IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received November 7, 2019, accepted November 23, 2019, date of publication November 28, 2019,
date of current version December 12, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2956550

Pancreatic Segmentation via Ringed Residual U-Net

LIN LU[1], LIQIONG JIAN[2], JUN LUO[1], AND BIN XIAO[3]
[1]School of Computer Science, Chongqing University, Chongqing 400044, China
[2]Ningxia Blood Center, Yinchuan 750001, China
[3]College of Computer Science and Technology, Chongqing University of Posts and Telecommunications, Chongqing 400065, China

Corresponding author: Bin Xiao (xiaobin@cqupt.edu.cn)

ABSTRACT Computer-aided diagnosis (CAD) has a wide range of clinical applications, and medical image segmentation is essential in CAD. In medical image segmentation, due to the high anatomical variability of the pancreas and the weak contrast of the environment, the segmentation of the pancreas has always been the most challenging task. This study proposes a novel segmentation method for pancreatic segmentation. On the basis of complete convolution, an attentional mechanism is added to enhance the information exchange between downsampling and upsampling. By using the ring residual module, the proposed segmentation method can yield satisfactory results via deep convolution and can consolidate the characteristics of traditional deep learning networks. At the same time, compared to previous methods that used the dice coefficient (DICE) as a loss function, this study proposes a new loss function in the proposed method. This new loss function focuses not only on the area coincidence degree but also on the focus shape similarity. In the present work, ten-fold cross-validation computed tomography (CT) data (82 samples) from the NIH public pancreas dataset was conducted. The average dice similarity coefficient (DSC) of the results is 88.32 ± 2.84, which is higher than the most advanced available methods and corresponds to higher robustness. Therefore, in practical applications, these methods can be used to provide more reliable auxiliary diagnostic data in the application of clinical medicine.

INDEX TERMS Attentional mechanism, deep neural networks, pancreas segmentation and ring residual module.

I. INTRODUCTION

Medical images contain a substantial amount of information that is highly useful for CAD. Hence, this has become a research hotspot in recent years. To apply this information, the organ segmentation of medical images is necessary. The present study mainly investigates the segmentation of a small organ— the pancreas—in abdominal computed tomography (CT). Pancreatic segmentation is an important component of CAD [1], [2], which has been used mainly for the quantitative imaging analysis of diabetic patients and the detection of pancreatic cancer.

With the improvement of deep learning algorithms [3], [4], convolutional neural networks (CNNs) have been increasingly used in medical image processing. For the organ segmentation of medical images, there are three general methods: manual segmentation, semiautomatic segmentation, and automatic segmentation. Manual segmentation requires the operator to have a high level of professional knowledge. Furthermore, the quality of the manual segmentation result depends mainly on the subjective judgment of the technician. Forming this judgment requires a long period of experience accumulation. In addition, manual segmentation is a highly time-consuming task. These shortcomings hinder the practical application of manual segmentation [5]. Semiautomatic segmentation requires various degrees of artificial intervention, which may cause subjective errors and biases in the segmentation result. This is an overprovisioning scheme that is mainly applied due to immaturity of the automatic segmentation method. For accelerating diagnosis and developing treatment plans, the demand for automatic segmentation methods is growing.

Excitingly, in the field of automatic segmentation, many other organs in CT, such as the lungs [6], [7], heart [8], [9], and kidneys [10], are segmented by deep learning algorithms.

Satisfactory results can be obtained, and the DSC can reach the level of $>90\%$ [11]–[13]. Regrettably, the segmentation of small organs, such as the pancreas, from CT scans has been

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

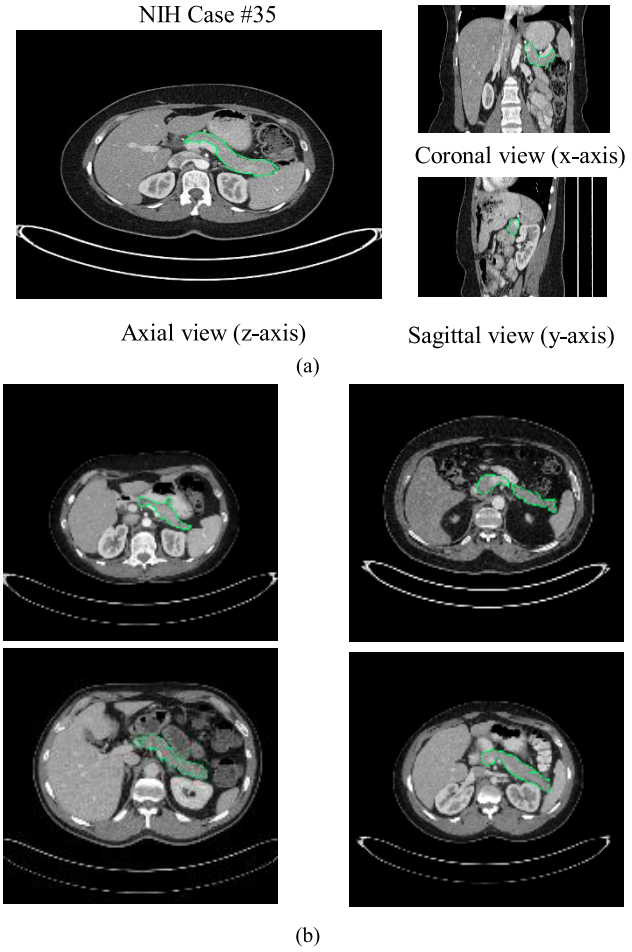

FIGURE 1. (a) The position of the pancreas is marked in green in the CT three different viewpoints. This is a small organ. (b) The pancreas is marked in green. It is an organ with an irregular shape and boundary.

challenging. This is mainly due to changes in the pancreas in large spaces, such as CT scans [14]. The primary factor is that the pancreas is only a small fraction of the entire CT image (Figure 1a), and the pancreas is a relatively small organ. Furthermore, the shape and behavior of the pancreas can be altered by changes in the motion of adjacent organs (Figure 1b), and the boundaries of the pancreas become blurred due to the surrounding organs. In addition, the limited number of available labeled medical images also limits the accuracy of the segmentation algorithm.

In deep learning segmentation, networks such as fully convolutional networks (FCNs) [15] and UNet [16] have been used for the segmentation of medical images. Zhou *et al.* [17] applied predictive segmentation masks to fixed-point models to narrow the input region. The average DSC of the segmentation increased to 83.18%. This segmentation DSC is much higher than the 81.14% segmentation DSC that was reported by Roth *et al.* [14], [18] in 2015 and 2016. In the latest pancreas segmentation task in 2018, Cai *et al.* [19] used the deep CNN method and obtained an average DSC value of 83.7%. In the study that was conducted by Zhu *et al.* [20], a coarse and thin network was used, and the coarse network was trained to obtain coarse segments and to remove the background regions. After this, the coarse regions were passed to a fine network for precise segmentation to obtain the average DSC. An improvement of 84.59% was realized. In contrast to the method that was used by Zhu et al., Ma *et al.* [21] proposed a new Bayesian model, which was combined with the segmentation method of a deep neural network and a statistical shape model and improved the average DSC to 85.32%.

From the perspective of current research, due to the rapid increase in the number of medical images, the use of manual segmentation or semiautomatic segmentation has become an almost-impossible task. In the field of automatic segmentation, many researchers have conducted studies. However, because the pancreas appears too small in a CT image and the pancreas is easily squeezed by other organs, the grayscale of the image is highly variable. Therefore, the above research performance is unsatisfactory, and the methods are not sensitive to the edge features of the pancreas. In the present study, the large deformation method was used to transfer the features that were extracted in the downsampling to the upsampling using the attention mechanism, which improved the accuracy of the segmentation results. Then, the ring residual module was used to extract additional features in the deep part of the network, which were used to address the boundary blur problem of the pancreas.

The remainder of this paper is organized as follows: Section 2 explains how the attention module and the ring residual module are integrated into UNet to obtain a new network structure. Section 3 describes the experimental process and the experimental results. Section 4 presents the conclusions of this study.

II. MATERIALS AND METHODS

A. THE ATTENTION MECHANISM AND THE CONVOLUTION BLOCK ATTENTION MODULE

The development of CNNs has substantially advanced the performance of computer vision tasks [22]–[24]. According to recent studies, the three main factors that affect the performance of a CNN are the depth, width, and cardinality of the network. LeNet [25] and VGGNet [26] have shown that a deeper network structure can yield improved performance. Resnet [27] overcomes the problem of gradient disappearance, which is caused by deep networks, by making the network more deeply designed, thereby improving the performance of CNNs. In GoogleNet [34], the network width is also a major factor that affects the CNN performance. Xception [28] and ResNeXt [29] increase the cardinality of the network and reduce the total number of parameters by increasing the cardinality. However, in addition to these three factors, in recent years, it was found that attention is also one of the main factors that affect the performance of a CNN. The attention mechanism determines the focus of the algorithm and improves the representation of the features. This process renders the CNN more similar to the human

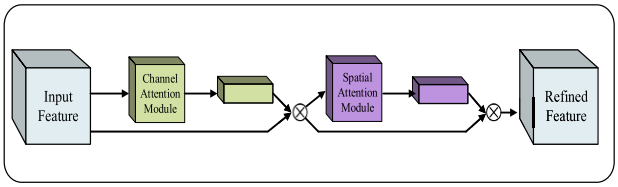

FIGURE 2. Overview of the CBAM.

eye when receiving image information. Although the human visual system does not process the information of the entire image at one time, human beings perform well at focusing on the local information of the image, thereby more effectively capturing the overall information structure.

The results of the experimental study that was conducted by Woo *et al.* [30] demonstrated that combining the attention mechanism with CNNs can improve the overall performance of the CNN, and they proposed a convolution block attention module (CBAM) (Figure 2). The convolution attention module mainly implements learning enhancement or reduces the influence of information by sequentially arranging the channel and space of two attention modules. The channel attention module solves the problem of learning ''what'', while the spatial attention module solves the problem of learning ''where.'' The whole process can be expressed as follows:

In the channel attention module, global maximum and global average pooling operations were performed on the input feature map based on the width and height, and two spatial context descriptions were generated. Then, these two sets of descriptions were entered into the same fully connected network, which contains a hidden layer. After passing through the shared network, these two sets of descriptions were summed element by element and activated using the sigmoid function to form a new feature map. The calculation process is as follows:

$$\begin{aligned} M_c(F) &= \sigma\left(MLP\left(AvgPool\left(F\right)\right) + MLP\left(MaxPool\left(F\right)\right)\right) \\ &= \sigma\left(W_1\left(W_0\left(F^c_{avg}\right)\right) + W_1\left(W_0\left(F^c_{max}\right)\right)\right) \end{aligned} \quad (1)$$

where F is an input feature and $M_c(F)$ is the corresponding output feature; W_1 and W_0 are the fully connected network weights, and the ReLU activation function is followed by W_0; F^c_{avg} is the feature after average-pooling; F^c_{max} is the feature after max-pooling; σ is the sigmoid function; and MLP is the fully connected network.

In the spatial attention module, focus was no longer directed to the effects of the channel. For the input feature map to perform the global maximum and global average pooling operations based on the channel, the generated two-pooling-layer information was linked. After a convolution operation by a 7-by-7 convolution kernel, the spatial attention feature was generated by the sigmoid function.

$$\begin{aligned} M_s(F) &= \sigma\left(f^{7\times7}\left(\left[AvgPool\left(F\right); MaxPool\left(F\right)\right]\right)\right) \\ &= \sigma\left(f^{7\times7}\left(\left[F^s_{avg}; F^s_{max}\right]\right)\right) \end{aligned} \quad (2)$$

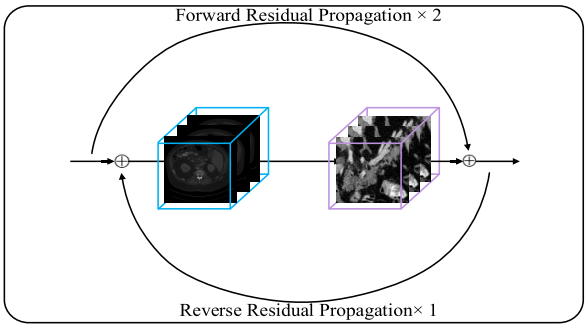

FIGURE 3. Overview of the ring residual model.

where F is the input feature and $M_s(F)$ is output feature, F^c_{avg} is the feature after average-pooling, F^c_{max} is the feature after max-pooling, σ is the sigmoid function, and $f^{7\times7}$ is the 7-by-7 convolution.

B. RING RESIDUAL MODEL

A deep network would cause gradient degradation problems, which would reduce the performance of the CNN. The ring residual module can overcome this problem because it can more effectively use the spatial information of the context and strengthen the learning mode of the CNN. This would further solve the degradation problem and ensure that the extracted feature image would not be weakened due to the depth of the network. The ring residual module (Figure 3) consists of two parts: a forward residual module and a reverse residual module. The equation is as follows:

$$y_b = G\left(F\left(x_f, \{W_i\}\right) + W_s * x_f\right) + x_b \quad (3)$$

where x_f is the input characteristic of the forward residual block, W_i is the weight of the i-th layer, function $F(x, \{W_i\})$ represents the mapping that must be learned, W_s is a linear projection, F is a mapping function, x_b is the input characteristic of the reverse residual, y_b is the output characteristic after the reverse residual module, and G is a linear projection.

In the back-propagation of deep neural networks, due to the chain-based differentiation rules, the gradient in the network can easily disappear. This is similar to the curve of human learning and learning: Humans forget old knowledge in the process of learning new knowledge. The process of forgetting is similar to the disappearance of the gradient. To overcome this problem, it is necessary to review the old knowledge to reinforce the old memory. In the deep network, the residuals were used to compensate for the disappearance of the gradient that is caused by the chain differentiation. The addition of the previous convolution and the nonlinear operation after the convolution operation (Figure 4) would slow the degradation of the network. This addition is the human review process. Forward convolutions are introduced after the convolution and nonlinear operations. The equation is as follows:

$$y_f = F\left(x, \{W_i\}\right) + W_s * x \quad (4)$$

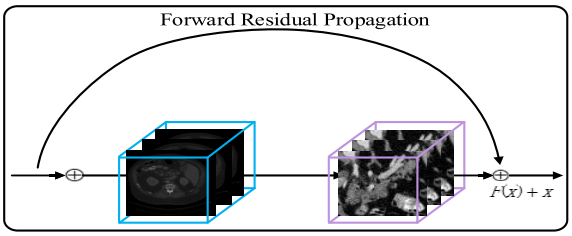

FIGURE 4. Overview of the forward residual propagation.

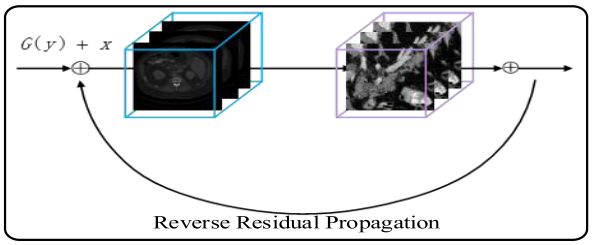

FIGURE 5. Overview of the reverse residual propagation.

where x is the input characteristic of the forward residual block and y_f is the output feature; W_i is the weight of the i-th layer; function $F(x, \{W_i\})$ represents the mapping that must be learned; and W_s is a linear projection that changes the x-size to match the size after the execution of the mapping function over F.

However, since the forward residual cannot make the spatial context information continuous and stable, the investigators propose a reverse residual for reinforcing the input information. The inverse residual (Figure 5) is an automated learning method. The equation is as follows:

$$y_b = G\left(y_f\right) + x \tag{5}$$

where x is the input characteristic of the reverse residual and y_b is the output characteristic after the reverse residual module; y_b is the feature vector after passing through the forward residual module; and G is a linear projection, which changes the size of y_f.

As an automated learning method, the inverse residual takes the features that were obtained by the forward residual output as inputs, examines these features, and stabilizes the information, thereby rendering continuous the characteristics that were obtained by the forward residual and suppressing required feature information. This accelerates the convergence of the entire network.

C. SPLIT NETWORK STRUCTURE

In the present study, the UNet network structure was used as the basis, without pretraining. The input to the network and the eigenvectors after each layer are convolved and are all three-dimensional arrays of size $h \times w \times d$. h and w represent the height and width of the feature vector, and d represents the dimension of the channel (Figure 6). The convolution operation is based on translation invariance and includes

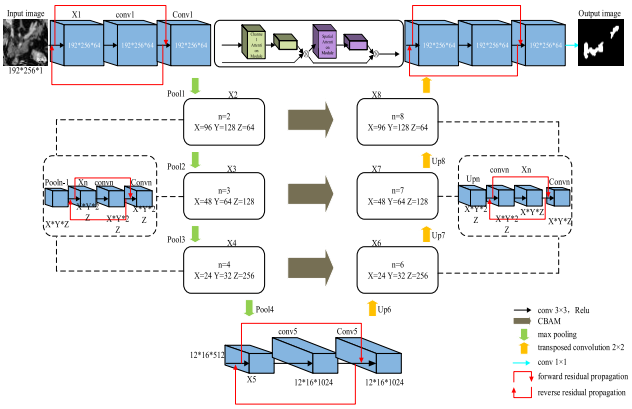

FIGURE 6. Overview of RRA-UNet network.

mainly convolution, pooling and activation functions.

$$y_{ij} = f_{ks}\left(\left\{X_{si+\delta i, sj+\delta j}\right\} 0 \le \delta i, \delta j \le k\right) \tag{6}$$

where y_{ij} is the output characteristic after each layer of convolution operation, X_{ij} is the input vector for each layer of special position, k is the size of the convolution kernel, s is the convolution step, and f_{ks} determines the type of convolution operation for each layer, such as matrix multiplication, averaging of the pooling function or multiplication of the maximum pooling function and the final activation function.

The deconvolution operation, or upsampling, is a backward convolution operation. Upsampling can be thought of as a normal convolution operation with an input step of $1/f$, where f is an integer.

Based on the problems that are encountered with the available pancreas segmentation approaches that are discussed in the previous section, regardless of whether upsampling or downsampling is being conducted, we will add the ring residual module and CBAM to the downsampling under the original UNet network.

In the medical image segmentation, the DICE loss function was the most commonly used metric and loss function. The DICE loss function mainly represents the ratio of the actual result to the total area of the predicted result intersection area. The present study improves the loss function, and the investigators named this improved function the complex-coefficient (CPCE) loss function. The CPCE loss function focuses not only on the ratio of the coincident area to the total area but also on the shape similarity between the real result and the predicted result. Hence, the CPCE loss function consists of two parts.

The first part is the area coincidence, and the equation is, as follows:

$$\mathrm{L}\left(Y, \hat{Y}\right) = \frac{-2\sum_i y_i \hat{y}_i}{\sum_i y_i + \sum_i \hat{y}_i} \tag{7}$$

where $y_i \in Y, \hat{y} \in \hat{Y}$, Y is the ground truth, $\hat{Y}$ is the prediction, and L is area coincidence function. However, problems are encountered in evaluating the segmentation results based only on the coincidence rate of the area. Therefore,

the actual results were compared with the shape of the prediction results, with the objective of obtaining a segmentation result that is more in line with the actual results. The degree of similarity between these two vectors was determined by determining whether the directions of the two vectors are consistent or similar and by calculating the cosine angle of these two vectors. The real results and the prediction results were flattened into two vectors, and the cosine similarity calculation method was used to calculate the shape similarity of these two results as the second part of the CPCE loss function. The equation is as follows:

$$\text{shape similarity} = \cos\left(Y, \hat{Y}\right) = \frac{Y \cdot \hat{Y}}{\| Y \| \cdot \| \hat{Y} \|} = \frac{\sum_i y_i \times \hat{y}_i}{\sqrt{\sum_i y_i^2} \times \sqrt{\sum_i \hat{y}_i^2}} \tag{8}$$

where: $y_i \in Y$, $\hat{y} \in \hat{Y}$, Y represents the ground truth, $\hat{Y}$ represents the prediction, and cosin is the shape similarity function. Therefore, the CPCE loss function is defined as:

$$\mathrm{C}\left(Y, \hat{Y}\right) = \lambda * \mathrm{L}\left(Y, \hat{Y}\right) + \xi * \cos\left(Y, \hat{Y}\right) \tag{9}$$

where λ is the penalty coefficient of the area coincidence, ξ is the penalty factor for the shape similarity, L is area coincidence function, and cosin is the shape similarity function. In the present study, we set λ to 0.8 and ξ to 0.2.

III. EXPERIMENTAL RESULTS

A. DATASET AND EXPERIMENT SETUP

The public NIH pancreatic dataset [14] was selected for the present experiment. The NIH pancreatic dataset included 82 abdominal contrast-enhanced 3D-CT data with a CT size of 512×512×D (D∈[181,146]). With 10-fold cross-validation, the bounding box was used to cut the clips, and the margins for the annotations were selected from the margins of the filled original image data that were reported in the literature [17]. The HU value of the cropped image was in the interval [−100, 240], and the image was scaled to the range [0, 1]. The DICE loss function and the CPCE loss function were used as the loss functions to conduct the two sets of controlled trials. The dice coefficient was also to evaluate the similarity between the present segmentation results and real results. $\mathrm{DSC} = 2\left(|Y \cap \hat{Y}|\right) / \left(|Y \cup \hat{Y}|\right)$.

The hardware that was used in the present study was a 12-core Intel core i7-8700k CPU 3.70 GHz with GeForce RTX 2070. During the training, the learning rate was set to 0.0001, the number of epochs was set to epochs = 20, and the size of each batch was set to 2. The size of the cross-validation set was set to 10, and it took approximately six hours to fully train the entire network structure. It took approximately 45 seconds to predict a case.

TABLE 1. Comparison of experimental results.

Method	Mean DSC (%)	Min DSC (%)	Max DSC (%)
A Fixed-Point Model[17]	83.18±4.81	65.10	91.03
Direct Loss Function[31]	82.40±6.70	60.00	90.10
Context Learning[19]	83.70±5.10	59.00	91.00
Novel Bayesian Model[21]	85.32±4.19	71.04	91.47
DenseUNet[21]	73.39±8.78	45.00	86.50
Cascaded 3D FCN[32]	76.80±9.40	43.70	89.40
Spatial Aggregation[33]	81.30±6.30	50.60	88.90
UNet[16]	79.70±7.60	43.40	89.30
RRA-UNet(DICE)	88.16±2.96	79.88	93.86
RRA-UNet(CPCE)	88.32±2.84	80.50	93.30

B. EXPERIMENTAL RESULTS AND ANALYSIS

Ten-fold cross-validation was used on the NIH dataset to compare the experimental results. The results are presented in Table 1.

According to the results in Table 1, the highest average DSC of the ring residual UNet with an attention mechanism (RRA-UNet) was 88.32%, and the standard deviation of only 2.84 was the smallest. The worst and best segmentation results were 80.50% and 93.88%, respectively. In the same case where the DICE loss function was also used, the present segmentation result is increased by five percentage points on the average DSC compared to that reported by Zhou *et al.* [17] by using the predicted segmentation mask to apply the fixed-point model. Compared with the results that were reported by Cai *et al.* [19], the present FCN-based method improved the data of the most difficult to segment pancreatic cases by 20.88 percentage points. The segmentation results of UNet [16] and DenseUNet [21] have demonstrated that the segmentation network of UNet is a reliable method for medical images, especially for pancreas segmentation. Ma *et al.* [21] integrated the Bayesian model into the deep neural network and statistical shape model to fine-tune the UNet network structure, thereby increasing the segmentation result of the pancreas to 85.32%, but this study was conducted a long time ago. In the most successful segmentation method, compared with that reported by Ma et al., the average segmentation result of the present biomedical segmentation method was still 2.83 percentage points higher. In the worst case and best cases, the performance of the proposed segmentation method is higher by 8.84 and 2.39 percentage points, respectively. Rotha *et al.* [32] proposed a general method for medical image segmentation. According to the experimental results, the segmentation performance was much poorer than that of the proposed method in pancreatic segmentation. The average

TABLE 2. Comparison of experimental results of the worst case and best case assessed by DSC.

Method	Case DSC (%)	Case DSC (%)	Best Case DSC (%)
RRA-UNet(DICE)	#50 79.88 (Worst)	#4 81.59	#72 93.86
RRA-UNet(CPCE)	#50 86.04	#4 80.50 (Worst)	#72 93.30

TABLE 3. Comparison of experimental results assessed by AUC.

Method	Mean DSC	Worst Case AUC	Best Case AUC
RRA-UNet(DICE)	0.8848	#33 0.7922	#74 0.9536
RRA-UNet(CPCE)	0.8855	#42 0.7924	#74 0.9667

segmentation DSC was 11.35 percentage points lower than that of the proposed method. Furthermore, compared to all the methods, the proposed method has the lowest standard deviation; hence, the proposed segmentation method has a similar segmentation performance on each cross-validation set, thereby demonstrating that the proposed segmentation method has higher stability. Therefore, for the segmentation of the pancreas, the proposed method is a superior solution.

As presented in Table 2, when using the DICE loss function, the segmentation performance on case #50 was the poorest, and after switching to the CPCE loss function, the segmentation performance increased by 6.16 percentage points.

DSC is a commonly used evaluation index for medical image segmentation models. However, we believe that by relying solely on DSC, one cannot fully evaluate the segmentation performance of our model and the role of our loss function in the improvement. Therefore, we also use the area under the ROC curve (AUC) to assess our model and loss function.

According to Table 3, the segmentation performance of our model is highly satisfactory. Moreover, using the CPCE loss function, the mean AUC is increased by 0.0007 compared to the DICE loss function. In case #74, the AUC is increased by 0.0131. The CPCE loss function improves the segmentation performance.

To determine whether our proposed technique avoids overfitting, we also use stratified 10-fold cross-validation. Our experimental results are as follows:

Comparing Table 4 with Table 1 and Table 3, the experimental results differ by a maximum of 0.0008 and a minimum

TABLE 4. Comparison of experimental results assessed by stratified 10-fold cross validation.

Method	Mean DSC(%)	Mean AUC
RRA-UNet(DICE)	88.13	0.8842
RRA-UNet(CPCE)	88.35	0.8863

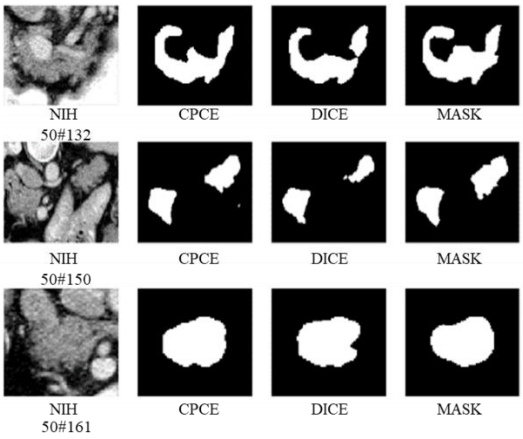

FIGURE 7. Segmentation results of the worst case under the RRA-UNet with the DICE loss function.

of 0.0003. No overfitting occurs in our model, and the experimental results demonstrate that our method is the best-performing method at present.

However, in case #4, which is challenging for the CPCE loss function, the division performance of the CPCE loss function was one percentage point lower than the division performance of the DICE loss function. However, in case #72, both loss functions performed equally well. Hence, the proposed segmentation method is a more effective method for the segmentation of the pancreas. After switching to the CPCE loss function, the algorithm obtains more accurate segmentation results. In Table 1, the average DSC value increased by 0.17 percentage points, while the standard deviation decreased.

In Figure 7, the segmentation results in case #50 are compared. DICE represents RRA-UNet with the Dice-coefficient loss function, while CPCE represents the RRA-UNet segmentation network with the CPCE loss function. The selected slices are located in the middle and back of case #50. When using the DICE function, the slices in each part of case #50 had an overfitting effect, thereby resulting in A lack of segmentation results. However, after switching to the CPCE loss function, the effect of overfitting was reduced. With the reduction, the result of the segmentation was closer to MASK.

In Figure 8, the segmentation results on case #04 are compared. Case #04 is the most challenging case of the RRA-UNet segmentation network using the CPCE loss

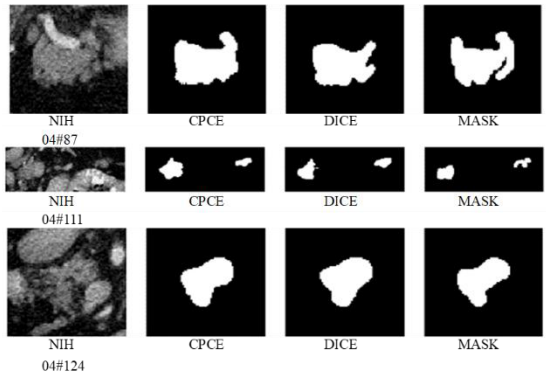

FIGURE 8. Segmentation results of the worst case under the RRA-UNet with the CPCE loss function.

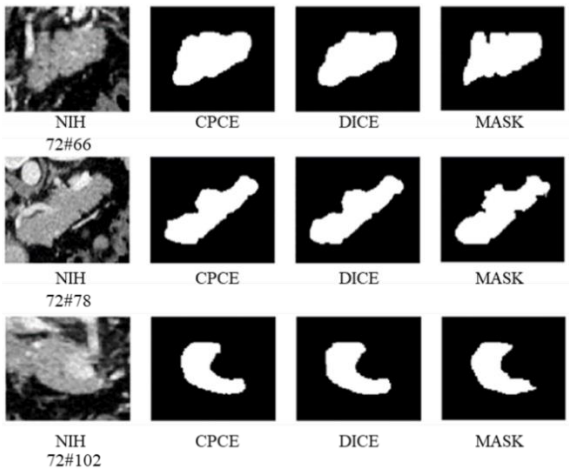

FIGURE 9. Segmentation results of the best case under the RRA-UNet with the DICE loss function and the RRA-UNet with the CPCE loss function.

function. Compared to the DICE loss function, the improved loss function is segmented. Furthermore, the effect of over-segmentation was reduced, which resulted in a smaller calculation of DSC coincidence, while subjectively, the shape of the segmentation was closer to MASK.

In Figure 9, the segmentation network of the two loss functions performed well in case #72, which proves that RRA-UNet outperforms other segmentation networks in pancreas segmentation. At the same time, after using the CPCE loss function, the segmentation result did not have more overfitting, and the segmentation result in the detail part was closer to the real result.

IV. SUMMARY

In pancreas segmentation, manual segmentation and segmentation are not the focus of our research due to their inefficiency. Automatic segmentation is a more reasonable choice, and our automatic segmentation method has been proven to outperform other methods in pancreatic segmentation. Hence, it can be applied to clinical medicine to facilitate doctors when determining a diagnosis, thereby reducing the total workload of doctors. This is the objective of our research.

However, pancreatic segmentation has always been one of the most challenging tasks in medical image segmentation. The main reason is that the pancreatic organ is small. Hence, the area of the CT that corresponds to the pancreas in the abdominal cavity is small, and this area is affected by the compression of the surrounding organs. Furthermore, the shape is highly diverse, and the gray-scale information of the pancreas in CT is similar to that of the surrounding environment, which increases the difficulty of segmentation. In the present study, a new segmentation network is proposed. Based on the UNet segmentation network, the CBAM was used to add an attention mechanism at the time of segmentation, which increased the context linkage in the segmentation network. In addition, the ring residual module was also added to the segmentation model, which made the whole segmentation network look forward and improved the segmentation performance of the network. These experimental results prove that the proposed method is more advanced and accurate than other methods.

Therefore, we believe that continuing along this research direction will be fruitful. In future research, our focus should be on improving the CBAM attention mechanism or replacing CBAM with other attention mechanisms. The method of context measurement will be used to calculate the shape similarity of the image. We think that this approach will more accurately compare the mask and our segmented image, and we will integrate this part of the method into our CPCE loss function.

ACKNOWLEDGMENT

(All the authors contributed equally to this work.)

REFERENCES

[1] M. Havaei, A. Davy, D. Warde-Farley, A. Biard, A. Courville, Y. Bengio, C. Pal, P. Jodoin, and H. Larochelle, "Brain tumor segmentation with deep neural networks," *Med. Image Anal.*, vol. 35, pp. 18–31, Jan. 2017.

[2] Y. Zhou, L. Xie, E. K. Fishman, and A. L. Yuille, "Deep supervision for pancreatic cyst segmentation in abdominal CT scans," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, 2017, pp. 222–230.

[3] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2012, pp. 1097–1105.

[4] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *Proc. Int. Conf. Learn. Representations*, 2015, pp. 1–14.

[5] W. Li, F. Jia, and Q. Hu, "Automatic segmentation of liver tumor in CT images with deep convolutional neural networks," *J. Comput. Commun.*, vol. 3, no. 11, pp. 146–151, 2015.

[6] T. Zhao, D. Gao, J. Wang, and Z. Tin, "Lung segmentation in CT images using a fully convolutional neural network with multi-instance and conditional adversary loss," in *Proc. IEEE 15th Int. Symp. Biomed. Imag. (ISBI)*, Washington, DC, USA, Apr. 2018, pp. 505–509.

[7] A. Mansoor, U. Bagci, Z. Xu, B. Foster, K. N. Olivier, J. M. Elinoff, A. F. Suffredini, J. K. Udupa, and D. J. Mollura, "A generic approach to pathological lung segmentation," *IEEE Trans. Med. Imag.*, vol. 33, no. 2, pp. 2293–2310, Dec. 2014.

[8] 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.

[9] Y. Zheng, A. Barbu, B. Georgescu, M. Scheuering, and D. Comaniciu, "Four-chamber heart modeling and automatic segmentation for 3-D cardiac CT volumes using marginal space learning and steerable features," *IEEE Trans. Med. Imag.*, vol. 27, no. 11, pp. 1668–1681, Nov. 2008.

[10] R. Cuingnet, R. Prevost, D. Lesage, L. D. Cohen, B. Mory, and R. Ardon, "Automatic detection and segmentation of kidneys in 3D CT images using random forests," in *Proc. Int. Conf. Med. Image Comput. Comput. Assist. Intervent.* Berlin, Germany: Springer, 2012, pp. 66–74.
[11] Z. Wang, K. K. Bhatia, B. Glocker, A. Marvao, T. Dawes, K. Misawa, K. Mori, and D. Rueckert, "Geodesic patch-based segmentation," in *Medical Image Computing and Computer-Assisted Intervention—MICCAI*. Cham, Switzerland: Springer, 2014, pp. 666–673.
[12] C. Chu, M. Oda, T. Kitasaka, K. Misawa, M. Fujiwara, Y. Hayashi, Y. Nimura, D. Rueckert, and K. Mori, "Multi-organ segmentation based on spatially-divided probabilistic atlas from 3D abdominal CT images," in *Medical Image Computing and Computer-Assisted Intervention—MICCAI*. Berlin, Germany: Springer, 2013, pp. 165–172.
[13] R. Wolz, C. Chu, K. Misawa, M. Fujiwara, K. Mori, and D. Rueckert, "Automated abdominal multi-organ segmentation with subject-specific atlas generation," *IEEE Trans. Med. Imag.*, vol. 32, no. 9, pp. 1723–1730, Sep. 2013.
[14] H. R. Roth, L. Lu, A. Farag, H.-C. Shin, J. Liu, E. B. Turkbey, and R. M. Summers, "DeepOrgan: Multi-level deep convolutional networks for automated pancreas segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput. Assist. Intervent.* Cham, Switzerland: Springer, 2015, pp. 556–564.
[15] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proc. IEEE Comput. Vis. Pattern Recognit.*, Jun. 2015, pp. 3431–3440.
[16] 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.* Cham, Switzerland: Springer, 2015, pp. 234–241.
[17] Y. Zhou, L. Xie, W. Shen, Y. Wang, E. K. Fishman, and A. L. Yuille, "A fixed-point model for pancreas segmentation in abdominal CT scans," in *Proc. Int. Conf. Med. Image Comput. Comput. Assist. Intervent.*, 2017, pp. 693–701.
[18] H. R. Roth, L. Lu, A. Farag, A. Sohn, and R. M. Summers, "Spatial aggregation of holistically-nested networks for automated pancreas segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput. Assist. Intervent.*, 2016, pp. 451–459.
[19] J. Cai, L. Le, F. Xing, and L. Yang, "Pancreas segmentation in CT and MRI images via domain specific network designing and recurrent neural contextual learning," Mar. 2018, *arXiv:1803.11303*. [Online]. Available: https://arxiv.org/abs/1803.11303
[20] Z. Zhu, Y. Xia, W. Shen, E. Fishman, and A. Yuille, "A 3D coarse-to-fine framework for volumetric medical image segmentation," in *Proc. Int. Conf. 3D Vis.*, Verona, Italy, Sep. 2018, pp. 682–690.
[21] J. Ma, F. Lin, S. Wesarg, and M. Erdt, "A novel Bayesian model incorporating deep neural network and statistical shape model for pancreas segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput. Assist. Intervent.*, 2018, pp. 480–487.
[22] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *Proc. Comput. Vis. Pattern Recognit. (CVPR)*, 2009, pp. 1–8.
[23] A. Krizhevsky and G. Hinton, "Learning multiple layers of features from tiny images," Univ. Toronto, Toronto, ON, Canada, Tech. Rep., vol. 1. no. 4, 2009.
[24] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick, "Microsoft COCO: Common objects in context," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2014, pp. 740–755.
[25] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, Nov. 1998.
[26] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*. [Online]. Available: https://arxiv.org/abs/1409.1556
[27] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.
[28] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," 2016, *arXiv:1610.02357*. [Online]. Available: https://arxiv.org/abs/1610.02357
[29] S. Xie, R. Girshick, P. Dollár, Z. Tu, and K. He, "Aggregated residual transformations for deep neural networks," 2016, *arXiv:1611.05431*. [Online]. Available: https://arxiv.org/abs/1611.05431
[30] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "CBAM: Convolutional block attention module," 2018, *arXiv:1807.06521*. [Online]. Available: https://arxiv.org/abs/1807.06521
[31] J. Cai, L. Lu, Y. Xie, F. Xing, and L. Yang, "Improving deep pancreas segmentation in CT and MRI images via recurrent neural contextual learning and direct loss function," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Honolulu, HI, USA, 2017, pp. 1–8.
[32] H. R. Rotha, H. Odaa, X. Zhoub, N. Shimizua, Y. Yanga, Y. Hayashia, M. Odaa, M. Fujiwarac, K. Misawad, and K. Moria, "An application of cascaded 3D fully convolutional networks for medical image segmentation," *Comput. Med. Imag. Graph.*, vol. 66, pp. 90–99, Jun. 2018.
[33] H. R. Roth, L. Lu, N. Lay, A. P. Harrison, A. Farag, and R. M. Summers, "Spatial aggregation of holistically-nested convolutional neural networks for automated pancreas localization and segmentation," *Med. Image Anal.*, vol. 45, no. 94, pp. 94–107, Apr. 2018.
[34] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proc. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 1–9.

LIN LU is currently pursuing the M.S. degree in computer technology with Chongqing University. His research interests include image processing and pattern recognition.

LIQIONG JIAN was born in 1982. She received the B.S. degree in computer science and technology from Ningxia University, Yinchuan, China, in 2005. She is currently working with the Ningxia Blood Center. Her research interests mainly include medical information processing and management.

JUN LUO received the B.S. degree from Xi'an Jiaotong University, Xi'an, China, in 1990. He is currently working as an Associate Professor with Chongqing University, China. His research interests include large medical system design, networks, and database.

BIN XIAO was born in 1982. He received the B.S. and M.S. degrees in electrical engineering from Shaanxi Normal University, Xi'an, China, in 2004 and 2007, respectively, and the Ph.D. degree in computer science from Xidian University, Xi'an, China. He is currently working as an Associate Professor with the Chongqing University of Posts and Telecommunications, Chongqing, China. His research interests include image processing, pattern recognition, and digital watermarking.