Right Ventricle Segmentation From Cardiac MRI: A Collation Study

Caroline Petitjean, Maria A. Zuluaga, Wenjia Bai, Jean-Nicolas Dacher, Damien Grosgeorge, Jérôme Caudron, Su Ruan, Ismail Ben Ayed, Jorge M. Cardoso, Hsiang-Chou Chen, et al.

Right Ventricle Segmentation From Cardiac MRI: A Collation Study

Caroline Petitjean[a,*], Maria A. Zuluaga[b], Wenjia Bai[d], Jean-Nicolas Dacher[c], Damien Grosgeorge[a], Jérôme Caudron[c], Su Ruan[a], Ismail Ben Ayed[h], M. Jorge Cardoso[b], Hsiang-Chou Chen[g], Daniel Jimenez-Carretero[f], Maria J. Ledesma-Carbayo[f], Christos Davatzikos[j], Jimit Doshi[j], Guray Erus[j], Oskar M.O. Maier[f], Cyrus M.S. Nambakhsh[i], Yangming Ou[j,k], Sébastien Ourselin[b], Chun-Wei Peng[g], Nicholas S. Peters[e], Terry M. Peters[i], Martin Rajchl[i], Daniel Rueckert[d], Andres Santos[f], Wenzhe Shi[d], Ching-Wei Wang[g], Haiyan Wang[d], Jing Yuan[i]

[a] *LITIS EA 4108, Université de Rouen, 76801 Saint-Etienne-du-Rouvray, France*
[b] *Centre for Medical Image Computing, University College London, London, UK*
[c] *INSERM U1096, Université de Rouen, 76031 Rouen Cedex, France*
[d] *Biomedical Image Analysis Group, Department of Computing, Imperial College London, UK*
[e] *National Heart and Lung Institute, St Mary's Hospital, Imperial College London, UK*
[f] *Biomedical Image Technologies, Universidad Politécnica de Madrid, and CIBERBBN, Spain*
[g] *Graduate institute of biomedical engineering, National Taiwan University of Science and Technology, Taipei, Taiwan*
[h] *GE Healthcare, London, Ontario, Canada*
[i] *Western University, Robarts Research Institute, London, Ontario, Canada*
[j] *Section of Biomedical Image Analysis (SBIA), Department of Radiology, University of Pennsylvania, USA*
[k] *A.A. Martinos Biomedical Imaging Center, Massachusetts General Hospital, Harvard Medical School, Charlestown, MA, USA*

Abstract

Magnetic Resonance Imaging (MRI), a reference examination for cardiac morphology and function in humans, allows to image the cardiac right ventricle (RV) with high spatial resolution. The segmentation of the RV is a difficult task due to the variable shape of the RV and its ill-defined borders

[*] Corresponding author. Tel: +33 232 955 215 - Fax: +33 232 955 022
Email address: Caroline.Petitjean@univ-rouen.fr (Caroline Petitjean)

in these images. The aim of this paper is to evaluate several RV segmentation algorithms on common data. More precisely, we report here the results of the Right Ventricle Segmentation Challenge (RVSC), concretized during the MICCAI'12 Conference with an on-site competition. Seven automated and semi-automated methods have been considered, along them three atlas-based methods, two prior based methods, and two prior-free, image-driven methods that make use of cardiac motion. The obtained contours were compared against a manual tracing by an expert cardiac radiologist, taken as a reference, using Dice metric and Hausdorff distance. We herein describe the cardiac data composed of 48 patients, the evaluation protocol and the results. Best results show that an average 80% Dice accuracy and a 1 cm Hausdorff distance can be expected from semi-automated algorithms for this challenging task on the datasets, and that an automated algorithm can reach similar performance, at the expense of a high computational burden. Data are now publicly available and the website remains open for new submissions (`http://www.litislab.eu/rvsc/`).

Keywords: Cardiac MRI, right ventricle segmentation, segmentation method evaluation, segmentation challenge, collation study

1. Introduction

Evaluation of right ventricular (RV) structure and function is of great importance in the management of most cardiac disorders, such as pulmonary hypertension, coronary heart disease, dysplasia and cardiomyopathies (Caudron et al., 2011). RV imaging is considered challenging, mainly because of the complex motion and anatomy of the RV. Magnetic resonance imaging (MRI) is increasingly used as a standard tool in the evaluation of the RV function (Haddad et al., 2008; Attili et al., 2010). As a prerequisite to the computation of functional parameters with MRI, the segmentation of the RV cavity on MR images is a necessary step.

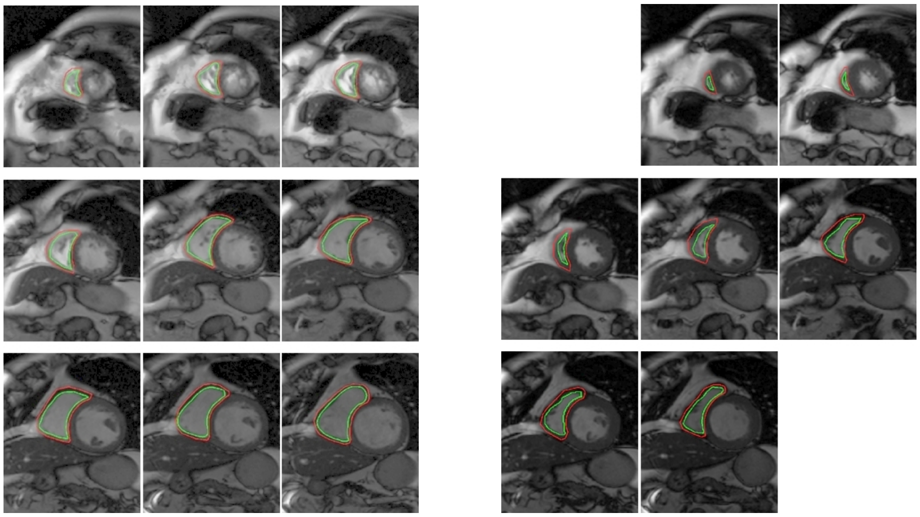
The RV segmentation is challenging because (i) fuzziness of the cavity borders due to blood flow and partial volume effect, (ii) the presence of trabeculations (wall irregularities) in the cavity, which have the same grey level as the surrounding myocardium, (iii) the complex crescent shape of the RV, which varies according to the imaging slice level. The segmentation of the RV is thus currently performed manually in clinical routine. This lengthy and tedious task requires about 15 min by a clinician and is also prone to intra and inter-expert variability (Caudron et al., 2012; Bonnemains et al., 2012).

As a consequence, RV functional assessment has long been considered secondary compared to that of the LV, leaving the problem of RV segmentation wide open. The segmentation of the LV in cardiac MRI has even given rise to three segmentation competitions[1]. The goal of such competitions is to compare different algorithms for a particular task on the same (clinically representative) data, using the same evaluation protocol. Indeed, medical image analysis papers require today solid experiments to prove the usefulness of their proposed methods. However, experiments are often performed on data selected by the researchers, which may come from different institutions, scanners and populations; evaluated with different measures, which make published methods difficult to compare. This has resulted in a growing interest in competitions in medical image analysis. The format of a competition or challenge is usually as follows: given a clinically relevant question, a set of data is collected by a research group, together with its gold standard (i.e. manual annotations). The image data is then made available to volunteering research groups and companies. After performing experiments on the image data, they return the results of their algorithms. Comprehensive and dedicated evaluation tools are then employed for an objective assessment of the algorithm performance, as compared to the gold standard.

RV segmentation algorithms have never been evaluated on common data. The aim of the Right Ventricle Segmentation Challenge (RVSC) is to propose a common evaluation framework, that includes MR datasets, a reference segmentation and standard evaluation measures. More precisely, the task is to delineate the RV endocardium, or endocardium and epicardium, on short-axis views, on end diastole (ED) and end systole (ES) phases (Fig. 1).

In this paper, based on the challenge results, we attempt to address the following questions: what accuracy can be expected from semi-automated and automated algorithms for RV endocardium and epicardium segmenta-

[1] The Cardiac MR Left Ventricle Segmentation Challenge during MICCAI'09: http://smial.sri.utoronto.ca/LV_Challenge/, STACOM'11 Cardiac Left Ventricular Segmentation Challenge during MICCAI'11: http://cilab2.upf.edu/stacom_cesc11/ and the SATA Segmentation Challenge for LV myocardium during MICCAI'13: https://masi.vuse.vanderbilt.edu/workshop2013/index.php/Segmentation_Challenge_Details. These websites, as well as up-to-date information about other challenges may be found at: http://www.grand-challenge.org/



(a) End diastole

(b) End systole

Figure 1: The RV endocardium and epicardium are manually delineated in green and red respectively, at (a) end diastole and (b) end systole, in consecutive SA slices. Note that apical and basal slices may differ between ED and ES.

tion, this latter case being known to be particularly problematic? How do automated algorithms compare to semi-automated ones? What type of methods performs best? Are 3D methods really the most appropriate? With the open availability of both the data and the evaluation framework, we hope to encourage researchers to contribute to this challenging task in the future.

The remainder of the paper is as follows. A brief state-of-the-art of RV segmentation in cardiac MRI is given in Section 2. In Section 4, the cardiac data and the manual reference are described. Evaluation measures and scoring methodology are given in Section 3. The outline of the challenge at MICCAI'12 is presented in Section 5. The 7 participating methods are introduced in Section 6 and the results are detailed and analyzed in Section 7. We finally conclude with a discussion about the methods and the results, and the perspectives in Sections 8 and 9.

2. Previous work

The literature of RV segmentation is much less abundant than the one of LV segmentation (Petitjean and Dacher, 2011; Zhuang, 2013). A number of techniques have indeed been applied to the LV segmentation: in particular thresholding, deformable models and level sets, graph cuts as well as knowledge-based approaches, such as active and appearance shape models or atlas-based methods. Among them, some take advantage of cardiac motion. All these methods are well suited to the MR image characteristics and to the LV geometry and still may give some insight about what techniques could be successful with RV segmentation. In the RV segmentation field, deformable models and, active shape models and their variants, are indeed particularly popular. Note that most of the methods are based on a joint segmentation of both ventricles – only a few methods focus exclusively on RV segmentation (Abi-Nahed et al., 2006; Mahapatra and Buhmann, 2013). Joint segmentation methods take benefit from the similarity of the gray levels in their respective blood cavities and from the stability of the relative positions of both ventricles, and can thus perform a joint segmentation of cardiac ventricles. This kind of information may be used within the active contours (Pluempitiwiriyawej et al., 2004; Grosgeorge et al., 2011) or the graph cuts (Mahapatra and Buhmann, 2013) framework, within an image-driven framework combining thresholding, clustering and morphological operations (Cocosco et al., 2008) or through prior anatomical information used to guide the segmentation process. A priori information can be introduced in the form of a biomechanical model (Sermesant et al., 2003), a 3D heart model (Peters et al., 2007), atlases (Lorenzo-Valdes et al., 2004; Kirisli et al., 2010; Bai et al., 2013), or statistical shape models (Mitchell et al., 2001; Ordas et al., 2003; Lötjönen et al., 2004; Abi-Nahed et al., 2006; Sun et al., 2010; ElBaz and Fahmy, 2012).

Atlas-based segmentation approaches make use of an intensity and a labelled image (denoted atlas) that describes the different structures present in a given type of image. The segmentation of the ventricles is obtained by registering a single (Lorenzo-Valdes et al., 2004) or multiple atlases (Kirisli et al., 2010; Bai et al., 2013) onto the image to be segmented. The main drawback of this technique is its dependence on the quality of the registration, particularly when a single atlas is used.

Statistical shape models have been widely explored in cardiac segmentation (Mitchell et al., 2001; Ordas et al., 2003). They typically consist of three steps: alignment of manually segmented contours, model construction through a technique such as principal component analysis (PCA) and usage of the model for segmentation. Statistical models have been used within the well-known active shape and appearance modelling framework (Cootes et al., 1995). This technique ensures to have a realistic solution since only shapes similar to the training set are allowed, but at the expense of building a training data set with manually generated segmentations.

With the emergence of machine learning techniques in the medical image domain, a novel method was proposed by (Lu et al., 2011) in which the a priori model is learnt via probabilistic boosting trees. In (Mahapatra and Buhmann, 2013), the random forests algorithm is used to generate RV probability maps that were then used within the graph cuts framework for segmentation. Although these methods have proven to be robust, they remain to depend on the quality and amount of annotated training data.

Although all of the works in the literature perform a quantitative evaluation of their methods, there is not a unique and common set of metrics among them. The Dice metric (Abi-Nahed et al., 2006; Grosgeorge et al., 2011; Kirisli et al., 2010; Mahapatra and Buhmann, 2013; Bai et al., 2013), surface-to-surface error (Lötjönen et al., 2004; Lorenzo-Valdes et al., 2004; Kirisli et al., 2010; ElBaz and Fahmy, 2012) (e.g. Hausdorff distance (Grosgeorge et al., 2011; Mahapatra and Buhmann, 2013)), point-to-mesh distance (Sun et al., 2010; Lu et al., 2011), false segmentation rate (Grosgeorge et al., 2011), area and shape similarity measures(Pluempitiwiriyawej et al., 2004), ventricle area difference (Mitchell et al., 2001; Ordas et al., 2003), linear regression analyses of volumes (Lorenzo-Valdes et al., 2004) and correlation with cardiac functional parameters (Sermesant et al., 2003; Cocosco et al., 2008) have been reported. This, in addition to the heterogeneity of the database size used for evaluation, complicates a fair comparison among methods.

3. Evaluation measures

In this challenge, we propose to analyze the performance of the methods technically, by computing the accuracy of the segmentation itself as compared to the gold standard, and clinically, by comparing global RV function indices.

Technical performance. A standard way to assess segmentation result when compared to a reference, is to compute an overlap measure, such as the Dice Metric, and a local, point-based distance measure, as they offer complementary information. For the latter, we chose the 2D Hausdorff Distance (HD), which is less sensitive to contour sampling contrary to some other measures such as mean point to curve error or perpendicular distance between contours, but is sensitive to outliers. The Hausdorff distance is a symmetric measure of distance between both contours (Huttenlocher et al., 1993). Let us denote by A and B the two contours. The HD is defined as:

$$HD(A,B) = \max\left(\max_{a\in A}(\min_{b\in B} d(a,b)), \max_{b\in B}(\min_{a\in A} d(a,b))\right) \tag{1}$$

where $d(\cdot,\cdot)$ denotes Euclidean distance. In the challenge, the Hausdorff distance is computed in mm with spatial resolution obtained from the `PixelSpacing` DICOM field.

The Dice Metric (DM), based on the pixel labeling as the result of a segmentation algorithm, is a measure of area overlap, defined as the ratio of the intersection by the sum of the two surfaces. Let us denote by U and V the areas enclosed by the two contours. The DM is defined as:

$$DM(U,V) = 2\frac{U\cap V}{U+V} \tag{2}$$

The DM varies from 0 (total mismatch) to 1 (perfect match).

The DM is computed from a polygon obtained from the contour points, which makes it little influenced by the contour sampling. HD is also little influenced by the contour sampling, since it is determined by the largest error between two curves. Both error measures (HD and DM) are computed in a multiple 2D way, i.e. one error computed for one slice and one phase, and independently for the endocardium and for the epicardium. Then, errors are averaged over slices, phases (i.e. ED and ES), and patients.

Clinical performance. Segmentation methods are also evaluated on the accuracy of the clinical indices based on the provided contours. One of the major clinical indices is the ejection fraction (EF), the best evaluation tool of RV systolic function. For instance, in young adults, RV EF can be used as a marker of systolic dysfunction, following tetralogy of Fallot surgery, to decide for secondary correction of pulmonary regurgitation. In right ventricular dysplasia, a RV EF value inferior to 40% as measured by MRI is one of the major diagnostic criteria for this pathology (Marcus et al., 2010). Another indicator is the RV mass, whose evaluation is required in some post-operative situations where the RV acts as the LV and vice versa ; for example in the case of a systemic RV, after senning or mustard correction of transposition of the great vessels (Lorenz et al., 1995). Tetralogy of Fallot, following a pulmonary stenosis, is also a major indication for RV mass evaluation (Davlouros et al., 2002). In this case, the free wall of the RV is often thickened, thus easing its segmentation on MR images.

Ventricular volumes are also of interest. Endocardial volumes at ED and ES (denoted resp. V_{endo}^{ED} and V_{endo}^{ES}) are computed (in ml) as the sum of all endocardial areas multiplied by the `SpaceBetweenSlices` value[2]. The definitions of the ejection fraction EF and ventricular mass vm (in g) are based on ventricular volumes, as follows:

$$EF = \frac{(V_{endo}^{ED} - V_{endo}^{ES})}{V_{endo}^{ED}} \tag{3}$$

$$vm = \rho * (V_{epi}^{ED} - V_{endo}^{ED}) \tag{4}$$

where ρ is the density equal to 1.05g/cm^3 (Bogaert et al., 2005). The vm is evaluated at ED, based on the convention used for the LV. RV volumes, EF and mass are obtained for both automated and manual contours. They may be compared through the computation of the correlation coefficient R, linear regression fitting and Bland-Altman analysis.

4. Cardiac data and manual reference

4.1. Cardiac MR data

Patients. From June 2008 to August 2008, all patients referred to our centre (Rouen University Hospital) with a clinical indication of cardiac MR were invited to participate in the study. The institutional review board approved the study and all patients gave written informed consent. Exclusion criteria were as follows: age < 18 years; contra indication to MR; arrhythmias during MR examination; congenital heart disease; and patients referred for an examination that did not include ventricular function analysis (i.e. MR angiography of pulmonary veins or thoracic aorta). A total of 48 patients were included; mean patients' age was 52.1 ± 18.1 years and 36 (75%) were males. Clinical indications were represented by a panel of the currently most frequent cardiac MRI indications in patients with acquired heart diseases: myocarditis, ischaemic cardiomyopathy, suspicion of arrhythmogenic right ventricular dysplasia, dilated cardiomyopathy, hypertrophic cardiomyopathy, aortic stenosis (Caudron et al., 2012).

[2] `SpaceBetweenSlices` = 8.4 mm for all patients. This value is the absolute difference between `SliceLocation` DICOM fields in 2 adjacent images.

Cardiac MR protocol. Cardiac MR examinations were performed at 1.5T (Symphony Tim, Siemens Medical Systems, Erlangen, Germany). A dedicated eight-element phased-array cardiac coil was used. Retrospectively synchronized balanced steady-state free precession sequences were performed for cine analysis, with repeated breath-holds of 10-15 s. Since the subject could not hold the breath at exactly the same position each time, there may be a shift in the slices. This inter-slice shift was not corrected. All conventional planes (2-, 3- and 4-chamber views) were acquired and a total of 10-14 contiguous cine short axis slices were performed from the base to the apex of the ventricles. Sequence parameters were as follows: TR = 50 ms; TE = 1.7 ms; flip angle = 55 ° ; slice thickness = 7 mm; matrix size = 256 × 216; Field of view (FOV) = 360 mm × 420 mm; 20 images per cardiac cycle.

Selection of MR datasets for training and test sets. Cardiac images have been zoomed and cropped to a 256 × 216 (or 216 × 256) pixel ROI. On each MRI dataset, the LV was left visible for joint ventricle segmentation, if necessary. Each patient examination typically includes between 200 and 280 images, with 20 images per cardiac cycle. Spatial resolution is originally 1.6 mm/pixel (as seen from the FOV and matrix size values above) but decreases down to around 0.75 mm/pixel depending on the patient, after zooming and cropping. The MR data is divided into a Training set (16 patients), a Test1 set (16 patients) and a Test2 set (16 patients). Data is anonymized, formatted and named following the naming convention of the MICCAI'09 LV segmentation challenge.

Selection of basal and apical slices. Basal and apical slices have been selected by a cardiac radiologist before the data were released to the participants. The basal and apical slice numbers were provided to the participants. This selection task was thus not part of the challenge.

4.2. Manual RV segmentation methodology

Even though conventions are used to guide cardiac radiologists for their manual delineation, manual segmentation is known to be quite observer-dependent[3] (Caudron et al., 2012; Bonnemains et al., 2012). The following conventions were used in this challenge:

[3] In particular, tracing tricuspid valve and pulmonary valve planes within SA images for the selection of basal and apical slices is a difficult task. Some guidelines may be found in (Prakken et al., 2008).

End-diastole (ED) and end-systole (ES) definitions. ED was defined as the first temporal image of each stack, i.e. the first cine phase of the R-wave triggered acquisition (Fig. 1a) whereas ES was defined on a mid short axis slice as the image with the smallest ventricular cavity area (Fig. 1b).

Definition of basal and apical slices. The basal slice of the RV at ED and ES was inferred from the position of the tricuspid annulus as defined on the 4-chamber view at ED/ES. Apical slice was defined as the last slice with a detectable ventricular cavity.

Manual endocardial and epicardial delineation. The expert manually delineated endocardial and epicardial borders of the RV on short axis slices at ED and ES. Trabeculae and papillary muscles were included in the ventricular cavity. On the septum specifically, the convention is not to include the interventricular septum in the RV mass, and thus to draw the epicardial border stuck to the endocardial one. Even it was the radiologist intention to follow this convention, the drawing software tool would not fully allow it, thus resulting in a minimal distance between the epicardial and the endocardial borders, especially as the duration of the delineation activity was kept compatible with clinical practice. Processing time per patient was indeed around 15 minutes.

5. MICCAI 2012 Challenge outline

The RVSC, organized by five of the authors (CP, DG, SR, JND and JC), was launched in March 2012 with the electronic invitation of a large number of researchers working on cardiac MR segmentation to visit the website and to participate in the challenge, and the announcements on various mailing lists. The RVSC went through different stages of data distribution and result submission and finally ended up with the "3D Cardiovascular Imaging: a MICCAI segmentation challenge" workshop that was organized in conjunction with the 15th International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI), held on October 1st, 2012 in Nice Sophia Antipolis, France, that included an on-site competition. 47 teams initially registered to the challenge and 7 of them submitted results and participated to the on-site challenge. The 7 evaluated algorithms are described in Section 6 and more details can be found in the full paper version, available on our website (http://www.litislab.eu/rvsc/).

Phase 1 (Training). In March 2012, participants were provided with a Training set that included the whole MR examination of 16 patients, i.e. all DICOM MR images, a list of images to be segmented (corresponding to selected images at ED and ES phases), and associated reference manual contours. It means that the participants did not have to choose by themselves apical and basal slices, as well as ED and ES phases. A Matlab evaluation code was provided to participants, intended to help them assessing their segmentation method performance on the Training dataset, with the same evaluation tool subsequently used by the organizers during Test1 and Test2 stages. This code yields DM and HD measures for each image and averaged (mean and standard deviation) for each patient and each phase (ED and ES), as well as correlation coefficient, linear regression fitting and Bland-Altman analysis as described in Section 3.

Phase 2 (Test1). At the beginning of June 2012, participants were provided with a Test1 set, that included MR images of other 16 patients, and a list of images to be segmented. Participants entered their best algorithm to find either the RV endocardium, or the RV endocardium and epicardium automatically, with little or no user intervention. In order to assess the performance of their algorithm on the Test1 set, participants were invited to send their automatic contours to the RVSC organizers, who in return, provided them with the performance measures using the same evaluation code proposed during the Training phase. Participants had then until July 5th to submit their papers describing their methods and results obtained on the Test1 set. These papers are now part of our workshop proceeding (available at `http://www.litislab.eu/rvsc/`). The results were published on the website, anonymously at that time (see Section 7 for the results).

Phase 3 (Test2). On the day of the workshop (October 1st 2012, at MICCAI), challengers were provided a new Test2 set of 16 patients. A 3-hour time-slot was dedicated to the on-site competition. For some algorithms, segmentation of large datasets could be technically challenging in terms of processing power and memory requirements. Thus challengers were allowed to perform the segmentations using remotely located hardware. Results were computed and presented by the organizers during the conference. Note that challengers were allowed to improve their algorithm between the Test1 submission and the day of the challenge.

6. Methods

Methods presented by the 7 challenger teams include three atlas-based methods, two prior based methods, and two prior-free, image-driven methods that make use of the temporal dimension of the data, as shown in Table 1, with some of them processing 3D data and some others 2D data. A majority of them (5) include prior knowledge in their segmentation framework while 2 algorithms were image driven, specifically designed for RV segmentation, but based on cardiac motion. Our panel of methods show the current interest for atlas registration based segmentation.

As stated in the Introduction, the RVSC offered the possibility to segment either the endocardium only (2 methods), or the epicardium and the endocardium of the RV (5 methods). For these two tasks, automated and semi-automated algorithms (3 vs 4 resp.) are distinguished. An automated algorithm does not require landmarks, ROIs, thresholds or similar settings to be defined by the user manually prior to starting the algorithm. A semi-automated algorithm would have some small number of manual steps prior to initiating the algorithm. The contours (unadjusted) that are output by the algorithm are the results that are evaluated.

6.1. CMIC (M. Zuluaga et al.)

This fully automated method is based on a coarse-to-fine strategy. The segmentation of an unseen image is incrementally refined by means of a multi-atlas propagation framework (Zuluaga et al., 2013). The coarser segmentation obtained at each propagation level is used as a mask to gradually improve the registration initialization and accuracy. Through a three level process, the algorithm first locates the heart, then obtains a rough segmentation of the RV and, finally, obtains a refined segmentation of the epi- and endocardium.

First, the unseen image is globally registered to the atlases using a block matching approach. The obtained transformations are applied to the atlas labels, which are all fused using majority voting. This fusion yields a binary mask, which is used next to suppress structures that are not of interest and that might bias the registration process (at this step, the mask covers the complete heart, i.e. LV and RV). Second, the atlas are rigidly registered to the masked unseen image, followed by a non-rigid alignment using a fast free form deformation algorithm. As the segmentation is performed on 2D slices and cardiac images can exhibit large variability, it is necessary to perform

Team	Method principle	A/SA	Contours
CMIC, UK (Zuluaga et al., 2013)	2D multi-atlas registration	A	Endo+Epi
NTUST, Taiwan (Wang et al., 2012)	2D clustering and motion	A	Endo+Epi
SBIA*, USA (Ou et al., 2011)	3D multi-atlas registration	A	Endo+Epi
BIT-UPM, Spain (Maier et al., 2012)	4D watershed graphcut segmentation	SA	Endo
GEWU*, Canada (Nambakhsh et al., 2013)	3D distribution matching prior	SA	Endo
ICL, UK (Bai et al., 2013)	3D multi-atlas registration	SA	Endo+Epi
LITIS, France (Grosgeorge et al., 2013)	2D shape prior graphcut segmentation	SA	Endo+Epi

Table 1: List of challengers. A: Automatic, SA: Semi-automatic. *Team not present at the workshop for the on-site challenge.

an atlas selection that chooses the best suited atlases for a particular unseen image slice. For this matter, a multi-label ranking criterion (Cardoso et al., 2013), based on the local normalized cross correlation, is used to select the best 10% atlases for label fusion. As a third step, all the label images are affinely aligned to the estimated rough segmentation, and the transformed label images are fused through majority voting. The newly obtained mask is used to remove surrounding structures in the final non-rigid registration step. The label images are non-rigidly transformed and fused using the same multi-label fusion algorithm. Typical computation time per patient is 12 min on a PC with a 2.13 GHz quad-core processor.

6.2. NTUST (C.-W. Wang et al.)

The principle of this automatic method is to use motion to detect the LV and the RV. The endocardium contour is segmented first on all images thanks to a binarization using the isodata algorithm, and cleaned up with morphological operations. Then the image sequence (denoted S) where ventricles are observed to be the largest of all slice levels is identified (empirically determined to be at the 4th slice level). For this S sequence, an exclusive or between all binarized images yields a motion map, and repeated motion maps are generated by overlapping consecutive motion maps. Then, the repeated motion map allows to select the two largest connected components from the binary image, which are the LV cavity on the right and the RV cavity on the left. The LV components of the S sequence are used to find the LV contours in other slice levels, by selecting components with the largest overlap. The RV endocardial contours over the remaining slices are found similarly, by finding components with a large overlap with the RV cavities of the S sequence and with a low overlap with the LV cavities of the S sequence. A dilatation of the endocardium contour allows to obtain the epicardium contour. Typical computation time per patient is 90.3 sec on a PC with a 3.1 GHz dual core processor.

6.3. SBIA (Y. Ou et al.)

This team has designed a fully automatic iterative segmentation framework based on multi-atlas registration and label fusion (Y.Ou et al., 2012). Manual segmentations of RV of the atlases were deformably registered onto the target image space using an attribute-based general-purpose non-rigid registration algorithm (Ou et al., 2011). A weighted majority voting strategy, which assigns higher weights to atlases locally more similar to the target image, is used for the label fusion. Note that there is no atlas selection. Within the iterative framework, the initial segmentation is used for a) restricting the focus area to the vicinity of the target anatomy, i.e. RV, and b) selecting a subset of atlases that are globally more similar to the target image within this restricted area, prior to a second round of registrations. In this way, the negative effects of large variations in images, mainly due to differences in field-of-view and/or the anatomic variability of structures surrounding the RV, have been partially reduced. In practice, the method converged to a stable final RV segmentation at the end of two iterations.

The multi-atlas segmentation framework has several hyperparameters. They are: the choice of registration algorithms, the weight for the smoothness of registration, the number of atlases to be used, the choice of atlas selection and label fusion strategies. A general-purpose attribute-based image registration algorithm with the default weight for registration smoothness is chosen here. The whole training set is used as atlases. Atlas-to-target registration each takes 2-3 minutes and the final label fusion takes around 10 seconds on a Linux OS with 2.8 GHz dual core CPU.

Team		Contours	User input
ICL	SA	Endo+Epi	5 landmarks per volume
LITIS	SA	Endo+Epi	2 landmarks per image
BIT-UPM	SA	Endo	rough contouring of 4 to 5 2D slices of ED phase
GEWU	SA	Endo	1 landmark per patient

Table 2: Amount of user input for semi-automatic (SA) methods

6.4. BIT-UPM (O. Maier et al.)

The proposed 4D semi-automatic segmentation approach is based on regions resulting from a watershed filter, merged through a graph cuts strategy (Maier et al., 2012). The watershed filtering is a popular solution to reduce the size of the graph, as voxel-based graph cuts are known to be memory consuming.

The user is required to trace a contour inside of the RV wall in four or five 2D slices of the ED phase (as shown in Table 2). The manual delineation inside the RV wall is dilated and eroded to create background and foreground markers respectively, and then propagated forward and backward along the temporal dimension, exploiting the cardiac cycle symmetry. Next, the 4D volume is pre-segmented into many small regions using the watershed transform. Finally, these regions are merged using 4D graph-cuts with an intensity based boundary-term. This approach extends the works of (Li et al., 2004) and (Stawiaski et al., 2008) to the fourth dimension. Typical cardiac MRI volumes exhibit a significant slice-to-slice discontinuity, because of the shift between two adjacent 2D slices (caused by breathing artifacts) and the large distance between two slices (8.4 mm) , while the temporal discontinuity is less obvious. Whereas a 3D GC approach might have failed to segment two neighboring slices, the proposed 4D GC approach takes advantage of the temporal consistency to impose a correct cut in the spatial dimension. A complete 4D segmentation of the RV is thus obtained in a single step. The method shows a strong robustness: since the approach is prior-free, it is suitable for any pathological cases and accounts for differences in MRI volumes originating from scanners and acquisition protocols. The hyperparameters concern the foreground and background marker extraction (three parameters linked to dilatation and erosion) and the graph cost function. They have been fixed once and for all on the training set. Thanks to the robust process to build background and foreground markers, results are little influenced by variable manual delineation, as shown in the study of inter and intra-observer variability presented in (Maier et al., 2012). The method medium runtime is 2 min 15 sec per patient on a 2.2GHz quadcore PC. Manual interaction requires an additional time of around 2 min.

6.5. GEWU (C. Nambakhsh et al.)

This method is a 3D segmentation via convex relaxation and distribution matching. The algorithm requires a single subject for training and a very simple user input, which amounts to one click at about centroid of LV in one of the 2D slices (as seen in Table 2). The RV endocardial contour is sought following the optimization of a functional containing shape as well as intensity priors, each based on a distribution matching measure, namely the Bhattacharyya measure. The shape prior evaluates the conformity between the distributions of some distance/angle features within the target right ventricular region (between RV contour points and the LV centroid) and fixed model distributions learned a priori from a single training subject. The intensity prior ensures that the image distribution within the target region most closely matches a model learned interactively from user inputs. These priors are used in conjunction with a standard total variation term, which regularizes the segmentation boundaries and attract them towards strong image edges. The overall functional is optimized with a convex relaxation technique.

The method involves the parameters balancing the contribution of the shape terms in the overall functional (four parameters involved) and two parameters for computing the distributions, one is the kernel width (a standard parameter in kernel density estimates based on the Gaussian kernel) and the other is the number of bins. The proposed algorithm relaxes the need of costly pose estimation (or registration) procedures and large, manually-segmented training sets. Furthermore, unlike related graph-cut approaches, it can be parallelized. The parallelized implementation on a graphics processing unit (GPU) demonstrates that the proposed algorithm requires about 5 seconds for a typical cardiac MRI volume.

6.6. ICL (W. Bai et al.)

This team presents a 3D multi-atlas based segmentation method which labels the RV myocardium and blood pool by ensembling opinions from multiple atlases. It only requires an initial input in the form of a few landmarks per volume (typically 5 landmarks, as specified in Table 2). Each atlas is aligned with the target image using landmark-based affine registration, followed by B-spline non-rigid image registration. In order to estimate the label at a target voxel x, the labels from the atlas voxels are combined using local weighted label fusion and defined as:

$$\tilde{L}(x) = \arg\max_{l} \sum_{n=1}^{N} \sum_{\Delta x \in S} P(I(x)|I_n(x + \Delta x)) \cdot P(L(x) = l|L_n(x + \Delta x)) \quad (5)$$

where N denotes the number of atlases, S denotes a search volume centered at voxel x. The first weight term $P(I(x)|I_n(x + \Delta x))$ is determined by the intensity similarity between the target voxel x and the atlas voxel $x + \Delta x$ and the second weight term $P(L(x) = l|L_n(x + \Delta x))$ is determined by the distance between the target and atlas voxels (Bai et al., 2012, 2013). An atlas voxel with a similar intensity to the target voxel and close to it will have a higher impact in determining its label than an atlas voxel less similar or far away from the target voxel. The label with the highest summed weight will be assigned to the target voxel. Regarding the parameter, both the intensity similarity weight and the distance weight are modelled using the Gaussian distribution. Main parameters of the method are thus the bandwidth of the Gaussian kernels, which are tuned on a small set of training images. Typical computation time per patient is 5 min with a parallel run on a 32-core computing server.

6.7. LITIS (D. Grosgeorge et al.)

This semi-automatic method is based on the 2D graph cut segmentation framework and uses a shape prior to guide the segmentation process. Each endocardial contour of the training set is transformed into a signed distance map (Tsai et al., 2003; Grosgeorge et al., 2013) and rigidly aligned on an arbitrary reference shape. All training shapes are averaged into a mean shape. Main variation axis of the endocardium are obtained via a PCA performed on the set of centered endocardial shapes. A single prior map is then derived from the PCA: areas of variations of the mean shape are first identified by generating several highly deformed shape instances for each variation axis and combined with the mean shape distance map, to form a single map. This endocardial shape prior is incorporated into the graph cut segmentation framework (Boykov and Jolly, 2001). The cost function of the graph classically includes a region (intensity-based) term and the boundary (regularization) term. In this approach, the shape prior contributes to both terms. The prior also allows to define object and background areas as hard constraints, and yields a probability model computed from the histogram of the mean shape, used in the region term. The image to be segmented is affinely registered to the shape model thanks to user input (two anatomical landmarks on the junction of the interventricular septum, as specified in Table 2). The prior-based graph cut approach allows to obtain the endocardium, and a combination of dilatations allows to obtain the epicardium. Parameters of the methods include the weighting of the graph cost terms, fixed using the training set, and the number of shape models built according to different the slice levels (6 for ED, 5 for ES). The graph cut framework is computationally efficient in 2D: typical computation time per patient is 45 sec on a PC laptop with 2.8 GHz processor.

7. Results and analysis

The results presented in this section have been obtained during Phase 2 (Test1) and Phase 3 (Test2). In particular, results obtained on Test2 have been obtained during an on-site competition, by the algorithms presented at the workshop. The method performance on the Training set was not evaluated, since this set of images was given only for training and parameter tuning purposes. As specified in Section 4.1, the 48 patients have been equally divided into the three datasets, yielding a number of 243 images for the Training set, 248 for Test1 and 252 for Test2. We denote by Test set the set of both Test1 and Test2.

7.1. Endocardium and epicardium segmentation accuracy

As a preamble, we have performed an inter-expert variability study in order to better interpret the obtained DM values. This study is expected to provide an indication of an acceptable accuracy for a (semi-)automated method. All ED phases from the Test1 set were delineated by another radiologist, using the same guidelines as specified in Section 4.2. Agreement between contours is measured with a DM equal to 0.90 ± 0.10. From Table 3, it can be seen that the DM values for endocardium ranges from 0.55 to 0.81, with high standard deviation, showing that performance may vary much from one patient to another. The best DM obtained by enrolled methods being around 0.8, one can say that room for improvement is left, as compared to inter-expert variability (0.90 ± 0.10). Nevertheless, the comparison to the companion task of LV segmentation, for which the state-of-the-art DM is about 0.8[4] to 0.9 (Bai et al., 2013; Zhuang et al., 2010) shows that the best DM obtained here for the RV is comparable to the state-of-the-art for LV. Figures 5, 6 and 7 help to visually grasp the difference between DM of values 0.8 and 0.9.

From Table 3, one can see that minimal HD values are close to 1 cm. When compared to the HD value obtained with the inter-expert study, i.e. 5.02 ± 2.87 mm, this value may seem large when considered alone, especially in comparison to the RV size. A HD value should actually be examined along with the corresponding DM value. Two images can have similar HD and different DM, see for example in Fig. 5, in which the LITIS-basal image (first row) and the SBIA-mid image (second row) have similar HD values (8.40 and 8.08 mm resp.), and a difference of 0.09 in their DM. The error between contours in SBIA is rather global (over the whole contour) whereas it is local in the LITIS image. In Fig. 6, mid slices in CMIC and LITIS have similar DM (0.94 and 0.95 resp.) but different HD values 5.18 and 7.17 mm. When anticipating about post-processing manual corrections, HD gives an idea of the correction amplitude, and DM of the amount of correction needed.

Out of the seven methods, five of them have reported segmentation for epicardium. Results reported in Table 4 shows that a Dice Metric can reach up to 0.82 (resp. 0.77) for automatic method (CMIC) and 0.83 (resp. 0.85) for semi-automatic ones (ICL, LITIS) on Test1 and Test2 respectively. Even if the segmentation of the epicardium might seem more challenging in terms of image content than the endocardium segmentation (the RV has a very thin wall, reaching the limit of MRI spatial resolution), comparison between epicardium and endocardium results show that they reach comparable accuracy, as seen Fig. 4. The quality of the epicardium as compared to the segmentation complexity may be due to the fact that no method segment the epicardium directly: all of them either apply a model (for atlas-based methods) or deduce the epicardium contour from the endocardium one. The superiority of the quality of epicardium segmentation was found significant with a one-tailed unpaired t-test only for the LITIS ($P < 0.01$).

[4] In the SATA Segmentation Challenge mentioned in the Introduction, the best-ranking DM for LV myocardium is about 0.8.

A separate analysis for ED and ES image (as shown in Fig. 2 for the endocardium and Fig. 3 for the epicardium) shows that segmentation results are superior for ED images than for ES images, for all methods: ED images are easier to process, as the heart is then the most dilated. ES images are also fuzzier because of partial volume effect. The difference of performance between ED and ES ranges from 0.05 up to 0.17 depending on the method and is shown to be significant ($P < 0.01$ for all methods thanks to an one-tailed, unpaired t-test). Note how the distribution is tightened around the median value for certain methods (BIT-UPM, ICL, LITIS), which indicates a stable behavior of the method.

It can be seen from Fig. 5, 6 and 7 that the erroneous behaviour of all methods depends on the slice level. We have thus performed a quantitative analysis of errors along the longitudinal axis of the RV, for all patients. Each volume having a different number of slices (ranging from 6 to 12, with a mean value of 8.94±1.53 for ED volumes), the DM values obtained for each slice have been interpolated over 12 values, so as to allow for a comparison between patients. Fig. 8 shows, for all methods, the average Dice metric in function of three normalized slice levels: basal, mid-ventricular and apical. It reveals that the error increases as most apical slices are processed. For the endocardial contour for example, the DM decreases by around 0.20 from base to apex: most basal slices have a score of 0.91 (when averaged over the three best methods CMIC, ICL, LITIS), whereas most apical ones have a score of 0.73. This indicates that the improvement of segmentation accuracy could be searched in apical slices, maybe with an emphasis of the model over the image content for these slices. Error on apical slices has eventually little influence on the volume computation but it can be a limiting factor in other fields such as studies of the fiber structure.

		Test1		Test2	
		DM	HD (mm)	DM	HD (mm)
CMIC	A	**0.78 ± 0.23**	**10.51 ± 9.17**	**0.73 ± 0.27**	**12.50 ± 10.95**
NTUST	A	0.57 ± 0.33	28.44 ± 23.57	0.61 ± 0.34	22.20 ± 21.74
SBIA	A	0.55 ± 0.32	23.16 ± 19.86	0.61 ± 0.29	15.08 ± 8.91
BIT-UPM	SA	**0.80 ± 0.19**	11.15 ± 6.62	0.77 ± 0.24	9.79 ± 5.38
GEWU	SA	0.59 ± 0.24	20.21 ± 9.72	0.56 ± 0.24	22.21 ± 9.69
ICL	SA	0.78 ± 0.20	**9.26 ± 4.93**	0.76 ± 0.23	9.77 ±5.59
LITIS	SA	0.76 ± 0.20	9.97 ± 5.49	**0.81 ± 0.16**	**7.28 ± 3.58**

Table 3: Endocardium segmentation: mean values (± standard deviation) of Dice Metric (DM) and Hausdorff Distance (HD) averaged over ED and ES. A: Automatic, SA: Semi-automatic

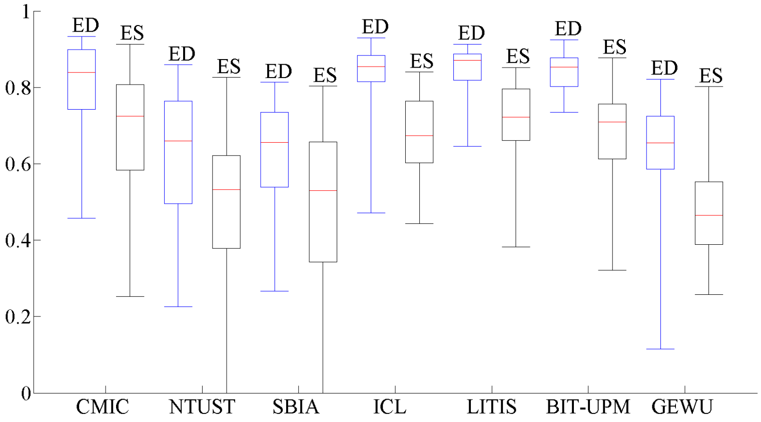


Figure 2: Endocardium segmentation: median DM value obtained for the Test patients. The median is the middle bar, in red. The box indicates the lower quartile (splits 25% of lowest data) and the upper quartile (splits 75% of highest data). The whiskers are the maximum and minimum values.

		Test1		Test2	
		DM	HD (mm)	DM	HD (mm)
CMIC	A	**0.82 ± 0.19**	**10.94 ± 8.32**	**0.77 ± 0.24**	**12.70 ± 10.44**
NTUST	A	0.62 ± 0.35	26.71 ± 22.90	0.64 ± 0.35	22.14 ± 21.61
SBIA	A	0.58 ± 0.29	22.53 ± 18.06	0.68 ± 0.25	15.17 ± 8.88
ICL	SA	**0.83 ± 0.14**	**9.64 ± 4.95**	0.80 ± 0.18	10.34 ± 5.41
LITIS	SA	0.82 ± 0.13	10.40 ± 5.45	**0.85 ± 0.11**	**8.32 ± 3.70**

Table 4: Epicardium segmentation: mean values (± standard deviation) of Dice Metric (DM) and Hausdorff Distance (HD) averaged over ED and ES. A: Automatic, SA: Semi-automatic

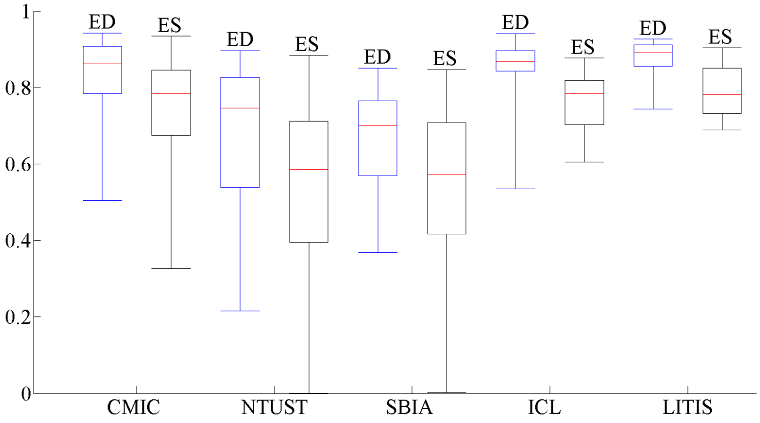


Figure 3: Epicardium segmentation: median DM value obtained for the Test set. The median is the middle bar, in red. The box indicates the lower quartile (splits 25% of lowest data) and the upper quartile (splits 75% of highest data). The whiskers are the maximum and minimum values.

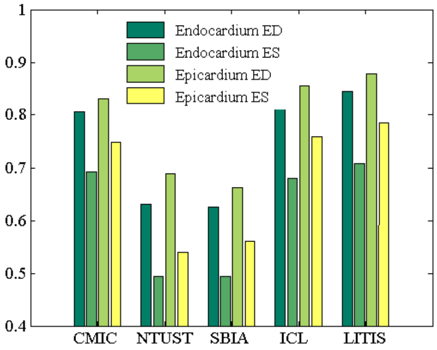


Figure 4: Endocardium vs. epicardium segmentation: mean DM value for the Test set.

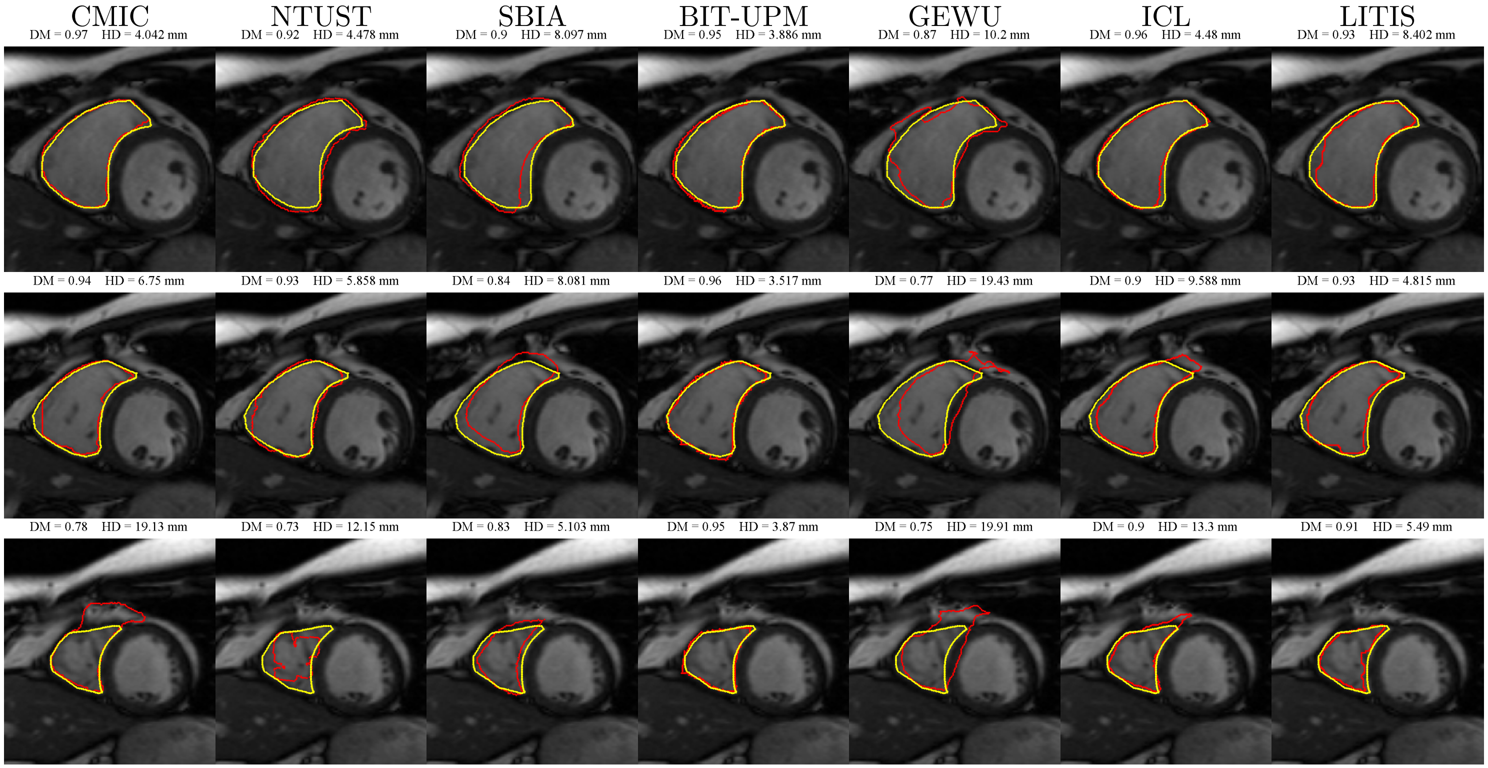


Figure 5: Endocardial contours at ED on one patient (P#33 from Test2) for all methods, from selected basal, mid-ventricular and apical slices (from top to bottom). Manual contours are shown in yellow, automatic contours in red. Corresponding DM and HD values are provided for each image.

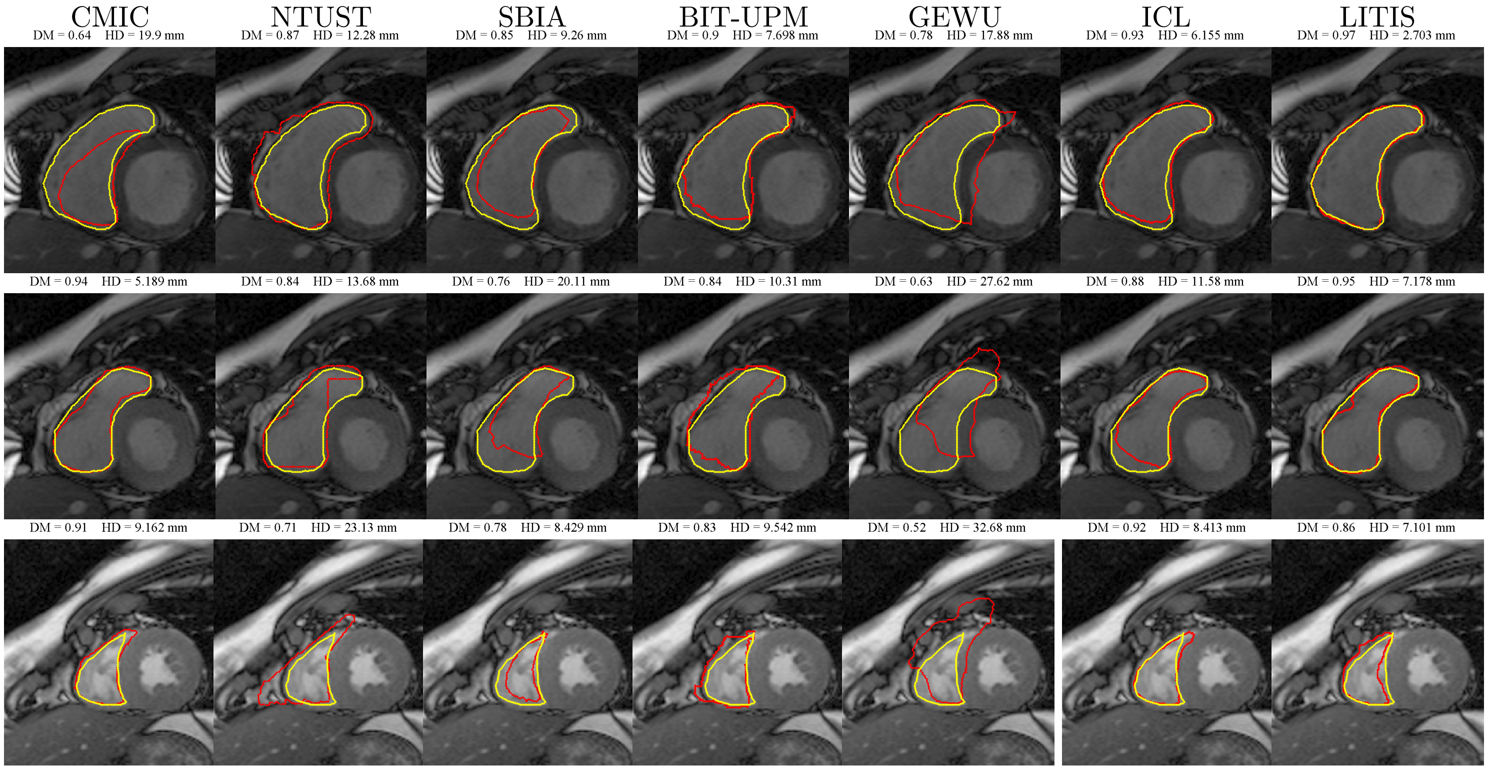


Figure 6: Endocardial contours at ES on one patient (P#42 from Test2) for all methods from selected basal, mid-ventricular and apical slices (from top to bottom). Manual contours are shown in yellow, automatic contours in red. Corresponding DM and HD values are provided for each image. For visualization purposes, the image contrast has been modified.

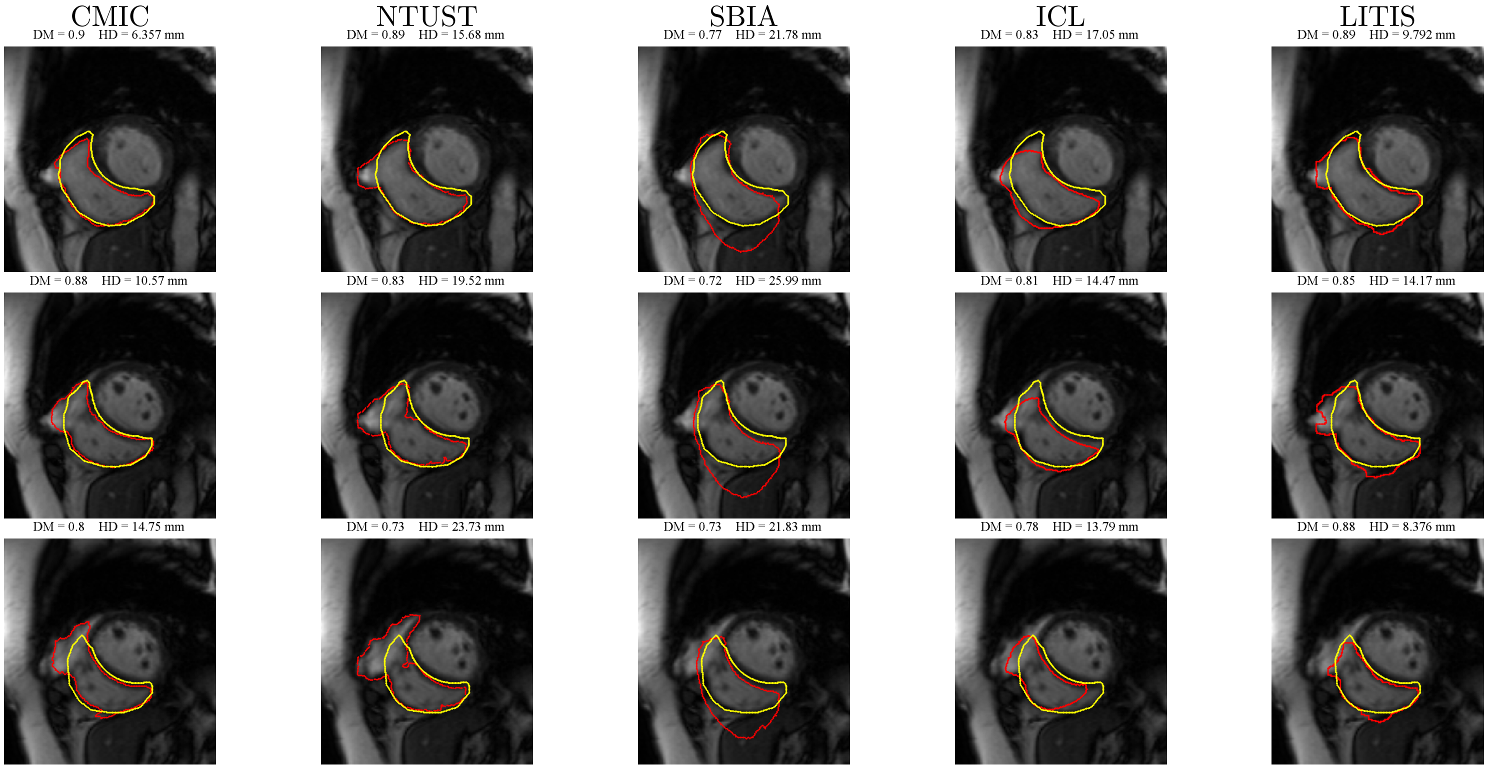
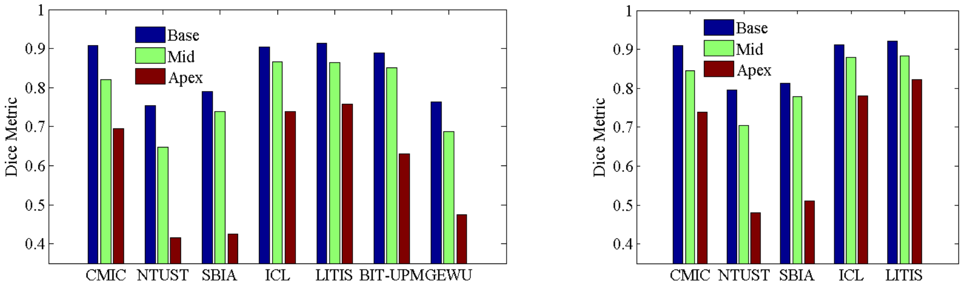


Figure 7: Epicardial contours at ED on one patient (P#38 from Test2) for all methods. Selected basal, mid-ventricular and apical slices, from top to bottom. Manual contours are shown in yellow, automatic contours in red. Corresponding DM and HD values are provided for each image.



(a) Endocardium

(b) Epicardium

Figure 8: Longitudinal distribution of DM values for all algorithms on the Test set for (a) endocardium and (b) epicardium segmentation.

7.2. Clinical performance

Endocardial volumes at ED and ES are computed as the sum of all endocardial areas enclosed by the contours, multiplied by a constant. Even if the comparison between volumes is conventionally made via correlation and linear regression analysis, these figures should be handled with caution as ventricular volumes measurements, as sum of areas, may be subject to compensation of contouring errors. More importantly, the correlation coefficient resulting from regression analysis is not directly related to the accuracy of the segmentation: very accurate segmentations will result in very high volume correlation, but not vice versa. The good correlation values between manual and automatic contours reported in Table 5 (coefficient R can reach up to 0.99 for semi-automatic methods (BIT-UPM) and 0.93 for automatic methods (CMIC)) show that automated contours behave similarly to manual contours.

Based on ventricular volumes, the ejection fraction and ventricular mass are computed. The analysis of ejection fraction and ventricular mass is a bit different from the volume values. As the EF is a ratio and vm is a difference, any existing, constant bias in the volume assessment may cause EF errors and vm errors to decrease. Yet EF correlation values is not that satisfying and there is a non-negligible fixed offset in the Bland-Altman plot, as shown in Fig. 9 and 10: the mean differences (red line) exhibit absolute values ranging from 0.06 to 0.19, with an average of 0.10. A two-tailed paired Student's t-test allowed to determine that there are indeed significant differences between manual and automated measurement of the EF ($P < 0.01$) for some of the

		R (ED)	R (ES)
CMIC	A	0.93	0.93
NTUST	A	0.71	0.78
SBIA	A	0.63	0.69
BIT-UPM	SA	0.99	0.97
GEWU	SA	0.81	0.81
ICL	SA	0.98	0.98
LITIS	SA	0.95	0.90

Table 5: Correlation coefficient R for RV volumes at ED and ES. A: Automatic, SA: Semi-automatic

teams (CMIC, ICL, GEWU). The same remarks holds for vm. Fig. 11 shows quite deceptive correlation and regression values. When performing the paired t-test, the null hypothesis was not rejected for only one team (LITIS); for the remaining teams, vm values were found significantly different ($P < 0.01$) from reference values. Room for improvement is thus left for the computation of clinical values. In addition to the significance test, we want to know if the estimated EF or vm values reach intra-expert variability, in order to assess whether they are clinically acceptable. We have thus compared them to intra-expert variability values obtained from (Caudron et al., 2012), where the EF Bland Altman plots reveal a bias close to zero and the 95% limits of agreement ($\pm 2\sigma$) are ±0.10 (Fig. 3 of (Caudron et al., 2012)). Looking at Fig. 9 and 10, one can see that there exists a non zero bias in general with the 95% limits closer to ±0.20. The same conclusion holds for the vm. The evaluation of EF and vm by (semi-)automated, although encouraging, cannot be fully satisfying in this study.

8. Discussion and conclusion

Let us now return to our introduction questions: what accuracy can be expected from semi-automated and automated algorithms for RV endocardium and epicardium segmentation? An overall reasonable accuracy 80% (in terms of DM) should be expected. Epicardium compared well to endocardium segmentation, with equivalent or even better results. ES phases exhibited more errors that ED phases. Apical slices were found to be difficult to process, with a DM of 0.62 for the most apical slices, while accuracy of basal slices reached 0.91. Clinical evaluation of the methods showed that ejection frac-

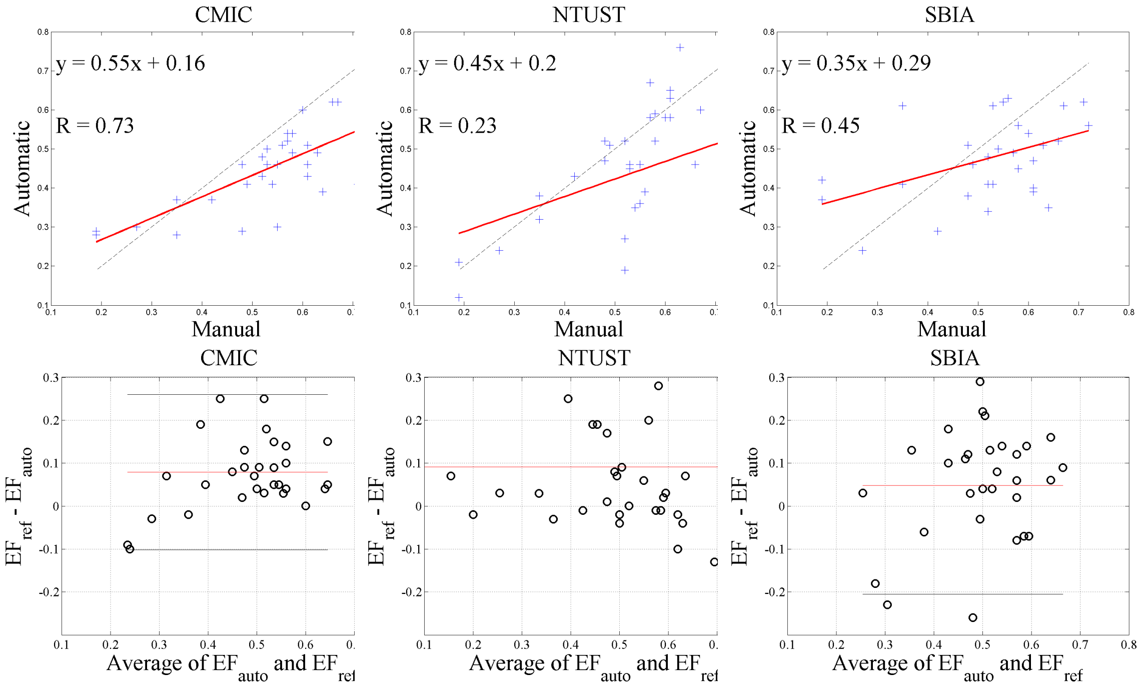


Figure 9: Analysis of EF for automatic methods on the Test set. Correlation coefficient (R). Linear regression: the black dotted line is the identity function. Bland-Altman plots: black lines indicate the 95% limits of agreement ($\pm 2\sigma$).

tion and ventricular mass were in some cases correctly estimated though room for improvement is left.

Second question was to know how do automated algorithms compare to semi-automated ones. The method presented by CMIC is a proof that automated can reach an accuracy very close to that of semi-automated algorithms. Difference with the closest semi-automatic methods is evaluated with a paired t-test: compared to BIT-UPM's DM on the endocardium contour, no significant difference was found ($P \approx 0.41$) ; yet CMIC's results were found significantly different from the ones of ICL and LITIS ($P < 0.05$) (same conclusion holds for epicardium contours). A fair comparison should include not only accuracy, but also complexity or computation time, which is an important matter for clinical use of the methods, and requested amount of user interaction. From the clinician point of view, a fast and semi-automatic method, involving for example the identification of landmarks such as the RV attachment to the LV, the triscupid or pulmonary valve, would be preferred to a fully automatic, lengthy one. Although computation times may not be fully compared, it seems that the good performance of CMIC algorithm is obtained at the expense of a higher computation time (see Section 6), in relation with the speed – degree of automation tradeoff. Although real time is not requested, the running time should be limited to a few minutes, which is the case for all methods here. For semi-automated methods, the amount of user interaction differs depending on the methods, from a few clicks per patient to full manual segmentation of some images, as shown in Table 2. Note that the robustness of the methods to user variability has not been assessed here.

As 3D methods are the state-of-the-art in many segmentation domain, one legimate question was if they really were the most appropriate for the task of RV segmentation. In our cardiac MR data, space between slices and slice thickness are quite large (8.4 mm and 7 mm resp.), and differ from the order of magnitude of spatial resolution within the image (1.6 mm per pixel). Most of the imaging centres still acquire cardiac MR data with 8 or even 10 mm slice thickness. This is still the main stream, and it has some advantages: the 2D short-axis acquisition is accompanied by 4-chamber and long-axis imaging acquisitions which allow for an easy identification of the pulmonary and tricuspid valves, thus preventing to include "out-of-RV" volumes (such as pulmonary artery or atrial volume). However, it seems that some groups have started to work on 3D isotropic MR images: voxel reported to be 1.4 × 1.4 × 1.4 mm in (Rajchl et al., 2014), 2.5 × 2.5 × 2.5 mm in (Uribe et al., 2007), 2 × 2 × 4 mm (reconstructed to 2 × 2 × 2 mm) in (Dawes et al., 2013). Authors of (Uribe et al., 2007) mentioned the drawback of the 3D SSFP sequence is that "current methods, even those that use undersampling techniques, involve breath-holding for periods that are too long for many patients." This might be the reason why in some studies 3D cine image acquisition is restrained to healthy volunteers (Uribe et al., 2007; Dawes et al., 2013). Also, it sacrifices some signal-to-noise-ratio. Definitely, isotropic imaging has some drawbacks which makes many radiologists still subscribe to the 2D imaging sequence. Our cardiac MR data may not be fully considered as 3D data, due to an anisotropic resolution and to a longitudinal shift between consecutive images since every phase image is acquired during a different breathhold (Attili et al., 2010). We can expect that, with the advances of MR imaging technique, it would become possible that cardiac images (for both LV and RV) can become truly 3D, like brain images. But for now, most of groups still process 2D cardiac image stacks and

3D segmentation methods may not be the best adapted to this data. This is also demonstrated in the paper empirically: some 2D methods (LITIS, BIT-UPM) exhibit good results.

At last, and as a conclusion question, we wanted to know what type of methods performs best, although this is an open and difficult question, and we are fully aware that the answer is limited by the framework of this challenge, and by the sample of methods that answered back to our challenge proposal. In the following, we summarize advantages and drawbacks of the methods.

Atlas-based methods (CMIC, ICL, SBIA). Atlas-based methods incorporate prior anatomical knowledge from multiple atlases. With good target-to-atlas registration, the resulting segmentation can be quite robust and accurate. For example, multi-atlas segmentation has been successfully applied to brain image segmentation in recent years and achieved very good results. Multi-atlas segmentation methods consist of two steps, namely image registration and label fusion. In this challenge, three methods are atlas-based, which differ on these two points, and also on how they handle the challenges arising from the direct registration of the entire cardiac images. ICL defines a ROI using the landmarks and performs registration only in the ROI for the RV, CMIC has a pre-processing step to remove non-relevant structures before the registration, and SBIA directly registers the whole atlas and the whole target images, which may account for the relatively low segmentation accuracy in the SBIA approach. Regarding the deformable registration algorithms, they all used free form deformation (FFD) as the transformation model, but ICL and CMIC used normalized mutual information (NMI) with continuous optimization strategies, whereas SBIA used a texture-attribute-based similarity metric with a discrete optimization strategy. ICL and CMIC use local weighting, where the label from each atlas voxel has its own weight: for ICL, the weight is determined by the intensity similarity to the target voxel and for CMIC, it is based on local normalized cross correlation. On the contrary, SBIA uses global weighting, where all the voxels from one atlas have the same weight, determined by the similarity between this atlas and the target image. The limitations of atlas-based methods include a high computational cost, which is associated with the registration between the target image and multiple atlases, and the dependence of the results on the quality of the atlas set. While CMIC and ICL exhibit some of the best results of this challenge, SBIA's results leave room for improvement. While being general, the SBIA framework was not specifically designed, and hence it is not necessarily optimal for, cardiac segmentation. For example, a purely registration based approach was used. Most general-purpose registration algorithms encounter challenges when directly applied to raw cardiac MRI data, mainly due to the complications caused by many neighboring structures in the image. Future studies may need to consider shape or anatomical priors as an initialization or constraint for the registration specifically for cardiac images. Also of interest would be the optimization of hyperparameters (registration algorithms, number of selected atlases, label fusion) specifically in the context of cardiac segmentation.

Shape prior-based approaches (LITIS). Based on a statistical shape prior model, the LITIS method yields quite accurate results, with an advantageous computation time. Main drawbacks are a heavy user interaction (2 landmarks per image) and the construction of the shape prior models.

Prior-based approaches (GEWU). The GEWU algorithm uses prior knowledge in its segmentation process, but not under the form of a shape model, which thus removes the need for costly pose estimation (or registration) procedures. As it uses a single subject for training, the need for large, manually segmented training sets is also relaxed. A good property is that performance is not significantly affected by the choice of the training subject (Nambakhsh et al., 2013). Another advantage of this algorithm is its possible parallelized implementations, which makes it run in near real time on typical graphics processing units can. This can accommodate interactive scenarios, where the user can correct the results or change the inputs. A limitation is that it is not straightforward to extend this formulation to train several subjects. When a large training set is available, as in this challenge, the shape prior cannot take full advantage of the available information (unlike standard statistical shape models) because the distribution matching measure provides summarized, not comprehensive, shape information. Therefore, in cases where massive training information is available, this algorithm is not expected to outperform standard statistical shape models. The results might depend on the user input.

Image-driven approaches (BIT-UPM, NTUST). These two approaches are based on cardiac motion to compensate the lack of a priori knowledge. One clear advantage is that they do not depend on a training set. Another is that images are segmented over full cardiac cycle. Apart from that, they principle are different. The automatic method of NTUST does not include any constraint and thus might fail in some difficult cases. The BIT-UPM approach circumvents the problem of discontinuity between slices by relying on temporal coherence with a 4D approach, with a certain efficacy. The BIT-UPM framework is flexible enough to include in the future a shape prior, such as the one proposed by the LITIS, incorporated as the graph-cuts regional term. On the other hand, it requires user interaction.

Conclusion. It is difficult to conclude on the best type method for this task. It may be surmised from the results, that at the present time and for this set of data, the best performing methods are CMIC for the automatic methods, ICL, LITIS and to a lesser extent BIT-UPM for the semi-automatic methods, whose performance are comparable (BIT-UPM performance is shown to be significantly inferior to ICL and LITIS with a paired t-test on DM values ($P < 0.01$), whereas the null hypothesis cannot be rejected between ICL and LITIS). We cannot conclude on whether the best approach should be 2D, 3D or 4D, or whether it should be prior shape based or data-driven. What we can say on the other side, is that the designed algorithm should contain some kind of spatial constraint (thanks to a model or a global, temporal approach), and that hyperparameters have to be somehow optimized for the context of cardiac segmentation. There is obviously a choice to make between between computational burden (required by CMIC) or a user interaction (required by LITIS, ICL and BIT-UPM). Yet, efforts still have to be made for the segmentation accuracy to reach inter-expert variability. Clinically acceptable accuracy has not been reached.

9. Perspectives

This paper has presented the results of the Right Ventricle Segmentation Challenge, provided over 48 patients and 7 different algorithms. Today, the challenge datasets are available to the MICCAI community, in order to encourage future investigations in this field. We hope these datasets will become reference datasets, and serve as standard performance tools for future segmentation methods. Ever since the challenge was finished, the datasets have been requested and downloaded around thirty times by research teams from all over the world, resulting in new publications already (Labrador et al., 2013; Ringenberg et al., 2014).

Future works concern the investigation of more reliable ground truth estimation, as manual segmentation is known to be quite observer-dependent. A new estimation of the reference segmentation could be generated, based for example on the well-known STAPLE algorithm (Warfield et al., 2004) or the more recent multi-STEPS approach (Cardoso et al., 2012). The STAPLE algorithm estimates a ground truth from the collection of rater segmentation results based on the Expectation-Maximization algorithm. Raters, in this case, can be a collection of automated segmentation results, manual assessment, or a mixture between the two. Other recent approaches regarding the evaluation of segmentation methods without ground truth will be profitably investigated (Kohlberger et al., 2012; Lebenberg et al., 2012).

Other perspectives also include investigation on the data. Following the standard protocol in use today, the RV is imaged based on the short-axis view perpendicular to the left ventricular long axis. Whereas the short-axis view is particularly well-suited to the LV, there might be better, alternative imaging orientations for right ventricular analysis. In recent research, it has been shown that axial slices (Attili et al., 2010) or 4-chamber view (Caudron et al., 2011) could be fruitfully used to evaluate the RV EF. Nonetheless, they require the patient to remain 15 additional minutes in the MR scanner and to perform around ten apneas for the RV only, which limits the use of these acquisitions in practice. The short-axis view should remain the standard protocol in the absence of consensus on optimal imaging orientation, with the advantage to allow functional assessment of both ventricles on the same slice stack. In this respect, a next and natural step after two challenges on the LV segmentation and one on the RV segmentation in MRI, would be the evaluation of the joint segmentation of both cardiac ventricles, whose outcome is known to be useful in the clinic.

Acknowledgments

The anonymous reviewers are acknowledged for their detailed, qualified and constructive reviews. W. Bai is supported by the BHF grant RG/10/11/28457 and NIHR Biomedical Research Centre funding. MA. Zuluaga and MJ Cardoso are funded by an EPSRC grant (EP/H046410/1). S. Ourselin receives funding from the EPSRC (EP/H046410/1, EP/J020990/1, EP/K005278), the MRC (MR/J01107X/1), the EU-FP7 project VPH-DARE@IT (FP7-ICT-2011-9-601055), the NIHR Biomedical Research Unit (Dementia) at UCL and the National Institute for Health Research University College London Hospitals Biomedical Research Centre (NIHR BRC UCLH/UCL High Impact Initiative). H.-C. Chen and C.-W. Peng are supported by the National Science Council of Taiwan with the project number NSC101-2628-E-011-006-MY3. O.M.O. Maier, D. Jimenez-Carretero, A. Santos and M.J. Ledesma-Carbayo's work is partially funded by Spanish Ministry of Science and Innovation through CDTI CENIT (AMIT) and projects TEC2010-21619-C04-03 & TEC2011-28972- C02-02; Comunidad de Madrid (ARTEMIS S2009/DPI-1802), and the European Funds (FEDER) and FP7-PEOPLE-2010-IRSES-269300 TAHITI.

References

Abi-Nahed, J., Jolly, M.-P., Yang, G.-Z., 2006. Robust active shape models: A robust, generic and simple automatic segmentation tool. In: Proc. of MICCAI. No. 2. pp. 1–8.

Attili, A., Schuster, A., Nagel, E., Reiber, J., van der Geest, R., 2010. Quantification in cardiac MRI: advances in image acquisition and processing. Int J Cardiovasc Imaging 26 Suppl 1, 27–40.

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 (7), 1302–1315.

Bai, W., Shi, W., Wang, H., Peters, N. S., Rueckert, D., 2012. Multi-atlas based segmentation with local label fusion for right ventricle MR images. In: Right Ventricle Segmentation Challenge at MICCAI 2012.

Bogaert, J., Dymarkowski, S., Taylor, A., 2005. Clinical Cardiac MRI: With Interactive CD-ROM. Springer.

Bonnemains, L., Mandry, D., Marie, P., Micard, E., Chen, B., Vuissoz, P., 2012. Assessment of right ventricle volumes and function by cardiac MRI: quantification of the regional and global interobserver variability. Magnetic Resonance in Medicine 67 (6), 1740–6.

Boykov, Y., Jolly, M., July 2001. Interactive graph cuts for optimal boundary and region segmentation of objects in n-d images. International Conference on Computer Vision 1, 105–113.

Cardoso, M., Modat, M., Keihaninejad, S., Cash, D., Ourselin, S., 2012. Multi-STEPS: multi-label similarity and truth estimation for propagated segmentations. In: IEEE Workshop on Mathematical Methods in Biomedical Image Analysis (MMBIA). pp. 153–8.

Cardoso, M. J., Leung, K., Modat, M., Keihaninejad, S., Cash, D., Barnes, J., Fox, N. C., Ourselin, S., 2013. STEPS: Similarity and truth estimation for propagated segmentations and its application to hippocampal segmentation and brain parcelation. Medical Image Analysis 17 (6), 671–684.

Caudron, J., Fares, J., Lefebvre, V., Vivier, P.-H., Petitjean, C., Dacher, J.-N., 2011. Diagnostic accuracy and variability of three semi-quantitative methods for assessing right ventricular systolic function from cardiac MRI in patients with acquired heart disease. European Radiology 21 (10), 2111–20.

Caudron, J., Fares, J., Lefebvre, V., Vivier, P.-H., Petitjean, C., Dacher, J.-N., 2012. Cardiac MR assessment of right ventricular function in acquired heart disease: Factors of variability. Academic Radiology 19 (8), 991–1002.

Cocosco, C., Wiro, W. N., Netsch, T., Vonken, E.-J., Lund, G., Stork, A., Viergever, M., 2008. Automatic image-driven segmentation of the ventricles in cardiac cine MRI. J Magn Reson Imaging 28 (2), 366–74.

Cootes, T., Cooper, D., Taylor, C., Graham, J., 1995. Active shape models - their training and application. Computer Vision and Image Understanding 61 (1), 38–59.

Davlouros, P. A., Kilner, P. J., Hornung, T. S., Li, W., Francis, J. M., Moon, J. C., Smith, G. C., Pennell, D. J., Gatzoulis, M. A., 2002. Right ventricular function in adults with repaired tetralogy of fallot assessed with cardiovascular magnetic resonance imaging: detrimental role of right ventricular outflow aneurysms or akinesia and adverse right-to-left ventricular interaction. Journal of the American College of Cardiology 40 (11), 20442052.

Dawes, T. J. W., de Marvao, A., Keenan, N. G., ORegan, D. P., 2013. High resolution 3D cine imaging: a novel approach for automated right ventricular phenotyping. Heart 99 (suppl 2), A56.

ElBaz, M. S., Fahmy, A. S., 2012. Active shape model with inter-profile modeling paradigm for cardiac right ventricle segmentation. In: Medical Image Computing and Computer Assisted Interventions (MICCAI). Vol. 1 of LNCS 7510. pp. 691–698.

Grosgeorge, D., Petitjean, C., Caudron, J., Fares, J., Dacher, J.-N., 2011. Automatic cardiac ventricle segmentation in MR images: a validation study. International Journal of Computer Assisted Radiology and Surgery 6 (5), 573–581.

Grosgeorge, D., Petitjean, C., Dacher, J.-N., Ruan, S., 2013. Graph cut segmentation with a statistical shape model in cardiac MRI. Computer Vision and Image Understanding 117, 1027–1035.

Haddad, F., Hunt, S., Rosenthal, D., Murphy, D., 2008. Right ventricular function in cardiovascular disease, part i. Circulation 117 (11), 1436–1448.

Huttenlocher, D., Klanderman, G., Rucklidge, W., 1993. Comparing images using the hausdorff distance. IEEE Transactions on Pattern Analysis and Machine Intelligence 15 (9), 850–863.

Kirisli, H., Schaap, M., Klein, S., Neefjes, L., Weustink, A., van Walsum, T., Niessen, W., 2010. Fully automatic cardiac segmentation from 3D CTA data: a multiatlas based approach. In: Proc. SPIE. Vol. 7623. pp. 762305–9.

Kohlberger, T., Singh, V., Alvino, C., Bahlmann, C., Grady, L., 2012. Evaluating segmentation error without ground truth. In: Proc. of MICCAI. LNCS 7510. Nice, France.

Labrador, A. A., Martínez, F., Castro, E., 2013. A novel right ventricle segmentation approach from local spatio-temporal MRI information. In: Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications (CIARP), La Havana, Cuba. Vol. 8259 of LNCS. pp. 206–213.

Lebenberg, J., Buvat, I., Lalande, A., Clarysse, P., Casta, C., Cochet, A., Constantinidès, C., Cousty, J., de Cesare, A., Jehan-Besson, S., Lefort, M., Najman, L., Roullot, E., Sarry, L., Tilmant, C., Garreau, M., Frouin, F.,

2012. Nonsupervised ranking of different segmentation approaches: Application to the estimation of the left ventricular ejection fraction from cardiac cine MRI sequences. IEEE Transactions on Medical Imaging 31 (8), 1651–60.

Li, Y., Sun, J., Tang, C.-K., Shum, H.-Y., 2004. Lazy snapping. ACM Trans. Graph. 23 (3), 303–308.

Lorenz, C. H., Walker, E. S., Graham, T. P., Powers, T. A., 1995. Right ventricular performance and mass by use of cine mri late after atrial repair of transposition of the great arteries. Circulation 92, 233–239.

Lorenzo-Valdes, M., Sanchez-Ortiz, G., Elkington, A., Mohiaddin, R., Rueckert, D., September 2004. Segmentation of 4D cardiac MR images using a probabilistic atlas and the EM algorithm. Medical Image Analysis 8 (3), 255–265.

Lötjönen, J., Kivistö, S., Koikkalainen, J., Smutek, D., Lauerma, K., September 2004. Statistical shape model of atria, ventricles and epicardium from short- and long-axis MR images. Medical Image Analysis 8 (3), 371–386.

Lu, X., Wang, Y., Georgescu, B., Littman, A., Comaniciu, D., 2011. Automatic delineation of left and right ventricles in cardiac MRI sequences using a joint ventricular model. In: Functional Imaging and Modeling of the Heart (FIMH). LNCS 6666. pp. 250–258.

Mahapatra, D., Buhmann, J. M., 2013. Automatic cardiac RV segmentation using semantic information with graph cuts. In: IEEE International Symposium Biomedical Imaging (ISBI'13). pp. 1094–1097.

Maier, O., Jimenez, D., Santos, A., Ledesma-Carbayo, M., 2012. Segmentation of RV in 4D Cardiac MR Volumes using Region-Merging Graph Cuts. In: Computing in Cardiology, Krakow (Poland). IEEE, pp. 697–700.

Marcus, F. I., McKenna, W. J., Zareba, W., 2010. Diagnosis of arrhythmogenic right ventricular cardiomyopathy/dysplasia (ARVC/D). Circulation 121 (13), 1533–1541.

Mitchell, S., Lelieveldt, B., van der Geest, R., Bosch, J., Reiber, J., Sonka, M., May 2001. Multistage hybrid active appearance model matching: segmentation of left and right ventricles in cardiac MR images. IEEE Transactions on Medical Imaging 20 (5), 415–423.

Nambakhsh, C. M., Yuan, J., Punithakumar, K., Goelaa, A., Rajchl, M., Peters, T. M., Ayed, I. B., 2013. Left ventricle segmentation in MRI via convex relaxed distribution matching. Medical Image Analysis 17, 1010–1024.

Ordas, S., Boisrobert, L., Huguet, M., Frangi, A., 2003. Active shape models with invariant optimal features (IOF-ASM) - application to cardiac MRI segmentation. In: Computers in cardiology. No. 30. pp. 633–636.

Ou, Y., Sotiras, A., Paragios, N., Davatzikos, C., 2011. DRAMMS: Deformable registration via attribute matching and mutual-saliency weighting. Medical Image Analysis 15 (4), 622–639.

Peters, J., Ecabert, O., Meyer, C., Schramm, H., Kneser, R., Groth, A., Weese, J., 2007. Automatic whole heart segmentation in static magnetic resonance image volumes. In: Proc. of MICCAI. Vol. 4792 of LNCS. pp. 402–410.

Petitjean, C., Dacher, J.-N., 2011. A review of segmentation methods in short axis cardiac MR images. Medical Image Analysis 15 (2), 169–184.

Pluempitiwiriyawej, C., Moura, J., Wu, Y., Ho, C., April 2004. Cardiac MR image segmentation: quality assessment of STACS. In: IEEE International Symposium on Biomedical Imaging: Macro to Nano. No. 1 in Medical Imaging. pp. 828–831.

Prakken, N., Velthuis, B., Vonken, E.-J., Mali, W., Crame, M.-J. J., 2008. Cardiac MRI: Standardized right and left ventricular quantification by briefly coaching inexperienced personnel. The Open Magnetic Resonance Journal 1, 104–111.

Rajchl, M., Yuan, J., White, J. A., Ukwatta, E., Stirrat, J., Nambakhsh, C. M. S., Li, F. P., Peters, T. M., 2014. Interactive hierarchical-flow segmentation of scar tissue from late-enhancement cardiac MR images. IEEE Transactions on Medical Imaging 33 (1), 154–172.

Ringenberg, J., Deo, M., Devabhaktuni, V., Berenfeld, O., Boyers, P., Gold, J., 2014. Fast, accurate, and fully automatic segmentation of the right ventricle in short-axis cardiac MRI. Computerized Medical Imaging and Graphics 38 (3), 190–201.

Sermesant, M., Forest, C., Pennec, X., Delingette, H., Ayache, N., 2003. Deformable biomechanical models: Application to 4d cardiac image analysis. Medical Image Analysis 7 (4), 475 – 488.

Stawiaski, J., Decenciere, E., Bidault, F., 2008. Interactive liver tumor segmentation using graph-cuts and watershed. The MIDAS Journal - Grand Challenge Liver Tumor Segmentation (2008 MICCAI Workshop).

Sun, H., Frangi, A., Wang, H., Sukno, F., Tobon-Gomez, C., Yushkevich, P., 2010. Automatic cardiac MRI segmentation using a biventricular deformable medial model. In: Proc. of MICCAI. LNCS 6361. pp. 468–475.

Tsai, A., Yezzi, A., Wells, W., Tempany, C., Tucker, D., Fan, A., Grimson, W., Willsky, A., 2003. A shape-based approach to the segmentation of medical imagery using level sets. IEEE Transactions on Medical Imaging 22 (2), 137–154.

Uribe, S., Muthurangu, V., Boubertakh, R., Schaeffter, T., Razavi, R., Hill, D. L., Hansen, M. S., 2007. Whole-heart cine MRI using real-time respiratory self-gating. Magnetic Resonance in Medicine 57, 606–613.

Wang, C.-W., Peng, C.-W., Chen, H.-C., 2012. A Simple and Fully Automatic Right Ventricle Segmentation Method for 4-Dimensional Cardiac MR Images. In: Proc. of 3D Cardiovascular Imaging: a MICCAI segmentation challenge, Nice, France.

Warfield, S., Zou, K., Wells, W., 2004. Simultaneous truth and performance level estimation (STAPLE): an algorithm for the validation of image segmentation. IEEE Trans Med Imaging 23 (7), 903–21.

Y.Ou, J.Doshi, Erus, G., C.Davatzikos, 2012. Multi-Atlas Segmentation of the Cardiac MR Right Ventricle. In: Proc. of 3D Cardiovascular Imaging: a MICCAI segmentation challenge, Nice, France.

Zhuang, X., 2013. Challenges and methodologies of fully automatic whole heart segmentation: A review. Journal of Healthcare Engineering 4 (2).

Zhuang, X., Rhode, K., Razavi, R., Hawkes, D., Ourselin, S., 2010. A registration-based propagation framework for automatic whole heart segmentation of cardiac mri. Medical Imaging, IEEE Transactions on 29 (9), 1612–1625.

Zuluaga, M., Cardoso, M., Modat, M., Ourselin, S., 2013. Multi-atlas propagation whole heart segmentation from MRI and CTA using a local normalised correlation coefficient criterion. In: Functional Imaging and Modeling of the Heart, FIMH 2013. Vol. 7945 of LNCS. London, UK, pp. 172–180.

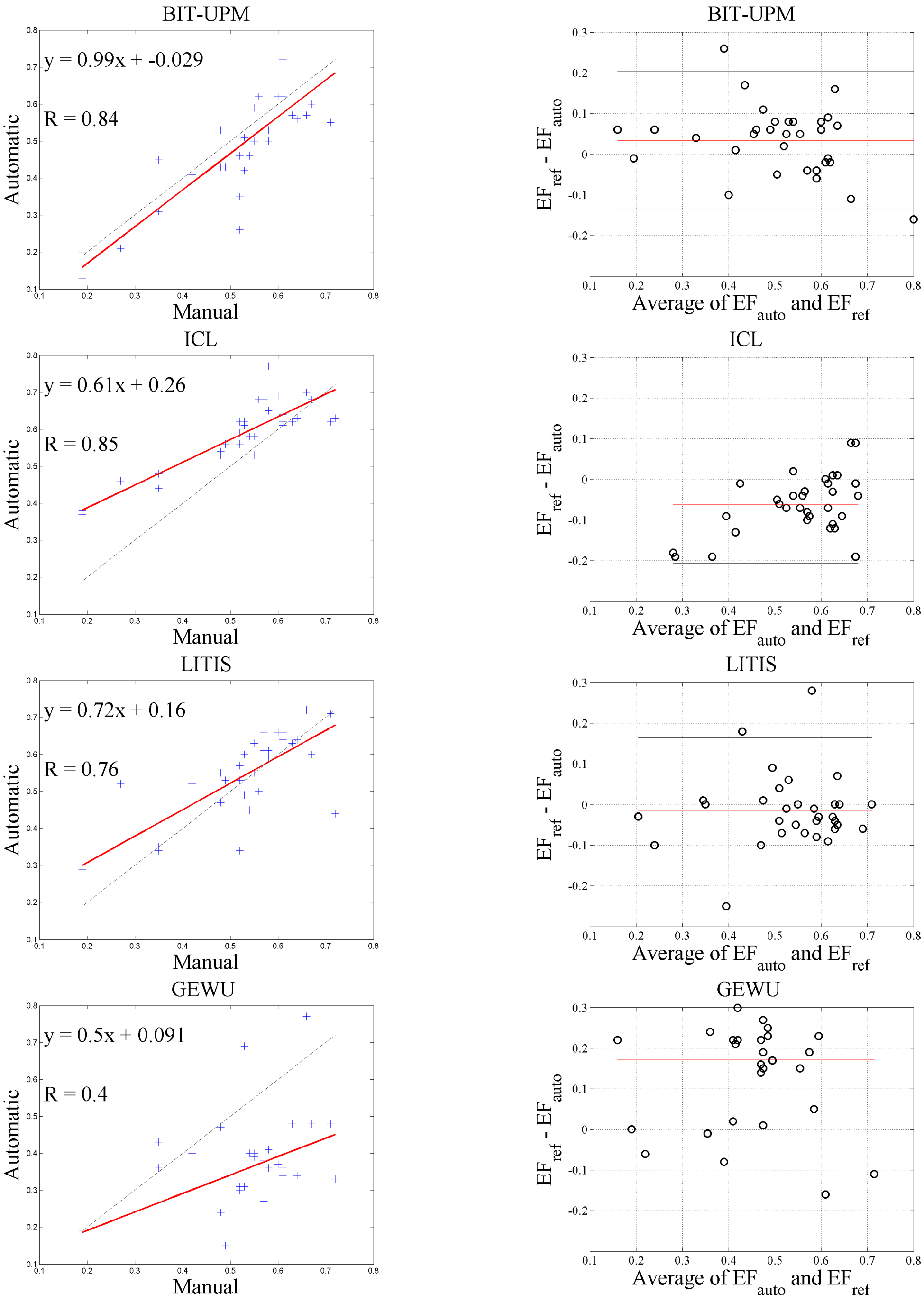


Figure 10: Analysis of EF for semi-automatic methods on the Test set. Correlation coefficient (R). Linear regression: the black dotted line is the identity function. Bland-Altman plots: black lines indicate the 95% limits of agreement ($\pm 2\sigma$).

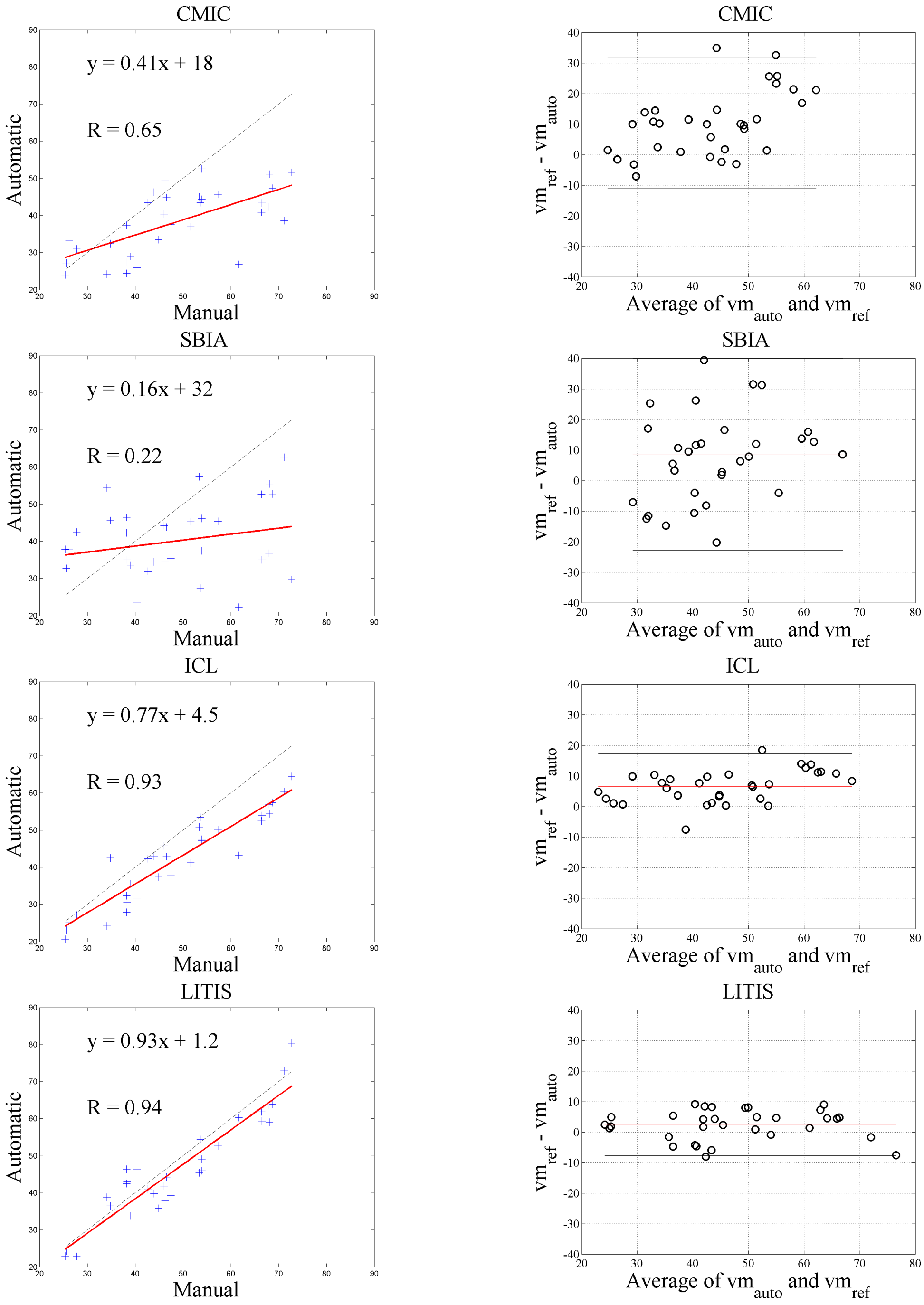


Figure 11: Analysis of ventricular mass (g) for epicardium-concerned methods on the Test set. Correlation coefficient (R). Linear regression: the black dotted line is the identity function. Bland-Altman plots: black lines indicate the 95% limits of agreement ($\pm 2\sigma$). The NTUST algorithm failed for some cases and the vm could not be computed.