

Mammographic mass segmentation using multichannel and multiscale fully convolutional networks

Shengzhou Xu^{1,2}  | Ehsan Adeli³ | Jie-Zhi Cheng⁴ | Lei Xiang⁵ | Yang Li² | Seong-Whan Lee⁶ | Dinggang Shen^{2,6}

¹College of Computer Science, South-Central University for Nationalities, Wuhan, China

²Department of Radiology and BRIC, University of North Carolina at Chapel Hill, North Carolina, USA

³Department of Psychiatry and Behavioral Sciences, Stanford University, California, USA

⁴Shanghai United Imaging Intelligence Co., Ltd., Shanghai, China

⁵School of Biomedical Engineering, Shanghai Jiao Tong University, Shanghai, China

⁶Department of Brain and Cognitive Engineering, Korea University, Seoul, Republic of Korea

Correspondence

Dinggang Shen, Department of Radiology and BRIC, University of North Carolina at Chapel Hill, NC 27599.
Email: dgshen@med.unc.edu

Funding information

Institute for Information & Communications Technology Promotion (IITP), Grant/Award Number: 2017-0-00451; the Fundamental Research Funds for the Central Universities, Grant/Award Number: CZY19011; the National Natural Science Foundation of China, Grant/Award Number: 61302192

Abstract

Breast cancer is one of the leading causes of death among women worldwide. Mammographic mass segmentation is an important task in mammogram analysis. This process, however, poses a prominent challenge considering that masses can be obscured in images and appear with irregular shapes and low image contrast. In this study, a multichannel, multiscale fully convolutional network is proposed and evaluated for mass segmentation in mammograms. To reduce the impact of surrounding unrelated structures, preprocessed images with a salient mass appearance are obtained as the second input channel of the network. Furthermore, to jointly conduct fine boundary delineation and global mass localization, we incorporate more crucial context information by learning multiscale features from different resolution levels. The performance of our segmentation approach is compared with that of several traditional and deep-learning-based methods on the popular DDSM and INbreast datasets. The evaluation indices consist of the Dice similarity coefficient, area overlap measure, area undersegmentation measure, area oversegmentation measure, and Hausdorff distance. The mean values of the Dice similarity coefficient and Hausdorff distance of our proposed segmentation method are 0.915 ± 0.031 and 6.257 ± 3.380 , respectively, on DDSM and 0.918 ± 0.038 and 2.572 ± 0.956 , respectively, on INbreast, which are superior to those of the existing methods. The experimental results verify that our proposed multichannel, multiscale fully convolutional network can reliably segment masses in mammograms.

KEYWORDS

fully convolutional network, mammogram, mass segmentation, multichannel, multiscale

1 | INTRODUCTION

Breast cancer alone accounts for 30% of all new cancer diagnoses in women in the United States.¹ Among the treatments proposed for breast cancer, carrying out early diagnosis and early treatment has been considered the major approach to boosting breast cancer survival. With

mammography being one of the standard technologies utilized for the early detection and diagnosis of breast cancer,² over 40 million mammographic exams are performed in the United States every year.^{3,4} Automatic segmentation of masses in mammograms is crucial for further quantitative and qualitative analysis. However, mass segmentation remains quite challenging since

mammograms are 2D projected images and a mass can be difficult to discern when the surrounding breast structures share a similar intensity distribution with the mass. Meanwhile, the masses may have irregular shapes, low contrast, and varying sizes.

There are many studies related to mass segmentation in mammograms. These studies propose methods that can be categorized into two major categories: traditional segmentation methods and deep-learning-based methods. The traditional methods can be further classified as region-based methods, edge-based methods, and hybrid methods. With respect to the region-based segmentation methods, the region growing algorithm,^{5,6} watershed transform,^{7,8} and morphological filter method⁹ have been adopted, improved, and applied to mass segmentation. These methods aim to detect regions that satisfy certain predefined homogeneity criteria of breast masses. Among the edge-based methods, active contour models¹⁰⁻¹³ and dynamic programming¹⁴⁻¹⁶ are the two most frequently used methods for mass segmentation. These methods search an optimal path that represents the mass edge by optimizing a predefined energy function. It is often difficult to obtain satisfactory results when using only one of these methods to perform mass segmentation, as each of them focuses only on a single aspect/characteristic of the mass. Therefore, some hybrid methods,¹⁷⁻¹⁹ which combine region-based and edge-based techniques, have been proposed to further improve segmentation accuracy. However, all these traditional methods for mass segmentation rely heavily on elaborated hand-crafted features.²⁰

Recent advances in deep learning techniques have made breakthroughs in breast mass segmentation in mammograms. For example, Dhungel et al²¹ demonstrated the benefit of combining multiple potential functions based on deep learning, a Gaussian mixture model, and shape prior for mass segmentation. They also made use of deep convolution and deep belief networks as potential functions in structured prediction models to improve the results of mass segmentation.²¹ One of the main deficiencies of these methods, however, is that these two-stage training procedures are prone to overfit the training data²⁰ and hence induce the generated models to form dependencies on the datasets used for training. Fully convolutional networks (FCNs)²² can produce accurate and detailed segmentations by combining global semantic information from a deep, coarse layer with local details from a shallow, fine layer. U-Net,²³ a more elegant architecture based on the FCN, consists of a contracting path used to learn image context and a symmetric expanding path used to maintain geometrical correctness. U-Net was originally proposed for biomedical image segmentation. It uses solely the original images as

the input of the network and does not make use of other complementary information from other sources. In order to comprehensively characterize the structural information of PVSS and segment PVSS more accurately, Lian et al²⁴ proposed a multi-channel multi-scale network as an extension of the original FCNs. Considering that both shape and appearance priori play important roles in mammogram mass segmentation, Zhu et al²⁰ empirically estimated a priori shape on each of the DDSM-BCRP²⁵ and INbreast²⁶ training datasets and assigned them to the bias values of the last layer of the FCN separately. However, it is important to note that the actual shapes of the masses clearly vary, even within the same dataset,²⁷ that is, the general priori shape does not adapt well to each individual. Singh et al²⁸ proposed a conditional generative adversarial network (cGAN) for mass segmentation. The algorithm was tested on INbreast and a private in-house dataset.

To overcome the difficulties inherent in mammograms and improve the limitations of the existing methods, in this article, we develop an end-to-end framework based on FCNs that can effectively fuse multichannel, multiscale cues to delineate the contour of masses in the regions of interest (ROIs) (see Figure 1). The major contributions of this study are 2-fold: (1) our proposed method takes the shape information into consideration and treats the preprocessed image with the salient appearance of the mass as a distinctive input channel of the network. That is, one channel loads the preprocessed images with the salient appearance of the mass, and another channel loads the original images to provide the image details. (2) The proposed FCN is equipped with the capability of exploring multiscale cues to jointly conduct both fine boundary delineation and global mass localization. Extensive experimental results on two public datasets, that is, DDSM and INbreast, demonstrate the efficiency of the multichannel, multiscale FCN compared with several state-of-the-art methods.

2 | MATERIALS AND METHODS

2.1 | Materials

The data used in this study are chosen from two publicly available and widely used mammographic datasets: the DDSM²⁵ and INbreast²⁶ datasets. The DDSM is the largest publicly available dataset of digitized mammograms.

Mammograms from the DDSM used in our study are acquired using the LUMYSIS scanner with a resolution of 50 μm . In our previous research, to reduce the computational complexity, mammograms from the DDSM were

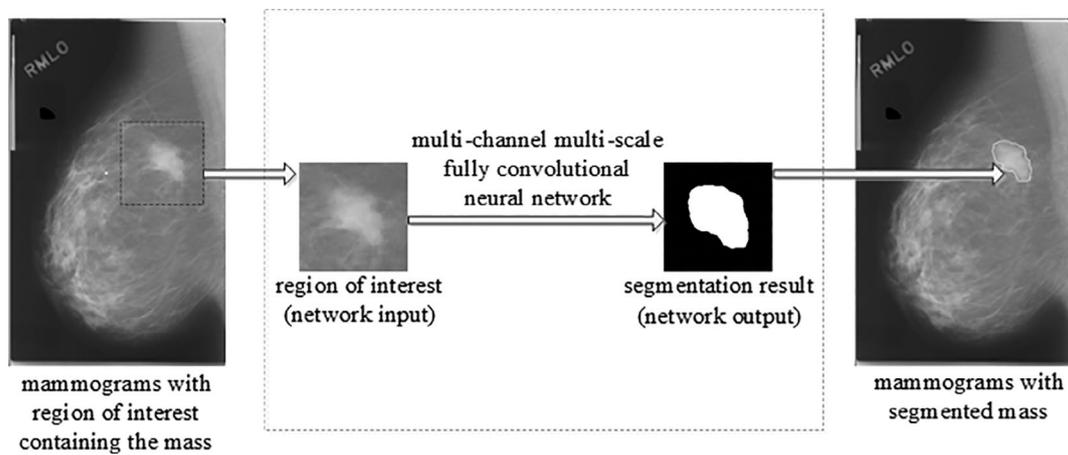


FIGURE 1 Mass segmentation from mammograms based on multichannel multiscale fully convolutional network (FCN)

modified as follows: they were scaled down from 50 to 400 μm per pixel, and the gray level was reduced from 4096 to 256 levels. As mentioned in previous works,^{15,16,26} the ground truth data provided by the DDSM provide only information about the locations and types and not the accurate outlines of suspicious regions. Therefore, outlines for each selected mass are delineated by one radiologist with 2 years of clinical experience on the scaled images; these outlines are then maintained or redrawn by another radiologist with more than 20 years of clinical experience and used as the ground truth. Furthermore, ROIs with a size of 128×128 pixels centered on the geometric center and estimated based on the radiologists' outlines are extracted from the downsampled and rescaled mammograms. In our experiments, 483 ROIs, including 214 benign and 269 malignant masses, are randomly selected from 473 mammograms of 328 patients in the DDSM. These masses have various margin types, sizes, and densities. The types of margins for the benign and malignant masses include 105 circumscribed, 111 spiculated, 141 ill-defined, 79 microlobulated, and 47 obscured masses. The masses are characterized by the effective radius of the area inside the contour delineated by radiologists. The maximum, mean, and minimum radii of these masses in the dataset are 20.1, 7.7, and 2.6 mm (ie, 50, 19, and 7 pixels), respectively.

Different from the digitized-mammogram DDSM, the digital-mammogram INbreast database provides accurately annotated contours of masses. For mass segmentation, it specifically provides 56 cases (11 benign and 45 malignant) containing 116 masses with a pixel size of 70 μm . The maximum, mean, and minimum radii of these masses in the dataset are 34.3, 12.5, and 2.2 mm, respectively. For ease of comparison with the state-of-the-art methods, we obtained the same ROIs provided by the authors of the previous works.^{20,21} That is, the rectangular

ROI is generated from the annotated bounding box of each mass by expanding the bounding box by 20%.²⁹ In doing so, the sizes of the ROIs are associated with the sizes of the masses they contain, that is, a large mass corresponds to a large ROI, and a small mass has a small ROI. Finally, all the ROIs are resized to 40×40 pixels using bicubic interpolation and used for the training set and test set. It is important to note that some cases in the DDSM and INbreast databases contain multiple masses.

2.2 | Method

In this section, our multichannel, multiscale fully convolutional network is outlined in detail and then used for the segmentation of masses in breast images. We describe the overall network architecture and then introduce each key component.

2.2.1 | Overall network architecture

Figure 2 illustrates the architecture of the proposed multichannel, multiscale network. Similar to the structure of U-Net,²³ our network consists of a contracting path (the left side of the network) and a symmetric expanding path (the right side of the network). There are a total of four input contracting paths, two for the scales, each of which has two channels for the filtered and original images. The contracting path consists of three steps (see Figure 2). The number of feature maps at each scale of the first step is 64, which is doubled for the second step with the concatenation of the two-scale features (see the \oplus symbol in Figure 2) and so on. Detailed descriptions of each step are discussed in the next subsection. Following the contracting path of multiscale features are four 3×3 convolutions

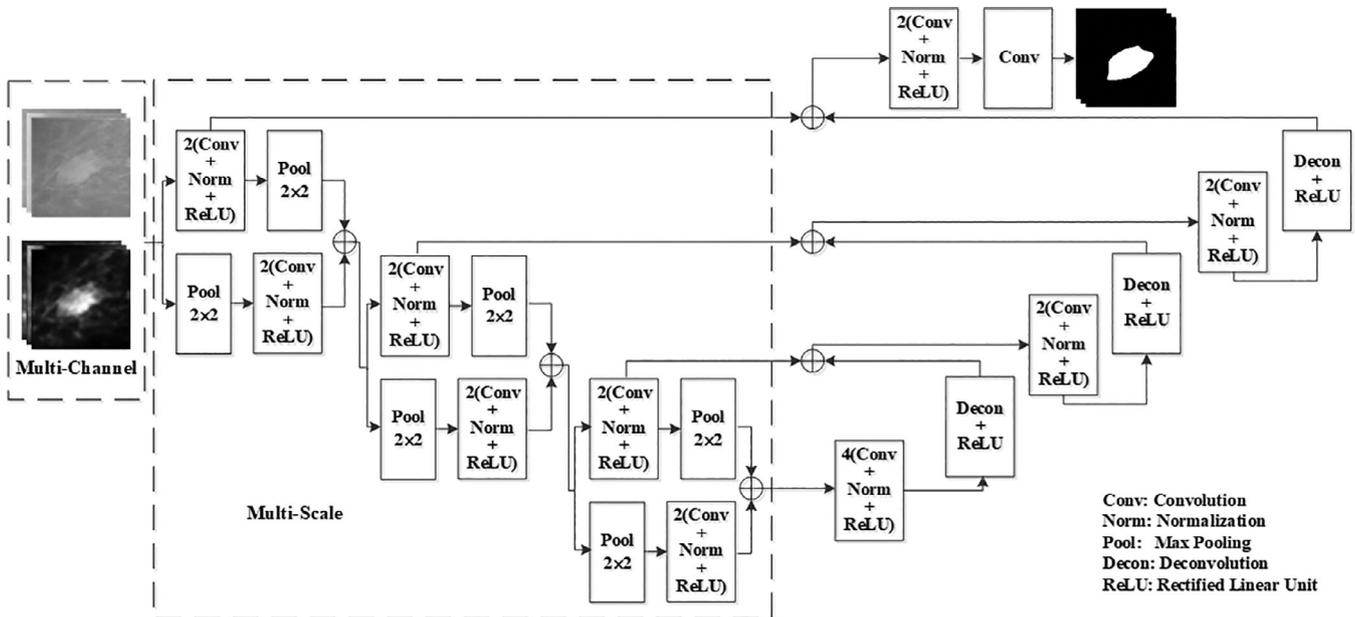


FIGURE 2 Architecture of the proposed multichannel multiscale fully convolutional network. Symbol \oplus denotes the concatenation of features along the channel axis

with a stride and padding of 1, each of which is followed by a batch normalization (Norm) and a rectified linear unit (ReLU), implemented as transition layers for the latter expanding path. It should be noted that the architecture of our network differs from the M^2EDN proposed by Lian et al²⁴ in the aspect of the transition layers and the contracting path (such as the number of convolution and normalization operations, and the way to obtain multiscale features, etc.). The expanding path is similarly composed of three steps, each of which consists of a 4×4 deconvolution with a stride of two (ie, halving the number of feature maps). These are then followed by a concatenation with the corresponding feature map from the contracting path and two 3×3 convolutions with a stride of 1, each followed by a batch normalization and a ReLU. At the last step of the expanding path, we obtain two channels of feature maps. Each channel has the same dimensions as those of the input image, representing the probability that the pixel in the corresponding position belongs to the background or the mass. During the training process, the Dice similarity coefficient (DSC)³⁰ (defined in Section 2.2.5.) is employed as the loss function to drive the learning of mass segmentation.

2.2.2 | Multiscale network architecture

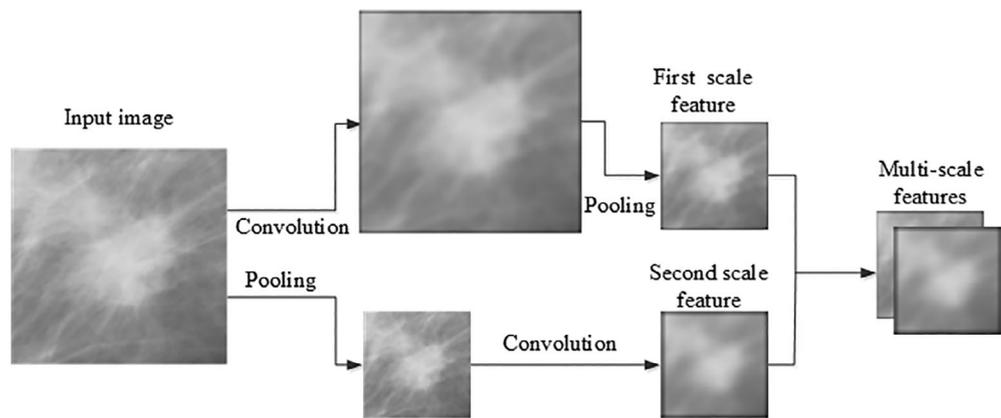
In FCNs³¹ (and similar models), the features near the input layer of the network are equipped with a small receptive field size to discover useful local cues and define fine boundary details, while the features near the

output layer, with large context, capture the global traits of the location. Therefore, by learning features at multiple scales, we can jointly conduct fine boundary delineation and global mass localization. This observation was the main motivation behind proposing a multiscale FCN architecture, as illustrated in Figure 2. Specifically, multiscale features are learned by concatenating two scales of features in the contracting path. An illustration of the concatenating procedure is shown in Figure 3. The first-scale features are obtained using two pairs of 3×3 convolutional operations with a stride and padding of 1, each followed by a batch normalization and a ReLU, applied to the outputs of the previous layer. Subsequently, these feature maps are downsampled by a 2×2 max pooling at a stride of two. The first-scale features provide useful information for the localization of image parts (ie, the mass). Meanwhile, outputs of the previous layer are downsampled by a max-pooling with a kernel size of 2×2 and stride of 2, after which the second-scale features are computed with serial operations of convolutions, batch normalizations, and ReLUs on the downsampled images. Hence, these features can encode larger context cues. Then, the first- and second-scale features are concatenated to form multiscale features, which carry fine boundary details and global mass localization.

2.2.3 | Multichannel inputs

Generally, in FCNs, the segmentation relies on the competition between the target region and the background

FIGURE 3 The schematic illustration of multiscale features acquisition. The convolution operations are always followed by batch normalization and ReLU



region (which is outside the target contour). Each pixel in the input image is estimated to be either the target or the background. In this process, most FCN models, such as U-Net,²³ use only the relevant information from the original images as the input to learn a segmentation model and often do not incorporate other complementary information from other sources. However, in many applications, including ours, the presence of the surrounding unrelated structures (in the input images), which may have similar gray values to those of the target, may hinder the estimation of the network model. Therefore, incorporating complementary knowledge can improve the segmentation performance. As argued by Jiang et al,³² shape and appearance priors are important cues for the task of mass segmentation. In mammograms, masses may vary in size and shape and can obscure appearance and yield low contrast. To properly separate background tissues from the masses for the segmentation task, we take the shape and appearance information into consideration through a distinct input channel in the network. To this end, we use one channel to load a preprocessed image with the salient appearance of the mass and another channel to input the corresponding original image. The former introduces the shape and appearance priors to the network, while the latter contains pixel-level details.

A schematic diagram of the preprocess carried out to form the multichannel input is shown in Figure 4. To reduce the influence of the background on the mass, a plane representing the trend of the grayscale distribution of the ROI is initially fitted and subtracted from the original ROI to form an enhanced ROI. Then, a template that exhibits the visual and statistical properties of masses 17 is designed to filter the enhanced ROI and make the mass more salient (as explained in detail below). Finally, the original and filtered images are combined as separate channels of an image to form the multichannel input. The aforementioned process is described in detail in the following.

The plane equation representing the trend of the grayscale distribution of the original image¹⁶ is denoted by:

$$z = f(x, y) = a_0x + a_1y + a_2, \quad (1)$$

where z is the gray value of the pixel with coordinates x and y in the ROI and a_0 , a_1 and a_2 are the polynomial coefficients. Letting n denote the number of pixels in the ROI, the n points (x_i, y_i, z_i) , $i = 0, 1, 2, \dots, n - 1$ used to fit the plane in Equation (1) should meet the following condition.

$$\min S = \sum_{i=0}^{n-1} (a_0x_i + a_1y_i + a_2 - z_i)^2. \quad (2)$$

Computing the partial derivative S with respect to a_0 , a_1 , and a_2 , we have:

$$\begin{vmatrix} \sum x_i^2 & \sum x_i y_i & \sum x_i \\ \sum x_i y_i & \sum y_i^2 & \sum y_i \\ \sum x_i & \sum y_i & n \end{vmatrix} \begin{pmatrix} a_0 \\ a_1 \\ a_2 \end{pmatrix} = \begin{pmatrix} \sum x_i z_i \\ \sum y_i z_i \\ \sum z_i \end{pmatrix} \quad (3)$$

Then, the plane image can be obtained by solving the above linear equations to obtain the values of the coefficients a_0 , a_1 , and a_2 . The original image is subtracted from the fitted plane image and then normalized, which leads to the enhanced image, as shown in Figure 4. By carrying out this procedure, the mass region in the image is obviously enhanced.

To further suppress small, bright tissues around the mass, a template that exhibits the visual and statistical properties of masses is designed to filter the image. In mammograms, masses tend to be brighter than their neighboring pixels and somewhat circular, although they exhibit weak fading boundaries with neighboring tissues.³³ This tendency has driven us to define the template as follows.

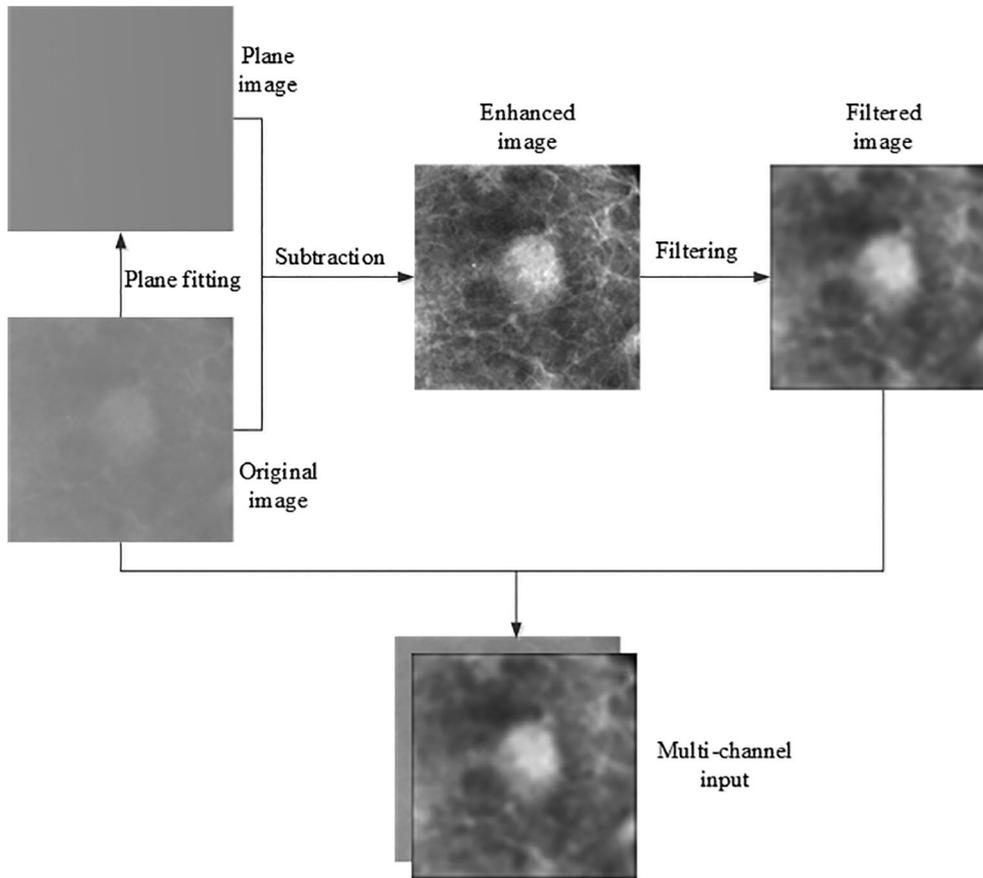


FIGURE 4 The schematic diagram of multichannel input of the network

$$T(x,y) = \frac{2}{e^{\beta \times \sqrt{x^2+y^2}} + e^{-\beta \times \sqrt{x^2+y^2}}} \quad (4)$$

The center of the template image is defined as the coordinate origin, and $T(x, y)$ is the gray value of the pixel, with coordinates x and y in the template image. The gray values of the template image gradually weaken from the center to the periphery with increasing values of x and y . The size of the template image and the parameter β , which controls the rate of changes in the gray values, are set to 5×5 pixels and 0.08 respectively (obtained by trial and error) in this study. We filter the original images with the template image to obtain the enhanced images. A sample result of the filtering is shown in Figure 4. Finally, we take the original and filtered images as multichannel inputs for the network. The filtered image provides the salient appearance of the mass, and the original image supplies the image details.

2.2.4 | Data augmentation

Similar to other deep learning methods, data augmentation is helpful to improve model training, particularly when only a few training samples are available.^{23,34} In

this study, we conduct a data augmentation strategy for both the original and the preprocessed images by flipping the images in the horizontal and vertical directions, rotating the images by 90, 180, and 270°, and deforming the images with three random elastic deformations. Therefore, there are another eight different copies of ROIs generated from the described augmentation scheme. For the random elastic deformations, first, three deformation fields with the same image size are produced, with the value of each of their elements being a random number between -1 and 1 . Meanwhile, a rotationally symmetric Gaussian lowpass filter with a size of 7×7 pixels and a SD of 4 is created. Then, the deformation fields are smoothed by the Gaussian filter and applied to the images to produce new deformed images.

2.2.5 | Segmentation evaluation

To facilitate comparison with other algorithms, the area overlap measure (AOM) and Dice similarity coefficient (DSC) are used to evaluate the performance of the segmentation method. These measures use the overlap to quantify the consistency between segmented results and the ground truth. The AOM is defined as the ratio of the

intersection to the union of the segmented and the ground-truth areas,¹⁷ that is,

$$AOM = \frac{A_{seg} \cap A_{gs}}{A_{seg} \cup A_{gs}} \quad (5)$$

where A_{seg} denotes the segmented area and A_{gs} is the ground-truth area. The DSC is defined similarly as³⁰

$$DSC = \frac{2 \times (A_{seg} \cap A_{gs})}{A_{seg} + A_{gs}} \quad (6)$$

We also use two additional metrics: area under-segmentation measure (AUM) and area oversegmentation measure (AVM). These metrics are used to evaluate the segmentation accuracy of the target and background, respectively. The AUM and AVM are defined as¹⁷

$$AUM = \frac{A_{gs} - A_{seg}}{A_{gs}} \quad (7)$$

$$AVM = \frac{A_{seg} - A_{gs}}{A_{seg}} \quad (8)$$

With these definitions, higher values of the AOM and DSC are desirable, while for the AUM and AVM, lower values are sought.

In addition, the Hausdorff distance (HD) is used to quantify the consistency between the contours obtained with the segmentation method and the radiologist's manually delineated outlines. Suppose that the two boundaries are composed of m and n pixels, respectively, and are denoted as $P = \{p_1, p_2, \dots, p_m\}$ and $Q = \{q_1, q_2, \dots, q_n\}$, respectively. The HD is defined as¹⁷

$$HD(P, Q) = \max \left(\max_{i \in \{1, \dots, m\}} \{d(p_i, Q)\}, \max_{j \in \{1, \dots, n\}} \{d(q_j, P)\} \right) \quad (9)$$

where $d(p_i, Q) = \min_{j \in \{1, \dots, n\}} \|p_i - q_j\|$ is the distance from p_i to the closest point on the contour Q .

3 | RESULTS

In this section, we first introduce the experimental settings and then present the results of our method in comparison with those of other baseline methods.

3.1 | Experimental settings

The proposed segmentation method is extensively evaluated on the DDSM²⁵ and INbreast.²⁶ For 483 masses with 5 different margin types (as described above) in the DDSM, 83 ROIs are selected from each subset of various margin types as the test set, and the remaining 400 ROIs are selected as the training set. We ensure that all the ROIs from the same case are divided into the same partition (ie, test set or training set) according to the file name of each ROI, which contains case information. On this basis, our experimental results are compared with the manual ground truth and those of several traditional and deep-learning-based segmentation methods.

In these state-of-the-art works in mammogram mass segmentation presented by Zhu et al and Dhungel et al,^{20,21} the 116 masses of INbreast are divided into mutually exclusive training and testing sets, containing 58 masses each. According to the file names of the ROIs, we ensure that ROIs from the same case are assigned to the same group. For direct comparison, we use the same experimental setup for those methods.

Our network is trained on Caffe³⁵ with the stochastic gradient descent (SGD) optimizer. During training, the values of the momentum, learning rate, weight decay and maximum iteration are set to 0.9, 0.001, 0.00005, and 14 000 by trial and error, respectively. To prevent overfitting, a dropout of 0.2 is used between two consecutive convolutional layers in our network. For the loss function, we use the Dice loss layer instead of the SoftMax, with the loss computed at the pixel level on the output of the network with respect to the ground-truth segmentation of the corresponding input image. All of the segmentation algorithms proposed by us in this study use the same settings for the hyperparameters described above.

3.2 | Segmentation results

We compared the results of the multichannel input with multiscale strategy (MCMS) with several baseline and state-of-the-art segmentation methods, including U-Net,²³ template matching and dynamic programming (TMDP),¹⁷ and marker-controlled watershed (MCWS).⁷ In our experiments, the MCWS and TMDP methods were implemented as described by Xu et al⁷ and Song et al,¹⁷ respectively. U-Net was implemented on the basis of the procedure described by Ronneberger et al.²³ The architecture used in U-Net is consistent with that of our MCMS, except that it does not have the second-scale features at each layer within the contracting path; its input is only one channel composed of the original images. In addition, its loss function (Dice) and training parameters are

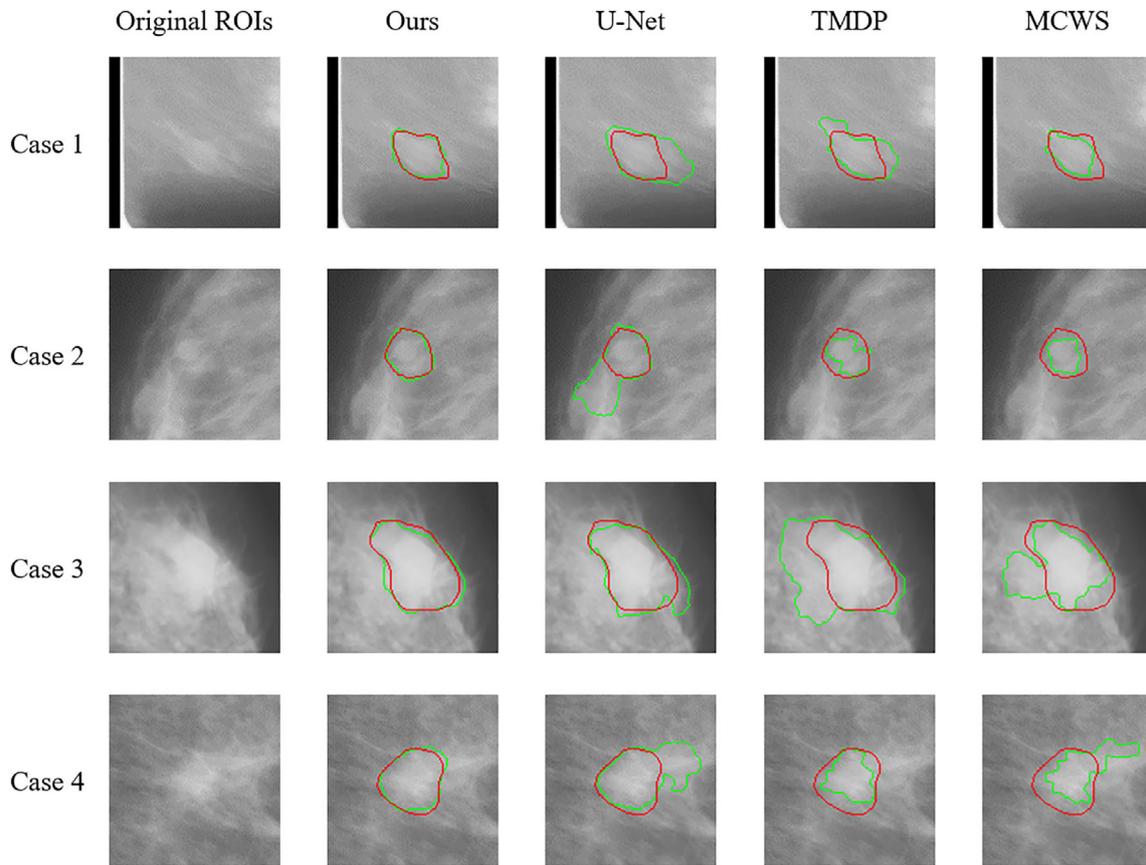


FIGURE 5 Segmentation results of different methods. Each row corresponds to one testing sample. Red lines denote ground-truth segmentation, and green lines indicate segmentation results for each method

consistent with those of our MCMS method, as described previously. Note that, for fair comparison, all baseline methods share the same data partition. Figure 5 demonstrates examples of the mass segmentation results of all comparison methods on four cases from the DDSM. As can be seen, compared to the other methods, there is much better alignment between our segmented contours and the ground truth.

For further quantitative analysis of the segmentation results on the DDSM, Table 1 summarizes the performances of all methods with several statistical indices, including the minimum, first quartile, mean, median, third quartile, and maximum values for the five metrics. The minimum DSC value of our model is 0.781, which is much higher than that of the other methods (ie, 0.465, 0.568, and 0.576). This value illustrates that our method is more robust than other comparison methods and can perform more favorably given the task of mass segmentation. Additionally, our results can better match the manual delineations than those of the other comparison methods, with relatively higher average DSC values and lower SDs: the mean \pm SD of the DSC value of our model is 0.915 ± 0.031 , while those for U-Net, TMDP and MCWS are 0.894 ± 0.066 , 0.845 ± 0.088 , and 0.822

± 0.083 , respectively. The comparison of the AOM measures of our model and the other methods is similar to that of the DSC. According to Student's *t*-test, the *P*-values for the distribution of the DSC and AOM metrics of our method and those of the other three algorithms are all $<.001$, which indicates that the difference between our method and the other three algorithms is statistically significant. For the HD metric, the mean value of our model is 6.257 pixels (2.503 mm), which is much lower than that of the other methods. The *P*-values in a paired Student's *t*-test for the distribution of the HD of our method and that of the TMDP, MCWS, and U-Net are 4.392E-09, 3.995E-12, and 0.013, respectively.

Furthermore, it is important to note that the results of U-Net listed in Table 1 are obtained with the same data augmentation and Dice loss function as used in our MCMS method. If no data augmentation is considered, the DSC of U-Net will drop dramatically from 0.894 ± 0.066 to 0.837 ± 0.103 . Although it can be seen from Table 1 that U-Net with data augmentation and the Dice loss function performs better than the MCWS and TMDP methods according to all five measures, its AVM value is still high and needs to be further improved to alleviate oversegmentation. By taking the shape prior and

TABLE 1 Performance comparison of different segmentation methods on DDSM dataset

Method	Minimum	First quartile	Mean	Median	Third quartile	Maximum
Descriptive statistics of DSC						
Ours (MCMS)	0.781	0.900	0.915 ± 0.031	0.921	0.938	0.968
U-Net*	0.465	0.872	0.894 ± 0.066	0.905	0.926	0.963
TMDP*	0.505	0.793	0.845 ± 0.088	0.868	0.910	0.965
MCWS*	0.576	0.742	0.822 ± 0.083	0.823	0.863	0.969
Descriptive statistics of AOM						
Ours (MCMS)	0.684	0.818	0.844 ± 0.051	0.853	0.883	0.938
U-Net*	0.302	0.774	0.811 ± 0.091	0.826	0.862	0.929
TMDP*	0.397	0.657	0.740 ± 0.122	0.767	0.834	0.933
MCWS*	0.404	0.603	0.718 ± 0.123	0.726	0.789	0.940
Descriptive statistics of AUM						
Ours (MCMS)	0.004	0.043	0.085 ± 0.061	0.066	0.118	0.307
U-Net	0.002	0.032	0.087 ± 0.096	0.061	0.126	0.694
TMDP*	0.002	0.052	0.148 ± 0.116	0.122	0.222	0.556
MCWS*	0.008	0.138	0.165 ± 0.123	0.145	0.257	0.608
Descriptive statistics of AVM						
Ours (MCMS)	0.002	0.047	0.081 ± 0.043	0.082	0.108	0.195
U-Net*	0.013	0.068	0.112 ± 0.076	0.097	0.139	0.434
TMDP*	0.000	0.042	0.133 ± 0.136	0.099	0.157	0.650
MCWS	0.000	0.012	0.107 ± 0.140	0.034	0.115	0.634
Descriptive statistics of HD						
Ours (MCMS)	1.414	3.803	6.257 ± 3.380	5.000	7.545	18.790
U-Net	2.000	4.298	7.257 ± 4.696	5.657	9.110	28.284
TMDP*	2.828	6.325	9.146 ± 5.070	8.062	10.507	33.287
MCWS*	3.000	7.071	11.040 ± 5.982	10.000	12.120	34.010

Note: Symbol * denotes the P -value for the distributions of this metric between our model and the corresponding method is lower than .01. Abbreviations: AOM, area overlap measure; AUM, area undersegmentation measure; AVM, area oversegmentation measure; DSC, Dice similarity coefficient; HD, Hausdorff distance; MCMS, multichannel multiscale fully convolutional network; MCWS, marker-controlled watershed; TMDP, template matching and dynamic programming.

multiscale features into consideration, our MCMS model yields an obvious improvement in the segmentation results compared to those of U-Net.

For easy comparison with other state-of-the-art methods in mammogram mass segmentation, we use the same protocol of these works^{20,21}: the first 58 masses of INbreast are used for training, and the other 58 masses are used for testing. Table 2 shows the mean values of the DSC for both our proposed method and other methods, including adversarial deep structured nets (ADSN),²⁰ deep learning and structured prediction (DLSP),²¹ and U-Net.²³ The P -values of a paired Student's t -test for the distribution of the DSC, AOM and HD of our method and those of U-Net are 6.277E-08, 2.942E-08, and 0.015, respectively. The results of the U-Net method are obtained with data augmentation and the Dice loss function, as in our settings. It can be

observed that our model still achieved higher DSC values than those of the other methods on the INbreast dataset. This is mainly due to the multichannel nature of our method, as we incorporate appearance and shape priors to facilitate the segmentation task. This is particularly important for the mass segmentation in mammograms because the mass is surrounded by structures with similar gray values in the original image. In addition, we learn features at multiple scales; thus, we can jointly conduct fine boundary delineation and global mass localization.

To evaluate each component of our proposed method, we also perform ablation studies in terms of multichannel and multiscale factors. Let SCMS-O denote the single input channel with the original ROI and the multiscale strategy scheme, and let SCMS-E denote the single input channel with the enhanced ROI and the multiscale

strategy scheme. The MCMS, SCMS-O, SCMS-E and MCSS methods share the same network structure and hyperparameters, but their input datasets are different; thus, they are trained independently and have their data optimized independently. First, we compare SCMS-O, SCMS-E, and the multichannel with both ROIs and the multiscale strategy scheme (MCMS) and report their performance in Table 3. Compared to the use of a single input channel (SCMS-O and SCMS-E), the undersegmentation rate (ie, AUM) and the oversegmentation rate (ie, AVM) can achieve a better balance when the salient appearance and the details of the mass are provided by the multichannel inputs (in our method, MCMS). Correspondingly, the DSC, AOM, and HD performance metrics of the segmentation are improved for both the DDSM and INbreast datasets. The P -values of a paired Student's t -test for the distribution of the DSC, AOM and AUM between the MCMS and either the SCMS-O or SCMS-E are all less than 0.01 on both the DDSM and INbreast datasets (see Table 3).

TABLE 2 Performance comparison of different segmentation methods on INbreast dataset

Method	DSC	AOM	HD
Ours (MCMS)	0.918 ± 0.038	0.851 ± 0.062	2.572 ± 0.956
U-Net	$0.891 \pm 0.040^*$	$0.806 \pm 0.064^*$	2.972 ± 0.971
DLSP	0.900	—	—
ADSN	0.909	—	—

Note: Symbol * denotes the P -value for the distributions of this metric between our model and the corresponding method is lower than .01.

Abbreviations: ADSN, adversarial deep structured nets; AOM, area overlap measure; DLSP, deep learning and structured prediction; DSC, Dice similarity coefficient; HD, Hausdorff distance; MCMS, multichannel multiscale fully convolutional network.

TABLE 3 The average values of the five measures for single-channel and multichannel methods

Dataset	Method	DSC	AOM	AUM	AVM	HD
DDSM	SCMS-O	$0.898 \pm 0.051^*$	$0.821 \pm 0.076^*$	$0.106 \pm 0.083^*$	0.084 ± 0.068	6.796 ± 4.298
	SCMS-E	$0.902 \pm 0.037^*$	$0.824 \pm 0.061^*$	$0.107 \pm 0.068^*$	0.077 ± 0.062	6.691 ± 4.115
	MCMS	0.915 ± 0.031	0.844 ± 0.051	0.085 ± 0.061	0.081 ± 0.043	6.257 ± 3.380
INbreast	SCMS-O	$0.902 \pm 0.045^*$	$0.825 \pm 0.072^*$	$0.046 \pm 0.035^*$	$0.139 \pm 0.080^*$	2.793 ± 1.286
	SCMS-E	$0.901 \pm 0.048^*$	$0.824 \pm 0.075^*$	$0.096 \pm 0.067^*$	0.094 ± 0.070	2.805 ± 1.335
	MCMS	0.918 ± 0.038	0.851 ± 0.062	0.076 ± 0.044	0.083 ± 0.069	2.572 ± 0.956

Note: Symbol * denotes the P -value for the distributions of this metric between our MCMS model and the corresponding method is lower than .01.

Abbreviations: AOM, area overlap measure; AUM, area undersegmentation measure; AVM, area oversegmentation measure; DSC, Dice similarity coefficient; HD, Hausdorff distance; MCMS, multichannel input with multiscale strategy; SCMS-E, single input channel with the enhanced ROI and multiscale strategy scheme; SCMS-O, single input channel with original ROI and multiscale strategy scheme.

For the second ablation study, we compare the multichannel input with the single-scale strategy scheme (MCSS) and the multichannel input with the multiscale strategy (MCMS) to illustrate the efficiency of our multiscale strategy. For the MCSS, we use only the first-scale features (with no second-scale features incorporated), that is, the network structure is similar to that of the original U-Net. The results of the ablation study based on the DDSM and INbreast are shown in Table 4. It can be seen that, compared with the use of single-scale features, both the undersegmentation rate and oversegmentation rate are reduced when the fine boundary details and global mass localization are jointly utilized by the multiscale features, the DSC and AOM values of the segmentation result are obviously improved, and the P -values of the paired Student's t -test for the distribution of the DSC and AOM between the MCMS and MCSS are all less than .01.

4 | DISCUSSION

The MCMS network is evaluated on two widely used datasets: the digitized-mammogram DDSM and the digital-mammogram INbreast dataset. For the DDSM dataset, we extract 483 ROIs and test our algorithm on it. The results are compared with those of other popular algorithms, such as U-Net, TMDP and MCWS, on these ROIs. For direct and fair comparison with other state-of-the-art methods (DLSP and ADSN) based on the INbreast datasets, we use the ROIs provided by the authors²⁰ and apply our MCMS network to these ROIs. For the DLSP and ADSN methods, we do not have access to their codes and quote only the results from their respective papers.

The results show that the MCMS network is very useful in mass segmentation on both digitized and digital mammograms. Segmentation of masses in digital mammograms is a practical application because almost all

TABLE 4 The average values of the five measures for single-scale and multiscale methods

Dataset	Method	DSC	AOM	AUM	AVM	HD
DDSM	MCSS	0.903 ± 0.036*	0.825 ± 0.059*	0.114 ± 0.073*	0.070 ± 0.060	6.410 ± 3.407
	MCMS	0.915 ± 0.031	0.844 ± 0.051	0.085 ± 0.061	0.081 ± 0.043	6.257 ± 3.380
INbreast	MCSS	0.902 ± 0.048*	0.823 ± 0.075*	0.096 ± 0.067*	0.094 ± 0.070	2.805 ± 1.335
	MCMS	0.918 ± 0.038	0.851 ± 0.062	0.076 ± 0.044	0.083 ± 0.069	2.572 ± 0.956

Note: Symbol * denotes the *P*-value for the distributions of this metric between our MCMS model and the corresponding method is lower than .01.

Abbreviations: AOM, area overlap measure; AUM, area undersegmentation measure; AVM, area oversegmentation measure; DSC, Dice similarity coefficient; HD, Hausdorff distance; MCMS, multichannel input with multiscale strategy; MCSS, multichannel input with single scale strategy scheme.

TABLE 5 The average values of the five measures for the generalization performance

Training set	Test set	DSC	AOM	AUM	AVM	HD
DDSM	INbreast	0.849 ± 0.068	0.741 ± 0.092	0.085 ± 0.117	0.195 ± 0.091	12.300 ± 5.816
INbreast	DDSM	0.853 ± 0.056	0.748 ± 0.080	0.223 ± 0.092	0.042 ± 0.047	3.838 ± 1.008

Abbreviations: AOM, area overlap measure; AUM, area undersegmentation measure; AVM, area oversegmentation measure; DSC, Dice similarity coefficient; HD, Hausdorff distance.

mammography systems are currently digital. Therefore, in this study, we verify whether the digitized-mammogram-trained MCMS network is generalizable to digital mammograms and vice versa. On the one hand, we deploy the DDSM-trained MCMS network to the 58 ROIs in the INbreast test set. These ROIs from the INbreast database and their corresponding ground-truth images are scaled to 128×128 pixels to fit the network input. On the other hand, we also deploy the INbreast-trained MCMS network to all the ROIs in the DDSM dataset. These ROIs are resized according to the size of the masses they contain and then scaled to 40×40 pixels to fit the network input. The performances of the generalization study between the DDSM and INbreast are shown in Table 5. It can be seen that, compared with the training set and the test set from the same dataset (see Tables 1 and 2), the average values of the DSC and AOM measures of the segmentation results are obviously lower and the values of the HD measure are higher when they are from different datasets. This may be because the mammograms come from different types of devices, the ROIs are extracted in different ways and the masses are delineated by different radiologists with different standards. Despite this, our results are still very close to the results of the traditional methods (see Tables 1 and 5). From Table 5, we can also find that the AUM value is low and that the AVM value is high when the DDSM-trained MCMS network is applied to INbreast. Conversely, when the INbreast-trained MCMS network is applied to the DDSM, the AUM value is high and the AVM is low. This result further illustrates the inconsistency criteria for

mass delineation between the two datasets: for the same mass, the area delineated by the DDSM may be larger than the area defined by the INbreast dataset. In future research, we will strive to unify the delineation standards, the radiologists and the ROI extraction process for these two datasets.

5 | CONCLUSIONS

A multichannel, multiscale fully convolutional network is introduced in this study to address the difficult problem of mass segmentation in mammograms. Through the proposed method, we obtained an accurate contour of the mass. On the one hand, we applied preprocessed images with a salient mass appearance as the second input channel of the network to reduce the impact of surrounding unrelated structures. On the other hand, multiscale features learned from various resolution levels were incorporated to conduct fine boundary delineation and global mass localization. As the results show, our proposed method outperforms other baseline and state-of-the-art methods in terms of several assessment metrics.

Despite the superior performance on the segmentation task, there are still some issues worthy of further research, such as analysis of the segmentation performance for the various types of margins and the consideration of benign vs malignant masses. Another direction for future research could be to develop fully automated detection and segmentation algorithms.

ACKNOWLEDGMENTS

Dr. S. Xu was supported in part by the National Natural Science Foundation of China (Grant No. 61302192), and the Fundamental Research Funds for the Central Universities (CZY19011). Dr. S.-W. Lee was partially supported by Institute for Information & Communications Technology Promotion (IITP) grant funded by the Korea government (No. 2017-0-00451).

ORCID

Shengzhou Xu  <https://orcid.org/0000-0003-4285-9588>

REFERENCES

- Siegel RL, Miller KD, Jemal A. Cancer statistics, 2018. *CA Cancer J Clin.* 2018;68(1):7-30.
- Samala RK, Chan HP, Hadjiiski L, Helvie MA, Wei J, Cha K. Mass detection in digital breast tomosynthesis: deep convolutional neural network with transfer learning from mammography. *Med Phys.* 2016;43(12):6654-6666.
- Haas JS. The complexity of achieving the promise of precision breast cancer screening. *J Natl Cancer Instit.* 2017;109(5):1-2.
- Yuan Y, Giger ML, Li H, Suzuki K, Sennett C. A dual-stage method for lesion segmentation on digital mammograms. *Med Phys.* 2007;34(11):4180-4193.
- Mencattini A, Rabottino G, Salmeri M, Lojacocono R, Colini E. Breast mass segmentation in mammographic images by an effective region growing algorithm. *International Conference on Advanced Concepts for Intelligent Vision Systems.* Springer, Berlin, Heidelberg; 2008.
- Rouhi R, Jafari M, Kasaei S, Keshavarzian P. Benign and malignant breast tumors classification based on region growing and CNN segmentation. *Expert Syst Appl.* 2015;42(3):990-1002.
- Xu S, Liu H, Song E. Marker-controlled watershed for lesion segmentation in mammograms. *J Digit Imaging.* 2011;24(5):754-763.
- Hsu W-Y. Improved watershed transform for tumor segmentation: application to mammogram image compression. *Expert Syst Appl.* 2012;39(4):3950-3955.
- Min H, Chandra SS, Dhungel N, Crozier S, Bradley AP. Multi-scale mass segmentation for mammograms via cascaded random forests, Paper presented at: IEEE 14th International Symposium on Biomedical imaging; 2017; IEEE.
- Dong M, Lu X, Ma Y, Guo Y, Ma Y, Wang K. An efficient approach for automated mass segmentation and classification in mammograms. *J Digit Imaging.* 2015;28(5):613-625.
- Rahmati P, Adler A, Hamarneh G. Mammography segmentation with maximum likelihood active contours. *Med Image Anal.* 2012;16(6):1167-1186.
- Xie W, Li Y, Ma Y. Breast mass classification in digital mammography based on extreme learning machine. *Neurocomputing.* 2016;173:930-941.
- Liu X, Zeng Z. A new automatic mass detection method for breast cancer with false positive reduction. *Neurocomputing.* 2015;152:388-402.
- Timp S, Karssemeijer N. A new 2D segmentation method based on dynamic programming applied to computer aided detection in mammography. *Med Phys.* 2004;31(5):958-971.
- Domínguez AR, Nandi AK. Improved dynamic-programming-based algorithms for segmentation of masses in mammograms. *Med Phys.* 2007;34(11):4256-4269.
- Song E, Jiang L, Jin R, Zhang L, Yuan Y, Li Q. Breast mass segmentation in mammography using plane fitting and dynamic programming. *Acad Radiol.* 2009;16(7):826-835.
- Song E, Xu S, Xu X, et al. Hybrid segmentation of mass in mammograms using template matching and dynamic programming. *Acad Radiol.* 2010;17(11):1414-1424.
- Abbas Q, Celebi ME, García IF. Breast mass segmentation using region-based and edge-based methods in a 4-stage multiscale system. *Biomed Signal Process Control.* 2013;8(2):204-214.
- Mencattini A, Salmeri M, Casti P, Raguso G, L'Abbate S, Chieppa L, et al., editors. Automatic breast masses boundary extraction in digital mammography using spatial fuzzy c-means clustering and active contour models. Paper presented at: IEEE International Workshop on Medical Measurements and Applications Proceedings (MeMeA); 2011, IEEE.
- Zhu W, Xiang X, Tran TD, Hager GD, Xie X. Adversarial Deep Structured Nets for Mass Segmentation from Mammograms. Paper presented at: IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018); 2018; IEEE.
- Dhungel N, Carneiro G, Bradley AP. Tree re-weighted belief propagation using deep learning potentials for mass segmentation from mammograms. Paper presented at IEEE 12th International Symposium on Biomedical Imaging (ISBI); 2015; IEEE, 2015.
- Shelhamer E, Long J, Darrell T. Fully convolutional networks for semantic segmentation. *IEEE Trans Pattern Anal Mach Intell.* 2017;39(4):640-651.
- Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation. Paper presented at: International Conference on Medical Image Computing and Computer-Assisted Intervention; 2015; Springer.
- Lian C, Zhang J, Liu M, et al. Multi-channel multi-scale fully convolutional network for 3D perivascular spaces segmentation in 7T MR images. *Med Image Anal.* 2018;46:106-117.
- Heath M, Bowyer K, Kopans D, Moore R, Kegelmeyer WP. The digital database for screening mammography. In Proceedings of the 5th International Workshop on Digital Mammography; 2000; Medical Physics Publishing, 2000.
- Moreira IC, Amaral I, Domingues I, Cardoso A, Cardoso MJ, Cardoso JS. Inbreast: toward a full-field digital mammographic database. *Acad Radiol.* 2012;19(2):236-248.
- Dalmış MU, Litjens G, Holland K, et al. Using deep learning to segment breast and fibroglandular tissue in MRI volumes. *Med Phys.* 2017;44(2):533-546.
- Singh VK, Rashwan HA, Romani S, et al. Breast Tmass segmentation and shape classification in mammograms using generative adversarial and convolutional neural network. *Expert Syst Appl.* 2020;139:112855.
- Cardoso JS, Domingues I, Oliveira HP. Closed shortest path in the original coordinates with an application to breast cancer. *Int J Pattern Recognit Artif Intell.* 2015;29(01):1555002.
- Kirby M, Heydarian M, Svenningsen S, et al. Hyperpolarized ³He magnetic resonance functional imaging semiautomated segmentation. *Acad Radiol.* 2012;19(2):141-152.

31. Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition; 2015.
32. Jiang M, Zhang S, Zheng Y, Metaxas DN. Mammographic mass segmentation with online learned shape and appearance priors. Paper presented at: International Conference on Medical Image Computing and Computer-Assisted Intervention; 2016; Springer.
33. Domínguez AR, Nandi AK. Toward breast cancer diagnosis based on automated segmentation of masses in mammograms. *Pattern Recognit.* 2009;42(6):1138-1148.
34. Kooi T, van Ginneken B, Karssemeijer N, den Heeten A. Discriminating solitary cysts from soft tissue lesions in mammography using a pretrained deep convolutional neural network. *Med Phys.* 2017;44(3):1017-1027.
35. Jia Y, Shelhamer E, Donahue J, Karayev S, Long J, Girshick R, et al. Caffe: Convolutional architecture for fast feature embedding. In Proceedings of the 22nd ACM International Conference on Multimedia; 2014; ACM.

How to cite this article: Xu S, Adeli E, Cheng J-Z, et al. Mammographic mass segmentation using multichannel and multiscale fully convolutional networks. *Int J Imaging Syst Technol.* 2020;1-13. <https://doi.org/10.1002/ima.22423>