



This is a repository copy of *MA-SOCRATIS: An automatic pipeline for robust segmentation of the left ventricle and scar*.

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/180842/>

Version: Accepted Version

Article:

Mamalakis, M. orcid.org/0000-0002-4276-4119, Garg, P., Nelson, T. et al. (3 more authors) (2021) MA-SOCRATIS: An automatic pipeline for robust segmentation of the left ventricle and scar. *Computerized Medical Imaging and Graphics*, 93. 101982. ISSN 0895-6111

<https://doi.org/10.1016/j.compmedimag.2021.101982>

Article available under the terms of the CC-BY-NC-ND licence (<https://creativecommons.org/licenses/by-nc-nd/4.0/>).

Reuse

This article is distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs (CC BY-NC-ND) licence. This licence only allows you to download this work and share it with others as long as you credit the authors, but you can't change the article in any way or use it commercially. More information and the full terms of the licence here: <https://creativecommons.org/licenses/>

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.



eprints@whiterose.ac.uk
<https://eprints.whiterose.ac.uk/>



ELSEVIER

Contents lists available at ScienceDirect

Computerized Medical Imaging and Graphics

journal homepage: www.elsevier.com/locate/media

MA-SOCRATIS: An automatic pipeline for robust segmentation of the left ventricle and scar.

Michail Mamalakis^{a,b,*}, Pankaj Garg^{c,1}, Tom Nelson^c, Justin Lee^c, Jim M. Wild^{a,d}, Richard H. Clayton^{a,b}

^aInsigneo Institute for in-silico Medicine, University of Sheffield, Sheffield, UK.

^bDepartment of Computer Science, University of Sheffield, Regent Court, Sheffield, S1 4DP, UK

^cDepartment of Cardiology, Sheffield Teaching Hospitals NHS Trust, Sheffield S5 7AU, UK

^dPolaris, Imaging Sciences, Department of Infection, Immunity and Cardiovascular Disease, University of Sheffield, Sheffield, UK.

ARTICLE INFO

Article history:

Received February 2021

Communicated by M. Mamalakis

MSC codes here, in the form:

2000 MSC: code, code or

2008 MSC: code, code (2000)

2000 MSC: 41A05, 41A10, 65D05,

65D17 Keywords

Keywords: cardiac MRI, unsupervised, automatic segmentation, machine learning, cardiac segmentation, left ventricle, scars

ABSTRACT

Multi-atlas segmentation of cardiac regions and total infarct scar (MA-SOCRATIS) is an unsupervised automatic pipeline to segment left ventricular myocardium and scar from late gadolinium enhanced MR images (LGE-MRI) of the heart.

We implement two different pipelines for myocardial and scar segmentation from short axis LGE-MRI. Myocardial segmentation has two steps; initial segmentation and re-estimation. The initial segmentation step makes a first estimate of myocardium boundaries by using multi-atlas segmentation techniques. The re-estimation step refines the myocardial segmentation by a combination of k-means clustering and a geometric median shape variation technique. An active contour technique determines the unhealthy and healthy myocardial wall. The scar segmentation pipeline is a combination of a Rician-Gaussian mixture model and full width at half maximum (FWHM) thresholding, to determine the intensity pixels in scar regions. Following this step a watershed method with an automatic seed-points framework segments the final scar region.

MA-SOCRATIS was evaluated using two different datasets. In both datasets ground truths were based on manual segmentation of short axis images from LGE-MRI scans. The first dataset included 40 patients from the MS-CMRSeg 2019 challenge dataset (STACOM at MICCAI 2019). The second is a collection of 20 patients with scar regions that are challenging to segment. MA-SOCRATIS achieved robust and accurate performance in automatic segmentation of myocardium and scar regions without the need of training or tuning in both cohorts, compared with state-of-the-art techniques (intra-observer and inter observer myocardium segmentation: 81.9 and 70% average Dice value, and scar (intra-observer and inter observer segmentation: 70.5 and 70.5% average Dice value).

© 2021 Elsevier B. V. All rights reserved.

1. Introduction

Sudden death associated with cardiac arrhythmias remains a significant global health problem. Patients with regions of ventricular fibrosis or scar are at elevated risk of ventricular arrhythmias (Nelson et al., 2019; Martin et al., 2018), and accurate identification

*Multi-atlas - Segmentation Of Cardiac Regions And Total Infarct Scar

*Corresponding author: Tel.: +44 (0) 7562 576321; email: mmamalakis1@sheffield.ac.uk

of scar is important for both risk stratification and guiding interventions (Pashakhanloo et al., 2018). Cardiac MRI with late gadolinium enhancement (LGE) is routinely used for imaging scar in patients (Kurzendorfer et al., 2017). However, scar segmentation from LGE images is challenging (Wu et al., 2001) because there is variability in the size, heterogeneity, and intensity of scar, as well as low resolution and image noise in LGE images (Tao et al., 2010).

The end point of the segmentation process is a set of contours for the left ventricular (LV) endocardium and epicardium, and the edge of scar regions. Current practice is manual segmentation, which relies on sketching contours slice by slice using a pointing device such as mouse or trackball. Three dimensional LGE-MRI can include as many as one hundred slices, and so manual segmentation is time consuming, tedious and limited by inter- and intra observer variability (Yushkevich et al., 2006; Top et al., 2011). Automated segmentation can overcome these limitations, and methods have been developed for automated scar segmentation in atrial (Xiong et al., 2020; Karim et al., 2013) and ventricular (Amado et al., 2004; Detsky et al., 2009; Lu et al., 2012) images.

Methods for automated myocardium segmentation include active appearance models (Cootes et al., 1995; Mitchell et al., 2002) and active contour methods (Kass et al., 1988; Pluempitiwiriyaewej et al., 2005). More recent developments include region based techniques (Chen et al., 1998; Zhang et al., 2001; Kang et al., 2012), deformable models (Rueckert and Burger, 1997; Ecabert et al., 2008; Ecabert et al., 2006), atlas models (Lorenzo-Valdés et al., 2002; van Rikxoort et al., 2010), statistical shape models (Kang et al., 2012), graph cut methods (Meyer, 1993; Lu et al., 2011), and deep learning (Zheng et al., 2018). Although combinations of these approaches can achieve improved performance (Hennemuth et al., 2008; Bai et al., 2016), segmentation of myocardium in noisy LGE-MR images remains challenging and the performance typically depends on the signal to noise ratio in the image (Ukwatta et al., 2016).

Proposed approaches for automated scar segmentation include fuzzy clustering (Detsky et al., 2009), models of pixel intensity distributions (HáKon and Samuel, 1995), threshold-based methods (Karim et al., 2016a), and deep learning (Moccia et al., 2018; Tao et al., 2019). While these approaches are promising, manual correction is often needed (Lu et al., 2012).

Quantitative methods that take into account variability in both manual and automatic observations are key for evaluating automated methods, and a thorough review of methods relevant to cardiovascular imaging is provided by (Popović and Thomas, 2017). Repeatability of measurements derived from images is a further concern, and this was addressed in a clinical study of infarct size (McAlindon et al., 2015). Manual identification of scar contours by experienced experts resulted in a low intra- and inter-observer variability compared to semi-automated methods (McAlindon et al., 2015).

The present paper describes the development and evaluation of a novel automatic segmentation pipeline to identify both the myocardium and regions of scar from LGE images of the left ventricle.

Our main contributions are:

- An unsupervised automatic segmentation pipeline, which does not need training or tuning.
- A robust unsupervised combination of Rician-Gaussian mixture models and watershed techniques for automatic scar segmentation of LV LGE-MRI images.
- Evaluation against an unbiased manual segmentation taking into account intra- and inter- observer variability.
- Overestimation or underestimation of the accuracy of an automatic pipeline can result if the evaluation is based on biased manual segmentation ground truth.

- Evaluation of the robustness and generalization of our myocardium automatic segmentation pipeline; tested in two unrelated and challenging LGE-MRI cohorts without any training, tuning or transfer learning.

2. Related work

The end point of the segmentation process is a set of contours for the left ventricular (LV) endocardium and epicardium, and the edge of scar regions. We review each of these steps in turn.

2.1. Myocardial imaging and segmentation

Myocardial segmentation in LGE-MRI images of LV with scars is difficult. Even expert radiologists need to examine the cine-MRI pair images of the patient to define the boundaries of LV endocardium and epicardium in LGE images. This is a result of the overlap intensity histogram pixels of dead tissue (scar) and blood pool regions, especially in cases of endocardial scar. Many different approaches to automated myocardium segmentation have been developed, and these include active appearance models, edge and contour detection, region-based, graph-cuts, model fitting and deep learning.

One of the earliest techniques described in the literature was three dimensional **active appearance models** (AAM) (Cootes et al., 1995; Mitchell et al., 2002). This technique was initially developed to locate rigid objects in images, and is a very good method for generalised detection of an object. However, it has been found to be less successful for identifying specific objects (Cootes et al., 1995; van Assen et al., 2008). The main limitation of 3D AAM is the assumption that the training data are 3D. However, typical MRI is a stack of 2D slices, which may not align perfectly because the patient may fail to hold their breath during the imaging procedure. Another important drawback, common to many approaches, is dependence on training set. The model learns only the variation of MRI images in the training data, so an outlier may not be detected correctly if it is not represented in the training data.

Active contour methods are another promising technique for edge and contour detection. The idea of 'snakes' was introduced by (Kass et al., 1988), who described a technique related to energy minimizing models that detect contours and edges in images. In this approach an energy function related to the spatial topology of the image is minimised. There is an extension of this technique applied to cardiac segmentation (Pluempitiwiriwawej et al., 2005; Kass et al., 1988; Zhuang et al., 2020). Plenty of studies used active contour for segmentation of Left Ventricle (LV) from short-axis cardiac magnetic resonance images (Yang et al. (2017b), Yang et al. (2017a), and left ventricle myocardial contrast echocardiography (LVMCE) images (Guo et al. (2017), Zhu et al. (2021)). There are some studies that highlights the robustness and high performance of level-set and active contour techniques contrary to CNN networks, without the need of any training dataset (Shi and Li (2021), Zhu et al. (2021)). Furthermore, Arrieta et al. (2017) studied the feasibility of applying a level-set based methods for parallel segmentation of both left and right ventricles from MR images fast and accurate in normal and abnormal heart regions (even in congenital heart diseases). Yang et al. (2017a) implemented a level set method. They used two different steps a convexity preserving mechanism to keep the shape of the evolving contour convex stable, and a two-layer level set method to define the endocardial and epicardial boundaries. They validated their methodology on a public dataset and they further observed that the level set methodology can be used in many other potential applications (medical and computer vision). Also Shi and Li (2021) utilized a convexity preserving mechanism to ensure the curvature, and non-linearity (convexity) in the relationship of the segmented contours of LV. They tested the performance of the level-set pipeline in the scenario where there is no training dataset available. Finally, variational level sets are extensions of the technique (Kass et al., 1988) and have been applied for segmentation of different organs from MR images (Göçeri et al. (2014); Göçeri (2013)). Even if the active contour and level-set method are very robust and high performance unsupervised techniques, the main drawback of these methods

are the need for manual intervention when the image intensity of epicardium and surrounding tissues (mostly lungs and liver) is similar. The reason for this was the sensitivity of computed gradient values to noise, since gray areas in scanning modality images tend to magnify noise.

Region based techniques are related to active contours, and define the region of interest based on global homogeneity within a region. Region based techniques include k-means, fuzzy c-means and the expectation-maximization (EM) algorithm (Goceri, 2016; Zhang et al., 2001; Kang et al., 2012). The main drawback of this method is the possibility to be trapped in local minima, which do not give correct image segmentation. Moreover, other limitations of region based techniques are sensitivity to initialisation, noise, and inhomogeneities in image intensities.

Another common automatic segmentation technique is **model fitting algorithms**. A predefined statistical shape model can be fitted to a specific patient image. The most common algorithms of this technique are deformable models (Rueckert and Burger, 1997; Ecabert et al., 2008; Ecabert et al., 2006), atlas models (Lorenzo-Valdés et al., 2002; van Rikxoort et al., 2010), and statistical shape models (Kang et al., 2012). The main advantage of model fitting algorithms is the ability to deform edges where there is poor resolution of boundaries (weak edges). Because of this, these approaches are more flexible than AAM. However, these methods have high computational cost in terms of memory and time, and they require landmarks from expert or other complex prior shape algorithms such as AAP or ASP. Finally, generalisation of these methods is difficult because performance can be tied to specific vendors and patient groups (Heimann and Meinzer, 2009).

The idea of **graph-cut techniques** was developed by (Greig et al., 1989), who used maximum posteriori estimation (MAP) to determine the maximum flow of a binary image. In graph cut techniques, the user determines seed points in the background and region of interest. Based on these, the algorithm solves MAP and determines the best optimal solution. The algorithm establishes a network with a graph of nodes (pixels) and edges (region of pixels) that are connected to each other, and an optimum cut is determined. Generally, these techniques are accurate. However, the main limitations are the need to manually determine seed points and the over-segmentation effect (Kang et al., 2012). A common technique linked with graph cut is watershed (Meyer, 1993; Lu et al., 2011). Several studies have combined watershed with other segmentation techniques, reducing the effects of noise and over-segmentation (Hennemuth et al., 2008; Bai et al., 2016).

There is an increase trend to use **deep learning algorithms** for cardiac automatic segmentation. The results of supervised methods are better than those obtained by the use of machine learning and computer vision algorithms. In last cardiac segmentation challenge of MICCAI 2017 nine out of ten segmentation methods were based on deep learning (Zheng et al., 2018). One of the main drawbacks of deep learning techniques is overfitting to training data (Goceri (2019)). As a result the model fits the noise of the training set and loses the generalisation of model that can be robust to variation in vendors and cardiac shape-structure. To achieve a valid generalisation a large labelled training set is required from different vendors and with variation of patients (such as abnormal or healthy cases, sex, age). Obtaining this kind of dataset is difficult.

2.2. Automated scar segmentation

Automated segmentation of cardiac scar is challenging, because there is variation of size, shape and location between patients. Several methods have been developed and evaluated for LV scar.

Others used fuzzy clustering techniques to identify infarct scar in multi-contrast delayed enhancement (MCDE) images of left ventricle (Detsky et al., 2009), and a graph cut segmentation technique to separate scar from healthy LV myocardium (Lu et al., 2012). The advantages of these automatic methods were high speed compared to manual segmentation, no false detection of scar region (common in FWHM manual methods) and automatic detection of scar. The limitation was the need for manual correction in

66/136 images, due to misalignment deformation between the DE-MRI (Lu et al., 2012).

Tao et al. (Tao et al., 2010) evaluated a segmentation method which combined intensity and spatial information to segment LV scar. Their approach was based on theoretical density model theory, which assumes that intensity of normal myocardium follows a Rician distribution (HáKon and Samuel, 1995) and scar follows a Gaussian distribution. The authors recommended this method for accurate identification of scar. Moreover they suggested this as useful tool for quantitative assessment of scar burden in late gadolinium enhancement (LGE) MRI.

Karim et al. (Karim et al., 2016a) evaluated six fixed-threshold methods and five scar segmentation algorithms using an LV data-set from human and porcine hearts. An overlap metric and volumetric-based metric (Dice metric) were used to evaluate the algorithms against expert manual segmentation. (Karim et al., 2016a) concluded that the best segmentation algorithm was the conditional Markov random field (MCG) (Kamnitsas et al., 2017; Krähenbühl and Koltun, 2011). A mixture model with watershed transformation algorithm (Hennemuth et al., 2008) had low accuracy (MV=0.48). Some limitations of this study were the small number of subjects (30 patients) and the bias effect of ground truth (one expert). Moreover there was high intensity variation across the images due to coil shading effect, and these segmentation methods measure the portion of scar tissue. However about cardiac disease treatment, the specific location of scar of LV is significant.

In the last decade, deep learning networks have been used for automatic supervised scar segmentation (Moccia et al., 2018; Tao et al., 2019). Deep learning is a promising method that is capable of accurate results. However there is a need for pre-processing steps with training-testing datasets. Moreover there can be a bias in the training dataset if labelled images for supervised approaches rely on only one expert.

2.3. Observer variability and evaluation metrics

Quantitative methods to take into account variability in both manual and automatic observations are key for evaluating automated methods, and a thorough review of methods relevant to cardiovascular imaging is provided by (Popović and Thomas, 2017). Repeatability of measurements derived from images is a further concern, and this was addressed in a clinical study of infarct size by (McAlindon et al., 2015). Manual identification of scar contours by experienced experts resulted in a low intra- and inter-observer variability compared to semi-automated methods.

Inter- and intra- observer variability determine the variation arising from different observers (inter), and from the same observer at different times (intra). The need for a generalized unbiased ground truth to evaluate an automatic segmentation technique is crucial. (Li et al., 2011) compared different methods of ground truth extraction with inter-observer variability from eight experts, and the best voting threshold method was found to provide the most accurate outcome.

2.4. Our pipeline

Based on the review of methods summarised above, we have constructed an automatic pipeline for LV myocardium and scar segmentation as described below.

Myocardium segmentation involves two steps; initial and re-estimate segmentation. The initial step includes a multi-atlas initial segmentation technique (Bai et al., 2013; Kurzendorfer et al., 2017). The re-estimate step uses a combination of k-means thresholding and a level-set active contour segmentation method (Ngo et al., 2017; Lee et al., 2010). Combining these methods enables their limitations to be overcome. The parameters of active contour are determined from the initial step, which overcomes a limitation of the active contour scheme, and the use of active contours overcomes the effects of noise and image artifacts on a k-means pixel-wise classifier.

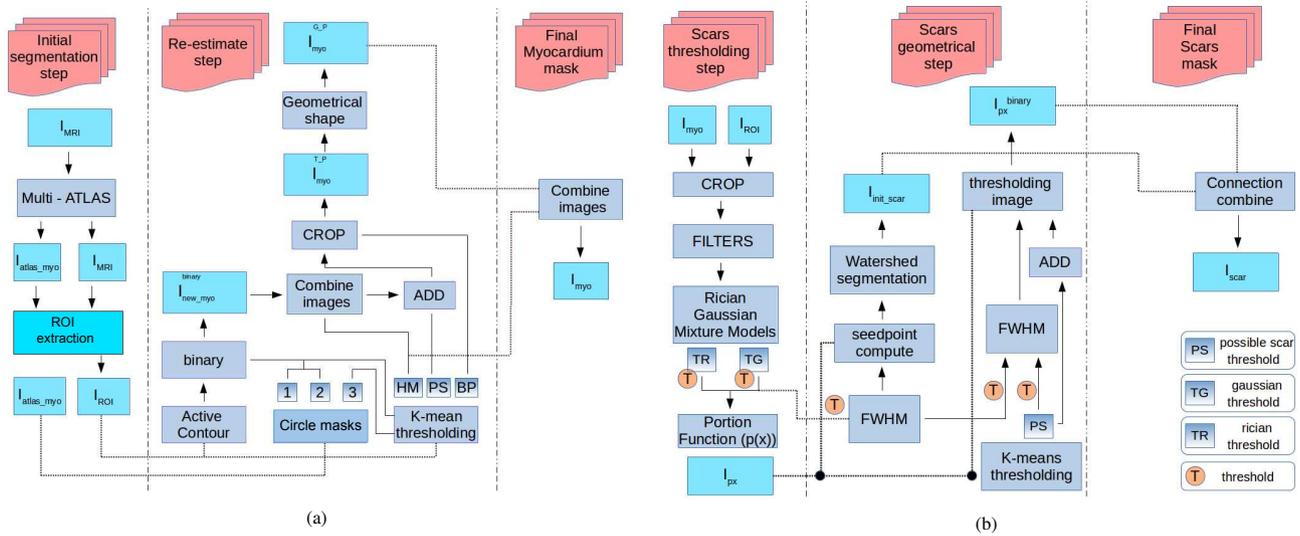


Fig. 1. Workflow for automatic unsupervised segmentation pipeline: (a) myocardium segmentation, (b) scar segmentation. In (a) 1,2,3, denote the three binary circle masks with different radius, BP is blood pool, PS are possible scars, and HM is healthy myocardium. In (b) TR is threshold Rician, TG is threshold Gaussian, and PS are possible scars.

Table 1. Improved accuracy of myocardial segmentation in MA-SOCRATIS with re-estimation compared with the Multi-Atlas segmentation alone, using the MS-CMRSeg 2019 dataset.

Intra-Inter observer results				
Region	Dice (%)	Sensitivity (%)	Specificity (%)	Compared against
MA-SOCRATIS	70.0 ± 5.5	90.5 ± 7.0	93.0 ± 1.7	MS-CMRSeg 2019 challenge
Multi-Atlas	65.4 ± 7.0	57.8 ± 4.6	97.8 ± 1.4	MS-CMRSeg 2019 dataset

For automated segmentation of scar we use an algorithm that combines mixture models histogram fitting and constrained watershed segmentation using an automatic seed point for the scar region. This approach is based on (Hennemuth et al., 2008) with some amendments, so the performance of watershed segmentation is improved. A limitation of the method described by (Hennemuth et al., 2008) was the use of Rayleigh-Gaussian mixture models. Other studies have assumed that healthy myocardium is described by a Rician distribution (Tao et al., 2010; Roy et al., 2012). Thus our approach involves the use of Rician-Gaussian mixture models. In addition, the re-estimation of myocardium can deliver automatic seed points for the scar region with good spatial accuracy.

3. Methodology development

The aim of this study was to create a computational pipeline to process a stack of LGE MR images with 8-bit resolution and size 256×256 pixels, to obtain contours representing the LV endocardium, LV epicardium, and scar edge. We separated the problem into two stages; myocardial segmentation and scar segmentation. These pipelines are illustrated in Figures 1a and 1b, and we describe each in turn. Full details are given in the supplementary information.

3.1. Myocardium segmentation

Myocardium segmentation (Figure 1a) involved two steps; an initial segmentation step and a re-estimate segmentation step.

3.1.1. Myocardium segmentation: Initial step

The pipeline for the initial segmentation step is shown in the left hand column of Figure 1a. The comprehensive review of (Chen et al., 2020b) describes a combination of machine learning techniques and shape models as a way to increase the robustness

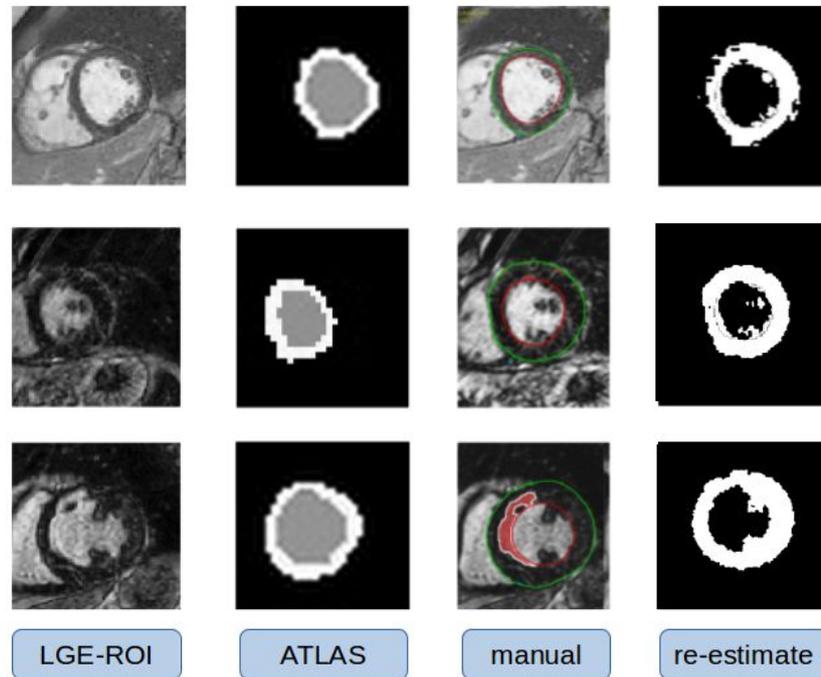


Fig. 2. Comparison of initial and re-estimation steps in three example images. LGE-ROI shows the region of interest, with the LV towards the centre of the image. ATLAS shows the result of initial segmentation with the multi-atlas method. Manual shows manual segmentation of myocardium (green) and scar (red). Re-estimate shows the result of the re-estimation step.

of a segmentation pipeline. This is a very common idea in the literature. The idea of using an atlas pipeline for initial myocardium segmentation was described by (Kurzendorfer et al., 2017). A cine-MRI atlas with deformable registration was used to estimate the LV endocardium and epicardium with scars from 3D LGE-MRI images. A multi-atlas segmentation technique was used by (Yang et al., 2017c) for left atrium segmentation. Since this approach yielded better performance than a single atlas, we used multi-atlas segmentation to initially estimate the myocardium in LGE-MRI images.

The multi-atlas segmentation was based on a method with registration refinement described by (Bai et al., 2013). Landmarks in the LGE-MRI images were manually annotated, although an automated process for generation of landmarks has been described by (Oktay et al., 2017). The multi-atlas segmentation yielded a 256×256 myocardial segmentation $I_{atlas-myocardium}$. A 100×100 pixels ROI image I_{ROI} was then obtained from the center of $I_{atlas-myocardium}$ image by cropping the raw image I_{MRI} .

3.1.2. Myocardium segmentation: Re-estimation step

Figure 2 shows three example LGE images in the first column, and the results of LV myocardium segmentation based on the initial step alone in the second column. The third column is the manual segmentation of myocardium based on expert radiologists and the fourth the MA-SOCRATIS results after the re-estimation step. Table 1 shows the Dice accuracy of MA-SOCRATIS and initial myocardium step (Multi-atlas), demonstrating improved accuracy of our pipeline (MA-SOCRATIS). This improvement of the resolution and performance of myocardium segmentation was based on the central column of Figure 1a. There were four main phases to this re-estimation step; circle estimation, active contour, k-means thresholding, and a geometric phase.

Circle estimation: The initial myocardium segmentation was used to estimate three circular binary masks enclosing different parts of the chamber. The centre of each circle was the origin of the endocardial contour in the initial multi-atlas segmentation. We defined two additional radii, a maximum and a minimum. The maximum was calculated as the half of Euclidean distance between minimum and maximum points of the initial epicardium plus an offset. The minimum radius was calculated from the endocardial

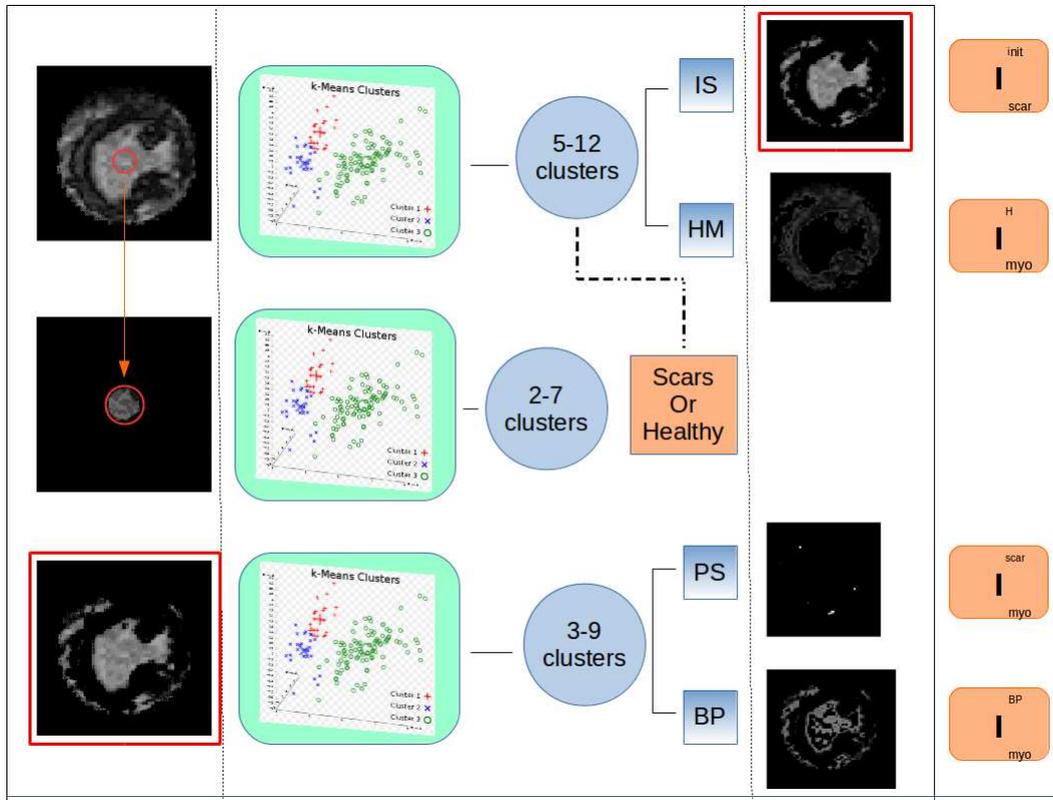


Fig. 3. This is the re-estimate myocardium phase based on k-means. There are three different k-mean clusters of a variation of different (5-12, 2-7 and 3-9) areas of determination. Each of these clusters are focusing in a different region of interest (ROI). Respectively from up to down, the first is focusing in the whole LV region histogram, the second in blood pool histogram and the third in initial scar region. Based on these 3 k-mean clusters the BP, PS, HM and IS are defined in the end of this step, where: BP: blood pool, PS: possible scars, HM: healthy myocardium, IS: initial scars

contour in a similar way (equations/Fig. in supplementary material). The endocardial and epicardial circles (first and second masks) were used as binary masks to constrain an active contour segmentation method described below. The third circle was used to define maximum intensity of the blood pool region as part of the k-means thresholding (Goceri and Dura (2018)).

K-means thresholding: Figure 3 shows the k-means structure we used in our pipeline. The framework is a combination of three k-means clusters with a variation in the initial number of regions. The first k-means cluster has 5 to 12 regions, the second 2 to 7, and third 3 to 9 (The k-means cluster are determined by fixed intensity pixel values from 10 to 200 bytes, equally split with respect of the number of clusters). The variation in number of regions is determined by the histogram intensity in each image. The first k-means cluster was used to create an homogeneous image, so an FWHM threshold could be computed (using $radius_{circle}^{max}$ mask). The second k-means cluster was used to define the maximum intensity of the blood pool area (using $radius_{circle}^{BP}$ mask). Given the homogeneous image from first k-means cluster, a third k-means cluster computed the FWHM threshold for possible scar. Based on this FWHM threshold, the blood pool and regions of possible scar were defined. If the threshold of the blood pool in the second k-means cluster was greater than the highest mean value of the first cluster then it was assumed that there was no scar present. In this case, the FWHM threshold of the first k-means cluster defines only healthy myocardium and blood pool regions (lower and higher respectively) and the third k-mean was not computed.

The outputs of the k-means framework were therefore estimates of the pixel intensity thresholds for healthy myocardium, blood pool and possible scar regions.

Geometric phase: The geometric phase is shown in Figure 4. The thresholds estimated from the k-means framework were used to compute a healthy myocardium image I_{myo}^H , a possible scar surface image I_{myo}^{scar} and a blood pool image I_{BP} . I_{ROI} was used

to initialise an active contour without edge detection method (Chan and Vese, 2001). This was combined with the first and second binary circle masks to restrict endocardium and epicardium areas. A framework to process redefined binary myocardium mask I_{newmyo}^{binary} based on an object removal filter (Lehmann, 2007), median smooth filter and binary filter was used. Lastly, pre-processing filters were used on I_{newmyo}^{binary} to reduce noise and artifacts in the image.

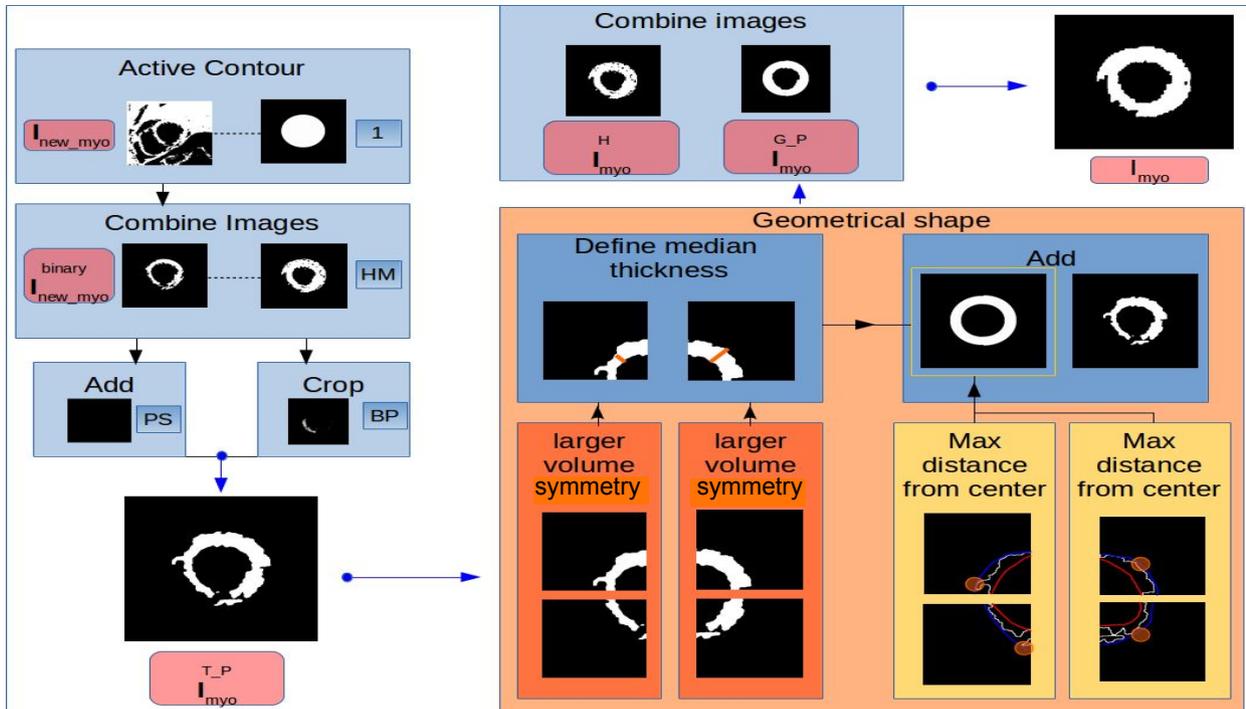


Fig. 4. Detailed workflow for final step of re-estimation of myocardium segmentation. The outer circle mask (1), healthy myocardium (HM), possible scar (PS), and blood pool (BP) were combined, a further geometrical mask was applied (see text for details), and a final myocardial binary mask I_{myo} was computed.

After *k-means thresholding* we estimated median myocardium thickness (Figure 4), the last component of the re-estimation step. We split I_{myo}^{TP} and its associated epicardial contour into four quadrants. We extracted two maximum volumes of I_{myo}^{TP} between the mirror sub-parts. A median ventricular wall thickness was then computed. Four arc masks were created based on the two median thicknesses. These arc masks were located at the maximum distance of the correspond epicardium sub-parts. Each arc mask was added to I_{myo}^{TP} so the I_{myo}^{GP} myocardium mask of geometric phase was extracted. Finally I_{myo}^H and I_{myo}^{GP} were combined, so the myocardial binary mask I_{myo} could be computed.

3.2. Scar segmentation

The pipeline for scar segmentation is shown in Figure 1b. The pipeline has two detection components, threshold and geometric, and is designed to take into account the variation in pixel intensity histograms and the geometrical shape-size of the LV.

Threshold detection component: We used a Rician-Gaussian mixture model (Tao et al., 2010; Roy et al., 2012) to estimate the intensity threshold of normal myocardium and scar. First, we defined a threshold for scar in each myocardium region. We used four different mixture models of Rician-Gaussian (RG) distributions (RG, RGG, RRG and RRRG), and these are shown in Figure 5. For each LGE-MRI image, we tested each combination of RG mixture model. Then we sorted the mean values of each RG model from maximum to minimum. The maximum value was defined as Gaussian threshold (TG) and the minimum as Rician threshold (TR). We assumed two different cases of mixture model initialisation; fixed mean values (at 0, 60, 100, 200) and initial mean values based on (Hennemuth et al., 2008). We defined the threshold for scar region as the FWHM of maximum (Gaussian mean value) and

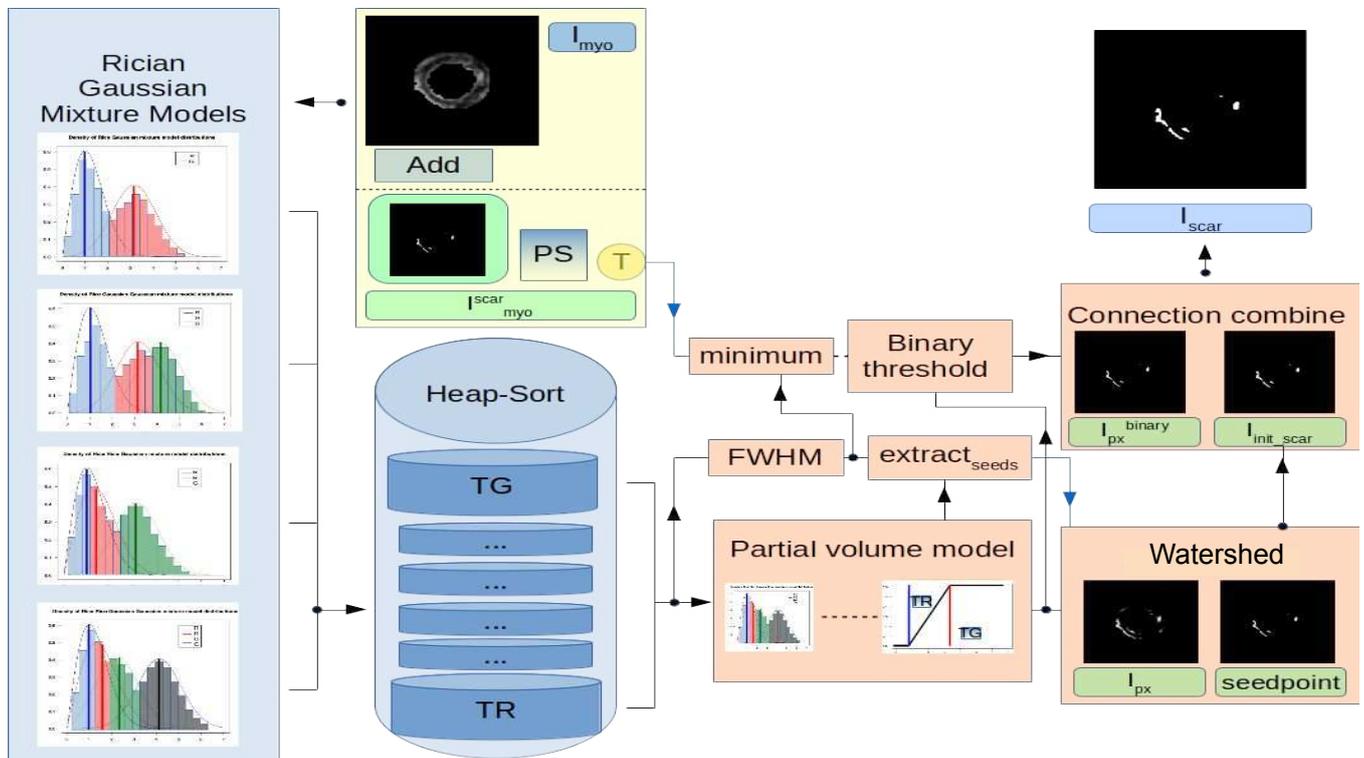


Fig. 5. Details of workflow for automatic scar segmentation, showing the Rician-Gaussian mixture model framework combined with watershed seed point technique. See text for details.

minimum (Rician mean value) sorted mean values. Lastly, these thresholds of Rician (TR) and Gaussian (TG) distributions were used to define a linear partial volume model (Hennemuth *et al.*, 2008) (more details are provided in supplementary material).

Geometric segmentation component: For the geometric component, we used a graph cut technique to identify the geometrical shape of scars. We used a morphological watershed segmentation technique, with initial seed points computed automatically with respect to the linear partial volume model equation and Euclidian distances $dist_{variation}$ between endocardium and epicardium contours (Beare R., 2006) (more details are provided in supplementary material).

The seed points and the image from the $p(x)$ equation were the input of a watershed segmentation algorithm (Meyer and Beucher, 1990), which provided an initial estimate of the scar region I_{init_scar} . The minimum of the RGMM threshold and the k-mean threshold obtained in the re-estimation phase of myocardial segmentation was used as a threshold to convert the image from the $p(x)$ equation to a binary image as shown in Figure 5. The final scar binary image I_{scars} was then extracted by combining the binary $p(x)$ image and I_{init_scar} .

4. Implementation

4.1. Cardiac MR images and image analysis

4.1.1. Training and test data

We assembled a new dataset of short-axis LGE-MR images from 20 patients. All images were acquired, anonymised and stored in DICOM format. All patients gave informed consent. The region of interest (ROI) was defined on the center of the endocardium region of each image. For myocardium regions, we evaluated our system on the data of 20 patients we collected and on the data of 40 patients of MS-CMRSeg 2019 challenge (STACOM at MICCAI 2019). The LGE-MRI dataset was obtained from a collection of three different MRI machine vendors with a variation in the initial settings used for a Philips MRI scanner. The images from

Table 2. Summary of automatic segmentation of myocardium and scar compared to intra- and inter-observer ground truth. Scar segmentation performance was measured with either manual (MM) or automatic myocardium segmentation (AM). MCC is Matthew's correlation coefficients

Intra-Inter observer results						
Region	Dice (%)	Sensitivity (%)	Specificity (%)	MCC (%)	F1 score (%)	Compared against
Myocardium	81.9 ± 9.4	91.9 ± 11.0	95.9 ± 1.2	80.0 ± 10.0	81.9 ± 9.4	Intra Ground Truth
Myocardium	78.1 ± 9.3	94.2 ± 11.8	95.1 ± 1.2	76.3 ± 9.6	78.1 ± 9.3	Inter Ground Truth
Myocardium	70.0 ± 5.5	90.5 ± 7.0	93.0 ± 1.7	68.1 ± 5.8	70.0 ± 5.5	MS-CMRSeg 2019
Scar	70.5 ± 11.7	75.0 ± 12.5	99.6 ± 0.3	68.2 ± 12.1	70.5 ± 11.7	Intra Ground Truth (MM)
Scar	29.5 ± 8.0	31.6 ± 9.0	98.1 ± 0.7	24.6 ± 8.9	29.5 ± 8.0	Intra Ground Truth (AM)
Scar	70.5 ± 17.2	85.4 ± 7.5	99.6 ± 0.2	68.3 ± 17.7	70.5 ± 17.2	Inter Ground Truth (MM)
Scar	24.4 ± 9.0	36.5 ± 11.2	98.0 ± 0.7	20.0 ± 10.0	24.4 ± 9.0	Inter Ground Truth (AM)

each patient comprised between 8 and 12 short axis slices. Three experts segmented these images by identifying the LV endocardial and epicardial surfaces, and the borders of scar. Segmentation was done using MASS (research version 2017; Leiden University Medical Center, Leiden, the Netherlands).

4.1.2. Ground truth estimation

To assess *inter-* observer variability, each of the three experts segmented the myocardium and scar in each slice from each of the 20 patients. The experts were independent and they did not know the structure of our automatic segmentation pipeline. To assess *intra-*observer variability, one expert repeated the segmentation three times. Ground truths (GTs) were then determined based on the best voting threshold assessment described by (Li et al., 2011). GT masks were given by $k = (J + 1)/2$ where k was the threshold values of agreement among observations, and J the total manual observers-observations.

To evaluate the performance of the processing pipeline, for each scar and myocardium region we computed the Dice, Sensitivity and Specificity metrics.

4.2. Software and hardware details

The code implementation is available on a public repository with url: <https://github.com/INSIGNEO/MA-SOCRATIS>. The code is implemented in C++ and python. For automatic segmentation of myocardium-scar regions of our dataset, we used an Intel(R) Core(TM) i5-4570 CPU @ 3.20GHz with 8GB RAM. The automatic segmentation pipeline needed 8.55 minutes for each patient of the test LGE-MRI data set (8-12 slices of LGE-MRI images, around 40 s per slice). This level of performance was considered suitable for the present proof of concept study, but we would expect significant improvement if the code is adapted for GPU or other multi-threading hardware.

5. Results

5.1. Automated segmentation of myocardium and scar

The automated segmentation of myocardium and scar was evaluated against both intra- and inter-observer GTs, and these comparisons are summarised in Table 2. The overall accuracy is given as a mean Dice, sensitivity, and specificity score with standard deviations, which were evaluated over all 167 MR images included in the test set. The Dice score measures the overall similarity of the pipeline and ground truth classifications, whereas the sensitivity and specificity measure how well the pipeline identifies regions of myocardium and scar relative to the ground truth. To further evaluate the accuracy and robustness of our myocardium automatic pipeline in different cohorts we tested it in the 40 patients cohort of MS-CMRSeg 2019 challenge dataset (STACOM at MICCAI 2019). Table 2 shows the results.

Overall, the Dice scores for myocardium segmentation were higher than for scar segmentation, with only small differences between comparisons with intra- and inter-observer ground truths. In MS-CMRSeg 2019 dataset the accuracy reduced to 70 %. MA-SOCRATIS achieved these results without the need of training, tuning or transfer learning in both cohorts. The Matthew's correlation coefficients for the three cases (inter, intra, and MS-CMRSeg-2019) is slightly lower than the average Dice value (-2%). The F1 score is the same as the average Dice values. As the F1 score is the harmonic mean of precision and sensitivity, and Dice is the harmonic mean of specificity and sensitivity, we assume that the specificity is equal with the precision.

The accuracy of automated scar segmentation was the same for both intra- and inter-observer ground truths. The accuracy of scar segmentation depends on the myocardium segmentation, and so we used the intra- and inter-observer ground truths for myocardium segmentation before applying the scar segmentation pipeline. The sensitivity of scar segmentation was lower than for myocardium segmentation, whereas the specificity was very high. This can be explained because the binary scar region has a smaller number of pixels compared to the number of pixels classified as myocardium. The sensitivity will therefore have a higher variance as sensitivity is calculated from the number of white pixels, and the shape, size and location of scars is less smooth than myocardium.

Figure 6 illustrates the performance of automated myocardium segmentation with respect to inter-observer variability in three cases where accuracy (mean Dice) was lower than average (<70%), close to the average (70-80%), and higher than average (>80%). Although the top two rows show good accuracy with respect to the manual observations, our automated method has identified papillary muscle as myocardium whereas this has been ignored by the manual observers. This is an advantage of the pipeline as in (Duchenne *et al.*, 2018) and (Rajiah *et al.*, 2019) verify the importance of papillary muscle as part of ventricle myocardium detection. In the bottom row the myocardium is very clear on the image, but the low intensity of surrounding tissue has resulted in patchy automated segmentation. These observations are consistent with the examples of myocardial segmentation compared with intra-observer variability shown in Figure 7, where the detection of papillary muscle by the automated segmentation is even more pronounced.

Scar segmentation is illustrated in Figures 8 and 9, for inter- and intra-observer variability respectively. For large scars (top row in both figures), the accuracy of the automated pipeline is high, and the consistency of between and within manual observers is also high. For smaller scars, it becomes harder for both automated and manual segmentation to distinguish between scar and noise in the image.

5.2. Accurate myocardium segmentation is important for subsequent scar segmentation

In our pipeline, segmentation of an LGE image proceeds in two stages: segmentation of the myocardium and then segmentation of scar regions. In the results reported above, we have treated these two parts of the pipeline separately. When we combined automated myocardium segmentation with automated scar segmentation then we found that errors in automated myocardium segmentation were associated with a loss of accuracy for scar segmentation, and these findings are summarised in Table 2. The accuracy of our scar segmentation pipeline based on manual segmentation was 70.5% with respect to inter- or intra-observer ground truth (see Table 2), but the accuracy fell to less than 30% when our automatic myocardial segmentation pipeline was used.

The reasons for this finding are illustrated in Figure 10, which shows a set of images for a single patient where myocardium was segmented automatically with high accuracy relative to manual myocardium segmentation. The four columns on the left of the figure highlighted in blue show that the myocardium is distinct in the raw LGE-MR image, and there is generally a good correspondence between the automatic segmentation and the inter- and intra-observer ground truths. The four columns highlighted in orange show the inter- and intra- observer ground truths for scar, together with the automatic scar segmentation based on the



Fig. 6. Myocardium segmentation and inter-observer ground truth. Comparison of automatic segmentation, manual segmentation by 1st and 3rd observer, and overall inter-observer ground truth for images where the automated myocardial segmentation showed high accuracy (top row), mid-range accuracy (middle row), and low accuracy (bottom row) compared to ground truth. Right hand column shows the image ROI in each case.

manual myocardial (MM) segmentations shown in columns 2 and 3. The final column shows automatic scar segmentation based on the automatic myocardium (AM) segmentation shown in column 1.

With automatic myocardium (AM) segmentation more scar pixels are identified close to the endocardial surface compared to both manual scar segmentation and automatic scar segmentation based on manual myocardium (MM) segmentation. In this example, the endocardial pixels account for the differences in scar segmentation with the AM and MM approaches, and this effect can be traced back to the myocardial segmentations in column 1 of Figure 10.

6. Discussion

6.1. Robustness of pipeline in different GT cases

To illustrate the differences in manual segmentation between and within individual experts, Figure 11 shows an example of manual myocardial and scar segmentation by three independent observers. In Figure 11(a), the inter-observer differences for myocardial segmentation can be clearly seen in the upper row, and the lower row illustrates how these differences were resolved using the best threshold approach to provide the ground truth. Figure 11(b) shows that the inter-observed differences for scar segmentation were greater than for myocardial segmentation. The intra-observer variability was generally less than inter-observer variability.

In addition Table 3 shows the effectiveness of GT determination. As it can be observed there is a consistent variability within the three observers (inter-variability) and the three observations (intra-variability). The manual myocardium segmentation was much more repeatable than scar segmentation.



Fig. 7. Myocardium segmentation and intra-observer ground truth. Comparison of automatic segmentation, manual segmentation with 1st and 2nd observations, and overall intra-observer ground truth for images where the automated myocardial segmentation showed high accuracy (top row), mid-range accuracy (middle row), and low accuracy (bottom row) compared to ground truth. Right hand column shows the image ROI in each case.

Table 3. Summary of intra- and inter-observer variability in manual segmentation of myocardium and scar compared to overall ground truths (GT)

Observation/er	Intra Myo Dice (%)	Intra Scar Dice (%)	Inter Myo Dice (%)	Inter Scar Dice (%)
First	90.9	50.5	90.5	44.4
Second	85.2	41.6	86.4	38.2
Third	88.7	44.8	89.6	43.6
GT	100	100	100	100

To check the robustness of our pipeline, we evaluated them with respect to both single expert observations and each intra observation of scar and myocardium lesions (single-case). By this method, we could capture variation in sensitivity, accuracy and specificity metrics of our two pipelines.

Figures 13 and 12 show the variation of sensitivity specificity and Dice value results by changing the ground truth inter and intra base. We used the terminology: Myo_Sens_GT for myocardial sensitivity total ground truth, Myo_Spec_O1 for myocardial specificity observer/ation one ground truth and Myo_Dice_O2 for myocardial Dice observer/ation two ground truth.

Figures 12 and 13 show that the sensitivity, specificity and Dice metrics are less robust in cases of both one observation and one observer compared to the total GT. The standard deviation, especially in the sensitivity metric was also higher in single observations and with a single observer compared to the total GT, and the mean values of the metrics are in some cases an overestimation or underestimation regarding the total GT.

Figures 6, 7 show examples of myocardium segmentation with a variation between different observers and observations. By the extraction of an unbiased truth we can evaluate our pipeline in a more generalized way. So our pipeline accuracy is less sensitive to observation bias, something that can be problem with other automated segmentation techniques such as deep learning networks.

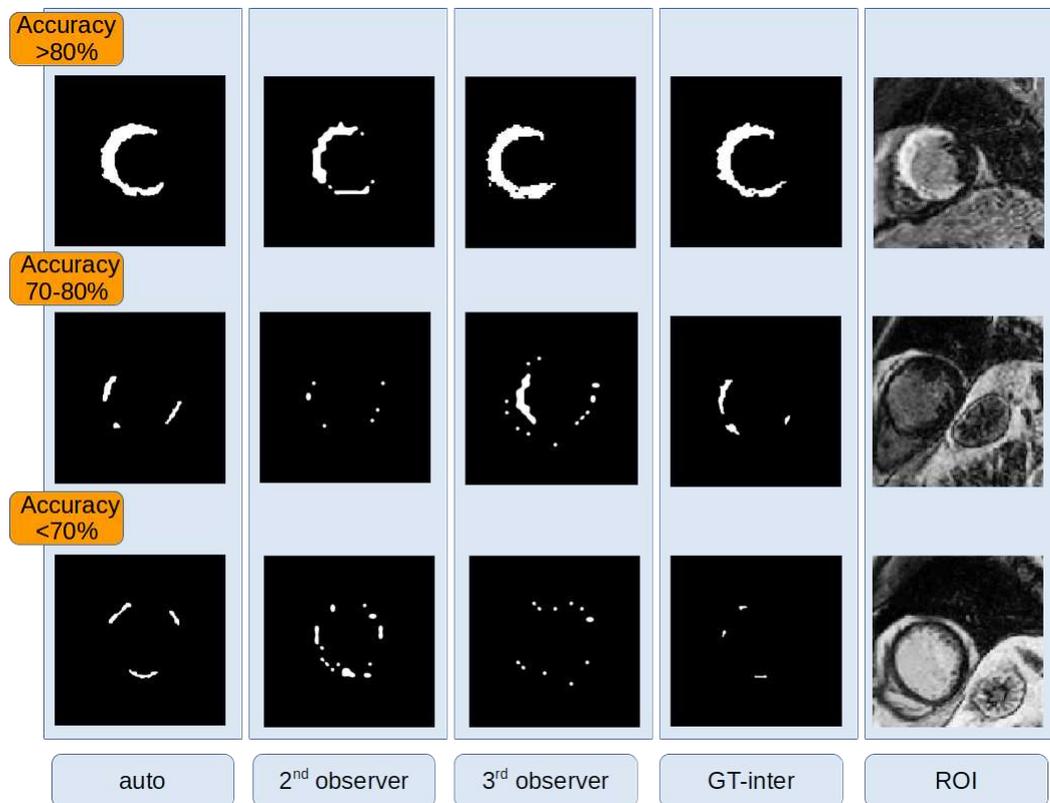


Fig. 8. Scar segmentation and inter-observer ground truth. Comparison of automatic segmentation of scar, manual segmentation of scar with 1st and 3rd observer, and overall inter-observer ground truth for images where the automated scar segmentation showed high accuracy (top row), mid-range accuracy (middle row), and low accuracy (bottom row) compared to ground truth. Right hand column shows the image ROI in each case.

Moreover, by the extraction of the unbiased GT the effect of noise is reduced between observations, especially in scar regions. In the examples shown in figures 8 and 9, the variability of the first, second and third observation-observer and the reduced variability of the GT can be seen in each case (intra-inter). The inter- and intra-observer variability that we observed highlights the difficulty of obtaining a GT, and the importance of at least two different experts and at least two observations by one expert for constructing a less biased GT.

These findings demonstrate the importance of reducing bias in GTs used to evaluate automatic segmentation pipelines.

6.2. Comparison with established and state of the art techniques

There are few publicly available cardiac LGE-MRI datasets with scar segmentation. Thus comparison of MA-SOCRATIS with established and state of the art techniques using the same dataset was challenging. We therefore compared our pipeline with established automatic or semi-automatic segmentation methods of LV with scars; focusing on datasets with identically or almost the same experimental conditions as our cohorts (20 patient testing cohort, LGE-MRI modalities). To the authors' knowledge the datasets used to evaluate the methods which we compared with MA-SOCRATIS are not published. An exception is the MS-CMRSeg challenge of 2019, which we included as a cross-institute validation dataset.

The established methods we used have been evaluated with datasets which followed a specific pre-processing and tuning methodology ((Karim et al., 2016b), (Moccia et al., 2018), (Liu et al., 2019)). So we had to utilize the same procedures or none in our datasets. We chose not to do any tuning in either our own LGE-MRI dataset of 20 patients, or in the MS-CMRSeg challenge of 2019. We did not use any supervision from experts, so the pipeline was totally automatic. Moreover, we used an initial noise reducing pre-processing step for both datasets, as described in the supplementary material (Pre-processing and post processing



Fig. 9. Scar segmentation and intra-observer ground truth. Comparison of automatic segmentation, manual segmentation with 1st and 2nd observations, and overall intra-observer ground truth for images where the automated scar segmentation showed high accuracy (top row), mid-range accuracy (middle row), and low accuracy (bottom row) compared to ground truth. Right hand column shows the image ROI in each case.

Table 4. Comparison of our myocardium and scar segmentation pipeline and established machine/deep learning techniques

Region	Dice (%)	Technique (cohort)
Myocardium	86.0	DL (Moccia et al., 2018) dataset
Myocardium	81.9	MA-SOCRATIS (our Intra GT dataset)
Myocardium	80.8	CNN (Liu et al., 2019) dataset
Myocardium	78.1	MA-SOCRATIS (our Inter GT dataset 80.5 median)
Region	Median (%)	Technique (cohort)
Scar	85.0	MCG (Karim et al., 2016b) dataset
Scar	75.5	MA-SOCRATIS (our Inter GT dataset)
Scar	74.0	KCL (Karim et al., 2016b) dataset
Scar	73.0	AIT (Karim et al., 2016b) dataset
Scar	72.5	MA-SOCRATIS (our Intra GT dataset)
Scar	44.0	MV (Karim et al., 2016b) dataset

frameworks section).

Table 4 compares the performance of our MA-SOCRATIS pipeline with state of the art techniques for myocardium segmentation of LV with scar from LGE-MRI images based on deep learning techniques. All of these state of the art techniques are supervised to some extent, and the performance of our myocardial segmentation pipeline was comparable or better than the Dice metrics quoted in these studies. The segmentation techniques described by (Moccia et al., 2018; Liu et al., 2019) all used deep learning myocardium segmentation networks applied to LGE-MRI images. They used supervised techniques and still they did not achieve more than 86.0 % Dice value. In contrast our method MA-SOCRATIS is an unsupervised method with no tuning or transfer learning techniques, yet achieved an 81.9 % Dice.

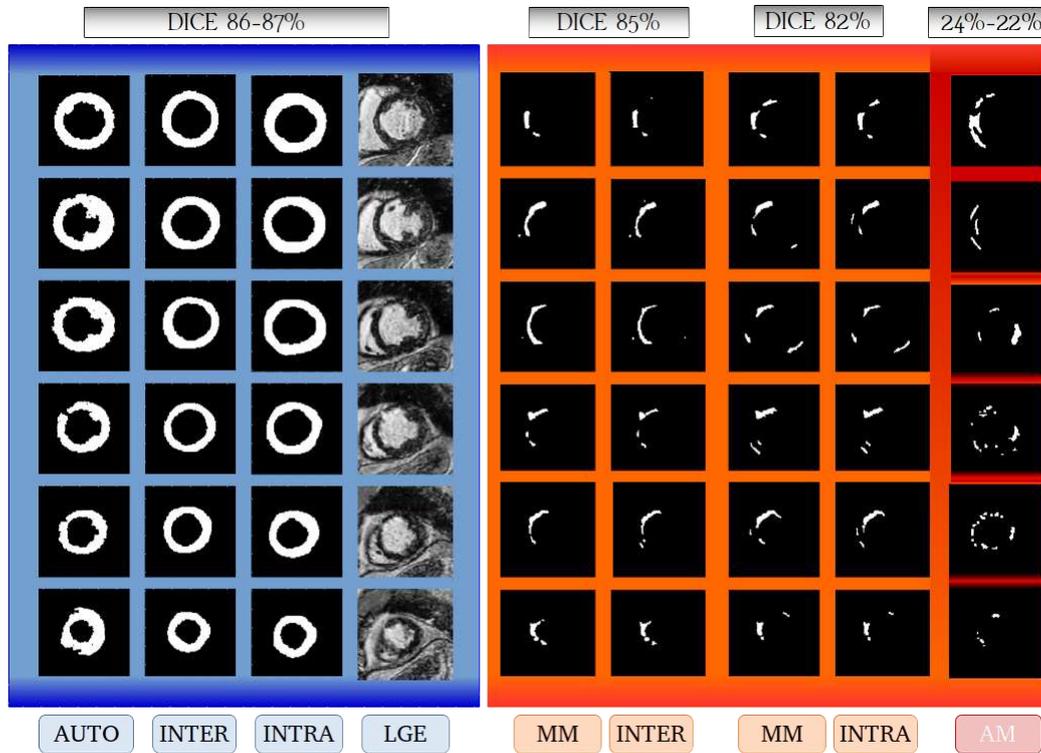


Fig. 10. Images from a single patient showing comparison of automated myocardium segmentation with manual inter- and intra- observer ground truth (blue); automated scar segmentation with manual myocardium segmentation (MM) compared to inter- and intra- observer ground truth (orange); and scar segmentation with automated myocardium segmentation (AM, red).

In one of the studies listed in table 4, (Karim et al., 2016b) evaluated four different techniques with a test dataset of 20 patients. Of the methods evaluated, the conditional Markov random field (MCG) had the best results following by Graph-cuts with EM-algorithm technique (KCL) and support vector machines combined with a level set technique (AIT). All of these techniques were supervised and some were semi-automatic. This means that there is a need of training in a specific cohort. The cohort has to be split in testing (20 patient) and training (10 patient) patients. The only unsupervised technique was the EM-algorithm and watershed transformation (MV), and these achieved a modest median accuracy of only 44%. Our scar segmentation pipeline is unsupervised, and with our test dataset of 20 patients, we achieved a 72.5 and 75.5 % median Dice accuracy for intra- and inter-observer ground truths respectively. The MV technique evaluated by (Karim et al., 2016b) and our scar segmentation pipeline are both based on the study of (Hennemuth et al., 2008), although we have modified the approach of (Hennemuth et al., 2008) as described in the methods section.

We also tested our pipeline using the open source MS-CMRSeg 2019 challenge dataset. Table 5 summarizes the results, and compares the performance of MA-SOCRATIS with other published methods. The approach described by (Chen et al., 2020a,b; Liu et al., 2020a; Chen et al., 2019; Wang et al., 2020b) used a multi-modal image translation network from balanced steady-state free-precession (bSSFP) images to LGE images. In all case these methods need tuning or supervised training with the training cohort of the MS-CMRSeg 2019 challenge dataset. (Chen et al., 2020a) used a cascaded segmentation for the bSSFP images. Although they achieved an unsupervised segmentation of the LGE-MRI images, the cascade network required supervised training. (Wang et al., 2020b) used manual ROI definition and they used all T2 cine-MRI and LGE in training process. They used an image post-processing process; such as hole filling technique, connected component analysis and the largest connected component of

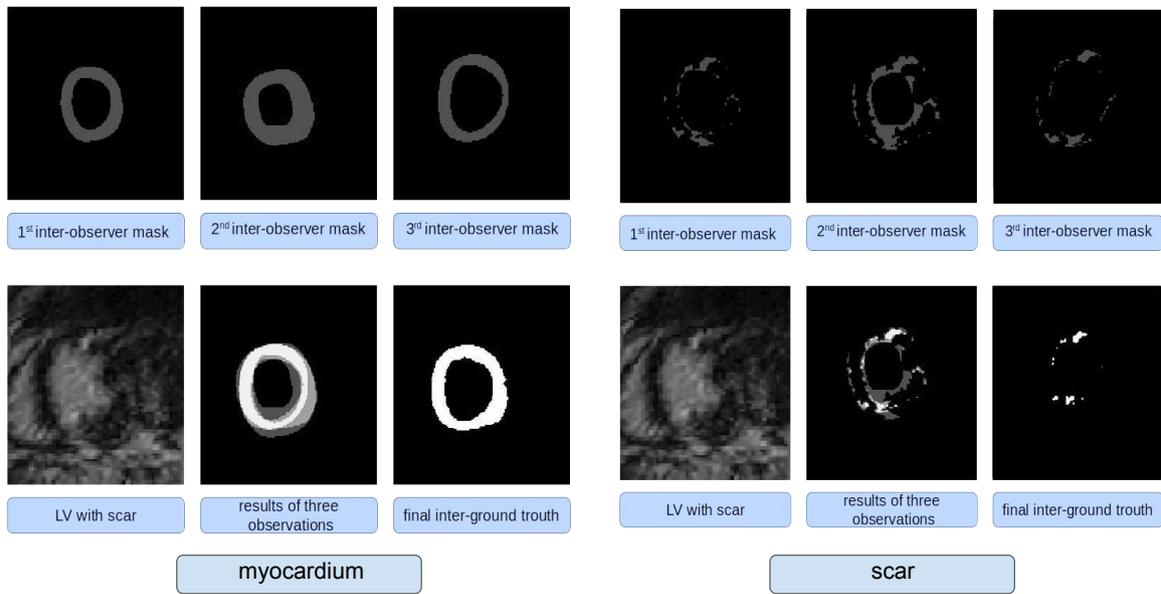


Fig. 11. Extraction of unbiased inter-observer ground truth for myocardium (left) and scar (right).

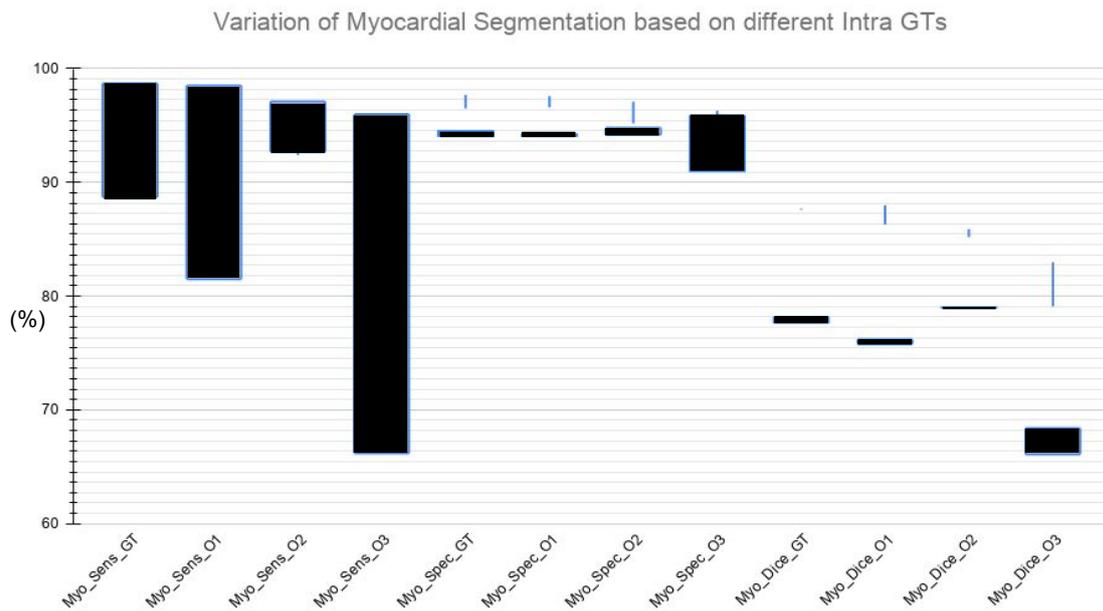


Fig. 12. Variation in sensitivity, Dice and specificity metrics of automatic myocardial segmentation by changing the intra observation ground truths.

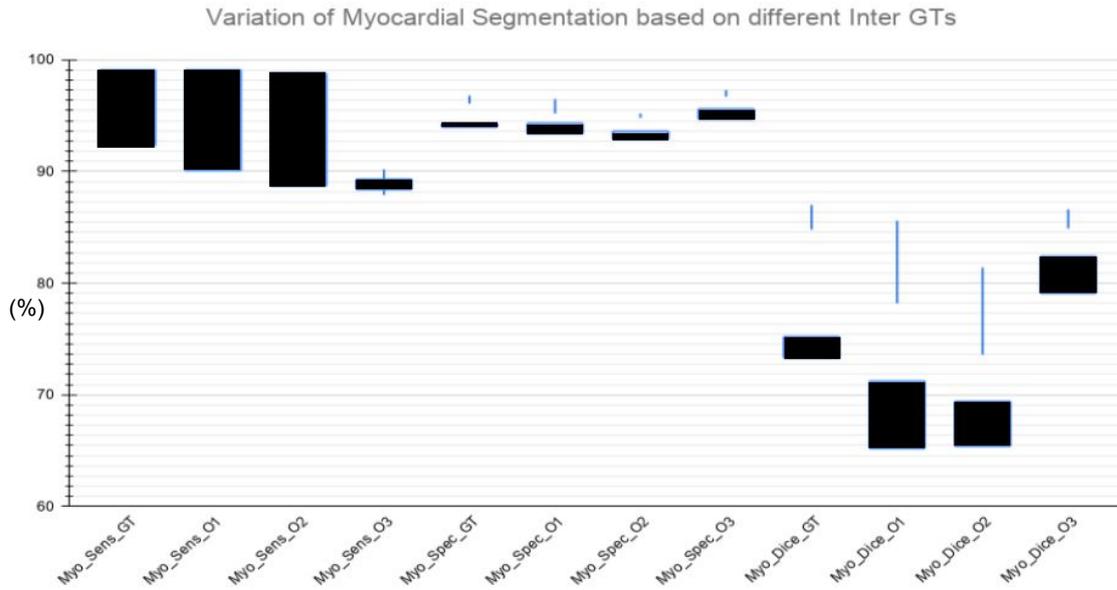


Fig. 13. Variation in sensitivity, Dice and specificity metrics of automatic myocardial segmentation by changing the inter observation ground truths

Table 5. Comparison of our myocardium segmentation pipeline and state of the art techniques in MS-CMRSeg 2019 published dataset based on average and standard deviation (SD) Dice values (NM: not mentioned).

Region	Dice +/- SD (%)	Technique (MS-CMRSeg 2019 dataset)
Myocardium	83.0 +/- NM	Unsupervised Multi-modal (Chen et al., 2020a)
Myocardium	82.7 +/- 6	SK-Unet (Wang et al., 2020b)
Myocardium	81.2 +/- 6.1	Combining Multi-Sequence and Synthetic-LGE (Campello et al., 2020)
Myocardium	80.8 +/- NM	Automatic Cardiac Segmentation Framework (Liu et al., 2020b)
Myocardium	80.1 +/- NM	Adversarial Domain Adaptation Network (GFRM) (Wang et al., 2020a)
Myocardium	78.0 +/- NM	Multi-Atlas-CNN noise images (Roth et al., 2020)
Myocardium	74.9 +/- 10.0	Supervised-Domain-Adaptation (Vesal et al., 2020)
Myocardium	71.4 +/- 10.0	Deep-Learning-Framework (Zheng et al., 2020)
Myocardium	70.0 +/- 5.5	MA-SOCRATIS
Myocardium	68.6 +/- 7.8	Shape-Transfer GAN network (Tao et al., 2020)
Myocardium	65.4 +/- 7.0	Multi-Atlas (Bai et al., 2013)
Myocardium	61.7 +/- 8.6	Pseudo-3D network (Liu et al., 2020a)
Myocardium	61.7 +/- NM	CNN weak Domain Transfer (Chen et al., 2019)
Myocardium	61.0 +/- NM	Adversarial-CNN-Domain-Adaptation (Chen et al., 2020c)
Myocardium	57.3 +/- NM	2D/3D-UNET Framework (Xu et al., 2020)
Myocardium	47.0 +/- 11.7	UNET++ Framework (Ren et al., 2020)

all slices in each patient. Segmentation areas that exceeded the largest connected range were removed. The approach described by (Chen et al., 1998; Liu et al., 2020a; Wang et al., 2020b) used a multi-modal image translation network from balanced steady-state free-precession (bSSFP) images to LGE images. In all case they need tuning or supervised training in T2, bSSFP and LGE-MR images in the training cohort of MS-CMRSeg 2019 challenge dataset.

In contrast to these studies, MA-SOCRATIS is free of any kind of multi-modal training in bSSFP or LGE-MRI images or tuning, yet it still achieved 70.0 % accuracy against the MS-CMRSeg 2019 challenge dataset. It is important to mention that our pipeline achieved the most robust results as the standard deviation of the results was the smallest (5.5 %) of all methods.

6.3. Limitations of pipeline

The detection of the initial position and size of endocardial and epicardial regions was not always accurate (multi-atlas initial segmentation step). This is the main limitation of our myocardium pipeline. A possible explanation could be the shape variation of multi-atlas process combined with the fact that we used the method for a cross-modality application (cine-MRI to LGE-MRI). As a result, the MRI slices of some of the patients were miss-align with respect of the multi-atlas slices (registration error). The low accuracy of myocardium segmentation in some patients, can explain the big difference between AM and MM scar segmentation results (29.5 to 70.5 % respectively). Finally, the time needed for the MA-SOCRATIS to deliver the myocardium and scar automatic segmentation, was a lot (almost an hour per patient), as the registration step of multi-atlas method was time consuming.

7. Conclusion

In this study, we have presented an automatic unsupervised segmentation pipeline (MA-SOCRATIS) which achieved impressive results when evaluated with two different datasets (STACOM at MICCAI 2019, our collection of 20 patients with challenging scar regions and LV area). Hence, we have tested the adaptability of our pipeline with new cohorts, and assessed performance without the need for training, tuning or changes to the initial hyper-parameters. The robustness and generalization of this pipeline are therefore demonstrated, and our unsupervised method achieved Dice metrics comparable with state-of-the-art supervised techniques.

In this study we have highlighted the importance of correct detection of myocardium boundaries in LGE-MRI from patients with infarct tissue, in both automatic and manual segmentation. Furthermore, we elucidated the sensitivity of our two pipelines with GT from both repeated observations and different experts. As a result, we clarified the robustness of our pipeline, which produces generalize and accurate results free of intra-inter observer bias effects. We demonstrated a robust unsupervised combination of Rician-Gaussian mixture models and watershed techniques for automatic scar segmentation of LV in LGE images. Lastly, we justified the importance of an unbiased ground truth to evaluate an automatic pipeline; thus no overestimation or underestimation effects of metrics results be occurred.

As future work, we will evaluate our pipeline in further cohorts of LGE images using again unbiased extractions of the ground truths so the results of our pipeline can be more robust and credible.

Acknowledgment

We acknowledge the use of facilities of the Research Software Engineering Sheffield (RSE) UK and the JADE Tier 2 HPC UK system specification. This work was funded by University of Sheffield Department of Computer Science Scholarship. We thank Gavin Bainbridge, Caroline Richmond, Margaret Saysell, and Petra Bijsterveld for their invaluable assistance in recruiting and collecting data for this study. All authors declare that they have no competing interests.

References

- Amado, L.C., Gerber, B.L., Gupta, S.N., Rettmann, D.W., Szarf, G., Schock, R., Nasir, K., Kraitchman, D.L., Lima, J.A., 2004. Accurate and objective infarct sizing by contrast-enhanced magnetic resonance imaging in a canine myocardial infarction model. *Journal of the American College of Cardiology* 44, 2383 – 2389.
- Arrieta, C., Uribe, S., Sing-Long, C., Hurtado, D., Andia, M., Irarrazaval, P., Tejos, C., 2017. Simultaneous left and right ventricle segmentation using topology preserving level sets. *Biomedical Signal Processing and Control* 33, 88–95. URL: <https://www.sciencedirect.com/science/article/pii/S174680941630180X>, doi:<https://doi.org/10.1016/j.bspc.2016.11.002>.
- Bai, J., Li, P., Wang, K., 2016. Automatic whole heart segmentation based on watershed and active contour model in ct images, in: 2016 5th International Conference on Computer Science and Network Technology (ICCSNT), pp. 741–744. doi:10.1109/ICCSNT.2016.8070256.
- Bai, W., Shi, W., O'Regan, D.P., Tong, T., Wang, H., Jamil-Copley, S., Peters, N.S., Rueckert, D., 2013. A probabilistic patch-based label fusion model for multi-atlas segmentation with registration refinement: Application to cardiac mr images. *IEEE Transactions on Medical Imaging* 32, 1302–1315. doi:10.1109/TMI.2013.2256922.
- Beare R., L.G., 2006. The watershed transform in itk - discussion and new developments.

- Campello, V.M., Martín-Isla, C., Izquierdo, C., Petersen, S.E., Ballester, M.A.G., Lekadir, K., 2020. Combining multi-sequence and synthetic images for improved segmentation of late gadolinium enhancement cardiac mri, in: Pop, M., Sermesant, M., Camara, O., Zhuang, X., Li, S., Young, A., Mansi, T., Suinesiaputra, A. (Eds.), *Statistical Atlases and Computational Models of the Heart. Multi-Sequence CMR Segmentation, CRT-EPiggy and LV Full Quantification Challenges*, Springer International Publishing, Cham. pp. 290–299.
- Chan, T.F., Vese, L.A., 2001. Active contours without edges. *IEEE Transactions on Image Processing* 10, 266–277. doi:10.1109/83.902291.
- Chen, C., Ouyang, C., Tarroni, G., Schlemper, J., Qiu, H., Bai, W., Rueckert, D., 2020a. Unsupervised multi-modal style transfer for cardiac mr segmentation, in: Pop, M., Sermesant, M., Camara, O., Zhuang, X., Li, S., Young, A., Mansi, T., Suinesiaputra, A. (Eds.), *Statistical Atlases and Computational Models of the Heart. Multi-Sequence CMR Segmentation, CRT-EPiggy and LV Full Quantification Challenges*, Springer International Publishing, Cham. pp. 209–219.
- Chen, C., Qin, C., Qiu, H., Tarroni, G., Duan, J., Bai, W., Rueckert, D., 2020b. Deep learning for cardiac image segmentation: A review. *Frontiers in Cardiovascular Medicine* 7, 25.
- Chen, C.W., Luo, J., Parker, K.J., 1998. Image segmentation via adaptive k-mean clustering and knowledge-based morphological operations with biomedical applications. *IEEE Transactions on Image Processing* 7, 1673–1683. doi:10.1109/83.730379.
- Chen, J., Li, H., Zhang, J., Menze, B., 2019. Adversarial convolutional networks with weak domain-transfer for multi-sequence cardiac mr images segmentation. [arXiv:1908.09298](https://arxiv.org/abs/1908.09298).
- Chen, J., Li, H., Zhang, J., Menze, B., 2020c. Adversarial convolutional networks with weak domain-transfer for multi-sequence cardiac mr images segmentation, in: Pop, M., Sermesant, M., Camara, O., Zhuang, X., Li, S., Young, A., Mansi, T., Suinesiaputra, A. (Eds.), *Statistical Atlases and Computational Models of the Heart. Multi-Sequence CMR Segmentation, CRT-EPiggy and LV Full Quantification Challenges*, Springer International Publishing, Cham. pp. 317–325.
- Cootes, T., Taylor, C., Cooper, D., Graham, J., 1995. Active shape models-their training and application. *Computer Vision and Image Understanding* 61, 38 – 59. doi:<https://doi.org/10.1006/cviu.1995.1004>.
- Detsky, J.S., Paul, G., Dick, A.J., Wright, G.A., 2009. Reproducible classification of infarct heterogeneity using fuzzy clustering on multicontrast delayed enhancement magnetic resonance images. *IEEE Transactions on Medical Imaging* 28, 1606–1614. doi:10.1109/TMI.2009.2023515.
- Duchenne, J., Turco, A., Bézy, S., Ünlü, S., Pagourelis, E.D., Beela, A.S., Degtiarova, G., Vunckx, K., Nuyts, J., Coudyzer, W., Claus, P., Rega, F., Gheysens, O., Voigt, J.U., 2018. Papillary muscles contribute significantly more to left ventricular work in dilated hearts. *European Heart Journal - Cardiovascular Imaging* 20, 84–91. URL: <https://doi.org/10.1093/ehjci/iej043>, doi:10.1093/ehjci/iej043, [arXiv:https://academic.oup.com/ehjci/article-pdf/20/1/84/27170997/iej043.pdf](https://academic.oup.com/ehjci/article-pdf/20/1/84/27170997/iej043.pdf).
- Ecabert, O., Peters, J., Schramm, H., Lorenz, C., von Berg, J., Walker, M.J., Vembar, M., Olszewski, M.E., Subramanian, K., Lavi, G., Weese, J., 2008. Automatic model-based segmentation of the heart in ct images. *IEEE Transactions on Medical Imaging* 27, 1189–1201. doi:10.1109/TMI.2008.918330.
- Ecabert, O., Peters, J., Weese, J., 2006. Modeling shape variability for full heart segmentation in cardiac computed-tomography images. doi:10.1117/12.652105.
- Goceri, E., 2016. Automatic labeling of portal and hepatic veins from mr images prior to liver transplantation., in: *Int J Comput Assist Radiol Surg*.
- Goceri, E., 2019. Challenges and recent solutions for image segmentation in the era of deep learning, in: *2019 Ninth International Conference on Image Processing Theory, Tools and Applications (IPTA)*, pp. 1–6. doi:10.1109/IPTA.2019.8936087.
- Goceri, E., Dura, E., 2018. Comparison of weighted k-means clustering approaches, in: *International Conference on Mathematics (ICOMATH2018), An Istanbul Meeting for World Mathematicians, Minisymposium on Approximation Theory Minisymposium on Math Education*, p. 98.
- Göçeri, E., 2013. A comparative evaluation for liver segmentation from spir images and a novel level set method using signed pressure force function, in: *Phd thesis, Izmir Institute of Technology, Izmir, Turkey*.
- Greig, D.M., Porteous, B.T., Seheult, A.H., 1989. Exact maximum a posteriori estimation for binary images. *Journal of the Royal Statistical Society: Series B (Methodological)* 51, 271–279.
- Guo, Y., Du, G.Q., Xue, J.Y., Xia, R., hang Wang, Y., 2017. A novel myocardium segmentation approach based on neutrosophic active contour model. *Computer Methods and Programs in Biomedicine* 142, 109–116. URL: <https://www.sciencedirect.com/science/article/pii/S0169260716302747>, doi:<https://doi.org/10.1016/j.cmpb.2017.02.020>.
- Göçeri, E., Gürcan, M.N., Dicle, O., 2014. Fully automated liver segmentation from spir image series. *Computers in Biology and Medicine* 53, 265–278. URL: <https://www.sciencedirect.com/science/article/pii/S001048251400211X>, doi:<https://doi.org/10.1016/j.combiomed.2014.08.009>.
- HáKón, G., Samuel, P., 1995. The rician distribution of noisy mri data. *Magnetic Resonance in Medicine* 34, 910–914. doi:10.1002/mrm.1910340618.
- Heimann, T., Meinzer, H.P., 2009. Statistical shape models for 3d medical image segmentation: A review. *Medical Image Analysis* 13, 543 – 563. doi:<https://doi.org/10.1016/j.media.2009.05.004>.
- Hennemuth, A., Seeger, A., Friman, O., Müller, S., Klumpp, B., Oeltze, S., Peitgen, H.O., 2008. A comprehensive approach to the analysis of contrast enhanced cardiac mr images. *IEEE Transactions on Medical Imaging* 27, 1592–1610. doi:10.1109/TMI.2008.2006512.
- Kamnitsas, K., Baumgartner, C.F., Ledig, C., Newcombe, V.F.J., Simpson, J.P., Kane, A.D., Menon, D.K., Nori, A.V., Criminisi, A., Rueckert, D., Glocker, B., 2017. Unsupervised domain adaptation in brain lesion segmentation with adversarial networks, in: *IPMI*.
- Kang, D., Woo, J., Kuo, C.C.J., Slomka, P.J., Dey, D., Germano, G., 2012. Heart chambers and whole heart segmentation techniques: review 21, 10901–10917.
- Karim, R., Bhagirath, P., Claus, P., Housden, R.J., Chen, Z., Karimaghloo, Z., Sohn, H.M., Rodríguez, L.L., Vera, S., Albà, X., Hennemuth, A., Peitgen, H.O., Arbel, T., Ballester, M.A.G., Frangi, A.F., Götte, M., Razavi, R., Schaeffter, T., Rhode, K., 2016a. Evaluation of state-of-the-art segmentation algorithms for left ventricle infarct from late gadolinium enhancement mr images. *Medical Image Analysis* 30, 95 – 107. doi:<https://doi.org/10.1016/j.media.2016.01.004>.
- Karim, R., Bhagirath, P., Claus, P., Housden, R.J., Chen, Z., Karimaghloo, Z., Sohn, H.M., Rodríguez, L.L., Vera, S., Albà, X., Hennemuth, A., Peitgen, H.O., Arbel, T., Ballester, M.A.G., Frangi, A.F., Götte, M., Razavi, R., Schaeffter, T., Rhode, K., 2016b. Evaluation of state-of-the-art segmentation algorithms for left ventricle infarct from late gadolinium enhancement mr images. *Medical Image Analysis* 30, 95 – 107. doi:<https://doi.org/10.1016/j.media.2016.01.004>.
- Karim, R., Housden, R.J., Balasubramaniam, M., Chen, Z., Perry, D., Uddin, A., Al-Beyatti, Y., Palkhi, E., Acheampong, P., Obom, S., Hennemuth, A., Lu, Y., Bai, W., Shi, W., Gao, Y., Peitgen, H.O., Radau, P., Razavi, R., Tannenbaum, A., Rueckert, D., Cates, J., Schaeffter, T., Peters, D., MacLeod, R., Rhode, K., 2013. Evaluation of current algorithms for segmentation of scar tissue from late gadolinium enhancement cardiovascular magnetic resonance of the left atrium: an open-access grand challenge. *Journal of Cardiovascular Magnetic Resonance* 15, 105. doi:10.1186/1532-429X-15-105.
- Kass, M., Witkin, A., Terzopoulos, D., 1988. Snakes: Active contour models. *International Journal of Computer Vision* 1, 321–331. doi:10.1007/BF00133570.
- Krähenbühl, P., Koltun, V., 2011. Efficient inference in fully connected crfs with gaussian edge potentials, in: *Shawe-Taylor, J., Zemel, R.S., Bartlett, P.L., Pereira, F., Weinberger, K.Q. (Eds.), Advances in Neural Information Processing Systems 24*. Curran Associates, Inc., pp. 109–117.
- Kurzendorfer, T., Forman, C., Schmidt, M., Tillmanns, C., Maier, A., Brost, A., 2017. Fully automatic segmentation of left ventricular anatomy in 3-d lge-mri. *Computerized Medical Imaging and Graphics* 59, 13 – 27. doi:<https://doi.org/10.1016/j.compmimag.2017.05.001>.
- Lee, H., Codella, N.C.F., Cham, M.D., Weinsaft, J.W., Wang, Y., 2010. Automatic left ventricle segmentation using iterative thresholding and an active contour model with adaptation on short-axis cardiac mri. *IEEE Transactions on Biomedical Engineering* 57, 905–913. doi:10.1109/TBME.2009.2014545.
- Lehmann, G., 2007. Label object representation and manipulation with itk. <http://hdl.handle.net/1926/584>.
- Li, X., Aldridge, B., Fisher, R., Rees, J., 2011. Estimating the ground truth from multiple individual segmentations incorporating prior pattern analysis with application to skin lesion segmentation, in: *2011 IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, pp. 1438–1441. doi:10.1109/

ISBI . 2011 . 5872670.

- Liu, T., Tian, Y., Zhao, S., Huang, X., Xu, Y., Jiang, G., Wang, Q., 2020a. Pseudo-3d network for multi-sequence cardiac mr segmentation, in: Pop, M., Sermesant, M., Camara, O., Zhuang, X., Li, S., Young, A., Mansi, T., Suinesiaputra, A. (Eds.), *Statistical Atlases and Computational Models of the Heart. Multi-Sequence CMR Segmentation, CRT-EPiggy and LV Full Quantification Challenges*, Springer International Publishing, Cham. pp. 237–245.
- Liu, Y., Wang, W., Wang, K., Ye, C., Luo, G., 2019. An automatic cardiac segmentation framework based on multi-sequence mr image. *ArXiv abs/1909.05488*.
- Liu, Y., Wang, W., Wang, K., Ye, C., Luo, G., 2020b. An automatic cardiac segmentation framework based on multi-sequence mr image, in: Pop, M., Sermesant, M., Camara, O., Zhuang, X., Li, S., Young, A., Mansi, T., Suinesiaputra, A. (Eds.), *Statistical Atlases and Computational Models of the Heart. Multi-Sequence CMR Segmentation, CRT-EPiggy and LV Full Quantification Challenges*, Springer International Publishing, Cham. pp. 220–227.
- Lorenzo-Valdés, M., Sanchez-Ortiz, G.I., Mohiaddin, R., Rueckert, D., 2002. Atlas-based segmentation and tracking of 3d cardiac mr images using non-rigid registration, in: Dohi, T., Kikinis, R. (Eds.), *Medical Image Computing and Computer-Assisted Intervention — MICCAI 2002*, Springer Berlin Heidelberg, Berlin, Heidelberg. pp. 642–650.
- Lu, Y., Connelly, K.A., Dick, A.J., Wright, G.A., Radau, P.E., 2011. Watershed segmentation of basal left ventricle for quantitation of cine cardiac mri function. *Journal of Cardiovascular Magnetic Resonance* 13, P4. doi:10.1186/1532-429X-13-S1-P4.
- Lu, Y., Yang, Y., Connelly, K.A., Wright, G.A., Radau, P.E., 2012. Automated quantification of myocardial infarction using graph cuts on contrast delayed enhanced magnetic resonance images. *Quantitative imaging in medicine and surgery* 2, 2, 81–6.
- Martin, R., Maury, P., Biscaglia, C., Wong, T., Estner, H., Meyer, C., Dallet, C., Martin, C.A., Shi, R., Takigawa, M., Rollin, A., Frontera, A., Thompson, N., Kitamura, T., Vlachos, K., Wolf, M., Cheniti, G., Duchâteau, J., Massoulié, G., Pambrun, T., Denis, A., Derval, N., Hocini, M., Bella, P.D., Haïssaguerre, M., Jaïs, P., Dubois, R., Sacher, F., 2018. Characteristics of scar-related ventricular tachycardia circuits using ultra-high-density mapping. *Circulation: Arrhythmia and Electrophysiology* 11, e006569. doi:10.1161/CIRCEP.118.006569.
- McAlindon, E., Pufulete, M., Lawton, C., Angelini, G.D., Bucciarelli-Ducci, C., 2015. Quantification of infarct size and myocardium at risk: evaluation of different techniques and its implications. *European Heart Journal - Cardiovascular Imaging* 16, 738–746. doi:10.1093/ehjci/jev001.
- Meyer, F., 1993. The morphological approach to segmentation: the watershed transformation.
- Meyer, F., Beucher, S., 1990. Morphological segmentation. *Journal of Visual Communication and Image Representation* 1, 21 – 46.
- Mitchell, S.C., Bosch, J.G., Lelieveldt, B.P.F., van der Geest, R.J., Reiber, J.H.C., Sonka, M., 2002. 3-d active appearance models: segmentation of cardiac mr and ultrasound images. *IEEE Transactions on Medical Imaging* 21, 1167–1178. doi:10.1109/TMI.2002.804425.
- Moccia, S., Banali, R., Martini, C., Moscogiuri, G., Pontone, G., Pepi, M., Caiani, E.G., 2018. Automated scar segmentation from cmr-lge images using a deep learning approach, in: *2018 Computing in Cardiology Conference (CinC)*, pp. 1–4. doi:10.22489/CinC.2018.278.
- Nelson, T., Garg, P., Clayton, R., Lee, J., 2019. The role of cardiac mri in the management of ventricular arrhythmias in ischaemic and non-ischaemic dilated cardiomyopathy. *Arrhythmia Electrophysiology Review* 8, 191–201. doi:10.15420/aer.2019.5.1.
- Ngo, T.A., Lu, Z., Carneiro, G., 2017. Combining deep learning and level set for the automated segmentation of the left ventricle of the heart from cardiac cine magnetic resonance. *Medical Image Analysis* 35, 159 – 171.
- Oktay, O., Bai, W., Guerrero, R., Rajchl, M., de Marvao, A., O'Regan, D.P., Cook, S.A., Heinrich, M.P., Glocker, B., Rueckert, D., 2017. Stratified decision forests for accurate anatomical landmark localization in cardiac images. *IEEE Transactions on Medical Imaging* 36, 332–342. doi:10.1109/TMI.2016.2597270.
- Pashkhanloo, F., Herzka, D.A., Halperin, H., McVeigh, E.R., Trayanova, N.A., 2018. Role of 3-dimensional architecture of scar and surviving tissue in ventricular tachycardia. *Circulation: Arrhythmia and Electrophysiology* 11, e006131. doi:10.1161/CIRCEP.117.006131.
- Pluempitwiriyawej, C., Moura, J.M.F., and, 2005. Stacs: new active contour scheme for cardiac mr image segmentation. *IEEE Transactions on Medical Imaging* 24, 593–603. doi:10.1109/TMI.2005.843740.
- Popović, Z.B., Thomas, J.D., 2017. Assessing observer variability: a user's guide. *Cardiovascular Diagnosis and Therapy* 7.
- Rajiah, P., Fulton, N.L., Bolen, M., 2019. Magnetic resonance imaging of the papillary muscles of the left ventricle: normal anatomy, variants, and abnormalities. *Insights into Imaging* 10, 83. URL: <https://doi.org/10.1186/s13244-019-0761-3>, doi:10.1186/s13244-019-0761-3.
- Ren, J., Sun, H., Huang, Y., Gao, H., 2020. Knowledge-based multi-sequence mr segmentation via deep learning with a hybrid u-net++ model, in: Pop, M., Sermesant, M., Camara, O., Zhuang, X., Li, S., Young, A., Mansi, T., Suinesiaputra, A. (Eds.), *Statistical Atlases and Computational Models of the Heart. Multi-Sequence CMR Segmentation, CRT-EPiggy and LV Full Quantification Challenges*, Springer International Publishing, Cham. pp. 280–289.
- van Rikxoort, E.M., Isgum, I., Arzhaeva, Y., Staring, M., Klein, S., Viergever, M.A., Pluim, J.P., van Ginneken, B., 2010. Adaptive local multi-atlas segmentation: Application to the heart and the caudate nucleus. *Medical Image Analysis* 14, 39 – 49. doi:<https://doi.org/10.1016/j.media.2009.10.001>.
- Roth, H., Zhu, W., Yang, D., Xu, Z., Xu, D., 2020. Cardiac segmentation of lge mri with noisy labels, in: Pop, M., Sermesant, M., Camara, O., Zhuang, X., Li, S., Young, A., Mansi, T., Suinesiaputra, A. (Eds.), *Statistical Atlases and Computational Models of the Heart. Multi-Sequence CMR Segmentation, CRT-EPiggy and LV Full Quantification Challenges*, Springer International Publishing, Cham. pp. 228–236.
- Roy, S., Carass, A., Bazin, P.L., Resnick, S., Prince, J.L., 2012. Consistent segmentation using a rician classifier. *Medical Image Analysis* 16, 524 – 535. doi:<https://doi.org/10.1016/j.media.2011.12.001>.
- Rueckert, D., Burger, P., 1997. Shape-based segmentation and tracking in 4d cardiac mr images, in: Trocraz, J., Grimson, E., Mösges, R. (Eds.), *CVRMed-MRCAS'97*, Springer Berlin Heidelberg, Berlin, Heidelberg. pp. 43–52.
- Shi, X., Li, C., 2021. Convexity preserving level set for left ventricle segmentation. *Magnetic Resonance Imaging* 78, 109–118. URL: <https://www.sciencedirect.com/science/article/pii/S0730725X21000151>, doi:<https://doi.org/10.1016/j.mri.2021.02.003>.
- Tao, Q., Milles, J., Zeppenfeld, K., Lamb, H.J., Bax, J.J., Reiber, J.H., van der Geest, R.J., 2010. Automated segmentation of myocardial scar in late enhancement mri using combined intensity and spatial information. *Magnetic Resonance in Medicine* 64, 586–594. doi:10.1002/mrm.22422.
- Tao, Q., Yan, W., Wang, Y., Paiman, E.H.M., Shamonin, D.P., Garg, P., Plein, S., Huang, L., Xia, L., Sramko, M., Tintera, J., de Roos, A., Lamb, H.J., van der Geest, R.J., 2019. Deep learning-based method for fully automatic quantification of left ventricle function from cine mr images: A multivendor, multicenter study. *Radiology* 290, 81–88. doi:10.1148/radiol.2018180513. PMID: 30299231.
- Tao, X., Wei, H., Xue, W., Ni, D., 2020. Segmentation of multimodal myocardial images using shape-transfer gan, in: Pop, M., Sermesant, M., Camara, O., Zhuang, X., Li, S., Young, A., Mansi, T., Suinesiaputra, A. (Eds.), *Statistical Atlases and Computational Models of the Heart. Multi-Sequence CMR Segmentation, CRT-EPiggy and LV Full Quantification Challenges*, Springer International Publishing, Cham. pp. 271–279.
- Top, A., Hamarneh, G., Abugharbich, R., 2011. Active learning for interactive 3d image segmentation, in: Fichtinger, G., Martel, A., Peters, T. (Eds.), *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2011*, Springer Berlin Heidelberg, Berlin, Heidelberg. pp. 603–610.
- Ukwatta, E., Arevalo, H., Li, K., Yuan, J., Qiu, W., Malamas, P., Wu, K.C., Trayanova, N.A., Vadakkumpadan, F., 2016. Myocardial infarct segmentation from magnetic resonance images for personalized modeling of cardiac electrophysiology. *IEEE Transactions on Medical Imaging* 35, 1408–1419. doi:10.1109/TMI.2015.2512711.
- van Assen, H.C., Danilouchkine, M.G., Dirksen, M.S., Reiber, J.H.C., Lelieveldt, B.P.F., 2008. A 3-d active shape model driven by fuzzy inference: Application to cardiac ct and mr. *IEEE Transactions on Information Technology in Biomedicine* 12, 595–605. doi:10.1109/TITB.2008.926477.
- Vesal, S., Ravikumar, N., Maier, A., 2020. Automated multi-sequence cardiac mri segmentation using supervised domain adaptation, in: Pop, M., Sermesant, M., Camara, O., Zhuang, X., Li, S., Young, A., Mansi, T., Suinesiaputra, A. (Eds.), *Statistical Atlases and Computational Models of the Heart. Multi-Sequence CMR Segmentation, CRT-EPiggy and LV Full Quantification Challenges*, Springer International Publishing, Cham. pp. 300–308.

- Wang, J., Huang, H., Chen, C., Ma, W., Huang, Y., Ding, X., 2020a. Multi-sequence cardiac mr segmentation with adversarial domain adaptation network, in: Pop, M., Sermesant, M., Camara, O., Zhuang, X., Li, S., Young, A., Mansi, T., Suinesiaputra, A. (Eds.), *Statistical Atlases and Computational Models of the Heart. Multi-Sequence CMR Segmentation, CRT-EPiggy and LV Full Quantification Challenges*, Springer International Publishing, Cham. pp. 254–262.
- Wang, X., Yang, S., Tang, M., Wei, Y., He, L., Zhang, J., Han, X., 2020b. Sk-unet: an improved u-net model with selective kernel for the segmentation of multi-sequence cardiac mr *arXiv:2001.00736*.
- Wu, E., Judd, R., Vargas, J., Klocke, F., Bonow, R., Kim, R., 2001. Visualisation of presence, location, and transmural extent of healed q-wave and non-q-wave myocardial infarction. *Lancet* 357, 21–28. doi:10.1016/S0140-6736(00)03567-4.
- Xiong, Z., Xia, Q., Hu, Z., Huang, N., Bian, C., Zheng, Y., Vesal, S., Ravikumar, N., Maier, A., Yang, X., Heng, P.A., Ni, D., Li, C., Tong, Q., Si, W., Puybareau, E., Khoudli, Y., Géraud, T., Chen, C., Bai, W., Rueckert, D., Xu, L., Zhuang, X., Luo, X., Jia, S., Sermesant, M., Liu, Y., Wang, K., Borra, D., Masci, A., Corsi, C., de Vente, C., Veta, M., Karim, R., Preetha, C.J., Engelhardt, S., Qiao, M., Wang, Y., Tao, Q., Nuñez-Garcia, M., Camara, O., Savioli, N., Lamata, P., Zhao, J., 2020. A global benchmark of algorithms for segmenting the left atrium from late gadolinium-enhanced cardiac magnetic resonance imaging. *Medical Image Analysis*, 101832.
- Xu, H., Xu, Z., Gu, W., Zhang, Q., 2020. A two-stage fully automatic segmentation scheme using both 2d and 3d u-net for multi-sequence cardiac mr, in: Pop, M., Sermesant, M., Camara, O., Zhuang, X., Li, S., Young, A., Mansi, T., Suinesiaputra, A. (Eds.), *Statistical Atlases and Computational Models of the Heart. Multi-Sequence CMR Segmentation, CRT-EPiggy and LV Full Quantification Challenges*, Springer International Publishing, Cham. pp. 309–316.
- Yang, C., Shi, X., Yao, D., Li, C., 2017a. A level set method for convexity preserving segmentation of cardiac left ventricle. 2017 IEEE International Conference on Image Processing (ICIP), 2159–2163.
- Yang, C., Wu, W., Su, Y., Zhang, S., 2017b. Left ventricle segmentation via two-layer level sets with circular shape constraint. *Magnetic Resonance Imaging* 38, 202–213. URL: <https://www.sciencedirect.com/science/article/pii/S0730725X17300115>, doi:<https://doi.org/10.1016/j.mri.2017.01.011>.
- Yang, G., Zhuang, X., Khan, H., Haldar, S., Nyktari, E., Li, L., Ye, X., Slabaugh, G., Wong, T., Mohiaddin, R., Keegan, J., Firmin, D., 2017c. Multi-atlas propagation based left atrium segmentation coupled with super-voxel based pulmonary veins delineation in late gadolinium-enhanced cardiac MRI, in: Styner, M.A., Angelini, E.D. (Eds.), *Medical Imaging 2017: Image Processing*, International Society for Optics and Photonics. SPIE. pp. 312 – 318. doi:10.1117/12.2250926.
- Yushkevich, P.A., Piven, J., Hazlett, H.C., Smith, R.G., Ho, S., Gee, J.C., Gerig, G., 2006. User-guided 3d active contour segmentation of anatomical structures: Significantly improved efficiency and reliability. *NeuroImage* 31, 1116 – 1128.
- Zhang, Y., Brady, M., Smith, S., 2001. Segmentation of brain mr images through a hidden markov random field model and the expectation-maximization algorithm. *IEEE Transactions on Medical Imaging* 20, 45–57. doi:10.1109/42.906424.
- Zheng, Q., Delingette, H., Duchateau, N., Ayache, N., 2018. 3-d consistent and robust segmentation of cardiac images by deep learning with spatial propagation. *IEEE Transactions on Medical Imaging* 37, 2137–2148.
- Zheng, R., Zhao, X., Zhao, X., Wang, H., 2020. Deep learning based multi-modal cardiac mr image segmentation, in: Pop, M., Sermesant, M., Camara, O., Zhuang, X., Li, S., Young, A., Mansi, T., Suinesiaputra, A. (Eds.), *Statistical Atlases and Computational Models of the Heart. Multi-Sequence CMR Segmentation, CRT-EPiggy and LV Full Quantification Challenges*, Springer International Publishing, Cham. pp. 263–270.
- Zhu, X., Wei, Y., Lu, Y., Zhao, M., Yang, K., Wu, S., Zhang, H., Wong, K.K., 2021. Comparative analysis of active contour and convolutional neural network in rapid left-ventricle volume quantification using echocardiographic imaging. *Computer Methods and Programs in Biomedicine* 199, 105914. URL: <https://www.sciencedirect.com/science/article/pii/S0169260720317478>, doi:<https://doi.org/10.1016/j.cmpb.2020.105914>.
- Zhuang, X., Xu, J., Luo, X., Chen, C., Ouyang, C., Rueckert, D., Campello, V.M., Lekadir, K., Vesal, S., RaviKumar, N., Liu, Y., Luo, G., Chen, J., Li, H., Ly, B., Sermesant, M., Roth, H., Zhu, W., Wang, J., Ding, X., Wang, X., Yang, S., Li, L., 2020. Cardiac segmentation on late gadolinium enhancement mri: A benchmark study from multi-sequence cardiac mr segmentation challenge *arXiv:2006.12434*.