Biomedical Physics & Engineering Express

ENDORSED BY IPEM Institute of Physics and Engineering in Medicine

PAPER

Automated segmentation of the left ventricle from MR cine imaging based on deep learning architecture

Biomedical Physics & Engineering Express

PAPER

Automated segmentation of the left ventricle from MR cine imaging based on deep learning architecture

RECEIVED
3 September 2019

REVISED
6 January 2020

ACCEPTED FOR PUBLICATION
6 February 2020

PUBLISHED
18 February 2020

Wenjian Qin[1,3], **Yin Wu**[1,3], **Siyue Li**[1], **Yucheng Chen**[2], **Yongfeng Yang**[1], **Xin Liu**[1], **Hairong Zheng**[1], **Dong Liang**[1] and **Zhanli Hu**[1]

[1] Lauterbur Research Center for Biomedical Imaging, Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518055, People's Republic of China
[2] Cardiology Division, West China Hospital, Sichuan University, Chengdu 610041, People's Republic of China
[3] W Qin and Y Wu contributed equally to this paper.

E-mail: zl.hu@siat.ac.cn

Keywords: Image segmentation, left ventricular segmentation, artificial neural networks, magnetic resonance cine imaging, deep learning.

Abstract

Background: Magnetic resonance cine imaging is the accepted standard for cardiac functional assessment. Left ventricular (LV) segmentation plays a key role in volumetric functional quantification of the heart. Conventional manual analysis is time-consuming and observer-dependent. Automated segmentation approaches are needed to improve the clinical workflow of cardiac functional quantification. Recently, deep-learning networks have shown promise for efficient LV segmentation. PURPOSE: The routinely used V-Net is a convolutional network that segments images by passing features from encoder to decoder. In this study, this method was advanced as DenseV-Net by replacing the convolutional block with a densely connected algorithm and dense calculations to alleviate the vanishing-gradient problem, prevent exploding gradients, and to strengthen feature propagation. Thirty patients were scanned with a 3 Tesla MR imager. ECG-free, free-breathing, real-time cines were acquired with a balanced steady-state free precession technique. Linear regression and the dice similarity coefficient (DSC) were performed to evaluate LV segmentation performance of the classic neural networks FCN, UNet, V-Net, and the proposed DenseV-net methods, using manual analysis as the reference. Slice-based LV function was compared among the four methods. *Results*: Thirty slices from eleven patients were randomly selected (each slice contained 73 images), and the LVs were segmented using manual analysis, UNet, FCN, V-Net, and the proposed DenseV-Net methods. A strong correlation of the left ventricular areas was observed between the proposed DenseV-Net network and manual segmentation ($R = 0.92$), with a mean DSC of 0.90 ± 0.12. A weaker correlation was found between the routine V-Net, UNet, FCN, and manual segmentation methods ($R = 0.77$, 0.74, 0.76, respectively) with a lower mean DSC (0.85 ± 0.13, 0.84 ± 0.16, 0.79 ± 0.17, respectively). Additionally, the proposed DenseV-Net method was strongly correlated with the manual analysis in slice-based LV function quantification compared with the state-of-art neural network methods V-Net, UNet, and FCN. *Conclusion*: The proposed DenseV-Net method outperforms the classic convolutional networks V-Net, UNet, and FCN in automated LV segmentation, providing a novel way for efficient heart functional quantification and the diagnosis of cardiac diseases using cine MRI.

Background

MR cine imaging is accepted as the standard for cardiac functional assessment in clinical practice [1, 2]. Conventional cine imaging is acquired with electrocardiograph (ECG) gating and breath-holding to reduce artifact from cardiac and respiratory motion because it is challenging for patients who cannot hold their breath or for those with arrythmias. ECG-free, free-breathing, real-time cine imaging has been proposed, and this greatly improves the robustness of cine imaging and extends the diagnostic capability of cardiac MRI [3, 4]. Compared to cine MRI, real-time ECG-free cine imaging has a shorter image acquisition

time for the same high-quality image, and it has been proven to detect diaphragmatic movement. However, image analysis usually relies on manual interaction to date, making cardiac functional quantification time-consuming. An efficient, preferably automated segmentation method is needed to improve the workflow of real-time cine analysis.

To achieve the segmentation of medical images, several studies have focused on different trials. The current methods used to segment anatomical medical images can be broadly divided into two categories: semi-automated and automated methods [5]. For the semi-automated category, thresholding is the simplest method. Ibrahim *et al* [6, 7] proposed the application known as salient isolated thresholding with ant colony optimization to achieve segmentation of the blood myocardium. Yang *et al* [8] developed a new automatic segmentation method based on template matching and improved region growing that requires manual feeding of a seed point. However, this method leaves holes in the region of interest. Bardinet *et al* [9] proposed a deformable model based on a superquadratic fit to process the cardiac images. Although these semi-automatic methods perform well in medical imaging applications, they require the initialization to be completed through human intervention. Differing levels of initialization for human variability may lead to different segmentation results. With the increasing development of deep learning, artificial neural networks have been proposed for use in medical image processing. The initial experiments were based on convolutional networks, particularly in the field of segmentation. Ciresan *et al* [10] proposed the use of deep neural networks to automatically segment the neuronal membranes. However, this method involves redundancy in the parameters, and the receptive field exhibits a reverse proportional relationship with the location accuracy.

With more data becoming available and recent advances in machine learning and parallel computing infrastructure, automatic segmentation algorithms for the left ventricle based on various 2D and 3D convolutional neural network architectures have been proposed [11]. Zotti *et al* proposed a fully automatic MRI cardiac segmentation method based on a novel deep convolutional neural network [12]. The fully convolutional neural network (FCN), involving pixel-to-pixel training, is widely applied in segmentation tasks because of its great accuracy. Based on the FCN, U-Net was developed with fewer training images to realize medical image segmentation. The summed convolutional part and feature channels allow the sufficient transfer of background information to higher-resolution layers. U-Net is designed without fully connected layers to ensure that the segmentation results are obtained with no missing context features [13, 14] .

In this work, we explored automatic segmentation of the left ventricle using a densely connected, fully convolutional network, DenseV-Net, which was developed to process 2D medical images based on V-Net, with the images being labeled by experienced physicians. The most important aspect of DenseV-Net is that it improves spatial consistency by fusing information from different slices. V-Net has performed well in prostate segmentation and in liver segmentation [15]. A dense function is proposed in reference [16], providing a pathway for original feature information to be maintained as compressed features after the convolutional layer has been obtained. The dense function enhances the flow of features and reduces the gradient-vanishing and gradient-exploding phenomenon. The dice coefficient is used to evaluate the performance of the proposed network. Our results show that the volume of the left ventricle predicted by the proposed networks is close to the true volume of the left ventricle. This finding shows that our network can release clinicians from the time-consuming work of segmenting the left ventricle. Moreover, the entire analysis procedure is fully automated, without the need for any human effort, and takes less than one second to produce an image.

Methods

MR cine imaging

Thirty cardiac patients were scanned on a 3 Tesla (T) scanner (MAGNETOM TIM Trio, Siemens, Erlangen, Germany). Short-axis planes were localized from scout imaging, and ECG-free, free-breathing, real-time cine imaging was conducted slice by slice. The real-time cines were acquired using balanced steady-state, free precession imaging with the following parameters: repetition time (TR) = 2.5 ms, echo time (TE) = 1.1 ms, matrix size = 176 × 96, fixed field of view (FOV) = 360 × 270 mm^2, phase resolution = 75% (which yielded a true in-plane resolution of 2.1 × 2.8 mm^2), and slice thickness = 10 mm, with a 2-mm slice gap for all the subjects. Parallel imaging with an acceleration rate of 4 using temporal generalized autocalibrating partially parallel acquisition (TGRAPPA) was performed. Images were reconstructed with standard Fourier transform of k-space data. The temporal resolution achieved was 61 ms, which was adequate for distinguishing different cardiac phases [17] without reliance on ECG gating and breathholds. The slice number was adapted to the heart size so that the whole LV was covered. Each slice was imaged for approximate 4.5 s to include at least one complete respiratory cycle, resulting in 73 images. A Karhunen–Loeve transform filter was applied in the temporal direction to improve the signal-to-noise ratio (SNR) of the real-time cines. This prospective study was approved by the local Institutional Ethics Committee.

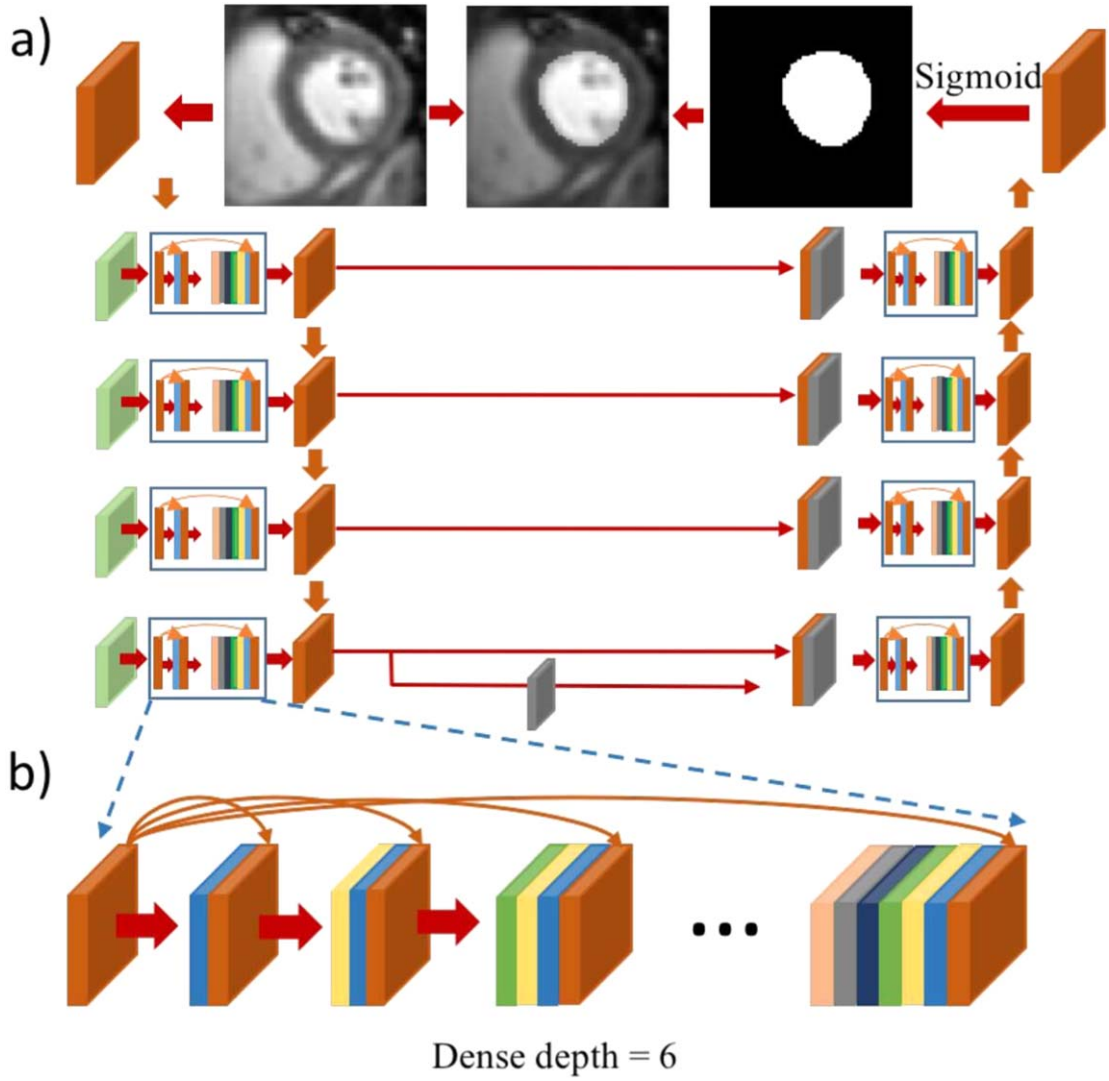

Figure 1. (a) Schematic representation of our proposed network architecture. Our custom implementation of Tensorflow processing of 2D images cut out by 3D MR cardiac images. (b) Schematic representation of the dense building block. Each layer is followed by the Elu active function. The output layer concatenates the original input.

Data analysis

The MR data of patients was collected, and MR sections containing left ventricular layers were selected for datasets. The left ventricle of the images was labeled manually through ITK-SNAP by an experienced radiologist (having more than 10 years of experience) and two researchers (each with more than 5 years of experience) to generate the ground truth mask using majority voting.

We adopted the area of the left ventricle labeled by the observers as ground truth. In total, we collected 86 MR slices of thirty patients (each slice contained 73 images), of which 46 were used as a training set, ten as a validation set, and the remainder as a test set. The size of the scanned images was 132 $\times$ 176. To better train the network, we rescaled all the images to the same size, 128 $\times$ 128, and they contained the information of interest, thus reducing data redundancy.

Network architecture

Figure 1(a)) shows the schematic representation of our proposed DenseV-Net network architecture. The network contains eight basic dense building blocks, as shown in figure 1(b)). The basic network architecture of DenseV-Net takes V-Net as a reference [15]; the network can easily be divided into two parts, namely, encoder and decoder. The encoder applies the basic dense building blocks and a lower convolutional layer to extract features of the images and to use different strides to reduce the resolution of these images. The decoder uses basic dense building blocks and upper convolutional layers to decompress the signal and reconstruct the image until the output is restored to the original size.

For the dense building blocks in the DenseV-Net, we proposed to apply the dense function [15] to alleviate the vanishing and exploding gradients problem and strengthen feature propagation (figure 1(b)); this function was iteratively concatenated six times. The

Table 1. Detailed setting of parameters for this network.

Layer	Kernel size	Stride	In channel	Out channel
Input layer	5	1	1	16
Down-stage 1	5	2	16	32
Down-stage 2	5	2	32	64
Down-stage 3	5	2	64	128
Down-stage 4	5	2	128	256
Up-stage 4	5	2	256	256
Up-stage 3	5	2	256	128
Up-stage 2	5	2	128	64
Up-stage 1	5	2	64	32
Output layer	5	2	32	1

dense function enables every input of the convolutional layer in the dense building blocks to be concatenated with all the previous output of the convolutional layers. The dense function can be written as follows:

$$X_l = H_l([x_0, x_1, ,,,x_{l-1}]) \quad (1)$$

where x denotes the input of the convolutional layer, F(x) denotes the convolutional operation, and H(x) represents the output of the convolutional layer, those being the features after learning. The dense function provides a shortcut connection to the input, which can reduce the increased training error along with an increasing number of layers. The output of each block is processed by a convolutional layer. For the dense building blocks, the input and output have the same shape, and we use exponential linear units (ELUs) to add non-linearity, which is introduced in reference [18].

Rectified linear units (ReLUs), leaky ReLUs (LReLUs), and parametrized ReLUs (PreLUs) are typical activation functions used to reduce the problem of vanishing gradient and to accelerate the training speed. However, the output of ReLUs with no negative values will cause the mean value of the activation output to be higher than zero, furthering the occurrence of bias shift in the last activation unit. Negative values push the mean shifts toward zero, increasing the learning speed. Compared to LReLUs and PreLUs, all with negative values, ELUs have improved robustness to noise and soft saturation when the input is small. The function of ELUs is akin to batch normalization; however, ELUs involve a lower time complexity. The dropout of these ELUs equals 0.8, and experiments show that this activation function produces quicker learning and less variability of forward propagation. To verify the performance of the densely connected dense function, we also implemented V-Net without the dense function to segment the left ventricle. As confirmed by our empirical observations, this DenseV-Net architecture accelerated the speed of convergence in contrast to a V-Net that does not learn dense functions.

For the encoder, the lower convolutional layer processes the input signal using kernels with a size of 5×5 voxels. Inside the dense building block, the convolutional layer with appropriate stride and padding can maintain the size of the image. However, in the process of compression to obtain features of the images, the resolution should be gradually decreased to further extract high-level characteristics. A convolutional layer with appropriate stride is used to replace max-pooling to reduce the resolution. The alternative convolutional layer generally outperforms the pooling layer, which is proposed in reference [19]. The stride of the lower convolutional layers in this network is 2. Consequently, the output of the compressed convolutional layer is halved. In addition, the number of feature maps doubles at each stage of the compressed convolutional layers of network. We applied these operations to obtain complex, high-level features of the left ventricle in medical images and to increase the receptive field to the subsequent layers.

For the decoder of this network, the objective is to extract features and, according to the acquired feature maps, to reconstruct the segmented image. Through mapping the feature maps to the spatial space of the pixels and gathering the low-level features of previous lower convolutional layers, the up-sampling part combines the low-class features to improve the accuracy of segmentation and high-class features, which are advanced and complex. Low-level features are important in the processing of medical data because they represent a minority of features, and this combination enhances the ability to recognize borders [10]. This step compensates for the lack of information in the conventional up-sampling layer. The decoder is completed when the output is the same size as the original input. After each stage of the dense building block is decompressed, the deconvolutional layer is applied to replace the traditional up-pooling layer with a kernel size of 5×5 voxels and a stride of 2. The up-sampling operation doubles the size of the image compared to the previous layer. This step is similar to that of the encoder, and the dense function is referenced in the convolutional dense building block. The details of convolutional layers are shown in table 1.

The architecture has two important improvements. The first is the use of multiple convolutional layers instead of a pooling layer and a deconvolution layer instead of an up-pooling layer, both of which use

smaller memory footprints during training without a need to switch mapping. The input images provided can be more easily understood and analyzed. The other improvement is the forwarding of the features extracted from the encoder to the decoder, which prevents the features being missed during the compression path. Specifically, ELUs are used as an activation function.

Dice coefficient

The prediction of the network is a single channel of a 128×128 matrix, which is the same size as the original input. We separated the foreground and background information via a soft-max algorithm, which was proposed to calculate the probability of classification, prediction, or segmentation. However, traditional loss functions such as cross-entropy loss and l2-loss are not common in evaluations of segmentation. These loss functions easily cause the convergence of data into the regional extreme, which makes the prediction deviate to the background with the missing foreground regions. The dice similarity coefficient (DSC) was proposed to evaluate the segmentation accuracy, as shown in reference [15]. The DSC is a supervised method employing ground truth to calculate the similarity or the divergence of ground truth and predictions. The DSC between two binary volumes can be written as follows:

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

where the predicted binary segmentation volume is $p_i \in P$ and the ground truth binary volume is $g_i \in G$. The range of DSC is from 0 to 1. A higher DSC indicates a greater similarity between the segmented image and the ground truth image. There is no need to focus on balancing the foreground information and background information with this loss function. Figure 3 shows that the segmentation performs well.

Implementation

The model was trained using an Adam optimized function, which is a combination between momentum and RMSprop. We subsequently set the momentum term of Adam to be $\beta = 0.5$. The learning rate was initiated with 1 and gradually reduced to 1e-4 during training. The dropout was 0.8. The network was implemented using the Tensorflow framework and Python language. We used a GTX 1080Ti graphic processor with 4.7 h to train the network.

Results

We obtained image-segmented results using four different deep-learning networks and the ground truth. Figure 2 shows that the left ventricle borders delineated with the proposed method (green) outperform the results of the V-Net method (blue), UNet (yellow), and FCN (cyan). Additionally, the proposed network can segment the left ventricle with an accuracy similar to that achieved by experienced observers (red).

We computed the area of LV segmented by four networks and the area of corresponding ground truth in each slice and plotted the error rate of one slice as a scatter plot (figure 3). The error rate was defined as the difference in value of the left ventricular volume between the proposed methods and ground truth, as a percentage.

The error rate of DenseV-Net is usually lower than that of V-Net, UNet and FCN. Linear regression (figure 4) was used to estimate the consistency between the proposed and manual methods. Based on the 73 slices from the cine MRI of each patient, the outcome of DenseV-Net ($R = 0.92$) was closer to the ground truth than the result of V-Net ($R = 0.77$), UNet ($R = 0.74$) and FCN(0.76), and the segmentation performance was better. The mean DSC of the network with dense function is 0.90, which is higher than that of V-Net (0.85).

In addition to the DSC, we also calculated the volume overlap error (VOE), RV coefficient (RVD), and average symmetric surface distance metric (ASD) as references (figure 5). The mean and median for each index across all cardiac phases for the proposed DenseV-Net and V-Net are shown. Note that the segmentation results and the stability of DenseV-Net outperform those of V-Net.

Discussion

In this paper, we describe an automatic segmentation method to delineate the left ventricle for cine magnetic resonance imaging. The method is based on a deep-learning neural network that combines the dense function, convolutional layer, and deconvolutional layer. The main contribution is that the network is equipped with the dense function so it can critically represent the features of the input image. The reconstructed images from cine MRI are labeled by experienced observers and serve as a training set. The network aims to produce images with delineation of the left ventricle and to separate the background and foreground. After testing the well-trained network on the test label set, the results show that the proposed method can effectively segment the left ventricle with cine MRI. This method can therefore be used to relieve clinicians from manual segmentation work.

In cardiac segmentation, the results based on a deep learning algorithm are affected by various factors including the selection of parameters and loss functions. Therefore, it is difficult to judge the merits of segmentation frameworks. Convolutional Neural Networks (CNNs) have been recently employed to solve problems in both the computer vision and medical image analysis fields. Despite their popularity, most

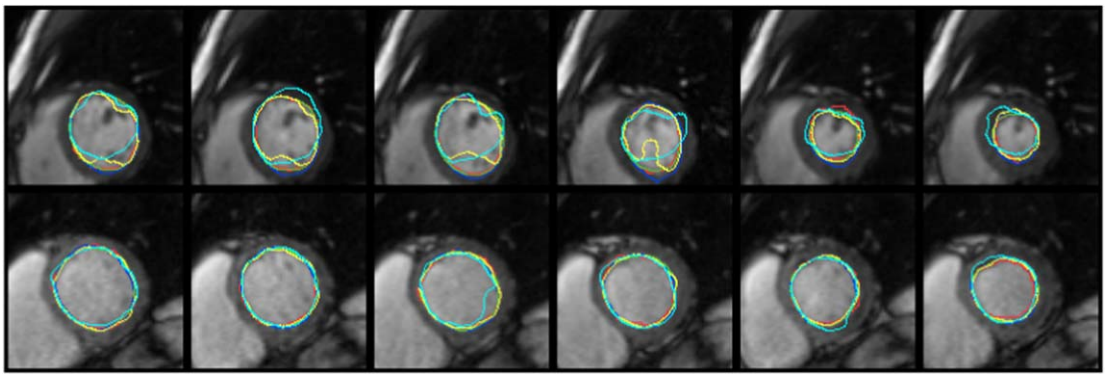

Figure 2. Left ventricle segmentation results from two patient datasets. The left ventricle borders delineated with the proposed method (green) outperformed the results obtained using V-Net (blue), UNet (yellow), and FCN (cyan); the proposed network can segment the left ventricle with an accuracy similar to that achieved by experienced observers (red).

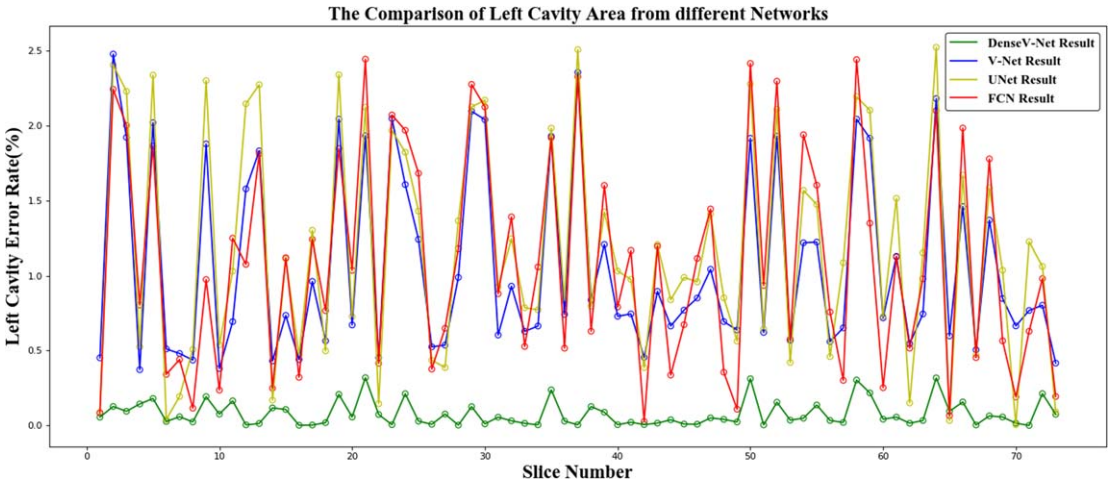

Figure 3. Comparison of error rate for segmentation of the left ventricle using four different deep-learning networks. A left ventricular volume of 73 slices from each cine MRI of patient 1 is provided. The proposed DenseV-Net outperformed the other neural networks.

approaches are only able to process 2D images while most medical data used in clinical practice consist of 3D volumes. V-net is a 3D image segmentation method based on a volumetric, fully convolutional neural network. It is useful in situations where there is a strong imbalance between the number of foreground and background voxels. V-net achieves good performance while requiring only a fraction of the processing time required by other methods. The purpose of our study was to improve upon the V-net method. Our results demonstrate that the proposed DenseV-Net method was strongly correlated with manual analysis in slice-based LV function quantification compared with the V-Net method.

The recently proposed methods efficiently perform segmentation tasks by employing deep-learning networks in the artificial intelligence field, increasing the capacity of computation and the available amount of data. Thus, the motivation for using segmentation for medical organs and tissues has increased [20]. The deep-learning network requires high-quality training data. Because the accuracy of 3D methods has not improved, according to references [21, 22], we chose to use 2D images. In contrast to normal networks which may have a limited amount of training data, the one we trained overcomes the problem of over-fitting. The volume of the left ventricle segmented by cardiac MR data is the critical indicator for calculating the EF, EDV, ESV, and other clinical indicators.

In a recent study about segmentation of the left ventricle, Ngo introduced the application of the Deep Belief Network (DSN) to segment the left ventricle by combining the DSN with a level set method. Of the two DSNs, one was for detecting the LV bounding box and the other was trained to produce an LV probability map upon which the generation of the final segmentation can be based [23]. The initial methods proposed to segment the left ventricle were performed voxel-wise, which leads to the appearance of false-positives. This means that when performing segmentation based on voxels, the potential exists that the performance is unconnected in 3D space. To address this challenge, the deep-learning method is combined with graph partitioning, CRF, MRF, and level setting

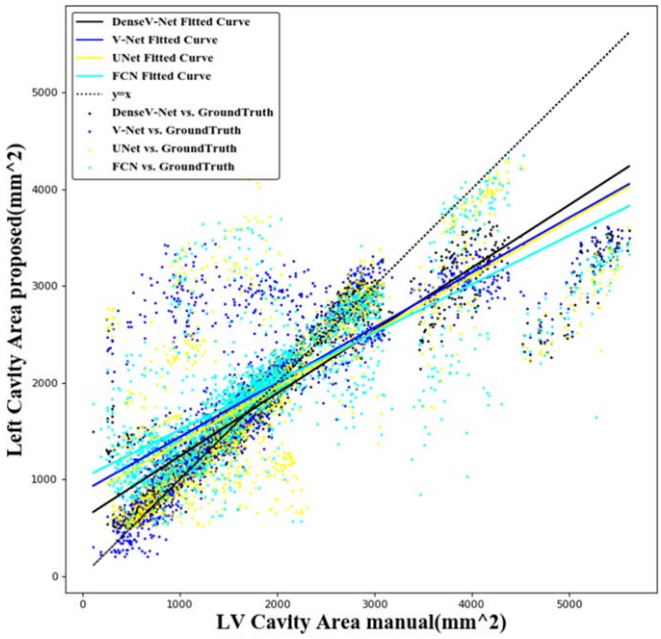

Figure 4. Linear regression analysis of the left ventricular area between the manual methods and four different networks. The correlation coefficients are 0.92, 0.77, 0.74, and 0.76, respectively.

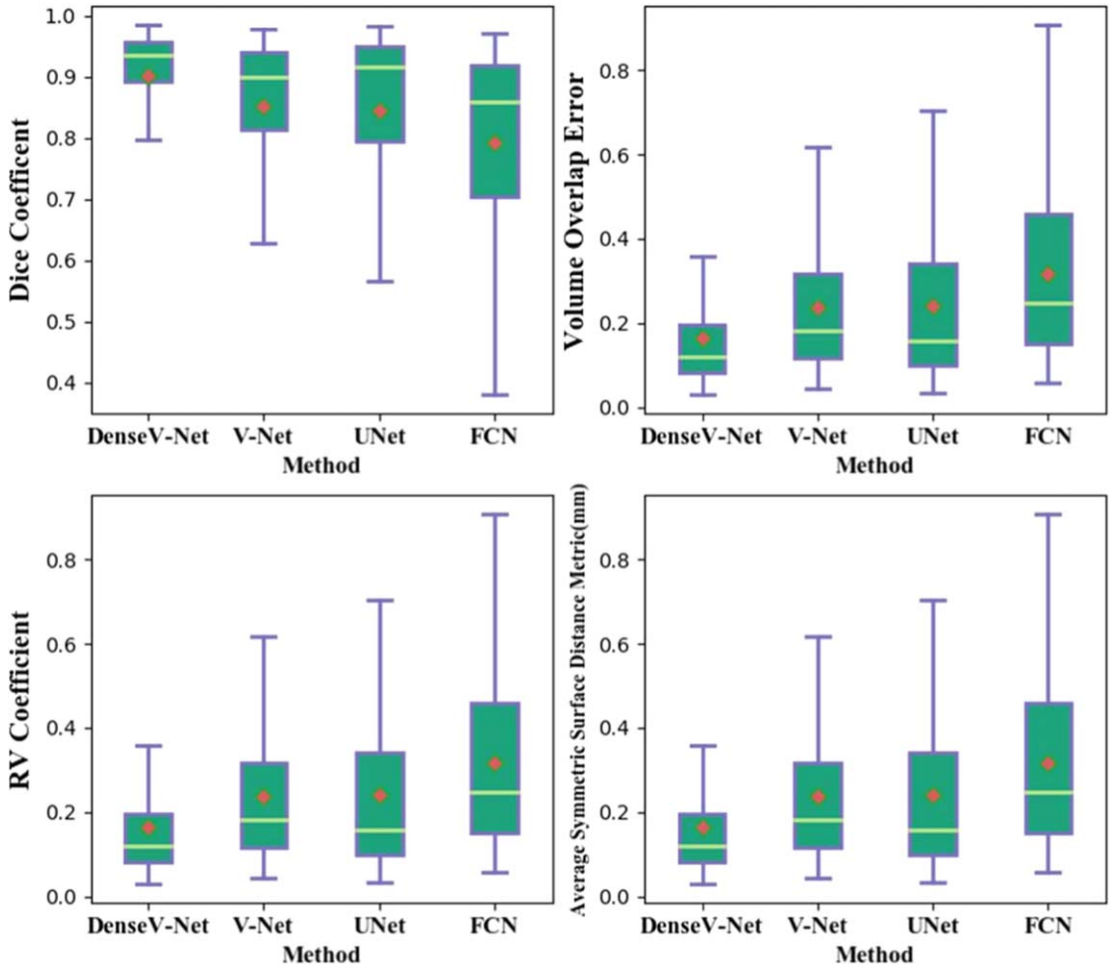

Figure 5. Boxplot for (a) dice similarity coefficient (DSC), (b) volume overlap error (VOE), (c) RV coefficient (RVD), and (d) average symmetric surface distance metric (ASD) for the proposed DenseV-Net, V-Net, UNet, and FCN across all the cardiac slices.

methods [24]. The variant shape of FCN, V-Net, has been employed in different medical applications. Combining the dense function with the convolutional layer enables the left ventricle to be segmented effectively.

Segmentation of the left ventricle is an important process. Considering the importance of the heart, in addition to the canonical long-axis view, there are also two-chamber, three-chamber, four-chamber, five-chamber, and short-axis views. When training all the images with different views in one network, the different segmentation tasks can share the low-level features but require classification layers after extraction of the high-level characteristics. Considering the limitations of our study, in future work we will focus on improving the quality of segmentation. First, the training data we collected are minor compared to the common amount used to train deep-learning networks for other tasks, such as classification and objective recognition, and the work of labeling images is laborious. We can consider a semi-supervised method that was recently introduced to use both labeled and unlabeled data to train the network. Second, real-time cine MRI can be referenced to cardiac functional analysis. Although it has been proposed that the artifacts of respiratory motion can be slightly lessened through the use of the average signal acquired from equipment, direct processing of images from cine MRI may cause unsatisfactory performance because it disregards patients who cannot hold their breath well. Therefore, this image blurring should be considered in the segmentation of the left ventricle in the future.

Additionally, for the limitation of medical imaging, different modalities of medical images can only represent the specific information of humans. Even when using the same medical imaging technique under different contrast enhancement methods and modalities for MRI, the acquired images are different. Hence, combinations of multiple modalities of images may be considered to enhance the accuracy of segmentation and to provide complementary information.

Conclusion

Automatic segmentation based on a deep convolutional network is a reliable method for segmenting the left ventricle for the quantification of left ventricular function. This approach could potentially be useful for releasing physicians from manual delineation, which is time-consuming and tedious.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (81871441, 61901463), the Shenzhen International Cooperation Research Project of China (GJHZ20180928115824168), the Guangdong International Science and Technology Cooperation Project of China (2018A050506064), the Guangdong Special Support Program of China (2017TQ04R395), the Shenzhen Science and Technology Program of China (JCYJ20170413161350892, JCYJ20170818160306270).

Abbreviations

LV: left ventricle; DSC: dice similarity coefficient; EDV: end-diastolic volume; ESV: end-systolic volume; EF: ejection fraction; ECG: electrocardiograph; FCN: fully convolutional neural network; ELUs: exponential linear units; ReLUs: rectified linear units; LReLUs: leaky exponential linear units; PreLUs: parametrized exponential linear units; VOE: volume overlap error (VOE); RVD: RV coefficient; ASD: average symmetric surface distance metric; DSN: deep belief network.

Declarations

Ethics approval and consent to participate

The study was approved by the ethics committee, and informed consent was obtained from all the subjects.

Consent for publication

Written informed consent for the publication was obtained from all the subjects.

Availability of data and material

All the data generated or analyzed during this study are included in this published article and its supplementary information files.

Competing interests

The authors declare that they have no competing interests.

ORCID iDs

Zhanli Hu https://orcid.org/0000-0003-0618-6240

References

[1] Masaki Y *et al* 2006 Cardiac functional analysis by free-breath real-time cine CMR with a spatiotemporal filtering method, TSENSE: comparison with breath-hold cine CMR *Journal of Cardiovascular Magnetic Resonance Official Journal of the Society for Cardiovascular Magnetic Resonance* **8** 801–7

[2] Shuichiro K 2001 Rapid evaluation of left ventricular volume and mass without breath-holding using real-time interactive cardiac magnetic resonance imaging system *Journal of the American College of Cardiology* **38** 527–33

[3] Wu Y et al 2015 *Efficient Method for Analyzing MR Real-Time Cines: Toward Accurate Quantification of Left Ventricular Function* **42** 972–80
[4] Yang P, Liu A C and Liang D H 1998 New real-time interactive cardiac magnetic resonance imaging system complements echocardiography *Journal of the American College of Cardiology* **32** 2049–56
[5] Pham D L X C and Prince J L 2000 *Current Methods In Image Mmedical Segmentation* **2** 315–37
[6] Bai W *et al* 2017 Semi-supervised learning for network-based cardiac MR image segmentation *Int. Conf. on Medical Image Computing & Computer-assisted Intervention* (https://doi.org/10.1007/978-3-319-66185-8_29)
[7] Ibrahim E S and Birchell S 2012 Automatic heart volume measurement from CMR images using ant colony optimization with iterative salient isolated thresholding *Journal of Cardiovascular Magnetic Resonance* **14** P286
[8] Yang B, Xiang D, Yu F and Chen X 2018 *Lung Tumor Segmentation Based on the Multi-Scale Template Matching and Region Growing* (Houston: SPIE) (https://doi.org/10.1117/12.2293065)
[9] Bardinet E, Cohen L D and Ayache N 1998 *A Parametric Deformable Model to Fit Unstructured 3D Data* **71** 39–54
[10] Ciresan D, Luca M G and Jr S 2012 *Deep Neural Networks Segment Neuronal Membranes in Electron Microscopy Images* (https://doi.org/10.1.1.300.2221)
[11] Baumgartner C *et al* 2017 An exploration of 2D and 3D deep learning techniques for cardiac MR Image segmentation *Int. Workshop on Statistical Atlases and Computational Models of the Heart* (https://doi.org/10.1007/978-3-319-75541-0_12)
[12] Zotti C, Lalande A and Luo Z 2017 GridNet with automatic shape prior registration for automatic MRI cardiac segmentation
[13] Ronneberger O, Fischer P and Brox T 2015 Int. Conf. on Medical Image Computing & Computer-assisted InterventionU-Net: convolutional networks for biomedical image segmentation
[14] Long J, Shelhamer E and Darrell T 2014 *Fully Convolutional Networks for Semantic Segmentation* arXiv:1411.4038
[15] Milletari F, Navab N and Ahmadi S A 2016 V-Net: fully convolutional neural networks for volumetric medical image segmentation *Fourth Int. Conf. on 3d Vision* (https://doi.org/10.1109/3DV.2016.79)
[16] He K Z X, Ren S and Sun J 2016 *Deep Residual Learning for Image Recognition* **1** 1–9
[17] Francone M, Dymarkowski S, Kalantzi M and Bogaert J 2005 Real-time cine MRI of ventricular septal motion: a novel approach to assess ventricular coupling *J. Magn. Reson. Imaging* **21** 305–9
[18] Clevert D-A, Unterthiner T and Hochreiter S 2015 *Fast and Accurate Deep Network Learning by Exponential Linear Units (ELUs)* arXiv:1511.07289
[19] Long J S E and Berkeley T D U 2015 Fully Convolutional Networks for Semantic Segmentation arXiv:1411.4038
[20] Lee J G *et al* 2017 Deep learning in medical imaging: general overview *Korean Journal of Radiology* **18** 570
[21] Litjens G *et al* 2017 A survey on deep learning in medical image analysis *Med. Image Anal.* **42** 60–88
[22] Shen D W G and Suk H I 2017 Deep learning in medical image analysis *Annu. Rev. Biomed. Eng.* **19** 221–48
[23] Ngo T A and Carneiro G 2014 Left ventricle segmentation from cardiac MRI combining level set methods with deep belief networks *IEEE Int. Conf. on Image Processing* (https://doi.org/10.1109/ICIP.2013.6738143)
[24] Ngo T A, Lu Z and Carneiro G 2017 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** 159–71