Hindawi
Scientific Programming
Volume 2021, Article ID 8666693, 8 pages
https://doi.org/10.1155/2021/8666693

Research Article

An Improved 2D U-Net Model Integrated Squeeze-and-Excitation Layer for Prostate Cancer Segmentation

Bingshuai Liu,[1] Jiawei Zheng,[1] Hongwei Zhang,[1] Peijie Chen,[2] Shipeng Li,[3] and Yuexian Wen[2]

[1] *School of Informatics Xiamen University, Xiamen University, Xiamen 361000, Fujian, China*
[2] *Zhongshan Hospital Affiliated of Xiamen University, Xiamen 361004, Fujian, China*
[3] *The Third Clinical Medical College of Fujian Medicial University, FuZhou 350122, Fujian, China*

Correspondence should be addressed to Peijie Chen; pajachen@xmu.edu.cn and Yuexian Wen; wyx@xmzsh.com

Received 23 September 2021; Revised 12 October 2021; Accepted 3 November 2021; Published 21 December 2021

Academic Editor: Tongguang Ni

In this paper, we proposed an improved 2D U-Net model integrated squeeze-and-excitation layer for prostate cancer segmentation. The proposed model combined a more complex 2D U-Net model and squeeze-and-excitation technique. The model consisted of an encoder stage and a decoder stage. The encoder stage aims to extract features of the input, which contains CONV blocks, SE layers, and max-pooling layers for improving the feature extraction capability of the model. The decoder aims to map the extracted features to the original image with CONV blocks, SE layers, and upsampling layers. The SE layer is implemented to learn more global and local features. Experiments on the public dataset PROMISE12 have demonstrated that the proposed model could achieve state-of-the-art segmentation performance compared with other traditional methods.

1. Introduction

Prostate cancer has become a high incidence cancer among men. Early medical detection and diagnosis of cancers could substantially improve the cure rate among patients. Currently, radiation therapy which uses medical ionizing radiation to kill cancer cells is a very common procedure to treat prostate cancers [1]. However, the worst disadvantage of the procedure is that the radiation may damage the cells of surrounding tissue when it kills prostate cancer. For the sake of raising the accuracy of radiation therapy and reducing the side effect in surrounding tissue such as bladder and rectum, more delicate prostate cancer diagnosis and more accurate prostate cancer localization methods are required.

At present, there are two main types of artificial and automatic to achieve prostate cancer segmentation on MRI (magnetic resonance imaging) [2]. The former, however, is gradually being displaced by the latter. Manual segment by radiologists is a time resuming work, and there are subjective differences among radiologists' diagnoses. For example, a radiologist may get a segmentation image differently, and different radiologists may obtain to different results on the same image.

Automatic segmentation methods can help radiologists achieve prostate cancer segmentation result faster with higher accuracy. There are two main methods usually utilized: atlas-based methods and deformable model-based methods [3]. As for the atlas-based method, training images accompanied with their corresponding manual labels are mixed together; then, through nonrigid registration (NRR), a reference image named as an Atlas and labeled Atlas is formed [3]. The Atlas is a trained image which represents the prostate and its surrounding tissue while its corresponding labeled Atlas shows the probability of a voxel being a part of the prostate [2, 3]. In model-based methods, the model can use the atlas-based segmentation for its initialization and use the grey-level information of the image to be deformed to match the boundaries of the prostate [4]. Then, a distance metric is utilized, usually the Mahalanobis distance to match the contour of the feature model with the contour extracted from the case images [3]. Both methods can be time-consuming since they require a good initialization to display better effects on prostate cancer segmentation [2].

Currently, the deep learning-based methods have made a remarkable performance in medical image segmentation. There are some research studies based on deep learning methods that have obtained accurate results in the segmentation, which prove that a well-trained deep learning model can improve the accuracy and velocity in medical image segmentation [5–7]. Karimi et al. put forward a two-step segmentation method which contains two convolutional neural networks (CNNs), where the first CNN determines a prostate bounding box and the second CNN provides accurate delineation of the prostate boundary [5]. Guo et al. designed a deformable MR prostate segmentation method by integrating deep feature learning with sparse patch matching [6]. Cheng et al. presented a supervised learning framework which merges the atlas-based active appearance model (AAM) and support vector machines (SVM) to achieve a high segmentation result of the prostate boundary [7]. However, all the methods mentioned above have a common disadvantage in which it is difficult to achieve a pixelwise level segmentation with high accuracy.

Fully convolutional networks (FCN) proposed by Long et al., where the last fully connected layer of regular CNN is replaced with a convolution layer, can obtain the classification information of every pixel; therefore, it solves the problem of pixelwise level segmentation [8]. Roneneberger et al. made a further optimisation based on FCN and presented a symmetric structure called U-Net, which is a regular CNN with an upsampling operation, where deconvolutions are utilized to increase the size of feature maps [9]. At present, FCN or U-Net becomes the most popular backbone network in the medical image segmentation field. There are many new structures derived from the FCN or U-Net model after that time. For example, Zhou et al. modified the skip connection between encoder layers and decoder layers based on U-Net and then designed a new model called U-Net++ [10] and Milletari et al. put forward a variant model named as V-Net which can realize 3D segmentation [11]. However, these methods have a common disadvantage that the similar low-level features are extracted by the model repeatedly which results in unnecessary waste of computational resources.

In order to solve the problems above, in this paper, we proposed a more effective model, which utilizes the U-Net as the backbone of our network, and a squeeze-and-excitation layer is added to every convolution operation to select the emphasize the features which are contributed to the prostate cancer segmentation.

2. Related Works

There are many research studies [5, 6, 10–12] took the deep learning method the same with as to achieve prostate cancer segmentation on MRI because it comes to more remarkable performance in the field compared to the traditional method. The idea of making an optimisation based on U-net has attracted much attention in recent years; many related research studies have made good results. For examples, the U-Net++ was proposed by Zhou et al. which modifies the skip connection between the encoder and the decoder to achieve an optimisation [10], and the 3D U-Net called V-Net was put forward by Milletari et al. based on 2D U-Net [11].

The application of the SE layer took much inspiration from the channel attention utilized in a biattention adversarial network designed by Zhang et al. [12], which proves to have a positive effect on improving model performance.

3. Background

3.1. Structure. Our proposed model refers to the U-Net model and fully convolutional network (FCN), which divide the model into the encoder stage and the decoder stage (autoencoder). The overall structure of our model can be seen in Figure 1. The encoder (also called the contraction path) is used to capture the context in the image, and the decoder (also called the symmetric expanding path) is used to enable precise localization. U-Net and FCN are actually very similar and both of them are published in 2015; however, U-Net is a little bit later than FCN. However, there are still some differences between them. Compared with FCN, U-Net is completely symmetrical whose encoder stage and decoder stage are similar while FCN's decoder stage structure is simpler which only uses one deconvolution operation and no more convolution structures such as U-Net. The second difference is about skip connection, FCN uses summation operation while U-Net uses concatenation operation.

3.2. The Activation Layer. An activation layer is always used after a convolution layer to choose if a particular neuron should be activated or not to be activated in U-Net. There are two most common activation functions used in U-Net. The first is rectified linear unit (ReLU) and the second is leaky rectified linear unit (Leaky ReLU). We are going to introduce these two functions in this section.

The ReLU formula is as follows:

$$f(x) = \begin{cases} 0, & \text{if } x \leq 0, \\ x, & \text{if } x > 0. \end{cases} \tag{1}$$

For the Leaky ReLU,

$$f(x) = \begin{cases} x, & \text{if } x < 0, \\ cx, & \text{if } x \geq 0. \end{cases} \tag{2}$$

Compared to the traditional activation function, such as logistic sigmoid, tanh, and other hyperbolic functions, the rectified linear function has the following advantages:

(1) Imitation of biological principles: brain studies have shown that the message encoding of biological neurons is relatively scattered and sparse [13]. There are about 1–4% of neurons working in the brain at the same time. With linear rectification and regularization, we can know the detailed activities in the machine neural network. The logic function reaches 12 at input 0, which is already half full and stable

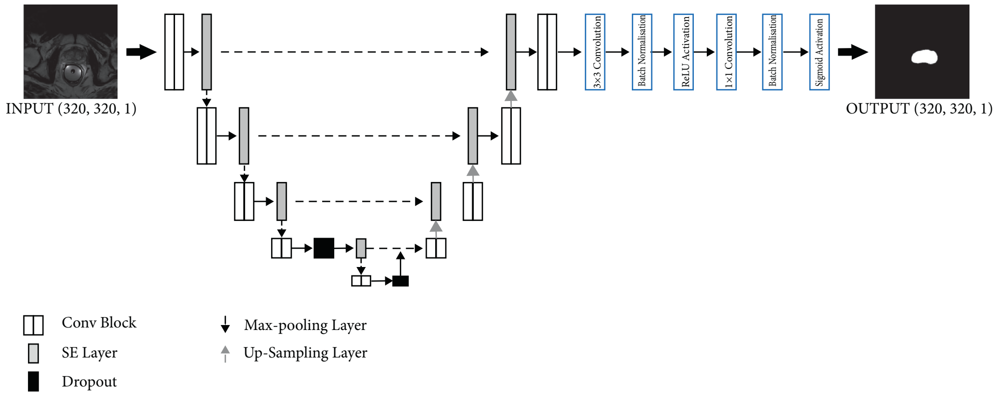

Figure 1: Overall structure of the segmentation network. The network includes the encoder and decoder stage connected with skip connection operation. The base components are Conv Block, SE layer, max-pooling layer, and upsampling layer. The first two components will be introduced later.

which is not the same as the expectation of the scientist who think a simulated neural network is the same as the real biology [14].

(2) More efficient gradient descent and backpropagation.

(3) Simplify the calculation: ReLU function can prevent the influence of complicated function, for example, exponential functions, and reduce the total computing cost of the model.

3.3. Dropout Layer. Dropout is a popular way to prevent overfitting in neural network training. In the training process of deep learning network, dropout temporarily discards neural network units from the network with a certain probability, which causes each batch to train a different network model. Use the average to improve the generalization ability of the model. In addition to overfitting, dropout also alleviates the problem of long training time for large-scale neural networks.

3.4. Skip Connections. Skip connection is an operation that skips some of the layer of the network and then takes the output of the layer to feed to the next layers. In U-Net, skip connections were used to fight the vanishing gradient problem and learn pyramid level features [9]. The main idea of skip connections in U-Net is to have the pretrained features and reuse them in the later layer to improve the performance. The features are transferred from the encoder layer to the decoder layer by skip connections which are combined with concatenation instead of summation.

4. Proposed Methods

In this paper, we proposed an improved 2D U-Net model integrated squeeze-and-excitation layer which is used to segment prostate cancer automatically. We are going to introduce our proposed model and the main blocks.

4.1. Model Structure. We did some improvements to the traditional U-Net. Inspired by [8, 9], we added some squeeze-and-excitation (SE) layers, which will be introduced later, based on U-Net. Our model is divided into the encoder stage and the decoder stage; on the encoder stage, the model can effectively extract the input image feature by continuous convolution layer and pooling layer; on the decoder stage, the model will step by step map the extracted features to the original image by the continuous upsampling layer and output predicted mask eventually. Figure 2 is our proposed model, which is more complex than the traditional U-Net. In particular, we added a SE layer before each encoder's pooling layer and after each decoder's upsampling layer.

4.2. CONV BLOCK. We use skip connection operation to concatenate two continuous convolution layer and activation layer and consist of a block and put them into a block which we named as CONV BLOCK. Figure 3 is its inner structure.

4.3. SE Layer. Inspired by [10], calculating the importance weights of each channel and then marking the more useful features, referring to Se-Net's [15] practice, we implemented a method which can extract important features from channels and named it the SE layer; Figure 4 shows its detailed structure.

First of all, we assume feature $F \in R^{H\times W\times C}$, H, W, and C represent the height, width, and channel and number of features is F, respectively, and the function of F is

$$F = [F_1, F_2, \ldots, F_i, \ldots, F_C]. \tag{3}$$

F_i is the i_{th} feature of the channel. For feature F, we use a global average pooling layer (GAP) to generate a vector and named it z_i whose function is

$$z_i = \frac{1}{H \times W} \sum_{x}^{H} \sum_{y}^{W} F_i(x, y). \tag{4}$$

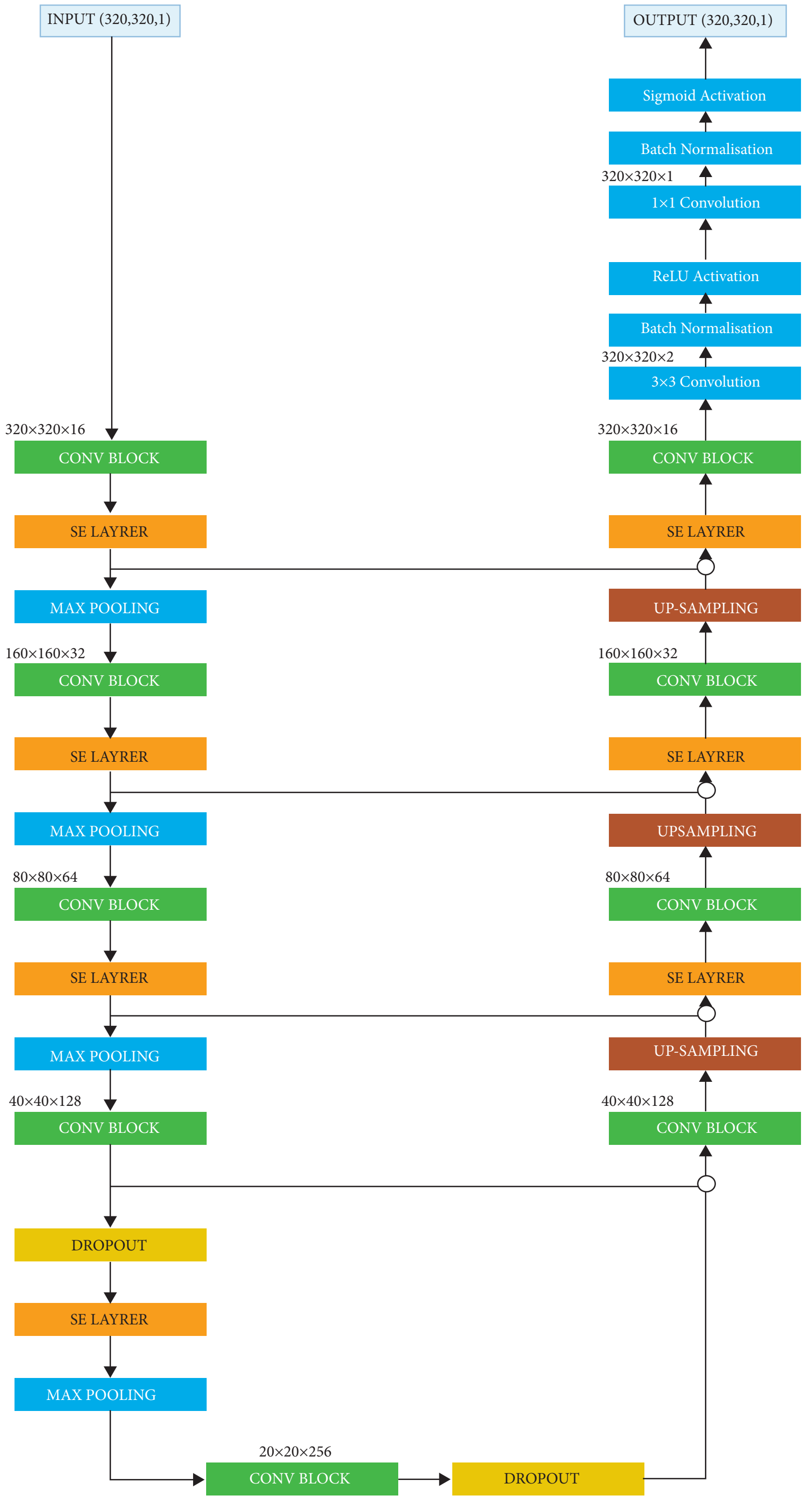

FIGURE 2: Our proposed model structure. The input size is 320 × 320. And the encoder consists of a series of CONV BLOCK, SE layer, max-pooling layer, and two dropout layers. The decoder consists of a series of CONV BLOCK, upsampling layer, and SE layer. In order to get a 320 × 320 output, the tail of the decoder consists of two convolution layers, two batch normalisation layers, a ReLU function, and a sigmoid.

z_i is the i_{th} global averaged channel. After that, we use a ReLU activation layer and a sigmoid activation layer to achieve information aggregation as

$$z' = \sigma(W_2\delta(W_1 z)), \quad (5)$$

where σ refers to the ReLU function, $W_1 \in R^{C\times C/r}$ and $W_2 \in R^{C/r\times C}$ refer to the two fully connected layers, and r is a ratio parameter to reduce the dimensional complexity which is set to 4. The importance of each feature channel can be learned and named as z'.

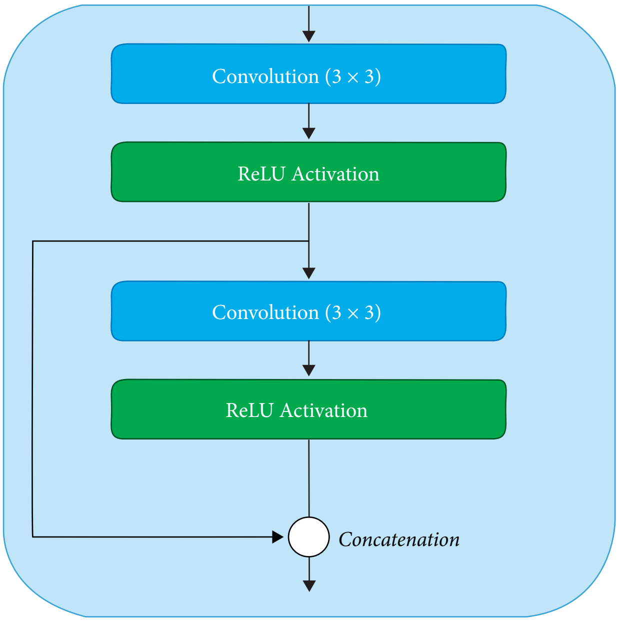

Figure 3: The inner structure of the CONV BLOCK. It consists of two continuous convolution layer and activation layer using concatenate operation.

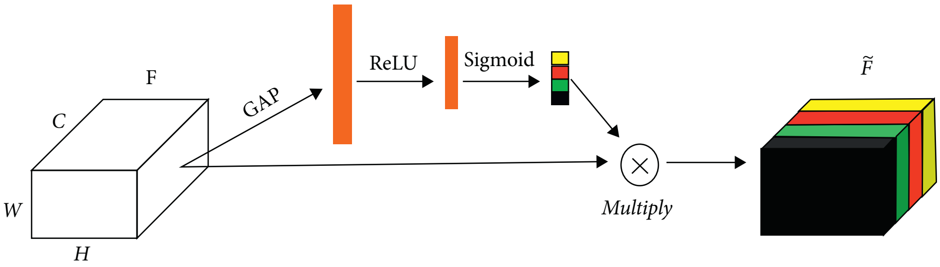

Figure 4: The structure of the SE (squeeze-and-excitation) layer, including a GlobalAveragePooling layer, a ReLU activation, and a Sigmoid activation. F will select important features by a multiply operation as $\widetilde{F}$.

We can extract important features by multiplying F with z', and it can be described as

$$\begin{aligned}\widetilde{F} &= F * z' \\ &= [F_1 * z'_1, F_2 * z'_2, \ldots, F_i * z'_i, \ldots, F_C * z'_C].\end{aligned} \tag{6}$$

The SE layer is a good way to enhance the ability to learn globally of the model, which is proved to be correct and valid in [15], by strengthening more important features. We use it in both the encoder stage and decoder stage; the detailed location is described in Section 3.1.

4.4. Evaluation Function. We choose Dice similarity coefficient (DSC) as our evaluation function according to [16]. Denote P the predicted mask and GT the ground truth:

$$\mathrm{DSC} = \frac{2|P \cap \mathrm{GT}|}{|P| + |\mathrm{GT}|}. \tag{7}$$

In addition to this, we also choose accuracy (AC), Jaccard index (JA), and sensitivity (SE). TP, FP, TN, and FN represent true positive, false positive, true negative, and false negative, respectively. Their functions can be described as

$$\begin{aligned}\mathrm{AC} &= \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{FP} + \mathrm{TN} + \mathrm{FN}}, \\ \mathrm{JA} &= \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP} + \mathrm{FN}}, \\ \mathrm{SE} &= \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}}.\end{aligned} \tag{8}$$

5. Results

5.1. Dataset. The performance of the model is evaluated on a public dataset, PROMISE12 dataset, which includes 50 training sets and 30 continuous T2 weighted MR images in each set. We will resize the original image to 320 × 320 as the input of the model after loading the origin images.

5.2. Training. The designed model is based Tensorflow-Keras library. Our test set and training set all run on 6 GB

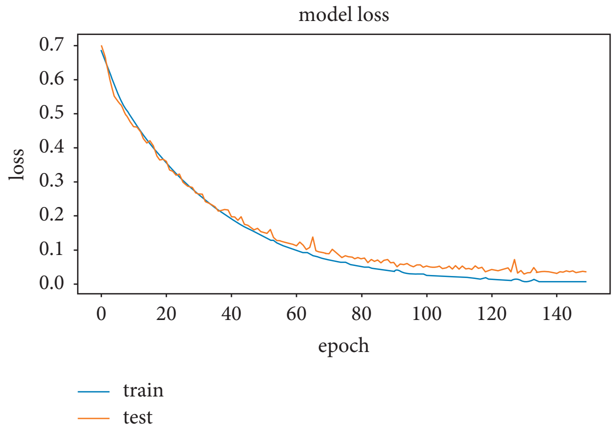

Figure 5: The loss curve of the model. The train loss curve dropped gradually to 0.05 and lower until 150 epochs.

NVIDIA GTX 1660TI GPU with Intel (R) Core (TM) i7-9750H CPU @ 2.60 GHz 16RAM. The initial learning rate is 10^{-4}, and the epoch is 150. Before training, we use random flip, rotation, and cropping to augment our training sets to get better training results.

We use an Adam optimizer [17] with a 10^{-4} learning rate as we mention above and a binary cross-entropy loss function [18], given by

$$\text{Loss} = -\left(y_i \log f\left(x_i, \theta\right) + \left(1 - y_i\right) \log\left(1 - f\left(x_i, \theta\right)\right)\right), \quad (9)$$

where $f(x_i, \theta)$ is the prediction of the network on sample i in a range between 0 and 1 and y_i is the ground truth of sample i in binary 0 or 1.

5.3. Results and Discussion. After the training of 150 epochs using five folds to pick each train set and test set, we can get the model loss and accuracy curves.

As can be seen in Figures 5 and 6, both the loss and accuracy curves perform well, and the effectiveness of the training was preliminarily proved. Two curves remain stable in dozens of epochs, which showed the model is not overfitted. And the gradual decline of the curve demonstrates good convergence of the model.

To show the effectiveness of our model, we implemented three traditional prostate segmentation methods [8, 9, 19]. The work in [8] is fully convolutional networks (FCN), [9] is traditional U-Net, and [19] is a multiatlas method. We will compare our model results to the other three model results mentioned above.

After examining the score in the whole dataset using five-fold cross validation, our model performed well compared to the other three models whose mean DSC is 0.87 and median DSC is 0.89. And the remaining three were also higher than the others.

The detailed five-fold cross-validation results can be seen in Figure 7.

As can be seen in Figure 7, our model performed well on five-fold cross validation. Most of its DSC scores are in the range of 0.70 to 0.95. On the first fold, the median DSC score is above 0.90 and the mean DSC score is a little lower in the range of 0.85 to 0.90. And the second, fourth, and fifth folds are almost like the first fold whose median DSC is around 0.9. And the mean DSC of all five folds is 0.87 which can be seen in Table 1.

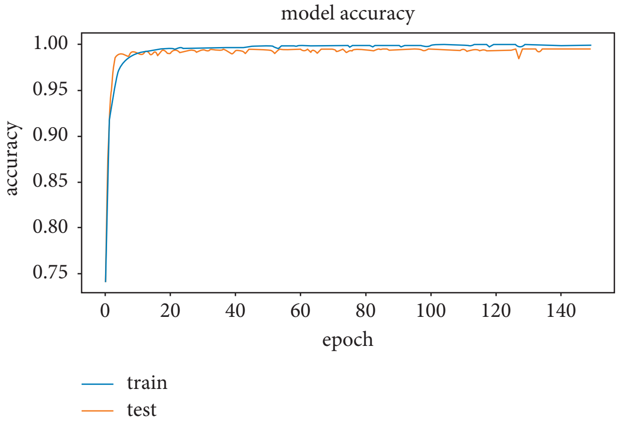

Figure 6: The accuracy curve of the model. The train accuracy curve increased gradually to about 0.95 until 150 epochs.

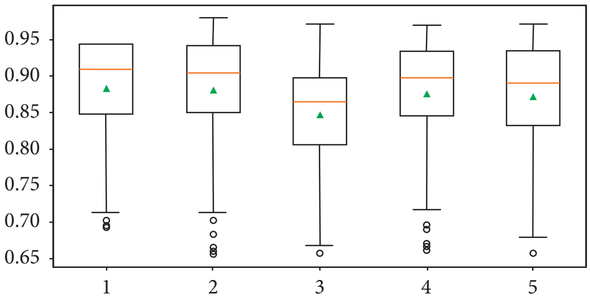

Figure 7: Five-fold cross-validation Dice's similarity coefficient (DSC) scores plotting with box-and-whisker. The orange line represents the median DSC score and the green triangle icon represents the mean DSC score on the first fold, the median DSC score is above 0.90, and the mean DSC score is about 0.88. The other folds performed well like the first fold, whose median DSC score is above 0.90 and the mean DSC is in the range of 0.85 to 0.90, except the third fold whose median DSC score is above 0.85 and mean DSC score is a little lower than 0.85.

Table 1: Performance comparison between our model and the traditional methods.

Method	Mean DSC	Median DSC	Mean AC (%)	Mean JA (%)	Mean SE (%)
FCN	0.79	0.81	84.6	72.5	90.6
U-Net	0.81	0.82	85.8	74.0	92.7
Multiatlas	0.80	0.82	85.0	73.5	90.1
Our model	0.87	0.89	87.3	75.3	93.2

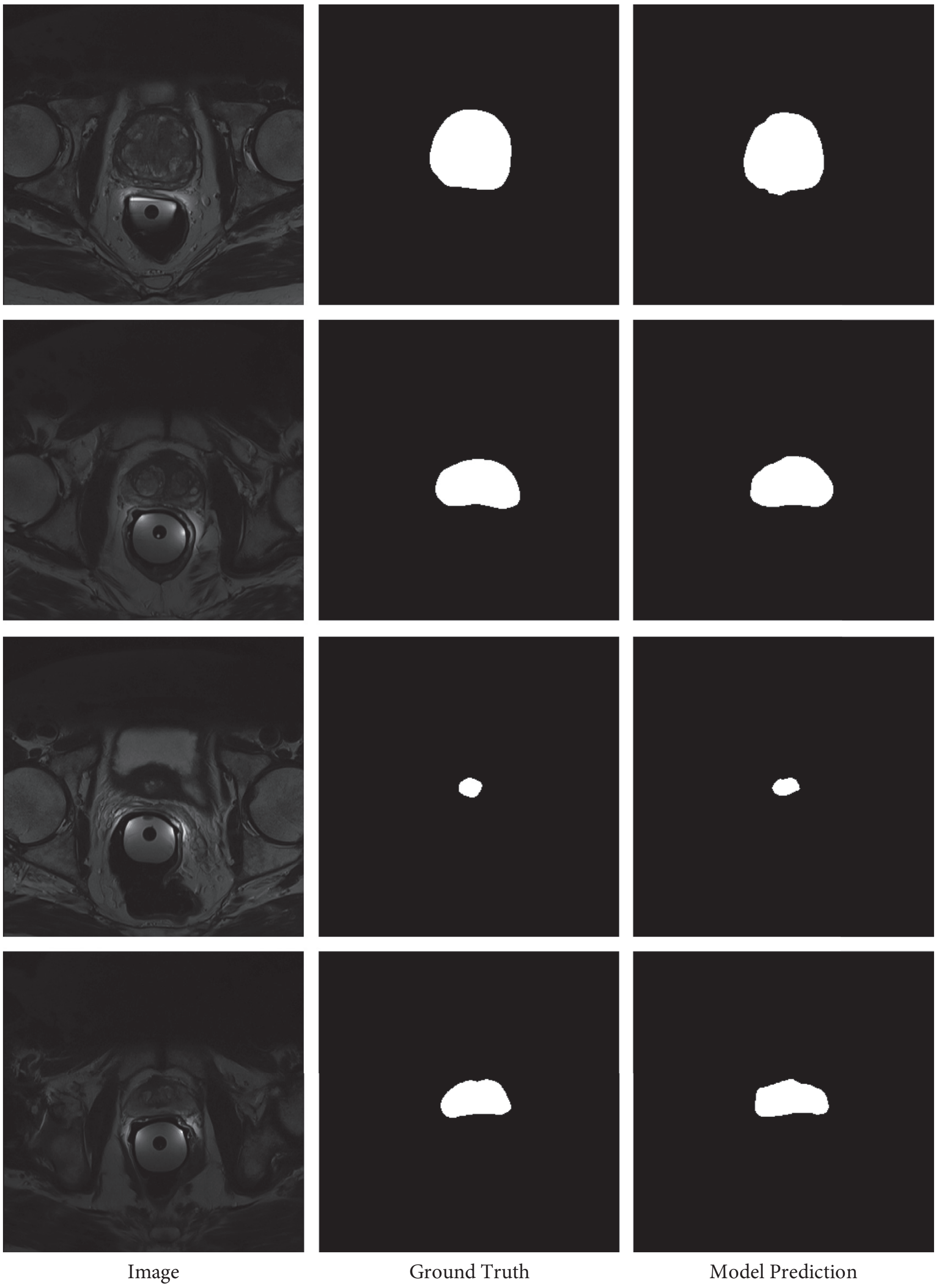

Figure 8: The visualization of some segmentation results in test sets.

6. Conclusion

In this paper, we develop an improved 2D U-Net model integrated Squeeze-and-excitation layer for prostate cancer segmentation. We divided two important components: SE layer and CONV BLOCK. With the SE layer, our model can learn more global and local features. In the CONV BLOCK, we combined feature maps and skip connection with a concatenation operation to bring further improvement in the model performance. In future work, different MRI modalities are going to be tried on our model to segment prostate cancer automatically.

Data Availability

The prostate MRI image dataset can be downloaded from the website (https://promise12.grand-challenge.org/Download/).

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Authors' Contributions

The authors contributed equally to this paper.

Acknowledgments

This study was funded by the Science and Technology Planning Project of Xiamen City (no. 3502Z20184036).

References

[1] N. Mishra, S. Petrovic, and S. Sundar, "A knowledge-light nonlinear case-based reasoning approach to radiotherapy planning," in *Proceedings of the 2009 21st IEEE International Conference on Tools with Artificial Intelligence*, pp. 776–783, Newark, NJ, USA, November 2009.

[2] I. P. Astono, J. S. Welsh, S. Chalup, and P. Greer, "Optimisation of 2D U-net model components for automatic prostate segmentation on MRI," *Applied Sciences*, vol. 10, no. 7, 2020.

[3] S. S. Chandra, J. A. Dowling, K. Shen et al., "PatientSpecific prostate segmentation in 3-D magnetic resonance images," *IEEE Transactions on Medical Imaging*, vol. 31, pp. 1955–1964, 2012.

[4] S. Martin, J. Troccaz, and V. Daanen, "Automated segmentation of the prostate in 3-D MR images using a probabilistic atlas and a spatially constrained deformable model," *Medical Physics*, vol. 37, no. 4, pp. 1579–1590, 2010.

[5] D. Karimi, G. Samei, Y. Shao, and T. Salcudean, "A Deep Learning-Based Method for Prostate Segmentation in T2-Weighted Magnetic Reso-Nance Imaging," 2019, https://arxiv.org/abs/1901.09462.

[6] Y. Guo, Y. Gao, and D. Shen, "Deformable MR prostate segmentation via deep feature learning and sparse patch matching," *IEEE Transactions on Medical Imaging*, vol. 35, no. 4, pp. 1077–1089, 2016.

[7] R. Cheng, B. Turkbey, W. Gandler et al., "Atlas based AAM and SVM model for fully automatic MRI prostatesegmentation," in *Proceedings of the 2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 2881–2885, Chicago, IL, USA, August 2014.

[8] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proceedings of the IEEE Conf. Comput. Vis. Pattern Recognit*, pp. 3431–3440, Silver Spring, MD, USA, June 2015.

[9] O. Ronneberger, P. Fischer, and T. Brox, "U-Net Convolutional networks for biomedical image segmentation," in *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 234–241, Springer, Cham, Switzerland, November 2015.

[10] Z. Zhou, M. Siddiquee, N. Tajbakhsh, and J. Liang, "U-Net++ A Nested U-Net Architecture for Medical Image Segmentation," in *Proceedings of the Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*, Cham, Switzerland, September 2018.

[11] F. Milletari, N. Navab, and S. A. Ahmadi, "Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation," in *Proceedings of the 2016 Fourth International Conference on 3D Vision (3DV)*, IEEE, Los Alamitos, CA, USA, June 2016.

[12] G. Zhang, Y. Luo, B. Zhao et al., "A Bi-attention adversarial network for prostate cancer segmentation," *IEEE Access*, vol. 7, 131458 pages, 2019.

[13] X. Glorot, B. Antoine, and Y. Bengio, "Deep sparse rectifier neural networks," in *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*, PMLR, Ft.Lauderdale, FL, USA, June 2011.

[14] D. Attwell and S. B. Laughlin, "An energy budget for signaling in the grey matter of the brain," *Journal of Cerebral Blood Flow and Metabolism*, vol. 21, no. 10, pp. 1133–1145, 2001.

[15] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vis. Pattern Recognit.*, pp. 7132–7141, Salt Lake, UT, USA, June 2018.

[16] V. Yeghiazaryan and I. Voiculescu, "An overview of current evaluation methods used in medical image segmentation," Technical Report RR-15-08, Department of Computer Science, Oxford, UK, 2015.

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

[18] A. Jain, A. Fandango, and A. Kapoor, *TensorFlow Machine Learning Projects: Build 13 Real-World Projects with Advanced Numerical Computations Using the Python Ecosystem*, Packt Publishing, Birmingham, UK, 2018.

[19] J. A. Dowling, "Automatic substitute computed tomography generation and contouring for magnetic resonance imaging (MRI)-Alone external beam radiation therapy from standard MRI sequences," *International Journal of Radiation Oncology, Biology, Physics*, vol. 93, no. 5, pp. 1144–1153, 2015.