PROCEEDINGS OF SPIE

SPIEDigitalLibrary.org/conference-proceedings-of-spie

Deep CNN task-based image quality assessment: application to digital breast tomosynthesis reconstruction and denoising

Mingjie Gao, Mark Helvie, Ravi Samala, Lubomir Hadjiiski, Jeffrey Fessler, et al.

Mingjie Gao, Mark A. Helvie, Ravi K. Samala, Lubomir M. Hadjiiski, Jeffrey A. Fessler, Heang-Ping Chan, "Deep CNN task-based image quality assessment: application to digital breast tomosynthesis reconstruction and denoising," Proc. SPIE 12463, Medical Imaging 2023: Physics of Medical Imaging, 1246319 (7 April 2023); doi: 10.1117/12.2655419

SPIE.

Event: SPIE Medical Imaging, 2023, San Diego, California, United States

Downloaded From: https://www.spiedigitallibrary.org/conference-proceedings-of-spie on 28 Apr 2023 Terms of Use: https://www.spiedigitallibrary.org/terms-of-use

Deep CNN Task-based Image Quality Assessment: Application to Digital Breast Tomosynthesis Reconstruction and Denoising

Mingjie Gao[*a,b], Mark A. Helvie[a], Ravi K. Samala[c], Lubomir M. Hadjiiski[a], Jeffrey A. Fessler[a,b], Heang-Ping Chan[a]

[a]Department of Radiology, University of Michigan, Ann Arbor, MI 48109; [b]Department of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor, MI 48109; [c]U.S. Food and Drug Administration, Silver Spring, MD 20993

ABSTRACT

Image noise in digital breast tomosynthesis (DBT) reduces the detectability of subtle signs of breast cancer such as microcalcifications (MC). This study investigated the potential of applying DNGAN, our previously developed deep convolutional neural network (CNN) DBT denoiser, to different reconstruction stages to improve the image quality, including projection views before reconstruction, intermediate images during iterative reconstruction, or final reconstructed images, and a combination of the different stages of denoising. We also proposed two CNNs as task-based image quality measures to compare different reconstructions: a CNN noise estimator (CNN-NE) trained to evaluate the noise level of a given DBT image, and a CNN MC classifier (CNN-MC) trained to estimate the detectability of MCs by classifying clustered MCs from MC-free backgrounds. The CNN-NE was trained with virtual DBTs reconstructed from projections generated by the VICTRE tool over a wide range of noise levels. The CNN-MC was trained with human subject DBTs. We adopted the training strategy of transfer learning to train CNN-NE and CNN-MC due to the limited training data. We found that the increase in AUC estimated by the CNN-MC classifier correlated well with the decrease in image noise by DNGAN estimated by CNN-NE on an independent human subject test set. A combination of DNGAN-regularized plug-and-play reconstruction and an additional DNGAN post-reconstruction denoising achieved the lowest noise level and the best MC detectability. The AUC and noise rankings from the CNNs matched our visual judgement that less noisy images had better MC conspicuity.

Keywords: deep learning, digital breast tomosynthesis, image reconstruction, microcalcification

1 INTRODUCTION

Digital breast tomosynthesis (DBT) is a low-dose breast imaging technique for both breast cancer screening and diagnosis [1]. Due to its low exposure and multiple projection acquisition, the image noise is high and subtle signs of early breast cancer such as microcalcifications (MC) are not easily visible. We previously developed a deep convolutional neural network (CNN), which we called DNGAN, for denoising reconstructed DBT images [2]. It reduced image noise and preserved subtle MCs and breast tissue texture. We have also shown that DNGAN could be applied not only to reconstructed DBT images, but also to projection views (PVs) before reconstruction [3] or to intermediate reconstructed images in iterative algorithms as a regularizer [4]. In this work, we investigated these denoising or reconstruction methods, or a combination of them, and compared their image quality.

For the comparison of DBT images, it is important to characterize image features that may affect the diagnosis of breast cancer. In this study, we focused on two DBT image features: noise level and MC detectability, and their relationship. We proposed two CNNs to provide task-based assessments. The first was a CNN noise estimator (CNN-NE) to evaluate the noise level of a given DBT image. The second was a CNN MC classifier (CNN-MC) to estimate the detectability of MCs by classifying clustered MCs from MC-free backgrounds. We investigated the potential of using the two CNNs as surrogates of image quality measures to compare different DBTs and to guide the development of the reconstruction and denoising methods.

[*] Corresponding author: gmingjie@umich.edu

Medical Imaging 2023: Physics of Medical Imaging, edited by Lifeng Yu, Rebecca Fahrig, John M. Sabol, Proc. of SPIE Vol. 12463, 1246319
© 2023 SPIE · 1605-7422 · doi: 10.1117/12.2655419

Downloaded From: https://www.spiedigitallibrary.org/conference-proceedings-of-spie on 28 Apr 2023
Terms of Use: https://www.spiedigitallibrary.org/terms-of-use

2 METHODS AND MATERIALS

2.1 DBT Reconstruction with DNGAN Denoising

We trained the DNGANs with digital breast phantom data. The training details can be found in [2]. The data set consisted of 70 virtual DBT scans generated by the VICTRE package [5] simulating a variety of breast densities, thicknesses, and x-ray exposure conditions. For the DNGAN applied to reconstructed DBT images, the DBT volumes were reconstructed using 3 iterations of simultaneous algebraic reconstruction technique (SART). For the DNGANs applied to PVs, we used similar training techniques and the same data set as described above but used PVs before reconstruction as input images.

To use DNGAN for regularization, we followed the plug-and-play (PnP) reconstruction framework [6]. The updates of the image variable x and the auxiliary variable z can be written as

$$x_n = \underset{x}{\operatorname{argmin}} \frac{1}{2}\|y - Ax\|^2 + \frac{\beta}{2}\|x - z_{n-1}\|^2 \quad (1)$$

$$z_n = D(x_n) \quad (2)$$

Where β is the regularization parameter, $D(\cdot)$ is the plugged-in trained DNGAN denoiser, y is the PV, A is the system matrix, and n is the iteration index. We ignored the PnP Lagrangian multiplier for simplicity. We applied one iteration of preconditioned gradient descent to the minimization problem in (1).

2.2 Task-based Image Quality Measures using CNN

The training sets for the CNN-NE and CNN-MC models were prepared as follows. The CNN-NE was trained with VICTRE-simulated data. We used the same set of virtual DBT images as that used for DNGAN training and extracted 256,194 128×128-pixel patches as input images. The labels were the standard deviations of patch pixel values after background reduction [7]. The CNN-MC was trained with human subject data previously collected with IRB approval. We collected 127 DBT views from 64 patients with biopsy-proven MCs [8]. We extracted 751 128×128×3-pixel patches with clustered MCs (6,008 after flipping and rotation augmentations) and 19,079 MC-free patches, and took maximum intensity projection (MIP) over 3 consecutive slices along the depth dimension for all patches to emphasize MC clusters, if any. We employed four-fold cross validation to select the training hyper-parameters and then trained the final model using the entire training set and the selected parameters.

An independent set of 52 human subject DBTs with 104 views was collected as a test set [8]. To test the CNN-NE, we extracted 1,955 MC-free patches and took the average of the CNN-NE outputs on these patches as an indicator of the noise level of the entire test set. To test the CNN-MC, we extracted 709 MC patches as positives and 1,955 MC-free patches from the same locations for CNN-NE as negatives and obtained MIP over 3 slices for all patches. The area under the receiver operating characteristic (ROC) curve (AUC) of the CNN-MC classification was used as an MC detectability measure.

Observing that the amount of CNN-MC data was relatively small for training a typical deep network, we adopted the training strategy of transfer learning, in which the weights that a model has learned from a source task in a different domain or similar domain are transferred as initial weights to train the model for the target task. Transfer learning could improve the model robustness when the training data was limited [9]. Specifically, for this study we first trained the CNN-NE model from scratch. Then, the CNN-NE model was further fine-tuned with a smaller learning rate to obtain the CNN-MC model.

We used the ResNet [10] as the backbone network structure of CNN-NE and CNN-MC. The CNN-MC was retrained for each image condition and tested accordingly. We also repeated the training for each condition 5 times, each with a different random initialization for training the CNN-NE, to account for the training uncertainties. The mean and standard deviation of the results from the 5 repeats during deployment were reported.

Downloaded From: https://www.spiedigitallibrary.org/conference-proceedings-of-spie on 28 Apr 2023
Terms of Use: https://www.spiedigitallibrary.org/terms-of-use

3 RESULTS

We compared a number of reconstruction schemes in this study, as shown in Table 1. They included SART at different iterations, SART with PVs denoised by DNGAN, DNGAN-regularized PnP with different β values, and PnP followed by DNGAN post-reconstruction denoising. Figure 1 plots the AUCs of the CNN-MC classification performance versus CNN-NE estimated noise levels for the different reconstructions deployed on the independent human subject test set. Figure 2 shows examples of MIP patches with clustered MCs and MC-free background from test cases.

Table 1. Configurations of reconstruction methods for image quality comparison.

Label	Description
(a)	SART iteration 2
(b)	SART iteration 3
(c)	DNGAN PV denoised SART iteration 3
(d)	DNGAN-regularized PnP iteration 3 with $\beta = 20$
(e)	DNGAN-regularized PnP iteration 3 with $\beta = 100$
(f)	DNGAN-regularized PnP iteration 3 with $\beta = 200$
(g)	DNGAN-regularized PnP iteration 3 with $\beta = 20$ and DNGAN post-reconstruction denoising

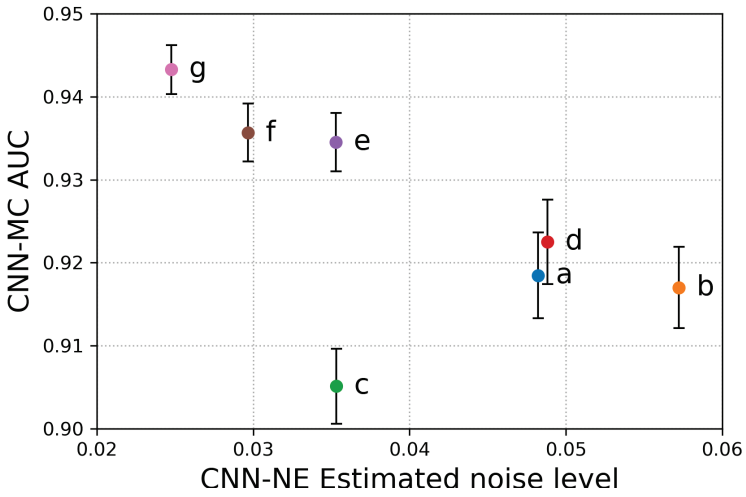

Figure 1. AUCs of CNN-MC classification versus CNN-NE estimated noise levels on the independent test set of human subject DBTs for the reconstruction schemes in Table 1. The vertical error bars indicate one standard deviation estimated from the 5 repeated CNN-MC models. The error bars for the estimated noise levels were two orders of magnitude smaller than the mean values.

Overall, lower image noise levels were correlated with higher AUCs. In addition, the CNN-MC was sensitive to the signal strength of the MCs. More iterations of the unregularized SART reconstruction accumulated noise and lowered the AUC, as shown by (a) and (b) for iterations 2 and 3, respectively. SART with DNGAN PV denoising before reconstruction reduced image noise but smoothed out some subtle MCs, thus lowering the AUC, as shown by (c) in comparison to (b). DNGAN-regularized PnP reconstructions had stronger signal enhancement and therefore higher AUCs compared with other reconstructions at the same noise level, such as between (c) and (e), or between (a) and (d). DNGAN-regularized PnP with the additional DNGAN post-reconstruction denoising, shown as (g), had the lowest noise level and the highest MC detectability among those studied. The improvement can also be clearly seen in the example patches in Figure 2. The AUC and noise rankings from the task-based CNN quality measures matched our visual judgement on the images.

Downloaded From: https://www.spiedigitallibrary.org/conference-proceedings-of-spie on 28 Apr 2023
Terms of Use: https://www.spiedigitallibrary.org/terms-of-use

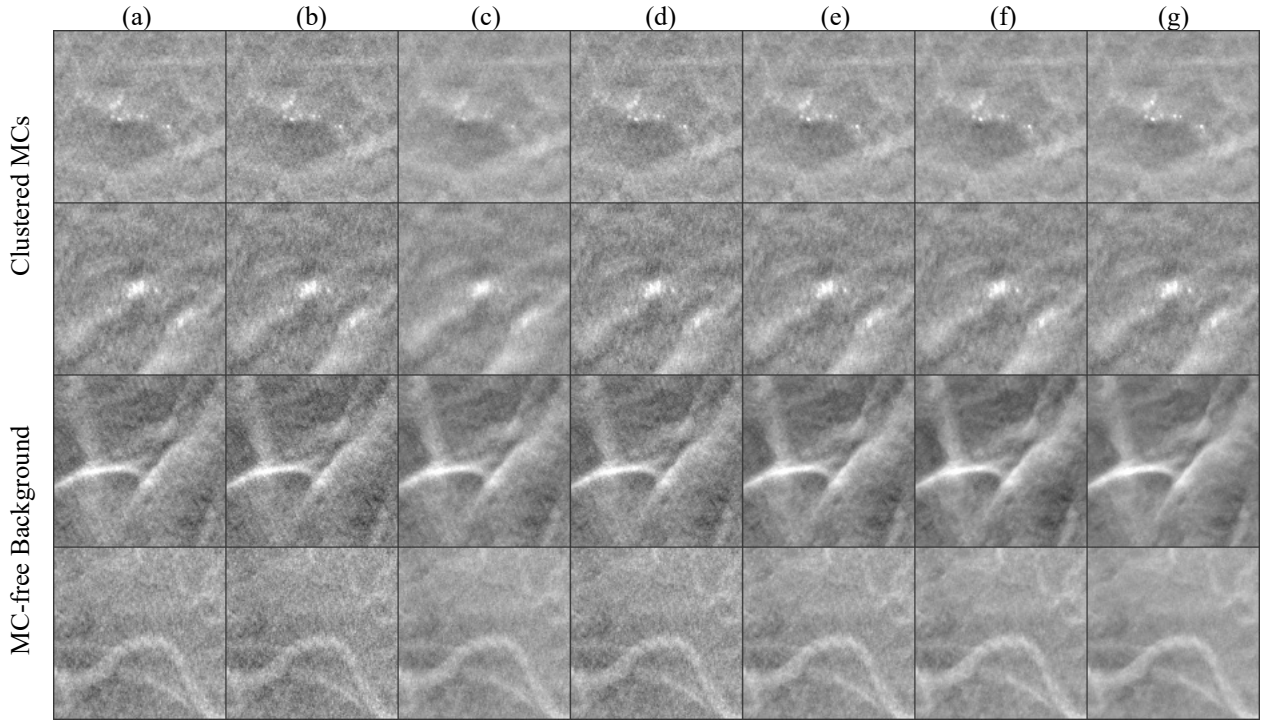

Figure 2. Example MIP patches with clustered MCs and MC-free background from the human subject test set for different reconstructions. The labels (a) to (g) are defined in Table 1.

Finally, we did a visual quality check on the spiculated masses. Figure 3 shows the example images of a spiculated mass from the human subject DBT for selected reconstruction schemes. The example images were ordered by decreasing noise. Among the conditions being compared, (g) was the best in terms of noise and the appearance of spiculations and tissue structures. It was also free of artifact as opposed to the patchy appearance that often occurs in conventional model-based reconstructed images.

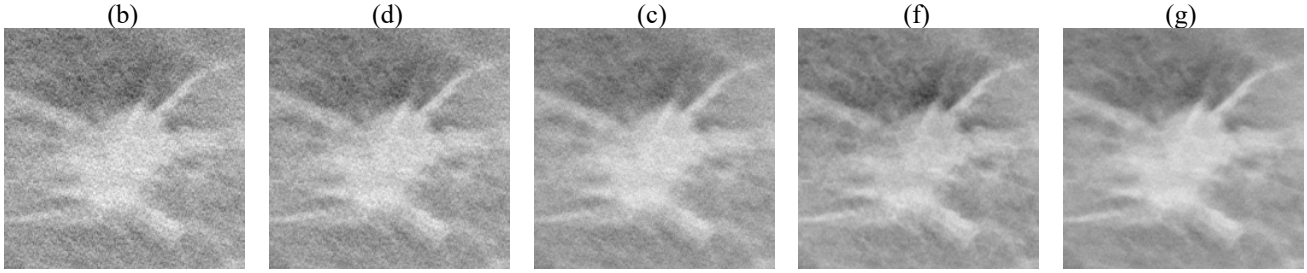

Figure 3. Example images of a spiculated mass from a human subject DBT for selected reconstruction schemes. The images are 20 mm × 20 mm in size and are ordered by decreasing noise.

4 CONCLUSION

This work presents promising results of training CNNs for task-based image noise and MC detectability assessments. The CNN-MC can distinguish MC clusters from noisy image background and the performance correlated well with the noise level estimated by the CNN-NE, which may be useful for guiding the development of imaging, reconstruction and image processing techniques. The MC detectability estimated by the CNN-MC takes into account the MCs as a cluster, which is different from the conventional signal detectability *d'* models that focus on individual microcalcifications.

In this study, we used the CNN-based image quality measures to compare the application of DNGAN denoising to different DBT reconstruction stages, either on PVs before reconstruction, intermediate reconstructed images, or final reconstructed outputs. The CNN measures indicated that the DNGAN reduced the image noise and improved the AUC

Downloaded From: https://www.spiedigitallibrary.org/conference-proceedings-of-spie on 28 Apr 2023
Terms of Use: https://www.spiedigitallibrary.org/terms-of-use

of MC classification on an independent human subject test set. A combination of DNGAN-regularized PnP reconstruction and DNGAN post-reconstruction denoising achieved the lowest noise level and the best MC detectability. Future work includes continued improvement of the DBT reconstruction techniques, conducting observer studies to confirm the effectiveness of DNGAN, and validating the correlation between CNN-NE/CNN-MC rankings and human detection.

ACKNOWLEDGMENT

This work is supported by the National Institutes of Health under Award Number R01 CA214981.

REFERENCES

[1] A. Chong, S. P. Weinstein, E. S. McDonald, and E. F. Conant, "Digital breast tomosynthesis: Concepts and clinical practice," *Radiology*, vol. 292, no. 1, pp. 1–14, Jul. 2019, DOI: 10.1148/radiol.2019180760.

[2] M. Gao, J. A. Fessler, and H.-P. Chan, "Deep convolutional neural network with adversarial training for denoising digital breast tomosynthesis images," *IEEE Transactions on Medical Imaging*, vol. 40, no. 7, pp. 1805–1816, Jul. 2021, DOI: 10.1109/TMI.2021.3066896.

[3] M. Gao, R. K. Samala, J. A. Fessler, and H.-P. Chan, "Deep convolutional neural network denoising for digital breast tomosynthesis reconstruction," in *Proceedings of SPIE*, 2020, 113120Q, DOI: 10.1117/12.2549361.

[4] M. Gao, J. A. Fessler, and H.-P. Chan, "Plug-and-play reconstruction with deep learning denoising for improving detectability of microcalcifications in digital breast tomosynthesis images," in *Radiological Society of North America Scientific Assembly and Annual Meeting*, 2021.

[5] A. Badano, C. G. Graff, A. Badal, D. Sharma, R. Zeng, F. W. Samuelson, S. J. Glick, and K. J. Myers, "Evaluation of digital breast tomosynthesis as replacement of full-field digital mammography using an in silico imaging trial," *JAMA Network Open*, vol. 1, no. 7, p. e185474, Nov. 2018, DOI: 10.1001/jamanetworkopen.2018.5474.

[6] S. V. Venkatakrishnan, C. A. Bouman, and B. Wohlberg, "Plug-and-play priors for model based reconstruction," in *IEEE Global Conference on Signal and Information Processing*, 2013, 945–948, DOI: 10.1109/GlobalSIP.2013.6737048.

[7] H.-P. Chan, D. Wei, M. A. Helvie, B. Sahiner, D. D. Adler, M. M. Goodsitt, and N. Petrick, "Computer-aided classification of mammographic masses and normal tissue: linear discriminant analysis in texture feature space," *Physics in Medicine and Biology*, vol. 40, no. 5, pp. 857–876, May 1995, DOI: 10.1088/0031-9155/40/5/010.

[8] R. K. Samala, H.-P. Chan, Y. Lu, L. M. Hadjiiski, J. Wei, and M. A. Helvie, "Computer-aided detection system for clustered microcalcifications in digital breast tomosynthesis using joint information from volumetric and planar projection images," *Physics in Medicine and Biology*, vol. 60, no. 21, pp. 8457–8479, Nov. 2015, DOI: 10.1088/0031-9155/60/21/8457.

[9] J. Yosinski, J. Clune, Y. Bengio, and H. Lipson, "How transferable are features in deep neural networks?," 2014. [Online]. Available: http://arxiv.org/abs/1411.1792.

[10] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2016, 770–778, DOI: 10.1109/CVPR.2016.90.

Downloaded From: https://www.spiedigitallibrary.org/conference-proceedings-of-spie on 28 Apr 2023
Terms of Use: https://www.spiedigitallibrary.org/terms-of-use