Article type   : Research Article

Vessel Segmentation from Volumetric Images: A Multi-scale Double-pathway Network with Class-balanced Loss at the Voxel Level

Running Title: Vessel segmentation in 3D medical images

Yibing Chen, Siqi Fan, Yongfeng Chen, Chang Che, Xin Cao, Xiaowei He, Xiaolei Song, and Fengjun Zhao

Xi'an Key Lab of Radiomics and Intelligent Perception, School of Information Sciences and Technology, Northwest University, Xi'an, Shaanxi 710069, China

Corresponding authors: Xiaolei Song (Address: School of Information Sciences and Technology, Northwest University, No. 1 Xuefu Avenue, Xi'an, Shaanxi 710127, China; E-mail: xlsong8@163.com), and Fengjun Zhao (Address: School of Information Sciences and Technology, Northwest University, No. 1 Xuefu Avenue, Xi'an, Shaanxi 710127, China; E-mail: fjzhao@nwu.edu.cn)

Abstract

Purpose: Vessel segmentation from volumetric medical images is becoming an essential pre-step in aiding the diagnosis, guiding the therapy and patient management for vascular-related diseases. Deep learning-based methods have drawn many attentions, but most of them did not fully utilize the multi-scale spatial information of vessels. To address this shortcoming, we propose a multi-scale network similar to the well-known multi-scale DeepMedic. It also includes a double-pathway architecture and a class-balanced

loss at the voxel level (MDNet-Vb) to achieve both the computation efficiency and segmentation accuracy.

Methods: The proposed network consists two parallel pathways to learn the multi-scale vessel morphology. Specifically, the pathway with a normal resolution uses three-dimensional (3D) U-Net fed with small inputs to learn the local details with relatively small storage and time consumption. The pathway with a low resolution employs 3D fully convolutional network (FCN) fed with down-sampled large inputs to learn the overall spatial relationships between vessels and adjacent tissues, and the morphological information of large vessels. To cope with the class-imbalanced issue in vessel segmentation, we propose a class-balanced loss at the voxel level with uniform sampling strategy. The class-balanced loss at the voxel level re-balances the loss function with a coefficient that is inversely proportional to the normalized effective number at the voxel level of each class. The uniform sampling strategy extracts training data by sampling uniformly from two classes in every epoch.

Results: Our MDNet-Vb outperforms several state-of-the-art methods including ResNet, DenseNet, 3D U-Net, V-Net and DeepMedic with the highest dice coefficients of 72.91% and 69.32% on cardiac computed tomography angiography (CTA) dataset and cerebral magnetic resonance angiography (MRA) dataset, respectively. Amongst four different double-pathway networks, our network (3D U-Net+3D FCN) not only has the fewest training parameters and shortest training time, but also gets competitive dice coefficients on both the CTA and MRA datasets. Compared with classical losses, our class-balanced focal loss (FL-Vb) and dice coefficient loss at the voxel level (Dsc-Vb) alleviates class imbalanced issue by improving both the sensitivity and dice coefficient on the CTA and MRA datasets. Moreover, simultaneously training on two datasets shows that our method has the highest dice coefficient of 73.06% and 65.40% on CTA and MRA datasets respectively, outperforming the commonly used methods, such as U-Net and DeepMedic, which demonstrates the generalization potential of our network for segmenting different blood vessels.

Conclusions: Our MDNet-Vb method demonstrates its superiority over other state-of-the-art methods, on both cardiac CTA and cerebral MRA datasets. For the network architecture, the MDNet-Vb combined the 3D U-Net and 3D FCN, which dramatically reduces the network parameters yet maintains the segmentation accuracy. The class-balanced loss at the voxel level further improves accuracy by properly alleviating the class-imbalanced issue between different classes. In summary, MDNet-Vb is promising for vessel segmentation from various volumetric medical images.

Keywords: Vessel segmentation, deep leaning, fully convolutional network, computed tomography angiography, magnetic resonance angiography

1. Introduction

Precise segmentation of blood vessels from volumetric medical images has an important role in examining vascular morphology, calculating hemodynamics, and diagnosing vascular-related diseases.[1] Manual delineation of blood vessels is a tedious and time-consuming process during which experts need to mark tiny vessels slice by slice. Over the last decade, a number of automatic methods have been proposed to segment vessels from volumetric medical images. These procedures can be categorized into rule-based methods and machine-learning-based methods.[2-4] The rule-based methods mainly include the Hessian matrix,[5-8] mathematical morphology,[9-11] and minimal cost path,[12-15] etc. The learning-based methods consist of conventional machine learning approaches,[16-18] such as support vector machines (SVM) and random forests, and deep learning methods. It is well known that most rule-based methods and conventional machine learning methods require the exquisite design of rule sets or quantitative features, which heavily depend on the user's domain knowledge.

Compared with conventional methods, deep learning methods can directly extract vessel features from raw medical images due to the efficient representation of complex image features through multiple layers of learning. For example, Kitrungrotsakul et al. [19] proposed a multi-view convolutional neural network (CNN) for hepatic vessel

segmentation. Haft-Javaherian et al. [20] proposed a CNN with a fully connected layer to segment 3D vessels within volumetric *in vivo* images acquired by multiphoton microscopy. Xu et al. [21] developed a stage-wise three-dimensional (3D) fully convolutional network (FCN) for accurate pulmonary vessel segmentation. However, all these methods do not consider multi-scale information to provide additional guidance for vessel segmentation.

The well-known DeepMedic,[22] a classical deep learning method for brain lesion segmentation, feeds input images of different scales into two parallel FCNs, allowing simultaneous learning of morphological details and the overall context information. However, the FCNs used in DeepMedic only includes the encoding module, which consists of convolutional layers, pooling layers, activation function and batch normalization. The size of outputs is smaller than that of inputs. The U-Net [23] encompasses both the encoding module and decoding module in the architecture to generate the output with the same size as the input, which could dramatically reduce the storage and time consumption compared to FCN. Based on these observations, in this study, we proposed a novel multi-scale double-pathway network to segment blood vessels from volumetric medical images (**Figure 1**). For the pathway with a normal resolution, a typical 3D U-Net fed with smaller-size images is used to learn the local vascular morphological details and reduce storage and time consumption. For the pathway with a low resolution, a 3D FCN fed with down-sampled larger-size images is employed to learn not only the overall spatial relationships between vessels and adjacent tissues, but also the morphological information of large vessels.

In addition to the network architecture, the performance of deep learning is vulnerable to the class-imbalance issue between different classes. Data re-sampling strategy (sample training data uniformly from each class) has been used to alleviate this issue[22]; yet, this method may lead to over-sampling from the minor class and under-sampling from the major class. Besides, loss re-balance strategy is also widely used to alleviate class-imbalance in classification,[24-27] object detection,[28] and semantic segmentation,[29-38] which assigns a weight to loss function in order to match a given data distribution.

For vessel segmentation, each voxel has to be classified as vessels or backgrounds, which causes a severe

class-imbalanced issue at the voxel level due to the sparse and thin vessel structures. To solve this issue, we integrated a class-balanced loss function at the voxel level with a uniform sampling strategy during the training process of the network. Inspired by the class-balanced loss at the image level,[37] the class-balanced loss function at the voxel level re-balanced loss functions based on the normalized effective number at the voxel level of each class during network training. For the sampling strategy, we sample training data uniformly from two classes in every epoch, instead of sampling only once as the traditional training process. This strategy efficiently increases the amount of training data, and ensures that the number of each class is almost balanced during each training epoch.

In summary, in this study, we proposed a novel multi-scale double-pathway network with class-balanced loss at the voxel level (MDNet-Vb) to segment blood vessels from volumetric medical images. The contributions of the present study are the following: (1) the proposed multi-scale double-pathway architecture can learn both the local vascular morphological details and overall context information and reduce the storage and time consumption. (2) To cope with the severe class-imbalanced issue at the voxel level in vessel segmentation, this new network is effectively trained end-to-end from scratch by using the proposed class-balanced loss function at the voxel level and uniform sampling strategy. (3) Experimental results on two datasets (cerebral magnetic resonance angiography (MRA) and cardiac computed tomography angiography (CTA) images) demonstrate the performance of the proposed method compared with several state-of-the-art methods. (4) Simultaneously training on the two datasets (MRA and CTA) shows that our method has the potential to be generalized to the segmentation of blood vessels from different volumetric images.

2. Materials and Methods

2.1 Materials

Two volumetric medical images datasets were included in this study as follows.

(1) *Cerebral vessels segmentation from magnetic resonance angiography (MRA) images*. Twenty-nine publicly

available MRA images from the MIDAS platform [39] were recruited for this task. The resolution of each image is 0.5×0.5×0.8 mm^3, and the size of them is 448×448×128. The provided vessel centerline and radius were used to generate the ground truth cerebral vessel (or label) for each MRA data [40]. We randomly partitioned this dataset into 12 training data, 3 validation data, and 14 testing data.

(2) *Coronary arteries segmentation from cardiac computed tomography angiography (CTA) images.* Twenty-seven cardiac CTA images were acquired with 120 KVp tube voltage and 55 mAs tube current from the Chinese PLA General Hospital. All the images were resampled to the same voxel size of 0.5×0.5×0.5 mm^3. Ground truth of coronary arteries were annotated by two experienced cardiologists with medical software MITK 2015.5.0. This dataset was randomly partitioned into 12 training data, 3 validation data, and 12 testing data.

2.2 Methods

2.2.1 Network architecture

Figure 1 depicts the architecture of the proposed MDNet-Vb network, which consists of two parallel pathways (i.e., 3D U-Net and FCN), followed by a fusion module (with two convolutional layers). Specifically, the pathway with the normal resolution follows the framework of 3D U-Net, encompassing an encoding module and a decoding module. In the encoding module, we first perform 3×3×3 convolution with the stride of 1 and zero padding, and then calculate the rectified linear unit (ReLU) activations and batch normalization (BN). Besides, successive 2×2×2 max pooling with the stride of 2 are performed to enlarge receptive fields after 3×3×3 convolution with the stride of 1 and zero padding. It can be observed that the inputs of this pathway are down-sampled two times to capture the global contextual information of vessels.

Symmetric to the encoding module, the feature maps of subsequent decoding module are up-sampled two times with de-convolutions to recovery spatial details. Specifically, the 2×2×2 de-convolutions are performed with the stride of 2, and then the same convolutional operations as those in the encoding module are performed. Furthermore, skip

connections are deployed to fuse the up-sampled feature maps with the feature maps of the same level obtained from the encoding module, and to complementarily combine global contextual information with spatial details.

The pathway with a low resolution adopts the architecture of the 3D FCN, consisting of a series of convolutional layers and an up-sampling layer. We first perform 3×3×3 convolution with the stride of 1 without zero padding, and then calculate ReLU and BN. Note that the inputs of this pathway are down-sampled to low resolution, so at the end of this pathway, an up-sampling operation implemented by nearest neighbor interpolation is used to restore the resolution of output feature maps.

The multi-scale features extracted from the two pathways are fused by two convolutional layers with the stride of 1, in which kernel size are 3×3×3 and 1×1×1, respectively. The first 3×3×3 convolutional layer helps combining the multi-scale features smoother [22]. The fully connected layer is replaced by a 1×1×1 convolutional layer to generate dense predictions. Finally, it is connected to a softmax function with temperature parameter to obtain the final segmentation results.

2.2.2 Multi-scale feature learning

As illustrated in **Figure 1**, our MDNet-Vb network simultaneously extracts multi-scale features from the multi-scale inputs in two pathways. For the normal-resolution pathway, the smaller-size inputs I_S are fed to 3D U-Net to capture the local vascular morphological details. If this pathway adopts FCN instead of U-Net, it will need larger inputs under the same output size because of the lack of up-sampled decoding module in the FCN, and lead to greater storage and time consumption.

When the inputs of U-Net are part of the whole volumetric images, the spatial context information learned by lager inputs is important for being able to discriminate voxels that otherwise appear very similar. Therefore, in order to learn the overall spatial relationships between vessels and adjacent tissues, the larger-size inputs I_L underwent down-sampling are fed to low-resolution pathway. For this pathway, a series of convolutional layers of FCN reduce

the spatial dimension of larger input to ensure that the size of output feature maps of two pathways are matched. To ensure that the output features of two pathways correspond to the same area of the volumetric image, the multi-scale inputs of two pathways should be extracted centered on the same image location. Moreover, according to the network architectures of two pathways, the relationship between the size of I_L and I_S can be formulated as:

$$\text{size}(I_L) = F \times [L \times (K - 1) + \text{size}(I_S)/F]$$ (1)

where $L \times (K - 1)$ indicates the reduced size by the convolutional operations of FCN. L and K denote the total number of convolutional layers and kernel size of those layers, respectively. F is the down-sampled factor used to reduce the resolution of the input in the low-resolution pathway.

2.2.3 class-balanced loss at the voxel level

Cui et al. [37] proposed an image-level class-balanced loss function to address the problem of long-tailed data distribution in natural image classification, in which they rebalanced the loss functions with a coefficient that was inversely proportional to the effective number of each class. The effective number of each class was calculated by the number of images belonging to each class based on random covering theory [41]. However, in this study, we have to classify each voxel instead of an image as vessels or backgrounds, which therefore will cause severe class-imbalanced issue at the voxel level due to the sparse and thin vessel structures (e.g., the ratio of the average number of background voxels to vessel voxels exceeds 900:1 on the CTA dataset). Therefore, inspired by the image-level class-balanced loss, we propose a class-balanced loss function at the voxel level to re-balance losses with a coefficient that is the inversely proportional to the effective number at the voxel level of each class. The effective number at the voxel level for class v can be written as:

$$E_{n_v} = \frac{1 - \beta_v^{n_v}}{1 - \beta_v}$$ (2)

$$\beta_v = \frac{N_v - 1}{N_v}$$ (3)

where n_v is the average number of voxels for class v in every mini batch, which reflects the distribution of the class v in the actual training process after data re-sampling. N_v is the number of all possible data in the feature space of the class v. Similar to the image-level class-balanced loss, we assume that for all classes in the dataset, N_v is a fixed number. According to the total number of voxels in each mini batch, N_v and β_v can be estimated as:

$$N_v = N = T \times N_b \tag{4}$$

$$\beta_v = \beta = \frac{N-1}{N} \tag{5}$$

where N_b denotes the total number of voxels in each mini batch. T is a non-zero hyper-parameter, which is utilized to bridge the gap between N_v and N_b. Moreover, to ensure the total loss roughly in the same scale when applying E_{n_v}, we introduce α_v that is equal to the normalized effective number at the voxel level of class v:

$$\alpha_v = \frac{E_{n_v}}{\Sigma_{v=0}^1 E_{n_v}} = \frac{(1-\beta^{n_v})/(1-\beta)}{\sum_{v=0}^1 \left[(1-\beta^{n_v})/(1-\beta)\right]} \tag{6}$$

The class-balanced loss at the voxel level (Vb) can be written as:

$$L_{Vb} = \frac{1}{\alpha_v} L_o \tag{7}$$

In practice, we choose the focal loss [28] and dice coefficient loss [34] as L_o to build our class-balanced loss function at the voxel level in the coronary artery and cerebral vessel segmentation task, respectively. Specifically, we assume y is the ground truth, $\hat{y}$ is the vessel probability map. The class-balanced cross-entropy loss at the voxel level is:

$$L_{Vb-crossEntropy} = \begin{cases} -\dfrac{1}{\alpha_v} \Sigma_{v=0}^1 \log(\hat{y}_v) & y_v = 1 \\ -\dfrac{1}{\alpha_v} \Sigma_{v=0}^1 \log(1-\hat{y}_v) & y_v = 0 \end{cases} \tag{8}$$

The class-balanced focal loss at the voxel level is:

$$L_{Vb-focal} = \begin{cases} -\dfrac{1}{\alpha_v} \Sigma_{v=0}^1 (1-\hat{y}_v)^\gamma \log(\hat{y}_v) & y_v = 1 \\ -\dfrac{1}{\alpha_v} \Sigma_{v=0}^1 \hat{y}_v^\gamma \log(1-\hat{y}_v) & y_v = 0 \end{cases} \tag{9}$$

where $\hat{y}_v^\gamma$ or $(1-\hat{y}_v)^\gamma$ is the modulating factor to automatically assign low weights to easy examples during training; γ adjusts the rate of the modulating factor.[28]

The class-balanced dice loss at the voxel level is:

$$L_{Vb-dice} = -\frac{1}{\alpha_v} \Sigma_{v=0}^1 \frac{2 \cdot y_v \cdot \hat{y}_v}{y_v + \hat{y}_v} \tag{10}$$

2.2.4 Uniform sampling strategy

In addition, we further integrate the uniform sampling strategy to correct the severe class-imbalanced issue. Instead of sampling only once in the training phase, we sample the training data uniformly from two classes in every epoch. This strategy can efficiently increase the number of training data, and ensure that the number of vascular and background voxels is balanced in every training epoch. Specifically, in every epoch, we randomly selected every voxel from vessels with equal probability, and a pair of cubes (normal- and low-resolution inputs) with this voxel as the center are extracted as the positive training data. Repeating this process $M/2$ times, we obtained $M/2$ pairs of positive training data. Similarly, an equal number of background training data are generated. After training with N epochs, a total of NM pairs of normal- and low-resolution inputs were generated from the volumetric medical images. Note that these training data are randomly sampled with replacement. In this study, referring to the DeepMedic[22], M was set to 1000 and N was set to 700 in the training phase. Therefore, the total number of input pairs is $NM = 700,000$.

2.2.5 Implementation

The proposed method was implemented using Python language and TensorFlow package [42] on the workstation with single graphics processing unit (NVIDIA GeForce GTX TITAN V). As mentioned in 2.2.2, the normal- and low-resolution inputs were concentric with the size of 36×36×36 and 84×84×84, respectively. The number of convolutional layers L and kernel size K were set to 8 and 3, respectively. The down-sampled factor F was set to 3. L_1 and L_2 regularization was employed to alleviate over-fitting. The batch size in each epoch was 8. The network was

trained by the RMSProp [43] optimizer with rho of 0.9 and e of 0.0001. The initial learning rate was set to 0.001. The search space of hyper-parameter T was $T \in \{0.01, 0.1, 1, 10, 100\}$, and the search space of parameter γ was $\gamma \in \{0.5, 1, 2\}$. The total number of voxels in each mini batch N_b was equal to $8\times36\times36\times36$. Note that we used both the training set and validation set to develop the proposed loss function. In the testing phase, we extracted non-overlapped multi-scale inputs from each testing image in order, and fed them into the trained network to generate the concentric $36\times36\times36$ likelihood maps. Then we mosaicked those likelihood maps to form the segmentation result for each test image. To alleviate the impact of random initialization during network training, we ran our MDNet-Vb network and other methods three times, and took the average value as the final results.

2.2.6 Evaluation Metrics

Using manual annotations as ground truth, the segmentation performance of our method was quantitatively evaluated with the following four metrics: (1) sensitivity (SEN), (2) specificity (SPE), (3) dice coefficient (DSC) and (4) Hausdorff distance (HD), defined as

$$SEN = TP/(TP + FN) \tag{11}$$

$$SPE = TN/(TN + FP) \tag{12}$$

$$DSC = 2TP/(2TP + FP + FN) \tag{13}$$

$$HD = \max\left\{\vec{d}_{HD}(A,B), \vec{d}_{HD}(B,A)\right\} \tag{14}$$

$$\vec{d}_{HD}(A,B) = \max_{x\in A}\left\{\min_{y\in B}\{d(x,y)\}\right\} \tag{15}$$

where TP and FP denote the numbers of true positives and false positives respectively, while TN and FN are the numbers of true negatives and false negatives respectively. HD is maximum distance between two voxel sets, in which $\vec{d}_{HD}(A,B)$ is the directed Hausdorff distance of ground truth A and prediction B, where $d(x,y)$ represents the Euclidean distance between two voxels.

To compare the statistical difference between two groups, we first calculated the *p*-values of all experiments using two-tailed paired t-test, and then used Holm-Bonferroni Method[44] (also called Holm's Sequential Bonferroni Procedure) for correcting the *p*-values. Specifically, the Holm-Bonferroni Method first sorts the original *p*-values from small to large, and then multiplies the original *p*-values by the corresponding order. Finally, the corrected *p*-values of the adjacent sequences are equal to the larger one between them. When the corrected *p*-values are less than 0.05, it represents there is a statistical difference between two groups.

3. Results

3.1 Comparison to state-of-the-art methods

The segmentation results on both the CTA and MRA dataset obtained by our method using the MDNet-Vb network, and the other five state-of-the-art methods including ResNet [45], DenseNet [46], 3D U-Net [23], V-Net [34], and DeepMedic [22] are shown in **Figure 2** and **3**, and **Table 1** and **2**. According to the experimental results of the validation set, the hyper-parameters T and γ of our loss function were set to 0.01 and 1, respectively. As can be seen from visualization results, the automatic segmentations obtained by the proposed MDNet-Vb were more consistent with the manual ground truth in these examples, especially for the relatively low-contrast blood vessels.

From **Table 1** and **2**, we concluded the following: firstly, compared with state-of-the-art methods, the proposed MDNet-Vb had the highest dice coefficients (72.91% and 69.32% on CTA and MRA datasets, respectively), which was significantly higher than others ($p<0.05$) excepted for DeepMedic ($p>0.05$) on the CTA dataset, and significantly higher than others ($p<0.05$) excepted for DeepMedic and 3D U-Net ($p>0.05$) on the MRA dataset. Secondly, the proposed MDNet-Vb efficiently reduced the Hausdorff distance to 17.52 voxels on the CTA dataset. Thirdly, compared with the well-known DeepMedic, our MDNet-Vb improved the sensitivity and dice coefficient on CTA/MRA dataset by 0.99% ($p>0.05$) /3.14% ($p<0.05$) and 0.27% ($p>0.05$) /0.68% ($p>0.05$), respectively, and reduced Hausdorff distance by 1.63 ($p>0.05$) /0.01 ($p>0.05$) voxels.

To evaluate the generalization ability, we simultaneously trained the MDNet-Vb with both MRA and CTA datasets. DeepMedic and 3D U-Net are two widely used segmentation networks, and had good performance in **Table 1** and **2**. Therefore, to reduce the time consumption, we only compared the proposed network with DeepMedic and 3D U-Net in this part. Compared to DeepMedic and 3D U-Net, our MDNet-Vb had a good performance in the segmentation of thick blood vessels on CTA and MRA (**Figure 4**). As indicated in **Table 3** and **4**, compared with 3D U-Net and DeepMedic, our MDNet-Vb had the highest dice coefficient values (73.06% and 65.40% on CTA and MRA datasets, respectively), which was significantly higher than others ($p<0.05$) on the CTA dataset, and significantly higher than others ($p<0.05$) excepted for DeepMedic ($p>0.05$) on the MRA dataset. It also achieved the significant lowest Hausdorff distance (21.02 voxels, $p<0.05$) on the CTA dataset and the significant highest sensitivity (59.37%, $p<0.05$) on the MRA dataset. These data suggest that the new MDNet-Vb may potentially a generalized 3D vessel segmentation approach.

3.2 Evaluation of network architectures

To assess the effectiveness of our multi-scale double-pathway network (MDNet-Vb) architecture, we performed the four possible combinations of the 3D FCN and 3D U-Net. The setting and the number of training parameters of the four architectures are shown in **Table 5**. The visualization results on two datasets are presented in **Figure 5**, and the corresponding quantitative results are shown in **Table 6** and **7**. After evaluating the results, we concluded the following: (1) compared with other models, our architecture (U-Net+FCN) had the fewest training parameters (~1.53 million) and the shortest training time (~42.22 hours and ~33.06 hours on CTA and MRA datasets, respectively). (2) The performance of our architecture was similar to ($p>0.05$) that of the architecture of the DeepMedic (FCN+FCN), but we reduced the parameters by ~0.15 million and saved the computation time by ~18.69 hours and ~11.08 hours on CTA and MRA datasets, respectively. (3) It can be seen from the visualization results, the architecture of FCN+U-Net was difficult to segment thin coronary blood vessels of the arteries and it also tended to generate dotted false positive

predictions on the MRA dataset. From **Table 6** and **7**, the architecture of FCN+U-Net had the significant lowest sensitivity ($p<0.05$), and significant highest dice coefficient ($p<0.05$) and the longest time consumption.

To further evaluate whether it would be beneficial to increase the amount of down-sampling, we compared the double-pathway networks with different down-sampling factors (F=3, 5, 7), and the results are given in **Table 8** and **9**. It is observed from the results that the larger down-sampling factors led the lower dice coefficients (F=5, 7), compared to smaller down-sampling factor (F=3). This may be due to that the larger F results in lower resolution, and the morphological information of medium to large vessels is weakened.

3.3 Evaluation of different loss functions

We compared class-balanced loss functions at the voxel level including cross-entropy (CE-Vb), focal loss (FL-Vb), and dice coefficient loss (Dsc-Vb) with their corresponding classical loss, i.e., CE, FL[28], and Dsc [34] (**Table 10** and **11**). The quantitative results indicated that our class-balanced loss functions at the voxel level consistently improved the sensitivity ($p>0.05$ and $p<0.05$ on CTA and MRA dataset, respectively) and dice coefficient ($p>0.05$ on both datasets) compared with three corresponding classical loss functions on both CTA and MRA datasets and decreased the Hausdorff distance ($p>0.05$) on the CTA dataset. Moreover, FL-Vb had the significant highest sensitivity (76.46%, $p<0.05$) and highest dice coefficient (72.58%, $p>0.05$) on the CTA dataset, while Dsc-Vb had the highest sensitivity (65.28%) and dice coefficient (69.06%) on the MRA dataset, which were significantly higher than others ($p<0.05$) excepted for CE-Vb ($p>0.05$). According to the dice coefficient, we chose FL-Vb and Dsc-Vb as our training losses on CTA and MRA datasets, respectively.

3.4 Evaluation of different sampling strategies

We evaluated the proposed sampling strategy, which uniformly extracted training data in every epoch, and compared these data with the traditional sampling strategy that uniformly extracted training data only once. As shown in **Table 12** and **13**, our sampling strategy significant outperformed the traditional strategy in terms of dice coefficient and

Hausdorff distance on both CTA and MRA dataset ($p<0.05$). Compared with the traditional strategy, our sampling strategy increased the dice coefficient by 4.05% and dramatically decreased Hausdorff distance by 11.38 voxels on the CTA dataset, and improved the dice coefficient by 2.91% and reduced Hausdorff distance by 0.18 voxels on the MRA dataset.

3.5 Evaluation of data augmentation

Both rotation and flipping are common data augmentation tricks during the training phase of deep neural networks. Therefore, we rotated and flipped the training data with a probability of 0.5, and the rotation angles were 45, 90, 180 and 270, respectively; the flip direction was along the first dimension of the images. After data augmentation, the cubic bilinear interpolation was used. We evaluated the effectiveness of rotation and flipping on CAT and MRA datasets **(Table 14** and **15)**. Compared to the results without data augmentation, the results with data augmentation significantly improved the sensitivity by 5.36% and 2.32% on CTA and MRA datasets respectively ($p<0.05$), but reduced the specificity on both datasets. It also increased the Hausdorff distance by 9.95 and 0.35 voxels on CTA and MRA datasets, respectively. According to the dice coefficient, we adopted the data augmentation on the MRA dataset, but we directly used the non-augmented data to train the proposed network on the CTA dataset.

4. Discussion

Our MDNet-Vb simultaneously learns both local morphological details and overall context information from multi-scale inputs. The results presented in **Table 1** and **2** have shown that our method has higher dice coefficient compared with single input networks (ResNet, DenseNet, 3D U-Net, V-Net). Amongst different multi-scale networks, our network has the fewest parameters and the shortest training time without sacrificing segmentation accuracy. The reason is that we use 3D U-Net to minimize the input size of the normal-resolution pathway, and use 3D FCN to minimize the output size of the low-resolution pathway (**Table 5**), which reduces the number of double-pathway parameters and training time.

Our class-imbalanced loss at the voxel level assigns higher weights to vessel voxels according to data distribution during the training process, and makes the loss higher when the vessel voxels are misclassified, thus it achieves higher sensitivity and dice coefficient compared with the classical loss function (**Table 10** and **11**). We also observe that the highest dice coefficients on the CTA and MRA datasets are obtained by FL-Vb and Dsc-Vb losses, respectively. Correspondingly, amongst the three classical losses, the highest dice coefficients on CTA and MRA datasets are obtained by FL and Dsc losses, respectively. These consistent experimental results may indicate that there are hard and easy examples in the CTA dataset, while there may not be obvious hard and easy examples in the MRA dataset. In theory, the FL loss studies the hard example mining, and automatically assigns higher weights to hard examples during training[28], while the Dsc loss directly optimizes dice coefficient metric and do not establishes the right balance between hard and easy examples. [34]

The data augmentation (rotation and flipping) improves dice coefficient on the MRA dataset, but it does not apply to the CTA dataset. The possible reason of this inconsistent results may be the morphological difference between two datasets. Zhang et al. [47] mentioned that data augmentation is dataset-dependent, and Cubuk et al. [48] indicated that horizontal flipping of images improved the performance on CIFAR-10, but it is not the case on MNIST due to the different symmetries present in these datasets. The segmentation tasks are actually to classify pixels according to the neighborhood information of each pixel, and the rotation and flipping may change the neighborhood information of pixels. From the visual observation of CTA and MRA, we find that cerebral vessels are approximately symmetrical layout, and coronary arteries are not.

The experimental results of generalizability have shown that our method obtains highest dice coefficients on two datasets, compared with well-known 3D U-Net and DeepMedic (**Table 3** and **4**). This may benefit from the combination of our multi-scale feature learning and class-balanced loss at the voxel level. Firstly, our double-pathway network can effectively learn multi-scale features from the different scales of cerebral vessel and coronary artery. It can also be seen from **Table 4** that the dice coefficients of DeepMedic and our network are both nearly 7% higher

than single-pathway 3D U-Net. Secondly, class-balanced loss at the voxel level further improves the performance of our method by alleviating class-imbalance. From the **Table 6** and **1**, we can find that our method improves the dice coefficient after adding class-balanced loss at the voxel level, so that it performs better than DeepMedic on the CTA dataset in **Table 1**.

Although the proposed MDNet-Vb obtained competitive segmentation performance compared with several state-of-the-art methods, there are still some shortcomings that need to be improved. (1) The proposed method has lower dice coefficient on finer cerebral vessel segmentation task. One direct way to overcome such difficulty is to combine more advanced blocks (attention blocks, dilated dense blocks and so on) with current network to pay more attention to these small objects. (2) The proposed method may produce some false positive predictions that do not connect with other blood vessels. We may need to add some connectivity constrains to guide the training of our network. (3) The rotation and flipping are not suitable for CTA dataset. We may need to find a generalized automatically learned data-augmentation methods, or develop our method to a semi-supervised method to improve accuracy further. (4) The performance of our method may not be sufficient for clinical application because of the relatively low metrics. This may be caused by the small datasets in this study, where the CTA and MRA datasets only contain 27 and 29 images, respectively. We need to collect more data to further improve the accuracy of our method and to evaluate it more comprehensively in the future.

5. Conclusions

In this study, we have proposed a multi-scale double-pathway network with class-balanced loss at the voxel level (MDNet-Vb) to automatically segment blood vessels from volumetric medical images. The double-pathway network can learn both the local morphological details and the overall context information from multi-scale inputs, while the class-balanced loss function at the voxel level alleviates the severe class-imbalanced issue at the voxel level between different classes. The results demonstrate that the proposed method outperforms state-of-the-arts with highest dice

coefficients on two datasets. Moreover, simultaneously training on two datasets shows that our method has potential to be generalized to segment different blood vessels.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Data Availability Statement

The MRA brain images that support the findings of this study are openly available in TubeTK at https://vtk.org/Wiki/TubeTK/Data. The CTA cardiac images that support the findings of this study are available from the corresponding author upon reasonable request.

References

1. Moccia S, De Momi E, El Hadji S, Mattos LS. Blood vessel segmentation algorithms - Review of methods, datasets and evaluation metrics. *Computer Methods and Programs in Biomedicine.* 2018;158:71-91.

2. Zhao F, Chen Y, Hou Y, He X. Segmentation of blood vessels using rule-based and machine-learning-based methods: a review. *Multimedia Systems.* 2019;25(2):109-118.

3. Bibiloni P, Gonzalez-Hidalgo M, Massanet S. A survey on curvilinear object segmentation in multiple applications. *Pattern Recognition.* 2016;60:949-970.

4. Lesage D, Angelini ED, Bloch I, Funka-Lea G. A review of 3D vessel lumen segmentation techniques: Models, features and extraction schemes. *Medical Image Analysis.* 2009;13(6):819-845.

5. Frangi R, Niessen WJ, Vincken K, Viergever M. Multiscale Vessel Enhancement Filtering. *Med Image Comput Comput Assist Interv.* 2000;1496.

6. Xiao C, Staring M, Wang Y, Shamonin DP, Stoel BC. Multiscale Bi-Gaussian Filter for Adjacent Curvilinear Structures Detection With Application to Vasculature Images. *IEEE Transactions on Image Processing.* 2013;22(1):174-188.

7. Krissian K, Malandain G, Ayache N, Vaillant R, Trousset Y. Model-based detection of tubular structures in 3D images. *Computer Vision and Image Understanding.* 2000;80(2):130-171.

8. Zhao F, Liang J, Chen D, et al. Automatic segmentation method for bone and blood vessel in murine hindlimb. *Medical Physics.* 2015;42(7):4043-4054.

9. Wang R, Li C, Wang J, et al. Threshold segmentation algorithm for automatic extraction of cerebral vessels from brain magnetic resonance angiography images. *Journal of Neuroscience Methods.* 2015;241:30-36.

10. Passat N, Ronse C, Baruthio J, Armspach JP, Foucher J. Watershed and multimodal data for brain vessel segmentation: Application to the superior sagittal sinus. *Image and Vision Computing.* 2007;25(4):512-521.

11. Bouraoui B, Ronse C, Baruthio J, Passat N, Germain P. 3D segmentation of coronary arteries based on advanced mathematical morphology techniques. *Computerized Medical Imaging and Graphics.* 2010;34(5):377-387.

12. Mohan V, Sundaramoorthi G, Tannenbaum A. Tubular Surface Segmentation for Extracting Anatomical Structures From Medical Imagery. *IEEE Transactions on Medical Imaging.* 2010;29(12):1945-1958.

13. Benmansour F, Cohen LD. Tubular Structure Segmentation Based on Minimal Path Method and Anisotropic Enhancement. *International Journal of Computer Vision.* 2011;92(2):192-210.

14. Cetin S, Demir A, Yezzi A, Degertekin M, Unal G. Vessel Tractography Using an Intensity Based Tensor Model With Branch Detection. *IEEE Transactions on Medical Imaging.* 2013;32(2):348-363.

15. Forkert ND, Schmidt-Richberg A, Fiehler J, et al. 3D cerebrovascular segmentation combining fuzzy vessel enhancement and level-sets with anisotropic energy weights. *Magnetic Resonance Imaging.* 2013;31(2):262-271.

16. Ochs RA, Goldin JG, Abtin F, et al. Automated classification of lung bronchovascular anatomy in CT using AdaBoost. *Medical Image Analysis.* 2007;11(3):315-324.

17. Zhao B, Cao Z, Wang S. Lung vessel segmentation based on random forests. *Electronics Letters.* 2017;53(4).

18. Xin H, Yuanzhi C, Deqiong D, Dianhui C. Axis-Guided Vessel Segmentation Using a Self-Constructing Cascade-AdaBoost-SVM Classifier. *BioMed Research International.* 2018;2018:1-12.

19. Kitrungrotsakul T, Han X-H, Iwamoto Y, et al. VesselNet: A deep convolutional neural network with multi pathways for robust hepatic vessel segmentation. *Computerized Medical Imaging and Graphics.* 2019;75:74-83.

20. Haft-Javaherian M, Fang L, Muse V, Schaffer CB, Nishimura N, Sabuncu MR. Deep convolutional neural networks for segmenting 3D in vivo multiphoton images of vasculature in Alzheimer disease mouse models. *Plos One.* 2019;14(3).

21. Xu Y, Mao Z, Liu C, Wang B. Pulmonary Vessel Segmentation via Stage-Wise Convolutional Networks With Orientation-Based Region Growing Optimization. *IEEE Access.* 2018;6:71296-71305.

22. Kamnitsas K, Ledig C, Newcombe VFJ, et al. Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation. *Medical Image Analysis.* 2017;36:61-78.

23. Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation. Paper presented at: Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015; 2015//, 2015; Cham.

24. Mahajan D, Girshick R, Ramanathan V, et al. Exploring the Limits of Weakly Supervised Pretraining. Paper presented at: Computer Vision – ECCV 2018; 2018//, 2018; Cham.

25. Huang C, Li Y, Loy CC, Tang X. Learning Deep Representation for Imbalanced Classification. Paper presented at: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 27-30 June 2016, 2016.

26. Koh PW, Liang P. Understanding Black-box Predictions via Influence Functions. Proceedings of the 34th International Conference on Machine Learning; 2017; Proceedings of Machine Learning Research.

27. Ren M, Zeng W, Yang B, Urtasun R. Learning to Reweight Examples for Robust Deep Learning. Proceedings of the 35th International Conference on Machine Learning; 2018; Proceedings of Machine Learning Research.

28. Lin T, Goyal P, Girshick R, He K, Dollár P. Focal Loss for Dense Object Detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence.* 2020;42(2):318-327.

29. Cai J, Lu L, Xie Y, Xing F, Yang L. Pancreas Segmentation in MRI Using Graph-Based Decision Fusion on Convolutional Neural Networks. Paper presented at: Medical Image Computing and Computer Assisted Intervention − MICCAI 2017; 2017//, 2017; Cham.

30. Rahman MA, Wang Y. Optimizing Intersection-Over-Union in Deep Neural Networks for Image Segmentation. Paper presented at: Advances in Visual Computing; 2016//, 2016; Cham.

31. Sudre CH, Li W, Vercauteren T, Ourselin S, Jorge Cardoso M. Generalised Dice Overlap as a Deep Learning Loss Function for Highly Unbalanced Segmentations. Paper presented at: Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support; 2017//, 2017; Cham.

32. Resheff YS, Mandelbaum A, Weinshall D. Every Untrue Label is Untrue in its Own Way: Controlling Error Type with the Log Bilinear Loss. *arXiv e-prints.* 2017.arXiv:1704.06062. https://ui.adsabs.harvard.edu/abs/2017arXiv170406062R Accessed April 01, 2017.

33. Ren M, Zemel RS. End-to-End Instance Segmentation with Recurrent Attention. Paper presented at: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 21-26 July 2017, 2017.

34. Milletari F, Navab N, Ahmadi S. V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation. Paper presented at: 2016 Fourth International Conference on 3D Vision (3DV); 25-28 Oct. 2016, 2016.

35. Huang Y, Tang Z, Chen D, Su K, Chen C. Batching Soft IoU for Training Semantic Segmentation Networks. *IEEE Signal Processing Letters.* 2020;27:66-70.

36. Nogueira K, Mura MD, Chanussot J, Schwartz WR, Santos JAd. Learning to semantically segment high-resolution remote sensing images. Paper presented at: 2016 23rd International Conference on Pattern Recognition (ICPR); 4-8 Dec. 2016, 2016.

37. Cui Y, Jia M, Lin T, Song Y, Belongie S. Class-Balanced Loss Based on Effective Number of Samples. Paper presented at: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR); 15-20 June 2019, 2019.

38. Zhou Z, Rahman Siddiquee MM, Tajbakhsh N, Liang J. UNet++: A Nested U-Net Architecture for Medical Image Segmentation. Paper presented at: Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support; 2018//, 2018; Cham.

39. Bullitt E, Zeng D, Gerig G, et al. Vessel Tortuosity and Brain Tumor Malignancy: A Blinded Study1. *Academic Radiology.* 2005;12(10):1232-1240.

40. Zhao F, Chen Y, Chen F, et al. Semi-Supervised Cerebrovascular Segmentation by Hierarchical Convolutional Neural Network. *IEEE ACCESS.* 2018;6:67841-67852.

41. Janson S. Random coverings in several dimensions. *Acta Mathematica.* 1986;156(1):83-118.

42. Abadi M, Agarwal A, Barham P, et al. TensorFlow: Large-scale machine learning on heterogeneous systems. 2015.

43. Dauphin YN, Vries Hd, Bengio Y. Equilibrated adaptive learning rates for non-convex optimization. *Computer Science.* 2015.

44. Holm SA. A Simple Sequentially Rejective Multiple Test Procedure. *Scandinavian Journal of Statistics.* 1979;6(1):65-70.

45. He K, Zhang X, Ren S, Sun J. Deep Residual Learning for Image Recognition. Paper presented at: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 27-30 June 2016, 2016.

46. Huang G, Liu Z, Maaten LVD, Weinberger KQ. Densely Connected Convolutional Networks. Paper presented at: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 21-26 July 2017, 2017.

47. Zhang H, Cissé M, Dauphin Y, Lopez-Paz D. mixup: Beyond Empirical Risk Minimization. *ArXiv.* 2018;abs/1710.09412.

48. Cubuk ED, Zoph B, Mané D, Vasudevan V, Le QV. AutoAugment: Learning Augmentation Strategies From Data. Paper presented at: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR); 15-20 June 2019, 2019.

Figure Legends

Fig. 1. MDNet consists of two parallel pathways. The pathway with the normal resolution is a 3D U-Net, for illustration to omit the short cuts (from max pooling to the de-conv layers) in the figure. The pathway with the low resolution is a 3D FCN with four 5×5×5 convolutional layers without zero-padding (for illustration to reduce the number of layers in the figure, in practice, the kernel size is 3×3×3).

Fig. 2. Cerebral vessel segmentation results on the MRA dataset obtained using six different methods. Second row zooms in the white box in the first row. The green arrows indicate low-contrast blood vessels and background voxels, which are easily misclassified on the original image and ground truth, and can be effectively segmented by the proposed MDNet-Vb. The white arrows indicate the wrong segmentation on the same region using the other five state-of-the-art methods.

Fig. 3. Coronary artery segmentation results on the CTA dataset obtained using six different methods. The yellow arrows in the first row indicate background voxels that are easily misclassified on the original image and ground truth, and can be effectively segmented by the proposed MDNet-Vb. The white arrows indicate the wrong segmentation on the same region using the other five state-of-the-art methods.

Fig. 4. 3D segmentation results on MRA (first row) and CTA (second row) images, respectively. The results achieved by the three methods simultaneously trained on both two datasets.

Fig. 5. Coronary artery (first row) and cerebral vessel (second row) segmentation results on two datasets obtained using four combinations of FCN and U-Net networks.

Table 1. Performance on the testing set with different segmentation methods on the CTA dataset

	SEN (%)	SPE (%)	DSC (%)	HD (voxel)
ResNet	**78.06**	99.91	59.41	46.52
DenseNet	76.92	99.92	61.59	42.69
3D U-Net	70.63	**99.97**	69.20	24.87
V-Net	70.85	99.96	67.92	27.54
DeepMedic	75.47	**99.97**	72.64	19.15
Ours	76.46	99.96	**72.91**	**17.52**

Table 2. Performance on the testing set with different segmentation methods on the MRA dataset

	SEN (%)	SPE (%)	DSC (%)	HD (voxel)
ResNet	**69.92**	99.84	60.98	4.80
DenseNet	45.15	99.90	49.62	4.83
3D U-Net	61.01	**99.95**	68.22	**1.73**
V-Net	64.46	99.94	69.21	1.90
DeepMedic	64.46	99.94	68.64	2.48
Ours	67.60	99.92	**69.32**	2.47

Table 3. Performance on the testing set by simultaneously training with both MRA and CTA images on the CTA dataset

	SEN (%)	SPE (%)	DSC (%)	HD (voxel)
3D U-Net	72.27	99.96	70.10	31.25
DeepMedic	**79.79**	99.95	70.06	27.76
Ours	75.50	**99.97**	**73.06**	**21.02**

Table 4. Performance on the testing set by simultaneously training with both MRA and CTA images on the MRA dataset

	SEN (%)	SPE (%)	DSC (%)	HD (voxel)
3D U-Net	43.96	**99.96**	58.35	**2.42**
DeepMedic	55.89	**99.96**	65.22	2.51
Ours	**59.37**	99.94	**65.40**	2.58

Table 5. The setting of four double-pathway networks

	$S(I_S)$	$S(O_S)$	$S(I_L)$	$S(O_L)$	#Parameters
FCN+FCN	52	36	28	12	1,679,427
U-Net+U-Net	36	36	28	28	2,233,024
FCN+U-Net	52	36	28	28	1,865,374
U-Net+FCN (ours)	36	36	28	12	1,530,524

Note: $S(I_S)$, $S(I_L)$ denote the sizes of inputs of normal- and low-resolution pathways described in II.A, respectively. $S(O_S)$,

$S(O_L)$ are the sizes of outputs of normal- and low-resolution pathways, respectively.

Table 6. Performance on the testing set with four double-pathway networks on the CTA dataset

	SEN (%)	SPE (%)	DSC (%)	HD (voxel)	Time (s)
FCN+FCN	**75.47**	**99.97**	**72.64**	19.15	219,175
U-Net+U-Net	72.12	99.96	68.36	28.22	176,144
FCN+U-Net	55.93	99.96	58.25	**19.02**	220,287
U-Net+FCN (ours)	72.56	**99.97**	71.82	19.95	**151,840**

Table 7. Performance on the testing set with four double-pathway networks on the MRA dataset

	SEN (%)	SPE (%)	DSC (%)	HD (voxel)	Time (s)
FCN+FCN	**64.46**	99.94	68.64	2.48	158,756
U-Net+U-Net	58.42	**99.96**	67.75	**1.76**	119,714
FCN+U-Net	31.90	99.64	23.95	8.25	159,938
U-Net+FCN (ours)	63.20	99.95	**68.92**	1.92	**118,870**

Table 8. Performance on the testing set with different down-sampling factors on the CTA dataset

	SEN (%)	SPE (%)	DSC (%)	HD (voxel)
3 (ours)	**76.46**	99.96	**72.91**	**17.52**
5	74.75	**99.97**	72.84	18.68
7	74.80	**99.97**	72.68	19.74

Table 9. Performance on the testing set with different down-sampling factors on the MRA dataset

	SEN (%)	SPE (%)	DSC (%)	HD (voxel)
3 (ours)	**67.60**	99.92	**69.32**	**2.47**
5	67.42	99.93	69.09	2.50
7	65.31	**99.94**	69.06	2.48

Table 10. Performance on the testing set with different loss functions on the CTA dataset

	SEN (%)	SPE (%)	DSC (%)	HD (voxel)
CE	72.56	**99.97**	71.82	19.95
FL	72.62	**99.97**	72.29	17.64
Dsc	70.99	**99.97**	71.56	17.39
CE-Vb	72.66	**99.97**	72.48	18.29
FL-Vb	**76.46**	99.96	**72.58**	17.52
Dsc-Vb	71.01	**99.97**	72.14	**16.60**

Table 11. Performance on the testing set with different loss functions on the MRA dataset

	SEN (%)	SPE (%)	DSC (%)	HD (voxel)
CE	63.20	**99.95**	68.92	1.92
FL	62.30	99.94	67.84	1.97
Dsc	63.55	**99.95**	69.03	**1.72**
CE-Vb	65.06	99.94	68.94	2.10
FL-Vb	65.01	99.93	68.35	2.21
Dsc-Vb	**65.28**	99.94	**69.06**	2.12

Note: CE denotes cross-entropy, and FL denotes focal loss, and Dsc denotes dice coefficient loss. CE-Vb, FL-Vb, Dsc-Vb

denote our class-balanced cross-entropy, focal loss, and dice coefficient loss at the voxel level, respectively.

Table 12. Performance on the testing set with different sampling strategies on the CTA dataset

	SEN (%)	SPE (%)	DSC (%)	HD (voxel)
traditional	**76.61**	99.95	67.77	31.33
ours	72.56	**99.97**	**71.82**	**19.95**

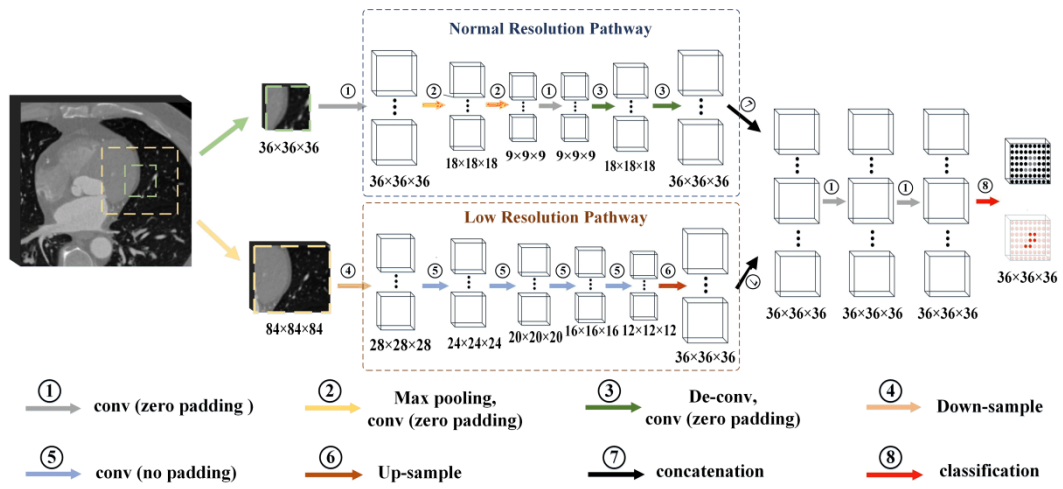**Table 13.** Performance on the testing set with different sampling strategies on the MRA dataset

	SEN (%)	SPE (%)	DSC (%)	HD (voxel)
traditional	59.04	**99.95**	66.01	2.10
ours	**63.20**	**99.95**	**68.92**	**1.92**

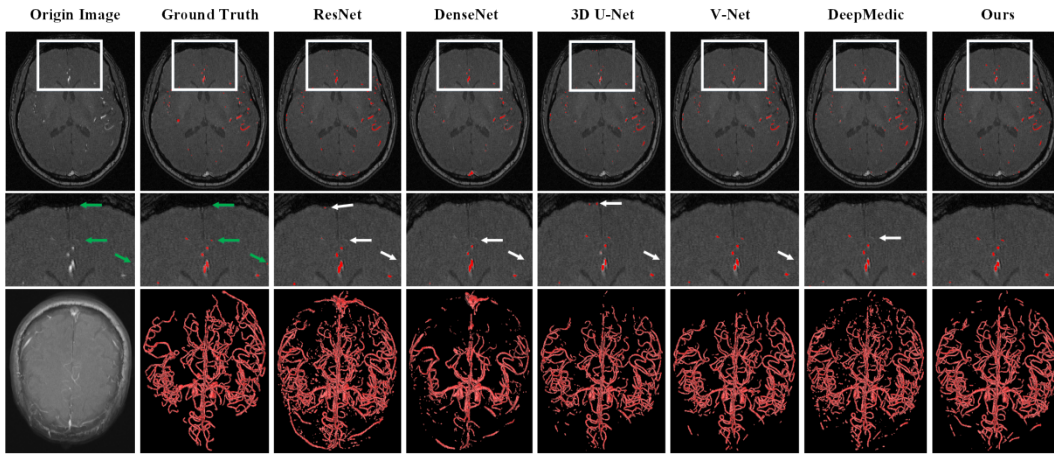**Table 14.** Performance on the testing set with/without augment on the CTA dataset

	SEN (%)	SPE (%)	DSC (%)	HD (voxel)
Without	76.46	**99.96**	**72.91**	**17.52**
With	**81.82**	99.95	71.02	27.47

Table 15. Performance on the testing set with/without augment on the MRA dataset
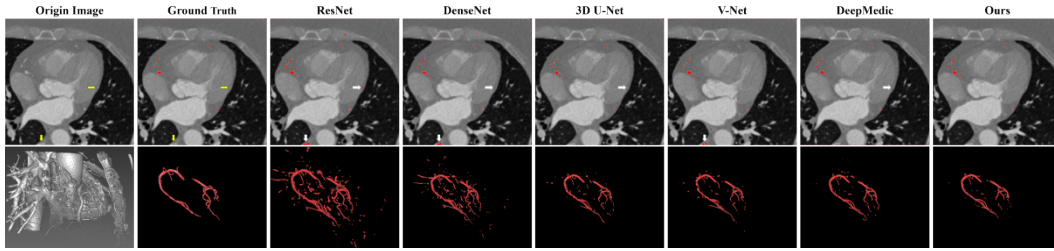
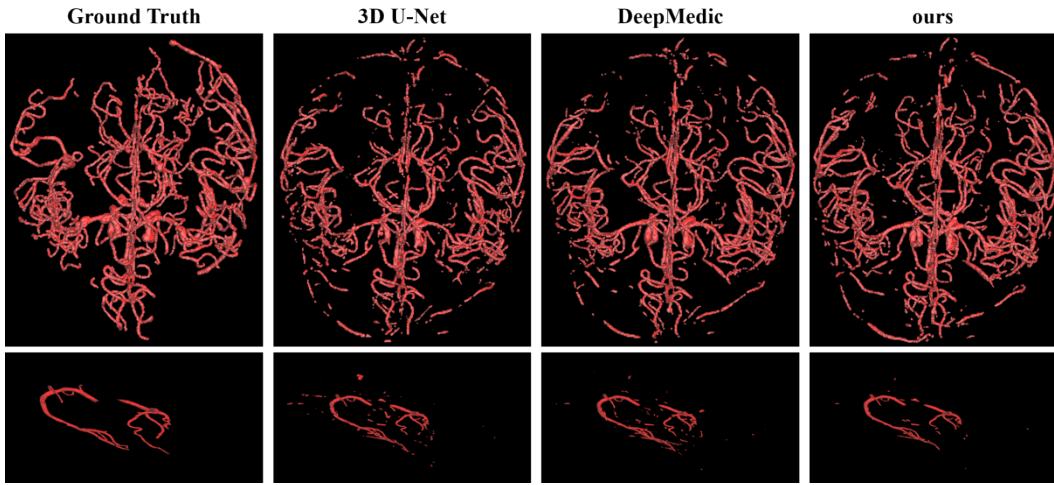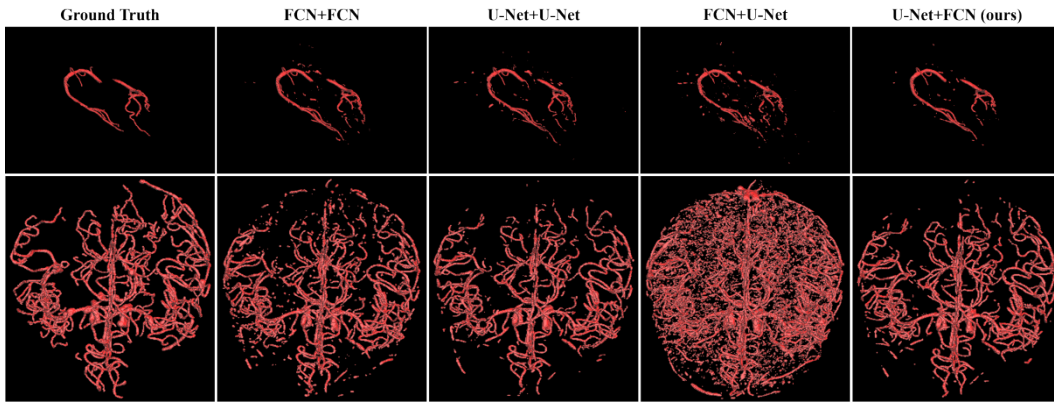|	SEN (%)	SPE (%)	DSC (%)	HD (voxel)
Without	65.28	**99.94**	69.06	**2.12**
With	**67.60**	99.92	**69.32**	2.47

mp_14934_f1.tif

| Origin Image | Ground Truth | ResNet | DenseNet | 3D U-Net | V-Net | DeepMedic | Ours |

mp_14934_f2.tif

| Origin Image | Ground Truth | ResNet | DenseNet | 3D U-Net | V-Net | DeepMedic | Ours |

mp_14934_f3.tif

Ground Truth   3D U-Net   DeepMedic   ours

mp_14934_f4.tif

| Ground Truth | FCN+FCN | U-Net+U-Net | FCN+U-Net | U-Net+FCN (ours) |

mp_14934_f5.tif