GLOBALLY OPTIMAL BREAST MASS SEGMENTATION FROM DCE-MRI USING DEEP SEMANTIC SEGMENTATION AS SHAPE PRIOR

Gabriel Maicas[†] *Gustavo Carneiro*[†] *Andrew P. Bradley*[⋆] [*]

[†] ACVT, School of Computer Science, The University of Adelaide
[⋆] School of ITEE, The University of Queensland

ABSTRACT

We introduce a new fully automated breast mass segmentation method from dynamic contrast-enhanced magnetic resonance imaging (DCE-MRI). The method is based on globally optimal inference in a continuous space (GOCS) using a shape prior computed from a semantic segmentation produced by a deep learning (DL) model. We propose this approach because the limited amount of annotated training samples does not allow the implementation of a robust DL model that could produce accurate segmentation results on its own. Furthermore, GOCS does not need precise initialisation compared to locally optimal methods on a continuous space (e.g., Mumford-Shah based level set methods); also, GOCS has smaller memory complexity compared to globally optimal inference on a discrete space (e.g., graph cuts). Experimental results show that the proposed method produces the current state-of-the-art mass segmentation (from DCE-MRI) results, achieving a mean Dice coefficient of 0.77 for the test set.

***Index Terms*—** breast cancer, deep learning, energy-based segmentation, shape prior, breast mass segmentation, breast MRI, global optimization.

1. INTRODUCTION

Breast screening based on dynamically contrast-enhanced magnetic resonance imaging (DCE-MRI) is particularly useful for patients with dense breasts [1], given that for this cohort, DCE-MRI allows an increase in sensitivity, compared to mammograms [2, 3]. Due to the necessity of interpreting 4D images (3D volumes over time), analysing DCE-MRI images is a complex task that requires medical expertise and is prone to large inter-user reading variability. As a result, computer assisted detection (CAD) systems are being developed to assist radiologists in this task [4]. The analysis used in these systems can be divided into the detection, segmentation and classification of masses where the main contribution of this paper lies in the segmentation of masses. Furthermore, we also propose a novel multimodal detection approach to allow the implementation of a fully automated segmentation methodology.

The particular problem of mass segmentation is challenging due to the variable size, appearance and shape of tumours [5], and the relatively low signal to noise ratio of the masses in DCE-MRI. In addition, fully automated methodologies need to address the usually inaccurate alignment of the initial region of interest (ROI) for the segmentation. State-of-the-art segmentation methods mostly rely on the development of hand-crafted features [6, 7] and methods based on globally optimal inference on a discrete space [7].

In this paper, we propose GOCS-DLP, a new breast mass segmentation methodology from DCE-MRI based on *globally optimal inference* on a *continuous space* that relies on a *shape prior based on the semantic segmentation computed from a deep learning* (DL) model (see Fig. 1). The method is inspired by a recent work [8] that explores *locally optimal inference* on continuous space and uses a shape prior based on the semantic segmentation computed from a DL model for the problem of left ventricle segmentation from MRI. We extend this method with the use of globally optimal inference, which shows robustness to the initialisation of the inference process [9]. Compared to the work by Cremers et al. [9], the main novelty lies in the use of a DL model as a shape prior. In order to make the segmentation fully automated, we extend the breast mass detection methodology (from mammograms) proposed by Dhungel et al. [10] with a model that is formed by a cascade of multimodal deep learning classifiers. While we employ DCE-MRI for segmentation, we combine T1-weighted, T2-weighted and DCE-MRI for detection .

We test our proposed methodology using a breast multimodal MRI dataset, containing 117 cases, with 141 annotated masses, with 46 being benign and 95 malignant, where 58 patients are for training and 59 for testing. We compare, in terms of the mean Dice coefficient ($\bar{D}$), the segmentation results produced by our proposed method ($\bar{D} = 0.77$) with several baselines: globally optimal inference on a discrete space [7] (GODS) ($\bar{D} = 0.74$), globally optimal inference on a continuous space without a DL shape prior (GOCS-MS) ($\bar{D} = 0.73$), locally optimal inference on a continuous space using DL shape priors [8] (LOCS) ($\bar{D} = 0.62$), and semantic segmentation from convolutional neural network (CNN) [11] ($\bar{D} = 0.68$). These results show that our method is significantly more accurate than the competition for the fully automated problem (i.e. using automated mass detection) and for the semi-automated problem (i.e., with manual mass detection).

1.1. Literature Review

The segmentation of breast masses from DCE-MRI has been addressed in several ways. Jayender et al. [6] use hand-crafted features in a voxel-wise classification, where the disadvantage lies in the relatively poor segmentation accuracy produced due to the lack of a shape model that can smooth the result, which is an issue that also affects region-growing segmentation methods [12]. Relying on hand-crafted features is another issue of the method above [6], which has recently been addressed by the computer vision community with the use of deep learning models that automatically learn features for particular classification/regression tasks [13]. More accurate results can be obtained with segmentation methods based on global infer-

[*]This work was partially supported by the Australian Research Council's Discovery Projects funding scheme (project DP140102794). Prof. Bradley is the recipient of an Australian Research Council Future Fellowship(FT110100623)

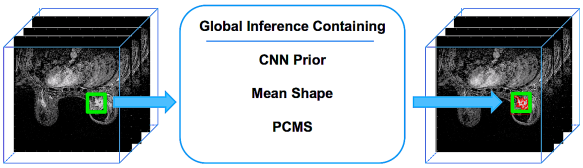

Fig. 1. Breast mass segmentation from DCE-MRI using a global minimisation on a continuous space based on the following energy functional terms: DL (CNN) prior [11], mean shape and the terms inherited from the piece-wise constant Mumford Shah [19] model (PCMS).

ence on discrete spaces [7] (i.e., graph cut [14]), which has high memory complexity that is circumvented with the use of techniques to reduce the number of graph nodes (e.g., superpixels [15]).

Deep learning models have been explored for semantic segmentation [11], which in general need large amounts of annotated training data and produce relatively poor results in terms of segmentation accuracy due to the lack of shape models. The use of shape prior models within deep learning [16] has addressed the accuracy issue at the expense of complex learning methods that require even larger annotated training sets. Given that in medical image analysis it is rare to come across problems with large annotated training sets, recent developments focus on the use of the shape produced by deep learning models as a weak shape prior that is combined with other segmentation cues, such as: strong edges, homogeneous grey-value intensities, contour smoothness, etc. In essence, this involves the combination of level set methods [17] and deep learning shape priors, which has been recently studied for the problem of left ventricle segmentation from MRI [8]; here the issue is the strong dependence on an accurate intialisation because level set produces a locally optimal segmentation result. Level set optimisation problems have been relaxed [18] to transform it into globally optimal inference, which in turn has been adapted to work with (non-deep learning) shape priors [9]. Therefore, the main novelty of our paper is the combination of the globally optimal inference on continuous space with the use of a deep learning-based shape prior.

For the mass detection problem, we extend the methodology proposed by Dhungel at al. [10], which consists of a cascade of deep learning and random forest classifiers (where the random forest classifiers use hand-crafted features) to detect masses on mammograms. This methodology currently holds the state-of-the-art results in a few publicly available mammogram datasets. Our proposed extension reduces the complexity of Dhungel at al.'s approach [10] by reducing the number of cascade stages and avoids the use of hand-designed features by automatically learning features with DL. Furthermore, similarly to [7], we also consider a multimodal approach for mass detection.

2. METHODOLOGY

In this section we explain the deep learning model used to produce the shape prior, the globally optimal inference on a continuous space that uses this deep learning shape prior, and the breast mass detection approach. Hereafter, let $\mathcal{D} = \{\mathbf{v}_i, \mathbf{y}_i\}_{i=1}^{|\mathcal{D}|}$ be the annotated dataset, where each DCE-MRI volume is represented by $\mathbf{v} : \Omega \subset \mathbb{R}^3 \rightarrow [0, 1]$ (where mass-like voxels have values closer to 1) and the corresponding breast mass annotation is denoted by $\mathbf{y} : \Omega \subset \mathbb{R}^3 \rightarrow \{0, 1\}$, where 0 represents background and mass is denoted by 1.

2.1. Deep Learning Model

The deep learning shape prior is produced by a convolutional neural network (CNN) that outputs a semantic segmentation [11] of the breast mass from a DCE-MRI. The CNN is defined by:

$$f(\mathbf{v}, \theta) = \mathbf{y}^{*,CNN} = f_{out} \circ f_L \circ ... \circ f_2 \circ f_1(\mathbf{v}(0)), \quad (1)$$

where $\mathbf{v}(0) = \mathbf{v}$ (i.e., the original DCE-MRI volume), $\circ$ denotes the composition operator, $\mathbf{y}^{*,CNN} \in [0, 1]$, and θ represents the CNN parameters (i.e., weights and biases). Note from (1) that the output $\mathbf{y}^{*,CNN}$ estimates a binary map with the mass segmentation. Each layer in (1) contains a set of filters, defined by

$$\mathbf{v}(l) = f_l(\mathbf{v}(l-1)) = \sigma(\mathbf{W}_l^\top \mathbf{v}(l-1) + \beta_l), \quad (2)$$

where $\sigma(.)$ represents a non-linearity [13], and the convolutional filters are represented by the weight matrix $\mathbf{W}_l$ and bias vector β_l. The modelling of the CNN is performed with a supervised learning process, where the goal is to approximate the annotation by minimising the following per-pixel binomial logistic loss:

$$L = \sum_{i=1}^{|\mathcal{D}|} \sum_{x \in \Omega} \log\left(1 + e^{(-\mathbf{y}_i(x) \times \mathbf{y}_i^{*,CNN}(x))}\right), \quad (3)$$

where x indexes the volume lattice Ω.

2.2. Globally Optimal Inference on a Continuous Space using a Deep Learning Shape Prior

The locally optimal inference on a continuous space (i.e., level set method [17]) can denote the segmentation function in various ways [18], such as the zero level set of a signed distance function, or as one of the two regions of a binary function. We assume the latter representation, with the level set function denoted by $\tilde{\mathbf{u}} : \Omega \rightarrow \{0, 1\}$, where the final segmentation is obtained with $\mathbf{y}^{*,LO} = \tilde{u}$, where LO stands for local optimisation. The energy functional of our approach extends the piece-wise constant Mumford-Shah (PCMS) [19] (Fig. 2):

$$E(\tilde{\mathbf{u}}) = \beta \int_\Omega (\bar{\mathbf{y}}(x) - \tilde{\mathbf{u}}(x))^2 dx + \alpha \int_\Omega (\mathbf{y}^{*,CNN}(x) - \tilde{\mathbf{u}}(x))^2 dx + \lambda \int_\Omega (\mathbf{v}(x) - \tilde{\mathbf{u}}(x))^2 dx + |\nabla \tilde{\mathbf{u}}(x)| \quad (4)$$

where $\bar{\mathbf{y}}(x) = \frac{1}{|\mathcal{D}|} \sum_{i=1}^{|\mathcal{D}|} \mathbf{y}_i(x)$ represents the mean shape prior [9] at position $x \in \Omega$, $\mathbf{y}^{*,CNN}(.)$ denotes the DL shape prior from (1). While the role of the DL shape prior is to capture mass variability, the mean shape prior attemps to constrain its translation and scale. Finally, the last two terms (from PCMS) penalise differences between $\mathbf{v}$ and $\tilde{\mathbf{u}}$ and large segmentation perimeters. The minimisation of the energy functional in (4) finds the steady state solution of the gradient flow by iteratively computing the solution of the equation $\frac{\partial \tilde{\mathbf{u}}}{\partial t} = -\frac{\partial E}{\partial \tilde{\mathbf{u}}}$, where the $\frac{\partial E}{\partial \mathbf{u}}$ denotes the Gâteaux derivative of $E(\tilde{\mathbf{u}})$.

The energy functional proposed in (4) is not convex because although the functional $E(\tilde{\mathbf{u}})$ is convex, the domain of optimisation is a non-convex set of functions. Following the approach by Chan et al. [18], we relax $\tilde{\mathbf{u}} : \Omega \rightarrow \{0, 1\}$ to $\mathbf{u} : \Omega \rightarrow [0, 1]$ so it can represent a convex set of functions (i.e., the domain of optimisation

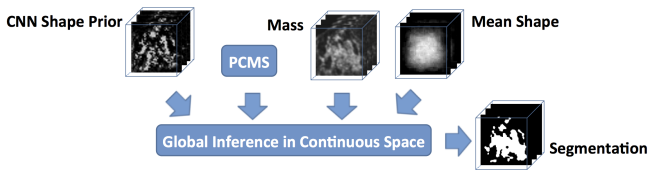

Fig. 2. Energy functional terms of the proposed method: shape prior from the CNN semantic segmentation (1), PCMS appearance and shape terms, and mean shape from training annotations. The final segmentation is estimated with a global inference on a continuous space.

is now convex), and as a result transform (4) into a convex optimisation problem. Theorem 1 in [18] assures the existence, but not uniqueness, of a global minimiser of the original problem reached by thresholding $\mathbf{u}(.)$:

$$\mathbf{y}^{*,GO} = \mathbf{1}_{\Sigma(x)}, \text{ with } \Sigma = \{x \in \Omega \mid \mathbf{u}(x) > \tau\}, \qquad (5)$$

where $\tau \in [0, 1]$ and $\mathbf{1}_{\Sigma(x)}$ denotes an indicator function that returns 1 if $x \in \Sigma$. Therefore, the inference based on the minimisation of $E(\mathbf{u})$ represents a globally optimal inference in the continuous space, where the main advantage lies in the use of arbitrary initialisation.

2.3. Breast Mass Detection based on a Cascade of Deep Learning Models

Initial mass regions are found automatically. Firstly, we follow [7] to obtain the breast region in the volume by applying Hayton's algorithm [20]. Then, the mass detection method extends the approach proposed by Dhungel at al. [10] that finds breast masses in mammograms. The proposed approach consists of a cascade of CNNs (1), where the first stage is a pixel-wise detection of mass candidates, using as input the 3-D data of the different MRI modalities: T1-weighted, T2-weighted and DCE-MRI. The second stage uses connected component analysis to merge the mass candidate voxels to form region candidates. Finally, these region candidates are fed into a cascade of two CNNs that process those candidates sequentially. The activations from the last layer of each CNN are concatenated and passed into a random forest classifier [21] to produce a final region classification. This approach improves Dhungel et al.'s approach [10] by reducing the complexity of the cascade in terms of the number of stages and by including multimodal image data.

3. EXPERIMENTS

This section introduces the dataset and experimental setup, the methods used in the comparison, the details of our proposed method, and the results.

3.1. Dataset

The dataset used to assess the accuracy of our methodology contains breast MRI studies of 117 patients, where the mean age is 48 ± 12 with age range between 22 and 84 years. Three image modalities are used in this study: DCE-MRI scans for segmentation and ROI detection, and T1-weighted anatomical and T2-weighted anatomical scans for ROI detection only. All images were acquired on a 1.5T GE Signa HDxt scanner, with the patient in prone position. T1-weighted anatomical volumes acqusition was performed axially (acquisition matrix of 512×512) and without fat suppression. T2-weighted anatomical volumes were obtained axially (acquisition matrix of 320×224) and with fat suppression. In order to obtain the DCE-MRI volumes, four or five volumes were acquired axially with fat suppression. The first (pre-contrast) volume is obtained before a contrast agent is injected to the patient. Then, several acquisitions are obtained at different time points (post-contrast volumes), where subtraction volumes are obtained by subtracting pre-contrast and post-contrast volumes. The acquisition matrix is 360x360 for these DCE-MRI volumes. All images were registered to the first post-contrast volume. There is at least one breast mass present in each DCE-MRI study, where the total number of masses is 141 (46 benign and 95 malignant) that were cyto- or histopathology confirmed. All masses were annotated by a radiographer using a region growing algorithm on the subtraction volumes [22].

The dataset was randomly divided into training and testing sets. In contrast to [7], where the training and testing data consisted of 35 (41 lesions) and 85 patients (93 lesions) respectively, our training set contains studies from 58 patients, with 72 lesions (23 benign and 49 malignant), and testing has the studies from 59 patients, with 69 lesions (23 benign and 46 malignant). Segmentation accuracy is assessed with the mean and median Dice coefficient on the training and testing sets.

3.2. Experimental Setup

We evaluate our segmentation methodology on the first subtraction image of DCE-MRI. As the initial ROI, we use both automated (as explained in Sec 2.3) and manual detection. A lesion is correctly detected when the Dice coefficient between the manual annotation and the ROI is at least 0.4, yielding a true positive rate (TPR) of 0.85 at 3.66 false positive regions per patient. T1-weighted, T2-weighted and the first two subtraction volumes are used for all deep learning models during the detection phase. In the case of the manual set-up, the initial region is the bounding box of the ground truth augmented by three voxels in each direction.

For the segmentation baseline methods, we use the results of the globally optimal inference on a discrete space (GODS) method [7] that holds the current state-of-the-art results for the dataset above. We re-implemented the locally optimal inference on a continuous space (LOCS) [8] that uses a DL shape prior model on a distance regularised level set method [23]. In addition, we also implemented the segmentation for globally optimal inference on a continuous space with a mean shape prior (GOCS-MS) [9]. Finally, we also implemented the CNN semantic segmentation [11] for comparison.

For our methodology, we use the training set to estimate the mean shape $\bar{\mathbf{y}}$ in (4), the weights and biases of the CNN in (1), and the weights of the terms in the energy functional (4). The CNN consists of 3 convolutional layers, with linear activation functions, and an output layer with two channels of size $30 \times 25 \times 18$ representing the probability of background (channel 1) or mass (channel 2). This fixed output size requires that the annotations are resized to fit that output layer during training. The first layer has 10 filters of size $5 \times 5 \times 3$, while the second and third layers contain 20 filters of size $3 \times 3 \times 3$. The learning rate is 0.1 and the input layer is a volume of size $30 \times 25 \times 18$, where the input volume is resized with cubic interpolation to fit this input layer. The CNN structure and its weights and biases in (1) are estimated using exclusively the training set, which is sub-divided into training (with 45 patients, containing 57 lesions) and validation (with 13 patients and 15 lesions) sets, for model selection. We augment the training data by flipping

	Mean Dice		Median Dice			
	Train	**Test**	**Train**	**Test**	**Detection**	**Inference Time**
GOCS-DLP(Ours)	0.80 ± 0.11	$\mathbf{0.77 \pm 0.14}$	0.82	**0.82**	Auto	7.78 ± 20.69 s
GOCS-DLP(Ours)	0.79 ± 0.13	$\mathbf{0.77 \pm 0.13}$	0.81	0.80	Manual	5.95 ± 18.07 s
GODS [7]	-	0.74 ± 0.12	-	0.76	Auto	-
LOCS [8]	0.64 ± 0.16	0.62 ± 0.15	0.65	0.64	Auto	14.37 ± 41.15 s
LOCS [8]	0.61 ± 0.15	0.59 ± 0.17	0.64	0.61	Manual	12.90 ± 28.13 s
GOCS-MS [9]	0.76 ± 0.18	0.73 ± 0.21	0.81	0.79	Auto	7.61 ± 21.44 s
GOCS-MS [9]	0.75 ± 0.18	0.72 ± 0.22	0.80	0.79	Manual	5.83 ± 15.02 s
CNN	0.69 ± 0.16	0.68 ± 0.19	0.70	0.74	Auto	0.12 ± 0.16 s
CNN	0.66 ± 0.16	0.66 ± 0.18	0.68	0.69	Manual	0.11 ± 0.03 s

Table 1. Mean, median and standard deviation for training/testing Dice coefficients and inference time (per lesion) for each methodology.

ROIs in each of the three possible axes. Finally, the weights in the energy functional in (4) are estimated (using the training set) over a grid of possible values for each term, where the estimated values are $\lambda = 55, \alpha = 2.5, \beta = 1.5$. We label our approach as globally optimal inference on a continuous space using DL shape prior (GOCS-DLP). Note that for the LOCS and GOCS-MS method, we run a similar method to estimate these weights.

For the inference of GOCS-DLP, GOCS-MS and LOCS the initial segmentation consists of a rectangular centered prism of 90% of the ROI volume. When the optimisation process has converged, we threshold the solution of the relaxed problem at the value of $\tau = 0.75$ as defined in (5).

The same training, validation and testing sets are employed to automatically detect masses. For the first stage, we use a three-scale CNN. The network architecture in every scale is composed of 4 convolutional layers, a fully connected layer and a softmax operation to normalize probabilities of being normal and mass tissue. The multi-scale CNN produces an average of 30 false positive ROIs per image. Such candidates are fed into a cascade of two CNNs, each containing two convolutional - max pooling layers, two convolutional layers, a fully connected layer and a softmax layer. Activations from both last layers are concatenated together to form the input to a random forest classifier to produce the final classification.

3.3. Results

We compare our proposed GOCS-DLP with GODS, LOCS, GOCS-MS, and the CNN semantic segmentation for the training and testing sets in Table 1, where the last column refers to the average time employed for the segmentation of one lesion. We measure the statistical significance of the results for both automated and manual detection in Table 1 with the Wilcoxon signed-rank test, and the results from the proposed GOCS-DLP are significant compared to all others, assuming a significance level of 0.01. In particular, p-values of fully automated GOCS-DLP with respect to fully automated GODS and GOCS-MS are 0.0081 and 0.0005 respectively. Fig.3 shows examples of mass segmentations achieved with our proposed GOCS-DLP after an automated detection.

4. DISCUSSION AND CONCLUSION

The experimental resuts show that the segmentation accuracy produced by our proposed GOCS-DLP is significantly better ($p < 0.01$) than the baselines GODS, LOCS and GOCS-MS and CNN semantic segmentation for the problem of breast mass segmentation from DCE-MRI. Note that the segmentation accuracy using the manual ROI detection is not better than its automated counterpart because masses that are not automatically detected are not passed to the segmentation stage, and these masses turn out to be the most challenging ones to be segmented.

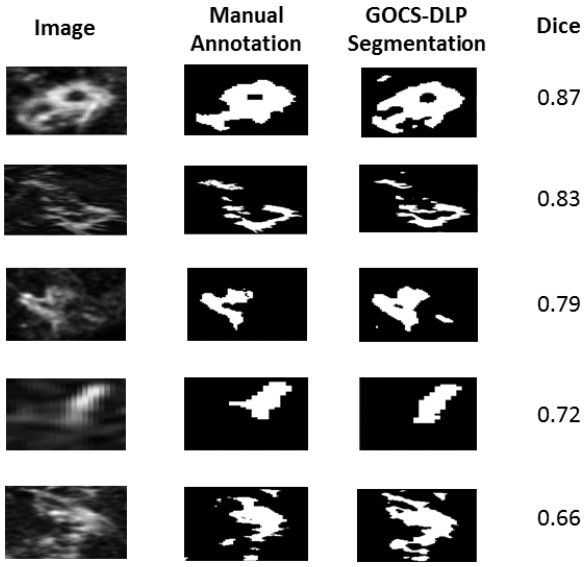

Fig. 3. Examples of mass segmentations produced by our proposed GOCS-DLP. GODS results not available

In Sec. 1, we hypothesise that the semantic segmentation from CNN would not be robust enough because of the small training set, and the evidence in Table 1 provides support for that hypothesis. Table 1 also shows that the inclusion of the DL shape prior significantly improves the quality of the segmentation (GOCS-DLP vs GOCS-MS). However, we expect that for large training sets, the CNN alone will be able to produce accurate segmentation on its own. In addition, one of the advantages of globally optimal inference compared to the locally optimal inference is the independence with respect to the initialisation: global methods produce significantly better segmentation than the local method. This advantage is particularly effective to avoid large errors in segmentation after an automated detection of the lesion, which might yield a misaligned ROI. In fact, even a more precise initialisation (the bounding box of the manual annotation) for LOCS does not achieve an accurate segmentation ($\bar{D} = 0.68$ for the test set) compared to the global methods.

The inference time in Table 1 shows that the global methods converge much faster. Note that the large variability in the inference time is due to the variation of shape and size of masses. Even though we do not have the training and inference time results for GODS [7], the comparison in terms of "Mean Dice Test" and "Median Dice Test" shows evidence of the disadvantage of using superpixels to decrease the memory complexity that implicitly assumes appearance homogeneity, which may not be correct for this problem. Finally, the visual results in Fig. 3 show that our proposed GOCS-DLP produces quite precise segmentation results for different types of masses.

In this work, we introduce a new segmentation method that combines global inference in the continuous space with deep learning for the problem of breast mass segmentation from DCE-MRI. Our results show a significant improvement over the state-of-art for this problem, where we also present results produced by several baseline methods based on DL alone, discrete global optimisation and continuous global optimisation. We intend to apply the proposed methodology in other medical imaging segmentation problems.

5. REFERENCES

[1] Albert L Siu, "Screening for breast cancer: Us preventive services task force recommendation statement," *Annals of internal medicine*, 2016.

[2] Per Skaane, Andriy I Bandos, Randi Gullien, Ellen B Eben, Ulrika Ekseth, Unni Haakenaasen, Mina Izadi, Ingvild N Jebsen, Gunnar Jahr, Mona Krager, et al., "Comparison of digital mammography alone and digital mammography plus tomosynthesis in a population-based screening program," *Radiology*, vol. 267, no. 1, pp. 47–56, 2013.

[3] Debbie Saslow, Carla Boetes, Wylie Burke, Steven Harms, Martin O Leach, Constance D Lehman, Elizabeth Morris, Etta Pisano, Mitchell Schnall, Stephen Sener, et al., "American cancer society guidelines for breast screening with mri as an adjunct to mammography," *CA: a cancer journal for clinicians*, vol. 57, no. 2, pp. 75–89, 2007.

[4] C Dromain, B Boyer, R Ferre, S Canale, S Delaloge, and C Balleyguier, "Computed-aided diagnosis (cad) in the detection of breast cancer," *European journal of radiology*, vol. 82, no. 3, pp. 417–423, 2013.

[5] Yading Yuan, Maryellen L Giger, Hui Li, Kenji Suzuki, and Charlene Sennett, "A dual-stage method for lesion segmentation on digital mammograms," *Medical physics*, vol. 34, no. 11, pp. 4180–4193, 2007.

[6] Jagadaeesan Jayender, Sona Chikarmane, Ferenc A Jolesz, and Eva Gombos, "Automatic segmentation of invasive breast carcinomas from dynamic contrast-enhanced mri using time series analysis," *Journal of Magnetic Resonance Imaging*, vol. 40, no. 2, pp. 467–475, 2014.

[7] Darryl McClymont, Andrew Mehnert, Adnan Trakic, Dominic Kennedy, and Stuart Crozier, "Fully automatic lesion segmentation in breast mri using mean-shift and graph-cuts on a region adjacency graph," *Journal of Magnetic Resonance Imaging*, vol. 39, no. 4, pp. 795–804, 2014.

[8] Tuan Ngo and Gustavo Carneiro, "Fully automated non-rigid segmentation with distance regularized level set evolution initialized and constrained by deep-structured inference," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 3118–3125.

[9] Daniel Cremers, Frank R Schmidt, and Frank Barthel, "Shape priors in variational image segmentation: Convexity, lipschitz continuity and globally optimal solutions," in *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*. IEEE, 2008, pp. 1–6.

[10] Neeraj Dhungel, Gustavo Carneiro, and Andrew P Bradley, "Automated mass detection in mammograms using cascaded deep learning and random forests," in *Digital Image Computing: Techniques and Applications (DICTA), 2015 International Conference on*. IEEE, 2015, pp. 1–8.

[11] Jonathan Long, Evan Shelhamer, and Trevor Darrell, "Fully convolutional networks for semantic segmentation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 3431–3440.

[12] Ali Qusay Al-Faris, Umi Kalthum Ngah, Nor Ashidi Mat Isa, and Ibrahim Lutfi Shuaib, "Computer-aided segmentation system for breast mri tumour using modified automatic seeded region growing (bmri-masrg)," *Journal of digital imaging*, vol. 27, no. 1, pp. 133–144, 2014.

[13] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, 2012, pp. 1097–1105.

[14] Yuri Boykov, Olga Veksler, and Ramin Zabih, "Fast approximate energy minimization via graph cuts," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 23, no. 11, pp. 1222–1239, 2001.

[15] Radhakrishna Achanta, Appu Shaji, Kevin Smith, Aurelien Lucchi, Pascal Fua, and Sabine Susstrunk, "Slic superpixels compared to state-of-the-art superpixel methods," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 34, no. 11, pp. 2274–2282, 2012.

[16] Shuai Zheng, Sadeep Jayasumana, Bernardino Romera-Paredes, Vibhav Vineet, Zhizhong Su, Dalong Du, Chang Huang, and Philip HS Torr, "Conditional random fields as recurrent neural networks," in *Proceedings of the IEEE International Conference on Computer Vision*, 2015, pp. 1529–1537.

[17] Stanley Osher and James A Sethian, "Fronts propagating with curvature-dependent speed: algorithms based on hamilton-jacobi formulations," *Journal of computational physics*, vol. 79, no. 1, pp. 12–49, 1988.

[18] Tony F Chan, Selim Esedoglu, and Mila Nikolova, "Algorithms for finding global minimizers of image segmentation and denoising models," *SIAM journal on applied mathematics*, vol. 66, no. 5, pp. 1632–1648, 2006.

[19] David Mumford and Jayant Shah, "Optimal approximations by piecewise smooth functions and associated variational problems," *Communications on pure and applied mathematics*, vol. 42, no. 5, pp. 577–685, 1989.

[20] Paul Hayton, Michael Brady, Lionel Tarassenko, and Niall Moore, "Analysis of dynamic mr breast images using a model of contrast enhancement," *Medical image analysis*, vol. 1, no. 3, pp. 207–224, 1997.

[21] Leo Breiman, "Random forests," *Machine learning*, vol. 45, no. 1, pp. 5–32, 2001.

[22] Antoine Rosset, Luca Spadola, and Osman Ratib, "Osirix: an open-source software for navigating in multidimensional dicom images," *Journal of digital imaging*, vol. 17, no. 3, pp. 205–216, 2004.

[23] Chunming Li, Chenyang Xu, Changfeng Gui, and Martin D Fox, "Distance regularized level set evolution and its application to image segmentation," *Image Processing, IEEE Transactions on*, vol. 19, no. 12, pp. 3243–3254, 2010.