



Research article

CT medical image segmentation algorithm based on deep learning technology

Tongping Shen^{1,2,*}, Fangliang Huang¹ and Xusong Zhang²

¹ School of Information Engineering, Anhui University of Chinese Medicine, Hefei, 230012, China

² Graduate School, Angeles University Foundation, Angeles 2009, Philippines

* **Correspondence:** Email: shentp2010@ahtcm.edu.cn.

Abstract: For the problems of blurred edges, uneven background distribution, and many noise interferences in medical image segmentation, we proposed a medical image segmentation algorithm based on deep neural network technology, which adopts a similar U-Net backbone structure and includes two parts: encoding and decoding. Firstly, the images are passed through the encoder path with residual and convolutional structures for image feature information extraction. We added the attention mechanism module to the network jump connection to address the problems of redundant network channel dimensions and low spatial perception of complex lesions. Finally, the medical image segmentation results are obtained using the decoder path with residual and convolutional structures. To verify the validity of the model in this paper, we conducted the corresponding comparative experimental analysis, and the experimental results show that the DICE and IOU of the proposed model are 0.7826, 0.9683, 0.8904, 0.8069, and 0.9462, 0.9537 for DRIVE, ISIC2018 and COVID-19 CT datasets, respectively. The segmentation accuracy is effectively improved for medical images with complex shapes and adhesions between lesions and normal tissues.

Keywords: residual structure; attention module; medical images; segmentation; deep learning

1. Introduction

Medical images can reflect the anatomical structure or functional tissues of the human body, mainly through imaging techniques such as Computed Tomography (CT), Magnetic Resonance Imaging (MRI), and X-ray. However, due to the constraints of the internal environment of organ

tissues, the formed medical images are usually characterized by low contrast, blurred boundaries, and inaccurate image edge recognition. Medical image segmentation is one of the key steps to realize medical image visualization, and different image segmentation regions are often closely related to the related diseases and organs [1]. By segmenting medical images into other regions and sections, they are then provided to physicians for different diagnostic tasks such as lesion location determination, symptom identification, tissue and organ localization, description of anatomical structures, and treatment planning [2]. Therefore, medical image segmentation assists doctors in disease diagnosis, reduces the workload, improves the effectiveness and quality of treatment, and is widely used in disease diagnoses such as liver segmentation [3], cell segmentation [4], brain tumor segmentation [5], and COVID-19 [6].

Early medical image segmentation algorithms mainly used manually formulated rules for segmenting medical images. Zhang et al. [7] modified and optimized the objective function of the traditional fuzzy c-mean algorithm and achieved better performance. Ng et al. [8] combined the k-means algorithm with an improved watershed algorithm and applied it to medical image segmentation to greatly reduce the number of segmentation maps generated. Mohamed et al. [9] improved the fuzzy c-mean classification for the segmentation of brain images. Prabin et al. [10] combined the region-growing algorithm and contextual clustering technique and applied it to the lung CT image segmentation task.

This method of relying on manual methods for segmentation is costly and time-consuming, and the accuracy of segmentation markers cannot be guaranteed [11]. With the widespread use of deep learning techniques in the field of medical image segmentation, a series of research results have been achieved. Zhang et al. and Irfan et al. [12,13] proposed SNELM and HDNNs network structures for COVID-19 recognition. Long et al. [14] proposed the Fully Convolutional Network (FCN) to obtain a finer segmentation map by replacing the fully connected layer with a convolutional layer. Ronneberger et al. [15] proposed a classical medical image segmentation algorithm for U-Net networks based on FCN networks. Zhou et al. [16] proposed a U-Net++ network for medical image segmentation by adding new jump connections to transmit more image feature information. Oktay et al. [17] introduced the attention module into the jump connection in U-Net to improve the accuracy of pancreas segmentation. Peng et al. [18] proposed a local context-perception Net (LCP-Net) to obtain rich image feature information by parallel inflated convolution. Chen et al. [19] proposed a cross-scale residual network (CSR-Net) to achieve feature fusion of different layers by cross-scale residual connections. Wang et al. [20] proposed an adaptive, fully dense connected network (AFD-UNet); this network is based on U-Net++, which adaptively and effectively uses shallow and deep features. Feng et al. [21] proposed U-Net based residual network (URN) for image denoising to extract more detailed image features. Ge et al. [22] proposed a multi-input dilated U-Net (MD-UNet) for segmenting bladder cancer.

Various improved networks based on U-Net have improved the effectiveness of medical image segmentation. Still, medical images inherently suffer from problems such as category imbalance and noise factors that require the introduction of an attention mechanism to improve network segmentation performance. The attention mechanism is characterized by helping U-Net to better learn the interrelationships between multiple content modalities and thus better represent this information, overcoming its uninterruptable and thus difficult to design drawbacks.

Lan et al. [23] proposed mixed-attention based residual U-Net (MARU), which uses lightweight mixed attention blocks in the encoder to enhance image features effectively and suppress noise in the

encoding stage. Li et al. [24] proposed an attention-based nested U-Net (ANU-Net), which can suppress background regions irrelevant to the segmentation task. Guo et al. [25] proposed Spatial Attention U-Net (SA-UNet); this model adds a spatial attention module in the bottleneck layer, which can help the network focus on important features, suppress unnecessary features, and improve the network's representation capability.

The above algorithms achieved high classification accuracy, but there are still areas for improvement, such as not fully exploiting the structural advantages of the coding network and the attention mechanism needs to learn the semantic features of the disease fully. In this paper, we improve both the network structure and attention mechanism, propose a dual coding network structure based on multi-scale modules, and design a fully convolutional neural network model combining spatial and channel attention and residual modules. In the encoding and decoding feature extraction stages, the spatial channel attention (SCA) module is added to highlight the key feature information of medical images and suppress the interference of noise factors in medical images.

The innovation points of this paper are as follows:

1) Based on spatial and channel attention mechanisms analysis, an integrated attention mechanism module SCA is proposed, which can give different attention weights from space and channel dimensions so that the model can focus more on the image segmentation task. The model can be integrated into the mainstream neural network segmentation task.

2) Considering that the increase in network depth will lead to network degradation and other problems, this paper introduces the residual network structure and integrates it into the U-net backbone network with the SCA module. While maintaining the network depth, it pays attention to the segmentation task of the target area, enhances the transmission of feature information and gradient information between different levels of networks, learns more detailed feature information of medical images, and improves the segmentation accuracy.

3) The multi-model and multi-perspective comparative analysis of three different medical datasets is conducted to objectively analyze and evaluate the algorithms proposed in this paper.

2. Related works

This section is about medical image segmentation algorithms, including traditional medical image segmentation algorithms and medical image segmentation algorithms based on deep learning techniques.

2.1. Traditional methods

Traditional medical image segmentation methods mainly segment images based on physical features such as the shape, angle, and edge structure of medical images.

The region-based image segmentation method mainly uses the similarity and difference of features in the image region for medical image segmentation. The threshold method is simple in calculation and fast in segmentation, but when the gray values of medical images are similar, the segmentation effect could be better. The region-growth method introduces spatial information based on the threshold method, but it must manually select the initial point. The clustering method is sensitive to parameters and easy to falls into local optimization. The random walk method iterates by randomly selecting the initial points, and its randomness leads to extremely unstable segmentation

performance.

Bernal et al. [26] designed a model based on the appearance of polyps using the median depth of the valley accumulation window. Soltani-Nabipour et al. [27] introduced an improved algorithm that can automatically detect the threshold, update the threshold information, and accurately segment medical images in a shorter time. Chong et al. [28] used a clustering-based lung nodule algorithm to classify the lung nodule. Savic et al. [29] proposed a segmentation algorithm based on a fast marching method to segment the lung nodule image.

Bruntha et al. [30] used an image segmentation algorithm with edge-free active contours. By preprocessing, segmenting, and detecting lung nodules in the CT image, the segmentation accuracy reached 91.5%. Manickavasagam et al. [31] proposed a gradient-driven active contour algorithm, which uses normalization and gray co-occurrence matrix to extract nodule shape, and finally uses a support vector machine algorithm to detect and classify pulmonary nodules.

To sum up, the traditional segmentation method has a fast segmentation speed. Still, it is sensitive to parameters and cannot accurately segment lung regions that adhere to each other, so it needs rich prior information to obtain more accurate segmentation results. In addition, this method has great uncertainty for the segmentation results of different case images. Because of the complex and changeable structure of medical images, noise, partial volume effect, and other factors, image segmentation is generally combined with many traditional methods.

2.2. Deep learning

The traditional segmentation methods can no longer meet the clinical requirements for the current medical image segmentation tasks with a large amount of data and higher accuracy and time requirements. But the deep learning algorithm has been greatly improved in the accuracy of segmentation and the degree of automation of the algorithm. The deep learning method is divided into three categories according to its network structure.

2.2.1. CNN

Lecun et al. [32] designed a LeNet network structure for handwritten digit recognition. The convolution layer is responsible for extracting image feature information, and the full connection layer synthesizes and classifies the previously removed features. For medical image segmentation, CNN continuously extracts target features and realizes feature dimensionality reduction by combining the convolution layer and pooling layer, then integrates local features into global features through a full connection layer. Finally uses activation functions, such as softmax for classification and output, to complete the task of segmentation. Simantiris et al. [33] used a dilated Convolutional Neural Network method for brain MRI. Thyreau et al. [34] used a cortical parcellation method for MR brain images based on Convolutional Neural Networks. Aslan et al. and Akila Agnes et al. [35,36] proposed a new CNN segmentation framework for the semantic segmentation of lung disease.

The CNN algorithm has a simple network structure and fast operation, but the ability to extract image feature information is limited, which affects the image segmentation effect.

2.2.2. FCN

Long et al. [14] proposed the full connection layer with convolution layer FCN based on CNN, which solved the problem that CNN can only extract local features. FCN uses convolution and pooling to achieve down-sampling in the coding process which can extract high-level semantic information, de-convolution to achieve up-sampling in the decoding process, and predict each class's score at the pixel level. Du et al. [37] introduced a novel framework for retinal vessel segmentation based on deep ensemble learning. Wu et al. and Xia et al. [38,39] redesigned the convolution layer. By expanding the convolution layer and adding multi-scale feature information, they segmented medical images and achieved good segmentation results. Liu et al. [40] used a residual network structure to segment the image of pulmonary nodules by extracting local features and context information. Roth et al. [41] used a two-stage method that can focus on the segmentation of organ and blood vessel images.

The segmentation method based on FCN adopts the idea of down sampling and up sampling path designed instead of full connection and combines deep semantic information with shallow appearance information by designing jump connection.

2.2.3. U-Net

The FCN network structure is easy to lose image details, which affects segmentation results. Based on FCN, Ronneberger et al. [15] proposed a U-Net network including encoding and decoding structures. Each time the decoding is performed, the feature extraction part with the same number of channels is fused.

Lin et al. [42] proposed a novel deep medical image segmentation framework called dual swin transformer U-Net (DS-TransUNet). Milletari et al. [43] used a suitable network V-Net for 3D image segmentation. Hoorali et al. [44] proposed the IRUNet segmentation network, which makes full use of inception and residual blocks in skip connections and combines multi-scale features to extract better features for segmentation. Huang et al. [45] proposed the UNet 3+ network, which uses full-scale jump connections to obtain multi-level feature information. Alom et al. [46] proposed the R2UNet network, which combines residual connectivity and circular convolution to extract multidimensional image information.

Shen et al. [47] used the U-Net network as the basic framework, combined with the HarDNet module and attention module, for polyp image segmentation. Han et al. [48] designed the ConvUNeXt model to reduce the number of parameters while retaining the advantage of excellent segmentation. Gu et al. [49] proposed an attention-based integrated CNN (CA-Net) to achieve more accurate and interpretable medical image segmentation. Zhang [50] proposed the AResU-Net structure, which adds attention and residual network modules for brain tumor segmentation. Tong et al. [51] proposed an image segmentation network incorporating a triple attention mechanism to allow the segmentation network to focus more on the segmentation task.

Among these three kinds of deep learning methods, U-Net is improved and developed on FCN, while FCN is developed and improved on CNN. Each kind of image segmentation method can be improved by single network architecture and segmented together with other network architectures. No matter which deep learning segmentation method is still faced with many challenges and tests in clinical application, we need further research.

3. The proposed architecture

For the problems of blurred edges, uneven background distribution, and much noise interference in medical image segmentation, we propose a medical image segmentation algorithm based on deep neural network technology, which adopts a U-Net-like backbone structure and includes two parts: encoding and decoding, as shown in Figure 1. In the encoding process, the training image is input into the model. The image is passed through an encoder path with residual and convolutional structures to extract image feature information. The number of feature map channels will be doubled; down-sampling uses 2×2 Max-pooling convolutional layers for feature integration, and half will reduce the feature map size in aspect size after each down-sampling module. In the decoding process, the number of feature map channels will be reduced by half as each multi-branch residual block passes through the perceptual field of different sizes and adaptively captures the image feature information of different sizes; up-sampling uses Upsampling2D to double the size of the feature map. There is a lot of redundant information in the low-level image feature information extracted in the encoding stage during the jump connection. If it is directly mapped to the corresponding layer in the decoding stage, it will affect the segmentation effect. With the SCA module, the image features relevant to the segmentation task can be learned intensively during the jump connection process to improve the segmentation accuracy. The predicted image is the same size as the input image in the final output.

The block-level representation of our proposed technique is shown in Figure 2.

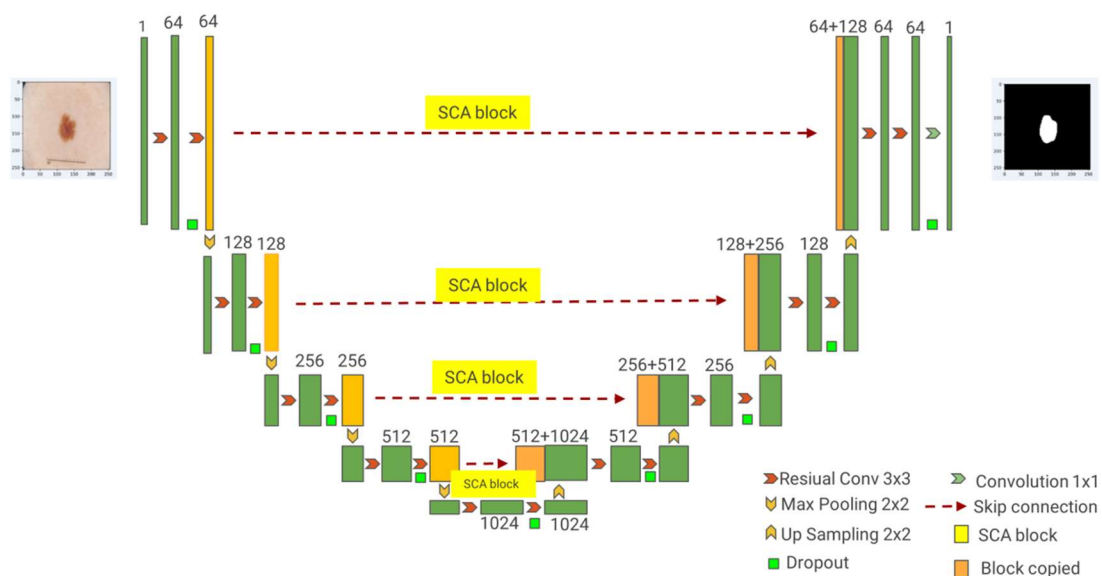


Figure 1. The proposed architecture.

3.1. Attention mechanism

The input information is selectively distinguished, located, and analyzed through the attention mechanism. The attention mechanism will also be applied in image segmentation, target tracking, and behavior detection in the neural network learning process. In deep learning, we can obtain the

image feature information of different spaces and latitudes by giving different weight information to the input image. How to build an attention mechanism model and integrate it into the mainstream neural network structure so that the simple neural network can achieve complex and high-precision image segmentation tasks is one of the problems that need to be solved.

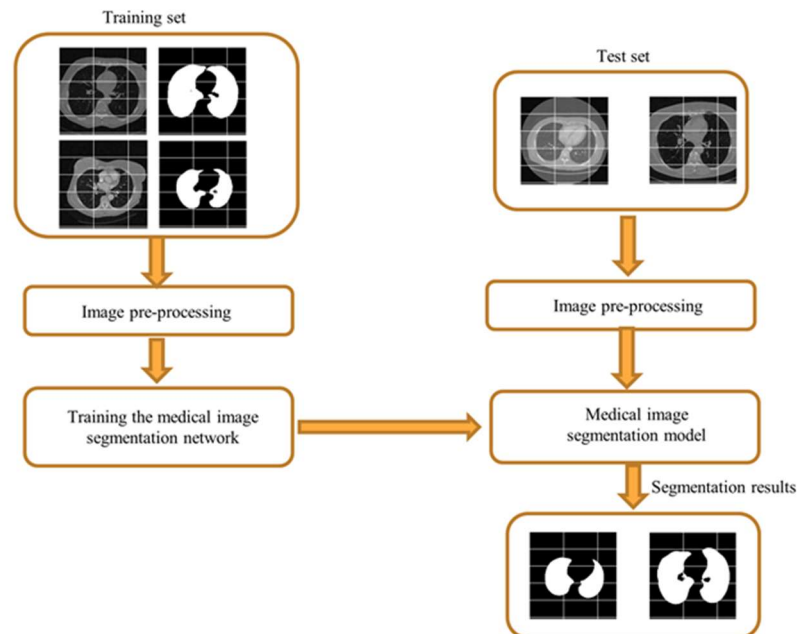


Figure 2. The block-level representation of our proposed technique.

3.1.1. Spatial attention

After analyzing the input image, the neural network assigns more weights to the regions closely related to the segmentation task, making the target segmentation region more prominent. At the same time, the image region feature information, which has nothing to do with the segmentation task, is suppressed. The input feature information of the image is multiplied by the spatial attention weight map to get the final output result.

$$x_m = \text{Maxpool}(x_{input}) \quad (1)$$

$$x_a = \text{Avgpool}(x_{input}) \quad (2)$$

$$x_{grap} = \text{Sigmoid}(w_1 * (\text{cat}[x_m, x_a]) + b_1) \quad (3)$$

$$x_{outputs} = x_{grap} * x_{input} \quad (4)$$

3.1.2. Channel attention

The feature information on each channel is different, and the importance to the global feature information of the whole image is also different. By analyzing the segmented image, we can assign weight information to each channel, it indicating the importance of the channel information to the

global feature description. The input feature information of the image is multiplied by the channel attention weight map to get the final output result.

$$x_m = \text{Maxpool}(x_{input}) \quad (5)$$

$$x_a = \text{Avgpool}(x_{input}) \quad (6)$$

$$x_1 = w_{fc3} * (w_{fc2} * (w_{fc1} * x_m + b_{fc1}) + b_{fc2}) + b_{fc3} \quad (7)$$

$$x_2 = w_{fc3} * (w_{fc2} * (w_{fc1} * x_a + b_{fc1}) + b_{fc2}) + b_{fc} \quad (8)$$

$$x_{graphc} = \text{Sigmoid}(w_f * (\text{cat}[x_1, x_2]) + b_f) \quad (9)$$

$$x_{outputc} = x_{graphc} * x_{input} \quad (10)$$

3.1.3. Spatial channel attention

In the paper, channel and spatial attention mechanisms are combined and given different weights. Finally, the output image feature information processed by the SCA module is as follows:

$$x_{output} = \text{cat}[x_{outputs}, x_{outputc}] \quad (11)$$

$x_{outputs}$ represent the image feature information processed by the spatial attention mechanism. $x_{outputc}$ represent the image feature information through the channel attention mechanism.

Our proposed SCA module, as shown in Figure 3, with a general design idea similar to the architecture proposed by Fu et al. [52] integrates spatial and channel attention integration modules into an improved U-net network structure. The SCA module combines spatial and channel attention mechanisms to get comprehensive attention mechanism information. This module enhances the significant features of the up-sampling process by applying attention weights to high-dimensional and low-dimensional features.

The SCA module proposed in this paper can address the feature information of medical images, highlight more of the key feature information of medical images, and suppress the interference of noise factors in medical images.

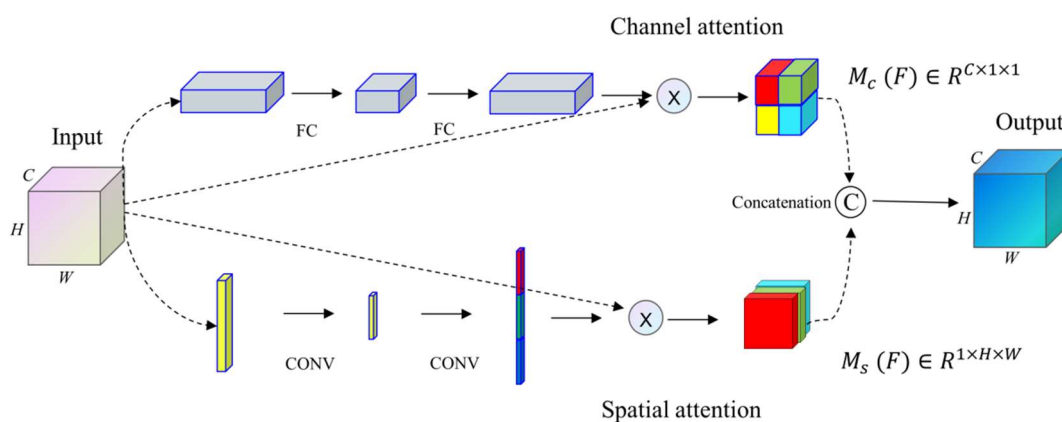


Figure 3. The SCA architecture.

3.2. Residual convolution module

He [53] first proposed the residual network, which effectively solves the contradiction between neural network depth and recognition accuracy, as shown in Figure 4.

When a neural network reaches a certain depth, the output x of that layer is already optimal, and further deepening the network will result in degradation. In a convolutional neural network, it is difficult to ensure the weight of the following layer network. In the residual structure, the $F(x)$ only needs to update a small part of the weight of $F(x)$. It is more sensitive to output changes, and the parameters are adjusted more widely, which can speed up the learning speed and improve the performance of the model.

$$y = F(x, \{W_i\}) + W_s x \quad (12)$$

where W_s is mainly a 1×1 convolution used to match the channel dimensions of the residual structure model input x and model output y . $F(x, \{W_i\})$ is the residual mapping that the network needs to learn. When the residual structure has the same input and output dimensions, the definition is as follows:

$$y = F(x, \{W_i\}) + x \quad (13)$$

The input information x is added to the feature calculation process, combined with the feature information of the upper layer to enrich the feature extraction of the network layer.

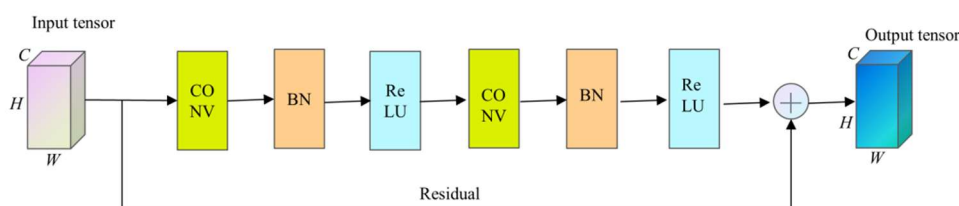


Figure 4. The residual architecture.

Through the residual structure design, the degradation problem in the process of deep structure network training can be well solved without adding additional parameters and calculations, improving model run speed and segmentation performance.

3.3. Loss function

The Loss function is the index of a neural network to find the optimal weight parameters [54]. There are many kinds of loss functions, including single loss function and mixed loss function. We combine the Dixeloss [55] and Focal loss [56] as the loss function in this paper. This function can combine the two functions' advantages and fully use semantic information to make the network better finds the optimal parameters for optimization learning.

$\mathcal{L}_{\text{Dixeloss}}$ is the loss function of Dice loss, it is mainly used to measure the degree of loss of similarity between the segmented image predicted by the model and the real segmented image, and the value range is $[0,1]$. The calculation formula of the function is shown in Eq (14). $|X \cap Y|$

represent the number of intersections between an actual segmented image and a model predicted image, $|X|$ and $|Y|$ represent the number of real segmented images and model predicted images respectively.

$$\mathcal{L}_{\text{DiceLoss}} = 1 - \frac{2|X \cap Y|}{|X| + |Y|} \quad (14)$$

$\mathcal{L}_{\text{FocalLoss}}$ is a loss function to deal with the unbalanced classification of samples. According to the difficulty of sample resolution, different weight coefficients α are added to the samples to reduce the adverse effects on training loss caused by the imbalance of sample classification.

$$\mathcal{L}_{\text{FocalLoss}} = -\alpha \times (1 - p)^{\gamma} \times \log(p) \quad (15)$$

$p \in [0,1]$ is the probability of the model predicting the positive sample. $\mathcal{L}_{\text{FocalLoss}}$ is a modification of the cross entropy loss function. The total loss function proposed in this paper is $\mathcal{L}_{\text{Total}}$, The formula is shown in Eq (16).

$$\mathcal{L}_{\text{Total}} = \mathcal{L}_{\text{DiceLoss}} + \mathcal{L}_{\text{FocalLoss}} \quad (16)$$

4. Experimental results

The main content of this section is to compare and verify the models based on the analysis of three medical image data sets. First of all, the three data sets are described and pre-processed. Then it describes the operating parameters of the model and related evaluation indicators. Finally, the segmentation results of different network structures on the three data sets are compared, analyzed, and displayed visually, including different model segmentation results, model ACC diagrams, Loss diagrams, and so on.

4.1. Datasets

The DRIVE data set is mainly used for the study of vascular segmentation in retinal images. There are 40 retinal vascular images, including 33 fundus images of healthy people and seven fundus images of diabetic retinal lesions. The neural network needs more data to learn, so we carry on the data expansion operation to the picture of the training set. First of all, the input image is segmented into different parts. We generate a total of 100000 parts, of which 90000 are used as training sets and 10000 as validation sets. According to the requirements of the neural network model, the size of each part of the picture is set to 32×32 .

The ISIC2018 data set is from the Kaggle [57]. The dataset has 2594 images, the original image size is 700×900 , and the image size is resized to 256×256 as needed.

The COVID-19 CT contains a series of CT images of lung image segmentation and the corresponding label data, released by Kaggle in 2017. The dataset has 301 images; the original image size is 512×512 , and the image size is resized to 256×256 as needed.

4.2. Experimental environment and parameter metrics

4.2.1. Implementation details

All models in this paper were run on the open-source TensorFlow 2.4.1 platform and used the NVIDIA GeForce Gtx1080 for experiments. We all used picture rotation and flipping operations for the samples participating in the training to increase the data. First, rotate each image 60 degrees apart once, followed by a horizontal pan of 30-pixel points, and finally, a random crop once. During the training, we set the picture of the DRIVE data set to 32×32 and the ISIC2018 data set and COVID-19 CT data set pictures to 256×256 . The model learning rate is 0.0001, the dropout ratio is 0.5, and use the Adam optimization algorithm. We set the DRIVE data set and ISIC2018 data set the batch size is 8 and the COVID-19 CT data set batch size to 32. 100 iterations were carried out in each experiment.

To further avoid the problem of over-fitting in the model training process and reasonably evaluate the model's segmentation performance, we use 5-fold cross-validation to optimize the whole network.

4.2.2. Metrics

We use several evaluation metrics, including Area under the ROC curve (AUC), Sensitivity (SENS), Precision (PRC), Jaccard similarity score (JS), Specificity (SPE), DICE, and IOU.

The JS coefficient measures the similarity between the two sets of predicted and valid segmented pixels.

$$JS = \frac{TP}{TP+FP+FN} \quad (17)$$

The DICE coefficient measures the similarity between the predicted segmented labeled graph and the true segmented labeled graph.

$$Dice = \frac{2TP}{2TP+FP+FN} \quad (18)$$

The IOU can reflect the overlap rate between the predicted segmented label map and the real segmented label map.

$$Iou = \frac{TP}{TP+FP+FN} \quad (19)$$

The sensitivity calculates the proportion of correctly segmented target pixels to the target class pixels in the true segmented label map.

$$Sens = \frac{TP}{TP+FN} \quad (20)$$

where TP is the pixels correctly segmented in the medical segmentation results, TN is the pixels incorrectly segmented in the medical segmentation results, FP is the background pixels incorrectly treated as medical pixels in the medical segmentation results, and FN is the medical pixels

incorrectly treated as background pixels in the segmentation results.

4.3. Analysis of experimental results

To verify the effectiveness and superiority of the proposed Residual-Attention-Unet network, the segmentation experiments are carried out on DRIVE, ISIC2018, and COVID-19 CT datasets and compared with Attention-Unet, Dense-Unet, U-Net, Unet++, and Residual-Unet. The model test results are shown in Tables 1–3.

Table 1. Segmentation results of different network structures on DRIVE.

Model	AUC	PRC	JS	SENS	SPE	DICE	IOU
Unet [15]	0.9559	0.8707	0.6628	0.7348	0.9728	0.7314	0.9452
Dense-Unet [58]	0.9725	0.8910	0.6764	0.7554	0.9867	0.7529	0.9523
Unet++ [16]	0.9699	0.8839	0.6755	0.7627	0.9848	0.7359	0.9483
Attention-Unet [59]	0.9759	0.8999	0.6953	0.7800	0.9858	0.7628	0.9504
Residual-Unet [59]	0.9751	0.8996	0.6799	0.7499	0.9888	0.7451	0.9593
Ours	0.9762	0.9034	0.7070	0.8070	0.9905	0.7826	0.9683

Table 2. Segmentation results of different network structures on ISIC2018.

Model	AUC	PRC	JS	SENS	SPE	DICE	IOU
Unet [15]	0.8557	0.8235	0.6648	0.7262	0.9452	0.8755	0.7786
Dense-Unet [58]	0.8733	0.8765	0.7244	0.7623	0.9843	0.8835	0.7937
Unet++ [16]	0.8770	0.8807	0.7324	0.7690	0.9850	0.8783	0.7831
Attention-Unet [59]	0.8741	0.8750	0.7239	0.7652	0.9829	0.8791	0.7851
Residual-Unet [59]	0.8604	0.8689	0.7026	0.7343	0.9801	0.8852	0.7743
Ours	0.8995	0.9419	0.7680	0.8974	0.9865	0.8904	0.8069

Table 3. Segmentation results of different network structures on COVID-19 CT.

Model	AUC	PRC	JS	SENS	SPE	DICE	IOU
Unet [15]	0.9678	0.9237	0.8478	0.9669	0.9487	0.7982	0.8253
Dense-Unet [58]	0.9916	0.9865	0.9709	0.9890	0.9942	0.9257	0.9418
Unet++ [16]	0.9913	0.9884	0.9740	0.9866	0.9960	0.8765	0.9023
Attention-Unet [59]	0.9905	0.9891	0.9747	0.9900	0.9970	0.8963	0.9116
Residual-Unet [59]	0.9859	0.9849	0.9645	0.9754	0.9965	0.8769	0.9008
Ours	0.9935	0.9915	0.9809	0.9904	0.9971	0.9462	0.9537

As can be seen from Table 1, for the DRIVE data set, the AUC, PRC, JS, SENS, SPE, DICE, and IOU of the U-Net network are 0.9559, 0.8707, 0.6628, 0.7348, 0.9728, 0.7314, and 0.9452, respectively. Compared with U-Net, the Residual-Attention-Unet proposed in this paper increased by 2.03, 3.27, 4.42, 7.22, 1.77, 5.12 and 2.31% on AUC, PRC, JS, SENS, DICE and IOU, respectively. This paper proposes that compared with these networks, the indicators of AUC, PRC, JS, SENS, SPE,

DICE, and IOU have been improved to a certain extent, up to 0.6, 1.95, 3.15, 5.71, 0.57, 4.67 and 2.0%, respectively.

As can be seen from Table 2, for the ISIC2018 data set, the AUC, PRC, JS, SENS, SPE, DICE, and IOU of the U-Net network are 0.8557, 0.8235, 0.6648, 0.7262, 0.9452, 0.8755, and 0.7786, respectively. Compared with U-Net, the Residual-Attention-Unet proposed in this paper increased by 4.38, 11.84, 10.32, 17.12, 4.13, 1.49 and 2.83% on AUC, PRC, JS, SENS, DICE and IOU, respectively. This paper proposes that compared with these networks, the indicators of AUC, PRC, JS, SENS, SPE, DICE, and IOU have been improved to a certain extent, up to 3.91, 7.3, 6.54, 16.31, 0.64, 1.21 and 3.26%, respectively.

As can be seen from Table 3, for the COVID-19 CT data set, the AUC, PRC, JS, SENS, SPE, DICE, and IOU of the U-Net network are 0.9678, 0.9237, 0.8478, 0.9669, 0.9487, 0.7982, and 0.8253, respectively. Compared with U-Net, the Residual-Attention-Unet proposed in this paper increased by 2.57, 6.78, 13.31, 2.35, 4.84, 14.8 and 12.84% on AUC, PRC, JS, SENS, DICE and IOU, respectively. This paper proposes that compared with these networks, the indicators of AUC, PRC, JS, SENS, SPE, DICE, and IOU have been improved to a certain extent, up to 0.76, 0.66, 1.64, 1.5, 0.29, 6.97 and 5.29%, respectively.

Table 4. Comparisons against existing approaches on DRIVE, ISIC2018 and COVID-19 CT.

Datasets	Methods	DICE	IOU	PRC	SENS
DRIVE	DFUNet [60]	0.7962	0.9605	0.9024	0.7863
	IterNet [61]	0.7891	0.9692	0.8973	0.7735
	RV-GAN [62]	-	0.9762	-	0.7927
	Ours	0.7826	0.9683	0.9034	0.8070
ISIC2018	TransFuse [63]	0.8927	0.8063	0.9466	0.9128
	SANet [64]	0.8859	0.7952	0.9439	0.8760
	UNeXt-S [65]	0.8833	0.7909	0.9348	0.8715
	Ours	0.8904	0.8069	0.9419	0.8974
COVID-19 CT	BCDU-Net [66]	0.9794	0.9477	0.9753	0.9979
	R2U-Net [46]	0.9431	0.9746	0.9729	0.9832
	MDA-Net [67]	0.9855	0.9536	0.9864	-
	Ours	0.9462	0.9537	0.9915	0.9904

We also compare the proposed method in this paper with several recently proposed segmentation methods for analysis. Table 4 shows the performance of different segmentation methods on the DRIVE, ISIC2018 and COVID-19 CT datasets. The method proposed in this paper achieves good results on all three datasets. Compared with recent research methods, the method in this paper achieves optimal performance on some evaluation metrics on different datasets, and some metrics are close to the optimal segmentation performance.

To verify the effectiveness and superiority of the proposed Residual-Attention-Unet network, the segmentation experiments are carried out on DRIVE, ISIC2018 and COVID-19 CT data sets and compared with Unet, Unet++, Dense-Unet, Residual-Unet, Attention-Unet, and Residual-Attention-Unet. To ensure the fairness of the experimental results, this paper runs the six comparison networks in the same experimental environment, and their visual effects are shown in Figures 5–7.

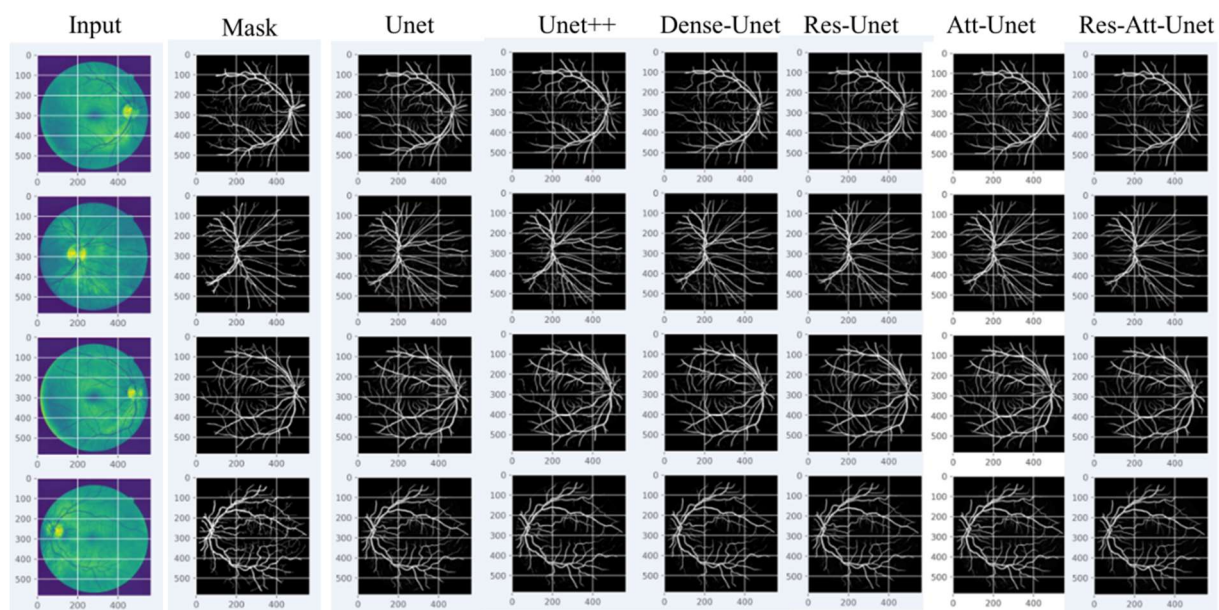


Figure 5. Model segmentation results in the DRIVE dataset.

Figure 5 shows the segmentation effect of various networks on the DRIVE data set. The first and second columns are the original picture and the segmentation result diagram, respectively, and the third and eighth columns are the resulting diagram of six network segmentation, respectively. As seen in Figure 5, several networks can segment the details of the main part of blood vessels. Still, the Residual-Attention-Unet network proposed in this paper can segment the most detailed information, and the segmentation effect is the best.

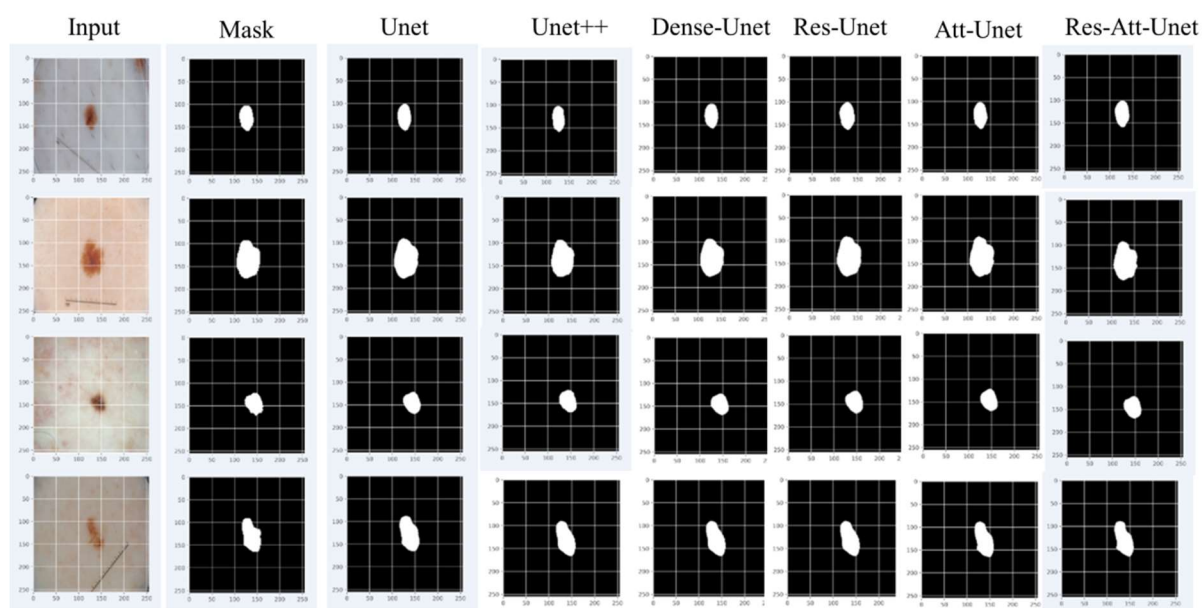


Figure 6. Model segmentation results in the ISIC2018 dataset.

Figure 6 shows the segmentation effect of various networks on the ISIC2018 data set. The third and eighth columns are the resulting diagram of six network segmentations. As seen in Figure 6, several networks can segment the edge information of skin cancer images. Still, the Residual-Attention-Unet network proposed in this paper is better than others in dealing with the edge part. The segmentation boundary is more precise, the structure is relatively complete, and it achieves the best segmentation performance.

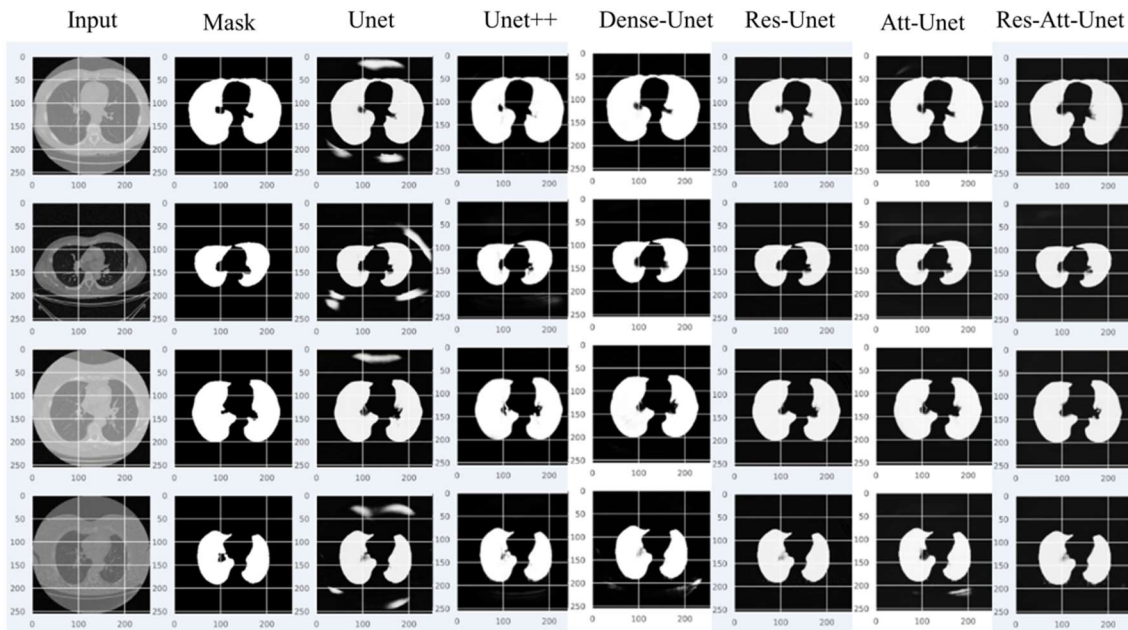


Figure 7. Model segmentation results in the COVID-19CT dataset.



Figure 8. Accuracy and loss on the DRIVE dataset.

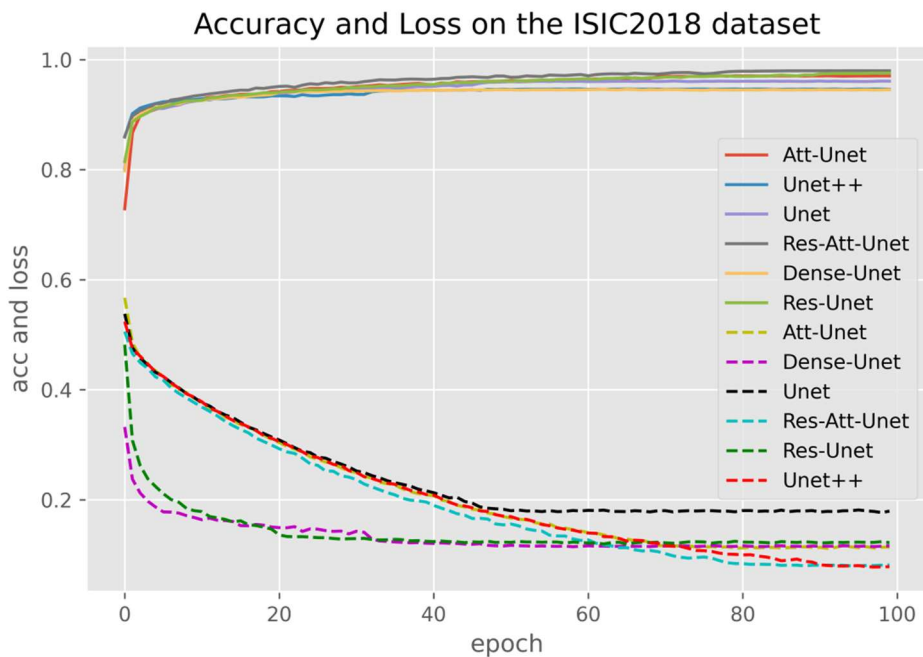


Figure 9. Accuracy and loss on the ISIC2018 dataset.

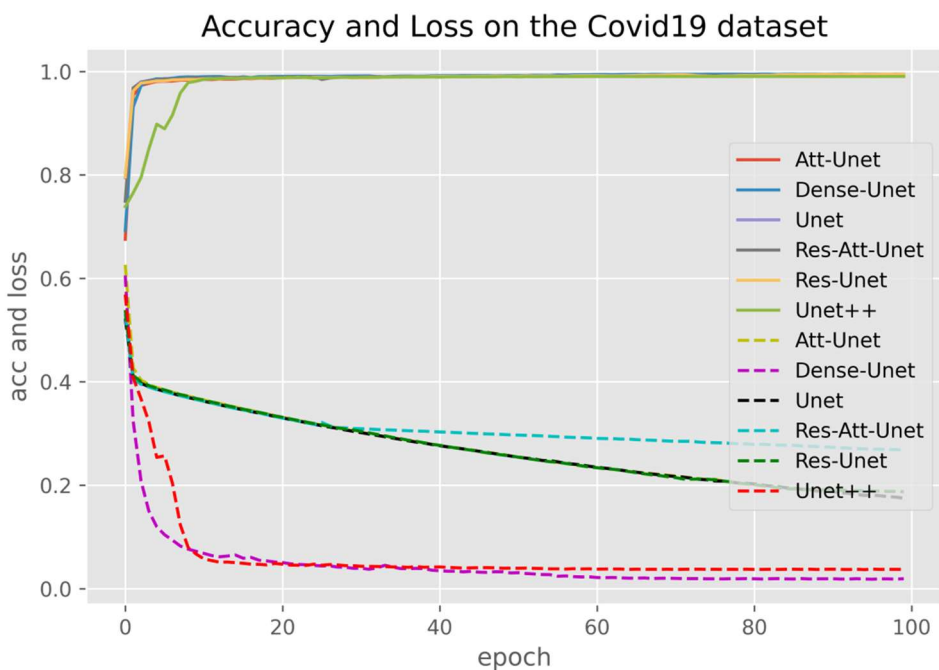


Figure 10. Accuracy and loss on the COVID-19 CT dataset.

Figure 7 shows the segmentation effect of various networks on the COVID-19 data set. The third and eighth columns are the resulting diagram of six network segmentations. From Figure 7, the U-Net has learned too many redundant features. There are always obvious noise points; several other networks also have good segmentation performance on the segmentation boundary, but it pays too

much attention to the image boundary, thus ignoring the internal features of the image. However, the Residual-Attention-Unet network proposed in this paper retains more image details, and the segmentation results are basically consistent with the standard segmented images.

We also compare the accuracy and loss of the different models on the three datasets, as shown in Figures 8–10.

The comparative analysis of the three figures shows that the model proposed in our paper converges fast on the three datasets. Finally, the model is almost converged and achieved a high accuracy rate.

From the loss diagrams of the above three data sets, we can find that after 100 rounds of experimental iterations, the network reaches the convergence state. The convergence rate of each model is also different on different data sets. Dense-Unet and Unet++ converge for DRIVE data sets after 40 rounds of experimental iterations. For ISIC2018 data sets, Dense-Unet and Res-Unet converge after 40 experimental iterations.

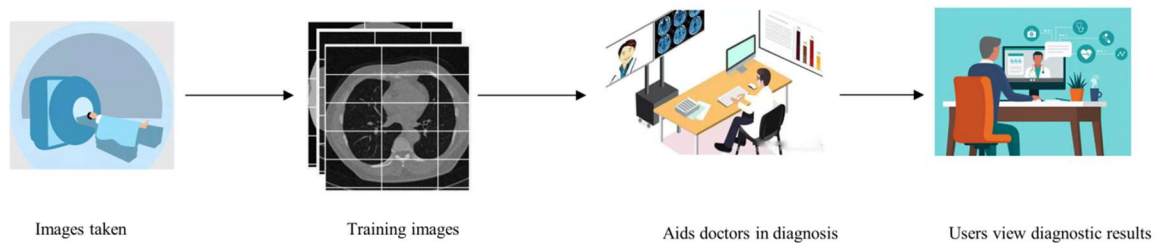


Figure 11. Deployed proposed deep-learning model in clinical application.

5. Discussion

The U-Net network structure has achieved excellent performance in the field of medical image processing, but the U-Net network itself has problems such as incomplete feature extraction and lack of multi-scale feature information processing capability, therefore, researchers continue to propose improved segmentation networks based on the U-Net structure, such as Dense-Unet, Unet++, Attention-Unet, Residual-Unet, etc. The Dense-Unet segmentation network, by expanding the convolution with different expansion rates, expands the perceptual field without increasing the computational cost and effectively prevents the loss of spatial information, Dense-Unet segmentation network, by convolving with different expansion rates of dilation, the expanded perceptual field does not increase the computational cost and effectively prevents the loss of spatial information. attention-Unet segmentation network, by eliminating the influence of noise and invalid information in the image through the attention mechanism, reconstructs the contextual features in the image.

These network structures further improve the image segmentation accuracy by extracting multi-scale feature information and deepening the segmentation network depth, etc. Based on these excellent segmentation networks, we further optimize them in two aspects: for the medical image feature information extracted by the small number of layers of the U-Net network is not sufficient, the original convolutional layers are replaced by the residual network structure, and more levels of

image feature information are extracted by deepening the network depth; the gradient disappearance problem caused by the deepening of the model is avoided while improving the network performance. On the other hand, in the encoding and decoding feature extraction, the network can be used to extract the image information. On the other hand, we add the SCA module in the encoding and decoding feature extraction stages. To make the network achieve a more accurate capture of the main part of medical images, a dual-channel attention mechanism is added to the expansion part of the model to highlight the key feature information of medical images and suppress the interference of noise factors in medical images, which can effectively improve the segmentation accuracy of the whole network.

Tables 1–3 show the evaluation results of the network models of Unet, Unet++, Dense-Unet, Residual-Unet, Attention-Unet and Residual-Attention-Unet on the DRIVE and ISIC2018 and COVID-19CT datasets, respectively. Figures 5–7 show the experimental results of segmentation of network models such as Unet, Unet++, Dense-Unet, Residual-Unet, Attention-Unet and Residual-Attention-Unet on DRIVE and ISIC2018 and COVID-19CT datasets, respectively. Through the analysis of the above graphical results, the proposed networks in this paper have high accuracy and acceptable segmentation results. Meanwhile, the accuracy and loss plots of different models on DRIVE, ISIC2018 and COVID-19CT datasets are compared and analyzed as in Figures 8–10, the proposed model achieves good results in terms of accuracy and convergence speed of loss.

Although we evaluated the performance of the network extensively on three different datasets, our network still needs some improvements. First, due to objective factors, we did not try to validate the effect of the connection method of different attention modules on the network structure; second, our network did not try to validate the 3D medical image segmentation dataset.

Meanwhile, the Transformer structure, which is popular in natural language processing tasks, has been widely used in medical image segmentation in recent years, and we will further investigate the fusion of Transformer and U-Net and other network structures.

6. Conclusions

Based on the analysis of the U-Net segmentation network and related improved medical image segmentation networks, we propose an optimized medical segmentation network, which is mainly manifested in two aspects: for the U-Net network with few layers, the medical image feature information extracted is not sufficient, the original convolutional layers are replaced by the residual network structure, and more layers of image feature information are extracted by deepening the network. The network performance is improved while avoiding the problem of gradient disappearance caused by the deepening of the model. On the other hand, we add the SCA module in the encoding and decoding feature extraction stages. To make the network achieve a more accurate capture of the main part of the medical image, a dual-channel attention mechanism is added to the expansion part of the model to highlight the key feature information of the medical image and suppress the interference of noise factors in the medical image, which can effectively improve the segmentation accuracy of the whole network. The algorithm is compared and analyzed on several medical image datasets. The proposed network structure has improved the evaluation indexes of DICE, Precision, and IOU compared with other improved U-Net segmentation networks, and better segmentation results are obtained. We will further expand the datasets and extend the method to 3D medical image segmentation and accurate segmentation of other diseases in the future.

Acknowledgments

This research was funded by Excellent Young Talents in Anhui Universities Project (Granted No. gxyq2022026, No. gxyq2020016), Anhui Province Quality Engineering Project (Granted No. 2021jyxm0801, No.2021jxjy035), and Science Foundation of universities in Anhui Province (Granted No. KJ2020A0394, No. 2022AH050428).

Conflict of interest

The authors declare there is no conflict of interest.

References

1. J. Chen, Y. Lu, Q. Yu, X. Luo, E. Adeli, Y. Wang, et al., TransUNet: Transformers make strong encoders for medical image segmentation, preprint, arXiv:2102.04306.
2. T. P. Shen, H. Q. Xu, Medical image segmentation based on Transformer and HarDNet structures, *IEEE Access*, **11** (2023), 16621–16630. <https://doi.org/10.1109/ACCESS.2023.3244197>
3. L. Han, Y. H. Chen, J. M. Li, B. W. Zhong, Y. Z. Lei, M. H. Sun, Liver segmentation with 2.5 D perpendicular UNets, *Comput. Electr. Eng.*, **91** (2021), 107118. <https://doi.org/10.1016/j.compeleceng.2021.107118>
4. H. Y. Li, X. Q. Zhao, A. Y. Su, H. T. Zhang, J. X. Liu, G. Y. Gu, Color space transformation and multi-class weighted loss for adhesive white blood cell segmentation, *IEEE Access*, **8** (2020), 24808–24818. <https://doi.org/10.1109/ACCESS.2020.2970485>
5. T. Magadza, S. Viriri, Deep learning for brain tumor segmentation: a survey of state-of-the-art, *J. Imaging*, **7** (2021), 19. <https://doi.org/10.3390/jimaging7020019>
6. Y. E. Almalki, A. Qayyum, M. Irfan, N. Haider, A. Glowacz, F. M. Alshehri, et al., A novel method for COVID-19 diagnosis using artificial intelligence in chest X-ray images, *Healthcare*, **9** (2021), 522. <https://doi.org/10.3390/healthcare9050522>
7. D. Q. Zhang, S. C. Chen, A novel kernelized fuzzy c-means algorithm with application in medical image segmentation, *Artif. Intell. Med.*, **32** (2014), 37–50. <https://doi.org/10.1016/j.artmed.2004.01.012>
8. H. P. Ng, S. H. Ong, K. W. C. Foong, Poh-Sun Goh, W. L. Nowinski, Medical image segmentation using k-means clustering and improved watershed algorithm, in *2006 IEEE southwest symposium on image analysis and interpretation*, (2006). <https://doi.org/10.1109/SSIAI.2006.1633722>
9. N. A. Mohamed, M. N. Ahmed, A. Farag, Modified fuzzy c-mean in medical image segmentation, in *1999 IEEE International Conference on Acoustics*, (1999). <https://doi.org/10.1109/ICASSP.1999.757579>
10. A. Prabin A, J. Veerappan, Automatic segmentation of lung ct images by CC based region growing, *J. Theor. Appl. Inf. Technol.*, **68** (2014), 63–69.

11. M. Negassi, R. Suarez-Ibarrola, S. Hein, A. Miernik, A. Reiterer, Application of artificial neural networks for automated analysis of cystoscopic images: a review of the current status and future prospects, *World J. Urol.*, **38** (2020), 2349–2358. <https://doi.org/10.1007/s00345-019-03059-0>
12. Y. Zhang, M. A. Khan, Z. Zhu, S. Wang, SNELM: SqueezeNet-guided ELM for COVID-19 recognition, *Comput. Syst. Sci. Eng.*, **46** (2023), 13–26. <https://doi.org/10.32604/csse.2023.034172>
13. M. Irfan, M. A. Iftikhar, S. Yasin, U. Draz, T. Ali, S. Hussain, et al., Role of hybrid deep neural networks (HDNNs), computed tomography, and chest X-rays for the detection of COVID-19, *Int. J. Environ. Res. Public Health*, **18** (2021), 3056. <https://doi.org/10.3390/ijerph18063056>
14. J. Long, E. Shelhamer, T. Darrell, Fully convolutional networks for semantic segmentation, in *Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition*, (2015). <https://doi.org/10.1109/CVPR.2015.7298965>
15. O. Ronneberger, P. Fischer, T. Brox, U-net: Convolutional networks for biomedical image segmentation, in *International Conference on Medical image computing and computer-assisted intervention*, (2015). https://doi.org/10.1007/978-3-319-24574-4_28
16. Z. W. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, J. M. Liang, Unet++: redesigning skip connections to exploit multiscale features in image segmentation, *IEEE Trans. Med. Imaging*, **39** (2020), 1856–1867. <https://doi.org/10.1109/TMI.2019.2959609>
17. O. Oktay, J. Schlemper, L. L. Folgoc, M. Lee, M. Heinrich, K. Misawa, et al., Attention U-Net: learning where to look for the Pancreas, preprint, arXiv:1804.03999.
18. D. L. Peng, S. Y. Xiong, W. J. Peng, J. P. Lu, LCP-net: a local context-perception deep neural network for medical image segmentation, *Expert Syst. Appl.*, **168** (2021), 114234. <https://doi.org/10.1016/j.eswa.2020.114234>
19. C. Chen, B. Liu, K. N. Zhou, W. Z. He, F. Yan, Z. L. Wang, R. X. Xiao, CSR-net: cross-scale residual network for multi- objective scaphoid fracture segmentation, *Comput. Biol. Med.*, **137** (2021), 104776. <https://doi.org/10.1016/j.compbiomed.2021.104776>
20. E. K. Wang, C. M. Chen, M. M. Hassan, A. Almogren, A deep learning based medical image segmentation technique in Internet-of-Medical-Things domain, *Future Gene. Comput. Sy.*, **108** (2020), 135–144. <https://doi.org/10.1016/j.future.2020.02.054>
21. T. Feng, C. S. Wang, X. W. Chen, H. T. Fan, K. Zeng, Z. Y. Li, URNet: A UNet based residual network for image dehazing, *Appl. Soft Comput.*, **102** (2020), 106884. <https://doi.org/10.1016/j.asoc.2020.106884>
22. R. Q. Ge, H. H. Cai, X. Yuan, F. W. Qin, Y. Huang, et al., MD-UNET: Multiinput dilated U-shape neural network for segmentation of bladder cancer, *Comput. Biol. Chem.*, **93** (2021), 107510. <https://doi.org/10.1016/j.compbiolchem.2021.107510>
23. Y. C. Lan, X. M. Zhang, Real-time ultrasound image despeckling using mixed-attention mechanism based residual UNet, *IEEE Access*, **8** (2020), 195327–195340. <https://doi.org/10.1109/ACCESS.2020.3034230>
24. C. Li, Y. S. Tan, W. Chen, X. Luo, Y. L. He, Y. M. Gao, F. Li, ANU-Net: Attention-based nested U-Net to exploit full resolution features for medical image segmentation, *Comput. Graph.*, **90** (2020), 11–20. <https://doi.org/10.1016/j.cag.2020.05.003>
25. C. L. Guo, M. Szemenyei, Y. G. Yi, W. L. Wang, B. Chen, C. Q. Fan, SA-UNet: Spatial attention U-Net for retinal vessel segmentation, in *25th International Conference on Pattern Recognition (ICPR)*, (2021). <https://doi.org/10.1109/ICPR48806.2021.9413346>

26. J. Bernal, F. J. Sánchez, G. Fernández-Esparrach, D. Gil, C. Rodríguez, F. Vilariño, WM-DOVA maps for accurate polyp highlighting in colonoscopy: validation vs. saliency maps from physicians, *Comput. Med. Imaging Graphics*, **43** (2015), 99–111. <https://doi.org/10.1016/j.compmedimag.2015.02.007>
27. J. Soltani-Nabipour, A. Khorshidi, B. Noorian, Lung tumor segmentation using improved region growing algorithm, *Nuclear Eng. Technol.*, **52** (2020), 2313–2319. <https://doi.org/10.1016/j.net.2020.03.011>
28. S. Y. Chong, M. K. Tan, K. B. Yeo, M. Y. Ibrahim, X. Hao, K. T. K. Teo, Segmenting nodules of lung tomography image with level set algorithm and neural network, in *2019 IEEE 7th Conference on Systems, Process and Control (ICSPC)*, (2019). <https://doi.org/10.1109/ICSPC47137.2019.9067987>
29. M. Savic, Y. Ma, G. Ramponi, W. Du, Y. Peng, Lung nodule segmentation with a region-based fast marching method, *Sensors*, **21** (2021), 1908. <https://doi.org/10.3390/s21051908>
30. P. M. Bruntha, S. I. A. Pandian, P. Mohan, Active Contour Model (without edges) based pulmonary nodule detection in low dose CT images, in *2019 2nd International Conference on Signal Processing and Communication (ICSPC)*, (2019). <https://doi.org/10.1109/ICSPC46172.2019.8976813>
31. R. Manickavasagam, S. Selvan, GACM based segmentation method for Lung nodule detection and classification of stages using CT images, in *2019 1st International Conference on Innovations in Information and Communication Technology (ICIICT)*, (2019). <https://doi.org/10.1109/ICIICT1.2019.8741477>
32. Y. LeCun, L. Bottou, Y. Bengio, P. Haffner, Gradient-based learning applied to document recognition, *Proc. IEEE*, **86** (1998), 2278–2324. <https://doi.org/10.1109/5.726791>
33. G. Simantiris, G. Tziritas, Cardiac MRI segmentation with a dilated CNN incorporating domain-specific constraints, *IEEE J. Selected Topics Signal Process.*, **14** (2020), 1235–1243. <https://doi.org/10.1109/JSTSP.2020.3013351>
34. B. Thyreau, Y. Taki, Learning a cortical parcellation of the brain robust to the MRI segmentation with convolutional neural networks, *Med. Image Anal.*, **14** (2020), 101639. <https://doi.org/10.1016/j.media.2020.101639>
35. M. F. Aslan, A robust semantic lung segmentation study for CNN-based COVID-19 diagnosis, *Chemom. Intell. Lab. Syst.*, **231** (2022), 104695. <https://doi.org/10.1016/j.chemolab.2022.104695>
36. S. Akila Agnes, J. Anitha, J. D. Peter, Automatic lung segmentation in low-dose chest CT scans using convolutional deep and wide network (CDWN), *Neural Comput. Appl.*, **32** (2020), 15845–15855. <https://doi.org/10.1007/s00521-018-3877-3>
37. L. L. Du, H. R. Liu, L. Zhang, Y. Lu, M. Y. Li, Y. Hu, et al., Deep ensemble learning for accurate retinal vessel segmentation, *Comput. Biol. Med.*, **158** (2023), 106829. <https://doi.org/10.1016/j.compbiomed.2023.106829>
38. Y. Wu, L. Lin, Automatic lung segmentation in CT images using dilated convolution based weighted fully convolutional network, *J. Phys. Confer. Ser.*, **1646** (2022), 012032. <https://doi.org/10.1088/1742-6596/1646/1/012032>
39. H. Xia, W. Sun, S. Song, X. Mou, Md-net: multi-scale dilated convolution network for CT images segmentation, *Neural Process. Lett.*, **51** (2020), 2915–2927. <https://doi.org/10.1007/s11063-020-10230-x>

40. H. Liu, H. Cao, E. Song, G. Ma, X. Xu, R. Jin, C. C. Hung, A cascaded dual-pathway residual network for lung nodule segmentation in CT images, *Phys. Med.*, **63** (2019), 112–121. <https://doi.org/10.1016/j.ejmp.2019.06.003>
41. H. R. Roth, H. Oda, X. Zhou, N. Shimizu, Y. Yang, Y. Hayash, et al., An application of cascaded 3D fully convolutional networks for medical image segmentation, *Comput. Med. Imaging Graphics*, **66** (2018), 90–99. <https://doi.org/10.1016/j.compmedimag.2018.03.001>
42. A. Lin, B. Chen, J. Xu, Z. Zhang, G. Lu, D. Zhang, Ds-transunet: Dual swin transformer u-net for medical image segmentation, *IEEE Trans. Instrum. Meas.*, **71** (2022), 1–15. <https://doi.org/10.1109/TIM.2022.3178991>
43. F. Milletari, N. Navab, S. A. Ahmadi, V-net: Fully convolutional neural networks for volumetric medical image segmentation, in *2016 Fourth International Conference on 3D Vision (3DV)*, (2016). <https://doi.org/10.48550/arXiv.1606.04797>
44. F. Hoorali, H. Khosravi, B. Moradi, IRUNet for medical image segmentation, *Expert Syst. Appl.*, **191** (2022), 116399. <https://doi.org/10.1016/j.eswa.2021.116399>
45. H. Huang, L. Lin, R. Tong, H. Hu, Q. Zhang, Y. Iwamoto, et al., UNet 3+: A full-scale connected UNet for medical image segmentation, in *2020 IEEE International Conference on Acoustics, Speech and Signal Processing*, (2020). <https://doi.org/10.48550/arXiv.2004.08790>
46. M. Z. Alom, C. Yakopcic, T. M. Taha, V. K. Asari, Nuclei segmentation with recurrent residual convolutional neural networks based U-net(R2U-net), *2018-IEEE National Aerospace and Electronics Conference*, (2018). <https://doi.org/10.1109/NAECON.2018.8556686>
47. T. Shen, X. G. Li, Automatic polyp image segmentation and cancer prediction based on deep learning, *Front. Oncol.*, **12** (2022), 1087438. <https://doi.org/10.3389/fonc.2022.1087438>
48. Z. Han, M. Jian, G. G. Wang, ConvUNeXt: An efficient convolution neural network for medical image segmentation, *Knowl. Based Syst.*, **253** (2022), 109512. <https://doi.org/10.1016/j.knosys.2022.109512>
49. R. Gu, G. Wang, T. Song, R. Huang, M. Aertsen, J. Deprest, et al., CA-Net: Comprehensive attention convolutional neural networks for explainable medical image segmentation, *IEEE Trans. Med. Imaging*, **40** (2020), 699–711. <https://doi.org/10.48550/arXiv.2009.10549>
50. J. Zhang, X. Lv, H. Zhang, B. Liu, AResU-Net: Attention residual U-Net for brain tumor segmentation, *Symmetry*, **12** (2020), 721. <https://doi.org/10.3390/sym12050721>
51. X. Tong, J. Wei, B. Sun, S. Su, Z. Zuo, P. Wu, ASCU-Net: attention gate, spatial and channel attention u-net for skin lesion segmentation, *Diagnostics*, **11** (2021), 501. <https://doi.org/10.3390/diagnostics11030501>
52. J. Fu, J. Liu, H. J. Tian, Y. Li, Y. J. Bao, Z. W. Fang, et al., Dual attention network for scene segmentation, in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, (2019). <https://doi.org/10.48550/arXiv.1809.02983>
53. K. M. He, X. Y. Zhang, S. Q. Ren, J. Sun, Deep residual learning for image recognition, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, (2016). <https://doi.org/10.1109/CVPR.2016.90>
54. M. Jun, J. N. Chen, M. Ng, R. Huang, Y. Li, C. Li, et al., Loss odyssey in medical image segmentation, *Med. Image Anal.*, **71** (2021), 102035. <https://doi.org/10.1016/j.media.2021.102035>

55. R. Wang, T. Lei, R. Cui, B. Zhang, H. Meng, A. K. Nandi, Medical image segmentation using deep learning: a survey, *IET Image Process.*, **162** (2022), 1243–1267. <https://doi.org/10.48550/arXiv.2009.13120>
56. T. Y. Lin, P. Goyal, R. Girshick, K. He, P. Dollár, Focal loss for dense object detection, in *Proceedings of the IEEE international conference on computer vision*, (2017). <https://doi.org/10.48550/arXiv.1708.02002>
57. N. Codella, V. Rotemberg, P. Tschandl, M. E. Celebi, S. Dusza, D. Gutman, et al., Skin lesion analysis toward melanoma detection 2018: A challenge hosted by the international skin imaging collaboration (isic), preprint, arXiv:1902.03368.
58. M. Yahyatabar, P. Jouvét, F. Chériet, Dense-Unet: a light model for lung fields segmentation in Chest X-Ray images, in *42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, (2020). <https://doi.org/10.1109/EMBC44109.2020.9176033>
59. Y. Sun, F. K. Bi, Y. T. Gao, L. Chen, S. T. Feng, A Multi-Attention UNet for Semantic Segmentation in Remote Sensing Images, *Symmetry*, **14** (2022), 906. <https://doi.org/10.3390/sym14050906>
60. Q. Jin, Z. Meng, T. D. Pham, Q. Chen, L. Wei, R. Su, Dunet: A deformable network for retinal vessel segmentation, *Knowledge-Based Systems*, **178** (2018), 149–162. <https://doi.org/10.48550/arXiv.1811.01206>
61. L. Li, M. Verma, Y. Nakashima, H. Nagahara, R. Kawasaki, Iternet: Retinal image segmentation utilizing structural redundancy in vessel networks, in *The IEEE Winter Conference on Applications of Computer Vision*, (2020). <https://doi.org/10.1109/WACV45572.2020.9093621>
62. S. A. Kamran, K. F. Hossain, A. Tavakkoli, S. L. Zuckerbrod, K. M. Sanders, S. A. Baker, RV-GAN: segmenting retinal vascular structure in fundus photographs using a novel multi-scale generative adversarial network, in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, (2021). <https://doi.org/10.48550/arXiv.2101.00535>
63. Y. Zhang, H. Liu, Q. Hu, Transfuse: Fusing transformers and cnns for medical image segmentation, in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, (2021). <https://doi.org/10.48550/arXiv.2102.08005>
64. J. Wei, Y. Hu, R. Zhang, Z. Li, S. K. Zhou, S. Cui, Shallow attention network for polyp segmentation, in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, (2021). <https://doi.org/10.48550/arXiv.2108.00882>
65. J. M. J. Valanarasu, V. M. Patel, Unext: Mlp-based rapid medical image segmentation network, preprint, arXiv:2203.04967.
66. R. Azad, M. Asadi-Aghbolaghi, M. Fathy, S. Escalera, Bi-directional ConvLSTM U-Net with densley connected convolutions, in *Proceedings of the IEEE/CVF international conference on computer vision workshops*, (2019). <https://doi.org/10.48550/arXiv.1909.00166>
67. X. G. Peng, D. L. Peng, MDA-Net: a medical image segmentation network combining dual-path attention mechanism, *Small Microcomputer Syst.*, **43** (2022), 1–9. <http://kns.cnki.net/kcms/detail/21.1106.tp.20220729.1534.034.html>.

