











Radiation Dose Reduction in Digital Mammography by Deep-Learning Algorithm Image Reconstruction: A Preliminary Study

딥러닝 알고리즘을 이용한 저선량 디지털 유방 촬영 영상의 복원: 예비 연구

Su Min Ha, MD^{1,2} , Hak Hee Kim, MD^{3*} , Eunhee Kang, PhD⁴ ,
Bo Kyoung Seo, MD⁵ , Nami Choi, MD⁶ , Tae Hee Kim, MD⁷ ,
You Jin Ku, MD⁸ , Jong Chul Ye, PhD⁴ 

¹Department of Radiology, Research Institute of Radiology, Chung-Ang University Hospital, Chung-Ang University College of Medicine, Seoul, Korea

²Department of Radiology, Research Institute of Radiology, Seoul National University Hospital, Seoul National University College of Medicine, Seoul, Korea

³Department of Radiology, Research Institute of Radiology, Asan Medical Center, University of Ulsan College of Medicine, Seoul, Korea

⁴Department of Bio and Brain Engineering, Korea Advanced Institute of Science and Technology (KAIST), Daejeon, Korea

⁵Department of Radiology, Korea University Ansan Hospital, Korea University College of Medicine, Ansan, Korea

⁶Department of Radiology, Konkuk University Medical Center, Konkuk University School of Medicine, Seoul, Korea

⁷Department of Radiology, Ajou University Hospital, Ajou University School of Medicine, Suwon, Korea

⁸Department of Radiology, Catholic Kwangdong University International St. Mary's Hospital, Catholic Kwangdong University, Incheon, Korea

Purpose To develop a denoising convolutional neural network-based image processing technique and investigate its efficacy in diagnosing breast cancer using low-dose mammography imaging.

Materials and Methods A total of 6 breast radiologists were included in this prospective study. All radiologists independently evaluated low-dose images for lesion detection and rated them for diagnostic quality using a qualitative scale. After application of the denoising network, the same radiologists evaluated lesion detectability and image quality. For clinical application, a consensus on lesion type and localization on preoperative mammographic examinations of breast cancer patients was reached after discussion. Thereafter, coded low-dose, reconstructed full-dose, and full-dose images were presented and assessed in a random order.

Received August 19, 2020

Revised October 28, 2020

Accepted July 23, 2021

*Corresponding author

Hak Hee Kim, MD

Department of Radiology,
Research Institute of Radiology,
Asan Medical Center,
University of Ulsan
College of Medicine,
88 Olympic-ro 43-gil, Songpa-gu,
Seoul 05505, Korea.

Tel 82-2-3010-4390

Fax 82-2-476-0090


E-mail radhhkim@hanmail.net

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<https://creativecommons.org/licenses/by-nc/4.0>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.


ORCID iDs

Su Min Ha 


<https://orcid.org/0000-0002-1833-0919>

Hak Hee Kim 

<https://orcid.org/0000-0002-2956-9212>

Eunhee Kang 


<https://orcid.org/0000-0001-9921-8103>

Bo Kyoung Seo 

<https://orcid.org/0000-0002-9512-5361>

Nami Choi 


<https://orcid.org/0000-0002-2703-8282>

Tae Hee Kim 

<https://orcid.org/0000-0002-0520-2395>

You Jin Ku 

<https://orcid.org/0000-0002-6727-0289>

Jong Chul Ye 

<https://orcid.org/0000-0001-9763-9609>

Results Lesions on 40% reconstructed full-dose images were better perceived when compared with low-dose images of mastectomy specimens as a reference. In clinical application, as compared to 40% reconstructed images, higher values were given on full-dose images for resolution ($p < 0.001$); diagnostic quality for calcifications ($p < 0.001$); and for masses, asymmetry, or architectural distortion ($p = 0.037$). The 40% reconstructed images showed comparable values to 100% full-dose images for overall quality ($p = 0.547$), lesion visibility ($p = 0.120$), and contrast ($p = 0.083$), without significant differences.

Conclusion Effective denoising and image reconstruction processing techniques can enable breast cancer diagnosis with substantial radiation dose reduction.

Index terms Deep Learning; Artificial Intelligence; Radiation; Mammography; Breast Neoplasm

INTRODUCTION

Although mammography is a reliable method for breast cancer screening (1), the increasing awareness of the need for radiation protection raises concern, which is particularly relevant when a patient needs to be repeatedly screened. It is standard practice that the radiation dose should be maintained as low as reasonably achievable (ALARA) while still providing an image quality adequate to enable accurate diagnosis (1, 2). A higher radiation dose will lower the noise level and improve image quality but impart an unnecessary radiation dose to the patient (2), while a lower radiation dose will lower the signal to noise ratio (SNR) of the image and may negatively affect diagnostic performance. Thus, the appropriate radiation dose for mammography should be dictated by the radiation required to achieve an adequate SNR for an accurate diagnosis. Indeed, digital mammography systems are equipped with automatic exposure control (AEC) to adjust the kilovoltage peak (kVp), milliamperere-second (mAs), and filter, to provide an acceptable radiation dose and image contrast. In addition, digital breast tomosynthesis (DBT), which is a three-dimensional imaging technology with multiple projections over a limited angle, presents a further issue regarding radiation exposure. DBT has rapidly disseminated as a routine breast screening tool with increased cancer detection rate, decreased recall rate and improved performance metrics of screening, specifically among women with dense breast (3). Repeated DBT for annual screening may increase the cumulative radiation exposure and lifetime attributable risk of radiation-induced breast cancer (4). Although many studies have examined the relationship between reduced-radiation dose and accuracy in the detection of lesions on mammography (5-7), the reduction of noise and artifacts arising from a decreased number of photons and increased Poisson noise remains a challenge. There has been previous attempts to compensate low image contrast by image enhancement and de-noising methods, such as the kernel-based iterative scatter correction method used for X-ray cone-beam CT (8) and iterative image de-noising methods used for mammography (9-11). However, these methods are rather impractical and its performances are not enough to use a low dose X-ray mammography images in real situations. Unlike CT, there are few researches related to X-ray mammography image quality improvement, because its resolution is too high incomparable to CT images and hard to modeling the measurement acquisition process with detector's characteristic such as grid. This problem can be solved by

machine learning approach which has network with huge capacity and good ability to estimate the non-linear model distribution

Thus, we hypothesized that the low-dose image reconstruction problem in mammography could be solved with a machine learning approach using a high capacity network. Deep-learning methods have shown great potential in medical imaging applications such as image reconstruction and de-noising (12, 13) and indeed, a previous study (14) proposed a generative adversarial network (GAN) for mammography synthesis that could be used for data augmentation processes. Recently, Liu et al. (15) investigated a deep-learning-based supervised image-processing technique for radiation dose reduction in DBT. They trained a convolutional neural network (CNN) on quarter-dose and corresponding high-dose DBT images of a breast phantom, and demonstrated a 79.0% dose reduction with the application of their deep-learning technique. However, to the best of our knowledge, no previous study has investigated deep-learning-based solutions for digital mammography image reconstruction.

Therefore, we here propose a de-noising CNN specialized for digital mammography. This CNN converts low-dose images to reconstructed full-dose images, with the noise and artifact present in the low-dose images largely eliminated. Our study involved two steps: first, we evaluated our CNN image-processing technique with breast cancer mastectomy specimens acquired at different mammographic radiation dose levels; and second, using the acceptable low dose determined in the first step, we examined breast cancer patients using the reduced dose and investigated reduced dosage images by applying our CNN technique and evaluated the detection performance in breast cancer diagnosis compared to original images.

MATERIALS AND METHODS

STUDY POPULATION

This prospective study was approved by the Institutional Review Boards of six institutions (Chung-Ang University Hospital [IRB No. 1771-002-288], Asan Medical Center [IRB No. S2018-0913-0002], Korea University Ansan Hospital [IRB No. 2018AS0189], Konkuk University Medical Center [IRB No. KUH 1140134], Ajou University Hospital [IRB No. AJIRB-MED-OBS-18-274], Catholic Kwangdong University International St. Mary's Hospital [IRB No. IS16OCMI0036]). After receiving the relevant permissions of the ethics committees and written informed patients' consents, mastectomy specimens were acquired from 31 patients with a recent diagnosis of breast cancer and who were treated with total mastectomy. The standard full-dose X-ray level was determined through the AEC, and the mastectomy specimens were subsequently imaged at five different radiation dose levels (80%, 60%, 40%, 20%, and 10% of AEC) on Selenia Dimension (Hologic Inc., Bedford, MA, USA; $n = 8$), Senographe DS, or Senographe Essential scanners (GE Healthcare, Milwaukee, WI, USA; $n = 23$). The low-dose images were acquired by changing the mAs with the tube voltage fixed (ranges, 26–30 kVp for the Selenia Dimension; 25–29 kVp for the Senographe DS/Essential). Three cases were excluded because of poor image quality with unavoidable artifact, and 28 mastectomy specimen cases were finally included in the analysis.

TRAINING, TESTING, AND VALIDATION OF THE CNN DE-NOISING METHOD USING THE MASTECTOMY SPECIMENS

The CNN was trained using low-dose and full-dose images as inputs. As there could be a mismatch between low-dose and full-dose images due to potential deformation of the specimen during multiple acquisitions, a de-noising method based on unsupervised learning using cyclic consistency was utilized. The proposed framework is described in Fig. 1.

For a subset of the data, the CNN de-noising method was validated by measuring the structural similarity index (SSIM) (16), which is a quantitative value allowing the evaluation of image quality according to noise characteristics in comparison with the full-dose image. SSIM is a widely accepted image quality measure that overcomes a limitation of the conventional SNR measure, which lacks the incorporation of spatial information in the evaluation. In the production of the “simulated” low-dose image, Poisson noise was added to the full-dose image. The SNR of the “simulated” low-dose image was 17 decibel (dB). The SSIM is represented by

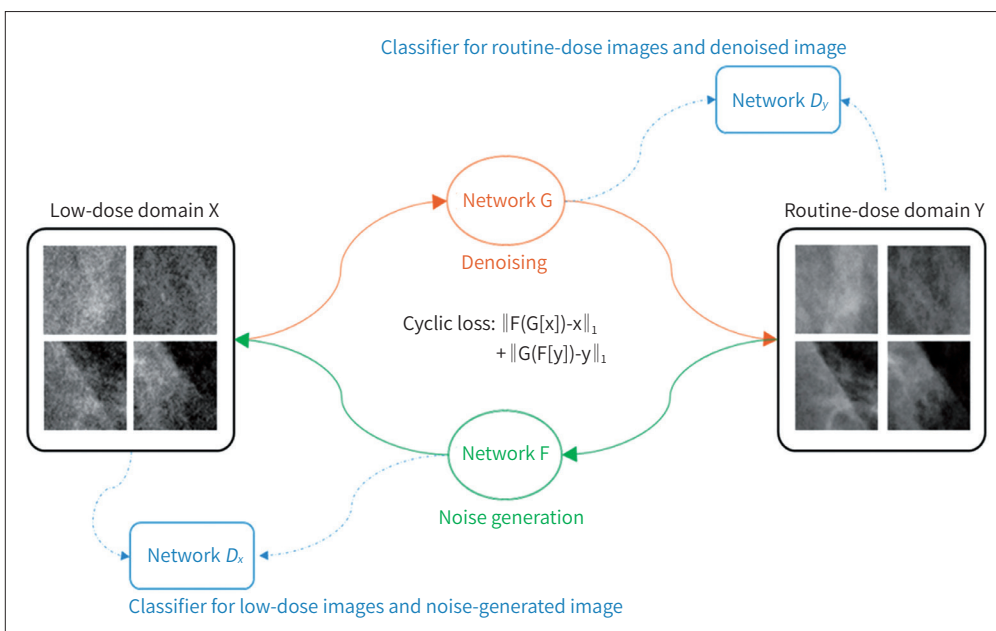
$$\text{SSIM}(x, y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)}$$

where μ_x , μ_y , and $\sigma_x^2\sigma_y^2\sigma_{xy}$ are the average, variance, and covariance of x and y , respectively; $c_1 = (k_1L)^2$ and $c_2 = (k_2L)^2$ are two variables that stabilize the division with a weak denominator; L is the dynamic range of the pixel-values; and $k_1 = 0.01$ and $k_2 = 0.03$ by default.

IMAGE INTERPRETATION USING THE MASTECTOMY SPECIMENS

Six breast radiologists (S.M.H., with 6 years of breast imaging experience; H.H.K, with 27 years of experience, B.K.S, with 21 years of experience, N.M.C, with 15 years of experience,

Fig. 1. Our network contained two functions, $G: X \rightarrow Y$ and $F: Y \rightarrow X$, wherein we trained two generators (Network G and Network F) and two discriminators (Network D_x and Network D_y), with the use of three loss functions.



T.H.K, with 12 years of experience, and Y.J.K. with 5 years of experience) were involved in the image evaluation procedure. All radiologists independently evaluated the low-dose images (80%, 60%, 40%, 20%, and 10% of AEC) for lesion detection (no detection, detection), and by comparing them with the respective reference full-dose images, rated them for diagnostic quality using a qualitative 3-point scale (equivalent, acceptable, or unacceptable) while blinded to dose-level information. Each radiologist interpreted the images stored in DICOM format and was allowed to alter the window/level settings as required. After application of the de-noising network to the low-dose images, the same radiologists again recorded whether the lesion was detectable or not, and they were presented with full-dose images and randomly corresponding low-dose and reconstructed full-dose images, and were again asked to rate their quality on the 3-point scale (decreased, equivalent, improved) compared to low-dose and full-dose images as reference standard.

APPLICATION FOR CLINICAL CASES

For the second step of our study, 102 patients with a recent diagnosis of breast cancer and who gave written informed consent, with approval by the institutional review boards of the above same six institution, underwent preoperative routine mammography consisting of cranio-caudal (CC) and mediolateral oblique (MLO) views with an additional MLO projection view using the pre-determined reduced low dose (obtained from the first step of our study). The patients were examined using Hologic ($n = 62$) or GE Healthcare systems ($n = 40$). The low-dose images were acquired by changing the mAs with the tube voltage fixed (ranges, 24–33 kVp for the Selenia Dimension; 26–34 kVp for the Senographe DS/Essential). All methods were carried out in accordance with relevant guidelines and regulations.

TRAINING, TESTING, AND VALIDATION OF THE CNN DE-NOISING ON CLINICAL CASES

The network was trained with low-dose and full-dose images as input. As there were mismatches between input and full-dose images during multiple acquisitions, a de-noising method based on unsupervised learning using cyclic consistency was applied. The proposed framework is described in Fig. 2. For a subset of the data, the CNN de-noising method was validated by measuring SSIM (16).

IMAGE INTERPRETATION ON CLINICAL CASES

For the second step of our study, before performing independent analysis, the same radiologists reviewed altogether and made consensus decisions on the lesion types (mass, calcification, mass with calcification, architectural distortion, asymmetry) and localizations (upper, middle, lower) on the preoperative mammographic examinations. Then, one radiologist (S.M.H.) provided the other five radiologists with coded low-dose, reconstructed full-dose, and full-dose images, with all dose-level and patient information being blinded. The three images were presented in a random order for each patient. The five readers then used a known-location paradigm to score all the images considering the overall image quality, visibility of the lesion, image contrast, image resolution, and diagnostic quality of the lesion (calcification, mass, asymmetry, architectural distortion) on 1–10 point scales. The diagnostic quality of the lesion

was defined as the ability to describe mammographic findings according to breast imaging-reporting and data system (BI-RADS) lexicon such as margin of the mass or morphology of calcifications. The type of breast tissue was also determined independently according to the American College of Radiology (ACR) guidelines (17).

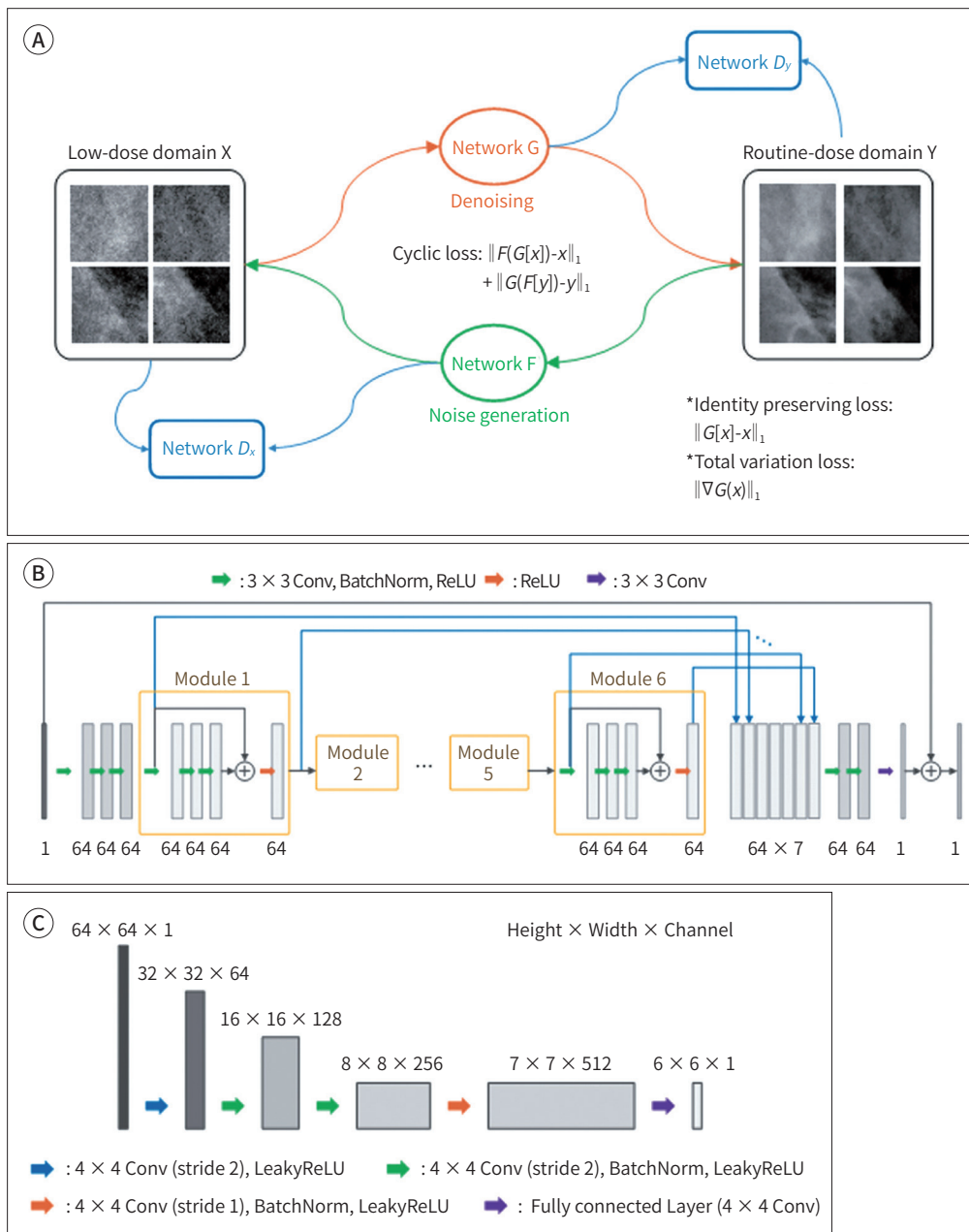
Fig. 2. De-noising method based on unsupervised learning using cyclic consistency.

A. Overview of the proposed framework (X, low-dose domain; Y, routine-dose domain), wherein the generator G denoted the mapping from X to Y, and F similarly defined the mapping from Y to X; there are two adversarial discriminators D_x and D_y , which distinguished between measured input images and reconstructed images from the generators.

B. Network architecture of the generators G and F.

C. Network architecture of the discriminators D_x and D_y .

Conv = convolutional, ReLU = rectified linear unit



STATISTICAL ANALYSIS

For the first step of the study, the mean percentage of the lesion detectability and diagnostic quality assessment scale from the readers was calculated for each lesion type (mass, calcification) according to each dose. The trends were analyzed using the Mantel–Haenszel statistic. After application of the de-noising and image reconstruction method, lesion detectability and diagnostic quality assessments were compared with McNemar's test or a marginal homogeneity test using a generalized estimating equations (GEE) model to account for data clustering effects. SSIM improvement was also assessed using the Wilcoxon signed-rank test or paired *t* test.

For the second step of the study, the observer results for each low-dose, reconstructed full-dose, and full-dose image were assessed using a two-way random effects model with provision of intraclass correlation coefficients (ICCs). An ICC reflects the reliability of ratings for data that have been collected as a group or sorted into groups. An ICC greater than 0.9 means excellent reliability, 0.75–0.9 good reliability, 0.5–0.75 moderate reliability, and less than 0.5 poor reliability. The multiple assessment criteria for each image were compared with analysis of variance (ANOVA) adjusted by Dunnett's method. Whether lesion size or parenchymal density affected the image assessment and diagnostic quality of the lesions was assessed using a linear mixed model. A *p* value of less than 0.05 was considered to indicate a statistically significant difference. Statistical analyses were performed using SPSS software (version 14.0; Statistical Package for the Social Sciences, Chicago, IL, USA).

RESULTS

Evaluation of Mastectomy Specimens before and after Application of the De-Noising Method Similar trends were observed in the qualitative assessments for lesion detection and diagnostic quality, except for mass quality ($p = 0.173$) (Supplementary Table 1 in the online-only Data Supplement). Mass and calcification detection rates decreased substantially for the 10% low-dose image. The 'not acceptable' quality rating for masses showed a notable increase with 10% low-dose images, while for calcifications, a notable increase was observed at 20%. As the lesion detection rate was maintained with a 20% low dose, and the diagnostic quality of masses and calcifications was maintained at 20% and 40% low-dose levels, respectively, the de-noising and image reconstruction method was applied to 20% and 40% low-dose images.

Masses and calcifications showed detection rates of 87.4% and 90.0%, respectively, on the 20% reconstructed full-dose images, and rates of 96.8% and 100.0% on the 40% reconstructed full-dose images (Table 1). The 'no detection' assessment for both lesion types was significantly higher for the 20% reconstructed full-dose images than for the 40% reconstructed full-dose images (12.6% vs. 3.2%, $p = 0.003$ for masses; 10.0% vs. 0.0%, $p < 0.001$ for calcifications). When 20% and 40% reconstructed full-dose images were compared with the reference full-dose images, there was a significantly higher proportion of 'equivalent' image quality assessments for both masses and calcifications at the 40% dose level (65.3% for masses; 65.0% for calcifications) than at the 20% dose level (41.1% for masses; 20.8% for calcifications; $p < 0.001$). With low-dose images as a reference, the lesions on the 40% reconstructed full-dose images were better perceived (20.0% for masses; 18.3% for calcifications), whereas the quality of the recon-

Table 1. Detection and Diagnostic Quality of Lesions on Reconstructed Full-Dose Mammography

Lesion Type	Detection Assessment	Mean %*		p-Value
		Reconstructed 20% Dose	Reconstructed 40% Dose	
Mass	No detection	12.6	3.2	0.003
	Detection	87.4	96.8	
Calcification	No detection	10.0	0.0	<0.001
	Detection	90.0	100.0	
Lesion Type	Quality Assessment	Mean %*		p-Value
		Reconstructed 20% Dose	Reconstructed 40% Dose	
Mass	Decrease [†]	54.7	14.7	<0.001
	Equivalent [†]	41.1	65.3	
	Improve [†]	4.2	20.0	
	Decrease [‡]	25.3	4.2	
	Equivalent [‡]	62.1	75.8	
	Improve [‡]	12.6	20.0	
Calcification	Decrease [†]	78.3	26.7	<0.001
	Equivalent [†]	20.8	65.0	
	Improve [†]	0.8	8.3	
	Decrease [‡]	19.2	5.8	
	Equivalent [‡]	70.8	75.8	
	Improve [‡]	10.0	18.3	

p value calculated with McNemar's test or marginal homogeneity test using a generalized estimating equations model to account for data clustering effect.

*Mean refers to the mean percentage of the lesion detectability and diagnostic quality assessment scale from the readers calculated for each lesion type (mass, calcification) according to each dose.

[†] Full-dose used as the reference standard for diagnostic quality assessment.

[‡] Low-dose (20% or 40%) was used as the reference standard for diagnostic quality assessment.

reconstructed full-dose images was rather decreased with the 20% dose (25.3% for masses; 19.2% for calcifications). In addition, CNN de-noising method was validated by improved SSIM values in eight subsets of our data (average value, 0.63 ± 0.05 ; range, 0.55–0.69 vs. 0.79 ± 0.07 ; range, 0.63–0.86; $p = 0.01$) (Fig. 3).

APPLICATION OF THE DE-NOISING METHOD TO CLINICAL CASES

From the first step result, additional MLO projection views with 40% reduced dose were obtained from breast cancer patients ($n = 102$). Among 102 lesions, the lesion types were composed of mass ($n = 49$, 48.0%), calcification ($n = 17$, 16.7%), mass with calcification ($n = 23$, 22.5%), architectural distortion ($n = 5$, 4.9%), and asymmetry ($n = 8$, 7.8%). The tumor stage according to pathologic tumor size for these patients were composed of T1 ($n = 29$, 28.4%), T2 ($n = 49$, 48.0%) and T3 ($n = 24$, 23.5%). Using 100% full-dose images as a reference, comparisons of assessment criteria concerning overall image quality, lesion visibility, contrast, resolution, diagnostic quality for calcifications, diagnostic quality of masses, asymmetry, or architectural distortion with low-dose images all revealed lower mean values with significant differences, with p values < 0.05 (Table 2). Though still lower than 100% full-dose images, reconstructed full-dose images showed comparable values for overall image quality ($p = 0.547$), visibility of

lesions ($p = 0.120$) and contrast ($p = 0.083$). The readers gave higher mean values for 100% full-dose images for image resolution ($p < 0.001$), diagnostic quality for calcifications ($p < 0.001$), and diagnostic quality for masses, asymmetry, or architectural distortion ($p = 0.037$). Reduc-

Fig. 3. Reconstructed full-dose (middle) results from specimen with calcification, which is improved in comparison to low-dose images and was deemed comparable to routine dose images by all readers. The magnified views ($\times 2$) indicated by the yellow boxes demonstrate how the conspicuity of the calcifications is maintained (routine dose image at right).

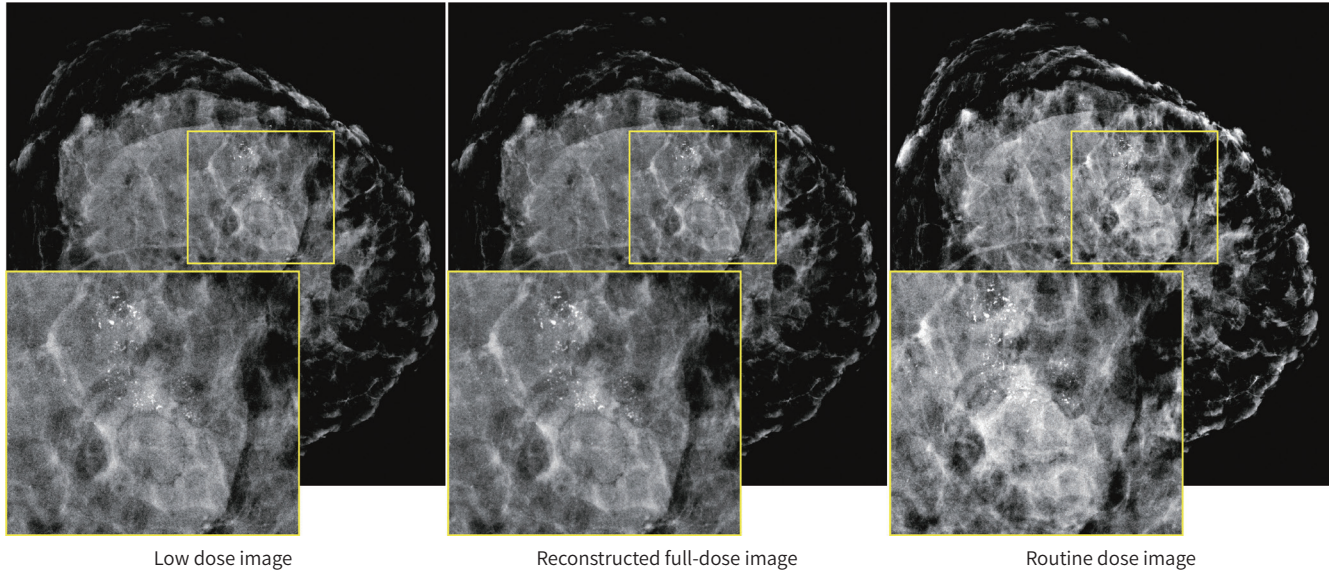


Table 2. Comparisons of Low-Dose, Full-Dose, and Reconstructed Full-Dose Images according to Different Assessment Parameters ($n = 102$)

Parameters	Image	Reader	p -Value	p -Value [†]	p -Value [*]
Overall image quality	100%	8.15 ± 1.02		0.547	
	40%	7.41 ± 1.07	< 0.001	0.900	< 0.001
	Reconstructed 40%	7.46 ± 1.10			0.547
Visibility of lesion	100%	8.48 ± 1.31		0.120	
	40%	7.94 ± 1.46	0.01	0.547	0.011
	Reconstructed 40%	8.12 ± 1.39			0.120
Contrast	100%	8.14 ± 1.23		0.083	
	40%	7.46 ± 1.31	< 0.001	0.150	< 0.001
	Reconstructed 40%	7.78 ± 1.38			0.083
Resolution	100%	8.26 ± 1.23		< 0.001	
	40%	7.41 ± 1.29	< 0.001	0.958	< 0.001
	Reconstructed 40%	7.45 ± 1.28			< 0.001
Diagnostic quality of calcification	100%	8.38 ± 1.20		< 0.001	
	40%	7.39 ± 1.52	< 0.001	0.883	0.002
	Reconstructed 40%	7.26 ± 1.31			< 0.001
Diagnostic quality of mass/ symmetry/architectural distortion	100%	8.02 ± 1.27		0.037	
	40%	7.43 ± 11.34	< 0.001	0.796	0.007
	Reconstructed 40%	7.54 ± 1.37			0.037

40% (low dose), 100% (full-dose), reconstructed 40% (reconstructed full-dose).

*100% image as the reference standard.

† Reconstructed 40% image as the reference standard.

tions in overall image quality, visibility of lesions, contrast, resolution, and diagnostic quality of masses, asymmetry, or architectural distortion were significant as parenchymal density increased ($p < 0.05$), but the diagnostic quality of calcifications did not show such a change ($p = 0.851$) (Table 3). Notably, the diagnostic quality of calcifications was affected along with the dose amount (low dose, full-dose, reconstructed full-dose) and parenchymal density ($p = 0.024$) (Table 3). As lower the dose amount and higher the parenchymal density, its diagnostic quality was decreased. As lesion size increased, better assessments were observed in all criteria ($p < 0.05$) (Table 3), although when assessed with parenchymal density, it lost its significance. In addition, de-noising method was validated by higher SSIM value obtained in our clinical assessment (0.34 ± 0.05 vs. 0.77 ± 0.02 ; $p < 0.001$) (Fig. 4). Comparisons of assessments of overall image quality among readers showed moderate reliability for 40% low-dose, full-dose and 40% reconstructed full-dose images (Supplementary Table 2 in the online-only Data Supplement). Inter-reader comparisons of lesion visibility, contrast, resolution, and diagnostic quality of calcifications also showed moderate to good reliability (Supplementary Table 2 in the online-only

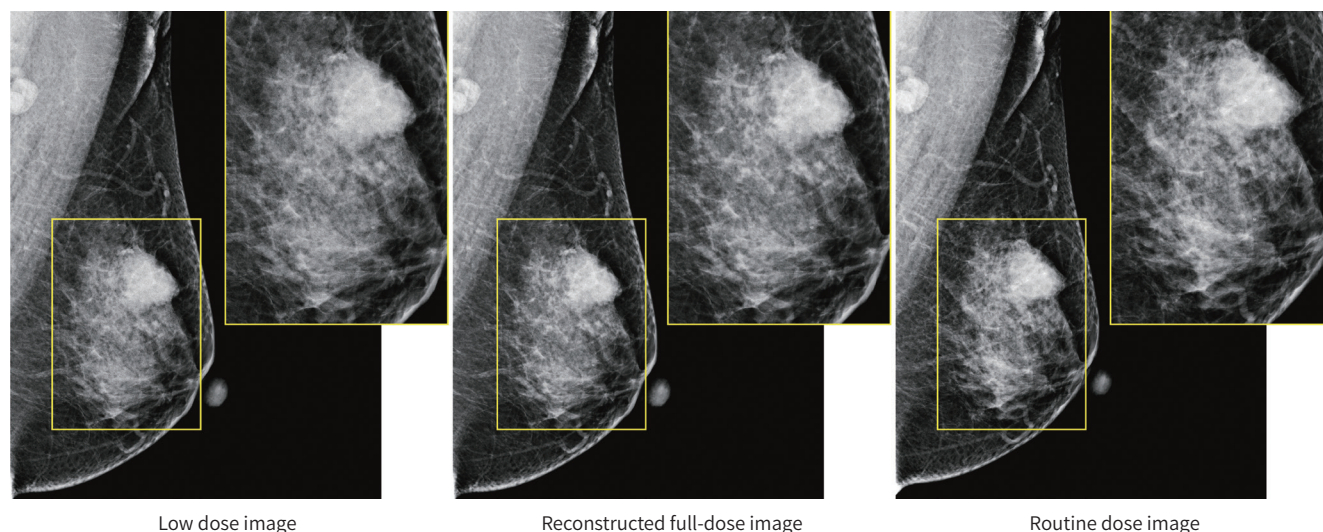
Table 3. Effect of Lesion Size and Parenchymal Density on Assessment Parameters

Parameters	Lesion Size			Parenchymal Density		
	β	p -Value	p -Value*	B	p -Value	p -Value*
Overall image quality	0.08	0.010	0.926	-0.58	<0.001	0.135
Visibility of lesion	0.18	<0.001	0.864	-0.83	<0.001	0.078
Contrast	0.15	<0.001	0.938	-0.83	<0.001	0.214
Resolution	0.12	<0.001	0.967	-0.81	<0.001	0.098
Diagnostic quality of calcification	0.26	<0.001	0.282	0.06	0.851	0.024
Diagnostic quality of mass/asymmetry/architectural distortion	0.12	0.013	0.920	-0.80	<0.001	0.112

$p < 0.05$ means as the lesion size increases, the assessment criteria (i.e. overall image quality) rate increases or decreases by β .

* p value is the interaction p value, with $p < 0.05$ meaning that the assessment criteria (i.e., diagnostic quality of calcification) is affected by two factors, the image (low dose, full-dose, reconstructed full-dose) and increased parenchymal density.

Fig. 4. Reconstructed full-dose (middle) results from a breast cancer patient with mass on mammography, which was rated as higher than routine dose images by two readers, but lower than routine dose images by three readers. The magnified views indicated by the yellow boxes show how the mass margin is better demarcated by application of the network (routine dose image at right).



Data Supplement). Furthermore, the diagnostic quality of masses, asymmetry, or architectural distortion showed moderate to good reliability and assessment of parenchymal density showed excellent reliability (Supplementary Table 2 in the online-only Data Supplement).

DISCUSSION

We demonstrated that application of a de-noising and image reconstruction method to low-dose mammographic images of mastectomy specimens improved their image quality. In particular, 40% reconstructed full-dose images showed an improvement in image quality compared with the original low-dose images, with rates of 20% for masses and 18.3% for calcifications. In our clinical study, application of the de-noising and image reconstruction algorithm reduced noise and achieved better structural preservation with higher SSIM value. Indeed, with 100% full-dose as reference, all parameters were assessed inferior on low-dose images, but after application of our reconstruction algorithm, overall image quality, lesion visibility and contrast were assessed as comparable on the 40% reconstructed full-dose images without significance differences. However, image resolution and the diagnostic quality for calcifications, masses, asymmetry, or architectural distortion on the 40% reconstructed full-dose images were judged significantly lower than on the full-dose images. With the combination of a substantial radiation dose reduction and image reconstruction using our de-noising and image reconstruction method, there is still limitation but our results imply an opportunity to potentially reduce the radiation hazard from digital mammography, which is especially important for high-risk young patients who are exposed to more frequent mammographic examinations.

Mammographic screening reduces mortality through early detection of breast cancer (18). However, the radiation dose associated with mammographic screening has always presented a concern (19-22). When optimizing the mammographic protocol, the radiation dose should be considered together with image quality, and dose optimization is therefore a prerequisite. Because of concern over the potential loss of image quality and its effect on diagnostic accuracy, clinical implementations of mammographic dose reduction have been limited. Two phantom studies suggested that dose reductions of up to 50% may be possible in digital mammography without affecting diagnostic accuracy (23, 24), but these studies used phantoms with homogeneous backgrounds, and phantoms are generally not designed to be sensitive to the large range of dose levels that exist in digital mammography (25). Thus, in our study, a closer match to clinical practice was achieved by using actual mammograms of breast mastectomy specimens with heterogeneous backgrounds. An earlier study using clinical images obtained on a Senographe 2000D mammography system during a screening program (26) indicated that half of the reference average glandular dose level was adequate for maintaining a clinically acceptable image quality. Another study also revealed no statistically significant difference in image quality between full- and half-dose levels, suggesting the possibility of reducing the AEC dose level by half (5). Similarly, we found that a dose reduction to 60% had a minimal effect on lesion detection and image quality. However, at the 40% level there was a gradual degradation in image quality with fewer details being clearly visualized. Mostly, the characterization of calcifications was more dose-dependent at levels below 40%, and the ef-

fect of dose reductions on masses became evident at the 20% dose level, concordant with a previous phantom study (5), which revealed a significant degradation in image quality of calcifications at a 50% dose level and of masses at a 30% dose level. Two studies using phantom and clinical images with calcifications found significant reductions in detection rates at a quarter-dose level but not at a half-dose level (6, 7). Yakabe et al. (6) demonstrated that calcification detection was not affected by a half-dose reduction, but that accuracy deteriorated at a quarter dose level. In agreement, Samei et al. (7) reported that a quarter dose had a significant effect on the screening detection of calcifications. Several previous studies investigated the effect of image processing on calcification detection (27-30), except one found no significant difference in detection due to image processing (29). In our results, the readers perceived masses and calcifications better on reconstructed 40% full-dose images of mastectomy specimens than on the unenhanced 40% dose images. In addition, our study explored the effect of our reconstruction algorithm on actual mammograms of breast mastectomy specimens obtained with varying doses, which means that our results should be more generalizable to clinical practice.

Image de-noising is essential to address the image quality degradation caused by minimizing the radiation dose (31) because it maximizes the restoration of the original image details by reducing noise, as well as by enhancing contrast (32). A previous study revealed that although calcification detection and mass discrimination decreased significantly from a full-dose to a quarter dose, the detection of malignant masses did not appear to be affected (7); thus we hypothesized that a 40% dose reduction would be possible in breast cancer patients. In our clinical study, we used a CNN-based de-noising method to convert 40% low-dose mammographic images to resemble images acquired at the higher standard-dose (AEC level). CNNs were previously used to classify calcifications and masses in computer aided detection (CAD) for mammography (33), while Samala et al. (34) demonstrated that mammography images could be useful for training CNNs to detect masses on DBT. Liu et al. (15) demonstrated a CNN to convert low-dose DBT images to virtual high-dose DBT images while preserving calcification and breast tissue structures, with their conversion technique resulting in a 79% radiation dose reduction and an increased SSIM index. However, in this study (15), the authors simply assessed conspicuity of the lesion with provision of SSIM index. In our clinical study, we demonstrated that the application of a de-noising and image reconstruction algorithm improved the SSIM value. Although resolution and diagnostic quality of calcifications, masses, asymmetry, or architectural distortions were rated as being higher on full-dose images than 40% reconstructed full-dose images, the slight inferiorities in the assessment criteria on the 40% reconstructed full-dose images, in relation to the notable reduction in radiation dose, our results imply that the radiation dose may potentially be decreased with possibility for improvement with reconstruction method, but should be validated with larger number of patients before clinical implementation. In addition, the image enhancement using our reconstruction method was more influential in patients with less-dense background parenchyma and a larger lesion size, although when we assessed the effect of radiation dose and parenchymal density, it lost its statistical significance. Effective de-noising and image reconstruction processing techniques can pave the way for a substantial dose reduction in mammographic screening, while maintaining image consistency.

There are several limitations to this study. First, we analyzed a relatively small number of specimens. As mentioned, we used mastectomy specimens rather than breast phantoms because phantoms have a uniform background with noise characteristics that are not representative of the true anatomical background encountered in clinical practice. Second, we could not evaluate diagnostic performance in our study because the breast cancer cases were already judged before assessment. Third, observer variability was generally the largest source of variability. Previously, Bernhardt et al. (35) studied SNR normalized to AEC levels and showed that a W/Rh node/filter combination was superior. However, we did not evaluate the impact of this combination. In addition, most diagnostic performance experiments involve rating images for the presence of abnormalities, with multiple grades ranging from 4 to 100 being used (36, 37). However, such an approach is limited for many diagnostic tasks when they are performed in a clinical setting. Also due to technical limitation, for both first and second steps of our study, the specimen and breast cancer patients had to be examined with additional paddle compression for additional mammographic images, which may have caused patient motion artifact. In addition, each mammography scanning system had an imaging characteristic, therefore, we had to divide overall dataset into two groups by scanning system. Lastly, in our clinical study, we investigated the effect of dose reduction using a signal-known-exactly paradigm, in which the interpreting radiologists knew the approximate location of a lesion. This strategy eliminated visual searching and kept other sources of variability under control; however, there could have been additional differences if we had used full images with visual searching. A clinical study involving a larger number of patients should be performed to confirm the preliminary conclusions of our study. Moreover, our image-processing technique needs to be shared and external validation needs to be performed.

In summary, the findings of our preliminary study suggest that by application of our CNN image reconstruction algorithm, a reduction in radiation dose to 40% of the standard digital mammography dose is possible, with only modest effects on diagnostic information in breast cancer patients. Whether there is a difference in the ability to accurately discriminate between benign and malignant masses between the full-dose and 40% dose enhanced images still needs to be investigated in more extensive studies. However, there appears to be a substantial potential for our CNN image-processing technique to improve low dose images for better breast cancer detection with reduced radiation hazard.

Supplementary Materials

The online-only Data Supplement is available with this article at <http://doi.org/10.3348/jksr.2020.0152>.

Author Contributions

Conceptualization, K.H.H., Y.J.C.; data curation, all authors; formal analysis, H.S.M., K.E., Y.J.C.; funding acquisition, all authors; investigation, S.B.K., C.N., K.T.H., K.Y.J.; software, K.E., Y.J.C.; supervision, K.H.H.; visualization, K.E., Y.J.C.; writing—original draft, H.S.M., K.E., S.B.K., C.N., K.T.H., K.Y.J., Y.J.C.; and writing—review & editing, H.S.M., K.E., S.B.K., C.N., K.T.H., K.Y.J., Y.J.C.

Conflicts of Interest

Hak Hee Kim has been a Section Editor of the Journal of the Korean Society of Radiology since 2012; however, she was not involved in the peer reviewer selection, evaluation, or decision process of this article. Otherwise, no other potential conflicts of interest relevant to this article were reported.

Funding

This research was supported by Korean Society of Breast Imaging & Korean Society for Breast Screening (KSBI&KSFBS-2017-03).

REFERENCES

1. Tang J, Rangayyan RM, Xu J, El Naqa I, Yang Y. Computer-aided detection and diagnosis of breast cancer with mammography: recent advances. *IEEE Trans Inf Technol Biomed* 2009;13:236-251
2. National Research Council. *Committee to assess health risks from exposure to low level of ionizing radiation. Health risks from exposure to low levels of ionizing radiation*. Washington DC: National Academies Press 2006
3. Friedewald SM, Rafferty EA, Rose SL, Durand MA, Plecha DM, Greenberg JS, et al. Breast cancer screening using tomosynthesis in combination with digital mammography. *JAMA* 2014;311:2499-2507
4. Yaffe MJ, Mainprize JG. Risk of radiation-induced breast cancer from mammographic screening. *Radiology* 2011;258:98-105
5. Svahn T, Hemdal B, Ruschin M, Chakraborty DP, Andersson I, Tingberg A, et al. Dose reduction and its influence on diagnostic accuracy and radiation risk in digital mammography: an observer performance study using an anthropomorphic breast phantom. *Br J Radiol* 2007;80:557-562
6. Yakabe M, Sakai S, Yabuuchi H, Matsuo Y, Kamitani T, Setoguchi T, et al. Effect of dose reduction on the ability of digital mammography to detect simulated microcalcifications. *J Digit Imaging* 2010;23:520-526
7. Samei E, Saunders RS Jr, Baker JA, Delong DM. Digital mammography: effects of reduced radiation dose on diagnostic performance. *Radiology* 2007;243:396-404
8. Sun M, Star-Lack JM. Improved scatter correction using adaptive scatter kernel superposition. *Phys Med Biol* 2010;55:6695-6720
9. Mencattini A, Salmeri M, Lojacono R, Frigerio M, Caselli F. Mammographic images enhancement and denoising for breast cancer detection using dyadic wavelet processing. *IEEE Trans Instrum Meas* 2008;57:1422-1430
10. Gorgel P, Sertbas A, Ucan ON. A wavelet-based mammographic image denoising and enhancement with homomorphic filtering. *J Med Syst* 2010;34:993-1002
11. Vikhe PS, Thool VR. Contrast enhancement in mammograms using homomorphic filter technique. Proceedings of the 2016 International Conference on Signal and Information Processing (IConSIP); 2016 October 6-8; Nanded, India: IEEE; 2016:1-5
12. Golkov V, Dosovitskiy A, Sperl JI, Menzel MI, Czisch M, Samann P, et al. Q-space deep learning: twelve-fold shorter and model-free diffusion MRI scans. *IEEE Trans Med Imaging* 2016;35:1344-1351
13. Zhu B, Liu JZ, Cauley SF, Rosen BR, Rosen MS. Image reconstruction by domain-transform manifold learning. *Nature* 2018;555:487-492
14. Sun Y, Liu X, Cong P, Li L, Zhao Z. Digital radiography image denoising using a generative adversarial network. *J Xray Sci Technol* 2018;26:523-534
15. Liu J, Zarshenas A, Qadir A, Wei Z, Yang L, Fajardo L, et al. Radiation dose reduction in digital breast tomosynthesis (DBT) by means of deep-learning-based supervised image processing. Proceedings of the SPIE 10574, Medical Imaging: image processing; 2018 Feb 10-15; Houston, TX, USA: SPIE; 2018
16. Wang Z, Bovik AC, Sheikh HR, Simoncelli EP. Image quality assessment: from error visibility to structural similarity. *IEEE Trans Image Process* 2004;13:600-612
17. American College of Radiology. *ACR BI-RADS breast imaging and reporting data system: breast imaging atlas*. 4th ed. Reston: American College of Radiology 2003
18. Nyström L, Andersson I, Bjurstram N, Frisell J, Nordenskjöld B, Rutqvist LE. Long-term effects of mammography screening: updated overview of the Swedish randomised trials. *Lancet* 2002;359:909-919
19. Parker MS, Hui FK, Camacho MA, Chung JW, Broga DW, Sethi NN. Female breast radiation exposure during CT pulmonary angiography. *AJR Am J Roentgenol* 2005;185:1228-1233
20. Doody MM, Lonstein JE, Stovall M, Hacker DG, Luckyanov N, Land CE. Breast cancer mortality after diagnostic radiography: findings from the U.S. Scoliosis Cohort Study. *Spine (Phila Pa 1976)* 2000;25:2052-2063
21. Law J, Faulkner K. Two-view screening and extending the age range: the balance of benefit and risk. *Br J Radiol* 2002;75:889-894

22. Law J, Faulkner K. Concerning the relationship between benefit and radiation risk, and cancers detected and induced, in a breast screening programme. *Br J Radiol* 2002;75:678-684
23. Hemdal B, Bay TH, Bengtsson G, Gangeskar L, Martinsen AC, Pedersen K, et al. *Comparison of screen-film, imaging plate and direct digital mammography with CD phantoms*. In Peitgen HO, ed. *Digital mammography*. Berlin, Heidelberg: Springer 2003:105-107
24. Gennaro G, Katz L, Souchay H, Alberelli C, di Maggio C. Are phantoms useful for predicting the potential of dose reduction in full-field digital mammography? *Phys Med Biol* 2005;50:1851-1870
25. Huda W, Sajewicz AM, Ogden KM, Scalzetti EM, Dance DR. How good is the ACR accreditation phantom for assessing image quality in digital mammography? *Acad Radiol* 2002;9:764-772
26. Hemdal B, Andersson I, Grahn A, Håkansson M, Ruschin M, Thilander-Klang A, et al. Can the average glandular dose in routine digital mammography screening be reduced? A pilot study using revised image quality criteria. *Radiat Prot Dosimetry* 2005;114:383-388
27. Cole EB, Pisano ED, Zeng D, Muller K, Aylward SR, Park S, et al. The effects of gray scale image processing on digital mammography interpretation performance. *Acad Radiol* 2005;12:585-595
28. Zanca F, Jacobs J, Van Ongeval C, Claus F, Celis V, Geniets C, et al. Evaluation of clinical image processing algorithms used in digital mammography. *Med Phys* 2009;36:765-775
29. Uematsu T. Detection of masses and calcifications by soft-copy reading: comparison of two postprocessing algorithms for full-field digital mammography. *Jpn J Radiol* 2009;27:168-175
30. Cole EB, Pisano ED, Kistner EO, Muller KE, Brown ME, Feig SA, et al. Diagnostic accuracy of digital mammography in patients with dense breasts who underwent problem-solving mammography: effects of image processing and lesion type. *Radiology* 2003;226:153-160
31. Chatterjee P, Milanfar P. Is denoising dead? *IEEE Trans Image Process* 2010;19:895-911
32. Cesarelli M, Bifulco P, Cerciello T, Romano M, Paura L. X-ray fluoroscopy noise modeling for filter design. *Int J Comput Assist Radiol Surg* 2013;8:269-278
33. Chan HP, Lo SC, Sahiner B, Lam KL, Helvie MA. Computer-aided detection of mammographic microcalcifications: pattern recognition with an artificial neural network. *Med Phys* 1995;22:1555-1567
34. Samala RK, Chan HP, Hadjiiski L, Helvie MA, Wei J, Cha K. Mass detection in digital breast tomosynthesis: deep convolutional neural network with transfer learning from mammography. *Med Phys* 2016;43:6654
35. Bernhardt P, Mertelmeier T, Hoheisel M. X-ray spectrum optimization of full-field digital mammography: simulation and phantom study. *Med Phys* 2006;33:4337-4349
36. Riedl CC, Jaromi S, Floery D, Pfarl G, Fuchsjaeger MH, Helbich TH. Potential of dose reduction after marker placement with full-field digital mammography. *Invest Radiol* 2005;40:343-348
37. Chesters MS. Human visual perception and ROC methodology in medical imaging. *Phys Med Biol* 1992;37:1433-1476

딥러닝 알고리즘을 이용한 저선량 디지털 유방 촬영 영상의 복원: 예비 연구

하수민^{1,2} · 김학희^{3*} · 강은희⁴ · 서보경⁵ · 최나미⁶ · 김태희⁷ · 구유진⁸ · 예종철⁴

목적 깊은 컨볼루션 신경망 기법을 결합한 영상 잡음 제거 알고리즘을 개발하고 이를 응용하여 저선량 유방 촬영 영상으로 유방암을 진단하는 데 그 효능을 조사하고자 한다.

대상과 방법 6명의 유방 영상 전문가가 전향적 연구에 참여하였다. 모든 영상 전문의는 병변 감지를 위해 저선량 영상을 독립적으로 평가하고 정성적 척도를 사용하여 진단 품질을 평가하였다. 영상 잡음 제거 알고리즘을 적용한 후, 동일한 영상 전문가가 병변 감지 가능성과 영상 품질에 대한 평가를 하였다. 임상 적용을 위해 동일한 영상 전문가가 병변 유형과 위치에 대한 합의 결정 후, 저선량 영상, 재구성된 영상, 기존 선량 영상을 무작위 순서로 제시하여 평가하였다.

결과 전 절제 표본의 저선량 영상을 참조로 40% 재구성된 영상에서 병변이 더 잘 인식되었다. 임상 적용단계에서 40% 재구성된 영상과 비교하여, 기존 선량 영상이 해상도($p < 0.001$), 석회에 대한 진단 품질($p < 0.001$), 유방 종괴, 비대칭, 구조왜곡의 진단 품질($p = 0.037$)에 대해 더 높은 평균값을 보였다. 40% 재구성된 영상은 100% 영상과 비교 시 전반적 화질($p = 0.547$), 병변의 가시성($p = 0.120$), 대조도($p = 0.083$)에서 비슷한 성적을 보였으며 유의미한 차이도 보이지 않았다.

결론 깊은 컨볼루션 신경망 기법을 결합한 효과적인 잡음 제거 및 영상 재구성 처리 알고리즘은 유방 촬영의 상당한 선량 감소를 위한 길을 열어 유방암 진단을 가능하게 할 것이다.

¹중앙대학교 의과대학 중앙대학교병원 영상의학과,

²서울대학교 의과대학 서울대학교병원 영상의학과,

³울산대학교 의과대학 서울아산병원 영상의학과,

⁴한국과학기술원 바이오 및 뇌공학과,

⁵고려대학교 의과대학 고려대학교 안산병원 영상의학과,

⁶건국대학교 의과대학 건국대학교병원 영상의학과,

⁷아주대학교 의과대학 아주대학교병원 영상의학과,

⁸가톨릭관동대학교 국제성모병원 영상의학과