THE UNIVERSITY
of ADELAIDE

Medical Image Segmentation Combining Level Set Method and Deep Belief Networks

by

Tuan Anh Ngo

A thesis submitted in fulfillment for the
degree of Doctor of Philosophy

in the

Faculty of Engineering, Computer and Mathematical Sciences
School of Computer Science

February 2016

Contents

List of Figures

Abstract

Medical image segmentation is an important step in medical image analysis, where the main goal is the precise delineation of organs and tumours from medical images. For instance there is evidence in the field that shows a positive correlation between the precision of these segmentations and the accuracy observed in classification systems that use these segmentations as their inputs.

Over the last decades, a vast number of medical image segmentation models have been introduced, where these models can be divided into five main groups: 1) image-based approaches, 2) active contour methods, 3) machine learning techniques, 4) atlas-guided segmentation and registration and 5) hybrid models. Image-based approaches use only intensity value or texture for segmenting (i.e., thresholding technique) and they usually do not produce precise segmentation. Active contour methods can use an explicit representation (i.e., snakes) with the goal of minimizing an energy function that forces the contour to move towards strong edges and maintains the contour smoothness. The use of implicit representation in active contour methods (i.e., level set method) embeds the contour as zero level set of a higher dimensional surface (i.e., the curve representing the contour does not need to be parameterized as in the Snakes model). Although successful, the main issue with active contour methods is the fact that the energy function must contain terms describing all possible shape and appearance variations, which is a complicated task given that it is hard to design by hand all these terms. Also, this type of active contour methods may get stuck at image regions that do not belong to the object of interest. Machine learning techniques address this issue by automatically learning shape and appearance models using annotated training images. Nevertheless, in order to meet the high accuracy requirements of medical image analysis applications, machine learning methods usually need large and rich training sets and also face the complexity of the inference process. Atlas-guided segmentation and registration use an atlas image, which is constructed based on manually segmentation images. The new image is segmented by registering it with the atlas image. These techniques have been applied successfully in many applications, but they still face some issues, such as their ability to represent the variability of anatomical structure and scale in medical image, and the complexity of the registration algorithms.

In this work, we propose a new hybrid segmentation approach by combining a level set method with a machine learning approach (deep belief network). Our main objective with this approach is to achieve segmentation accuracy results that are either comparable or better than the ones produced with machine learning methods, but using relatively smaller training sets. These weaker requirements on the size of training sets is compensated by the hand designed segmentation

terms present in typical level set methods, that are used as prior information on the anatomy to be segmented (e.g., smooth contours, strong edges, etc.). In addition, we choose a machine learning methodology that typically requires smaller annotated training sets, compared to other methods proposed in this field. Specifically, we use deep belief networks, with training sets consisting to a large extent of un-annotated training images. In general, our hybrid segmentation approach uses the result produced by the deep belief network as a prior in the level set evolution. We validate this method on the Medical Image Computing and Computer Assisted Intervention (MICCAI) 2009 left ventricle segmentation challenge database and on the Japanese Society of Radiological Technology (JSRT) lung segmentation dataset. The experiments show that our approach produces competitive results in the field in terms of segmentation accuracy. More specifically, we show that the use of our proposed methodology in a semi-automated segmentation system (i.e., using a manual initialization) produces the best result in the field in both databases above, and in the case of a fully automated system, our method shows results competitive with the current state of the art.

Declaration

I certify that this work contains no material which has been accepted for the award of any other degree or diploma in my name, in any university or other tertiary institution and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made in the text. In addition, I certify that no part of this work will, in the future, be used in a submission in my name, for any other degree or diploma in any university or other tertiary institution without the prior approval of the University of Adelaide and where applicable, any partner institution responsible for the joint-award of this degree.

I give consent to this copy of my thesis when deposited in the University Library, being made available for loan and photocopying, subject to the provisions of the Copyright Act 1968.

The author acknowledges that copyright of published works contained within this thesis resides with the copyright holder(s) of those works.

I also give permission for the digital version of my thesis to be made available on the web, via the Universitys digital research repository, the Library Search and also through web search engines, unless permission has been granted by the University to restrict access for a period of time.

Signed: _____

Date: _____

Publications

My thesis is based on the content of the following conference and journal papers:

- Tuan Anh Ngo, Gustavo Carneiro. Left Ventricle Segmentation from Cardiac MRI Combining Level Set Methods with Deep Belief Networks. In *IEEE International Conference on Image Processing (ICIP)*, 2013.
  (DOI: 10.1109/ICIP.2013.6738143)
  (This paper was selected among the top 10% of the accepted papers of the conference)

- Tuan Anh Ngo, Gustavo Carneiro. Fully Automated Non-rigid Segmentation with Distance Regularized Level Set Evolution Initialized and Constrained by Deep-structured Inference. *IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)*, 2014.
  (DOI: 10.1109/CVPR.2014.399)

- Tuan Anh Ngo, Gustavo Carneiro. Lung Segmentation in Chest Radiographs using Distance Regularized Level Set and Deep-structured Learning and Inference. *IEEE International Conference on Image Processing (ICIP)*, 2015.

- Tuan Anh Ngo, Gustavo Carneiro. Combining Deep Learning and Level Set for the Automated Segmentation of the Left Ventricle of the Heart from Cardiac Cine MR. *Submitted to Medical Image Analysis* .
  (under review)

Acknowledgements

I would like to sincerely thank all those who support and encourage me in completing this Ph.D. thesis.

Foremost, I would like to express my deepest gratitude to my principal supervisor, Assoc Prof. Gustavo Carneiro, for his excellent guidance. In the first year of my research journey, I was facing huge amount of difficulties when I could not find the dataset for experiments. Therefore, I had to change the topic research and also supervisor. I have been grateful to him for accepting me as his student despite the fact that he knew that I did not have a strong background in computer vision and medical image processing. He has always been a great advisor and given me the directions to solve all challenges that I have faced when doing the research. I wish to thank for his patience when discussing, guiding my proposal, setting up the experiences, checking the results, and writing the papers. I believe that without his support, this thesis would not have been complete.

I very much appreciate Dr. Tat-Jun Chin, Dr. Alfred Brown, Prof. David Suter, Assoc Prof. Yung Ngothai (School of Chemical Engineering) for their encouragement and help. I would like to express my grief and gratitude to Dr. Trent Kroeger, who passed away, for supporting me in the early stages of the research. I also would like to thank my friends Trung Thanh Pham, Neeraj Dhungel, Quoc-Huy Tran, and Anh Quang Nguyen who have been sharing with me the knowledge as well as experience in other areas of life.

My sincere thanks go to the Vietnamese Government and the University of Adelaide for providing me the scholarship to do my thesis. I wish to thank the Australia Centre of Visual Technologies (ACVT) and all ACVT's staff for academic and ethical work environment.

Last and not at least, I am very grateful to my wife, Kieu Oanh Le, and my son, Anh Kha Ngo, for always being by my side and sharing difficulties during my four-year Ph.D. journey. I also would like to place on record my sincere thanks to my sister, Hien Anh Ngo, my father, my parents-in-law, who have been motivating and caring to me and my family.

July, 2015

Tuan Anh Ngo

Dedicated to my family,
for their unconditional love and endless support.

Chapter 1

Introduction

Image segmentation is an important and difficult stage that is usually present in medical image analysis systems. A loose definition of image segmentation is that it is a process of partitioning a digital image into regions of interest (e.g., an organ or a tumour and background). The result of image segmentation is a set of regions, or contours, extracted from an image. Medical image segmentation faces several challenges, including: strict requirements in terms of segmentation accuracy, low signal-to-noise ratio between the objects that need to be segmented and the background, and large shape and appearance variations of the visual object of interest.

In the last decades, there have been several segmentation methods proposed in medical image analysis, which produced significant results. One of the simplest techniques is the image-based approach, such as thresholding [1]. The main goal of thresholding is to find some threshold points (e.g., gray values) that divide the image into separate regions (e.g., objects and background). One well-known thresholding technique based on a discriminant criterion was introduced by Otsu [2], which estimates optimal threshold points that maximizes the between-class variance. The results of thresholding techniques are usually not precise enough, which means that these techniques are commonly combined with other more sophisticated segmentation techniques.

One of the most successful segmentation methods is the active contour model using explicit contour representation (also know as Snake) [3]. The snake model produces the object contour by minimising an energy function that includes internal and external energy constraints, with internal constraints denoting contour smoothness, and external constraints representing image features (e.g., edges) that drive the contour to segment the object of interest. The contour is represented with a parametric form, which can be called explicit contour representation. Deformable models [4] improve the accuracy of the snake model by integrating prior knowledge of the objects, such as intensity, texture, colour, orientation, location and shape into the segmentation process. In general, for deformable models, if the initial guess is not close enough

to the boundary ground truth, then the external energy cannot force the contour to move to the correct place. Cohen et al. [5, 6] augmented the accuracy of deformable models by changing the external term to include an inflation force, which makes this model more robust to the initial guess of the segmentation contour. Another solution for deformable models relating to the initial guess and the concavity of the object border is introduced by Xu and Prince [7]. This model incorporates a new class of external forces that are named gradient vector flow (GVF) fields, which represent dense vector fields computed from image features (e.g., gray level or edges) that minimise an energy functional. The segmentation result is improved, especially at concave areas of the segmentation boundary, and the final result is less sensitive to initial guess.

Another variety of active contour models is the level set method [8], which has been proposed in order to improve the performance of active contour techniques. One of the main differences of level sets is the use of the implicit contour representation, where the contour of the object is embedded as the zero level set of a surface. Level set improves over previous active contour models by allowing topological changes of the object and reducing the dependence on a parametric curve representation. Cremers et al. [9] presents an overview of applying level set methods to image segmentation, where they show that one advantage of level set methods is that they can easily integrate many types of prior information such as grey values, colour, texture and motion. Furthermore, by using Bayesian inference, it is also possible to incorporate prior knowledge about the shape of the object as a constraint in the level set evolution. In general, active contour models depend heavily on the a priori information about the segmentation, which has to be defined such that it can provide a robust representation for the shape and appearance variations present in the data. The main trouble with this approach is that this is an extremely complex task, which is rarely achieved successfully in practice.

In recent years, machine learning approaches have produced state-of-the-art results in segmentation problems in medical image analysis, by exploring large datasets annotated by clinicians (these annotated datasets are known as training sets). One typical example of a machine learning approach is the active shape model (ASM) [10]. ASM is a statistical model of the shape of objects estimated from an annotated training set. The shape of an object is represented by a set of contour points (landmarks), which is used by ASM to learn the shape model. Active appearance model (AAM) [11–13] improves ASM with the use of both the shape and texture models, also estimated from training sets. Other machine learning approaches are based on discriminative models for segmentation that use large datasets of annotated examples [14], such as boosting [15, 16], and support vector machines [17–20]. Machine learning techniques face a potential large inference complexity because of the search process. Specifically, the segmentation contours are usually explicitly represented by a set of contour points, which defines the dimensionality of the search space, and since contours generally need a large number of points, the dimensionality of the search space will become proportionally large [21]. Machine learning

models also typically make little use of a priori information in a segmentation process, which means that these models require extensive training sets to work robustly.

Some other machine learning techniques are based on graphical models. For instance, image segmentation can be considered as a classification task that assigns labels (e.g., objects and background) for observable elements of the image (e.g., pixels or voxels). Graphical models represent a suitable tool to model the dependencies between labels and observable data, where the label of an observable element depends not only on the corresponding observed data, but also the labels of its neighbouring nodes. Markov random field (MRF) [22–27] is a well-known undirected graphical model used in image segmentation that can be classified as a generative model, which estimates the joint probability distribution of labels and observable data. Another graphical model that has been explored in image segmentation is conditional random field (CRF) [28–33]. The difference between MRF and CRF is that the latter is based on a discriminative model that measures the conditional probability of labels given observable data. Similarly to the machine learning methods presented before, graphical model-based methods also need large annotated datasets, but the fact that they explore prior information (e.g., label consistency among neighbouring pixels) reduces this requirement to a certain extent.

Atlas-based segmentation [34–38] is another popular approach in medical image segmentation. The main idea explored in atlas-based segmentation is that the visual objects (e.g., tissues, organs) in medical images are reasonably consistent in terms of shape and localization. This allows the use of a reference model (manually built or automatically learned from training data) of the sought visual object that is used as a template. The segmentation process usually works by registering this template model to the test image using a non-rigid deformation model. However, this technique also faces some challenges, such as the variability of anatomical structure and scale in medical images, and the complexity of the registration process.

Some other works combine machine learning techniques with deformable models, which use either explicit or implicit contour representations, such as the integration of MRFs and explicit deformable models [23, 24], CRF and implicit deformable models [29], and embedded CRF in level set methods [30, 39]. The combination of CRF and SVM is described in [20, 32]. Nevertheless, these hybrid models are usually quite complex in terms of running time inference and they also need large annotated training sets.

1.1 Motivation

This thesis proposes a new model for medical image segmentation by combining the distance regularised level set method (DRLS) [40] with the deep belief network (DBN) [41]. This combination is relevant in the context of medical image segmentation because it tries to aggregate

the advantages of both techniques: 1) the high accuracy segmentation results produced by level set methods using virtually no annotated training sets, and 2) the robustness of DBNs to small datasets, with the use of unsupervised pre-training. More specifically, we use the DBN model result as a prior appearance term for the level set optimisation, which addresses one of the issues identified above for level set methods: the difficulty in hand-designing terms that model robustly the shape and appearance variations of the visual object of interest. Another interesting advantage with the specific use of DBN lies in its need of relatively small annotated training sets, compared with other machine learning models [23, 24, 29, 30, 39]. This advantage is based on the fact that the training of a DBN model involves both unsupervised and supervised learning: an unsupervised learning stage that uses massive amounts of un-annotated training data and a supervised stage that fine-tunes the learning process to converge with a small number of annotated images. Finally, level set method can incorporate other prior shape and appearance terms into its optimisation.

The proposed hybrid model is tested in two datasets with different imaging techniques: endocardial and epicardial left ventricle segmentation in MR images (Medical Image Computing and Computer Assisted Intervention (MICCAI) 2009 LV segmentation challenge database [42]) and lung segmentation in chest X-ray images (Japanese Society of Radiological Technology (JSRT) dataset [43]). Our experiments show that we can train robust DBN models with a limited number of annotated training images. In case of semi-automated segmentation (see Chapter 4 and 7), our proposed method produces the best results in the field in both datasets, and for the case of fully automated LV segmentation (see Chapter 5 and 6), the result is on par with other current state of the art results on the same dataset. We also show in the experiments that the DBN inference produces a segmentation based on the appearance information that improves significantly the result of DRLS model.

The motivation for using these datasets is based on the following reasons: 1) we want to show that our method can work with different organ appearances and shapes and imaging modalities; and 2) the importance of these two segmentation problems, as explained in more detail below. Cardiovascular disease is the current leading cause of deaths in the world, and cardiac cine magnetic resonance (MR) imaging is one of the most effective imaging for diagnosing heart disease [44]. Cardiac cine MR is a non-invasive medical imaging technology that is used for assessing the function and structure of the cardiovascular system, with the following advantages (compared to other imaging technologies, such as ultrasound or computerised tomography): image quality, non-invasiveness, accuracy, and no ionising radiation. The assessment of function and structure is based on the segmentation of several heart structures (see Fig. 1.1), like the left ventricle (LV) endocardium and epicardium borders. This segmentation is then used for the computation of the LV volume during the end systole (ES - greatest contraction) and end diastole (ED - greatest expansion) phases of the cardiac cycle, where the ratio of these volumes is then used to compute the ejection fraction, which is useful to assess the health of the heart.

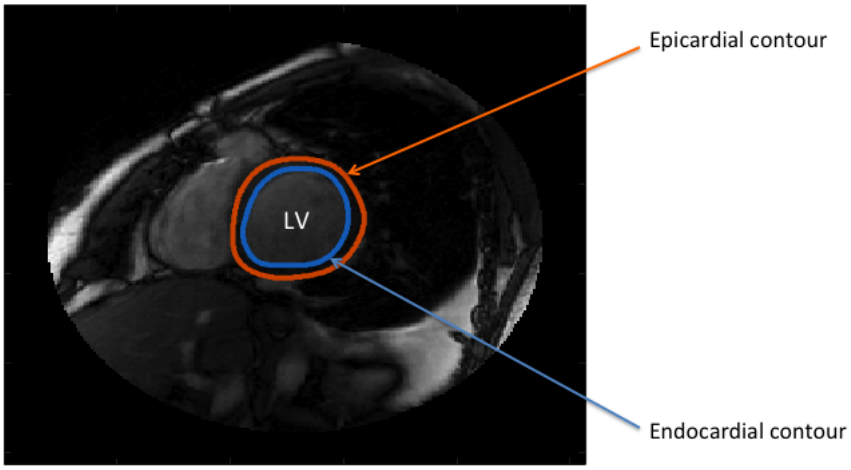


Figure 1.1: Left ventricle segmentation.

Since the manual segmentation of these structures has proven to be tedious, time consuming and subjective (i.e., it lacks reproducibility), several automated segmentation methodologies have been proposed [45] and tested in a challenge recently proposed by Radau et al. [42].

Another medical image segmentation that is used to test our methodology is the lung segmentation from chest X-ray (CXR). The automated segmentation of lung boundaries from digital CXR is one of the main stages in the computer-aided diagnosis (CAD) of lung health [36]. Lung boundaries can be used for computing lung volume or estimating shape irregularities [46], but it is also used as one of the stages in several CAD systems [47]. These CAD systems are particularly important for screening and detecting pulmonary pathologies, but with a major focus on tuberculosis, which is the second leading cause of death from infectious disease worldwide [48]. The highest incidences of these diseases occur in places of the world with inadequate health care infrastructure, so the deployment of such CAD systems in these places is important because they can help local clinicians in the screening and diagnosis processes mentioned above [36]. However, the automated segmentation of lung boundaries is a challenging task because of the following reasons (see Fig. 1.2) [36]: 1) the edges present at the rib cage and clavicle represent a challenge for optimisation methods that can get stuck at local minima; 2) the appearance inconsistencies caused by the clavicle bone at the lung apex also represent an issue for most optimisation approaches for the same reason above; and 3) the lack of a consistent lung shape among different individuals is a challenge for the use of shape priors.

1.2 Contributions of This Thesis

The main contributions of my thesis are as following:

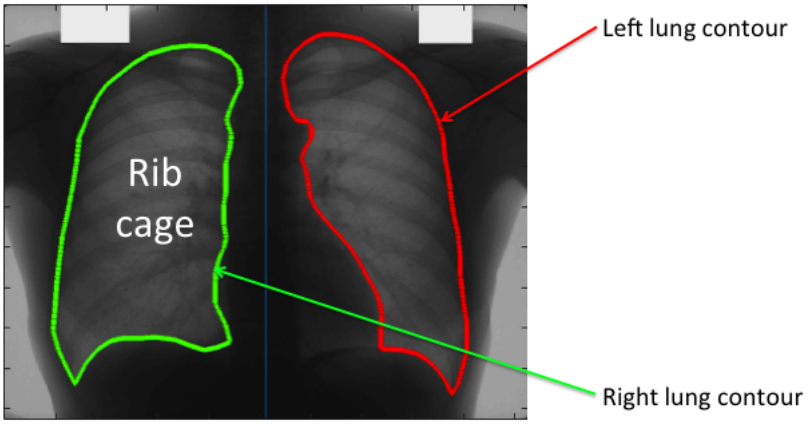


Figure 1.2: Lung segmentation.

1. An appearance model learned with a structured output DBN model that is used to detect the region of interest (ROI) containing the LV directly from grey-value image using structured inference. This is presented in Chapter 5 and 6.

2. Another structured output DBN model that is used to segment anatomical organs that will be used as an appearance based prior in the level set evolution. The structured outputs for endocardial and epicardial borders are presented in Chapter 4, 5 and 6 and for the left and right lung borders are shown in Chapter 7.

3. An extension of the distance regularised level set method (DRLS) [40] with the integration of the results of the DBN output from the innovation (2) with the level set evolution. This combination significantly improves the final segmentation results of level set, which is mentioned in Chapter 4, 5, 6 and 7.

1.3 Thesis Outline

In Chapter 2, we review robust segmentation techniques, which have been applied in medical image segmentation. In this chapter, we divide these techniques into five groups: image-based approaches, active contour methods, machine learning techniques, atlas-guided segmentation and hybrid models. Our proposed methodology is explained in Chapter 3. In Chapters 4, 5 and 6, we show the application of the proposed methodology to the segmentation of endocardial and epicardial borders of the left ventricle of the heart from cardiac cine magnetic resonance imaging (MRI). In these chapters, we show the results from semi-automated segmentation (Chapter 4) and fully automated segmentation (Chapters 5 and 6). The experiment results show that the combination of level set and DBN leads to competitive accuracy results on the MICCAI 2009 LV

segmentation challenge database [42]. In Chapter 7, we also assess the proposed methodology in the lung segmentation problem from chest radiographs. In this chapter, we only show the results for the semi-automated segmentation problem, which uses manual initial guess. The evaluation of the accuracy of our methodology uses the publicly available Japanese Society of Radiological Technology (JSRT) dataset [43]. Finally, discussion and conclusions are presented in Chapter 8.

Chapter 2

Medical Image Segmentation Techniques

We can divide medical image segmentation techniques into five categories: image-based approaches (e.g., thresholding), active contour methods (e.g., snake or level set methods), machine learning techniques (e.g., pixel classification, region classification, Markov random field, and active shape and appearance models), atlas-guided segmentation and registration, and hybrid models. In this chapter, we provide a brief introduction to these methods in order to motivate our proposed hybrid model.

2.1 Image-based Approaches

Image-based methods use techniques that consider only grey values or texture features to segment images, which usually involves the estimation of a threshold that separates (or segments) the foreground objects from the background. Thresholding techniques can be classified as local or global, where global techniques are sub-divided into point-dependent and region-dependent techniques [1]. If the threshold detection is based on features extracted from individual image pixels (e.g., grey value), then it is point-dependent thresholding. On the other hand, region-dependent methods replace the individual pixel information by a distribution involving a group of neighbouring pixels, and use that information for the thresholding process. In global thresholding, the goal is to find the thresholds for the whole image, whereas local thresholding is applied to local parts of the image [1]. In this section, we discuss Otsu's method [2] because it provides a good representation of image-based approaches and also because we apply it in a couple of stages of our methodology.

Otsu [2] proposed a method using a discriminant criterion, by assuming that a grey-value image is represented by $I : \Omega \to \mathbb{R}$, with $\Omega \subseteq \mathbb{R}^2$ denoting the image coordinate space. $I(x, y)$ denotes the intensity value (e.g., grey level) of the pixel with coordinate (x, y). Let $\mathcal{M} = \{1, 2, 3, \cdots, l\}$ represent the grey levels for image I. Given n_i is number of pixels at level $i \in \mathcal{M}$, the total number of pixels in image I is

$$n = \sum_{i=1}^{l} n_i. \tag{2.1}$$

We can then compute the probability distribution of pixels at grey level i as

$$p_i = n_i/n, \tag{2.2}$$

where $p_i \geq 0$, $\sum_{i=1}^{l} p_i = 1$. Suppose that we would like to classify the pixels of image I into two classes: foreground object with grey levels $\mathcal{C}_1 = \{1, 2, \cdots, m\}$ and background with grey values $\mathcal{C}_2 = \{m + 1, m + 2, \cdots, l\}$ by using a threshold at level m. This threshold is automatically estimated using the steps below. The probabilities of occurrence for the two classes, given the threshold m, are:

$$\omega_1(m) = \sum_{i=1}^{m} p_i \text{ and } \omega_2(m) = \sum_{i=m+1}^{l} p_i = 1 - \omega_1(m). \tag{2.3}$$

The probabilities in (2.3) allow us to compute the mean grey value for both classes as follows:

$$\mu_1 = \sum_{i=1}^{m} ip_i/\omega_1(m) \text{ and } \mu_2 = \sum_{i=m+1}^{l} ip_i/\omega_2(m), \tag{2.4}$$

and the mean grey value for whole image:

$$\mu_T = \sum_{i=1}^{l} ip_i. \tag{2.5}$$

After computing the mean values, the class variances are computed with:

$$\sigma_1^2 = \sum_{1}^{m} (i - \mu_1)^2 p_i/\omega_1 \tag{2.6}$$

$$\sigma_2^2 = \sum_{m+1}^{l} (i - \mu_2)^2 p_i/\omega_2. \tag{2.7}$$

Otsu then proposes the following discriminant criterion factor [2]:

$$\lambda = \sigma_B^2/\sigma_W^2, \tag{2.8}$$

where σ_B^2 and σ_W^2 in (2.8) are referred to as the within-class and between-class variances, respectively, which are computed as:

$$\begin{aligned}\sigma_W^2 &= \omega_1\sigma_1^2 + \omega_2\sigma_2^2,\\ \sigma_B^2 &= \omega_1(\mu_1 - \mu_T)^2 + \omega_2(\mu_2 - \mu_T)^2\\ &= \omega_1\omega_2(\mu_2 - \mu_1)^2.\end{aligned} \tag{2.9}$$

Finally, the optimal threshold m^* is found by maximising λ, as in

$$m^* = \arg\max_{1\leq m\leq l} \lambda \tag{2.10}$$

The results of thresholding techniques are usually not precise enough, which is the reason why these techniques are commonly combined with other more sophisticated segmentation techniques.

2.2 Active Contour Methods

In this section, we discuss active contour methods with explicit contour representation, also known as the Snake method [3], and with implicit contour representation, referred to as level set methods [8].

2.2.1 Active Contours with Explicit Contour Representation

Active contour methods with explicit contour representation is generally referred to as the Snake model [3]. The goal of Snake is to minimise an energy function in order to segment the foreground object from the background. Given an image $I : \Omega \to \mathbb{R}$, the segmentation is obtained with the contour $\mathbf{c}$ that denotes the explicit contour representation, defined as follows:

$$\begin{aligned}\mathbf{c} : [0,1] &\to \boldsymbol{\Omega}\\ s &\mapsto \mathbf{c}(s) = (x(s), y(s))\end{aligned} \tag{2.11}$$

The energy function associated with the contour is:

$$E(\mathbf{c}) = E_{int}(\mathbf{c}) + E_{ext}(\mathbf{c}), \tag{2.12}$$

where

$$E_{int}(\mathbf{c}) = \int_0^1 (\alpha(s)|\mathbf{c}_s(s)|^2 + \beta(s)|\mathbf{c}_{ss}(s)|^2)ds \tag{2.13}$$

denotes the internal energy of the contour due to the amount of stretch (first term) and the amount of curvature (second term) in the contour, with $\mathbf{c}_s$ and $\mathbf{c}_{ss}$ denoting the first and second order derivatives of $\mathbf{c}$ with respect to s (that is, the first-order term in E_{int} makes the contour act like a membrane and the second-order term makes it act like a thin plate); and

$$E_{ext}(\mathbf{c}) = \int_0^1 w_{\text{line}} E_{\text{line}} + w_{\text{edge}} E_{\text{edge}} \tag{2.14}$$

represents the external energy of the contour, which $E_{\text{line}} = I(x, y)$ is attracted to either dark or light regions of the image, depending on the sign of w_{line}, and $E_{\text{edge}} = -| \bigtriangledown I(x, y)|^2$ is drawn to regions with large image gradients. Note that in order to find the sought contour that segments the image, we need to solve the following energy minimisation problem:

$$\mathbf{c}^* = \arg\min_{\mathbf{c}} E(\mathbf{c}). \tag{2.15}$$

We can also define the energy in (2.12) using a probabilistic framework, by defining the joint probability $p(I, \mathbf{c})$ as a Gibbs distribution with energy function [49]:

$$p(I, \mathbf{c}) = p(I|\mathbf{c})p(\mathbf{c}) = k \exp\left(-E_{int}(\mathbf{c}) - E_{ext}(\mathbf{c})\right) = k \exp\left(-E(\mathbf{c})\right), \tag{2.16}$$

where k is a normalising constant and $p(.)$ is a probability density function defined by $p(x) \geq 0$ and $\int_{-\infty}^{+\infty} p(x)dx = 1$. Minimising the energy function of the contour in (2.15) is equivalent to maximising the joint probability $p(I, \mathbf{c})$ in (2.16). The internal energy $E_{int}(\mathbf{c})$ is converted to Gibbs prior distribution by the following formulation [24]:

$$p(\mathbf{c}) = \frac{1}{Z_{int}} \exp\left(-E_{int}(\mathbf{c})\right), \tag{2.17}$$

where Z_{int} is a normalisation factor. Following the same idea, the external energy $E_{ext}(\mathbf{c})$ is converted into the conditional probability:

$$p(I|\mathbf{c}) = \frac{1}{Z_{ext}} \exp\left(-E_{ext}(\mathbf{c})\right), \tag{2.18}$$

where Z_{ext} is a normalisation factor. Finally, the minimisation of the energy function of deformable models in (2.15) can be derived by solving the maximum a posteriori (MAP) problem:

$$\mathbf{c}^* = \arg\max_{\mathbf{c}} p(\mathbf{c}|I), \tag{2.19}$$

where

$$p(\mathbf{c}|I) \propto p(I|\mathbf{c})p(\mathbf{c}). \tag{2.20}$$

The solution to this optimisation problem is usually based on gradient descent, which means that it is prone to local minima given that the optimisation in (2.19) is non-convex in general.

Although relatively successful in some specific tasks, active contours with explicit contour representation are sensitive to image conditions, priors, and the initialisation to the optimisation process in (2.15) and (2.19). In particular, the prior models assumed in both energy terms in (2.12) are unlikely to capture the variations presented in most medical image segmentation problems because of the usual large variability found in the foreground objects in terms of appearance and shape. Moreover, this method is not adaptable to foreground objects that change topology. Finally, the parameterisation of the curve $\mathbf{c}$ is another weak point of this type of active contour model because it has to be compact and at the same time, represent well the contour details.

2.2.2 Active Contours with Implicit Contour Representation

The issues mentioned above with active contours with explicit contour representation motivated the development of active contour models with implicit contour representation, which are known as level set methods [8]. This approach uses an implicit contour representation that can adapt for topological changes of the foreground object, such as splitting and merging. Furthermore, this implicit representation avoids altogether the issue with the curve parameterisation. The main idea of segmentation based on level set methods is to embed an initial curve as the zero level set of a higher dimensional surface. The segmentation process evolves the surface in order to propagate the zero level set, which converges to the boundary of the object [50–52]. The level set formulation also allows the introduction of several new terms, such as the shape and region models (estimated from training sets) [9], which is another advantage of this segmentation approach.

The level set method takes an image $I : \Omega \rightarrow \mathbb{R}$ and produces a segmentation represented by the zero level set of an embedding function $\phi : \Omega \rightarrow \mathbb{R}$, as follows:

$$\mathcal{C} = \{\mathbf{x} \in \Omega | \phi(\mathbf{x}) = 0\}\,, \tag{2.21}$$

where the function ϕ is usually assumed to be the signed distance function, defined by [29]:

$$\begin{aligned} \phi(\mathbf{x}) &= \begin{cases} 0 & \mathbf{x} \in \mathcal{C} \\ +d(\mathbf{x}, \mathcal{C}) & \mathbf{x} \in \Re_{\mathcal{C}}, \\ -d(\mathbf{x}, \mathcal{C}) & \mathbf{x} \in \Omega \setminus \Re_{\mathcal{C}}, \end{cases} \\ d(\mathbf{x}, \mathcal{C}) &= \min \{\| \mathbf{x} - \mathbf{x}_{\mathcal{C}} \|_2 | \, \mathbf{x}_{\mathcal{C}} \in \mathcal{C}\}\,, \end{aligned} \tag{2.22}$$

where $\| \mathbf{x} - \mathbf{x}_{\mathcal{C}} \|_2$ is Euclidean distance between two pixels $\mathbf{x}$ and $\mathbf{x}_{\mathcal{C}}$. We assume that the contour $\mathcal{C}$ divides the image domain Ω into two parts, namely the foreground $\Re_{\mathcal{C}}$ enclosed by the contour $\mathcal{C}$ and the background $\Omega \setminus \Re_{\mathcal{C}}$. The result of active contour optimisation can be

reached by solving the following energy minimisation problem:

$$\phi^* = \arg\min_{\phi} E(\phi), \tag{2.23}$$

where this energy function takes into account contour specific measures (length, curvature, etc.) and image-based measures (e.g., edges, region grey values distribution, etc.). For example, a possible energy function to minimise is the following [30, 53]:

$$E(\phi) = \int_{\Omega} -H(\phi) \log\, p_1(f(\mathbf{x}), \mathbf{w}_1) - (1 - H(\phi) \log\, p_2(f(\mathbf{x}), \mathbf{w}_2) + v|\nabla H(\phi)| g(I, \alpha) d\mathbf{x}, \tag{2.24}$$

where $f(\mathbf{x})$ defines the feature values (e.g. image gradient, colour, intensity) at each image location, the functions p_1 and p_2 are probability functions of the random process $f(\mathbf{x})$ in $\Re_{\mathcal{C}}$ and $\Omega \setminus \Re_{\mathcal{C}}$ respectively, with parameters $\mathbf{w}_1$ and $\mathbf{w}_2$. In the last term of (2.24), $g(I, \alpha) = \frac{1}{1+\alpha|\nabla I|}$ denotes the non-decreasing function with the parameter α that takes low values at image edges, and $|\nabla H(\phi)|$ represents a contour smoothness prior term, where the parameter v controls the strength of this term. Finally in (2.24), the function $H(\phi)$ is the Heaviside step function, defined by

$$H(\phi) = \begin{cases} 1 & \phi \geq 0 \\ 0 & otherwise. \end{cases} \tag{2.25}$$

In summary, the first two terms in (2.24) model the inside and outside regions of the contour and the last term models the smoothness of the contour (with the edge attraction). Also note that similarly to (2.15), the energy minimisation problem in (2.23) can also be converted into a MAP problem, as in

$$\phi^* = \arg\max_{\phi} p(\phi|I), \tag{2.26}$$

with

$$p(\phi|I) \propto p(I|\phi)p(\phi). \tag{2.27}$$

In order to minimise the energy function E in (2.24), the common approach [54] is to find the steady state solution of the following gradient flow equation:

$$\frac{\partial \phi}{\partial t} = -\frac{\partial E}{\partial \phi}, \tag{2.28}$$

where $\frac{\partial E}{\partial \phi}$ is the Gâteaux derivative of the functional $E(\phi)$, and t is temporal parameter. The main idea of this method is then to iteratively follow the steepest descent direction of the functional $E(\phi)$.

Active contour methods with implicit contour representation produce some remarkable results,

such as topology robustness, less dependence on an initial contour, independence of curve parameterisation, and integration of local and region distributions. However, these methods still suffer from several drawbacks, namely, sensitiveness to image conditions and the a priori information about the segmentation process. Specifically, the shape and appearance models are usually based on prior information that are hand designed, and is unlikely to cover all variation present in the real data. In addition, level set methods generally can only represent closed contours, except the method proposed by Schaeffer [55].

2.3 Machine Learning Techniques

The main idea behind machine learning methods is the use of a manually annotated training set to estimate the parameters of high capacity models about shape and appearance information. This implies that it is necessary to first adopt a model that has a capacity that is large enough to represent the variations of the foreground object being studied. Moreover, the capacity of this model will determine the amount of training data that is needed to estimate robustly the model parameters, so it can generalize well to unseen test data. In general, the higher the capacity, the better the model can represent the variations in the dataset, but the larger the annotated dataset set needed. Therefore, machine learning methods solve the problem of hand-designing the shape and appearance prior models (of the active contour models above) with an automatic learning of such models, where the cost is the need for large annotated datasets.

2.3.1 Boosting Methods

Discriminative classifiers are machine learning techniques, which model the posterior distribution directly. In medical image segmentation, assuming that $\mathbf{x}$ represents an observable information (e.g., an image region) and $\mathbf{y}$ denotes the corresponding labeling (e.g., foreground or background pixels of the image $\mathbf{x}$), the discriminative classifier models the conditional probability distribution $p(\mathbf{y}|\mathbf{x})$. In contrast, generative classifier models the joint distribution $p(\mathbf{y}, \mathbf{x})$. Some discriminative classifier techniques have been applied to medical image segmentation, such as support vector machines (SVMs) [18–20], neural networks [56–59], and boosting techniques [16, 60–62]. Below, we explain the application of boosting in a particular medical image analysis problem.

Boosting methods, one group of discriminative models, are generally known as ensemble methods that are formed by combining a set of weak classifiers in order to produce a stronger classifier [15, 63, 64]. In this section, we explain the method proposed by Carneiro et al. [61], where the authors use the probabilistic boosting tree (PBT) [60] to detect and segment fetal anatomies from ultrasound images automatically. More specifically, given a image $I : \Omega \rightarrow \mathbb{R}$, assume

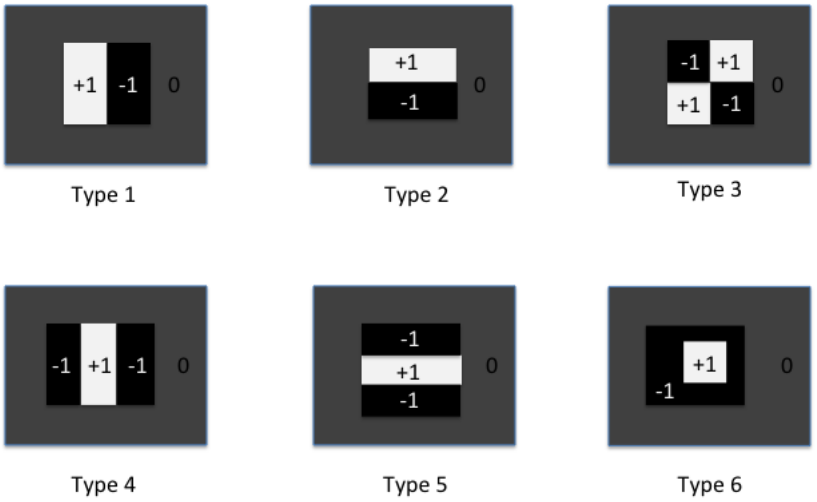


Figure 2.1: Image feature types.

that $S \subset I$ denotes the region of interest (i.e., the foreground) containing the fetal anatomy, and $B \subset I$ represents the background, where $S \cup B = I$ and $S \cap B = \emptyset$. The geometrical properties of the foreground region S is represented by a vector of parameters:

$$\theta = [x, y, \alpha, \sigma_x, \sigma_y], \tag{2.29}$$

where (x, y), α, (σ_x, σ_y) denote the top-left coordinate, orientation and scale of the region of interest S, respectively. The appearance of image region S is represented by the following set of Haar wavelet features:

$$\theta_f = [t, x_f, y_f, d_x, d_y, s], \tag{2.30}$$

where t is type of image features (see Fig. 2.1 [61]), (x_f, y_f) is the top-left coordinate of feature within S, (d_x, d_y) denotes the length and width of feature, and $s \in \{-1, +1\}$ is the original feature or its inverted sign. Both orientations of θ and θ_f are the same. Essentially, the Haar wavelet features in Fig. 2.1 assigns a feature value representing the difference between the number of pixels in white area (value +1) and in black area (value -1).

The inference process to find the region of interest S consists of:

$$\theta^* = \arg\max_{\theta} p(y|S), \tag{2.31}$$

where $y \in \{-1, +1\}$, $p(y|S)$ is a PBT classifier [60] that computes the probability of the appearance of fetal anatomies ($y = +1$) or background ($y = -1$) in the interest region S, extracted based on the geometric parameter θ. Basically, PBT is a binary tree (see Fig. 2.2 [61]), where each of its node is a strong classifier that can be trained using the AdaBoost algorithm [15],

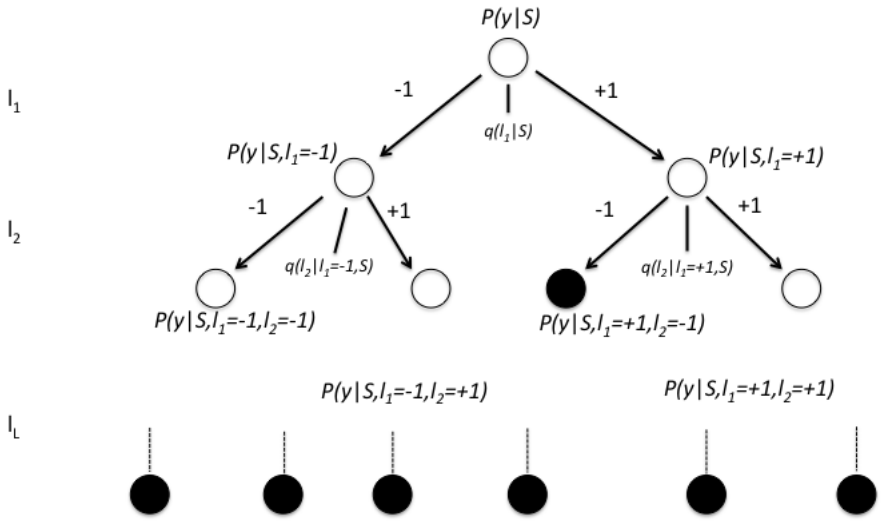


Figure 2.2: PBT binary tree.

which learns a strong classifier by combining a set of weak classifiers:

$$H(S) = \sum_{t=1}^{T} \omega_t h_t(S), \tag{2.32}$$

where $h_t(S)$ is a weak classifier and ω_t is its corresponding weight.

In the training process of the PBT, the binary tree is constructed recursively, where the output of each strong classifier is computed with the following equations:

$$q(+1|S) = \frac{e^{2H(S)}}{1 + e^{2H(S)}}, \text{ and } q(-1|S) = \frac{e^{-2H(S)}}{1 + e^{-2H(S)}}. \tag{2.33}$$

The posterior probability that the region S is labelled as foreground or background is computed with [60]:

$$p(y|S) = \sum_{l_1, l_2, \ldots, l_n} p(y|l_n, \ldots, l_1, S) \ldots q(l_2|l_1, S) q(l_1|S), \tag{2.34}$$

where $l \in \{-1, +1\}$, and

$$p(y|l_i, \ldots, l_1, S) = \sum_{l_{i+1}} \delta(y = l_{i+1}) q(l_{i+1}|l_i, \ldots, l_1, S), \tag{2.35}$$

with $\delta(x)$ is the Dirac delta function.

Similarly to other discriminative classifiers presented in the field [18–20, 56–60, 62], the approach by Carneiro et al. [16, 61] is able to solve the important problem of providing robust appearance and shape models estimated from annotated training data. However, the tradeoff is

the fact that a very large training set is needed to make this methodology work properly (in their paper they mention the need for thousands of annotated data). The main issue is that such large training sets are rarely available in medical image analysis problems. Furthermore, the search space in the inference problem in (2.31) can be large due to the number of dimensions of θ and the data distribution in that space, resulting in a potentially slow inference.

2.3.2 Markov and Conditional Random Fields

Image segmentation can be considered a pixel labelling task that can be solved with the Markov Random Field (MRF) model, which is one of the most explored machine learning techniques in medical image segmentation [22–25]. MRF is a probabilistic graphical model that uses undirected graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ to label image pixels, exploring the information present locally at each pixel location and semi-locally in the vicinity of each pixel (see Fig. 2.3 [24]). This graph uses two types of nodes to represent the variables: the circles representing the hidden labels (e.g. object and background) and squares denoting the observed image pixel values (e.g. grey values of pixels). Assume that n is the number of pixels of the image I, $\mathcal{L}$ is the set of region labels (e.g. $\mathcal{L} = \{$"foreground", "background"$\}$) and $\mathcal{D}$ is the set of pixel values (e.g. grey values). The variables of the observed layer are defined as:

$$\mathbf{x} = (x_1, x_2, \cdots, x_n), x_i \in \mathcal{D}, i = 1, ..., n, \tag{2.36}$$

and the variables of the hidden labels are represented by:

$$\mathbf{y} = (y_1, y_2, \cdots, y_n), y_i \in \mathcal{L}, i = 1, ..., n, \tag{2.37}$$

where n denotes the number of nodes in observed/hidden layer of the model. The MRF essentially models the joint probability between $\mathbf{x}$ and $\mathbf{y}$ with

$$p(\mathbf{x}, \mathbf{y}) = \frac{1}{Z} \prod_{s \in \mathcal{F}} \psi_a(\mathbf{x}_a, \mathbf{y}_a), \tag{2.38}$$

where Z is a normalisation constant, $s \in \mathcal{F} \subset (\mathbf{x} \bigcup \mathbf{y})$ is an index to a graph clique, and $\psi_a : \mathcal{D} \times \mathcal{L} \rightarrow \mathbb{R}^+$. Note in (2.38) that we model the joint probability between $\mathbf{x}$ and $\mathbf{y}$, but we are only interested in finding the label assignment $\mathbf{y}$ given the image represented by $\mathbf{x}$. This is the main reasoning behind the development of conditional random field, which models only the conditional probability $p(\mathbf{y}|\mathbf{x})$.

Conditional Random Field (CRF) is a type of MRF represented by a probabilistic graphical model that facilitates computation of the posterior distribution $p(\mathbf{y}|\mathbf{x})$, i.e., conditional probability of labelling layer $\mathbf{y}$ given observable layer $\mathbf{x}$. Note that while MRF is a generative

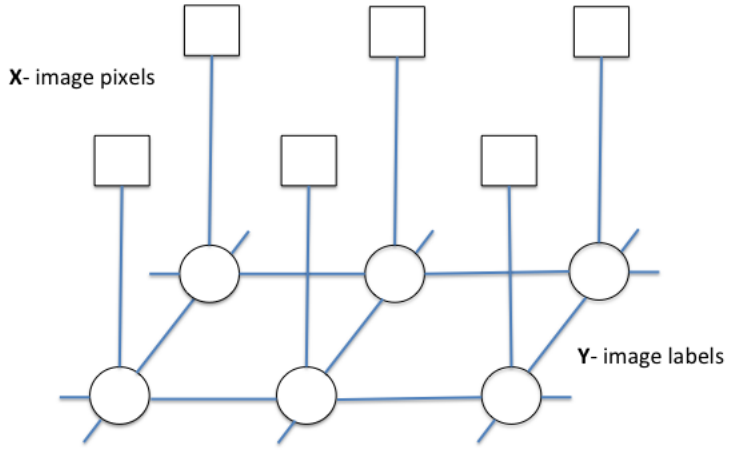


Figure 2.3: Markov random field model.

model, CRF is a discriminative model. Specifically, MRF approximates the posterior distribution $p(\mathbf{y}|\mathbf{x})$ by modelling the joint probability distribution $p(\mathbf{y}, \mathbf{x})$, while CRF models the posterior distribution $p(\mathbf{y}|\mathbf{x})$ directly. CRF was introduced by Lafferty et al.[28] and has been used in medical image segmentation in several problems [29–31].

CRF definition [28]: Given a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where $\mathbf{y} = (\mathbf{y}_v)_{v \in \mathcal{V}}$ is indexed by vertices of graph $\mathcal{G}$. Then $(\mathbf{y}, \mathbf{x})$ is a CRF, if the conditional probability of $\mathbf{y}$ given $\mathbf{x}$ satisfies the Markov property: $p(\mathbf{y}_v|\mathbf{x}, \mathbf{y}_w, w \neq v) = p(\mathbf{y}_v|\mathbf{x}, \mathbf{y}_w, w \sim v)$, where $w \sim v$ means that w and v are neighbours in $\mathcal{G}$. By applying the Hammersley Clifford theorem, the conditional probability $p(\mathbf{y}|\mathbf{x})$ can be expressed as a factorization over cliques as following:

$$p(\mathbf{y}|\mathbf{x}) = \frac{1}{Z} \exp \left\{ \sum_{i \in S} \phi_a(y_i, \mathbf{x}) + \sum_{i \in S} \sum_{j \in \mathcal{N}_i} \phi_i(y_i, y_j, \mathbf{x}) \right\}, \tag{2.39}$$

where S is number of sites, $\mathcal{N}_i$ is the set neighbours of the site i, Z is normalisation constant, and ϕ_a and ϕ_i are association and interaction potential functions, respectively. The association potential function ϕ_a describes the correlation between a given site and a specific class without neighbouring information, and the interaction term ϕ_i models data dependency.

The number of possible labelling configurations in (2.39) is $|\mathcal{L}|^n$, where $\mathcal{L}$ is set of pixel labels and n is number of image pixels, so solving directly this optimisation is computationally expensive. Various approximating techniques were presented to solve this problem: simulated annealing [65], iterated conditional modes (ICM) [66], graph cuts [67, 68], etc. The most dominant inference algorithm for solving (2.39) is arguably graph cuts, which is based on the $s - t$ graph cut theory, where minimising an energy function is equivalent to maximising a flow in a graph. Graph cuts was proposed by Greig et al. [67], and applied to image segmentation problems [68, 69].

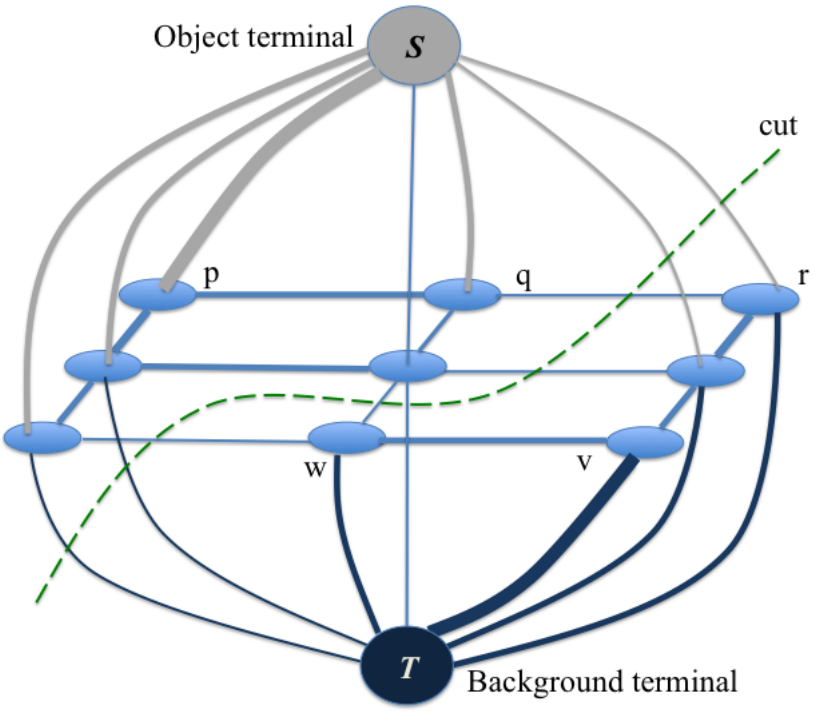


Figure 2.4: Undirected graph for image segmentation.

In graph cuts, $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ represents a graph, where each node in the vertex set $\mathcal{V}$ represents an image pixel or voxel, and a single edge $e = \{p, q\}$ in edge set $\mathcal{E}$ expresses neighboring nodes. Two terminal nodes, $S(source)$ and $T(sink)$, denoting object and background, respectively, are added (see Fig. 2.4 and 2.5 [69]). An $n - link$ represents an edge between two pixels, whereas a $t - link$ denotes an edge connecting an image pixel and a terminal node. Each edge $e \in \mathcal{E}$ is assigned a nonnegative weight w_e. A subset of edges $C \subset \mathcal{E}$ is called an $s - t$ cut if the terminals S and T are completely separated on the induced graph $\mathcal{G}(C) = (\mathcal{V}, \mathcal{E} \setminus C)$. The cost of the cut C is defined as:

$$|C| = \sum_{e \in C} w_e \tag{2.40}$$

Assume that $\mathcal{P}$ is the set of image elements (such as pixels or voxels) and $\mathcal{N}$ is the set of pairs $\{p, q\}$, where p and q are neighboring elements in $\mathcal{P}$. Let $\mathbf{y} = (y_1, y_2, \cdots, y_{|\mathcal{P}|}), y_i \in \mathcal{L} =$ {"foreground", "background"} be the labelling configuration, which is a segmentation result. The segmentation cost function is defined as following:

$$E(\mathbf{y}) = \lambda . R(\mathbf{y}) + B(\mathbf{y}), \tag{2.41}$$

where $R(\mathbf{y})$ and $B(\mathbf{y})$ are regional term and boundary terms, respectively:

$$\begin{aligned} R(\mathbf{y}) &= \sum_{p \in \mathcal{P}} R_p(y_p) \\ B(\mathbf{y}) &= \sum_{\{p,q\}} B_{p,q}.\delta_{y_p \neq y_q} \end{aligned} \tag{2.42}$$

where $\delta_{y_p \neq y_q}$ is Dirac delta function:

$$\delta_{y_p \neq y_q} = \begin{cases} 1 & \text{if } y_p \neq y_q \\ 0 & \text{otherwise.} \end{cases} \tag{2.43}$$

The regional term $R(\mathbf{y})$ aims to cope with how to assign a specific image element to "foreground" or " background", which uses negative log-likelihood form:

$$\begin{aligned} R_p(\text{"foreground"}) &= -\log\, p(I_p|\text{"foreground"}) \\ R_p(\text{"background"}) &= -\log\, p(I_p|\text{"background"}) \end{aligned} \tag{2.44}$$

On the other hand, the boundary term $B(\mathbf{y})$ explains the discontinuity between image elements p and q via coefficient $B_{p,q} \geq 0$. More specifically, the larger $B_{p,q}$ value means pixels p and q are more similar and this term is usually represented as:

$$B_{p,q} \propto \exp\left(-\frac{(I_p - I_q)^2}{2\sigma^2}\right)\frac{1}{dist(p,q)}, \tag{2.45}$$

where I_p and I_q are intensity values at image positions p and q, and $dist(p,q)$ is Euclidean distance between p and q.

The segmentation problem is solved by maximising the posterior probability $p(\mathbf{y}|\mathbf{x})$, which is equivalent to find minimum graph cut solution for graph $\mathcal{G}$ in order to minimise the segmentation cost function $E(\mathbf{y})$ in (2.41). Minimising the cut of the graph is transformed to maximising the flow from source S to sink T [70, 71], which allows the application of max-flow algorithms to segmenting the image, such as Ford-Fulkerson algorithm [72], Push-Relabel algorithm [73], and algorithms proposed by Boykov et al. [68, 69, 74]. Although graph cut is a useful technique for some computer vision applications, the graph construction is in general specific to particular applications and energy functions [75].

Conditional and Markov random fields have been successfully explored in the field in several medical image analysis applications, but they present two issues: 1) the learning of the graph model parameters (edge weights) and the potential function parameters also depends on large amounts of annotated training data [29–31], similarly to other discriminative learning problems; and 2) the large running time of the inference (e.g., graph cuts) limits the application of such

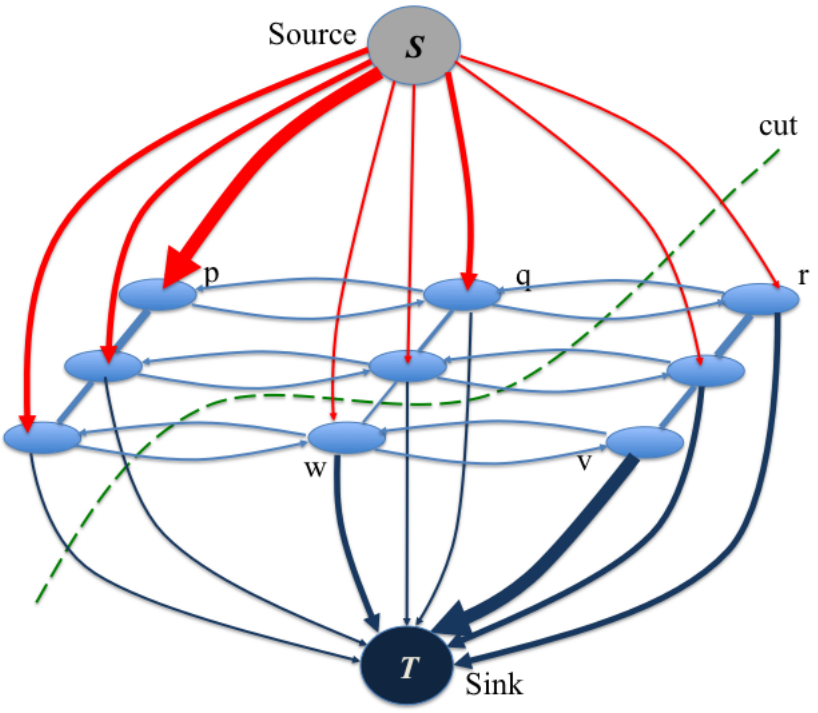


Figure 2.5: Directed graph for image segmentation.

methods to input images of relatively small sizes, which is not always possible in medical image analysis problems.

2.3.3 Active Shape and Appearance Models

Active Shape Model (ASM) is a 'top-down' method that automatically learns the object shape model to be applied in image segmentation problems. This approach differs from 'bottom-up' method that uses low level image features (e.g. pixel intensity, colour, edge) to segment the objects from background. ASM is a generative model that estimates the visual object shape model based on a set of training images. The segmentation is performed by searching for the model parameters in which the ASM produces the best plausible match with the test image. ASM has been proposed by Cootes et al. [10, 13], and have been applied successfully in many computer vision problems, especially in image segmentation [76–78]. ASM is represented by a point distribution model [10, 13], where each shape is represented as a vector of landmark points. Assume that $\mathcal{S} = \{\mathbf{x}_i\}_{i=1}^{N}$ is the training set containing N shapes, where each shape $\mathbf{x}_i \in \mathbb{R}^{dn}$ is represented by n landmarks points in d dimensions (note that we will consider that $d = 2$, but this method can easily be generalized to higher dimensions). For instance, each shape is represented by the following vector of coordinates:

$$\mathbf{x}_i = (x_{i0}, y_{i0}, x_{i1}, y_{i1}, \ldots, x_{ik}, y_{ik}, \ldots, x_{i(n-1)}, y_{i(n-1)})^{\mathbf{T}}. \tag{2.46}$$

Then, all shapes are aligned by applying the Procrustes method [79] in order to minimise the sum of distances between corresponding landmark point pairs in shape set $\mathcal{S}$, as explained below. Let $M(s, \theta)$ be a linear transformation combining a rotation θ, scale s, and translation $\mathbf{t} = (t_x, t_y)$. The goal is to minimise the following objective function, by assuming that we have two shapes $\mathbf{x}_i$ and $\mathbf{x}_j$, where the aim is to find the parameters θ_j, s_j and $\mathbf{t}_j = (t_{xj}, t_{yj})$ of the transformation $M(s_j, \theta_j)[\mathbf{x}_j] + \mathbf{t}_j$ that maps $\mathbf{x}_j$ onto $\mathbf{x}_i$:

$$D_j = (\mathbf{x}_i - M(s_j, \theta_j)[\mathbf{x}_j] - \mathbf{t}_j)^{\mathbf{T}} \mathbf{W} (\mathbf{x}_i - M(s_j, \theta_j)[\mathbf{x}_j] - \mathbf{t}_j), \tag{2.47}$$

where

$$M(s, \theta) \begin{bmatrix} x_{jk} \\ y_{jk} \end{bmatrix} = \begin{pmatrix} (s\cos\theta)x_{jk} - (s\sin\theta)y_{jk} \\ (s\sin\theta)x_{jk} + (s\cos\theta)y_{jk} \end{pmatrix}, \tag{2.48}$$

$$\mathbf{t}_j = (t_{xj}, t_{yj}, \ldots, t_{xj}, t_{yj})^{\mathbf{T}}, \text{and} \tag{2.49}$$

$\mathbf{W}$ is a diagonal weight matrix of points defined as follows: given that d_{kj} denotes the distance between two points $\mathbf{x}_k$ and $\mathbf{x}_j$ and $V_{d_{kj}}$ is variance of d_{kj} over the set $\mathcal{S}$, then the weight of the kth point is computed as:

$$w_k = (\sum_{j=0}^{n-1} V_{d_{kj}})^{-1}. \tag{2.50}$$

The following algorithm is used for aligning the set of shapes $\mathcal{S}$:

Algorithm 2.1: Aligning shapes [10]

- Choose an initial shape $\mathbf{x} \in \mathcal{S}$
- Apply transformations (e.g, rotation, scale, translation) to align each another shape with $\mathbf{x}$

repeat

- Compute the mean shape of all aligned shapes
- Normalise the current mean
- Realign every shape with the current mean

until the process converges

Suppose that N shapes have been aligned, so now each shape can be considered as a point in this space. The point distribution model (PDM) estimates the variance of these points in d dimensions with principle component analysis (PCA). The PDM can generate new points (i.e., new shapes) by varying the model's parameter values. The details of this algorithm is as following:

- Compute the mean shape

$$\bar{\mathbf{x}} = \frac{1}{N} \sum_{i=1}^{N} \mathbf{x}_i \tag{2.51}$$

• Compute $d \times d$ covariance matrix of the shapes,

$$\mathbf{C} = \frac{1}{N-1}\sum_{i=1}^{N}(\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^{\mathbf{T}} \tag{2.52}$$

• Compute the set of all eigenvectors $\{p_k|k = 1, \ldots, d\}$ and corresponding eigenvalues $\{\lambda_k|k = 1, \ldots, d; \lambda_k \geq \lambda_{k+1}\}$:

$$\mathbf{C}p_k = \lambda_k p_k. \tag{2.53}$$

Let $\mathbf{P} = (p_1|p_2|\ldots|p_t)$ represent the matrix of t first eigenvectors corresponding to the t largest eigenvalues, then each shape in $\mathcal{S}$ can be represented as follows:

$$\mathbf{x} = \bar{\mathbf{x}} + \mathbf{P}\mathbf{b}, \tag{2.54}$$

where $\mathbf{b} = (b_1, b_2, \ldots, b_t)$ is vector of weight parameters of the model.

By changing the value of parameter $\mathbf{b}$ in (2.54), the model can generate new shapes. In order to guarantee that the new generated shapes are similar to those in $\mathcal{S}$, the values of the parameter vector $\mathbf{b}$ are usually selected within ± 3 standard deviations of the mean as following:

$$-3\sqrt{\lambda_k} \leq b_k \leq 3\sqrt{\lambda_k}. \tag{2.55}$$

The inference process takes an image I containing the visual object that has a similar shape compared to the shape generated by the PDM model using (2.54) and estimates the values for the model parameters that generates a shape that matches the object shape in image I. Assume that $\mathbf{X}$ is the result of the model after a rotation by θ, scaling by s and translation by $\mathbf{t}_c$:

$$\mathbf{X} = M(s, \theta)[\mathbf{x}] + \mathbf{t}_c, \tag{2.56}$$

where $\mathbf{t}_c$ is the vector that translates shape $\mathbf{x}$ to the model centre $\mathbf{t}_c = (x_c, y_c, x_c, y_c, \ldots, x_c, y_c,)^{\mathbf{T}}$, with (x_c, y_c) representing the coordinates of the centre of the model shape. Furthermore, the boundary points of $\mathbf{X}$ are fit to the object in the image I by adjusting each point along a line by an amount $d\mathbf{X}$ to the model boundary to reach a more plausible position (see Fig. 2.6 [10]). Algorithm 2.2 explains each iteration of the ASM inference.

Algorithm 2.2: ASM algorithm [13]

- Searching a region around each boundary point $\mathbf{X}_i \in \mathbf{X}$ to get the best matching point $\mathbf{X}_i^{'}$
- Update the parameters $(x_c, y_c, s, \theta, \mathbf{b})$ in order to fit to new found points $\mathbf{X}$
- Apply constrains to parameter $\mathbf{b}$(e.g,. $-3\sqrt{\lambda_k} \leq b_k \leq 3\sqrt{\lambda_k}$), which guarantee that new shape is similar with the shapes in training set.
- Repeat until convergence.

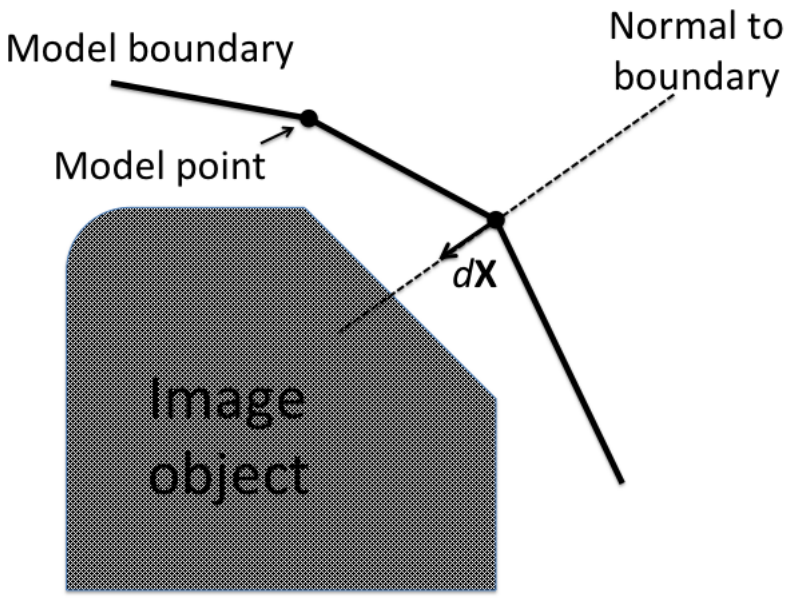


Figure 2.6: Adjustment model points to object.

Active Appearance Model (AAM) is an extension of ASM, where the texture information (e.g,. grey-level intensity) is included [11–13]. AAM uses statistical shape information and also local appearance knowledge around each landmark point. Similarly to ASM, AAM is a generative model that can generate synthetic images by combining both shape and appearance knowledge. This model can solve the image segmentation problem by minimising the differences between the synthesized image and the unseen test image.

Assume that $\mathcal{S} = \{\mathbf{x}_i\}_{i=1}^{N}$ is set of N training shapes, where each shape $\mathbf{x}_i$ includes n landmarks points. In order to model the texture information, each image in the training set is warped to match its landmark point with corresponding point in the mean shape $\bar{\mathbf{x}}$ in (2.51). Let $\mathbf{g}_{im}$ be the texture vector of a new image after warping (i.e., shape-normalised image), then each texture vector $\mathbf{g}_{im}$ is normalised as following:

$$\mathbf{g} = (\mathbf{g}_{im} - \beta\mathbf{1})/\alpha, \tag{2.57}$$

where $\mathbf{1}$ is vector of ones, α is a scaling term and β is an offset term that are chosen to achieve the best match with respect to each vector $\mathbf{g}_{im}$ with the normalised mean. Given that the normalised mean is $\bar{\mathbf{g}}$, then two terms α and β are selected as:

$$\alpha = \mathbf{g}_{im}.\bar{\mathbf{g}}, \text{ and } \beta = (\mathbf{g}_{im}.\mathbf{1})/n, \tag{2.58}$$

where n is the number of elements of vector $\mathbf{g}_{im}$.

This normalisation guarantees that $\mathbf{g}^T\mathbf{1} = 0$ and $|\mathbf{g}| = 1$. By applying PCA to both shape dataset and texture dataset, we have:

$$\mathbf{x} = \bar{\mathbf{x}} + \mathbf{P}_s\mathbf{b}_s, \tag{2.59}$$

where $\mathbf{P}_s$ and $\mathbf{b}_s$ are the matrix of modes of shape variance and the vector of shape parameters, respectively, and

$$\mathbf{g} = \bar{\mathbf{g}} + \mathbf{P}_g\mathbf{b}_g, \tag{2.60}$$

where $\mathbf{P}_g$ is the matrix of modes of appearance variance (e.g., grey-level), and $\mathbf{b}_g$ is the vector of appearance parameters.

In order to model the relationship between the shape and texture variations, the PCA method is applied on the set of concatenated vectors:

$$\mathbf{b} = \begin{pmatrix} \mathbf{W}_s\mathbf{b}_s \\ \mathbf{b}_s \end{pmatrix} = \begin{pmatrix} \mathbf{W}_s\mathbf{P}_s^T(\mathbf{x} - \bar{\mathbf{x}}) \\ \mathbf{P}_g^T(\mathbf{g} - \bar{\mathbf{g}}) \end{pmatrix} \tag{2.61}$$

where $\mathbf{W}_s$ is a diagonal matrix of weights (see (2.47)). After applying PCA, a new model of parameters is produced:

$$\mathbf{b} = \mathbf{Q}\mathbf{c}, \tag{2.62}$$

where $\mathbf{Q} = \begin{pmatrix} \mathbf{Q}_s \\ \mathbf{Q}_g \end{pmatrix}$ is the matrix of eigenvectors and $\mathbf{c}$ is an appearance vector comprising both shape and texture information. New shape and texture samples can then be generated by changing the value of parameter $\mathbf{c}$:

$$\begin{aligned} \mathbf{x} &= \bar{\mathbf{x}} + \mathbf{P}_s\mathbf{W}_s\mathbf{Q}_s\mathbf{c}, \\ \mathbf{g} &= \bar{\mathbf{g}} + \mathbf{P}_g\mathbf{Q}_g\mathbf{c}. \end{aligned} \tag{2.63}$$

Let $\mathbf{I}_i$ be the target image having the object that should be segmented, and $\mathbf{I}_m$ is synthetic image generated by AAM. The segmentation can be done by searching the appearance parameter $\mathbf{c}$ to minimise the difference between grey-level between two images:

$$\delta\mathbf{I} = \mathbf{I}_i - \mathbf{I}_m. \tag{2.64}$$

Active shape and appearance models can be considered the first machine learning models applied to medical image analysis problems. They have shown some success in few applications, but they have three main drawbacks: 1) the reliance on a generative model is problematic in the sense that it needs a large amount of training data to provide a robust estimation of the shape and appearance models (more so than discriminative models, in general); 2) the fact that it essentially uses global statistics to represent shape (and appearance) makes this model inaccurate

when dealing with test cases that present slightly different shape (and appearance) characteristics from the ones observed in the training set; and 3) it is quite sensitive to the initial conditions of the inference process.

2.4 Atlas-guided Segmentation and Registration

The use of geometric and appearance constraints (either learned or assumed in a prior model) provides a powerful cue in medical image segmentation problems. For example, level set (Section 2.2.2) and active shape models (Section 2.3.3) incorporate these types of constraints to drive the segmentation process. Chen et al. [80] use the shape prior learned by deep Boltzmann machine to do object extraction and segmentation. Atlas-guided segmentation represents another way of exploring a similar idea [34–38, 81], which includes two main steps: first, the atlas templates are constructed based on manually segmented images; and second, the visual object in a test image (i.e., target or reference image) is segmented by registering the atlas image to the test image.

An overview of atlas-guided segmentation and registration in medical image segmentation problems is given in [34, 35]. Basically, these techniques can be classified into four groups, depending on the atlas selection strategies:

- Single individual atlas: Segmentation is based on the registration with a random manually annotated image, which is chosen by user [82].
- Most similar individual image: Given a set of atlas templates, the target image is registered to each atlas and a similarity measure assesses the accuracy of the registration processes (e.g., mutual information [83, 84] or normalised mutual information [85]) in order to select the best segmentation result.
- Average shape atlas: This approach is based on the computation of an average model of all available atlases, where the segmentation is performed based on the registration between the test image and that average model [36, 86–88]. ASM [10] and AAM [11] represent typical models of atlas-guided segmentation using average shape atlas.
- Multiple atlases: Given a set of atlas templates, the target image is registered to each atlas template, which produces a set of segmentation results. Then, these segmentation results are combined using a fusion technique (e.g., "Vote Rule" decision fusion [89], partial volume interpolation [84]) in order to get the final segmentation result.

Atlas-based segmentation techniques have been applied successfully in many applications, but they also have some issues, such as their ability to represent the variability of anatomical structure and scale in medical images, and the complexity of registration algorithms.

2.5 Hybrid Models

The combination of machine learning and active contour methods has been one of the main ideas being recently explored in medical image analysis. In this section we describe some of the works that follow this idea.

Huang et al. [24] proposes how to combine MRF and active contour model in medical image segmentation. Using the formulations explained in Sections 2.2.1 and 2.3.2, image segmentation based on MRF and active contour model can be obtained by solving the two MAP estimation problems as follows:

- MRF model (see Section. 2.3.2):

$$\mathbf{y}^* = \arg\max_{\mathbf{y}} p(\mathbf{y}|\mathbf{x}), \tag{2.65}$$

  where

$$p(\mathbf{y}|\mathbf{x}) \propto p(\mathbf{x}|\mathbf{y})p(\mathbf{y}) \tag{2.66}$$

- Active contour model using explicit contour representation (see Section. 2.2.1):

$$\mathbf{c}^* = \arg\max_{\mathbf{c}} p(\mathbf{c}|I), \tag{2.67}$$

  where

$$p(\mathbf{c}|I) \propto p(I|\mathbf{c})p(\mathbf{c}). \tag{2.68}$$

The hybrid model is shown in Fig. 2.7 [24], where the authors [24] added a new hidden node to the original MRF model in Fig. 2.3. This new node represents the underlying contour $\mathbf{c}$ from active contour model, and this combination produces the extended MRF model. This MRF model estimates pixel labels in a band area around the contour that is the result of the active contour model. Then, the active contour model uses the result of hybrid model to improve its result. The hybrid model can be viewed as a joint MAP estimation problem:

$$(\mathbf{c}, \mathbf{y})_{MAP} = \arg\max_{\mathbf{c},\mathbf{y}} p(\mathbf{c}, \mathbf{y}|\mathbf{x}), \tag{2.69}$$

where

$$p(\mathbf{c}, \mathbf{y}|\mathbf{x}) \propto p(\mathbf{x}|\mathbf{y})p(\mathbf{y}|\mathbf{c})p(\mathbf{c}). \tag{2.70}$$

The likelihood term $p(\mathbf{x}|\mathbf{y})$ is defined as:

$$p(\mathbf{x}|\mathbf{y}) = \prod_i p(x_i|y_i). \tag{2.71}$$

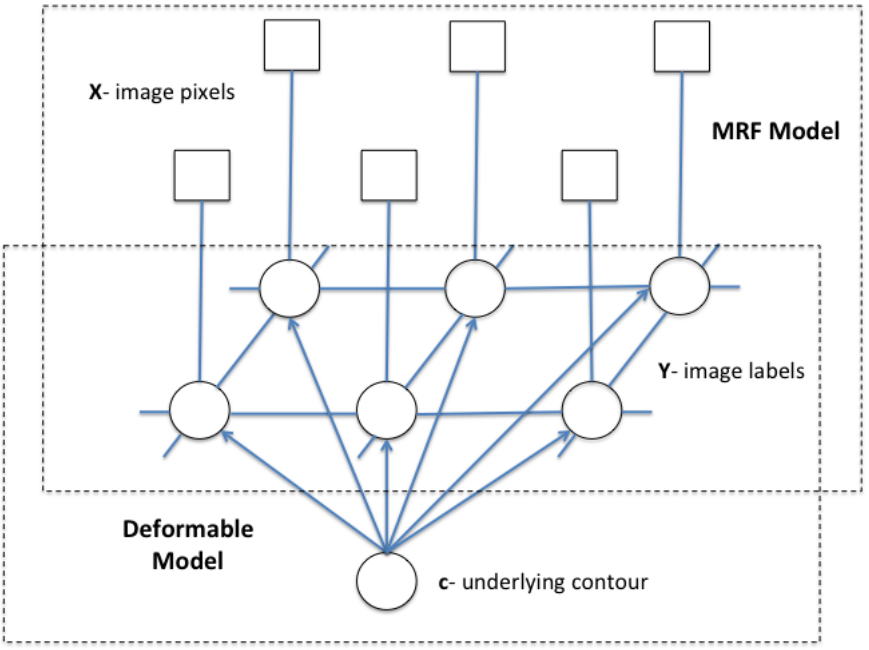


Figure 2.7: Combining MRF and explicit deformable model.

The second term $p(\mathbf{y}|\mathbf{c})$ measures the probability of pixel labels given the contour $\mathbf{c}$, where $\mathbf{c}$ is the result of deformable model as following:

$$p(\mathbf{y}|\mathbf{c}) = \prod_{(i,j)} p(y_i, y_j) \prod_i p(y_i|\mathbf{c}) \tag{2.72}$$

Note that the contour $\mathbf{c}$ is used as a prior constraint to region labels $\mathbf{y}$. The pixel label can be "foreground" or "background" depending on this pixel is inside or outside of the contour $\mathbf{c}$, where the dependence between pixel label and contour $\mathbf{c}$ can be defined via softmax function:

$$p(y_i = \text{"foreground"}|\mathbf{c}) = \frac{1}{1 + \exp\left(-dist(i, \mathbf{c})\right)} \tag{2.73}$$

$$p(y_i = \text{"background"}|\mathbf{c}) = 1 - p(y_i = \text{"foreground"}|\mathbf{c}) \tag{2.74}$$

$$dist(i, \mathbf{c}) = sign(i, \mathbf{c}) \min_{s \in [0,1]} ||loc(i) - \mathbf{c}(s)||_2, \tag{2.75}$$

where $sign(i, \mathbf{c})$ is 1 or -1 if pixel i is inside or outside contour $\mathbf{c}$, respectively, and $loc(i)$ is the spatial coordinates of pixel i. Finally, the last term $p(\mathbf{c})$ can be sampled from Gibbs distribution.

The active contour, presented in Section 2.2.1 usually lacks a term that denotes the dependence between pixels and image regions and among image regions, which can make the final result sensitive to imaging conditions. The combination of CRF and level set, proposed in [29, 30],

addresses this issue. The pixel label result of CRF, which contains information about pixels and regions dependences, are used as a constraint in the level set evolution. Tsechpenakis and Metaxas [29] also proposes hybrid model that is solved with an MAP problem.

In essence, these methods work as follows: given an image $I : \Omega \rightarrow \mathbb{R}$, the implicit contour representation $\mathcal{C}$ is the zero level set of an signed distance function $\phi : \Omega \rightarrow \mathbb{R}$ (see Section. 2.2.2).

Assuming that $\mathbf{y}$ is a labelling layer of the image I. We have to estimate the following MAP:

$$(\phi, \mathbf{y})_{MAP} = \arg\max_{(\phi,\mathbf{y})} p(\phi, \mathbf{y}|I), \tag{2.76}$$

where

$$\begin{aligned} p(\phi, \mathbf{y}|I) &\propto p(I|\mathbf{y})p(\mathbf{y}|\phi)p(\phi) \\ &= p(\phi)p(I)p(\mathbf{y}|\phi)p(\mathbf{y}|I), \end{aligned} \tag{2.77}$$

with the term $p(I)$ denoting the image prior term that is represented by a Gaussian distribution, as follows:

$$p(I(\mathbf{x}_i)) = \frac{1}{\sqrt{2\pi\sigma_0^2}} \exp\{-I(\mathbf{x}_i)^2/\sigma_0^2\}, \tag{2.78}$$

where $I(\mathbf{x}_i)$ denotes the intensity value of image I at the pixel position $\mathbf{x}_i$. The term $p(\mathbf{y}|\phi)$ in (2.77) is defined using the softmax function, such that the zero level set $\mathcal{C}$ of signed distance function ϕ is used as a prior constraint for labelling the layer $\mathbf{Y}$, as

$$p(y_i|\phi) = \frac{1}{(1 + \exp\{-\phi(\mathbf{x}_i)\})}, \tag{2.79}$$

where y_i is the label of the pixel at position $\mathbf{x}_i$. The term $p(\phi)$ in (2.77) represents the internal energy $E_{int}(\phi)$ of contour $\mathcal{C}$:

$$E_{int}(\phi) = \varepsilon_1 \mathcal{A}(\Re_{\mathcal{C}}) + \varepsilon_2 \iint_{\partial\Re_{\mathcal{C}}} ||\nabla\phi(\mathbf{x})|| d\mathbf{x}, \tag{2.80}$$

where ε_1 and ε_2 are constants, and $\partial\Re_{\mathcal{C}}$ is a narrow band of contour $\mathcal{C}$. The first term in equation 2.80 is area constraint, defined by

$$\mathcal{A}(\Re_{\mathcal{C}}) = \iint_{\Omega} H(\phi(\mathbf{x})) d\mathbf{x}, \tag{2.81}$$

with $H(\phi)$ representing the Heaviside step function, and the second term is a smoothness constraint. Finally, maximising $p(\phi)$ in (2.77) is equivalent to minimising $E_{int}(\phi)$, so $p(\phi)$ can be defined as a Gibbs prior:

$$p(\phi) = \frac{1}{Z_{int}} \exp\{-E_{int}(\phi)\}, \tag{2.82}$$

where Z_{int} is partition function. Finally, the term $p(\mathbf{y}|I)$ in (2.77) is modelled as a CRF framework. More precisely, let $\mathbf{y} = \{y_i\}$ be the set of labels corresponding to image sites $\mathbf{s} = \{s_i\}$ (e.g. pixels or image regions), so $p(\mathbf{y}|I)$ can be measured as:

$$p(\mathbf{y}|I) = \frac{1}{Z} \exp \left\{ \sum_{i \in S} \psi_a(y_i, I) + \sum_{i \in S} \sum_{j \in \mathcal{N}_i} \psi_i(y_i, y_j, I) \right\}, \tag{2.83}$$

where S is number of sites, $\mathcal{N}_i$ is the set neighbors of the site (y_i, s_i), Z is normalisation constant, and ψ_a and ψ_i are association and interaction potential functions, respectively. These terms are defined by:

$$\psi_a(y_i, I) = \log\, p(y_i|I), \tag{2.84}$$

$$\psi_i(y_i, y_j, I) = \frac{1}{z_i} \exp \left\{ \frac{\delta(y_i - y_j)}{\sigma^2} \right\}, \tag{2.85}$$

where δ is Dirac delta function, z_i is normalisation constant and σ^2 decides the relationship between neighboring labels. The term $p(y_i|I)$ can be defined as following:

$$p(y_i|I) = p_1(\mathbf{x}_i), \tag{2.86}$$

where $p_1(\mathbf{x}_i) = p(\mathbf{x}_i \in \Re_{\mathcal{C}}|I)$ represents the probability density distribution of the intensity values at the location $\mathbf{x}_i$ that is inside the contour $\mathcal{C}$. The probability density function p_1 can be estimated by using Gaussian mixture model:

$$p_1(\mathbf{x}_i) = \sum_{i=1}^{k} \omega_i . g(\mu_i, \sigma_i^2), \tag{2.87}$$

where the parameters $\{(\omega_i, \mu_i, \sigma_i), i = 1, \ldots, k\}$ can be learnt with EM algorithm.

These hybrid models address the same issues we address in this thesis, which is the combination of active contour models with machine learning models. The main issue that are present here is that the machine learning models used are standard discriminative classifiers and Markov conditional random fields, which in general require a large amount of annotated training data, which limits their application to problems that have such datasets available.

2.6 Conclusions

The medical image segmentation techniques presented above have been successful in several applications. Nevertheless, each segmentation technique still presents some issues, as discussed above. Essentially, active contour models can produce solid segmentation results with no or small annotated training sets, but these models are not able to represent well all the appearance and shape variations present in the visual object of interest. On the other hand, machine learning

methods produce much more powerful models, at the expense of needing quite large annotated training sets. Hybrid models tend to bring together the advantages of both methods, but they have to be carefully designed in order to avoid the need of large annotated training sets, which is a point not observed in the approaches explained above.

Our proposed hybrid model combines active contour models with and deep belief network, which in general needs a relatively small annotated training set (but still uses a large un-annotated training set). This is the first approach in the field to successfully combine these two methods. This combination aims to get the state of the art results in terms of segmentation accuracy. The performance of this approach is assessed in different medical imaging methods and datasets, and the experimental results show that our method produces the best results of the field for semi-automated segmentation (see results in Chapter 4 and 7), and competitive results in terms of fully automated segmentation (see results in Chapter 5 and 6).

Chapter 3

Methodology

In this chapter, we explain the main components of our methodology, which are the distance regularised level set [40], the deep belief network [41] and the combination of these two methodologies. Assume that a database of annotated images is denoted by $\mathcal{D} = \{(I, \mathbf{c})_i\}_{i=1}^{|\mathcal{D}|}$ (i.e., $\mathcal{D}$ is a database of cardiac MRI images in Chapters 4, 5, 6, and a database of chest radiographies in Chapter 7), where $I : \Omega \rightarrow \mathbb{R}$ represents an image (with $\Omega \subseteq \mathbb{R}^2$ denoting the pixel address space) and $\mathbf{c} : [0, 1] \rightarrow \Omega$ denotes the contour representation of the segmentation. Also, a segmentation map can be obtained from $\mathbf{c}$ and is represented by $\mathbf{z_c} : \Omega \rightarrow \{0, 1\}$, where 1 represents foreground (i.e., region inside the contour $\mathbf{c}$) and 0 denotes background (region outside the contour).

3.1 Distance Regularised Level Set (DRLS)

In the original level set formulation [8], the evolution of the level set function tends to develop irregularities in the signed distance function, which are fixed with periodic re-initialisations of this distance function, presenting practical and theoretical issues [40], such as numerical problems and the scheduling of re-initializations. The main issue is that the magnitude of the gradient of the distance function becomes different from one during the optimization of the level set method, and the re-initialisations of the distance function guarantee that the magnitude over the domain is restored to one. By including a term in the level set formulation that guarantees that the signed distance function remains regularised (i.e., with gradient magnitude equal to one), Li et al. [40] eliminates the need for re-initialisations and consequently the issues involved with them, which is the reason why we use this level set implementation.

The implicit contour representation that denotes the segmentation in the level set method is the zero level set of a signed distance function $\phi : \Omega \rightarrow \mathbb{R}$, as in $\mathcal{C} = \{\mathbf{x} \in \Omega | \phi(\mathbf{x}) = 0\}$, with $\phi(\mathbf{x})$

denoting the signed Euclidean distance from $\mathbf{x}$ to $\mathcal{C}$ taking negative values for points inside the contour and positive values outside this contour. In order to find the contour $\mathcal{C}$, we define an energy functional

$$\mathcal{E}(\phi) = \mu\mathcal{R}_p(\phi) + \mathcal{E}_{\text{ext}}(\phi, \phi_{\text{DBN}}), \tag{3.1}$$

where $\mathcal{R}_p(\phi) = \int_\Omega p(|\nabla\phi|)d\mathbf{x}$ is the level set regularisation term that guarantees that $|\nabla\phi| \approx 1$ with $p(s) = 0.5(s-1)^2$ [40], $\mu > 0$ is a constant, and $\mathcal{E}_{\text{ext}}$ is the external energy term with ϕ_{DBN} representing the shape produced by the DBN model explained later in Sec. 3.2. The second term in (3.1) is defined as:

$$\mathcal{E}_{\text{ext}}(\phi, \phi_{\text{DBN}}) = \lambda\mathcal{E}_{\text{lng}}(\phi) + \alpha\mathcal{E}_{\text{area}}(\phi) + \beta\mathcal{E}_{\text{shp}}(\phi, \phi_{\text{DBN}}) + \gamma\mathcal{E}_{\text{shp}}(\phi, \phi_{\text{PRIOR}}), \tag{3.2}$$

where $\lambda, \beta, \gamma > 0, \alpha \in \mathbb{R}$. The first term in (3.2) minimises the length of the contour as follows:

$$\mathcal{E}_{\text{lng}}(\phi) = \int_\Omega g\delta(\phi)|\nabla\phi|dx, \tag{3.3}$$

where $\delta(\phi)$ represents the Dirac delta function, $g = \frac{1}{1+|\nabla G_\alpha \star I|}$ denotes the edge indicator function, with $\star$ denoting the convolution operator and ∇G_α representing the gradient of the Gaussian kernel with zero mean and standard deviation σ. The energy term in (3.3) is minimised when the contour is located at image edges, and when the contour has small length. The second term in (3.2) is defined with:

$$\mathcal{E}_{\text{area}}(\phi) = \int_\Omega gH(-\phi)dx, \tag{3.4}$$

which speeds up the level set evolution process by quickly increasing or decreasing the contour area (depending on the value of α), with $H(-\phi) = 1$ when $\phi < 0$, and $H(-\phi) = 0$ otherwise (i.e., this is the Heaviside step function), and g defined above in (3.3). Note that in (3.2), we have two more energy terms: 1) $\mathcal{E}_{\text{shp}}(\phi, \phi_{\text{DBN}})$ that takes into consideration the shape estimated by the DBN ϕ_{DBN} [52, 90]; and 2) $\mathcal{E}_{\text{shp}}(\phi, \phi_{\text{PRIOR}})$ that takes into consideration the prior geometrical shape estimated from the training set [52]. Both terms are defined by:

$$\begin{aligned} \mathcal{E}_{\text{shp}}(\phi, \phi_{\text{DBN}}) &= \int_\Omega (\phi - \phi_{\text{DBN}})^2 dx, \\ \mathcal{E}_{\text{shp}}(\phi, \phi_{\text{PRIOR}}) &= \int_\Omega (\phi - \phi_{\text{PRIOR}})^2 dx, \end{aligned} \tag{3.5}$$

where ϕ_{DBN} is the signed distance function returned by the deep belief network explained below and ϕ_{PRIOR} is the signed distance function returned by the prior model estimated from the training set. The objective with these two terms is to approximate the level set function ϕ to the functions ϕ_{DBN} and ϕ_{PRIOR}.

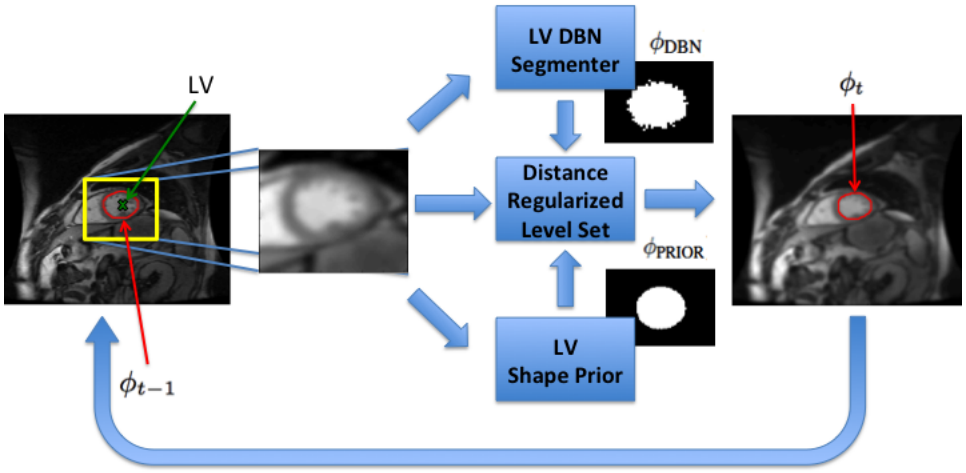


Figure 3.1: Endocardium segmentation.

The gradient flow for minimising the energy functional $\mathcal{E}(\phi)$ in (3.1) is based on finding the steady state solution of the gradient flow equation for the signed distance function $\phi(.)$ as in:

$$\begin{aligned}\frac{\partial\phi}{\partial t} &= -\frac{\partial\mathcal{E}}{\partial\phi},\\ &= \mu\text{div}(d_p(|\nabla\phi|)\nabla\phi) + \lambda\delta(\phi)\text{div}(g\frac{\nabla\phi}{|\nabla\phi|}) + \alpha g\delta(\phi) + \\ &2\beta(\phi(\mathbf{x}) - \phi_{\text{DBN}}(\mathbf{x})) + 2\gamma(\phi(\mathbf{x}) - \phi_{\text{PRIOR}}(\mathbf{x})),\end{aligned} \tag{3.6}$$

where div$(.)$ denotes the divergence operator, $\phi(\mathbf{x})$ denotes the current level set function, and $d_p(|\nabla\phi|) = p'(|\nabla\phi|)/|\nabla\phi|$.

The estimated segmentation is obtained from the minimisation of the energy functional in (3.1). In practice, the segmentation is obtained from the steady solution of the gradient flow equation [40] in (3.6). The main idea of the DRLS [40] is then to iteratively follow the steepest descent direction (3.6) until convergence, resulting in the final steady solution ϕ^*.

Fig. 3.1 shows an illustration of our methodology applied to endocardium segmentation of left ventricle (LV) in cardiac MRI.

3.2 Deep Belief Network (DBN)

One of the main recent advances that has happened in machine learning is the development of deep learning techniques, consisting of a hierarchical representation that can learn complicated functions, representing several levels of abstractions [91]. One of the breakthroughs that

enabled the exploration of deep learning architectures was the development of the contrastive divergence learning algorithm [92] that could estimate reliably the parameters of these deep hierarchies, with several levels of non-linear operators. Deep learning architecture has been applied not only to classic learning problems, producing better results than competing methodologies [41], but also to new learning problems, previously too difficult to be handled by traditional machine learning methodologies [91]. For instance, the image segmentation problem based on the structured output inference problem from raw pixel data is a problem that can be naturally and effectively handled by deep learning methodologies, and we propose a solution based on DBN in this section. Note that this solution is used to build the distance function ϕ_{DBN} for the level set energy function in (3.1).

The DBN shape detection is based on the maximisation of the following joint probability function representing a DBN model:

$$\mathbf{y}^* = \arg\max_{\mathbf{y}} \int_{\mathbf{h}_1} ... \int_{\mathbf{h}_K} p(\mathbf{v}, \mathbf{h}_1, ..., \mathbf{h}_K, \mathbf{y}; \Theta) d\mathbf{h}_1 ... d\mathbf{h}_K, \tag{3.7}$$

where $\mathbf{h}_k \in \{0,1\}^{|\mathbf{h}_k|}$ represents the $|\mathbf{h}_k|$ hidden nodes of layer $k \in \{1,..,K\}$ of the DBN, $\mathbf{v}$ is a vector representation of the input image I or a sub-image $I' : \Omega' \subset \Omega \rightarrow \mathbb{R}$, $\mathbf{y} : \Omega/\Omega' \rightarrow \{0,1\}$ represents the maps with $1s$ for objects and $0s$ for background, and Θ denotes the DBN parameters (weights and biases). The probability term in (3.7) is computed as:

$$p(\mathbf{v}, \mathbf{h}_1, ..., \mathbf{h}_K, \mathbf{y}) = p(\mathbf{h}_K, \mathbf{h}_{K-1}, \mathbf{y}) \left(\prod_{k=1}^{K-2} p(\mathbf{h}_{k+1} | \mathbf{h}_k) \right) p(\mathbf{h}_1 | \mathbf{v}), \tag{3.8}$$

where $p(\mathbf{h}_K, \mathbf{h}_{K-1}, \mathbf{y}) \propto \exp\{-\mathcal{E}_{\text{RBM}}(\mathbf{h}_K, \mathbf{h}_{K-1}, \mathbf{y})\}$ with

$$\begin{aligned} \mathcal{E}_{\text{RBM}}(\mathbf{h}_K, \mathbf{h}_{K-1}, \mathbf{y}) = &- \mathbf{b}_K^\top \mathbf{h}_K - \mathbf{a}_{K-1}^\top \mathbf{h}_{K-1} - \mathbf{a}_y^\top \mathbf{y} - \\ &(\mathbf{h}_K)^\top \mathbf{W}_K \mathbf{h}_{K-1} - (\mathbf{h}_K)^\top \mathbf{W}_{\mathbf{y}} \mathbf{y}, \end{aligned} \tag{3.9}$$

representing the energy function of a restricted Boltzmann machine (RBM) [41], where $\mathbf{b}_K, \mathbf{a}_{K-1}, \mathbf{a}_y$ denote the bias vectors and $\mathbf{W}_K, \mathbf{W}_{\mathbf{y}}$ are the weight matrices. In (3.8), we also have:

$$p(\mathbf{h}_{k+1} | \mathbf{h}_k) = \prod_j p(\mathbf{h}_{k+1}(j) = 1 | \mathbf{h}_k), \tag{3.10}$$

with $p(\mathbf{h}_{k+1}(j) = 1 | \mathbf{h}_k) = \sigma(\mathbf{b}_{k+1}(j) + \mathbf{h}_k^\top \mathbf{W}_{k+1}(:, j))$, $p(\mathbf{h}_1(j) = 1 | \mathbf{v}) = \sigma(\mathbf{b}_1(j) + \mathbf{v}^\top \mathbf{W}_1(:, j))$ [1], where $\sigma(x) = \frac{1}{1+e^{-x}}$, the operator (j) returns the j^{th} vector value, and $(:, j)$ returns the j^{th} matrix column.

This DBN is trained layer by layer by stacking RBMs up to layer $(K-1)$ [41]. The error being minimised during this unsupervised training is the reconstruction error of the visible input,

[1] That is, we assume Gaussian visible units for the DBN with mean zero and standard deviation one.

which means that this is an unsupervised learning problem. Note that as each layer k is added to the network, the result obtained from the first layer I_L (i.e., I_L is image I or sub-image I') up to layer $k-1$ is used as the "visible" input for training the RBM formed by layers $k-1$ and k. The supervised training takes place only at the highest layer K, when the manual segmentation $\mathbf{z_c}$ is provided as visible inputs to the top RBM. Each RBM is trained with contrastive divergence (CD) [92], which provides a maximum likelihood estimation of the network parameters (i.e., weights and biases) using a stochastic gradient descent algorithm (thus very efficient for large scale problems).

The inference process that produces the segmentation is achieved by first taking an input test image at the input visible layer I_L, and then computing the probability of activation up until the layer $K-1$ using the bottom-up conditional probabilities in (3.10). Then the algorithm performs Gibbs sampling in order to achieve a stable value for the hidden layers $\mathbf{h}_{K-1}$ and $\mathbf{h}_K$, and the segmentation that is denoted by $\mathbf{y}^*$. The initialisation of this sampling process is based on the probability distribution for layer $K-1$ with $P(\mathbf{h}_{K-1}|\mathbf{h}_{K-2})$ and $\mathbf{y} = \mathbf{0}$ (for all input nodes in the segmentation layer). The signed distance function $\phi_{\text{DBN}} : \Omega \rightarrow \mathbb{R}$ representing the DBN shape is then defined by:

$$\phi_{\text{DBN}}(\mathbf{x}) = \begin{cases} -d(\mathbf{x}, \Omega^{out}), & \text{if } \mathbf{x} \in \Omega^{in} \\ +d(\mathbf{x}, \Omega^{in}), & \text{if } \mathbf{x} \in \Omega^{out} \end{cases}, \tag{3.11}$$

where $\Omega^{in} = \{\mathbf{x} \in \Omega | \mathbf{z}(\mathbf{x}) = 1\}$, $\Omega^{out} = \{\mathbf{x} \in \Omega | \mathbf{z}(\mathbf{x}) = 0\}$, and $d(\mathbf{x}, \Omega) = \inf_{\mathbf{y} \in \Omega} \|\mathbf{x} - \mathbf{y}\|_2$, and where $\mathbf{z}$ is defined as:

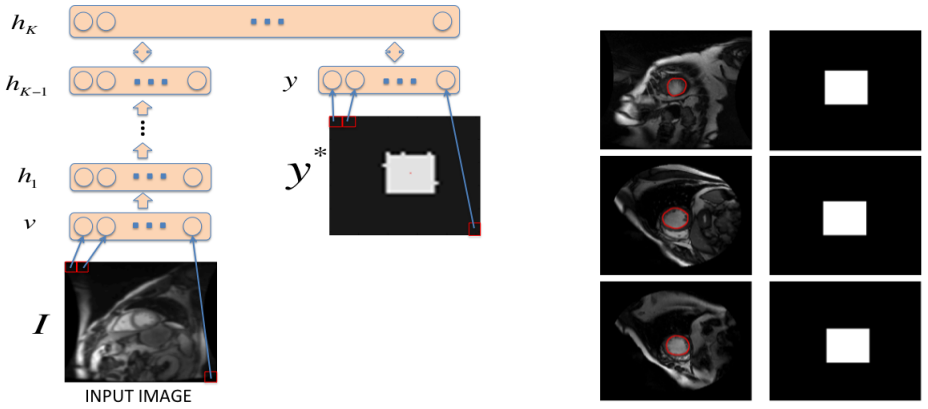
$$\mathbf{z}(\mathbf{x}) = \begin{cases} 1, & \mathbf{y}^*(\mathbf{x}) > 0.5 \\ 0, & \text{otherwise} \end{cases}. \tag{3.12}$$

Fig. 3.2, 3.3, 3.4 represent the application of DBN models for detecting the region of interest (ROI) containing the LV (Chapter 5 and 6), the endocardium and epicardium shape prior of the LV (Chapter 4, 5, and 6), and the lung shape prior (Chapter 7) respectively.

3.2.1 Shape Prior

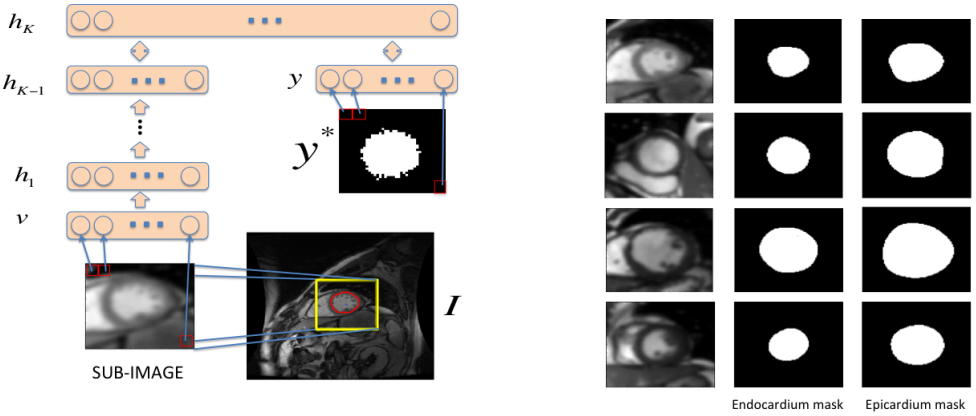
The shape prior is computed with the mean of the manual annotations $\mathbf{z_c}$, which are binary masks having "1" or "0" value depending on the pixel is inside or outside of the manual contour (e.g., epicardium masks in the panel (b) of Fig. 3.3). The shape prior is calculated as follows: $\bar{\mathbf{z}}(j) = \frac{1}{|\mathcal{D}|} \sum_{s=1}^{|\mathcal{D}|} \mathbf{z_c}(j)$, where the index $j \in \Omega$ represents a pixel address. Assuming that each element of the mean map $\bar{\mathbf{z}}$ is between 0 and 1, the shape prior is computed as

$$\mathbf{z}_{\text{PRIOR}}(j) = \begin{cases} 1, & \text{if } \bar{\mathbf{z}}_{\text{PRIOR}}(j) > 0.5 \\ 0, & \text{if } \bar{\mathbf{z}}_{\text{PRIOR}}(j) \leq 0.5 \end{cases}. \tag{3.13}$$



(a) Left ventricle ROI DBN model (b) Training samples

Figure 3.2: Left ventricle ROI DBN model (a) and training samples for the DBN (b).



(a) Endocardium and epicardium DBN model (b) Training samples

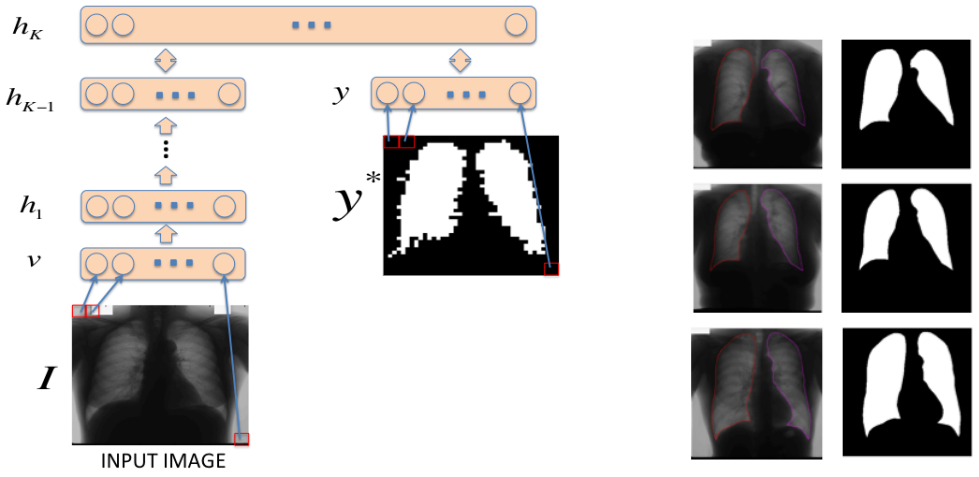
Figure 3.3: Endocardium and epicardium DBN model (a) and training samples for the DBN (b).

The signed distance function is then defined by $\phi_{\text{PRIOR}} = f_\phi(\mathbf{z}_{\text{PRIOR}}, \mathbf{m}^H, M, I)$.

$$\phi_{\text{PRIOR}}(\mathbf{x}) = \begin{cases} -d(\mathbf{x}, \Omega^{out}), & \text{if } \mathbf{x} \in \Omega^{in} \\ +d(\mathbf{x}, \Omega^{in}), & \text{if } \mathbf{x} \in \Omega^{out} \end{cases}, \tag{3.14}$$

where $\Omega^{in} = \{\mathbf{x} \in \Omega | \mathbf{z}_{\text{PRIOR}}(\mathbf{x}) = 1\}$, $\Omega^{out} = \{\mathbf{x} \in \Omega | \mathbf{z}_{\text{PRIOR}}(\mathbf{x}) = 0\}$, and $d(.)$ is defined in (3.11).

An example of computing the shape priors of endocardium and epicardium at both end diastole (ED) and end systole (ES) cardiac phases in LV segmentation is illustrated in Fig. 3.5.



(a) Lung DBN model (b) Training samples

Figure 3.4: Lung DBN model (a) and training samples for the DBN (b).

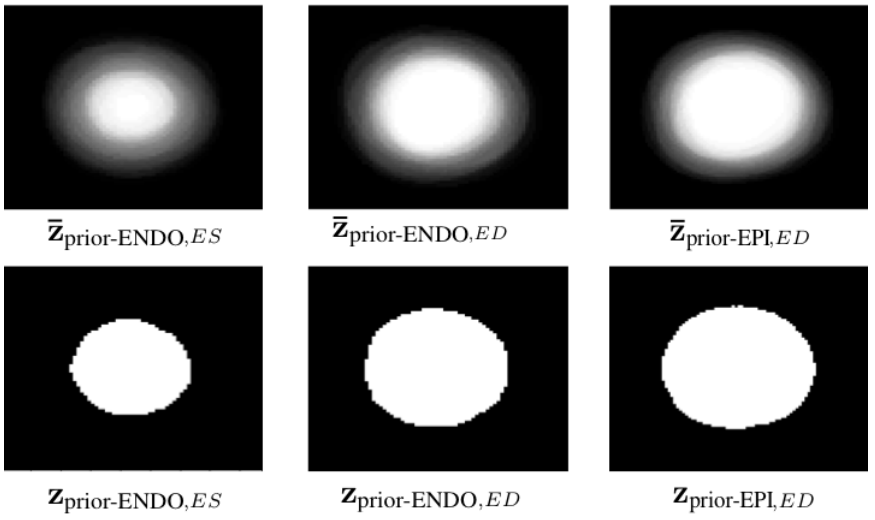


Figure 3.5: Shape priors for the endocardium and epicardium segmentation in ES and ED cardiac cycles.

3.2.2 DRLS Initialisation

The DRLS methodology needs a good initial level set function ϕ_0 because the energy function in (3.1) is non-convex, and as a result the methodology is prone to local minima. This initial guess depends on the application, but we also use a structured output DBN model combined with an image-based segmentation technique, based on Otsu's segmentation [2]. More specifically, we train a DBN model to predict the bounding box of the annotation $\mathbf{z_c}$ using the same methodology described in Section 3.2. Then, we apply Otsu's thresholding within this bounding box that

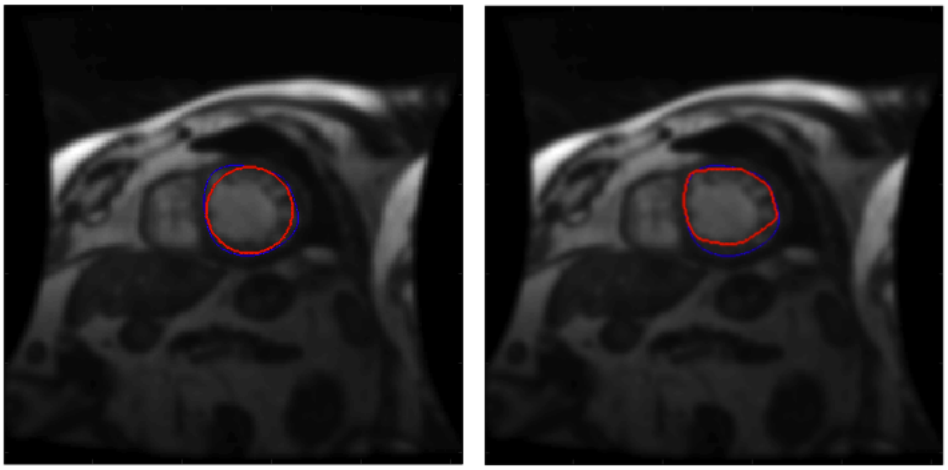


Figure 3.6: Manual and automated initial guesses of endocardium segmentation (red contour denotes our proposed initial guess and the blue contour represents the manual annotation).

produces a binary map, where

$$\phi_0(\mathbf{x}) = \begin{cases} -d(\mathbf{x}, \Omega^{out}), & \text{if } \mathbf{x} \in \Omega^{in} \\ +d(\mathbf{x}, \Omega^{in}), & \text{if } \mathbf{x} \in \Omega^{out} \end{cases}, \tag{3.15}$$

where Ω^{in} and Ω^{out} are the pixel locations, where the binary map is equal to "1" and "0", respectively.

Fig. 3.6, 3.7, 3.8 show some examples of the initial guesses used in our experiments, where the red contour denotes our proposed initial guess and the blue contour represents the manual annotation.

3.3 Segmentation Algorithm Combining DRLS and DBN

The segmentation process (detailed in Alg. 3.1), consists of a level set evolution explained in Sec. 3.1, where we assume that ϕ_{PRIOR} has been computed from the training images. The initial guess ϕ_0 is a user-defined contour or a automated detection depending on the segmentation method is semi-automated or fully-automated respectively. With ϕ_0, we compute the dynamic window L that is a sub-region of image I ((in Chapter 7, the dynamic window L is image I). With the window L, we also form the input window I_L for the DBN and run an inference process in order to find the map $\mathbf{y}^*$, as explained in Sec. 3.2. We then use $\mathbf{y}^*$ to compute the distance function ϕ_{DBN}. At this point, we run the level set iteration, minimizing the energy in (3.1), which updates the distance function ϕ.

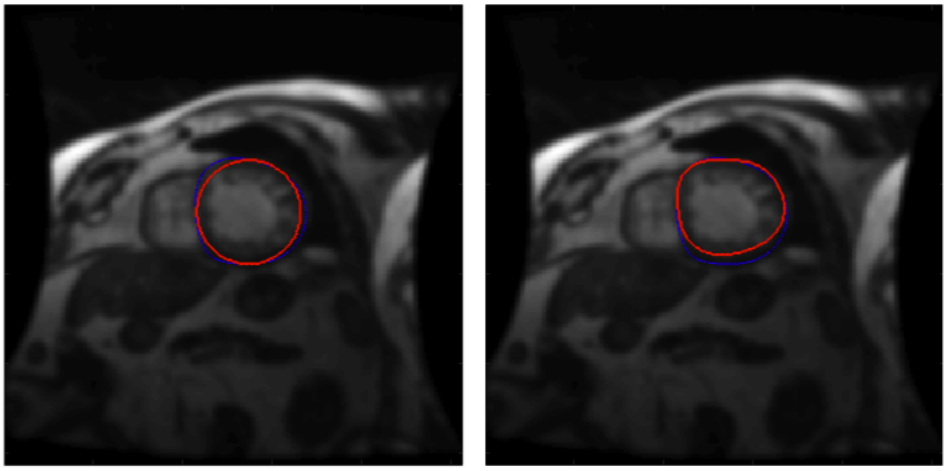


Figure 3.7: Manual and automated initial guesses of epicardium segmentation (red contour denotes our proposed initial guess and the blue contour represents the manual annotation).

Figure 3.8: Manual initial guesses of lung segmentation (red contour denotes our proposed initial guess and the blue contour represents the manual annotation).

Algorithm 3.1: Combined Level Set and DBN Segmentation

- Given test image I, ϕ_0 from I, and ϕ_{PRIOR} from $\mathcal{D}$

for t = 1:T **do**

- Compute the dynamic window L from ϕ_{t-1}
- From L extract image region I_L for the DBN, and infer $\mathbf{y}^*$
- Compute distance function ϕ_{DBN} from map $\mathbf{y}^*$
- Run DRLS using $\phi_{t-1}, \phi_{\text{PRIOR}}, \phi_{\text{DBN}}$ to produce updated distance function ϕ_t

end for

- Segmentation is the zero level set $\mathcal{C} = \{\mathbf{x} \in \Omega | \phi_T(\mathbf{x}) = 0\}$

3.4 Conclusions

In this section, we presented the general methodology proposed in this thesis. This methodology has been adapted for each problem being dealt in this thesis, as shown in the next chapters. Nevertheless, the main ideas proposed here are present in all methods below.

Chapter 4

Left Ventricle Segmentation from Cardiac MRI Combining Level Set Methods with Deep Belief Networks

---

Tuan Anh Ngo, Gustavo Carneiro

Australian Centre for Visual Technologies
University of Adelaide, Australia

The work contained in this chapter has been published in
IEEE International Conference on Image Processing (ICIP), 2013

The final publication is available at
http://dx.doi.org/10.1109/ICIP.2013.6738143

Tuan Anh Ngo, Gustavo Carneiro. Left Ventricle Segmentation from Cardiac MRI Combining Level Set Methods with Deep Belief Networks. In *IEEE International Conference on Image Processing (ICIP)*, 2013.

Statement of Authorship

Title of Paper	Left Ventricle Segmentation from Cardiac MRI Combining Level Set Methods with Deep Belief Networks
Publication Status	☒ Published ☐ Accepted for Publication ☐ Submitted for Publication ☐ Publication Style
Publication Details	Tuan Anh Ngo, Gustavo Carneiro. Left Ventricle Segmentation from Cardiac MRI Combining Level Set Methods with Deep Belief Networks. In IEEE International Conference on Image Processing (ICIP), 2013.

Principal Author

Name of Principal Author (Candidate)	Tuan Anh Ngo		
Contribution to the Paper	- Checked the database - Coded and ran the experiments - Checked the manuscript - Acted as corresponding author		
Overall percentage (%)	50%		
Signature		Date	04/08/2015

Co-Author Contributions

By signing the Statement of Authorship, each author certifies that:

i. the candidate's stated contribution to the publication is accurate (as detailed above);

ii. permission is granted for the candidate in include the publication in the thesis; and

iii. the sum of all co-author contributions is equal to 100% less the candidate's stated contribution.

Name of Co-Author	Gustavo Carneiro		
Contribution to the Paper	- Checked the database - Supervised the development of this work - Wrote the manuscript		
Signature		Date	04-08-2015

Name of Co-Author			
Contribution to the Paper			
Signature		Date	

Please cut and paste additional co-author panels here as required.

LEFT VENTRICLE SEGMENTATION FROM CARDIAC MRI COMBINING LEVEL SET METHODS WITH DEEP BELIEF NETWORKS

Tuan Anh Ngo *Gustavo Carneiro*

Australian Centre for Visual Technologies
University of Adelaide, Australia

ABSTRACT

This paper introduces a new semi-automated methodology combining a level set method with a top-down segmentation produced by a deep belief network for the problem of left ventricle segmentation from cardiac magnetic resonance images (MRI). Our approach combines the level set advantages that uses several a priori facts about the object to be segmented (e.g., smooth contour, strong edges, etc.) with the knowledge automatically learned from a manually annotated database (e.g., shape and appearance of the object to be segmented). The use of deep belief networks is justified because of its ability to learn robust models with few annotated images and its flexibility that allowed us to adapt it to a top-down segmentation problem. We demonstrate that our method produces competitive results using the database of the MICCAI grand challenge on left ventricle segmentation from cardiac MRI images, where our methodology produces results on par with the best in the field in each one of the measures used in that challenge (perpendicular distance, Dice metric, and percentage of good detections). Therefore, we conclude that our proposed methodology is one of the most competitive approaches in the field.

1. INTRODUCTION

The leading cause of death in the world is cardiovascular disease [17], and one of the best methods to improve the survival rate is based on the early diagnosis using imaging technologies. Over the last few years, there has been significant developments of imaging technologies that have enabled physicians to analyze better some parameters to assess the health of the heart (e.g., ejection fraction, wall motion, etc.). One of the current dominant imaging technologies is the cardiac magnetic resonance imaging (MRI), using the short axis view, but the segmentation of the left ventricle (LV) is a crucial first step in this analysis. Manual LV segmentation is still the standard clinical practice, but it suffers from operator bias, poor reproducibility and relatively large inter- and intra-observer variability. One possible solution to these issues is the development of a (semi-)automated LV segmentation. However, there are a few issues that must be solved before it can be accepted in a clinical setting, such as [8]: 1) the precise segmentation of the LV when the outflow tract is present, reducing the strength of edge information; and 2) the variability of the LV shape across slices, phases and patients.

Automated and semi-automated LV segmentation from cardiac MRI images has been an intensive area of research, and it is possible to classify current techniques into three categories: 1) active contour models, 2) machine learning models, and 3) combined active contour and machine learning models. Active contours with explicit contour representation [13] segments an object by minimizing an energy function with internal constraints denoting contour smoothness, and external constraints usually represented by strong edges. The use of implicit contour representation with active contours, known as the level set method [20], allowed the implementation of a similar optimization function directly on the fixed Cartesian grid without having to parameterize the curve representing the segmentation, which also allowed the delineation of objects that change topology. The main issue with active contour models is the fact that the energy function must contain all terms that are needed to segment an object, requiring substantial hand-engineering of functions (and their parameters). It is important to mention that the task of writing such energy function addressing all possible shape and appearance variations of the LV from cardiac MRI is a complicated, if not impossible task. The introduction of machine learning models has addressed exactly this issue with the use of an annotated training set to automatically learn the parameters of statistical appearance and shape models [4]. However, the automatic learning of the model parameters either requires a large training set or models with relatively low capacity, so it is clear that it would be advantageous to include some of the segmentation priors used by active contour models in order to alleviate the model learning issues. This is the idea behind the combination of active contour and machine learning models, such as the combination of Markov random field with active contour models [9], or conditional random field with active contour models [2,23]. Nevertheless, the proper training of the random fields is usually considered intractable, but there has been some progress regarding the implementation of an efficient training of random field models [22]. Our proposal focus on solving the LV segmentation problem from cardiac MRI images using this combination of active contour and machine learning models, but the machine learning model used is based on deep belief network [7], which offers: 1) efficient training and inference approaches; 2) advantages in terms of model flexibility (which means that it can be easily adapted to different types of classification and segmentation problems); and 3) relatively small annotated training sets for a robust estimation of model parameters.

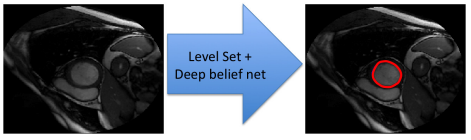


Fig. 1. Proposed methodology for cardiac MRI segmentation.

In this paper we propose a novel semi-automated LV segmentation from MRI images. Our method uses a level set method, which has a constraint provided by an LV segmentation estimated by a deep belief network (DBN), as depicted in Fig. 1. The main novelties of our paper are:

- The combination of level sets and DBN for image segmentation,
- The DBN segmentation model that produces a segmenta-

tion directly from raw pixels.

We test our approach on the MICCAI grand challenge on left ventricle segmentation from cardiac MRI images using the 15 training, 15 validation and 15 testing datasets [21]. The results show that our approach produces results on par with the best in the field in each one of the measures used in that challenge (perpendicular distance, Dice metric, and percentage of good detections). Since other approaches never achieve the top results in all three measures, we conclude that our proposed method is one of the most competitive approaches in the field.

2. PROBLEM FORMULATION

In this section, we first explain the level set method used in our methodology and then we explain how the DBN segmentation model is formulated. Assume that a database of annotated cardiac MRI images is denoted by $\mathcal{D} = \{(I, \mathbf{c})_i\}_{i=1}^{|\mathcal{D}|}$, where $I : \Omega \rightarrow \mathbb{R}$ represents an image (with $\Omega \subseteq \mathbb{R}^2$ denoting the pixel address space) and $\mathbf{c} : [0, 1] \rightarrow \Omega$ denotes the explicit contour representation of the segmentation.

2.1. Distance Regularized Level Set (DRLS)

In the original level set formulation [20], the evolution of the level set function tends to develop irregularities in the signed distance function, which are fixed with periodic re-initializations of this distance function, presenting practical and theoretical issues [14], such as numerical problems and the scheduling of re-initializations. By including a term in the level set formulation that guarantees that the signed distance function remains regularized, Li et al. [14] eliminates the need for re-initializations and consequently the issues involved with them. Because of this advantage, we use this level set implementation, and the implicit contour representation is the zero level set of a signed distance function $\phi : \Omega \rightarrow \mathbb{R}$, as in $\mathcal{C} = \{\mathbf{x} \in \Omega | \phi(\mathbf{x}) = 0\}$, with $\phi(\mathbf{x})$ denoting the signed Euclidean distance from $\mathbf{x}$ to $\mathcal{C}$ taking negative values for points inside the contour and positive values outside this contour.

In order to find the contour $\mathcal{C}$, we define an energy functional

$$\mathcal{E}(\phi, \phi_{\text{DBN}}, L) = \mu \mathcal{R}_p(\phi) + \mathcal{E}_{\text{ext}}(\phi, \phi_{\text{DBN}}, L), \tag{1}$$

where $\mathcal{R}_p(\phi) = \int_\Omega p(|\nabla\phi|) d\mathbf{x}$ is the level set regularization term, $\mu > 0$ is a constant, and $\mathcal{E}_{\text{ext}}$ is the external energy term with ϕ_{DBN} representing the shape produced by the DBN model explained later in Sec. 2.2. By solving the curve evolution equation $\frac{\partial \phi}{\partial t} = -\frac{\partial \mathcal{E}}{\partial \phi}$ using a temporal variable $t \in [0, \infty)$ (note that we assume that ϕ is parameterized not only by $\mathbf{x}$, but also by t), we can find the zero level set using an initial guess ϕ_0. The evolution of the time-dependent function ϕ follows the steepest descent direction of the energy functional $\mathcal{E}$. The derivative of the regularization term $\frac{\partial \mathcal{R}_p}{\partial \phi} = -\text{div}(d_p(|\nabla\phi|)\nabla\phi)$, where $d_p(|\nabla\phi|) = p'(|\nabla\phi|)/|\nabla\phi|$ and div is the divergence operator. The idea of using a distance regularizer is based on the fact that $p(|\nabla\phi|)$ should have two local minima at $|\nabla\phi| = \{0, 1\}$, which maintains the signed distance property $|\nabla\phi| = 1$ in a vicinity of the zero level set, and $|\nabla\phi| = 0$ at locations far away from the zero level set [14].

The second term in (1) is defined as:

$$\mathcal{E}_{\text{ext}}(\phi, \phi_{\text{DBN}}, L) = \lambda \mathcal{E}_{\text{lng}}(\phi) + \alpha \mathcal{E}_{\text{area}}(\phi) + \gamma \mathcal{E}_{\text{shp}}(\phi, \phi_{\text{prior}}, L) + \beta \mathcal{E}_{\text{shp}}(\phi, \phi_{\text{DBN}}, L), \tag{2}$$

where $\lambda, \gamma, \beta > 0$, $\alpha \in \mathbb{R}$, $\mathcal{E}_{\text{lng}}(\phi) = \int_\Omega g\delta(\phi)|\nabla\phi| dx$ returns small value for g and $|\nabla\phi|$ at edges (with $\delta(\phi)$ denoting the Dirac delta function), $\mathcal{E}_{\text{area}}(\phi) = \int_\Omega gH(-\phi) dx$ speeds up the level set evolution process by quickly increasing or decreasing the contour area (depending on the value of α, with $H(-\phi) = 1$ when $\phi < 0$, and $H(-\phi) = 0$ otherwise), $g = \frac{1}{1+|\nabla G_\sigma * I|^2}$ is a function that is small at edges and close to one elsewhere (with ∇G_σ being the gradient of a Gaussian kernel, and $*$ being the convolution operator). Also in (2), we add two energy terms to take into consideration the prior geometrical shape [5,16] learned from the annotated training set and the result of the segmentation produced by the DBN, with $\mathcal{E}_{\text{shp}}$ defined as:

$$\mathcal{E}_{\text{shp}}(\phi, \phi_{\text{k}}, L) = \int_\Omega (\phi - \phi_{\text{k}})^2 (L + 1)^2 dx, \tag{3}$$

for $k \in \{\text{prior}, \text{DBN}\}$, where $L : \Omega \rightarrow \{-1, +1\}$ is known as the dynamic labeling function that assumes the values $+1$ or -1, indicating that the prior must be enforced or not [5], respectively. Note that the the size of the window where $L = +1$ is a rectangle of $M \times N$ pixels. In practice, this dynamic labeling defines a window around the region of interest where the object of interest is believed to be localized, and this means that initially, $L = +1$ will be around the initialization $\phi_0 < 0$, and after each iteration, $L = +1$ will be around the updated $\phi < 0$. Finally, the ϕ_{prior} is computed from the training set $\mathcal{D}$ defined above by: 1) centering the training contours $\mathbf{c}$ at the origin $(0, 0)$, and 2) defining a bounding box of size $M \times N$ pixels around the centered contours. This means that all contours will have the same center, which are represented by $\widetilde{\mathbf{c}}_i$ (for $i \in \{1, 2, ..., |\mathcal{D}|\}$). The ϕ_{prior} is then the distance function computed from $\bar{\mathbf{c}}$, which is the mean aligned contour in the $M \times N$ window, calculated as $\bar{\mathbf{c}} = \frac{1}{|\mathcal{D}|} \sum_{i=1}^{|\mathcal{D}|} \widetilde{\mathbf{c}}_i$.

2.2. Deep Belief Network (DBN)

One of the main recent advances that has happened in machine learning is the development of deep learning techniques, consisting of a hierarchical representation that can learn complicated functions, representing several levels of abstractions [1]. One of the breakthroughs that enabled the exploration of deep learning architectures was the development of the contrastive-divergence learning algorithm [6] that could estimate reliably the parameters of these deep hierarchies, with several levels of non-linear operators. Deep learning architecture has been applied not only to classic learning problems, producing better results than competing methodologies [7], but also to new learning problems, previously too difficult to be handled by traditional machine learning methodologies. For instance, image segmentation from raw pixel data is a problem that can be effectively handled by deep learning methodologies, and we propose a solution based on DBN in this section. Moreover, this solution is used to build the distance function ϕ_{DBN} for the level set energy function (2).

Specifically, we exploit the model depicted in Fig. 2 with the following joint probability:

$$\begin{aligned} &P(I_L, \mathbf{h}_1, ..., \mathbf{h}_K, \mathbf{y}) \\ &= P(\mathbf{h}_K, \mathbf{h}_{K-1}, \mathbf{y}) \left(\prod_{k=1}^{K-2} P(\mathbf{h}_{k+1} | \mathbf{h}_k) \right) P(\mathbf{h}_1 | I_L) \\ &= P(\mathbf{h}_K, \mathbf{h}_{K-1}, \mathbf{y}) \left(\prod_{k=1}^{K-2} P(\mathbf{h}_k | \mathbf{h}_{k+1}) \right) P(I_L | \mathbf{h}_1) \end{aligned} \tag{4}$$

where I_L represents the raw pixel extracted from the window defined by $L = +1$ (3), $\mathbf{y} \in \{0, 1\}^{M \times N}$ represents the segmentation map of I_L, $\mathbf{h}$ denotes the hidden variables, and

$$\begin{aligned} -\log P(\mathbf{h}_K, \mathbf{h}_{K-1}, \mathbf{y}) \propto &\mathcal{E}_{\text{RBM}}(\mathbf{h}_K, \mathbf{h}_{K-1}, \mathbf{y}) \\ = &-\mathbf{b}_K^\top \mathbf{h}_K - \mathbf{a}_{K-1}^\top \mathbf{h}_{K-1} - \mathbf{a}_y^\top \mathbf{y} - \\ &(\mathbf{h}_K)^\top \mathbf{W} \mathbf{h}_{K-1} - (\mathbf{h}_K)^\top \mathbf{W}_y \mathbf{y} \end{aligned} \tag{5}$$

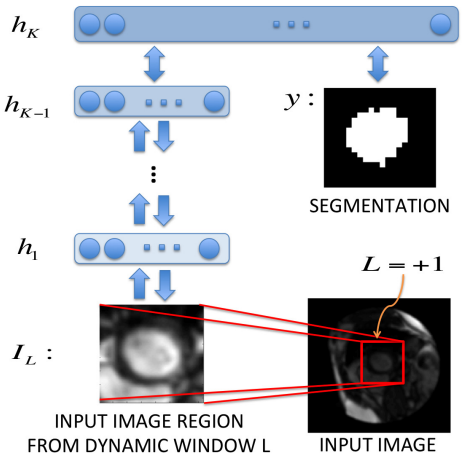


Fig. 2. Deep belief network model.

is known as a restricted Boltzmann machine (RBM) [7], where the energy function in (5) is defined by the bias vectors $\mathbf{b}_K, \mathbf{a}_{K-1}, \mathbf{a}_y$ and weight matrices $\mathbf{W}, \mathbf{W}_y$. Note from (4) and Fig. 2 that the conditional probabilities outside the top pair of layers (representing the RBM) can either be top-down or bottom-up. Also in (4), the remaining terms are related to the probability of hidden given visible variables and vice versa, which in general has the form

$$P(\mathbf{h}_{k+1}|\mathbf{h}_k) = \prod_j P(\mathbf{h}_{k+1}(j) = 1|\mathbf{h}_k),$$
$$\text{where } P(\mathbf{h}_{k+1}(j) = 1|\mathbf{h}_k) = \sigma(\mathbf{b}(j) + \mathbf{h}_k^\top \mathbf{W}(:, j))$$
$$P(\mathbf{h}_k|\mathbf{h}_{k+1}) = \prod_i P(\mathbf{h}^k(i) = 1|\mathbf{h}_{k+1}),$$
$$\text{where } P(\mathbf{h}_k(i) = 1|\mathbf{h}_{k+1}) = \sigma(\mathbf{a}(i) + \mathbf{W}(i, :)\mathbf{h}_{k+1}), \quad (6)$$

where $\sigma(x) = \frac{1}{1+e^{-x}}$, the operator (j) returns the j^{th} vector value, $(i, :)$ returns the i^{th} matrix row, and $(:, j)$ returns the j^{th} matrix column. The definition for $P(\mathbf{h}_1|I_L)$ is the same as $P(\mathbf{h}_{k+1}|\mathbf{h}_k)$ and $P(I_L|\mathbf{h}_1)$ is the same as $P(\mathbf{h}_k|\mathbf{h}_{k+1})$ in (6).

This DBN is trained layer by layer in an unsupervised way by stacking RBMs up to layer $K-1$ [7]. The error being minimized during this unsupervised training is the reconstruction error of the visible input. Note that as each layer k is added to the network, the result obtained from the first layer I_L up to layer $k-1$ is used as the "visible" input for training the RBM formed by layers $k-1$ and k. The supervised training takes place only at the highest layer K, when the segmentation $\mathbf{y}$ is provided as visible inputs to the top RBM, as depicted in Fig. 2. Note that the segmentation map is computed from the annotation $\mathbf{c}$, where pixels inside the contour are labeled "1", and outside are labeled "0". Each RBM is trained with contrastive divergence (CD) [6], which provides a maximum likelihood estimation of the network parameters (i.e., weights and biases) using a stochastic gradient descent algorithm (thus very efficient for large scale problems).

The inference process that produces the segmentation and classification is achieved by first taking an input test image at the input visible layer I_L, and then computing the probability of activation up until the layer $K-1$ using the bottom-up conditional probabilities in (6). Then the algorithm performs Gibbs sampling in order to achieve a stable value for the segmentation $\mathbf{y}$, and hidden layers $\mathbf{h}_{K-1}$ and $\mathbf{h}_K$. The initialization of this sampling process is based on the probability distribution for layer $K-1$ with $P(\mathbf{h}_{K-1}|\mathbf{h}_{K-2})$ and $\mathbf{y} = \mathbf{0}$ (for all input nodes in the segmentation layer).

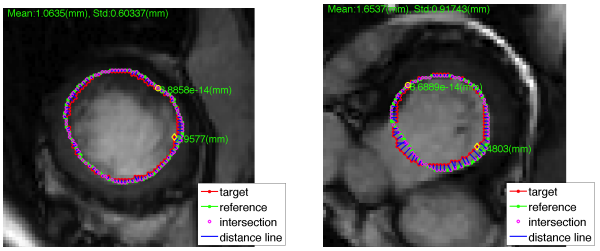


Fig. 3. Segmentation results. The left image shows a relatively simple case, but the image on the right shows a challenging case with the outflow tract present. In the legend, 'target' (red) denotes the detection and 'reference' (green) shows the manual annotation.

2.3. Segmentation Algorithm Combining DRLS and DBN

The segmentation process (detailed in Alg. 1), consists of a level set evolution explained in Sec. 2.1, where we assume that ϕ_{prior} has been computed from the training images. The first step takes the user-defined input center and scale, which forms a circle ϕ_0 that is used to initialize the level set evolution. With ϕ_0, we compute the dynamic window L in (2) by taking the region where $\phi_0 < 0$ and extend it with a fixed margin. With the window L, we also form the input window I_L for the DBN and run an inference process in order to find the map $\mathbf{y}$, as explained in Sec. 2.2. We then use $\mathbf{y}$ to compute the distance function ϕ_{DBN}. At this point, we run the level set iteration, minimizing the energy in (1), which updates the distance function ϕ.

Algorithm 1 Combined Level Set and DBN Segmentation

- Given test image I, ϕ_0 from I, and ϕ_{prior} from $\mathcal{D}$

for t = 1:T **do**

- Compute the dynamic window L from ϕ_{t-1}
- From L extract image region I_L for the DBN, and infer $\mathbf{y}$
- Compute distance function ϕ_{DBN} from map $\mathbf{y}$
- Run DRLS using $\phi_{t-1}, \phi_{\text{prior}}, \phi_{\text{DBN}}, L$ to produce updated distance function ϕ_t

end for

- Segmentation is the zero level set $\mathcal{C} = \{\mathbf{x} \in \Omega | \phi_T(\mathbf{x}) = 0\}$

3. EXPERIMENTAL RESULTS

In order to assess the performance of our algorithm, we use the MICCAI 2009 challenge database [21], consisting of three data sets (online, validation and training) obtained from the Sunnybrook Health Sciences Centre, Toronto, Canada. Each data set contains 15 cases (4 ischemic heart failures, 4 non-ischemic heart failures, 4 LV hypertrophies and 3 normal cases), thus forming 45 cardiac short axis cine-MR (SAX-MR) datasets with expert contours for the endocardial and epicardial contours in all slices at end diastole (ED) and end systole (ES) phases[1], but note that in this paper, we focus on the endocardial segmentation problem. All the images were obtained during 10-15 second breath-holds with a temporal resolution of 20 cardiac phases over the heart cycle, and scanned from the ED phase. Six to 12 SAX images were obtained from the atrioventricular ring to the apex. Three measures

[1] Only endocardial contours are available for ES.

Table 1. Quantitative experiments comparing the performance of several competing approaches on the MICCAI 2009 challenge database [21]. Each cell is formatted as "mean (standard deviation) [min value - max value]". The best result for each measure on each dataset is highlighted, and '?' means that the result is not available.

Method	Endocardial AVP	Endocardial ADM	"Good" Percentage
Training set (15 sequences)			
DBN+LS	$\mathbf{1.96(0.35)[1.43-2.55]}$	$\mathbf{0.90(0.03)[0.84-0.94]}$	$\mathbf{98.45(3.11)[91.66-100]}$
LS ONLY	$2.58(0.27)[2.28-3.08]$	$0.86(0.03)[0.79-0.91]$	$\mathbf{98.61(3.57)[88.88-100]}$
Jolly [12]	$2.09(0.53)[1.35-3.23]$	$0.88(0.06)[0.75-0.95]$	$96.93(7.59)[72-100]$
Validation set (15 sequences)			
DBN+LS	$2.22(0.46)[1.69-3.30]$	$\mathbf{0.89(0.03)[0.83-0.93]}$	$\mathbf{96.58(9.58)[63.15-100]}$
LS ONLY	$2.91(0.35)[2.41-3.73]$	$0.84(0.04)[0.77-0.90]$	$\mathbf{97.01(6.97)[73.68-100]}$
Jolly [12]	$2.26(0.59)[1.35-3.68]$	$0.88(0.04)[0.75-0.95]$	$95.62(8.83)[62-100]$
Wijnhout [24]	$2.29(0.57)[1.67-3.93]$	$\mathbf{0.89(0.03)[0.82-0.94]}$	$86.47(11)[68.4-100]$
Lu [15]	$\mathbf{2.07(0.61)[1.32-3.77]}$	$\mathbf{0.89(0.03)[0.84-0.94]}$	$72.45(18.86)[?-?]$
Huang [10]	$2.10(0.44)[?-?]$	$\mathbf{0.89(0.04)}[?-?]$	?
Marak [18]	?	$0.86(0.04)[?-?]$	?
O'Brien [19]	?	$0.81(?)[?-?]$	?
Online set (15 sequences)			
DBN+LS	$\mathbf{2.04(0.35)[1.53-2.67]}$	$\mathbf{0.90(0.04)[0.83-0.95]}$	$\mathbf{98.71(3.66)[86.66-100]}$
LS ONLY	$2.66(0.38)[2.24-3.49]$	$0.85(0.04)[0.80-0.92]$	$\mathbf{99.33(2.58)[90-100]}$
Full set (45 sequences)			
DBN+LS	$2.08(0.40)[1.43, 3.30]$	$\mathbf{0.90(0.03)[0.83, 0.95]}$	$\mathbf{97.91(6.18)[63.15, 100]}$
LS ONLY	$2.72(0.36)[2.24, 3.73]$	$0.85(0.04)[0.77, 0.92]$	$\mathbf{98.31(4.78)[73.68, 100]}$
Constantinides (full) [3]	$2.44(0.56)[1.31-4.20]$	$0.86(0.05)[0.72-0.94]$	$80(16.00)[29-100]$
Constantinides (semi) [3]	$\mathbf{1.94(0.42)[1.47-3.03]}$	$0.89(0.04)[0.80-0.96]$	$91.00(8.00)[61-100]$
Hu [8]	$2.24(0.40)[?-?]$	$0.89(0.03)[?-?]$	$91.06(9.42)[?-?]$
Huang [11]	$2.03(0.34)[?-?]$	$\mathbf{0.90(0.04)}[?-?]$	?

are computed for each data set in order to assess the performance of the proposed methodology: percentage of "good" contours, average perpendicular distance (AVP) and the average Dice metric (ADM). A contour is considered good if its AVP is less than 5mm, where each measure was computed for the annotated slices and a mean value for all the slices is given. However, AVP and ADM are computed only for good contours.

For the combined model proposed in this paper, the DBN parameters/structure and level set weights are learned using the training set, and validated with the online set. The validation set is used exclusively for testing. Note that this setup is implemented to enable a comprehensive comparison with other approaches that used the validation set for testing. With this setup, the DBN configuration achieved is: 2 hidden layers with 100 nodes in the first layer and 1000 in the second, the segmentation layer has size 20×20 in (2). The level set weights learned are: $\mu = 0.12, \lambda = 4, \alpha = -2, \gamma = 0.0005$, and $\beta = 0.001$, and the window size of $L = +1$ in (3) is $M \times N = 73 \times 73$.

In Table 1, we show quantitative results (mean, standard deviation and range) for the proposed approach combining level sets and DBN (labeled as "DBN+LS") and for the approach that uses the original DRLS formulation [14] (i.e., without the term $\mathcal{E}_{\text{shp}}$ in Eq. 2), which is labeled as "LS ONLY". The goal of comparing "DBN+LS" and "LS ONLY" is to show the influence of the DBN in the level set formulation. Moreover, we also show the results of several methodologies proposed in the literature for comparison purposes. In general, most of the approaches can be considered to be some variation of the active contour model [3,10–12,15,18], and few can be classified as machine learning methods [19,24], and one can be classified as a combination of both methods [8]. In Fig. 3, we show a couple of segmentation results produced by our approach.

From these results, we can conclude that the influence of DBN in the level set evolution is important, producing significantly more accurate results in terms of AVP and ADM, and decreasing slightly (but not significantly so) the "Good" percentage results. In general, our method is comparable or superior to all other competing methods in almost all measures, except for the AVP in the valiadation and full sets. These results place our approach among the most competitive in the field. In terms of running time per patient, the approaches vary from one minute [3,12,24] to anything in between two and three minutes [8,15]. Our approach currently takes 167.25 ± 32.71 seconds per patient (i.e., between two and three minutes), which is similar to the state of the art.

4. CONCLUSIONS AND FUTURE WORK

In this paper we present a technique for the LV segmentation in cardiac MRI images that combines level sets with deep belief networks. This is the first time such combination is proposed. Moreover, the DBN segmentation model proposed is also new, with promising results. We apply our methodology on the MICCAI 2009 challenge database [21], and the results show that the proposed methodology is one of the most accurate among the approaches that have used such database. We plan to extend our approach in several ways. First, we are currently working to make it fully automatic, with a method to detect the LV blood pool. Second, we are also extending the methodology for the detection of the epicardial contour. Finally, we also plan to work on the reduction of the running time of our approach to be closer to one minute per patient.

5. REFERENCES

[1] Y. Bengio, "Learning deep architectures for AI," *Foundations and Trends in Machine Learning*, Vol. 2. no.1 pp. 1-127, 2009. 2

[2] D. Cobzas and M. Schmidt, "Increased discrimination in level set methods with embedded conditional random fields," *CVPR*, 2009. 1

[3] C. Constantinides et al., "Fully automated segmentation of the left ventricle applied to cine MR images: Description and results on a database of 45 Subjects," *Engineering in Medicine and Biology Society (EMBC)*, 2012. 4

[4] T. Cootes, C. Taylor, D. Cooper, J. Graham et al. "Active shape models-their training and application," *Computer vision and image understanding*, 61(1):3859, 1995 1

[5] D. Cremers, "Dynamical statistical shape priors for level set-based tracking," *IEEE TPAMI* 28(8):12621273, 2006. 2

[6] M. Carreira-Perpina and G. Hinton, "On contrastive divergence learning," *Artificial Intelligence and Statistics*, vol. 2005, p. 17, 2005. 2, 3

[7] G. Hinton and R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *Science*, Vol. 313. no. 5786, pp. 504 - 507, 2006. 1, 2, 3

[8] H. Hu, H. Liu, Z. Gao and L. Huang, "Hybrid segmentation of left ventricle in cardiac MRI using gaussian-mixture model and region restricted dynamic programming," *Magnetic Resonance Imaging*, 2012. 1, 4

[9] R. Huang, V. Pavlovic, and D. Metaxas, "A graphical model framework for coupling mrfs and deformable models." *CVPR*, 2004. 1

[10] S. Huang et al., "Segmentation of the Left Ventricle from Cine MR Images Using a Comprehensive Approach," *The MIDAS Journal - Cardiac MR Left Ventricle Segmentation Challenge*, 2009. 4

[11] S. Huang et al., "An image-based comprehensive approach for automatic segmentation of left ventricle from cardiac short axis cine MR images," *Journal of Digital Imaging* Vol. 24, No.4, pp. 598-608, (2011). 4

[12] M. Jolly, "Fully automatic left ventricle segmentation in cardiac cine MR images using registration and minimum surfaces," *The MIDAS Journal - Cardiac MR Left Ventricle Segmentation Challenge*, 2009. 4

[13] M. Kass, A. Witkin, D. Terzopoulos, "Snakes: Active contour models," *International Journal of Computer Vision*, 1(4):321331, 1988. 1

[14] C. Li, C. Xu, C. Gui and M. Fox, "Distance regularized level set evolution and its application to image segmentation," *IEEE TIP*, Vol. 19, no. 12, pp. 3243-3254, 2010. 2, 4

[15] Y. Lu et al. "Automatic Image-Driven Segmentation of Left Ventricle in Cardiac Cine MRI, *The MIDAS Journal - Cardiac MR Left Ventricle Segmentation Challenge*, 2009. 4

[16] M. Lynch et al. "Left-ventricle myocardium segmentation using a coupled level-set with a priori knowledge, *Computerized Medical Imaging and Graphics*, 30(4):255-262, 2006. 2

[17] J.P. Mackenbach et al., "Socioeconomic inequalities in health in 22 European countries," *N Engl J Med*, 358, pp. 24682481, 2008. 1

[18] L. Marak et al., "4D Morphological segmentation and the MICCAI LV-segmentation grand challenge, *The MIDAS Journal - Cardiac MR Left Ventricle Segmentation Challenge*, 2009. 4

[19] S. O'Brien et al., "Segmenting the Left Ventricle in 3D Using a Coupled ASM and a Learned Non-Rigid Spatial Model, *The MIDAS Journal - Cardiac MR Left Ventricle Segmentation Challenge*, 2009. 4

[20] S. Osher and J. Sethian, "Fronts propagating with curvature-dependent speed: algorithms based on hamilton-jacobi formulations," *Journal of Computational Physics*, 79(1):1249, 1988. 1, 2

[21] P. Radau et al., "Evaluation framework for algorithms segmenting short axis cardiac MRI, *The MIDAS Journal - Cardiac MR Left Ventricle Segmentation Challenge*, 2009. 2, 3, 4

[22] M. Szummer, P. Kohli and D. Hoiem, "Learning CRFs using Graph Cuts," *ECCV*, 2008. 1

[23] G. Tsechpenakis and D. Metaxas, "CRF-driven implicit deformable model," *CVPR*, 2007. 1

[24] J. Wijnhout et al., "LV Challenge LKEB contribution: fully automated myocardial contour detection, *The MIDAS Journal - Cardiac MR Left Ventricle Segmentation Challenge*, 2009. 4

Chapter 5

Fully Automated Non-rigid Segmentation with Distance Regularized Level Set Evolution Initialized and Constrained by Deep-structured Inference.

Tuan Anh Ngo, Gustavo Carneiro

Australian Centre for Visual Technologies
University of Adelaide, Australia

The work contained in this chapter has been published in
IEEE International Conference on Computer Vision and Pattern Recognition (CVPR), 2014

The final publication is available at http://dx.doi.org/10.1109/CVPR.2014.399

Tuan Anh Ngo, Gustavo Carneiro. Fully Automated Non-rigid Segmentation with Distance Regularized Level Set Evolution Initialized and Constrained by Deep-structured Inference. *IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)*, 2014.

Statement of Authorship

Title of Paper	Fully Automated Non-rigid Segmentation with Distance Regularized Level Set Evolution Initialized and Constrained by Deep-structured Inference.
Publication Status	☒ Published ☐ Accepted for Publication ☐ Submitted for Publication ☐ Publication Style
Publication Details	Tuan Anh Ngo, Gustavo Carneiro. Fully Automated Non-rigid Segmentation with Distance Regularized Level Set Evolution Initialized and Constrained by Deep-structured Inference. IEEE International Conference on Computer Vision and Pattern Recognition (CVPR), 2014.

Principal Author

Name of Principal Author (Candidate)	Tuan Anh Ngo		
Contribution to the Paper	- Checked the database - Coded and ran the experiments - Checked the manuscript - Acted as corresponding author		
Overall percentage (%)	50%		
Signature		Date	04/08/2015

Co-Author Contributions

By signing the Statement of Authorship, each author certifies that:

i. the candidate's stated contribution to the publication is accurate (as detailed above);

ii. permission is granted for the candidate in include the publication in the thesis; and

iii. the sum of all co-author contributions is equal to 100% less the candidate's stated contribution.

Name of Co-Author	Gustavo Carneiro		
Contribution to the Paper	- Checked the database - Supervised the development of this work - Wrote the manuscript		
Signature		Date	04-08-2015

Name of Co-Author			
Contribution to the Paper			
Signature		Date	

Please cut and paste additional co-author panels here as required.

Fully Automated Non-rigid Segmentation with Distance Regularized Level Set Evolution Initialized and Constrained by Deep-structured Inference *

Tuan Anh Ngo Gustavo Carneiro
Australian Centre for Visual Technologies
The University of Adelaide, Australia

Abstract

We propose a new fully automated non-rigid segmentation approach based on the distance regularized level set method that is initialized and constrained by the results of a structured inference using deep belief networks. This recently proposed level-set formulation achieves reasonably accurate results in several segmentation problems, and has the advantage of eliminating periodic re-initializations during the optimization process, and as a result it avoids numerical errors. Nevertheless, when applied to challenging problems, such as the left ventricle segmentation from short axis cine magnetic ressonance (MR) images, the accuracy obtained by this distance regularized level set is lower than the state of the art. The main reasons behind this lower accuracy are the dependence on good initial guess for the level set optimization and on reliable appearance models. We address these two issues with an innovative structured inference using deep belief networks that produces reliable initial guess and appearance model. The effectiveness of our method is demonstrated on the MICCAI 2009 left ventricle segmentation challenge, where we show that our approach achieves one of the most competitive results (in terms of segmentation accuracy) in the field.

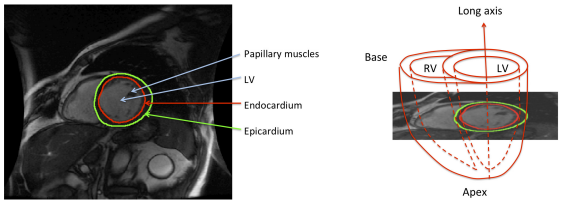
1. Introduction

Fully automated non-rigid segmentation has been one of the main research subjects in the analysis of medical images. In general, these segmentation problems involve the delineation of different types of anatomies from several imaging modalities, such as magnetic resonance imaging (MRI), ultrasound (US) or computed tomography (CT). Compared to typical segmentation applications in computer vision problems [1], the problems in medical imaging present the following challenges: more restrictive requirements in terms of the segmentation accuracy, weaker appearance models, and generally stronger shape and context models based on the consistency of human body anatomy. Given these idiosyncrasies, the most competitive methodologies developed in medical image analysis revolved around three main approaches, which are: active contour models, machine learning models, and integrated active contour and machine learning models.

The active contour model [2] is based on an optimization approach that uses an explicit representation of a contour and minimizes an energy function composed of internal and external constraints. The internal constraint represents the energy required to bend the contour, while the external constraint denotes the energy used to attract or repulse the contour towards certain appearance or shape features. The active contour model was then extended to use an implicit representation of the contour [3], which allowed the segmentation of objects that change topology. The main issue affecting active contour models lies in the design and estimation of the parameters of all the terms involved, which usually requires a substantial amount of hand tuning that rarely models all variations in terms of the shape and appearance of the visual object of interest studied in several medical image analysis problems. Machine learning models [4, 5] address exactly this issue by automatically learning these shape and appearance parameters using an annotated training set. However, it has been observed that only highly complex machine learning models are able to meet the precision requirements of medical imaging segmentation problems. Consequently, the success of machine learning models is tightly linked to large and rich training sets. Given that the task of acquiring such comprehensive training sets is complicated, particularly in medical image analysis, several researchers started looking at the alternative of combining active contour models and machine learning approaches that could be trained with smaller training sets. The most dominant approach in this direction is the integration of active contour models and Markov random fields [6, 7, 8], but the main issue of these approaches is that the training of these random fields are in general complex, requiring large amounts of training data and hand tuning.

In this paper, we propose a new fully automated segmentation approach that combines an active contour model (distance regularized level sets [9]) with a machine learning approach (deep belief network [10]). Our main objective with this approach is to obtain the most competitive segmentation results (in terms of accuracy) for the problem of

*This work was partially supported by the Australian Research Council's Discovery Projects funding scheme (project DP140102794). Tuan Anh Ngo acknowledges the support of the 322 Program - Vietnam International Education Development, Ministry of Education and Training (VIED-MOET).



(a) mid-ventricular image (b) heart model

Figure 1. LV segmentation from cine MR images [11] (a), and a 3-D model of the heart with respective MR image.

automated delineation of the left ventricle (LV) from short axis cine magnetic ressonance (MR) images [11]. The main innovations proposed are the following: 1) an appearance model learned with a deep belief network (DBN) that is used to detect the rough location and scale of the LV directly from the gray-value image using structured inference; 2) another DBN-based appearance model that is used to delineate the LV from the gray-value image using structured inference; and 3) an extension to the distance regularized level set method (DRLS) [9] that takes the estimated LV location and scale from innovation 1 (above) to initialize the optimization process and the LV delination from innovation 2 to constrain the level set evolution. The main advantage of using DBN in models 1 and 2 is that the requirements in terms of the size and richness of the annotated training set tend to be less restrictive compared to more common machine learning methods [6, 7, 8, 12, 13]. These less restrictive requirements stem from the fact that the training of a DBN involves two stages: an unsupervised learning stage that can use massive amounts of un-annotated training data, and a supervised stage that relies on relatively small training sets to converge [10]. Therefore, this addresses one of the main issues of machine learning methods listed above. We test the accuracy of our approach on the MICCAI 2009 left ventricle segmentation challenge [11], and the results show that our approach produces one of the most competitive segmentation results (in terms of segmentation accuracy) for the problem of automated LV segmentation from short axis cine MR images.

1.1. Literature Review

In this section, we summarize the main techniques proposed for the problem of left ventricle segmentation from short axis cine MR images, and for the problem of structured inference using DBNs.

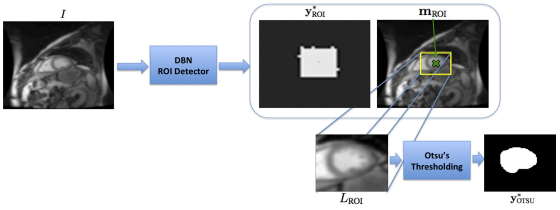
The methodology proposed in this paper can in principle be applied to most segmentation problems in medical image analysis, but we focus on the segmentation of the left ventricle (LV) endocardium from short axis cine MR images [14] (see Fig. 1). The main challenges involved in this problem are the gray level inhomogeneities of LV (because of the presence of blood flow, presence of papillary muscles and trabeculations) and the lower resolution of the apical and basal slice images when compared to the mid-ventricular images [14]. The main goal of this application is the computation of the LV volume during the end systole (ES - greatest contraction) and end diastole (ED - greatest expansion) phases of the cardiac cycle, where the ratio of these volumes is then used to compute the ejection fraction, which is used to assess the health of the heart.

According to recent review by Petitjean and Dacher [14], the approaches that address this problem can be classified in terms of the segmentation method (region and edge based, pixel classification, deformable models, active appearance and shape models), prior information (none, weak, and strong), and automated localization of the heart (time-based or object detection). They discuss the results of the MICCAI 2009 challenge [11], and reach the conclusion that the image-based methodologies [15, 16] (e.g., thresholding, or dynamic programming applied to image segmentation results) produce the best accuracy, but have the drawbacks of requiring user interaction and of being unable to assess the ventricular surface in all cardiac phases. On the other hand, other methods based on more sophisticated methodologies [17, 18, 19] do not present such issues, but show slightly less accurate results. Moreover, the remaining methodologies [15, 16, 20] present reasonably accurate results, but are too specific to the LV segmentation problem, as opposed to the approaches by O'Brien et al. [17] and Wijnhout et al. [21] that are not as accurate in general, but can be generalized to other applications. The main conclusion reached by the authors of the review [14] is that the methodology presented by Jolly [19] is the most competitive because it is fully automatic and offers the best compromise between accuracy and generability. Therefore, we regard Jolly's approach [19] as our main competitor.

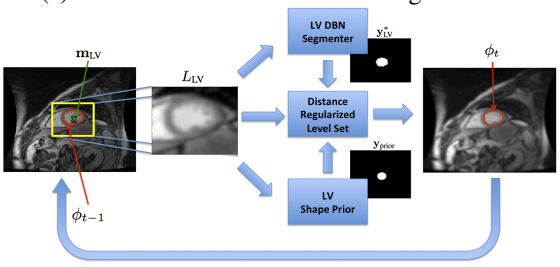
Another important point of this paper is the formulation of the image segmentation problem as a structured inference using deep belief networks (DBN) [10], where the input consists of a gray level image and the output is denoted by a binary segmentation. Most of the recent work in this field is focused on recognizing (and generating) shapes from input binary images (as opposed to gray level images) containing partially occluded or noisy shapes [22, 23]. The only methods (that we are aware of) proposing a structured inference from gray level images using DBNs are the extraction of tongue contours [24] and the segmentation of the left ventricle from MRI images [25]. In fact, we extend the work of [25] in order to make it fully automated as opposed to the semi-automated approach proposed in that paper, and also to make it robust to the ED and ES phases of the cardiac cycle. We do not consider the image parsing methods based on DBNs [26] relevant because the goals of such approaches are different from the ones in our paper, but notice that they also show structured inference using deep learning.

2. Methodology

Our methodology can be divided into two steps. The first step detects the region of interest (ROI) using a structured inference on a deep belief network (DBN), which outputs a rectangular region containing the left ventricle (LV), followed by an initial delineation of the LV using Otsu's thresholding [27] (Fig. 2-(a)). The second step takes this



(a) ROI Detection and Initial LV Segmentation



(b) LV Segmentation

Figure 2. Initial guess and level set.

initial LV segmentation and runs the distance regularized level set method [9] with the original terms plus two new terms, one based on shape prior and another based on the structured inference obtained from the DBN proposed in this paper (Fig. 2-(b)). An important contribution about the proposed DBNs is that both take as input the original gray-value image. The full segmentation algorithm is shown in Alg. 1, and we explain each step below.

Algorithm 1 Combined Level Set and DBN Segmentation

- Given image I, cardiac phase $q \in \{\text{ED,ES}\}$, and the cardiac phase dependent shape priors $\mathbf{y}_{\text{prior,ED}}$ and $\mathbf{y}_{\text{prior,ES}}$
- Estimate $\mathbf{y}^*_{ROI}$ with (1) using I
- Extract sub-image $(L_{\text{ROI}}, \mathbf{m}_{\text{ROI}}) = f_L(\mathbf{y}^*_{\text{ROI}}, I, M_{\text{ROI}})$
- Compute initial LV segmentation from L_{ROI} using Otsu's thresholding, which produces $\mathbf{y}^*_{\text{OTSU}}$
- Compute initial distance function $\phi_0 = f_\phi(\mathbf{y}^*_{\text{OTSU}}, \mathbf{m}_{\text{ROI}}, I)$

for t = 1 to T **do**

- $(L_{\text{LV}}, \mathbf{m}_{\text{LV}}) = f_L(H(-\phi_{t-1}), I, M_{LV})$, where $H(-\phi_{t-1})$ is the Heaviside step function, which effectively transforms ϕ_{t-1} into a segmentation map $\Omega \rightarrow \{0,1\}$
- Estimate $\mathbf{y}^*_{\text{LV},q}$ from (7) using L_{LV}
- $\phi^*_{\text{LV},q} = f_\phi(\mathbf{y}^*_{\text{LV},q}, \mathbf{m}_{\text{LV}}, I)$
- $\phi_{\text{prior,q}} = f_\phi(\mathbf{y}_{\text{prior},q}, \mathbf{m}_{\text{LV}}, I)$
- Run DRLS using $\phi_{t-1}, \phi_{\text{prior},q}, \phi^*_{\text{LV},q}$ to produce updated distance function ϕ_t

end for

- Final LV segmentation: $\mathcal{C} = \{\mathbf{x} \in \Omega | \phi_T(\mathbf{x}) = 0\}$

2.1. Notation

A gray-value image is represented by $I : \Omega \rightarrow \mathbb{R}$, with $\Omega \subseteq \mathbb{R}^2$ denoting the image coordinate space, the explicit contour representation of a segmentation is denoted by $\mathbf{c} : [0,1] \rightarrow \Omega$, the implicit contour representation is formed with the zero level set of an Euclidean signed distance function $\phi : \Omega \rightarrow \mathbb{R}$, represented by $\mathcal{C} = \{\mathbf{x} \in \Omega | \phi(\mathbf{x}) = 0\}$, where points inside the contour have $\phi(\mathbf{x}) < 0$ and outside, $\phi(\mathbf{x}) > 0$. Assume that a set of annotated sequences is represented by $\mathcal{D} = \{(I, \mathbf{c}, i, q)_s\}_{i \in \{1,...,N\}, s \in \{1,...,S\}, q \in \{\text{ED,ES}\}}$, where $i \in \{1, ..., N\}$ is an index to an image within a sequence, $q \in \{\text{ED,ES}\}$ is the annotation of the cardiac phase, $s \in \{1, ..., S\}$ is an index to a sequence and S is the number of sequences in $\mathcal{D}$. A segmentation map is represented by $\mathbf{y} : \Omega \rightarrow \{0,1\}$, where 1 represents foreground (i.e., the segmentation of the object of interest) and 0 denotes background. Also, assume that we have a function $(L, \mathbf{m}) = f_L(\mathbf{y}, I, M)$ that takes a segmentation map $\mathbf{y}$, an image I and parameter M, and returns $L : \Omega \rightarrow \mathbb{R}$ of size $M \times M$ pixels, where L is a sub-image of I centered at the center of mass $\mathbf{m} \in \mathbb{R}^2$ of the segmentation map where $\mathbf{y} = \mathbf{1}$. Finally, we also have a function $\phi = f_\phi(\mathbf{y}, \mathbf{m}, I)$ that returns a signed Euclidean distance function using the segmentation $\mathbf{y}$ (note that the map $\mathbf{y}$ in this case has size smaller than the size of image I) centered at position $\mathbf{m}$ on image I.

2.2. ROI DBN and Initial Segmentation

The first step of our approach is to use structured deep inference to detect the region of interest (ROI) containing the visual object and then run a simple and fast segmentation approach that will produce the initial segmentation for the distance regularized level set method, described below in Sec. 2.3. The ROI is estimated using the maximization of the following joint probability function:

$$\mathbf{y}^*_{\text{ROI}} = \arg\max_{\mathbf{y}_{\text{ROI}}} \int ... \int P(\mathbf{v}, \mathbf{h}_1, ..., \mathbf{h}_K, \mathbf{y}_{\text{ROI}}; \Theta) d\mathbf{h}_1 ... d\mathbf{h}_K, \tag{1}$$

where $\mathbf{h}_k \in \{0,1\}^{|\mathbf{h}_k|}$ represents the $|\mathbf{h}_k|$ hidden nodes of layer $k \in \{1, .., K\}$ of the deep belief network, $\mathbf{v}$ is a vector representation of the input image I, and Θ denotes the DBN parameters (weights and biases). The probability term in (1) is computed as

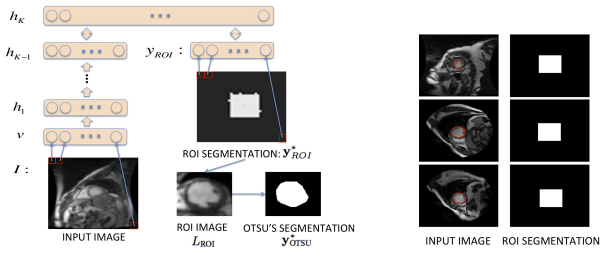
$$P(\mathbf{v}, \mathbf{h}_1, ..., \mathbf{h}_K, \mathbf{y}) = P(\mathbf{h}_K, \mathbf{h}_{K-1}, \mathbf{y}) \left(\prod_{k=1}^{K-2} P(\mathbf{h}_{k+1} | \mathbf{h}_k) \right) P(\mathbf{h}_1 | \mathbf{v}), \tag{2}$$

where $-\log P(\mathbf{h}_K, \mathbf{h}_{K-1}, \mathbf{y}) \propto \mathcal{E}_{\text{RBM}}(\mathbf{h}_K, \mathbf{h}_{K-1}, \mathbf{y})$ with

$$\mathcal{E}_{\text{RBM}}(\mathbf{h}_K, \mathbf{h}_{K-1}, \mathbf{y}) = -\mathbf{b}_K^\top \mathbf{h}_K - \mathbf{a}_{K-1}^\top \mathbf{h}_{K-1} - \mathbf{a}_y^\top \mathbf{y} - (\mathbf{h}_K)^\top \mathbf{W}_K \mathbf{h}_{K-1} - (\mathbf{h}_K)^\top \mathbf{W}_y \mathbf{y} \tag{3}$$

representing the energy function of a restricted Boltzmann machine (RBM) [10], where $\mathbf{b}_K, \mathbf{a}_{K-1}, \mathbf{a}_y$ denote the bias vectors and $\mathbf{W}_K, \mathbf{W}_y$ are the weight matrices. Also in (2), we have

$$P(\mathbf{h}_{k+1} | \mathbf{h}_k) = \prod_j P(\mathbf{h}_{k+1}(j) = 1 | \mathbf{h}_k), \tag{4}$$



(a) ROI DBN & Otsu's segmentation (b) Training samples

Figure 3. ROI DBN Model and Otsu's segmentation (a) and training set samples for the ROI DBN (b).

with $P(\mathbf{h}_{k+1}(j)=1|\mathbf{h}_k)=\sigma(\mathbf{b}_{k+1}(j)+\mathbf{h}_k^\top\mathbf{W}_{k+1}(:,j))$, $P(\mathbf{h}_1(j)=1|\mathbf{v})=\sigma(\mathbf{b}_1(j)+\frac{\mathbf{v}^\top\mathbf{W}_1(:,j)}{\sigma^2})$ [1], where $\sigma(x)=\frac{1}{1+e^{-x}}$, the operator (j) returns the j^{th} vector value, and $(:,j)$ returns the j^{th} matrix column.

The DBN represented by (1) is trained with a dataset containing the training image I and a segmentation map represented by $\mathbf{y}_{\text{ROI}}$, which is a map with 0's everywhere except around the center of mass $\mathbf{m}$ of the annotation $\mathbf{c}$, which is used as the center of a square of 1's with size M_{ROI}, as shown in Fig. 3-(b). The training process follows the same scheme proposed by Hinton et al. [10], which consists of an unsupervised bottom-up training of each pair of layers, where the weights and biases of the network are learned to build an auto-encoder for the values at the bottom layer, and a top RBM is trained with an additional input containing the segmentation map $\mathbf{y}_{\text{ROI}}$ (see training samples in Fig. 3-(b)). The inference process consists of taking the input image and performing bottom-up inferences, until reaching the top two layers, which form an RBM, and then initialize the layer $\mathbf{y}_{\text{ROI}}=\mathbf{0}$, and perform Gibbs sampling on the layers $\mathbf{h}_K$ and $\mathbf{h}_{K-1}$, $\mathbf{y}_{\text{ROI}}$ until convergence [10].

Once the ROI segmentation map $\mathbf{y}^*_{\text{ROI}}$ is estimated, then we run a simple segmentation algorithm that quickly produces an initialization for the level set method described below in Sec. 2.3. For this task, we first extract a sub-image of size M_{ROI} from I, representing the detected ROI with $(L_{\text{ROI}},\mathbf{m}_{\text{ROI}})=f_L(\mathbf{y}^*_{\text{ROI}},I,M_{\text{ROI}})$, which also returns the center of mass $\mathbf{m}_{\text{ROI}}$ of the detected ROI. Then we apply the Otsu's thresholding [27] on this sub-image L_{ROI}, and only take the connected component at the center of the ROI to build the segmentation $\mathbf{y}^*_{\text{OTSU}}$, as shown in Fig. 3-(a). This segmentation is then used to build the initial Euclidean distance function as in $\phi_0=f_\phi(\mathbf{y}^*_{\text{OTSU}},\mathbf{m}_{\text{ROI}},I)$.

2.3. Segmentation Combining DRLS and DBN

The final segmentation is obtained with the distance regularized level set (DRLS) formulation [9], where the energy functional is represented by

$$\mathcal{E}(\phi)=\mu\mathcal{R}_p(\phi)+\mathcal{E}_{\text{ext}}(\phi), \tag{5}$$

[1] That is, we assume zero-mean Gaussian visible units for the DBN.

with the distance regularization $\mathcal{R}_p(\phi)=\int_\Omega p(|\nabla\phi|)dx$ with $p(s)=0.5(s-1)^2$ (this guarantees that $|\nabla\phi|\approx 1$); and the $\mathcal{E}_{\text{ext}}(\phi)$ defined as [25]:

$$\mathcal{E}_{\text{ext}}(\phi,\phi_{\text{prior}},\phi_{\text{LV}},q)= \lambda\mathcal{L}(\phi)+\alpha\mathcal{A}(\phi)+\beta\mathcal{S}(\phi,\phi_{\text{LV},q})+\gamma\mathcal{S}(\phi,\phi_{\text{prior},q}), \tag{6}$$

where the length term $\mathcal{L}(\phi)=\int_\Omega g\delta(\phi)|\nabla\phi|dx$ (with $\delta(.)$ denoting the Dirac delta function and $g=\frac{1}{1+|\nabla G_\sigma * I|}$ representing the edge indicator function), the area $\mathcal{A}(\phi)=\int_\Omega gH(-\phi)dx$ (with $H(.)$ denoting the Heaviside step function), and $\mathcal{S}(\phi,\phi_\kappa)=\int_\Omega(\phi-\phi_\kappa)^2dx$ (with $\kappa\in\{(\text{prior},q),(\text{LV},q)\}$ and $q\in\{\text{ED,ES}\}$) represents the shape term that drives the ϕ towards the shape $\phi_{\text{LV},q}$ inferred from the LV DBN described below in Sec. 2.3.1 and also towards the shape prior $\phi_{\text{prior},q}$ learned from the training set (see Sec. 2.3.2 below). This formulation presents three extensions compared to [25], which are: 1) the cardiac phase dependent LV DBN, 2) the cardiac phase dependent shape prior, and 3) the elimination of the sub-window L in the formulation of the shape term.

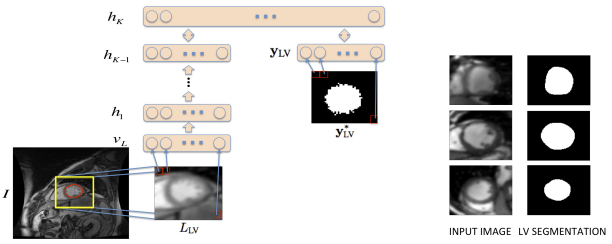
The minimization of the energy functional in (5) is achieved by finding the steady solution of the gradient flow equation [9] $\frac{\partial\phi}{\partial t}=-\frac{\partial\mathcal{E}}{\partial\phi}$, where $\partial\mathcal{E}/\partial\phi$ is the Gâteaux derivative of the functional $\mathcal{E}(\phi)$. The main idea of the DRLS [9] is then to iteratively follow the steepest descent direction of the functional $\mathcal{E}(\phi)$.

2.3.1 LV DBN

The DBN used in this stage follow the same steps as the ROI DBN, described in Equations 1-3, with a few differences, highlighted below. First, we no longer use the whole image I as the DBN input; instead, we use the square sub-image L_{LV} of size M_{LV} extraced with $(L_{\text{LV}},\mathbf{m}_{\text{LV}})=f_L(H(-\phi_{t-1}),I,M_{LV})$, where the visible layer $\mathbf{v}_L$ receives a vectorized version of this sub-image. Second, the segmentation $\mathbf{y}_{\text{LV},q}$ is a mapping with points inside the annotation contour represented by 1 and points outside denoted by 0 (note that this contour is more complicated than the rectangle represented by $\mathbf{y}_{\text{ROI}}$). Third, two DBNs will be trained: one with images belonging to the $q=\text{ES}$ phase and another with images from the $q=\text{ED}$ phase of the cardiac cycle. The segmentation from LV DBN is obtained with (see Fig. 4):

$$\mathbf{y}^*_{\text{LV},q}= \arg\max_{\mathbf{y}_{\text{LV}}}\int\!\!\ldots\!\!\int P(\mathbf{v}_L,\mathbf{h}_1,...,\mathbf{h}_K,\mathbf{y}_{\text{LV}};\Theta,q)d\mathbf{h}_1...d\mathbf{h}_K. \tag{7}$$

The two DBNs are trained in two stages (similarly to the training described in Sec. 2.2), with the first stage comprising an unsupervised bottom-up training of each pair of layers, and the second stage consisting of the training of the



(a) LV DBN (b) Training samples

Figure 4. Model for the LV DBN and training set samples.

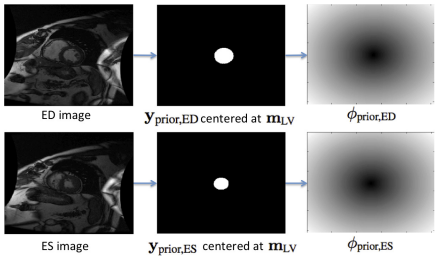


Figure 5. Examples of shape priors for ES and ED images.

top RBM with an additional input containing the segmentation map $\mathbf{y}_{\text{LV},q}$ [10]. In both stages, the objective function minimizes the reconstruction error of the visible input. The inference also follows the same process described in Sec. 2.2, which is a bottom-up inference starting from $\mathbf{v}_L$ until reaching the top two layers, followed by a Gibbs sampling inference on the layers $\mathbf{h}_K$ and $\mathbf{h}_{K-1}$, $\mathbf{y}_{\text{LV}}$ ($\mathbf{y}_{\text{LV}}$ is initialized at $\mathbf{0}$) that runs until convergence [10]. Note that this inference process runs iteratively, as shown in Alg. 1, where the sub-image L_{LV} is extracted based on the zero level set of the distance function computed from the previous iteration.

2.3.2 Shape Prior

The shape priors $\mathbf{y}_{\text{prior},q}$ (for $q \in \{\text{ED,ES}\}$) are computed based on the manual annotations present in the training set. Specifically, we take the maps $\mathbf{y}_{\text{LV},q}$ used in Sec. 2.3.1 (see Fig. 4-(b)) and compute the mean map $\bar{\mathbf{y}}_{\text{LV},q}$ using all annotations in the training set belonging to one of the two cardiac cycle phases $q \in \{\text{ED,ES}\}$. Assuming that each element of the mean map $\bar{\mathbf{y}}_{\text{LV},q}$ is between 0 and 1, the shape prior is the computed as

$$\mathbf{y}_{\text{prior},q}(j) = \begin{cases} 1, & \text{if } \bar{\mathbf{y}}_{\text{LV},q}(j) > 0.5 \\ 0, & \text{if } \bar{\mathbf{y}}_{\text{LV},q}(j) \leq 0.5 \end{cases}, \qquad (8)$$

where j indexes each element of the shape prior mapping. Notice that this shape prior map has size M_{LV}, so in order to build the Euclidean signed distance function, we need to use the center of mass $\mathbf{m}_{\text{LV}}$ (from the LV detection) in the function $\phi_{\text{prior},q} = f_\phi(\mathbf{y}_{\text{prior},q}, \mathbf{m}_{\text{LV}}, I)$.

3. Experiments

We first described the data set used and the evaluation measures proposed by Radau et al. [11]. This is followed by a detailed description of the training and inference procedures, and then we show the experimental results.

3.1. Data Set and Evaluation Measures

We assess the accuracy of our methodology using the MICCAI 2009 challenge database [11], which contains three data sets (online, testing and training sets) obtained from the Sunnybrook Health Sciences Centre, Toronto, Canada. Each of these data sets consists of 15 sequences, divided into four ischemic heart failures, four non-ischemic heart failures, four LV hypertrophies and three normal cases. Therefore, we have a total of 45 cardiac short axis (SAX) cine-MR data sets annotated with expert contours for the endocardial and epicardial contours in all slices at ED and ES cardiac phases (note that for ES images, only endocardial contours are available). As mentioned before, in this paper we focus on the segmentation of the endocardium border. Each sequence has been acquired during a 10-15 second breath-holds, with a temporal resolution of 20 cardiac phases over the heart cycle, starting from the ED cardiac phase, and containing six to 12 SAX images obtained from the atrioventricular ring to the apex (thickness=$8mm$, gap=$8mm$, FOV=$320mm \times 320mm$, matrix= 256×256). Finally, the evaluation of the segmentation accuracy is based on the following three measures: 1) percentage of "good" contours, 2) the average Dice metric (ADM) of the "good" contours, and 3) average perpendicular distance (APD) of the "good" contours. A segmentation is classified as good if the APD is less than $5mm$.

This data set was used for the MICCAI 2009 LV Segmentation Challenge [11], where the organizers first released the training and test sets, where the training set had the manual annotation, but the test set did not include the manual annotation. However, participants could submit the segmentation computed from the test set, so that they could get the evaluation results. A few days before the contest, the online set became available, and the participants could submit their segmentation results for assessment. The authors of the challenge reported all segmentation results that were available from the participants. Currently all three data sets are available with the respective manual annotations.

Given that most of the test results from the contest participants are available for the test set, we decided to use the training set to estimate the DBN parameters, and the online set for validation. The test set is then used exclusively for testing.

3.2. Experimental Setup

We use the training set for training and the online set for validation in order to estimate the ROI DBN parameters, the LV DBN parameters, the shape prior signed distance functions $\phi_{\text{prior,ED}}, \phi_{\text{prior,ES}}$, and the level set weights $\mu, \lambda, \alpha, \beta, \gamma$ in (5-6). The DBN parameters consist of the weights and biases of the network, the number of hidden

layers (we test from two to four hidden layers), and the number of nodes per hidden layer (we consider ranges from 100 to 2000 nodes per layer in intervals of 100 nodes). Note that the weights and biases are estimated with the training set only, but all other parameters are cross validated with the online set. For the ROI DBN, we reach the following configuration: 2 hidden layers with 1300 nodes in the first layer and 1500 in the second, and the input and segmentation layers with 40×40 nodes (i.e., the image is resized from 256×256 to 40×40). For the LV DBN trained with ED annotations, the following configuration is achieved: 2 hidden layers with 1000 nodes in the first layer and 1000 in the second, and the input and segmentation layers with size 40×40. The LV DBN for ES cycle has the following configuration: 2 hidden layers with 700 nodes in the first layer and 1000 in the second, and the input and segmentation layers with size 40×40. In order to estimate the shape prior, we only use the training set as described in Sec. 2.3.2. Finally, the level set weights are learned using the training set, and the result achieved are as follows: $\mu = 0.12, \lambda = 4, \alpha = -2, \gamma = 0.001$, and $\beta = 0.02$.

For the inference procedure (Alg. 1), we set the constants as follows: $T = 10$, $M_{\text{ROI}} = 100$, $M_{\text{LV}} = 100$. The segmentation results are stable if these constants are within the ranges: $T \in [5, 20]$, $M_{\text{ROI}} \in [80, 120]$, $M_{\text{LV}} \in [80, 120]$.

3.3. Results

The results in Tab. 1 show the role that each step of the proposed algorithm has in the accuracy of the resulting segmentation. In this table, "Proposed model" displays the result with all steps described in Sec. 2, while "Model without shape prior" shows the result with $\gamma = 0$ in (6), which means that the shape prior is "switched off". Similarly, "Model without DBN" represents (6) with $\beta = 0$, "Model without DBN/shape prior" denotes $\gamma = 0$ and $\beta = 0$ in (6), and "Initial guess only" means the accuracy of the initial guess alone (i.e., without running the level set method described in Sec. 2.3).

Table 2 shows a comparison between our methodology (labeled "Proposed model") and the state of the art. Most of the approaches on that table are based on active contour models [20, 16, 28, 19, 15, 29], machine learning models [17, 21], or a combination of both models [30]. Furthermore, Table 2 also shows a semi-automated version of our method (labeled "Proposed model (semi)") using the same initial guess as in [25]. Fig. 6 shows segmentation results produced by our approach.

4. Discussion and Conclusions

From the results in Table 1 we can reach three conclusions. First, the DRLS method alone [9] (i.e., without the prior and LV DBN terms) improves the result from the initial segmentation explained in Sec. 2.2. Second, the LV DBN described in Sec. 2.3.1 plays an important role given the dramatic improvements observed when it is added to the model. Third, the shape prior seems to matter only marginally since the performance without the shape prior

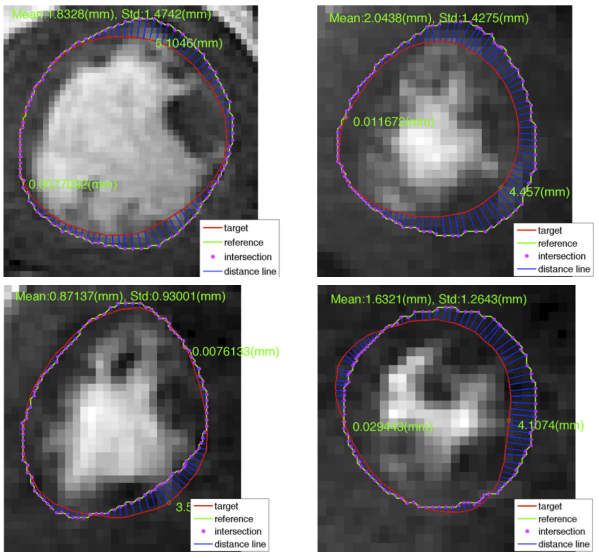


Figure 6. Segmentation results with challenging cases, such as images from apical and basal slice images and presenting papillary muscles and trabeculations. The red contour denotes the automated detection, and green shows the manual annotation.

is almost identical as with it. Nevertheless, the full model with all terms presents the most accurate results.

The comparison with the state of the art in Table 2 shows that among the fully automated methods, our approach presents the most competitive results when looking at the three measures together. The current state of the art (according to [14]) is the method proposed by Jolly [19], but our results seem to be comparable or better, particularly in terms of "Good" percentage (also notice our smaller standard deviation and larger minimum value). It is important to mention that while some approaches appear to be more accurate in terms of APD or ADM [15], they also present low values for "Good" percentage, indicating that they produce a relatively large number of segmentations with APD larger than $5mm$ (see Sec. 3.1), so it is expected that methods with larger "Good" percentage also present larger APD and smaller ADM. Another important comparison is with respect to method by Ngo and Carneiro [25], which also shows a combination of DRLS and DBN with a manual initialization (i.e., it is a semi-automated approach). Note that when compared to that approach, our proposed methodology shows slightly less competitive results, which is expected when one compares a fully automated method and a semi-automated one. However, when using a semi-automated version of our method, we achieve considerably better results than [25], which shows the improvement brought by the new bi-modal model presented in Sec. 2.3. Finally, our approach runs on (mean) average in 175 ± 33 seconds per patient (i.e., between two and three minutes) using a non-optimized Matlab program, which is comparable to other aproaches proposed that run between one minute [20, 19, 21] and three minutes [30, 15].

Table 1. Quantitative experiments on the MICCAI 2009 challenge database [11] showing the influence of each step of the proposed methodology. Each cell is formatted as "mean (standard deviation) [min value - max value]". For each measure and dataset, we highlight the most accurate measure.

Method	"Good" Percentage	Endocardial ADM	Endocardial APD
Training set (15 sequences)			
Proposed model	$97.22(3.16)[91.67 - 100]$	$\mathbf{0.88(0.05)[0.76 - 0.95]}$	$\mathbf{2.13(0.46)[1.27 - 2.73]}$
Model without shape prior	$\mathbf{97.42(4.63)[83.33 - 100]}$	$\mathbf{0.88(0.04)[0.76 - 0.95]}$	$2.14(0.43)[1.28 - 2.63]$
Model without DBN	$89.42(11.83)[61.11 - 100]$	$0.85(0.06)[0.71 - 0.93]$	$2.61(0.66)[1.74 - 3.65]$
Model without DBN/shape prior	$88.11(13.84)[50.00 - 100]$	$0.84(0.06)[0.70 - 0.93]$	$2.57(0.62)[1.72 - 3.53]$
Initial guess only	$89.61(11.57)[55.56 - 100]$	$0.85(0.06)[0.71 - 0.93]$	$2.71(0.57)[1.78 - 3.49]$
Test set (15 sequences)			
Proposed model	$\mathbf{95.91(5.28)[84.62 - 100]}$	$\mathbf{0.88(0.03)[0.82 - 0.93]}$	$\mathbf{2.34(0.46)[1.62 - 3.24]}$
Model without shape prior	$95.71(6.96)[78.95 - 100]$	$\mathbf{0.88(0.03)[0.83 - 0.93]}$	$\mathbf{2.34(0.45)[1.67 - 3.14]}$
Model without DBN	$85.89(18.00)[36.84 - 100]$	$0.84(0.04)[0.77 - 0.92]$	$2.77(0.58)[1.73 - 3.74]$
Model without DBN/shape prior	$84.49(18.31)[36.84 - 100]$	$0.84(0.04)[0.78 - 0.92]$	$2.78(0.58)[1.72 - 3.81]$
Initial guess only	$85.18(15.83)[47.37 - 100]$	$0.85(0.04)[0.79 - 0.92]$	$2.81(0.47)[2.07 - 3.58]$

According to the results shown above, we can conclude that the methodology proposed here is competitive with the state of the art in the challenging problem of LV segmentation from cine-MR images mainly in terms of accuracy. This methodology can be extended in several ways, such as the incorporation of a motion model or a 3-D geometric model that can constrain the segmentation process. Finally, another important point that we plan to address in the future is the segmentation of the epicardial contour in all slices at end diastole (ED) cardiac phase.

References

[1] R. Szeliski, *Computer vision: algorithms and applications*. Springer, 2011. 1

[2] M. Kass, A. Witkin, and D. Terzopoulos, "Snakes: Active contour models," *International journal of computer vision*, vol. 1, no. 4, pp. 321–331, 1988. 1

[3] S. Osher and J. 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. 1

[4] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham, "Active shape models-their training and application," *Computer vision and image understanding*, vol. 61, no. 1, pp. 38–59, 1995. 1

[5] B. Georgescu, X. S. Zhou, D. Comaniciu, and A. Gupta, "Databased-guided segmentation of anatomical structures with complex appearance," in *CVPR*, 2005. 1

[6] D. Cobzas and M. Schmidt, "Increased discrimination in level set methods with embedded conditional random fields," in *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*. IEEE, 2009, pp. 328–335. 1, 2

[7] R. Huang, V. Pavlovic, and D. N. Metaxas, "A graphical model framework for coupling mrfs and deformable models," in *Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on*, vol. 2. IEEE, 2004, pp. II–739. 1, 2

[8] G. Tsechpenakis and D. N. Metaxas, "Crf-driven implicit deformable model," in *Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on*. IEEE, 2007, pp. 1–8. 1, 2

[9] C. Li, C. Xu, C. Gui, and M. 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. 1, 2, 3, 4, 6

[10] G. Hinton and R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *Science*, vol. 313, no. 5786, pp. 504–507, 2006. 1, 2, 3, 4, 5

[11] P. Radau, Y. Lu, K. Connelly, G. Paul, A. Dick, and G. Wright, "Evaluation framework for algorithms segmenting short axis cardiac mri," *MIDAS J.Cardiac MR Left Ventricle Segmentation Challenge*, 2009. 2, 5, 7, 8

[12] C. Cortes and V. Vapnik, "Support vector machine," *Machine learning*, vol. 20, no. 3, pp. 273–297, 1995. 2

[13] Y. Freund and R. E. Schapire, "A desicion-theoretic generalization of on-line learning and an application to boosting," in *Computational learning theory*. Springer, 1995, pp. 23–37. 2

[14] C. Petitjean and J.-N. Dacher, "A review of segmentation methods in short axis cardiac mr images," *Medical Image Analysis*, vol. 15, no. 2, pp. 169–184, 2011. 2, 6

[15] Y. Lu, P. Radau, K. Connelly, A. Dick, and G. Wright, "Automatic image-driven segmentation of left ventricle in cardiac cine mri," *The MIDAS Journal*, vol. 49, 2009. 2, 6, 8

[16] S. Huang, J. Liu, L. Lee, S. Venkatesh, L. Teo, C. Au, and W. Nowinski, "Segmentation of the left ventricle from cine mr images using a comprehensive approach," *The MIDAS Journal*, vol. 49, 2009. 2, 6, 8

[17] S. O'Brien, O. Ghita, and P. Whelan, "Segmenting the left ventricle in 3d using a coupled asm and a learned non-rigid spatial model," *The MIDAS Journal*, vol. 49, 2009. 2, 6, 8

[18] J. Schaerer, C. Casta, J. Pousin, and P. Clarysse, "A dynamic elastic model for segmentation and tracking of the heart in mr image sequences," *Medical Image Analysis*, vol. 14, no. 6, pp. 738–749, 2010. 2

[19] M. Jolly, "Fully automatic left ventricle segmentation in cardiac cine mr images using registration and minimum surfaces," *The MIDAS Journal*, vol. 49, 2009. 2, 6, 8

[20] C. Constantinides, E. Roullot, M. Lefort, and F. Frouin, "Fully automated segmentation of the left ventricle applied to cine mr images: Description and results on a database of 45 subjects," in *Engineering in Medicine and Biology Society (EMBC), 2012 Annual International Conference of the IEEE*. IEEE, 2012, pp. 3207–3210. 2, 6, 8

[21] J. Wijnhout, D. Hendriksen, H. Assen, and R. der Geest, "Lv challenge lkeb contribution: Fully automated myocardial contour detection," *The MIDAS Journal*, vol. 43, 2009. 2, 6, 8

[22] S. A. Eslami, N. Heess, and J. Winn, "The shape boltzmann machine: a strong model of object shape," in *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*. IEEE, 2012, pp. 406–413. 2

Table 2. Quantitative experiments on the MICCAI 2009 challenge database [11] comparing the performance of our proposed approach with the state of the art. Notice that the methods are classified into fully or semi-automated. The cell formatting is the same as in Tab. 1, but note that '?' means that the result is not available in the literature.

Method	"Good" Percentage	Endocardial ADM	Endocardial APD
Test set (15 sequences)			
Fully Automated			
Proposed model	$\mathbf{95.91(5.28)[84.62-100]}$	$0.88(0.03)[0.82-0.93]$	$2.34(0.46)[1.62-3.24]$
Jolly [19]	$94.33(9.93)[62.00-100]$	$0.88(0.03)[0.84-0.94]$	$2.44(0.62)[1.36-3.68]$
Wijnhout [21]	$86.47(11.00)[68.4-100]$	$\mathbf{0.89(0.03)[0.82-0.94]}$	$2.29(0.57)[1.67-3.93]$
Lu [15]	$72.45(19.52)[42.11-100]$	$\mathbf{0.89(0.03)[0.84-0.94]}$	$\mathbf{2.07(0.61)[1.32-3.77]}$
Marak [29]	?	$0.86(0.04)[?-?]$	?
O'Brien [17]	?	$0.81(?)[?-?]$	?
Semi Automated			
Proposed model (semi)	$\mathbf{100(0)[100-100]}$	$\mathbf{0.91(0.03)[0.83-0.95]}$	$\mathbf{1.79(0.36)[1.28-2.75]}$
Ngo [25]	$96.58(9.58)[63.15-100]$	$0.89(0.03)[0.83-0.93]$	$2.22(0.46)[1.69-3.30]$
Huang [16]	?	$\mathbf{0.89(0.04)}[?-?]$	$2.10(0.44)[?-?]$

Training set (15 sequences)			
Fully Automated			
Proposed model	$\mathbf{97.22(3.16)[91.67-100]}$	$\mathbf{0.88(0.05)[0.76-0.95]}$	$2.13(0.46)[1.27-2.73]$
Jolly [19]	$96.93(7.59)[72-100]$	$\mathbf{0.88(0.06)[0.75-0.95]}$	$\mathbf{2.09(0.53)[1.35-3.23]}$
Semi Automated			
Proposed model (semi)	$\mathbf{100(0)[100-100]}$	$\mathbf{0.91(0.03)[0.85-0.95]}$	$\mathbf{1.63(0.40)[1.29-2.70]}$
Ngo [25]	$98.45(3.11)[91.66-100]$	$0.90(0.03)[0.84-0.94]$	$1.96(0.35)[1.43-2.55]$
Huang [16]	?	$\mathbf{0.90(0.04)}[?-?]$	$2.03(0.34)[?-?]$

Online set (15 sequences)			
Fully Automated			
Proposed model	$\mathbf{90.54(14.40)[46.67-100]}$	$\mathbf{0.89(0.03)[0.82-0.94]}$	$\mathbf{2.17(0.46)[1.62-3.46]}$
Semi Automated			
Proposed model (semi)	$\mathbf{100(0)[100-100]}$	$\mathbf{0.91(0.03)[0.85-0.96]}$	$\mathbf{1.78(0.49)[1.17-3.15]}$
Ngo [25]	$98.71(3.66)[86.66-100]$	$0.90(0.04)[0.83-0.95]$	$2.04(0.35)[1.53-2.67]$

Full set (45 sequences)			
Fully Automated			
Proposed model	$\mathbf{94.55(9.31)[46.67-100]}$	$0.88(0.04)[0.76-0.95]$	$2.22(0.46)[01.27-3.46]$
Constantinides [20]	$80.00(16.00)[29-100]$	$0.86(0.05)[0.72-0.94]$	$2.44(0.56)[1.31-4.20]$
Hu [30]	$91.06(9.42)[?-?]$	$\mathbf{0.89(0.03)}[?-?]$	$2.24(0.40)[?-?]$
Huang [28]	$79.20(19.00)[?-?]$	$\mathbf{0.89(0.04)}[?-?]$	$\mathbf{2.16(0.46)}[?-?]$
Semi Automated			
Proposed model (semi)	$\mathbf{100(0)[100,100]}$	$\mathbf{0.91(0.03)[0.83,0.96]}$	$\mathbf{1.73(0.31)[1.17,3.15]}$
Ngo [25]	$97.91(6.18)[63.15,100]$	$0.90(0.03)[0.83,0.95]$	$2.08(0.40)[1.43,3.30]$
Constantinides [20]	$91.00(8.00)[61-100]$	$0.89(0.04)[0.80-0.96]$	$1.94(0.42)[1.47-3.03]$

[23] F. Chen, H. Yu, R. Hu, and X. Zeng, "Deep learning shape priors for object segmentation," in *Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on*. IEEE, 2013, pp. 1870–1877. 2

[24] I. Fasel and J. Berry, "Deep belief networks for real-time extraction of tongue contours from ultrasound during speech," in *Pattern Recognition (ICPR), 2010 20th International Conference on*. IEEE, 2010, pp. 1493–1496. 2

[25] T. A. Ngo and G. Carneiro, "Left ventricle segmentation from cardiac mri combining level set methods with deep belief networks," in *Image Processing (ICIP), 2013 International Conference on*. IEEE, 2013. 2, 4, 6, 8

[26] C. Farabet, C. Couprie, L. Najman, and Y. LeCun, "Scene parsing with multiscale feature learning, purity trees, and optimal covers," *arXiv preprint arXiv:1202.2160*, 2012. 2

[27] N. Otsu, "A threshold selection method from gray-level histograms," *Automatica*, vol. 11, no. 285-296, pp. 23–27, 1975. 2, 4

[28] S. Huang, J. Liu, L. C. Lee, S. K. Venkatesh, L. L. San Teo, C. Au, and W. L. Nowinski, "An image-based comprehensive approach for automatic segmentation of left ventricle from cardiac short axis cine mr images," *Journal of digital imaging*, vol. 24, no. 4, pp. 598–608, 2011. 6, 8

[29] L. Marak, J. Cousty, L. Najman, H. Talbot, *et al.*, "4d morphological segmentation and the miccai lv-segmentation grand challenge," in *MICCAI 2009 Workshop on Cardiac MR Left Ventricle Segmentation Challenge*, no. 1, 2009, pp. 1–8. 6, 8

[30] H. Hu, H. Liu, Z. Gao, and L. Huang, "Hybrid segmentation of left ventricle in cardiac mri using gaussian-mixture model and region restricted dynamic programming," *Magnetic resonance imaging*, 2012. 6, 8

Chapter 6

Combining Deep Learning and Level Set for the Automated Segmentation of the Left Ventricle of the Heart from Cardiac Cine Magnetic Resonance

---

Tuan Anh Ngo, Gustavo Carneiro

Australian Centre for Visual Technologies
University of Adelaide, Australia

The work contained in this chapter is under review by
Medical Image Analysis (MedIA)

Tuan Anh Ngo, Gustavo Carneiro. Combining Deep Learning and Level Set for the Automated Segmentation of the Left Ventricle of the Heart from Cardiac Cine Magnetic Resonance. *Submitted to Medical Image Analysis (MedIA)*

Statement of Authorship

Title of Paper	Combining Deep Learning and Level Set for the Automated Segmentation of the Left Ventricle of the Heart from Cardiac Cine MR.
Publication Status	☐ Published ☐ Accepted for Publication ☒ Submitted for Publication ☐ Publication Style
Publication Details	Tuan Anh Ngo, Gustavo Carneiro. Combining Deep Learning and Level Set for the Automated Segmentation of the Left Ventricle of the Heart from Cardiac Cine MR. Submitted to Medical Image Analysis .

Principal Author

Name of Principal Author (Candidate)	Tuan Anh Ngo		
Contribution to the Paper	- Checked the database - Coded and ran the experiments - Checked the manuscript - Acted as corresponding author		
Overall percentage (%)	50%		
Signature		Date	04/08/2015

Co-Author Contributions

By signing the Statement of Authorship, each author certifies that:

i. the candidate's stated contribution to the publication is accurate (as detailed above);

ii. permission is granted for the candidate in include the publication in the thesis; and

iii. the sum of all co-author contributions is equal to 100% less the candidate's stated contribution.

Name of Co-Author	Gustavo Carneiro		
Contribution to the Paper	- Checked the database - Supervised the development of this work - Wrote the manuscript		
Signature		Date	04-08-2015

Name of Co-Author			
Contribution to the Paper			
Signature		Date	

Please cut and paste additional co-author panels here as required.

Combining Deep Learning and Level Set for the Automated Segmentation of the Left Ventricle of the Heart from Cardiac Cine Magnetic Resonance

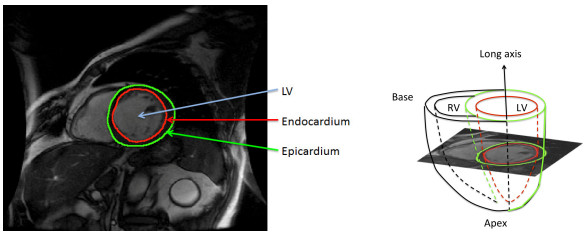
Tuan Anh Ngo and Gustavo Carneiro[1]

Australian Centre for Visual Technologies
The University of Adelaide
North Terrace, Ingkarni Wardli Building
Adelaide, SA 5005, Australia

Abstract

We introduce a new methodology that combines deep learning and level set for the automated segmentation of the left ventricle of the heart from cardiac cine magnetic resonance (MR) data. This combination is relevant for segmentation problems, where the visual object of interest presents large shape and appearance variations, but the annotated training set is small, which is the case for various medical image analysis applications, including the one considered in this paper. In particular, level set methods are based on shape and appearance terms that use small training sets, but present limitations for modelling the visual object variations. Deep learning methods can model such variations using relatively small amounts of annotated training, but they often need to be regularised to produce good generalisation. Therefore, the combination of these methods brings together the advantages of both approaches, producing a methodology that needs small training sets and produces accurate segmentation results. We test our methodology on the MICCAI 2009 left ventricle segmentation challenge database (containing 15 sequences for training, 15 for validation and 15 for testing), where our approach achieves the most accurate results in the semi-automated problem and state-of-the-art results for the fully automated challenge.

[1]Corresponding author.



(a) mid-ventricular image (b) heart model

Figure 1: LV segmentation from cardiac cine MR imaging (Radau et al. (2009)) (a), and a 3-D model of the heart with respective MR image, representing one of the volume slices.

Keywords: Deep learning, Level set method, Segmentation of the Left Ventricle of the Heart, Cardiac Cine Magnetic Resonance.

1. Introduction

Medical image analysis segmentation problems are unique in the sense that they require highly accurate results, but at the same time provide relatively small annotated training sets. A typical example is the segmentation of the endocardium and epicardium from the left ventricle (LV) of the heart using cardiac cine Magnetic Resonance (MR), as shown in Fig. 1. The LV segmentation is necessary for the assessment of the cardiovascular system function and structure and needs to be accurate for a precise diagnosis, but current public databases do not present large annotated training sets (Petitjean and Dacher (2011); Radau et al. (2009)). Therefore, one of the main research topics in this field is how to obtain the precision required with these small training sets.

The main techniques being explored for the automated segmentation of the endocardium and epicardium from cardiac cine MR are based on active contour models, machine learning models, and integrated active contour and machine learning models. Active contour models (Kass et al. (1988); Osher and Sethian (1988)) represent one of the most successful methodologies in the field, and they are based on an optimisation that minimises an energy functional that varies the shape of a contour using internal and external constraints. The energy to

bend, stretch or shrink a contour is represented by the internal constraints, while the external constraints use the observed data (e.g., image) to move the contour towards (or away from) certain appearance features (such as edges). These constraints are usually designed by hand based on shape and appearance priors that use small or no annotated training sets. Although successful, active contour models are based on low-complexity shape and appearance models that are usually unable to robustly model all variation present in the visual object of interest studied in several medical image analysis problems.

The advent of machine learning methods to medical image analysis (Cootes et al. (1995); Georgescu et al. (2005)) has addressed this issue by estimating more complex shape and appearance models using annotated training sets. However, the accuracy requirements found in medical image analysis applications usually mean that these models need to be quite complex in order to allow the learning of all appearance and shape variations found in the annotated training set, and as a consequence, this training set has to be large and rich. The issue in machine learning based models then becomes centred on the acquisition of comprehensive annotated training sets, which is a particularly complicated task in medical image analysis. Therefore, in order to reduce the model complexity and consequently, the need for large and rich training sets, a natural idea is combine the prior information of active contour models with the learned information of machine learning models. The most dominant approach in this direction is the integration of active contour models and Markov random fields (Cobzas and Schmidt (2009); Huang et al. (2004); Tsechpenakis and Metaxas (2007)), but the main issue of these approaches is that these models are in general quite complex, and as a result they still require large amounts of training data.

In this paper, we propose a new automated segmentation approach for the endocardial and epicardial borders of the left ventricle (LV) from all slices of the end diastole (ED) and end systole (ES) cardiac phases of an MR cine study. This proposed approach combines an active contour model (distance regularised level sets) (Li et al. (2010)) with a machine learning approach (deep belief network) (Hinton and Salakhutdinov (2006)). This is a sensible combination because

this problem does not usually have comprehensive training sets available, but still requires high segmentation accuracy (Radau et al. (2009)). Specifically, we explore the fact that the prior information explored by the level set method reduces the need of using highly complex machine learning models (requiring large training sets), but the limitations of this prior information indicates the need of a machine learning method that can reliably model the shape and appearance of the LV. However, this method must be able to be robustly trained with a limited number of annotated training images, which is the exactly one of the advantages behind deep belief network training (Carneiro et al. (2012); Carneiro and Nascimento (2013)). We show that this combination leads to competitive segmentation accuracy results on the MICCAI 2009 LV segmentation challenge database (Radau et al. (2009)), which does not contain a large training set and that has been tested by several different methodologies. Specifically, our experiments show that our approach produces the best result in the field when we rely on a semi-automated segmentation (i.e., with manual initialisation). Also, our fully automated approach produces a result that is on par with the current state of the art on the same database (Jolly (2009)).

1.1. Contributions

The main contributions of our approach are the following: 1) structured output for the region of interest (ROI) of the LV using a deep belief network (DBN), 2) structured output for the delineation of the endocardial and epicardial borders using another DBN, and 3) extension to the distance regularised level set method (DRLS) (Li et al. (2010)) that takes the estimated ROI from innovation (1) (above) to initialise the optimisation process and the delineation from innovation (2) to constrain the level set evolution. One advantage of using DBN models lies in the need of smaller training sets (Hinton and Salakhutdinov (2006)) compared to other machine learning methods (Cobzas and Schmidt (2009); Huang et al. (2004); Tsechpenakis and Metaxas (2007); Cortes and Vapnik (1995); Freund and Schapire (1995)). Another advantage of our method is the improved accuracy brought by the integration of the DBN and DRLS,

when compared to the accuracy of the DBN and DRLS independently. Finally, compared to our preliminary papers (Ngo and Carneiro (2013, 2014)), this work presents the following contributions: 1) detection and segmentation of the epicardial border, and 2) comparison of our epicardium segmentation results (in addition to the endocardium segmentation already presented in (Ngo and Carneiro (2013, 2014))) with the state of the art.

2. Literature Review

We focus this work on the segmentation of the endocardial and epicardial borders of the LV from short axis cine MR images (see Fig. 1), so we explore the literature for this application, but in principle our proposed methodology is general enough to be extended to other applications (this extension is out of the scope of this paper). This segmentation has several challenges, which include the lack of gray level homogeneity of LV among different cases (due to blood flow, papillary muscles and trabeculations) and the low resolution of the apical and basal images (Petitjean and Dacher (2011)). According to (Petitjean and Dacher (2011)), current LV segmentation approaches can be classified based on three characteristics: 1) segmentation method (region and edge based, pixel classification, deformable models, active appearance and shape models), 2) prior information (none, weak, and strong), and 3) automated localisation of the heart (time-based or object detection). Furthermore, their analysis (Petitjean and Dacher (2011)) of the MICCAI 2009 challenge results (Radau et al. (2009)) indicates that image-based methodologies (Lu et al. (2009); Huang et al. (2009)) (e.g., thresholding, or dynamic programming applied to image segmentation results) produce the highest accuracy, but have the drawbacks of requiring user interaction and of being unable to assess the ventricular surface in all cardiac phases. More sophisticated methodologies (O'Brien et al. (2009); Schaerer et al. (2010); Jolly (2009)) demonstrate how to handle these challenges, but they show slightly less accurate results. Also, by making the technique specific to the LV segmentation, some methodologies (Lu et al. (2009); Huang et al. (2009); Con-

stantinides et al. (2012)) present more accurate results when compared to more general approaches (O'Brien et al. (2009); Wijnhout et al. (2009)). The main conclusion reached by the authors of the review (Petitjean and Dacher (2011)) is that the methodology presented by (Jolly (2009)) is the most competitive because it is fully automatic and offers the best compromise between accuracy and generalisation. Therefore, we regard Jolly's approach (Jolly (2009)) as our main competitor for the fully automated case. For the semi-automated case, the most competitive method in the MICCAI 2009 challenge was developed by (Huang et al. (2009)), so we consider it to be our main competitor for the semi-automated case.

Structured inference and learning is the classification problem involving a structured output (BakIr (2007)), such as the case for segmentation tasks, where the classification is represented by a multi-dimensional binary vector. Although most of the current work in computer vision and machine learning is focused on the large margin structured learning formulation (Tsochantaridis et al. (2005)), one of the most natural ways to represent a structured learning is with a multi-layer perceptron (MLP), where the output layer consists of a multi-dimensional binary vector denoting the segmentation (Collins (2002)). One of the recent breakthroughs in the field was the discovery of an efficient learning algorithm for training DBN (Hinton and Salakhutdinov (2006)), which allowed the development of structured inference and learning with DBN, as demonstrated by several works recently proposed in the field (Fasel and Berry (2010); Farabet et al. (2012); Ngo and Carneiro (2013, 2014)). The method proposed by (Farabet et al. (2012)) shows a method to parse a scene into several visual classes. Fasel et al. (Fasel and Berry (2010)) propose a DBN that takes as input an ultrasound image of the mouth and outputs a segmentation of the tongue, and (Ngo and Carneiro (2013, 2014)) propose the segmentation of the endocardium of the LV from cardiac MR cine study.

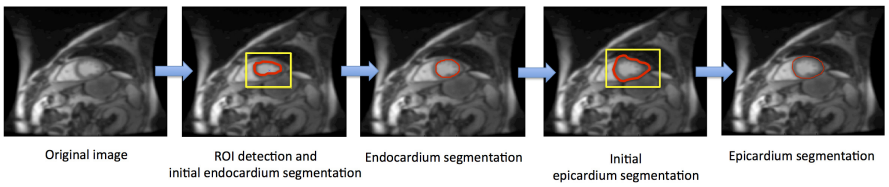


Figure 2: All steps involved in our methodology - Fig. 3 depicts each step in more detail.

3. Methodology

3.1. Notation

A cardiac cine MR series consists of a sequence of K volumes $\{V_i\}_{i=1}^{K}$, each representing a particular cardiac phase. In turn, each volume comprises a set of L images $\{I_i\}_{i=1}^{L}$ (also known as volume slices), where each image is represented by $I : \Omega \rightarrow \mathbb{R}$, with $\Omega \subseteq \mathbb{R}^2$ denoting the image coordinate space. We assume to have annotation only at the ED and ES cardiac phases (i.e., only two out of the K phases available) for all L images in these two volumes. In each of these annotated images, the explicit endocardial and epicardial contour representations are denoted by $\mathbf{c}_{\text{ENDO}} : [0,1] \rightarrow \Omega$ and $\mathbf{c}_{\text{EPI}} : [0,1] \rightarrow \Omega$, respectively. The implicit contour representation is formed with the zero level set of an Euclidean signed distance function $\phi : \Omega \rightarrow \mathbb{R}$, represented by $\mathcal{C} = \{\mathbf{x} \in \Omega | \phi(\mathbf{x}) = 0\}$, where points inside the contour have $\phi(\mathbf{x}) < 0$ and outside, $\phi(\mathbf{x}) > 0$. Assume that a set of annotated sequences is represented by $\mathcal{D} = \{(I, \mathbf{c}_{\text{ENDO}}, \mathbf{c}_{\text{EPI}}, i, q)_s\}_{i \in \{1,...,N_s\}, s \in \{1,...,S\}, q \in \{\text{ED,ES}\}}$, where $i \in \{1, ..., N_s\}$ is an index to an image within the sequence s, $q \in \{\text{ED,ES}\}$ is the annotation of the cardiac phase, $s \in \{1, ..., S\}$ is an index to a sequence and S is the number of sequences in $\mathcal{D}$. A segmentation map is represented by $\mathbf{y}_{\text{ENDO}} : \Omega \rightarrow \{0,1\}$ (or $\mathbf{y}_{\text{EPI}} : \Omega \rightarrow \{0,1\}$), where 1 represents foreground (i.e., the region inside the contour $\mathbf{c}_{\text{ENDO}}$ or $\mathbf{c}_{\text{EPI}}$) and 0 denotes background (region outside the contour). For the explanation of our methodology below, please assume that we run our segmentation slice by slice in each of the ED and ES volumes, using a sequence of steps displayed in Fig. 2.

3.2. Endocardium Segmentation

The endocardium segmentation is divided into two steps, with the first step comprising the ROI detection using structured inference on a DBN, which produces a rectangular region. Using this region as input, an initial endocardium segmentation is produced using Otsu's thresholding (Otsu (1975)) (Fig. 3-(a)). Note that Otsu's thresholding (Otsu (1975)) is a method that binarizes a gray-level image using a threshold value that is estimated in order to minimise the intra-class variance of the grey values, where the classes are defined by the pixel values above and below this threshold. The second step uses this initial segmentation to initialise an optimisation using the distance regularised level set method (DRLS) (Li et al. (2010)), which is based on an energy functional using length, area, shape prior and DBN-based appearance terms (Fig. 3-(b)). We give details about both steps below.

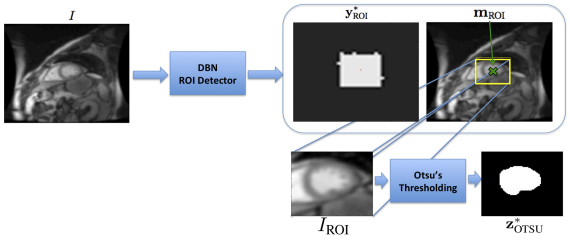
3.2.1. ROI DBN Detection and Initial Endocardium Segmentation

The ROI detection is based on a structured output inference using a DBN, which is a generative model composed of several layers of unsupervised networks, known as restricted Boltzmann machines (RBM). These RBMs have connections between layers but not between units within each layer, which facilitates the training procedure (Hinton and Salakhutdinov (2006)). The visible layers in this DBN are composed of the input image and the segmentation map (see Fig. 4). The ROI DBN detection is based on the maximisation of the following joint probability function representing a DBN model:

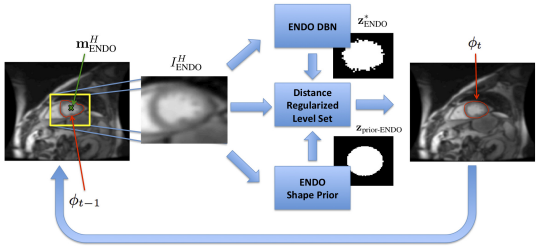
$$\mathbf{y}^*_{\mathrm{ROI}} = \arg\max_{\mathbf{y}} \int_{\mathbf{h}_1} ... \int_{\mathbf{h}_K} P(\mathbf{v}, \mathbf{h}_1, ..., \mathbf{h}_K, \mathbf{y}; \Theta_{\mathrm{ROI}}) d\mathbf{h}_1 ... d\mathbf{h}_K, \qquad (1)$$

where $\mathbf{h}_k \in \{0,1\}^{|\mathbf{h}_k|}$ represents the $|\mathbf{h}_k|$ hidden nodes of layer $k \in \{1,..,K\}$ of the DBN, $\mathbf{v}$ is a vector representation of the input image I, $\mathbf{y} : \Omega \rightarrow \{0,1\}$, and Θ_{ROI} denotes the DBN parameters (weights and biases). The probability term in (1) is computed as

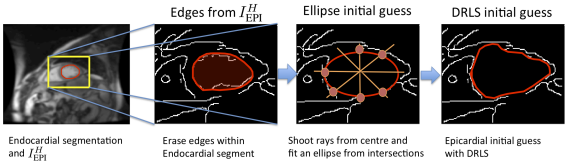
$$P(\mathbf{v}, \mathbf{h}_1, ..., \mathbf{h}_K, \mathbf{y}) = P(\mathbf{h}_K, \mathbf{h}_{K-1}, \mathbf{y}) \left(\prod_{k=1}^{K-2} P(\mathbf{h}_{k+1} | \mathbf{h}_k) \right) P(\mathbf{h}_1 | \mathbf{v}), \qquad (2)$$



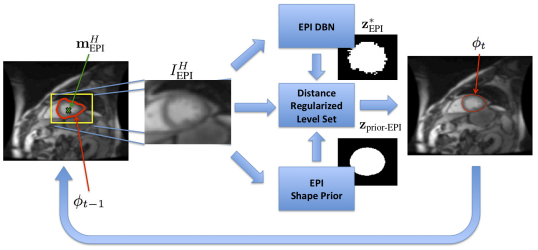
(a) ROI Detection and Initial Endocardium Segmentation



(b) Endocardium Segmentation



(c) Initial Epicardium Segmentation



(d) Epicardium Segmentation

Figure 3: Models of the ROI detection and initial endocardium segmentation (a), final endocardium segmentation (b), initial epicardium segmentation (c) and final epicardium segmentation (d).

where $P(\mathbf{h}_K, \mathbf{h}_{K-1}, \mathbf{y}) \propto \exp\{-\mathcal{E}_{\text{RBM}}(\mathbf{h}_K, \mathbf{h}_{K-1}, \mathbf{y})\}$ with

$$\begin{aligned}\mathcal{E}_{\text{RBM}}(\mathbf{h}_K, \mathbf{h}_{K-1}, \mathbf{y}) = &- \mathbf{b}_K^\top \mathbf{h}_K - \mathbf{a}_{K-1}^\top \mathbf{h}_{K-1} - \mathbf{a}_y^\top \mathbf{y} - \\ &(\mathbf{h}_K)^\top \mathbf{W}_K \mathbf{h}_{K-1} - (\mathbf{h}_K)^\top \mathbf{W}_{\mathbf{y}} \mathbf{y},\end{aligned} \tag{3}$$

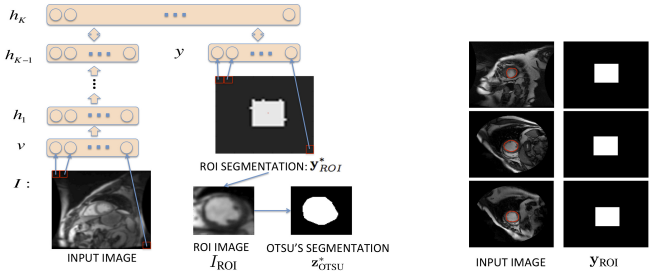
representing the energy function of an RBM (Hinton and Salakhutdinov (2006)), where $\mathbf{b}_K, \mathbf{a}_{K-1}, \mathbf{a}_y$ denote the bias vectors and $\mathbf{W}_K, \mathbf{W}_\mathbf{y}$ are the weight matrices. In (2), we also have

$$P(\mathbf{h}_{k+1}|\mathbf{h}_k) = \prod_j P(\mathbf{h}_{k+1}(j) = 1|\mathbf{h}_k), \tag{4}$$

with $P(\mathbf{h}_{k+1}(j) = 1|\mathbf{h}_k) = \sigma(\mathbf{b}_{k+1}(j) + \mathbf{h}_k^\top \mathbf{W}_{k+1}(:, j))$, $P(\mathbf{h}_1(j) = 1|\mathbf{v}) = \sigma(\mathbf{b}_1(j) + \mathbf{v}^\top \mathbf{W}_1(:, j))$ [2], where $\sigma(x) = \frac{1}{1+e^{-x}}$, the operator (j) returns the j^{th} vector value, and $(:, j)$ returns the j^{th} matrix column.

The estimation of the DBN parameter in (1) uses a training set comprising images I and their respective ROI segmentation maps $\mathbf{y}_{\text{ROI}}$. This annotation is automatically built from the manual endocardial border delineations $\mathbf{c}_{\text{ENDO}}$ (from $\mathcal{D}$, defined in Sec. 3.1), by producing a segmentation map with 0's everywhere except at a square of 1's with size M_{ROI}, centred at the centre of gravity of the annotation $\mathbf{c}_{\text{ENDO}}$ (see training samples in Fig. 4-(b)). The training process is based on the initial unsupervised bottom-up training of each pair of layers, where the DBN parameters are estimated in order to build an auto-encoder, and the top RBM is trained with an additional input containing the segmentation map $\mathbf{y}_{\text{ROI}}$ (Hinton and Salakhutdinov (2006)). The main algorithm used in this training process is the contrastive divergence, which is an approximation to gradient descent (Hinton and Salakhutdinov (2006)). Note that the DBN is a generative model, so the inference process to produce a segmentation map given an input image is based on the generation of a segmentation map when the input $\mathbf{v}$ is clamped at this input image values. More specifically, using the input image at the bottom layer, bottom-up inferences are realised with mean-field approximation until reaching the top two layers, which form an RBM. The segmentation map layer is then initialised at $\mathbf{y} = \mathbf{0}$ and we then run Gibbs sampling on the layers $\mathbf{y}$ and $\mathbf{h}_K$ until convergence (Hinton and Salakhutdinov (2006)), with $\mathbf{h}_{K-1}$ clamped from the mean-field approximation. The stable

[2] That is, we assume Gaussian visible units for the DBN with mean zero and standard deviation one.



(a) ROI DBN & Otsu's segmentation (b) Training samples

Figure 4: ROI DBN Model and Otsu's segmentation (a) and training samples for the ROI DBN (b).

vector for the layer $\mathbf{y}$ is labelled $\mathbf{y}^*_{\text{ROI}}$.

After estimating the ROI segmentation map $\mathbf{y}^*_{\text{ROI}}$, a rough endocardial border delineation is estimated by first applying the following function:

$$(I_{\text{ROI}}, \mathbf{m}_{\text{ROI}}, \mathbf{z}_{\text{ROI}}) = f_R(\mathbf{y}^*_{\text{ROI}}, I, M_{\text{ROI}}), \tag{5}$$

where $\mathbf{m}_{\text{ROI}}$ is the centre of gravity of $\mathbf{y}^*_{\text{ROI}}$ computed as $\mathbf{m}_{\text{ROI}} = \int_\Omega \mathbf{x} h(\mathbf{y}^*_{\text{ROI}}) d\mathbf{x}$, with $h(\mathbf{y}^*_{\text{ROI}}) = \frac{H(\mathbf{y}^*_{\text{ROI}})}{\int_\Omega H(\mathbf{y}^*_{\text{ROI}}) d\mathbf{x}}$ and $H(.)$ denoting the Heaviside step function, the binary map $\mathbf{z}_{\text{ROI}} : \Omega \rightarrow \{0, 1\}$ from $\mathbf{y}^*_{\text{ROI}}$ is computed with

$$\mathbf{z}_{\text{ROI}}(\mathbf{x}) = \begin{cases} 1, & \mathbf{y}^*_{\text{ROI}}(\mathbf{x}) > 0.5 \\ 0, & \text{otherwise} \end{cases}, \tag{6}$$

and I_{ROI} is a sub-image of size $M_{\text{ROI}} \times M_{\text{ROI}}$ extracted with $I_{\text{ROI}} = I(\mathbf{m}_{\text{ROI}} \pm M_{\text{ROI}}/2)$. Then, Otsu's thresholding (Otsu (1975)) is run on sub-image I_{ROI}, where the convex hull of the connected component linked to the centre $M_{\text{ROI}}/2$ is returned as the rough endocardial border delineation with $\mathbf{z}^*_{\text{OTSU}} = f_O(I_{\text{ROI}})$, as displayed in Fig. 4-(a). This segmentation is used to form the initial signed distance function, as follows:

$$\phi_0 = f_\phi(\mathbf{z}^*_{\text{OTSU}}, \mathbf{m}_{\text{ROI}}, M_{\text{ROI}}, I), \tag{7}$$

where we first create a temporary binary map $\mathbf{z} : \Omega \rightarrow \{0, 1\}$ with a map of the size of I containing only zeros, as in $\mathbf{z} = \mathbf{0}_{size(I)}$ (the function $size(i)$ returns

the size of the image), then we fill this map with the result from $\mathbf{z}^*_{\text{OTSU}}$ centred at $\mathbf{m}_{\text{ROI}}$, with $\mathbf{z}(\mathbf{m}_{\text{ROI}} \pm M_{\text{ROI}}/2) = \mathbf{z}^*_{\text{OTSU}}(M_{\text{ROI}}/2 \pm M_{\text{ROI}}/2)$. Finally, the signed distance function $\phi_0 : \Omega \rightarrow \mathbb{R}$ in (7) is computed with

$$\phi_0(\mathbf{x}) = \begin{cases} -d(\mathbf{x}, \Omega^{out}), & \text{if } \mathbf{x} \in \Omega^{in} \\ +d(\mathbf{x}, \Omega^{in}), & \text{if } \mathbf{x} \in \Omega^{out} \end{cases}, \tag{8}$$

where $\Omega^{in} = \{\mathbf{x} \in \Omega | \mathbf{z}(\mathbf{x}) = 1\}$, $\Omega^{out} = \{\mathbf{x} \in \Omega | \mathbf{z}(\mathbf{x}) = 0\}$, and $d(\mathbf{x}, \Omega) = \inf_{\mathbf{y} \in \Omega} \|\mathbf{x} - \mathbf{y}\|_2$.

3.2.2. Endocardium Segmentation Combining DRLS and DBN

Given the initial segmentation ϕ_0 defined in (7), we run an optimisation algorithm to estimate the final endocardial border using the distance regularised level set (DRLS) formulation (Li et al. (2010)), where the energy functional is represented by

$$\mathcal{E}(\phi) = \mu \mathcal{R}_p(\phi) + \mathcal{E}_{\text{ext}}(\phi, \phi_{\text{ENDO-DBN},q}, \phi_{\text{ENDO-PRIOR},q}), \tag{9}$$

where $\mathcal{R}_p(\phi) = \int_\Omega p(|\nabla \phi|) d\mathbf{x}$ (with $p(s) = 0.5(s-1)^2$) is a regularisation term that guarantees $|\nabla \phi| \approx 1$; and $\mathcal{E}_{\text{ext}}(\phi)$ is defined as (Ngo and Carneiro (2013)):

$$\begin{aligned} &\mathcal{E}_{\text{ext}}(\phi, \phi_{\text{ENDO-DBN},q}, \phi_{\text{ENDO-PRIOR},q}) = \\ &\quad \lambda \mathcal{L}(\phi) + \alpha \mathcal{A}(\phi) + \beta \mathcal{S}(\phi, \phi_{\text{ENDO-DBN},q}) + \gamma \mathcal{S}(\phi, \phi_{\text{ENDO-PRIOR},q}), \end{aligned} \tag{10}$$

where the length term $\mathcal{L}(\phi) = \int_\Omega g \delta(\phi) |\nabla \phi| d\mathbf{x}$ (with $\delta(.)$ denoting the Dirac delta function and $g = \frac{1}{1+|\nabla G_\sigma * I|}$ representing the edge indicator function), the area $\mathcal{A}(\phi) = \int_\Omega g H(-\phi) d\mathbf{x}$, and $\mathcal{S}(\phi, \phi_\kappa) = \int_\Omega (\phi(\mathbf{x}) - \phi_\kappa(\mathbf{x} + \mathbf{m}_\phi))^2 d\mathbf{x}$ (with $\kappa \in \{(\text{ENDO-DBN}, q), (\text{ENDO-PRIOR}, q)\}$, and $q \in \{\text{ED,ES}\}$) represents the shape term that drives ϕ either towards the shape $\phi_{\text{ENDO-DBN},q}$ inferred from the ENDO DBN (described below in Sec. 3.2.3) or towards the shape prior $\phi_{\text{ENDO-PRIOR},q}$ estimated from the training set (see Sec. 3.4 below). Notice that the shape term $\mathcal{S}(\phi, \phi_\kappa)$ matches the two signed distance functions using the translation invariance by intrinsic alignment (Cremers et al. (2006)), where $\mathbf{m}_\phi = \int_\Omega \mathbf{x} h(\phi(\mathbf{x})) d\mathbf{x}$ with $h(\phi) = \frac{H(-\phi)}{\int_\Omega H(-\phi) d\mathbf{x}}$ is the centre of

gravity of the segmentation from ϕ, and assuming that the shape prior represented by ϕ_κ has its centre of gravity at the origin. Note that this translation aligns the centre of gravity of ϕ_κ and ϕ. It is important to mention that when $\kappa \in \{\text{ENDO-PRIOR}, q, \text{EPI-PRIOR}, q\}$, then $\phi_\kappa(\mathbf{x} + \mathbf{m}_\phi)$ is essentially the same signed distance function translated according to $\mathbf{m}_\phi$, but when $\kappa \in \{\text{ENDO-DBN}, q, \text{EPI-DBN}, q\}$, the shape of the signed distance function changes as a function of $\mathbf{m}_\phi$. This happens because the result from the DBN segmentation changes as a function of where it is applied in the input image. The gradient flow of the energy $\mathcal{E}(\phi)$ is then defined as follows:

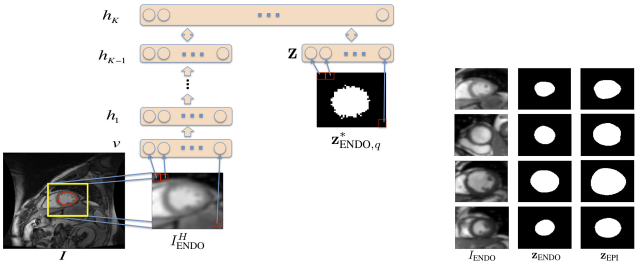
$$\begin{aligned}\frac{\partial \phi}{\partial t} =& \mu \text{div}(d_p(|\nabla\phi|)\nabla\phi) + \lambda\delta(\phi)\text{div}(g\frac{\nabla\phi}{|\nabla\phi|}) + \alpha g\delta(\phi) + \\ & 2\beta(\phi(\mathbf{x}) - \phi_{\text{ENDO-DBN},q}(\mathbf{x} + \mathbf{m}_\phi)) + \\ & 2\gamma(\phi(\mathbf{x}) - \phi_{\text{ENDO-PRIOR},q}(\mathbf{x} + \mathbf{m}_\phi)),\end{aligned} \tag{11}$$

where div(.) denotes the divergence operator, $\phi(\mathbf{x})$ denotes the current level set function, $\phi_{\text{ENDO-DBN},q}(\mathbf{x}+\mathbf{m}_\phi)$ denotes the translated signed distance function produced by the ENDO-DBN (similarly for ENDO-PRIOR), and $d_p(.)$ denotes the derivative of the function $p(.)$ defined in (9).

The estimated final endocardium segmentation is obtained from the minimisation of the energy functional in (9). In practice, the segmentation is obtained from the steady solution of the gradient flow equation (Li et al. (2010)) $\frac{\partial\phi}{\partial t} = -\frac{\partial\mathcal{E}}{\partial\phi}$, where $\partial\mathcal{E}/\partial\phi$ is the Gâteaux derivative of the functional $\mathcal{E}(\phi)$ and $\frac{\partial\phi}{\partial t}$ is defined in (11). The main idea of the DRLS (Li et al. (2010)) is then to iteratively follow the steepest descent direction (11) until convergence, resulting in the final steady solution $\phi^*_{\text{ENDO},q}$.

3.2.3. ENDO DBN

The ENDO DBN used at this stage is similar to the ROI DBN from Sec. 3.2.1, but with the following differences: 1) instead of using the whole image I as the input, we use the sub-image I^H_{ENDO} of size M_{ENDO} (centred at $\mathbf{m}^H_{\text{ENDO}}$) extracted with $(I^H_{\text{ENDO}}, \mathbf{m}^H_{\text{ENDO}}, \mathbf{z}^H_{\text{ENDO}}) = f_R(H(-\phi_{t-1}), I, M_{\text{ENDO}})$, where $f_R(.)$ is defined in (5), $H(-\phi_{t-1})$ is a binary image containing the estimation for



(a) ENDO DBN (b) Training samples

Figure 5: Graphical model for the ENDO DBN (a) and respective training samples (b).

the endocardium map from DRLS (at iteration $t-1$), I denotes the original image, and $\mathbf{z}^H_{\text{ENDO}}$ represents the binary segmentation map (of size M_{ENDO}) from $H(-\phi_{t-1})$ on sub-image I^H_{ENDO}. We estimate the parameters of two distinct DBNs, one to segment images for q = ES phase and another for q = ED phase of the cardiac cycle, where the training set is formed by samples $\{(I_{\text{ENDO}}, \mathbf{z}_{\text{ENDO}}, i, q)_s\}_{i\in\{1,...,N_s\},s\in\{1,...,S\},q\in\{\text{ED,ES}\}}$ extracted from the original training set with $(I_{\text{ENDO}}, \mathbf{m}_{\text{ENDO}}, \mathbf{z}_{\text{ENDO}}) = f_R(\mathbf{y}_{\text{ENDO}}, I, M_{\text{ENDO}})$, where $f_R(.)$ is defined in (5), and $\mathbf{y}_{\text{ENDO}}$ is the binary map formed from the original endocardium annotation $\mathbf{c}_{\text{ENDO}}$ (see Sec. 3.1). The segmentation from ENDO DBN is obtained with (see Fig. 5):

$$\mathbf{z}^*_{\text{ENDO},q} = \arg\max_{\mathbf{z}} \int ... \int P(\mathbf{v}, \mathbf{h}_1, ..., \mathbf{h}_K, \mathbf{z}; \Theta_{\text{ENDO,q}}) d\mathbf{h}_1 ... d\mathbf{h}_K, \tag{12}$$

which is defined in (1), with $\mathbf{v}$ receiving the vectorised sub-image I^H_{ENDO}. The segmentation $\mathbf{z}^*_{\text{ENDO},q}$ can then be used to define the signed distance function $\phi_{\text{ENDO-DBN},q}$ in (9) with $\phi_{\text{ENDO-DBN},q} = f_\phi(\mathbf{z}^*_{\text{ENDO},q}, \mathbf{m}^H_{\text{ENDO}}, M_{\text{ENDO}}, I)$, with $f_\phi(.)$ defined in (7). The training and inference processes for these ENDO DBNs for $q \in \{\text{ES}, \text{ED}\}$ are the same as described for the ROI DBN in Sec. 3.2.1.

3.3. Epicardium Segmentation

The epicardium segmentation also follows two steps, comprising an initial epicardium segmentation, which produces a square region containing the epicardium and an initial estimation of its border, similarly to the approach in

Sec. 3.2.1 (Fig. 3-(c)). The second step involves an optimisation with DRLS (Li et al. (2010)), similar to the one presented above in Sec. 3.2.2 (Fig. 3-(d)).

3.3.1. Initial Epicardium Segmentation

The epicardium segmentation process is initialised with a rough delineation based on the endocardium detection (see Figure 3-(c)). Specifically, after the endocardium segmentation is finalized, we estimate the borders of the epicardium segmentation by first running the Canny edge detector (Canny (1986)) that outputs the edges within the window I^H_{EPI}, produced with $(I^H_{\text{EPI}}, \mathbf{m}^H_{\text{EPI}}, \mathbf{z}^H_{\text{EPI}}) = f_R(H(-\phi^*_{\text{ENDO},q}), I, M_{\text{EPI}})$, where $\phi^*_{\text{ENDO},q}$ represents the result from the DRLS, described in Sec. 3.2.2, and $f_R(.)$ is defined in (5). The edges lying in the region where $H(-\phi^*_{\text{ENDO},q})$ equals to one (this region represents blood pool found by the endocardium segmentation) are then erased and then, by "shooting" 20 rays (18 degrees apart from each other) from the centre $\mathbf{m}^H_{\text{EPI},q}$ and recording the intersection position between each ray and the first edge it crosses, we form a set of points that are likely to belong to the endocardial border. At this stage, since it is expected that the endocardial border will be relatively close to the epicardial border, we only record the points that are within a limited range from the original endocardial border (specifically, we expect the epicardial border to be within 1.05 and 1.1 of the length of the ray from $\mathbf{m}^H_{\text{EPI}}$ to the endocardial border; otherwise no point is recorded - these numbers are estimated from the 95% confidence interval of the distance between the endocardium and epicardium annotations from the training set). Finally, by fitting an ellipse to these points and running a small number of iterations of the original DRLS (Li et al. (2010)) (which is the model in (9)-(10) with $\beta = \gamma = 0$), we form the initial epicardium segmentation that is represented by a map $\mathbf{z}^*_{\text{EPI-initial}}$, which is then used to form the initial signed distance function $\phi_0 = f_\phi(\mathbf{z}^*_{\text{EPI-initial}}, \mathbf{m}^H_{\text{EPI}}, M_{\text{EPI}}, I)$, as defined in (7).

3.3.2. Epicardium Segmentation Combining DRLS and DBN

Using the initial epicardium segmentation ϕ_0 from Sec. 3.3.1 above, we run the optimisation function as defined in (9), but with the following external energy function: $\mathcal{E}_{\text{ext}}(\phi, \phi_{\text{EPI-DBN},q}, \phi_{\text{EPI-PRIOR},q})$, with $q \in \{\text{ED,ES}\}$, where $\phi_{\text{EPI-DBN},q}$ and $\phi_{\text{EPI-PRIOR},q}$ are defined below. The final steady solution of this optimisation is represented by $\phi^*_{\text{EPI},q}$.

3.3.3. EPI DBN

The EPI DBN runs similarly to the network defined above in Sec. 3.2.3, where the input sub-image I^H_{EPI} (centred at $\mathbf{m}^H_{\text{EPI}}$) of size M_{EPI} is extracted with $(I^H_{\text{EPI}}, \mathbf{m}^H_{\text{EPI}}, \mathbf{z}^H_{\text{EPI}}) = f_R(H(-\phi_{t-1}), I, M_{\text{EPI}})$, defined in (5). We can estimate the parameters of two DBNs for $q \in \{\text{ED,ES}\}$ with the following training set $\{(I_{\text{EPI}}, \mathbf{z}_{\text{EPI}}, i, q)_s\}_{i\in\{1,\ldots,N_s\}, s\in\{1,\ldots,S\}, q\in\{ED,ES\}}$ also extracted from the original training set with $(I_{\text{EPI}}, \mathbf{m}_{\text{EPI}}, \mathbf{z}_{\text{EPI}}) = f_R(\mathbf{y}_{\text{EPI}}, I, M_{\text{EPI}})$, with $\mathbf{y}_{\text{EPI}}$ representing the binary map computed from the epicardium annotation $\mathbf{c}_{\text{EPI}}$. The inference process is the same as the one defined in (12), resulting in $\mathbf{z}^*_{\text{EPI},q}$. The signed distance function is then defined by $\phi_{\text{EPI-DBN},q} = f_\phi(\mathbf{z}^*_{\text{EPI},q}, \mathbf{m}^H_{\text{EPI}}, M_{\text{EPI}}, I)$.

3.4. Shape Prior

The shape priors are computed with the mean of the manual annotations $\mathbf{z}_{\text{ENDO}}$ and $\mathbf{z}_{\text{EPI}}$, respectively, as follows: $\bar{\mathbf{z}}_{\text{ENDO-PRIOR}}(j) = \frac{1}{SN_s}\sum_{s=1}^{S}\sum_{i=1}^{N_s}\mathbf{z}_{\text{ENDO}}(j)$, where the index j represents as specific pixel address in the window $\mathbf{z}_{\text{ENDO}}$ of size $M_{\text{ENDO}} \times M_{\text{ENDO}}$. Assuming that each element of the mean map $\bar{\mathbf{z}}_{\text{ENDO}}$ is between 0 and 1, the shape prior is computed as

$$\mathbf{z}_{\text{ENDO-PRIOR}}(j) = \begin{cases} 1, & \text{if } \bar{\mathbf{z}}_{\text{ENDO-PRIOR}}(j) > 0.5 \\ 0, & \text{if } \bar{\mathbf{z}}_{\text{ENDO-PRIOR}}(j) \leq 0.5 \end{cases} . \tag{13}$$

Fig. 6 shows $\bar{\mathbf{z}}_{\text{ENDO-PRIOR}}$ and $\mathbf{z}_{\text{ENDO-PRIOR}}$ for the ED and ES cycles (and also the epicardium prior for the ED cycle). The signed distance function for the endocardium segmentation at cardiac cycle $q \in \{\text{ED,ES}\}$ is then defined by $\phi_{\text{ENDO-PRIOR},q} = f_\phi(\mathbf{z}_{\text{ENDO-PRIOR},q}, \mathbf{m}^H_{\text{ENDO}}, M_{\text{ENDO}}, I)$. This process works in the same way for the case of epicardial shape prior.

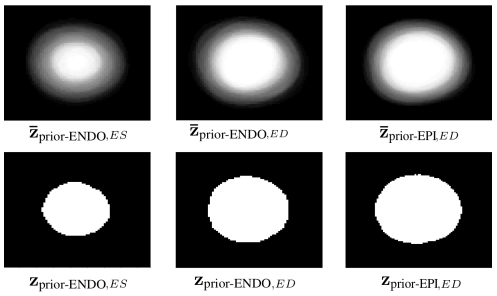


Figure 6: Shape priors for the endocardium and epicardium segmentation in ES and ED cardiac cycles. Note that for the epicardium case, the MICCAI 2009 challenge database (Radau et al. (2009)) does not contain training samples for the ES cycle.

4. Experiments

4.1. Data Set and Evaluation Measures

The accuracy of the endocardium and epicardium segmentation results produced by our methodology is assessed using the database and the evaluation proposed in the MICCAI 2009 LV segmentation challenge (Radau et al. (2009)), obtained from the Sunnybrook Health Sciences Centre, Toronto, Canada. In total, 45 cardiac short axis (SAX) cine-MR data sets are available, which are divided into three sets (online, testing and training sets) of 15 sequences, where each sequence contains four ischemic heart failures, four non-ischemic heart failures, four LV hypertrophies and three normal cases. Each of those sequences has been acquired during a 10-15 second breath-holds, with a temporal resolution of 20 cardiac phases over the heart cycle, starting from the ED cardiac phase, and containing six to 12 SAX images obtained from the atrioventricular ring to the apex (thickness=8mm, gap=8mm, FOV=320mm $\times$ 320mm, matrix= 256×256). Expert annotations are provided for endocardial contours in all slices at ED and ES cardiac phases, and for epicardial contours only at ED cardiac phase. The evaluation proposed for assessing the algorithms submitted to the MICCAI 2009 LV segmentation challenge is based on the following three measures: 1) percentage of "good" contours, 2) the average Dice metric (ADM)

of the "good" contours, and 3) average perpendicular distance (APD) of the "good" contours. A segmentation is classified as good if APD $< 5mm$.

During the MICCAI 2009 LV Segmentation Challenge (Radau et al. (2009)), the organisers first released the training and test sets, where the training set contained the manual annotation, but the test set did not include the manual annotation. The online dataset only became available a few days before the challenge day, so that the participants could submit their segmentation results for assessment. The challenge organisers reported all segmentation results for all datasets that were available from the participants. Currently all three data sets with their respective expert annotations are publicly available. Given that most of the results from the challenge participants are available for the training and test sets, we decided to use the training set to estimate all DBN parameters, the online set for validating some DBN parameters (e.g., number of layers and number of nodes per layer), and the test set exclusively for testing (since this is the set which has the majority of results from the participants).

4.2. Experimental Setup

The training set is used for estimating the ROI DBN, ENDO DBN and EPI DBN parameters (network weights and biases), the shape priors (as described in Sec. 3.4) and for estimating the weights of the DRLS method (i.e., $\mu, \lambda, \alpha, \beta, \gamma$ in (9) and (10)); while the online set is used for the model selection of the DBNs (i.e., estimation of the number of DBN hidden layers and number of nodes per layer). Specifically, we use the online set for cross validating the number of hidden layers (we test from two to four hidden layers), and the number of nodes per hidden layer (we consider ranges from 100 to 2000 nodes per layer in intervals of 100 nodes). For the ROI DBN, we reach the following configuration: 2 hidden layers with 1300 nodes in the first layer and 1500 in the second, and the input and segmentation layers with 40×40 nodes (i.e., the image is resized from 256×256 to 40×40 using standard blurring and downsampling techniques). For the ENDO DBN trained for the ED cycle, the following configuration is achieved: 2 hidden layers with 1000 nodes in the first

layer and 1000 in the second, and the input and segmentation layers with size 40×40 nodes (again, image is resized from $M_{\text{ENDO}} \times M_{\text{ENDO}}$ to 40×40). The ENDO DBN for the ES cycle has the following configuration: 2 hidden layers with 700 nodes in the first layer and 1000 in the second, and the input and segmentation layers with size 40×40. The EPI DBN for the ED cycle has the following configuration: 2 hidden layers with 1000 nodes in the first layer and 1000 in the second, and the input and segmentation layers with size 40×40 nodes (image resized from $M_{\text{EPI}} \times M_{\text{EPI}}$ to 40×40). Note that all these DBN models are trained using an augmented training set, where for each annotated training image, we generate additional ones by translating the original image (and its annotation) within a range of ± 10 pixels. More specifically, we have 105 ED images and 75 ES annotated training images (from the 15 training volumes), and in addition to the original image, we generate 40 additional images with the translations mentioned above. Therefore, in total we have 105x41=4305 annotated images for training the ED endocardial DBN and epicardial DBN, and 75x41=3075 annotated images for training the ES endocardial DBN. The segmentation accuracy on training saturates with this augmented training data (i.e., adding more translated training images no longer improves the training results).

The level set weights in (9) learned with the training set for the endocardium segmentation are $\Delta t = 2$ (time step in the level set formulation), $\mu = \frac{0.24}{\Delta t} = 0.12, \lambda = 4, \alpha = -2, \beta = 0.02$, and $\gamma = 0.001$; and for the epicardium segmentation, we have $\Delta t = 2, \mu = \frac{0.24}{\Delta t} = 0.12, \lambda = 4, \alpha = -4, \beta = 0.015$, and $\gamma = 0.001$. Note that we follow the recommendation by (Li et al. (2010)) in defining the values for Δt, and μ (the recommendations are $\Delta t > 1$ and $\mu < \frac{0.25}{\Delta t}$). For the inference procedure,the number of level set (DRLS) iterations is $T = 10$, the size of the sub-windows are set as $M_{\text{ROI}}, M_{\text{ENDO}}, M_{\text{EPI}} = 100$. We found that the segmentation results are stable if these constants are within the ranges: $T \in [5, 20]$, $M_{\text{ROI}}, M_{\text{ENDO}}, M_{\text{EPI}} \in [80, 120]$.

4.3. Results of Each Stage of the Proposed Methodology

The role of each stage of our algorithm for the endocardium segmentation is presented in Table 1. The "Initial endocardium segmentation" shows the result produced by the zero level set ϕ_0 in (7) (i.e., the result from the ROI detection, followed by the initial endocardium segmentation). The "ENDO DBN alone" displays the accuracy results of the endocardium segmentation produced by the ENDO DBN (Sec. 3.2.3) alone. The "Model without DBN/shape prior" represents the energy functional in (10) with $\beta = \gamma = 0$, which effectively represents our model without the influence of the shape prior and the ENDO DBN. Similarly the "Model without DBN" denotes the case where the functional in (10) has $\beta = 0$ (i.e., with no influence from ENDO DBN) and the "Model without shape prior" has $\gamma = 0$ (no influence from the shape prior). Finally, the "Proposed model" displays the result with all steps described in Sec. 3.2, and "Proposed model (semi)" represents our model using a manual initialisation instead of the automated initialisation described in Sec. 3.2.1. This manual initialisation consists of a circle, where the centre is the manual annotation centre of gravity and the radius is the minimum distance between the manual annotation and this centre. Table 2 shows the result of the "initial epicardium segmentation" explained in Sec. 3.3.1, and the result of the segmentation produced by the complete model described in Sec. 3.3.2 (labelled as "Proposed model"). We also show the result of the semi-automated epicardium segmentation with manual initialisation (defined in the same way as the manual initialisation above for the endocardium segmentation), labelled as "Proposed model (semi)". Note that we do not show all steps in Table 2 because the results are similar to the initial epicardium segmentation. Finally, we show that the combination of DBN and DRLS provides an accuracy improvement by running the independent two-sample t-test for the three measures considered in this paper (i.e., the good percentage, APD and ADM) for the endocardium segmentation, where the first experiment compares the measures from the proposed model (combining DBN and DRLS) and from a method consisting only of the level set without the DBN, and the second experiment compares the proposed model and the segmentation

result produced by the DBN segmentation alone. In both experiments and for all measures, the null hypothesis that the measures are drawn from independent samples from normal distributions with equal means is rejected at the 5% significance level.

Table 1: Quantitative experiments on the MICCAI 2009 challenge database (Radau et al. (2009)) showing the influence of each step of the proposed methodology for the endocardium segmentation. Each cell is formatted as "mean (standard deviation) [min value - max value]".

Method	"Good" Percentage	Endocardium ADM	Endocardium APD
Test set (15 sequences)			
Proposed model (semi)	100(0)[100 − 100]	0.91(0.03)[0.83 − 0.95]	1.79(0.36)[1.28 − 2.75]
Proposed model	95.91(5.28)[84.62 − 100]	0.88(0.03)[0.82 − 0.93]	2.34(0.46)[1.62 − 3.24]
Model without shape prior	95.71(6.96)[78.95 − 100]	0.88(0.03)[0.83 − 0.93]	2.34(0.45)[1.67 − 3.14]
Model without DBN	85.89(18.00)[36.84 − 100]	0.84(0.04)[0.77 − 0.92]	2.77(0.58)[1.73 − 3.74]
Model without DBN/shape prior	84.49(18.31)[36.84 − 100]	0.84(0.04)[0.78 − 0.92]	2.78(0.58)[1.72 − 3.81]
ENDO DBN alone	18.31(19.46)[0 − 100]	0.87(0.02)[0.84 − 0.89]	3.81(0.64)[2.97 − 4.88]
Initial endocardium segmentation	85.18(15.83)[47.37 − 100]	0.85(0.04)[0.79 − 0.92]	2.81(0.47)[2.07 − 3.58]
Training set (15 sequences)			
Proposed model (semi)	100(0)[100 − 100]	0.91(0.03)[0.85 − 0.95]	1.63(0.40)[1.29 − 2.70]
Proposed model	97.22(3.16)[91.67 − 100]	0.88(0.05)[0.76 − 0.95]	2.13(0.46)[1.27 − 2.73]
Model without shape prior	97.42(4.63)[83.33 − 100]	0.88(0.04)[0.76 − 0.95]	2.14(0.43)[1.28 − 2.63]
Model without DBN	89.42(11.83)[61.11 − 100]	0.85(0.06)[0.71 − 0.93]	2.61(0.66)[1.74 − 3.65]
Model without DBN/shape prior	88.11(13.84)[50.00 − 100]	0.84(0.06)[0.70 − 0.93]	2.57(0.62)[1.72 − 3.53]
ENDO DBN alone	48.09(38.42)[0 − 100]	0.86(0.05)[0.73 − 0.90]	3.23(0.44)[2.70 − 4.05]
Initial endocardium segmentation	89.61(11.57)[55.56 − 100]	0.85(0.06)[0.71 − 0.93]	2.71(0.57)[1.78 − 3.49]

4.4. Comparison with the State of the Art

Tables 3 and 4 shows a comparison between our methodology (labelled "Proposed model") and the state of the art for the endocardium segmentation prob-

Table 2: Quantitative experiments on the MICCAI 2009 challenge database (Radau et al. (2009)) compared different versions of the proposed methodology for the epicardium segmentation. Each cell is formatted as "mean (standard deviation) [min value - max value]".

Method	"Good" Percentage	Epicardium ADM	Epicardium APD
Test set (15 sequences)			
Proposed model (semi)	$100(0)[100-100]$	$0.94(0.01)[0.92-0.97]$	$1.73(0.28)[1.16-2.17]$
Proposed model	$94.65(6.18)[85.71-100]$	$0.93(0.02)[0.88-0.96]$	$2.08(0.60)[1.27-3.74]$
Initial epicardium segmentation	$94.65(6.18)[85.71-100]$	$0.93(0.02)[0.88-0.96]$	$2.19(0.58)[1.32-3.68]$
Training set (15 sequences)			
Proposed model (semi)	$100.00(0.00)[100-100]$	$0.94(0.01)[0.91-0.96]$	$1.64(0.34)[1.17-2.47]$
Proposed model	$98.52(5.74)[77.78-100]$	$0.93(0.02)[0.89-0.96]$	$1.99(0.46)[1.35-3.13]$
Initial epicardium segmentation	$96.83(6.92)[77.78-100$	$0.93(0.02)[0.89-0.95]$	$1.99(0.40)[1.46-3.14]$

lem, while Tables 5 and 6 displays a similar comparison for the epicardium problem for different subsets of the MICCAI 2009 challenge databases (Radau et al. (2009)). Most of the approaches on that table are based on active contour models (Constantinides et al. (2012); Huang et al. (2009, 2011); Jolly (2009); Lu et al. (2009); Marak et al. (2009)), machine learning models (O'Brien et al. (2009); Wijnhout et al. (2009)), or a combination of both models (Hu et al. (2012)). Furthermore, Tables 3-6 also show a semi-automated version of our method (labelled "Proposed model (semi)") using the same initial guess described above in Sec. 4.3.

Fig. 7 shows a few endocardium and epicardium segmentation results produced by our approach for challenging cases, such as with images from apical and basal slice images and presenting papillary muscles and trabeculations (please see supplementary material for more results). Finally, Fig. 8 shows a few unprocessed 3-D models of the endocardial and epicardial borders obtained with our proposed methodology.

Table 3: Quantitative experiments on the **training and test sets** of the MICCAI 2009 challenge databases (Radau et al. (2009)) comparing the performance of our proposed approach with the state of the art on the **endocardium segmentation problem**. Notice that the methods are classified into fully or semi-automated. The cell formatting is the same as in Tab. 1, but note that '?' means that the result is not available in the literature. The top performance for each measure and dataset is highlighted.

Method	"Good" Percentage	Endocardium ADM	Endocardium APD
Test set (15 sequences)			
Semi Automated			
Proposed model (semi)	**100(0)[100 − 100]**	**0.91(0.03)[0.83 − 0.95]**	**1.79(0.36)[1.28 − 2.75]**
Ngo and Carneiro (2013)	96.58(9.58)[63.15 − 100]	0.89(0.03)[0.83 − 0.93]	2.22(0.46)[1.69 − 3.30]
Huang et al. (2009)	?	0.89(0.04)[?−?]	2.10(0.44)[?−?]
Fully Automated			
Proposed model	**95.91(5.28)[84.62 − 100]**	0.88(0.03)[0.82 − 0.93]	2.34(0.46)[1.62 − 3.24]
Jolly (2009)	94.33(9.93)[62.00 − 100]	0.88(0.03)[0.84 − 0.94]	2.44(0.62)[1.36 − 3.68]
Wijnhout et al. (2009)	86.47(11.00)[68.4 − 100]	**0.89(0.03)[0.82 − 0.94]**	2.29(0.57)[1.67 − 3.93]
Lu et al. (2009)	72.45(19.52)[42.11 − 100]	**0.89(0.03)[0.84 − 0.94]**	**2.07(0.61)[1.32 − 3.77]**
Marak et al. (2009)	?	0.86(0.04)[?−?]	?
O'Brien et al. (2009)	?	0.81(?)[?−?]	?

Training set (15 sequences)			
Semi Automated			
Proposed model (semi)	**100(0)[100 − 100]**	**0.91(0.03)[0.85 − 0.95]**	**1.63(0.40)[1.29 − 2.70]**
Ngo and Carneiro (2013)	98.45(3.11)[91.66 − 100]	0.90(0.03)[0.84 − 0.94]	1.96(0.35)[1.43 − 2.55]
Huang et al. (2009)	?	**0.90(0.04)**[?−?]	2.03(0.34)[?−?]
Fully Automated			
Proposed model	**97.22(3.16)[91.67 − 100]**	**0.88(0.05)[0.76 − 0.95]**	2.13(0.46)[1.27 − 2.73]
Jolly (2009)	96.93(7.59)[72 − 100]	**0.88(0.06)[0.75 − 0.95]**	**2.09(0.53)[1.35 − 3.23]**

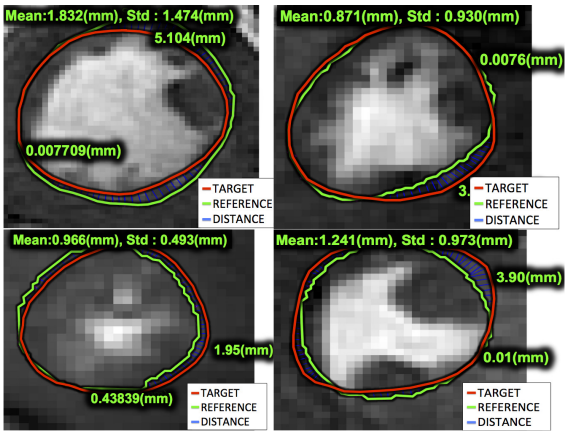
Table 4: Quantitative experiments on the **online and full sets** of the MICCAI 2009 challenge databases (Radau et al. (2009)) comparing the performance of our proposed approach with the state of the art on the **endocardium segmentation problem**. Notice that the methods are classified into fully or semi-automated. The cell formatting is the same as in Tab. 1, but note that '?' means that the result is not available in the literature. The top performance for each measure and dataset is highlighted.

Method	"Good" Percentage	Endocardium ADM	Endocardium APD
Online set (15 sequences)			
Semi Automated			
Proposed model (semi)	**100(0)[100 − 100]**	**0.91(0.03)[0.85 − 0.96]**	**1.78(0.49)[1.17 − 3.15]**
Ngo and Carneiro (2013)	98.71(3.66)[86.66 − 100]	0.90(0.04)[0.83 − 0.95]	2.04(0.35)[1.53 − 2.67]
Fully Automated			
Proposed model	**90.54(14.40)[46.67 − 100]**	**0.89(0.03)[0.82 − 0.94]**	**2.17(0.46)[1.62 − 3.46]**

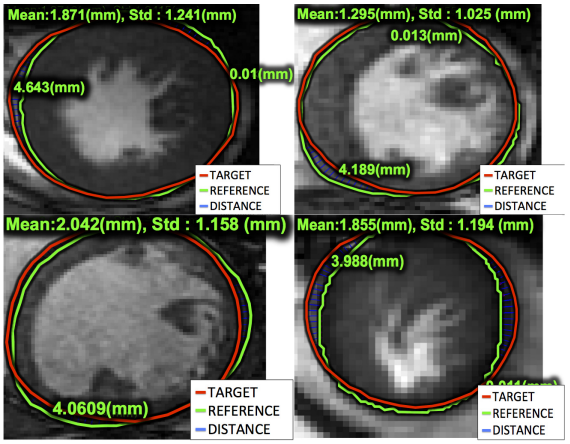
Full set (45 sequences)			
Semi Automated			
Proposed model (semi)	**100(0)[100 − 100]**	**0.91(0.03)[0.83 − 0.96]**	**1.73(0.31)[1.17 − 3.15]**
Ngo and Carneiro (2013)	97.91(6.18)[63.15 − 100]	0.90(0.03)[0.83 − 0.95]	2.08(0.40)[1.43 − 3.30]
Constantinides et al. (2012)	91.00(8.00)[61 − 100]	0.89(0.04)[0.80 − 0.96]	1.94(0.42)[1.47 − 3.03]
Fully Automated			
Proposed model	**94.55(9.31)[46.67 − 100]**	0.88(0.04)[0.76 − 0.95]	2.22(0.46)[01.27 − 3.46]
Constantinides et al. (2012)	80.00(16.00)[29 − 100]	0.86(0.05)[0.72 − 0.94]	2.44(0.56)[1.31 − 4.20]
Hu et al. (2012)	91.06(9.42)[?−?]	**0.89(0.03)**[?−?]	2.24(0.40)[?−?]
Huang et al. (2011)	79.20(19.00)[?−?]	**0.89(0.04)**[?−?]	**2.16(0.46)**[?−?]

5. Discussion and Conclusions

The role of each stage of our methodology for the endocardium segmentation becomes clear with the results presented in Table 1. For instance, the DRLS method alone (Li et al. (2010)) (i.e., without the prior and ENDO DBN terms)



a) Results of endocardium segmentation on the test set



b) Results of epicardium segmentation on the test set

Figure 7: Epicardium and endocardium segmentation results with challenging cases, such as images from apical and basal slice images and presenting papillary muscles and trabeculations. The red contour denotes the automated detection, and green shows the manual annotation. For more results, please see the supplementary material.

is not able to improve significantly the result from the initial endocardium segmentation. The addition of the shape prior term improves slightly the accuracy (see row "Model without DBN"), but not significantly so; therefore we can removed it from the framework in order to obtain small gains in terms of efficiency.

Table 5: Quantitative experiments on the **training and test sets** of the MICCAI 2009 challenge databases (Radau et al. (2009)) comparing the performance of our proposed approach with the state of the art on the **epicardium segmentation problem**. Notice that the methods are classified into fully or semi-automated. The cell formatting is the same as in Tab. 1, but note that '?' means that the result is not available in the literature. The top performance for each measure and dataset is highlighted.

Method	"Good" Percentage	Epicardium ADM	Epicardium APD
Test set (15 sequences)			
Semi Automated			
Proposed model (semi)	**100(0)[100 − 100]**	**0.94(0.01)[0.92 − 0.97]**	**1.73(0.28)[1.16 − 2.17]**
Huang et al. (2009)	?	**0.94(0.01)**[?−?]	1.95(0.34)[?−?]
Fully Automated			
Proposed model	94.65(6.18)[85.71 − 100]	0.93(0.02)[0.88 − 0.96]	2.08(0.60)[1.27 − 3.74]
Jolly (2009)	**95.60(6.90)[80.00 − 100]**	0.93(0.02)[0.90 − 0.96]	2.05(0.59)[1.28 − 3.29]
Wijnhout et al. (2009)	94.20(7.00)[80.00 − 100]	0.93(0.01)[0.90 − 0.96]	2.28(0.39)[1.57 − 2.98]
Lu et al. (2009)	81.11(13.95)[57.14 − 100]	**0.94(0.02)[0.90 − 0.97]**	**1.91(0.63)[1.06 − 3.26]**

Training set (15 sequences)			
Semi Automated			
Proposed model (semi)	**100.00(0.00)[100 − 100]**	**0.94(0.01)[0.91 − 0.96]**	**1.64(0.34)[1.17 − 2.47]**
Huang et al. (2009)	?	0.93(0.02)[?−?]	2.28(0.42)[?−?]
Fully Automated			
Proposed model	98.52(5.74)[77.78 − 100]	**0.93(0.02)[0.88 − 0.96]**	1.99(0.46)[1.35 − 3.13]
Jolly (2009)	**99.07(3.61)[86.00 − 100]**	**0.93(0.01)[0.91 − 0.95]**	**1.88(0.40)[1.20 − 2.55]**

ENDO DBN (see row "Model without shape prior") is the term that provides the largest gain in terms of accuracy, even though its performance as a stand alone segmentation system is not competitive. This indicates that the results produced by ENDO DBN complements the results from DRLS using the information available (and automatically learned) from the training set. Putting

Table 6: Quantitative experiments on the **online and full sets** of the MICCAI 2009 challenge databases (Radau et al. (2009)) comparing the performance of our proposed approach with the state of the art on the **epicardium segmentation problem**. Notice that the methods are classified into fully or semi-automated. The cell formatting is the same as in Tab. 1, but note that '?' means that the result is not available in the literature. The top performance for each measure and dataset is highlighted.

Method	"Good" Percentage	Epicardium ADM	Epicardium APD
Online set (15 sequences)			
Semi Automated			
Proposed model (semi)	**100.00(0.00)[100 − 100]**	**0.94(0.02)[0.88 − 0.96]**	**1.90(0.53)[1.22 − 3.16]**
Fully Automated			
Proposed model	**84.32(23.45)[12.50 − 100]**	**0.93(0.03)[0.84 − 0.95]**	**2.05(0.61)[1.39 − 3.63]**
Full set (45 sequences)			
Semi Automated			
Proposed model (semi)	**100(0)[100 − 100]**	**0.94(0.02)[0.88 − 0.97]**	**1.76(0.40)[1.16 − 3.16]**
Constantinides et al. (2012)	91.00(10.00)[70 − 100]	0.92(0.02)[0.84 − 0.95]	2.38(0.57)[1.28 − 3.79]
Fully Automated			
Proposed model	**92.49(15.31)[12.50 − 100]**	0.93(0.02)[0.84 − 0.96]	**2.04(0.55)[1.27 − 3.70]**
Constantinides et al. (2012)	71.00(26.00)[0 − 100]	0.91(0.03)[0.81 − 0.96]	2.80(0.71)[1.37 − 4.88]
Hu et al. (2012)	91.21(8.52)[?−?]	**0.94(0.02)**[?−?]	2.21(0.45)[?−?]
Huang et al. (2011)	83.90(16.80)[?−?]	0.93(0.02)[?−?]	2.22(0.43)[?−?]

all terms together, the "Proposed model" displays the best performance of our method, which is shown to be statistically significantly superior to both the DRLS and DBN methods. It is important to notice the relative small accuracy differences between the training and test sets, which indicates good generalisation capabilities of our method (even with the relatively small training set of the MICCAI 2009 challenge database (Radau et al. (2009))). Finally, by using a manual initialization, note that we obtain the best result in the field.

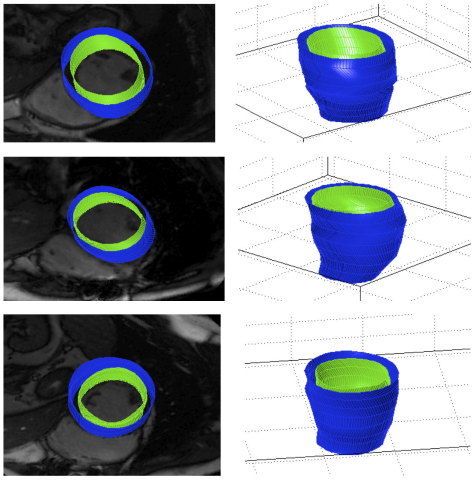


Figure 8: 3D Model formed by linking the slice by slice results for the endocardial (green surface) and epicardial (blue) borders in ED cycle (note that we focus on ED cardiac cycle because no annotated epicardium training set is available for the ES cycle from the MICCAI 2009 challenge database (Radau et al. (2009))).

Table 2 shows that the initial epicardium segmentation already produces a result that is close to the result produced by our proposed model. Therefore, even though we notice that the use of the EPI DBN also improves the result, it is only a slight improvement that mostly happens on the training set. Furthermore, similarly to the endocardium segmentation, the use of manual initialisation also shows the best result in the field. Finally, given the similar appearance of the endocardium and epicardium images, it is important to justify the need for learning two separate DBN models, that is the ENDO and EPI DBNs, instead of a single one estimated with all training sets. The main reason for these two models lies in the empirical evidence that they produce more accurate segmentation results, as shown in Tab. 3-4, where the rows labelled by **Proposed model (semi)** show the results with the two separate DBNs, while the rows labelled by Ngo and Carneiro (2013) display results using a single classifier.

The comparison with the state of the art in terms of the endocardium seg-

mentation in Tables 3-4 and the epicardium segmentation in Tables 5-6 shows that our approach produces the best results in the field for the semi-automated segmentation problem. For the fully automated segmentation problem our results is on par with the result of the method proposed by (Jolly (2009)), which is considered to be the current state of the art by a recent review paper by (Petitjean and Dacher (2011)). In general, for the endocardium segmentation, our results are better in terms of "Good" percentage than other methods but comparable to the best ones with respect to ADM and APD. For the epicardium segmentation our results are comparable to the method proposed by (Jolly (2009)), but better than all others. Note that while some approaches are more accurate in terms of APD or ADM (Lu et al. (2009)), they also present low values for "Good" percentage, which means that these methods also produce a large number of segmentations with APD larger than $5mm$, but the few ones that survive the "Good" percentage test are reasonably accurate. Another important observation is the relatively worse performance of the fully automated compared to semi-automated segmentation (not only for our proposed method, but other methods proposed in the literature), indicating that there is still room for improving the accuracy of the initial endocardium and epicardium segmentations. It is also important to mention that our approach runs on (mean) average in 175 ± 35 seconds for the endocardium segmentation and 119 ± 20 seconds for the epicardium segmentation using a *non-optimised Matlab program*, which is slower or comparable to other aproaches that run between one minute (Constantinides et al. (2012); Jolly (2009); Wijnhout et al. (2009)) and three minutes (Hu et al. (2012); Lu et al. (2009)).

There are several points that can be explored in order to improve the results of the endocardium and epicardium segmentation. First, instead of running the segmentation algorithm slice by slice, we can run it over the whole volume and use a 3-D shape model to constrain the search process. Second, we can also use a motion model as another constraint for the segmentation process. Third, if new training sets become available in the field, we can train more complex DBN models that can potentially produce more accurate segmentation

results. Finally, we can decrease the running time of our approach by parallelizing the segmentation processes since the segmentation of each slice is done independently of all others (roughly this means that we can in principle make our approach 10 times faster).

Acknowledgements

This work was partially supported by the Australian Research Council's Discovery Projects funding scheme (project DP140102794). Tuan Anh Ngo acknowledges the support of the 322 Program - Vietnam International Education Development, Ministry of Education and Training (VIED-MOET).

References

BakIr, G., 2007. Predicting structured data. MIT press.

Canny, J., 1986. A computational approach to edge detection. Pattern Analysis and Machine Intelligence, IEEE Transactions on , 679–698.

Carneiro, G., Nascimento, J.C., 2013. Combining multiple dynamic models and deep learning architectures for tracking the left ventricle endocardium in ultrasound data. Pattern Analysis and Machine Intelligence, IEEE Transactions on 35, 2592–2607.

Carneiro, G., Nascimento, J.C., Freitas, A., 2012. The segmentation of the left ventricle of the heart from ultrasound data using deep learning architectures and derivative-based search methods. Image Processing, IEEE Transactions on 21, 968–982.

Cobzas, D., Schmidt, M., 2009. Increased discrimination in level set methods with embedded conditional random fields, in: Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on, IEEE. pp. 328–335.

Collins, M., 2002. Discriminative training methods for hidden markov models: Theory and experiments with perceptron algorithms, in: Proceedings of the

ACL-02 conference on Empirical methods in natural language processing-Volume 10, Association for Computational Linguistics. pp. 1–8.

Constantinides, C., Roullot, E., Lefort, M., Frouin, F., 2012. Fully automated segmentation of the left ventricle applied to cine mr images: Description and results on a database of 45 subjects, in: Engineering in Medicine and Biology Society (EMBC), 2012 Annual International Conference of the IEEE, IEEE. pp. 3207–3210.

Cootes, T.F., Taylor, C.J., Cooper, D.H., Graham, J., 1995. Active shape models-their training and application. Computer vision and image understanding 61, 38–59.

Cortes, C., Vapnik, V., 1995. Support vector machine. Machine learning 20, 273–297.

Cremers, D., Osher, S.J., Soatto, S., 2006. Kernel density estimation and intrinsic alignment for shape priors in level set segmentation. International Journal of Computer Vision 69, 335–351.

Farabet, C., Couprie, C., Najman, L., LeCun, Y., 2012. Scene parsing with multiscale feature learning, purity trees, and optimal covers. arXiv preprint arXiv:1202.2160 .

Fasel, I., Berry, J., 2010. Deep belief networks for real-time extraction of tongue contours from ultrasound during speech, in: Pattern Recognition (ICPR), 2010 20th International Conference on, IEEE. pp. 1493–1496.

Freund, Y., Schapire, R.E., 1995. A desicion-theoretic generalization of on-line learning and an application to boosting, in: Computational learning theory, Springer. pp. 23–37.

Georgescu, B., Zhou, X.S., Comaniciu, D., Gupta, A., 2005. Databased-guided segmentation of anatomical structures with complex appearance, in: CVPR.

Hinton, G., Salakhutdinov, R., 2006. Reducing the dimensionality of data with neural networks. Science 313, 504–507.

Hu, H., Liu, H., Gao, Z., Huang, L., 2012. Hybrid segmentation of left ventricle in cardiac mri using gaussian-mixture model and region restricted dynamic programming. Magnetic resonance imaging .

Huang, R., Pavlovic, V., Metaxas, D.N., 2004. A graphical model framework for coupling mrfs and deformable models, in: Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on, IEEE. pp. II–739.

Huang, S., Liu, J., Lee, L., Venkatesh, S., Teo, L., Au, C., Nowinski, W., 2009. Segmentation of the left ventricle from cine mr images using a comprehensive approach. The MIDAS Journal 49.

Huang, S., Liu, J., Lee, L.C., Venkatesh, S.K., San Teo, L.L., Au, C., Nowinski, W.L., 2011. An image-based comprehensive approach for automatic segmentation of left ventricle from cardiac short axis cine mr images. Journal of digital imaging 24, 598–608.

Jolly, M., 2009. Fully automatic left ventricle segmentation in cardiac cine mr images using registration and minimum surfaces. The MIDAS Journal 49.

Kass, M., Witkin, A., Terzopoulos, D., 1988. Snakes: Active contour models. International journal of computer vision 1, 321–331.

Li, C., Xu, C., Gui, C., Fox, M.D., 2010. Distance regularized level set evolution and its application to image segmentation. Image Processing, IEEE Transactions on 19, 3243–3254.

Lu, Y., Radau, P., Connelly, K., Dick, A., Wright, G., 2009. Automatic image-driven segmentation of left ventricle in cardiac cine mri. The MIDAS Journal 49.

Marak, L., Cousty, J., Najman, L., Talbot, H., et al., 2009. 4d morphological segmentation and the miccai lv-segmentation grand challenge, in: MICCAI 2009 Workshop on Cardiac MR Left Ventricle Segmentation Challenge, pp. 1–8.

Ngo, T.A., Carneiro, G., 2013. Left ventricle segmentation from cardiac mri combining level set methods with deep belief networks, in: Image Processing (ICIP), 2013 20th IEEE International Conference on, IEEE. pp. 695–699.

Ngo, T.A., Carneiro, G., 2014. Fully automated non-rigid segmentation with distance regularized level set evolution initialized and cosntrained by deep-structured inference, in: Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on, IEEE.

Osher, S., Sethian, J.A., 1988. Fronts propagating with curvature-dependent speed: algorithms based on hamilton-jacobi formulations. Journal of computational physics 79, 12–49.

Otsu, N., 1975. A threshold selection method from gray-level histograms. Automatica 11, 23–27.

O'Brien, S., Ghita, O., Whelan, P., 2009. Segmenting the left ventricle in 3d using a coupled asm and a learned non-rigid spatial model. The MIDAS Journal 49.

Petitjean, C., Dacher, J.N., 2011. A review of segmentation methods in short axis cardiac mr images. Medical Image Analysis 15, 169–184.

Radau, P., Lu, Y., Connelly, K., Paul, G., Dick, A., Wright, G., 2009. Evaluation framework for algorithms segmenting short axis cardiac mri. MIDAS J.Cardiac MR Left Ventricle Segmentation Challenge .

Schaerer, J., Casta, C., Pousin, J., Clarysse, P., 2010. A dynamic elastic model for segmentation and tracking of the heart in mr image sequences. Medical Image Analysis 14, 738–749.

Tsechpenakis, G., Metaxas, D.N., 2007. Crf-driven implicit deformable model, in: Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on, IEEE. pp. 1–8.

Tsochantaridis, I., Joachims, T., Hofmann, T., Altun, Y., Singer, Y., 2005. Large margin methods for structured and interdependent output variables. Journal of Machine Learning Research 6.

Wijnhout, J., Hendriksen, D., Assen, H., der Geest, R., 2009. Lv challenge lkeb contribution: Fully automated myocardial contour detection. The MIDAS Journal 43.

Biography

Tuan Anh Ngo received the B.S. degree in mathematics from Hanoi National University of Education, Hanoi, Vietnam, the M.S. degree in computer science from Asian Institute of Technology, Bangkok, Thailand, and currently he is a PhD candidate at the School of Computer Science of the University of Adelaide. He was a lecturer at Department of Computer Science, Faculty of Information Technology, Vietnam National University of Agriculture, from 1996 to 2003. He became deputy head from 2005 to 2007, and head of this department from 2007 to 2011. His research interests include medical image analysis and optimisation methods based on deep learning and level sets.

Gustavo Carneiro received the BS and MSc degrees in computer science from the Federal University of Rio de Janeiro and the Military Institute of Engineering, Brazil, in 1996 and 1999, respectively, and the PhD degree in computer science from the University of Toronto, Canada, in 2004. Currently, he is an associate professor at the School of Computer Science of the University of Adelaide, Australia. Previously, he worked at the Instituto Superior Tecnico of Lisbon, from 2008 to 2011 as a visiting researcher and assistant professor, and from 2006 to 2008, he worked at Siemens Corporate Research in Princeton, New Jersey. His research interests include medical image analysis, image

feature selection and extraction, content-based image retrieval and annotation, and general visual object classification. He has authored more than 50 peer reviewed publications in international journals and conferences.

Chapter 7

Lung Segmentation in Chest Radiographs using Distance Regularized Level Set and Deep-structured Learning and Inference

---

Tuan Anh Ngo, Gustavo Carneiro

Australian Centre for Visual Technologies
University of Adelaide, Australia

The work contained in this chapter has been accepted by
IEEE International Conference on Image Processing (ICIP), 2015

Tuan Anh Ngo, Gustavo Carneiro. Lung Segmentation in Chest Radiographs using Distance Regularized Level Set and Deep-structured Learning and Inference. *IEEE International Conference on Image Processing (ICIP)*, 2015.

Statement of Authorship

Title of Paper	Lung Segmentation in Chest Radiographs using Distance Regularized Level Set and Deep-structured Learning and Inference.
Publication Status	☒ Published ☐ Accepted for Publication ☐ Submitted for Publication ☐ Publication Style
Publication Details	Tuan Anh Ngo, Gustavo Carneiro. Lung Segmentation in Chest Radiographs using Distance Regularized Level Set and Deep-structured Learning and Inference. IEEE International Conference on Image Processing (ICIP), 2015.

Principal Author

Name of Principal Author (Candidate)	Tuan Anh Ngo		
Contribution to the Paper	- Checked the database - Coded and ran the experiments - Checked the manuscript - Acted as corresponding author		
Overall percentage (%)	50%		
Signature		Date	04/08/2015

Co-Author Contributions

By signing the Statement of Authorship, each author certifies that:

i. the candidate's stated contribution to the publication is accurate (as detailed above);

ii. permission is granted for the candidate in include the publication in the thesis; and

iii. the sum of all co-author contributions is equal to 100% less the candidate's stated contribution.

Name of Co-Author	Gustavo Carneiro		
Contribution to the Paper	- Checked the database - Supervised the development of this work - Wrote the manuscript		
Signature		Date	04-08-2015

Name of Co-Author			
Contribution to the Paper			
Signature		Date	

Please cut and paste additional co-author panels here as required.

LUNG SEGMENTATION IN CHEST RADIOGRAPHS USING DISTANCE REGULARIZED LEVEL SET AND DEEP-STRUCTURED LEARNING AND INFERENCE

Tuan Anh Ngo *Gustavo Carneiro*

Australia Centre for Visual Technologies
The University of Adelaide, Australia

ABSTRACT

Computer-aided diagnosis of digital chest X-ray (CXR) images critically depends on the automated segmentation of the lungs, which is a challenging problem due to the presence of strong edges at the rib cage and clavicle, the lack of a consistent lung shape among different individuals, and the appearance of the lung apex. From recently published results in this area, hybrid methodologies based on a combination of different techniques (e.g., pixel classification and deformable models) are producing the most accurate lung segmentation results. In this paper, we propose a new methodology for lung segmentation in CXR using a hybrid method based on a combination of distance regularized level set and deep structured inference. This combination brings together the advantages of deep learning methods (robust training with few annotated samples and top-down segmentation with structured inference and learning) and level set methods (use of shape and appearance priors and efficient optimization techniques). Using the publicly available Japanese Society of Radiological Technology (JSRT) dataset, we show that our approach produces the most accurate lung segmentation results in the field. In particular, depending on the initialization used, our methodology produces an average accuracy on JSTR that varies from 94.8% to 98.5%.

Index Terms— Lung segmentation, Deep learning, Level set methods

1. INTRODUCTION

The automated segmentation of lung boundaries from digital chest X-ray (CXR) is one of the main stages in the computer-aided diagnosis (CAD) of lung health [1]. Lung boundaries can be used for computing lung volume or estimating shape irregularities [2], but it is also used as one of the stages in several CAD systems [6]. These CAD systems are particularly important for screening and detecting pulmonary pathologies, but with a major focus on tuberculosis, which is the second leading cause of death from infectious disease worldwide [3].

This work was partially supported by the Australian Research Council's Discovery Projects funding scheme (project DP140102794). Tuan Anh Ngo acknowledges the support of the 322 Program - Vietnam International Education Development, Ministry of Education and Training (VIED-MOET).

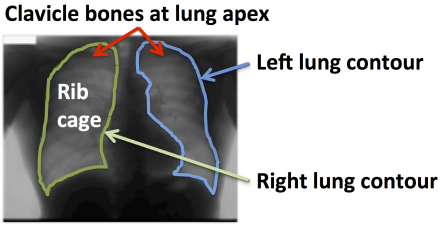


Fig. 1. Left and right lungs segmentation.

The highest incidences of these diseases occur in places of the world with inadequate health care infrastructure, so the deployment of such CAD systems in these places is important because they can help local clinicians in the screening and diagnosis processes mentioned above [1]. However, the automated segmentation of lung boundaries is a challenging task because of the following reasons (see Fig. 1) [1]: 1) the edges present at the rib cage and clavicle represent a challenge for optimization methods that can get stuck at local minima; 2) the appearance inconsistencies caused by the clavicle bone at the lung apex also represent an issue for most optimization approaches for the same reason above; and 3) the lack of a consistent lung shape among different individuals is a challenge for the use of shape priors.

There has been considerable effort applied in the development of automated lung segmentation methods [13], and the most successful approaches are usually based on hybrid methods that combine several techniques, such as methods that combine landmark learning with active shape and appearance models [11,12] or graph cuts with non-rigid registration [1]. Similarly, we propose a hybrid method based on a recent methodology that we developed for the problem of left ventricle segmentation from magnetic resonance image (MRI) [8], which has recently achieved the best results in the field. The extension of this methodology to this new problem requires some modifications to the original algorithm, but it is interesting to note that the core steps have remained almost unaltered, showing that this algorithm can be potentially applied to other similar problems.

The method being proposed in this paper for segmenting lungs from CXR images is based on the combination of distance regularized level set (DRLS) [7] and deep structured learning and inference using a deep belief network (DBN) [5]. Essentially, we use the DRLS [7] optimization with the usual shape and appearance terms, but with an additional term based on the top-down segmentation produced by a deep structured inference. This combination aims at exploring the advantages of both approaches, which are the efficient optimization and the prior shape and appearance terms from DRLS, and the robust statistical segmentation models produced by deep learning methods. We test our approach using the Japanese Society of Radiological Technology (JSRT) dataset, and our results show that, depending on the initialization used, our methodology can produce an average accuracy on JSTR that varies from 94.8% to 98.5%, which is significantly better than the current best approach in the field [1] (that reported an accuracy of 95.4%).

2. METHODOLOGY

In this section, assume that the annotated chest radiograph database is represented by $\mathcal{D} = \{(I, \mathbf{c}, q)_i\}_{i=1}^{|\mathcal{D}|}$, where $I : \Omega \rightarrow \mathbb{R}$ represents an image (with $\Omega \subseteq \mathbb{R}^2$ denoting the image lattice), $q \in \{\text{left lung,right lung}\}$ and $\mathbf{c} : [0, 1] \rightarrow \Omega$ denotes the explicit contour representation of the segmentation. Also assume that the implicit contour representation is the zero level set of a signed distance function $\phi : \Omega \rightarrow \mathbb{R}$, and the lung segmentation map is represented by $\mathbf{y} : \Omega \rightarrow \{0, 1\}$, where 1 represents the foreground (i.e., left or right lung) and 0 denotes the background. Below, we first explain the DRLS method, then we describe the DBN segmentation model, and finally explain the combined inference algorithm.

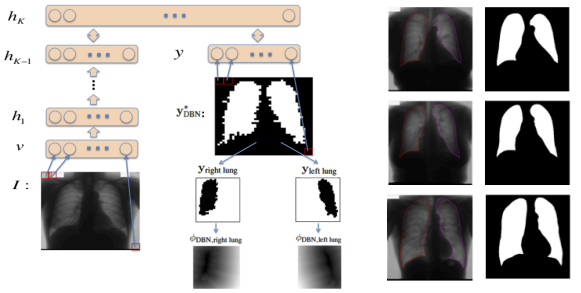
The main optimization method for producing the segmentation is based on the DRLS formulation [7], where the energy functional is represented by:

$$\mathcal{E}(\phi, \phi_{\text{DBN},q}) = \mu\mathcal{R}_p(\phi) + \mathcal{E}_{\text{ext}}(\phi, \phi_{\text{DBN},q}), \quad (1)$$

with the distance regularization $\mathcal{R}_p(\phi) = \int_\Omega 0.5(|\nabla\phi| - 1)^2 dx$ (this guarantees that $|\nabla\phi| \approx 1$), and

$$\mathcal{E}_{\text{ext}}(\phi, \phi_{\text{DBN},q}) = \lambda\mathcal{L}(\phi) + \alpha\mathcal{A}(\phi) + \gamma\mathcal{S}(\phi, \phi_{\text{DBN},q}), \quad (2)$$

where the length term $\mathcal{L}(\phi) = \int_\Omega g\delta(\phi)|\nabla\phi|dx$ (with $\delta(.)$ denoting the Dirac delta function and $g = \frac{1}{1+|\nabla G_\sigma * I|}$ representing the edge indicator function), the area $\mathcal{A}(\phi) = \int_\Omega gH(-\phi)dx$ (with $H(.)$ denoting the Heaviside step function), and $\mathcal{S}(\phi, \phi_{\text{DBN},q}) = \int_\Omega(\phi - \phi_{\text{DBN},q})^2 dx$ represents the shape term that drives the ϕ towards the shape $\phi_{\text{DBN},q}$, which is the distance function inferred from the deep belief network (DBN) structured inference described below (see Fig. 2-(a)). The minimization of the energy functional in (1) is achieved by finding the steady solution of the gradient flow equation $\frac{\partial\phi}{\partial t} = -\frac{\partial\mathcal{E}}{\partial\phi}$ [7].



(a) DBN model (b) Training samples

Fig. 2. DBN Model (a) and training samples for the DBN (b).

The DBN structured inference produces the following segmentation map (Fig, 2-(a)):

$$\mathbf{y}^*_{\text{DBN},q} = \arg\max_{\mathbf{y}} \sum_{\mathbf{h}_1} .. \sum_{\mathbf{h}_K} P(\mathbf{v}, \mathbf{h}_1, ..., \mathbf{h}_K, \mathbf{y}; \Theta_q), \quad (3)$$

where $\mathbf{h}_k \in \{0, 1\}^{|\mathbf{h}_k|}$ represents the $|\mathbf{h}_k|$ hidden nodes of layer $k \in \{1, .., K\}$ of the deep belief network, $\mathbf{v}$ is a vector representation of the input image I, and Θ_q denotes the DBN parameters (weights and biases). The probability term in (3) is computed as

$$\begin{aligned} P(\mathbf{v}, \mathbf{h}_1, ..., \mathbf{h}_K, \mathbf{y}) =& P(\mathbf{h}_K, \mathbf{h}_{K-1}, \mathbf{y}) \\ & \left(\prod_{k=1}^{K-2} P(\mathbf{h}_{k+1}|\mathbf{h}_k)\right) P(\mathbf{h}_1|\mathbf{v}), \end{aligned} \quad (4)$$

where $-\log P(\mathbf{h}_K, \mathbf{h}_{K-1}, \mathbf{y}) \propto \mathcal{E}_{\text{RBM}}(\mathbf{h}_K, \mathbf{h}_{K-1}, \mathbf{y})$ with

$$\begin{aligned} \mathcal{E}_{\text{RBM}}(\mathbf{h}_K, \mathbf{h}_{K-1}, \mathbf{y}) = & -\mathbf{b}_K^\top\mathbf{h}_K - \mathbf{a}_{K-1}^\top\mathbf{h}_{K-1} - \mathbf{a}_y^\top\mathbf{y} - \\ & (\mathbf{h}_K)^\top\mathbf{W}_K\mathbf{h}_{K-1} - (\mathbf{h}_K)^\top\mathbf{W}_y\mathbf{y} \end{aligned} \quad (5)$$

representing the energy function of a restricted Boltzmann machine (RBM) [5], where $\mathbf{b}_K, \mathbf{a}_{K-1}, \mathbf{a}_y$ denote the bias vectors and $\mathbf{W}_K, \mathbf{W}_y$ are the weight matrices. Also in (4), we have

$$P(\mathbf{h}_{k+1}|\mathbf{h}_k) = \prod_j P(\mathbf{h}_{k+1}(j) = 1|\mathbf{h}_k), \quad (6)$$

with $P(\mathbf{h}_{k+1}(j) = 1|\mathbf{h}_k) = \sigma(\mathbf{b}_{k+1}(j) + \mathbf{h}_k^\top\mathbf{W}_{k+1}(:, j))$, $P(\mathbf{h}_1(j) = 1|\mathbf{v}) = \sigma(\mathbf{b}_1(j) + \frac{\mathbf{v}^\top\mathbf{W}_1(:,j)}{\sigma^2})$ [1], where $\sigma(x) = \frac{1}{1+e^{-x}}$, the operator (j) returns the j^{th} vector value, and $(:, j)$ returns the j^{th} matrix column.

The DBN in (3) is trained with a dataset containing training image I and respective segmentation map $\mathbf{y}$, as shown in Fig. 2-(b). The training process is based on the unsupervised bottom-up training of each pair of layers, where the weights

[1] That is, we assume zero-mean Gaussian visible units for the DBN.

and biases of the network are learned to build an auto-encoder for the values at the bottom layer, and a top RBM is trained with the segmentation map $\mathbf{y}$ [5]. The structured inference process consists of taking the input image and performing bottom-up inferences, until reaching the top two layers, which form an RBM, and then initialize the layer $\mathbf{y} = \mathbf{0}$ and perform Gibbs sampling on the layers $\mathbf{h}_K$ and $\mathbf{h}_{K-1}$, $\mathbf{y}$ until convergence [5].

The combination of DRLS and DBN is explained in the Alg. 1, where essentially, we iteratively run the DRLS method until convergence using the segmentation result from the DBN as one of the optimization terms.

Algorithm 1 Combined Level Set and DBN Segmentation

- INPUT: test image I and initial segmentation ϕ_0
- Infer $\mathbf{y}^*_{\text{DBN},q}$ from I using (3) for $q \in \{\text{left lung,right lung}\}$
- Compute distance function $\phi_{\text{DBN},q}$ from map $\mathbf{y}^*_{\text{DBN},q}$ (Fig. 2-(a))

for t = 1:T **do**

- Run DRLS using $\phi_{t-1}, \phi_{\text{DBN},q}$ to produce ϕ_t

end for

- Segmentation is the zero level set $\mathcal{C} = \{\mathbf{x} \in \Omega | \phi_T(\mathbf{x}) = 0\}$

3. EXPERIMENTS

The evaluation of the accuracy of our methodology uses the publicly available Japanese Society of Radiological Technology (JSRT) dataset [10], which contains manual segmentations of lung fields, heart and clavicles [12]. The JSRT database contains 247 chest radiographs, where 154 contain lung nodules (100 malignant, 54 benign) and 93 have no nodules, and each sample is represented by 12-bit gray scale image with size 2048×2048 pixels and $0.175mm$ pixel resolution. This database is randomly split into three sets: training (84 images), validation (40 images) , and test (123 images), and the assessment is based on following three measures: Jaccard Similarity Coefficient (Ω), Dice's Coefficient (DSC), and Average Contour Distance (ACD) [1]. We use the training set for the estimation of the DBN and DRLS parameters and the validation set for the DBN model selection (e.g. select the number of layers and number of nodes per layer in the network). The model selection estimated the following configuration for DBN: each hidden layer has 1000 nodes, with the input and segmentation layers with 1600 nodes. The initial guess ϕ_0 in Alg. 1 used by our approach is not automatically produced, so we show how the performance of our approach is affected by initial guesses of different accuracies, which are generated by random perturbations from the manual annotation. We denote the different initial guesses by the index $k \in \{1, 2, 3\}$, where $k = 1$ indicates the highest precision and $k = 3$ means the lowest precision initial guess. The estimation of the level set parameters is performed separately for each type of initial guess, and we achieve the following result: $\mu = 0.12, \lambda = 2, \alpha = -3, \gamma = 0.0005$ for $k = 1$; $\mu = 0.12, \lambda = 2, \alpha = -10, \gamma = 0.003$ for $k = 2$; and $\mu = 0.12, \lambda = 2, \alpha = -15, \gamma = 0.007$ for $k = 3$.

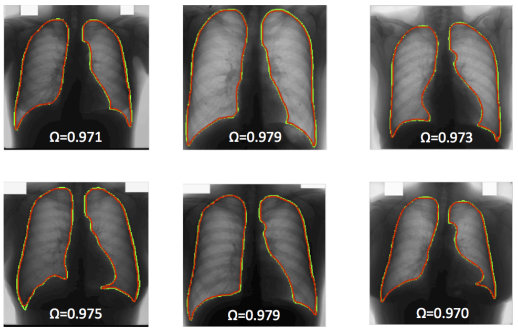


Fig. 3. Lung segmentation results with initial guess $k = 2$. The green contour shows expert annotation and the red illustrates the final result.

3.1. Results

Table 1 shows the results of our proposed methodology for lung segmentation with the different types of initial guesses. In this table, we also show the results when $\gamma = 0$, which is denoted by "Model without DBN" (this shows the influence of the DBN in the proposed methodology); and we also show the results for the initial guess, represented by "Initial guess only". Table 2 compares our results with the ones produced by the current state of the art on the JSRT database. Finally, Fig. 3.1 shows a few lung segmentation results using initial guess $k = 2$ on images of the test set. Using a standard computer (Intel(R) Core(TM) i5-2500k 3.30GHz CPU with 8GB RAM), and processing an input image with size 256×256 pixels, our method runs on average in 20.68 sec/image, which is comparable to the result by Candemir et al. [1], who report a running time of between 20 and 25 sec/image using the same input resolution and similar computer configuration.

4. DISCUSSIONS AND CONCLUSIONS

The results show that our proposed method using the initial guesses $k \in \{1, 2\}$ produces the best results in the field and comparable running times to the current state of the art. The main step that is missing from our approach is an automated initial guess, and we plan to address this issue by using the initial guess proposed by Candemir et al. [1]. We also plan to extend this method to other lung segmentation databases [1] and other segmentation problems (i.e., other anatomies) from different imaging techniques.

5. REFERENCES

[1] S. Candemir, et al. Lung segmentation in chest radiograps using anatomical atlases with non-rigid registration. 2014.

[2] F. Carrascal, et al. Automatic calculation of total lung capacity from automatically traced lung boundaries in

Table 1. Quantitative experiments on the JSRT database [10] showing the performance of the proposed method as a function of the initial guess used, where each cell is formatted as "mean (standard deviation) [min value - max value]".

Initial guess	Method	Ω	DSC	ACD
$k = 1$	Proposed model	$0.985(0.003)[0.972 - 0.991]$	$0.992(0.002)[0.986 - 0.996]$	$1.075(0.065)[0.825 - 1.267]$
	Model without DBN	$0.984(0.003)[0.969 - 0.990]$	$0.992(0.002)[0.984 - 0.995]$	$1.376(0.221)[1.234 - 6.184]$
	Initial guess only	$0.955(0.006)[0.919 - 0.968]$	$0.977(0.003)[0.958 - 0.984]$	$1.392(0.006)[1.372 - 1.404]$
$k = 2$	Proposed model	$0.973(0.007)[0.944 - 0.985]$	$0.986(0.004)[0.971 - 0.993]$	$1.120(0.165)[0.628 - 1.916]$
	Model without DBN	$0.946(0.007)[0.910 - 0.961]$	$0.972(0.004)[0.953 - 0.980]$	$2.408(0.232)[0.021 - 7.232]$
	Initial guess only	$0.912(0.013)[0.844 - 0.935]$	$0.954(0.007)[0.916 - 0.967]$	$2.519(0.041)[2.369 - 2.621]$
$k = 3$	Proposed model	$0.948(0.012)[0.893 - 0.970]$	$0.973(0.006)[0.943 - 0.985]$	$1.852(0.286)[1.120 - 3.708]$
	Model without DBN	$0.866(0.018)[0.790 - 0.900]$	$0.928(0.010)[0.883 - 0.947]$	$4.695(0.276)[3.792 - 9.112]$
	Initial guess only	$0.828(0.024)[0.712 - 0.873]$	$0.906(0.014)[0.832 - 0.932]$	$4.936(0.105)[4.391 - 5.200]$

Table 2. Quantitative experiments on the JSRT database [10] comparing our results with the state of the art on the same database, sorted from best (top) to worst (bottom). The symbol '?' indicates that the result is not available.

Method	Ω	DSC	ACD
Proposed model, $k = 1$	$0.985(0.003)[0.972 - 0.991]$	$0.992(0.002)[0.986 - 0.996]$	$1.075(0.065)[0.825 - 1.267]$
Proposed model, $k = 2$	$0.973(0.007)[0.944 - 0.985]$	$0.986(0.004)[0.971 - 0.993]$	$1.120(0.165)[0.628 - 1.916]$
Candemir, Hybrid [1]	$0.954(0.015)[?-?]$	$0.967(0.008)[?-?]$	$1.321(0.316)[?-?]$
Ginneken, Hybrid voting [12]	$0.949(0.020)[0.818 - 0.978]$	$?(?)[?-?]$	$1.62(0.66)[0.95 - 7.72]$
Proposed model, $k = 3$	$0.948(0.012)[0.893 - 0.970]$	$0.973(0.006)[0.943 - 0.985]$	$1.852(0.286)[1.120 - 3.708]$
Ginneken, PC post-processed [12]	$0.945(0.022)[0.823 - 0.972]$	$?(?)[?-?]$	$1.61(0.80)[0.83 - 8.34]$
Dawoud [4]	$0.940(0.053)[?-?]$	$?(?)[?-?]$	$2.46(2.06)[?-?]$
Ginneken, PC [12]	$0.938(0.027)[0.823 - 0.968]$	$?(?)[?-?]$	$3.25(2.65)[0.93 - 15.59]$
Ginneken, Hybrid ASM/PC [12]	$0.934(0.037)[0.706 - 0.968]$	$?(?)[?-?]$	$2.08(1.40)[0.91 - 11.57]$
Seghers [9]	$0.930(?)[?-?]$	$?(?)[?-?]$	$?(?)[?-?]$
Ginneken, AAM whiskers BFGS [12]	$0.922(0.029)[0.718 - 0.961]$	$?(?))[?-?]$	$2.39(1.07)[1.15 - 12.09]$
Yu [14]	$0.907(0.033)[?-?]$	$?(?)[?-?]$	$?(?)[?-?]$
Ginneken, Mean shape [12]	$0.713(0.075)[0.460 - 0.891]$	$?(?)[?-?]$	$10.06(3.18)[3.50 - 23.77]$

postero-anterior and lateral digital chest radiographs. *Medical physics*, 25(7):1118–1131, 1998.

[3] Charles L Daley, M Gotway, and R Jasmer. Radiographic manifestation of tuberculosis. *A primer for clinicians*, 1, 2003.

[4] A Dawoud. Lung segmentation in chest radiographs by fusing shape information in iterative thresholding. *IET Computer Vision*, 5(3):185–190, 2011.

[5] G. Hinton and R. Salakhutdinov. Reducing the dimensionality of data with neural networks. *Science*, 313(5786):504–507, 2006.

[6] S. Jaeger, et al. Automatic tuberculosis screening using chest radiographs. 2013.

[7] C. Li, et al. Distance regularized level set evolution and its application to image segmentation. *IEEE TIP*, 19(12):3243–3254, 2010.

[8] T. Ngo and G. Carneiro. Fully automated non-rigid segmentation with distance regularized level set evolution initialized and cosntrained by deep-structured inference. In *CVPR*. 2014.

[9] D. Seghers, et al. Minimal shape and intensity cost path segmentation. *IEEE TMI*, 26(8):1115–1129, 2007.

[10] J. Shiraishi, et al. Development of a digital image database for chest radiographs with and without a lung nodule: receiver operating characteristic analysis of radiologists' detection of pulmonary nodules. *American Journal of Roentgenology*, 174(1):71–74, 2000.

[11] B. Ginneken et al. Active shape model segmentation with optimal features. *IEEE TMI*, 21(8):924–933, 2002.

[12] B. Ginneken, M. Stegmann, and M. Loog. Segmentation of anatomical structures in chest radiographs using supervised methods: a comparative study on a public database. *MIA*, 10(1):19–40, 2006.

[13] B. Ginneken, Bart M ter Haar Romeny, and Max A Viergever. Computer-aided diagnosis in chest radiography: a survey. *IEEE TMI*, 20(12):1228–1241, 2001.

[14] T. Yu, J. Luo, and N. Ahuja. Shape regularized active contour using iterative global search and local optimization. In *CVPR*, 2005.

Chapter 8

Conclusion and Future Works

The segmentation problems present in medical image analysis applications represent an important and difficult challenge because of their strict accuracy requirements and the lack of large annotated datasets. The methodology proposed in this thesis addresses the accuracy aspect of these segmentation problems without requiring large annotated datasets, and we believe that the results produced by our system show that it is able to achieve the current state-of-the-art results in quite different medical image analysis applications. We believe that this represents a strong evidence that supports our accuracy claims. In this final chapter we discuss the main contributions of our work, its limitations and directions for the future.

8.1 Summary of Contributions

The methodology proposed in this thesis, based on the combination of level set methods and deep learning techniques, is the first approach in the field to successfully combine these two approaches. This hybrid model has been tested on two datasets containing images of different organs (heart and lungs) from different modalities (MRI and X-Ray), and we show that we can obtain state-of-the-art results on both problems. The main contributions of this thesis are the following:

1. In Chapter 4, we introduce the use of a DBN for the structured output prediction that represents the endocardium segmentation from MRI, which is a novelty to the best of our knowledge. In addition, we also introduce the combination of level set and deep learning, which is also another novelty. The experiments are based on a semi-automatic system that uses manually defined regions of interest (ROI) represented by a bounding box around the

endocardium. Results show that we obtain state-of-the-art segmentation results in the public dataset MICCAI 2009 [42] used for assessing competing segmentation methodologies for the problem of endocardium segmentation;

2. The introduction of an automatic initialisation of the level set method has been proposed in Chapter 5, where we use a DBN as a structured output model for the ROI detection, which is also another novelty in the computer vision and medical image analysis fields. This ROI detection allowed us to proposed a fully automatic segmentation system that can produce the current best results in the MICCAI 2009 challenge dataset;

3. In Chapter 6, we propose a fully automatic methodology capable of localising the left ventricle ROI, then the endocardium border and finally the epicardium border. The results on the MICCAI 2009 challenge dataset show that our methodology is the best methodology in the field for the semi-automatic problem and on par with the best methodology in the field for the fully automatic segmentation problem. Note that in the papers presented in Chapters 4-6, we show that the DBN terms provide significant improvements in the segmentation accuracy (see Table 1 in Chapter 4, and Table 1 in Chapter 5); and

4. The methodology is also tested in a new problem: lung segmentation from chest radiographies. The experimental segmentation results show that the DBN term again improves the final results of level set evolution (see Table 1 in Chapter 7), which can be considered to be the current state of the art in the field.

8.2 Future Work

Our method successfully shows how to use a deep learning structured output model to initialise and constrain the evolution of the level set methods. Nevertheless, we believe that there are some points that could be improved in our methodology, as follows:

1. Besides using intensity and shape constraint, we will incorporate motion information into the segmentation process of the left ventricle of the heart: this can be done with the insertion of a motion model in our methodology;

2. We will introduce the DBN method in the level set formulation in such way that it can be updated during the level set evolution (i.e., re-estimate the model parameters during inference);

3. We will use convolutional neural networks [93] for the detection and segmentation tasks, which have been shown to produce more accurate results, compared with DBN models [94];

4. We will implement a fully automated version for the lung segmentation task, using the same steps as in the left ventricle segmentation task; and

5. We will apply our model to other problems in medical image analysis.

Bibliography

[1] Prasanna K Sahoo, SAKC Soltani, and Andrew KC Wong, "A survey of thresholding techniques," *Computer vision, graphics, and image processing*, vol. 41, no. 2, pp. 233–260, 1988.

[2] Nobuyuki Otsu, "A threshold selection method from gray-level histograms," *Automatica*, vol. 11, no. 285-296, pp. 23–27, 1975.

[3] Michael Kass, Andrew Witkin, and Demetri Terzopoulos, "Snakes: Active contour models," *International journal of computer vision*, vol. 1, no. 4, pp. 321–331, 1988.

[4] Tim McInerney and Demetri Terzopoulos, "Deformable models in medical image analysis," in *Mathematical Methods in Biomedical Image Analysis, 1996., Proceedings of the Workshop on*. IEEE, 1996, pp. 171–180.

[5] Laurent D Cohen, "On active contour models and balloons," *CVGIP: Image understanding*, vol. 53, no. 2, pp. 211–218, 1991.

[6] Laurent D Cohen and Isaac Cohen, "Finite-element methods for active contour models and balloons for 2-d and 3-d images," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 15, no. 11, pp. 1131–1147, 1993.

[7] Chenyang Xu and Jerry L Prince, "Snakes, shapes, and gradient vector flow," *Image Processing, IEEE Transactions on*, vol. 7, no. 3, pp. 359–369, 1998.

[8] 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.

[9] Daniel Cremers, Mikael Rousson, and Rachid Deriche, "A review of statistical approaches to level set segmentation: integrating color, texture, motion and shape," *International journal of computer vision*, vol. 72, no. 2, pp. 195–215, 2007.

[10] Timothy F Cootes, Christopher J Taylor, David H Cooper, and Jim Graham, "Active shape models-their training and application," *Computer vision and image understanding*, vol. 61, no. 1, pp. 38–59, 1995.

[11] Timothy F Cootes, Gareth J Edwards, and Christopher J Taylor, "Active appearance models," in *Computer VisionECCV98*, pp. 484–498. Springer, 1998.

[12] Timothy F Cootes, Gareth J Edwards, and Christopher J Taylor, "Active appearance models," *IEEE Transactions on Pattern Analysis & Machine Intelligence*, , no. 6, pp. 681–685, 2001.

[13] Timothy F Cootes, Cristopher J Taylor, et al., "Statistical models of appearance for computer vision," 2004.

[14] J Alison Noble, Nassir Navab, and H Becher, "Ultrasonic image analysis and image-guided interventions," *Interface focus*, vol. 1, no. 4, pp. 673–685, 2011.

[15] Yoav Freund and Robert E Schapire, "A desicion-theoretic generalization of on-line learning and an application to boosting," in *Computational learning theory*. Springer, 1995, pp. 23–37.

[16] Gustavo Carneiro, Bogdan Georgescu, Sara Good, and Dorin Comaniciu, "Automatic fetal measurements in ultrasound using constrained probabilistic boosting tree," in *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2007*, pp. 571–579. Springer, 2007.

[17] Christopher JC Burges, "A tutorial on support vector machines for pattern recognition," *Data mining and knowledge discovery*, vol. 2, no. 2, pp. 121–167, 1998.

[18] Constantine Kotropoulos and Ioannis Pitas, "Segmentation of ultrasonic images using support vector machines," *Pattern Recognition Letters*, vol. 24, no. 4, pp. 715–727, 2003.

[19] Zhiwen Yu, Hau-San Wong, and Guihua Wen, "A modified support vector machine and its application to image segmentation," *Image and Vision Computing*, vol. 29, no. 1, pp. 29–40, 2011.

[20] Stefan Bauer, Lutz-P Nolte, and Mauricio Reyes, "Fully automatic segmentation of brain tumor images using support vector machine classification in combination with hierarchical conditional random field regularization," in *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2011*, pp. 354–361. Springer, 2011.

[21] Gustavo Carneiro, Jacinto C Nascimento, and António Freitas, "The segmentation of the left ventricle of the heart from ultrasound data using deep learning architectures and derivative-based search methods," *Image Processing, IEEE Transactions on*, vol. 21, no. 3, pp. 968–982, 2012.

[22] Karsten Held, E Rota Kops, Bernd J Krause, WMIII Wells, Ron Kikinis, and H-W Muller-Gartner, "Markov random field segmentation of brain mr images," *Medical Imaging, IEEE Transactions on*, vol. 16, no. 6, pp. 878–886, 1997.

[23] Ting Chen and Dimitris Metaxas, "Image segmentation based on the integration of markov random fields and deformable models," in *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2000*. Springer, 2000, pp. 256–265.

[24] Rui Huang, Vladimir Pavlovic, and Dimitris N Metaxas, "A graphical model framework for coupling mrfs and deformable models," in *Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on*. IEEE, 2004, vol. 2, pp. II–739.

[25] Yongyue Zhang, Michael Brady, and Stephen Smith, "Segmentation of brain mr images through a hidden markov random field model and the expectation-maximization algorithm," *Medical Imaging, IEEE Transactions on*, vol. 20, no. 1, pp. 45–57, 2001.

[26] Huawu Deng and David A Clausi, "Unsupervised image segmentation using a simple mrf model with a new implementation scheme," *Pattern recognition*, vol. 37, no. 12, pp. 2323–2335, 2004.

[27] Zoltan Kato and Josiane Zerubia, *Markov random fields in image segmentation*, Now, 2012.

[28] John Lafferty, Andrew McCallum, and Fernando CN Pereira, "Conditional random fields: Probabilistic models for segmenting and labeling sequence data," 2001.

[29] Gabriel Tsechpenakis and Dimitris N Metaxas, "Crf-driven implicit deformable model," in *Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on*. IEEE, 2007, pp. 1–8.

[30] Dana Cobzas and Mark Schmidt, "Increased discrimination in level set methods with embedded conditional random fields," in *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*. IEEE, 2009, pp. 328–335.

[31] Sanjiv Kumar and Martial Hebert, "Discriminative fields for modeling spatial dependencies in natural images," NIPS, 2003.

[32] Chi-Hoon Lee, Mark Schmidt, Albert Murtha, Aalo Bistritz, Jöerg Sander, and Russell Greiner, "Segmenting brain tumors with conditional random fields and support vector machines," in *Computer vision for biomedical image applications*, pp. 469–478. Springer, 2005.

[33] Charles Sutton and Andrew McCallum, "An introduction to conditional random fields for relational learning," *Introduction to statistical relational learning*, pp. 93–128, 2006.

[34] Mariano Cabezas, Arnau Oliver, Xavier Lladó, Jordi Freixenet, and Meritxell Bach Cuadra, "A review of atlas-based segmentation for magnetic resonance brain images," *Computer methods and programs in biomedicine*, vol. 104, no. 3, pp. e158–e177, 2011.

[35] Torsten Rohlfing, Robert Brandt, Randolf Menzel, Daniel B Russakoff, and Calvin R Maurer Jr, "Quo vadis, atlas-based segmentation?," in *Handbook of Biomedical Image Analysis*, pp. 435–486. Springer, 2005.

[36] Sema Candemir, Stefan Jaeger, Kannappan Palaniappan, Jonathan P Musco, Rahul K Singh, Zhiyun Xue, Alexandros Karargyris, Sameer Antani, George Thoma, and Clement J McDonald, "Lung segmentation in chest radiographs using anatomical atlases with nonrigid registration," *Medical Imaging, IEEE Transactions on*, vol. 33, no. 2, pp. 577–590, 2014.

[37] Xiao Han, Mischa S Hoogeman, Peter C Levendag, Lyndon S Hibbard, David N Teguh, Peter Voet, Andrew C Cowen, and Theresa K Wolf, "Atlas-based auto-segmentation of head and neck ct images," in *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2008*, pp. 434–441. Springer, 2008.

[38] Li Zhang, Eric Hoffman, Joseph M Reinhardt, et al., "Atlas-driven lung lobe segmentation in volumetric x-ray ct images," *Medical Imaging, IEEE Transactions on*, vol. 25, no. 1, pp. 1–16, 2006.

[39] Xuqing Wu and Shishir K Shah, "Level set with embedded conditional random fields and shape priors for segmentation of overlapping objects," in *Computer Vision–ACCV 2010*, pp. 230–241. Springer, 2011.

[40] 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.

[41] Geoffrey E Hinton and Ruslan R Salakhutdinov, "Reducing the dimensionality of data with neural networks," *Science*, vol. 313, no. 5786, pp. 504–507, 2006.

[42] P Radau, Y Lu, K Connelly, G Paul, A Dick, and G Wright, "Evaluation framework for algorithms segmenting short axis cardiac mri," *MIDAS J.Cardiac MR Left Ventricle Segmentation Challenge*, 2009.

[43] Junji Shiraishi, Shigehiko Katsuragawa, Junpei Ikezoe, Tsuneo Matsumoto, Takeshi Kobayashi, Ken-ichi Komatsu, Mitate Matsui, Hiroshi Fujita, Yoshie Kodera, and Kunio Doi, "Development of a digital image database for chest radiographs with and without a lung nodule: receiver operating characteristic analysis of radiologists' detection of pulmonary nodules," *American Journal of Roentgenology*, vol. 174, no. 1, pp. 71–74, 2000.

[44] Sharon W Kirschbaum and Robert-Jan van Geuns, "Cardiac magnetic resonance imaging to detect and evaluate ischemic heart disease.," *Hellenic journal of cardiology: HJC= Hellenike kardiologike epitheorese*, vol. 50, no. 2, pp. 119–126, 2008.

[45] Caroline Petitjean and Jean-Nicolas Dacher, "A review of segmentation methods in short axis cardiac mr images," *Medical Image Analysis*, vol. 15, no. 2, pp. 169–184, 2011.

[46] Francisco M Carrascal, José M Carreira, Miguel Souto, Pablo G Tahoces, Lorenzo Gómez, and Juan J Vidal, "Automatic calculation of total lung capacity from automatically traced lung boundaries in postero-anterior and lateral digital chest radiographs," *Medical Physics*, vol. 25, no. 7, pp. 1118–1131, 1998.

[47] Stefan Jaeger, Alexandros Karargyris, Sema Candemir, Les Folio, Jenifer Siegelman, Fiona Callaghan, Zhiyun Xue, Kannappan Palaniappan, Rahul K Singh, Sameer Antani, et al., "Automatic tuberculosis screening using chest radiographs," *Medical Imaging, IEEE Transactions on*, vol. 33, no. 2, pp. 233–245, 2014.

[48] Charles L Daley, M Gotway, and R Jasmer, "Radiographic manifestation of tuberculosis," *A primer for clinicians*, vol. 1, 2003.

[49] Anil K Jain, Yu Zhong, and Marie-Pierre Dubuisson-Jolly, "Deformable template models: A review," *Signal processing*, vol. 71, no. 2, pp. 109–129, 1998.

[50] Michael E Leventon, W Eric L Grimson, and Olivier Faugeras, "Statistical shape influence in geodesic active contours," in *Computer Vision and Pattern Recognition, 2000. Proceedings. IEEE Conference on*. IEEE, 2000, vol. 1, pp. 316–323.

[51] Nikos Paragios, "A level set approach for shape-driven segmentation and tracking of the left ventricle," *Medical Imaging, IEEE Transactions on*, vol. 22, no. 6, pp. 773–776, 2003.

[52] Daniel Cremers, "Dynamical statistical shape priors for level set-based tracking," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 28, no. 8, pp. 1262–1273, 2006.

[53] Tony F Chan and Luminita A Vese, "Active contours without edges," *Image processing, IEEE transactions on*, vol. 10, no. 2, pp. 266–277, 2001.

[54] Gilles Aubert and Pierre Kornprobst, *Mathematical problems in image processing: partial differential equations and the calculus of variations*, vol. 147, Springer Science & Business Media, 2006.

[55] Hayden Schaeffer, "Active arcs and contours," *UCLA CAM Report*, pp. 12–54, 2012.

[56] Giuseppe Coppini, Stefano Diciotti, Massimo Falchini, Natale Villari, and Guido Valli, "Neural networks for computer-aided diagnosis: detection of lung nodules in chest radiograms," *Information Technology in Biomedicine, IEEE Transactions on*, vol. 7, no. 4, pp. 344–357, 2003.

[57] Lawrence O Hall, Amine M Bensaid, Laurence P Clarke, Robert P Velthuizen, Martin S Silbiger, and James C Bezdek, "A comparison of neural network and fuzzy clustering techniques in segmenting magnetic resonance images of the brain," *Neural Networks, IEEE Transactions on*, vol. 3, no. 5, pp. 672–682, 1992.

[58] Shan Shen, William Sandham, Malcolm Granat, and Annette Sterr, "Mri fuzzy segmentation of brain tissue using neighborhood attraction with neural-network optimization," *Information Technology in Biomedicine, IEEE Transactions on*, vol. 9, no. 3, pp. 459–467, 2005.

[59] Guo Dong and Ming Xie, "Color clustering and learning for image segmentation based on neural networks," *Neural Networks, IEEE Transactions on*, vol. 16, no. 4, pp. 925–936, 2005.

[60] Zhuowen Tu, "Probabilistic boosting-tree: Learning discriminative models for classification, recognition, and clustering," in *Computer Vision, 2005. ICCV 2005. Tenth IEEE International Conference on*. IEEE, 2005, vol. 2, pp. 1589–1596.

[61] Gustavo Carneiro, Bogdan Georgescu, Sara Good, and Dorin Comaniciu, "Detection and measurement of fetal anatomies from ultrasound images using a constrained probabilistic boosting tree," *Medical Imaging, IEEE Transactions on*, vol. 27, no. 9, pp. 1342–1355, 2008.

[62] Azhar Quddus, Paul Fieguth, and Otman Basir, "Adaboost and support vector machines for white matter lesion segmentation in mr images," in *Engineering in Medicine and Biology Society, 2005. IEEE-EMBS 2005. 27th Annual International Conference of the*. IEEE, 2005, pp. 463–466.

[63] Jerome Friedman, Trevor Hastie, Robert Tibshirani, et al., "Additive logistic regression: a statistical view of boosting (with discussion and a rejoinder by the authors)," *The annals of statistics*, vol. 28, no. 2, pp. 337–407, 2000.

[64] Robert E Schapire, "The boosting approach to machine learning: An overview," in *Nonlinear estimation and classification*, pp. 149–171. Springer, 2003.

[65] Stuart Geman and Donald Geman, "Stochastic relaxation, gibbs distributions, and the bayesian restoration of images," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, , no. 6, pp. 721–741, 1984.

[66] Julian Besag, "On the statistical analysis of dirty pictures," *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 259–302, 1986.

[67] DM Greig, BT Porteous, and Allan H Seheult, "Exact maximum a posteriori estimation for binary images," *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 271–279, 1989.

[68] Yuri Y Boykov and M-P Jolly, "Interactive graph cuts for optimal boundary & region segmentation of objects in nd images," in *Computer Vision, 2001. ICCV 2001. Proceedings. Eighth IEEE International Conference on*. IEEE, 2001, vol. 1, pp. 105–112.

[69] Yuri Boykov and Gareth Funka-Lea, "Graph cuts and efficient nd image segmentation," *International journal of computer vision*, vol. 70, no. 2, pp. 109–131, 2006.

[70] Peter Elias, Amiel Feinstein, and Claude E Shannon, "A note on the maximum flow through a network," *Information Theory, IRE Transactions on*, vol. 2, no. 4, pp. 117–119, 1956.

[71] Leslie R Foulds, *Graph theory applications*, Springer Science & Business Media, 2012.

[72] Lester Randolph Ford and Delbert R Fulkerson, *A simple algorithm for finding maximal network flows and an application to the Hitchcock problem*, Citeseer, 1955.

[73] Boris V Cherkassky and Andrew V Goldberg, "On implementing the pushrelabel method for the maximum flow problem," *Algorithmica*, vol. 19, no. 4, pp. 390–410, 1997.

[74] 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.

[75] Vladimir Kolmogorov and Ramin Zabin, "What energy functions can be minimized via graph cuts?," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 26, no. 2, pp. 147–159, 2004.

[76] Jyrki Lötjönen, Sari Kivistö, Juha Koikkalainen, Daniel Smutek, and Kirsi Lauerma, "Statistical shape model of atria, ventricles and epicardium from short-and long-axis mr images," *Medical image analysis*, vol. 8, no. 3, pp. 371–386, 2004.

[77] Bram Van Ginneken, Mikkel B Stegmann, and Marco Loog, "Segmentation of anatomical structures in chest radiographs using supervised methods: a comparative study on a public database," *Medical Image Analysis*, vol. 10, no. 1, pp. 19–40, 2006.

[78] Tobias Heimann and Hans-Peter Meinzer, "Statistical shape models for 3d medical image segmentation: a review," *Medical image analysis*, vol. 13, no. 4, pp. 543–563, 2009.

[79] Colin Goodall, "Procrustes methods in the statistical analysis of shape," *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 285–339, 1991.

[80] Fei Chen, Huimin Yu, Rose Hu, and Xunxun Zeng, "Deep learning shape priors for object segmentation," in *Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on*. IEEE, 2013, pp. 1870–1877.

[81] Ivana Išgum, Marius Staring, Annemarieke Rutten, Mathias Prokop, Max Viergever, Bram Van Ginneken, et al., "Multi-atlas-based segmentation with local decision fusionapplication to cardiac and aortic segmentation in ct scans," *Medical Imaging, IEEE Transactions on*, vol. 28, no. 7, pp. 1000–1010, 2009.

[82] Ron Kikinis, Martha E Shenton, Dan V. Iosifescu, Robert W McCarley, Pairash Saiviroonporn, Hiroto H Hokama, Andre Robatino, David Metcalf, Cynthia G Wible, Chiara M Portas, et al., "A digital brain atlas for surgical planning, model-driven segmentation, and teaching," *Visualization and Computer Graphics, IEEE Transactions on*, vol. 2, no. 3, pp. 232–241, 1996.

[83] Paul Viola and William M Wells III, "Alignment by maximization of mutual information," *International journal of computer vision*, vol. 24, no. 2, pp. 137–154, 1997.

[84] Frederik Maes, Andre Collignon, Dirk Vandermeulen, Guy Marchal, and Paul Suetens, "Multimodality image registration by maximization of mutual information," *Medical Imaging, IEEE Transactions on*, vol. 16, no. 2, pp. 187–198, 1997.

[85] Colin Studholme, Derek LG Hill, and David J Hawkes, "An overlap invariant entropy measure of 3d medical image alignment," *Pattern recognition*, vol. 32, no. 1, pp. 71–86, 1999.

[86] Baojun Li, Gary E Christensen, Eric A Hoffman, Geoffrey McLennan, and Joseph M Reinhardt, "Establishing a normative atlas of the human lung: intersubject warping and registration of volumetric ct images," *Academic Radiology*, vol. 10, no. 3, pp. 255–265, 2003.

[87] Maria Lorenzo-Valdes, Gerardo I Sanchez-Ortiz, Raad Mohiaddin, and Daniel Rueckert, "Atlas-based segmentation and tracking of 3d cardiac mr images using non-rigid registration," in *Medical Image Computing and Computer-Assisted InterventionMICCAI 2002*, pp. 642–650. Springer, 2002.

[88] Anil Rao, Gerardo I Sanchez-Ortiz, Raghavendra Chandrashekara, Maria Lorenzo-Valdés, Raad Mohiaddin, and Daniel Rueckert, "Construction of a cardiac motion atlas from mr using non-rigid registration," in *Functional Imaging and Modeling of the Heart*, pp. 141–150. Springer, 2003.

[89] Josef Kittler and Fuad M. Alkoot, "Sum versus vote fusion in multiple classifier systems," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 25, no. 1, pp. 110–115, 2003.

[90] Michael Lynch, Ovidiu Ghita, and Paul F Whelan, "Left-ventricle myocardium segmentation using a coupled level-set with a priori knowledge," *Computerized Medical Imaging and Graphics*, vol. 30, no. 4, pp. 255–262, 2006.

[91] Yoshua Bengio, "Learning deep architectures for ai," *Foundations and trends® in Machine Learning*, vol. 2, no. 1, pp. 1–127, 2009.

[92] Miguel A Carreira-Perpinan and Geoffrey E Hinton, "On contrastive divergence learning," in *Artificial Intelligence and Statistics*, 2005, vol. 2005, p. 17.

[93] 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.

[94] Neeraj Dhungel, Gustavo Carneiro, and Andrew P Bradley, "Deep learning and structured prediction for the segmentation of mass in mammograms," in *Medical Image Computing and Computer-Assisted Intervention–MICCAI*. 2015, Springer.