의학박사 학위논문

딥러닝 알고리즘을 이용한 영상복원을 통한 저선량 디지털 유방촬영술 플랫폼 개발

Development of Low Dose Digital Mammography Platform by Image Reconstruction using Deep Learning Algorithm

울 산 대 학 교 대 학 원

의 학 과

하 수 민

Development of Low Dose Digital Mammography Platform by Image Reconstruction using Deep Learning Algorithm

지 도 교 수　　김　학　희

이 논문을 의학박사 학위 논문으로 제출함

2018 년　12 월

울 산 대 학 교 대 학 원

의　　학　　과

하　수　민

하수민의 의학박사학위 논문을 인준함

심사위원장	차 주 희	(인)
심사위원	김 학 희	(인)
심사위원	도 경 현	(인)
심사위원	이 종 원	(인)
심사위원	문 우 경	(인)

울 산 대 학 교 대 학 원

2018 년 12 월

English Abstract

Purpose:

First, we analyze qualitatively the detection and characterization of breast cancer on low dose mammographic images and find the acceptable low dose levels compared to the full dose image. Second, we use the convolutional neural network (CNN) de-noising method to convert the acceptable low dose images and find the lowest dose to reconstruct the acceptable image quality compared to synthesized full dose image.

Materials and methods:

This prospective study was approved by institutional review board. The true full dose level was determined through AEC, and images were subsequently acquired at 5 different radiation dose levels (80% of AEC, 60%, 40%, 20%, and 10%) of breast cancer mastectomy specimen with digital mammography. For the first purpose, five radiologists evaluated low dose images by comparison to the reference true full dose image and scaled as equivalent, acceptable or unacceptable for each dose level. For the second purpose, the same five radiologists blindly rated three images (low dose, synthesized full dose, and true full dose).

We analyzed the trend using Mantel-Haenszel statistic. In addition, we compared the quantitative assessment using McNemar's or marginal homogeneity test.

Results:

Mass and calcification detection rate decreased substantially at 10% reduced dose. Regarding characterization, the 'not acceptable' rate of mass and calcification increased at 10% (86.3%) and 20% (83.8%) respectively. The 80% and 60% images were equivalent to full dose images regarding both mass and calcification detection and characterization. The synthetized images showed high detection rate of 87.4-90.0% and 96.8-100.0% for 20% and 40% respectively. There was significantly higher 'equivalent' image quality for both mass and calcification at synthesized 40% dose level (65.3% and 65.0% respectively) compared to synthesized 20% dose level (41.1% and 20.8% respectively) ($p<0.001$). Compared to low dose images, synthesized image quality improved in 40% dose whereas at 20% dose, it did not increase in synthesized images ($p<0.001$).

Conclusion:

There is a potential for modest dose reduction retaining diagnostic information with de-

noising algorithm in digital mammography. Our results provide a baseline for future studies on reducing the radiation dose with lesion preservation on mammography.

Key Words

Mammography

Deep Learning

Breast neoplasms

Radiation

CONTENTS

LIST of TABLES AND FIGURES

Introduction

The goal of mammography is early detection of suspicious mass and calcification and is reliable method for screening breast cancer (1). One of the limitations of mammography screening is a radiation exposure (2). In particular, when a patient needs to be repeatedly monitored, it is important to minimize the radiation dose. Indeed, mammography has automatic exposure control (AEC) system which chooses an appropriate current of X-ray to get an appropriate image. There have been many attempts to examine the relationship between reduced radiation doses and accuracy in detection of lesions on digital mammography (3-5). Indeed, it is hard to achieve optimal trade-off between lesion conspicuity and noise reduction in medical images. It is a challenge to reduce noise and artifact while maintaining important diagnostic information.

The several decades of research into statistical learning methods, in particular, deep learning (6) with many Artificial Intelligence (AI) systems claim human or even superhuman performance in variety of tasks. Machine learning and deep learning method have shown great potential in medical imaging applications such as reconstruction and de-noising (7, 8).

Recently, Liu et al (9) had investigated deep-learning based supervised image processing technique for radiation dose reduction in digital breast tomosynthesis (DBT). They trained network with quarter dose images and corresponding high dose images of breast phantom acquired with a DBT system, demonstrating a 79.0% dose reduction with application of deep-learning technique. Another study using DBT, noise reduction techniques have been developed, but it smooth out image details and subtle patterns such as calcifications (10).

Our purposes were two-fold. First, we analyzed qualitatively regarding detection and characterization of lesion on various low dose mammographic images and find the lower dose than AEC level with maintenance of equivalent image quality. Second, we used a convolutional neural network (CNN) de-noising method based on unsupervised learning using cyclic consistency. This method converts lower doses less than 'equivalent dose level' from the first step of our study and reconstructs to synthesized full dose image with equivalent quality to the true full dose, and finally investigate the lowest acceptable mammography dose level with this de-nosing method. To our knowledge, there has been no study applying deep learning technique to acquire synthesized full dose image while

maintaining the image quality and contrast information of full dose image with radiation dose reduction using digital mammography.

Materials and Methods

Study design

This prospective study from two institutions was approved by institutional review board. The true full dose level was determined through AEC, and images were subsequently acquired at 5 different radiation dose levels (80% of AEC, 60%, 40%, 20%, 10%) of breast cancer mastectomy specimen with mammography (Selenia Dimension, Hologic, Inc Bedford, MA (n=8) or Senographe DS or Senographe Essential scanner; GE Healthcare, Milwaukee, WI (n=23)) from two institutions. We used the mastectomy specimen to use AEC dose and to simulate the real clinical practice rather than using the breast phantom. The radiation doses were altered by changing tube-current-time product (mAs), while tube voltage was fixed (range, 26-30kVp for Selenia Dimension and 25-29kVp for Senograph DS/Essential). The reduced dose images were obtained by changing tube current-time product manually, while the tube voltage was maintained in each exam. Three cases were excluded due to poor image quality with unavoidable artifact at low radiation dose; finally 28 cases were included in analysis.

Image Interpretation

Five breast radiologists with 5-20 years of experience were involved in image interpretation regarding detection and characterization of mass (n=12), calcification (n=7), mass with calcification (n=9). For the first purpose of our study, to assess subjective quality, five radiologists separately evaluated low dose images (80%, 60%, 40%, 20%, and 10%) compared to the reference true full dose image blinded to the dose level, and scaled as equivalent, acceptable or unacceptable for each dose level. Each observer interpreted the images on DICOM and was allowed to alter the optimal window/level settings. For the second purpose, after deep learning method application to low dose images, the same five radiologists separately and blindly rated three images (low dose, synthesized full dose, and true full dose). Specifically, the readers were independently presented the full dose and randomized version of low dose and synthesized images and asked to rate their quality.

Training, testing and validation of CNN de-noising method

The network was trained with input low dose images and true full dose images. Since there is a mismatch between the input and the true full dose image due to the potential deformation of the specimen during multiple acquisitions, we proposed de-noising method based on unsupervised learning using cyclic consistency. The proposed framework is described in Fig.1. We trained two generators (Network G and Network F) and two discriminators (Network Dx and Network Dy) (11, 12). We used three losses to train the networks. First, we employed adversarial loss with discriminator which distinguishes images are real or fake. The adversarial loss is that forces the generated images to be indistinguishable from real image. However, this adversarial loss alone cannot guarantee the learned function to map an individual input to a desired output. Thus, secondly we applied the cyclic loss to guarantee an inverse relation described in Fig. 1. Finally, identity loss was added to enforce the network does not generate any artificial components.

Loss functions are as follows:

$$\mathcal{L}(\mathrm{G},\mathrm{F},\mathrm{D_x},D_y) = \mathcal{L}_{adversarial}(G,F,X,Y) + \mathcal{L}_{adversarial}(F,G,Y,X) + \mathcal{L}_{cyclic}(G,F) + \mathcal{L}_{identity}(G,F)$$

$$\mathcal{L}_{\mathrm{adversarial}}(G,F,X,Y) = \mathbb{E}_{x\sim P_x}\left[\log\left(1 - D_y\big(G(x)\big)\right)\right] + \mathbb{E}_{y\sim P_y}[\log D_y(y)]$$

$$\mathcal{L}_{\mathrm{cyclic}}(G,F) = ||F\big(G(x)\big) - x||_1 + ||G\big(F(y)\big) - y||_1$$

$$\mathcal{L}_{\mathrm{identity}}(G,F) = ||G(y) - y||_1 + ||F(x) - x||_1$$

Training was performed by minimizing the loss functions. We used the ADAN optimization method to train the all networks with $\beta_1 = 0.5$ and $\beta_2 = 0.999$. The number of epochs was 300, which was divided into two phases to control the learning rate during training. In the first 100 epochs, we set the learning rate to 0.00002, and linearly decreased it to zero over the next 200 epochs. The size of patch was 56×56 and the size of mini-batch was 10. Kernels were initialized randomly from a Gaussian distribution. We updated the generator and discriminator with the ratio 3/1 in iterations. We normalized the intensity of the input low dose images and the target full dose images using the maximum intensity value of the input images. We trained the networks twice with one half of the dataset and the other half of the dataset. The proposed method was implemented in Python with the PyTorch (13) and NVIDIA GeForce GTX 1080 Ti GPU was used to train and the test the network. When we tested the network, we applied the trained generator G to reduce the noises and we did soft-thresholding to the image differences between input and output. After that, we added it to the output images to remove the unrealistic components.

For subset of data, we validated the CNN de-noising method by measuring

structural similarity index (SSIM) (14), which is quantitative value to evaluate the image quality, by noise simulation to full dose and evaluate its value compared to full dose image. When we make simulated low dose image, we added Poisson noise to the routine dose image. The signal to noise ratio (SNR) of simulated low dose images was 17dB. SSIM is a widely accepted image quality measure that overcomes the limitation of conventional SNR which lack spatial information in evaluation. The SSIM index is represented by

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

where μ_x μ_y σ_x^2 σ_y^2 σ_{xy} are the average, the variance and the covariance of x and y. $c_1 = (k_1 L)^2$, $c_2 = (k_2 L)^2$ are two variables to stabilized the division with weak denominator;L is the dynamic range of the pixel-values, $k_1 = 0.01$ and $k_2 = 0.03$ by default.

Statistical Analysis

To assess the qualitative analysis from five radiologists, the mean percentage of each assessment scale was calculated for lesion type and each dose. We analyzed the trend

using Mantel-Haenszel statistic. In addition, after deep learning method application, we compared the quantitative assessment between true full dose, low dose versus synthesized image using McNemar's or marginal homogeneity test. To show the quantitative performance of the proposed deep learning method, we computed SSIM measurement and compared using Wilcoxon signed ranked test. Calculations for statistical analyses were performed using SPSS software (version 14.0, Statistical Package for the Social Sciences, Chicago, IL). P values less than .05 were considered to indicate statistically significance differences.

Results

Diagnostic quality evaluated from radiologists

Detailed qualitative metrics assessed by the five radiologists are shown in Table 1. Similar trend in assessment percentage regarding mass detection and characterization were observed, except for the mass characterization ($p=0.173$). Mass and calcification detection rate decreased substantially at 10% dose (76.8% and 60.8% respectively). Regarding characterization, the 'not acceptable' rate of mass and calcification notably increased at 10% (86.3%) and 20% (83.8%) respectively. The 80% and 60% showed equivalent image quality compare to the full dose regarding both mass and calcification detection and characterization. Since the lesion detection was maintained till 20% dose and the characterization of mass and calcification gradually decreased from 20% and 40% dose level respectively, the radiologists made consensus to apply de-noising method for the low dose level of 40% and 20%, the second purpose of our study.

Application of de-noising method

First, we trained and validated our de-noising method based on unsupervised

learning using cyclic consistency using simulated images with addition of Poisson noise for 8 subsets of our data. Table 2 shows the improvement of SSIM value ($p=0.012$). The higher SSIM value indicates that our technique can not only substantially reduce noise but also achieve better structure preservation performance of image details and subtle patterns such as calcifications.

When we compared the synthetized images with true full dose images, the calcification and mass showed high detection rate of 87.4%, 90.0% and 96.8%, 100.0% for 20% and 40% respectively. However, the 'no detection' rate for both lesion types was significantly higher at 20% synthetized image compared to 40% synthetized image (12.6% vs 3.2%, $p=0.003$ for mass, 10.0% vs 0.0% $p<0.001$ for calcification) (Table 3, Fig. 2).

Regarding lesion characterization, when we compared true full dose images to synthetized images, there was significantly higher 'equivalent' image quality for both mass and calcification at 40% dose level (65.3% and 65.0% respectively) compared to 20% dose level (41.1% and 20.8% respectively) ($p<0.001$). Compared to low dose images, synthesized image quality improved in 40% dose (20.0% for mass and 18.3% for

calcification) whereas at 20% dose, it did not increase in synthesized images (25.3% for mass and 19.2% for calcification) ($p<0.001$) (Table 4, Fig. 3).

Table 1. Detection and characterization according to lesion type on low dose images by five observers

Detection	Dose	Assessment	Observer 1	Observer 2	Observer 3	Observer 4	Observer 5	Mean %	P value*
Lesion Type									
Mass	80%	No detection	0.0 (0/20)	0.0 (0/20)	0.0 (0/20)	0.0 (0/20)	0.0 (0/20)	0.0	0.010
		Detection	100.0 (20/20)	100.0 (20/20)	100.0 (20/20)	100.0 (20/20)	100.0 (20/20)	100.0	
	60%	No detection	0.0 (0/20)	0.0 (0/20)	0.0 (0/20)	0.0 (0/20)	5.0 (1/20)	1.0	
		Detection	100.0 (20/20)	100.0 (20/20)	100.0 (20/20)	100.0 (20/20)	95.0 (19/20)	99.0	
	40%	No detection	5.0 (1/20)	0.0 (0/20)	0.0 (0/20)	0.0 (0/20)	5.0 (1/20)	2.0	
		Detection	95.0 (19/20)	100.0 (20/20)	100.0 (20/20)	100.0 (20/20)	95.0 (19/20)	98.0	
	20%	No detection	5.0 (1/20)	5.0 (1/20)	5.0 (1/20)	0.0 (0/20)	5.0 (1/20)	4.0	
		Detection	95.0 (19/20)	95.0 (19/20)	95.0 (19/20)	100.0 (20/20)	95.0 (19/20)	96.0	

	10%	No detection	26.3(5/19)	15.8(3/19)	26.3(5/19)	26.3(5/19)	21.1(4/19)	23.2	
		Detection	73.7(14/19)	84.2(16/19)	73.7(14/19)	73.7(14/19)	78.9(15/19)	76.8	
Calcification	80%	No detection	0.0 (0/26)	0.0 (0/26)	0.0 (0/26)	0.0 (0/26)	0.0 (0/26)	0.0	0.001
		Detection	100.0 (26/26)	100.0 (26/26)	100.0 (26/26)	100.0 (26/26)	100.0 (26/26)	100.0	
	60%	No detection	3.9 (1/26)	0.0 (0/26)	0.0 (0/26)	3.9 (1/26)	0.0 (0/26)	1.5	
		Detection	96.1 (25/26)	100.0 (26/26)	100.0 (26/26)	96.1 (25/26)	100.0 (26/26)	98.5	
	40%	No detection	7.7 (2/26)	0.0 (0/26)	0.0 (0/26)	3.9 (1/26)	0.0 (0/26)	2.3	
		Detection	92.3 (24/26)	100.0 (26/26)	100.0 (26/26)	96.1 (25/26)	100.0 (26/26)	97.7	
	20%	No detection	19.2 (5/26)	15.4 (4/26)	0.0 (0/26)	11.5 (3/26)	7.7 (2/26)	10.8	
		Detection	80.8 (21/26)	84.6 (22/26)	100.0 (26/26)	88.5 (23/26)	92.3 (24/26)	89.2	
	10%	No detection	44.0 (11/25)	36.0 (9/25)	28.0 (7/25)	44.0 (11/25)	44.0 (11/25)	39.2	
		Detection	56.0 (14/25)	64.0 (16/25)	72.0 (18/25)	56.0 (14/25)	56.0 (14/25)	60.8	

Characterization	Dose	Assessment	Observer 1	Observer 2	Observer 3	Observer 4	Observer 5	Mean %	P value*
Lesion Type									
Mass	80%	Not acceptable	0.0 (0/20)	0.0 (0/0)	0.0 (0/0)	0.0 (0/0)	0.0 (0/0)	0.0	0.173
		Acceptable	25.0 (5/20)	5.0 (1/20)	0.0 (0/0)	10.0 (2/20)	5.0 (1/20)	9.0	
		Equivalent	75.0 (15/20)	95.0 (19/20)	100.0 (20/20)	90.0 (19/20)	95.0 (19/20)	91.0	
	60%	Not acceptable	0.0 (0/0)	0.0 (0/0)	0.0 (0/0)	0.0 (0/0)	0.0 (0/0)	0.0	
		Acceptable	90.0 (18/20)	30.0 (6/20)	45.0 (9/20)	50.0 (10/20)	30.0 (6/20)	49.0	
		Equivalent	10.0 (2/20)	70.0 (14/20)	55.0 (11/20)	50.0 (10/20)	70.0 (14/20)	51.0	
	40%	Not acceptable	20.0 (4/20)	0.0 (0/0)	15.0 (3/20)	20.0 (4/20)	5.0 (1/20)	12.0	
		Acceptable	80.0 (16/20)	90.0 (18/20)	70.0 (14/20)	75.0 (15/20)	60.0 (12/20)	75.0	
		Equivalent	0.0 (0/0)	10.0 (2/20)	15.0 (3/20)	5.0 (1/20)	35.0 (7/20)	13.0	
	20%	Not acceptable	60.0 (12/20)	50.0 (10/20)	50.0 (10/20)	35.0 (7/20)	35.0 (7/20)	46.0	

		Acceptable	40.0 (8/20)	45.0 (9/20)	45.0 (9/20)	65.0 (13/20)	50.0 (10/20)	49.0	
		Equivalent	0.0 (0/0)	5.0 (1/20)	5.0 (1/20)	0.0 (0/0)	15.0 (3/20)	5.0	
	10%	Not acceptable	89.5 (17/19)	89.5 (17/19)	100.0 (19/19)	68.4 (13/19)	84.2 (16/19)	86.3	
		Acceptable	10.5 (2/19)	10.5 (2/19)	0.0 (0/0)	31.6 (6/19)	15.8 (3/19)	12.6	
		Equivalent	0.0 (0/0)	0.0 (0/0)	0.0 (0/0)	0.0 (0/0)	0.0 (0/0)	1.0	
Calcification	80%	Not acceptable	0.0 (0/0)	0.0 (0/0)	0.0 (0/0)	0.0 (0/0)	0.0 (0/0)	0.0	0.001
		Acceptable	30.8 (8/26)	7.7 (2/26)	3.8 (1/26)	38.5 (10/26)	7.7 (2/26)	17.7	
		Equivalent	69.2 (18/26)	92.3 (24/26)	96.2 (25/26)	61.5 (16/26)	92.3 (24/26)	82.3	
	60%	Not acceptable	26.9 (7/26)	0.0 (0.0)	0.0 (0.0)	7.7 (2/26)	0.0 (0.0)	7.0	
		Acceptable	61.5 (16/26)	50.0 (13/13)	65.4 (17/26)	73.1 (19/26)	57.7 (15/26)	61.5	
		Equivalent	11.5 (3/26)	50.0 (13/13)	34.6 (9/26)	19.2 (5/26)	42.3 (11/26)	31.5	
	40%	Not acceptable	65.4 (17/26)	23.1 (6/26)	30.8 (8/26)	61.5 (16/26)	30.8 (8/26)	42.3	

		Acceptable	34.6 (9/26)	69.2 (8/26)	61.5 (16/26)	38.5 (10/26)	61.5 (16/26)	53.1	
		Equivalent	0.0 (0/0)	7.7 (2/26)	7.7 (2/26)	0.0 (0/0)	7.7 (2/26)	4.6	
	20%	Not acceptable	96.2 (25/26)	80.8 (21/26)	76.9 (20/26)	88.5 (23/26)	76.9 (20/26)	83.8	
		Acceptable	3.8 (1/26)	19.2 (5/26)	19.2 (5/26)	11.5 (3/26)	19.2 (5/26)	14.6	
		Equivalent	0.0 (0/0)	0.0 (0/0)	3.9 (1/26)	0.0 (0/0)	3.9 (1/26)	1.5	
	10%	Not acceptable	96.0 (24/25)	92.0 (23/25)	100.0 (25/25)	96.4 (24/25)	100.0 (25/25)	96.8	
		Acceptable	4.0 (1/25)	8.0 (2/25)	0.0 (0/0)	4.0 (1/25)	0.0 (0/0)	3.2	
		Equivalent	0.0 (0/0)	0.0 (0/0)	0.0 (0/0)	0.0 (0/0)	0.0 (0/0)	0.0 (0/0)	

*agreement trend test

Table 2. Structural Similarity Index Measure (SSIM) values for de-noising algorithm validation

No.	Input	Output	P value
1	0.55255	0.70162	0.012
2	0.5962	0.80514	
3	0.66913	0.83729	
4	0.68076	0.85729	
5	0.69183	0.86143	
6	0.66412	0.81149	
7	0.59455	0.64519	
8	0.60425	0.81822	

Table 3. Comparison of detection of lesion on full dose and synthesized full dose mammography

Lesion Type	Assessment	Mean %		P value
		Synthesized 20% Dose	Synthesized 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	

*Full dose as the reference standard

Table 4. Comparison of lesion characterization of full dose, low dose and synthesized full dose mammography

Lesion type	Assessment	Mean %		P value
Mass		Synthesized 20% Dose	Synthesized 40% Dose	<0.001
	Decrease	54.7	14.7	
	Equivalent	41.1	65.3	
	Improve	4.2	20.0	
				<0.001
	Decrease[+]	25.3	4.2	
	Equivalent[+]	62.1	75.8	
	Improve[+]	12.6	20.0	
Calcification				<0.001
	Decrease	78.3	26.7	
	Equivalent	20.8	65.0	
	Improve	0.8	8.3	
				<0.001
	Decrease[+]	19.2	5.8	
	Equivalent[+]	70.8	75.8	
	Improve[+]	10.0	18.3	

[+]Reference standard is low dose (20%, 40%) images.

Figure 1. Our network contains two functions G: X→Y, and F: Y→X, We trained two generators (Network G and Network F) and two discriminators (Network Dx and Network Dy). We used three losses to train the networks. First, we employed adversarial loss with discriminator which distinguishes images are real or fake. Secondly we applied the *cyclic loss* to guarantee an inverse relation. Finally, identity loss was added to enforce the network does not generate any artificial components.

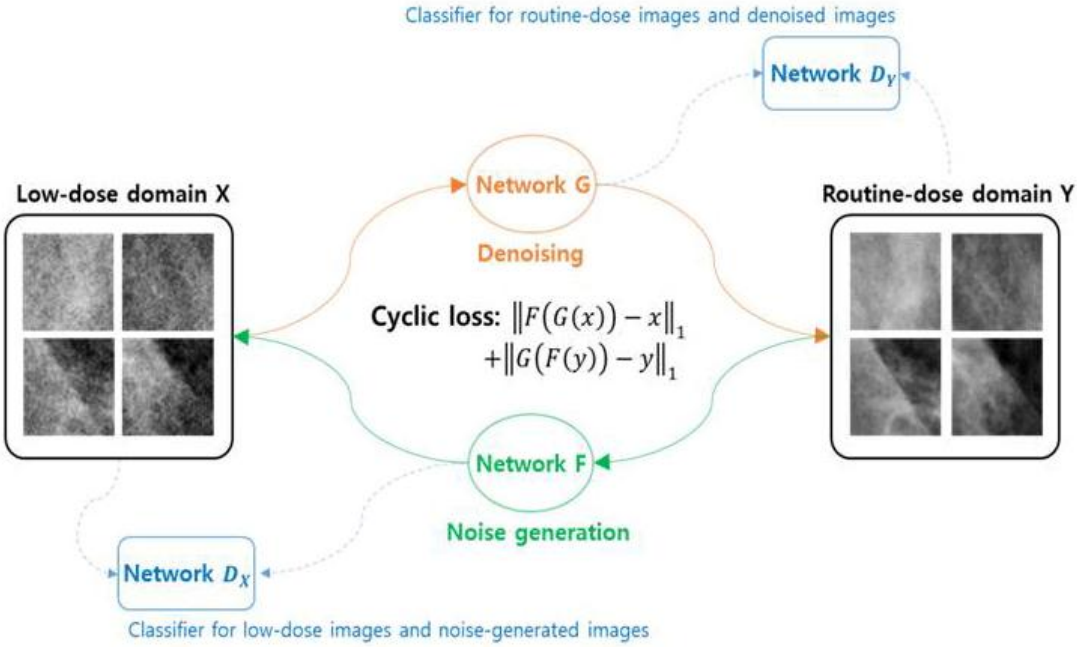


Figure 2. Comparison of five radiologist's average ratings of lesion detection on synthesized images compared to full dose images. The synthesized 40% image shows highest detection rate for mass (96.8%) and calcification (100.0%).

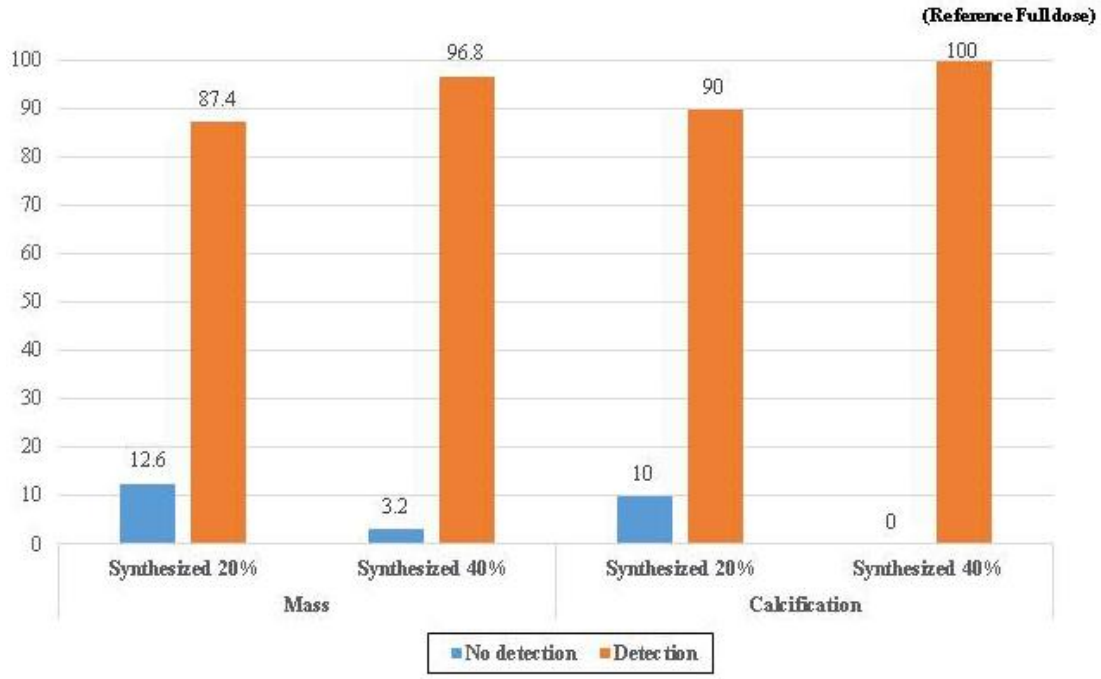
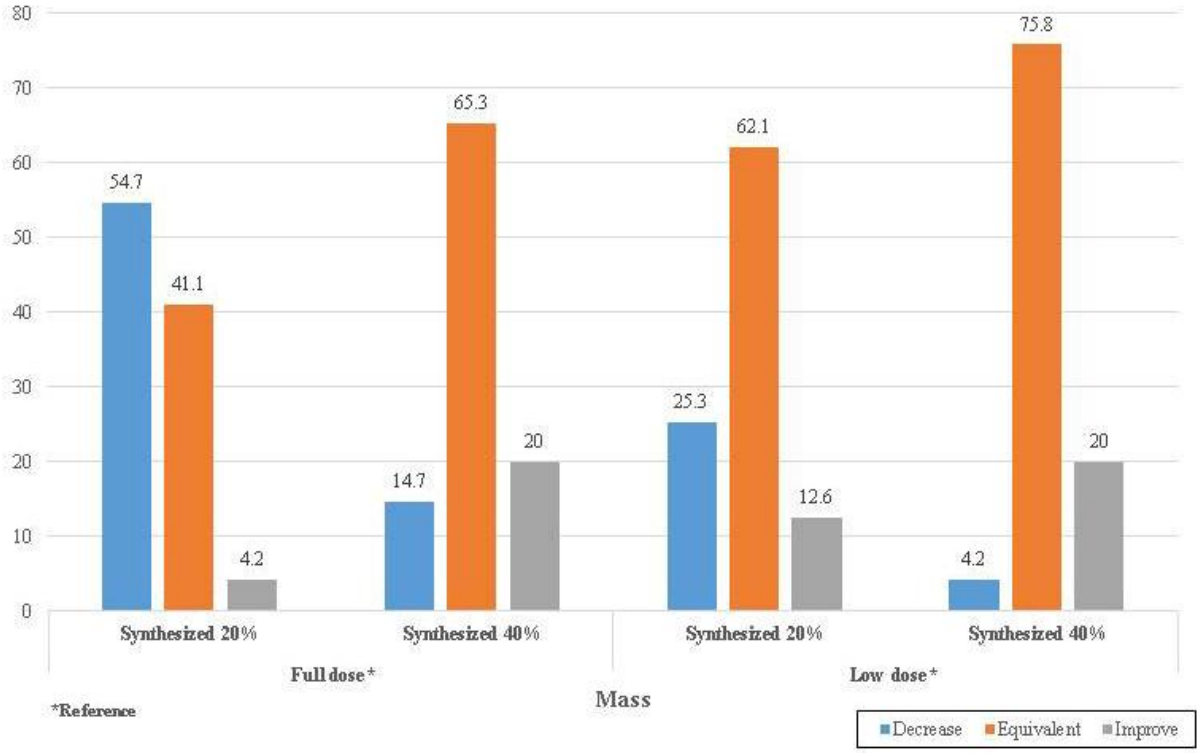
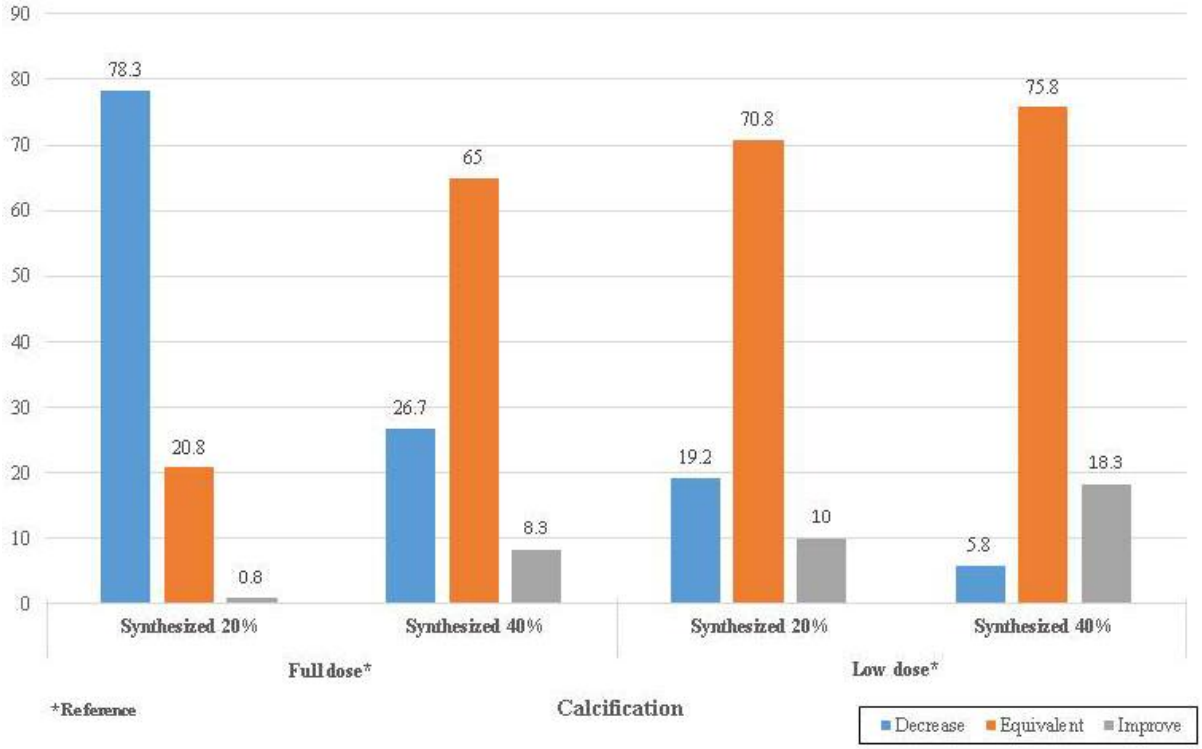


Figure 3. Bar graph shows the effect of synthesized images on the characterization of mass (a) and calcification (b) with reference to true full dose and corresponding low dose (20% and 40%) images. Synthesized 40% images show significantly higher proportion of 'equivalent' and 'improved' quality for mass and calcification. Synthesized 20% images show significantly higher proportion of 'decreased' quality for mass and calcification.

a



b


90
80
70
60
50
40
30
20
10
0
78.3
20.8
0.8
26.7
65
8.3
19.2
70.8
10
5.8
75.8
18.3
Synthesized 20%
Synthesized 40%
Synthesized 20%
Synthesized 40%
Full dose*
Low dose*
*Reference
Calcification
Decrease
Equivalent
Improve

Discussion

With application of de-noising method based on unsupervised learning using cyclic consistency, we demonstrated that it can successfully improve the reduced dose image quality. To be clear, the synthesized image quality was mostly equivalent to true full dose images and showed improvement at 40% dose. Our network for reconstruction of low dose image was revealed in terms of increased SSIM values. If this method can be applicable in patients, the cumulative radiation hazard by digital mammography, especially in high risk, young patients who are exposed to more frequent mammographic examinations can be avoided.

Screening mammography has been shown to reduce mortality through earlier detection (15). Its radiation dose should be regarded together with image quality, thus dose optimization is required. This refers to the determination of the lowest average glandular dose that yields a sufficient level of clinical image quality. Earlier study using clinical images from a screening program (16) with a Senographe 2000D (GE Healthcare, Milwaukee, WI) have indicated that doses down to 50% reference average glandular dose

level are adequate for maintaining clinically acceptable image quality. Another study also revealed that there is no statistically significant difference between the 100% and the 50% dose levels, suggesting possibility of reducing the dose to half of the dose level of AEC (3). Similarly, we found that there is minimal effect on lesion conspicuity up to 60% dose reduction. However, lower doses with threshold of 40% level, there was gradual degradation in image quality, with fewer details clearly visualized. The characterization of calcification was more dose dependent at the level below 40% level, while the effect of dose reduction on mass became evident at the 20% level, concordant with previous study with phantom study (3). It revealed the there was a significant degradation in image quality at the 30% dose level. Regarding calcification, Yakabet et al (4) demonstrated that calcification was not affected by a dose reduction of 50% and Samei et al (5) reported a dose reduction to 50% did not have statistically significant effect on ability of the screening to detect calcification while 25% dose did have a significant effect. With regard to ability to detect masses, 30-50% dose reductions are reported to be possible (5, 17, 18). Another study also revealed that calcification detection and mass discrimination decreased from full dose to

quarter dose significantly, however detection of malignant mass did not appear to be affected by dose reduction (5).

Image de-noising is essential to address image quality degradation caused by minimizing radiation dose (19). It is to maximize the restoration of the original image details by noise reduction as well as contrast enhancement (20). Recently, image de-noising based on deep learning has been shown to outperform. An effective de-noising could allow further dose reduction (21). Our CNN based de-noising method was employed to convert lower-dose to higher-dose mammographic images. We used cyclic consistency in unsupervised framework to eliminate the need for perfectly paired images (11). Originally, CNNs have been used for microcalcification and mass classifications in computer aided detection (CAD) for mammography (22). Samala et al (23) have demonstrated that mammography images can be useful for training of CNN for mass detection in DBT.

There are several limitations in our study. We analyzed relatively small number of cases and were all breast cancer patients. Second, observer variability is generally the largest source of variability. Previously Bernhardt et al (24) studied signal difference to

noise ratios normalized to AEC and showed that the W/Rh node/filter combination is superior. However, we did not evaluate the impact of these combinations. Lastly, most diagnostic performance experiments involve rating images for the presence of one type of abnormality and assign each image a grade ranging from absent to present, with multiple grades in between. The number of gradations range from four to 100 (25, 26). However, this approach is limited for many of the diagnostic tasks performed in clinical setting. In our study, we used categorical approach to emulate the clinical paradigm by rating images for the detection and characterization of abnormalities. A clinical study involving adequate number of patient images should be performed to investigate the preliminary conclusions of our study further.

We were able to convert low dose images to synthesized full dose images in the evaluation of mastectomy specimens of breast cancer patients. Our results suggest that there is a potential for modest dose reduction retaining diagnostic information with de-noising algorithm in digital mammography. Our results provide a baseline for future studies on reducing the radiation dose with lesion preservation on mammography. Our future work

includes extending the approach to clinical practices with breast cancer patients using reduced dose mammography together with de-noising deep learning algorithm.

Acknowledgement

We would like to thank for the support of EunHee Kang, Ph.D and Jong Chul Ye, Ph.D, Department of Bio and Brain Engineering, Korea Advanced Institute of Science and Technology (KAIST).

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 transactions on information technology in biomedicine : a publication of the IEEE Engineering in Medicine and Biology Society. 2009;13(2):236-51.

2. Pisano ED, Gatsonis C, Hendrick E, Yaffe M, Baum JK, Acharyya S, et al. Diagnostic performance of digital versus film mammography for breast-cancer screening. N Engl J Med. 2005;353(17):1773-83.

3. 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. The British journal of radiology. 2007;80(955):557-62.

4. 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. Journal of digital imaging. 2010;23(5):520-6.

5. Samei E, Saunders RS, Jr., Baker JA, Delong DM. Digital mammography: effects of reduced radiation dose on diagnostic performance. Radiology. 2007;243(2):396-404.

6. LeCun Y, Bengio Y, Hinton G. Deep learning. Nature. 2015;521(7553):436-44.

7. 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 transactions on medical imaging. 2016;35(5):1344-51.

8. Zhu B, Liu JZ, Cauley SF, Rosen BR, Rosen MS. Image reconstruction by domain-transform manifold learning. Nature. 2018;555(7697):487-92.

9. Liu J, Zarshenas A, Qadir A, Wei Z, Yang L, Fajardo L, et al., editors. Radiation dose reduction in digital breast tomosynthesis (DBT) by means of deep-learning-based supervised image processing. SPIE Medical Imaging; 2018: SPIE.

10. Dabov K, Foi A, Katkovnik V, Egiazarian K. Image denoising by sparse 3-D transform-domain collaborative filtering. IEEE transactions on image processing : a publication of the IEEE Signal Processing Society. 2007;16(8):2080-95.

11. Zhou T, Krähenbühl, Philipp, Aubry, Mathieu, Huang, Qi-Xing, Efros, Alexei A.

Learning Dense Correspondence via 3D-Guided Cycle Consistency. 2016 IEEE Conference on Computer Vision and Pattern Recognition 117-26.

12. Zhu J-Y, Park, Taesung, Isola, Phillip, Efros, Alexei A. Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks. 2017 IEEE International Conference on Computer Vision (ICCV).2242-51.

13. A. Paszke SG, S. Chintala, G. Chanan, E. Yang, Z. DeVito, Z. Lin, A. Desmaison, L. Antiga, and A. Lerer. Automatic differentiation in pytorch. 31st Conference on Neural Information Processing Systems. 2017.

14. Wang Z, Bovik AC, Sheikh HR, Simoncelli EP. Image quality assessment: from error visibility to structural similarity. IEEE transactions on image processing : a publication of the IEEE Signal Processing Society. 2004;13(4):600-12.

15. Nystrom L, Andersson I, Bjurstam N, Frisell J, Nordenskjold B, Rutqvist LE. Long-term effects of mammography screening: updated overview of the Swedish randomised trials. Lancet (London, England). 2002;359(9310):909-19.

16. Hemdal B, Andersson I, Grahn A, Hakansson 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. Radiation protection dosimetry. 2005;114(1-3):383-8.

17. Saunders RS, Jr., Baker JA, Delong DM, Johnson JP, Samei E. Does image quality matter? Impact of resolution and noise on mammographic task performance. Medical physics. 2007;34(10):3971-81.

18. Ruschin M, Timberg P, Bath M, Hemdal B, Svahn T, Saunders RS, et al. Dose dependence of mass and microcalcification detection in digital mammography: free response human observer studies. Medical physics. 2007;34(2):400-7.

19. Chatterjee P, Milanfar P. Is denoising dead? IEEE transactions on image processing : a publication of the IEEE Signal Processing Society. 2010;19(4):895-911.

20. Cesarelli M, Bifulco P, Cerciello T, Romano M, Paura L. X-ray fluoroscopy noise modeling for filter design. International journal of computer assisted radiology and surgery. 2013;8(2):269-78.

21. Lee D, Choi S, Kim H-J. Performance evaluation of image denoising developed

using convolutional denoising autoencoders in chest radiography. Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment. 2018;884:97-104.

22. Chan HP, Lo SC, Sahiner B, Lam KL, Helvie MA. Computer-aided detection of mammographic microcalcifications: pattern recognition with an artificial neural network. Medical physics. 1995;22(10):1555-67.

23. 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. Medical physics. 2016;43(12):6654.

24. Bernhardt P, Mertelmeier T, Hoheisel M. X-ray spectrum optimization of full-field digital mammography: simulation and phantom study. Medical physics. 2006;33(11):4337-49.

25. Riedl CC, Jaromi S, Floery D, Pfarl G, Fuchsjaeger MH, Helbich TH. Potential of dose reduction after marker placement with full-field digital mammography. Investigative radiology. 2005;40(6):343-8.

26. Chesters MS. Human visual perception and ROC methodology in medical imaging. Physics in medicine and biology. 1992;37(7):1433-76.

국문요약

목적:

본 연구는 딥러닝을 이용하여 저선량 디지털 유방촬영술 영상의 화질 개선을 위한 알고리즘을 개발하고, 저선량 디지털 유방촬영술로 얻은 이미지를 개발한 딥러닝 알고리즘을 이용하여 복원하여 이미지의 판독에 영향을 주지 않는 범위 내에서 얼마의 선량을 감소시킬 수 있는지 알아보고자 하였다.

연구 대상 및 방법:

본 연구는 전향적 연구로 본원 임상연구심의위원회의 심의를 통과하였다. 유방암 수술 과정 중 얻은 유방 전 절제 조직에 방사선 조사량을 여러 단계(Automatic exposure control, 이로부터 10%, 20%, 40%, 60%, 80% 감소한 선량)로 달리하여 저선량 디지털 유방촬영술을 시행하고 평가하였다. Full dose 영상과 비교하여 acceptable radiation dose level 을 찾았다. 이 dose 이하의 영상에 대해 딥러닝 알고리즘을 통해 복원한 후 정보를 모르는 상태에서 화질을 평가하여 판독에 영향을 주지 않는 최소의 선량치를 구하였다.

결과:

True full dose 영상과 비교하였을 때 유방 병변의 발견은 10% 선량 영상에서 약 60.8%-76.8%로 감소하였고, 병변의 질은 mass 는 10% 선량에서 86.3%로, 석회화는 20% 선량에서 83.8%로 저하되었다. 딥러닝 알고리즘을 적용하여 20%선량과 40%선량 영상을 복원하였고 True full 영상과 비교하였을 때 복원된 영상에서 병변의 발견은 두 선량에서 87.4%-100.0%로 유지되었으나 병변의 질이 복원된 40% 선량 영상에서 향상된 것에 반해 복원된 20% 영상에서는 병변의 질 향상은 보이지 않았다 ($p<0.001$)

결론:

본 연구에서 딥러닝 알고리즘을 이용하여 약 40%까지 유방촬영술의 선량을 감소하면서 영상의 질을 유지하고 향상시킬 수 있음을 확인하였다. 본 예비연구의 결과는 딥러닝 알고리즘을 이용한 유방촬영술의 선량 감소의 기초 연구로 추후 이를 임상 적용 가능성이 있음을 시사한다.

중심단어

유방촬영술

딥러닝

유방암

방산선량