Published in final edited form as:

IEEE Trans Biomed Eng. 2019 June ; 66(6): 1567–1579. doi:10.1109/TBME.2018.2875955.

Three-Dimensional Whole Breast Segmentation in Sagittal and Axial Breast MRI with Dense Depth Field Modeling and Localized Self-Adaptation for Chest-Wall Line Detection

Dong Wei, Susan Weinstein, Meng-Kang Hsieh, Lauren Pantalone, Despina Kontos*

Department of Radiology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA.

Abstract

Objective: Whole breast segmentation is an essential task in quantitative analysis of breast MRI for cancer risk assessment. It is challenging mainly because the chest-wall line (CWL) can be very difficult to locate due to its spatially varying appearance—caused by both nature and imaging artifacts—and neighboring distracting structures. This paper proposes an automatic three-dimensional (3-D) segmentation method, termed DeepSeA, of whole breast for breast MRI.

Methods: DeepSeA distinguishes itself from previous methods in three aspects. First, it reformulates the challenging problem of CWL localization as an equivalent problem that optimizes a smooth depth field and so fully utilizes the CWL's 3-D continuity. Second, it employs a localized self-adapting algorithm to adjust to the CWL's spatial variation. Third, it applies to breast MRI data in *both* sagittal and axial orientations equally well without training.

Results: A representative set of 99 breast MRI scans with varying imaging protocols are used for evaluation. Experimental results with expert-outlined reference standard show that DeepSeA can segment breasts accurately: the average Dice similarity coefficients, sensitivity, specificity, and CWL deviation error are 96.04%, 97.27%, 98.77%, and 1.63mm. In addition, the configuration of DeepSeA is generalized based on experimental findings, for application to broad prospective data.

Conclusion: A fully automatic method—DeepSeA—for whole breast segmentation in sagittal and axial breast MRI is reported.

Significance: DeepSeA can facilitate cancer risk assessment with breast MRI.

Keywords

Breast cancer; breast MRI; dense depth field; localized self-adaptation; three-dimensional (3-D) whole breast segmentation

I. Introduction

BREAST cancer is the most commonly diagnosed cancer and the second leading cause of cancer death in women in the United States [1]. Regular screening among women of suitable

*Corresponding author: Despina.Kontos@uphs.upenn.edu.

ages has proven effective in reducing the mortality [2]. Despite being established as the current imaging standard for such screening, mammography suffers low sensitivity in dense breasts, as the excessive fibroglandular tissue (FGT) can mask lesions in the two-dimensional (2-D) projection of a mammogram. Breast magnetic resonance imaging (MRI) greatly alleviates this problem via three-dimensional (3-D) visualization of the breast. MRI has achieved higher sensitivity than mammography for women at high risk for breast cancer [3], and is recommended as a supplement to screening mammography for this cohort by the American Cancer Society [4]. MRI screening also improves early diagnosis of prognostically pertinent breast cancer for average-risk women [5].

In addition to lesion detection, breast MRI is also highly valuable for cancer risk assessment in cancer-unaffected breasts. Knowledge of individual risk is critical for personalized care planning, e.g., to initiate preventive strategies [6]. MRI allows measurement of the volumetric percentage of FGT in the breast (FGT%); this measure has been shown to be a risk factor for breast cancer [7], especially in the high-risk cohort [8]. MRI also allows measurement of background parenchymal enhancement (BPE) in dynamic contrast-enhanced breast MRI; this is the contrast uptake in normal breast tissue (other than tumors), and the level of this uptake has lately been suggested as another strong index for breast cancer risk assessment [9].

Despite the desirable clinical value, manual quantitative analysis of breast MRI is labor-intensive and time-consuming due to the task complexity and large amount of data; Wu *et al.* reported that ~90 minutes were needed for manual quantification of FGT% in a bilateral MRI of 56 slices [10], [11]. As well, the subjectivity of human readers brings inter- and intraobserver variation into the analysis. Therefore, development of automated methods for breast MRI quantification has recently drawn great research interest. To achieve the quantification, two (groups of) tasks need to be accomplished: whole breast segmentation, and regions of interests segmentation—including FGT [11], [12] and BPE [13], [14]—within the breast. In this paper we focus on the task of whole breast segmentation, often a prerequisite to the other.

Automatic segmentation of whole breast in clinical MR images is challenging primarily because it can be very difficult to localize the chest-wall line (CWL). The appearance of the CWL varies through the image, with both nature—the varying properties of the CWL, and complex neighboring structures—and imaging artifacts—most prominently, the bias-field artifact—contributing to this difficulty. The bias field artifact refers to image intensity non-uniformity caused by magnetic field variations rather than tissue properties (e.g., Fig. 6d). Various approaches have been tried; Gubern-Mérida *et al.* compared three methods quantitatively [15]. Most previous methods have been developed for *axial* breast MRI only, despite tremendous existence of *sagittal* data. Axial-specific methods may not be directly applicable to sagittal data due to the different protocols used. Although sagittal breast MRI can be reformatted into axial slices, the reformatted slices still may not provide the same information as data originally acquired in the axial orientation. For example, [12] used the sternum, and [16] used the thoracic spine as landmarks to initiate the segmentation process; these landmarks are lost or suboptimally captured in sagittal data when two breasts are imaged separately (Fig. 1) or only one breast is imaged.

In contrast, the method presented in this paper applies to both axial and sagittal breast MRI. So far as we know this is the only breast segmentation method that has been thoroughly validated with both sagittal and axial data in their original imaging orientations rather than in reconstructed views. We were motivated to develop an approach that is applicable to both views, and able to develop such an approach, as our institution captured sagittal views through 2015, providing a large amount of data in both views and necessitating the need of a tool to analyze them. Sagittal orientation was the main stream of breast MRI; the shift to axial was fairly recent at our institution and is still happening at many others. The method here can facilitate analysis of both the more recent axial data and the substantial amount of archived sagittal data. Among the few methods developed for sagittal data, the one in [10] shows good experimental results; its key components include two edge-enhancing nonlinear filters and a CWL candidate evaluation/selection scheme. However, this method largely depends on a 2-D slice-wise edge-linking algorithm and does not fully utilize the CWL's property of being a smooth 3-D surface, and so can result in incorrect segmentation in the presence of strong edges other than the CWL [17].

Another approach segments breasts based on images from MRI sequences that are not commonly acquired in standard clinical settings [18], [19], e.g., the breast fat and water images of the Dixon sequences. Using these sequences, however, limits wider application as it is unclear how these methods can generalize to routinely acquired breast MR sequences.

Atlas- [12], [18] and machine learning-based [19] methods have also been applied to whole breast segmentation; both of these approaches utilize knowledge embedded in pre-labeled data. A recent trend in machine learning is the use of deep convolutional neural networks (CNNs), which have gained attention due to their success in various applications such as image classification, object recognition, and medical image analysis. Dalmi *et al.* were the first to publish results on deep learning for breast MRI segmentation, and showed improved results over an atlas-based method [20]. However, it can be labor-intensive and time-consuming to manually label adequate data for atlas construction/classifier training.

In this paper, we propose a non-machine learning-based, automated 3-D method for whole breast segmentation in *both* sagittal and axial breast MRI. Our method distinguishes itself from previous methods in: 1) reformulating the challenging problem of localizing the CWL as an equivalent dual problem of optimizing a smooth depth field, thus exploiting 3-D continuity; 2) employing a localized self-adapting approach towards valid local neighbors to adjust to local patterns; and 3) applying to both sagittal and axial data equally well without the need for training. We refer to this method as “DeepSeA” for its **DE**nse **dEP**th field modeling and **SE**lf-Adaptation features. This paper is a substantial and comprehensive extension of our previous preliminary work [17] from three aspects:

- It extends the key concepts of dense depth field modeling and localized self-adaptation to support axial breast MRI, in addition to the sagittal support in [17].
- It conducts thorough experiments with more data for validation with both sagittal and axial data, as well as method generalization.

- It supplements [17] by 1) complementing the localized self-adaptation scheme with newly proposed global self-adaptation, 2) implementing improvements such as randomized initialization, and 3) providing more details.

The rest of this paper is organized as follows. Section II describes DeepSeA. Section III presents the experiments, results, and discussion, followed by conclusion in Section IV.

II. Method

A. Materials

1) Study Population and MRI Acquisition: A representative set of 99 3-D bilateral breast MRI scans were used in this work. The subjects were randomly selected from cancer-unaffected women in the high-risk screening population at the Hospital of the University of Pennsylvania. This set of data were selected so that they: 1) cover a long time span (2003–2017) to include both recent and archive data, 2) include breasts of varying sizes and densities, and 3) contain data scanned in both sagittal and axial orientations. Based on the year of study and imaging orientation, we divide the data into three subsets: archive data in sagittal orientation (the Sag^I subset; this is the same set of data used in [10] with one scan excluded for its different size from all others'), recent data in sagittal orientation (the Sag^{II} subset), and recent data in axial orientation (the Ax subset). A scan was in either sagittal or axial orientation. Table I shows the subjects' demographic information. As we can see, the body mass index (BMI) spanned quite a wide range, so did the breast volume. According to the four-point American College of Radiology (ACR) BI-RADS breast density categorization (almost entirely fat, scattered, heterogeneous, and extreme FGT) [21], the data set had a balanced distribution of breast density.

As most related works ([10], [12], [16], to name a few), this work uses T1-weighted nonfat-suppressed images since fat is brighter in such images making segmentation of the breast easier. Subjects were imaged prone in a 1.5T scanner using a dedicated surface breast coil array. The MRI acquisition parameters are presented in Table II. It can be seen that along with the replacement of imaging systems, the imaging protocol also evolved over the time span of our data, resulting in variations in not only orientation but also acquisition parameters.

We assign internal identification numbers to the data in each subset; also, for clarity, we will indicate slice numbers where applicable in the rest of this paper. For instance, the ninth slice of the third case in Sag^{II} is referred to as Slc₉ of Sag₃^{II}.

2) Reference Standard Generation: The ACR BI-RADS density for each breast MRI was visually assessed by a board-certified breast-imaging radiologist (S.W., 20 years of experience). In addition, breasts in each MRI were manually segmented by the same radiologist.

Breasts in Sag^I were segmented using the ITK-SNAP [22] software slice by slice; we refer readers to [10] for more details. Breasts in Sag^{II} and Ax were segmented using an inhouse developed interactive software. The software presented one slice at a time, and modeled

breast contours with smooth splines controlled by series of points; the radiologist defined breast contours by placing and replacing these control points. The software interpolated breast contours for slices inbetween manually segmented slices. The interpolation quality was up to: 1) the extent of breast change across slices and the sparsity/density of manually segmented slices. In case of unsatisfactory interpolation, the radiologist segmented more slices where needed and the interpolation was updated. The software generated a 3-D breast mask at the end.

The radiologist used her discretion to ensure accuracy of the manual segmentation. This set of manual segmentation is considered as the golden reference standard in our study.

B. Method Overview

Fig. 2 charts the DeepSeA work flow. Based on functionalities, processing steps in the flowchart can be grouped into three major tasks: air-breast boundary (ABB) detection (Fig. 2b), breast bounding plane delineation (Figs. 2c, 2e, and 2f), and CWL localization (Fig. 2d). Of the three, CWL localization is most challenging and it is in this task that the major contributions of this paper lie. We will describe only the outcomes of the other two tasks and their relation to the entire work flow, but refer readers to the online supplement for concrete approaches.

C. ABB and Breast Bounding Plane Delineation

DeepSeA starts by detecting the ABB, which defines the anterior boundary of the breasts. This task is relatively straightforward thanks to the high contrast between fat tissue and air in nonfat-suppressed images, and can be achieved with a combination of k-means clustering, morphological operations, and connected component analysis. A general body mask is obtained from the ABB detection (Fig. 2b). Usually, this mask encloses most of the breasts and some adjacent thoracic regions such as pectoral muscles; occasionally it can also miss a posterior portion of the breasts, especially in the presence of severe bias-field artifact. The general body mask will be used to define the region of interest (ROI) for CWL localization (Section II-D.2).

In addition to the anterior and posterior boundaries defined by the ABB and CWL, the segmented breast should also be bounded in the superior/inferior and medial/lateral directions to exclude surrounding subcutaneous fat that does not belong to the breast. Specifically, the following additional bounding planes are determined by DeepSeA: the lateral, to exclude fat near the armpits (Fig. 2c); the inferior, to exclude fat at the upper abdominal region (Fig. 2e); and the medial, to exclude fat between the two breasts (Fig. 2f). Currently DeepSeA does not define a superior boundary as there is usually not as much fat above the breast as in the abdominal region.

The bounding planes are derived from the projected 2-D general body masks, e.g., the inferior bounding plane is derived from the 2-D mask projected into the sagittal plane (Fig. 2e). As shown in the flowchart, the bounding planes are determined at different stages in the segmentation process: the lateral are determined prior to CWL localization (Fig. 2c), and the inferior and medial determined after the CWL localization (Figs. 2e and 2f). This is because, as part of detecting a smooth CWL surface, CWLs between the two breasts and beyond the

inferior boundary can contribute to our 3-D localization, so these areas are included in the ROI for detecting the CWL; but the very lateral portion of the breast MRI volume often goes beyond the pectoral muscles and can distract CWL localization, so this area is excluded from the ROI prior to CWL detection.

D. CWL Localization

The CWL localization includes three procedures: CWL initialization, self-adaptation, and depth field optimization. In this section we will first introduce our novel problem formulation of this challenging task, then define the ROI for CWL localization, and finally describe the procedures in detail.

1) Problem Formulation: Since DeepSeA will be applied to both sagittal and axial breast MRI, a coordinate system independent of imaging orientation is needed. For this purpose, an intuitive reference coordinate system (RCS) is defined with respect to the patient body: the x -axis points from right to left, y -axis anterior to posterior, and z -axis feet to head. Figs. 3a and 3b illustrate how the sagittal and axial breast MR images used in this work align with the RCS. Let $\Omega \subset \mathbb{R}^3$, then a 3-D breast MR image can be written as a real-valued intensity function defined in the RCS: $I(x, y, z): \Omega \rightarrow \mathbb{R}$.

Unlike existing methods that treat CWL localization as a byproduct of labeling different structures, as in atlas-based methods [12], or that directly link edges returned by edge detectors [10], we reformulate this problem into an equivalent dual problem of optimizing a smooth depth field. Looking at Fig. 2d, it can be seen that the CWL can be represented by its distances from the image boundary along the y -axis, or, in other words, its *depths* along the y -axis. Furthermore, all the CWLs in all slices of a breast MR volume $I(x, y, z)$ should form a smooth surface because the anatomical structure of the pectoral muscle dictates a smooth boundary. Therefore, the CWL surface of a breast can be expressed with a dense, smooth (slowly varying), depth field: $y = D(x, z): \Phi \rightarrow \mathbb{R}$, where $\Phi \subset \mathbb{R}^2$. The depth field D specifies the 3-D coordinates of all voxels on the CWL surface with $(x, D(x, z), z)$. We argue that the dense depth field model is a better option than many parametric models because individuals' chests vary widely, and it would be difficult to find a parametric model that fit the variety of real-world data. Depth fields have been popular in computer vision tasks (e.g., scene reconstruction), but, to the best of our knowledge, this is the first application of the idea to localizing the CWL in 3-D.

In addition to the depth field D , another representation of the CWL is also defined—the CWL mask $M_{\text{CWL}}(x, y, z): \Omega \rightarrow \{0 \text{ or } 1\}$. M_{CWL} is a binary mask of the same size as the breast image $I(x, y, z)$, and marks CWL locations with white voxels (1s) on black background (0s). D and M_{CWL} express the same information— $(x, D(x, z), z)$ describe the CWL locations in M_{CWL} , but the M_{CWL} allows us to perform certain operations conveniently. The converting function from D (in $\mathbb{R}^2$) to M_{CWL} (in $\mathbb{R}^3$) is denoted by $M_{\text{CWL}} = \mathcal{f}[D]$. To avoid potential confusion, we will refer to elements in the 2-D domain Φ (such as D) as *pixels*, and those in the 3-D domain Ω (I and M_{CWL}) as *voxels*.

To impose a smoothness constraint on $D(x, z)$, the quadratic functional

$$E_{\text{smth}}[D] = \frac{1}{2} \int_{\Phi} |\nabla D|^2 dx dz \quad (1)$$

is used as a measure of the total *un*-smoothness of D , where ∇D denotes gradients of D . This form of E_{smth} is preferred to the popular total variation [23] ($\int_{\Phi} |\nabla D| dx dz$) because it does not allow discontinuities across lines and thus no edge should be present in the resulting D . When combined with an energy functional $E_{\text{patn}}[D]$ which favors depth values at locations with desired CWL intensity patterns (described in Section II-D.4), the optimal D can be solved for by minimizing

$$E[D] = E_{\text{patn}}[D] + \lambda E_{\text{smth}}[D], \quad (2)$$

where λ is a weighting factor.

2) CWL ROI: A volumetric ROI for CWL search is defined using the general body mask obtained when detecting the ABB (Fig. 2b). For sagittal breast MRI, this mask is projected into the yz plane producing a *silhouette* (Fig. 3a). Three points are identified on the silhouette: the frontmost point as the nipple y_{nip} ; the frontmost point of all the points on either the superior or inferior side of the silhouette bounding box, denoted by y_{base} ; and the rearmost point y_{rear} . Then a front bounding plane is defined as the xz plane passing through the midpoint of y_{nip} and y_{base} , and a back bounding plane as the xz plane passing through the midpoint of y_{rear} and the image boundary. Given these, the CWL ROI is defined as all body parts posterior to the ABB confined within these two bounding planes (Fig. 4b). This ROI reduces the volume that must be examined to detect the CWL for computational efficiency and reduced interference, but still allows for a wide variation in depth of the CWL. The CWL ROI for axial data is defined almost identically, with the only notable difference being that symmetry between the breasts is utilized. Concretely, y_{base} is identified on the front surface of the silhouette (projected into the xy plane this time) as the point that splits the general body mask closest to equally (Fig. 3b).

3) CWL Initialization: Using Fig. 4a as an example, it can be observed that most CWL locations follow a general pattern across the CWL: a transition from relatively bright fat to dark chest, with the darkest CWL in the center. Thus, we search for such a pattern in the CWL ROI. For this purpose, a generic one-dimensional (1-D) template T_{gen} is devised, with its anterior half set to a constant high value, its posterior half a constant medium value, and the center a low value (corresponding to the CWL). To determine these high, medium, and low values we cluster all voxels in I into three clusters, and use the cluster centers for the values in T_{gen} . To account for the partial volume effect, a single-pixel transition value is set on both sides of T_{gen} 's center using the averages of the high and low values, and of the low and medium values. T_{gen} 's size w can be expressed as a combination of two halves of size v and the center: $w = 2v + 1$, dictating it must be odd-numbered. T_{gen} 's mean is shifted to zero by $T_{\text{gen}} - \bar{T}_{\text{gen}}$. As a 1-D filter is vulnerable to noise, T_{gen} is made into a square patch by

stacking repeated 1-D templates along the orthogonal direction in the image plane (Fig. 4c top). (Similarly, a 3-D template can be constructed by stacking repeated 2-D templates along the direction orthogonal to the image plane. We will compare the performance of 2-D versus 3-D templates in Section III-A.4; here we consider the 2-D generic template as a special case of its 3-D counterpart—whose out-of-image-plane values are zero—so that we can use a single formula to describe the generic filtering process for both 2-D and 3-D templates.) Finally, as a first pass at CWL detection, T_{gen} is used to filter through the CWL ROI (regions outside the ROI are simply set to 0), producing a volumetric response map $M_{\text{resp}}: \Omega \rightarrow \mathbb{R}$ corresponding to the probability of being CWL:

$$M_{\text{resp}}(x, y, z) = T_{\text{gen}} \otimes I = \sum_{l=-v}^v \sum_{m=-v}^v \sum_{n=-v}^v T_{\text{gen}}(l, m, n) I(x+l, y+m, z+n), \quad (3)$$

where $(x, y, z) \in \text{CWL ROI}$, and T_{gen} 's origin is its geometrical center. In addition, as T_{gen} favors dark voxels at the center, we further enhance this preference by weighing M_{resp} :

$$M_{\text{resp}}^{\text{wtd}}(x, y, z) = M_{\text{resp}}(x, y, z) \left[\frac{\max(I) - I(x, y, z)}{\max(I) - \min(I)} \right]. \quad (4)$$

The CWL ROI in $M_{\text{resp}}^{\text{wtd}}$ is then linearly scaled to $[0, 1]$, yielding $M_{\text{resp}}^{\text{scl}}$. An example slice of $M_{\text{resp}}^{\text{scl}}$ is shown in Fig. 4d.

Next, the CWL is initialized by finding the y value that maximizes the filter response for each combination of x and z . $D^0(x, z) = \arg \max_y M_{\text{resp}}^{\text{scl}}(x, y, z)$, and also represented with a CWL mask $M_{\text{CWL}}^0 = f[D^0]$ for subsequent operations. False positives—voxels incorrectly identified as being part of the chest wall—will usually be present in the initialization. These may be caused by: 1) distracting structures of similar intensity patterns, such as chest-lung or fat-dense tissue interfaces; 2) spatial variation of the CWL's intensity patterns; and 3) bias-field artifacts. These false positives need to be removed before the next step. Assuming most of the initial CWL to be correct, a 3-D connected component analysis is performed on M_{CWL}^0 , and only the largest component, presumably the set of all true positives, is retained. All other components are considered false positives (Fig. 4e) and removed by setting them to 0 in M_{CWL}^0 . We denote the true positive set by $S_{\text{TP}}^0 = \{(x, z) \mid M_{\text{CWL}}^0(x, D^0(x, z), z) = 1\}$.

4) Localized and Global Self-Adaptation: Although intensity patterns change spatially on a CWL surface, the change usually happens gradually. Hence it is reasonable to learn local templates for regions where we have not identified the CWL—regions where the

initial CWL voxels have been discarded as false positives—from intensity patterns extracted at the nearest available true positive CWL voxels. Locally adapted templates can also be used to refine the locations of true positive CWL voxels—those which have already been detected via the primitive generic template. To do this, for each pixel (x_0, z_0) in D^0 , we search for a nearest “neighbor” in each of the four upright directions (up: $+z$, down: $-z$, left: $+x$, and right: $-x$), denoted by $\{(x_n, z_n) | n = 1, 2, 3, 4\}$, that belongs in the true positive set: $(x_n, z_n) \in S_{TP}^0$. Then a patch P_n centered at $(x_n, D^0(x_n, z_n), z_n)$ is extracted from I , and the locally adapted template $T_{\text{adpt}}^{\text{loc}}$ is defined as:

$$T_{\text{adpt}}^{\text{loc}}(x_0, z_0) = \sum_{n=0}^4 \omega_n^{\text{loc}} P_n / \sum_{n=0}^4 \omega_n^{\text{loc}}, \quad (5)$$

where

$$\omega_n^{\text{loc}} = \frac{M_{\text{resp}}(x_n, D(x_n, z_n), z_n)}{1 + d((x_n, z_n), (x_0, z_0))}, \quad (6)$$

and $d((x_n, z_n), (x_0, z_0))$ is the Manhattan distance—selected for its low computational cost—between (x_n, z_n) and (x_0, z_0) . In other words, the adaptive template is a weighted average of the local intensity patterns extracted at the nearest available true positive CWL voxels (including $(x_0, D^0(x_0, z_0), z_0)$, if $(x_0, z_0) \in S_{TP}^0$). The weights are in proportion to the strengths of the filter responses and in inverse proportion to the distances between the pixel (x_0, z_0) and its neighbors. Note that there may be no true positive neighbor available for a given pixel in one or more directions, or the pixel itself is not a true positive; in this case the corresponding weight ω_n^{loc} is set to 0, meaning no contribution from that direction or itself. Size of $T_{\text{adpt}}^{\text{loc}}$ is increased to $4v + 1$ as we are more confident of its fidelity to the real data. An example $T_{\text{adpt}}^{\text{loc}}$ is shown in Fig. 4c middle.

In the extreme case when $\omega_n^{\text{loc}} = 0$ for all $n = 0, \dots, 4$ —there is no true positive neighbor in any direction and the pixel itself is not a true positive— $T_{\text{adpt}}^{\text{loc}}$ cannot be defined with (5).

This most frequently occurs in images of large fatty breasts affected by severe bias-field artifact, in which the CWLs are difficult to discern (e.g., Fig. 6d). In such cases a *globally* adapted template is defined as a backup:

$$T_{\text{adpt}}^{\text{glob}} = \sum_{n \in S_{TP}^0} \omega_n^{\text{glob}} P_n / \sum_{n \in S_{TP}^0} \omega_n^{\text{glob}}, \quad (7)$$

where $n \in S_{TP}^0$ denotes all true positive CWL voxels identified in M_{CWL}^0 , and

$$\omega^{\text{glob}}(x, z) = M_{\text{resp}}(x, D(x, z), z). \quad (8)$$

In other words, the globally adapted template is a weighted average of intensity patterns extracted at *all* true positive CWL voxels of M_{CWL}^0 , and the weights are in proportion to the strengths of the filter responses. Fig. 4c bottom shows an example $T_{\text{adpt}}^{\text{glob}}$. Then the unified self-adapted template T_{adpt} is defined as:

$$T_{\text{adpt}}(x, z) = \begin{cases} T_{\text{adpt}}^{\text{loc}}(x, z), & \text{if } \omega_n^{\text{loc}} \neq 0 \text{ for any } n = 0, \dots, 4, \\ T_{\text{adpt}}^{\text{glob}}, & \text{otherwise.} \end{cases} \quad (9)$$

Now for each location (x_0, z_0) within the depth map D , we have a self-adapted template $T_{\text{adpt}}(x_0, z_0)$ which is the desired CWL intensity pattern at (x_0, y, z_0) . The next step is to examine how different each local patch $P(x_0, y, z_0)$ is from $T_{\text{adpt}}(x_0, z_0)$ by traversing all possible y values within the CWL ROI. Meanwhile, if $(x_0, z_0) \in S_{\text{TP}}^0$, the CWL ROI at (x_0, y, z_0) is reduced to $\{(x_0, y, z_0) | D^0(x_0, z_0) - 2v \leq y \leq D^0(x_0, z_0) + 2v\}$ (see Fig. 4f; the narrow regions of the bright strap are those with reduced ROI). This is to ensure only fine adjustment of the CWL in true-positive areas as well as for computational efficiency. Since T_{adpt} is adapted to the MR image, a direct difference metric other than correlation can be used to enforce higher fidelity. This work employs the ℓ^2 norm (denoted by $\|\cdot\|$) of the raw intensity difference between the image patch P and template T_{adpt} . Then we can generate a difference map $M_{\text{diff}}: \Omega \rightarrow \mathbb{R}$ by computing the difference between the local image patch P extracted at each possible (x, y, z) location within the reduced CWL ROI and the corresponding $T_{\text{adpt}}(x, z)$:

$$M_{\text{diff}}(x, y, z) = \begin{cases} 1, & \text{if } (x, y, z) \notin \text{CWL ROI}, \\ \|P(x, y, z) - T_{\text{adpt}}(x, z)\|, & \text{otherwise.} \end{cases} \quad (10)$$

Similar to $M_{\text{resp}}^{\text{wtd}}$, values of M_{diff} within the CWL ROI are scaled to the unit range $[0, 1]$, yielding $M_{\text{diff}}^{\text{scl}}$. Fig. 4f shows an example slice of $M_{\text{diff}}^{\text{scl}}$. Then E_{patn} is defined as the summation of all differences at the current CWL locations (represented by D):

$$E_{\text{patn}}[D] = \sum_{(x, z) \in \Phi} M_{\text{diff}}^{\text{scl}}(x, D(x, z), z). \quad (11)$$

5) Depth Field Optimization: To minimize E respect to D using a gradient descent method, we deduce the associated Euler-Lagrange equation for D .

$$\frac{\partial D}{\partial t} = -\frac{\partial E}{\partial D} = -\frac{\partial M_{\text{diff}}^{\text{scl}}}{\partial D} + \lambda \nabla^2 D, \quad (12)$$

where $t = 0$ is an artificial time parameterizing the descent direction and $\nabla^2 D$ the Laplacian of D . Using a finite difference scheme, D can be updated iteratively by the evolution equation below to find a local minimum of E :

$$D^{i+1} = D^i + \mu \cdot (-\partial M_{\text{diff}}^{\text{scl}} / \partial D^i + \lambda \nabla^2 D^i), \quad (13)$$

where $i = 1$ is a positive integer indicating the iteration number, D^i is the depth field after the $(i-1)^{\text{th}}$ iteration, and μ is the step size. For numerical implementation, $\nabla^2 D$ can be calculated by a Laplacian operator. For $\partial M_{\text{diff}}^{\text{scl}} / \partial D$, a lookup table can be built: denoting the y component of $M_{\text{diff}}^{\text{scl}}$'s 3-D numerical gradients by G_y , then the value of $\partial M_{\text{diff}}^{\text{scl}} / \partial D$ at (x, z) can be looked up by $G_y(x, D(x, z), z)$.

D^1 is initialized much as the CWL was initialized in Section II-D.3. First, a temporary depth field D^{tmp} is extracted by $D^{\text{tmp}}(x, z) = \text{argmin}_y M_{\text{diff}}^{\text{scl}}(x, y, z)$. Then, D^{tmp} is converted into a temporary CWL mask: $M_{\text{CWL}}^{\text{tmp}} = f[D^{\text{tmp}}]$. The true positive set in $M_{\text{CWL}}^{\text{tmp}}$ is again identified as the largest connected component, with all others deemed false positives and set to 0. Denoting the true positive set by $S_{\text{TP}}^{\text{tmp}} = \{(x, z) \mid M_{\text{CWL}}^{\text{tmp}}(x, D^{\text{tmp}}(x, z), z) = 1\}$, then

$$D^1(x, z) = \begin{cases} D^{\text{tmp}}(x, z), & \text{if } (x, z) \in S_{\text{TP}}^{\text{tmp}}, \\ D^{\text{tmp}}(x_{\text{rdm}}, z_{\text{rdm}}), & \text{otherwise,} \end{cases} \quad (14)$$

where $(x_{\text{rdm}}, z_{\text{rdm}})$ is a point randomly selected from $S_{\text{TP}}^{\text{tmp}}$. In other words, for a pixel in D^{tmp} that does not correspond to a true positive CWL voxel in $M_{\text{CWL}}^{\text{tmp}}$, its initial depth value is initialized with that of a randomly selected pixel that does.

The step size, μ , is initially set to 1 and halved when E_{smth} does not decrease for three consecutive iterations; we focus on E_{smth} here because E_{patn} was initialized close to a local minimum when the iteration started. Iteration stops when the absolute update to D is smaller than 0.001 for every pixel. Fig. 5a illustrates the iterative process of minimizing $E[D]$, and Fig. 5b shows a depth field after optimization.

E. Three-Dimensional Breast Mask Generation

The closed 3-D breast mask is the volume enclosed by the ABB, CWL, medial, lateral, and inferior bounding planes of the breast (see Fig. 2g).

F. Implementation

DeepSeA is implemented in the MATLAB R2014b environment, on a PC with an Intel Core i5-2500 processor and 16 GB RAM, and running 64-bit Microsoft Windows 7 Service Pack 1 OS.¹ We did not optimize for speed in this work except for using the MATLAB built-in *parallel for loop* for the localized adaptation, which invokes four workers on the specific processor.

III. Experiments, Results and Discussion

A. Experimental Settings

1) General Settings: Of the three subsets of data, Sag^{II} and Ax are used as exploration sets for method generalization—selecting optimal values for parameters and template dimensionality—besides being used for validation; Sag^I is solely used for validation of the generalized method, serving as a separate, stand-alone test set.

To allow consistent calculation of the smoothness constraint on the depth field, breast MRI volumes are interpolated to be isotropic using cubic interpolation for the CWL localization step in DeepSeA. Data in Sag^{II} are interpolated to 0.5mm isotropic and those in Ax to 1mm isotropic, which are close or equal to their respective finest physical resolutions. Data in Sag^{II} are also interpolated to 1mm isotropic to allow investigation of the consistency of the method when applied to sagittal and axial data of the same resolution.

Both breasts of a subject are segmented separately for data in Sag^I and Sag^{II}, jointly for those in Ax.

2) Evaluation Metrics: The primary performance measure is the association between which voxels are designated as breast by DeepSeA, versus those designated as breast in the reference standard, as measured by three metrics—Dice similarity coefficient (DSC), sensitivity, and specificity. However, this measure is breast volume-based and insensitive to CWL locations, especially for large breasts. To specifically evaluate the automatically detected CWL, we calculate the CWL average deviation distance (CDD) [10] from the reference standard. All metrics are reported case-wise. The *t*-test (paired-/two-sample, where appropriate) is used to test for statistically significant difference between two groups of measurements. However, if the Lilliefors test [24] rejects the null hypothesis for any of the two groups, that the measurement comes from a distribution in the normal family, nonparametric test (Wilcoxon signed rank/Mann-Whitney U, where appropriate) is used, instead. 5% significance level is used in this study, with the Bonferroni correction where appropriate.

3) Parameter Optimization: Only two parameters need to be set for DeepSeA: ν and λ .

The parameter ν determines the sizes of the templates used in CWL localization. Too small a template will be indistinct and subject to noise, too large a template will be distracted by interference. In addition, larger templates slow computation.

¹Source code available at <https://github.com/CBICA/DeepSeA>.

The parameter λ weights the relative importance of the two energy terms, E_{smth} against E_{patn} , and a proper value ensures that the resulting CWL is neither rigid (dominated by E_{smth}) nor irregular (dominated by E_{patn}). However, a value that works well on one data set may not generalize well to another, particularly because E_{smth} and E_{patn} do not scale proportionally when the MRI changes its physical resolution, i.e., voxel size. From a different perspective, however, voxel size may be the only factor that can be controlled of the many factors that affect the ratio $E_{\text{smth}}/E_{\text{patn}}$; via interpolation, it can be normalized across patients and data sets. Therefore, in order to apply DeepSeA to potentially diverse breast MRI data, this work explores λ values at 0.5mm and 1mm voxel sizes, two representative physical resolutions of modern breast MRI. In addition, to constrain the grid search, optimization only evaluates λ at powers of 10, i.e., $\lambda = 10^q$.

Coarse-to-fine grid search is employed to optimize v and q (Table III)—the delegates that determine the sizes of the templates and the weight λ . Since both parameters affect only the CWL localization procedures, optimization does not require evaluating breast volume. Performance is evaluated in terms of CDD after $E[D]$ has been optimized but before determining the medial and inferior bounding planes. To avoid overfitting, 5-fold cross validation is employed. Each fold includes four cases, one from each of the four ACR BI-RADS density categories. For each fold, parameters are optimized (via the grid search) based on the data from the other four folds, and performance is evaluated on the held-out fold.

4) 3-D versus 2-D Templates: DeepSeA is a 3-D method in that it segments a breast volume all at once rather than slice by slice. The update to the depth field D updates the 3-D CWL surface in each iteration. The localized self-adaptation is 3-D, too, as it incorporates information from neighboring slices. Despite that, both the generic and self-adapted templates were 2-D in our preliminary work [17]. In the current work, we compare the performance of 2-D templates (in the image planes) to 3-D templates on Sag^{II} and Ax. The 3-D templates, are generated by stacking repeated 2-D templates in the third dimension for the generic templates, and by extracting 3-D cubes instead of 2-D patches for self-adapted templates. Again, because this design option only affects CWL localization, breast volumes are not compared in this experiment. The comparison is based on the cross validation results (for CDD) obtained from the grid search described above.

5) Comparison with Alternatives: We compare the performance of DeepSeA on Sag^I with that of the method presented in [10]. We will refer to this method as the “edge-linking” method for its major steps in edge enhancing, evaluating, selecting, and linking. This method was exclusively developed for sagittal breast MRI, thus inapplicable to the data in Ax.

We also test whether the self-adaptation scheme improves on the performance provided by solely the 3-D continuity constraint on D . To do this, after initialization and false-positive removal (see Section II-D.3), we fit a nonparametric smooth surface to the true positive set S_{TP}^0 using natural neighbor interpolation [25]:

$$D^{\text{fit}}(x, z) = \sum_{(x_j, z_j) \in S_{\text{TP}}^0} \omega_j D^0(x_j, z_j), \quad (15)$$

where $(x_j, z_j) \in S_{\text{TP}}^0$ denotes all pixels in S_{TP}^0 , $D^{\text{fit}}(x, z)$ is the interpolated depth value at a query pixel (x, z) , $D^0(x_j, z_j)$ the initialized depth value of a true positive pixel in S_{TP}^0 , and ω_j a weight determined by how much area the query point (x, z) “steals” from the known data point (x_j, z_j) (please refer to [25] for details). The resulting D^{fit} represents a smooth surface and provides smooth CWL estimates in areas of false positives, without employing the self-adaptation portion of DeepSeA. We compare performance of DeepSeA versus this surface-fitting method on Sag^{II} and Ax.

B. Results

1) Optimal Configurations and Program Efficiency: Cross validation results are tabulated in Table IV. The optimal values for (v, q) are quite consistent across imaging orientations and voxel sizes. The optimal pairs found in the coarse grid are consistently $(2, -1)$ in all settings and folds. Over the fine grid, the optimal values obtained in different folds of a specific setting (i.e., a unique combination of imaging orientation, voxel size, and template dimensionality) are relatively consistent: v is still a single value for all five folds in almost all settings, and the q 's are within a narrow range. Results are also very similar for the coarse and fine grids for all settings. As to the difference between 2-D and 3-D templates, we see similar results across both types of templates with axial data being the exception and showing improvement from 2-D to 3-D templates. We also see that the resulting optimal parameters are more consistent when 3-D templates are used.

Based on the cross validation results, $v = 2\text{mm}$, $q = -1$, and 2-D templates will be used for Sag^{II} (voxel size = 0.5mm, corresponding typical volume size: 512×512×384), and $v = 2\text{mm}$, $q = -1.25$, and 3-D templates for Ax (voxel size = 1mm, corresponding typical volume size: 320×320×224) for performance reporting in the next section. The average running time to segment a bilateral sagittal or axial breast MRI volume using the above settings is 4.76 minutes and 51 seconds, respectively. The time for an axial MRI is shorter because the volume size is exponentially smaller for a larger voxel size. In contrast, studies reported that manual segmentation of a bilateral sagittal breast MRI scan of size 256×256×56 required ~35 minutes [10], and that of a bilateral axial scan of size 448×448×144 required ~53 minutes [26].

As to Sag^I, $v = 2\text{mm}$, $q = -1$ and 3-D templates will be used along with 0.5mm voxel size. Such a configuration is determined solely with the cross validation results on Sag^{II} and Ax, thus the performance of DeepSeA on Sag^I in this case can provide a fair evaluation of the method's generalizability.

2) Performance Reporting: Qualitatively, we observe that segmentation results produced by DeepSeA are consistently accurate for nearly all slices. Sample breast segmentations are shown in Fig. 6; also shown is the manual segmentation by the

radiologist. As we can see, for most parts the automatic and manual contours overlap with or are close to each other, except for minor discrepancies near the superior/inferior (for sagittal data) or medial/lateral (for axial data) bounding planes.

Quantitative evaluation results are tabulated in Table V. The overall breast segmentation accuracy of DeepSeA measured against the expert-produced reference standard on the total 99 scans appears to be high, with the average DSC, sensitivity, and specificity above 90%, and the average CDD at 1.63mm. Comparing DeepSeA's accuracy on the Sag^I and Sag^{II} subsets, no statistically significant difference is found for DSC, sensitivity, or CDD ($\alpha = 0.05$); although statistically significant differences are found for specificity, the difference is insubstantial (98% versus 99%) and the results are good for both data. The comparison between DeepSeA's accuracy on Sag^{II} and Ax is similar: there is no statistically significant difference for specificity or CDD; while statistically significant differences are found for DSC and sensitivity, again the differences are insubstantial and both results are good.

Compared to the alternative methods, DeepSeA exhibited superior accuracy. On Sag^I, DeepSeA outperformed the previously validated edge-linking method [10] in DSC, sensitivity, and CDD with statistical significance, with the most notable improvement in CDD. On Sag^{II} and Ax, breasts segmented by DeepSeA and the surface-fitting method both seem highly consistent with those in the reference standard in high DSCs, sensitivity and specificity. When looking into CWL locations, however, while DeepSeA gives small CDDs, the surface fitting is significantly less accurate ($p = 0.005$ and 0.008).

C. Discussion

1) Accuracy and Applicability.: Versus manually-labelled reference standard from an experienced breast-imaging radiologist, DeepSeA produced accurate whole breast segmentation in our experiments.

DeepSeA does not require explicit correction of the bias-field artifact, which by itself is a challenging problem; instead, it spontaneously adapts to the spatially varying bias field and the varying appearance of the CWL. The significance of this self-adaptation scheme was demonstrated by showing that fitting a nonparametric smooth surface with the true positive set S_{TP}^0 after initialization resulted in CWLs much worse than those detected by DeepSeA.

This is because the fitted surface, despite being smooth, does not reflect the actual shape of the CWL at false positive locations as DeepSeA does (Fig. 7).

In addition, DeepSeA is genuinely 3-D and applicable to both sagittal and axial breast MRI; it produced similarly good results on Sag^{II} and Ax. The consistent accuracy across imaging orientations indicates DeepSeA's potential as a reliable tool to aid clinicians in quantitative analysis of breast MRI. Also, DeepSeA can be used to facilitate development of deep learning-based breast segmentation methods, such as deep CNNs, which often require a large number of labeled training samples. For instance, Dalmi *et al.* used an interactive software to manually outline breasts in [20]; we expect similar processes to be expedited in the future by using DeepSeA.

Using the Sag^I subset, we showed that DeepSeA outperformed the previously validated edge-linking method [10] with statistically significant improvements in DSC, sensitivity, and CWL location (Table V); we attribute the improvement to the 3-D continuity imposed by the smooth depth field and the self-adaptation scheme of DeepSeA. We also applied the edge-linking method to Sag^{II}; however it failed to produce meaningful CWLs for many slices, and we think it may be unfair to compare DeepSeA to this method on Sag^{II} considering that the older method was developed and validated with data acquired more than 10 years ago.

To compare the accuracy of DeepSeA with that of its sagittal-exclusive predecessor, we applied the method presented in our previous preliminary work [17] to Sag^I and Sag^{II} (combined $N=79$), and obtained the results $95.15 \pm 2.78\%$, $96.82 \pm 4.10\%$, $98.38 \pm 0.77\%$, and $2.42 \pm 2.75\text{mm}$ for DSC, sensitivity, specificity and CDD, respectively, in contrast to DeepSeA's accuracy of $96.41 \pm 1.55\%$, $97.73 \pm 1.68\%$, $98.60 \pm 0.80\%$, and $1.72 \pm 1.13\text{mm}$ on the combined sagittal data; all differences are statistically significant ($p = 0.004$ for sensitivity and $p \ll 0.01$ for others). DeepSeA's improvements on its predecessor may be attributed to 1) the experimentally determined optimal configuration, and 2) the change in technique for D^I initialization. The earlier technique simply initialized D^I to minima in $M_{\text{diff}}^{\text{scl}}$ (Fig. 8b). This sometimes trapped some regions of D at incorrect local minima if they were initialized to false positive CWL voxels (Fig. 8c). With the extra steps of true/false positive identification and randomized substitution (Fig. 8d), the current D^I initialization scheme reduces the chances of trapping (Fig. 8e). Therefore, in this work we not only extended the applicability of our method to axial data, but also improved its accuracy on sagittal data.

In [20], Dalmin *et al.* compared breast segmentation accuracies of four methods with 66 cases of breast MRI in axial view (44 of which were reconstructed from coronal orientation). Of the four methods, the 2C U-net segmented the breasts best by large margins over the atlas- and sheetness-based methods; its average DSC, sensitivity, and specificity against manual segmentation were $94.4 \pm 2.6\%$, $93.7 \pm 3.7\%$, and $99 \pm 0.5\%$, respectively. These numbers are similar to the corresponding values for DeepSeA. However, it may be inappropriate to directly compare to numbers reported in that related work, given that different data were used. As well, [20] focused on breast volume evaluation and provided no exclusive evaluation of CWL location. As demonstrated with the surface fitting method in our experiment, CWL locations are much more sensitive to segmentation defects and can be unsatisfactory even when breast volume-based metrics look good. A crucial differentiation between DeepSeA and the 2C U-net is that DeepSeA is not machine learning-based, and can be readily applied to new data using generalized configuration in the absence of labelled training samples (to be discussed shortly in Sec. III-C.3). A pretrained 2C U-net can also be directly applied to new data, but more likely to be subject to the overfitting problem which is common to the machine learning community (and more so in the context of medical image analysis due to inadequacy of high-quality manual labeling and dependency on acquisition parameters).

We observe that larger CWL errors mostly happen in far lateral regions on the two sides of the body, where the boundary is between breast and subcutaneous fat (e.g., Fig. 9) and much less anatomically well-defined than a boundary between breast and pectoral muscle. In this

case, DeepSeA separates breast from subcutaneous fat with a smooth CWL surface constrained by (1). This is similar to what human readers do when they smoothly extend the CWL along the pectoral muscle into the boundary-less lateral regions. The radiologist affirmed this behaviour of DeepSeA and its similarity to that of humans, despite numerical differences between the automatic and manual CWLs.

2) 3-D versus 2-D Templates: DeepSeA is a genuinely 3-D method. This study investigated how 3-D versus 2-D templates would affect DeepSeA's performance. As presented in Table IV axial data was the only exception in which 3-D templates noticeably outperformed 2-D ones. This distinction may result from the anisotropic voxels of the sagittal data. When the image voxels are close to isotropic, as in our axial data, 3-D templates may improve upon 2-D by incorporating extra information across slices. In contrast, when the voxels are very anisotropic (and subject to appreciable interpolation), little useful information for CWL localization may be captured by 3-D templates in the additional dimension due to the relatively big distance between neighboring slices.

Independent of segmentation accuracy, 3-D templates showed more consistent optimal parameters than 2-D in cross validation. This is likely because more information is included in a single 3-D unit than a 2-D counterpart of the same side length, leading to superior robustness and reliability.

Speed is another factor to consider. Setting $v = 2\text{mm}$ and $q = -1$, the average times to segment a sagittal MRI volume of 0.5mm isotropic voxels (typical size: $512 \times 512 \times 384$) using 3-D versus 2-D templates are 8.55 and 4.76 min, respectively; the corresponding times for 1mm voxels (typical size: $256 \times 256 \times 192$) are almost identical—66s versus 65s. Only at the finer resolution are 2-D templates considerably faster.

Based on this analysis, 3-D templates are preferred for the potential gain in accuracy, robustness, and reliability. However, if the data are appreciably anisotropic and running time is a problem, 2-D templates should also be considered.

3) Generalizability: DeepSeA is not machine learning-based so does not assume ground truth training data. Instead, we suggest generalized parameters so that users will not need to search for suitable parameters with a validation set before applying DeepSeA to their data.

The optimal parameter values identified in our cross-validation experiment showed generalizability across a variety of settings (Table IV): the optimal pair of (v, q) is consistently $(2, -1)$ for all folds in the coarse grid. In the fine grid the optimal q values sometimes deviate a bit from -1 and only in one setting—axial data processed with 2-D templates—the optimal v values deviate from 2, but the resulting CWL localization results do not meaningfully differ. We therefore expect the combination of 3-D templates, $v = 2\text{mm}$ and $\lambda = 10^{-1}$ to produce reasonable segmentation when applied to new data. (One caution is that these values are based on isotropic voxel sizes of 0.5 and 1mm. Therefore it will be necessary to interpolate the data to either of the two voxel sizes for the suggested configuration to work best. Notwithstanding, physical resolutions of modern breast MRI are usually around 0.5–1mm.) We applied DeepSeA with the above-described generalized

configuration to our test set—the Sag^I subset. Compared to the exploration sets, Sag^I was acquired much earlier with different imaging protocol and systems and included a larger cohort; also, the manual segmentation of Sag^I was done years earlier with a different software. Therefore, it could be a challenging test set. Nonetheless, DeepSeA achieved good accuracy on Sag^I which were comparable to that on the exploration set Sag^{II}, and better than that of a previous validated method [10]. We consider this good performance an indication of DeepSeA's generalizability.

Lastly, when ground truth is available, users can use the values presented in as Table IV reasonable starts to search nearby for parameters better fitted to their specific data. Considering that only two parameters are involved in this process, it should not be a difficult task.

4) Limitations and Directions for Future Work: One limitation of this work was that it was based on single-institution data, although these data covered reasonable variations in subjects' age, BMI, breast size and density (Table I), and some variations in imaging system, orientation, and protocol (Table II). In future work we plan to build a larger validation set with multi-institution data, preferably with diversified population composition, for a more rigorous assessment of DeepSeA.

As most related works, DeepSeA uses nonfat-suppressed breast MRI for whole breast segmentation and is not directly applicable to fat-suppressed images. Nevertheless, its key concepts of dense depth field modeling and self-adaptation can be directly generalized to fat-suppressed images. Hence, we believe that with minor adjustments DeepSeA will also work for fat-suppressed images and plan it as a future work.

Manual segmentation of breast MRI is extremely timeconsuming and labor-intensive. In this study, only a single set of manual segmentation was available and we were unable to investigate intra- or interobserver reliability. So far as we are aware of, only few works in the literature reported interobserver variation [27], [28], and none reported intraobserver variation, in MRI-based whole breast segmentation. Noticeably, Rosado-Toro *et al.* reported interobserver variability among three tracers with 266 axial slices of 14 subjects [28]. Using tracer 1 as reference and for tracers 2 and 3, the *slice-wise* average DSCs are $92.6 \pm 4.6\%$ and $92.9 \pm 5.4\%$, and 84.8% and 86.6% of the CWL pixels were within three pixels of the reference CWLs. With the due caution that the imaging protocols used were very different (registered fat and water MRIs were used in [28]), we computed the same metrics between DeepSeA and our manual reference on the Ax subset and obtained numbers comparable to the reported interobserver variations: $95.0 \pm 3.8\%$ for slice-wise average DSC and 92.7% CWL pixels within three pixels of the reference.

In this paper we compared the accuracy of DeepSeA with that of a previously validated method [10] on the Sag^I subset and showed statistically significant improvements. In future work we plan to more broadly compare DeepSeA to other methods. As we rarely see comparable methods applied to breast MRI in both sagittal and axial views for thorough assessment in the literature, we will select competing methods that can be expected to work well in their preferred views.

As mentioned in Section I, whole breast segmentation often serves as the prerequisite for quantitative analysis within the breast. To move towards clinically applicable technologies, we are currently particularly interested in indices that are indicative of breast cancer risk for normal (i.e., cancer-unaffected) screening breasts. Therefore, we plan to extend DeepSeA to automated methods for quantitative assessment of FGT and BPE as the immediate next step in methodology development. After that, we will use DeepSeA for clinically relevant cancer risk assessment and evaluate its effectiveness against actual patient outcomes (e.g., cancer incidence/recurrence).

IV. Conclusions

In this paper we presented DeepSeA, an automatic 3-D segmentation method of whole breasts applicable to *both* sagittal and axial breast MRI. DeepSeA features two novelties. First, it reformulates the challenging task of CWL localization to the equivalent task of optimizing a smooth depth field, so that the natural continuity of the CWL surface is fully utilized. Second, it implements a localized self-adaptation scheme so that the search for the CWL at different locations is locally adjusted to the varying appearance of the CWL surface. Experimental results on real patient data showed that DeepSeA can successfully segment breasts in both sagittal and axial views with good accuracy and reliability. Furthermore, we generalized the configuration of DeepSeA parameter settings and template type choices to allow for broad application of DeepSeA to a wide range of breast MRI data.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgment

This study was supported by the NIH/NCI grants 1R01CA 197000-01A1, 5R01CA183086-03, and 5U24CA189523-02. We thank Eric Cohen for discussion on statistical analysis.

References

- [1]. Siegel RL et al., "Cancer statistics, 2018," *CA Cancer J. Clin.*, vol. 68, no. 1, pp. 7–30, 2018. [PubMed: 29313949]
- [2]. Berry DA et al., "Effect of screening and adjuvant therapy on mortality from breast cancer," *N. Engl. J. Med.*, vol. 353, no. 17, pp. 1784–1792, 2005. [PubMed: 16251534]
- [3]. Lehman CD et al., "Screening women at high risk for breast cancer with mammography and magnetic resonance imaging," *Cancer*, vol. 103, no. 9, pp. 1898–1905, 2005. [PubMed: 15800894]
- [4]. Saslow D et al., "American cancer society guidelines for breast screening with MRI as an adjunct to mammography," *CA Cancer J. Clin.*, vol. 57, no. 2, pp. 75–89, 2007. [PubMed: 17392385]
- [5]. Kuhl CK et al., "Supplemental breast MR imaging screening of women with average risk of breast cancer," *Radiology*, vol. 283, no. 2, pp. 361–370, 2017. [PubMed: 28221097]
- [6]. Smith KL and Isaacs C, "Management of women at increased risk for hereditary breast cancer," *Breast Disease*, vol. 27, no. 1, pp. 51–67, 2007.
- [7]. Tice JA and Kerlikowske K, "Screening and prevention of breast cancer in primary care," *Prim. Care: Clinics in Office Practice*, vol. 36, no. 3, pp. 533–558, 2009. [PubMed: 19616154]
- [8]. Klifa C et al., "Magnetic resonance imaging for secondary assessment of breast density in a high-risk cohort," *Magn. Reson. Imaging*, vol. 28, no. 1, pp. 8–15, 2010. [PubMed: 19631485]

- [9]. Telegrafo M et al., "Breast MRI background parenchymal enhancement (BPE) correlates with the risk of breast cancer," *Magn. Reson. Imaging*, vol. 34, no. 2, pp. 173–176, 2016. [PubMed: 26597834]
- [10]. Wu S et al., "Automated chest wall line detection for whole-breast segmentation in sagittal breast MR images," *Med. Phys.*, vol. 40, no. 4, p. 042 301, 2013.
- [11]. Wu S et al., "Automated fibroglandular tissue segmentation and volumetric density estimation in breast MRI using an atlas-aided fuzzy C-means method," *Med. Phys.*, vol. 40, no. 12, p. 122 302, 2013.
- [12]. Gubern-Mérida A et al., "Breast segmentation and density estimation in breast MRI: A fully automatic framework," *IEEE J. Biomed. Health Inform.*, vol. 19, no. 1, pp. 349–357, 1 2015. [PubMed: 25561456]
- [13]. Klifa C et al., "Quantification of background enhancement in breast magnetic resonance imaging," *J. Magn. Reson. Imaging*, vol. 33, no. 5, pp. 1229–1234, 2011. [PubMed: 21509883]
- [14]. Wu S et al., "Quantitative assessment of background parenchymal enhancement in breast MRI predicts response to risk-reducing salpingo-oophorectomy: Preliminary evaluation in a cohort of BRCA1/2 mutation carriers," *Breast Cancer Res.*, vol. 17, no. 1, p. 67, 2015. [PubMed: 25986460]
- [15]. Gubern-Mérida A et al., "Breast segmentation in MRI: Quantitative evaluation of three methods," in *Proc. SPIE*, vol. 8669, 2013, 86693G.
- [16]. Lin M et al., "Template-based automatic breast segmentation on MRI by excluding the chest region," *Med. Phys.*, vol. 40, no. 12, p. 122 301, 2013.
- [17]. Wei D et al., "Three-dimensional whole breast segmentation in sagittal MR images with dense depth field modeling and localized self-adaptation," in *Proc. SPIE*, vol. 10133, 2017, p. 1 013 314.
- [18]. Ortiz CG and Martel A, "Automatic atlas-based segmentation of the breast in MRI for 3D breast volume computation," *Med. Phys.*, vol. 39, no. 10, pp. 5835–5848, 2012. [PubMed: 23039622]
- [19]. Wang Y et al., "3D multi-parametric breast MRI segmentation using hierarchical support vector machine with coil sensitivity correction," *Acad. Radiol.*, vol. 20, no. 2, pp. 137–147, 2013. [PubMed: 23099241]
- [20]. Dalmı MU et al., "Using deep learning to segment breast and fibroglandular tissue in MRI volumes," *Med. Phys.*, vol. 44, no. 2, pp. 533–546, 2017. [PubMed: 28035663]
- [21]. Morris EA et al., "ACR BI-RADS® Magnetic Resonance Imaging," in *ACR BI-RADS® Atlas, Breast Imaging Reporting and Data System*, Reston, VA: American College of Radiology, 2013.
- [22]. Yushkevich PA et al., "User-guided 3D active contour segmentation of anatomical structures: Significantly improved efficiency and reliability," *Neuroimage*, vol. 31, no. 3, pp. 1116–1128, 2006. [PubMed: 16545965]
- [23]. Rudin LI et al., "Nonlinear total variation based noise removal algorithms," *Physica D: Nonlinear Phenomena*, vol. 60, no. 1, pp. 259–268, 1992.
- [24]. Lilliefors HW, "On the Kolmogorov-Smirnov test for normality with mean and variance unknown," *J. Am. Stat. Assoc.*, vol. 62, no. 318, pp. 399–402, 1967.
- [25]. Sibson R, "A brief description of natural neighbour interpolation," *Interpreting multivariate data*, 1981.
- [26]. Milenkovi J et al., "Automated breast-region segmentation in the axial breast MR images," *Comput. Biol. Med.*, vol. 62, pp. 55–64, 2015. [PubMed: 25912987]
- [27]. Ortiz CG and Martel AL, "Automatic atlas-based segmentation of the breast in MRI for 3D breast volume computation," *Med. Phys.*, vol. 39, no. 10, pp. 5835–5848, 2012. [PubMed: 23039622]
- [28]. Rosado-Toro JA et al., "Automated breast segmentation of fat and water MR images using dynamic programming," *Acad. Radiol.*, vol. 22, no. 2, pp. 139–148, 2015. [PubMed: 25572926]

Fig. 1.

(a) A reformatted axial view of a sagittal breast MRI, versus (b) a native axial slice. In the sagittal imaging protocol, the two breasts were imaged separately and their images were not aligned with each other, thus there is an obvious gap in the middle. This results in the absence of certain mid-thoracic landmarks—e.g., the sternum, which is pointed out in (b) for comparison—and uncertainty over bilateral symmetry.

Fig. 2.

Flowchart of DeepSeA. (a) Input: a volumetric breast MRI; here a slice of the volume is used for illustration (Slc_{61} of Ax_{11}). (b) Air-breast boundary (ABB) detection. (c) Breast lateral bounding plane delineation. (d) CWL localization. (e) Breast inferior bounding plane delineation. (f) Breast medial bounding plane delineation. (g) Output: 3-D segmentation of the breasts, viewed from two perspectives. Here body parts beyond the medial and inferior breast bounding planes (locations of the bounding planes are marked with red asterisks) are displayed too for visual perception; the ABB surface is displayed semi-transparently. The task of breast bounding plane delineation, including steps (c), (e), and (f) enclosed in the dashed box, is performed in projected 2-D views for the ease of breast boundary observation in such views.

Fig. 3.

Illustration of the RCS and CWL ROI bounding planes using (a) Sag_I^{II} and (b) Ax_3 . In both figures, the three triangles from the anterior to posterior indicate y_{nip} , y_{base} and y_{rear} , respectively, and the two dashed lines indicate the front and back bounding planes of the ROI.

Fig. 4.

(a) SIC_{50} of Sag_1^{II} . (b) The ROI (outlined in green) for CWL search. (c) Top: the 2-D generic template T_{gen} , which is constructed by stacking repeated 1-D templates along the orthogonal direction in the image plane; middle: the locally adapted template $T_{\text{adpt}}^{\text{loc}}$ for the red line in (b); bottom: the globally adapted template $T_{\text{adpt}}^{\text{glob}}$. (d) The weighted and then scaled filter response image by T_{gen} . (e) CWL initialization (false positives in red). (f) Inverse of the difference image after application of the self-adapted templates T_{adpt} (inverted for visibility).

Fig. 5.

(a) Changes of $E[D]$, E_{patn} , and E_{smth} during evolution of the depth field D . Vertical axes: $E[D]$, E_{patn} , and E_{smth} , respectively; horizontal axes: number of iterations. (b) The depth field after optimization for Ax_{11} shown in Fig. 2; this is an inverted display of the original D for inspection, i.e., shallow depths appear brighter.

Fig. 6.

Example breast segmentation results. Top row: the original MR images; bottom row: segmentations by DeepSeA (in red), compared to those by the radiologist (in green); overlaps between the two contours are displayed in yellow. (a) Slc₆₇ of Sag₁₆^{II}, scattered FGT (BI-RADS density rating). (b) Slc₁₆ of Sag₁₂^{II}, heterogeneous FGT. (c) Slc₁₆ of Sag₁^{II}, extreme FGT. (d) Slc₁₂₄ of Ax₉, almost entirely fat. (e) Slc₈₂ of Ax₁₂, heterogeneous FGT. (f) Slc₆₆ of Ax₃, extreme FGT. As we can see, for most part the two contours overlap or are close, except for minor discrepancies near the superior/inferior (for sagittal slices) or medial/lateral (for axial slices) bounding planes.

Fig. 7.

(a) Slc_{52} of Sag_5^{Π} . (b) The fitted smooth surface does not follow the CWL where false positives are identified in M_{CWL}^0 . (c) In contrast, DeepSeA locates the CWL correctly with the self-adaptation scheme.

Fig. 8.

(a) Slc_{21} of Sag_4^{II} . (b) Previously in [17], D^{l} was simply initialized to minima in $M_{\text{diff}}^{\text{scl}}$ and (c) the CWL could be trapped at incorrect local minima by false positive CWL voxels during depth field optimization. (d) In this work, D^{l} is initialized with a random process where false positive CWL voxels are identified, and (e) the trapping is effectively eliminated.

Fig. 9.

(a) SLC_{14} of Sag_9^{II} : an example body-side slice in which the boundary between breast and subcutaneous fat is ambiguous. (b) Breast outlined by DeepSeA. (c) Breast outlined by the radiologist. Despite numerical differences between the automatic and manual CWLs, they both give reasonable breast segmentation.

TABLE I

Demographic information of the study population.

	Sag ^I	Sag ^{II}	Ax
No. of subjects	59	20	20
Age (y)			
Range	26–63	24–64	25–64
Median	45	51.5	44.5
IQR ^a	10	28.5	20.0
BMI ^b (kg/m ²)			
Range	18.75–36.61	19.14–39.87	18.26–46.21
Median	24.91	23.39	22.81
IQR	6.62	6.73	8.4
Breast density ^c (No. of cases)			
Almost entirely fat	17	5	5
Scattered FGT	15	5	5
Heterogeneous FGT	14	5	5
Extreme FGT	13	5	5
Breast size ^d (cm ³)			
Range	377–3529	401–4460	403–6787
Median	1220	1461	1615
IQR	1241	1856	1518

^aIQR: interquartile range.^bBMI: body mass index.^cAccording to the four-point American College of Radiology BI-RADS breast density categorization [21].^dComputed from manually outlined breast regions.

TABLE II

MRI acquisition parameters.

	Sag^I	Sag^{II}	Ax
Orientation	Sagittal	Sagittal	Axial
Scanner	GE LX echo speed or Siemens Sonata	Siemens Espree	
Field strength (T)	1.494	1.5	1.5
Flip angle (°)	20	14.88	25
Echo time (ms)	4.38	4.78	4.78
Repetition time (ms)	9.7	8.12	6.93
Slice size	256×256	512×512	256×256
Pixel size, isotropic (mm)	0.70–0.86	0.47 or 0.51	1.09–1.41
Slice thickness (mm)	2.3–4	2.7 or 3	1
Slices per MRI	56	64–104	192–256
Field of view (cm)	18–22	24 or 26	28–36

TABLE III

The coarse and fine grids used in the search for optimal parameters.

Grid	Voxel size (mm)	r (mm)	q
Coarse	0.5	$\{1, {}^a 2, 3, 4\}$	$\{-3, -2, -1, 0\}$
	1	$\{2, {}^a 4, 6, 8\}$	
Fine ^b	0.5	Step size 0.5 ^c	Step size 0.25
	1	Step size 1 ^c	

^a 1mm and 2mm are the smallest r values that are allowed given the voxel sizes of 0.5mm and 1mm, respectively.

^b The fine grid is centered at the optimal points obtained from the coarse grid.

^c 0.5mm and 1mm are the smallest step sizes that are allowed given the voxel sizes of 0.5mm and 1mm, respectively.

TABLE IV

Cross-validation results using the grid search for optimal parameters. The unit for voxel size, v and cross validation results (CDD) is mm; CDD is presented as mean $\pm$ std.

Data	Voxel size	Template dimensionality	Grid	Optimal parameters (v, q) for Fold					Cross validation results (CDD);
				No. 1	No. 2	No. 3	No. 4	No. 5	
Sag ^{II}	0.5	2-D	Coarse	(2, -1)	(2, -1)	(2, -1)	(2, -1)	(2, -1)	1.41 $\pm$ 0.95
			Fine	(2, -1)	(2, -1)	(2, -1)	(2, -1)	(2, -1.25)	1.41 $\pm$ 0.97
		3-D	Coarse	(2, -1)	(2, -1)	(2, -1)	(2, -1)	(2, -1)	1.48 $\pm$ 0.88
			Fine	(2, -1)	(2, -1)	(2, -1)	(2, -1)	(2, -1)	1.48 $\pm$ 0.88
	1	2-D	Coarse	(2, -1)	(2, -1)	(2, -1)	(2, -1)	(2, -1)	2.01 $\pm$ 1.22
			Fine	(2, -0.75)	(2, -0.75)	(2, -0.75)	(2, -1)	(2, -1)	2.02 $\pm$ 1.22
Ax	1	3-D	Coarse	(2, -1)	(2, -1)	(2, -1)	(2, -1)	(2, -1)	2.00 $\pm$ 1.15
			Fine	(2, -0.75)	(2, -0.75)	(2, -0.75)	(2, -0.75)	(2, -0.75)	1.99 $\pm$ 1.13
		2-D	Coarse	(2, -1)	(2, -1)	(2, -1)	(2, -1)	(2, -1)	1.67 $\pm$ 1.96
			Fine	(2, -1.25)	(3, -1)	(3, -1.25)	(3, -1.25)	(3, -1.25)	1.67 $\pm$ 1.82
		3-D	Coarse	(2, -1)	(2, -1)	(2, -1)	(2, -1)	(2, -1)	1.25 $\pm$ 1.26
			Fine	(2, -1)	(2, -1.25)	(2, -1.25)	(2, -1)	(2, -1.25)	1.26 $\pm$ 1.26

For the same imaging orientation (sagittal or axial) and voxel size (0.5mm or 1mm), no statistically significant difference is found at the 5% significance level for varying template dimensionality or grid coarseness, except for Ax 2-D versus 3-D over the fine grid ($p = 0.005$).

TABLE V

Quantitative evaluation results (presented as mean±std).

	Breast volume			CWL location	
	DSC (%)	Sensitivity (%)	Specificity (%)	CDD (mm)	
Overall: N=99					
DeepSeA ^a	96.04±1.68	97.27±2.07	98.77±0.80	1.63±1.16	
Sag ^I : N=59					
DeepSeA	96.51±1.57	97.72±1.35	98.38±0.73	1.82±1.17	} ^d
Edge linking [10]	95.08±5.97	95.61±8.23	98.32±0.88	2.81±1.80	
Sag ^{II} : N=20					
DeepSeA	96.12±1.48	97.77±2.45	99.24±0.64	1.42±0.96	} ^f
Surface fitting	95.27±2.07	97.33±2.50	98.89±1.79	2.94±4.47	
Ax: N=20					
DeepSeA	94.60±1.41	95.44±2.47	99.43±0.31	1.26±1.24	} ^g
Surface fitting	94.34±1.63	95.06±2.76	99.41±0.36	1.80±2.62	

^aFor DeepSeA:

1) No statistically significant difference ($\alpha = 0.05$) between Sag^I and Sag^{II} in DSC, sensitivity, or CDD; although significant difference is found in specificity ($p \ll 0.01$), the difference is insubstantial and the results are good for both data.

2) No statistically significant difference between Sag^{II} and Ax in specificity or CDD; although significant differences are found in DSC and sensitivity ($p = 0.002$ for both), again the differences are insubstantial and both results are good.

^{b-d}Statistically significant difference: $p \ll 0.01$.

^{e-g}Statistically significant difference: $p = 0.001, 0.005, \text{ and } 0.008$, respectively.