ResD-Unet Research and Application for Pulmonary Artery Segmentation

HONGFANG YUAN[1], ZHENHONG LIU [ID][1], YAJUN SHAO[1], AND MIN LIU[2]

[1]College of Information Science and Technology, Beijing University of Chemical Technology, Beijing 100029, China
[2]Department of Radiology, China-Japan Friendship Hospital, Beijing 100029, China

Corresponding author: Hongfang Yuan (yuanhf@mail.buct.edu.cn)

ABSTRACT In the three-dimensional reconstruction of the pulmonary artery and the identification of pulmonary embolism, experts find it difficult to accurately estimate the severity of the embolism in the pulmonary artery, due to its irregular shape and complex adjacent tissues. In effect, segmenting the pulmonary artery accurately is the basis for assessing the severity of pulmonary embolism, and it is also a challengeable task. To solve this problem, this study proposes a ResD-Unet architecture for pulmonary artery segmentation. To begin with, the U-Net network is used as the basic structure, which allows efficient information flow and good performance in the absence of a sufficiently large dataset. In what follows, novel Residual-Dense blocks are introduced in the ResD-Unet architecture to refine image segmentation and build a deeper network while improving the gradient circulation of the network. Finally, a novel hybrid loss function is utilized to make full use of the advantages of the binary cross entropy loss, Dice loss and SSIM loss. Equipped with the hybrid loss, the proposed architecture is able to effectively segment the object areas and accurately predict the structures with clear boundaries. The experimental results show that the proposed framework can achieve high segmentation accuracy and efficiency, and the segmentation results are comparable to that of manual segmentation.

INDEX TERMS Neural network, ResD-Unet, Residual-dense block, image segmentation, deep learning.

I. INTRODUCTION

Pulmonary embolism (PE) refers to the pathological and clinical conditions caused by impacting substances entering the pulmonary artery and blocking the blood supply to the tissues, and its morbidity is only lower than that of coronary heart disease and hypertension. Early detection, early diagnosis and timely treatment are the keys to effectively reduce the risk of death. In recent years, contrast-enhanced Computed Tomography (CT) is the most commonly used modality for PE screening [1]. Three-dimensional visualization is a technology that uses two-dimensional image sequence to reconstruct three dimensional model and perform analysis. The 3D construction of lung is achieved by stacking the evolved contours of individual slices over one another [2]. Compared with two-dimensional (2D) images, the 3D visualization of pulmonary artery CT images can provide more information

to predict the morphological changes of the pulmonary artery, thereby bringing convenience to the diagnosis of PE. Since PE exists only in the pulmonary artery, the segmentation of pulmonary artery is particularly important. On the one hand, manual segmentation of pulmonary artery from a CT sequence is very time-consuming and easy to misdiagnose [3]. On the other hand, the segmentation accuracy of the pulmonary artery is of great significance to the three-dimensional reconstruction of the pulmonary artery, which can help physicians diagnose the embolism and evaluate the effectiveness of embolism treatments. Therefore, automatic segmentation methods are highly desirable in clinical practice.

In the early days, people followed simple feature-based methods, such as regional growth methods, active contour models and gradient vector flow [4]. Since the grayscale of pulmonary artery is similar to surrounding tissues and the structure is complex, the automatic segmentation of pulmonary artery is a realistic and challenging issue.

The associate editor coordinating the review of this manuscript and approving it for publication was Valentina E. Balas [ID].

Nevertheless, existing traditional methods of medical image segmentation have shown limited success in solving this difficult problem. At present, owing to the increasing popularity of deep neural networks, scholars have gradually adopted deep learning methods to solve complex problems in the field of medical images. The use of deep networks has shown significant advances and reliable results in many computer vision tasks, including image segmentation, target recognition, motion tracking and image classification [5]. U-Net is an image segmentation network proposed in 2015 by Ronneberger *et al.* [6]. It has been widely studied and applied in the field of medical image segmentation by virtue of its simple structure, strong generalization and strong segmentation abilities. In this process, many variants of U-Net have been proposed. Liu *et al.* [7] proposed an IU-Net architecture that directly applied the maximum pooling layer to the deconvolution layer, which reduced the loss of image features and achieved better results than U-Net. Zhou *et al.* [8] proposed a new general purpose image segmentation architecture UNet++. Attributed to its nested structure and re-designed skip connections, it demonstrates well performance over various state-of-the-art backbones. Xu *et al.* [9] proposed a new liver segmentation framework using ResUnet with geometric post-processing. Gu *et al.* [10] proposed a context encoder network named CE-Net, which is used to obtain higher-level information and preserve spatial information for 2D medical image segmentation.

When training the neural network model, as the network deepening, the model will encounter problems such as slow convergence and vanishing gradient, which will affect the accuracy of segmentation [11]. Besides, the fuzzy edges of the pulmonary artery and the variability of its shape and size make it prone to mis-segmentation and low segmentation accuracy when using U-Net for pulmonary artery segmentation. Based on this, we introduce a novel framework for pulmonary artery segmentation, termed ResD-Unet. The framework that includes a new hybrid loss function performs well. The main contributions of this paper are as follows:

(1) A novel semantic segmentation architecture is proposed that can address problems faced by the traditional U-Net network, i.e., vanishing gradients and an insufficient feature extraction. The architecture is mainly composed of Residual-Dense block which combines the advantages of residual connections and dense connections.

(2) A new hybrid loss function that combines binary cross entropy (BCE) loss, Dice loss and structural similarity (SSIM) loss is used. It makes full use of the advantages of the three loss functions, i.e., smoothing gradients, handling imbalanced categories, and thinning boundaries.

II. METHODOLOGY

A. NETWORK FOUNDATION
The ResD-Unet network proposed in this paper is based on U-Net structure. U-Net is one of the most well-known deep learning networks with an encoder-decoder architecture, it is widely used in the field of medical image segmentation.
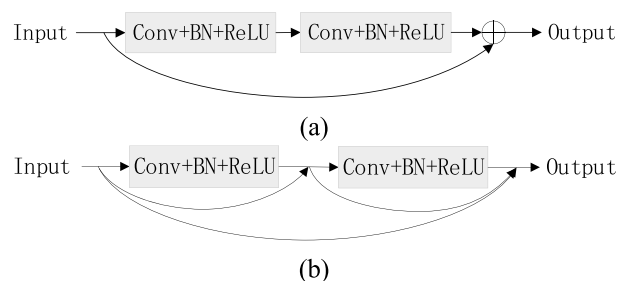

FIGURE 1. Convolution module of network structure.

Although in the absence of a large number of data sets, U-Net can also achieve good performance.

The concept of "the deeper the better" is regarded as the key principle of neural networks. Training deep neural networks with increased network depth can improve accuracy. However, when dealing with deep architecture, it will hinder convergence during training and cause vanishing gradients. In order to solve these limitations, ResNet is proposed by Hang *et al.* [12]. The residual connection (as shown in Fig 1(a)) is implemented in the form of a skip connection, where the input of the unit is directly added to the output of the unit and activated. The residual network solves the problem of gradient disappearance when the network is deepened and it also converges faster under the premise of the same number of layers. The dense connection (as shown in Fig 1(b)) is derived from the densely-connected network (Dense Net) [13]. For dense connection, any layer is added direct connection to all subsequent layers in a feed-forward manner to maximize the information flow in the network and make training easier. This is motivated by three observations. First, each layer in the network accepts the features of all the layers before it as input, helps to improve the flow of information and gradients throughout the entire network. Secondly, there is an implicit deep supervision thanks to short paths to all feature maps in the architecture. And third, dense connection has the effect of regularization, which can reduce the overfitting on training sets. However, excessive use of dense connection can increase network complexity.

Drawing lessons from residual connection and dense connection, we propose Residual-Dense blocks (as shown in Fig 2), which attenuate to a great extent the problem of degradation and vanishing gradients that are present in deep architectures. To achieve optimal network performance, the ordinary convolution blocks of the classical U-Net architecture are replaced with modified Residual-Dense blocks to form ResD-Unet.

B. RESD-UNET ARCHITECTURE
To fully extract high-level intra-slice features, we designed an efficient deep network called ResD-Unet, which combines the advantages of residual connections [12], dense connections [13] and U-Net [6]. Residual-Dense connection adds the features of all the previous layers to the bottom layer to achieve feature reuse and facilitate the back-propagation of the gradient during the training process. Besides, it effectively
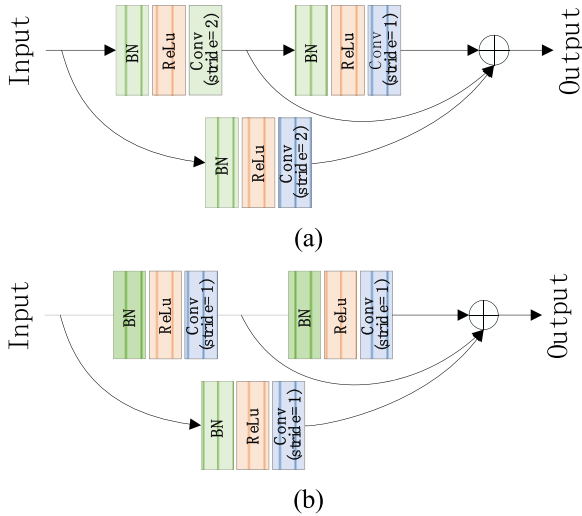
FIGURE 2. Residual-dense blocks. (a) Encode block; (b) Decoder block.

solves the problem of information loss and avoids the disappearance of gradients. The use of batch normalization (BN) layer [14] avoids overfitting and speeds up the training and convergence of the network. The above-mentioned strategies can make the network deeper, thus faster to converge.

An overview of ResD-Unet architecture is shown in Fig 3. It can be divided into three parts, namely, the encoding path, the central part and the decoding path. The left half of the network is the encoding path, while the right one is the decoding path. The encoding path consists of Encoder-blocks (as shown in Fig 2(a)). Its backbone consists of two successive 3×3 convolution blocks and an identity mapping. Each convolution block includes a batch normalization layer, a rectified linear unit (ReLU) [15] activation layer and two convolution layers with the stride of 2 and 1 respectively. A convolutional layer with the stride of 2 and a BN layer are used in identity mapping to connect the input and output of the Encoder-block. Then, the output of the first convolution block is connected to the output of the Encoder-block. In this paper, a convolutional layer with the stride of 2 is used to reduce the spatial dimension of the feature maps by half, rather than the pooling layer. Some of the latest network architectures [16] have demonstrated that with this approach, network performance can be roughly consistent, or even slightly improved. Residual-block contains two successive 3×3 convolutional layers, and the input of the block is added to the output. The structure of the Decoder-block (as shown in Fig 2(b)) is similar to the Encoder-block, but the stride of all convolutional layers in it is 1. At the beginning of each Decoder-block, there is an up-sampling of the lower-level feature map.

We increase network layers for more precise positioning. We construct ResD-Unet with four Encoder-blocks and Decoder -blocks instead of three as in [17]. At the end of the network, a 1×1 convolution and a sigmoid activation layer are used to convert the channel map into a feature map with two categories.

C. LOSS FUNCTION

The hybrid loss function is proposed to improve the similarity between the segmented image and the labelled image,
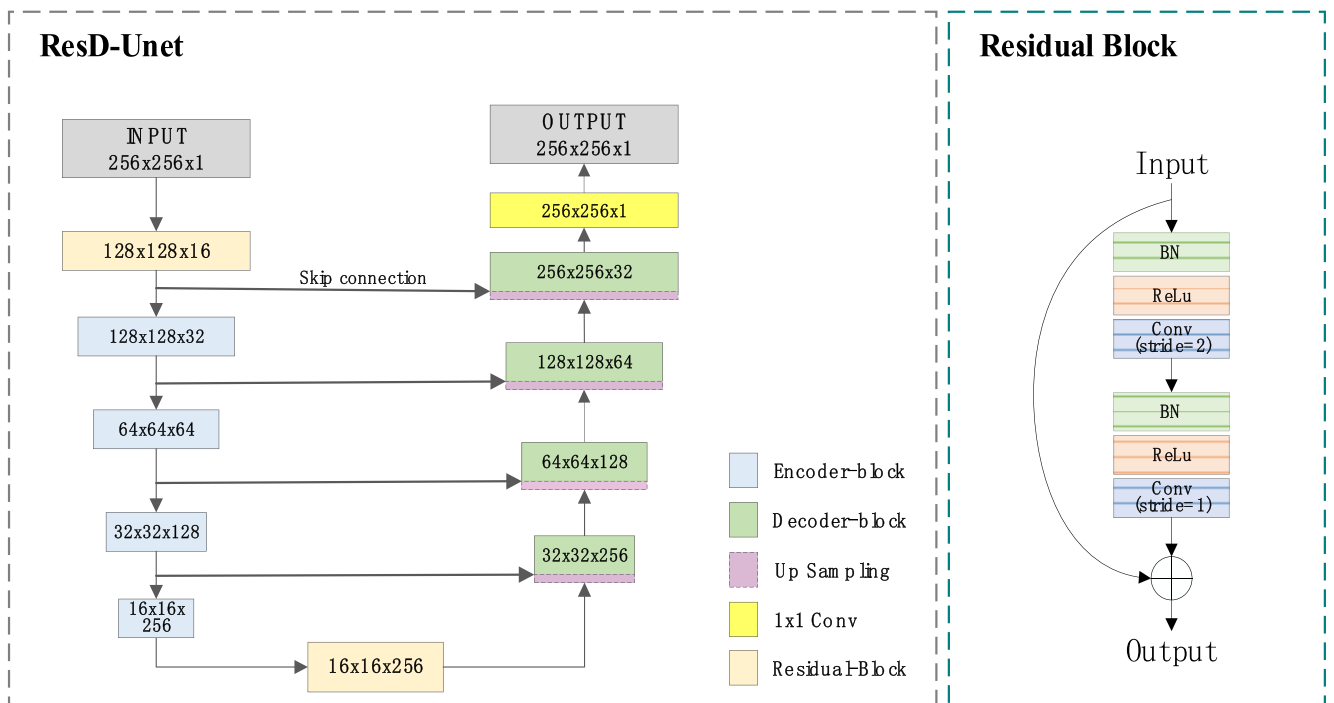

FIGURE 3. The ResD-Unet architecture.

which is defined as:

$$Loss = \beta_1 Loss_{Dice} + \beta_2 Loss_{CE} + \beta_3 Loss_{SSIM} \qquad (1)$$

Among them, $Loss_{Dice}$, $Loss_{CE}$ and $Loss_{SSIM}$ represent the Dice loss function [18], BCE loss function [19] and SSIM loss function [20] respectively, and β represents each hyper-parameter of the loss function.

Dice coefficient is a similarity measurement, and is quite effective in dealing with imbalanced categories. It is defined as follows:

$$Loss_{Dice} = 1 - \frac{2\sum_i^N p_i g_i}{\sum_i^N P_i^2 + \sum_i^N g_i^2} \qquad (2)$$

where N denotes the sum of pixels, p_i is the predicted label and g_i is the ground-truth label, $p_i, g_i \in [0, 1]$.

Dice loss is sometimes unstable during the training process, which is not conducive to network convergence [21]. Especially when p_i and g_i are too small, the calculated gradient of Dice changes drastically. BCE loss is widely used in neural network training by calculating the similarity between the ground-truth label and the predicted distribution. The cross entropy is used to stabilize training. Its formula is as follows:

$$Loss_{CE} = -\frac{1}{N} \sum_i^N (g_i \log p_i + (1 - g_i) \log(1 - p_i)) \quad (3)$$

where p_i is the predicted label, g_i is the ground truth label, $p_i, g_i \in [0, 1]$.

SSIM is used for image quality evaluation. It captures the structure and contrast information in an image. When it is integrated into the loss function of the network, the structural information of the annotated image can be obtained. $x = \{x_j : j = 1, \ldots, N^2\}$ and $y = \{y_j : j = 1, \ldots, N^2\}$ denote the pixel values of the corresponding regions of the segmented picture and the expert-labelled picture respectively. N is the number of pixels. SSIM is defined as:

$$SSIM = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \qquad (4)$$

$$Loss_{SSIM} = 1 - \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \qquad (5)$$

Among them, μ_x, μ_y and σ_x, σ_y are the mean and standard deviation of x and y, respectively. σ_{xy} represents their covariance. C1 and C2 are small constants chosen to prevent instability when the denominator is equal to zero. In our paper, C1 and C2 are set as 0.04 and 0.088 respectively. SSIM loss is a patch-level metric that considers the local neighborhood of each pixel. It gives the border a higher weight. In other words, the loss is greater near the border, even if the predicted probabilities on the boundary and the rest of the foreground are the same. The loss on the boundary is the largest at the beginning of training [20]. Therefore, the use of SSIM helps to optimize the boundary features.

Dice coefficient evaluates the degree of overlap between our network segmentation output and the ground truth label,

which is our ultimate optimization goal. BCE loss maintains a smooth gradient for all pixels and helps the convergence of the loss function. SSIM loss function is used to optimize the segmentation details. Therefore, the combination of these three losses can make full use of their advantages, which can deal with unbalanced categories, smooth gradients, optimize the segmentation details, and ultimately obtain higher network performance.

III. EXPERIMENT

The lung CT images provided by the China-Japan Friendship Hospital and the public dataset CHAOS were utilized in the experiment to train and test ResD-Unet architecture, and finally evaluate its performance. To make the results more reliable, we also compared the performance of ResD-Unet model with other models.

A. DATASET

To evaluate ResD-Unet, we used the lung CT image sequence data set provided by the China-Japan Friendship Hospital. The data set consists of a total of about 4000 CT images of 80 patients, and the corresponding ground truth mask was marked by professional physicians. This dataset is not public.

The second data set we used was the CHAOS, which is a public data set. CHAOS [22] challenge was held in the IEEE International Symposium on Biomedical Imaging (ISBI) in April, 2019 Venice, ITALY. Two databases (Abdominal CT and MRI) were used in this challenge but we only choose one, i.e., Liver Segmentation (CT only), which corresponds to a series of DICOM images belonging to a single patient and contains 4,282 slices of CT images of 40 different patients.

B. PRE-PROCESSING

The input images in this article are two-dimensional CT images, and the format of the original CT images are DICOM [23]. The value range of Hounsfield is large, and the target area is not obvious. Therefore, we preprocessed each image in the CT sequence. Firstly, we read all DICOM files and extracted the pixel values. Then, the pixel values were normalized to the [0, 1] interval, and were stored in the. BMP format. Finally, the histogram equalization was performed on the image, which can enhance the contrast of the organ boundary. Fig 4 illustrates a flowchart of image preprocessing, and Fig 5 gives an example of this preprocessing method.

C. IMPLEMENTATION DETAILS

The experimental environment of this research is Linux. We ran our experiments on a NVIDIA Tesla V100 system with 256G of memory. All models are implemented using Keras framework [24] with Tensorflow 2.1.0 [25] as backend. In the meantime, the proposed architecture is optimized by Adam (adaptive moment estimation) [26] optimizer. The initial learning rate of the neural network is 2e-4. The batch size is set to be four. For the lung CT dataset, the hyperparameters $\beta1$, $\beta2$, and $\beta3$ of the loss function are set as 0.6, 0.3, 0.1 respectively through many experiments and

FIGURE 4. Flow chart of data preprocessing.

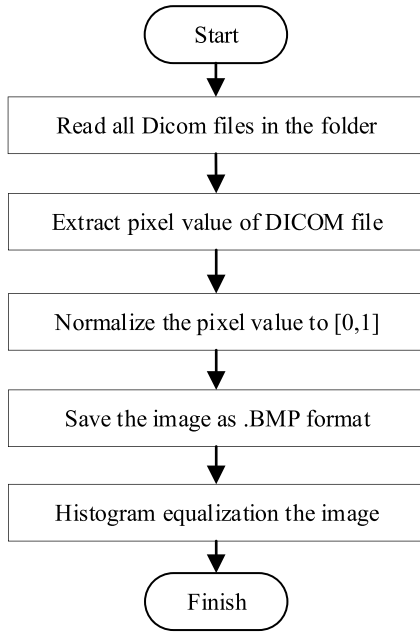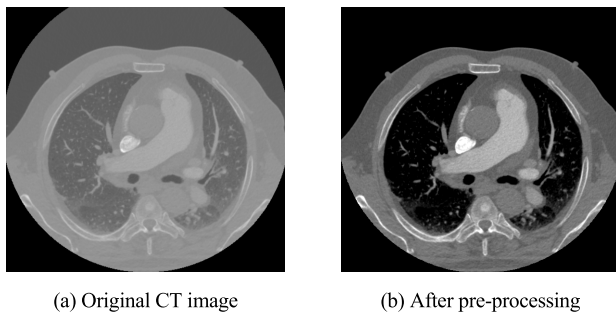(a) Original CT image   (b) After pre-processing

FIGURE 5. Sample of CT image after preprocessing.

adjustments, and the number of epochs is set as 100. For the CHAOS dataset, the hyper-parameters $\beta1$, $\beta2$, and $\beta3$ are set as 0.7, 0.2, 0.1 respectively through many experiments and adjustments, and the number of epochs is set as 120. In both datasets, we used 80% of dataset for training, 10% for validation, and 10% for testing. The size of input images is 256 × 256.

D. RESULTS AND COMPARISONS

To make the results more reliable and illustrate ResD-Unet's high performance in pulmonary artery segmentation, the dataset were segmented simultaneously by the following models, namely the original U-Net, ResUnet, DenseUnet as in [27], CE-Net as in [10], UNet++ as in [8] and ResD-Unet.

Fig 6 shows the results of different networks to segment the pulmonary artery. Fig 6(b) is a mask generated from the labelled file, representing the label value. Fig 6(f) represents the result of the prediction of the CE-Net model as in [9]. Fig 6(g) concerns the result of the prediction of UNet++. Fig 6(h) is the segmentation result of ResD-Unet. It can

be seen from Fig 6 that the segmentation results obtained by ResD-Unet model designed in this paper (Fig 6(h)) are very close to that of the expert annotations (Fig 6(b)). Compared with other algorithms, this method can effectively display pulmonary artery boundary details and optimize the flaws of other algorithms, e.g., over-segmentation and under-segmentation.

To further verify the effectiveness of ResD-Unet, we also conducted experiments on the public data set CHAOS.

Fig 7 displays the segmentation results on CHAOS data set. Fig 7(a) represents the input image. Fig 7(b) is the mask generated from the labelled file, representing the label value. Fig 7(c) represents the segmentation result of U-Net. Fig 7(d) is the segmentation result of ResUnet. Fig 7(e) represents the segmentation result of DenseUnet. Fig 7(f) represents the result of the prediction of the CE-Net. Fig 7(g) concerns the result of the prediction of UNet++. Fig 7(h) is the result of the prediction of ResD-Unet model in this paper. It can be seen from the figure that the model proposed in this paper can accurately segment in the range of lung. Segmentation effects of CE-Net and UNet++ are almost as good as that of ResD-Unet, but they need some subtle improvements. Other algorithms can misidentify a small amount of soft tissue as lung. The segmentation results of ResD-Unet architecture (as shown in Fig 7(h)) is high in quality and has good effect, which is better than U-Net, ResUnet, CE-Net and DenseUnet. Its results have a high degree of similarity with the results marked by the physician (As shown in Fig 7(b)).

After building a model, it is of utmost importance to evaluate its performance. There is a peculiar high class unbalance, and the selection of the correct metrics is critical to resolve this problem. According to [28], recall and precision are suitable and useful metrics when working with unbalanced classes. Recall discloses how many relevant samples are selected. Precision unveils how many predicted samples are relevant. They are based on true positives TP (i.e., the sample label is positive and is classified as positive), false positives FP (i.e., the sample label is negative, but is classified as positive) and false negatives FN (i.e., the sample label is positive, but is classified as negative) [29].

Based on the above values, the recall and precision indexes can be calculated.

Recall is defined as follows:

$$\mathrm{Re}\,call = \frac{TP}{TP + FN} \tag{6}$$

Precision is defined as follows:

$$\mathrm{Pr}\,ecision = \frac{TP}{TP + FP} \tag{7}$$

To make the results more convincing, SSIM value and Dice coefficient are also used to evaluate the results [30]. SSIM is a metric used to capture structure and contrast information in an image for image quality assessment, its calculation formula is shown in equation (4). We divide the image into N blocks with a sliding window, calculate the structural similarity SSIM of each block, and finally take the average
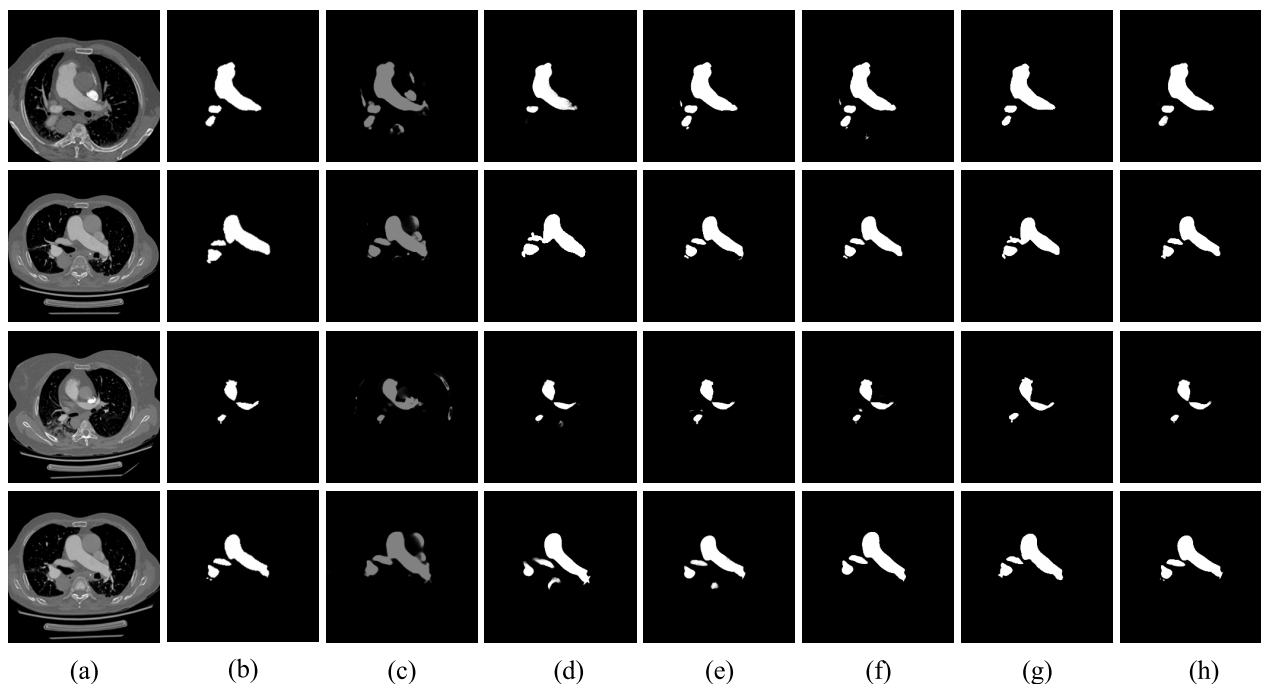
FIGURE 6. Segmentation results of pulmonary artery with different algorithms. (a) Input image; (b) mask; (c) U-Net; (d) ResUnet; (e) DenseUnet; (f) CE-Net; (g) UNet++;(h) ResD-Unet FIGURE 6. Segmentation results of pulmonary artery with differ.
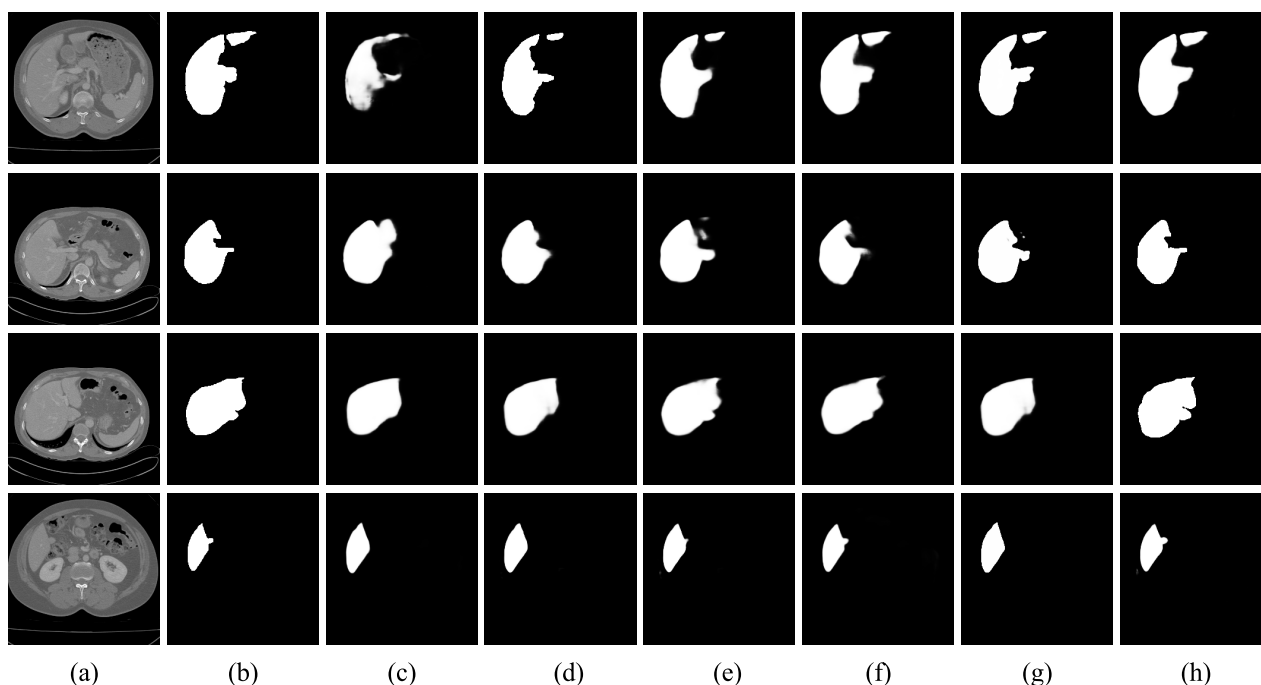

FIGURE 7. Segmentation results of lung with different algorithms. (a) Input image; (b) mask; (c) U-Net; (d) ResUnet; (e) DenseUnet; (f) CE-Net; (g) UNet++;(h) ResD-Unet.

value as the structural similarity measure of the ground true image and the predicted image. Dice coefficient is a similarity measurement, usually used to calculate the similarity of two samples. In image segmentation, this formula can be refined as:

$$DSC = \frac{2\,|G \cap S|}{|G| + |S|} \tag{8}$$

where $|G|$ represents the pixels of the ground truth image, $|S|$ represents the pixels of the segmented image. $|G \cap S|$ represents the intersect pixels of two images.

When experimenting with lung CT image dataset, the performances of different models are shown in Table 1.

Then we apply the proposed architecture to CHAOS dataset. The performances of different models are shown in Table 2.

TABLE 1. The performance comparison of different algorithms on lung CT image dataset.

Architecture	DSC	Precision	Recall	SSIM
U-Net	0.949	0.958	0.965	0.922
ResUnet in[12]	0.963	0.969	0.972	0.937
CE-Net in[10]	0.970	0.961	0.979	0.934
DenseUnet in[27]	0.981	0.980	0.977	0.941
UNet++ in[8]	0.976	0.979	0.978	0.958
ResD-Unet	**0.982**	**0.985**	**0.980**	**0.961**

TABLE 2. The performance comparison of different algorithms on CHAOS dataset.

Architecture	DSC	Precision	Recall	SSIM
U-Net	0.931	0.928	0.925	0.909
ResUnet in[12]	0.942	0.944	0.943	0.918
CE-Net in[10]	0.962	0.965	0.958	0.923
DenseUnet in[27]	0.965	0.966	0.961	0.932
UNet++ in[8]	0.961	0.959	0.966	**0.952**
ResD-Unet	**0.969**	**0.966**	**0.968**	0.951

TABLE 3. The Running time comparison of four algorithms on lung CT dataset.

Architecture	Epochs	Running time(Seconds)
DenseUnet	100	2198.9
CE-Net	100	1633.4
UNet++	100	2276.1
ResD-Unet	100	**1096.7**

It can be seen from Table 1 and Table 2 that our ResD-Unet architecture shows pretty good performance in DSC, Recall, Precision, and SSIM, thus verifying the effectiveness of the proposed framework. From the last three rows of Table 1 and Table 2, it can be seen that the segmentation effect of Dense-Unet and UNet++ are almost as good as that of ResD-Unet. However, DenseUnet and UNet++ use dense connection, which is time-consuming when converging the model and may lead to high computational complexity. In addition, dense connection also requires more complex hardware conditions.

We compared the running time of ResD-Unet architecture with CE-Net, UNet++ and DenseUnet architecture on the lung CT dataset in Table 3. It can be seen that the running time of DenseUnet and UNet++ are almost two times that of ResD-Unet. Our ResD-Unet architecture can still complete the segmentation well without increasing the network complexity. Compared with other networks, the method proposed in this paper is higher in accuracy and efficiency. Therefore, the ResD-Unet architecture combined with Residual-Dense module and the hybrid loss function of boundary refinement can effectively optimize the segmentation results and obtain more accurate segmentation images.

IV. CONCLUSION

In this study, a ResD-Unet architecture was proposed to segment the pulmonary artery region accurately. In the very beginning, to solve U-Net's problem of insufficient feature extraction, we added Residual-Dense blocks to the feature extraction layer composed of the original fixed-scale convolution kernel, which to a great extent resolved the problems in deep architectures, i.e., network degradation and vanishing gradients. Then, a BN layer was added to the network to accelerate model convergence and enhance the model's generalization ability. Finally, a hybrid loss function was proposed to give higher weight to boundary pixels and improve the definition of boundary segmentation. The experimental results showed that the ResD-Unet network combined with the boundary thinning hybrid loss function and Residual-Dense module can segment the pulmonary arteries of the lung CT images accurately, which facilitated the subsequent research about three-dimensional reconstruction of the pulmonary arteries and the evaluation of the severity of PE.

REFERENCES

[1] F. G. Meinel, J. W. Nance, U. J. Schoepf, V. S. Hoffmann, K. M. Thierfelder, P. Costello, S. Z. Goldhaber, and F. Bamberg, "Predictive value of computed tomography in acute pulmonary embolism: Systematic review and meta-analysis," *Am. J. Med.*, vol. 128, pp. 747–759, Jul. 2015.

[2] J. Li, H. Wang, and L. Song, "A novel sparse feature extraction method based on sparse signal via dual-channel self-adaptive TQWT," *Chin. J. Aeronaut.*, 2020, doi: 10.1016/j.cja.2020.06.013.

[3] X. Han, "Automatic liver lesion segmentation using a deep convolutional neural network method," in *Proc. Comput. Vis. Pattern Recognit.*, 2017.

[4] C. Xu and J. L. Prince, "Snakes, shapes, and gradient vector flow," *IEEE Trans. Image Process.*, vol. 7, no. 3, pp. 359–369, Mar. 1998.

[5] H. Zhang, K. Dana, J. Shi, Z. Zhang, X. Wang, A. Tyagi, and A. Agrawal, "Context encoding for semantic segmentation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 7151–7160.

[6] 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.*, in Lecture Notes in Computer Science, vol. 9351, 2015, pp. 234–241.

[7] Z. Liu, X. L. Zhang, Y. Zhu, D. Q. Yuan, and Y. Q. Song, "Liver segmentation with improved U-Net and Morphsnakes algorithm," *J. Image Graph.*, vol. 23, pp. 1254–1262, Aug. 2018.

[8] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, and J. Liang, "UNet++: Redesigning skip connections to exploit multiscale features in image segmentation," *IEEE Trans. Med. Imag.*, vol. 39, no. 6, pp. 1856–1867, Jun. 2020.

[9] W. Xu, H. Liu, X. Wang, and Y. Qian, "Liver segmentation in CT based on ResUNet with 3D probabilistic and geometric post process," in *Proc. IEEE 4th Int. Conf. Signal Image Process. (ICSIP)*, Jul. 2019, pp. 685–689.

[10] Z. Gu, J. Cheng, H. Fu, K. Zhou, H. Hao, Y. Zhao, T. Zhang, S. Gao, and J. Liu, "CE-Net: Context encoder network for 2D medical image segmentation," *IEEE Trans. Med. Imag.*, vol. 38, no. 10, pp. 2281–2292, Oct. 2019.

[11] A. Pezeshk, S. Hamidian, N. Petrick, and B. Sahiner, "3-D convolutional neural networks for automatic detection of pulmonary nodules in chest CT," *IEEE J. Biomed. Health Inform.*, vol. 23, no. 5, pp. 2080–2090, Sep. 2019.

[12] Z. Zhang, Q. Liu, and Y. Wang, "Road extraction by deep residual U-Net," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 5, pp. 749–753, May 2018.

[13] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 4700–4708.

[14] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," 2015, *arXiv:1502.03167*. [Online]. Available: http://arxiv.org/abs/1502.03167

[15] F. Tian, J. Wang, and J. Li, "A deep convolutional learning method for blind recognition of channel codes," *J. Phys., Conf. Ser.*, vol. 1621, Aug. 2020, Art. no. 012088.

[16] J. T. Springenberg, A. Dosovitskiy, T. Brox, and M. Riedmiller, "Striving for simplicity: The all convolutional net," 2014, *arXiv:1412.6806*. [Online]. Available: http://arxiv.org/abs/1412.6806

[17] X. Han, "Automatic liver lesion segmentation using a deep convolutional neural network method," 2017, *arXiv:1704.07239*. [Online]. Available: http://arxiv.org/abs/1704.07239

[18] C. H. Sudre, W. Li, T. Vercauteren, S. Ourselin, and M. J. Cardoso, "Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations," in *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*. Cham, Switzerland: Springer, 2017, pp. 240–248.

[19] P.-T. D. Boer, D. P. Kroese, S. Mannor, and R. Y. Rubinstein, "A tutorial on the cross-entropy method," *Ann. Oper. Res.*, vol. 134, no. 1, pp. 19–67, Feb. 2005.

[20] X. Qin, Z. Zhang, C. Huang, C. Gao, M. Dehghan, and M. Jagersand, "BASNet: Boundary-aware salient object detection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 7479–7489.

[21] (2018). *Dice-Coefficient Loss Function vs Cross Entropy*. [Online]. Available: https://stats.stackexchange.com/questions/321460/dice-coefficient-loss-function-vs-cross-entropy

[22] (2019). *CHAOS-Combined (CT-MR) Healthy Abdominal Organ Segmentation*. [Online]. Available: https://chaos.grand-challenge.org/Combined_Healthy_Abdominal_Organ_Segmentation/

[23] L. G. Wang, W. Q. Liu, Y. H. Yu, and G. Z. Wang, "Research on the file format analysis and application of DICOM medical image," *Comput. Eng. Appl.*, vol. 42, no. 29, pp. 210–215, 2006.

[24] A. Li, Y.-X. Li, and X.-H. Li, "Tensor flow and keras-based convolutional neural network in CAT image recognition," in *Proc. 2nd Int. Conf. Comput. Modeling, Simulation Appl. Math. (CMSAM)*. Science and Engineering Research Center, 2017, p. 5.

[25] M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving, M. Isard, and M. Kudlur, "Tensorflow: A system for large-scale machine learning," in *Proc. USENIX Symp. Operating Syst. Design Implement. (OSDI)*, 2016, pp. 28–265.

[26] S. Andradóttir, "A method for discrete stochastic optimization," *Manage. Sci.*, vol. 41, no. 12, 1995.

[27] A. Kaku *et al.*, "DARTS: DenseUnet-based automatic rapid tool for brain segmentation," 2019.

[28] A. Tharwat, "Classification assessment methods," in *Applied Computing and Informatics*. Riyadh, Saudi Arabia: King Saud Univ., 2018.

[29] H. Xue, D. Ding, Z. Zhang, M. Wu, and H. Wang, "A fuzzy system of operation safety assessment using multi-model linkage and multi-stage collaboration for in-wheel motor," *IEEE Trans. Fuzzy Syst.*, 2021, doi: 10.1109/TFUZZ.2021.3052092.

[30] Z. Zeng, W. Xie, Y. Zhang, and Y. Lu, "RIC-UNet: An improved neural network based on UNet for nuclei segmentation in histology images," *IEEE Access*, vol. 7, pp. 21420–21428, 2019.
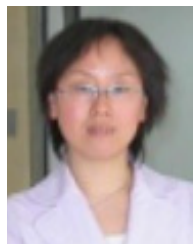
HONGFANG YUAN received the Ph.D. degree in mechatronics from the Beijing Institute of Technology, Beijing, China, in 2001. She is currently an Associate Professor with the School of Beijing University of Chemical Technology. Her research interests include intelligent diagnosis of equipment failure, data processing, and information fusion technology.

ZHENHONG LIU received the B.E. degree in communication engineering from the Beijing University of Chemical Technology, Beijing, China, in 2018, where she is currently pursuing the M.E. degree in control engineering. Her research interests include medical image processing and computer-aided diagnosis.

YAJUN SHAO received the B.E. degree in communication engineering from the Beijing University of Chemical Technology, Beijing, China, in 2019, where she is currently pursuing the M.E. degree in communication engineering. Her research interests include medical image identification and computer-aided diagnosis.

MIN LIU received the Ph.D. degree in imaging medicine and nuclear medicine from Xi'an Jiaotong University, Xi'an, China, in 2007.

She is currently the Deputy Director with the Department of Radiology, China-Japan Hospital. Her research interests include cardiothoracic imaging diagnosis, and pulmonary vascular disease imaging diagnosis and recognition. She is also a member of the Youth Group of the Chinese Medical Association Radiology Branch.

• • •