Received: 30 August 2021 | Revised: 25 June 2022 | Accepted: 23 July 2022

DOI: 10.1002/mp.16007

RESEARCH ARTICLE

MEDICAL PHYSICS

Multi-scale cascaded networks for synthesis of mammogram to decrease intensity distortion and increase model-based perceptual similarity

Gongfa Jiang[1] | Zilong He[2] | Yuanpin Zhou[1] | Jun Wei[3,4] | Yuesheng Xu[5] | Hui Zeng[2] | Jiefang Wu[2] | Genggeng Qin[2] | Weiguo Chen[2] | Yao Lu[1,6]

[1] School of Computer Science and Engineering, Sun Yat-sen University, Guangzhou, P. R. China

[2] Department of Radiology, Nanfang Hospital, Southern Medical University, Guangzhou, P. R. China

[3] Perception Vision Medical Technology Company Ltd., Guangzhou, P. R. China

[4] Department of Radiology, University of Michigan, Ann Arbor, Michigan, USA

[5] Department of Mathematics and Statistics, Old Dominion University, Norfolk, Virginia, USA

[6] Key Laboratory of Machine Intelligence and Advanced Computing, Ministry of Education, Guangzhou, P. R. China

Correspondence
Yao Lu, School of Computer Science and Engineering, Sun Yat-sen University, Guangzhou 510275, China.
Email: luyao23@mail.sysu.edu.cn

Weiguo Chen, Department of Radiology, Nanfang Hospital, Southern Medical University, Guangzhou 510515, China.
Email: chenweiguo1964@21cn.com

Funding information
The China Department of Science and Technology under Key Grant, Grant/Award Numbers: 210YBXM, 2020109002; National Science Foundation, Grant/Award Numbers: 12126610, 81971691, 81801809, 81830052, 81827802, DMS-1912958; Construction Project of Shanghai Key Laboratory of Molecular Imaging, Grant/Award Number: 18DZ2260400; Guangdong Province Key Laboratory of Computational Science at the Sun Yat-sen University, Grant/Award Number: 2020B1212060032; National Cancer Institute of the National Institutes of Health, Grant/Award Number: R21CA263876

Abstract

Purpose: Synthetic digital mammogram (SDM) is a 2D image generated from digital breast tomosynthesis (DBT) and used as a substitute for a full-field digital mammogram (FFDM) to reduce the radiation dose for breast cancer screening. The previous deep learning-based method used FFDM images as the ground truth, and trained a single neural network to directly generate SDM images with similar appearances (e.g., intensity distribution, textures) to the FFDM images. However, the FFDM image has a different texture pattern from DBT. The difference in texture pattern might make the training of the neural network unstable and result in high-intensity distortion, which makes it hard to decrease intensity distortion and increase perceptual similarity (e.g., generate similar textures) at the same time. Clinically, radiologists want to have a 2D synthesized image that feels like an FFDM image in vision and preserves local structures such as both mass and microcalcifications (MCs) in DBT because radiologists have been trained on reading FFDM images for a long time, while local structures are important for diagnosis. In this study, we proposed to use a deep convolutional neural network to learn the transformation to generate SDM from DBT.

Method: To decrease intensity distortion and increase perceptual similarity, a multi-scale cascaded network (MSCN) is proposed to generate low-frequency structures (e.g., intensity distribution) and high-frequency structures (e.g., textures) separately. The MSCN consist of two cascaded sub-networks: the first sub-network is used to predict the low-frequency part of the FFDM image; the second sub-network is used to generate a full SDM image with textures similar to the FFDM image based on the prediction of the first sub-network. The mean-squared error (MSE) objective function is used to train the first sub-network, termed low-frequency network, to generate a low-frequency SDM image. The gradient-guided generative adversarial network's objective function is to train the second sub-network, termed high-frequency network, to generate a full SDM image with textures similar to the FFDM image.

Results: 1646 cases with FFDM and DBT were retrospectively collected from the Hologic Selenia system for training and validation dataset, and 145 cases with masses or MC clusters were independently collected from the Hologic Selenia system for testing dataset. For comparison, the baseline network has the same architecture as the high-frequency network and directly generates a full SDM image. Compared to the baseline method, the proposed MSCN improves the peak-to-noise ratio from 25.3 to 27.9 dB and improves the

Gongfa Jiang, Zilong He, and Yuanpin Zhou contributed equally to this work.

© 2022 American Association of Physicists in Medicine.

structural similarity from 0.703 to 0.724, and significantly increases the perceptual similarity.

Conclusions: The proposed method can stabilize the training and generate SDM images with lower intensity distortion and higher perceptual similarity.

KEYWORDS

breast cancer, deep learning, generative adversarial networks (GAN), digital breast tomosynthesis (DBT), synthetic mammogram

1 INTRODUCTION

A full-field digital mammogram (FFDM) is a widely used technique for breast cancer screening. However, FFDM suffers from overlapping tissue problems. Digital breast tomosynthesis (DBT), which is a 3D volume reconstructed from a series of low-dose projection images in a limited angle range, is used to address the overlapping tissue issue in FFDM. Large-scale studies have shown that using an additional DBT volume has higher accuracy in breast cancer detection.[1] However, using both DBT and FFDM for screening approximately doubles the radiation dose compared to FFDM alone.[2] However, to reduce the radiation dose, generating a synthetic digital mammogram (SDM) image from DBT and replacing FFDM with SDM is one possible solution.

Most of the previous studies on SDM solution development focused on using handcrafted features extracted from the DBT volume. Some of these studies used an edge-detection filter, gradient information, or a computer-aided detection (CAD) system to detect conspicuous points in DBT volume and combine them into a 2D image as the SDM.[3–5] These methods only used a part of the information in DBT to generate SDM images and might miss textural abnormalities. Another study required additional projection data to construct an SDM with enhanced microcalcifications (MCs), while the conspicuity of masses on the SDM was degraded.[6] There are several FDA-approved commercial SDM systems, such as the Hologic C-view, GE V-Preview, Siemens Insight, and Fujifilm S-View. The C-view image is created by re-projecting and filtering the central projection data and/or the stack of reconstructed DBT slices.[7] The image has an intrinsically different appearance and low overall resolution and noise properties compared to the FFDM image.[8,9] Clinical studies have shown that C-view + DBT have a performance similar to that of standard FFDM + DBT.[27,28,39] Intelligent 2D was newly developed by Hologic to further improve the performance of C-view. However, the enhancement may result in false positives due to pseudo-calcifications.[29,35,36] Besides, the C-view image provides poor overall resolution and noise properties compared to the FFDM image.[30,31,40,41] More importantly, large-scale clinical studies reported that more breasts' density is categorized as non-dense than dense when using C-view + DBT compared to using FFDM + DBT,[32,33,37,38] probably due to inherently different visual appearance between the C-view image and FFDM image. Since breast density has both imaging and risk implications[34] and is an important component of mammography reports and BI-RADS category classification, the C-view image may result in an inconsistent mammography report to FFDM image and unreliable risk assessment, which increases the recall of patients.

Deep learning has become a widely used solution for image-to-image translation.[10–12] Recent studies[13,14] proposed using gradient-guided generative adversarial networks (GGGAN) as the objective function for training a single deep convolutional neural network (DCNN) to directly generate an SDM image from the DBT volume. In GGGAN, a discriminator network using gradient maps as additional inputs is trained to distinguish between the generated images and ground-truth images. The generator network is trained to minimize the objective function based on the discriminator network to generate images with similar appearances and textures to the ground-truth FFDM image. The GGGAN was designed to maintain high-frequency structures, such as MCs in FFDM images. Other image-to-image regression objective functions, which are based on full-reference distortion measures,[15] such as mean-squared error (MSE) and perceptual loss,[16] can decrease the intensity distortion between the generated images and the ground truth. It has been shown in low-dose CT denoising tasks[17] that combining MSE with generative adversarial networks (GAN)[18] has a lower intensity distortion compared to using GAN only. Thus, by combining GGGAN with MSE and perceptual loss, we might generate SDM images with high-frequency structures (such as textures) and low-frequency structures (such as intensity distribution) similar to FFDM images.

However, FFDM has a different texture pattern from DBT because of the different radiation doses and detectors used in FFDM and DBT acquisition, and different post-processing techniques applied on acquired FFDM images and DBT volumes.[19] While using the FFDM image as the target, the difference in texture pattern might make the training of the DCNN unstable and result in high-intensity distortion, which makes it difficult to decrease intensity distortion and increase perceptual similarity (e.g., generate similar textures) simultaneously.

For FFDM, images contain low-frequency components such as mass and tissue background and high-frequency components such as calcifications and edges. Low-frequency components preserve intensity distribution and high-frequency components represent texture patterns. To overcome the above high-intensity distortion, we propose to apply two neural networks to generate low-frequency components and full images individually. Low-frequency component is generated first and then when a full image is generated to reproduce the texture pattern, a specific penalty is applied to force the previously generated low-frequency component to be preserved.

In this study, we proposed a multi-scale cascaded network (MSCN) that comprises two subnetworks to decrease intensity distortion and increase perceptual similarity. The first sub-network is trained to predict the low-frequency part (i.e., intensity distribution) of the FFDM image. Thus, the first subnetwork is called the low-frequency network. U-net[10] is used for the network architecture, and MSE is used as the objective function to decrease intensity distortion. U-net is widely used in the image-to-image translation task.[11] The second sub-network is trained to generate a full SDM image with high-frequency structures (e.g., textures) similar to the FFDM image based on the prediction of the first sub-network, and it is called the high-frequency network. A state-of-the-art network architecture in the image super-resolution task called residual-in-residual dense block (RRDB)[12] network is used for network architecture. The residual connections and dense connections in the RRDB network lead to the effective fusion of global and local features which can recover sharper edges and finer details. GGGAN is used as the objective function to preserve high-frequency structures and textures.

In the experiments, a single RRDB network that directly generated full SDM images was used as the baseline to show the performance gain of the proposed MSCN. The baseline network used the same network architecture and objective function as the high-frequency network in the MSCN. The peak-to-noise ratio (PSNR) and structural similarity (SSIM) were used to measure the intensity distortion between the generated SDM images and the ground-truth FFDM images. The learned perceptual image patch similarity (LPIPS),[20] which correlates well with human perceptual similarity judgments, was used to measure the perceptual similarity of the generated SDM images. In addition, a mass segmentation task was used to measure the ability of the proposed method to preserve the mass.

TABLE 1 The distribution of the 627 tumorous cases' features

Features	Number of cases
Mass	219
Micro-calcification	146
Architectural distortion	56
Asymmetry	58
Multiple types	148
Total	627

2 | MATERIALS AND METHODS

2.1 | Data

We retrospectively collected 1646 cases (1019 cases were normal and 627 cases were tumorous) with FFDM and DBT for the training and validation datasets. The distribution of the tumorous cases is shown in Table 1. Fifty normal cases and 31 tumorous cases were randomly selected for validation, and the validation dataset was only used for hyperparameter selection. We set normal cases as the blank group and cancer cases as the control group. The blank group has 42 cases in BI-RADS breast density type A, 114 cases in type B, 756 cases in type C, and 107 cases in type D. The control group has 24 cases in BI-RADS breast density type A, 98 cases in type B, 448 cases in type C, and 57 cases in type D. Overall, the dense type of breasts for the largest proportion (84.69%) in the blank group, and the same as in the control group is 80.54%. The average age of the blank group is 46.9 ± 9.67 years old. The control group is 51.1 ± 9.06 years old. We independently collected 145 cases with masses or MC clusters for testing. The masks of 239 masses in the 145 cases were manually annotated by a radiologist and used for mass quality evaluation. All the data were collected from the Hologic Selenia System. The slices of DBT volumes were resized to have the same pixel spacing as the FFDM image, and DBT volumes were padded with all zero slices on one side until each DBT had 96 slices. The gray level of the DBT volume is 10-bit, that is, ranging from 0 to 1023, and the gray level of the FFDM image is 12-bit, that is, ranging from 0 to 4095. The gray level of both DBT and FFDM images was rescaled to between -1 and 1 using a linear transformation.

2.2 | Multi-scale cascaded networks

In the proposed MSCN, we first trained the first subnetwork, which is denoted by the low-frequency network G^L, to generate low-frequency SDM. The low-frequency part of the FFDM images was extracted from the

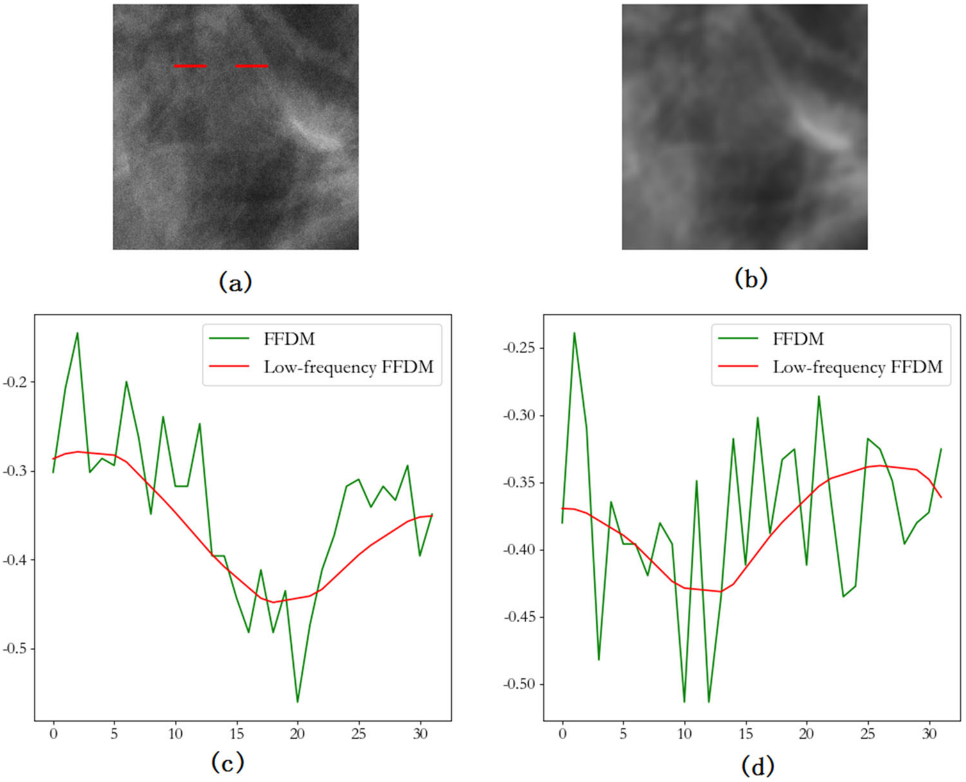

FIGURE 1 (a) FFDM image, which has a size of 256 × 256. (b) Low-frequency FFDM image (resized to have a size of 256 × 256). (c) The line profile of the FFDM image and low-frequency FFDM image of the left red line in (a). (d) The line profile of the FFDM image and low-frequency FFDM image of the right red line in (a). Best viewed in color

FFDM images and used as the ground truth to train the network G^L. We proposed an image processing operator, denoted by φ, to extract the low-frequency part of the FFDM image. A low-frequency FFDM image, denoted byI_D^L, is extracted by

$$I_D^L = \varphi(I_D) = \phi(\phi(I_D)), \quad (1)$$

where I_D is the original FFDM image, ϕ is a Gaussian smoothing operator (Gaussian kernel with a mean of 0 and a standard deviation of 1) followed by a bilinear down-sample with a factor of 2. An example of a low-frequency FFDM image is shown in Figure 1. Given a trained low-frequency network G^L, the low-frequency SDM is generated by

$$I_S^L = G^L(I_T), \quad (2)$$

where I_T is the DBT volume. We generated low-frequency SDM for all cases in the training, validation, and testing datasets. Subsequently, we trained an RRDB network, which is denoted by the high-frequency network G^H, to generate the high-frequency part of the FFDM image. In this study, we used residual learning[21] to train networkG^H. In residual learning, given a DBT volume I_T and a low-frequency SDM I_S^L, the output of the network G^H is added to the low-frequency SDM to derive a full SDM image, that is

$$I_S = G^H\left(I_T, I_S^L\right) + u\left(I_S^L\right), \quad (3)$$

where u is bilinear up-sampling with a factor of 4. Thereafter, the original FFDM images were used as the ground truth to train the network G^H.

2.3 | Network architecture

The network architecture of the low-frequency network G^L is shown in Figure 2. To extract 3D information in the input DBT volume, we used shared weight group convolution (SWGC)[13,14] in the encoder path of U-net. To increase the network's capacity, we used 16 RRDB blocks[12] in the U-net's lowest level. We used layer normalization[22] instead of batch normalization.

The network architecture of the high-frequency network G^H is shown in Figure 3. We used a similar architecture network as the state-of-the-art RRDB network in the image super-resolution task.[12] The input DBT was fed into a feature extraction truck to extract high-frequency features. High-frequency features were concatenated with low-frequency features in the

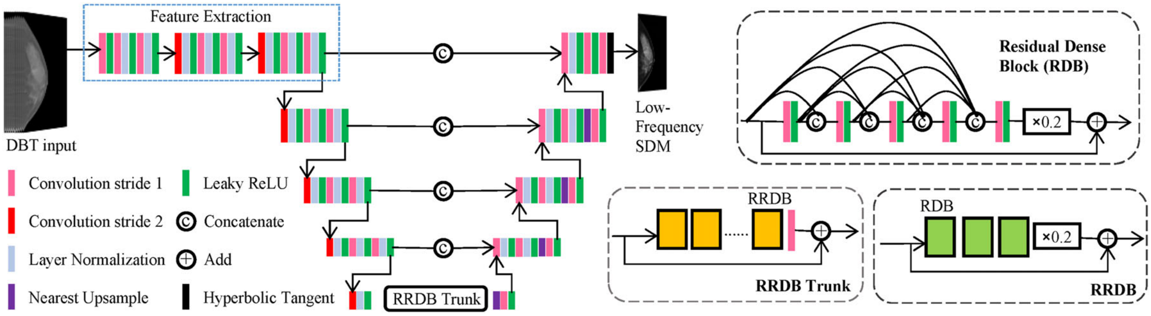

FIGURE 2 The diagram of the low-frequency network for low-frequency SDM generation. All the convolution layers in the feature extraction part (outlined with a blue dashed line) are shared weight group convolution (SWGC) layer. All the convolution layers have a kernel size of 3 × 3 and 64 output channels. There are 16 RRDB blocks in the RRDB trunk. Best viewed in color

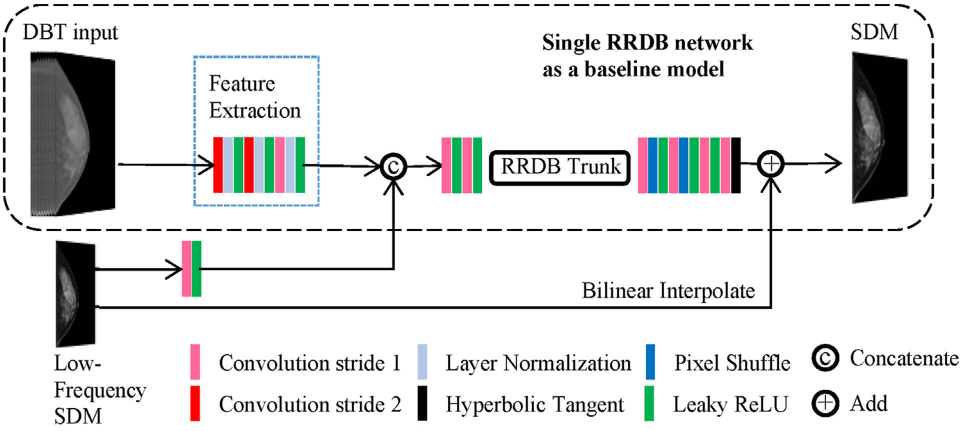

FIGURE 3 The diagram of the high-frequency network for full SDM generation. The first convolution layer in the feature extraction part (outlined with a blue dashed line) is the SWGC layer and has 256 output channels. All the other convolution layers have 64 output channels. All the convolution layers have a kernel size of 3 × 3. Low-frequency SDM is bilinear up-sampled with a factor of 4 and added to the output of the RRDB network to derive the full SDM. There are 16 RRDB blocks in the RRDB trunk. Best viewed in color

proposed MSCN. The features were then fed into the RRDB Trunk whose architecture was shown in Figure 3 to generate a high-frequency image. The high-frequency image was added to the low-frequency image and output the final SDM. To show the performance gain of the proposed MSCN, we used an RRDB network, which has the same network architecture as the high-frequency network, to directly generate full SDM images and used the RRDB network as the baseline model. The baseline network has no additional low-frequency SDM input or residual learning.

2.4 | Objective function

To train the low-frequency network G^L, we used the MSE objective function. The MSE loss is given by

$$L_{MSE}\left(G^L\right) = \frac{1}{|S|}\sum_{\left(I_D^L, I_T\right)\in S} \left\| I_D^L - G^L\left(I_T\right)\right\|_2^2, \qquad (4)$$

where S is the training dataset, $|S|$ is the size of the training dataset, and $\|\cdot\|_2$is l_2-norm.

To train the high-frequency networkG^H, we used GGGAN[13] with perceptual loss[14,16] and multi-frequency MSE loss as the regularization terms. In GGGAN, a discriminator network, which is denoted by D, is trained to distinguish between generated SDM image I_S and the ground truth imageI_D. The loss function for the discriminator D is given by

$$L_{GGGAN}(D) = \frac{1}{|S|}\sum_{\left(I_D^L, I_T\right)\in S, I_S} \left(\left\|\vec{1} - D\left(I_T, \left[I_D, I'_D\right]\right)\right\|_2^2 + \left\|\vec{0} - D\left(I_T, \left[I_S, I'_S\right]\right)\right\|_2^2\right), \qquad (5)$$

where $\vec{1} = [1, 1, \ldots, 1]^T$ and $\vec{0} = [0, 0, \ldots, 0]^T$, both have the same size as $D(\cdot)$, and I'_D/I'_S are the gradient maps of I_D/I_S. Sobel operators are used to extract the gradient

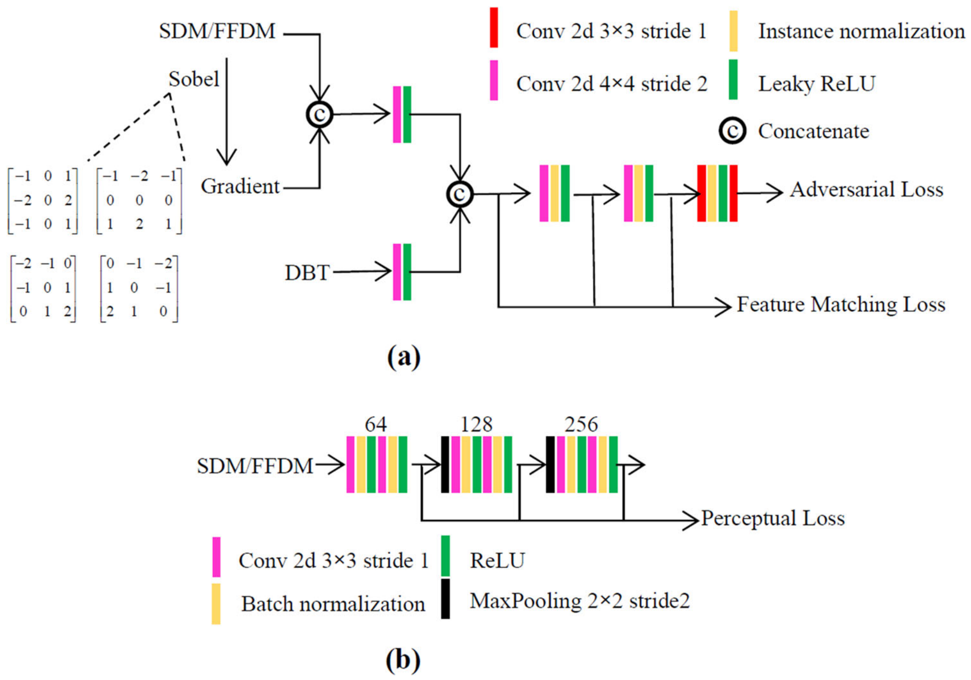

FIGURE 4 (a) The diagram of the discriminator network. In the first layer, the features with a channel size of 32 are extracted from SDM/FFDM and DBT separately and then concatenated. The output features of the next four convolution layers have a channel size of 128, 256, 512, and 1, respectively. (b) The diagram of the first seven convolution layers of the VGG-16 network. The output features of the first seven convolution layers have a channel size of 64, 64, 128, 128, 256, 256, and 256, respectively. Max-pooling layers with a stride of two are used to down-sample the features with a factor of 2. Best viewed in color

maps[13]. The architecture of the discriminator network is shown in Figure 4. Given a trained discriminator D, the adversarial loss and feature matching loss for the network G^H is given by

$$L_{Adv}\left(G^H\right)=\frac{1}{|S|}\sum_{I_T\in S,I_S}\left(\left\|\vec{1}-D\left(I_T,\left[I_S,I'_S\right]\right)\right\|_2^2\right),\quad(6)$$

$$L_{FM}\left(G^H\right)=\frac{1}{|S|}\sum_{\left(I^L_D,I_T\right)\in S,I_S}\sum_{j=1}^{T_D}\frac{1}{N^j_D}\left\|D^j\left(I_T,\left[I_D,I'_D\right]\right)-D^j\left(I_T,\left[I_S,I'_S\right]\right)\right\|_1,\quad(7)$$

where $D^j(\cdot)$ is the feature of jth layer of $D(\cdot)$, T_D is the total number of layers, N^j_D is the number of elements of the feature in the jth layer, and $\|\cdot\|_1$ is l_1-norm. The GGGAN objective function used to train the generator G^H is given by

$$L_{GGGAN}\left(G^H\right)=L_{Adv}\left(G^H\right)+\lambda_{FM}L_{FM}\left(G^H\right)\quad(8)$$

where λ_{FM} is a weighting factor for balancing the adversarial loss and feature matching loss. We empirically set $\lambda_{FM}=10$, which is the same as that in previous work[14]. The discriminator and generator were trained by minimizing Equations (5) and (8) in an alternative way.

In the perceptual loss, a pre-trained network is used to extract features from the generated SDM images and the ground-truth FFDM images; the distance between the features is then minimized. We used the VGG-16 network[23] pre-trained on ImageNet.[24] Given the pre-trained VGG-16 network, which is denoted by V, the perceptual loss is given by

$$L_{Percep}\left(G^H\right)=\frac{1}{|S|\,T_V}\sum_{I_D\in S,I_S}\sum_{j=1}^{T_V}\frac{1}{N^j_V}\left\|V^j\left(I_D\right)-V^j\left(I_S\right)\right\|_1,\quad(9)$$

where $V^j(\cdot)$ is the feature of jth layers of $V(\cdot)$, T_V is the total number of layers, and N^j_V is the number of elements of the feature in the jth layer. Due to the assumption that natural images only share low-level feature space with medical imaging, we only used the 2nd, 4th, and 7th convolution layers of the VGG-16 network (which has 13 convolution layers in total), which are the three lower layers of the VGG-16 network, to extract features. Thus, we set the number of layers $T_V=3$. The diagram of the first seven convolution layers of the VGG-16 network is shown in Figure 4.

For multi-frequency MSE loss, we used the MSE loss of full SDM images and the MSE loss of the low-frequency part of the full SDM images. Assuming that the residual learning cannot significantly change the intensity of the low-frequency SDM, we also used the MSE loss between the low-frequency part of the full SDM images and low-frequency SDM images. The multi-frequency MSE loss is

$$L_{MFMSE}\left(G^{H}\right) = \frac{1}{|S|} \sum_{I_D \in S, I_S^L, I_S} \left(\|I_D - I_S\|_2^2 + \|\varphi(I_D) - \varphi(I_S)\|_2^2 + \left\| I_S^L - \varphi(I_S) \right\|_2^2 \right). \tag{10}$$

Then the objective function for the high-frequency network G^H is

$$L\left(G^{H}\right) = L_{GGGAN}\left(G^{H}\right) + \lambda_{percep} L_{Precp}\left(G^{H}\right) + L_{MFMSE}\left(G^{H}\right) \tag{11}$$

where λ_{percep} is a weighting factor that balances the GGGAN loss and perceptual loss. We empirically set λ_{percep} to 10, which is the same as that in previous work.[14] We used the same objective function for the training of the baseline model, except that the third term of the multi-frequency MSE loss (Equation 10) was not included.

2.5 | Training details

The low-frequency network (U-net) was trained on full-sized images, whereas the high-frequency network (RRDB network) was trained on patches. For the RRDB network training, the DBT volumes and FFDM images in the training and validation datasets were cut into patches at a resolution of 512 × 512 without overlapping. Patches with over 50% background were discarded for better training convergence. Note that the RRDB network is a fully convolutional network. Thus, the RRDB network can use full-sized DBT as the input and generate a full-sized SDM in the testing phase.

For the training of the U-net, we used the Adam[25] solver with a learning rate of 1×10^{-4}, $\beta_1 = 0.9$, and $\beta_2 = 0.999$. The batch size was set to 1 owing to the limitations of the GPU memory. Horizontal flip augmentation was used for all images, and vertical flip augmentation was used for images from the CC-view mammogram. The U-net was trained for 400 000 iterations, and the learning rate was set to 1×10^{-5} after 300 000 iterations. The training takes about 10 days on an NVIDIA RTX 8000 GPU.

For the training of the RRDB network, we used the Adam solver with a learning rate of 1×10^{-4}, $\beta_1 = 0.5$, and $\beta_2 = 0.9$. The batch size was set to four owing to the limitations of the GPU memory. Horizontal flip augmentation was used for all patches, and vertical flip augmentation was used for patches from the CC-view mammogram. The RRDB network was trained for 100 000 iterations, and the learning rate was set to 1×10^{-5} after 50 000 iterations. The training took approximately two days on an NVIDIA TitanX GPU.

3 | RESULTS

3.1 | Evaluation

To show the performance gain of the proposed MSCN, we trained an RRDB network to directly generate full SDM images and used the RRDB network as the baseline model (denoted by RRDB below). For the baseline model, we used the same network architecture and the same objective function (Equation 11) as the high-frequency network in the proposed MSCN (shown in Figure 3), except that the baseline model has no residual learning and additional low-frequency SDM input. Thus, the third term of the multi-frequency MSE loss for the baseline model was not included. The main difference between the proposed MSCN and the baseline model is that the proposed MSCN uses the low-frequency SDM generated by the low-frequency network.

In the experiment, we measured the intensity distortion, perceptual similarity, and mass quality of the SDM images derived using the proposed and baseline methods. To measure the intensity distortion of the generated SDM images, we used the PSNR and SSIM. To measure the perceptual similarity of the generated SDM images, we used the LPIPS,[20] which correlates well with human perceptual similarity judgments. To measure the mass quality of the generated SDM images, we trained a U-net to predict the mask of masses in the generated SDM images, and then measured the dice similarity coefficient between the mask of the generated SDM images and the mask of FFDM images.

3.2 | MSE loss

The MSE loss curves of training and validation on the full SDM images and low-frequency parts of the full SDM images are shown in Figure 5. As can be seen, the proposed MSCN has more stable training and a lower MSE loss than the RRDB in both full SDM images and low-frequency parts of the full SDM images. Using the low-frequency SDM generated by the low-frequency network can stabilize the training and might result in a better local minimum.

3.3 | Intensity distortion

We used PSNR and SSIM to measure the intensity distortion of the generated SDM images. The average

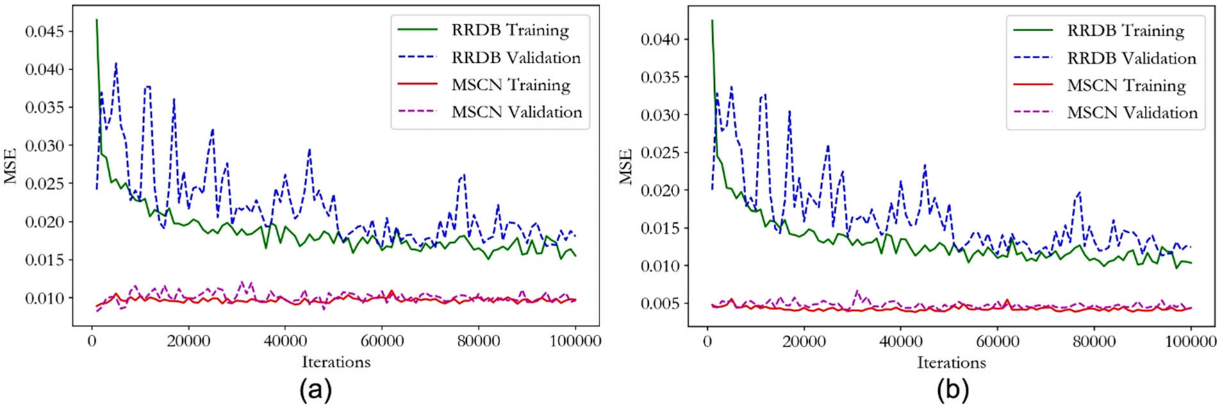

FIGURE 5 (a) MSE loss curves of training and validation on full SDM images. (b) MSE loss curves of training and validation on the low-frequency part of the full SDM images. Best viewed in color

values of PSNR and SSIM of the SDM images generated by the proposed MSCN and RRDB are shown in Table 1. The line profiles of the two representative cases are shown in Figure 6. As can be seen, the proposed MSCN significantly ($p < 1 \times 10^{-6}$) outperforms the RRDB in terms of PSNR and SSIM and has a lower intensity distortion than the RRDB. The proposed MSCN can decrease intensity distortion and make the SDM images more similar to those of the FFDM images.

3.4 | Perceptual similarity

We used the LPIPS,[20] which correlates well with human perceptual similarity judgments, to measure the perceptual similarity of the generated SDM images. In LPIPS, a neural network was trained on a dataset of human perceptual similarity judgments, and the neural network was used to extract features to measure the dissimilarity between the generated images and the ground-truth images. Lower LPIPS values were related to higher perceptual similarity. The results are listed in Table 2. We used paired *t*-test to compare the perceptual similarity between MSCN and RRDB. As can be seen, the proposed MSCN has a significantly ($p < 1 \times 10^{-6}$) higher perceptual similarity than the RRDB.

3.5 | Mass quality

To measure the mass quality of the generated SDM images, we trained the same U-net as in the previous work[14] for the mass segmentation task. We used an independently collected in-house dataset including 673 masses and manually annotated the masks of masses as the training dataset, which did not overlap with the dataset described before. For each mass, a patch with the size of 1024 × 1024 and a mass in the center was cropped out from the FFDM image and used as the input image. During the training, smoothed dice loss was used as the objective function. The Adam solver with a learning rate of 2×10^{-4}, $\beta_1 = 0.5$, and $\beta_2 = 0.9$ was used. The batch size was set to eight. Horizontal flip, vertical flip, random rotation of 90°,180°,270°, and random resizing were used for training augmentation. The network was trained for 150 epochs (approximately 12 500 iterations). We used an ensemble of five trained U-net[26] to derive a more robust segmentation result.

We used the mean dice scores of the predicted masks of the SDM and FFDM images, using the manually annotated mask as the ground truth. In addition, we used a semantic similarity score[14] to take the predicted mask of FFDM images as the ground truth for the calculation of a dice score, to directly evaluate the similarity of masses between SDM and FFDM images. The results are presented in Table 3. As can be seen, the SDM images derived from the proposed MSCN and the RRDB have the same segmentation results as the FFDM images. However, the SDM image derived from the proposed MSCN has significantly ($p < 1 \times 10^{-4}$) more similar to the FFDM image than that derived from the RRDB (see Table 4).

A representative result is shown in Figure 7. As can be seen, the mass of the SDM image derived from the RRDB has a higher intensity than the mass of the FFDM image, which results in over-segmentation and a mask inconsistent with the mask of the FFDM image. However, the proposed MSCN can derive an SDM image with an intensity more similar to the FFDM image than the RRDB, resulting in a mask that is more consistent with the mask of the FFDM image (see Figure 8).

3.6 | FFDM versus SDM reader study

In order to demonstrate the improved performance of the proposed reading by radiologists, a reader study was provided as an additional experiment. To reflect the differences between the two imaging methods, FFDM and SDM were reviewed by six radiologists with different

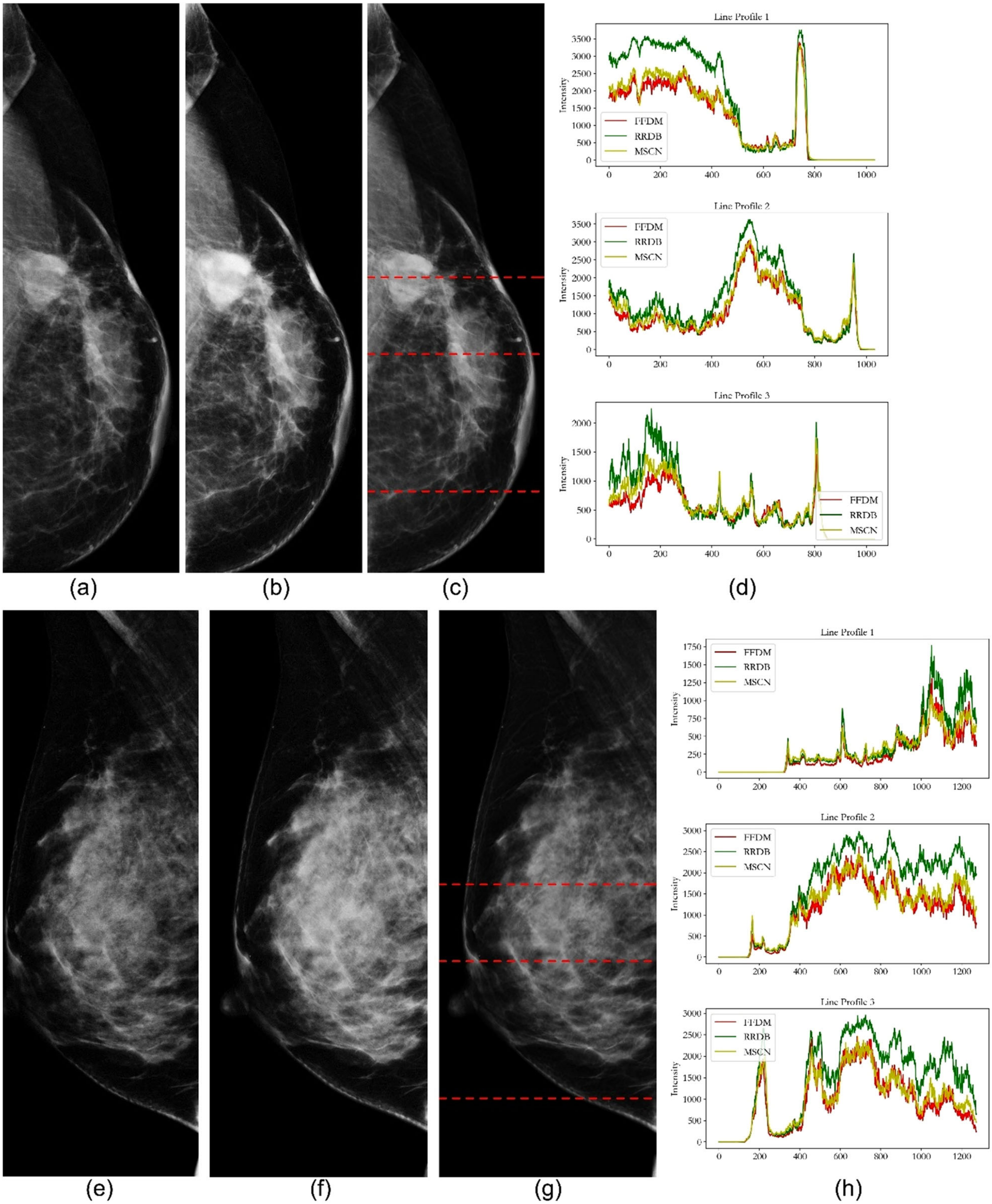

FIGURE 6 (a,e) FFDM image of the two representative cases. (b,f) SDM image derived by the RRDB network. (c,g) SDM image derived by the proposed MSCN. (d,h) Line profiles (FFDM (red), RRDB network (green), proposed MSCN (yellow)) of the three red dashed lines in (c) and (g). Best viewed in color

TABLE 2 PSNR and SSIM (mean±standard deviation) of the RRDB network and the proposed MSCN

	PSNR	SSIM
RRDB	25.333±2.15	0.7028±0.069
MSCN	**27.892±1.99**	**0.7238±0.072**

TABLE 3 LPIPS (mean±standard deviation) of the RRDB network and the proposed MSCN

	LPIPS
RRDB	0.1160±0.033
MSCN	**0.1077±0.033**

TABLE 4 Dice score (mean±standard deviation) of FFDM images, the RRDB network, and the proposed MSCN on the mass segmentation task. Significance (*p*-value) of the difference between the results of FFDM and SDM is listed. Semantic similarity is the dice score (mean±standard deviation) between the predicted mask of SDM and the predicted mask of FFDM images

	Dice score	*p*-Value	Semantic similarity
FFDM	0.7939±0.147	-	-
RRDB	0.7938±0.141	0.977	0.9312±0.093
MSCN	0.7981±0.145	0.0594	**0.9546±0.043**

years of experience, and also analyzed the differences between the detection effect and diagnostic efficiency of lesions in the images in this study.

Readers consisted of six breast diagnostic radiologists, including three groups, the junior group: reader 1 and reader 2 had 2 years of breast image diagnosis experience; the middle-seniority group: reader 3 and reader 4 had 4 years of breast image diagnosis experience; the senior group: reader 5 and reader 6 had 6 years of breast image diagnosis experience.

Data are collected according to the requirements of the reading experiment of multiple readers. In order to ensure sufficient diagnostic power, 120 cases are required when the diagnostic power is set to 1 and the readers are set to six. Therefore, 145 cases with unilateral lesions were randomly selected between 1 January 2014 to 31 December 2020, from Nanfang Hospital, Southern Medical University, Guangdong, Guangzhou, China. This retrospective study was approved by the Institutional Review Board (IRB) approved protocol (code number NFEC-2018-037), and informed consent was waived. Also, FFDM images, radiologists' reports and pathological gold standard results were collected. All images were numbered and desensitized after collection. The inclusion criteria are that the case must have unilateral breast lesions. The exclusion criteria are: (1). unilateral multiple lesions; (2). no pathological gold standard results or surgical puncture results were found.

In the first session, all cases were desensitized and numbered after collection. The image is put into the deep learning model for processing, and the corresponding SDM image is the output, which is divided into two groups, one group is the FFDM image, and the other is the SDM image. After merging the two groups of images, they were randomized and numbered to obtain the experimental data set.

In the second session, preparation for the reading test: (1) train the readers before film reading, and extract an image of another case for trial; (2) description of diagnostic rules: diagnosis is carried out on a case-by-case basis, and the main lesions are located and evaluated by BI-RADS (benign below 4a and malignant above 4a).

If the following conditions are met, it is the correct label: (1) the lesion is located correctly (left and right, quadrant and depth of the lesion); (2) the main lesions were diagnosed correctly, corresponding to pathological results. If the following conditions occur, it will be judged as wrong marking: (1) the localization of the lesion is wrong marking; (2) the diagnosis of main lesions is wrong.

In the third session, the reading experiment: images of 290 cases were reviewed by six readers. All cases are interpreted using a three-monitor Hologic diagnostic workstation (SecurViewDx, Hologic MA), which was calibrated to the DICOMGSDF and enabled zooming in or out. As the database was randomized and merged, each reader would have a different case order.

For each case, the location (left and right), quadrant, depth (front, middle and rear), type of lesion (mass, calcification, architecture distortion, asymmetry), image type (FFDM, SDM), benign and malignant, BI-RADS category (0–5), and probability of malignancy (%) were recorded by the readers. The reading time was not limiting.

ROC curves, sensitivity, specificity, accuracy, PPV, and NPV of six readers were calculated and analyzed, as well as the consistency of lesion detection in two image types of each reader. All statistical analyses were based on the R language and SPSS 25.00.

The age distribution, BI-RADS breast density and benign and malignant distribution of lesions of 145 patients are shown in Table 5.

The specific location, depth of distribution, and quadrant of the lesions in the 145 cases and the assessment of the BI-RADS category in the radiologists' report are shown in Table 6.

All 145 cases underwent pathological biopsy, including 142 malignant lesions and 3 benign lesions. Most of the malignant lesions were invasive ductal carcinoma. The specific pathological types were distributed as shown in Table 7.

The detection by six readers in FFDM and SDM is shown in Table 8. It can be found that readers 1, 3, and 5 all performed well in FFDM and SDM, while reader 2's detection ability was relatively low in six readers, and 15 cases of errors were detected in FFDM and SDM (see Table 9–10).

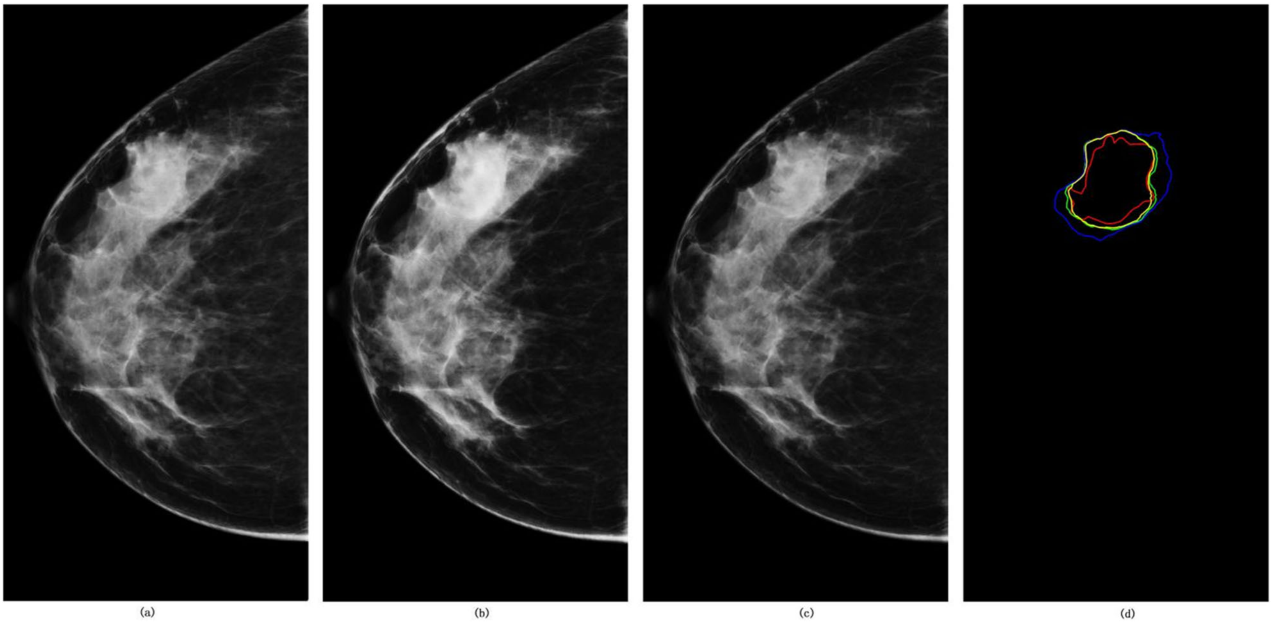

FIGURE 7 (a)FFDM image. (b) SDM image derived from the RRDB network. (c) SDM image derived from the proposed MSCN. (d) Contours of the masks: ground truth (red), FFDM image (green), RRDB network (blue), and proposed MSCN (yellow). Best viewed in color

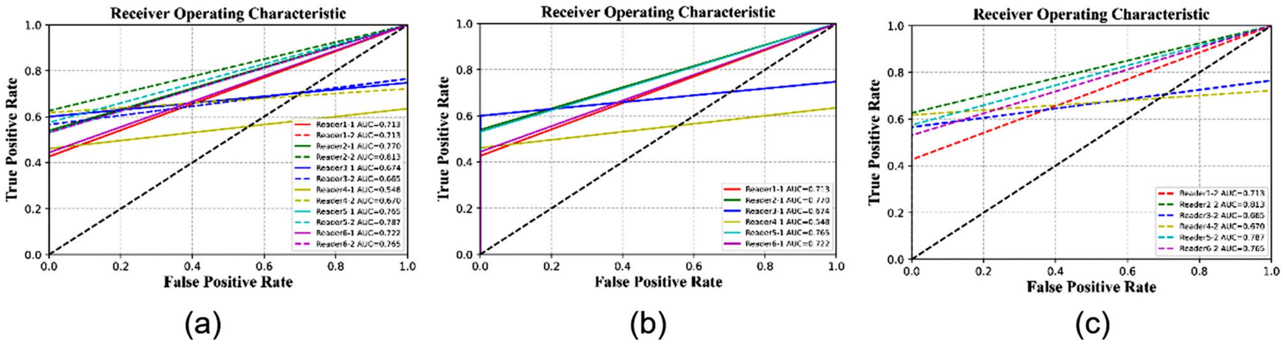

FIGURE 8 ROC curves for six readers in three groups diagnosed with FFDM and SDM. (a) Six readers' ROC curves in two image types; (b) six readers' ROC curves in FFDM; (c) six readers' ROC curves in SDM. We can observe that ROCs have similar upward trends in the two image types.

In the group of six readers in FFDM and SDM, there was statistical significance for the detection of lesions, which indicated that there was a correlation between the readers and the detection of the lesion and also indicated no clear influence between the detection of lesions and the image type.

Comparing the diagnostic efficacy of FFDM and SDM of six viewers, we found that the sensitivity and PPV of the two image types of readers 1, 2, 3, 5, and 6 were consistent. The specificity and NPV could not be calculated due to the large difference between benign and malignant proportions. The AUC values of the two image types of the same reader were roughly the same. The senior group had higher diagnostic efficiency, with an average of 0.760, the average AUC value of the middle seniority group was 0.639, and the average AUC value of the junior group was 0.750. The junior group performed better than the middle-seniority group.

Two image types of FFDM and SDM in different readers, different image types have no influence on the detection of lesions. The consistency of the same reader in different image types was high, and the kappa values of readers 1, 2, 4, 5, and 6 were all greater than 0.7, indicating that the coincidence degree of the two images in the detection of lesions was statistically significant and had a strong consistency. However, the kappa value of reader 3 was only 0.5, indicating that its ability to detect lesions in two different images was slightly weaker than that of other readers. On the other hand, different image modes have no significant influence on the detection of lesions in the viewers, and the detection of lesions in FFDM and SDM is roughly the same.

TABLE 5 Characteristics of the population for this study

Variable	Test set (*n* = 145)
Patient age (y)	
Mean	48.89
Median	48
Range	24–73
Interquartile range	48–73
BI-RADS breast density	
a	4
b	24
c	107
d	10
Number of each class	
Benign	3
Malignant	142

TABLE 6 The characteristics of mass and associated features in three group

Main features		Test set (*n* = 145)
Laterality	Left	75
	Right	70
Quadrant	Outer upper	78
	Inner upper	12
	Outer lower	7
	Inner lower	20
	Axillary region	2
	Central area	2
	Subareolar region	5
	Other	19
Depth	Front	15
	Middle	59
	Rear	71
BI-RADS category	5	92
	4c	31
	4b	13
	4a	9
	3	0
	2	0
	1	0
	0	0

When chi-square analysis was performed between groups with different image modes, it was found that there was a significant correlation between image types' lesions detection and the readers, suggesting that image modes did not affect the lesion detection ability of the readers.

TABLE 7 Histopathology results

Variable	Test set (*n* = 145)
Histopathology	
Adenosis of breast	1
Basal-like breast carcinoma	5
Chronic suppurative inflammation	0
Cyst of galactostasia	0
Ductal carcinoma in situ (DICS)	8
Epidermal cyst	0
Fibroadenoma	0
Fibroadenosis	2
Fibrous adipose tissue and breast ducts	0
Granulomatous mastitis	0
Interstitial fibers proliferate	0
Intraductal papilloma	0
Paget disease	1
Invasive ductal carcinoma	120
Invasive lobular carcinoma	3
Leukemia	0
Mammary neuroendocrine carcinoma	0
Mixed invasive carcinoma (ILC+IDC)	1
Metaplastic breast carcinoma	3
Mucinous carcinoma	1
Papilloma	0
Phyllodes tumors	0
Pure cyst	0
Sclerosing adenosis	0
Suppurative mastitis	0
Tubular carcinoma	0

TABLE 8 The detection and each reader's kappa valve in FFDM versus SDM

	FFDM		**SDM**	
	Y	**N**	**Y**	**N**
reader 1	140	5	140	5
reader 2	130	15	130	15
reader 3	144	1	142	3
reader 4	138	7	135	10
reader 5	140	5	142	3
reader 6	141	4	139	6

	Kappa value	**Approximate significance**
reader 1	1.000	0.000
reader 2	0.930	0.000
reader 3	0.500	0.000
reader 4	0.812	0.000
reader 5	0.743	0.000
reader 6	0.793	0.000

TABLE 9 The group's chi-square Test in FFDM versus SDM

	FFDM					
	reader 1	reader 2	reader 3	reader 4	reader 5	reader 6
Y	140	130	144	138	140	141
N	5	15	1	7	5	4
Total	145	145	145	145	145	145
	SDM					
	reader 1	**reader 2**	**reader 3**	**reader 4**	**reader 5**	**reader 6**
Y	140	130	142	135	142	139
N	5	15	3	10	3	6
Total	145	145	145	145	145	145

Chi-square tests	FFDM	SDM
Pearson chi-square	19.11	16.511
Likelihood ratio	17.968	15.42
Asymptotic significance (2-sided)	0.002	0.006

TABLE 10 The diagnostic efficacy parameters in the observer study of each reader

	Sensitivity	Specificity	AUC	Accuracy	PPV	NPV
Read 1_1	0.991	0.000	0.713	0.983	0.991	0.000
Read 1_2	0.991	0.000	0.713	0.983	0.991	0.000
Read 2_1	0.974	0.000	0.770	0.965	0.991	0.000
Read 2_2	0.974	0.000	0.813	0.966	0.991	0.000
Read 3_1	0.991	0.000	0.674	0.983	0.991	0.000
Read 3_2	0.991	0.000	0.665	0.983	0.991	0.000
Read 4_1	0.965	0.000	0.548	0.957	0.991	0.000
Read 4_2	0.974	0.000	0.670	0.966	0.991	0.000
Read 5_1	0.965	0.000	0.765	0.957	0.991	0.000
Read 5_2	0.965	0.000	0.787	0.957	0.991	0.000
Read 6_1	0.991	0.000	0.722	0.983	0.991	0.000
Read 6_2	0.991	0.000	0.765	0.983	0.991	0.000

Analyzing the diagnostic results of each reader under FFDM and SDM image type, we found that the sensitivity PPV of viewers 1, 2, 3, 5, and 6 were consistent. The AUC values of the two image types are roughly the same for the same viewer.

In conclusion, under FFDM and SDM image types, the detection and diagnosis abilities of the six readers were roughly the same, indicating that FFDM and SDM had roughly the same screening and diagnosis effects on breast lesions.

3.7 | Observer study about pseudo-calcifications in SDM

In order to reflect the influence of pseudo-calcifications, we make an observer study about pseudo-calcifications in SDM. A total of 20 normal cases were reviewed by two radiologists with the same experience (3 years), and also analyzed the consistency in this study.

Readers consist of two breast diagnostic radiologists who had 3 years of breast image diagnosis experience. Data are collected according to the requirements of the reading experiment. In order to ensure sufficient diagnostic power, 20 cases are required when the diagnostic power is set to 1 and the readers are set to six. Therefore, 20 normal cases, which were confirmed by follow-up 1 year, were randomly selected between 1 January 1 2022 to 31 March 31 2022, from Nanfang Hospital, Southern Medical University, Guangdong, Guangzhou, China. This retrospective study was approved by the Institutional Review Board (IRB) approved protocol (code number NFEC-2018-037), and informed consent was waived. Also, FFDM images, radiologists' reports and

TABLE 11 The group's McNemar test in SDM of detection of pseudo-calcifications

	SDM	
	reader 1	reader 2
Y	2	1
N	18	19
Total	20	20
McNemar tests		
		SDM
Pearson chi-square		0.36
Likelihood ratio		0.37
Asymptotic significance (2-sided)		0.55
Kappa		0.05
Approximate significance		0.55

pathological gold standard results were collected. All images were numbered and desensitized after collection. The inclusion criteria are: (1) BI-RADS 1 and follow-up at least 2 years. The exclusion criteria are: (1) multiple lesions; (2) BI-RADS equal to or greater than 2.

In the first session, all cases were desensitized and numbered after collection. The image is put into the deep learning model for processing, and the corresponding SDM image is output. They were randomized and numbered to obtain the experimental data set.

In the second session, preparation for reading test: (1) train the readers before film reading and extract an image of another case for trial; (2) description of diagnostic rules: diagnosis is carried out on a case-by-case basis, and find out pseudo-calcifications (Y/N).

In the third session, the reading experiment: images of 20 cases were reviewed by two readers. All cases are interpreted using a three-monitor Hologic diagnostic workstation (SecurViewDx, Hologic MA). For each case, yes (have pseudo-calcifications) or no (do not have) were recorded by the readers.

The consistency and kappa value of the two readers were calculated and analyzed, as well as the consistency of pseudo-calcifications detection in the SDM of each reader. All statistical analyses were based on the R language and SPSS 25.00. The results are shown in Table 11.

In the group of two readers in SDM, there was no statistical significance for the detection of pseudo-calcifications, which indicates that there was no correlation between the readers and the detection of pseudo-calcifications and also indicates indirect proof that the proposed model might overcome this tissue.

4 DISCUSSION

In this work, we proposed an MSCN to make the training more stable, decrease intensity distortion, and increase perceptual similarity. In the proposed method, low-frequency structures (e.g., intensity distribution) and high-frequency structures (e.g., textures) were generated separately. A single network that directly generated the full SDM images was used as the baseline model. The experiment results show that the training curve of the proposed MSCN is more stable than that of the baseline model. The baseline model has a PSNR of 25.3 dB, SSIM of 0.703, and LPIPS of 0.116, while the proposed MSCN has a PSNR of 27.9 dB, SSIM of 0.724, and LPIPS of 0.1077. An additional reader study was performed to compare the difference between FFDM and SDM. Six radiologists with different years of experience reviewed and analyzed the differences between the detection effect and diagnostic efficiency of lesions in the images. Radiologists found these two images were roughly the same. The proposed MSCN decreases the intensity distortion and increases the perceptual similarity. It can also generate SDM images with masses that are more similar to FFDM images than the baseline model. The proposed MSCN can stabilize the training process and improve the image quality of SDM images.

The low-frequency network was proposed to generate low-frequency structures (e.g., intensity distribution). The output target of the low-frequency network was the low-frequency FFDM which was smoothed by a Gaussian smoothing operator (Gaussian kernel with a mean of 0 and a standard deviation of 1) followed by a bilinear down-sample with a factor of 2. We called the generated image low-frequency SDM. Since high-frequency structures were removed from the target, it is easier for the low-frequency network to learn important low-frequency structures. Experimental results showed using the low-frequency SDM generated by the low-frequency network can stabilize the training and might result in a better local minimum.

We found that disentangling FFDM images into low-frequency images and high-frequency images is beneficial to learning multi-scale structures more efficiently. We use two DNNs to generate the low-frequency SDM images and high-frequency SDM images successively, which allows the two DNNs to learn structures under different scales independently. More importantly, disentangling low-frequency structures and high-frequency structures and using two different networks to learn the structures allow us to use low-frequency specific loss function and high-frequency specific loss function for different networks respectively. With the cascaded networks design, the low-frequency structures are combined with the high-frequency structures, which fusion into the target SDM images.

In order to reflect the differences between the two imaging methods, a reader study of FFDM and SDM was reviewed by six radiologists with different years of experience, and also analyzed the differences between the detection effect and diagnostic efficiency of lesions in the images in this study. By analyzing the diagnostic results of each reader under FFDM and SDM image type, we found that the sensitivity PPV of viewers 1, 2, 3, 5, and 6 were consistent. The AUC values of the two image types are roughly the same for the same viewer, which means the detection and diagnosis abilities of the six readers were roughly the same, indicating that FFDM and SDM had roughly the same screening and diagnosis effects on breast lesions.

The major limitation of this work is that statistical results of comparisons between the proposed method and C-view/Intelligent 2D, such as a comparison in terms of breast density consistency, were not provided. Since both C-view and Intelligent 2D were not approved by the FDA of China, we could only collect a limited number of images from one hospital which conducted a clinical trial for Hologic C-view. The findings in the visual comparison are preliminary and the conclusions lack significance due to insufficient available cases. However, we do think it is valuable to provide this result in order to make audiences have enough confidence and interest to try our method if they have enough data. We will further quantify the performance of the proposed method compared with commercial SDM solutions when we collect sufficient C-view data in the future.

Another major limitation is that the trained generator DCNN can only be used for data acquired from the Hologic system since it only learned the transformation from DBT volume to FFDM image of the Hologic system. To investigate the potential capacity of the proposed method to transfer to an unseen machine system, more work is needed in the future to further quantify the cross-vendor potential of the proposed method.

In future work, we will provide a human observer study, reader detection test, comparison with C-View image and further analysis of the low-frequency SDM image to provide more insight into the proposed method's pros and cons. Cross-vendor data will also be collected to evaluate the potential of our method on different vendor systems.

There are several possible directions for improving the performance of the proposed method. In this study, we used two state-of-the-art network architectures in image generation tasks for low-frequency SDM image generation and high-frequency SDM image generation. There might be other network architectures that have a higher performance in image-to-image regression or texture generation. Using these networks in the proposed method might further improve the quality of SDM images. In addition, the vertical projection image acquired in DBT acquisition has the same geometry as the FFDM. Replacing the generated low-frequency SDM image with the vertical projection image might reduce the error introduced by the low-frequency network in the proposed MSCN and obtain a better image quality. Similar to progressive reconstruction in the Gaussian pyramid, progressively generating SDM images from low frequency to high frequency might make the training more stable and derive an SDM image with lower intensity distortion.

The proposed method might be beneficial for other image-synthesis tasks in medical imaging. For example, in the low-dose CT denoising task, low-dose CT, and full-dose CT have different noise patterns and texture patterns because of the different radiation doses used in the acquisition, although they might have equal intensity (tissue) distribution. In the MRI-to-CT translation task, the ground-truth CT images derived by image registration might introduce regression errors owing to the mismatching of subtle details. In these tasks, the proposed MSCN might decrease the regression error introduced by the texture pattern difference or image registration and make the training more stable.

5 | CONCLUSIONS

In this study, we proposed an MSCN for SDM generation. The experiments showed that the proposed method could decrease intensity distortion, increase perceptual similarity, and improve mass quality, resulting in SDM images with higher image quality. In future work, we will conduct a human observer study and further analysis to provide more insight into the proposed method's pros and cons.

ACKNOWLEDGMENTS

This work was supported in part by the China Department of Science and Technology under Key Grant 210YBXM 2020109002. The work of Yao Lu was supported in part by the NSFC under grants 12126610, 81971691, 81801809, 81830052, and 81827802; in part by the Construction Project of Shanghai Key Laboratory of Molecular Imaging under grant 18DZ2260400; and in part by the Guangdong Province Key Laboratory of Computational Science at the Sun Yat-sen University under grant 2020B1212060032. The work of Yuesheng Xu was supported in part by the U.S. National Science Foundation under grant DMS-1912958 and in part by theNational Cancer Institute of the National Institutes of Health under Award R21CA263876.

REFERENCES

1. Hodgson R, Heywang-Köbrunner SH, Harvey SC, et al. Systematic review of 3D mammography for breast cancer screening. *Breast*. 2016;27:52-61. https://doi.org/10.1016/j.breast.2016.01.002
2. Svahn TM, Houssami N, Sechopoulos I, Mattsson S. Review of radiation dose estimates in digital breast tomosynthesis relative to those in two-view full-field digital mammography. *Breast*. 2015;24(2):93-99. https://doi.org/10.1016/j.breast.2014.12.002
3. van Schie G, Wallis MG, Leifland K, Danielsson M, Karssemeijer N. Mass detection in reconstructed digital breast tomosynthesis volumes with a computer-aided detection system trained on 2D mammograms. *Med Phys*. 2013;40(4):041902. https://doi.org/10.1118/1.4791643
4. Kim ST, Kim DH, Ro YM. Generation of conspicuity-improved synthetic image from digital breast tomosynthesis. In: *2014 19th*

International Conference on Digital Signal Processing. IEEE; 2014:395-399. https://doi.org/10.1109/ICDSP.2014.6900693
5. Homann H, Bergner F, Erhard K. Computation of synthetic mammograms with an edge-weighting algorithm. In: Medical Imaging 2015: Physics of Medical Imaging. Vol 9412. SPIE; 2015. https://doi.org/10.1117/12.2081797
6. Wei J, Chan H-P, Helvie MA, et al. Synthesizing mammogram from digital breast tomosynthesis. *Phys Med Biol*. 2019;64(4):045011. https://doi.org/10.1088/1361-6560/aafcda
7. Ruth C, Smith A, Stein J, inventors; Hologic Inc, assignee System and method for generating a 2D image from a tomosynthesis data set. US patent 7760924B2. July 20, 2010.
8. Nelson JS, Wells JR, Baker JA, Samei E. How does c-view image quality compare with conventional 2D FFDM? *Med Phys*. 2016;43(5):2538-2547. https://doi.org/10.1118/1.4947293
9. Barca P, Lamastra R, Aringhieri G, Tucciariello RM, Traino A, Fantacci ME. Comprehensive assessment of image quality in synthetic and digital mammography: a quantitative comparison. *Australas Phys Eng Sci Med*. 2019;42(4):1141-1152. https://doi.org/10.1007/s13246-019-00816-8
10. Ronneberger O, Fischer P, Brox T. U-Net: convolutional networks for biomedical image segmentation. In: Navab N, Hornegger J, Wells WM, Frangi AF, eds. *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*. Springer International Publishing; 2015:234-241.
11. Isola P, Zhu J-Y, Zhou T, Efros AA. Image-to-image translation with conditional adversarial networks. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE; 2017.
12. Wang X, Yu K, Wu S, et al. ESRGAN: enhanced super-resolution generative adversarial networks. In: *Proceedings of the European Conference on Computer Vision (ECCV) Workshops*. Vol 11133. Springer; 2019
13. Jiang G, Lu Y, Wei J, Xu Y. Synthesize mammogram from digital breast tomosynthesis with gradient guided cGANs. In: Shen D, Liu T, Peters TM, et al., eds. *Medical Image Computing and Computer Assisted Intervention – MICCAI 2019*. Lecture Notes in Computer Science. Springer International Publishing; 2019:801-809. https://doi.org/10.1007/978-3-030-32226-7_89
14. Jiang G, Wei J, Xu Y, et al. Synthesis of mammogram from digital breast tomosynthesis using deep convolutional neural network with gradient guided cGANs. *IEEE Trans Med Imaging*. 2021;40(8):1-1. https://doi.org/10.1109/TMI.2021.3071544
15. Blau Y, Michaeli T. The Perception-Distortion Tradeoff. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*; 2018.
16. Johnson J, Alahi A, Fei-Fei L. Perceptual losses for real-time style transfer and super-resolution. In: Leibe B, Matas J, Sebe N, Welling M, eds. *Computer Vision – ECCV 2016*. Springer International Publishing; 2016:694-711.
17. Yang Q, Yan P, Zhang Y, et al. Low-dose CT image denoising using a generative adversarial network with Wasserstein distance and perceptual loss. *IEEE Trans Med Imaging*. 2018;37(6):1348-1357. https://doi.org/10.1109/TMI.2018.2827462
18. Mirza M, Osindero S. Conditional generative adversarial nets. arXiv:1411.1784, 2014.
19. Smith AP, Chen B, Jing Z, inventors; Hologic Inc, assignee. Mammography/tomosynthesis systems and methods automatically deriving breast characteristics from breast x-ray images and automatically adjusting image processing parameters accordingly. US patent 8170320B2. May 1, 2012.
20. Zhang R, Isola P, Efros AA, Shechtman E, Wang O. The unreasonable effectiveness of deep features as a perceptual metric. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. IEEE; 2018:586-595.
21. Zhang K, Zuo W, Chen Y, Meng D, Zhang L. Beyond a Gaussian denoiser: residual learning of deep CNN for image denoising. *IEEE Trans Image Process*. 2017;26(7):3142-3155. https://doi.org/10.1109/TIP.2017.2662206
22. Ba JL, Kiros JR, Hinton GE. Layer normalization. arXiv:1607.06450, 2016.
23. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. arXiv:1409.1556, 2015.
24. Deng J, Dong W, Socher R, Li L-J, Li K, Fei-Fei L. ImageNet: a large-scale hierarchical image database. In: *2009 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE; 2009:248-255. https://doi.org/10.1109/CVPR.2009.5206848
25. Kingma DP, Ba J. Adam: A method for stochastic optimization. arXiv:1412.6980, 2017.
26. Li H, Jiang G, Zhang J, et al. Fully convolutional network ensembles for white matter hyperintensities segmentation in MR images. *NeuroImage*. 2018;183:650-665. https://doi.org/10.1016/j.neuroimage.2018.07.005
27. Skaane P, Bandos AI, Eben EB, et al. Two-view digital breast tomosynthesis screening with synthetically reconstructed projection images: comparison with digital breast tomosynthesis with full-field digital mammographic images. *Radiology*. 2014;271(3):655-663.
28. Gilbert FJ, Tucker L, Gillan MG, et al. Accuracy of digital breast tomosynthesis for depicting breast cancer subgroups in a UK retrospective reading study (TOMMY Trial). *Radiology*. 2015;277(3):697-706.
29. Ratanaprasatporn L, Chikarmane SA, Giess CS. Strengths and weaknesses of synthetic mammography in screening. *RadioGraphics*. 2017;37(7):1913-1927.
30. Nelson JS, Wells JR, Baker JA, Samei E. How does c-view image quality compare with conventional 2D FFDM? *Med Phys*. 2016;43(5):2538-2547.
31. Barca P, Lamastra R, Aringhieri G, Tucciariello RM, Traino A, Fantacci ME. Comprehensive assessment of image quality in synthetic and digital mammography: a quantitative comparison. *Australas Phys Eng Sci Med*. 2019;42(4):1141-1152.
32. Aujero MP, Gavenonis SC, Benjamin R, Zhang Z, Holt JS. Clinical performance of synthesized two-dimensional mammography combined with tomosynthesis in a large screening population. *Radiology*. 2017;283(1):70-76.
33. Gastounioti A, McCarthy AM, Pantalone L, Synnestvedt M, Kontos D, Conant EF. Effect of mammographic screening modality on breast density assessment: digital mammography versus digital breast tomosynthesis. *Radiology*. 2019;291(2):320-327.
34. Destounis SV, Santacroce A, Arieno A. Update on breast density, risk estimation, and supplemental screening. *Am J Roentgenol*. 2020;214(2):296-305.
35. Mackenzie A, Thomson EL, Mitchell M, et al. Virtual clinical trial to compare cancer detection using combinations of 2D mammography, digital breast tomosynthesis and synthetic 2D imaging. *Eur Radiol*. 2022;32(2):806-814.
36. Khanani S, Xiao L, Jensen MR, et al. Comparison of breast density assessments between synthesized C-ViewTM & intelligent 2DTM mammography. *Br J Radiol*. 2022;95:20211259.
37. Horvat JV, Keating DM, Rodrigues-Duarte H, Morris EA, Mango VL. Calcifications at digital breast tomosynthesis: imaging features and biopsy techniques. *Radiographics*. 2019;39(2):307-318.
38. Zeng B, Yu K, Gao L, Zeng X, Zhou Q. Breast cancer screening using synthesized two-dimensional mammography: a systematic review and meta-analysis. *Breast*. 2021;59:270-278.
39. Kang HJ, Chang JM, Lee J, et al. Replacing single-view mediolateral oblique (MLO) digital mammography (DM) with synthesized mammography (SM) with digital breast tomosynthesis (DBT) images: comparison of the diagnostic performance and radiation dose with two-view DM with or without MLO-DBT. *Eur J Radiol*. 2016;85(11):2042-2048.

40. You C, Zhang Y, Gu Y, et al. Comparison of the diagnostic performance of synthesized two-dimensional mammography and full-field digital mammography alone or in combination with digital breast tomosynthesis. *Breast Cancer*. 2020;27(1):47-53.
41. Abdullah P, Alabousi M, Ramadan S, et al. Synthetic 2D mammography versus standard 2D digital mammography: a diagnostic test accuracy systematic review and meta-analysis. *AJR Am J Roentgenol*. 2021;217(2):314-325.

How to cite this article: Jiang G, He Z, Zhou Y, et al. Multi-scale cascaded networks for synthesis of mammogram to decrease intensity distortion and increase model-based perceptual similarity. *Med. Phys*. 2023;50:837–853. https://doi.org/10.1002/mp.16007