

Article Type: Research Article

Automated Pectoral Muscle Identification on MLO-view Mammograms:
Comparison of Deep Neural Network to Conventional Computer Vision

Xiangyuan Ma, B.S^{1, 2, 3}

Jun Wei, PhD^{1*}

Chuan Zhou, PhD¹

Mark A. Helvie, MD¹

Heang-Ping Chan, PhD¹

Lubomir M. Hadjiiski, PhD¹

Yao Lu, PhD^{2,3}

¹Department of Radiology, University of Michigan, Ann Arbor

²School of Data and Computer Science, Sun Yat-Sen University, Guangzhou 510275, P.R. China

³Guangdong Province Key Laboratory Computational Science, Sun Yat-Sen University, Guangzhou
510275, P.R. China

Correspondence:

Jun Wei, Ph.D.

This is the author manuscript accepted for publication and has undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the [Version of Record](#). Please cite this article as [doi: 10.1002/MP.13451](https://doi.org/10.1002/MP.13451)

This article is protected by copyright. All rights reserved

Department of Radiology
University of Michigan
25 CGC B2103
1500 E Medical Center Drive
Ann Arbor, MI 48109
Phone: 734-647-8553
Fax: 734-615-5513
30 Email: jvwei@med.umich.edu

Abstract

Objectives: To develop a fully automated deep learning approach for identification of the pectoral muscle on mediolateral oblique (MLO) view mammograms and evaluate its performance in
35 comparison to our previously developed texture-field orientation (TFO) method using conventional image feature analysis. Pectoral muscle segmentation is an important step for automated image analyses such as breast density or parenchymal pattern classification, lesion detection, and multi-view correlation.

Materials and Methods: Institutional Review Board (IRB) approval was obtained before data
40 collection. A dataset of 729 MLO-view mammograms including 637 digitized film mammograms (DFM) and 92 digital mammograms (DM) from our previous study were used for the training and validation of our deep convolutional neural network (DCNN) segmentation method. In addition, we collected an independent set of 203 DMs from 131 patients for testing. The film mammograms were digitized at a pixel size of $50 \mu\text{m} \times 50 \mu\text{m}$ with a Lumiscan digitizer. All DMs were acquired
45 with GE systems at a pixel size of $100 \mu\text{m} \times 100 \mu\text{m}$. An experienced MQSA radiologist manually drew the pectoral muscle boundary on each mammogram as the reference standard. We trained the DCNN to estimate a probability map of the pectoral muscle region on mammograms. The DCNN consisted of a contracting path to capture multi-resolution image context and a symmetric expanding path for prediction of the pectoral muscle region. Three DCNN structures were compared for
50 automated identification of pectoral muscles. Ten-fold cross-validation was used in training of the

DCNNs. After training, we applied the 10 trained models during cross validation to the independent DM test set. The predicted pectoral muscle region of each test DM was obtained as the mean probability map by averaging the ensemble of probability maps from the 10 models. The DCNN-segmented pectoral muscle was evaluated by three performance measures relative to the reference standard: 1) the percent overlap area (POA) of the pectoral muscle regions, 2) the Hausdorff distance (Hdist), and 3) the average Euclidean distance (AvgDist) between the boundaries. The results were compared to those obtained with the TFO method, used as our baseline. A two-tailed paired t-test was performed to examine the significance in the differences between the DCNN and the baseline.

Results: In the 10 test partitions of the cross-validation set, the DCNN achieved a mean POA of $96.5\pm 2.9\%$, a mean Hdist of 2.26 ± 1.31 mm, and a mean AvgDist of 0.78 ± 0.58 mm, while the corresponding measures by the baseline method were $94.2\pm 4.8\%$, 3.69 ± 2.48 mm, and 1.30 ± 1.22 mm, respectively. For the independent DM test set, the DCNN achieved a mean POA of $93.7\pm 6.9\%$, a mean Hdist of 3.80 ± 3.21 mm, and a mean AvgDist of 1.49 ± 1.62 mm comparing to $86.9\pm 16.0\%$, 7.18 ± 14.22 mm, and 3.98 ± 14.13 mm, respectively, by the baseline method.

Conclusion: In comparison to the TFO method, DCNN significantly improved the accuracy of pectoral muscle identification on mammograms ($p < 0.05$).

KEYWORDS:

Mammogram, Mediolateral oblique (MLO) view, Pectoral muscle, Deep convolutional neural network (DCNN)

I. INTRODUCTION

Breast cancer is the most common cancer diagnosed in American women. The American Cancer Society estimates that there will be about 266,120 new breast cancer cases and about 40,920 deaths in the United States for 2018¹. Early detection increases the chance of cancer free survival. At present, screening mammography is the most effective method for early detection of breast cancer². Dense breast parenchyma is a risk factor for breast cancer. The combined relative risks of incident breast cancer in the general population are estimated to be about 4-6 times higher for women whose mammograms have parenchymal densities over 75% of the breast area than women with less than 5% of parenchymal densities^{3,4,5}.

Accurate segmentation of breast region is an essential step for quantitative analysis of breast parenchyma on mammograms. However, due to the high x-ray attenuation of pectoral muscle on mediolateral oblique (MLO) view mammograms, its pixel values overlap with the range of the fibroglandular tissue in the breast region. If the pectoral muscle is not excluded as a part of the segmented fibroglandular tissue, the quantitative analysis of breast parenchyma on MLO-view mammograms will not be consistent with that on craniocaudal (CC) view mammograms. Therefore, accurate identification of the pectoral muscle on MLO-view mammograms is important in quantitative analysis of breast parenchyma on MLO-view mammograms.

Automated pectoral muscle identification on MLO-view mammograms remains a challenging problem, especially for the improperly positioned breasts and those that contain dense glandular tissue overlapping with the pectoral muscle region. Figure 1 shows image examples illustrating three different challenging pectoral muscle patterns.

A number of studies have been conducted to identify the pectoral muscle. Karthikeyan et al.⁶ reviewed methods for pectoral muscle segmentation. Some assumed that the intensity of the pectoral muscle was higher than that of the breast tissue⁷⁻⁹ and were successful to different degrees by intensity-based segmentation techniques. Karssemeijer¹⁰ estimated the edge of the pectoral muscle with a straight line and applied Hough transform to detect the straight line as the pectoral muscle. Ma et al.¹¹ applied graph theory in conjunction with active contour methods for identifying the pectoral muscle. We have previously developed a texture-field orientation (TFO) method¹² to estimate pectoral muscle boundary on MLO-view mammograms. Taghanaki et al.¹³

105 combined geometric rules with a region growing algorithm to support the segmentation of all types
of pectoral muscles.

Convolution neural network (CNN) was introduced for recognition of handwritten numeral patterns
in the early 1980's^{14,15} and applied to medical image pattern recognition in the early 1990's¹⁶⁻²².
The CNN structures had relatively few layers in these early applications due to the limitations in
110 computational speed and training sample sizes. CNN applications became much more versatile in
recent years due to the development of more efficient regularization methods and fast graphical
processing units, allowing CNN structures to grow in both depth and width which increase the
learning capacity tremendously. The deep convolution neural networks (DCNN) had been shown
to be particularly successful in the task of classifying natural images for which millions of training
115 samples can be collected relatively easily^{23,24}. A recent conference proceeding paper²⁵ presented a
preliminary study that trained a deep learning model for pectoral muscle segmentation using only
136 digital breast tomosynthesis (DBT) images. In contrast to conventional methods, DCNN can
automatically extract image features through unsupervised or supervised learning if a large training
set is available. We hypothesize that DCNN can accurately segment pectoral muscle on MLO-
120 view mammograms. To test this hypothesis, we develop a supervised deep learning approach for
automated identification of the pectoral muscle on MLO-view mammograms and evaluate its
performance in comparison to our previous TFO method in this study.

125 **II. MATERIALS AND METHOD**

II.A. Data sets

Institutional Review Board (IRB) approval was obtained prior to the collection of the image data.
Two datasets were used in this study. Dataset I comprised a total of 729 MLO-view mammograms,
including 637 digitized film mammograms (DFMs) and 92 digital mammograms (DMs) that were
used to test the TFO method in our previous study¹². Of the 637 DFMs, 531 from 463 patients
130 were randomly selected from a de-identified data set collected at the Old Order Amish population of
Lancaster County, Pennsylvania, and the remaining 106 from 99 patients were randomly selected
from the patient files at our institution. All film mammograms were digitized with a LUMISYS 85

laser film scanner at a pixel size of $50 \mu\text{m} \times 50 \mu\text{m}$ and 12-bit gray levels. The 92 MLO-view DMs from 92 patients were collected from the patient files at our institution. Dataset II with 203 DMs from 131 patients was independently collected from screening patients at our institution. This dataset was kept independent of the training process and sequestered for testing the trained system after the training was completed and all parameters were frozen to validate its robustness and generalizability to unknown cases²⁶. All of the DMs were acquired with a GE Senographe 2000D system at a pixel size of $100 \mu\text{m} \times 100 \mu\text{m}$ and 16-bit gray levels. The “for presentation” DMs were used in this study. To reduce processing time and noise, the resolution of the DFMs and DMs was reduced to $800 \mu\text{m} \times 800 \mu\text{m}$ by averaging every 16×16 pixels and 8×8 adjacent pixels, respectively, before the pectoral muscle analysis.

An in-house developed graphical user interface was used for interactive tracking of pectoral boundary on mammograms. An experienced Mammography Quality Standards Act (MQSA) radiologist used the interface to draw the pectoral boundary on each MLO-view mammogram. The manually segmented boundary was then used to generate a binary mask separating the pectoral muscle and the breast region. For the training set, the binary mask was used to guide the training of the DCNN. For the test set, the boundary and the binary mask were used as the reference standard to evaluate the segmentation accuracy of the performances of the automated methods.

150

II.B. Methods

Figure 2 shows our deep learning approach for pectoral muscle identification on MLO-view mammograms. We first applied a preprocessing step to an input mammogram, which was designed to normalize the image for reliable feature extraction. For a given image, zero ranks were first removed. The maximum pixel intensity in each image was determined and used to scale the gray level dynamic range to $[0.0, 1.0]$. Then we resized all images and the reference binary masks to a uniform size of 256×256 pixels.

155

II. B.1 Deep Convolutional Neural Network

160 We configured a DCNN that has a structure similar to a U-Net²⁷ to generate the likelihood map of the pectoral muscle on MLO-view mammogram. Keras²⁸ with Tensorflow backend was used to implement the neural network. Figure 3 illustrated the architecture of our DCNN. The DCNN consisted of a contracting path to capture multi-resolution image context and a symmetric expanding path for prediction of the pectoral muscle region. The reference binary mask of the pectoral muscle region for each training mammogram was used to formulate the loss function to guide the training.

165 The contracting path consisted of repeated application of two 5×5 convolution kernels, each followed by a rectified linear unit (ReLU) and a 2×2 max pooling operation with stride 2. Unlike the original U-Net, we applied the same padding to the convolution kernels in order to keep the spatial dimensions of the output feature map the same as those of the input feature map. After each

170 max pooling operation, the number of feature channels was doubled to include both band-pass and low-pass information. In the expanding path, each layer consisted of an upsampling of the feature map followed by a 5×5 deconvolution operation with a ReLU activation function. Then two 5×5 convolution operations each followed by a ReLU were applied and the feature map was concatenated with the corresponding feature map from the contracting path.

175 In order to investigate the stability of this DCNN with different parameters, we conducted two reliability studies: 1) to use different network architectures, and 2) to use different loss functions. For the first study, we varied the number of channels to generate three different DCNN architectures. Table 1 summarized the parameters of these three architectures. According to the total number of parameters, the three different models were denoted as the large parameter DCNN

180 (LP-DCNN) model, the medium parameter DCNN (MP-DCNN) model and the small parameter DCNN (SP-DCNN) model. The LP-DCNN, MP-DCNN, and SP-DCNN had 15.1 million, 3.8 million and 0.9 million parameters, respectively.

185 Table 1. The number of channels at the different layers of the DCNN for the three models studied

In our second reliability study, we compared three different loss functions including the binary cross-entropy loss function (BCEL), the Dice loss function (DL) and the percent overlap area loss

function (POAL) during back-propagation of DCNN training. We treated these loss functions as the optimizer in a binary classification formulation. Let R be the reference foreground area with pixel values r_n , and P be the predicted probabilistic map for the foreground area with values p_n . Assuming that the input has N image pixels, the formulations of three loss functions are expressed as follows.

1) Binary cross-entropy loss function²⁹ (BCEL):

$$BCEL = -\frac{1}{N} \sum_{n=1}^N [r_n \ln(p_n) + (1 - r_n) \ln(1 - p_n)]$$

2) Dice loss function³⁰ (DL):

$$DL = 1 - \left[\frac{2 \sum_{n=1}^N r_n p_n + \varepsilon}{\sum_{n=1}^N (r_n + p_n) + \varepsilon} + \frac{2 \sum_{n=1}^N (1 - r_n)(1 - p_n) + \varepsilon}{\sum_{n=1}^N (2 - r_n - p_n) + \varepsilon} \right]$$

The ε term is used as a smooth term to make sure the stability of the loss function by avoiding the problem of dividing by zero.

3) Percent overlap area loss function³¹ (POAL):

The percent overlap area of two binary images can be written as

$$P = \frac{\sum_{n=1}^N r_n p_n}{\sum_{n=1}^N r_n^2 + \sum_{n=1}^N p_n^2 - \sum_{n=1}^N r_n p_n} .$$

P can be differentiated with respect to the j -th pixel of the prediction, yielding the gradient

$$\frac{\partial P}{\partial p_j} = \frac{r_j (\sum_{n=1}^N r_n^2 + \sum_{n=1}^N p_n^2 - \sum_{n=1}^N r_n p_n) - (2p_j - r_j) (\sum_{n=1}^N r_n p_n)}{(\sum_{n=1}^N r_n^2 + \sum_{n=1}^N p_n^2 - \sum_{n=1}^N r_n p_n)^2} .$$

Therefore, we can express the two-class form of POAL as

$$POAL = 1 - \left\{ \frac{\sum_{n=1}^N r_n p_n + \varepsilon}{\sum_{n=1}^N (r_n + p_n - r_n p_n) + \varepsilon} + \frac{\sum_{n=1}^N (1 - r_n)(1 - p_n) + \varepsilon}{\sum_{n=1}^N [2 - r_n - p_n - (1 - r_n)(1 - p_n)] + \varepsilon} \right\} .$$

II. B.2 DCNN Training

205 The networks were trained with a mini-batched gradient decent algorithm. To alleviate the limitation of training samples and to improve the invariance and robustness properties of the networks, we used a combined affine transformation including rotation, shift and zoom for online data augmentation. Each combined affine transformation is a composition of elementary transformations from the following list: 1) rotation: rotation of the image by an angle up to 5
210 degrees; 2) shift: vertical or horizontal shift by a distance within 0.08 of the image size; 3) zoom: zooming of the focal lengths by a factor between 0.92 and 1.08. Points outside the boundaries of the input are filled by nearest neighbor interpolation. For each transformed image, the corresponding binary mask is transformed in the same way. We used a threshold value of 0.5 to keep the transformed mask a binary image after the data augmentation transformations. Previous work³² has
215 demonstrated that data augmentation using a combination of affine transformations to manipulate the training data was effective to increase the accuracy and generalizability of classification tasks. Rather than generating and saving augmentation data to hard disk, we generated them on the fly during training. With online data augmentation, the image after applying a combined affine transformation replaced the original image in a given epoch of training and each image was varied
220 randomly from epoch to epoch. Although the training samples did not increase in number, the “jittering” of the images by the combined affine transformation increased the variability of the training samples seen by the DCNN, thereby reducing the risk of overfitting to the limited training set. Figure 4 shows an image example with different transformations.

All weights of the network were initialized by randomly drawing from a normal distribution with a
225 mean of 0 and a standard deviation of 0.02. The batch size of images was 8 and the learning rate was 0.0001. We experimentally chose the number of epochs for training. Figure 5 shows an example of the training and validation loss of DCNN as a function of training epochs using the LP-DCNN with BCEL function from one of the 10-fold cross validation cycles. We observed that the loss of the network for the validation set had a broad minimum around 100 epochs and increased as
230 the training continued beyond about 200 epochs, indicating a trend of overfitting to the training set. We therefore chose to train with a fixed number of epochs at 100 and selected the best model based on the validation loss within 100 epochs for each cycle.

II. B.3 Boundary Identification of Pectoral Muscle

The neural network output a probability map, each pixel of which indicated the likelihood of the pixel being in the pectoral muscle region. We chose 0.5 as the threshold to segment the mammogram into a binary image with 1 being in the pectoral muscle candidate region and 0 otherwise. A connected component analysis was applied to the pectoral muscle candidate region to identify the largest connected component as the final pectoral muscle area. The boundary of the pectoral muscle was then generated by applying the Moore-Neighbor tracing algorithm modified by Jacob's stopping criteria to the pectoral muscle binary image. The Moore-Neighbor tracing algorithm is designed to find an ordered outline of a contour in an image. More details of the algorithm can be found in the literature³³.

II. C. Baseline Computer Vision Method

We have previously developed a computer vision method, which we referred to as the TFO method¹², for automated pectoral muscle identification. In this study, the TFO method was used as the baseline for comparison with our new DCNN approach. Briefly, we first used a gradient-based directional kernel (GDK) filter to enhance the linear texture structures with the breast region. The texture orientation image was further enhanced by a second GDK filter to extract the ridge point and remove noise. Finally, we applied a shortest-path finding method to track the ridge points with the highest cumulated probability that were likely lying on the pectoral boundary.

II.D. Performance Evaluation

We used three quantitative measures to evaluate the performance of the pectoral muscle segmentation using the experienced MQSA radiologist's manual segmentation as the reference standard. Let $H = \{h_1, h_2, \dots, h_p\}$ be the point set that contains p connected reference standard points representing the radiologist-drawn boundary, and $D = \{d_1, d_2, \dots, d_q\}$ be the DCNN-identified pectoral boundary point set that contains q connected points. The Euclidean distance

between a reference standard point h_i and a point d_j in D is denoted by $Dist(h_i, d_j)$. The three
 260 performance measures are defined as:

1) Percent overlap area³⁴ (POA)

$$POA(H, D) = \frac{A_H \cap A_D}{A_H \cup A_D}$$

where A_H and A_D are the reference standard pectoral muscle area and the DCNN-identified
 pectoral muscle area enclosed by the boundary point sets H and D , respectively.

2) Hausdorff distance³⁵ (Hdist)

$$Hdist = \max \left\{ \max_{i \in \{1, 2, \dots, p\}} \left\{ \min_{j \in \{1, 2, \dots, q\}} \{Dist(h_i, d_j)\} \right\}, \max_{j \in \{1, 2, \dots, q\}} \left\{ \min_{i \in \{1, 2, \dots, p\}} \{Dist(d_j, h_i)\} \right\} \right\}$$

265 3) Average Euclidean distance³⁶ (AvgDist)

$$AvgDist = \frac{1}{2} \left(\frac{1}{p} \sum_{i=1}^p \min_{j \in \{1, 2, \dots, q\}} \{Dist(h_i, d_j)\} + \frac{1}{q} \sum_{j=1}^q \min_{i \in \{1, 2, \dots, p\}} \{Dist(d_j, h_i)\} \right)$$

We used both the test set within DCNN cross-validation training and the independent test set for
 evaluation of different computer methods. The two-tailed paired t-test was used to estimate the
 statistical significance in the difference between the two methods and a p -value of less than 0.05 is
 270 considered statistically significant.

III. EXPERIMENTS AND RESULTS

III.A. DCNN Training

Previously, we used dataset I for testing the TFO methods¹². In this study, we used dataset I to
 275 train our DCNN because it was relatively large. During the training, 10-fold cross-validation was
 used to train and evaluate the performance of the DCNN. The training set consisted of both DFMs
 and DMs. We first randomly partitioned the DFM set and DM set separately into ten folds, and

then randomly combined one DFM fold and one DM fold to form the 10 final folds. In each training cycle, we used nine folds for training and validation, the remaining one fold was used for testing. Of the cases in the nine folds, 90% of them were randomly selected as training set and the remaining cases were used for validation. The trained model was applied to the left-out test fold. The results of the ten test folds were then pooled together to form a complete set for performance evaluation and compared to the baseline TFO method on the same test set. We also separately analyzed the results for the DFM and DM sets for comparison.

Table 2 showed the test results during 10-fold cross validations in which we compared three different network configurations. The BCEL function was used in this comparison. The differences in performances of both LP-DCNN vs MP-DCNN and LP-DCNN vs SP-DCNN were statistically significant for all three performance measures ($p < 0.05$). For the differences in performance of MP-DCNN vs SP-DCNN, the POA and Hdist were statistically significant while AvgDist did not reach statistical significance. We therefore chose the LP-DCNN as the final configuration for the rest of the study.

Table 2. Comparison of DCNN-identified pectoral muscle boundary with an experienced radiologist's manually identified pectoral muscle boundary on the test images from 10-fold cross validation using 637 MLO-view DFMs and 92 MLO-view DMs. The BCEL loss function was used for training the DCNNs.

We compared three loss functions including BCEL, DL and POAL during network training. Table 3 summarized the test performance measures for the loss functions. The differences in performances of both BCEL vs DL and BCEL vs POAL were statistically significant for all three performance measures ($p < 0.05$). The differences in performance between DL and POAL did not reach statistical significance.

Table 3. Comparison of three loss functions on the test images from 10-fold cross validation using 637 MLO-view DFMs and 92 MLO-view DMs. The LP-DCNN model was used.

III.B. Single-modality vs Mixed-modality Training

300 Our previous studies in breast cancer detection and diagnosis^{37,38} demonstrated the strong potential that DFMs can be effective supplemental training samples for DCNN method in mammographic breast cancer detection tasks when training samples of DMs are limited. To demonstrate the validity of this approach for the pectoral muscle segmentation task, we compared the performance of the DCNN trained on mixed DFM and DM with the performance trained on only DFM or DM alone.
 305 The LP-DCNN and BCEL function were used in this comparison. The 10 folds generated in III.A were used for the training and testing of the DCNN models used in this comparison. Each fold consisted of both DFMs and DMs. For the training and testing of DCNN on mixed DFM and DM, the entire fold was used. For the training and testing of the DCNN on either DFM or DM alone, only the DFM or only the DM in each fold was used, respectively. We also combined the test
 310 results from the DCNN trained on DFM alone and the DCNN trained on DM alone for comparison.

Table 4 showed the test results obtained from 10-fold cross validations in which we compared the DCNN trained on mixed DFM and DM to the DCNN trained on either DFM or DM. The differences in performance between the DCNN trained on mixed DFM and DM and the DCNN trained on only DFM or DM reached statistical significance for any of the three performance metrics ($p < 0.05$).

Table 4. Comparison of the DCNN trained on mixed DFM and DM with the DCNN trained on either DFM or DM on the test images from 10-fold cross validation. The LP-DCNN and BCEL function were used. The test results from the DCNN trained on DFM alone and the DCNN trained on DM alone were also combined into a whole data set for comparison.

315

III.C. DCNN vs Baseline Computer Vision Approach

We compared the DCNN method with our previous TFO method that was based on manually designed feature extraction techniques. The LP-DCNN trained with BCEL function was used for the comparison. Table 5 summarized the three performance measures between the DCNN and the TFO methods on dataset I, and the p -values estimated from two-tailed paired t-test for the differences in the corresponding performance measures. The differences in performance between DCNN and TFO reached statistical significance for any of the three performance metrics ($p < 0.05$) on dataset I.

Table 5. Comparison of DCNN method and TFO method on test images from 10-fold cross validation using 637 MLO-view DFMs and 92 MLO-view DMs. The LP-DCNN model and BCEL loss function were used.

Figure 6 showed the cumulative percentage of images as a function of a) percent overlap area, (b) Hausdorff distance, and (c) average distance on the test images from the 10-fold cross validation, corresponding to the results in Table 5. Of the 637 DFMs, 97.6% (622/637), 99.5% (634/637), 99.8% (636/637) of the DCNN-identified pectoral muscle boundaries had a POA greater than 90%, 85% and 80%, respectively. For the distance measures, 95.7% (610/637) and 99.3% (633/637) of the DCNN-identified pectoral muscle boundaries had Hausdorff distances within 5 and 10 mm from the reference boundaries, respectively, and 99.8% (636/637) of the DCNN-identified pectoral muscle boundaries had average distances within 5 mm from the reference boundaries. For DMs, 92.3% (85/92), 93.4% (86/92), 95.6% (88/92) of the DCNN-identified pectoral muscle boundaries had a POA greater than 90%, 85% and 80%, respectively. For the distance measures, 90.2% (83/92) and 95.6% (88/92) of the DCNN-identified pectoral muscle boundaries had Hausdorff distances within 5 and 10 mm from the reference boundaries, respectively, and 97.8% (90/92) of the DCNN-identified pectoral muscle boundaries had average distances within 5 mm from the reference boundaries.

III.D. Evaluation on Independent Test Set

340 We applied the ten LP-DCNN models trained with the BCEL function from the ten-fold cross
 validation to the 203 independent test DMs (dataset II). The independent test set was unknown to
 both the DCNN and the TFO methods so that the performance on this test set represented a fair
 comparison and validation of the robustness and the generalizability of the two methods²⁶. Table 6
 summarized the three performance measures obtained by comparing the pectoral muscle boundary
 345 identified by the two methods to the radiologist-provided reference standard. The DCNN method
 was significantly better than the TFO method for all three performance measures ($p < 0.05$).

Table 6. Comparison of the performance measures obtained with the DCNN method and TFO
 method on an independent test set of 203 DMs.

III.E. Observer Variability in Identifying Pectoral Muscle on MLO-view Mammogram

To evaluate the effect of the variability in the radiologist's manually identified pectoral muscle
 350 boundary on the relative ranking of the segmentation methods, we used a randomly selected subset
 of 106 DFMs for which the same experienced MQSA radiologist had manually drawn the pectoral
 muscle twice (denoted as R1 and R2) in an interval of time separated by about one year. Table 7
 summarized the agreement between pectoral muscle boundaries by the LP-DCNN trained with the
 BCEL function and the TFO method.

Table 7. Effect of the variability in the radiologist's manually identified pectoral muscle boundary.
 The comparison between pectoral muscle boundaries by the DCNN and TFO methods using 106
 MLO-view mammograms. R1 and R2 denote the first reading and second reading of the same
 experienced MQSA radiologist.

355 For the intra-observer variability evaluation, the R2 reading compared to the R1 reading showed a
 mean POA of $92.8\% \pm 4.85\%$, a mean Hdist of 4.44 ± 2.62 mm, and a mean AvgDist of 1.73 ± 1.31 mm.
 Two-tailed paired t-test was used to examine the differences in the corresponding performance
 measures between pairs of the comparisons. With either the radiologist's R1 reading or R2 reading

as the reference standard, the performance of the LP-DCNN was significantly better than the TFO
360 for any of the three performance metrics ($p < 0.05$).

IV. DISCUSSION & CONCLUSION

Breast density is one of the strong factors for breast cancer risk and more than half of the states in
the United States have passed the density notification law. Quantitative analysis of breast density
365 will improve the consistency and reproducibility as compared to visual assessment by radiologists
based on the Breast Imaging Reporting and Data System (BI-RADS) density categories, and
therefore will facilitate screening management as well as breast cancer risk estimation by computer-
assisted decision support tools. Pectoral muscle segmentation on MLO-view mammograms is an
essential process in automated multiple-view mammographic analysis and directly affects the
370 accuracy of the quantitative analysis of breast density and parenchymal patterns. Our previous
study found that the average difference between the pectoral areas segmented by two traditional
methods was larger than 160 mm^2 and the difference in the estimated percent breast density was
statistically significant ($p < 0.05$). In this study, we presented a detailed study of the deep learning
approach for automated identification of the pectoral muscle on MLO-view mammograms and
375 demonstrated that: (1) a DCNN could significantly outperform a previously developed method based
on conventional image processing techniques, (2) the pectoral muscle segmentation task requires a
U-Net with a relatively large number of channels to achieve high accuracy, (3) a mixed modality
training approach using both FFDM and DFMs is effective for the pectoral muscle segmentation
task and could alleviate the limited training sample problem in digital mammography, and (4)
380 among the three cost functions that were designed for segmentation tasks, the binary cross-entropy
loss function was the most effective for pectoral muscle segmentation. Although U-Net is a known
method for image segmentation, to our knowledge, this is the first study applying U-Net to pectoral
muscle segmentation on MLO-view mammograms and studied the impact of various network
configurations, loss functions, and training with DM and/or DFMs.

385 Our experimental results showed that the performance of the DCNN was superior to that of the
conventional computer vision approach on both the test set in cross validation and an independent

test set. The differences between the DCNN and the TFO methods were statistically significant ($p < 0.05$). In our previous study¹², we compared the TFO method with five reported methods. The five published methods were based on Hough transform¹⁰, Gabor filter^{39,40}, two methods using graph theory in conjunction with active contour¹¹, and Radon transform⁴¹, respectively. The comparison showed that the TFO method achieved higher performance than the previous methods. The results of the current study support our hypothesis that DCNN can accurately segment pectoral muscle on MLO-view mammograms, as indicated by the significantly higher performance measures in comparison to these conventional methods.

395 The false positive rate (FPR) and false negative rate (FNR) were calculated for pixel-based segmentation performance. The DCNN achieved a mean FPR of $0.11 \pm 0.11\%$ and a mean FNR of $2.06 \pm 4.15\%$, while the corresponding performance measures for TFO were $2.33 \pm 3.10\%$ and $2.88 \pm 3.19\%$, respectively. In the previous study³⁹ using Gabor filter, pectoral muscle segmentation was considered to be accurate when both FPR and FNR were less than 5%, acceptable when FPR and FNR were between 5% and 10%, and unacceptable if both FPR and FNR were more than 10%. The performance of our trained DCNN is therefore accurate and has better performance than the conventional methods in terms of FPR and FNR as shown in Table IV by Zhou *et al*¹². As previous study⁴ point out that patients with PD less than 5% has about 4 times lower in the relative risks of breast cancer than those with PD over 75%, the improvement of our DCNN is especially useful in
400
405 differentiate the fatty breast ($PD < 5\%$) from other breasts.

By examining the test results, we found that both the DCNN and the TFO methods performed well even when there was skin fold in the upper region of the breast mimicking pectoral muscle boundary. However, in the cases that dense glandular tissue overlapping with the pectoral muscle region, DCNN was more robust while TFO might fail. For some cases with severe fuzzy boundaries in the
410 lower region of the pectoral muscle, neither DCNN nor TFO provided good segmentation. This may be caused by insufficient training of this type of cases, which constituted only a small fraction of the training set. Figure 7 showed segmentation results by the two methods on several example mammograms with three types of challenging pectoral muscle patterns.

Generally, the training of DCNN models need a large number of annotated data. We had a
415 relatively small training set, especially DMs, to train our DCNN in this study. To overcome the
limitation of training samples, we used combined affine transformations for online data
augmentation in order to increase the robustness of the DCNN. In addition, we chose the U-net
architecture which was trained in pixel-wise manner so that the localized image features can be
effectively learned in each training cycle. The U-Net is trained for pixel-wise predictions in which
420 the value of each pixel in the output image represents the likelihood that the pixel is in the pectoral
muscle region. The contracting path in the U-Net model was able to capture multi-resolution image
context feature. The average training time of 100 epochs for this application was 1.5 hours using a
256 x 256 input image size on an NVIDIA Tesla K40 GPU with cuDNN v3 acceleration.

From the study of the effect of loss functions, we found that the DCNN trained using the BCEL
425 function performed the best in terms of all three performance measures, while the DCNN trained
using the POAL function did not perform the best in term of the POA metric. A DCNN trained with
different parameters may reach different local optima. Since different loss functions may provide
different gradient during the training process of the DCNN, it is possible that the DCNN trained
using the POAL loss function may reach a local optimum that does not perform the best in terms of
430 the POA metric. In addition, even if the network trained using the POAL loss function performed the
best in terms of the POA metric on the training set, the network may not perform the best on the test
set.

From the study of intra-observer variability, the agreement between the two readings by the same
radiologist is similar to the agreement between LP-DCNN and the radiologist's R2 reading. Both
435 the LP-DCNN and the TFO method showed better agreement with the radiologist's R1 reading.
Compared to the TFO method, LP-DCNN performed better in any of the three performance metrics
with either the radiologist's R1 reading or R2 reading as the reference standard. The small
variability of the radiologist's reference standard therefore did not affect the conclusions of our
study.

440 All the performance measures for the DM set were slightly, but consistently, lower than those of the
DFM set. Considering that the majority of the images were DFMs (637 out of 729), the DCNN did

adapt quite well for the DMs. The DFMs and DMs were images of the same anatomical structures, although the physical characteristics of the image detectors were different, which may result in different gray level contrast, noise, and resolution properties of the images. It appears that the DCNN was able to extract the relevant features despite the differences in the two types of images, whereas conventional image processing methods are generally more sensitive to these differences, as also observed from the larger drop in performance in DMs for the TFO method. This study further shows that DFMs can be effective supplemental training samples for DCNN in mammographic image analysis tasks when the primary image samples, e.g., DMs, are limited as demonstrated in our previous study³⁸.

There are several limitations in this study. First, we used an unbalanced mix of DFM set and DM set for the training of DCNN. The majority of the training set were DFMs. Although the knowledge learned from DFMs transferred well to DMs, as discussed above, one can expect that training the DCNN with a larger set of DMs will make it even more robust for DMs. We will continue to enlarge our DM set to improve the DCNN since DM systems have replaced the screen-film mammography systems in clinical practice. Second, the segmentation of pectoral muscle is only a first step in our image analysis pipeline. One of our applications is to analyze the density and pattern of the breast parenchyma on mammograms. We will study the effect of pectoral muscle segmentation on parenchymal analysis in the next step. Third, we did not compare U-net to other DCNN architectures such as FCN⁴², PSP net⁴³, and SegNet⁴⁴. These limitations will be addressed in future studies.

In summary, we have developed a DCNN for automated pectoral muscle identification on MLO-view mammograms. The results demonstrated that the pectoral muscle can be segmented accurately by our DCNN method. The mixed-modality approach to training DCNN is effective in alleviating the limited sample size problem of DMs. The trained DCNN is applicable to both DFMs and DMs, and achieves high accuracy for both modalities.

Acknowledgments

This work is supported by National Institutes of Health award number U01 CA195599 and China

470 Department of Science and Technology Key grant (No. 2016YFB0200602). Xiangyuan Ma, B.S.
 and Yao Lu, Ph.D. are also supported by grants from NSFC (Grant No. 81830052, 11401601), the
 Science and Technology Innovative Project of Guangdong Province, China (Grant Nos.
 2016B030307003, 2015B010110003, and 2015B020233008), Guangdong Provincial Science and
 475 Technology Key Grant (No. 2017B020210001), Guangzhou Science and Technology Creative Key
 Grant (No. 201604020003). The content of this paper does not necessarily reflect the position of
 the government and no official endorsement of any equipment and product of any companies
 mentioned should be inferred.

480

References

- 485 1. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2017. *CA: A Cancer Journal for Clinicians*.
 2017;67(1) :7-30.
2. Zuckerman HC. The role of mammography in the diagnosis of breast cancer. *Breast cancer,
 diagnosis and treatment*. 1987.152-172.
3. McCormack VA, dos Santos Silva I. Breast Density and Parenchymal Patterns as Markers
 490 of Breast Cancer Risk: A Meta-analysis. *Cancer Epidemiology Biomarkers & Prevention*.
 2006;15(6) :1159-1169.
4. Boyd NF, Guo H, Martin LJ, et al. Mammographic Density and the Risk and Detection of
 Breast Cancer. *New England Journal of Medicine*. 2007;356(3) :227-236.

5. Boyd NF, Martin LJ, Yaffe MJ, Minkin S. Mammographic density and breast cancer risk:
495 current understanding and future prospects. *Breast Cancer Research*. 2011;13(6) :223-
223.
6. Ganesan K, Acharya UR, Chua KC, Min LC, Abraham KT. Pectoral muscle segmentation: A
review. *Computer Methods and Programs in Biomedicine*. 2013;110(1) :48-57.
7. Goodsitt MM, Chan HP, Liu B, et al. Classification of compressed breast shapes for the
500 design of equalization filters in x-ray mammography. *Medical Physics*. 1998;25(6) :937-
948.
8. Hong BW, Sohn BS. Segmentation of Regions of Interest in Mammograms in a Topographic
Approach. *IEEE Transactions on Information Technology in Biomedicine*.
2010;14(1) :129-139.
- 505 9. Mirzaalian H, Ahmadzadeh MR, Sadri S. Pectoral Muscle Segmentation on Digital
Mammograms by Nonlinear Diffusion Filtering. Paper presented at: 2007 IEEE Pacific Rim
Conference on Communications, Computers and Signal Processing; 22-24 Aug. 2007.
10. Karssemeijer N. Automated classification of parenchymal patterns in mammograms.
Physics in Medicine & Biology. 1998;43(2) :365.
- 510 11. Ma F, Bajger M, Slavotinek JP, Bottema MJ. Two graph theory based methods for
identifying the pectoral muscle in mammograms. *Pattern Recognition*. 2007;40(9) :2592-
2602.

12. Zhou C, Wei J, Chan HP, et al. Computerized image analysis: Texture-field orientation method for pectoral muscle identification on MLO-view mammograms. *Medical Physics*. 2010;37(5) :2289-2299.
- 515
13. Taghanaki SA, Liu Y, Miles B, Hamarneh G. Geometry-Based Pectoral Muscle Segmentation From MLO Mammogram Views. *IEEE Transactions on Biomedical Engineering*. 2017;64(11) :2662-2671.
14. Fukushima K, Miyake S, Ito T. Neocognitron: A neural network model for a mechanism of visual pattern recognition. *IEEE Transactions on Systems, Man, and Cybernetics*. 1983;SMC-13:826-834.
- 520
15. Fukushima K. Neocognitron: A hierarchical neural network capable of visual pattern recognition. *Neural Networks*. 1988;1(2) :119-130.
16. Lo S-CB, Lin J-S, Freedman MT, Mun SK. Computer-assisted diagnosis of lung nodule detection using artificial convolution neural network. *Proc SPIE*. 1993;1898:859-869.
- 525
17. Chan HP, Lo SCB, Helvie MA, Goodsitt MM, Cheng S, Adler DD. Recognition of mammographic microcalcifications with artificial neural network. *Radiology*. 1993;189((P)) :318.
18. Chan H-P, Sahiner B, Lo S-C, et al. Computer-aided diagnosis in mammography: Detection of masses by artificial neural network. Paper presented at: 36th Annual Meeting of the American Association of Physicists in Medicine. *Medical Physics*. 1994;21:875-876.
- 530

19. Lo S-CB, Lou S-LA, Lin J-S, Freedman MT, Chien MV, Mun SK. Artificial convolution neural network techniques and applications for lung nodule detection. *IEEE transactions on medical imaging*. 1995;14 4:711-718.
- 535 20. Chan HP, Lo SCB, Sahiner B, Lam KL, Helvie MA. Computer - aided detection of mammographic microcalcifications: Pattern recognition with an artificial neural network. *Medical Physics*. 1995;22(10) :1555-1567.
21. Lo S-CB, Chan HP, Lin J-S, Li H, Freedman MT, Mun SK. Artificial convolution neural network for medical image pattern recognition. *Neural Networks*. 1995;8(7) :1201-1214.
- 540 22. Sahiner B, Chan HP, Petrick N, et al. Classification of mass and normal breast tissue: a convolution neural network classifier with spatial domain and texture images. *IEEE Transactions on Medical Imaging*. 1996;15(5) :598-610.
23. Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. Proceedings of the 25th International Conference on Neural Information
545 Processing Systems (Advances in Neural Information Processing Systems 25) ; 2012; 1097-1105.
24. Szegedy C, Ioffe S, Vanhoucke V, Alemi AA. Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning. Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence (AAAI-17) ; 2017: 4278-4284.
- 550 25. Rodriguez-Ruiz A, Teuwen J, Chung K, et al. Pectoral muscle segmentation in breast tomosynthesis with deep learning. *Proc SPIE*. 2018;10575:7.

26. Petrick N, Sahiner B, Armato SG, et al. Evaluation of computer-aided detection and diagnosis systems. *Medical Physics*. 2013;40(8) :087001.
27. Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image
555 Segmentation. 2015. arXiv:1505.04597.
28. Chollet F. Keras: <https://github.com/fchollet/keras>. *GitHub repository*. 2015.
29. Nasr GE, Badr EA, Joun C. Cross Entropy Error Function in Neural Networks: Forecasting Gasoline Demand. Proceedings of the Fifteenth International Florida Artificial Intelligence Research Society Conference; 2002.
- 560 30. Milletari F, Navab N, Ahmadi S-A. V-net: Fully convolutional neural networks for volumetric medical image segmentation. Paper presented at: 3D Vision (3DV) , 2016 Fourth International Conference 2016. IEEE.
31. Atiqur Rahman M, Wang Y. *Optimizing Intersection-Over-Union in Deep Neural Networks for Image Segmentation*. Vol 10072;2016.
- 565 32. Perez L, Wang J. The effectiveness of data augmentation in image classification using deep learning. *arXiv preprint arXiv:171204621*. 2017.
33. Ghuneim AG. Moore-Neighbor Tracing: http://www.imageprocessingplace.com/downloads_V3/root_downloads/tutorials/contour_tracing_Abeer_George_Ghuneim/moore.html.
- 570 34. Jaccard P. *The Distribution of the Flora in the Alpine Zone*. Vol 11;1912; 37-50.
35. Rockafellar RT, Wets RJ-B. *Variational analysis*. Vol 317: Springer-Verlag.; 2005; 117.
36. Anton H. *Elementary Linear Algebra (7th ed.)* . John Wiley & Sons; 1994; 170-171.

37. Samala RK, Chan H-P, Hadjiiski L, Helvie MA, Wei J, Cha K. Mass detection in digital breast tomosynthesis: Deep convolutional neural network with transfer learning from mammography. *Medical Physics*. 2016;43(12) :6654-6666.
- 575
38. Samala RK, Chan H-P, Hadjiiski LM, Helvie MA, Cha KH, Richter CD. Multi-task transfer learning deep convolutional neural network: application to computer-aided diagnosis of breast cancer on mammograms. *Physics in medicine and biology*. 2017;62 23:8894-8908.
39. Ferrari RJ, Rangayyan RM, Desautels JL, Borges R, Frere AF. Automatic identification of the pectoral muscle in mammograms. *IEEE transactions on medical imaging*. 2004;23(2) :232-245.
- 580
40. Burgess A. On the noise variance of a digital mammography system. *Medical physics*. 2004;31(7) :1987-1995.
41. Kinoshita SK, Azevedo-Marques PM, Pereira RR, Rodrigues JAH, Rangayyan RM. Radon-domain detection of the nipple and the pectoral muscle in mammograms. *Journal of digital imaging*. 2008;21(1) :37-49.
- 585
42. Shelhamer E, Long J, Darrell T. Fully Convolutional Networks for Semantic Segmentation. *IEEE Trans Pattern Anal Mach Intell*. 2017;39(4) :640-651.
43. Zhao H, Shi J, Qi X, Wang X, Jia J. Pyramid Scene Parsing Network. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) .* 2017.6230-6239.
- 590
44. Badrinarayanan V, Kendall A, Cipolla R. SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2017;39:2481-2495.

Author Manuscript

Table 1. The number of channels at the different layers of the DCNN for the three models studied.

Parameter	LP-DCNN	MP-DCNN	SP-DCNN
a	32	16	8
b	64	32	16
c	64	32	16
d	128	64	32
e	128	64	32
f	256	128	64
g	256	128	64
h	256	128	64
i	256	128	64
j	128	64	32
k	128	64	32
l	64	32	16
m	64	32	16

Table 2. Comparison of DCNN-identified pectoral muscle boundary with an experienced radiologist's manually identified pectoral muscle boundary on the test images from 10-fold cross validation using 637 MLO-view DFMs and 92 MLO-view DMs. The BCE loss function was used for training the DCNNs.

	Method	POA (%)	Hdist (mm)	AvgDist (mm)
DFM	LP-DCNN	96.8±2.0	2.18±1.17	0.73±0.41
	MP-DCNN	95.7±4.4	2.66±1.71	0.98±0.99
	SP-DCNN	95.5±4.1	2.78±1.64	0.99±0.75
DM	LP-DCNN	94.9±5.0	2.85±1.78	1.11±1.03
	MP-DCNN	93.9±4.5	3.78±2.70	1.39±1.02
	SP-DCNN	93.4±5.1	3.89±2.80	1.42±0.99

	LP-DCNN	96.5±2.9	2.26±1.31	0.78±0.58
Whole dataset*	MP-DCNN	95.4±4.8	2.80±2.00	1.03±1.10
	SP-DCNN	95.2±4.5	2.92±1.93	1.04±0.85

^aData are mean ± standard deviation.

* Two-tailed paired t-test was used to examine the differences in the pairwise performances in the whole data set between the DCNN models (LP-DCNN vs MP-DCNN, LP-DCNN vs SP-DCNN, and MP-DCNN vs SP-DCNN):

P-values of POA are p<0.001, p<0.001 and p=0.046.

P-values of Hdist are p<0.001, p<0.001 and p=0.044.

P-values of AvgDist are p<0.001, p<0.001 and p=0.662.

Table 3. Comparison of three loss functions on the test images from 10-fold cross validation using 637 MLO-view DFMs and 92 MLO-view DMs. The LP-DCNN model was used.

	Loss function	POA (%)	Hdist (mm)	AvgDist (mm)
DFM	BCEL	96.8±2.0	2.18±1.17	0.73±0.41
	DL	95.8±4.5	2.59±1.96	1.02±1.49
	POAL	96.0±3.5	2.63±1.78	0.93±0.96
DM	BCEL	94.9±5.0	2.85±1.78	1.11±1.03
	DL	94.0±4.6	3.58±2.54	1.30±0.94
	POAL	94.1±4.7	3.63±2.76	1.29±0.99
Whole dataset*	BCEL	96.5±2.9	2.26±1.31	0.78±0.58
	DL	95.6±4.8	2.72±2.22	1.05±1.53
	POAL	95.7±3.9	2.76±2.08	0.98±1.03

^aData are mean ± standard deviation.

*Two-tailed paired t-test was used to examine the differences in the pairwise performances in the whole data set between the loss functions (BCEL vs DL, BCEL vs POAL and DL vs POAL):

P-values of POA are p<0.001, p<0.001 and p=0.274.

P-values of Hdist are p<0.001, p<0.001 and p=0.561.

P-values of AvgDist are p<0.001, p<0.001 and p=0.173.

Table 4. Comparison of the DCNN trained on mixed DFM and DM with the DCNN trained on either DFM or DM on the test images from 10-fold cross validation. The LP-DCNN and BCEL function were used. The test results from the DCNN trained on DFM alone and the DCNN trained on DM alone were also combined into a whole data set for comparison.

Test	Training	POA (%)	Hdist (mm)	AvgDist (mm)
DFM	Mixed DFM and DM	96.8±2.0	2.18±1.17	0.73±0.41
	Only DFM	95.7±4.5	2.64±2.15	0.95±1.32
DM	Mixed DFM and DM	94.9±5.0	2.85±1.78	1.11±1.03
	Only DM	87.9±11.8	5.41±3.72	2.09±1.54
Whole dataset*	Mixed DFM and DM	96.5±2.9	2.26±1.31	0.78±0.58
	Only DFM and only DM	94.7±6.5	2.99±2.57	1.09±1.40

^aData are mean ± standard deviation.

*Two-tailed paired t-test was used to examine the differences in the pairwise performance measures in the whole data set between the DCNN trained on mixed DFM and DM and the DCNN trained on either DFM or DM:

P-values of POA, Hdist and AvgDist are $p < 0.001$, $p < 0.001$ and $p < 0.001$.

Table 5. Comparison of DCNN method and TFO method on test images from 10-fold cross validation using 637 MLO-view DFMs and 92 MLO-view DMs. The LP-DCNN model and BCEL loss function were used.

	Method	POA (%)	Hdist (mm)	AvgDist (mm)
DFM	LP-DCNN	96.8±2.0	2.18±1.17	0.73±0.41
	TFO	95.0±3.6	3.45±2.16	1.12±0.82
DM	LP-DCNN	94.9±5.0	2.85±1.78	1.11±1.03
	TFO	89.3±12.7	5.37±4.74	2.54±4.04
Whole dataset*	LP-DCNN	96.5±2.9	2.26±1.31	0.78±0.58

TFO	94.2±6.1	3.69±2.79	1.30±1.69
-----	----------	-----------	-----------

^aData are mean ± standard deviation.

* Two-tailed paired t-test was used to examine the differences in the pairwise performance measures in the whole data set between the LP-DCNN and the TFO methods:

P-values of POA, Hdist and AvgDist are $p < 0.001$, $p < 0.001$ and $p < 0.001$.

Table 6. Comparison of the performance measures obtained with the DCNN method and TFO method on an independent test set of 203 DMs.

	Method	POA (%)	Hdist (mm)	AvgDist (mm)
Independent test set [*]	LP-DCNN	93.7±6.9	3.80±3.21	1.49±1.62
	TFO	86.9±16.0	7.18±14.22	3.98±14.13

^aData are mean ± standard deviation.

* Two-tailed paired t-test was used to examine the differences in the pairwise performance measures in the independent test set between the LP-DCNN and the TFO methods.

P-values of POA, Hdist and AvgDist are $p < 0.001$, $p < 0.001$, and $p = 0.013$.

Table 7. Effect of the variability in the radiologist's manually identified pectoral muscle boundary. The comparison between pectoral muscle boundaries by the DCNN and TFO methods using 106 MLO-view mammograms. R1 and R2 denote the first reading and second reading of the same experienced MQSA radiologist.

	POA (%)	Hdist (mm)	AvgDist (mm)
R1 vs R2	92.8±4.9	4.44±2.62	1.73±1.31
TFO vs R1	93.3±5.1	4.13±2.60	1.46±1.18
LP-DCNN vs R1	95.3±4.6	2.80±2.01	0.96±0.87
TFO vs R2	91.3±6.2	4.28±2.96	1.97±2.09

LP-DCNN vs R2

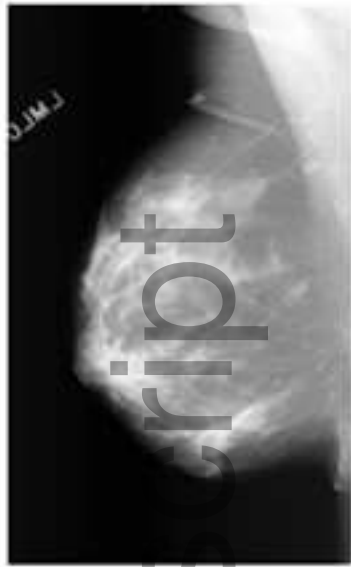
92.0±5.9

3.68±2.43

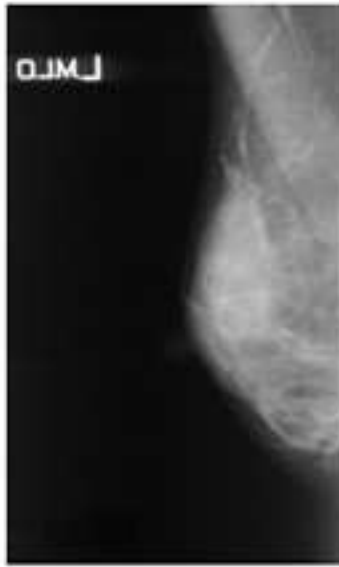
1.66±1.41

^aData are mean ± standard deviation.

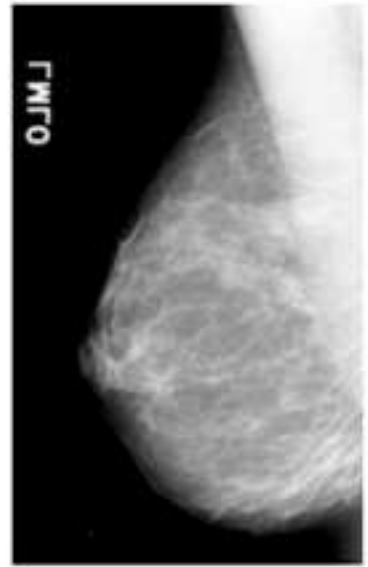
Author Manuscript



(a)



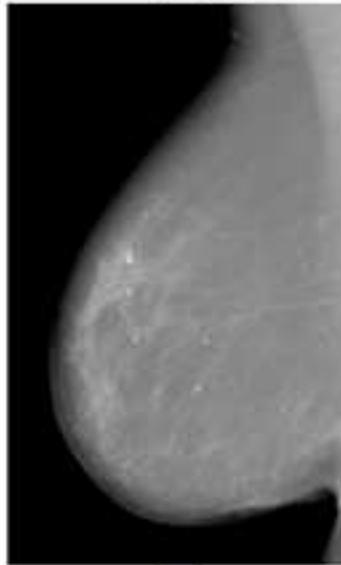
(b)



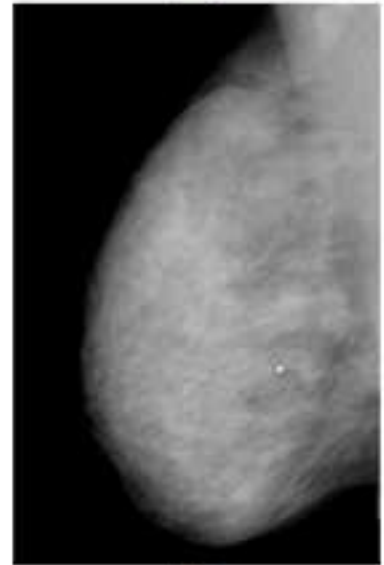
(c)



(d)



(e)



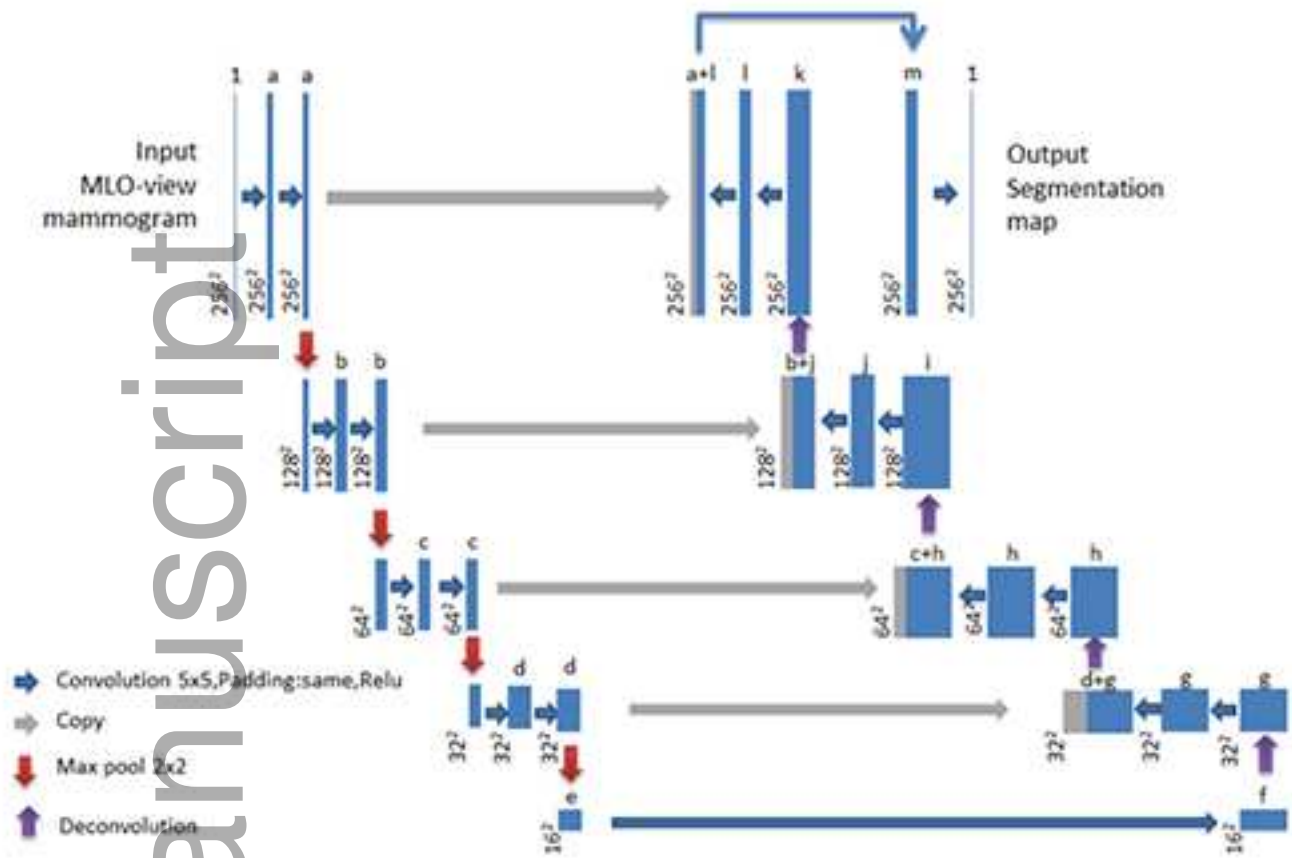
(f)

mp_13451_f1.tif

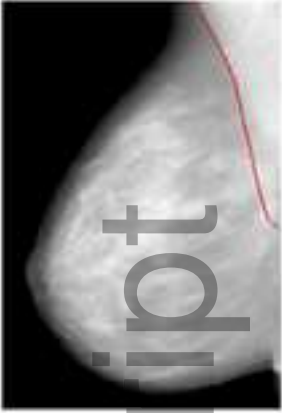


mp_13451_f2.tif

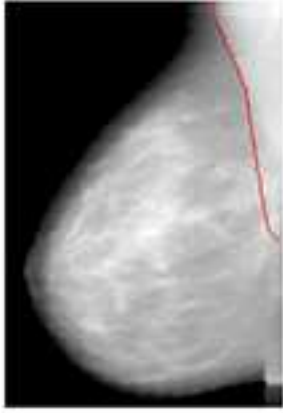
Author Manuscript



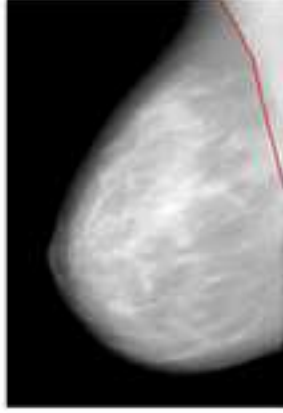
mp_13451_f3.tif



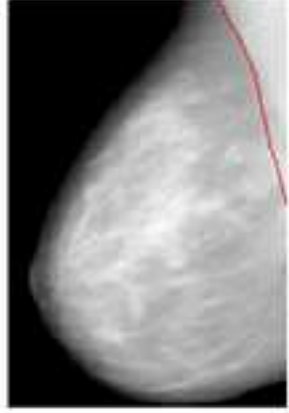
(a)



(b)

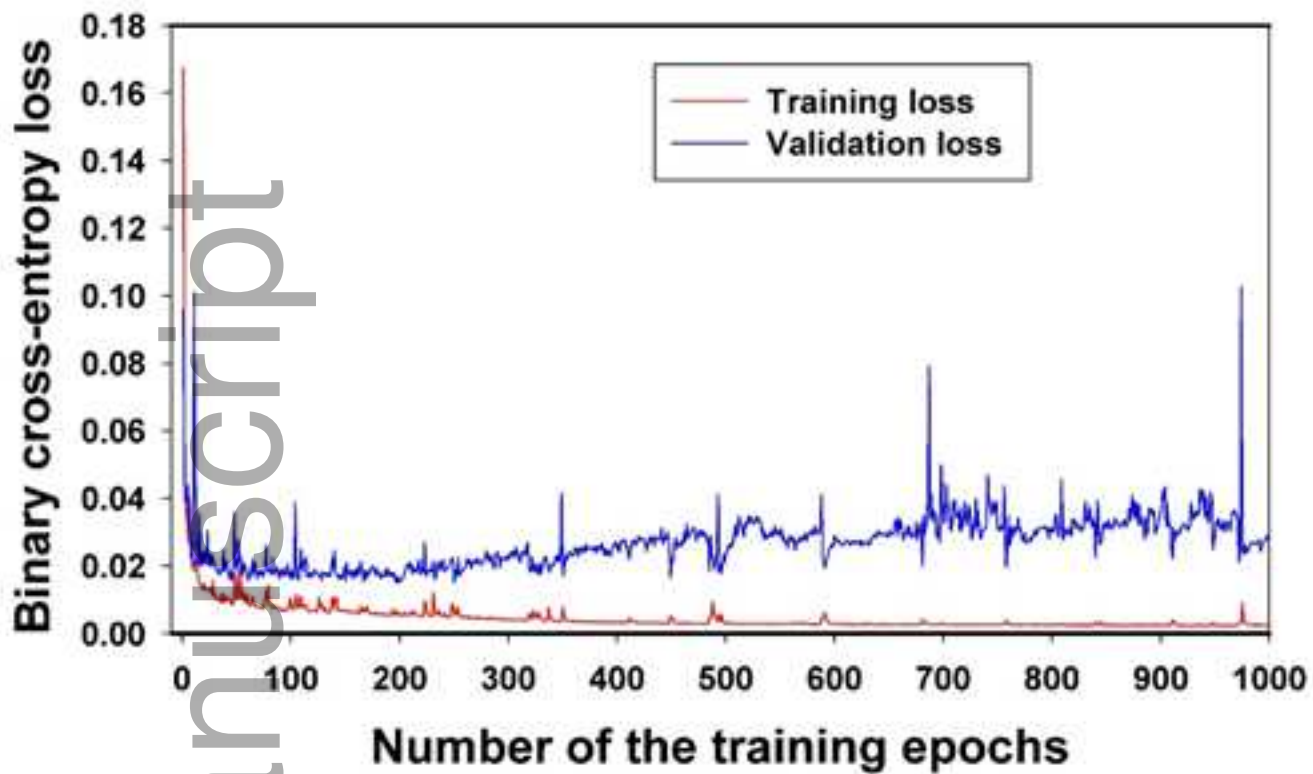


(c)

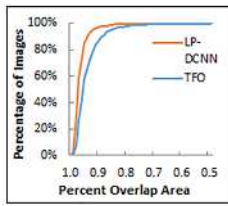


(d)

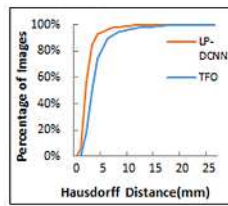
mp_13451_f4.tif



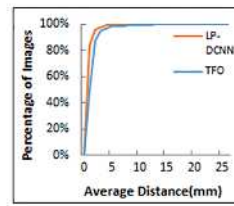
mp_13451_f5.tif



(a)

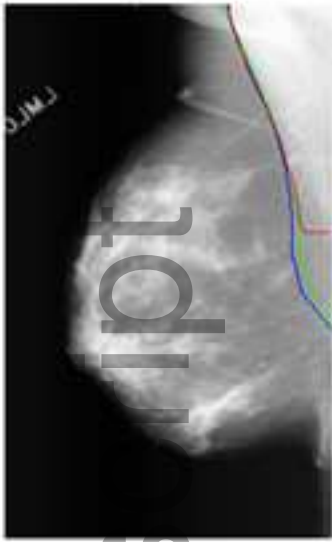


(b)

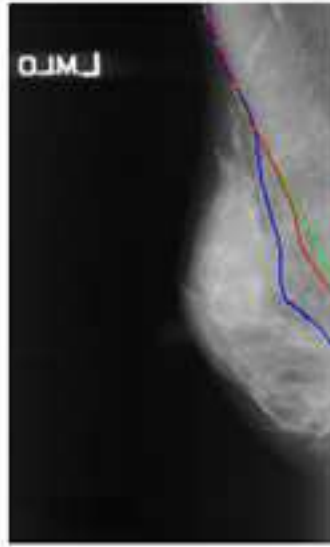


(c)

mp_13451_f6.tif



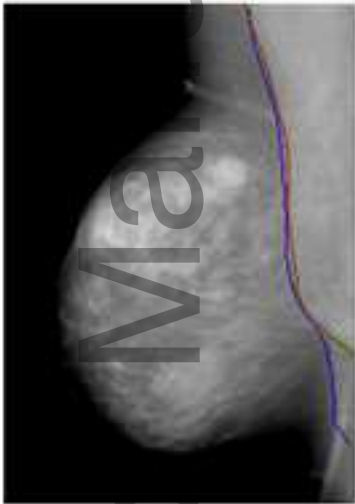
(a)



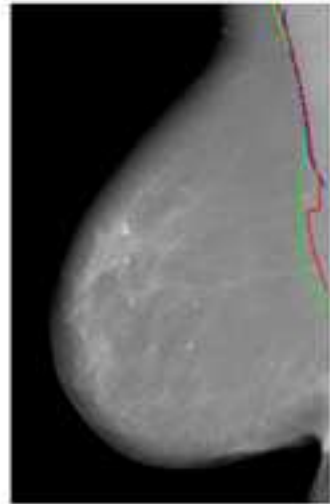
(b)



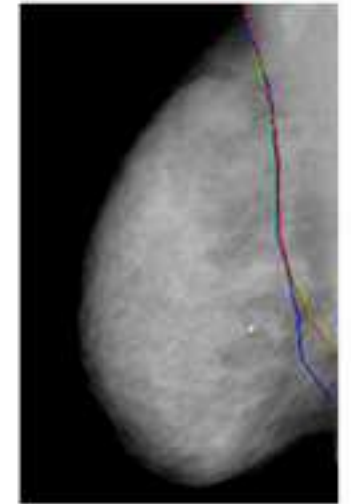
(c)



(d)



(e)



(f)

mp_13451_f7.tif