Digital Breast Tomosynthesis Denoising Using Deep Convolutional Neural Network: Effects of Dose Level of Training Target Images

Mingjie Gao[*a,b], 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

Abstract

This paper investigates the training of deep convolutional neural networks (DCNNs) to denoise digital breast tomosynthesis (DBT) images. In our approach, the DCNN was trained with high dose (HD)/low dose (LD) image pairs, in which the HD image was used as reference to guide the DCNN to learn to reduce noise of a corresponding input LD image. In the current work, we studied the effect of the dose level of the HD target images on the effectiveness of the trained denoiser. We generated a set of simulated DBT data with different dose levels using CatSim, a dedicated x-ray imaging simulator, in combination with virtual anthropomorphic breast phantoms simulated by VICTRE. We found that a DCNN trained with higher dose level of HD target images led to less noisy denoised images. We also acquired DBT images of real physical phantoms containing simulated microcalcifications (MCs) for training and validation. The denoisers trained with either simulated or physical phantom data improved significantly ($p < 0.0001$) the contrast-to-noise ratio of MCs in the validation phantom images. In an independent test set of human subject DBTs, the MCs became more conspicuous, and the mass margins and spiculations were well preserved. The study showed that the denoising DCNN was robust in that the denoiser could be trained with either simulation or physical phantom data. Moreover, the denoiser trained with CatSim simulation data was directly applicable to human DBTs, allowing flexibility in terms of the training data preparation, especially the HD images.

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

1 INTRODUCTION

Digital breast tomosynthesis (DBT) is the newest screening tool for detection of early breast cancer [1]. It alleviates the problem of overlapping tissues in two-dimensional (2D) mammography by acquiring multiple projection views (PVs) and reconstructing a quasi-3D image. The noise level of the DBT images is high due to the fact that the total dose level of DBT is kept comparable to that of a 2D mammogram. Meanwhile, it is challenging to denoise DBT images because of the need to preserve important signs of malignancy such as subtle microcalcifications (MCs) and the ill-defined or spiculated margins of masses. Our previous study used a deep convolutional neural network (DCNN) to denoise the PVs before DBT reconstruction [2].

In this study, we investigated the feasibility of training a DCNN denoiser for the reconstructed DBT slices. The training required paired low dose (LD) and high dose (HD) DBT images. We studied the effect of the dose level of the training HD images on the effectiveness of the trained denoiser. We generated virtual DBT data from a dedicated x-ray imaging simulator [3]. To demonstrate the robustness of the denoising method, we also trained a denoiser using physical phantom data and compared it with the denoiser from digital phantom data. The performance of the denoiser was evaluated in terms of the improvement in MC conspicuity and the preservation of background texture.

* Corresponding author: gmingjie@umich.edu

Medical Imaging 2021: Physics of Medical Imaging, edited by Hilde Bosmans, Wei Zhao, Lifeng Yu, Proceedings of SPIE Vol. 11595, 115951K · © 2021 SPIE
CCC code: 1605-7422/21/$21 · doi: 10.1117/12.2580900

2 MATERIALS AND METHODS

2.1 Digital breast phantom data

We prepared 24 heterogeneous dense digital breast phantoms at a voxel resolution of 0.05 mm using an analytical anthropomorphic model from VICTRE [4]. We used a simulation tool by GE Global Research, CatSim [5], to generate PVs of the digital breast phantoms by modeling the GE Pristina DBT system. The flat panel detector model and its signal and noise characteristics as validated for CatSim by Carvalho [6] were used in the simulations. Focal spot blur, scattered radiation, and detector noise were not considered. The Pristina system acquires 9 PVs in a 25º scan and has a detector pixel size of 0.1 mm × 0.1 mm. To simulate low dose PVs, we set the x-ray exposure in CatSim to 24 mAs, which was close to the value from automatic exposure control (AEC) for a 4.5 cm breast for the Pristina system [7]. The mean glandular dose (MGD) was 1.42 mGy under this exposure, estimated by a Monte Carlo simulation tool called CatDose in the CatSim package. We also simulated the high dose DBT scans with 72 mAs (3× AEC), 120 mAs (5× AEC), 360 mAs (15× AEC) and noiseless (∞ × AEC) settings. After obtaining the PVs, we used 3 iterations of the simultaneous algebraic reconstruction technique (SART) with the segmented separable footprint projector [8] to reconstruct the DBT volumes. The reconstructed DBT voxel size was 0.1 mm × 0.1 mm × 1 mm. We extracted pairs of 24mAs/target non-overlapping 32×32-pixel patches from the DBT images to form four LD/HD paired training sets, each of which used the HD target images at one of the four dose settings (72, 120, 360, ∞ mAs) while the LD images (24 mAs) were common to all sets. Each set contained approximately 200,000 pairs of patches.

2.2 Physical phantom data

We had six 1-cm-thick heterogeneous slabs with 50% glandular/50% adipose breast-tissue-equivalent material mixed in random patterns to construct five 5-cm-thick phantoms [9]. For each phantom, five of the six slabs were used and stacked in a different order or orientation to form a different phantom. To generate physical phantom data, each phantom was scanned twice by a Pristina system under the same compression. The LD scans used the standard dose (STD) setting, which automatically chose a technique of Rh/Ag 34 kVp, with a mean exposure of 31.7 mAs and a mean MGD of 1.30 mGy. We manually set the exposure for the HD scans to Rh/Ag 34 kVp, 125 mAs. The DBT reconstruction settings were the same as before. We extracted about 209,000 pairs of LD/HD non-overlapping 32×32-pixel patches from four phantoms to form the training set and held out one phantom for validation.

To evaluate the denoising effect on subtle objects, clusters of simulated calcifications (glass beads) of three nominal diameters (0.150-0.180 mm, 0.180-0.212 mm, 0.212-0.250 mm) were randomly sandwiched between the slabs in the validation phantom. We marked the MCs in the HD volume and had a total of 236, 227, 159 MCs of the three sizes, respectively.

2.3 DCNN denoiser

We used a 10-layer convolutional neural network as our denoiser. Its input is a 32×32-pixel image region extracted from DBT slice. The corresponding HD patch was used as a reference (target) image during training. The denoiser output had the same shape as the input. Because the denoiser was fully convolutional, we applied it directly to the full DBT slices during deployment. The training loss was composed of the mean-squared-error (MSE) loss and the adversarial loss [10]. The MSE loss compares the pixel-wise difference between the denoised and target images. The adversarial loss, which we found effective for denoising PVs of DBT images [2], preserves texture details by comparing the distributions of the denoised images and the target images.

2.4 Figures of merit

To quantitatively evaluate the conspicuity of MCs in the images, we calculated the contrast-to-noise ratio (CNR) and full width half maximum (FWHM) as figures of merit for each MC. We obtained the contrast and FWHM of an MC from a Gaussian function fitted to the MC pixels. The background noise was estimated as the root-mean-squared noise in the surrounding area after removing the local background mean gray level using a box-rim filter [11] and excluding the MC pixels.

3 RESULTS

To study the effect of the dose level of the training HD target images on the effectiveness of denoising by the trained DCNN, we trained four denoisers using the four training sets from the digital breast phantoms with the same training settings, and deployed it on the validation physical phantom. As shown in Figure 1, the denoised images were smoother when the training targets were less noisy. To compare with the denoiser trained with LD/HD image pairs from the physical phantoms, we selected the 24mAs/120mAs denoiser that had a dose ratio closest to that of the physical phantoms.

We trained a denoiser with the physical phantom data, denoted as LD/HD denoiser, using the same training settings as before. Figure 2 compares the visibility of MCs in the validation phantom for the LD/HD denoiser as well as the 24mAs/120mAs denoiser. For the 0.180-0.212 mm and 0.212-0.250 mm groups, the MCs were more visible than those in the original LD DBT slice for both denoisers. This was supported by the quantitative CNR and FWHM results. The CNR improved by an average of 64% (77%) and 49% (62%) for the 0.180-0.212 mm (0.212-0.250 mm) size group by the 24mAs/120mAs denoiser and the LD/HD denoiser, respectively, and the improvements were statistically significant ($p < 0.0001$). The FWHMs indicated that the MCs maintained their sharpness. However, for the 0.150-0.180 mm group, although the CNRs of the analyzable MCs were also significantly ($p < 0.0001$) improved by an average of 35% and 17%, respectively, by the two denoisers, some MCs were blurry and failed the Gaussian fitting after denoising, especially for the very subtle ones, as demonstrated in Figure 2. The 24mAs/120mAs denoised images had higher CNRs than the LD/HD denoised images and were visually less noisy, probably because 24mAs/120mAs had a dose ratio of 5, while LD/HD had a dose ratio of 4 and contained scatter and detector noises in both the LD/HD images. Figure 2 also shows the denoised images from the 24mAs/noiseless denoiser for comparison.

Figure 3 shows the results of applying the 24mAs/120mAs, LD/HD and 24mAs/noiseless denoisers to independent test human subject DBTs. The images were acquired using a GE prototype GEN2 DBT system. The prototype system acquires 21 PVs in a scan angle of 60°. We used the central 9 PVs that correspond to a scan angle of 24°, similar to the GE commercial Pristina DBT system. The test images in Figure 3 contained an invasive ductal carcinoma (mass) and a ductal carcinoma in situ (MC cluster). After denoising, the mass margins and spiculations were well preserved. The background noise was reduced so the MCs were more conspicuous.

4 DISCUSSION AND CONCLUSION

We trained a DCNN to denoise reconstructed DBT slices. We demonstrated by CatSim simulation data that using higher dose training target led to less noisy output from the denoiser, but the use of noiseless images as the training target may lead to denoised images that appear overly smooth. The degree of smoothness in the parenchymal texture on DBT images that is acceptable to radiologists warrants further investigation. Our DCNN denoising method was robust in that the denoiser could be trained with either digitally simulated or physical phantom DBT data. Moreover, the denoiser trained with CatSim simulation data was directly applicable to the real data, allowing flexibility in terms of the training data preparation. The simulation data has some advantages over the real data. For example, once the simulation model is formed, it is inexpensive to generate a large set of data and there is no exposure limit. On the other hand, it is difficult to make a large number of realistic physical phantoms and the tube loading of the DBT system limits the dose level of imaging a physical phantom. Most importantly, the DCNN denoiser trained with phantom data was applicable to DBT human subjects, thus avoiding the need to train using LD/HD human DBT that may be infeasible to collect. Future work includes improving the conspicuity of subtle MCs in the denoised images, study of the robustness of the denoiser across reconstruction methods or DBT systems, and further validation of the effectiveness of the trained denoisers in clinical DBTs.

ACKNOWLEDGMENT

This work is supported by the National Institutes of Health under Award Number R01 CA214981. The authors would like to thank GE Global Research for providing the CatSim simulation tool, and Christian Graff, Ph.D., for the open-source digital breast phantom generation program.

REFERENCES

[1] J. T. Dobbins, "Tomosynthesis imaging: at a translational crossroads," *Medical Physics*, vol. 36, no. 6, pp. 1956–1967, May 2009, DOI: 10.1118/1.3120285.

[2] 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.

[3] M. Gao, J. A. Fessler, and H.-P. Chan, "Training deep convolutional neural network with in silico data for denoising digital breast tomosynthesis images," in *Radiological Society of North America Scientific Assembly and Annual Meeting*, 2020, RSNA Program Book, SSPH01.

[4] 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.

[5] B. De Man, S. Basu, N. Chandra, B. Dunham, P. Edic, M. Iatrou, S. McOlash, P. Sainath, C. Shaughnessy, B. Tower, and E. Williams, "CatSim: a new Computer Assisted Tomography SIMulation environment," in *Proceedings of SPIE*, 2007, 65102G, DOI: 10.1117/12.710713.

[6] P. Milioni de Carvalho, "Low-dose 3D quantitative vascular x-ray imaging of the breast," Université Paris Sud, 2014.

[7] "Technical evaluation of GE Healthcare Senographe Pristina digital breast tomosynthesis system," *Public Health England*, 2019. [Online]. Available: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/771830/GE_Pristina_Tomo.pdf.

[8] J. Zheng, J. A. Fessler, and H.-P. Chan, "Segmented separable footprint projector for digital breast tomosynthesis and its application for subpixel reconstruction," *Medical physics*, vol. 44, no. 3, pp. 986–1001, Mar. 2017, DOI: 10.1002/mp.12092.

[9] H.-P. Chan, M. M. Goodsitt, M. A. Helvie, S. Zelakiewicz, A. Schmitz, M. Noroozian, C. Paramagul, M. A. Roubidoux, A. V. Nees, C. H. Neal, P. Carson, Y. Lu, L. Hadjiiski, and J. Wei, "Digital breast tomosynthesis: observer performance of clustered microcalcification detection on breast phantom images acquired with an experimental system using variable scan angles, angular increments, and number of projection views," *Radiology*, vol. 273, no. 3, pp. 675–685, 2014, DOI: 10.1148/radiol.14132722.

[10] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, and A. Courville, "Improved training of Wasserstein GANs," 31-Mar-2017. [Online]. Available: http://arxiv.org/abs/1704.00028.

[11] B. Sahiner, H.-P. Chan, L. M. Hadjiiski, M. A. Helvie, J. Wei, C. Zhou, and Y. Lu, "Computer-aided detection of clustered microcalcifications in digital breast tomosynthesis: a 3D approach," *Medical Physics*, vol. 39, no. 1, pp. 28–39, Jan. 2012, DOI: 10.1118/1.3662072.

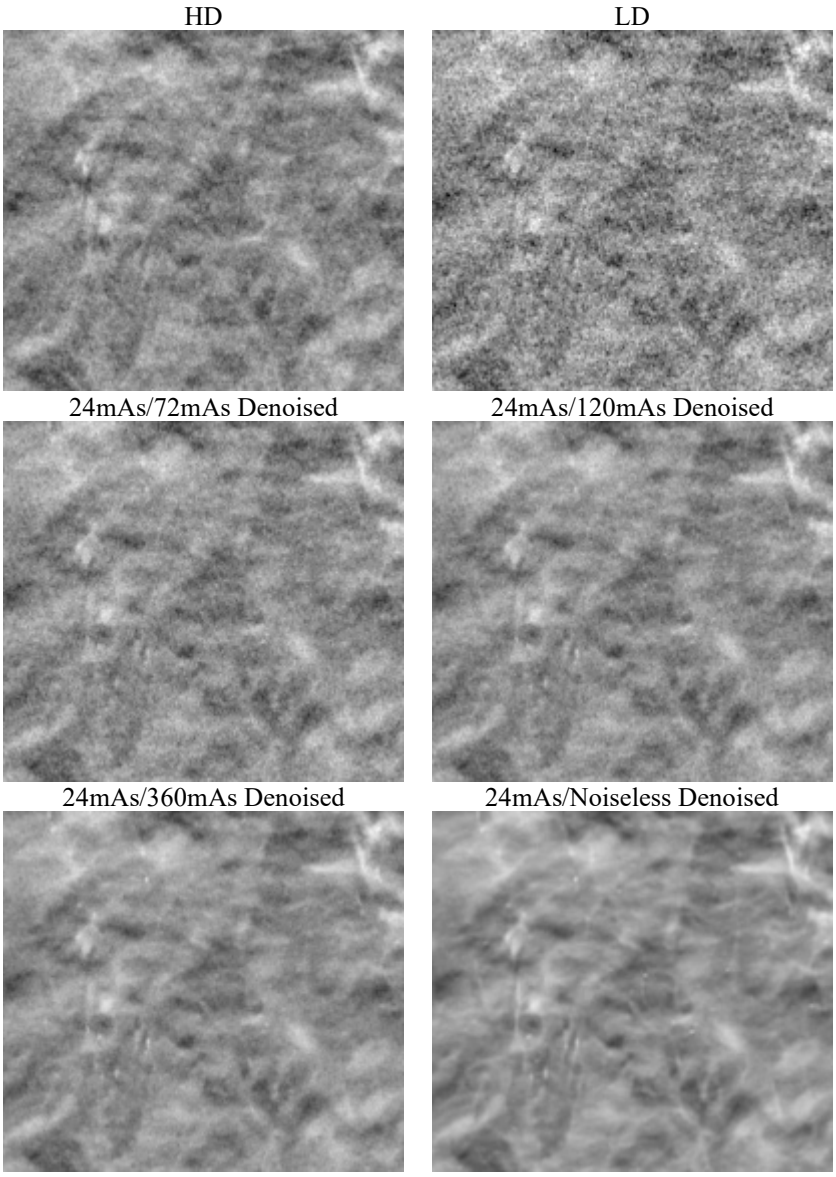

Figure 1. An example of a background 20 mm × 18 mm region in the validation physical phantom. Top row: high dose (HD) and low dose (LD) pair. Middle and bottom row: denoised with DCNN models based on different training/target pairs. The images are displayed with the same window/level settings.

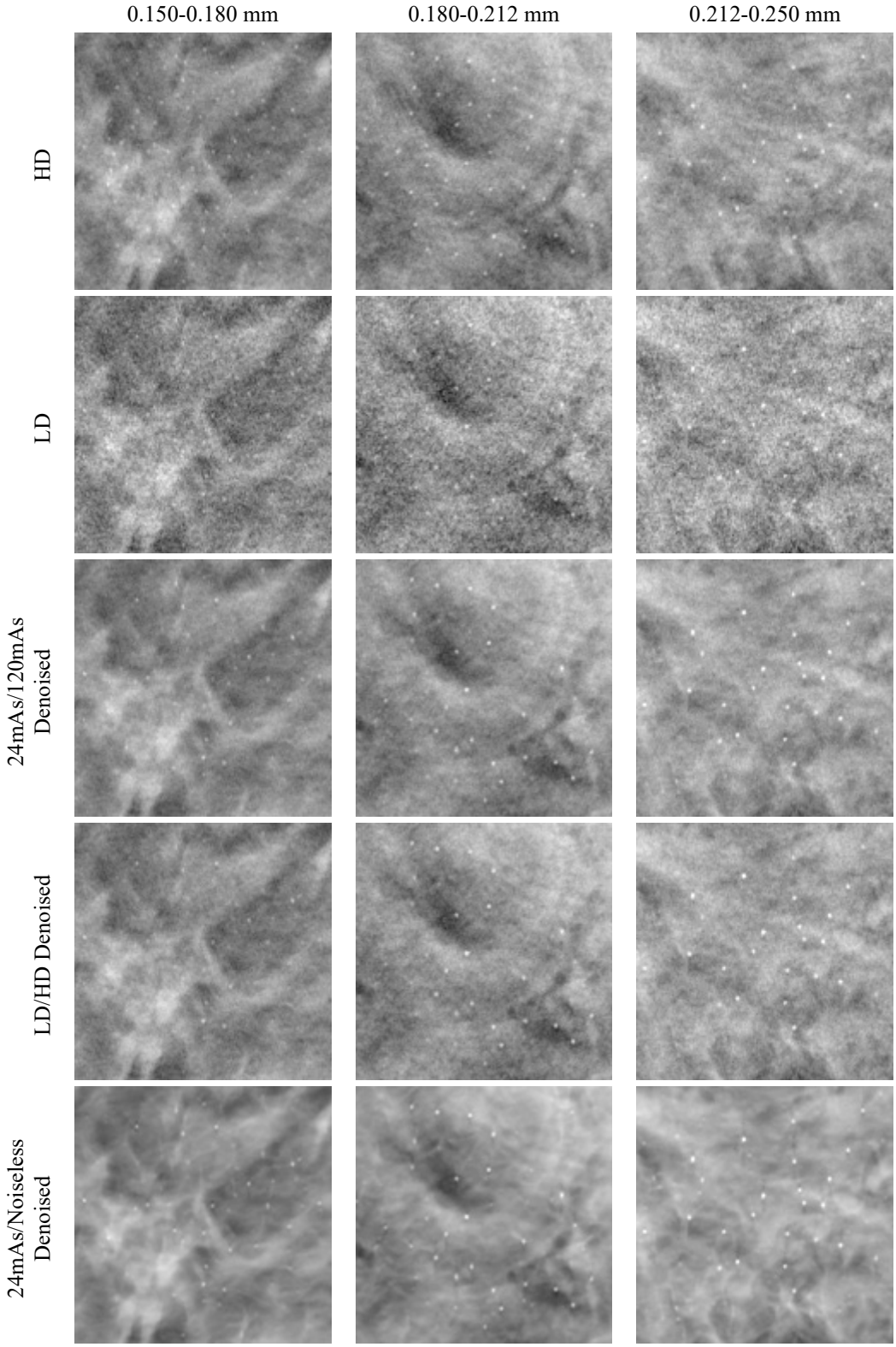

Figure 2. Three MC clusters of different sizes in the validation phantom. All images show a 15 mm × 15 mm region. The images in the same column are displayed with the same window/level settings.

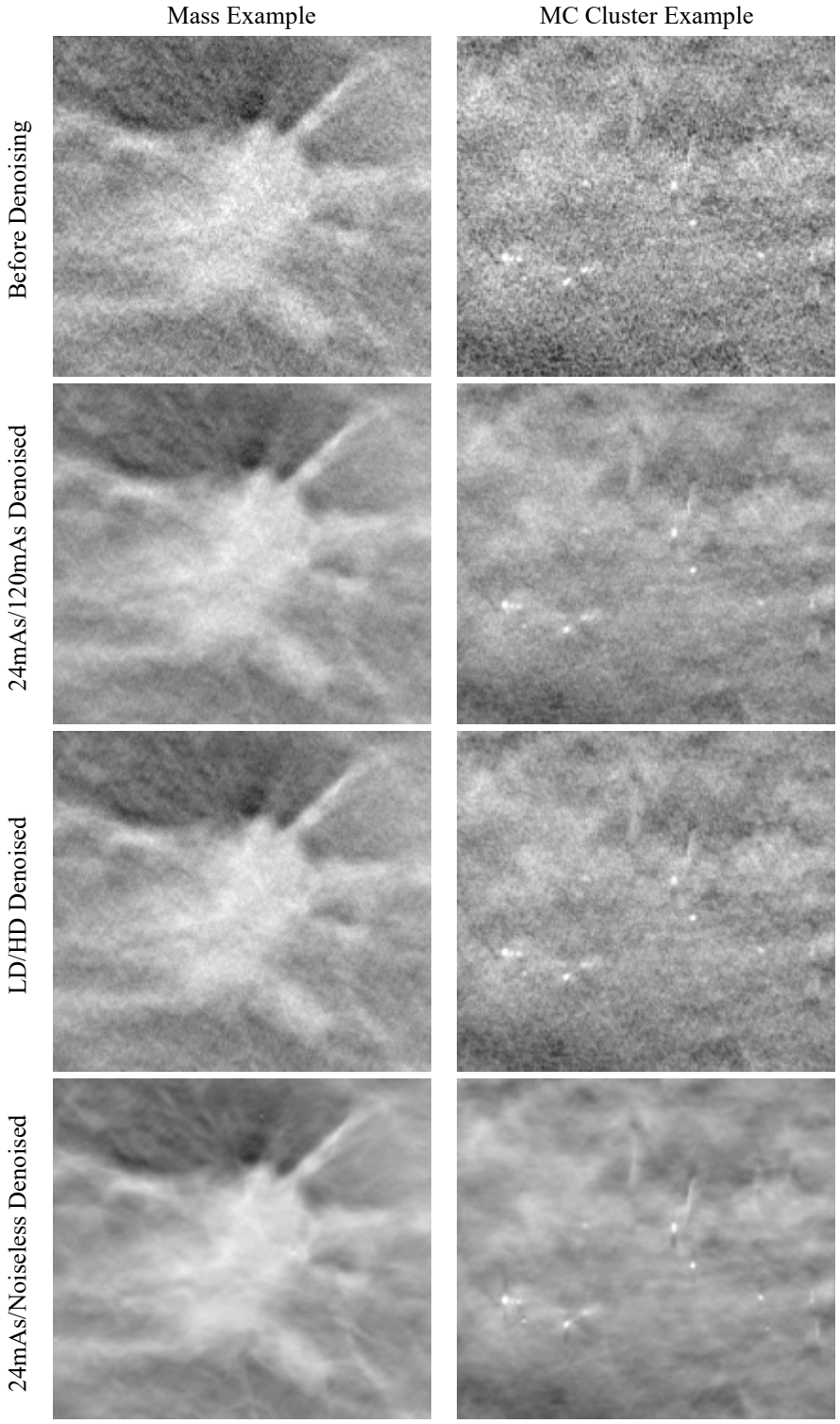

Figure 3. Example images of human subject DBTs with an invasive ductal carcinoma (mass) and a ductal carcinoma in situ (MC cluster). All images show a 20 mm × 18 mm region. The images in the same column are displayed with the same window/level settings.