Computers in Biology and Medicine 142 (2022) 105191

Contents lists available at ScienceDirect

Computers in Biology and Medicine

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

Deep learning-based whole-heart segmentation in 4D contrast-enhanced cardiac CT

Steffen Bruns [a,b,*], Jelmer M. Wolterink [c], Thomas P.W. van den Boogert [d], Jurgen H. Runge [e], Berto J. Bouma [f], José P. Henriques [d], Jan Baan [f], Max A. Viergever [g], R. Nils Planken [e], Ivana Išgum [a,b,e]

[a] Department of Biomedical Engineering and Physics, Amsterdam UMC, University of Amsterdam, Meibergdreef 9, 1105AZ, Amsterdam, the Netherlands
[b] Amsterdam Cardiovascular Sciences, Amsterdam UMC, Meibergdreef 9, 1105 AZ, Amsterdam, the Netherlands
[c] Department of Applied Mathematics, Technical Medical Centre, University of Twente, Drienerlolaan 5, 7522 NB, Enschede, the Netherlands
[d] Heart Centre, Academic Medical Centre, Amsterdam Cardiovascular Sciences, University of Amsterdam, Meibergdreef 9, 1105 AZ, Amsterdam, the Netherlands
[e] Department of Radiology and Nuclear Medicine, Amsterdam UMC, Meibergdreef 9, 1105AZ, Amsterdam, the Netherlands
[f] Department of Cardiology, Amsterdam UMC, Meibergdreef 9, 1105 AZ, Amsterdam, the Netherlands
[g] Image Sciences Institute, University Medical Center Utrecht, Heidelberglaan 100, 3584 CX, Utrecht, the Netherlands

ARTICLE INFO

Keywords:
Whole-heart segmentation
Transcatheter aortic valve implantation
Convolutional neural network
Deep learning
4D contrast-enhanced cardiac CT

ABSTRACT

Automatic cardiac chamber and left ventricular (LV) myocardium segmentation over the cardiac cycle significantly extends the utilization of contrast-enhanced cardiac CT, potentially enabling in-depth assessment of cardiac function. Therefore, we evaluate an automatic method for cardiac chamber and LV myocardium segmentation in 4D cardiac CT.

In this study, 4D contrast-enhanced cardiac CT scans of 1509 patients selected for transcatheter aortic valve implantation with 21,605 3D images, were divided into development (N = 12) and test set (N = 1497). 3D convolutional neural networks were trained with end-systolic (ES) and end-diastolic (ED) images. Dice similarity coefficient (DSC) and average symmetric surface distance (ASSD) were computed for 3D segmentations at ES and ED in the development set via cross-validation, and for 2D segmentations in four cardiac phases for 81 test set patients. Segmentation quality in the full test set of 1497 patients was assessed visually on a three-point scale per structure based on estimated overlap with the ground truth.

Automatic segmentation resulted in a mean DSC of 0.89 ± 0.10 and ASSD of 1.43 ± 1.45 mm in 12 patients in 3D, and a DSC of 0.89 ± 0.08 and ASSD of 1.86 ± 1.20 mm in 81 patients in 2D. The qualitative evaluation in the whole test set of 1497 patients showed that automatic segmentations were assigned grade 1 (clinically useful) in 98.5%, 92.2%, 83.1%, 96.3%, and 91.6% of cases for LV cavity and myocardium, right ventricle, left atrium, and right atrium.

Our automatic method using convolutional neural networks performed clinically useful segmentation across the cardiac cycle in a large set of 4D cardiac CT images, potentially enabling in-depth assessment of cardiac function.

1. Introduction

Segmentation of the cardiac chambers and left ventricular (LV) myocardium over the whole cardiac cycle enables the assessment of cardiac morphology and function. Cardiac MRI and echocardiography are standardly used to determine cardiac function. However, for a growing number of applications, 3D contrast-enhanced cardiac CT scans are also acquired over the whole cardiac cycle, resulting in 4D cardiac CT images. These have great potential for in-depth cardiac function assessment and could be used to derive the LV ejection fraction (LVEF)

* Corresponding author. Department of Biomedical Engineering and Physics, Amsterdam UMC, University of Amsterdam, Meibergdreef 9, 1105AZ, Amsterdam, the Netherlands.

E-mail addresses: s.bruns@amsterdamumc.nl (S. Bruns), j.m.wolterink@utwente.nl (J.M. Wolterink), t.p.vandenboogert@amsterdamumc.nl (T.P.W. van den Boogert), j.h.runge@amsterdamumc.nl (J.H. Runge), b.j.bouma@amsterdamumc.nl (B.J. Bouma), j.p.henriques@amsterdamumc.nl (J.P. Henriques), j.baan@amsterdamumc.nl (J. Baan), m.viergever@umcutrecht.nl (M.A. Viergever), r.n.planken@amsterdamumc.nl (R.N. Planken), i.isgum@amsterdamumc.nl (I. Išgum).

https://doi.org/10.1016/j.compbiomed.2021.105191
Received 14 November 2021; Received in revised form 27 December 2021; Accepted 27 December 2021
Available online 30 December 2021

Table 1
Patient population and image acquisition parameters for the development set and the test set. Median values and range for continuous variables.

Specification	Development set	Test set
Patients	12	1497
Male	6	704
Female	6	793
Age (y)	81 (range 69–87)	81 (range 36–96)
Scanners		
SOMATOM Force, Siemens Healthcare, Erlangen, Germany	8	898
Brilliance 64, Philips Healthcare, Best, The Netherlands	4	576
SOMATOM Definition AS+, Siemens	–	14
SOMATOM Definition Flash, Siemens	–	9
Low-pitch spiral scan	8	576
Axial scan	4	921
Tube potential (kVp)	90 (range 70–120)	90 (range 70–120)
Tube current-time product (mAs)	399 (range 310–628)	406 (range 95–901)
In-plane resolution (mm)	0.41 (range 0.31–0.61)	0.41 (range 0.28–0.79)
Slice thickness (mm)	0.6 (range 0.6–0.9)	0.6 (range 0.6–3.0)
Increment (mm)	0.45	0.45
Reconstructed 3D images	162	21,443
Number of phases available	11 (range 8–21)	11 (range 2–21)
Interval between phases		
5%	8	891
10%	4	554
Only end-systolic and end-diastolic phases	–	52

[1], for automatic LV wall motion analysis [2], or for the assessment of right ventricular function [3].

Manual segmentation of the cardiac chambers and the LV myocardium, especially over the whole cardiac cycle, is time-consuming and practically unfeasible. To overcome this limitation, automatic methods for the segmentation of cardiac structures in contrast-enhanced cardiac CT images have been developed, including model-based approaches [4,5] and atlas-based approaches [6–8]. Over the recent years, deep learning-based approaches have become state-of-the-art for cardiac segmentation [9–11]. Recent advances include using a deeply supervised 3D U-Net and a combination of focal loss and Dice loss for a better extraction of contextual information [12], combining deep learning with volumetric shape models [13], end-to-end training of one localization and one segmentation network [14], knowledge distillation for cross-modality segmentation in MRI and CT [15], using two variational autoencoders with segmentation modules for unsupervised domain adaptation between MRI and CT [16], and disentangling domain-invariant and domain-specific features for cross-modality cardiac segmentation [17]. However, 4D cardiac CT images pose several challenges for automatic segmentation methods. Unlike whole-heart segmentation in one specific cardiac phase, heart chambers may have a different appearance depending on the phase in the cardiac cycle. This is challenging for a learning method that is only trained on images of one cardiac phase, such as the end-diastolic (ED) phase in Ref. [9]. To achieve temporally consistent segmentations, Myronenko et al. have used a 4D CNN with a sparse loss function to segment the LV cavity and myocardium in 4D cardiac CT images [18]. Kong et al. use an end-to-end trainable architecture consisting of a voxel encoder, a voxel decoder, and a mesh decoder that take an image and a template mesh as input and directly output surface meshes of cardiac chambers, and also applied it to 4D cardiac CT images [19]. However, imaging artifacts caused by metal implants, large atherosclerotic calcifications, or other pathology might complicate automatic segmentation in a clinical dataset of 4D cardiac CT images.

Therefore, we present a detailed evaluation of an automatic method for cardiac chamber and LV myocardium segmentation in 4D contrast-enhanced cardiac CT images in a large clinical dataset that uses both end-systolic and ED images for training. In a preliminary study, we presented the automatic segmentation method and focused on the automatic selection of ES and ED phases as the main mode of evaluation [1]. In the present study, we have significantly increased our dataset to 1509 patients who were scanned in the work-up for transcatheter aortic valve implantation (TAVI) [20,21]. In this clinical dataset with larger anatomical variability and variability in image quality, we perform a detailed qualitative evaluation to investigate the performance of the method for potential clinical use.

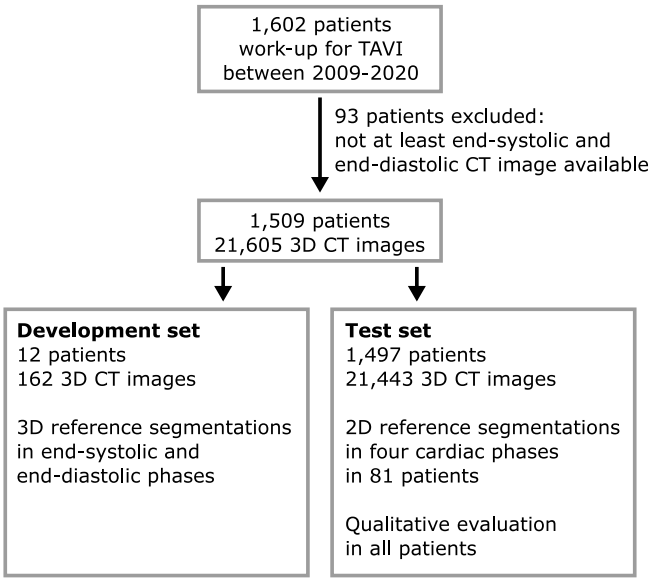

Fig. 1. Dataset overview and available reference segmentations per set.

2. Material and methods

2.1. Data and annotations

This retrospective study included 1602 consecutive patients who underwent the work-up for TAVI at the Amsterdam University Medical Center between 2009 and 2020. We excluded 93 patients who did not have a contrast-enhanced cardiac CT scan with at least ES and ED reconstructed images. For the remaining 1509 patients, at least ES and ED images were available and a total of 21,605 3D images were collected under a waiver from the local ethics committee. In our preliminary study, we reported on a subset of 472 patients included in this current study [1]. Further details on patient population and image acquisition parameters can be found in Table 1.

The data set was divided into a set for method development and a test set (Fig. 1). To develop a method that generalizes well while keeping the annotation time attainable, the development set contained contrast-enhanced cardiac CT scans of twelve patients selected to represent data variability in terms of scanners, patient sex, body weight, heart rate, and the presence of pacemakers. Per patient, full 3D reference segmentations for method development were obtained in the ES and ED images for a total of 24 full 3D reference segmentations. First, automatic segmentations of LV myocardium, LV cavity, right ventricle (RV), left atrium (LA), and right atrium (RA) were obtained using a method trained on a different dataset, which has been published for LV myocardium segmentation and spectral CT augmentation [22]. Where necessary, these were manually corrected by a radiology resident (TPWB), three years of experience, level II cardiac CT reader) by voxel-wise annotation. While this approach significantly reduces the manual workload, manual correction of the full 3D segmentation of all structures still took roughly 5 h per 3D image.

Table 2
Overview of CNN architecture. 3D convolutional layers (Conv3D) with 3D batch normalization (3D BatchNorm) and rectified linear units (ReLU) are used. Strided convolutions are used in the downsampling path, followed by residual ResNet blocks, and transposed convolutions for upsampling.

Layer	Type	Kernel size	Padding	Stride	Channels	3D BatchNorm	ReLU
1	Conv3D	7	3	1	8	yes	yes
2	Conv3D	3	1	2 (strided)	16	yes	yes
3	Conv3D	3	1	2 (strided)	32	yes	yes
4–9	ResNet block	3	1	1	32	yes	yes
10	Conv3D	3	1	2 (transposed)	16	yes	yes
11	Conv3D	3	1	2 (transposed)	8	yes	yes
12	Conv3D	7	3	1	6	no	no
13	Softmax	–	–	–	6	–	–

Therefore, it was infeasible to obtain 3D reference segmentations in all 21,443 3D images in the test set. To obtain reference segmentations in a representative number of patients across different cardiac phases, 4D cardiac CT images of 81 consecutive patients were selected, and 2D reference segmentations for the LV myocardium, LV cavity, RV, LA, and RA were obtained in these images. For each patient, a central axial 2D image was manually segmented by a medical student at four different time points in the cardiac cycle, viz. the mid-systolic, ES, mid-diastolic, and ED phases. Annotations were verified and, if necessary, corrected by a radiology resident (TPWB). All manual annotations, both 2D and 3D, were performed with 3DSlicer (3D Slicer 4.8.1, http://www.slicer.org).

For additional external evaluation of the method's performance and for comparison to previously published methods, we used the CT data from the publicly available multi-modality whole heart segmentation challenge [9] consisting of 20 3D contrast-enhanced cardiac CT images for training and 40 3D contrast-enhanced cardiac CT images as a test set.

2.2. Segmentation with convolutional neural network

A 3D convolutional neural network (CNN) architecture was used for automatic multi-class segmentation of the LV myocardium, LV cavity, RV, LA, and RA across the entire cardiac cycle. The 3D CNN architecture was based on a previously proposed 2D architecture [23] and consisted of an encoding path with two downsampling layers with strided convolutions, six residual ResNet blocks, and a decoding path composed of two upsampling layers with transposed convolutions [24]. 3D batch normalization is used for stabilized network training. The softmax function as the output layer is used to predict per-class probabilities that sum up to 1. A detailed overview of all layers of the network can be found in Table 2. All 3D CT images were resampled to $0.8 \times 0.8 \times 0.8$ mm^3 isotropic resolution, and CT numbers were linearly rescaled to a [0, 1] range before processing with the CNN. No data augmentation was performed. CNN training was carried out with mini-batches of eight $128 \times 128 \times 128$ voxel patches. For each patch, one training image was randomly selected and padded with its edge values in each direction that was smaller than 128 voxels. After that, the center of the cube-shaped patch was randomly selected such that the entire patch was inside the (padded) image. Adam [25] was used as the optimizer and the negative sum of soft Dice similarity coefficients as the loss function for an equal weighting of structures of different sizes [26]. CNN training was performed with an initial learning rate of 0.001 that was reduced by 70% every 4000 iterations. Early stopping was applied after 10,000 iterations to reduce overfitting. On average, CNN training took 8.5 h. These hyperparameters were selected based on previous work on whole-heart segmentation in non-contrast-enhanced cardiac CT [27]. During testing, 3D CNN output probability maps were combined across image patches by averaging, and the largest connected component of each structure was retained in the final segmentation mask. On average, CNN inference took 11 s per image.

We performed leave-one-patient-out cross-validation in the development set to reduce overfitting. For 3D CT images of ES and ED phases separately, twelve CNNs were trained. Each CNN was trained using images of eleven patients and its performance was evaluated using the image of the patient not used for training. For evaluation of segmentation in the independent test set, an ensemble of all 24 trained CNNs was used. Each CNN independently segmented the 3D CT image, and output probabilities of all CNNs were averaged to obtain final segmentations. This approach increases the robustness of the method. CNN training and testing were performed with PyTorch 1.4.0 (https://www.pytorch.org) and Python 3.6 (https://www.python.org) on graphics processing units (GeForce RTX 2080 Ti, NVIDIA, Santa Clara, CA, USA). The code can be found at https://github.com/qurAI-amsterdam/cardiacSegmentationCCTA.

2.3. Quantitative evaluation

Automatic 3D segmentations across the whole cardiac cycle in the cross-validation in the development set were evaluated against the available 3D reference segmentations in ES and ED phases using the Dice similarity coefficient (DSC) and the average symmetric surface distance (ASSD). Moreover, automatically obtained volumes were compared to reference volumes through Bland-Altman analysis.

In the test set, all cardiac phases were segmented, too, and the ES and ED phases were automatically selected as the ones with minimal and maximal LV volume, respectively. The mid-systolic and mid-diastolic phases were selected between the automatically identified ES and ED phases. Automatic segmentations in these four cardiac phases were evaluated against the 2D reference segmentations in axial slices in 81 patients using DSC and ASSD.

To facilitate comparison with other automatic segmentation methods, we re-trained our method with the training data from the publicly available multi-modality whole heart segmentation challenge [9] and segmented their test set of 40 3D contrast-enhanced cardiac CT images.

2.4. Qualitative evaluation

Our method automatically segmented all 21,443 3D CT images in the test set. Due to the extensive manual workload required for manual segmentations of all 3D images, we could not perform a complete quantitative evaluation. Therefore, we performed a qualitative analysis on all patients in the test set in addition to the quantitative evaluation on the subset of patients. A 4th year radiology resident (JHR, with 10 years of experience in image analysis) inspected six central 2D images (axial, sagittal, coronal for ES and ED phases) with corresponding automatic segmentation masks in each patient and graded image quality and segmentation quality.

Eight image quality categories were assessed separately on a three-point scale (1: no issues, 2: mild issues, 3: severe issues): correct image orientation, heart coverage in the field-of-view, contrast enhancement level, image noise level, metal artifacts, step-and-shoot artifacts, cardiac motion artifacts, and anatomical abnormalities.

Segmentation quality of the five segmented structures was graded separately on a three-point scale. Because previously published methods for whole-heart segmentation in contrast-enhanced cardiac CT that perform well report DSC values between 0.85 and 0.95, and inter-

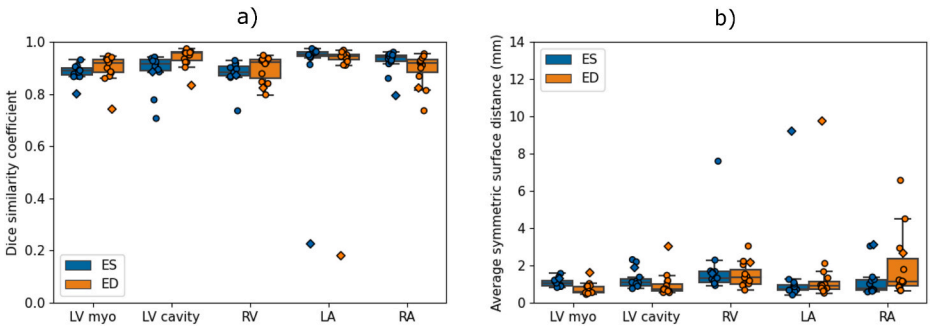

Fig. 2. Dice similarity coefficient (a) and average symmetric surface distance (b) of automatic segmentation for left ventricular (LV) myocardium, LV cavity, right ventricle (RV), left atrium (LA), and right atrium (RA) in end-systolic (ES) and end-diastolic (ED) phases against 3D reference segmentations in leave-one-patient-out cross-validation in the development set. Diamond markers indicate results for one patient with substantially lower contrast enhancement.

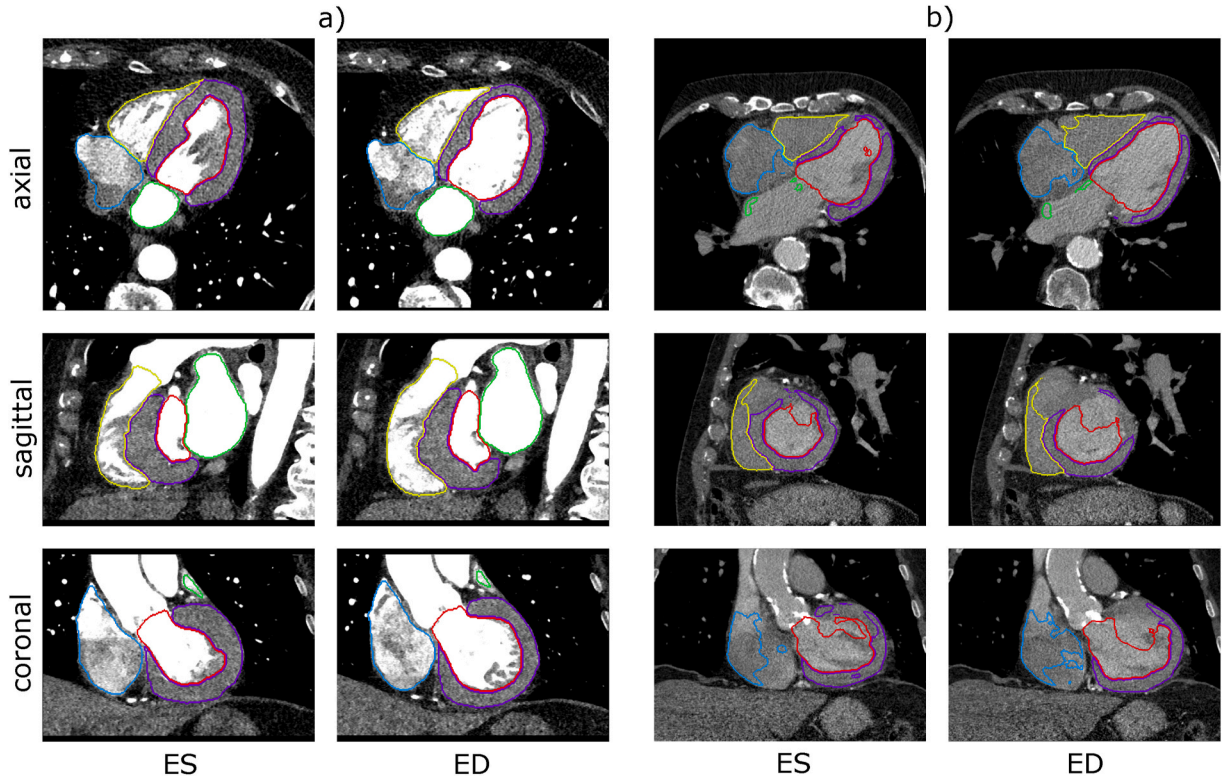

Fig. 3. Automatic segmentations in end-systolic (ES) and end-diastolic (ED) images in the leave-one-patient-out cross-validation in the development set. a) images of a patient with a representative contrast enhancement, b) images of a patient with lower contrast enhancement level than the rest of the images in the development set resulting in a poor segmentation of the left ventricle and atria.

observer DSC are found to be between 0.81 and 0.94 [9], we define grade 1 as a visually estimated >85% segmentation overlap between automatic segmentation and expected ground truth, clinically useful for volume quantification. Grade 2 is defined as 60–85% overlap, clinically useful after minor manual correction, and grade 3 is defined as <60% overlap, failed segmentation. In a subset of 274 patients stratified for the CT scan date, a second observer (TPWB) performed the same qualitative analysis.

Spearman rank-order correlation coefficients were used to investigate correlations between image quality and quality of automatic segmentations.

3. Results

3.1. Quantitative evaluation

Fig. 2 shows DSC and ASSD for 3D segmentation in the cross-validation experiments in the development set. These results demonstrate that automatic segmentation of the LV myocardium and the LV cavity is more accurate in the ED phase than in the ES phase, while segmentation of the RA is more accurate in the ES phase. Overall, the results show that the median DSC is always above 0.85, and the median ASSD is always below 1.5 mm, with the left atrium showing the highest DSC and lowest ASSD values.

Fig. 3 illustrates the best and worst segmentations obtained in our cross-validation experiments. For the patient shown in Fig. 3 (a), automatic segmentation was very accurate in both ES and ED phases across all structures. For the patient in Fig. 3 (b), the contrast enhancement

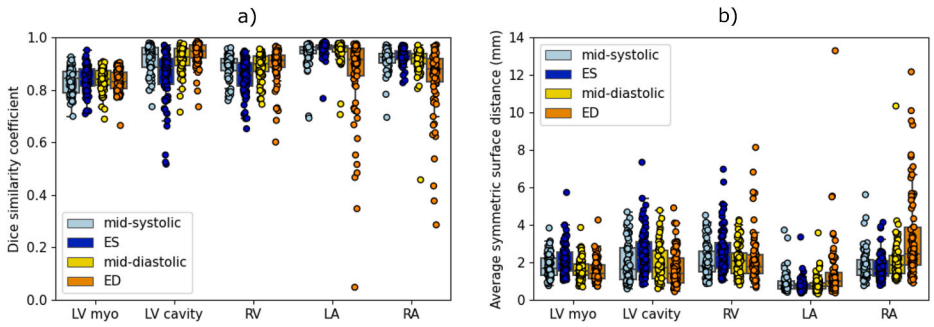

Fig. 4. Dice similarity coefficient (a) and average symmetric surface distance (b) of automatic segmentations for left ventricular (LV) myocardium, LV cavity, right ventricle (RV), left atrium (LA), and right atrium (RA) in mid-systolic, end-systolic (ES), mid-diastolic, and end-diastolic (ED) phases against 2D reference segmentations in 81 patients in the test set.

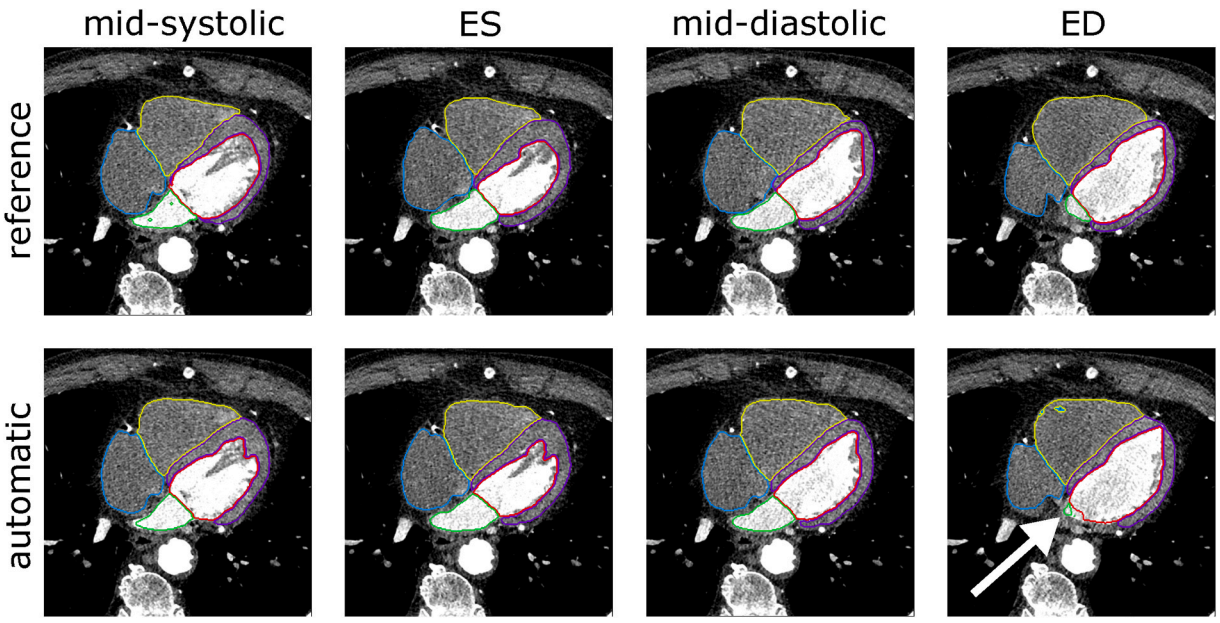

Fig. 5. Reference and automatic segmentations on mid-systolic, end-systolic (ES), mid-diastolic, and end-diastolic (ED) phase of a patient from the test set in which the automatic segmentation of the left atrium has low Dice similarity coefficient with respect to the 2D reference segmentation due to the small area covered by the left atrium in the chosen 2D image (corresponding to the most extreme outlier in Fig. 4). Note that segmentation in axial images in the transverse plane leads to foreshortening of the left ventricle.

level is substantially lower than in the other images from the development set, and the segmentation failed to a large extent. Quantitative results for this patient are indicated with diamond markers in Fig. 2. After excluding this patient with lower contrast enhancement level, bias and limits of agreement between reference and automatically obtained volumes were −2.2 [−16.6; 12.1] mL for LV cavity, 4.4 [−16.8; 25.5] mL for RV, 3.9 [−5.2; 13.0] mL for LA, 8.9 [−33.4; 51.3] mL for RA, and 10.1 [−12.3; 32.4] mL for LV myocardium in the ES images, and 3.7 [−8.1; 15.5] mL for LV cavity, 12.6 [−31.4; 56.7] mL for RV, 2.3 [−12.2; 16.9] mL for LA, 1.3 [−37.3; 39.9] mL for RA, and 2.9 [−19.0; 24.7] mL for LV myocardium in the ED images, respectively.

Fig. 4 shows DSC and ASSD for the 2D segmentations in mid-systolic, ES, mid-diastolic, and ED images of 81 patients in the test set. For all five structures, the automatic segmentation method produced accurate segmentations with median DSC above 0.8 and median ASSD below 2.5 mm in the four different cardiac phases. Outliers are mainly observed when structures only cover a small area in the 2D reference image as can be seen in Fig. 5.

Fig. 6 shows segmentations in a patient from the test set in comparison with the manual reference standard in these four cardiac phases, and the volume of the LV over the whole cardiac cycle. The automatically derived LVEF was 49.1% for this patient. Note that despite the imaging artifacts in the right atrium, the images were accurately segmented. A video of an automatically segmented beating heart can be found in the online version of this manuscript. The external evaluation of segmentation performance on the multi-modality whole-heart segmentation challenge dataset [9] yielded a DSC of 0.915 ± 0.025 for whole-heart segmentation, which is slightly higher than the current best method from the challenge (0.908 ± 0.086).

3.2. *Qualitative evaluation*

The qualitative evaluation was performed on the full test set. In 68 out of 1497 patients, the heart was not fully covered by the field-of-view as indicated by the main observer. These patients were excluded from further analysis because accurate volume quantification would not be feasible. Fig. 7 shows the results of our qualitative analysis for the remaining 1429 patients. Reduced contrast enhancement, especially in the right side of the heart, image noise level, and metal artifacts were the most frequently observed image quality issues, while step-and-shoot artifacts, cardiac motion artifacts, and anatomical abnormalities occurred less often. For all structures, in more than 80% of the cases,

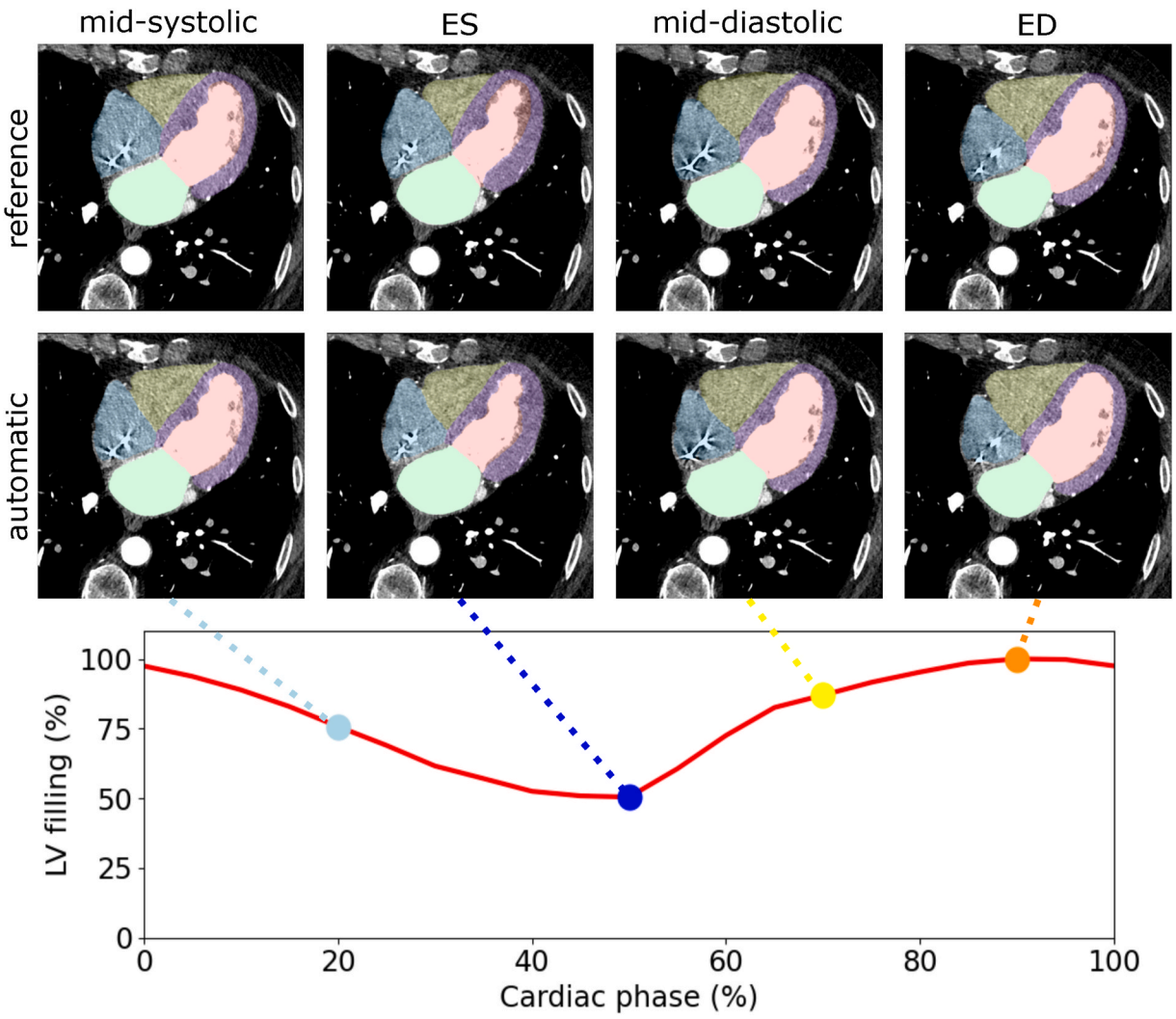

Fig. 6. Reference and automatic segmentations on mid-systolic, end-systolic (ES), mid-diastolic, and end-diastolic (ED) phases of a patient from the test set and the automatically derived left ventricular (LV) filling over the cardiac cycle.

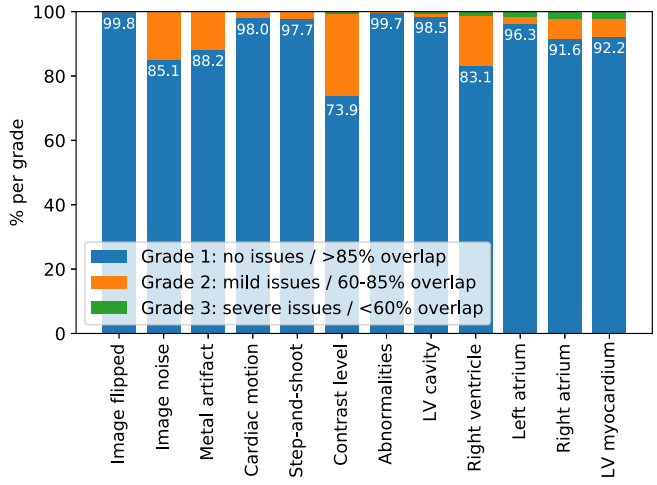

Fig. 7. Results of the qualitative evaluation in the test set. Percentage of scans assigned "grade 1: no issues" through "grade 3: severe issues" for different categories. Percentages of automatic segmentations assigned "grade 1: >85% segmentation overlap, clinically useful for volume quantification" through "grade 3: <60% overlap, failed segmentation" for all cardiac structures.

Table 3
Spearman rank-order correlation coefficients between qualitative grades for image quality categories and segmentation quality per structure. Values in bold are statistically significant ($p < 0.05$). LV = left ventricular, RV = right ventricle, LA = left atrium, RA = right atrium.

	LV cavity	RV	LA	RA	LV myo
Image orientation	0.370	0.121	0.239	0.157	0.167
Noise level	0.043	−0.009	0.076	0.072	0.122
Metal artifacts	0.026	−0.009	0.089	0.102	0.017
Cardiac motion	0.066	0.007	0.054	0.051	0.036
Step-and-shoot	0.020	0.019	0.044	0.021	0.042
Contrast enhancement level	0.133	0.265	0.101	0.043	0.022
Abnormalities	−0.007	0.087	0.061	0.031	0.085

clinically useful segmentations were obtained. The right ventricle was most challenging to segment (83.1% grade 1), the LV cavity was least challenging (98.5% grade 1).

Table 3 shows the Spearman rank-order correlation coefficients between image quality grades and segmentation quality grades. We find a statistically significant correlation between contrast enhancement level and segmentation quality. Moreover, we find a strong correlation between image orientation and segmentation quality for almost all structures. High image noise, metal artifacts, cardiac motion, and anatomical abnormalities only slightly influence segmentation quality of some

Table 4
Spearman rank-order correlation coefficient between qualitative grades for segmentation quality of different structures. Values in bold are statistically significant ($p < 0.05$). LV = left ventricular, RV = right ventricle, LA = left atrium, RA = right atrium.

	LV cavity	RV	LA	RA	LV myo
LV cavity	**1.0**	**0.182**	**0.495**	**0.241**	**0.362**
RV	**0.182**	**1.0**	**0.157**	**0.178**	**0.232**
LA	**0.495**	**0.157**	**1.0**	**296**	**0.291**
RA	**0.241**	**0.178**	**0.296**	**1.0**	**0.214**
LV myo	**0.362**	**0.232**	**0.291**	**0.214**	**1.0**

structures, whereas no statistically significant correlation is found for step-and-shoot artifacts.

Table 4 shows the Spearman rank-order correlation coefficient between the image quality grades of different structures. The correlation is significant across all structures and highest between LV cavity and left atrium. In the subset of 274 patients, the first and second observers agreed on a segmentation quality grade in 94.9%, 77.4%, 90.5%, 83.6%, and 81.4% of the cases for LV, RV, LA, RA, and LV myocardium, respectively. The largest agreement between observers was found for the LV, which also received grade 1 most often, while the lowest agreement was found for the RV, which was most challenging to be segmented and was therefore likely more often on the edge between grade 1 and grade 2. Over all structures, the observers agreed in 85.5%, differed by one grade in 13.9%, and differed by two grades in 0.6% of the cases. All reported grades for the qualitative evaluation can be found in Table 5.

4. Discussion

Automatic segmentation of the cardiac chambers and LV myocardium over the whole cardiac cycle in contrast-enhanced cardiac CT might enable in-depth assessment of cardiac morphology and function. Given that manual segmentation is not feasible due to high workload in routine clinical practice, we developed a deep learning-based method for whole-heart segmentation in 4D contrast-enhanced cardiac CT and evaluated it in a large clinical dataset for TAVI planning. In this set of 1429 patients, we found that our automatic method produced clinically useful segmentations in 98.5%, 92.2%, 83.1%, 96.3%, and 91.6% of the cases for LV cavity and myocardium, right ventricle, left atrium, and right atrium, respectively.

Our quantitative evaluation showed that the automatic method produces very accurate segmentations for all cardiac structures. This indicates that the routinely acquired high-resolution contrast-enhanced cardiac CT scans could also be used for a more comprehensive assessment of cardiac morphology and function. Results were on par with the results achieved by atlas-based and learning-based methods in a recent challenge on whole-heart cardiac CT segmentation in the ED phase [9]. Poor segmentation in the development set was found only in one patient whose scan showed low contrast enhancement not represented in the training data. Diversity in the levels of contrast enhancement in training data will likely improve the method's robustness [22,28].

Furthermore, the results in the test set indicate that the method generalizes well to cardiac phases not included in the training. Segmentation outliers were mainly observed when the corresponding structures covered only a small area in the 2D image that was manually segmented. Differences in segmentation quality between different cardiac phases may be due to changing morphology: The LV cavity and myocardium might be more difficult to segment in the ES phase due to their strong contraction which has also been reported for automatic segmentation in cardiac MRI [29]. On the other hand, RA segmentation performance might depend strongly on the mixing of contrast agent and blood throughout the cardiac cycle.

In this study, cardiac CT images of severely diseased patients who are not eligible for surgical valve replacement were analyzed. These scans are clinically used to accurately size the valve prosthesis, quantify valve calcifications, and determine the optimal access route [30,31]. The most frequently occurring image quality issues hampering segmentation were low contrast enhancement in the right side of the heart, image noise, and metal artifacts. Despite these issues, the method produced clinically useful segmentations in the majority of patients. The high number of clinically useful LV cavity and LA segmentations might be due to excellent contrast enhancement distinguishing them from neighboring structures, which is in line with existing literature [9]. On the other hand, the RV was affected the most by segmentation errors due to lower contrast enhancement and unclear boundaries.

Functional markers such as the LVEF can directly be derived from the ES and ED segmentations of the LV cavity. Cardiac MR images, which are the reference standard for LVEF, were not clinically acquired during the workup for TAVI and could therefore not be used as a reference. In

Table 5
Qualitative evaluation results in scans of 1429 patients after excluding the ones in which the heart was not fully covered. Image quality scored for different categories as grade 1: no issues, grade 2: mild issues, and grade 3: severe issues. Segmentation quality scored for different structures as grade 1: >85% segmentation overlap between automatic segmentation and expected ground truth, clinically useful for volume quantification, grade 2: 60–85% overlap, clinically useful after minor manual correction, grade 3: <60% overlap, failed segmentation.

	Grade	LV cavity			Right ventricle			Left atrium			Right atrium			LV myo			Total
Grade		1	2	3	1	2	3	1	2	3	1	2	3	1	2	3	
Image orientation	1	1407	11	8	1187	225	14	1376	29	21	1309	87	30	1318	78	30	1426
	3	0	0	3	0	0	3	0	0	3	0	1	2	0	0	3	3
Contrast level	1	1048	5	3	938	112	6	1027	19	10	973	65	18	975	63	18	1056
	2	358	4	2	247	112	5	347	10	7	333	22	9	343	14	7	364
	3	1	2	6	2	1	6	2	0	7	3	1	5	0	1	8	9
Noise level	1	1200	8	8	1008	195	13	1178	23	15	1124	69	23	1138	56	22	1216
	2	206	3	3	179	29	4	198	6	8	184	19	9	180	22	10	212
	3	1	0	0	0	1	0	0	0	1	1	0	0	0	0	1	1
Metal artifacts	1	1242	8	10	1045	200	15	1221	20	19	1167	72	21	1164	69	27	1260
	2	165	2	1	141	25	2	155	8	5	141	16	11	154	8	6	168
	3	0	1	0	1	0	0	0	1	0	1	0	0	0	1	0	1
Step-and-shoot	1	1375	11	10	1161	219	16	1346	27	23	1280	85	31	1290	74	32	1396
	2	31	0	0	26	5	0	29	2	0	28	3	0	28	3	0	31
	3	1	0	1	0	1	1	1	0	1	1	0	1	0	1	1	2
Cardiac motion	1	1381	11	9	1164	222	15	1351	28	22	1286	86	29	1294	76	31	1401
	2	26	0	1	23	3	1	25	1	1	23	2	2	24	2	1	27
	3	0	0	1	0	0	1	0	0	1	0	0	1	0	0	1	1
Abnormalities	1	1403	11	11	1186	223	16	1373	29	23	1306	87	32	1316	77	32	1425
	2	4	0	0	1	2	1	3	0	1	3	1	0	2	1	1	4
	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Total		1407	11	11	1187	225	17	1376	29	24	1309	88	32	1318	78	33	1429

future work, we will collect a set of 4D contrast-enhanced cardiac CT images in patients for whom cardiac MRI is available as well, to compare the automatically determined LVEF from CT against the clinical reference standard. Besides LVEF, our method allows quantification of other markers for cardiac function such as ED LV volume, LV stroke volume, LV wall thickness, right ventricular ejection fraction, myocardial mass, and other cardiac chamber volumes for assessment of cardiac function and cardiovascular risk prediction [32–34], which need to be compared to the respective clinical reference standards in future work. This could also improve risk stratification in patients undergoing transcatheter aortic valve implantation [35]. Based on 4D segmentation, more sophisticated markers beyond volumes could be derived, such as myocardial wall motion, which is usually assessed using cardiac MRI [36].

Our study had several limitations. The CT scan protocol and contrast injection protocol were tailored to assess the left side of the heart and aorta. Minor changes to the contrast injection protocol would substantially improve the attenuation of the right side of the heart and therewith improve automatic segmentation. Due to the high workload of manual segmentation, we did not quantitatively evaluate segmentation on all 3D images in all patients. The quantitative and qualitative analysis of 2D images that we performed instead can only be generalized to 3D with caution. A large part of the analysis was carried out on automatically determined ES and ED images. Although segmentation of ES and ED images already provide a wealth of information, valuable insights for the assessment of 4D cardiac function might lie beyond ES and ED images. In this work, 4D images were analyzed with 3D CNNs. With further hardware improvements and increased availability of full 4D reference segmentations, future work should explore true 4D CNN segmentation [18] in a large clinical dataset.

To conclude, we developed an automatic deep learning-based method to segment cardiac chambers and left ventricular myocardium in 4D contrast-enhanced cardiac CT. We evaluated the method with a large set of routine clinical scans for transcatheter aortic valve implantation planning. Our method produced clinically useful segmentations across the cardiac cycle and thus holds promise for automatic extraction of morphological and functional cardiac measures in any 4D contrast-enhanced cardiac CT.

CRediT authorship contribution statement

Steffen Bruns: Conceptualization of this study, analysis of data, development of software, drafting the article, revising the article. Jelmer M. Wolterink: Conceptualization of this study, development of software, drafting the article, revising the article. Thomas P.W. van den Boogert: Conceptualization of this study, acquisition of data, analysis of data, revising the article. Jurgen H. Runge: Analysis of data, revising the article. Berto J. Bouma: Analysis of data, revising the article. José P. Henriques: Acquisition of data, revising the article. Jan Baan: Acquisition of data, revising the article. Max A. Viergever: Conceptualization of this study, revising the article. R. Nils Planken: Conceptualization of this study, acquisition of data, revising the article. Ivana Išgum: Conceptualization of this study, drafting the article, revising the article.

Declaration of competing interest

Ivana Išgum received institutional research projects by the Dutch Technology Foundation co-funded by Pie Medical Imaging and Philips Healthcare (P15-26), the Netherlands Organisation for Health Research and Development with participation of Pie Medical Imaging (104003009), institutional research grant by Dutch Cancer Foundation (BRAGATSTON study), institutional research grant by Dutch Heart Foundation (CVON2015-17), and institutional research grants funded by Pie Medical Imaging. Ivana Išgum is a cofounder and scientific lead of Quantib BV, the Netherlands. Jelmer M. Wolterink acknowledges funding from the 4TU Precision Medicine program supported by High Tech for a Sustainable Future, a framework commissioned by the four Universities of Technology of the Netherlands. Jan Baan received an independent research grant from Edwards Lifesciences. Berto J. Bouma received an unrestricted research grant from Abbott. The authors declare that the research was conducted in the absence of any additional commercial or financial relationships that could be construed as a potential conflict of interest.

Acknowledgements

This study was funded by the Dutch Technology Foundation (STW, perspectief, P15-26) with participation of Philips Healthcare, Haifa, Israel.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.compbiomed.2021.105191.

References

[1] Steffen Bruns, Jelmer M. Wolterink, Thomas PW. van den Boogert, José P. Henriques, Baan Jan, R Nils Planken, Ivana Išgum, Automatic whole-heart segmentation in 4D TAVI treatment planning CT, in: Medical Imaging 2021: Image Processing, vol. 11596, International Society for Optics and Photonics, 2021, 115960B.
[2] Kai Uwe Juergens, Roman Fischbach, Left ventricular function studied with MDCT, Eur. Radiol. 16 (2) (2006) 342–357.
[3] Norbert F. Voelkel, Robert A. Quaife, Leslie A. Leinwand, Robyn J. Barst, Michael D. McGoon, Daniel R. Meldrum, Jocelyn Dupuis, Carlin S. Long, Lewis J. Rubin, Frank W. Smart, et al., Right ventricular function and failure, Circulation 114 (17) (2006) 1883–1891.
[4] Yefeng Zheng, Barbu Adrian, Bogdan Georgescu, Michael Scheuering, Dorin Comaniciu, Four-chamber heart modeling and automatic segmentation for 3-D cardiac CT volumes using marginal space learning and steerable features, IEEE Trans. Med. Imag. 27 (11) (2008) 1668–1681.
[5] Olivier Ecabert, Jochen Peters, Hauke Schramm, Cristian Lorenz, Jens von Berg, Matthew J. Walker, Mani Vembar, Mark E. Olszewski, Subramanyan Krishna, Lavi Guy, et al., Automatic model-based segmentation of the heart in CT images, IEEE Trans. Med. Imag. 27 (9) (2008) 1189–1201.
[6] Maria A. Zuluaga, M Jorge Cardoso, Marc Modat, Sébastien Ourselin, Multi-atlas propagation whole heart segmentation from MRI and CTA using a local normalised correlation coefficient criterion, in: International Conference on Functional Imaging and Modeling of the Heart, Springer, 2013, pp. 174–181.
[7] Xiahai Zhuang, Wenjia Bai, Jingjing Song, Songhua Zhan, Xiaohua Qian, Wenzhe Shi, Yanyun Lian, Daniel Rueckert, Multiatlas whole heart segmentation of CT data using conditional entropy for atlas ranking and selection, Med. Phys. 42 (7) (2015) 3822–3833.
[8] Tarun Kanti Ghosh, Md Kamrul Hasan, Shidhartho Roy, Md Ashraful Alam, Eklas Hossain, Mohiuddin Ahmad, Multi-class probabilistic atlas-based whole heart segmentation method in cardiac CT and MRI, IEEE Access 9 (2021) 66948–66964.
[9] Xiahai Zhuang, Lei Li, Christian Payer, Darko Štern, Martin Urschler, Mattias P. Heinrich, Julien Oster, Chunliang Wang, Örjan Smedby, Bian Cheng, et al., Evaluation of algorithms for multi-modality whole heart segmentation: an open-access grand challenge, Med. Image Anal. 58 (2019) 101537.
[10] Marija Habijan, Danilo Babin, Irena Galić, Hrvoje Leventić, Krešimir Romić, Lazar Velicki, Aleksandra Pižurica, Overview of the whole heart and heart chamber segmentation methods, Cardiovasc. Eng. Technol. (2020) 1–23.
[11] Chen Chen, Chen Qin, Huaqi Qiu, Giacomo Tarroni, Jinming Duan, Wenjia Bai, Daniel Rueckert, Deep learning for cardiac image segmentation: a review, Front. Cardiovasc. Med. 7 (25) (2020).
[12] Chengqin Ye, Wei Wang, Shanzhuo Zhang, Kuanquan Wang, Multi-depth fusion network for whole-heart CT image segmentation, IEEE Access 7 (2019) 23421–23429.
[13] Chunliang Wang, Örjan Smedby, Automatic whole heart segmentation using deep learning and shape context, in: International Workshop on Statistical Atlases and Computational Models of the Heart, Springer, 2017, pp. 242–249.
[14] Christian Payer, Darko Štern, Horst Bischof, Martin Urschler, Multi-label whole heart segmentation using CNNs and anatomical label configurations, in: International Workshop on Statistical Atlases and Computational Models of the Heart, Springer, 2017, pp. 190–198.
[15] Dou Qi, Quande Liu, Pheng Ann Heng, Ben Glocker, Unpaired multi-modal segmentation via knowledge distillation, IEEE Trans. Med. Imag. 39 (7) (2020) 2415–2425.
[16] Fuping Wu, Xiahai Zhuang, Unsupervised domain adaptation with variational approximation for cardiac segmentation, IEEE Trans. Med. Imag. 40 (12) (2021) 3555–3567.
[17] Chenhao Pei, Fuping Wu, Liqin Huang, Xiahai Zhuang, Disentangle domain features for cross-modality cardiac image segmentation, Med. Image Anal. 71 (2021) 102078.

[18] Andriy Myronenko, Yang Dong, Varun Buch, Daguang Xu, Alvin Ihsani, Sean Doyle, Mark Michalski, Neil Tenenholtz, Holger Roth, 4D CNN for semantic segmentation of cardiac volumetric sequences, in: International Workshop on Statistical Atlases and Computational Models of the Heart, Springer, 2019, pp. 72–80.
[19] Fanwei Kong, Nathan Wilson, Shawn C. Shadden, A deep-learning approach for direct whole-heart mesh reconstruction, Med. Image Anal. 74 (2021) 102222.
[20] Martin B. Leon, Craig R. Smith, Michael Mack, D Craig Miller, Jeffrey W. Moses, Lars G. Svensson, E. Murat Tuzcu, John G. Webb, Gregory P. Fontana, Raj R. Makkar, et al., Transcatheter aortic-valve implantation for aortic stenosis in patients who cannot undergo surgery, N. Engl. J. Med. 363 (17) (2010) 1597–1607.
[21] Michael J. Mack, Martin B. Leon, Vinod H. Thourani, Makkar Raj, Susheel K. Kodali, Mark Russo, Samir R. Kapadia, S. Chris Malaisrie, David J. Cohen, Philippe Pibarot, et al., Transcatheter aortic-valve replacement with a balloon-expandable valve in low-risk patients, N. Engl. J. Med. 380 (18) (2019) 1695–1705.
[22] Steffen Bruns, Jelmer M. Wolterink, Robbert W van Hamersvelt, Majd Zreik, Tim Leiner, Ivana Išgum, Improving myocardium segmentation in cardiac CT angiography using spectral information, in: Medical Imaging 2019: Image Processing, vol. 10949, International Society for Optics and Photonics, 2019, 109492M.
[23] Justin Johnson, Alahi Alexandre, Li Fei-Fei, Perceptual losses for real-time style transfer and super-resolution, in: European Conference on Computer Vision, Springer, 2016, pp. 694–711.
[24] Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun, Deep residual learning for image recognition, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 770–778.
[25] Diederik P. Kingma, Ba Jimmy, Adam: A Method for Stochastic Optimization, 2014 arXiv preprint arXiv:1412.6980.
[26] Fausto Milletari, Nassir Navab, Ahmadi Seyed-Ahmad, V-net: fully convolutional neural networks for volumetric medical image segmentation, in: 2016 Fourth International Conference on 3D Vision (3DV), IEEE, 2016, pp. 565–571.
[27] Steffen Bruns, Jelmer M. Wolterink, Richard AP. Takx, Robbert W van Hamersvelt, Dominika Suchá, Max A. Viergever, Tim Leiner, Ivana Išgum, Deep learning from dual-energy information for whole-heart segmentation in dual-energy and single-energy non-contrast-enhanced cardiac CT, Med. Phys. 47 (10) (2020) 5048–5060.
[28] Pierre-Jean Lartaud, Aymeric Rouchaud, Jean-Michel Rouet, Olivier Nempont, Loic Boussel, Spectral CT based training dataset generation and augmentation for conventional CT vascular segmentation, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2019, pp. 768–775.
[29] Mariana Bustamante, Vikas Gupta, Daniel Forsberg, Carl-Johan Carlhäll, Engvall Jan, Tino Ebbers, Automated multi-atlas segmentation of cardiac 4D flow MRI, Med. Image Anal. 49 (2018) 128–140.
[30] Stephan Achenbach, Victoria Delgado, Jörg Hausleiter, Schoenhagen Paul, James K. Min, Jonathon A. Leipsic, SCCT expert consensus document on computed tomography imaging before transcatheter aortic valve implantation (TAVI)/ transcatheter aortic valve replacement (TAVR), J. Cardiovasc. Comput. Tomogr. 6 (6) (2012) 366–380.
[31] Thomas PW. van den Boogert, Jeroen Vendrik, Bimmer EPM. Claessen, Baan Jan, Marcus A. Beijk, Jacqueline Limpens, S.A. Matthijs Boekholdt, R. Hoek, R Nils Planken, José P. Henriques, CTCA for detection of significant coronary artery disease in routine TAVI work-up, Neth. Heart J. 26 (12) (2018) 591–599.
[32] Kumar Narayanan, Kyndaron Reinier, Carmen Teodorescu, Audrey Uy-Evanado, Aleong Ryan, Harpriya Chugh, Gregory A. Nichols, Karen Gunson, Barry London, Jonathan Jui, et al., Left ventricular diameter and risk stratification for sudden cardiac death, J. Am. Heart Assoc. 3 (5) (2014), e001193.
[33] Marco R. Di Tullio, Donna R. Zwas, Ralph L. Sacco, Robert R. Sciacca, Shunichi Homma, Left ventricular mass and geometry and the risk of ischemic stroke, Stroke 34 (10) (2003) 2380–2384.
[34] S Nagarajarao Harsha, D Penman Alan, Herman A. Taylor, Thomas H. Mosley, Kenneth Butler, Thomas N. Skelton, Tandaw E. Samdarshi, Giorgio Aru, Ervin R. Fox, The predictive value of left atrial size for incident ischemic stroke and all-cause mortality in African Americans: the atherosclerosis risk in communities (ARIC) study, Stroke 39 (10) (2008) 2701–2706.
[35] Philippe Généreux, Philippe Pibarot, Björn Redfors, Michael J. Mack, Raj R. Makkar, Wael A. Jaber, Lars G. Svensson, Samir Kapadia, E Murat Tuzcu, Vinod H. Thourani, et al., Staging classification of aortic stenosis based on the extent of cardiac damage, Eur. Heart J. 38 (45) (2017) 3351–3358.
[36] Chen Qin, Wenjia Bai, Schlemper Jo, Steffen E. Petersen, Stefan K. Piechnik, Stefan Neubauer, Daniel Rueckert, Joint learning of motion estimation and segmentation for cardiac MR image sequences, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2018, pp. 472–480.