SSCA-Net: Simultaneous Self- and Channel-attention Neural Network for Multi-scale Structure-Preserving Vessel Segmentation

Jiajia Ni
Shenzhen Institutes of Advanced Technology Chinese Academy of Sciences
Jianhuang Wu (jh.wu@siat.ac.cn)
Shenzhen Institutes of Advanced Technology https://orcid.org/0000-0002-0158-0890
Jing Tong
Hohai University - Changzhou Campus
Mingqiang Wei
Nanjing University of Aeronautics and Astronautics
Zhengming Chen
Hohai University - Changzhou Campus

Research

Keywords: Multi-scale structure-preserving vessel segmentation, Simultaneous self- and channel-attention, Deep learning, Attention mechanisms, Pyramid pooling

Posted Date: December 10th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-122970/v1

SSCA-Net: Simultaneous Self- and Channel-attention Neural Network for Multi-scale Structure-preserving Vessel Segmentation

Jiajia Ni[1,2], Jianhuang Wu[1], Jing Tong[2], Mingqiang Wei[3] and Zhengming Chen[2]

1. Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, China

2.College of Internet of Things Engineering, HoHai University Changzhou, China

3. Nanjing University of Aeronautics and Astronautics, China

***Corresponding author:**

Jianhuang Wu

Address: Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, 1068 Xueyuan Boulevard, Xili Nanshan, Shenzhen, 518055 China.

Email: jh.wu@siat.ac.cn.

Abstract

Background: Vessel segmentation is a fundamental, yet not well-solved problem in medical image analysis, due to complicated geometrical and topological structures of human vessels. Unlike existing rule- and conventional learning-based techniques, which hardly capture the location of tiny vessel structures and perceive their global spatial structures,

Methods: we propose Simultaneous Self- and Channel-attention Neural Network (termed SSCA-Net) to solve the multi-scale structure-preserving vessel segmentation (MSVS) problem. SSCA-Net differs from the conventional neural networks in modeling image global contexts, showing more power to understand the global semantic information by both self- and channel-attention (SCA) mechanism, and offering high performance on segmenting vessels with multi-scale structures. Specifically, the SCA module is designed and embedded in the feature decoding stage to learn SCA features at different layers, which the self-attention is used to obtain the position information of the feature itself, and the channel attention is designed to guide the shallow features to obtain global feature information.

Results: Three blood vessel data sets are train and validate the models. our SSCA-Net achieves 96.21% in Dic and 92.70% in Mean IoU on the intracranial vessel dataset and achieved 98.20 %, 83.52% and 96.14% in AUC, Sen and Acc respectively on retinal vessel dataset. The obtain model can segment the leg arteries and Dic score is 97.21% and the Mean IoU score is 94.42%.

Conclusions: The results demonstrated that the proposed SSCA-Net clear improvements of our method over the state-of-the-arts in terms of preserving vessel details and global spatial structures.

Keywords: Multi-scale structure-preserving vessel segmentation, Simultaneous self- and channel-attention, Deep learning, Attention mechanisms, Pyramid pooling

1. Background

Vessel segmentation aims to automatically or semi-automatically detect the boundaries (consisting of pixels) of blood vessels within 2D or 3D medical images such as Computed Tomography (CT) or Magnetic Resonance Angiography (MRA) images. As one of the most challenging tasks in medical image segmentation (MIS), vessel segmentation can deliver significant information about the shapes and volumes of vessels, which are critical to the diagnosis and treatment of vessel diseases [1, 2].

The most successful type of models for vessel segmentation is deep learning-based techniques, especially convolutional neural network (CNN) based frameworks, which shown to be a powerful and robust tool in separating homogeneous areas of medical images [3-8], as shown in Figure 1. Although those methods improve the accuracy of semantic segmentation, it is often done by increasing the number of feature maps or the number of skip connections. Intuitively, maintaining lots of feature maps or complex network structures can boost the segmentation performance. However, doing that is not optimal to both reduce network over-fitting and improve the segmentation accuracy. In addition, due to the characteristics of CNN, those method use shared weights, which may cause the network to lose some spatial feature information and channel feature information. To address these problems, people began to use the attention mechanism.

Recent studies have validated that the attention-based global features are important for semantic segmentation [9]. For instance, the pyramid attention network [10] exploits the impact of global contextual information on semantic segmentation, and global attention up-sampling to replace the bilinearly up-sampling structure. Yuan et al. [11] presented an object context pooling scheme to improve the segmentation accuracy of the entire grid. Attention U-Net [12] demonstrate the implementation of attention gate in a standard U-Net architecture and apply it for the segmentation of abdominal organs. However, these methods of capturing global feature

information often utilize consecutive pooling and stride convolutional operations. As known, this kind of operation can lead to the loss of location and spatial information.

We argue that the attention mechanism is helpful for vessel segmentation and assume that taking a more comprehensive use of attention would boost the segmentation performance. In this paper, we propose novel building blocks including a self-attention and channel attention module (SCA) to connect shallow features and high-level features. Compared to the original U-Net, which uses simple skip connections to connect shallow and high-level features. Our proposed SCA block could capture wider and deeper semantic features by infusing the attention mechanism. Additionally, we also use squeeze and excitation pyramid pooling (SEPP) [13], which can aggregate feature maps in the same convolutional layer to extract enriched feature representations. It can better increase the resilience of the entire network and the robustness of the network. Furthermore, to decrease the structures, the network replaces ResNet [14] with the Relu Feature Unit (RFU). In summary, the RFU block is proposed to reduce model parameters and optimize network structure, followed by the SEPP block for further context information with multi-scale pooling operations. Integrating the RFU block and the SEPP block with the backbone encoder-decoder structure and use the SCA module as skip connection structure, we develop an end-to-end vessel segmentation neural network named SSCA-Net. It relies on the SAC block and the SEPP block to get more abstract spatial and channel features and preserve more multi-scale spatial information to boost the performance of vessel segmentation. The main contributions of this work are summarized as follows:

1) We propose an SCA block to better integrate shallow features and deep features to obtain more abstract features and preserve more spatial information.
2) We propose novel building networks including a multi-scale spatial and feature attention module, a novel multi-scale feature fuse module and a simple feature extraction block that decreasing model parameters.
3) We apply the proposed method to three different datasets, namely intracranial blood vessel, retinal vessel data, and leg vessel data. Results show that the proposed method outperforms the state-of-the-art methods in these different tasks.

The paper is organized as follows: Section II discusses related work. The architectures of the proposed SSCA-Net models are presented in Section III. Section IV, explains experiments, results, Discussion and ablation study. The conclusion is discussed in Section V

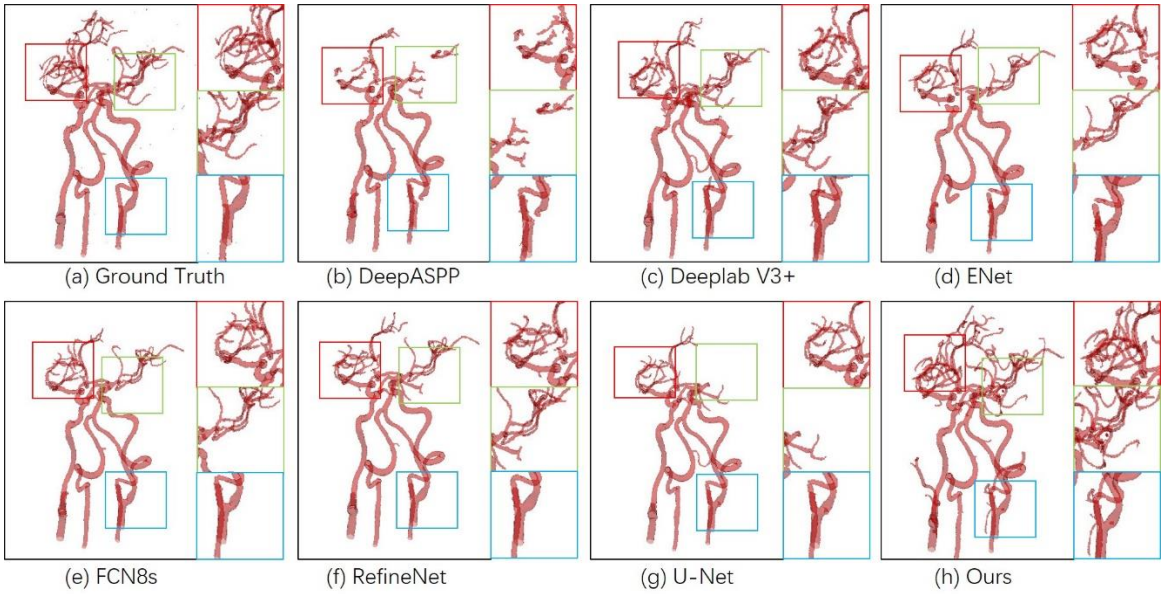


Figure 1: Medical image segmentation results tested in the dataset of intracranial artery. From (a)-(h): (a) ground truth, and the segmentation results of (b) DeepASPP, (c) DeepLab V3+, (d) ENet, (e) FCN8s, (f) RefineNet, (g) U-Net, and (h) ours, respectively. Our SSCA-Net can perform the segmentation of intracranial arteries effectively while preserving multi-scale structures of vessels, especially the tinny-scale structures.

2. Related Work

Traditional Deep Learning: Deep learning-based semantic segmentation methods can be roughly divided into two categories: FCN-based and U-Net-based. The FCN-based methods are characterized by the direct use of high-level semantic segmentation, which remove the last two full-connect layers to classify each pixel. Due to solely using high-level features, FCN-based methods perform not well and even leads to less accuracy on some data sets. To address this problem, a variety of improvements are proposed, e.g., combining FCN with graphical models like Markov Random Fields (MRFs) [13] and Conditional Random Fields (CRFs) [14] [15] to refine the segmentation prediction. Furthermore, the U-Net-based methods have been proposed, which can be characterized by using a skip connect to combine low- and high-level features to predict the segmentation. It has become a popular neural network architecture and shown promising results on different medical image segmentation tasks [8, 16, 17].

Context Aggregation: In recent years, various methods have explored contextual information by many researchers, which are more complicated than the original U-Net, for example, the Deeplab series [18-21]. The Deeplab method introduces the atrous convolution and atrous spatial pyramid pooling (ASPP) network structure. The latest Deeplab v3+ [19] introduces decoder and Xception [22] to improve the network performance and reduce the computational complexity. PSPNet [23] adopts the pyramid pooling module to partition the feature map into different scale regions. Yu et al.[24] developed a Context Prior to distinguish the intra-class and interclass context clearly. Lin et al. [17] proposed a multi-path refinement network, which contains residual convolution unit, multi-resolution fusion, and chained residual pooling. Yang et al. [25] proposed the densely connected atrous spatial pyramid pooling (DenseASPP), which connects a set of atrous convolutional layers densely. Furthermore, to improve the resilience of the network, the pyramid structure of the space is applied to the semantic segmentation [26, 27].

Attention model: The attention mechanism was first successfully applied in natural language processing tasks, and then it was well extended to solve image processing tasks. Zhao et al. [28] proposed the pointwise spatial attention network to guide contextual information collection. The squeeze-and-excitation networks [29] adopted a channel-wise relationships attention mechanism to enhanced the representational power of the network. Woo et al. [30] proposed Convolutional Block Attention Module for feed-forward convolutional neural networks. OCNet [11] and CCNet [31] utilized the self-attention mechanism to obtain the contextual information. Zhong et al. [32] proposed a novel squeeze-and-attention network architecture for obtaining an enhanced pixel-wise prediction. Niu et al. [33] proposed a novel attention-based framework to adaptively capture global correlations from the perspective of space, channel and category. BAM [34] used a simple yet effective attention module, which infers an attention map along channel and spatial. Ni et al. [35] proposed a spatial and channel-based attention-based convolutional neural network (GC-Net) to segment medical image data. Our SSCA-Net network is different from the methods mentioned above. The contextual information is aggregated by both self-attention and channel attention modules.

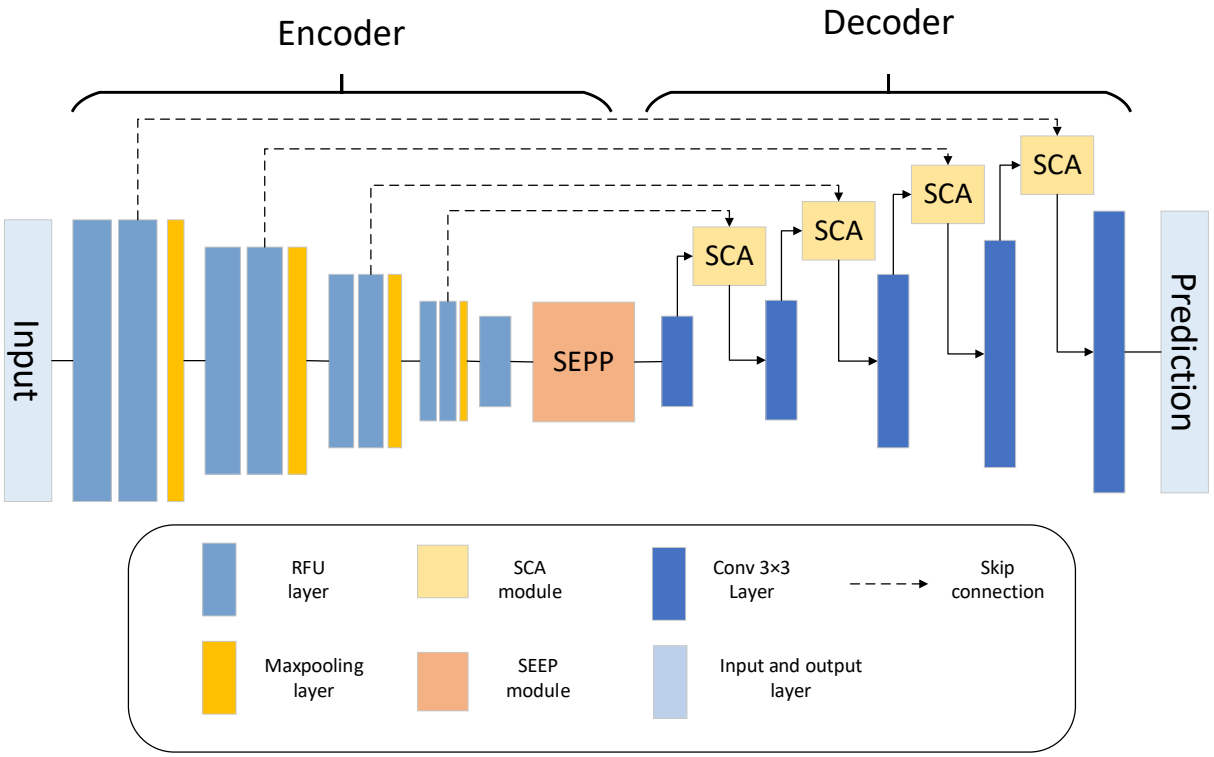


Figure 2. Illustration of the proposed SSCA-Net. We use multiple times of Relu Feature Unit (RFU) module as Feature Encoder, Then the feature maps are fed into a feature decoder module. It contains an self- and channel-attention (SCA) block and a squeeze and excitation pyramid pooling (SEPP) block. Moreover, we adopt skip connection to connect the low-level feature maps and high-level feature maps.

3. Methodology

We propose a new framework that provides multiple modules over which information from feature encoder module and decoder module, is assimilated using a generic building block, the SSCA-Net, as shown in Figure 2. We begin by describing the self- and channel-attention module in Sec. 3.1 followed by a detailed description of each SSCA-Net block in Sec. 3.2.

3.1 Self- and channel-attention module

As noted previously, we aim to exploit attention features for prediction with long-range residual connections. Hence, we propose an SCA module, as shown in Figure 3. In the classic image segmentation network model, multiple convolutional layers are used to preserve local neighborhood information of the image. However, the modeling of long-range dependence of images by convolutional neural networks is inefficient [36]. Therefore, we adapt the non-local model [37] to introduce self-attention to the image semantic segmentation framework.

In addition, the up-sampling portion of the image generation network typically uses a

deconvolution network. Besides, convolution kernel sizes and step sizes can cause deconvolution operations to generate checkerboard artifacts. To avoid the checkerboard effect, we use bilinear interpolation as an up-sampling operation.

The image features from the previous hidden layer $x \in R^{H\times W}$ are first transformed into two feature spaces $F(x);\ G(x)$ to calculate the attention:

$$F(x) = W_f x \tag{3}$$

$$G(x) = W_G x \tag{4}$$

$$\beta_{i,j} = \frac{\exp(s_{i,j})}{\sum_{i=1}^{N}\exp(s_{i,j})} \tag{5}$$

$$s_{i,j} = F(x_i)G(x_j) \tag{6}$$

where $\beta_{i,j}$indicates the extent to which the model attends to the i^{th}location when synthesizing the j^{th} region. Then the output of the attention layer is $\beta = (\beta_1, \beta_2, \beta_3 \dots \dots \beta_j \dots \dots \beta_N)$ where,

$$\beta_j = \sum_{i=1}^{N}\beta_{j.i}H(x_i), \tag{7}$$

$$H(x_i) = W_H x_i \tag{8}$$

In the above formulation, $W_f \in R^{H\times H}, W_G \in R^{H\times H}, W_H \in R^{H\times H}$are the learned weight matrices, which are implemented as 1×1 convolution. Finally, the features are again element-wise multiplication operation with the feature x_i. In short, the operation is computed as: $\mu_i = \beta_j x_i$.

In addition, the feature map should be aggregated in each channel. To this end, we take global average pooling on the attention layer feature map μ_j and produce a channel vector $X_c \in R^{c\times 1\times 1}$. Then, X_c and X_{low} perform the element-wise multiplication operation and produce a multiplication vector $X_m \in R^{C\times H\times w}$. This last obtained feature vector X_m is combined with a bilinearly interpolated feature vector X_{hig}. Therefore, the final output is given by

$$\mathrm{y_i} = (aver(\mu_i) \otimes X_{low}) + upsame(X_{hig}) \tag{9}$$

where $aver$ is the global average pooling and $upsame$ is the up-sampling operation. ⊗ denotes element-wise multiplication. X_{low} and X_{hig} represent low-level feature maps and high-level feature maps, respectively.

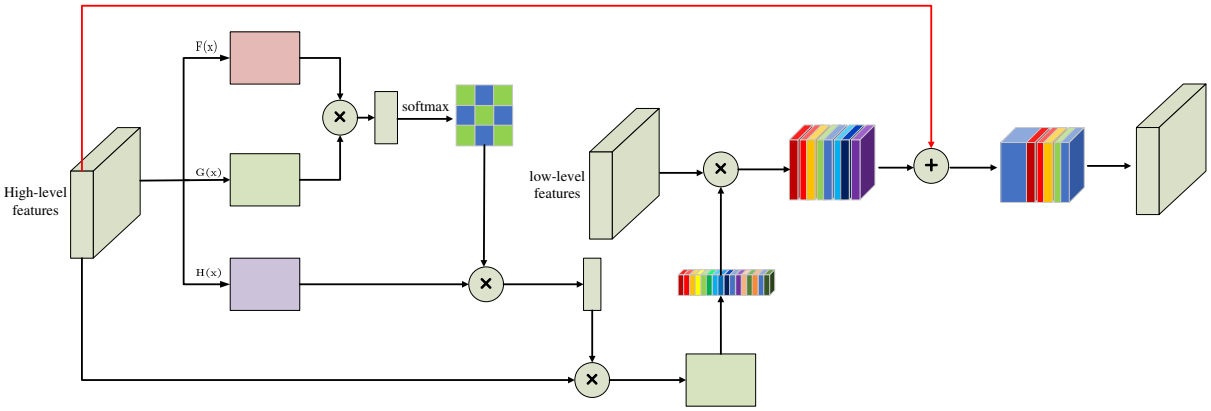


Figure 3. The designed Self- and channel-attention module (SCA) module for aggregating high-level features and low-level features. "⊗" denotes spatial element-wise multiplication and "⊕" denotes element-wise sum. The red lines represent the up-sampling operators.

3.2 SSCA-Net block

The architecture of SSCA-Net is illustrated in Figure 2. our architecture is generic and each SSCA-Net block can be easily modified to accept an arbitrary number of feature maps with arbitrary resolutions and depths.

Relu Feature Unit: The first part of each SSCA-Net block consists of the relu feature unit (RFU) that mainly for feature learning**.** We do not use ResNet in this task, since the medical image is not included in the category of the pre-trained model. And it can also prevent over-fitting and reduce both network parameters and training time. The RFU (see Figure 2) can reduce the training time and accelerate the network convergence.

Mathematically, the RFU block can be formulated as

$$RFU = Relu(BN(Conv_{3\times3}(x))) \quad (1)$$

where $Relu$ is an activation function and BN denotes the batch normalization. *Conv* 3×3 is the convolution operation with the kernel size of 3.

Therefore, the feature encoder module network structure can be expressed as follows:

$$\begin{aligned} layer_0 &= RFU(x) \\ layer_1 &= MaxPooling(layer_0) \\ layer_i &= RFU(layer_{i-1}) \end{aligned} \quad (2)$$

where i is the number of down-sampling, e.g., it takes "4" in the intracranial artery and the leg bone artery, and it takes "3" on the Retinal vessel set.

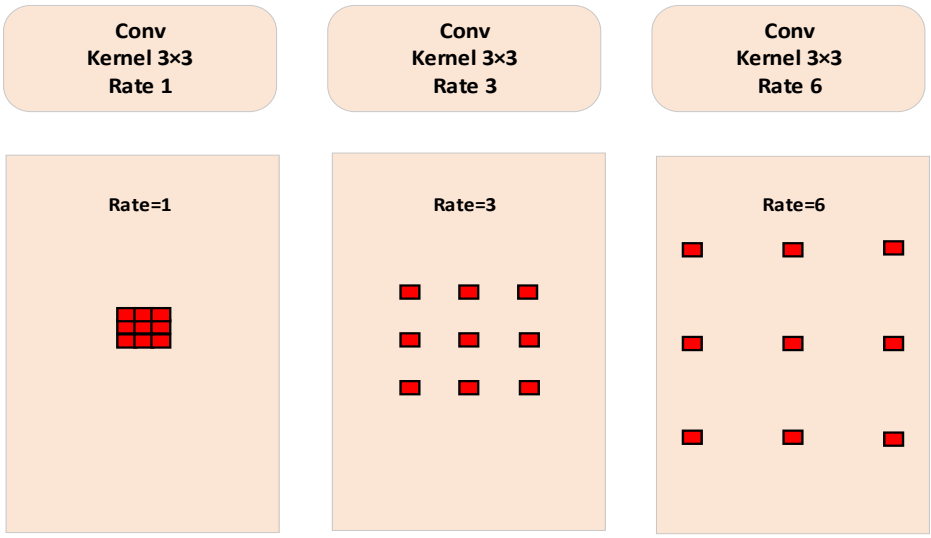


Figure 4. The illustrations of atrous convolution.

Squeeze-and-excitation pyramid pooling. In semantic segmentation, multiple convolution and pooling operations may lead to the reduction of the receptive field of the network and the loss of information features in different layers. To overcome this limitation, atrous convolution (see Figure 4) and spatial pyramid model (see Figure 5) are adopted for semantic segmentation, Due to pyramid pooling can counteract the shrunken receptive field by assembling multi-scale context [38].For example, Pyramid Scene Parsing Net (PSP) [23] performs spatial pooling at several grid scales and demonstrates outstanding performance on several semantic segmentation benchmarks. In the classic ASPP network [39], there are four parallel atrous convolutions with different atrous rates in the feature coding stage. Different from [31], we combine the SE operation into the residual block in ASPP to re-adjust the dynamic channel characteristics.

In this case, the SEPP module is also different from GC-Net [35]. Here the SEPP module has four cascaded branches with the gradual increment of the number of atrous convolution and SE network structure (see Figure 5). Since a large receptive field behaves well to acquire much contextual information, we present 4 dilated convolutions whose dilation scales are 1, 6, 12 and 12 in SEPP. In each branch, we apply 1×1 convolution for rectified linear activation after each atrous convolution and SE network.

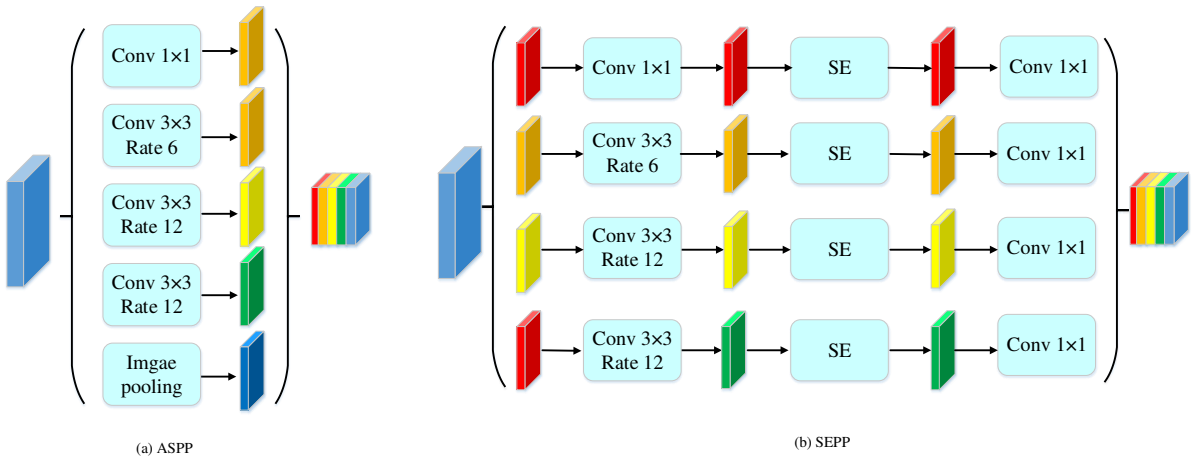


Figure 5. Illustration of (a) atrous spatial pyramid pooling (ASPP) and (b) Squeeze-and-excitation pyramid pooling (SEPP).

Mathematically, the SEPP block can be formulated as

$$SEPP = Cat\begin{bmatrix} Conv_{1\times1,}(SE\{Conv_{1\times1}(x)\}), \\ Conv_{1\times1,}\left(SE\left\{Conv_{3\times3,d_y}(x)\right\}\right) \end{bmatrix} y \in [6,12,12] \quad (10)$$

$$SE = Sigmod(ReLU(aver(x))) \quad (11)$$

where $Conv_{1\times1}$ denotes the 1×1 convolutions and $Conv_{3\times3,d_x}$ denotes the dilation convolutions with the kernel size of 3×3 and the dilation scale is *y*. $Cat(*)$ is a concatenating operation and *x* is the input feature. $Sigmod$ is the full connection with the sigmod activation function. $Relu$ is the full connection with the $Relu$ activation function.

Therefore, the feature decoder module network structure can be expressed as follows:

$$\begin{gathered} layer_0 = SCA(SEPP(conv_{i+1}), Conv_i) \\ layer_1 = BN(Conv_{1\times1}(layer_0)) \\ layer_x = SCA(layer_{x-1}, Conv_{i-x+1}) \end{gathered} \quad (12)$$

where x is the number of up-sampling, i is the number of down-sampling, which should be equal to x. We use BN to reduce internal covariate shift [40] and $Conv_{1\times1}$ to reduce feature dimensions and complexity of training. $SCA(*), SEPP(*)$ represent the SCA module and the SEPP module, respectively.

4. Results

4.1 Experimental settings

To show the effectiveness of our approach, we carry out comprehensive experiments on three datasets: intracranial blood vessel dataset, Digital Retinal Images for Vessel Extraction (DRIVE) [41] and leg arteries. The segmentation quality is measured by the dice coefficient[35] , mean intersection-over-union (IoU) score [35, 42], the sensitivity (Sen) and the accuracy (Acc) [43]. In addition, we also introduce the area under receiver operation characteristic curve (AUC) to measure segmentation performance on DRIVE. We apply simple data augmentation during training on the intracranial blood vessel dataset and leg arteries, including affine transformation, rotation, and vertical flip operations. we also performed data augmentation on DRIVE, including gray-scale conversion, standardization, contrast-limited adaptive histogram equalization, and gamma adjustment.

Intracranial blood vessel dataset: We first present our results on the intracranial blood vessel dataset in this work is courtesy of a local hospital in Shenzhen, China. The imaging modality of the dataset is computed tomography angiography (CTA). There are 4326 CTA images of intracranial blood vessels with the dimension of 512×512 in the original dataset. During the training, 20% of images are used as the validation set, while the remainder 80% as the training set. In addition, we also use two new patient data as the test data which are not included in the training and validation set.

DRIVE: The second application is the retinal vessel detection. The DRIVE dataset has been obtained from a diabetic retinopathy screening program in the Netherlands which contains 40 photographs. These are equally divided into 20 images for training and the other 20 images for testing. Due to the limited amount of data, we use sub-images for training. Each 128×128 patch is obtained by randomly selecting its center inside the full image.

Leg arteries: The next application is leg arteries segmentation task. The imaging modality of the leg blood vessel dataset is CTA from a local hospital in Shenzhen, China. There are 6545 CTA images of leg blood vessels with the dimension of 512×512 in the original dataset. During the training, 20% of images are used as the validation set, while the remainder 80% as the training set. In addition, we use two new patient data as the test data which are not included in

the training and validation set.

Training details: In the training stage, we use the ADAM [44] optimizer with the initial learning rate of 1e–3, β1 = 0.5, and β2 = 0.999 and the initial rate lr=1e–3. The initial learning rate is multiplied by $(1-\left(\frac{epoch-1}{totalepoch}\right)^{power})$ where the power is set to 0.9. The maximum number of epochs is 200. In this work, the loss function same as GC-Net [35]. The implementation is based on the public Keras [45] platform with TensorFlow [46] as backend. The training and testing bed are an Ubuntu 16.04 system.

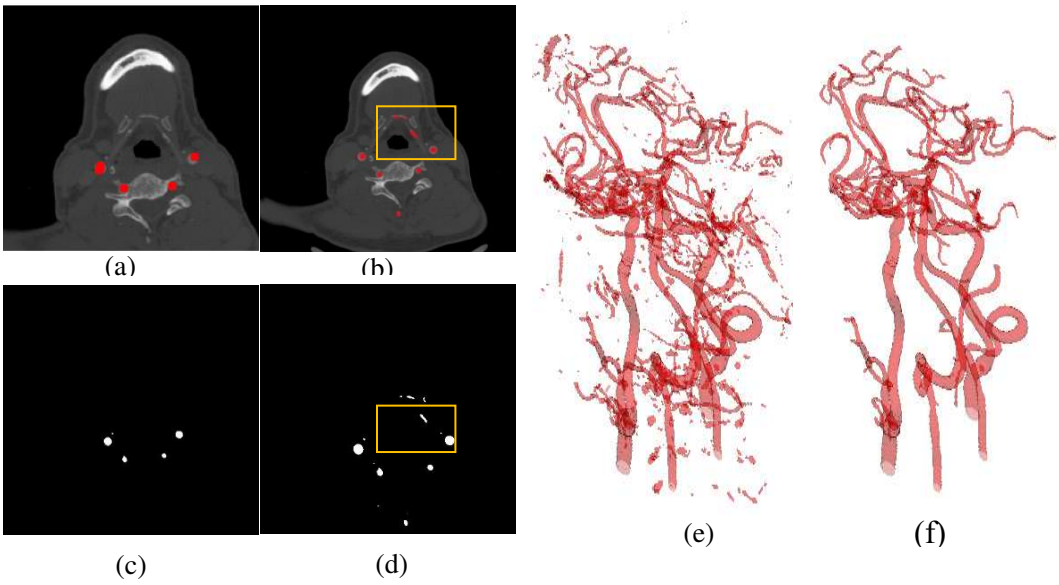


Figure 6. The effectiveness of post-processing. (a) and (c) show some non-vessel areas have been removed after post-processing, compared to the results of (b) and (d) before post-processing. (e) and (f) show the 3D results before and after post-processing, respectively.

4.2 Test on the intracranial blood vessel dataset

The 3D reconstruction of segmented vessels (consisting of 2D CT slices) can validate the segmentation quality by visually demonstrating their spatial information. It can be observed in Figure 6 and Figure 7 that there are some noises on the surface as isolated objects, arising from the misclassifications.

It is known that the entire intracranial arteries are interconnected. However, the mis-segmented noise is not connected to the entire blood vessel. As shown in Figure 6, there are some unconnected noises near each blood vessel. Therefore, we removed some areas or noises, accounting for less than 0.03% of the entire blood vessel.

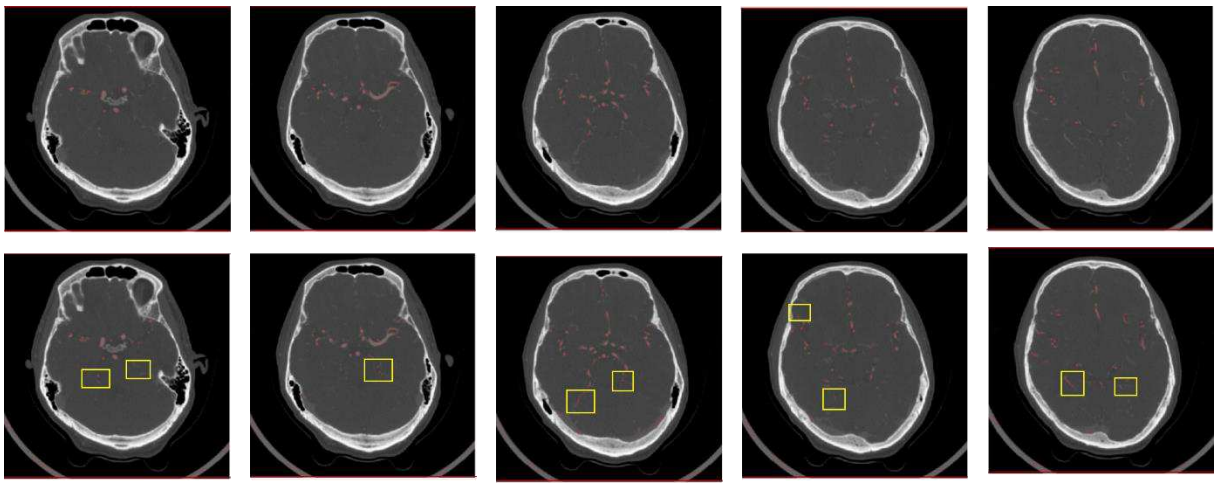
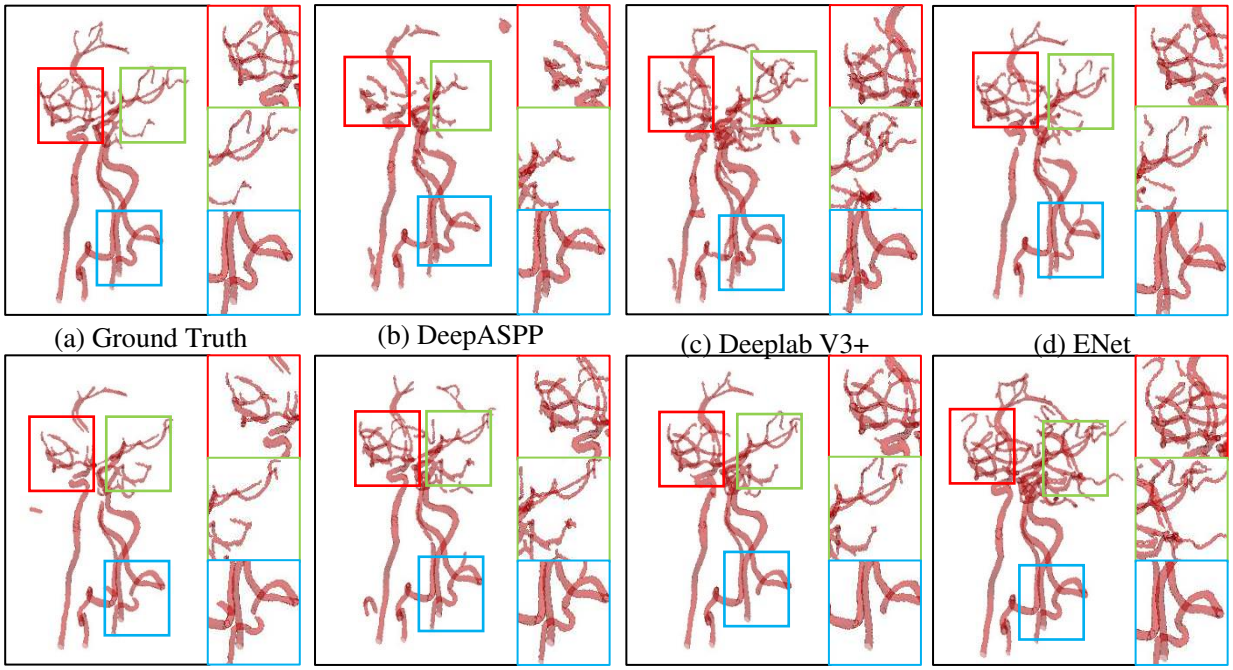


Figure 7. 2D visualization of segmentation results on the intracranial blood vessel data. The top row shows the ground truth; and the bottom row is our method

Post-processing is not performed to better explain the effect of SSCA-Net, as shown in Figure 8. As pointed out in the yellow circles, some segmented areas are either noise or real vessels. We can see some more small-scale structures produced by our SSCA-Net, in terms of ground truths. After post-processing, we can see more clearly. This is because the ground truth is manually labeled, and some of the vessels are too small, the marker does not notice. This result also demonstrates the SSCA-Net can effectively perform semantic segmentation.

Numerical results of our SSCA-Net and the state-of-the-art semantic segmentation solutions on intracranial blood vessel dataset are summarized in Table 1. These results are obtained under the same experimental conditions and the same data pretreatment. The Dic coefficient of segmentation accuracy increased from 76.14% to 96.21%, and the accuracy of Mean IoU increased from 66.53% to 92.70%. In particular, as we can see in Figure 1 and Figure 8, SSCA-Net produces more multi-scale structures than other methods. The reason is that the SCA module and the SEPP module can well preserve the information of medical images.



(a) Ground Truth (b) DeepASPP (c) Deeplab V3+ (d) ENet

(e) FCN8s (f) RefineNet (g) U-Net (h) Ours

Figure 8. Medical image segmentation results tested in the dataset of intracranial artery. From (a)-(h): (a) ground truth, and the segmentation results of (b) DeepASPP, (c) DeepLab V3+, (d) ENet, (e) FCN8, (f) RefineNet, (g) U-Net and (h) ours, respectively. Our SSCA-Net can perform segmentation of intracranial arteries effectively while preserving more vessel tinny-scale structures.

Table 1. Comparison with the state-of-the-art methods on the intracranial blood vessel training dataset

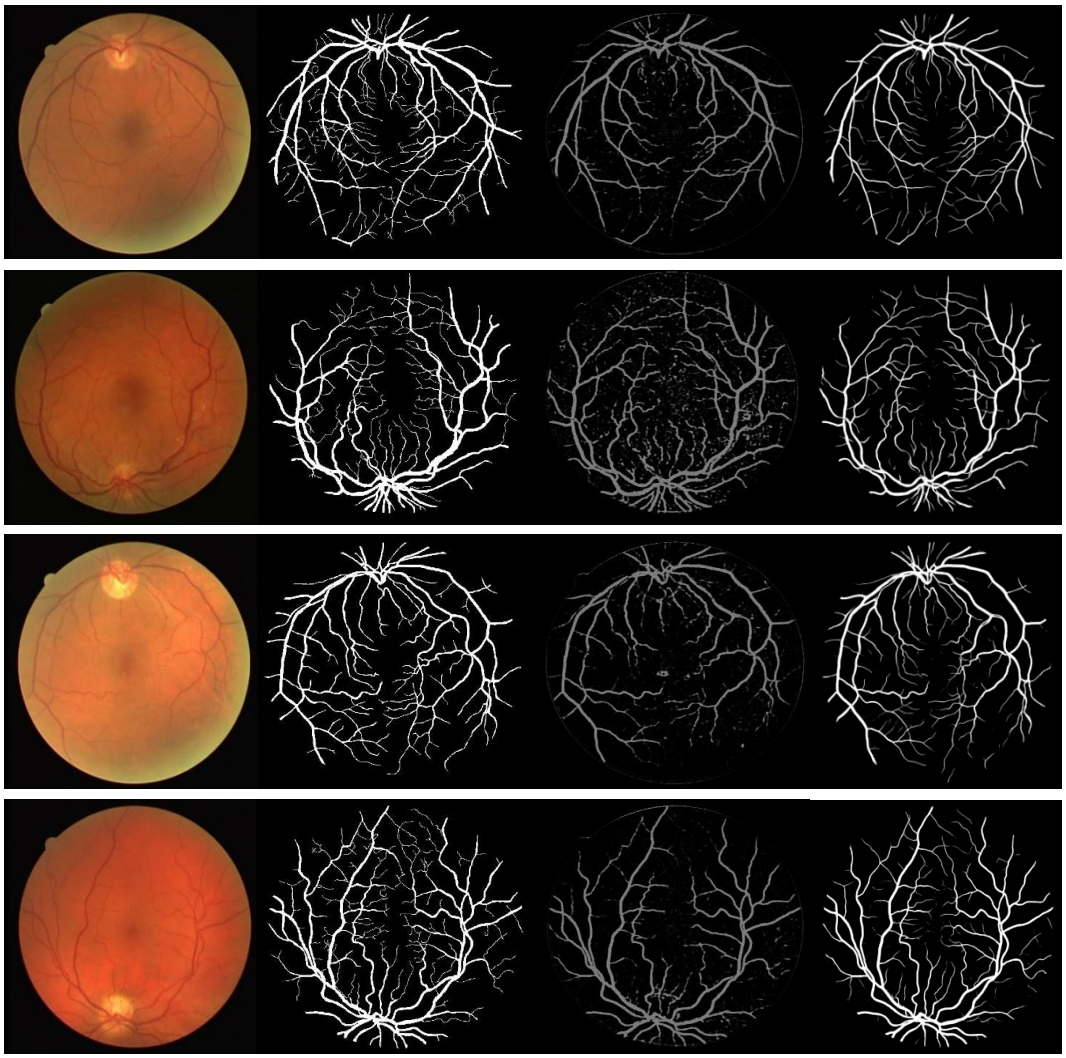
Method	Dic (%)	Mean IoU (%)
U-Net [6]	87.32	86.48
SegNet [30]	88.40	81.63
FCN8s [3]	84.23	67.72
FCN16s [3]	76.14	66.53
DenseASPP [25]	84.38	81.80
Deeplab V3+ [19]	90.70	87.83
RefineNet [17]	91.68	76.72
ENet [47]	85.97	81.72
GC-Net [35]	96.35	91.89
SSCA-Net	**96.21**	**92.70**

4.3 Test on retinal vessel segmentation

We have compared the proposed SSCA-Net with CNN based algorithms [6, 35, 48, 49] and some classical methods [50-54]. Table 2 shows the comparison of our method to those methods. From the comparison, the SSCA-Net achieved 98.20 %, 83.52% and 96.14% in AUC, Sen and Acc respectively, which are better than the other methods. Comparing with the CE-Net, the AUC increases from 97.79% to 98.20 % and that the sensitivity score increases from 83.09% to 83.52% while the accuracy increases from 95.45% to 96.14%, which shows that the SSCA-Net are beneficial for retina vessel detection. We show some examples for visual comparisons in Figure 9.

Table 2. Performance comparison of the competing methods on retina vessel data using different performance metrics

Method	Sen (%)	Acc (%)	AUC (%)
Azzopadi et al. [50]	76.55	94.42	96.14
Roychowdhury et al. [51]	72.50	95.20	96.72
Zhao et al. [52]	74.20	95.40	86.20
U-Net [6]	73.44	95.23	97.44
DeepVessel [48]	76.03	95.23	97.52
HED [49]	73.64	94.34	97.23
Li et al. [53]	75.69	95.27	97.38
Melinscak et al. [54]	-	94.66	97.49
CE-Net[43]	83.09	95.45	97.79
GC-Net [35]	78.44	95.51	97.77
SSCA-Net	**83.52**	**96.14**	**98.20**



(a) (b) (c) (d)

Figure 9. Visualization results on the DRIVE dataset. From (a)-(d): (a) test image, (b) ground truth, and results of (c) U-Net and (d) SSCA-Net, respectively.

4.4 Test on leg arteries

We have compared our SSCA-Net with the state-of-the-art algorithms as shown in Table 3, our proposed method achieves the performance, which the Dic score is 97.21% and the Mean IoU score is 94.42%. Comparing with the FCN16, the Dic score increases from 83.41% to 97.21% by 16.5%, the Mean IoU score increases from 65.12% to 94.42%, which shows that the skip connected is beneficial for semantic segmentation. Besides, comparing with U-Net the Dic score increases from 91.25% to 97.21%, the Mean IoU score increases from 76.26% to 94.42%, which shows that the proposed SCA and SEPP blocks are beneficial for leg vessel segmentation as well. We also compared some of the existing excellent methods, and the results show that SSCA-Net can perform blood vessel segmentation more effectively. We show some examples for visual comparisons in Figure 11.

Table 3 Comparisons with state-of-the-arts on leg blood vessel training dataset

method	Dic (%)	Mean IoU (%)
U-Net [6]	91.25	76.26
FCN8 [3]	88.52	80.11
FCN16 [3]	83.41	65.12
ENet [47]	90.01	86.31
Refinenet [17]	95.21	91.85
DeepASPP[25]	88.36	87.02
DeepLab V3+[19]	92.05	90.57
SSCA-Net	**97.21**	**94.42**

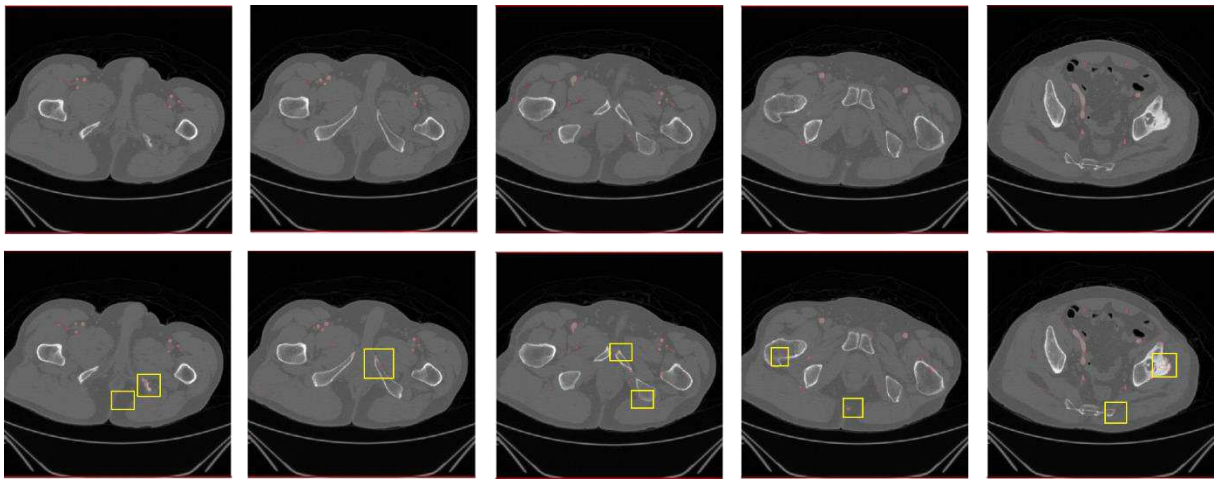


Figure. 10. 2D visualization of segmentation results on the leg artery data. The top row shows the ground truth; and the bottom row is our method.

.

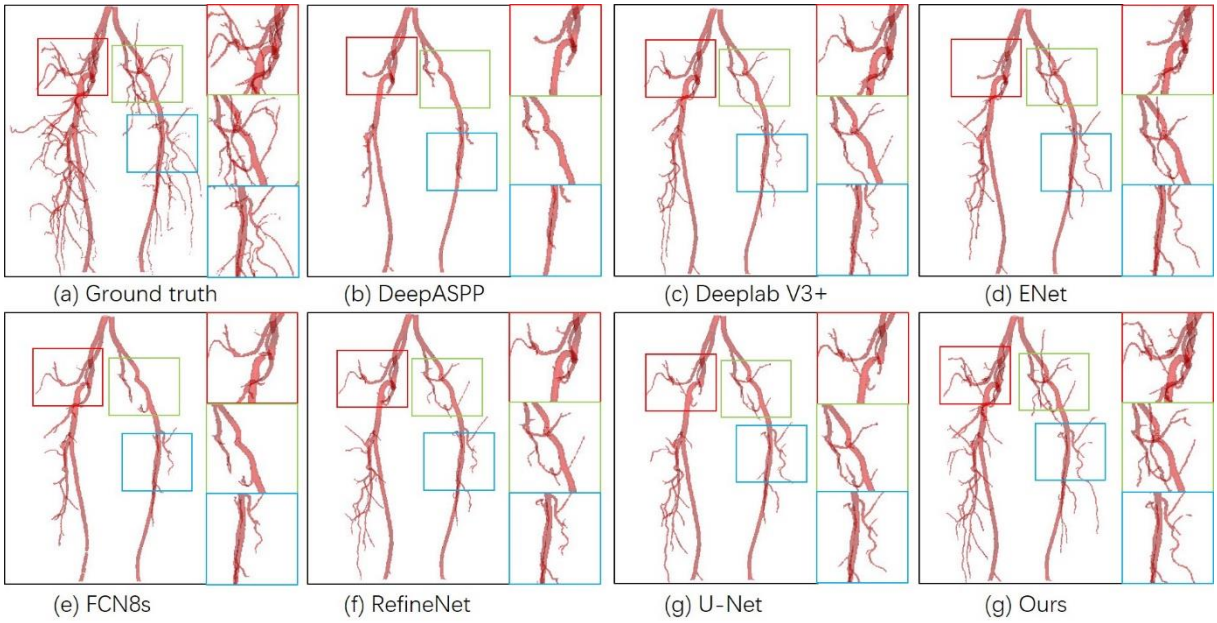


Figure. 11. Comparative visualization of 3D results achieved on the test dataset1 and test datset2. Compared to the ground truth, all the four state-of-the-art methods (DeepASPP, DeepLap V3+, ENet, FCN8s, RefineNet and U-Net) miss fine features (e.g. small vessels in rectangle), whereas the proposed method preserves fine vessels well.

5. Discussion and ablation study

To verify the efficacy of different modules in our method, we conduct the ablation study. We also give several design choices and show their influences on the results.

Backbone: The modified U-Net without the pre-trained ResNet50 and with the SCA block.

Backbone + ASPP: The network with the SCA block but without the SEPP block and replaces it with ASPP

ResNet50 + SCA: The network with the pre-trained ResNet50 and SCA.

ResNet50 + SCA+SEPP: The network with the pre-trained ResNet50, SCA and SEPP.

ResNet50 + SCA+ASPP: The network with the pre-trained ResNet50, SCA and ASPP.

Analysis on pre-trained weight:

Recent work [55] points out that ImageNet pre-training is no better than the original feature encoder in terms of model training accuracy. We do ablation learning on two datasets because the two data sets contain a large amount of data, which can better verify the potential of the network. On the intracranial arterial blood vessel dataset and the leg arterial blood vessel dataset, we can see that **ResNet50 + SCA+SEPP** has increased from 95.79% and 96.78% to 96.21% and 97.21% in Dic and Mean IoU increased from 91.70% and 93.75% to 92.70% and 94.42%.

The results in Figure 12, Figure 13, Table 3, and Table 4 have demonstrated the effectiveness of without pre-training weights which is not worse than using weights.

Analysis on SEPP and SCA:

SEPP: In Tab.4 and Tab.5, we experiment with the effect of incorporating SE to the improved ASPP module. Considering the characteristics of the network, we designed two experiments. One is the use of pre-training weights, and the other is the use of pre-training weights. The results can be seen on the two arterial blood vessel data. It demonstrates that the receptive field plays a significant role in semantic segmentation. We can get in Tables 4 and 5 that the network structure using the SEPP module has improved in both the Dic and Mean IoU evaluation standards compared to the use of ASPP and networks that do not use similar structures. This is because medical images contain very little information compared to natural images, and it is easy to cause information loss when using a large number of convolution and pooling operations.

SCA: Similarly, we apply experiments to verify the effectiveness of the SCA module. In this paper, if the pre-training weights and SCA modules are not used, this network can be regarded as a U-Net network. In Table 1 and Table 3, we can see that the SSCA-Net network structure is better than U-Net.

Comparison with GC-Net:

Both our best Segmentation model (in Table 1) and SEPP model. We can see that SSCA-Net has a slight decrease compared to GC-Net in the Dic evaluation standard, but it has an improvement in Mean IoU value. These subtle differences are in the range of tolerance, and the reason for this is mainly comes from the fine-tuning batch normalization parameters [40].

Ablation study:

Finally, it has been proved that the proposed algorithm is accurate and robust in medical segmentation from various CT images (see Figure 6 to Figure 11). The average Dic value and Mean IoU of the proposed method on intracranial blood vessel data were 96.21% and 92.70% respectively, which are shown in Table 1. On the leg bone artery dataset, the average Dic value and Mean IoU were 97.21% and 94.42% respectively, which are shown in Table 3. On the retinal vessel dataset, Sen, ACC and AUC obtained 85.32%, 96.14% and 98.20% respectively on the three evaluation criteria, and the results were better than other methods, which are shown in Table 2.

Table 4. Performance comparisons of context aggregation approaches on leg blood vessel data

Method	Dic (%)	Mean IoU (%)
ResNet50+SCA+SEPP	96.78	93.75
ResNet50+SCA+ASPP	96.51	93.53
ResNet50+SCA	96.67	93.37
Backbone+ASPP	96.92	94.19
Backbone	96.98	94.07
SSCA-net	97.21	94.42

Table 5. Performance comparisons of context aggregation approaches on intracranial blood vessel data

Method	Dic (%)	Mean IoU (%)
ResNet50+SCA+SEPP	95.79	91.70
ResNet50+SCA+ASPP	95.72	90.69
ResNet50+SCA	95.38	91.08
Backbone	95.90	91.90
Backbone+ASPP	95.99	92.47
SSCA-Net	96.21	92.70

Limitations: We have introduced two new modules to deal with the problem of medical image segmentation from CT images. To some extent, the SSCA-Net network structure can better improve the segmentation accuracy of CT images. But compared to the classic U-net network, it requires more parameters and takes a little more time to train the network. In different experiments, we have observed that the richer the information is, the more beneficial the segmentation of the image is, but in this work, the experiment is performed on 2D images. In order to get more segmentation image information, 3D data may be used in future.

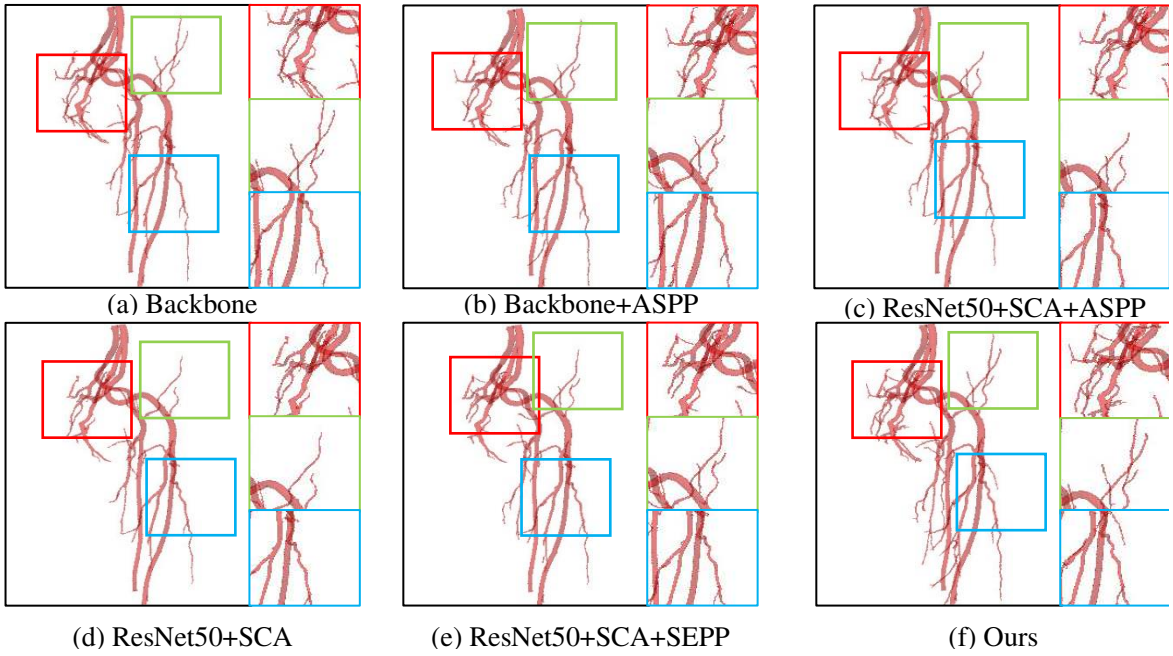


Figure. 12. Medical image segmentation results tested in the dataset of leg artery. From (a)-(f): (a) Backbone,(b)Backbone+ASPP,(c)ResNet50+SCA+ASPP,(d)ResNet50+SCA,(e)ResNet50+SCA+SEPP, (f) SSCA-Net respectively. Our SSCA-Net can perform segmentation of intracranial arteries effectively while preserving more vessel details.

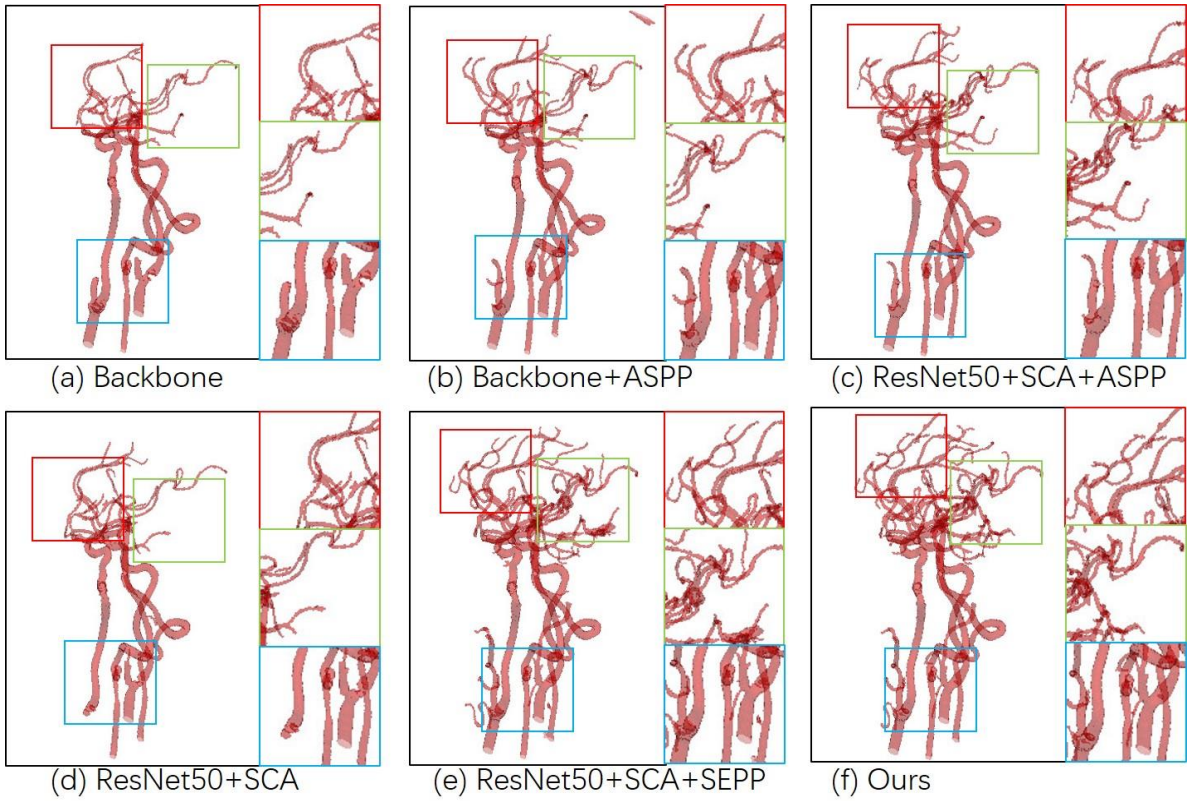


Figure. 13. Medical image segmentation results tested in the dataset of intracranial artery. From (a)-(f): (a) Backbone,(b)Backbone+ASPP, (c) ResNet50+SCA+ASPP, (d) ResNet50+SCA, (e) ResNet50+SCA+SEPP, (f) SSCA-Net. respectively. Our SSCA-Net can segment intracranial arteries effectively while preserving more vessel details.

6. Conclusion

This paper presents a novel network, called SSCA-Net, for multi-scale structure-preserving vessel segmentation. SSCA-Net mainly uses two attention mechanisms to analyze the context information of the entire network. To obtain global contextual information, we introduce the SCA attention module which applies two attention modes to obtain the feature information of the image, the SEPP module is devised to increase the size of the receptive field of the network while learning more features, and design a weighted cross-entropy loss function to make the training process be more effective. These operations are beneficial for improving the accuracy of vessel segmentation with multi-scale structures. Furthermore, we also experimented with the feature encoder module instead of the ResNet50 pre-training model. This greatly reduces the training time and also reduces the problem of network overfitting. Our method can be applied to different tasks by fine-tuning our model using the new training data, and test on three benchmark datasets, and is compared with various state-of-the art methods with respect to the Dic, Mean IoU and AUC metrics.

Ethics approval and consent to participate

Not applicable

Competing interests

The authors declare that they have no competing interests.

Availability of data and material

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request

Consent for publication

Not applicable

Funding

National Natural Science Foundation of China (No. 61672510)

Shenzhen Basic Research Program (No. JCYJ20180507182441903).

Author Contributions

JHW conceived and designed the study, JJN preformed the experiment and analyzed the data. made discussions and composed the manuscript together with JT, QMW and MZC. All authors read and approved the manuscript

Acknowledgements

Not applicable

References

1. Carmeliet, Peter and Rakesh K Jain, *Angiogenesis in cancer and other diseases.* nature, 2000. **407**(6801): p. 249.
2. Campochiaro, Peter A, *Molecular pathogenesis of retinal and choroidal vascular diseases.* Progress in retinal and eye research, 2015. **49**: p. 67-81.
3. Long, Jonathan, Evan Shelhamer, and Trevor Darrell. *Fully convolutional networks for semantic segmentation*. in *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2015.
4. Korez, Robert, Boštjan Likar, Franjo Pernuš, and Tomaž Vrtovec. *Model-based segmentation of vertebral bodies from MR images with 3D CNNs*. in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. 2016. Springer.
5. Moeskops, Pim, Jelmer M Wolterink, Bas HM van der Velden, Kenneth GA Gilhuijs, Tim Leiner, Max A Viergever, and Ivana Išgum. *Deep learning for multi-task medical image segmentation in multiple modalities*. in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. 2016. Springer.
6. Ronneberger, Olaf, Philipp Fischer, and Thomas Brox. *U-net: Convolutional networks for biomedical image segmentation*. in *International Conference on Medical image computing and computer-assisted intervention*. 2015. Springer.
7. Çiçek, Özgün, Ahmed Abdulkadir, Soeren S Lienkamp, Thomas Brox, and Olaf Ronneberger. *3D U-Net: learning dense volumetric segmentation from sparse annotation*. in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. 2016. Springer.
8. Milletari, Fausto, Nassir Navab, and Seyed-Ahmad Ahmadi. *V-net: Fully convolutional neural networks for volumetric medical image segmentation*. in *3D Vision (3DV), 2016 Fourth International Conference on*. 2016. IEEE.
9. Liu, Wei, Andrew Rabinovich, and Alexander C Berg, *Parsenet: Looking wider to see better.* arXiv preprint arXiv:1506.04579, 2015.
10. Li, Hanchao, Pengfei Xiong, Jie An, and Lingxue Wang, *Pyramid attention network for semantic segmentation.* arXiv preprint arXiv:1805.10180, 2018.
11. Yuan, Yuhui and Jingdong Wang, *Ocnet: Object context network for scene parsing.* arXiv preprint arXiv:1809.00916, 2018.
12. Oktay, Ozan, Jo Schlemper, Loic Le Folgoc, Matthew Lee, Mattias Heinrich, Kazunari Misawa, Kensaku Mori, Steven McDonagh, Nils Y Hammerla, and Bernhard Kainz, *Attention u-net: Learning where to look for the pancreas.* arXiv preprint arXiv:1804.03999, 2018.

13. Shakeri, Mahsa, Stavros Tsogkas, Enzo Ferrante, Sarah Lippe, Samuel Kadoury, Nikos Paragios, and Iasonas Kokkinos. *Sub-cortical brain structure segmentation using F-CNN's*. in *2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI)*. 2016. IEEE.
14. Alansary, Amir, Konstantinos Kamnitsas, Alice Davidson, Rostislav Khlebnikov, Martin Rajchl, Christina Malamateniou, Mary Rutherford, Joseph V Hajnal, Ben Glocker, and Daniel Rueckert. *Fast fully automatic segmentation of the human placenta from motion corrupted MRI*. in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. 2016. Springer.
15. Christ, Patrick Ferdinand, Mohamed Ezzeldin A Elshaer, Florian Ettlinger, Sunil Tatavarty, Marc Bickel, Patrick Bilic, Markus Rempfler, Marco Armbruster, Felix Hofmann, and Melvin D'Anastasi. *Automatic liver and lesion segmentation in CT using cascaded fully convolutional neural networks and 3D conditional random fields*. in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. 2016. Springer.
16. Norman, Berk, Valentina Pedoia, and Sharmila Majumdar, *Use of 2D U-Net convolutional neural networks for automated cartilage and meniscus segmentation of knee MR imaging data to determine relaxometry and morphometry.* Radiology, 2018. **288**(1): p. 177-185.
17. Lin, Guosheng, Anton Milan, Chunhua Shen, and Ian Reid. *Refinenet: Multi-path refinement networks for high-resolution semantic segmentation*. in *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017.
18. Chen, Liang-Chieh, George Papandreou, Iasonas Kokkinos, Kevin Murphy, and Alan L Yuille, *Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs.* IEEE transactions on pattern analysis and machine intelligence, 2017. **40**(4): p. 834-848.
19. Chen, Liang-Chieh, Yukun Zhu, George Papandreou, Florian Schroff, and Hartwig Adam, *Encoder-decoder with atrous separable convolution for semantic image segmentation.* arXiv preprint arXiv:1802.02611, 2018.
20. Chen, Liang-Chieh, George Papandreou, Iasonas Kokkinos, Kevin Murphy, and Alan L Yuille, *Semantic image segmentation with deep convolutional nets and fully connected crfs.* Computing, 2014. **1**(1): p. 12-17.
21. Chen, Liang-Chieh, George Papandreou, Florian Schroff, and Hartwig Adam, *Rethinking atrous convolution for semantic image segmentation.* arXiv preprint arXiv:1706.05587, 2017.
22. Chollet, François. *Xception: Deep learning with depthwise separable convolutions*. in *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017.
23. Zhao, Hengshuang, Jianping Shi, Xiaojuan Qi, Xiaogang Wang, and Jiaya Jia. *Pyramid scene parsing network*. in *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*. 2017.
24. Yu, Changqian, Jingbo Wang, Changxin Gao, Gang Yu, Chunhua Shen, and Nong Sang. *Context Prior for Scene Segmentation*. in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2020.
25. Yang, Maoke, Kun Yu, Chi Zhang, Zhiwei Li, and Kuiyuan Yang. *Denseaspp for semantic segmentation in street scenes*. in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2018.
26. Giusti, Alessandro, Dan C Cireşan, Jonathan Masci, Luca M Gambardella, and Jürgen Schmidhuber. *Fast image scanning with deep max-pooling convolutional neural networks*. in *2013 IEEE International Conference on Image Processing*. 2013. IEEE.
27. Papandreou, George, Iasonas Kokkinos, and Pierre-André Savalle. *Modeling local and global*

deformations in deep learning: Epitomic convolution, multiple instance learning, and sliding window detection. in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2015.

28. Zhao, Hengshuang, Yi Zhang, Shu Liu, Jianping Shi, Chen Change Loy, Dahua Lin, and Jiaya Jia. *Psanet: Point-wise spatial attention network for scene parsing*. in *Proceedings of the European Conference on Computer Vision (ECCV)*. 2018.
29. Hu, Jie, Li Shen, and Gang Sun. *Squeeze-and-excitation networks*. in *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2018.
30. Woo, Sanghyun, Jongchan Park, Joon-Young Lee, and In So Kweon. *Cbam: Convolutional block attention module*. in *Proceedings of the European Conference on Computer Vision (ECCV)*. 2018.
31. Huang, Zilong, Xinggang Wang, Lichao Huang, Chang Huang, Yunchao Wei, and Wenyu Liu, *Ccnet: Criss-cross attention for semantic segmentation.* arXiv preprint arXiv:1811.11721, 2018.
32. Zhong, Zilong, Zhong Qiu Lin, Rene Bidart, Xiaodan Hu, Ibrahim Ben Daya, Zhifeng Li, Wei-Shi Zheng, Jonathan Li, and Alexander Wong. *Squeeze-and-Attention Networks for Semantic Segmentation*. in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2020.
33. Niu, Ruigang, *Hmanet: Hybrid multiple attention network for semantic segmentation in aerial images.* arXiv preprint arXiv:2001.02870, 2020.
34. Park, Jongchan, Sanghyun Woo, Joon-Young Lee, and In So Kweon, *Bam: Bottleneck attention module.* arXiv preprint arXiv:1807.06514, 2018.
35. Ni, Jiajia, Jianhuang Wu, Jing Tong, Zhengming Chen, and Junping Zhao, *GC-Net: Global context network for medical image segmentation.* Computer Methods and Programs in Biomedicine, 2019: p. 105121.
36. Wu, Zifeng, Chunhua Shen, and Anton van den Hengel, *Bridging category-level and instance-level semantic image segmentation.* arXiv preprint arXiv:1605.06885, 2016.
37. Wang, Xiaolong, Ross Girshick, Abhinav Gupta, and Kaiming He. *Non-local neural networks*. in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2018.
38. Tao, Andrew, Karan Sapra, and Bryan Catanzaro, *Hierarchical Multi-Scale Attention for Semantic Segmentation*. arXiv preprint arXiv:2005.10821, 2020.
39. Soares, João VB, Jorge JG Leandro, Roberto M Cesar, Herbert F Jelinek, and Michael J Cree, *Retinal vessel segmentation using the 2-D Gabor wavelet and supervised classification.* IEEE Transactions on medical Imaging, 2006. **25**(9): p. 1214-1222.
40. Ioffe, Sergey and Christian Szegedy, *Batch normalization: Accelerating deep network training by reducing internal covariate shift*. arXiv preprint arXiv:1502.03167, 2015.
41. Staal, Joes, Michael D Abràmoff, Meindert Niemeijer, Max A Viergever, and Bram Van Ginneken, *Ridge-based vessel segmentation in color images of the retina.* IEEE transactions on medical imaging, 2004. **23**(4): p. 501-509.
42. Everingham, Mark, Luc Van Gool, Christopher KI Williams, John Winn, and Andrew Zisserman, *The pascal visual object classes (voc) challenge.* International journal of computer vision, 2010. **88**(2): p. 303-338.
43. Gu, Zaiwang, Jun Cheng, Huazhu Fu, Kang Zhou, Huaying Hao, Yitian Zhao, Tianyang Zhang, Shenghua Gao, and Jiang Liu, *CE-Net: Context Encoder Network for 2D Medical Image Segmentation*. IEEE transactions on medical imaging, 2019.
44. Kingma, Diederik P and Jimmy Ba, *Adam: A method for stochastic optimization.* arXiv preprint

arXiv:1412.6980, 2014.

45. Chollet, François, *Keras*. 2015.
46. Abadi, Martın, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S Corrado, Andy Davis, Jeffrey Dean, and Matthieu Devin, *TensorFlow: Large-scale machine learning on heterogeneous systems, 2015.* Software available from tensorflow. org, 2015. **1**(2).
47. Paszke, Adam, Abhishek Chaurasia, Sangpil Kim, and Eugenio Culurciello, *Enet: A deep neural network architecture for real-time semantic segmentation.* arXiv preprint arXiv:1606.02147, 2016.
48. Fu, Huazhu, Yanwu Xu, Stephen Lin, Damon Wing Kee Wong, and Jiang Liu. *Deepvessel: Retinal vessel segmentation via deep learning and conditional random field*. in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. 2016. Springer.
49. Xie, Saining and Zhuowen Tu. *Holistically-nested edge detection*. in *Proceedings of the IEEE international conference on computer vision*. 2015.
50. Azzopardi, George, Nicola Strisciuglio, Mario Vento, and Nicolai Petkov, *Trainable COSFIRE filters for vessel delineation with application to retinal images*. Medical image analysis, 2015. **19**(1): p. 46-57.
51. Roychowdhury, Sohini, Dara D Koozekanani, and Keshab K Parhi, *Blood vessel segmentation of fundus images by major vessel extraction and subimage classification.* IEEE journal of biomedical and health informatics, 2015. **19**(3): p. 1118-1128.
52. Zhao, Yitian, Lavdie Rada, Ke Chen, Simon P Harding, and Yalin Zheng, *Automated vessel segmentation using infinite perimeter active contour model with hybrid region information with application to retinal images*. IEEE transactions on medical imaging, 2015. **34**(9): p. 1797-1807.
53. Li, Qiaoliang, Bowei Feng, LinPei Xie, Ping Liang, Huisheng Zhang, and Tianfu Wang, *A cross-modality learning approach for vessel segmentation in retinal images.* IEEE transactions on medical imaging, 2015. **35**(1): p. 109-118.
54. Melinščak, Martina, Pavle Prentašić, and Sven Lončarić. *Retinal vessel segmentation using deep neural networks*. in *10th International Conference on Computer Vision Theory and Applications (VISAPP 2015)*. 2015.
55. He, Kaiming, Ross Girshick, and Piotr Dollár, *Rethinking imagenet pre-training.* arXiv preprint arXiv:1811.08883, 2018.

Figures

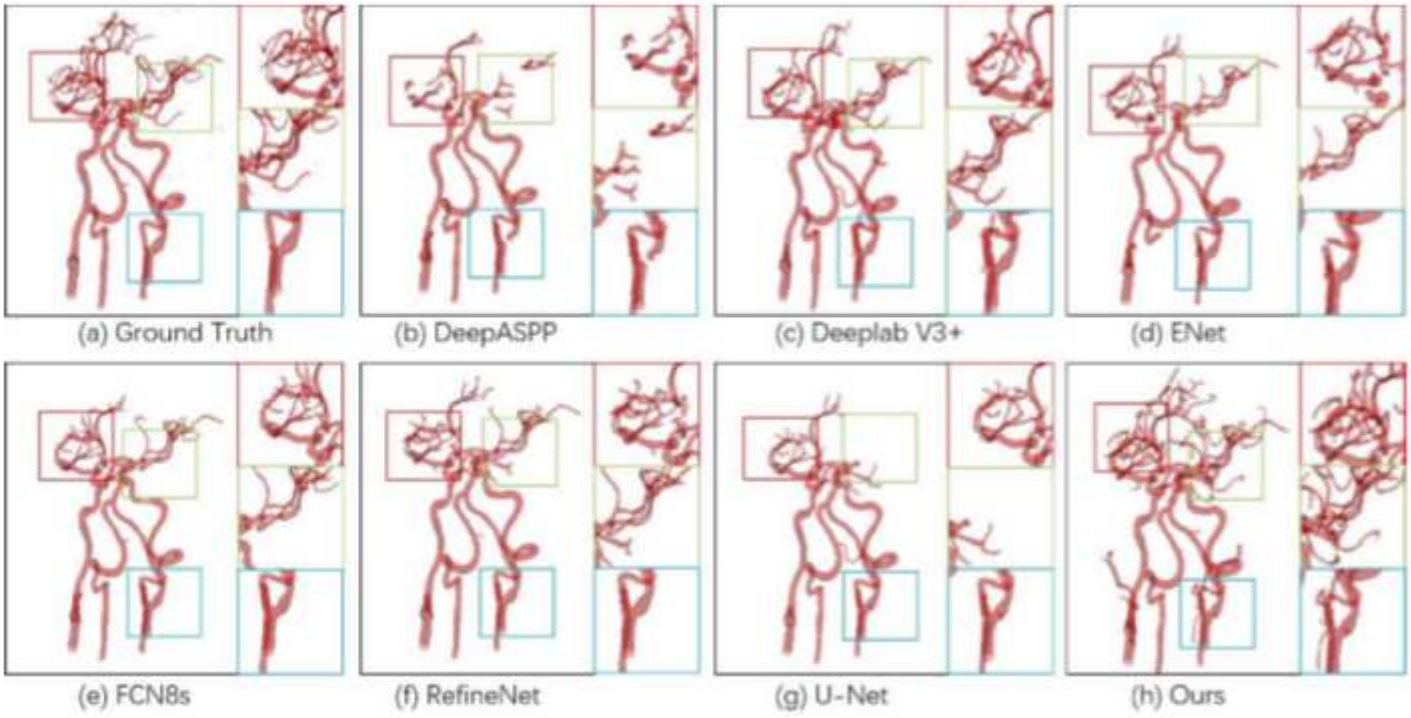


Figure 1

Medical image segmentation results tested in the dataset of intracranial artery. From (a)-(h): (a) ground truth, and the segmentation results of (b) DeepASPP, (c) DeepLab V3+, (d) ENet, (e) FCN8s, (f) RefineNet, (g) U-Net, and (h) ours, respectively. Our SSCA-Net can perform the segmentation of intracranial arteries effectively while preserving multi-scale structures of vessels, especially the tinny-scale structures.

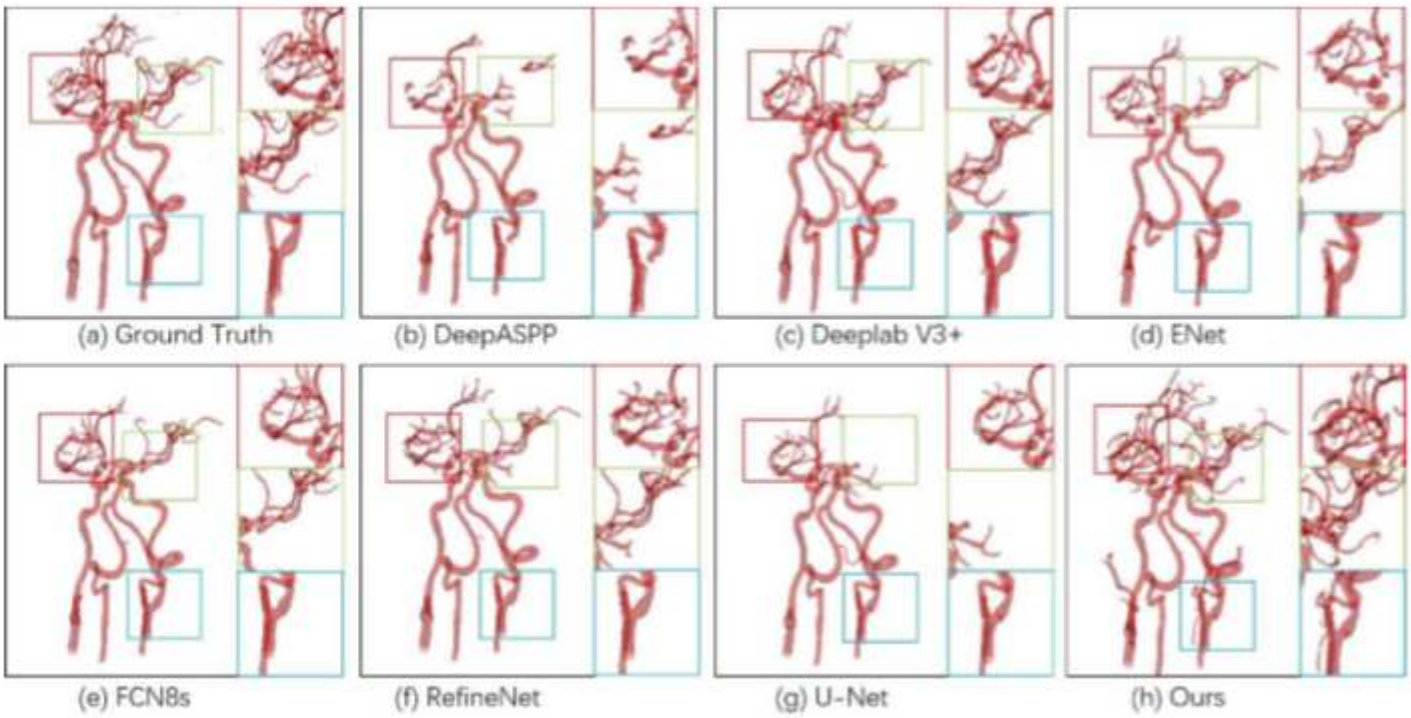


Figure 1

Medical image segmentation results tested in the dataset of intracranial artery. From (a)-(h): (a) ground truth, and the segmentation results of (b) DeepASPP, (c) DeepLab V3+, (d) ENet, (e) FCN8s, (f) RefineNet, (g) U-Net, and (h) ours, respectively. Our SSCA-Net can perform the segmentation of intracranial arteries effectively while preserving multi-scale structures of vessels, especially the tinny-scale structures.

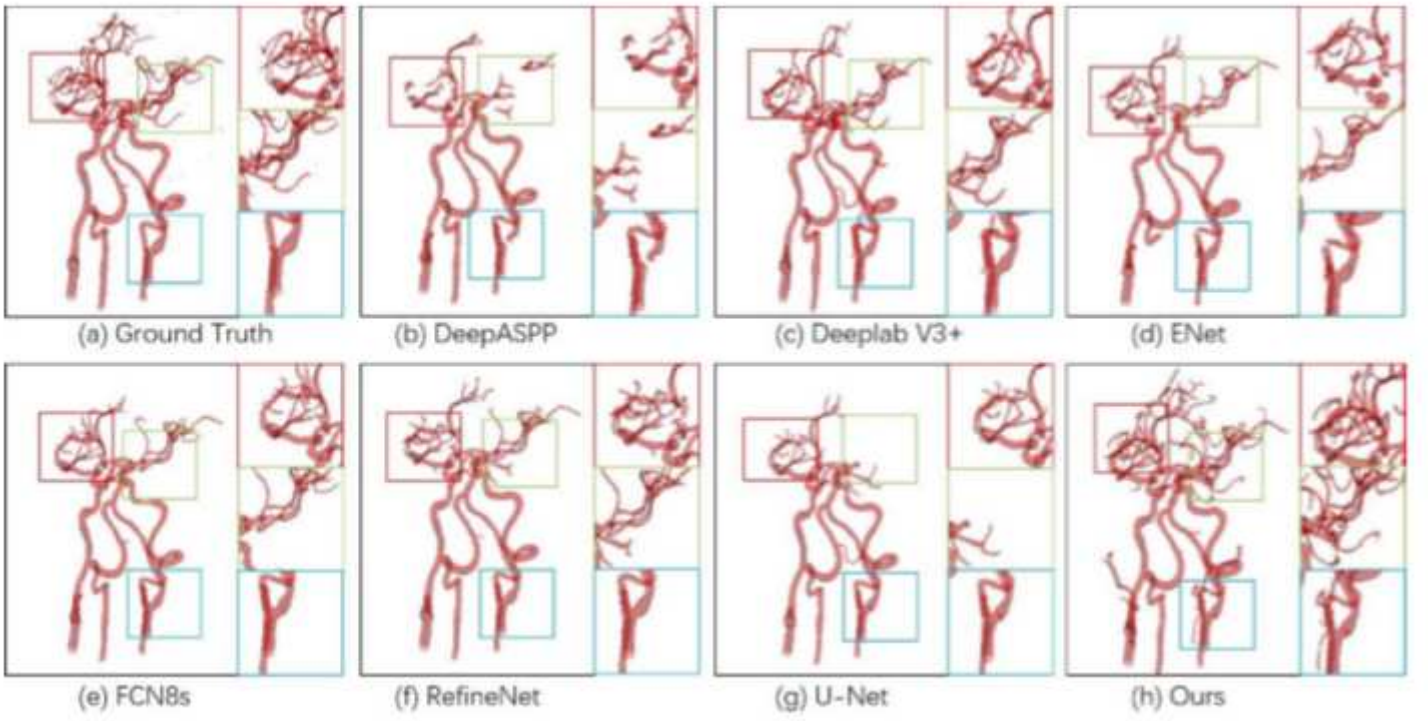


Figure 1

Medical image segmentation results tested in the dataset of intracranial artery. From (a)-(h): (a) ground truth, and the segmentation results of (b) DeepASPP, (c) DeepLab V3+, (d) ENet, (e) FCN8s, (f) RefineNet, (g) U-Net, and (h) ours, respectively. Our SSCA-Net can perform the segmentation of intracranial arteries effectively while preserving multi-scale structures of vessels, especially the tinny-scale structures.

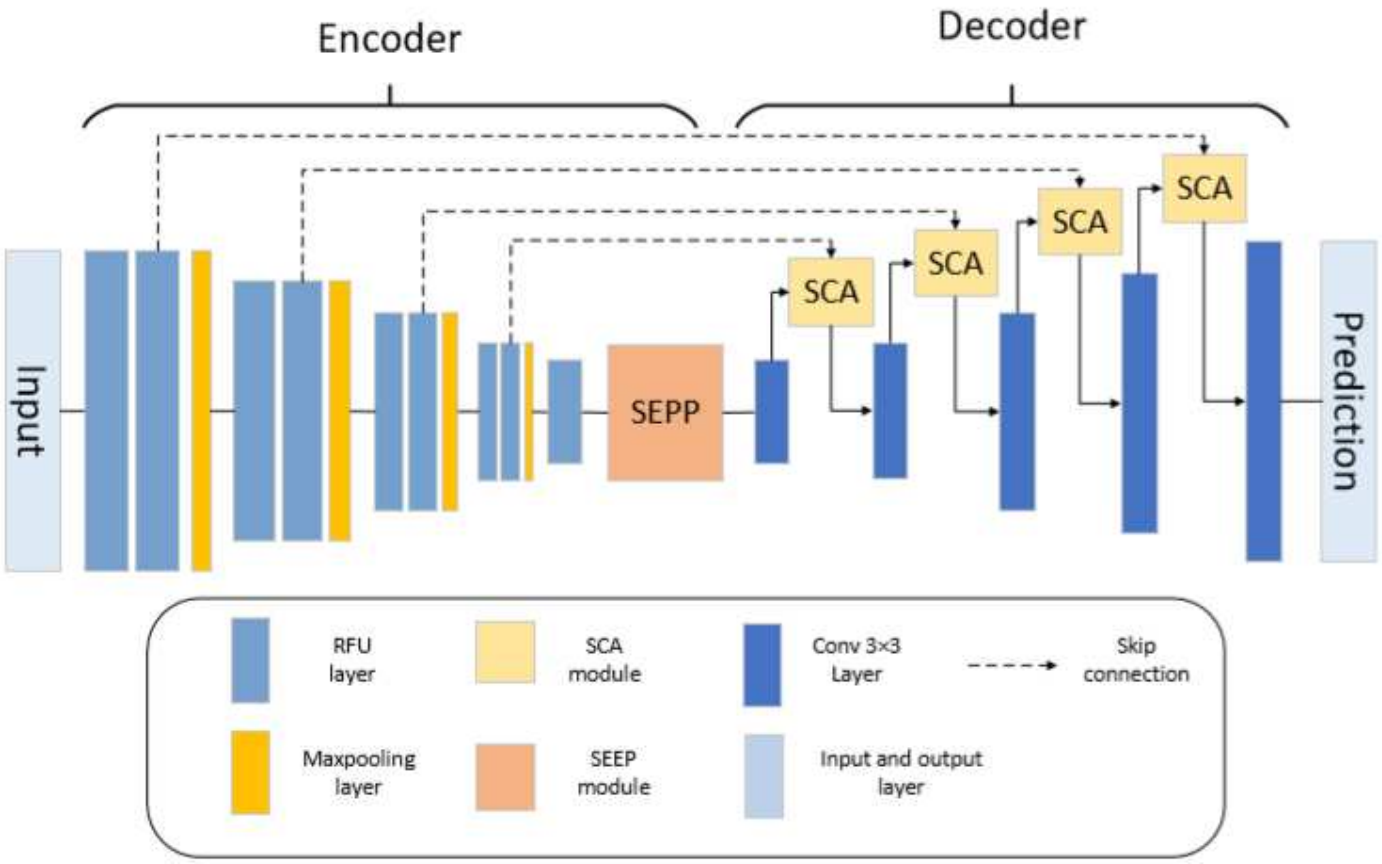


Figure 2

Illustration of the proposed SSCA-Net. We use multiple times of Relu Feature Unit (RFU) module as Feature Encoder, Then the feature maps are fed into a feature decoder module. It contains an self- and channel-attention (SCA) block and a squeeze and excitation pyramid pooling (SEPP) block. Moreover, we adopt skip connection to connect the low-level feature maps and high-level feature maps.

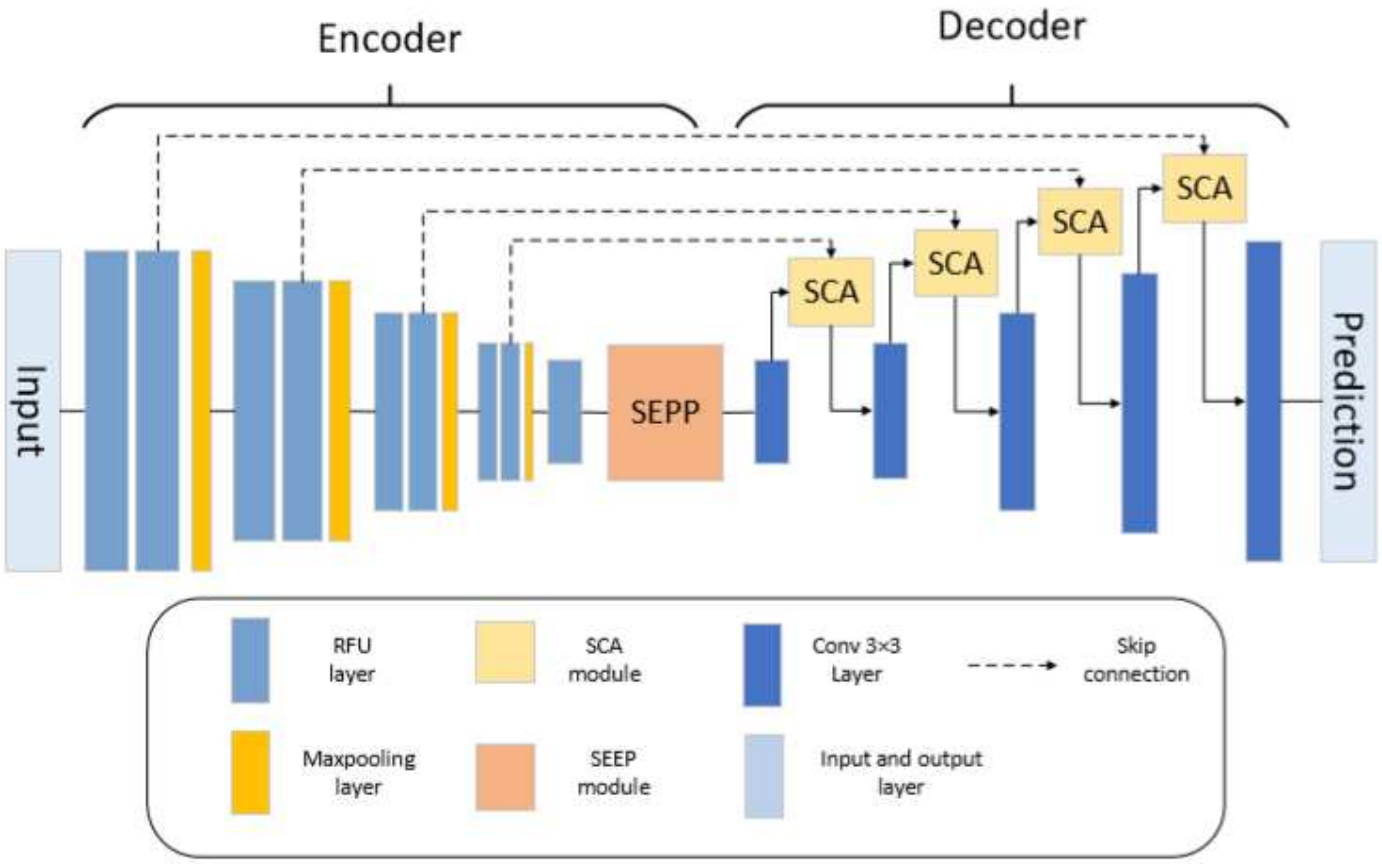


Figure 2

Illustration of the proposed SSCA-Net. We use multiple times of Relu Feature Unit (RFU) module as Feature Encoder, Then the feature maps are fed into a feature decoder module. It contains an self- and channel-attention (SCA) block and a squeeze and excitation pyramid pooling (SEPP) block. Moreover, we adopt skip connection to connect the low-level feature maps and high-level feature maps.

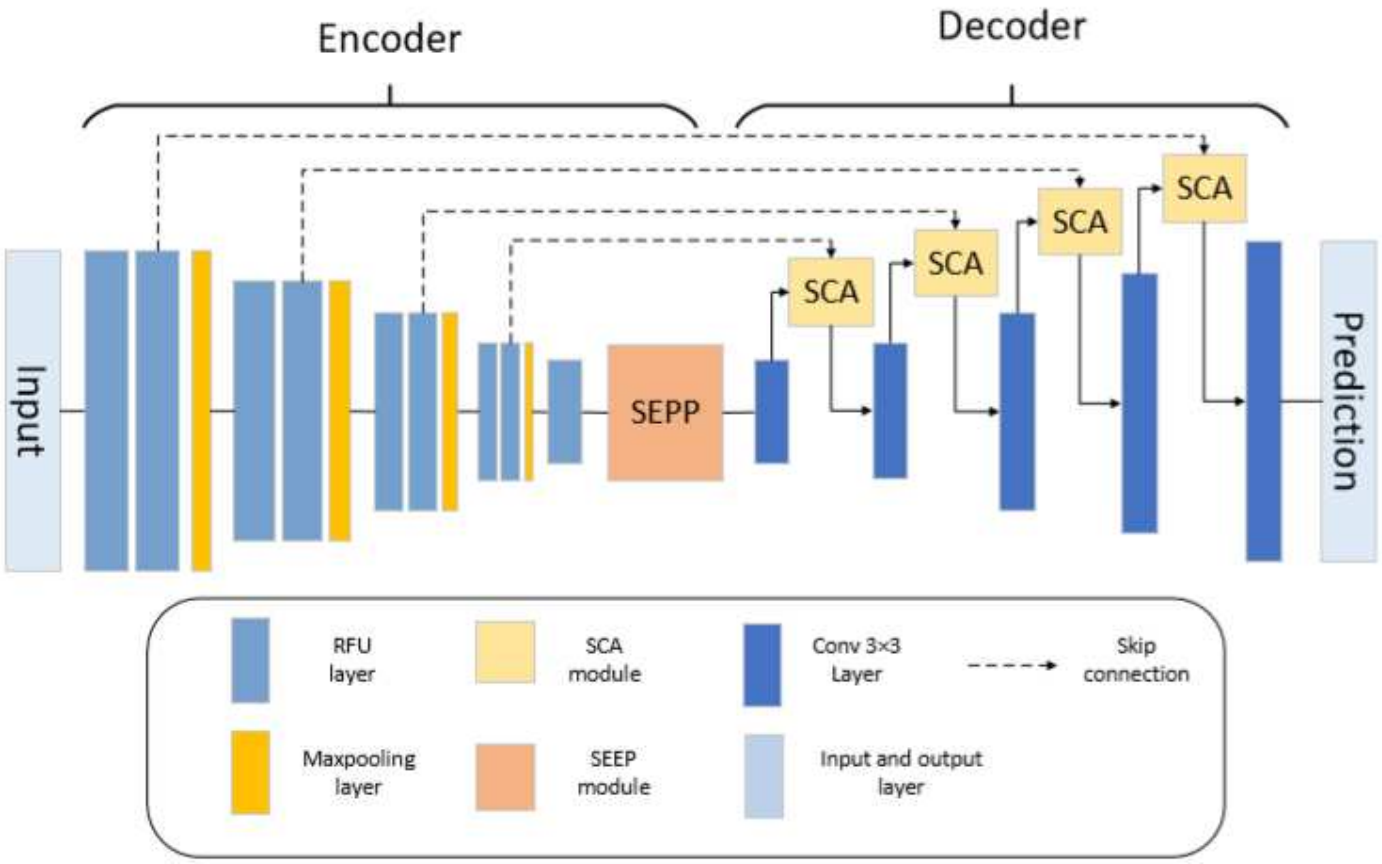


Figure 2

Illustration of the proposed SSCA-Net. We use multiple times of Relu Feature Unit (RFU) module as Feature Encoder, Then the feature maps are fed into a feature decoder module. It contains an self- and channel-attention (SCA) block and a squeeze and excitation pyramid pooling (SEPP) block. Moreover, we adopt skip connection to connect the low-level feature maps and high-level feature maps.

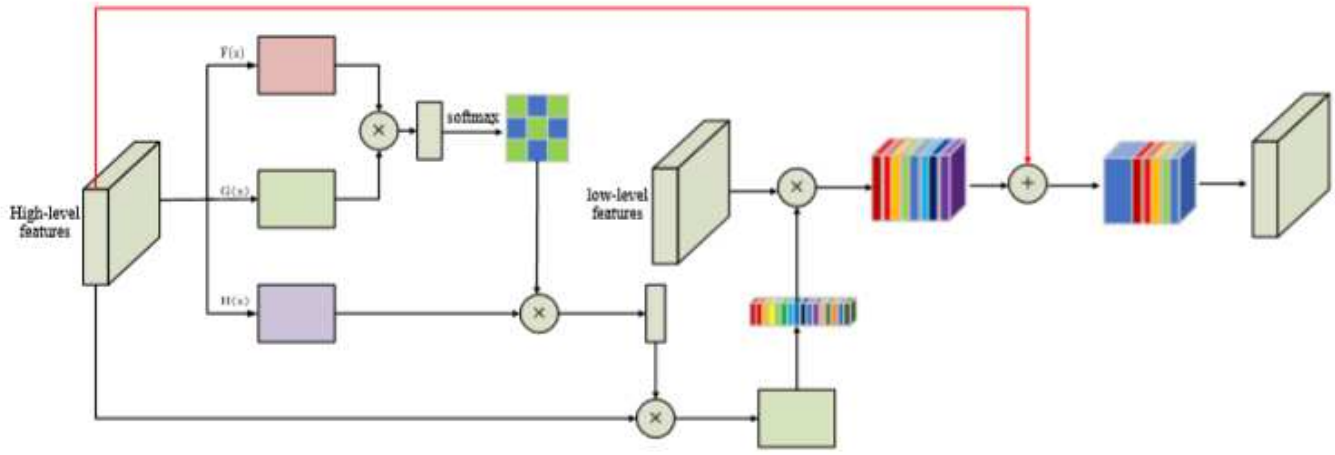


Figure 3

The designed Self- and channel-attention module (SCA) module for aggregating high-level features and low-level features. “⊗” denotes spatial element-wise multiplication and “⊕” denotes element-wise sum. The red lines represent the up-sampling operators.

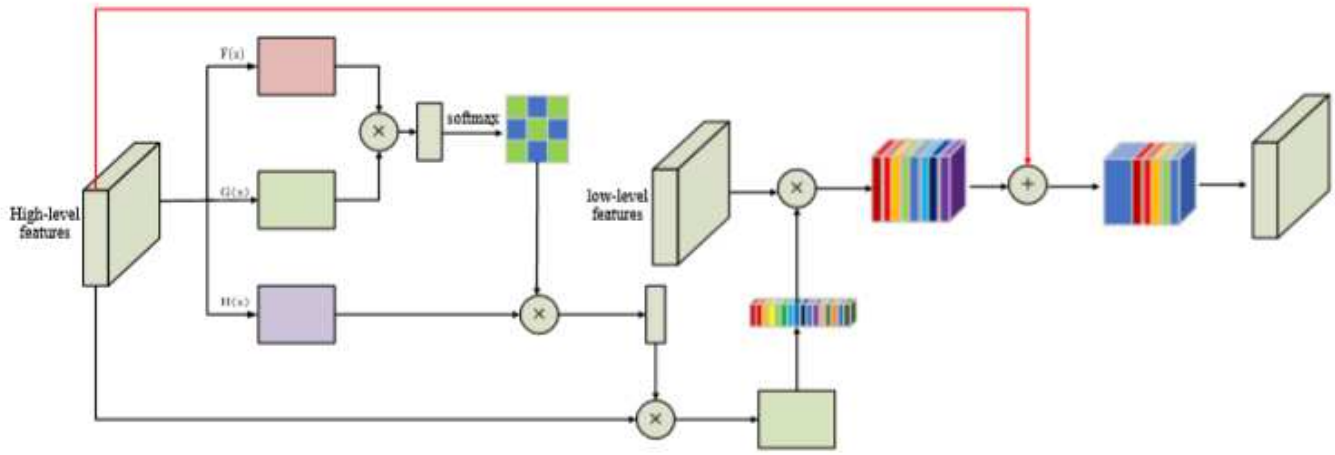


Figure 3

The designed Self- and channel-attention module (SCA) module for aggregating high-level features and low-level features. “⊗” denotes spatial element-wise multiplication and “⊕” denotes element-wise sum. The red lines represent the up-sampling operators.

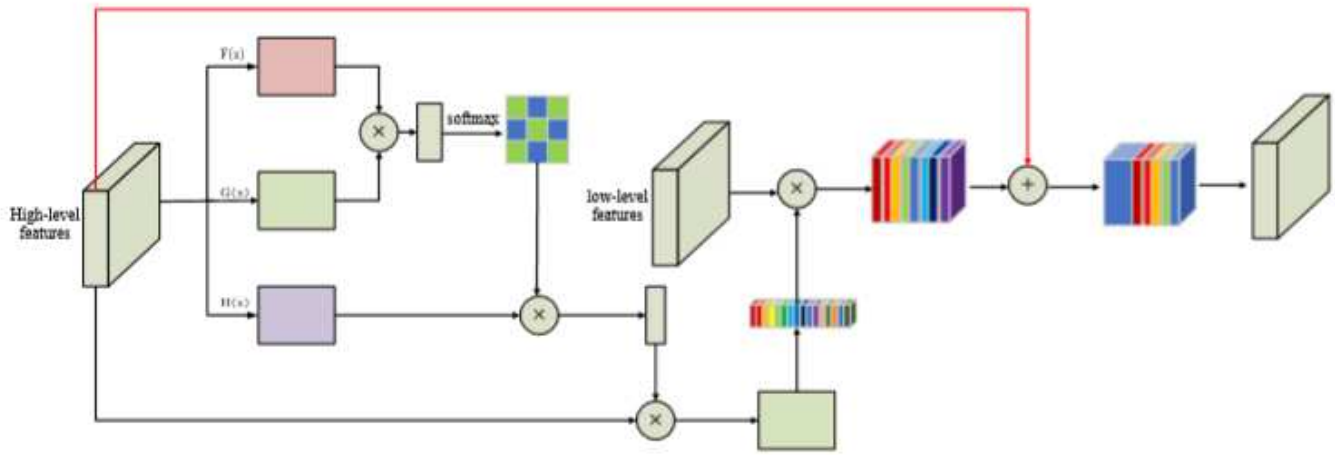


Figure 3

The designed Self- and channel-attention module (SCA) module for aggregating high-level features and low-level features. “⊗” denotes spatial element-wise multiplication and “⊕” denotes element-wise sum. The red lines represent the up-sampling operators.

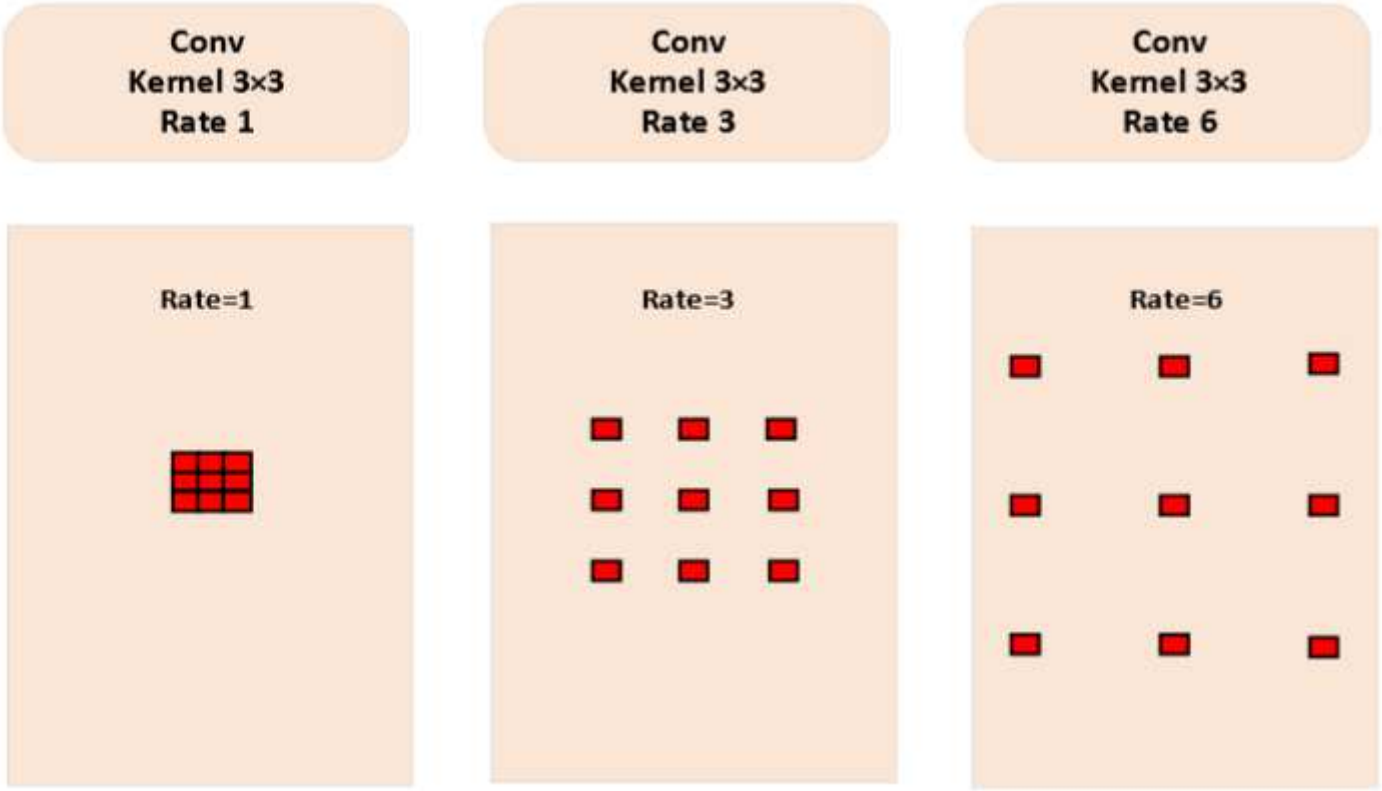


Figure 4

The illustrations of atrous convolution.

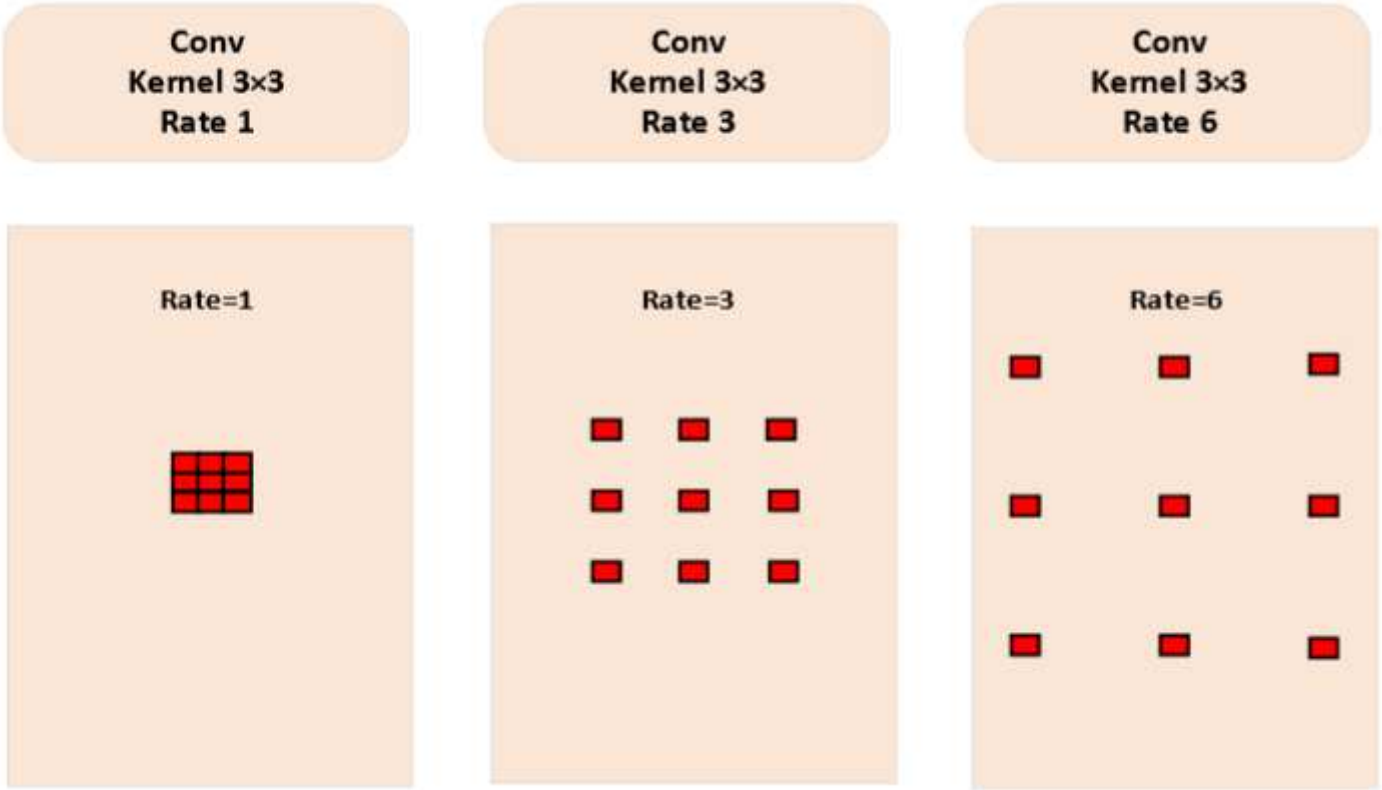


Figure 4

The illustrations of atrous convolution.

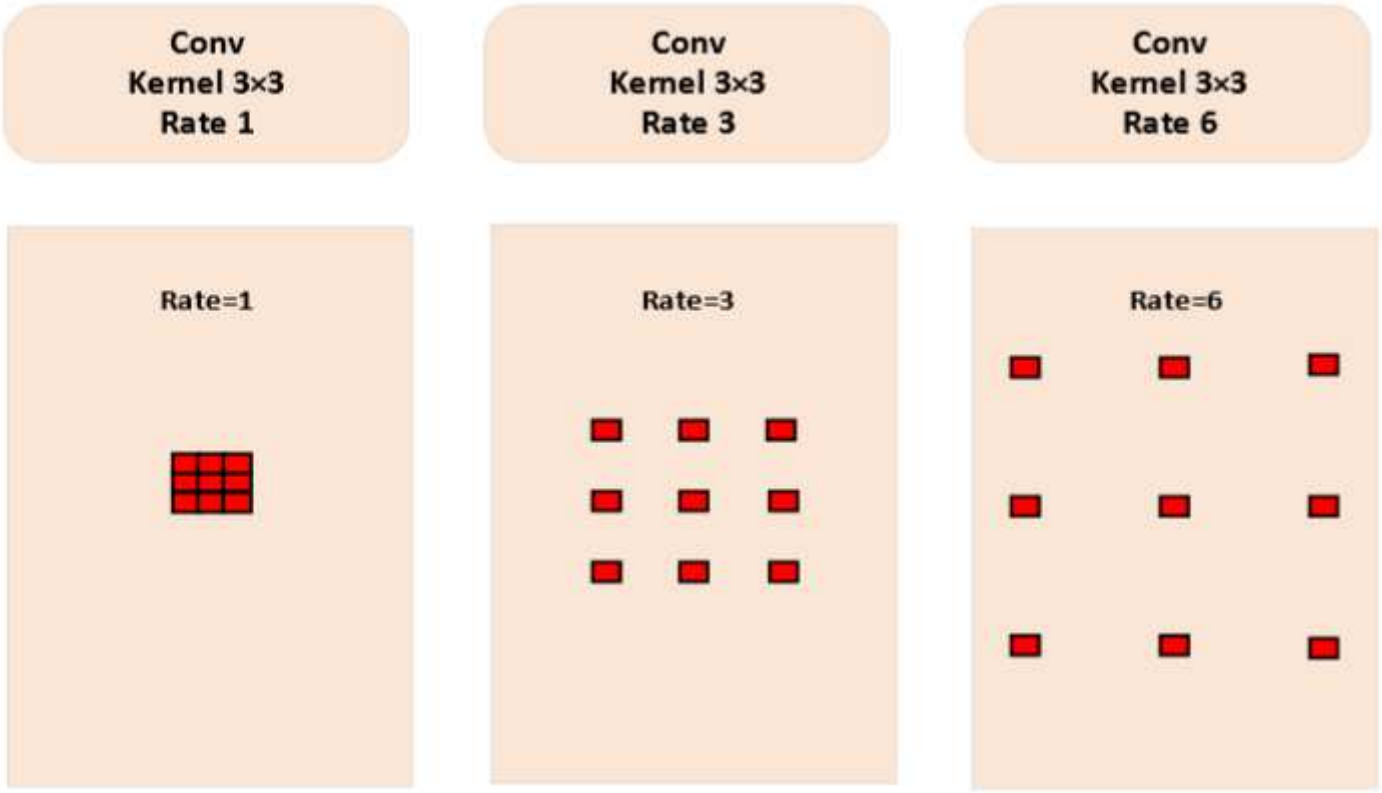


Figure 4

The illustrations of atrous convolution.

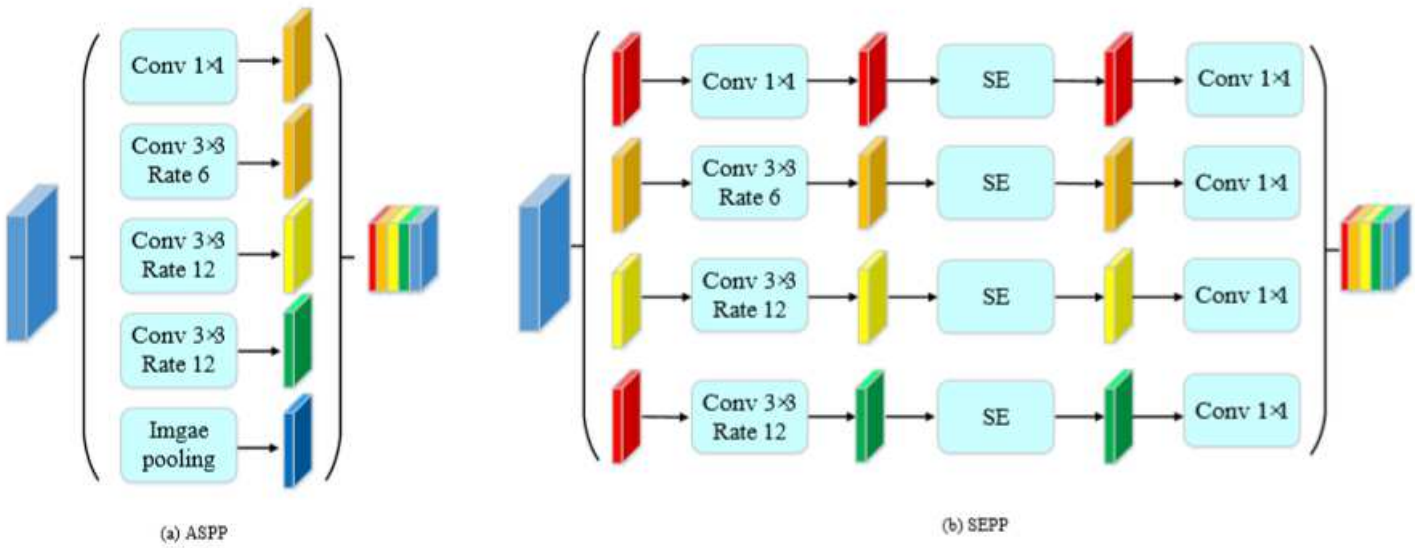


Figure 5

Illustration of (a) atrous spatial pyramid pooling (ASPP) and (b) Squeeze-and-excitation pyramid

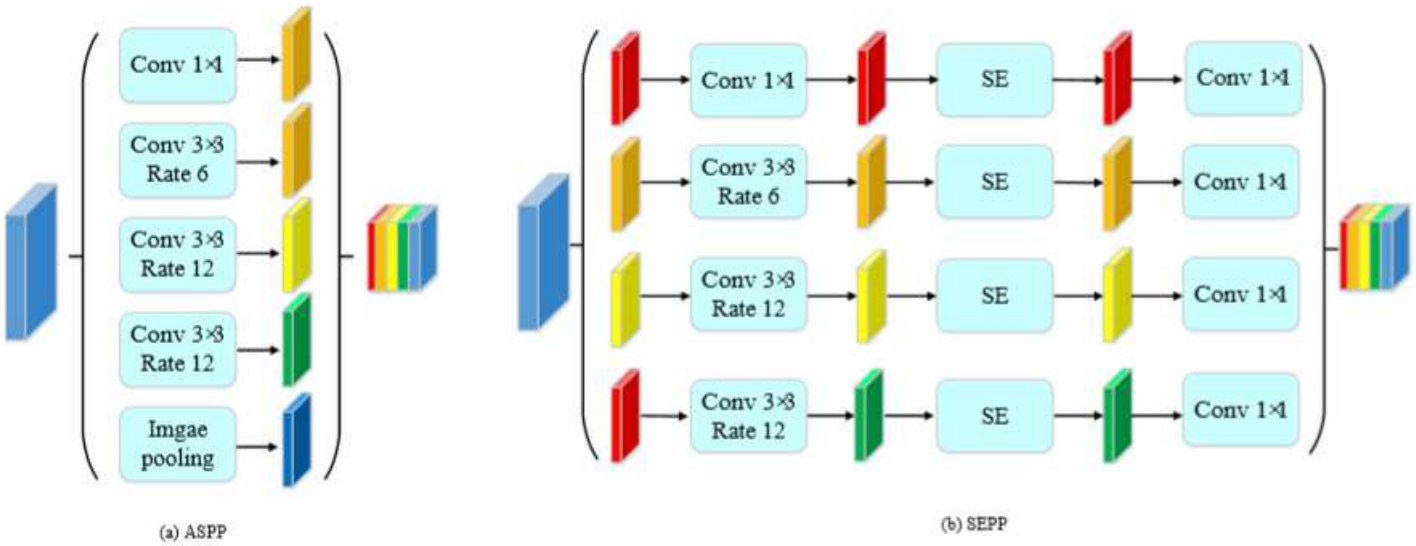


Figure 5

Illustration of (a) atrous spatial pyramid pooling (ASPP) and (b) Squeeze-and-excitation pyramid

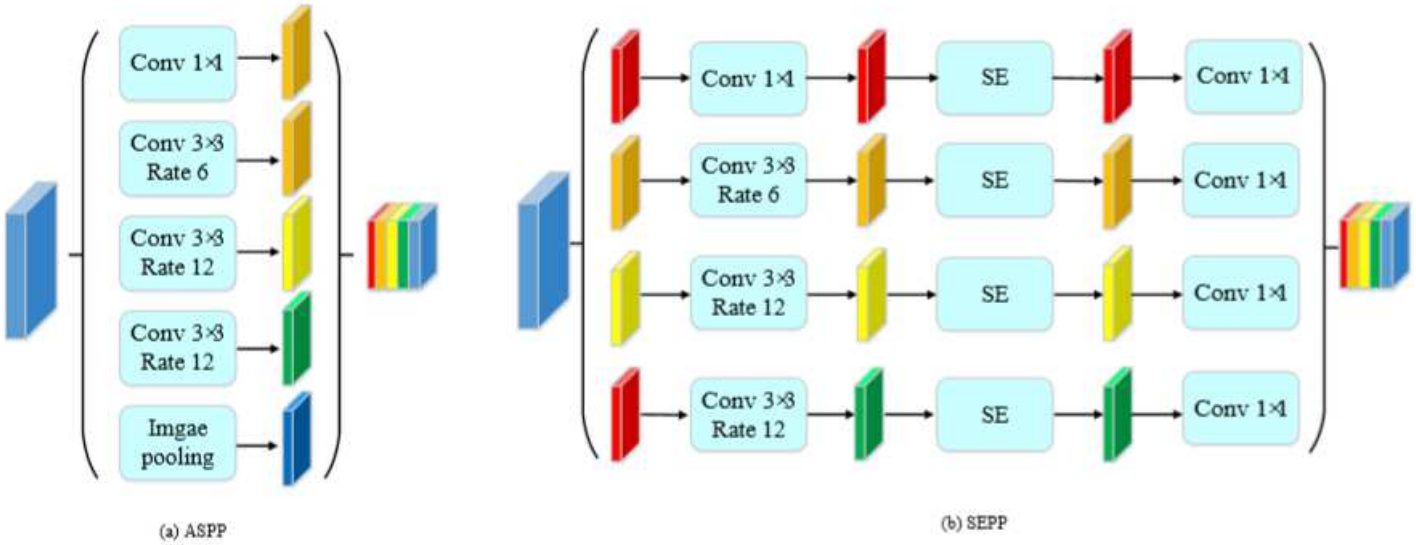


Figure 5

Illustration of (a) atrous spatial pyramid pooling (ASPP) and (b) Squeeze-and-excitation pyramid

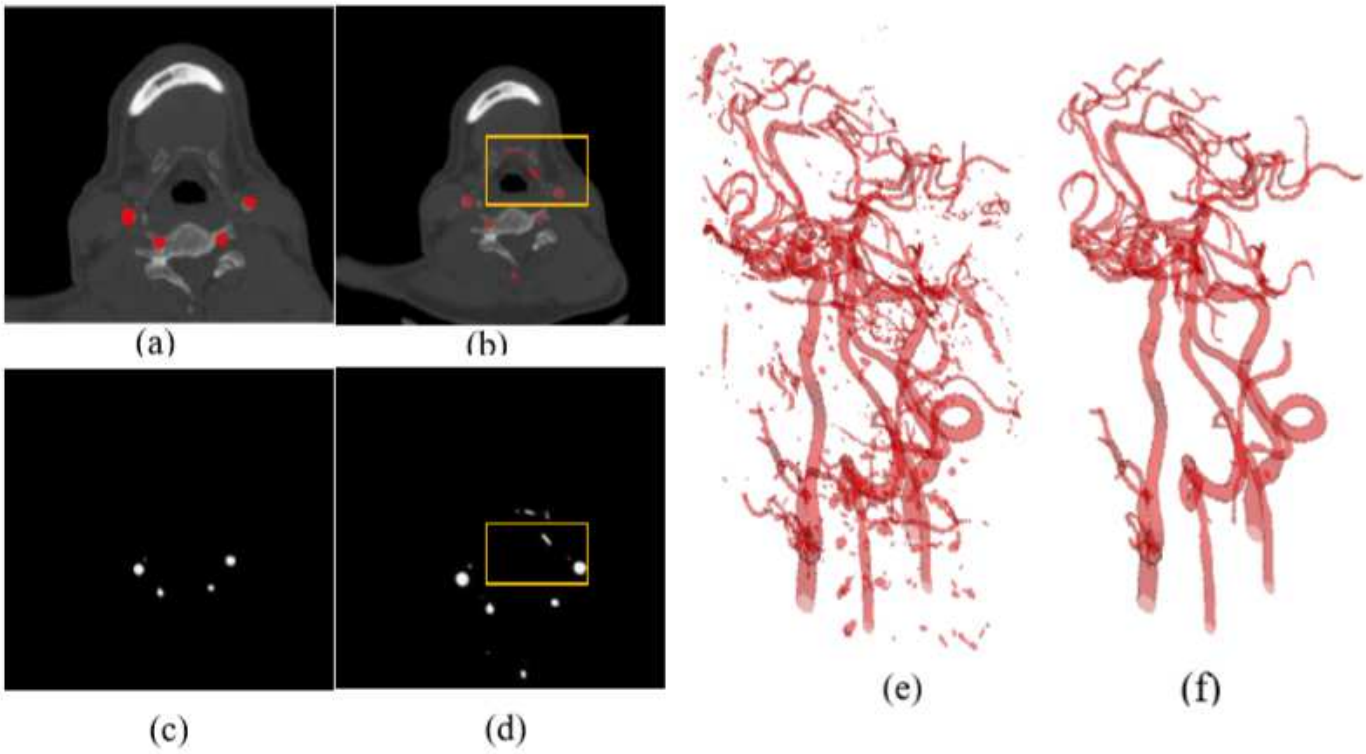


Figure 6

The effectiveness of post-processing. (a) and (c) show some non-vessel areas have been removed after post-processing, compared to the results of (b) and (d) before post-processing. (e) and (f) show the 3D results before and after post-processing, respectively.

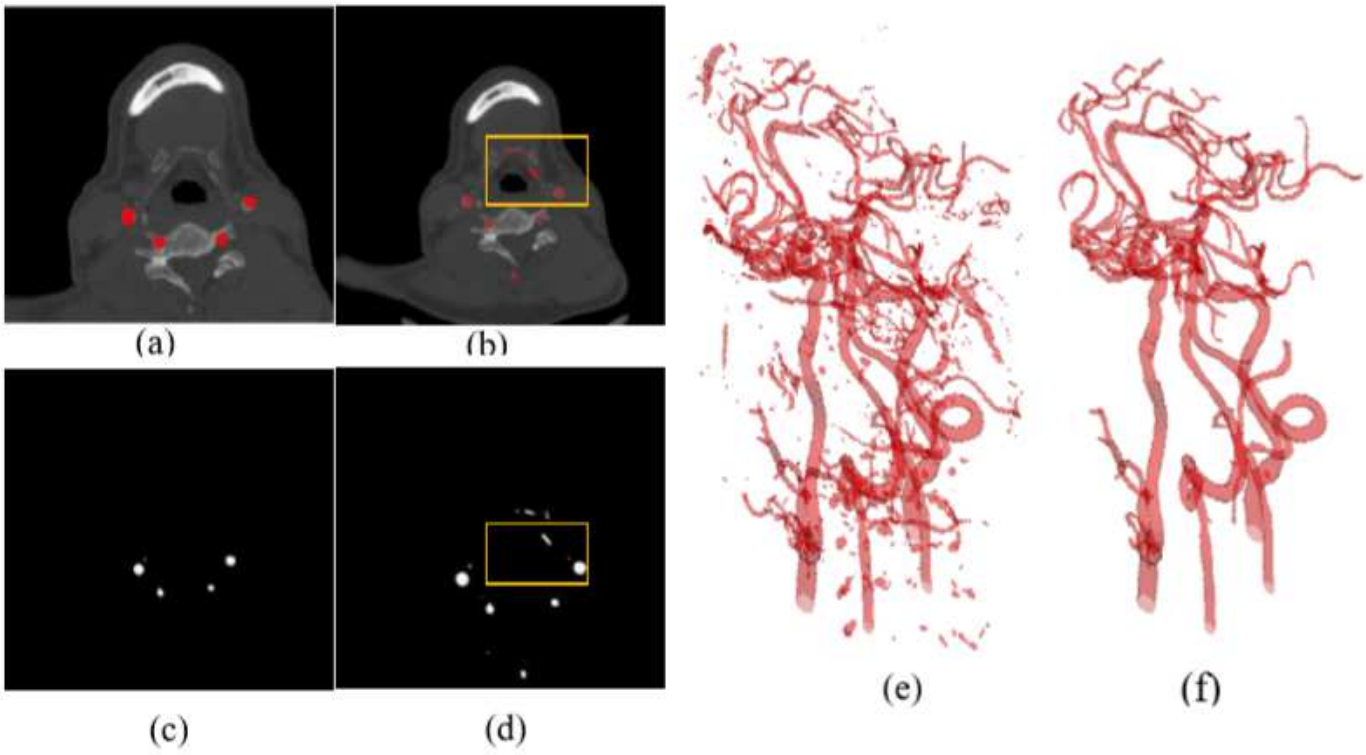


Figure 6

The effectiveness of post-processing. (a) and (c) show some non-vessel areas have been removed after post-processing, compared to the results of (b) and (d) before post-processing. (e) and (f) show the 3D results before and after post-processing, respectively.

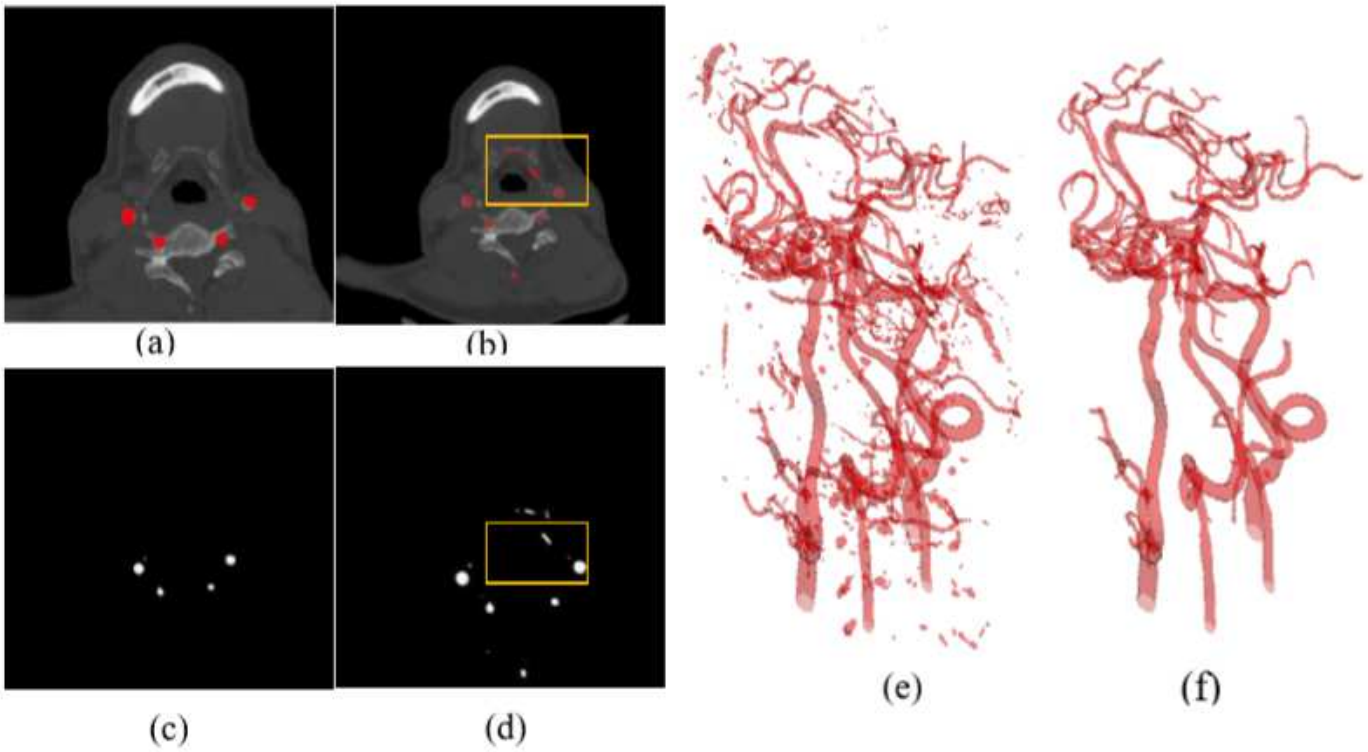


Figure 6

The effectiveness of post-processing. (a) and (c) show some non-vessel areas have been removed after post-processing, compared to the results of (b) and (d) before post-processing. (e) and (f) show the 3D results before and after post-processing, respectively.

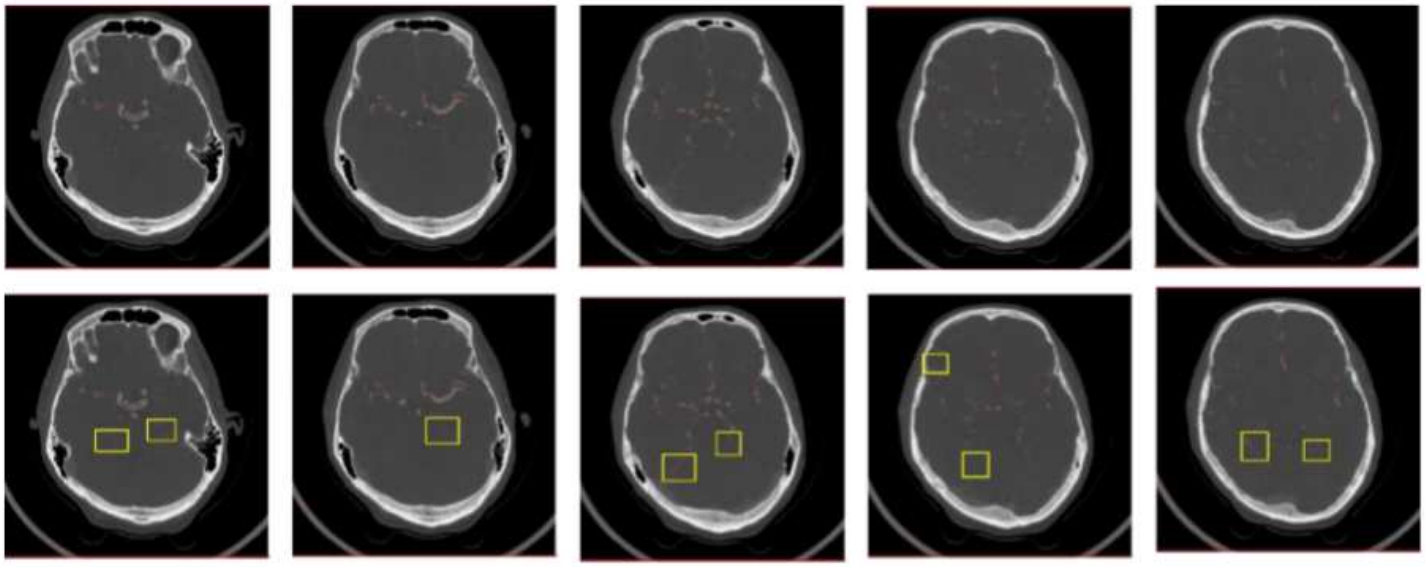


Figure 7

2D visualization of segmentation results on the intracranial blood vessel data. The top row shows the ground truth; and the bottom row is our method

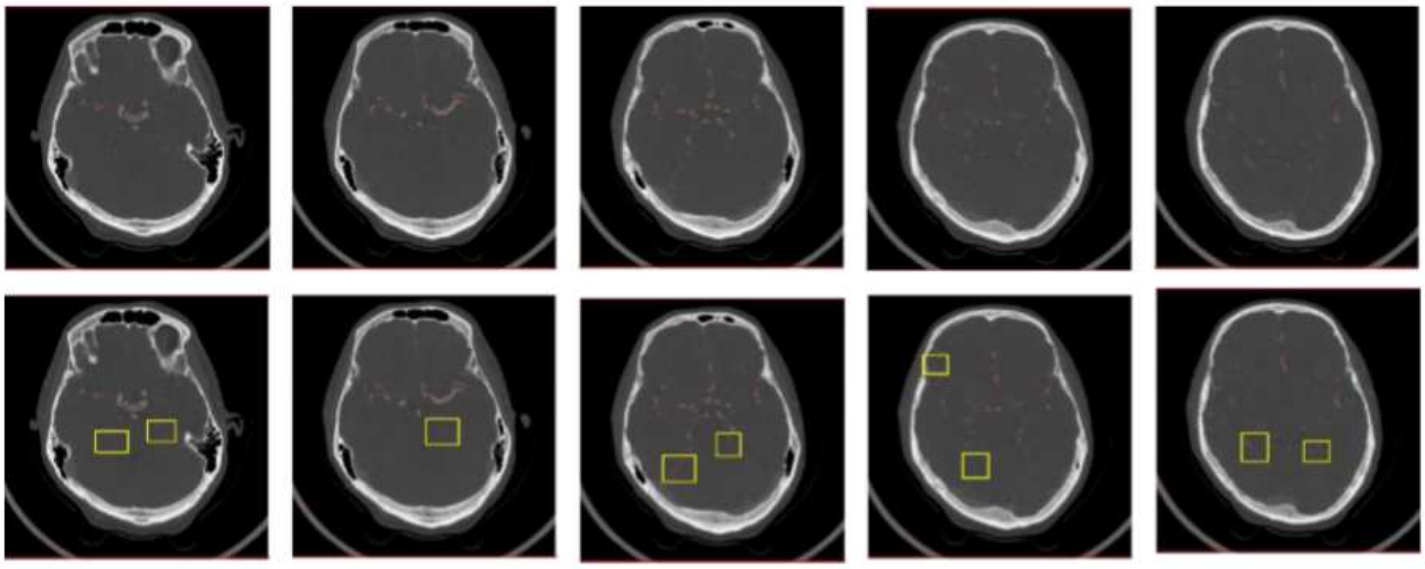


Figure 7

2D visualization of segmentation results on the intracranial blood vessel data. The top row shows the ground truth; and the bottom row is our method

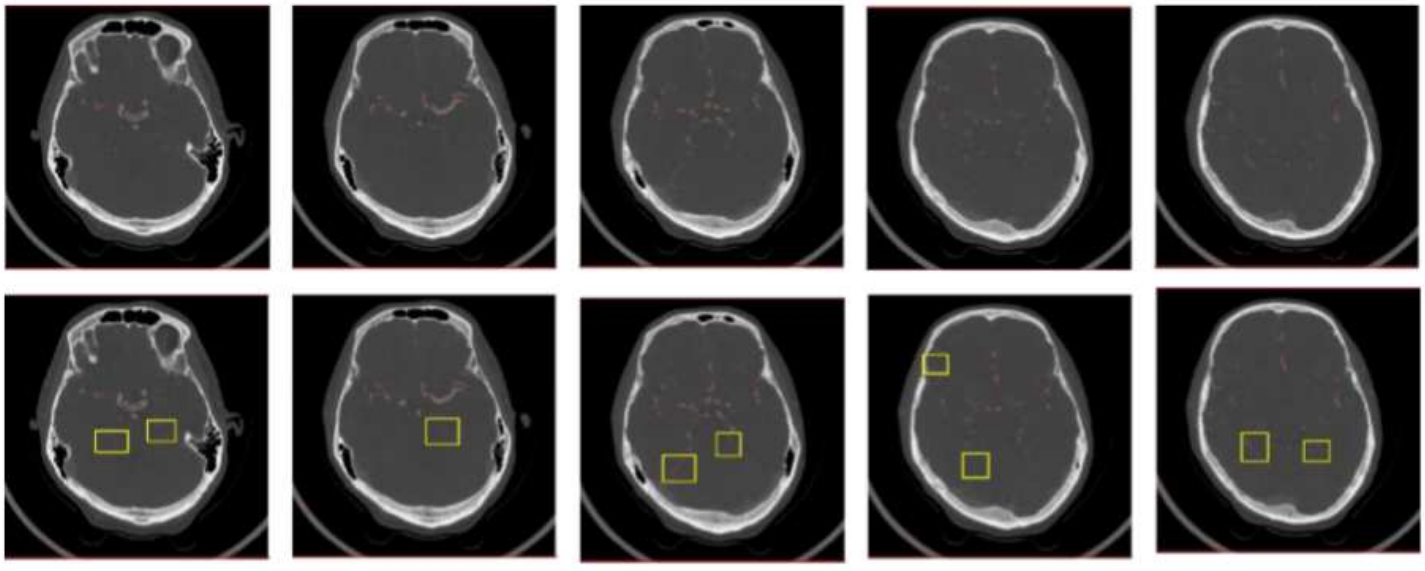


Figure 7

2D visualization of segmentation results on the intracranial blood vessel data. The top row shows the ground truth; and the bottom row is our method

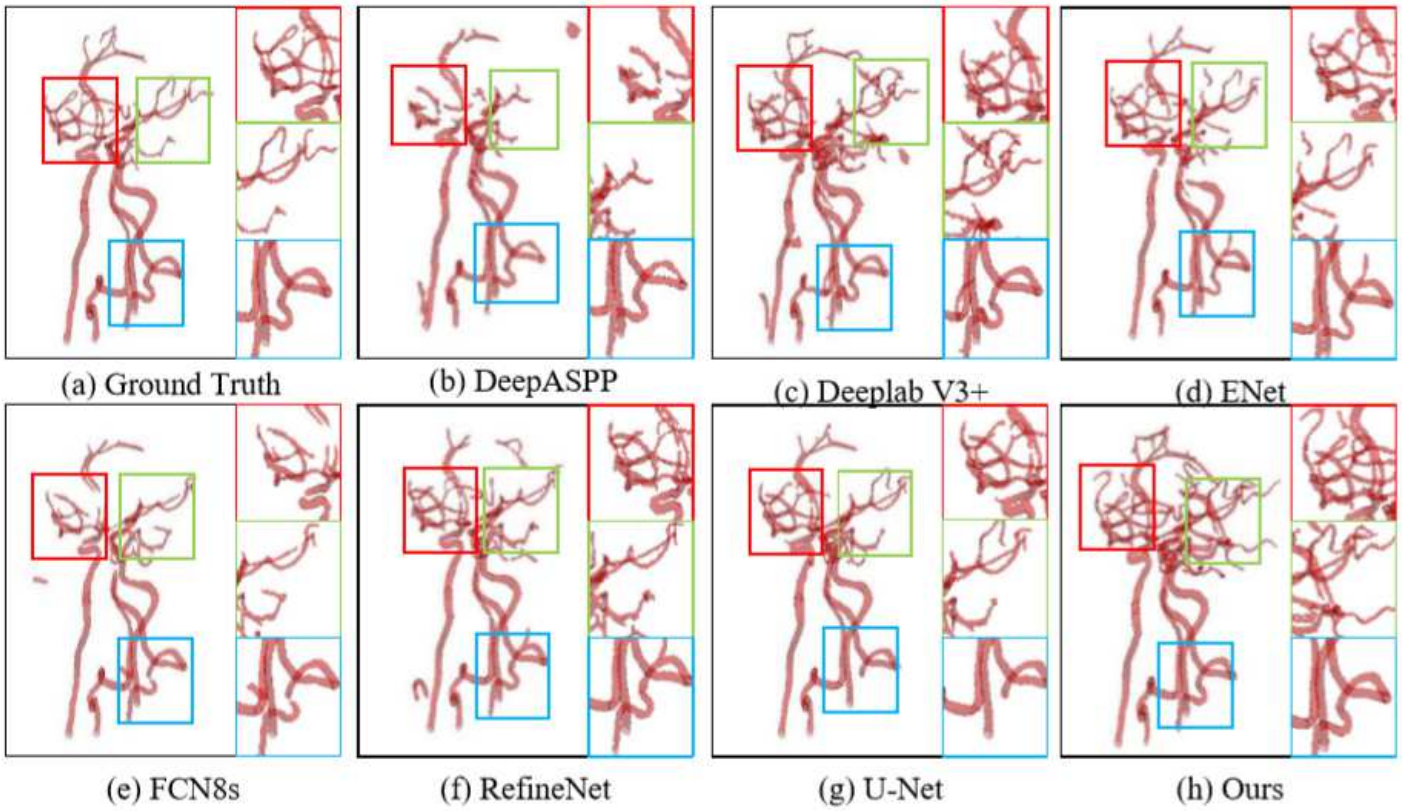


Figure 8

Medical image segmentation results tested in the dataset of intracranial artery. From (a)-(h): (a) ground truth, and the segmentation results of (b) DeepASPP, (c) DeepLab V3+, (d) ENet, (e) FCN8, (f) RefineNet, (g) U-Net and (h) ours, respectively. Our SSCA-Net can perform segmentation of intracranial arteries effectively while preserving more vessel tinny-scale structures.

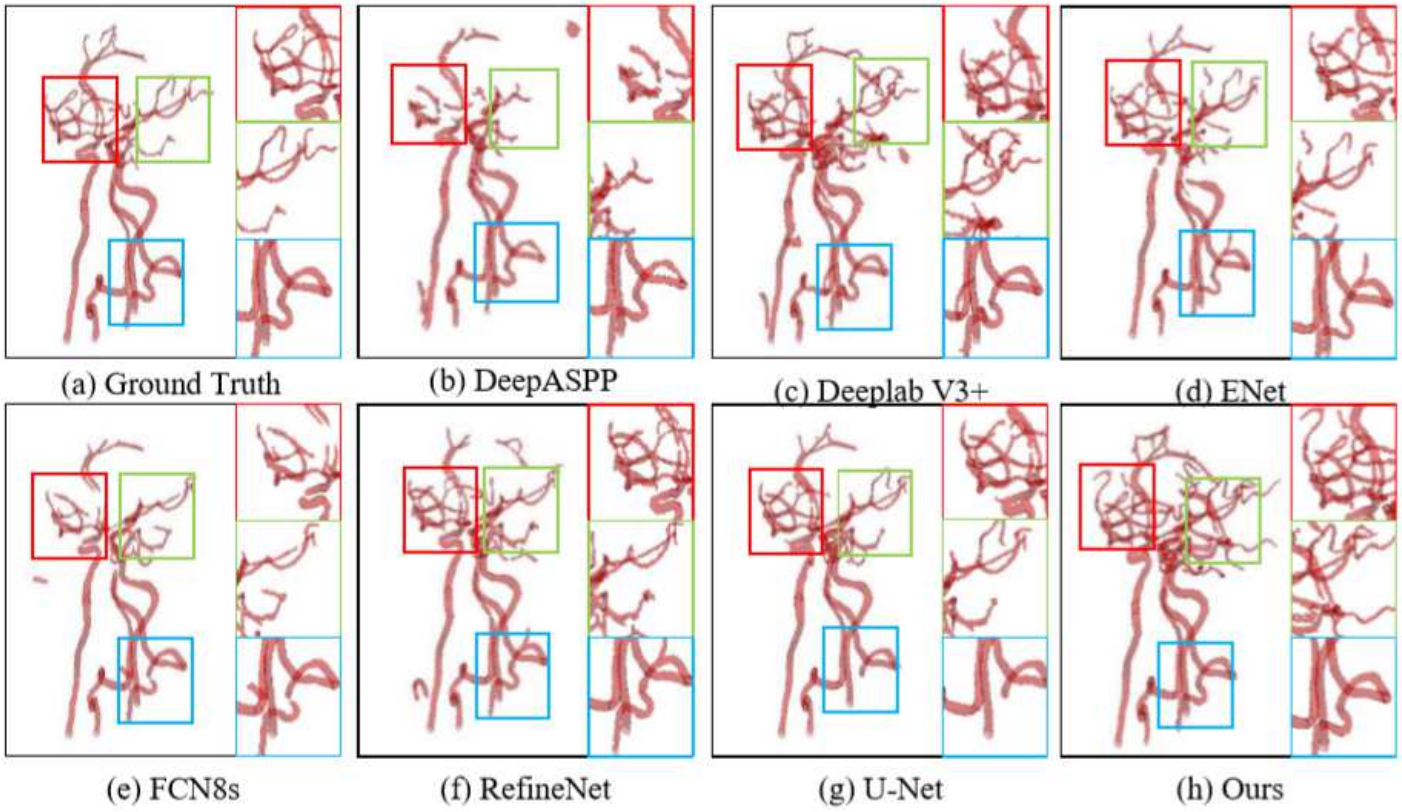


Figure 8

Medical image segmentation results tested in the dataset of intracranial artery. From (a)-(h): (a) ground truth, and the segmentation results of (b) DeepASPP, (c) DeepLab V3+, (d) ENet, (e) FCN8, (f) RefineNet, (g) U-Net and (h) ours, respectively. Our SSCA-Net can perform segmentation of intracranial arteries effectively while preserving more vessel tinny-scale structures.

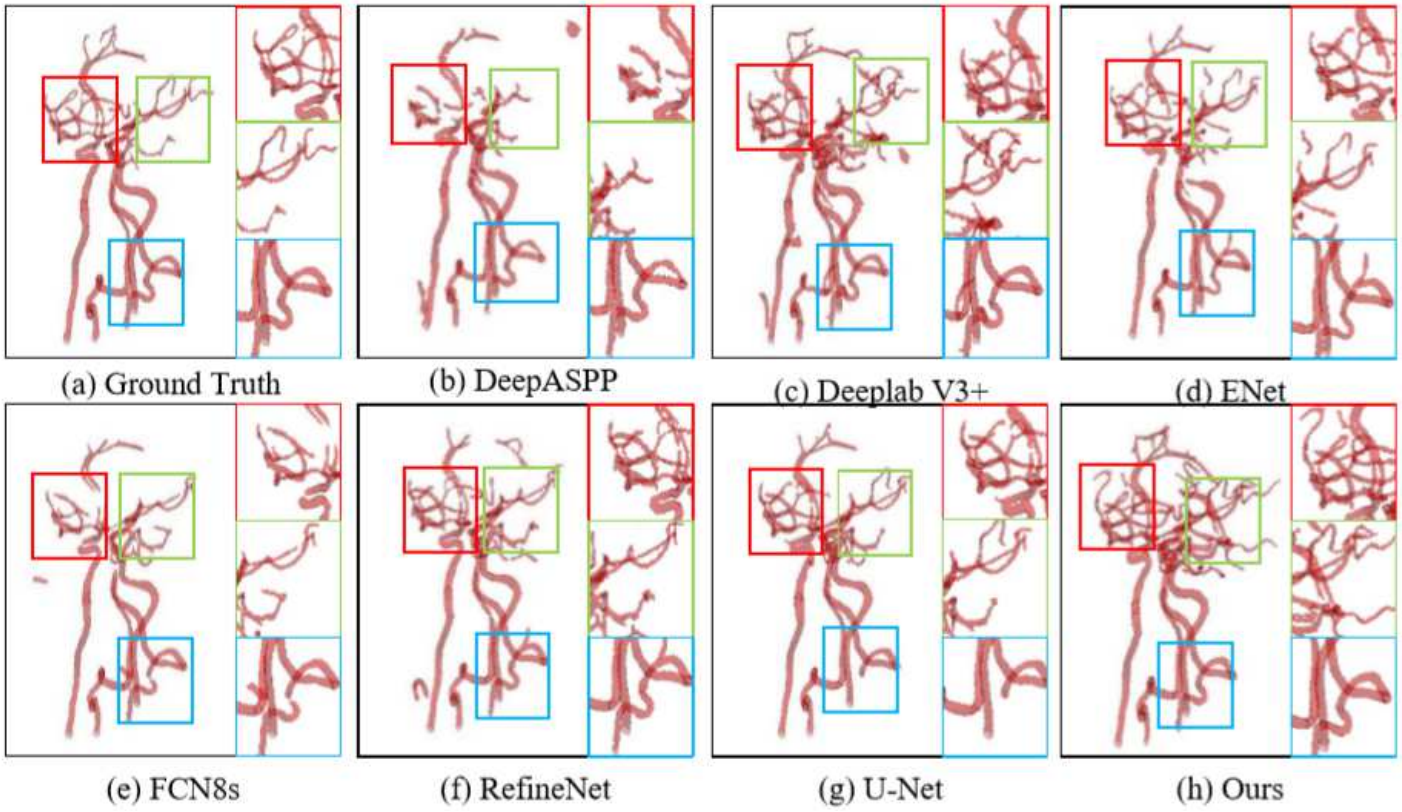


Figure 8

Medical image segmentation results tested in the dataset of intracranial artery. From (a)-(h): (a) ground truth, and the segmentation results of (b) DeepASPP, (c) DeepLab V3+, (d) ENet, (e) FCN8, (f) RefineNet, (g) U-Net and (h) ours, respectively. Our SSCA-Net can perform segmentation of intracranial arteries effectively while preserving more vessel tinny-scale structures.

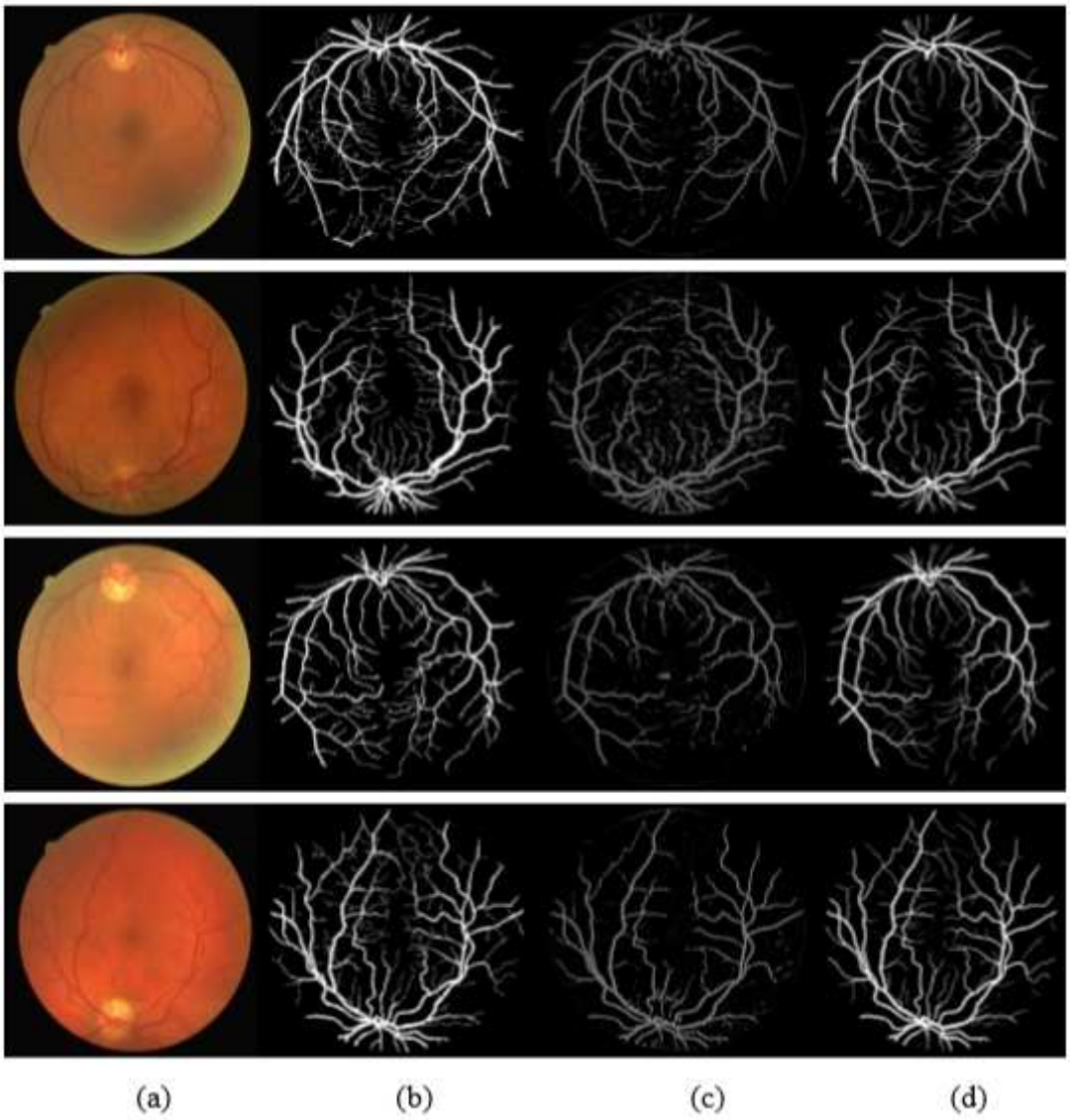


Figure 9

Visualization results on the DRIVE dataset. From (a)-(d): (a) test image, (b) ground truth, and results of (c) U-Net and (d) SSCA-Net, respectively.

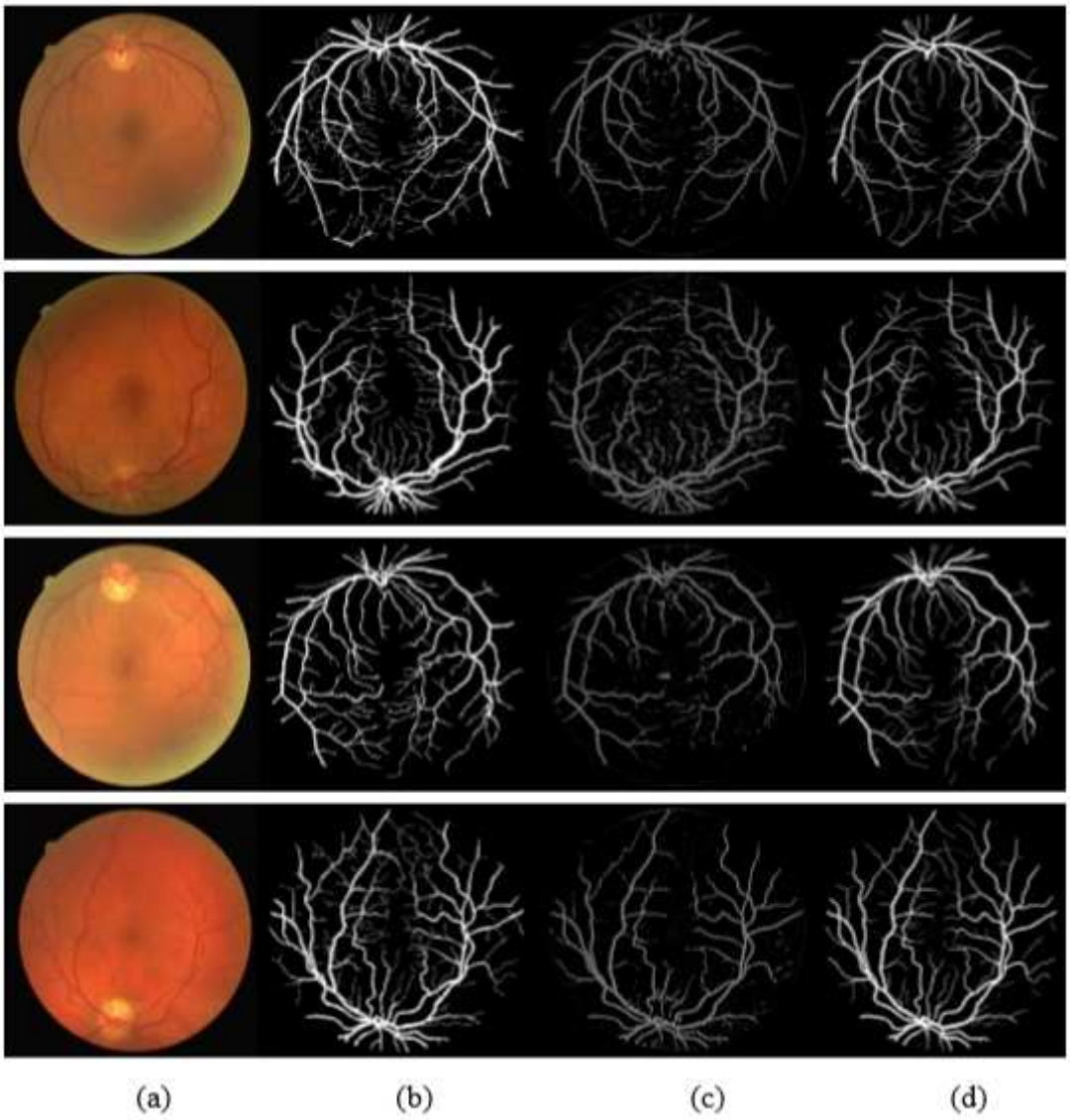


Figure 9

Visualization results on the DRIVE dataset. From (a)-(d): (a) test image, (b) ground truth, and results of (c) U-Net and (d) SSCA-Net, respectively.

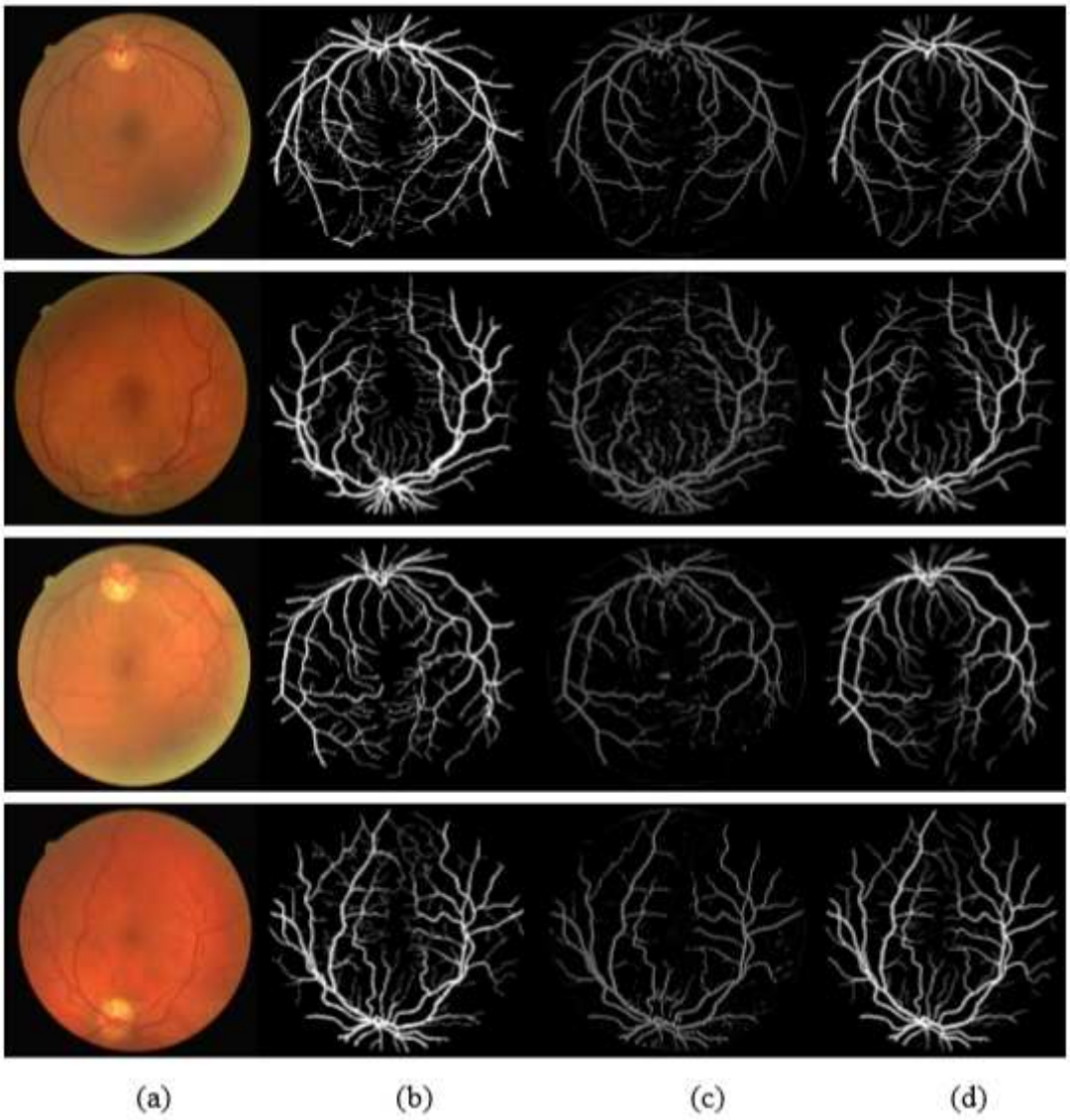


Figure 9

Visualization results on the DRIVE dataset. From (a)-(d): (a) test image, (b) ground truth, and results of (c) U-Net and (d) SSCA-Net, respectively.

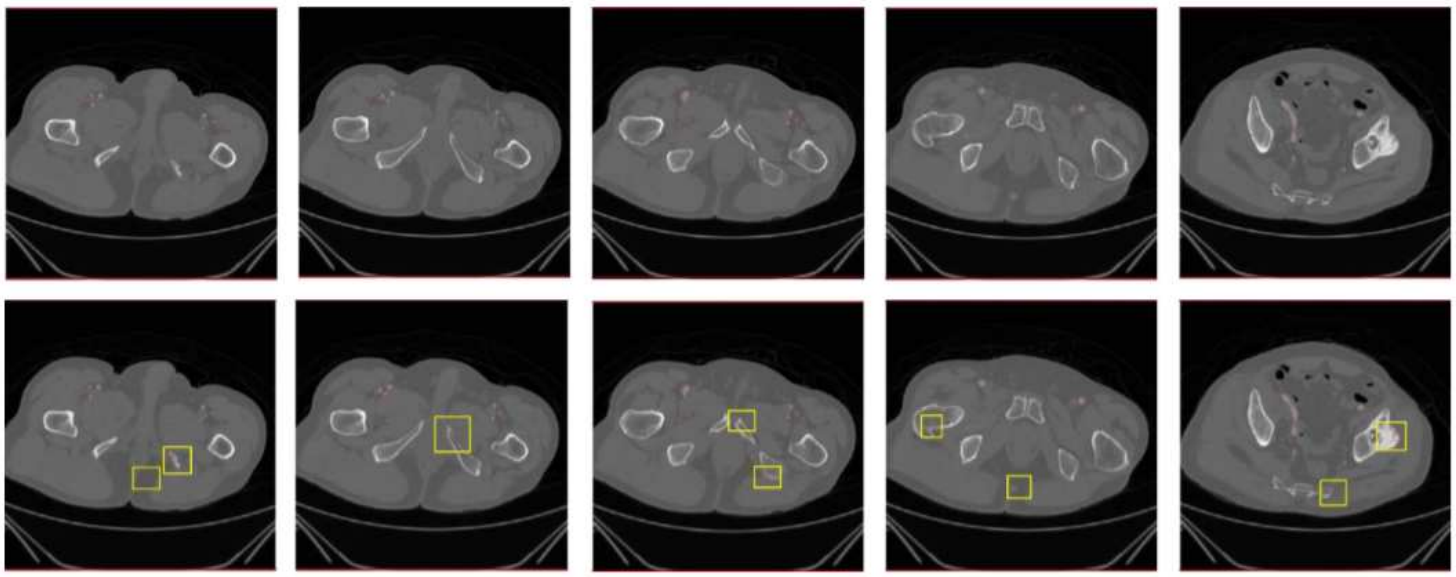


Figure 10

2D visualization of segmentation results on the leg artery data. The top row shows the ground truth; and the bottom row is our method.

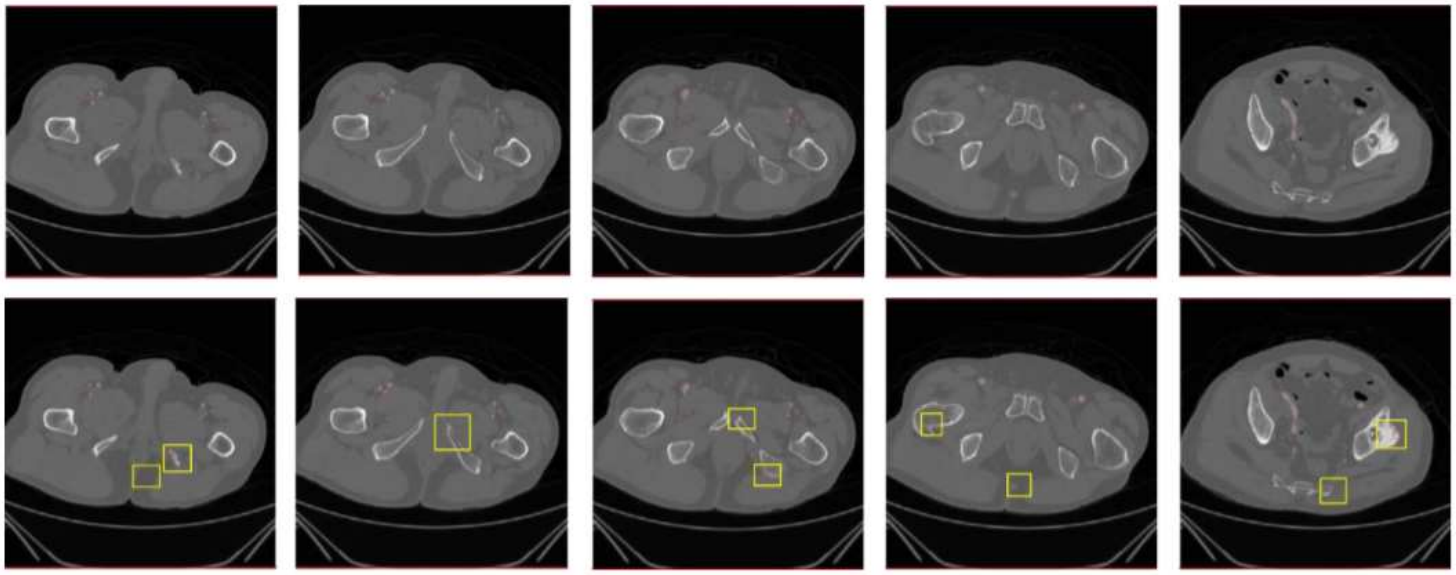


Figure 10

2D visualization of segmentation results on the leg artery data. The top row shows the ground truth; and the bottom row is our method.

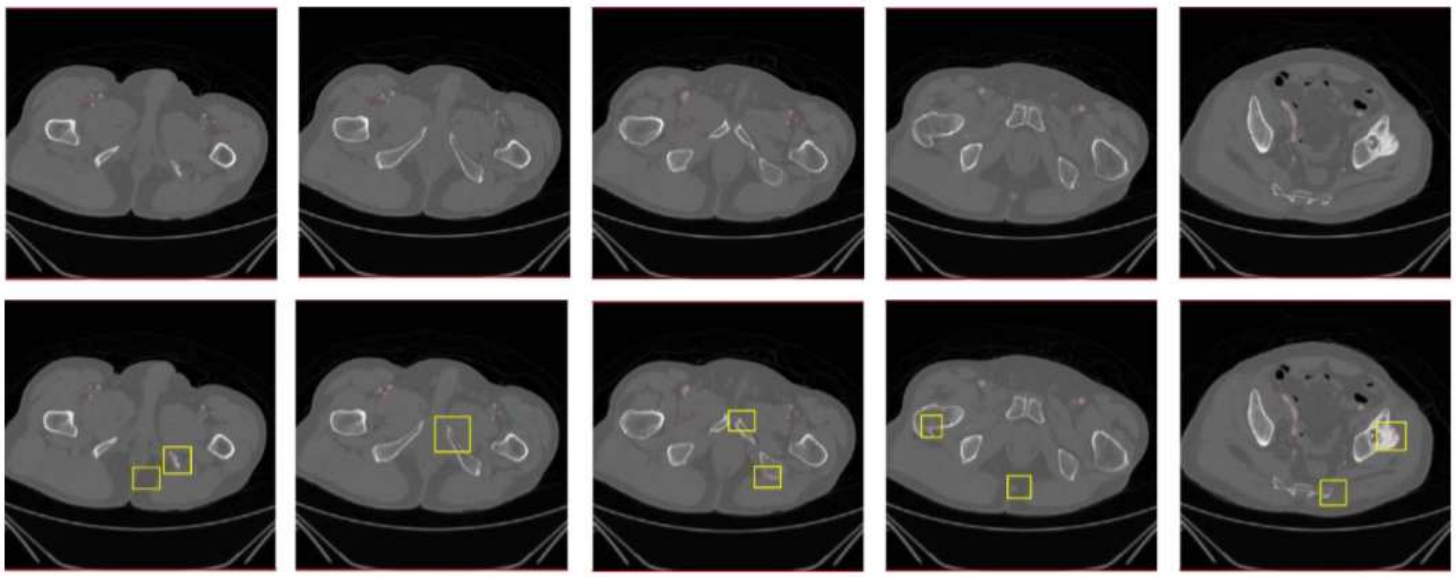


Figure 10

2D visualization of segmentation results on the leg artery data. The top row shows the ground truth; and the bottom row is our method.

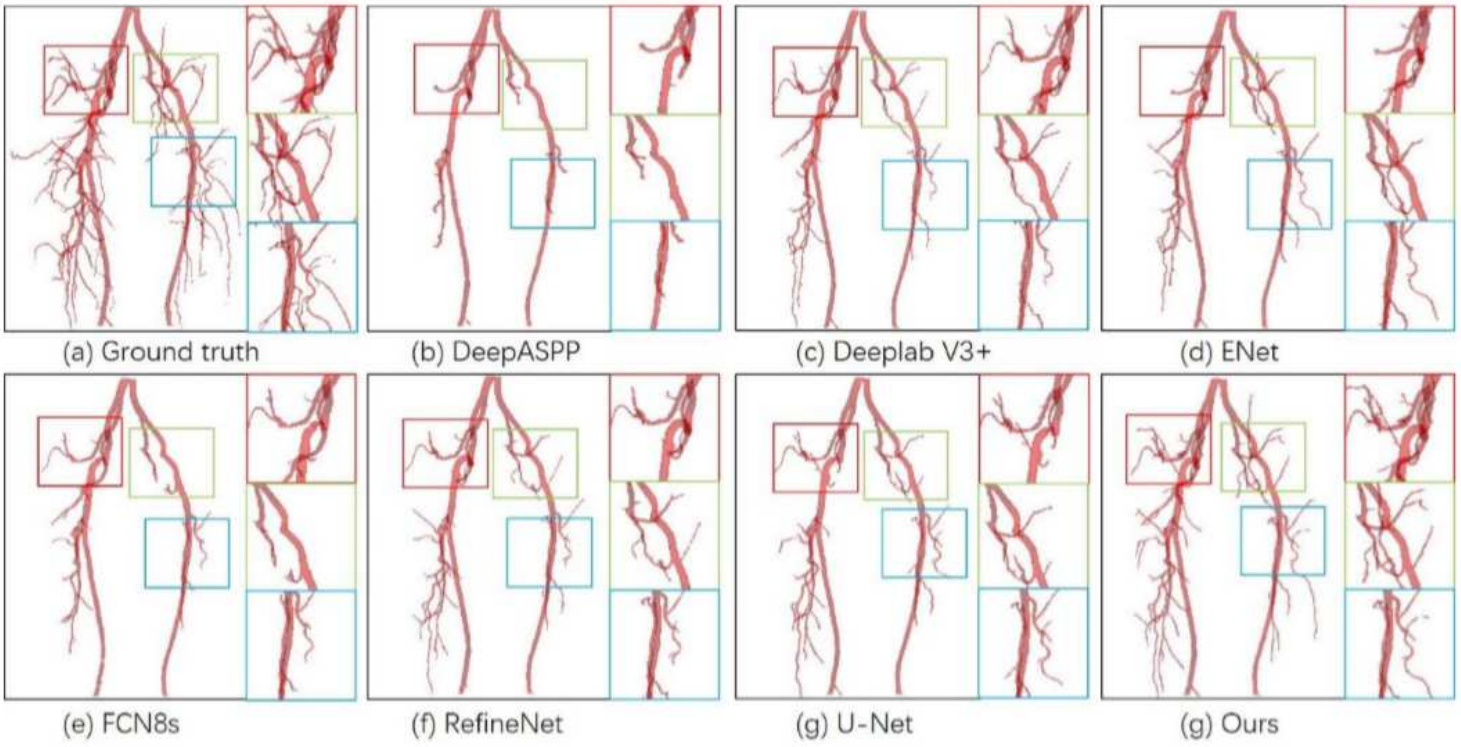


Figure 11

Comparative visualization of 3D results achieved on the test dataset1 and test datset2. Compared to the ground truth, all the four state-of-the-art methods (DeepASPP, DeepLap V3+, ENet, FCN8s, RefineNet and U-Net) miss fine features (e.g. small vessels in rectangle), whereas the proposed method preserves fine vessels well.

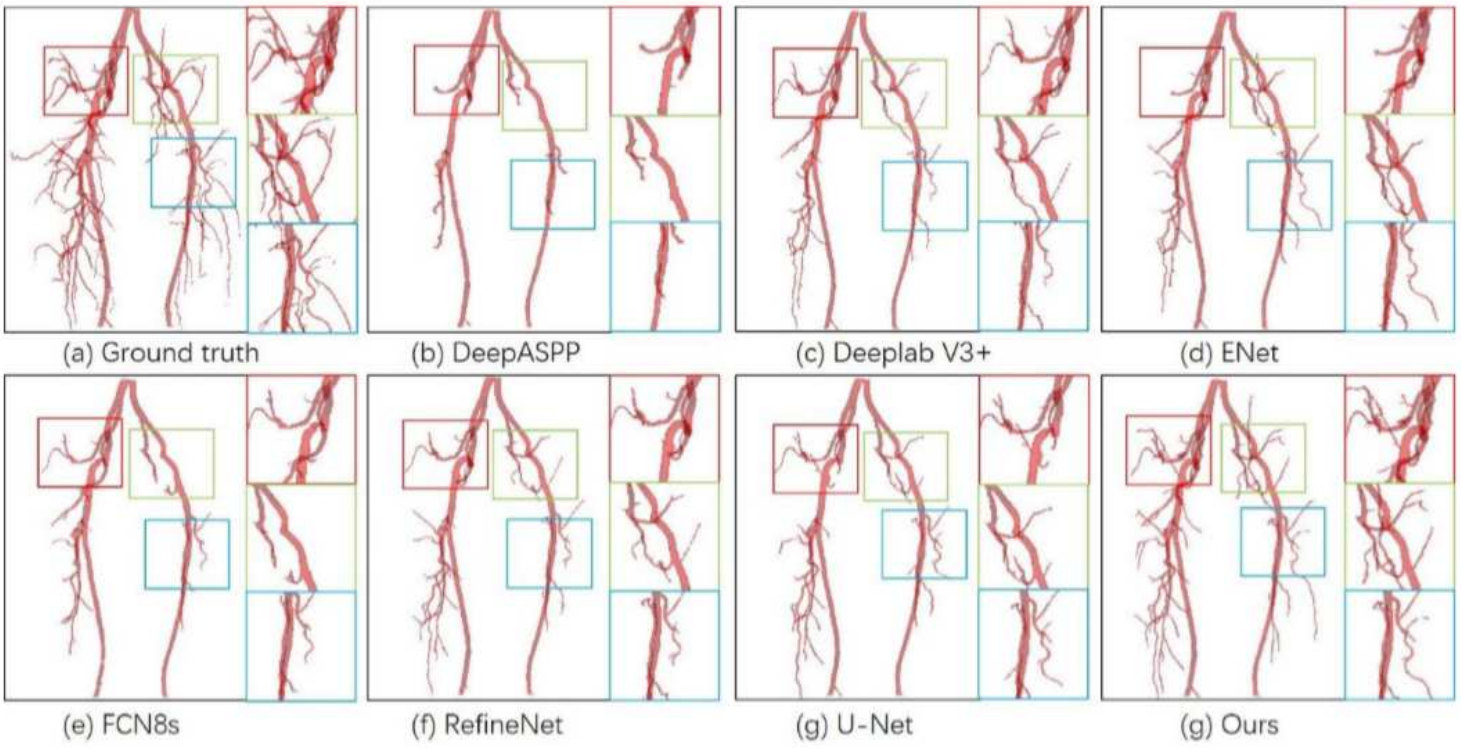


Figure 11

Comparative visualization of 3D results achieved on the test dataset1 and test datset2. Compared to the ground truth, all the four state-of-the-art methods (DeepASPP, DeepLap V3+, ENet, FCN8s, RefineNet and U-Net) miss fine features (e.g. small vessels in rectangle), whereas the proposed method preserves fine vessels well.

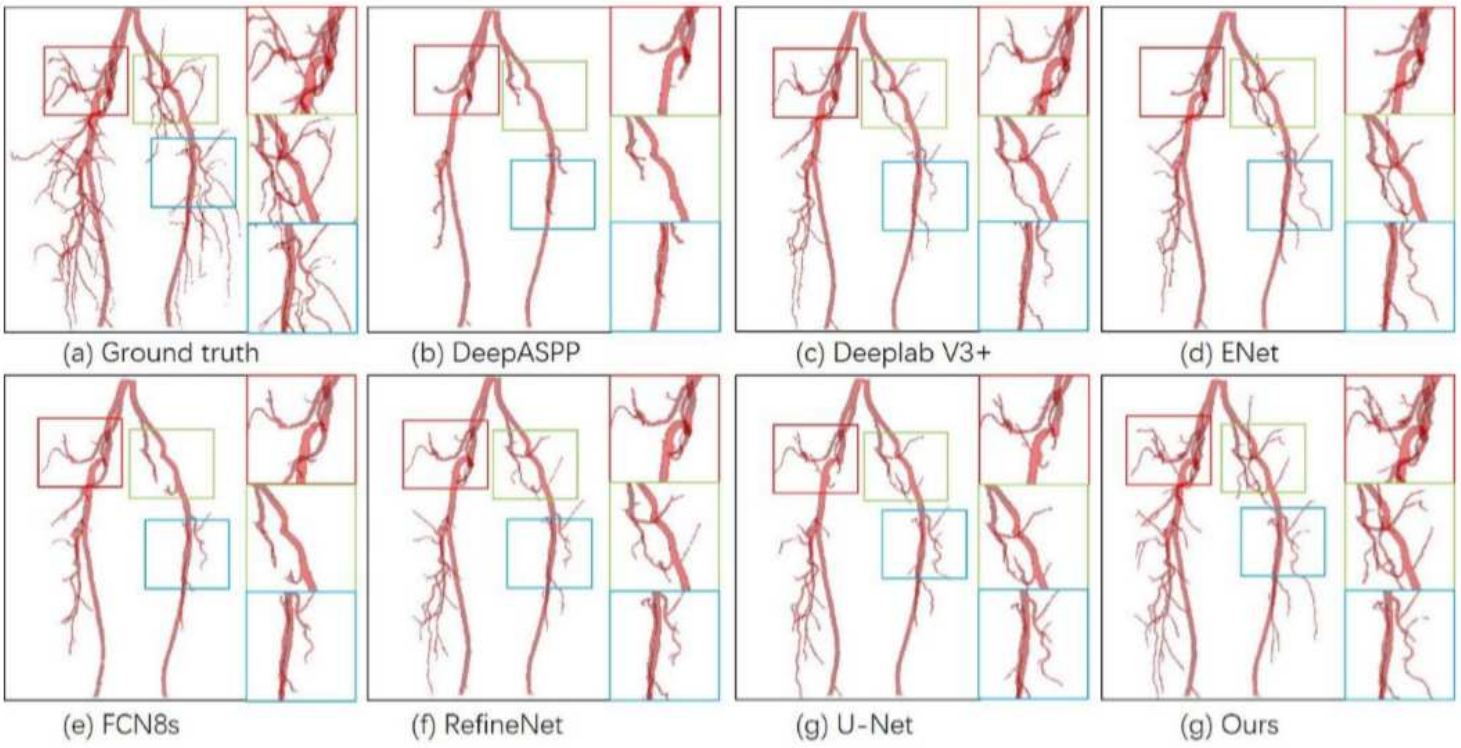


Figure 11

Comparative visualization of 3D results achieved on the test dataset1 and test datset2. Compared to the ground truth, all the four state-of-the-art methods (DeepASPP, DeepLap V3+, ENet, FCN8s, RefineNet and U-Net) miss fine features (e.g. small vessels in rectangle), whereas the proposed method preserves fine vessels well.

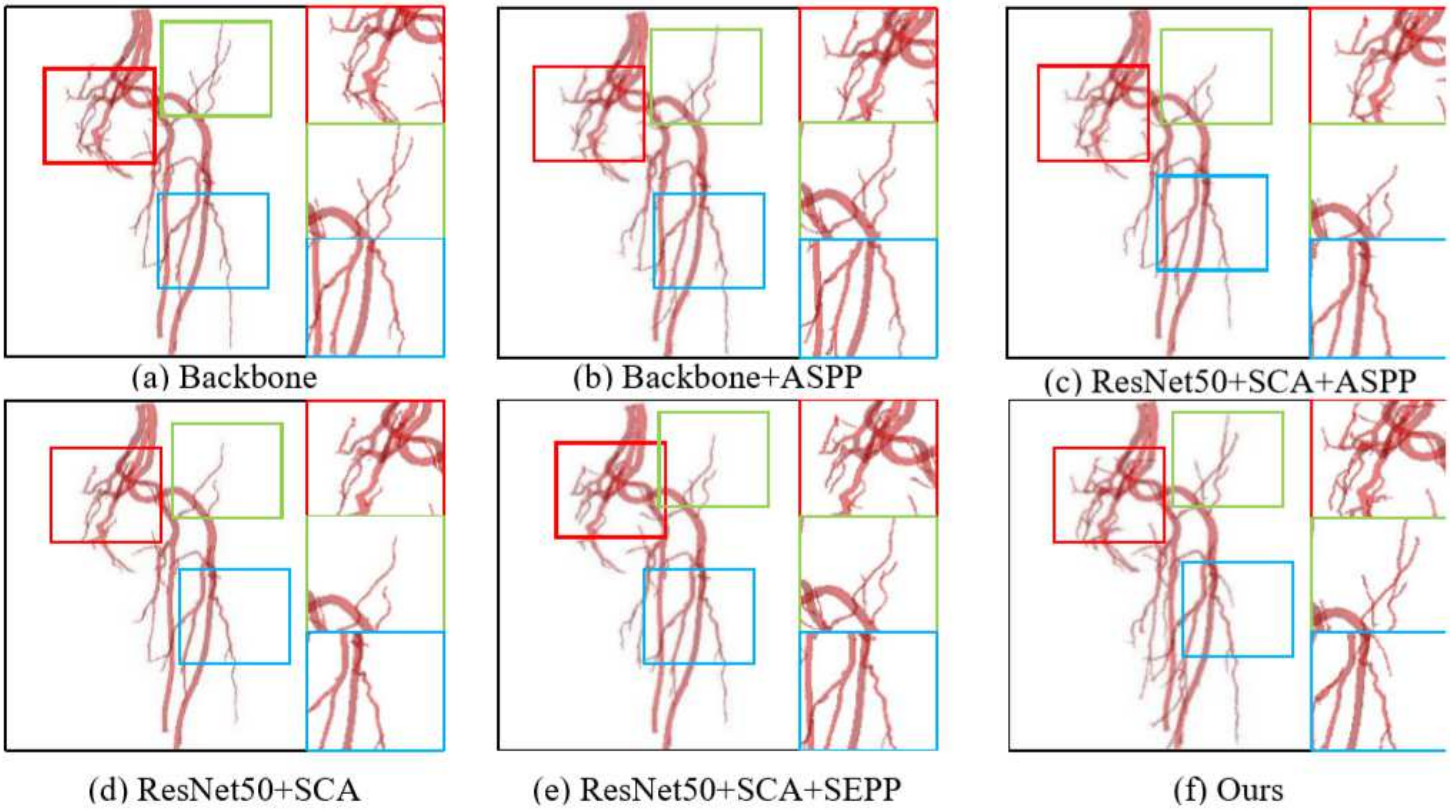


Figure 12

Medical image segmentation results tested in the dataset of leg artery. From (a)-(f): (a) Backbone, (b)Backbone+ASPP,(c)ResNet50+SCA+ASPP,(d)ResNet50+SCA,(e)ResNet50+SCA+SEP P, (f) SSCA-Net respectively. Our SSCA-Net can perform segmentation of intracranial arteries effectively while preserving more vessel details.

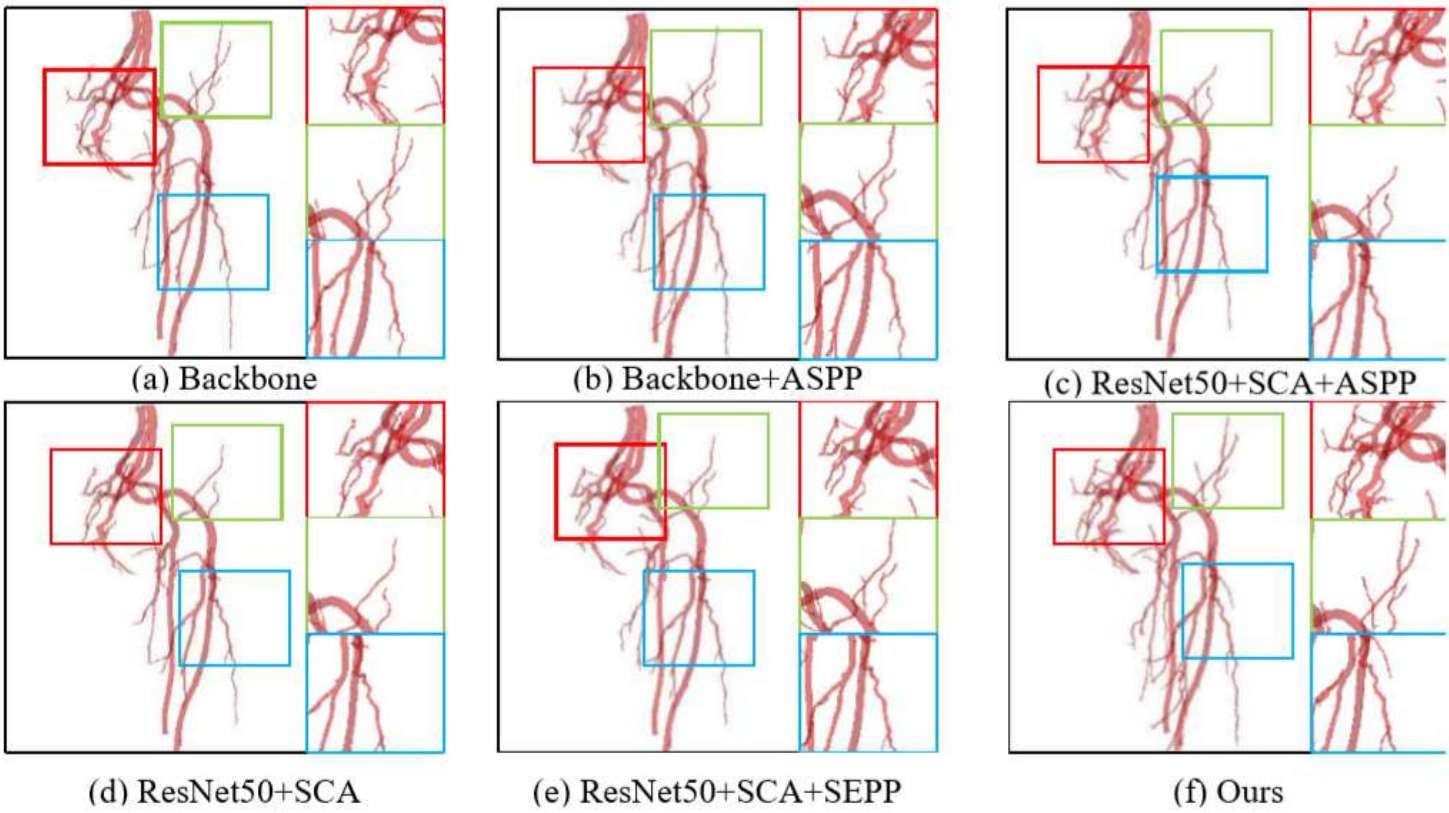


Figure 12

Medical image segmentation results tested in the dataset of leg artery. From (a)-(f): (a) Backbone, (b)Backbone+ASPP,(c)ResNet50+SCA+ASPP,(d)ResNet50+SCA,(e)ResNet50+SCA+SEP P, (f) SSCA-Net respectively. Our SSCA-Net can perform segmentation of intracranial arteries effectively while preserving more vessel details.

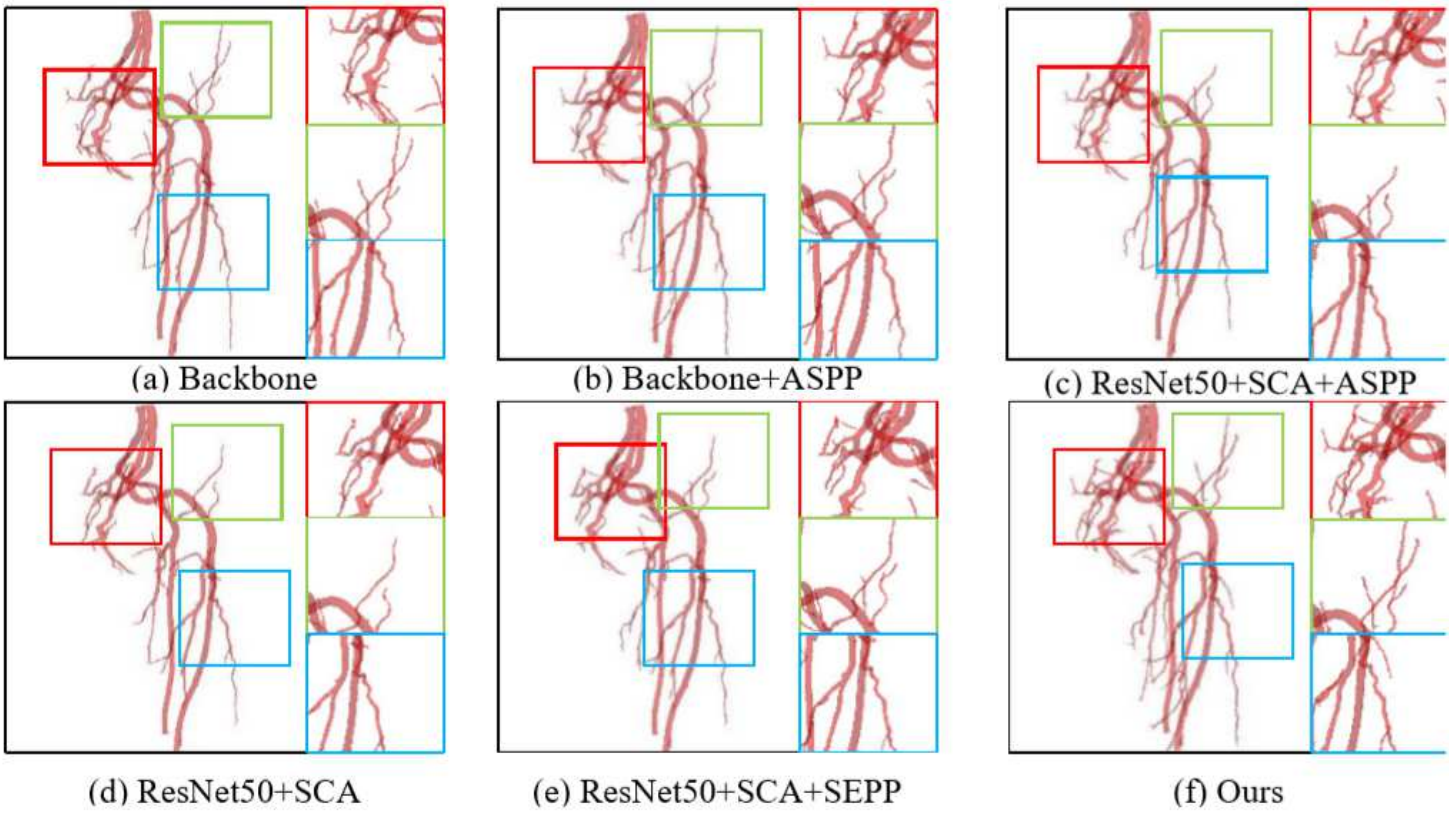


Figure 12

Medical image segmentation results tested in the dataset of leg artery. From (a)-(f): (a) Backbone, (b)Backbone+ASPP,(c)ResNet50+SCA+ASPP,(d)ResNet50+SCA,(e)ResNet50+SCA+SEP P, (f) SSCA-Net respectively. Our SSCA-Net can perform segmentation of intracranial arteries effectively while preserving more vessel details.

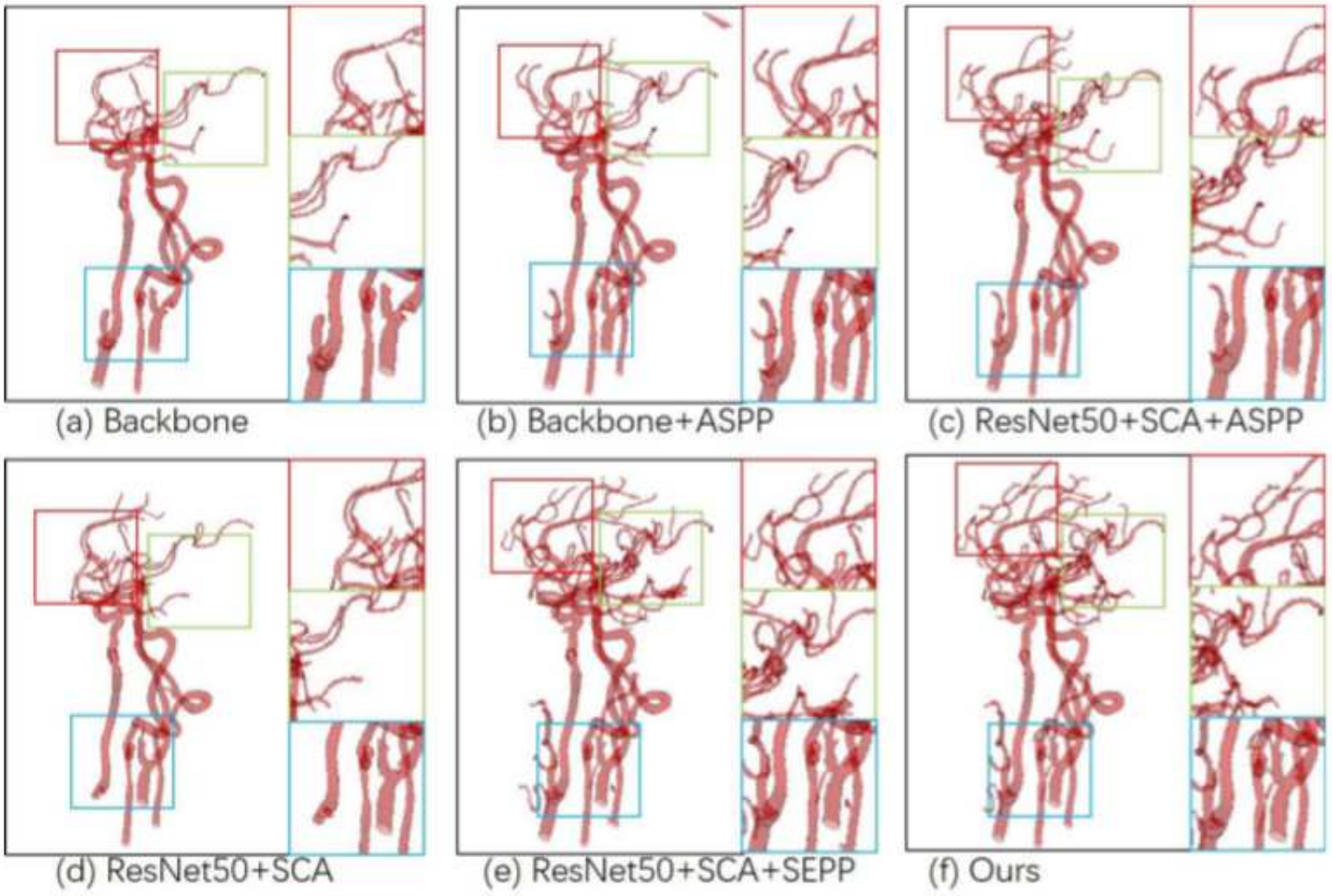


Figure 13

Medical image segmentation results tested in the dataset of intracranial artery. From (a)-(f): (a) Backbone,(b)Backbone+ASPP, (c) ResNet50+SCA+ASPP, (d) ResNet50+SCA, (e) ResNet50+SCA+SEPP, (f) SSCA-Net. respectively. Our SSCA-Net can segment intracranial arteries effectively while preserving more vessel details. (

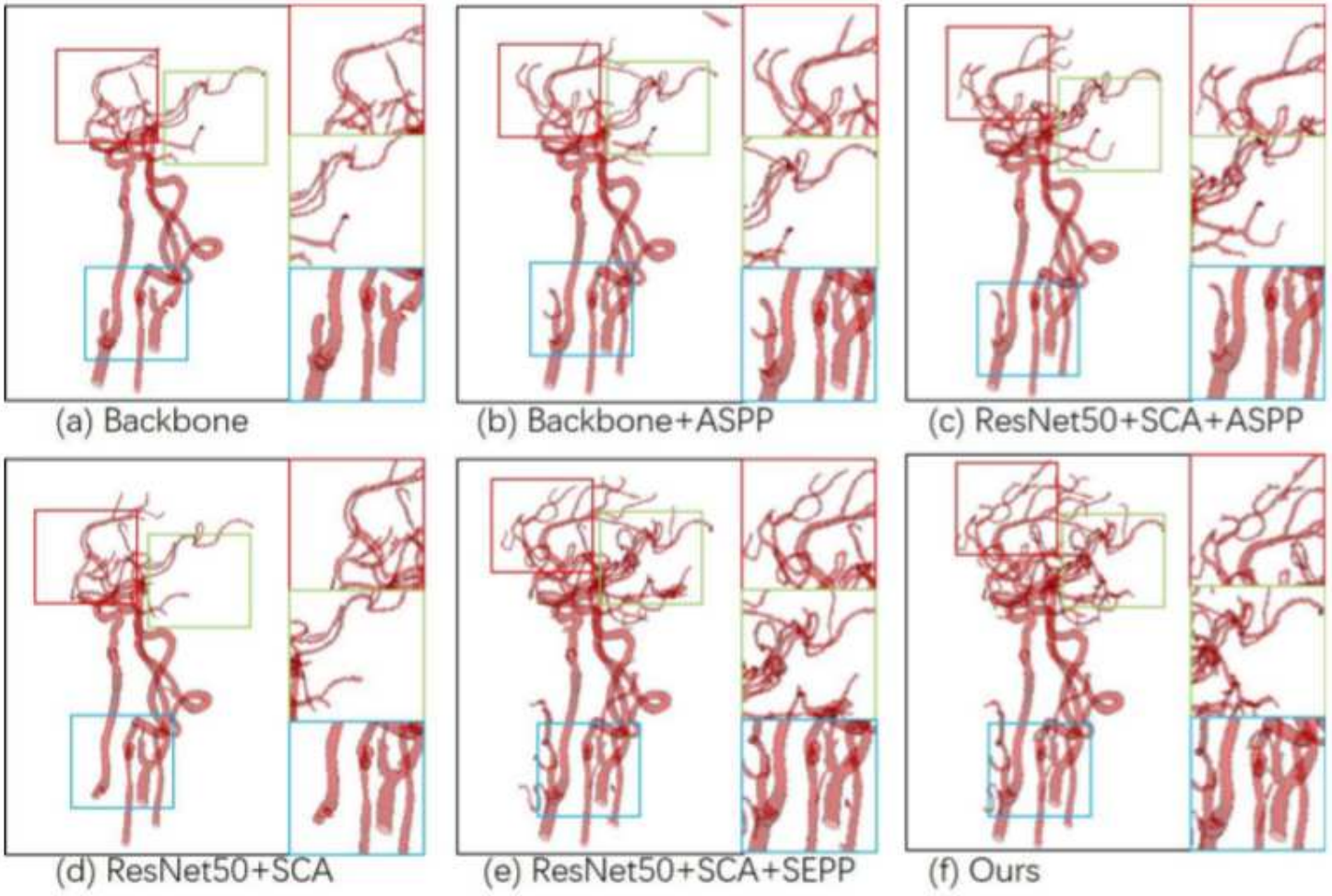


Figure 13

Medical image segmentation results tested in the dataset of intracranial artery. From (a)-(f): (a) Backbone,(b)Backbone+ASPP, (c) ResNet50+SCA+ASPP, (d) ResNet50+SCA, (e) ResNet50+SCA+SEPP, (f) SSCA-Net. respectively. Our SSCA-Net can segment intracranial arteries effectively while preserving more vessel details. (

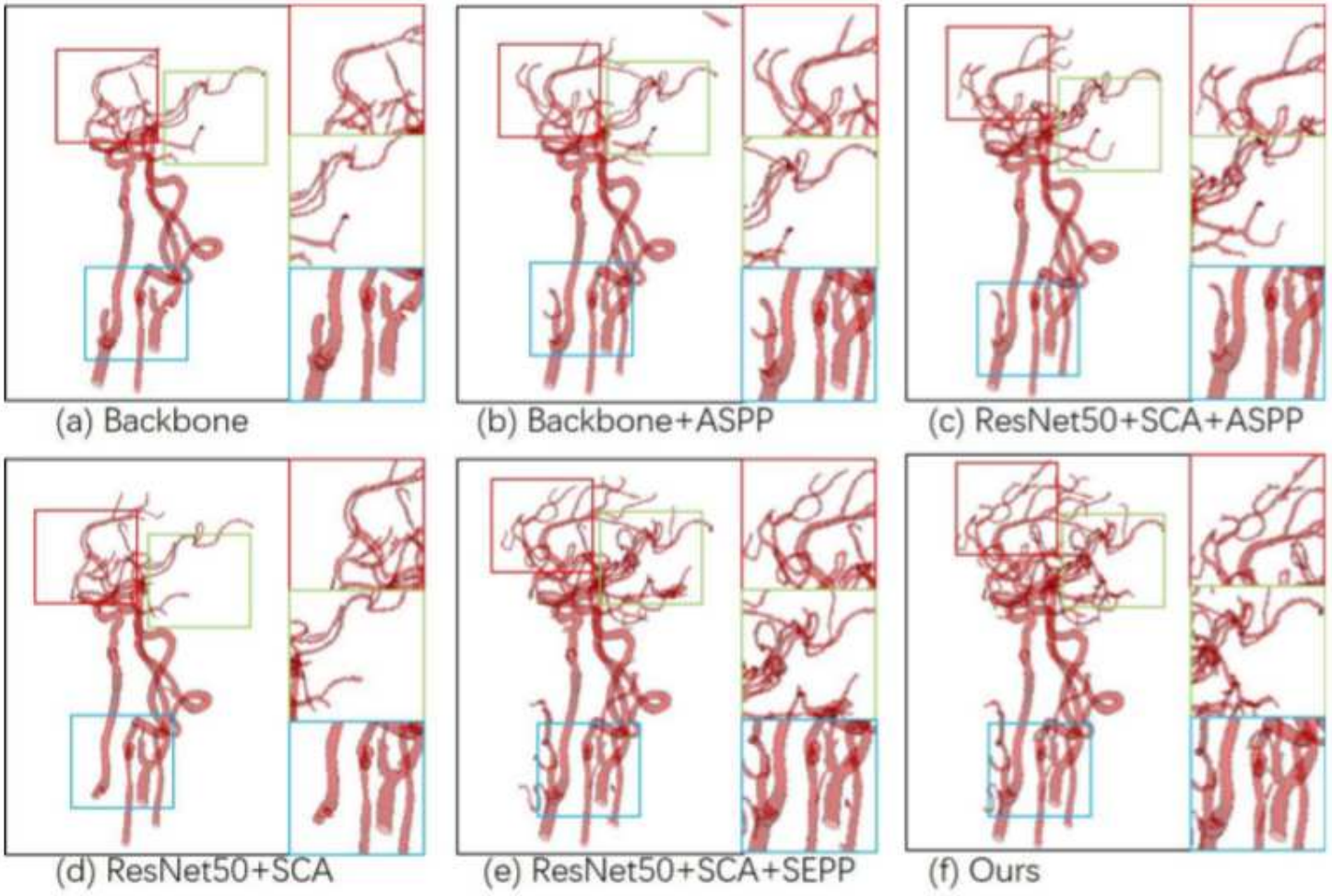


Figure 13

Medical image segmentation results tested in the dataset of intracranial artery. From (a)-(f): (a) Backbone,(b)Backbone+ASPP, (c) ResNet50+SCA+ASPP, (d) ResNet50+SCA, (e) ResNet50+SCA+SEPP, (f) SSCA-Net. respectively. Our SSCA-Net can segment intracranial arteries effectively while preserving more vessel details. (