Mihaela Pop · Maxime Sermesant
Pierre-Marc Jodoin · Alain Lalande
Xiahai Zhuang · Guang Yang
Alistair Young · Olivier Bernard (Eds.)

Statistical Atlases and Computational Models of the Heart

ACDC and MMWHS Challenges

LNCS 10663

8th International Workshop, STACOM 2017
Held in Conjunction with MICCAI 2017
Quebec City, Canada, September 10–14, 2017, Revised Selected Papers

Springer

Lecture Notes in Computer Science 10663

Commenced Publication in 1973
Founding and Former Series Editors:
Gerhard Goos, Juris Hartmanis, and Jan van Leeuwen

Editorial Board

More information about this series at http://www.springer.com/series/7412

Mihaela Pop · Maxime Sermesant
Pierre-Marc Jodoin · Alain Lalande
Xiahai Zhuang · Guang Yang
Alistair Young · Olivier Bernard (Eds.)

Statistical Atlases and Computational Models of the Heart

ACDC and MMWHS Challenges

8th International Workshop, STACOM 2017
Held in Conjunction with MICCAI 2017
Quebec City, Canada, September 10–14, 2017
Revised Selected Papers

Springer

Editors
Mihaela Pop
Sunnybrook Research Institute
University of Toronto
Toronto, ON
Canada

Maxime Sermesant
Inria-Asclepios
Sophia Antipolis
France

Pierre-Marc Jodoin
Université de Sherbrooke
Quebec City, QC
Canada

Alain Lalande
University Bourgogne
Dijon
France

Xiahai Zhuang
Fudan University
Shanghai
China

Guang Yang
Cardiovascular Research Center, National Heart and Lung Institute
Royal Brompton Hospital
London
UK

Alistair Young
University of Auckland
Auckland
New Zealand

Olivier Bernard
CREATIS, INSA-Lyon
Villeurbanne
France

ISSN 0302-9743 ISSN 1611-3349 (electronic)
Lecture Notes in Computer Science
ISBN 978-3-319-75540-3 ISBN 978-3-319-75541-0 (eBook)
https://doi.org/10.1007/978-3-319-75541-0

Library of Congress Control Number: 2018935903

LNCS Sublibrary: SL6 – Image Processing, Computer Vision, Pattern Recognition, and Graphics

Printed on acid-free paper

This Springer imprint is published by the registered company Springer International Publishing AG part of Springer Nature
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

Integrative models of cardiac function are important for understanding disease, evaluating treatment, and planning intervention. In recent years, there has been considerable progress in cardiac image analysis techniques, cardiac atlases, and computational models, which can integrate data from large-scale databases of heart shape, function, and physiology. However, significant clinical translation of these tools is constrained by the lack of complete and rigorous technical and clinical validation, as well as benchmarking of the developed tools. To this end, common and available ground-truth data capturing generic knowledge on the healthy and pathological heart are required. Several efforts are now established to provide Web-accessible structural and functional atlases of the normal and pathological heart for clinical, research, and educational purposes. We believe that these approaches will only be effectively developed through collaboration across the full research scope of the cardiac imaging and modeling communities.

STACOM 2017 was held in conjunction with the MICCAI 2017 international conference (Quebec City, Canada), following the past seven editions: STACOM 2010 (2010, Beijing, China), STACOM 2011 (Toronto, Canada), STACOM 2012 (Nice, France), STACOM 2013 (Nagoya, Japan), STACOM 2014 (Boston, USA), STACOM 2015 (Munich, Germany), and STACOM 2016 (Athens, Greece). STACOM 2017 provided a forum in which to discuss the latest developments in various areas of computational imaging and modeling of the heart as well as statistical cardiac atlases. The topics of the workshop included: cardiac imaging and image processing, atlas construction, statistical modeling of cardiac function across different patient populations, cardiac computational physiology, model customization, atlas-based functional analysis, ontological schemata for data and results, integrated functional and structural analyses, as well as the pre-clinical and clinical applicability of these methods. Besides regular contributing papers, additional efforts of this year's STACOM workshop were also focused on two challenges: ACDC and MM-WHS, described here. A total of 27 papers (regular papers and from the two challenges) were accepted to be presented at STACOM 2017, and are published in this LNCS proceedings volume.

ACDC Automatic Cardiac Diagnostic Challenge

The overarching objective of this challenge is two-fold:

(1) To compare the performance of automatic MRI segmentation methods on the left ventricular endocardium and epicardium as well as the right ventricular endocardium for both the end-diastolic and end-systolic phase instances

(2) To compare the performance of automatic methods for the classification of examinations in five classes (normal case, myocardial infarction with altered left ventricular ejection fraction, dilated cardiomyopathy, hypertrophic cardiomyopathy, abnormal right ventricle)

The overall AC-DC dataset contains real clinical examinations from 150 patients all acquired at the University Hospital of Dijon (France). Each patient dataset comes with two ground truth information: (1) the pathology the patient suffers from and (2) a pixel-accurate delineation of each cardiac region (the endocardial wall of the left ventricle and of the right ventricles, and the epicardial wall of the left ventricle). The segmentation ground truths were manually drawn by two experts (a cardiologist and a nuclear medicine physician with experience in cardiology and MRI). The delineation was done for the most important phases of the cardiac cycle, i.e., diastole and systole. The diastolic phase is the first image acquired after the R-wave of the ECG while the systolic phase is the moment when the mitral valve reaches its maximum excursion and the myocardium reaches its maximum contraction. The dataset as well as the results obtained by the challengers can be found here: http://acdc.creatis.insa-lyon.fr/.

MM-WHS — Multi-Modal Whole-Heart Segmentation Challenge

Accurate computing, modeling, and analysis of the whole-heart substructures from 3D medical image scans is important in the development of clinical applications. Segmentation and registration of whole-heart images is, however, still challenging. The extraction and modeling of whole-heart substructures currently relies heavily on manual delineation, which is a time-consuming task and is also prone to errors and dependent on the expertise of the observer; therefore, fully automated methods are highly desirable. Over the past decade, many techniques have been proposed to solve this ill-posed problem, particularly for whole-heart segmentation, such as atlas-based methods, statistical shape model-based methods, and recently emerged deep-learning-based methods. The organized MM-WHS Challenge provided an open and fair competition for various research groups to test and validate their methods, particularly whole-heart segmentation, on datasets that were acquired in real clinical environments. The aim of the MM-WHS Challenge was not only to benchmark various whole-heart segmentation algorithms, but also to cover the topic of general cardiac image segmentation, registration, and modeling, and to raise discussions for further technical development and clinical deployment.

The organizers provided more than 120 datasets from multiple sites, including 60 cardiac CT/CTA and 60 cardiac MRI volumes in 3D that cover whole-heart substructures for multi-modality whole-heart segmentation. All these clinical data received institutional ethic approval and were anonymized. Both cardiac CT and cardiac MRI data were acquired in real clinical environment for patients that cover a wide range of cardiac diseases as well as normal controls. We received great interest from participants all over the world and the proposed methods have achieved substantial methodological innovations and significant performance improvement. We aim at keeping the MM-WHS Challenge as a long-term event for participants who may not be able to enter the competition, but are interested in further developments. All the relevant information and challenge results can be found at: http://www.sdspeople.fudan.edu.cn/zhuangxiahai/0/mmwhs/.

We would like to thank all organizers, reviewers, authors, and sponsors for their time, efforts, contributions, and financial support in making STACOM 2017 a successful event. We hope that the results obtained by these two challenges, along with the regular paper contributions, will act to accelerate progress in the important areas of cardiac image analysis, heart function, and structure analysis.

September 2017

Mihaela Pop
Maxime Sermesant
Pierre-Marc Jodoin
Alain Lalande
Xiahai Zhuang
Guang Yang
Alistair Young
Olivier Bernard

Organization

Chairs and Organizers

STACOM

Mihaela Pop	University of Toronto, Canada
Alistair Young	University of Auckland, New Zealand
Maxime Sermesant	Inria-Asclepios, France
Tommaso Mansi	Siemens, USA
Kawal Rhode	KCL, London, UK
Kristin McLeod	Simula, Norway

ACDC Challenge

Pierre-Marc Jodoin	University of Sherbrook, Canada
Alain Lalande	University of Bourgogne, France
Olivier Bernard	University of Lyon, France

MM-WHS Challenge

Xiahai Zhuang	Fudan University
Guang Yang	Imperial College London, UK
Lei Li	Shanghai Jiao Tong University, China

OCS - Springer Conference Submission/Publication System

Mihaela Pop	Medical Biophysics, University of Toronto, Sunnybrook Research Institute, Toronto, Canada

Webmaster

Avan Suinesiaputra	University of Auckland, New Zealand

Workshop Website

stacom2017.cardiacatlas.org

Sponsors

We are extremely grateful for the industrial funding support. The STACOM 2017 workshop received financial support from the following sponsors:

SciMedia Ltd (http://www.scimedia.com/) for STACOM
Imeka (http://imeka.ca) for ACDC challenge
Nvidia (http://nvidia.com) for MM-WHS challenge
Arterys (http://arterys.com) for MM-WHS challenge

Contents

Regular Papers

Multiview Machine Learning Using an Atlas of Cardiac Cycle Motion

Esther Puyol-Antón[1(✉)], Matthew Sinclair[1], Bernhard Gerber[3], Mihaela Silvia Amzulescu[3], Hélène Langet[2,3], Mathieu De Craene[2], Paul Aljabar[1], Julia A. Schnabel[1], Paolo Piro[2], and Andrew P. King[1]

[1] Division of Imaging Sciences and Biomedical Engineering, King's College London, London, UK
esther.puyol_anton@kcl.ac.uk
[2] Philips Research, Medisys, Paris, France
[3] Division of Cardiology, Cliniques Universitaires St-Luc, Avenue Hippocrate 10-2881, 1200 Brussels, Belgium

Abstract. A cardiac motion atlas provides a space of reference in which the cardiac motion fields of a cohort of subjects can be directly compared. From such atlases, descriptors can be learned for subsequent diagnosis and characterization of disease. Traditionally, such atlases have been formed from imaging data acquired using a single modality. In this work we propose a framework for building a multimodal cardiac motion atlas from MR and ultrasound data and incorporate a multiview classifier to exploit the complementary information provided by the two modalities. We demonstrate that our novel framework is able to detect non ischemic dilated cardiomyopathy patients from ultrasound data alone, whilst still exploiting the MR based information from the multimodal atlas. We evaluate two different approaches based on multiview learning to implement the classifier and achieve an improvement in classification performance from 77.5% to 83.50% compared to the use of US data without the multimodal atlas.

Keywords: Multimodal cardiac motion atlas
Multiview dimensionality reduction · Classification

1 Introduction

Heart disease is the leading cause of death globally, and evaluation of the function of the left ventricle (LV) can provide useful information for diagnosis and characterization of disease. Magnetic resonance imaging (MRI) is increasingly accepted as the gold standard for assessing global as well as regional heart function due to its excellent image contrast. However, its clinical use is limited by lack of access to scanners, high acquisition cost and a lack of expertise in acquiring the data. Instead, ultrasound (US) is commonly used in the clinic for assessing cardiac function due to its low acquisition cost and portability. For example,

M. Pop et al. (Eds.): STACOM 2017, LNCS 10663, pp. 3–11, 2018.
https://doi.org/10.1007/978-3-319-75541-0_1

the typical cost of a cardiac MR scan is US$600 whereas the corresponding cost for US is US$200. Most cardiac US is 2D, making estimation of volumetric disease biomarkers prone to error. However, 3D US probes are becoming more widely available and have the potential to achieve relatively low cost volumetric assessment of cardiac function.

In recent years, cardiac motion atlases have been proposed as a way of analyzing cardiac cycle motion in the context of population variation. Motion atlases have been used for detecting disease [5], predicting response to treatment [4] and estimating scar location in the LV [2].

Traditionally, such techniques have been based on motion information estimated from a single imaging modality. However, information from different modalities may be complementary due to the different characteristics of the acquisition devices, so using the combined information may provide a richer description of cardiac function.

The problem of analyzing imaging data from multiple sources was investigated in [3], which proposed a method to correct shape bias between different MRI sequence acquisitions. However, after bias correction the data were not combined into a single atlas. Furthermore, this was limited to single-modality imaging, rather than combining information from different modalities. Recently, we described a motion atlas technique for use in a multimodal context [6], but the atlas was only formed from single modality imaging data (MRI). A truly multimodal atlas (MRI and US) was recently presented in [7], but was based upon only healthy volunteer data.

In this work we extend the work proposed in [7] to include patient data and a novel framework that incorporates a multiview classifier to enable predictions to be made using test data from a single modality (US) whilst exploiting the multimodal (MRI and US) training data. We apply our novel framework to the problem of detecting patients with non ischemic dilated cardiomyopathy (DCM). DCM is a common and mostly irreversible form of heart muscle disease causing LV motion abnormality. It is the third most common cause of heart failure and the most frequent cause of heart transplantation. As cardiac motion is an early predictor of cardiac dysfunction, our eventual aim is to use the motion atlas to detect patients suffering from DCM before shape changes occur.

Section 2 describes the multimodal database used to generate the motion atlas, then Sect. 3 focuses on the construction of the atlas and the classification methods proposed. Section 4 presents the major results of the work, while a discussion of the results and conclusions are reported in Sect. 5.

2 Materials

Two databases of MRI and US data of the LV were combined for evaluation in this paper. The first is the database used for the cardiac motion analysis challenge at the 2011 MICCAI STACOM workshop [10]. The second database was acquired at the Division of Cardiology, Cliniques Universitaires St-Luc, Avenue Hippocrate 10-2881, B-1200 Brussels, Belgium. In total, these databases

include MRI and US datasets from 41 healthy volunteers and 19 patients with DCM (aged 42 ± 17 years, 40 male and 20 female). For both databases, the MRI acquisitions were performed using a 3T Philips Achieva System (Philips Healthcare, Best, The Netherlands), and the US datasets were acquired using an iE33 echocardiography system (Philips Medical Systems, Andover, MA, United States) with a 1–5 MHz transthoracic matrix array transducer (xMATRIX X5.1).

In particular, the two datasets contain for each subject:

- **cine SA**: a multi-slice short-axis (SA) cine-MR sequence (TR/TE = 3.0/1.5 ms, flip angle = 60°). Typical slice thickness was between 8.0–10.0 mm with an in-plane resolution ≈ 1.0 mm × 1.0. Typical temporal resolution was between 25 and 30 consecutive short axis images per cycle covering the entire LV. In our framework, only the end-diastolic (ED) frame was used for geometry definition (see Sect. 3.1). The cine SA sequence was not used for motion estimation.
- **TAG**: a 3D tagged MRI sequence in three orthogonal directions (TR/TE = 7.5/3.2 ms, flip angle = 19°, tag distance between 7.7–8.8 mm). The images have reduced field-of-view enclosing the LV, with typical isotropic 3D spatial resolution between 2.5 mm and 1.1 mm, and typical temporal resolution between 22 and 30 frames per cycle. The TAG data were used for motion estimation (see Sect. 3.1).
- **US**: Full-volume acquisition (FVA) mode was used in which several smaller imaging sectors are combined to form a large composite volume. Each smaller sector was acquired in a single heart cycle. FVA was performed during breath-hold to minimize translation artefacts between the acquired sectors. Apical FVAs of the left ventricle were obtained with all subjects on their side. Typical slice thickness was between 0.7 mm–1.0 mm with an in-plane resolution ≈ 0.65 mm × 0.8 mm. Typical temporal resolution was between 15 and 23 frames per cycle. The US data were used for geometry and motion estimation.

3 Methods

The main novelty of the proposed method lies in the application of a multimodal cardiac motion atlas for classification of patients with DCM. To allow an unbiased comparison of the motions of different subjects, a spatiotemporal motion atlas of the LV was generated using a similar framework to that proposed in [7]. The proposed framework is illustrated in Fig. 1. Details of the atlas formation are reported in Sect. 3.1, while Sect. 3.2 describes the different multiview algorithms used to classify patients with DCM.

3.1 Motion Atlas Formation

LV Geometry Definition. For each subject, the LV myocardium was manually segmented in the ED cine SA and US images. A statistical shape model (SSM) was optimized to fit to the endocardial and epicardial surfaces of the LV

binary segmentation [1], providing point-correspondences between subjects and modalities.

MRI/US Alignment. MRI and US images were aligned by registering their respective LV meshes (≈2200 vertices) using a Generalized Procrustes analysis. The American Heart Association (AHA) segment delineations generated in both meshes were used to ensure that the mid-septum was in the same location in both modalities.

Motion Tracking. A 3D GPU-based B-spline free-form deformation (FFD) was used [8] to estimate LV motion between consecutive frames of the MRI and US sequences. Subsequently, the inter-frame transformations were estimated and then composed using a $3D + t$ B-spline to estimate a full cycle $3D + t$ transformation.

Temporal Normalization. This stage establishes temporal correspondence between all subjects based on specific cardiac events and transforms them to a normalized timescale [7]. Note that this normalization is applied separately to the MRI and US data, but both are transformed to the same normalized timescale. We automatically identified the following cardiac events based on volume curves computed using the estimated geometries and motions for all subjects: (1) *Beginning of the isovolumic contraction phase:* the point at ED at which the volume in the cavity is maximized; (2) *Beginning of the isovolumetric relaxation phase:* the point at the end of systole (ES) at which the volume in the cavity is minimized; (3) *Beginning of the atrial systole phase:* The point when the atrium contracts and increases the volume in the ventricle with only a small amount of blood. The point of maximum atrial contraction is the minimum of the second derivative of the volume of the LV cavity. The events were used as landmarks in a piecewise linear temporal transformation to align the sequences to a normalized timescale.

Spatial Normalization. This stage aims to remove bias towards differences in subject-specific LV geometries from the motion analysis. Similar to [6], we transform each mesh to an unbiased atlas coordinate system using a combination of Procrustes alignment and Thin Plate Spline transformation. We denote the transformation for each subject n from its subject-specific coordinate system to the atlas by ϕ_n.

Medial Surface Generation. A medial surface mesh with regularly sampled vertices (≈1000) was generated from the SSM epicardial and endocardial meshes using ray-casting and homogeneous downsampling followed by cell subdivision. The use of a medial surface enables a more robust motion estimation compared to the endo- and epicardial surfaces of the SSM as it is likely to be less affected by motion tracking errors which can be caused by any potential inaccuracies in the ED segmentation.

Motion Reorientation. With the aim of comparing the LV motions from different subjects in the atlas coordinate system, displacements for MRI ($\boldsymbol{u}_n$) and US ($\boldsymbol{v}_n$) for each subject n were reoriented into the atlas coordinate

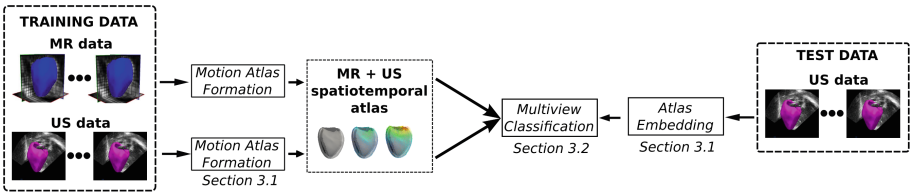


Fig. 1. Overview of the proposed framework for forming and applying a multimodal motion atlas.

system under a small deformation assumption using a push-forward action: $\boldsymbol{u}_n^{atlas} = \boldsymbol{J}(\phi_n(\boldsymbol{r}_n))\boldsymbol{u}_n$ and $\boldsymbol{v}_n^{atlas} = \boldsymbol{J}(\phi_n(\boldsymbol{r}_n))\boldsymbol{v}_n$ [4], where $\boldsymbol{J}(\cdot)$ refers to the Jacobian matrix of the subject-to-atlas mapping ϕ_n, and $\boldsymbol{r}_n$ represents the positions of the vertices in the ED frame.

Transform to Local Coordinate System. For a more intuitive understanding of the LV motion, for each subject n MRI displacements in the atlas coordinate system $\boldsymbol{u}_n^{atlas}$ were projected onto a local cylindrical coordinate system $\boldsymbol{x}_n^{atlas}$, providing radial, longitudinal, and circumferential information. The long axis of the LV ED atlas medial mesh was used as the longitudinal direction. Similarly, US displacements $\boldsymbol{v}_n^{atlas}$ were projected onto a local cylindrical coordinate system $\boldsymbol{z}_n^{atlas}$.

3.2 Multiview Classification

Once all of the MRI and US motion data (i.e. vertex displacements) are in the same reference space, the next task is to perform a classification. Similar to a number of previous works [4,5] we tackle this problem using dimensionality reduction techniques. To enable the dimensionality reduction the local displacements for each modality are concatenated into a single column vector such that for subject n, $\boldsymbol{x}_n^{atlas}$ and $\boldsymbol{z}_n^{atlas} \in \mathbb{R}^L$, where $L = 3MT$, T is the number of cardiac phases and M is the number of points in the atlas medial surface mesh. The column vectors for each subject are combined to produce matrices $X = [\boldsymbol{x}_1^{atlas}, ..., \boldsymbol{x}_N^{atlas}]$ and $Z = [\boldsymbol{z}_1^{atlas}, ..., \boldsymbol{z}_N^{atlas}]$, where N is the number of subjects, X is the MRI-derived motion matrix and Z is the US-derived motion matrix. With the aim of classifying patients, we propose two approaches: (1) apply a multiview dimensionality reduction algorithm followed by a standard classifier from the low dimensional embedding, (2) apply a multiview classifier, which both reduces the dimensionality and performs the classification. Figure 2 shows an overview of the proposed methods for classification along with two comparative monomodal approaches.

1. Multiview Dimensionality Reduction. We propose to use the nonlinear iterative partial least squares (PLS) algorithm [11], which finds linear projections to form a common space between both modalities and extracts meaningful descriptors to model the inter-modality variability. More specifically, PLS tries to

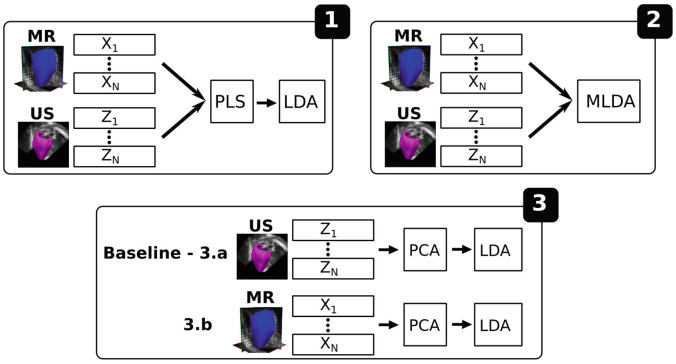


Fig. 2. Different approaches for classification: (1) proposed multiview dimensionality reduction followed by a standard classifier. (2) proposed multiview classifier. (3) Comparative approaches: 3.a - US-only pipeline, 3.b - MRI-only pipeline.

find an orthogonal basis, such that the covariance between the set of projections onto these basis vectors is mutually maximized:

$$\max_{\boldsymbol{w_x}, \boldsymbol{w_z}} \quad \boldsymbol{w_x}^T \boldsymbol{C_{xz}} \boldsymbol{w_z} \quad \text{s.t.} \quad \begin{cases} \boldsymbol{w_x}^T \boldsymbol{w_x} = 1 \\ \boldsymbol{w_z}^T \boldsymbol{w_z} = 1 \end{cases} \tag{1}$$

where $\boldsymbol{C_{xz}}$ corresponds to the cross-covariance matrix of the two views (i.e. MRI/US), $\boldsymbol{w_x}$ is the projection matrix for the first view (i.e. MRI), and $\boldsymbol{w_z}$ is the projection matrix for the second view (i.e. US). Hence, a subject k can be embedded in the low dimensional space using only MR data as $\boldsymbol{h}_k = \boldsymbol{w}_x^T \boldsymbol{x}_k$ and using US data as $\boldsymbol{q}_k = \boldsymbol{w}_z^T \boldsymbol{z}_k$. From the common embedding space between MRI and US, we apply a linear discriminant analysis (LDA) classifier that can be formulated as follows:

$$\max_{\boldsymbol{\alpha}} \quad \boldsymbol{\alpha}^T \boldsymbol{S_b} \boldsymbol{\alpha} \quad \text{s.t.} \quad \boldsymbol{\alpha}^T \boldsymbol{C_{qq}} \boldsymbol{\alpha} = 1 \tag{2}$$

where $\boldsymbol{\alpha}$ denotes a projection matrix, $\boldsymbol{C_{qq}}$ corresponds to the covariance matrix of the US embedding, and $\boldsymbol{S_b} = \sum_c N_c (\boldsymbol{\mu}_c - \overline{\boldsymbol{v}})(\boldsymbol{\mu}_c - \overline{\boldsymbol{v}})^T$ is the between-class variation [9]. N_c is the number of cases in class c, $\boldsymbol{\mu}_c$ is the mean of the class c, and $\overline{\boldsymbol{v}}$ is the mean of the class means.

2. Multiview Classifier. In this approach we used a multiview classifier based on the multiview linear discriminant analysis (MLDA) [9] algorithm. MLDA seeks to find a common space while simultaneously preserving the correlation between modalities and the discriminating information in each modality. Compared to the multiview dimensionality reduction approach, this method includes the inter-class information in the optimization process. The optimization problem of MLDA is given by:

$$\begin{aligned} \max_{\boldsymbol{w_x}, \boldsymbol{w_z}} \quad & \boldsymbol{w_x}^T \boldsymbol{S_{bx}} \boldsymbol{w_x} + \boldsymbol{w_z}^T \boldsymbol{S_{bz}} \boldsymbol{w_z} + 2\gamma \boldsymbol{w_x}^T \boldsymbol{C_{xz}} \boldsymbol{w_z} \\ \text{subject to} \quad & \boldsymbol{w_x}^T \boldsymbol{C_{xx}} \boldsymbol{w_x} = 1, \boldsymbol{w_z}^T \boldsymbol{C_{zz}} \boldsymbol{w_z} = 1 \end{aligned} \tag{3}$$

where $\boldsymbol{S_{bx}}$, $\boldsymbol{S_{bz}}$ denote the between-class matrices [9]. Similar to Eqs. (1) and (2), $\boldsymbol{C_{xx}}$, $\boldsymbol{C_{zz}}$ are the covariance matrices, and $\boldsymbol{C_{xz}}$ is the cross-covariance matrix. We can see that the leftmost two terms of Eq. 3 are related to the LDA algorithm, and attempt to minimize the within-class distance, while the rightmost term attempts to find a common space between the two views. γ is a regularization parameter that balances the relative significance between these two constraints.

3. Comparative Approaches. To analyze the improvement of using both modalities to classify patients, we compared the results of the two previous methods to single modality methods using Principal Component Analysis (PCA) followed by a LDA classifier (see Fig. 2). We consider the PCA technique trained using US motion data only to be our baseline comparative approach: this represents the current state-of-the-art in the use of US data alone for statistical analysis of motion. We also compare our novel approaches with the PCA technique trained and applied using MRI data alone (i.e. no US data is involved). Since MRI is considered to be the gold standard for analysis of cardiac function, we consider this approach to be the true state-of-the-art using the best available data, regardless of cost or other considerations.

4 Experiments and Results

Before applying any of the proposed methods, as a preprocessing step, we estimated the intrinsic dimensionality of both matrices X and Z using PCA. The estimated dimensionality ($d = 20$) was used for all compared approaches. To evaluate the performances of the different classification methods, we used 8-fold repeated stratified cross-validation with balanced classes and 100 repetitions. More specifically, in each fold for training there are 35 healthy volunteers and 17 patients, and for testing there are 2 patients and 6 healthy volunteers. We computed the classification accuracy (i.e. the proportion of subjects correctly classified), as well as the sensitivity (the proportion of healthy subjects correctly classified) and the specificity (the proportion of patients with DCM correctly classified). The average classification accuracies, sensitivities and specificities as well as their standard deviations across all folds were computed. In addition, the individual accuracy percentages for each fold were recorded and used as the input to Welch's $t-$tests to evaluate the statistical significance of the results. All methods were compared with the performance of PCA trained on US data (i.e. the baseline technique) with a significance level of 0.05. Table 1 shows the results of the proposed algorithms. Note that, although techniques 1 and 2 are capable of performing classification using data from either view, we only report here the results of classifying using the US data. Similar to [9] the regularization parameter γ in MLDA was optimized across values [1, 5, 10, 15, 20]. The stratified cross-validation was performed for each of these values of γ, and based on the results, the optimal value of γ was chosen.

Table 1. Classification accuracy, sensitivity and specificity of the proposed and comparative methods and Welch's $t-$test results. An asterisk indicates a statistically significant improvement in accuracy over the baseline comparative approach (i.e. 3.a). Bold text indicates the US-based method with the highest classification accuracy.

Method	ACC (%)	SEN (%)	SPE (%)
1. PLS + LDA	81.5 ± 13.2*	84.2 ± 15.3	73.5 ± 31.2
2. **MLDA**	**83.5 ± 12.8***	**94.5 ± 7.7**	**81.4 ± 23.2**
Baseline - 3.a. PCA_{US} + LDA	77.5 ± 15.5	78.7 ± 17.2	74.5 ± 22.9
3.b. PCA_{MRI} + LDA	87.8 ± 9.8*	88 ± 11.1	87.5 ± 23.9

Our results show that the highest accuracy is achieved using the MRI-only approach, which is consistent with clinical perception of MRI as a gold standard for analysis of cardiac function. However, the use of a multiview dimensionality reduction algorithm followed by a classifier increases the classification accuracy compared to the baseline approach (i.e. using only US data). Moreover, we can observe that using US data alone, the highest classification performance (around 84%) is achieved using a multiview classifier.

5 Discussion

We have presented a method for building a multimodal cardiac motion atlas from MRI and US data and demonstrated its application for classification between patients suffering from DCM and healthy volunteers. The novelty of this paper lies in the application of a multimodal motion atlas using multiview machine learning algorithms. We investigated two approaches, the first one based on multiview dimensionality reduction followed by a standard classifier, and the second one based on a multiview classifier. We compared both to single modality approaches. Furthermore, our proposed framework opens up the possibility of even better performance than the MRI-only approach by exploiting *both* MRI and US data as input to a multimodal atlas. We are planning to investigate this possibility in future work. We also plan to investigate the incorporation of geometry information into the atlas. DCM is characterized by a dilation of the LV so geometry information may provide useful information for the classification. We would also like to increase the database size (currently limited to 60 subjects), extend the pipeline to non-linear classification methods and investigate other applications in cardiology.

References

1. Bai, W., Shi, W., et al.: A bi-ventricular cardiac atlas built from 1000+ high resolution MR images of healthy subjects and an analysis of shape and motion. Med. Image Anal. **26**(1), 133–145 (2015)
2. Duchateau, N., De Craene, M., et al.: Infarct localization from myocardial deformation: prediction and uncertainty quantification by regression from a low-dimensional space. IEEE Trans. Med. Imaging **35**(10), 2340–2352 (2016)
3. Medrano-Gracia, P., Suinesiaputra, A., Cowan, B., Bluemke, D., Frangi, A., Lee, D., Lima, J., Young, A.: An atlas for cardiac MRI regional wall motion and infarct scoring. In: Camara, O., Mansi, T., Pop, M., Rhode, K., Sermesant, M., Young, A. (eds.) STACOM 2012. LNCS, vol. 7746, pp. 188–197. Springer, Heidelberg (2013). https://doi.org/10.1007/978-3-642-36961-2_22
4. Peressutti, D., Sinclair, M., et al.: A framework for combining a motion atlas with non-motion information to learn clinically useful biomarkers: application to cardiac resynchronisation therapy response prediction. Med. Image Anal. **35**, 669–684 (2017)
5. Perperidis, D., Mohiaddin, R., Rueckert, D.: Construction of a 4D statistical atlas of the cardiac anatomy and its use in classification. In: Duncan, J.S., Gerig, G. (eds.) MICCAI 2005. LNCS, vol. 3750, pp. 402–410. Springer, Heidelberg (2005). https://doi.org/10.1007/11566489_50
6. Puyol-Antón, E., Peressutti, D., et al.: Towards a multimodal cardiac motion atlas. In: ISBI, pp. 32–35. IEEE (2016)
7. Puyol-Antón, E., Sinclair, M., Gerber, B., Amzulescu, M., Langet, H., De Craene, M., Aljabar, P., Piro, P., King, A.: A multimodal spatiotemporal cardiac motion atlas from MR and ultrasound data. Med. Image Anal. **40**, 96–110 (2017)
8. Rueckert, D., Sonoda, L., et al.: Nonrigid registration using free-form deformations: application to breast MR images. IEEE Trans. Med. Imaging **18**(8), 712–721 (1999)
9. Sun, S., Xie, X., Yang, M.: Multiview uncorrelated discriminant analysis. IEEE Trans. Cybern. **18**, 712–721 (2015)
10. Tobon-Gomez, C., De Craene, M., et al.: Benchmarking framework for myocardial tracking and deformation algorithms: an open access database. Med. Image Anal. **17**(6), 632–648 (2013)
11. Wold, H.: Partial Least Squares. Encyclopedia of Statistical Sciences. Wiley, New York (1985)

Joint Myocardial Registration and Segmentation of Cardiac BOLD MRI

Ilkay Oksuz[1,2(✉)], Rohan Dharmakumar[3], and Sotirios A. Tsaftaris[2]

[1] IMT Institute for Advanced Studies Lucca, Lucca, Italy
ilkay.oksuz@imtlucca.it
[2] The University of Edinburgh, Edinburgh, UK
[3] Biomedical Imaging Research Institute, Cedars-Sinai Medical, Los Angeles, USA

Abstract. Registration and segmentation of anatomical structures are two well studied problems in medical imaging. Optimizing segmentation and registration jointly has been proven to improve results for both challenges. In this work, we propose a joint optimization scheme for registration and segmentation using dictionary learning based descriptors. Our joint registration and segmentation aims to solve an optimization function, which enables better performance for both of these ill-posed processes. We build two dictionaries for background and myocardium for square patches extracted from training images. Based on dictionary learning residuals and sparse representations defined on these pre-trained dictionaries, a Markov Random Field (MRF) based joint optimization scheme is built. The algorithm proceeds iteratively updating the dictionaries in an online fashion. The accuracy of the proposed method is illustrated on Cardiac Phase-resolved Blood Oxygen-Level-Dependent (CP-BOLD) MRI and standard cine Cardiac MRI data from MICCAI 2013 SATA Segmentation Challenge. The proposed joint segmentation and registration method achieves higher dice accuracy for myocardium segmentation compared to its variants.

Keywords: Segmentation · Registration · Markov Random Fields
Joint optimization · BOLD · CINE MR

1 Introduction

Cardiac Phase-resolved Blood Oxygen-Level-Dependent (CP-BOLD MRI) is a new imaging technique, free of stress and contrast agents, used for the assessment of myocardial ischemia at rest [19]. The specific segmentation and registration among the cardiac phases in this cine type acquisition is crucial for automated analysis approaches of this technique, since it potentially leads to better specificity of ischemia detection [4]. To achieve this, precise segmentation and nonlinear registration of the myocardium among the frames (the cardiac phases) in the cine stack would be required. Unfortunately, at present due to BOLD contrast variations, classical approaches to segmentation [15] and registration [14] fail to reach sufficient accuracy.

M. Pop et al. (Eds.): STACOM 2017, LNCS 10663, pp. 12–20, 2018.
https://doi.org/10.1007/978-3-319-75541-0_2

In this paper, we propose a joint registration and segmentation scheme to generate accurate timeseries information for cardiac sequence. We adopt a joint optimization scheme [11] to optimize the registration term on sparse representations and segmentation terms for dictionary learning residuals. The motivation behind this choice is the mutual benefit of both functions, which can be directly translated to accurate registration and segmentation.

There are two major contributions of this work. First, we define a joint optimization scheme based on dictionary learning residuals and sparse representations for the first time. Second, we introduce an iterative dictionary update stage, which takes the spatial smoothness into account to increase discriminative power of the dictionary learning structure. With this, the dictionaries are ensured to be subject-specific and more robust for classifying the myocardium region.

The remainder of the paper is organized as follows: Sect. 2 investigates the prior art on joint registration and segmentation. Section 3 discusses the dictionary learning based methods for image registration and segmentation and presents the proposed joint optimization scheme for myocardial timeseries generation. The experimental results are described in Sects. 4 and 5 concludes the paper.

2 Background

Registration and segmentation of organs in medical imaging are two major tasks, which are processed with two independent optimization schemes in most applications. One approach of solving both problems is using a sequential strategy to address both challenges, which results in concatenation of errors of both processes. Instead of a sequential segmentation and registration scheme, which uses estimated solution of one sub-problem as a prior knowledge to the other, joint optimization of two problems can be defined [21], where both problems are solved simultaneously. Early works merged the two processes with partial differential equations [20] and in particular within level-set formulations [22]. More recent literature relies on joint optimization with single function simultaneously using Markov Random Fields (MRF)s [7]. MRFs are suitable for discrete labeling problems and the labels are defined as segmentation classes and discrete displacement vectors. The concept of utilizing mutual benefits between the registration and segmentation has been studied for the problem of atlas-based tumor segmentation for brain MRI [16]. Alchatzidis et al. [3] proposes to couple segmentation and registration scheme for classifying multiple regions in brain MRI and compared with standard post-registration label fusion strategies [2]. Mahapatra et al. [10] used a joint optimization scheme to detect the left ventricle (LV) in standard cine and perfusion MR images.

In joint registration and segmentation, the estimate of one set of parameters of registration should not adversely affect parameters of segmentation. An appropriate optimization scheme aims to balance these influences. Graph cuts is based on maximum-flow approach and is very effective in finding the global minimum or a strong local minimum of discrete MRF energy formulations [5].

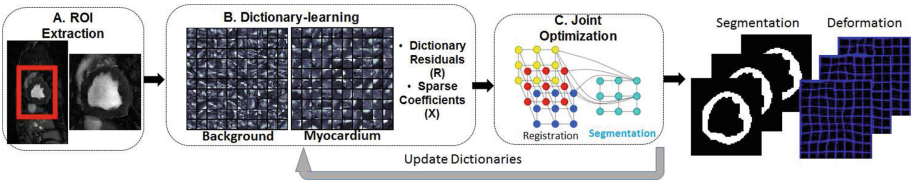


Fig. 1. Algorithm design for joint segmentation and registration. Region of interest extraction (Panel A). Dictionary learning from training images and calculation of residuals (R) and sparse coefficients (X) (Panel B). Multi-resolution deformation grids and exemplary connections with segmentation grid (Panel C)

However, a number of issues have to be addressed in using segmentation information for MRF-based registration. Registration and segmentation energies have to be combined such that there is no bias for a particular term. The mutual dependence of registration and segmentation has to be factored in the objective function.

In this work, we propose a joint optimization scheme for myocardial registration and segmentation to generate accurate deformations and segmentation masks for the entire cine stack. Our method builds upon the externally trained dictionaries of myocardium and background and uses priors on each problem jointly to extract and register the myocardial region. We introduce a dictionary update scheme to fuse subject-specific local information. Our algorithm generates deformations and segmentations for the entire cardiac sequence.

3 Methods

The details of our method are visualized in Fig. 1. We extract a region of interest around LV blood pool using a similar preprocessing strategy to [17]. We use externally trained dictionaries of myocardium and background to define registration and segmentation terms for joint optimization. Then, these terms are optimized using a discrete graphical model over the labels of registration and segmentation. Finally, we update our dictionaries to enrich subject-specific information in the dictionaries and run the optimization process again.

3.1 Dictionary Learning Based Image Segmentation

Dictionary learning based approaches have been used for segmentation of medical images [9]. In our specific algorithmic design, given some sequences of training images and corresponding ground truth labels, we can obtain two sets of matrices, Y^B and Y^M, where the matrix Y^B contains the background information and Y^M is the corresponding matrix referring to the myocardium. Squared patches are sampled around each pixel of the training images from both regions. The i-th column of the matrix Y^B (and similarly for the matrix Y^M) is obtained by concatenating the normalized patch vector of pixel intensities, taken around the

i-th pixel in the background, along with the Gabor and HOG features of the same patch. The method detailed in [13] trains two dictionaries, D_k^B and D_k^M, and two sparse feature matrices, X_k^B and X_k^M using the K-SVD algorithm [1] for each class $C = \{B, M\}$:

$$\underset{D^C, X^C}{\text{minimize}} \|Y^C - D^C X^C\|_2^2, \text{ subject to } \|x_i^C\|_0 \leq \text{sparsity}$$

After the training given a new subject, a certain patch will be assigned to the class that gives the smallest dictionary approximation error using Orthogonal Matching Pursuit [18]. If $R_B = \|\hat{y}_i - D^B \hat{x}_i^B\|_2$ is less than $R_M = \|\hat{y}_i - D^M \hat{x}_i^M\|_2$, the patch is assigned to the background; otherwise, it is considered belonging to the myocardial region.

3.2 Graph-Based Joint Optimization

In this section, we introduce our dictionary learning based joint optimization scheme for registration and segmentation of the myocardium. The general term for energy of a second-order MRF is defined as:

$$E(L) = \sum_{p \in \Omega} D_p(l_p) + \lambda \sum_{p,q \in N} V_{pq}(l_p, l_q)$$

where p and q denote the pixels, l_p and l_q denotes the registration and segmentation labels of the pixels p and q. λ controls the interaction between data term and smoothness term. The function is optimized over the labels $L = \{C, u\}$, which consists of the segmentation label C and discrete deformation u. We define the general data term $D_p(l_p)$ similar to [10]:

$$D_p(l_p) = D_{l_p}^1 + \gamma D_{l_p}^2$$

which consists of two terms, namely segmentation and registration data terms. Segmentation of the myocardium is defined over the dictionary learning residuals R_B and R_M. The penalty of the pixel p to be classified as myocardium is : $\kappa_M(p) = \frac{R_M(p)}{R_M(p)+R_B(p)}$. Similarly, the penalty for the same pixel to be classified as background is $\kappa_B(p) = \frac{R_B(p)}{R_M(p)+R_B(p)}$. Using these penalty definitions D_p^1 is defined as:

$$D_{l_p}^1 = \begin{cases} \sqrt{\kappa_M^r(p) * \kappa_M^f(p+u)}, & \text{if } C^r(p) = C^f(p+u) = M \\ \sqrt{\kappa_B^r(p) * \kappa_B^f(p+u)}, & \text{if } C^r(p) = C^f(p+u) = B \\ \sqrt{\kappa_B^r(p) * \kappa_B^f(p+u)} + \sqrt{\kappa_M^t(r) * \kappa_M^f(p+u)}, & \text{otherwise} \end{cases}$$

where $\kappa_M^f(p+u)$ corresponds to the penalty associated with myocardium class for the deformed floating image with displacement u. Similarly, $\kappa_B^r(p)$ corresponds to penalty of the reference image for the background class. This term ensures

a low penalty for same labels of the displaced image and the reference image. If the floating image and the reference image do favor different segmentation classes the penalty will be high.

The registration penalty term $D^2_{l_p}$ of the data term D_{l_p} is defined as:

$$D^2_{l_p} = \begin{cases} \| X^r_M(p) - X^f_M(p+u) \|_1, & \text{if } C^r(p) = C^f(p+u) = M \\ \| X^r_B(p) - X^f_B(p+u) \|_1, & \text{if } C^r(p) = C^f(p+u) = B \\ \| X^r_M(p) - X^f_M(p+u) \|_1 + \| X^r_B(p) - X^f_B(p+u) \|_1, & \text{otherwise} \end{cases}$$

where $X^r_M(p)$ corresponds to the sparse representation defined for D^M for the reference image and $X^f_M(p+u)$ defines sparse representation defined for the floating image at location $p+u$. This penalty is increased for dissimilar representation and also for the points with different segmentation labels.

Regularization term ensures the smoothness of segmentation labels and deformation field. The term favors the same segmentation labels in local neighborhoods N and smooth deformations. The regularization term is defined as:

$$V_{pq}(l_p, l_q) = \begin{cases} 1, & \text{if } (C_p = C_q \text{ and } \| u_p - u_q \| \leq \varepsilon) \\ 1, & \text{if } (C_p \neq C_q \text{ and } \| u_p - u_q \| \leq \tau) \\ 100, & \text{otherwise} \end{cases}$$

where ε and τ restrict high displacements for local neighborhoods N when segmentation labels agree or disagree respectively. To optimize the energy functional $E(L)$, we use graph cuts [5] on discrete labels of registration and segmentation.

3.3 Dictionary Update

We propose a dictionary update, which refines the dictionaries to inject subject-specific information. After every run of the MRF-based optimization scheme the estimated segmentation labels C are subject to change. We only extract patches that are corresponding to the points of label changes to update our dictionaries. We add square patches Y_u concatenated with Gabor and HOG features and train our dictionaries with Online Dictionary Learning (ODL) algorithm [12], which uses mini-batches to update the dictionaries. We add the new patches with changed labels for updating dictionaries we trained before. During the update the dictionary learning is initialized with the pre-trained dictionaries and this approach improves the discriminative power of the dictionaries in the next iteration.

4 Experimental Results

This section offers qualitative and quantitative comparison of our proposed method w.r.t. state-of-the-art methods, to demonstrate its effectiveness for myocardial segmentation and registration. Note that our method outperforms all supervised methods from current literature in both baseline and ischemia cases.

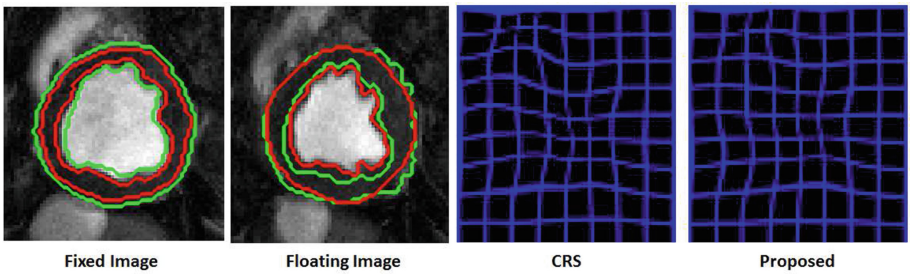


Fig. 2. Segmentation masks (red contours) and registration grid of proposed approach compared to CRS [11] (green contours) for an exemplary subject under baseline conditions in between end diastole and end systole frames. (Color figure online)

4.1 Data Preparation and Implementation Details

2D short-axis images of the whole cardiac cycle were acquired at baseline and severe ischemia (inflicted as stenosis of the left-anterior descending coronary artery (LAD)) on a 1.5T Espree (Siemens Healthcare) in the same 10 canines along mid ventricle using both standard CINE and a flow and motion compensated CP-BOLD acquisition within few minutes of each other. The image resolution is 192×114 and each cardiac cycle has 25 frames approximately. We have utilized a strict leave one out cross validation experiment, where the patch size is defined as 11×11, dictionary size as 100 and sparsity threshold as 8. The parameters of deformation $\varepsilon = \sqrt{2}$ and $\tau = 3$ are optimized to ensure smooth labels for deformations. $\gamma = 0.7$ gave the optimal contribution of the data terms and $\lambda = 0.9$ ensures the balance of regularization and data terms. We have utilized three scales from coarse to fine for registration. The influence of the control points on each pixel is calculated using cubic B-Splines [8]. The displacement ranges from 2 to 6 pixels.

4.2 Visual Evaluation

We demonstrate a set of contours and a deformation grid to highlight the performance of our joint optimization and registration framework. In Fig. 2, we visualize the deformation grid in between the end systole and end diastole for an exemplary subject under baseline condition. We also illustrate the segmentation and deformation results of CRS [11] compared with our algorithm. Our method generates smooth deformation fields and smooth contours compared to CRS [11].

4.3 Quantitative Comparison

Table 1 summarizes our results for Dice overlap measure for myocardium. We compare our algorithm with an atlas-based segmentation technique, which relies on discrete registration performance using mutual information as a similarity

Table 1. Dice overlap comparison of myocardial segmentation

Methods	Baseline		Ischemia	
	Standard cine	CP-BOLD	Standard cine	CP-BOLD
Atlas-based [8]	0.60 ∓ 0.03	0.54 ∓ 0.08	0.54 ∓ 0.08	0.45 ∓ 0.06
CRS [11]	0.71 ∓ 0.06	0.70 ∓ 0.06	0.69 ∓ 0.05	0.68 ∓ 0.07
Segmentation only	0.70 ∓ 0.05	0.71 ∓ 0.04	0.69 ∓ 0.04	0.68 ∓ 0.04
Sequential Seg. and Reg.	0.74 ∓ 0.06	0.72 ∓ 0.07	0.71 ∓ 0.07	0.68 ∓ 0.08
Proposed w.o update	0.75 ∓ 0.04	0.76 ∓ 0.04	0.75 ∓ 0.05	0.74 ∓ 0.04
Proposed	0.81 ∓ 0.04	0.80 ∓ 0.04	0.79 ∓ 0.04	0.80 ∓ 0.05

metric [8]. Moreover, we include a recent joint registration and segmentation scheme CRS [11], which relies on sum of squared distances and edge-based differences as similarity term for registration. We generate results based on dictionary residuals for each pixel just for segmentation (Segmentation only). In addition, we used a sequential segmentation and registration (Sequential Seg. and Reg.), which first segments myocardium based on residuals and then refines the contours with propagation of the contours via registration based on sparse representations. Finally, we generate a variant of our algorithm, without using the dictionary update (Proposed w.o. update) to highlight the performance increase.

The proposed method outperforms all variants and other techniques in all four datasets. Segmentation information alone shows low performance compared to the variants, which incorporate registration. The sequential segmentation and registration has low performance compared to the proposed method. This low performance is due to the mutual dependence of registration and segmentation that has not been factored in the objective function, which is ensured with the proposed approach. Our method is superior to CRS [11], which relies on edge-based terms for myocardial registration. Ischemia condition generates a slight decrease in the performance for all methods. The proposed dictionary update enables a performance improvement thanks to less coherent dictionaries. The coherence of two dictionaries is calculated before and after the single dictionary update. The average coherence of two dictionaries 0.850 is reduced to 0.780 with the update. We illustrate an example set of dictionaries before and after the update in Fig. 3.

4.4 CAP Dataset

To demonstrate that our method works also non-BOLD, clinical data, we have tested our algorithm on cine cardiac training data set from the MICCAI 2013 SATA Segmentation Challenge. The dataset is part of the Cardiac Atlas Project (CAP) [6] and consists of 83 subjects with a varying in plane resolution from 0.7 mm to 2 mm and a varying range of 19 to 30 frames per subject. On mid-ventricular level, we train our algorithm on 30 subjects to learn dictionaries for background and myocardium. Then, we test on the remaining 53 subjects and we

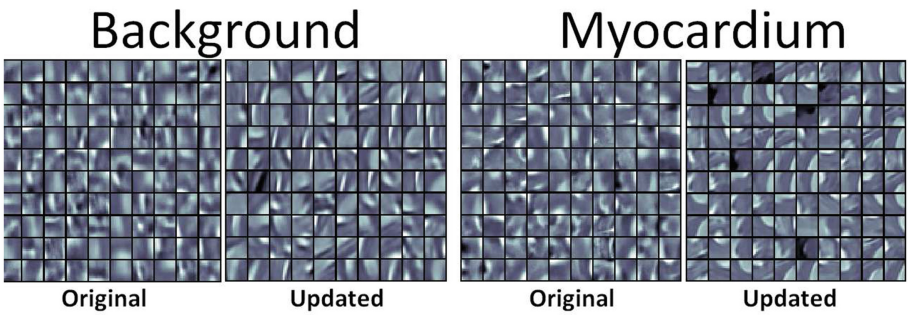


Fig. 3. Background and myocardium dictionaries before and after the dictionary update. Observe the increased number of unique myocardial patterns after the dictionary update.

achieve a dice score of 0.81 ∓ 0.04 compared to 0.80 ∓ 0.05 of CRS [11] algorithm (where standard deviation refers to variation among subjects and not on leave one out cross validation).

5 Conclusion

In this paper, we propose a joint registration and segmentation scheme based on sparse representation. Our algorithm uses a MRF-based optimization scheme defined on dictionary learning residuals and at each iteration the dictionaries are updated using patches corresponding to the points that changed segmentation labels. This not only improves the performance by introducing subject specific information, but also adds more discriminative power as showcased with experiments. Currently, our algorithm works on 2D and we would like to extend our method to 3D. Moreover, currently we evaluate the deformation field visually and not quantitatively. One way to evaluate the registration performance is the target registration error, which will be available with the definition of landmark points for each frame. In the future, we would like to evaluate our approach on perfusion images that show stronger spatio-temporal variations.

References

1. Aharon, M., et al.: K-SVD: an algorithm for designing overcomplete dictionaries for sparse representation. IEEE TSP **54**(11), 4311–4322 (2006)
2. Alchatzidis, A., et al.: A discrete MRF framework for integrated multi-atlas registration and segmentation. IJCV **121**(1), 169–181 (2017)
3. Alchatzidis, S., et al.: Discrete multi atlas segmentation using agreement constraints. In: BMVC (2014)
4. Bevilacqua, M., et al.: Dictionary-driven ischemia detection from cardiac phase-resolved myocardial BOLD MRI at rest. IEEE TMI **35**(1), 282–293 (2016)

5. Boykov, Y., et al.: Fast approximate energy minimization via graph cuts. IEEE PAMI **23**(11), 1222–1239 (2001)
6. Fonseca, C.G., et al.: The cardiac atlas projectan imaging database for computational modeling and statistical atlases of the heart. Bioinformatics **27**(16), 2288–2295 (2011)
7. Gass, T., et al.: Simultaneous segmentation and multiresolution nonrigid atlas registration. IEEE TIP **23**(7), 2931–2943 (2014)
8. Glocker, B., et al.: Dense image registration through MRFs and efficient linear programming. MedIA **12**(6), 731–741 (2008)
9. Huang, X., et al.: Contour tracking in echocardiographic sequences via sparse representation and dictionary learning. MedIA **18**, 253–271 (2014)
10. Mahapatra, D., Sun, Y.: Joint registration and segmentation of dynamic cardiac perfusion images using MRFs. In: Jiang, T., Navab, N., Pluim, J.P.W., Viergever, M.A. (eds.) MICCAI 2010. LNCS, vol. 6361, pp. 493–501. Springer, Heidelberg (2010). https://doi.org/10.1007/978-3-642-15705-9_60
11. Mahapatra, D., et al.: Integrating segmentation information for improved MRF-based elastic image registration. IEEE TIP **20**(1), 170–183 (2012)
12. Mairal, J., et al.: Online dictionary learning for sparse coding. In: ICML, pp. 689–696 (2009)
13. Mukhopadhyay, A., Oksuz, I., Bevilacqua, M., Dharmakumar, R., Tsaftaris, S.A.: Data-driven feature learning for myocardial segmentation of CP-BOLD MRI. In: van Assen, H., Bovendeerd, P., Delhaas, T. (eds.) FIMH 2015. LNCS, vol. 9126, pp. 189–197. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-20309-6_22
14. Oksuz, I., Mukhopadhyay, A., Bevilacqua, M., Dharmakumar, R., Tsaftaris, S.A.: Dictionary learning based image descriptor for myocardial registration of CP-BOLD MR. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) MICCAI 2015. LNCS, vol. 9350, pp. 205–213. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-24571-3_25
15. Oksuz, I., et al.: Unsupervised myocardial segmentation for cardiac BOLD. IEEE TMI **36**, 2228–2238 (2017)
16. Parisot, S., et al.: Concurrent tumor segmentation and registration with uncertainty-based sparse non-uniform graphs. MedIA **18**(4), 647–659 (2014)
17. Queiros, S., et al.: Fast automatic myocardial segmentation in 4D cine CMR datasets. MedIA **18**(7), 1115–1131 (2014)
18. Tropp, J.A., et al.: Signal recovery from random measurements via orthogonal matching pursuit. IEEE Trans. Inf. Theory **53**(12), 4655–4666 (2007)
19. Tsaftaris, S.A., et al.: Detecting myocardial ischemia at rest with cardiac phase-resolved blood oxygen level-dependent cardiovascular magnetic resonance. Circ. Cardiovasc. Imaging **6**(2), 311–319 (2013)
20. Wang, F., et al.: Joint registration and segmentation of neuroanatomic structures from brain MRI. Acad. Radiol. **13**(9), 1031–1044 (2006)
21. Wyatt, P.P., et al.: MAP MRF joint segmentation and registration of medical images. MedIA **7**(4), 539–552 (2003)
22. Yezzi, A., et al.: A variational framework for integrating segmentation and registration through active contours. MedIA **7**(2), 171–185 (2003)

Transfer Learning for the Fully Automatic Segmentation of Left Ventricle Myocardium in Porcine Cardiac Cine MR Images

Antong Chen[1(✉)], Tian Zhou[2], Ilknur Icke[3], Sarayu Parimal[4], Belma Dogdas[1], Joseph Forbes[3], Smita Sampath[4], Ansuman Bagchi[1], and Chih-Liang Chin[4]

[1] Applied Mathematics and Modeling, Merck & Co., Inc., Kenilworth, USA
antong.chen@merck.com
[2] Department of Chemistry and Chemical Biology, Rutgers University, New Brunswick, USA
[3] Global Research Information Technology, Merck & Co., Inc., Kenilworth, USA
[4] Merck Sharp & Dohme, Translational Biomarkers, Singapore, Singapore

Abstract. A fully automatic approach for the segmentation of the left ventricle (LV) myocardium in porcine cardiac cine MRI images is proposed based on deep convolutional neural networks (CNN). We trained a 56-layer residual learning CNN (ResNet-56) from scratch on a set of porcine cine MRI images acquired internally, and another CNN via transfer learning by fine tuning a network previously trained on a public human cine MRI dataset. A leave-one-out validation was performed on an 8-specimen porcine cardiac cine MRI dataset (3,600 slices). Comparisons with manual segmentations show that both CNN models are able to produce precise results (99.94% "good" segmentations), while the CNN trained through transfer learning performs better by achieving Dice similarity coefficient (DSC) of 0.86, Hausdorff distance (HD) of 4.01 mm, and overall average perpendicular distance (APD) of 1.04 mm on average.

Keywords: Cardiac imaging · Transfer learning
Convolutional neural networks

1 Introduction

Short axis cine MRI has been one of the most important structural and functional MRI protocols owing to its capability to examine the entire heart spatially and the full cardiac cycle temporally. In pharmaceutical research, particularly in development of translational cardiac therapies via preclinical animal models, cardiac cine MRI studies are often conducted on porcine (pig) hearts due to their anatomical and pathological likeness to human hearts [1]. As in the standard practice on human scans, a comprehensive study on porcine cine MRI requires delineation of the complete myocardium boundaries (epicardium and endocardium), on all cine MRI frames, leading to a labor-intensive and error-prone task of manually delineating a large number of contours.

M. Pop et al. (Eds.): STACOM 2017, LNCS 10663, pp. 21–31, 2018.
https://doi.org/10.1007/978-3-319-75541-0_3

To alleviate human workload and reduce intra- and inter-rater variability, various automatic image segmentation techniques have been introduced for the delineation of the myocardium in human cine MRI images, which could also be useful on the similar porcine scans. As a plethora of approaches have been reviewed in recent years [2], most of the fully automatic approaches can be categorized as intensity-based, model-based, or atlas-based. Intensity-based methods often group similar pixels into structures by growing image regions or propagating deformable contours/surfaces. These methods are easy to implement, however they are often undermined by image heterogeneities and noise. Model-based approaches extract the statistical distribution of shape and intensity from a well-segmented dataset and find the optimal statistical fit in the new image space. Atlas-based approaches find correspondences between the new and selected image(s), known as the atlas(es), through rigid and/or deformable transformations to propagate segmentations onto the new image, and the consensus of multiple atlases is generally known to yield more precise segmentations. Model-based and atlas-based methods are usually much more robust against heterogeneities because they incorporate strong prior knowledge, although performance can be highly sensitive to the selection of fitting mechanisms and registration algorithms, respectively. Recent advances in deep neural networks also led to novel approaches for cine MRI segmentation, either used as the primary classification model [3–5], or combined with existing techniques to improve the performance [6–8].

Herein we introduce a fully automatic deep convolutional neural network (CNN) approach to cardiac LV myocardium segmentation on cine MRI images. We utilize a state-of-the-art 56-layer residual learning network (ResNet) [9] trained on porcine cine MRI images acquired in house, in addition to a network fine-tuned from an existing CNN. A leave-one-out validation study is performed on 3,600 cine MRI slices from 8 specimens. Model predictions are compared to the manual segmentations using 2D Dice similarity coefficient (DSC), Hausdorff distance (HD), and average point-to-curve distance (APD) as measures.

2 Method

2.1 Data Description

The porcine cardiac cine MRI images were obtained in house via a multi-phase steady state free precession acquisition using the following protocol: Field of view: 290×290 mm^2, imaging matrix: 256×256, slice thickness: 8 mm, echo spacing: 3.6 ms, bandwidth: ± 125 kHz, TR/TE: 36.4 ms/1.6 ms. Scans were performed on 8 porcine specimens (all female, weight = 52 ± 6 kg, age = 4–6 months) longitudinally at three time points, nominally week 0, 1, and 4, as a surgery to introduce infarction was performed for each specimen after the baseline (week 0) scan. At each time point, scans were performed at about 13 consecutive short axis locations to cover the whole heart with 25 frames acquired at each location to cover the complete cardiac cycle. Manual delineations of both the endocardium and epicardium were performed by one of our co-authors (SP) on 6 successive locations from the base to the apex of the LV. Contours were reconstructed and saved as binary masks for the myocardium to jointly

form a dataset consisting of 3,600 image-manual segmentation pairs. All intervention and imaging experiments were reviewed and approved by the IACUC of Merck & Co., Inc. (West Point, PA, USA) and National University of Singapore.

2.2 Image Preprocessing

To eliminate influence from structures outside the heart region, image slices were resampled to 1 × 1 mm^2 pixel size and cropped symmetrically to 184 × 184 matrices. Images were then enhanced through contrast limited adaptive histogram equalization (CLAHE, www.opencv.org). Further LV localization was then performed using a method similar to the one proposed in [10]. Specifically, for each pixel, its intensity over the entire cardiac cycle was extracted and low pass filtered based on the Fourier transform. Following an inverse Fourier transform, pixels associated with heart motion were identified and enclosed by a bounding box. To exclude RV motion, a Hough transform filter was applied in this box to localize the LV center (x_c, y_c) as

$$(x_c, y_c) = \arg\max_{x,y} \sum_{n=1}^{N} -\exp(((x - x_n)^2 + (y - y_n)^2)/\sigma^2) \quad (1)$$

where (x_n, y_n) was the center of the nth circle among all the N detected circles, and the variance σ^2 controlled the decay of the kernels placed in the circle. In the last step, the final cropping box was assigned to contain all circles whose centers were within 8-pixel radius to the LV center, with a 10% margin added to both directions. Pixels outside the cropping box were set to 0 to maintain the 184 × 184 matrix size.

2.3 CNN Architecture and Training Setup

The ResNet deep CNN [9] was constructed with units shown in the left panel of Fig. 1, where two convolutional layers were connected through batch normalization (BN) and a rectified linear unit (ReLu). Unlike most conventional CNN architectures that stack convolutional layers directly, ResNet allows the input to the unit to be fed forwardly through a shortcut and added to the output of the two convolutional layers. The shortcut allows the convolutional layers to learn residual functions with reference to the layers' input, thereby eliminates the degradation problem in training deeper networks and allows easy optimization.

In the right panel of Fig. 1, each of the colored blocks 1-3 represents a concatenation of 9 units with N = 16, 32, or 64 3 × 3 convolution filters, respectively. Note that when switching to the next colored block, the number of convolution filters is doubled and the feature map is downsampled by a factor of 2 by increasing the convolution stride to 2 in the very first convolutional layer. Finally, two deconvolution layers were added to upsample the feature maps and a Softmax layer was added to generate the segmentation. Excluding the Softmax layer, the network contained a total of 56 convolutional and deconvolutional layers.

For training the CNN, (−DSC) between the manual and estimated segmentations was minimized as the loss function with an Adam optimizer [11] using an initial learning rate of 0.001 subsequently reduced by a factor of 10 every 50,000 iterations. A highly regularized 1-pixel deformation was introduced for training data

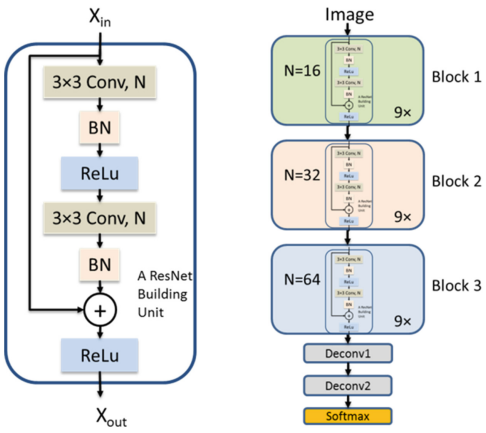


Fig. 1. One ResNet building unit (left), and a ResNet-56 CNN architecture (right)

augmentation. The training was set to run for 50 epochs, although by observation the loss function would reach a steady state after around 35 epochs.

2.4 Transfer Learning

Transfer learning [12] is an alternative to training a network from scratch, allowing the use of a previously trained network as a feature extractor or fine-tuning with new training data even when the size of this training dataset is limited as it is in our case. We acquired a ResNet CNN trained with around 9,300 short axis cardiac cine MRI images from about 500 human subjects [13] for segmenting only the LV endocardium. In our application, although highly similar, cardiac images of the two species could show discrepancies due to their anatomical differences [14], e.g. the appearance of the surrounding tissue caused by the heart orientation due to the unguligrade posture of specimens. Therefore, to receive better performance, we elected to fine-tune the network by removing the Softmax layer from the original network and continuing the training on porcine images with our myocardium labels. The training was started with the same parameters as it was described in Sect. 2.3, and as we expected fast convergence in transfer learning, the ending epoch was set to 30 and the learning rate was set to decay by a factor of 10 every 10,000 iterations instead.

3 Experiments and Results

Under a leave-one-out framework, the images from one of the 8 specimens were set as the testing dataset in each run, and the remaining images were divided into training and validation sets at an 80/20 ratio through random sampling and then fed into the CNNs for training. In order to improve the accuracy and robustness, an ensemble was composed by averaging results from 5 same CNNs initialized with different random seeds. The averaged results were thresholded at intensity 127 out of 255 and saved.

For validation, segmentation results obtained using both CNN models, i.e. CNN trained from scratch using porcine images, denoted CNN_P, and CNN transferred from the network trained by human images and fine-tuned using porcine images (denoted CNN_T), are compared with the manual delineations. Metrics for the evaluation are Dice similarity coefficient (DSC) [15] which evaluates the overlap of the automatic and manual segmentation masks, Hausdorff distance (HD) [16] which evaluates the maximum distance between the two segmentation contours, and average perpendicular distance (APD) [17] which evaluates the average distance between the two contours. Based on these metrics, Table 1 shows results for all 8 specimens, Table 2 demonstrates results with respect to the 6 short axis locations from the base (location 1) to the apex (location 6) of the ventricle, and Table 3 summarizes the entire study and shows the percentage of slices with high segmentation quality. Qualitative examples are given in Fig. 2.

Table 1. DSC, HD, and APD between automatic and manual segmentations for the 8 porcine specimens under a leave-one-out framework. DSC is unitless and all the distance measures are in mm. Better results between CNN_P and CNN_T that are statistically significant ($p < 0.01$ in paired t-tests) are in bold.

Specimen No.		1	2	3	4	5	6	7	8
DSC	CNN_P	0.88 (0.04)	**0.88 (0.06)**	0.84 (0.12)	0.85 (0.06)	0.88 (0.06)	**0.85 (0.05)**	0.87 (0.06)	0.79 (0.10)
	CNN_T	**0.89 (0.04)**	0.87 (0.07)	0.84 (0.13)	0.85 (0.06)	**0.89 (0.04)**	0.83 (0.12)	**0.88 (0.05)**	**0.85 (0.06)**
HD	CNN_P	3.36 (1.11)	3.57 (2.18)	5.43 (3.05)	3.41 (0.98)	3.97 (2.23)	**3.94 (2.30)**	3.60 (1.51)	17.49 (10.81)
	CNN_T	3.30 (1.26)	3.61 (2.12)	5.39 (2.92)	3.37 (0.85)	**3.57 (1.23)**	4.52 (3.58)	**3.37 (1.23)**	**4.95 (5.50)**
APD (endo)	CNN_P	1.06 (0.44)	0.96 (0.54)	0.95 (0.52)	0.94 (0.35)	0.94 (0.63)	0.88 (0.43)	0.88 (0.43)	0.95 (0.45)
	CNN_T	**1.00 (0.45)**	0.95 (0.58)	0.94 (0.48)	0.96 (0.32)	**0.82 (0.39)**	0.90 (0.48)	**0.81 (0.38)**	**0.79 (0.34)**
APD (epi)	CNN_P	0.97 (0.46)	**0.95 (0.41)**	**1.62 (1.25)**	1.10 (0.52)	0.97 (0.38)	**1.15 (0.37)**	1.18 (0.59)	1.10 (0.43)
	CNN_T	**0.93 (0.46)**	1.05 (0.55)	1.65 (1.30)	1.10 (0.50)	0.99 (0.40)	1.27 (0.49)	**1.11 (0.49)**	**0.98 (0.32)**
APD (all)	CNN_P	1.01 (0.40)	**0.95 (0.38)**	**1.35 (0.89)**	1.03 (0.32)	0.94 (0.33)	**1.03 (0.32)**	1.06 (0.46)	1.03 (0.38)
	CNN_T	**0.96 (0.40)**	1.01 (0.49)	1.36 (0.91)	1.04 (0.31)	0.92 (0.30)	1.12 (0.41)	**0.98 (0.39)**	**0.89 (0.26)**

The CNN models were implemented using the MXNet deep learning framework [18] on NVIDIA Tesla K80 GPU accelerated hardware. Each training run of 50 epochs took $\sim$7 h while a 30-epoch transfer learning run completed within 4.5 h. When applying a trained model on a new specimen, each run on 450 images consumed 30–40 s.

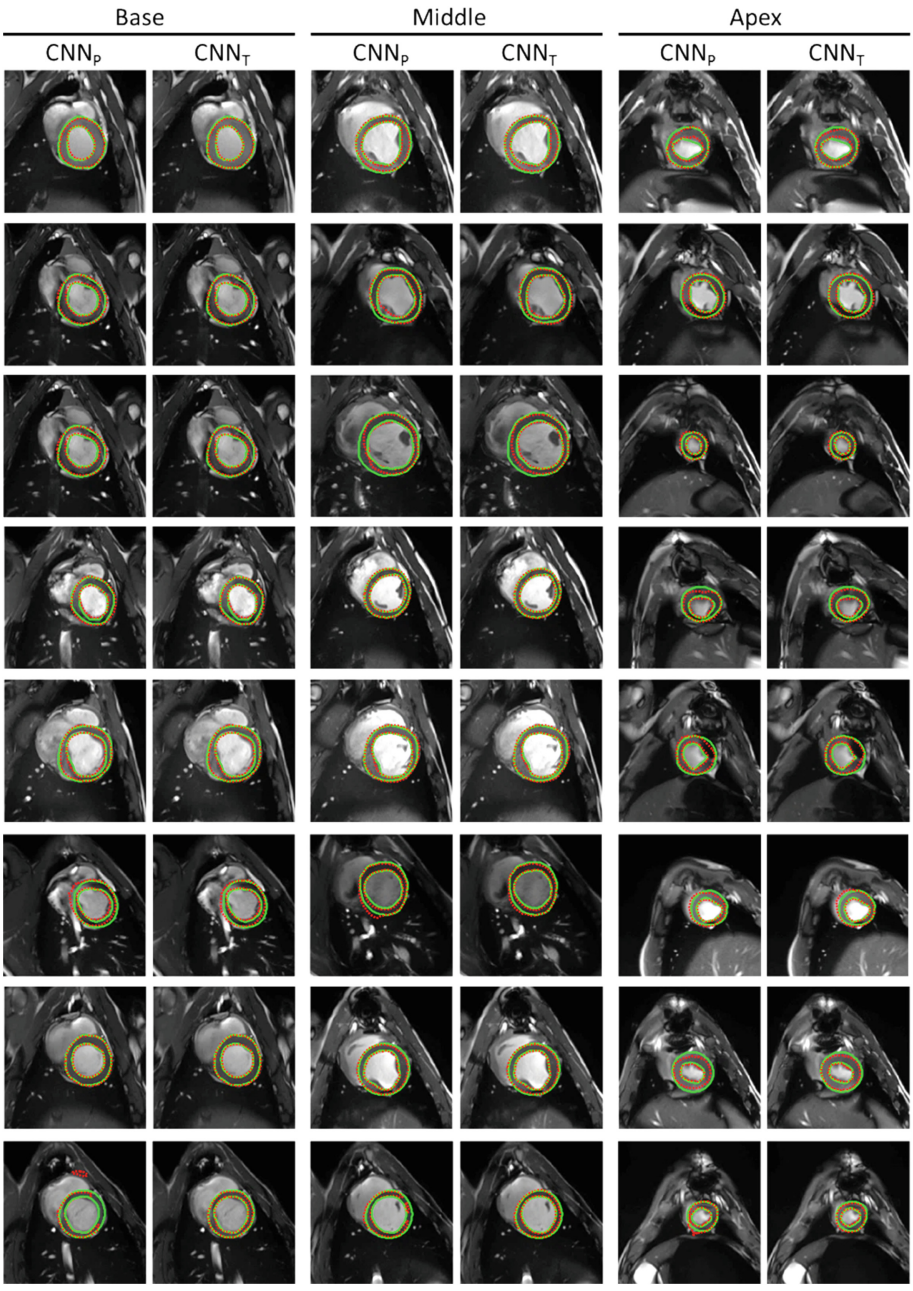


Fig. 2. Automatic delineations obtained using CNN_P and CNN_T (in dotted red lines) compared with manual delineations (in solid green lines) for cine image slices at the base, middle, and apex of the LV. Each row from row 1 (top) to 8 (bottom) is for one specimen from no. 1 to 8. (Color figure online)

Table 2. DSC, HD, and APD between automatic and manual segmentations for the 6 short axis locations from the base to the apex of the left ventricle under a leave-one-out framework. DSC is unitless and all the distance measures are in mm. Better results between CNN_P and CNN_T that are statistically significant ($p < 0.01$ in paired t-tests) are in bold.

Location No.		1	2	3	4	5	6
DSC	CNN_P	0.84 (0.08)	0.89 (0.06)	0.89 (0.06)	0.87 (0.05)	0.85 (0.06)	0.79 (0.10)
	CNN_T	**0.85 (0.06)**	**0.90 (0.04)**	**0.90 (0.04)**	**0.88 (0.04)**	**0.85 (0.06)**	0.79 (0.13)
HD	CNN_P	6.96 (7.83)	5.30 (7.39)	4.24 (5.50)	4.78 (4.82)	5.14 (4.23)	7.17 (6.21)
	CNN_T	**4.40 (2.61)**	**3.41 (2.91)**	**3.43 (3.24)**	**3.51 (2.00)**	**4.02 (2.01)**	**5.30 (3.61)**
APD_ (endo)	CNN_P	1.24 (0.54)	0.73 (0.31)	0.69 (0.24)	0.85 (0.33)	0.97 (0.41)	1.20 (0.63)
	CNN_T	**1.15 (0.54)**	**0.67 (0.26)**	0.69 (0.23)	**0.81 (0.31)**	0.97 (0.42)	**1.10 (0.52)**
APD (epi)	CNN_P	1.19 (0.44)	0.88 (0.31)	0.84 (0.26)	**0.99 (0.39)**	1.23 (0.62)	1.64 (1.09)
	CNN_T	1.17 (0.45)	0.87 (0.30)	0.85 (0.27)	1.01 (0.39)	1.23 (0.63)	1.69 (1.13)
APD (all)	CNN_P	1.20 (0.39)	0.81 (0.24)	0.78 (0.19)	0.93 (0.27)	1.12 (0.45)	1.45 (0.75)
	CNN_T	**1.16 (0.41)**	**0.78 (0.20)**	0.78 (0.20)	0.92 (0.26)	1.12 (0.46)	1.45 (0.77)

Table 3. Summary for all segmentations. Better results that are statistically significant ($p < 0.01$ in paired t-tests) are in bold. Percentage of slices in high quality standards are also shown.

	DSC	HD	APD (endo)	APD (epi)	APD (all)
CNN_P	0.85 (0.08)	5.60 (6.23)	0.95 (0.48)	1.13 (0.65)	1.05 (0.48)
CNN_T	**0.86 (0.08)**	**4.01 (2.87)**	**0.90 (0.44)**	1.13 (0.67)	**1.04 (0.49)**
	DSC $\geq$ 0.90	DSC $\geq$ 0.80	HD $\leq$ 3 mm	APD (all) $\leq$ 1 mm	APD (all) $\leq$ 5 mm
CNN_P	29.97%	82.14%	42.50%	57.20%	99.94%
CNN_T	32.83%	86.50%	47.39%	59.03%	99.94%

4 Conclusion and Discussions

Both models (CNN_P and CNN_T) performed adequately well considering both led to 99.94% success rate (Table 2) on LV segmentation in porcine cine MRI slices, which is categorized as "good" by the convention of APD $\leq$ 5 mm [22]. However, CNN_T excelled in almost all accuracy measures (higher DSC, lower APDs, and substantially higher HD) at a lower computation cost. It is also observed that CNN_T produced more

regularized contours than CNN_P, which is an important advantage against image heterogeneities (Fig. 2). Intensity anomalies such as darker LV on specimen 8, caused CNN_P to produce incomplete contours, and part of the segmentation fell into the myocardium of the right ventricle, while CNN_T performed significantly better as it is shown in Table 1. This shows how transfer learning from a more comprehensive model can prevent overfitting in the case of limited training data. Furthermore, both models were robust against the pathological cases (here in the form of LV enlargement after week 0), as segmentations of week 1 have only 0.02 (CNN_P) and 0.01 (CNN_T) less mean DSC than that of week 0, and mean DSCs for week 0 and 4 are comparable.

Location-wise assessment (Table 2) showed that the lowest accuracy was observed at the apex (location 6) region. CNN_T achieved mean DSC > 0.85 and mean APD < 1.16 mm for LV locations 1-5, while the mean DSC for location 6 was 0.79 and the mean APD was 1.45. This observation is expected since segmenting the apex would be particularly difficult due to the small volume of the myocardium at the location.

Even though a head-to-head evaluation against existing approaches is challenging due to different input datasets, a general comparison with reported results [2] shows that our approach performs on par with the most accurate ones (Table 4 upper section). Our approach only requires regular manual delineation of the training set and a simple preprocessing before the training, which is more straightforward and less prone to potential bias introduced by human interactions than other methods. Compared with other primarily deep learning-based methods (Table 4 lower section), our segmentations are more accurate in general, due to its capabilities to learn more profound feature combinations through a deeper network and training based on a more comprehensive set of features extracted from a richer dataset via transfer learning.

Although the overall performance of CNN_T is higher than that of CNN_P, the specimen-wise comparison shows that not all cases received improvements especially when using APD (epi) and APD (all) as measures of accuracy. For the cases showing statistically significant improvements the difference was not particularly large except for the case of specimen 8. The limiting factor for the CNN_T here could be the moderately large human cardiac cine MRI dataset on which the pre-trained CNN was obtained, whose size was only less than 3 times the size of the porcine dataset. Due to the different purposes of the studies, the human dataset only contained delineation of the diastole and systole frames for each subject [13], which would not be able to represent the complete range of dynamics in the cardiac cycle. Moreover, the lack of epicardial contours in the human dataset could limit the improvement of performance in the transfer learning since the continuing training on the porcine dataset required a relabeling of the region of interest. When experimenting on a more comprehensive dataset that is substantially larger than the human dataset here, the size limitation of the human dataset could become even more evident. To further evaluate the robustness of the transfer learning approach, we are acquiring images on more porcine specimens, which together with the existing porcine data here will form a larger dataset for future experiments. Also we are seeking for opportunities to test on properly labeled public databases that are of larger scale. With substantially expanded datasets, training data would be able to cover more variability, and it could become feasible to construct models that are more specific to the different locations of the ventricle.

Table 4. A subset of up-to-date methods and their accuracy. Rows 3-6 are deep learning based.

Authors	Method description	Mean accuracy
Alba *et al.* [19]	Graph cuts with smoothness and inter-slice constraints	Endo: APD = 2.76 mm Epi: APD = 2.58 mm
Bai *et al.* [20]	Multi-atlas with local patch-based label fusion	Endo: APD = 1.26 mm, HD = 7.27 mm Epi: APD = 1.49 mm, HD = 9.35 mm
Zhu *et al.* [21]	3D inter-subject & intra-cycle shape models with manual mesh refinement	Endo: MAD = 0.69 mm (3D) Epi: MAD = 1.27 mm (3D)
Li *et al.* [3]	3D fully convolutional network with deeply supervised connections	Myocardium DSC = 0.70
Tran [4]	17-layer CNN with max pooling	Endo: DSC = 0.84, HD = 8.86 mm Epi: DSC = 0.86, HD = 9.33 mm
Poudel *et al.* [5]	U-Net [22] based CNN with recurrent connection at the most compact layer	Endo: DSC = 0.94, APD = 1.56

It is noticed that the CNN models herein were designed to process 2D cine MRI slices with each slice treated as an independent data point. Although a 3D solution could be achieved by extending the 2D convolutional models into 3D, the large slice thickness could hinder its performance, which is a problem that could not be solved easily through 3D resampling. An alternative solution is to incorporate 3D slice-wise correlation via a recurrent neural network (RNN) like what was proposed in [5]. An RNN model could also be introduced to describe the dynamics of heart motion throughout the cardiac cycle [23]. Both RNN-based solutions could lead to a 2.5D-plus-time model that could further improve the myocardium segmentation accuracy. The segmentation result will be used to extract cardiac LV biomarkers that can help progressively study structural remodeling of the heart during heart failure and develop novel therapeutics for treatment of cardiac disease [24].

References

1. Suzuki, Y., Yeung, A.C., Ikeno, F.: The representative porcine model for human cardiovascular disease. Biomed Res. Int. **2011**, 1–10 (2010)
2. Peng, P., Lekadir, K., Gooya, A., Shao, L., Petersen, S.E., Frangi, A.F.: A review of heart chamber segmentation for structural and functional analysis using cardiac magnetic resonance imaging. Magn. Reson. Mater. Phy. Biol. Med. **29**(2), 155–195 (2016)
3. Li, J., Zhang, R., Shi, L., Wang, D.: Automatic whole-heart segmentation in congenital heart disease using deeply-supervised 3D FCN. In: Zuluaga, M.A., Bhatia, K., Kainz, B., Moghari, M.H., Pace, D.F. (eds.) RAMBO/HVSMR -2016. LNCS, vol. 10129, pp. 111–118. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-52280-7_11

4. Tran, P.V.: A fully convolutional neural network for cardiac segmentation in short-axis MRI. arXiv preprint arXiv:1604.00494 (2016)
5. Poudel, R.P., Lamata, P., Montana, G.: Recurrent fully convolutional neural networks for multi-slice MRI cardiac segmentation. arXiv:1608.03974 (2016)
6. Zhen, X., Wang, Z., Islam, A., Bhaduri, M., Chan, I., Li, S.: Multi-scale deep networks and regression forests for direct bi-ventricular volume estimation. Med. Image Anal. **30**, 120–129 (2016)
7. Ngo, T.A., Lu, Z., Carneiro, G.: Combining deep learning and level set for the automated segmentation of the left ventricle of the heart from cardiac cine magnetic resonance. Med. Image Anal. **35**, 159–171 (2017)
8. Avendi, M.R., Kheradvar, A., Jafarkhani, H.: A combined deep-learning and deformable-model approach to fully automatic segmentation of the left ventricle in cardiac MRI. Med. Image Anal. **30**, 108–119 (2016)
9. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778 (2016)
10. Lin, X., Cowan, B.R., Young, A.A.: Automated detection of left ventricle in 4D MR images: experience from a large study. In: International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 728–735 (2006)
11. Kingma, D., Ba, J.: Adam: a method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014)
12. Pan, S.J., Yang, Q.: A survey on transfer learning. IEEE Trans. Knowl. Data Eng. **22**(10), 1345–1359 (2010)
13. Zhou, T., Icke, I., Dogdas, B., Parimal, S., Sampath, S., Forbes, J., Bagchi, A., Chin, C., Chen, A.: Automatic segmentation of left ventricle in cardiac cine MRI images based on deep learning. In: Proceedings of SPIE 10133, Medical Imaging: Image Processing (2017)
14. Crick, S.J., Sheppard, M.N., Ho, S.Y., Gebstein, L., Anderson, R.H.: Anatomy of the pig heart: comparisons with normal human cardiac structure. J. Anat. **193**(1), 105–119 (1998)
15. Dice, L.R.: Measures of the amount of ecologic association between species. Ecology **26**(3), 297–302 (1945)
16. Huttenlocher, D.P., Klanderman, G.A., Rucklidge, W.J.: Comparing images using the Hausdorff distance. IEEE Trans. Patt. Anal. Mach. Intell. **15**(9), 850–863 (1993)
17. Radau, P., Lu, Y., Connelly, K., Paul, G., Dick, A., Wright, G.: Evaluation framework for algorithms segmenting short axis cardiac MRI. MIDAS J. Card. MR Left Ventricle Segmentation Challenge **49**, 134 (2009)
18. Chen, T., Li, M., Li, Y., Lin, M., Wang, N., Wang, M., Xiao, T., Xu, B., Zhang, C., Zhang, Z.: Mxnet: a flexible and efficient machine learning library for heterogeneous distributed systems. arXiv preprint arXiv:1512.01274 (2015)
19. Alba, X., Ventura, F., Rosa, M., Lekadir, K., Tobon-Gomez, C., Hoogendoorn, C., Frangi, A.F.: Automatic cardiac LV segmentation in MRI using modified graph cuts with smoothness and interslice constraints. Magn. Reson. Med. **72**(6), 1775–1784 (2014)
20. Bai, W., Shi, W., O'Regan, D.P., Tong, T., Wang, H., Jamil-Copley, S., Peters, N.S., Rueckert, D.: A probabilistic patch-based label fusion model for multi-atlas segmentation with registration refinement: application to cardiac MR images. IEEE Trans. Med. Imaging **32**(7), 1302–1315 (2013)
21. Zhu, Y., Papademetris, X., Sinusas, A.J., Duncan, J.S.: Segmentation of the left ventricle from cardiac MR images using a subject-specific dynamical model. IEEE Trans. Med. Imaging **29**(3), 669–687 (2010)

22. Ronneberger, O., Fischer, P., Brox, T.: U-net: convolutional networks for biomedical image segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 234–241 (2015)
23. Xue, W., Nachum, I.B., Pandey, S., Warrington, J., Leung, S., Li, S.: Direct estimation of regional wall thicknesses via residual recurrent neural network. In: Niethammer, M., Styner, M., Aylward, S., Zhu, H., Oguz, I., Yap, P.-T., Shen, D. (eds.) IPMI 2017. LNCS, vol. 10265, pp. 505–516. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-59050-9_40
24. Parimal, S., Sampath, S., Mazlan, I., Croft, G., Totman, T., Zheng, Y.T.W., Manigbas, E., Klimas, M., Evelhoch, J.L., Kleijn, D.P.V., Chin, C.: Early prediction of chronic infarct size by acute strain: a cardiac MRI study of myocardial infarction in pigs. In: SMRT 26th Annual Meeting of International Society of Magnetic Resonance in Medicine (2017)

Left Atrial Appendage Neck Modeling for Closure Surgery

Cheng Jin(✉), Heng Yu(✉), Jianjiang Feng(✉), Lei Wang(✉), Jiwen Lu(✉), and Jie Zhou(✉)

Tsinghua National Laboratory for Information Science and Technology, Department of Automation, Tsinghua University, Beijing, China
{jin-c12,h-yu14,w-l14}@mails.tsinghua.edu.cn,
{jfeng,lujiwen,jzhou}@tsinghua.edu.cn

Abstract. The left atrial appendage closure surgery is the main treatment of thrombosis in patients with atrial fibrillation. Left atrial appendage neck is the region of implanting occluder in the surgery. The occluder is strictly matched with the neck to prevent the occluder piercing the endocardium or falling off and avoid threatening patients' lives. So we build a model of left atrial appendage neck based on the segmentation result of maximal volume phase from CT data. The model automatically generated by our approach is a circumscribed cylinder to represent the irregular columnar neck. This circumscribed cylinder can help to determine the diameter of the closure device before the surgery, the implanted position and pose of the device during the surgery. Specifically, we successfully solved these problems including the auto-detection of the boundary points of left atrial appendage ostium, the establishment of the standard coordinate system, the auto-calculation of the cylinder height and the polychromatic display of occlusive tension on the neck. We built our model on 100 patients' data and dissected the three pig hearts to do comparative experiments. Tests were performed in 67 occlusion surgeries with the success rate of 97.01%. These indicate that our approach can precisely and non-invasively model the left atrial appendage neck for assisting closure surgeries.

Keywords: Left atrial appendage neck · Closure surgery
Circumscribed cylinder · Standard coordinate system
Occlusive tension

1 Introduction

The left atrial appendage (LAA) is a substructure of the cardiac connected to the left atrium (LA). It is often tubular, hooked and with a few lobes. As a cardiovascular disease, atrial fibrillation (AF) leads to LAA's emptying obstruction and induces thrombosis. About 80% of the cardiogenic thrombi navigate to the brain by the coursing blood, which causes strokes [1]. The LAA closure can reduce the risk of strokes in patients with AF. During closure surgery, the

M. Pop et al. (Eds.): STACOM 2017, LNCS 10663, pp. 32–41, 2018.
https://doi.org/10.1007/978-3-319-75541-0_4

clinician implants the closure device into LAA neck to avoid the deposition of thrombi. The key of the surgery is the closure device, whose parameters should exactly consistent with the size, shape and contraction of LAA neck. Improper parameters choices for the closure device would lead to surgical failure. If the size is too large, the cardiac membrane would burst, while too small the device would slip off. Improper position and pose of the closure device could also lead to the same consequence. We address the closure problems based on CT data.

Studying the small and variable structure of the LAA is a tough job on the CT data. Few researchers focus on the LAA neck in the image processing domain. Fan *et al*. [2] proposes an occlusion procedure for a double-lobed LAA using 3-D printing, but it is lack of practical use. The LAA neck ostium is the boundary between the LA and LAA. In clinical medicine, the boundary is defined as the link between the left superior pulmonary vein (LSPV) ostium and the mitral valve (MV). In the image processing domain, one possible method to draw the boundary is through the surface curvature changes between LA and LAA [3]. In addition, the measurement of the LAA neck is usually confined to the 2-D space, which cannot truly reflect the 3-D structure. However, the neck structure must be accurately understood before LAA closure surgery to improve the success rate. Thus, to provide a precise solution, we fit a circumscribed cylinder to the local segmentation model of the LAA neck to achieve precise measurement. And the tension level between the closure device and LAA neck is also calculated and displayed. The height and placement angle of the cylinder are optimized for the selection of closure device diameter before the surgery and both the placement and pose of the closure device in the surgery.

2 LAA Segmentation

Our group has already segmented the LAA precisely in the 45% phase based on learning to rank segmentation proposals [4]. Compared to the model based approaches such as ASM, this segmentation approach can handle large LAA shape variations because of no explicit shape constraints. Compared to graph-cut, this approach can set seed points automatically and generates a pool of segmentation proposals to pick out the best segmentation, instead of generating a single result which cannot guarantee good performance under all image conditions.

3 LAA Neck Modeling

After the segmentation of LAA, we focus on the LAA model's neck where closure device is implanted in closure surgery. In fact, there is no obvious anatomical boundary between the LAA and LA. The appropriate position of closure in this region is deemed as the beginning of the neck of the LAA. The shape of the neck is an irregular columnar. The purpose of neck modeling is twofold. On the one hand, the model can assist clinicians to learn more about the size and pose of the neck before closure surgery [5,6]. On the other hand, the model can be

fused with real-time three-dimensional transesophageal echocardiography (RT-3D-TEE) to provide real-time and stable navigation during the surgery. We select the phase with the largest LAA volume in the whole cardiac cycle as the experimental subject and model its neck. The modeling process consists of three steps: (1) Auto-detection of the ostium of the LAA; (2) Establishment of the standard coordinate system based on the ostium plane; (3) Auto-building of circumscribed cylindrical model of LAA neck.

3.1 Auto-Detection of the Ostium of the LAA

In practice, we prefer to find the suitable location for closure rather than the accurate neck border. There is a large surface curvature change in the transitional region between the LA and LAA. So we propose a self-adaptive method based on optimization to find a smooth closed boundary of the highest curvature, which is the ostium of the LAA neck. The specific method is as follows [3].

We set a proximal ring of LAA and calculate the center C of the ring by performing high-density resampling in the transitional region. Then a plane perpendicular to the ring is determined, which passes through R_i (a point on the ring) and C. The plane rotates around the axis which is the normal vector passing C. The intersection of these planes with the surface of LAA are produced by high-density resampling and they are the contour lines of LAA, as shown by the green part in the Fig. 1(a). We constrain the longitudinal size of the contours to 31 mm (the length of transitional region between the LA and LAA is about 20–30 mm anatomically), and then calculate the sum of the maximum curvature of all sampling points on every contour line.

$$E = (S_0^{J(0)}, \cdots, S_{n-1}^{J(n-1)}) = \arg \max_{J(0),\cdots,J(n-1)} \sum_{i=0}^{n-1} C(S_i^{J(i)}). \tag{1}$$

Here, S_i^j indicates the jth point on the ith contour line and $S_i^0 = R_i$. $C(S_i^j)$ is the curvature at point S_i^j, which is defined as

$$C(S_i^{J(i)}) = ||(S_i^{J(i+1)} - S_i^{J(i)}) - (S_i^{J(i)} - S_i^{J(i-1)})||. \tag{2}$$

When we calculate the curvature using the differential geometry, the second order derivatives of the contour line need to be obtained, and the calculated curvature sometimes tilts in the wrong direction. The final boundary may be stuck far away from the LAA in some data. Therefore, in order to improve robustness, we add a bias, which is the 2-norm of the inner product of N and $(S_i^{J(i)} - S_i^0)$, to pull the boundary toward the LAA.

$$||N \cdot (S_i^{J(i)} - S_i^0)||. \tag{3}$$

N is the normal vector of the ring. The weight ω of the bias is set to 0.002 by experience. To ensure the smoothness of the boundary, we require the 2-norm of the adjacent sampling point J_i, $J_{(i+1)}$ less than 1. Similarly, to ensure the boundary is closed, we require the 2-norm of the start and end points less than 1.

$$||J(i) - J(i+1)|| \leq 1, \qquad ||J(n-1) - J(0)|| \leq 1. \tag{4}$$

$$E = \arg \max_{J(0),\cdots,J(n-1)} \sum_{i=0}^{n-1} C(S_i^{J(i)}) + \omega ||N \cdot (S_i^{J(i)} - S_i^0)||. \tag{5}$$

We obtain the corresponding boundary (as shown by the blue line in Fig. 1(a), which is the ostium of the LAA neck. The normal direction of the boundary cross-section goes along with the trend of LAA neck, as shown in Fig. 1(b) and (c).

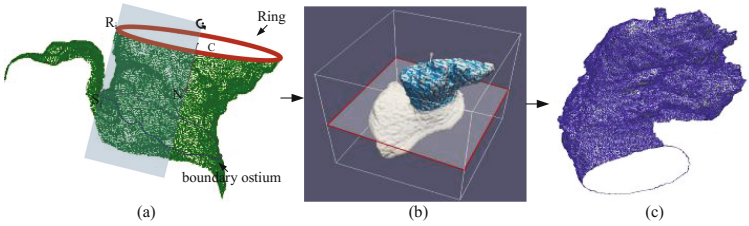


Fig. 1. Cutting off the LA to extract the LAA. (Color figure online)

3.2 Establishment of the Standard Coordinate System Based on the Ostium Plane

The establishment of a standard coordinate system is the basis for the automatic building of the LAA neck model, and also convenient for RT-TEE during surgery (the further extension of our work). To establish a standard coordinate system, we define four necessary elements: origin O, X axis, Y axis and Z axis.

We first find a point, which is nearest to all points of the ostium boundary in terms of Euclidean distance. This point is called the centroid O, that is:

$$\sum dist(B, O) \rightarrow MIN, \tag{6}$$

where B is an arbitrary point on the boundary, O is the centroid. The centroid is the origin of the coordinate system. X axis is defined as the line along with which we have the longest chord of the ostium closure curve through origin O. Y axis is defined as the line vertical to the X axis through O point. By generating the normal vector n of the ostium plane through origin O, the vector direction is defined as the orientation of Z axis. Thus the standard coordinate system is established, as shown in Fig. 2.

All points of the LAA model are mapped to this coordinate system as follows:

$$P' = R(P - C), \tag{7}$$

$$R = [\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{Z}]^T, \tag{8}$$

where P is the point in the original coordinate, P' is the point in the mapped coordinate, R is the rotation matrix and C is the translation matrix. In this way, the coordinates of all LAA model points are normalized.

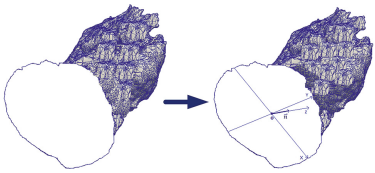


Fig. 2. The diagram of standard coordinate system.

3.3 Auto-Building of Circumscribed Cylindrical Model of LAA Neck

We use the LAA ostium plane as the basal plane of the circumscribed cylindrical, which is the first slice of the neck irregular cylinder. We traverse each slice along the Z axis to figure out the centroid and area. We evaluate each slice to calculate the cylinder height automatically by the following formula:

$$E = \lg[a(|S_n - \frac{1}{n}\sum_{n=1}^{n} S_n|) + bd_n^2], \tag{9}$$

where S_n is the area of the n-th slice, d_n is the distance from the n-th slice centroid to the Z axis, a and b are the weights, and E is the cut-off evaluation function. Since the distance from the centroid of each slice to the Z axis is the key factor that determines the direction of the neck cylinder, we set the value of b larger than a. We experimentally set b as 0.7, and a as 0.3. Logarithmic function is sensitive to the difference of data, so we use it for evaluating. We define $E_0 = 1.6$.

These constraints are established by visually examining anatomical features of LAA neck in some CT data. The LAA neck generally consists of around 20 stacked slices and the distance between adjacent slices is around 0.4 mm. These slices are approximately coaxial circles, whose radii are around 12 mm and the distances of whose centers and Z axis are less than 4 mm. These constraints have been also tested in our dataset. The traversal ends when E is greater than or equal to E_0. All traversed slices are projected onto the $X{-}Y$ plane. And the minimal circumcircle of these projections is taken as the bottom of the cylinder. The thickness of all the traversed slices is the height of the cylinder. The surface of LAA neck is painted with different colors to characterize the tension. And the tension is indicated by the vertical distance from each point of the LAA to the surface of the cylinder. The distance is calculated by:

$$D_i = |x_{\mathrm{LAA}} - x_{\mathrm{cyl}}|. \tag{10}$$

The tension levels of which the closure devices brace cardiac membrane is plotted. The dark color indicates that the cardiac membrane is easily burst. This can assist the clinician to select the size of the closure devices (Fig. 3).

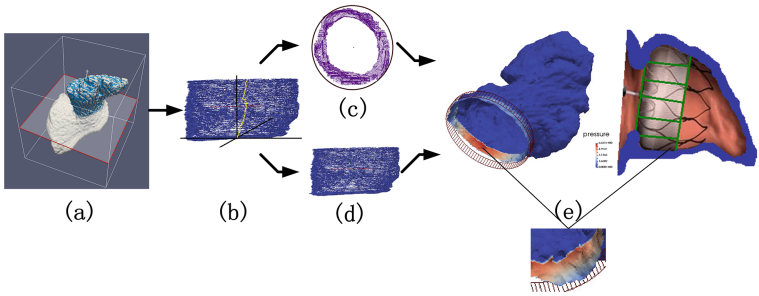


Fig. 3. The circumscribed cylinder establishing process (a), Determining the datum plane, (b) Traversing the slices, (c) Determining the bottom of the cylinder, (d) Determining the cylinder height, (e) Generating the circumscribed cylinder (the colors represent the tensions of the closure devices).

4 Experiments and Results

4.1 Dataset

To show the superiority of the proposed modeling method, we conduct experiments on 100 patients who have comprehensive examinations of Philip 256-iCT (0.35 mm × 0.35 mm × 0.44 mm, 512 × 512, Philips 256-iCT) and transthoracic echocardiography (200 frames TDI images per second, 30 volumes per second, Philips iE33). In addition, we do comparative experiments on the anatomized heart of 3 pigs. And experimental occlusion surgeries are performed on other 67 patients for validation.

4.2 Ground Truth

With radiologists assistant, we annotated the LAA using a paintbrush tool 3-D SLICER 4.4 slice by slice at the voxel level on 100 sets of CT data. To obtain the ostium of LAA neck by calculating the curvature, a small part of the LA inside the region of interest (ROI) was also included. Further the technicians manually measured the shape and size of LAA neck with the corresponding echocardiography. We anatomized 3 pigs hearts to measure their sizes and manually built the neck model. Then we implanted closure in the transitional region between the LA and LAA to evaluate the result of occlusion, although this method is static and does not reflect the contractile characteristics of the living body. Finally, to ensure the clinical feasibility of our approach, we also tested it in 67 LAA closure surgeries.

4.3 Evaluation

We evaluate the advantages of our modeling approach. The proposed modeling approach can provide a precise 3D model of the LAA rather than 2D images

to assist closure surgeries. Figure 4 **L** shows the images with conventional 2D measurement, which is only one slice of the 3D LAA. However, it is difficult to the best slice for the measurement of LAA neck parameters [7]. In contrast, the established 3D LAA model by the proposed method can provide the detailed structure of the LAA neck. Moreover, the color of the circumscribed surface represents the tension between closure devices and cardiac membrane. 3 examples of 100 CT models are shown in Fig. 4 **R**. The detailed LAA neck model can help to choose the appropriate size of closure devices.

Our coordinate system is established on the ostium of LAA rather than conventional arteriae aorta. It's for the convenience of the following: 1. Auto-modeling LAA neck; 2. Real-time fusing of preoperative CT-model and intraoperative ultrasound images in the following LAA occlusion surgery. The establishment of the standard coordinate system can provide guidance for the position and pose of the closure devices. The heights, radii and bottom centers' coordinates of 3D models are listed in Table 1, which correspond to the three examples in Fig. 4 **R**. The presented parameters of the cylinder models are critical for the selection of the closure devices.

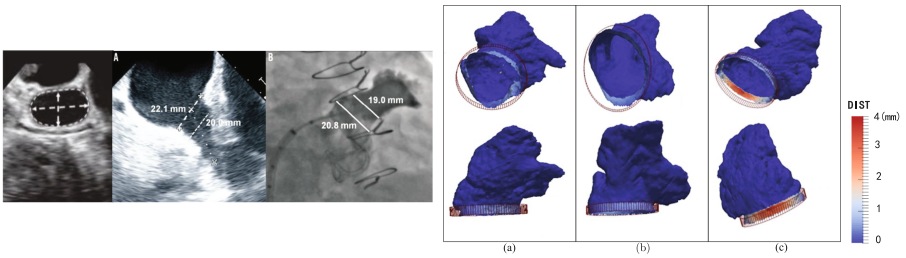


Fig. 4. L The label of the LAA neck in the two-dimensional images. **R** Generating cylindrical model on three different LAA necks (the tension levels of which the closure devices brace cardiac membrane are plotted)

Table 1. The parameters of LAA neck cylinder models

SN	Ostium plane centroid			Radius	Height
	X (mm)	Y (mm)	Z (mm)	(mm)	(mm)
a	4.88	−2.12	0	15.19	12.98
b	1.53	0.63	0	14.30	10.86
c	−1.07	−2.57	0	11.44	8.97

We respectively measure the sizes of LAA necks of 3 pigs and build cylindrical models manually. The average diameter of which is (24±3.9) mm and the average height is (18 ± 4.3) mm (Fig. 5a and b). Then the occluder is implanted into the neck with strong occlude and moderate tension. The average pulling force to

pull it out of the LAA neck is (5 ± 0.67) N (Fig. 5c). The experiment of stress and stretch shows that our approach is anatomically feasible.

Table 2. Comparing the proposed approach with the traditional approach. The compression ratio is $(\mathrm{Dia_{bef}} - \mathrm{Dia_{aft}})/\mathrm{Dia_{bef}}$. $\mathrm{Dia_{bef}}$ is the diameter of occluder before placement, and $\mathrm{Dia_{aft}}$ is the diameter of occluder after placement.

	The traditional approach		The proposed approach
Watchman size [6] (mm)	Max ostium	Occluder size	The occluder diameter approximately equals the diameter of the 3D-model
	17–19	21	
	20–22	24	
	23–25	27	
	26–28	30	
	29–31	33	
Position	Positioning subjectively		Relative position is determined based on the coordinate system
Compression ratio and occlusion effect	Measuring subjectively		The compression ratio is limited to $15\% - 30\%$ based on the distribution of the tension

In the follow-up of 67 cases of LAA closure surgeries, the clinician implants the occluder in the region of the neck based on the size of the 3D-model, the simulated posture of the neck, and the distribution of the tension (Table 2, Fig. 6). The standards of the occluder implantation are: i. Position: the occluder is placed in the LAA neck. ii. Anchor: all anchors have been embedded in the LAA wall. iii. Size: the occluder compression ratio is limited to $15\% - 30\%$. iv. Seal: the occluder holds up the neck, and all LAA lobes are blocked. The occluder selected according to our approach is better than the traditional approach (the occluder diameter is 30% greater than the short diameter of LAA ostium) and the success rate of implantation reached 97.01%. There are two cases of failure because

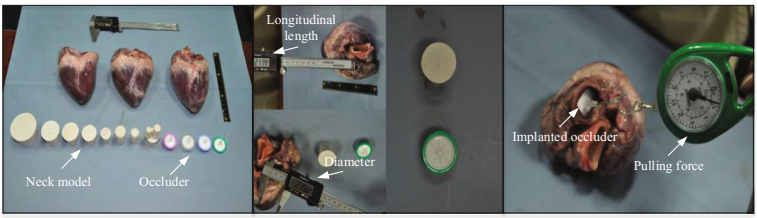


Fig. 5. Simulating the implant of the occluder in the LAA of 3 pigs. (a) 3 hearts of pigs, the obtained 8 types of neck-models based on 100 CT data and 4 simulate occluders. (b) Making the external cylindrical model by hand. (c) The experiment of stress and stretch after the occluder is implanted.

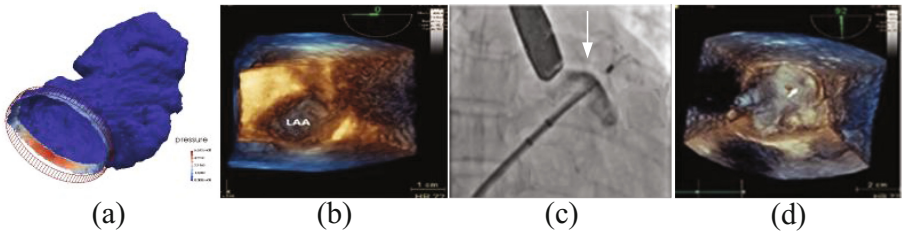


Fig. 6. A case of LAA closure surgery. (a) The preoperative CT-model of LAA (the diameter of the ostium is 26.5 mm, the longitudinal length is 25.3 mm). (b) the intra-operative LAA ostium in 3D-TEE. The 27 mm occluder is chosen based on (a) and (b). (c) There is no leakage of re-injected contrast agent after the occluder is implanted. (d) Observed by 3D-TEE, the ostium is completely closed.

the longitudinal length of LAA neck is too short and two holes immediately appear after the neck, so we should build the "***Y***" model for such special cases. Meanwhile, this static analysis can be improved by looking at the whole cardiac cycle.

5 Conclusion

In this paper, we propose a circumscribed cylinder model with occlusive tension display to fit the irregular columnar LAA neck. This provides the foundation for the choice of the diameter, the implanted position and the pose of the closure devices. In complementary to current work, we can use "***Y***" models with certain curvature in some special cases. In addition, it is very promising to achieve precise surgical navigation by the fusion of CT and RT 3D-TEE, which is considered as the future work.

Acknowledgements. This work is supported by the National Natural Science Foundation of China under Grants 61622207, 61373074, 61225008, and 61572271.

References

1. Yamamoto, M., Seo, Y., et al.: Complex left atrial appendage morphology and left atrial appendage thrombus formation in patients with atrial fibrillation. Circ. Cardiovasc. Imaging **7**(2), 337–343 (2014)
2. Fan, Y., Kwok, K.W., Zhang, Y., Cheung, G.S.H., Chan, A.K.Y., Lee, A.P.W.: Three-dimensional printing for planning occlusion procedure for a double-lobed left atrial appendage. Circ. Cardiovasc. Interv. **9**(3), e003561 (2016)
3. Zheng, Y., Yang, D., et al.: Multi-part modeling and segmentation of left atrium in C-arm CT for image-guided ablation of atrial fibrillation. IEEE Trans. Med. Imaging **33**(2), 318–331 (2014)
4. Wang, L., Feng, J., et al.: Left atrial appendage segmentation based on ranking 2-D segmentation proposals. In: Workshop at the 19th International Conference on Medical Image Computing and Computer Assisted Intervention (2016)

5. Wang, Y., Di Biase, L., et al.: Left atrial appendage studied by computed tomography to help planning for appendage closure device placement. J. Cardiovasc. Electrophysiol. **21**(9), 973–982 (2010)
6. Akinapelli, A., Bansal, O., Chen, J.P., Pflugfelder, A., Gordon, N., Stein, K., Huibregtse, B., Hou, D.: Left atrial appendage closure-the watchman device. Curr. Cardiol. Rev. **11**(4), 334–340 (2015)
7. Budge, L.P., Shaffer, K.M., Moorman, J.R., Lake, D.E., Ferguson, J.D., Mangrum, J.M.: Analysis of in vivo left atrial appendage morphology in patients with atrial fibrillation: a direct comparison of transesophageal echocardiography, planar cardiac CT, and segmented three-dimensional cardiac ct. J. Intervent. Cardiac Electrophysiol. **23**(2), 87–93 (2008)

Detection of Substances in the Left Atrial Appendage by Spatiotemporal Motion Analysis Based on 4D-CT

Cheng Jin(✉), Heng Yu(✉), Jianjiang Feng(✉), Lei Wang(✉), Jiwen Lu(✉), and Jie Zhou(✉)

Tsinghua National Laboratory for Information Science and Technology, Department of Automation, Tsinghua University, Beijing, China
{jin-c12,h-yu14,w-l14}@mails.tsinghua.edu.cn,
{jfeng,lujiwen,jzhou}@tsinghua.edu.cn

Abstract. The detection of substances in the left atrial appendage (LAA) is essential in evaluating disease development and treatment planning in patients with atrial fibrillation. The advent of 4D-CT bringing high spatiotemporal resolution, we present a new approach for the detection of substances in the LAA by spatiotemporal motion analysis and make a detailed judgment and analysis of spatial distribution and classification of most objects in the LAA. The noise interference is also eliminated properly. This approach requires the extraction of the optical flow field for all adjacent phases in a cardiac cycle of 20 phases. According to the optical flow information of 19 optical flow fields, we adopt the nearest neighbor interpolation method to establish the motion trajectory of the key voxels in a whole cardiac cycle. Then we create a hierarchical clustering tree by calculating the similarity between the tracks based on hierarchical clustering and find the corresponding classification for every track. Different classifications of tracks represent the divisions of substances in the LAA. Finally, we perform the stress and strain detection of the critical lump using time-frequency analysis of their trajectories. Tests and validations of our approach were performed on 32 data sets (artificial diagnosis of echocardiography and 4-D CT). The frequency responded range to stress and strain of different substances was obtained, which included normal circulation blood, mild, moderate and severe SEC blood, initial jelling thrombi, old or calcified thrombi, organic thrombi and pectinate muscles. Our results are consistent with the two artificial diagnoses. Furthermore, they can refine the identification of substances such as their texture and tiny sizes.

Keywords: Left atrial appendage (LAA) · Thrombus · 4D-CT
Optical flow · Hierarchical clustering · Time-frequency analysis

1 Introduction

Many atrial fibrillation (AF) events are associated with thrombi in the left atrial appendage (LAA). The LAA is a long and curved chamber and there are many

M. Pop et al. (Eds.): STACOM 2017, LNCS 10663, pp. 42–50, 2018.
https://doi.org/10.1007/978-3-319-75541-0_5

pectinate muscles in the intima, which make it an anatomical foundation for the formation of thrombi [1]. In the treatment of AF, clinicians need to evaluate the risk of cardiogenic thrombosis and its complications, which is important before the cardiac cardioversion and LAA occlusion [2]. At present, the diagnostic modalities using imaging mainly include ultrasound and spectral Doppler. These modalities determine hematic features by observing with the naked eyes or calculating blood flow rate that cannot be quantitatively evaluated.

We present a new approach that detects substances by spatiotemporal motion analysis based on 4D-CT and make a detailed analysis of spatial distribution and classification of most objects in the LAA. The noise interference is also eliminated properly (Paragraph 1 Sect. 2.4, Paragraph 3 Sect. 3.2). The whole 20 phase 4D-CT data is collected from Philips 256-iCT and the interval between adjacent phases is around 0.04 s in a cardiac cycle. The voxels' movement of position is little and their gray level is almost unchanged between adjacent phases in the smooth "3D + t" heart movie, which meets the requirements of our approach. We detect the motion trajectories of most voxels and they are automatically clustered by their motion law and corresponding morphological features. Finally, the composition, spatial distribution and texture of the objects are obtained, whose physical basis is the biological tissues with different hardness can produce different responses under the external or internal force.

The estimation of deformation body motion from dynamic imaging is a difficult task. Optical flow (OF) has been used to produce elastograms, which is a noninvasive quantitative method of vascular elasticity in atherosclerotic images. i.e., a local phase-based optical flow (LPBOF) was used to estimate the longitudinal and radial motion amplitudes of the artery, which improved the tracking performance and beam forming strategies recommended in vivo [3]. Meanwhile, Goncalves et al. [4] used Short-time Fourier transforms and continuous wavelet transform in the evaluation of Doppler ultrasound embolic signals, which indicates that the time-frequency analysis can detect emboli more accurately.

We present a four-step approach (Fig. 1). Tests and validations were performed on 32 data sets (artificial diagnosis of echocardiography and 4-D CT). The frequency response to stress and strain of different substances were obtained, which included normal blood, mild, moderate and severe SEC blood, initial jelling thrombi, calcified thrombi, organic thrombi and pectinate muscles.

The 4-D CT data of 20 phases				
STEP	1. Extraction of optical flow fields of adjacent phase	2. The tracking of key voxels in whole cardiac cycle	3. Clustering of all trajectory	4. Time-frequency analysis of the tracks curve of critical lump
METHOD	Optical flow	Nearest Neighbour interpolation	Hierarchical clustering	Discrete-Time FourierTransform (DTFT)

Fig. 1. Framework of the auto-spatiotemporal motion analysis.

2 Method

2.1 Extraction of Optical Flow Fields of Adjacent Phase

Our group has designed a number of LAA segmentation methods [5], which were used to extract regions of interest (ROI), the LAA volume of 5%, 10%, ..., 90%, 95% phases in the whole cardiac cycle. Then, we use optical flow to track all voxels in the ROI of the adjacent phase in turn. Its constraint equation is

$$I(x, y, z, t) = I(x + \delta_x, y + \delta_y, z + \delta z, t + \delta_t). \tag{1}$$

$I(x, y, z, t)$ are the voxels in positions (x, y, z), we assume that the movement is sufficiently small, then the constraint equations can be obtained using the Taylor formula:

$$I(x+\delta_x, y+\delta_y, z+\delta z, t+\delta_t) = I(x, y, z, t)+\frac{\partial I}{\partial x}\delta_x+\frac{\partial I}{\partial y}\delta_y+\frac{\partial I}{\partial z}\delta_z+\frac{\partial I}{\partial t}\delta_t+H.O.T. \tag{2}$$

H.O.T. stands for higher order terms, which can be neglected when the movement is small. From this equation we can obtain:

$$\frac{\partial I}{\partial x}\frac{\delta_x}{\delta_t}+\frac{\partial I}{\partial y}\frac{\delta_y}{\delta_t}+\frac{\partial I}{\partial z}\frac{\delta_z}{\delta_t}+\frac{\partial I}{\partial t}\frac{\delta_t}{\delta_t}=\frac{\partial I}{\partial x}V_x+\frac{\partial I}{\partial y}V_y+\frac{\partial I}{\partial z}V_z+\frac{\partial I}{\partial t}=0, \tag{3}$$

V_x, V_y, V_z are optical flow vectors in the x, y, z composition of $I(x, y, z, t)$, then

$$I_x * V_x + I_y * V_y + I_z * V_z = -I_t. \tag{4}$$

Finally,

$$\boldsymbol{V} = \begin{bmatrix} V_x \\ V_y \\ V_z \end{bmatrix} = \begin{bmatrix} \sum I_{x_i}^2 & \sum I_{x_i} I_{y_i} & \sum I_{x_i} I_{z_i} \\ \sum I_{x_i} I_{y_i} & \sum I_{y_i}^2 & \sum I_{y_i} I_{z_i} \\ \sum I_{x_i} I_{z_i} & \sum I_{y_i} I_{z_i} & \sum I_{z_i}^2 \end{bmatrix}^{-1} \begin{bmatrix} -\sum I_{x_i} I_{t_i} \\ -\sum I_{y_i} I_{t_i} \\ -\sum I_{z_i} I_{t_i} \end{bmatrix}, \tag{5}$$

where the sum is from 1 to n [6].

Therefore, from 5% to 95%, 19 OF fields are displayed through the Munsell color system. The arrow's length represents the moving velocity of the voxel, the arrow's color represents the gray of the voxel, and the arrow's direction represents the direction of moving voxel (Fig. 3).

2.2 The Tracking of Key Voxels in Whole Cardiac Cycle

We obtain the OF elements of 19 optical flow fields, and connect them in terms of their positions, colors, and directions and every complete trajectory curve of the voxel is generated. The end coordinates of each vector $(P_{(V_{x_0}, V_{y_0}, V_{z_0})})$ in the previous of yields are decimal values while the coordinates of the starting voxels of the next optical flow to be connected are integers. To address thus we use the Nearest Neighbour interpolation [7]. First, the gray similarity levels are compared

between $P_{(V_{x_0},V_{y_0},V_{z_0})}$ and its adjacent eight voxels, then the $P_{(V_{x_0},V_{y_0},V_{z_0})}$ is connected with the most similar voxel. But, if there are two or more voxels similar to its gray, we choose the nearest voxel as the connecting point. The formula can be written as:

$$P'(x', y', z') = P(\text{int}(x + 0.5), \text{int}(y + 0.5), \text{int}(z + 0.5)). \tag{6}$$

$P'(x', y', z')$ is the adjacent voxel's coordinates, and $\text{int}(x)$ is the integer arithmetic.

2.3 Hierarchical Clustering of All Trajectory Curves

We create a hierarchical nested clustering tree by calculating the similarity between the tracks based on hierarchical clustering [8], and find the corresponding classification for every track. Different classifications of tracks represent the divisions of the substances in the LAA.

19 OF fields are generated between 20 phases in a cardiac cycle, and there are 20 stationary points (single one at each phase) in each track of a voxel and their positions are known. We calculated the metric distance (similarity) of the hierarchical clustering as follow:

$$D = \frac{1}{20}\sum_{n=1}^{20}\sqrt{((x_{n,i} - x_{n,j})^2 + (y_{n,i} - y_{n,j})^2 + (z_{n,i} - z_{n,j})^2)} + (|G_i - G_j|). \tag{7}$$

N is the number of stagnation points. (x, y, z) are coordinates of the stationary point, $i, j \in \Phi$, Φ are the set of the tracks of all voxels, and G_i is the grayscale of the track i. We combine the two tracks with the highest similarity and iterate the process to generate the clustering tree. The reasons we chose hierarchical clustering are: first, it can deal with outliers and noise in CT data well; second, a termination condition can be artificially set, thus we can flexibly set the desired number of clusters or define a threshold of the distance between two nearest clusters depending on the patient's CT development (with or without "Filling Defect", etc.).

2.4 Time-Frequency Analysis of the Track Curve of Critical Lumps — to Realize the Stress and Strain Detection of Lumps

In our approach, the voxel can generate a track only when it is present in all 20 phases, otherwise it will be discarded, such as noises and isolated points that accidentally appear in the CT volume. All the tracks are obvious in jitter, which meet the requirements of the time-frequency analysis. In the tracks of each cluster, the features of normal circulation blood are: the total number is large, the journey is long, the tortuous degree is not high. Yet, there are different laws of motion for some coagulated lumps of platelets such as the thrombi attached to the endocardium, whose tracks are short, generally in situ vibratile and are different from each other in the way of vibration. Initial jelling thrombosis is very

elastic and its track is anisotropic rapid vibration on X, Y, Z. The movement of calcific thrombi is close to rigid vibration, and the vibration direction is relatively simple. During a cardiac cycle, the LAA experiences a regular contraction movement of filling, emptying, refilling, and emptying. The organic thrombosis and pectinate muscles regularly move along with the whole LAA. Therefore, we can make a time-frequency analysis of the trajectories of 3-D voxels and try to distinguish their subordinate categories. Also, we can calculate the total number of tracks of the lump to evaluate its size.

We use Discrete-Time Fourier Transform (DTFT) to do the time-frequency analysis and add 81 zero-value points to overcome the "Fence Effect" and fully observe the details on all the frequency points [9]. So the 3-D coordinate matrix of 20 stationary points of each track is:

$$T = \begin{bmatrix} x_1 \; x_2 \; \ldots \; x_{19} \; 0 \ldots 0 \\ y_1 \; y_2 \; \ldots \; y_{19} \; 0 \ldots 0 \\ z_1 \; x_2 \; \ldots \; z_{19} \; 0 \ldots 0 \end{bmatrix}^T_{3\times 100}, \tag{8}$$

which can be simplified as:

$$T = \left[\boldsymbol{x} \; \boldsymbol{y} \; \boldsymbol{z} \right]. \tag{9}$$

By N-point Discrete-Time Fourier Transform (DTFT), we obtain:

$$\tilde{T} = \left[\tilde{x} \; \tilde{y} \; \tilde{z} \right]. \tag{10}$$

The spectrums of three directions "$\tilde{x} - k$", "$\tilde{y} - k$", "$\tilde{z} - k$" are obtained. The lump is the initial jelling thrombi if the frequency in most of the three directions is much higher, and it is calcific thrombi if the frequency is high in one direction present but low in the other two directions. It may be organic thrombi or pectinate muscles if the distribution of spectrum on three directions is unified and simple.

3 Experiment and Discussion

3.1 Dataset

Our database consists of 32 patients who have combined examinations of 4-D CT volumes (0.35 mm × 0.35 mm × 0.44 mm, 512 × 512, 20 phases, Philips 256-iCT) and transthoracic echocardiography (200 frames TDI images per second, 30 volumes per second, Philips iE33) from February 2017 to May 2017. 11 of them are spontaneous echocardiographic contrast (SEC) and 6 of them have thrombi in the LAA which are mainly attached to the necks of the LAA.

Ground Truth Annotation. The images of CT and echocardiography were analyzed by 2 experienced physicians respectively. In echocardiography, the "cloudy" echoes of the whirlpool motion are defined as SEC. According to the SEC echo density, they are divided into mild, moderate and severe. In fact, the velocity of blood flow in LAA is the most significant parameter related to SEC.

Thrombosis is defined as a lump of parenchymal echo in the LAA. Slightly lower echoes are initial jelling thrombi, and slightly higher and stronger echoes are old or calcific thrombi. Small or new thrombi are easily missed into the pectinate muscles, which is the limitation of echocardiography. In CT, the thrombosis is defined as a multi-phase "Filling Defect" in the LAA after the injection of radiopaque contrast.

3.2 Evaluation and Results

In AF, the fluid characteristics of the blood in the LAA will produce a series of changes of different degrees, which has a corresponding presentation in 4D-CT volumes. We tested our approach on 32 CT data sets and compared the previous results of diagnosis by two doctors. The results are shown in Table 1.

Table 1. Comparison of the proposed auto-approach with the diagnostic results.

Approach		Diagnosis						
		Normal	Abnormal					
			SEC			Thrombi		
Artifical diagnosis	Echo-cardiography	16	Mild	Moderate, severe		Initial jelling	Old, calcific	
			6	5		3	2	
	4D-CT	16	9			7		
Auto-approach	4D-CT	16	Mild	Moderate	Severe	Initial jelling	Old, calcific	Organic, muscles
			4	4	2	3	2	1

In Echocardiography, an initial thrombus was wrongly diagnosed as pectinate muscles, which tallied with the examination of Intracardiac Echocardiography (ICE) in this patient's radiofrequency ablation operation. In 4D-CT, the lump of SEC was misjudged as thrombosis due to the slowness of contrast agent absorbing. We analyzed the motion characteristics of the substances and obtained a more detailed classification and maturity qualitation. The size of thrombi can also be evaluated through the number of trajectories of voxels.

Figures 2 and 3 shows the whole test process of the subjects. We generated the whole tracks of all voxels originated from the 5% phase in the ROI (Fig. 3(a)). The voxels which entered or outflowed in the LAA in other subsequent phases in the whole cardiac cycle were abandoned. The reasons are that the mural thrombi in LAA exist at any phase and the fluid characteristics of voxels in the 5% phase can represent the condition of intracardiac blood for the isolated noises occurring randomly in certain phases are filtered.

In 3-D space, the tracks corresponding to the voxels of normal circulating blood (Fig. 3(b)): i. In the time domain, the curve in "$x-t$", "$y-t$", "$z-t$" changes smoothly; ii. In the frequency domain, the responses in "$\tilde{x}-k$", "$\tilde{y}-k$", "$\tilde{z}-k$" are distributed in the low frequency region. In the echocardiography, the track of SEC blood is long, the transverse ("$x-t$", "$y-t$") tortuosity is high, and

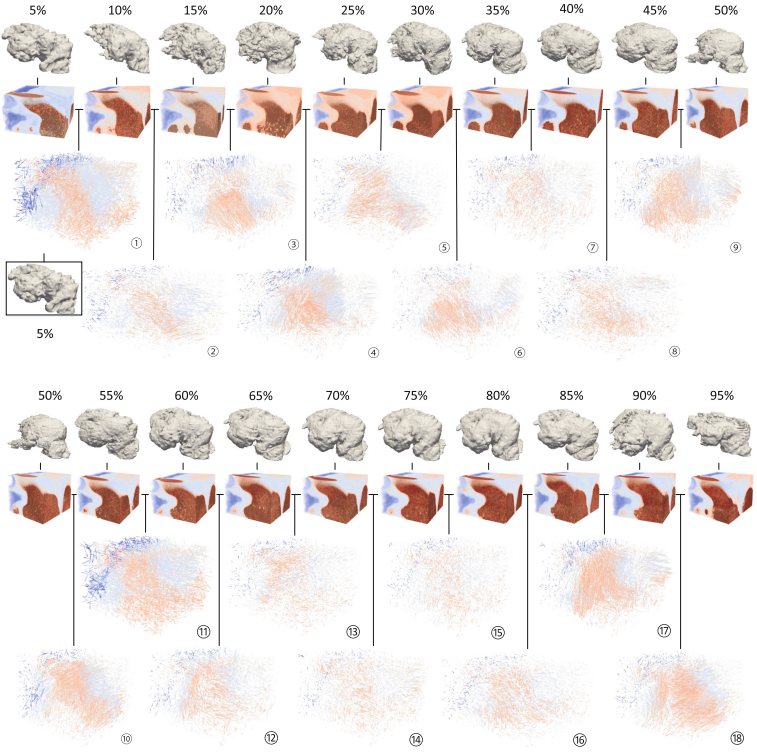


Fig. 2. Extraction of optical flow fields of adjacent phase.

the longitudinal curve ("$z-t$") changes more gently (Fig. 3(c)): i. In the time domain, the "$x-t$", "$y-t$" curve oscillates severely, the "$z-t$" oscillates more smoothly; ii. In the frequency domain, the responses are in the high frequency region in "$\tilde{x}-k$", "$\tilde{y}-k$" and in the lower frequency region in "$\tilde{z}-k$". In the moderate and severe SEC, the response increase in time and frequency domain. The diagnostic parameters of this dataset are in Table 2. Thrombosis diagnosis is consistent with the description in the last paragraph of Sect. 2.4. According to the standard diagnosis (Fig. 3(d), (e), (f)), the diagnostic parameters of this dataset in the preliminary validation are in Table 2.

Although the test is performed on small data sets from 32 people, the corresponding laws are worth exploring. Our approach took about 90 min for the tracking of pivotal voxels, about 4 min for Hierarchical Clustering, about 2 min for DTFT on 4 Intel Core i7 processors at 4.0 GHz with 16 GB of RAM.

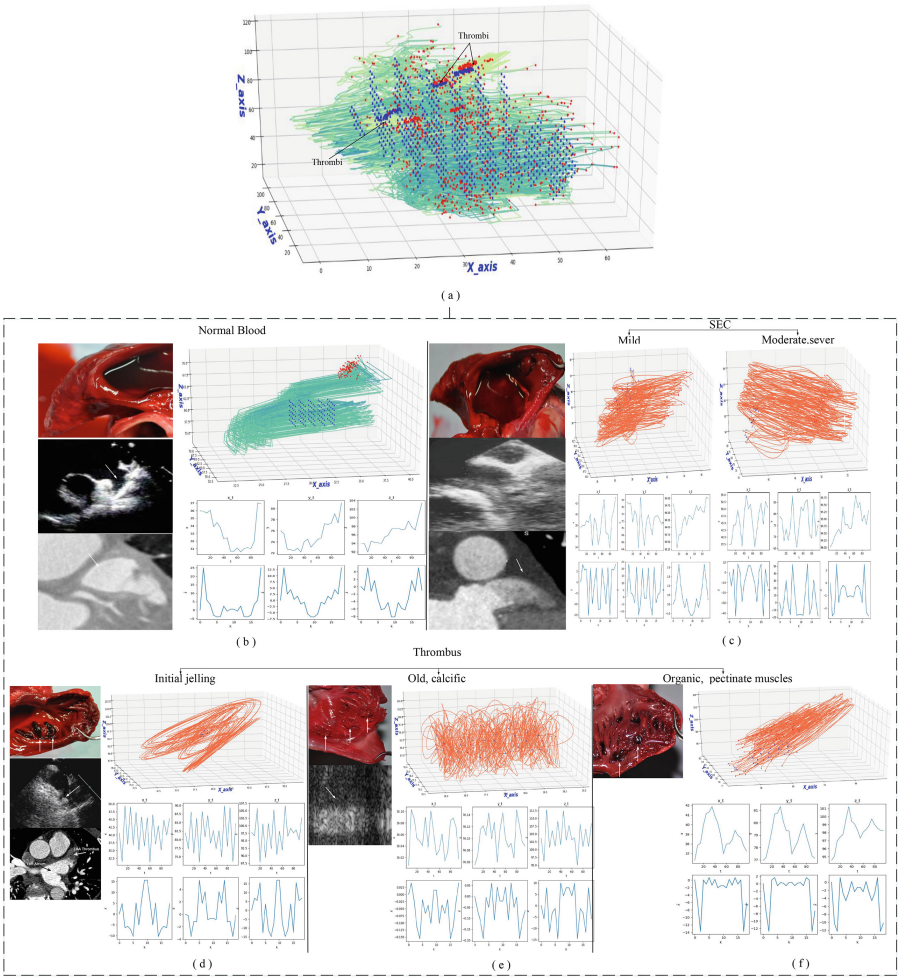


Fig. 3. The whole test of time-frequency analysis of the substances. In the dashed box, from top to bottom, from left to right, the real map, ultrasound, CT, trajectory, time-domain response, frequency-domain response. (a) The whole tracks of all voxels originated from the 5% phase. (b) The tracks corresponding to the voxels of normal circulation blood. (c) The track of SEC blood. (d), (e), (f) The track of thrombi (Initial jelling, old or calcific, organic or pectinate muscles).

Table 2. The corresponding diagnostic parameters for this dataset.

	Normal blood	Mild SEC	Moderate, severe SEC
Frequency response region	$k_{x,y,z} < 4 \pm 1$	$k_{x,y} < 8 \pm 2, k_z < 4 \pm 1$	$k_{x,y} < 9 \pm 1, k_z < 3 \pm 2$
Actual blood flow velocity	v > 24 cm/s	v < 24 cm/s	v < 12 cm/s
	Initial jelling	Old, calcific	Organic, pectinate muscles
Frequency response region	$k_{x,y,z} < 9 \pm 1$	$k_{x,y} \approx 0, k_z < 9 \pm 1$	$k_{x,y,z} \approx 5$

4 Conclusion

We established an automatic discrimination system for the LAA, which judged the characteristics of blood flow, identified the presence of thrombi, and further determined the texture and size of the thrombi. The improvement of the OF method may generate the trajectory of voxels with less error. DTFT transform may not be the most appropriate for analysing non-stationary and mutated thrombi signals, so we will do further research to achieve better results.

Acknowledgements. This work is supported by the National Natural Science Foundation of China under Grants 61622207, 61373074, 61225008, and 61572271.

References

1. Holmes, D.R., Reddy, V.Y., Turi, Z.G., Doshi, S.K., Sievert, H., Buchbinder, M., Mullin, C.M., Sick, P.: Percutaneous closure of the left atrial appendage versus warfarin therapy for prevention of stroke in patients with atrial fibrillation: a randomised non-inferiority trial. Lancet **374**(9689), 534–542 (2009)
2. Calvert, P.A., Rana, B.S., Begley, D.A., Shapiro, L.M.: Occlusion of left atrial appendage to treat atrial fibrillation. Lancet **374**(9703), 1742–1743 (2009)
3. Zahnd, G., Salles, S., Liebgott, H., Vray, D., Serusclat, A., Moulin, P.: Real-time ultrasound-tagging to track the 2D motion of the common carotid artery wall in vivo. Med. Phys. **42**(2), 820–830 (2015)
4. Goncalves, I.B., Leiria, A., Moura, M.M.M.: STFT or CWT for the detection of Doppler ultrasound embolic signals. Int. J. Numer. Methods Biomed. Eng. **29**(9), 964–976 (2013)
5. Wang, L., Feng, J., Jin, C., Lu, J., Zhou, J.: Left atrial appendage segmentation based on ranking 2-D segmentation proposals. In: Mansi, T., McLeod, K., Pop, M., Rhode, K., Sermesant, M., Young, A. (eds.) STACOM 2016. LNCS, vol. 10124, pp. 21–29. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-52718-5_3
6. Baker, S., Scharstein, D., Lewis, J.P., Roth, S., Black, M.J., Szeliski, R.: A database and evaluation methodology for optical flow. Int. J. Comput. Vis. **92**(1), 1–31 (2011)
7. Lehmann, T.M., Gonner, C., Spitzer, K.: Survey: interpolation methods in medical image processing. IEEE Trans. Med. Imaging **18**(11), 1049–1075 (1999)
8. Liu, A.-A., Yu-Ting, S., Nie, W.-Z., Kankanhalli, M.: Hierarchical clustering multi-task learning for joint human action grouping and recognition. IEEE Trans. Patt. Anal. Mach. Intell. **39**(1), 102–114 (2017)
9. Gröchenig, K.: Foundations of Time-Frequency Analysis. Springer (2013)

Estimation of Healthy and Fibrotic Tissue Distributions in DE-CMR Incorporating CINE-CMR in an EM Algorithm

Susana Merino-Caviedes[1(✉)], Lucilio Cordero-Grande[2], M. Teresa Sevilla-Ruiz[3], Ana Revilla-Orodea[3], M. Teresa Pérez Rodríguez[4], César Palencia de Lara[4], Marcos Martín-Fernández[1], and Carlos Alberola-López[1]

[1] Laboratorio de Procesado de Imagen, Universidad de Valladolid, Valladolid, Spain
smercav@lpi.tel.uva.es
[2] Department of Biomedical Engineering, King's College London, London, UK
[3] Unidad de Imagen Cardiaca, Hospital Clínico Universitario de Valladolid, CIBER de Enfermedades Cardiovasculares (CIBERCV), Valladolid, Spain
[4] Dpto. de Matemática Aplicada, Universidad de Valladolid, Valladolid, Spain

Abstract. Delayed Enhancement (DE) Cardiac Magnetic Resonance (CMR) allows practitioners to identify fibrosis in the myocardium. It is of importance in the differential diagnosis and therapy selection in Hypertrophic Cardiomyopathy (HCM). However, most clinical semiautomatic scar quantification methods present high intra- and interobserver variability in the case of HCM. Automatic methods relying on mixture model estimation of the myocardial intensity distribution are also subject to variability due to inaccuracies of the myocardial mask. In this paper, the CINE-CMR image information is incorporated to the estimation of the DE-CMR tissue distributions, without assuming perfect alignment between the two modalities nor the same label partitions in them. For this purpose, we propose an expectation maximization algorithm that estimates the DE-CMR distribution parameters, as well as the conditional probabilities of the DE-CMR labels with respect to the labels of CINE-CMR, with the latter being an input of the algorithm. Our results show that, compared to applying the EM using only the DE-CMR data, the proposed algorithm is more accurate in estimating the myocardial tissue parameters and obtains higher likelihood of the fibrosis voxels, as well as a higher Dice coefficient of the subsequent segmentations.

Keywords: Scar segmentation · EM algorithm
Hypertrophic cardiomyopathy

This work was partially supported by the Spanish Ministerio de Ciencia e Innovación and the European Regional Development Fund (ERDF-FEDER) under Research Grant TEC2014-57428-R and the Spanish Junta de Castilla y León under Grant VA069U16.

M. Pop et al. (Eds.): STACOM 2017, LNCS 10663, pp. 51–59, 2018.
https://doi.org/10.1007/978-3-319-75541-0_6

1 Introduction

Hypertrophic Cardiomyopathy (HCM) is the most prevalent cardiomyopathy of non-ischemic origin, with mortality rates between 1% and 5% [1]. One of the clinical indicators in HCM is the volume of fibrosis in the myocardium. It may be measured using Delayed Enhancement (DE) Cardiac Magnetic Resonance (CMR), where fibrosis appears hyperenhanced with respect to myocardial healthy tissue. Current clinical segmentation methods for DE-CMR employ a prior mask of the myocardium, either obtained from CINE-CMR and aligned to DE-CMR if needed, or manually delineated on the image. A small remote myocardium region of interest (ROI) is also chosen to compute a threshold above which a voxel is considered fibrosis. However, the presence of diffuse fibrosis complicates selecting a threshold in HCM. In [2], it was observed that both inter- and intraobserver variability of manual and semiautomatic clinical methods were significantly higher in HCM than in either acute or chronic infarction.

Several proposed automatic scar segmentation methods are based on estimating the probability distribution of the healthy and scarred tissue [3,4] with the expectation maximization (EM) algorithm [5]. This approach is better suited to the use of more complex segmentation methods for identifying the scar, but misalignments in the myocardial prior mask may alter the estimated distributions. Moreover, the fibrosis distribution may be similar to the blood distribution. In [6], a multivariate mixture model to segment the myocardial contours in DE-CMR including CINE-CMR as well as T2-weighted images was proposed, where complementary modalities were used to solve ambiguities in borders. However, this method uses the same label partition for all the modalities and relies on an atlas for its initialization. Therefore, its adaptation to fibrosis segmentation in HCM is not straightforward.

Here, the variability of the tissue distribution estimations with respect to displacements in the prior myocardial mask is explored. We propose an EM algorithm for the estimation of all tissue distributions in DE-CMR using CINE-CMR and a prior myocardial mask. In the proposed algorithm, the label partitions of DE-CMR and CINE-CMR do not need to be the same. Additionally, small misalignments between both modalities are considered. The experimental results show that the proposed EM parameter estimations have lower error with respect to the classic EM algorithm on the DE-CMR myocardial mask. The scar segmentations obtained by applying the maximum likelihood (ML) criterion with the parameters estimated by the proposed method achieved a higher Dice coefficient than the classic EM method as well.

The rest of the document is structured as follows: in Sect. 2, the proposed EM method is described, followed by Sect. 3, where the experimental setup and results are shown. Finally, some conclusions are drawn in Sect. 4.

2 Methods

Let $I_C(\mathbf{x}, t) : \Omega \times [0, T] \rightarrow \mathbb{R}$ be a spatiotemporal CINE-CMR image, where T is the cardiac cycle period and $\Omega \subset \mathbb{R}^D$ is the image spatial domain. Let $I_R(\mathbf{x})$

be a DE-CMR image acquired at an instant $t = t_R \in [0, T]$. We define the label sets $\mathcal{L} = \{L_i\}_{i=1}^{N_L} = \{\mathrm{C}, \mathrm{H}, \mathrm{S}, \mathrm{B}\}$ for DE-CMR and $\mathcal{A} = \{A_i\}_{i=1}^{N_A} = \{\mathrm{C}, \mathrm{M}, \mathrm{B}\}$ for CINE-CMR, where $N_L = 4$, $N_A = 3$, and the labels C, M, H, S and B stand respectively for the blood cavity enclosed by the endocardium, the myocardium, the healthy tissue, the scar and the background. Finally, let $\widehat{A}(\mathbf{x}) : \Omega \to \mathcal{A}$ be an anatomical segmentation that assigns to $\mathbf{x}$ its estimated CINE-CMR label at time t_R.

The Rice distribution is employed to model a single tissue intensity distribution in magnetic resonance imaging. The Rayleigh and the Gaussian distributions, appropriate particular cases of the general Rice distribution, have been used to model the healthy tissue and the blood intensity distributions, respectively [3,4]. The intensity distribution of each particular tissue in DE-CMR is assumed to have invariant parameters with respect to the CINE-CMR intensity $I_C(\mathbf{x}, t_R)$ (from now on, $I_C(\mathbf{x}, t_R) = I_C(\mathbf{x})$ for clarity) and anatomical segmentation $\widehat{A}(\mathbf{x})$; that is, if $L_i(\mathbf{x})$ is known for a given pixel $\mathbf{x}$, then $P(I_R(\mathbf{x})|I_C(\mathbf{x}), L_i(\mathbf{x}), A_k(\mathbf{x})) = P(I_R(\mathbf{x})|L_i(\mathbf{x}))$. However, both the labels L_i and the distribution parameters for each L_i are unknown. For this reason, it seems reasonable to use the Expectation Maximization (EM) algorithm [5] to estimate these parameters. Let $\bar{\theta}$ be the parameters required in order to characterize all DE-CMR tissue distributions. The image loglikelihood is:

$$\begin{aligned} \log L(\bar{\theta}) &= \sum_{\mathbf{x}_n \in \Omega} \log \sum_{j=1}^{N_L} P(I_R(\mathbf{x}_n), I_C(\mathbf{x}_n), L_j; \bar{\theta}) \\ &= \sum_{\mathbf{x}_n \in \Omega} \log \sum_{j=1}^{N_L} \sum_{k=1}^{N_A} P(I_R(\mathbf{x}_n), I_C(\mathbf{x}_n), L_j, A_k; \bar{\theta}) \end{aligned} \tag{1}$$

where the second identity comes from applying the law of total probability using the CINE-CMR labels. In order to estimate the $\bar{\theta}$ that maximizes the log likelihood, we modify the EM method so that it takes into account the CINE-CMR label probabilities. Therefore, let $Q_{jk}(\mathbf{x})$ be such that $Q_{jk}(\mathbf{x}) \geq 0$ and $\sum_{j=1}^{N_L} \sum_{k=1}^{N_A} Q_{jk}(\mathbf{x}) = 1$. Then:

$$\begin{aligned} \log L(\bar{\theta}) &= \sum_{\mathbf{x}_n \in \Omega} \log \sum_{j=1}^{N_L} \sum_{k=1}^{N_A} P(I_R(\mathbf{x}_n), I_C(\mathbf{x}_n), L_j, A_k; \bar{\theta}) \\ &= \sum_{\mathbf{x}_n \in \Omega} \log \sum_{j=1}^{N_L} \sum_{k=1}^{N_A} Q_{jk}(\mathbf{x}_n) \frac{P(I_R(\mathbf{x}_n), I_C(\mathbf{x}_n), L_j, A_k; \bar{\theta})}{Q_{jk}(\mathbf{x}_n)} \\ &\geq \sum_{\mathbf{x}_n \in \Omega} \sum_{j=1}^{N_L} \sum_{k=1}^{N_A} Q_{jk}(\mathbf{x}_n) \log \frac{P(I_R(\mathbf{x}_n), I_C(\mathbf{x}_n), L_j, A_k; \bar{\theta})}{Q_{jk}(\mathbf{x}_n)} = J(\bar{\theta}) \end{aligned} \tag{2}$$

by the application of Jensen's inequality[1]. The joint probabilities may be expressed as (dropping the $\mathbf{x}_n$ dependence for clarity):

$$P(I_R, I_C, L_j, A_k; \bar{\theta}) = P(I_R|I_C, L_j, A_k; \bar{\theta})P(L_j|I_C, A_k)P(I_C|A_k)P(A_k) \quad (3)$$

The $P(I_C(\mathbf{x}_n)|A_k), k = 1, 2, 3$ are estimated from $\widehat{A}(\mathbf{x})$ and the CINE-CMR intensity distributions as smoothed normalized histograms. Regarding $P(A_k)$, it is modeled so that the probability decays when the distance to the considered CINE-CMR ROI increases. Its computation uses a Gaussian filter $g(\mathbf{x}, \sigma)$ with $\sigma = d/3$, where d is a parameter for the maximum distance the myocardial contours are expected to be misaligned. Then, $P(A_k) = (\chi(\widehat{A}(\mathbf{x}), A_k) * g(x, d/3))/(\sum_{l=1}^{N_A} \chi(\widehat{A}(\mathbf{x}), A_l) * g(x, d/3))$, where $\chi(z, A_l) = 1$ if $z = A_l$ and 0 otherwise. Since $P(L_j|A_k)$ are also unknown, the method should also provide their estimates, which will be referred to as π_{jk}. Since they are conditional probabilities, they obey $\pi_{jk} \geq 0$ and $\sum_{j=1}^{N_L} \pi_{jk} = 1$. $Q_{jk}(\mathbf{x})$ is chosen so that $J(\bar{\theta})$ is as close as possible to $\log L(\bar{\theta})$, and it takes the expression:

$$Q_{jk}(\mathbf{x}_n) = \frac{P(I_R(\mathbf{x}_n), L_j, A_k; \bar{\theta})}{\sum_{\hat{j}=1}^{N_L} \sum_{\hat{k}=1}^{N_A} P(I_R(\mathbf{x}_n), L_{\hat{j}}, A_{\hat{k}}; \bar{\theta})} = P(L_j, A_k|I_R(\mathbf{x}_n); \bar{\theta}) \quad (4)$$

Regarding the maximization step, a new value for $\bar{\theta}$ is chosen as the argument that maximizes $J(\bar{\theta})$, considering $Q_{ij}(\mathbf{x}_n)$ as fixed.

$$\begin{aligned} \widehat{\bar{\theta}}, \widehat{\bar{\pi}} &= \arg\max_{\bar{\theta}, \bar{\pi}} J(\bar{\theta}) \\ &= \arg\max_{\bar{\theta}, \bar{\pi}} \sum_{\mathbf{x}_n \in \Omega} \sum_{j=1}^{N_L} \sum_{k=1}^{N_A} Q_{jk}(\mathbf{x}_n) \big(\log P(I_R(\mathbf{x}_n)|L_j, A_k; \bar{\theta}) + \log \pi_{jk} \\ &\quad + \log P(A_k) - \log Q_{jk}(\mathbf{x}_n) \big) \end{aligned} \quad (5)$$

In order to estimate $\widehat{\bar{\theta}}$, the first derivatives are set to zero. Given the assumption $P(I_R(\mathbf{x}_n), L_j, A_k; \bar{\theta}) = P(I_R(\mathbf{x}_n), L_j; \bar{\theta})$, problem (5) is equivalent to:

$$\widehat{\bar{\theta}} = \arg\max_{\bar{\theta}} \sum_{\mathbf{x}_n \in \Omega} \sum_{j=1}^{N_L} \log P(I_R(\mathbf{x}_n)|L_j; \bar{\theta}) \sum_{k=1}^{N_A} Q_{jk}(\mathbf{x}_n) \quad (6)$$

For the computation of $\widehat{\bar{\pi}}$, the method of the Lagrange multipliers is employed, so that the augmented problem is:

$$\widehat{\pi_{jk}} = \arg\max_{\pi_{ml}} \sum_{\mathbf{x}_n \in \Omega} \sum_{m=1}^{N_L} \sum_{l=1}^{N_A} Q_{ml}(\mathbf{x}_n) \log \pi_{ml} - \lambda \Big(\sum_{m=1}^{N_L} \pi_{mk} - 1 \Big) \quad (7)$$

[1] Jensen's inequality states that if f is a concave function and X is a random variable, then $E[f(X)] \leq f(E[X])$.

The solution of (7) has the following closed form:

$$\widehat{\pi_{jk}} = \frac{\sum_{\mathbf{x}_n \in \Omega} Q_{jk}(\mathbf{x}_n)}{\sum_{\mathbf{x}_n \in \Omega} \sum_{m=1}^{N_L} Q_{mk}(\mathbf{x}_n)} \tag{8}$$

Since the EM algorithm finds local maxima, the choice of initial values for the parameters greatly influences the output estimates. In our method, these initial parameters are computed using the CINE-CMR labels, sometimes combined with heuristics based on clinical criteria.

- The blood tissue is modeled by a Gaussian distribution whose parameters are estimated by the maximum likelihood (ML) criterion on the voxels labeled as cavity in CINE-CMR.
- The healthy myocardial tissue is modeled by a Rayleigh distribution. Its parameter is estimated from the mode of the histogram computed from the voxels labeled as myocardium in CINE-CMR [7].
- The scar intensity distribution is assumed to be close enough to a Gaussian. Its mean is initialized at 5 standard deviations over the mean of the estimated healthy myocardium, and its standard deviation is initialized to the same initial standard deviation of the healthy tissue distribution. This makes use of the findings in [8].
- The background is composed of a number of different tissues. For this reason, this distribution is not estimated as a parametric model, but as a normalized histogram, smoothed by a Gaussian kernel with a standard deviation of 0.8, of the voxels labeled as background.
- The π_{ij} are initialized as the (i, j)-th element of the matrix $\begin{pmatrix} 0.7 & 0.1 & 0.1 \\ 0.1 & 0.4 & 0.1 \\ 0.1 & 0.4 & 0.1 \\ 0.1 & 0.1 & 0.7 \end{pmatrix}$.

Table 1. Imaging parameters of the acquired CMR sequences.

Settings	SAx-C	SAx-LE	2C-C	4C-C
Acquisition sequence	sBTFE BH	PSIR_TFE BH	sBTFE BH	sBTFE BH
View	SAx	SAx	2C LAx	4C LAx
Temporal phases	30	1	30	30
FOV/Frequency encoding steps	1.98–2.01 mm	1.98–2.01 mm	1.98–2.00 mm	1.98–2.00 mm
In-plane pixel spacing	0.93–1 mm	0.55–0.62 mm	1.18–1.25 mm	0.81–1 mm
Slice thickness (gap)	10 (0) mm	10 (0) mm	8 (0) mm	8 (0) mm
Number of slices	9–13	9–12	1	3
Echo time	1.60–1.79	2.99	1.59–1.83	1.71–1.89
Repetition time	3.18–3.57	6.09–6.14	3.18–3.66	3.41–3.79
Flip Angle	45	25	45	45

BH: Breath Hold. FOV: Field of View.

3 Experimental Results

For this work, 21 patients with hypertrophic cardiomyopathy (HCM) underwent a cardiac MRI study, each of which contained CINE sequences in short axis (SAx), two chamber (2C) and four chamber (4C) long axis (LAx), and a SAx DE-CMR sequence. All sequences were acquired with a 3T Philips Achieva MR scanner, and their main acquisition parameters are summarized in Table 1.

The epicardial and endocardial SAx-C contours at end-diastole were delineated by cardiologists. From them and SAx-C, the $P(I_C|A_k)$ distributions were estimated. In DE-CMR, the fibrosis ROI (ROI_s) was defined by a manually selected threshold, applied on a myocardial ROI (ROI_m) conservatively drawn to avoid including false positives after thresholding and scars of other pathologies. Additionally, a small ROI (ROI_r) of remote myocardium (healthy tissue far away from the septum) was drawn, and a voxel whose brightness was considered by the cardiologists as the maximum fibrosis brightness $I_R^{S,\max}$.

In order to simulate the variability in the myocardial contours of registration methods and to study the effect of false positives due to contour misalignments in the distributions estimation, a second set of myocardial masks were drawn on the SAx-LE images for each patient and 30 realizations of in-plane translations with random orientation uniformly distributed in the range $[0, 2\pi)$ and norm of 3 mm were applied to the myocardial contours of each slice. The resulting masks were our test set of myocardial masks. The SAx-C and SAx-LE volumes were spatially aligned using the framework described in [9], which also corrects breath hold misalignments. The SAx-C volume at t_R was transformed to the SAx-LE space and resolution.

For all patients and the test set of myocardial masks, a Rayleigh-Gaussian mixture (RGM) was estimated on the myocardium, and the proposed EM algorithm (PROP) was run with an expected maximum distance to misalignments $d = 1, 2, 3, 4$. Additionally, the intensity distribution parameters of ROI_s and ROI_r were estimated on each patient by the ML criterion. The absolute deviation of the EM estimated parameters with respect to the parameters estimated on ROI_s and ROI_r, normalized by the latter, are visualized as boxplots in Fig. 1. It may be observed that the proposed method provides parameters more similar to the ones computed on ROI_s and ROI_r. In addition, the estimations present

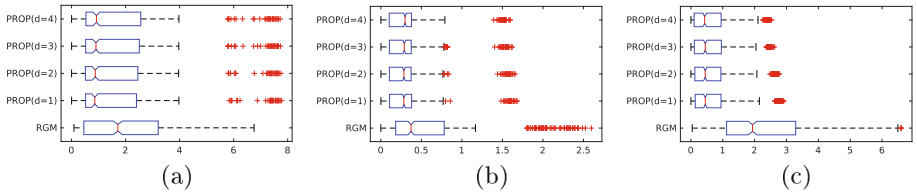


Fig. 1. Boxplots of the normalized absolute deviations of (a) the healthy tissue intensity mode, (b) the fibrosis mean and (c) the fibrosis standard deviation computed by the EM methods with respect to the respective ML estimates on ROI_r and ROI_s.

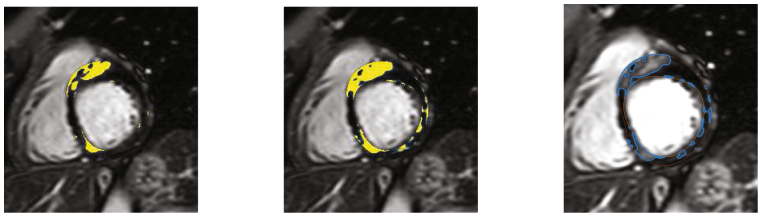
less error variance with respect to the RGM method, and similar results for every d employed.

The mean loglikelihood (mLL) yielded by the estimated parameters was computed on ROI_s and ROI_r. Their mean and standard deviation are given in Table 2. For each experiment, a fibrosis mask was generated as the voxels contained in ROI_m where $P(L_S|I_R(\mathbf{x}), \bar{\theta}) > P(L_H|I_R(\mathbf{x}), \bar{\theta})$. This mask was compared with the ROI_s by means of the Dice coefficient. Their means and standard deviations are also included in Table 2. Figure 2 shows an example of the averaged segmentations with the test myocardial set with RGM and PROP($d = 3$).

It may be observed that the proposed method achieved higher mean mLL in the fibrosis ROI, but lower mean mLL in the remote myocardium ROI. With respect to the Dice coefficient, the proposed method achieved higher Dice mean values, as well as lower standard deviations than the RGM method. Kruskal-Wallis tests at a 1% significance level on the mLL of each ROI and the Dice coefficient rejected the null hypothesis of having the same median for the mLL of fibrosis and the Dice coefficient with $p < 10^{-12}$ and $p < 10^{-5}$ respectively, and accepted it for the mLL of healthy tissue with $p = 0.81$. Conducting paired Mann-Whitney U tests at 5% significance level between the RGM and the PROP method with all d values indicated that the median Dice coefficient with the

Table 2. Mean loglikelihood (mLL) on ROI_s and ROI_r, and Dice coefficient of the ensuing segmentations. Measures are given as mean ± standard deviation.

Method	Extra inputs	mLL in ROI_r	mLL in ROI_s	Dice coefficient
RGM	—	-5.551 ± 0.851	-4.287 ± 0.671	0.473 ± 0.304
PROP(d = 1)	—	-5.603 ± 1.270	-4.254 ± 1.472	0.541 ± 0.254
PROP(d = 2)	—	-5.597 ± 1.244	-4.220 ± 1.402	0.545 ± 0.257
PROP(d = 3)	—	-5.602 ± 1.229	-4.198 ± 1.328	0.543 ± 0.258
PROP(d = 4)	—	-5.615 ± 1.235	-4.218 ± 1.302	0.537 ± 0.261
5SD	ROI_r	—	—	0.482 ± 0.254
FWHM	$ROI_r, I_R^{S,\max}$	—	—	0.589 ± 0.287



(a) (b) (c)

Fig. 2. SAx-LE slice with overlapped mean outcome of the ML segmentation using (a) the RGM method and (b) the PROP method. (c) Manual delineations.

PROP method was higher than the median Dice coefficient with the RGM, with $p < 10^{-9}$ in all tests. From these values it may be inferred that the proposed EM method is better at identifying fibrosis than the classic RGM EM method in the presence of false positives introduced by myocardial contour delineation errors, even if the mLL of healthy tissue is slightly decreased. In Table 2, the Dice coefficient results of two methods used in clinical practice, which require additional user interaction, are also given. The 5 standard deviations over the remote myocardium mean method (5SD) behaves slightly better than the RGM, but yields lower Dice coefficient values than the PROP method. The Full Width at Half Maximum (FWHM) method achieves the best results, at the expense of requiring the cardiologist to delineate ROI_r and selecting a fibrosis voxel with maximum DE-CMR intensity $I_R^{S,\max}$, which the proposed method does not need.

4 Conclusions

In this work, an EM algorithm that takes into account the information of two images from different cardiac CMR modalities (DE-CMR and CINE-CMR) has been proposed. This algorithm is designed so that the number of labels on each modality do not need to be the same, and without assuming perfect alignment between modalities. Compared to the Rayleigh-Gaussian mixture model often used in the literature, the proposed method achieved lower normalized absolute deviations in the parameter estimates, as well as improved Dice coefficient of the segmentations performed by the ML criterion. Our future work includes studying the partial volume effect influence on the algorithm.

References

1. Bruder, O., Wagner, A., Jensen, C.J., Schneider, S., Ong, P., Kispert, E.M., Nassenstein, K., Schlosser, T., Sabin, G.V., Sechtem, U., Mahrholdt, H.: Myocardial scar visualized by cardiovascular magnetic resonance imaging predicts major adverse events in patients with hypertrophic cardiomyopathy. J. Am. Coll. Cardiol. **56**(11), 875–887 (2010)
2. Flett, A.S., Hasleton, J., Cook, C., Hausenloy, D., Quarta, G., Ariti, C., Muthurangu, V., Moon, J.C.: Evaluation of techniques for the quantification of myocardial scar of differing etiology using cardiac magnetic resonance. JACC Cardiovas. Imaging **4**(2), 150–156 (2011)
3. Hennemuth, A., Seeger, A., Friman, O., Miller, S., Klumpp, B., Oeltze, S., Peitgen, H.O.: A comprehensive approach to the analysis of contrast enhanced cardiac MR images. IEEE Trans. Med. Imaging **27**(11), 1592–1610 (2008)
4. Elagouni, K., Ciofolo-Veit, C., Mory, B.: Automatic segmentation of pathological tissues in cardiac MRI. In: IEEE International Symposium on Biomedical Imaging, Rotterdam, Netherlands, pp. 472–475, April 2010
5. Dempster, A.P., Laird, N.M., Rubin, D.B.: Maximum likelihood from incomplete data via the EM algorithm. J. Roy. Stat. Soc. Ser. B (Methodol.) **39**(1), 1–38 (1977)
6. Zhuang, X., Shen, J.: Multi-scale patch and multi-modality atlases for whole heart segmentation of MRI. Med. Image Anal. **31**, 77–87 (2016)

7. Aja-Fernández, S., Tristán-Vega, A., Alberola-López, C.: Noise estimation in single- and multiple-coil magnetic resonance data based on statistical models. Magn. Reson. Imaging **27**(10), 1397–1409 (2009)
8. Aquaro, G., Positano, V., Pingitore, A., Strata, E., Di Bella, G., Formisano, F., Spirito, P., Lombardi, M.: Quantitative analysis of late gadolinium enhancement in hypertrophic cardiomyopathy. J. Cardiovasc. Magn. Reson. **12**(1), 21 (2010)
9. Cordero-Grande, L., Merino-Caviedes, S., Alba, X., Figueras i Ventura, R.M., Frangi, A.F., Alberola-Lopez, C.: 3D fusion of cine and late-enhanced cardiac magnetic resonance images. In: 9th IEEE International Symposium on Biomedical Imaging, Barcelona, Spain, pp. 286–289 (2012)

Multilevel Non-parametric Groupwise Registration in Cardiac MRI: Application to Explanted Porcine Hearts

Mia Mojica[1], Mihaela Pop[2], Maxime Sermesant[3], and Mehran Ebrahimi[1(✉)]

[1] Faculty of Science, University of Ontario Institute of Technology, Oshawa, ON, Canada
{mia.mojica,mehran.ebrahimi}@uoit.ca
[2] Department of Medical Biophysics, Sunnybrook Research Institute, University of Toronto, Toronto, ON, Canada
mihaela.pop@utoronto.ca
[3] Asclepios Team, INRIA, Sophia Antipolis, France
maxime.sermesant@inria.fr

Abstract. Statistical atlases of myocardial fiber directions have great utility in modelling applications. The first step in building atlases requires a registration of the hearts to a template. In this paper, we performed groupwise registration on a small database of explanted pig hearts ($N = 4$) and coupled it with a multilevel pairwise registration framework in order to generate an average cardiac geometry. The scheme implemented in our experiments effectively registers and normalizes the hearts despite a high variability in cardiac measurements. In addition, we adopted an intuitive averaging technique on the transformed versions of each heart to obtain a new reference geometry at every iteration. This reduces biases that may be introduced by the selection of an initial reference geometry in the construction of an average cardiac geometry. The next step will focus on improving current results by using a larger database of heart samples.

Keywords: Image registration · Inverse problems · Cardiac MRI
Multilevel registration · Non-parametric registration
Groupwise registration · Cardiac atlas

1 Introduction

Cardiovascular disease continues to be the leading cause of death, accounting for 30% of mortality worldwide [1]. There has been an increasing demand to understand the mechanical and electrical activities of the heart through the construction of atlases that model healthy hearts, against which pathological hearts can be compared. However, availability of explanted human hearts is scarce. Thus, studying large hearts (e.g. canine and pig hearts) could provide a good alternative as the cardiac anatomies and functions of the three species

M. Pop et al. (Eds.): STACOM 2017, LNCS 10663, pp. 60–69, 2018.
https://doi.org/10.1007/978-3-319-75541-0_7

are very similar. A statistical comparison between canine and human hearts in terms of their fiber orientations and their variability is provided in [2].

In this paper, we aim to lay the foundation of a framework for constructing a statistical atlas of a healthy porcine heart. To do this, we first need a registration framework that allows us to calculate transformations that would normalize their geometries. Specifically, we focus on the construction of an average cardiac geometry through groupwise registration of anatomical magnetic resonance (MR) images obtained using a diffusion-weighted method applied to a small database of porcine hearts (Fig. 1).

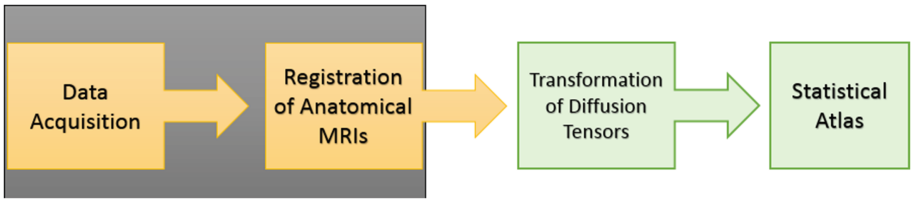


Fig. 1. Workflow diagram for building a statistical porcine cardiac atlas. Diffusion-weighted MR images of porcine hearts were acquired as discussed in Sect. 2.1, and then an average cardiac geometry was constructed through groupwise registration framework initialized by a multilevel affine and non-parametric pairwise registration scheme.

2 Methods

In this section, we discuss both the data acquisition methods and the registration framework used for our experiments.

2.1 Data Acquisition

All diffusion-weighted (DW) MRI studies were performed on a dedicated 1.5T GE Signa Excite scanner using freshly explanted healthy pig hearts [5]. In the current study, we used the following MR parameters: TE = 35 ms, TR = 700 ms, echo train length = 2, b-value = 0 for the unweighted MR images and $b = 500$ s/mm^2 when the seven diffusion gradients were applied, respectively. We used the same field of view (FOV) and a 256×256 k-space. The total scan time for DW imaging was approximately 8 h.

2.2 Pairwise Registration of the Anatomical MR Images

Constructing an average geometry for cardiac data is not trivial due to its variability in terms of scaling and spatial structures. Thus, we need to use an approach that caters to large deformations and assures the correct alignment of corresponding cardiac structures. In [2], a combination of constrained affine registration and hybrid intensity-based and feature-based non-rigid registration algorithm was used for the pairwise registration step. The constrained affine registration step

served to normalize the heights and radii of the hearts, while non-rigid registration ensured precise matching of cardiac structures. It is important to note, however, that the approach in [2] implied the need for selection of landmarks.

We followed the image registration outline proposed in [2] to align the hearts in the data set and come up with an average geometry. However, instead of using a non-rigid registration algorithm for pairwise registration, we used a combination of affine and non-parametric registration to align each subject to the current reference geometry. At every iteration, a reference geometry is obtained by computing the average of the transformations that register each of the hearts to the current reference geometry.

2.2.1 Mathematical Model

Given a template image $\mathcal{T} : \Omega \subset \mathbb{R}^3 \to \mathbb{R}$ and a reference image $\mathcal{R} : \Omega \subset \mathbb{R}^3 \to \mathbb{R}$, our goal is to find a reasonable transformation such that a transformed version of the template image $\mathcal{T}$ is similar to the reference $\mathcal{R}$ [10]. Mathematically, this can be modelled by solving the optimization problem

$$\min_{y} \mathcal{J}[y] = \mathcal{D}[\mathcal{T}[y], \mathcal{R}] + \mathcal{S}[y], \tag{1}$$

where $y : \Omega \to \mathbb{R}^3$ is the transformation that aligns $\mathcal{T}$ to $\mathcal{R}$, and $\mathcal{T}[y]$ is a transformed version of the template image $\mathcal{T}$.

The first term ($\mathcal{D}$) in the joint functional $\mathcal{J}$ measures the similarity between the two images and thus helps determine if there is a reasonable match between the image features. The second term ($\mathcal{S}$) is the regularization term, which makes the registration problem well-posed.

In our implementation, we used the sum of squared differences (SSD) for similarity measure, i.e.,

$$\mathcal{D}[\mathcal{T}[y], \mathcal{R}] = \frac{1}{2} \int_{\Omega} (\mathcal{T}[y](x) - \mathcal{R}(x))^2 \, dx. \tag{2}$$

The above integral is approximated using a midpoint quadrature rule on a cell-centered grid with mesh spacing h_i in each dimension $i \in \{1, 2, 3\}$. Its discretized form is given by

$$D^{\mathrm{SSD},h}(T^h, R^h) = \frac{1}{2} hd \|T^h - R^h\|^2,$$

with $hd = h_1 \cdot h_2 \cdot h_3$.

The regularization term $\mathcal{S}[y]$ enforces the functional to lead to a unique minimizer. In our experiments, we used the elastic potential of the transformation y for our regularization term [10]. It is given by

$$\mathcal{S}[y] = \text{Elastic Potential}[y - y^{\mathrm{ref}}].$$

2.2.2 Multilevel Representation of Data

In this paper, we used a multilevel registration scheme to initialize each iteration in the groupwise registration. With a multilevel approach, we start by solving

the minimization problem on a coarser level and then progress onto finer levels. The solutions on the coarser levels serve as starting guesses for the next (finer) levels. This is an efficient method for aligning two 3D images since computations on coarser levels are cheaper relative to those on finer levels. This approach also helps avoid running into local minimizers.

To obtain a smoothed measurement of an image in different levels, we get the average of the intensity values of adjacent cells. A detailed discussion on the computation of a multilevel representation of a 3D MR image can be found in [10].

2.2.3 Affine and Non-parametric Registration

From the coarsest to the finest level, we solve the discretized form of the optimization problem in (1), which is given by

$$\min_{y^h} \mathcal{J}^h \left[y^h\right] = \mathcal{D}\left(T^h, R^h; y^h\right) + \mathcal{S}\left(y^h - y^{\mathrm{ref},h}\right). \tag{3}$$

At the coarsest level, an affine parametric registration is performed. An affine transformation is one that preserves points, lines and planes. It allows for translation, rotation, scaling, and shearing.

An affine transformation $y = [y^1, y^2, y^3]^T$ of a point $[x^1, x^2, x^3]^T \in \mathbb{R}^3$ may be parametrized as

$$\begin{aligned} y^1 &= w_1 x^1 + w_2 x^2 \ + w_3 x^3 \ + w_4 \\ y^2 &= w_5 x^1 + w_6 x^2 \ + w_7 x^3 \ + w_8 \\ y^3 &= w_9 x^1 + w_{10} x^2 + w_{11} x^3 + w_{12}. \end{aligned}$$

The solution $y(w, x)$ to the affine registration problem [10] at the coarsest level will serve as the initial guess for the reference transformation y^{ref} for the elastic regularizer $\mathcal{S}[y]$ in the non-parametric registration step. That is, $y^{\mathrm{ref}} = y(w, x)$.

At every level, the minimization problem (3) is solved using a Gauss-Newton approach with an Armijo line search. The initial guess at every level is given by the prolongated version of the solution y^h from the preceding coarser level.

3 Groupwise Registration

Various methods have already been used to normalize cardiac geometries. In [4], a new reference geometry is computed each time an image in the data set is registered. Here, we adopted the method used in [8] and [2], where biases introduced by using one of the anatomical MR images as the first reference image were eliminated by registering the images to the same current reference geometry.

At every iteration, the reference geometry is updated using an averaging technique that takes into account all the transformations that align each subject to the current reference geometry. The update to the current reference geometry is given by

$$I^{n+1}_{\text{mean}}(\mathrm{x}^h) = \frac{1}{N}\sum_{i=1}^{N} I_i\left(T^n_i \circ [T^n_{\text{mean}}]^{-1}(\mathrm{x}^h)\right), \tag{4}$$

where N is number of images in the data set, x^h is the original grid, I_i are the anatomical MR images, $i = 1, 2, \ldots, N$, T^n_i is the mapping that registers the ith subject to the nth reference geometry, and T^n_{mean} is the average of the transformations T^n_i at the nth iteration defined as $T^n_{\text{mean}} = \frac{1}{N}\sum_{i=1}^{N} T^n_i$. The term $[T^n_{\text{mean}}]^{-1}$ is the inverse of the average of the transformations T^n_{mean}, and $\circ$ denotes the composition of transformations.

Repeating the update process in (4) leads to an average geometry I_{mean} and a collection of transformations aligning the anatomical MR images to I_{mean}. We can then use these transformations to transform the diffusion tensors of all the diffusion-weighted MR images in the data set.

In our implementations, we assumed that the transformation T and displacement d obtained when aligning a template image to a reference image are related by the equation $T(\mathrm{x}^h) = \mathrm{x}^h + d(\mathrm{x}^h)$. An inverse for the transformation T may be approximated by

$$\begin{aligned}\left[T(\mathrm{x}^h)\right]^{-1} &\approx \mathrm{x}^h - d(\mathrm{x}^h)\\ &= \mathrm{x}^h - \left(T(\mathrm{x}^h) - \mathrm{x}^h\right)\\ &= -T(\mathrm{x}^h) + 2\mathrm{x}^h.\end{aligned}$$

Thus, an approximation of the inverse for the average transformation field T^n_{mean} is

$$\left[T^n_{\text{mean}}(\mathrm{x}^h)\right]^{-1} \approx -T^n_{\text{mean}}(\mathrm{x}^h) + 2\mathrm{x}^h. \tag{5}$$

An outline of the groupwise registration framework is given in Algorithm 1.

Algorithm 1. The Groupwise Registration Framework

1. Initialize $n = 0$.
2. Set an arbitrary image in the data set as the initial reference image I^n_{mean}.
3. Use multilevel non-parametric registration to register each image in the data set to I^n_{mean} and store the resulting transformation field T^n_i that aligns each pair of images.
4. Compute the average transformation field T^n_{mean} at the nth step.
5. Approximate the inverse of T^n_{mean} using the formula in (5).
6. Transform each image in the data set by interpolating the intensity values of each subject I_i over the composition $T^n_i \circ [T^n_{\text{mean}}]^{-1}(\mathrm{x}^h)$.
7. Compute the average of transformed images to obtain the new current reference geometry $I^{n+1}_{\text{mean}}(\mathrm{x}^h)$.
8. Update $n \leftarrow n + 1$.
9. Repeat steps 3 to 8 until the method converges.

4 Results

In this section, we present some of the results obtained after implementing the algorithm discussed in the previous section on a small database of three healthy porcine hearts. We will also comment on the efficiency of the pairwise registration method used to align the full hearts to the reference geometries and on how fast the groupwise registration algorithm converges to a stable average geometry.

The unweighted center axial slices of the 3D MR images of the three healthy pig hearts used in our experiments are displayed in Fig. 2(a)–(c). Figure 2(d) shows the subjects superimposed on each other. Observe how varied the subjects are in terms of their heights and radii. Note that the image relating to Fig. 2(b) was set as the initial reference geometry, i.e., I^0_{mean}.

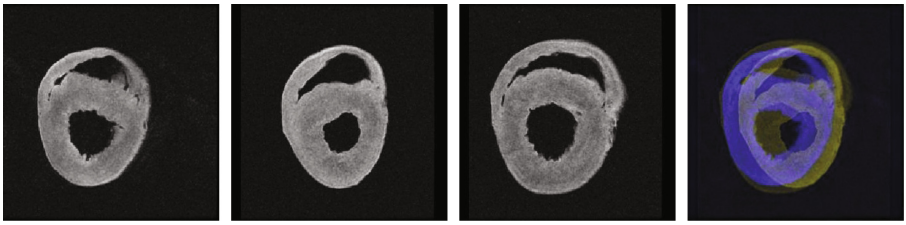


Fig. 2. Variability of heart geometries. (a)–(c) The three anatomical MR images used in our data set and (d) the three hearts overlaid onto each other.

We now observe the efficiency of the pairwise registration method implemented in our experiments. Shown in Fig. 3(a) and (e) are the initial and fourth reference geometries, respectively. The hearts in green masks in Fig. 3(b)–(d) and (f)–(h) are the transformed versions of the three subjects, superimposed against I^0_{mean} and I^4_{mean}, respectively. The pairwise registration algorithm that we used was able to determine reasonable transformations registering the full hearts to the current reference geometries, as demonstrated by the overlap between the template and reference images.

We observed that the method needs only a few iterations until it converges to a reasonable and stable average geometry. In our experiments, a stable average geometry was achieved after 5 to 7 iterations. Presented in Fig. 4 are the absolute changes $|I^n_{\text{mean}} - I^{n-1}_{\text{mean}}|$ in the reference geometries for iterations $n = 1, 2, 3, 4$. Figure 4 (c)–(d) being "almost black" indicates that there is minimal update made to the previous reference geometry. We also calculated the average change in the intensity values of the $256 \times 256 \times 128$ array $|I^n_{\text{mean}} - I^{n-1}_{\text{Imean}}|$ for the first seven iterations. The average change in intensity values dropped from more approximately 0.055 to 0.010, where the intensity change was is in the interval $[0, 1]$. The results are displayed in Fig. 5.

In Fig. 6, we show the first reference geometry and the computed average geometry and compared how the reference geometry changed after 7 iterations.

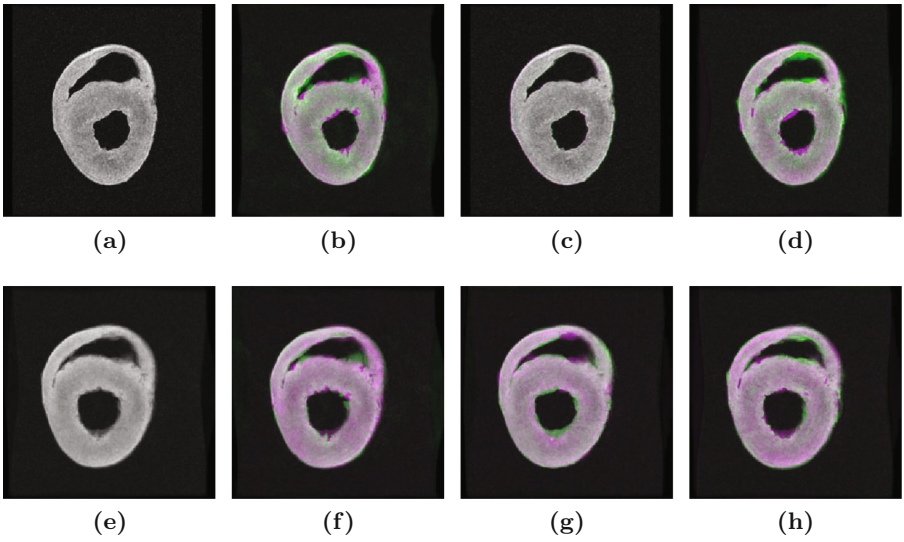
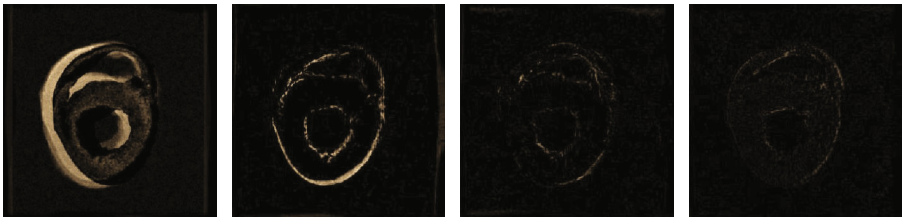


Fig. 3. Efficiency of the Pairwise Registration. (a) Initial reference geometry, (b)–(d) [in green mask] Transformed versions of the hearts in Fig. 2 vs I^0_{mean} [in pink mask], (e) Fourth reference geometry, (f)–(h) [in green mask] Transformed versions of the subjects vs I^4_{mean} [in pink mask] (Color figure online)



(a) $|I^1_{\text{mean}} - I^0_{\text{mean}}|$ **(b)** $|I^2_{\text{mean}} - I^1_{\text{mean}}|$ **(c)** $|I^3_{\text{mean}} - I^2_{\text{mean}}|$ **(d)** $|I^4_{\text{mean}} - I^3_{\text{mean}}|$

Fig. 4. Illustration of the rate of convergence of the groupwise framework to a stable reference geometry. (Color figure online)

Visually, the groupwise registration framework was able to normalize the heights and radii of the three subjects and converged to a reasonable average geometry.

Finally, we added an extra fourth subject to our data set, which significantly increased the variability of the cardiac geometries used in the above experiments. The additional subject is shown in Fig. 7(a), and again in Fig. 7(b) superimposed against the initial reference geometry. In Fig. 7(c), we present the result after aligning the aforementioned subject to the first reference image (Fig. 3(a)). Despite the differences in the cardiac features and geometries between I^0_{mean} and the new heart, the algorithm was still able to find a reasonable transformation that registers the two images. Groupwise registration converged to a stable and

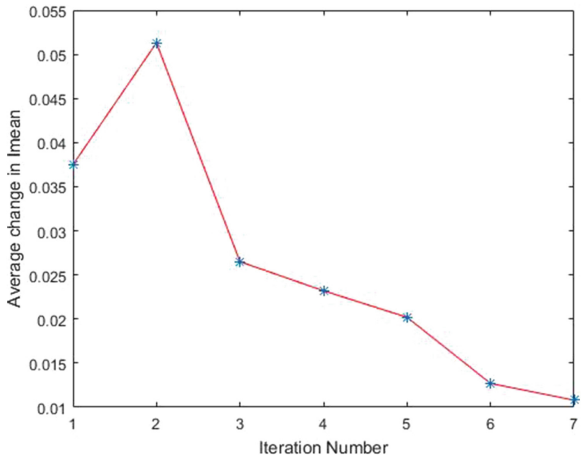
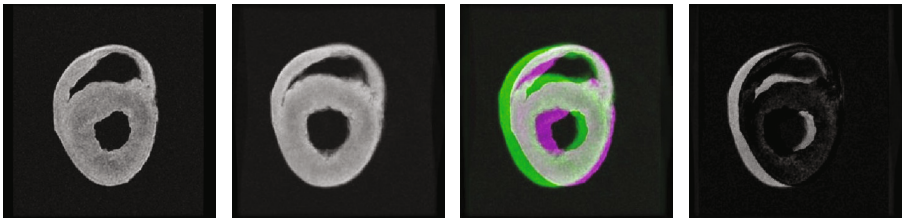
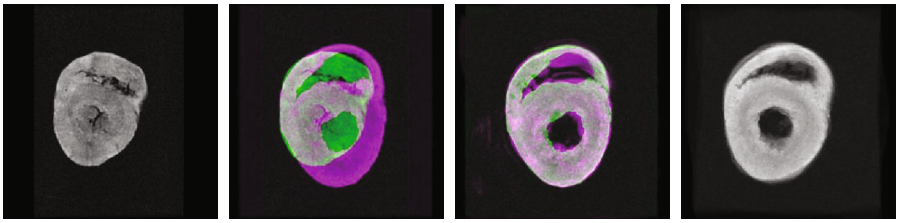


Fig. 5. Average change in I^n_{mean} after each iteration.



(a) I^0_{mean} **(b)** I^7_{mean} **(c)** I^0_{mean} vs I^7_{mean} **(d)** $|I^7_{\text{mean}} - I^0_{\text{mean}}|$

Fig. 6. Evolution of the reference geometries. (a) The initial reference geometry, (b) final average geometry, (c) final reference geometry overlaid onto the initial average geometry, and (d) the absolute change in the reference geometries.



(a) **(b)** **(c)** **(d)**

Fig. 7. Introduction of an Additional Cardiac Data. The same framework was implemented with an outlier cardiac data ($N = 4$). (a) The newly added cardiac MR image, (b) New subject vs I^0_{mean}, (c) Result of the pairwise registration of (a) to I^0_{mean}, and (d) The average geometry computed from the database of the 4 porcine hearts.

reasonable average geometry after 9 iterations. The iterations were stopped when the absolute change in successive reference geometries was less than 0.01, i.e., $|I^n_{\text{mean}} - I^{n-1}_{\text{mean}}| < 0.01$.

5 Future Work and Conclusions

In this paper, we laid the foundation of a framework for building a statistical atlas for healthy porcine hearts by computing an average cardiac geometry from a small database of four freshly explanted healthy porcine hearts. We also demonstrated that the groupwise registration framework that we used converges to a stable average geometry. In addition, the multilevel non-parametric-based registration algorithm was able to successfully normalize the heart geometries and find reasonable transformations registering the subjects to each current reference geometry.

The next step would be to include more hearts in our experiments so that the average geometry would be a more accurate representation of a healthy porcine heart. Along with this, we are planning to compare the efficiency of the Diffeomorphic Log-Demons registration algorithm in [7] with the one we have implemented in this paper.

After building an average geometry, we aim to transform the diffusion tensors of the diffusion-weighted images of the anatomical MRIs to better understand porcine cardiac fiber and laminar sheet orientations needed in building a statistical atlas. The diffusion tensors can be transformed with the same deformations obtained in the pairwise registration step.

References

1. World Health Organization: Cardiovascular Diseases (2017)
2. Peyrat, J.M., Sermesant, M., Pennec, X., et al.: A computational framework for the statistical analysis of cardiac diffusion tensors: application to a small database of canine hearts. IEEE Trans. Med. Imag. **26**, 1500–14 (2007)
3. Lombaert, H., Peyrat, J.M., Croisille, P., et al.: Human atlas of the cardiac fiber architecture: study on a healthy population. IEEE Trans. Med. Imag. **31**(7), 1436–47 (2012)
4. Avants, B., Gee, J.C.: Shape averaging with diffeomorphic flows for atlas creation. In: 2nd IEEE International Symposium on Biomedical Imaging, vol. 1, 595–598 (2004)
5. Pop, M., Ghugre, N.R., Ramanan, V., et al.: Quantification of fibrosis in infarcted swine hearts by ex vivo late gadolinium-enhancement and diffusion-weighted MRI methods. Phys. Med. Biol. **58**(15), 5009–28 (2013)
6. Beg, M.F., Helm, P.A., McVeigh, E., Miller, M.I., Winslow, R.L.: Computational cardiac anatomy Using MRI. Magn. Reson. Med. Off. J. Soc. Magn. Reson. Med./ Soc. Magn. Reson. Med. **52**(5), 1167–1174 (2004)
7. Vercauteren, T., Pennec, X., Perchant, A., Ayache, N.: Symmetric log-domain diffeomorphic registration: a demons-based approach. In: Metaxas, D., Axel, L., Fichtinger, G., Székely, G. (eds.) MICCAI 2008, Part I. LNCS, vol. 5241, pp. 754–761. Springer, Heidelberg (2008). https://doi.org/10.1007/978-3-540-85988-8_90

8. Helm, P.: A Novel Technique for Quantifying Variability of Cardiac Anatomy: Application to the Dyssynchronous Failing Heart. Johns Hopkins University, Baltimore (2005)
9. Modersitzki, J.: Numerical Methods for Image Registration. Oxford University Press, Oxford (2004)
10. Modersitzki, J.: FAIR: Flexible Algorithms for Image Registration. SIAM, Philadelphia (2009)
11. Nocedal, J., Wright, S.J.: Numerical Optimization, 2nd edn. Springer, New York (2006)

ACDC Challenge

GridNet with Automatic Shape Prior Registration for Automatic MRI Cardiac Segmentation

Clément Zotti[1(✉)], Zhiming Luo[1,4], Olivier Humbert[3], Alain Lalande[2], and Pierre-Marc Jodoin[1]

[1] Computer Science Department, Université de Sherbrooke, Sherbrooke, Canada
clement.zotti@usherbrooke.ca
[2] Le2i, Université de Bourgogne Franche-Comté, Dijon, France
[3] Department of Nuclear Medicine, Centre Antoine Lacassagne, Nice, France
[4] Cognitive Science Department, Xiamen University, Xiamen, China

Abstract. In this paper, we propose a fully automatic MRI cardiac segmentation method based on a novel deep convolutional neural network (CNN) designed for the 2017 ACDC MICCAI challenge. The novelty of our network comes with its embedded shape prior and its loss function tailored to the cardiac anatomy. Our model includes a cardiac center-of-mass regression module which allows for an automatic shape prior registration. Also, since our method processes raw MR images without any manual preprocessing and/or image cropping, our CNN learns both high-level features (useful to distinguish the heart from other organs with a similar shape) and low-level features (useful to get accurate segmentation results). Those features are learned with a multi-resolution conv-deconv "grid" architecture which can be seen as an extension of the U-Net.

Experimental results reveal that our method can segment the left and right ventricles as well as the myocardium from a 3D MRI cardiac volume in 0.4 s with an average Dice coefficient of 0.90 and an average Hausdorff distance of 10.4 mm.

Keywords: Convolutional neural networks · MRI · Heart Segmentation

1 Introduction

MRI is the gold standard modality for cardiac assessment [1,2]. In particular, the use of kinetic MR images along the short axis orientation of the heart allows accurate evaluation of the function of the left and right ventricles. For these examinations, one has to delineate the left ventricular endocardium (LV), the left ventricular epicardium (or myocardium - MYO) and the right ventricular endocardium (RV) in order to calculate the volume of the cavities in diastole

M. Pop et al. (Eds.): STACOM 2017, LNCS 10663, pp. 73–81, 2018.
https://doi.org/10.1007/978-3-319-75541-0_8

and systole (and thus the ejection fraction), as well as the myocardial mass [3]. These parameters are mandatory to detect and quantify different pathologies.

As of today, the clinical use of cardiovascular MRI is hampered by the amount of data to be processed (often more than 10 short axis slices and more than 20 phases per slice). Since the manual delineation of all 3D images is clinically impracticable, several semi-automatic methods have been proposed, most of which being based on active contours, dynamic programming, graph cut or some atlas fitting strategies [3–7]. Unfortunately, these methods are far from real time due to the manual interaction which they require. Also, most of them are ill-suited for segmenting simultaneously the LV, the RV, and the MYO.

So far, a limited number of fully-automatic cardiac segmentation methods have been proposed. While some use traditional image analysis techniques like the Hough transform [8] or level sets [9], fully-automatic segmentation methods are usually articulated around a machine learning method [4], and more recently deep learning (DL) and convolutional neural networks (CNN) [10]. Among the best CNN segmentation models are those involving a series of convolutions and pooling layers followed by one [11] or several [12] deconvolution layers. In 2015, Ronneberger et al. [13] proposed the U-Net, a CNN which involves connections between the conv and deconv layers and whose performances on medical images are astonishing. Recently, DL methods have been proposed to segment cardiac images [10,14,15]. While Tran [10] applied the well-known fCNN [11] on MR cardiac images, Tan et al. [14] used CNN to localize (but not segment) the LV, and Ngo et al. [15] use deep belief nets again to localize but not segment the LV. These methods are doing only one or two class segmentation and do not incorporate the shape prior inside the network.

In this paper we propose the first CNN method specifically designed to segment the LV, RV and MYO without a third party segmentation method. Our approach incorporates a shape prior whose registration on the input image is learned by the model.

2 Our Method

The goal of our method is to segment the LV, the RV and the MYO of a 3D $N \times M \times H$ raw MR image X. This is done by predicting a 3D label map T also of size $N \times M \times H$ and whose voxels $\boldsymbol{v} = (i, j, k)$ contain a label $T_{\boldsymbol{v}} \in \{\text{Back, LV, RV, MYO}\}$, where "Back" stands for tissues different than the other three. Following the ACDC structure, X is a series of short axis slices starting from the mitral valve down to the apex [16] (please refer to the ACDC website for more details [17]). In order to enforce a clinically plausible result, a shape prior S is provided which encapsulates the relative position of the LV, RV and MYO. The main challenge when using a shape prior such as this one is to align it correctly onto the input data X [18]. Since the size and the orientation of the heart does not vary much from one patient to another, we register S on X by translating the center of S on the cardiac center of mass (CoM) $\boldsymbol{c}$ of X. The CoM is computed based on the location of the pericardium (obtained from MYO and

RV) in each slice. Since $\boldsymbol{c}$ is not provided with the input image, our method has a regression module designed to predict it. To our knowledge, our approach is the first to incorporate a shape prior as well as its registration within an end-to-end trainable structure.

2.1 Shape Prior

The shape prior S is a 3D volume which encodes the probability of a 3D location $\boldsymbol{v} = (i, j, k)$ of being part of a certain class (Back, LV, RV, or MYO). We estimate this probability by computing the pixel-wise empirical proportion of each class based on the groundtruth label fields T_i of the training dataset:

$$P(\mathcal{C}|\boldsymbol{v}) = \frac{1}{N_t} \sum_{i=1}^{N_t} \mathbb{1}_{\mathcal{C}} \left(T_{i,\boldsymbol{v}}\right)$$

where $\mathbb{1}_{\mathcal{C}} \left(T_{i,\boldsymbol{v}}\right)$ is an indicator function which returns 1 when $T_{i,\boldsymbol{v}} = \mathcal{C}$ and 0 otherwise, and N_t is the total number of training images.

These probabilities are put into a $3 \times 20 \times 100 \times 100$ volume S where 3 stands for the 3 classes (RV, MYO, LV)[1], 20 stands for the number of interpolated slices (from the base to the apex) and 100×100 is the inplane size. Note that prior to compute S, we realign the CoM of all training label fields T_i into a common space and crop a 100×100 region around that center.

2.2 Loss

The goal of our system is to predict a correct label field T given an input image X while automatically aligning the shape prior S on X by aligning their center of masses c. In order to do so, our loss incorporates the following four terms:

$$\mathcal{L} = \sum_i \underbrace{-\gamma_T \sum_{l=1}^{4} \sum_{\boldsymbol{v}} T_{i,l,\boldsymbol{v}} \ln \hat{T}_{l,\boldsymbol{v}}}_{\mathcal{L}_T} \underbrace{-\gamma_C \sum_{l=1}^{4} \sum_{\boldsymbol{v}} C_{i,l,\boldsymbol{v}} \ln \hat{C}_{i,l,\boldsymbol{v}}}_{\mathcal{L}_C} + \underbrace{\gamma_c ||c_{i,\mathbf{w}} - \hat{c}_i||^2}_{\mathcal{L}_c} + \underbrace{\gamma_w ||\boldsymbol{w}||^2}_{\mathcal{L}_w}. \quad (1)$$

Here $\mathcal{L}_T$ and $\mathcal{L}_C$ are the cross-entropies of the predicted labels and the predicted contours. In this equation, l stands for the class index, $\boldsymbol{v}$ is a pixel location, and γ_T and γ_C are constants. $T_{i,l,\boldsymbol{v}}$ is the true probability that pixel $\boldsymbol{v}$ is in class l, and $\hat{T}_{i,l,\boldsymbol{v}}$ is the output of our model for pixel $\boldsymbol{v}$ and class l while C_i and $\hat{C}_i$ are contours extracted from T_i and $\hat{T}_i$. Note that the use of a contour loss has been shown by Luo et al. [19] to enforce a better precision. As for $\mathcal{L}_c$, it is the Euclidean distance between the predicted CoM $c_{i,\mathbf{w}}$ and the true CoM, and $\mathcal{L}_w$ is the prior loss.

[1] No need to store the probability of "Back" since the 4 probabilities sum up to 1.

2.3 Proposed Network

The goal of our CNN is to learn good features for predicting the label field T_i as well as the CoM $\boldsymbol{c}_i$ which is used to align the shape prior S on the input image X_i. In other words, a network that has the ability of minimizing the loss of Eq.(1). With that objective, good features must account for both the global and the local context: the global context to differentiate the heart from the surrounding organs and estimate its CoM, and the local context to ensure accurate segmentation and prediction of contours.

In that perspective, we implemented a grid-like CNN network with 3 columns and 5 rows (c.f. Fig. 1). The input to our model (upper left) is an 256×256 MR image X_i and the shape prior S_i of the corresponding slice, while the output is the CoM $\boldsymbol{c}_i$ (bottom right) and a label field T_i (top right) also of size 256×256. Note that a common issue with MRI cardiac images is the fact that along the 2D short-axis, the location of the heart sometimes get shifted from one slice to another due to different breath-holds during successive acquisitions. As a consequence, instead of processing a 3D volume as a whole, we feed the network with 2D slices as shown in Fig. 1 and reshape the 3D volume T_i by stacking up the resulting 2D label fields.

As we get deeper in the network (from CONV-1 to CONV-5), the extracted features involve a larger context of the input image. Since the CONV-5 layer includes high-level features from the entire image, we use it to predict the cardiac CoM $\boldsymbol{c}_i$ of the input image X_i. The second column contains 4 convolution layers (all without max-pooling) used to compute features at various resolutions. The last column aggregates features from the lowest to the highest resolution. The UNCONV-4 layer contains both global and local features which we use to segment the image. Note that this grid structure is similar to the U-Net except

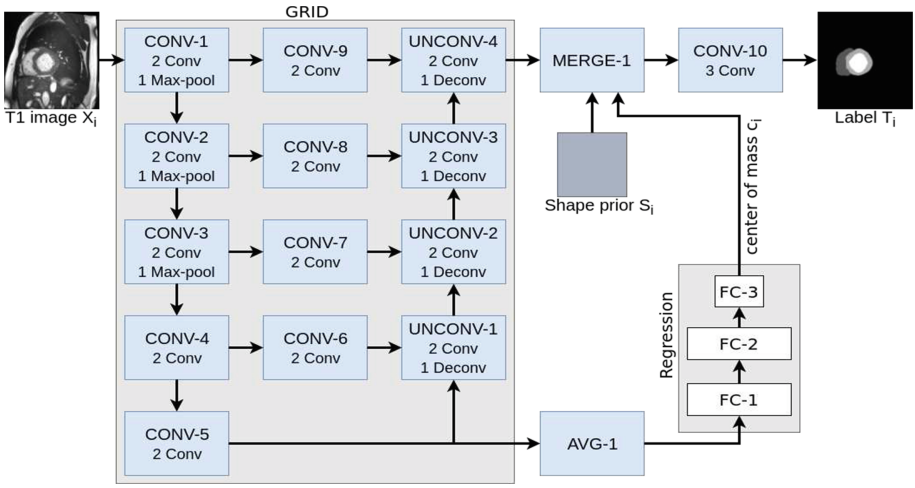


Fig. 1. Our network architecture.

for the middle CONV-6 to 9 layers and the fact that we use the CONV-5 features to estimate a CoM $\boldsymbol{c}_i$. Each conv layer has a 3×3 receptive field and its feature maps have the same size than their input (zero padding). We also batch normalize each feature map, use the ReLU activation function, and dropout [20] to have a better generalization. Please note that so instead of having 32 millions parameters like the U-Net, our gridNet has approximately 8 millions parameters.

Estimating the Center of Mass $\boldsymbol{c}_i$. The CoM $\boldsymbol{c}_i$ is estimated with a regression module located after the CONV-5 layer. An average pooling layer (AVG-1) is used to reduce the number of features fed to the FC-1 layer. The FC-1, FC-2 and FC-3 layers are all fully connected and the output of FC-3 is the (x_i, y_i) prediction of $\boldsymbol{c}_i$.

Estimating the Label Field $\boldsymbol{T}_i$. The output of the UNCONV-4 layer has 4 256×256 feature maps which we append to the shape prior S. The MERGE-1 layer realigns S based on the estimated CoM $\boldsymbol{c}_i$ and use zero padding to make sure S has a 256×256 inplane size. In this way, the output of MERGE-1 has 7 feature maps: 4 from UNCONV-4 and 3 from S. The last CONV-10 layer is used to squash those 7 feature maps down to a 4D output.

Training. The model is trained by minimizing the loss $\mathcal{L}$ of Eq.(1). We use a batch size of 10 2D MR images taken from the ED or ES phase independently and the ADAM optimizer [21]. The model is trained with a learning rate of 10^{-4} for a total of 100 epochs.

Pre and Post Processing. The input 3D images X are pre-processed by clamping the 4% outlying grayscale values, zero-centering the grayscales and normalize it by their standard deviation. Once training is over, we remove outliers by keeping the largest connected component for each class on the overall predicted 3D volume. Note that these pre and post processes are applied to every model tested in Sect. 3.

3 Experimental Setup and Results

3.1 Dataset, Evaluation Criteria, and Other Methods

Our system was trained and tested on the 2017 ACDC dataset. Since the testing dataset was not available as of the paper submission deadline, we trained our system on 75 exams and validated it on the remaining 25 exams. The exams are divided into 5 evenly distributed groups: dilated cardiomyopathy, hypertrophic cardiomyopathy, myocardial infarction with altered left ventricular ejection fraction, abnormal right ventricle and patients without cardiac disease. Cine MR images were acquired in breath hold with a retrospective or prospective gating

and with a SSFP sequence in 2-chambers, 4-chambers and in short axis orientations. A series of short axis slices cover the LV from the base to the apex, with a thickness of 5 to 8 mm and often an interslice gap of 5 mm. The spatial resolution goes from 0.83 to 1.75 mm^2/pixel. For more details on the dataset, please refer to the ACDC website [17].

In order to gauge performances, we report the clinical and geometrical metrics used in the ACDC challenge. The clinical metrics are the correlation coefficients for the cavity volume and the ejection fraction (EF) of the LV and RV, as well as correlation coefficient of the myocardial mass for the End Diastolic (ED) phase. As for the geometrical metrics, we report the Dice coefficient [22] and the Hausdorff distance [23] for all 3 regions and phases.

We compared our method with two recent CNN methods: the conv-deconv CNN by Noh et al. [12] and the U-Net by Ronneberger et al. [13]. We chose those methods based on their excellent segmentation capabilities but also because their architecture can be seen as a particular case of our approach.

3.2 Experimental Results

As can be seen in Table 1, our method outperforms both the conv-deconv and the U-Net. The Dice coefficient is better by an average of 5%, and the Hausdorff distance is lower for our method by an average of 4.4 mm. This is a strong indication that the contour loss and the shape prior help improving results, especially close to the boundaries. Without much surprise, the RV is the most challenging organ, mostly because of its complicated shape, the partial volume effect close to the free wall, and intensity inhomogeneities. The clinical metrics are also in favor of our method. Based on the correlation coefficient, our method is overall better than conv-deconv and UNet, especially for the myocardium (bottom of Table 1).

Table 1. Results for the validation dataset.

	Dice LV		Dice RV		Dice MYO	
	ED	ES	ED	ES	ED	ES
ConvDeconv	0.92	0.87	0.82	0.64	0.76	0.81
UNet	0.96	0.92	0.88	0.79	0.78	0.79
Our method	**0.97**	**0.94**	**0.92**	**0.82**	**0.88**	**0.89**
	HD LV (mm)		HD RV (mm)		HD MYO (mm)	
	ED	ES	ED	ES	ED	ES
ConvDeconv	8.77	10.34	22.59	28.45	13.92	11.64
UNet	6.17	8.29	20.51	21.20	15.25	17.92
Our method	**4.77**	**6.93**	**16.07**	**18.68**	**7.12**	**8.91**

	Corr EF LV	Corr EF RV	Corr MYO ED	Corr LV vol	Corr RV vol
ConvDeconv	0.988	0.764	0.927	0.990	0.957
UNet	0.991	**0.824**	0.921	0.995	0.821
Our method	**0.994**	0.819	**0.963**	**0.998**	**0.967**

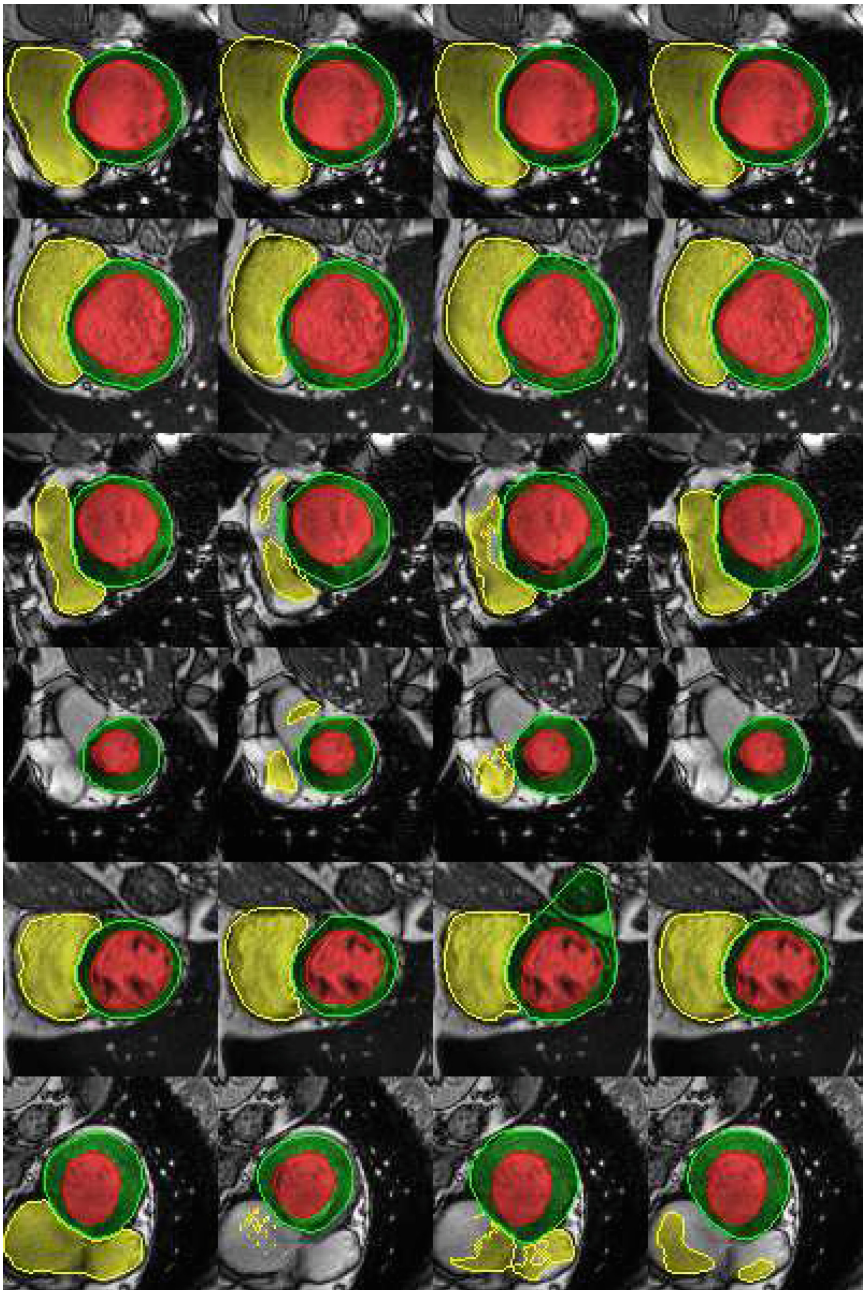


Fig. 2. From left to right: ground truth, conv-deconv, U-Net, and our method.

Careful inspection reveal that errors are not uniformly distributed. Interestingly, conv-deconv and U-Net produce accurate results on most slices of each 3D volume as illustrated in the first two rows of Fig. 2. That said, they often get to generate a distorted result for 1 or 2 slices (out of 7 to 17) which end up decreasing the Dice score and increasing the Hausdorff distance. This situation is shown in rows 3, 4, and 5 of Fig. 2. Overall, the right ventricle is the most challenging region for all three methods. It is especially true at the base of the heart, next to the mitral valve where the RV is connected to the pulmonary artery. This is illustrated in the last row of Fig. 2.

4 Conclusion

We proposed a new CNN method specifically designed for MRI cardiac segmentation. It implements a "grid" architecture which is a generalization of the U-net. It uses a shape prior which is automatically aligned on the input image with a regression method. The shape prior forces the method to produce anatomically plausible results. Experimental results reveal that our method produces an average DICE score of 0.9. This shows that an approach such as our's can be seen as a decisive step towards a fully-automatic clinical tool used to compute functional parameters of the heart. In the future, we plan on generalizing our method to other modalities (such as echocardiography or CT-scan) as well as other organs such as the brain. In that case, we shall propose a more elaborate registration module which shall implement an affine transformation.

References

1. Epstein, F.H.: MRI of left ventricular function. J Nucl. Cardiol **14**(5), 729–744 (2007)
2. Vick, G.W.: The gold standard for noninvasive imaging in coronary heart disease: magnetic resonance imaging. Curr. opin. cardiol. **24**(6), 567–579 (2009)
3. Peng, P., et al.: A review of heart chamber segmentation for structural and functional analysis using cardiac magnetic resonance imaging. MAGMA **29**(2), 155–195 (2016)
4. Petitjean, C., et al.: Right ventricle segmentation from cardiac MRI: a collation study. Med. Image Anal. **19**(1), 187–202 (2015)
5. Auger, D.A., et al.: Semi-automated left ventricular segmentation based on a guide point model approach for 3D cine DENSE cardiovascular magnetic resonance. J. Cardiovasc. Magn. Reson. **16**(1), 8 (2014)
6. Grosgeorge, D., Petitjean, C., Dacher, J.-N., Ruan, S.: Graph cut segmentation with a statistical shape model in cardiac MRI. CVIU **117**(9), 1027–1035 (2013)
7. Petitjean, C., Dacher, J.: A review of segmentation methods in short axis cardiac MR images. Med. Image Anal. **15**(2), 169–184 (2011)
8. Wang, L., Pei, M., Codella, N.C.F., et al.: Left ventricle: fully automated segmentation based on spatiotemporal continuity and myocardium information in cine cardiac magnetic resonance imaging (LV-FAST). BioMed Res. Int. **2015**, 9 (2015). https://doi.org/10.1155/2015/367583. Article ID 367583

9. Liu, Y., Captur, G., et al.: Distance regularized two level sets for segmentation of left and right ventricles from cine-MRI. Magn. Reson. Img. **34**(5), 699–706 (2016)
10. Tran, P.V.: A fully convolutional neural network for cardiac segmentation in short-axis MRI. arXiv preprint arXiv:1604.00494 (2016)
11. Long, J., Shelhamer, E., Darrell, T.: Fully convolutional networks for semantic segmentation. In: Proceedings of CVPR, pp. 3431–3440 (2015)
12. Noh, H., Hong, S., Han, S.: Learning deconvolution network for semantic segmentation. In: Proceedings of ICCV (2015)
13. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: Proceedings of MICCAI, pp. 234–241 (2015)
14. Tan, L.K., et al.: Cardiac left ventricle segmentation using convolutional neural network regression. In: Proceedings of IECBES, pp. 490–493. IEEE (2016)
15. Ngo, T.A., Lu, Z., Carneiro, G.: Combining deep learning and level set for the automated segmentation of the left ventricle of the heart from cardiac cine magnetic resonance. Med. Image Anal. **35**(1), 159–171 (2017)
16. Kastler, B.: Cardiovascular anatomy and atlas of MR normal anatomy. MRI of Cardiovascular Malformations, pp. 17–39. Springer, Heidelberg (2011). https://doi.org/10.1007/978-3-540-30702-0_2
17. ACDC-MICCAI challenge. http://acdc.creatis.insa-lyon.fr/
18. Tavakoli, V., Amini, A.A.: A survey of shaped-based registration and segmentation techniques for cardiac images. CVIU **117**(9), 966–989 (2013)
19. Luo, Z., Mishra, A., Achkar, A., Eichel, J., Li, S.-Z., Jodoin, P.-M.: Non-local deep features for salient object detection. In: proceeding of CVPR (2017)
20. Srivastava, N., Hinton, G., et al.: Dropout: a simple way to prevent neural networks from overfitting. J. of Mach. Learn. Res. **15**, 1929–1958 (2014)
21. Kingma, D., Ba, J.: Adam: a method for stochastic optimization. In: Proceedings of ICLR (2015)
22. Zou, K.H., et al.: Statistical validation of image segmentation quality based on a spatial overlap index 1: scientific reports. Acad. rad. **11**(2), 178–189 (2004)
23. Huttenlocher, D., Klanderman, G., Rucklidge, W.J.: Comparing images using the Hausdorff distance. IEEE Trans PAMI **15**(9), 850–863 (1993)

A Radiomics Approach to Computer-Aided Diagnosis with Cardiac Cine-MRI

Irem Cetin[1(✉)], Gerard Sanroma[1], Steffen E. Petersen[3], Sandy Napel[4], Oscar Camara[1], Miguel-Angel Gonzalez Ballester[1,2], and Karim Lekadir[1]

[1] BCN MedTech, Universitat Pompeu Fabra, Barcelona, Spain
irem.cetin01@estudiant.upf.edu
[2] Catalan Institution for Research and Advanced Studies (ICREA), Barcelona, Spain
[3] William Harvey Research Institute, Queen Mary University of London, London, UK
[4] Department of Radiology, School of Medicine, Stanford University, Stanford, USA

Abstract. Computer-aided diagnosis of cardiovascular diseases (CVDs) with cine-MRI is an important research topic to enable improved stratification of CVD patients. However, current approaches that use expert visualization or conventional clinical indices can lack accuracy for borderline classications. Advanced statistical approaches based on eigendecomposition have been mostly concerned with shape and motion indices. In this paper, we present a new approach to identify CVDs from cine-MRI by estimating large pools of radiomic features (statistical, shape and textural features) encoding relevant changes in anatomical and image characteristics due to CVDs. The calculated cine-MRI radiomic features are assessed using sequential forward feature selection to identify the most relevant ones for given CVD classes (e.g. myocardial infarction, cardiomyopathy, abnormal right ventricle). Finally, advanced machine learning is applied to suitably integrate the selected radiomics for final multi-feature classification based on Support Vector Machines (SVMs). The proposed technique was trained and cross-validated using 100 cine-MRI cases corresponding to five different cardiac classes from the ACDC MICCAI 2017 challenge (https://www.creatis.insa-lyon.fr/Challenge/acdc/index.html). All cases were correctly classified in this preliminary study, indicating potential of using large-scale radiomics for MRI-based diagnosis of CVDs.

Keywords: Cardiac MRI · Machine learning · SVM · Diagnosis Radiomics

1 Introduction

Despite continuous progresses both in clinical research and practice, cardiovascular diseases (CVDs) remain the leading cause of mortality and morbidity in the world [1]. In this context, cardiac imaging such as cine-MRI (Magnetic Resonance Imaging) is expected to play an important role due to its ability

M. Pop et al. (Eds.): STACOM 2017, LNCS 10663, pp. 82–90, 2018.
https://doi.org/10.1007/978-3-319-75541-0_9

to quantify in detail structural and functional properties of the beating heart [2]. However, visual assessment of CVDs using cine-MRI remains challenging and labor-intensive due to the complexity of these diseases, in particular when the structural and functional disorders are subtle [3]. Moreover, quantitative assessment through existing clinical indices such as volumetric measures, ejection fraction, and thickening measures can be suboptimal for borderline cases. Consequently, more advanced automated techniques are needed to exploit the richness of the cardiac data to estimate diagnosis, as well as severity of the phenotype which often is associated with prognosis. Over the years, several methods have been proposed based on eigendecomposition of the moving cardiac shapes [4–10] but they have mostly used geometrical information.

In this paper, we propose instead a radiomics approach to automated image-based diagnosis of complex CVDs. Radiomics is the task of calculating a large number of imaging descriptors from delineated images, which has been widely used in cancer imaging [11]. In the context of cine-MRI, radiomics describing changes in image appearence due to CVDs have not been exploited. The proposed method estimates a large number of radiomic features including statistical, shape, and textural descriptors and assess their ability to discriminate between different CVDs automtically and robustly within a machine learning framework based on SVMs. Training and cross-validation are carried out in this study based on a database of 100 cine-MRI cases corrresponding to five different subclasses, from the ACDC challenge of MICCAI 2017.

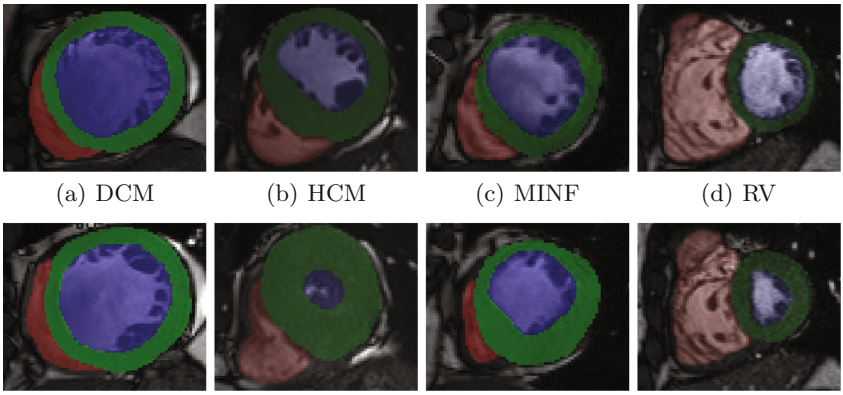
2 Method

2.1 Data Description

This study was conducted in the context of the MICCAI 2017 challenge on Automated Cardiac Diagnosis (the ACDC challenge). The database consists of 100 cases comprising cine-MRI data, height and weight information, as well as the diastolic and systolic phase instants for each subject. Five subclasses were included, namely (see examples in Fig. 1):

(1) Normal subjects (NOR).
(2) Patients with dilated cardiomyopathy (DCM).
(3) Hypertrophic cardiomyopathy (HCM).
(4) Abnormal right ventricle (RV)
(5) Myocardial infarction (MINF).

The 100 images were acquired at the University Hospital of Dijon (France) by using 1.5T or 3T MRI scans (Siemens Medical Solutions, Germany) with the following parameters depending on the examination: image sequence = SSFP cine-MRI, slice thickness = 5 mm or 8 mm, inter-slice gaps = 5 mm or 10 mm, spatial resolution = 1.37 to 1.68 mm2/pixel, number of frames = 28 to 40. This training dataset was then manually segmented for the left ventricle, myocardium and right ventricle by an experienced manual observer at both end-diastole and end-systole time frames. The test data of the challenge will be segmented as explained in the next section.



(a) DCM (b) HCM (c) MINF (d) RV

Fig. 1. Examples of cine-MRI images for the four abnormalities classified in this study (Top: ED, bottom: ES)

2.2 Semi-automatic Segmentation

To segment the test datasets semi-automatically, we propose an atlas-based approach using the publicly available cardiac atlas [12]. To this end, we firstly define manually six anatomical landmarks on each cine-MRI case, more specifically at:

(1) Mid-ventricular slice: RV insertion point next to the liver.
(2) Mid-ventricular slice: A point on the RV free wall.
(3) Mid-ventricular slice: RV insertion point next to the lung.
(4) Mid-ventricular slice: A point on the LV free wall.
(5) Apical slice: Apex.
(6) Basal slice: Center of the base.

We then use the atlas-based technique described in [13] to extract the cardiac structures of interest, namely the LV, RV and myocardium. This is followed by user-friendly manual correction of the segmented contours to correct for potential errors using the ITK-SNAP tool[1]. Note that this segmentation approach will be only used to segment the test data of the challenge and that this paper focuses only on the classification part of the ACDC challenge.

2.3 Radiomics Features for Cardiac Diagnosis

As mentioned in the introduction, most existing techniques included in clinical practice use shape and motion indices such as ejection fraction, ventricular volumes and myocardial thickening to classify the subjects under investigation. This means that a lot of information produced by the image is lost during this operation, in particular imaging evidence in relation to the tissue appearence in the blood pool, myocardium and right ventricle, as well as more complex morphological and functional information. But it is unclear which advanced indices

[1] http://www.itksnap.org/pmwiki/pmwiki.php.

could contribute to improved classification of cardiovascular cases. To address these issues, we propose a radiomics approach for computer-aided diagosis in cine-MRI. Radiomic features have been used so far mostly for cancer image quantification [11,13], such as for the estimation of patient prognosis and treatment response based on the characteristics of the tumors as encoded by the image data. The suffix -omics attached to radiology refers to the use of large amounts of imaging features, from which the most relevant set can be selected for the specific task in question. In this paper, we estimate a large pool of new radiomic features from the segmented cine-MRI images, which will be then analyzed to extract the most powerful features for classification.

In other words, we augment the set of indices to be leveraged for cardiac diagnosis by considering more complex shape/motion radiomic features, as well as advanced textural radiomic features. Specifically, we use 567 features (height, weight, ED-ES duration, plus 188 features per structure: LV, MYO, RV at ED and ES) based on five categories using the PyRadiomics library [14], namely:

(1) Shape based (Volume, surface area, sphericity, compactness, diameters, elongation, etc.).
(2) Intensity first order statistics (e.g. mean, standard deviation, energy, entropy, etc.).
(3) Gray level cooccurence matrix (GLCM) (autocorrelation, contrast, dissimilarity, homogeneity, inverse difference moment, maximum probability, etc.).
(4) Gray level run length matrix (GLRLM) (short/long run emphasis, gray-level/run-length non-uniformity, etc.).
(5) Gray level size zone matrix (GLSZM) (smal/large area emphasis, zone percentage, etc.).

Note that the first group of radiomics consist of pure shape information, while the four remaining groups are intensity-based features, describing the intensity variations inside the cardiac structures, as well as the complexity and repeatability of the tissue texture. Our hypothesis is that some of these radiomic values will be modified in the presence of cardiac abnormality in a way that is unique to each subgroup of patients when compared to normal individuals.

2.4 Classification Method

The next step of our approach is to combine the heterogeneous radiomic features within a classification scheme that will learn to discriminate between the different patient subgroups and normal individuals. In this paper, we choose to use Support Vector Machines (SVMs) [15] due to well-known performance when classifying image data, in particular in the case of small sample size. An SVM model corresponds to a transformation of the examples to a hyperspace where a good separation is achieved by the hyperplanes that have the largest distance to the nearest training-data point of any class (so-called functional margin). This ensures that the examples belonging to the different classes are separated as clearly as possible. New cases are then mapped onto that same hyperspace and

classified based on their location with respect to the hyperplanes separating the different classes. As such, it is suitable for cardiovascular disease classification as the challenge is precisely to identify subtle changes and differences between normal cardiac characteristics and those of pathological cases.

2.5 Radiomic Feature Selection

Due to the large number of radiomic features, radiomic-based classification can easily suffer from over-fitting due to the limited number of examples that can be realistically collected for training. As a result, it is of paramount importance to identify a smaller subset of radiomic features that is optimal for the cardiac diagnosis task. In this paper, we do this by using sequential forward feature selection [16], through which radiomic features will be added to the final subset one at a time until the classification becomes negatively impacted as a result of adding new radiomic features.

3 Results

For all experiments, we used leave-one-out tests to evaluate the proposed method and measured accuracy as proportion correct classifications. Firstly, we evaluated the accuracy of the CVD classifications by using only intensity radiomics or only shape radiomic features. For intensity radiomics, we obtained a maximal accuracy of 0.98 (two misclassifications) when using 13 optimal features. For shape radiomics, we obtained an accuracy of 1.0 (all cases correctly classified) but by using a total of 32 features. Subsequently, we combined intensity, shape and patient information (height and weight) all together and the forward feature selection results are provided in Fig. 2. It can be seen that the best single feature only achieves a 0.62 accuracy. However, after adding three selected features to the classification task, the accuracy is improved beyond the 0.90 accuracy line to reach 0.91 and even 0.94 after five selected features. A maximum accuracy of 1.0 (all cases correctly classified) is reached by combining 10 features only when combining intensity, shape and patient information.

The shape of the curve in Fig. 2 indicates the importance of feature selection, as after reaching the maximum accuracy, incorporating additionnal features leads to model over-fitting and reduced accuracy. While these preliminary results are obtained on a small and controlled study, they are encouraging. In comparison, we obtained an accuracy of 0.84 when using all radiomic features and 0.86 when combining conventional clinical indices only such as ejection fraction, cavity volumes, and body-mass index.

To understand the behavior of the model, we evaluated the precision, recall and confusion matrix after selecting five features, at an accuracy of 0.94. It can be seen that the least accurately detected classes are for the HCM and DCM patients. However, after adding all optimal features, we finally reach a maximal accuracy of 1.0 (Fig. 2).

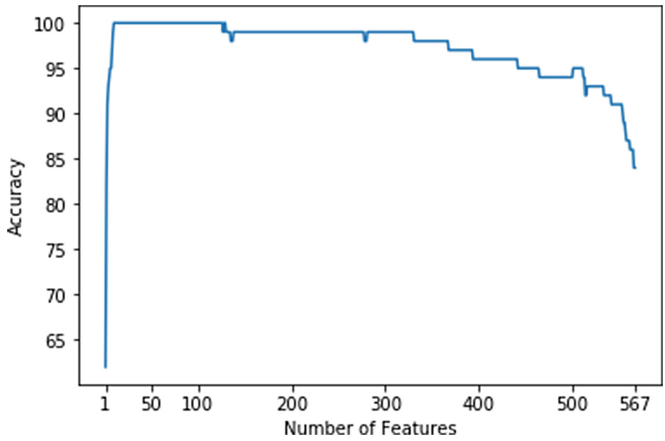


Fig. 2. Accuracy of the proposed CVD classification as a function of the number of radiomic features trained in the model.

Table 1. Precision, recall and confusion matrix obtained by using the five first optimal radiomic features at accuracy of 0.94.

	NOR	DCM	HCM	MINF	RV
Precision	1	0.85	0.9	0.95	1
Recall	0.87	1	0.86	1	1

	NOR	DCM	HCM	MINF	RV
NOR	20	0	0	0	0
DCM	0	17	0	3	0
HCM	2	0	18	0	0
MINF	1	0	0	19	0
RV	0	0	0	0	20

This selected list of optimal features is given in Table 2, which include one conventional shape index (volume), seven advanced shape radiomic features (e.g. compactness, least axis, surface area), one patient information (height) and one textural radiomic feature (GLCM inverse difference). This shows how mutiple radiomics of different nature can be complimentary to each other, which enables to identify correctly all cases. Also, the table shows that the features are well distributed among the three cardiac structures (LV, myocardium, RV), as well as for the ED and ES frames.

To show the relevance of the selected features, we have added to the table the accuracy results by removing each feature from the SVM model (column W/O). It can be seen that the removal of each of these features negatively affects the final accuracy, which is reduced from 1.0 to 0.88 by removing the Surface Area to Volume feature, and to 0.96 by removing the Inverse Difference Intensity (GLCM) feature. This shows how these features can play an role in discriminating some of the challenging and ambiguous cases.

Table 2. List of 10 selected radiomic features as selected by the proposed technique for CVD classification. W/O: Accuracy without the feature. Alone: Accuracy using only this feature.

Name	Type	Frame	Structure	W/O	Alone
Volume	Conventional shape	ED	MYO	0.92	0.5
Surface area to volume	Advanced shape	ES	LV	0.88	0.62
Least axis	Advanced shape	ES	LV	0.95	0.42
Maximum 2D diameter	Advanced shape	ED	LV	0.95	0.41
Maximum 3D diameter	Advanced shape	ES	RV	0.97	0.36
GLCM inverse difference	Intensity/textural	ES	RV	0.96	0.34
Compactness 2	Advanced shape	ES	LV	0.91	0.40
Maximum 3D diameter	Advanced shape	ES	MYO	0.96	0.47
Surface area	Advanced shape	ED	RV	0.97	0.29
Height	Patient Information	-	-	0.91	0.18

To further show the relevance of combining all of the selected, we have also added to the table in the last column the accuracy by using a single radiomic feature. It can be seen that their on their own, these features do not enable a satisfactory classification, with the accuracy values vaying between 0.18 (Height) and 0.62 (Volume). In particular, the Height variable is not capable of producing any meaninful classification on its own, but contributes to the overall accuracy of the multi-radiomic model by normalizing with respect to size.

4 Conclusions

In this paper, we proposed the use of large amounts of radiomic features, integrating advanced shape and textural descriptors, to predict cardiac subgroups. The obtained results suggest that radiomics are indeed capable to encode alterations in the anatomy and tissues of the affected cardiac structures. Furthermore, the feature selection results indicate that shape and intensity descriptors compliment each other and their combinations enable to enhance the prediction power of the system in particular for uncertain cases situated close to the boundary between two disease classes. However, the high accuracy of 1.0 suggests that further evaluations with additional datasets are required to test this radiomics model in larger and more variable data samples. In particular, inter-subject variability due to semi-automatic segmentation of the boundaries will need to be assessed. Future work also includes the testing of additional radiomic features (e.g. fractals, wavelets) and clincal interpretation of the features and results.

Acknowledgments. IC and KL are funded by a Ramon y Cajal research grant (Ryc-2015-17183) from the Spanish Ministry of Economy and Competitiveness. SN is partly funded by a National Institute of Health grant (NIH U01 CA187947). The work of SEP

forms part of the translational research portfolio of the NIHR Biomedical Research Unit at Barts.

References

1. Santulli, G.: Epidemiology of cardiovascular disease in the 21st century: updated numbers and updated facts. J. Cardiovasc. Dis. **1**(1), 1–2 (2013)
2. Peng, P., Lekadir, K., Gooya, A., Shao, L., Petersen, S.E., Frangi, A.F.: A review of heart chamber segmentation for structural and functional analysis using cardiac magnetic resonance imaging. MAGMA **29**, 155–195 (2016)
3. Pennell, D.J.: Cardiovascular magnetic resonance. Circulation **121**(5), 692–705 (2010)
4. Bosch, J.G., Nijland, F., Mitchell, S.C., Lelieveldt, B.P.F., Kamp, O., Reiber, J.H.C., Sonka, M.: Computer-aided diagnosis via model-based shape analysis: automated classification of wall motion abnormalities in echocardiograms. Acad. Radiol. **12**(3), 358–367 (2005)
5. Zhao, F., Wahle, A., Thomas, M.T., Stolpen, A.H., Scholz, T.D., Sonka, M.: Congenital aortic disease: 4D magnetic resonance segmentation and quantitative analysis. Med. Image Anal. **13**(3), 483–493 (2009)
6. Suinesiaputra, A., et al.: Statistical shape modeling of the left ventricle: myocardial infarct classification challenge. IEEE J. Biomed. Health Inf. **PP**(99), 1 (2017)
7. Lekadir, K., Albà, X., Pereañez, M., Frangi, A.F.: Statistical shape modeling using partial least squares: application to the assessment of myocardial infarction. In: Camara, O., Mansi, T., Pop, M., Rhode, K., Sermesant, M., Young, A. (eds.) STACOM 2015. LNCS, vol. 9534, pp. 130–139. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-28712-6_14
8. Lekadir, K., Hoogendoorn, C., Pereanez, M., Alba, X., Pashaei, A., Frangi, A.F.: Statistical personalization of ventricular fiber orientation using shape predictors. IEEE Trans. Med. Imag. **33**(4), 882–890 (2014)
9. Suinesiaputra, A., Frangi, A.F., Kaandorp, T., Lamb, H.J., Bax, J.J., Reiber, J., Lelieveldt, B.: Automated detection of regional wall motion abnormalities based on a statistical model applied to multislice short-axis cardiac MR images. IEEE Trans. Med. Imag. **28**(4), 595–607 (2009)
10. Bai, W., Oktay, O., Rueckert, D.: Classification of myocardial infarcted patients by combining shape and motion features. In: Camara, O., Mansi, T., Pop, M., Rhode, K., Sermesant, M., Young, A. (eds.) STACOM 2015. LNCS, vol. 9534, pp. 140–145. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-28712-6_15
11. Aerts, H.J., Velazquez, E.R., Leijenaar, R.T., Parmar, C., Grossmann, P., Cavalho, S., Bussink, J., Monshouwer, R., Haibe-Kains, B., Rietveld, D., Hoebers, F.: Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. Nature commun. **5**, 4006 (2014)
12. Bai, W., Shi, W., de Marvao, A., Dawes, T.J., O'Regan, D.P., Cook, S.A., Rueckert, D.: A bi-ventricular cardiac atlas built from 1000+ high resolution MR images of healthy subjects and an analysis of shape and motion. Med. Image Anal. **26**(1), 133–145 (2015)
13. Gevaert, O., Xu, J., Hoang, C.D., Leung, A.N., Xu, Y., Quon, A., Rubin, D.L., Napel, S., Plevritis, S.K.: Non-small cell lung cancer: identifying prognostic imaging biomarkers by leveraging public gene expression microarray data-methods and preliminary results. Radiology **264**(2), 387–396 (2012)

14. van Griethuysen, J., Fedorov, A., Parmar, C., Hosny, A., Aucoin, N., Narayan, V., Beets-Tan, R.G.H., Fillion-Robin, J.-C., Pieper, S., Aerts, H.J.W.L.: Computational radiomics system to decode the radiographic phenotype. Cancer Res. **77**(21), 104–107 (2017)
15. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J.: Scikit-learn: machine learning in python. J. Mach. Learn. Res. **12**, 2825–2830 (2011)
16. Guyon, I., Elisseeff, A.: An introduction to variable and feature selection. J. Mach. Learn. Res. **3**, 1157–1182 (2003)

Fast Fully-Automatic Cardiac Segmentation in MRI Using MRF Model Optimization, Substructures Tracking and B-Spline Smoothing

Elias Grinias and Georgios Tziritas(✉)

Department of Computer Science, University of Crete, Heraklion, Greece
tziritas@csd.uoc.gr

Abstract. We present a fast fully automatic method for cardiac segmentation in ED and ES short axis MRI. At first we extract a region where the whole heart is situated, using a new, time-based approach. Then, the segmentation in LV, myocardium and right ventricle (RV) is obtained for a slice in a basal ED slice where both cavities are well distinguished. The extracted regions are tracked for the whole slice sequence backwards and forwards in ED. In all cases the segmentation is based on MRF optimization in four classes, two for the blood areas, and one for the myocardium and the background. Subsequently the segmentation in the ES images is based on the result of ED segmentation. As the epicardium is not well delineated, a smoothing process based on spline curves is used for obtaining the final result. We consider that, with an unsupervised method, we have obtained good results for LV and satisfactory for the RV and the myocardium on the ACDC 2017 datasets.

1 Introduction

Automatic or semi-automatic heart segmentation is a challenging problem. A recent review [1] of the main existing algorithms for the whole heart segmentation shows that the state-of-the-art methods use prior models, either atlas-based or generic deformable models. More specifically, a review on segmentation in short axis cardiac MR images shows also the importance of prior knowledge [2] and emphasizes in different levels of information for different categories of methods. Globally it is concluded that the last decade results useful for the clinical practice have been obtained for LV segmentation, while the RV segmentation is a more difficult and open problem. The RV segmentation has been addressed in a recent MICCAI competition [3]. On the datasets of the competition best results shown that 80% Dice accuracy and a Hausdorff distance less than 1 cm can be expected from semi-automated algorithms. More recently convolutional neural networks have been used in cardiac segmentation with improved performance.

In this paper we present an image-driven fully automatic and fast method, without strong priors, for heart segmentation in ED and ES short axis MRI.

M. Pop et al. (Eds.): STACOM 2017, LNCS 10663, pp. 91–100, 2018.
https://doi.org/10.1007/978-3-319-75541-0_10

We give in the following sections an extended summary of our method illustrated by some results on the MICCAI 2017 ACDC challenge.

2 Automatic Localization of the Heart

Our approach is time-based as the heart is the only moving organ in the image sequences. The mean absolute difference between the corresponding images in the two phases (diastole and systole) is computed. The main region with important intensity change being the heart, taking into account the relative size of this region and for an almost sure decision we extract the biggest connected component C_0 of the change area determined by a threshold at 3% largest change values. We then use the Chan Vese active contours method [5] for extracting an extended region of the heart, with initial region the C_0 component and value the square root of the mean absolute difference between the two phases. A result of the proposed method is illustrated in Fig. 1.

(a) (b) (c)

Fig. 1. Patient 001: (a) Square root of the mean absolute difference between ED and ES phases, (b) the biggest connected component C_0 and (c) the extracted region of interest using active contours algorithm.

3 Segmentation of an ED Phase Slice in Between Base and Mid-Ventricle

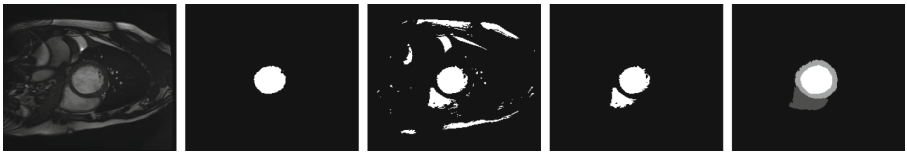
A slice where the two ventricles are well distinguished is needed for starting the recursive segmentation process. It seems possible to determine a slice near the base, where the two cavities appear clearly, taking into account the above determined region of interest. However, for simpicity reasons and without significant quality loss we take the slice at 1/3 position in the z axis. In any case, the segmentation procedure could be initialized by the user for a good first slice segmentation.

We use again the Chan Vese active contours method with initial region the C_0 component for localizing blood pool candidate regions from the intensity image. In addition, the intensity on the region of interest C_0 is quantized in two levels using the *k-means* algorithm with 'cityblock' distance. The connected components of the detected, by the level set algorithm, map are extracted. Then,

the resulting connected components are characterized by their size, their width and the average intensity change between the two phases, V_d. A threshold on V_d is automatically computed, while thresholds are given for the size and the width of regions candidate for being either left, or right, ventricle. The thresholds take into account the spatial resolution of the images and the ground truth from the training dataset. For all regions satisfying the criteria, their position is determined.

After that, we consider three cases accordingly to the number of connected components which could be candidate regions for the two ventricles, with the objective to select almost sure initial segments for the LV and RV cavities.

One component. This component may be related either to the LV or the RV cavity, or it may intersect both. A case where only the LV cavity is initially localized is illustrated in Fig. 2(b). For covering the three possible situations, a sufficient large box around the extracted componet is considered, the intensity value in this box is quantized in four levels and a threshold is obtained for separating blood regions from all other. More than one regions are then obtained, as illustrated in Fig. 2(c). Taking into account the relative position of the regions, their shape and their size the initial localization for LV and RV cavities is extracted, as shown in Fig. 2(d). The procedure is started with the segment that has the maximum intersection area with the initially extracted component. According to its shape and relative position it is classified as LV or RV cavity, assuming that the shape of the LV cavity is close to a circle.



(a) (b) (c) (d) (e)

Fig. 2. (a) Original image of slice S_3 (patient 001), (b) the only extracted component using the active contours algorithm, (c) the blood candidate regions after quantization, (d) the selected LV and RV initial regions and (d) the final MRF segmentation result.

Two components. As we have two regions that almost surely correspond to the two cavities their shape is used for the discrimination. The only shape index used is an eccentricity measure defined by the ratio of the mean distance of the region points to the border over the mean radius of the region. Considering also the relative position of the two regions, the initial segments of the two ventricles are obtained.

More than two components. At first the distance between the candidate regions is measured and compared to their respective size. Only pairs of regions that could likely be at a distance resulting from an acceptable myocardium thickness are considered. The shape and the relative position of the regions is used for their classification and the initial positioning of the LV and RV cavities.

From these initial regions a ring around the LV cavity gives a rough localization of the myocardium. The initial localization of the myocardium and subsequently the statistical description of its intensity is aided by the distance between the two cavities. Finally, we consider four intensity classes and estimate their mean and variance from the initial regions: LV, RV, myocardium and other. We have considered two intensity classes for the blood pool, because of the existing inhomogeneity and variability in its appearance. After the estimation of the four probability density functions, assumed to be Gaussians, the segmentation map is obtained using a Markov random field (*MRF*) model described hereafter.

The segmentation is posed as a probabilistic optimization problem using a discrete MRF in order to obtain a regularized label field. In this work a second order model with 8 connections is employed. The singleton potentials, or priors, are based on the computed probability density functions. The pairwise potentials are set according to the Potts function, where the regularization constant is data adapted.

For minimizing the MRF energy we make use of the *primal-dual* method [4], which casts the MRF optimization problem as an integer programming problem and then makes use of the duality theory of linear programming in order to derive solutions that have been proved to be almost optimal. A result after the last algorithmic step is illustrated in Fig. 2(e). The two blood classes have been merged in one for the final segmentation and region classification.

The segmentation map is then tracked for all slices in ED phase. The procedure described hereafter is applied to the bounding box of the extacted region of the whole heart in slice S_3 for less computational cost and better segmentation accuracy.

4 Segmentation Based on Tracking the Cardiac Substructures in ED Phase

The whole sequence is divided in three parts: basal, mid-ventricular and apical. The segmentation is based on a prediction from the nearest already segmented slice. In the basal part at first the active contour method of Chan Vese is applied with initial mask the region of interest extracted as described in Sect. 2. Using *k-means* algorithm we also obtain the quantization of the intensity in the region of interest. The main connected components candidate for being ventricular cavities are characterized by their size, shape, intensity change between phases, and their relative position. Then the detection of initial LV and RV cavities is based on similarity measures to the nearest extracted cavities and an estimated thickness of the myocardium. Finally, an MRF model is completely estimated from the data and optimized.

In the mid-ventricular zone, initially the LV cavity is localized based on estimated parameters (intensity range, size) from the segmentation map of the nearest already segmented slice. An adaptive quantization method based on LV cavity appearance, on relative position and size is used for localizing the RV cavity. The thickness of the myocardium as estimated from already segmented slices

is used for fixing the inter-ventricles distance. Then, again intensity parameters are estimated and an MRF model is optimized. A similar procedure is adopted for the apical slices, where a specific detection mechanism is also provided as the exact position of the apex is unknown.

An illustration is given in Fig. 3. In the last (apical) slice the RV cavity, which is relatively small and inhologeneous, is not detected.

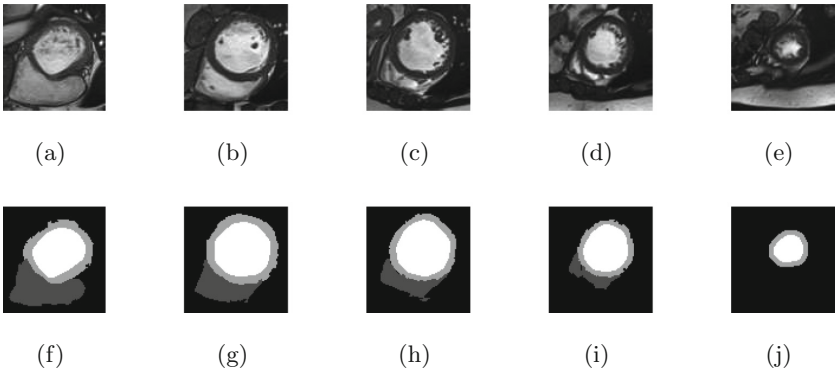


Fig. 3. A sequence of segmentation maps for patient 001.

5 Segmentation in the ES Phase

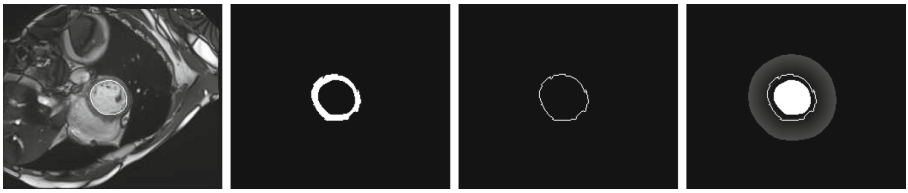
Slice S_3 is again considered first. Having from the ED phase the localization of the whole heart and that of the LV cavity, we search firstly the best threshold for a rough separation of the blood pools. At the same time an initial localization of the LV in the ES phase is obtained. A second threshold is also estimated by the quantization process, as the intensity in the LV and the RV cavities is inhomogeneous and maybe different. Having regions candidates as the locations of RV cavity, we use the Jaccard distance to the ED localized RV and a measure of their relative position to the extracted initial LV cavity, and select the best fitting region as initial RV cavity. The final step is again the MRF optimization in four classes.

The next step is the segmentation performed slice by slice in the basal, mid-ventricular and apical zones. The tracking of the LV and RV cavities is based on the corresponding ED cavities and the nearest ES slice with the cavities localized. Specific detectors consider the cases where the RV or both cavities disappear. The rough registration process gives the initial cavity locations. The final step is again the MRF optimization in four classes.

6 Left Ventricle Epicardial Boundary Smoothing

After segmentation, the "holes" of region R_{LV} of left ventricle are removed, since in many cases the dark papillary muscles are not included in R_{LV}. A boundary

of R_{LV} superimposed on the original image is depicted in Fig. 4(a). Smoothing of the epicardial boundary which surrounds R_{LV} involves the boundary $\mathbf{B}$ of myocardial region R_{epi} of the segmentation result. The boundary between R_{epi} and R_{LV} is not included in $\mathbf{B}$. In image Fig. 4(b), the region R_{epi} which surrounds R_{LV} is shown.



(a) (b) (c) (d)

Fig. 4. Boundary of R_{LV} imposed on original image (a), region R_{epi} (b), boundary points b of R_{epi} (c), distance τ_s of pixels s with $\tau_s < 30$ and (d) extracted boundary of epicardium.

Epicardial smooth boundary is obtained by slightly modifying a method that has been proposed by the authors before, for robustly extracting epicardial boundary after left ventricle endocardium detection [6]. The algorithm fits a cubic B-spline [7] of weighted minimum square error on boundary points. The computation of weights w_b for the points b in set $\mathbf{B}$ is based on the mean distance μ_τ and variance σ_τ^2. Computation of μ_τ and σ_τ^2 involves the pixels s in R_{epi}. In Fig. 4(d), the distance τ_s of pixels from the boundary of R_{LV} is shown. R_{LV} is the white region, while, for the needs of demonstration, pixels with distance greater than 30 are depicted black. The truncated boundary set of points is also drawn in white. The histogram of τ_s for R_{epi} of Fig. 4(b) is shown in Fig. 5. The red bar indicates μ_τ. Estimate σ_τ^2 is used to progressively factor out the impact of boundary points b as their distance τ_b from μ_τ increases.

Cubic spline minimization takes place in the polar (θ, ρ) space defined by the centroid $c = (x_c, y_c)$ of R_{LV}. Thus, boundary points are first transformed to polar coordinates and are sorted according to their value of θ in the range $[-\pi, \pi)$. Boundary data is reproduced in the range $[\pi, 3\pi)$, to reduce the effect of interpolation that is applied by the spline minimization algorithm on the first and last boundary points. Then, a cubic B-spline curve of t segments is fitted to the boundary points in the range $[-\pi, 3\pi)$. The resulting boundary function $\hat{\rho} = f(\theta)$, $\theta \in [0, 2\pi]$, is transformed back to cartesian coordinates. In Fig. 6, the red curve delineates the computed function $\hat{\rho} = f(\theta)$, for $\theta \in [0, 2\pi]$. The transformed boundary points of Fig. 4(c) are the blue crosses, while red circles are the *knots* of cubic polynomial segments.

We present experimental results for a number of patients and for both ED and ES phases. In the images of Fig. 7, the resulting boundaries of epicardium and endocardium are depicted for 4 patient cases and for ED (first and third columns) and ES phases (second and fourth columns), respectively. It is shown

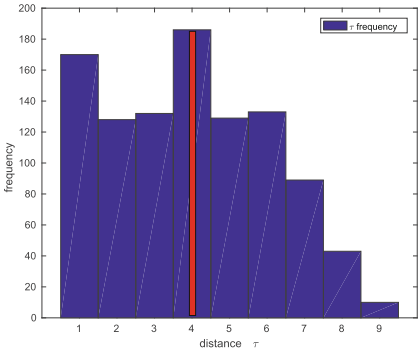


Fig. 5. Histogram of euclidean distance τ of pixels from the boundary of R_{endo} for region R_{epi} of Fig. 4(b). The red bar delineates the estimated mean value μ_τ. (Color figure online)

that the final LV segmentation results are stable and accurate. Furthermore, in all cases the papillary muscles are included in the endocardium region.

The epicardial boundary smoothing is only applied on boundary points separating the myocardium from the background. The endocardial boundary of the RV cavity is not smoothed in the current implementation. The RV segmentation might be improved by smoothing the resulting boundary, as we have observed an improvement of about 5% on Dice measure with the LV epicardial boundary post-processing.

Table 1. Classification and geometrical metrics for the ACDC 2017 training dataset.

	Dice ED	Dice ES	HD ED	HD ES
LV	0.95	0.86	10.16	14.52
Myocardium	0.81	0.76	12.26	16.84
RV	0.81	0.71	24.69	29.90

7 Global Results and First Conclusions

For the ACDC training dataset we give Table 1 the Dice coefficient and the Hausdorff distance for the left and right ventricles and the myocardium. The results on the ejection fraction and the volume in ED and ES phases for the left ventricle are given in Table 2. The results on the mass ED and the volume ES for the myocardium are given in Table 3. All the segmentation maps obtained by our method on the ACDC training dataset are given in https://drive.google.com/file/d/0B1hiv6qJQ9b9THpSem5vQm9tYmc/view?usp=sharing.

For the ACDC testing dataset we give in Table 4 the Dice coefficient and the Hausdorff distance for the left and right ventricles and the myocardium. The

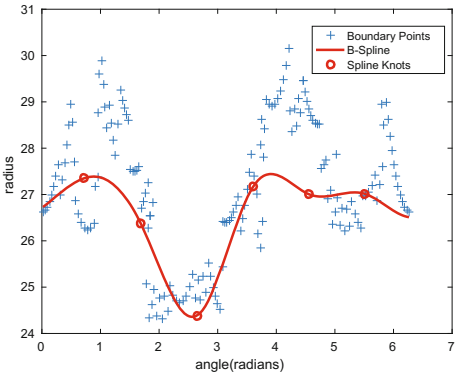


Fig. 6. Cubic B-Spline weighted least squares minimization for the boundary points of Fig. 4(c). (Color figure online)

Table 2. Clinical indice metrics for the left ventricle.

	Ejection fraction	Volume ED	Volume ES
Correlation coefficient	0.901	0.991	0.977
Bias	−0.05	1.43	2.61
Limits of agreement	[−16.93; 16.83]	[−18.07; 20.93]	[−31.89; 37.11]

results on the ejection fraction and the volume in ED phase for the left and right ventricles are given in Table 5. All the segmentation maps obtained by our method on the ACDC testing dataset are given in https://drive.google.com/file/d/0B1hiv6qJQ9b9d3ZoMlRZMGl1U2s/view?usp=sharing.

We consider our results as preliminary and encouraging. The results on the left ventricle could be considered as good, while it would be possible to improve the myocardium result, which also depends on the right ventricle segmentation. We continue to work on the RV segmentation, which is more difficult because of important inhomogeneities in the intensity value and absence of clear separation. As expected, globally the results are better in the ED phase than in the ES phase, as the size of the cavities and their appearance change betwen the two phases. We continue to work on improving the ES phase results. We think that a global post-processing could remove some outlying segmentation results.

Table 3. Clinical indice metrics for the myocardium.

	Mass ED	Volume ES
Correlation coefficient	0.917	0.836
Bias	−13.21	−0.32
Limits of agreement	[−62.55; 36.13]	[−65.45; 64.81]

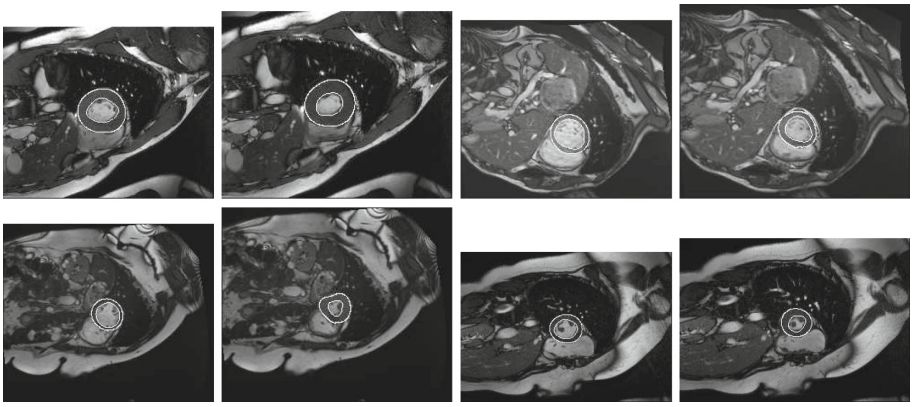


Fig. 7. LV and myocardium segmentation results for 4 patient cases, for ED (first and third column) and ES (second and fourth column) phases respectively.

Table 4. Classification and geometrical metrics for the ACDC 2017 testing dataset.

	Dice ED	Dice ES	HD ED	HD ES
LV	0.95	0.87	8.90	11.57
Myocardium	0.79	0.80	12.59	14.77
RV	0.86	0.74	21.02	25.70

In addition, in Table 6 we give the Dice measure for the five groups of patients for the three structures in the two considered phases. The best result is obtained in the ED phase, where the performance almost does not depend on the group of patients for all structures.

Therefore automatic heart segmentation should be possible without strong priors. More work should be done in the ES phase. The worst results in the ES phase are obtained for patients with previous myocardial infarction (Myo and RV) and for patients with hypertrophic cardiomyopathy (both LV and RV).

An important strong point of our algorithm is its computational efficiency. Processing a sequence of 10 slices for one patient in both ED and ES phases on a laptop Intel Core i-7 2.6 GHz takes less than 10 s for the whole process,

Table 5. Clinical indice metrics for the left and right ventricles (testing dataset).

	LV ejection fraction	Volume LV ED	RV ejection fraction	Volume RV ED
Bias	−1.560	2.002	−0.462	18.566
Standard deviation	4.972	11.699	9.064	25.407

including the localization of the heart, the segmentation in both phases and the epicardium smoothing.

It should be noticed that the hyperparameters of the algorithmic modules have been empirically fixed. We consider that there is place for learning algorithms for some important parameters of the classification and segmentation modules. On the other hand, as the parameters' tuning is often the work of an expert, it would be possible to extend our approach in an interactive mode, for gaining in robustness and limiting outlying results.

Table 6. The Dice measure by group of patients.

	Average	Healthy	Previous myocardial infarction	Dilated cardio-myopathy	Hypertrophic cardio-myopathy	Abnormal right ventricle
ED LV	0.95	0.95	0.94	0.96	0.93	0.95
ED Myo	0.81	0.82	0.78	0.81	0.82	0.81
ED RV	0.81	0.84	0.80	0.79	0.80	0.83
ES LV	0.86	0.86	0.90	0.92	0.76	0.86
ES Myo	0.76	0.79	0.71	0.75	0.75	0.79
ES RV	0.71	0.75	0.66	0.72	0.65	0.78

References

1. Zhuang, X.: Challenges and methodologies of fully automatic whole heart segmentation: a review. J. Healthcare Eng. **4**, 371–407 (2013)
2. Petitjean, C., Dacher, J.-N.: A review of segmentation methods in short axis cardiac MR images. Med. Image Anal. **15**, 169–184 (2011)
3. Petitjean, C., et al.: Right ventricle segmentation from cardiac MRI: a collation study. Med. Image Anal. **19**, 187–202 (2015)
4. Komodakis, N., Tziritas, G.: Approximate labeling via graph cuts based on linear programming. IEEE Trans. Pattern Anal. Mach. Intell. **29**, 1436–1453 (2007)
5. Chan, T.F., Vese, L.A.: Active contours without edges. IEEE Trans. Image Process. **10**, 266–277 (2001)
6. Mazonakis, M., Grinias, E., Pagonidis, K., Tziritas, G., Damilakis, J.: Development and evaluation of a semi-automatic segmentation method for the estimation of LV parameters on cine MR images. Phys. Med. Biol. **55**, 1127–1150 (2010)
7. Bartles, R.H., Beatty, J.C., Barsky, B.A.: An Introduction to Splines for Use in Computer Graphics and Geometric Modeling. Morgan Kaufmann Publishers, Los Altos (1987)

Automatic Segmentation and Disease Classification Using Cardiac Cine MR Images

Jelmer M. Wolterink[1(✉)], Tim Leiner[2], Max A. Viergever[1], and Ivana Išgum[1]

[1] Image Sciences Institute, University Medical Center Utrecht, Utrecht, The Netherlands
j.m.wolterink@umcutrecht.nl

[2] Department of Radiology, University Medical Center Utrecht, Utrecht, The Netherlands

Abstract. Segmentation of the heart in cardiac cine MR is clinically used to quantify cardiac function. We propose a fully automatic method for segmentation and disease classification using cardiac cine MR images. A convolutional neural network (CNN) was designed to simultaneously segment the left ventricle (LV), right ventricle (RV) and myocardium in end-diastole (ED) and end-systole (ES) images. Features derived from the obtained segmentations were used in a Random Forest classifier to label patients as suffering from dilated cardiomyopathy, hypertrophic cardiomyopathy, heart failure following myocardial infarction, right ventricular abnormality, or no cardiac disease.

The method was developed and evaluated using a balanced dataset containing images of 100 patients, which was provided in the MICCAI 2017 automated cardiac diagnosis challenge (ACDC). Segmentation and classification pipeline were evaluated in a four-fold stratified cross-validation. Average Dice scores between reference and automatically obtained segmentations were 0.94, 0.88 and 0.87 for the LV, RV and myocardium. The classifier assigned 91% of patients to the correct disease category. Segmentation and disease classification took 5 s per patient.

The results of our study suggest that image-based diagnosis using cine MR cardiac scans can be performed automatically with high accuracy.

Keywords: Deep learning · Random Forest
Convolutional neural networks · Cardiac MR · Automatic diagnosis

1 Introduction

Quantification of volumetric changes in the heart during the cardiac cycle is essential for diagnosis and monitoring of cardiac diseases. To this end, quantitative indices such as the ejection fraction and myocardial mass are typically

M. Pop et al. (Eds.): STACOM 2017, LNCS 10663, pp. 101–110, 2018.
https://doi.org/10.1007/978-3-319-75541-0_11

extracted based on segmentations of the ventricular cavities and myocardium and used to identify patients suffering from cardiac diseases [6].

However, segmentation of the ventricular cavities and myocardium in cine MR is a challenging problem [9]. Cine MR images are highly anisotropic, contrast in these images may be poor, and cardiac diseases may cause large variations in patient anatomy. The development of accurate cine MR segmentations methods is an ongoing endeavor [10], which has recently seen contributions from deep learning methods, e.g. [5,11].

We propose a method for fully automatic segmentation of the LV cavity, the RV cavity and the myocardium in cardiac cine MR images. We use a deep learning method for cardiac cine MR segmentation and show that this method achieves high overlap with manual reference segmentations. Furthermore, we show how basic quantitative features extracted from the automatically obtained segmentations can be combined with patient information in a forest of randomized decision trees. This allows fast and accurate disease classification in cardiac patients, and detection of patients with ambiguous indications.

2 Data

The proposed method was developed and evaluated using data from the MICCAI 2017 *a*utomated *c*ardiac *d*iagnosis *c*hallenge (ACDC). This challenge provides a dataset consisting of cine MR images of 150 patients who have been clinically diagnosed in five classes: normal, dilated cardiomyopathy (DCM), hypertrophic cardiomyopathy (HCM), heart failure with infarction (MINF), or right ventricular abnormality (RVA). Thirty cases are provided in each class. The data set was separated by the challenge organizers into 100 training cases for which a reference standard was provided, and 50 test cases for which no reference standard was provided. Here, we describe experiments and results using the 100 training cases.

For each patient, short axis (SA) cine MR images with 12–35 frames are available, in which the end-diastole (ED) and end-systole (ES) frame have been indicated. The image slices cover the LV from the base to the apex. In-plane voxel spacing varies from 1.37 to 1.68 mm and inter-slice spacing varies from 5 to 10 mm. Manual reference segmentations of the LV cavity, RV cavity and myocardium at ED and ES are provided.

To correct for differences in voxel size, all 2D image slices were resampled to 1.4×1.4 mm^2 spacing. Furthermore, to correct for image intensity differences between images, each MR volume was normalized between $[0.0, 1.0]$ according to the 5th and 95th percentile of intensities in the image.

3 Methods

We propose a fully automatic method for segmentation and diagnosis in cardiac cine MR images. The method uses a convolutional neural network (CNN) to segment the LV cavity, the RV cavity, and the myocardium in 2D short-axis cine MR slices. Quantitative indices are extracted from the obtained segmentations and combined with patient information in a Random Forest classifier that assigns patients to one of five classes (normal, DCM, HCM, MINF, RVA).

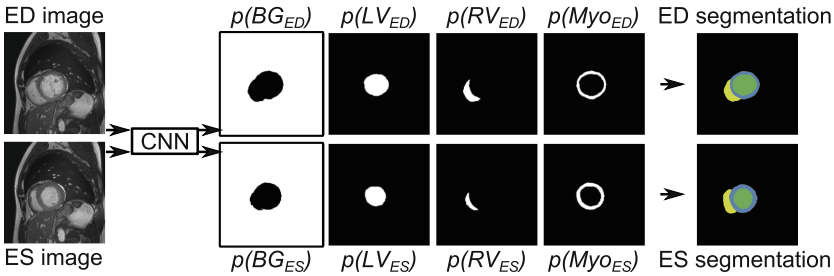


Fig. 1. Convolutional neural network for segmentation. The CNN uses anatomically aligned end-diastole (ED) and end-systole (ES) 2D image slices as input and simultaneously predicts probability maps for the background (BG), LV cavity (LV), RV cavity (RV) and myocardium (Myo) at ED and ES. These are combined into two multi-class segmentations.

3.1 Segmentation

A CNN was trained to segment the LV cavity, the RV cavity, and the myocardium in 2D short-axis cine MR slices. Motivated by [12,13], the network was designed to contain a number of convolutional layers with increasing levels of dilation. This ensures a large receptive field with few trainable parameters and high resolution feature maps. The final receptive field for each voxel was 131×131 voxels, or 18.3×18.3 cm^2 in the resampled 2D slices. Potential overfitting of the network was mitigated by the inclusion of Batch Normalization layers [4].

Cine cardiac MR slices obtained throughout the cardiac cycle are anatomically aligned, but cardiac motion causes differences between images at different time points. These differences are more pronounced in the heart than in other areas [1]. We allowed the CNN to leverage this information for heart localization by simultaneously providing anatomically corresponding ED and ES slices in two input channels (Fig. 1). The CNN had eight output channels; four for ED labels (BG_{ED}, LV_{ED}, RV_{ED}, Myo_{ED}), and four for ES labels (BG_{ES}, LV_{ES}, RV_{ES}, Myo_{ES}). ED and ES output channels were separately normalized through softmax functions. Hence, for both the ED and the ES image there were four probability maps summing to 1. A segmentation was obtained for both images by assigning the class with the highest probability to each voxel.

The trainable parameters in the CNN were optimized using a loss function based on the Dice similarity coefficient [7]. This partly corrects for class imbalance in the voxel labels. A soft Dice loss was used,

$$Dice_c = \frac{\sum_i^N R_c(i) A_c(i)}{\sum_i^N R_c(i) + \sum_i^N A_c(i)}, \tag{1}$$

where R_c is the binary reference image for class c, A_c is the probability map for class c, N is the number of voxels, and $Dice_c$ is the Dice coefficient for class c. This coefficient was computed for all eight classes (BG_{ED}, RV_{ED}, Myo_{ED},

LV_{ED}, BG_{ES}, RV_{ES}, Myo_{ES}, LV_{ES}) and averaged to ensure joint optimization for all classes.

The combination of multiple trained CNN models in an ensemble typically results in more accurate predictions, but at the cost of repeated training. To obtain multiple models with a single training phase, we used the snapshot ensemble technique proposed in [3]. Hence, the learning rate followed a cyclic scheme according to the equation

$$\alpha_i = \frac{\alpha_0}{2}\left(\cos\left(\frac{\pi \mathrm{mod}(t-1, M)}{M}\right) + 1\right), \tag{2}$$

where α_i is the current learning rate, α_0 is the initial learning rate and M is the cycle length, i.e. the number of iterations before a reset of the learning rate to the initial value. We set the total number of iterations to 150,000 and reset the learning rate to $\alpha_0 = 0.2$ after every $M = 10,000$ iterations. A copy of the model was stored before each learning rate reset. Stochastic gradient descent was used for training, with $L2$-regularization on the parameters of the CNN. In each iteration, the network was optimized with a mini-batch containing 4 images with 151×151 voxel samples, padded to 281×281 to accommodate the 131×131 voxel receptive field. The training data was augmented by 90° rotations of the images and reference segmentations.

During testing, pairs of ED and ES images were processed by the six stored versions of the model between 100,000 and 150,000 iterations. The six predicted probability maps for each class were averaged before the class label with the largest probability was assigned to each voxel. No post-processing was applied other than selection of the largest 3D 6-connected component for each class.

3.2 Diagnosis

Each patient was described by patient and image characteristics. Patient characteristics were patient weight (in kg) and patient height (in cm). Image characteristics were extracted from the automatically obtained segmentations: LV, RV and myocardial volume at ED and ES (in ml), the LV and RV ejection fraction (EF), the ratio between RV and LV volume at ED and ES, and the ratio between myocardial and LV volume at ED and ES. Hence, 14 features were used in total: 2 patient-based and 12 image-based features.

A five-class (normal, HCM, DCM, MINF, RVA) Random Forest classifier [2] was trained, consisting of 1,000 decision trees that were grown to full depth. For each case, a posterior probability distribution was obtained. Patients were assigned to the class with the highest probability, and the entropy in the probability distribution was determined to estimate uncertainty of the classifier.

4 Experiments and Results

The proposed method was evaluated using the ACDC training set in a stratified four-fold cross-validation experiment. For each fold the system was trained using

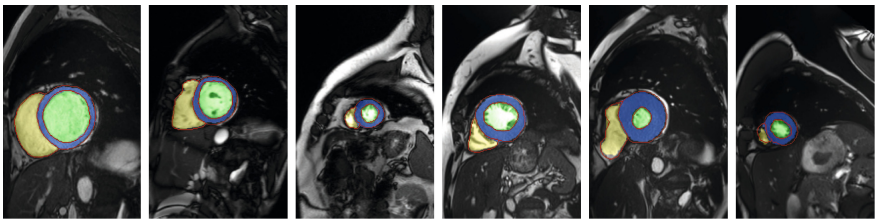


Fig. 2. Example segmentations obtained by the CNN in six different patients, showing the LV cavity in green, the RV cavity in yellow, and the myocardium in blue. Reference delineations are shown in red (Color figure online).

Table 1. Average Dice coefficients (Dice) and Hausdorff distances (HD, in mm) for LV, RV and myocardium segmentation at ED and ES.

	LV		RV		Myocardium	
	Dice	HD	Dice	HD	Dice	HD
ED	0.96 ± 0.02	8.35 ± 4.63	0.92 ± 0.04	13.39 ± 5.68	0.86 ± 0.04	11.77 ± 6.38
ES	0.91 ± 0.07	9.01 ± 4.39	0.84 ± 0.09	15.03 ± 6.30	0.88 ± 0.04	10.85 ± 4.74
Total	0.93 ± 0.05	8.68 ± 4.51	0.88 ± 0.08	14.21 ± 6.04	0.87 ± 0.04	11.31 ± 5.62

15 training patients from each of the five classes and evaluated using five test patients from each of the five classes. Quantitative indices used to train the Random Forest in a fold were obtained with the trained CNN for that fold. Hence, training and validation set were completely separated throughout both stages. We here present combined results on all 100 training images.

4.1 Segmentation Results

Figure 2 shows example segmentations and the corresponding reference delineations. All obtained segmentations were evaluated using the online platform provided by the organizers of the ACDC challenge. Table 1 shows Dice coefficients and Hausdorff distances for LV, RV and myocardium segmentation at ED and ES. Agreement with the reference standard was highest for the left ventricle at ED, and lowest for the right ventricle at ES. Performance was substantially higher in ED than in ES for the LV and RV, but not for the myocardium. Average Hausdorff distances were lower for the LV than for the RV and myocardium. This might be caused by the strong contrast that is typically present between the LV and the surrounding myocardium, and the poorer contrast between the myocardium and its surrounding structures. Furthermore, the shape of the RV is more irregular than that of the LV (Fig. 2).

In addition to the Dice coefficient and Hausdorff distance, the agreement between reference and automatically derived quantitative indices was determined

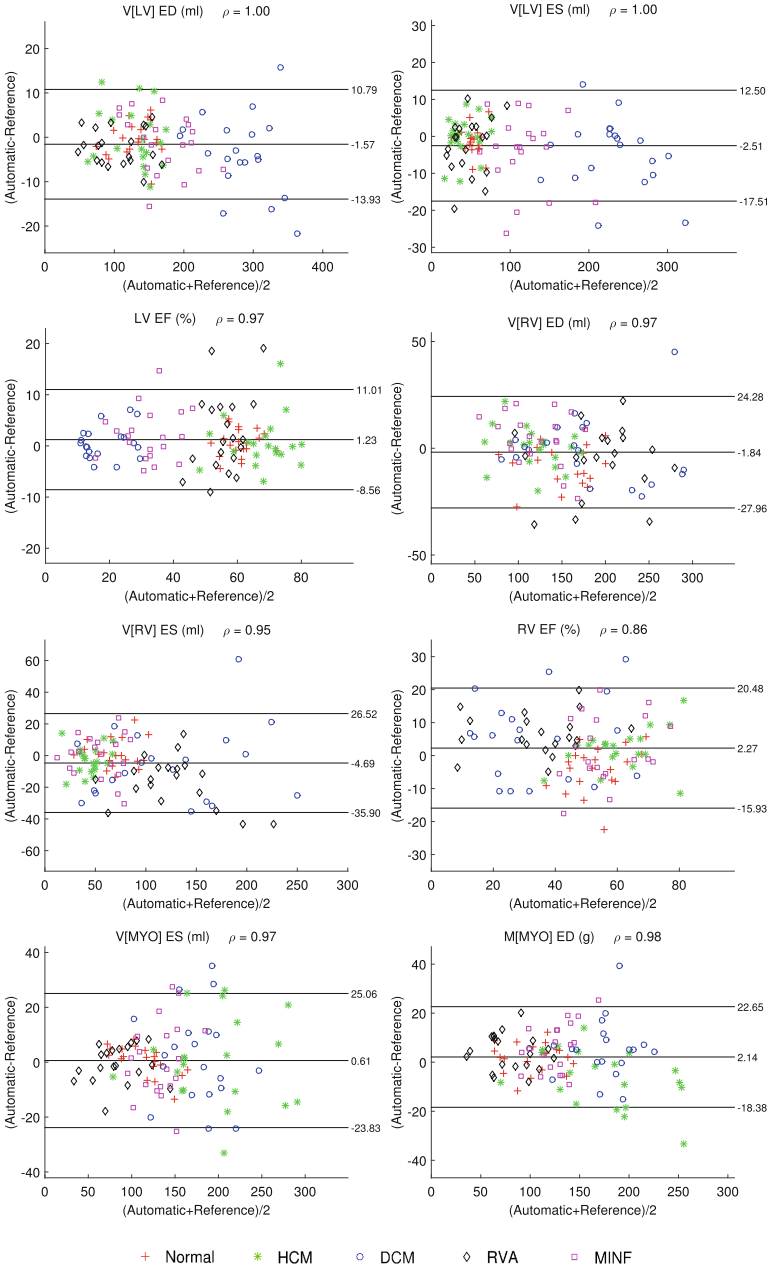


Fig. 3. Bland-Altman plots showing the agreement between reference and automatic quantification of the end-diastolic (ED) and end-systolic (ES) volume (V [in ml]) or mass (M [in g]) of the left ventricle (LV), right ventricle (RV) and myocardium (MYO). Bland-Altman limits of agreement and Pearson correlation values are listed. Points correspond to patients, with markers indicating classes.

Table 2. Agreement in diagnosis between the reference standard and automatic classification. Patients were classified as normal, dilated cardiomyopathy (DCM), hypertrophic cardiomyopathy (HCM), heart failure with infarction (MINF), or right ventricular abnormality (RVA). Overall classification accuracy was 91%.

		Automatic					
		Normal	DCM	HCM	MINF	RVA	Total
Reference	Normal	**20**	0	0	0	0	20
	DCM	0	**18**	0	2	0	20
	HCM	2	0	**18**	0	0	20
	MINF	1	2	0	**17**	0	20
	RVA	2	0	0	0	**18**	20
	Total	25	20	18	19	18	100

using the ACDC challenge online platform. Figure 3 shows Bland-Altman plots with limits of agreement and Pearson correlations for the agreement between LV, RV and myocardium volume or mass quantification at ED and ES, as well as the LV and RV EF. There was a slight underestimation of the LV and RV at both ED and ES, and a slight overestimation of the myocardium at both ED and ES. The Pearson correlation between reference and automatically determined LV EF values was 0.97, while this correlation was 0.86 for the RV EF, reflecting lower segmentation accuracy for the RV.

The CNN was implemented in Theano and Lasagne. Segmentation with an ensemble of six trained CNNs took 4 s per patient on a NVIDIA Titan X GPU.

4.2 Diagnosis Results

The obtained diagnoses were evaluated in the online platform provided by the ACDC organizers. Table 2 shows the confusion matrix for classification into five categories, with an overall accuracy of 91%. Sensitivity was 100% for the normal class, 90% for the DCM, HCM and RVA classes, and 85% for the MINF class. Four out of nine errors were made by confusion between myocardial infarction and dilated cardiomyopathy, both of which are characterized by low LV EF values.

The three most important features as determined by the Random Forest were the left ventricular ejection fraction (LV EF), the ratio between right and left ventricular volume at ED (V[RV]/V[LV] ED), and the ratio between myocardial and left ventricular volume at ES (V[MYO]/V[LV] ES). Figure 4 shows the feature value distribution over the five different classes, showing several clear patterns. RVA patients generally have a large RV to LV volume ratio compared with normal patients. Patients with DCM and MINF have a reduced LV EF, while this value is higher for normal patients and patients with HCM. The myocardial volume is relatively small compared with the LV volume in patients with DCM, indicating thinning of the myocardium, but large in patients with HCM.

However, not all cases can be clearly separated using these features. Based on the entropy in the posterior probabilities provided by the Random Forest classifier, several patients with high classification uncertainty could be identified. Figure 5 shows automatically obtained segmentations in two such patients.

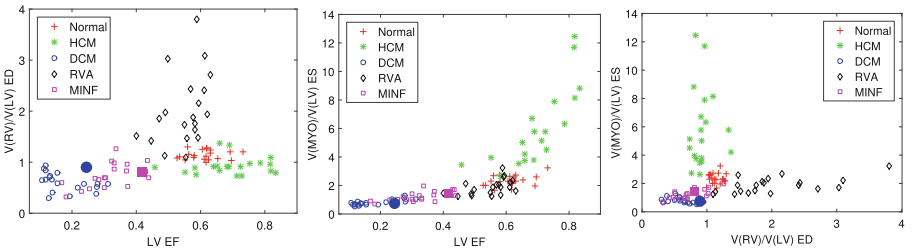
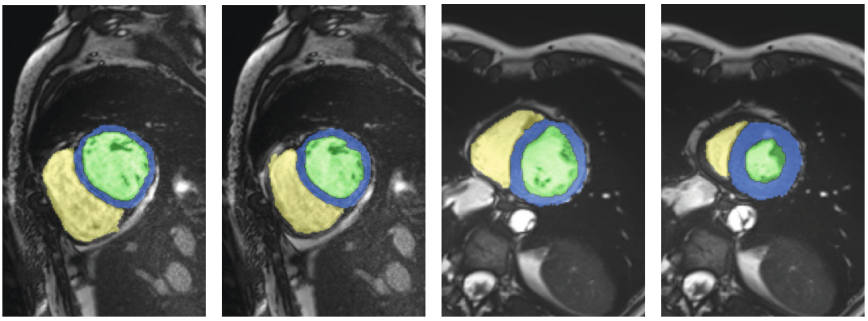


Fig. 4. Three most important features according to the Random Forest classifier: left ventricular ejection fraction (LV EF), volume ratio between right and left ventricle at ED (V[RV]/V[LV] ED), volume ratio between myocardium and left ventricle at ES (V[MYO]/V[LV] ES). Each point corresponds to a patient.



(a) Patient 18 ED (b) Patient 18 ES (c) Patient 44 ED (d) Patient 44 ES

Fig. 5. Two cases in which the classifier showed high uncertainty. The reference diagnosis for Patient 18 was DCM, but the patient was classified as MINF. The reference diagnosis for Patient 44 was MINF (LV EF 41.7%), but the patient was classified as normal.

Patient 18 (indicated by a blue circle in Fig. 4) was incorrectly diagnosed as MINF (classification probability $p = 0.45$), while the reference diagnosis was DCM ($p = 0.25$). In addition, there was a substantial probability for RVA ($p = 0.21$). In this patient, LV EF and the ratio between RV and LV at ED were both relatively high compared with other DCM patients. Patient 44 (indicated by a magenta square in Fig. 4) was incorrectly diagnosed as normal ($p = 0.44$), while the reference diagnosis was MINF ($p = 0.31$). This patient had a high LV EF value of 41.7% compared with other MINF patients.

Extraction of features based on segmentations and classification using the Random Forest classifier took around 1 s per patient.

5 Discussion and Conclusion

We have presented a method for fully automatic segmentation and diagnosis in cardiac cine MR images. The results show that automatically obtained segmentations of the left ventricle, right ventricle, and myocardium have good overlap with manual reference segmentations. Furthermore, based on these segmentations patients can be diagnosed with 91% multi-class accuracy.

While disease classification based on quantitative descriptors extracted from cine MR typically follows clinical guidelines, we have shown here that these guidelines can partially be captured in a Random Forest classifier. Furthermore, the posterior probability distribution of the classifier can be used to identify patients which cannot easily be assigned to a single disease. In future work, we will further investigate to what extent uncertainty of the classifier corresponds to uncertainty in the clinical diagnosis.

Deep learning methods have been shown to provide state-of-the-art results in a wide range of medical imaging problems. Here, we used deep learning for segmentation of the cine MR images. However, we opted for a more conventional Random Forest approach for patient classification because of the small training dataset at hand. A potential limitation of this two-stage approach is that errors in the segmentation stage may affect performance in the classification stage. However, a Bland-Altman comparison of quantitative indices derived from reference and automatic segmentations (Fig. 3) showed only very small bias values, comparable to interstudy and intraobserver differences in cardiac MR of normal healthy adults [8]. Moreover, patient classification using quantitative indices derived from the reference segmentations instead of the automatic segmentations resulted in only minor improvements (classification accuracy 92% instead of 91%) with a considerable increase in time and effort.

In the current study, we only used the ED and ES images. However, cine MR contains a whole sequence of images. In future work we will investigate whether inclusion of the complete sequence of images as input to the CNN could improve segmentation, and whether features derived from this sequence could provide additional value for disease diagnosis.

References

1. Atehortúa, A., Zuluaga, M.A., García, J.D., Romero, E.: Automatic segmentation of right ventricle in cardiac cine MR images using a saliency analysis. Med. Phys. **43**(12), 6270–6281 (2016)
2. Breiman, L.: Random forests. Mach. Learn. **45**(1), 5–32 (2001)
3. Huang, G., Li, Y., Pleiss, G., Liu, Z., Hopcroft, J.E., Weinberger, K.Q.: Snapshot Ensembles: Train 1, Get M for Free. arXiv preprint arXiv:1704.00109 (2017)
4. Ioffe, S., Szegedy, C.: Batch normalization: accelerating deep network training by reducing internal covariate shift. In: ICML (2015)
5. Lieman-Sifry, J., Le, M., Lau, F., Sall, S., Golden, D.: FastVentricle: cardiac segmentation with ENet. In: Pop, M., Wright, G.A. (eds.) FIMH 2017. LNCS, vol. 10263, pp. 127–138. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-59448-4_13

6. Marcus, F.I., McKenna, W.J., Sherrill, D., Basso, C., Bauce, B., Bluemke, D.A., Calkins, H., Corrado, D., Cox, M.G., Daubert, J.P., et al.: Diagnosis of arrhythmogenic right ventricular cardiomyopathy/dysplasia. Eur. Heart J., ehq025 (2010)
7. Milletari, F., Navab, N., Ahmadi, S.A.: V-net: Fully convolutional neural networks for volumetric medical image segmentation. In: Fourth International Conference on 3D Vision (3DV), 2016, pp. 565–571. IEEE (2016)
8. Moody, W., Edwards, N., Chue, C., Taylor, R., Ferro, C., Townend, J., Steeds, R.: Variability in cardiac MR measurement of left ventricular ejection fraction, volumes and mass in healthy adults: defining a significant change at 1 year. Br. J. Radiol. **88**(1049), 20140831 (2015)
9. Petitjean, C., Dacher, J.N.: A review of segmentation methods in short axis cardiac MR images. Med. Image Anal. **15**(2), 169–184 (2011)
10. Petitjean, C., Zuluaga, M.A., Bai, W., Dacher, J.N., Grosgeorge, D., Caudron, J., Ruan, S., Ayed, I.B., Cardoso, M.J., Chen, H.C., et al.: Right ventricle segmentation from cardiac MRI: a collation study. Med. Image Anal. **19**(1), 187–202 (2015)
11. Tran, P.V.: A fully convolutional neural network for cardiac segmentation in short-axis MRI. arXiv preprint arXiv:1604.00494 (2016)
12. Wolterink, J.M., Leiner, T., Viergever, M.A., Išgum, I.: Dilated convolutional neural networks for cardiovascular MR segmentation in congenital heart disease. In: Zuluaga, M.A., Bhatia, K., Kainz, B., Moghari, M.H., Pace, D.F. (eds.) RAMBO/HVSMR -2016. LNCS, vol. 10129, pp. 95–102. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-52280-7_9
13. Yu, F., Koltun, V.: Multi-scale context aggregation by dilated convolutions. In: ICLR (2016)

An Exploration of 2D and 3D Deep Learning Techniques for Cardiac MR Image Segmentation

Christian F. Baumgartner[1], Lisa M. Koch[2(✉)], Marc Pollefeys[2], and Ender Konukoglu[1]

[1] Computer Vision Lab, ETH Zurich, Zürich, Switzerland
lisa.koch@inf.ethz.ch
[2] Computer Vision and Geometry Group, ETH Zurich, Zürich, Switzerland

Abstract. Accurate segmentation of the heart is an important step towards evaluating cardiac function. In this paper, we present a fully automated framework for segmentation of the left (LV) and right (RV) ventricular cavities and the myocardium (Myo) on short-axis cardiac MR images. We investigate various 2D and 3D convolutional neural network architectures for this task. Experiments were performed on the ACDC 2017 challenge training dataset comprising cardiac MR images of 100 patients, where manual reference segmentations were made available for end-diastolic (ED) and end-systolic (ES) frames. We find that processing the images in a slice-by-slice fashion using 2D networks is beneficial due to a relatively large slice thickness. However, the exact network architecture only plays a minor role. We report mean Dice coefficients of 0.950 (LV), 0.893 (RV), and 0.899 (Myo), respectively with an average evaluation time of 1.1 s per volume on a modern GPU.

1 Introduction

Cardiovascular diseases are a major public health concern and currently the leading cause of death in Europe [12]. Automated segmentation of cardiac structures from medical images is an important step towards analysing normal and pathological cardiac function on a large scale, and ultimately towards developing diagnosis and treatment methods.

Until recently, the field of anatomical segmentation was dominated by atlas-based techniques (e.g. [2]), which have the advantage of providing strong spatial priors and yielding robust results with relatively little training data. With more data becoming available and recent advances in machine learning and parallel computing infrastructure, segmentation techniques based on deep convolutional neural networks (CNN) are emerging as the new state-of-the-art [9,15].

This paper is dedicated to the segmentation of cardiac structures on short-axis MR images and is accompanied by a submission to the automated cardiac

C. F. Baumgartne and L. M. Koch contributed equally.

M. Pop et al. (Eds.): STACOM 2017, LNCS 10663, pp. 111–119, 2018.
https://doi.org/10.1007/978-3-319-75541-0_12

diagnosis challenge (ACDC) 2017. Short-axis MR images consist of a stack of 2D MR images acquired over multiple cardiac cycles which are often not perfectly aligned and typically have a low through-plane resolution of $5 - 10$ mm.

In this paper, we investigate the suitability of state-of-the-art 2D and 3D CNNs for the segmentation of three cardiac structures. A specific focus is to answer the question if 3D context is beneficial for this task in light of the low through-plane resolution. Furthermore, we explore different network architectures and employ a variety of techniques which are known to enhance training and inference performance in deep neural network such as batch normalisation [6], and different loss functions [11]. The proposed framework was evaluated on the training set for the ACDC 2017 segmentation challenge. Accurate segmentation results were obtained with a fast inference time of 1.1 s per 3D image.

2 Method

In the following, we will outline the individual steps focusing on the preprocessing, network architectures, optimisation and post-processing of the data.

2.1 Pre-Processing

Since the data were recorded at varying resolutions, we resampled all images and segmentations to a common resolution. For the networks operating in 2D, the images were resampled to an in-plane resolution of 1.37×1.37 mm. We did not perform any resampling in the through-plane direction to avoid any losses in accuracy in the up- and downsampling steps. Part of the data had a relatively low through-plane resolution of 10 mm and we found that losses incurred by resampling artefacts can be significant. For the 3D network we chose a resolution of $2.5 \times 2.5 \times 5$ mm. Higher resolutions were not possible due to GPU memory restrictions. We then placed all the resampled images centrally into images of constant size, padding with zeros where necessary. The exact image size depended on the network architecture and will be discussed below. Lastly, each image was intensity-normalised to zero mean and unit variance.

2.2 Network Architectures

We investigated four different network architectures. The fully convolutional segmentation network (FCN) proposed by [10] is a 2D segmentation network widely used for natural images. In this architecture deep, and thus coarse, feature maps are upsampled to the original image resolution by using transposed convolutions. In order to fuse the semantic information available in the deeper layers with the spatial information available in the shallower stages, the authors proposed to use skip connections. In the present work, we used the best performing incarnation which is based on the VGG-16 architecture and uses three skip connections (FCN-8) [10]. We used an image size of 224×224 pixels for this architecture.

Another popular segmentation architecture is the 2D U-Net initially proposed for the segmentation of neuronal structures in electron microscopy stacks and cell tracking in light microscopy images [15]. Inspired by [10] the authors employ an architecture with symmetric up- and downsampling paths and skip connections within each resolution stage. Since this architecture does not employ padded convolutions, a larger image size of 396×396 pixels was necessary, which led to segmentation masks of size 212×212 pixels.

Inspired by the fact that the FCN-8 produces competitive results despite having a simple upsampling path with few channels, we speculated that the full complexity of the U-Net upsampling path may not be necessary for our problem. Therefore, we additionally investigated a modified 2D U-Net with number of feature maps in the transpose convolutions of the upsampling path set to the number of classes. Intuitively, each class should have at least one channel.

Çiçek et al. recently extended the U-Net architecture to 3D [4] by following the same symmetric design principle. However, for data with few slices in one orientation, the repeated pooling and convolving may be too aggressive. We found that using the 3D U-Net for our data all spatial information in the through-plane direction was lost before the third max pooling step. We thus also investigated a slightly modified version of the 3D U-Net in which we performed only one max-pooling (and upsampling) step in the through-plane direction. This had two advantages: (1) The spatial information in the through-plane was retained and thus available in the deeper layers, (2) it allowed us to work with a slightly higher image resolution because less padding in the through-plane direction (and thus less GPU memory) was required. In preliminary experiments we found that the modified 3D U-Net led to improvements of around 0.02 of the average Dice score over the standard 3D U-Net. In the interest of brevity we only included the modified version in the final results of this paper. Here, we used an input image size of $204 \times 204 \times 60$, which led to output masks of size $116 \times 116 \times 28$.

We used batch normalisation [6] on the outputs of every convolutional and transposed convolutional layer for all architectures. We found that this not only led to faster convergence, as reported in [4], but also consistently yielded better results and allowed the training of some networks to converge that did not converge otherwise.

2.3 Optimisation

We trained the networks introduced above (i.e. FCN-8, 2D U-Net, 2D U-Net (mod.) and 3D U-Net (mod.)) from scratch with the weights of the convolutional layers initialised as described in [5].

We investigated three different cost functions. First, we used the standard pixel-wise cross entropy. To account for the class imbalance between the background and the foreground classes, we also investigated a weighted cross entropy loss. We used a weight of 0.1 for the background class, and 0.3 for the foreground classes in all experiments in this paper, which corresponds approximately to the inverse prevalence of each label in the dataset. Lastly, we investigated optimising

the Dice coefficient directly. In order to get more stable gradients we calculated the Dice loss on the softmax output as follows:

$$\mathcal{L}_{dice} = 1 - \frac{\sum_{k=2}^{K} \sum_{n=1}^{N} t_{nk} y_{nk}}{\sum_{k=2}^{K} \sum_{n=1}^{N} t_{nk} + y_{nk}},$$

where K is the number of classes, N the number of pixels/voxels, y is the softmax output, t is a one-hot vector encoding the true label per location.

To minimise the respective cost functions we used the ADAM optimiser [7] with a learning rate of 0.01, $\beta_1 = 0.9$ and $\beta_2 = 0.999$. The best results were obtained without using any weight regularisation. The training of each of the models took approximately 24 h on a Nvidia Titan Xp GPU.

2.4 Post-Processing

Since training and inference were performed in a different resolution, the predictions had to be resampled to each subject's initial resolution. To avoid resampling artefacts, this step was carried out on the softmax (i.e. continuous) network outputs for each label using linear interpolation. The final discrete segmentation was then obtained in the final resolution by choosing the label with the highest score at each voxel. Interpolation on the softmax output, rather than the output masks, led to consistent improvements of around 0.005 in the average Dice score.

We occasionally observed spurious predictions of structures in implausible locations. To compensate for this, we applied simple post-processing to the segmentation results by keeping only the largest connected component for every structure. Since the segmentations are already quite accurate without post-processing this only lead to an average Dice increase of approximately 0.0003, however, it reduced the Hausdorff distance considerably, which by definition is very sensitive to outliers. Other post-processing techniques such as the commonly used spatial regularisation method based on fully connected conditional random fields [8] did not yield improvements in our experiments.

3 Experiments and Results

3.1 Data

The experiments in this paper were performed on cardiac cine-MRI training data of the ACDC challenge[1]. The publicly available training dataset consists of 100 patient scans each including a short-axis cine-MRI acquired on 1.5T and 3T systems with resolutions ranging from 0.70×0.70 mm to 1.92×1.92 mm in-plane and 5 mm to 10 mm through-plane. Furthermore, segmentation masks for the myocardium (Myo), the left ventricle (LV) and the right ventricle (RV) are available for the end-diastolic (ED) and end-systolic (ES) phases of each

[1] https://www.creatis.insa-lyon.fr/Challenge/acdc (last accessed 26 July 2017).

Table 1. Segmentation accuracy obtained by optimising the modified 2D U-Net using different cost functions.

	Dice (LV)	ASSD (LV)	Dice (RV)	ASSD (RV)	Dice (Myo)	ASSD (Myo)
Crossentropy	**0.950 (0.029)**	**0.43 (0.41)**	0.891 (0.084)	1.06 (1.04)	0.888 (0.031)	0.52 (0.22)
W. Crossentropy	**0.950 (0.036)**	0.52 (0.75)	**0.893 (0.083)**	**1.04 (1.06)**	**0.899 (0.032)**	**0.51 (0.35)**
Dice Loss	0.944 (0.051)	0.56 (0.77)	0.843 (0.137)	2.13 (2.03)	0.891 (0.029)	0.55 (0.24)

patient. The dataset includes, in equal numbers, patients diagnosed with previous myocardial infarction, dilated cardiomyopathy, hypertrophic cardiomyopathy, abnormal right ventricles, as well as normal controls. We did not employ any external data for training or pre-training of the networks.

The dataset was divided into a training and validation set comprising 80 and 20 subjects, respectively, with a stratified split w.r.t. patient diagnosis. All images were pre-processed as described in Sect. 2.1.

3.2 Evaluation Measures

We evaluated the segmentation accuracy achieved with the different network architectures and optimisation techniques using three measures: the Dice coefficient, the Hausdorff distance and the average symmetric surface distance (ASSD). Furthermore, for the best performing experiment configuration, the correlations to commonly measured clinical variables were calculated.

3.3 Experiment 1: Comparison of Loss Functions

In the first experiment we focused on the modified 2D U-Net architecture for which we obtained good initial results, and compared the performance using the different cost functions introduced in Sect. 2.3. In Table 1 we report the Dice score and ASSD averaged over both cardiac phases. It can be seen that using cross entropy led to better results than optimising the Dice directly. Weighted and unweighted cross entropy performed similarly, with the weighted loss function leading to marginally better results. We conclude that for the task at hand, the class imbalance does not seem to be an issue. Nevertheless, for the comparison of the network architectures in the next section we continued using the unweighted cross entropy as a loss function due to the slightly better results.

3.4 Experiment 2: Comparison of Network Architectures

This experiment focuses on the comparison of the different 2D and 3D network architectures described in Sect. 2.2. The results are shown in Table 2. It can be seen that the 2D U-Net (both the original and modified version) outperformed FCN-8 and the (modified) 3D U-Net. While both versions of the 2D U-Net perform similarly, the modified version leads to slightly better results.

Clinical measures for the best performing method (the modified 2D U-Net) are shown in Table 3. A detailed description of the measures is provided by

Table 2. Segmentation accuracy measures for different network architectures. Each table entry depicts the mean (std) value accuracy measure obtained for a specific structure and cardiac phase.

	Left ventricle (ED)			Left ventricle (ES)		
	Dice	ASSD	HD	Dice	ASSD	HD
FCN-8	0.960 (0.018)	0.41 (0.49)	5.77 (3.05)	0.926 (0.061)	0.64 (0.80)	7.31 (3.39)
2D U-Net	0.965 (0.014)	**0.36 (0.38)**	**5.63 (2.79)**	**0.937 (0.051)**	**0.54 (0.64)**	**6.85 (3.52)**
2D U-Net (mod.)	**0.966 (0.017)**	0.37 (0.48)	5.71 (4.22)	0.935 (0.042)	0.67 (0.92)	8.23 (8.29)
3D U-Net (mod.)	0.939 (0.022)	0.63 (0.50)	8.69 (4.25)	0.905 (0.039)	0.70 (0.38)	9.13 (4.10)
	Right ventricle (ED)			Right ventricle (ES)		
	Dice	ASSD	HD	Dice	ASSD	HD
FCN-8	0.932 (0.025)	**0.57 (0.45)**	12.24 (5.51)	0.835 (0.100)	1.63 (1.07)	13.89 (4.24)
2D U-Net	**0.936 (0.028)**	0.65 (0.48)	12.43 (6.13)	0.838 (0.085)	1.72 (1.22)	14.52 (5.28)
2D U-Net (mod.)	0.934 (0.039)	0.66 (0.74)	**12.17 (6.02)**	**0.852 (0.095)**	**1.42 (1.19)**	**13.46 (6.24)**
3D U-Net (mod.)	0.888 (0.069)	1.17 (1.21)	14.91 (5.02)	0.781 (0.101)	2.26 (1.40)	16.24 (5.39)
	Myocardium (ED)			Myocardium (ES)		
	Dice	ASSD	HD	Dice	ASSD	HD
FCN-8	0.869 (0.029)	0.55 (0.23)	9.16 (6.74)	0.890 (0.027)	0.62 (0.24)	9.69 (5.28)
2D U-Net	0.885 (0.027)	0.52 (0.29)	9.01 (7.66)	0.904 (0.029)	**0.55 (0.28)**	10.06 (5.79)
2D U-Net (mod.)	**0.892 (0.027)**	**0.45 (0.22)**	**8.65 (6.02)**	**0.906 (0.034)**	0.56 (0.44)	**9.66 (6.21)**
3D U-Net (mod.)	0.802 (0.053)	0.91 (0.34)	11.87 (6.25)	0.839 (0.066)	0.90 (0.42)	10.95 (3.47)

Table 3. Clinical measurements: correlation, bias and limits of agreement (LoA) for the LV and RV ejection fraction (EF) and all structure volumes.

	Correlation			Bias [LoA]		
	EF	Vol (ED)	Vol (ES)	EF	Vol (ED)	Vol (ES)
LV	0.972	0.998	0.994	−0.45[−9.68; 8.78]	1.28[−7.27; 9.83]	2.55[−14.86; 19.96]
RV	0.868	0.961	0.965	6.25[−13.08; 25.58]	−0.45[−28.89; 27.99]	−8.08[−33.87; 17.71]
Myo	-	0.995	0.988	-	−5.24[−15.27; 4.79]	−0.71[−17.89; 16.47]

ACDC (See footnote 1). Figure 1 shows example segmentation results at three slice positions using the above method. Inference on a single volume took approximately 1.1 s for the 2D networks and 2.2 s for the 3D networks using a Nvidia Titan Xp GPU.

3.5 Discussion and Conclusion

In this work we evaluated the suitability of state-of-the-art neural network architectures for the task of fully automatic cardiac segmentation. We also investigated modified versions of those networks which yielded marginal improvements in performance. In particular, we found that using fewer feature maps in the upsampling path of the 2D U-Net yielded minor but consistent improvements. We speculate that for this problem the full complexity of the upsampling path is not necessary. Furthermore, the "bottlenecks" may force the downsampling layers to learn more semantically meaningful features. Lastly, having fewer

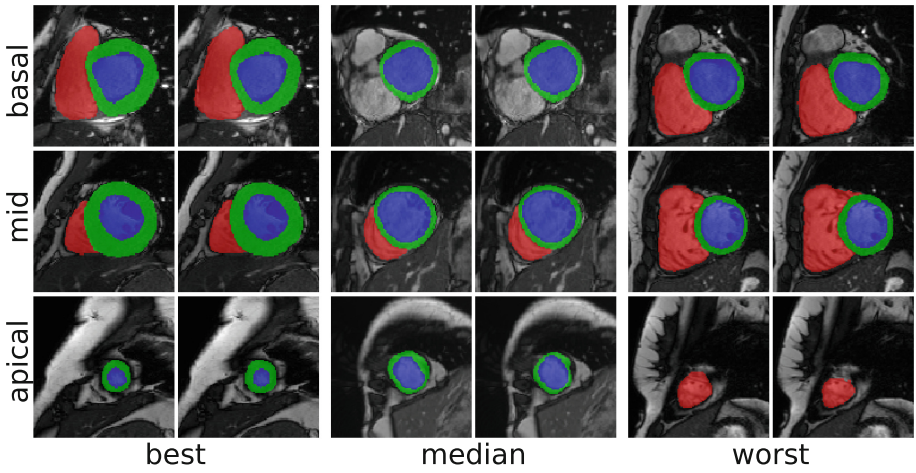


Fig. 1. Example segmentations at ED obtained using the 2D U-Net (mod.) for subjects with the highest, median, and lowest Dice coefficients on the Myocardium (left to right). Ground truth (left) and predicted segmentation (right) are shown for a basal, mid-ventricular and apical slice (top to bottom).

parameters may also make the problem easier to optimise. Further investigation into the significance of the upsampling path complexity will be necessary.

Overall we found that the exact architecture played a minor role in the accuracy of the system. However, the use of batch normalisation as well as the choice of the cost function had a big impact on the performance. Moreover, we found that resampling of the predictions to the original image resolution was a significant source of errors. This could be reduced by resampling the softmax output with linear interpolation, rather than the predicted masks.

One goal of this paper was to investigate if 3D context is helpful for the segmentation of short-axis MR images. Our experiments revealed that all 2D approaches consistently outperformed the (modified) 3D U-Net. There are at least three possible reasons for this: (1) when using 3D data, the amount of training images is drastically reduced which complicates training. (2) Since the through-plane resolution is low (and the cardiac structures typically appear in the top and bottom slices already), border effects from 3D convolutions may compromise the information available at intermediate representations. (3) GPU memory restrictions required a substantial downsampling of the data for training and prediction, potentially leading to a loss of information.

The segmentation scores reported in this work compare favourably to the related literature. However, it should be noted that a direct comparison is complicated by the fact that different datasets were used in the different works. For the LV cavity two recent deep learning methods [1,14] report Dice scores of around 0.94, while the modified 2D U-Net discussed here achieved a slightly higher value of 0.95. For automated segmentation of the RV cavity, [3,13] report

similar results to ours. Segmentation of the myocardium is a more challenging task than the LV and RV cavities, which is reflected by lower Dice scores of around 0.81 reported in recent literature [2,14]. We achieved substantially higher results using all 2D architectures. In particular, the modified 2D U-Net architecture produced a Dice score of 0.899 for this structure. While these results are encouraging, further analysis on common datasets is necessary. Specifically, we observed that the field of view in many images of the ACDC challenge dataset does not include the apex and basal region of the heart, which are particularly challenging to segment.

The code and pretrained models for all examined network architectures are publicly available at https://github.com/baumgach/acdc_segmenter.

References

1. Avendi, R.M.R., Kheradvar, A., Jafarkhani, H.: A combined deep-learning and deformable-model approach to fully automatic segmentation of the left ventricle in cardiac MRI. Med. Image Anal. **30**, 108–119 (2016)
2. Bai, W., Shi, W., Ledig, C., Rueckert, D.: Multi-atlas segmentation with augmented features for cardiac MR images. Med. Image Anal. **19**(1), 98–109 (2015)
3. Bai, W., Shi, W., O'Regan, D.P., Tong, T., Wang, H., Jamil-Copley, S., Peters, N.S., Rueckert, D.: A probabilistic patch-based label fusion model for multi-atlas segmentation with registration refinement: application to cardiac MR images. IEEE Trans. Med. Imaging **32**(7), 1302–15 (2013)
4. Çiçek, Ö., Abdulkadir, A., Lienkamp, S.S., Brox, T., Ronneberger, O.: 3D U-Net: learning dense volumetric segmentation from sparse annotation. In: Ourselin, S., Joskowicz, L., Sabuncu, M.R., Unal, G., Wells, W. (eds.) MICCAI 2016. LNCS, vol. 9901, pp. 424–432. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-46723-8_49
5. He, K., Zhang, X., Ren, S., Sun, J.: Delving deep into rectifiers: surpassing human-level performance on ImageNet classification. In: ICCV, pp. 1026–34 (2015)
6. Ioffe, S., Szegedy, C.: Batch normalization: accelerating deep network training by reducing internal covariate shift. In: ICML, pp. 448–456 (2015)
7. Kingma, D.P., Ba, J.L.: ADAM: a method for stochastic optimization. In: ICLR (2015)
8. Krähenbühl, P., Koltun, V.: Efficient inference in fully connected CRFs with Gaussian edge potentials. In: NIPS, pp. 109–117 (2011)
9. Litjens, G., Kooi, T., Bejnordi, B.E., Setio, A.A.A., Ciompi, F., Ghafoorian, M., van der Laak, J.A.W.M., van Ginneken, B., Sánchez, C.I.: A Survey on Deep Learning in Medical Image Analysis. arXiv:1702.05747 (2017)
10. Long, J., Shelhamer, E., Darrell, T.: Fully convolutional networks for semantic segmentation. In: CVPR, pp. 343–3440 (2015)
11. Milletari, F., Navab, N., Ahmadi, S.A.: V-Net: fully convolutional neural networks for volumetric medical image segmentation. In: 3D Vision, pp. 565–571 (2016)
12. Nichols, M., Townsend, N., Scarborough, P., Rayner, M.: Cardiovascular disease in Europe 2014: epidemiological update. Eur. Heart J. **35**, 2950–2959 (2014)
13. Oktay, O., Bai, W., Guerrero, R., Rajchl, M., de Marvao, A., O'Regan, D.P., Cook, S.A., Heinrich, M.P., Glocker, B., Rueckert, D.: Stratified decision forests for accurate anatomical landmark localization in cardiac images. IEEE Trans. Med. Imaging **36**(1), 332–342 (2017)

14. Oktay, O., Ferrante, E., Kamnitsas, K., Heinrich, M., Bai, W., Caballero, J., Guerrero, R., Cook, S., de Marvao, A., Dawes, T., O'Regan, D., Kainz, B., Glocker, B., Rueckert, D.: Anatomically Constrained Neural Networks (ACNN): Application to Cardiac Image Enhancement and Segmentation. arXiv:1705.08302 (2017)
15. Ronneberger, O., Fischer, P., Brox, T.: U-Net: convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) MICCAI 2015. LNCS, vol. 9351, pp. 234–241. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-24574-4_28

Automatic Cardiac Disease Assessment on cine-MRI via Time-Series Segmentation and Domain Specific Features

Fabian Isensee[1], Paul F. Jaeger[1(✉)], Peter M. Full[2,3], Ivo Wolf[3], Sandy Engelhardt[2,3], and Klaus H. Maier-Hein[1]

[1] Medical Image Computing, German Cancer Research Center (DKFZ), Heidelberg, Germany
p.jaeger@dkfz.de
[2] Division of Computer-assisted Medical Interventions, German Cancer Research Center (DKFZ), Heidelberg, Germany
[3] Department of Computer Science, Mannheim University of Applied Science, Mannheim, Germany

Abstract. Cardiac magnetic resonance imaging improves on diagnosis of cardiovascular diseases by providing images at high spatiotemporal resolution. Manual evaluation of these time-series, however, is expensive and prone to biased and non-reproducible outcomes. In this paper, we present a method that addresses named limitations by integrating segmentation *and* disease classification into a fully automatic processing pipeline. We use an ensemble of UNet inspired architectures for segmentation of cardiac structures such as the left and right ventricular cavity (LVC, RVC) and the left ventricular myocardium (LVM) on each time instance of the cardiac cycle. For the classification task, information is extracted from the segmented time-series in form of comprehensive features handcrafted to reflect diagnostic clinical procedures. Based on these features we train an ensemble of heavily regularized multilayer perceptrons (MLP) and a random forest classifier to predict the pathologic target class. We evaluated our method on the ACDC dataset (4 pathology groups, 1 healthy group) and achieve dice scores of 0.945 (LVC), 0.908 (RVC) and 0.905 (LVM) in a cross-validation over the training set (100 cases) and 0.950 (LVC), 0.923 (RVC) and 0.911 (LVM) on the test set (50 cases). We report a classification accuracy of 94% on a training set cross-validation and 92% on the test set. Our results underpin the potential of machine learning methods for accurate, fast and reproducible segmentation and computer-assisted diagnosis (CAD).

Keywords: Automated cardiac diagnosis challenge
Cardiac magnetic resonance imaging · Disease prediction
Deep learning · CNN

F. Isensee and P. F. Jaeger—Contributed equally.

M. Pop et al. (Eds.): STACOM 2017, LNCS 10663, pp. 120–129, 2018.
https://doi.org/10.1007/978-3-319-75541-0_13

1 Introduction

Cardiac remodeling plays an inherent role in the progressive course of heart failure. The process results in poor prognosis for the patient due to diminished contractile systolic function, reduced stroke volume or malignant arrhythmia. Clinical manifestations are changes in size, mass, geometry, regional wall motion and function of the heart [1], which can be assessed timely and monitored non-invasively by cardiac magnetic resonance imaging (CMRI). In today's clinical routine, the huge benefits of comprehensive quantitative measurements are still not exploited due to the associated labour time, subjective biases and lack of reproducibility. Accurate automatic approaches for simultaneous multi-structure segmentation and CAD are thus desirable assets for a large spectrum of cardiac diseases.

Convolutional neural networks (CNN) have recently shown outstanding performance in medical image segmentation [2], where they typically take the form of UNet-like architectures [3]. So far, a limited number of fully-automatic cardiac CNN segmentation methods have been proposed [4–6], however, they did not consider to segment all volumes of the cardiac cycle. Highly accurate multi-structural segmentations on the entire cardiac cycle are crucial for automatic pathology assessment, especially when considering geometrical and dynamical changes of anatomical structures.

Computer-assisted diagnosis (CAD) approaches originate from the field of lesion detection and classification [7], which primarily focuses on texture information to discriminate healthy from pathological tissue. Medrano-Gracia et al. investigated global shape variations of the left ventricle in a large cohort of an asymptomatic population [8]. They found the major principal modes of shape variation to be associated with known clinical indices of adverse remodelling, including heart size, sphericity and concentricity. Later, Zhang et al. used a supervised method to extract the most discriminatory global shape changes associated with remodeling after myocardial infarction [9]. The resulting shape model was able to discriminate patients from asymptomatic subjects with 95% accuracy. However, to the best of our knowledge, a comprehensive CAD system for different cardiac remodelling pathologies and myocardial infarct patients has not been proposed before.

In this paper we present an approach for automatic classification of cardiac diseases associated with pathological remodelling. Based on multi-structure segmentation for each time step of the CMRI, we extract domain-specific features, which are motivated by a cardiologist's workflow, to then train an ensemble of classifiers for disease prediction (see Fig. 1). We evaluated our methods for segmentation *and* classification on the MICCAI ACDC data set [10].

2 Methods

2.1 Cardiac cine-MRI Dataset

The ACDC dataset [10] comprises short-axis cine-MRI of 150 patients acquired at the University Hospital of Dijon using two MR scanners of different magnetic

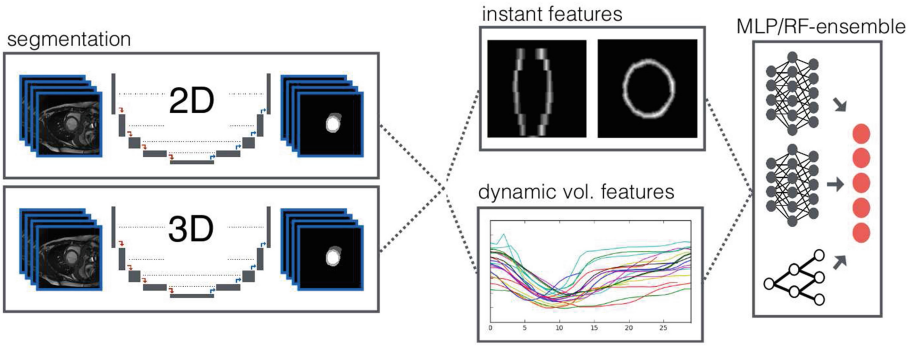


Fig. 1. Overview of the proposed pipeline: Segmentation predictions from a 2D and a 3D model are averaged and used to extract instant and dynamic volume features, which are fed into an ensemble of classifiers for disease prediction.

strengths (1.5 T and 3.0 T). Each time-series is composed of 28 to 40 3D volumes, which partially or completely cover the cardiac cycle. As typical for CMRI, the data is characterized by a high in-plane resolution ranging from 0.49 to 3.69 mm^2 and a low resolution in the direction of the long axis of the heart (5−10mm slice thickness). Note that some data exhibit severe slice misalignments, which originate from different breath hold positions between slice stack acquisitions. Structures of interest, namely LVC, LVM and RVC were segmented manually by clinical experts on end diastolic (ED) and end sstolic (ES) phase instants. Four pathological groups and one group of healthy patients are evenly distributed in the dataset: patients with previous myocardial infarction (MINF), dilated cardiomyopathy (DCM), hypertrophic cardiomyopathy (HCM), abnormal right ventricle (ARV) and normal (healthy) subjects (NOR). Additional information for all patients is provided in form of height and weight. The dataset is split into 100 training and 50 test patients. Segmentation and classification ground truth is provided only for the 100 training cases. All reported test set results were obtained by submitting our predictions to the online evaluation platform.

2.2 Segmentation

Data Preprocessing. For the segmentation part of the ACDC challenge we resampled all volumes to $1.25 \times 1.25 \times 10$ mm per voxel (for 3D UNet and feature extraction) and $1.25 \times 1.25 \times Z_{orig}$ mm per voxel (for 2D UNet) to account for varying spatial resolutions. The grey level information of every image was normalized to zero-mean and unit-variance.

Network Architecture. We tackle the segmentation using an ensemble of modified 2D and 3D UNets [3,11]. The 3D segmentation model consists of a context aggregating pathway followed by a localization pathway. Both are interconnected at various scales to allow for recombination of abstract context features

with the corresponding local information. We carefully adapted the architecture to cope with specific challenges of CMRI (see Fig. 2): Due to low z-resolution of the input, pooling and upscaling operations are carried out only in the x-y-plane. Context in the z-dimension is solely aggregated through the 3D convolutions. Each feature extraction block (shown in gray) consists of two padded $3 \times 3 \times 3$ convolutions, followed by batch normalization and a leaky ReLU nonlinearity. Due to the shallow nature of the network (18 layers) no residual connections are utilized. The initial number of 26 feature maps is doubled (halved) with each of the 4 pooling (upscaling) operations, resulting in a maximum of 416 feature maps at the bottom of the U-shape. Deep supervision (as in [12]) is implemented by generating low resolution segmentation outputs via $1 \times 1 \times 1$ convolutions before each of the last two upscaling operations, which are upscaled and aggregated for the final segmentation.

The 3D model was trained for 300 epochs in a 5-fold cross validation using the ADAM solver and a pixel-wise categorical cross-entropy loss. The initial learning rate of $5 \cdot 10^{-4}$ was decayed by 0.98 per epoch, where an epoch was defined as 100 batches, each comprising four training examples. Training examples were generated as random crops of size $224 \times 224 \times 10$ voxels taken from a randomly chosen training patient and phase instance (ED/ES).

The 2D model's architecture is equivalent to the 3D approach except 2D convolutions. Due to the lower memory requirements, we increased the number of initial feature maps to 48. The network is trained with a batch size of 10 and input patches of size 352×352 pixels using a multiclass dice loss:

$$\mathcal{L}_{\mathrm{dc}} = -\frac{2}{|K|} \sum_{k \in K} \frac{\sum_i u_i^k v_i^k}{\sum_i u_i^k + \sum_i v_i^k}, \tag{1}$$

where u is the softmax output of the network and v denotes a one hot encoding of the ground truth segmentation map. Both u and v are of size $i \times k$ with i being the number of pixels in the training patch and $k \in K$ being the classes.

To accomplish the training of a well generalizing model on limited data, we used a broad range of data augmentation techniques, such as mirroring along the x and y axes, random rotations, gamma-correction and elastic deformations. Due to the low z-resolution all data augmentation was performed only in the x-y-plane. To account for the presence of slice misalignments, we artificially increased the number of misaligned slices by motion augmentation for the training of the 3D model: All slices within the training batch were perturbed with a probability of 10% and a random offset drawn from $\mathcal{N}(0, 20)$.

To obtain the final segmentations, softmax outputs of both networks were resampled to the original voxel resolution of the input image and then averaged.

2.3 Cardiac Disease Classification

Feature Extraction. We extract two sets of features from the previously segmented structures to perform disease classification. All features were designed to quantify the traditional assessment procedures of expert cardiologists by describing static and dynamic properties of the structures of interest (see Table 1).

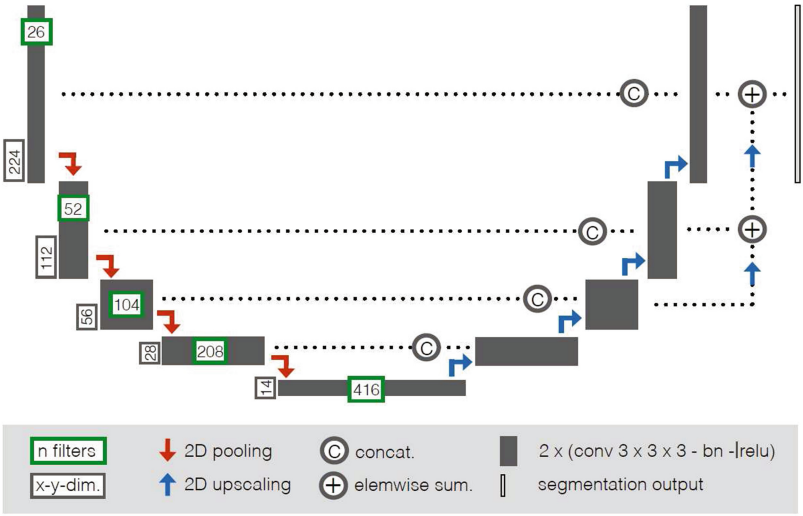


Fig. 2. Architecture of the 3D segmentation network. The 2D network is equivalent, but uses 2D convolutions, patch size 352 × 352 and 48 initial features. (Color figure online)

Instant features. Extracted from the two labeled ED and ES time instants as provided by the ACDC dataset, these features cover local and global shape information (circumference, circularity, LVM thickness, etc.), local variations (size of RVC at the apex, LVM thickness between RVC and LVC), simple texture descriptors (mass) as well as additional meta information (body mass index, weight, height). Notably, all thicknesses, circumferences and circularities are computed on the individual x-y-planes and aggregated over the z-dimension. The body surface is estimated from weight and height using the Mosteller formula.

Dynamic volume features. We deployed the trained segmentation model to predict the anatomical structures in all time steps of the CMRI. This allows for exploitation of volume dynamics throughout the entire cardiac cycle independent of the predefined ED/ES. These volume dynamics are quantified in form of first order statistics (median, standard deviation, kurtosis, skewness) complemented by characteristics of the cardiac cycle's minimum and maximum volumes: We found the time instants of these extrema to not match the predefined ED/ES instants in the majority of patients. This finding is accounted for by computing volume, volume ratios and ejection fractions based on the determined actual minimum (v_{min}) and maximum (v_{max}) volume of the cardiac cycle. Finally, the synchrony of contraction between LVC and RVC is measured in form of the time step differences between their corresponding v_{min} and v_{max}.

Classification. The features described in Sect. 2.3 were used to train an ensemble of 50 multilayer perceptrons (MLP) and a random forest for pathology

Table 1. The two sets of features extracted for disease classification and the corresponding cardiovascular structure (RVC, LVM, LVC). All instant features (except for additional patient information) are extracted on both ED and ES.

Instant features	RVC	LVM	LVC
Max thickness*		x	
Min thickness*		x	
Std thickness*		x	
Mean thickness*		x	
Std thickness of LVM between LVC and RVC*			
Mean thickness of LVM between LVC and RVC*			
Mean circularity*	x	x	
Max circumference*	x	x	
Mean circumference*	x	x	
RVC size at most apical LVM slice*			
RVC to LVC size ratio at most apical LVM slice*			
Volume per m^2 body surface	x	x	x
Mass		x	
Patient weight			
Patient height			
Patient body mass index			
Dynamic volume features	**RVC**	**LVM**	**LVC**
v_{max}	x	x	x
v_{min}	x	x**	x
Dynamic ejection fraction	x	x**	x
Volume median	x	x	x
Volume kurtosis	x	x	x
Volume skewness	x	x	x
Volume standard deviation	x	x	x
Volume ratio $v_{min,LVC}/v_{min,RVC}$			
Volume ratio $v_{min,LVM}/v_{min,LVC}$			
Volume ratio $v_{min,RVC}/v_{min,LVM}$			
Time step difference $t(v_{min,LVC})$-$t(v_{min,RVC})$			
Time step difference $t(v_{max,LVC})$-$t(v_{max,RVC})$			

*This feature was calculated in the x-y-plane and aggregated over slices in z.
** $v_{min,LVM}$ was determined at $t(v_{min,LVC})$.

classification. The MLP's architecture consists of four hidden layers, each containing 32 units, followed by batch normalization, leaky ReLU nonlinearity and a Gaussian noise layer ($\sigma = 0.1$). Each MLP was trained on a random subset of

75% of the training data, while the remaining 25% were used for epoch selection. Further regularization was provoked by only presenting a random subset of 2/3 of the features to each MLP. We trained all MLPs for 400 epochs (with a patience of 40 epochs) using the ADAM solver with an initial learning rate of $5 \cdot 10^{-4}$, decayed by 0.97 per epoch. An epoch was defined as a set of 50 batches containing 20 patients each. Additionally, we trained a random forest with 1000 trees. During testing, the softmax outputs of all MLPs were averaged to obtain an overall MLP score, which was recombined subsequently with the random forest output to obtain the final ensemble prediction.

3 Results

Segmentation. With regard to the expert segmentations on the original ED and ES phase instants, individual dice scores of 0.945 for the LVC, 0.905 for the LVM and 0.908 for the RVC were achieved in 5-fold cross-validation (see Table 2 for detailed results including Hausdorff distances). When comparing performances of the 2D and 3D approach, the 3D model was largely outperformed by the 2D model (see Table 3). A marginal increase in RVC dice was observed when ensembling the models. Note that cardiac phase instances other than ED and ES were not considered in the scores due to unavailable ground truth labels. Qualitatively, the 4D segmentation yielded convincing results, which were smooth and robust in time for all substructures. CMRI with slice misalignments were segmented successfully by the model when using motion augmentation (3D network only). Based on the cross-validation, we observed only little overfitting, most of which occurred for the RVC region. The main mode of failure was the basal part of the RVC region, where the model struggled to distinguish between the right atrium and the right outflow tract or the RVC. This occurred mostly in ES images, resulting in a lower average dice score compared to ED. Other failures occurred in the LVC region, where papillary muscles were anatomical correctly classified as LVM but should have been classified as LVC to meet the convention of the challenge. This was especially observed in HCM patients. Simultaneous segmentation of multiple structures in a 3D volume of size $320 \times 320 \times 10$ voxels took less than one second for the 3D model and 1–2 seconds for the 2D model on a Pascal Titan X GPU (Fig. 3).

Classification. We trained the classification ensemble (see Sect. 2.3) on the ACDC training data using the features described in Sect. 2.3. In a five fold cross-validation, a classification accuracy of 94% was achieved. The individual performance of the MLP ensemble and random forest were 93% and 92%, respectively. The test set accuracy was 92%. Confusion matrices are provided in Fig. 4, indicating equal performance among classes in the cross-validation, and difficulties in distinguishing DCM from MINF patients on the test set. Feature computation took 15 s for instant features and less than one second for the dynamic volume features.

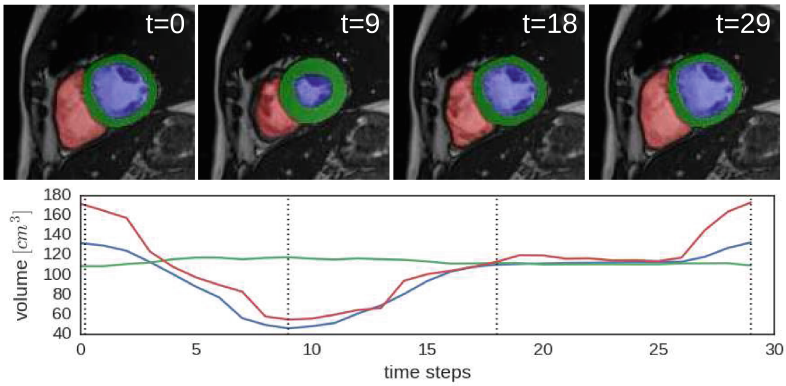


Fig. 3. Time-series segmentation for RVC (red), LVM (green), LVC (blue) and their corresponding volume dynamics. The example shows the central slice in z direction of a healthy patient (NOR). (Color figure online)

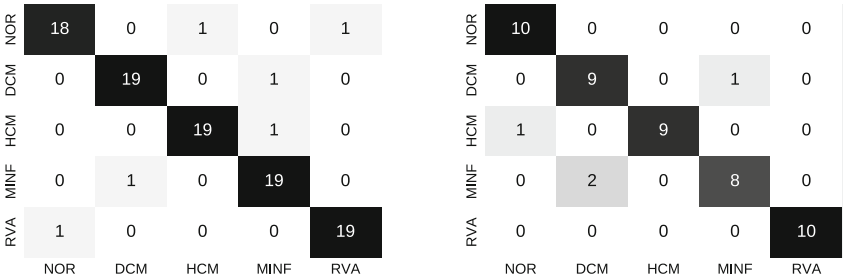


Fig. 4. Confusion matrices of the ensemble predictions from cross-validation on the training set (left) and on the test set (right). Rows correspond to the predicted class and columns to the target class, respectively.

4 Discussion

In this paper we presented a fully automatic processing pipeline for pathology classification on cardiac cine-MRI. First, we developed an accurate multi-structure segmentation method trained solely on ED and ES phase instances, but capable of processing the entire cardiac cycle. Our approach revolves around the use of both a 2D and 3D model, leveraging their respective advantages through ensembling. The resulting pipeline is robust against slice misalignments, different CMRI protocols as well as various pathologies. We achieve dice scores of 0.950 (LVC), 0.923 (RVC) and 0.911 (LVM) on the ACDC test set, which earned us the first place in the segmentation part of the challenge. Based on the segmentations generated by our model, geometrical features are extracted and utilized by an ensemble of classifiers to predict the diagnosis, yielding promising outcomes. We ranked second in the classification part of the challenge with an accuracy of 92%. Our fully automatic processing pipeline constitutes an attractive software

Table 2. Dice scores and Hausdorff distances of the segmentation model for pathological subgroups (results of 5-fold cross validation).

	Instance	Dice			Hausdorff (mm)		
		RVC	LVM	LVC	RVC	LVM	LVC
DCM	ED	0.942	0.906	0.968	20.87	8.21	7.117
	ES	0.872	0.913	0.916	17.944	8.161	5.886
	Total	**0.907**	**0.910**	**0.942**	**19.830**	**8.153**	**6.544**
HCM	ED	0.938	0.901	0.968	12.721	8.709	7.256
	ES	0.878	0.907	0.935	18.326	11.355	14.514
	Total	**0.908**	**0.904**	**0.952**	**15.321**	**10.105**	**11.022**
MINF	ED	0.937	0.896	0.961	13.385	9.63	6.882
	ES	0.889	0.907	0.907	18.639	11.74	9.599
	Total	**0.913**	**0.901**	**0.934**	**16.107**	**10.730**	**8.116**
NOR	ED	0.939	0.887	0.971	9.765	7.231	4.626
	ES	0.884	0.901	0.941	11.407	9.164	7.665
	Total	**0.911**	**0.898**	**0.956**	**10.615**	**8.397**	**6.330**
RV	ED	0.948	0.909	0.964	14.728	10.716	9.392
	ES	0.852	0.911	0.922	15.126	11.769	11.224
	Total	**0.900**	**0.91**	**0.943**	**15.133**	**11.605**	**10.620**

Table 3. Dice scores and Hausdorff distances of the segmentation. Results from cross-validation (CV) on the training set are shown for the 2D model, the 3D model and the corresponding ensemble. On the test set, only ensemble results are shown.

		Dice			Hausdorff (mm)		
		RVC	LVM	LVC	RVC	LVM	LVC
CV	2D model	0.902	0.905	0.945	14.294	8.899	7.055
	3D model	0.879	0.872	0.928	16.289	10.438	9.778
	ensemble	**0.908**	**0.905**	**0.945**	**15.291**	**9.668**	**8.416**
Test	**ensemble**	**0.923**	**0.911**	**0.950**	**11.134**	**8.696**	**7.145**

for clinical decision support due to the visualization of intermediate segmentation maps, the comprehensive quantification of cardiologic assessment and the rapid processing speed of less than 40 s. Possible future improvements of the model concern data augmentation and the architecture of the segmentation network as well as a regularization objective as used in [5]. Training the pipeline end-to-end in a multitask architecture could yield further improvement.

Acknowledgements. The author Sandy Engelhardt was funded by the German Research Foundation (DFG) as part of project B01, SFB/TRR 125 Cognition-Guided Surgery.

References

1. Cohn, J.N., Ferrari, R., Sharpe, N.: Cardiac remodeling-concepts and clinical implications: a consensus paper from an international forum on cardiac remodeling. JACC **35**, 569–582 (2000)
2. Litjens, G., Kooi, T., Bejnordi, B.E., Setio, A.A.A., Ciompi, F., Ghafoorian, M., van der Laak, J.A., van Ginneken, B., Sánchez, C.I.: A Survey on Deep Learning in Medical Image Analysis. arXiv preprint arXiv:1702.05747 (2017)
3. Ronneberger, O., Fischer, P., Brox, T.: U-Net: convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) MICCAI 2015, Part III. LNCS, vol. 9351, pp. 234–241. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-24574-4_28
4. Zotti, C., Luo, Z., Lalande, A., Humbert, O., Jodoin, P.-M.: Novel Deep Convolution Neural Network Applied to MRI Cardiac Segmentation. arXiv preprint arXiv:1705.08943 (2017)
5. Oktay, O., Ferrante, E., Kamnitsas, K., Heinrich, M., Bai, M., Caballero, M., Guerrero, R., Cook, S., de Marvao, A., O'Regan, D., et al.: Anatomically Constrained Neural Networks (ACNN): Application to Cardiac Image Enhancement and Segmentation. arXiv preprint arXiv:1705.08302 (2017)
6. Tran, P.V.: A Fully Convolutional Neural Network for Cardiac Segmentation in Short-axis MRI. arXiv preprint arXiv:1604.00494 (2016)
7. Doi, K.: Computer-aided diagnosis in medical imaging: historical review, current status and future potential. CMIG **31**(4), 198–211 (2008)
8. Medrano-Gracia, P., Cowan, B.R., Ambale-Venkatesh, B., Bluemke, D.A., Eng, J., Finn, J.P., Fonseca, C.G., Lima, J.A., Suinesiaputra, A., Young, A.A.: Left ventricular shape variation in asymptomatic populations: the multi-ethnic study of atherosclerosis. JCMR **16**(1), 56 (2014)
9. Zhang, X., Ambale-Venkatesh, B., Bluemke, D.A., Cowan, B.R., Finn, J.P., Kadish, A.H., Lee, D.C., Lima, J.A.C., Hundley, W.G., Suinesiaputra, A., Young, A.A., Medrano-Gracia, P.: Information maximizing component analysis of left ventricular remodeling due to myocardial infarction. JTM **13**(1), 343 (2015)
10. Automated Cardiac Diagnosis Challenge. https://www.creatis.insa-lyon.fr/Challenge/acdc. Accessed 23 Jun 2017
11. Çiçek, Ö., Abdulkadir, A., Lienkamp, S.S., Brox, T., Ronneberger, O.: 3D U-Net: learning dense volumetric segmentation from sparse annotation. In: Ourselin, S., Joskowicz, L., Sabuncu, M.R., Unal, G., Wells, W. (eds.) MICCAI 2016, Part II. LNCS, vol. 9901, pp. 424–432. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-46723-8_49
12. Kayalibay, B., Jensen, G., van der Smagt, P.: CNN-Based Segmentation of Medical Imaging Data. arXiv preprint arXiv:1701.03056 (2017)

2D-3D Fully Convolutional Neural Networks for Cardiac MR Segmentation

Jay Patravali(✉), Shubham Jain, and Sasank Chilamkurthy

Qure.ai, Mumbai, India
{jay.patravali,shubham.jain,sasank.chilamkurthy}@qure.ai
http://www.qure.ai

Abstract. In this paper, we develop a 2D and 3D segmentation pipelines for fully automated cardiac MR image segmentation using Deep Convolutional Neural Networks (CNN). Our models are trained end-to-end from scratch using the ACD Challenge 2017 dataset comprising of 100 studies, each containing Cardiac MR images in End Diastole and End Systole phase. We show that both our segmentation models achieve near state-of-the-art performance scores in terms of distance metrics and have convincing accuracy in terms of clinical parameters. A comparative analysis is provided by introducing a novel dice loss function and its combination with cross entropy loss. By exploring different network structures and comprehensive experiments, we discuss several key insights to obtain optimal model performance, which also is central to the theme of this challenge.

Keywords: Deep learning · Medical image analysis
Computer vision · MR segmentation

1 Introduction

MR imaging is an effective non-invasive procedure for diagnosis and treatment of known or suspected Cardiac diseases. Cardiac MR images can produce highly detailed pictures of different structures within the heart. Delineation of these structures can provide relevant diagnostic information and evaluate the overall functioning of the heart. Segmentation of left ventricle, right ventricle and the myocardium can be used to calculate relevant diagnostics parameters such as ejection fraction and myocardial mass. Due to massive volumes of cardiac image data, relying on manual delineations can be a time-consuming process, often prone to error and rater variability. Hence there is a critical need for accurate, reproducible and fully-automated methods for cardiac segmentation.

In recent works, Deep Learning and Convolutional Neural Networks (CNNs) have shown tremendous progress in fully-automated segmentation tasks. The growing success of CNNs in solving computer vision problems such as image recognition and classification [1,2] can be attributed to its ability in learning a hierarchical representation of the input data, without relying on hand-crafted

M. Pop et al. (Eds.): STACOM 2017, LNCS 10663, pp. 130–139, 2018.
https://doi.org/10.1007/978-3-319-75541-0_14

features. Deep learning techniques for segmentation have defined the state-of-the-art using Fully Convolutional Networks (FCN) [3]. The idea behind FCN is to use a contracting path to extract features at different spatial scales followed by an expanding path to upsample and increase the spatial resolution of learned features.

For segmentation in medical images, U-Net [6] is a well established 2D CNN architecture that builds upon the FCN. By adding skip connections between the contracting and expanding paths, the U-Net model showed reasonable segmentation accuracy with very few training samples. Cardiac segmentation based on original FCN have been proposed [4] with modifications to make it faster and memory efficient [5]. As raw 3D MR volumes are fed slice-by-slice as inputs to these 2D CNN models, they fail to capture the spatial contextual information required to segment the whole heart. To that end, U-Net3D [7] extends the 2D U-Net model by replacing its 2D convolutional operations with its 3D counterparts. In similar way, V-Net [8] employs a 3D CNN model with a novel dice loss function showing convincing results in medical image segmentation. To our best knowledge, there are very few methods that have applied 3D CNNs for Cardiac Segmentation and have obtained satisfactory performance.

In our work, we develop a fully-automated 2D and 3D CNN models designed to segment the Left Ventricle, Right Ventricle and Myocardium. This segmentation task is part of the Automatic Cardiac Detection Challenge 2017 [9]. The 2D segmentation model is trained slice-by-slice, whereas we compute volumetric segmentation for the 3D model. Our models are easy to implement, have modular architecture, and relatively short training and testing times. We introduce a new dice loss function, and compare its performance with traditional cross entropy loss and combined cross entropy-dice loss. Through our experiments we also compare and analyze the performance of our 2D and 3D models, both which achieve near state-of-the-art accuracy scores in terms of geometric metrics and clinical validity.

2 Method

2.1 Network Architecture

We develop a 2D segmentation model architecture that is adapted from U-Net [6] as illustrated in Fig. 1. On left side is the "contracting" stage and on the right side is "expanding" stage. At the bottom is a base layer. We provide an option to feed varying number (N) of image slices that can be passed as input channels to the model. Here, N can be 1 for single image slice or more. Every step on the contracting path consists of a series of a 3x3 convolutions (conv 3x3), batch normalization (bn) [10], rectified linear unit (ReLU) and conv 3x3 in a sequence that forms a *conv_bn_relu* block. Two of *conv_bn_relu* blocks in succession forms a *Conv* block that doubles the number of feature channels. The contracting path downsamples the image with a 2x2 maxpool operation of stride 2. Similar to contracting stage, every step in expanding stage has a sequence of conv 3x3, bn, ReLU and conv 3x3 in a series of two consecutive

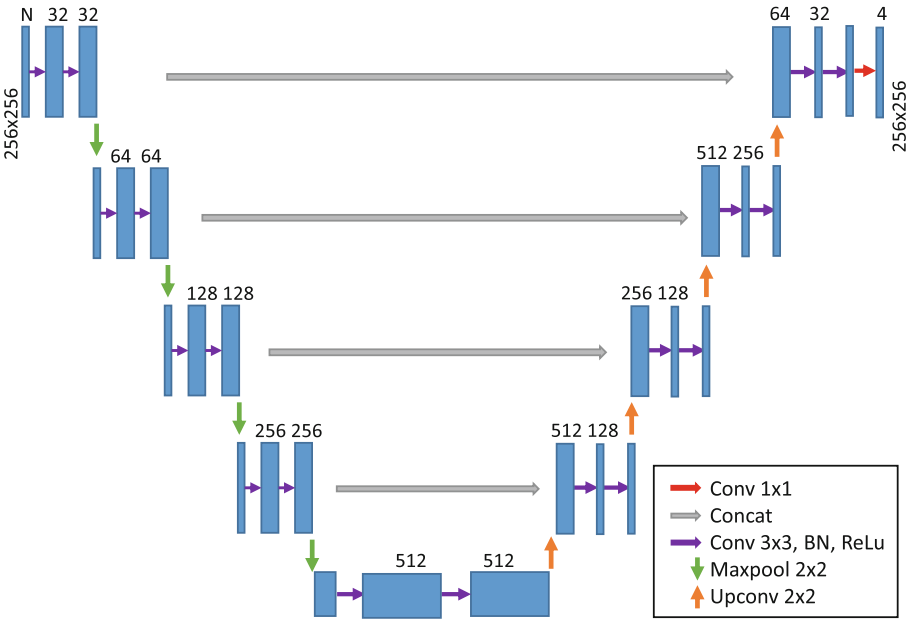


Fig. 1. 2D model architecture.

blocks. The images are upsampled with a 2x2 up-convolution (upconv 2x2) with stride 2. For upsampling of images, a sequence of 2x2 up-convolution (upconv 2x2) with stride 2, concatenations and a *Conv* block forms a *deconv* block. A final 1x1 convolution layer maps the 32 feature channels to 4 classes.

Our 3D model is an extension to our 2D model with few modifications, and finds similarities with U-Net 3d [7]. First, we replace all 2D operations with its 3D counterparts. Second, due to memory constraints and less number of training examples we limit the maximum number of feature maps to 256. Overall since the number of slices are less across the dataset (9 avg.), we apply 3D 1x2x2 maxpooling operation only in X and Y leaving out the Z dimension. This allows our 3D model to accept input volumes of varying slices at training or inference stage. Similar to our 2D model, every step in contracting and expanding stage consists of two repeating blocks, where each block is a sequence of conv 3x3x3, bn, ReLU, conv 3x3x3. Due to symmetric nature of the model, we can simultaneously add, remove or modify blocks across both paths. Additionally, the size of input data and final outputs images remains the same for both our 2D and 3D models. Thus, we are able to maintain modularity for faster experimentation and modification at the time of training and design. To prevent overfitting, we add dropout layers [11] with probability values of 0.5 in the last and 0.3 in the second last layer of the contracting stage, in both 2D and 3D models (Fig. 2).

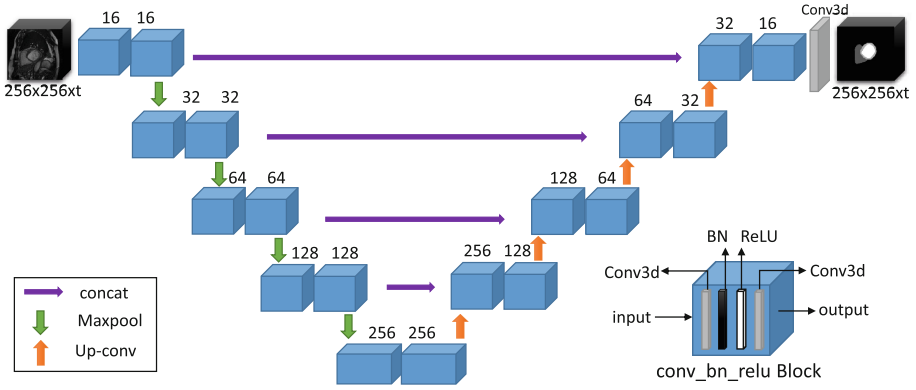


Fig. 2. 3D model architecture.

2.2 Dataset, Preprocessing and Augmentation

Our models are trained end-to-end from scratch using MICCAI's ACD Challenge 2017 dataset. It contains 150 exams of fully-annotated cardiac MRI's. Out of these 100 are used for training phase and 50 for testing phase of this challenge. These exams are obtained from multiple patients, each consisting of scans from End Diastole and End Systole phase taken in short axis orientation.

Since the data acquisition can bring inconsistencies in dataset, its necessary to carry preprocessing steps to ensure that the model receives uniform inputs. To remove noise and enhance contrast, we use contrast limited adaptive histogram localization (CLAHE) [12]. Then, we normalize the intensity values of all images between the range of 1–99 percentiles. Finally, we clip the image pixel values between 0 and 1.

To ensure both 2D and 3D model perceive heart features in similar proportion, we do a resampling operation across all input volumes to a common voxel spacing of $1.5 \times 1.5 \times 10$ mm. For 2D segmentation we resize and crop the images to fixed size of 256×256. Due to maxpool operation applied only on height and width in our 3D segmentation model, when testing the 3D model, we can feed it image volumes of varying number of slices. At training, the 3D model is fed with raw input volumes that are resized and cropped to $256 \times 256 \times 12$.

We apply a light data augmentation techniques on-the-fly to efficiently feed the input data volumes into our model. On a random basis, the data is rotated between -15 to $+15°$, and scaled between 0.9–1.1 range. This ensures slight robustness and variability in training the network.

2.3 Training

We train the 2D segmentation model by feeding it raw input images slice-by-slice. Whereas the 3D segmentation is trained by feeding it with entire 3D input volumes. Both the 2D and 3D models are trained with different optimization functions described more in detail in the following subsection.

During the training process, weights are updated using stochastic gradient descent with a momentum of 0.99. The initial learning rate is decayed by a factor of 10 every 30 epochs. For the training phase of this challenge, we do a 5-fold cross validation, which leaves 80 patients for training and 20 for validation. Training is completed after 300 epochs. Best models are checkpointed and stored for testing. For 2D segmentation we use a batch size of 8, whereas for 3D segmentation we use a batch size of 4.

Our models are implemented entirely using the PyTorch [13] framework, due to its flexibility in experimentation. We run all experiments on a standard workstation equipped with 64 GB of memory, Intel(R) Core(TM) i7-6700K CPU clocking at 4.00 GHz, with a 12 GB NVidia Titan X Pascal GPU.

2.4 Optimization Function

In this section we describe three optimization functions that are used for training our 2D and 3D segmentation models. We use these functions to compare the performance of 2D and 3D models. At training, we apply a pixel-wise softmax activation in the final layer of the model to get the predicted probabilities $p(x, i)$ for each class i at each pixel x. The targets at location x is denoted by $t(x)$.

Cross Entropy Loss. In segmentation tasks, the standard practice is to apply cross entropy loss function to measure the pixel-wise probability error between the predicted output and target and sum the errors across all the pixels. In addition to this, we apply weights to each class (w_i for class i) to offset the imbalance of pixel frequency across different classes. Concretely, the weighted cross entropy loss L_{CE} is defined as,

$$L_{\text{CE}} = -\sum_{x} w_{t(x)} \log(p(x, t(x))) \tag{1}$$

Dice Loss. The Dice's Coefficient is a metric to measure the similarity between two given samples. Extending it as a loss function as shown in [8], improves the performance when dealing with situations where background pixels are higher than the labels. Here, we introduce a novel dice loss L_{dice} that is a logarithmic value of the dice score, making it easier to optimize. Similar to weighted cross entropy, we use weights to offset the class imbalance. Dice loss L_{dice} is weighted sum of dice losses l_i for each class i is given as,

$$L_{\text{dice}} = \sum_{i} w_i l_i \tag{2}$$

For class i, let's denote its binary map by t_i. i.e.,

$$t_i(x) = \begin{cases} 1, & \text{if } t(x) = i \\ 0, & \text{otherwise} \end{cases} \tag{3}$$

Then, dice loss for class i is given by,

$$l_i = \log\left(2 - \frac{\sum_x t_i(x)p(x,i) + \epsilon}{\sum_x t_i(x) + p(x,i) + \epsilon}\right) \tag{4}$$

Combined Dice Cross Entropy Loss. While cross-entropy loss optimizes for pixel-level accuracy, the Dice loss function enhances the segmentation quality. Combining these two objective functions, we define a weighted average of cross entropy L_{CE} and dice loss function L_{dice} formulated as Cross Entropy-Dice Loss $L_{\text{CE+dice}}$ in Eq. 5. Here, λ_{CE} and λ_{dice} are weight parameters for cross entropy loss and Dice loss function respectively.

$$L_{\text{CE+dice}} = \lambda_{\text{CE}} * L_{\text{CE}} + \lambda_{\text{dice}} * L_{\text{dice}} \tag{5}$$

3 Results

In this section we evaluate the performance of the proposed 2D and 3D segmentation models in terms of geometric or distance metrics and clinical metric scores for all 100 studies provided in the training phase of ACDC 2017 contest. Tables 1 and 2 presents the distance metric scores for our 2D model and 3D model respectively. For distance metric, we utilize the Dice Score and the Hausdorff Distance to measure the accuracy of segmented Left Ventricle (LV), Right Ventricle (RV) and Myocardium (MYO) in both end diastole (ED) and end systole (ES) phase. For each model, we compare the performance of the three optimization functions namely, the Cross Entropy Loss (CE Loss), Dice Loss and Combined Cross Entropy-Dice Loss (Dice-CE Loss). We observe that our proposed dice loss function outperforms CE Loss and CE-Dice Loss functions across all metrics in both 3D and 2D models. The Hausdorff distances in 3D models is observed to be much higher than 2D models, due to false positives as far-off speckles in 3D space. Illustration of results for both 2D and 3D models are provided in Figs. 3 and 4. Overall, we achieve near equal distance metric scores when compared to [14].

For clinical metrics, we use Correlation Coefficient (CC), Bias and Limits of Agreement (LOA). Our clinical metric results for 2D and 3D models are

Table 1. 2D segmentation: distance metric results

	Dice score						Hausdorff distance					
	LV		RV		MYO		LV		RV		MYO	
	ED	ES	ED	ES	ED	ES	ED	ES	ED	ES	ED	ES
CE loss	0.95	0.90	0.87	0.76	0.79	0.82	13.92	17.67	27.40	27.73	23.81	22.11
Dice loss	**0.95**	**0.90**	**0.90**	0.79	**0.86**	**0.88**	9.51	12.29	16.1	20.38	**13.45**	**14.88**
Dice-CE loss	0.95	0.90	0.89	**0.81**	0.83	0.84	**9.15**	**11.7**	**16.0**	**18.22**	13.87	15.35

Table 2. 3D segmentation: distance metric results

	Dice score						Hausdorff distance					
	LV		RV		MYO		LV		RV		MYO	
	ED	ES	ED	ES	ED	ES	ED	ES	ED	ES	ED	ES
CE loss	0.94	0.89	0.86	0.73	0.76	0.81	12.36	14.41	25.85	29.57	43.47	43.82
Dice loss	**0.95**	**0.90**	**0.91**	**0.83**	**0.85**	**0.86**	14.95	14.35	**23.15**	**22.14**	**37.75**	**38.50**
Dice-CE loss	0.94	0.89	0.91	0.81	0.83	0.85	**10.71**	**11.52**	38.01	32.26	43.28	44.98

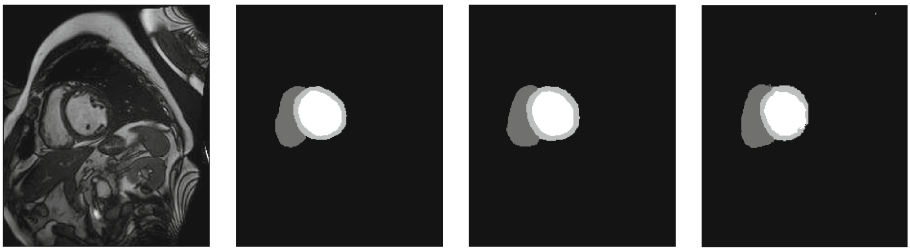


Fig. 3. Segmentation results for 2D and 3D model. **From Left to Right:** Raw MR input image slice, corresponding ground truth annotation, output predictions from 2D segmentation model and output predictions from 3D segmentation model.

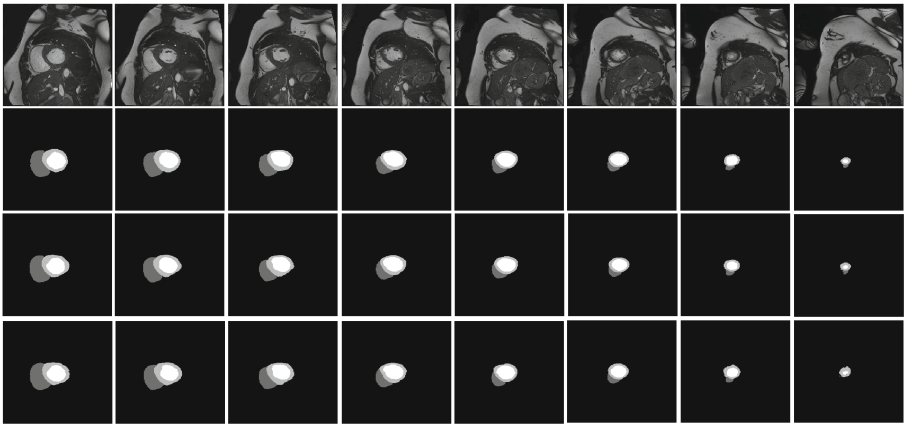


Fig. 4. Segmentation results for a full MR image (Complete Phase) from slices 0–8. **First Row:** Raw MR input images **Second Row:** Corresponding ground truth annotations. **Third Row:** Output predictions from 2D segmentation model. **Fourth Row:** Output predictions from 3D segmentation model.

presented in Tables 3 and 4 respectively. For both 2D and 3D models, the performance using cross-entropy and dice-loss functions for LV and RV is fairly similar, however the difference in performance is significant for MYO where dice-loss outperforms cross-entropy optimization. While distance metric scores

are fairly similar for both 2D and 3D model, in clinical parameters we observe that the 3D model outperforms the 2D model.

Although accuracy scores are important when making clinical decisions, runtime efficiency and memory usage of the algorithm are also crucial to apply it in real-world applications. Our 2D model takes 2.9 h to train using 4 GB of GPU memory. At testing, it takes 0.3 s and 1.2 GB GPU memory to generate a single output (whole phase). Whereas our 3D model requires 2.6 h for training and 4 GB of GPU memory. At test time, it can generate output within 0.3 s using 2 GB GPU memory. This shows that our 2D and 3D models are efficient to train, are light-weight and relatively easy to deploy in clinical settings.

Table 3. 2D segmentation: clinical metric results

	Ejection fraction						Myocardial mass		
	LV			RV			MYO		
	CC	Bias	LOA	CC	Bias	LOA	CC	Bias	LOA
CE loss	**0.95**	−0.74	**−12.48, 11.00**	0.822	9.79	**−10.89, 30.47**	0.93	−43.85	−92.12,4.42
Dice loss	0.88	1.06	−18.10, 20.22	**0.822**	**9.35**	−11.67, 30.37	**0.95**	**−6.32**	**−39.46, 26.82**
Dice-CE loss	0.93	**−0.46**	−15.67, 14.75	0.813	5.66	−16.59, 27.91	0.94	−29.31	−68.20, 9.58

Table 4. 3D segmentation: clinical metric results

	Ejection fraction						Myocardial mass		
	LV			RV			MYO		
	CC	Bias	LOA	CC	Bias	LOA	CC	Bias	LOA
CE loss	**0.975**	**1.04**	**−7.67, 9.75**	0.756	9.62	−15.11, 34.35	0.922	−48.17	−97.51, 1.17
Dice loss	0.956	1.51	−10.09, 13.11	0.825	**4.99**	−17.17, 27.15	**0.958**	**3.77**	**−24.91, 32.45**
Dice-CE loss	0.956	1.25	−10.37, 12.87	**0.867**	6.19	**−12.09, 24.47**	0.950	−10.08	−42.85, 22.69

4 Discussion

Model Structures. Since 2D segmentation model is trained by splitting MR volumes into slices, it lacks the spatial context across the 3D volume. This reduces the model's performance for slices at end of the phase, where the ratio of heart structure to background pixels is less. To solve this, we design our 2D model to accept a stack of image slices as input channels where the output is predicted for the middle slice. Given this design, we have 3 input options to pass as inputs: 1, 3 and 5. Among the three, we obtain the best performance with 3-input slices and report the scores for it in our results section.

Analyzing the 3D CNN segmentation model, we observe that the 3D model doesn't meet the expectations in performance improvement over 2D, given its ability to exploit 3D structure from input volumes. We see that Dice Score for RV in 3D is better 2D model given the fact that it has complex shape and intensity inhomogeneities. Thus, predicting RV using single slice is much more difficult

compared to looking at complete 3D context. Further improvements in performance can be brought about using post-processing techniques. Due to the modularity of 3D model architecture, we were able to quickly explore several designs and concepts. For example, we tried the recently introduced subpixel CNN's [15] that proposes using subpixel layers as opposed to transposed convolutions. We executed this by replacing the *deconv* blocks in the expanding paths with a *subpixel* block that comprises of a *subpixel* layer between two *conv_bn_relu* blocks. However, no performance improvements were to be observed.

Data Augmentation. At the initial stages of model design and training, we applied variety of data augmentation techniques demonstrated in [7,8], to incorporate randomness and robustness in the training the network. These include elastic deformations, random intensity jitter and affine transformations that include rotation, shearing, translation and flipping. Acquiring sub-par performances at testing, we found that applying heavy augmentations might misrepresent the anatomical structure of the heart. Instead, we opt to apply light data augmentations consisting of random rotations and scaling that can naturally match the variability of taking MR scans in real-world settings. With this modification, we observed a jump in performance scores for both 2D and 3D segmentation model.

Optimization Functions. Using our Dice loss as objective function improves the performance significantly for 2D and 3D models. As compared to pixel-level error optimization, dice loss is more robust and better at capturing the spatial context over the entire image. As shown in Tables 1 and 2, the best dice scores are achieved by using dice loss function.

5 Conclusion

This paper introduces a 2D and 3D convolutional neural network for fully-automated cardiac MR segmentation. Our models have light-weight modular architecture, easy implementation and run-time efficiency. Using multiple loss criterion, we compare and analyze the performance of 2D and 3D model pipelines and show that both our models achieve near state-of-the-art accuracy scores in terms of distance metrics. With convincing performance in clinical accuracy metrics, we also prove our model's viability in real-world practical applications. Through our discussions, we derive several insights that can be used for optimizing overall performance of these segmentation models. For future work, we plan to utilize our segmentation models to learn and classify different cardiac diseases.

References

1. Girshick, R.: Fast R-CNN. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 1440–1448 (2015)
2. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778 (2016)
3. Long, J., Shelhamer, E., Darrell, T.: Fully convolutional networks for semantic segmentation. In: Proceedings of CVPR, pp. 3431–3440 (2015)
4. Tran, P.V.: A fully convolutional neural network for cardiac segmentation in short-axis MRI. arXiv preprint arXiv:1604.00494 (2016)
5. Lieman-Sifry, J., Le, M., Lau, F., Sall, S., Golden, D.: FastVentricle: cardiac segmentation with ENet. arXiv preprint arXiv:1704.04296 (2017)
6. Ronneberger, O., Fischer, P., Brox, T.: U-Net: convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) MICCAI 2015. LNCS, vol. 9351, pp. 234–241. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-24574-4_28
7. Çiçek, Ö., Abdulkadir, A., Lienkamp, S.S., Brox, T., Ronneberger, O.: 3D U-Net: learning dense volumetric segmentation from sparse annotation. In: Ourselin, S., Joskowicz, L., Sabuncu, M.R., Unal, G., Wells, W. (eds.) MICCAI 2016. LNCS, vol. 9901, pp. 424–432. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-46723-8_49
8. Milletari, F., Navab, N., Ahmadi, S.A.: V-Net: Fully convolutional neural networks for volumetric medical image segmentation. In: Proceedings of 3DV 2016, pp. 565–571 (2016)
9. Automatic Cardiac Detection Challenge 2017. http://www.creatis.insa-lyon.fr/Challenge/acdc/
10. Ioffe, S., Szegedy, C.: Batch normalization: accelerating deep network training by reducing internal covariate shift. CoRR abs/1502.03167 (2015)
11. Srivastava, N., Hinton, G.E., Krizhevsky, A., Sutskever, I., Salakhutdinov, R.: Dropout: a simple way to prevent neural networks from overfitting. J. Mach. Learn. Res. **15**(1), 1929–1958 (2014)
12. Pizer, S.M., et al.: Adaptive histogram equalization and its variations. Comput. Vis. Graph. Image Process. **39**(3), 355–368 (1987)
13. PyTorch. http://pytorch.org/
14. Zotti, C., Luo, Z., Lalande, A., Humbert, O., Jodoin, P.M.: Novel deep convolution neural network applied to MRI cardiac segmentation. arXiv preprint arXiv:1705.08943 (2017)
15. Dong, C., Loy, C.C., He, K., Tang, X.: Image super-resolution using deep convolutional networks. IEEE Trans. Pattern Anal. Mach. Intell. **38**(2), 295–307 (2016)

Densely Connected Fully Convolutional Network for Short-Axis Cardiac Cine MR Image Segmentation and Heart Diagnosis Using Random Forest

Mahendra Khened, Varghese Alex, and Ganapathy Krishnamurthi(✉)

Indian Institute of Technology Madras, Chennai 600036, Tamil Nadu, India
gankrish@iitm.ac.in

Abstract. In this paper, we propose a fully automatic method for segmentation of left ventricle, right ventricle and myocardium from cardiac Magnetic Resonance (MR) images using densely connected fully convolutional neural network. Dense Convolutional neural network (DenseNet) facilitates multi-path flow for gradients between layers during training by back-propagation and feature propagation. DenseNet also encourages feature reuse & thus substantially reduces the number of parameters while maintaining good performance, which is ideal in scenarios with limited data. The training data was subjected to Fourier analysis and classical computer vision (CV) techniques for Region of Interest (ROI) extraction. The parameters of the network were optimized by training with a dual cost function i.e. weighted cross-entropy and Dice co-efficient. For the task of automated heart diagnosis, cardiac parameters such as ejection fraction, volumes of ventricles etc. where calculated from segmentation masks predicted by the network at the end systole and diastole phases. Further these parameters were used as features to train a Random forest classifier. On the exclusively held-out test set (10% of training set) the proposed method for segmentation task achieved a mean dice score of 0.92, 0.87 and 0.86 for left ventricle, right ventricle and myocardium respectively. For automated cardiac disease diagnosis, the Random Forest classifier achieved an accuracy of 90%.

Keywords: Cardiac MRI · Segmentation · CNN · FCN · DenseNet Inception · Dice loss

1 Introduction and Related Work

In clinical practice, MRI is preferred over ultrasound and CT due to its superior spatial-temporal resolution and non-ionizing radiation. Cardiac parameters such as left ventricular ejection fraction, volumes of the left ventricle and right ventricle, myocardial thickness are calculated routinely to diagnose a subject as healthy or diseased. For the aforementioned reason, segmentation of the structures such as left ventricle, right ventricle and myocardium from MR images

M. Pop et al. (Eds.): STACOM 2017, LNCS 10663, pp. 140–151, 2018.
https://doi.org/10.1007/978-3-319-75541-0_15

becomes a pivotal step. Manual segmentation of the structures from the surrounding tissues is a tedious task and often introduces inter-rater variability.

Convolutional neural networks (CNNs) [1] have been applied to wide variety of pattern recognition tasks, most common ones are image classification [2–4] and semantic segmentation using fully convolutional networks (FCN) [5]. CNNs have also been applied to medical image segmentation and classification [7,8]. In this paper, we propose a CNN based architecture for segmentation of the left ventricle, right ventricle and myocardium from short-axis view of cardiac MR images. Our network's connectivity pattern was inspired from DenseNets [10]. DenseNets connects each layer to every other layer in a feed-forward fashion by concatenation of all feature outputs. The output of the l^{th} layer is defined as

$$x_l = H_l([x_{l-1}, x_{l-2}, \cdots, x_0]) \tag{1}$$

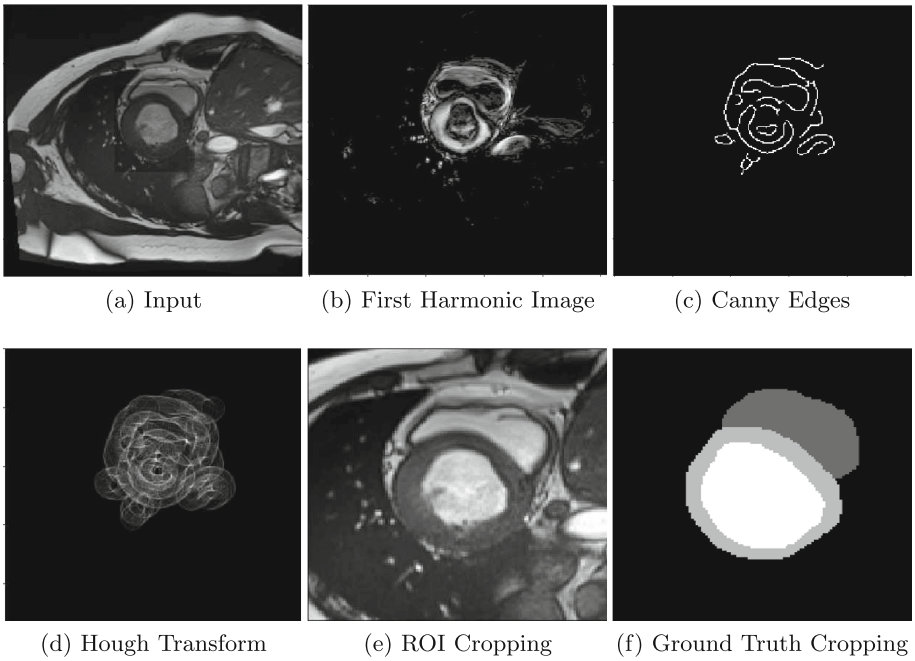
where x_l represents the feature maps at the l^{th} layer and $[\cdots]$ represents the concatenation operation. In our case, H is the layer comprising of Batch Normalization (BN) [22], followed by Exponential Linear Unit (ELU) [23], a convolution and dropout [21]. This kind of connectivity pattern aids in reuse of features and allows implicit deep supervision during training. The output dimension of each layer has k (growth rate parameter) feature maps. The number of feature maps in DenseNets grow linearly with depth. A Transition Down layer in DenseNets is introduced for reducing spatial dimension of feature maps which is accomplished by using a 1×1 convolution (depth preserving) followed by a 2×2 max-pooling operation. A Dense-Block refers to concatenation of new feature maps created at a given resolution.

2 Our Method

2.1 Data Pre-processing Pipeline

Region of Interest (ROI) Detection. The cardiac MR images of the patient comprises of the heart and various surrounding structures like the lungs and diaphragm. Since the task at hand was segmentation of various heart structures, an automated method for region of interest detection was carried out to delineate the heart structures from the surrounding tissues. The Fourier based techniques [13], employ the fact that each slice sequence in time captures one heartbeat. Fourier analysis was done to extract first harmonic images which captured the maximal activity at the corresponding heartbeat frequency. Assuming that the left ventricle approximates a circle, the first harmonic images were subjected to canny edge detector. The approximate radius & center of the left ventricle were calculated from the edge maps using circular Hough transform approach [14]. Figure 1 shows the extraction of ROI using the proposed technique.

Data Augmentation. Data augmentation was done to artificially increase the size of the dataset. Pixel-Spacing information was used to rescale the images to 1 mm spacing. The ROI detection estimates the approximate center of the left



(a) Input (b) First Harmonic Image (c) Canny Edges

(d) Hough Transform (e) ROI Cropping (f) Ground Truth Cropping

Fig. 1. Fourier based Region of interest (ROI) detection scheme is based on capturing pixel regions where there is maximal intensity variation over one full cardiac cycle. These pixels mostly correspond to ventricular regions of heart. Cropping a patch of fixed size centered around the left ventricle (LV) leads to removal of irrelevant structures. The steps involved in ROI detection are (a) Temporal slices, (b) Estimation of first harmonic image using Fourier Analysis, (c) Canny edge-detection on the harmonic image, (d) Circular Hough Transform on edge-map to localize LV, (e)–(f) ROI cropping on the input & ground-truth image

ventricle C, further a patch of size 128×128 centered around C was extracted from the rescaled image. This method helped in alleviating the huge class-imbalance problem associated with labels for heart structures seen in the full sized cardiac MR images. In addition, the proposed technique enables the network to precisely learn the fine-grained structures of the heart. Most importantly, this approach reduces the computation time required for learning the parameters of network and also during inference. The data augmentation scheme employed were:

- rotation: random angle between -5 and $5°$ (uniform)
- translation x-axis: random shift between -5 and 5 mm (uniform)
- translation y-axis: random shift between -5 and 5 mm (uniform)
- rescaling: random zoom factor between .6 and 1.4 (uniform)
- horizontal and vertical flipping: yes or no (bernoulli)

Normalization. Each slice of the patient's voxel intensities were normalized to the range of 0–1 using Eq. (2)

$$X_{norm} = \frac{X - X_{min}}{X_{max} - X_{min}} \tag{2}$$

where X is voxel intensity.

2.2 Proposed Network Architecture: Densely Connected Fully Convolutional Network (DFCN)

Figure 2 illustrates the schematic diagram of our proposed network for segmentation. The down-sampling and up-sampling components adopts the fully convolutional DenseNets architecture for semantic segmentation as described in [9]. Each layer in the dense block is sequentially composed of BN-ELU and a 3×3 convolution layers. The first Dense-Block was prefixed with a naive version of Inception module [11] comprising of convolution filters of size 3×3, 5×5 and 7×7. In the down-sampling path, the input to a dense block was concatenated with its output, leading to a linear growth of the number of feature maps. The Transition-Down block (TD) consists of BN-ELU a 1×1 convolution and a 2×2 max-pooling layers. The last layer of the down-sampling path is referred to as Bottleneck.

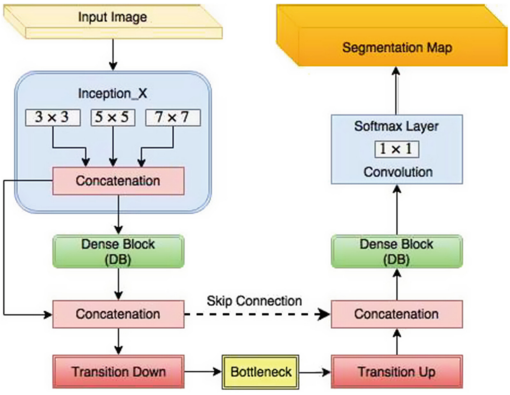


Fig. 2. Architecture of DFCN.

In the up-sampling path, the input of a Dense-Block is not concatenated with its output. Transition-Up (TU) block comprises of 3×3 transposed convolution layer with a stride of 2. The output feature maps of the TU block was concatenated (via skip connection) with the feature maps corresponding to those DB's from down-sampling path. The feature maps of the hindmost up-sampling component was convolved with a 1×1 convolution layer followed by a soft-max layer

to generate the final label map of the segmentation. To prevent over-fitting, a dropout of 0.2 was implemented following each convolution layer.
Table 1 summaries the individual blocks of our architecture. For the segmentation task, the proposed network's architecture is summarized in Table 2. The number of trainable parameters is about 4×10^6 (4M) in total, which is far lesser than number of trainable parameters in U-Net [6] (30M parameters). It was observed that using exponential linear units (ELUs) instead of rectified linear units (ReLUs) led to faster convergence.

Table 1. Building blocks of DFCN. From left to right: layer used in the model, Transition Down (TD) and Transition Up (TU).

Layer
Batch Normalization
Exponential Linear Unit
3×3 Convolution
Dropout $p = 0.2$

TD
Batch Normalization
Exponential Linear Unit
1×1 Convolution
Dropout $p = 0.2$
2×2 Max Pooling

TU
3×3 Transposed Convolution stride = 2

Table 2. Architecture details of model used in our experiments. The growth rate parameter $k = 8$

DFCN Architecture
Input_Size: (128×128), channels=1
Inception_X: 3×3 (16), 5×5 (4), 7×7 (4) convolutions, features = 24
Dense Block (3 layers) + Transition Down, features = 48
Dense Block (4 layers) + Transition Down, features = 80
Dense Block (5 layers) + Transition Down, features = 120
BottleNeck (8 layers), features = 176
Transition Up + Dense Block (5 layers), features = 216
TransitionUp + Dense Block (4 layers), features = 144
Transition Up + Dense Block (3 layers), features = 104
1×1 convolution, channels = 4
Softmax Layer

Number of Convolutional layers : 42
Number of parameters : 374292

2.3 Loss Function

The anatomical structures of interest in the medical images are sparsely represented in whole volume. This leads to class imbalance in the dataset, thereby

making it hard for the network to learn subtle structures in the region of interest. In order to address this issue, the loss function used, weighting mechanism based on class frequencies. A weighted combination of two loss function, namely:- cross-entropy loss and a loss function based on Dice overlap co-efficient [12] was used to train the network.

The dice co-efficient is an overlap metric used for assessing the quality of segmentation maps. The dice coefficient between two binary volumes can be written as:

$$DICE = \frac{2\sum_i^N p_i g_i}{\sum_i^N p_i^2 + \sum_i^N g_i^2} \tag{3}$$

where the sums run over the N voxels, of the predicted binary segmentation volume $p_i \in P$ and the ground truth binary volume $g_i \in G$.

For multi-class problem the dice loss can be written as weighted sum of Eq. 3. The weights are empirically determined based on the their relative class frequencies. The total dice loss for multi-class segmentation problem is given in Eq. (4):

$$dice\_loss = \frac{W_{class1} DICE_{class1} + \cdots + W_{classN} DICE_{classN}}{W_{class1} + \cdots + W_{classN}} \tag{4}$$

where W_{classN} is the empirically assigned weight based on its relative frequency, smaller the frequency higher the assigned weight.

The parameters of the network were optimized so as to minimize the $total\_loss$, Eq. (5).

$$total\_loss = \lambda(cross\_entropy\_loss) + \gamma(1 - dice\_loss) + L2\_loss \tag{5}$$

where λ and γ are empirically assigned weights to individual losses. During training it was observed that the Dice loss allowed higher overlap scores than when trained with the loss function based on the cross entropy loss alone. In this work we set $\gamma = 0.75$ and $\lambda = 0.25$.

The proposed model was trained on a batch size of 10 2D-MR images for 200 epochs using ADAM [20] as the optimizer. The learning rate was set to 10^{-4} and additionally a $L2$ weight decay of 10^{-4} was added to the cost function as a regularizer.

2.4 Post-processing

The results of segmentation predicted by DFCN network were subjected to connected component analysis to remove false positives. The largest the component (heart structures) was retained, while the rest were discarded.

2.5 Cardiac Disease Diagnosis

To develop an automated cardiac diagnosis system the following 10 attributes from the training dataset were used:

- Ejection fraction of left ventricle and right ventricle
- Volume of the left ventricle at end systole and end diastoles phases
- Volume of the right ventricle at end systole and end diastole phases
- Mass of the myocardium at end diastole and its volume at end systole
- Patient height and weight

Initially, these 11 attributes were calculated from the training set and were used for training a Random Forest classifier [19]. The proposed Random Forest classifier comprises of 100 trees. On the test set, the segmentation maps predicted from the trained neural network was used for calculating the above listed 11 cardiac parameters. These parameters were fed as input to the trained Random Forest classifier to diagnose the patient as: Dilated cardiomyopathy (DCM), Hypertrophic cardiomyopathy (HCM), Myocardial infarction (MNF), Abnormal right ventricle (ARV) and normal patients (NOR).

3 Experimental Setup and Results

3.1 Dataset and Evaluation Criteria

The network was trained and tested on the ACDC STACOM 2017 challenge dataset, comprising of 100 patients. The patients were divided into 5 evenly distributed groups: dilated cardiomyopathy (DCM), hypertrophic cardiomyopathy (HCM), myocardial infarction (MNF), abnormal right ventricle (ARV) and normal patients (NOR). The end diastolic (ED) and the end systolic (ES) phases come with a pixel-accurate manual delineation by two independent medical experts. The dataset was split into 70 for training, 20 for validation and 10 for testing using stratified sampling (strata for sampling is based on the cardiac disease). In order to gauge performances on the held out test set, we report the clinical metrics such as the average ejection fraction (EF) error, the average left ventricle (LV) and right ventricle (RV) systolic and diastolic volume errors, and the average myocardium (MYO) mass error. For the geometrical metrics, we report the Dice and the Hausdorff distances for all 3 regions at the ED and ES phases. For the cardiac disease diagnosis the metrics used was accuracy, precision and recall.

3.2 Experimental Results

The proposed model was evaluated on the exclusively held-out testing data ($n = 10$). Table 3 shows the average dice scores achieved by the model at ED and ES phases of the heart. The model achieves a relatively higher dice score for LV when compared to RV and MYO. The proposed method relies on localization of the LV and cropping a patch of fixed size from the LV region's center as pre-processing step before feeding into the network. So, in cases of abnormally large RV, the model slightly under-performs when RV region extends beyond the patch size. The aforementioned reasons & irregular shape of RV when compared to LV leads to a dip in dice score & higher Hausdorff distance of RV.

Figures 3 and 4 shows the results of segmentation produced by the proposed network at ED and ES phase of the heart. It was observed that model generates good segmentations throughout the volume. However due to small structures at the base and close proximity of valves such as aorta at the apex regions leads to erroneous segmentation. For example, in Fig. 3(h) myocardium is slightly over-segmented at the apex region of the heart, while in Fig. 4(b) the model does some erroneous segmentation in the basal slices near valves of the heart.

The model's geometric metrics are slightly better at ED phase than at ES phase, whereas the clinical metrics (Tables 4 and 5) namely the LV & RV volume error and MYO mass error are relatively better at ES phase.

Table 6 compares the effect of training a network with proposed loss function as opposed to training with the vanilla cross-entropy. It was observed that the proposed loss function manifested in producing better segmentations when compared to vanilla cross entropy and thus led to improvement of dice score by 2%.

Table 7 shows the result of the cardiac disease diagnosis on the testing data. The Random Forest classifier's accuracy heavily depends on the clinical metrics, which in-turn depends on the segmentation results of the proposed model.

Table 3. Results of geometrical metrics on the testing dataset.

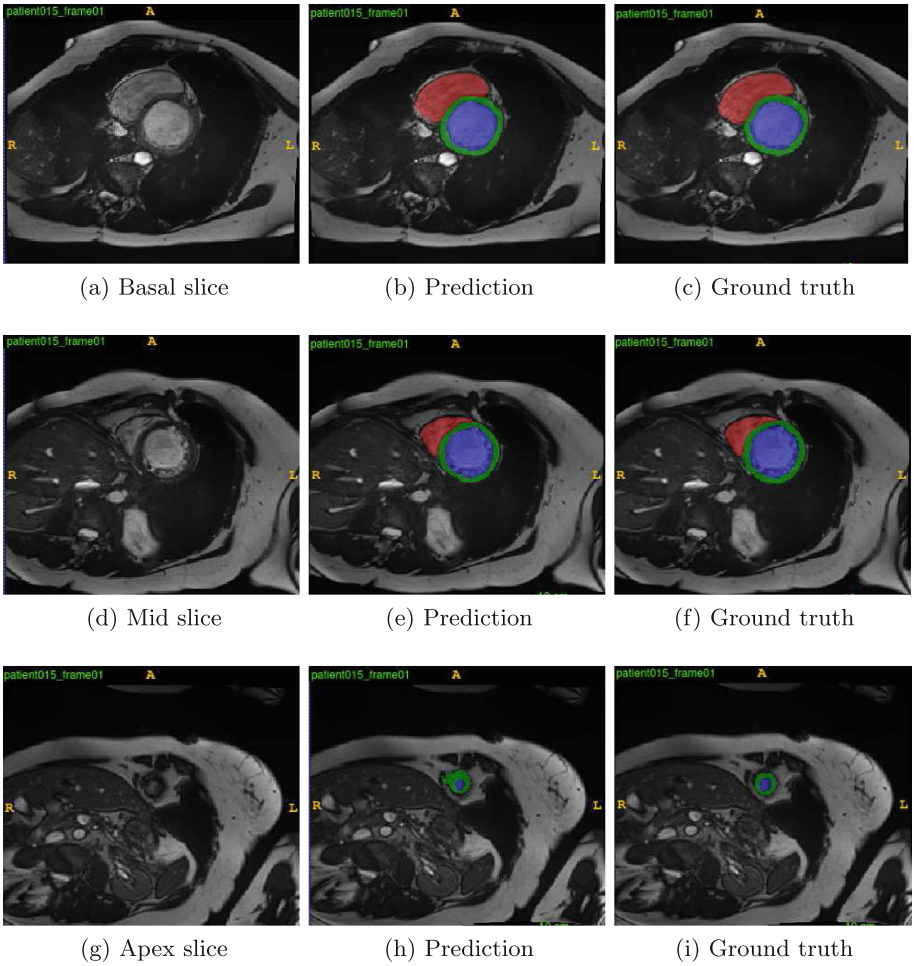
	Cardiac phase					
	End diastole			End systole		
Geometric metric	LV	RV	MYO	LV	RV	MYO
Dice score	.94	.89	.84	.89	.84	.87
Hausdorff distance	12.13	18.97	17.05	12.04	23.97	12.92

Table 4. Results of clinical metrics on the testing dataset.

Ejection fraction error (%)		Left ventricle volume error (mL)		Right ventricle volume error (mL)		MYO mass error (g)	
Left ventricle	Right ventricle	Diastole	Systole	Diastole	Systole	Diastole	Systole
2	11.1	3.2	1.7	9.7	4	14.1	10.6

Table 5. Results of clinical metrics on the testing dataset.

Clinical metric	Ejection fraction (%)		Volume ED (ml)		Volume ES (ml)			Mass ED (g)
	LV	RV	LV	RV	LV	RV	MYO	MYO
Correlation coefficient	0.980	0.889	0.968	0.903	0.983	0.960	0.894	0.859
BIAS	3.77	11.73	9.68	13.26	2.07	−6.98	−4.03	−10.09
LOA	[−6.21; 13.75]	[−9.37; 32.83]	[−11.86; 31.22]	[−28.63; 55.15]	[−27.02; 31.16]	[−29.03; 15.07]	[−39.07; 31.01]	[−48.53; 28.35]



(a) Basal slice (b) Prediction (c) Ground truth

(d) Mid slice (e) Prediction (f) Ground truth

(g) Apex slice (h) Prediction (i) Ground truth

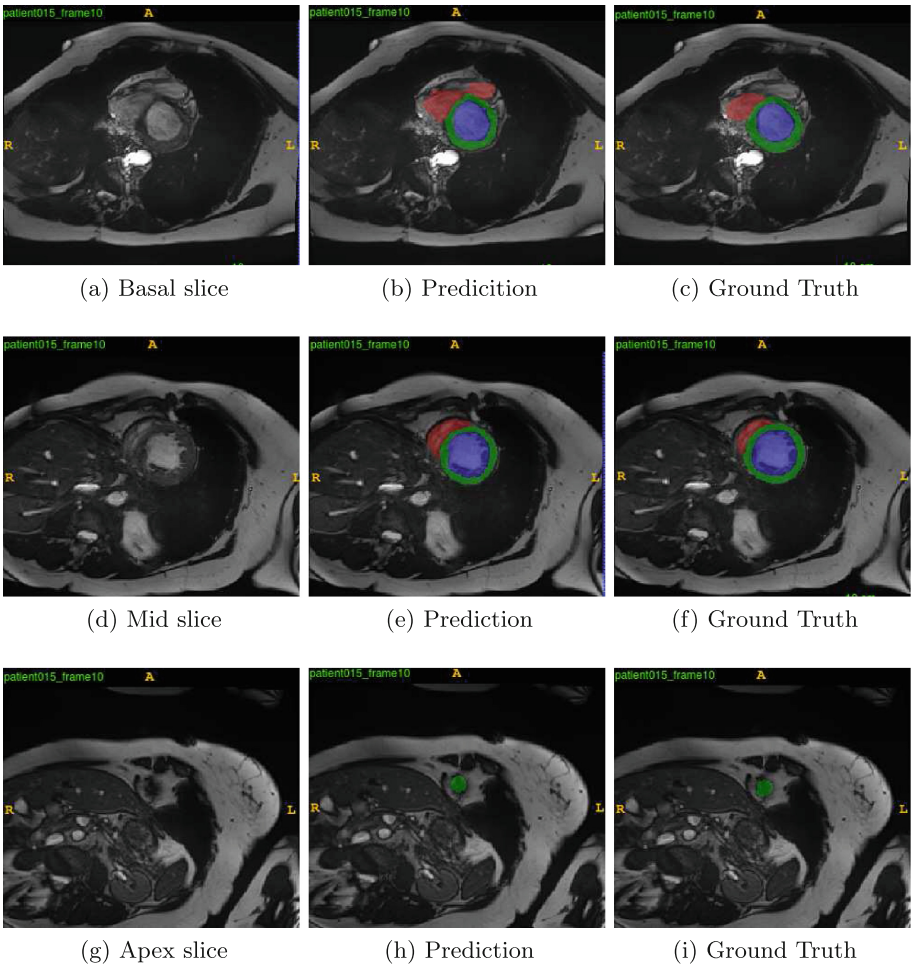
Fig. 3. The figure shows the segmentation results generated by the proposed model on test-set at ED phase of heart. The columns from left to right indicate: the input images, segmentations generated by the model and their associated ground-truths. The rows from top to bottom indicate: short axis slices of the heart at basal, mid and apex. In all figures the colors red, green and blue indicate RV, MYO and LV respectively. (Color figure online)

Table 6. Evaluation comparison for the proposed loss function

	Average dice score		
Heart Structure	Left ventricle	Right ventricle	Myocardium
Proposed loss function	.92	.87	.86
Vanilla cross-entropy loss	.90	.85	.83

Table 7. Results of automated cardiac diagnosis on the testing dataset.

Disease →	NOR	DCM	HCM	MINF	ARV
Recall	1	1	1	1	1
Precision	1	0.67	0.5	1	1
Overall classification accuracy: 0.90					



(a) Basal slice (b) Predicition (c) Ground Truth

(d) Mid slice (e) Prediction (f) Ground Truth

(g) Apex slice (h) Prediction (i) Ground Truth

Fig. 4. The figure shows the segmentation results generated by the proposed model on test-set at ES phase of heart. The columns from left to right indicate: the input images, segmentations generated by the model and their associated ground-truths. The rows from top to bottom indicate: short axis slices of the heart at basal, mid and apex. In all figures the colors red, green and blue indicate RV, MYO and LV respectively. (Color figure online)

3.3 Conclusion

We propose a new CNN based method for cardiac MR image segmentation which is based on DenseNet connectivity pattern and inception modules.

- The proposed architecture showed that higher performance can be achieved with fewer trainable parameters by properly designing the network connectivity pattern and loss function.
- The customized loss function was observed to perform to better when compared to vanilla cross-entropy loss.
- Replacing ReLUs with ELUs manifested in faster convergence & improved segmentation metrics.

The proposed model was implemented in Tensorflow [17] and Theano [15,16]. The network was trained on NVIDIA K20c GPU. The entire pipeline (ROI extraction, prediction and post-processing) takes about 3 s for one patient's heart volume comprising of 10 SAX-slices at ED and ES phases of heart.

References

1. LeCun, Y., et al.: Gradient-based learning applied to document recognition. Proc. IEEE **86**(11), 2278–2324 (1998)
2. Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. In: Advances in Neural Information Processing Systems (2012)
3. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556 (2014)
4. He, K., et al.: Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (2016)
5. Long, J., Shelhamer, E., Darrell, T.: Fully convolutional networks for semantic segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (2015)
6. Ronneberger, O., Fischer, P., Brox, T.: U-net: convolutional networks for biomedical image segmentation. arXiv preprint arXiv:1505.04597 (2015)
7. Ciresan, D., et al.: Deep neural networks segment neuronal membranes in electron microscopy images. In: Advances in Neural Information Processing Systems (2012)
8. Menze, B.H., et al.: The multimodal brain tumor image segmentation benchmark (BRATS). IEEE Trans. Med. Imaging **34**(10), 1993–2024 (2015)
9. Jgou, S., et al.: The one hundred layers tiramisu: fully convolutional DenseNets for semantic segmentation. arXiv preprint arXiv:1611.09326 (2016)
10. Huang, G., et al.: Densely connected convolutional networks. arXiv preprint arXiv:1608.06993 (2016)
11. Szegedy, C., et al.: Going deeper with convolutions. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (2015)
12. Milletari, F., Navab, N., Ahmadi, S.-A.: V-net: fully convolutional neural networks for volumetric medical image segmentation. In: 2016 Fourth International Conference on 3D Vision (3DV). IEEE (2016)
13. https://www.kaggle.com/c/second-annual-data-science-bowl/details/fourier-based-tutorial

14. http://irakorshunova.github.io/2016/03/15/heart.html
15. Theano Development Team. Theano: a Python framework for fast computation of mathematical expressions (2016)
16. Lasagne Development Team. Lasagne: First release (2015)
17. Abadi, M., et al.: TensorFlow: large-scale machine learning on heterogeneous systems (2015). http://www.tensorflow.org
18. van der Walt, S., Schnberger, J.L., Nunez-Iglesias, J., Boulogne, F., Warner, J.D., Yager, N., Gouillart, E., Yu, T.: scikit-image: image processing in Python. PeerJ **2**, e453 (2014)
19. Liaw, A., Wiener, M.: Classification and regression by random forest. R News **2**(3), 18–22 (2002)
20. Kingma, D., Ba, J.: Adam: a method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014)
21. Srivastava, N., et al.: Dropout: a simple way to prevent neural networks from overfitting. J. Mach. Learn. Res. **15**(1), 1929–1958 (2014)
22. Ioffe, S., Szegedy, C.: Batch normalization: accelerating deep network training by reducing internal covariate shift. In: International Conference on Machine Learning (2015)
23. Clevert, D.-A., Unterthiner, T., Hochreiter, S.: Fast and accurate deep network learning by exponential linear units (elus). arXiv preprint arXiv:1511.07289 (2015)

Class-Balanced Deep Neural Network for Automatic Ventricular Structure Segmentation

Xin Yang[1(✉)], Cheng Bian[2], Lequan Yu[1], Dong Ni[2], and Pheng-Ann Heng[1,3]

[1] Department of Computer Science and Engineering, The Chinese University of Hong Kong, Sha Tin, Hong Kong
xinyang@cse.cuhk.edu.hk

[2] National-Regional Key Technology Engineering Laboratory for Medical Ultrasound, School of Biomedical Engineering, Health Science Center, Shenzhen University, Shenzhen, China

[3] Guangdong Provincial Key Laboratory of Computer Vision and Virtual Reality Technology, Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen, China

Abstract. Segmenting ventricular structures from cardiovascular MR scan is important for quantitative evaluation of heart. Manual delineation is time-consuming and tedious and lack of reproductivity. Considering MR image quality, heart variance, spatial inconsistency and motion artifacts during scanning, it is still a non-trivial task for automatic segmentation methods. In this paper, we propose a general and fully automatic solution to concurrently segment three important ventricular structures. Rooting in the deep learning trend, our method starts from 3D Fully Convolutional Network (3D FCN). We then enhance the 3D FCN with two well-verified blocks: (1) we conduct transfer learning between a pre-trained C3D model and our 3D FCN to get good initialization and thus suppress overfitting. (2) since boosting the gradient flow in network is beneficial to promote segmentation performance, we attach several auxiliary loss functions so as to expose early layers to better supervision. Because the volume size imbalance among different ventricular structures often biases the training of our 3D FCN, to this end, we investigate the capacity of different loss functions and propose a Multi-class Dice Similarity Coefficient ($mDSC$) based loss function to re-weight the training for all classes. We verified our method, especially the significance of $mDSC$, on the Automated Cardiac Diagnosis Challenge 2017 datasets for MR image segmentation. Extensive experimental results demonstrate the promising performance of our method.

1 Introduction

Cardiovascular diseases are leading cause of death around world. MR imaging serves as a versatile and indispensable tool for radiologist to noninvasively inspect anatomical and functional defects of heart. Segmenting the ventricular structures

X. Yang and C. Bian contributed equally to this work.

M. Pop et al. (Eds.): STACOM 2017, LNCS 10663, pp. 152–160, 2018.
https://doi.org/10.1007/978-3-319-75541-0_16

is fundamental and even a prerequisite for many following quantitative analysis, such as ejection fraction and myocardial mass [10]. However, facing with the explosive growth of volumetric data, corrupted image quality and inevitable user dependency, manually delineating the ventricular structures is prone to low efficiency and low reproducibility. Therefore, there is a strong need for automatic solutions to segment the different ventricular structures.

Developing automatic solutions for efficient and accurate segmentation of ventricular structure is a non-trivial task. As observed from the Challenge datasets, there exist large shape and size variations of heart. Due to the movement and breath-hold during image acquisition, spatial inconsistency caused by shift between slices are common. Also, the boundary uncertainty and spatial inhomogeneities severely corrupt the image quality [19]. Early attempts for heart structure segmentation have been contributed from active contours [4], shape and texture prior knowledge based deformable models [11], non-rigid registration [20] and level sets [7]. However, the empirically designed features and limited training data set performance bound for previous studies. As it comes to the era of deep learning, segmenting ventricular structures also benefits a lot from deep neural networks. Fully Convolutional Networks (FCN) presents attractive capability in end-to-end mapping, and has been explored in a 2D manner for MR cardiac image segmentation in [16,21]. Recently, utilizing 3D FCN for direct volumetric segmentation becomes the trend [2]. Variations of 3D FCN achieve superior performance in segmenting cardiovascular volumes [3,18]. Promising as they are, combating with vanishing gradient and overfitting on limited training data are still important concerns for 3D FCN. More importantly, the class imbalance among ventricular structures are amplified in volumes and tend to bias the training of neural networks only towards major classes.

In this paper, we propose a general and fully automatic solution for ventricular structures segmentation in MR volumes, which can localizes the ventricular region and simultaneously label the region with 3 pre-defined categories, including left ventricular endocardium (LVEDO), left ventricular epicardium (LVEPI) and right ventricular endocardium (RVEDO). The proposed framework originates from 3D Fully Convolutional Network (3D FCN) for an efficient end-to-end mapping and is then reinforced from the following aspects. First, by inheriting the trainable filters from a C3D model which is trained on the large scale Sports-1M video dataset [15], our network launches with a good initialization and is able to cope with overfitting. Second, we further exploit deep supervision mechanism to promote the gradient flow within the network by shortening the backpropagation path and exposing early layers to the direct supervision of auxiliary loss functions [3,6]. Considering the bias caused by volume imbalance among classes, we discard the trivial cross-entropy loss and investigate different loss functions in removing the bias. We finally choose to extend the Dice Similarity Coefficient based loss function proposed in [9] into a Multi-class variant ($mDSC$) to significantly re-weight the training for all classes. We evaluated our method on the Automated Cardiac Diagnosis Challenge (ACDC) 2017 datasets for segmentation. Extensive experimental results demonstrated the promising performance of our method, especially the $mDSC$.

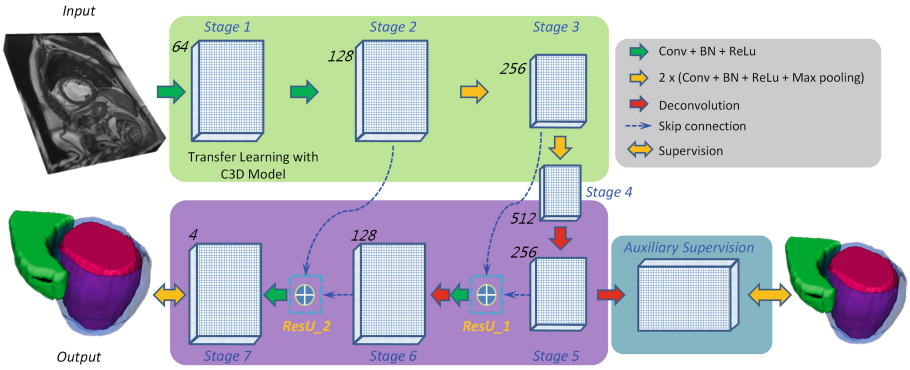


Fig. 1. Illustration of proposed framework. Digits are the number of feature volume channel. Down- and up-sampling path are denoted with green and purple rectangles. (Color figure online)

2 Methodology

Figure 1 is the schematic view of our proposed framework. System input is the complete volume of cardiovascular MR scanning. Our 3D FCN firstly extracts abstract from the volume with a downsampling path. A following upsampling path then gradually decodes the encoded feature maps into the labeling results. Skip connections are constructed with a residual unit to blend features from different semantic levels. The downsampling path is transferred from C3D model. Our proposed *mDSC* is adopted in main and auxiliary loss functions. System output is the semantic labeling of 3 ventricular structures.

2.1 Efficient Semantic Labeling with 3D FCN

Characterized with end-to-end mapping, Fully Convolutional Network (FCN) [8] is popular in semantic segmentation. By building skip connections between down- and up-sampling paths, U-net [13] promotes FCN in enhancing segmentation details. Shown as Fig. 1, running in a 3D fashion, we customize a 3D FCN, which is similar with that in [2] but substitutes the concatenation operator with a residual unit (ResU) to smooth gradient flow. To reduce computation cost, we use small convolution kernels with size of $3 \times 3 \times 3$ in convolutional layers (Conv). Limited by volume dimension, we only insert 2 pooling layers. Each Conv layer is followed by a batch normalization (BN) layer and a rectified linear unit (ReLU). Our tailored 3D FCN outputs probability volumes for 4 classes (including background), and finally the labeling volumes.

2.2 Transfer Learning from C3D Model

For deep neural networks, the filters learned by shallow layers can be general across different tasks. When facing with limited training data, leveraging parameters of well-trained model proves to be beneficial in improving generalization

ability, even the pre-trained model is obtained in a different domain [1]. Popular models, like ImageNet [5] and VGG16 [14] are only tractable for 2D applications. The C3D model proposed in [15] for video recognition sheds light on transfer learning in 3D networks, since it is equipped with 3D convolution operators to capture spatial and temporal abstract across consecutive frames. Therefore, we initialize the downsampling path of our 3D FCN with the same parameters from layers *conv1, conv2, conv3a, conv3b, conv4a and conv4b* in C3D model. During fine-tuning, learning rate for all layers are set as the same 0.001.

2.3 Promote Training with Deep Supervision

Subject to gradient vanishing issue, the parameter tuning processes of our 3D-FCN is at high risks of low efficiency and overfitting. In this paper, we adopt the deep supervision strategy introduced in [3,6], which promotes training by exposing shallow convolutional layers to the direct supervision of $\mathcal{M}$ auxiliary classifiers. The final loss function for our deeply supervised 3D FCN is formulated as Eq. 1, where $\mathcal{X}$, $\mathcal{Y}$ are training pairs, $\mathcal{W}$ is the weight of main network. $w = (w^1, w^2, .., w^m)$ are the weights of auxiliary classifiers, α_m is the corresponding ratio in final loss, $m = 1$ in this paper. The specific format for main loss $\mathcal{L}$ and auxiliary $\mathcal{L}_m$ will be explained in Sect. 2.4.

$$\mathcal{L}(\mathcal{X}, \mathcal{Y}; \mathcal{W}, w) = \mathcal{L}(\mathcal{X}, \mathcal{Y}; \mathcal{W}) + \sum_{m \in \mathcal{M}} \alpha_m \mathcal{L}_m(\mathcal{X}, \mathcal{Y}; \mathcal{W}, w^m) + \lambda(||\mathcal{W}||^2) \quad (1)$$

2.4 Investigation of Class-Balanced Loss

The feedback from loss function drives the latent mapping that deep neural network needs to fit. Popular choice of loss function for differentiable optimization is cross entropy. However, utilizing cross entropy based loss function is improper in classification or segmentation occasions where classes present significantly different proportions [17]. Class imbalance is obvious in ventricular structure segmentation. Classical cross-entropy based loss function trivially summarizes the error of each voxel without giving associated significances for specific classes, which will lead the network to ignore minor classes and only focus on major ones so as to minimize the loss. Based on [9,17], in this paper, we conduct investigation on different loss functions in re-weighting classes.

Volume Size Weighted Cross Entropy. As proposed in [17] for edge extraction which is rare compared to background, weighting the cross entropy loss for different classes can reduce class imbalance. In this paper, we extend the formulation in [17], and propose a patch-wise weighted cross entropy (*wCross*) for multiple classes. Mathematically, the formulation of *wCross* is shown as Eq. 2. $\mathcal{X}$ represents the training samples and $p(y_i = \ell(x_i)|x_i; W)$ is the probability of

target class label $\ell(x_i)$ corresponding to sample $x_i \in \mathcal{X}$. $|\mathcal{X}^{\ell(x_i)}|$ is the volume size of class $\ell(x_i)$ in patch $|\mathcal{X}|$. Minor class get larger weight with $\eta_{\ell(x_i)}$.

$$\mathcal{L}_{wCross}(\mathcal{X}; W) = \sum_{x_i \in \mathcal{X}} -\eta_{\ell(x_i)} \log p(y_i = \ell(x_i)|x_i; W), \eta_{\ell(x_i)} = 1 - \frac{|\mathcal{X}^{\ell(x_i)}|}{|\mathcal{X}|} \quad (2)$$

Multi-class Dice Similarity Coefficient. Dice Similarity Coefficient (DSC) based loss function is another novel attempt to alleviate class imbalance [9]. By focusing on the global shape similarity, DSC based loss for each class is normalized within class to [0, 1], so the bias in final loss caused by class imbalance is reduced. We extend the DSC loss in [9] and propose a multi-class Dice Similarity Coefficient (*mDSC*) based loss function to balance the training for multiple classes. Given the segmentation ground truth $G^{w \times h \times d}$, we firstly encode it into a one-hot format for C classes $\mathcal{G}^{C \times w \times h \times d}$, $C = 8$ for our task. With probability volumes $\mathcal{P}^{C \times w \times h \times d}$, our proposed *mDSC* can be written as

$$\mathcal{L}_{mDSC} = -\sum_{c \in C} \frac{\frac{2}{N} \sum_i^N \mathcal{G}_c^i \mathcal{P}_c^i}{\sum_i^N \mathcal{G}_c^i \mathcal{G}_c^i + \sum_i^N \mathcal{P}_c^i \mathcal{P}_c^i}, \quad (3)$$

where $N = w \times h \times d$, $\mathcal{G}_c^i$ and $\mathcal{P}_c^i$ are the i^{th} voxel of c^{th} volume in $\mathcal{G}$ and $\mathcal{P}$. Same with [9], for differentiable optimization, the formulation of *mDSC* is not strictly

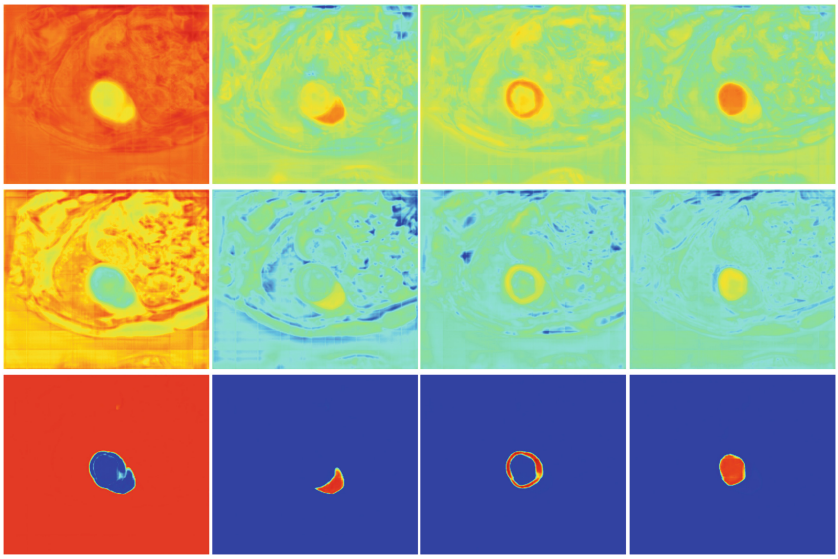


Fig. 2. From left to right: probability map of background, right ventricular endocardium, left ventricular epicardium and left ventricular endocardium. From top to bottom: training with cross-entropy, *wCross* and *mDSC*.

implemented as the definition of DSC. Also, for the numerator in $\mathcal{L}_{mDSC}$, we take the mean over N voxels rather than the summation, since we get much more noise segmentation results with the latter choice.

With experiments, we find that, *mDSC* guided model tend to generate more compact and clean probability maps than traditional cross entropy based loss and the *wCross* based loss, as the explicit comparison shown in Fig. 2, where the warmer the color, the higher the probability belonging to the class. Further illustration of improvement caused by *mDSC* is provided in Sect. 3.

3 Experimental Results

Dataset and Pre-processing. We evaluated our network on the segmentation tasks of ACDC 2017 Challenge. The training dataset contains data from 100 anonymized patients, each has 2 MR scans, and totally 200 volumes. Testing dataset is not available now, so we evenly split the training dataset into non-overlapped two parts, one for training and the rest for validation. We calibrate the intensity of original MR volumes with CLAHE algorithm [12], and normalize them as zero mean and unit variance. Random rotation with probability 0.30 is used to augment training dataset. We resize the third dimension of volume to 32 during training and testing to facilitate consecutive 3D convolution and pooling.

Implementation Details. The proposed 3D FCN was implemented in *Tensorflow*, using 2 NVIDIA GeForce GTX TITAN X GPUs. *Code will be available upon publication*[1]. Given the limited memory in one GPU, we assign the down- and up-sampling paths to 2 GPUs. We update the weights of network with a Adam optimizer (batch size is 1). We utilize 1 auxiliary classifier with $\alpha_0 = 0.8$. Randomly cropped $96 \times 96 \times 24$ sub-volumes serve as input to train our network. We adopt sliding window and overlap-tiling stitching strategies to generate predictions for the whole volume. We further remove the small isolated connected components in final labeling result. As we take highly overlapped sliding windows, it takes about 50 s to process one volume.

Quantitative and Qualitative Analysis. We adopt 4 metrics to evaluate the proposed framework, including DSC, Jaccard Index, Precision and Recall. Sharing the basic configuration of deep supervision (DS) and transfer learning (TL), we conduct extensive experiments to compare performances of models based on cross-entropy (*DS+TL+Cross*), *wCross* (*DS+TL+wCross*) and *mDSC* (*DS+TL+mDSC*). We also compare with the *wCross* guided model without DS (*TL+mDSC*). Table 1 shows the quantitative evaluation on validation set (including 100 volumes). As shown, *wCross* improve the result of *Cross*, especially for the ignored class RVEDO, while *mDSC* based model performs better than all other 3 methods, which proves the significance of *mDSC* in balancing the training. We further provide explicit visualization of the segmentation

[1] https://github.com/xy0806/miccai17-acdc.

Table 1. Quantitative evaluation for ventricular structure segmentation in MR

Method	Metrics	Structures			
		RVEDO	LVEPI	LVEDO	*mean*
DS+TL+***Cross***	*DSC[%]*	76.14	77.24	83.28	78.87
	Jaccard[%]	64.83	66.19	74.77	68.60
	Precision[%]	92.03	83.75	91.40	89.06
	Recall[%]	69.27	75.50	79.77	74.85
DS+TL+***wCross***	*DSC[%]*	80.68	77.20	84.75	80.88
	Jaccard[%]	68.92	64.97	76.50	70.13
	Precision[%]	81.97	75.61	91.23	82.94
	Recall[%]	83.16	81.61	81.50	**82.09**
TL+***mDSC***	*DSC[%]*	75.54	79.54	85.73	80.27
	Jaccard[%]	64.23	68.34	77.55	70.06
	Precision[%]	90.32	86.14	92.11	89.52
	Recall[%]	68.98	77.31	82.56	76.28
DS+TL+***mDSC***	*DSC[%]*	80.62	80.37	85.80	**82.27**
	Jaccard[%]	69.51	69.15	77.59	**72.08**
	Precision[%]	91.87	85.49	92.09	**89.81**
	Recall[%]	74.83	79.10	81.60	78.51

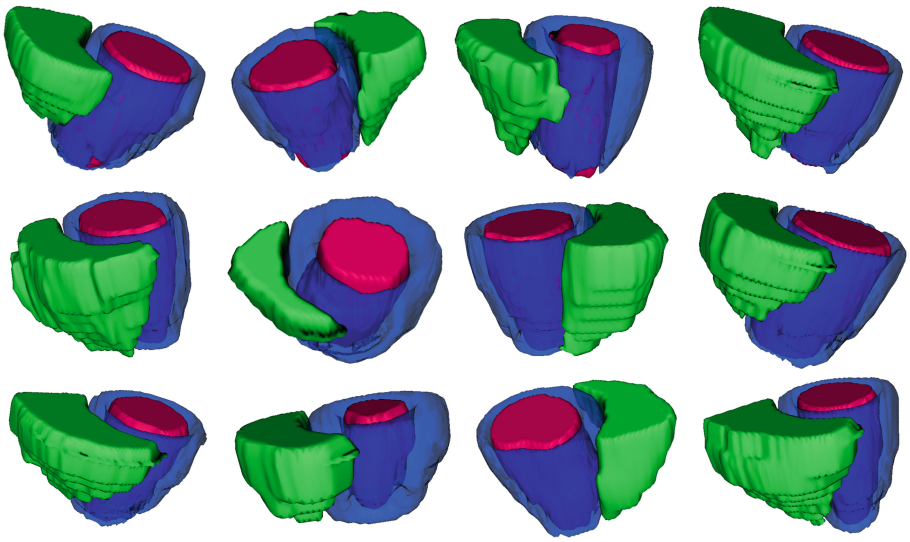


Fig. 3. Visualization of our segmentation results. Right ventricular endocardium, left ventricular epicardium and left ventricular endocardium are rendered with green, blue and red color. (Color figure online)

results for 3 ventricular structures in 12 testing MR volumes in Fig. 3. Our proposed method conquers varying shape and size, floating spatial relationship and presents promising performance.

4 Conclusions

We present a fully automatic framework for ventricular structure segmentation in MR volumes, which could be helpful for quantitative analysis of heart in clinic. Transfer learning and deep supervision are utilized to enhance the training of our customized 3D FCN. We propose the *mDSC* based loss function which performs better than traditional cross entropy and *wCross* based function in balancing the training procedure, and remarkable improvement are achieved on different metrics. Our proposed framework and class-balanced loss is general and can be explored in other segmentation tasks.

Acknowledgments. The work in this paper was supported by the grant from National Natural Science Foundation of China under Grant 61571304, a grant from Hong Kong Research Grants Council (Project no. GRF 14203115), a grant from the National Natural Science Foundation of China (Project No. 61233012) and a grant from Shenzhen Science and Technology Program (JCYJ20170413162256793).

References

1. Chen, H., Ni, D., Qin, J., et al.: Standard plane localization in fetal ultrasound via domain transferred deep neural networks. IEEE JBHI **19**(5), 1627–1636 (2015)
2. Çiçek, Ö., Abdulkadir, A., et al.: 3D U-net: learning dense volumetric segmentation from sparse annotation. arXiv preprint arXiv:1606.06650 (2016)
3. Dou, Q., Yu, L., et al.: 3D deeply supervised network for automated segmentation of volumetric medical images. Med. Image Anal. **41**, 40–54 (2017)
4. Kaus, M.R., von Berg, J., Weese, J., et al.: Automated segmentation of the left ventricle in cardiac MRI. Med. Image Anal. **8**(3), 245–254 (2004)
5. Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. In: NIPS, pp. 1097–1105 (2012)
6. Lee, C.Y., Xie, S., Gallagher, P., Zhang, Z., Tu, Z.: Deeply-supervised nets (2015)
7. Liu, Y., Captur, G., Moon, J.C., Guo, S., Yang, X., Zhang, S., Li, C.: Distance regularized two level sets for segmentation of left and right ventricles from cine-MRI. Magn. Reson. Imaging **34**(5), 699–706 (2016)
8. Long, J., Shelhamer, E., Darrell, T.: Fully convolutional networks for semantic segmentation. In: CVPR, pp. 3431–3440 (2015)
9. Milletari, F., Navab, N., et al.: V-net: fully convolutional neural networks for volumetric medical image segmentation. In: 2016 Fourth International Conference on 3D Vision (3DV), pp. 565–571. IEEE (2016)
10. Peng, P., Lekadir, K., Gooya, A., Shao, L., Petersen, S.E., Frangi, A.F.: A review of heart chamber segmentation for structural and functional analysis using cardiac magnetic resonance imaging. Magma **29**, 155 (2016)

11. Peters, J., Ecabert, O., Meyer, C., Schramm, H., Kneser, R., Groth, A., Weese, J.: Automatic whole heart segmentation in static magnetic resonance image volumes. In: Ayache, N., Ourselin, S., Maeder, A. (eds.) MICCAI 2007. LNCS, vol. 4792, pp. 402–410. Springer, Heidelberg (2007). https://doi.org/10.1007/978-3-540-75759-7_49
12. Pizer, S.M., Amburn, E.P., et al.: Adaptive histogram equalization and its variations. Comput. Vis. Graph. Image Process. **39**(3), 355–368 (1987)
13. Ronneberger, O., Fischer, P., Brox, T.: U-Net: convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) MICCAI 2015. LNCS, vol. 9351, pp. 234–241. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-24574-4_28
14. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556 (2014)
15. Tran, D., Bourdev, L., Fergus, R., Torresani, L., Paluri, M.: Learning spatiotemporal features with 3D convolutional networks. In: ICCV, pp. 4489–4497 (2015)
16. Tran, P.V.: A fully convolutional neural network for cardiac segmentation in short-axis MRI. arXiv preprint arXiv:1604.00494 (2016)
17. Xie, S., Tu, Z.: Holistically-nested edge detection. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 1395–1403 (2015)
18. Yu, L., Yang, X., Qin, J., Heng, P.-A.: 3D FractalNet: dense volumetric segmentation for cardiovascular MRI volumes. In: Zuluaga, M.A., Bhatia, K., Kainz, B., Moghari, M.H., Pace, D.F. (eds.) RAMBO/HVSMR -2016. LNCS, vol. 10129, pp. 103–110. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-52280-7_10
19. Zhuang, X.: Challenges and methodologies of fully automatic whole heart segmentation: a review. J. Healthc. Eng. **4**(3), 371–408 (2013)
20. Zhuang, X., Rhode, K.S., Razavi, R.S., Hawkes, D.J., Ourselin, S.: A registration-based propagation framework for automatic whole heart segmentation of cardiac MRI. IEEE Trans. Med. Imaging **29**(9), 1612–1625 (2010)
21. Zotti, C., Luo, Z., et al.: Novel deep convolution neural network applied to MRI cardiac segmentation. arXiv preprint arXiv:1705.08943 (2017)

Automatic Segmentation of LV and RV in Cardiac MRI

Yeonggul Jang[1], Yoonmi Hong[2(✉)], Seongmin Ha[2], Sekeun Kim[2], and Hyuk-Jae Chang[2,3]

[1] Brain Korea 21 PLUS Project for Medical Science, Yonsei University, Seoul, South Korea
jygl722@gmail.com
[2] Integrative Cardiovascular Imaging Research Center, Yonsei University College of Medicine, Seoul, South Korea
yoonmhong@gmail.com
[3] Division of Cardiology, Severance Cardiovascular Hospital, Yonsei University College of Medicine, Seoul, South Korea

Abstract. Automatic and accurate segmentation of Left Ventricle (LV) and Right Ventricle (RV) in cine-MRI is required to analyze cardiac function and viability. We present a fully convolutional neural network to efficiently segment LV and RV as well as myocardium. The network is trained end-to-end from scratch. Average dice scores from five-fold cross-validation on the ACDC training dataset were 0.94, 0.89, and 0.88 for LV, RV, and myocardium. Experimental results show the robustness of the proposed architecture.

Keywords: Cardiac segmentation · Convolutional neural network
Cardiac MRI · Automated Cardiac Diagnosis Challenge

1 Introduction

Cardiac image segmentation plays an important role for the diagnosis of cardiac diseases, quantification of volume, and image-guided interventions [1]. Due to the advancement of the echocardiogram, Computed Tomography (CT), Magnetic Resonance Imaging (MRI), quantitative and qualitative measurements from the cardiac imaging can be proceeded easily. Accurate segmentation of Left Ventricle (LV) and Right Ventricle (RV) are particularly valuable for the extraction of ventricular function information such as stroke volume and ejection fraction. In clinic, MRI becomes a reference modality to evaluate the cardiac function.

Various methods are developed to automate the segmentation of the ventricles in cardiac MRI [2–4]. Recently, the advancement of the deep learning shows high performances in object detection, recognition, as well as segmentation. There are many attempts to use deep learning technique, especially Convolutional Neural Networks

Y. Jang and Y. Hong—both authors contributed equally.

M. Pop et al. (Eds.): STACOM 2017, LNCS 10663, pp. 161–169, 2018.
https://doi.org/10.1007/978-3-319-75541-0_17

(CNN), in cardiac image segmentation problems. 2D CNN was applied in cardiac images for LV segmentation with auto-encoder [5]. Also, 3D CNN is applied to detect coronary calcium in gated cardiac CT Angiography [6]. Recurrent Neural Net-work (RNN) is applied for LV segmentation in multi-slice MR images [7]. Multi-scale convolutional deep belief network is proposed to estimate bi-ventricular volume in cardiac MRI [8]. The largest challenge in the cardiac imaging was the 2015 Kaggle Data Science Bowl that aims to automatically measure end-systolic and end-diastolic volumes in cardiac MRI [9].

In this paper, we introduce a fully automated segmentation method for LV, LV myocardium, and RV in cine MRI. Our Architecture is based on the M-net where we only use 2D CNNs without 3D-to-2D converter [10]. The M-net architecture proposed in [10] has 3D filtering layer to utilize 3D information, and the authors applied the architecture to Brain MR images with slice thickness 1 mm to 1.5 mm. However, we observe that our training dataset has relatively large slice thickness from 5 mm to 10 mm. The experimental results also show that utilizing 3D information degrades the performance.

This paper is organized as follows. In Sect. 2, we present the proposed architecture as well as the pre-processing method. We show the experimental results from five-fold cross-validation using given 100 training dataset in Sect. 3. Conclusions and discussions are given in Sect. 4.

2 Methods

2.1 Dataset

The training datasets come from clinical exams acquired at the University Hospital of Dijon (Dijon, France), Automated Cardiac Diagnosis Challenge (ACDC) in MICCAI challenge 2017. This datasets contain cardiac short-axis MRI images with the corresponding manual reference images of LV, LV myocardium, and RV for 100 patients. Each case contains all phases of 4D images; however, manual reference images are provided only in ED (end-diastole) and ES (end-systole) phases.

The dataset is divided into 5 evenly distributed subgroups: normal case (NOR), heart failure with infarction (MINF), dilated cardiomyopathy (DCM), hypertrophic cardiomyopathy (HCM), and abnormal right ventricle (ARV).

2.2 Preprocessing

We observe that the MRI datasets provided by ACDC challenge have a wide range of in-plane dimensions from 154×224 to 428×512. We re-scale all dataset to 256×256 by fitting maximum size of X and Y to 256 and padding residual regions with minimum value of each image. Also, the 16-bit MRI datasets have a wide range of voxel intensity that results from different scanner types or acquisition protocols. This variety can affect the performance of a segmentation model. We normalize the voxel intensity of each image by subtracting its mean then dividing it by its standard deviation.

2.3 Architecture

The provided datasets in this challenge have a large slice-thickness (5 to 10 mm) and the connectivity between adjacent slices is insufficient in this case. In short, 3D information is not considered necessarily since it can impede generalization of a model. We propose an end-to-end fully convolutional network (FCN) architecture, which is based on M-net [10]. This architecture is inspired by U-net [11]. The proposed FCN architecture has the same layers with M-net excluding the 3D convolution filter. Figure 1 illustrates our proposed FCN architecture.

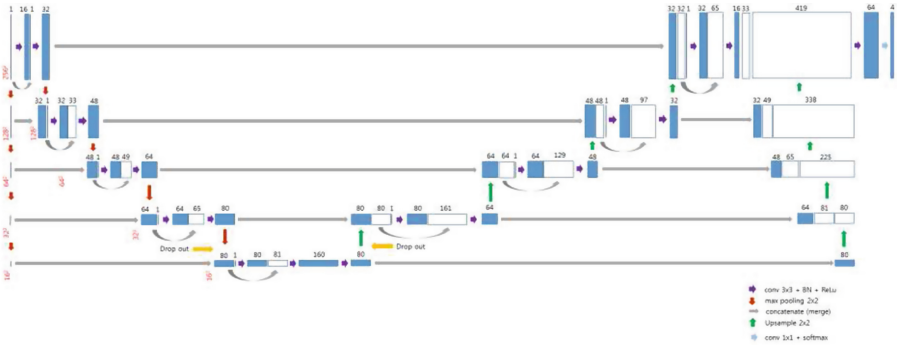


Fig. 1. The proposed FCN architecture

Our architecture has two main paths in common with M-net: Contraction and Expansion paths. Contraction path has 5 cascade steps. Each step in this path has 2 convolution layers of size 3×3 and max-pooling layer of size 2×2. This path reduces the size of input by half and allows network to capture contextual information. As input size is reduced by max-pooling, the number of filters gradually increases to avoid bottleneck (information loss). Expansion path has the symmetric steps and layers but for replacing max-pooling with up-sampling to double the size of input. For precise localization, previous feature maps are concatenated to the corresponding next feature maps. The final layer is processed by 1×1 convolution layer with 4 channels (Background, RV endocardium, LV myocardium, and LV endocardium) and pixel-wise softmax which gives the probability of 4 classes to every pixel. The final segmentation labels are assigned to the classes with maximum probability for every pixel. Batch normalization layers are applied after each convolution layer before Relu activation. Dropout with probability 0.5 is applied to contraction and expansion only once, respectively.

To resolve bad training of a certain class due to the class imbalance (especially LV-Myo), we use weighted cross-entropy as loss function. The weight of loss function is defined based on the number of voxels in a certain class.

3 Experimental Results

3.1 Implemented Details

We divide the ACDC datasets which have ED and ES volumes of 100 patients into 80 training sets and 20 test sets to train our FCN and test its performance with five-fold cross validation. Five-fold cross validation was performed with the following details: (1) select 4 volumes sequentially in each of 5 disease classes, which gives 20 volumes in total, (2) The selected 20 volumes and the remaining 80 ones are considered as test sets and training sets, respectively, (3) Training and Testing, (4) conduct the whole process iteratively five times for ED and ES, respectively.

Considering each image consists of approximately 10 slices, it is not enough to train a model without overfitting. Also, a CNN architecture is not invariant to rotation though it is partially invariant to translation. Therefore, we perform rotation transformation from −60° to 60° at uniform intervals of 15° to augment the training datasets. For post-processing, we apply morphological operations to fill the small gap or to remove small volumes. We also apply convex hull to remove concavities only for LV.

The proposed FCN was trained on NVIDIA TITAN X, with 12 GB of RAM for 150 epochs through the training set of about 6,800 images (80 volumes), which took 18 h for training. The FCN was implemented by Tensorflow r0.11 and trained using RMSprop Optimizer with following hyper parameters: learning rate = 10^{-3}, decay = 0.9, momentum = 0.0, and epsilon = 10^{-10}.

3.2 Results and Quantitative Analysis with Other Methods

The segmentation performance is evaluated with the mean Dice Similarity Coefficient (DSC) and Hausdorff distance (HD). Let S_R and S_{GT} be the segmentation result and ground truth, respectively. The $\mathrm{DSC}(S_R, S_{GT})$ is defined as $\frac{2|S_R \cap S_{GT}|}{|S_R| + |S_{GT}|}$, where 0 signifies the zero overlap between the ground truth and the derived segmentation result, and 1 signifies the complete overlap between ground truth and segmentation result in both the foreground and background. The $\mathrm{HD}(S_R, S_{GT})$ is defined as $\max\left(\max_{x \in C_R} \min_{y \in C_{GT}} d(x,y), \max_{x \in C_{GT}} \min_{y \in C_R} d(x,y)\right)$, where C_R and C_{GT} are contour point sets of S_R and S_{GT}, respectively, and $d(x,y)$ is the distance between two points. It is the longest distances of all which are measured from a contour point in one to the closest contour point in the other.

We compared the segmentation results of the proposed FCN with U-net, and U-net with 3D to 2D converter. U-net with converter includes 3D to 2D convolution layer in front of existing U-net layers. It takes 2n + 1 slices as input image, a central slice and its neighboring 2n slices, for using 3D context. In this paper, the number of neighboring slices n was empirically assigned to one in the light of large slice thickness.

For fair comparison, we did not consider data augmentation, and the comparison is conducted only for one cross-validation subset at ED phase. The mean DSC values of each structure for three architectures are listed in Table 1. Based on these results, it is shown that the proposed FCN architecture segments three structures of interest on MRI images slightly better than two other architectures.

Table 1. Comparison of segmentation results of the proposed FCN with U-net and U-net with 3D to 2D converter.

	U-net			U-net with converter			The proposed FCN		
	RV	LV-Myo	LV	RV	LV-Myo	LV	RV	LV-Myo	LV
DSC	0.900	0.878	0.967	0.878	0.829	0.959	0.908	0.883	0.962

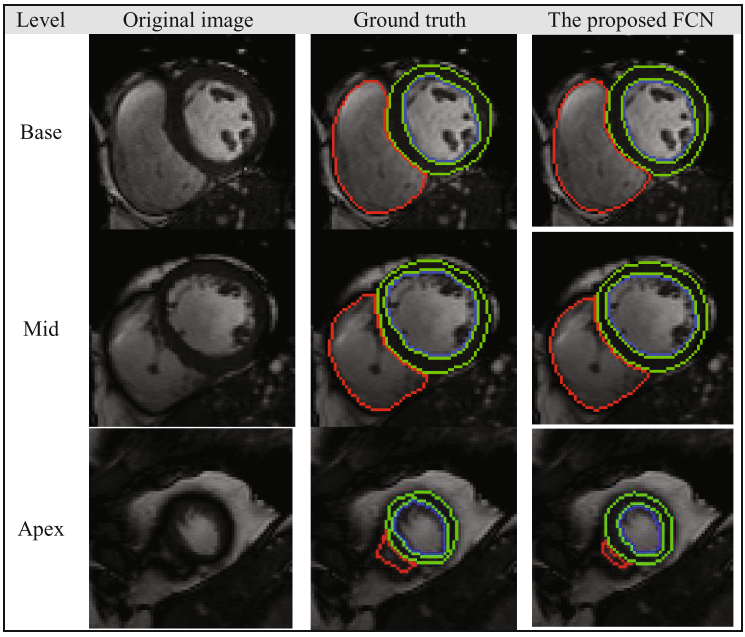
It should be noted that U-net with 3D to 2D converter for using 3D context produced lower mean DSC values than two other architectures. It is due to relatively large slice thickness of the provided datasets or image-shift by different breath-hold during acquisition. As previously mentioned, 3D information is not considered necessarily in images with thick slices and it can impede generalization of a model.

Finally, we trained the proposed FCN with the augmented datasets by rotation and gained the increased segmentation results for RV, LV-Myo and LV. Table 2 shows mean DSC values and Hausdorff distances of the proposed FCN with augmentation for each subset of five-fold cross validation and its average. There is major improvement in DSC for RV and minor improvement for LV-myo and LV on the same datasets (ED CV#5). It is noted that average Hausdorff distances for RV are higher than LV-Myo and LV due to many false positives especially in basal slices. We also evaluated group-based cross-validation, and the results are summarized in Table 3.

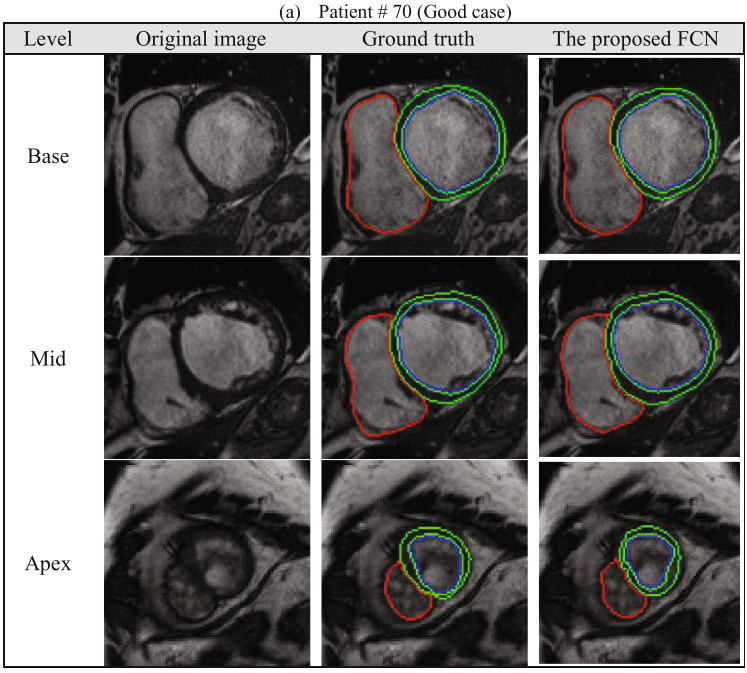
Table 2. Cross-validation (CV) results of our model on the 100 cases (training datasets are 80, test datasets are 20). Values correspond to the mean and standard deviation.

		RV		LV-Myo		LV	
		ED	ES	ED	ES	ED	ES
DSC	CV#1	0.934 ± 0.03	0.837 ± 0.09	0.880 ± 0.02	0.903 ± 0.02	0.960 ± 0.02	0.924 ± 0.05
	CV#2	0.924 ± 0.03	0.853 ± 0.07	0.866 ± 0.03	0.882 ± 0.04	0.959 ± 0.01	0.919 ± 0.04
	CV#3	0.917 ± 0.03	0.851 ± 0.10	0.861 ± 0.05	0.884 ± 0.04	0.956 ± 0.02	0.898 ± 0.09
	CV#4	0.918 ± 0.06	0.843 ± 0.09	0.853 ± 0.04	0.873 ± 0.04	0.958 ± 0.02	0.911 ± 0.08
	CV#5	0.932 ± 0.04	0.891 ± 0.04	0.892 ± 0.03	0.900 ± 0.03	0.970 ± 0.02	0.920 ± 0.05
	Avg	0.925 ± 0.04	0.855 ± 0.08	0.870 ± 0.04	0.888 ± 0.04	0.961 ± 0.02	0.915 ± 0.06
	Total	0.890 ± 0.07		0.879 ± 0.04		0.938 ± 0.05	
HD (mm)	CV#1	12.49 ± 5.87	15.54 ± 5.05	9.23 ± 5.93	8.68 ± 3.91	7.29 ± 4.36	7.15 ± 4.27
	CV#2	14.63 ± 5.55	16.20 ± 5.89	10.88 ± 9.02	11.86 ± 6.78	6.94 ± 4.50	9.52 ± 4.10
	CV#3	15.61 ± 7.62	15.34 ± 5.57	9.52 ± 5.27	10.49 ± 3.80	8.73 ± 5.66	8.78 ± 4.69
	CV#4	15.30 ± 7.60	16.98 ± 8.80	9.68 ± 6.21	11.83 ± 6.28	7.49 ± 8.29	9.33 ± 5.94
	CV#5	11.82 ± 4.91	14.64 ± 7.49	7.84 ± 6.50	9.22 ± 5.36	3.99 ± 2.83	8.90 ± 6.55
	Avg	12.60 ± 6.02	14.78 ± 6.36	9.47 ± 6.64	10.05 ± 5.27	6.27 ± 4.74	8.28 ± 4.67
	Total	13.69 ± 6.30		9.76 ± 6.02		7.27 ± 4.83	

Figure 2 shows segmentation results for three different levels (position) of two sample volumes aligned by short axis of heart. Figure 2(a) and (b) represent a good case without any and a case with LV trabeculations and partial volume effect at apical level, respectively. The segmentation results by our FCN are coterminous with the provided ground truth for both two cases as shown in Fig. 2.



(a) Patient # 70 (Good case)



(b) Patient#17 (with LV trabeculations at apex)

Fig. 2. The segmentation results for three different levels (base, middle and apex) of two sample slices aligned by short axis of heart. (a) and (b) represent a good case without any and a case with LV trabeculations and partial volume effect at apical level, respectively. Red: Right Ventricle, Blue: Left Ventricle, Green: Myocardium. (Color figure online)

Table 3. Group-base analysis results of our model on the 100 cases with five-fold cross-validation (training datasets are 80, test datasets are 20).

Group	DSC			HD (mm)		
	RV	LV-Myo	LV	RV	LV-Myo	LV
DCM	0.898 ± 0.07	0.879 ± 0.03	0.966 ± 0.01	15.37 ± 8.28	8.84 ± 4.72	5.83 ± 2.58
HCM	0.867 ± 0.09	0.903 ± 0.03	0.889 ± 0.08	12.89 ± 7.43	11.98 ± 6.73	9.57 ± 5.15
MINF	0.865 ± 0.08	0.863 ± 0.04	0.952 ± 0.02	15.06 ± 6.71	9.40 ± 4.25	7.02 ± 3.32
NOR	0.909 ± 0.07	0.896 ± 0.03	0.945 ± 0.04	10.86 ± 4.80	7.55 ± 4.59	6.31 ± 4.12
ARV	0.907 ± 0.04	0.857 ± 0.04	0.930 ± 0.05	13.23 ± 4.85	11.47 ± 8.18	7.96 ± 6.98

We note that the mean DSC values for RV and LV on ES are relatively degraded compared with ED as shown in Table 2. The issues about it and LV trabeculations will be discussed on Sect. 4 in detail.

The CPU and GPU run times to segment one volume using the proposed FCN are approximately 8.09 s and 0.62 s, respectively.

4 Conclusion and Discussion

In this paper, we proposed a new FCN architecture for three structures (RV, LV-Myo and LV) segmentation on MRI images. It has the same layers as M-net excluding 3D-to-2D converter layer. As we observe that the datasets provided by ACDC challenge have large slice thickness and image-shift due to different breath-hold during acquisition, we think that considering 3D information can impede generalization of a model. Therefore, the proposed FCN architecture combines U-net architecture and skip connections of M-net to learn better features. Experimental results on the provided datasets showed that the proposed FCN has better performance for RV, LV-Myo and LV segmentation than other current state-of-the-art models. It is well known that CNN is not invariant to rotation. We found that the orientation of the provided datasets varies approximately from $-60°$ to $60°$ by volume. For this reason, we applied data augmentation for rotation and the DSC values were slightly improved especially for RV, which has variant shape to rotation unlike LV-Myo and LV.

Most segmentation errors occur at basal and apical slices of volumes aligned by short-axis of heart as shown in Figs. 2 and 3. Usually, LV contours are approximately delineated as an ellipse which includes LV Trabeculations. These are located near boundary at basal and middle level, but near center at apical level. As shown in Fig. 2(b), it seems that there exist two different structures in LV cavity, which become faint with partial volume effect. These reasons make it difficult to segment structures of interest at apical level.

On the other hand, there are some cases without RV labels for ground truth at basal level as shown in Fig. 3. These mismatch between RV labels by our FCN and ground truth at basal level causes performance decreases in DSC and Hausdorff measures. RV and LV are connected to pulmonary artery and ascending aorta, respectively, and these connected regions are usually observed at basal level. Although regions which belong to RV or LV still remain at basal level, these regions are occasionally not included in

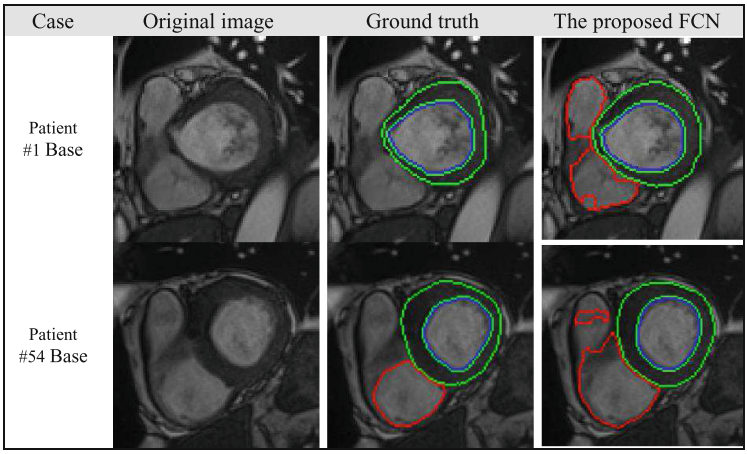


Fig. 3. The mismatch between RV labels by our FCN and ground truth at basal level.

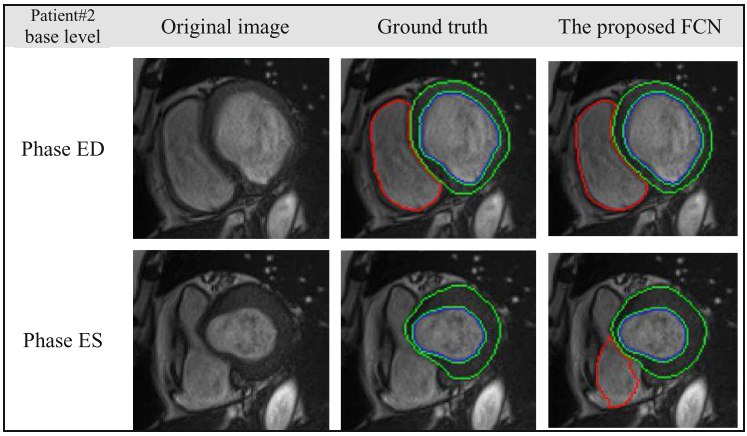


Fig. 4. The different shapes of structures of interest on ED and ES at the same basal level.

clinical setting. It is very difficult to make the distinction on a static slice, although the regions belonging to RV look similar in the first and second rows in Fig. 3. Also, it can be explained why the mean DSC values for RV and LV on ES are relatively poor from the similar reason. In case of patient#2, at the same basal level of ED and ES, these connected regions are observed only on the image on ES as shown in Fig. 4. Thus, it needs to consider for the additional processes at basal level.

In the future, we will apply further sophisticated post-processing algorithms using the probability maps for structures of interest produced by the proposed FCN as well as level classification methods for additional performance improvement at basal and apical level.

Acknowledgement. This work was supported by Institute for Information & communications Technology Promotion (IITP) grant funded by the Korea government (MSIT) (2017-0-00255, Autonomous digital companion framework and application).

References

1. Kang, D., et al.: Heart chambers and whole heart segmentation techniques. J. Electr. Imaging **21**(1), 010901-1–010901-16 (2012)
2. Petitjean, C., et al.: A review of segmentation methods in short axis cardiac MR images. Med. Image Anal. **15**(2), 169–184 (2011)
3. Bai, W., et al.: A probabilistic patch-based label fusion model for multi-atlas segmentation with registration refinement: application to cardiac MR images. IEEE Trans. Med. Imaging **32**(7), 1302–1315 (2013)
4. Bai, W., et al.: Multi-atlas segmentation with augmented features for cardiac MR images. Med. Image Anal. **19**(1), 98–109 (2015)
5. Avendi, M., et al.: A combined deep-learning and deformable-model approach to fully automatic segmentation of the left ventricle in cardiac MRI. Med. Image Anal. **30**, 108–119 (2016)
6. Wolterink, J.M., Leiner, T., Viergever, M.A., Išgum, I.: Automatic coronary calcium scoring in cardiac CT angiography using convolutional neural networks. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) MICCAI 2015. LNCS, vol. 9349, pp. 589–596. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-24553-9_72
7. Poudel, R.P.K., Lamata, P., Montana, G.: Recurrent fully convolutional neural networks for multi-slice MRI cardiac segmentation. In: Zuluaga, M.A., Bhatia, K., Kainz, B., Moghari, M.H., Pace, D.F. (eds.) RAMBO/HVSMR -2016. LNCS, vol. 10129, pp. 83–94. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-52280-7_8
8. Zhen, X., et al.: Multi-scale deep networks and regression forests for direct bi-ventricular volume estimation. Med. Image Anal. **30**, 120–129 (2016)
9. Litjens, G., et al.: A survey on deep learning in medical image analysis. arXiv:1702.05747 (2017)
10. Mehta, R., et al.: M-Net: a convolutional neural network for deep brain structure segmentation. In: International Symposium on Biomedical Imaging, pp. 437–440 (2017)
11. Ronneberger, O., Fischer, P., Brox, T.: U-Net: convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) MICCAI 2015. LNCS, vol. 9351, pp. 234–241. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-24574-4_28

Automatic Multi-Atlas Segmentation of Myocardium with SVF-Net

Marc-Michel Rohé(✉), Maxime Sermesant, and Xavier Pennec

Université Côte d'Azur, Inria, Sophia-Antipolis, France
{marc-michel.rohe,maxime.sermesant}@inria.fr

Abstract. Segmentation of the myocardium is a key step for image guided diagnosis in many cardiac diseases. In this article, we propose an automatic multi-atlas segmentation framework which relies on a very fast registration algorithm trained with convolutional neural networks. The speed of this registration method allows us to use a high number of templates in the multi-atlas segmentation while remaining computationally tractable. The performance of the propose approach is evaluated on a dataset of 100 end-diastolic and end-systolic MRI images of the STACOM 2017 Automated Cardiac Diagnosis Challenge (ACDC).

1 Introduction

Both ventricles play a fundamental role for the circulation of oxygenated blood to the body. To evaluate their functions, clinicians rely on indices that are based on geometrical measurements of regions of the hearts [2,3]: the blood pool volume, the wall thickness of the myocardium, or the myocardial mass. These indices are usually estimated using a manual segmentation of the contours of the myocardium and the blood pool. However, this task is very time-consuming, requires clinical experience and is prone to large inter-rater variability [8] which will impact the measures derived. For these reasons, there is an important clinical need to define segmentation methods that are fast and fully automatic.

Main challenges to develop such a fully automated segmentation method from medical images are the large variability of the shape of the myocardium, the artifacts and the noise in the images, and the difficulty to chose the most basal slice to segment. In this paper, we propose a method based on multi-atlas segmentation (MAS), an extension of atlas-based methods with multiple templates [7]. With respect to the state-of-the-art MAS methods, our contribution rely on the use of a very fast and robust registration algorithm [6] specifically trained to perform inter-patient heart registration. This registration method leverages recent advances in the field of convolutional neural networks and uses a machine learning approach to the task of registration. The speed of the registration method used paves the way for the use of a high number of atlas which increases the range of anatomy that can be predicted with such a model. One can expect that, for each target geometry we want to segment, at least one or multiple atlases will have a similar shape.

M. Pop et al. (Eds.): STACOM 2017, LNCS 10663, pp. 170–177, 2018.
https://doi.org/10.1007/978-3-319-75541-0_18

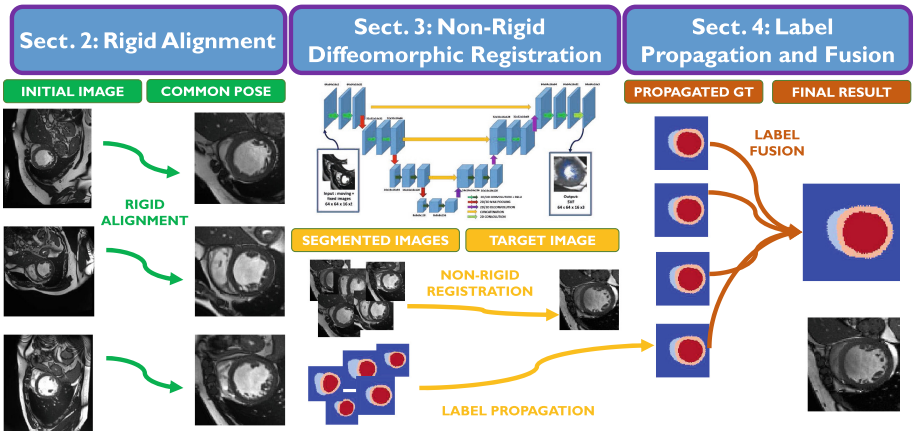


Fig. 1. Overview of our proposed method. Firstly, all the images are aligned in a common pose by computing a $2D$ transformation with 2 landmarks: the center of the LV and the RV. In Sect. 3, we present our non-rigid registration method that is used to register each of the template images with the target image. Finally, in Sect. 4, the labels are propagated and fused to get our final estimation.

The rest of this article describes our segmentation method step by step. First, as a pre-processing step, we detect the location and the orientation of the heart in the image. We perform this task thanks to a CNN trained to detect two landmarks. One landmark gives the position of the heart and the other one is used to get the orientation. Using these landmarks, the target image is rigidly aligned with respect to the database of templates with ground-truth segmentation. This is a mandatory step before applying non-rigid diffeomorphic registration. In the following section, we present and adapt the SVF-Net registration method [6] to the specific data of the challenge. This registration algorithm performs the non-rigid registration part of our MAS method. Then, we define a method to fuse the label of the estimation from the different templates using specific weights. The pipeline is schematically represented in Fig. 1. Finally, we evaluate our proposed method on the training dataset of the Automatic Cardiac Diagnosis Challenge held in STACOM 2017.

2 Rigid Alignment by Landmarks Detection

For the images of the training database, one can easily define a common pose and alignment by using the barycenter of the LV and the RV computed with the segmentation information. The center of the LV is used to defined a region of interest of the heart while we use the center of the RV to get the axis of the LV to the RV defining orientation of the heart in the X/Y plan. A $2D$ rotation on this plan is applied to all images of the atlases (see Fig. 3) so that RV and LV are aligned.

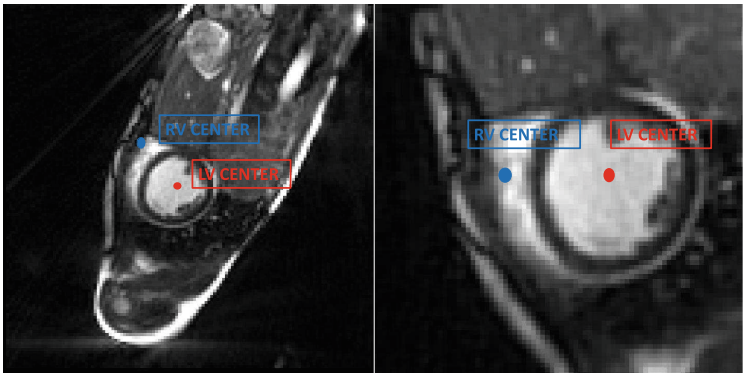


Fig. 2. (Left): Image from a slice of a 3D acquisition. (Right): Same image after pre-processing (cropping the ROI and $2D$ rotation around the Z axis to align with the pose of the atlases). Most of the background has been removed from the image and only the important information remains. The pose of the LV/RV ventricles and the heart position is aligned with the atlas making the registration step easier. To do so, a landmark corresponding to the LV and one corresponding to the RV are detected.

For the target image, for which we do not have the ground truth segmentation, we need to define a method to detect these 2 landmarks in order to perform the same pre-processing that was done with the atlases. Inspired by recent works [1], we propose to use heatmaps regression for landmark detection to detect both landmarks (LV and RV centers) using CNNs. In particular, the work of [5] investigates the idea of directly estimating multiple landmark locations from 3D image using a single fully-convolutional CNN, trained in an end-to-end manner to regress heatmaps for landmarks instead of absolute landmark coordinates. This approach has multiple advantages. It is a learning-based method so that we can efficiently leverage our large database of 200 images with ground truth landmarks derived from segmentations. Also, the prediction of heatmaps is an easier task for a CNNs than the prediction of absolutes coordinates of landmarks, as the localization of the responses in the successive layers can be directly used to predict the heatmap (Fig. 3).

For all the images of the training set, the heatmap of a landmark with position p is defined on the image grid as:

$$H_p(\mathbf{x}) = \exp\left(-\frac{\|\mathbf{x} - p\|^2}{\sigma^2}\right),$$

where σ is the decaying factor of the heatmap. Examples of such heatmap for both the RV center and LV center landmarks are shown in Fig. 3. An CNN U-Net architecture similar to the one presented in [6] is used. The input of the network is the complete image and the output is the predicted heatmap. At test time,

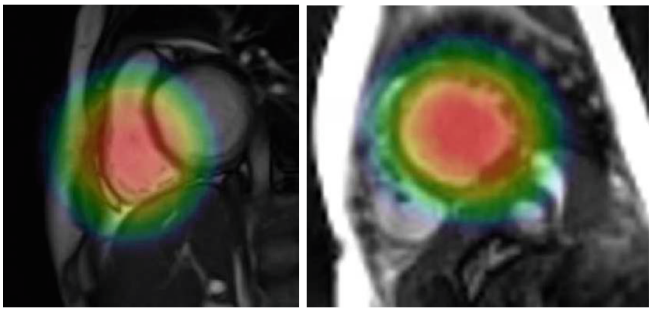


Fig. 3. Heatmaps for both landmarks (RV center: left image, LV center: right image).

the landmark position p is inferred by computing the point that minimizes the least-square distance to the predicted map H_{pred}:

$$p = \text{argmin} \sum_{\mathbf{x}} \|H_{pred}(\mathbf{x}) - H_p(\mathbf{x})\|^2.$$

3 Non-rigid Diffeomorphic Registration with SVF-Net

In [6], the authors propose to use Fully-Convolutional Neural Networks (CNN) (illustrated in [6]) to predict directly the deformation from a pair of images. With respect to traditional patch-based approaches, this fully convolutional architecture has the benefit to be faster at test time (registration taking less than 6 s/30 ms with CPU/GPU) as the whole image is passed in a single stream to the network instead of passing multiple streams corresponding to the patches of the image in a sliding-window approach. The computational efficiency of this method makes it particularly suitable for our MAS approach paving the way for the use of a large number of templates.

To train this kind of CNN, one needs to compute ground truth registrations from pairs of segmented images. In [6], reference deformations are computed using the result of a registration algorithm previously run on pairs of segmented shape. With respect to the use of the result of the registration on the images, this method to define reference deformations tends to be more robust as the segmentations can be corrected manually. In the dataset provided by the challenge, the segmentations were given in the form of binary masks rather than segmented shapes. Therefore, we adapt the method and perform pair-wise registration of the binary masks using the *LCC-log demons algorithm* [4]. The deformation fields are computed with an iterative optimization run successively on each of the 3 regions of interest.

4 Label Fusion Method

We consider a database of M training images I_j, $j = 1, ..., N$, or atlases for which we have ground truth segmentations (which are images with 3 channels

corresponding to the 3 regions of interest). We perform the registration of each of these images with respect to the target image I with the method described previously. The resulting deformation field is applied to the binary mask of the atlas M_j and we need to define a method to combine these estimations $\hat{M}_j$ to get $\hat{M}_j$: the estimation of the segmentation of the target image.

A straight-forward method to combine the estimations is by majority voting. In this work, we chose to use varying decision weights in order to combine these estimations using a local assessment of the registration success. Therefore, these weights will have a higher value for registrations in which we have a higher confidence. This confidence is evaluated at each point of the image using 3 different metrics. The first one is the *Local Correlation Coefficient* (LCC) [4] which locally estimates the similarity between the voxel intensities of the images. The second metric is the square norm of the displacement of the transformation because we consider that we have stronger confidence in small transformations (corresponding to similar images) than in large transformations. Finally, the last metric corresponds to the Jacobian of the deformation field, meaning we give more confidence to smooth displacement fields over non-regular ones.

To get the function $d_j(x)$ representing the local assessment of the registration success (between the target image and atlas j), these 3 metrics are combined linearly. The coefficients are learned on the training set as to minimize the square difference of the distance versus the ground truth labeling error. Then, we define the local weight of each point of each atlas as a function of the local distance with a kernel σ_{metric}. Furthermore, we also smooth the weights spatially with a kernel $\sigma_{spatial}$ in order to ensure spatial consistency of the estimations. Finally, the weights are normalized in order to sum to 1:

$$\tilde{\omega}_j(x) = G_{\sigma_{spatial}} \star \exp\left(-d_j(x)/\sigma^2_{metric}\right),$$

$$\omega_j(x) = \frac{\tilde{\omega}_j(x)}{\sum_k \tilde{\omega}_k(x)}.$$

The kernel σ_{metric} corresponds to the confidence on our estimation of the local distance $d(\hat{p}^j_k)$. Large values for the kernel corresponds to small confidence. At the limit when σ_{metric} becomes large enough, all the weights become equal and we get the simple method of averaging the labels and the local distance does not have any impact. The spatial kernel $\sigma_{spatial}$ is to ensure spatial consistency of the resulting segmentation. Large values of $\sigma_{spatial}$ will make the weights more global whereas small values will make them more local.

5 Results and Discussion

Our method is applied to the database of the STACOM 2017 Automated Cardiac Diagnosis Challenge (ACDC). This challenge provides the community with a comprehensive set of $3D$ cine-MRI images (100 patients divided in 5 groups: 4 pathological plus 1 healthy control groups) acquired at the University Hospital of Dijon. For each of these patients, manual expert segmentation was performed for

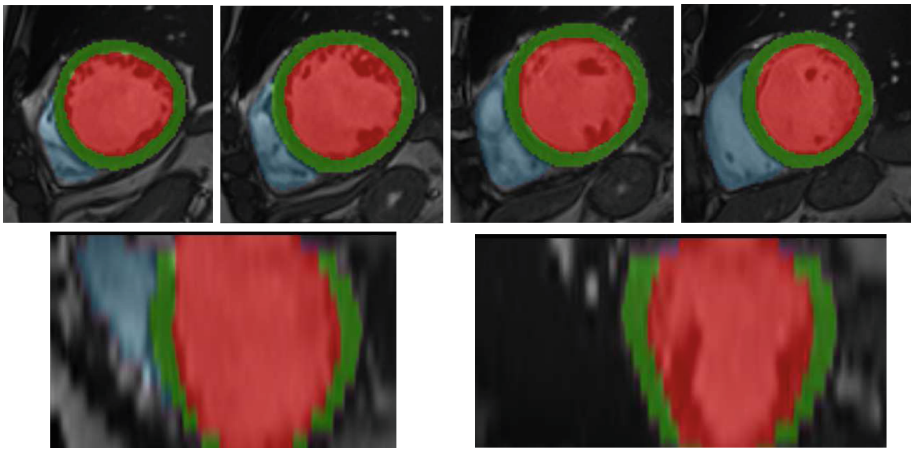


Fig. 4. Example of a segmentation using proposed method. (Top): short axis view with 4 slices. (Bottom): 4CH and 2CH axis views.

the end-diastolic (ED) and end-systolic (ES) frame to trace the LV endocardial and epicardial borders giving 3 regions of interest: RV cavity, LV cavity and left ventricle myocardium. We use 80 patients to train the CNN networks and 20 patients as testing test to evaluate the accuracy of the segmentation.

Training. Reference deformations are computed for each of the possible combination of pairs of our dataset of 80 ED and 80 ES images for a total of $160^2 = 40,000$ reference deformations which took 2 min per pair on a single core CPU (a cluster of CPU was used). Because our method already gives us a large database of ground truth data, we only use small translations in the X and Y axis for data augmentation (this also improves the robustness of the learned network over slight rigid misalignment of both images). For the loss function, we used the sum of squared difference between the predicted SVF parametrization and the ground truth. We implement the network using Tensorflow[1] and we train it on a NVIDIA TitanX GPU with $100,000$ iterations using the ADAM solver which took approximately 24 h. The CNN to detect the position and the orientation of the heart is trained similarly using these 160 images. The coefficients of the function $d_j(x)$ are then learned using leave-one-out for each of the testing images (19 testing images are used to estimate the optimal coefficients that is applied to the other image). Finally, σ_{metric} and $\sigma_{spatial}$ are estimated with a trial and error approach to balance accuracy and smoothness of the result.

Testing. We evaluate the method on the 20 ED and ES testing images. For each of these images, we perform the registration using SVF-net with respect to the 80 training images and fuse the warped labels using the method described

[1] www.tensorflow.org.

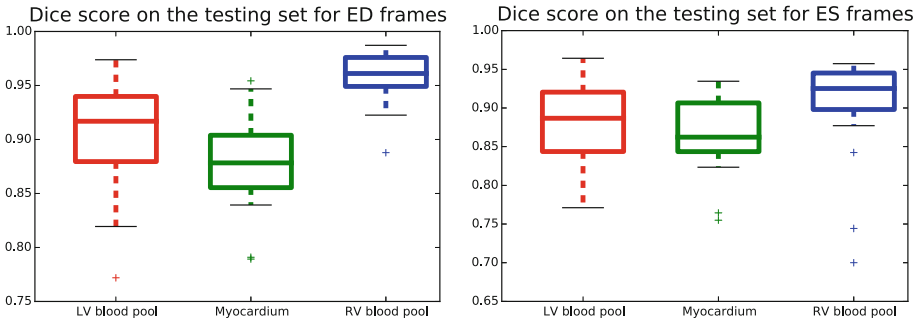


Fig. 5. Results on the 20 patients used for testing. (Left): dice scores for the ED frames. (Right): dice scores for the ES frames. The three different regions of interests are shown.

in Sect. 4 to get the final estimation of the label corresponding to the regions of interest. An example of the segmentation with our method can be seen in Fig. 4. One can see that our method produces a segmentation that is smooth and spatially consistent in the Z axis. When compared qualitatively to ground truth segmentations, most of the differences were seen at the base, where our method did not always segment the same basal slice as the ground truth. Additional work could be done to our method so that we come up with a more consistent evaluation of the first slice to segment. Finally, we have evaluated quantitatively the results using dice scores in Fig. 5. As expected, dice scores for LV blood pool (median of 0.97/0.93 for the ED/ES frames) tend to be higher than for the two other regions of interest with myocardium at 0.87/0.87 and RV blood pool at 0.92/0.89 for ED/ES. These results are promising and need to be confirmed and compared to other state-of-the-art methods on the final testing set of the challenge.

Extension to Classification. The function d_j defined in Sect. 4 represents the distance between pairs of myocardium shapes. This distance can be used, together with more advanced statistics of deformation fields, to perform classification of the target patient. For example by looking for the closest patients with respect to this distance or by running classical machine learning algorithms on the deformation fields corresponding to the pairwise registrations.

6 Conclusion

In this article, we present a method for the segmentation of the myocardium using multi-atlas segmentation. The method we present has several important qualities. It is completely automatic and does not even require the location of the heart as user input, thanks to the landmarks detection network. With respect to traditional multi-atlas segmentation algorithm, the speed of the registration

method allows us to use a large database of atlases while keeping the method computationally tractable. To combine the different segmentation into the final result, we define local weights that are a priori learned on a training sample. These weights are based on an estimation of the confidence of the evaluation of a specific point by each atlas. These weights and the deformation fields of the registration result could be used to perform classification of patients with respect to the 5 classes provided by the challenge dataset. The method is evaluated on the training set of the ACDC challenge and is ready to be applied to the final testing dataset.

References

1. Bulat, A., Tzimiropoulos, G.: Convolutional aggregation of local evidence for large pose face alignment. In: British Machine Vision Conference (2016)
2. Kilner, P.J., Geva, T., Kaemmerer, H., Trindade, P.T., Schwitter, J., Webb, G.D.: Recommendations for cardiovascular magnetic resonance in adults with congenital heart disease from the respective working groups of the European society of cardiology. Eur. Heart J. **31**, 794–805 (2010). ehp586
3. Kramer, C.M., Barkhausen, J., Flamm, S.D., Kim, R.J., Nagel, E.: Standardized cardiovascular magnetic resonance (CMR) protocols 2013 update. J. Cardiovasc. Magn. Reson. **15**(1), 91 (2013)
4. Lorenzi, M., Ayache, N., Frisoni, G.B., Pennec, X.: LCC-Demons: a robust and accurate symmetric diffeomorphic registration algorithm. NeuroImage **81**, 470–483 (2013)
5. Payer, C., Štern, D., Bischof, H., Urschler, M.: Regressing heatmaps for multiple landmark localization using CNNs. In: Ourselin, S., Joskowicz, L., Sabuncu, M.R., Unal, G., Wells, W. (eds.) MICCAI 2016. LNCS, vol. 9901, pp. 230–238. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-46723-8_27
6. Rohé, M.-M., Datar, M., Heimann, T., Sermesant, M., Pennec, X.: SVF-Net: learning deformable image registration using shape matching. In: Descoteaux, M., Maier-Hein, L., Franz, A., Jannin, P., Collins, D.L., Duchesne, S. (eds.) MICCAI 2017. LNCS, vol. 10433, pp. 266–274. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-66182-7_31
7. Rohlfing, T., Brandt, R., Menzel, R., Maurer, C.R.: Evaluation of atlas selection strategies for atlas-based image segmentation with application to confocal microscopy images of bee brains. NeuroImage **21**(4), 1428–1442 (2004)
8. Suinesiaputra, A., Bluemke, D.A., Cowan, B.R., Friedrich, M.G., Kramer, C.M., Kwong, R., Plein, S., Schulz-Menger, J., Westenberg, J.J., Young, A.A., et al.: Quantification of LV function and mass by cardiovascular magnetic resonance: multicenter variability and consensus contours. J. Cardiovasc. Magn. Reson. **17**(1), 63 (2015)

MM-WHS Challenge

3D Convolutional Networks for Fully Automatic Fine-Grained Whole Heart Partition

Xin Yang[1(✉)], Cheng Bian[2], Lequan Yu[1], Dong Ni[2], and Pheng-Ann Heng[1,3]

[1] Department of Computer Science and Engineering,
The Chinese University of Hong Kong, Hong Kong, China
xinyang@cse.cuhk.edu.hk
[2] National-Regional Key Technology Engineering Laboratory for Medical Ultrasound, School of Biomedical Engineering, Health Science Center, Shenzhen University, Shenzhen, China
[3] Shenzhen Key Laboratory of Virtual Reality and Human Interaction Technology, Shenzhen Institutes of Advanced Technology,
Chinese Academy of Sciences, Shenzhen, China

Abstract. Segmenting cardiovascular volumes plays a crucial role for clinical applications, especially parsing the whole heart into fine-grained structures. However, conquering fuzzy boundaries and differentiating branchy structures in cardiovascular volume images still remain a challenging task. In this paper, we propose a general and fully automatic solution for fine-grained whole heart partition. The proposed framework originates from the 3D Fully Convolutional Network, and is reinforced in the following aspects: (1) By inheriting the knowledge from a pre-trained C3D Network, our network launches with a good initialization and gains capabilities in coping with overfitting. (2) We triggered several auxiliary loss functions on shallow layers to promote gradient flow and thus alleviate the training difficulties associated with deep neural networks. (3) Considering the obvious volume imbalance among different substructures, we introduced a Multi-class Dice Similarity Coefficient based metric to efficiently balance the training for all classes. We evaluated our method on the MM-WHS Challenge 2017 datasets. Extensive experimental results demonstrated the promising performance of our method. Our framework achieves promising results across different modalities and is general to be referred in other volumetric segmentation tasks.

1 Introduction

Noninvasive cardiac imaging is an indispensable tool for the scanning and inspection of cardiovascular structures. Segmenting the cardiovascular volumes plays a crucial role and even a prerequisite for clinical applications. Parsing the whole heart into fine-grained and well-defined structures opens further opportunities

X. Yang and C. Bian contributed equally to this work.

M. Pop et al. (Eds.): STACOM 2017, LNCS 10663, pp. 181–189, 2018.
https://doi.org/10.1007/978-3-319-75541-0_19

for radiologists to analyze the functionalities of heart in a more precise manner. However, facing with the explosive growth of volumetric data and inevitable user dependency, manually delineating the whole heart are prohibited in practice. Under this situation, automatic solutions to extract the whole heart and further decompose it into fine-grained parts are highly desired.

Subjecting to the large variation of heart, boundary incompleteness, low contrast between different substructures and spatial inhomogeneities [16], it's a arduous challenge to develop automatic solutions for efficient whole heart segmentation across different modalities. Driven by hand-crafted descriptors, previous methods typically resort to active contours [5], shape and texture prior knowledge encoded deformable models [10] and label fusion based non-rigid registration [17]. However, the handcrafted descriptors and limited training data put the performance bottleneck on aforementioned attempts. Based on learned compact features, [15] proposes to map volumetric images to four-chamber segmentations with a regressor. Thriving with large network and relighted by discriminative feature learning capability, Deep Neural Networks (DNNs), especially the Fully Convolutional Networks in 3D version, attract surge of interest in whole heart segmentation [14]. Promising as they are, conquering the difficulties in combating with gradient vanishing and limited medical training data is still an important concern for DNNs. Specifically, simultaneously partitioning the whole heart into substructures with low inter-class variances requires the model to be more discriminative and equal for all classes, while the class imbalance among heart substructures often biases the training of model.

In this paper, we propose a general and fully automatic solution for fine-grained whole heart partition on multi-modality volumetric data, which can simultaneously decompose the whole heart into 7 well-defined substructures. The proposed framework originates from the 3D Fully Convolutional Network (3D FCN) for an efficient end-to-end mapping, and is then reinforced in the following aspects. First, by inheriting the knowledge from a C3D model which is trained on the large scale Sports-1M video dataset [13], our network launches with a good initialization and thus gains capabilities in coping with limited training data and avoiding overfitting. Second, motivated by the success of deep supervision mechanism [3,7], we promote the gradient flow within the network by shortening the backpropagation path and exposing shallow layers to the direct supervision of auxiliary loss functions. (3) Considering the obvious volume imbalance among different classes, we discard the traditional cross-entropy based loss function and extend the Dice Similarity Coefficient based loss function proposed in [9] into a Multi-class variant (mDSC) to significantly balance the training for all classes. We evaluated our method on the MM-WHS Challenge 2017 datasets, including CT and MR. Extensive experimental results demonstrated the promising performance of our method in differentiating substructures of heart.

2 Methodology

Figure 1 is the schematic view of our proposed framework. System input is a complete volume of cardiovascular scanning. Our 3D FCN firstly extracts

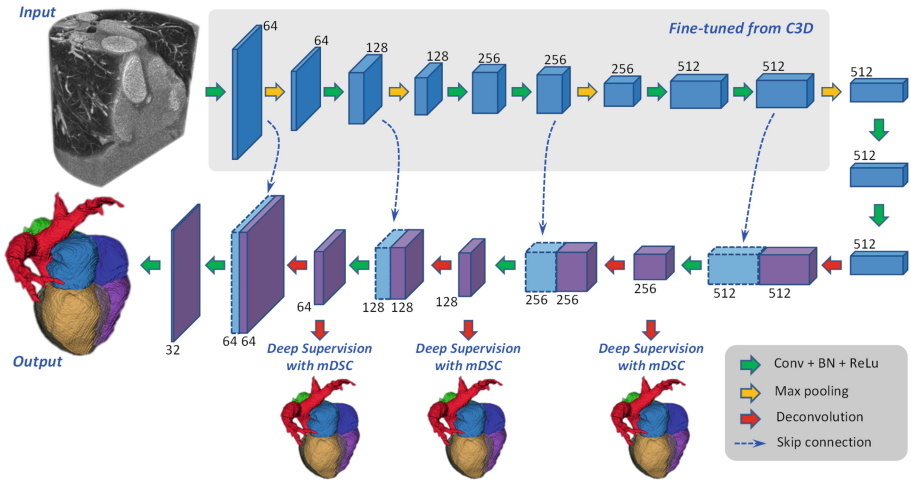


Fig. 1. Schematic view of our proposed framework. Digits represent the number of feature volumes in each layer. Blue volume with dotted line is for concatenation. (Color figure online)

compact gist from the volume with a downsampling path. A following upsampling path then decodes the encoded features into the segmentation results. Skip connections are constructed to blend the features from different semantic levels that with same resolutions. The downsampling path is well initialized with the C3D model. Driven by our proposed Multi-class Dice Similarity Coefficient metric, auxiliary loss functions are properly triggered along the upsampling path. The system output is the semantic labeling of 7 substructures for the whole heart.

2.1 Dense Semantic Labeling with 3D FCN

Characterized with end-to-end mapping, Fully Convolutional Network (FCN) [8] is popular in semantic segmentation. By building skip connections between down- and up-sampling paths, and thus combining detailed feature maps in shallow semantic levels with coarse feature maps in deep levels, U-net [11] promotes FCN in preserving segmentation details. To thoroughly distill contextual cues in volumetric data and get a high throughput, digesting volumetric data with explicit 3D manner is nowadays the trend [2,3]. Therefore, similar with U-net, as shown in Fig. 1, by equipping all layers with 3D operators, we customize a deep 3D FCN to efficiently conduct dense semantic labeling. To reduce computation cost, we use small convolution kernels with size of $3 \times 3 \times 3$ in convolutional layers (Conv). Each Conv layer is followed by a batch normalization (BN) layer and a rectified linear unit (ReLU). Our 3D FCN outputs probability volumes for 8 classes (including background), and finally the labeling results.

Effective as it is, our 3D FCN is still at risk of being degraded by improper initialization, overfitting and class imbalance. In following sections, we will elaborate our solution details to alleviate these problems.

2.2 Knowledge Transfer from C3D Model

Proper initialization is crucial for DNNs to avoid being trapped in local minima and gain generalization ability. Knowledge transfer proves to be momentous in addressing this problem [1]. For DNNs, the compact and task-specific representations in high semantic levels are successively built upon the basic features learned in shallow layers, in which the filters are generic across different domains. Consequently, inheriting the knowledge from models which are pre-trained on large-scale annotated datasets in other domain can be beneficial for DNNs.

Many successful large networks, such as ImageNet [6] and VGG16 [12], can not be exploited for knowledge transfer in our task, because they are trained on 2D images. Recently, the action recognition in video gains benefit from DNNs and sheds light on transfer learning for 3D FCN. Associated with 3D convolution operators, the remarkable C3D model proposed in [13] achieves state-of-the-art recognition performance by simultaneously exploring spatial and temporal representations across consecutive frames. Therefore, we propose to conduct knowledge transfer between C3D model and our 3D FCN by implanting the filters from the former into the latter. Specifically, shown as Fig. 1, we only take the filters of 6 shallow layers, including *conv1, conv2, conv3a, conv3b, conv4a and conv4b*, in C3D to initialize the down-sampling path of our 3D FCN.

2.3 Promote Training with Deep Supervision

Directly training the deep 3D FCN will be degraded by low efficiency and overfitting due to the gradient vanishing problem [4] in shallow layers. Motivated by its success in [3,7], we adopt the deep supervision mechanism proposed in [7], which shortens the backpropagation path and exposes shallow convolutional layers to the direct supervision of $\mathcal{M}$ auxiliary classifiers ($1\times1\times1$ convolutional layer and softmax layer). Because the output of auxiliary classifiers have different dimensions from that of ground truth, so we insert deconvolutional layers to accordingly upsample auxiliary classifiers' output before softmax layer.

Let $\mathcal{X}^{w\times h\times d}$ be the volumetric input, W be the weights of main network, $w = (w^1, w^2, .., w^m)$ be the weights of auxiliary classifiers and w^m denotes the weight of the m^{th} classifier. $\mathcal{L}_m$ denote the m^{th} auxiliary loss function. The final loss function $\mathcal{L}$ of our deeply supervised 3D FCN is:

$$\mathcal{L}(\mathcal{X}; W, w) = \mathcal{L}(\mathcal{X}; W) + \sum_{m\in\mathcal{M}} \beta_m \mathcal{L}_m(\mathcal{X}; W, w^m) + \lambda(||W||^2 + \sum_{m\in\mathcal{M}} ||w^m||^2), \quad (1)$$

where β_m is the weight of different auxiliary classifiers. Because layers in down-sampling path have smaller receptive fields than deep layers, so we only attach auxiliary classifiers in our upsampling path, shown as Fig. 1. We will explain our design of loss function $\mathcal{L}$ and $\mathcal{L}_m$ in Sect. 2.4.

2.4 Multi-class Balanced Loss Function

The design of loss function defines the mapping that DNNs need to learn, and determines the bound that DNNs can finally achieve. For classification and segmentation tasks, the classical choice is the differentiable cross-entropy based loss function. However, when comes to handle the segmentation of multiple classes, the cross-entropy based loss function tends to sacrifice the minor classes and only fit the dominant ones. For FCN based segmentation, the reason for this phenomenon is that the classical cross-entropy function trivially summarizes the error of each voxel without giving associated significances for specific classes. Recently, a novel Dice Similarity Coefficient (DSC) based loss function which can circumvent the flaw is proposed in [9]. DSC based loss function can inherently address the class-imbalance, since it roots in the global shape similarity and thus the cost for each class is self-normalized before being counted into the total loss.

Attractive as it is, the DSC based loss function in [9] is only applicable to binary segmentation tasks. Therefore, we propose to introduce a multi-class Dice Similarity Coefficient (mDSC) based loss function to guide and balance the training for multiple classes. Mathematically, given the segmentation ground truth $G^{w\times h\times d}$, we firstly encode it into a one-hot format for C classes $\mathcal{G}^{C\times w\times h\times d}$, $C = 8$ for our task. With the probability volumes $\mathcal{P}^{C\times w\times h\times d}$ generated by 3D FCN, our proposed mDSC based differentiable loss function can be written as

$$\mathcal{L}_{mDSC} = -\sum_{c\in C} \frac{\frac{2}{N}\sum_i^N \mathcal{G}_c^i \mathcal{P}_c^i}{\sum_i^N \mathcal{G}_c^i \mathcal{G}_c^i + \sum_i^N \mathcal{P}_c^i \mathcal{P}_c^i}, \tag{2}$$

where $N = w \times h \times d$, $\mathcal{G}_c^i$ and $\mathcal{P}_c^i$ are the i^{th} voxel of c^{th} volume in $\mathcal{G}$ and $\mathcal{P}$. The $1/N$ in denominator is empirically devised to suppress prediction noise. With this formulation, all classes gain opportunities to be fairly treated during the training process. In Fig. 2 We illustrate the improvement of probability maps for CT segmentation when we change from cross-entropy based loss function to mDSC based loss function. The warmer the color, the higher the probability. As we can observe, mDSC based loss function makes the probability maps to be more compact and thus enlarges the intra-class gaps. Quantitative improvement caused by mDSC will be illustrated in Sect. 3.

3 Experimental Results

Dataset and Pre-processing: We evaluated our networks on two tasks: whole heart segmentation in CT and MR volumes which come from Multi-Modality Whole Heart Segmentation Challenge 2017 datasets. The datasets consist of 60 CT and 60 MR volumes from anonymized healthy patients (each contains 20 for training and 40 for testing). Note that the ground truth of testing dataset is held out by the organizer for independent evaluation. Before training networks, we pre-process the training dataset by normalizing them as zero mean and unit variance. In order to tackle the insufficiency of training data and avoid overfitting,

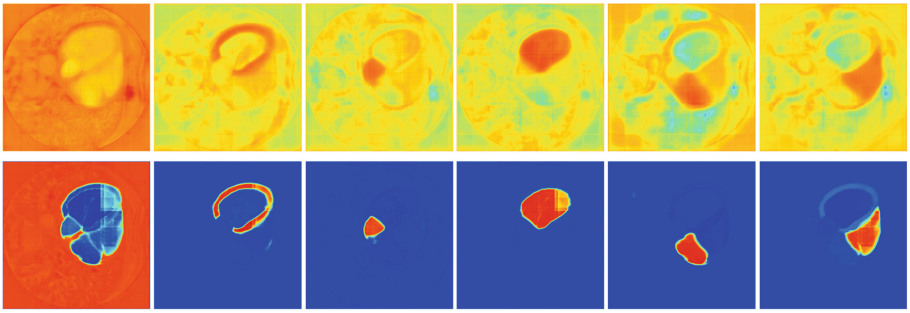


Fig. 2. From left to right: probability of background, myocardium of the left ventricle, left atrium blood cavity, left ventricle blood cavity, right atrium blood cavity and right ventricle blood cavity. From top to bottom: training with cross-entropy and mDSC.

we augment the training dataset with rotation. Because there is no alignment between CT and MR data, so we just train two networks to segment these two modalities independently.

Implementation Details: The proposed 3D FCN was implemented in *Tensorflow*, using a standard PC with a 2.60 GHz Intel(R) Xeon(R) E5-2650 CPU and 2 NVIDIA GeForce GTX TITAN X GPUs. Given the limited memory in one GPU, we assign the down- and up-sampling paths to different GPUs. We update the weights of network with a Adam optimizer (batch size = 1, initial learning rate is 0.001, total iteration = 30000). We utilize 3 auxiliary classifiers. Since early auxiliary classifiers often generate coarse labeling results, we set $\beta_0 = 0.2, \beta_1 = 0.4, \beta_2 = 0.8$ from coarse to fine levels. Randomly cropped $96 \times 96 \times 96$ cubes serve as input to train our network. We adopt sliding window and overlap-tiling stitching strategies to generate predictions for the whole volume, and remove small isolated connected components in final labeling result.

Table 1. Quantitative evaluation for whole heart segmentation in CT volume

Method	Metrics	Substructures of heart							Mean
		MLV	LABC	LVBC	RABC	RVBC	ASA	PUA	
DS+KT+Cross	*DSC[%]*	81.31	79.07	90.86	85.39	81.61	71.71	76.38	**80.90**
	HDB[voxel]	26.13	71.01	25.51	40.58	50.84	48.70	67.15	47.13
	ADB[voxel]	2.954	29.01	3.345	6.178	7.371	3.509	5.378	8.249
DS+KT+mDSC	*DSC[%]*	68.97	90.00	83.42	84.36	62.50	91.51	80.84	80.22
	HDB[voxel]	47.53	37.42	30.31	33.06	72.41	27.77	65.27	**44.825**
	ADB[voxel]	3.185	5.415	5.666	5.875	8.245	2.692	3.875	**4.993**

Table 2. Quantitative evaluation for whole heart segmentation in MR volume

Method	Metrics	Substructures of heart							Mean
		MLV	LABC	LVBC	RABC	RVBC	ASA	PUA	
DS+KT+Cross	*DSC[%]*	71.98	76.96	87.05	78.60	73.38	63.50	70.85	74.62
	Jaccard[%]	58.06	65.98	78.34	68.44	62.94	50.20	58.92	63.27
	ADB[voxel])	1.323	1.679	1.587	2.062	5.901	2.075	1.781	2.344
DS+KT+mDSC	*DSC[%]*	66.54	74.62	86.80	86.16	71.43	71.24	70.19	**75.28**
	Jaccard[%]	52.07	64.23	77.71	75.97	60.59	58.13	57.88	**63.80**
	ADB[voxel]	1.509	1.761	1.646	1.773	3.300	1.560	1.587	**1.864**

Quantitative and Qualitative Analysis: To consider both region and boundary similarities, we adopt 4 metrics to evaluate the proposed framework on segmentation, including DSC ($DSC = 2(A \cap B)/(A+B)$), Jaccard Index, Hausdorff Distance of Boundaries (HDB), and Average Distance of Boundaries (ADB). Sharing the basic configuration of deep supervision (DS) and knowledge transfer (KT), we mainly conduct experiments to compare the performance of cross-entropy (denoted as *DS+KT+Cross*) and mDSC (denoted as *DS+KT+mDSC*) based models. By taking 10 volumes in training dataset as validation, we show the segmentation results in CT and MR volumes for 7 substructures in Tables 1 and 2. The substructures are myocardium of the left ventricle (MLV), left atrium blood cavity (LABC), left ventricle blood cavity (LVBC), right atrium blood cavity (RABC), right ventricle blood cavity (RVBC), ascending aorta (ASA) and pulmonary artery (PUA). As we can see, the mDSC based models is general in achieving improvements on most classes and metrics for both two imaging

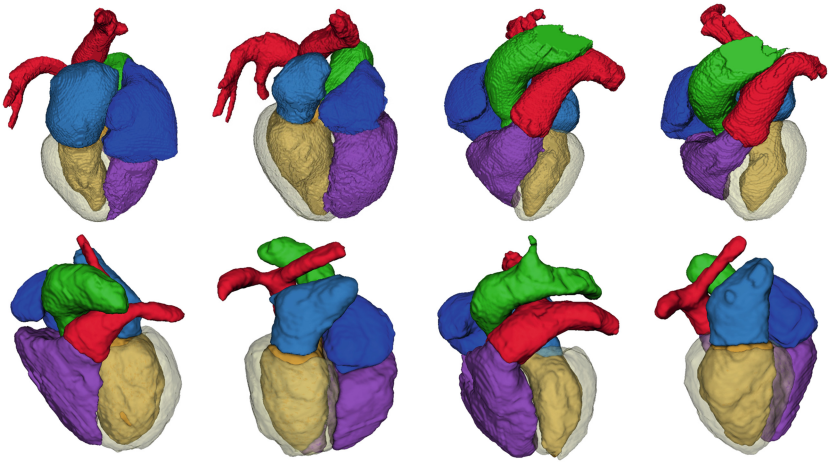


Fig. 3. Visualization of our segmentation results. From top to bottom: substructure partition results in CT, MR volumes.

modalities. We further provide explicit visualization of the fine-grained partition results for 4 testing CT and MR volumes in Fig. 3. Our proposed method conquers large shape and size variances, boundary uncertainty and low inter-class variance, and presents promising performance.

4 Conclusions

We present a fully automatic framework to segment the whole heart from CT and MR volumes, and further decompose it into several substructures, which would potentially promote clinical studies about heart. Leveraging the advanced techniques, including knowledge transfer and deep supervision, we propose to use the mDSC based loss function to effectively balance the training procedure and therefore improve the segmentation performance. Promising quantitative and qualitative results are achieved on a large dataset.

Acknowledgments. The work in this paper was supported by the grant from National Natural Science Foundation of China under Grant 61571304, a grant from Hong Kong Research Grants Council (Project no. CUHK 412513) and grants from the National Natural Science Foundation of China (Project No. 61233012 and No. 81601576).

References

1. Chen, H., Ni, D., Qin, J., et al.: Standard plane localization in fetal ultrasound via domain transferred deep neural networks. IEEE JBHI **19**(5), 1627–1636 (2015)
2. Çiçek, Ö., Abdulkadir, A., et al.: 3D U-NET: learning dense volumetric segmentation from sparse annotation. arXiv preprint arXiv:1606.06650 (2016)
3. Dou, Q., Yu, L., et al.: 3D deeply supervised network for automated segmentation of volumetric medical images. Med. Image Anal. **41**, 40–54 (2017)
4. Glorot, X., Bengio, Y.: Understanding the difficulty of training deep feedforward neural networks. Aistats. **9**, 249–256 (2010)
5. Kaus, M.R., von Berg, J., Weese, J., et al.: Automated segmentation of the left ventricle in cardiac MRI. Med. Image Anal. **8**(3), 245–254 (2004)
6. Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. In: NIPS, pp. 1097–1105 (2012)
7. Lee, C.Y., Xie, S., Gallagher, P., Zhang, Z., Tu, Z.: Deeply-supervised nets (2015)
8. Long, J., Shelhamer, E., Darrell, T.: Fully convolutional networks for semantic segmentation. In: CVPR, pp. 3431–3440 (2015)
9. Milletari, F., Navab, N., et al.: V-net: fully convolutional neural networks for volumetric medical image segmentation. In: 2016 Fourth International Conference on 3D Vision (3DV), pp. 565–571. IEEE (2016)
10. Peters, J., Ecabert, O., Meyer, C., Schramm, H., Kneser, R., Groth, A., Weese, J.: Automatic whole heart segmentation in static magnetic resonance image volumes. In: Ayache, N., Ourselin, S., Maeder, A. (eds.) MICCAI 2007. LNCS, vol. 4792, pp. 402–410. Springer, Heidelberg (2007). https://doi.org/10.1007/978-3-540-75759-7_49

11. Ronneberger, O., Fischer, P., Brox, T.: U-Net: convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) MICCAI 2015. LNCS, vol. 9351, pp. 234–241. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-24574-4_28
12. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556 (2014)
13. Tran, D., Bourdev, L., Fergus, R., Torresani, L., Paluri, M.: Learning spatiotemporal features with 3d convolutional networks. In: ICCV, pp. 4489–4497 (2015)
14. Yu, L., Yang, X., Qin, J., Heng, P.-A.: 3D FractalNet: dense volumetric segmentation for cardiovascular MRI volumes. In: Zuluaga, M.A., Bhatia, K., Kainz, B., Moghari, M.H., Pace, D.F. (eds.) RAMBO/HVSMR -2016. LNCS, vol. 10129, pp. 103–110. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-52280-7_10
15. Zhen, X., Zhang, H., Islam, A., Bhaduri, M., Chan, I., Li, S.: Direct and simultaneous estimation of cardiac four chamber volumes by multioutput sparse regression. Med. Image Anal. **36**, 184–196 (2017)
16. Zhuang, X.: Challenges and methodologies of fully automatic whole heart segmentation: a review. J. Healthc. Eng. **4**(3), 371–408 (2013)
17. Zhuang, X., Rhode, K.S., Razavi, R.S., Hawkes, D.J., Ourselin, S.: A registration-based propagation framework for automatic whole heart segmentation of cardiac MRI. IEEE Trans. Med. Imaging **29**(9), 1612–1625 (2010)

Multi-label Whole Heart Segmentation Using CNNs and Anatomical Label Configurations

Christian Payer[1(✉)], Darko Štern[2], Horst Bischof[1], and Martin Urschler[2,3]

[1] Institute for Computer Graphics and Vision, Graz University of Technology, Graz, Austria
christian.payer@icg.tugraz.at
[2] Ludwig Boltzmann Institute for Clinical Forensic Imaging, Graz, Austria
[3] BioTechMed-Graz, Graz, Austria

Abstract. We propose a pipeline of two fully convolutional networks for automatic multi-label whole heart segmentation from CT and MRI volumes. At first, a convolutional neural network (CNN) localizes the center of the bounding box around all heart structures, such that the subsequent segmentation CNN can focus on this region. Trained in an end-to-end manner, the segmentation CNN transforms intermediate label predictions to positions of other labels. Thus, the network learns from the relative positions among labels and focuses on anatomically feasible configurations. Results on the MICCAI 2017 Multi-Modality Whole Heart Segmentation (MM-WHS) challenge show that the proposed architecture performs well on the provided CT and MRI training volumes, delivering in a three-fold cross validation an average Dice Similarity Coefficient over all heart substructures of 88.9% and 79.0%, respectively. Moreover, on the MM-WHS challenge test data we rank first for CT and second for MRI with a whole heart segmentation Dice score of 90.8% and 87%, respectively, leading to an overall first ranking among all participants.

Keywords: Heart · Segmentation · Multi-label
Convolutional neural network · Anatomical label configurations

1 Introduction

The accurate analysis of the whole heart substructures, i.e., left and right ventricle, left and right atrium, myocardium, pulmonary artery and the aorta, is highly relevant for cardiovascular applications. Therefore, automatic segmentation of these substructures from CT or MRI volumes is an important topic in medical image analysis [1,11,12]. Challenges for segmenting the heart substructures are their large anatomical variability in shape among subjects, the potential indistinctive boundaries between substructures and, especially for MRI data, artifacts

This work was supported by the Austrian Science Fund (FWF): P28078-N33.

M. Pop et al. (Eds.): STACOM 2017, LNCS 10663, pp. 190–198, 2018.
https://doi.org/10.1007/978-3-319-75541-0_20

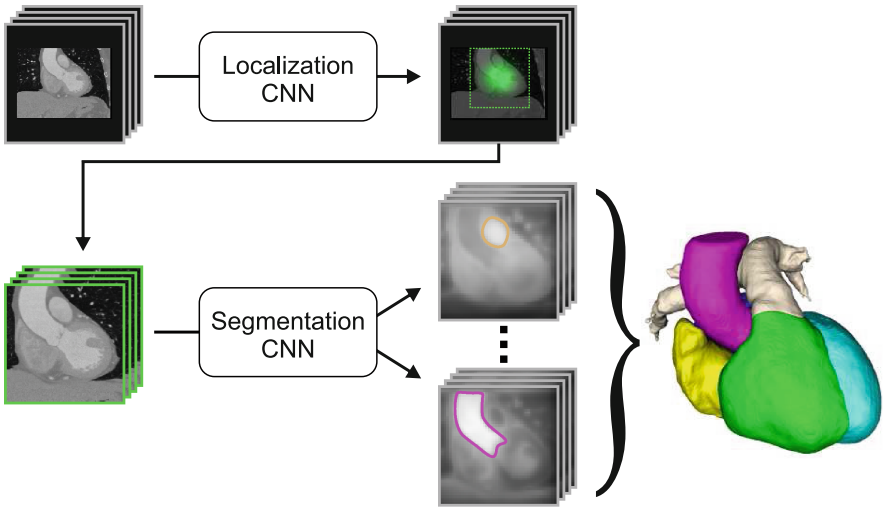


Fig. 1. Overview of our fully automatic two-step multi-label segmentation pipeline. The first CNN uses a low resolution volume as input to localize the center of the bounding box around all heart substructures. The second CNN crops a region around this center and performs the multi-label segmentation.

and intensity inhomogeneities resulting from the acquisition process. To objectively compare and analyze whole heart substructure segmentation approaches, efforts like the MICCAI 2017 Multi-Modality Whole Heart Segmentation (MM-WHS) challenge are necessary and important for potential future application of semi-automated and fully automatic methods in clinical practice.

In this work, we propose a deep learning framework for fully automatic multi-label segmentation of volumetric images. The first convolutional neural network (CNN) localizes the center of the bounding box around all heart substructures. Based on this bounding box, the second CNN predicts the label positions, i.e., the spatial region each label occupies in the volume. By transforming intermediate label predictions to positions of other labels, this second CNN learns the relative positions among labels and focuses on anatomically feasible positions. We evaluate our proposed method on the MM-WHS challenge dataset consisting of CT and MRI volumes.

2 Method

We perform fully automatic multi-label whole heart segmentation from CT or MRI data with CNNs using volumetric kernels. Due to the increased memory and runtime requirements when applying such CNNs to 3D data, we use a two-step pipeline that first localizes the heart on lower resolution volumes, followed by obtaining the final segmentation on a higher resolution. This pipeline is illustrated in Fig. 1.

Localization CNN: As a first step, we localize the approximate center of the heart. Although different localization strategies could be used for this purpose, e.g., [9], to stay within the same machine learning framework for all steps we perform landmark localization with a U-Net-like fully convolutional CNN [5,8] using heatmap regression [6,7,10], trained to regress the center of the bounding box around all heart substructure segmentations. Due to memory restrictions, we downsample the input volume and let the network operate on a low resolution. Then, we crop a fixed size region around the predicted bounding box center and resample voxels from the original input volume on a higher resolution than for localizing bounding box centers. We define the fixed size of this region, such that it encloses all segmentation labels on every image from the training set, thus covering all the anatomical variation occurring in the training data.

Segmentation CNN: A second CNN for multi-label classification predicts the labels of each voxel inside the cropped region from the localization CNN (see Fig. 1). For this segmentation task, we use an adaptation of the fully convolutional end-to-end trained SpatialConfiguration-Net from [6] that was originally proposed for landmark localization. The main idea in [6] is to learn from relative positions among structures to focus on anatomically feasible configurations as seen in the training data. In a three stage architecture, the network generates accurate intermediate label predictions, transforms these predictions to positions of other labels, and combines them by multiplication.

In the first stage, a U-Net-like architecture [8], which has as many outputs as segmentation labels, generates the intermediate label predictions. For each output voxel, a sigmoid activation function is used to restrict the values between 0 and 1, corresponding to a voxel-wise probability prediction of all labels. Then, in the second stage, the network transforms these probabilities to the positions of other labels, thus allowing the network to learn feasible anatomical label configurations by suppressing infeasible intermediate predictions. As the estimated positions of other labels are not precise, for this stage we can downsample the outputs of the U-Net to reduce memory consumption and computation time without losing prediction performance. Consecutive convolution layers transform these downsampled label predictions to the estimated positions of other labels. Upsampling back to the input resolution leads to transformed label predictions, which are entirely based on the intermediate label probabilities of other labels. Finally, in the last stage, multiplying the intermediate predictions from the U-Net with the transformed predictions results in the combined label predictions. For more details on the SpatialConfiguration-Net, we refer the reader to [6]. Without any further postprocessing, choosing the maximum value among the label predictions for each voxel leads to the final multi-label segmentation.

3 Experimental Setup

Dataset: We evaluated the networks on the datasets of the MM-WHS challenge. The organizers provided 20 CT and 20 MRI volumes with corresponding manual segmentations of seven whole heart substructures. The volumes were

acquired in clinics with different scanners, resulting in varying image quality, resolution and voxel spacing. The maximum physical size of the input volumes for CT is $300 \times 300 \times 188$ mm^3 while for MRI it is $400 \times 360 \times 400$ mm^3. The maximum size of the bounding box around the segmentation labels for CT is $155 \times 151 \times 160$ mm^3 (MRI: $180 \times 153 \times 209$ mm^3).

Implementation Details: We train and test the networks with Caffe [3] where we perform data augmentations using ITK[1], i.e., intensity scale and shift, rotation, translation, scaling and elastic deformations. We apply these augmentations on the fly during network training. We optimize the networks using Adam [4] with learning rate 0.001 and the recommended default parameters from [4]. Due to memory restrictions coming from the volumetric inputs and the use of 3D convolution kernels, we choose a mini-batch size of 1. Hyperparameters for training and network architecture were chosen empirically from the cross validation setup. All experiments were performed on an Intel Core i7-4820K based workstation with a 12 GB NVidia Geforce TitanX.

Input Preprocessing: The intensity values of the CT volumes are divided by 2048 and clamped between -1 and 1. For MRI, the intensity normalization factor is different for each image. We divide each intensity value by the median of 10% of the highest intensity values of each image to be robust to outliers. In this way, all voxels are in the range between 0 and 1; we multiply them with 2, shift them by -1, and clamp them between -1 and 1. For random intensity augmentations during training, we shift intensity values by $[-0.1, 0.1]$ and scale them by $[0.9, 1.1]$. As we know the voxel spacing of each volume, we resample the images trilinearly to have a fixed isotropic voxel spacing for each network. In training, we randomly scale the volumes by $[0.8, 1.2]$ and rotate the volumes by $[-10°, 10°]$ in each dimension. We additionally employ elastic deformations by moving points on a regular $8 \times 8 \times 8$ voxel grid randomly by up to 10 voxels, and interpolating with 3rd order B-splines. All random operations sample from a uniform distribution within the specified intervals. During testing, we do not employ any augmentations.

Localization CNN: We localize the bounding box centers with a U-Net-like network using heatmap regression. The U-Net has an input voxel size of $32 \times 32 \times 32$ voxels and 4 levels. For the CT images, we resample the input volumes to have an isotropic voxel size of 10 mm^3 (MRI: 12 mm^3), which leads to a maximum input volume size of $320 \times 320 \times 320$ mm^3 (MRI: $384 \times 384 \times 384$ mm^3). Then, we feed the resampled, centered volumes as input to the network. Each level of the contracting path as well as the expanding path consists of two consecutive convolution layers with $3 \times 3 \times 3$ kernels and zero padding. Each convolution layer, except the last one, has a ReLU activation function. The next deeper levels with half the resolution are generated with average pooling; the next higher levels with twice the resolution are generated with trilinear upsampling. Starting from 32 outputs at the first level, the number of outputs of each convolution layer at the same level is identical, while it is doubled at the next deeper level.

[1] The Insight Segmentation and Registration Toolkit https://www.itk.org/.

We employ dropout of 0.5 after the convolutions of the contracting path in the deepest two levels. A last convolution layer at the highest level with one output generates the predicted heatmap. The final output of the network is resampled back to the original input volume size with tricubic interpolation, to generate more precise localization. The networks are trained with L2 loss on each voxel to predict a Gaussian target heatmap with $\sigma = 1.5$. We initialize the convolution layer weights with the method from [2], except for the last layer, where we sample from a Gaussian distribution with standard deviation 0.001. All biases are initialized with 0. We train the network for 30000 iterations.

Segmentation CNN: The segmentation network is structured as follows. The intermediate label predictions are generated with a similar U-Net as used for the localization, but twice the input voxel size, i.e., $64 \times 64 \times 64$, and twice the number of convolution layer outputs. For the CT images, we resample the input images trilinearly to have an isotropic voxel size of 3 mm^3 (MRI: 4 mm^3), which leads to a maximum input volume size of $192 \times 192 \times 192$ mm^3 (MRI: $256 \times 256 \times 256$ mm^3). The final layer of this U-Net generates eight outputs, which correspond to the number of segmentation labels, i.e., seven heart substructures and the background. This layer has a sigmoid activation function to predict intermediate probabilities. Then in the subsequent label transformation stage, the outputs of the previous stage are downsampled with average pooling by a factor of 4 in each dimension. Four consecutive convolution layers with kernel size $5 \times 5 \times 5$ and zero padding transform the downsampled outputs of the U-Net. The intermediate layers have 64 outputs with a ReLU activation function, while the last layer has eight outputs with linear activation. A final trilinear upsampling resizes the output back to the resolution of the first stage. After multiplying the predictions of the U-Net and the label transformation stage, a softmax with multinomial logistic loss on each voxel is used as a target function. The final output of the network is resampled back to the original input volume size with tricubic interpolation, to generate more precise segmentations. The weights of each layer are initialized as proposed in [2]; the biases are initialized with 0. We train the network for 50000 iterations. To show the impact of the label transformation stage on the total performance, we additionally train a U-Net that is identical to the first stage of the segmentation CNN, without the subsequent label transformation.

4 Results and Discussion

To evaluate our proposed approach, we performed a three-fold cross validation on the training images of the MM-WHS challenge for both imaging modalities, such that each image is tested exactly once. Additionally, the organizers of the MM-WHS challenge provided the ranked results of the challenge participants on the undisclosed manual segmentations of the test set.

The localization network achieved a mean Euclidean distance to the ground truth bounding box centers of 13.2 mm with 5.4 mm standard deviation for CT, and 20.0 mm $\pm$ 30.5 mm for MRI, respectively. Despite the larger standard

Table 1. Dice Similarity Coefficients in % for the U-Net-like CNN (U-Net) and our proposed segmentation CNN (Seg-CNN). The values show the mean (± standard deviation) of all images from the CT and MRI cross validation setup for each segmentation label. Label abbreviations: LV - left ventricle blood cavity, Myo - myocardium of the left ventricle, RV - right ventricle blood cavity, LA - left atrium blood cavity, RA - right atrium blood cavity, aorta - ascending aorta, PA - pulmonary artery, μ - average of the seven whole heart substructures.

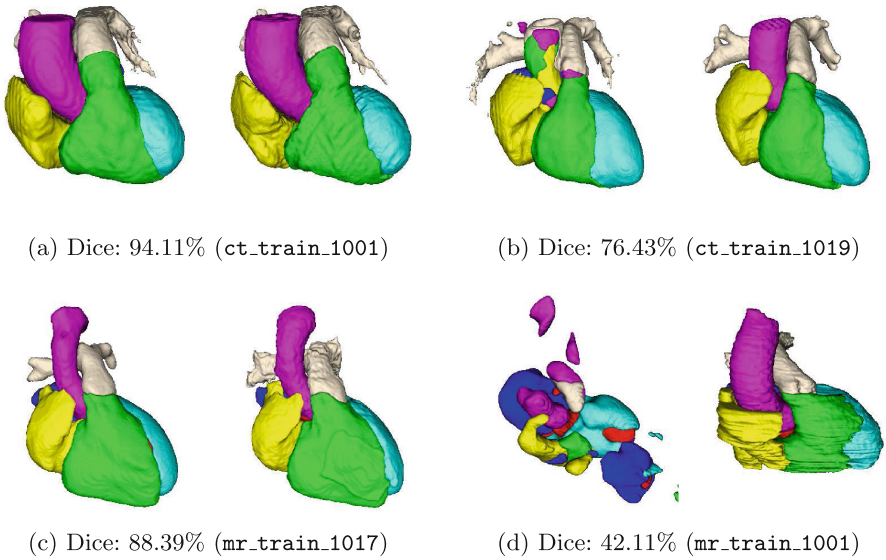
		LV	Myo	RV	LA	RA	aorta	PA	μ
CT	U-Net	91.0 (± 4.3)	86.1 (± 4.2)	**88.8** (± 3.9)	91.0 (± 5.2)	86.5 (± 6.0)	**94.0** (± 6.2)	**83.7** (± 7.7)	88.7 (± 3.3)
	Seg-CNN	**92.4** (± 3.3)	**87.2** (± 3.9)	87.9 (± 6.5)	**92.4** (± 3.6)	**87.8** (± 6.5)	91.1 (± 18.4)	83.3 (± 9.1)	**88.9** (± 4.3)
MRI	U-Net	81.1 (± 23.8)	68.1 (± 25.3)	76.2 (± 24.9)	74.0 (± 24.7)	77.0 (± 22.1)	70.6 (± 20.2)	68.7 (± 16.5)	73.7 (± 21.4)
	Seg-CNN	**87.7** (± 7.7)	**75.2** (± 12.1)	**77.7** (± 19.5)	**81.1** (± 13.8)	**82.7** (± 15.8)	**76.6** (± 13.8)	**72.0** (± 16.1)	**79.0** (± 11.7)

Table 2. Dice Similarity Coefficients on the CT and MRI test sets of the MM-WHS challenge for all participants in %, ranked by highest score. The values show the mean of all images for each segmentation label. The results of our approach are highlighted in yellow. Label abbreviations: same as Table 1, WHS - whole heart segmentation.

		LV	Myo	RV	LA	RA	aorta	PA	WHS
CT	1.	91.8	**88.1**	**90.9**	92.9	**88.8**	**93.3**	**84.0**	**90.8**
	2.	**92.3**	85.6	85.7	**93.0**	87.1	89.4	83.5	89.0
	3.	90.4	85.1	88.3	91.6	83.6	90.7	78.4	87.9
	4.	90.1	84.6	85.6	88.4	83.7	91.4	80.0	87.0
	5.	90.8	87.4	80.6	90.8	85.5	83.5	67.7	86.6
	6.	89.3	83.7	81.0	88.9	81.2	86.8	69.8	84.9
	7.	88.0	81.5	84.9	84.5	79.9	83.9	73.7	83.8
	8.	59.3	53.3	70.6	72.0	51.5	60.1	63.7	62.3
MRI	1.	**91.8**	**78.1**	**87.1**	**88.6**	87.3	**87.8**	**80.4**	**87.0**
	2.	91.6	77.8	86.8	85.5	**88.1**	83.8	73.1	86.3
	3.	87.1	74.7	83.0	81.1	75.9	83.9	71.5	81.8
	4.	89.7	76.3	81.9	76.5	80.8	70.8	68.5	81.7
	5.	83.6	72.1	80.5	74.2	83.2	82.1	69.7	79.7
	6.	85.5	72.8	76.0	83.2	78.2	77.1	57.8	79.2
	7.	75.0	65.8	75.0	82.6	85.9	80.9	72.6	78.3
	8.	70.2	62.3	68.0	67.6	65.4	59.9	47.0	67.4

deviation for MRI, we observed that this is sufficient for the subsequent cropping, i.e., the input for the multi-level segmentation network, as the cropped region encloses the segmentation labels of all heart substructures of all tested images from the training set.

We provide evaluation results of our proposed multi-label segmentation CNN and of our implementation of the U-Net. The Dice Similarity Coefficients are



(a) Dice: 94.11% (`ct_train_1001`) (b) Dice: 76.43% (`ct_train_1019`)

(c) Dice: 88.39% (`mr_train_1017`) (d) Dice: 42.11% (`mr_train_1001`)

Fig. 2. Segmentation results of volumes with best and worst Dice scores for CT (top row) and MRI (bottom row) datasets. Volumes on the left show predictions; volumes on the right show corresponding ground truth segmentation.

shown in Table 1, where both approaches perform similar for the CT dataset. However, in the MRI dataset, which shows more variation in anatomical field of view, intensity ranges and acquisition artifacts compared to CT data, the improvements when adding the label configuration stage are very prominent. We assume that the larger variability of MRI data would require more training data for the U-Net, while our proposed label transformation stage compensates the lack of training data by focusing on anatomically feasible configurations. Figure 2 shows qualitative segmentation results of the best and worst cases for CT and MRI datasets, respectively. The wrong labels in the ascending aorta of Fig. 2b were caused by acquisition artifacts in the CT volume, whereas a failing intensity value normalization of the MRI volume resulted in wrong segmentations in Fig. 2d.

For generating the segmentations on the test set, we trained the networks on all training images with the same hyperparameters as used for the cross validation. Table 2 shows the results on the test set of the MM-WHS challenge, ranked for all participants selected for the final comparison. By achieving the first place on the CT dataset and the second place on the MRI dataset, our method was the best in overall ranking. Although in CT the results of our own cross validation and the test images of the challenge are similar, in MRI the results on the test set are better than for the cross validation. We think the reason for this is the larger variability in the MRI dataset, such that increasing the number of training images improves the results more drastically as compared to CT.

In future work we are planning to evaluate our method on datasets coming from different scanners and sites.

5 Conclusion

We have presented a method for fully automatic multi-label segmentation from CT and MRI data, using a pipeline of two fully convolutional networks, performing coarse localization of a bounding box around the heart, followed by multi-label segmentation of the heart substructures. Results on the MICCAI 2017 Multi-Modality Whole Heart Segmentation challenge show top performance of our proposed method among the contesting participants. Achieving the first place in the CT and the second place in the MRI dataset, our method was the best performing in overall ranking.

References

1. Grbic, S., Ionasec, R., Vitanovski, D., Voigt, I., Wang, Y., Georgescu, B., Comaniciu, D.: Complete valvular heart apparatus model from 4D cardiac CT. Med. Image Anal. **16**(5), 1003–1014 (2012)
2. He, K., Zhang, X., Ren, S., Sun, J.: Delving deep into rectifiers: surpassing human-level performance on imagenet classification. In: Proceedings of International Conference on Computer Vision, pp. 1026–1034. IEEE (2015)
3. Jia, Y., Shelhamer, E., Donahue, J., Karayev, S., Long, J., Girshick, R., Guadarrama, S., Darrell, T.: Caffe: convolutional architecture for fast feature embedding. In: Proceedings of ACM International Conference on Multimedia, pp. 675–678. ACM (2014)
4. Kingma, D.P., Ba, J.: Adam: a method for stochastic optimization. In: International Conference for Learning Representations. CoRR, abs/1412.6980 (2015)
5. Long, J., Shelhamer, E., Darrell, T.: Fully convolutional networks for semantic segmentation. In: Proceedings of Computer Vision Pattern Recognition, pp. 3431–3440. IEEE (2015)
6. Payer, C., Štern, D., Bischof, H., Urschler, M.: Regressing heatmaps for multiple landmark localization using CNNs. In: Ourselin, S., Joskowicz, L., Sabuncu, M.R., Unal, G., Wells, W. (eds.) MICCAI 2016. LNCS, vol. 9901, pp. 230–238. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-46723-8_27
7. Pfister, T., Charles, J., Zisserman, A.: Flowing convnets for human pose estimation in videos. In: Proceedings of International Conference on Computer Vision, pp. 1913–1921 (2015)
8. Ronneberger, O., Fischer, P., Brox, T.: U-Net: convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) MICCAI 2015. LNCS, vol. 9351, pp. 234–241. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-24574-4_28
9. Štern, D., Ebner, T., Urschler, M.: From local to global random regression forests: exploring anatomical landmark localization. In: Ourselin, S., Joskowicz, L., Sabuncu, M.R., Unal, G., Wells, W. (eds.) MICCAI 2016. LNCS, vol. 9901, pp. 221–229. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-46723-8_26

10. Tompson, J., Jain, A., LeCun, Y., Bregler, C.: Joint training of a convolutional network and a graphical model for human pose estimation. In: Proceedings of Neural Information Processing System, pp. 1799–1807 (2014)
11. Zhuang, X., Rhode, K., Razavi, R., Hawkes, D.J., Ourselin, S.: A registration-based propagation framework for automatic whole heart segmentation of cardiac MRI. IEEE Trans. Med. Imaging **29**(9), 1612–1625 (2010)
12. Zhuang, X., Shen, J.: Multi-scale patch and multi-modality atlases for whole heart segmentation of MRI. Med. Image Anal. **31**, 77–87 (2016)

Multi-Planar Deep Segmentation Networks for Cardiac Substructures from MRI and CT

Aliasghar Mortazi[1], Jeremy Burt[2], and Ulas Bagci[1(✉)]

[1] Center for Research in Computer Vision (CRCV), University of Central Florida, Orlando, FL, USA
a.mortazi@knights.ucf.edu, ulasbagci@gmail.com
[2] Diagnostic Radiology Department, Florida Hospital, Orlando, FL, USA

Abstract. Non-invasive detection of cardiovascular disorders from radiology scans requires quantitative image analysis of the heart and its substructures. There are well-established measurements that radiologists use for diseases assessment such as ejection fraction, volume of four chambers, and myocardium mass. These measurements are derived as outcomes of precise segmentation of the heart and its substructures. The aim of this paper is to provide such measurements through an accurate image segmentation algorithm that automatically delineates seven substructures of the heart from MRI and/or CT scans. Our proposed method is based on multi-planar deep convolutional neural networks (CNN) with an adaptive fusion strategy where we automatically utilize complementary information from different planes of the 3D scans for improved delineations. For CT and MRI, we have separately designed three CNNs (the same architectural configuration) for three planes, and have trained the networks from scratch for voxel-wise labeling for the following cardiac structures: myocardium of left ventricle (Myo), left atrium (LA), left ventricle (LV), right atrium (RA), right ventricle (RV), ascending aorta (Ao), and main pulmonary artery (PA). We have evaluated the proposed method with 4-fold-cross-validation on the multi-modality whole heart segmentation challenge (MM-WHS 2017) dataset. A precision and dice index of 0.93 and 0.90, and 0.87 and 0.85 were achieved for CT and MR images, respectively. Cardiac CT volume was segmented in about 50 s, with cardiac MRI segmentation requiring around 17 s with multi-GPU/CUDA implementation.

Keywords: Cardiovascular disorders · Computed tomography
Cardiac magnetic resonance imaging · Convolutional neural network
Whole heart segmentation

1 Introduction

According to the World Health Organization [1], cardiovascular diseases (CVDs) are the most common cause of death globally. About 17.7 million people died from CVDs in 2015, which was 31% of total global deaths from diseases. Almost

M. Pop et al. (Eds.): STACOM 2017, LNCS 10663, pp. 199–206, 2018.
https://doi.org/10.1007/978-3-319-75541-0_21

7.4 million of these deaths were due to CVDs and about 6.7 million were due to the stroke. Extensive research and clinical applications have shown that both CT and MRI have vital roles in non-invasive assessment of CVDs. CT is used more frequently than MRI due to its fast acquisition and cheaper cost. On the other hand, MRI has excellent soft tissue contrast and no ionizing radiation. However, most commercially available image analysis methods have been either tuned for CT or MRI only. Furthermore, many studies are focused on only one substructure of the heart (for instance, the left ventricle or left atrium). Surprisingly, there is very little published research on segmenting all substructures of the heart despite the fact that clinically established markers rely on shape, volumetric, and tissue characterization of all the cardiac substructures. Our study is concerned with this open problem from a machine learning perspective. We have investigated architectural designs of deep learning networks to solve multi-label and multi-modality image segmentation challenges within the scope of limited GPU processing power and limited imaging data.

Related Works. Literature related to cardiac image segmentation is vast. Among these works, atlas-based methods have been quite popular and favored for many years. For instance, multi-atlas based whole-heart segmentation using MRI and CT by [2] and atlas propagation based method using prior information by [3] are a few key examples. Despite their accuracy, those methods often lack efficiency due to heavy computations on the registration algorithms (e.g., from 13 min to 11 hours of computations reported in the literature). Interested readers can find a survey paper on cardiac image segmentation methods in [4] for a full list of methods and their comparative evaluations.

More recently, deep learning based approaches are replacing the conventional methods in medical image segmentation fields in general, and cardiac field in particular. For instance, in [5], a multi-planar deep learning has been utilized to segment LA and pulmonary veins from MR images. A recurrent fully convolutional neural network has been proposed to segment LV from MRI in [6]. In a similar fashion, a deep learning algorithm combined with a deformable-model approach was used to segment LV from MRI [7]. In [8], RV segmentation has been accomplished through a joint localization and segmentation algorithm within a deep learning framework. To date, the majority of deep learning methods have segmented only one or two structures of the heart and constrained to only one modality, unlike what is presented herein.

Our Contributions. We have constructed a network structure similar to the one devised in [5], which segments the left atrium and proximal pulmonary veins from MRI. In this paper, we have extended this segmentation engine in several different ways as follows. (1) A deeper CNN has been utilized as compared to [5]. (2) We have used both CT and MRI to test and evaluate the proposed system while Mortazi et al. used only MRI [5]. (3) We have extended the binary segmentation problem into a multi-label segmentation problem. (4) We have devised a rank based adaptive fusion method to assess effective information from different imaging planes for all delineated objects and select the best fusion strategies for highly accurate and efficient delineation results.

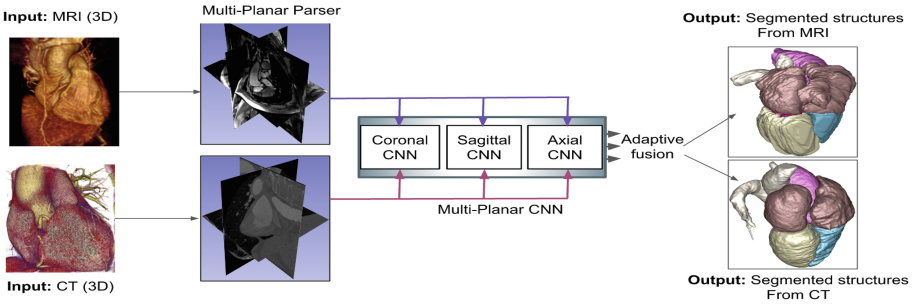


Fig. 1. Overall view of Mo-MP-CNN with adaptive fusion

2 Multi-Object Multi-Planar CNN (MO-MP-CNN)

The proposed method is called multi-object multi-planar convolutional neural networks (MO-MP-CNN), and its modules are illustrated in Fig. 1. MO-MP-CNN takes 3D CT or MR scans as an input and parses them into three perpendicular planes: Axial(A), Coronal(C), and Sagittal(S). For each plane (and modality), a 2D CNN is trained to label pixels. CNNs have been trained from scratch to adapt into CT and MRI context. After training each of the 2D CNNs separately, adaptive fusion strategy is utilized by combining the probability maps of each of the CNNs. The details of the CNN and adaptive fusion method are explained in the following.

CNN network. The proposed encoder-decoder based network architecture is illustrated in Fig. 2. Twelve *convolution* layers have been used in encoder and decoder separately. In the encoder part, two *max-pooling* layers have been used to reduce the dimension of the image by half and in decoder part two *upsampling* layers (bilinear interpolation) have been used to get the image back to its original size. The size of all filters were set as 3×3. Each convolution layer is followed by a *batch normalization* and *Rectified Linear Unit (ReLU)* as an activation function. The number of filters in the last convolution layer is equal to the number of classes (i.e., 8 (background + 7 objects)) and is followed by a *softmax* function to make a final probability map for each object. Similar to [5], the simplified z-loss [9] function has been used to train the network. To provide a sufficient number of

Table 1. Data augmentation parameters.

Data augmentation		
Methods	Parameters	
Zoom in	Scale $\varepsilon[1.1, 1.3]$	
Rotation	$k \times 45, k\,\varepsilon[-1, 1]$	
Training images (CT)		
CNN	# of images	Image size
Sagittal	40,960	350×350
Axial	21,417	350×350
Coronal	40,960	350×350
Training images (MRI)		
CNN	# of images	Image size
Sagittal	20,074	288×288
Axial	29,860	288×160
Coronal	19,404	288×288

training images for the networks, data augmentation has been applied to the training images by rotation and zoom-in operations. The details of the augmentation and the number of data for each CNN are summarized in Table 1.

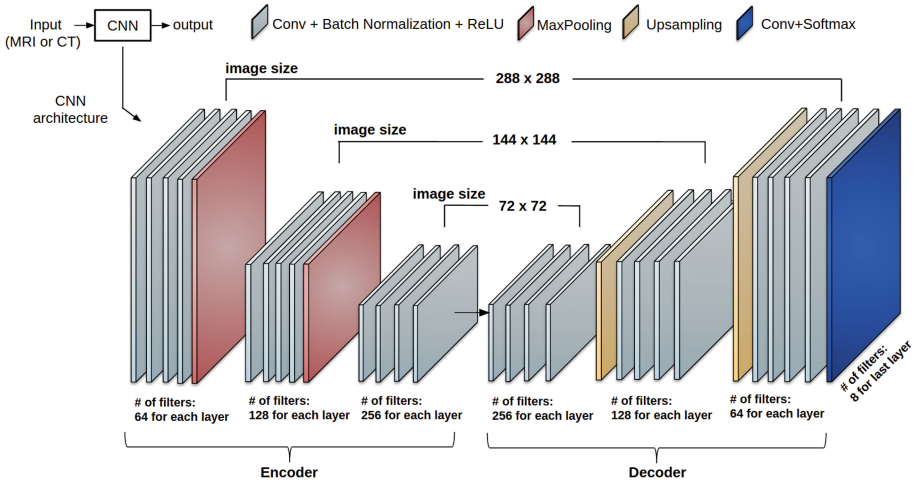


Fig. 2. Details of the CNN architecture. Note that image size is not necessarily fixed for each plane's CNN.

Multi-object Adaptive Fusion. An adaptive fusion strategy has been extended in the way that it can be applied to multi-object segmentation instead of binary segmentation. Let **I** and **P** denote an input and output image pair, where output is the probability map of the CNN. Also, let the final segmentation be denoted as **o**. As shown in Fig. 3, **o** is obtained from the probability map **P** by taking the maximum probability of each pixel in all classes (labels). Then, a connected component analysis (CCA) is applied to **o** to select reliable and unreliable regions, where unreliable regions are considered to come from false positive findings. Although this approach gives a "rough" estimation of the object, this information can well be used for assessing the quality of segmentations from different planes. If it is assumed that **n** is the number of classes (structures) in the images and **m** is the number of components in each class, then connected component analysis can be performed as follows: $CCA(\mathbf{o}) = \{o_{11}, \ldots, o_{nm} | \cup o_{ij} = \mathbf{o}$, and $o_{11}, \ldots, o_{nm} | \cap o_{ij} = \phi\}$. For each class **n**, we can now assign reliability parameters (weights) to increase the influence of planes that have more reliable (trusted) segmentations as follows: $w = \sum_i \{max_j\{|o_j|\}\} / \sum_{ij} |o_{ij}|$, where w indicates a weight parameter. In our interpretation of the CCA, the difference between trusted and non-trusted regions have been used to guide the reliability of the segmentation process: the higher the difference, the more reliable the segmentation (See Fig. 3, weight distribution w.r.t the difference). In test phase, we have simply used those predetermined weights from the training stage.

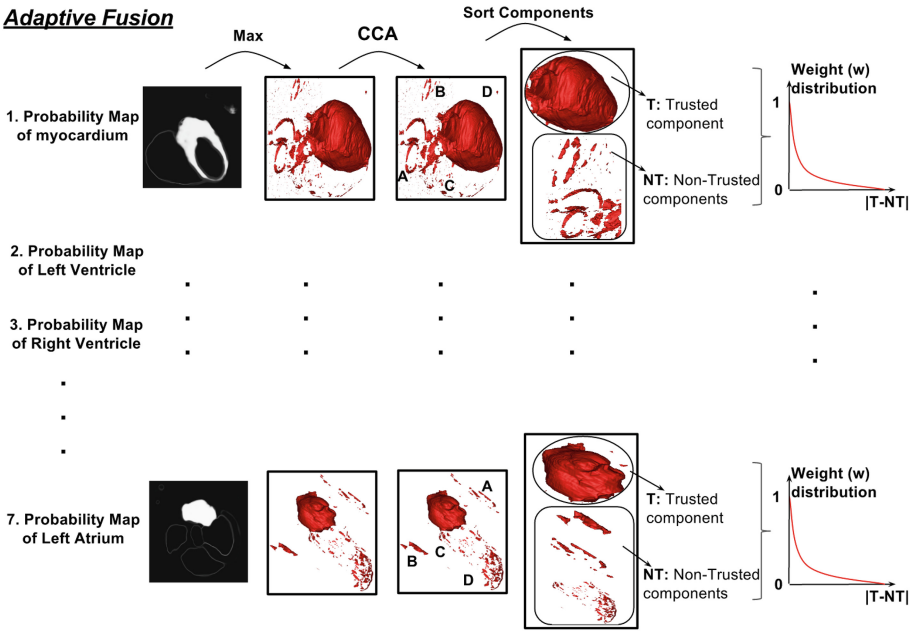


Fig. 3. Connected components obtained from each plane were computed and the residual volume (T-NT) was used to determine the strength for fusion with the other planes.

3 Experimental Results

Dataset and preprocessing: For the experiments and evaluations of the proposed method, we used the STACOM 2017 for whole heart segmentation challenge dataset, containing 20 MR and 20 CT images for training (with ground-truth) and 40 test images without ground-truth for each modality. We performed a 4 fold cross-validation on the dataset such that 15 subjects were used for training and 5 subjects have been chosen for validation for each fold. The CT images were obtained from routine cardiac CT angiography and to cover the whole heart, extending from the upper abdomen to the aortic arch. Axial in-plane resolution was 0.78×0.78 mm and slice thickness was 1.6 mm. The MR images were acquired by using 3D balanced steady state free precession (b-SSFP) sequences, with about 2 mm acquisition resolution in each direction. In preprocessing step, anisotropic smoothing filtering was applied to both CT and MR images prior to segmentation. In addition, histogram matching was used for MR images to alleviate intensity non-standardness issues.

Evaluation: Five metrics were assessed: sensitivity, specificity, precision, dice index (DI), and surface to surface (S2S) distance. A summary of the findings for each structure and also for the whole heart are reported in Table 2. The WHS is the average of all structures. The box-plot for sensitivity, precision, and DI

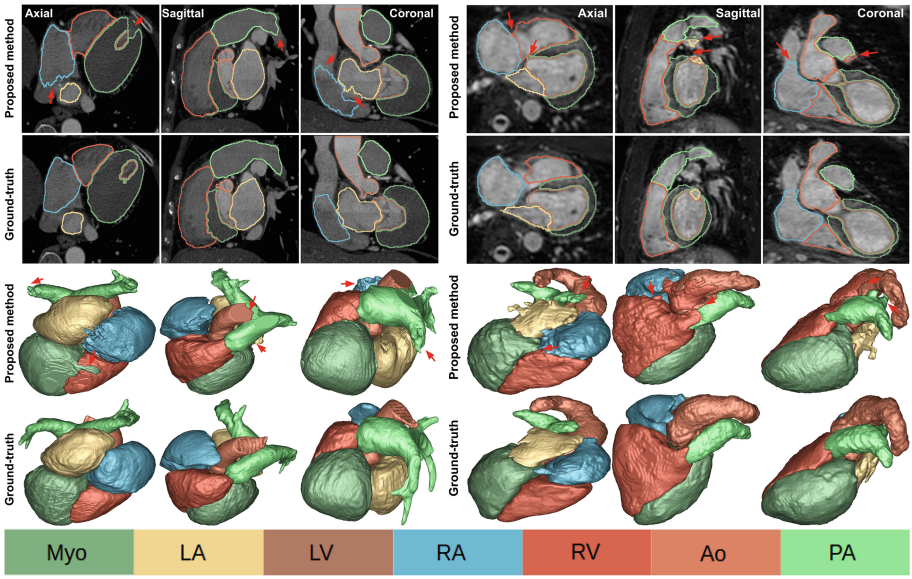


Fig. 4. First two rows show axial, sagittal, and coronal planes of the CT (first three columns) and MR images (last three columns), annotated cardiac structures, and their corresponding surface renditions (last two rows). Red arrows indicate some of mis-segmentations. (Color figure online)

Table 2. Quantitative evaluations of the proposed segmentation method for both CT and MRI are summarized.

	MRI								***CT***							
Structures:	*Myo*	*LA*	*LV*	*RA*	*RV*	*Aorta*	*PA*	*WHS*	*Myo*	*LA*	*LV*	*RA*	*RV*	*Aorta*	*PA*	*WHS*
Sensitivity	0.816	0.856	0.928	0.827	0.878	0.728	0.782	0.831	0.888	0.903	0.918	0.835	0.872	0.86	0.783	0.866
Specificity	0.999	0.999	0.999	0.998	0.999	0.999	0.998	0.999	0.999	0.999	0.999	0.999	0.998	0.999	0.999	0.999
Precision	0.842	0.88	0.936	0.909	0.846	0.873	0.791	0.868	0.912	0.931	0.944	0.914	0.911	0.983	0.912	0.929
DI	0.825	0.887	0.932	0.874	0.884	0.772	0.784	0.851	0.898	0.925	0.93	0.877	0.888	0.909	0.851	0.897
S2S(mm)	1.152	1.130	1.084	1.401	1.825	1.977	2.287	1.551	0.903	1.386	1.142	2.019	1.895	1.023	1.781	1.450

for both CT and MRI and for all structures are shown in Fig. 5. The qualitative results (including difficult cases for segmentation) for CT and MR modalities are illustrated in Fig. 4. Algorithms were implemented on the Nvidia TitanXp GPUs using Tensorflow [10]. The average time for segmenting the whole heart from the cardiac CT volume using three TitanXp GPUs was about 50 s. Segmenting using the cardiac MR volume took about 17 s. For comparison, the time on the Intel Xeon Processor E5-2620 with 8 cores for CT images was about 30 min and for MR images was about 8 min.

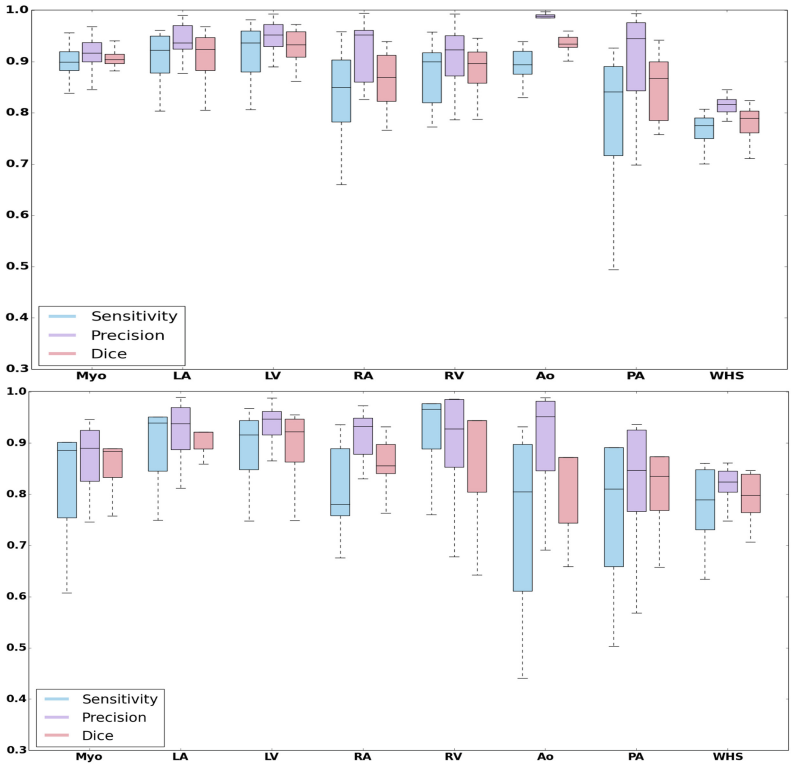


Fig. 5. Box plots for sensitivity, precision, and Dice index for each structure and WHS. Top figure is for CT dataset and bottom figure is for MR dataset

4 Discussion and Conclusion

The main goal of the current study is to develop a framework for accurate and efficient segmentation of all cardiac substructures from both cardiac CT and MR images. The main strength of the proposed method is to train multiple CNNs from scratch and to allow an adaptive fusion strategy for information maximization in pixel labeling despite the limited data and hardware support. Our findings indicate that MO-MP-CNN can be used as an efficient tool to delineate cardiac structures with high precision, accuracy, and efficiency.

Technically, one may question why we did not employ a completely 3D CNN approach instead of utilizing a multi-planar fusion of multiple 2D CNNs. As discussed in [5], the lack of a large number of 3D images restricts the depth of CNN training, which may result in sub-optimal implementation. Hence, training large number of 2D slices is much more feasible than utilizing a 3D approach with the current algorithm. In the instance of plentiful GPU processing power and 3D imaging data, training would be optimized using a 3D CNN.

Another limitation of our work stems from the use of the *softmax* function in the last layer of the proposed network. To explore whether the information loss due to class normalization in this step is significant, further research should be undertaken using information from the layer before the *softmax* in fusion part and with comparison to the current system. Finally, further work is needed to establish comparative evaluation of different deep neural network approaches such as *ResNet*, *U-net*, and others. While deeper networks are desirable to achieve higher precision in segmentation tasks, lack of 3D data is a significant limitation for training such a system. Data augmentation and transfer learning have been shown to adequately address such challenges to a certain degree, but there is currently no research proving the optimality of such networks relative to the availability of data at hand.

References

1. Cardiovascular Diseases (CVDs) (2007). http://www.who.int/mediacentre/factsheets/fs317/en/. Accessed 30 June 2017
2. Zhuang, X., Shen, J.: Multi-scale patch and multi-modality atlases for whole heart segmentation of MRI. Med. Image Anal. **31**, 77–87 (2016)
3. Zhuang, X., Ourselin, S., Razavi, R., Hill, D.L.G., Hawkes, D.J.: Automatic whole heart segmentation based on atlas propagation with a priori anatomical information. In: Medical Image Understanding and Analysis-MIUA, pp. 29–33 (2008)
4. Peng, P., Lekadir, K., Gooya, A., Shao, L., Petersen, S.E., Frangi, A.F.: A review of heart chamber segmentation for structural and functional analysis using cardiac magnetic resonance imaging. MAGMA **29**, 155–195 (2016)
5. Mortazi, A., Karim, R., Rhode, K., Burt, J., Bagci, U.: *CardiacNET*: segmentation of left atrium and proximal pulmonary veins from MRI using multi-view CNN. In: Descoteaux, M., Maier-Hein, L., Franz, A., Jannin, P., Collins, D.L., Duchesne, S. (eds.) MICCAI 2017. LNCS, vol. 10434, pp. 377–385. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-66185-8_43
6. Poudel, R.P.K., Lamata, P., Montana, G.: Recurrent fully convolutional neural networks for multi-slice MRI cardiac segmentation. arXiv preprint arXiv:1608.03974 (2016)
7. Avendi, M.R., Kheradvar, A., Jafarkhani, H.: A combined deep-learning and deformable-model approach to fully automatic segmentation of the left ventricle in cardiac MRI. Med. Image Anal. **30**, 108–119 (2016)
8. Luo, G., An, R., Wang, K., Dong, S., Zhang, H.: A deep learning network for right ventricle segmentation in short-axis MRI. In: 2016 Computing in Cardiology Conference (CinC), pp. 485–488. IEEE (2016)
9. de Brébisson, A., Vincent, P.: The Z-loss: a shift and scale invariant classification loss belonging to the spherical family. arXiv preprint arXiv:1604.08859 (2016)
10. Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G.S., Davis, A., Dean, J., Devin, M., et al.: TensorFlow: large-scale machine learning on heterogeneous distributed systems. arXiv preprint arXiv:1603.04467 (2016)

Local Probabilistic Atlases and a Posteriori Correction for the Segmentation of Heart Images

Gaetan Galisot(✉), Thierry Brouard, and Jean-Yves Ramel

LI Tours, Université Francois-Rabelais,
64 Avenue Jean Portalis, 37000 Tours, France
{gaetan.galisot,thierry.brouard,jean-yves.ramel}@univ-tours.fr

Abstract. Atlas-based segmentation is a well-known method for segmentation of medical images. In particular, this method could be used in an efficient way to automatically segment heart structures in MRI or CT scans. We propose, in this paper a more adaptive and interactive atlas-based segmentation method. The model presented combines several local probabilistic atlases with a topological graph. The local atlases provide more refined information about the structures' shape while the spatial relationships between the atlases are learned and stored in a graph. Hence, local registrations need less computational time and the image segmentation can be guided by the user in an incremental way. Following this step, a pixel classification is performed with a hidden Markov random field that integrates the learned a priori information with the pixel intensities that originate from different modalities. Finally, an a posteriori correction is performed using Adaboost classifiers in order to correct voxels in the border of the seek region and improve the precision of the results. The proposed method is tested on CT scan and MRI images of the heart coming from the MM-WHS challenge.

1 Introduction

Whole heart segmentation is an important problem in medical image analysis. As for the other applications of medical image segmentation, the previous work are mostly based on landmarks detection, atlas-based segmentation or learning-based segmentation [1–3]. The atlas-based methods are mainly used for the segmentation of anatomical structures. They provide a priori spatial information that can then drive lots of different methods. The probabilistic atlas is one kind of atlas composed of an average MRI image denoted as *template* and membership probability maps for each region. It has been used successfully in several articles [4,5]. In this article, we proposed a semi-automatic method of segmentation of the whole heart images which can be driven by the user to increase the adaptability. This method was build for the segmentation of MRI brain images. A training database is used to create different local probabilistic atlases for each anatomical structure. They are only defined on a subpart of the image

M. Pop et al. (Eds.): STACOM 2017, LNCS 10663, pp. 207–214, 2018.
https://doi.org/10.1007/978-3-319-75541-0_22

around the anatomical structures. Spatial relationships are also learned in order to automatically position the atlases based on the previously detected structures. A Markov random field is used for the voxel classification using the information from the probability maps. Finally, an a posteriori correction is performed with Adaboost classifiers. The proposed method is described in Sect. 2. The method is finally evaluated, in Sect. 3, on two datasets of heart images (MRI and CT) from the MM-WHS challenge.

2 Methods

2.1 Construction of the a Priori Information

A priori information is made of atlases and spatial relationships. It is represented by a topological graph where the nodes (atlases) describe the anatomical structures and the edges represent the spatial relationships between the structures. Now, let's see how such a graph is constructed.

Local Probabilistic Atlases. The construction of the local atlases is performed in several steps (cf. Fig. 1) starting from training data composed of N couples of medical images (MRI or CT scan) and associated labeled images.

Region delineation: The process of atlas creation is initialized by the delineation of the bounding box associated to each region represented in the training dataset. Based on the N available labeled images and for each region r, the volume inside the bounding box of r and the volume associated in the real image (MRI or CT scan) are extracted and denoted L_r and B_r, respectively. A margin is added around each bounding box in order to better tolerate the possible variability; i.e., smoothing the edges inside the local atlas. This margin is a percentage of the real size of the bounding box. Throughout this paper, the bounding box will refer to this extended bounding box.

Normalization: An intensity normalization is applied on the N sub-images B_r. The method described in [10] is used and transformed the intensity of the sub-images to an average histogram. This intensity normalization is local and achieved separately for each region r available in the training dataset.

Reference image selection: The construction of a probabilistic atlas needs a reference space which is obtained, in this case, with the choice of a reference sub-image among the sub-images B_r. The sub-images minimizing the Euclidean distance to all the others is selected as the reference for the local atlas of the region r. The couple of reference images of the region r is denoted as L_r^0 and B_r^0. Both sub-images are used to compute the first iteration of the template and the probability map of membership to r denoted as T_r^0 and P_r^0 respectively. T_r^0 is the same sub-images than B_r^0, P_r^0 is equal to one 1 if the voxel is labeled as *region* r in L_r^0, 0 otherwise.

Probability map and template construction: The template T_r and the probability map P_r are built incrementally. The transformations are done region by region and image by image considering all the available images in the training set. The atlas at the iteration $I+1$ is performed with the registration of the Ith couple of sub-images (B_r and L_r) to T_r^I and P_r^I. First, the labeled sub-images L_r is registered to the probability map P_r^I in two steps. A linear transformation (for a dimension adjustment) followed by a nonlinear registration is performed using Bsplines [7]. The transformation applied to B_r is also applied to T_r^I. Then, the template T_r^{I+1} and the probability map P_r^{I+1} are updated by averaging the current value with the registered information. The local atlas of the region r is built when this process is done for all the sub-images.

At the end of the process, each couple of images $\{T_r, P_r\}$ describes the local atlas of the considered region r and is stored inside the node of the graph corresponding to that region.

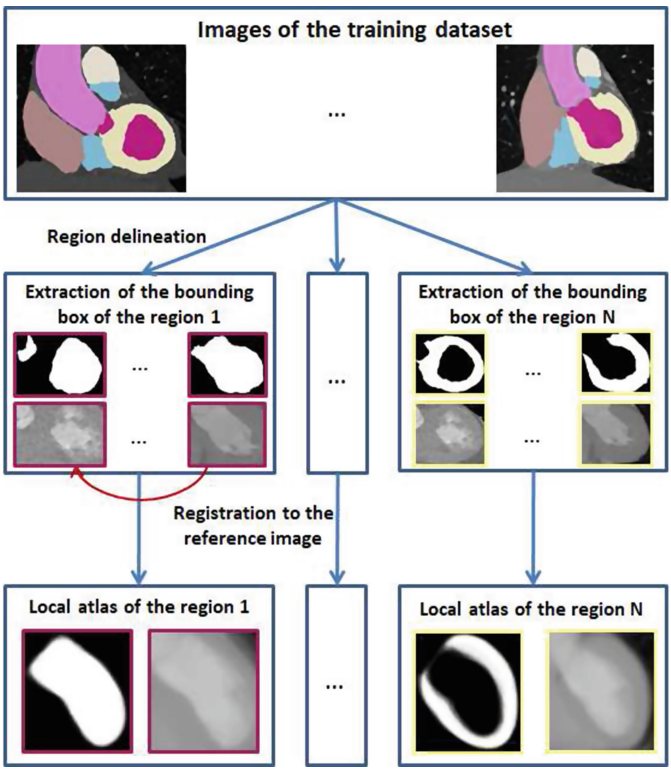


Fig. 1. Atlas construction (CT scan)

Topological Information. By using local atlases instead of a global one, the position and size of the regions are lost. In order to store this information, spatial relationships between each region are learned and incorporated into the edges of the so-called *topological graph*. The goal of these spatial relationships is to position the local atlases in a good way. It means that we have to determine where the borders of the regions are. So, in 3D, no less than twelve distances between two structures to be linked are learned and stored in one graph edge. The distance values are stored as relative distances compared to the size of the source region, making the relation independent of the space reference of the image. For each of the 12 spatial relationships, the minimum and maximum relative distances observed in the training set, are stored as an interval. These spatial relationships allow positioning the bounding box of a target region from the position of all the regions previously localized in the image.

2.2 Segmentation

The segmentation of an image is performed in an incremental way using all the information learned in the topological graph. First, the user has to choose the region to segment among the ones available in the topological graph. Second, the bounding box of the region is positioned. Third, the local atlas is registered to the volume inside the bounding box. Finally, an hidden Markov random field (HMRF) is used to classify the voxels.

The position of the bounding box has to be defined in order to use a local atlas for the segmentation. This position can be defined automatically or manually. Because the spatial relationships are information between the regions, at least one region has to be already segmented in order to use this spatial information. So the bounding box of the first region is necessarily positioned manually. But, for the other regions, the spatial relationships provide a positioning of the bounding which can be corrected or validated by the user. A margin is also taken around the bounding box. The margin is the same percentage than during the creation of the atlas. At the moment, a volume V inside the extended bounding box is extracted for the following steps.

Probability map and template for region R are available in the nodes of the topological graph. The atlas is registered to Ve_R in the same way as during the atlas construction (cf. Fig. 1). But the first image registered is the template (MRI or CT scan) and then the same transformation is applied to the probability map. Each voxel among the volume Ve_R is now linked to a membership probability allowing to drive a classification process.

The classification of voxels is performed with a HMRF. It can merge the information coming from the atlas, the intensity of the image and also from the neighborhood of each voxel. The HMRF is initialized with a Kmeans in K classes on the voxels of Be_r; attributes are the intensity and the membership probability. One class is defined as the *region* class, the others are defined as *non-region* class. The class defined as *region* is the class which has the highest number of voxels whose membership probability to the region R is important. The atlas information is given through an external field which is equal to $-log(atlas_i)$ for

the *region* class and equal to $-log(1-atlas_i)$ for the other classes defined as *non-region*. The intensity of each class is also modeled by a Gaussian. An expectation maximization (EM) algorithm is then used in order to maximized the likehood by optimizing the parameters of the Gaussians. At the end of this process, the voxel belonging to the class defined as region are definitively assigned to the region R.

2.3 A Posteriori Correction

The authors in [6] described a method for the automatic correction of systematic segmentation errors from a host method using several classifiers that reclassify misclassified voxels. This correction is applied on the result of a previous segmentation. The correction uses *Adaboost* [8] classifiers which are based in several weak classifiers. Each weak classifier is corresponding to a threshold on a selected feature and is merged. In our case, each region r is associated to a couple of classifiers: one dedicated to the background and one dedicated to the foreground (anatomical structure). The attributes of each voxel are computed on a patch around this voxel with a size determined by the user (usually $5 \times 5 \times 5$). The attributes of one voxel are composed of the relative intensity, distance from the centroid and the membership probability from the HMRF of each voxel among the patch.

The training is performed independently for each region. In this case, it is learned on the same database used for the topological graph construction. The training database is segmented with the method described in the previous section. The probability maps from the HMRF are used as input for the training of the Adaboost classifiers. The ground truth and the image (MRI or CT) are also needed. The classifier of the region r is trained on a volume of interest (VOI), defined by the thresholding of the probability map at 0.5 and by a dilation. The background classifier of the region r is trained on the intersection of the VOI and a dilated volume around de background.

During the classification, the same volumes are computed by thresholding the probability maps. Each Adaboost classifier of the region $r \in R$ is used on its own VOI. It provides a probability of membership to the region P_r and to the background $P_{\overline{r}}$. Let R be the set of all the previously segmented regions; the a posteriori probability value (and associated region label L_r) associated to the voxel x is defined as follows:

$$R_{Max}(x) = \arg\max_{r \in R} P_r(x)$$

Thus, if the obtained region membership probability P_r is higher than the non-region membership probability $P_{\overline{r}}$, then the voxel belongs to the region L_r.

$$I(x) = \begin{cases} R_{Max}(x) & if\ P_{R_{Max}}(x) > P_{\overline{R_{Max}}}(x) \\ 0 & if\ P_{R_{Max}}(x) \leq P_{\overline{R_{Max}}}(x) \end{cases}$$

A small adjacency correction is also added after the computation of the Adaboost correction step. Empty space can be present between two adjacent

regions because of the incremental segmentation. In order to limit this effect, if a voxel is surrounded by a lot of voxels which are already classified to a region, the probability to belong to a region is promoted compared to the background classifier. If Nb_{class} is the number of voxels classified as *region* inside the patch, Nb_{patch} is the total number of voxels inside the patch. So we defined :
$voxRatio = \frac{Nb_{class}}{Nb_{patch}}$

$$I(x) = \begin{cases} R_{Max} \; if P_{R_{Max}} > (\overline{P_{R_{Max}}} * (1 - voxRatio)) \\ 0 \quad if P_{R_{Max}} \leq (\overline{P_{R_{Max}}} * (1 - voxRatio)) \end{cases}$$

This operation has to be performed iteratively after each segmentation of a new region.

3 Experiments

The experiments were performed on the dataset coming from the MM-WHS challenge. The database is composed of 20 MRI images and 20 CT scans images of the heart with the associated labeled images. Seven regions are labeled in the ground truth with the left ventricle blood cavity (LV), the right ventricle blood cavity (RV), the left atrium blood cavity (LA), the right atrium blood cavity (RA), the myocardium (Myo) of the left ventricle, the ascending aorta (AA) and the pulmonary artery (PA). Each dataset was divided into four datasets of 5 images for the cross validation. All the a priori information, including the local atlases, the topological information and the Adaboost classifiers, was learned on three datasets. The MRI a priori information were used to segment the five other MRI images which were not used during the training steps. The same process was performed for the CT images.

The proposed method is interactive and needs the user assistance. In our case and with the ground truth, two experiments were performed in order to test separately the local atlases and topological information. In the first case (E1), all the bounding boxes of all the regions are positioned with the ground truth in order to simulate a perfect interactive positioning of the borders of the region by an expert user. The second experiment (E2) is done with only one region (the first one, left ventricle) positioned. The other regions are positioned with the learned spatial relationships.

The metric used during registration is the mutual information for CT images and mean square distance for the MRI images. Intensity values are scaled from 0 to 255 by the normalization step. In order not to lose information for the CT scan and nor take into account the void outside. For the CT scan, in order to not lose too much information and ignore the outside parts, the intensities below -800 and above $+800$ are respectively set to -800 and 800 before the normalization. The coefficient of the Markov random field are set for the MRI and the CT scan. The number of class K is set to 3. The size of the dilation for the VOI of the correction is set to 3 for the MRI images and 1 for the CT scan because of the size of the CT images.

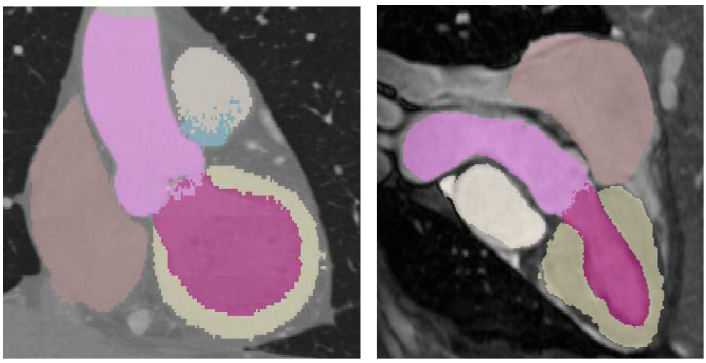
4 Results

The Table 1 shows the results on the CT images for both experiments. When the bounding boxes are positioned with the labeled images, the Dice ratio is above 88 % for six regions. The important deviation in the Dice results is because of some images with no segmentation coming from a wrong registration of the atlas. With the spatial relationships, the segmentation is decreased especially for the pulmonary artery (PA) but remain close to the results from the experiment E1 for the other regions. The variability of the arteries is sometimes not well managed by the spatial relationships compared to the ventricles and atrium cavities. Figure 2a shows the median case for the segmentation of the CT images. As explained in the qualitative results, the segmentation of most of the anatomical structures seems satisfactory. Only the borders of the regions is not smooth enough. The results on MRI images have Dice ratio lower than CT scan especially for the myocardium and the left atrium. The variability seems to be higher in the MRI images. The local registration is sometimes not enough accurate. The

Table 1. Dice Ratio

Regions	LV	RV	Myo	RA
E1 CT	88.2 ± 21.1	88.8 ± 3.8	88.2 ± 3.5	89.0 ± 4.4
E2 CT	88.1 ± 21.6	86.3 ± 7.4	87.4 ± 7.9	81.3 ± 20.2
E1 MR	88.4 ± 6.5	82.4 ± 12.4	75.5 ± 9.7	85.3 ± 5.6

Regions	AA	PA	LA
E1 CT	88.1 ± 0.21	82.0 ± 12.1	92.3 ± 3.5
E2 CT	81.3 ± 20.2	72.3 ± 24.8	86.0 ± 19.4
E1 MR	77.1 ± 10.9	67.7 ± 15.5	78.4 ± 5.4



(a) (b)

Fig. 2. Segmentation images for the (2a) median case of the CT images, (2b) median case of the MR images

Fig. 2b shows the median case of the segmentation. In this image, the segmentation is globally correct but most of the regions have a volume lower than it really is. Without optimization, the whole segmentation process takes about 1–4 min per region, depending of the image size and the region size.

5 Conclusion

In this paper, a new incremental segmentation method based on local atlases was presented. This segmentation technique allows the user to perform partial and incremental segmentation. It can be applied on a various types of medical imaging. Experimental results have been done on the images of the MM-WHS challenge. The quality of the segmentation is satisfactory on the CT scan images even when the spatial relationships are used. The segmentation of the MRI images is not as good as the segmentation of the CT images. The registration of the MRI atlases can be inaccurate. The segmentation results remain lower than [1] but it could be interesting to use also a Joint Label Fusion method [9] rather than the probabilistic atlas and HMRF. It could also be interesting to make a better optimization in order to improve the computational time.

References

1. Zhuang, X., Shen, J.: Multi-scale patch and multi-modality atlases for whole heart segmentation of MRI. Med. Image Anal. **31**, 77–87 (2016)
2. Zhuang, X., Rhode, K., Razavi, R., Hawkes, D., Ourselin, S.: A registration-based propagation framework for automatic whole heart segmentation of cardiac MRI. IEEE Trans. Med. Imaging **29**, 1612–1625 (2010)
3. Bai, W., Shi, W., Lediq, C., Rueckert, D.: Multi-atlas segmentation with augmented features for cardiac MR images. Med. Image Anal. **19**, 98–109 (2014)
4. Fischl, B., Salat, D.H., Busa, E., Albert, M., Dieterich, M., Haselgrove, C., Van Der Kouwe, A., Killiany, R., Kennedy, D., Klaveness, S., Montillo, A., Makris, N., Rosen, B., Dale, A.M.: Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. Neuron **33**, 341–355 (2002)
5. Scherrer, B., Forbes, F., Garbay, C., Dojat, M.: Distributed local MRF models for tissue and structure brain segmentation. Trans. Med. Imaging **28**, 1278–1295 (2009)
6. Wang, H., Das, S.R., Altinay, M., Pluta, J., Suh, J., Avants, B., Yushkevich, P.: A general-purpuse learning-based wrapper method to correct systematic errors in automatic image segmentation: consistently improved performance in hippocampus, cortex brain segmentation. NeuroImage **55**(3), 968–985 (2011)
7. Fornefett, M., Rohr, K., Stiehl, H.S.: Radial basis functions with compact support for elastic registration of medical images. Image Vis. Comput. **19**, 87–96 (2001)
8. Freund, Y., Schapire, R.: A decision-theoretic generalization of on-line learning and an application to boosting. In: Proceedings of the 2nd European Conference on Computational Learning Theory, vol. 55(3), pp. 23–27 (1995)
9. Wang, H., Yushkevich, P.A.: Multi-atlas segmentation with joint label fusion and corrective learning - an open source implementation. Front. Neuroinformatics **7**(27), 27 (2013)
10. Nyul, L.G., Udupa, J.K., Zhang, X.: New variants of a method of MRI scale standardization. Trans. Med. Imaging **19**, 143–150 (2000)

Hybrid Loss Guided Convolutional Networks for Whole Heart Parsing

Xin Yang[1(✉)], Cheng Bian[2], Lequan Yu[1], Dong Ni[2], and Pheng-Ann Heng[1,3]

[1] Department of Computer Science and Engineering, The Chinese University of Hong Kong, Sha Tin, Hong Kong
xinyang@cse.cuhk.edu.hk

[2] National-Regional Key Technology Engineering Laboratory for Medical Ultrasound, School of Biomedical Engineering, Health Science Center, Shenzhen University, Shenzhen, China

[3] Shenzhen Key Laboratory of Virtual Reality and Human Interaction Technology, Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen, China

Abstract. CT and MR are dominant imaging modalities in cardiovascular inspection. Segmenting the whole heart from CT and MR volumes, and parsing it into distinctive substructures are highly desired in clinic. However, traditional methods tend to be degraded by the large variances of heart and image, and also the high requirement in simultaneously distinguishing several substructures. In this paper, we start with the well-founded Fully Convolutional Network (FCN), and closely couple the FCN with 3D operators, transfer learning and deep supervision mechanism to distill 3D contextual information and attack potential difficulties in training deep neural networks. We then focus on a main concern in our enhanced FCN. As the number of substructures to be distinguished increases, the imbalance among different classes will emerge and bias the training towards major classes and therefore should be tackled seriously. Class-balanced loss function is useful in addressing the problem but at the risk of sacrificing the segmentation details. For a better trade-off, in this paper, we propose a hybrid loss which takes advantage of different kinds of loss functions to guide the training procedure to equally treat all classes, and at the same time preserve boundary details, like the branchy structure of great vessels. We verified our method on the MM-WHS Challenge 2017 datasets, which contain both CT and MR. Our hybrid loss guided model presents superior results in concurrently labeling 7 substructures of heart (*ranked as second in CT segmentation Challenge*). Our framework is robust and efficient on different modalities and can be extended to other volumetric segmentation tasks.

1 Introduction

Cardiovascular disease is the leading cause of death in the world. CT and MR are dominant imaging modalities for noninvasive cardiovascular anatomy inspection.

X. Yang and C. Bian—Contributed equally to this work.

M. Pop et al. (Eds.): STACOM 2017, LNCS 10663, pp. 215–223, 2018.
https://doi.org/10.1007/978-3-319-75541-0_23

Segmenting the whole heart, especially labeling pre-defined substructures, is fundamental for the following structural and functional analysis, diagnosis and treatment. Affected by experience level and huge scale of volumetric data, manual delineation is severely confined in low time efficiency, low inter- and intra-expert reproductivity. Featured with efficiency and promising accuracy, automatic solutions to segment the whole heart are becoming viable in the field.

Concurrently labeling all the voxels in cardiovascular volume with pre-defined categories (8 classes in this paper) retains intensive research interest. However, previous automatic segmentation methods tend to be hindered by many factors, such as the highly varying structure of heart, low inter-class appearance variance and boundary uncertainty. Deformable models associated with statistical shape and texture abstract have been proposed to segment chambers and vessels [5,10]. Other popular streams are the multi-atlas based method [1] and non-rigid registration based method [18]. However, those model based methods convey the difficulties in designing discriminative descriptors and constructing the specific model on limited annotation data. The huge surge of deep neural networks gets less dependency on pre-defined modeling and pushes the bound of cardiovascular volume segmentation. 2D Fully Convolutional Network (FCN) and its variants have been actively studied to segment heart in MR images [15,19]. By substituting 2D operators with 3D ones, 3D FCN and its extension [3,9] earn population for medical image segmentation. By coupling with transfer learning [2] and deep supervision [8] techniques, 3D FCN presents superior performance for whole heart segmentation in MR volumes [4,17]. However, as the number of class to categorize increases, the imbalance among different classes will emerge. Subjecting to the improper guidance of loss function, the training of networks may be biased only towards major classes and ignore the minor classes [9].

In this paper, we propose a fully automatic solution for whole heart partition, which can simultaneously decompose the whole heart into 7 substructures. We start with a tailored 3D FCN to construct an efficient end-to-end mapping and thoroughly exploit 3D context. We closely couple the 3D FCN with transfer learning and deep supervision mechanism to attack potential training difficulties caused by overfitting and vanishing gradient. We then focus on the main concern in tackling the class imbalance. Motivated by [9,16], to address the problem of class imbalance, we investigate the characteristics of different loss functions, and propose a hybrid loss function, which blends weighted cross-entropy (*wCross*) loss and our proposed multi-class Dice Similarity Coefficient (*mDSC*) loss, to fairly guide the training process. We verified our method on the MM-WHS Challenge 2017 datasets, including CT and MR. Our method achieves promising results in concurrently labeling 7 substructures of heart (*ranked as second in CT segmentation Challenge*) and a proper compromise between balancing distinctive classes and preserving segmentation details. Our framework is robust and efficient on different modalities and can be extended to other volumetric segmentation tasks.

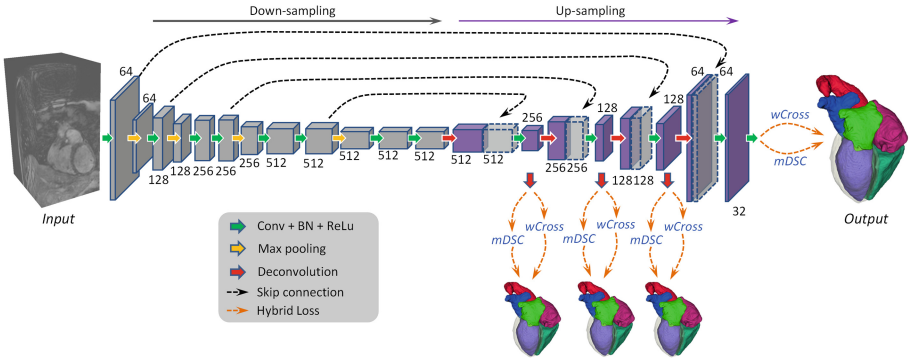


Fig. 1. Schematic view of our proposed framework.

2 Methodology

Figure 1 is the schematic illustration of our proposed framework. Without any auxiliary heart localization module, our system takes the original whole volume of heart as input. There is a preprocessing module to conduct intensity calibration. Our tailored 3D FCN originates from the famous U-net [12], the down-sampling path benefits from transfer learning. The proposed hybrid loss function are adopted in a stratified deep supervision format. The output of our framework is the volume-wise labeling result for 7 substructures of heart.

2.1 Intensity Calibration as Preprocessing

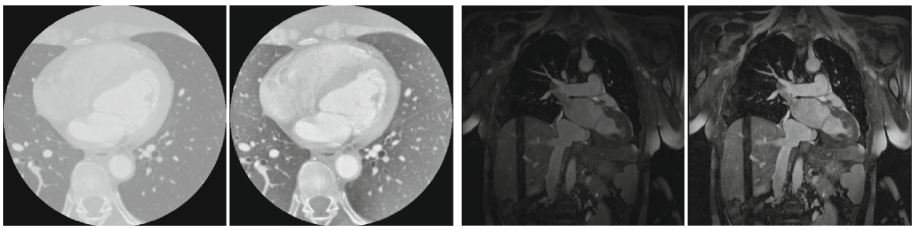
Because the Challenge datasets are collected on different subjects in different sites, the image quality are varying greatly and subject to imaging parameters and machines. Low contrast and inhomogeneity are common around the volumes. So, we adopt the Contrast Limited Adaptive Histogram Equalization (CLAHE) technique [11] to enhancing local contrast and reduce the inhomogeneity. Specifically, we apply CLAHE in each slice and set the block size as 8×8. Shown as Fig. 2, with CLAHE, slices in CT and MR get significant visual quality improvement.

2.2 Enhance the Training of 3D FCN

Shown as Fig. 1, we customize a 3D FCN from [3] with skip connections and take it as the core of our design. We enhance the 3D FCN with transfer learning and deep supervision to promote the training from different aspects.

Transfer Learning from Video Recognition. For vision tasks, the filters learned by shallow layers of neural networks tend to be general across different tasks. When facing with limited training data, leveraging parameters of

well-trained model proves to be beneficial in improving generalization ability of networks, even the pre-trained model is generated in a different domain [2]. Popular models, like ImageNet [7] and VGG16 [13] are only tractable for 2D applications. The C3D model proposed in [14] for video recognition sheds light on transfer learning in 3D networks, since it is equipped with 3D convolution operators to concurrently extract spatial and temporal abstract across consecutive frames. Therefore, we initialize the downsampling path of our 3D FCN with the same parameters from layers *conv1, conv2, conv3a, conv3b, conv4a and conv4b* in C3D model. The configuration of our upsampling path is symmetric with downsampling path but is initialized from normal distribution.



(a) (b)

Fig. 2. Intensity calibration with CLAHE. (a) CT slice without and with CLAHE, (b) MR slice without and with CLAHE

Stratified Deep Supervision. Gradient vanishing problem associated with deep neural networks often causes the early layers to be under-tuned [6], where the fundamental representations are extracted. Enhancing the gradient flow for shallow layers with the deep supervision mechanism proves to be effective in improving the segmentation [4,8]. Shown as Fig. 1, the core idea of deep supervision is to add several stratified side-paths and thus shorten the backpropagation path of gradient flow and expose shallow layers to more direct supervision. Because the feature maps in shallow layers contains inadequate semantic information for segmentation, it is not suggested to attach side-paths in downsampling path.

Let $\mathcal{X}^{w\times h\times d}$ be the volumetric input, W be the weights of main network, $w = (w^1, w^2, .., w^{\mathcal{S}})$ be the weights of side-paths and w^s denotes the weight of the s^{th} side-path. $\tilde{\mathcal{L}}$ and $\mathcal{L}_s$ are main loss function and loss function in s^{th} side-path, respectively. Both $\tilde{\mathcal{L}}$ and $\mathcal{L}_s$ are further composed of hybrid loss functions which will be detailed in Sect. 2.3. The final loss function $\mathcal{L}$ for our 3D FCN with stratified deep supervision is elaborated in Eq. 1, where β_s is the weight of different side-paths.

$$\mathcal{L}(\mathcal{X}; W, w) = \tilde{\mathcal{L}}(\mathcal{X}; W) + \sum_{s\in\mathcal{S}} \beta_s \mathcal{L}_s(\mathcal{X}; W, w^s) + \lambda(||W||^2 + \sum_{s\in\mathcal{S}} ||w^s||^2) \quad (1)$$

2.3 Hybrid Loss Guided Class-Balanced Segmentation

The feedback generated by loss function determines the latent mapping that the deep neural network needs to fit. Considering the feasibility in differentiable optimization, the popular choice for loss function is cross entropy. However, utilizing cross entropy based loss function is improper in classification or segmentation occasions where classes present significantly different proportions [16]. Class imbalance becomes more obvious in whole heart partition since structures, like left ventricle blood cavity and myocardium, often have less volume size than others. Classical cross-entropy based loss function trivially summarizes the error of each voxel without giving corresponding significances for specific classes, which will lead the network to oversee minor classes and only focus on major ones. Based on [9,16], in this paper, we conduct investigation on different loss functions in balancing different classes and preserving segmentation details, and finally propose a hybrid loss function as a optimal choice.

Volume Size Weighted Cross Entropy. As proposed in [16] for rare edge extraction, weighting the cross entropy loss for different classes is helpful in removing the imbalance. In this paper, we extend the formulation in [16], and propose a patch-wise weighted cross entropy (*wCross*). Mathematically, the formulation of *wCross* is shown as Eq. 2. $\mathcal{X}$ represents the training samples and $p(y_i = \ell(x_i)|x_i; W)$ is the probability of target class label $\ell(x_i)$ corresponding to sample $x_i \in \mathcal{X}$. $|\mathcal{X}^{\ell(x_i)}|$ is the volume size of class $\ell(x_i)$ in patch $\mathcal{X}$. Minor class get larger weight with $\eta_{\ell(x_i)}$.

$$\mathcal{L}_{wCross}(\mathcal{X}; W) = \sum_{x_i \in \mathcal{X}} -\eta_{\ell(x_i)} \log\, p(y_i = \ell(x_i)|x_i; W), \eta_{\ell(x_i)} = 1 - \frac{|\mathcal{X}^{\ell(x_i)}|}{|\mathcal{X}|} \quad (2)$$

Multi-class Dice Similarity Coefficient. Dice Similarity Coefficient (DSC) based loss function is another novel attempt to alleviate class imbalance [9]. DSC based loss function roots in the metric of global shape similarity and thus the cost for each class is self-normalized before being equally counted into the total loss. We extend the loss in [9] and propose a differentiable multi-class Dice Similarity Coefficient (*mDSC*) based loss function to balance the training for multiple classes. Given the segmentation ground truth $G^{w \times h \times d}$, we firstly encode it into a one-hot format for C classes $\mathcal{G}^{C \times w \times h \times d}$, $C = 8$ for our task. With probability volumes $\mathcal{P}^{C \times w \times h \times d}$, our proposed *mDSC* can be written as

$$\mathcal{L}_{mDSC} = -\sum_{c \in C} \frac{\frac{2}{N} \sum_i^N \mathcal{G}_c^i \mathcal{P}_c^i}{\sum_i^N \mathcal{G}_c^i \mathcal{G}_c^i + \sum_i^N \mathcal{P}_c^i \mathcal{P}_c^i}, \quad (3)$$

where $N = w \times h \times d$, $\mathcal{G}_c^i$ and $\mathcal{P}_c^i$ are the i^{th} voxel of c^{th} volume in $\mathcal{G}$ and $\mathcal{P}$. The $1/N$ in denominator is introduced to suppress prediction noise. Improvement caused by *mDSC* is illustrated in Sect. 3. As illustrated in Sect. 3, both *wCross* and *mDSC* can reduce class imbalance. *wCross* often guides networks to preserve complex boundary details but bring about many noise, while *mDSC* tend to generate more compact and clear ones, but at sacrifice of losing branchy details. Therefore, we propose to blend this two kinds of complementary loss functions as a hybrid one, shown as Eq. 4, $\alpha = 100$, so as to get segmentation results in a compact but details-enhanced format.

$$\mathcal{L}_{hybrid} = \mathcal{L}_{wCross} + \alpha \mathcal{L}_{mDSC} \tag{4}$$

3 Experimental Results

Experiment Materials: We evaluated our network on Multi-Modality Whole Heart Segmentation Challenge 2017 datasets, segmenting whole heart in CT and MR volumes. The datasets consist of 60 CT and 60 MR volumes (contains 20 for training and 40 for testing). All the training samples are normalized as zero mean and unit variance. We augment the training dataset with 30% rotated samples. We trained two networks to segment the two modalities independently.

Implementation Details: We implemented our 3D FCN in *Tensorflow*, using 2 NVIDIA GeForce GTX TITAN X GPUs. *Code will be available upon publication*[1]. Given the limited memory of one GPU, we assign the down- and up-sampling paths to different GPUs. We update the weights of network with a Adam optimizer (batch size = 1, initial learning rate is 0.001). We totally attached 3 side-paths to the upsampling branch and set $\beta_0 = 0.2, \beta_1 = 0.4, \beta_2 = 0.8$ for feature maps from coarse to fine. Randomly cropped $96 \times 96 \times 96$ sub-volumes serve as input to train our network. To avoid shallow layers being over-tuned during fine-tuning, we set smaller initial learning rate for *conv1, conv2, conv3a, conv3b, conv4a and conv4b* as *1e–6, 1e–6, 1e–5, 1e–5, 1e–4, 1e–4*. We adopt sliding window with high overlapping ratio and overlap-tiling stitching strategies to generate predictions for the whole volume, and remove small isolated connected components in final labeling result.

Quantitative and Qualitative Analysis: We use 3 metrics to evaluate the proposed framework on segmentation, including DSC, Jaccard and Average Distance of Boundaries (ADB). Transfer learning (TL) and deep supervision (DS) are configured for both compared methods. Limited by time, we only conduct experiments to compare the model driven by *mDSC* loss function (denote as TL+DS+mDSC) and hybrid loss function (denote as TL+DS+hybrid).

[1] https://github.com/xy0806/miccai17-mmwhs-hybrid.

Table 1. Quantitative evaluation for whole heart segmentation in CT volume

Method	Metrics	Substructures of heart							Mean
		MLV	LABC	LVBC	RABC	RVBC	ASA	PUA	
TL+DS+mDSC	*DSC[%]*	68.97	90.00	83.42	84.36	62.50	91.51	80.84	80.22
	Jaccard[%]	55.67	82.19	73.33	74.05	49.31	85.03	68.87	69.77
	ADB[voxel]	3.185	5.415	5.666	5.875	8.245	2.692	3.875	**4.993**
TL+DS+hybrid	*DSC[%]*	81.86	84.54	87.76	81.53	77.80	94.12	82.62	**84.32**
	Jaccard[%]	69.93	76.12	78.66	70.28	65.90	89.20	71.13	**74.46**
	ADB[voxel]	2.987	22.67	4.609	6.502	8.609	2.237	5.086	7.529

Table 2. Quantitative evaluation for whole heart segmentation in MR volume

Method	Metrics	Substructures of heart							Mean
		MLV	LABC	LVBC	RABC	RVBC	ASA	PUA	
TL+DS+mDSC	*DSC[%]*	66.54	74.62	86.80	86.16	71.43	71.24	70.19	75.28
	Jaccard[%]	52.07	64.23	77.71	75.97	60.59	58.13	57.88	63.80
	ADB[voxel]	1.509	1.761	1.646	1.773	3.300	1.560	1.587	**1.864**
TL+DS+hybrid	*DSC[%]*	74.17	78.66	85.83	81.99	81.91	72.60	69.83	**77.86**
	Jaccard[%]	60.27	67.53	76.59	71.89	71.27	58.43	55.17	**65.88**
	ADB[voxel]	1.404	1.950	2.045	2.733	4.483	3.346	4.367	2.904

Because the ground truth of testing dataset is held out by the organizer for independent evaluation, we get our current evaluation results by taking 10 volumes from training dataset to train and another 10 volumes as validation. We denote the substructures as myocardium of the left ventricle (MLV), left atrium blood cavity (LABC), left ventricle blood cavity (LVBC), right atrium blood cavity (RABC), right ventricle blood cavity (RVBC), ascending aorta (ASA) and pulmonary artery (PUA). Quantitative results are shown in Tables 1 and 2. With the hybrid loss function, for both CT and MR, our proposed method gets significant improvement on DSC and Jaccard metrics, especially in classes which fail in *mDSC* based model. With segmentation for same CT volumes, we provide an explicit proof about how the *wCross* based model can preserve more branchy details, and how the *mDSC* based model get the compromise in Fig. 3. We also visualize the segmentation results in MR volumes. Our proposed method conquers complex variance of heart and achieves promising performance in both two modalities.

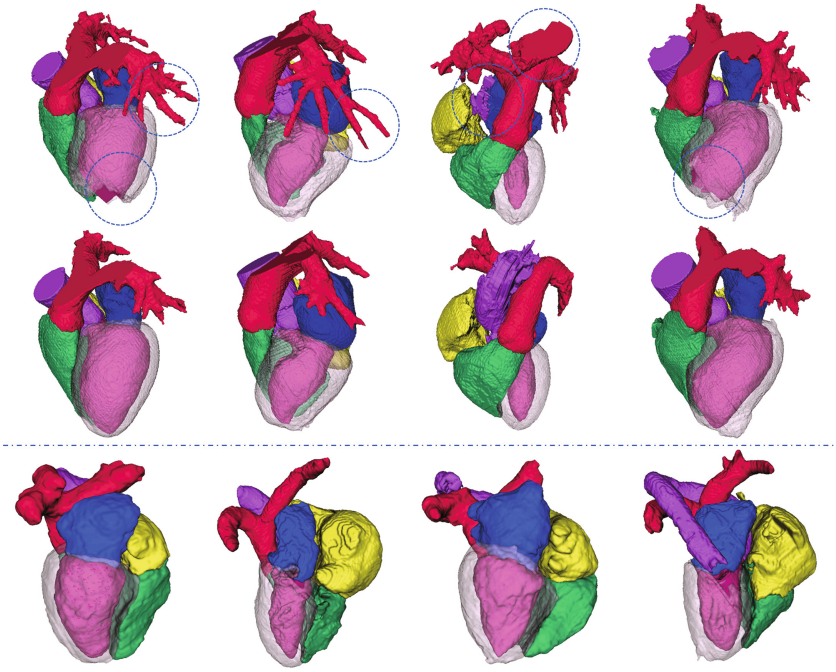


Fig. 3. Visualization of our segmentation results in testing datasets. From top to bottom: segmentation results for same CT volumes from *wCross* based model, hybrid loss based model and MR segmentation results from hybrid loss based model. Blue circles denote the branchy details enhanced by *wCross*, and also the flaws caused by *wCross*. (Color figure online)

4 Conclusions

In this paper, we present a fully automatic framework to partition the whole heart from CT and MR volumes into substructures. We first tailor and enhance the 3D FCN, and then investigate the impact of different loss functions in attacking class imbalance. *wCross* based loss function and our proposed *mDSC* prove to be complementary, and the final hybrid loss function gets a proper compromise in generating compact and detail enhanced segmentation. As validated on two modalities, our proposed hybrid loss function is promising for complete heart partition and applicable to other volumetric segmentation tasks.

Acknowledgments. The work in this paper was supported by the grant from National Natural Science Foundation of China under Grant 61571304, a grant from Hong Kong Research Grants Council (Project no. CUHK 412513), a grant from Shenzhen Science and Technology Program (No. JCYJ20160429190300857) and a grant from Guangdong province science and technology plan project (No. 2016A020220013).

References

1. Bai, W., Shi, W., Ledig, C., Rueckert, D.: Multi-atlas segmentation with augmented features for cardiac MR images. Med. Image Anal. **19**(1), 98–109 (2015)
2. Chen, H., Ni, D., Qin, J., et al.: Standard plane localization in fetal ultrasound via domain transferred deep neural networks. IEEE JBHI **19**(5), 1627–1636 (2015)
3. Çiçek, Ö., Abdulkadir, A., et al.: 3D U-Net: learning dense volumetric segmentation from sparse annotation. arXiv preprint arXiv:1606.06650 (2016)
4. Dou, Q., Yu, L., et al.: 3D deeply supervised network for automated segmentation of volumetric medical images. Med. Image Anal. (2017)
5. Ecabert, O., et al.: Segmentation of the heart and great vessels in CT images using a model-based adaptation framework. Med. Image Anal. **15**(6), 863–876 (2011)
6. Glorot, X., Bengio, Y.: Understanding the difficulty of training deep feedforward neural networks. In: AISTATS, vol. 9, pp. 249–256 (2010)
7. Krizhevsky, A., Sutskever, I., Hinton, G.E.: ImageNet classification with deep convolutional neural networks. In: NIPS, pp. 1097–1105 (2012)
8. Lee, C.Y., Xie, S., Gallagher, P., Zhang, Z., Tu, Z.: Deeply-supervised nets (2015)
9. Milletari, F., Navab, N., et al.: V-Net: fully convolutional neural networks for volumetric medical image segmentation. In: 2016 Fourth International Conference on 3D Vision (3DV), pp. 565–571. IEEE (2016)
10. Peters, J., Ecabert, O., Meyer, C., Schramm, H., Kneser, R., Groth, A., Weese, J.: Automatic whole heart segmentation in static magnetic resonance image volumes. In: Ayache, N., Ourselin, S., Maeder, A. (eds.) MICCAI 2007. LNCS, vol. 4792, pp. 402–410. Springer, Heidelberg (2007). https://doi.org/10.1007/978-3-540-75759-7_49
11. Pizer, S.M., Amburn, E.P., et al.: Adaptive histogram equalization and its variations. Comput. Vis. Graph. Image Process. **39**(3), 355–368 (1987)
12. Ronneberger, O., Fischer, P., Brox, T.: U-Net: convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) MICCAI 2015. LNCS, vol. 9351, pp. 234–241. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-24574-4_28
13. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556 (2014)
14. Tran, D., Bourdev, L., Fergus, R., Torresani, L., Paluri, M.: Learning spatiotemporal features with 3D convolutional networks. In: ICCV, pp. 4489–4497 (2015)
15. Tran, P.V.: A fully convolutional neural network for cardiac segmentation in short-axis MRI. arXiv preprint arXiv:1604.00494 (2016)
16. Xie, S., Tu, Z.: Holistically-nested edge detection. In: ICCV, pp. 1395–1403 (2015)
17. Yu, L., Yang, X., Qin, J., Heng, P.-A.: 3D FractalNet: dense volumetric segmentation for cardiovascular MRI volumes. In: Zuluaga, M.A., Bhatia, K., Kainz, B., Moghari, M.H., Pace, D.F. (eds.) RAMBO/HVSMR -2016. LNCS, vol. 10129, pp. 103–110. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-52280-7_10
18. Zhuang, X., et al.: A registration-based propagation framework for automatic whole heart segmentation of cardiac MRI. IEEE TMI **29**(9), 1612–1625 (2010)
19. Zotti, C., Luo, Z., et al.: Novel deep convolution neural network applied to MRI cardiac segmentation. arXiv preprint arXiv:1705.08943 (2017)

3D Deeply-Supervised U-Net Based Whole Heart Segmentation

Qianqian Tong[1], Munan Ning[1], Weixin Si[2], Xiangyun Liao[3(✉)], and Jing Qin[2]

[1] School of Computer, Wuhan University, Wuhan, China
[2] School of Nursing, The Hong Kong Polytechnic University, Hong Kong, China
[3] Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Beijing, China
xyunliao@gmail.com

Abstract. Accurate whole-heart segmentation from multi-modality medical images (MRI, CT) plays an important role in many clinical applications, such as precision surgical planning and improvement of diagnosis and treatment. This paper presents a deeply-supervised 3D U-Net for fully automatic whole-heart segmentation by jointly using the multi-modal MRI and CT images. First, a 3D U-Net is employed to coarsely detect the whole heart and segment its region of interest, which can alleviate the impact of surrounding tissues. Then, we artificially enlarge the training set by extracting different regions of interest so as to train a deep network. We perform voxel-wise whole-heart segmentation with the end-to-end trained deeply-supervised 3D U-Net. Considering that different modality information of the whole heart has a certain complementary effect, we extract multi-modality features by fusing MRI and CT images to define the overall heart structure, and achieve final results. We evaluate our method on cardiac images from the multi-modality whole heart segmentation (MM-WHS) 2017 challenge.

Keywords: Whole heart segmentation · 3D deeply-supervised U-Net
Multi-modal cardiac images

1 Introduction

The morphological and pathological information acquired from three-dimensional (3D) medical imaging (such as magnetic resonance imaging (MRI), computed tomography (CT), etc.) is essential to improve the diagnosis and treatment of cardiac vascular diseases. Accurate whole-heart 3D segmentation can be beneficial for generating patient-specific 3D heart models thus to obtain these clinical information. However, it is very tedious to annotate large volumes slice-by-slice [1]. More importantly, it may be prone to errors to delineate all the substructures of the whole heart manually. Therefore, automatically preforming the whole-heart 3D segmentation becomes increasingly popular in medical image analysis.

M. Pop et al. (Eds.): STACOM 2017, LNCS 10663, pp. 224–232, 2018.
https://doi.org/10.1007/978-3-319-75541-0_24

Different from many cardiac segmentation works, which focus solely on a single structure such as the left ventricle myocardium [2,3], the goal of the whole-heart segmentation is to extract the volume and shapes of all the substructures of the heart and there are many challenges to perform this task fully automatically [4]: the geometry of the heart is complex because the heart shape varies significantly for different subjects and it may be different for the same subject at different cardiac conditions, some boundaries between anatomical substructures are visually indistinct based on the intensity distributions of the medical images, and the image data may contain heavy motion artifacts, intensity inhomogeneity and noise because of the complex motions and blood flow within the heart.

In recent years, some researches have been involved in the whole-heart segmentation. Zhuang et al. [5] built a mean cardiac MRI atlas from 10 healthy subjects and developed a comprehensive registration algorithm for the atlas-based WHS of MRI. Pace et al. [6] proposed an efficient interactive segmentation method for the whole-heart segmentation (WHS) in congenital heart disease (CHD). For this method, the workload can be significantly reduced (from 4 to 8 h to less than an hour) for the reason that only a small set of annotated slice regions are needed to delineate the whole volume. Nevertheless, manual annotation is still needed in the interactive segmentation. Zhuang et al. [7] employed multi-modality atlases from MRI and CT and achieved an overall Dice score of 89.9. Li er al. [8] proposed to adopt a 3D fully convolutional network (3D FCN) to ensure an efficient voxel-wise labeling for the automatic whole-heart segmentation in cardiac magnetic resonance (CMR) images with CHD.

With the development of deep neural networks [9], deep neural architectures have been successfully applied for many computer vision task such as object classification and detection etc. [10,11] in the last few years. To tackle the challenge that the success of typical convolutional networls was limited due to the size of the available training sets and the size of the considered networks, Ronneberger er al. [12] modified and extended the architecture of fully convolutional network (FCN) [13], yielding a U-Net, which can work with very few training images and produce more precise segmentations. Subsequently, Çiçek et al. [1] extended the U-Net architecture from Ronneberger et al. [12] by replacing all 2D operations with their 3D counterparts for volumetric segmentation, and the 3D U-Net performed well on the highly variable structures of the Xenopus kidney.

In this paper, we present to using deeply-supervised 3D U-Net for fully automatic whole-heart segmentation. Firstly, we coarsely detect the whole heart and segment its region of interest (ROI) using a 3D U-Net (namely detection U-Net) trained by the raw 3D image data, which is used to alleviate the impact of surrounding tissues. Large and rich training dataset is essential for training a competitive deep learning model [2]. However, acquiring amounts of training set is very complicated particularly for medical image analysis. To train a deeply U-net and achieve precise WHS, we enlarge the training set from two aspects. The first strategy is to artificially enlarge the training set by extracting different regions of interest based on the trained detection U-Net. The second strategy

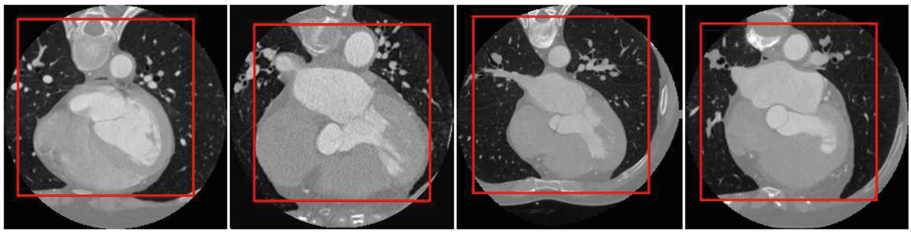


Fig. 1. Results of ROI detection. (Color figure online)

mainly takes full advantage of multi-modality information and we train a 3D U-Net by fusing MRI and CT images to extract multi-modality features. Consequently, the trained 3D U-Net can perform voxel-wise whole-heart segmentation and we evaluate our method on the dataset of the multi-modality whole heart segmentation (MM-WHS) 2017 challenge.

2 Method

We present to use 3D U-Net for the multi-modality whole-heart segmentation. In the training stage, to alleviate the impact of surrounding tissues, we firstly employ a 3D U-Net to coarsely detect the whole heart and segment its region of interest. We then artificially augment the training dataset by extracting different regions of interest. Finally, a refined 3D U-Net is trained using the augmented training dataset. It is worth noting that we fuse MRI and CT images as the training data, which has twofold advantages. It can not only enlarge the training set so as to train a more precise deep learning model, but also take full advantage of multi-modality information so that the volume and shapes of different substructures can be better extracted. In the testing stage, we first coarstly detect the ROI of the whole heart using the trained detection U-Net. Then, the ROI is finely segmented and yield final segmentation results, which is performed voxel-wisely using the end-to-end refined 3D U-Net.

2.1 Data Pre-processing

Region of Interest Detection. In the whole-heart segmentation, we are only interested in the heart organ. However, the images are generally acquired by scanning in clinical practice. Thus the raw image dataset usually include the heart and its surrounding organs. From our former experiments, we found that the background is in a great rates (almost 93%) and the model was trained to output only zero (the label of background). Therefore, it is also very important for machine learning models to alleviate the impact of surrounding tissues and reduce the computational complexity, which helps to improve the segmentation efficiency. In this work, we detect the ROI of the whole heart at first.

To train a detection U-Net, we design the label according to the manual segmentation of the seven whole heart substructures for the training data sets.

For the whole-heart segmentation, there are 8 kinds of voxels (background and seven whole heart substructures) in a 3D object D. For the voxel (i, j, k) in D, supposing that the label of background voxel is $d(i, j, k) = b$ and the label of seven whole heart substructures is $d(i, j, k) = s_m(m = 1, 2, \ldots, 7)$. To train a detection U-Net, we denote the label for ROI detecting as follows for the voxel (i, j, k) in D.

$$l(i,j,k) = \begin{cases} 0 & d(i,j,k) = b \\ 1 & d(i,j,k) = s_m \, (m = 1, 2, \cdots, 7) \end{cases} \tag{1}$$

The results of ROI detection are depicted in Fig. 1. The red square denotes the corresponding predicted ROI. We can see that the predicted ROI contains all the whole heart substructures.

Data Augmentation. We augment the training dataset by segmenting different ROIs in different regions. The raw 3D image is firstly rotated, we translate a cuboid window using different steps in different directions according the detected ROI using the trained detection U-Net and yield different ROIs, which containing the whole heart. Finally, the cropped image is down-sampled, which is beneficial for reducing the complexity of models. We apply the above operation on both data and ground truth labels.

Multi-modality Fusion. In traditional methods, multiple networks should to be trained to segment medical images of different modalities, which has two-fold disadvantages. Firstly, the limited training dataset will affect the performance of a single segmentation network. Moreover, features of different modalities can not be fully utilized. To take full advantage of multi-modality information so that features of different substructures can be better extracted, we obtain multi-modality medical images by fusing MRI and CT images.

Because the size of medical images in different modalities is different, we firstly normalize these images to the same size. Besides, the range of voxels value in medical images of different modalities varies. Therefore, it can not train a good network by using multi-modality medical images. To tackle this issue, we present a fusion strategy. Firstly, we calculate the mean value ν of 3D training data, including MRI anf CT. We then calculate the value of each voxel $\mathbf{I}(i, j, k)$ in 3D image $\mathbf{I}$, and the calculated voxel value $\widehat{\mathbf{I}}(i, j, k)$ is

$$\widehat{\mathbf{I}}(i,j,k) = \begin{cases} \mathbf{I}(i,j,k) - \nu & if \ \mathbf{I} \ is \ MRI \\ \nu - \mathbf{I}(i,j,k) & if \ \mathbf{I} \ is \ CT \end{cases} \tag{2}$$

2.2 Network Architecture

Our base net is a 3D U-Net implanted by Keras Çiçek et al. [1]. U-net is a classic segment net based on traditional FCN [13], which is widely used for medical image segmentation. It is a multi-scale net and has two path. One path is a

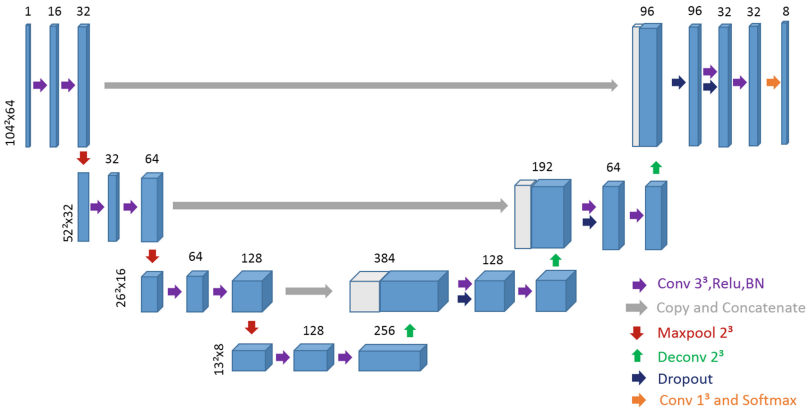


Fig. 2. Architecture of our 3D U-Net for the whole-heart segmentation.

serial of symmetry downsampleing (maxpool) and upsampling (deconv) units, another is to concatenate the "origin" (without deconv operation) images with the upsampling images in each scale. As a result, by considering both "origin" and upsampling images, it can both classific and locate each pixel well, and then yield a good segmentation. The 3D U-Net [12] is accomplished by replacing the 2D layers with 3D layers, and we choose it as our base model to achieve the whole-heart segmentation accurately.

As shown in Fig. 2, we have made many adaptations to train a deep 3D U-Net so as to perform the task of the whole-heart segmentation. Fisrt, we gathered the Conv-Relu-BN layer as our basic conventional unit. To avoid the trained U-Net only producing zero, we adopt BN layers to help the net fit rather than stay at a local optimal solution. What is more, we use padding to make the output size is equal to the input size of conventional units. Second, we add 4 dropout layers and set a 4-scale net (104-52-26-13) to avoid overfitting. We add dropout layers and reduce the depth of our net to adapt to our training dataset. We used the input (128,128,64) at first, which is generated by downsampling origin images by 4 times. To segment and augment the training dataset, our input is changed as (104,104,64). At last, we used Softmax as our activation layer to fit our model and finally generated a (104,104,64,8) output. The 8 maps respectively correspond to 8 kinds of voxels (including background). We also changed our label into one-hot code, then we can calculate the categorical_crossentropy as our loss function, as well as the Adam optimizer because it is self-adaptive and can automatically change the learning rate.

3 Experiments and Results

3.1 Data

We trained and tested our method on the MM-WHS 2017 Challenge dataset, containing 120 multi-modality whole heart images from multiple sites, including 60 cardiac CT/CTA and 60 cardiac MRI in 3D that cover the whole heart

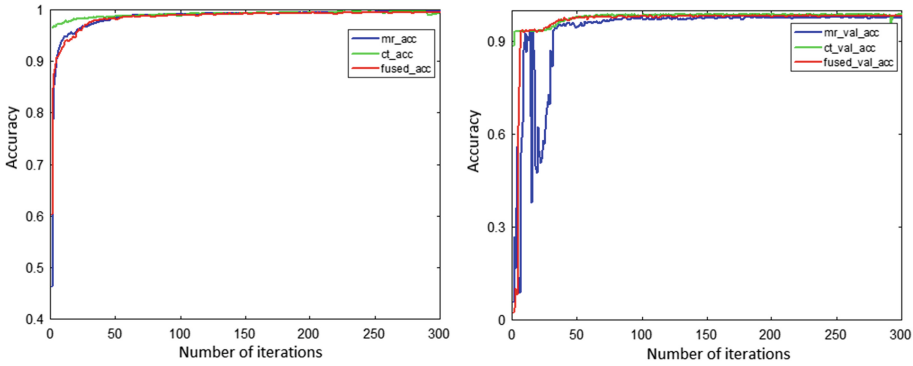


Fig. 3. Results of training and cross validation on training set.

substructures. The datasets is split into training (20 CT and 20 MRI representative) and test (40 CT and 40 MRI) data sets. For the training data sets, manual segmentation of the seven whole heart substructures is provided.

3.2 Performance on Training Set

To evaluate the performance of our multi-modality fusion segmentation method, we adopt the cross validation method to train our multi-modality segmentation network. We utilize 20 CT training data, 20 MRI training data and 40 fused training data to train three 3D U-Net, respectively. The training results are shown in Fig. 3. The left figure denotes the accuracy on three above training data and the right figure depicts the accuracy of cross validation. We can see that although the training performance by MRI training data is good, the results of cross validation is bad. Fortunately, the performance of cross validation for the trained network using our fused training data has been greatly improved, which is validated on both CT and MRI data.

3.3 Performance on Testing Set

Figure 4 depicts the segmentation results of our multi-modality 3D U-Net for the whole-heart segmentation. Seven substructures were of interest in the WHS study, including:

(1) LV: the left ventricle blood cavity (red);
(2) RV: the right ventricle blood cavity (green);
(3) LA: the left atrium blood cavity (blue);
(4) RA: the right atrium blood cavity (cyan);
(5) Myo: the myocardium of the left ventricle (magenta);
(6) AO: the ascending aorta (yellow);
(7) PA: the pulmonary artery (white).

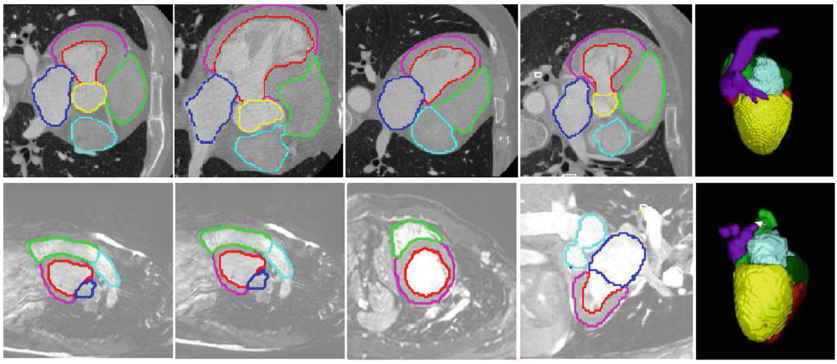


Fig. 4. Segmentation results of our method for the whole-heart segmentation. The top row shows the segmentation results for CT images and the bottom shows the segmentation results for MRI images. (Color figure online)

We can see that our method can segment CT images relatively accurately. The top row shows the segmentation results for CT images and the bottom shows the segmentation results for MRI images. For each row, the first four images denote different slices and the most right denotes the reconstructed 3D model of our result.

Furthermore, our method is quantitatively evaluated using the evaluation metrics Dice score. The average Dice score of our method for seven substructures and the whole heart segmentation (WHS) is shown in Table 1. We can see that our method can achieve good segmentation for CT images and the average Dice score of the whole heart segmentation is 0.849. Besides, except for the right atrium blood cavity, the Dice scores of six other substructures are all bigger than 0.80. The average Dice score of the whole heart segmentation from MRI is 0.674, which may result from the poor quality of MRI.

Table 1. The average Dice score of our method.

Modality	LV	Myo	RV	LA	RA	AO	PA	WHS
CT	0.893	0.837	0.810	0.889	0.812	0.868	0.698	0.849
MRI	0.702	0.623	0.680	0.676	0.654	0.599	0.470	0.674

4 Discussion and Conclusion

In this paper, we present to use 3D U-Net to perform fully automatic whole-heart segmentation. Firstly, we coarsely detect the ROI of the whole heart by a trained detection U-Net. Then, we artificially enlarge the training set by extracting different regions of interest so as to train a deeply supervised network. Subsequently, we train a deeply-supervised 3D U-Net by fusing MRI and CT images. Finally,

we perform voxel-wise whole-heart segmentation with the end-to-end trained 3D U-Net and achieve final results. The experimental results of our method demonstrates that our method can yield precise segmentation results. However, our segmentation performance is affected by the limited display memory. In future work, we will focus on the distributed approach to handle the task of the whole-heart segmentation.

Acknowledgements. This work was supported by grants from Shenzhen Science and Technology Program (No. JCYJ20160429190300857), China Posdoctoral Science Foundation (2017M622831) and SIAT Innovation Program for Excellent Young Researchers (No. 2017059).

References

1. Çiçek, Ö., Abdulkadir, A., Lienkamp, S.S., Brox, T., Ronneberger, O.: 3D U-Net: learning dense volumetric segmentation from sparse annotation. In: Ourselin, S., Joskowicz, L., Sabuncu, M.R., Unal, G., Wells, W. (eds.) MICCAI 2016. LNCS, vol. 9901, pp. 424–432. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-46723-8_49
2. Anh, T., Carneiro, G.: Fully automated non-rigid segmentation with distance regularized level set evolution initialized and constrained by deep-structured inference. In: Proceedings of CVPR, pp. 3118–3125 (2014)
3. Avendi, R., Kheradvar, A., Jafarkhani, H.: A combined deep-learning and deformable-model approach to fully automatic segmentation of the left ventricle in cardiac MRI. Med. Image Anal. **30**, 108–119 (2016)
4. Zhuang, X., Shen, J.: Challenges and methodologies of fully automatic whole heart segmentation: a review. J. Healthc Eng. **4**(3), 371–407 (2013)
5. Zhuang, X., Rhode, S., Razavi, S., Hawkes, J., Ourselin, S.: A registration-based propagation framework for automatic whole heart segmentation of cardiac MRI. IEEE Trans. Med. Imaging **29**(9), 1612–1625 (2010)
6. Pace, D.F., Dalca, A.V., Geva, T., Powell, A.J., Moghari, M.H., Golland, P.: Interactive whole-heart segmentation in congenital heart disease. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) MICCAI 2015. LNCS, vol. 9351, pp. 80–88. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-24574-4_10
7. Zhuang, X., Shen, J.: Multi-scale patch and multi-modality atlases for whole heart segmentation of MRI. Med. Image Anal. **31**, 77–87 (2016)
8. Li, J., Zhang, R., Shi, L., Wang, D.: Automatic whole-heart segmentation in congenital heart disease using deeply-supervised 3D FCN. In: Zuluaga, M.A., Bhatia, K., Kainz, B., Moghari, M.H., Pace, D.F. (eds.) RAMBO/HVSMR -2016. LNCS, vol. 10129, pp. 111–118. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-52280-7_11
9. LeCun, Y., Bengio, Y., Hinton, G.: Deep learning. Nature **521**(7553), 436–444 (2015)
10. Krizhevsky, A., Sutskever, I., Hinton, E.: ImageNet classification with deep convolutional neural networks. In: NIPS, pp. 1097–1105 (2012)
11. Eigen, D., Fergus, R.: Predicting depth, surface normals and semantic labels with a common multi-scale convolutional architecture. In: Proceedings of ICCV, pp. 2650–2658 (2015)

12. Ronneberger, O., Fischer, P., Brox, T.: U-Net: convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) MICCAI 2015. LNCS, vol. 9351, pp. 234–241. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-24574-4_28
13. Long, J., Shelhamer, E., Darrell, T.: Fully convolutional networks for semantic segmentation. In: Proceedings of CVPR, pp. 3431–3440 (2015)

MRI Whole Heart Segmentation Using Discrete Nonlinear Registration and Fast Non-local Fusion

Mattias P. Heinrich[1(✉)] and Julien Oster[2]

[1] Institute of Medical Informatics, University of Lübeck, Lübeck, Germany
heinrich@imi.uni-luebeck.de
[2] IADI, U947, Inserm, CHRU de Nancy, Vandoeuvre les Nancy, France
http://www.mpheinrich.de

Abstract. We present a robust and accurate method for multi-atlas segmentation of whole heart MRI scans. After preprocessing, which includes resampling to isotropic voxel sizes and cropping or padding to same dimensions, all training scans are registered linearly and nonlinearly to an unseen set of test scans. We employ the efficient discrete registration framework called *deeds* that captures large shape variations across scans, performed best in a recent registration comparison on abdominal scans and requires less than 2 min of computation time per scan. Subsequently, we perform multi-atlas label fusion using a non-local means approach with a normalised SSD metric and a fast implementation using boxfilters. Subsequently, a multi-label random walk is performed on the obtained probability maps for an edge-preserving smoothing. Without performing any domain-specific parameter tuning, we obtained a Dice accuracy of 86.0% (averaged across 7 labels) and 87.0% for the whole heart on the MRI test dataset, which is the first rank of the MICCAI 2017 challenge. The segmentations are also visually very smooth using this fully automatic method.

1 Introduction and Related Work

The automatic segmentation of patient geometry from MRI scans has numerous applications in clinical practice and research of cardiac physiology. Cardiac MRI enables an accurate estimation of the cardiac function [1], by acquiring CINE images throughout the cardiac cycle. Radiologists are then required to manually delineate the left ventricle blood cavity, for at least the end-diastole and end-systole phase and all (≈10) slices covering the heart. The left ventricle blood cavity volume can then be computed and the left ventricle ejection fraction can be estimated. This process is thus very cumbersome for the radiologists and automatic delineation and segmentation of the MRI scans is therefore appealing for clinicians.

We are in particular interested in deriving a robust anatomical surface model from tomographic scans to enable electro-physiological modelling and perspectively electrocardiographic imaging (ECGI) [2]. ECGI consists in fusing images

M. Pop et al. (Eds.): STACOM 2017, LNCS 10663, pp. 233–241, 2018.
https://doi.org/10.1007/978-3-319-75541-0_25

(morphological models) with electrophysiological signals (hundreds of ECG signals acquired across the patient torso) in order to estimate a map of the myocardial electrical biopotentials. Applications of ECGI range from ablation procedure planning for both atrial and ventricular arrhythmias, and localisation of premature ventricular contractions. Cardiac MRI images are rarely used for ECGI, but may offer advantages over CT scans, namely dynamic images for myocardial motion estimation [3], but also offer an in-depth tissue characterization, such as ischemic scars [4] and fiber orientation with Diffusion Tensor Images [5].

The goal of the *"MM-WHS 2017: Multi-Modality Whole Heart Segmentation"* challenge held in conjunction with MICCAI 2017 is the delineation of the following seven anatomical structures in medical scans: the myocardium of the left ventricle, the left atrium blood cavity, the left ventricle blood cavity, the right atrium blood cavity, the right ventricle blood cavity, the ascending aorta, and the pulmonary artery. The challenge offers two datasets of each 20 CT and MRI scans for training with corresponding ground truth labels and 40 scans each for testing and benchmarking of different methods. We are mainly interested in MRI scans and do not consider the CT data at all.

Numerous automatic algorithms have been proposed for heart segmentation in the past (see e.g. [6] for an overview of a recent left atrium challenge), which can be roughly divided into three classes: registration-based, model-based approaches (cf. [7]) or deep learning with convolutional neural networks (CNN). The strengths of statistical models and deep learning are their low processing time when applying the model to unseen data and a computational demand in test that is generally independent of the number of atlases. Whereas registration-based approaches require fewer labelled atlases and often achieve the highest accuracies [8]. Since, individual registrations sometimes fail, it is important to find a robust label fusion strategy, e.g. by employing patch-similarity [9]. This may, however, result in more irregular surfaces for difficult to segment subjects (see Fig. 5 of [6] for the registration approach UCL-1C). Statistical shape models or locally adaptive models can more robustly capture physiologically plausible shapes, but sometimes are inaccurate when certain rare anatomies are under-represented in the training data [7]. CNN models have only very recently been successfully addressed to cardiac image segmentation (e.g. in [10]) and excel in directly deriving clinically relevant information from cine MR (as evident from https://www.kaggle.com/c/second-annual-data-science-bowl). But, deep models require large training datasets, additional data augmentation strategies and substantial computing resources. Furthermore, they cannot directly encode shape or deformation priors as well as registration- or statistical-model based approaches. Newer work, therefore, focuses on incorporating shape priors [11] to improve accuracy for scans with low quality.

In this work, we aim to address the remaining challenges of MRI cardiac segmentation, which include: varying image quality, (relatively) low numbers of training scans and inconsistent image contrast. Our approach relies on a very effective discrete deformable registration approach that is combined with a specifically adapted non-local fusion strategy and probability map

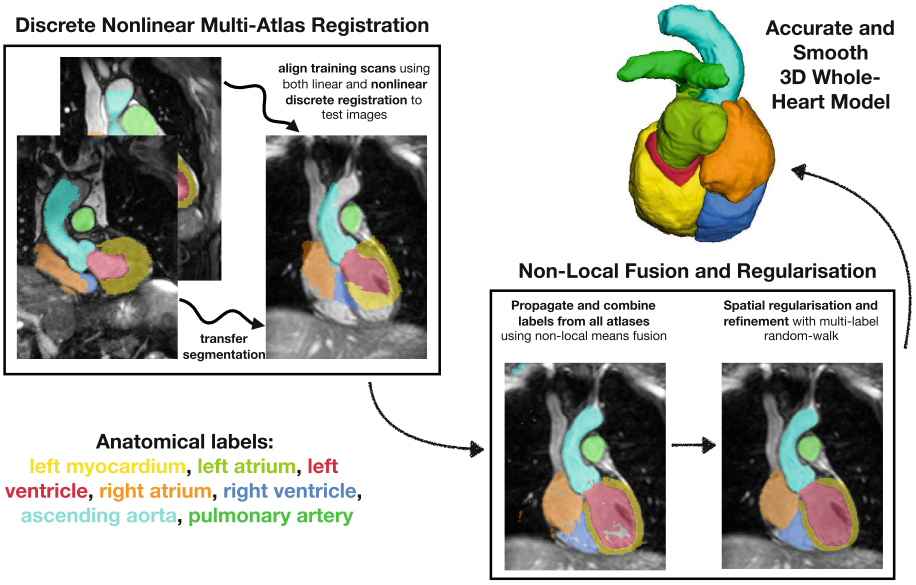


Fig. 1. Flow-chart of our proposed algorithm for accurate segmentation of the whole heart in MRI. First, all 20 training scans are linearly and nonlinearly aligned to the unseen test scan. Second, the transformed segmentations from all atlases are weighted (over a non-local search region) using their patch similarity. Finally, a spatial edge-preserving regularisation is applied to compensate for minor inconsistencies.

regularisation. The discrete registration *deeds* [12] overcomes the limitations of long-processing times for registration-based and performs substantially better than common methods using continuous optimisation e.g. ANTS, IRTK and NiftyReg. A detailed evaluation of these state-of-the-art approaches for non-brain scans was presented in [13]. A graphical overview of our approach is given in Fig. 1. Each part of the approach is then described in detail in Sect. 2 (discrete registration), Sect. 3 (non-local fusion) and Sect. 4 (multi-label random walk regularisation). A preliminary evaluation of the influence of those consecutive parts on the segmentation accuracy and robustness of surface estimation is presented alongside the description of the methods and discussed in Sect. 5.

2 Discrete Registration

We resampled the scans to an isotropic resolution of $\approx 1.0\,\text{mm}^3$ and padded or cropped them to have same dimensions of $288 \times 288 \times 180$ and enable pairwise registration of all images. We employ the publicly available discrete deformable registration tool *deeds* [12] based on the implementation found at github.com/mattiaspaul/deedsBCV. The choice is motivated by its very high computational efficiency and the encouraging accuracies for inter-subject abdominal registration in [13]. An affine pre-registration is performed using discrete

block-matching on four scale levels as defined per default. Subsequently the non-linear registration is called using with the following (default) parameters that where found to be very suitable for abdominal registration in [13]: number of displacement steps $l_{\max} = [8, 7, 6, 5, 4]$, quantisation steps $q = [5, 4, 3, 2, 1]$, and B-spline grid spacings of $[8, 7, 6, 5, 4]$ voxels. Using a dense displacement sampling covering a large range of potential displacements together with a combinatorial optimisation based on dynamic programming, enables this approach to capture strong shape deformations across patients. The self-similarity context (SSC) [14] was used as a similarity metric with a weighting of $\alpha = 1.6$. Initially developed for multi-modal registration, SSC enables contrast invariance, which is particularly beneficial for MRI scans and focuses the alignment on image edges. The computations are performed in less than two minutes per scan-pair (and many registrations can be run in parallel on a multi-core system). We store both the deformed MRI scans and the transformed segmentations for the following label fusion.

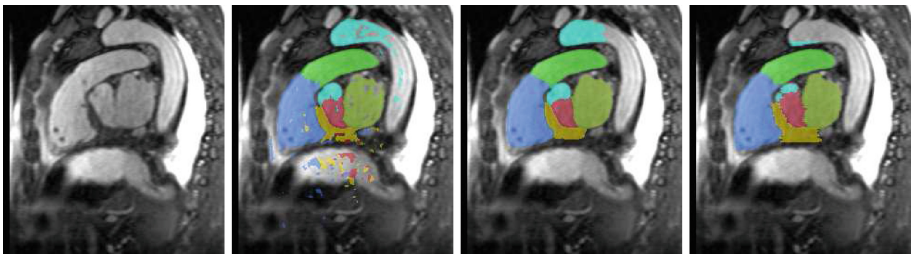
3 Non-local Label Fusion

Label fusion of multiple atlases is a common approach in multi-atlas segmentation [9,15,16]. The rationale behind this is to exclude poorly registered atlases and refine the segmentation boundary. Due to the regularisation constraint within deformable registration, the deformation of the anatomies is restricted and certain parts are therefore often not perfectly aligned. One could directly incorporate the displacement uncertainty into a multi-atlas fusion [17], but usually a sequential step that performs an (unregularised) non-local search is simpler to implement.

We follow the approach of [18] and calculate non-local weights, independently for each patch (and atlas scan), based on a similarity metric, the normalised sum of squared differences (NSSD), between the central patch of the target image and a patch within the non-local search region $\boldsymbol{n} \in \mathcal{N}$ of the atlas scan. Let $\boldsymbol{x}$ be a 3D location in an image I and $X(\boldsymbol{x})$ an intensity patch in I with a patch radius r, which defines $\Omega_{\boldsymbol{x}}$ with a size of $R = (2r+1)^3$. The corresponding patch at $\boldsymbol{n}$ within the search region of the atlas scan is defined as $Y(\boldsymbol{x}+\boldsymbol{n})$. To reduce the impact of local contrast variations (cf. [15]), we subtract the mean of each patch and divide by their standard deviation to obtain a normalised patch $X'(\boldsymbol{x})$, $Y'(\boldsymbol{x})$, using:

$$X'(\boldsymbol{x}) = \frac{X(\boldsymbol{x}) - \mu_X}{\sigma_X} \text{ with } \mu_X = \frac{1}{R}\sum_{\boldsymbol{y}\in\Omega_{\boldsymbol{x}}} X(\boldsymbol{y}) \text{ and } \sigma_X = \sqrt{\frac{1}{R}\sum_{\boldsymbol{y}\in\Omega_{\boldsymbol{x}}} (X(\boldsymbol{y}) - \mu_X)^2} \quad (1)$$

In practice, we use a fast implementation to calculate local means and standard deviations using the boxfilter approach of [19]. It can be shown that the computation time is then independent of the patch size R and thus substantially lower than in [15,18]. To calculate the NSSD between all patches in two scans, only one additional pointwise multiplication between the intensity images followed by



a) MRI scan b)non-local fusion c) random walk d) ground truth

Fig. 2. Visual comparison of segmentation results. ■ myocardium of the left ventricle, ■ left atrium, ■ left ventricle, ■ right atrium, ■ right ventricle, ■ ascending aorta, and ■ pulmonary artery

another boxfilter is required. The value at each location can then be computed using binomial expansion as (the sum being implemented as filter):

$$\mathrm{NSSD}(\boldsymbol{x}, \boldsymbol{n}) = 2R - \frac{2(\sum_{\boldsymbol{y}\in\Omega_x}(X(\boldsymbol{y}) \cdot Y(\boldsymbol{y}+\boldsymbol{n})) - \mu_X \mu_Y R)}{\sigma_X \sigma_Y} \tag{2}$$

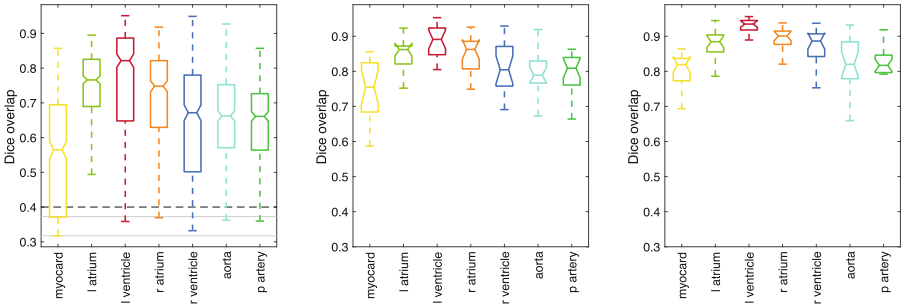
We shift the atlas image by each of the $\boldsymbol{n} \in \mathcal{N}$ displacements within the non-local search region and achieve around 25 boxfilter operations (for scans with 15 million voxels) per second using advanced vector instructions. Finally the weights are computed by the exponential term $w(\boldsymbol{x}, \mathbf{n}) = \exp\left(-\mathrm{NSSD}(\boldsymbol{x}, \boldsymbol{n})/(2\sigma^2_{nlm})\right)$, with $\sigma_{nlm} = 0.25$. The size of the local patch and search region are empirically chosen to be both $7 \times 7 \times 7$ voxels.

We anticipate that considering fusion weights jointly as proposed in [15] would lead to better results. However, this algorithm requires much longer computation times and, due to limited processing resources, we would have been unable to finish these experiments in time. Alternatively, the SIMPLE approach for atlas pre-selection as e.g. employed in [20,21] could further reduce the computational demand and negative influence of poorly aligned atlases.

4 Multi-label Random Walk Regularisation

The non-local fusion helps to correct minor registration inaccuracies, but its normalised sum of squared difference metric is also more sensitive to image edges. Therefore, the output may yield some isolated misclassifications or holes in solid objects, hence the segmentation is not necessarily spatially consistent. To counter these effects it is therefore beneficial to regularise the probability maps $P^c(\boldsymbol{x})$ spatially that were obtained using the label fusion in the previous section. We aim to obtain smooth maps $P(\boldsymbol{x})^c_{reg}$ for every label $c \in C$, while adhering to image edges and employ the multi-label random walk [22] to minimise the objective $E(P(\boldsymbol{x})^c_{reg})$:

$$\sum_{\boldsymbol{x}} \frac{1}{2}(P(\boldsymbol{x})^c - P(\boldsymbol{x})^c_{reg})^2 + \sum_{\boldsymbol{x}} \frac{\lambda}{2}||\nabla P(\boldsymbol{x})^c_{reg}||^2 \tag{3}$$



a) deeds registration b) non-local fusion c) spatial regularisation

Fig. 3. Numerical results for Dice overlap, of the three different steps in our multi-atlas segmentation framework. Note that the first boxplot (a) consists of the results for single-atlas propagation (380 nonlinear registrations) and has a larger range due to sporadic outliers. It is evident that both non-local multi-atlas fusion and random walk regularisation can substantially improve the registration outcome.

The implementation follows [23] and is publicly available. To preserve edges, the gradient of the probability map is weighted by $w_j = \exp(-(I(\boldsymbol{x}_i) - I(\boldsymbol{x}_j))^2/(2\sigma_w^2))$. We empirically chose $\sigma_w = 50$ and a regularisation weight of $\lambda = 50$ as suggested in [23] for multi-organ segmentation. This weight depends on the differences of image intensities I of $\boldsymbol{x}_i$ and its neighbouring voxels $\boldsymbol{x}_j \in \mathcal{N}_i$. Alternatively, optimisation techniques such as fully-connected conditional random fields (CRF) have been proposed in the literature, which are particularly powerful when enforcing long-range spatial clues [24]. However, for the segmentation task at hand, we found the random walk algorithms provided visually very good results (see Fig. 2) with computation times as low as a few seconds per scan.

Numerical results for segmentation accuracy: We use the Dice overlap $D = 2|A \cap M|/(|A| + |M|)$ to measure the accuracy of our automatic segmentations A compared to the provided manual label images M. For 20 training scans of the challenge, we performed a leave-one-out validation. The distribution of

Table 1. Numerical results of the label fusion (NLM) and spatial regularisation in Dice overlap in % and mean surface distance (MSD) in mm, obtained using leave-one-out cross validation on the MRI training dataset and the submitted results for the test set of the challenge. The spatial random walk (RW) regularisation improves all labels. On the test dataset only the complete method including RW is employed.

Label	myocard	l atrium	l ventricle	r atrium	r ventricle	aorta	p artery	Average
NLM train	74.9%	84.3%	86.7%	84.8%	79.5%	78.8%	79.0%	**81.1%**
RW train	79.6%	87.8%	92.0%	89.3%	85.8%	81.7%	80.3%	**85.2%**
Dice test	78.1%	88.6%	91.8%	87.3%	87.1%	87.8%	80.4%	**86.0%**
MSD test	1.46	1.38	1.32	1.80	1.74	1.23	1.64	**1.51**

scores for the discrete registration (single-atlas propagation), the non-local label fusion and the spatial regularisation are presented in Fig. 3. The average over all 20 training cases is additionally reported in Table 1.

5 Discussion and Conclusion

The obtained segmentations are visually very convincing, in particular for the left ventricle and atrium. By using a relatively strong regularisation, we can obtain smooth surfaces that are important for mesh generation and motion or electrophysiological modelling. Depending on the available number of processing cores, the complete multi-atlas fusion for one test scan could be accomplished in less than 5 min. The Dice overlap for all 7 labels is given in Table 1. Averaged across all structures, we achieve a Dice of 85.2% (training) and 86.0% (test dataset), which is the first rank for the *"MM-WHS 2017: Multi-Modality Whole Heart Segmentation"* challenge. It is, however, slightly below the current state-of-the-art that reached 88.3% on the same 7 structures [9]. We obtained a relatively low mean surface result across all structures of 1.51 mm.

We argue that part of the remaining inaccuracies are due to ambiguities for the definition of the ascending aorta and pulmonary artery (see Fig. 2), which are hard to consistently segment manually (leading to inter-observer errors of up to 24% in [9]). Furthermore, the papillary muscles are often visually hard to distinguish from the myocardium, which leads to further uncertainties.

Prospectively, a more elaborated parameter tuning and adaptation to this specific domain of registration, label fusion and regularisation could lead to further improvements. In addition, the automatic cropping of the MRI scans sometimes led to a field-of-view that did not completely cover the whole heart and this should be addressed in the future. Two areas of particular interest would be a better delineation of the myocardium and the exact separation between left and right atria e.g. by employing a shape prior during label fusion [25] or considering contextual features [26].

Acknowledgements. We would like to thank the organisers of the MM-WHS 2017 for providing this rich new dataset to the public, which enables the evaluation of new algorithms for the problem of detailed 3D heart segmentation.

References

1. Grothues, F., Smith, G.C., Moon, J.C., Bellenger, N.G., Collins, P., Klein, H.U., Pennell, D.J.: Comparison of interstudy reproducibility of cardiovascular magnetic resonance with 2D echocardiography in normal subjects and in patients with heart failure or left ventricular hypertrophy. Am. J. Cardiol. **90**(1), 29–34 (2002)
2. Ramanathan, C., Ghanem, R.N., Jia, P., Ryu, K., Rudy, Y.: Noninvasive electrocardiographic imaging for cardiac electrophysiology and arrhythmia. Nat. Med. **10**(4), 422–428 (2004)

3. Vuissoz, P.A., Odille, F., Fernandez, B., Lohezic, M., Benhadid, A., Mandry, D., Felblinger, J.: Free-breathing imaging of the heart using 2D cine-GRICS with assessment of ventricular volumes and function. J. Magn. Reson Imaging **35**(2), 340–351 (2012)
4. Nazarian, S., Bluemke, D.A., Lardo, A.C., Zviman, M.M., Watkins, S.P., Dickfeld, T.L., Meininger, G.R., Roguin, A., Calkins, H., Tomaselli, G.F., et al.: Magnetic resonance assessment of the substrate for inducible ventricular tachycardia in non-ischemic cardiomyopathy. Circulation **112**(18), 2821–2825 (2005)
5. Nielles-Vallespin, S., Mekkaoui, C., Gatehouse, P., Reese, T.G., Keegan, J., Ferreira, P.F., Collins, S., Speier, P., Feiweier, T., Silva, R., et al.: In vivo diffusion tensor MRI of the human heart: reproducibility of breath-hold and navigator-based approaches. Magn. Reson. Med. **70**(2), 454–465 (2013)
6. Tobon-Gomez, C., Geers, A.J., Peters, J., Weese, J., Pinto, K., Karim, R., Ammar, M., Daoudi, A., Margeta, J., Sandoval, Z., et al.: Benchmark for algorithms segmenting the left atrium from 3D CT and MRI datasets. IEEE Trans. Med. Imag. **34**(7), 1460–1473 (2015)
7. Kutra, D., Saalbach, A., Lehmann, H., Groth, A., Dries, S.P.M., Krueger, M.W., Dössel, O., Weese, J.: Automatic multi-model-based segmentation of the left atrium in cardiac MRI scans. In: Ayache, N., Delingette, H., Golland, P., Mori, K. (eds.) MICCAI 2012. LNCS, vol. 7511, pp. 1–8. Springer, Heidelberg (2012). https://doi.org/10.1007/978-3-642-33418-4_1
8. Zhuang, X., Rhode, K.S., Razavi, R.S., Hawkes, D.J., Ourselin, S.: A registration-based propagation framework for automatic whole heart segmentation of cardiac MRI. IEEE Trans. Med. Imag. **29**(9), 1612–1625 (2010)
9. Zhuang, X., Shen, J.: Multi-scale patch and multi-modality atlases for whole heart segmentation of MRI. Med. Imag. Anal. **31**, 77–87 (2016)
10. Wolterink, J.M., Leiner, T., Viergever, M.A., Išgum, I.: Dilated convolutional neural networks for cardiovascular MR segmentation in congenital heart disease. In: Zuluaga, M.A., Bhatia, K., Kainz, B., Moghari, M.H., Pace, D.F. (eds.) RAMBO/HVSMR -2016. LNCS, vol. 10129, pp. 95–102. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-52280-7_9
11. Oktay, O., Ferrante, E., Kamnitsas, K., Heinrich, M., Bai, W., Caballero, J., Guerrero, R., Cook, S., de Marvao, A., O'Regan, D., et al.: Anatomically constrained neural networks (ACNN): Application to cardiac image enhancement and segmentation. arXiv preprint arXiv:1705.08302 (2017)
12. Heinrich, M., Jenkinson, M., Brady, J., Schnabel, J.: MRF-based deformable registration and ventilation estimation of lung CT. IEEE Trans. Med. Imag. **32**(7), 1239–1248 (2013)
13. Xu, Z., Lee, C., Heinrich, M., Modat, M., Rueckert, D., Ourselin, S., Abramson, R., Landman, B.: Evaluation of six registration methods for the human abdomen on clinically acquired CT. IEEE Trans. Biomed. Eng. 1–10 (2016)
14. Heinrich, M.P., Jenkinson, M., Papież, B.W., Brady, S.M., Schnabel, J.A.: Towards realtime multimodal fusion for image-guided interventions using self-similarities. In: Mori, K., Sakuma, I., Sato, Y., Barillot, C., Navab, N. (eds.) MICCAI 2013. LNCS, vol. 8149, pp. 187–194. Springer, Heidelberg (2013). https://doi.org/10.1007/978-3-642-40811-3_24
15. Wang, H., Suh, J.W., Das, S.R., Pluta, J.B., Craige, C., Yushkevich, P.A.: Multi-atlas segmentation with joint label fusion. IEEE Trans. Patt. Anal. Mach. Intell. **35**(3), 611–623 (2013)
16. Asman, A.J., Landman, B.A.: Non-local statistical label fusion for multi-atlas segmentation. Med. Imag. Anal. **17**(2), 194–208 (2013)

17. Heinrich, M.P., Simpson, I., Papież, B., Brady, J., Schnabel, J.: Deformable image registration by combining uncertainty estimates from supervoxel belief propagation. Med. Imag. Anal. **27**, 57–71 (2016)
18. Coupé, P., Manjón, J.V., Fonov, V., Pruessner, J., Robles, M., Collins, D.L.: Patch-based segmentation using expert priors: application to hippocampus and ventricle segmentation. NeuroImage **54**(2), 940–954 (2011)
19. Heinrich, M.P., Papież, B.W., Schnabel, J.A., Handels, H.: Non-parametric discrete registration with convex optimisation. In: Ourselin, S., Modat, M. (eds.) WBIR 2014. LNCS, vol. 8545, pp. 51–61. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-08554-8_6
20. Langerak, T., Van Der Heide, U., Kotte, A., Viergever, M., Van Vulpen, M., Pluim, J.: Label fusion in atlas-based segmentation using a selective and iterative method for performance level estimation (SIMPLE). IEEE Trans. Med. Imag. **29**(12), 2000–2008 (2010)
21. Xu, Z., Asman, A.J., Shanahan, P.L., Abramson, R.G., Landman, B.A.: SIMPLE is a good idea (and better with context learning). In: Golland, P., Hata, N., Barillot, C., Hornegger, J., Howe, R. (eds.) MICCAI 2014. LNCS, vol. 8673, pp. 364–371. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-10404-1_46
22. Grady, L.: Multilabel random walker image segmentation using prior models. In: CVPR, pp. 763–770 (2005)
23. Heinrich, M.P., Blendowski, M.: Multi-organ segmentation using vantage point forests and binary context features. In: Ourselin, S., Joskowicz, L., Sabuncu, M.R., Unal, G., Wells, W. (eds.) MICCAI 2016. LNCS, vol. 9901, pp. 598–606. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-46723-8_69
24. Krähenbühl, P., Koltun, V.: Efficient inference in fully connected CRFs with Gaussian edge potentials. In: Proceedings of NIPS, pp. 2–9 (2011)
25. Oguz, I., Kashyap, S., Wang, H., Yushkevich, P., Sonka, M.: Globally optimal label fusion with shape priors. In: Ourselin, S., Joskowicz, L., Sabuncu, M.R., Unal, G., Wells, W. (eds.) MICCAI 2016. LNCS, vol. 9901, pp. 538–546. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-46723-8_62
26. Bai, W., Shi, W., Ledig, C., Rueckert, D.: Multi-atlas segmentation with augmented features for cardiac MR images. Med. Imag. Anal. **19**(1), 98–109 (2015)

Automatic Whole Heart Segmentation Using Deep Learning and Shape Context

Chunliang Wang(✉) and Örjan Smedby

School for Technology and Health (STH), KTH Royal Institute of Technology, Hälsovägen 11C, 14152 Huddinge, Stockholm, Sweden
{chunliang.wang,orjan.smedby}@sth.kth.se
http://www.kth.se/sth

Abstract. To assist 3D cardiac image analysis, we propose an automatic whole heart segmentation using a deep learning framework combined with shape context information that is encoded in volumetric shape models. The proposed processing pipeline consists of three major steps: scout segmentation with orthogonal 2D U-nets, shape context estimation and refining segmentation with U-net and shape context. The proposed method was evaluated using the MMWHS challenge data. Two sets of networks were trained separately for contrast-enhanced CT and MRI. On the 20 training datasets, using 5-fold cross-validation, the average Dice coefficients for the left ventricle, the right ventricle, the left atrium, the right atrium and the myocardium of the left ventricle were 0.895, 0.795, 0.847, 0.821, 0.807 for MRI and 0.935, 0.825, 0.908, 0.881, 0.879 for CT, respectively. Further improvement may be possible given more training data or advanced data augmentation strategy.

Keywords: Deep learning · Fully convolutional network
Heart segmentation · Shape context · Statistic shape model

1 Introduction

Cardiovascular disease (CVD) is currently the leading cause of death worldwide. Approximately one in five deaths is currently related to cardiac disease in Europe and the US. Nearly 500,000 deaths caused by CVD are reported every year in the US, and over 600,000 in Europe. Approximately half of men and one third of women over 40 years old will develop CVD [1]. Both computed tomography (CT) and magnetic resonance imaging (MRI) are important diagnostic tools for CVD, allowing medical doctors to directly access morphological and functional changes of the heart. In addition to their clinical use, they are also widely used in clinical trials to study the remodeling of the heart due to various cardiac diseases [2]. Several large cardiac cohort studies include CT and/or MRI imaging in their data collection protocols. One recent example is the Swedish cardiopulmonary bioimage study (SCAPIS), where CT, MRI and ultrasound images as well as blood samples will be collected from 30,000 subjects [3].

M. Pop et al. (Eds.): STACOM 2017, LNCS 10663, pp. 242–249, 2018.
https://doi.org/10.1007/978-3-319-75541-0_26

Both in clinical diagnostic procedures and in the data analysis process in clinical trials, segmentation of the heart is often one of the primary steps required to generate any useful quantitative measurements. Heart segmentation in 3D is known to be time-consuming if performed manually. Considerable efforts have been made to automate the procedure and to reduce the involvement of the radiologist during the segmentation. Promising results have been reported in the literature, in particular with the statistical shape-model based methods [4] and atlas-based methods [5]. On the other hand, deep learning based methods are gaining more and more attention due to their superior performance in several image segmentation challenges [6,7].

In this study, we propose a hybrid method that attempts to integrate the statistical shape model into the deep learning process. This is done by feeding the estimated volumetric shape models of several cardiac structures as context layers to a fully convolutional network (FCN). In our preliminary experiments, the shape context layers seem to increase the segmentation accuracy of most structures compared to the plain FCN, when validated on a publicly available database of 20 contrast-enhanced CT scans and 20 contrast-enhanced MR scans with manually created ground truth.

2 Methods

As illustrated in Fig. 1, the proposed processing pipeline consists of three major steps: scout segmentation with orthogonal 2D U-nets, shape context estimation and refining segmentation with U-net and shape context. The individual steps are explained in the following sections.

2.1 2.5D Segmentation Using Orthogonal U-Nets

Convolutional neural networks (CNN) have been successfully used in many segmentation tasks [6]. In this study, we adopted a somewhat more sophisticated FCN architecture, called U-net, proposed by Ronneberger et al. [7]. In FCNs, the fully connected layers of classical CNNs are replaced by convolutional layers [6], which allows FCNs to be applied to images of any size and output label maps proportional to the input image. In combination with "skips" and up-sampling layers [6], the FCNs are often designed to produce the same size output image as the input image, thus eliminating the need for the time-consuming sliding window process used by classical CNN-based methods. To perform segmentation in 3D volumes, we used three U-nets that were trained independently to segment multiple structures in 2D slices acquired in three orthogonal projects, i.e., in axial, coronal and sagittal views. The final probability map of each structure is generated by averaging the outputs of these three U-nets.

2.2 Shape Context Generation

In this study, we use the volumetric statistical shape models proposed by proposed by Leventon et al. [8]. As described in [8], the statistical model is created

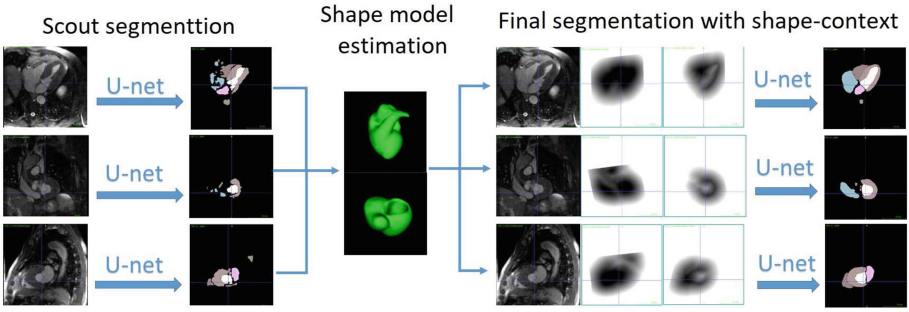


Fig. 1. An overview of the segmentation pipeline

by taking the mean of the signed distance functions of each segmented region and n prominent variations extracted via Principal Component Analysis (PCA). Then the model M that matches the current case is estimated by solving a level set function

$$\frac{\partial \phi}{\partial t} = \alpha F(x) + \beta M(T(x)) + \gamma \kappa(x) \left| \nabla \phi \right| \tag{1}$$

where F is the image force, which is simply a threshold function on the probability maps from U-nets in this case, M is the statistical model as a weighted sum of the mean shape and PCA components, T is the global transformation and κ is the mean curvature. The transformation T and the weighting factors of PCA components are updated iteratively by minimizing the squared distance between the model and the level set function, which is also a signed distance map. The weighting factors, α, β and γ are determined empirically. To speedup the process, a fast level set method using coherent propagation is used [9].

Since the heart is a complex structure with four chambers and several vessels attached to it, we use a hierarchical approach to model the whole heart, similar to the method reported in [10]. At the higher level, we create the overall shape model of the surface of the whole heart (Fig. 2A), and at a lower level the detailed structures are modeled separately with the relative position to the parent shape model preserved (e.g. Fig. 2C). To save computation time and memory consumption in this preliminary study, we merged the right ventricle with the myocardium of the left ventricle into a single structure (Fig. 2B) and ignored all other second level structures. For simplicity, we refer to this fused structure as "ventricle surface".

2.3 Shape-Context Guided U-Net

After estimating the shape models that fit to the probability map of the scout segmentation, the distance maps of the heart surface and the ventricle surface, together with the input images, are fed into another three U-nets that will perform the segmentation again in three orthogonal views. The architecture of these

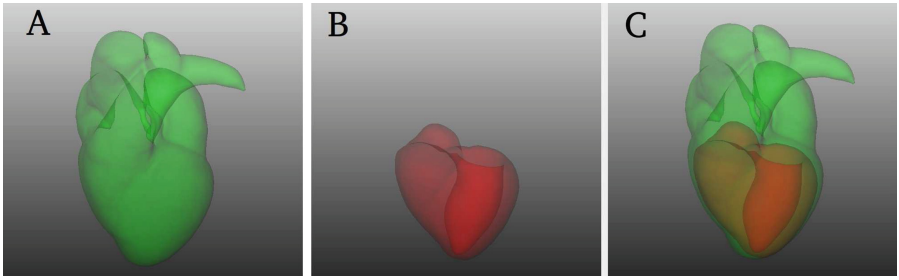


Fig. 2. Shape models used in this study. A, heart surface. B, ventricle surface. C, the relative position of two level models. (Volumetric shape models are used in this study; the illustrations above are created by taking the iso-surface of level 0)

U-nets is identical to the ones used in the scouting step, but retrained from scratch. The final probability map of each structure is again generated by taking the mean value of the outputs from these three U-nets.

2.4 Implementation Details

Our U-net implementation was based on the Keras framework with Theano backend (http://keras.io). The U-net architecture is identical to the one proposed in the original paper, with a little modification to the size of the input image. In our implementation, the segmentation is performed at 1 mm isotropic resolution, which means the CT images are down-sampled to about half the original resolution before being processed. This down-sampling step is introduced only to reduce the GPU memory consumption and reduce training time. The same resolution is used for both the scout segmentation and the refinement segmentation. For the MR scans, an additional landmark detection step using random forest [11] is used to first crop the image into a smaller region of interest ($256 \times 256 \times 256$) to reduce the size of the input image to the U-nets.

Different data normalization methods are used for CT and MRI images and the shape context images. For the CT image and the shape context channels, the voxel intensity is simply divided the group standard deviation (SD) without changing the reference point of 0. For MRI, the intensity of all MRI scans are normalized first individually and then together. During the individual normalization, the lower 5% cutting point of each subject's histogram is mapped to 0 and the upper 5% cutting point is mapped to 1.0. In the group normalization, all subjects are normalized together by subtracting the group mean and divided by group SD, i.e., the normalized images will have 0 mean and SD of 1.

The categorical cross-entropy is used as the loss function for multi-structure segmentation. Stochastic gradient descent (SGD) is used as the optimizer in all training process, and the number of epochs is fixed at 150 for all U-nets. The shape models were created using 10 randomly selected CT cases and used for both the CT and the MRI experiments.

3 Results

The proposed method was evaluated using the training data of the Multi-Modality Whole Heart Segmentation (MMWHS) challenge, which consists 20 contrast-enhanced CT scans and 20 contrast-enhanced T1-weighted MRI scans. In all cases, seven structures, namely the left ventricle blood cavity (LV), the right ventricle blood cavity (RV), the left atrium blood cavity (LA), the right atrium blood cavity (RA), the myocardium of the left ventricle (Myo), the ascending aorta (AA) and the pulmonary artery (PA) were manually delineated. The U-nets were trained separately for the CT images and MRI images. The evaluation was done using five-fold cross-validation, in each fold 16 cases were used for training and 4 cases were used for testing. Shape models, however, were not recreated in every fold.

The average accuracy (Dice coefficient) of the scout segmentation and the refined segmentation is compared in Tables 1 and 2 for MRI images and CT images, respectively. In most cases, the combined approach yielded somewhat higher accuracy than U-nets alone. Table 3 shows the segmentation accuracy of the scout segmentation on the testing datasets of the MMWHS challenge. The accuracy of the refined segmentation with shape context was not submitted by the submission deadline due to an implementation error. Because re-submission is not allowed after the deadline, the scores of the entire pipeline on the testing datasets are not available by the time this paper being accepted. Figure 3 shows the scout segmentation and refined segmentation results of three example cases. The overall processing time for both CT and MRI was about 5–7 min on a personal computer with an Nvidia GTX 1080 graphic card.

Table 1. Comparing the accuracy (Dice coefficient) of the scout segmentation with U-net and the final segmentation with U-net and shape context in MR images

Structures	U-net	U-net & Shape context
Myocardium	0.785 ± 0.091	0.807 ± 0.059
Left atrium	0.877 ± 0.026	0.847 ± 0.061
Left ventricle	0.873 ± 0.082	0.895 ± 0.057
Right atrium	0.749 ± 0.225	0.821 ± 0.087
Right ventricle	0.688 ± 0.205	0.795 ± 0.102
Ascending aorta	0.708 ± 0.257	0.679 ± 0.180
Pulmonary artery	0.622 ± 0.232	0.743 ± 0.204

4 Discussion and Conclusion

Overall, adding the shape context as additional input to the U-nets seems to increase the segmentation accuracy, especially in the case of MRI images where the Dice coefficients of the original U-nets are relatively low compared to the

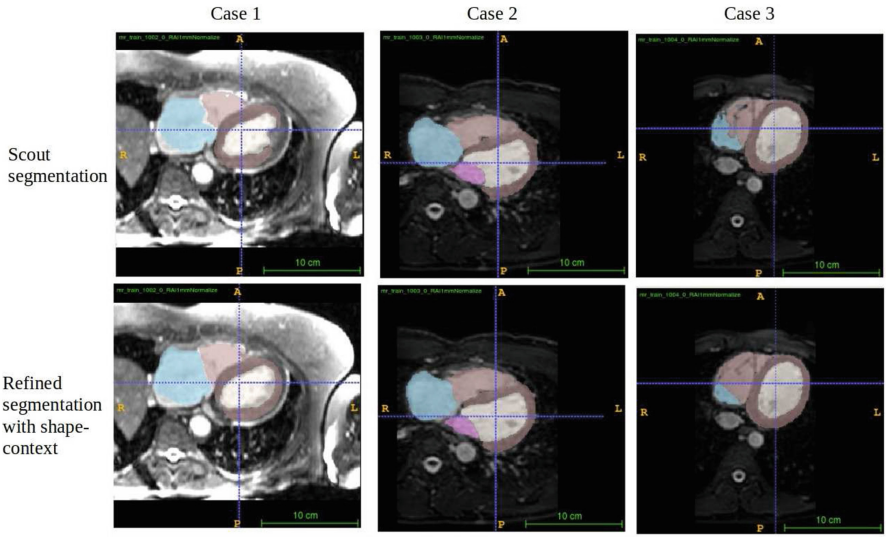


Fig. 3. Three example cases. The upper row shows the scout segmentation results, the lower row shows the refined results with shape context

Table 2. Comparing the accuracy (Dice coefficient) of the scout segmentation with U-net and the final segmentation with U-net and shape context in CT images

Structures	U-net	U-net & Shape context
Myocardium	0.872 ± 0.084	0.879 ± 0.068
Left atrium	0.903 ± 0.062	0.908 ± 0.067
Left ventricle	0.911 ± 0.035	0.935 ± 0.046
Right atrium	0.858 ± 0.105	0.881 ± 0.082
Right ventricle	0.884 ± 0.075	0.825 ± 0.082
Ascending aorta	0.956 ± 0.021	0.959 ± 0.023
Pulmonary artery	0.830 ± 0.125	0.815 ± 0.131

Table 3. Dice coefficient of the scout segmentation on the testing datasets

Structures	MRI	CT
Myocardium	0.728 ± 0.142	0.874 ± 0.051
Left atrium	0.832 ± 0.093	0.908 ± 0.044
Left ventricle	0.855 ± 0.136	0.908 ± 0.059
Right atrium	0.782 ± 0.131	0.855 ± 0.064
Right ventricle	0.760 ± 0.174	0.806 ± 0.082
Ascending aorta	0.771 ± 0.219	0.835 ± 0.231
Pulmonary artery	0.578 ± 0.246	0.677 ± 0.240
Whole heart	0.792 ± 0.246	0.866 ± 0.048

scores for CT images. For CT images, small improvements were also observed for a majority of the structures; however, the segmentation accuracy of RV declined considerably when the shape context channels were added. This may be due to an over-fitting problem, as the U-nets rely too much on the shape context channels. Adding dropout layers may be helpful to overcome this problem. Also, including more augmented training samples that are designed to model the uncertainty of the shape context layer could be even more important for the network to learn to cope with the cases where shape context is not very precise.

In previous studies, researchers have also proposed the auto-context method [12], which uses the output of the first classifier to train a second classifier. The advantage of using the shape context is that the statistical shape could help to eliminate some of the false positive regions that are produced by the first classifier but do not fit to the overall shape of the heart or the ventricles. However, direct comparison between auto-context approaches and the proposed shape context was not performed due to time constraints.

In conclusion, we have proposed a hybrid image segmentation methods based on deep neural network and statistical shape modeling. In our preliminary experiments, the proposed method delivered promising results on cardiac structure segmentation in both CT and MRI.

Acknowledgments. This research has been partially funded by the Swedish Research Council (VR), grant no. 2014-6153, and the Swedish Heart-Lung Foundation (HLF), grant no. 2016-0609.

References

1. Lloyd-Jones, D.M., Larson, M.G., Beiser, A., Levy, D.: Lifetime risk of developing coronary heart disease. Lancet **353**(9147), 89–92 (1999)
2. Zhang, X., Cowan, B.R., Bluemke, D.A., Finn, J.P., Fonseca, C.G., Kadish, A.H., Lee, D.C., Lima, J.A.C., Suinesiaputra, A., Young, A.A., et al.: Atlas-based quantification of cardiac remodeling due to myocardial infarction. PLoS ONE **9**(10), e110243 (2014)
3. Bergström, G., Berglund, G., Anders Blomberg, J., Brandberg, G.E., Jan Engvall, M., Eriksson, U.F., Flinck, A., Hansson, M.G., et al.: The swedish cardiopulmonary bioimage study: objectives and design. J. Intern. Med. **278**(6), 645–659 (2015)
4. Zheng, Y., Barbu, A., Georgescu, B., Scheuering, M., Comaniciu, D.: Four-chamber heart modeling and automatic segmentation for 3-D cardiac CT volumes using marginal space learning and steerable features. IEEE Trans. Med. Imaging **27**(11), 1668–1681 (2008)
5. Zhuang, X., Shen, J.: Multi-scale patch and multi-modality atlases for whole heart segmentation of MRI. Med. Image Anal. **31**, 77–87 (2016)
6. Long, J., Shelhamer, E., Darrell, T.: Fully convolutional networks for semantic segmentation. In: 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 3431–3440 (2015)
7. Ronneberger, O., Fischer, P., Brox, T.: U-Net: convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) MICCAI 2015. LNCS, vol. 9351, pp. 234–241. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-24574-4_28

8. Leventon, M.E., Grimson, W.E.L., Faugeras, O.: Statistical shape influence in geodesic active contours. In: 2000 Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, vol. 1, pp. 316–323. IEEE (2000)
9. Wang, C., Frimmel, H., Smedby, Ö.: Fast level-set based image segmentation using coherent propagation. Med. Phys. **41**(7), 073501 (2014)
10. Wang, C., Smedby, Ö.: Automatic multi-organ segmentation in non-enhanced CT datasets using hierarchical shape priors. In: Proceedings of the 22nd International Conference on Pattern Recognition (ICPR). IEEE (2014)
11. Wang, C., Wang, Q., Smedby, Ö.: Automatic heart and vessel segmentation using random forests and a local phase guided level set method. In: Zuluaga, M.A., Bhatia, K., Kainz, B., Moghari, M.H., Pace, D.F. (eds.) RAMBO/HVSMR -2016. LNCS, vol. 10129, pp. 159–164. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-52280-7_16
12. Tu, Z., Bai, X.: Auto-context and its application to high-level vision tasks and 3D brain image segmentation. IEEE Trans. Patt. Anal. Mach. Intell. **32**(10), 1744–1757 (2010)

Automatic Whole Heart Segmentation in CT Images Based on Multi-atlas Image Registration

Guanyu Yang[1,4(✉)], Chenchen Sun[1,4], Yang Chen[1,4], Lijun Tang[3], Huazhong Shu[1,4], and Jean-louis Dillenseger[2,4]

[1] Lab of Image Science and Technology, School of Computer Science and Engineering, Southeast University, Nanjing, China
yang.list@seu.edu.cn
[2] INSERM-U1099, LTSI, Université de Rennes 1, 35000 Rennes, France
[3] Department of Radiology, The First Affiliated Hospital of Nanjing Medical University, Nanjing, China
[4] Centre de Recherche en Information Biomedicale Sino-Francais (CRIBs), Nanjing, China

Abstract. Whole heart segmentation in CT images is a significant prerequisite for clinical diagnosis or treatment. In this work, we present a three-step multi-atlas-based method for obtaining a segmentation of the whole heart. In the first step, the region of the heart was detected by aligning the down-sampled patient CT with the low-resolution atlas images. The detected region of heart was used to crop the original patient image. In the second step, the registration between high-resolution atlas images and cropped original patient images was performed to obtain the precise segmentation of the heart. In the third step, the registration was performed again by minimizing the dissimilarity within the heart region. Finally, the labels of four cardiac chambers, aorta and pulmonary artery were generated according to the similarity between the deformed atlas images and the patient image. A leave-one-out experiment has been performed on the 20 training datasets of MM-WHS 2017 challenge. The average Dice coefficient between our segmentation results and the manual segmentation results is 0.9051. The mean and standard deviation of Dice coefficients of each structure (i.e. LV, RV, LA, RA, Myo, Ao, PA) are 0.9601 ± 0.0324, 0.9344 ± 0.0418, 0.9594 ± 0.0316, 0.8836 ± 0.0826, 0.8724 ± 0.0707, 0.9295 ± 0.0883, 0.7966 ± 0.1149 respectively.

Keywords: Whole heart segmentation · Heart · Segmentation
Multi-atlas · Cardiac chambers

This research was supported by National Natural Science Foundation under Grant (No. 31571001), and Science Foundation for The Excellent Youth Scholars of Southeast University.

M. Pop et al. (Eds.): STACOM 2017, LNCS 10663, pp. 250–257, 2018.
https://doi.org/10.1007/978-3-319-75541-0_27

1 Introduction

Cardiovascular disease, with increasing mortality and morbidity, has been threatening human health globally. Whole heart segmentation in CT images is a significant prerequisite for clinical diagnosis or treatment, providing not only the anatomical morphology but also the functional information of each cardiac chamber. Some approaches have been proposed for whole heart segmentation in CT images. Ecabert [1] introduced a method based on statistical shape model, with model adaption to increase segmentation accuracy. Kirişli et al. [2] proposed a multi-atlas based method and did an evaluation with a leave-one-out strategy. Recently, Zhuang [3] proposed a framework M^3AS for whole heart segmentation of MRI. The method is based on a multi-scale patch strategy and a label fusion algorithm using both global and local weights.

In this work, we present a three-step multi-atlas-based method for obtaining a segmentation of whole heart, including left and right ventricles, left and right atrium, aorta, pulmonary artery in CT images. This method mainly relies on image registration between the patient image and multiple atlas images, which is to generate segmentation results in each step. Experimental results show that our proposed multi-atlas based method can segment whole heart accurately.

2 Methodology

2.1 A Three-Step Multi-atlas-Based Whole Heart Segmentation

In this multi-atlas registration method, the patient dataset is aligned with each atlas image to map the reference labels of the atlas images to the patient image. A coarse-to-fine segmentation result is obtained during the registration process. Then, the deformed labels from several atlas image, which is selected by a strategy, are fused to generate the final segmentation result according to a specified fusion criterion.

This multi-atlas registration method chiefly relies on the image registration process which can be described as an optimization problem to find an optimal transformation $\hat{T}$ between a fixed image $I_P(x)$ and a moving image $I_A(x)$, and the expression (1) illustrates this transformation.

$$\hat{T} = \arg\min_{T} C(I_P(x), I_A(T(x))) \tag{1}$$

In this expression, T(x) is a transformation function which deforms $I_A(x)$ to align with $I_P(x)$ spatially. C is the cost function to measure the dissimilarity between $I_P(x)$ and deformed $I_A(T(x))$. C is minimized iteratively by an optimization algorithm. In this work, the mutual information defined in [4] is adopted to measure the dissimilarity due to a large difference of intensity distribution between CT images. The adaptive stochastic gradient descent method [5] is used to optimize the cost function. Owing to the high computational cost in multi-atlas registration, we applied some techniques to reduce the computation time. In each iteration, a randomly sampled subset of image voxels is selected to be measurement of the dissimilarity and the number of voxels in

the sampled subset is set to 2048. A multi-resolution strategy based on Gaussian pyramid is also used during the registration process.

In the cardiac CT images, the strong boundaries of tissues surrounding the heart, such as lung and rib cage, the difference scanning range will lead to an inaccurate transformation parameters of image registration. Therefore, the segmentation results could be inaccurate if we register the whole CT image with the atlas image as the method in [6] did. So, in this paper, we introduce a three-step method to generate the coarse-to-fine segmentation result, in order to generate more accurate results. Figure 1 describes the framework of the proposed method which consists of three steps: (1) patient image ROI definition, (2) precise segmentation within ROI, (3) similarity computation, no-grid registration and label fusion.

In the first step, a ROI detection based on registration is performed to locate the region of heart in the patient image. In this step, the atlas images and label images are down-sampled and resized to generate low-resolution atlas images $I_{A_i}^L (i = 1, \ldots, N)$ which are then aligned with the down-sampled patient image to obtain the coarse segmentation of heart. Both the atlas and patient images will be down-sampled to an isotropic image with the size of each axial slice of 128 × 128. The region of interest (ROI) is defined by the low resolution label images to locate the region of heart. In the second step, the heart is cropped from the original patient image by the ROI defined in the previous step. Then the cropped high-resolution patient image are registered with the high-resolution atlas images $I_{A_i}^H (j = 1, \ldots, M)$ to obtain the precise segmentation of heart. In this step, two registration methods are applied. Affine registration is used to align the patient CT image with atlas images roughly and this is followed by B-spline registration which is to refine the spatial transformation. In the third step, only B-spline non-rigid registration is re-performed with the mask of heart estimated in the previous step in order to refine the transformation from atlas images to patient images. The MI between affine registration results and patient image is computed to measure the dissimilarity. Then the atlas images is ranked according to the dissimilarity. To generate a final segmentation of the patient image, many methods adopted majority voting. Regarding some situation that several atlas images have low similarity with patient images, we employ a strategy to do a selection in these atlas images by comparing the similarity $\mathrm{S}(I_{D_i}, I_p)$ between its deformed results and patient image. The similarity $\mathrm{S}(\mathrm{X}, \mathrm{Y})$ is defined as $\mathrm{S}(\mathrm{X}, \mathrm{Y}) = \sum_{i=1}^{n} (X_i - Y_i)^2$, in which n is the total number of voxels in the ROI of the image. An atlas image will be selected if its similarity with the patient image $\mathrm{S}(I_{D_i}, I_p)$ is above the mean value $S_M = (1/M) \sum_i^M \mathrm{S}(I_{D_i}, I_p)$. According to this strategy, m atlas images will be selected and m is usually less than $M/2$ in our experiment. Then a global fusion is performed on the deformed labels of the selected atlas images to generate the final segmentation of whole heart. The global weight w_i of each selected atlas is defined as

$$w_i = s(I_{D_i}, I_P) \Big/ \sum_{j}^{m} s(I_{D_i}, I_P) \tag{2}$$

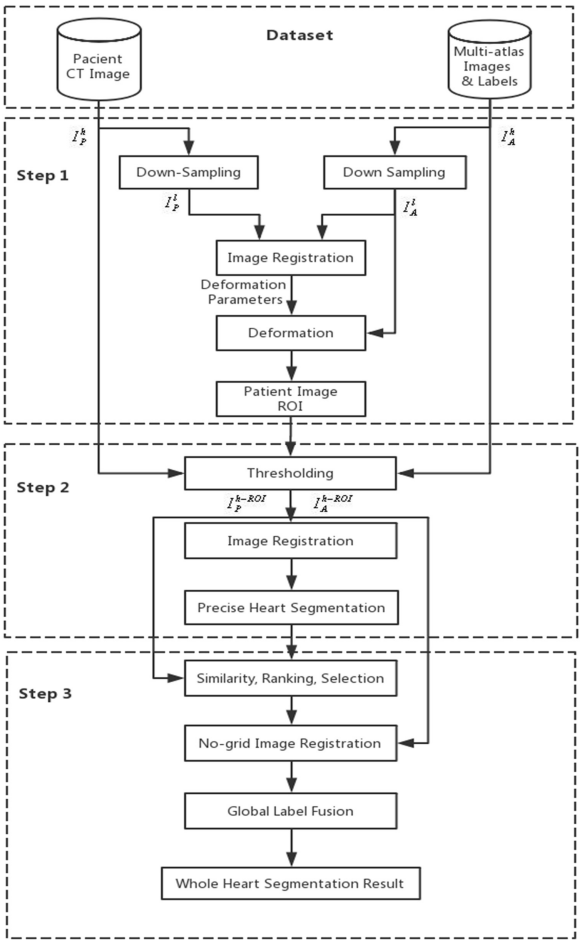


Fig. 1. The framework of the proposed method for automatic whole heart segmentation. I_P^l, I_P^h: low resolution and high resolution patient images, I_A^l, I_A^h: low resolution and high resolution atlas images, I_P^{h-ROI}, I_A^{h-ROI}: patient images and atlas images cropped from high resolution images by the ROI.

During the second and third step, registrations are executed on the ROI of the images which is defined in the first step. Because the whole heart is the major part of the ROI in images, it can be matched more accurately between patient CT images and atlas images. In all the three steps, the limitation of iteration times in the optimization of cost function C is fixed to 500. The cubic B-spline interpolator is used to estimate the values of the voxels at non-voxel positions. The B-spline grid is defined by control points with 16 mm interval. To generate the segmentation results, the labels of selected atlas images are then fused by the global weights which are defined according to the similarity.

2.2 Multiple Atlas Images

Sixty multi-modality cardiac images were used in the experiment, which are provided by the Challenge Multi-Modality Whole Heart Segmentation (MM-WHS 2017). These data was acquired in real clinical environment, which indicates its various quality. Some images with poor quality were included to test the robustness of the algorithms in real clinical situation.

The cardiac CT/CTA images were obtained at Shanghai Shuguang Hospital, China, using routine cardiac CT angiography. Each data consists of the whole heart, covering the region from upper abdominal to the aortic arch. The slices were obtained in axial view and the inplane resolution is about 0.78 × 0.78 mm and the slice thickness is 1.60 mm on average.

The data comprises two types, which are training data (20 CT images) and test data (40 CT images). Training data were segmented manually into seven whole heart substructures [1, 2]:

- the left ventricular cavity;
- the right ventricular blood cavity;
- the left atrial cavity;
- the right atrial blood cavity;
- the myocardium of the left ventricle;
- the ascending aorta trunk from the aortic valve to the superior level of the atria;
- the pulmonary artery trunk between the pulmonary valve and the bifurcation point.

These structures above in training data were manually labeled with different values. Due to labels of test data is not provided, we also employed training data for evaluation of this method, using a leave-one-out test.

As described in the last section, the proposed method is based on multi-atlas registration and processed images are used. The patient images and atlas images used in this method have two scales: low resolution whole images and high resolution cropped images. Original atlas images and label images were down-sampled to generate low resolution whole images. The axial sizes were different among these original images and they were resized to low resolution images whose axial size is 128 × 128. To acquire high resolution images, the whole heart was cropped from the original images using region of heart defined by the labels of heart. The same option is performed on patient images. Not all atlas images are suitable for segmenting a patient image because the dissimilarity between them is large. In this paper, we did a selection among the atlas images, N and M in $I_{A_i}^{L}(i = 1, \ldots, N)$ and $I_{A_i}^{H}(j = 1, \ldots, M)$ are fixed to be 8.

3 Experimental Results

The three-step multi-atlas-based whole heart segmentation framework is implemented in Mevislab (http://www.mevislab.de/). The image registration is performed by an open-source package named ELASTIX. The images used in the experiment were provided and segmented by MM-WHS 2017 challenge.

First, we performed our method on the 40 test images to yield segmentation results and the test results were submitted to organizers of MM-WHS 2017 challenge for evaluation. Figure 2 illustrates the segmentation results from a test patient image. The results demonstrates that this proposed method has the capability to segment the whole heart accurately. The evaluation results of test datasets will be provided on the day of MM-WHS2017 challenge event. We have to provide quantitative evaluation of our proposed method with leave-one-out test of training datasets.

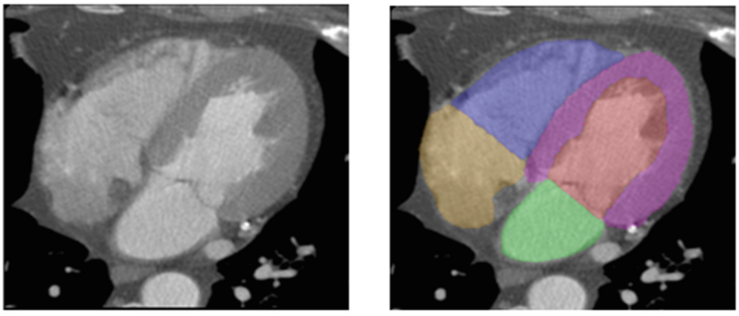


Fig. 2. Segmentation results of whole heart by our proposed method. (a) cropped high resolution patient image. (b) segmentation result obtained in the cropped image.

Second, we applied leave-one-out test with 20 labeled atlas images to evaluate our proposed segmentation method. Each atlas image was segmented by the proposed method with the other 19 atlas images. The segmentation result of each structure (i.e. LV, RV, LA, RA, Myo, Ao, PA) from a training image is displayed by Fig. 3. At the same time, Dice similarity coefficient (DSC) between the segmentation results and reference atlas labels was calculated to evaluate the accuracy of this method.

$$\mathrm{DSC} = \frac{2(M \cap R)}{M + R} \tag{3}$$

The DSCs of each structure in 20 training images are displayed in Table 1. The mean and standard deviation of Dice coefficients of each structure (i.e. LV, RV, LA, RA, Myo, Ao, PA) are 0.9601 ± 0.0324, 0.9344 ± 0.0418, 0.9594 ± 0.0316, 0.8836 ± 0.0826, 0.8724 ± 0.0707, 0.9295 ± 0.0883, 0.7966 ± 0.1149 respectively. These quantitative evaluations show that our method can generate accurate whole heart segmentation in 3D CT images.

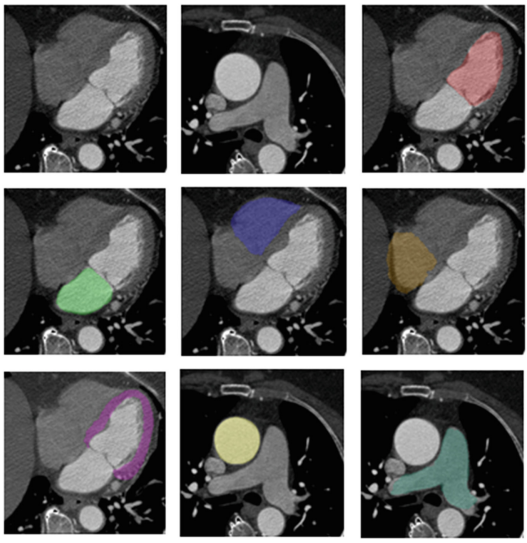


Fig. 3. The original training image and segmentation result of each structure (i.e. LV, RV, LA, RA, Myo, Ao, PA).

Table 1. Dice similarity coefficient between reference labels and segmentation results of each structure of heart.

DSC of each structure of proposed method							
Atlas	LV	RV	LA	RA	Myo	Ao	PA
No. 1	0.9738	0.9932	0.9985	0.9452	0.8949	0.9902	0.9845
No. 2	0.9293	0.8244	0.9680	0.9600	0.7934	0.9859	0.9279
No. 3	0.9895	0.9446	0.9508	0.8706	0.9227	0.9802	0.8177
No. 4	0.9068	0.9241	0.9939	0.8951	0.8184	0.9732	0.8408
No. 5	0.9915	0.9530	0.9079	0.6784	0.8860	0.9941	0.7487
No. 6	0.8683	0.9428	0.9768	0.9651	0.8656	0.7442	0.9046
No. 7	0.9586	0.9398	0.9284	0.8710	0.8698	0.9855	0.8127
No. 8	0.9847	0.9987	0.9802	0.8444	0.8292	0.9910	0.7908
No. 9	0.9585	0.9443	0.9361	0.8963	0.9039	0.9858	0.8074
No. 10	0.9721	0.9526	0.9809	0.9748	0.9620	0.9593	0.8605
No. 11	0.9930	0.8941	0.9796	0.9657	0.9248	0.7952	0.7995
No. 12	0.9852	0.9630	0.9630	0.9008	0.8886	0.9943	0.7747
No. 13	0.9636	0.9338	0.9872	0.8951	0.6786	0.9824	0.8124
No. 14	0.9419	0.8957	0.9342	0.9224	0.7518	0.9095	0.7040
No. 15	0.9673	0.8910	0.9513	0.8332	0.9053	0.9374	0.5936
No. 16	0.9745	0.9646	0.9897	0.8909	0.8627	0.8373	0.7211
No. 17	0.9928	0.9409	0.9568	0.8481	0.9030	0.7465	0.7805
No. 18	0.9701	0.9039	0.9054	0.8344	0.9665	0.9816	0.5177
No. 19	0.9222	0.9896	0.8999	0.6953	0.8741	0.8184	0.7377
No. 20	0.9573	0.8928	0.9984	0.9850	0.9467	0.9979	0.9957
Avg.	0.9601	0.9344	0.9594	0.8836	0.8724	0.9295	0.7966

4 Conclusion

In this paper, an automatic method for segmenting whole heart in CT images is presented. The method mainly relies on a three-step multi-atlas image registration to acquire the segmentation results. In the first step, down-sampled patient image is registered with low resolution atlas images to detect the region of heart which was used to crop the original patient image. In the second step, the registration between cropped original patient images and high-resolution atlas images was performed to obtain the precise segmentation of the heart. In the third step, the registration was performed again by minimizing the dissimilarity within the heart region. Finally, the labels of four cardiac chambers, aorta and pulmonary artery were acquired respectively according to the similarity between patient image and deformed atlas images. Experimental results in 20 training images and 40 test images show that this method can generate a accurate whole heart segmentation.

References

1. Ecabert, O., Peters, J., Schramm, H., et al.: Automatic model-based segmentation of the heart in ct images. IEEE Trans. Med. Imaging **27**(9), 1189–1201 (2008)
2. Kirişli, H.A., Schaap, M., Klein, S., et al.: Evaluation of a multi-atlas based method for segmentation of cardiac CTA data: a large-scale, multicenter, and multivendor study. Med. Phys. **37**(12), 6279–6291 (2010)
3. Zhuang, X., Shen, J.: Multi-scale patch and multi-modality atlases for whole heart segmentation of MRI. Med. Image Anal. **31**, 77 (2016)
4. Thevenaz, P., Unser, M.: Optimization of mutual information for multiresolution image registration. IEEE Trans. Image Process. **9**(12), 2083–2099 (2000)
5. Klein, S., Pluim, J.P.W., et al.: Adaptive stochastic gradient descent optimisation for image registration. Int. J. Comput. Vis. **81**(3), 227 (2009)
6. Kirisli, H.A., Klein, S., et al.: Fully automatic cardiac segmentation from 3D CTA data: a multi-atlas based approach. In: Proceedings of SPIE - The International Society for Optical Engineering, vol. 7623 (2010)

Author Index