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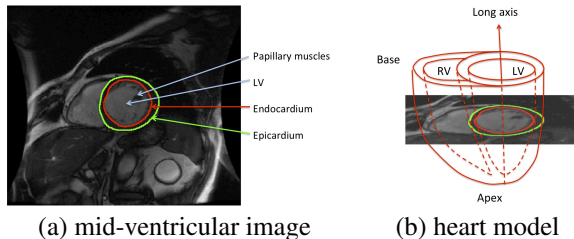
(a) mid-ventricular image   (b) heart model

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

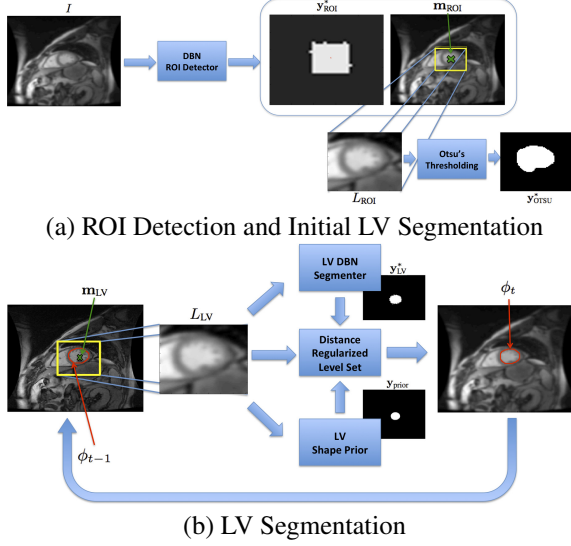(a) ROI Detection and Initial LV Segmentation

(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 $q \in \{ED,ES\}$, 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 $\Omega \to \{0, 1\}$
- 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 $I : \Omega \to \mathbb{R}$, with $\Omega \subseteq \mathbb{R}^2$ denoting the image coordinate space, the explicit contour representation of a segmentation is denoted by $\mathbf{c} : [0, 1] \to \Omega$, the implicit contour representation is formed with the zero level set of an Euclidean signed distance function $\phi : \Omega \to \mathbb{R}$, 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 $\mathcal{D} = \{(I, \mathbf{c}, i, q)_s\}_{i \in \{1,...,N\}, s \in \{1,...,S\}, q \in \{ED,ES\}}$, where $i \in \{1, ..., N\}$ is an index to an image within a sequence, $q \in \{ED,ES\}$ 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 $\mathbf{y} : \Omega \to \{0, 1\}$, 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 $L : \Omega \to \mathbb{R}$ 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:

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

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

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

$$P(\mathbf{h}_{k+1} | \mathbf{h}_k) = \prod_j P(\mathbf{h}_{k+1}(j) = 1 | \mathbf{h}_k),$$
(4)
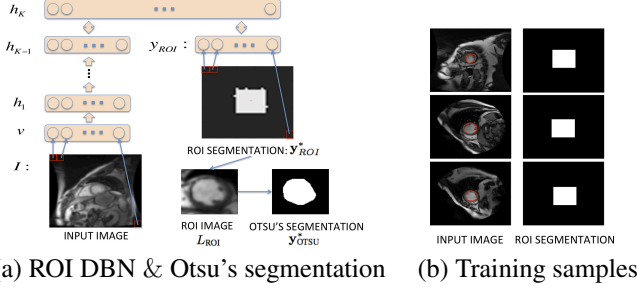
(a) ROI DBN & Otsu's segmentation (b) Training samples

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 jth vector value, and $(:,j)$ returns the jth 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

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

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]:

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

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 $\mathcal{S}(\phi, \phi_\kappa) = \int_\Omega (\phi - \phi_\kappa)^2 dx$ (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):

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

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
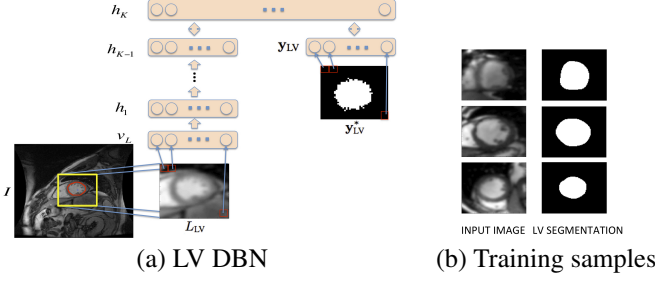
(a) LV DBN   (b) Training samples

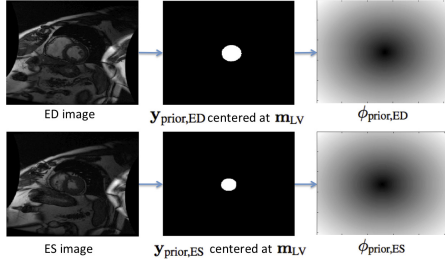Figure 4. Model for the LV DBN and training set samples.

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: $T = 10$, $M_{\mathrm{ROI}} = 100$, $M_{\mathrm{LV}} = 100$. The segmentation results are stable if these constants are within the ranges: $T \in [5, 20]$, $M_{\mathrm{ROI}} \in [80, 120]$, $M_{\mathrm{LV}} \in [80, 120]$.

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
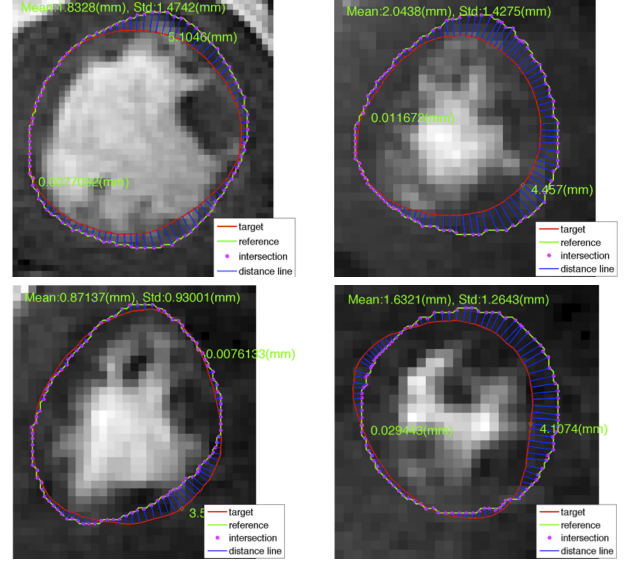

Figure 6. Segmentation results with challenging cases, such as images from apical and basal slice images and presenting papillary muscles and trabeculations. The red contour denotes the automated detection, and green shows the manual annotation.

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 approaches 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	$\mathbf{97.42(4.63)[83.33-100]}$	$0.88(0.04)[0.76-0.95]$	$2.14(0.43)[1.28-2.63]$
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