Combining CRF and Multi-hypothesis Detection for Accurate Lesion Segmentation in Breast Sonograms

Zhihui Hao, Qiang Wang, Yeong Kyeong Seong,
Jong-Ha Lee, Haibing Ren, and Ji-yeun Kim

Samsung Advanced Institute of Technology (SAIT), Samsung Electronics

Abstract. The implementation of lesion segmentation for breast ultrasound image relies on several diagnostic rules on intensity, texture, *etc.* In this paper, we propose a novel algorithm to achieve a comprehensive decision upon these rules by incorporating image over-segmentation and lesion detection in a pairwise `CRF` model, rather than a term-by-term translation. Multiple detection hypotheses are used to propagate object-level cues to segments and a unified classifier is trained based on the concatenated features. The experimental results show that our algorithm can avoid the drawbacks of separate detection or bottom-up segmentation, and can deal with very complicated cases.

1 Introduction

Breast cancer is the second leading cause of cancer death for women. Currently, early detection is the only solution to reduce the death rate. Ultrasonography is widely used in the diagnosis and observation of breast abnormality because of the convenience, safety and high accuracy rate [1]. However, it is also widely acknowledged that ultrasound image interpretation is highly reliant on medical expertise. Designing a computer-aided system to assist ultrasound practitioners in recognizing lesion and delineating the boundary is becoming necessary.

A typical breast ultrasound image is shown in the upper left of Fig. 1. Generally, object segmentation in ultrasound images is much more difficult than that in natural images due to the following aspects: poor quality of the image with low contrast and heavy speckle noise; large variation of lesion in shape and appearance, especially between the benign and the malignant; existence of similar tissues or acoustic shadows; irregular and poorly defined lesion boundaries.

In this paper, we put forward an automatic lesion segmentation algorithm. Unlike most previous works which solve the problem by translating the diagnostic rules into computer language term by term, we propose to achieve it in an integrated framework of all image cues. To this end, features from segments and multiple lesion detection hypotheses are combined together to train a single classifier. The lesion segmentation is then accomplished by optimizing a segment based `CRF` model. Fig. 1 gives an overview of our approach.
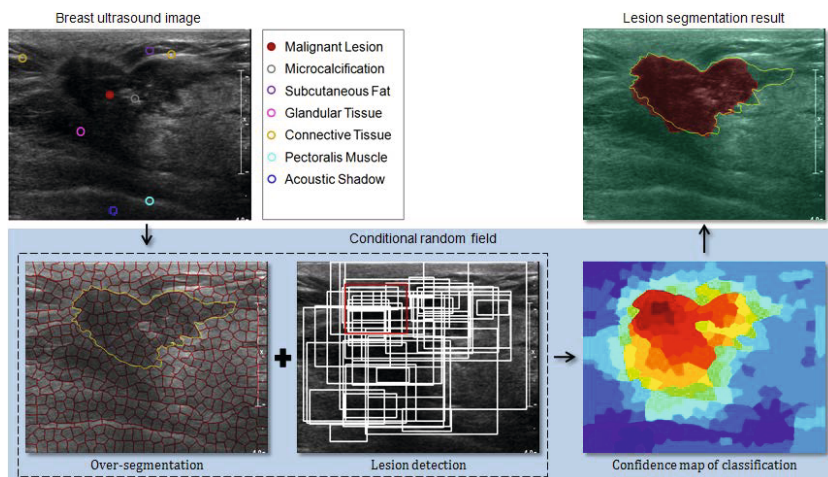
Fig. 1. An overview of the system. The process is depicted anticlockwise. The yellow contours show the groundtruth of lesion. Rectangles in the lower middle image show the detection windows, where the red one has the maximal confidence.

2   Motivations

A sound wave is sent by the sonographic transducer into the human breast, absorbed in or scattered from tissues and structures in it. The reflected wave is captured and processed into a sonogram by the ultrasonic instrument. Intensive research has been done in both fields of radiology and biomedicine [2] to distinguish lesions (both the benign and the cancerous) in ultrasound images from normalities and shadowing artifacts.

The diagnostic criteria can be generalized into the following terms [1]. First, the different echogenicity that nodule and the surrounding area show. A portion of fibrous lesions are hyperechoic with respect to isoechoic fat, while another portion of benign lesions and most of the malignant are markedly hypoechoic. And also, distinguishable internal echotexture can be observed in many cases. Second, the border and the shape of nodule. Benign nodules usually have a thin echogenic pseudocapsule with an ellipsoid shape or several gentle lobulations, and malignant nodules could show radially with spiculations and angular margins. Third, the position of the nodule. Most lesions appear in the middle mammary layer and shadows are produced under the nodules.

These criteria have been translated into computer vision language in many different ways for the design of computer-aided diagnosis system [2]. In [3], Madabhushi and Metaxas build probability distribution models for intensity and echotexture of lesion, based on which they estimate the seed point followed by a region growing procedure. To eliminate the spurious seeds, spatial arrangement together with other rules are then used. At last, the boundaries are located and shaped successively. In [4], Liu *etc.* divide the image into lattices and classify

them based on texture descriptors. After that, they have to use rules like "lesions are more likely to be around the image center and occupy larger areas" to select the true regions of interest. Another common idea is towards deformable shape modeling, for example, the level set method [5]. The interior texture information can also be incorporated during the model deformation [6]. This type of algorithms usually requires a careful initialization and thus cannot avoid the selection of seed region.

From the perspective of an ultrasound practitioner, however, the recognition (and the delineation) of breast lesion in sonogram is perhaps not a product of an assembly line, which consists of several small and inaccurate rules with unclear relationships. Instead, it should be an integrated decision upon these rules. The importance of each rule is learned and adjusted through training and practicing with a large number of samples. A notable work comes from Siemens researchers recently [7]. They train discriminative models of both texture and boundary, and combine them in the framework of Markov random field. However, they still treat lesion detection and segmentation as two steps of one problem, and leave the situation that detection fails undiscussed.

We found that in practical applications, even the state-of-the-art lesion detector such as the deformable part based detection algorithm [8] still can not provide perfect detection result. So in this paper, we propose a novel algorithm to achieve a joint optimization of lesion detection and segmentation. Empirical rules are implicitly contained in different units. Specifically, rules about shape and structure are in the lesion detector, boundary and texture are in the image over-segmentation and the segment classifier, and position cues are in both of the sliding windows and segments.

The study on the combination of detection and segmentation is not new for computer vision community. In [9], Gao *et al.* augment the bounding box with internal cells to enrich the representation ability. Although outputting tighter masks, their motivation is still to improve the detection results by handling the problem of object occlusion. The most similar work to ours is [10], where bounding boxes are enforced as higher order potentials on the conditional random filed (CRF) model of image pixel. Detection hypothesis could be accepted or rejected depends on its agreement with confidences from pixels and segments. The difference of this paper is, we use over-redundant outputs of detection without considering their validity. We believe that many detection windows more than the true positive can provide higher level cues to the area they cover. These cues are collected and retrained with features from segments into a unified classification model, and thus the final segmentation is able to avoid the mistake which possibly occurs in lesion detection.

3   Algorithm

We define the problem of lesion segmentation on a pairwise CRF model. Let $\mathcal{X}$ denote the set of random variable which takes the label of either lesion or not, and $\mathcal{E}$ denote the set of edge which connects each pair of nodes. A typical pairwise

CRF is modeled as the sum of a unary potential ψ and a pairwise potential ϕ, and minimizes the energy function with the form of

$$E(\mathcal{C}|\mathcal{X}) = \sum_{x_i \in \mathcal{X}} \psi(c_i|x_i) + \mu \sum_{(x_i,x_j) \in \mathcal{E}} \phi(c_i, c_j|x_i, x_j), \qquad (1)$$

where c_i denotes the label (*i.e.* lesion or non-lesion) of the node x_i.

The node in the `CRF` model could correspond to a pixel or a segment in the image. We use the segment here. The over-segmentation tool we have used is a hierarchical method proposed by G. Mori [11]. The method starts with a normalized cut [12], and then builds the following layers iteratively by applying the `k-means` clustering on the premise of respecting the existing boundaries. The normalized cut tends to produce roughly equal size of patches and the `k-means` algorithm ensures a low internal variation of intensity. In this way, we obtain a set of segments with a multiple-layer structure. Features from different layers are collected together to enhance the representation ability. Other less time-consuming methods could be used for over-segmentation as well, for example, the quick shift [13], but the system performance would deteriorate with the loss of larger-scale information. We will show the related experiments in section 4.

The unary potential of the `CRF` model is defined based on the response of features of segment. Before solving the `CRF` problem, we will discuss first how the features are collected, especially from lesion detection.

3.1   Lesion Detection and Feature Propagation

When the image is broken into segments, some critical information of the lesion such as the shape or the context is possibly lost and can hardly be retrieved with this bottom-up fashion. Then the object detector becomes a complementary tool which is capable of providing higher-level supports. For this reason, we introduce the deformable part model (`DPM`) [8], one of the most successful detectors currently.

The deformable part-based detector produces over-complete sliding windows for potential lesion area. Many of them are retained after non-maximum suppression. Then, the `MAP` estimation is usually applied to select the window with maximum confidence and discard all of the others. For ultrasound image, the problem is becoming more difficult because of the existence of similar tissues and artifacts, which makes the detection confidence less reliable. So it is risky to select one of the windows as the bounding box of the lesion [7] before investigating the interior.

On the other hand, however, it is also difficult to re-rank the windows based on the ambiguous local cues. Here we propose a new mechanism to avoid the comparison. We propagate the information provided by these rectangular windows to amorphous segments. Specifically, the feature received by the segment x_i includes 4 pieces, *i.e.* $\mathbf{f}(x_i) = [\mathbf{f}_{rect}(x_i), \mathbf{f}_{dist}(x_i), \mathbf{f}_{prop}(x_i), \mathbf{f}_{acc}(x_i)]$.

$\mathbf{f}_{rect}(x_i) = [g(s_i), r(s_i), g(s_m), r(s_m)]$ records the original information of rectangles. s_i is the maximal confidence among the windows that cover x_i, and
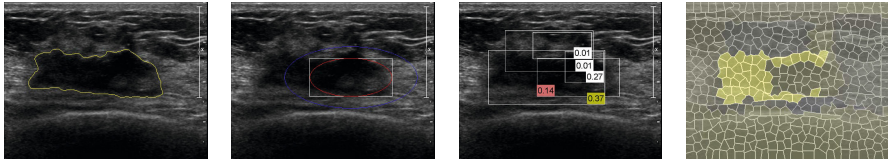
(a) Ultrasound im- (b) Foreground and (c) Detection win- (d) Score map of
age and groundtruth background masks   dows with scores   segments

Fig. 2. An illustration of how the intensity contrast score is computed and propagated to segments. 5 windows are shown in (c) where clearly, the window with the maximal score (in yellow) is not the one with the maximal detection confidence (in red).

$g(s_i) = 1/(1 + \exp(-2s_i))$. For ease of exposition, we call the detection window with s_i as the proxy of x_i. $r(s_i)$ contains the position and the size of the proxy, which are regularized by the format of the image. s_m is the maximal score in the full image.

$\mathbf{f}_{dist}(x_i) = [d(s_i), d(s_m)]$ records the distances from the segment to its proxy and the most supported window. Since tissues in windows usually appear as dense nodules, a segment around the center is more likely to inherit the property of the window. $d(s_i) = 2/(1 + \exp(2t))$ and $t = max(t_x, t_y)$, where t_x, t_y are the distances along x and y axes and regularized by the size of the window.

$\mathbf{f}_{prop}(x_i)$ contains the extended properties of detection window. Here we introduce the intensity contrast to measure the dissimilarity of a window to its surrounding area. The score equals to the Chi-square distance between their histograms of intensity. Different with [14], we modify the area masks from rectangle to ellipse and again pass on these scores from proxy windows to segments. See Fig. 2 as an example. Note that other objectness measures in [14] can be readily used as well.

$\mathbf{f}_{acc}(x_i)$ measures the total strength of detection confidence and intensity contrast score of all the windows covering x_i, and then regularized by the maximal value in the current image.

3.2   Problem Solving

We also extract the following features from segments: their positions in the image, histograms of intensity and texture descriptors derived from grey-level co-occurrence matrix [15]. These features and those propagated from detections are concatenated to train a segment classifier.

Finally, we solve the lesion segmentation problem by optimizing Eq. 1. The unary potential ψ in CRF model is defined based on the probability given by the output of the segment classifier: $\psi(c_i|x_i) = -\log(P(c_i|x_i))$. The pairwise potential ϕ is defined as

$$\phi(c_i, c_j|x_i, x_j) = \exp\left(-\frac{\|h_i - h_j\|^2}{2\sigma^2}L(x_i, x_j)\right)\delta(c_i \neq c_j), \qquad (2)$$

where $\|h_i - h_j\|$ is the distance between two histograms of intensity, and $L(x_i, x_j)$ is the strength of the shared boundary, which is set to infinity when x_i and x_j are not contiguous. $\delta(\pi)$ is a bool expression which takes 1 when π holds and 0 otherwise. The problem of minimizing Eq. 1 could be solved by using the popular min-cut/max-flow algorithm.

4   Experiments

We collect 480 breast ultrasound images to evaluate the proposed algorithm. All the ultrasound images are grayscale, produced by the ultrasound machine Philips IU22. 320 images are randomly selected for training, the other 160 for testing, but both of them covers all kinds of lesions presented in the full dataset. Lesion boundaries are delineated by experienced ultrasound practitioners.

We create about 400 segments for each image. The segment classifier is trained by support vector machines with RBF kernels. For the DPM detector, the number of parts is set to 8 and the minimal size of part template is 6×6. From several to dozens of detection windows are used for feature propagation depending on the complexity of the image.

To show the functionalities of different units, we have carefully designed two competitors. The first (**Fulkerson09**) is proposed in [16], where a segment based CRF model is used to solve the segmentation problem but without the intervention of any object detector. Segments are produced by the quick shift as in [16]. Another competitor (DPM-**Levelset**) is a cascade of lesion detection and level set segmentation. Similar to [7], the result of detection is used as the bounding box of the lesion to initialize the shape model of level set. The maximal number of dynamic iterations is set to 500.

For lesions with homogeneous appearance and distinct border, all these approaches perform well. We show some complicated cases with their results in Fig. 3. Lesions are benign in upper 4 rows and malignant in lower 4 rows. The proposed algorithm works best in these cases. The approach of **Fulkerson09** lacks of object-level information and thus has troubles in discriminating spurious segments and recovering the lesion boundary. Post-processing of these results is not straightforward. DPM-**Levelset** works quite well if the level set model is initialized properly, but it is totally confused when the lesion detection fails such as in the 4th and 8th cases. The proposed algorithm is able to find an optimal balance between the bottom-up and the top-down image cues, therefore avoids being trapped by any of these problems.

The segmentation performance is reported in Table 1. 10% outliers are removed for all methods as in [7]. Let S be the segmented lesion region and G be the lesion in groundtruth. The Jaccard coefficient is defined as $(S \cap G)/(S \cup G)$.

Table 1. Statistical results of lesion segmentation

	Fulkerson09	DPM-Levelset	The proposed
Average Jaccard	0.57 ± 0.24	0.69 ± 0.26	0.75 ± 0.17
Median Jaccard	0.65	0.76	0.81

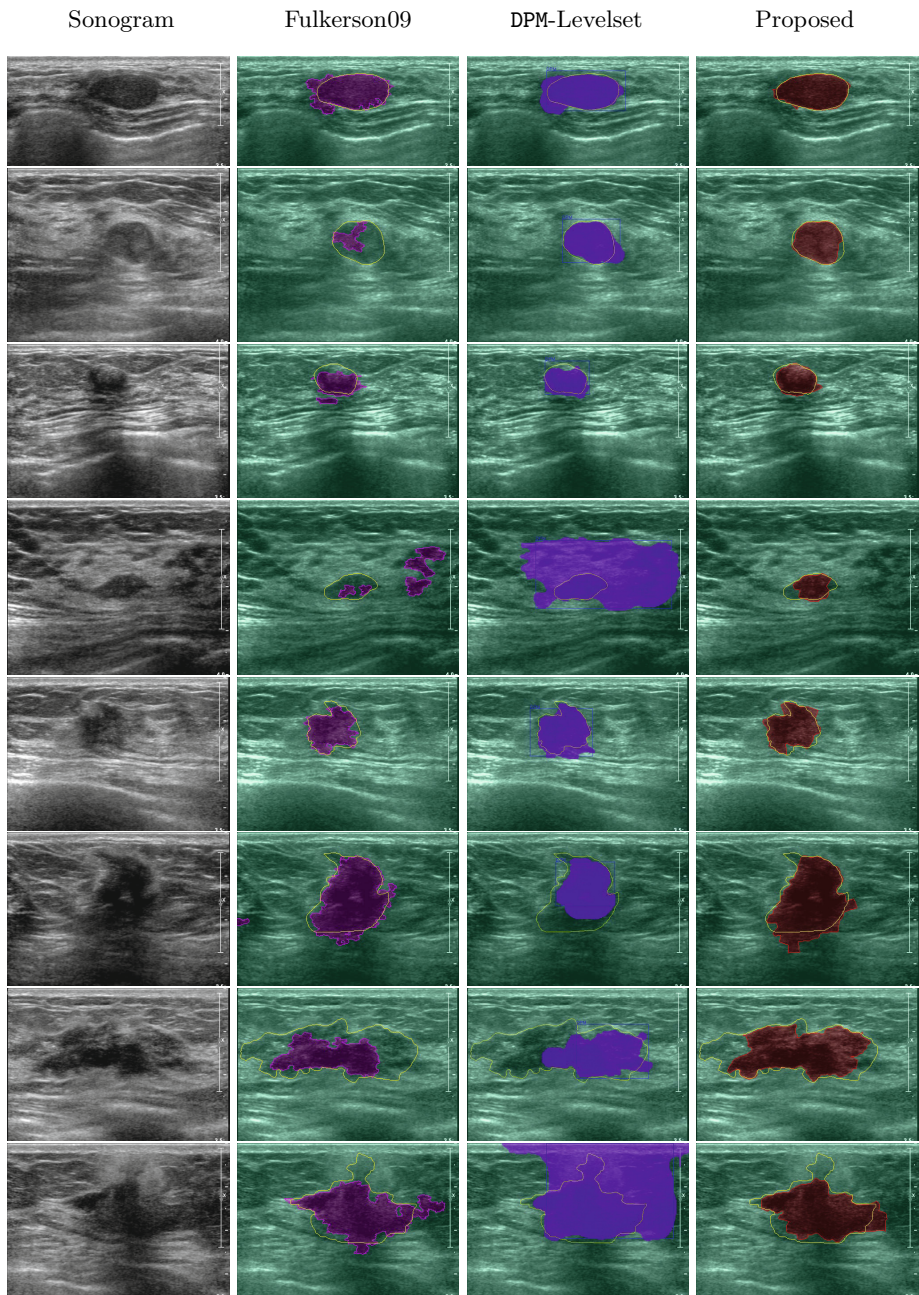| Sonogram	Fulkerson09	DPM-Levelset	Proposed

Fig. 3. Lesion segmentation results on breast ultrasound images. The upper 4 rows are benign cases and the lower 4 are malignant. Contours of groundtruth are shown in yellow. Detection windows of DPM with maximal confidences are shown as blue rectangles in the 3rd column. Note that the DPM detector works in both **DPM-Levelset** and the proposed algorithm. When it fails in the 4th and the last cases, the proposed algorithm ignores the detection mistakes and outperforms its counterparts.

5   Conclusion

We present a new algorithm for lesion segmentation in breast sonograms. The integration of image over-segmentation and lesion detection into a `CRF` model is proposed to achieve a comprehensive optimization on different diagnostic rules. The segmentation is driven by the segment-based `CRF` model and not aligned well with the local edge, which provides us a direction of the future work.

References

1. Stavros, A., Thickman, D., Rapp, C., Dennis, M., Parker, S., Sisney, G., et al.: Solid breast nodules: use of sonography to distinguish between benign and malignant lesions. Radiology 196(1), 123 (1995)
2. Noble, J., Boukerroui, D.: Ultrasound image segmentation: A survey. TMI 25(8), 987–1010 (2006)
3. Madabhushi, A., Metaxas, D.: Combining low-, high-level and empirical domain knowledge for automated segmentation of ultrasonic breast lesions. TMI 22(2), 155–169 (2003)
4. Liu, B., Cheng, H., Huang, J., Tian, J., Tang, X., Liu, J.: Fully automatic and segmentation-robust classification of breast tumors based on local texture analysis of ultrasound images. PR 43(1), 280–298 (2010)
5. Chan, T., Vese, L.: Active contours without edges. TIP 10(2), 266–277 (2001)
6. Huang, X., Metaxas, D.: Metamorphs: Deformable shape and appearance models. PAMI 30(8), 1444–1459 (2008)
7. Zhang, J., Zhou, S., Brunke, S., Lowery, C., Comaniciu, D.: Database-guided breast tumor detection and segmentation in 2d ultrasound images. In: SPIE Medical Imaging, vol. 7624, p. 3. Citeseer (2010)
8. Felzenszwalb, P., Girshick, R., McAllester, D., Ramanan, D.: Object detection with discriminatively trained part-based models. PAMI, 1627–1645 (2009)
9. Gao, T., Packer, B., Koller, D.: A segmentation-aware object detection model with occlusion handling. In: CVPR, pp. 1361–1368. IEEE (2011)
10. Ladický, Ľ., Sturgess, P., Alahari, K., Russell, C., Torr, P.H.S.: What, Where and How Many? Combining Object Detectors and CRFs. In: Daniilidis, K., Maragos, P., Paragios, N. (eds.) ECCV 2010, Part IV. LNCS, vol. 6314, pp. 424–437. Springer, Heidelberg (2010)
11. Mori, G.: Guiding model search using segmentation. In: ICCV, vol. 2, pp. 1417–1423. IEEE (2005)
12. Shi, J., Malik, J.: Normalized cuts and image segmentation. PAMI 22(8), 888–905 (2000)
13. Vedaldi, A., Soatto, S.: Quick Shift and Kernel Methods for Mode Seeking. In: Forsyth, D., Torr, P., Zisserman, A. (eds.) ECCV 2008, Part IV. LNCS, vol. 5305, pp. 705–718. Springer, Heidelberg (2008)
14. Alexe, B., Deselaers, T., Ferrari, V.: What is an object? In: CVPR, pp. 73–80. IEEE (2010)
15. Oliver, A., Freixenet, J., Martí, R., Zwiggelaar, R.: A Comparison of Breast Tissue Classification Techniques. In: Larsen, R., Nielsen, M., Sporring, J. (eds.) MICCAI 2006, Part II. LNCS, vol. 4191, pp. 872–879. Springer, Heidelberg (2006)
16. Fulkerson, B., Vedaldi, A., Soatto, S.: Class segmentation and object localization with superpixel neighborhoods. In: CVPR, pp. 670–677. IEEE (2009)