing of the portal venous phase scan was optimized using the fixed time-delay technique of 90 seconds after contrast administration.

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.

All LD CT images were reconstructed using both FBP and ADMIRE methods and LD-DLA images were obtained from LD-FBP images using DLAs without raw projection data.

Phantom Studies

We performed CT scans of the American College of Radiology (ACR) CT accreditation phantom (Gammex 464, Sun Nuclear, Middleton, WI, USA) to quantitatively evaluate image quality. The phantom studies were also conducted under the same scan protocol with fixed current-time values (50 mAs or 200 mAs to approximate a 25% or 100% dose, respectively) without dose modulation.

To assess the objective image quality, the NPS and MTF were calculated using the ACR phantom. Because of the ethical issues of performing CT scans at multiple doses in a single patient, quantitative image analysis was performed using phantoms. NPS is a comprehensive metric of noise

properties. The NPS curve shows the variance in image intensity over spatial frequencies in an image. The mean image noise defined as the standard deviation of the CT attenuation value within a uniform region of interest (ROI) is correlated with the area under the curve (AUC) of the NPS curve and the image texture is determined by the shape of the NPS curve. The higher the AUC of the NPS curve, the greater the image noise. Reduced peak frequency of the NPS (left-shifted curve) represents a smoother and plastic-like image texture (20). MTF is a representative metric of spatial resolution. The MTF curve provides a measure of how well the system transfers the contrast ratio of the original object (y-axis) across spatial frequencies (x-axis) (21). If the system has perfect spatial resolution, the object will be able to retain 100% contrast of its original contrast in any spatial frequency. However, in practice, the system is not perfect; therefore, the object's contrast is more attenuated, and it is difficult to distinguish the object at higher spatial resolutions. Therefore, a higher MTF value means better spatial resolution.

To measure the NPS, we used the method proposed in a previous study (22). The peak frequency and AUC of the NPS were calculated. We used the circular edge method for MTF measurements, based on a study by Friedman et al. (22), because conventional methods could not reflect the spatial resolution dependence of the image contrast in the IR algorithm. The ACR phantom contains three inserts and one air cavity for the measurement of Hounsfield unit accuracy. We measured the MTF in the three inserts using different contrasts, which provided nominal Hounsfield unit values of -95, 955, and 120 at 120 kVp for polyethylene (Disc 1), bone (Disc 2), and acrylic (Disc 3), respectively. We obtained the average MTF value after 21 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 NPS and MTF of ACR phantom images were calculated using MATLAB.

In addition, the detailed method of another phantom study for assessing the training performance of the DLA and results of the training performance are described in the Supplementary Materials.

Patient Study

To objectively assess image quality in the clinical environment, a 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 from 18 patients. The mean image noise was measured by manually placing oval ROIs (range, 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 location for the LD-DLA, LD-FBP, and LD-ADMIRE images using a copy-and-paste function.

Qualitative image analysis of image noise, spatial resolution, and overall image quality was performed. Another abdominal radiologist with 25 years of experience independently evaluated the 90 datasets of the portal phase abdomen CT images (five different CT reconstructions or processed methods, i.e., DLA-1, DLA-2, DLA-3, FBP, and ADMIRE, from each of the 18 patients).

The five image sets from each patient were ranked against one another on a comparative scale from 1 to 5 with regard to image noise, spatial resolution, and overall image quality, and average ranks were compared. A score of 1 was given to the images with the best quality and 5 was given to the images with the worst quality. Reconstruction or processed methods were blinded and the order of image sets was randomized for each patient.

The presence of additional artifacts that were not noted on LD-FBP and LD-ADMIRE CT images was also evaluated in the LD-DLA images of patients by a different abdominal radiologist with 8 years of experience. For the evaluation of the additional artifacts, reconstruction or processed methods were not blinded and LD-DLA images were compared with paired LD-FBP and LD-ADMIRE images. All noise measurements and qualitative image evaluations 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, IBM Corp., Armonk, NY, USA).

A paired t test was used to compare the MTF values of phantom images. Repeated measures analysis of variance followed by pairwise comparisons was used to compare image noise levels in the 18 patients' images with different reconstruction methods and DLAs. After Bonferroni adjustment for the ten comparisons (all pairwise comparisons for FBP, ADMIRE, and DLA-1, -2, and -3.), a p value < 0.005 was considered significant. The Friedman test was used to evaluate the qualitative image analysis. Dunn's pairwise post

hoc test was performed for multiple pairwise comparisons, then a Bonferroni correction for multiple testing was applied. The modified Wald method was used to determine the confidence interval (CI) for the presence of image artifacts in the 2998 LD CT image slices of 18 patients.

RESULTS

The baseline patient characteristics and radiation dose of the training and test set are shown in Table 1. The age of patients (mean ± standard deviation) was 63.5 ± 13.0 years in the training population and 35.0 ± 12.2 years in the test population. Because LD abdominal CT was performed in adults under 45 years of age in this institution, the test population was younger than the training population. The mean volume CT dose index (CTDIvol), dose-length product, and peak mAs in the test population were approximately one-quarter of those in the training population.

Phantom Studies

LD-DLA achieved a 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, the peak frequency and AUC of the NPS curves in the LD-DLA images were lower than those of the LD-FBP and LD-ADMIRE images at all dose levels. In addition, the NPS curves were shifted towards lower spatial frequencies in the LD-DLA images compared with those of the ADMIRE images.

Table 3 shows the mean MTF-50 value of each reconstructed CT image for the three different disc objects of the phantom CT with a 25% radiation dose level. The MTF-50 was significantly lower in the LD-DLA images than in the LD-ADMIRE and LD-FBP images (all $p < 0.001$).

Figure 3 shows the MTF for the DLA, ADMIRE, and FBP methods acquired at the 25% dose level for various contrast

inserts. When the training radiation dose decreased, the AUC of the NPS curve and MTF values of the LD-DLA images decreased. DLA-1 showed the least noise and worst spatial resolution among the three DLA methods.

Patient Study

In the 18 patients, the mean image noise in all the LD-DLA images was significantly lower than that in the LD-ADMIRE and LD-FBP images (all $p < 0.001$). Table 4 shows the mean
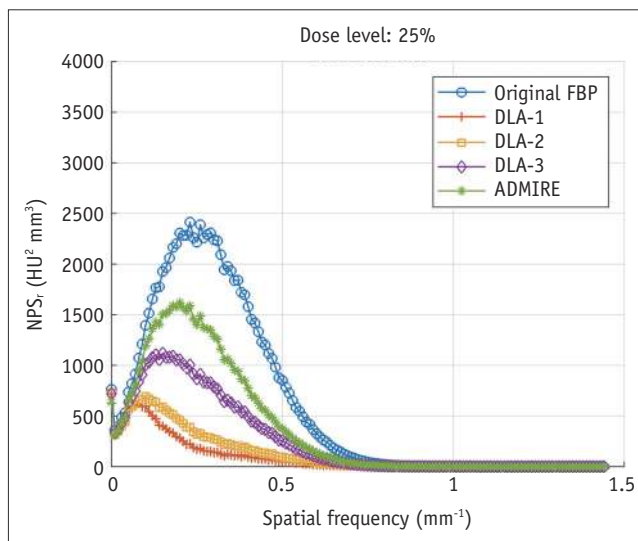
Fig. 2. NPS of 25% dose CT of phantom according to reconstruction method. NPS curves are shifted towards lower spatial frequencies in images produced by DLAs trained at lower radiation dose level. HU = Hounsfield unit, NPS = noise power spectrum

Table 2. AUC and Peak Frequency of NPS Curve in Each Reconstruction Method

NPS Property	DLA-1	DLA-2	DLA-3	ADMIRE	FBP
AUC (x 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

ADMIRE = advanced modeled iterative reconstruction, AUC = area under curve, DLA = deep learning-based denoising algorithm, FBP = filtered back projection, HU = Hounsfield unit, NPS = noise power spectrum

Table 1. Baseline Characteristics of Study Population

Charecteristics	Training Population (n = 100)	Test Population (n = 18)
Demographics		
Age (year)	63.5 ± 13.0	35.0 ± 12.2
Body mass index	22.9 ± 3.6	23.0 ± 2.5
Radiation dose		
CTDIvol (mGy)	6.5 ± 1.3	1.4 ± 0.3
DLP (mGycm)	312.8 ± 74.3	72.1 ± 20.4
Peak tube current (mAs)	240.1 ± 58.6	52.8 ± 9.2

Data are mean ± standard deviation. CTDIvol = volume CT dose index, DLP = dose-length product

Table 3. MTF-50s (mm^{-1}) of 25% ACR Phantom CT according to Different Discs and Reconstruction Methods

Reconstruction Method	Disc 1 (Polyethylene)	Disc 2 (Bone)	Disc 3 (Acrylic)	$P*$
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 comparisons. ACR = American College of Radiology, MTF = modulation transfer function

image noise values according to the image reconstruction method. As the training radiation dose of DLA decreased, the mean image noise also significantly decreased ($p < 0.001$). Representative images are shown in Figure 4.

Table 5 shows the mean ranking scores for image quality according to the image reconstruction method. In the qualitative image analysis, the overall image quality was the best for DLA-3, although it was not significantly different from that of ADMIRE ($p > 0.999$). Although DLA-1 and DLA-2 had lower noise than the other methods, they provided worse spatial resolution and lower overall image quality relative to the other methods.

The DLA images showed no additional image artifacts in the 2998 image slices of each DLA (0%, 95% CI, 0–4 slices).

DISCUSSION

We found that DLA can achieve lower image noise than ADMIRE, which is a state-of-the-art commercially available IR method. Furthermore, there are several strengths of the DLA used in our study. First, while the IR algorithm is vendor-specific and has limited applicability to other CT systems, FBP-based DLAs can be used in a more general way because they allow image-based reconstruction without raw data in the FBP method (7). Second, DLAs require shorter

computational time compared with IR (14). Although a direct comparison was difficult in our experiment because we could not test the ADMIRE method in the same computing environment as that processing the DLAs, the DLAs might potentially reduce computing load relative to the ADMIRE method. Finally, we evaluated DLA performance using quarter dose CT images, which is a frequently used radiation dose in clinical LD CT (23, 24). To the best of our knowledge, this is the first study suggesting alternative approaches for CT denoising in a clinical setting, specifically comparing DLAs with the FBP and IR algorithms in real low-dose CT images of patients.

In the test study with 18 patients, we were able to test the DLAs with LD CT in a real clinical environment. We found that DLAs effectively reduced image noise in the test study and did not show any additional image artifacts. Due to several negative aspects of deep learning techniques, such as their "black-box" nature and the complexity of neural networks, we are concerned that the processed images might show unknown artifacts. However, there were no deep learning-related image artifacts in approximately 3000 images. Therefore, we cautiously concluded that artifacts are highly unlikely under these test conditions. However, additional evaluations will be required under different conditions (e.g., images with beam hardening
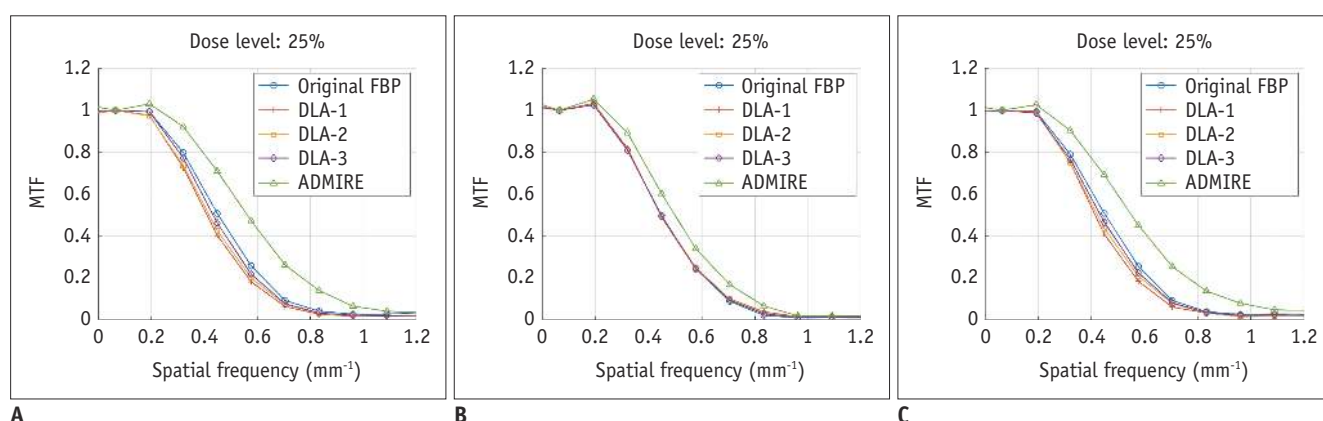

Fig. 3. Comparison of MTF with five different CT reconstruction or processing methods for three different discs.
A. Polyethylene. **B.** Bone. **C.** Acrylic. MTF = modulation transfer function

Table 4. Mean Image Noise (HU) according to Image Reconstruction Method

Measuring Site	Reconstruction Method					
	DLA-1	DLA-2	DLA-3	ADMIRE	FBP	P^*
Liver	6.3 (1.2)	8.5 (1.6)	11.4 (2.0)	16.2 (1.6)	23.0 (2.5)	
Subcutaneous fat	6.0 (1.2)	7.1 (1.5)	9.9 (1.5)	13.2 (1.3)	20.4 (2.5)	
Paraspinal muscle	5.3 (0.9)	6.6 (1.3)	9.9 (1.7)	14.5 (1.1)	22.0 (3.0)	< 0.001
Abdominal aorta	6.3 (1.1)	7.7 (1.3)	11.1 (2.4)	15.5 (2.0)	21.3 (2.3)	

Data are mean noise (standard deviation). *p value for all pairwise comparisons.
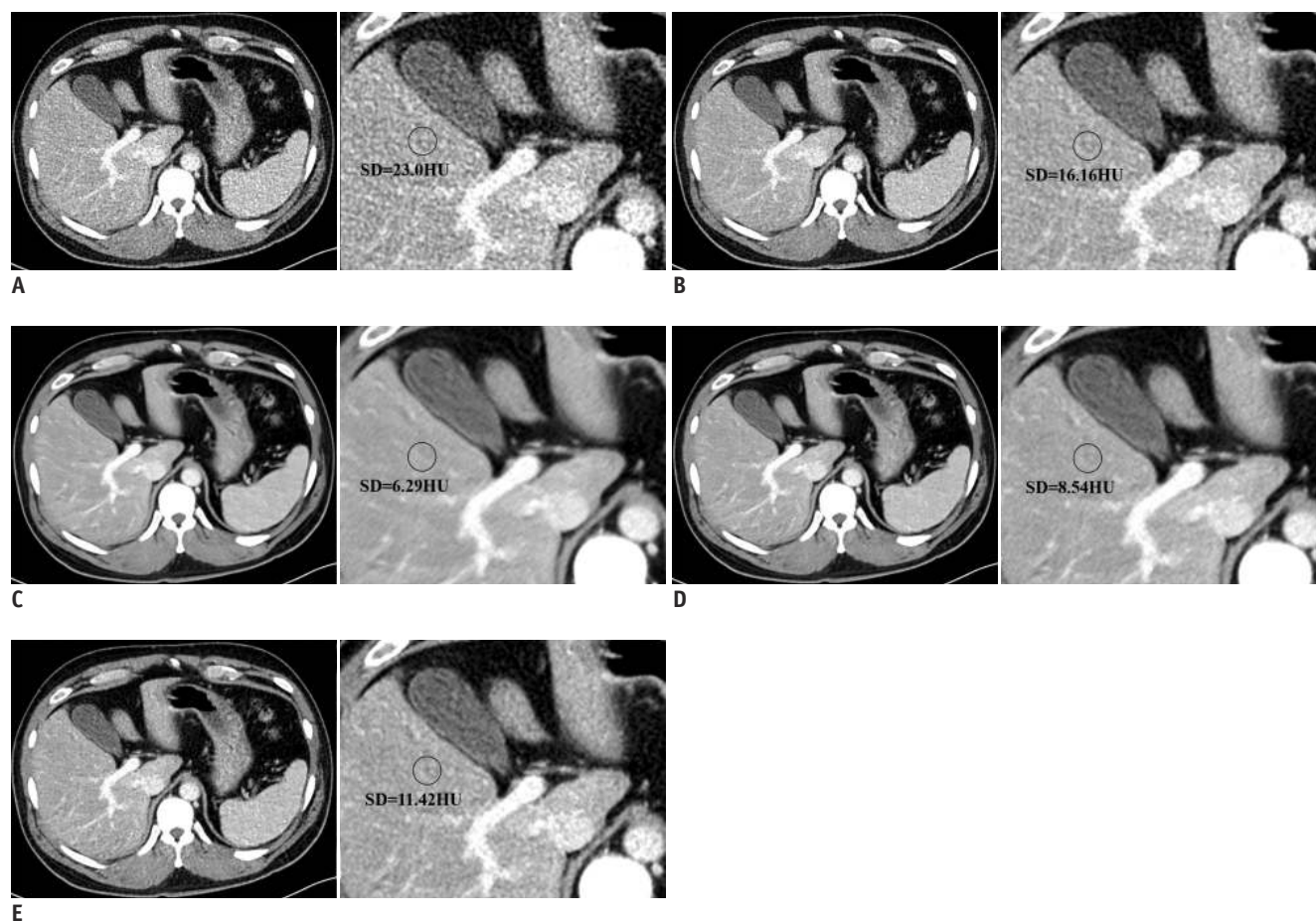
Fig. 4. LD abdominal CT images of test set with conventional reconstruction methods (A, B) and with DLA (C-E). A. FBP. **B.** ADMIRE. **C.** DLA-1. **D.** DLA-2. **E.** DLA-3. First column of each image shows LD (25%) abdominal CT using five different methods and enlarged image of first column is in second column. Mean image noise of all LD-DLA images was lower than that of LD-ADMIRE and LD-FBP images, and DLA-1 image showed lowest mean image noise. As training radiation dose of DLA increased, mean image noise of processed CT images increased.

Table 5. Mean Image Quality Assessment Ranking of Image Reconstruction Methods

Reconstruction Method	DLA-1	DLA-2	DLA-3	ADMIRE	FBP
Image noise	1.1 (0.2)*†	1.9 (0.2)*†	3.2 (0.4)†	3.8 (0.4)	5.0 (0)
Spatial resolution	4.8 (0.5)*†	4.0 (0.3)*	1.8 (0.7)	1.8 (1.1)	2.6 (0.9)
Overall image quality	4.6 (0.6)*	3.7 (0.7)*	1.4 (0.5)†	1.7 (0.7)	3.7 (0.9)

Data are mean ranking score (standard deviation). *$p < 0.05$ in pairwise comparison with ADMIRE, †$p < 0.05$ in pairwise comparison with FBP.

artifacts, images in elderly patients with motion artifacts, or images from other parts of the body) that we did not consider during either training or testing.

In this study, we compared the image quality of LD-DLA CT with that of LD-FBP and LD-ADMIRE CT using quantitative metrics such as NPS, MTF, and mean image noise. We found that 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, the AUC of the NPS curve, MTF values, and mean image noise level of the LD-DLA images decreased as DLA training

radiation dose decreased. This tendency is difficult to explain but it might be related to the intrinsic limitation of deep learning, which requires proper training intensity.

In terms of noise properties, the processed phantom images using DLA were less noisy than FBP. However, DLA images showed an over-smooth texture, just like those produced by the first-generation IR algorithm, which is currently used to lower the radiation dose by lowering the image noise with better spatial resolution (7). Because CT images reconstructed using IR algorithms can produce plastic-like texture, radiologists who are familiar with CT

images reconstructed using FBP might be unfamiliar to those images and IR has its weakness in this regard (7, 9). Since we trained the DLAs with both LD-FBP and RD-FBP images, we assumed that DLA could maintain the image texture of the FBP method in the resultant images. However, the NPS curve was left-shifted in the DLA images, i.e, noise texture of DLAs was different from FBP and rather similar to that of ADMIRE. It is difficult to find a convincing explanation of these results, and further investigations are needed.

Conventional MTF measurements used highly dense materials and very low noise to improve measurement precision, and MTF tended to be greater in IR than in FBP for high-contrast objects (6). However, in real clinical settings, most CT images have lower contrast and higher noise than in an ideal environment. In fact, Richard et al (21). showed how the MTF in IR could be different for varying contrast and noise levels. Although the task-based MTF for model-based IR was shown to be greater than for FBP in high-dose and high-contrast objects, the MTF performance decreased in LD and low-contrast objects (21). Therefore, we evaluated the MTF for three different contrast materials in the phantom study. Our results showed a little loss of spatial resolution by DLAs for all disc objects, but the difference from FBP was marginal. Although the MTF increased as the training radiation dose increased, it may be difficult to go beyond the limits of FBP because the DLA training was based on RD-FBP CT images. If the DLAs were trained using CT images with improved spatial resolution, the MTF could be further improved in the DLA compared with that of FBP-based DLAs.

Our qualitative image analysis revealed that DLA-3, which trained with 50% radiation dose images, showed the best overall image quality among the five image sets (DLA-1, DLA-2, DLA-3, FBP, and ADMIRE). Interestingly, the spatial resolution of DLA-3 was not significantly different from that of ADMIRE, which showed the highest MTF values, although the quantitative MTF values of DLA-3 were significantly lower than those of conventional methods (FBP and ADMIRE). Overall image quality is a complex combination of image noise and spatial resolution. Indeed, a previous study assessed image quality using the detectability index, which is a function of the NPS, MTF, and other factors (25). In this context, the combination of low noise and low loss of spatial resolution is likely to result in a favorable outcome for DLA-3. On the contrary, although DLA-1 and DLA-2 had less noise, the overall image quality was not good because of a large loss of spatial resolution.

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 measurements, the noise texture of the DLA image can be changed, affecting the diagnostic performance of DLA, such as in lesion detection. Because this is a preliminary study aiming to investigate the performance of DLAs in image noise and spatial resolution, future studies are needed to confirm this exploratory result as well as to determine the effects of DLAs on task-based performance. Second, because the DLA used in this study was trained using simulated LD CT images, we could not expect how the results would differ using actual LD CT images as training data. In addition, we did not evaluate the effects of image artifacts because we included only qualified images without significant artifacts in the training set. Finally, because we only included adult patients under the age of 45 and obtained quarter dose abdominal CT images, further studies using various LD CT protocols in different body parts may be required.

In conclusion, DLAs could improve image quality in terms of noise reduction in CT. Although DLAs achieved less noise than FBP and ADMIRE in LD CT images, they did not maintain the spatial resolution. Regarding the effect of the radiation dose in training conditions, a 50% simulated radiation dose for training images led to the best overall image quality.

Supplementary Materials

The Data Supplement is available with this article at https://doi.org/10.3348/kjr.2019.0413.

ORCID iDs
Won Chang
　https://orcid.org/0000-0001-7367-9841
Yoon Joo Shin
　https://orcid.org/0000-0001-7872-5552
Jong Chul Ye
　https://orcid.org/0000-0001-9763-9609

Eunhee Kang
   https://orcid.org/0000-0001-9921-8103
Dong Yul Oh
   https://orcid.org/0000-0002-8393-7960
Yoon Jin Lee
   https://orcid.org/0000-0002-3572-029X
Ji Hoon Park
   https://orcid.org/0000-0002-6794-4909
Young Hoon Kim
   https://orcid.org/0000-0001-5554-3828

REFERENCES

1. Smith-Bindman R, Lipson J, Marcus R, Kim KP, Mahesh M, Gould R, et al. Radiation dose associated with common computed tomography examinations and the associated lifetime attributable risk of cancer. *Arch Intern Med* 2009;169:2078-2086

2. Brenner DJ, Hall EJ. Computed tomography — An increasing source of radiation exposure. *N Engl J Med* 2007;357:2277-2284

3. Berrington de González A, Mahesh M, Kim KP, Bhargavan M, Lewis R, Mettler F, et al. Projected cancer risks from computed tomographic scans performed in the United States in 2007. *Arch Intern Med* 2009;169:2071-2077

4. Hara AK, Paden RG, Silva AC, Kujak JL, Lawder HJ, Pavlicek W. Iterative reconstruction technique for reducing body radiation dose at CT: feasibility study. *AJR Am J Roentgenol* 2009;193:764-771

5. Nakayama Y, Awai K, Funama Y, Hatemura M, Imuta M, Nakaura T, et al. Abdominal CT with low tube voltage: preliminary observations about radiation dose, contrast enhancement, image quality, and noise. *Radiology* 2005;237:945-951

6. Sagara Y, Hara AK, Pavlicek W, Silva AC, Paden RG, Wu Q. Abdominal CT: comparison of low-dose CT with adaptive statistical iterative reconstruction and routine-dose CT with filtered back projection in 53 patients. *AJR Am J Roentgenol* 2010;195:713-719

7. Geyer LL, Schoepf UJ, Meinel FG, Nance JW Jr, Bastarrika G, Leipsic JA, et al. State of the art: iterative CT reconstruction techniques. *Radiology* 2015;276:339-357

8. Holmquist F, Nyman U, Siemund R, Geijer M, Söderberg M. Impact of iterative reconstructions on image noise and low-contrast object detection in low kVp simulated abdominal CT: a phantom study. *Acta Radiol* 2016;57:1079-1088

9. Prakash P, Kalra MK, Kambadakone AK, Pien H, Hsieh J, Blake MA, et al. Reducing abdominal CT radiation dose with adaptive statistical iterative reconstruction technique. *Invest Radiol* 2010;45:202-210

10. Yu L, Liu X, Leng S, Kofler JM, Ramirez-Giraldo JC, Qu M, et al. Radiation dose reduction in computed tomography: techniques and future perspective. *Imaging Med* 2009;1:65-84

11. Jain V, Seung S. *Natural image denoising with convolutional networks*. 23rd annual conference on neural information processing systems 22;2009 December 7-10;Vancouver, Canada

12. Nasri M, Nezamabadi-pour H. Image denoising in the wavelet domain using a new adaptive thresholding function. *Neurocomputing* 2009;72:1012-1025

13. Xie J, Xu L, Chen E. *Image denoising and inpainting with deep neural networks*. 26th annual conference on neural information processing systems 25;2012 December 3-6;Lake Tahoe, NV, USA

14. Kang E, Min J, Ye JC. A deep convolutional neural network using directional wavelets for low-dose X-ray CT reconstruction. *Med Phys* 2017;44:e360-e375

15. Chen H, Zhang Y, Zhang W, Liao P, Li K, Zhou J, et al. Low-dose CT via convolutional neural network. *Biomed Opt Express* 2017;8:679-694

16. McCollough C. TU-FG-207A-04: overview of the low dose CT grand challenge. *Med Phys* 2016;43(Part 35):3759-3760

17. Chen H, Zhang Y, Kalra MK, Lin F, Chen Y, Liao P, et al. Low-dose CT with a residual encoder-decoder convolutional neural network. *IEEE Trans Med Imaging* 2017;36:2524-2535

18. Kang E, Chang W, Yoo J, Ye JC. Deep convolutional framelet denosing for low-dose CT via wavelet residual network. *IEEE Trans Med Imaging* 2018;37:1358-1369

19. Ellmann S, Kammerer F, Brand M, Allmendinger T, May MS, Uder 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. *Invest Radiol* 2016;51:331-339

20. Ehman EC, Yu L, Manduca A, Hara AK, Shiung MM, Jondal D, et al. Methods for clinical evaluation of noise reduction techniques in abdominopelvic CT. *Radiographics* 2014;34:849-862

21. Richard S, Husarik DB, Yadava G, Murphy SN, Samei E. Towards task-based assessment of CT performance: system and object MTF across different reconstruction algorithms. *Med Phys* 2012;39:4115-4122

22. Friedman SN, Fung GS, Siewerdsen JH, Tsui BM. 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. *Med Phys* 2013;40:051907

23. Kim K, Kim YH, Kim SY, Kim S, Lee YJ, Kim KP, et al. Low-dose abdominal CT for evaluating suspected appendicitis. *N Engl J Med* 2012;366:1596-1605

24. Kim SY, Lee KH, Kim K, Kim TY, Lee HS, Hwang SS, et al. Acute appendicitis in young adults: low- versus standard-radiation-dose contrast-enhanced abdominal CT for diagnosis. *Radiology* 2011;260:437-445

25. Christianson O, Chen JJ, Yang Z, Saiprasad G, Dima A, Filliben JJ, et al. An improved index of image quality for task-based performance of CT iterative reconstruction across three commercial implementations. *Radiology* 2015;275:725-734