Advances in Image Reconstruction for Digital Breast Tomosynthesis

by

Mingjie Gao

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Electrical and Computer Engineering)
in the University of Michigan
2024

Doctoral Committee:

Professor Heang-Ping Chan, Co-Chair
Professor Jeffrey A. Fessler, Co-Chair
Professor Mitchell Goodsitt
Associate Professor Zhongming Liu

Mingjie Gao

gmingjie@umich.edu

ORCID iD: 0000-0002-9129-0060

To my grandfather, in loving memory

Acknowledgements

This thesis would not have been possible without the guidance of my advisors, Prof. Heang-Ping Chan and Prof. Jeffrey Fessler. Dr. Chan has been a true role model, exemplifying scientific rigor and self-discipline in her daily work. Her availability, patience, and attention to every detail have been instrumental in shaping my ideas and refining my work. I began working with Jeff when I was an undergraduate student, and I have been constantly amazed by his insights and organized thinking. He has not only been a mentor but also a source of valuable emotional support, especially during the challenging times of the Covid-19 pandemic. I feel fortunate to have had the opportunity to learn from both Dr. Chan and Jeff. Their diverse expertise and research interests have added stimulating dynamics and interactions to our collaborations. Their guidance and mentorship will serve as lifelong assets in my pursuit of a career as an imaging scientist.

I would also like to extend my sincere thanks to the members of my thesis committee, Prof. Mitchell Goodsitt and Prof. Zhongming Liu, for their insightful comments and constructive feedback that enhanced the quality of this thesis.

I am grateful to the members of the CAD-AI Lab. I want to thank Prof. Lubomir Hadjiyski for his technical support and advice about life, Prof. Chuan Zhou for sharing his stories and humor, Prof. Ravi Samala for his encouragement when I started my PhD, Jiabei Zheng for sharing his code and documentations, Yifan Wang and Di Sun for their camaraderie. I am also thankful to the members of Jeff's lab for discussing research with me and organizing fun group activities. Special thanks to Gopal Nataraj and Steven Whitaker who mentored my undergraduate research and

motivated me to pursue a PhD. I would also like to thank my friends outside the labs in Ann Arbor: Jiaren Zou, Lise Wei, Lili Chen, Dehao Zhang, and others.

I acknowledge the financial support from National Institutes of Health with grant number R01 CA214981 for my PhD study.

Finally, I am deeply indebted to my parents in China, Xiong Gao and Yinghong Xu. Their unconditional support, endless encouragement, and constant love mean the world to me. I would also like to express my heartfelt gratitude to my wife, Yuhua Zhang. Meeting her has been the greatest blessing on this journey.

Table of Contents

Dedication ..... **ii**

Acknowledgements ..... **iii**

List of Figures ..... **ix**

List of Tables ..... **xiii**

List of Abbreviations ..... **xiv**

Abstract ..... **xvii**

Chapter 1 Introduction **1**

Chapter 2 Background **6**

2.1 Digital Breast Tomosynthesis 6
2.1.1 DBT Imaging System 6
2.1.2 Masses and Microcalcifications 8
2.2 Physics of X-ray Imaging 10
2.2.1 X-ray Attenuation and Beer's Law 10
2.2.2 Image Reconstruction 11
2.3 Tomographic Reconstruction Methods 13
2.3.1 Analytical Methods 13
2.3.2 Algebraic Methods 13
2.3.3 Statistical Methods 14
2.4 Deep Learning 16
2.4.1 Neural Networks 16
2.4.2 Backpropagation 17
2.4.3 Convolutional Neural Networks 18

Chapter 3 Deep Convolutional Neural Network with Adversarial Training for Denoising Digital Breast Tomosynthesis Images **20**

3.1 Introduction 20
3.2 Methods and Materials 22
3.2.1 DCNN Training 22
3.2.1.1 DNGAN Framework 22

3.2.1.2 Network Structures 23
3.2.1.3 Training Loss Function 25
3.2.1.4 Fine-tuning with MC Patches 27
3.2.2 Data Sets 27
3.2.2.1 Digital Phantom Data 28
3.2.2.2 Physical Phantom Data 30
3.2.2.3 Training Set Generation 31
3.2.2.4 Human Subject DBTs 33
3.2.3 Figures of Merit 34
3.2.3.1 Noise Power Spectrum 34
3.2.3.2 Contrast-to-noise Ratio and Full Width at Half Maximum 34
3.2.3.3 Detectability Index 36
3.2.4 DCNN Training Setup 37
3.2.5 Comparison Method 38
3.3 Results 38
3.3.1 Effect of Tuning Parameter λ_{adv} 38
3.3.2 Effects of Training Batch Size, Learning Rate, and Number of Epochs 40
3.3.3 Effect of Dose Level of Targets on DNGAN Training 42
3.3.4 Effect of MC Fine-tuning and Layer Freezing 45
3.3.5 Effect of Training Sample Sizes 47
3.3.6 Denoising Performance on Validation Physical Phantom 48
3.3.7 Denoising Performance on Human Subject DBTs 50
3.3.8 Denoising Pristina-reconstructed Images Using SART Denoiser 51
3.4 Discussion 53
3.5 Conclusion 56

Chapter 4 Model-based Deep CNN-Regularized Reconstruction for Digital Breast Tomosynthesis with a Task-based CNN Image Assessment Approach 58

4.1 Introduction 58
4.2 Methods and Materials 61
4.2.1 DBT Reconstruction 61
4.2.1.1 Background of MBIR and DBCN Reconstruction 61
4.2.1.2 DNGAN Denoising 62
4.2.1.3 Model-based DCNN-regularized Reconstruction (MDR) 63
4.2.2 Task-based Image Quality Measures 64
4.2.2.1 CNN MC Classifier (CNN-MC) 64
4.2.2.2 CNN Noise Estimator (CNN-NE) 66
4.2.2.3 Transfer Learning 67
4.2.3 Data Sets 67

4.2.3.1 Virtual Phantom DBTs 69
4.2.3.2 DNGAN Training Set 70
4.2.3.3 CNN-NE Training Set 71
4.2.3.4 Human Subject DBTs 71
4.2.3.5 CNN-MC Training Set and CNN-NE/CNN-MC Validation and Test Set 72
4.2.3.6 Physical Phantom DBTs 73
4.2.4 Implementation 74
4.2.5 Image Conditions 75
4.3 Results 75
4.3.1 Effectiveness of DNGAN, CNN-NE and CNN-MC 75
4.3.2 Parameter Selection for MDR Regularization 78
4.3.3 MDR Ablation Study 79
4.3.4 Comparisons on Validation Set and Independent Test Set 81
4.3.5 Qualitative Comparison Between MDR and Commercial Reconstruction 85
4.4 Discussion 87
4.5 Conclusion 92

Chapter 5 X-ray Source Motion Blur Modeling and Deblurring with Generative Diffusion for Digital Breast Tomosynthesis 93

5.1 Introduction 93
5.2 Methods 95
5.2.1 Source Blur Modeling 95
5.2.1.1 DBT Imaging System 95
5.2.1.2 Constructing Source Blur Matrix $\boldsymbol{B}$ 97
5.2.1.3 Challenges in Using $\boldsymbol{AB}$ for DBT Reconstruction 99
5.2.2 Nonblind Image Deblurring 100
5.2.2.1 Denoising Diffusion Probabilistic Models (DDPM) 101
5.2.2.2 Image Deblurring with Generative Diffusion 102
5.2.3 X-ray Source Motion Deblurring for Reconstructed DBTs 104
5.3 Materials 104
5.3.1 DBT System Configuration 104
5.3.2 Source Blur Modeling 105
5.3.2.1 Verification Study of Blur Kernel Modeling 105
5.3.2.2 Investigating Using $\boldsymbol{AB}$ or $\boldsymbol{A}_{\text{sub}}$ for DBT Reconstruction 106
5.3.3 Data Sets 107
5.3.3.1 VICTRE Phantoms 107
5.3.3.2 Line Pair Test Object and Image Quality Metrics 108
5.3.4 DDPM Implementation 110
5.3.5 Comparison Methods 111

5.4 Results 112
5.4.1 Source Blur Modeling 112
5.4.1.1 Verification Study of Blur Kernel Modeling 112
5.4.1.2 Limitations of Using $\boldsymbol{AB}$ or $\boldsymbol{A}_{\text{sub}}$ for DBT Reconstruction 113
5.4.2 DDPM Unconditional Image Generation 114
5.4.3 Image Deblurring with Generative Diffusion 115
5.4.3.1 Parameter Selection of DDPM Posterior Sampling 115
5.4.3.2 Effect of Breast Densities on Deblurring 116
5.4.3.3 Effect of Test Object Heights above Detector 117
5.4.3.4 Comparison of Deblurring Methods 118
5.5 Discussion 120
5.6 Conclusion 122

Chapter 6 Future Work 123

6.1 Computerized Detection for Image Quality Evaluation 123
6.2 DCNN Training with Task-based Losses 124
6.3 Potential Dose Reduction for DBT Scans 124
6.4 Limited-angle Reconstruction with Generative Models 125

Bibliography 127

List of Figures

1.1 (a) 2D mammogram. Cancer is not visible even after it is known to be in the marked region due to overlapping fibroglandular tissue. (b) A reconstructed slice from DBT of the same breast. Invasive cancer is visible as spiculated mass. Image credit: [7]. 2

2.1 The geometry and workflow of the GE Senographe Pristina DBT system. 7

2.2 (a) An example 5 mm spiculated mass (invasive ductal carcinoma). (b) An example MC cluster (ductal carcinoma in situ) marked by the white box. Its zoomed-in view is shown in (c). 9

3.1 The framework of the denoising DCNN with adversarial training (DNGAN). 23

3.2 The network structure of (a) the denoiser and (b) the discriminator. The numbers on the left of each layer (rectangle) represent the width and height of the tensors. The numbers on the top of the rectangles represent the number convolution filters for the tensors (blue) or the length of the vectors (black). 24

3.3 (a) The x-ray fluence spectrum of a 34 kVp Rh anode after 0.03 mm Ag filtration in the CatSim simulation. (b) The system MTF for the central PV. The edge for MTF calculation was placed on the breast support plate and was parallel to the chest wall with 3° tilt. 30

3.4 Examples of (a) DBT patches in the 24mAs/120mAs training set and (b) MC patches in the accompanying MC fine-tuning set. In each pair, the LD patch is shown on the left and the HD patch on the right. The images are from the digital simulated phantoms..... 32

3.5 An illustration of CNR and FWHM calculation for an MC patch..... 35

3.6 (a) The training MSE losses and (b) the WD estimates for different λ_{adv} versus training epochs. (c) The NPS curves of the denoised validation digital phantom volumes for different λ_{adv}. 39

3.7 The CNR, FWHM, fit success rates, and *d'* of the MCs in the validation physical phantom for different λ_{adv}. The error bars represent one standard deviation. 40

3.8 The training losses versus epochs for illustrating the effects of (a) the batch size, (b) the initial learning rate, and (c) the number of epochs. 42

3.9 The NPS curves of the validation digital phantom volumes for comparing the dose levels of the training targets. The solid lines are calculated from the volumes by deploying the denoisers to the 24mAs volume. The dashed lines are from the CatSim simulated volumes with the corresponding dose levels. 43

3.10 The CNR, FWHM, fit success rates, and *d'* of the MCs in the validation physical phantom for the different dose levels of the targets used in the DNGAN training. 43

3.11 An example 18 mm × 18 mm region in the validation physical phantom for the different dose levels of the targets used in the DNGAN training. The images are displayed with the same window/level settings. The HD scan of the validation physical phantom (34 kVp, 125 mAs) is also shown for reference. 44

3.12 An example 20 mm × 15 mm region in the validation physical phantom showing the effect of fine-tuning and layer freezing. The region contains a background MC-free area on the left and a 0.180-0.212 mm MC cluster on the right. The images are displayed with the same window/level settings. 46

3.13 The CNR, FWHM, fit success rates, and *d'* of the MCs in the validation physical phantom showing the effect of fine-tuning and layer freezing. 46

3.14 The box plots of the CNRs for different training sample sizes. Each box contains 10 data points. In the box plot, the red bar represents the median; the length of the box equals the interquartile range; the whiskers extend to the minimum and maximum data points. The boxes are slightly shifted horizontally to avoid overlap. 47

3.15 Example MC clusters in the validation physical phantom for comparing the denoising results. Top row: 0.150-0.180 mm cluster. Bottom row: 0.180-0.212 mm cluster. All images show a 15 mm × 15 mm region. The images in the same row are displayed with the same window/level settings. 49

3.16 The CNR, FWHM, fit success rates, and *d'* of the MCs in the validation physical phantom for comparing the denoising results. 49

3.17 Example images of human subject DBTs with a spiculated mass (invasive ductal carcinoma) and an MC cluster (ductal carcinoma in situ). All images show an 18 mm × 18 mm region. The images in the same row are displayed with the same window/level settings. 50

3.18 The CNR scatter plot of MCs in the human subject DBTs for the DNGAN denoised images and the DBCN reconstructed images versus the LD SART images. 51

3.19 Example MC clusters in the Pristina-reconstructed images of the validation physical phantom for comparing the mismatched and matched denoisers. Top row: 0.150-0.180 mm cluster. Bottom row: 0.180-0.212 mm cluster. All images show a 15 mm × 15 mm region. The images in the same row are displayed with the same window/level settings. 52

3.20 The CNR, FWHM, fit success rates, and *d'* of the MCs in the Pristina-reconstructed images of the validation physical phantom for comparing the mismatched and matched denoisers. 52

3.21 Scatter plot of *d'* versus CNR, including all data points from Figures 3.7, 3.10, 3.13, 3.16, and 3.20 for three nominal MC sizes and different conditions. 55

4.1 Example 3-slice MIP patches with MCs (top two rows) and without MCs (bottom two rows) from the CNN-MC training set of human subject DBTs. The patch size is 128×128 pixels (12.8 mm × 12.8 mm). 72

4.2 The scatter plot of the CNN-NE estimated RMS noise vs the analytically calculated RMS noise for the individual patches in the CNN-NE validation set (human subject DBT images). The dotted line is the diagonal line. The RMS noise is plotted in terms of pixel values. 77

4.3 Example images (18 mm × 18 mm) with MC clusters from the physical phantom DBTs acquired and reconstructed by a GE Pristina DBT system. The dnSTD images were obtained by DNGAN denoising on the STD images. The STD+ mode used 54% more dose than the STD mode. 77

4.4 AUCs of CNN-MC for the classification of image patches with and without MC clusters in the physical phantom DBTs of 3 image conditions. The data points and error bars were obtained from the mean and standard deviation of the 5 repeated models trained with different random initialization. The points are slightly shifted horizontally to avoid overlap..... 78

4.5 The task-based image quality assessment plot showing the tradeoff between MC cluster detectability (CNN-MC AUC) versus image noise (CNN-NE RMS noise) on the validation set reconstructed with different regularization parameters β_{EP} and β_{RED} in MDR. The data point obtained with a fixed β_{EP} are linked by a solid line and data points for a fixed β_{RED} are shown with the same symbol. 79

4.6 The task-based image quality assessment plot on the validation set showing the tradeoffs of the different terms in the MDR ablation study. 80

4.7 Example 3-slice MIP patches (12.8 mm × 12.8 mm) from the human subject validation set for the MDR ablation study. In each pair, a positive patch of MC cluster is shown on the left and a negative MC-free patch on the right. 81

4.8 The task-based image quality assessment plot on the human subject validation set (first row) and test set (second row) using the VGG-Net (left column), ResNet (middle column) and ConvNeXt (right column) for comparing the reconstruction and denoising approaches. The labels (a) to (h) are defined in Section 4.3.4 and Table 4.3. 83

4.9 Example 3-slice MIP patches (12.8 mm × 12.8 mm) of two MC clusters (example 1: ductal carcinoma in situ; example 2: invasive ductal carcinoma) from the human subject test set. The labels (a) to (h) are defined in Section 4.3.4 and Table 4.3. 84

4.10 Example images (20 mm × 20 mm) of a spiculated mass (invasive ductal carcinoma) of a human subject DBT. The labels (a) to (h) are defined in Section 4.3.4 and Table 4.3.. 85

4.11 Qualitative comparison between the commercial GE Pristina reconstruction and the proposed MDR. Top row: full DBT slices (84.4 mm × 170 mm). Bottom row: zoomed-in views of the marked regions (20 mm × 15 mm) containing an MC cluster (ductal carcinoma in situ)..... 86

4.12 The scatter plot of the analytically calculated RMS noise for the CNN-NE training image patches with the exposure factor of 1 using either quadratic fitting or noise-free images for background removal. The RMS noise is plotted in terms of pixel values. 89

5.1 Diagram of DBT imaging system. The finite-sized rectangular x-ray source is exaggerated to show details. 96

5.2 The simplified 1D rectangle source and the derivation of in-plane PSF. (a) In the forward scan process, given a blurry source w_0 and an impulse at height z_s above the detector, its PSF at the detector is w_d. (b) In the reconstruction process, the in-plane PSF at the impulse location is w_s. 99

5.3 The simulation study to verify the in-plane blur kernel modeling. In this example, the impulse is centered in y direction, so its PSF and OTF are both real and symmetric. In general, the system OTF is complex-valued if the impulse PSF is asymmetric, but the source blur OTF is always real-valued. 106

5.4 (a) LP test object with 5 LP/mm, 3.33 LP/mm, 2.5 LP/mm, and 2 LP/mm (top to bottom) and vertical shifts (left to right). (b) Example reconstructed DBT slice with embedded LP test object. (c) Illustration of LP contrast calculation (gray area: bar regions; yellow area: space regions). 110

5.5 (a) Example impulse OTFs and source blur OTFs for three distances (z_s) between the impulse and the detector in the simulation study of blur kernel modeling. (b) The comparison of sinc-fit estimated and analytically calculated blur kernel widths. The shaded region shows the range of w_s values, the rectangle widths from individual scan angles. 113

5.6 The impulse OTFs that demonstrate the limitations of using AB or A_{sub} to model source motion blur for DBT reconstruction. The conditions (1) to (4) are defined in Section 5.3.2.2. 114

5.7 (a) DBT slice from the virtual phantoms used in the DDPM training set. (b) DBT slice generated by unconditional DDPM sampling. Image sizes are 300×400 pixels (25.5 mm × 34 mm). (c) Structural NPS of the two example images. 115

5.8 Contrast-vs-noise plots of the LP test object showing the dependence on the parameters of the proposed deblurring method with generative diffusion. 116

5.9 Contrast-vs-noise plots of the LP test object demonstrating the effect of breast density on the performance of the proposed deblurring method with generative diffusion. 117

5.10 Contrast-vs-frequency plots of the LP test object demonstrating the effect of the object height above detector z_s on the performance of the proposed deblurring method with generative diffusion. 118

5.11 Contrast-vs-noise plots of the LP test object for comparing the deblurring performance of different methods on the DBCN-reconstructed images. 119

5.12 Example ROIs of the LP test object in the test phantom for comparing different deblurring methods on the DBCN-reconstructed images. From top to bottom, the LPs have spatial frequencies of 5 LP/mm, 3.33 LP/mm, 2.5 LP/mm, and 2 LP/mm. The ROI sizes are 60×60 pixels (5.1 mm × 5.1 mm). 119

List of Tables

3.1 Summary of data sets for DNGAN investigation. 28

4.1 Network structures of CNN-MC and CNN-NE. 66

4.2 Summary of data sets for DNGAN, CNN-NE and CNN-MC in MDR study. 69

4.3 Summary of rankings for different reconstruction and denoising approaches using DCNN task-based image quality assessment in Figure 4.8. 83

5.1 Density and size characteristics of the virtual breast phantoms for DDPM training. 108

List of Abbreviations

AEC	Automatic exposure control
AUC	Area under the receiver operating characteristic curve
CAD	Computer-aided detection
CatSim	Computer assisted tomography simulator (GE Global Research)
CHO	Channelized Hotelling observer
CNN	Convolutional neural network
CNN-MC	CNN MC classifier
CNN-NE	CNN noise estimator
CNR	Contrast-to-noise ratio
CT	Computed tomography
d'	Detectability index
DBCN	Detector blur and correlated noise modeling
DBT	Digital breast tomosynthesis
DCNN	Deep convolutional neural network
DDPM	Denoising diffusion probabilistic models
DFT	Discrete Fourier transform
DNGAN	Denoising generative adversarial network
EP	Edge-preserving regularizer
FBP	Filtered backprojection

FDA	U.S. Food and Drug Administration
FP	False positives
FWHM	Full width at half maximum
GAN	Generative adversarial network
GE	General Electric Company
GVF	Glandular volume fraction
HD	High dose
IRB	Institutional Review Boards
LD	Low dose
LP	Line pair
MAP	Maximizing a posteriori
MBIR	Model-based iterative reconstruction
MC	Microcalcification
MC-GPU	GPU-accelerated Monte Carlo x-ray imaging simulator (U.S. FDA)
MDR	Model-based DCNN-regularized reconstruction
MGD	Mean glandular dose
MIP	Maximum intensity projection
ML	Maximum likelihood
MRI	Magnetic resonance imaging
MSBF	Multiscale bilateral filtering
MSE	Mean squared error
MTF	Modulation transfer function
NPS	Noise power spectrum

NPWE	Nonprewhitening model observer with eye filter
OTF	Optical transfer function
PSF	Point spread function
PV	Projection view
RED	regularization by denoising
ReLU	Rectified linear unit
RMS	Root-mean-square
ROC	Receiver operating characteristic
ROI	Region of interest
SART	Simultaneous algebraic reconstruction technique
SG	Segmented separable footprint projector
SKS	Signal-known-statistically task
SSIM	Structural similarity
STD	Standard dose exposure mode
TV	Total variation
VICTRE	Virtual imaging clinical trial for regulatory evaluation (U.S. FDA)
WD	Wasserstein distance

Abstract

Digital breast tomosynthesis (DBT) is an important imaging modality for breast cancer screening and diagnosis. It acquires a sequence of projection views within a limited angle and provides quasi-three-dimensional images of the breasts, allowing for improved lesion visualization and reduced false positives compared with two-dimensional mammography. Despite its advantages, DBT suffers from noise and blur problems that can compromise image quality and reduce its sensitivity in detecting subtle signs of breast cancer such as microcalcifications (MCs). The primary objective of this thesis is to push the state-of-the-art of DBT imaging by developing advanced DBT image reconstruction and processing methods. By reducing image noise, enhancing spatial resolution, optimizing reconstruction methods and evaluating them based on clinical tasks, our ultimate goal is to make DBT an even more effective tool for breast cancer screening and diagnosis.

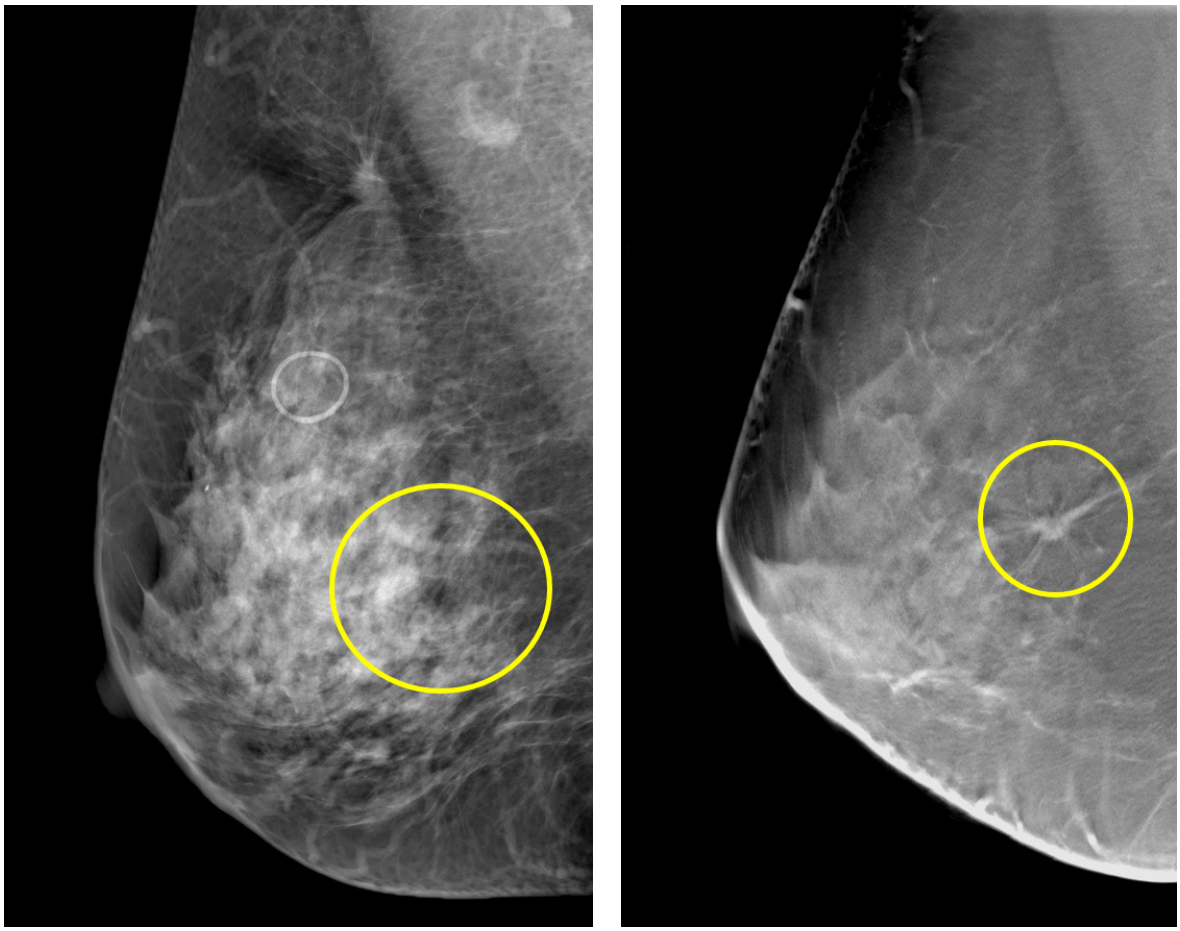
In this thesis, we first developed a deep convolutional neural network (DCNN) for denoising reconstructed DBT images. We trained the DCNN using a weighted combination of mean squared error loss and the adversarial loss based on generative adversarial network (GAN), and therefore called it DNGAN. The DNGAN improved the contrast-to-noise ratio, detectability index, and human observer detection sensitivity of the MCs in DBT images of breast simulating phantoms. Promising denoising results were also observed on a small test set of human subject DBTs. Then, we introduced a model-based DCNN-regularized reconstruction (MDR) method for DBT. It combined a model-based iterative reconstruction method with the DNGAN denoiser. To

facilitate task-based image quality assessment, we also proposed two DCNN tools: CNN-NE for noise estimation, and CNN-MC as a model observer for MC cluster detectability measure. We demonstrated the effectiveness of CNN-NE and CNN-MC using phantom DBTs. The MDR method achieved low noise and the highest detection rankings on a test set of human subject DBTs. Finally, we presented our work on modeling the x-ray source motion blur of the DBT imaging system. We derived an analytical in-plane source blur kernel for DBT images based on imaging geometry and showed that it could be approximated by a shift-invariant kernel over the DBT slice at a given height above the detector. We proposed a post-processing image deblurring method with a generative diffusion model as an image prior and successfully enhanced spatial resolution of the reconstructed DBT images.

Chapter 1
Introduction

Breast cancer is a significant global public health concern. In 2020, female breast cancer surpassed lung cancer and became the most commonly diagnosed cancer in the world [1]. In the U.S. in 2023, approximately 297,790 new cases of female breast cancer are expected to be diagnosed, and 43,170 deaths will occur [2]. Detection of breast cancer at an early stage is crucial for improving treatment outcomes and reducing mortality rates. Mammography has been the gold standard for breast cancer screening for several decades. However, conventional two-dimensional (2D) mammography has its limitations, including overlapping breast tissue and reduced sensitivity in dense breast tissue. These limitations have led to the development of more advanced breast imaging techniques, among which digital breast tomosynthesis (DBT) has emerged as a promising solution.

DBT is an innovative imaging modality that has gained popularity in recent years as it provides quasi-three-dimensional images of the breasts, allowing for improved lesion visualization (Figure 1.1) and reduced false positives. It works by acquiring a series of low-dose x-ray projection views (PVs) at different angles around the breast and then reconstructing these PVs into a stack of cross-sectional images. Clinical studies have shown that DBT offers increased cancer detection rates and decreased recall rates and has the potential to become a cornerstone of breast cancer screening [3]–[6]. While there are other breast imaging techniques being developed such as breast computed tomography (bCT), DBT remains more commonly available than bCT in clinical practice due to its lower radiation dose, less susceptibility to patient motion, and lower costs.



(a) (b)

Figure 1.1 (a) 2D mammogram. Cancer is not visible even after it is known to be in the marked region due to overlapping fibroglandular tissue. (b) A reconstructed slice from DBT of the same breast. Invasive cancer is visible as spiculated mass. Image credit: [7].

Despite its advantages, DBT suffers from noise and blur problems that can compromise image quality and reduce its sensitivity in detecting early-stage cancer. In breast imaging, minimizing radiation dose is paramount given the recommended annual screening for women of certain age groups. For this reason, the overall patient radiation dose of DBT is maintained at a level similar to that of a 2D mammogram. As a result, each PV has a much lower dose and higher quantum noise level compared to the mammogram. The use of multiple PVs also introduces additional detector readout noise. At larger PV scan angles, oblique x-ray incidence to the detector causes not only greater detector blur, but also longer x-ray beam paths through the compressed

breast, resulting in increased x-ray attenuation and higher quantum noise levels. Additionally, DBT systems using continuous-motion x-ray sources exhibit source motion blur. These noise and blur effects are propagated to the DBT volume through reconstruction, potentially degrading the visibility of subtle signs of breast cancer such as microcalcifications (MCs). These challenges motivate the need for continuous research and innovation in DBT imaging.

In recent years, deep learning has sparked a revolution in image processing and computer vision [8]. It is a subset of machine learning that focuses on the development of artificial neural networks that are capable of automatically learning hierarchical features from data. Among these neural network designs, deep convolutional neural network (deep CNN or DCNN) has been proven particularly successful in handling grid-like data, such as images. Deep learning has also emerged as a powerful tool in the field of medical imaging. Deep learning has been shown to improve image reconstruction in various 3D imaging modalities, enhancing image quality and reducing dose. However, deep learning has not been widely studied in DBT, probably due to the challenges in limited-angle reconstruction and special signal characteristics such as the subtle noise-like MCs and spiculated and ill-defined margins of masses. The possibility of using deep learning to reduce noise and enhance signals, potentially improving cancer detection and characterization and reducing radiation dose, is yet to be fully explored for DBT [9].

The primary objective of this thesis is to push the state-of-the-art of DBT imaging by addressing the aforementioned challenges. To do so, we investigated and developed advanced DBT image reconstruction and processing methods by leveraging deep learning techniques. By reducing image noise, enhancing spatial resolution, optimizing reconstruction methods and evaluating them based on clinical tasks, our ultimate goal is to make DBT an even more effective tool for breast cancer screening and diagnosis. The rest of this thesis is organized as follows.

- Chapter 2 provides the relevant background materials.
- Chapter 3 discusses the development of a DCNN framework for denoising reconstructed DBT images. We trained the DCNN using a weighted combination of mean squared error loss and the adversarial loss based on generative adversarial network (GAN), and therefore called it DNGAN. The DNGAN improved the contrast-to-noise ratio, detectability index, and human observer detection sensitivity of the MCs in DBT images of breast simulating phantoms. Promising denoising results were also observed on a small test set of human subject DBTs. This chapter is based on the published journal paper [10] that extends the conference paper [11] and abstract [12].
- Chapter 4 introduces a model-based DCNN-regularized reconstruction (MDR) method for DBT. It combined a model-based iterative reconstruction method with the DNGAN denoiser. To facilitate task-based image quality assessment, we also proposed two DCNN tools: CNN-NE for noise estimation, and CNN-MC as a model observer for MC cluster detectability measure. We demonstrated the effectiveness of CNN-NE and CNN-MC using phantom DBTs. The MDR method achieved low noise and the highest detection rankings on a test set of human subject DBTs. This chapter is based on the published journal paper [13] that extends the conference papers [14][15] and abstracts [16][17].
- Chapter 5 presents our work on modeling the x-ray source motion blur of the DBT imaging system. We derived an analytical in-plane source blur kernel for DBT images based on imaging geometry and showed that it could be approximated by a shift-invariant kernel over the DBT slice at a given height above the detector. We validated the accuracy of our blur kernel modeling by simulation. We proposed a post-processing image deblurring method with a

generative diffusion model as an image prior and successfully enhanced spatial resolution of the reconstructed DBT images. This chapter is based on a submitted journal manuscript.

- Chapter 6 suggests potential directions for future work.

Chapter 2
Background

2.1 Digital Breast Tomosynthesis

2.1.1 DBT Imaging System

DBT was approved by the U.S. Food and Drug Administration (FDA) for screening and diagnosis in 2011. As of 2023, FDA-approved DBT systems from four vendors are available clinically in the U.S.: GE, Hologic, Siemens, and Fuji. The specifications of the DBT system vary among vendors. We mainly use the GE Senographe Pristina DBT system in our work, so we focus our discussion on this system here, but the principle is applicable to other systems. Please refer to [18]–[21] for comparisons of the technical details and imaging performances of different DBT systems.

Figure 2.1 shows the geometry and workflow of the GE Senographe Pristina DBT system [22]. In a DBT scan, the patient stands in front of the scanner and the breast is placed on the support plate that is 2.3 cm above the digital detector plane. The breast is then compressed from the top by the compression paddle parallel to the detector. The thickness of the compressed breast is recorded and used for reconstruction. The gantry of the DBT system can be rotated to any angle to compress the breast at an angle to the vertical direction. In screening, usually two views are taken for each breast: the cranial-caudal view where the breast is compressed in vertical direction and the

mediolateral-oblique view where the gantry is rotated to about 45 degrees from the vertical direction to compress the breast approximately parallel to the pectoral muscle.

We use x-y-z coordinate for the breast object or the image volume, and t-s coordinate for the projection views (PVs). The origin of the coordinate system is located at the fulcrum or center of rotation of the x-ray source, denoted as O, and is 2 cm above the breast support plate. The GE Pristina system uses an x-ray source that moves in a step-and-shoot mode and a stationary detector. Other vendors may use a continuously moving source with pulsed x-rays; one vendor additionally tilts the detector in synchrony with the movement of the x-ray source. The Pristina DBT system takes 9 projections over an angular range of 25° ($\pm$12.5°). This gives 9 PVs with a size of 2394×2850 pixels at a pixel size of 0.1 mm × 0.1 mm. The Pristina system is also equipped with a Bucky grid with a grid density of 67 lines/cm and a grid ratio of 11:1 moving in the direction perpendicular to the chest wall [19]. However, the grid and scattered radiation are not considered in this thesis.

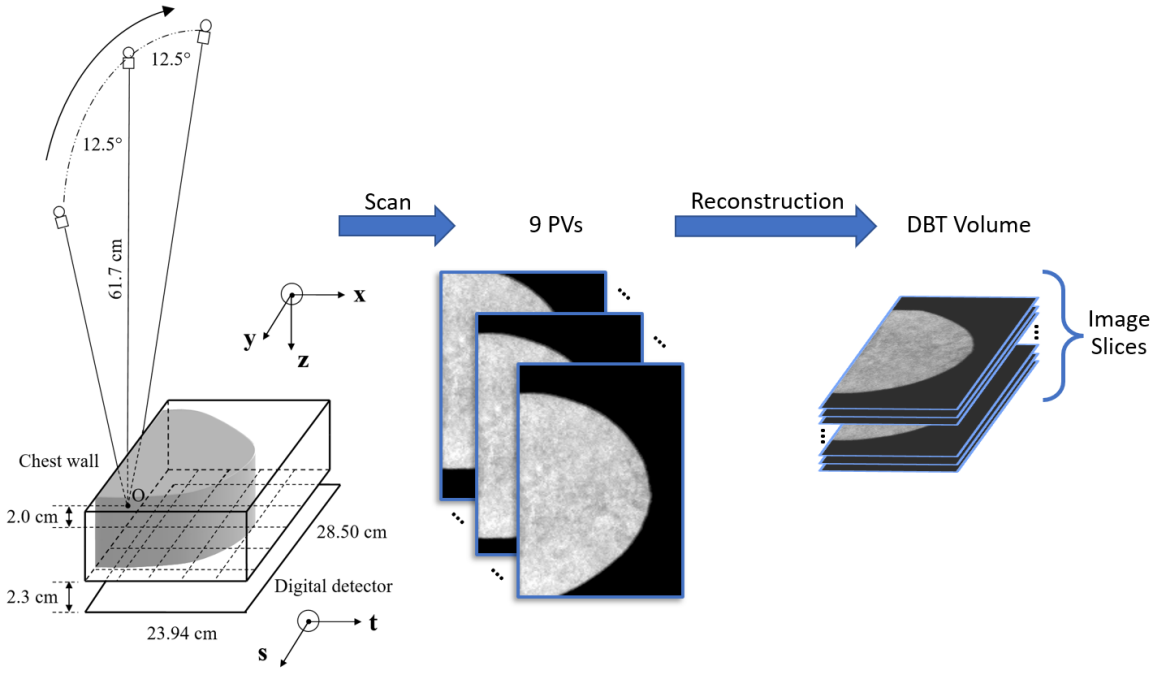


Figure 2.1 The geometry and workflow of the GE Senographe Pristina DBT system.

Reconstruction from the limited-angle acquisition produces tomographic image volumes with anisotropic voxel sizes such that the resolution is superior in the x-y plane but is very limited in the z direction. Therefore, the DBT volume is always displayed as slices (or "planes" and "slabs" depending on the slice interval, as referred to by GE) in the x-y plane. We use a reconstruction voxel size of 0.1 mm × 0.1 mm × 1 mm, corresponding to a slice interval of 1 mm. Several slices are usually added above and below the volumes by the reconstruction algorithm to ensure adequate coverage of the breast. For a compressed breast thickness of 5 cm, a typical reconstructed image volume contains about 60 slices, each of which can be larger than 1500×2500 pixels in area.

2.1.2 Masses and Microcalcifications

Radiologists look for various signs and features in breast images and use a combination of these features and clinical information to make a diagnosis. When evaluating the image quality for the DBT reconstruction and processing techniques, we focus on two key signs of breast cancer: masses and microcalcifications (MCs).

Breast masses are one of the primary signs of breast cancer. They are localized low-contrast objects that have different tissue density or composition from the surrounding tissue. Masses can vary in size and shape, but irregularly shaped masses with spiculated or ill-defined margins are more likely to be cancerous. Figure 2.2(a) shows an example of a malignant mass (invasive ductal carcinoma). When evaluating image quality, we are concerned about whether the images preserve fine textural details and natural appearance of masses and are free of any undesirable artifacts. To our knowledge, there is no established quantitative measure that correlates with human visual preference for this aspect. Therefore, we resort to visual assessment to monitor the textural changes of masses in the DBT images reconstructed or processed by various methods.

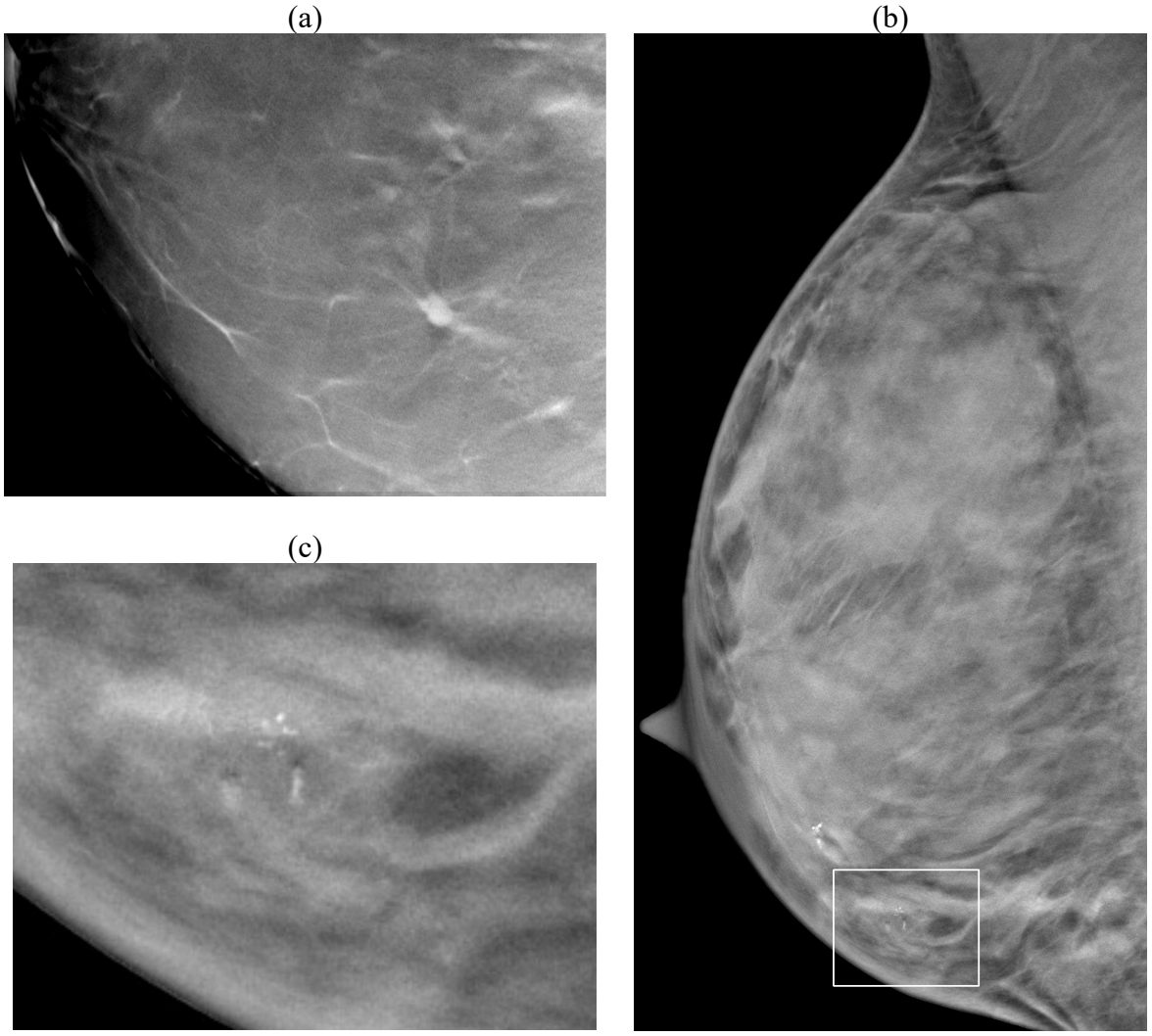


Figure 2.2 (a) An example 5 mm spiculated mass (invasive ductal carcinoma). (b) An example MC cluster (ductal carcinoma in situ) marked by the white box. Its zoomed-in view is shown in (c).

MCs are tiny deposits of calcium that manifest as white specks or dots in breast x-ray images. While some MCs can be non-cancerous, certain patterns of MCs may suggest the presence of cancer. MCs can be extremely tiny and appear in clusters or dispersed across the breast tissue. Clinically significant MCs seen in breast x-ray images have diameters of less than 0.5 mm. Although MCs contain calcium that has relatively high x-ray attenuation, the small sizes can result in overall low conspicuity. Figure 2.2(b)(c) show an example of an MC cluster (ductal carcinoma in situ). Subtle MCs can be easily mistaken for image noise, so it is important to reduce image

noise and enhance MCs. In this thesis, we use metrics such as contrast-to-noise ratio (CNR), detectability index (*d'*), and a CNN-based model observer to quantify the conspicuity of MCs in DBT images.

2.2 Physics of X-ray Imaging

2.2.1 X-ray Attenuation and Beer's Law

X-rays are a form of high-energy ionizing electromagnetic radiation. They have shorter wavelengths and higher frequencies than visible light, making it capable of penetrating matter, including human tissues. When x-ray photons pass through an object, they are absorbed, scattered, or transmitted. The combined effect of absorption and scattering is called x-ray attenuation. The photons that are not absorbed or scattered will exit from the object and are measured by the detector. Different materials have different attenuations, which give rise to contrasts representing the internal structures inside the object in the x-ray projection image. Some of the scattered photons may also escape from the object but they do not carry useful structural information. The scattered photons reduce the contrasts of the x-ray image if recorded by the detector.

The linear attenuation coefficient of a material characterizes the extent to which it attenuates x-rays as they pass through a unit thickness of the material. It is typically measured in units of cm^{-1} or mm^{-1}. The attenuation coefficient depends on the type of material and the energy of incident photons. In breast imaging, the materials of interest include glandular tissue, adipose tissue (fat), fibrous tissue, as well as MC, among others.

Assume that the spatial distribution of the attenuation coefficients of the object being imaged is $\mu(x, y, z; \mathcal{E})$ where $\mathcal{E}$ is the photon energy. For a ray l that connects the source and the detector pixel at location (s, t), the measured x-ray intensity $I(s, t)$ is given by the Beer's law

$$I(s,t) = \int_0^{\mathcal{E}_{\max}} \eta(\mathcal{E}) I_0(s,t;\mathcal{E}) \exp\left(-\int_0^l \mu(\tilde{l};\mathcal{E}) d\tilde{l}\right) d\mathcal{E} \tag{2.1}$$

where $I_0(s,t;\mathcal{E})$ is the incident x-ray intensity, $\eta(\mathcal{E})$ is the detector quantum absorption efficiency. Equation (2.1) is the general expression for a polyenergetic x-ray source. In this thesis, we consider a monoenergetic x-ray source to simplify our analysis. We assume that the acquired data has been preprocessed to correct for factors such as beam hardening and detector artifacts. Then the Beer's law (2.1) simplifies to

$$I(s,t) = I_0(s,t) \exp\left(-\int_0^l \mu(\tilde{l}) d\tilde{l}\right) \tag{2.2}$$

where $\mu(x,y,z)$ is the attenuation at the source energy. We can take log transform of (2.2) to obtain the post-log measurement which is linear in μ

$$\log \frac{I_0(s,t)}{I(s,t)} = \int_0^l \mu(\tilde{l}) d\tilde{l}. \tag{2.3}$$

2.2.2 Image Reconstruction

Image reconstruction is the process of obtaining the spatial distribution of the attenuation coefficients of the object from the measured data which can be later viewed as cross-sectional images or slices. In practice, the continuous-space object μ is discretized into voxels for easier computation and storage. Assume that the voxelized (and vectorized) representation of μ, i.e., the image, is $x \in \mathbb{R}^N$ where N is the number of image voxels. This leads to the following discretization of (2.3)

$$\log \frac{I_0(s,t)}{I(s,t)} = \int_0^l \mu(\tilde{l}) d\tilde{l} \approx a'x \tag{2.4}$$

where the entries of $a \in \mathbb{R}^N$ represent the lengths of ray l within each voxel, and $'$ denotes matrix transpose. At a given scan angle, by stacking (2.4) from all detector pixels, we obtain the system

of equations that describes the relationship between the (vectorized) post-log PV measurement $y_i \in \mathbb{R}^M$ and the unknown image x

$$y_i = A_i x, \qquad A_i = \begin{bmatrix} a_1' \\ \vdots \\ a_M' \end{bmatrix} \tag{2.5}$$

where $A_i \in \mathbb{R}^{M \times N}$ is called the system matrix or forward projector, M is the number of detector pixels, $i = 1, \dots, N_p$ is the scan angle index, N_p is the total number of scan angles. Furthermore, the overall DBT system of equations is the stack of (2.5) from all scan angles

$$y = Ax, \qquad y = \begin{bmatrix} y_1 \\ \vdots \\ y_{N_p} \end{bmatrix}, \; A = \begin{bmatrix} A_1 \\ \vdots \\ A_{N_p} \end{bmatrix}. \tag{2.6}$$

Zheng *et al*. developed an accurate and efficient implementation of A for DBT called segmented separable footprint (SG) projector that is capable of calculating the average of x-ray paths over the finite-sized detector pixels from an image with anisotropic voxels [23].

Image reconstruction involves solving the system of equations (2.6) for x. This task is challenging due to several fundamental reasons. First, the problem is underdetermined because there are more unknowns (image voxels) than equations (PV measurement). Consequently, a unique solution cannot be guaranteed, and regularization is essential to constrain the solution space. Second, the reconstruction problem is ill-posed and sensitive to noise. Medical imaging inherently contains noise, which introduces errors into the measurements and can drastically affect the reconstructed images. Third, the system matrix A is typically large and sparse, so calculating its (pseudo) inverse directly is computationally infeasible and can lead to numerical instability. With these challenges in mind, the subsequent section provides a review of tomographic reconstruction methods.

2.3 Tomographic Reconstruction Methods

2.3.1 Analytical Methods

Filtered backprojection (FBP) is one of the most commonly used analytical methods for tomographic reconstruction. It is rooted from the Fourier-slice theorem in continuous space. As the name suggests, FBP first filters the PVs by a ramp filter, and then backprojects these filtered PVs to form the images. FBP is mathematically straightforward, making it easy to implement and understand. It is also computationally efficient and can produce images quickly.

Nonetheless, it is important to recognize the drawbacks of FBP. FBP can produce certain artifacts in the reconstructed images, such as streaks. Moreover, it is less suitable for low dose settings because it ignores measurement noise in the problem formulation. The ramp filter amplifies high frequency noise, so one must apodize it to balance the enhancement of noise and edges. Analytical methods like FBP also rely on certain standard imaging geometries, such as parallel-beam, for deriving the point spread function, which make them not robust when the application changes to other geometries, such as fan-beam or cone-beam. For DBT, most commercial systems at present use iterative reconstruction methods.

2.3.2 Algebraic Methods

Algebraic methods solve the system of equations (2.6) by iteratively refining the unknown image to minimize the difference between the forward projection of the current estimate and the measured data. Algebraic methods for tomographic reconstruction include algebraic reconstruction technique (ART) [24], simultaneous iterative reconstruction technique (SIRT) [25], and simultaneous algebraic reconstruction technique (SART) [26]. Notably, research has demonstrated that a single iteration of SART is well-suited for DBT reconstruction [27]. The GE

Pristina system uses a version of SART commercially named as Adaptive Statistical Iterative Reconstruction for DBT (ASIR^{DBT}) [19].

We use SART as the baseline reconstruction method in this thesis. The update equation of SART is

$$x_n^{i+1} = x_n^i + \left(A_i' \left(\left(y_i - A_i x_n^i\right) \oslash \left(A_i \mathbf{1}_N\right) \right) \right) \oslash \left(A_i' \mathbf{1}_M\right) \tag{2.7}$$

where n is the SART iteration index and is incremented after i goes through 1 to N_p, $\oslash$ represents element-wise division, $\mathbf{1}_N$ denotes the vector of ones, and $'$ denotes matrix transpose. Much like the analytical methods, the algebraic reconstruction methods do not incorporate noise modeling into the problem formulation.

2.3.3 Statistical Methods

Statistical methods used in image reconstruction integrate both the statistical properties of the measurement noise and the prior knowledge about the images.

Maximum likelihood (ML) estimation aims to find x by maximizing the log of the likelihood of observing the measurement y given x

$$\underset{x}{\operatorname{argmax}}\, L(x), \qquad L(x) \coloneqq \log P(y|x). \tag{2.8}$$

ML approaches allow for better handling of data noise and uncertainty and results in improved image quality. In cases where the measurement noise is additive Gaussian, $y = Ax + \varepsilon$ with $\varepsilon \sim \mathcal{N}(0, K)$, where K is the noise covariance matrix, the log-likelihood function $L(x)$ becomes

$$L(x) = -\frac{1}{2}\|y - Ax\|_K^2 \tag{2.9}$$

where we have omitted constant and scaling terms, and $\|v\|_W^2 = v'Wv$. Research has also shown that SART converges to the geometrically weighted least squares estimator $\underset{x}{\operatorname{argmin}} \frac{1}{2}\|y -$

$Ax\|^2_{\text{Diag}\{A\mathbf{1}_N\}^{-1}}$ [28], making it a special case of ML estimation. ML estimators tend to have high noise and are unsuitable for low-dose DBTs.

In addition to ML estimation, another statistical method known as penalized-likelihood estimation incorporates a regularization function $R(x)$ into the ML framework to control image noise level. The choice of $R(x)$ is driven by our prior knowledge about the nature of images, such as their sparsity or smoothness. This approach seeks to find the solution to the following optimization problem

$$\underset{x}{\text{argmin}} -L(x) + \beta \cdot R(x) \tag{2.10}$$

where $\beta > 0$ represents the regularization parameter. Penalized-likelihood methods can also be viewed as the maximizing a posteriori (MAP) estimation from a Bayesian perspective. To illustrate this connection, MAP estimation aims to find x by maximizing the log of the image posterior distribution given the observed PV y

$$\underset{x}{\text{argmax}} \log P(x|y) = \underset{x}{\text{argmax}} \log \left(\frac{P(y|x)P(x)}{P(y)} \right) = \underset{x}{\text{argmax}} L(x) + \log P(x). \tag{2.11}$$

In this context, $R(x) \coloneqq \log P(x)$ is the log-prior of the image distribution.

The penalized-likelihood formulation (2.10) can account for noise characteristics and regularization terms and can also incorporate mathematical models of the imaging physics and processes. This problem does not typically have a closed-form solution, necessitating the use of iterative algorithms. Consequently, iterative optimization algorithms for solving (2.10) are commonly called model-based iterative reconstruction (MBIR) algorithms. Chapter 3 of this thesis develops a DBT image denoiser that implicitly defines a regularizer that is subsequently incorporated into MBIR in Chapter 4. Chapter 5 is motivated by the idea of modifying the system

matrix A to model the x-ray source motion blur for DBT MBIR, although this approach turns out to be challenging and we resort to post-processing deblurring.

2.4 Deep Learning

2.4.1 Neural Networks

Neural networks are a class of machine learning models inspired by the structure of the human brain. They consist of interconnected nodes, often referred to as neurons, organized into layers. Neural networks are powerful function approximators that are capable of modeling complex and nonlinear relationships in data [29]. Their joint, end-to-end training also makes them automatically learn important features from data instead of manually crafting them for the intended tasks.

A neural network typically has an input layer, several hidden layers, and an output layer. To give a concrete example, a two-layer neural network with input $x \in \mathbb{R}^{D_0}$, one hidden layer $h \in \mathbb{R}^{D_1}$, and output $y \in \mathbb{R}^{D_2}$ can be formulated as

$$h = \sigma(W_1 x + b_1), \qquad y = W_2 h + b_2 \tag{2.12}$$

where $W_1 \in \mathbb{R}^{D_1 \times D_0}$ and $W_2 \in \mathbb{R}^{D_2 \times D_1}$ are the learnable weights of the network, $b_1 \in \mathbb{R}^{D_1}$ and $b_2 \in \mathbb{R}^{D_2}$ are the learnable biases. The nonlinear function $\sigma(\cdot)$ is called the activation function that is applied element-wise to the vector input. A commonly used activation function is the Rectified linear unit (ReLU) defined as $\sigma(t) = \max(0, t)$. For classification problems, the softmax function is often used as the last activation to normalize the output. For regression problems, the output layer usually does not have activation.

2.4.2 Backpropagation

The goal of training a neural network is to adjust the weights and biases so that the network can learn and extract meaningful information from data to make accurate predictions. With a properly defined training loss function based on the specific tasks, gradient-based optimizers can be used to optimize the loss, such as stochastic gradient descent or Adam [30]. Optimizing neural networks is a nonconvex problem in general, where gradient-based optimizers may return local minima. However, empirical evidence strongly suggests that local minima tend to perform well for large networks [8].

Backpropagation is a fundamental learning algorithm in training neural networks. It applies the chain rule to compute the gradients for all intermediate variables and parameters. This process begins at the output layer and works backward until the input layer is reached. At each node, the upstream gradient is multiplied by the local gradient to determine the gradient that is used not only for optimizing the node but also for downstream gradient computations. For example, for the two-layer network (2.12), assuming the training loss is $L(y) \in \mathbb{R}$, the gradients of L with respect to W_1, W_2, b_1, and b_2 are calculated using backpropagation as follows. For the second layer,

$$\frac{\partial L}{\partial W_2} = \frac{\partial y}{\partial W_2}\frac{\partial L}{\partial y} = \nabla L(y)h', \qquad \frac{\partial L}{\partial b_2} = \frac{\partial y}{\partial b_2}\frac{\partial L}{\partial y} = \nabla L(y) \tag{2.13}$$

where $\nabla L(y) \in \mathbb{R}^{D_2}$ is the derivate of the loss, and $'$ denotes matrix transpose. Although we do not optimize over h, we also compute its gradient

$$\frac{\partial L}{\partial h} = \frac{\partial y}{\partial h}\frac{\partial L}{\partial y} = W_2'\nabla L(y). \tag{2.14}$$

Then, for the first layer,

$$\frac{\partial L}{\partial W_1} = \frac{\partial h}{\partial W_1}\frac{\partial L}{\partial h} = \big(W_2'\nabla L(y) \odot \dot{\sigma}(W_1 x + b_1)\big)x', \tag{2.15}$$

$$\frac{\partial L}{\partial b_1} = \frac{\partial h}{\partial b_1}\frac{\partial L}{\partial h} = W_2' \nabla L(y) \odot \dot{\sigma}(W_1 x + b_1)$$

where $\odot$ represents element-wise multiplication. Note that we have used the upstream gradient $\frac{\partial L}{\partial h}$ in (2.14). The function $\dot{\sigma}(\cdot)$ is the derivative of the activation function. For ReLU, $\dot{\sigma}(t) = \chi_{(0,\infty)}(t)$ is an indicator function of positive numbers.

Since backpropagation operates with a modular approach, it is readily extensible to more complex, deeper neural networks and, in principle, can be applied to any computational graph. Modern deep learning frameworks, such as PyTorch[1] and TensorFlow[2], offer highly efficient implementations of backpropagation for diverse layers, modules, losses, and operators, enabling fast and flexible prototyping of deep neural networks.

2.4.3 Convolutional Neural Networks

Convolutional neural networks (CNNs) are a specialized type of neural network designed for tasks involving grid-like data such as images. CNNs use a specific layer called a convolution layer that applies convolution operations to the input data. These convolution operations involve small filters (also called kernels) that slide over the input. Consider an input with size $H \times W \times C_{\text{in}}$ (height, width, channels). The multiple channels can be either the color channels of an image, or the multiple feature maps from the output of another convolution layer. A 2D convolution layer uses filters with size $K_W \times K_H \times C_{\text{in}}$ to convolve across the spatial dimensions of the input. The filters always have the same number of channels as the input, so there is no sliding in the channel dimension. Also, the input is usually zero padded so that the spatial dimensions remain the same

[1] https://github.com/pytorch/pytorch
[2] https://github.com/tensorflow/tensorflow

after convolution. By using a total of C_{out} filters, the convolution layer produces a total of C_{out} feature maps with size $H \times W \times 1$, or $H \times W \times C_{\text{out}}$ by stacking them together.

Like neural networks, activation functions are incorporated in CNNs to introduce nonlinearity. Pooling layers are often used to reduce the spatial dimensions. The use of multiple filters allows the convolution layer to capture diverse features and patterns in parallel. Perhaps most notably, the convolution layers can be stacked to create a deep CNN. Research has demonstrated that deep CNNs can learn and represent features at various levels of abstraction [31]. In particular, lower layers of a deep CNN tend to identify basic edges and textures, while higher layers can combine these features to recognize more complex shapes and structures. This hierarchical representation makes deep CNNs remarkably effective in image processing and computer vision tasks. In this thesis, we develop deep CNNs for DBT image denoising in Chapter 3 as well as noise estimation and clustered MC detection in Chapter 4. We also use CNNs for constructing the diffusion models in Chapter 5.

Chapter 3
Deep Convolutional Neural Network with Adversarial Training for Denoising Digital Breast Tomosynthesis Images

3.1 Introduction

It[1] is a challenge to denoise DBT images because conventional noise smoothing methods may also smooth out the subtle MCs. Researchers have tried various methods to suppress noise in DBT images. PV filtration was performed using a linear filter [32] or a neural network filter with one convolutional layer [33]. Model-based iterative reconstruction (MBIR) has attracted much attention because of its potential for handling noise. Statistical noise models, such as noise variance [27][34][35], detector blur and correlated noise (DBCN) [36] and scattered noise [37], were incorporated into the DBT system model for MBIR. Gradient-based regularizers, such as selective-diffusion regularizer [38], total variation (TV) [39][40] and its variants [41]–[43], were also used in MBIR of DBT for noise reduction. However, MBIR techniques may introduce "plastic appearance" to the soft tissue structures as observed in CT reconstruction [44][45]. Several denoising methods were proposed for the reconstructed DBT images. Das *et al*. used a 3D Butterworth filter to improve MC detection [46]. Abdurahman *et al*. iteratively applied a smoothing filter to improve the contrast-to-noise ratio (CNR) of MCs [47]. Lu *et al*. applied

[1] This chapter is based on the published journal paper [10] that extends the conference paper [11] and abstract [12].

multiscale bilateral filtering [48] either to the reconstructed images as post-processing or between reconstruction iterations, improving the CNRs of MCs without distorting the masses.

Recently, deep convolutional neural network (DCNN) methods have shown state-of-the-art performances in natural image restoration tasks. Zhang *et al*. constructed a feed-forward denoising convolutional neural network (DnCNN) for Gaussian noise removal [49]. The DCNN training loss was the mean squared error (MSE) between the network output and clean training target. Dong *et al*. trained a three-layer convolutional network [50] and Kim *et al*. trained a 20-layer network [51] with MSE loss for single-image super-resolution. However, there is a perception-distortion tradeoff: the MSE loss tends to produce overly smoothed images that are not visually satisfactory even if their MSE, peak signal-to-noise ratio or structural similarity are high [52]. Alternative training losses were designed to address this problem. For example, Johnson *et al*. used feature-level MSE loss, called the perceptual loss, for image transformation and super-resolution [53]. Inspired by the generative adversarial network (GAN) [54], Ledig *et al*. introduced the adversarial loss for image super-resolution and greatly increased the mean opinion scores [55]. The adversarial training stability was further improved as the Wasserstein GAN (WGAN) was proposed [56][57]. The adversarial loss was applied to medical image processing and achieved promising results, including CT denoising and artifacts correction [58]–[62] and MRI de-aliasing [63][64].

We conducted a preliminary study that used a DCNN to denoise PVs before DBT reconstruction and achieved moderate CNR improvement for MCs [65]. In this chapter, we trained a DCNN using a weighted combination of MSE loss and adversarial loss to denoise reconstructed DBT images. We focused on improving the conspicuity of MCs and preserving the natural appearance of soft tissues and masses in DBT images.

A DCNN having millions of parameters requires a large amount of data to learn complex image patterns. However, in medical imaging fields, training data is limited due to the high costs of collecting and annotating the data. For a denoising task using a supervised approach, the DCNN training requires high dose (HD) images as references or targets to learn to reduce noise of a corresponding input low dose (LD) images, but we cannot scan a patient with a HD technique. To overcome these problems, we studied the feasibility of using two methods for generating data to train DCNN for DBT denoising. The first method is to generate *in silico* training data. The virtual imaging clinical trial for regulatory evaluation (VICTRE) project [66] conducted a computer-simulated imaging trial to evaluate DBT as a replacement for digital mammography. It provides an anthropomorphic breast model to generate digital breast phantoms[2] [67]. We incorporated the digital breast phantom into an x-ray imaging simulation tool developed by GE Global Research, named Computer Assisted Tomography SIMulator[3] (CatSim) [68][69], to generate relatively realistic breast images from a clinical DBT system and use them for DCNN training. The second method is to prepare physical heterogeneous breast phantoms using tissue-mimicking materials and scan them with a DBT imaging system [70][71]. We trained the DCNNs using the two types of data and compared their denoising performances.

3.2 Methods and Materials

3.2.1 DCNN Training

3.2.1.1 DNGAN Framework

To reduce the noise in the reconstructed DBT images, we would like to obtain a mapping function, called a denoiser, that maps the noisy images to clean or less noisy ones. The denoiser

[2] Available at https://github.com/DIDSR/VICTRE.
[3] We used an earlier version of CatSim that is Matlab-based. A Python-based CatSim is recently available at https://github.com/xcist/CatSim.

was implemented as a DCNN with trainable weights. In the training phase, the denoiser learned how to denoise by adjusting the trainable weights to minimize the training loss function. In the deployment phase, the denoiser with frozen weights served as a well-trained function that could be applied to noisy DBT images.

During training, the system took pairs of low dose (LD) and high dose (HD) images as input and target, respectively. The HD target images could be obtained from either simulation or physical phantom scans with a manual exposure setting (see Section 3.2.2 for details). They were used to guide the denoiser to generate denoised images from the LD images by minimizing a weighted combination of MSE loss and adversarial loss, where the adversarial loss was derived by training a discriminator to distinguish between the denoised LD and the target HD images as in a GAN. As demonstrated below, the GAN-based adversarial training was crucial to constrain the degree of smoothing and maintain the sharpness of the denoised DBT images. We therefore call our training framework DNGAN. Figure 3.1 shows the framework of the DNGAN.

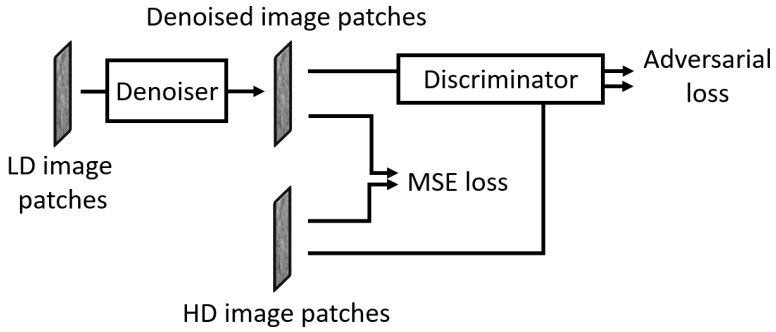


Figure 3.1 The framework of the denoising DCNN with adversarial training (DNGAN).

3.2.1.2 Network Structures

We designed the network structure of the denoiser based on the DnCNN [49]. The original DnCNN structure consisted of 17 to 20 convolutional layers, each including 64 filters of 3×3 kernels, and used rectified linear units (ReLU) between the layers. Each convolution layer was followed by batch normalization [72] before the ReLU except for the first convolution layer. For

our DBT denoising task, we chose to use LD/HD image regions, or patches, 32×32 pixels in size as inputs. This patch size was determined so that it could cover the noise autocorrelation and was also small to allow the DCNN to focus on the local image structures in the adversarial training [73]. Our pilot studies found that the structure could be reduced to 10 convolutional layers, each with 32 filters, without substantial difference in performance. We also removed the batch normalization layers without experiencing training instability. We therefore used a much smaller structure, as shown in Figure 3.2(a), to improve computational efficiency. This structure had a total of 74,593 trainable weights. To ensure that the output of the convolution layer had the same size as the input, the input was padded with values that were mirrored from the inner region of the input. Because the denoiser was fully convolutional, we could directly apply it to the full DBT slices during deployment.

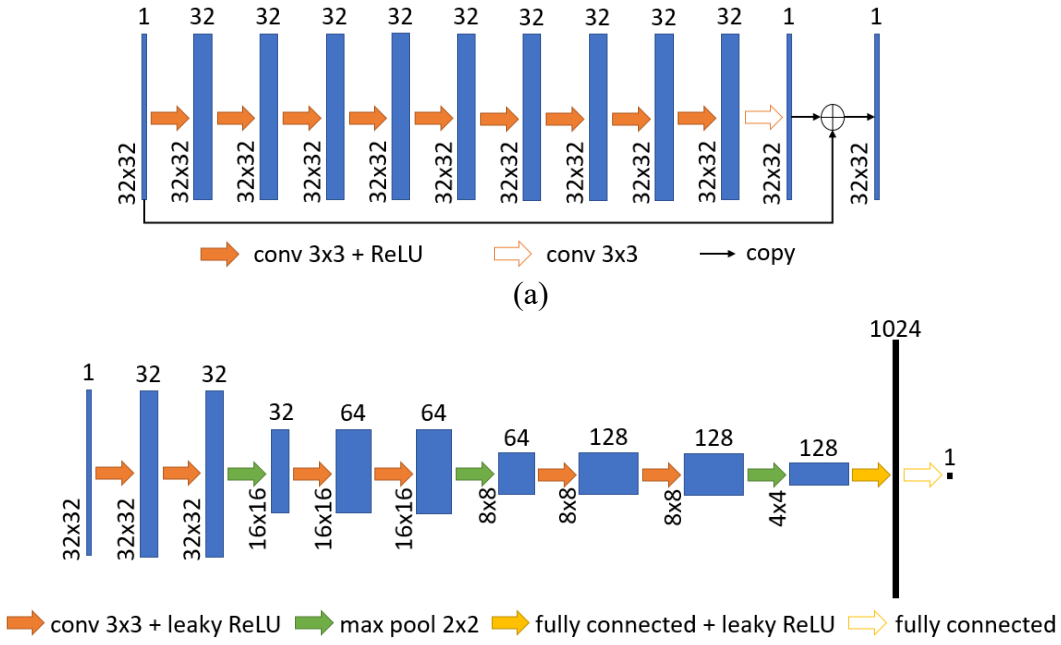


Figure 3.2 The network structure of (a) the denoiser and (b) the discriminator. The numbers on the left of each layer (rectangle) represent the width and height of the tensors. The numbers on the top of the rectangles represent the number convolution filters for the tensors (blue) or the length of the vectors (black).

Figure 3.2(b) shows the network structure of the discriminator. We used the VGG-Net [74], with a reduced number of downsampling blocks due to the small input patch size, as the backbone of our discriminator. The discriminator had a total of 2,385,633 trainable weights.

3.2.1.3 Training Loss Function

The training loss function is composed of the MSE loss L_{MSE} and the adversarial loss L_{adv}

$$\underset{G}{\operatorname{argmin}}\, L_{\text{MSE}}(G) + \lambda_{\text{adv}} \cdot L_{\text{adv}}(G) \tag{3.1}$$

where G denotes the denoiser, λ_{adv} is a tuning parameter controlling the weighting between the MSE loss, which contributes to image smoothness, and the adversarial loss, which contributes to preserving high frequency image textures.

The MSE loss compares the pixel-wise difference between the denoised image patches and the corresponding HD target image patches x_{target} as follows

$$L_{\text{MSE}}(G) = \mathbb{E}_{(x_{\text{noisy}},\, x_{\text{target}})}\left[\frac{1}{N_{\text{pixel}}}\left\|G\left(x_{\text{noisy}}\right) - x_{\text{target}}\right\|^2\right] \tag{3.2}$$

where x_{noisy} is the LD noisy input patch, N_{pixel} is the number of pixels in an image patch.

We implemented the adversarial loss as the WGAN with gradient penalty [57]. The key idea behind it is summarized as follows. Assume $\Omega = \mathbb{R}^{N_{\text{pixel}}}$ is the sample space of DBT images, $\text{Prob}(\Omega)$ is the space of probability measures defined on Ω. We can view the denoiser $G: \Omega \to \Omega$ as a function parameterized by its trainable weights. It generates denoised images following a distribution $\mathbb{P}_G \in \text{Prob}(\Omega)$ from the input samples that follow a noisy image distribution $\mathbb{P}_{\text{noisy}} \in \text{Prob}(\Omega)$. The Wasserstein distance (WD) between the distribution of the denoised images $\mathbb{P}_G$ and the distribution of the HD target images $\mathbb{P}_{\text{target}} \in \text{Prob}(\Omega)$ is defined as [56]

$$d\left(\mathbb{P}_G, \mathbb{P}_{\text{target}}\right) = \inf_{\gamma \in \Pi\left(\mathbb{P}_G,\, \mathbb{P}_{\text{target}}\right)} \mathbb{E}_{(p,\, q)\sim\gamma}[\|p - q\|] \tag{3.3}$$

where $\Pi(\mathbb{P}_G, \mathbb{P}_{\text{target}})$ denotes the set of all joint distributions $\gamma(p, q)$ whose marginals are $\mathbb{P}_G$ and $\mathbb{P}_{\text{target}}$, respectively. Arjovsky *et al*. [56] showed that, instead of directly solving (3.3), which is intractable, one could solve

$$d\left(\mathbb{P}_G, \mathbb{P}_{\text{target}}\right) = \max_{\|D\|_L \leq 1} \mathbb{E}_{x_{\text{target}}}\left[D\left(x_{\text{target}}\right)\right] - \mathbb{E}_{x_{\text{noisy}}}\left[D\left(G\left(x_{\text{noisy}}\right)\right)\right]. \tag{3.4}$$

In other words, to calculate the WD, we need to find a 1-Lipschitz function $D: \Omega \to \mathbb{R}$ that maximizes the objective function. D is also called a discriminator, or a critic, to output a similarity score that assesses whether the input image patch comes from the target distribution. The discriminator is approximated by a DCNN in our implementation. Its training loss function is [56]

$$\underset{\|D\|_L \leq 1}{\operatorname{argmax}} \hat{d}_G(D) \coloneqq \mathbb{E}_{x_{\text{target}}}\left[D\left(x_{\text{target}}\right)\right] - \mathbb{E}_{x_{\text{noisy}}}\left[D\left(G\left(x_{\text{noisy}}\right)\right)\right]. \tag{3.5}$$

Arjovsky *et al*. showed that $\hat{d}_G(D)$ can be interpreted as an estimation of the WD. Note that $\hat{d}_G(D)$ also depends on G. Gulrajani *et al*. proposed a gradient penalty to constrain the 1-Lipschitz condition [57], so the overall training loss function for the discriminator becomes

$$\underset{D}{\operatorname{argmin}} -\hat{d}_G(D) + \lambda_D \cdot \mathbb{E}_{\bar{x}}[(\|\nabla_{\bar{x}} D(\bar{x})\| - 1)^2] \tag{3.6}$$

where $\bar{x} = t \cdot x_{\text{target}} + (1 - t) \cdot G(x_{\text{noisy}})$ is an interpolated image, $t \sim \text{Unif}([0,1])$, λ_D is the penalty weight. To promote the denoised images to be perceptually similar to the target images, we maximize the term $\mathbb{E}_{x_{\text{noisy}}}\left[D\left(G\left(x_{\text{noisy}}\right)\right)\right]$ in $\hat{d}_G(D)$ when optimizing G, which gives the corresponding adversarial loss of the denoiser training

$$L_{\text{adv}}(G) = -\mathbb{E}_{x_{\text{noisy}}}\left[D\left(G\left(x_{\text{noisy}}\right)\right)\right]. \tag{3.7}$$

In practice, the denoiser and the discriminator are trained alternately in DNGAN so that the discriminator is always up to date for estimating $d\left(\mathbb{P}_G, \mathbb{P}_{\text{target}}\right)$ and the denoiser improves through iterations.

3.2.1.4 Fine-tuning with MC Patches

For DNGAN training, the DBT images with any kind of breast structures can be used for training, as long as the LD/HD image pairs contain matched structures to be preserved and residual differences to be reduced. In our training set preparation, we extracted non-overlapping patches from the DBT slices using a shifting window. A DBT volume mainly consisted of tissue background and there were very few MCs in a volume. Consequently, the training set was dominated by background patches. To emphasize the MC images so that the denoiser could focus on the MC signals and learn to preserve or enhance them, we investigated the feasibility of a second training stage that fine-tuned the DNGAN only with patches centered at individual MCs.

In the fine-tuning stage, we adopted the training technique of layer freezing [75]. For example, freezing m layers means that layers 1 to m in the denoiser were frozen, and layers $(m+1)$ to the last layer were active. Our denoiser had a total of 10 layers, so $m \leq 10$. Note that the freezing was only applied to the denoiser network. All layers in the discriminator were active during training.

3.2.2 Data Sets

We prepared several data sets to investigate the effects of the dose level of the HD training target, the training sample size, and the underlying reconstruction algorithm on the performance of our proposed DNGAN. We also prepared a data set for the aforementioned MC fine-tuning experiment. Table 3.1 summarizes the data sets and their use in the experiments.

As introduced in Section 3.1, we generated two types of data, namely the digital phantom data and the physical phantom data, for DNGAN training and validation. The digital phantom data provided a wide range of x-ray exposures including noiseless images for the study of the effect of the dose level of the HD training target. We also used the digital data to generate the MC fine-

tuning set because the coordinates of the simulated MCs were known exactly. Compared with the digital phantom data, the physical phantom data contained all the imaging degradation factors of a DBT system and were considered to be more realistic, and only physical phantom data imaged with the DBT system could be reconstructed with the manufacturer's proprietary reconstruction technique. We therefore used the physical phantom data for the other experiments and for study of the transferability of a digital-data-trained DNGAN denoiser to real DBT images acquired from physical phantoms and human subjects.

Table 3.1 Summary of data sets for DNGAN investigation.

Purpose	Name	Phantom type	Recon algorithm	No. of patch pairs	Comments
Training	24mAs/target	Digital	SART	199,850	target = 72mAs, 120mAs, 360mAs, noiseless. For investigating the effect of the dose level of the HD training target.
	MC fine-tuning set	Digital	SART	3,048	Patches centered at individual MCs generated at known locations. For investigating the feasibility of a second fine-tuning stage.
	LD/HD$_k$	Physical	SART	$k \times 400{,}000$	k = 20%, 35%, 50%, 65%, 80%, 100%. For investigating the effect of the training sample size.
	LD/HD$_{Pristina}$	Physical	Pristina	400,000	For training a matched denoiser when evaluating the generalizability of DNGAN in terms of the reconstruction algorithms.
Validation	24mAs as input, higher dose levels as reference truth	Digital	SART	/	Has ground truth scans simulated at multiple dose levels. Used for NPS comparison.
	LD as input, HD as reference for performance comparison	Physical	SART	/	Has individually marked MCs of three nominal diameters. Used for CNR, FWHM, fit success rate, *d'*, and visual comparisons.
			Pristina	/	For evaluating the generalizability of DNGAN in terms of the reconstruction algorithms.
Test	Human subject DBTs	/	SART	/	An independent test set. For demonstrating the robustness and the feasibility of applying a denoiser trained with phantom data to human DBTs.

* The names of the training sets (24mAs/target, LD/HD$_k$, and LD/HD$_{Pristina}$) refer to the image conditions of training image pairs.

3.2.2.1 Digital Phantom Data

We prepared 25 heterogeneous dense (34% glandular volume fraction) 4.5-cm-thick digital phantoms at a voxel resolution of 0.05 mm using the VICTRE breast model [66][67]. We inserted simulated MCs consisting of calcium oxalate in clusters into the digital phantoms. Each cluster

had 12 MCs arranged on a 3-by-4 grid parallel to the detector plane with a small offset in the direction perpendicular to the chest wall to avoid in-plane artifacts interfering with each other during reconstruction. The MCs had three diameters: 0.150 mm, 0.200 mm, 0.250 mm. We used 24 phantoms for training data preparation and held out one phantom for validation.

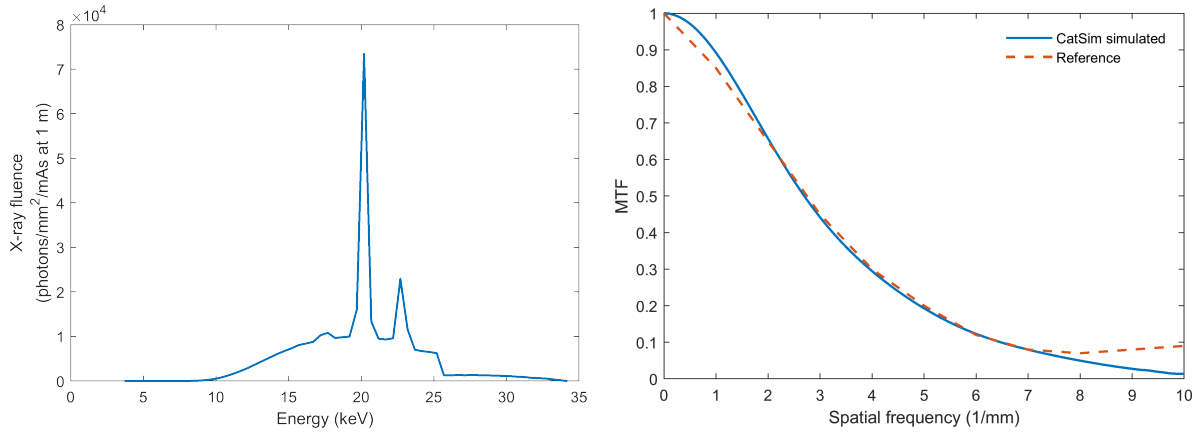
Next we configured CatSim [68][69] to model the GE Pristina DBT system (GE Healthcare) [22] as follows: set the acquisition geometry as 9 PVs within $\pm 12.5°$ at a detector pixel size of 0.1 mm × 0.1 mm; set the x-ray fluence spectrum at 34 kVp from a Rh anode [76] with a 0.03 mm Ag filter (Figure 3.3(a)); used the x-ray detection model developed by Carvalho [77] for the CsI/Si flat panel indirect detector. The simulated signal at the ith detector pixel $Y(i)$ is

$$Y(i) = c \cdot h_{\text{scint}} * \left(\sum_{E \in \mathcal{E}} E \cdot \text{Poisson}\{\eta(E,i) \cdot I_{\text{inc}}(E,i)\} \right) \tag{3.8}$$

where c is a conversion factor from photons to electrons, h_{scint} is the scintillator blur kernel, $*$ represents spatial convolution, $\mathcal{E}$ is the set of energy bins in the input x-ray spectrum, Poisson{·} denotes Poisson distribution, $\eta(E,i)$ is the energy-dependent detection efficiency at detector element i, $I_{\text{inc}}(E,i)$ is the x-ray intensity spectrum incident on the detector element i. Focal-spot blur, scattered radiation, grid, and electronic noise were not considered in the simulation for this study. To validate the system response of CatSim simulation, we calculated its presampled modulation transfer function (MTF) for the central PV using the edge method [78]. It agreed well with the measured Pristina MTF in the literature [79], as shown in Figure 3.3(b).

To simulate LD PVs for the digital phantoms, we set the total x-ray exposure of 9 PVs to 24 mAs in CatSim, which was close to the value from automatic exposure control (AEC) for a 4.5 cm breast for the Pristina system [79]. The estimated mean glandular dose (MGD) was 1.42 mGy

under this exposure, calculated by a Monte Carlo simulation tool called CatDose in the CatSim package. We reconstructed the DBT volumes at a voxel size of 0.1 mm × 0.1 mm × 1 mm using three iterations of simultaneous algebraic reconstruction technique (SART) [27] with the segmented separable footprint (SG) projector [23].



(a) (b)

Figure 3.3 (a) The x-ray fluence spectrum of a 34 kVp Rh anode after 0.03 mm Ag filtration in the CatSim simulation. (b) The system MTF for the central PV. The edge for MTF calculation was placed on the breast support plate and was parallel to the chest wall with 3° tilt.

3.2.2.2 Physical Phantom Data

We used seven 1-cm-thick heterogeneous slabs with 50% glandular/50% adipose breast-tissue-equivalent material to construct the physical phantoms [70][71]. By arranging five slabs in different orders and orientations, we formed nine 5-cm-thick phantoms. Clusters of simulated MCs (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. Glass beads are used in some commercial breast phantoms but have lower x-ray attenuation than calcium oxalate specks of the same size. We used eight phantoms for training data preparation and held out one phantom for validation.

Each phantom was scanned twice, one at LD and the other at HD, by a Pristina DBT system under the same compression. The LD scans were acquired with the standard dose (STD) setting, which automatically chose a technique of Rh/Ag 34 kVp. The exposures ranged between 30.4 mAs

and 32.6 mAs with a mean of 31.4 mAs for the nine phantoms. We manually set the exposure for the HD scans to Rh/Ag 34 kVp, 125 mAs. The reconstruction parameters were the same as those for the digital phantoms.

We marked the MCs in the SART-reconstructed HD volume of the hold-out validation phantom for denoiser evaluation. There was a total of 236 MCs of size 0.150-0.180 mm, 227 MCs of 0.180-0.212 mm, and 159 MCs of 0.212-0.250 mm.

3.2.2.3 Training Set Generation

Training sets with different target dose levels: Using CatSim with the digital phantoms, we prepared HD images over a range of dose levels to study the effect of the dose level of the training target images on the effectiveness of the trained denoiser. Specifically, we simulated the HD DBT scans with 72 mAs (3× AEC), 120 mAs (5× AEC), 360 mAs (15× AEC) and noiseless (∞× AEC) settings. We paired these HD scans with the 24 mAs scans to form four training sets, referred to as 24mAs/72mAs, 24mAs/120mAs, 24mAs/360mAs and 24mAs/noiseless, respectively. We extracted 199,850 pairs of patches from the 24 pairs of SART-reconstructed DBT volumes of the digital phantoms as the training set for each dose condition. Figure 3.4(a) shows the example patches in the 24mAs/120mAs training set.

MC fine-tuning set: We prepared the training set with patches centered at each MC for fine-tuning using the digital phantom data. There were 1,032 MCs of 0.150 mm, 1,008 MCs of 0.200 mm, 1,008 MCs of 0.250 mm in the 24 digital phantoms, giving a total of 3,048 MC patches in the fine-tuning set. Figure 3.4(b) shows the example MC fine-tuning set patches of the 24mAs/120mAs training set.

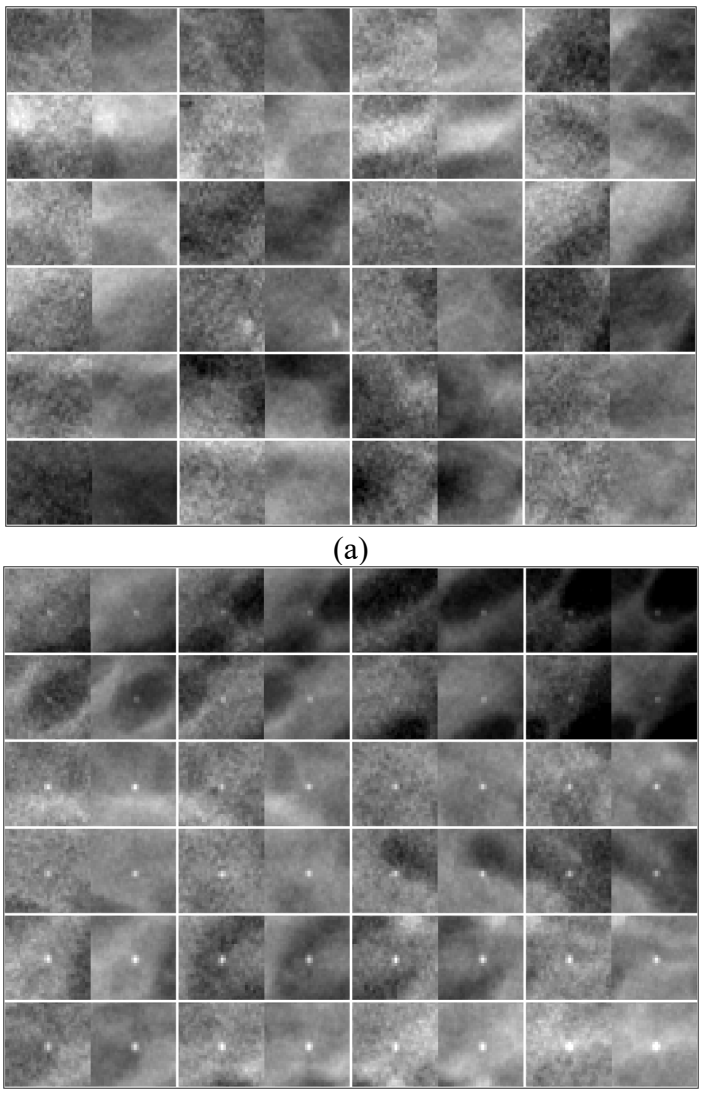


Figure 3.4 Examples of (a) DBT patches in the 24mAs/120mAs training set and (b) MC patches in the accompanying MC fine-tuning set. In each pair, the LD patch is shown on the left and the HD patch on the right. The images are from the digital simulated phantoms.

Training sets with different sample sizes: The generalizability of a trained DCNN depends on the training sample size [80]. We designed an experiment to study the effect of training sample sizes for the DBT denoising task using the physical phantom data. Specifically, we first extracted 400,000 pairs of patches from the eight physical phantoms to form the pool of training patches. Then we randomly drew 20%, 35%, 50%, 65%, 80% of patches from the pool to simulate five training set sizes in addition to the 100% set. These training sets were referred to as LD/HD$_k$,

where the subscript k is the drawing percentage. The subset drawing at each percentage was repeated 10 times with different random seeds. Although the independence among the drawn subsets decreased as k increased, the simulation study would provide some understanding of the trend and variation of the training.

Training set of Pristina-reconstructed images: The Pristina DBT system has a built-in commercial reconstruction algorithm. We refer to it as Pristina algorithm. To evaluate the generalizability of the DNGAN denoiser in terms of the reconstruction algorithms, we directly deployed the denoiser that was trained with SART-reconstructed images to the Pristina-reconstructed images. We also prepared a training set using Pristina-reconstructed images to train a matched denoiser for comparison. This training set was referred to as $LD/HD_{Pristina}$ and had 400,000 pairs of patches extracted from the eight training physical phantoms.

For all the training sets, we subtracted the mean from each DBT volume to center its histogram before patch extraction.

3.2.2.4 Human Subject DBTs

We used eleven de-identified human subject DBT scans, previously collected for another study with IRB approval, as an independent test set to evaluate the denoising effect on the CNR of MCs and the appearances of breast tissue and cancerous masses in real breasts. They contained biopsy-proven invasive ductal carcinomas (masses) and ductal carcinomas in situ (MC clusters). The images were acquired using a GE prototype GEN2 DBT system. The prototype system acquired 21 PVs in a scan angle of 60° with a 29 kVp Rh/Rh x-ray beam. We used the central 9 PVs that corresponded to a scan angle of 24° to simulate an LD DBT with scan parameters like the Pristina DBT system. Also note that the prototype system did not have a grid whereas Pristina

had. The DBTs were reconstructed with SART with the same parameters as described above. 301 MCs were marked in the DBT volumes.

3.2.3 Figures of Merit

3.2.3.1 Noise Power Spectrum

The structural noise power spectrum (NPS) of a breast image quantifies both the structured noise of the object being imaged and other noise on the imaging chain. It has been shown to have a power-law form for mammograms [81]. We used the NPS to quantify the change of textures and noise in the denoised DBT.

For NPS calculation, we first extracted a total of 40 background slice patches, 200×200 pixels each, from the DBT slices parallel to the detector plane, and then calculated the 2D NPS defined as [82]

$$\mathrm{NPS}_{2\mathrm{D}} = \frac{p_x p_y}{N_x N_y} \langle |\mathrm{DFT}_{2\mathrm{D}}\{x_i - \bar{x}_i\}|^2 \rangle_i \tag{3.9}$$

where $p_x = p_y = 0.1$ mm is the image pixel size, $N_x = N_y = 200$ is the patch size, $\langle \cdot \rangle_i$ means averaging over all patches, $\mathrm{DFT}_{2\mathrm{D}}\{\cdot\}$ denotes 2D discrete Fourier transform, x_i is the image patch, $\bar{x}_i$ is the mean pixel value of the patch, $i = 1, \ldots, 40$. Finally, the 1D NPS was calculated by taking the rotational average of the 2D NPS and was plotted for comparison.

3.2.3.2 Contrast-to-noise Ratio and Full Width at Half Maximum

To quantitatively evaluate the MCs in the images, we calculated the contrast-to-noise ratio (CNR) and full width at half maximum (FWHM) as figures of merit for each MC. CNR indicates the conspicuity of MCs within the local surroundings, while FWHM measures their sharpness. Given a 32×32 patch with an MC at the center, we used a 2D Gaussian plus a 2D first order plane as the fitting function to fit the signal and the background in the central 13×13-pixel region. The background was thus removed when the height of the fitted Gaussian was taken as the contrast of

the MC. The shapes of MCs are normally spherical or elliptical (see Figures in the Results section), which can be well-fitted by 2D Gaussian functions. We define

$$\mathrm{CNR} = \frac{I_{\mathrm{MC}}}{\sigma_{\mathrm{bg}}},\ \mathrm{FWHM} = 2\sqrt{2\ln 2}\cdot\sigma_{\mathrm{MC}} \tag{3.10}$$

where I_{MC} is the maximum value of the 2D fitted Gaussian on the pixel grids, σ_{bg} is the root-mean-square noise of the surrounding area after removing the local background mean gray level using a box-rim filter [83] and excluding the MC pixels, σ_{MC} is the standard deviation of the fitted Gaussian. Figure 3.5 illustrates the process of the CNR and FWHM calculation.

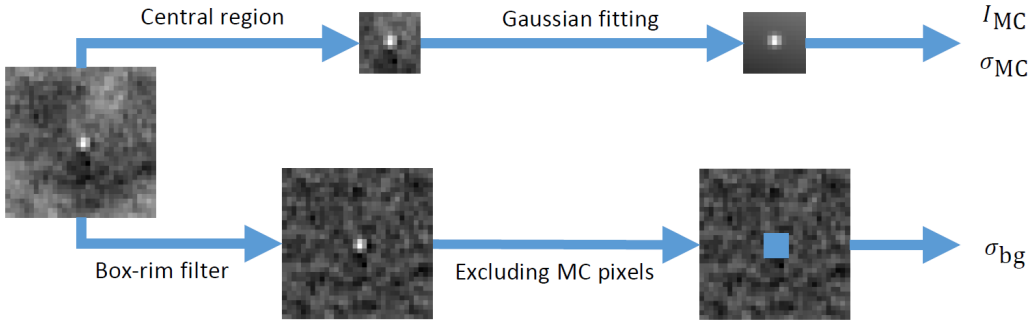


Figure 3.5 An illustration of CNR and FWHM calculation for an MC patch.

For comparison of the performance of the denoiser at different conditions, we calculated the mean and standard deviation of the CNR and the FWHM over the marked MCs at each speck size on the validation physical phantom. The denoiser inevitably smoothed out some subtle MCs as if they were noise. If an MC was very blurred, the fitting program would fit to the background. We chose three criteria experimentally to automatically mark the fitting as a failure: (1) if the FWHM was larger than twice of the nominal MC size, or (2) if the fitted Gaussian was off centered by two pixels, or (3) if the fitting error was larger than a threshold (if R-squared is less than 0.86). These failed MCs were excluded from the mean CNR or mean FWHM calculations, but the fit success rate was counted for each MC speck size and considered one of the indicators of the

denoiser performance. The value (1.0 – fit success rate) indicated the fraction of MCs that were blurred out under the given condition.

3.2.3.3 Detectability Index

We also calculated the task-based detectability index (*d'*) from the nonprewhitening model observer with eye filter (NPWE) as an image quality metric [84][85]. The NPWE observer performance was shown to correlate well with human observer performance [85]–[89]. In this study, we considered the task of detecting MCs of different nominal sizes in the heterogeneous background of breast phantom DBT.

We considered the 2D in-plane *d'* of the DBT slices in the validation physical phantom for each of the study conditions

$$d' = \frac{\iint S^2(u,v)\, E^2(u,v)\, du\, dv}{(\iint W(u,v)\, S^2(u,v)\, E^4(u,v)\, du\, dv)^{1/2}} \tag{3.11}$$

where u, v are spatial frequencies in 1/mm, S is the (blurred) signal spectrum (the product of the task function and the task transfer function), W is the 2D NPS, E is the eye filter or the visual response function of a human observer. Similar to the CNR calculation, we assumed the MC shape to be Gaussian. We calculated the average *d'* for each MC speck size group by using the averages of the fitted parameters, namely the contrast $\bar{I}_{\mathrm{MC}}$ and standard deviation $\bar{\sigma}_{\mathrm{MC}}$, obtained from the Gaussian fitting to the individual MCs. The signal spectrum was thus given by

$$\begin{gathered} s(x,y) = \bar{I}_{\mathrm{MC}} \cdot \exp\left(-\frac{x^2+y^2}{2\bar{\sigma}_{\mathrm{MC}}^2}\right) \\ \updownarrow \text{Fourier transform} \\ S(u,v) = \bar{I}_{\mathrm{MC}} \cdot 2\pi\bar{\sigma}_{\mathrm{MC}}^2 \cdot \exp\left(-2\pi^2\bar{\sigma}_{\mathrm{MC}}^2(u^2+v^2)\right). \end{gathered} \tag{3.12}$$

The 2D NPS was calculated by (3.9), which characterized the structured noise of the image background as well as other noise from the imaging chain. We used the theoretical model of the eye filter that was proposed by Kelly [90]

$$E(u,v) = (u^2 + v^2) \cdot \exp\left(-c\sqrt{u^2 + v^2}\right). \quad (3.13)$$

Considering that radiologists usually search for MCs in zoomed-in mode because of their small size, we set the viewing distance to 12.5 cm, which corresponds to 4 times higher magnification than the usual 50 cm viewing distance. Under this condition, the value of c was set to 1 so that the eye filter had its peak at 4 cycles/deg [84].

3.2.4 DCNN Training Setup

For the DNGAN training, we randomly initialized all the kernel weights. We set the mini-batch size to 512, and λ_D to 10 as suggested [57]. We set $\lambda_{\text{adv}} = 10^{-2}$. The denoiser and the discriminator were trained alternately, and the discriminator had 3 steps of updates for every step of the denoiser update. The discriminator and the denoiser both used Adam optimizer [30] and shared the same learning rate. The learning rate started with 10^{-3} and dropped by a factor of 0.8 for every 10 epochs. The learning rate started with 10^{-4} and dropped by a factor of 0.8 for every 50 epochs in the fine-tuning stage. We selected 300 epochs for stage one training and 1,000 epochs for fine-tuning. The training parameters, including λ_{adv}, batch size, learning rate, and the number of epochs, were chosen experimentally based on the training convergence and efficiency as discussed in Section 3.3.1 and Section 3.3.2. The DCNN model was implemented in Python 2.7 and TensorFlow 1.4.1. The training was run on one Nvidia GTX 1080 Ti GPU. The training time depended on the sizes of the training sets. For example, it took 36 hours for the 24mAs/target sets and 74 hours for the LD/HD$_{100\%}$ and LD/HD$_{\text{Pristina}}$ sets.

3.2.5 Comparison Method

We included a DBT MBIR algorithm developed in our laboratory that models the detector blur and correlated noise (DBCN) with an edge-preserving regularizer [36] for comparison. The parameters ($\beta = 70, \delta = 0.002/\text{mm}$, 10 iterations) that were chosen in [36] were used for reconstructing the DBTs from the GE prototype system. We adapted the DBCN to the Pristina system by adjusting β to 40.

3.3 Results

3.3.1 Effect of Tuning Parameter λ_{adv}

To demonstrate the effect of λ_{adv} in (3.1), we trained six denoisers using $\lambda_{\text{adv}} = 0$, 10^{-3}, 10^{-2}, 10^{-1}, 1, ∞ in the DNGAN. The condition $\lambda_{\text{adv}} = 0$ is equivalent to using the MSE loss only, and the condition $\lambda_{\text{adv}} = \infty$ is equivalent to using the adversarial loss only for DNGAN training. The training set was 24mAs/noiseless. We used the same random seeds for weight initialization and data batching for all conditions.

Figure 3.6(a) and Figure 3.6(b) shows the training MSE losses and the WD estimate $\hat{d}_G(D)$ defined in (3.5) versus training epochs. For small λ_{adv} (0 and 10^{-3}), even though they converged to low training MSE values, they had high WD estimates which means that the denoisers produced images that were perceptually dissimilar to the noiseless targets in the training. Note that the WD estimate $\hat{d}_G(D)$ could be increasing or decreasing versus training epochs because the adversarial training aimed at maximizing it over D and minimizing it over G. Figure 3.6(c) shows that the denoised validation digital phantoms of $\lambda_{\text{adv}} = 0$ and 10^{-3} had low NPS compared to the ground truth noiseless image. This is evidence that the images were overly smoothed and lost structural details.

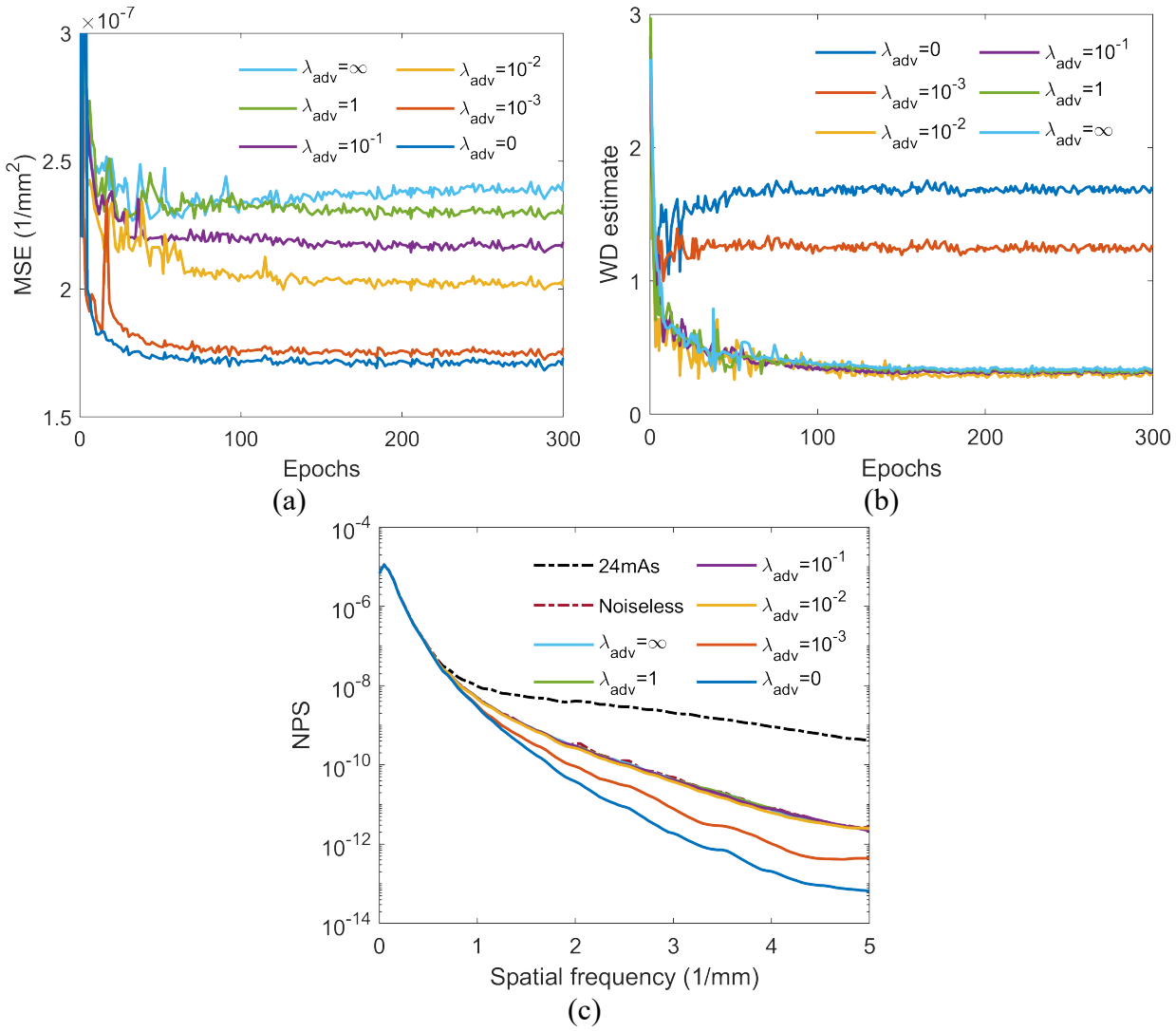


Figure 3.6 (a) The training MSE losses and (b) the WD estimates for different λ_{adv} versus training epochs. (c) The NPS curves of the denoised validation digital phantom volumes for different λ_{adv}.

For $\lambda_{adv} = 10^{-2}$, 10^{-1}, 1, ∞, although these conditions had similar WD estimates and NPS, the converged training MSE values monotonically increased as λ_{adv} increased in Figure 3.6(a). Moreover, as shown in Figure 3.7, for the MCs in the denoised validation physical phantoms, the FWHM increased and the fit success rate dropped substantially as λ_{adv} increased beyond 10^{-2}, indicating the blurring and loss of MC signals. Therefore, when the image smoothness was comparable, we preferred $\lambda_{adv} = 10^{-2}$ for a smaller training MSE and MC preservation.

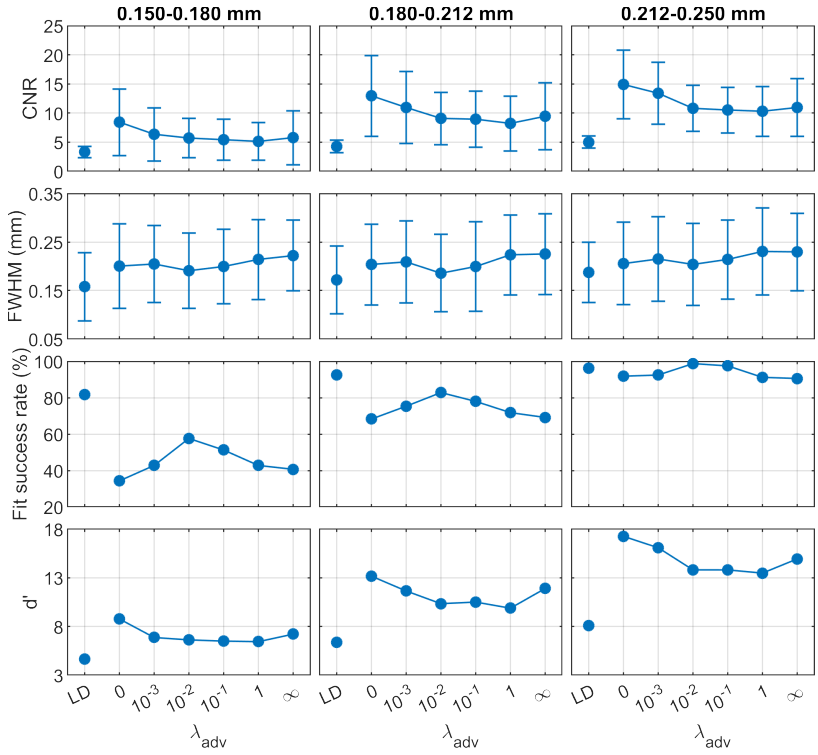


Figure 3.7 The CNR, FWHM, fit success rates, and *d'* of the MCs in the validation physical phantom for different λ_{adv}. The error bars represent one standard deviation.

3.3.2 Effects of Training Batch Size, Learning Rate, and Number of Epochs

We studied the effects of several other training parameters including batch size, learning rate, the number of epochs. The training set was the 24mAs/120mAs digital phantom set. If not specified, the rest of training parameters were the same as those described before. We used the same random seeds for weight initialization and data batching for all conditions.

To study the effect of training batch size, we trained four denoisers using the batch sizes of 1024, 512, 256, and 128 while keeping all other training parameters the same. Figure 3.8(a) shows the training losses (3.1) of the four conditions. For a small batch size such as 128, the training loss decreased more quickly in early epochs than others because the weights were updated more frequently. However, in later epochs, the gradient estimated from a small batch had larger variations than that estimated from a larger batch, so the training loss had a large oscillation. For

a large batch size such as 1024, the training occupied more GPU memory and also converged more slowly than the others. In our study, we used the batch size of 512.

To study the effect of learning rate, we trained four denoisers using the initial learning rates of 10^{-2}, 10^{-3}, 10^{-4}, and 10^{-5} while keeping all other training parameters the same. Figure 3.8(b) shows the training losses of the four conditions. For a large initial learning rate such as 10^{-2}, the training oscillated and did not converge. For a small initial learning rate such as 10^{-5}, the training step size was small, so the training might be trapped by local minima. In our study, we used the initial learning rate of 10^{-3}.

To study the effect of the number of epochs, we ran the denoiser training up to 500 epochs. We also prepared a validation set to monitor if overfitting occurred in our training. The validation set had 53,141 paired patches that were extracted from the LD/HD validation physical phantom pair in the same way as for the training set. Figure 3.8(c) shows the training and validation losses. We did not observe overfitting because both the training and validation losses were stable and the gap between them remained approximately constant after about 200 epochs. We determined that 300 epochs were sufficient to achieve training convergence. The stability and convergence of the training loss achieved within 300 epochs were also observed for other conditions. More importantly, the robustness of many of the trained denoisers was validated by their across-phantom performance in the independent validation set (i.e., digital-phantom-data-trained denoiser applied to physical phantom images) and further on the independent unseen test set of human subject DBT images, as discussed below.

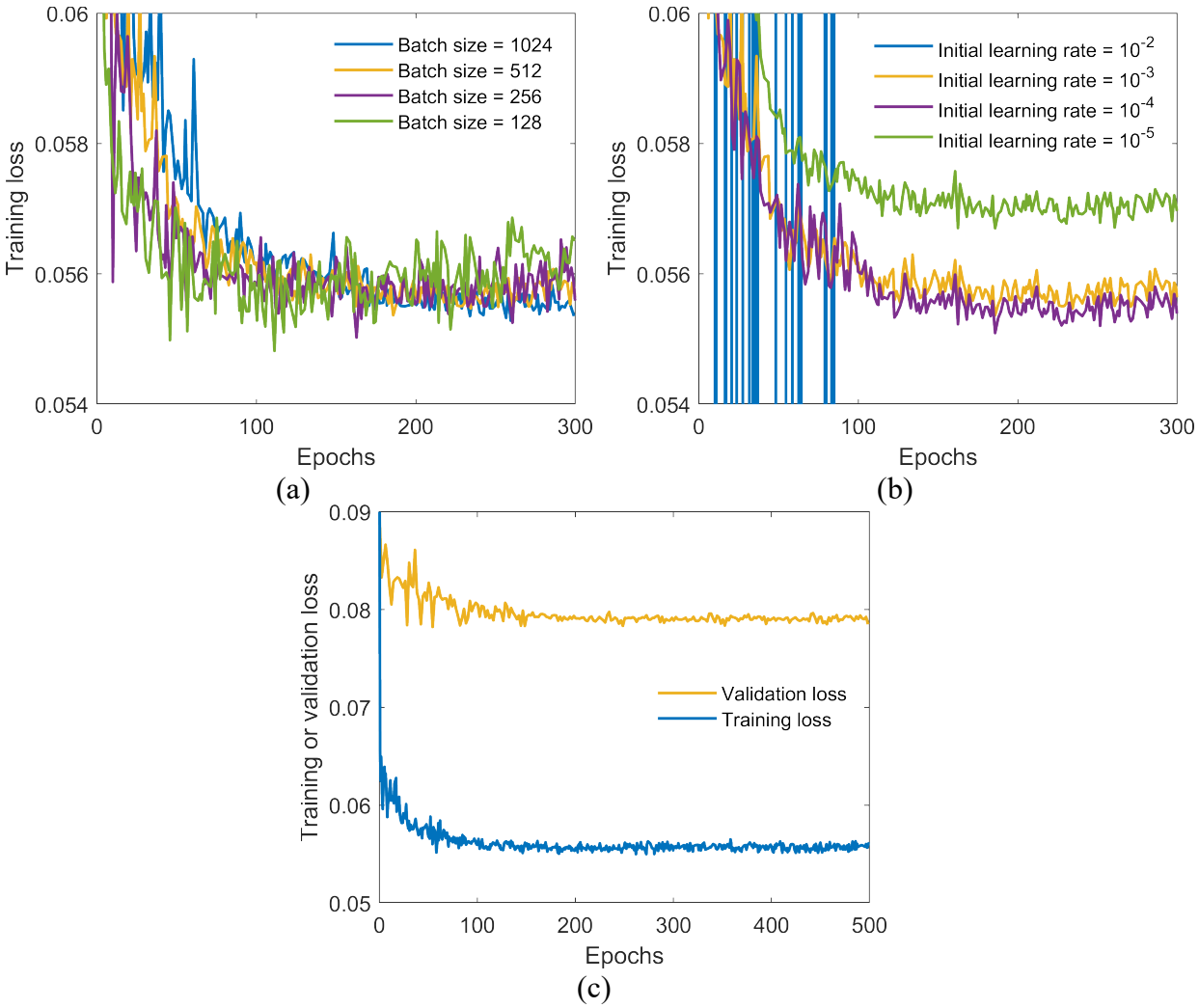


Figure 3.8 The training losses versus epochs for illustrating the effects of (a) the batch size, (b) the initial learning rate, and (c) the number of epochs.

3.3.3 Effect of Dose Level of Targets on DNGAN Training

To study the effect of the dose level of the training target images on the effectiveness of the denoiser, we trained the DNGAN using the 24mAs/72mAs, 24mAs/120mAs, 24mAs/360mAs, 24mAs/noiseless sets with $\lambda_{\text{adv}} = 10^{-2}$. We used the same random seeds for weight initialization and data batching for all conditions.

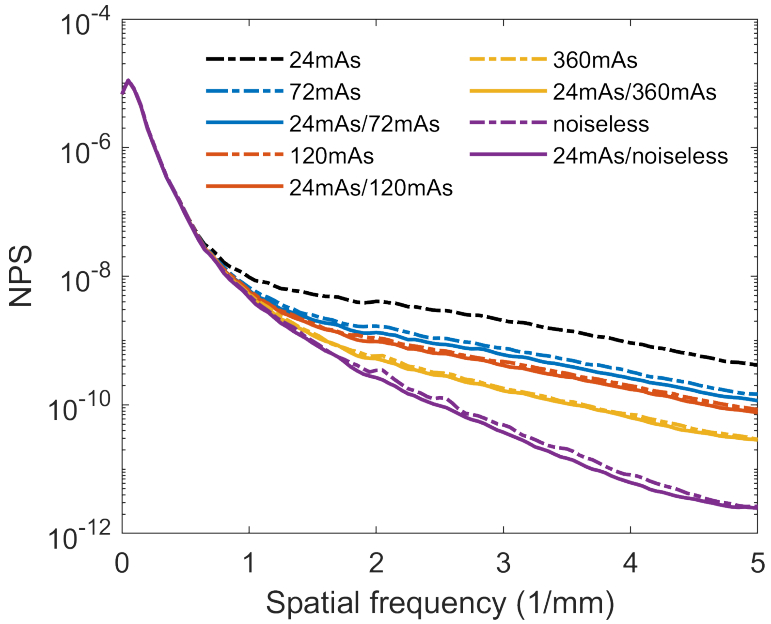


Figure 3.9 The NPS curves of the validation digital phantom volumes for comparing the dose levels of the training targets. The solid lines are calculated from the volumes by deploying the denoisers to the 24mAs volume. The dashed lines are from the CatSim simulated volumes with the corresponding dose levels.

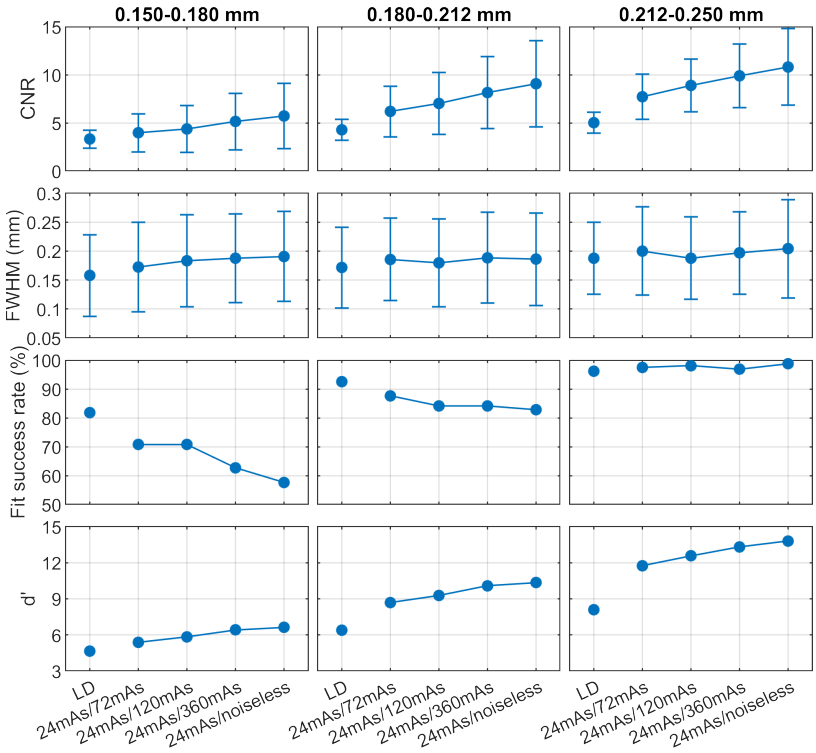


Figure 3.10 The CNR, FWHM, fit success rates, and *d'* of the MCs in the validation physical phantom for the different dose levels of the targets used in the DNGAN training.

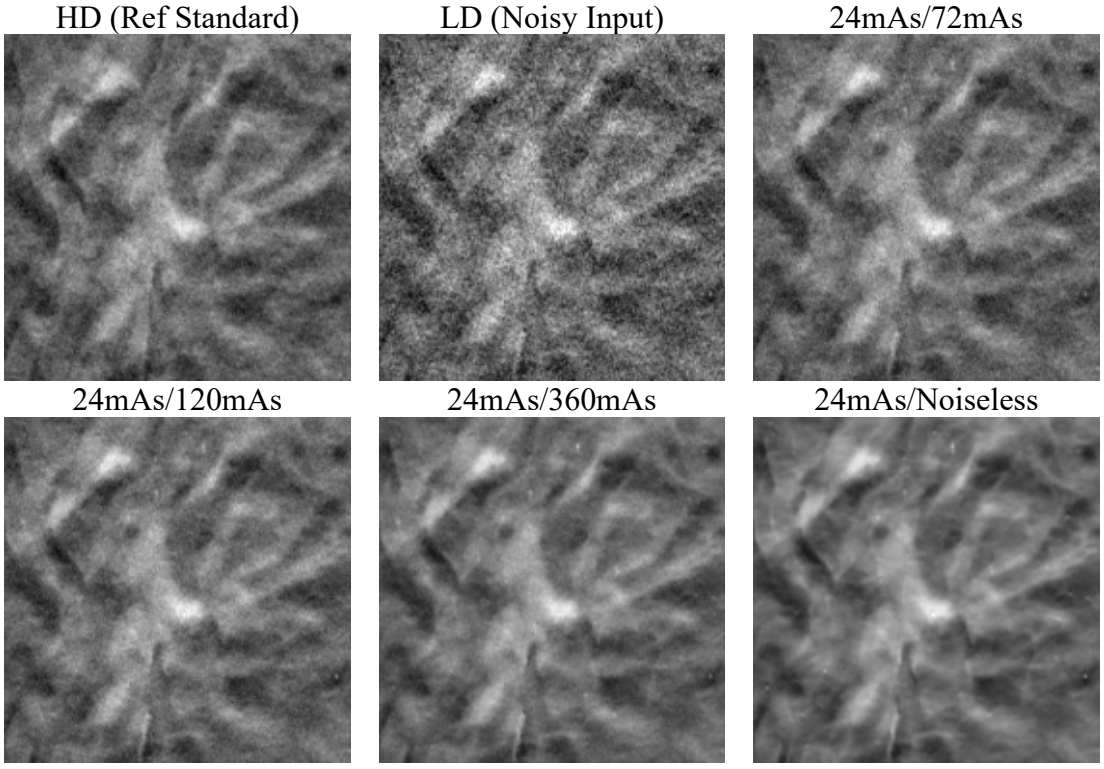


Figure 3.11 An example 18 mm × 18 mm region in the validation physical phantom for the different dose levels of the targets used in the DNGAN training. The images are displayed with the same window/level settings. The HD scan of the validation physical phantom (34 kVp, 125 mAs) is also shown for reference.

Figure 3.9 shows that the NPS of the denoised validation digital phantoms matched the NPS of the corresponding CatSim-simulated ground truth volumes. Figure 3.10 shows that the DNGAN achieved higher CNR and *d'* for the MCs in the validation physical phantoms when the training targets were acquired with a higher dose. However, as the target dose level increased, the fit success rate decreased for the two smaller MC groups. Figure 3.11 shows that if the target dose level was very high, for example, 360 mAs or infinity, the tissue backgrounds of the denoised validation physical phantoms images could be too smooth. The smoothing effect of the 24mAs/noiseless denoiser is further demonstrated in human subject DBTs in Section 3.3.7. We used the 24mAs/120mAs for training the DNGAN in the following studies since its dose ratio was closer to that of the LD/HD physical phantom images from a real scan.

3.3.4 Effect of MC Fine-tuning and Layer Freezing

We selected the DNGAN trained with 24mAs/120mAs in Section 3.3.3 as the initial model and fine-tuned it using the MC data set. We set the number of frozen layers of the denoiser to 8, 6, 4, 2, 0 while all layers in the discriminator were allowed to be fine-tuned under all conditions. We obtained five fine-tuned denoisers in addition to the base denoiser that was equivalent to freezing 10 layers.

Figure 3.12 shows that the fine-tuned denoisers improved the visibility of the subtle MCs in the denoised validation physical phantoms compared to the base denoiser. Figure 3.13 shows that the CNR and *d'* values increased and the FWHM values decreased as the number of frozen layers decreased, indicating that the MCs became brighter and sharper. The fit success rate also increased for the two smaller MC groups. However, the improvements leveled off when fewer than about 6 layers were frozen. In addition, as seen in the examples in Figure 3.12, the fine-tuning not only enhanced the subtle MCs but also some MC-like noise and background structures in the denoised images. The false MCs were obvious and distracting for all the fine-tuned denoisers even though freezing layers mitigated the problem to some extent. The fine-tuning was excluded from further discussions below because we concluded that it was unsuitable for practical use at this point.

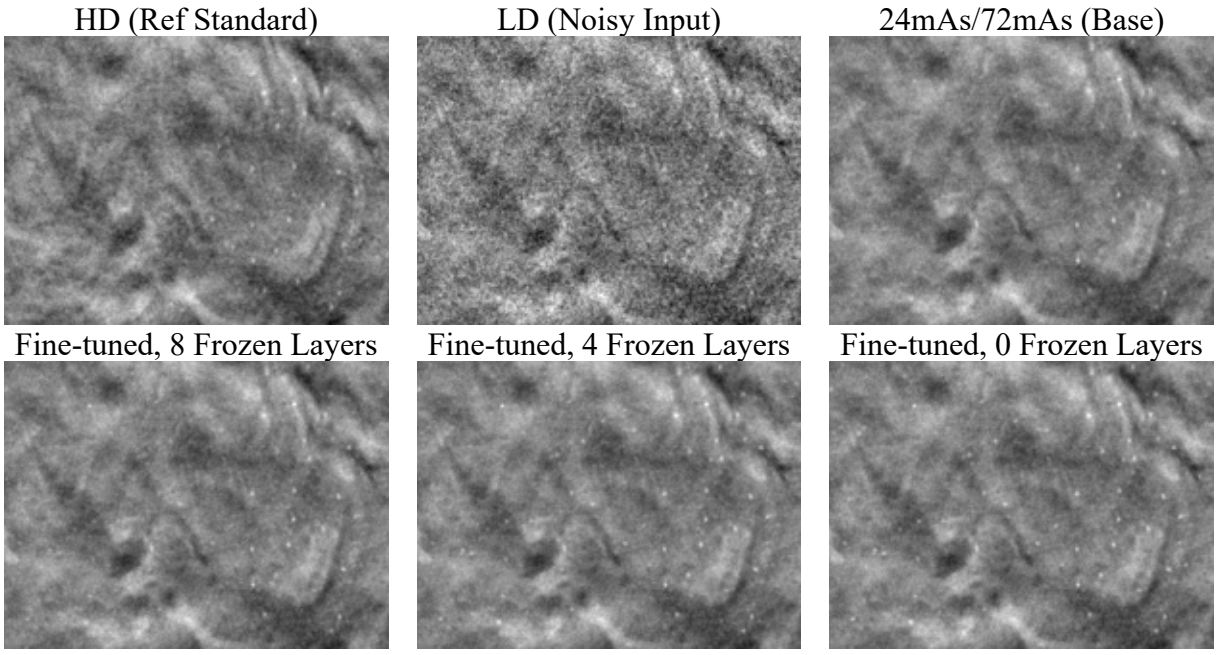


Figure 3.12 An example 20 mm × 15 mm region in the validation physical phantom showing the effect of fine-tuning and layer freezing. The region contains a background MC-free area on the left and a 0.180-0.212 mm MC cluster on the right. The images are displayed with the same window/level settings.

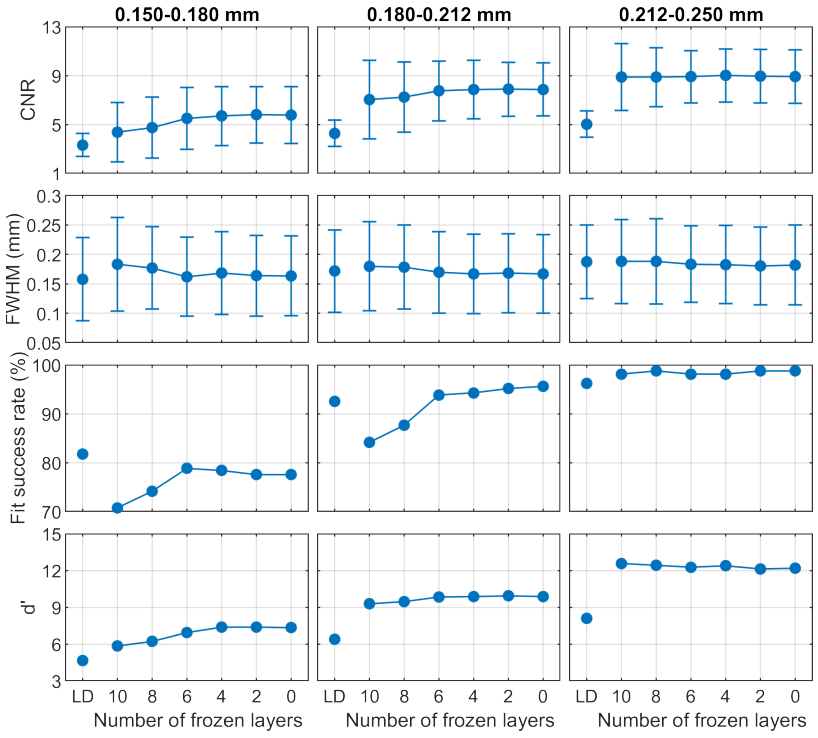


Figure 3.13 The CNR, FWHM, fit success rates, and *d'* of the MCs in the validation physical phantom showing the effect of fine-tuning and layer freezing.

3.3.5 Effect of Training Sample Sizes

We trained the DNGAN using the LD/HD$_k$ datasets from the physical phantoms. A different random seed was used for the weight initialization and data batching in each repeated experiment to account for the training randomness. After training, we deployed the denoiser to the validation physical phantom and calculated the mean CNRs for the MCs.

Figure 3.14 shows box plots using the 10 repeated experiments versus training data percentage. The general trend was that, when the training sample size increased, the training variation became smaller, and the median CNR increased and became stable. The CNR variations were large at 20% and 35%. This is especially undesirable for DBT because a denoiser with large performance variations can have unpredictable effect on subtle MCs. The large variation can be attributed mainly to the insufficient representation of the imaging characteristics by the small training set and the overfitting of the DCNN to each set of samples. The training randomness from weight initialization and data batching also contributed substantially to the variations, as can be seen from the 100% data point where the training set was the same for all repeated experiments.

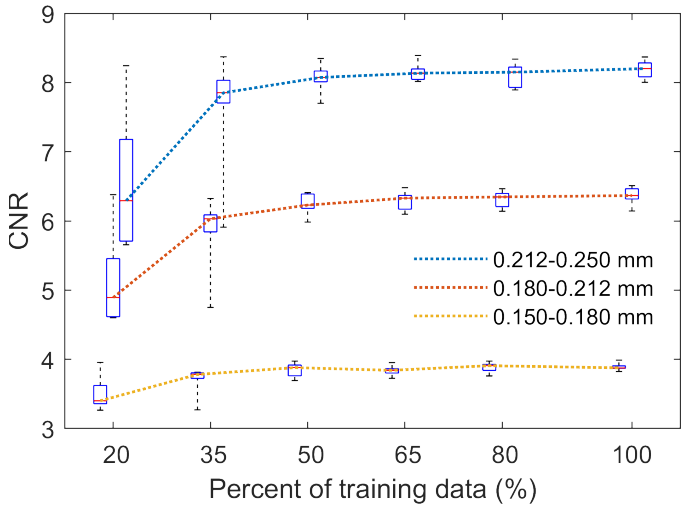


Figure 3.14 The box plots of the CNRs for different training sample sizes. Each box contains 10 data points. In the box plot, the red bar represents the median; the length of the box equals the interquartile range; the whiskers extend to the minimum and maximum data points. The boxes are slightly shifted horizontally to avoid overlap.

3.3.6 Denoising Performance on Validation Physical Phantom

We compared the DNGAN-denoised LD images where the DNGAN was trained with the digital phantom data (24mAs/120mAs) or the physical phantom data (LD/HD$_{100\%}$), and the LD images reconstructed from the DBCN algorithm. The LD/HD$_{100\%}$ model that was closest to the mean performance among the 10 repeated trainings in Section 3.3.5 was used in this comparison.

Figure 3.15 shows that the backgrounds in the 24mAs/120mAs and LD/HD$_{100\%}$ denoised validation physical phantoms were perceptually similar to the HD references, with the former being less noisy. Both denoisers improved the CNRs significantly ($p < 0.001$ for all three MC sizes, two-tailed paired t-test) compared to the LD images, as shown in Figure 3.16. Moreover, 24mAs/120mAs had significantly higher CNRs than LD/HD$_{100\%}$ ($p < 0.001$ for all three MC sizes) with the *d'* values showing the same trend. The reason may be that 24mAs/120mAs had a dose ratio of five, while LD/HD$_{100\%}$ had a dose ratio of four and contained scatter and detector noises. A denoiser trained with a higher dose ratio or a less noisy target produced a smoother background, thus larger CNR values. For 24mAs/120mAs and LD/HD$_{100\%}$, a few more percentages of MCs failed the Gaussian fitting than those in the LD images for the two smaller MC groups, indicating a greater loss of the relative subtle MCs, as also evident in Figure 3.15. Figure 3.16 shows that the CNRs of MCs in the DBCN images were comparable to those in the 24mAs/120mAs images but were sharper and had higher fit success rates. However, the backgrounds in the DBCN images in Figure 3.15 appeared patchier and were noisier than the DNGAN images, which might have contributed to the perceived noise or even bright pixels mimicking MCs. The high noise of DBCN led to lower *d'* than those of 24mAs/120mAs and LD/HD$_{100\%}$ for the two larger MC groups. The CNRs of MCs in the DBCN images were significantly higher than those in the LD images ($p < 0.001$ for all three MC sizes).

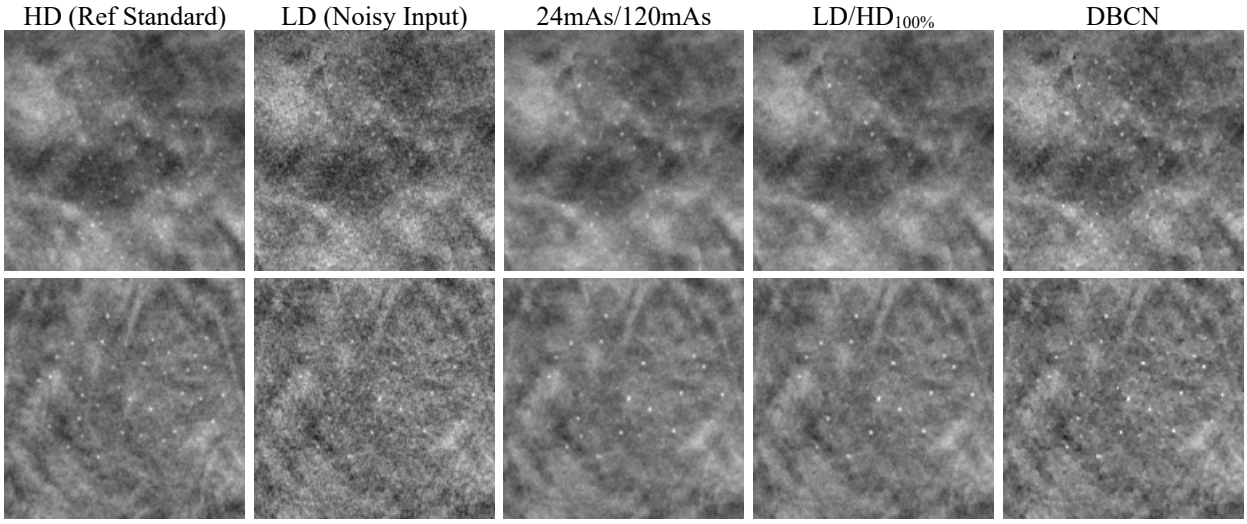


Figure 3.15 Example MC clusters in the validation physical phantom for comparing the denoising results. Top row: 0.150-0.180 mm cluster. Bottom row: 0.180-0.212 mm cluster. All images show a 15 mm × 15 mm region. The images in the same row are displayed with the same window/level settings.

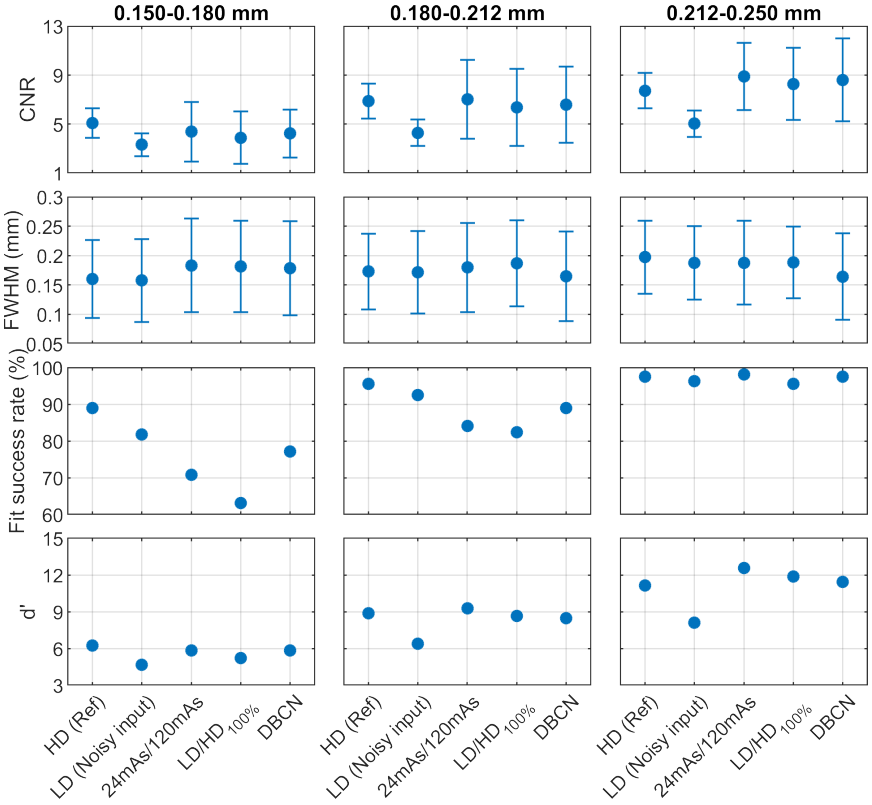


Figure 3.16 The CNR, FWHM, fit success rates, and *d'* of the MCs in the validation physical phantom for comparing the denoising results.

3.3.7 Denoising Performance on Human Subject DBTs

We deployed the DNGAN denoisers (24mAs/120mAs and 24mAs/noiseless) to the human subject DBTs for independent testing. Figure 3.17 shows that both denoisers and the DBCN were capable of reducing noise and maintaining the margins of the spiculated mass (invasive ductal carcinoma) and improving the conspicuity of the MC cluster (ductal carcinoma in situ). Although the background tissue of the 24mAs/noiseless denoised images was smooth as we discussed in Section 3.3.3, the spiculations were still well preserved. The DBCN images had a patchy and noisier breast parenchyma than the DNGAN denoised images.

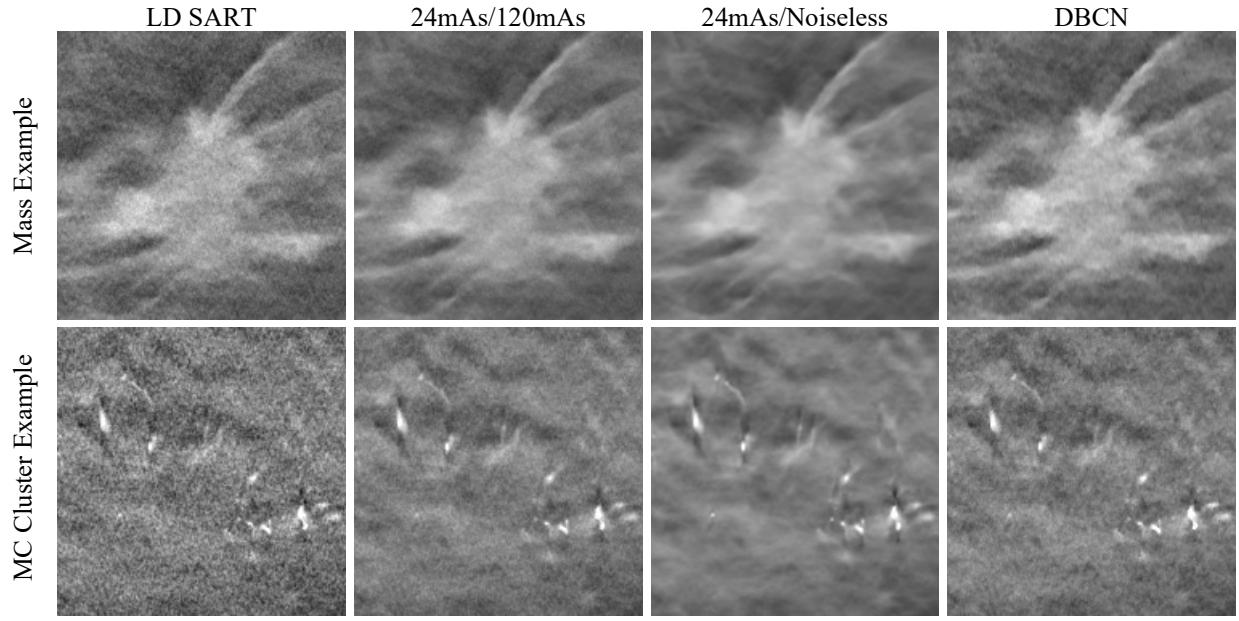


Figure 3.17 Example images of human subject DBTs with a spiculated mass (invasive ductal carcinoma) and an MC cluster (ductal carcinoma in situ). All images show an 18 mm × 18 mm region. The images in the same row are displayed with the same window/level settings.

Because the MCs in human subjects did not have nominal sizes, instead of comparing the *d'* or average CNR values, we generated the CNR scatter plot of individual MCs, as shown in Figure 3.18. The CNRs of most MCs were improved after DNGAN denoising. The CNRs of the 24mAs/120mAs denoised images were comparable to those of the DBCN images. The 24mAs/noiseless denoised images had the highest CNRs. However, whether the smooth

appearance of the breast parenchyma is acceptable to radiologists and whether it has any effect on diagnosis will warrant future investigations.

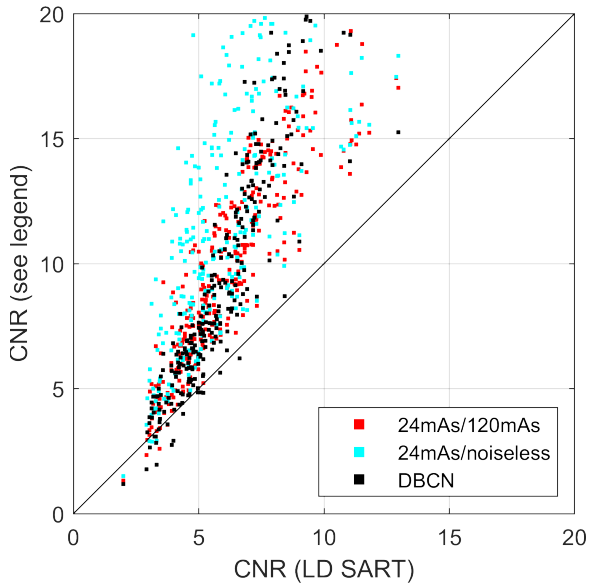


Figure 3.18 The CNR scatter plot of MCs in the human subject DBTs for the DNGAN denoised images and the DBCN reconstructed images versus the LD SART images.

3.3.8 Denoising Pristina-reconstructed Images Using SART Denoiser

To evaluate the generalizability of our DNGAN denoisers in terms of the reconstruction algorithms, we deployed the denoisers trained with SART-reconstructed images to the LD validation physical phantom image that was reconstructed by the Pristina algorithm. Specifically, we selected the 24mAs/120mAs and LD/$HD_{100\%}$ denoisers that were used in Section 3.3.6. We also trained a matched denoiser using the LD/$HD_{Pristina}$ set.

Figure 3.19 shows that the 24mAs/120mAs and LD/$HD_{100\%}$ denoisers worked to certain extent even though they were trained using SART-reconstructed images. The background texture of LD/$HD_{Pristina}$ denoised images was visually more similar to that of the HD Pristina reference, whereas the 24mAs/120mAs and LD/$HD_{100\%}$ denoisers produced smoother appearance. Figure 3.20 shows that all three denoisers reduced the noise and improved the CNRs significantly

($p < 0.001$ for all three MC sizes) compared to the LD Pristina-reconstructed images. LD/HD$_{Pristina}$ had a higher MC fit success rate and lower FWHM than the other two mismatched denoisers.

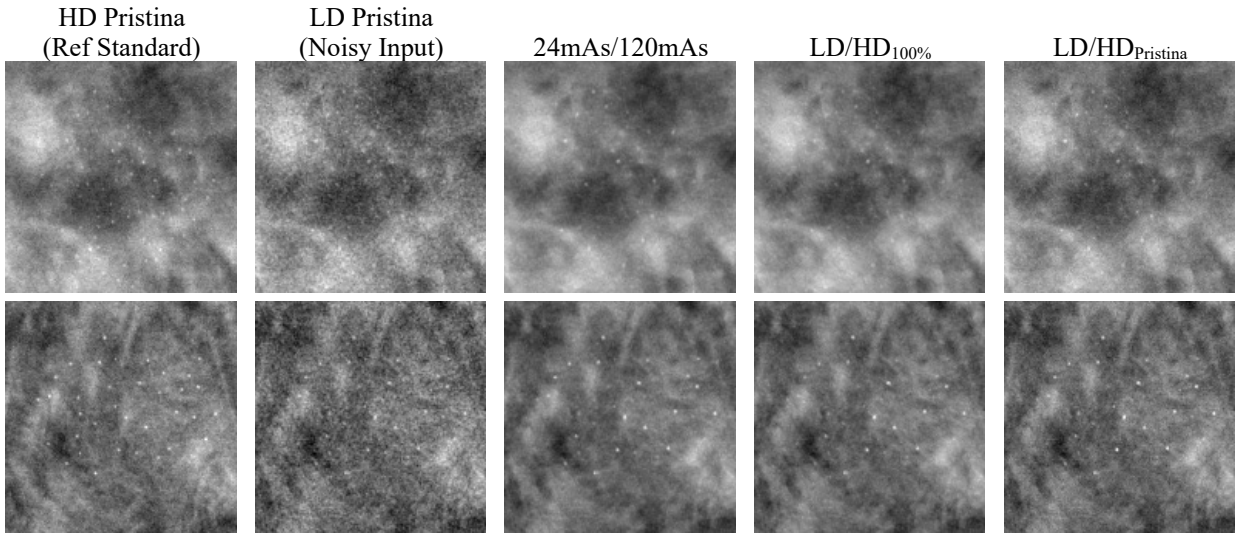


Figure 3.19 Example MC clusters in the Pristina-reconstructed images of the validation physical phantom for comparing the mismatched and matched denoisers. Top row: 0.150-0.180 mm cluster. Bottom row: 0.180-0.212 mm cluster. All images show a 15 mm × 15 mm region. The images in the same row are displayed with the same window/level settings.

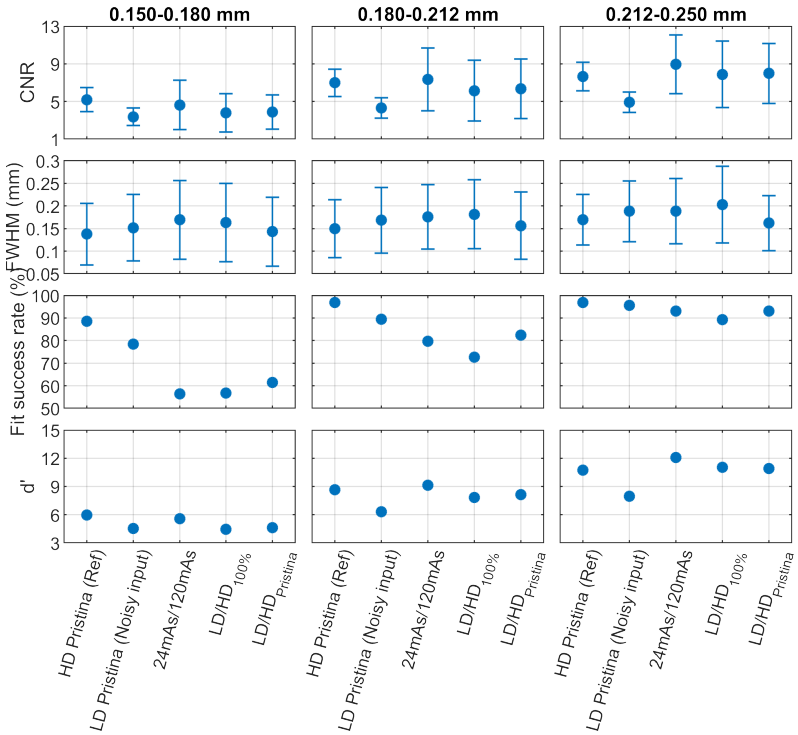


Figure 3.20 The CNR, FWHM, fit success rates, and *d'* of the MCs in the Pristina-reconstructed images of the validation physical phantom for comparing the mismatched and matched denoisers.

3.4 Discussion

The proposed DNGAN enjoys three aspects of robustness. First, the DNGAN trained with phantom data is applicable to human subject DBTs. This avoids the need to train using HD human DBTs, which may be impossible to collect. Second, the DNGAN can be trained with either digital phantom data or physical phantom data. This allows much flexibility in terms of the training data preparation. The digital phantom data has some advantages over the physical phantom data. For example, the software packages for producing the digital phantom data are open-source. It is inexpensive to generate a large set of data once the simulation model is formed, whereas making a large number of realistic physical phantoms is difficult. The high dose level of imaging a physical phantom is also limited by the tube loading of the DBT system. Third, the DNGAN trained with SART-reconstructed images is transferable to denoise other types of images such as Pristina-reconstructed images, although the denoising performance is not as good as that obtained with a denoiser trained with data from matched reconstruction. This makes training DNGAN with *in silico* data applicable to clinical DBT images for which the reconstruction algorithm is proprietary such as those used in the commercial systems.

For DBT denoising, it seems to be less strict for the training data to be statistically representative of the patient population to achieve generalizability, as we demonstrated that the denoisers trained with digital phantoms were quite effective for physical phantoms and, most importantly, human subject DBTs. One explanation is that the mapping function for DBT denoising is simpler than those for predicting diseases or other clinical tasks. Nevertheless, we only tested the denoiser on a small set of human subject DBTs, so follow-up studies with DBTs of a wide range of properties are needed.

DNGAN still smoothed out a substantial fraction of the very subtle MCs. We studied the feasibility of a second fine-tuning stage to improve the CNR and *d'* of subtle MCs, but the gain was offset by the increased spurious enhancement of noise and background structures (Section 3.3.4). Using our current fine-tuning approach within the DNGAN framework, we have not found a good training condition that could balance between MC enhancement and spurious noise suppression. Further investigations of the training framework to enable the denoiser to distinguish MCs more effectively from noise and selectively enhance the true MCs are warranted.

We compared the image quality obtained from the proposed DNGAN denoising and our DBCN reconstruction. The two approaches represent two different directions to enhance the subtle signals in DBT. The DBCN models the detector blur and correlated noise of the imaging system, which was simplified to essentially a high-frequency-boosting filter on the PVs. To control the high-frequency noise, the DBCN was implemented with an edge-preserving regularizer. However, the reconstructed image quality was sensitive to the choice of the parameters of the regularizer and improper parameters may cause patchy soft tissue texture as discussed by Zheng *et al*. [36]. In contrast, DNGAN smoothed the background around the signals to improve their conspicuity, similar to the role of a regularizer, but it also smoothed out some subtle MCs. Chapter 4 discusses our follow-up work that combines the DNGAN denoiser and the DBCN model into one reconstruction framework. Another noteworthy difference between the two methods is that DNGAN is a post-processing approach, whereas DBCN is a reconstruction algorithm. The DNGAN training is relatively flexible and, once fully trained, the DNGAN denoiser is readily deployable to the reconstructed DBT images and potentially applicable to DBT from different reconstruction techniques, as demonstrated in our study. In contrast, DBCN models a given DBT imaging system and requires raw PVs that may not be stored or accessible in clinical practice. Both

approaches have their advantages and disadvantages, and the choice will require future studies to compare the overall cancer detection accuracy and assess the preference of the image appearance by radiologists.

We observed a good correlation between CNR and *d'* that we calculated to assess the conspicuity of MCs. Figure 3.21 shows a scatter plot of the mean CNR and the corresponding *d'* from the results in Section 3.3. The Spearman rank correlation coefficient between CNR and *d'* was $\rho = 0.96$ and the correlation was statistically significant ($p < 0.001$). This observation suggests that the simple CNR might be a good surrogate for the more sophisticated *d'* as an image quality metric of MCs for the task in this study.

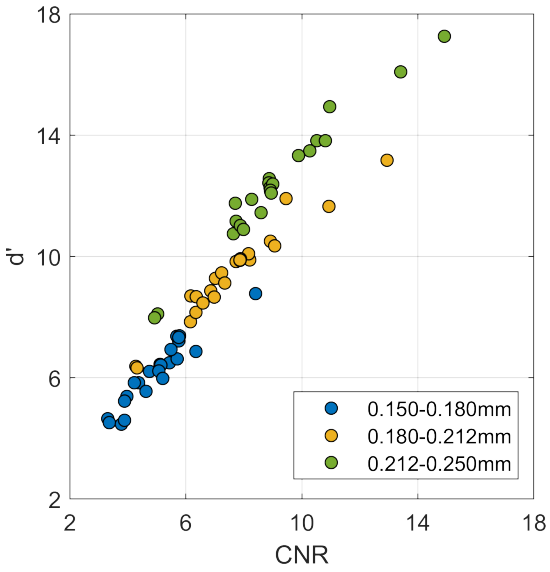


Figure 3.21 Scatter plot of *d'* versus CNR, including all data points from Figures 3.7, 3.10, 3.13, 3.16, and 3.20 for three nominal MC sizes and different conditions.

The figures of merit we used in this study have their limitations. First, CNR was used as an indicator for the conspicuity of individual MCs, but we only calculated CNR at known locations. The clinical usefulness of an image enhancement method has to consider both the true signals and the falsely enhanced noise or structures in an image. Alternative methods, such as computerized detection [91], can be used to study the tradeoff between the increase in detectability

of MCs and false positive detection in the future. Second, NPS and *d'* are Fourier-based, but the DCNN denoiser is nonlinear and DBT is shift-variant. Third, NPS provides a relative ranking of the noise level of the images, but it does not reflect the visual quality of the soft tissues or masses. To our knowledge, there is no figure of merit available to describe the fine textural appearance of an image or a soft-tissue lesion and correlate it with human visual preference. This makes it difficult to objectively optimize the balance between image smoothness and MC enhancement (Section 3.3.3 and Section 3.3.7). The acceptability of the image quality or image appearance for clinical reading will have to be judged by radiologists in human subject DBTs. Reader studies with radiologists can provide more clinically relevant assessments about the pros and cons of each condition but it is impractical to conduct reader studies for many conditions because of the limited availability of radiologists' time.

3.5 Conclusion

We developed a DNGAN framework based on adversarial training for denoising reconstructed DBT images. A properly weighted combination of an MSE training loss and an adversarial loss was found to be effective for noise reduction and texture preservation. We demonstrated the impacts of the dose level of the training targets and the training sample size on the performance of DBT denoising. We evaluated a fine-tuning stage to further enhance subtle MCs but found that it also enhanced false positives and was unsuitable for practical use. The DNGAN could be trained using *in silico* data and applied to physical phantom images even from a different reconstruction algorithm. Promising preliminary results were observed in deploying the trained denoiser to a small set of human subject DBTs. The DNGAN was also used in an observer study with radiologists and was shown to improve the detection of MCs in breast phantom images with increased conspicuity ratings and confidence levels [92]. Future work includes comparing

DNGAN with other classical and deep learning denoising methods in the literature, and investigating 3D DBT volume denoising as opposed to the current 2D slice-by-slice denoising.

Chapter 4
Model-based Deep CNN-Regularized Reconstruction for Digital Breast Tomosynthesis with a Task-based CNN Image Assessment Approach

4.1 Introduction

DBT[1] reconstruction is an under-determined and ill-posed inverse problem due to the limited-angle scan and incomplete sampling. Iterative reconstruction with or without regularization have been developed for DBT [27][34][41]. Model-based iterative reconstruction (MBIR) is a tomographic reconstruction approach that models the imaging physics and noise statistics and includes a regularization term as a prior of the unknown image [93][94]. MBIR has been applied to DBT and has shown good reconstruction quality. For example, Haneda *et al.* [95] used MBIR with regularization to improve the quality of spherical signals in a uniform background. Xu *et al.* [35] employed a Poisson likelihood function and a Gaussian Markov random field prior and improved the detectability of MCs. Zheng *et al.* [36] developed MBIR for DBT by incorporating detector blur and correlated noise (DBCN) modeling and an edge-preserving (EP) regularizer to improve the contrast-to-noise ratio (CNR) and sharpness of MCs.

Medical image denoising and restoration have made remarkable progresses by using deep convolutional neural networks (deep CNNs or DCNNs) [58][62][63]. This data-driven approach

[1] This chapter is based on the published journal paper [13] that extends the conference papers [14][15] and abstracts [16][17].

learns the image and noise features from training data for denoising. In the field of DBT, denoising projection views (PVs) before reconstruction was performed using DCNN trained with the mean squared error (MSE) loss [96], the combination of CNR, perceptual, and adversarial loss [65], generative adversarial network (GAN) [97], and conditional GAN [98]. In Chapter 3, we developed a denoising network for the reconstructed DBT images called DNGAN and improved the MC conspicuity in terms of the detectability index (*d'*) and human observer detection sensitivity [92]. Although straightforward and fast, image domain and PV domain denoising methods have limitations because they do not fully exploit measurement statistics and imaging physics.

There is a growing interest in combining DCNN with MBIR [99][100]. One idea is to unroll the iterative reconstruction loops and replace some steps in the iterations with networks [101]–[103]. Teuwen *et al.* [104] applied a learned primal-dual method to DBT reconstruction for breast density and dose estimation. Wu *et al.* [105] unrolled the proximal gradient descent algorithm for DBT reconstruction and achieved better in-depth resolution. Su *et al.* [106] proposed DIR-DBTnet and reduced artifacts. However, all these studies worked on low-resolution or downsampled 3D DBT images because the full-resolution images were too large to fit into memory. Another approach is to use pre-trained denoisers as priors in the reconstruction. Such frameworks include plug-and-play [107] and regularization by denoising (RED) [108]. These methods take advantage of the DCNN denoisers for denoising between data-consistency steps and allow one to train the denoisers separately to reduce computation compared to end-to-end training. This approach has shown promising results for MRI [109], CT [110][111], and PET [112]. In this study, we adopted the RED framework to combine DNGAN and DBCN modeling and proposed a new model-based DCNN-regularized reconstruction (MDR) method for DBT.

Task-based image quality assessment is an approach that evaluates the quality of medical images based on their specific utility for clinical tasks, such as tumor detection, lesion segmentation, or disease classification [113]. It aligns with the ultimate goal of medical imaging to support accurate diagnosis and treatment. Model observers have been developed as a surrogate for human observers to provide task-based image quality assessment in research and development stages where systematic and controlled evaluations are required [114]. Their usage offers several advantages over human observers such as consistency, objectiveness, and computational efficiency. They are also more informative than the generic metrics such as MSE or structural similarity (SSIM) which primarily focus on pixel-level differences between images without considering the specific diagnostic goals.

The clinical task of interest of this work is the detection of MC clusters in DBT. For signal-known-exactly detection tasks, channelized Hotelling observer (CHO) and non-prewhitening observer with eye filter (NPWE) are commonly used and have shown good correlation with human observers [88][115][116]. However, as a signal-known-statistically (SKS) task, MC clusters have various number of MCs, shapes and spatial distributions. Some studies designed CHO and NPWE for MCs but considered either a single spherical MC [117][118] or an artificial MC cluster with a fixed layout [119][120]. Recently, Zhang *et al.* [121] trained a DCNN to approximate the ideal observer for simulated MC clusters in synthetic mammograms. In the current study, we proposed a DCNN model observer called CNN-MC to evaluate the detectability of MC clusters in human subject DBTs, and a DCNN noise estimator called CNN-NE to evaluate the DBT noise levels. We used CNN-NE and CNN-MC as task-based image quality measures to guide the optimization of DBT reconstruction and to compare several reconstruction and denoising methods for DBT.

4.2 Methods and Materials

4.2.1 DBT Reconstruction

4.2.1.1 Background of MBIR and DBCN Reconstruction

Assume the post-log PV at the ith scan angle is $y_i \in \mathbb{R}^M$ where M is the number of detector pixels, $i = 1, \dots, N_p$, N_p is the number of scan angles, and the unknown DBT volume is $x \in \mathbb{R}^N$ where N is the number of voxels. As introduced in Section 2.3.3, MBIR formulates image reconstruction as an optimization problem with the following cost function:

$$\hat{x} = \underset{x}{\operatorname{argmin}}\, L(x) + \beta \cdot R(x) \tag{4.1}$$

where we define $L(x)$ as the negative log-likelihood (i.e., we absorb the negative sign into $L(x)$ compared to (2.10) to simplify notation), $R(x)$ is the regularization term, and β is a regularization parameter. If the measurement noise is additive Gaussian, $y_i = A_i x + \varepsilon_i$, $\varepsilon_i \sim \mathcal{N}(0, K_i)$, where $A_i \in \mathbb{R}^{M \times N}$ is the linear system matrix at the ith scan angle, $K_i \in \mathbb{R}^{M \times M}$ is the noise covariance matrix, then the data-fit term becomes an inverse-covariance weighted Euclidean norm:

$$L(x) = \frac{1}{2} \sum_{i=1}^{N_p} \|y_i - A_i x\|^2_{K_i^{-1}} \tag{4.2}$$

where $\|v\|^2_W = v'Wv$ and $'$ denotes matrix transpose. Usually, A_i does not model the detector blur caused by finite detector pixel size, crosstalk, or the light spread in the scintillator of an indirect detector. Zheng *et al.* [36] introduced a detector blur matrix $B \in \mathbb{R}^{M \times M}$ in front of A_i and modeled the resulting noise correlation in the covariance matrix K_i:

$$L_{\text{DBCN}}(x) = \frac{1}{2} \sum_{i=1}^{N_p} \|y_i - BA_i x\|^2_{\left(BK_{q,i}B' + K_r\right)^{-1}}, \tag{4.3}$$

where $K_{q,i} \in \mathbb{R}^{M\times M}$ is the diagonal quantum noise matrix of the ith scan angle, $K_r \in \mathbb{R}^{M\times M}$ is the diagonal detector readout noise matrix. In the actual implementation, Zheng *et al.* assumed the quantum noise variances and readout noise variances to be constant across all detector pixels, i.e., $K_{q,i} = \sigma_{q,i}^2 \cdot \boldsymbol{I}$ and $K_r = \sigma_r^2 \cdot \boldsymbol{I}$ where $\boldsymbol{I}$ denotes the identity matrix. Let $S_i = \left(BK_{q,i}B' + K_r\right)^{-1/2}$. Then $L_{\text{DBCN}}(x)$ has the following equivalent form using the prewhitened PV, $\tilde{y}_i$, and the DBCN system matrix $\tilde{A}_i$:

$$L_{\text{DBCN}}(x) = \frac{1}{2}\sum_{i=1}^{N_p}\left\|\tilde{y}_i - \tilde{A}_i x\right\|_2^2, \qquad \text{where } \tilde{y}_i = S_i y_i,\ \ \tilde{A}_i = S_i B A_i. \tag{4.4}$$

Here, S_i serves as a prewhitening matrix that boosts high frequency signals but also amplifies noise. Zheng *et al*. introduced an EP regularizer to control the noise level in DBCN reconstruction:

$$R_{\text{EP}}(x) = \frac{1}{1+\gamma}\sum_{j=1}^{N}\left(\eta\left([C_{\rightarrow}x]_j\right) + \eta\left([C_{\downarrow}x]_j\right) + \gamma\left(\eta\left([C_{\nearrow}x]_j\right) + \eta\left([C_{\searrow}x]_j\right)\right)\right) \tag{4.5}$$

where $C_{\rightarrow}, C_{\downarrow}, C_{\nearrow}, C_{\searrow} \in \mathbb{R}^{N\times N}$ are the finite differencing operators between neighboring pixels along the horizontal, vertical, and two diagonal directions in the DBT slices, γ is an adjustable weight in the diagonal directions, and $\eta(t) = \delta^2\left(\sqrt{1+(t/\delta)^2} - 1\right)$ is the hyperbola potential function.

4.2.1.2 DNGAN Denoising

This section briefly reviews the DNGAN introduced in Chapter 3. The DNGAN denoiser is a DCNN with trainable weights designed to denoise DBT images. DNGAN was trained with a supervised approach using pairs of noisy DBT image and target low-noise image of the same objects. In the training stage, the denoiser took pairs of image patches extracted from the two sets of DBT slices as input and target output and learned to produce the denoised image patch. The denoiser training loss was a weighted combination of the MSE loss and adversarial loss. The

adversarial loss was derived from the Wasserstein GAN with gradient penalty [57] where the denoiser acted as the generator. We implemented the discriminator in the GAN as a trainable VGG-Net [74]. After training, the DNGAN denoiser can be deployed to full DBT slices of any size since it is fully convolutional. The performance of DNGAN denoiser was validated with both phantom and human subject DBT [10].

4.2.1.3 Model-based DCNN-regularized Reconstruction (MDR)

We integrated the DNGAN denoiser into the iterative reconstruction loop in the general reconstruction optimization problem (4.1). RED defines a regularizer based on a general image filter or denoiser [108], which in this work is the trained DNGAN denoiser denoted as G:

$$R_{\mathrm{RED}}(x) = \frac{1}{2}x'\big(x - G(x)\big). \tag{4.6}$$

This regularizer promotes the cross-correlation between the residual after denoising and the image, or the residual itself, to be small.

By combining the DBCN data-fit term, the EP regularizer, and the RED regularizer with DNGAN denoiser, we formulated the overall optimization problem for the proposed model-based DCNN-regularized reconstruction (MDR):

$$\hat{x} = \underset{x}{\operatorname{argmin}}\, L_{\mathrm{DBCN}}(x) + \beta_{\mathrm{EP}} \cdot R_{\mathrm{EP}}(x) + \beta_{\mathrm{RED}} \cdot R_{\mathrm{RED}}(x) \tag{4.7}$$

where β_{EP} and β_{RED} are the regularization parameters.

A variety of optimization algorithms exist for inverse problems with $R_{\mathrm{RED}}(x)$ [108][122]. We used the RED proximal gradient method [122] to solve (4.7). It introduces a secondary image variable $z \in \mathbb{R}^N$ and updates the primary image variable x and the secondary image variable z alternately:

$$x_n = \underset{x}{\operatorname{argmin}}\, L_{\mathrm{DBCN}}(x) + \beta_{\mathrm{EP}} \cdot R_{\mathrm{EP}}(x) + \frac{\beta_{\mathrm{RED}}}{2} \|x - z_{n-1}\|^2 \tag{4.8}$$

$$z_n = G(x_n) \tag{4.9}$$

where $n = 1, \ldots, N_{\mathrm{iter}}$ is the iteration index. Both $L_{\mathrm{DBCN}}(x)$ and $R_{\mathrm{EP}}(x)$ are convex and differentiable in x. We used the diagonally preconditioned gradient descent with ordered subsets for the inner minimization problem (4.8):

$$x_n^{i+1} = x_n^i - \alpha \cdot P\left(N_p \cdot \tilde{A}_i{}'\left(\tilde{A}_i x_n^i - \tilde{y}_i\right) + \beta_{\mathrm{EP}} \cdot \nabla R_{\mathrm{EP}}\left(x_n^i\right) + \beta_{\mathrm{RED}} \cdot \left(x_n^i - z_{n-1}\right)\right) \tag{4.10}$$

where α is the step size, n is incremented after i goes through N_{inner} cycles of $1, \ldots, N_p$, $x_n^0 = x_{n-1}^{N_{\mathrm{inner}} N_p}$. We used the following preconditioning matrix P whose inverse majorizes the Hessian of the cost function (4.8):

$$P = \left(\left(\sum_{i=1}^{N_p} \frac{1}{\sigma_{q,i}^2 + \sigma_r^2} \operatorname{diag}\{A_i' A_i \mathbf{1}_N\}\right) + 8\beta_{\mathrm{EP}} \cdot \boldsymbol{I} + \beta_{\mathrm{RED}} \cdot \boldsymbol{I}\right)^{-1} \tag{4.11}$$

where $\mathbf{1}_N$ denotes the vector of ones of length N.

4.2.2 Task-based Image Quality Measures

4.2.2.1 CNN MC Classifier (CNN-MC)

Clustered MCs are one of the important signs of early breast cancer and image noise can negatively impact its diagnosis. The detectability of MC clusters and the image noise are therefore important indicators of image quality for DBT reconstruction. We developed an MC classifier as a model observer for the detection task of differentiating breast structured background with and without clustered MCs in human subject DBTs for image quality evaluation. Importantly, reconstruction or other image processing processes can inadvertently enhance artifacts mimicking calcifications in the normal tissue background. Image quality assessment by MC classifiers takes into account the false positives (FPs) while other measures such as the detectability index (*d'*)

focus only on the visibility of the target objects. The MC classifier was implemented as a DCNN with trainable weights, and therefore called CNN-MC. It took an image region of interest (ROI) or patch as input. The training set included MC patches as positives with label of 1 and MC-free background patches as negatives with label of 0. The training loss was the binary cross entropy loss, which has been shown to give the maximum likelihood estimation of the DCNN weights from the training data [123]. The CNN-MC output a score between 0 and 1 indicating the likelihood that an input patch contained clustered MCs. After training, the CNN-MC model with frozen weights was applied to the test patches.

We trained the CNN-MC for each image condition and tested it accordingly. The area under the receiver operating characteristic (ROC) curve (AUC) of the test set scores was used as an MC detectability measure. The ROC curve characterizes the tradeoff between sensitivity and (1 – specificity) of a classifier across the entire range of decision thresholds. AUC is a widely used metric in medical imaging analysis that condenses the classifier performance into a single value, simplifying the comparison of different models or algorithms. It represents the classifier's average sensitivity over the range of specificity, or the average specificity over the range of sensitivity. If the underlying image condition enhanced the MCs effectively, the CNN-MC should learn the MC features more accurately during training and have better classification ability between image patches with and without MCs during testing, resulting in a higher AUC metric. Any MC-like artifacts in the background patches would increase false positive detection and degrade the classification performance of CNN-MC, thereby reducing the AUC. We studied the use of CNN-MC to rank the relative performances of different DBT reconstruction and denoising methods.

There are variations in the CNN-MC observer modeling. DCNN classifiers with different network structures learn image features differently. Even for the same network structure, the

randomness in DCNN training such as kernel initialization or data batching can lead to different local minima. These are akin to the interobserver and intraobserver variabilities of human observers. To account for these variations, we investigated the VGG-Net [74], the ResNet [124] and the ConvNeXt [125] of similar sizes, as shown in Table 4.1, as the backbone structures of CNN-MC. We also repeated the training multiple times with different random initialization for each structure. The image condition rankings from the individual models and the final combined rankings were analyzed.

Table 4.1 Network structures of CNN-MC and CNN-NE.

	VGG-Net	**ResNet**	**ConvNeXt**
Stem	/	$1 \times 1, 32$; BN; ReLU	$1 \times 1, 32$; LN
Block 1	$3 \times 3, 32$; ReLU	$\begin{bmatrix} 3 \times 3, 32 \\ 3 \times 3, 32 \end{bmatrix}$	$\begin{bmatrix} \mathrm{d}7 \times 7, 32 \\ 1 \times 1, 128 \\ 1 \times 1, 32 \end{bmatrix}$
Downsample 1	2×2 Max Pool, s2	/	LN; $2 \times 2, 64$, s2
Block 2	$3 \times 3, 64$; ReLU $3 \times 3, 64$; ReLU	$\begin{bmatrix} 3 \times 3, 64, \mathrm{s}2 \\ 3 \times 3, 64 \end{bmatrix}$	$\begin{bmatrix} \mathrm{d}7 \times 7, 64 \\ 1 \times 1, 256 \\ 1 \times 1, 64 \end{bmatrix}$
Downsample 2	2×2 Max Pool, s2	/	LN; $2 \times 2, 128$, s2
Block 3	$3 \times 3, 128$; ReLU $3 \times 3, 128$; ReLU	$\begin{bmatrix} 3 \times 3, 128, \mathrm{s}2 \\ 3 \times 3, 128 \end{bmatrix}$	$\begin{bmatrix} \mathrm{d}7 \times 7, 128 \\ 1 \times 1, 512 \\ 1 \times 1, 128 \end{bmatrix}$
Head	Global Avg Pool, Fully Connected Layer, Softmax		
No. of parameters	2.8×10^5	3.1×10^5	2.3×10^5

* The input to the models is a 128×128-pixel image patch. BN: batch normalization. LN: layer normalization. Convolutional layers are specified by the kernel size and number of filters. "d" denotes the depth-wise convolution. "s2" denotes the stride-2 operation. The brackets denote ResNet block or ConvNeXt block.

4.2.2.2 CNN Noise Estimator (CNN-NE)

It is important to assess the image noise for an image processing technique because enhancing the signals is always associated with a change in noise. We developed a DCNN noise estimator (CNN-NE) to quantify the root-mean-square (RMS) noise level of its input image patch. We designed their network structures to be the same as those for CNN-MC (Table 4.1), except for without the exit softmax function, to facilitate transfer learning (discussed next). The training set contained DBT image patches of breast structured background with the training labels calculated by an ROI-based method for each patch as follows: the patch was divided into 10×10-pixel ROIs

with 25-pixel spacing on a grid, the ROI background trend was removed by a quadratic fitting, the RMS variations of the ROI pixel values were then calculated and averaged over all ROIs as the RMS noise. The CNN-NE was implemented as a regression model. The training loss was the MSE loss between the estimated RMS noise and the training labels. After training, the CNN-NE model was applied to the test set patches. The average of the estimated RMS noise over all patches was used as a noise measure of the entire set.

4.2.2.3 Transfer Learning

For training CNN-NE and CNN-MC, we adopted the training technique of transfer learning [75][80]. The CNN-NE required only breast structured background images with a range of noise levels so that a large training set could be obtained relatively easily. The trained CNN-NE then served as the pre-trained model to be further fine-tuned as CNN-MC by transfer learning. Since the availability of MC-positive samples was limited, transfer learning reduced the training sample size required and improved the model robustness. It also reused the weights that the model learned from the source task of noise estimation to encourage the CNN-MC to focus on the background noise patterns in the downstream task of MC detection.

4.2.3 Data Sets

We modeled three DCNNs: DNGAN, CNN-NE and CNN-MC, for denoising, noise estimation, and MC classification, respectively, in this work. To facilitate the training, validation, and testing of these DCNNs, we prepared two primary DBT sources: virtual phantom DBTs and human subject DBTs. This section first provides an overview of these data sources and how they were used for the data sets, as summarized in Table 4.2, and then delves into the details in the following subsections.

In Chapter 3, we performed training and validation of DNGAN using virtual phantom DBTs and demonstrated its transferability to human subject DBTs. In the current work, we used the virtual phantom DBTs for the DNGAN training set. This choice was motivated by their flexibility and controllability in generating a wide range of image noise levels with different breast densities and thicknesses. Similar to Chapter 3, we chose a small training patch size of 32×32 pixels for DNGAN, allowing it to concentrate on local image structures during adversarial training [73]. The denoiser was fully convolutional at deployment so that the training patch size would not affect the image sizes for which it could be used.

The training set, validation set and independent test set of CNN-MC were prepared using human subject DBTs because we were interested in evaluating the detectability of MC clusters in real patient data and it is difficult to simulate the wide variations in features of real MC clusters by virtual phantom software. The input patch size of CNN-MC was 128×128 pixels. This patch size was chosen to cover typical MC clusters in patient cases while keeping the processing time modest.

The CNN-NE training set was extracted from the virtual phantom DBTs that had a wide noise range, similar to the motivation of the DNGAN training set. For the CNN-NE validation and test sets, we had to use the same human subject DBTs as those for CNN-MC. This enabled us to calculate the RMS noise for each reconstruction, which complemented the AUC metric for MC detectability by CNN-MC. These two CNNs therefore provided the two metrics for the task-based image quality assessment plot, as detailed in the Results section below. The input patch size of CNN-NE was the same as CNN-MC because the two networks shared the same structures for transfer learning.

In addition, Chan *et al.* [92] conducted an observer study with radiologists using a set of physical phantom DBTs with simulated MC clusters to demonstrate the improvement of MC

detection by DNGAN denoising. In the current study, we reused their data set to test the capability of CNN-MC in ranking MC detectability relative to human readers. We calculated the relative rankings of the CNN-MC AUC for the MC clusters in the physical phantom DBTs, and then compared them with the relative performance of radiologists' reading under the image conditions of the observer study.

Table 4.2 Summary of data sets for DNGAN, CNN-NE and CNN-MC in MDR study.

DCNN Name	Purpose	DBT Source	Patch Size (pixels)	No. of Patches	Comments
DNGAN	Training/ Validation	Virtual Phantoms	32 × 32	Avg: 449,245 Min: 343,498 Max: 547,595	12 groups of training/validation data for different noise levels.
	Test	Human Subjects	/	/	Applied to full DBT slices for denoising and reconstruction.
CNN-NE	Training	Virtual Phantoms	128 × 128	256,194	Quantitative RMS noise patch label for training regression model.
	Validation	Human Subjects		1,955	From the same locations as the negative CNN-MC patches without MIP.
	Test			6,176	
CNN-MC	Training	Human Subjects	128 × 128 (MIP over 3 slices of 128 × 128 × 3)	Pos: 751 Neg: 19,079	Positive patches augmented 8 times by flipping and rotation during training to approximate the sample ratio of the validation and test sets.
	Validation			Pos: 709 Neg: 1,955	Validation set also used for tuning the MDR parameters.
	Test			Pos: 2,289 Neg: 6,176	Independent test set.
	Test	Physical Phantoms	128 × 128 (MIP over 3 slices of 128 × 128 × 3)	For each image set and MC size: Pos: 36 Neg: 75	To compare CNN-MC rankings with radiologists' performance of observer study. - Three image sets: STD, dnSTD, STD+ - Four speck sizes of simulated MC clusters

* Pixel size of all patches is 0.1 mm × 0.1 mm.

4.2.3.1 Virtual Phantom DBTs

We generated a collection of virtual breast phantoms at a voxel resolution of 0.05 mm using the anthropomorphic breast model from the Virtual Imaging Clinical Trial for Regulatory Evaluation (VICTRE) package [66]. The VICTRE project considered four breast density categories and one compressed thickness for each category. They stated that this design mirrored the cohort demographics of their comparative clinical trial. In our study, we used the same breast density setting as VICTRE, and slightly varied the thicknesses to make it more varied like human data. In particular, a total of 70 phantoms were simulated at a range of glandular volume fractions

(GVF), including 10 almost entirely fatty (5% GVF), 10 scattered fibroglandular dense (15% GVF), 25 heterogeneously dense (34% GVF), and 25 extremely dense (60% GVF). The thicknesses of the compressed phantoms for the four density categories were 52-70 mm at every 2 mm, 46-64 mm at every 2 mm, 36-60 mm at every 1 mm, and 31-55 mm at every 1 mm, respectively.

We generated the PVs using the Monte Carlo x-ray imaging simulator [126]. We configured the scan geometry as 9 PVs in 25° scan angle and 3.125° increments and the x-ray spectrum as 34 kVp Rh/Ag to model the Pristina DBT system (GE Healthcare, Waukesha, WI). The distances between source and isocenter, isocenter and breast support, breast support and detector were 617 mm, 20 mm, 23 mm, respectively. Scatter and electronic noise were not simulated. The exposure was adjusted for each phantom so that the estimated mean glandular dose matched the measured value for that breast thickness from the automatic exposure control (AEC) of Pristina [79]. To simulate a wide range of noise levels, we repeated the simulations and varied the exposure by factors of 0.25, 0.3, 0.35, 0.4, 0.5, 0.6, 0.75, 1, 1.4, 2, 3. These exposure factors were selected such that the noise standard deviations of the post-log x-ray intensity on the PVs, and hence the reconstructed DBT images, were evenly spaced. Finally, we reconstructed the DBT images using 3 iterations of the simultaneous algebraic reconstruction technique (SART) [27].

4.2.3.2 DNGAN Training Set

For DNGAN training, we used the virtual phantom DBTs with an exposure factor of 3 as the training high dose (HD) targets and the DBTs with the remaining 10 exposure levels as the training low dose (LD) inputs. To ensure that DNGAN covered a wide range of DBT noise levels due to variations in patient sizes and other factors in clinical settings, we trained a suite of DNGAN denoisers, each of which can handle a certain range of image noise levels. To do so, we calculated

the average RMS noise for each DBT volume using the method described in Section 4.2.2.2 and obtained the range of the noise values for the entire set of virtual DBT volumes. We empirically grouped the volumes into 12 groups so that the noise interval of each group was reasonably small. We then extracted training patches within the breast regions from each group. This resulted in 12 groups of DNGAN training data with an average of 449,245 (min: 343,498; max: 547,595) LD/HD pairs of 32×32-pixel training patches for training a denoiser for each group.

4.2.3.3 CNN-NE Training Set

We randomly extracted 128×128-pixel patches within the breast regions from all the 70 virtual phantom DBTs and 11 exposure levels to form the CNN-NE training set. This gave a total of 256,194 patches. The low-frequency background was removed from the patches to reduce the nonuniformity caused by heterogeneous breast structures [127].

4.2.3.4 Human Subject DBTs

We had 238 human subject cases with biopsy-proven MCs collected with IRB approval and written informed consent. Two-viewed DBTs of the breast with MCs were acquired for each case using the GEN2 prototype DBT system (GE Global Research, Niskayuna, NY). The system acquired 21 PVs in 60° scan angle and 3° increments with a 29 kVp Rh/Rh x-ray beam. The distances between source and isocenter and between isocenter and detector were 640 mm and 20 mm, respectively. The breast support was at the same height as the isocenter. For our reconstructions, we used the central 9 PVs in 24° angle so that the scan geometry was similar to that of the GE commercial Pristina DBT system. The radiation doses at 9 PVs were similar to those of the Pristina system. A Mammography Quality Standards Act (MQSA) approved radiologist marked the biopsied MCs with 3D bounding boxes in the reconstructed DBT images based on all available clinical information.

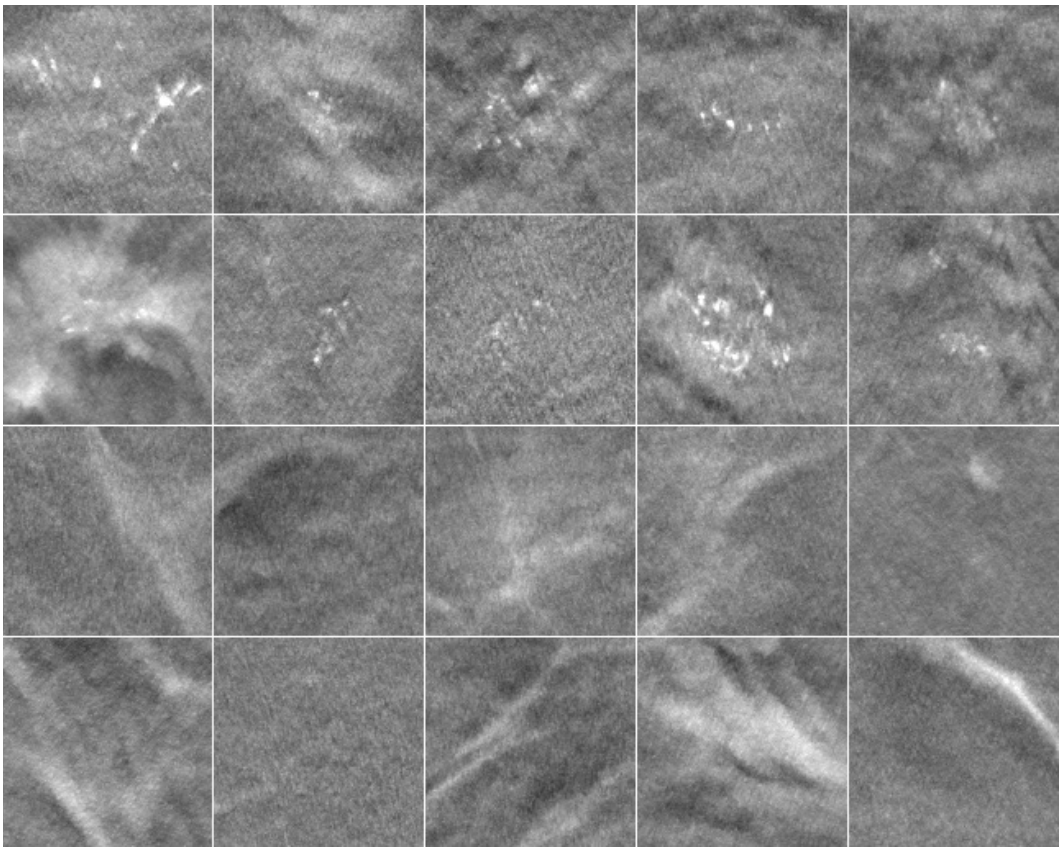


Figure 4.1 Example 3-slice MIP patches with MCs (top two rows) and without MCs (bottom two rows) from the CNN-MC training set of human subject DBTs. The patch size is 128×128 pixels (12.8 mm × 12.8 mm).

4.2.3.5 CNN-MC Training Set and CNN-NE/CNN-MC Validation and Test Set

We split the human subject DBTs by case into three disjoint sets for training, validation and testing. The training set consisted of 64 cases, or 127 DBT views (one view was lost due to technical issue). The validation set consisted of 52 cases, or 104 DBT views. The remaining 122 cases with 246 views (one case had bilateral MCs) were sequestered as independent test set. For each data set, positive patches of size 128×128×3 pixels containing clustered MCs were extracted inside the radiologist's 3D boxes on a regular grid with centers separated by 128 pixels in the two directions along a slice and separated by 1 slice with offset centers along the depth direction. An MC cluster therefore might be extracted multiple times in part or in whole but spatially shifted in each patch, which served as augmented samples to reduce the imbalance between the two classes.

Non-overlapping negative patches of the same size were randomly extracted outside the 3D boxes in the structured breast background. For the CNN-MC, we obtained 751 positive and 19,079 negative patches for the training set, 709 positive and 1,955 negative patches for the validation set, and 2,289 positive and 6,176 negative patches for the test set. For each patch, we removed the low-frequency background and took maximum intensity projection (MIP) over the three slices to emphasize the MCs, if any. Figure 4.1 shows example MIP patches from the CNN-MC training set. For the validation and test sets of CNN-NE, the same sets of negative patches for CNN-MC were extracted except that only the central slices without MIP were used.

4.2.3.6 Physical Phantom DBTs

The physical phantom DBTs were used in a previous observer study by Chan *et al.* to detect MC clusters under three image conditions with radiologists [92]. Each of the six breast phantoms were 5-cm thick and made of 50% glandular/50% adipose heterogeneous breast-tissue-equivalent materials. Clusters of simulated MCs of four nominal speck diameters (0.125-0.150 mm, 0.150-0.180 mm, 0.180-0.212 mm, 0.212-0.250 mm) were embedded in the phantoms. There were 36 clusters for each diameter range, giving a total of 144 clusters. The phantoms were imaged twice using two automatic exposure modes of the Pristina DBT system: the standard (STD) mode, which was for routine patient imaging, and the STD+ mode, which used about 54% more dose than that of STD. The DBT images were reconstructed by the built-in commercial algorithm. A third image set called dnSTD was obtained by denoising the STD set using the LD/HD$_{\text{Pristina}}$ DNGAN denoiser developed in Chapter 3. The denoiser was trained and validated using a separate set of physical phantoms that shared the same materials but had different designs from the test phantoms in the Chan *et al*. study. For each image set and each MC speck size, we extracted a total of 36 128×128-pixel background-corrected 3-slice MIP patches centered at the MC clusters as positives and paired

them with 75 MC-free background MIP patches as negatives for CNN-MC testing. The set of 75 negative patches was extracted from different random locations for each MC positive set.

4.2.4 Implementation

For MDR, the implementation of $L_{\mathrm{DBCN}}(x)$ and $R_{\mathrm{EP}}(x)$ was described in Zheng *et al.* [36]. The 12 DNGAN denoisers were trained separately following the process described in Chapter 3. We retrained DNGAN using our newly generated training set that had a wider range of noise levels and breast characteristics while keeping all training settings unchanged. When DNGAN was called during reconstruction, the algorithm automatically estimated the RMS noise of the DBT volume at that iteration and selected the denoiser with the closest-matched noise from the 12 trained denoisers for deployment. We chose N_{iter} and N_{inner} to 3 experimentally. Section 4.3.2 discusses the selections of β_{EP} and β_{RED}. Both x and z were initialized to 0. The image variable z was saved as the final reconstructed image. All the PVs had a pixel size of 0.1 mm × 0.1 mm. The reconstructed images had a voxel size of 0.1 mm × 0.1 mm × 1 mm and were saved in DICOM format with a pixel value range of $[0, 4095]$. Metal artifact reduction [128] and truncated projection artifact reduction [129] were implemented to minimize these artifacts in the reconstructed DBT volumes.

For the CNN-NE (CNN-MC) training, we set the number of epochs to 600 (300) and the mini-batch size to 2048 (512). The Adam optimizer was set with an initial learning rate of 10^{-3} (10^{-5}) and dropped by a factor of 0.8 for every 20 (10) epochs. We initialized the kernel weights of CNN-NE randomly and initialized CNN-MC with the trained CNN-NE. In CNN-MC training, the positive patches were augmented by a factor of 8 (4 rotations by 90 degrees and another 4 rotations after flipping). We weighed the positive class in the cross entropy loss by the negative-to-positive training sample ratio to compensative for the imbalanced training data. During CNN-

MC deployment, we augmented every patch and averaged the 8 output scores as the patch score. We further averaged the patch scores of all positive patches from the same MC cluster by DBT view for the ROC analysis. We repeated the training of CNN-NE and fine-tuning of CNN-MC 5 times. Each repeated experiment used a different random seed for the weight initialization of CNN-NE and the data batching during CNN-NE training and CNN-MC fine-tuning. We reported the mean and standard deviation of the 5 CNN-MC AUCs. We reported only the mean of the 5 CNN-NE RMS estimates because all standard deviations were smaller than 2% of one pixel value.

4.2.5 Image Conditions

We compared MDR with other approaches, including SART [27], SART with multiscale bilateral filtering (MSBF) regularization [48], DBCN reconstruction [36], RED reconstruction without DBCN modeling, and post-reconstruction DNGAN denoising on SART and DBCN images. We previously evaluated DNGAN denoising of PVs for SART reconstruction and observed that it suffered from blurry MCs [15]. Therefore, PV denoising was not considered in this study.

4.3 Results

4.3.1 Effectiveness of DNGAN, CNN-NE and CNN-MC

As an example, we deployed one of the five CNN-NE models with VGG-Net backbone trained with different random initializations to the CNN-NE validation set from 2 iterations of SART images. Figure 4.2 shows that the CNN-NE RMS noise of the individual patches correlated well with the analytical calculations (correlation coefficient = 0.993, $p<0.0001$), indicating that CNN-NE can accurately estimate the RMS noise of an input patch. Similar correlations were observed for other reconstructions (not shown). While CNN-NE was trained using VICTRE images from 3 iterations of SART, it was applicable to other reconstructions and human subject

images at deployment because it was designed to assess the RMS variations in noisy pixels without relying on the specific image conditions.

Chan *et al.* [92] conducted an observer study using physical phantom DBTs and had seven radiologists detect the MC clusters in the STD, dnSTD, and STD+ conditions. Figure 4.3 shows example of the MC clusters. The study found that the average sensitivities of detecting MC clusters in dnSTD, obtained from DNGAN denoising of the STD images, were higher than those in STD for all 4 MC speck sizes and were comparable to those in STD+. This shows the effectiveness of the DNGAN denoiser to reduce noise and enhance MCs in DBT.

To demonstrate the potential of using CNN-MC in estimating the relative detectability of MC clusters, we applied the CNN-MC models to the extracted patches from the phantom DBTs used in the observer study. The CNN-MC models used VGG-Net backbone and were trained for 2 iterations of SART images. The solid lines in Figure 4.4 show that the AUCs increased with MC speck size. The AUCs of dnSTD were higher than STD and close to STD+. The relative rankings for the speck sizes and the image conditions given by the CNN-MC were consistent with the findings of the observer study. To show the benefit of transfer learning, we repeated the experiment but trained the CNN-MC models from scratch instead of from the pre-trained CNN-NE. In this case, the CNN-MC failed to learn and likely estimated the patch scores based on image features unrelated to MC detectability. This led to very low AUCs and a lack of correlation between MC size and AUC, as shown by the dotted lines in Figure 4.4.

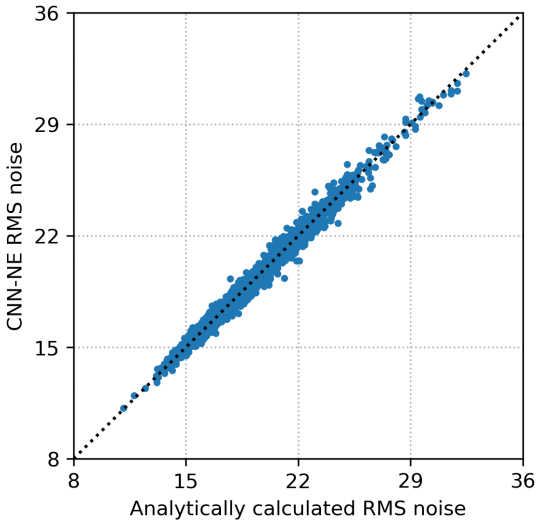


Figure 4.2 The scatter plot of the CNN-NE estimated RMS noise vs the analytically calculated RMS noise for the individual patches in the CNN-NE validation set (human subject DBT images). The dotted line is the diagonal line. The RMS noise is plotted in terms of pixel values.

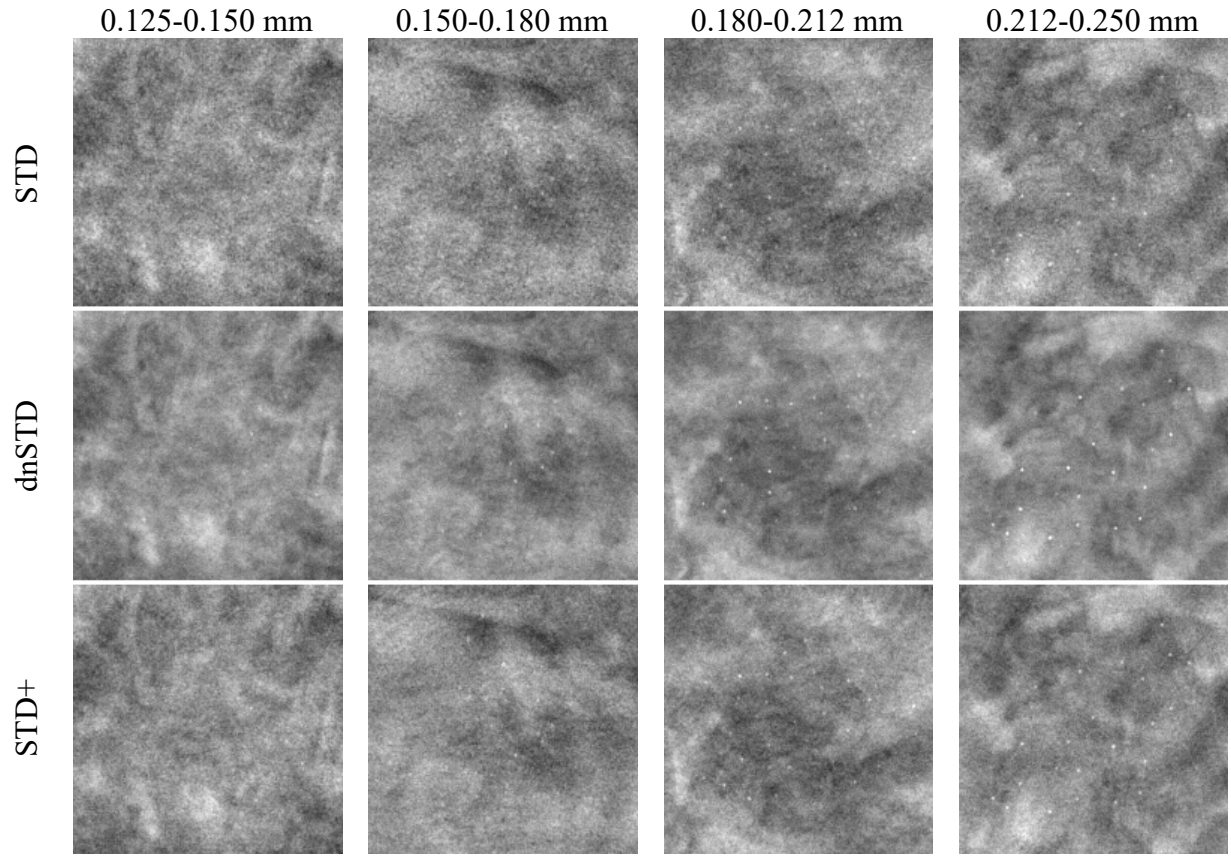


Figure 4.3 Example images (18 mm × 18 mm) with MC clusters from the physical phantom DBTs acquired and reconstructed by a GE Pristina DBT system. The dnSTD images were obtained by DNGAN denoising on the STD images. The STD+ mode used 54% more dose than the STD mode.

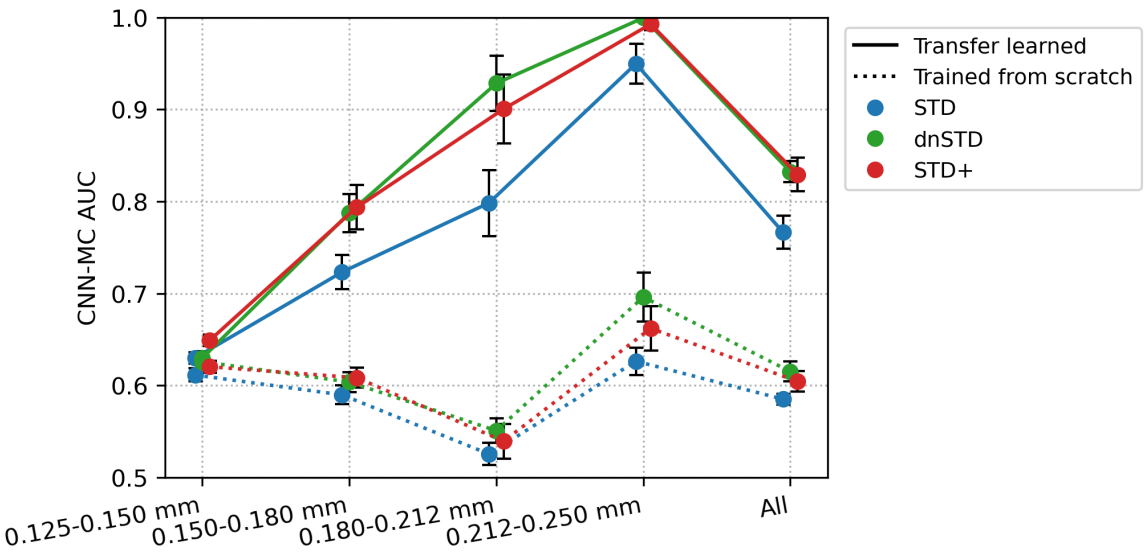


Figure 4.4 AUCs of CNN-MC for the classification of image patches with and without MC clusters in the physical phantom DBTs of 3 image conditions. The data points and error bars were obtained from the mean and standard deviation of the 5 repeated models trained with different random initialization. The points are slightly shifted horizontally to avoid overlap.

4.3.2 Parameter Selection for MDR Regularization

To select the regularization parameters β_{EP} and β_{RED} in MDR, we performed a grid search with β_{EP} = 30, 50, 70, 90 and β_{RED} = 100, 500, 1000, 5000. Figure 4.5 shows the task-based image quality assessment plot of CNN-MC AUC versus CNN-NE RMS noise for the VGG-Net backbone on the validation set for the different parameters. At a low β_{EP} value of 30, the AUCs were low because the noise was poorly suppressed and thus obscured MC conspicuity. As β_{EP} increased to 50, the MC signals were enhanced relative to noise and the AUCs were improved. But if β_{EP} further increased to 70 or 90, the regularization became so strong that the true MCs were smoothed, resulting in lower AUCs. For a given β_{EP}, decreasing β_{RED} meant a weaker regularization, resulting in higher image noise. On the other hand, reduced regularization placed greater emphasis on the data-fit term, thus enhancing the MC signals. This explains why we observed slight increase in AUC as β_{RED} decreased. We selected β_{RED} = 500 to balance the tradeoff between noise reduction

and signal enhancement. In other parts of this study, we set the regularization parameter β_{EP} to 50 and β_{RED} to 500 if not specified.

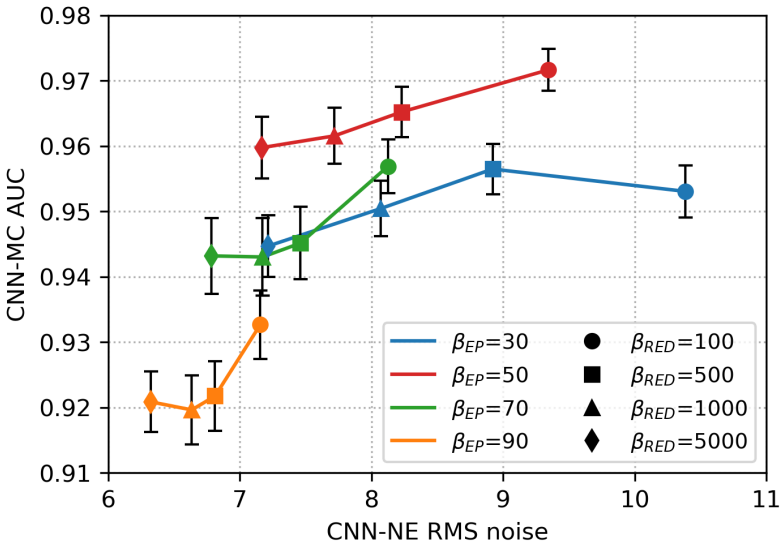


Figure 4.5 The task-based image quality assessment plot showing the tradeoff between MC cluster detectability (CNN-MC AUC) versus image noise (CNN-NE RMS noise) on the validation set reconstructed with different regularization parameters β_{EP} and β_{RED} in MDR. The data point obtained with a fixed β_{EP} are linked by a solid line and data points for a fixed β_{RED} are shown with the same symbol.

4.3.3 MDR Ablation Study

We conducted an ablation study to illustrate the effects of the three terms in the MDR cost function (4.7). The three terms could be turned on or off separately to create 6 partial models. To turn off $L_{\text{DBCN}}(x)$, we used the regular data-fit term $L_{\text{regular}}(x) = \frac{1}{2}\sum_{i=1}^{N_p}\|y_i - A_i x\|_2^2$ instead of (4.4). $R_{\text{EP}}(x)$ was turned off by setting β_{EP} to 0. $R_{\text{RED}}(x)$ was turned off by setting β_{RED} to 0 and skipping all DNGAN denoising.

Figure 4.6 shows the task-based image quality assessment plot for the VGG-Net backbone on the validation set for the different partial MDR models in comparison to the full MDR model. Figure 4.7 shows examples of MIP patches with and without MCs from the human subject validation set. From "R_{EP} Only", we found that the EP regularizer (β_{EP}=50) encouraged the

smoothness of the images but sacrificed the sharpness of the MCs. The L_{DBCN} term contributed to signal enhancement, as can be seen from the AUC improvement between "R_{EP} Only" and "$L_{\mathrm{DBCN}} + R_{\mathrm{EP}}$", although it produced exceptionally noisy images if used alone. The "L_{DBCN} Only" condition (RMS noise: 101.1, AUC: 0.846 ± 0.004) is outside the plot range of Figure 4.6 and is not shown in Figure 4.7. The "$L_{\mathrm{DBCN}} + R_{\mathrm{EP}}$" condition had a high AUC but also a high RMS noise. From "R_{RED} Only", we observed that the RED regularizer (β_{RED}=500) powered by DNGAN was good at noise reduction and signal preservation, but sometimes it falsely enhanced the background tissue structures. Those MC-like enhancements were detected as false positives by CNN-MC and led to an overall low AUC. The false enhancement was even more severe for "$L_{\mathrm{DBCN}} + R_{\mathrm{RED}}$" which had the lowest AUC in Figure 4.6. With "$R_{\mathrm{EP}} + R_{\mathrm{RED}}$", the false positives in the "R_{RED} Only" condition were suppressed and the RMS noise was the lowest, but it overly smoothed the MCs. Finally, the full MDR model took advantage of all three terms and achieved reasonably low RMS noise and the highest AUC.

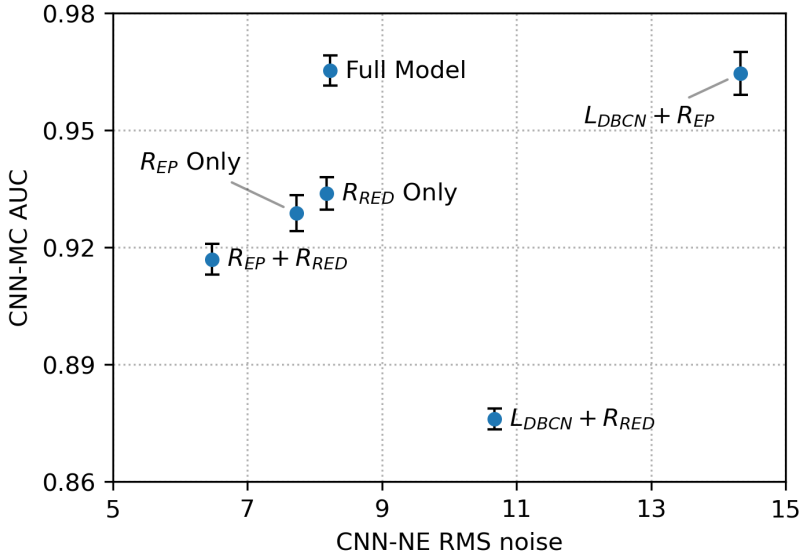


Figure 4.6 The task-based image quality assessment plot on the validation set showing the tradeoffs of the different terms in the MDR ablation study.

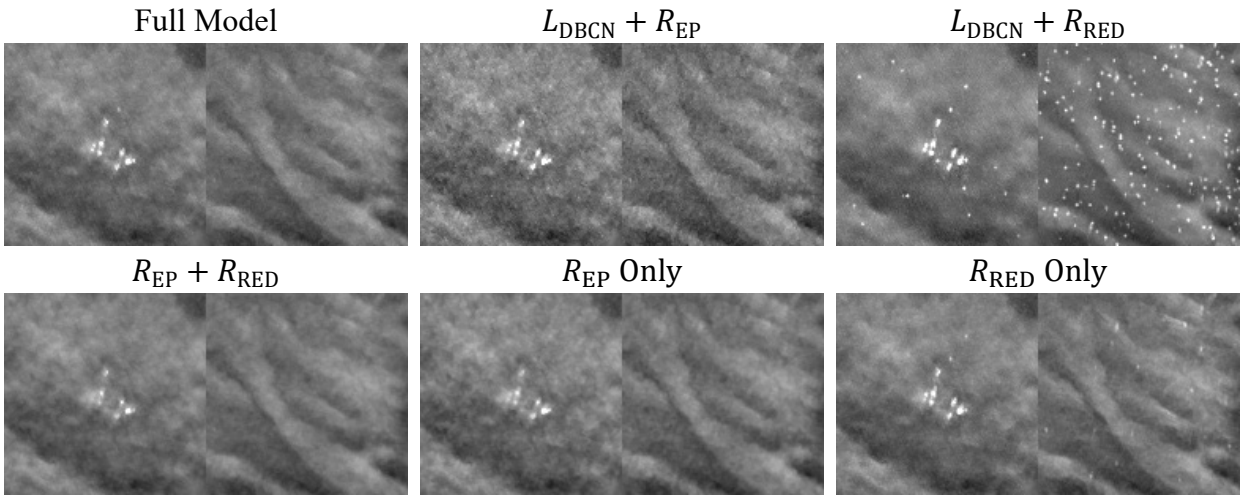


Figure 4.7 Example 3-slice MIP patches (12.8 mm × 12.8 mm) from the human subject validation set for the MDR ablation study. In each pair, a positive patch of MC cluster is shown on the left and a negative MC-free patch on the right.

4.3.4 Comparisons on Validation Set and Independent Test Set

We compared eight image conditions in this section: (a) SART iteration 2; (b) SART iteration 3; (c) SART iteration 3 with post-reconstruction DNGAN denoising; (d) RED without DBCN modeling, equivalent to "R_{RED} Only" in Section 4.3.3; (e) DBCN (β=70); (f) DBCN (β=50) with post-reconstruction DNGAN denoising; (g) MSBF; (h) MDR. Figure 4.8 shows the task-based image quality assessment plot for the validation set and the independent test set using the VGG-Net, ResNet and ConvNeXt backbones. The VGG-Net and ResNet had smaller error bars for AUCs and thus better reproducibility than ConvNeXt. The AUC and RMS noise rankings given by the three networks had some variations but were mostly consistent within the error bars. Table 4.3 summarizes the RMS noise and AUC ranking results in Figure 4.8. The final rankings were obtained by summing the individual rankings. The rankings on the validation set largely agreed with those on the independent test set with minor variations between adjacent ranks except for DBCN, indicating reasonable generalizability of the proposed DCNN image quality assessment approach to the unseen test set.

Figure 4.9 shows the example MIP patches of MC clusters and Figure 4.10 shows example images of a spiculated mass from the human subject test set. The visual judgment of the noise levels matched the quantitative RMS noise estimation. The SART images (conditions (a) and (b)) were noisy with poor MC conspicuity as it is an unregularized reconstruction method [28]. Increasing the number of SART iterations enhanced the signal, but it also amplified noise, resulting in comparable signal-to-noise ratios and thus AUCs, as illustrated in Figure 4.8. This observation reaffirmed our understanding of SART that there was no benefit of continuing SART iterations without regularization. Furthermore, it highlighted the advantage of our task-based metric using CNN-MC, which focused on the detectability of the MC clusters and provided a more informative assessment. DBCN (condition (e)) had relatively large variations of AUC rankings among the three CNN-MCs and two data sets, possibly because its patchy and noisy background confused the image features and made the training set less representative of the data distribution. Post-reconstruction denoising with DNGAN was flexible for application to different reconstruction techniques, reducing image noise and improving the AUC as shown in the examples for both SART (condition (c) in comparison to (b)) and DBCN (condition (f) in comparison to (e)). Note that, with DNGAN denoising, SART and DBCN could use parameters of stronger MC enhancement (3 iterations instead of 2 in SART, and β=50 instead of 70 in DBCN). Although RED (condition (d)) reached a very low noise level and its visual signal quality was comparable to that of MDR, its AUC was low due to false enhancement as discussed in Section 4.3.3. The conspicuous bright spots resembling MCs in the examples shown in Figure 4.7 (lower right) and Figure 4.10(d) reiterated this downside of RED. The AUC ranking of MSBF (condition (g)) was high, but its noise also remained high. The proposed MDR (condition (h)) achieved one of the lowest RMS noise and the highest AUC rankings than other image conditions. The high visibility

of the subtle MCs and smooth background of the MDR images can also be clearly seen in Figure 4.9. Moreover, it preserved the fine texture details and had a satisfactory visual quality of mass margins without creating artifact in Figure 4.10.

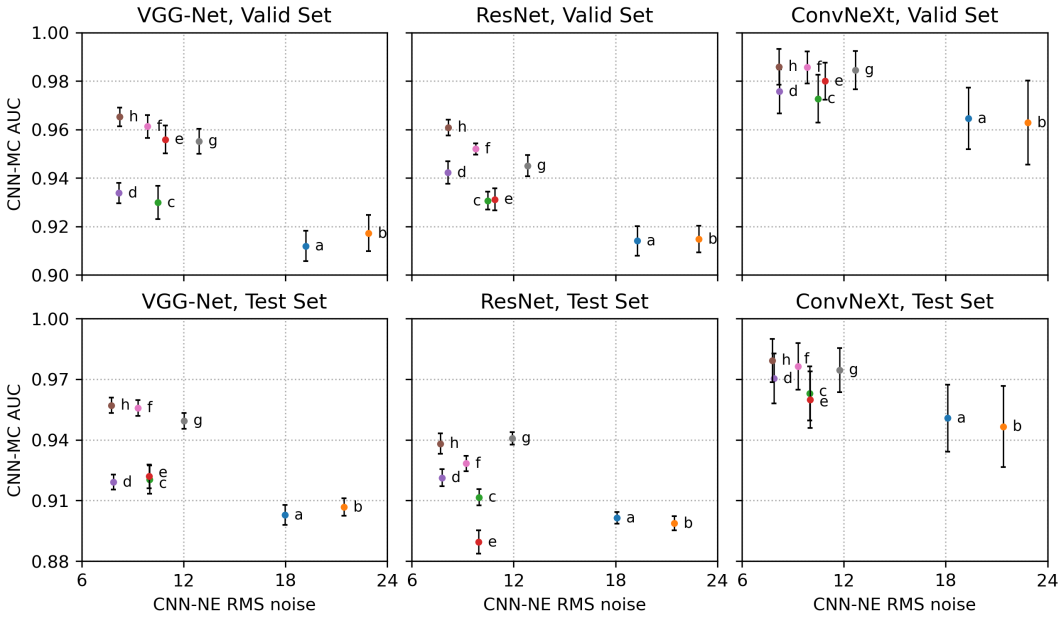


Figure 4.8 The task-based image quality assessment plot on the human subject validation set (first row) and test set (second row) using the VGG-Net (left column), ResNet (middle column) and ConvNeXt (right column) for comparing the reconstruction and denoising approaches. The labels (a) to (h) are defined in Section 4.3.4 and Table 4.3.

Table 4.3 Summary of rankings for different reconstruction and denoising approaches using DCNN task-based image quality assessment in Figure 4.8.

		CNN-NE RMS Noise Rankings					CNN-MC AUC Rankings				
		VGG-Net	ResNet	ConvNeXt	Sum	**Final**	VGG-Net	ResNet	ConvNeXt	Sum	**Final**
Validation Set	(a) SART Iter 2	7	7	7	21	**7**	8	8	7	23	**8**
	(b) SART Iter 3	8	8	8	24	**8**	7	7	8	22	**7**
	(c) SART Iter 3 + DNGAN	4	4	4	12	**4**	6	6	6	18	**6**
	(d) RED	1	1	2	4	**1**	5	4	5	14	**5**
	(e) DBCN	5	5	5	15	**5**	3	5	4	12	**4**
	(f) DBCN + DNGAN	3	3	3	9	**3**	2	2	2	6	**2**
	(g) MSBF	6	6	6	18	**6**	4	3	3	10	**3**
	(h) MDR	2	2	1	5	**2**	1	1	1	3	**1**
		VGG-Net	ResNet	ConvNeXt	Sum	**Final**	VGG-Net	ResNet	ConvNeXt	Sum	**Final**
Test Set	(a) SART Iter 2	7	7	7	21	**7**	8	6	7	21	**7**
	(b) SART Iter 3	8	8	8	24	**8**	7	7	8	22	**8**
	(c) SART Iter 3 + DNGAN	5	5	4	14	**5**	5	5	5	15	**5**
	(d) RED	2	2	2	6	**2**	6	4	4	14	**4**
	(e) DBCN	4	4	5	13	**4**	4	8	6	18	**6**
	(f) DBCN + DNGAN	3	3	3	9	**3**	2	3	2	7	**2**
	(g) MSBF	6	6	6	18	**6**	3	1	3	7	**2**
	(h) MDR	1	1	1	3	**1**	1	2	1	4	**1**

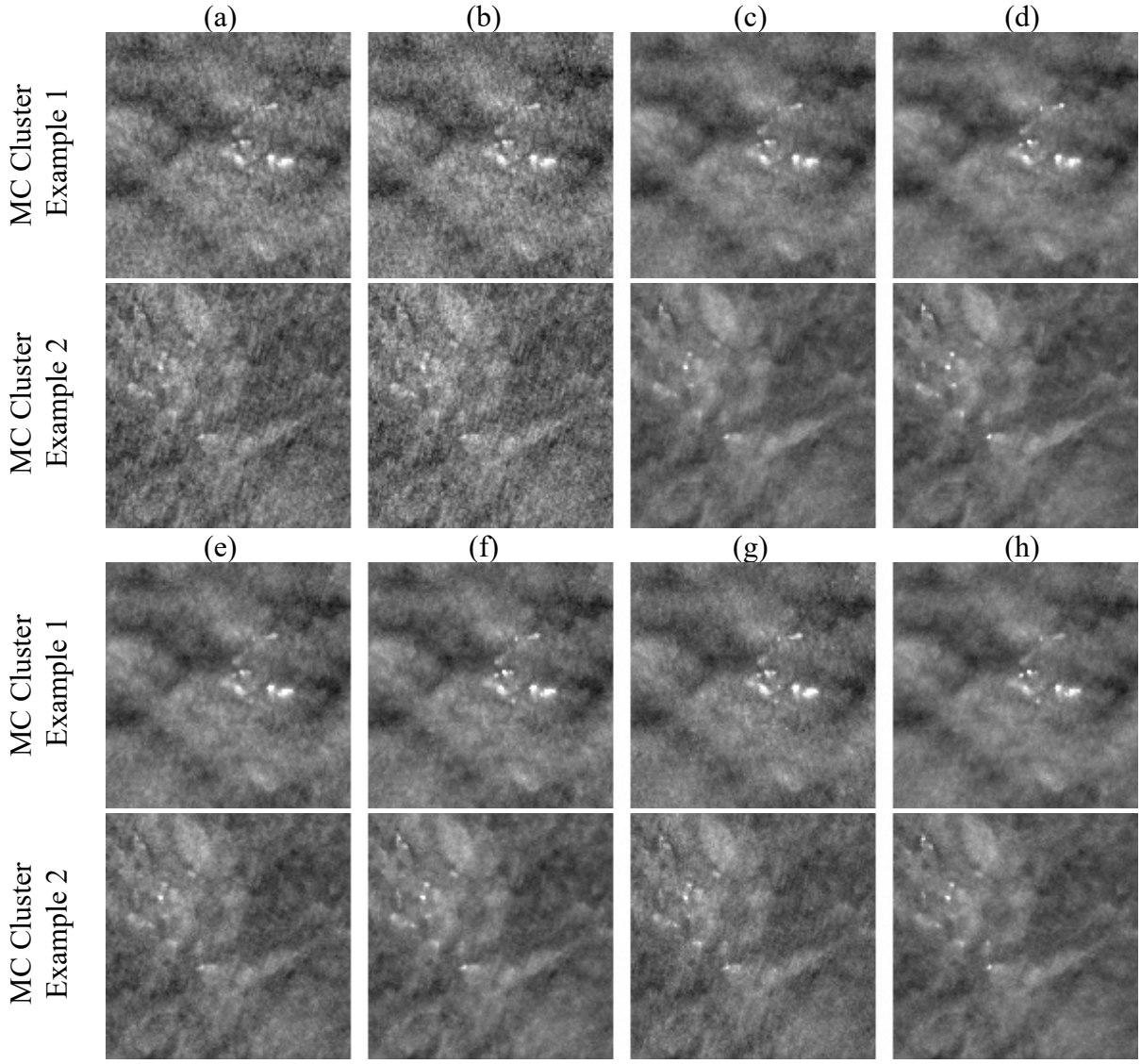


Figure 4.9 Example 3-slice MIP patches (12.8 mm × 12.8 mm) of two MC clusters (example 1: ductal carcinoma in situ; example 2: invasive ductal carcinoma) from the human subject test set. The labels (a) to (h) are defined in Section 4.3.4 and Table 4.3.

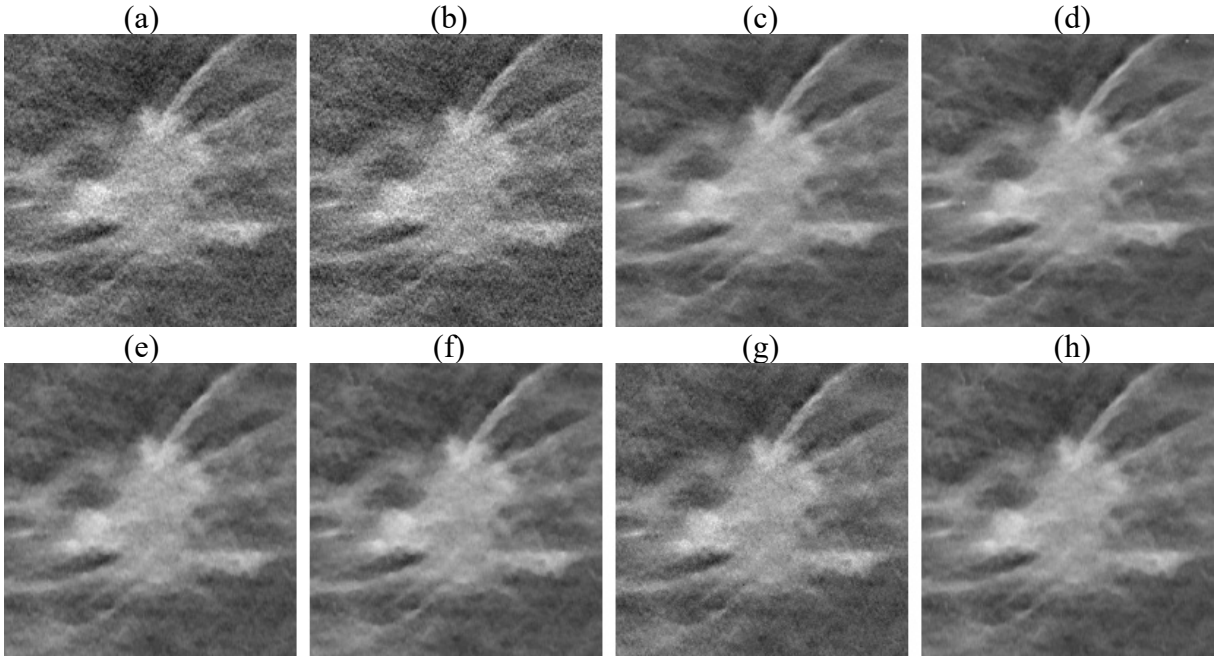


Figure 4.10 Example images (20 mm × 20 mm) of a spiculated mass (invasive ductal carcinoma) of a human subject DBT. The labels (a) to (h) are defined in Section 4.3.4 and Table 4.3.

4.3.5 Qualitative Comparison Between MDR and Commercial Reconstruction

The GEN2 DBT system used for acquiring the human subject DBTs was developed as a prototype for research purposes, so it lacked a built-in reconstruction method. To compare the proposed MDR with gold standard reconstruction, we reconstructed a human subject DBT case acquired with the commercial GE Pristina DBT system. Figure 4.11 shows a qualitative comparison of the full DBT slices and the zoomed-in views of an MC cluster obtained with the Pristina reconstruction and our MDR approach. The full slice image of MDR had reduced noise levels compared to the Pristina reconstruction. The zoomed images facilitated the visualization of the nuances of the MC signals, surrounding tissue structures, and noise. The MDR not only achieved a substantial reduction in noise, but also preserved details and enhanced the subtle MC specks, demonstrating superior image quality than the commercial reconstruction.

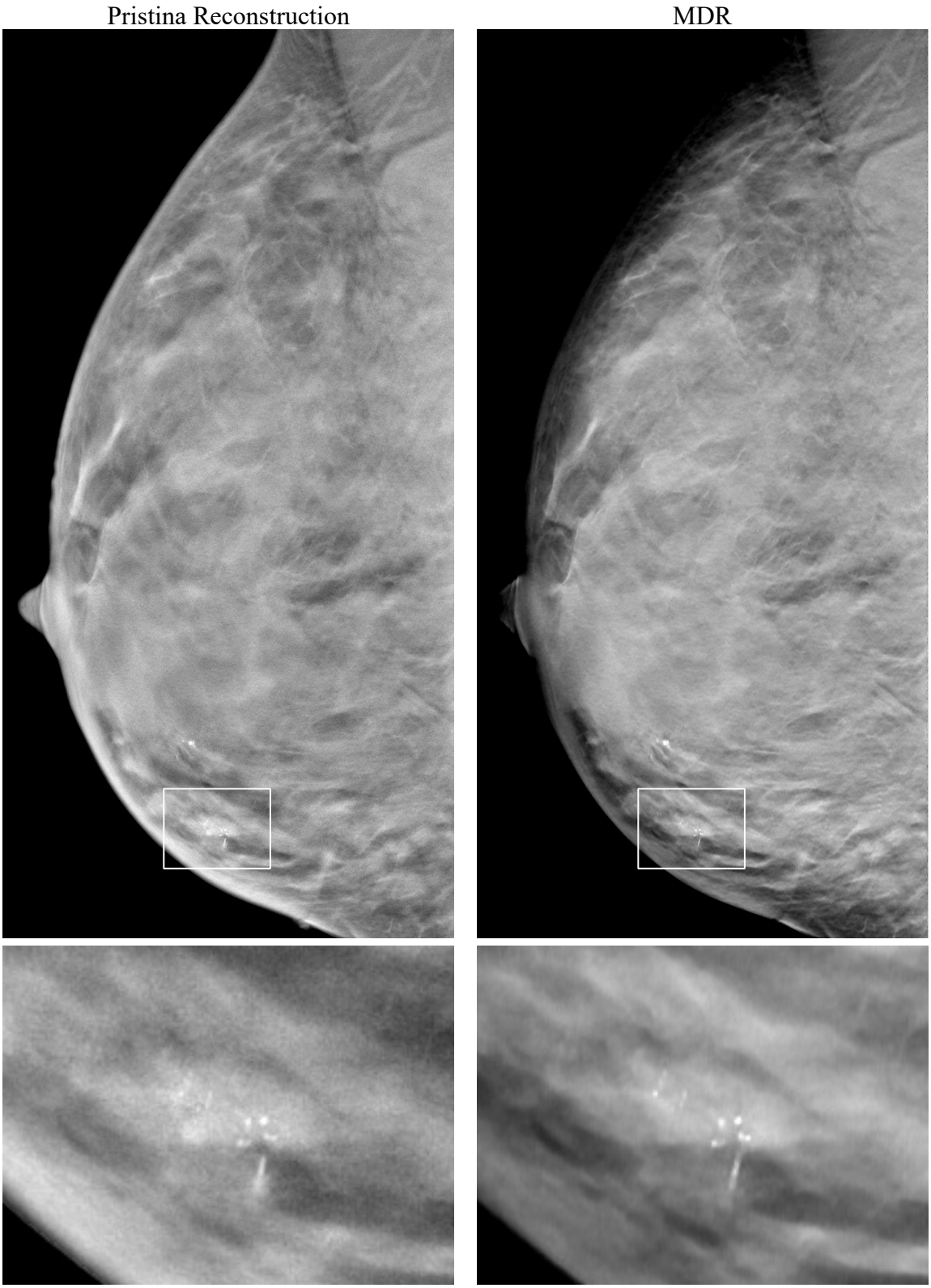


Figure 4.11 Qualitative comparison between the commercial GE Pristina reconstruction and the proposed MDR. Top row: full DBT slices (84.4 mm × 170 mm). Bottom row: zoomed-in views of the marked regions (20 mm × 15 mm) containing an MC cluster (ductal carcinoma in situ).

The darkening of breast periphery as seen in the MDR reconstructed image in Figure 4.11 resulted from the physics of x-ray imaging. The reduction in tissue thickness and thus increasing penetration of x-ray exposure to the detector caused varying image intensities near the breast boundaries. Commercial DBT images are generally processed with peripheral equalization software to facilitate radiologists' reading. We did not attempt to address this issue because we did not have access to any breast peripheral equalization software and its development was outside the scope of this study. One can always view the peripheral regions by adjusting the display contrast and brightness properly. As demonstrated in Figure 4.11, the proposed MDR method enhanced the MC signals regardless of their location. Despite being near the breast periphery, the enlarged ROI still shows the improvement of image quality by the MDR.

4.4 Discussion

DBCN and DNGAN represent two distinct approaches for enhancing the subtle signals in DBT. The DBCN method approximates the deblurring and noise decorrelation by the inverse of the detector point spread filter and the prewhitening filter. Both operations boost the high frequency signals, which inevitably amplify high frequency noise at the same time. DBCN therefore is incorporated with an EP regularizer to control the noise. DNGAN emphasizes the signals by smoothing the surrounding backgrounds like a regularizer. It also increases the contrast of the signals to some extent, but sometimes falsely enhances the high frequency structures and creates MC-like artifacts. We explored combining the two methods to take advantage of both. We implemented the newly proposed MDR method in an efficient way so that it could reconstruct the full-sized 3D DBT images at full resolution. Quantitative results using our newly developed DCNN image quality measures on the independent test set of human subject DBTs revealed that MDR achieved the highest AUC rankings for MC detectability and low noise among the

reconstruction and denoising approaches studied. This study shows the promise of combining the physics-based MBIR and the learning-based DCNNs for reconstruction.

The CNN-MC model observer can characterize the detectability of the MC clusters in human subject DBTs. Because of the complexity of the SKS MC clusters, we trained the CNN-MC to learn the signal representations through back-propagation, unlike CHO and *d'* where the target signal is trained to be fitted with channelized functions or mathematically defined. Moreover, the CNN-MC can include the assessment of image background with and without MC-like false enhancement (shown in Section 4.3.3) that the traditional image quality measures such as *d'* cannot achieve because they focus on individual signals. Assessing potential false information generated by an image processing method is crucial for evaluating its feasibility in clinical applications.

The role of CNN-NE can be understood from two aspects. First, CNN-NE played a pivotal role in our training and processing pipeline because it served as the pre-trained model of CNN-MC, which worked well only when transfer-learned from CNN-NE (Figure 4.4). The advantages of transfer learning were elaborated in Section 4.2.2.3. Second, despite the possibility of analytically calculating the RMS noise of an image patch, our work showcased the ability of deep CNN to approximate this calculation. There is also a growing interest in applying CNNs for noise quantification in medical imaging [130]. The development of CNN-NE provides users the flexibility of choosing between analytical and CNN calculations for estimating the RMS noise. It is noteworthy that CNN-NE has superior computational efficiency when run on a GPU, compared to the time-consuming grid-based calculation method. We proposed to use CNN-NE and CNN-MC together to guide the selection of the reconstruction parameters and compare different reconstruction techniques and denoising approaches. The task-based image quality assessment plot

of CNN-MC AUC versus CNN-NE RMS noise clearly illustrates the tradeoff between the detectability and noise of the different image conditions.

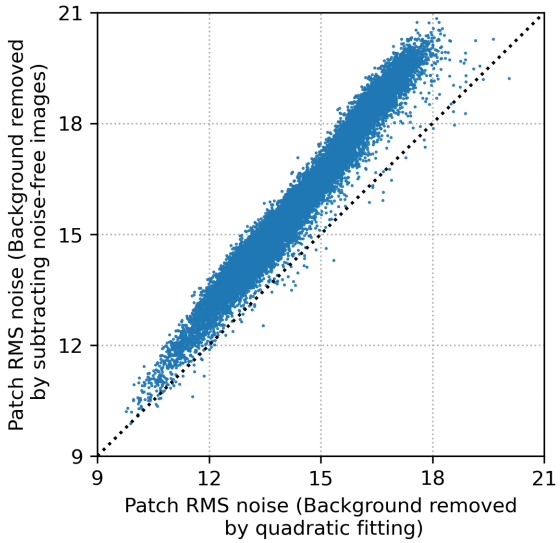


Figure 4.12 The scatter plot of the analytically calculated RMS noise for the CNN-NE training image patches with the exposure factor of 1 using either quadratic fitting or noise-free images for background removal. The RMS noise is plotted in terms of pixel values.

The RMS noise calculation introduced in Section 4.2.2.2 involved a quadratic fitting step to remove the background trend. Given that we used VICTRE simulation data for CNN-NE training and simulations can produce (nearly) noise-free images, it is possible to remove the background structures using such noise-free images. To verify that the RMS noise in images could be effectively estimated by the quadratic fitting background removal method, we simulated the nearly noise-free images using an exposure factor of 10, and then calculated the RMS values for the CNN-NE training image patches with the exposure factor of 1 using the two approaches. Figure 4.12 shows that the two sets of RMS values were close and exhibited a strong linear correlation (correlation coefficient = 0.983, $p<0.0001$), indicating the effectiveness of quadratic fitting for background removal. While using noise-free images may be the ideal approach for accurate RMS calculations, its real-world impact is limited due to the linearity and small

differences. More importantly, having the noise-free images is a special case and they are typically unavailable for actual scanned images. Therefore, we employed the fitting method in our work due to its practicality in general scenarios in case others would want to adopt the method for their applications.

A suboptimal CNN-MC model observer is preferred for evaluating the effectiveness of image processing techniques since image processing is not expected to improve the performance of an ideal observer. Zhang *et al.* [121] showed that a DCNN could be trained to approximate an ideal observer or a suboptimal observer, depending on its learning capacity and the available training sample size, for the task of detecting MC clusters in a simulation study. This is consistent with the observations in our study. We found that DCNNs of very deep structures reached an AUC of 1 regardless of the image conditions. We had to constrain the learning capacity by limiting the network sizes such that they could discriminate between the image conditions. In addition, it is known that human observers are suboptimal, as also demonstrated by our observer study [92]. Since there are no specific guides for selecting the DCNN structure for a given task, we investigated three CNN-MC backbones. The pooled performance of the multiple models may partly alleviate the learning variability, in analogy to the use of multiple observers to account for the interobserver variability in human reader studies. The CNN-NE trained as a regression model was much more stable, as observed from the small variances among the different initializations and different CNN backbones.

We did not analyze the convergence property of MDR in this study. Due to the limited-angle nature of the DBT scan and the fact that radiologists are accustomed to the tissue texture appearance of the 2D projection mammograms, we usually do early stopping on the DBT reconstruction so that the soft tissue structured backgrounds are not overly enhanced. The

truncation artifact correction [129] during reconstruction essentially changes the system matrices and also complicates the convergence analysis. Furthermore, the RED assumptions are not satisfied by many popular denoisers including DNGAN [122]. Without the explicit expression and minimization of a cost function, RED serves only as a motivation and one can at best hope to obtain the fixed-point convergence or an equilibrium of the reconstruction [131][132]. For these reasons, we focused on image quality instead of seeking the convergence of the iterates.

There are limitations in the current study. First, the DNGAN denoisers were trained on the SART images but were directly applied to the intermediate images of MDR. This introduced a noise mismatch between the denoiser and the images to be denoised. The fact that the DNGAN works well in MDR indicates the flexibility of the application, but it also leaves room for improvement by fine-tuning the DNGAN denoisers at every MDR iteration, which will tradeoff computational efficiency, however. Second, the number of MC clusters in our data set is limited. A larger CNN-MC training set may further improve the generalizability of the CNN-MC model observer and reduce the AUC variations among the different DCNN models or between training and deployment to unseen cases. Third, although MDR achieved the highest MC detectability ranking and low noise, post DBCN reconstruction with DNGAN denoising was also among the top three rankings. Post-reconstruction denoising is the only practical approach if the raw PVs are not available for user-designed reconstruction. It will be of interest to evaluate the effectiveness of DNGAN denoising for the DBT images reconstructed by the various commercial DBT systems when large human subject data sets with subtle MCs from these systems become available. Fourth, the proposed DCNN image evaluation approach only focused on image noise and MC detectability. The evaluations of mass and tissue textures still relied on visual judgment. It will be useful if DCNN models can be developed to assess the quality of these important image features

as well in the future. Finally, we do not address the time efficiency of MBIR as it falls outside the primary scope of this paper. We encourage future research to tackle this issue, which may complement and enhance the utility of the proposed method.

4.5 Conclusion

We proposed MDR by combining DBCN modeling and DNGAN denoising using the RED framework for DBT reconstruction. To facilitate the task-based image quality assessment, we also proposed two DCNN tools. The CNN-NE was trained to estimate the RMS noise of the DBT images. The CNN-MC was trained to be a model observer to evaluate the detectability of clustered MCs. The efficacies of DNGAN, CNN-NE and CNN-MC were demonstrated using physical phantom DBTs and human subject DBTs. We investigated the impacts of the MDR regularization parameters and the cost function terms. MDR reduced image noise and improved MC detectability on an independent test set of human subject DBTs. The proposed CNN-NE and CNN-MC evaluation method can serve as a surrogate for human observers to provide task-specific metrics and rank the imaging systems in a cost-effective way. The proposed reconstruction method may potentially lead to lower dose and higher sensitivity and specificity for MC detection in breast cancer screening and diagnosis.

Chapter 5
X-ray Source Motion Blur Modeling and Deblurring with Generative Diffusion for Digital Breast Tomosynthesis

5.1 Introduction

One[1] of the primary challenges in DBT imaging is the x-ray source motion blur that can degrade the quality of DBT images, reducing their sharpness and potentially affecting the visibility of subtle lesions such as MCs [133]–[136]. The x-ray tube motion in DBT can be carried out in two modes: step-and-shoot mode and continuous-motion mode [5][18]. In the step-and-shoot mode, the x-ray source essentially stops at each angle, acquires a PV, then moves to the next angle. The focal spot size of this mode equals the nominal size of a stationary source. Compared with the ideal point source, the nominal source has a negligible effect on the image sharpness [133]. In continuous-motion mode, the x-ray source moves continuously along the designated arc while capturing PVs at the respective angles by pulsing the x-rays. This approach introduces source motion blur that depends on the pulse width, causing geometric unsharpness and substantially degrading the image resolution in the source motion direction [133][136]. Very recently, Siemens introduced a flying focal spot technology for DBT to prevent blurring effect of a continuously moving source [137]. This new technology is outside the scope of our study.

[1] This chapter is based on a submitted journal manuscript.

X-ray source motion blur is not unique to DBT as a similar problem occurs in computed tomography (CT). Several image processing and reconstruction methods have been developed to address this issue. In CT, Tilley *et al.* proposed to deconvolve the projection data for focal spot blur before reconstruction [138]. In another study, Tilley *et al.* modified the CT system forward model by incorporating focal spot blur as a shift-invariant convolution applied to the reconstructed images and used it in model-based iterative reconstruction (MBIR) [139]. Fu *et al.* and Majee *et al.* also modified the CT system forward model, but took a different approach by subsampling the focal spot and then averaging the projections [140][141]. In DBT, Michielsen *et al.* treated the PV as a summation of the projections of each DBT slice convolved with a blur kernel dependent on the slice height, and proposed to update the DBT slices sequentially for solving the reconstruction problem [142]. While their method increased the peak contrast-to-noise ratio of a simulated MC in a uniform background, the improvement was somewhat limited when applied to images with heterogeneous backgrounds.

Nonblind image deblurring is an important topic in image processing and computer vision research. Its task is to estimate sharp images from the blurred images given the blur kernel. Classic methods for nonblind image deblurring include the renowned Wiener filter [143] and Richardson-Lucy deconvolution [144]. Model-based methods construct mathematical models and priors to estimate the latent sharp images using statistical methods like maximum a posterior (MAP). Great efforts have been devoted to designing image priors for MAP [145]–[147]. In recent years, deep convolutional neural network (CNN) methods have emerged as powerful tools, leveraging the capacity of deep networks to learn complex mappings from blurry to sharp images [148][149]. More recently, denoising diffusion probabilistic models (DDPM) and score-based generative models have gained significant attention for their ability to generate high-quality samples [150]–

[152]. Their remarkable success in image generation facilitates various inverse problems including image deblurring [153]–[155] and image reconstruction [156][157].

X-ray source blur modeling for DBT remains a challenging problem due to its shift-variant nature. Furthermore, the low-dose exposure of DBT introduces a high level of image noise. In this study, we first analytically derived the in-plane blur kernel for the reconstructed DBT slices using imaging geometry and showed that it could be approximated by a shift-invariant kernel for a given slice. Then, we investigated several choices of problem formulations and converted the problem into post-processing deblurring. When deblurring, it is important to control the image noise level by means of regularization. We proposed an effective nonblind deblurring approach with DDPM as an image prior and applied it to the reconstructed DBT images.

5.2 Methods

5.2.1 Source Blur Modeling

5.2.1.1 DBT Imaging System

This section briefly reviews the DBT imaging system introduced in Chapter 2, but with slightly different notations. Figure 5.1 shows the diagram of the DBT imaging system. We use x-y-z coordinates for the imaging volume and t-s coordinates for the digital detector. During the imaging process, the x-ray source moves around the compressed breast in the chest wall plane. The rotation center is denoted as O. The source rotates over a limited angular range, and a total of N_p PVs are captured. We denote the source location for the central PV as S and its vertical projection onto the detector plane as D. Let $v \in \mathbb{R}^{N_x N_y N_z}$ denote the (vectorized) DBT volume and $y_i \in \mathbb{R}^{N_t N_s}$ denote the post-log PV at the ith scan angle. The forward imaging process can be characterized as a system of equations by the system matrices $A_i \in \mathbb{R}^{N_t N_s \times N_x N_y N_z}$

$$y_i = A_i v + n_{i,y} \qquad i = 1, \dots, N_p \tag{5.1}$$

where $n_{i,y} \sim \mathcal{N}\left(0, \sigma_{i,y}^2 I\right)$ is the additive Gaussian PV noise. We stack all the A_i to define the overall system of equations

$$y = Av + n_y \qquad \text{where } A = \begin{bmatrix} A_1 \\ \vdots \\ A_{N_p} \end{bmatrix}, \quad y = \begin{bmatrix} y_1 \\ \vdots \\ y_{N_p} \end{bmatrix}, \quad n_y = \begin{bmatrix} n_{1,y} \\ \vdots \\ n_{N_p,y} \end{bmatrix}. \tag{5.2}$$

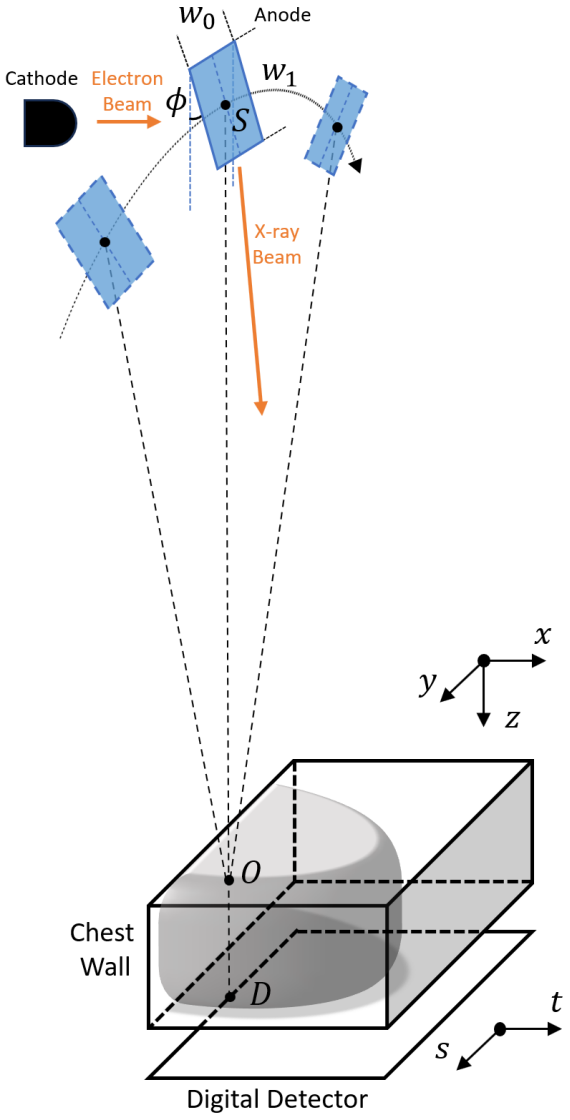


Figure 5.1 Diagram of DBT imaging system. The finite-sized rectangular x-ray source is exaggerated to show details.

5.2.1.2 Constructing Source Blur Matrix B

The x-ray source is not a point source but has a finite size and shape. For simplicity, it is modeled as a rectangular source (edges w_0 and w_1, target angle ϕ) with x-rays emitted uniformly across the anode target, as shown in Figure 5.1. When the source rotates around O, the edge w_0 stays parallel to the tangent line of the source trajectory. In x-ray imaging, a small ϕ is often used to keep the effective focal spot size small according to the line focus principle [158]. This study uses a nominal focal spot size of $w_0 = w_1 \sin\phi = 0.3$ mm, which is the nominal size for most mammography systems.

For the step-and-shoot mode, the source size is the same as the stationary nominal source size. However, for the continuous-motion mode, w_0 is elongated and equal to the nominal size convolved with the source traveling distance during the x-ray pulse. For example, the Siemens Mammomat Inspiration DBT system uses a continuous-motion x-ray source with a 0.18° motion angle per pulse [159]. Given that the distance between the source and rotation center is 600 mm, the effective source size w_0 is $0.18^\circ \times (\pi/180^\circ) \times 600 \text{ mm} + 0.3 \text{ mm} = 2.185 \text{ mm}$.

Zheng *et al.* conducted a simulation study and demonstrated that a focal spot of the nominal size has negligible effect on the DBT image resolution compared with the ideal point source, while the extra blur caused by source motion leads to a substantial loss of image resolution in the source motion direction [133]. Therefore, we further simplify our source model and ignore the dimension w_1 in the x-direction (the anode-cathode direction, i.e., the chestwall-anterior direction) because $w_1 \sin\phi$ remains at 0.3 mm at the chest wall regardless of the motion of x-ray source and decreases as the distance from the chest wall increases, causing a negligible effect on image resolution. We focus on the continuous-motion x-ray source and consider a simplified 1D rectangular source of length w_0 tangential to the source motion trajectory, as shown in Figure 5.2(a).

During the imaging process, for the scan angle α and an impulse with a distance z_s from the detector, the point spread function (PSF) at the detector is a rectangle whose width can be obtained using triangle similarity (Figure 5.2(a))

$$w_d(z_s, \alpha) = \frac{w_0}{\cos\alpha} \cdot \frac{z_s}{D_{SO} \cdot \cos\alpha + D_{OD} - z_s} \tag{5.3}$$

where D_{SO} is the distance between S and O, and D_{OD} is the distance between O and D.

During reconstruction, the point source projector is normally used for forward and backward projections. Then, as shown in Figure 5.2(b), the in-plane PSF of an impulse at the reconstructed DBT slice at z_s is a rectangle of width

$$w_s(z_s, \alpha) = w_d(z_s, \alpha) \cdot \frac{D_{SO} \cdot \cos\alpha + D_{OD} - z_s}{D_{SO} \cdot \cos\alpha + D_{OD}} = \frac{w_0}{\cos\alpha} \cdot \frac{z_s}{D_{SO} \cdot \cos\alpha + D_{OD}}. \tag{5.4}$$

This is the source blur PSF from one scan angle. The source blur PSF is independent of the location of the impulse along both the x- and y-directions, i.e., shift-invariant, at a given z_s.

The source motion blur in the reconstructed DBT images is a combined effect of the blurs from all scan angles. As shown in Section 5.4.1.1, the widths w_s from different scan angles are close to their mean averaged over all angles

$$\bar{w}_s(z_s) = \frac{1}{N_p} \sum\nolimits_{\alpha \in \{\alpha_1, \ldots, \alpha_{N_p}\}} w_s(z_s, \alpha). \tag{5.5}$$

Therefore, we treat the aggregated PSF over all scan angles also as a rectangle with a width of $\bar{w}_s$.

To summarize, for a continuous-motion DBT system, if the moving source is modeled as a 1D rectangle tangential to the source motion trajectory, then the in-plane source blur in the reconstructed DBT images can be approximated by a 1D rectangle, the size of which is shift-invariant over a DBT plane at a given slice height. Mathematically, we define a block-diagonal matrix B to characterize the source motion blur and incorporate it into the DBT system of equations (5.2)

$$y = ABv + n_y \qquad \text{where } B = \begin{bmatrix} B_1 & 0 & \cdots & 0 \\ 0 & B_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & B_{N_z} \end{bmatrix}. \tag{5.6}$$

The matrices $B_z \in \mathbb{R}^{N_x N_y \times N_x N_y}$ represent the shift-invariant blur for each slice $z = 1, \ldots, N_z$ and can be efficiently implemented as convolution with kernel size $\bar{w}_s$.

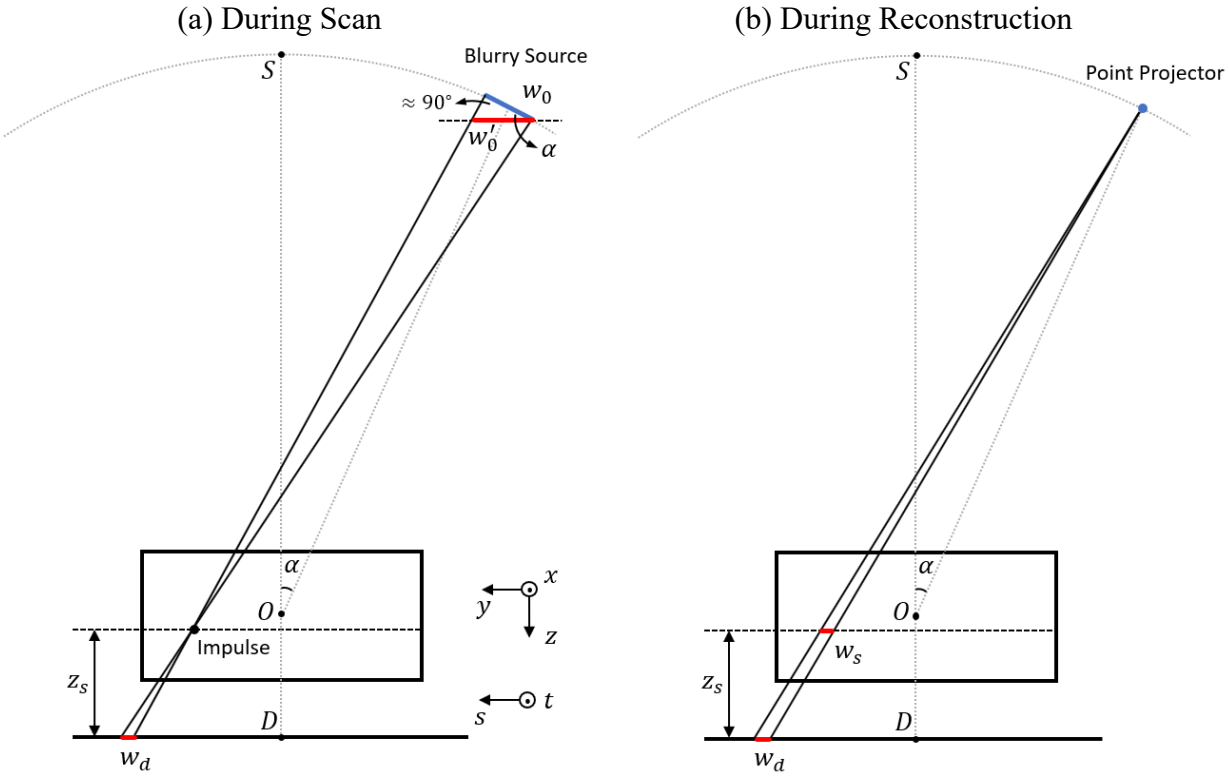


Figure 5.2 The simplified 1D rectangle source and the derivation of in-plane PSF. (a) In the forward scan process, given a blurry source w_0 and an impulse at height z_s above the detector, its PSF at the detector is w_d. (b) In the reconstruction process, the in-plane PSF at the impulse location is w_s.

5.2.1.3 Challenges in Using AB for DBT Reconstruction

Incorporating B into the DBT system model presents an opportunity to replace A with a new system matrix AB in existing DBT reconstruction algorithms for forward and backward projections. However, the inherent limited-angle design of DBT makes the reconstruction inverse problem highly underdetermined. We found that this posed a substantial challenge, and using AB

in DBT reconstruction did not improve image sharpness. Section 5.3.2.2 and Section 5.4.1.2 introduce a simulation study to demonstrate the limitation of this approach.

The work of Fu *et al.* [140] and Majee *et al.* [141] addressed source motion blur in CT imaging by introducing sub-focal spots within the blurry source and used the subsampled system matrix for CT reconstruction. We also explored a similar idea for DBT by putting N_{sub} point sources as sub-focal spots distributed uniformly within w_0. The corresponding subsampled system matrix is

$$A_{\text{sub}} = \frac{1}{N_{\text{sub}}} \sum_{i_{\text{sub}}=1}^{N_{\text{sub}}} A\left(\delta_{i_{\text{sub}}}\right) \tag{5.7}$$

where $A(\delta)$ is the perturbed system matrix of A obtained by offsetting the scan angles by δ, and $\delta_{i_{\text{sub}}} = \frac{w_0}{D_{SO}} \frac{1}{N_{\text{sub}}} \left(i_{\text{sub}} - \frac{N_{\text{sub}}+1}{2}\right)$ in radians for $i_{\text{sub}} = 1, \dots, N_{\text{sub}}$. This approach was conceptually similar to the use of AB, and we found that it did not improve image sharpness in DBT reconstruction for similar reasons. The simulation study in Section 5.3.2.2 and Section 5.4.1.2 also shows the limitation associated with this approach.

5.2.2 Nonblind Image Deblurring

The negative results of using AB and A_{sub} motivated us to separate B from A and turn the source blur modeling problem into post-processing deblurring. We formulate the problem as

$$v_{\text{blur}} = B v_{\text{true}} + n_B \tag{5.8}$$

where v_{blur} denotes the reconstructed DBT images without source blur modeling, v_{true} is the sharp and clean images, $n_B \sim \mathcal{N}(0, \sigma_B^2 I)$ represents noise modeled as being additive Gaussian. This deblurring problem is nonblind because B is known.

When deblurring DBT images, it is crucial to control the image noise level through regularization due to the low-dose exposure of DBT scans. In this work, we investigated applying

DDPM as an image prior for regularizing the deblurring process. The upcoming sections first give a brief review of DDPM, and then introduce the proposed deblurring method with generative diffusion.

5.2.2.1 Denoising Diffusion Probabilistic Models (DDPM)

DDPM is a class of generative models that use a diffusion process to model complex probability distributions [150][160]. These are Bayesian methods that assume that the images of interest can be represented as random vectors characterized by some probability distribution. The 3D DBT images are typically viewed as slices parallel to the detector plane by radiologists in clinical practice, so we consider the 2D slices $x_{\text{true}} \in \mathbb{R}^{N_x N_y}$ taken from v_{true} as our images of interest with the associated distribution $p(x_{\text{true}})$. The diffusion process is a Markov chain that progressively adds noise to the image until a tractable distribution, such as a standard Gaussian, is achieved [150]. Mathematically, for a sequence of T diffusion steps, at each step $t = 1, \dots, T$,

$$x_t = \sqrt{1-\beta_t}\, x_{t-1} + \sqrt{\beta_t}\, \epsilon \tag{5.9}$$

where $\beta_t \in (0,1)$ is the prescribed noise variance schedule, $\epsilon \sim \mathcal{N}(0, I)$ is the standard Gaussian noise. By the design of DDPM, we have $x_0 = x_{\text{true}}$ and $x_T \sim \mathcal{N}(0, I)$. Equation (5.9) can also be written in a non-iterative form

$$x_t = \sqrt{\bar{\alpha}_t}\, x_0 + \sqrt{1-\bar{\alpha}_t}\, \epsilon \tag{5.10}$$

where $\bar{\alpha}_t = \prod_{t_0=1}^{t} (1-\beta_{t_0})$.

To reverse the diffusion process, DDPM trains a deep neural network $\epsilon_\theta(x_t, t)$ parameterized by θ to learn to predict the added noise from x_t. The training loss is a variant of a variational lower bound, and intuitively speaking, is the mean squared error between the predicted noise and actual added noise [150]

$$\underset{\theta}{\operatorname{argmin}}\, \mathbb{E}_{t,x_0,\epsilon}\left[\left\|\epsilon - \epsilon_\theta\left(\sqrt{\bar{\alpha}_t}\, x_0 + \sqrt{1-\bar{\alpha}_t}\, \epsilon, t\right)\right\|^2\right]. \tag{5.11}$$

Once $\epsilon_\theta(x_t, t)$ is trained, we can generate images by randomly initializing a sample with pure Gaussian noise $x_T \sim \mathcal{N}(0, I)$ and then iteratively removing noise from it following the DDPM sampling procedure [150]. In our implementation, we use a variant of DDPM sampling called denoising diffusion implicit models (DDIM) sampling [161]

$$x_{t-1} = \sqrt{\bar{\alpha}_{t-1}}\left(\frac{x_t - \sqrt{1-\bar{\alpha}_t}\, \epsilon_\theta(x_t, t)}{\sqrt{\bar{\alpha}_t}}\right) + \sqrt{1-\bar{\alpha}_{t-1}}\, \epsilon_\theta(x_t, t) \tag{5.12}$$

for $t = T, \ldots, 1$.

5.2.2.2 Image Deblurring with Generative Diffusion

The goal of image deblurring is to estimate the unknown sharp and clean images x_{true} from the observed corrupted images $x_{\text{blur}} \in \mathbb{R}^{N_x N_y}$ taken from the DBT volume v_{blur}. In Bayesian image deblurring, commonly used techniques include sampling from the posterior $p(x_{\text{true}}|x_{\text{blur}})$ and MAP estimation. Note that the prior $p(x_{\text{true}})$ is the distribution of true DBT slices, which can be effectively sampled by a well-trained DDPM using DBT images with no noise and blur.

We propose to perform posterior sampling to estimate x_{true} from x_{blur}. This requires us to modify the unconditional DDPM sampling to be a conditional sampling process. To do so, we exploit the connection between DDPM and score-based generative modeling following the derivation of Dhariwal and Nichol [162]. First, it has been shown that the DDPM network $\epsilon_\theta(x_t, t)$ approximates the gradient of the log probability, also called the score function, of the distribution $p(x_t)$

$$\nabla_{x_t} \log p(x_t) = -\frac{\epsilon_\theta(x_t, t)}{\sqrt{1-\bar{\alpha}_t}}. \tag{5.13}$$

To see this, recall that $p(x_t|x_0) \sim \mathcal{N}\left(\sqrt{\bar{\alpha}_t}\, x_0, (1-\bar{\alpha}_t)I\right)$ from (5.10), so

$$\nabla_{x_t} \log p(x_t|x_0) = -\frac{x_t - \sqrt{\bar{\alpha}_t}\, x_0}{1-\bar{\alpha}_t} = -\frac{(x_t - \sqrt{\bar{\alpha}_t}\, x_0)/\sqrt{1-\bar{\alpha}_t}}{\sqrt{1-\bar{\alpha}_t}} \approx -\frac{\epsilon_\theta(x_t,t)}{\sqrt{1-\bar{\alpha}_t}}$$

and thus, using the law of iterated expectation:

$$\nabla_{x_t} \log p(x_t) = \mathbb{E}_{p(x_0)}\left[\nabla_{x_t} \log p(x_t|x_0)\right] \approx \mathbb{E}_{p(x_0)}\left[-\frac{\epsilon_\theta(x_t,t)}{\sqrt{1-\bar{\alpha}_t}}\right] = -\frac{\epsilon_\theta(x_t,t)}{\sqrt{1-\bar{\alpha}_t}},$$

which gives (5.13). Then, the score function of the conditional distribution $p(x_t|x_{\text{blur}})$ becomes

$$\begin{aligned} \nabla_{x_t} \log p(x_t|x_{\text{blur}}) &= \nabla_{x_t} \log p(x_t) + \nabla_{x_t} \log p(x_{\text{blur}}|x_t) \\ &= -\frac{1}{\sqrt{1-\bar{\alpha}_t}} \underbrace{\left(\epsilon_\theta(x_t,t) - \sqrt{1-\bar{\alpha}_t} \cdot \nabla_{x_t} \log p(x_{\text{blur}}|x_t)\right)}_{\tilde{\epsilon}_\theta(x_t,t) :=} \end{aligned} \tag{5.14}$$

where Bayes rule gives $p(x_t|x_{\text{blur}}) = \frac{p(x_t)\, p(x_{\text{blur}}|x_t)}{p(x_{\text{blur}})}$ and the gradient of $\log p(x_{\text{blur}})$ with respect to x_t vanishes because it does not depend on x_t.

When t is small, the structural details of x_t are close to x_0, so we assume $x_{\text{blur}} \approx Bx_t + n_{B,t}$, $n_{B,t} \sim \mathcal{N}\left(0, \sigma_{B,t}^2 I\right)$. We omit the subscript of B to simplify the notation, but it should be clear that B here is the blur matrix for a slice rather than a volume. The gradient of the likelihood is therefore $\nabla_{x_t} \log p(x_{\text{blur}}|x_t) = -\frac{1}{\sigma_{B,t}^2} B'(Bx_t - x_{\text{blur}})$ where $'$ denotes matrix transpose. Finally, we insert this gradient into (5.14) and define the modified DDPM output

$$\tilde{\epsilon}_\theta(x_t,t) = \epsilon_\theta(x_t,t) + \frac{\sqrt{1-\bar{\alpha}_t}}{\sigma_{B,t}^2} B'(Bx_t - x_{\text{blur}}). \tag{5.15}$$

We use this function in DDPM sampling to draw samples from the posterior $p(x_{\text{true}}|x_{\text{blur}})$ instead of $p(x_{\text{true}})$. The sampling equation (5.12) now becomes

$$x_{t-1} = \sqrt{\bar{\alpha}_{t-1}}\left(\frac{x_t - \sqrt{1-\bar{\alpha}_t}\,\tilde{\epsilon}_\theta}{\sqrt{\bar{\alpha}_t}}\right) + \sqrt{1-\bar{\alpha}_{t-1}}\,\tilde{\epsilon}_\theta$$

$$= \sqrt{\bar{\alpha}_{t-1}}\left(\frac{x_t - \sqrt{1-\bar{\alpha}_t}\,\epsilon_\theta}{\sqrt{\bar{\alpha}_t}}\right) + \sqrt{1-\bar{\alpha}_{t-1}}\,\epsilon_\theta - \frac{\sqrt{1-\bar{\alpha}_t}}{\sigma_{B,t}^2}\left(\frac{\sqrt{1-\bar{\alpha}_t}}{\sqrt{1-\beta_t}} - \sqrt{1-\bar{\alpha}_{t-1}}\right)\cdot B'(Bx_t - x_{\text{blur}})$$

$$\approx \sqrt{\bar{\alpha}_{t-1}}\left(\frac{x_t - \sqrt{1-\bar{\alpha}_t}\,\epsilon_\theta}{\sqrt{\bar{\alpha}_t}}\right) + \sqrt{1-\bar{\alpha}_{t-1}}\,\epsilon_\theta - \lambda\cdot B'(Bx_t - x_{\text{blur}}) \quad (5.16)$$

where we isolate $B'(Bx_t - x_{\text{blur}})$ and replace its positive coefficient with a tuning parameter λ to control the balance between noise reduction and data fidelity. We drop the t-dependency of λ for easier tuning while still achieving empirically good results.

5.2.3 X-ray Source Motion Deblurring for Reconstructed DBTs

We have introduced a post-processing deblurring method for reconstructed DBT image slices through DDPM posterior sampling. We only run the sampling equation (5.16) for a small number of steps $t = \bar{T}, \ldots, 1$, where $\bar{T} \ll T$ to satisfy the small t assumption. To deblur the entire DBT volume, we do slice-by-slice deblurring.

The proposed deblurring method is applicable to DBTs obtained from any reconstruction method. In this study, we investigate applying it to the model-based image reconstruction with detector blur and correlated noise modeling (DBCN) approach [36]. By deblurring DBCN reconstructed images, the overall image reconstruction and post-processing pipeline represents a framework that employs both source motion blur and detector blur modeling.

5.3 Materials

5.3.1 DBT System Configuration

We focus our study on the Siemens Mammomat Inspiration DBT system that takes 25 PVs from -25° to 25° scan angles uniformly with $D_{SO} = 600$ mm and $D_{OD} = 50$ mm. The gap between the detector plane and the bottom of the compressed breast is 20 mm. The Siemens system uses a

continuous-motion x-ray source with a 0.18° source motion angle, and we modeled it as a rectangle of width $w_0 = 2.185$ mm, as described in Section 5.2.1.2. The detector pixel size is 0.085 mm × 0.085 mm.

5.3.2 Source Blur Modeling

5.3.2.1 Verification Study of Blur Kernel Modeling

We designed a simulation study to verify the in-plane blur kernel modeling and rectangular width estimation. Figure 5.3 shows the simulation workflow. First, we created an impulse object in the voxelized image with background value of zero. The impulse value was 0.05 mm^{-1}, close to the typical attenuation coefficient of breast tissues [163]. Then, we simulated the PVs using the point and blurry sources, denoted as y_{pt} and y_{blur}, respectively. The pixel values of y_{pt} were generated with the segmented separable footprint (SG) projector [23] instead of simple ray-tracing. The generation of y_{blur} used 50 SG projectors as sub-focal spots within w_0 uniformly. Because the blur occurs after the x-ray is attenuated in actual scans, we summed the sub-PVs in the pre-log domain to match the physics process and then took log to get the post-log blurry PVs. The simulation was noise-free.

Next, we reconstructed the PSFs with a point source SG projector for the following two conditions: $\min_v \frac{1}{2}\left\|Av - y_{\text{pt}}\right\|^2$ and $\min_v \frac{1}{2}\|Av - y_{\text{blur}}\|^2$. The reconstructed PSFs had spoke-like inter-plane artifact due to the limited-angle nature of DBT. We took the 1D PSFs through the impulse in y direction and took Fourier transform to obtain the impulse optical transfer functions (OTFs). Although the OTFs are complex-valued in general because the PSFs are asymmetric (except when the impulse is centered in y), their ratio, which represents the source blur OTF, is always real-valued.

Finally, recall that the Fourier transform of a rectangular function $\frac{1}{w}\text{rect}\left(\frac{x}{w}\right)$ of width w and unit area is a sinc function $\text{sinc}(wx)$. We fit the source blur OTF with $\text{sinc}(wx)$ to estimate the blur kernel width w and compared it with our analytically derived blur kernel width $\bar{w}_s$ defined in (5.5). We moved the impulse object to different heights above the detector and repeated the experiment.

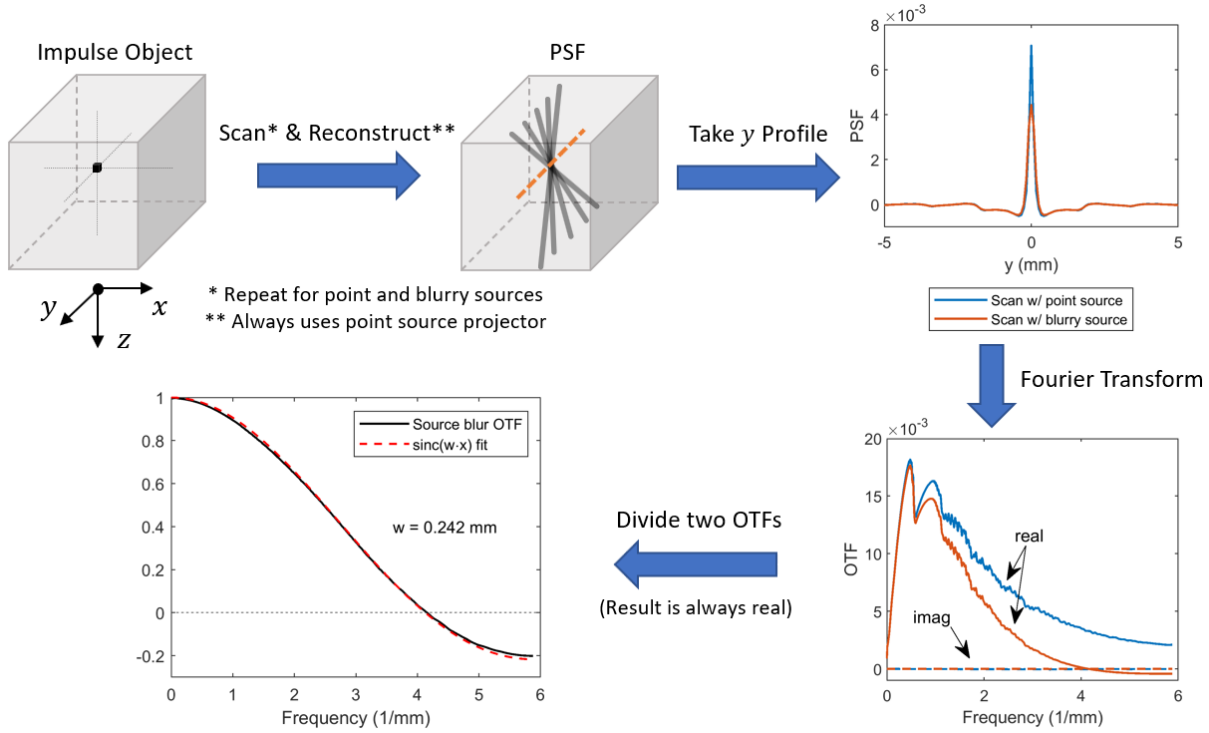


Figure 5.3 The simulation study to verify the in-plane blur kernel modeling. In this example, the impulse is centered in y direction, so its PSF and OTF are both real and symmetric. In general, the system OTF is complex-valued if the impulse PSF is asymmetric, but the source blur OTF is always real-valued.

5.3.2.2 Investigating Using AB or A_{sub} for DBT Reconstruction

We conducted a simulation study to demonstrate the limitations of using AB or A_{sub} as system matrices for DBT reconstruction. The simulation setup was the same as Section 5.3.2.1, where we created an impulse object and generated noise-free y_{pt} and y_{blur} PVs. We reconstructed the PSFs using gradient descent until convergence for the following four conditions:

(1) $\min_v \frac{1}{2}\left\|Av - y_{\text{pt}}\right\|^2$, (2) $\min_v \frac{1}{2}\|Av - y_{\text{blur}}\|^2$, (3) $\min_v \frac{1}{2}\|ABv - y_{\text{blur}}\|^2$, (4) $\min_v \frac{1}{2}\|A_{\text{sub}}v - y_{\text{blur}}\|^2$ with $N_{\text{sub}} = 50$. We extracted the reconstructed PSFs in y direction through the impulse and compared their OTFs.

5.3.3 Data Sets

5.3.3.1 VICTRE Phantoms

We used the Virtual Imaging Clinical Trial for Regulatory Evaluation (VICTRE) package [66] to create virtual phantoms with breast tissue backgrounds to train the DDPM network and test the deblurring methods. The VICTRE package can generate anthropomorphic breast phantoms that were used for virtual clinical trials. The PVs of the virtual phantoms were simulated by MC-GPU, a Monte Carlo x-ray imaging simulator in the package. MC-GPU was configured for the Siemens Mammomat Inspiration DBT system, and its simulation accuracy was validated in terms of noise and resolution [126]. It also provided an option of either using an ideal point source or a blurry source with 0.3 mm nominal size and 0.18° motion angle for the scans. The x-ray exposure was adaptively determined for each phantom by first running a quick scan with a small exposure, and then the full scan with a scaled exposure so that the mean glandular dose matched that of a real scan under automatic exposure control (AEC) [164]. Scatter was also simulated by MC-GPU.

For DDPM training, we created 70 virtual phantoms whose density and size characteristics are shown in Table 5.1. The glandular volume fraction (GVF) setting followed that of the VICTRE study [66]. The voxel size of the virtual phantoms was 0.05 mm × 0.05 mm × 0.05 mm. We used the point source and increased the exposure to be 5 times the AEC in MC-GPU so that the simulated images were sharp and less noisy to better represent the prior DBT image distribution $p(x_{\text{true}})$. We reconstructed the DBTs using DBCN (5 iterations, regularization strength = 70) [36] and the SG projector. The reconstructed image voxel size was 0.085 mm × 0.085 mm × 1.0 mm.

Due to the large sizes of DBT images, we trained the DDPM network using image patches instead of full DBT slices to reduce memory cost. We randomly extracted 128,401 64×64-pixel non-overlapping 2D slice patches from the reconstructed DBT images to form the DDPM training set.

To test the deblurring methods, we created four extra virtual phantoms, one for each breast density, and compressed them to 80 mm in thickness. The phantoms were embedded with a line pair test object discussed next. We scanned the phantoms twice under AEC exposure in MC-GPU, first using the blurry source, and then using the ideal point source to serve as a reference standard. We reconstructed the test phantoms using DBCN (5 iterations, regularization strength = 70) and the SG projector.

Table 5.1 Density and size characteristics of the virtual breast phantoms for DDPM training.

Density	Almost entirely fatty	Scattered fibroglandular dense	Heterogeneously dense	Extremely dense
GVF	5%	15%	34%	60%
No. of Phantoms	10	10	25	25
Thickness (mm)	52-70 (2 mm intervals)	46-64 (2 mm intervals)	36-60 (1 mm intervals)	31-55 (1 mm intervals)

5.3.3.2 Line Pair Test Object and Image Quality Metrics

To quantitatively evaluate the image resolution, we designed a test object consisting of line pairs (LPs) with a range of spatial frequencies, as shown in Figure 5.4(a). The test object was 35 mm × 35 mm × 0.2 mm in size and voxelized with size of 0.05 mm × 0.05 mm × 0.05 mm. The LPs were 0.2 mm in thickness and 2 mm in length and were made of calcium oxalate to mimic the attenuation of small MCs. Each LP group contained five horizontally placed bars with equal width and spacing. From top to bottom of the test object, the four bar width settings were 0.1 mm, 0.15 mm, 0.2 mm, and 0.25 mm. These were equivalent to the spatial frequencies of 5 LP/mm, 3.33 LP/mm, 2.5 LP/mm, and 2 LP/mm, respectively. The line pair groups were placed 10 mm apart in the vertical directions. Since the alignment between the test object and reconstruction voxel grid

could affect the resolution of the reconstructed LPs, we duplicated the LPs in five columns from left to right of the test object with an accumulation of 0.05 mm vertical shift, simulating the possible random alignments of small object with the detector pixel array during imaging. The columns of the line pair groups were 7 mm apart in the horizontal direction.

We inserted the LP test object into the virtual test phantoms such that it was parallel to the detector and the bars were perpendicular to the source motion direction. The object was against the chest wall and centered in the y direction. Figure 5.4(b) shows an example reconstructed DBT slice containing the test object.

We calculated the LP contrast and image noise of the reconstructed DBT images as image quality metrics. For each reconstructed LP, we averaged the central 1.5 mm bar region along x direction to get the LP profile in y direction. Then, we overlapped the LP profile with the ground truth locations of the five bars and four spaces on the continuous y coordinate, as illustrated in Figure 5.4(c). The reconstructed LP profile was linearly interpolated (lines connecting adjacent pixel values). If the reconstructed LP was well-resolved, it should have peaks and valleys matching the corresponding locations in the ideal profile. We followed Zheng *et al*. [133] and defined the contrast of the reconstructed LP as the difference between the mean value of the five ground truth bar regions (gray area in Figure 5.4(c)) and the mean value of the four space regions (yellow area in Figure 5.4(c)), normalized to the contrast of the input ideal profile. Since the input ideal profile had the same contrast for all LP frequencies, the relative contrast of the LP frequencies would be the same with or without normalization. The final contrast of each LP frequency was averaged over the five shifted instances in the test object. To quantify image noise, we took 20 10×10-pixel LP-free regions of interest (ROIs) near the LPs and calculated the root-mean-square pixel variation

of the ROIs after background removal using quadratic fitting (see Chapter 4). The overall image noise level was the average over all noise ROIs.

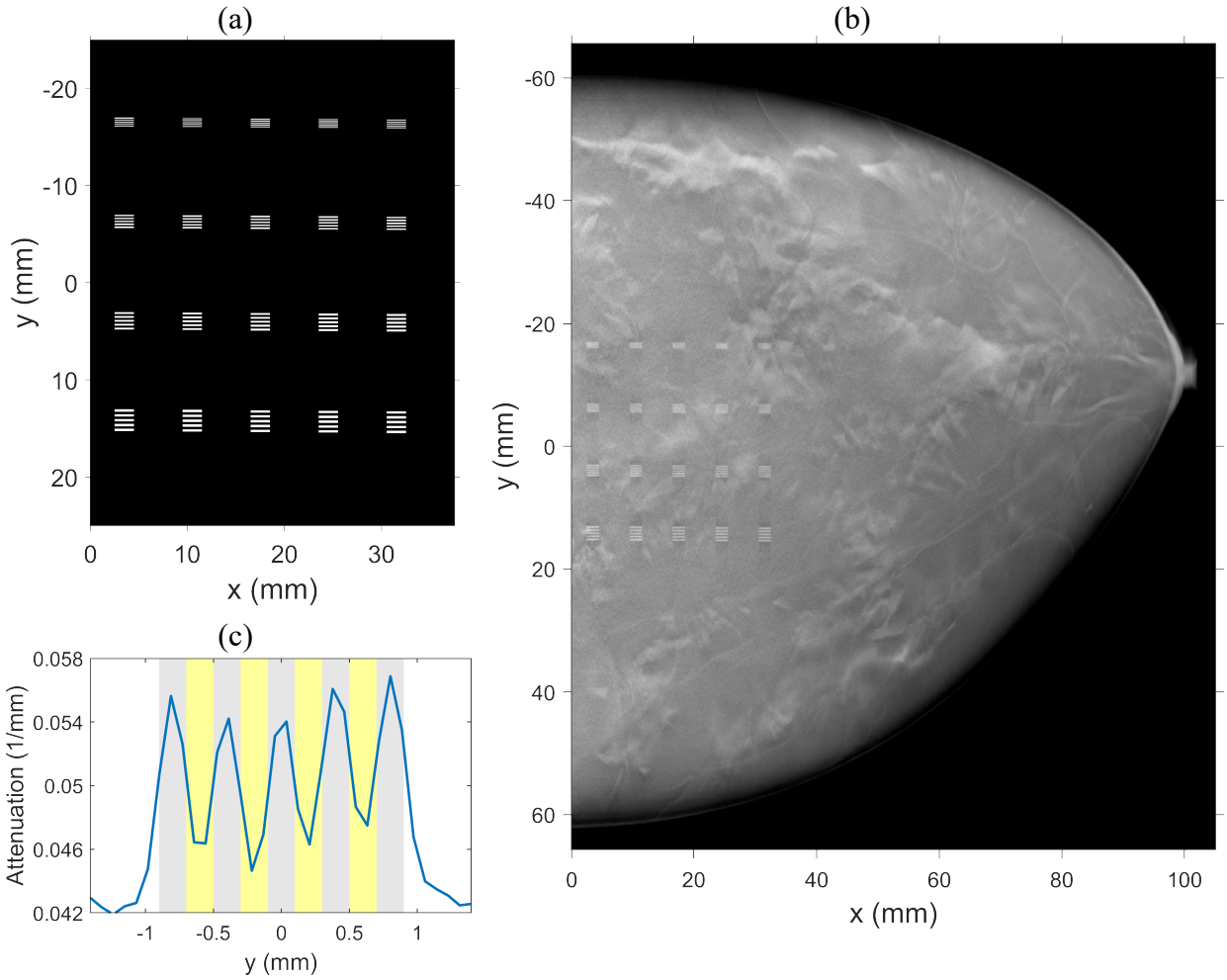


Figure 5.4 (a) LP test object with 5 LP/mm, 3.33 LP/mm, 2.5 LP/mm, and 2 LP/mm (top to bottom) and vertical shifts (left to right). (b) Example reconstructed DBT slice with embedded LP test object. (c) Illustration of LP contrast calculation (gray area: bar regions; yellow area: space regions).

5.3.4 DDPM Implementation

The structure of the DDPM network was a modified U-Net [165] as described in [150]. The U-Net had four downsampling scales, each with three ResNet blocks [124]. The numbers of 3×3 convolutional kernels for the downsampling scales were 64, 128, 128, 128, respectively. Following the DDPM paper, we trained the DDPM using 60 training epochs, a batch size of 256,

and a learning rate of 10^{-4} with Adam optimizer [30]. The noise variance schedule β_t was evenly spaced between $\beta_0 = 10^{-4}$ and $\beta_T = 0.02$ with T = 1000 [150]. The diffusion steps t were encoded with sinusoidal positional encoding [166] and then added to the feature maps of the ResNet blocks. We removed all the attention layers, so the network contained only convolution and up/downsampling layers, and thus could sample images of arbitrary sizes. Section 5.4.3.1 discusses the parameter selection of $\bar{T}$ and λ in the DDPM posterior sampling for deblurring.

5.3.5 Comparison Methods

We compared the proposed deblurring method with the following nonblind deblurring methods: Tikhonov regularized deblurring, total variation (TV) regularized deblurring, and the unfolding super-resolution network (USRNet) [167]. Tikhonov regularized deblurring was formulated as $\hat{v} = \underset{v}{\operatorname{argmin}} \frac{1}{2}\|Bv - v_{\text{blur}}\|^2 + \frac{\mu_{\text{Tik}}}{2}\|v - v_{\text{blur}}\|^2$, which has an analytical solution that uses an inverse filter: $\hat{v} = (B'B + \mu_{\text{Tik}}I)^{-1}(B' + \mu_{\text{Tik}}I)v_{\text{blur}}$. TV regularized deblurring was formulated as $\hat{v} = \underset{v}{\operatorname{argmin}} \frac{1}{2}\|Bv - v_{\text{blur}}\|^2 + \mu_{\text{TV}}\|Dv\|_1$ where D contains the finite difference operators in x and y directions. USRNet was an end-to-end trainable CNN, and we trained it using the paired virtual phantom images with low-quality images being the simulated 1× AEC exposure using a blurry source and high-quality images being the simulated 5× AEC exposure using a point source.

5.4 Results

5.4.1 Source Blur Modeling

5.4.1.1 Verification Study of Blur Kernel Modeling

Figure 5.5(a) shows the example impulse OTFs and source blur OTFs in the simulation study of the in-plane blur kernel modeling. The impulse was placed against the chest wall at different heights z_s. The impulse was centered in y direction, so its PSF and OTF were both real and symmetric. Figure 5.5(b) shows the scatter plot of the sinc-fit estimated blur kernel width versus z_s. The data points exhibited a good linear relationship (correlation coefficient = 0.998, $p < 0.0001$). The linear fit of the data points had an almost perfect alignment with the analytically calculated kernel width $\bar{w}_s$. We also experimented with other reconstructions or moved the impulse to different x and y locations, and always obtained the same kernel widths with difference smaller than 5%, confirming that the blur kernel only depended on the height of the impulse above the detector. This alignment strongly supported our modeling approach, which treated source motion blur as a 1D rectangle for the reconstructed DBT slices.

The shaded region in Figure 5.5(b) shows the range of w_s, the rectangle widths from individual scan angles. For the Siemens Mammomat Inspiration DBT system that acquired 25 PVs from -25° to 25°, w_s differed from $\bar{w}_s$ by -6.7% at 0° (lower bound in Figure 5.5(b)) and 12.7% at 25° (upper bound in Figure 5.5(b)). In other words, the variation of w_s was small compared to the mean $\bar{w}_s$. Together with the good alignment mentioned before, this observation justified our simplification of blur kernel by averaging the rectangle widths over all scan angles, resulting in a shift-invariant 1D kernel over the area of a reconstructed DBT plane at a given z_s.

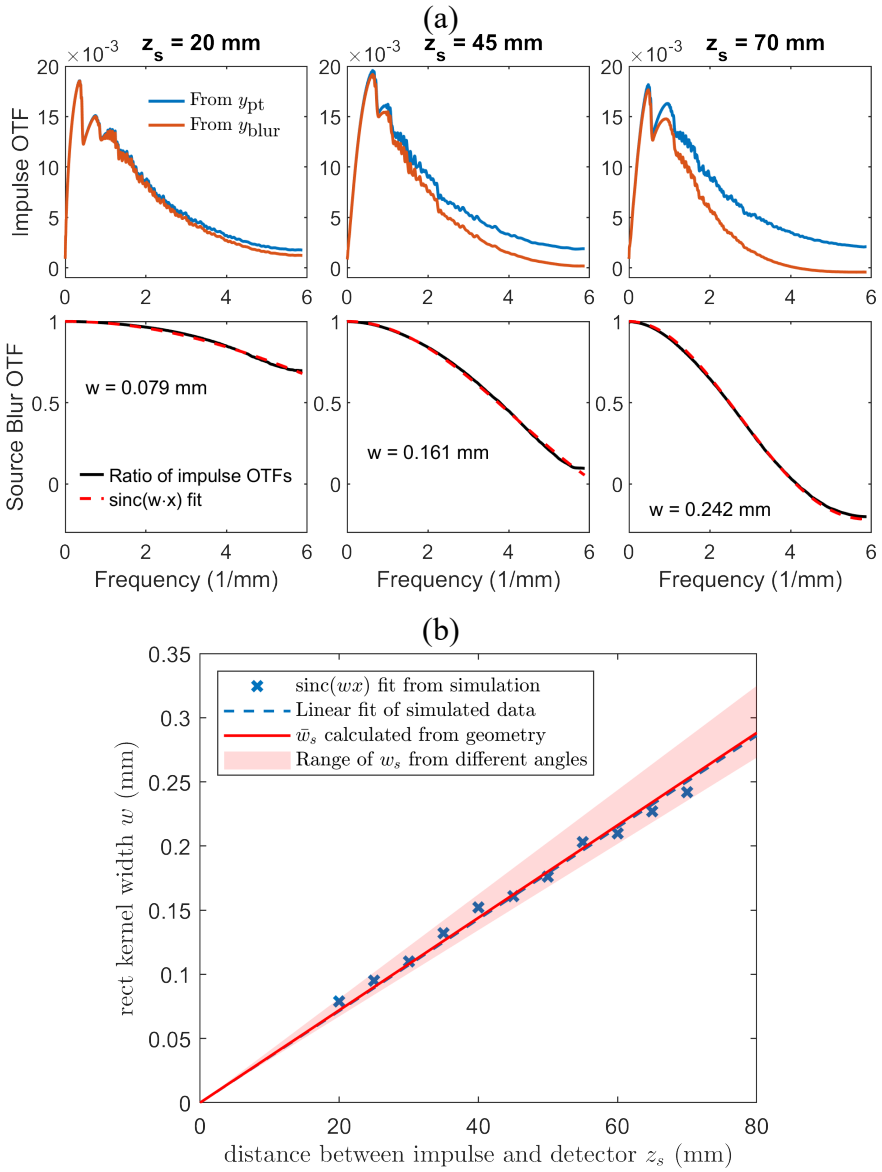


Figure 5.5 (a) Example impulse OTFs and source blur OTFs for three distances (z_s) between the impulse and the detector in the simulation study of blur kernel modeling. (b) The comparison of sinc-fit estimated and analytically calculated blur kernel widths. The shaded region shows the range of w_s values, the rectangle widths from individual scan angles.

5.4.1.2 Limitations of Using AB or A_{sub} for DBT Reconstruction

Figure 5.6 shows the impulse OTFs for the four image conditions outlined in Section 5.3.2.2. The impulse was placed against the chest wall and centered in y at z_s = 70 mm. The decrease in OTF from condition (1) to condition (2) was due to the source motion blur. The ratio of these two OTFs corresponded to the smooth source blur OTF that was fitted by sinc

functions in Section 5.4.1.1. Condition (3) and (4) demonstrated that both AB and A_{sub} were able to correct the negative phases at high frequency bands caused by the blur. However, the OTF magnitudes were not significantly improved and remained considerably lower than condition (1). This result suggested that straightforwardly using AB or A_{sub} for DBT reconstruction cannot recover the loss in resolution caused by source motion blur.

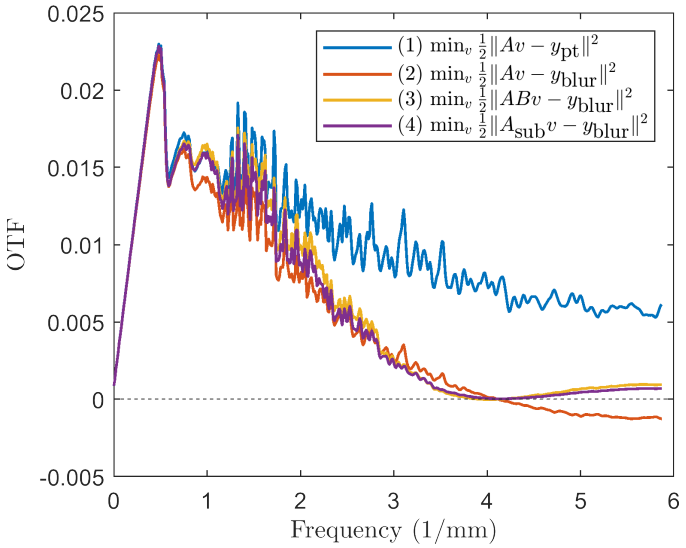


Figure 5.6 The impulse OTFs that demonstrate the limitations of using AB or A_{sub} to model source motion blur for DBT reconstruction. The conditions (1) to (4) are defined in Section 5.3.2.2.

5.4.2 DDPM Unconditional Image Generation

To demonstrate the ability of DDPM to produce high-quality DBT images, we ran unconditional DDPM sampling to draw samples from the prior distribution $p(x_{\text{true}})$. Figure 5.7(a) shows an example DBT slice from the DDPM training set. Figure 5.7(b) shows an example of DDPM generated sample. The DDPM generated image had natural heterogeneous background textures resembling the characteristics of the training image and was free from artifacts. The structural noise power spectrum (NPS) [10] of the DDPM generated image exhibited a power-law form as for mammograms [81] and was close to the NPS of the training image, as shown in Figure 5.7(c).

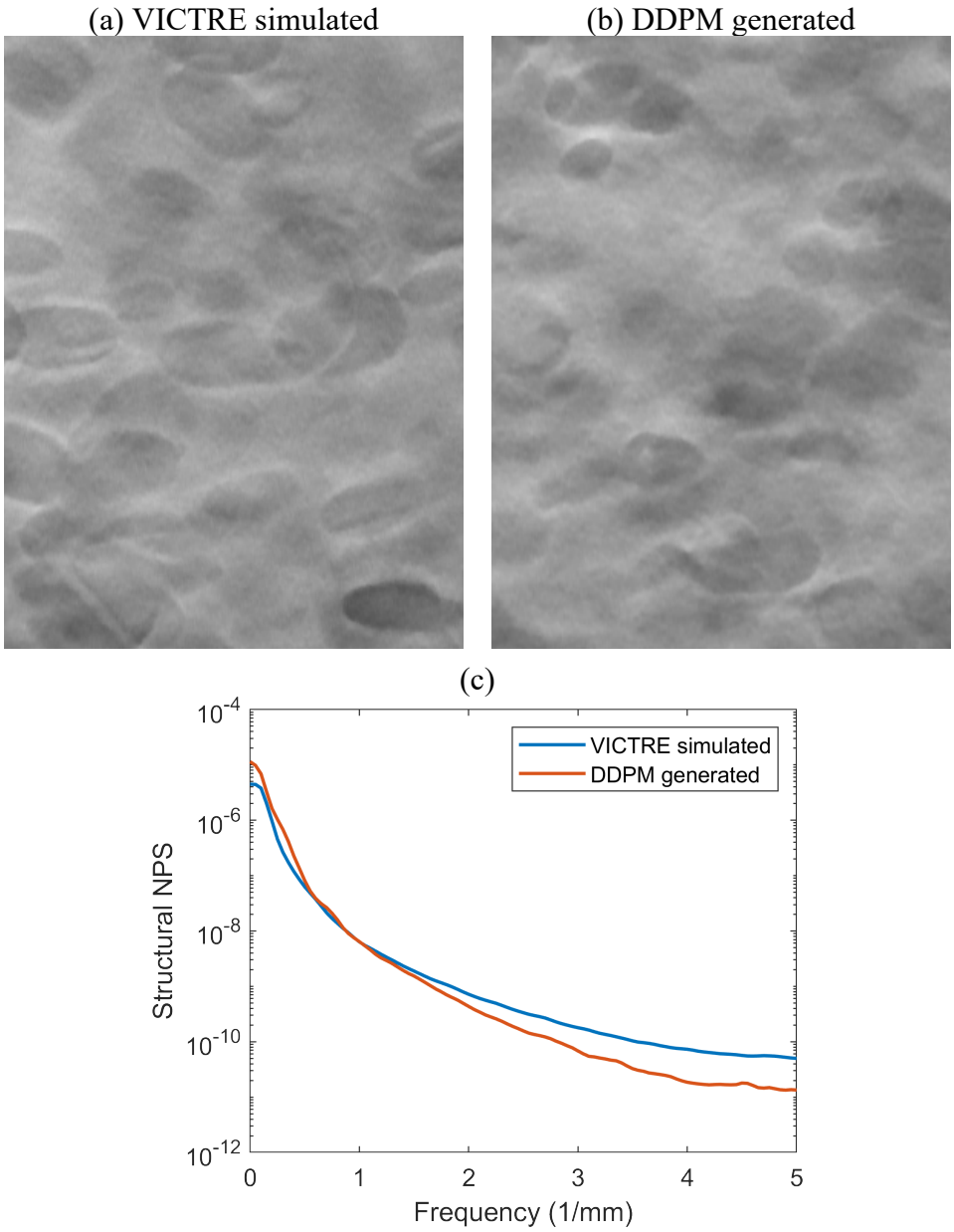


Figure 5.7 (a) DBT slice from the virtual phantoms used in the DDPM training set. (b) DBT slice generated by unconditional DDPM sampling. Image sizes are 300×400 pixels (25.5 mm × 34 mm). (c) Structural NPS of the two example images.

5.4.3 Image Deblurring with Generative Diffusion

5.4.3.1 Parameter Selection of DDPM Posterior Sampling

To select the number of sampling steps $\bar{T}$ and the weight parameter λ in DDPM posterior sampling, we did a grid search by varying $\bar{T}$ = 5, 10, 20, 50, and λ = 0.0, 0.2, 0.4, 0.8, 1.4. We

positioned the LP test object at z_s = 70 mm in the scattered test phantom and deblurred its DBCN reconstructed image using these parameter settings. Figure 5.8 shows the contrast-vs-noise plots of the LPs. Notice that the LPs with 5 LP/mm had severe blurring due to their narrow spacing. Deblurring could not recover their resolution, resulting in always negative contrasts. Therefore, we focused our attention on LP frequencies lower than 5 LP/mm.

When λ = 0.0, the deblurring method simplified to unconditional DDPM sampling. In this situation, the blur of LPs became more severe as $\bar{T}$ increased. As we increased λ, the LP contrast improved due to the high frequency boosting of the likelihood gradient. However, this enhancement also amplified background noise. Additionally, the enforcement of data fidelity at each sampling step became stronger, making the impact of $\bar{T}$ less apparent. To balance between contrast enhancement and noise control, we selected λ = 0.4 and $\bar{T}$ = 20 for the subsequent sections of this study. This parameter setting improved the image resolution after deblurring while maintaining the same image noise level as the blurry input.

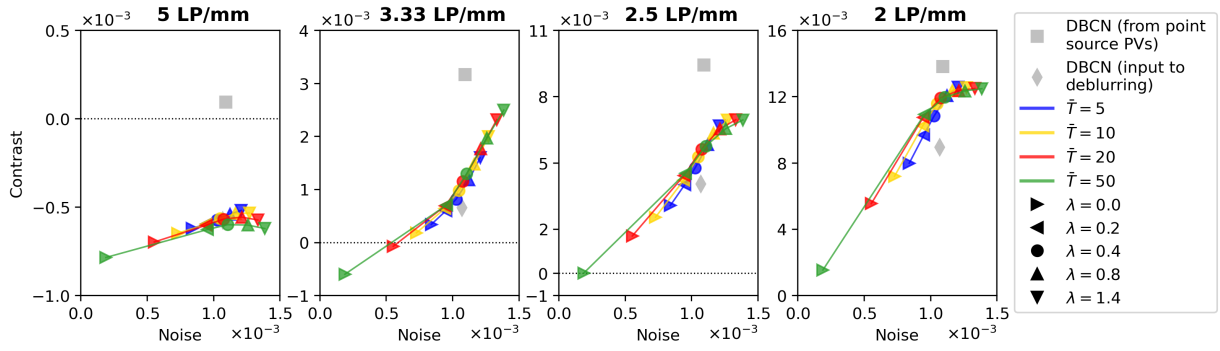


Figure 5.8 Contrast-vs-noise plots of the LP test object showing the dependence on the parameters of the proposed deblurring method with generative diffusion.

5.4.3.2 Effect of Breast Densities on Deblurring

To investigate the effect of breast density on the deblurring performance, we applied the proposed deblurring method to the four test phantoms. All the test phantoms were 8 cm thick but with distinct GVF settings: 5%, 15%, 34%, and 60%. We placed the LP test object at 70 mm above

the detector in these phantoms. Figure 5.9 shows the contrast-vs-noise plots of the LPs for the DBCN images and the deblurred images. As GVF increased, the image noise increased due to the AEC mechanism in the MC-GPU simulation. In particular, the same breast thickness of the four test phantoms resulted in the same mean glandular dose adjusted by the AEC. Hence, dense breasts absorbed more x-rays and had fewer transmitted x-rays, leading to higher image noise. However, the proposed deblurring method with generative diffusion consistently improved the LP contrasts, demonstrating its robustness and flexibility to handle various image noise levels and breast densities. While we mainly used the scattered dense test phantom with GVF = 15% in other sections of this study, our findings affirmed the applicability of the proposed method to a broader range of breast densities.

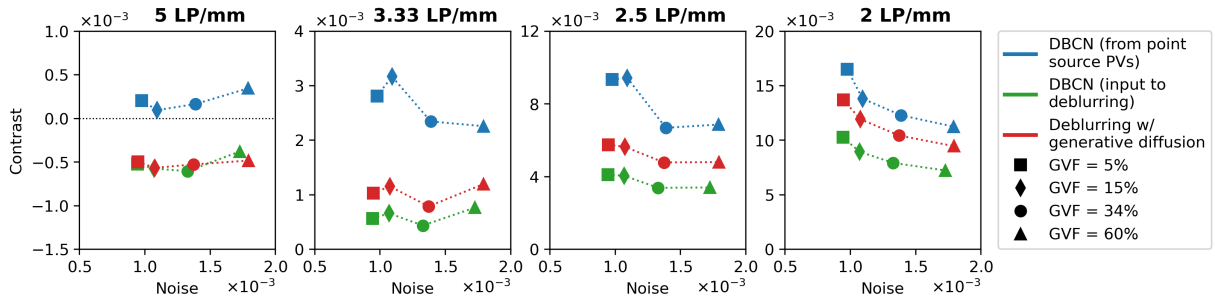


Figure 5.9 Contrast-vs-noise plots of the LP test object demonstrating the effect of breast density on the performance of the proposed deblurring method with generative diffusion.

5.4.3.3 Effect of Test Object Heights above Detector

Due to the increased geometric unsharpness, the x-ray source blur increased as the DBT slice became closer to the x-ray source. To assess its impact on the proposed deblurring method, we placed the LP test object at z_s = 50 mm, 70 mm, and 90 mm in the scattered dense test phantom. The magnification factors of these slices were 1.08, 1.12, and 1.16, respectively. Figure 5.10 shows the contrasts versus LP frequency for the DBCN images and the deblurred images. These contrast-vs-frequency plots resembled the modulation transfer function (MTF) curves commonly used in

assessing radiographic systems, albeit with the signals represented by rectangular LPs instead of sine waves. Also, our LPs were made with calcium oxalate instead of lead in the MC-GPU simulation.

The DBCN images reconstructed from point source PVs served as the reference standard, where the LP contrasts remained almost the same irrespective of z_s due to the absence of source blur. The DBCN images reconstructed from blurry source PVs had a reduction in LP contrasts compared to the reference standard. This reduction was more pronounced when z_s increased. The proposed deblurring method with generative diffusion successfully enhanced the LP contrasts and improved the image spatial resolution across all three conditions. Nevertheless, there remained room for improvement with respect to the reference standard, especially for the challenging scenarios where the test object was closer to the x-ray source. The deblurring method was not effective for LPs that were entirely blurred and had nearly zero or negative contrasts.

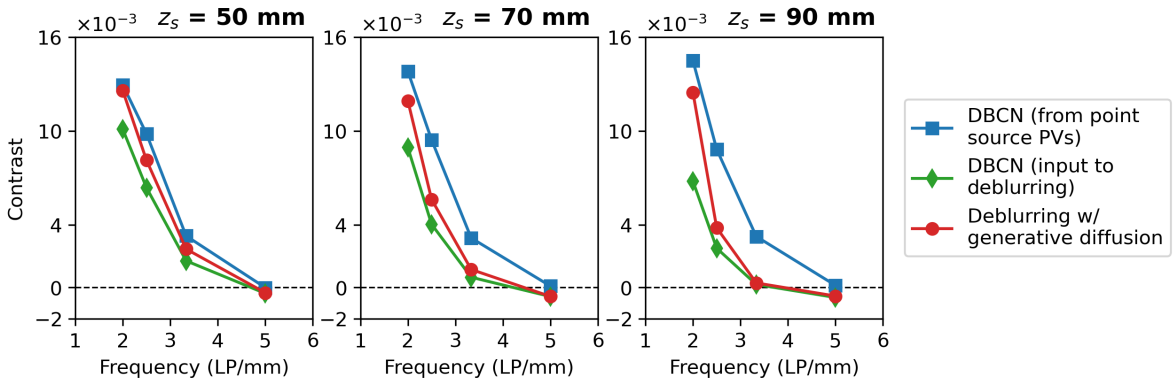


Figure 5.10 Contrast-vs-frequency plots of the LP test object demonstrating the effect of the object height above detector z_s on the performance of the proposed deblurring method with generative diffusion.

5.4.3.4 Comparison of Deblurring Methods

We evaluated the performance of different deblurring methods on the DBCN-reconstructed images using the scattered dense test phantom with the LP test object placed at z_s = 70 mm. Figure 5.11 shows the contrast-vs-noise plots and Figure 5.12 displays the example LP ROIs.

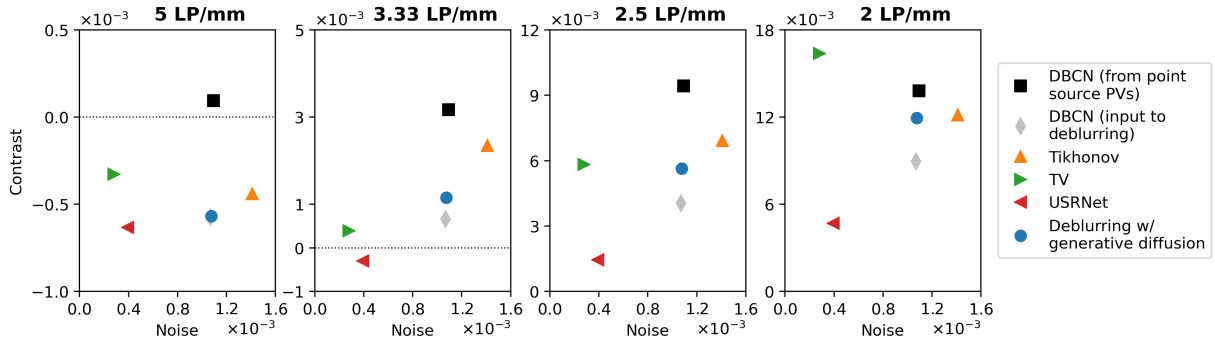


Figure 5.11 Contrast-vs-noise plots of the LP test object for comparing the deblurring performance of different methods on the DBCN-reconstructed images.

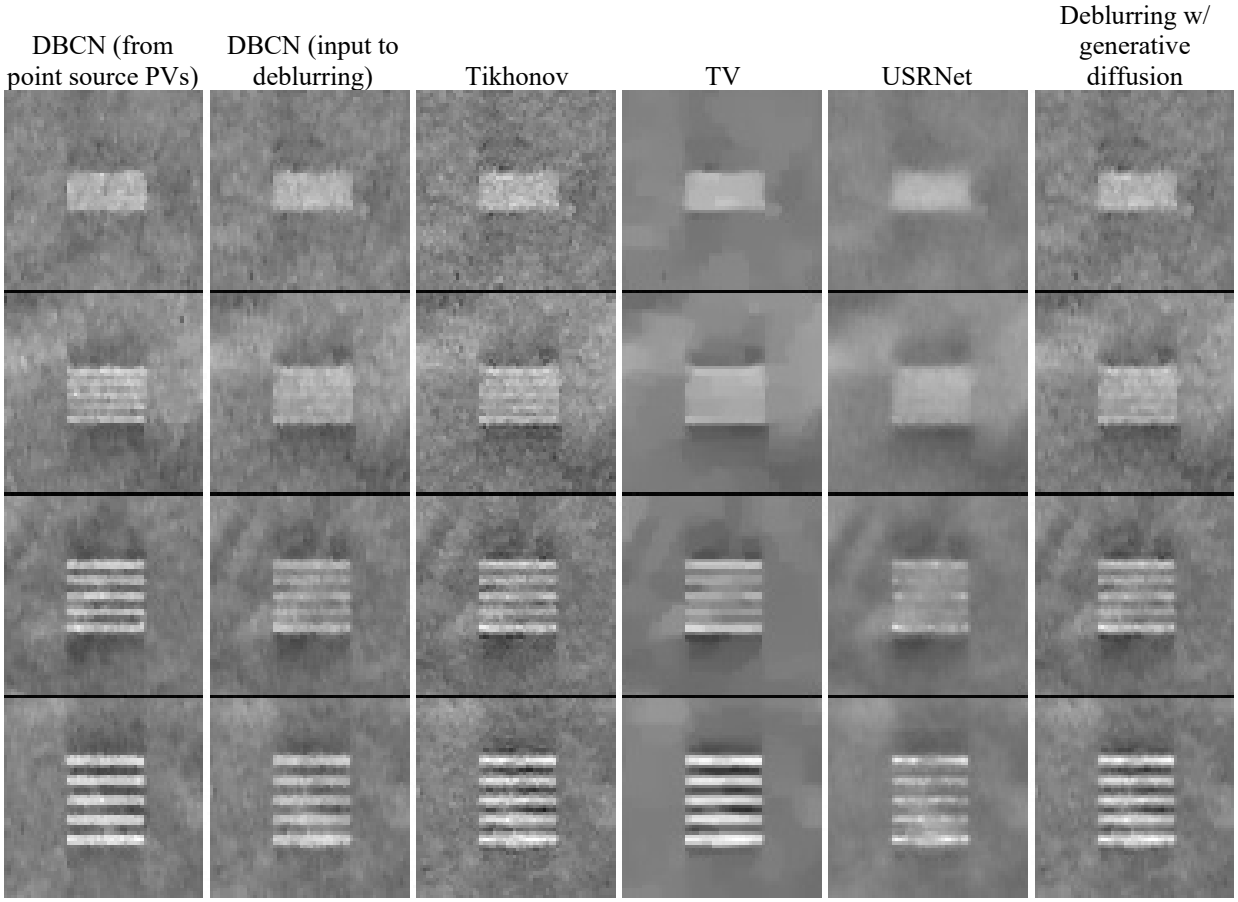


Figure 5.12 Example ROIs of the LP test object in the test phantom for comparing different deblurring methods on the DBCN-reconstructed images. From top to bottom, the LPs have spatial frequencies of 5 LP/mm, 3.33 LP/mm, 2.5 LP/mm, and 2 LP/mm. The ROI sizes are 60×60 pixels (5.1 mm × 5.1 mm).

Tikhonov regularized deblurring involved an inverse filter that inevitably enhanced the LP contrast and image noise at the same time. We tried a range of μ_{Tik} values and set it to 0.05. TV regularization ($\mu_{\text{TV}} = 0.0005$) performed very poorly on breast images, mainly because TV caused piecewise-constant patchy artifacts and could not characterize the ill-defined boundaries of soft tissue. While USRNet effectively smoothed the images, it failed to preserve the LP signals in the images. The proposed deblurring method with generative diffusion achieved an improvement in the LP contrast while maintaining a similar image noise level as the DBCN images. According to our visual judgement, it also retained the natural appearance of the tissue background without introducing artifacts.

5.5 Discussion

The DDPM network uses an unsupervised training approach that solely requires high-quality images. To apply DDPM for deblurring, we simply need to integrate the likelihood gradient into the DDPM sampling process, requiring no re-training or fine-tuning of the DDPM network. This unique feature endows DDPM much flexibility in terms of training data preparation because there is no need for paired low-quality and high-quality images. Moreover, it also makes the DDPM regularization very robust in that a single trained network can be applied to not only deblurring, but also other image restoration tasks as long as the specific task can be defined by a degradation operator like the blur matrix B in (5.8).

There are no LPs or bar patterns in our simulated training images as they do not exist in actual patients. Nonetheless, the LPs simulated closely spaced objects with sizes and x-ray attenuation in the range of microcalcifications along the source motion direction to allow quantitative evaluation of contrast and spatial resolution in the DBT planes. The proposed deblurring method with generative diffusion was able to preserve the LP test objects and enhance

their contrasts by using the data likelihood term. The use of multiple diffusion steps in each deblurring process also ensures more gradual alterations of image content. In contrast, the end-to-end trained USRNet processed the images in a single step, resulting in an abrupt change in image content and a failure to retain the LPs. This advantage is crucial, especially considering that MC signals in DBT images are sparse and small so they may be difficult for a network to learn. Deblurring by DDPM posterior sampling may help preserve the signals of interest such as MCs when using a potentially biased network or when a discrepancy exists between the training data and test data.

The limited-angle nature of DBT poses complexity and challenges for DBT image reconstruction. As illustrated by our simulation study in Section 5.4.1.1 and Section 5.4.1.2, although the impulse OTF was reconstructed from the point source PV using the point source projector, it still largely deviated from the ideal OTF (a horizontal line with a value of one). It also had large oscillations even in the absence of noise. As a result, incorporating the source motion modeling directly into an iterative reconstruction method is far from trivial. We made a compromise by deblurring the reconstructed images and achieved moderate improvement in image sharpness. Post-processing deblurring has the advantage that it is applicable to DBT obtained from any reconstruction techniques. Nonetheless, post-processing deblurring may not be the optimal solution because the measured PVs are not exploited in the deblurring step. Future research is required to further improve the image resolution, especially for the image slices closer to the x-ray source where the blur is more severe.

We trained and tested the deblurring method using VICTRE simulated images. Besides the realism of the VICTRE phantoms, one of the main advantages of using simulation data is the availability of DBT scans from an ideal x-ray source. The point source DBT images can be used

as ground truth, which are otherwise impossible to obtain from real scans, for either network training or algorithm evaluation. Our deblurring method has not been tested with real patient images due to the unavailability of data. Future work should apply this method to real patient DBT images with x-ray source motion blur to evaluate its effectiveness in clinical scenarios.

5.6 Conclusion

In this study, we introduced a new approach for modeling x-ray source motion blur in DBT imaging. We derived the in-plane source blur kernel for the reconstructed DBT slices based on imaging geometry and showed that it could be approximated by a shift-invariant point spread function over the DBT plane at a given depth. We conducted a simulation study to validate its accuracy. Our simulation also underscored the limitations of modifying the system matrix to model source blur in DBT reconstruction, whether by incorporating the source blur matrix or introducing subsampling focal spots. In view of these limitations, we proposed a post-processing deblurring method with generative diffusion for the reconstructed DBT images using the known blur kernel. The quantitative results demonstrated that our deblurring method improved spatial resolution while maintaining the same level of image noise when applied to DBT images reconstructed with detector blur and correlated noise modeling. Future research can explore further refinements of the deblurring technique and investigate its application to human subject data for improving the diagnostic accuracy of DBT imaging.

Chapter 6
Future Work

This chapter suggests possible directions for future research. The discussion here takes a broader perspective, whereas the Discussion sections in Chapter 3, Chapter 4, and Chapter 5 delve into chapter-specific limitations and areas for improvement in more details.

6.1 Computerized Detection for Image Quality Evaluation

Medical images should be evaluated based on their specific utility for clinical tasks. In Chapter 4, we introduced CNN-MC as a surrogate to human observers to assess DBT images for the task of MC cluster detection. CNN-MC works on image patches and is able to classify whether a given patch contains clustered MCs or not. Besides MCs, masses and architectural distortions are also essential in breast cancer diagnosis. Alternative methods, such as computerized lesion detection programs for computer-aided detection (CAD), can be developed to encompass the broader spectrum of breast lesions and used to study the tradeoff between their detectability and FPs. Unlike CNN-MC, CAD systems involve free search and localization of the lesions in the entire images, a process more closely resembling clinical readings by radiologists. The results of CAD are often analyzed using the free response receiver operating characteristic (FROC) as opposed to the conventional ROC methodology.

In an initial pilot study, we explored the use of a CAD program for MC cluster detection developed by Samala *et al.* [168] to guide the development of image reconstruction. It detected

MC clusters using the following steps: identifying small objects in the DBT image, forming clusters, reducing FP clusters, and reporting the detected clusters. Unfortunately, the program could not distinguish the image conditions because it had not been trained to distinguish true MCs from various types of noisy background and MC-like artifacts generated by the reconstruction methods. Despite these limitations, the idea of using CAD for image evaluation remains of interest. As CAD programs mature and offer improved accuracy, we plan to revisit their utilization in the future. We will also extend beyond MC detection and include masses and architectural distortions in our analysis.

6.2 DCNN Training with Task-based Losses

In our current approaches, the training of DCNNs is handled separately from image evaluation. Specifically, the DNGAN in Chapter 3 and Chapter 4 are trained using MSE and adversarial losses. These training objectives, however, do not reflect the performance of clinical tasks. The task-based image quality metrics such as CNN-MC are only used for final evaluation after the DCNNs are trained. A promising direction for future research is the integration of task-based metrics into DCNN training, thereby establishing a direct link between DBT image processing and the high-level clinical tasks. For example, it could be desirable to optimize the DCNN weights with the explicit objective of maximizing the AUC of MC cluster detection. Importantly, the CNN-MC is composed of convolutional layers and is amendable for backpropagation during DCNN training.

6.3 Potential Dose Reduction for DBT Scans

The DBT images we have worked on so far are collected using the standard exposure (STD) setting. We showed that the DNGAN and the MDR were able to improve image quality for

this scan setting. To adhere to the principle of as low as reasonably achievable (ALARA) in medical imaging with ionizing radiation, we are interested in reducing the DBT exposure below the current clinical standard, by reducing either the exposures of all views or the number of views, while ensuring that the image quality remains at an acceptable level. Dose reduction is also one of the main goals of the NIH grant that funded my PhD research. We want to determine whether either MDR or post-denoising can permit a dose reduction compared to the current STD setting without reducing the detectability of subtle cancers such as those manifested as MCs. Given the ethical constraints that prevent us from scanning patients using experimental low-dose techniques, we can either digitally add noise to the STD scans to simulate lower doses, or we can employ advanced simulation software such as VICTRE and MC-GPU. We will need to validate the realism of these synthetic scans, especially the synthetic soft-tissue lesions such as spiculated masses or architectural distortion, before developing image processing algorithms for further analysis.

6.4 Limited-angle Reconstruction with Generative Models

The DBT imaging system matrix has a nonempty null space due to its inherent underdetermined nature. Furthermore, the limited-angle design and incomplete sampling of DBT scans make the null space even larger and intensify the issue. This means there are infinitely many viable reconstructions that satisfy the measured PVs. When using iterative reconstruction algorithms, the images tend to have inter-plane artifacts or leakages across the DBT image slices due to the forward and backward projection operations. These degradations not only cause discrepancies between the reconstructed images and the actual objects, but also cause undesired signal bleeding that substantially reduces the contrasts of MCs.

Recent advancements in generative artificial intelligence, particularly in the domain of image generation, offer a promising direction for addressing the challenges associated with

limited-angle scans. The core idea involves leveraging generative models to synthesize and supplement the missing information of the DBT system matrix, thus mitigating the inter-plane artifacts. To achieve this, we can make use of VICTRE that is capable of emulating complete breast phantom sampling akin to a breast CT scan. The resulting fully sampled breast images can serve as ground truth for training the generative models. Then, the trained generative models can be used as regularization during the reconstruction. Their role is to encourage more faithful reconstruction with fewer inter-plane artifacts and signal bleeding, therefore enhancing the fidelity and quality of the reconstructed DBT images.

Bibliography

[1] H. Sung, J. Ferlay, R. L. Siegel, M. Laversanne, I. Soerjomataram, A. Jemal, and F. Bray, "Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries," *CA Cancer J Clin*, Feb. 2021, DOI: 10.3322/caac.21660.

[2] American Cancer Society, "Cancer facts & figures," 2023. [Online]. Available: https://www.cancer.org/content/dam/cancer-org/research/cancer-facts-and-statistics/annual-cancer-facts-and-figures/2023/2023-cancer-facts-and-figures.pdf.

[3] F. J. Gilbert, L. Tucker, and K. C. Young, "Digital breast tomosynthesis (DBT): A review of the evidence for use as a screening tool," *Clin Radiol*, vol. 71, no. 2, pp. 141–150, Feb. 2016, DOI: 10.1016/j.crad.2015.11.008.

[4] 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.

[5] Y. Gao, L. Moy, and S. L. Heller, "Digital breast tomosynthesis: Update on technology, evidence, and clinical practice," *RadioGraphics*, vol. 41, no. 2, pp. 321–337, Mar. 2021, DOI: 10.1148/rg.2021200101.

[6] E. F. Conant, M. M. Talley, C. R. Parghi, B. C. Sheh, S.-Y. Liang, S. Pohlman, A. Rane, Y. Jung, L. A. S. Stevens, J. K. Paulus, and N. Alsheik, "Mammographic screening in routine

practice: Multisite study of digital breast tomosynthesis and digital mammography screenings," *Radiology*, vol. 307, no. 3, May 2023, DOI: 10.1148/radiol.221571.

[7] H.-P. Chan, M. A. Helvie, L. Hadjiiski, D. O. Jeffries, K. A. Klein, C. H. Neal, M. Noroozian, C. Paramagul, and M. A. Roubidoux, "Characterization of breast masses in digital breast tomosynthesis and digital mammograms: an observer performance study," *Acad Radiol*, vol. 24, no. 11, pp. 1372–1379, Nov. 2017, DOI: 10.1016/j.acra.2017.04.016.

[8] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, May 2015, DOI: 10.1038/nature14539.

[9] J. E. Goldberg, B. Reig, A. A. Lewin, Y. Gao, L. Heacock, S. L. Heller, and L. Moy, "New horizons: Artificial intelligence for digital breast tomosynthesis," *RadioGraphics*, vol. 43, no. 1, pp. 1–13, Jan. 2023, DOI: 10.1148/rg.220060.

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

[11] M. Gao, J. A. Fessler, and H.-P. Chan, "Digital breast tomosynthesis denoising using deep convolutional neural network: effects of dose level of training target images," in *Proceedings of SPIE*, 2021, 115951K, DOI: 10.1117/12.2580900.

[12] 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 *RSNA Program Book*, 2020, SSPH01-10.

[13] M. Gao, J. A. Fessler, and H.-P. Chan, "Model-based deep CNN-regularized reconstruction for digital breast tomosynthesis with a task-based CNN image assessment approach," *Phys Med Biol*, vol. 68, no. 24, p. 245024, Dec. 2023, DOI: 10.1088/1361-6560/ad0eb4.

[14] M. Gao, J. A. Fessler, and H.-P. Chan, "Deep convolutional neural network regularized digital breast tomosynthesis reconstruction with detector blur and correlated noise modeling," in *Proceedings of SPIE*, 2022, 1203108, DOI: 10.1117/12.2611933.

[15] M. Gao, M. A. Helvie, R. K. Samala, L. M. Hadjiyski, J. A. Fessler, and H.-P. Chan, "Deep CNN task-based image quality assessment: Application to digital breast tomosynthesis reconstruction and denoising," in *Proceedings of SPIE*, 2023, 1246319, DOI: 10.1117/12.2655419.

[16] 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 *RSNA Program Book*, 2021, SSPH09-6.

[17] M. Gao, M. A. Helvie, R. K. Samala, J. A. Fessler, and H.-P. Chan, "Deep learning denoising and assessment of detectability of microcalcifications in digital breast tomosynthesis: A task-based image evaluation approach using CNN," in *RSNA Program Book*, 2022, R5B-SPPH.

[18] I. Sechopoulos, "A review of breast tomosynthesis. Part I. The image acquisition process," *Med Phys*, vol. 40, no. 1, p. 014301, Jan. 2013, DOI: 10.1118/1.4770279.

[19] O. Ortenzia, R. Rossi, M. Bertolini, A. Nitrosi, and C. Ghetti, "Physical characterisation of four different commercial digital breast tomosynthesis systems," *Radiat Prot Dosimetry*, vol. 181, no. 3, pp. 277–289, 2018, DOI: 10.1093/RPD/NCY024.

[20] N. W. Marshall and H. Bosmans, "Performance evaluation of digital breast tomosynthesis systems: Physical methods and experimental data," *Phys Med Biol*, vol. 67, no. 22, p. 22TR03, Nov. 2022, DOI: 10.1088/1361-6560/ac9a35.

[21] V.-M. Sundell, M. Jousi, K. Hukkinen, R. Blanco, T. Mäkelä, and T. Kaasalainen, "A phantom study comparing technical image quality of five breast tomosynthesis systems," *Physica Medica*, vol. 63, pp. 122–130, Jul. 2019, DOI: 10.1016/j.ejmp.2019.06.004.

[22] "GE Senographe Pristina operator manual," *General Electric Company*, 2019. [Online]. Available: https://customer-doc.cloud.gehealthcare.com/.

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

[24] R. Gordon, R. Bender, and G. T. Herman, "Algebraic Reconstruction Techniques (ART) for three-dimensional electron microscopy and X-ray photography," *J Theor Biol*, vol. 29, no. 3, pp. 471–481, Dec. 1970, DOI: 10.1016/0022-5193(70)90109-8.

[25] P. Gilbert, "Iterative methods for the three-dimensional reconstruction of an object from projections," *J Theor Biol*, vol. 36, no. 1, pp. 105–117, Jul. 1972, DOI: 10.1016/0022-5193(72)90180-4.

[26] A. H. Andersen and A. C. KaK, "Simultaneous Algebraic Reconstruction Technique (SART): A superior implementation of the ART algorithm," *Ultrason Imaging*, vol. 6, no. 1, pp. 81–94, Jan. 1984, DOI: 10.1016/0161-7346(84)90008-7.

[27] Y. Zhang, H.-P. Chan, B. Sahiner, J. Wei, M. M. Goodsitt, L. M. Hadjiiski, J. Ge, and C. Zhou, "A comparative study of limited-angle cone-beam reconstruction methods for breast tomosynthesis," *Med Phys*, vol. 33, no. 10, pp. 3781–3795, Oct. 2006, DOI: 10.1118/1.2237543.

[28] M. Jiang and G. Wang, “Convergence of the Simultaneous Algebraic Reconstruction Technique (SART),” *IEEE Transactions on Image Processing*, vol. 12, no. 8, pp. 957–961, 2003.

[29] K. Hornik, M. Stinchcombe, and H. White, “Multilayer feedforward networks are universal approximators,” *Neural Networks*, vol. 2, no. 5, pp. 359–366, Jan. 1989, DOI: 10.1016/0893-6080(89)90020-8.

[30] D. Kingma and J. Ba, “Adam: A method for stochastic optimization,” 2015. [Online]. Available: http://arxiv.org/abs/1412.6980.

[31] M. D. Zeiler and R. Fergus, “Visualizing and understanding convolutional networks,” in *European Conference on Computer Vision*, 2014, 818–833, DOI: 10.1007/978-3-319-10590-1_53.

[32] J. Ludwig, T. Mertelmeier, H. Kunze, and W. Härer, “A novel approach for filtered backprojection in tomosynthesis based on filter kernels determined by iterative reconstruction techniques,” in *International Workshop on Digital Mammography*, 2008, 612–620, DOI: 10.1007/978-3-540-70538-3_85.

[33] J. Liu, A. Zarshenas, Z. Wei, L. Yang, L. Fajardo, K. Suzuki, and A. Qadir, “Radiation dose reduction in digital breast tomosynthesis (DBT) by means of deep-learning-based supervised image processing,” in *Proceedings of SPIE*, 2018, 105740F, DOI: 10.1117/12.2293125.

[34] M. Das, H. C. Gifford, J. M. O’Connor, and S. J. Glick, “Penalized maximum likelihood reconstruction for improved microcalcification detection in breast tomosynthesis,” *IEEE Trans Med Imaging*, vol. 30, no. 4, pp. 904–914, Apr. 2011, DOI: 10.1109/TMI.2010.2089694.

[35] S. Xu, J. Lu, O. Zhou, and Y. Chen, "Statistical iterative reconstruction to improve image quality for digital breast tomosynthesis," *Med Phys*, vol. 42, no. 9, pp. 5377–5390, Sep. 2015, DOI: 10.1118/1.4928603.

[36] J. Zheng, J. A. Fessler, and H.-P. Chan, "Detector blur and correlated noise modeling for digital breast tomosynthesis reconstruction," *IEEE Trans Med Imaging*, vol. 37, no. 1, pp. 116–127, Jan. 2018, DOI: 10.1109/TMI.2017.2732824.

[37] K. Kim, T. Lee, Y. Seong, J. Lee, K. E. Jang, J. Choi, Y. W. Choi, H. H. Kim, H. J. Shin, J. H. Cha, S. Cho, and J. C. Ye, "Fully iterative scatter corrected digital breast tomosynthesis using GPU-based fast Monte Carlo simulation and composition ratio update," *Med Phys*, vol. 42, no. 9, pp. 5342–5355, Aug. 2015, DOI: 10.1118/1.4928139.

[38] Y. Lu, H.-P. Chan, J. Wei, and L. M. Hadjiiski, "Selective-diffusion regularization for enhancement of microcalcifications in digital breast tomosynthesis reconstruction," *Med Phys*, vol. 37, no. 11, pp. 6003–6014, Nov. 2010, DOI: 10.1118/1.3505851.

[39] E. Y. Sidky, I. S. Reiser, R. Nishikawa, and X. Pan, "Image reconstruction in digital breast tomosynthesis by total variation minimization," in *Proceedings of SPIE*, 2007, 651027, DOI: 10.1117/12.713663.

[40] I. Kastanis, S. Arridge, A. Stewart, S. Gunn, C. Ullberg, and T. Francke, "3D digital breast tomosynthesis using total variation regularization," in *International Workshop on Digital Mammography*, 2008, 621–627, DOI: 10.1007/978-3-540-70538-3_86.

[41] E. Y. Sidky, X. Pan, I. S. Reiser, R. M. Nishikawa, R. H. Moore, and D. B. Kopans, "Enhanced imaging of microcalcifications in digital breast tomosynthesis through improved image-reconstruction algorithms," *Med Phys*, vol. 36, no. 11, pp. 4920–4932, Nov. 2009, DOI: 10.1118/1.3232211.

[42] J. Zheng, J. A. Fessler, and H.-P. Chan, "Digital breast tomosynthesis reconstruction using spatially weighted non-convex regularization," in *Proceedings of SPIE*, 2016, 978369, DOI: 10.1117/12.2216414.

[43] M. Sghaier, E. Chouzenoux, J. Pesquet, and S. Muller, "A new spatially adaptive TV regularization for digital breast tomosynthesis," in *IEEE International Symposium on Biomedical Imaging*, 2020, 629–633, DOI: 10.1109/ISBI45749.2020.9098497.

[44] F. E. Boas and D. Fleischmann, "CT artifacts: causes and reduction techniques," *Imaging Med*, vol. 4, no. 2, pp. 229–240, Apr. 2012, DOI: 10.2217/iim.12.13.

[45] S. Kligerman, K. Lahiji, E. Weihe, C. T. Lin, S. Terpenning, J. Jeudy, A. Frazier, R. Pugatch, J. R. Galvin, D. Mittal, K. Kothari, and C. S. White, "Detection of pulmonary embolism on computed tomography: improvement using a model-based iterative reconstruction algorithm compared with filtered back projection and iterative reconstruction algorithms," *J Thorac Imaging*, vol. 30, no. 1, pp. 60–68, Jan. 2015, DOI: 10.1097/RTI.0000000000000122.

[46] M. Das, C. Connolly, S. J. Glick, and H. C. Gifford, "Effect of postreconstruction filter strength on microcalcification detection at different imaging doses in digital breast tomosynthesis: human and model observer studies," in *Proceedings of SPIE*, 2012, 831321, DOI: 10.1117/12.912807.

[47] S. Abdurahman, F. Dennerlein, A. Jerebko, A. Fieselmann, and T. Mertelmeier, "Optimizing high resolution reconstruction in digital breast tomosynthesis using filtered back projection," in *International Workshop on Digital Mammography*, 2014, 520–527, DOI: 10.1007/978-3-319-07887-8_73.

[48] Y. Lu, H.-P. Chan, J. Wei, L. M. Hadjiiski, and R. K. Samala, "Multiscale bilateral filtering for improving image quality in digital breast tomosynthesis," *Med Phys*, vol. 42, no. 1, pp. 182–195, Jan. 2015, DOI: 10.1118/1.4903283.

[49] K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang, "Beyond a Gaussian denoiser: residual learning of deep CNN for image denoising," *IEEE Transactions on Image Processing*, vol. 26, no. 7, pp. 3142–3155, Jul. 2017, DOI: 10.1109/TIP.2017.2662206.

[50] C. Dong, C. C. Loy, K. He, and X. Tang, "Image super-resolution using deep convolutional networks," *IEEE Trans Pattern Anal Mach Intell*, vol. 38, no. 2, pp. 295–307, Feb. 2016, DOI: 10.1109/TPAMI.2015.2439281.

[51] J. Kim, J. K. Lee, and K. M. Lee, "Accurate image super-resolution using very deep convolutional networks," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2016, 1646–1654, DOI: 10.1109/CVPR.2016.182.

[52] Y. Blau and T. Michaeli, "The perception-distortion tradeoff," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2018, 6228–6237, DOI: 10.1109/CVPR.2018.00652.

[53] J. Johnson, A. Alahi, and L. Fei-Fei, "Perceptual losses for real-time style transfer and super-resolution," in *European Conference on Computer Vision*, 2016, 694–711, DOI: 10.1007/978-3-319-46475-6_43.

[54] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial networks," in *Advances in Neural Information Processing Systems*, 2014, 2672–2680.

[55] C. Ledig, L. Theis, F. Huszar, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, and W. Shi, "Photo-realistic single image super-resolution using

a generative adversarial network," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2017, 105–114, DOI: 10.1109/CVPR.2017.19.

[56] M. Arjovsky, S. Chintala, and L. Bottou, "Wasserstein GAN," in *International Conference on Machine Learning*, 2017, 214–223.

[57] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, and A. Courville, "Improved training of Wasserstein GANs," in *Proceedings of International Conference on Neural Information Processing Systems*, 2017, 5769–5779.

[58] J. M. Wolterink, T. Leiner, M. A. Viergever, and I. Isgum, "Generative adversarial networks for noise reduction in low-dose CT," *IEEE Trans Med Imaging*, vol. 36, no. 12, pp. 2536–2545, Dec. 2017, DOI: 10.1109/TMI.2017.2708987.

[59] Q. Yang, P. Yan, Y. Zhang, H. Yu, Y. Shi, X. Mou, M. K. Kalra, Y. Zhang, L. Sun, and G. Wang, "Low-dose CT image denoising using a generative adversarial network with Wasserstein distance and perceptual loss," *IEEE Trans Med Imaging*, vol. 37, no. 6, pp. 1348–1357, Jun. 2018, DOI: 10.1109/TMI.2018.2827462.

[60] Z. Hu, C. Jiang, F. Sun, Q. Zhang, Y. Ge, Y. Yang, X. Liu, H. Zheng, and D. Liang, "Artifact correction in low-dose dental CT imaging using Wasserstein generative adversarial networks," *Med Phys*, vol. 46, no. 4, pp. 1686–1696, Apr. 2019, DOI: 10.1002/mp.13415.

[61] H. Shan, Y. Zhang, Q. Yang, U. Kruger, M. K. Kalra, L. Sun, W. Cong, and G. Wang, "3-D convolutional encoder-decoder network for low-dose CT via transfer learning from a 2-D trained network," *IEEE Trans Med Imaging*, vol. 37, no. 6, pp. 1522–1534, Jun. 2018, DOI: 10.1109/TMI.2018.2832217.

[62] H. Shan, A. Padole, F. Homayounieh, U. Kruger, R. D. Khera, C. Nitiwarangkul, M. K. Kalra, and G. Wang, "Competitive performance of a modularized deep neural network

compared to commercial algorithms for low-dose CT image reconstruction," *Nat Mach Intell*, vol. 1, no. 6, pp. 269–276, 2019, DOI: 10.1038/s42256-019-0057-9.

[63] G. Yang, S. Yu, H. Dong, G. Slabaugh, P. L. Dragotti, X. Ye, F. Liu, S. Arridge, J. Keegan, Y. Guo, and D. Firmin, "DAGAN: Deep de-aliasing generative adversarial networks for fast compressed sensing MRI reconstruction," *IEEE Trans Med Imaging*, vol. 37, no. 6, pp. 1310–1321, Jun. 2018, DOI: 10.1109/TMI.2017.2785879.

[64] M. Mardani, E. Gong, J. Y. Cheng, S. S. Vasanawala, G. Zaharchuk, L. Xing, and J. M. Pauly, "Deep generative adversarial neural networks for compressive sensing MRI," *IEEE Trans Med Imaging*, vol. 38, no. 1, pp. 167–179, Jan. 2019, DOI: 10.1109/TMI.2018.2858752.

[65] 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.

[66] 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 Netw Open*, vol. 1, no. 7, p. e185474, Nov. 2018, DOI: 10.1001/jamanetworkopen.2018.5474.

[67] C. G. Graff, "A new, open-source, multi-modality digital breast phantom," in *Proceedings of SPIE*, 2016, 978309, DOI: 10.1117/12.2216312.

[68] 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.

[69] B. De Man, J. Pack, P. FitzGerald, and M. Wu, “CatSim manual version 6.0,” *GE Global Research*. 2015.

[70] 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, Dec. 2014, DOI: 10.1148/radiol.14132722.

[71] M. M. Goodsitt, H.-P. Chan, A. Schmitz, S. Zelakiewicz, S. Telang, L. Hadjiiski, K. Watcharotone, M. A. Helvie, C. Paramagul, C. Neal, E. Christodoulou, S. C. Larson, and P. L. Carson, “Digital breast tomosynthesis: studies of the effects of acquisition geometry on contrast-to-noise ratio and observer preference of low-contrast objects in breast phantom images,” *Phys Med Biol*, vol. 59, no. 19, pp. 5883–5902, Oct. 2014, DOI: 10.1088/0031-9155/59/19/5883.

[72] S. Ioffe and C. Szegedy, “Batch normalization: Accelerating deep network training by reducing internal covariate shift,” in *International Conference on Machine Learning*, 2015, 448–456.

[73] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, “Image-to-image translation with conditional adversarial networks,” in *IEEE Conference on Computer Vision and Pattern Recognition*, 2017, 5967–5976, DOI: 10.1109/CVPR.2017.632.

[74] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” in *International Conference on Learning Representations*, 2015.

[75] J. Yosinski, J. Clune, Y. Bengio, and H. Lipson, "How transferable are features in deep neural networks?," in *Proceedings of International Conference on Neural Information Processing Systems*, 2014, 3320–3328.

[76] J. M. Boone, T. R. Fewell, and R. J. Jennings, "Molybdenum, rhodium, and tungsten anode spectral models using interpolating polynomials with application to mammography," *Med Phys*, vol. 24, no. 12, pp. 1863–1874, Dec. 1997, DOI: 10.1118/1.598100.

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

[78] E. Samei, M. J. Flynn, and D. A. Reimann, "A method for measuring the presampled MTF of digital radiographic systems using an edge test device," *Med Phys*, vol. 25, no. 1, pp. 102–113, Jan. 1998, DOI: 10.1118/1.598165.

[79] "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.

[80] R. K. Samala, H.-P. Chan, L. Hadjiiski, M. A. Helvie, C. D. Richter, and K. H. Cha, "Breast cancer diagnosis in digital breast tomosynthesis: effects of training sample size on multi-stage transfer learning using deep neural nets," *IEEE Trans Med Imaging*, vol. 38, no. 3, pp. 686–696, Mar. 2019, DOI: 10.1109/TMI.2018.2870343.

[81] A. E. Burgess, "Mammographic structure: data preparation and spatial statistics analysis," in *Proceedings of SPIE*, 1999, 3661, DOI: 10.1117/12.348620.

[82] J. H. Siewerdsen, I. A. Cunningham, and D. A. Jaffray, "A framework for noise-power spectrum analysis of multidimensional images," *Med Phys*, vol. 29, no. 11, pp. 2655–2671, Nov. 2002, DOI: 10.1118/1.1513158.

[83] 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," *Med Phys*, vol. 39, no. 1, pp. 28–39, Jan. 2012, DOI: 10.1118/1.3662072.

[84] A. E. Burgess, "Statistically defined backgrounds: performance of a modified nonprewhitening observer model," *Journal of the Optical Society of America A*, vol. 11, no. 4, p. 1237, Apr. 1994, DOI: 10.1364/JOSAA.11.001237.

[85] L.-N. D. Loo, K. Doi, and C. E. Metz, "A comparison of physical image quality indices and observer performance in the radiographic detection of nylon beads," *Phys Med Biol*, vol. 29, no. 7, pp. 837–856, Jul. 1984, DOI: 10.1088/0031-9155/29/7/007.

[86] A. E. Burgess, X. Li, and C. K. Abbey, "Visual signal detectability with two noise components: anomalous masking effects," *Journal of the Optical Society of America A*, vol. 14, no. 9, p. 2420, Sep. 1997, DOI: 10.1364/JOSAA.14.002420.

[87] S. Richard and J. H. Siewerdsen, "Comparison of model and human observer performance for detection and discrimination tasks using dual-energy x-ray images," *Med Phys*, vol. 35, no. 11, pp. 5043–5053, Oct. 2008, DOI: 10.1118/1.2988161.

[88] G. J. Gang, J. Lee, J. W. Stayman, D. J. Tward, W. Zbijewski, J. L. Prince, and J. H. Siewerdsen, "Analysis of Fourier-domain task-based detectability index in tomosynthesis and cone-beam CT in relation to human observer performance," *Med Phys*, vol. 38, no. 4, pp. 1754–1768, Mar. 2011, DOI: 10.1118/1.3560428.

[89] O. Christianson, J. J. S. Chen, Z. Yang, G. Saiprasad, A. Dima, J. J. Filliben, A. Peskin, C. Trimble, E. L. Siegel, and E. Samei, "An improved index of image quality for task-based performance of CT iterative reconstruction across three commercial implementations," *Radiology*, vol. 275, no. 3, pp. 725–734, Jun. 2015, DOI: 10.1148/radiol.15132091.

[90] D. H. Kelly, "Spatial frequency selectivity in the retina," *Vision Res*, vol. 15, no. 6, pp. 665–672, Jun. 1975, DOI: 10.1016/0042-6989(75)90282-5.

[91] 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," *Phys Med Biol*, vol. 60, no. 21, pp. 8457–8479, Nov. 2015, DOI: 10.1088/0031-9155/60/21/8457.

[92] H.-P. Chan, M. A. Helvie, M. Gao, L. Hadjiiski, C. Zhou, K. Garver, K. A. Klein, C. McLaughlin, R. Oudsema, W. T. Rahman, and M. A. Roubidoux, "Deep learning denoising of digital breast tomosynthesis: Observer performance study of the effect on detection of microcalcifications in breast phantom images," *Med Phys*, vol. 50, no. 10, pp. 6177–6189, Oct. 2023, DOI: 10.1002/mp.16439.

[93] K. Lange and J. A. Fessler, "Globally convergent algorithms for maximum a posteriori transmission tomography," *IEEE Transactions on Image Processing*, vol. 4, no. 10, pp. 1430–1438, Oct. 1995, DOI: 10.1109/83.465107.

[94] J. Nuyts, B. De Man, J. A. Fessler, W. Zbijewski, and F. J. Beekman, "Modelling the physics in the iterative reconstruction for transmission computed tomography," *Phys Med Biol*, vol. 58, no. 12, p. R63, Jun. 2013, DOI: 10.1088/0031-9155/58/12/R63.

[95] E. Haneda, J. E. Tkaczyk, G. Palma, R. Iordache, S. Zelakiewicz, S. Muller, and B. De Man, "Toward a dose reduction strategy using model-based reconstruction with limited-angle tomosynthesis," in *Proceedings of SPIE*, 2014, 90330V, DOI: 10.1117/12.2042897.

[96] A. Badal, K. H. Cha, S. Divel, C. G. Graff, R. Zeng, and A. Badano, "Virtual clinical trial for task-based evaluation of a deep learning synthetic mammography algorithm," in *Proceedings of SPIE*, 2019, 109480O, DOI: 10.1117/12.2513062.

[97] P. Sahu, H. Huang, W. Zhao, and H. Qin, "Using virtual digital breast tomosynthesis for de-noising of low-dose projection images," in *International Symposium on Biomedical Imaging*, 2019, 1647–1651, DOI: 10.1109/ISBI.2019.8759408.

[98] T. Gomi, Y. Kijima, T. Kobayashi, and Y. Koibuchi, "Evaluation of a generative adversarial network to improve image quality and reduce radiation-dose during digital breast tomosynthesis," *Diagnostics*, vol. 12, no. 2, p. 495, Feb. 2022, DOI: 10.3390/diagnostics12020495.

[99] G. Wang, J. C. Ye, and B. De Man, "Deep learning for tomographic image reconstruction," *Nat Mach Intell*, vol. 2, no. 12, pp. 737–748, Dec. 2020, DOI: 10.1038/s42256-020-00273-z.

[100] G. Wang, M. Jacob, X. Mou, Y. Shi, and Y. C. Eldar, "Deep tomographic image reconstruction: Yesterday, today, and tomorrow—Editorial for the 2nd special issue 'Machine learning for image reconstruction,'" *IEEE Trans Med Imaging*, vol. 40, no. 11, pp. 2956–2964, Nov. 2021, DOI: 10.1109/TMI.2021.3115547.

[101] V. Monga, Y. Li, and Y. C. Eldar, "Algorithm unrolling: Interpretable, efficient deep learning for signal and image processing," *IEEE Signal Process Mag*, vol. 38, no. 2, pp. 18–44, Mar. 2021, DOI: 10.1109/MSP.2020.3016905.

[102] Y. Yang, J. Sun, H. Li, and Z. Xu, "Deep ADMM-Net for compressive sensing MRI," in *Advances in Neural Information Processing Systems*, 2016, 10–18.

[103] H. K. Aggarwal, M. P. Mani, and M. Jacob, "MoDL: Model-based deep learning architecture for inverse problems," *IEEE Trans Med Imaging*, vol. 38, no. 2, pp. 394–405, Feb. 2019, DOI: 10.1109/TMI.2018.2865356.

[104] J. Teuwen, N. Moriakov, C. Fedon, M. Caballo, I. Reiser, P. Bakic, E. García, O. Diaz, K. Michielsen, and I. Sechopoulos, "Deep learning reconstruction of digital breast tomosynthesis images for accurate breast density and patient-specific radiation dose estimation," *Med Image Anal*, vol. 71, p. 102061, Jul. 2021, DOI: 10.1016/j.media.2021.102061.

[105] D. Wu, K. Kim, and Q. Li, "Digital breast tomosynthesis reconstruction with deep neural network for improved contrast and in-depth resolution," in *International Symposium on Biomedical Imaging*, 2020, 656–659, DOI: 10.1109/ISBI45749.2020.9098661.

[106] T. Su, X. Deng, J. Yang, Z. Wang, S. Fang, H. Zheng, D. Liang, and Y. Ge, "DIR-DBTnet: Deep iterative reconstruction network for three-dimensional digital breast tomosynthesis imaging," *Med Phys*, vol. 48, no. 5, pp. 2289–2300, May 2021, DOI: 10.1002/mp.14779.

[107] 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.

[108] Y. Romano, M. Elad, and P. Milanfar, "The little engine that could: regularization by denoising (RED)," *SIAM J Imaging Sci*, vol. 10, no. 4, pp. 1804–1844, Jan. 2017, DOI: 10.1137/16M1102884.

[109] R. Ahmad, C. A. Bouman, G. T. Buzzard, S. Chan, S. Liu, E. T. Reehorst, and P. Schniter, "Plug-and-play methods for magnetic resonance imaging: Using denoisers for image recovery," *IEEE Signal Process Mag*, vol. 37, no. 1, pp. 105–116, Jan. 2020, DOI: 10.1109/MSP.2019.2949470.

[110] D. H. Ye, S. Srivastava, J.-B. Thibault, K. Sauer, and C. Bouman, "Deep residual learning for model-based iterative CT reconstruction using plug-and-play framework," in *IEEE International Conference on Acoustics, Speech and Signal Processing*, 2018, 6668–6672, DOI: 10.1109/ICASSP.2018.8461408.

[111] J. He, Y. Yang, Y. Wang, D. Zeng, Z. Bian, H. Zhang, J. Sun, Z. Xu, and J. Ma, "Optimizing a parameterized plug-and-play ADMM for iterative low-dose CT reconstruction," *IEEE Trans Med Imaging*, vol. 38, no. 2, pp. 371–382, 2019, DOI: 10.1109/TMI.2018.2865202.

[112] Z. Xie, R. Baikejiang, T. Li, X. Zhang, K. Gong, M. Zhang, W. Qi, E. Asma, and J. Qi, "Generative adversarial network based regularized image reconstruction for PET," *Phys Med Biol*, vol. 65, no. 12, p. 125016, Jun. 2020, DOI: 10.1088/1361-6560/ab8f72.

[113] ICRU, "Medical imaging: The assessment of image quality (Report 54)," Apr. 1996, DOI: 10.1093/jicru/os28.1.Report54.

[114] H. H. Barrett, J. Yao, J. P. Rolland, and K. J. Myers, "Model observers for assessment of image quality," *Proceedings of the National Academy of Sciences*, vol. 90, no. 21, pp. 9758–65, 1993.

[115] J. B. Solomon and E. Samei, "Correlation between human detection accuracy and observer model-based image quality metrics in computed tomography," *Journal of Medical Imaging*, vol. 3, no. 03, p. 035506, Sep. 2016, DOI: 10.1117/1.JMI.3.3.035506.

[116] D. Petrov, N. W. Marshall, K. C. Young, and H. Bosmans, "Systematic approach to a channelized Hotelling model observer implementation for a physical phantom containing mass-like lesions: Application to digital breast tomosynthesis," *Physica Medica*, vol. 58, pp. 8–20, 2019, DOI: 10.1016/j.ejmp.2018.12.033.

[117] Y. H. Hu and W. Zhao, "The effect of angular dose distribution on the detection of microcalcifications in digital breast tomosynthesis," *Med Phys*, vol. 38, no. 5, pp. 2455–2466, 2011, DOI: 10.1118/1.3570580.

[118] C. Balta, R. W. Bouwman, I. Sechopoulos, M. J. M. Broeders, N. Karssemeijer, R. E. van Engen, and W. J. H. Veldkamp, "Can a channelized Hotelling observer assess image quality in acquired mammographic images of an anthropomorphic breast phantom including image processing?," *Med Phys*, vol. 46, no. 2, pp. 714–725, Feb. 2019, DOI: 10.1002/mp.13342.

[119] K. Michielsen, J. Nuyts, L. Cockmartin, N. Marshall, and H. Bosmans, "Design of a model observer to evaluate calcification detectability in breast tomosynthesis and application to smoothing prior optimization," *Med Phys*, vol. 43, no. 12, pp. 6577–6587, 2016, DOI: 10.1118/1.4967268.

[120] R. Zeng, F. W. Samuelson, D. Sharma, A. Badal, G. G. Christian, S. J. Glick, K. J. Myers, and A. Badano, "Computational reader design and statistical performance evaluation of an in-silico imaging clinical trial comparing digital breast tomosynthesis with full-field digital mammography," *Journal of Medical Imaging*, vol. 7, no. 04, p. 042802, Feb. 2020, DOI: 10.1117/1.JMI.7.4.042802.

[121] X. Zhang, V. A. Kelkar, J. Granstedt, H. Li, and M. A. Anastasio, "Impact of deep learning-based image super-resolution on binary signal detection," *Journal of Medical Imaging*, vol. 8, no. 06, p. 065501, Nov. 2021, DOI: 10.1117/1.JMI.8.6.065501.

[122] E. T. Reehorst and P. Schniter, "Regularization by denoising: Clarifications and new interpretations," *IEEE Trans Comput Imaging*, vol. 5, no. 1, pp. 52–67, Mar. 2019, DOI: 10.1109/TCI.2018.2880326.

[123] M. A. Kupinski, D. C. Edwards, M. L. Giger, and C. E. Metz, "Ideal observer approximation using Bayesian classification neural networks," *IEEE Trans Med Imaging*, vol. 20, no. 9, pp. 886–899, 2001, DOI: 10.1109/42.952727.

[124] 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.

[125] Z. Liu, H. Mao, C.-Y. Wu, C. Feichtenhofer, T. Darrell, and S. Xie, "A ConvNet for the 2020s," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2022, 11966–11976, DOI: 10.1109/CVPR52688.2022.01167.

[126] A. Badal, D. Sharma, C. G. Graff, R. Zeng, and A. Badano, "Mammography and breast tomosynthesis simulator for virtual clinical trials," *Comput Phys Commun*, vol. 261, p. 107779, Apr. 2021, DOI: 10.1016/j.cpc.2020.107779.

[127] 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," *Phys Med Biol*, vol. 40, no. 5, pp. 857–876, May 1995, DOI: 10.1088/0031-9155/40/5/010.

[128] Y. Lu, H.-P. Chan, J. Wei, L. M. Hadjiiski, and R. K. Samala, "Improving image quality for digital breast tomosynthesis: an automated detection and diffusion-based method for metal artifact reduction," *Phys Med Biol*, vol. 62, no. 19, pp. 7765–7783, Sep. 2017, DOI: 10.1088/1361-6560/aa8803.

[129] Y. Lu, H.-P. Chan, J. Wei, and L. M. Hadjiiski, "A diffusion-based truncated projection artifact reduction method for iterative digital breast tomosynthesis reconstruction," *Phys Med Biol*, vol. 58, no. 3, pp. 569–587, Feb. 2013, DOI: 10.1088/0031-9155/58/3/569.

[130] N. R. Huber, J. Kim, S. Leng, C. H. Mccollough, and L. Yu, "Deep learning-based image noise quantification framework for computed tomography," *J Comput Assist Tomogr*, vol. 47, no. 4, pp. 603–607, 2023, DOI: 10.1097/RCT.0000000000001469.

[131] G. T. Buzzard, S. H. Chan, S. Sreehari, and C. A. Bouman, "Plug-and-play unplugged: Optimization-free reconstruction using consensus equilibrium," *SIAM J Imaging Sci*, vol. 11, no. 3, pp. 2001–2020, Jan. 2018, DOI: 10.1137/17M1122451.

[132] E. K. Ryu, J. Liu, S. Wang, X. Chen, Z. Wang, and W. Yin, "Plug-and-play methods provably converge with properly trained denoisers," in *International Conference on Machine Learning*, 2019.

[133] J. Zheng, J. A. Fessler, and H.-P. Chan, "Effect of source blur on digital breast tomosynthesis reconstruction," *Med Phys*, vol. 46, no. 12, pp. 5572–5592, Dec. 2019, DOI: 10.1002/mp.13801.

[134] E. Shaheen, N. Marshall, and H. Bosmans, "Investigation of the effect of tube motion in breast tomosynthesis: continuous or step and shoot?," in *Proceedings of SPIE*, 2011, 79611E, DOI: 10.1117/12.877348.

[135] C. Lee and J. Baek, "Effect of optical blurring of X-ray source on breast tomosynthesis image quality: Modulation transfer function, anatomical noise power spectrum, and signal detectability perspectives," *PLoS One*, vol. 17, no. 5, p. e0267850, May 2022, DOI: 10.1371/journal.pone.0267850.

[136] N. W. Marshall and H. Bosmans, “Measurements of system sharpness for two digital breast tomosynthesis systems,” *Phys Med Biol*, vol. 57, no. 22, pp. 7629–7650, Nov. 2012, DOI: 10.1088/0031-9155/57/22/7629.

[137] F. Michelfeit, “Siemens Healthineers presents mammography system with groundbreaking new imaging technology,” *Siemens Healthineers*, 2023. [Online]. Available: https://www.siemens-healthineers.com/press/releases/bbrilliant. [Accessed: 18-Dec-2023].

[138] S. Tilley, J. H. Siewerdsen, and J. W. Stayman, “Model-based iterative reconstruction for flat-panel cone-beam CT with focal spot blur, detector blur, and correlated noise,” *Phys Med Biol*, vol. 61, no. 1, pp. 296–319, Jan. 2016, DOI: 10.1088/0031-9155/61/1/296.

[139] S. Tilley, M. Jacobson, Q. Cao, M. Brehler, A. Sisniega, W. Zbijewski, and J. W. Stayman, “Penalized-likelihood reconstruction with high-fidelity measurement models for high-resolution cone-beam imaging,” *IEEE Trans Med Imaging*, vol. 37, no. 4, pp. 988–999, Apr. 2018, DOI: 10.1109/TMI.2017.2779406.

[140] L. Fu, J. Wang, X. Rui, J. B. Thibault, and B. De Man, “Modeling and estimation of detector response and focal spot profile for high-resolution iterative CT reconstruction,” in *IEEE Nuclear Science Symposium and Medical Imaging Conference*, 2013, 1–5, DOI: 10.1109/NSSMIC.2013.6829044.

[141] S. Majee, S. Aslan, D. Gursoy, and C. A. Bouman, “CodEx: A modular framework for joint temporal de-blurring and tomographic reconstruction,” *IEEE Trans Comput Imaging*, vol. 8, pp. 666–678, 2022, DOI: 10.1109/TCI.2022.3197935.

[142] K. Michielsen, K. Van Slambrouck, A. Jerebko, and J. Nuyts, “Patchwork reconstruction with resolution modeling for digital breast tomosynthesis,” *Med Phys*, vol. 40, p. 031105, Feb. 2013, DOI: 10.1118/1.4789591.

[143] N. Wiener, *Extrapolation, Interpolation, and Smoothing of Stationary Time Series*. The MIT Press, 1949.

[144] W. H. Richardson, "Bayesian-based iterative method of image restoration," *J Opt Soc Am*, vol. 62, no. 1, p. 55, Jan. 1972, DOI: 10.1364/JOSA.62.000055.

[145] D. Krishnan and R. Fergus, "Fast image deconvolution using hyper-Laplacian priors," in *Advances in Neural Information Processing Systems*, 2009, 1033–1041.

[146] L. Xu, S. Zheng, and J. Jia, "Unnatural L0 sparse representation for natural image deblurring," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2013, 1107–1114, DOI: 10.1109/CVPR.2013.147.

[147] D. Zoran and Y. Weiss, "From learning models of natural image patches to whole image restoration," in *International Conference on Computer Vision*, 2011, 479–486, DOI: 10.1109/ICCV.2011.6126278.

[148] K. Zhang, W. Ren, W. Luo, W.-S. Lai, B. Stenger, M.-H. Yang, and H. Li, "Deep image deblurring: A survey," *Int J Comput Vis*, vol. 130, no. 9, pp. 2103–2130, Sep. 2022, DOI: 10.1007/s11263-022-01633-5.

[149] J. Dong, S. Roth, and B. Schiele, "DWDN: Deep Wiener deconvolution network for non-blind image deblurring," *IEEE Trans Pattern Anal Mach Intell*, vol. 44, no. 12, pp. 9960–9976, Dec. 2022, DOI: 10.1109/TPAMI.2021.3138787.

[150] J. Ho, A. Jain, and P. Abbeel, "Denoising diffusion probabilistic models," in *Advances in Neural Information Processing Systems*, 2020, 6840–6851.

[151] Y. Song and S. Ermon, "Generative modeling by estimating gradients of the data distribution," *Adv Neural Inf Process Syst*, vol. 32, no. NeurIPS, 2019.

[152] Y. Song, J. Sohl-Dickstein, D. P. Kingma, A. Kumar, S. Ermon, and B. Poole, "Score-based generative modeling through stochastic differential equations," in *International Conference on Learning Representations*, 2021.

[153] B. Kawar, M. Elad, S. Ermon, and J. Song, "Denoising diffusion restoration models," in *Advances in Neural Information Processing Systems*, 2022, 23593–23606.

[154] C. Saharia, J. Ho, W. Chan, T. Salimans, D. J. Fleet, and M. Norouzi, "Image super-resolution via iterative refinement," *IEEE Trans Pattern Anal Mach Intell*, vol. 45, no. 4, pp. 4713–4726, 2023, DOI: 10.1109/TPAMI.2022.3204461.

[155] Y. Wang, J. Yu, and J. Zhang, "Zero-shot image restoration using denoising diffusion null-space model," in *International Conference on Learning Representations*, 2023.

[156] Y. Song, L. Shen, L. Xing, and S. Ermon, "Solving inverse problems in medical imaging with score-based generative models," in *International Conference on Learning Representations*, 2022.

[157] H. Chung, D. Ryu, M. T. McCann, M. L. Klasky, and J. C. Ye, "Solving 3D inverse problems using pre-trained 2D diffusion models," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2023, 22542–22551.

[158] J. T. Bushberg, J. A. Seibert, E. M. Leidholdt, and J. M. Boone, "Diagnostic Radiology," in *The Essential Physics of Medical Imaging*, 3rd ed., Lippincott Williams & Wilkins, 2011, pp. 169–576.

[159] A. Mackenzie, N. W. Marshall, A. Hadjipanteli, D. R. Dance, H. Bosmans, and K. C. Young, "Characterisation of noise and sharpness of images from four digital breast tomosynthesis systems for simulation of images for virtual clinical trials," *Phys Med Biol*, vol. 62, no. 6, pp. 2376–2397, Mar. 2017, DOI: 10.1088/1361-6560/aa5dd9.

[160] J. Sohl-Dickstein, E. A. Weiss, N. Maheswaranathan, and S. Ganguli, "Deep unsupervised learning using nonequilibrium thermodynamics," in *International Conference on Machine Learning*, 2015, 2246–2255.

[161] J. Song, C. Meng, and S. Ermon, "Denoising diffusion implicit Models," in *International Conference on Learning Representations*, 2021.

[162] P. Dhariwal and A. Nichol, "Diffusion Models Beat GANs on Image Synthesis," *Adv Neural Inf Process Syst*, vol. 11, pp. 8780–8794, May 2021.

[163] P. C. Johns and M. J. Yaffe, "X-ray characterisation of normal and neoplastic breast tissues," *Phys Med Biol*, vol. 32, no. 6, pp. 675–695, Jun. 1987, DOI: 10.1088/0031-9155/32/6/002.

[164] "Technical evaluation of Siemens Mammomat Inspiration digital breast tomosynthesis system," *Public Health England*, 2018. [Online]. Available: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/997619/Siemens_InspirationVB60_PostCPG.pdf.

[165] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention*, 2015, 234–241, DOI: 10.1007/978-3-319-24574-4_28.

[166] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need," in *Advances in Neural Information Processing Systems*, 2017.

[167] K. Zhang, L. Van Gool, and R. Timofte, "Deep unfolding network for image super-resolution," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2020, 3217–3226.

[168] R. K. Samala, H.-P. Chan, Y. Lu, L. Hadjiiski, J. Wei, B. Sahiner, and M. A. Helvie, "Computer-aided detection of clustered microcalcifications in multiscale bilateral filtering regularized reconstructed digital breast tomosynthesis volume," *Med Phys*, vol. 41, no. 2, p. 021901, Jan. 2014, DOI: 10.1118/1.4860955.