Deep learning for automated analysis of cardiac imaging: applications in Cine and 4D flow MRI

Sun, X.

Chapter 4 Right Ventricle Segmentation via Registration and Multi-input Modalities in Cardiac Magnetic Resonance Imaging from Multi-disease, Multi-view and Multi-center

This chapter was adapted from:

Xiaowu Sun, Li-Hsin Cheng, Rob J. van der Geest. **Right Ventricle Segmentation via Registration and Multi-input Modalities in Cardiac Magnetic Resonance Imaging from Multi-disease, Multi-view and Multi-center.** International Workshop on Statistical Atlases and Computational Models of the Heart (STACOM). Springer, Cham, 2021.

Abstract

Quantitative assessment of cardiac function requires accurate segmentation of cardiac structures. Convolutional Neural Networks (CNNs) have achieved immense success in automatic segmentation in cardiac magnetic resonance imaging (cMRI) given sufficient training data. However, the performance of CNN models greatly degrade when the testing data is from different vendors or different centers. In this paper, we introduce the use of image registration to propagate annotation masks from labeled images to unlabeled images as to enlarge the training dataset. Furthermore, we investigated various input modalities including 3D volume, single-channel 2D image, multi-channel 2D image constructed from spatial and temporal stack to extract more features to improve do-main generalization in cMRI segmentation. We evaluated our method in M&Ms-2 challenge testing data (https://www.ub.edu/mnms-2/), achieving averaged Dice scores of 0.925, 0.919 and Hausdorff Distance of 10.587 mm, 6.045 mm in right ventricular segmentation in short-axis view and long-axis view respectively.

4.1 Introduction

In clinical routine, cardiac magnetic resonance imaging (cMRI) is considered a standard reference for the diagnosis of cardiac disease. Accurate segmentation of cardiac structures such as left ventricle (LV), myocardium and right ventricle (RV) is essential to quantitatively assess the cardiac function. Traditional manual segmentation method not only is time-consuming but also prone to inter-rater experience.

In recent years, deep learning based automatic segmentation approaches have been achieved immense success in cardiac segmentation. Tran et al. was the first to employ the fully convolutional neural (FCN) network for LV and RV segmentation in short-axis MRI [1]. Poudel proposed a recurrent FCN network ensembling the spatial information for LV segmentation [2]. However, the performance of most of those deep learning based models degrade dramatically when the trained model is applied directly on other unseen datasets from different centers or vendors. Differences in image protocols, disease characteristics, scanner-specific bias and the other factors remain even after careful pre-processing [3]. In addition, the RV has a more complex shape and border characteristics compared to the LV. Hence, the M&Ms-2 challenge is motivated to build a method to segment the RV using multi-center, multi-disease and multi-view cMRI data.

The most straight forward approach to tackle this problem is to collect and annotate data from multiple centers, vendors and patient pathologies. Tao used a large heterogeneous data with 41,593 images from different centers and different vendors to train a CNN model and achieved a good generalization [4]. Chen demonstrated that applying data augmentation strategies on a single-site single-scanner dataset could improve the performance on an unseen dataset across different sites or scanners [5]. Based on those studies, we hypothesize that a large-scaled pooling data from different domains could improve a model's performance on an unseen dataset. Additionally, in the conventional CNN models, the information derived from the neighboring images is usually ignored. Hence, we introduced two stack model to extract the spatiotemporal features to improve the performance.

In this paper, given limited data, we investigated several methods to generate more training data and extract more features including 1): The use of image registration to propagate annotation masks to unlabeled phases 2): Introducing the spatial and temporal neighboring images to construct a multi-channel 2D image to integrate more spatiotemporal information for the RV segmentation task.

4.2 Data

Table.4.1. Description of training, validation and testing dataset

Pathology	Num. of training	Num. of validation	Num. of testing
Normal subjects	40	5	30
Dilated Left Ventricle	30	5	25
Hypertrophic Cardiomyopathy	30	5	25
Congenital Arrhythmogenesis	20	5	10
Tetralogy of Fallot	20	5	10
Interatrial Communication	20	5	10
Dilated Right Ventricle	0	5	25
Tricuspid Valve Regurgitation	0	5	25
Total	160	40	160

The M&Ms-2 challenge provides 360 cases (160 for training, 40 for validation and 160 for testing) in both short-axis (SA) and long-axis (LA) views from four different centers, acquired with three vendors (General Electric, Philips and Siemens). As shown in Table.4.1, except the normal subjects, there are five pathologies in the training dataset, two pathologies are not present in the training dataset but only in the validation and testing dataset. In addition, only end-diastolic (ED) and end-systolic (ES) phases in the training data are annotated by experienced experts, including LV, RV and left ventricular myocardium (MYO). Although this challenge focused on the RV segmentation, in our experiments, LV and MYO annotations were also used to constrain the RV segmentation.

4.3 Method

4.3.1. Registration

In the available dataset, only the ED and ES phases are labeled, while the other phases are continuous in time consistent with the ED and ES phases. All the phases from the same case have an almost identical intensity distribution, which will alleviate the errors caused by inter-subject variability [6]. Hence, we used intensity-based registration method to propagate the labels, regarding the ED and ES as the template. The progress is described in Figure.4.1. Given three phases (ED, ES and unlabeled), the ED and ES are firstly registered to the unlabeled phase, generating two geometric transformation matrixes named ES-Tf and ED-Tf, then the transformation matrix with smaller norm was used to propagate the mask. Matlab inbuilt functions imregtform and imwarp were used to implement the registration

[11]. Mean square error (MSE) and Affine were set as the similarity metric and transformation type.

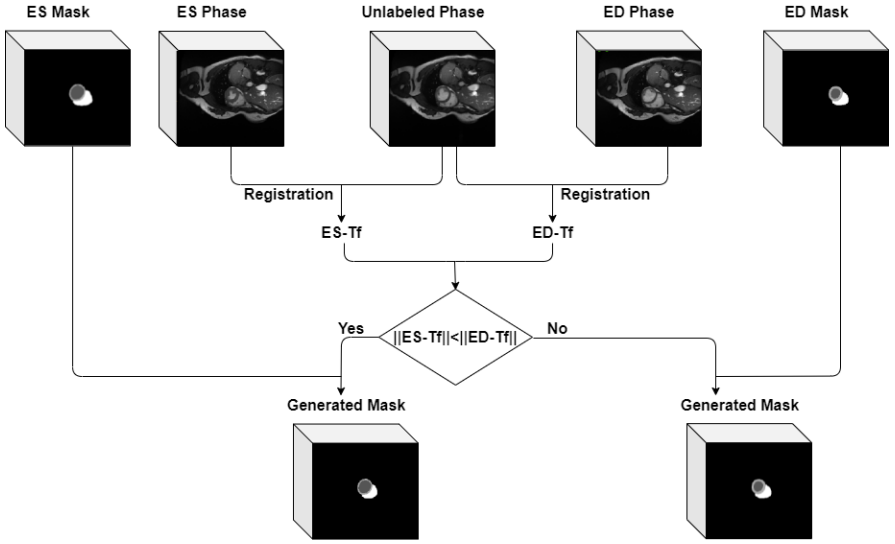

Figure.4.1. Registration method to generate a mask for a unlabeled phase in SA view.

4.3.2. Input modality of network

A illustrated in Figure.4.2 a short-axis cine MRI scan contains multiple slices and multiple phases. Images from the same slice level describe a cardiac cycle, while images from the same phase describe the complete heart structure. In conventional methods [4,5], each single-channel 2D image or a 3D volume with the whole images from the same phase is usually considered as the input of a network. Although using a single-channel 2D image as the input could enlarge the training dataset, a 3D volume can provide more spatial information for the segmentation than using a 2D image. As a compromise, a spatial stack or temporal stack model, as proposed in our previous work [7], can be used to build a multi-channel 2D image which can provide accompanying spatial or temporal information respectively.

Spatial Multi-channel 2D image (SMI). The slices from the same phase are used to construct a SMI. As illustrated in Figure.4.2, three 2D images from Phase 9 Slice 5,6,7 are used to build a 3-channels SMI for the image of Slice 6 Phase 9.

Temporal Multi-channel 2D image (TMI). In a similar way, a TMI consists of several neighboring phases of a particular slice. As shown in Figure.4.2, images from Slice 6 Phase 8,9,10 are used to construct a 3-channels TMI for the image of Slice 6 Phase 9.

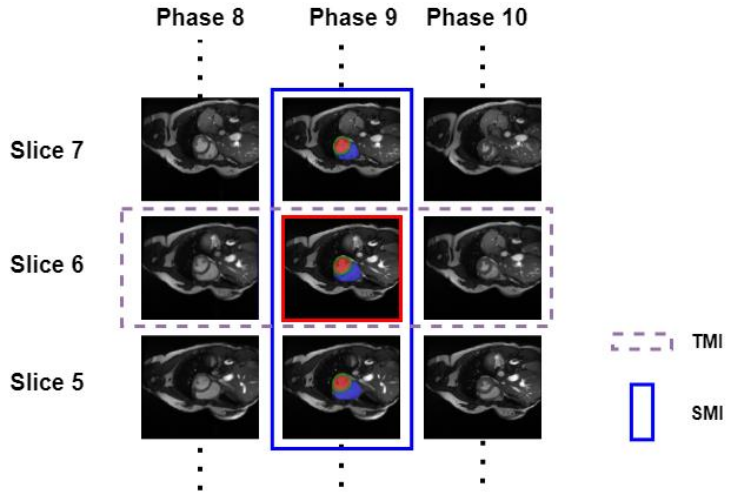

Figure.4.2. An example of constructing a spatial multi-channel 2D image and temporal multi-channel 2D image. The image in the red box is the target image which will be segmented. The three spatial neighboring images in the blue box is called an SMI with three channels, where the top one is the first channel, the middle one is the second channel and so on. The TMI consists of three temporal neighboring images in the dash-line box, the left one is the first channel and the right is the last channel.

Table.4.2 shows a brief summary of the training data size in SA view after combining the registration, SCI (single-channel 2D image), SMI and TMI. The original MnMS-2 dataset contains 320 3D volumes and 2,704 2D images with labeled annotation for training, when applying the registration approach to propagate the annotation masks, the data size of 3D volume and single-channel 2D image increased to 4,152 and 32,330 respectively. The LA-view images were acquired as single slice, resulting in the LA images being multi-phase single-slice. The 3D volumes and SMI cannot be constructed in LA view. Hence, the model achieving the best performance in SA view was used as the pre-trained model for the LA view instead of using different input modalities.

Table.4.2. Training dataset description in SA MRI. **SCI**: single-channel 2D image. The number of channel in SMI and TMI is set to 5.

Data modality	Used registration	Training Data Size
SCI	No	2,704
SCI	Yes	32,330
3D volume	No	320
3D volume	Yes	4,152
SMI	No	2,704
SMI	Yes	32,330
TMI	No	2,704
TMI	Yes	32,330

4.3.3. Network Architecture

nnUNet [8] based on the U-Net architecture is a fully automatic and out-of-the-box medical image segmentation framework. To improve the robustness of domain shift in cardiac MRI, nnUNet_MMS [9] was specially designed by investigating various data augmentation techniques and ranked first at the first edition of M&Ms [10]. Hence, we introduced nnUNet as the baseline, and built our method upon nnUNet_MMS. All the models in this study are based on a 2D network. The data augmentation methods are the same in nnUNet_MMS model.

Since the propagated masks are not as accurate as the manual masks, those pseudo data was used to pre-train the model and the manually labeled data was applied to fine-tune the pre-trained network. The results are reported using Dice and Hausdorff Distance (HD). All experiments were executed on an NVIDIA Quadro RTX 6000 GPU with 24 GB internal memory.

4.4 Experiments and Results

4.4.1. Validation Set Results

We first evaluated the performance of different networks with different input modalities in the SA view from the validation dataset. Then we compared the results in the LA view with or without pre-training from SA view.

Table.4.3. Segmentation results generated from different networks with different input modalities in the validation dataset in SA view.

Network	Input	Used registration data to pre-train network	Dice	HD (mm)
nnUNet(Baseline)	3D volume	No	0.912	10.318
nnUNet_MMS	3D volume	No	0.915	10.475
	3D volume	Yes	**0.922**	**9.472**
	SMI	No	0.919	9.577
	SMI	Yes	0.920	9.539
	TMI	No	0.917	10.343
	TMI	Yes	0.914	12.221
	SCI	No	0.915	11.354
	SCI	Yes	0.914	10.515

Table.4.3 shows that using 3D volume without registration processing as the input, nnUNet_MMS achieved a slightly better dice than nnUNet, but yielded worse HD. However, when the registration method is applied to generate more 3D volume data to pre-train nnUNet_MMS, it achieved the best performance with a Dice of 0.922 and HD of 9.472 mm. It also can be observed that the segmentation results derived from the two stack models (SMI and TMI) are better than that from SCI, which

confirmed that SMI and TMI could provide more spatial and temporal information for the segmentation task.

The results in LA views presented in Table.4.4 illustrates that the performance increased by 0.01 and 0.7 in terms of Dice and HD as a result of transferring the pre-trained model from SA view to LA view. Figure.4.3 and Figure.4.4 show some segmentation examples derived from the best model.

Table.4.4. Segmentation results in the validation dataset in LA view.

Network	Transfer from SA	Dice	HD
nnUnet(Baseline)	No	0.910	6.004
nnUNet_MMS	No	0.908	6.081
	Yes	**0.920**	**5.343**

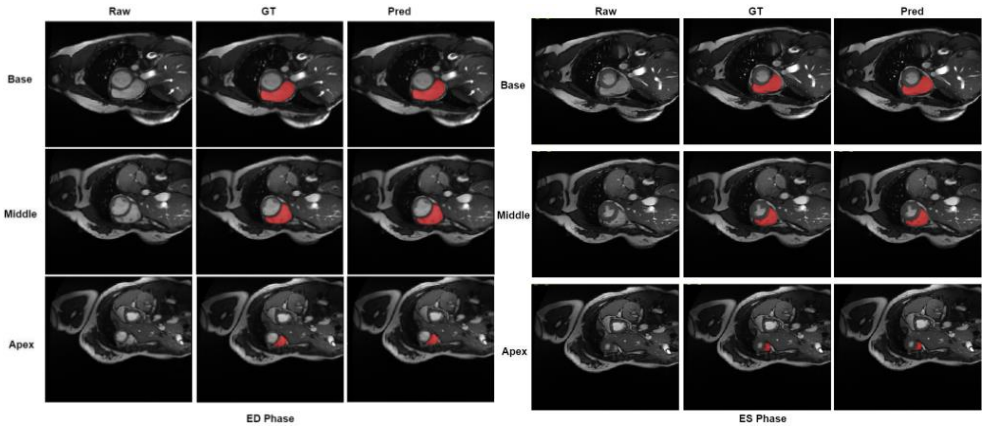

Figure.4.3. A visual example from the apex, middle and base levels at ED (left) and ES (right) phases in SA view.

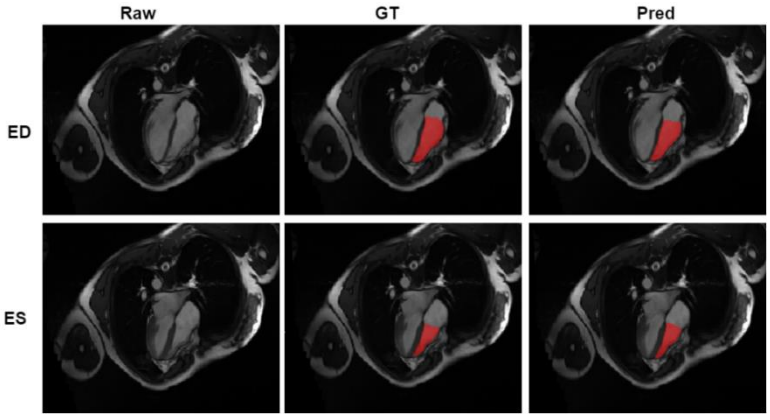

Figure.4.4. A visual example at ED and ES phase in LA view.

4.4.2. Testing Set Results

We chose the model which performs best in the validation data as the final model. As the testing dataset is hidden by the organizer, we submitted our final model to the organizer and evaluated the performance online. Table.4.5 shows the details of our method on the hidden test data. In the SA view our method performed best in congenital arrhythmogenesis yielding 0.949, 8.45 mm for Dice and HD. The best results in LA view are generated from the normal subjects with 0.935 and 5.006 mm for Dice and HD. In addition, two pathologies (dilated right ventricle and tricuspid valve regurgitation) are not present in the training data but only in the testing data. The results on those two pathologies reveal that our approach obtains promising performance on an unseen pathology.

Table.4.5. Segmentation results on 8 pathologies of the hidden test set. The mean and standard deviation are reported.

	Dice		HD (mm)	
Pathology	SA	LA	SA	LA
Normal subjects	0.922±0.050	**0.935± 0.035**	8.999±4.540	**5.006±2.657**
Dilated Left Ventricle	0.922±0.084	0.915± 0.052	13.257±13.134	5.944±3.547
Hypertrophic Cardiomyopathy	0.934±0.057	0.932± 0.033	10.214±5.842	5.343±2.916
Congenital Arrhythmogenesis	**0.949±0.028**	0.934± 0.031	**8.450±4.838**	5.125±1.738
Tetralogy of Fallot	0.920±0.034	0.914± 0.037	14.157±8.232	7.404±3.673
Interatrial Communication	0.910±0.048	0.906±0.066	12.045±4.189	8.021±6.089
Dilated Right Ventricle	0.924±0.045	0.897± 0.121	10.397±5.223	7.064±5.091
Tricuspidal Regurgitation	0.923±0.040	0.914± 0.039	9.236±3.675	6.112±3.349
Overall	0.925±0.055	0.919±0.063	10.587±7.241	6.045±3.824

4.5 Conclusion

In this paper, we investigated label propagation and multiple input modalities to increase the robustness in right ventricle segmentation from multi-disease, multi-view and multi-center cMRI data. To enlarge the training dataset, we explored the use of image registration to propagate annotation masks to unlabeled phases. We further systematically investigated the effect of using different input modalities including 3D volumes, single-channel 2D image, spatial stack and temporal stack. The results illustrate that spatial stack and temporal stack provide more information for the segmentation task, and using 3D volume with label propagation could further improve the generalization ability in a unseen dataset.

Declaration. The authors of this paper declare that the segmentation methods implemented in this challenge has not used any pre-trained models nor additional MRI datasets other than those provided by the organizers.

References

1. Tran, Phi Vu. "A fully convolutional neural network for cardiac segmentation in short-axis MRI." arXiv preprint arXiv:1604.00494 (2016).
2. Poudel, Rudra PK, Pablo Lamata, and Giovanni Montana. "Recurrent fully convolutional neural networks for multi-slice MRI cardiac segmentation." Reconstruction, segmentation, and analysis of medical images. Springer, Cham, 2016. 83-94.
3. Glocker, Ben, et al. "Machine learning with multi-site imaging data: An empirical study on the impact of scanner effects." arXiv preprint arXiv:1910.04597 (2019).
4. Tao, Q., et al.: Deep learning-based method for fully automatic quantification of left ventricle function from cine MR images: a multivendor, multicenter study. Radiology 290(1), 81–88 (2019)
5. Chen, Chen, et al. "Improving the generalizability of convolutional neural network-based segmentation on CMR images." Frontiers in cardiovascular medicine 7 (2020): 105.
6. Zhang, Yao, et al. "Semi-supervised Cardiac Image Segmentation via Label Propagation and Style Transfer." International Workshop on Statistical Atlases and Computational Models of the Heart. Springer, Cham, 2020.
7. Sun, Xiaowu, et al. "SAUN: Stack attention U-Net for left ventricle segmentation from cardiac cine magnetic resonance imaging." Medical Physics 48.4 (2021): 1750-1763.
8. Isensee, Fabian, et al. "nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation." Nature methods 18.2 (2021): 203-211.
9. Full, Peter M., et al. "Studying robustness of semantic segmentation under domain shift in cardiac MRI." International Workshop on Statistical Atlases and Computational Models of the Heart. Springer, Cham, 2020.
10. Campello, Víctor M., et al. "Multi-centre, multi-vendor and multi-disease cardiac segmentation: The M&Ms challenge." IEEE Transactions on Medical Imaging (2021).
11. “Intensity-based automatic image registration - MATLAB& Simulink.” [Online]. Available: https://www.mathworks.com/help/images/intensity-based-automatic-image-registration.html.