SEGMENTING THE LEFT VENTRICLE IN CARDIAC IN CARDIAC MRI: FROM HANDCRAFTED TO DEEP REGION BASED DESCRIPTORS

Daniela O. Medley, Carlos Santiago, and Jacinto C. Nascimento

Institute for Systems and Robotics, Instituto Superior Técnico, Lisbon, Portugal

ABSTRACT

Achieving robust object segmentations requires the ability to discard *outliers* (invalid observations) in the segmentation process. When considering an object model, such as an active shape model (ASM), described by a set of points, a *one-to-one* mapping from *one* model point to *one* valid observation would be ideal. However, in general, a *one-to-many* mapping is necessary to ensure the valid observation is detected and, thus, to have a reliable and robust fitting process. In this work, we compare three observation detectors: two of them based on handcrafted texture features and another based on a deep CNN classifier. Furthermore, to reduce from many detections to just one valid observation, we incorporate the Generalized Expectation-Maximization (GEM) algorithm in the ASM framework. This algorithm is able to neglect outliers by assigning them low weights. The proposed methodology exhibits remarkable accuracy with all the detectors in the context of the segmentation of the left ventricle in two publicly available MRI datasets, for which the proposed approach is competitive with other state-of-the-art methods.

1. INTRODUCTION

Cardiac magnetic resonance (CMR) is part of the gold standard procedure to assess cardiac function and morphology, namely by computing the ejection fraction [1]. This assessment requires cardiologists to manually delineate the left ventricle (LV) border, which is a slow and prone to poor repeatability procedure. Therefore, significant research has been focus on automatizing the segmentation process.

Shape models have become a widely used approach in this context [2]. The Active Shape Model (ASM), introduced by Cootes et al. [3], became a popular method for the segmentation of medical images due to its ability to model the organs contours through their mean shape and deformations patterns. This is particularly advantageous in the LV segmentation because of its roughly circular shape. The segmentation with an ASM framework has two main phases. In the first phase (training), the statistical information about the expected shape of the LV and its main modes of deformation is learned from the training set, creating a shape model. In the second phase (testing), the segmentation in a new (test) image takes place by iteratively fitting the learned model to observation points extracted from the test image.

Traditionally, the observations in the ASM framework are extracted by finding the strongest edge along profile lines for each model point. Although this approach may work in simple cases, it fails in more complex scenarios such as the segmentation of the LV. In this case, the strongest edge is not located along the LV border, which leads to the detection of noisy observations (outliers) [4]. Since the model is not able to distinguish between valid observations and outliers, the estimation of the model parameters is hampered, leading to poor segmentations.

In this paper, we solve this problem by: 1) considering two-dimensional search regions (instead of profile lines), which capture more information from the neighborhood of the model points; and 2) incorporating a robust statistical method for the estimation of the ASM parameters. For the former, we propose three new region-based detectors using either handcrafted texture features or using a deep CNN. Regarding the latter, the Generalized Expectation Maximization (GEM) [5] is used to reduce the effect of outliers that may still persist.

2. ACTIVE SHAPE MODEL

The ASM algorithm [3] represents the shape of an object by a set of N landmark points, $\mathbf{x} = [\mathbf{x}_i^\top]_{i=1,\dots,N}$, where $\mathbf{x}_i \in \mathbb{R}^{2\times 1}$ represents the coordinates of the *i*-th point. In order to characterize the shape of an object, ASM uses prior information of the shape statistics that comes from a the set of annotated images in the training set. Formally, a shape $\mathbf{x}$ can be represented in the shape space by a learned mean shape $\bar{\mathbf{x}} \in \mathbb{R}^{2N\times 1}$, deformed by a linear combination of the K main modes of deformation $\mathbf{D} \in \mathbb{R}^{2N\times K}$ associated to each landmark point,

$$\mathbf{x}_i = \bar{\mathbf{x}}_i + \mathbf{D}_i\mathbf{b} \ , \tag{1}$$

where $\mathbf{b} \in \mathbb{R}^{K\times 1}$ contains the deformation coefficients. The position of this shape $\mathbf{x}$ in the image is then given by a similarity transformation, $\mathbf{T}_\vartheta$, with parameters $\vartheta = \{\mathbf{a}, \mathbf{t}\}$. This transformation accounts for the translation $\mathbf{t} = (t_1, t_2)^T$, and scale s and rotation θ in $\mathbf{a} = (a_1, a_2)^T =$

This work was supported by FCT UID/EEA/50009/2019.

$(s\ cos(\theta), s\ sen(\theta))^T$, of the contour as

$$\begin{aligned} \tilde{\mathbf{x}}_i &= \mathbf{T}_\vartheta(\mathbf{x}_i) \\ &= \begin{bmatrix} a_1 & -a_2 \\ a_2 & a_1 \end{bmatrix} \begin{bmatrix} x_i^1 \\ x_i^2 \end{bmatrix} + \begin{bmatrix} t_1 \\ t_2 \end{bmatrix} \\ &= \mathbf{A}(\bar{\mathbf{x}}_i + \mathbf{D}_i\mathbf{b}) + \mathbf{t}\ . \end{aligned} \tag{2}$$

Let us assume we have an initial guess of the contour position, $\tilde{\mathbf{x}}$. To fit the model to a new image, the ASM iteratively searches for new observation points in the vicinity of each model point, $\tilde{\mathbf{x}}_i$, and estimates the set of parameters ($\mathbf{b}$, $\mathbf{a}$, $\mathbf{t}$) that minimize the distance between the model points and the new observations.

3. PROPOSED APPROACH

In this paper we propose a better ASM-based framework that provides more reliable observations, while also giving robust estimates of the model parameters that disregard the remaining outliers when fitting the model to a new image.

The proposed search method is a region-based (2D patches) approach whose goal is to provide reliable confidence maps of the correct location of each model point. Observations (candidate positions) are then extracted from these confidence maps by applying a threshold and non-maximum suppression. Section 3.1 describes three approaches on how these confidence maps are obtained. The probabilistic framework used to estimate the ASM parameters is described in Section 3.2.

3.1. Novel ASM Search Method

The proposed search method is based on the computation of confidence maps of the correct location of each model point. To obtain these maps, three different approaches are compared using: (i) *fixed texture templates* (FTT), (ii) *variable texture templates* (VTT) and (iii) *convolutional neural networks* (CNN) (see Fig.1). In (i) and (ii) the detection relies on handcrafted texture features, while in the CNN approach automatically learns the best features to locate the best observations. The following sections describe these three methodologies.

3.1.1. ASM Using Fixed Texture Templates

The first method consists of learning a texture template for each model point, which we denoted as ASM-FTT. This template should be able to capture the expected texture of the image around each model point. Formally, given a dataset $\mathcal{D}$ of annotated images, the template for the i-th model point, $\mathcal{T}^i$, of size $T \times T$ is given by

$$\mathcal{T}^i = \frac{1}{|\mathcal{D}|} \sum_{j=1}^{|\mathcal{D}|} \mathbf{P}_d^i, \quad i = 1, ..., N\ , \tag{3}$$

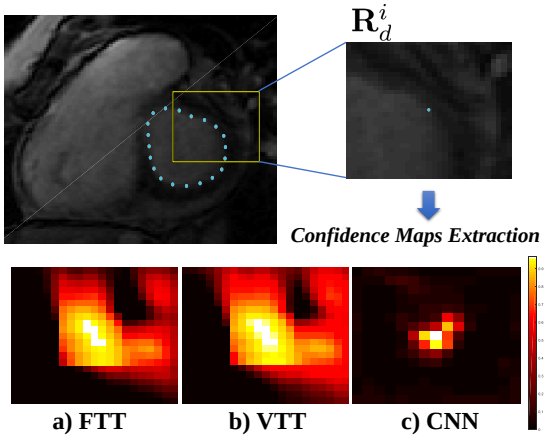

Fig. 1. Representation of the different confidence maps extraction from the test image. For each search region $\mathbf{R}_d^i$ (yellow rectangle) sampled around each current model point (blue) we compute confidence maps of the correct location by three different methods.

where $|\mathcal{D}|$ denotes the number of images in the dataset, and $\mathbf{P}_d^i$ is a patch (image crop) of the d-th image centered at $\tilde{\mathbf{x}}_i$ (see (2)).

The learned templates are then used to compute confidence maps by comparing them with a search region $\mathbf{R}_d^i$ of the test image, of size $R \times R$, with $R > T$. The comparison is performed by applying the normalized cross-correlation (see [6] for details) between the template $\mathcal{T}^i$ and each possible position in the search region $\mathbf{R}_d^i$ (see Fig. 1a).

3.1.2. ASM Using Variable Texture Templates

The second approach, denoted by ASM-VTT, uses a statistical template methodology instead of just using the average patch as described in Sec. 3.1.1. Formally, besides learning the template for each model point, $\mathcal{T}^i$, computed by (3), we also learn the main modes of variation of this template, $\mathbf{D}_g \in \mathbb{R}^{T^2 \times K}$. This matrix of variation in obtained through PCA and allows the template to adjust to each new image.

The templates are now given (in vectorized form) by the average texture plus a linear combination of its K main modes of variation as follows

$$\text{vec}\left[\mathcal{T}^i\right] \simeq \overline{\mathbf{g}}^i + \mathbf{D}_g^i \mathbf{b}_g^i, \quad i = 1, ..., N\ , \tag{4}$$

where $\overline{\mathbf{g}}^i \in \mathbb{R}^{T^2 \times 1}$ is the vectorized average texture, computed as in (3), and $\mathbf{b}_g^i \in \mathbb{R}^{k \times 1}$ is the vector containing the variation coefficients.

Given a test image, we first need to compute the corresponding coefficients, $\widehat{\mathbf{b}}_g^i$, for $i = 1, \ldots, N$, by fitting the template to the patch, $\mathbf{P}_d^i$, centered at the current model point position, leading to

$$\widehat{\mathbf{b}}_g^i = \arg\min_{\mathbf{b}} \left\| \overline{\mathbf{g}}^i + \mathbf{D}_g^i \mathbf{b} - \text{vec}[\mathbf{P}_d^i] \right\|^2 \text{ s.t. } \sum_{l=1}^{L} \frac{\widehat{b}_l^2}{\lambda_l} \leq d_{\max}^2 \tag{5}$$

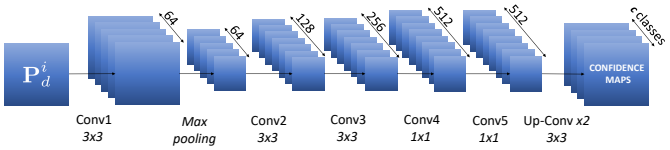

Fig. 2. CNN architecture used to compute the confidence maps of each patch $\mathbf{P}_d^i$.

where $\widehat{b}_l$ denotes the l-th component of $\mathbf{b}$, λ_l is the eigenvalue associated to the l-th variation mode, and $d_{\max}$ a pre-defined threshold. After updating the templates, the confidence maps are computed as previously, through the normalization cross-correlation (see Fig. 1b).

This method resembles the Constrained Local Model (CLC) [7] algorithm. However, here we use these templates to compute the confidence maps, instead of using them directly in the fitting strategy.

3.1.3. ASM Using Convolutional Neural Networks

The methods described previously are based on handcrafted texture features, which may not be the best approach. With the recent development of deep learning methods [8, 9], a greater interest has been given to algorithms capable of automatically learning the features that optimally represent the data, leading to more efficient feature extraction models than the traditional handcrafted approaches. Therefore, as the final method, we propose a different approach to compute the confidence maps: a simple fully convolutional network architecture that detects the location of each model point in an image.

Figure 2 illustrates the overall architecture of the network used to compute the confidence maps from each patch. This system takes as input a $T \times T$ patch and produces $T \times T \times (N+1)$ confidence map as output, representing the probability each pixel being the correct position of the model point corresponding to class c, where $c = 0, \ldots, N$. Note that each class here represents each one of the model points, and class $c = 0$ represents the background.

We train this network to output the set of $N+1$ confidence maps. Applying softmax to these maps should classify each pixel with $c = 0$, except for a small region around each model point, which are assigned the corresponding class. The maps with the pixel classes, denoted by $\mathbf{L}_d^i$, are fed to the network together with the corresponding input patches, $\mathbf{P}_d^i$, for $d = 1, \ldots, |\mathcal{D}|$ and the network weights are optimized using the cross-entropy loss.

During the test phase, given a new image, we feed each sub-region of size $T \times T$, $\mathbf{P}_d^i$, within the search region to the network and extract the corresponding confidence maps (after the softmax) for each model point, $i = 1, \ldots, N$ (see Fig. 1c). These maps will have local maxima in the most probable locations of the corresponding model points. This proposed method is denoted as ASM-CNN.

3.2. Model parameters estimation

Now the goal is to fit the learned model to the object. This is done by iteratively update the shape model parameters $\Theta = (\mathbf{a}, \mathbf{t}, \mathbf{b})$, given the new set of observations.

Classical ASM based approaches rely on a *one-to-one* mapping $\tilde{\mathbf{x}}^i \mapsto \mathbf{y}^i$, *i.e.*, each model point $\tilde{\mathbf{x}}^i$ is assigned to only one observation $\mathbf{y}^i$. However, the presence of outliers makes that a *one-to-many* mapping, *i.e.*, $\tilde{\mathbf{x}}^i \mapsto \mathbf{Y}^i = \{\mathbf{y}^{ij}\}_{j=1}^{M^i}$ is obtained, where $\mathbf{Y^i}$ stands for the set of M^i observations in the i-th search region $\mathbf{R}^i$. The estimation of the model parameters should not be affected by the presence of outliers. In order to do so, the estimation of the model parameters are accomplished by using an extension of the GEM-RASM (see [10]). The objective Q function is given as follows:

$$Q(\theta; \widehat{\theta}) = \sum_{i=1}^{N} \sum_{j=1}^{M^i} \omega_0^{ij} \log \ell_0 + \omega_1^{ij} \log \ell_1 \tag{6}$$

where $\ell_q = p(\mathbf{y}^{ij} | k^{ij} = q, \theta)$ is the log-likelihood, with $q \in \{0, 1\}$. In the E-step the weights of the observations are computed depending on the value of the binary label that is assigned. Thus, if the observation $\mathbf{y}^{ij}$ is valid (*i.e.* label $k^{ij} = 1$), the likelihood ℓ_1 is modeled as a normal distribution, as well as the corresponding weights, *i.e.*

$$w_1^{ij} \propto \mathcal{N}(\mathbf{y}^{ij};\, \tilde{\mathbf{x}}_i,\, \Sigma^{ij}) \tag{7}$$

Otherwise, if $\mathbf{y}^{ij}$ is invalid (*i.e.* label $k^{ij} = 0$), the likelihood ℓ_0 and its corresponding weights w_1^{ij} are modeled as a uniform distribution, *i.e.* $\mathcal{U}(\tilde{\mathbf{x}}_i)$ (see [10]). The effect of the strategy is to decrease the influence (weights) of the invalid observations simplifying the above mapping from the model to the observation space.

Now, given the most recent estimates of the unknown parameters Θ_t at iteration t, the model parameters can then be updated in the M-step. This can be achieved by minimizing a weighted least square between both the current model point and the corresponding observation, as

$$\widehat{\Theta}_{t+1} = \arg\min_{\Theta} \sum_{i=1}^{N} \sum_{j=1}^{M^i} \omega^{ij} [(\tilde{\mathbf{x}}_i - \mathbf{y}^{ij})^T \, (\Sigma^{ij})^{-1} \, (\tilde{\mathbf{x}}_i - \mathbf{y}^{ij})], \tag{8}$$

where Σ^{ij} is a covariance matrix of the point $\tilde{\mathbf{x}}_i$.

4. EXPERIMENTAL SETUP

The proposed framework was validated using two different publicly available datasets concerning cardiac MRI short axis volumes. The first dataset [11], contains data from 33 different patients. For each patient, the CMR data is a sequence of 20 volumes, with the corresponding manual LV segmentations, which is used as the ground truth. Among the different patients, only two are healthy, whereas the remaining ones display a variety of heart abnormalities. The second dataset

used was the MICCAI 2009 LV segmentation challenge [12] which comprises data from 45 patients also with a variety of different heart conditions, all coupled with expert segmentations.

For both datasets, each volume is segmented independently and the leave-one-sequence-out cross validation is used for each volume of each patient. The initialization of the first slice of each volume of the test sequence is performed using the ground truth. The remaining slices are initialized by successively propagating the previous slice segmentation.

The segmentation is quantitatively evaluated by comparing the estimated contour with the ground truth using two metrics: the Dice coefficient, d_{Dice} [13] and the average perpendicular distance, d_{AV}. The percentage of the good contours (GC), corresponding to the percentage of segmentations for which $d_{AV} < 5mm$, is also used for the performance evaluation.

5. RESULTS

Several search region sizes were tested, where the best overall results were obtained for $R = 15$ and $T = 7$ pixels. Figure 3 shows several examples of the segmentation obtained with these settings using with the different methods. Table 1 shows the average performance results obtained for both datasets. Comparison with other recent state of the art approaches is also presented. From the results it can be concluded that the proposed framework reaches the top performance using the CNN descriptors (see d_{Dice} and d_{AV} metrics). For the second dataset, we also see the superiority of the CNN detector. In this dataset, the methods [14, 15] also have a similar d_{Dice} metric. The results also show that the proposed methodology significantly outperforms the standard ASM [3] as well as the EM-ASM [10] that relies on the edge detection to find the observations. Overall, the proposed method achieves competitive results compared to the other state of the art methods.

6. CONCLUSION

In this paper we have proposed a new ASM framework that uses a more reliable observation detector, combined with a GEM statistical framework as a way to perform robust segmentations in the presence of outliers. The rationale behind this approach is to reduce the detection of outliers, and ensuring that the remaining outliers do not compromise the segmentation. This is accomplished by appropriately assigning a weight to each observation by switching the corresponding statistical model (normal vs. uniform distribution), which decreases the weights of the outliers. This framework was validated using three different detectors, namely: FTT, VTT and CNN. The results show that all the above detectors achieve competitive segmentation results, with the CNN achieving the best segmentation accuracy.

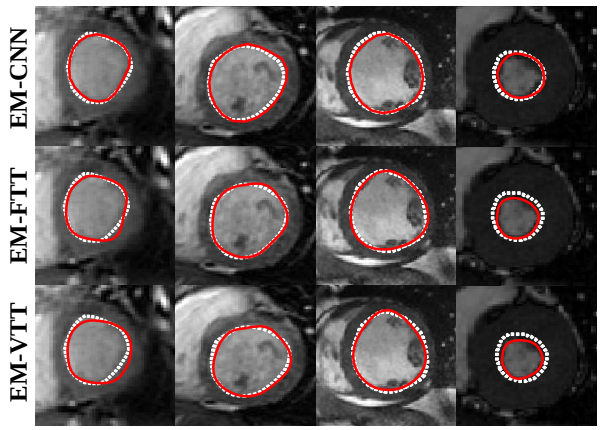

Fig. 3. Examples of LV segmentation with the different methods in four CMR images from different patients. The two first patients corresponds to dataset 1 [11], whereas the last two columns correspond to dataset 2 [12]. Each row shows the result of each method. The white dashed line shows the ground truth, whereas the red line represents the estimated segmentation.

Table 1. Comparison of the overall segmentation performance using the proposed algorithms with other approaches using both datasets (average value and standard deviation). Dashed entries in GC means all the contours were considered.

METHOD	d_{Dice}	d_{AV} (mm)	GC (%)
DATASET 1 [11]			
EM-CNN	**0.90 (0.02)**	**1.6 (0.2)**	**99.1**
EM-FTT	0.89 (0.03)	1.7 (0.3)	95.8
EM-VTT	0.88 (0.03)	1.9 (0.3)	95.3
Huang et al. [16]	0.89 (0.04)	2.2 (0.5)	82.5
Gopal et al. [17]	0.84 (0.04)	3.7 (0.6)	-
AAM [18]	0.84 (0.09)	2.6 (1.7)	-
EM-ASM [10]	0.82 (0.04)	2.9 (0.8)	-
ASM [3]	0.73 (0.10)	4.7 (2.1)	-
DATASET 2 [12]			
EM-CNN	**0.90 (0.04)**	2.0 (0.5)	87.3
EM-FTT	0.89 (0.04)	2.1 (0.5)	89.2
EM-VTT	0.89 (0.05)	2.1 (0.4)	87.7
Queiros et al. [14]	**0.90 (0.02)**	**1.8 (0.5)**	92.7
Ngo et al.[15]	**0.90 (0.03)**	2.1 (0.4)	**93.2**
Hu et al. [19]	0.89 (0.03)	2.2 (0.4)	91.1
Huang et al. [16]	0.89 (0.04)	2.2 (0.5)	79.2
Liu et al. [20]	0.88 (0.03)	2.4 (0.4)	91.2
Schaerer et al. [21]	0.87 (0.04)	3.0 (0.4)	-
AAM [22]	0.80 (0.05)	4.1 (1.1)	-
EM-ASM [10]	0.77 (0.11)	3.6 (0.5)	-
ASM [3]	0.65 (0.13)	3.7 (0.5)	-

7. REFERENCES

[1] W. Gregory Hundley, David A. Bluemke, J. Paul Finn, Scott D. Flamm, Mark A. Fogel, Matthias G. Friedrich, Vincent B. Ho, and Michael Jerosch-Herold, "ACCF/ACR/AHA/-NASCI/SCMR 2010 Expert Consensus Document on Cardiovascular Magnetic Resonance: A Report of the American College of Cardiology Foundation Task Force on Expert Consensus Documents," *J AM Coll Cardiol*, vol. 55, no. 23, pp. 2614–2662, 2010.

[2] 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.

[3] 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.

[4] Mike Rogers and Jim Graham, *Robust active shape model search*, Springer-Verlag, Berlin, Germany, 2006.

[5] A.P. Dempster, N.M. Laird, and Donald B Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *Journal of the Royal Statistical Society Series B Methodological*, vol. 39, no. 1, pp. 1–38, 1977.

[6] J .P. Lewis, "Fast Normalized Cross-Correlation," *Vision Interface*, vol. 1995, no. 1, pp. 1–7, 1995.

[7] D. Cristinacce and T. F. Cootes, "Feature Detection and Tracking with Constrained Local Models," *Procedings of the British Machine Vision Conference 2006*, pp. 95.1–95.10, 2006.

[8] Yann Lecun, Yoshua Bengio, and Geoffrey Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[9] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," *Advances In Neural Information Processing Systems*, pp. 1–9, 2012.

[10] Carlos Santiago, Jacinto Nascimento, and Jorge Marques, "2D Segmentation Using a Robust Active Shape Model With the EM Algorithm.," *IEEE transactions on image processing : a publication of the IEEE Signal Processing Society*, vol. 24, no. 8, pp. 2592–2601, 2015.

[11] Alexander Andreopoulos and John K. Tsotsos, "Efficient and generalizable statistical models of shape and appearance for analysis of cardiac MRI," *Medical Image Analysis*, vol. 12, no. 3, pp. 335–357, 2008.

[12] P. Radau, Lu Y., K. Connelly, G. Paul, Dick A. J., and G. A. Wright, "Evaluation Framework for Algorithms Segmenting Short Axis Cardiac MRI.," *The MIDAS Journal - Cardiac MR Left Ventricle Segmentation Challenge*, 2009.

[13] Lee R. Dice, "Measures of the Amount of Ecologic Association Between Species," *Journal of the Neurological Sciences*, vol. 26, no. 3, pp. 297–302, 1945.

[14] Sandro Queirós, Daniel Barbosa, Brecht Heyde, Pedro Morais, João L. Vilaça, Denis Friboulet, Olivier Bernard, and Jan D'hooge, "Fast automatic myocardial segmentation in 4D cine CMR datasets," *Medical Image Analysis*, vol. 18, no. 7, pp. 1115–1131, 2014.

[15] Tuan Anh Ngo and Gustavo Carneiro, "Left Ventricle Segmentation From Cardiac Mri Combining Level Set Methods With Deep Belief Networks," *International Conference on Image Processing(ICIP)*.

[16] Su Huang, Jimin Liu, Looi Chow Lee, Sudhakar K. Venkatesh, Lynette Li San Teo, Christopher Au, and Wieslaw 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.

[17] Sharath Gopal and Demetri Terzopoulos, "A Unified Statistical / Deterministic Deformable Model for LV Segmentation in Cardiac MRI," *Springer- Verlag Berlin Heidelberg*, pp. 180–187, 2014.

[18] Carlos Santiago, Jacinto C. Nascimento, and Jorge S. Marques, "A new ASM framework for left ventricle segmentation exploring slice variability in cardiac MRI volumes," *Neural Computing and Applications*, vol. 28, no. 9, pp. 2489–2500, 2017.

[19] Huaifei Hu, Haihua Liu, Zhiyong Gao, and Lu Huang, "Hybrid segmentation of left ventricle in cardiac MRI using gaussian-mixture model and region restricted dynamic programming," *Magnetic Resonance Imaging*, vol. 31, no. 4, pp. 575–584, 2013.

[20] Hong Liu, Huaifei Hu, Xiangyang Xu, and Enmin Song, "Automatic left ventricle segmentation in cardiac MRI using topological stable-state thresholding and region restricted dynamic programming," *Academic Radiology*, vol. 19, no. 6, pp. 723–731, 2012.

[21] Jol Schaerer, Christopher Casta, Jrme Pousin, and Patrick 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.

[22] Georgios Tzimiropoulos and Maja Pantic, "Optimization problems for fast AAM fitting in-the-wild," *Proceedings of the IEEE International Conference on Computer Vision*, pp. 593–600, 2013.