Segmentation and Quantification of Bi-Ventricles and Myocardium Using 3D SERes-U-Net

Marija Habijan[1(✉)] , Irena Galić[1] , Hrvoje Leventić[1] , Krešimir Romić[1] , and Danilo Babin[2]

[1] Faculty of Electrical Engineering, Computer Science and Information Technology Osijek, Josip Juraj Strossmayer University of Osijek, Osijek, Croatia
marija.habijan@ferit.hr
[2] TELIN-IPI, Faculty of Engineering and Architecture, Ghent University – imec, Ghent, Belgium

Abstract. Automatic cardiac MRI segmentation, including left and right ventricular endocardium and epicardium, has an essential role in clinical diagnosis by providing crucial information about cardiac function. Determining heart chamber properties, such as volume or ejection fraction, directly relies on their accurate segmentation. In this work, we propose a new automatic method for the segmentation of myocardium, left, and right ventricles from MRI images. We introduce a new architecture that incorporates SERes blocks into 3D U-net architecture (3D SERes-U-Net). The SERes blocks incorporate squeeze-and-excitation operations into residual learning. The adaptive feature recalibration ability of squeeze-and-excitation operations boosts the network's representational power while feature reuse utilizes effective learning of the features, which improves segmentation performance. We evaluate the proposed method on the testing dataset of the MICCAI Automated Cardiac Diagnosis Challenge (ACDC) dataset and obtain highly comparable results to the state-of-the-art methods.

Keywords: Cardiac MRI segmentation · Left ventricle · Right ventricle · Myocardium · Residual learning · Squeeze and excitation · 3D SERes-U-Net

1 Introduction

Cardiovascular diseases (CVDs) cause major health complications that often lead to death [19]. An evaluation of cardiac function and morphology plays an essential role for CVDs' early diagnosis, risk evaluation, prognosis setting, and therapy decisions. Magnetic resonance imaging (MRI) has a high resolution, contrast and great capacity for differentiating between types of tissues. This makes MRI the gold standard of cardiac function analysis [2]. Delineations of the myocardium (Myo), left ventricle (LV), and right ventricle (RV) are necessary for quantitative

assessment and calculation of clinical indicators such as volumetric measures at end-systole and at end-diastole, ejection fraction, thickening measures, as well as mass. Semi-automatic delineation is still commonly present in clinical practice. That is often a laborious, and time-consuming process, prone to intra- and inter-observer variability. Hence, accurate, reliable, and automated segmentation methods are required to facilitate cardiovascular disease diagnosis.

Various image processing methods have been proposed to automatize segmentation tasks in the medical field [4,10,21]. While some of these approaches use more traditional techniques like level sets [17], registration and atlases [5,8], fully-automatic methods mostly employ fully convolutional neural networks (FCNNs) [6]. Commonly used approaches include structures that consist of a series of convolutional, pooling, and deconvolutional layers such as U-Net architecture [7,22]. Generally, various deep learning methods have shown outstanding performance on medical images for segmentation purposes [3,9,13–16,20,23,24]. Promising as they are, the appearance of overfitting on limited training data, vanishing and exploding gradients, and network degradation are significant concerns for FCNs. The residual learning, introduced in ResNets [11], overcomes the above problems by enhancing information flow over through the network using identity shortcut connections. Squeeze and excitation operations, introduced in SeNets, [12] improve the network's representational power by modeling interdependencies of channel-wise features and by dynamically recalibrating them.

Motivated by previously described advancements, we propose a 3D U-Net-based network that incorporates residual and squeeze and excitation blocks (SERes blocks). We introduce the squeeze and excitation (SE) blocks at 3D U-Nets' encoder and decoder paths after each residual block. We provide experimental results of the proposed network for the task of LV, RV, and Myo segmentation and show that our proposed approach obtains highly comparable results to the state-of-the-art.

2   Method

2.1   Squeeze and Excitation Residual Block

The SERes block takes the advantages of the squeeze and excitation operations [12] for adaptive feature recalibration and residual learning for feature reuse [11].

The 3D SERes block can be expressed with the following expression:

$$\mathbf{X}^{res} = F_{res}(\mathbf{X}) \tag{1}$$

where $\mathbf{X}$ refers to the input feature, $\mathbf{X}^{res}$ is the residual feature, and $F_{res}(\mathbf{X})$ is residual mapping that needs to be learned. The squeeze function which groups channel-wise statistics and global spatial information using global average pooling can be expressed with:

$$F_{sq}(\mathbf{x}_n^{res}) = p_n = \frac{1}{L \times H \times W} \sum_{i=1}^{L} \sum_{j=1}^{H} \sum_{k=1}^{W} x_n^{res}(i,j,k) \tag{2}$$

where $\mathbf{p} = [p_1, p_2, ..., p_n]$ and p_n is the $n-th$ element of $\mathbf{p} \in R^n$, where, $L \times H \times W$ is the spatial dimension of $\mathbf{F}^{res}$, $x_n^{res} \in R^{L \times H \times W}$ represents the feature map of the $n - th$ channel from the feature $\mathbf{X}^{res}$, and N referst to the residual mapping channels'. Scale values for the residual feature channels $\mathbf{s} \in R^N$ can be expressed with:

$$\mathbf{s} = F_{ex}(\mathbf{p}, \mathbf{W}) = \sigma(\mathbf{W}_2 \delta(\mathbf{W}_1 \mathbf{p})) \tag{3}$$

where F_{ex} is the excitation function which generates them. It is parameterized by two fully connected layers (FCNs) with parameters $\mathbf{W}_1 \in R^{\frac{N}{r} \times N}$ and $\mathbf{W}_2 \in R^{N \times \frac{N}{r}}$, the sigmoid function σ, the $ReLU$ function δ and has reduction ration determened with r. The multiplication between feature map and learned scale value s_n across channel can be expressed with:

$$\widetilde{\mathbf{X}}_n^{res} = F_{scale}(\mathbf{X}_n^{res}, s_n) = s_n \cdot \mathbf{X}_n^{res}, \in R^{H \times W \times L} \tag{4}$$

Finally, applying the squeeze and excitation operations obtains the calibrated residual feature, which can be expressed with:

$$\widetilde{\mathbf{X}}^{res} = [\widetilde{\mathbf{X}}_1^{res}, \widetilde{\mathbf{X}}_2^{res}, ..., \widetilde{\mathbf{X}}_n^{res}] \tag{5}$$

The output feature $\mathbf{Y}$ after the $ReLU$ function δ is obtained as:

$$\mathbf{Y} = \delta(\widetilde{\mathbf{X}}^{res} + \mathbf{X}) \tag{6}$$

where $(\widetilde{\mathbf{X}}^{res} + \mathbf{X})$ refers to element-wise addition and a shortcut connection.

An illustration of the 3D ResNet block and 3D SERes block is shown in Fig. 1.

2.2   3D SERes-U-Net Architecture

Our proposed network architecture is based on the standard 3D U-Net [7] which follows encoder-decoder architecture. The encoder or contracting pathway encodes the input image and learns low-level features, while the decoder or expanding pathway learns high-level features and gradually recovers original image resolution.

Like 3D U-Net, our contracting pathway consist of three downsampling layers. We replace initially used pooling layers in the original 3D U-Net with convolutional layers with stride equal to 2. Instead of plain units, we adopt SERes blocks consisting of squeeze and excitation operations followed by a residual block, as described in 2.1, to accelerate convergence and training. Each residual block inside the SERes block has two convolutional layers that are followed by ReLU activation, and batch normalization (BN) as shown in Fig. 1(b). Similarly, three SERes blocks are used in the expanding pathway. This pathway has three up-sampling layers, each of which doubles the size of the feature maps, and are followed by a $2 \times 2 \times 2$ convolutional layer. The network can acquire the importance degree of each residual feature channel through the feature recalibration strategy. Based on the importance degree, the less useful channel features are

(a) 3D residual block   (b) 3D SERes block
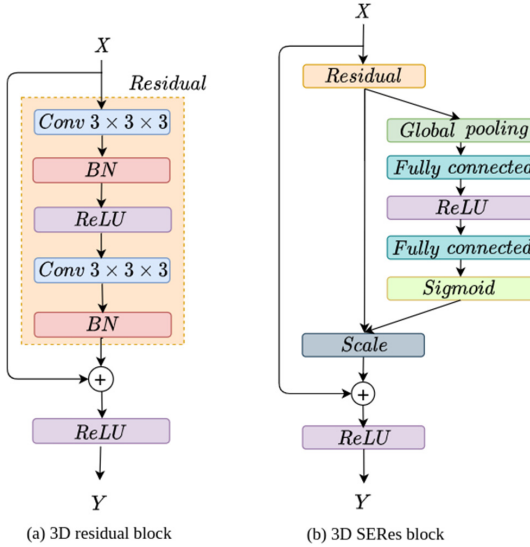
Fig. 1. An illustration of used residual blocks. (a) The original 3D ResNet block and (b) structure of the 3D SERes block

suppresed while useful features are enhanced. Therefore, by modeling the inter-dependencies between channels, the 3D SERes block performs dynamic recali-bration of residual feature responses in a channel-wise manner. In this way, the network can capture every residual feature channel's importance degree, which improves its representational power. SERes-U-Net architecture is presented in Fig. 2.

3   Implementation Details

3.1   Dataset and Evaluation Metrics

The Automated Cardiac Diagnosis Challenge (ACDC) dataset consists of real-life clinical cases obtained from an everyday clinical setting at the University Hospital of Dijon (France). The dataset includes cine-MRI images of patients suf-fering from different pathologies, including myocardial infarction, hypertrophic cardiomyopathy, dilated cardiomyopathy, abnormal right ventricle, and normal cardiac anatomy. Dataset has been evenly divided based on the pathological condition and includes 100 cases with corresponding ground truth for training, and 50 cases for testing through an online evaluation platform. Clinical experts manually annotated LV, RV, and Myo at systolic and diastolic phases, for which the weight and height information was provided as well. Images are acquired as a series of short-axis slices covering the LV from the base to the apex. The spatial resolution goes from 1.37 to 1.68 mm^2/pixel, slice thickness is between 5–8 mm, while an inter-slice gap is 5 or 10 mm.
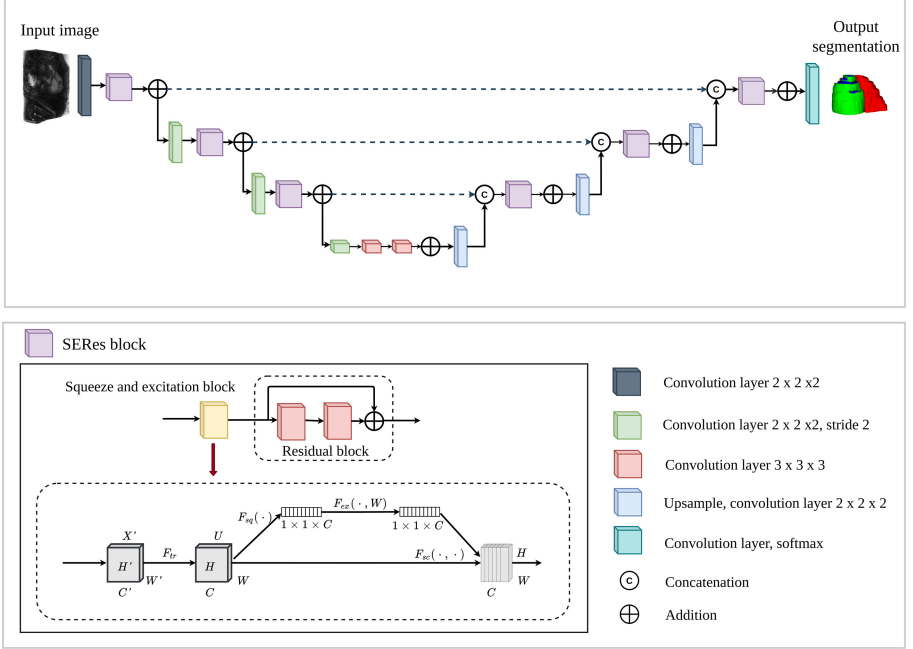
Fig. 2. Illustration of SERes-U-Net architecture for LV, RV, Myo segmentation.

3.2   Network Training

To overcome high intensity irregularities of MRI images, we normalize each volume based on the standard and mean deviation of their intensity values. The volumes were center-cropped to a fixed-size and zero-padded to provide fine ROI for the network input. For data augmentation, we apply random axis mirror flip with a probability of 0.5, random scale, and intensity shift on input image channel. We use $L2$ norm regularization with a weight of 10^{-5} and employ the spatial dropout with a rate of 0.2 after the initial encoder convolution. We use Adam optimizer with initial learning rate of $\alpha_0 = 10^{-4}$ and gradually decrease it according to following expression:

$$\alpha = \alpha_0 * \left(1 - \frac{e}{T_e}\right)^{0.9} \tag{7}$$

where T_e is a total number of epochs and e is an epoch counter. We employ a smoothed negative Dice score [18] loss function, defined with:

$$D_{loss} = -\frac{2\sum_{i=1}^{N} p_i g_i + 1}{\sum_{i=1}^{N} p_i + \sum_{i=1}^{N} g_i + 1} \tag{8}$$

where p_i is probability of predicted regions, g_i is the ground truth classification for every i voxel.
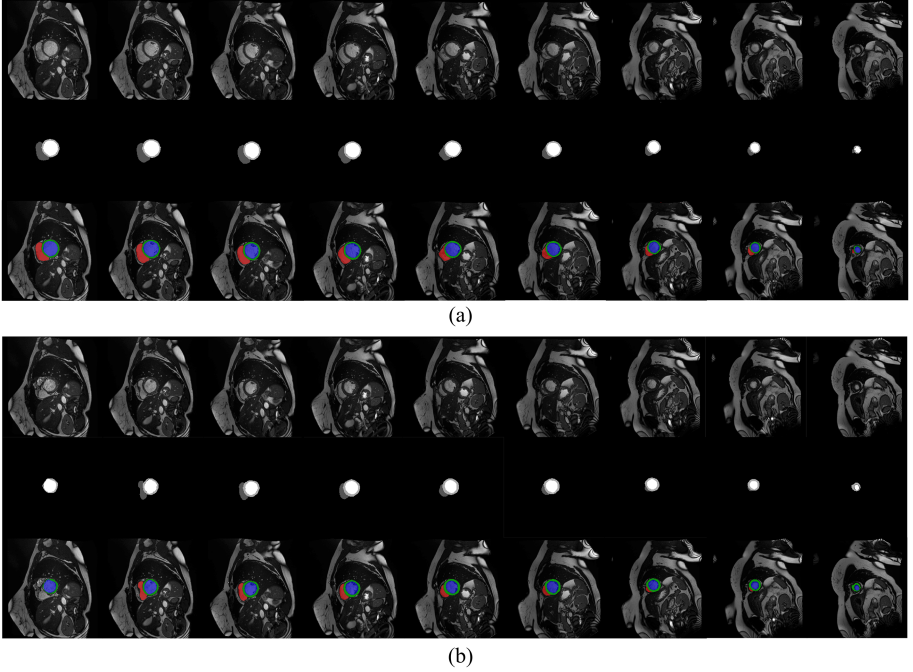
We use 80%-20% training and validation split, respectively. Final segmentation accuracy testing was done on an online ACDC Challenge submission page on 50 patient subjects [1]. The total training time took approximately 34 h for 400 epochs using a two NVIDIA Titan V GPU, simultaneously.

4   Results

To evaluate the segmentation performance of the proposed method, we observe distance and clinical indices metrics. Distance measures include calculation of Dice score (DSC) and Hausdorff distance (HD) which provides information of similarity between obtained segmentations for LV, RV, and Myo with their reference ground truth. The 3D Res-U-Net network achieves an average DSC for LV, RV and Myo at end diastole of 93%, 86, 80 respectively. The addition of squeeze and excitation operations, i.e., use of proposed SERes blocks, improves DSC and HD for 2%, 4% and 3%, respectively. Similarly, the 3D Res-U-Net network achieves an average DSC for LV, RV and Myo at end systole of 86%, 77, 81 respectively. The addition of squeeze and excitation operations, i.e., use of proposed SERes blocks, improves DSC and HD for 0.2%, 6% and 4%, respectively. Therefore, obtained results using proposed 3D SERes-U-Net shows significant improvements in DSC in comparison to network without squeeze and excitation operations (3D Res-U-Net). Detailed qualitative results are shown in Table 1 and Table 2 while Fig. 3 provides visual example of obtained segmentation predictions. Clinical metrics include calculation of the most widely used indicators of hearts' function, including volume of the left ventricle at end-diastole (LVEDV), volume of the left ventricle at end-systole (LVESV), left ventricles' ejection fraction (LVEF), volume of the right ventricle at end-diastole (RVEDV), volume of the right ventricle at end-systole (RVESV), right ventricles' ejection fraction (RVEF), myocardium volume at end-systole (MyoLVES), and myocardium mass at end-diastole (MyoMED).

Table 1. The segmentation accuracy results for LV, RV and Myo expressed in Dice score (DSC) and Hausdorff distance (HD) for the proposed method at end diastole for 3D Res-U-Net and proposed 3D SERes-U-Net.

| Methods	ED					
	LV		RV		Myo	
	D_{sc}	H_d	D_{sc}	H_d	D_{sc}	H_d
3D Res-U-Net	0.93	38.2	0.86	52.9	0.8	32.95
	(0.0636)	(4.8721)	(0.0919)	(12.4414)	(0.0636)	(5.6003)
3D SERes-U-Net	0.95	11.53	0.9	23.41	0.83	13.77
	(0.0071)	(0.4101)	(0.0212)	(12.3571)	(0.0071)	(1.9871)

(a)

(b)

Fig. 3. An example of obtained results. (a) Top row: original MRI image at end diastolic phase of cardiac cycle. Middle row: Obtained segmentation. Bottom row: an overlay of original image and obtained segmentation prediction. (b) Top row: original MRI image at end systolic phase of cardiac cycle. Middle row: Obtained segmentation. Bottom row: an overlay of original image and obtained segmentation prediction.

Table 2. The segmentation accuracy results for LV, RV and Myo expressed in Dice score (DSC) and Hausdorff distance (HD) for the proposed method at end systole for 3D Res-U-Net and proposed 3D SERes-U-Net.

Methods	ES					
	LV		RV		Myo	
	D_{sc}	H_d	D_{sc}	H_d	D_{sc}	H_d
3D Res-U-Net	0.86	29.77	0.77	36.99	0.81	30.29
	(0.0283)	(1.7748)	(0.0424)	(5.3952)	(0.0283)	(1.1031)
3D SERes-U-Net	0.86	11.94	0.83	21.49	0.85	15.00
	(0.1273)	(8.4994)	(0.0283)	(5.7558)	(0.0071)	(1.9799)

The Pearson correlation coefficient (R) and Bland-Altman and analysis of the results obtained using proposed methed for LV, RV and Myo are shown in Figs. 5, 6, 4.
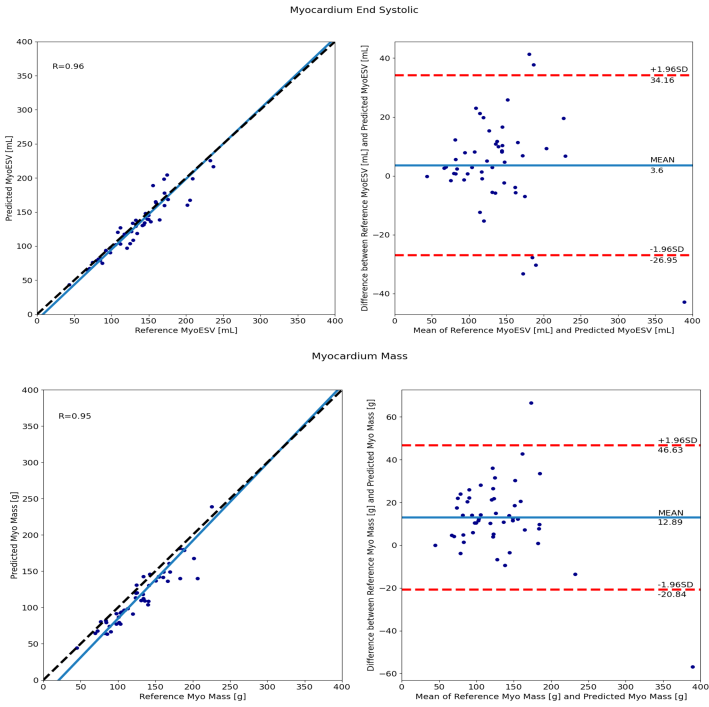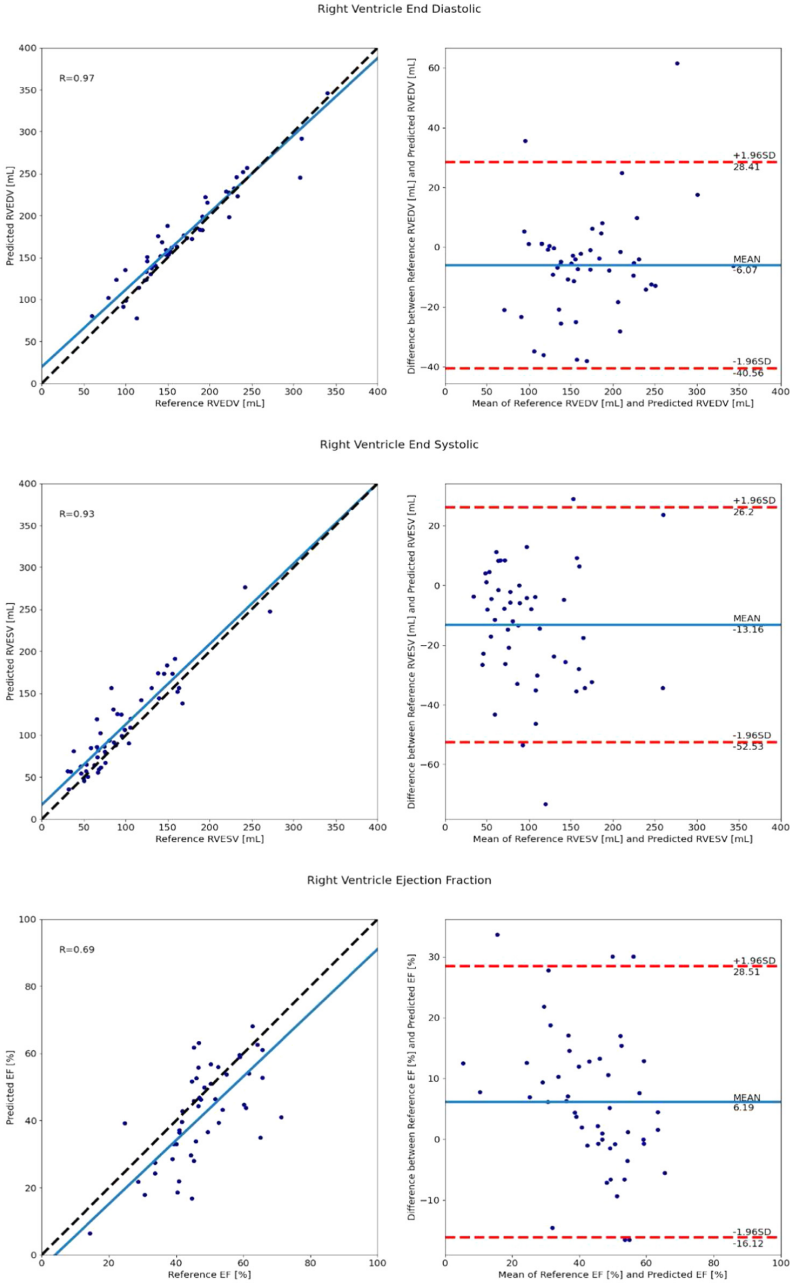

Fig. 4. Comparison of the automatically obtained segmentations and the reference volume of the myocardium end systolic volume and myocardium mass. The image showns correlation and Bland-Altman plots to compare automatically obtained segmentation and the reference values.
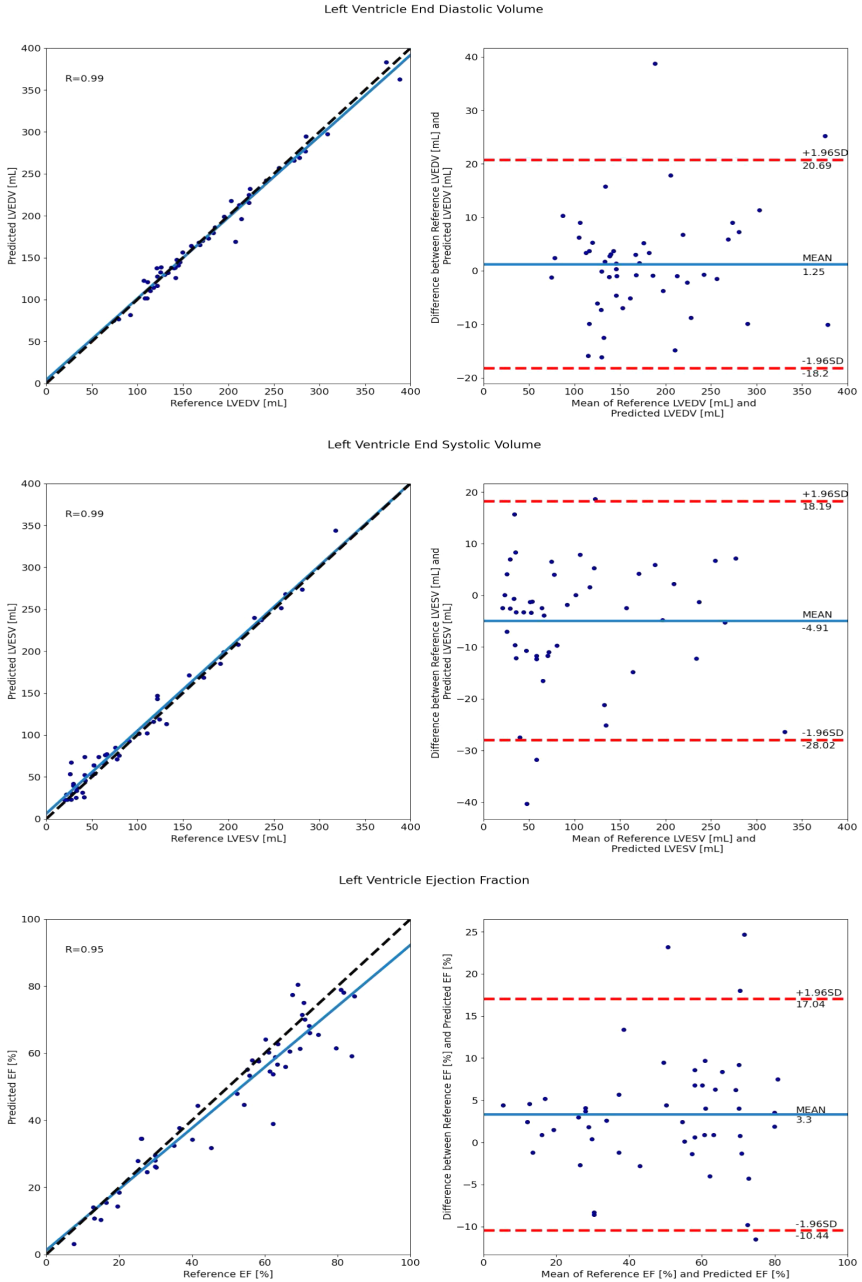
Fig. 5. Comparison of the automatically obtained segmentations and the reference volumes of the MRI scans. The image shows correlation and Bland-Altman plots for the LV volumes at and diastole and at the end systole as well as ejection fraction.

Fig. 6. Comparison of the automatically obtained segmentations and the reference volumes of the MRI scans. The image showns correlation and Bland-Altman plots for the RV volumes at and diastole and at the end systole as well as ejection fraction.

5   Conclusion

In this work, a deep neural network architecture named 3D SERes-U-Net was introduced for automatic segmentation of LV, RV, and Myo from MRI images. The significance of the proposed approach is in the two main characteristics. First, the approach is based on 3D deep neural networks, which are suitable for volumetric medical image processing. Second, the network introduces SERes blocks which optimizes the deep network and extracts distinct features. By taking advantage of the 3DSERes block, the proposed method learns the features with high discrimination capability, which is favorable to identify cardiac structures from the complex environment.

References

1. A.C.D.C.A.M.C.: Post-2017-miccai-challenge testing phase. https://acdc.creatis.insa-lyon.fr/#challenges (2017)
2. Arnold, J.R., McCann, G.P.: Cardiovascular magnetic resonance: applications and practical considerations for the general cardiologist. Heart **106**(3), 174–181 (2020)
3. Baumgartner, C.F., Koch, L., Pollefeys, M., Konukoglu, E.: An exploration of 2d and 3d deep learning techniques for cardiac MR image segmentation. ArXiv arXiv:1709.04496 (2017)
4. Bernard, O., et al.: Deep learning techniques for automatic MRI cardiac multi-structures segmentation and diagnosis: is the problem solved? IEEE Trans. Med. Imaging **37**(11), 2514–2525 (2018). https://doi.org/10.1109/TMI.2018.2837502
5. Cetin, I., et al.: A radiomics approach to computer-aided diagnosis with cardiac cine-MRI. In: Pop, M., et al. (eds.) STACOM 2017. LNCS, vol. 10663, pp. 82–90. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-75541-0_9
6. Cheng, F., et al.: Learning directional feature maps for cardiac MRI segmentation. In: Martel, A.L., et al. (eds.) MICCAI 2020. LNCS, vol. 12264, pp. 108–117. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-59719-1_11
7. Click, O., Abdulkadir, A., Lienkamp, S.S., Brox, T., Ronneberger, O.: 3d u-net: Learning dense volumetric segmentation from sparse annotation. CoRR abs/1606.06650 (2016). http://arxiv.org/abs/1606.06650
8. Duan, J., et al.: Automatic 3d bi-ventricular segmentation of cardiac images by a shape-refined multi- task deep learning approach. IEEE Trans. Med. Imaging **38**(9), 2151–2164 (2019)
9. Habijan, M., Leventić, H., Galić, I., Babin, D.: Estimation of the left ventricle volume using semantic segmentation. In: 2019 International Symposium ELMAR, pp. 39–44 (2019). https://doi.org/10.1109/ELMAR.2019.8918851
10. Habijan, M., et al.: Overview of the whole heart and heart chamber segmentation methods. Cardiovasc. Eng. Technol. **11**(6), 725–747 (2020)
11. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 770–778 (2016). https://doi.org/10.1109/CVPR.2016.90

12. Hu, J., Shen, L., Sun, G.: Squeeze-and-excitation networks. In: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 7132–7141 (2018). https://doi.org/10.1109/CVPR.2018.00745

13. Isensee, F., Jaeger, P., Full, P.M., Wolf, I., Engelhardt, S., Maier-Hein, K.H.: Automatic cardiac disease assessment on cine-MRI via time-series segmentation and domain specific features. CoRR abs/1707.00587 (2017). http://arxiv.org/abs/1707.00587

14. Jang, Y., Hong, Y., Ha, S., Kim, S., Chang, H.-J.: Automatic segmentation of LV and RV in cardiac MRI. In: Pop, M., et al. (eds.) STACOM 2017. LNCS, vol. 10663, pp. 161–169. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-75541-0_17

15. Khened, M., Varghese, A., Krishnamurthi, G.: Densely connected fully convolutional network for short-axis cardiac cine MR image segmentation and heart diagnosis using random forest. In: STACOM@MICCAI (2017)

16. Liu, T., Tian, Y., Zhao, S., Huang, X., Wang, Q.: Residual convolutional neural network for cardiac image segmentation and heart disease diagnosis. IEEE Access **8**, 82153–82161 (2020). https://doi.org/10.1109/ACCESS.2020.2991424

17. Liu, Y., et al.: Distance regularized two level sets for segmentation of left and right ventricles from cine-MRI. Magn. Reson. Imaging **34** (2015). https://doi.org/10.1016/j.mri.2015.12.027

18. Lu, J.-T., et al.: DeepAAA: clinically applicable and generalizable detection of abdominal aortic aneurysm using deep learning. In: Shen, D., et al. (eds.) MICCAI 2019. LNCS, vol. 11765, pp. 723–731. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-32245-8_80

19. Organization, W.H.: Mortality database (2018). Accessed 19 Jan 2021

20. Patravali, J., Jain, S., Chilamkurthy, S.: 2d–3d fully convolutional neural networks for cardiac MR segmentation. ArXiv arXiv:1707.09813 (2017)

21. 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. Phys. Biol. Med. **29**(2), 155–195 (2016). https://doi.org/10.1007/s10334-015-0521-4

22. Ronneberger, O., Fischer, P., Brox, T.: U-net: convolutional networks for biomedical image segmentation. CoRR abs/1505.04597 (2015). http://arxiv.org/abs/1505.04597

23. Vigneault, D.M., Xie, W., Ho, C.Y., Bluemke, D.A., Noble, J.A.: Omega-net (omega-net): fully automatic, multi-view cardiac MR detection, orientation, and segmentation with deep neural networks. Med. Image Anal. **48**, 95–106 (2018). https://doi.org/10.1016/j.media.2018.05.008

24. Zotti, C., Luo, Z., Humbert, O., Lalande, A., Jodoin, P.M.: Gridnet with automatic shape prior registration for automatic MRI cardiac segmentation. In: STACOM@MICCAI (2017)