Pseudo-3D Network for Multi-sequence Cardiac MR Segmentation

Tao Liu[1] , Yun Tian[1(✉)] , Shifeng Zhao[1], XiaoYing Huang[1], Yang Xu[1], Gaoyuan Jiang[1], and Qingjun Wang[2]

[1] Beijing Normal University, Beijing 100875, China
`liutao@mail.bnu.edu.cn, tianyun@bnu.edu.cn`
[2] Sixth Medical Center of PLA General Hospital, Beijing 100048, China

Abstract. Deep learning approaches have been regarded as a powerful model for cardiac magnetic resonance (CMR) image segmentation. However, most current deep learning approaches do not fully utilize the information from multi-sequence (MS) cardiac magnetic resonance. In this work, the deep learning method is used to fully-automatic segment the MS CMR data. The balanced-Steady State Free Precession (bSSFP) cine sequence is used to perform left ventricular positioning as a priori knowledge, and then the Late Gadolinium Enhancement (LGE) cine sequence is used for precise segmentation. This segmentation strategy makes full use of the complementary information from the MS CMR data. Moreover, to solve the anisotropy of volumetric medical images, we employ the Pseudo-3D convolution neural network structure to segment the LGE CMR data, which combines the advantage of 2D networks and preserving the spatial structure information in 3D data without compromising segmentation accuracy. Experimental results of the Multi-sequence Cardiac MR Segmentation Challenge (MS-CMRSeg 2019) show that our approach has achieved gratifying results even with limited GPU computing resources and small amounts of annotated data. The full implementation and configuration files in this article are available at https://github.com/liut969/Multi-sequence-Cardiac-MR-Segmentation.

Keywords: Multi-sequence · Pseudo-3D network · Segmentation

1 Introduction

Heart disease is the leading cause of death globally, cardiac magnetic resonance (CMR) imaging is the gold-standard for assessment and diagnosis of a wide range of heart diseases. Usually, the ventricle and myocardium need to be manually segmented from the CMR data by clinicians, and then ventricle volume, mass and ejection fraction can be calculated from the segmentation results to diagnose the heart disease. With the increasing medical image data, time-consuming, laborious and tedious manual segmentation methods are considered to be inefficient. Therefore, it is imperative to develop computer-aided techniques to analyze medical images automatically [6].

Multi-sequence (MS) CMR usually include three-sequence CMR images: the Late Gadolinium Enhancement (LGE) cine sequence, the T2-weighted (T2) and the balanced-Steady State Free Precession (bSSFP) cine sequence. The difficulties of MS CMR segmentation have been mainly composed of the following points [12,13]: (i) CMR image presence poor contrast between the myocardium and the surrounding structure, for example, in LGE CMR, the infarcted myocardium is similar to the blood pools, and the healthy myocardium is similar to the adjacent liver or lung; (ii) the location, size and shape of the heart are different in different people, and the lesions exacerbate this difference; (iii) efficient fusion strategies are lacking to take fully utilize the information from MS CMR data; (iv) some other factors, such as the inherent noise caused by motion artifacts and cardiac dynamics. Therefore, ventricular segmentation based on MS CMR data is still a challenging task.

Automatic heart segmentation and diagnosis has become more and more necessary. In the last decade, the international challenge has released a large number of CMR datasets and brought together the state-of-the-art methods. The automatic CMR data segmentation method based on deep learning has achieved gratifying results. For example, in the Automated Cardiac Diagnosis Challenge - MICCAI 2017[1], the 8 highest-ranked segmentation methods were all neural network-based methods, so deep learning approaches have been regarded as a powerful model for CMR image segmentation.

In this work, we employed the deep learning method to fully-automatic segment the MS CMR data. The main contributions of this study consist of the following:

– We segment the ventricles combining the complementary information from two-sequence CMR data. The bSSFP cine sequence is used to perform left ventricular positioning as a priori knowledge, and then the LGE cine sequence is used for precise segmentation. Our segmentation strategy makes full use of the complementary information in the MS CMR data.
– In order to solve the anisotropy of volumetric medical images [1], the Pseudo-3D [8] convolution neural network structure is used to segment the LGE CMR data. Compared to 2D convolution and 3D convolution, the Pseudo-3D convolution neural network structure combining the advantage of 2D networks and preserving the spatial structure information in 3D data without compromising segmentation accuracy.

2   Related Work

Typically for MS CMR data, two-sequence CMR is widely used for automated myocardial segmentation. For example, Rajchl et al. [9] used the segmentation results of the bSSFP cine CMR as a priori knowledge, and then performed ventricular segmentation on the LGE CMR, which compensates for differences between slices of different sequences. In [13], a unified framework combining

---

three-sequence CMR (bSSFP, T2 and LGE) was proposed to align the MS CMR data from the same patient into a common space for segmentation.

Present, MS CMR data segmentation based on deep learning also has a good performance. In [4], a multi-task deep learning network for automatic 3D bi-ventricular segmentation of CMR was proposed, this network combines high-resolution and low-resolution CMR volume. However, it should be noted that this network requires additional landmark localization information, which undoubt-edly increased the requirements for data. Also, Tseng et al. [11] proposed a deep encoder-decoder structure with cross-modality convolution layers to incorporate different modalities of MRI data. However, this multi-modal encoder method does not apply to MS CMR data due to misalignment between image slices, the resolution is not uniform, the difference in slice thickness between the short-axis.
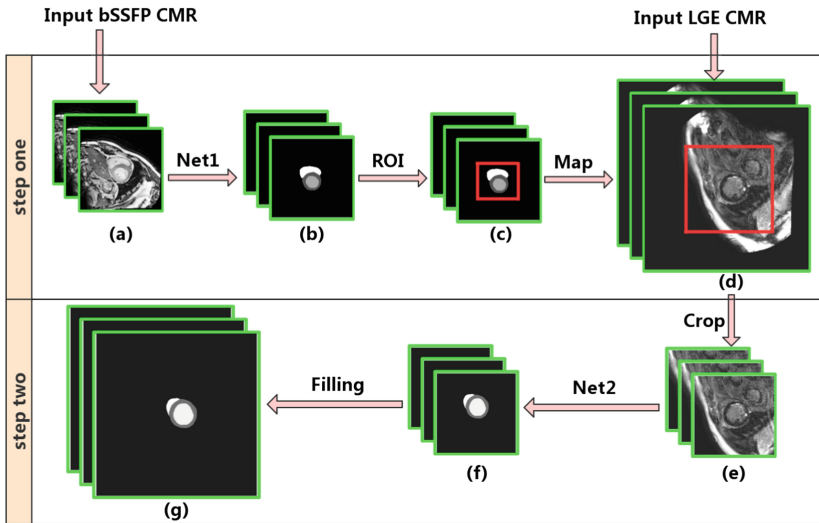

Fig. 1. Proposed pipeline for multi-sequence cardiac MR segmentation. (a) Input bSSFP CMR; (b) ventricular segmentation from bSSFP CMR; (c) the ROI (marked with a red square) obtained after positioning; (d) the ROI on the bSSFP CMR is mapped onto the input LGE CMR; (e) cropped LGE CMR; (f) segmentation result; (g) final output. Different sized rectangles represent different resolutions, smaller rect-angular represents low-resolution CMR volume, and conversely, larger rectangular rep-resents high-resolution CMR volume. (Color figure online)

For the segmentation of volumetric medical image data, a slice-by-slice learn-ing strategy is frequently used. This method processed the 3D volumetric medical image data into multiple 2D slices and then performed semantic segmentation on each 2D slice. However, simply connecting 2D segmentations into 3D will lose spatial correlation between the $z-$direction. A straightforward way to learn spatial structure information in volumetric medical image data is to extend the
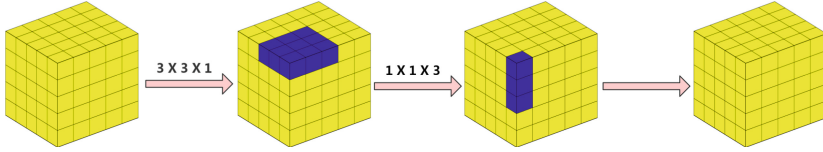
Fig. 2. The Pseudo-3D convolution.

2D convolution kernel to 3D convolution kernel, such as 3D U-Net [3] or V-Net [7]. Although 3D convolutional networks can learn more information, 3D convolutional networks require more computing resources (high memory consumption and more learning parameters) than 2D convolutional networks. Furthermore, volumetric medical image data are usually anisotropic [1]. For example, the Multi-sequence Cardiac MR Segmentation Challenge (MS-CMRSeg 2019[2]) data used in this work, the LGE CMR consisting of 10 to 18 slices, typically, the voxel scale in depth (the $z-$direction, 5 mm) is much larger than that in the xy plane (0.75 mm). To solve the above problems we employ the Pseudo-3D [8] convolution neural network structure to segment the LGE CMR data.

In [8], the Pseudo-3D network was first proposed and applied to learn spatio-temporal video representation. The Pseudo-3D convolution factorizes a standard $3 \times 3 \times 3$ convolution into two successive convolutional layers: $3 \times 3 \times 1$ convolutional filter to learn spatial features and $1 \times 1 \times 3$ convolutional filter to learn temporal features. This spatio-temporal separation network structure has been widely applied for video processing. Chen et al. [2] extended the Pseudo-3D network structure to the medical image field and segmented the small cell lung cancer, inspired by this, our study used this lightweight network structure to segment the ventricles in LGE CMR data.

3   Methods

Figure 1 illustrates the framework for multi-sequence cardiac MR segmentation, which can be roughly divided into two steps: (i) left ventricular positioning. First, the bSSFP CMR is taken as input, the left ventricle is obtained by segmentation network, and then the center position and radius of the left ventricle are obtained by Gaussian kernel-based circular Hough transform approach. Finally, the left ventricle position in the bSSFP CMR is mapped into the LGE CMR, and the region of interest (ROI) is obtained by the cropping operation. (ii) ventricle and myocardium segmentation. First, the ROI of the LGE CMR is taken as input, and the ventricular and myocardial segmentation results are obtained through a customized Pseudo-3D network structure. Finally, the filled image is used as the final output result.

---

[2] https://zmiclab.github.io/mscmrseg19/data.html.

3.1 Left Ventricular Positioning

In this work, we choose to use bSSFP CMR for left ventricular positioning for the following reasons: (i) compared with other modal CMRs, the bSSFP CMR captures cardiac motions and presents clear boundaries; (ii) compared to LGE CMR, the bSSFP CMR has more manual labels; (iii) each set of bSSFP CMR has more slices than the T2 CMR. Typically, the T2 CMR slice has a thickness of 20 mm, and a set of data usually consists of 3 to 5 slices, but the bSSFP CMR slice has a thickness of 8–13 mm and a set of data consists of 8 to 12 slices, more slices help the left ventricle to locate.
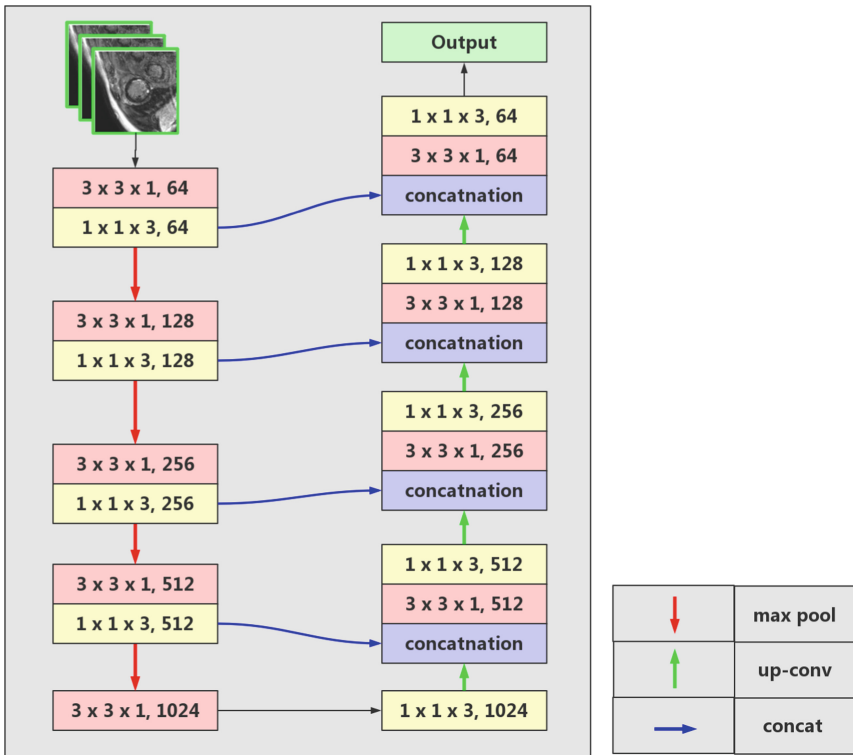

Fig. 3. An overview of the customized Pseudo-3D network framework.

First, the ventricle is segmented from the bSSFP CMR through a segmentation network. U-Net [10] only needs a small number of annotations to get better results, and is widely used in medical image segmentation, so U-Net is selected as our first segmentation network.

Next, the left ventricle is positioned on the segmented result to obtain a ROI. There are many interfering tissues around the ventricle. By locating the position of the left ventricle center point and extracting a ROI with a size of 256×256

centered around it, the interference tissue can be effectively reduced. At the same time, ROI operations can reduce computing resources and normalize data sizes. In this study, the left ventricular center point is extracted by Gaussian kernel-based circular Hough transform approach. The main idea of the algorithm we implemented is from [5], a little different from [5] is that this study calculates the left ventricular center point directly from the 3D data. In addition, it is difficult to calculate the left ventricular center point directly from the original data. Therefore, we first segment the original data and then perform ventricular positioning. Our left ventricle position method greatly improves the calculation accuracy of the center point.

Finally, the LGE CMR is used as the input, because each set of MS CMR is from the same location of the same patient, so the ROI of the bSSFP CMR can be mapped to the LGE CMR, and the original LGE CMR can be cropped according to the ROI. The cropped image is used as the input for the next stage (Sect. 3.2).

3.2   Ventricle and Myocardium Segmentation

Using the cropped LGE CMR as input, the customized Pseudo-3D network structure is used to obtain the segmentation result. Now, we explain the details of the customized Pseudo-3D network structure used in this study. The Pseudo-3D convolution, as shown in Fig. 2, splits one $3 \times 3 \times 3$ convolution into a $3 \times 3 \times 1$ convolution to learn intra-slice features and a $1 \times 1 \times 3$ convolution to learn inter-slice features. Such decoupled 3D convolutions not only reduce the model size significantly, but also address the problem of anisotropic dimensions.

Here, 3D U-Net [3] is used as a submodule of our customized Pseudo-3D network framework. As shown in Fig. 3, in this study, the original framework of 3D U-Net is preserved, and the 3D convolutional layer in the network structure is replaced by the Pseudo-3D structure (as shown in Fig. 2). This lightweight network structure is more suitable for CMR data of different heterogeneities.

4   Materials and Experiments

We validated the algorithm in this study on the MS CMRSeg 2019[3] (Multi-sequence Cardiac MR Segmentation Challenge). MS CMRSeg 2019 not only provides a multi-sequence ventricle and myocardium dataset with manual labels, but also provides an open and fair competitive platform to validate the ventricular segmentation algorithm. We implemented our framework using Keras with cuDNN, and ran all experiments on a personal computer with NVIDIA-GeForce-GTX-1080-Ti GPU, Intel Core i7–4790 CPU @ 3.60 GHz and 32 GB RAM.

The MS CMRSeg 2019 consisted of 45 patients with cardiomyopathy, and each set of patient data consists of three CMR sequences (the LGE, T2, and bSSFP), all of which were breath-hold, multi-slice, acquired in the ventricular

---

[3] https://zmiclab.github.io/mscmrseg19/.

short-axis views. In this study, our use of data is roughly divided into two steps. In the first step, 45 sets of bSSFP CMR were used for left ventricular positioning, 35 of which had manually labeled data as training sets, and the remaining 10 sets contained only image data as test sets. In the second step, 45 sets of LGE CMRs were used for fine segmentation, of which 5 sets of data containing tags were used as training sets, and the remaining 40 sets of unlabeled data were used as test sets. Finally, we send the segmentation results of the test set to the organizer of MS CMRSeg 2019. The test performance of the organizer feedback is shown in Table 1.

Table 1. Segmentation accuracy. Note: LV: Left ventricle; RV: Right ventricle; Myo: Myocardium.

Segmentation accuracy	LV blood cavity	Myo of LV	RV blood cavity
Dice	0.807 ($\pm$0.074)	0.617 ($\pm$0.086)	0.680 ($\pm$0.118)
Jaccard	0.683 ($\pm$0.103)	0.452 ($\pm$0.090)	0.526 ($\pm$0.129)
Surface distance (mm)	4.094 ($\pm$1.493)	5.687 ($\pm$1.794)	8.730 ($\pm$3.437)
Hausdorff distance (mm)	40.773 ($\pm$11.361)	43.351 ($\pm$7.186)	56.379 ($\pm$15.505)

From the Table 1 we can see that our method left ventricular Dice score is 0.807. Here, the Dice score, Jaccard, average surface distance and Hausdorff distance will be used as evaluation metrics, the Dice score and Jaccard can be computed as:

$$Dice(V_{manual}, V_{auto}) = \frac{2\,|V_{manual} \cap V_{auto}|}{|V_{manual}| + |V_{auto}|} \tag{1}$$

$$Jaccard(V_{manual}, V_{auto}) = \frac{|V_{manual} \cap V_{auto}|}{|V_{manual}| + |V_{auto}| - |V_{manual} \cap V_{auto}|} \tag{2}$$

where, V_{auto} is the segmented volume and the V_{manual} is the manual marker result. The scores of Dice and Jaccard represent the amount of overlap between the automatic segmentation results and the manually labeled results, which give a measurement value between 0 and 1. The average surface distance and the Hausdorff distance measure the distance between the automatic segmentation result and the manual marker result, and the smaller distance value represents a better segmentation result. It should be noted here that the LGE CMR data only provides 5 sets of training sets with labels, such a small amount of training data is also one of the challenges of this segmentation task.

5   Conclusions

This study detailed a simple but effective approach for automatic ventricle and myocardium segmentation from MS CMR, which uses the bSSFP CMR to perform left ventricular positioning, and use the LGE CMR to precise segmentation.

This segmentation method combines multiple sequences of CMR information. In addition, for the segmentation of LGE CMR, we used a customized Pseudo-3D convolution neural network, this framework not only reduces the size of the network, but also learns spatial structure information. In future work, we will continue to challenge the issue of multi-sequence CMR segmentation.

References

1. Chen, J., Yang, L., Zhang, Y., Alber, M., Chen, D.Z.: Combining fully convolutional and recurrent neural networks for 3D biomedical image segmentation. In: Advances in Neural Information Processing Systems, pp. 3036–3044 (2016)
2. Chen, W., Wei, H., Peng, S., Sun, J., Qiao, X., Liu, B.: HSN: hybrid segmentation network for small cell lung cancer segmentation. IEEE Access **7**, 75591–75603 (2019)
3. 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
4. Duan, J., et al.: Automatic 3D bi-ventricular segmentation of cardiac images by a shape-constrained multi-task deep learning approach. arXiv preprint (2018)
5. Khened, M., Kollerathu, V.A., Krishnamurthi, G.: Fully convolutional multi-scale residual densenets for cardiac segmentation and automated cardiac diagnosis using ensemble of classifiers. Med. Image Anal. **51**, 21–45 (2019)
6. Liu, T., Tian, Y., Zhao, S., Huang, X., Wang, Q.: Automatic whole heart segmentation using a two-stage u-net framework and an adaptive threshold window. IEEE Access **7**, 83628–83636 (2019). https://doi.org/10.1109/ACCESS.2019.2923318
7. 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), pp. 565–571. IEEE (2016)
8. Qiu, Z., Yao, T., Mei, T.: Learning spatio-temporal representation with pseudo-3D residual networks. In: proceedings of the IEEE International Conference on Computer Vision, pp. 5533–5541 (2017)
9. Rajchl, M., et al.: Interactive hierarchical-flow segmentation of scar tissue from late-enhancement cardiac MR images. IEEE Trans. Med. Imaging **33**(1), 159–172 (2013)
10. 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
11. Tseng, K.L., Lin, Y.L., Hsu, W., Huang, C.Y.: Joint sequence learning and cross-modality convolution for 3D biomedical segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 6393–6400 (2017)

12. Zhuang, X.: Multivariate mixture model for cardiac segmentation from multi-sequence MRI. In: Ourselin, S., Joskowicz, L., Sabuncu, M.R., Unal, G., Wells, W. (eds.) MICCAI 2016. LNCS, vol. 9901, pp. 581–588. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-46723-8_67
13. Zhuang, X.: Multivariate mixture model for myocardial segmentation combining multi-source images. In: Proceedings of IEEE Transactions on Pattern Analysis and Machine Intelligence (2018)