IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received May 22, 2019, accepted June 14, 2019, date of publication June 26, 2019, date of current version July 26, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2925060

LU-NET: An Improved U-Net for Ventricular Segmentation

JUN ZHANG, JIAZHUO DU, HONGPU LIU, XIANGDAN HOU, YIHAO ZHAO, AND MENGYUAN DING

School of Artificial Intelligence, Hebei University of Technology, Tianjin 300401, China
Hebei Provincial Key Laboratory of Big Data Computing, Tianjin 300401, China

Corresponding author: Hongpu Liu (liuii@scse.hebut.edu.cn)

This work was supported in part by the Tianjin Natural Science Foundation under Grant 16JCYBJC15600, in part by the Key Project of National Natural Science Fundation under Grant U1813222, and in part by the Smart Robot Key Project of National Key Research and Development Plan under Grant 2018YFB1306900.

ABSTRACT In order to solve the problem of low accuracy of the U-Net in cardiac ventricular segmentation, we propose an improved U-net named LU-Net by the following three methods. First, in order to improve the efficiency and effectiveness of extracting the features of the original image, we combine U-net with SE-Net model. This model reweights the channels of the feature map, which can give higher weight to the useful information and lower weight to the invalid information. Second, in order to alleviate the extent of losing the pixel-location information when using the encoder to dawn sample, we combine multi-scale input with U-net's encoder. Third, in order to solve the problem of low accuracy in traditional U-net, we replace the transposed convolution layer, used by the traditional U-Net's encoder during upsampling, with an unsampling layer. During the process of unsampling, it can put pixels to their original location using the pixel-location information reserved by the encoder during the sampling process, which can reduce errors caused by losing pixel-location information. Besides, using the unsampling layer during unsampling can also avoid producing checkerboard artifacts during transposed convolution and improve the segmentation accuracy. To verify the effectiveness of LU-Net, we apply it to the ACDC Stacom 2017 dataset. The experimental results show that the evaluation criteria of prediction results are 92.4%, 86.4%, and 92.5% on Dice coefficient, Jaccard similarity coefficient, and F1-ccore respectively, which are better than U-Net, SegNet, and IU-Net and remarkably better than the traditional neural convolution network model, FCN8s.

INDEX TERMS Cardiac ventricular segmentation, U-Net, reweight, multi-scale input, unsampling.

I. INTRODUCTION

Medical images play an extremely important role in disease diagnosis. In recent years, with the continuous development of medical imaging technology and computing technology, image processing technology for medical images has gradually become an important research field, in which medical image segmentation is a research direction with high clinical application value [1]. Medical image segmentation is a complex and essential step in the field of medical image processing and analysis. Its purpose is to segment the parts of the medical image with certain special meanings and extract relevant features to provide disease diagnosis with a basis and help doctors to diagnosis more accurately [2].

The associate editor coordinating the review of this manuscript and approving it for publication was Haruna Chiroma.

Imaging methods commonly used in cardiovascular diseases include computed tomography magnetic resonance imaging, echocardiography, etc. The current gold standard is based on non-invasive MRI (cine MRI) calculation of clinical indicators, such as ejection fraction, ventricular volume, myocardial mass, etc., to analyze and infer cardiac function. Calculating clinical indicators relies on the doctor's precise manual segmentation of cardiac MRI images, which is time consuming and laborious. Moreover, different doctors may not have the same segmentation results on the same MRI image and the same doctor may also have different results in the two segmentation processes [3]–[5]. Therefore, the fast and accurate automatic segmentation of cardiac MRI images will help doctors diagnose cardiovascular disease accurately.

In recent years, with the development of deep learning, medical image processing models based on convolutional

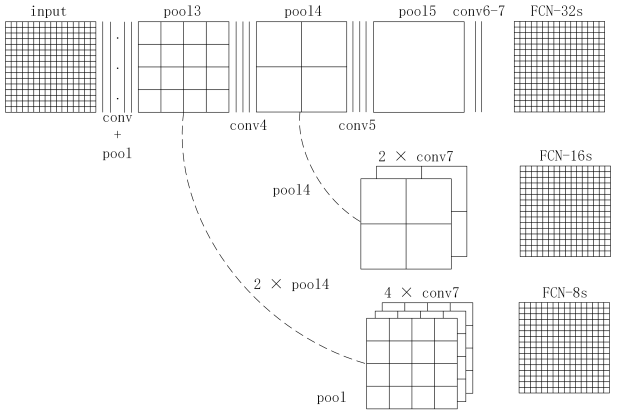


FIGURE 1. Full convolutional neural network learn to combine coarse, high layer information with fine, low layer information. Layers are shown as grids that reveal relative spatial coarseness.FCN-32s upsamples stride 32 predictions back to pixels in a single step.FCN-16s combines predictions from both the final layer and the pool4 layer,lets our net predict finer details, while retaining high-level semantic information. FCN-8s combines the predictions of the last layer and the pool4 layer, and combines the results with the pool3 layer prediction to further improve the accuracy.

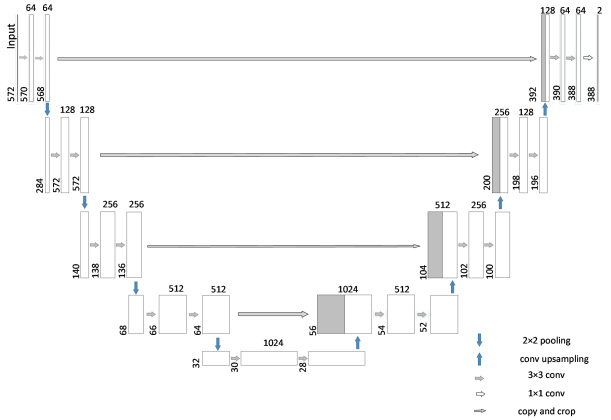


FIGURE 2. Original U-net structure diagram. The arrow pointing to the right means convolution, the downward arrow means pooling, the upward arrow means upconvolution.

neural networks have been widely used in different medical fields and have achieved quite good results. Medical image segmentation is the most common research direction among them and convolutional neural network is one of the most common automatic segmentation models. Compared with traditional algorithms, image segmentation methods based on convolutional neural networks, such as FCN [6], U-Net [7], SegNet [8], IU-Net [9] etc., have achieved better performance in accuracy. The convolutional neural network algorithm is characterized by the ability to perform multiple nonlinear transformations on images by convolution kernels to extract higher-level abstract features without the need for artificial feature design for different scenes. In addition, in the image segmentation process, the convolutional neural network enhances the robustness of feature extraction through techniques such as local join and weight sharing.

II. RELATED WORKS

This section discusses the deep learning models associated with LU-Net, including FCN, U-Net, SegNet and SE-Net.

Fully Convolutional Network (FCN) [6], shown in Fig. 1, is an end-to-end network model that classifies images at the pixel level to solve semantic level image classification problems.The FCN uses the transposed convolution at the end of the network to upsample the feature map generated by the last convolution to the resolution of the original picture and segment the image by producing a prediction for each pixel. Palit *et al.* [10] proposed the use of improved FCN for bioglial cell segmentation, greatly reducing image segmentation time and significantly improving accuracy. Li *et al.* [11] proposed combining the traditional watershed algorithm with FCN to eliminate useless non-edge pixels and improve the efficiency of segmentation.

Although the FCN model is simple and efficient, there are still some problems. For example, the FCN uses the convolution pooling operation multiple times during the downsampling process, which results in loss of the picture details and the unsatisfied segmentation result. Many subsequent image segmentation models are based on FCN, such as U-Net, SegNet.

Ronneberger *et al.* [7] proposes an improved full convolutional neural network structure U-Net, which was originally used for cell segmentation. Its structure is shown in Fig. 2. U-Net consists of the encoder on the left and the decoder on the right. The encoder is a typical convolutional neural network structure, which is composed of convolutional layer, batch normalization layers and pooling layers. The decoder consists of convolutional layers and deconvolution layers. The U-Net encoder is downsampled in total of 4 times. Symmetrically, the decoder also performs corresponding upsampling 4 times to restore the feature map obtained by the encoder to a picture with the same size of original picture. U-Net adopts the idea of cascading higher-order features and shallow features. When the right-side decoder performs upsampling, the feature maps on the encoder side of the corresponding level are concatenated by matrix cascading. This operation is beneficial to supplement the missing pixel position information during the convolution process, thereby improving the segmentation accuracy. Yang *et al.* [12] proposed to segment the short-axis MRI of the heart by coupling a U-shaped convolutional neural network and a deformation model. Cong and Zhang [13] explored the relationship between the image segmentation mechanism of deep neural networks and the channel in the network model. Based on the conclusions obtained, the hyperparameters in the U-shaped convolutional neural network were artificially designed. Zheng *et al.* [14] proposed a heart segmentation algorithm combining U-Net with spatial propagation. These methods have high accuracy and robustness for heart slices at different locations and different periods.

Badrinarayanan *et al.* [8] proposes a kind of U-Net network model SegNet, which is similar in structure to U-Net. It also

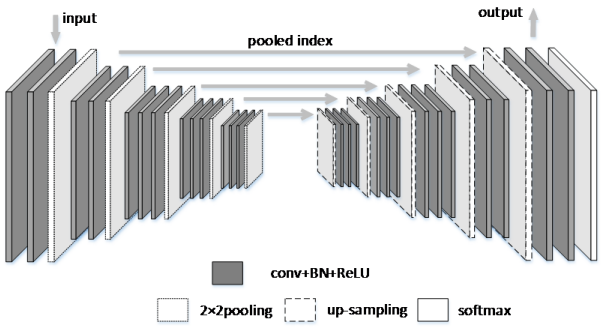


FIGURE 3. An illustration of the SegNet architecture. There are no fully connected layers and hence it is only convolutional. A decoder upsamples its input using the transferred pool indices from its encoder to produce a sparse feature map(s). It then performs convolution with a trainable filter bank to densify the feature map. The final decoder output feature maps are fed to a soft-max classifier for pixel-wise classification.

contains an encoder and a decoder. Its structure is shown in Fig.3. The left encoder uses the first 13 layers of the VGG16 model for downsampling and the right decoder maps the encoded feature map to the input resolution. Different from U-Net, the SegNet encoder retains the position information of the pixel in the upper-level feature map during the downsampling process, that is, the pooled index. The decoder combines the pooling during the upsampling process. The index is used to perform nonlinear upsampling, which eliminates the need for the decoder to learn the upsampling process. Therefore, compared with U-Net, SegNet can not only accurately restore the position of pixels, improve the segmentation accuracy, but also reduce the training parameters and shorten the training time. Tang *et al.* [15] applied SegNet to gland medical image segmentation to improve the clinical diagnosis efficiency of colon cancer. Kumar *et al.* [16] combined U-Net and SegNet to propose a SegNet-based U-Net hopping connection to combine fine multi-scale information for better tissue boundary recognition.

Hu *et al.* [17] proposes a network structure Squeeze-and-Excitation Networks (SE-Net) that considers the relationship between feature channels. The module structure is shown in Fig.4. Roy *et al.* [18] integrated the SE-Net module into multiple existing image segmentation network structures. The segmentation capabilities of different network structures have been improved to varying degrees. SE-Net adopts the ''feature recalibration'' strategy, which is to learn the importance of each channel in the feature map through network training, to enhance the useful features according to the importance level and suppress the features that are not useful. In SE-Net, it mainly includes three operations: Squeeze, Excitation and Reweight. The Squeeze operation uses the global pooling method to convert the feature map into a feature vector. The Excitation operation uses two fully connected layers to form a gating mechanism for each feature channel. Finally the Reweight operation weights the weight of the Excitation output and the feature map.

When the FCN8s, U-Net, SegNet and other network models use the convolution kernel to extract features in the

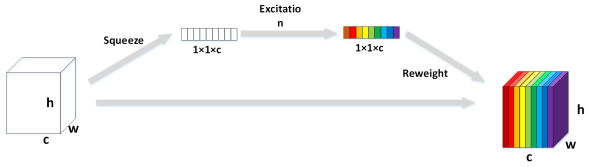


FIGURE 4. The squeeze-excitation block is a computational unit that can be constructed for any given transformation. The model calculates the feature map weights by squeeze and excitation to obtain the feature vector. Finally, the weighting operation of the two is performed.

encoder, the image pixel position information will be lost to some extent, which leads to inaccurate pixel restoration position in the decoder and the inaccurate segmentation boundary.

III. LU-NET MODEL

In order to improve the accuracy of neural network image segmentation, the following two aspects should be emphasized. Firstly, the encoder needs to learn how to extract more efficient abstract information during the downsampling process to make the pixel classification more accurate. Secondly, during the upsampling process, when the decoder expands the high-order feature map to the original image resolution, it is necessary to more accurately restore the pixel points to the corresponding positions of the original image.

In order to improve the efficiency of U-Net in extracting features during the downsampling process, We integrate the SE-Net module into the U-Net encoder. The method performs weighting operations on the channel, emphasizes effective information and suppresses invalid information, which not only improves the validity of the extracted features, but also reduces the amount of calculation.

During the encoding and decoding process of U-Net, convolution and pooling operations will lose some valuable details, resulting in inaccurate segmentation boundaries. In response to this problem, we make improvements to the U-Net model. Firstly, the model generates feature maps of different scales at the input end of the U-Net encoder, and then concatenates them with the feature maps of the corresponding scales generated by the pooling layer in the encoder to realize multi-scale input. Secondly, in the upsampling process of the U-Net decoder, the pixel is restored according to the pixel pooling information retained by the downsampling, which can reduce the error of the pixel position in the upsampled feature map and compensate for the loss of pixel position information caused by the convolution and pooling operations.

A. FUSION OF U-NET AND SE-NET MODULES

The main function of the SE-Net module is to improve the effectiveness of feature extraction. Therefore,the fusion of U-Net and SE-Net is implemented by performing a feature weighting operation before each pooling operation in the U-Net downsampling process. As shown in Fig.5, Two 3×3 convolutions extract the input features, obtain feature maps and then pass the feature maps to the SE-Net module. Firstly, global pooling is performed to obtain the feature vector

FIGURE 5. In the encoder, U-Net and SE-Net are combined by adding SE-Net modules in front of each pool layer.

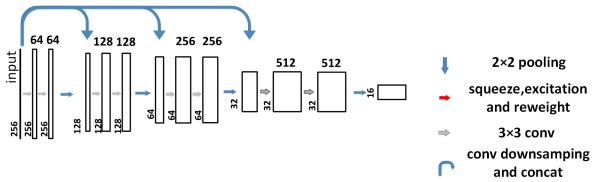


FIGURE 6. In the encoder, the input is convolved and downsampled to form feature maps of different scales, and then merged into corresponding layers respectively.

corresponding to the feature maps, and then the feature vector is input to two consecutive fully connected layers to obtain a set of weight values between 0 and 1. Finally, the weight values are multiplied with the feature maps to obtain new feature maps with weight information.

B. MULTI-SCALE INPUT

Multi-scale input can compensate for missing pixel location information for convolution and pooling operations. We concatenate the feature maps after each pooling operation in the U-Net downsampling process. The structure is shown in Fig.6. The convolution layer with a kernel size of 2 × 2 and a step size of 2 is subjected to downsampling operation to obtain feature maps of different scales. The feature maps of different scales are combined with the output of the previous pooled layer to generate the input of the corresponding layer.

C. IMPROVED UPSAMPLING METHOD

The original U-Net upsampling process uses deconvolution to magnify the feature map resolution. In the deconvolution process, some pixels are likely to cause uneven overlap, that is, some regions have more abstract results than other regions. Especially when the size of the convolution kernel cannot be divisible by the step size, the deconvolution will have uneven overlap. When calculated in two-dimensional space, uneven pixel overlap occurs on the horizontal axis and the vertical axis, causing chessboard artifacts and affecting the final segmentation result. We replace the deconvolution in the original U-Net with a combined convolution of 3 × 3 and 1 × 1. In addition, the network model is upsampled by the inverse unpooling operation. Structure shown in Fig. 7. The improved network records the position information of the pixel when the encoder is downsampling. When the upsampling is performed, the unpooling operation directly returns the pixel data to the original position, which compensates for the loss of detail caused by the pooling layer in the encoder reducing the feature map.

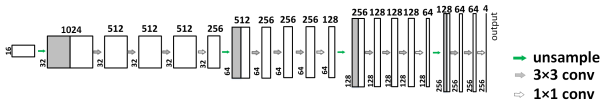


FIGURE 7. In the decoder, we replace the deconvolution layer in the original U-Net with the combined convolutional layer of 3 × 3 and 1 × 1, and use unsample to restore the feature map size.

In addition, in order to further alleviate the problem of gradient disappearance, we use batch normalization plus ReLU activation processing for the convolutional layer in the network. In summary, the overall structure of the model is shown in Fig.8.

IV. EXPERIMENT SETTINGS

A. DATASET

The dataset uses the ACDC Stacom 2017 Challenge dataset, which is short-axis cine MRI of 150 patients obtained from an MR scanner using different magnetic field strengths (1.5T and 3.0T) at the University Hospital of Dijon. The cine MRI of 100 patients was tagged by two medical experts. We divided the 100 patients' MRI with pixel annotations in the ACDC Stacom 2017 Challenge dataset into a training set and validation set in a 7:3 ratio. In addition, the cine MRI of remaining 50 patients is used as a test set.

1) DATA ENHANCEMENT

In order to make the training samples cover more target features, avoid over-fitting caused by insufficient training samples and improve the generalization ability of the algorithm, we have enhanced the data of the training set samples, including image flip, brightness and contrast random transformation. The above method is randomly added to each picture in the training set.

2) RESIZE

The MRI images in the dataset are collected by different medical imaging devices, so the image resolution is inconsistent. We use bilinear interpolation to unify image resolution to 256 × 256 [19].

B. CONFIGURATIONS

The deep learning framework we use is pytorch [20] 0.4.0. The experiment uses GPU acceleration, the GPU is configured as a single GeForce GTX 1080TI, CUDA is 9.0 and cuDNN is 7.1. We did not load the pre-training model during the training process. The convolution kernel initialized the parameters using the 0-1 random normal distribution method. The model optimizer uses Adam [21], the initial learning rate is 0.001, the loss function uses CrossEntropyLoss, the batch input size is 8 and the training round is 100 rounds. The learning rate attenuation method is used in the training process, the initial learning rate is used in 1 to 80 rounds, 10% of the initial learning rate in 81-90 rounds and 1% of the initial learning rate in 91-100 rounds.

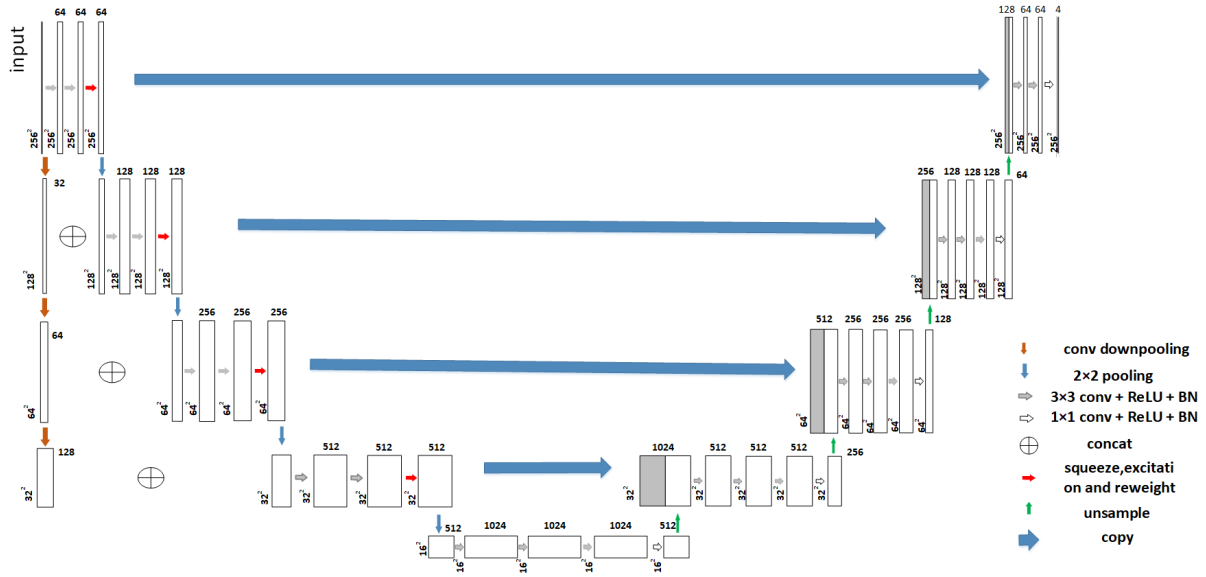


FIGURE 8. LU-Net model.

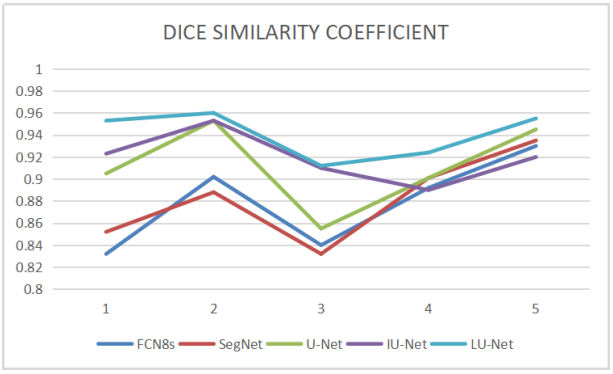


FIGURE 9. Average dice similarity coefficients of different network models on five test samples.

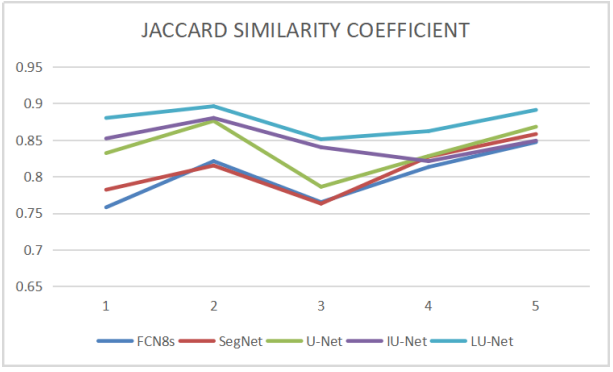


FIGURE 10. Average Jaccard similarity coefficients of different network models on five test samples.

C. EVALUATION CRITERIA

1) DICE SIMILARITY COEFFICIENT

Dice Similarity Coefficient (DSC) can be used to measure the coincidence degree between the actual segmentation result and the theoretical segmentation result. Its value ranges from 0 to 1. The closer to 1, the better the pixel classification effect of the model. The definition is as follows:

$$DSC = \frac{2\,|A \cap B|}{|A| + |B|} \tag{1}$$

where A and B respectively represent the theoretical segmentation results of expert annotations and the actual segmentation results of the proposed method.

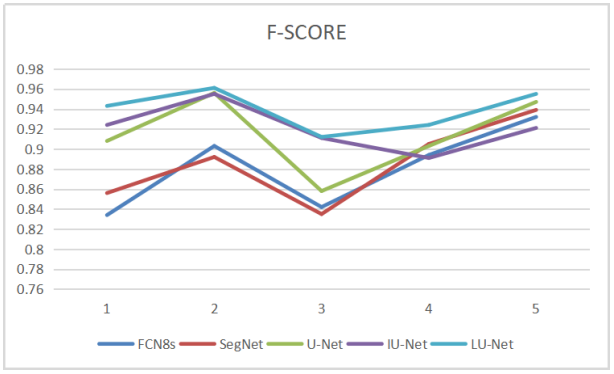


FIGURE 11. Average F-Score of different network models on five test samples.

2) JACCARD SIMILARITY COEFFICIENT

The Jaccard similarity coefficient can be used to measure the similarity between the actual segmentation result and the theoretical segmentation result. Its value ranges from 0 to 1. The closer to 1, the better the pixel classification effect of the model. The definition is as follows:

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|} \tag{2}$$

where A and B respectively represent the theoretical segmentation results of expert annotations and the actual segmentation results of the proposed method.

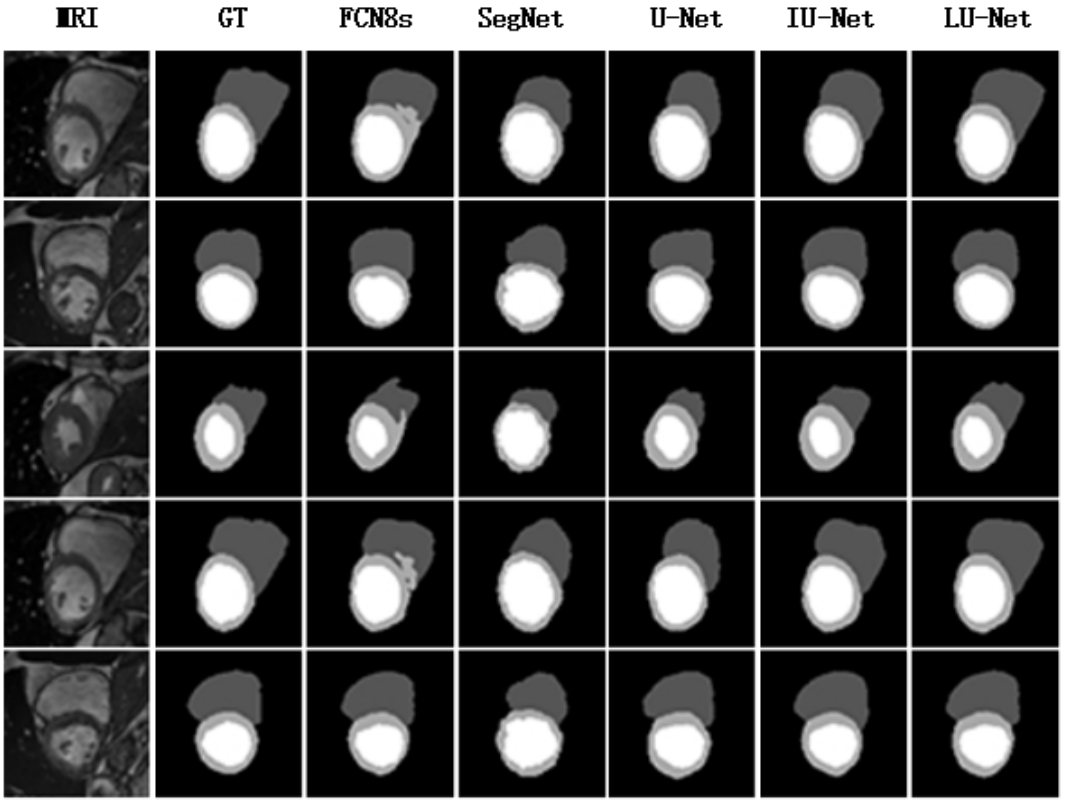


FIGURE 12. Different model test results.

3) F-SCORE

F-scores can be used to measure model accuracy. The F1 score takes into account the accuracy and recall rate of the pixel classification model and can be regarded as a weighted average of the model accuracy and recall rate. The value ranges from 0 to 1. The closer to 1, the better the pixel classification effect of the model. The definition is as follows:

$$F_\beta = \frac{(1+\beta^2)PR}{\beta^2(P+R)} \tag{3}$$

where β is the parameter, we set $\beta = 1$. P is the accuracy rate,

$$P = \frac{A \cap B}{B} \tag{4}$$

R is the recall rate,

$$R = \frac{A \cap B}{A} \tag{5}$$

A and B respectively represent the theoretical segmentation result of the expert labeling and the actual segmentation result of the method.

D. EXPERIMENTAL RESULTS AND DISCUSSION

To verify the effectiveness of the improved method in this paper, we used four network models for comparative experiments. The comparison network models are FCN8s, SegNet, U-Net, and IU-Net.

TABLE 1. Comparison of test results of different models.

Model	Dice	Jaccard	F1-Score
FCN8s	0.892	0.813	0.894
SegNet	0.901	0.827	0.905
U-Net	0.905	0.832	0.908
IU-Net	0.910	0.840	0.911
LU-Net	0.924	0.863	0.925

The above four networks were trained on the ACDC Stacom 2017 Challenge data set and ventricular MRI image segmentation verification was performed. The test results of different network models are shown in TABLE 1. In the case of the Dice coefficient as the evaluation standard, the method of this paper is 3.2%, 2.3%, 1.9% and 1.4% higher than FCN8s, SegNet, U-Net and IU-Net respectively. In the case of the Jaccard similarity coefficient as the evaluation standard, the method of this paper increased by 5.0%, 3.6%, 3.1% and 2.3% compared with FCN8s, SegNet, U-Net and IU-Net respectively. In the case of F1-Score as the evaluation standard, the method of this paper increased by 3.1%, 2.0%, 1.7% and 1.4% compared with FCN8s, SegNet, U-Net and IU-Net respectively.

The above experimental results were obtained on the entire complete dataset. In order to more intuitively demonstrate the image segmentation performance of different network models on individual samples, we used a short-axis cine MRI of five patients to segment. Figure 9-11 show the quantitative evaluation criteria for the segmentation results of the

various network models on these five test samples. As can be seen from the figure, LU-Net has the lowest Dice coefficient of 0.912 in the five examples, the highest is 0.960, the lowest Jaccard coefficient is 0.851, the highest is 0.896, the lowest F-Score is 0.912, and the highest is 0.961. The three evaluation indicators obtained by LU-Net are significantly better than the other four network models.

Figure 12 shows the comparison of the optimal predictions for different models. Experiments show that the segmentation results of IU-Net and our method are significantly better than those of FCN8s, SegNet and U-Net. Compared with FCN8s, SegNet, U-Net and IU-Net, LU-Net is more sensitive to edge pixels and the segmentation contour is clearer. Comparing the segmentation results in the third column and the seventh column in Fig. 9, it can be seen that in the segmentation tasks of the left ventricular cavity and the left ventricular wall, the edge of the segmentation results obtained by FCN8s, SegNet and U-Net is not high. Experiments show that during the encoding and decoding process, FCN8s, SegNet and U-Net will lose valuable detail information, which will result in poor recognition of the boundary of adjacent regions. The multi-scale input and anti-pooling layer replacement transposed convolution adopted in LU-Net enhances the left ventricular and left ventricular wall edge information by making up for the missing pixel position information of convolution and pooling operations, which improves the model to the left ventricle. The edge recognition ability of the cavity and left ventricular wall. In the segmentation task of the right ventricle, because the color of the right ventricle is close to the background color, IU-Net cannot make accurate segmentation. However, the SE-Net module used in the LU-Net encoder can extract more effective information, which makes the segmentation effect significantly better.

E. CONCLUSION

We propose an improved U-shaped convolutional network model and apply it to the ventricular segmentation problem in practical medical image analysis. The FCN8s, SegNet, U-Net and IU-Net models lose image detail information during continuous upsampling and downsampling, making the segmentation effect of the model on the edge unsatisfactory. The improved U-shaped convolutional neural network improves the feature extraction efficiency by multi-scale input and fusion SE-Net module on the encoder side. The anti-pooling method is used for upsampling at the decoder end to obtain more accurate feature information and pixel position information. In addition, the use of batch normalization and ReLU activation functions can further optimize network weights and improve model performance. The final experiment proves that the method has 92.4%, 86.3% and 92.5% of the Dice coefficient, Jaccard detailed coefficient and F1-Score, respectively, which is better than SegNet, U-Net and IU-Net, which is obviously better than FCN.

REFERENCES

[1] P. K. Pan, Y. Wang, and Y. Luo, "Automatic segmentation of nasopharyngeal neoplasm in MR images based on U-Net model," *J. Comput. Appl.*, Dec. 2018.

[2] Y. Wan, Y. Wu, P. Tao, and Y. Yao, "Left ventricle segmentation in transesophageal echocardiography based on supervised descent method," *J. Comput. Appl.*, vol. 38, pp. 545–549, Feb. 2018.

[3] C. Petitjean and J.-N. Dacher, "A review of segmentation methods in short axis cardiac MR images," *Med. Image Anal.*, vol. 15, pp. 169–184, Apr. 2011.

[4] C. A. Miller, P. Jordan, A. Borg, R. Argyle, D. Clark, K. Pearce, and M. Schmitt, "Quantification of left ventricular indices from SSFP cine imaging: Impact of real-world variability in analysis methodology and utility of geometric modeling," *J. Magn. Reson. Imag.*, vol. 37, pp. 1213–1222, Nov. 2013.

[5] V. Tavakoli and A. A. Amini, "A survey of shaped-based registration and segmentation techniques for cardiac images," *Comput. Vis. Image Understand.*, vol. 117, no. 9, pp. 966–989, Sep. 2013.

[6] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 3431–3440.

[7] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, Oct. 2015, pp. 234–241.

[8] V. Badrinarayanan, A. Kendall, and R. Cipolla, "SegNet: A deep convolutional encoder-decoder architecture for image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 12, pp. 2481–2495, Dec. 2017.

[9] G. Tong, Y. Li, H. Chen, Q. Zhang, and H. Jiang, "Improved U-NET network for pulmonary nodules segmentation," *Optik*, vol. 174, pp. 460–469, Dec. 2018.

[10] I. Palit, L. Yang, Y. Ma, D. Chen, M. Niemier, J. Xiong, and X. S. Hu, "Biomedical image segmentation using fully convolutional networks on TrueNorth," in *Proc. IEEE 31st Int. Symp. Comput.-Based Med. Syst. (CBMS)*, Jul. 2018, pp. 375–380.

[11] K. Li, G. Ding, and H. Wang, "L-FCN: A lightweight fully convolutional network for biomedical semantic segmentation," in *Proc. IEEE Int. Conf. Bioinf. Biomed. (BIBM)*, Jan. 2018, pp. 2363–2367.

[12] D. Yang, Q. Huang, L. Axel, and D. Metaxas, "Multi-component deformable models coupled with 2D-3D U-Net for automated probabilistic segmentation of cardiac walls and blood," in *Proc. IEEE 15th Int. Symp. Biomed. Imag. (ISBI)*, May 2018, pp. 479–483.

[13] C. Cong and H. Zhang, "Invert-U-Net DNN segmentation model for MRI cardiac left ventricle segmentation," *J. Eng.*, vol. 2018, pp. 1463–1467, Nov. 2018.

[14] Q. Zheng, N. Delingette, H. Duchateau, and N. Ayache, "3-D consistent and robust segmentation of cardiac images by deep learning with spatial propagation," *IEEE Trans. Med. Imag.*, vol. 37, no. 9, pp. 2137–2148, Sep. 2018.

[15] J. Tang, J. Li, and X. Xu, "Segnet-based gland segmentation from colon cancer histology images," in *Proc. IEEE 33rd Youth Academic Annu. Conf. Chin. Assoc. Automat. (YAC)*, May 2018, pp. 1078–1082.

[16] P. Kumar, P. Nagar, C. Arora, and A. Gupta, "U-SegNet: Fully convolutional neural network based automated brain tissue segmentation tool," in *Proc. IEEE 25th Int. Conf. Image Process. (ICIP)*, Oct. 2018, pp. 3503–3507.

[17] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2018, pp. 7132–7141.

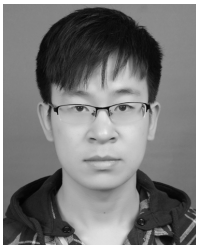
[18] A. G. Roy, N. Navab, and C. Wachinger, "Recalibrating fully convolutional networks with spatial and channel 'squeeze and excitation' blocks," *IEEE Trans. Med. Imag.*, vol. 38, no. 2, pp. 540–549, Feb. 2019.

[19] W. Xue, A. Islam, M. Bhaduri, and S. Li, "Direct multitype cardiac indices estimation via joint representation and regression learning," *IEEE Trans. Med. Imag.*, vol. 36, no. 10, pp. 2057–2067, Oct. 2017.

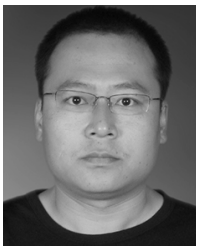
[20] N. Ketkar. *Deep Learning With Python*. [Online]. Available: https://link.springer.com/chapter/10.1007/978-1-4842-2766-4_12

[21] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*. [Online]. Available: https://arxiv.org/abs/1412.6980

JUN ZHANG received the master's and Ph.D. degrees from the Hebei University of Technology (HEBUT), in 2005 and 2011, respectively, where he is currently an Associate Professor with the School of Artificial Intelligence. His main researches include embedded systems and intelligent computing.

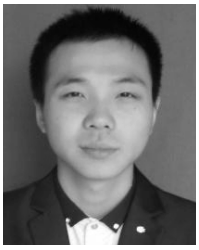


JIAZHUO DU received the B.S. degree from the School of Computer Science and Technology, Tangshan University (TSC), in 2017. He is currently pursuing the master's degree with the School of Artificial Intelligence, Hebei University of Technology (HEBUT). He works closely with Prof. X. Hou, and focuses on medical image segmentation using deep learning methods.



HONGPU LIU received the bachelor's degree in information systems and information management from Heilongjiang University (HLJU), in 2015. He is currently pursuing the master's degree with the School of Artificial Intelligence, Hebei University of Technology (HEBUT). His research interest includes digital image processing.

XIANGDAN HOU received the Ph.D. degree in theory and new technology of electrical engineering from the Hebei University of Technology (HEBUT), in 2007, where she is currently a Professor with the School of Artificial Intelligence. Her research interests include digital image processing and intelligent algorithm.



YIHAO ZHAO received the B.S. degree from the School of Artificial Intelligence, Hebei University of Technology (HEBUT), in 2017, where he is currently pursuing the master's degree. His research interest includes medical image analysis using deep learning methods.

MENGYUAN DING received the B.S. degree in information science and engineering majored in information engineering from Hebei North University (HBNU), in 2018. She is currently pursuing the master's degree with the School of Artificial Intelligence, Hebei University of Technology (HEBUT). Her research interest includes semantic segmentation of image processing using deep learning methods.