A survey on breast lesion segmentation in Ultra-Sound images and its assessment: A survey

Joan Massich[1,☯,¤a], Guillaume Lemaître[1,2,☯,¤a], Mojdeh Rastrgoo[1,2,☯,¤a], Joan Martí[2,¤b], Fabirce Mérieaudeau[1,‡,¤a],

1 Le2i-UMR CNRS 6306, Université de Bourgogne, 71200 Le Creusot, France
2 ViCOROB, Universitat de Girona, 17071 Girona, Spain

☯These authors contributed equally to this work.
‡These authors also contributed equally to this work.
¤a 12 rue de la Fondaire, 71200 Le Creusot, France ¤b Campus Montilivi Ed.P4, 17071 Girona, Spain
†Deceased
¶Membership list can be found in the Acknowledgments section.
*** joan.massich@u-bourgogne.fr**

Abstract

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

Introduction

sec:intro

- Narrow to the need of accurate delineations
- What is reviewed and what is not
- Article Objectives
- similar works
- paper structure

sec:intro:to_delinations

the need of accurate delineations

1. breast cancer kills
2. screening is needed for early detection
3. Health from images is like any other visual inspection
   - Radiologic diagnisis error rates are similar to any other human visual inspection manning2005perception [1]
   - Utilization of computers to aid the Radiologists during the diagnosis process giger2008anniversary [2]
4. Advantage of Ultra-Sound (US)
5. Breast Imaging-Reporting and Data System (BI-RADS)
6. need of accurate delineations

Breast cancer is the leading cause of cancer deaths among females worldwide cancerStatisti [3]. Nevertheless, death by breast cancer are highly reduced when early treated. Thus, to run a chance of surviving breast cancer, it is uttermost important the early detection of malignant tumors. This has motivated the establishment of Breast Screening PolicyBreast Screening Policies (BSPs) to facilitate this breast cancer detection at an early stage. Despite X-ray Digital Mammography (DM) is considered the gold standard technique for BSP, other screening techniques like US and Magnetic Resonance Imaging are being investigated to overcome DM limitations due to tissue superposition which can either mimic or obscure malignant pathology, and avoid X-ray radiation all together.

Medical imaging contributes to its early detection through screening programs, non-invasive diagnosis, follow-up, and similar procedures. Despite Breast Ultra-Sound (BUS) imaging not being the imaging modality of reference for breast cancer screening smith2003american [4], US imaging has more discriminative power when compared with other image modalities to visually differentiate benign from malignant solid lesions Stavros:1995p12392 [5]. In this manner, US screening is estimated to be able to reduce between $65 \sim 85\%$ of unnecessary biopsies, in favour of a less traumatic short-term screening follow-up using BUS images. As the standard for assessing these BUS images, the American College of Radiology (ACR) proposes the BI-RADS lexicon for BUS images piradsus [6]. This US BI-RADS lexicon is a set of standard markers that characterizes the lesions encoding the visual cues found in BUS images and facilitates their analysis. Further details regarding the US BI-RADS lexicon descriptors proposed by the ACR, can be found in Sect. sec:methodApp ??, where visual cues of BUS images and breast structures are discussed to define feature descriptors.

The incorporation of US in screening policies and the emergence of clinical standards to assess image like the US BI-RADS lexicon, encourage the development of Computer Aided Diagnosis (CAD) systems using US to be applied to breast cancer diagnosis. However, this clinical assessment using lexicon is not directly applicable to CAD systems. Shortcomings like the location and explicit delineation of the lesions need to be addressed, since those tasks are intrinsically carried out by the radiologists during their visual assessment of the images to infer the lexicon representation of the lesions.

sec:intro:what_is_reviewed

What is reviewed

sec:intro:article_objective

Article objective The objective of this article is - to provide an exhaustive list of segmentation methodologies that have been developed for delineating breast lesions in

US images. - to collect a set of terms that facilitate placing each work with respect of the rest of the State-of-the-Art. - to provide an overview of how each methodology has been assessed. - to clarify assessment assumptions that influence a fair results comparison.

A secondary objective of the authors, but equally essential to us, is - to ensure comprehensive coverage of the available segmentation methodologies. - fair treatment. - reusability of the efforts put in this article. Thus, the set of queries that generate the pool of methodologies present in this work are available at the *us-breast-lesion-delineation-survey* open repository as scripts. These queries screen all the publications included by targeted journals during the last 5 years for articles matching any of the search terms in order to build a bibliographic dataset. This dataset is complemented collecting all the articles that expand the citations tree, both forward and backwards, up to 2 levels. More details about this process can be found at the website of the project. Automatic criteria are used for rough pruning of the dataset while the final pruning has been carried out manually. Details of both are provided at the website.

project website in github

sec:intro:analysis_of_the_methods

Analysis of the methods

sec:intro:similar_works

similar works

sec:intro:paper_structure

Paper structure

1 Segmentation assessment

section:assessment

Comparing all the methodologies reviewed in section section:userInteraction ?? is rather cumbersome. The lack of a common framework for assessing the methodologies remains unaddressed, especially due to the absence of a public image dataset despite its being highly demanded by the scientific community [Noble:2006p1734,Noble:2009p14330,Cheng:2009pI0580 ? ? ?]. However, the lack of a common dataset is not the only aspect complicating the comparisons. Here is a list of some of the feasible aspects complicating direct comparison of the works reviewed.

Make a study refering to confusion matrix and all roc curves

- Uncommon database
- Uncommon assessing of criteria and metrics
- Different degrees of user interaction
- Inability to quantify the user effort when interacting with a method
- Correctness of the Ground Truth (GT) used when assessing
- Uncommon treatment of missegmentation due to unpropper detection

The dificulty of comparing the methodologies using distinct datasets, distinct assessing criteria and distinct metrics is clear. Section section:evalCriteria 1.1 analyzes the criteria and metrics used to analyze the different methodology proposals. In order to conduct a discussion comparing the methodologies in section discussion ??, when enough information is available, the reported results are set to a common framework for comparison purposes despite being assessed with different datasets. The assessment regarding user interaction is not further analyzed other than the already described interactive and automatic classification along with their respective subcategories (see section section:userInter ?? and fig fig:segmentationInteractivityTable ??). The correctness of the GT for assessing the segmentations refers to the huge

		Segmentation Ground Truth (GT) (reference)	
		Positive	Negative
Segmentation Outcome (prediction)	Positive	True Positive (TP)	False Positive (FP)
	Negative	False Negative (FN)	True Negative (TN)

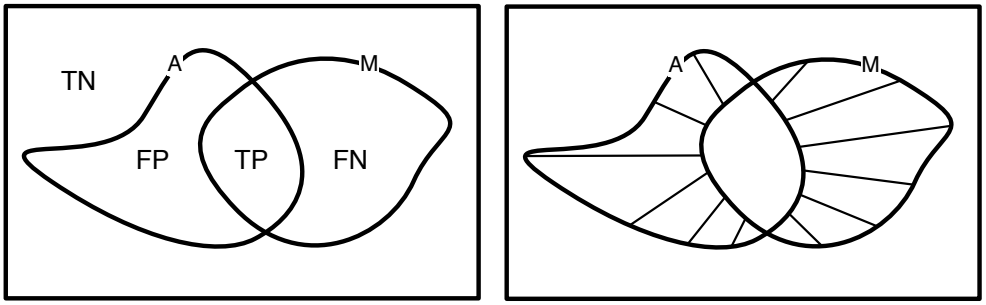

Figure 1. Methodology evaluation. (a) Statistical hypothesis test errors confusion matrix. (b) Graphic representation of the statistical hypothesis test errors for assessing the performance in terms of area. (c) Graphical representation of the boundary distance performance measures.

fig:gtEval

variability of the delineations found when analyzing intra expert and inter expert variability on the segmentations [7]. In this regard, later in this article (see section: 1.2), a short discussion about the work that took intra and inter-observer delineation variability into account for assessing segmentation proposals can be found. Finally, the frontier between segmentation errors and errors due to the detection process is unclear and a proper criterion is not set. Massich et al. [8] take all the segmentations into account even if the segmentation has been wrongly initialized by the automatic detection procedure. Meanwhile, Zhang et al. [9] only use 90% of the best segmentations to perform the segmentation assessment, arguing that the remaining segmentations suffered poor detection and that segmentation result assessment should not be subject to wrong initializations.

The rest of this section describes different area and boundary metrics collected from the works cited above, comments on the correctness of the assessing GT, based on intra- and inter-observer GT, variability and discusses the results reported.

1.1 Evaluation criteria

section:evalCriteria

Although multiple criteria arise when assessing segmentations, these criteria can be grouped into two families depending on whether they are area or distance based metrics as illustrated in figure 1. Area based metrics assess the amount of area shared (Area Overlap (AOV)) between the obtained segmentation and the reference. On the other hand, distance based metrics quantify the displacement or deformation between the obtained and the desired delineations.

For the sake of simplicity, the name of the reported similarity indexes has been unified.

Area based segmentation assessment metrics When analyzing the areas described by the segmented region to be assessed, A and the manually delineated reference region M (see fig. 1b), 4 areas become evident: True Positive (TP), True

Negative (TN), False Positive (FP), and False Negative (FN); corresponding to the regions of the confusion matrix in figure 1a.

Area metrics (or indexes) for assessing the segmentation are defined as a dimensionless quotient relating the 4 regions (TP, FP, FN and TN) described by the segmentation outcome being assessed (denoted A in fig:1a) and the reference GT segmentation (denoted M). Most of the indexes are defined within the interval $[0, 1]$ and some works report their results as a percentage.

Area Overlap (AOV), also known as overlap ratio, the Jaccard Similarity Coefficient (JSC) [10] or Similarity Index (SI) [11][1], is a common similarity index representing the percentage or amount of area common to the assessed delineation A and the reference delineation M according to equation 1. The AOV metric has been used to assess the following works: [8, 11, 14, 15, 16, 17, 18, 19]

$$AOV = \frac{TP}{TP + FP + FN} = \frac{|A| \wedge |M|}{|A| \vee |M|} \qquad \in [0, 1] \tag{1}$$

Dice Similarity Coefficient (DSC), also found under the name of SI [12, 13][2], is another widely used overlap metric similar to AOV. The difference between DSC and AOV is that DSC takes into account the TP area twice, one for each delineation. The DSC index is given by equation 2 and the relation between AOV or JSC and the DSC similarity indexes is expressed by equation 3. Notice that the DSC similiarity index is expected to be greater than the AOV index [7]. The DSC metric has been used to assess the following works:[7, 9, 12, 13]

$$DSC = \frac{2 \cdot TP}{2 \cdot TP + FP + FN} = \frac{2|A \wedge M|}{|A| + |M|} \qquad \in [0, 1] \tag{2}$$

$$DSC = \frac{2 \cdot AOV}{1 + AOV} \tag{3}$$

True-Positive Ratio (TPR), also known as the recall rate, sensitivity (at pixel level) [7, 20] or Overlap Fraction [12], quantifies the amount of properly labeled pixels as lesion with respect to the amount of lesion pixels from the reference delineation (eq: 4). Notice that like the DSC, this value always remains greater than AOV (or equal when the delineations are identical). The TPR metric has been used to assess the following works: [11, 12, 13, 19, 20, 21, 22, 23]

$$TPR = \frac{TP}{TP + FN} = \frac{TP}{|M|} = \frac{|A| \wedge |M|}{|M|} \qquad \in [0, 1] \tag{4}$$

Positive Predictive Value (PPV) corresponds to the probability that the pixel is properly labeled when restricted to those with positive test. It differentiates from TPR since here the TP area is regularized by the assessing delineation and not the reference, as can be seen in equation 5. PPV is also greater than AOV. The PPV metric is also used to assess the work in [7].

$$PPV = \frac{TP}{FP + TP} = \frac{TP}{|A|} = \frac{|A| \wedge |M|}{|A|} \qquad \in [0, 1] \tag{5}$$

[1] Notice that Similarity Index (SI) is also used formulated as the Dice Similarity Coefficient (DSC) in [12, 13] which differs from the SI definition in [11].

[2] Notice that Similarity Index (SI) is also used formulated as the Area Overlap (AOV) in [11] which differs from the SI definition in [12, 13].

Normalized Residual Value (NRV), also found as the Precission Ratio(PR) [24], corresponds to the area of disagreement between the two delineations regularized by the size of the reference delineation, as described in equation: 6. Notice that the NRV coefficient differs from $1 - AOV$ since it is regularized by the reference delineation and not the size of the union of both delineations. The NRV metric has been used to assess the following works: [15, 24, 25].

$$NRV = \frac{|A \oplus M|}{|M|} \qquad \in \left[0, 1 + \frac{A}{|M|}\right] \tag{6}$$

False-Positive Ratio' (FPR'), as reported in the presented work, is the amount of pixels wrongly labeled as lesion with respect to the area of the lesion reference, as expressed in equation 7. The FPR' metric has been used to assess the following works:[11, 19, 21, 22, 23] The FPR' has also been found in its complementary form $1 - TPR$ under the name of Match Rate (MR) [24].

$$FPR' = \frac{FP}{TP + FN} = \frac{FP}{|M|} = \frac{|A \vee M - M|}{|M|} \qquad \in \left[0, \frac{A}{|M|}\right] \tag{7}$$

Notice that the FPR' calculated in equation 7 differs from the classic False-Positive Ratio (FPR) obtained from the table in figure 1a, which corresponds to the ratio between FP and its column marginal $(FP + TN)$, as indicated in equation 8. The FPR, when calculated according to equation 8, corresponds to the complement of specificity (described below).

$$FPR = \frac{FP}{FP + TN} = 1 - SPC \qquad \in [0, 1] \tag{8}$$

False-Negative Ratio (FNR) corresponds to the amount of pixels belonging to the reference delineation that are wrongly labeled as background, as expressed in equation 9. Notice that it also corresponds to the inverse of the TPR since $TP \cup FN = M$. The FNR metric has been used to assess the following works: [21, 22, 23]

$$FNR = \frac{FN}{|M|} = \frac{|A \vee M - A|}{|M|} = 1 - TPR \qquad \in [0, 1] \tag{9}$$

Specificity corresponds to the amount of background correctly labeled. Specificity is described in equation 10 and is usually given as complementary information on the sensitivity (TPR). Specificity corresponds to the complementary of the FPR when calculated according to equation 8. The specificity index is also used to assess the work in [7, 20].

$$SPC = \frac{TN}{TN + FP} = \frac{|\overline{A} \wedge \overline{M}|}{|\overline{M}|} = 1 - FPR \qquad \in [0, 1] \tag{10}$$

Boundary based segmentation assessment metrics Although the boundary assessment of the segmentations is less common than area assessment, it is present in the following works: [9, 10, 11, 15, 16, 21, 22]. Like when assessing the segmentations in terms of area, the criteria for assessing disagreement between outlines are also heterogeneous which makes the comparison between works difficult. Unlike the area indexes, with the exception of the further introduced Average Radial Error (ARE) coefficient, which is also a dimensionless quotient, the rest of the boundary indexes or

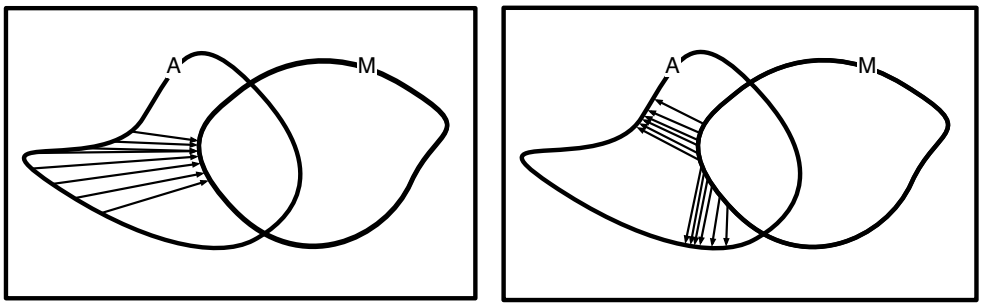

Figure 2. Illustration of the non-symmetry property of the Minimum Distance (MD) metric. (a) $\mathrm{MD}(a_i, M)$, (b) $\mathrm{MD}(m_i, A)$

fig:md

metrics are physical quantitative error measures and are assumed to be reported in pixels. Although some of the reported measures are normalized, they are not bounded by any means.

Zhang et al. [9] propose using average contour-to-contour distance (E_{cc}) for assessing their work. However, no definition or reference is found on it. Huang et al. [22] propose using ARE, defined in equation 11, where a set of n radial rays are generated from the center of the reference delineation C_0 intersecting both delineations. The ARE index consists of averaging the ratio between the distance of the two outlines $|C_s(i) - C_r(i)|$ and the distance between the reference outline and its center $|C_r(i) - C_0|$.

$$ARE = \frac{1}{n}\sum_{i=1}^{n}\frac{|C_s(i) - C_r(i)|}{|C_r(i) - C_0|} \tag{11}$$

The rest of the works base their similitude indexes on the analysis of the Minimum Distance (MD) coefficients. The MD is defined in equation 12 and corresponds to the minimum distance between a particular point a_i within the contour A (so that $a_i \in A$) and any other point within the delineation M.

$$\mathrm{MD}(a_i, M) = \min_{m_j \in M} \|a_i - m_j\| \tag{12}$$

Hausdorff Distance (HD), or Hausdorff error, measures the worst possible discrepancy between the two delineations A and M as defined in 13. Notice that it is calculated as the maximum of the worst discrepancy between (A, M) and (M, A) since MD is not a symmetric measure, as can be observed in figure 2. The HD as defined in equation 13 has been used for assessing the segmentation results in Gao et al. [10]. Meanwhile, Madabhushi and Metaxas [21] and Shan et al. [11] only take into account the discrepancy between the assessed delineation A with reference delineation M, here denoted as HD' (see eq. 14). In [11, 21], the HD' is also reported in a normalized form $\frac{\mathrm{HD}'}{\eta}$, where η is the length of the contour of reference M.

$$\mathrm{HD}(A, M) = \max\left\{\max_{a_i \in A}\mathrm{MD}(a_i, M), \max_{m_i \in M}\mathrm{MD}(m_i, A)\right\} \tag{13}$$

$$\mathrm{HD'}(A, M) = \max_{a_i \in A}\mathrm{MD}(a_i, M) \tag{14}$$

Average Minimum Euclidian Distance (AMED), defined in equation 15, is the average MD between the two outlines. [10]. Similar to the case of the HD' distance, Madabhushi and Metaxas [21] and Shan et al. [11] only take into account the discrepancy between the assessed delineation A with reference to the delineation M to

calculate the AMED' index (see eq. 16). The AMED index can be found under the name of Mean Error (ME) in [21] and Mean absolute Distance (MD) in. [11].

$$\mathrm{AMED}(A,M) = \frac{1}{2}\cdot\left[\frac{\sum_{a_i\in A}\mathrm{MD}(a_i,M)}{|A|} + \frac{\sum_{m_i\in M}\mathrm{MD}(m_i,A)}{|M|}\right] \quad (15)$$

$$\mathrm{AMED'}(A,M) = \frac{\sum_{a_i\in A}\mathrm{MD}(a_i,M)}{|A|} \quad (16)$$

Proportional Distance, used in [15, 16], takes into account the AMED regularized with the area of the reference delineation according to equation 17

$$\mathrm{PD}(A,M) = \frac{1}{2\sqrt{\frac{Area(M)}{\pi}}}\cdot\left[\frac{\sum_{a_i\in A}\mathrm{MD}(a_i,M)}{|A|} + \frac{\sum_{m_i\in M}\mathrm{MD}(m_i,A)}{|M|}\right] * 100 \quad (17)$$

1.2 Multiple grader delineations (Study of inter- and intra-observer segmentation variability)

sec:multipleGT

Assessing the true performance of a medical imaging segmentation procedure is, at least, difficult. Although method comparison can be achieved by assessing the methodologies with a common dataset and metric, true conclusions about the performance of the segmentation are questionable. Assessing segmentations of medical images is challenging because of the difficulty of obtaining or estimating a known true segmentation for clinical data. Although physical and digital phantoms can be constructed so that reliable GT are known, such phantoms do not fully reflect clinical imaging data. An attractive alternative is to compare the segmentations to a collection of segmentations generated by expert raters.

Pons et al. [7] analyzed the inter- and intra-observer variability of manual segmentations of breast lesions in US images. In the experiment, a subset of 50 images is segmented by an expert radiologist and 5 expert biomedical engineers with deep knowledge of a breast lesion appearance in US data. The experiment reported an AOV rate between 0.8 and 0.852 for the 6 actors. This demonstrates the large variability between GT delineations; a fact that needs to be taken into account in order to draw proper conclusions about the performance of a segmentation methodology. However, having multiple GT delineations to better assess the segmentations performance is not always possible. When possible, several strategies have been used to incorporate such information.

Cui et al. [17] tested the segmentation outcome against 488 images with two delineations provided by two different radiologists. The dataset is treated as two different datasets and the performance on both is reported. Yeh et al. [23] used a reduced dataset of 6 images with 10 different delineations accompanying each image. The performance for each image was studied in terms of reward average and variation of the 10 reference delineations. Aleman-Flores et al. [16], where a dataset of 32 image dataset with 4 GT delineations provided by 2 radiologists (2 each) was available, assessed the segmentation method as if there were 128 (32×4) images.

A more elaborate idea to estimate the underlying true GT is proposed by Massich et al. [8] and Pons et al. [7]. Both works propose the use of Simultaneous Truth and Performance Level Estimation (STAPLE) in order to determine the underlying GT from the multiple expert delineations. STAPLE states that the ground truth and performance levels of the experts can be estimated by formulating the scenario as a missing-data problem, which can be subsequently solved using an Expectation Maximization (EM)

algorithm. The EM algorithm, after convergence, provides the Hidden Ground Truth (HGT) estimation that has been inferred from the segmentations provided by the experts as a probability map. Massich et al. [8] propose to assess the segmentation against a thresholded HGT and weight the AOV index with the HGT. The authors in [8] argued that apart from comparing the segmentation resulting from binarizing the greaders segmentation agreement, the amount of agreement the needs to be taken into account. This way, properly classifying a pixel with large variability within the graders produces less reward and miss classifying a pixel with great consensus penalizes.

Todo list

References

manning2005perception
1. Manning D, Gale A, Krupinski E. Perception research in medical imaging. British journal of radiology. 2005;78(932):683–685.

giger2008anniversary
2. Giger ML, Chan HP, Boone J. Anniversary paper: History and status of CAD and quantitative image analysis: the role of Medical Physics and AAPM. Medical physics. 2008;35(12):5799.

cancerStatistics2011
3. Jemal A, et al. Global cancer statistics. CA: A Cancer Journal for Clinicians. 2011;61.

smith2003american
4. Smith RA, et al. American Cancer Society guidelines for breast cancer screening: update 2003. CA: a cancer journal for clinicians. 2003;53(3):141–169.

Stavros:1995p12392
5. Stavros AT, Thickman D, Rapp CL, Dennis MA, Parker SH, Sisney GA. Solid breast nodules: Use of sonography to distinguish between benign and malignant lesions. Radiology. 1995;196(1):123–34.

biradsus
6. Mendelson E, Baum J, WA B, et al. BI-RADS: Ultrasound, 1st edition in: D'Orsi CJ, Mendelson EB, Ikeda DM, et al: Breast Imaging Reporting and Data System: ACR BIRADS – Breast Imaging Atlas. American College of Radiology; 2003.

gerard2013
7. Pons G, Martí J, Martí R, Ganau S, Vilanova JC, Noble JA. Evaluating Lesion Segmentation in Breast Ultrasound Images Related to Lesion Typology. Journal of Ultrasound in Medicine. 2013;.

massich2010lesion
8. Massich J, Meriaudeau F, Pérez E, Martí R, Oliver A, Martí J. Lesion segmentation in breast sonography. Digital Mammography. 2010;p. 39–45.

Zhang:2010p14317
9. Zhang J, Zhou SK, Brunke S, Lowery C, Comaniciu D. Database-Guided Breast Tumor Detection and Segmentation in 2D Ultrasound Images. In: SPIE Medical Imaging. vol. 7624. International Society for Optics and Photonics; 2010. p. 762405–762405.

Gao:2012p14336
10. Gao L, Liu X, Chen W. Phase- and GVF-Based Level Set Segmentation of Ultrasonic Breast Tumors. Journal of Applied Mathematics. 2012;2012:1–22.

Shan:2012p14347 11. Shan J, Cheng HD, Wang Y. Completely Automated Segmentation Approach for Breast Ultrasound Images Using Multiple-Domain Features. Ultrasound in Medicine & Biology. 2012;38(2):262–275.

Huang:2007p6100 12. Huang YL, Jiang YR, Chen DR, Moon WK. Level set contouring for breast tumor in sonography. Journal of digital imaging. 2007;20(3):238–247.

Huang:2005p11636 13. Huang YL, Chen DR. Automatic contouring for breast tumors in 2-D sonography. In: Engineering in Medicine and Biology Society, 2005. IEEE-EMBS 2005. IEEE; 2006. p. 3225–3228.

Horsch:2001p6028 14. Horsch K, Giger ML, Venta L, Vyborny C. Automatic segmentation of breast lesions on ultrasound. Medical Physics. 2001;.

Gomez:2010p14339 15. Gómez W, Leija L, Alvarenga AV, Infantosi AFC, Pereira WCA. Computerized lesion segmentation of breast ultrasound based on marker-controlled watershed transformation. Medical Physics. 2010;37(1):82.

AlemanFlores:2007p14310 16. Alemán-Flores M, Álvarez L, Caselles V. Texture-Oriented Anisotropic Filtering and Geodesic Active Contours in Breast Tumor Ultrasound Segmentation. J Math Imaging Vis. 2007;28(1):81–97.

Cui:2009p14325 17. Cui J, Sahiner B, Chan HP, Nees A, Paramagul C, Hadjiiski LM, et al. A new automated method for the segmentation and characterization of breast masses on ultrasound images. Medical Physics. 2009;36(5):1553.

hao2012combining 18. Hao Z, Wang Q, Seong YK, Lee JH, Ren H, Kim Jy. Combining CRF and multi-hypothesis detection for accurate lesion segmentation in breast sonograms. In: Medical Image Computing and Computer-Assisted Intervention–MICCAI 2012. Springer; 2012. p. 504–511.

Liu:2010p14328 19. Liu B, Cheng HD, Huang J, Tian J, Tang X, Liu J. Probability density difference-based active contour for ultrasound image segmentation. Pattern Recognition. 2010;.

Jiang:2012p14354 20. Jiang P, Peng J, Zang G, Cheng E, Megalooikonomou V, Ling H. Learning-Based Automatic Breast Tumor Detection And Segmentation in Ultrasound Images. 2012;p. 1–4.

Madabhushi:2003p6036 21. Madabhushi A, Metaxas D. Combining low-, high-level and empirical domain knowledge for automated segmentation of ultrasonic breast lesions. IEEE Transactions on medical imaging. 2003;.

Huang:2012p14313 22. Huang QH, Lee SY, Liu LZ, Lu MH, Jin LW, Li AH. A robust graph-based segmentation method for breast tumors in ultrasound images. Ultrasonics. 2012;52(2):266–275.

Yeh:2009p11985 23. Yeh C, Chen Y, Fan W, Liao Y. A disk expansion segmentation method for ultrasonic breast lesions. Pattern Recognition. 2009;.

Huang:2004p2092 24. Huang YL, Chen DR. Watershed segmentation for breast tumor in 2-D sonography. Ultrasound in Medicine & Biology. 2004;30(5):625–32.

Liu:2005p14341 25. Liu X, Huo Z. Automated Segmentation of Breast Lesions in Ultrasound Images. 2005;p. 1–3.