Original Article

Residual Learning: A New Paradigm to Improve Deep Learning-Based Segmentation of the Left Ventricle in Magnetic Resonance Imaging Cardiac Images

Abstract

Background: Recently, magnetic resonance imaging (MRI) has become a useful tool for the early detection of heart failure. A vital step of this process is a valid measurement of the left ventricle's properties, which seriously depends on the accurate segmentation of the heart in captured images. Although various schemes have been tested for this segmentation so far, the latest proposed methods have used the concept of deep learning to estimate the range of the left ventricle in cardiac MRI images. While deep learning methods can lead to better results than their classical alternatives, but unfortunately, the gradient vanishing and exploding problems may hamper their efficiency for the accurate segmentation of the left ventricle in MRI heart images. **Methods:** In this article, a new concept called residual learning is utilized to improve the performance of deep learning schemes against gradient vanishing problems. For this purpose, the Residual Network of Residual Network (i.e., Residual of Residual) substructure is utilized inside the main deep learning architecture (e.g., Unet), which provides more significant detection indexes. **Results and Conclusion:** The proposed method's performances and its alternatives were evaluated on Sunnybrook Cardiac Data as a reliable dataset in the left ventricle segmentation. The results show that the detection parameters are improved at least by 5%, 3.5%, 8.1%, and 11.4% compared to its deep alternatives in terms of Jaccard, Dice, precision, and false-positive rate indexes, respectively. These improvements were made when the recall parameter was reduced to a negligible value (i.e., approximately 1%). Overall, the proposed method can be used as a suitable tool for more accurate detection of the left ventricle in MRI images.

Keywords: *Deep learning, left ventricle, magnetic resonance imaging, residual learning, semantic segmentation*

Maral Zarvani[1],
Sara Saberi[1],
Reza Azmi[1],
Seyed Vahab
Shojaedini[2]

[1]*Faculty of Computer, Engineering Alzahra University,* [2]*Iranian Research Organization for Science and Technology, Tehran, Iran*

Introduction

The cardiovascular system is one of the essential organs in the body, which through it, the oxygenated blood is pumped into the arteries. Due to congestive heart failure, the ventricles may not be fully drained, increasing the adjacent atria and veins' pressure. Since features that affect the cardiac function, especially the left ventricle, are practically treatable and reversible, measuring left ventricular volume is particularly important in cardiovascular monitoring and cure. Diagnostic devices for evaluating the cardiac function are generally divided into invasive and noninvasive types; the noninvasive devices are more common in practice. Angiographic computed tomography (CT) scan and magnetic resonance imaging (MRI) are the most common noninvasive heart monitoring methods, although angiographic MRI is more attractive owing to its better temporal resolution.[1] Furthermore, angiography is more suitable for atrial fibrillation as usually CT imaging is banned due to the irregular and high heart rate and risk of high radiation doses.

In angiographic MRI for accurate measuring of the functional heart behavior should be distinguished from other adjacent organs in the captured chest images. Manual or automatic image segmentation are two main approaches to perform such separation. Manual segmentation (i.e., visual analysis) is time-consuming and entirely depends on the expertise of the technician; therefore, the automated segmentation scheme has been more preferred method during recent years.

Address for correspondence:
Dr. Seyed Vahab Shojaedini,
Iranian Research Organization for Science and Technology,
Tehran, Iran.
E-mail: shojadini@irost.ir

Access this article online

Website: www.jmssjournal.net

DOI: 10.4103/jmss.JMSS_38_20

Quick Response Code:

However, the automatic segmentation of angiographic MRI images also suffers from some problems which hamper its performance. The most important limiting factors are (i) cardiac MRI images exhibit a significant variation in terms of either gray levels or structural shapes, (ii) the gray levels of images may also differ due to the use of different MRI scans, (iii) blood flow in parallel with respiration motion may cause considerable image fuzziness, mostly due to blood flow, (iv) the shape of the ventricle varies both over the patient and time, and finally, (v) the low contrast of angiographic MRI images, which may cause low performance of the automated techniques for distinguishing the left ventricle in the images.[2] Several studies tried to solve these problems and improve the automated detecting left ventricle algorithms' performance in angiographic MRI images.[3,4] In general, the MRI cardiac image segmentation methods can be divided into bellow five categories:[5]

Thresholding based models

In the earliest methods, the main idea of thresholding was combined with several complementary image processing methods to extract the region of the left ventricle.[6] In such scenarios, the locations of the endocardial surfaces are initially approximated by intensity thresholding. Then, each approximated surface points are replaced with the nearest locally maximum gradient magnitude points to fit a cylinder. As the brightness of the pixels varies significantly for the limiting factors mentioned above, the high dependence on the threshold is the most significant weakness of these methods.

Deformable models

Deformable models are widely used in medical image segmentation, especially for detecting the heart's borders and segmenting the left and right ventricles. This method is a combination of geometry, physics, and estimation theory which its basic idea is considering an initial shape for the object and then minimizing the energy function. The initial shape is formulated according to the extent of some prior knowledge, such as location, shape, and size. One of the most popular schemes in this group is the active contour algorithm. Unfortunately, these algorithms have different problems, such as sensitivity to artifacts and noise, depending on the initial information.[7-9]

Learning appearance and shape

These statistical algorithms, which were widely used for left ventricle detection, use the training datasets that include the main image and its variations.[10,11] The primary strategy is to minimize the energy function considering some parameters, including edge, texture, and elasticity. The performance is highly dependent on the training data, especially training manual contours, leading to low generalization power and reduced efficiency in the real applications.[12]

Atlas-based schemes

The algorithm and its derivatives, with capable of 3-D and 4-D heart segmentation,[13,14] received a manually segmented image as the base image (Atlas), compare it to other images, and then measure their similarities. A mapping between the new image and the Atlas is done for the segmentation of the new image. Then the transforms on the new image are applied to the Atlas to obtain the final segmentation.[5,15] These algorithms cannot be accurate for new images as they suffer from weak generalization power.

Deep learning

Compared to other methods, deep neural networks have been widely applied to analyze various types of medical image processing schemes, including heart segmentation, thanks to their superior features.[4]

In some researches, Deep Belief Networks have been utilized as a nonrigid classifier for heart segmentation in the medical images.[16] The disadvantage of this approach can be the lack of a dynamic model. In some other studies, the proposed methods classified the heart based on Convolutional Neural Networks (CNNs).[17] In these schemes, the spatial information lost in the local processing schemes may be improved by applying a fully connected CNN layer. Unet architecture was also introduced for heart segmentation in some researches.[18] The main difference between Unet and fully convolutional networks is the symmetry of Unet architecture along with adding concatenation operation in the decoder path, as is explained in section 2-1. Subsequently, various improvements in Unet are addressed in numerous articles that were mainly focusing on the innovations in the following three areas (i) changing the encoder network to extract more abstract features, (ii) modifying the up-sampling strategy, and (iii) changing the skip connections. For example, in[18] Unet++ was introduced. In the Unet network, data in skip connection path transmit directly from the encoder to the decoder, whereas, in Unet++, it passes through a series of Conv-blocks and transfers the feature maps.[19] Some other researchers used Feature Pyramid Networks advantages, including their flexibility and robustness, to propose the MFP-Unet model for semantic segmentation of the left ventricle.[20] Leclerc *et al*.[17] performed left ventricle image segmentation so that the training data were annotated by the Kalman filter for the Unet network. Furthermore, methods based on detecting the Region of Interest using deep neural networks were proposed in some other studies.[21] Finally, in some recent studies, magnetic resonance (MR) sequences of several subjects were investigated using Deep Multitask Relationship Learning Network, which contains a CNN for feature extraction and two recursive parallel networks for dynamic temporal modeling of cardiac sequences. In these techniques, sometimes Bayesian-based multitask learning was utilized for estimating left ventricle indices and a softmax classification

for cardiac phase detection.[22] Yan *et al*.[23] reported that the accuracy of LV spatial segmentation improved as a result of (1) modification of the Unet encoder with additional temporal coherence features in cardiac images and (2) usage of dilated convolution layers. Furthermore, a more comprehensive classification of cardiac segmentation deep learning methods is presented in[24] including Recurrent Neural Network, Generative Adversarial Networks, Auto-Encoders, and 3-D segmentation networks. Zhang *et al*. considered both temporal and spatial properties using the Resnet-based LSTM networks, exhibiting more robust model compared to image inhomogeneity.[25]

Proposed Method

As the main idea of this research was to improve the performance of the Unet in heart segmentation by using the concept of residual learning, in this section, the Unet and ResNet-Unet structures are firstly described. Then, the proposed structure based on Residual of Residual (ROR)-Unet is explained in the next section.

Unet

The procedure of Unet involves two steps. In the first step, feature maps are extracted using the convolution layers, followed by applying Max Pooling. These maps contain the full context of the input image pixels; therefore, they are saved for use in the next step, as shown in Figure 1. In the second step, the stored feature maps are concatenated with applying the zero-padding, convolution, and up-sampling. This scheme may improve the recovery of spatial resolution at the network output.[26]

This structure's remarkable fact is that skip connections improve the lost spatial features in the first step (encoding) because spatial information is crucial in semantic segmentation. Better representation and abstraction of the data encoding may lead to better network performance in the second step (i.e., decoding) and improve the semantic segmentation task. For this purpose, it is possible to use pretrained networks such as VGG[20] as the encoder of Unet architecture. ResNet-Unet architecture is discussed in the next section.

ResNet-unet

Although an increase in the depth of the neural networks leads to higher classification performance,[20,21,27] but such an increase may cause gradient vanishing problem during back-propagation. ResNet networks may partially alleviate this problem by performing identity mapping, which skips from one or more layers. Using this technique caused the gradient was reinforced in the deeper layers in parallel with some increase in convergence speed. The composite structure, called ResNet-Unet, is shown in Figure 2 including two types of blocks: identity-block and convolutional-block. The identity block has not convolution layer in shortcut, and the output possesses the same
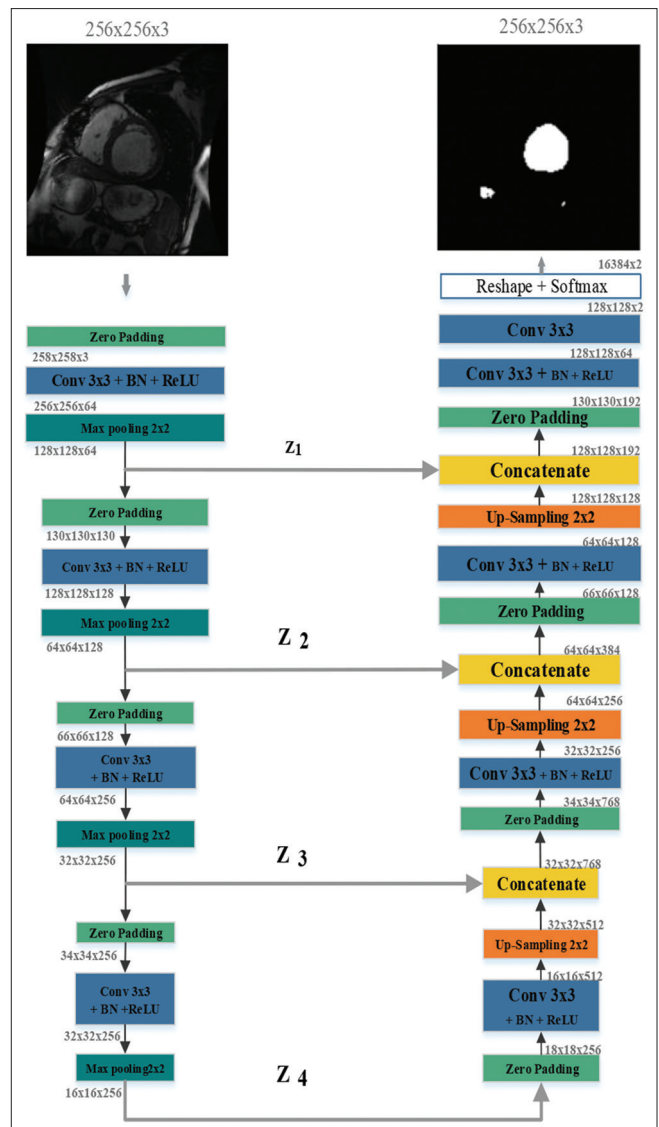

Figure 1: Schematic representation of the Unet architecture

dimension as the input dimension.[28] As shown in Figure 3 both these blocks consist of two 3 × 3 convolution layers, followed by ReLU and Batch normalization layers.

As demonstrated in Figure 3, in the encoding path, four long skip connections transfer the copy of feature maps into the decoding path, which includes some identity and convolutional blocks. Furthermore, an up-sampling layer is utilized in the decoder in which the feature maps are transformed into high-resolution images. Finally, the output is concatenated with the feature maps of the corresponding encoding path. Equations (1-2) briefly describe the logics of identity and convolutional blocks f representing the nonlinear ReLU function, furthermore $F_1(x)$, and $F_2(x)$ demonstrate the identity mapping and residual mapping functions, respectively. Finally $O_i(x)$ and $O_c(x)$ show identity and convolutional outputs, respectively.

$$O_i(x) = f\left(F_1(x) + x\right) \qquad (1)$$

$$O_c(x) = f\left(F_1(x) + F_2(x)\right) \tag{2}$$

Residual of residual-Unet

Taking into account that the problem of gradient vanishing is only partially reduced by using ResNet-Unet, in this study, the ROR (*so-called* ROR) structure is utilized as a type of residual learning to improve the performance of Unet. In the ROR structure, level-wise shortcut connections were added to the network, which may provide information exchange between the branches (contains a series of identity and Conv-block). This scheme may improve learning ability, thanks to the reduction of the gradient vanishing problem. As shown in Figure 4, the proposed ROR-Unet structure contains a similar encoding path to the ResNet-Unet. This path consists of identity-blocks
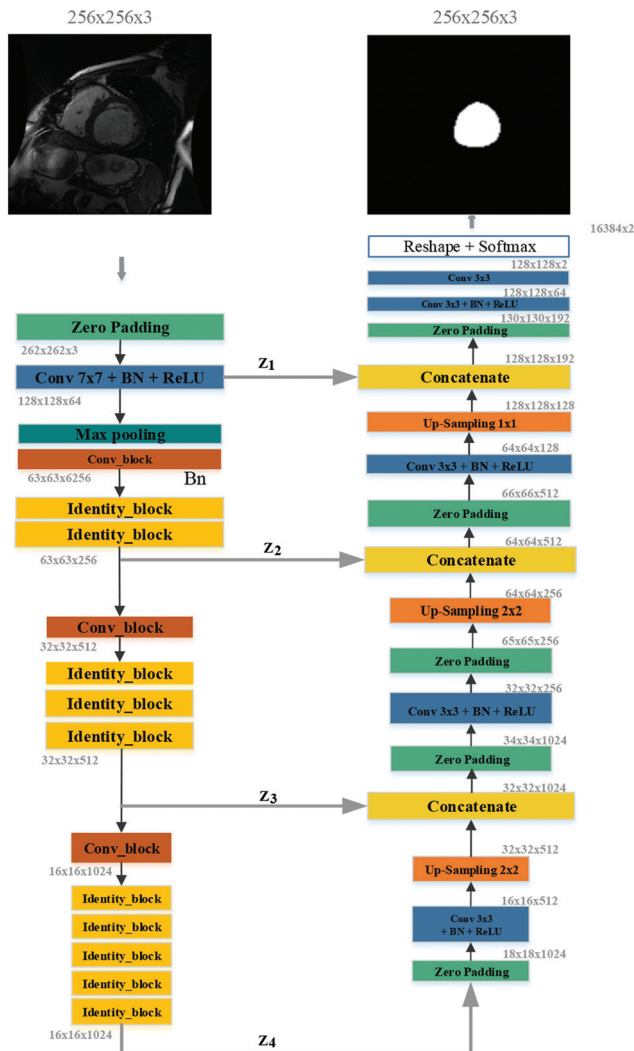
and Conv-blocks, but the difference is that the shortcut connections have been added to the level by level residual blocks. The number of ahortcut levels is a hyper-parameter, which can significantly affect the results. Therefore, it should be determined by the multiple experiments. The proposed structure follows three shortcut levels as described below:

1. Level-0 (*so-called* root shortcut): This shortcut maps the result of applying 3×3 convolution and zero padding on the input image into the end of the encoding path. This task is performed by making use of a 1×1 convolution

2. Level-1: At this level, the ResNet architecture is divided into three branches based on the total number of identity-blocks and conv-blocks. Each component consists of one convolutional-block and more than ten identity-blocks. All convolution layers in every three shortcuts possess the same structure (e.g., 1×1 kernel size with stride one and padding the same). The only difference is the number of the feature channels, respectively, are 64, 128, and 256 from the beginning to the end

3. Level-2: These shortcuts are original in ResNet structure and are used in identity and convolution blocks. Equations (3-8) show how the result of each of the mentioned branches is calculated. If Z_1, Z_2, Z_3, and Z_4 denote the outputs of branches as shown in equations (3)-(6), then equation (7)-(8) may illustrate that they are transported into the decoder layer and concatenate with the output of the transposed convolution layer (i.e., Deconvolution block). The deconvolution block applies a 3×3 convolution layer, followed by batch normalization, ReLU, and 2×2 up-sampling layers on its input data. Each up-sampling layer doubles the feature maps and halves the number of feature channels correspondingly.

$$Z_1 = X \tag{3}$$

$$Z_2 = f\left(O_{i1}^{16}\left(O_{c1}(X)\right)\right) + G(X) = \tag{4}$$
$$f\left(O_{i1}^{16}\left(f\left(F_1(X) + F_2(X)\right)\right)\right) + G(X)$$

$$Z_3 = f\left(O_{i2}^{16}\left(O_{c2}(Z_2)\right)\right) + G(Z_2) \tag{5}$$

$$Z_4 = f\left(O_{i3}^{16}\left(O_{c3}(Z_3)\right)\right) + G(Z_3) + G(X) \tag{6}$$

Figure 2: ResNet-Unet architecture in which the input image is applied by zero paddings, a 7 × 7 convolution layer with step 2 and 64 feature channels followed by batch normalization layer and nonlinear ReLu function. Then, conv-block and the number of identity blocks are applied in each branch. The internal architecture is illustrated in Figure 3. The decoding part is the same as Unet
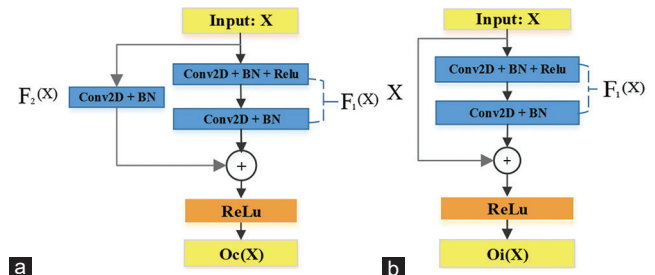

Figure 3: (a) The Conv-Block architecture (b) The architecture of Identity-Block

$$O_c(X) = f(F_1(X) + F_2(X)) \tag{7}$$

$$O_i(X) = f(F_1(X) + X) \tag{8}$$

The above X equations are considered the result of applying consecutive zero padding and convolutional layers on the input image. Furthermore, $O_c(X)$ and $O_i(X)$ represent the output of convolutional-block and identity-block in the first branch. $G(\ )$ is defined as a residual mapping function for each input and $f(\ )$ demonstrates the ReLU activation function. Finally, $O_i^{16}(X)$ shows the result obtained from 16 identity blocks (also may be shown as $(F_1 o...F_1)$).

The proposed method by strengthens the residual learning concept (ROR) improved the gradient vanishing and exploding lead to makes it possible to deepen the network by adding more branches and layers, which leads to more abstract features that ultimately increase the accuracy of segmentation in comparison with Unet and ResNet-Unet.
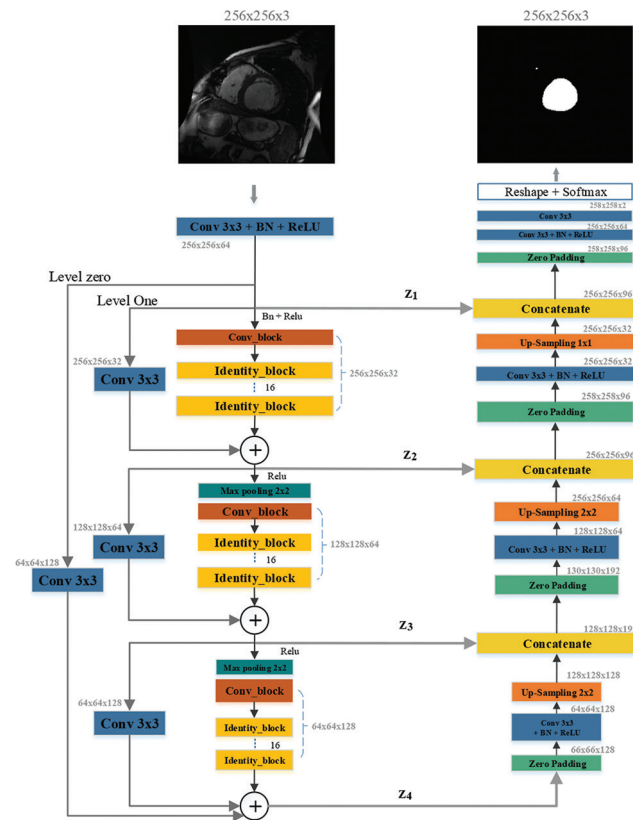

Figure 4: The encoder of Residual of Residual-Unet consists of 3 Branches. Each Branch contains 1 conv-block and 15 identity-blocks that has the same number of feature channels in identity and conv blocks. Each branch contains a Level 1 shortcut that contains one 1 × 1 convolution layer. A Level shortcut 0 by one 1 × 1 convolution layer, map the features from the start to the end of the Branch. After applying each max pooling layer, the size of the feature maps is halved. Z1, Z2, Z3, and Z4 are the outputs of branches which transport to the decoder layer and concatenate with the output of the transposed convolution layer

Experiments and Results

The proposed algorithm was applied to a set of cardiac MRI images known as Sunnybrook Cardiac Data[29] to evaluate its performance. This dataset, also known as the 2009 Cardiac MR Left Ventricle Segmentation Challenge data, includes 805 images of 45 cine-MRI with the same sizes (256 × 256 pixel). The captured images may be classified into four categories: healthy (N), hypertrophy, heart failure with infarction (HF-I), and heart failure without infarction (HF-NI).[29] The data set was split into five subsets (i.e., k1 to k5) with an equal percentage (20%) to make possible using a 5-fold cross-validation strategy. In the Table 1, the number of samples of each category is shown for k1.k5, respectively. The test procedure and some of the images and training essential parameters are shown in Table 1.

The proposed method was implemented on a testbed prepared using the Keras framework on a computer equipped with a GeForce GTX 1070 Ti with 8 GB RAM. Furthermore, other methods consisting of Unet[30] and ResNet-Unet[31,32] were used in parallel and applied on the same data for comparison to the proposed algorithm. However, different parameters were used for each case. The specifications of our best structure are reported in Table 2 for each case.

The images were firstly processed by experts to obtain a ground truth for comparison with the automatic methods. For better understanding, some results from the proposed

Table 1: Specifications of train and test data

Categories	n	HYP	HF-I	HF-NI
Number of sample				
k1				
Train	113	143	195	193
Test	29	41	50	41
k2				
Train	116	144	200	184
Test	26	40	45	50
k3				
Train	113	151	191	189
Test	29	33	54	45
k4				
Train	103	142	205	194
Test	39	42	40	40
k5				
Train	123	156	189	176
Test	19	28	56	58
Number of gender				
Male	6	7	11	8
Female	3	5	1	4
Average of age	60	57	61	64
Average contrast (%)	34.46	36.30	35.91	44.62

HYP: Hypertrophy, HF-I: Heart failure with infarction, HF-NI: Heart failure without infarction

and alternative schemes are graphically shown in this section. However, the complete statistics of the test results are discussed in the next section. Figure 5 shows some original heart MRI images, whereas Figure 6 shows their manual segmentation results (i.e., ground truth). Besides, Figure 7 shows the results obtained by using the proposed method. It may be noted that regions of the left ventricle in these images were detected with striking similarity with their ground truth. Lower matched results are observing in Figures 8 and 9 are corresponding to Unet and ResNet-Unet algorithms, respectively. These show that the different performances may happen due to applying proposed and alternative methods, especially when the left ventricle possesses regions with different visual properties or when there is no good contrast between the left ventricle and

other parts of the heart. For example, Figure 9a-i shows that ResNet-Unet method misidentified some additional components and labeled them as the left ventricle. A similar, more pronounced finding was observed in Figure 8a-i as two or three additional segments were extracted as the left ventricle. The above failures occurred in the alternative algorithms, while the proposed method could identify the left ventricle without these effects, as shown in Figure 7.

Another inspiring example of how the proposed method was superior to its alternatives can be observed by comparing Figures 7c, 8c and 9c. Figure 7c shows that the proposed method extracted the left ventricle similar to Figure 6c. However, the alternative methods could not perform a proper
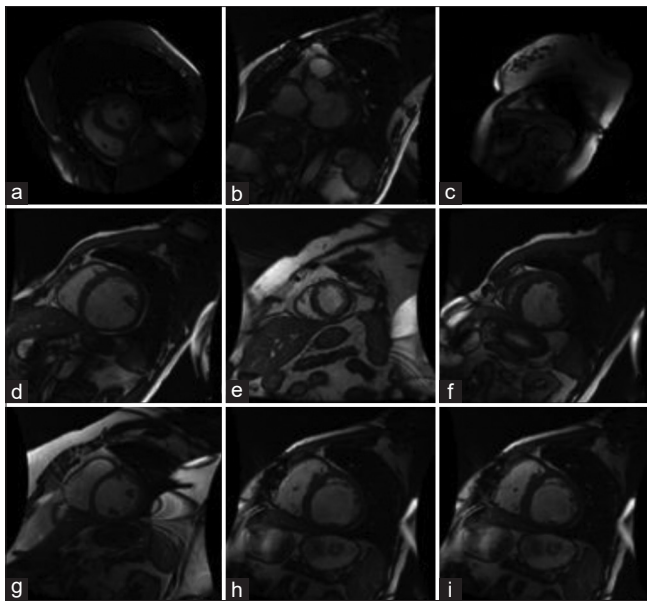

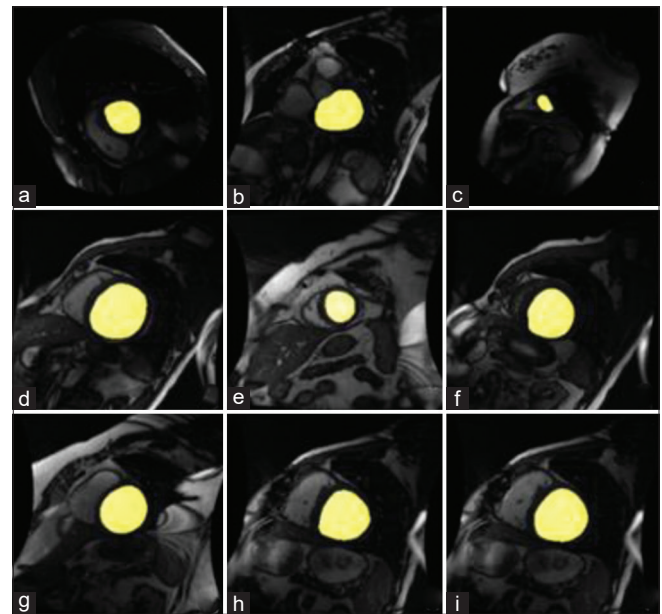Figure 5: Some original heart magnetic resonance imaging images

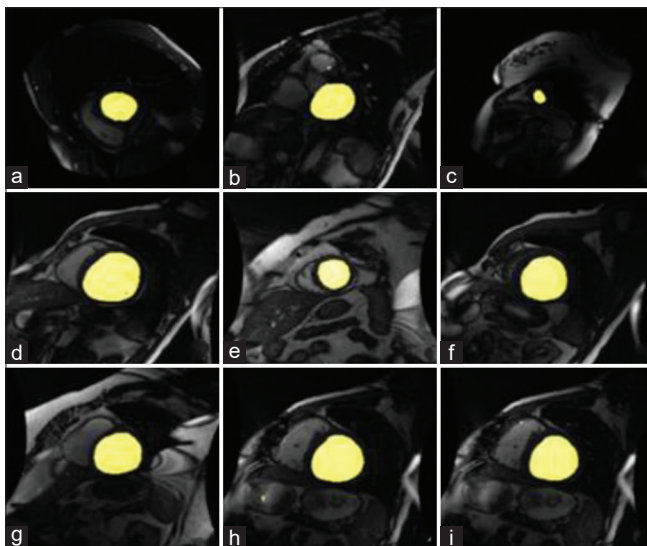Figure 6: The ground truth belonging to images shown in Figure 5

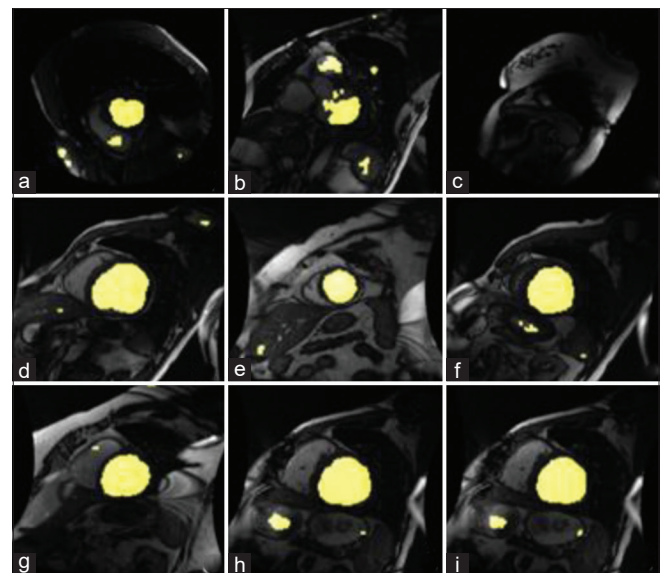Figure 7: The results of applying the proposed method on images shown in Figure 5

Figure 8: The results of applying the Unet method on images shown in Figure 5

segmentation without accurate detection of the left ventricle. It should be noted that all the results obtained by the proposed

Figure 9: The results of applying the ResNet-Unet method on images shown in Figure 5

and alternative methods were not necessarily very different. There was a situation that both examined approaches achieved similar results. Examples of this situation are shown in Figures 7e, g and 9e, g and to some extent in Figure 8e and g, all showing that the examined methods achieved acceptable results in the segmentation of the left ventricle. The examples mentioned above highlight that the examined algorithms were delivering different performances for different types of heart images as fully quantified in the next section.
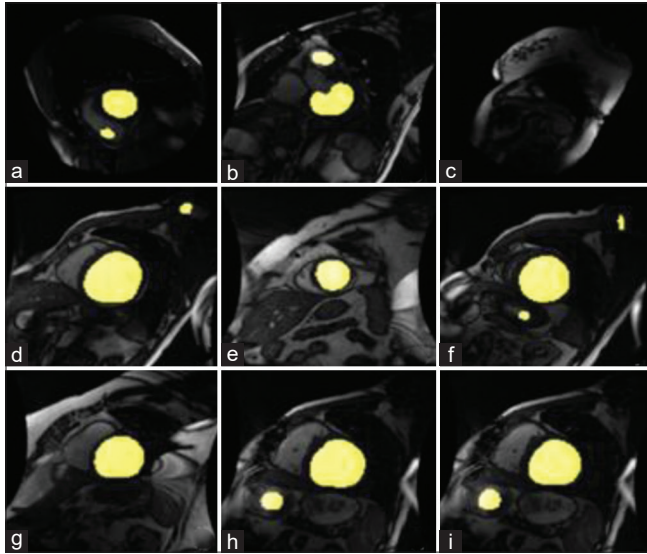
Discussion

To define the evaluation parameters, let *AS* and *MS* donate the result of automatic and ground truth segmentations, respectively. As mentioned in section 3, both *AS* and *MS* are binary images as the pixels belong to the left ventricle have label 1, and all the others have label 0. To perform a more convincing evaluation, five common metrics were used in this study. These parameters have frequently been applied in semantic segmentation, especially in the medical image segmentation.[33] These parameters include:

i. True positive rate measures the rate of pixels labeled as left ventricle by both the automatic and the ground truth as:

Model	Unet	ResNet-Unet		ROR-Unet		
Table 2: Specifications of the optimal structure for examined models						
Learning parameters						
Input size	256×256	256×256		256256		
Convolution layers#	10	57		110		
Trainable parameters (million)#	4.46	16.34		63.46		
Drop out	0	0		0		
Level of shortcuts	-	1		3		
Max epoch	500	250		250		
Mini batch size	150	150		150		
Optimizer	Adadelta	Adadelta		Adadelta		
Initial weights	-	ImageNet		ImageNet		
Layers parameters						
Feature map order	3->64 ->128 ->256 ->512 ->256 ->128 ->64 ->2	3->64 ->256 ->512 ->1024 ->512 ->256 ->128 ->64 ->2		3->32 ->64 ->128 ->64 ->32 ->2		
Activation function	All layers expect last are ReLU, the last layer is softmax	All layers expect last are ReLU, the last layer is softmax		All layers expect last are ReLU, the last layer is softmax		
Kernel size of convolution layers	All convolution layers are 3×3	Branch1	First convolution 7×7	Branch1	Convolution 1 in each block	3×3
			Others 3×3		Convolution 2 in each block	3×3
		Branch2	All convolutions: 3×3	Branch 2 Convolution 2 in each block	Convolution 1 in each block 5×5	7×7
		Branch3	All convolutions: 3×3	Branch3 Convolution 2 in each block	Convolution 1 in each block 11×11	9×9

ROR: Residual of residual

$$TPR = \frac{AS \cap MS}{MS} \qquad (9)$$

ii. False positive rate (FPR), also is known as *Fall-Out*, shows the rate of pixels labeled as the left ventricle by an automatic algorithm that was not indicated as left ventricle by the ground truth as:

$$FPR = \frac{AS \cap M\tilde{S}}{MS} \qquad (10)$$

iii. Precision, measures the rate of pixels correctly labeled to all predicted pixels as left ventricle by automatic:

$$Precision = \frac{TP}{TP + FP} \qquad (11)$$

iv. The Intersection-Over-Union (IoU), also known as the Jaccard Index, and dice coefficient[34] computes two standard parameters for measuring the ability of heart segmentation algorithm[35] with the equations (4) and (5):

$$Dice = \frac{2 \times |MS \cap AS|}{|AS| + |MS|} = \frac{2 \times (TP)}{(TP + FP) + (TP + FN)} \qquad (12)$$

$$Jaccard - Index = \frac{|MS \cap AS|}{|AS \cup MS|} = \frac{TP}{TP + FN + FP} \qquad (13)$$

The 5-fold cross-validation strategy was used for the evaluation of all methods. As shown in Table 3, the proposed method outperforms Unet in terms of Jaccard-index, precision, dice co-efficient, and FPR.

As shown in Table 2, the proposed structure outperformed most of the examined parameters compared to its alternatives. The obtained Jaccard-index revealed that the average value gained using the proposed structure is better than those obtained by ResNet-Unet and Unet by extents of 5.1% and 6.1%, respectively. A similar trend is observed for the Dice parameter as its average value for the proposed scheme captured 3% and 4.1% higher values for ResNet-Unet and Unet, respectively. The proposed structure's superiorities against its alternatives are more significant in terms of precision and FPR indexes. The above table shows that the precision of the proposed method was 8.6% and 8.2% higher than its alternatives. In the same manner, the FPR index demonstrates that the ROR-Unet structure obtained less false detection zones in the range of 14.3% and 11.5% against ResNet-Unet and Unet, respectively. However, in terms of the Recall parameter, the result was somewhat different, despite these mentioned substantial advantages. The last column in the Table 3 indicates that the best Recall was obtained by using ResNet-Unet structures among three examined models so that its detection rate was 3.6% better than the proposed structure. This result is maybe since the proposed, and Unet methods have shown almost a similar recall. However, the above analyzed results indicate that the benefit of the proposed method in the other four parameters is far superior to its weakness in the Recall parameter compared to its alternatives.

Table 3: Comparison of the evaluation indexes obtained for examined algorithms

Method	Jaccard-index	Dice	Precision	FPR	Recall
Unet					
k1	0.808	0.884	0.840	0.248	0.941
k2	0.799	0.876	0.831	0.260	0.930
k3	0.818	0.894	0.837	0.262	0.958
k4	0.811	0.881	0.875	0.155	0.906
k5	0.793	0.867	0.826	0.246	0.940
Mean	0.806	0.880	0.842	0.234	0.940
ResNet-Unet					
k1	0.812	0.889	0.826	0.271	0.964
k2	0.811	0.888	0.842	0.284	0.956
k3	0.820	0.894	0.831	0.268	0.968
k4	0.8261	0.900	0.845	0.244	0.964
k5	0.811	0.885	0.845	0.244	0.951
Mean	0.816	0.891	0.838	0.262	0.960
ROR-Unet					
k1	0.864	0.922	0.931	0.083	0.949
k2	0.879	0.936	0.926	0.12	0.948
k3	0.871	0.938	0.928	0.097	0.944
k4	0.857	0.913	0.917	0.187	0.939
k5	0.863	0.921	0.914	0.115	0.945
Mean	0.866	0.926	0.923	0.120	0.945

ROR: Residual of residual, FPR: False-positive rate

Consequently, the proposed method can still be considered more successful than detecting the boundary of the left ventricle. Some factors can justify the efficiency of the proposed process over other deep learning-based techniques. In the decoder of Unet, the feature maps generated by the encoder will be up-sampled so that they can ultimately produce a high-resolution output image with the same size as the input image. This step aims to determine the accurate location of the object in the input image (also called object localization).

This requires semantic coherence between object pixels or the quality of spatial features extracted from the original image. Therefore, a higher quality of encoder feature maps results in better generating of segmentation masks. Due to the low number of convolution layers, the feature maps generated in the encoder of the original Unet and transferred to the decoder are not of high quality; thus, the performance of generating actually pixel-wise classification segmentation mask is low. The residual network structure makes it possible to add more convolution layers, thanks to the residual shortcuts. These shortcuts reinforce the feature maps in the deeper layers and effectively prevents the gradient vanishing problem. In the proposed model, adding different levels of shortcuts to the ResNet structure leads to higher controlling of the gradient vanishing problem, consequently allowing adding more convolution layers. Furthermore, in the encoder path, these shortcuts strengthen the feature maps of the deeper layers by aggregating shallow layers feature maps.

Conclusion

In this study, a new method was introduced to improve the performance of deep neural networks for detecting the left ventricle in MRI images. We developed the residual network by adding some branches and used it as the Unet encoder in the proposed structure ROR-Unet network to overcome some challenges in the deep learning paradigm, mainly include gradient vanishing and exploding. To evaluate the performance of the proposed algorithm, real cardiac MRI images were examined parallel with two existing methods (i.e., Unet and ResNet-Unet). The results were interpreted based on five well-known detection evaluating indexes consisting of Jaccard, Dice, Precision, FPR, and Recall parameters. The results clearly demonstrated a significant superiority of the proposed method over its closest alternative (between 3.5% and 11% for the first four parameters) for accurate estimation of the left ventricle boundaries. The results also showed that the above significant improvements were achieved when there was no significant drop in the fifth parameter (i.e., Recall). Based on the results achieved in this study, it may be concluded that the proposed method can be used as a suitable alternative tool for determining the left ventricle region in MRI images, helpful for the early detection of heart failure.

Financial support and sponsorship

None.

Conflicts of interest

There are no conflicts of interest.

References

1. Peterzan MA, Rider OJ, Anderson LJ. The role of cardiovascular magnetic resonance imaging in heart failure. Cardiac Fail Rev 2016;2:115.
2. Mazaheri S, Sulaiman PSB, Wirza R, Khalid F, Kadiman S, Dimon MZ, *et al*. Echocardiography image segmentation: A survey. In: 2013 International Conference on Advanced Computer Science Applications and Technologies. Kuching, Malaysia: IEEE; 2013.
3. Merjulah R, Chandra J. Segmentation technique for medical image processing: A survey. In: 2017 International Conference on Inventive Computing and Informatics (ICICI). Coimbatore, India: IEEE; 2017.
4. Wolterink JM. Left ventricle segmentation in the era of deep learning. In: Virginia, Springer: Springer; 2019.
5. Petitjean C, Dacher JN. A review of segmentation methods in short axis cardiac MR images. Med Image Anal 2011;15:169-84.
6. Goshtasby A, Turner DA. Segmentation of cardiac cine MR images for extraction of right and left ventricular chambers. IEEE Trans Med Imaging 1995;14:56-64.
7. Kass M, Witkin A, Terzopoulos D. Snakes: Active contour models. Int J Comput Vis 1988;1:321-31.
8. Avendi MR, Kheradvar A, Jafarkhani H. A combined deep-learning and deformable-model approach to fully automatic segmentation of the left ventricle in cardiac MRI. Med Image Anal 2016;30:108-19.
9. Metaxas DN, Yan Z. Deformable models, sparsity and learning-based segmentation for cardiac MRI based analytics. In: Handbook of Medical Image Computing and Computer Assisted Intervention. Cambridge, Massachusetts, United States: Elsevier; 2020. p. 273-92.
10. Van Der Geest RJ, Lelieveldt BP, Angelié E, Danilouchkine M, Swingen C, Sonka M, *et al*. Evaluation of a new method for automated detection of left ventricular boundaries in time series of magnetic resonance images using an active appearance motion model. J Cardiovasc Magn Reson 2004;6:609-17.
11. Mitchell SC, Lelieveldt BP, van der Geest RJ, Bosch HG, Reiber JH, Sonka M. Multistage hybrid active appearance model matching: Segmentation of left and right ventricles in cardiac MR images. IEEE Trans Med Imaging 2001;20:415-23.
12. Cootes TF, Hill A, Taylor CJ, Haslam J. Use of active shape models for locating structures in medical images. Image Vis Comput 1994;12:355-65.
13. Lorenzo-Valdés M, Sanchez-Ortiz GI, Elkington AG, Mohiaddin RH, Rueckert D. Segmentation of 4D cardiac MR images using a probabilistic atlas and the EM algorithm. Med Image Anal 2004;8:255-6.
14. Zhuang X, Hawkes DJ, Crum WR, Boubertakh R, Uribe S, Atkinson D, *et al*. Robust registration between cardiac MRI images and atlas for segmentation propagation. In: Medical Imaging 2008: Image Processing. California, United States: International Society for Optics and Photonics; 2008.
15. Iglesias JE, Sabuncu MR. Multi-atlas segmentation of biomedical images: A survey. Med Image Anal 2015;24:205-19.
16. Carneiro G, Nascimento JC, Freitas A. The segmentation of the left ventricle of the heart from ultrasound data using deep learning architectures and derivative-based search methods. IEEE Trans Image Process 2011;21:968-82.
17. Romaguera LV, Vázquez L, Costa MG, Romero FP, Filho CF. Left ventricle segmentation in cardiac MRI images using fully convolutional neural networks. In: Medical Imaging 2017: Computer-Aided Diagnosis. Orlando, Florida, United States: International Society for Optics and Photonics; 2017.
18. Leclerc S, Sarah, Smistad E, Grenier T, Lartizien C, Ostvik A, *et al*. Deep learning applied to multi-structure segmentation in 2D echocardiography: A preliminary investigation of the required database size. In: 2018 IEEE International Ultrasonics Symposium (IUS). Kobe, Japan: IEEE; 2018.
19. Zyuzin V, Chumarnaya T. Comparison of unet architectures for segmentation of the left ventricle endocardial border on two-dimensional ultrasound images. In: 2019 Ural Symposium on Biomedical Engineering, Radioelectronics and Information Technology (USBEREIT). Yekaterinburg, Russia: IEEE; 2019.
20. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition.arXiv;2014:1409.1556.
21. He K, Zhang X, Ren S, Sun J. Identity mappings in deep residual networks. In: European Conference on Computer Vision. Amsterdam, Netherlands: Springer; 2016.
22. Litjens G, Kooi T, Bejnordi BE, Setio AA, Ciompi F, Ghafoorian M, *et al*. A survey on deep learning in medical image analysis. Med Image Anal 2017;42:60-88.
23. Yan W, Wang Y, Li Z, Geest RJ, Tao Q. Left ventricle segmentation via optical-flow-net from short-axis cine MRI: Preserving the temporal coherence of cardiac motion. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Granada, Spain: Springer; 2018.
24. Chen C, Qin C, Qiu H, Tarroni G, Duan J, Bai W, *et al*. Deep learning for cardiac image segmentation: A review. Front Cardiovasc Med 2020;7:25.
25. Zhang D, Dongqing, Icke I, Dogdas B, Parimal S,

Sampath S, *et al*. A multi-level convolutional LSTM model for the segmentation of left ventricle myocardium in infarcted porcine cine MR images. In: 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018). Washington, DC, USA: IEEE; 2018.

26. Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Munich, Germany: Springer; 2015.

27. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Montreal, Canada; 2016.

28. Ji Q, Huang J, He W, Sun Y. Optimized deep convolutional neural networks for identification of macular diseases from optical coherence tomography images. Algorithms 2019;12:51.

29. Radau P, Lu Y, Connelly K, Paul G, Dick AJ, Wright G. Evaluation framework for algorithms segmenting short axis cardiac MRI. The MIDAS Journal-Cardiac MR Left Ventricle Segmentation Challenge 49 (2009).

30. Abdelmaguid E, Huang J, Kenchareddy S, Singla D, Wilke L, Nguyen MH, et al. Left ventricle segmentation and volume estimation on cardiac mri using deep learning. arXiv;2018:1809.06247.

31. Chen A, Antong, Zhou T, Icke I, Parimal S, Dogdas B, *et al*. Transfer learning for the fully automatic segmentation of left ventricle myocardium in porcine cardiac cine MR images. In: International Workshop on Statistical Atlases and Computational Models of the Heart. Quebec City, Canada: Springer; 2017.

32. Chen M, Fang L, Liu H. FR-NET: Focal loss constrained deep residual networks for segmentation of cardiac MRI. In: 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019). Venice, Italy: IEEE; 2019.

33. Taha AA, Hanbury A. Metrics for evaluating 3D medical image segmentation: Analysis, selection, and tool. BMC Med Imaging 2015;15:29.

34. Zhang J, Du J, Liu H, Hou X, Zhao Y, Ding M. LU-NET: An Improved U-Net for ventricular segmentation. IEEE Access 2019;7:92539-46.

35. Kaus MR, von Berg J, Weese J, Niessen W, Pekar V. Automated segmentation of the left ventricle in cardiac MRI. Med Image Anal 2004;8:245-54.