

**** *****1, ***** *****1,2, **** *****2, and ***** *****1
 1 *****

 *****@*****
 2 *****

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

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 [5].

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 [9], Ultra-Sound (US) imaging has more discriminative

```
* **** *
***** ***** **
*****
*****
```

power compared to other image modalities to visually differentiate benign from malignant solid lesions [10]. In this manner, US screening is estimated to be able to reduce between 65 ~ 85% of unnecessary biopsies [11], 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 [7]. 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 [7].

2 Description of the segmentation methodology

Optimization methodologies offer a standardized manner to approach segmentation by minimizing an application-driven cost function [3]. Figure 1 illustrates a generic representation of the segmentation strategy, whereas concrete examples of its terms can be found in section 3 applied to BUS. 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

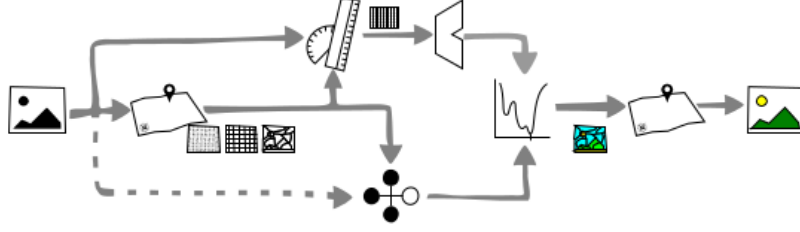


Fig. 1: Conceptual block representation of the segmentation methodology.

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

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

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.

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

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} . 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.

The set \mathcal{S} can be, in general, any discrete set representing the image (i.e. pixels, overlapping or non overlapping windows, super-pixels, etc.).

The data term $D(\cdot)$, given a label configuration $\omega \in \mathcal{W}$, penalizes the labelling of a particular image element or site ($\omega_s = l$) based on the data associated to s . In this manner, $D_s(\omega_s = l_{\checkmark}) \ll D_s(\omega_s = l_{\times})$. Figure 2b illustrates the data cost associated to some arbitrary labelling configurations to clarify the desired effect (or behaviour) of this data 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 design this data term $D(\cdot)$ with the help of Machine Learning (ML) because it provides a systematic process that is flexible enough to encode any desired behaviour based on a training stage. This concept is in fact depicted in the upper row in Fig. 1. For each site $s \in \mathcal{S}$, features

² Figure 2 illustrates the \mathcal{S} , $D(\cdot)$, and $V(\cdot)$ for the applied case of delineating breast structures in US data. (final version the figure elements should be labelled)

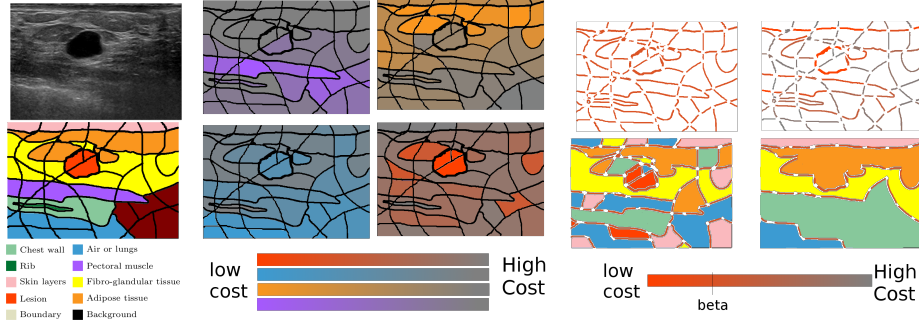


Fig. 2: Methodology: (a) Problem definition, (b) Data term, (c) Pairwise term.

describing s are designed. 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 $D(\cdot)$ is encoded based on ML classifiers, the features and a training step. Thus, the data term $D(\cdot)$ can be seen as a distance or goodness measure reflecting the likelihood for s to belong to class l .

The pairwise term $V(\cdot, \cdot)$ 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.

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

Figure 2c shows 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.

The minimization strategy is determined by the nature of $U(\cdot)$ and \mathcal{W} , since not all the minimization strategies are applicable or adequate to find $\hat{\omega}$. The size of the labelling space $|\mathcal{W}| = |\mathcal{L}|^{|\mathcal{S}|}$, discontinuities in $U(\omega)$ due to \mathcal{W} or the problem of local minima, along with all the particular constrains of all the different minimization methodology. Need to be taken into account while choosing the most desirable minimization strategy.

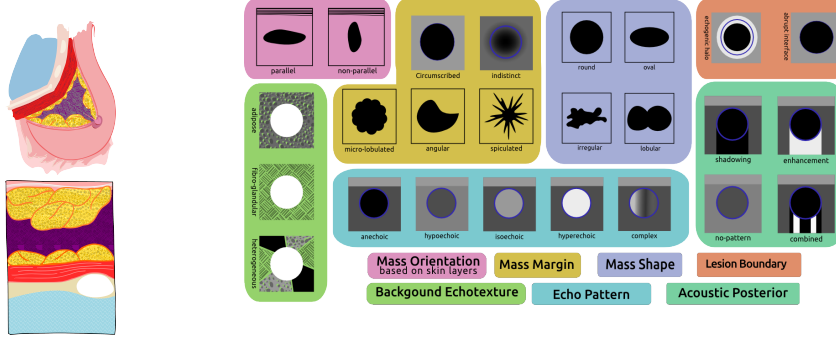


Fig. 3: Visual reference: (a) breast structures, (b) US BI-RADS lexicon

3 BUS images segmentation using optimization

In this section, the problem of delineating structures in BUS images is defined as an optimization problem that can be solved applying the framework presented in Sect. 2. The segmentation here proposed aims at tying a label $l \in \mathcal{L}$ (*i.e.* $\{\text{lesion}, \overline{\text{lesion}}\}$ or $\{\text{chest wall}, \text{lungs}, \dots, \text{lesion}\}$) to each element of \mathcal{S} by simultaneously optimizing the data and pairwise terms as illustrated in Fig. 2. Choices made regarding these different elements: the representation \mathcal{S} , the data term $D(\cdot)$, the pairwise term $V(\cdot)$, and the optimizer choice are summarized in Table 1 and justified thereafter (see Fig. 1 for reference).

\mathcal{S} is considered the result from an over-segmentation of the image using Quick-shift super-pixel [2]. The structures of the breast and their rendering when using a hand-held 2D US probe are sketched in Fig. 3a. Figure 3b illustrates the lexicon proposed by the ACR [7] 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 Ground Truth (GT), a Median Absolute Deviation (MAD) histogram model for every tissue label is built. The

Table 1: Design choices summary

\mathcal{S}	Quick-Shift super-pixels		
	Appearance	SIFT-BoW	Atlas, Brightness
$D(\cdot)$			
$V(\cdot, \cdot)$	homogeneity as Eq. (3)		
$\arg \min U(\cdot)$	Graph-Cuts		

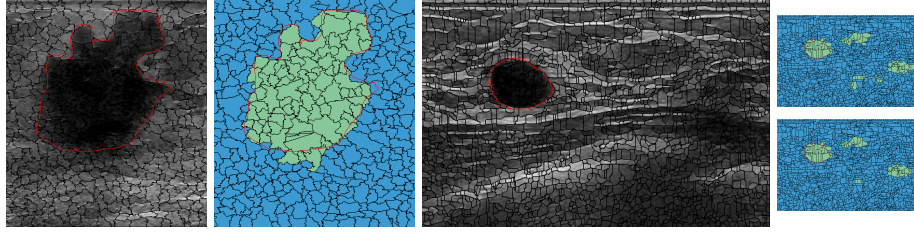


Fig. 4: Qualitative results. (a) Example 1: original image, super-pixels' delineations and GT. (b) Differences between GT and the delineation resulting from super-pixels' boundary. (c) Ex. 2. (d) weak $V(\cdot, \cdot)$ (e) strong $V(\cdot, \cdot)$

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 [6]. The dictionary is built with 36 words.

The relationships between the lexicon and the descriptors previously described is depicted in Table 1. 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 in 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 [4]. 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 a variable extension. The size of the lesions ranges from under 1/100 to over 1/5 of the image size. The dataset is composed of cysts, Fibro-Adenomas (FAs), Ductal Infiltrating Carcinomas (DICs) and Infiltrating Lobular Carcinomas (ILCs). This dataset is now publicly available at http://*****

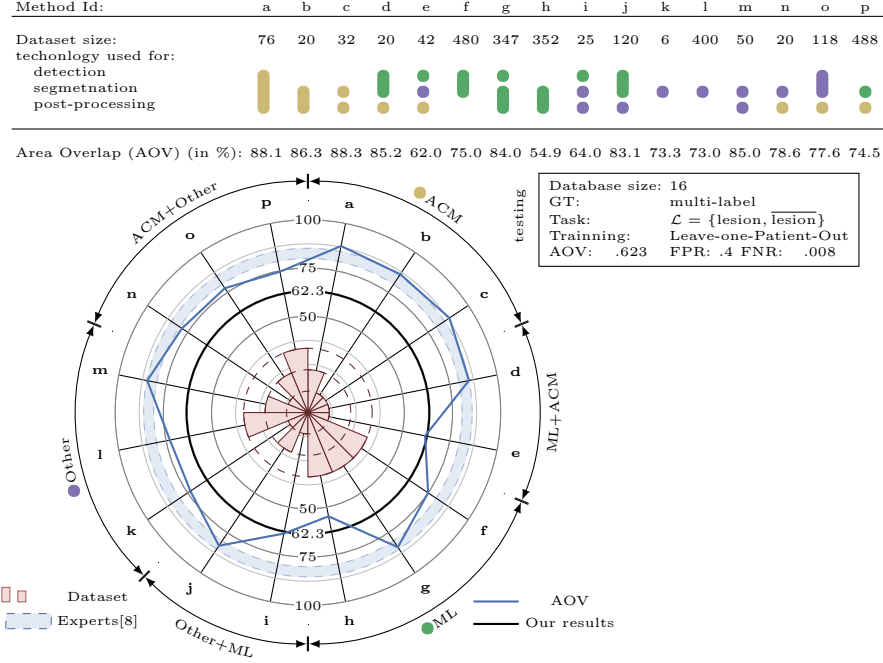


Fig. 5: Quantitative results compilation and comparison

Due to the lack of publicly available dataset and source code, a comparison between the different methods can only be limited to a compilation of results provided by the different authors on their own data. ***** *et al.* [1] compile such information in terms of AOV as depicted in Fig. 5.

Figure 5 is divided in three main parts: (i) a table on the top summarizes the core stages of each study framework, (ii) a legend box on the right side informing of our testing setup and (iii) a comparison of the different metrics in a radial manner. An extra element is also represented in this radial representation: a blue swatch delimited by two blue dashed lines. The boundaries of this swatch correspond to the performance of some expert radiologists in terms of AOV based on an inter- and intra-observer experiments carried out by Pons *et al.* on the same dataset used in our experiments [8].

It is interesting to note that the performance of some of the state-of-the-art methods outperform the performance of our radiologists. A publicly available dataset should allow a better comparison in that regard. Furthermore, ML methods have been developed in conjunction with larger dataset, whereas Active Contour Model (ACM) methods lead to a better segmentation due to the capability of these methods to better the lesion boundary.

Notice that for our proposed framework, the performance in terms of AOV is limited by the capacity of the super-pixels to snap the desired boundary. Figure 4

