임상의과학석사 학위논문

딥러닝 기반 잡음제거 알고리즘을 적용한 전산화단층촬영 영상에서의 저대조도검출률

Low-Contrast Detectability of Image-based Denoising Algorithm using Deep Learning in CT

2020 년 2 월

서울대학교 대학원

임상의과학과 석사 과정

김영준

Abstract

Low-Contrast Detectability of Image-based Denoising Algorithm using Deep Learning in CT

Youngjune Kim

Collage of Medicine, Department of Clinical Medical Sciences
The Graduate School
Seoul National University

Purpose: To compare the low-contrast detectability of deep learning algorithm (DLA), with those of advanced modeled iterative reconstruction (ADMIRE) and filtered back projection (FBP).

Materials and Methods: Using abdomen and pelvis CT images of 100 patients reconstructed with both ADMIRE and FBP, we trained DLA by feeding FBP images as input and ADMIRE images as the ground truth. To compare the low-contrast detectability of the DLA with that of ADMIRE and FBP, randomized repeat scans of Catphan® low-contrast phantom module were performed under various conditions of radiation exposures, and twelve radiologists evaluated the presence/absence of a target on a five-point confidence scale.

Multi-reader multi-case area under the receiver operating characteristic curve (AUC) was calculated and noninferiority tests were performed. To evaluate the spatial resolution and noise, the target transfer function (TTF50), area under the noise power spectrum (AUNPS) and noise power spectrum (NPS) peak were compared across the algorithms.

Results: The AUC of DLA in low contrast detectability was noninferior to that of ADMIRE (P <.001) and superior to that of FBP (P <.001). TTF50 of DLA was higher than that of FBP and ADMIRE in various conditions. AUNPS of DLA was lower than that of ADMIRE and FBP (P <.001). The NPS peak frequency of DLA was not different from that of ADMIRE (P >.99) and was lower than that of FBP (P <.001).

Conclusion: The low-contrast detectability of the proposed deep learning-based denoising algorithm was noninferior to that of ADMIRE and superior to that of FBP.

Keywords: Deep learning, Low-contrast detectability, Human observer study, Noise reduction, Computed tomography

Student Number: 2018-20093

Table of contents

List of tables

List of figures

Introduction

The utilization of CT examinations in the medical field is increasing annually. In the United States, 245 CT examinations were performed per 1000 people in 2016 [1]. Although controversial, there are concerns that CT examination may cause carcinogenesis [2, 3]. Correspondingly, there are ongoing attempts for CT dose optimization. While the simplest way to reduce CT dose is to reduce the tube voltage current, denoising reconstruction algorithms such as iterative reconstruction (IR) can also be used to reduce CT dose.

The most popular reconstruction algorithms for CT examinations are filtered back projection (FBP) and IR. Efforts have been made to reduce the radiation dose, especially using IR, while maintaining the diagnostic performance. Recently, model-based iterative reconstruction (MBIR), which utilizes a statistical model for the raw projection data and image data and a system model for the forward projection, was developed and applied in clinical practice [4]. Several studies reported that MBIR could reduce the radiation dose while maintaining diagnostic performance [5-8]. However, MBIR has some limitations as it requires heavy computational power and longer

reconstruction time. As a result, hybrid IR techniques such as Advanced modeled iterative reconstruction (ADMIRE, Siemens Healthcare, Forchheim, Germany) are gaining their popularity offering the shorter reconstruction time and comparable image quality to MBIR.

Along with the developments in the deep learning technology, several attempts have been made recently to reduce CT noise using deep learning. Deep learning algorithms have been used to reduce the noise of CT images either in the image reconstruction domain [9–11] or the image post-processing domain [12–16]. They may also contribute to reducing the time spent on noise reduction while maintaining or improving the image quality in comparison to the commercially available IR [14]. Recently, Advanced intelligent Clear-IQ Engine (AiCE, Canon Medical Systems, Otawara, Japan) and TrueFidelity (GE Healthcare, Waukesha, WI), deep learning-based reconstruction algorithms approved by the U.S. Food and Drug Administration, was developed for commercial purposes. However, no previous studies have focused on the performance of human observers in the images denoised with deep learning algorithm in comparison to those reconstructed with commercially available MBIR

or FBP.

This study aimed to compare the low-contrast detectability of deep learning-based denoising algorithm (DLA) with that of ADMIRE and FBP.

Materials and Methods

This retrospective study was conducted in a single tertiary urban hospital. Our institutional review board waived the patient informed consent regarding the use of axial images of the abdomen and pelvis CT of 100 patients.

Deep learning algorithm and CT images

We used the DLA introduced by Kang et al. [12]. To train the DLA, we used the abdomen and pelvis CT images of 100 patients (age in mean ± standard deviation, 63.5 ± 13.0 years; 42 females), which were examined from August 2017 to January 2018 and reconstructed using both ADMIRE and FBP. We trained the DLA by feeding images reconstructed using FBP and ADMIRE as the input and ground truth, respectively. None of the patients had metallic hardware in their CT images. The CT protocol used in these patients has been summarized in Table 1.

Human observer study

To compare the reader's performance in detecting low-contrast targets in DLA, ADMIRE, and FBP, we performed randomized repeat scans of Catphan® low-contrast phantom module (CTP 515) under various conditions of radiation exposures (100 kVp; 200, 100, 50, 26 mAs) using a single CT machine (Somatom Definition Edge, Siemens Healthcare, Erlangen, Germany) (Table 1). Among the various targets included in the phantom module, we used the 9 and 5 mm supra-slice targets with +10 Hounsfield unit difference to the background. All images were reconstructed with ADMIRE and FBP. We then denoised the images produced with FBP using the DLA trained earlier. We cropped the images with dimensions 5 × 5 cm2 to ensure that the target would either be absent in the image or present in random locations instead of only in the center. Next, we manually replaced the pixels of target of sizes other than 5 or 9 mm with the target-absent image patches. A total of 960 images were produced (40 images [20 images with target and 20 images without target] × 3 different algorithms × 4 different radiation doses × 2 different target sizes).

A graphical user interface (GUI) was designed using MATLAB (Mathworks, Natick, Mass), as shown in Fig. 1. Default display settings were set to a window level of 70, which was the mean Hounsfield unit value of the image background, and a window width of 100. All images were randomly arrayed, the even-numbered images were displayed on the left side of the screen, and the odd-numbered images were displayed on the right side of the screen to eliminate potential bias due to change detection in flicker paradigm [17-19]. The background of the GUI was black because most picture archiving and communication system (PACS) programs use a black background.

Twelve radiologists with various experiences in radiology (six attending radiologists from three different institutions with clinical experience of 6–24 years in abdominal radiology [including three co-authors: W.C., J.H.P., Y.H.K.] and six radiologists in training from a single institution [including one co-author: Y.K.]) evaluated the presence or absence of target on a five-point confidence scale (1: definitely absent, 2: probably absent, 3: indeterminate, 4: probably present, and 5: definitely present). First, the readers underwent two sessions of tutorials, each of which

consisted of ten questions and received instant feedbacks with the answers. After the tutorials, they each reviewed the 960 images independently. The radiologists were asked to use a display calibrated to the DICOM standard and minimize the reading room light as low as possible. Their response time for each image was automatically measured in seconds.

Physical measurement

We used the target transfer function (TTF) to compare the spatial resolution of the algorithms and noise power spectrum (NPS) to compare the distribution of the noise [20]. We performed 21 repeated scans using the American College of Radiology CT accreditation phantom module 1 under the condition of 100 kVp and 200 mAs. Other scan parameters were the same as those listed in Table 1. All images were reconstructed with ADMIRE and FBP. The DLA denoised the images produced with FBP. The TTF and NPS of four objects (disc 1, polyethylene; disc 2, bone; disc 3, acrylic; and disc 4, air) were calculated using the method proposed by Friedman et al. [21]. The mean Hounsfield unit of each disc was −98, 980, 120,

and −990, respectively, under our radiation dose condition. All physical measurements were performed by a medical physicist (D.O.) with six years of experience with a MATLAB program (MathWorks, Natick, Mass).

Statistical analysis

We performed a pilot study to calculate the required number of readers and cases. Two readers (Y.K., a fourth−year radiologist in training and W.C., an attending radiologist with nine years of clinical experience in abdominal radiology) reviewed the same dataset for calculating the sample size. We hypothesized that the low contrast detectability of DLA would be noninferior to that of ADMIRE and superior to that of FBP. We set the noninferiority margin to be −0.10, accounting for heterogeneous CT dose settings, a large number of readers, and high interobserver variability due to the various experience levels of readers. We set the superiority margin to be zero. Based on the pilot study results, 12 readers with 960 cases in total (320 cases for each algorithm) were required to ensure the significance level of 0.05 and statistical power of 80% in the

noninferiority test between the DLA and ADMIRE and superiority test between the DLA and FBP [22, 23].

Our primary endpoint was to compare the pooled area under the receiver operating characteristic curve (AUC) of human observers using each algorithm irrespective of radiation dose or target size. For such purpose, a multi-reader multi-case area under the receiver operating characteristic curve (AUC) was calculated. Noninferiority tests using Wald statistics were performed to compare the DLA with ADMIRE and FBP. We compared the TTF and NPS of the algorithms in terms of TTF50, area under the NPS (AUNPS), and NPS peak frequency, using paired t-tests. We then tested whether the noninferiority/superiority of the DLA in terms of AUC was consistently noted across various radiation doses and target sizes. We compared the mean response time between the three algorithms using paired t-tests. We used Holm's correction to correct the familywise type I error in the subgroup analysis. All statistical analyses were performed using iMRMC, software version 4.0.0 (Division of Imaging, Diagnostics, and Software Reliability, OSEL/CDRH/FDA, Silver Spring, MD), and R, version 3.5.2 (The R Foundation for Statistical Computing, Vienna, Austria). The P-value

< 0.05 was considered statistically significant.

Results

Human observer study

The AUC of the DLA in detecting low-contrast targets was noninferior to that of ADMIRE (AUC difference [95% confidence interval]: −0.013 [−0.038–0.012]; P < 0.001) and superior to that of FBP (0.039 [0.017–0.060], P < 0.001) (Table 2). Using the 9 mm target, the AUC of the DLA was noninferior to that of ADMIRE irrespective of the radiation doses. It was also superior to that of FBP under the reference tube current-time product of 200, 100, and 50 mAs, and noninferior under 26 mAs. Using the 5 mm target, the AUC of the DLA was noninferior to that of ADMIRE under the reference tube current-time product of 100, 50, and 26 mAs. It was noninferior to that of FBP in all radiation doses.

The mean response times ± standard deviations of DLA, ADMIRE, and FBP were 2.67 ± 2.45, 2.73 ± 2.56, and 2.57 ± 2.44 s, respectively. No significant difference was noted considering the response time for the DLA vs. ADMIRE (P = 0.26) and DLA vs. FBP (P = 0.068). The response time was shorter in the target-present

images than the target-absent images, and when a target size of 9 mm or a reference tube current-time product of 200 mAs was used (Table 3).

Physical measurement

TTF50, AUNPS, and NPS peak frequency of the four objects using the three algorithms are summarized in Table 4. TTF50 of DLA was significantly higher than that of FBP and ADMIRE, except when compared to ADMIRE using the disc simulating bone (the average TTF50 of DLA vs. that of ADMIRE: 0.484 vs. 0.487; $P = 0.002$) (Fig. 3). The AUNPS of the DLA was significantly lower than that of ADMIRE and FBP ($P < 0.001$). The NPS peak frequency of DLA was not different from that of ADMIRE ($P > 0.99$) and was significantly lower than that of FBP ($P < 0.001$) (Fig. 4).

Discussion

We developed an image-based denoising deep learning algorithm and tested it under various conditions of radiation doses and target sizes. While most iterative reconstruction algorithms are vendor-specific, the proposed denoising deep learning algorithm is a vendor-neutral algorithm. Although previous papers have already shown the potential of deep learning in imaging denoising [9-16], our study is distinguished from those studies that merely focused on physical measurements. We performed physical measurements as well as compared the radiologists' diagnostic performance using the proposed DLA vs. other conventional reconstruction algorithms. We showed that our DLA could denoise the CT images and improve the radiologists' diagnostic performance in low-contrast detectability. Our DLA showed that the pooled AUC for low-contrast detectability was comparable to that of ADMIRE and superior to that of FBP. In subgroup analysis, this trend was consistently observed irrespective of the target size and dose level.

In contrast to the results of our study, a previous study suggested that the low-contrast detectability in the images

reconstructed with ADMIRE was not significantly different from that in the images reconstructed with FBP [5]. This is probably due to several factors as follows. First, our GUI was different from the one used in the previous study. We intentionally cropped the images to ensure that the targets would appear in random locations instead of in the center of the PACS monitor. This was done to reflect the real practice where radiologists have to detect lesions that appear randomly (e.g., hepatic metastasis) instead of at a certain location, constantly. Furthermore, all images were consecutively numbered after being randomly arrayed, after which the even-numbered images were displayed on the left side of the screen while the odd-numbered images were displayed on the right side of the screen. This was done to eliminate the potential bias regarding change detection in flicker paradigm [17, 18]. Second, we used two different sizes of targets: 5 mm and 9 mm. Previous studies used targets that were usually around 10 mm, whereas in practice we often encounter lesions of smaller size that have subtle attenuation differences to the background. Third, the radiation dose used in this study was much lower than that in previous studies [5, 24], with a tube potential of 100 kVp and reference tube current-time product of 200 mAs. In the

liver, where low contrast detectability is particularly important, it was suggested that a tube potential of 100 kVp should be used instead of 120 kVp to increase the contrast of lesions [25–27]. According to Kanal et al. [28], the achievable dose of contrast–enhanced abdomen and pelvis CT in patients with a body size of our phantom (20 cm) is approximately 7 mSv, with the diagnostic reference level of approximately 9 mSv in the United States. To achieve such dose levels, we adjusted our reference tube current–time product to 200 mAs. Finally, we informed the readers in advance that we were going to measure their response time, which seemingly encouraged the radiologists to read the images more quickly.

There was no significant difference in the mean response time across the algorithms. Radiologists did not have any difficulty in adjusting to image texture and judging target presence using the images denoised with DLA. The response time was instead prolonged when the target size was smaller, radiation dose was lower, and when the target was absent in the given image.

Several deep learning algorithms that could decrease the image noise are being introduced in the medical field. The AUTOMAP, which is representative of such algorithms [29], outperformed the

conventional reconstruction methods in terms of noise reduction as well as motion artifact reduction. Our DLA also outperformed ADMIRE in terms of several physical measurements, such as TTF50 and AUNPS.

In contrast to the results of our study, however, a previous study that utilized the same deep learning architecture as ours showed a decrement in spatial resolution in the CT images denoised by the deep learning algorithm [13]. However, the deep learning algorithm used in that study was trained with simulated low-dose FBP images as the input images and full-dose FBP images as the ground truth images. We could improve the spatial resolution while reducing noise by using full-dose FBP images as the input and full-dose ADMIRE images as the ground truth.

The proposed deep learning denoising algorithm could reduce high spatial frequency noise more effectively than low spatial frequency noise, showing an NPS similar to that of ADMIRE. It has been known that the noise difference between MBIR and FBP is relatively smaller at lower spatial frequencies than at higher spatial frequencies [30]. Our deep learning algorithm mimicked ADMIRE in terms of the pattern of noise but it could not reduce the low spatial

frequency noise as effectively as high spatial frequency noise.

There were several limitations in our study. First, we only tested the hyperattenuating targets. Because the lesions in real practice may have a lower Hounsfield unit in comparison to that of the background (e.g., hepatic metastasis of colon cancer), further studies using hypoattenuating targets are required. Second, our study used phantoms and it is yet unknown whether our results can be reproduced using images of actual human body. Third, it was difficult to compare the time required for reconstruction and denoising across the algorithms because while FBP and ADMIRE images were reconstructed on real CT scanner consoles, our deep learning algorithm was implemented in an experimental setting, which has differences in terms of both hardware and software.

In conclusion, the low-contrast detectability of our deep learning-based denoising algorithm was noninferior to that of ADMIRE and superior to that of FBP.

References

1. Papanicolas I, Woskie LR, Jha AK (2018) Health care spending in the United States and other high-Income countries. JAMA 319:1024-1039

2. Brenner DJ, Hall EJ (2007) Computed tomography-an increasing source of radiation exposure. N Engl J Med 357:2277-2284

3. Smith-Bindman R, Lipson J, Marcus R et al (2009) Radiation dose associated with common computed tomography examinations and the associated lifetime attributable risk of cancer. Arch Intern Med 169:2078-2086

4. Fleischmann D, Boas FE (2011) Computed tomography-old ideas and new technology. Eur Radiol 21:510-517

5. Solomon J, Mileto A, Ramirez-Giraldo JC, Samei E (2015) Diagnostic performance of an Advanced Modeled Iterative Reconstruction algorithm for low-contrast detectability with a third-generation dual-source multidetector CT scanner: potential for radiation dose reduction in a multireader study. Radiology 275:735-745

6. Volders D, Bols A, Haspeslagh M, Coenegrachts K (2013) Model-based iterative reconstruction and adaptive statistical iterative reconstruction techniques in abdominal CT: comparison of image quality in the detection of colorectal liver metastases. Radiology 269:469-474

7. Chang W, Lee JM, Lee K et al (2013) Assessment of a model-based, iterative reconstruction algorithm (MBIR) regarding image quality and dose reduction in liver computed tomography. Invest Radiol 48:598-606

8. Fontarensky M, Alfidja A, Perignon R et al (2015) Reduced radiation dose with model-based iterative reconstruction versus standard dose with adaptive statistical iterative reconstruction in abdominal CT for diagnosis of acute renal colic. Radiology 276:156-166

9. Akagi M, Nakamura Y, Higaki T et al (2019) Deep learning reconstruction improves image quality of abdominal ultra-high-resolution CT. Eur Radiol

10. Tatsugami F, Higaki T, Nakamura Y et al (2019) Deep learning-based image restoration algorithm for coronary CT angiography. Eur

Radiol

11. Nakamura Y, Higaki T, Tatsugami F et al (2019) Deep learning-based CT image reconstruction: initial evaluation targeting hypovascular hepatic metastases. Radiology: Artificial Intelligence 1:e180011

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

13. Shin YJ, Chang W, Ye JC et al ([Under revision]) Improvement of image quality in low-dose computed tomography using a deep learning-based denoising algorithm. Korean J Radiol

14. Shan H, Padole A, Homayounieh F et al (2019) Competitive performance of a modularized deep neural network compared to commercial algorithms for low-dose CT image reconstruction. Nature Machine Intelligence 1:269

15. Du W, Chen H, Wu Z, Sun H, Liao P, Zhang Y (2017) Stacked competitive networks for noise reduction in low-dose CT. PLoS One 12:e0190069

16. Chen H, Zhang Y, Zhang W et al (2017) Low-dose CT via convolutional neural network. Biomed Opt Express 8:679-694

17. Rensink RA (2002) Change detection. Annu Rev Psychol 53:245-277

18. Ro T, Russell C, Lavie N (2001) Changing faces: a detection advantage in the flicker paradigm. Psychol Sci 12:94-99

19. Kim B, Lee H, Kim KJ et al (2011) Comparison of three image comparison methods for the visual assessment of the image fidelity of compressed computed tomography images. Med Phys 38:836-844

20. Verdun FR, Racine D, Ott JG et al (2015) Image quality in CT: From physical measurements to model observers. Phys Med 31:823-843

21. Friedman SN, Fung GS, Siewerdsen JH, Tsui BM (2013) 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 40:051907

22. Hillis SL, Obuchowski NA, Berbaum KS (2011) Power estimation for multireader ROC methods an updated and unified approach. Acad

Radiol 18:129-142

23. Chen W, Petrick NA, Sahiner B (2012) Hypothesis testing in noninferiority and equivalence MRMC ROC studies. Acad Radiol 19:1158-1165

24. Viry A, Aberle C, Racine D et al (2018) Effects of various generations of iterative CT reconstruction algorithms on low-contrast detectability as a function of the effective abdominal diameter: A quantitative task-based phantom study. Phys Med 48:111-118

25. Gonzalez-Guindalini FD, Ferreira Botelho MP, Tore HG, Ahn RW, Gordon LI, Yaghmai V (2013) MDCT of chest, abdomen, and pelvis using attenuation-based automated tube voltage selection in combination with iterative reconstruction: an intrapatient study of radiation dose and image quality. AJR Am J Roentgenol 201:1075-1082

26. Raman SP, Mahesh M, Blasko RV, Fishman EK (2013) CT scan parameters and radiation dose: practical advice for radiologists. J Am Coll Radiol 10:840-846

27. Schindera ST, Diedrichsen L, Muller HC et al (2011) Iterative

reconstruction algorithm for abdominal multidetector CT at different tube voltages: assessment of diagnostic accuracy, image quality, and radiation dose in a phantom study. Radiology 260:454–462

28. Kanal KM, Butler PF, Sengupta D, Bhargavan-Chatfield M, Coombs LP, Morin RL (2017) U.S. diagnostic reference levels and achievable doses for 10 adult CT examinations. Radiology 284:120–133

29. Zhu B, Liu JZ, Cauley SF, Rosen BR, Rosen MS (2018) Image reconstruction by domain-transform manifold learning. Nature 555:487–492

30. Nishizawa M, Tanaka H, Watanabe Y, Kunitomi Y, Tsukabe A, Tomiyama N (2015) Model-based iterative reconstruction for detection of subtle hypoattenuation in early cerebral infarction: a phantom study. Jpn J Radiol 33:26–32

Table 1. CT Scan and Reconstruction Parameters

Parameter	Data
Reference tube current−time product (mAs)	
Deep learning algorithm training data	200
Human observer study	200, 100, 50, 26
Physical measurement	200
Automated tube current modulation	
Deep learning algorithm training data	On
Human observer study, physical measurement	Off
Volume CT dose index (mGy)	
Deep learning algorithm training data	6.45 (5.49–7.20)
Human observer study	7.85, 3.94, 1.97, 1.05
Physical measurement	7.85
Tube potential (kVp)	100
Collimation (mm)	128 × 0.6
Section thickness (mm)	4
Rotation time (sec)	0.5
Pitch (mm)	0.6
Scan field of view (cm)	60
Display field of view (cm)	50
Scan option	Helical
Reconstruction algorithm	ADMIRE−3, FBP
Reconstruction kernel	I40f, B40f

Note.—ADMIRE denotes Advanced Modeled Iterative Reconstruction; FBP, filtered back projection.
CT dose index of the deep learning algorithm training data is shown in median and interquartile range in the parenthesis.

Table 2. Results of Noninferiority Tests

	DLA	ADMIRE			FBP		
	AUC	AUC	Difference (95% CI)	*P*–value	AUC	Difference (95% CI)	*P*–value
9 mm target	0.801	0.813	−0.012 (−0.044–0.019)	< 0.001*	0.744	0.057 (0.029–0.085)	< 0.001*
100% dose	0.976	0.982	−0.006 (−0.020–0.008)	< 0.001*	0.953	0.023 (0.002–0.045)	< 0.001*
50% dose	0.893	0.908	−0.015 (−0.066–0.037)	0.005*	0.826	0.068 (0.012–0.123)	< 0.001*
25% dose	0.723	0.717	0.006 (−0.075–0.086)	0.025*	0.647	0.075 (0.010–0.141)	< 0.001*
13% dose	0.626	0.652	−0.025 (−0.097–0.046)	0.040*	0.558	0.068 (−0.018–0.155)	< 0.001*
5 mm target	0.586	0.599	−0.013 (−0.051–0.026)	< 0.001*	0.564	0.022 (−0.007–0.050)	< 0.001*
100% dose	0.658	0.697	−0.039 (−0.130–0.052)	0.094	0.633	0.026 (−0.042–0.093)	0.001*
50% dose	0.604	0.614	−0.011 (−0.078–0.057)	0.025*	0.543	0.061 (−0.003–0.124)	< 0.001*
25% dose	0.549	0.540	0.009 (−0.077–0.095)	0.025*	0.559	−0.009 (−0.065–0.046)	0.005*
13% dose	0.536	0.555	−0.019 (−0.069–0.031)	0.005*	0.526	0.010 (−0.038–0.058)	< 0.001*
Total	0.692	0.705	−0.013 (−0.038–0.012)	< 0.001*	0.653	0.039 (0.017–0.0601)	< 0.001*

Note.—DLA denotes Deep Learning Algorithm; ADMIRE, Advanced Modeled Iterative Reconstruction; FBP, Filtered Back Projection; AUC, Area Under receiver operating characteristics Curve; CI, Confidence Interval.

*P–value < 0.05

Table 3. Multivariable Regression Analysis to Identify Factors Associated with Response Time

	Coefficient	*P*–value
Algorithm		
DLA	Reference	
ADMIRE	0.06 (−0.05–0.17)	0.26
FBP	−0.10 (−0.21–0.01)	0.068
Target presence		
Present	Reference	
Absent	0.20 (0.11–0.29)	<0.001*
Target size		
9 mm	Reference	
5 mm	0.16 (0.07–0.25)	0.001*
Radiation dose		
100% dose	Reference	
50% dose	0.21 (0.08–0.34)	0.001*
25% dose	0.20 (0.07–0.33)	0.002*
13% dose	0.39 (0.26–0.51)	<0.001*

Note.—DLA denotes Deep Learning Algorithm; ADMIRE, Advanced Modeled Iterative Reconstruction; FBP, Filtered Back Projection.
*P–value <0.05

Table 4. TTF_{50}, AUNPS, and NPS peak frequency of three algorithms

	DLA	ADMIRE		FBP	
	Data	Data	*P*-value	Data	*P*-value
TTF_{50}					
Disc 1, polyethylene	0.541 ± 0.008	0.539 ± 0.006	0.049*	0.441 ± 0.008	<0.001*
Disc 2, bone	0.484 ± 0.004	0.487 ± 0.004	0.002*	0.445 ± 0.002	<0.001*
Disc 3, acrylic	0.541 ± 0.006	0.534 ± 0.006	<0.001*	0.451 ± 0.010	<0.001*
Disc 4, air	0.448 ± 0.003	0.444 ± 0.003	<0.001*	0.437 ± 0.003	<0.001*
Area under NPS (HU^2mm^2)	11603.6 ± 307.0	13614.8 ± 337.8	<0.001*	22629.2 ± 659.5	<0.001*
NPS_{peak} (mm^{-1})	0.185 ± 0.027	0.185 ± 0.027	>0.99	0.230 ± 0.036	<0.001*

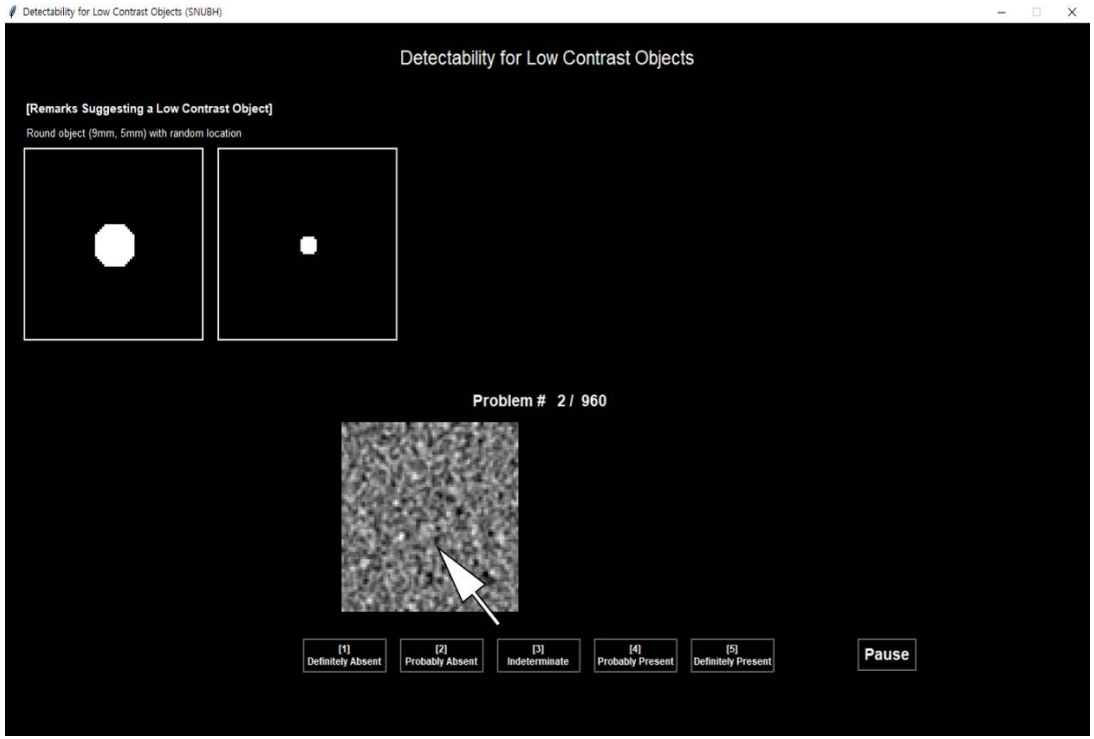
Note.—Data are mean ± standard deviation. *P*-values were calculated using paired t-test.

DLA denotes deep learning algorithm; ADMIRE, advanced modeled iterative reconstruction; FBP, filtered back projection; TTF, target transfer function; NPS, noise power spectrum.

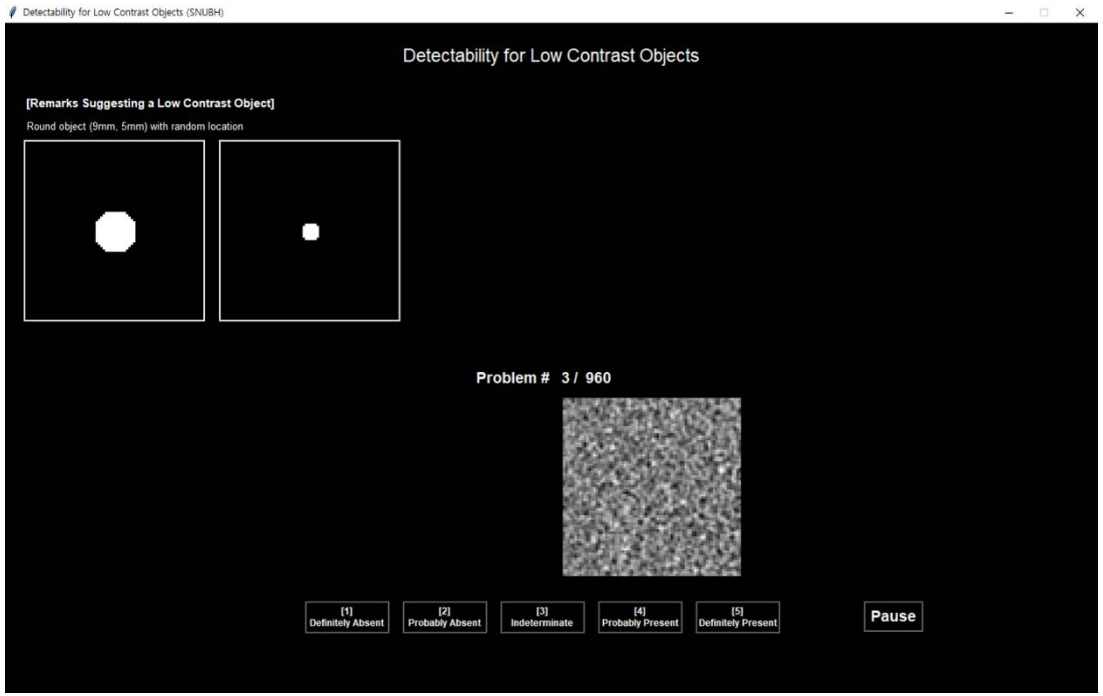
*P-value < 0.05

Figure 1.

Screen capture images of the graphic user interface used in this study.



(A) The even-numbered images appeared on the left side of the screen and the target appeared in random locations of the images (arrow).



(B) The odd-numbered images appeared on the right side of the screen.

Figure 2.

Representative images of deep learning algorithm (A), advanced modeled iterative reconstruction (ADMIRE) (B), and filtered back projection (FBP) (C).

(A) (B) (C)

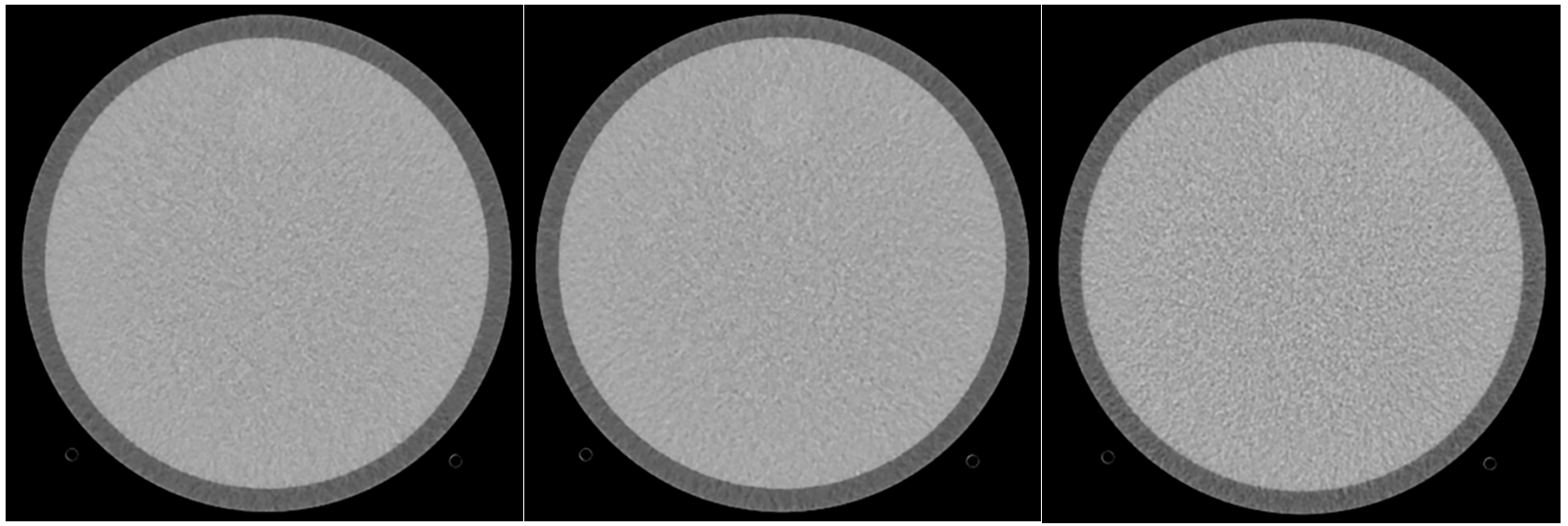
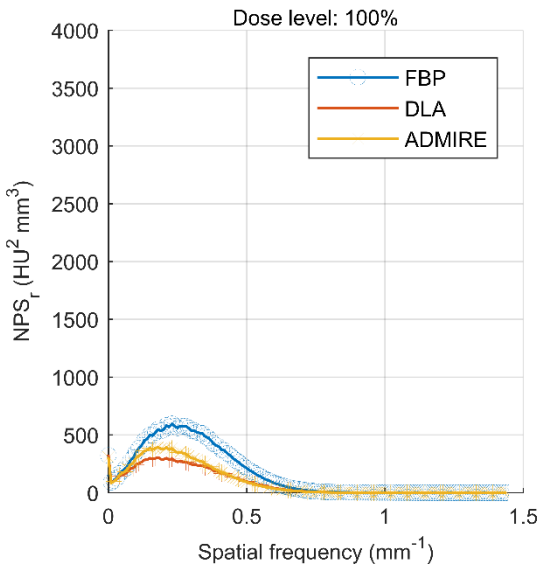


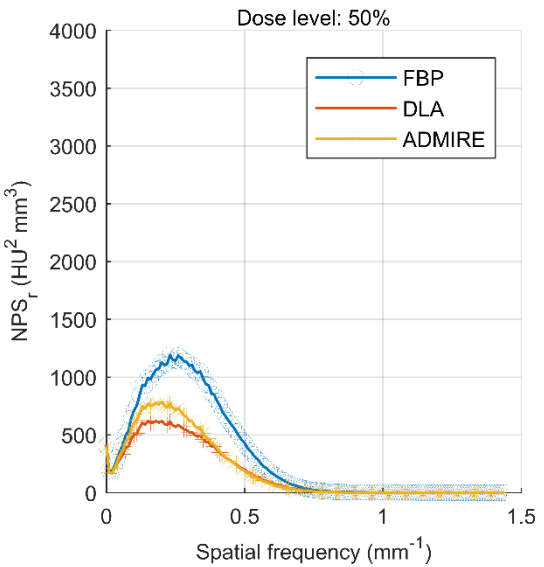
Figure 3.

Noise power spectrum (NPS) of the deep learning algorithm, advanced modeled iterative reconstruction (ADMIRE), and filtered back projection (FBP) in 100% dose (A), 50% dose (B), 25% dose (C), and 13% dose (D).

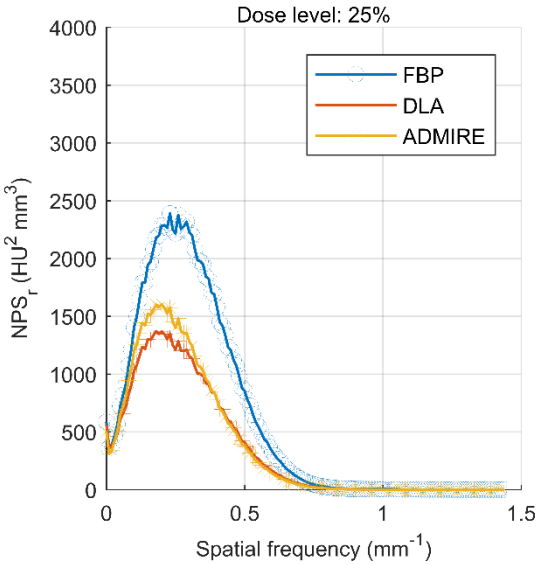
(A)



(B)



(C)


Dose level: 25%
FBP
DLA
ADMIRE
NPS_r (HU^2 mm^3)
Spatial frequency (mm^-1)
4000
3500
3000
2500
2000
1500
1000
500
0
0
0.5
1
1.5

(D)

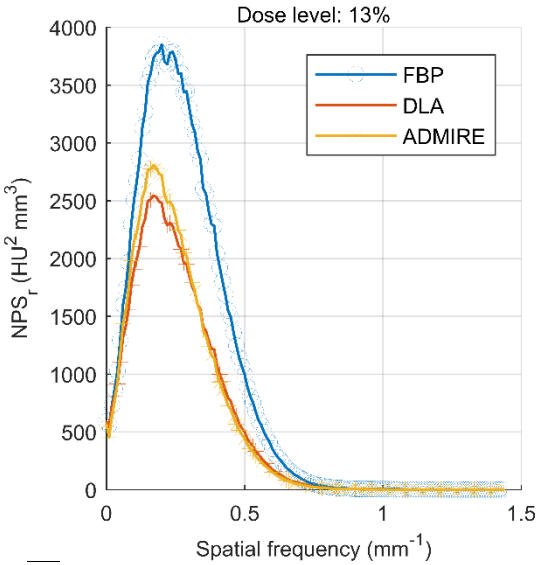
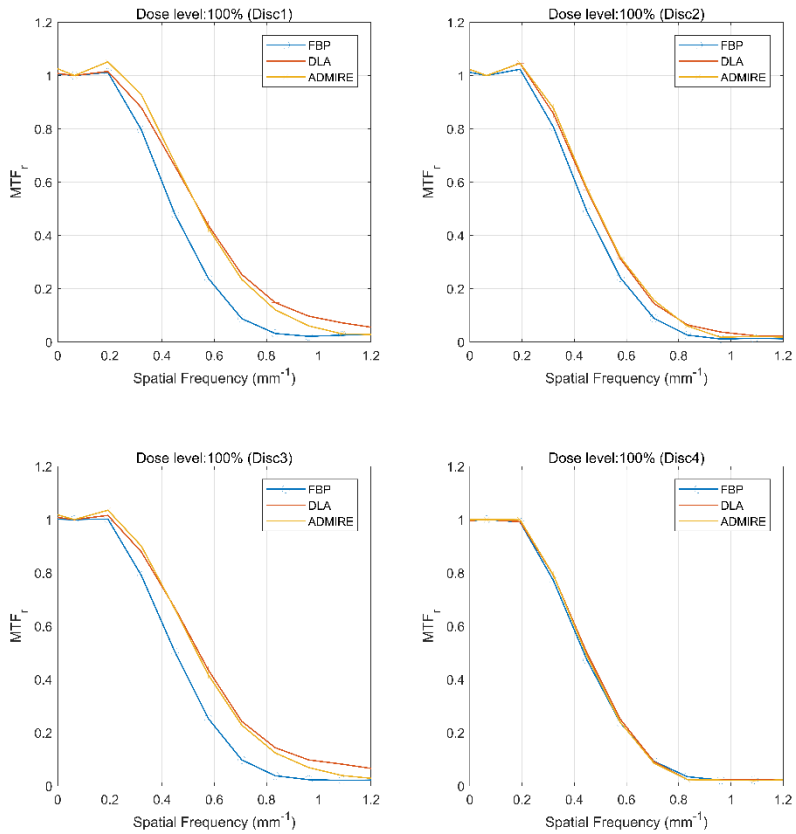

Dose level: 13%
FBP
DLA
ADMIRE
NPS_r (HU^2 mm^3)
Spatial frequency (mm^-1)
4000
3500
3000
2500
2000
1500
1000
500
0
0
0.5
1
1.5

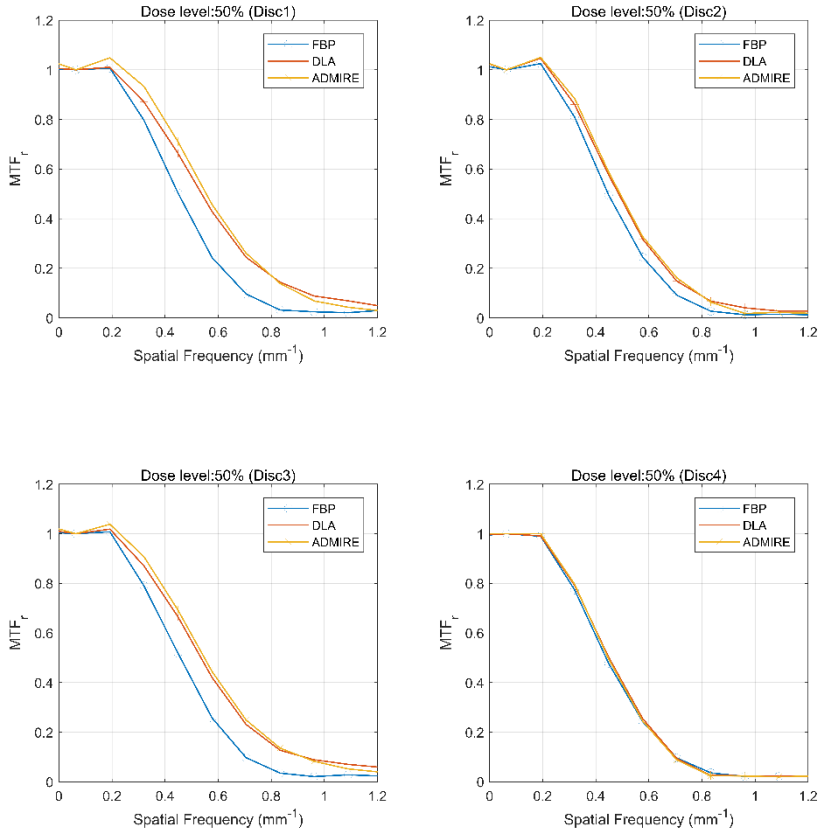
Figure 4.

Target transfer function (TTF) of the deep learning algorithm, advanced modeled iterative reconstruction (ADMIRE), and filtered back projection (FBP) in 100% dose (A), 50% dose (B), 25% dose (C), and 13% dose (D). Disc 1 denotes polyethylene disc; Disc 2, hypothetical bone disc; Disc 3, acrylic disc; and Disc 4, hypothetical air disc.

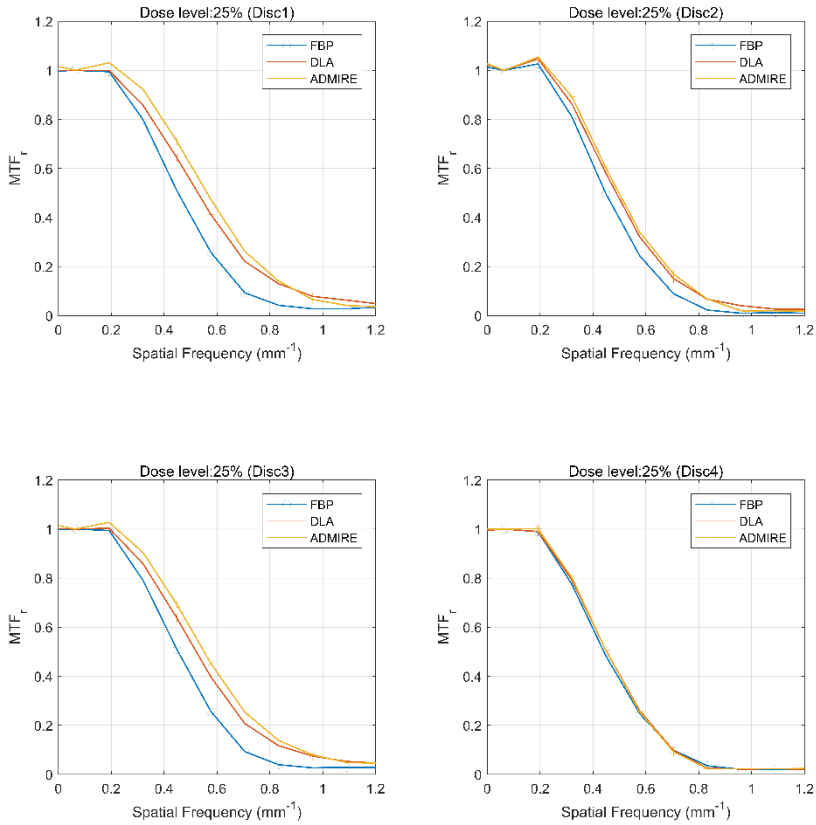
(A)



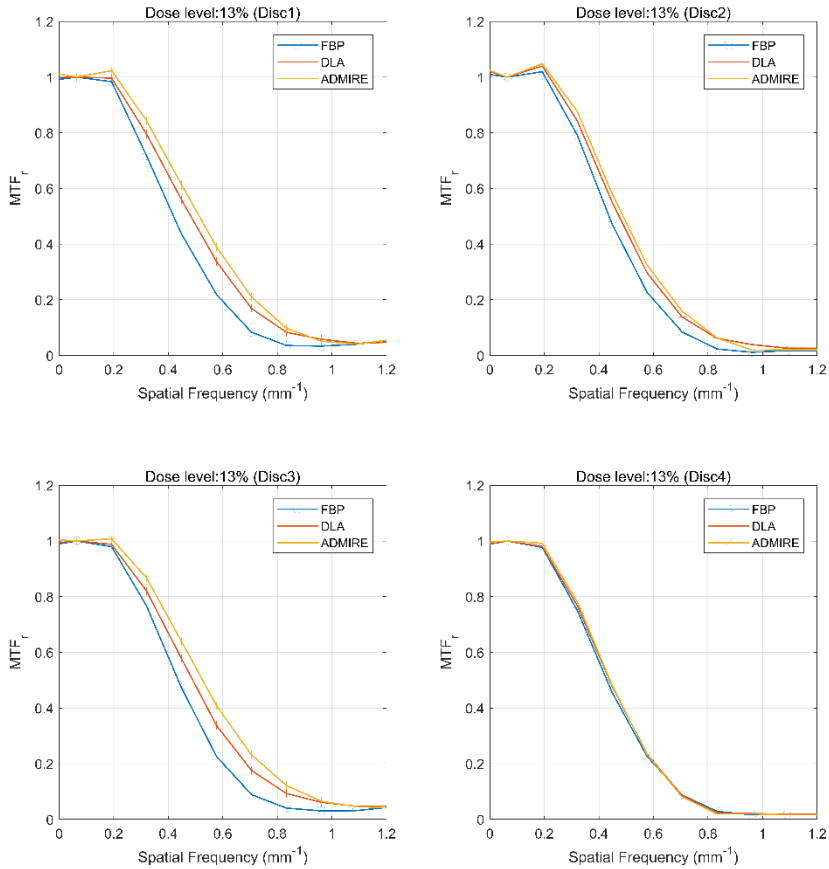
(B)


Dose level:50% (Disc1)
Dose level:50% (Disc2)
Dose level:50% (Disc3)
Dose level:50% (Disc4)
FBP
DLA
ADMIRE
MTF r
Spatial Frequency (mm-1)

(C)


Dose level:25% (Disc1)
Dose level:25% (Disc2)
Dose level:25% (Disc3)
Dose level:25% (Disc4)
FBP
DLA
ADMIRE
MTF r
Spatial Frequency (mm-1)
0
0.2
0.4
0.6
0.8
1
1.2

(D)


Dose level:13% (Disc1)
Dose level:13% (Disc2)
Dose level:13% (Disc3)
Dose level:13% (Disc4)
FBP
DLA
ADMIRE
MTF r
Spatial Frequency (mm-1)

논 문 초 록

목적: 딥러닝 기반의 전산화단층 촬영 잡음 제거 알고리즘에서의 저대조도검출률을 상용화된 알고리즘인 ADMIRE (advanced modeled iterative reconstruction)과 FBP (filtered back projection)에서의 저대조도검출률과 비교하고자 한다.

대상 및 방법: ADMIRE 와 FBP 로 재구성된 환자 100 명의 복부골반전산화단층촬영 영상이 딥러닝 알고리즘 기반의 잡음 제거 알고리즘 훈련에 이용되였다. 우리는 FBP 를 입력영상으로, ADMIRE 를 출력영상으로 설정하여 딥러닝을 훈련하였다. 딥러닝의 저대조도검출률은 다양한 방사선량 조건에서 측정되었다 (100 kVp, 200, 100, 50, 26 mAs). 세 종류의 알고리즘에서 저대조도검출률을 비교하기 위해 12 명의 영상의학과 의사가 저대조도 물체 존재 유무에 대해 평가하였다. 다중리더 다중케이스 곡선하면적이 측정되었으며 알고리즘 간에 비열등성 검정을 수행하였다. 목표 전달 함수 (target transfer function) 및 잡음 밀도 스펙트럼 (noise power spectrum)은 목표 전달 함수 중위값 진동수, 잡음 밀도 스펙트럼 곡선 하 면적, 잡음 밀도 스펙트럼 최빈 진동수로 비교되었다.

결과: 딥러닝 알고리즘의 저대조도검출률을 ADMIRE 와 비교하여 비열등 (P<.001)하였으며 FBP 와 비교하여 우열하였다 (P<.001}. 목표 전달 함수 중위값 진동수는 딥러닝 알고리즘이 ADMIRE 및 FBP 와 비교하여 다양한 실험조건에서 높은 결과를 보였다. 잡음 밀도 스펙트럼

곡선 하 면적은 딥러닝 알고리즘이 ADMIRE 및 FBP 보다 낮은 값을 보였다 (P<.001). 잡음 밀도 스펙트럼 최빈 진동수 측면에서 딥러닝 알고리즘은 ADMIRE 와 비슷하였고 (P>.99), FBP 와 비교해서는 낮은 값을 보였다 (P<.001).

결론: 딥러닝 기반의 전산화단층 촬영 잡음 제거 알고리즘에서의 저대조도검출률을 ADMIRE 와 비교하여 비열등하였으며 FBP 와 비교하여 우열하다.

주요어: 딥러닝, 저대조도검출률, 인간 관찰자 연구, 잡음 제거, 전산화단층촬영
학번: 2018-20093