



WRANet: wavelet integrated residual attention U-Net network for medical image segmentation

Yawu Zhao¹ · Shudong Wang¹ · Yulin Zhang² · Sibao Qiao¹ · Mufei Zhang³

Received: 9 January 2023 / Accepted: 16 May 2023
© The Author(s) 2023

Abstract

Medical image segmentation is crucial for the diagnosis and analysis of disease. Deep convolutional neural network methods have achieved great success in medical image segmentation. However, they are highly susceptible to noise interference during the propagation of the network, where weak noise can dramatically alter the network output. As the network deepens, it can face problems such as gradient explosion and vanishing. To improve the robustness and segmentation performance of the network, we propose a wavelet residual attention network (WRANet) for medical image segmentation. We replace the standard downsampling modules (e.g., maximum pooling and average pooling) in CNNs with discrete wavelet transform, decompose the features into low- and high-frequency components, and remove the high-frequency components to eliminate noise. At the same time, the problem of feature loss can be effectively addressed by introducing an attention mechanism. The combined experimental results show that our method can effectively perform aneurysm segmentation, achieving a Dice score of 78.99%, an IoU score of 68.96%, a precision of 85.21%, and a sensitivity score of 80.98%. In polyp segmentation, a Dice score of 88.89%, an IoU score of 81.74%, a precision rate of 91.32%, and a sensitivity score of 91.07% were achieved. Furthermore, our comparison with state-of-the-art techniques demonstrates the competitiveness of the WRANet network.

Keywords Medical image segmentation · Discrete wavelet transform · Downsampling · Residual attention module · Aneurysm image

Introduction

Intracranial aneurysm (IAS) is a common disease with a high mortality rate, and timely and accurate identification and treatment are essential for patients [1]. In medical imaging, computed tomography and magnetic resonance angiography are convenient, effective, and reliable methods for detecting IAS. However, professionally trained physicians are required to analyze and interpret them. Undoubtedly, this will be very time consuming and increase the burden on doctors [2, 3]. Therefore, a robust and reliable artificial intelligence algorithm model is urgently needed to improve this problem.

Medical image segmentation algorithms have been a hot problem for research, which is a practical guide to facilitate pathological assessment and subsequent disease diagnosis and treatment and a very challenging task [4]. Traditional medical image segmentation methods usually use features such as grayscale values, shapes, and textures to segment images; however, the best segmentation effect cannot be achieved by relying on pixels and contours alone. In recent years, with the rapid rise of artificial intelligence technol-

✉ Shudong Wang
shudongwang2013@souhu.com

Yawu Zhao
zhaoyawu9608@163.com

Yulin Zhang
zhangyulin@sdust.edu.cn

Sibao Qiao
siboqiao@126.com

Mufei Zhang
zhangmufeilc@inspur.com

- ¹ School of Computer Science and Technology, China University of Petroleum, Qingdao, Shandong, China
- ² College of Mathematics and System Science, Shandong University of Science and Technology, Qingdao, Shandong, China
- ³ Inspur Cloud Information Technology Co, Inspur, Jinan, Shandong, China

ogy, the emergence of deep learning techniques has widely advanced medical image segmentation, such as brain tumor segmentation [5, 6], skin damage segmentation [7], aneurysm segmentation [8, 9], and ventricular segmentation [10]. Compared with traditional methods, it can capture meaningful contextual information in images, which leads to more accurate diagnosis and segmentation, laying the foundation for further development of medicine [11].

In the field of computer vision, some advanced network structures have been proposed, such as FCN [12], SegNet [13], Deeplabv3+ [14], GoogleNet [15], Residual network [16], and DenseNet [17]. Most of these network structures are improved on the CNN structure [18] and have achieved excellent performance. Compared with CNN methods, FCN methods can classify images at the pixel level while preserving the spatial information of the original image and improving computational efficiency. SegNet consists of an encoder, decoder, and pixel-level classification layers that can independently compute the class probabilities of pixels. Deeplabv3+ improves Xception [19] and depth-separated convolution to improve the semantic performance of the segmentation task. GoogleNet uses the Inception module to break the limits of network depth and width to achieve deeper feature extraction and multi-scale feature processing. Skip connections are used for the Residual block in the Residual network to alleviate the gradient vanishing and network degradation associated with increasing depth in deep neural networks. DenseNet establishes dense connections at shallow and deep layers to improve model performance by enabling feature reuse through feature channel connections.

However, when dealing with medical image segmentation tasks, the U-Net model reduces errors in feature extraction by connecting the encoder and decoder through the skip layer and has achieved great success in medical image segmentation [20]. Inspired by the Residual network, He et al. [21] proposed a Residual learning framework integrated into the U-Net model to extract the features of the deeper network. Zhang et al. [22] proposed a residual context network (ConResNet) to improve the accuracy of pancreatic and brain tumor segmentation. In addition, the proposed attention module was initially used for image classification [23] and recently also used for medical image segmentation [24, 25]. Gu et al. [7] proposed a scale attention module to obtain multiple-scale feature maps. Yu et al. [26] constructed a six-layer residual neural network to fully extract the features of mechanical vibration signals and visualize them using gradient and feature vector-based class activation maps. Cao et al. [27] proposed a transformer-based U-shaped encoder–decoder structure called Swin-Unet, which fuses the extracted contextual features with the multi-scale features of the encoder through a jump connection to compensate for the spatial loss caused by downsampling. Xu et al. [28] proposed a novel adversarial discriminative network (segAN)

with a multi-scale L1 loss function that forces the critic and segmenter to learn to capture both global and local features of long- and short-range spatial relationships between pixels simultaneously, outperforming state-of-the-art methods in terms of Dice scores and accuracy.

More importantly, recent studies show that common image noise will affect the networks judgment of final results to a certain extent, proving the networks weak robustness [29, 30]. Robustness is closely related to standard downsampling methods such as max pooling and average pooling. The traditional downsampling technique cannot achieve the effect of denoising and cannot improve the robustness of the model. In signal processing, wavelet analysis is joint image compression and denoising technology, which can separate the low-frequency and high-frequency data information (some image noise). It has been widely applied in the fields of image processing [31, 32] and signal processing [33, 34].

Inspired by the above work, we proposed a residual attention network model (WRANet) based on wavelet integration to suppress noise propagation, gradient vanishes, and network feature reduction. During the downsampling period, low-frequency information propagates through the network to obtain higher-level features, while high-frequency information is discarded as noise. We used different discrete wavelets to remove image noise and evaluate the Dice coefficient (Dice), intersection over union (IoU), precision (PR), and sensitivity (SE) of the WRANet network on four data sets.

Our main contributions are summarized as follows:

- (1) In the encoder, we proposed a new downsampling module with different wavelet functions instead of pooling layers for downsampling to extract multi-scale tumor features and remove noise efficiently.
- (2) A jump connection layer is used between the encoder and decoder, and a residual attention model is incorporated to mitigate the performance degradation caused by gradient disappearance and feature information loss.
- (3) In this paper, we tested on the aneurysm and polyp datasets and demonstrated through experimental results that WRANet achieves better segmentation performance and is a very effective strategy.

Related work

Wavelets application

Wavelet analysis is widely used in signal analysis, image processing, medical imaging and diagnosis, seismic exploration data processing, etc. In image processing, discrete wavelet transform (DWT) is usually used to decompose 2D images [35]. DWT decomposes the image into LL, LH, HL, and

HH subbands through high-low pass filters. Among them, LL represents the low-frequency coefficient, representing the primary structural information of the picture. At the same time, LH, HL, and HH are the high-frequency coefficients, describing the details of horizontal, vertical, and diagonal coefficients, respectively.

In recent work, wavelets have also been integrated into deep learning models for image reconstruction [36], down-sampling operation [37], and noise suppression [38]. For example, in [39], the author proposed a new wavelet hybrid network (WH-NET) for single image defogging and DWT as a feature extraction layer to achieve a multilevel representation of fuzzy images. Liu et al. [40] proposed a novel multilevel wavelet convolutional neural network model (MWCNN), which introduced wavelet transform to reduce the size of feature images and reconstructed high-resolution feature images using inverse wavelet transform. Verma et al. [41] proposed a wavelet-based convolutional neural network architecture to detect SARS-NCOV, using mother wavelet functions from different families to perform discrete wavelet transform (DWT) and two-stage DWT decomposition to suppress the noise in chest X-ray images. Kang et al. [42] proposed a residual wavelet network, which synergistically combined the expression ability of deep learning with the denoising performance of the wavelet framework. Ma et al. [43] used a trained iWave++ wavelet transform as a new end-to-end method for optimizing images with lossy iWave++ to achieve state-of-the-art compression efficiency compared to deep network-based methods. Huang et al. [44] proposed a wavelet-inspired reversible network (WINNet) by combining wavelet and neural network-based methods to construct a denoising of the sparse coding process, thus recovering the noisy image to a clean one. In our work, we use different wavelet functions to replace the max-pooling layer for down-sampling operation, aiming to eliminate the noise caused by the upper-layer network and extract tumor multi-scale image features to improve the interpretability of the network. The experimental results confirmed that it improved the performance and retained the details and texture features of the original image.

Attention mechanism

The attention mechanism is widely used in deep learning models. It originates from human research on vision and focuses on multiple details by generating context vectors. It has been applied to different scenes, such as text translation [45], image description [46], and speech recognition [47], and achieved great success.

Due to the excellent performance of the attention mechanism, it is gradually applied to medical segmentation tasks. Hu et al. [48] proposed a dense convolutional network with a mixed attention mechanism to calibrate feature maps from

the upper layer using channel and spatial attention. Wang et al. [49] proposed a mixed dilated attentional convolution (HDAC) framework for liver tumor segmentation to fuse information from receptive fields of different sizes. Poudel et al. [50] used a compound-scaled EfficientNet to capture multi-scale global features to resolve limited long-range feature dependencies while exploiting an attention mechanism to suppress noisy and useless features. Zhuang et al. [51] designed a multi-mode cross-latitude attention (MCDA) module to automatically capture valid information from all dimensions of the multi-mode image, achieving excellent segmentation performance in the CEREBRO spinal fluid region. More importantly, we proposed a residual attention network model (RAM), which is used to establish skip connections and improve the context information fusion between encoder and decoder to effectively utilize the context information and the characteristics of the region of concern.

Our method

Overview

The WRANet model mainly comprises an encoder, decoder, and RAM module. We proposed a new down-sampling strategy to mine more compelling features in the image, which used a two-dimensional discrete wavelet transform (DWT) instead of the maximum pooling layer to extract features and remove noise. The decoder integrates the image information extracted by the encoder in the decoder. The bilinear interpolation method is used to restore the image information to reduce the loss of image features. We proposed a residual attention module (RAM) to use better image features, which can automatically focus on areas with significant features while ignoring irrelevant sites during training. Meanwhile, the residual structure can alleviate the problems of feature loss, gradient explosion, and network degradation. Figure 1 shows our proposed WRANet architecture.

In this work, we use bold letters and letters to denote matrices and scalars, e.g., input image \mathbf{X} , residual attention module output \mathbf{x}^o , and wavelet transform functions $\psi(x)$, and $\tilde{\varphi}(x)$, etc.

Downsampling module

Figure 2 shows the details of the down-sampling module. We designed a discrete wavelet transform layer for feature sampling and applied it to improve the performance of the deep neural network for aneurysm image segmentation. Max pooling is a standard down-sampling method in a deep neural network, but it is easy to destroy the structure information of the feature graph. Therefore, we integrate a 2D discrete

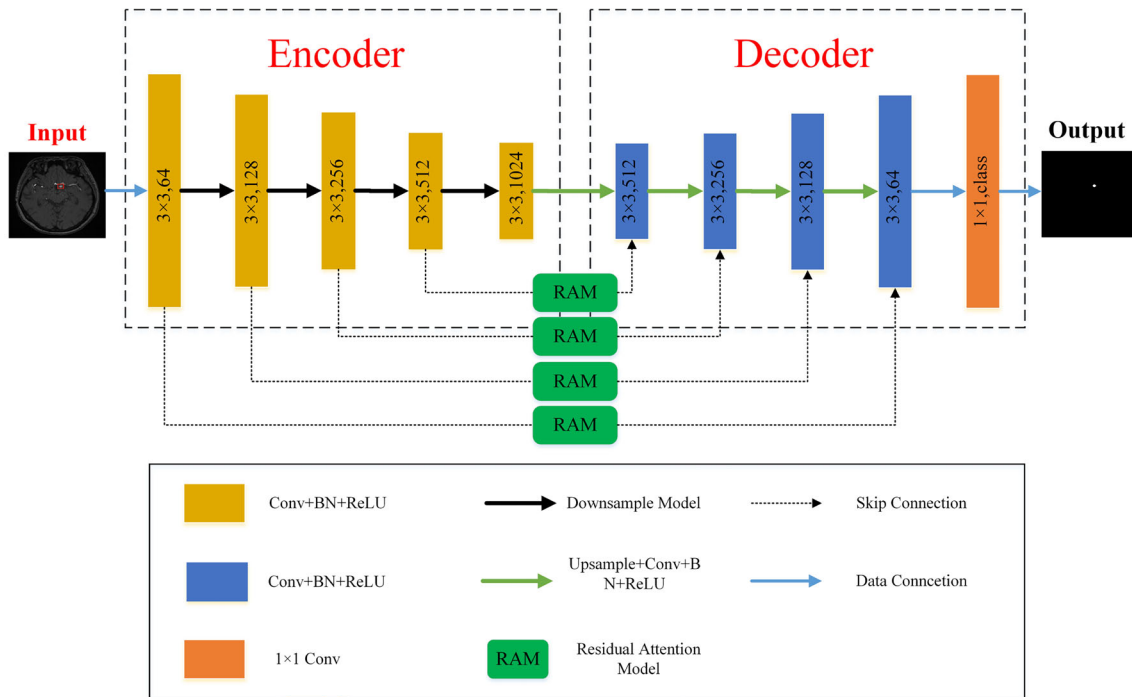
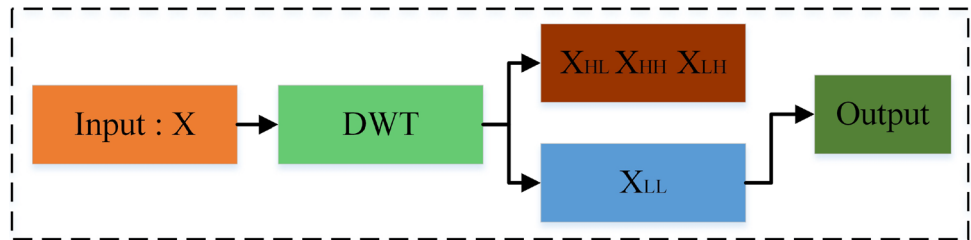


Fig. 1 Our proposed wavelet residual attention network (WRANet). 1×1 and 3×3 represent the size of the convolution kernel, while 64, 128, 256, 512, and 1024 represent the number of output channels. We use 4 downsampling modules, 4 upsampling modules, and 4 residual attention modules

Fig. 2 Details of downsampling module (DWT represents discrete wavelet transform, wavelet subsampling in the WRANet network filters out noise and facilitates information propagation)



wavelet transform into the network to replace the max-pooling layer, which can extract features more effectively and remove unnecessary noise information. The subsampled module we have rewritten can be adapted to various orthogonal and biorthogonal wavelets, such as Haar, Daubechies, Symlets, biorthogonal, and reverse biorthogonal. We first introduce the basic theory of 2D discrete wavelet transform.

2D discrete wavelet transform is closely related to scale function $\varphi(x)$, and wavelet function $\psi(x)$, which form the stable basis of signal space R . In discrete wavelet transform, scale and wavelet function correspond to low-pass filter $l = \{l_i\}_{i \in \mathbb{Z}}$, and high-pass filter $h = \{h_i\}_{i \in \mathbb{Z}}$, and low-pass and high-pass filter decompose data to obtain low-frequency and high-frequency information. Haar, Daubechies, and Symlets are orthogonal wavelets. The corresponding $l = \{l_i\}_{i \in \mathbb{Z}}$ value varies with the corresponding wavelet scale S . We choose the size of the scale S is $1 \leq S \leq 7$, when $S = 1$ is Haar wavelet.

Its filter length is $2S$, while the high-pass filter can be defined as:

$$h_i = (-1)^i l_{2n+1-i}, \tag{1}$$

where, $n \in \{0, 1, 2, 3, \dots\}$, i denotes the size of the filter, z denotes a positive integer.

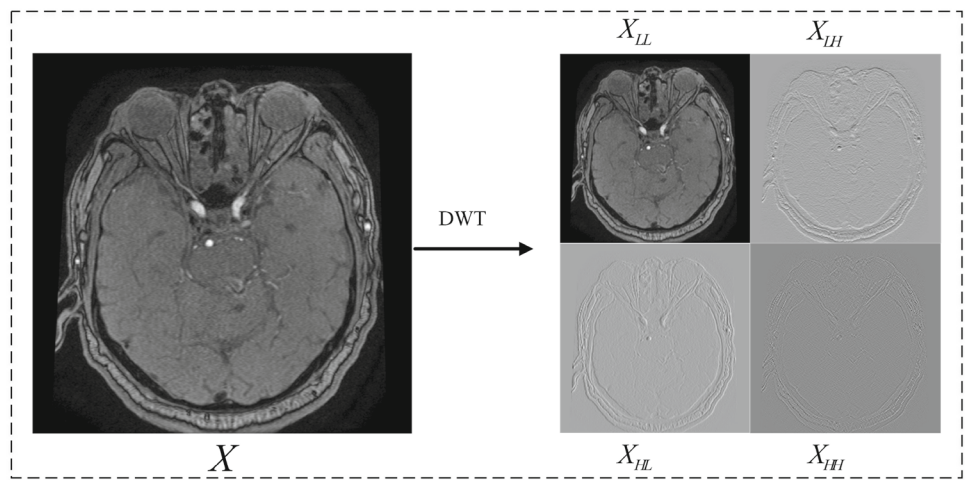
Biorthogonal and reverse biorthogonal wavelets are biorthogonal. If the two dual wavelet functions $\psi(x)$, and $\tilde{\psi}(x)$ satisfy:

$$\psi_{m,n}(x), \tilde{\psi}_{j,k}(x) = \delta(m-n)\delta(n-k), \tag{2}$$

Then $\psi(x)$, and $\tilde{\psi}(x)$ are biorthogonal, and the corresponding scale functions and must also satisfy:

$$\varphi_{j,m}(x), \tilde{\varphi}_{j,n}(x) = \delta(m-n), \tag{3}$$

Fig. 3 2D discrete wavelet transform decomposes aneurysm image



where j represents the scale of wavelet. Then $\psi(x)$ and $\tilde{\varphi}(x)$ are a pair of orthogonal wavelet bases, and there is orthogonality between them. After constructing the biorthogonal wavelet, the original two basis functions are changed into four. Accordingly, $l, \tilde{l}, h,$ and \tilde{h} together constitute orthogonal filter banks, and filter banks can be used for image reconstruction. Similarly, we choose the size of scale S as $1 \leq S \leq 4$, and different scales correspond to different values of $l = \{l_i\}_{i \in \mathbb{Z}}$ of the low-pass filter. A high-pass filter can be defined as:

$$h_i = (-1)^i l_{2n+1-i}, \tag{4}$$

$$\tilde{h}_i = (-1)^i \tilde{l}_{2n+1-i}, \tag{5}$$

2D-DWT of the image can be described as follows: first, 1D-DWT is performed on each row of the image to obtain the low-frequency component L and high-frequency component H of the original image in the horizontal direction; second, 1D-DWT is performed on each column of the obtained data to obtain the low-frequency component LL of the original image in the horizontal and vertical directions. Low frequency in the horizontal direction and high frequency in the vertical direction LH, high frequency in the horizontal direction and vertical direction HL, and low frequency in the horizontal and vertical direction HH. Given 2D image \mathbf{X} , its 2D-DWT can be defined as:

$$\begin{cases} \mathbf{X}_{LL} = l\mathbf{X}l^T, \\ \mathbf{X}_{LH} = l\mathbf{X}h^T, \\ \mathbf{X}_{HL} = h\mathbf{X}l^T, \\ \mathbf{X}_{HH} = h\mathbf{X}h^T, \end{cases} \tag{6}$$

After the input image \mathbf{X} executes the lower sampling block, its output consists of four parts. Where \mathbf{X}_{LL} is the low-frequency component, representing the primary characteristic information of the image; \mathbf{X}_{HL} , \mathbf{X}_{HH} , and \mathbf{X}_{LH} are the

three high-frequency components, representing the vertical, diagonal, and horizontal detail components of the input data \mathbf{X} , respectively, which reflect the noise of the image.

In training the network, we discard the high-frequency information of the image and only keep the low-frequency information for transmission in the network. Taking the Haar wavelet as an example, Fig. 3 shows the decomposition process of the 2D discrete wavelet for aneurysm images.

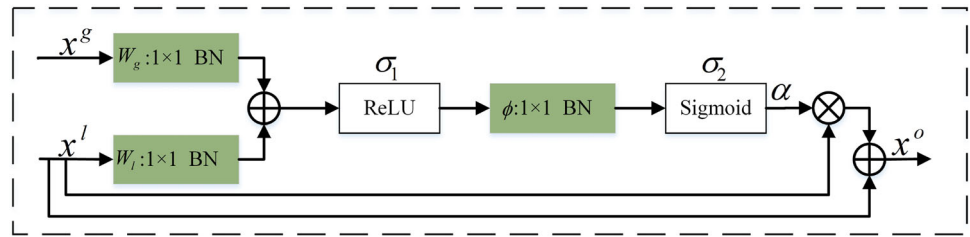
Residual attention module

In our network, the design of the residual attention module is inspired by AG [24], which used an attention gate to recalibrate the feature graph. In addition, the addition of residual structure also avoids feature loss, as shown in Fig. 4.

Let \mathbf{x}^s represent the advanced features from the decoder, the size of the input feature graph is $C \times H \times W$, \mathbf{x}^l is from the low-level features of the encoder, and the size of the input feature graph is $C \times H \times W$, where C represents the feature channel, H and W represent the height and width of the image, respectively. First, we used 1×1 convolution to reduce the dimension of the input feature graph of the two parts so that the number of channels is the same. Then after the batch normalization layer (to ensure relatively stable data distribution and accelerate network convergence), the results of the two parts are added together, and the ReLU activation function is performed (to increase the nonlinear relationship between each layer). One-dimensional convolution is used to reduce the number of feature channels to 1. After the Sigmoid function, the pixel-level attention coefficient α is obtained and multiplied by \mathbf{x}^l to obtain the feature map after calibration. The attention coefficient is as follows:

$$\alpha = \sigma_2 \left(\phi^T \left(\sigma_1 \left(\mathbf{x}^s \mathbf{W}_g^T + \mathbf{x}^l \mathbf{W}_l^T \right) \right) \right), \tag{7}$$

Fig. 4 Residual attention module



For a given matrix ϕ , T denotes its transpose. σ_1 represents ReLU function, and σ_2 represents Sigmoid function so that the value of α falls in $(0,1)$. α assigned different weights to each pixel feature to ensure the interaction between all feature image pixels. Therefore, the recalibrated feature graph \hat{x} is expressed as:

$$\hat{x} = \alpha \cdot x^l, \tag{8}$$

where the number of output channels of \hat{x} is C , and its value is 64, 128, 256, and 512, respectively, at different convolution layers of the encoder. In addition, the residual connection is used to improve information transmission during network training, and the output is as follows:

$$x^o = \hat{x} + x^l, \tag{9}$$

Upsampling model

Figure 5 shows the details of the upsampling module. In this work, the bilinear interpolation method is used to enlarge the feature graph, and 3×3 convolution with a step size of 1 and ReLU function can increase the nonlinear expression ability of the network. Therefore, the primary purpose of the upsampling module proposed is to avoid feature loss and improve the network’s performance.

In the decoding process, after each layer of convolution, the number of channels in the feature graph will be reduced by half, and after the upsampling module, the size of the feature graph will be twice that of the previous layer. Finally,

the network output will be changed to the original input size. We let X be the input of the module so that the output x_{h+1} can be defined as:

$$x_{h+1} = \text{BR} \left(C_{3 \times 3}^1 \left(U^2(x_h) \right) \right), \tag{10}$$

where C represents the 3×3 convolution layer with a step size of 1, U^2 represents the upsampling layer with a step size of 2, and BR represents the BN layer and the ReLU activation layer.

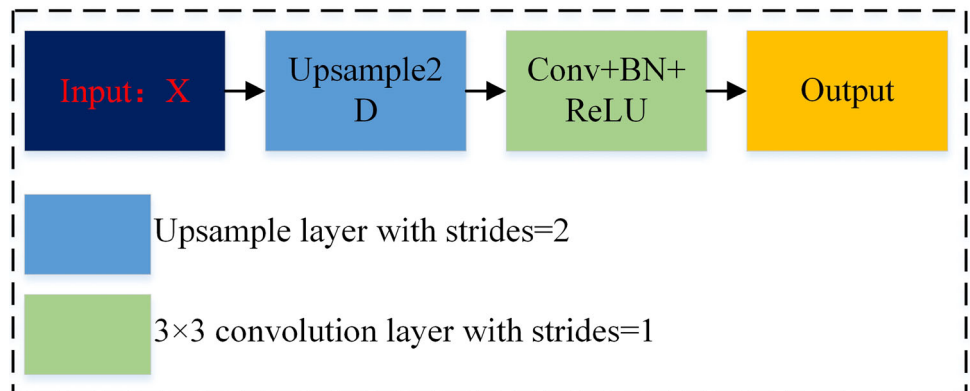
Hybrid loss function

In the medical image segmentation task, the loss function comprises classification and segmentation loss [52]. In this paper, to improve the segmentation performance of medical images, the cross-entropy loss function and Dice loss function are combined as the loss function of this paper. Cross-entropy loss describes the distance between the predicted value and ground truth, while Dice loss measures the degree of consistency between the predicted value and ground truth. Cross-entropy loss L_{BCE} and Dice loss function L_{Dice} are shown below:

$$L_{\text{BCE}} = - \frac{1}{n} \sum_{i=1}^n [y_i \cdot \log(p_i) + (1 - y_i) \log(1 - p_i)], \tag{11}$$

$$L_{\text{Dice}} = 1 - \frac{2 \sum_{i=1}^n p_i y_i}{\sum_{i=1}^n (p_i + y_i) + \lambda}, \tag{12}$$

Fig. 5 Upsampling model structure



where n represents the number of pixels, y_i represents the real label of the i -th pixel, and p_i represents the prediction probability of the i -th pixel belonging to the tumor location. To prevent the denominator from being zero, a smoothing factor λ is added. Although Dice losses are highly compatible with class-unbalanced data, it is challenging to achieve good segmentation performance using Dice losses alone in network training. Due to the pixel imbalance between tumor and normal tissue in medical images, to segment tumor sites more accurately, we combined the advantages of the two-loss functions to construct a mixed loss function L_{seg} , which is defined as follows:

$$L_{\text{seg}} = \alpha \cdot L_{\text{BCE}} + \beta \cdot L_{\text{Dice}}, \quad (13)$$

where, L_{BCE} is the cross-entropy loss, L_{Dice} is the Dice loss, α and β are the hyperparameters in the mixed loss function, $\alpha, \beta \in [0, 1]$. In this paper, $\alpha = 0.4$, $\beta = 0.6$ are selected.

Experimental results

In this work, we used three data sets containing different aneurysm diameters to verify the segmentation performance of the WRANet network for three aneurysms of different sizes. To evaluate the proposed model, the segmentation results were compared with U-Net [20], SegNet [13], DeepLabv3+ [14], ResUnet [21], R2U-net [53], Swin-Unet [27], Att-UNet [24], CE-Net [54], and HarDNet-MSEG [55].

Implementation details and evaluation methods

In this work, we used a combination of Dice and cross-entropy functions as a loss function for the WRANet network, which was implemented based on the Pytorch framework under a GPU server with two Intel(R) Xeon(R) Gold 6226R CPUs. It runs at 2.9 Ghz, has 384 G of RAM, and has two Tesla V100 GPUs with 32GB of RAM. Adam method [56] was used to optimize the parameters of this model, $\beta_1 = 0.9$, $\beta_2 = 0.999$, $\text{eps} = 1\text{e}-8$, $\text{weight decay} = 5\text{e}-4$, the initial learning rate was 0.0001, batch size was 4, and iteration was 200 times. We used fivefold cross-validation and a final evaluation of the test set.

To evaluate the segmentation performance of the model, the following indicators are used for evaluation, including Dice coefficient (Dice), intersection over union (IoU), precision (PR), and sensitivity (SE). They are defined as follows:

$$\text{Dice} = \frac{2 | \mathbf{R}_{\text{gt}} \cap \mathbf{R}_{\text{pred}} |}{| \mathbf{R}_{\text{gt}} | + | \mathbf{R}_{\text{pred}} |}, \quad (14)$$

$$\text{IoU} = \frac{| \mathbf{R}_{\text{gt}} \cap \mathbf{R}_{\text{pred}} |}{| \mathbf{R}_{\text{gt}} \cup \mathbf{R}_{\text{pred}} |}, \quad (15)$$

$$\text{PR} = \frac{\text{TP}}{\text{TP} + \text{FP}}, \quad (16)$$

$$\text{SE} = \frac{\text{TP}}{\text{TP} + \text{FN}}, \quad (17)$$

where, \mathbf{R}_{gt} represents the tumor region, \mathbf{R}_{pred} represents the segmentation result predicted by the model, and TP, FP, FN, and TN refer to a true positive, false positive, true negative, and false negative.

Aneurysm dataset

We follow the principles of the Declaration of Helsinki to conduct our research and have received approval from the Ethics Committee of the Affiliated Hospital of Qingdao University.

We collected 3D-TOF-MRA images of 953 patients with unruptured cystic aneurysm (IAS positive) and 150 regular patients without aneurysm (IAS negative) who underwent physical examination or visited the Affiliated Hospital of Qingdao University from January 2013 to May 2020. After the assessment, exclusion, and screening, 679 patients were identified for the study, including 579 IAS positive (the number of aneurysms was 636) and 100 IAS negative.

Since each patient contained an unbalanced number of sections, we selected five sections from each patient's image for the experiment. The aneurysm was annotated manually by experienced doctors as the ground truth. The size of the slice was adjusted to 256×256 . To verify the segmentation performance of the model for aneurysms of different sizes, we manually selected three different data sets and divided the training set and test set into a 7:3 ratio. Aneurysm images in the training set and test set were not repeated. IAS1 included 89 cases with aneurysms and 20 cases without aneurysms, IAS2 included 290 cases with aneurysms and 50 cases without aneurysms, and IAS3 included 200 cases with aneurysms and 30 cases without aneurysms. The images collected in this paper are from three different equipment manufacturers (Philips, Siemens, and GE), and their details are shown in Table 1.

CVC-ClinicDB [57] is an open polyp dataset consisting of 612 images with a resolution of 384×288 , which we cropped to a size of 256×256 and used as input to the model.

Results and discussion

In this subsection, first, we compare WRANet and SOTA methods' segmentation performance. Second, we perform adequate ablation experiments to verify each component's contribution in WRANet and compare the number of parameters, FLOPs, and inference times. Finally, we validate the

Table 1 Aneurysm datasets overview

	IAS1	IAS2	IAS3
Size (mm)	Size > 7	$3 \leq \text{Size} \leq 7$	Size < 3
Number of IAS	93 (14.6%)	329 (51.7%)	214 (33.7%)
Male (age)		245 (36.1%), 62 ± 12	
Female (age)		434 (63.9%), 65 ± 11	
Field strength		315 (49.5%), 1.5T	
Field strength		321 (50.5%), 3.0T	
Device (Philips)		190 (29.9%)	
Device (Siemens)		69 (10.8%)	
Device (GE)		377 (59.3%)	

generalization performance of the WRANet method with a public dataset.

Comparison with SOTA models. In this work, we used three datasets containing different aneurysm diameters and the CVC-ClinicDB dataset to validate the segmentation performance of the WRANet network for three different sizes of aneurysms. To evaluate the performance of the proposed model, all comparison methods use the same training dataset as the proposed method and compare their segmentation results with U-Net [20], SegNet [13], DeepLabv3+ [14], ResUnet [21], R2U-net [53], Swin-Unet [27], Att-UNet [24], CE-Net [54], and HarDNet-MSEG [55] for comparison. Our WRANet and most of the above segmentation methods are based on CNN, attention mechanism, and residual structure, while the Swin-Unet model is constructed based on Transformer structure. Table 2 shows the experimental results for different diameter-size aneurysm datasets, and Table 3 shows the results for the CVC-ClinicDB dataset.

WRANet segmentation performance analysis. Based on the results in Table 2, our proposed WRANet method outperforms most of the methods on several metrics, indicating our proposed module's validity. Specifically, by comparing with other Dice, IoU, PR, and SE metrics methods, our method outperformed all comparative methods on the aneurysm dataset. Compared to the baseline method, the U-Net model, Dice, IoU, PR, and SE scores improved by 2.42, 2.42, 5.9, and 4.29%, respectively, when the aneurysm diameter was > 7 mm. When the aneurysm diameter was between 3 and 7 mm, Dice, IoU, PR, and SE scores improved by 3.26, 2.48, 8.68, and 5.04%, respectively. When the diameter of the aneurysm was less than 3 mm, Dice, IoU, PR, and SE scores improved by 1.64, 0.39, 1.66, and 1.93%, respectively. This good performance was achieved thanks to the stability of U-Net and the effectiveness of the proposed module, which shows that our method is better segmented and has a lower false alarm rate, revealing that our proposed module is very effective.

To demonstrate the superiority of our method, we have also performed an experimental comparison on the public dataset CVC-ClinicDB. In Table 3, we perform a similar comparison of the CVC-ClinicDB dataset with other methods. Table 3 shows that our method significantly improves the Dice, PR, and SE metrics, with Dice, IoU, PR, and SE scores improving by 1.28, 1.49, 1.86, and 3.42%, respectively. Compared to the baseline method U-Net, the robustness of feature extraction was enhanced by different wavelet basis functions for feature extraction.

To further demonstrate the performance of WRANet, we used nine SOTA models for comparison. Notably, our network still achieves better segmentation performance for aneurysm diameters smaller than 3 mm, demonstrating that the wavelet downsampling module plays a vital role in extracting features and that the RAM module pays more attention to the tumor region and can distinguish to a large extent between the boundaries of tumor and normal tissue. As can be seen in Tables 2 and 3, WRANet achieved the best segmentation performance with Dice, IoU, PR, and SE of 78.99, 68.96, 85.21, and 62.64%, respectively, in the aneurysm dataset IAS2 and 62.64, 51.83, 72.87, and 65.17%, in the aneurysm dataset IAS3 Dice, IoU, PR, and SE were 39.71, 30.40, 46.37, and 40.54% respectively, and in the CVC-ClinicDB dataset, Dice, IoU, PR, and SE were 88.89, 81.74, 91.32, and 91.07%, respectively. However, the DeepLabv3+ network achieved the worst performance on the Dice metric on all datasets.

Ablation studies. We conducted ablation studies separately to verify the effectiveness of the loss function and wavelet transform. When performing feature extraction, some feature information is always lost to a greater or lesser extent, and how to reduce the loss of information during feature propagation is the main task considered in this paper. Therefore, we investigate how the wavelet transform affects the network's performance in feature extraction. As different types of wavelets have different scale functions, which results in different extracted features, which will have a massive impact on the model's performance. We used wavelet basis functions

Table 2 Quantitative comparison of the aneurysm dataset with the SOTA method

Models	IAS1					IAS2					IAS3				
	Dice (%)	IoU (%)	PR (%)	SE (%)		Dice (%)	IoU (%)	PR (%)	SE (%)		Dice (%)	IoU (%)	PR (%)	SE (%)	
U-Net (baseline) [20]	76.57	66.54	79.31	76.69		59.38	49.35	64.19	60.13		38.07	30.01	44.71	38.61	
SegNet [13]	74.52	63.97	76.87	71.59		48.19	32.90	52.92	37.75		26.78	20.36	25.94	36.28	
DeepLabv3+ [14]	68.49	56.93	78.40	68.21		43.15	32.19	41.93	55.86		25.16	17.41	21.38	37.12	
ResUNet [21]	76.12	65.32	75.49	78.97		60.14	49.63	62.99	64.38		32.69	24.89	41.49	32.93	
R2U-net [53]	72.88	62.58	81.01	75.85		57.54	47.64	72.47	60.04		34.17	28.43	42.76	38.22	
Att-UNet [24]	73.64	63.29	73.52	75.13		57.36	46.79	59.01	61.57		35.49	25.97	42.99	38.98	
CE-Net [54]	75.45	64.83	77.31	79.15		51.78	40.59	63.07	55.76		36.53	29.38	44.77	35.42	
HardNet-MSEG [55]	68.73	56.39	65.36	75.16		47.82	44.78	52.45	62.45		27.93	20.54	32.20	36.56	
Swin-UNet [27]	77.52	66.59	85.16	79.08		61.38	50.59	72.52	62.01		39.03	29.76	45.92	39.58	
WRANet	78.99	68.96	85.21	80.98		62.64	51.83	72.87	65.17		39.71	30.40	46.37	40.54	

Bold represent the optimal experimental results

Table 3 Quantitative comparison of the CVC-ClinicDB dataset with the SOTA method

Models	Dice (%)	IOU (%)	PR (%)	SE (%)
U-Net (baseline) [20]	87.61	80.25	89.46	87.65
SegNet [13]	77.59	66.43	88.44	77.82
DeepLabv3+ [14]	78.26	66.57	84.58	83.33
ResUNet [21]	88.18	80.93	89.04	88.34
R2U-net [53]	87.91	80.41	90.56	88.25
Att-UNet [24]	88.32	81.13	89.23	88.50
CE-Net [54]	88.26	81.36	89.19	88.05
HardNet-MSEG [55]	87.59	82.08	89.60	90.38
Swin-UNet [27]	88.35	81.06	89.65	90.72
WRANet	88.89	81.74	91.32	91.07

Bold represent the optimal experimental results

Table 4 Based on a comparison between different wavelet variants and RAM modules

Moedls	Dice (%)	IoU (%)	PR (%)	SE (%)
U-Net (baseline)	76.57	66.54	79.31	76.69
+RAM	76.92	67.34	82.06	76.59
+haar	76.29	66.73	83.70	77.01
+db2	77.95	67.94	83.40	77.47
+db3	77.71	68.09	82.95	78.33
+db4	76.97	67.42	83.53	78.08
+db5	77.61	67.91	84.65	77.24
+db6	77.39	67.69	85.03	77.64
+db7	76.94	67.29	84.10	77.47
+sym2	78.23	68.85	84.49	78.57
+sym3	77.18	67.65	84.07	77.57
+sym4	77.68	68.13	84.04	77.62
+sym5	77.63	68.00	82.64	78.70
+sym6	78.72	68.46	84.93	80.31
+sym7	78.65	68.39	84.86	78.52
+bior1.1	78.41	68.66	82.41	79.07
+bior2.2	77.14	67.90	85.18	76.03
+bior3.3	78.48	68.87	83.66	78.85
+bior4.4	78.10	68.67	83.08	78.94
+rbio1.1	77.65	67.97	83.48	78.58
+rbio2.2	77.08	67.25	82.40	78.77
+rbio3.3	75.46	65.83	83.73	75.26
+rbio4.4	78.29	68.73	83.45	79.78
+RAM+sym6+L _{BCE}	78.16	68.67	82.97	79.64
+RAM+sym6+L _{Dice}	78.63	68.89	84.71	77.97
+RAM+sym6+L _{seg}	78.99	68.96	85.21	80.98

Bold represent the optimal experimental results

replacing the maximum pooling layer in the U-Net network to investigate their respective performance, such as haar, db, sym, bior, and rbio. We used U-Net as a baseline model to compare with our proposed WRANet, and the experimental results are shown in Table 4. Since direct observation of the model segmentation result images does not accurately judge the model structure, minor differences are not directly observable using the naked eye. Therefore, we used a series of evaluation metrics to assess the model performance, such as the Dice coefficient, IoU, PR, and SE.

Table 4 shows different wavelet basis functions as the scores of various indicators of the pooling layer, and wavelet 'sym6' achieved the highest Dice, IoU, PR, and SE scores. 'Daubechies' wavelet can improve the performance of the network at low order ('db2'), and the Dice score is 77.95%. However, it will reduce the learning ability of the network with the increase of the order, resulting in a decrease in performance, such as high order ('db7') 'Daubechies' wavelet has a Dice score of 76.94%. However, the 'Symlets' wavelet was accompanied by an increase in order ('sym6'), and the performance of the network became better with a Dice score of 78.72%; the lower the order ('sym3', 'sym4', and 'sym5'), the worse the network performs, with Dice scores of 77.18, 77.68, and 77.63%. The biorthogonal wavelets 'Biorthogonal' and 'Reversebior' also improved the segmentation performance. The downsampling process using symmetric wavelets generally performs better than asymmetric wavelets. We choose L_{seg} as the loss function to study wavelet denoising performance. To make the experiment more convincing, we study the influence of L_{BCE} and L_{Dice} on the accuracy of segmentation results, respectively. In addition, we also introduce a RAM module, which can make the feature maps more focused on the target region, making its predictions more closely match the ground truth.

We validate our proposed method on private and public datasets, respectively. In addition, we found that it is more effective to perform the wavelet downsampling process in the encoder stage because its feature map is more concentrated, which can extract more helpful feature information, remove unnecessary noise, reduce the false positive rate, and improve the network of robustness and segmentation performance.

Complexity analysis. We analyzed the number of parameters, FLOPs, and inference time for the WRANet and SOTA models, and the analysis results are shown in Table 5. Although the WRANet method has higher complexity, it has fewer parameters and a shorter inference time than R2U-net, Att-UNet, HarDNet-MSEG, and Swin-UNet, due to its improved performance in aneurysm and polyp segmentation, making it acceptable in our study.

Visualization of segmentation results. Figures 6, 7, and 8 show the segmentation visualization results for the IAS1, IAS2, and IAS3 datasets. It is clear that our method yields

more accurate results. The first column represents the original image, the second column represents Ground Truth, and the remaining columns are U-Net, SegNet, DeepLabv3+, ResUnet, R2U-net, Att-UNet, CE-Net, HarDNet-MSEG, and WRANet. Visually, our WRANet achieves good segmentation performance while accurately identifying tumor locations. It can be seen in the IAS3 data set that when the aneurysm is small, our method still achieves better performance, while the SegNet, U-Net, ResUnet, and R2U-net models cannot accurately segment the aneurysm. The proposed model can learn more detailed features from the dataset. Thus, the results show better performance compared to other SOTA models.

Figure 9 represents the visual segmentation results for the CVC-ClinicDB dataset. As shown in Fig. 9, the U-Net, Att-UNet, and Swin-UNet methods have inaccurate boundaries for polyp segmentation, and our WRANet method performs better visually. This indicates that our proposed RAM module and wavelet feature extraction layer play an important role in extracting more complete tumor features and improving tumor segmentation accuracy.

Conclusion

In this paper, we proposed a wavelet residual attention network WRANet to improve aneurysm segmentation performance. We design a discrete wavelet transform layer (DWT) to replace conventional downsampling operations (max pooling and average pooling) to capture tumor features and remove noise information effectively. We used the RAM module to capture contextual information and fuse multi-scale features, which can recalibrate attention weights to make the network pay more attention to tumor regions, use skip connections to enhance fusion features, and make full use of standard features to improve the generalization performance of the model, which reflects the superiority of attention mechanism and residual connection. In addition, the proposed method overcomes the propagation of high-frequency noise of images in the network and the loss of information to a certain extent. Compared with the standard U-Net and its variants ResUnet, Att-UNet, and R2U-net, the performance of the WRANet network is significantly improved.

The experimental results of image segmentation on the IAS1, IAS2, and IAS3 datasets containing different tumor sizes show that the WRANet network improves the segmentation performance of small tumors and has the potential to help doctors improve the efficiency of diagnosis in clinical practice. It can be seen that the proposed WRANet network is a promising method for medical segmentation. The technique can be easily applied to 3D medical image segmentation in future work.

Table 5 Complexity comparison of our WRANet against SOTA methods model

Methods	Parameters (M)	FLOPs (B)	Inference time (s)
U-Net [20]	8.6	0.3	0.2826
SegNet [13]	29.4	0.6	0.2893
DeepLabv3+ [14]	31.6	0.8	0.3081
ResUnet [21]	13.9	0.5	0.2917
R2U-net [53]	39.1	2.4	0.5863
Att-UNet [24]	34.9	1.1	0.3119
CE-Net [54]	29.0	0.1	0.3529
HarDNet-MSEG [55]	35.4	0.2	0.4127
Swin-Unet [27]	37.2	2.3	0.5025
WRANet	34.6	1.1	0.3046

Bold represent the optimal experimental results

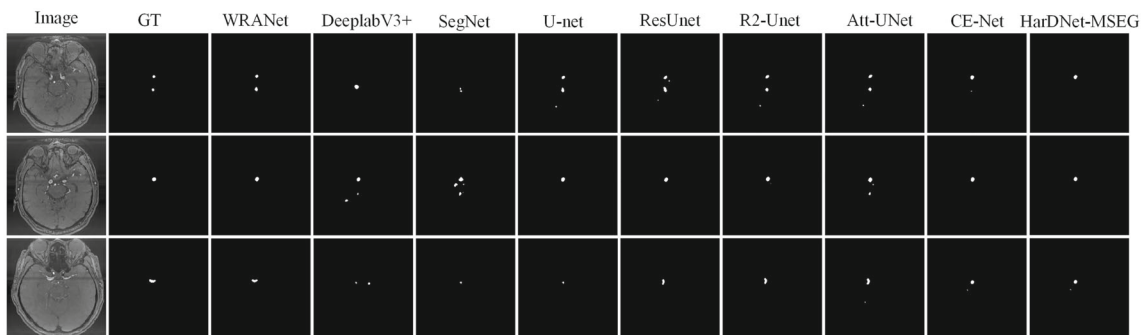


Fig. 6 Segmentation visualization results for each SOTA model on dataset IAS1

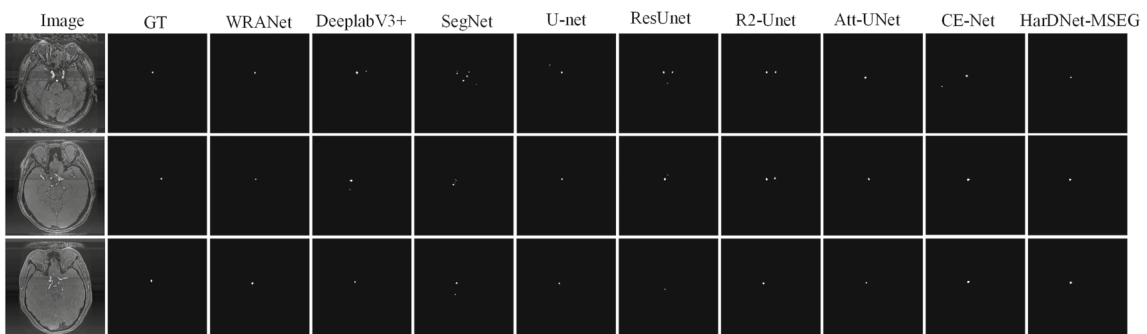


Fig. 7 Segmentation visualization results for each SOTA model on dataset IAS2

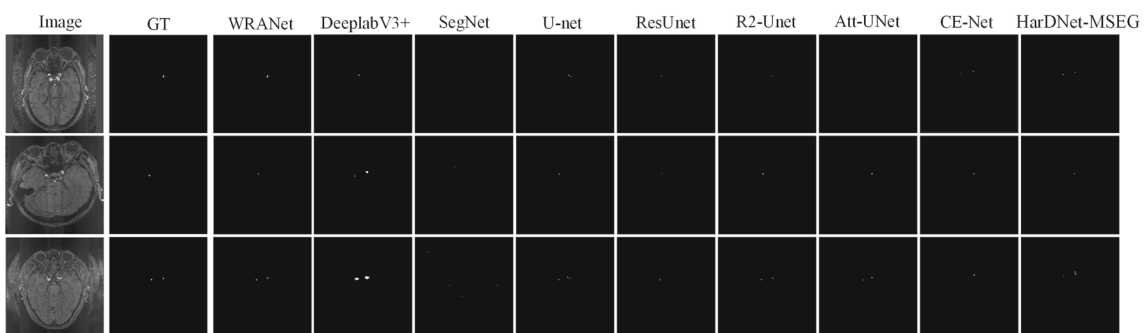


Fig. 8 Segmentation visualization results for each SOTA model on dataset IAS3

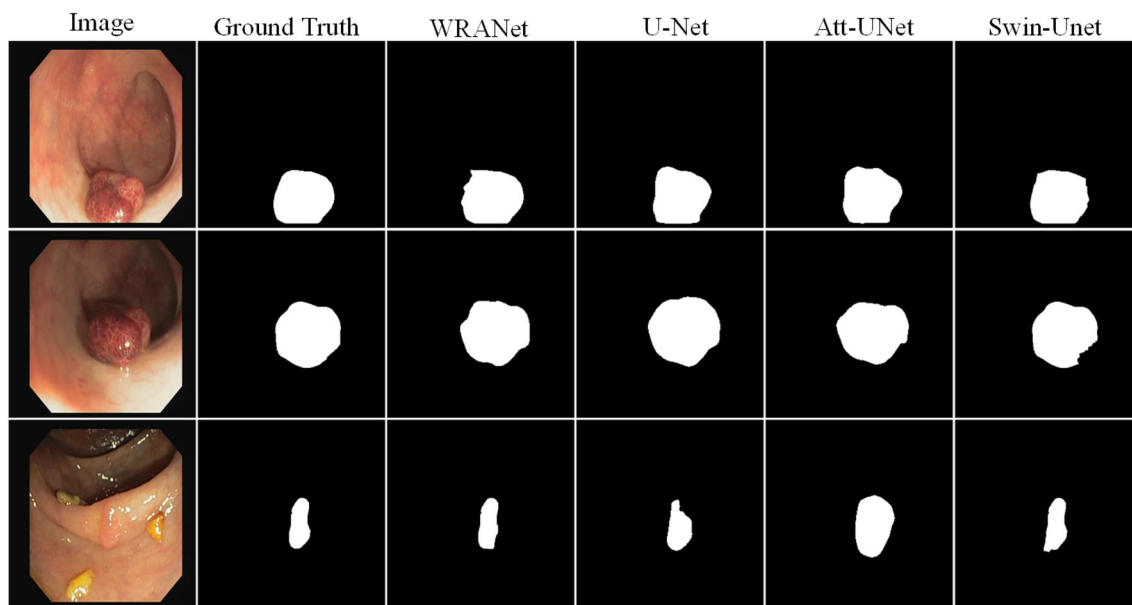


Fig. 9 Segmentation visualization results for each SOTA model on dataset CVC-ClinicDB

Acknowledgements This work was funded by the National Key Research and Development Program (No. 2021YFA10000102-3), National Natural Science Foundation of China (Grant no. 61873281).

Availability of data and materials The datasets generated during and analyzed during the current study are not publicly available as we have signed a non-disclosure agreement with the hospital to protect patient information, but are available from the corresponding author on reasonable request.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Agid R, Andersson T, Almqvist H et al (2010) Negative CT angiography findings in patients with spontaneous subarachnoid hemorrhage: when is digital subtraction angiography still needed? *Am J Neuroradiol* 31(4):696–705
2. Yang ZL, Ni QQ, Schoepf UJ et al (2017) Small intracranial aneurysms: diagnostic accuracy of CT angiography. *Radiology* 285(3):941952
3. Bullitt E, Gerig G, Pizer SM et al (2003) Measuring tortuosity of the intracerebral vasculature from MRA images. *IEEE Trans Med Imaging* 22(9):1163–1171
4. Litjens G, Kooi T, Bejnordi BE et al (2017) A survey on deep learning in medical image analysis. *Med Image Anal* 42:60–88
5. Zhao X, Wu Y, Song G et al (2018) A deep learning model integrating FCNNs and CRFs for brain tumor segmentation. *Med Image Anal* 43:98–111
6. Wang L, Nie D, Li G et al (2019) Benchmark on automatic six-month-old infant brain segmentation algorithms: the iSeg-2017 challenge. *IEEE Trans Med Imaging* 38(9):2219–2230
7. Gu R, Wang G, Song T et al (2020) CA-Net: comprehensive attention convolutional neural networks for explainable medical image segmentation. *IEEE Trans Med Imaging* 40(2):699–711
8. Shi Z, Miao C, Schoepf UJ et al (2020) A clinically applicable deep-learning model for detecting intracranial aneurysm in computed tomography angiography images. *Nat Commun* 11(1):1–11
9. Sichtermann T, Faron A, Sijben R et al (2019) Deep learning based detection of intracranial aneurysms in 3D TOF-MRA. *Am J Neuroradiol* 40(1):25–32
10. Leclerc S, Smistad E, Pedrosa J et al (2019) Deep learning for segmentation using an open large-scale dataset in 2D echocardiography. *IEEE Trans Med Imaging* 38(9):2198–2210
11. Panayides AS, Amini A, Filipovic ND et al (2020) AI in medical imaging informatics: current challenges and future directions. *IEEE J Biomed Health Inform* 24(7):1837–1857
12. Long, Jonathan, Evan Shelhamer, and Trevor Darrell. Fully convolutional networks for semantic segmentation. *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2015
13. Badrinarayanan V, Kendall A, Cipolla R (2017) Segnet: a deep convolutional encoder–decoder architecture for image segmentation. *IEEE Trans Pattern Anal Mach Intell* 39(12):2481–2495
14. Chen, Liang-Chieh, et al. Encoder-decoder with atrous separable convolution for semantic image segmentation. *Proceedings of the European conference on computer vision (ECCV)*. 2018, pp.801–818
15. Szegedy, Christian, et al. Going deeper with convolutions. *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2015, pp.1–9

16. He, Kaiming, et al. Deep residual learning for image recognition. Proceedings of the IEEE conference on computer vision and pattern recognition. 2016, pp.770-778
17. Huang, Gao, et al. (2017) Densely connected convolutional networks. Proceedings of the IEEE conference on computer vision and pattern recognition. pp.4700-4708
18. Gu J, Wang Z, Kuen J et al (2018) Recent advances in convolutional neural networks. Pattern Recogn 77:354–377
19. Chollet F (2017) Xception: Deep learning with depthwise separable convolutions[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 1251-1258
20. Ronneberger O, Fischer P, Brox T (2015) U-net: Convolutional networks for biomedical image segmentation[C]//International Conference on Medical image computing and computer-assisted intervention. Springer, Cham, 234-241
21. He K, Zhang X, Ren S, et al (2016) Deep residual learning for image recognition[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 770-778
22. Zhang J, Xie Y, Wang Y et al (2020) Inter-slice context residual learning for 3D medical image segmentation. IEEE Trans Med Imaging 40(2):661–672
23. Hu J, Shen L, Sun G (2018) Squeeze-and-excitation networks[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 7132-7141
24. Oktay O, Schlemper J, Folgoc L L, et al (2018) Attention u-net: Learning where to look for the pancreas. arXiv preprint [arXiv:1804.03999](https://arxiv.org/abs/1804.03999)
25. Sinha A, Dolz J (2020) Multi-scale self-guided attention for medical image segmentation. IEEE J Biomed Health Inform 25(1):121–130
26. Yu S et al (2023) TDMSAE: a transferable decoupling multi-scale autoencoder for mechanical fault diagnosis. Mech Syst Signal Process 185:109789
27. Cao H, Wang Y, Chen J, et al (2021) Swin-unet: Unet-like pure transformer for medical image segmentation. arXiv preprint [arXiv:2105.05537](https://arxiv.org/abs/2105.05537)
28. Xue Y, Xu T, Zhang H et al (2018) Segan: adversarial network with multi-scale l1 loss for medical image segmentation. Neuroinformatics 16:383–392
29. Kurakin A, Goodfellow IJ, Bengio S (2018) Adversarial examples in the physical world. Artificial intelligence safety and security. Chapman and Hall/CRC, Boca Raton, pp 99–112
30. Goodfellow IJ, Shlens J, Szegedy C (2014) Explaining and harnessing adversarial examples. arXiv preprint [arXiv:1412.6572](https://arxiv.org/abs/1412.6572)
31. Chang SG, Yu B, Vetterli M (2000) Adaptive wavelet thresholding for image denoising and compression. IEEE Trans Image Process 9(9):1532–1546
32. Li Q, Shen L, Guo S et al (2021) WaveCNet: wavelet integrated CNNs to suppress aliasing effect for noise-robust image classification. IEEE Trans Image Process 30:7074–7089
33. Singh BN, Tiwari AK (2006) Optimal selection of wavelet basis function applied to ECG signal denoising. Digit Signal Process 16(3):275–287
34. Cohen R (2012) Signal denoising using wavelets. Department of Electrical Engineering Technion, Israel Institute of Technology, Haifa, Project Report, p 890
35. Mallat S (1999) A wavelet tour of signal processing. Elsevier, Amsterdam
36. Lai Z, Qu X, Liu Y et al (2016) Image reconstruction of compressed sensing MRI using graph-based redundant wavelet transform. Med Image Anal 27:93–104
37. Williams T, Li R (2018) Wavelet pooling for convolutional neural networks[C]//International Conference on Learning Representations
38. Duan Y, Liu F, Jiao L et al (2017) SAR image segmentation based on convolutional-wavelet neural network and Markov random field. Pattern Recogn 64:255–267
39. Dharejo FA, Zhou Y, Deeba F et al (2021) A deep hybrid neural network for single image dehazing via wavelet transform. Optik 231:166462
40. Liu P, Zhang H, Zhang K, et al. (2018) Multi-level wavelet-CNN for image restoration[C]//Proceedings of the IEEE conference on computer vision and pattern recognition workshops. 773-782
41. Verma AK, Vamsi I, Saurabh P et al (2021) Wavelet and deep learning based detection of SARS-nCoV from thoracic X-ray images for rapid and efficient testing. Expert Syst Appl 185:115650
42. Kang E, Chang W, Yoo J et al (2018) Deep convolutional framelet denoising for low-dose CT via wavelet residual network. IEEE Trans Med Imaging 37(6):1358–1369
43. Ma H, Liu D, Yan N et al (2020) End-to-end optimized versatile image compression with wavelet-like transform. IEEE Trans Pattern Anal Mach Intell 44(3):1247–1263
44. Huang JJ, Dragotti PL (2022) WINNet: wavelet-inspired invertible network for image denoising. IEEE Trans Image Process 31:4377–4392
45. Bahdanau D, Cho K, Bengio Y (2014) Neural machine translation by jointly learning to align and translate. arXiv preprint [arXiv:1409.0473](https://arxiv.org/abs/1409.0473)
46. Xu K, Ba J, Kiros R, et al (2015) Show, attend and tell: Neural image caption generation with visual attention[C]//International conference on machine learning. PMLR, 2048-2057
47. Chorowski JK, Bahdanau D, Serdyuk D, Cho K, Bengio Y (2015) Attention-based models for speech recognition. Adv Neural Inf Process Syst 28
48. Hu H, Li Q, Zhao Y et al (2020) Parallel deep Learning algorithms with hybrid attention mechanism for image segmentation of lung tumors. IEEE Trans Ind Inf 17(4):2880–2889
49. Wang Z, Zou Y, Liu PX (2021) Hybrid dilation and attention residual U-Net for medical image segmentation. Comput Biol Med 134:104449
50. Poudel S, Lee SW (2021) Deep multi-scale attentional features for medical image segmentation. Appl Soft Comput 109:107445
51. Zhuang Y, Liu H, Song E, Ma G, Xu X, Hung C-C (2022) APRNet: a 3D anisotropic pyramidal reversible network with multi-modal crossdimension attention for brain tissue segmentation in MR images. IEEE J Biomed Health Inform 26(2):749–761. <https://doi.org/10.1109/JBHI.2021.3093932>
52. Xie Y, Zhang J, Xia Y et al (2020) A mutual bootstrapping model for automated skin lesion segmentation and classification. IEEE Trans Med Imaging 39(7):2482–2493
53. Alom M Z, Hasan M, Yakopcic C, et al (2018) Recurrent residual convolutional neural network based on u-net (r2u-net) for medical image segmentation. arXiv preprint [arXiv:1802.06955](https://arxiv.org/abs/1802.06955)
54. Gu Z, Cheng J, Fu H et al (2019) Ce-net: context encoder network for 2d medical image segmentation. IEEE Trans Med Imaging 38(10):2281–2292
55. Huang CH, Wu HY, Lin YL (2021) Hardnet-mseg: A simple encoder-decoder polyp segmentation neural network that achieves over 0.9 mean dice and 86 fps. arXiv preprint [arXiv:2101.07172](https://arxiv.org/abs/2101.07172)
56. Kingma DP, Ba J (2014) Adam: A method for stochastic optimization. arXiv preprint [arXiv:1412.6980](https://arxiv.org/abs/1412.6980)
57. Bernal J, Snchez FJ, Fernandez-Esparrach G et al (2015) WMDOVA maps for accurate polyp highlighting in colonoscopy: validation vs. saliency maps from physicians. Comput Med Imaging Graph 43:99–111