An optimization approach to segment breast lesions in ultra-sound images using clinically validated visual cues

**** ********⋆[1], ********* ********[1,2], **** *****[2], and ******* **********[1]

[1] **********, ********** ** *********, ** *** ** ** ********,
******* ** *******, ******;
****.*******@***********.**
[2] *******, *********** ** ******, ****** *********, ******* **, ******* ******,
*****

Abstract. Due to the fact that breast cancer remains the leading cause of cancer death among female population world-wide, any attempts to create tools for assisting radiologists during the diagnosis process are necessary. However, most of the technologies developed in the imaging laboratories are expelled from this process because these technologies are not based on the existing findings and procedures undertaken by the clinicians. To address this issue, and to develop Computer Aided Diagnosis (CAD) systems that take advantage of this clinical findings already adopted by clinicians, these systems only lack strategies to accurately delineate the breast structures and its lesions. Therefore, this article proposes a highly modular and flexible framework for segmenting breast tissues and lesions present in Breast Ultra-Sound (BUS) images using optimization strategies and high-level descriptors designed analogously to the visual cues used by radiologists.

Keywords: Breast Ultra-Sound, BI-RADS lexicon, Optimization based Segmentation, Machine-Learning based Segmentation, Graph-Cuts

1 Introduction

Breast cancer is the second most common cancer. In terms of mortality, breast cancer is the fifth most common cause of cancer death. However, it is ranked as the leading cause of cancer death among females in both western countries and economically developing countries [4].

Medical imaging contributes to its early detection through screening programs, non-invasive diagnosis, follow-up and suchlike procedures. Despite Breast Ultra-Sound (BUS) imaging not being the imaging modality of reference for breast cancer screening [8], Ultra-Sound (US) imaging has more discriminative

⋆ **** **** *** ********* ********* ** *** ******** ******* ** ********** **
*** ******* ********* *** ***** **.*******_****_***_**

power compared to other image modalities to visually differentiate benign from malignant solid lesions [9]. In this manner, US screening is estimated to be able to reduce between 65 $\sim$ 85% of unnecessary biopsies [10], in favour of a less traumatic short-term screening follow-up using BUS images. As the standard for assessing this BUS images, the American College of Radiology (ACR) proposes the Breast Imaging-Reporting and Data System (BI-RADS) lexicon for BUS images [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 this document in Sect. 3, 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 lexicon are 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. Therefore, developing accurate segmentation methodologies for breast lesions and structures are crucial to take advantage of this already validated clinical tools.

This article proposes a highly modular and flexible framework for segmenting lesions and tissues present in BUS images. The proposal takes advantage of an energy-based strategy to perform segmentation based on discrete optimization using super-pixels and a set of novel features analogous to the elements encoded by the US BI-RADS lexicon [6].

2 Description of the segmentation methodology

Optimization methodologies offer a standardized manner to approach segmentation by minimizing an application-driven cost function [2]. Figure 1 illustrates a generic representation of the segmentation strategy here adopted to delineate breast tissues or lesions in US images. The overall segmentation can be seen as a three-steps strategy: (1) a mapping of the image into a discrete set of elements $\mathcal{S}$, (2) the optimization stage which is formulated as a *metric labelling* problem, and (3) a re-mapping the labels obtained from the previous stage to produce the final delineation.

In order to formulate the segmentation like a metric labelling problem, the image is conceived as a discrete set of elements $\mathcal{S}$ that need to be labelled using a label l from the labelling set $\mathcal{L}$. Let $\mathcal{W}$ be all the possible labelling configurations of the set $\mathcal{S}$, given $\mathcal{L}$. Let $U(\cdot)$ be a cost function encoding the goodness of the labelling configuration $\omega \in \mathcal{W}$ based on the appearance of the elements in $\mathcal{S}$, their inner relation and some designing constraints. Then, the desired segmentation

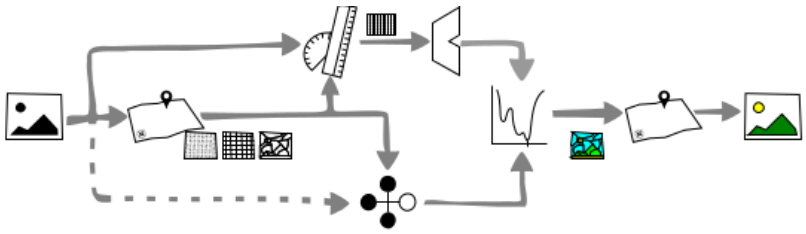

Fig. 1: Conceptual block representation of the segmentation methodology

$\hat{\omega}$ corresponds to the labelling configuration that minimizes this cost function, as described in Eq. (1).

$$\hat{\omega} = \arg\min_{\omega} U(\omega) \tag{1}$$

It is worth to mention here that not all the minimization strategies are applicable or adequate to find $\hat{\omega}$. The convenience of a particular minimization strategy is determined by the nature of $U(\cdot)$ and $\mathcal{W}$ (see Sect. 2.3).

This goodness measure $U(\cdot)$ must be defined to take into account the appearance of the target region, its relation with other regions and other designing constraints. Equation (2) describes this cost function as the combination of two independent costs that need to be simultaneously minimized as a whole.

$$U(\omega) = \sum_{s \in \mathcal{S}} D_s(\omega_s) + \sum_{s \in \mathcal{S}} \sum_{r \in \mathcal{N}_s} V_{s,r}(\omega_s, \omega_r) \tag{2}$$

Where, the left hand side of the expression integrates the so-called *data* term, while the right hand side integrates the *pairwise* term, which is also referred as the *smoothing* term. Both terms are shaped by $\mathcal{S}$ and evaluated in the labelling space $\mathcal{W}$. Figure 2 uses the problem of delineating the tissues present in a BUS image to represent the working principles of the data and pairwise terms in Eq. (2). In general, $\mathcal{S}$ can be any discrete set representing the image (i.e. pixels, overlapping or non overlapping windows, super-pixels, etc.). Figure **??** illustrates one such representation $\mathcal{S}$, applied to a BUS image example using super-pixels. The super-pixels are coloured according to the image's Ground Truth (GT).

In our quest to optimize the cost function $U(\cdot)$, it is required to define a representation for the set $\mathcal{S}$, a data term $D(\cdot)$, a pairwise term $V(\cdot)$, and a proper minimization methodology.

2.1 The data term

Given a label configuration $\omega \in \mathcal{W}$, the data term penalizes the labelling of a particular image element or site ($\omega_s = l$) based on the data associated to s. In this manner, $D_s(\omega_s = l_{\checkmark}) << D_s(\omega_s = l_{\times})$. Figure **??** illustrates the data cost associated to some arbitrary labelling configurations to clarify the desired effect (or behaviour) of this data term (Fig. **??** shows the GT of each site s).

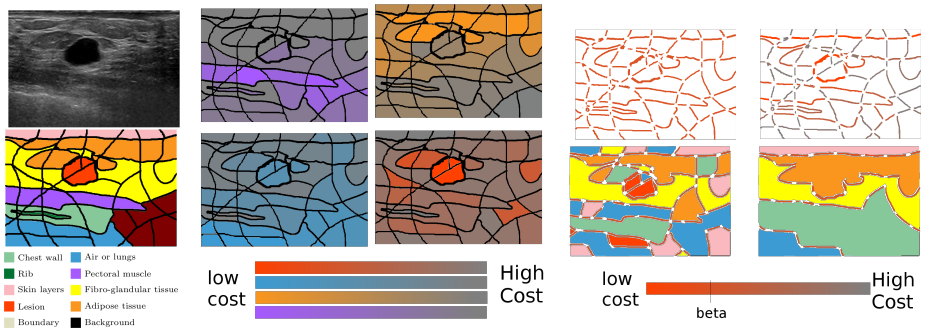

Fig. 2: Methodology: (a) Problem definition, (b) Data term, (c) Pairwise term.

Designing an obscure heuristic to comply with the desired behaviour of $D(\cdot)$ out of the box, is rather a complicated task. Therefore, an easier and cleaner approach is to take advantage of Machine Learning (ML) techniques to design this data cost in a systematic manner based on a training stage. The idea is to generate a data model for each label (or class) in $\mathcal{L}$ from training samples, and let $D(\cdot)$ be a distance or goodness measure reflecting the likelihood for s to belong to class l. ML offers a systematic but flexible approach to customize $D(\cdot)$. Different features to represent the data, custom construction of the data models by using different classifiers, training techniques or including arbitrary constraints; can be used to achieve the desired data term without changing the overall scheme.

Defining the data term $D(\cdot)$ with the help of ML follows a systematic process. For each site $s \in \mathcal{S}$, features describing s have to be detected. Then, different optional steps can be applied to this set of feature: (i) features normalization, (ii) features selection or (iii) features extraction. Finally, the data term $D(.)$ is encoded based on ML classifiers, the features and a training step. This representation is in fact depicted in the upper row in Fig. 1.

2.2 The pairwise term

The pairwise term represents the cost associated to ω_s taking into account the labels of its neighbour sites, ω_r, $r \in \mathcal{N}_s$. This term models a Markov Random Fields (MRFs) or a Conditional Random Fields (CRFs). The typical form of this term, given in Eq. (3), is called homogenization which acts as a regularization factor favouring configurations that have a coherent labelling.

$$V_{s,r}(\omega_s, \omega_r) = \begin{cases} \beta, & \text{if } \omega_s \neq \omega_r \\ 0, & \text{otherwise} \end{cases} \tag{3}$$

Figure **??** offers a visual interpretation of this cost. The more fragmented is the segmentation ω, the higher the overall pairwise term; since every boundary

brings a penalization β to the total cost $U(\omega)$. In this manner the regularization term can be seen as a post-processing or denoising stage since that some sites will flip their labelling if the cost of fragmenting the regions is larger than the cost of adopting their neighbour's label.

2.3 Searching the best labelling configuration

Once defined $U(\omega)$ so that the cost for a particular labelling configuration ω can be computed, the problem of finding $\hat{\omega}$ corresponding to the global minimum of the space $\mathcal{W}$ of all possible labelling configurations needs to be faced.

3 This generic framework applied to BUS imaging

In this section, the generic framework presented in the previous section is applied to BUS imaging. We need to define our problem properly. We recall that the aim in segmentation is to affect to a discrete set of elements $\mathcal{S}$, a label l from a labelling set $\mathcal{L}$. In our case, our labelling set $\mathcal{L} = \{\text{chest wall}, \text{lungs}, \ldots, \text{lesion}\}$ (see Fig **??** for the entire set of labels) and the set $\mathcal{S}$ is chosen to be a super-pixels representation of the image [1]. In our case, $\mathcal{S}$ is the result from an over-segmentation of the image using Quick-shift super-pixel. Figure **??** illustrates one such representation $\mathcal{S}$, applied to a BUS image example. The super-pixels are coloured according to the image's GT. Bear in mind that given an unseen BUS image, the ultimate goal is to represent the image as a set of super-pixels and infer the appropriated labelling for each of them.

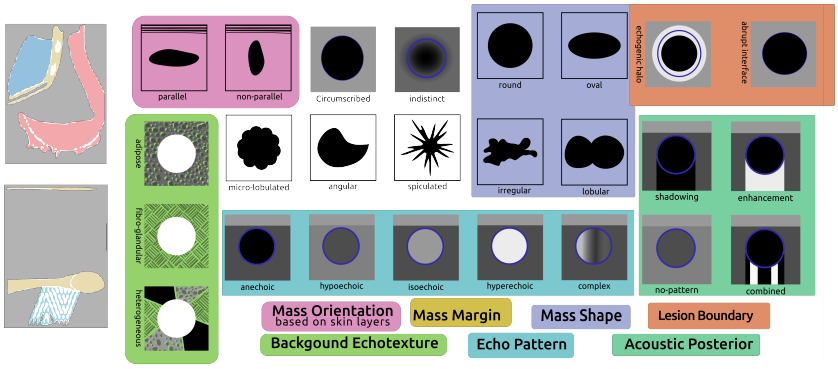

Fig. 3: Visual reference: (a) breast structures, (b) US BI-RADS lexicon, (c) encoded visual cues.

In this section, the generic framework presented in the previous section is applied to BUS imaging. We need to define our problem properly. We recall that the aim in segmentation is to affect to a discrete set of elements $\mathcal{S}$, a label l from

a labelling set $\mathcal{L}$. In our case, our labelling set $\mathcal{L} = \{\text{chest wall}, \text{lungs}, \ldots, \text{lesion}\}$ (see Fig **??** for the entire set of labels).

As illustrated in Fig. ??, choices have to be made regarding the different elements: the set $\mathcal{S}$, the data term $D(\cdot)$, the pairwise term $V(.)$, and the optimizer choice. These preferences are summarized in Table ?? and justified thereafter.

$\mathcal{S}$ is chosen to be the result from an over-segmentation of the input image using the Quick-shift super-pixel. The structures of the breast and their rendering when using a hand-held 2D US probe are sketched in Fig. 3. Figure **??** illustrates the lexicon proposed by the ACR [6] and used by clinicians to perform their diagnosis. Thus, our aim is to generate a set of computer vision features which is able to encode the characteristic described in the lexicon. The selected features are the following:

Appearance Based on the multi-labelled GT, a Median Absolute Deviation (MAD) histogram model for every tissue label is built. The Appearance feature is computed as the χ^2 distance between a histogram of s and the models generated.

Atlas Based on the multi-labelled GT an atlas is build to encode the labels likelihood based on the location of s.

Brightness Intensity descriptors are computed based on statistics of s (*i.e:* mean, median, mode) and are compared with some intensity markers of the set $\mathcal{S}$ such as the minimum intensity value, the maximum, its mean, etc.

Self-Invariant Feature Transform (SIFT)-Back-of-Features (BoF) s is described as an histogram of visual words based on SIFT [5]. The dictionary is built with 36 words.

The relationships between the lexicon and the descriptors previously described is depicted in Table ??. More precisely, we highlight the corresponding elements of the lexicon which is encoded by each feature. A choice regarding the encoding of the data term $D(\cdot)$ has to be made by using a ML classifier. An Support Vector Machines (SVM) classifier with an Radial Basis Function (RBF) kernel is selected to determine the data model during the training stage. The pairwise term is our framework was defined as in Eq. (3). The optimization method used as solver to minimize our cost function $U(\cdot)$ is Graph-Cuts (GC). GC when applicable allows to rapidly find a strong local minima guaranteeing that no other minimum with lower energies can be found [3]. GC is applicable if, and only if, the pairwise term favours coherent labelling configurations and penalizes labelling configurations where neighbours labels differs; such is our case, given by Eq. (3).

4 Method evaluation and comparison

A 16 BUS images dataset with accompanying multi-label GT delineating all the structures present in the images has been used to evaluate the proposed methodology for lesion segmentation application. Every image in the dataset presents a single lesion with variable extension. The size of the lesions ranges

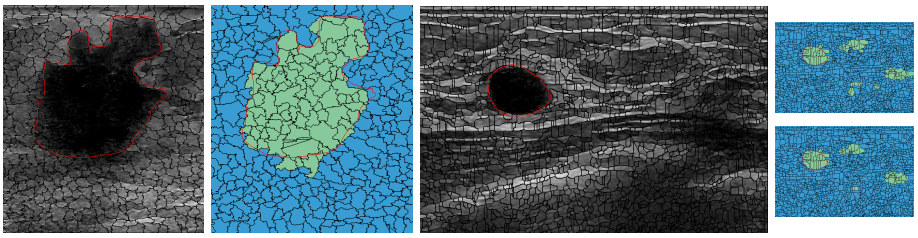

Fig. 4: Qualitative results. (a) Example 1: orignal image, super-pixels' delineations and GT. (b) Differences between GT and the delination resulting from super-pixels' boundary. (c) Ex. 2. (d) weak $V(\cdot,\cdot)$ (e) strong $V(\cdot,\cdot)$

from under 1/100 to over 1/5 of the image. The dataset composed of cysts, Fibro-Adenomas (FAs), Ductal Inflating Carcinomas (DICs) and Inflating Lobular Carcinomas (ILCs).

Figure 4 shows qualitative results, whereas the quantitative results from the best configuration are reported as a table in Fig. **??**. Notice that for the proposed framework, the performance in terms of Area Overlap (AOV) is limited by the capacity of the super-pixels to snap the desired boundary. Figure 4 shows how the delineation resulting from a proper labeling of the super-pixels differs from the GT.

Due to the lack of publicly available data (and code) there is no manner to perform a methodology comparison further than compiling the results reported in the literature. Fig. 5 resumes some of the methodologies found in the literature and traslates the reported results into AOV for comparison purposes. The iconography used in Fig. **??** relates methodology stages, technology used and if the stages have been treated separately or as a single stage.

Figure **??** arrange the information for direct visual comparison. An extra element is also represented in Fig. **??** as blue swatch delimited by two blue dashed lines. The boundaries of this swatch correspond to performance of expert radiologists in terms of AOV based on an inter- and intra-observer experiment carried out by Pons et al. [7].

When comparing the results it is clear the inconvenience of unexciting public data, since several of the results outperform the manual delineations studied in [7]. It can also be seen that the category tested in larger datasets is ML, whereas Active Contour Model (ACM) lead to better segmentations since the lesion boundary is easier to model in ACM compared to ML based techniques.

5 Conclusions

This work presents a segmentation strategy to delineate lesions in BUS images using an optimization framework that takes advantage of all the facilities available when using ML techniques. Despite the limitation that the final segmentation is subject to the super-pixels' boundaries, the AOV results here reported are similar to those reported by other methodologies in the literature. A higher

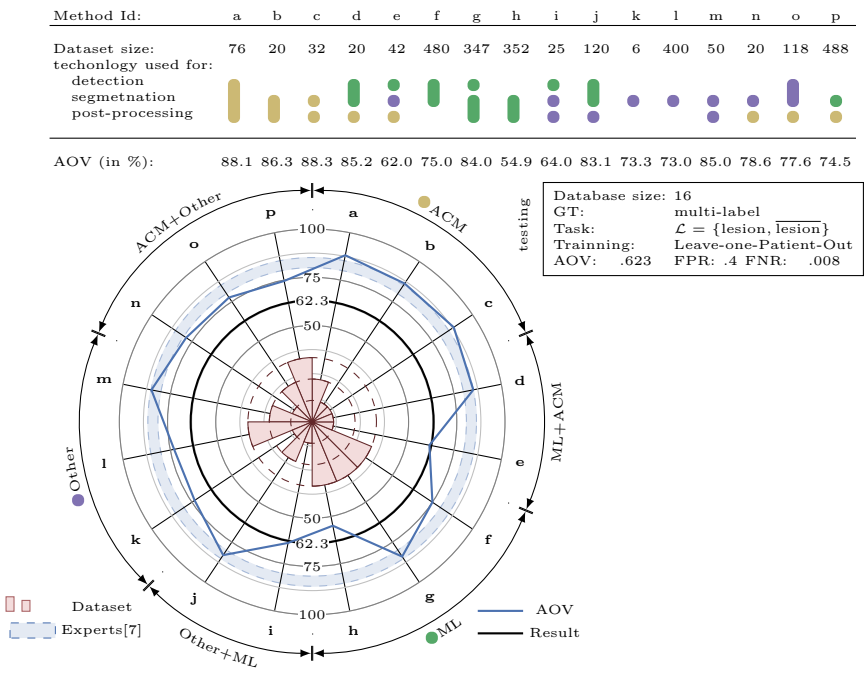

Fig. 5: Quantitative results compilation and comparison

AOV result can be achieved by deforming the delineation resulting from the proposed framework using this to initialize a second post-processing step based on ACM. In this manner the contour constrains could be applied to achieve a more natural delineation.

References

1. Achanta, R., et al.: SLIC superpixels compared to state-of-the-art superpixel methods (2012)
2. Cremers, D., Rousson, M., Deriche, R.: A review of statistical approaches to level set segmentation: integrating color, texture, motion and shape. International journal of computer vision 72(2) (2007)
3. Delong, A., Osokin, A., Isack, H.N., Boykov, Y.: Fast approximate energy minimization with label costs. International Journal of Computer Vision 96(1), 1–27 (2012)
4. Jemal, A., et al.: Global cancer statistics. CA: A Cancer Journal for Clinicians 61 (2011)
5. Massich, J., et al.: Sift texture description for understanding breast ultrasound images. In: Breast Imaging, Lecture Notes in Computer Science, vol. 8539, pp. 681–688. Springer International Publishing (2014)
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)
7. Pons, G., Martí, J., Martí, R., Ganau, S., Vilanova, J., Noble, J.: Evaluating lesion segmentation in breast ultrasound images related to lesion typology. Journal of Ultrasound in Medicine (2013)
8. Smith, R.A., et al.: American cancer society guidelines for breast cancer screening: update 2003. CA: a cancer journal for clinicians 53(3), 141–169 (2003)
9. Stavros, A.T., Thickman, D., Rapp, C.L., Dennis, M.A., Parker, S.H., Sisney, G.A.: Solid breast nodules: Use of sonography to distinguish between benign and malignant lesions. Radiology 196(1), 123–34 (1995)
10. Yuan, Y., Giger, M.L., Li, H., Bhooshan, N., Sennett, C.A.: Multimodality computer-aided breast cancer diagnosis with ffdm and dce-mri. Academic radiology 17(9), 1158 (2010)