



3D vessel-like structure segmentation in medical images by an edge-reinforced network

Likun Xia^{a,1}, Hao Zhang^{a,b,d,1}, Yufei Wu^c, Ran Song^d, Yuhui Ma^b, Lei Mou^b, Jiang Liu^e, Yixuan Xie^{a,b}, Ming Ma^f, Yitian Zhao^{c,b,*}

^a College of Information Engineering, Capital Normal University, Beijing, China

^b Cixi Institute of Biomedical Engineering, Ningbo Institute of Materials Technology and Engineering, Chinese Academy of Sciences, Ningbo, China

^c The Affiliated People's Hospital of Ningbo University, Ningbo, China

^d School of Control Science and Engineering, Shandong University, Jinan, China

^e Department of Computer Science and Engineering, Southern University of Science and Technology, Shenzhen, China

^f Department of Computer Science, Winona State University, Winona, USA

ARTICLE INFO

Keywords:

Vessel-like structure
3D segmentation
Attention
Deep network
Loss function

ABSTRACT

The vessel-like structure in biomedical images, such as within cerebrovascular and nervous pathologies, is an essential biomarker in understanding diseases' mechanisms and in diagnosing and treating diseases. However, existing vessel-like structure segmentation methods often produce unsatisfactory results due to challenging segmentations for crisp edges. The edge and nonedge voxels of the vessel-like structure in three-dimensional (3D) medical images usually have a highly imbalanced distribution as most voxels are non-edge, making it challenging to find crisp edges. In this work, we propose a generic neural network for the segmentation of the vessel-like structures in different 3D medical imaging modalities. The new edge-reinforced neural network (ER-Net) is based on an encoder-decoder architecture. Moreover, a reverse edge attention module and an edge-reinforced optimization loss are proposed to increase the weight of the voxels on the edge of the given 3D volume to discover and better preserve the spatial edge information. A feature selection module is further introduced to select discriminative features adaptively from an encoder and decoder simultaneously, which aims to increase the weight of edge voxels, thus significantly improving the segmentation performance. The proposed method is thoroughly validated using four publicly accessible datasets, and the experimental results demonstrate that the proposed method generally outperforms other state-of-the-art algorithms for various metrics.

1. Introduction

The visualization and quantitative analysis of vessel-like structures in biomedical images are of the utmost importance, where such structures indicate the presence of cerebrovascular, nerve fiber, bronchi, or other functional vessel-like structures. The geometric and morphological changes of these structures are closely associated with many diseases (Rivest-Henault and Cheriet, 2013; Zhao et al., 2020; Fu et al., 2020).

3D medical images contain relatively rich depth information, which is more conducive to observing the morphology of organs and other tissues (Li et al., 2018). The top row of Fig. 1 illustrates two examples of 3D vessel-like structure image types, which include the circle of Willis (CoW) region in magnetic resonance angiography (MRA) and

olfactory projection fibers (OPF) in the two-photon microscope image. Cerebrovascular is a vital part of the brain, and its morphological changes are closely related to stroke, aneurysm, arteriovenous malformation and other diseases (Zhao et al., 2017b). MRA is a commonly used cerebrovascular imaging technique, and accurate detection and quantification of the cerebrovasculature from MRA imagery are essential for many clinical applications to support early diagnosis, optimal treatment and neurosurgery planning. However, due to varying scales of noise, low image contrast and sophisticated vessel geometry and topology variations, it is challenging to obtain accurate cerebrovascular segmentation.

The axonal and dendritic nerve is another important vessel-like structure in biomedical images. Two-photon microscopy is considered one

* Corresponding author at: Cixi Institute of Biomedical Engineering, Ningbo Institute of Materials Technology and Engineering, Chinese Academy of Sciences, Ningbo, China.

E-mail address: yitian.zhao@nimte.ac.cn (Y. Zhao).

¹ Likun Xia and Hao Zhang contribute equally.

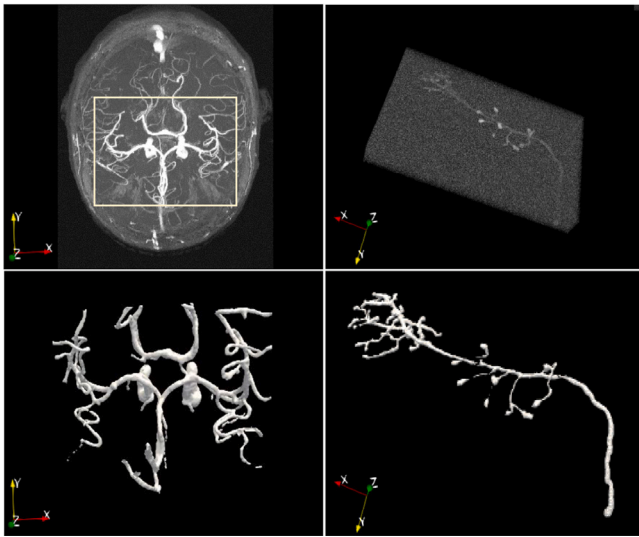


Fig. 1. 3D volume (top row) and their manual annotations of vessel-like structures (bottom row) in two sample medical imaging modalities. Left: Circle of Willis in MRA; Right: Olfactory projection fibers in two-photon microscope image.

of the best candidates to image nerves (Callara et al., 2020). Accurate reconstruction of these nerves from microscope imagery is necessary to analyze nerve morphological characterization and diagnose nervous system diseases (Li et al., 2018). However, it remains challenging to reconstruct axonal and dendritic nerves due to the low intensity and nonuniform intensity in a noisy background.

The bottom row of Fig. 1 demonstrates the manual annotations of two types of 3D vessel-like structure images. However, manual annotation for vascular or nerve structure is an exhausting task, and most existing computer-aided systems cannot reliably extract and segment these structures due to the high degree of anatomical variation (Zhao et al., 2017b). To this end, an automatic method for 3D vessel-like structure segmentation is highly desirable. In the last two decades, automatic vessel-like structure segmentation for 2D vessel-like images has rapidly developed, as evidenced by several survey papers (Katouzian et al., 2012; Salazar-Gonzalez et al., 2014; Xia et al., 2019). Recently, vessel-like structure segmentation from 3D images, e.g., cerebrovascular from MRA and Computed Tomography Angiography (CTA) (Phellan et al., 2017; Livne et al., 2019; Sanches et al., 2019) and nerve from optical microscopy images (Peng et al., 2015; Li and Shen, 2019; Callara et al., 2020), has drawn much attention. However, the extraction of 3D vessel-like structures is an underexplored topic. Most techniques tend to oversegment or missegment, primarily due to varying scales of noise, low image contrast, and sophisticated vessel geometry and topology variations.

Recent deep learning-based methods (Yang et al., 2014; Çiçek et al., 2016; Milletari et al., 2016; Li and Shen, 2019; Sanches et al., 2019; Mou et al., 2021) have demonstrated excellent performance in the medical image segmentation field, and most of them are based on the encoder-decoder architecture, among which the most representative methods are 3D U-Net (Çiçek et al., 2016) and its variants (Milletari et al., 2016; Sanches et al., 2019). However, these methods cannot effectively extract the edge part of the vessel-like structure because the distribution of edge voxels and nonedge voxels of vessel-like structures in 3D medical images is extremely unbalanced, where most of them are nonedge voxels. These deep learning-based methods treat all voxels equally, which makes it challenging to find crisp edges. In addition, the slender shape of the vessel-like structure increases the difficulty in extracting edge voxels. In summary, extraction of the edge is particularly important for the continuity of the vessel-like structure. The segmentation results of the vessel-like structure are heavily affected by

the extraction of the edge part. Existing segmentation methods often fail to effectively detect microvascular or microneurons, primarily due to the difficulty in extracting crisp edges. Therefore, edge extraction improvements are crucial for vessel-like structure segmentation.

In this paper, we propose an edge-reinforced network (ER-Net) for 3D vessel-like structure segmentation based on an encoder-decoder architecture, which can capture edge information effectively. Specifically, a reverse edge attention module (REAM) (Zhang et al., 2020a) is embedded between adjacent layers of the encoder to discover edge information by increasing the weight of the voxels on the edge part and sequentially exploiting complementary regions and details by erasing the estimated object regions from high-level side-output features. In addition, since not all the features are effective for extracting vessel-like structures, a feature selection module (FSM) is used to adaptively select effective features from the recovery features of the decoder and the coding features of the encoder to obtain more details of the vessel-like structure. Finally, we propose a novel piecewise loss called edge-reinforced loss (ERloss) to constrain the edge and nonedge voxels of the vessel-like structures to significantly improve the segmentation performance.

The proposed method considerably extends our previous work, namely, RE-Net, published in MICCAI 2020 (Zhang et al., 2020a). In the present work in contrast to the past work, a feature selection module is designed to adaptively select effective features, and an edge-reinforced loss function is introduced to preserve the spatial edge information. In addition, RE-Net was trained and validated on only one cerebrovascular dataset in the previous work, while we have expanded our data pool to four different datasets with different imaging modalities in this work. Overall, the research contributions of our work can be summarized as follows:

(1) A novel edge-reinforced neural network is introduced for 3D vessel-like structure segmentation over different types of biomedical images. It employs reverse edge attention and edge-reinforced optimization loss to constrain both the edge and nonedge voxels of vessel-like structures and adaptively selects features that significantly increase the edge voxels' weight and that improve the segmentation performance.

(2) The proposed method is validated quantitatively and qualitatively using two cerebrovascular and two nerve datasets. A comparative analysis shows that the proposed ER-Net achieves state-of-the-art performances in segmenting 3D vessel-like structures. The implementation code of our method can be found at: <https://github.com/iMED-Lab/ER-Net>.

2. Related work

In this section, we briefly review the start-of-the-art approaches for vessel-like structure segmentation and 3D attention mechanisms for medical images.

2.1. Vessel-like structure segmentation

Vessel-like structure segmentation has been studied for decades in medical image analysis, and the developed methods can roughly be classified into two categories: traditional and deep learning-based methods.

Traditional methods, also known as model-driven methods, focus on handcrafted features, including vessel-like structure intensity distribution, gradient features, morphological features and many others. For example, Wang et al. (2011) proposed an open-curve snake-based approach for automated tracing of neuronal structures in 3D microscopy images. Yang et al. (2014) segmented cerebrovasculature from CTA images using active contour models. Forkert et al. (2013) performed small and malformed cerebrovascular segmentation from TOF-MRA images by geometric models. In addition, a variety of filtering methods have been proposed including Hessian matrix-based filters (Frangi

et al., 1998), symmetry filters (Zhao et al., 2017b) and tensor-based filters (Zhao et al., 2017b), that aim to remove undesired intensity variations in the images and suppress background structures and noise. However, these methods require an elaborate design, which depends heavily on domain knowledge. Moreover, their performance degrades when the structures become irregular and the images become noisy.

In recent years, deep learning-based methods have demonstrated excellent performance in medical image analyses without any requirement of manual feature extraction. Phellan et al. (2017) constructed a simple neural network to segment cerebrovascular tissue on a 2D slice of TOF-MRA images. Mou et al. (2019) developed CS-Net based on attention mechanisms for the segmentation of corneal nerves in microscope images. Yang et al. (2021) proposed a multidiscriminator adversarial convolutional network for nerve fiber segmentation in vivo corneal confocal microscopy images. However, the above methods are based on the feature extractions from slices by a 2D CNN, and they discard valuable 3D context information crucial for tracking vessel-like structures. To this end, Çiçek et al. (2016) extended 2D U-Net to 3D U-Net by replacing all 2D operations with their 3D counterparts to segment 3D volume medical images, which has become a baseline model for segmenting 3D volume data. Tetteh et al. (2018) designed a 3D vessel segmentation network called DeepVesselNet to extract vessel networks and features in both MRA images of the human brain and CTA microscopy images of the rat brain. Li and Shen (2019) proposed using 3D U-Net Plus to segment the neuron from the surrounding fibers in volumetric optical microscopy neuronal images. Sanches et al. (2019) presented the Uception network, combined with the inception architecture based on the 3D U-Net model, for segmentation of an arterial cerebrovascular network in TOF-MRA images. Yang et al. (2020) proposed a voxelwise multilevel fully convolutional network to obtain 3D neuron image segmentation maps. Mou et al. (2021) introduced a two-channel attention-based segmentation network (CS²-Net) for 2D and 3D vessel-like structures in five different medical imaging modalities. Above all, the existing methods pay more attention to the skeleton segmentation of vessel-like structures. However, they do not prioritize the extraction of the edge parts in vessel-like structures, which severely limits the segmentation performance and leads to failures in identifying the weak blood vessels and edge information.

2.2. Attention mechanism for 3D medical images

The attention mechanism in computer vision applications has been successfully applied to various tasks, especially in medical image segmentation, due to its powerful feature selection ability. Oktay et al. (2018) proposed an attention gate (AG) model that automatically focuses on the target structures of varying shapes and sizes for CT abdominal image segmentation. Wang et al. (2019a) presented a volumetric attention (VA) module that can infer 3D enhanced attention maps along the channel and spatial dimensions for liver tumor segmentations in CT images. Mou et al. (2020) proposed using the self-attention mechanism in the channel and spatial dimensions to segment 3D curve structures, such as nerves and blood vessels, in medical imaging. Zhang et al. (2020b) proposed employing attention guidance to strengthen the 3D context perceiving ability of the segmentation decoder for brain tumor segmentation in MRI images. The above attention mechanisms recalibrate the feature maps of 3D medical images in the spatial and channel dimensions and then give them weights and fuse them to obtain the final volume features. In this way, the attention mechanism enables CNNs to focus attention on a certain object consciously and actively. However, there is a lack of research on using attention mechanisms to discover edge information in 3D vessel-like structure images. On the one hand, the extraction of such structures is a less explored topic. On the other hand, edge information is more difficult to extract due to the slender shape in the vessel-like structures.

3. Methodology

3.1. Overall architecture

The overall structure of the proposed segmentation method is illustrated in Fig. 2. It consists of a contracting path (encoder) and an expansive path (decoder). The encoder of the network is stacked by four layers that process feature maps at different resolutions. Each layer extracts features via convolutions followed by a $2 \times 2 \times 2$ maxpool layer with a stride of 2 to reduce the resolution to half. The residual block (ResBlock) (He et al., 2016) with two $3 \times 3 \times 3$ convolutional layers is adopted as each encoder layer, where each convolutional layer is followed by batch normalization (BN) and a rectified linear unit (ReLU) layer. With the shortcut connection in ResBlock, gradient vanishing can be alleviated during training, which is conducive to accelerating convergence. The channels of the feature maps are doubled after each encoder layer and increase from 32 to 256 through the encoder.

In contrast to the encoder, the decoder is employed to progressively upsample the feature maps extracted by the encoder to the resolution of the input image. The decoder is composed of three layers, each of which consists of a $2 \times 2 \times 2$ deconvolution with a stride of 2, followed by a novel feature selection module (FSM) and ResBlock. The FSM enables the adaptive selection of effective features from the recovery features of the decoder and the coding features of the encoder to obtain more details of the vessel-like structure. The ResBlock is then introduced for better recovery features. Correlated to the encoder, the channels of the feature maps are halved after each decoder layer and decrease from 256 to 32 through the decoder. In addition, reverse edge attention module (REAM) is embedded in each skip connection to capture the edge information from the corresponding encoder layer. At the end of the network, a $1 \times 1 \times 1$ convolutional layer with a sigmoid function is utilized to produce the segmentation probability map, followed by the Otsu method (Otsu, 1979) to produce the binary result.

To refine both the region and edge of the vessel-like structures, we propose a novel ERloss to train the segmentation framework. The ERloss is designed in the form of a piecewise function, where the optimal objective can be adjusted according to the performance of the framework during the training phase. Specifically, the framework is expected to focus on improving segmentations based on the regions first; only when a region segmentation is satisfactory enough will the edge refining of the structures begin to be considered. More details are described in Section 3.4.

3.2. Reverse edge attention module

Consecutive pooling and striding convolutional operations may enlarge the receptive field and obtain more global information. However, some details are lost after downsampling during the encoder stage, which implies that predictions near the edge tend to be less accurate (Gu et al., 2019). One possible solution to learn a more precise output is to apply a skip connection between each layer of the encoder and decoder in a U-shaped network (Ronneberger et al., 2015; Milletari et al., 2016; Ma et al., 2020). Such a connection can combine features from different levels to compensate for information caused by the down-sampling operations. Inspired by the reverse attention model in Chen et al. (2018), we introduce the REAM module to capture the edge information from the feature maps generated by the encoding layers. As illustrated in Fig. 3, it is embedded between adjacent layers of the encoder to discover the edge information, sequentially exploiting complementary regions and details by erasing the estimated object regions from the high-level side-output features, where the existing estimation is upsampled from the deeper layer.

The upsampling stage is shown in Eq. (1), where $F_i \in \mathbb{R}^{h \times w \times d \times c}$, $2 \leq i \leq 4$ denotes the high-level features generated in the i^{th} encoder stage; h , w , and d represent the height, width and depth of the feature map, respectively; and c represents the number of channels of the

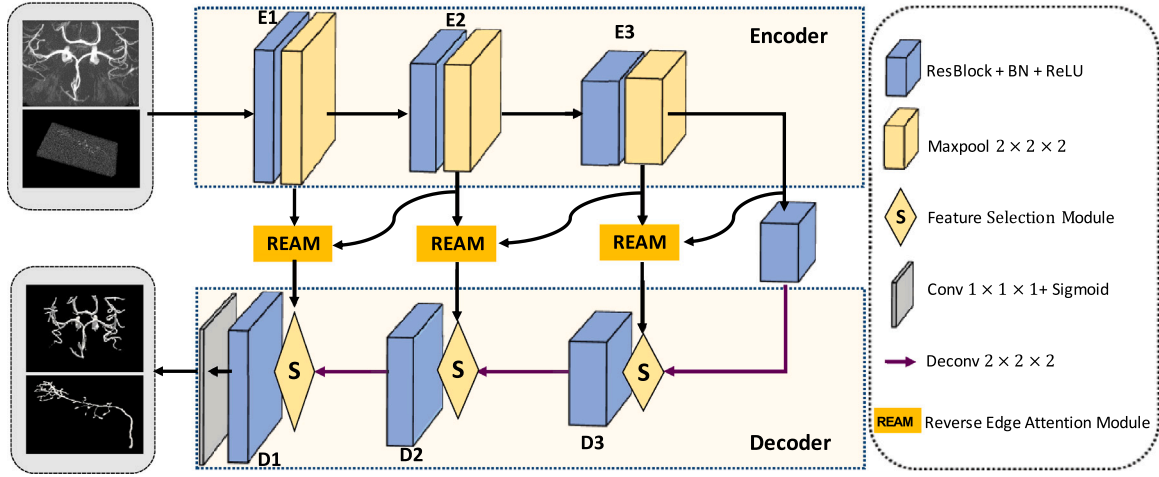


Fig. 2. The architecture of the proposed ER-Net for vessel-like structures segmentation: an encoder, a reverse edge attention module, a feature selection module and a decoder.

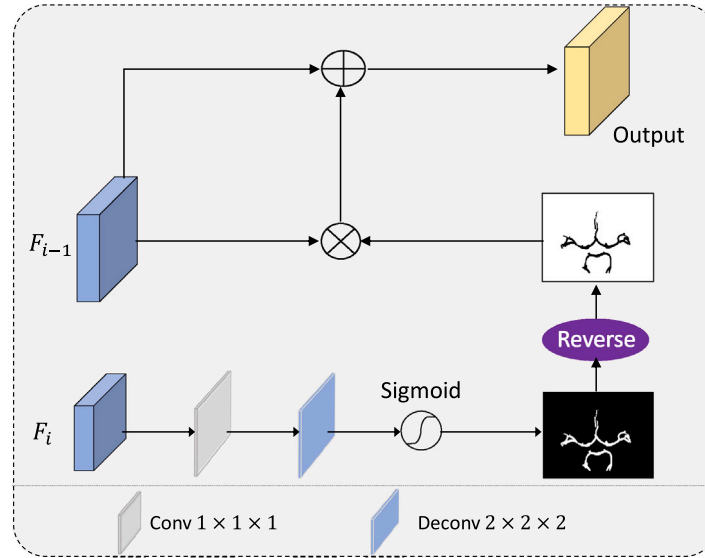


Fig. 3. The diagram of the reverse edge attention module (REAM). The low- and high-level features in the encoder, denoted as F_{i-1} and F_i , respectively are input to REAM to generate an output containing the edge information.

feature map. A $1 \times 1 \times 1$ convolution, $Conv_1$, is initially used to fuse the features into a single channel. The features are then upsampled to the same resolution as the outputs of the $(i-1)^{th}$ encoder stage by a $2 \times 2 \times 2$ deconvolution with a stride of 2, thus generating $X_i \in \mathbb{R}^{h \times w \times d \times 1}$:

$$X_i = Up(Conv_1(F_i)) \quad (1)$$

The corresponding weight A_{i-1} in the $(i-1)^{th}$ encoder stage is generated by subtracting the upsampled prediction of the i^{th} stage from 1, as follows:

$$A_{i-1} = 1 - \text{Sigmoid}(X_i) = 1 - \frac{1}{1 + e^{-X_i}} \quad (2)$$

The feature map F_i output by the i^{th} layer corresponds to the foreground region, and the probability map A_{i-1} of the background region in the $(i-1)^{th}$ layer can be generated by $(1 - \text{Sigmoid}(X_i))$. Let $F_{i-1} \in \mathbb{R}^{h \times w \times d \times c}$ denote the low-level features generated in the $(i-1)^{th}$ encoder stage. Then, the edge feature E_{i-1} can be captured by elementwise multiplication on F_{i-1} and A_{i-1} , expressed as follows:

$$E_{i-1} = F_{i-1} * A_{i-1} \quad (3)$$

The intersection of the foreground and background of different layers can be regarded as the target edge feature. These features are fused into the corresponding features of the decoder by a skip connection after a summing with F_{i-1} . That is,

$$F_{sum} = E_{i-1} + F_{i-1} \quad (4)$$

The edge feature is discovered by increasing the weight of voxels on the edge part. Comparably, the feature map after a concatenation at the same position has more edge information with the aid of REAM, thus improving the segmentation accuracy.

3.3. Feature selection module

In some popular encoder-decoder architectures such as U-Net (Ronneberger et al., 2015) and V-Net (Milletari et al., 2016), the concatenation at the end of the skip connection combines feature maps from the encoder layer and symmetric decoder layer. However, feature fusion via simple channel stacking may lead to feature redundancy in that not all the features are effective for extracting the vessel-like structures. To this end, we introduce a novel module, the feature selection module

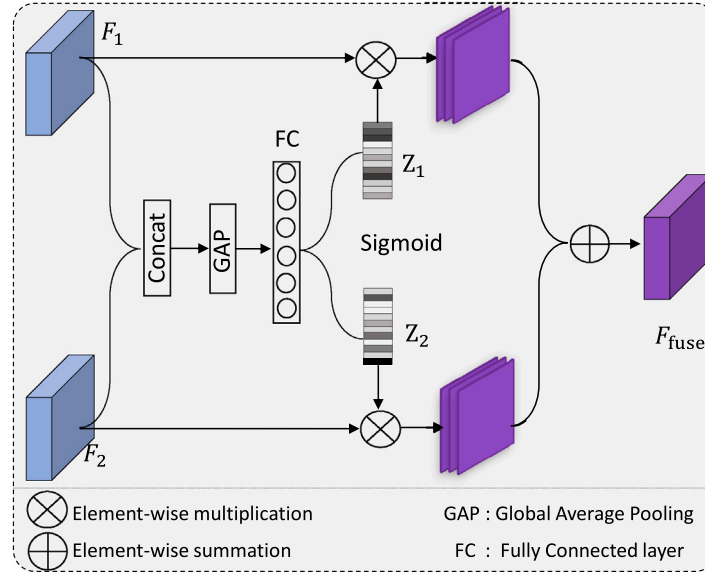


Fig. 4. The diagram of the feature selection module (FSM). Two features from the encoder layer and the symmetric decoder layer, denoted as F_1 and F_2 , respectively, are adaptively combined by FSM for optimal representations of the vessel-like structures.

(FSM), to adaptively select the effective features from the recovery features of the decoder and the coding features of the encoder to obtain more details of the vessel-like structure. The module can adaptively combine features from the encoder layer and the symmetric decoder layer for optimal representations of the vessel-like structures.

As shown in Fig. 4, the two feature maps from the encoder layer and the symmetric decoder layer (denoted as $F_1 \in \mathbb{R}^{h \times w \times d \times c}$ and $F_2 \in \mathbb{R}^{h \times w \times d \times c}$, respectively) are first fused by concatenation to generate the concatenated feature $U \in \mathbb{R}^{h \times w \times d \times 2c}$. The FSM weights the information from F_1 and F_2 using a channel attention mechanism; the channelwise weight $s \in \mathbb{R}^{1 \times 1 \times 1 \times 2c}$ is generated via global average pooling (GAP). The c th element of s can be expressed as:

$$s_c = \text{GAP}(U_c) = \frac{1}{h \times w \times d} \sum_{i=1}^h \sum_{j=1}^w \sum_{k=1}^d (U_c(i, j, k)) \quad (5)$$

where $U_c(i, j, k)$ is the value at coordinate position (i, j, k) of the c th channel of U . Afterward, two fully connected (FC) layers followed by the sigmoid function are applied to the channelwise weights to achieve the final attention weights:

$$z = \sigma(FC_2(\max(0, FC_1(s)))) \quad (6)$$

where σ denotes the sigmoid function. z can be regarded as the weight of the feature map of the original $2c$ channels, and it is learned from the two FC layers, which can encode the interdependencies between the different feature channels. The weights are calculated according to the features; thus the FSM can adaptively adjust the values with different inputs.

The final attention weights z are then split into two parts, i.e., $z_1 \in \mathbb{R}^{1 \times 1 \times 1 \times c}$ and $z_2 \in \mathbb{R}^{1 \times 1 \times 1 \times c}$ for F_1 and F_2 , respectively. Finally, the fused feature map is obtained as a channel-weighted sum of F_1 and F_2 :

$$F_{\text{fuse}} = z_1 \cdot F_1 + z_2 \cdot F_2 \quad (7)$$

where \cdot represents the channelwise product. The FSM utilizes the self-learned feature-oriented weights z_1 and z_2 to selectively integrate features from the encoder and symmetric decoder layers. It calculates the feature-oriented importance scores by using a channel attention mechanism and a sigmoid function. In the channel-weighted mechanism, features significantly associated with the vessel-like structures will be assigned larger weights, conducive to the segmentation task.

3.4. Edge reinforced loss

Although numerous deep learning methods have been proposed for vessel-like structure segmentation, most of them are based on pixel/voxel-level classification and do not pay particular attention to the extraction of edge features of these structures (Cao et al., 2020). In other words, such methods ignore the boundary of the object to be extracted, which is prone to rough and fuzzy mask prediction results. Compared with nonedge voxels, edge voxels are much fewer and less detailed, resulting in more difficult detections. To this end, we proposed a novel piecewise loss called edge-reinforced loss (ERloss) to constrain both the edge and nonedge voxels of the predicted vessel-like structures to significantly improve the segmentation performance. The ERloss is formulated as follows:

$$\mathcal{L}_{ER} = \begin{cases} \mathcal{L}_{Dice}(p, l) & \text{if DSC} < \lambda, \\ \mathcal{L}_{Dice}(p, l) + \mathcal{L}_{edge}(p_{edge}, l_{edge}) & \text{if DSC} \geq \lambda. \end{cases} \quad (8)$$

where p_{edge} and l_{edge} denote the soft edge maps of the predicted probability map p and ground truth l , respectively, both of which are generated by a $3 \times 3 \times 3$ Laplace operator. It is natural for humans to distinguish objects by perceiving both their shape and edge. Inspired by this, we explore the relationships between shapes and edges in an explicit way to obtain the final semantic segmentation result. The shape feature is supervised by the mask, where the edges are ignored during training, while the edge feature is supervised by an edge mask for edge prediction.

The ERloss provides two conditional objective option mask-supervised loss and edge-refined loss, for optimization of the vessel-like regions and tubular edges. Considering that the refining of edges is based on a satisfactory vessel-like structure segmentation result, we calculate the Dice Similarity Coefficient (DSC) of the predicted vessel-like structures in every iteration and set a switching condition of the two options during the training phase. If the calculated DSC is smaller than a threshold λ , the mask-supervised loss is adopted in the current iteration; otherwise, the edge-refined loss is adopted.

Mask-supervised loss adopts the Dice loss \mathcal{L}_{Dice} , which is defined as:

$$\mathcal{L}_{Dice}(p, l) = 1 - \frac{2 \sum_{i=1}^N p_i l_i + \epsilon}{\sum_{i=1}^N p_i^2 + \sum_{i=1}^N l_i^2 + \epsilon} \quad (9)$$

where N denotes the number of voxels; $p_i \in [0, 1]$ and $l_i \in \{0, 1\}$ accordingly denote the i th voxel of the predicted probability map p and

ground truth l . The Dice loss measures the nonoverlapping ratio between the prediction and ground truth and is insensitive to the number of foreground/background voxels, thus alleviating the class-imbalance issue.

The edge-refined loss consists of two terms: Dice loss (\mathcal{L}_{Dice}) and edge loss (\mathcal{L}_{edge}). The \mathcal{L}_{edge} measures the edge dissimilarity between the prediction and the ground truth, which is defined as the combination of BCE loss and Dice loss. As a pixel-wise loss, the BCE loss does not consider the global structure. In contrast, the Dice loss aims to optimize the global structure of the segmented object rather than focusing on a single pixel. Furthermore, the segmentation performance is strongly dependent on the relative weighting between each loss. Tuning these weights by hand is a difficult and expensive process. The performance is highly dependent on an appropriate choice of weighting between each loss. Searching for optimal weighting is prohibitively expensive and difficult to resolve with manual tuning. In this work, we propose a principled way of combining multiple loss functions to simultaneously learn multiple objectives using homoscedastic uncertainty in an adaptive weight update way. Inspired by the adaptive weight loss function in (Cao et al., 2020), \mathcal{L}_{edge} is defined as:

$$\mathcal{L}_{edge}(p_{edge}, l_{edge}) = \frac{\zeta}{\kappa^2} \mathcal{L}_{Dice}(p_{edge}, l_{edge}) + \frac{1}{\tau^2} \mathcal{L}_{bce}(p_{edge}, l_{edge}) + \log(1 + \kappa\tau) \quad (10)$$

where \mathcal{L}_{bce} represents the binary cross entropy (BCE) loss, which is expressed as: $\mathcal{L}_{bce}(p, l) = -\sum_{i=1}^N l_i \log p_i + (1 - l_i) \log(1 - p_i)$. The parameter ζ is a weighted balance parameter between the Dice loss and the BCE loss, which is empirically set as $\zeta = 1$. The auxiliary term $\log(1 + \kappa\tau)$ can be regarded as a weight learning term, where κ and τ are trainable parameters regularized by the auxiliary term and are automatically updated during training. During training, they mutually interact and restrict each other; thus, automatic balancing occurs.

4. Materials and segmentation evaluation

4.1. Materials

We verify the wide applicability of ER-Net in 3D vessel-like structure segmentation on four publicly accessible datasets: two cerebrovascular datasets and two nerve datasets.

4.1.1. Cerebrovascular datasets

MIDAS² is a public cerebrovascular dataset that contains a total of 100 MRA volumes that were acquired from healthy volunteers aged 18 to 60+ years. These volumes were captured by a 3T MRI scanner under standardized protocols, with a voxel size of $0.5 \times 0.5 \times 0.8 \text{ mm}^3$. The scan resolution of each volume is $448 \times 448 \times 128$.

- **MIDAS-I** is a subset of MIDAS including 50 manual annotations of Circles of Willis (CoW) in MRA. Manual annotations of CoW are provided by Prof. A. F. Frangi from the University of Leeds, UK, where vasculatures are generated by tracing the centerlines of the vessels, and the vessel surfaces were extracted using the geodesic active contour method (Bogunović et al., 2011). To focus only on the CoW region of interest, patches with a size of $224 \times 208 \times 64$ are cropped from the raw data.

- **MIDAS-II** is another version of MIDAS containing intra-cranial vasculature (centerline + radius), which was extracted from MRA images by Aylward and Bullitt (2002). Forty-two volume annotations are generated by the open-source Insight Toolkit (ITK). In the experiment, a cerebral mask without a skull was extracted from the original data by the FSL-BET (Brain Extraction Tool) (Smith, 2002). The images were resized to make them isotropic with a voxel size of 1 mm with a trilinear interpolation for a final size of $224 \times 224 \times 96$. During training, patches

with a size of $96 \times 96 \times 96$ were randomly cropped from the volume, followed by a 90-degree rotation for data augmentation. Patches of this size can both reduce computing costs and include a vascular region with a large area.

4.1.2. Nerve datasets

Two nerve datasets used in the paper are **OPF** (Olfactory Projection Fibers)³ and **NL1A** (Neocortical Layer1 Axons),⁴ both of which are subsets of the Digital Reconstruction of Axonal and Dendritic Morphology (DIADEM) challenge datasets.

- The OPF dataset consists of 9 separate Drosophila olfactory axonal projection image stacks, which are captured by two-channel confocal microscopy.

- The NL1 A dataset contains 16 image stacks involving numerous axonal trees, which are captured by two-photon laser scanning microscopy in vivo. It mainly illustrates axons in the nervous system region of neocortical layer 1, and each image stack within this dataset represents one tile for a mosaic that contains all of the separate axonal fibers.

All image stacks in both datasets are 512×512 pixels in X and Y coordinates. The provided gold standard reconstruction of each image stack is in the form of coordinate points and excludes spatial voxel information. Therefore, we applied the algorithm proposed in (Otsu, 1979) to segment nerves and then utilized the most connected domain for postprocessing to remove the outliers. Finally, the processed results were adjusted manually, and spatial voxel-level annotations were obtained.

4.2. Experimental setting

Our ER-Net was implemented on a PyTorch framework with a single GPU (RTX 3090). To improve convergence, a normalization step was applied where the intensities of the volumes were manipulated to have a zero-mean and a unit variance. During training, K-fold cross-validation was used to split the dataset, where K is 5, 5, 4 and 3 over the MIDAS-I, MIDAS-II, OPF and NL1 A datasets, respectively. Each dataset is split into K consecutive folds. Each fold is then used once as a validation while the K - 1 remaining folds form the training set. Adaptive moment estimation (Adam) (Kingma and Ba, 2014) was employed for network optimization. The initial learning rate was set to 0.0001, with a weight decay of 0.0005. A poly learning rate policy (Zhao et al., 2017) with a power of 0.9 was used. The maximum epoch was 1000. In addition, the λ in ERloss was experimentally set as 0.8, 0.6, 0.8 and 0.7 over the four datasets.

4.3. Evaluation metrics

To facilitate better observation and objective evaluation of the vessel-like structure segmentation method, we adopted voxel-wise metrics including sensitivity (Sen), specificity (Spe) and DSC and a distance metric, including the average Hausdorff distance (AHD).

Sensitivity measures the ability of the method to recognize positive samples. Specificity measures the ability of the method to recognize negative samples. DSC generally measures the intersection over the union between prediction and ground truth as a comprehensive indicator to evaluate the sparse vessel segmentation in a large portion of the background. AHD calculates the mean of the average directed distances from point set segmentation results P to point set labels L and the average directed distances from L to P . It is defined as:

$$\text{AHD} = \frac{1}{2} \left(\frac{1}{P} \sum_{p \in P} \min_{l \in L} d(p, l) + \frac{1}{L} \sum_{l \in L} \min_{p \in P} d(p, l) \right) \quad (11)$$

Compared with other metrics, AHD has the advantage of taking voxel location information into account, which reflects the edge error of segmentation results.

² <https://public.kitware.com/Wiki/TubeTK/Data>

³ https://diadem.janelia.org/olfactory_projection_fibers_readme.html

⁴ https://diadem.janelia.org/neocortical_layer_1_axons_readme.html

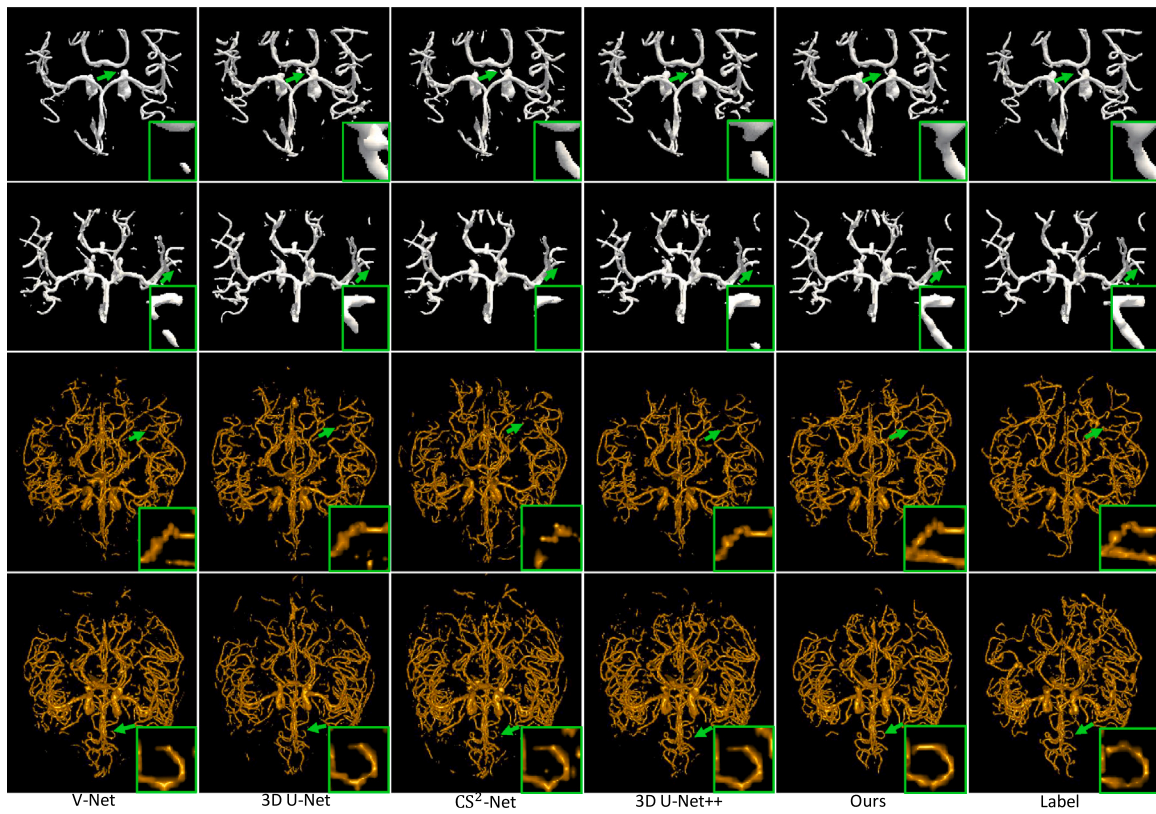


Fig. 5. Segmentation results on cerebrovascular datasets by different methods: V-Net (Milletari et al., 2016), 3D U-Net (Çiçek et al., 2016), CS²-Net (Mou et al., 2020) and 3D U-Net++ (Zhou et al., 2019).

Table 1

Cerebrovascular segmentation results obtained by different methods on MIDAS-I and II datasets.

Datasets	Methods	Sen	Spe	DSC	AHD
MIDAS-I	3D U-Net (Çiçek et al., 2016)	0.8913	0.9978	0.8459	0.5082
	V-Net (Milletari et al., 2016)	0.8727	0.9980	0.8423	0.5037
	3D U-Net ++ (Zhou et al., 2019)	0.8729	0.9981	0.8485	0.5011
	Uception (Sanches et al., 2019)	0.8888	0.9985	0.8605	0.4818
	CS²-Net (Mou et al., 2020)	0.8984	0.9960	0.8090	0.5740
	nnU-Net (Isensee et al., 2020)	0.9173	0.9984	0.8708	0.4411
	BANet (Hatamizadeh et al., 2019)	0.8821	0.9983	0.8256	0.5429
	RE-Net (Zhang et al., 2020a)	0.9066	0.9979	0.8574	0.4462
	Ours	0.9212	0.9985	0.8614	0.4369
MIDAS-II	3D U-Net (Çiçek et al., 2016)	0.6518	0.9985	0.6569	1.1500
	V-Net (Milletari et al., 2016)	0.6720	0.9983	0.6528	1.1354
	3D U-Net ++ (Zhou et al., 2019)	0.6709	0.9984	0.6625	1.1304
	Uception (Sanches et al., 2019)	0.6343	0.9982	0.6305	1.1775
	CS²-Net (Mou et al., 2020)	0.6691	0.9983	0.6322	1.1718
	nnU-Net (Isensee et al., 2020)	0.6703	0.9985	0.6624	1.1402
	BANet (Hatamizadeh et al., 2019)	0.6343	0.9982	0.6201	1.1793
	RE-Net (Zhang et al., 2020a)	0.6785	0.9983	0.6837	1.1045
	Ours	0.7134	0.9987	0.6876	1.0713

5. Experimental results

5.1. Segmentation results in different datasets

In this section, we compare our proposed method on four datasets with six different methods, including five state-of-the-art 3D medical image segmentation networks (3D U-Net (Çiçek et al., 2016), V-Net (Milletari et al., 2016), 3D U-Net++ (Zhou et al., 2019), boundary aware network (BANet) (Hatamizadeh et al., 2019), nnU-Net (Isensee et al., 2020)) and three state-of-the-art 3D vessel-like image segmentation networks (Uception (Sanches et al., 2019), CS²-Net (Mou et al., 2020), as well as our previous work RE-Net (Zhang et al., 2020a)). The evaluation results are all obtained based on complete volume images rather than patches.

5.1.1. Segmentation results in cerebrovascular datasets

In this section, we evaluate the proposed ER-Net on cerebrovascular MRA images. The qualitative results are presented, and the visual results obtained by the different methods are provided in Fig. 5. The upper two rows show the segmentation results on two randomly selected MRA volumes from MIDAS-I, and the lower two rows present the segmentation results on two randomly selected MRA volumes from MIDAS-II. Overall, all the methods demonstrate similar performance in segmenting significant vessels. However, for thin/small vessels with small diameters, the selected comparison methods suffer from either undersegmentation or oversegmentation, as indicated by the green arrows. For the visual results from MIDAS-I, V-Net, CS²-Net and U-Net++ can detect the most significant vessels but they performed poorly in identifying the continuity of the thin vessels, which is manifested in

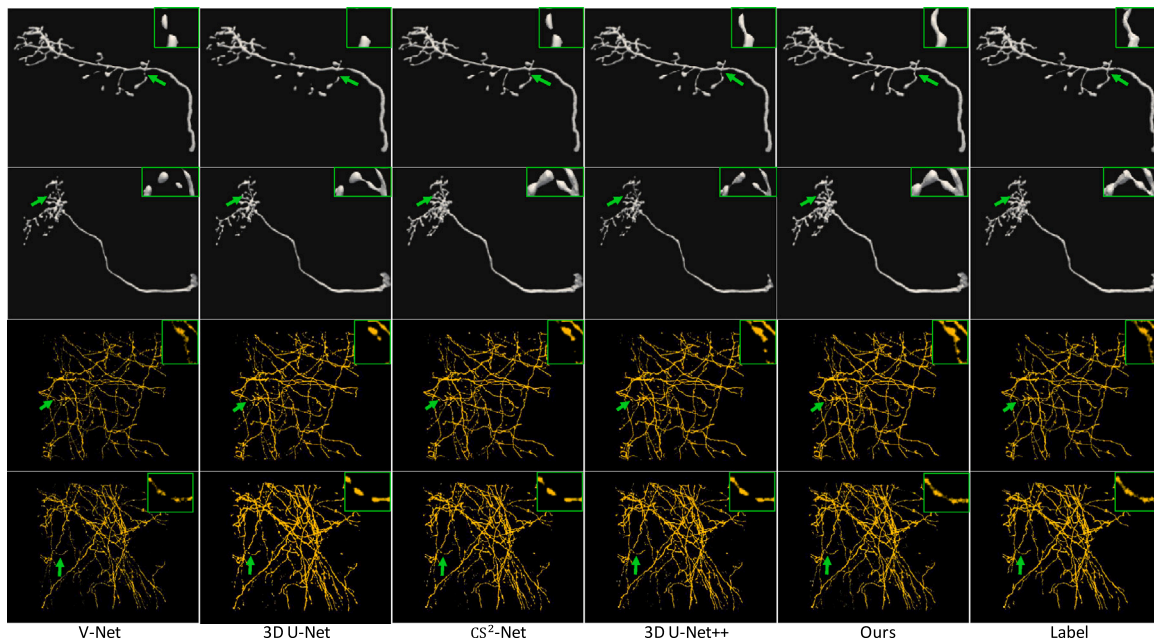


Fig. 6. Segmentation results on nerve datasets by different methods: V-Net (Milletari et al., 2016), 3D U-Net (Çiçek et al., 2016), CS²-Net (Mou et al., 2020) and 3D U-Net++ (Zhou et al., 2019).

the fracture of the vessel junctions. 3D U-Net falsely enhances the background features, resulting in false-positive results. In contrast, the proposed ER-Net achieved better segmentation results than the others in preserving the thin vessels and maintaining vessel continuity, as highlighted by the green arrows with close-ups shown in the green rectangles in Fig. 5. For the visual results from MIDAS-II, all methods present visually appealing results, and V-Net and 3D U-Net misdetect vessels with small diameters, which leads to a relatively lower sensitivity. In addition, CS²-Net and 3D U-Net++ incorrectly detect parts of the background (indicated by the green arrows) as vessels, resulting in oversegmentation. In contrast, the proposed ER-Net adaptively selects features to increase the weight of the edge voxels and constrains both the edge and nonedge voxels of vessel-like structures. Hence, it shows superior performance in detecting thin/small vessels and provides higher sensitivity. Above all, it can be confirmed that ER-Net is beneficial for vessel-like detection in MRA images.

The above findings are reinforced by the evaluation scores in Table 1, and the proposed method in the MIDAS-I dataset achieves 92.12%, 99.85%, 86.14% and 0.4369 in terms of Sen, Spe, DSC and AHD, respectively, which are approximately 2.28%, 0.25%, 5.24%, 0.1371 better than CS²-Net. The above segmentation results in the MIDAS-I dataset show the performance of our method in segmenting large arteries. Moreover, the performance of segmenting microlevel vessels in the MIDAS-II dataset is verified by the experimental results in Table 1. Our method achieves the best scores: 71.34%, 99.87%, 68.76% and 1.0713 in terms of Sen, Spe, DSC and AHD, respectively, indicating that ER-Net has superior performance in the overall segmentation results. In particular, the Sen score of our previous RE-Net (ER-Net without FSM and ERloss) is approximately 3.49% lower than that of ER-Net, which constitutes additional proof that the FSM and ERloss are beneficial to the performance of cerebrovascular segmentation. The state-of-the-art segmentation network nnU-Net outperforms our method only in terms of the DSC score on the MIDAS-I dataset. However, on the whole, the results on both the MIDAS-I and MIDAS-II datasets reported in Table 1 show that our method is superior to nnU-Net. The reason that contributes to this situation may be that cerebrovascular dataset properties of sophisticated vessel geometry and topology variations are yet unconsidered by nnU-Net, which may cause suboptimal segmentation performance. The values of the Spe metric

in both datasets are extremely high ($\geq 99\%$) for all methods, and the value is slightly different among various methods. The same is true for the other two OPF and NL1 A datasets. The reason for this is that the background voxels (true negatives) are dominant in the calculation of SPE, and the target of our segmentation task (such as vessels) is highly sparse. Accordingly, the best Sen and AHD scores obtained by our proposed method reflect the best ability to identify vessel voxels. The lowest AHD reflects the highest classification ability of our proposed method on edge part voxels. Therefore, it can be confirmed that the proposed ER-Net is more conducive to the extraction of the cerebrovascular edge part.

5.1.2. Segmentation results in nerve datasets

We further verify the effectiveness of ER-Net on two nerve datasets: OPF and NL1 A. A random crop operation with a size of $64 \times 128 \times 256$ is adopted to reduce the training cubes, and the batch size is set to 2 during training. To investigate how the proposed method behaves in detecting such vessel-like structures in noise-corrupted data, Gaussian noise with a variance of 10 ($\sigma^2 = 10$), is added to the OPF dataset.

Similar to the evaluation on the cerebrovascular MRA images, we again employ 3D U-Net, V-Net, 3D U-Net++, Uception, CS²-Net and RE-Net as the baselines for comparison. The upper two rows in Fig. 6 illustrate the segmentation results on two randomly selected volumes from the OPF dataset, and the lower two rows illustrate the segmentation results on two randomly selected volumes from the NL1 A dataset. All the methods can generate the skeleton of the nerve. However, segmentation results on weak/thin nerve regions received diverse performances. For visual results from OPF, compared with other methods, our ER-Net ensures continuous fiber tracing in the thin nerve regions and represents more resistance to the interference caused by the added noise, as indicated by the representative regions (green boxes in the upper right corner of each picture) in Fig. 6. For visual results from NL1 A, other comparison methods have difficulty segmenting weak/thin nerves, and these nerves appear to be broken, mainly due to a failure to recognize the edge effectively. In contrast, more completed tiny nerves have been identified by ER-Net. Compared with cerebrovascular, the 3D spatial structure of nerves is relatively simple, so the difference in the visualization results of using different methods may not be that very significant. Thus, the visualization results shown in

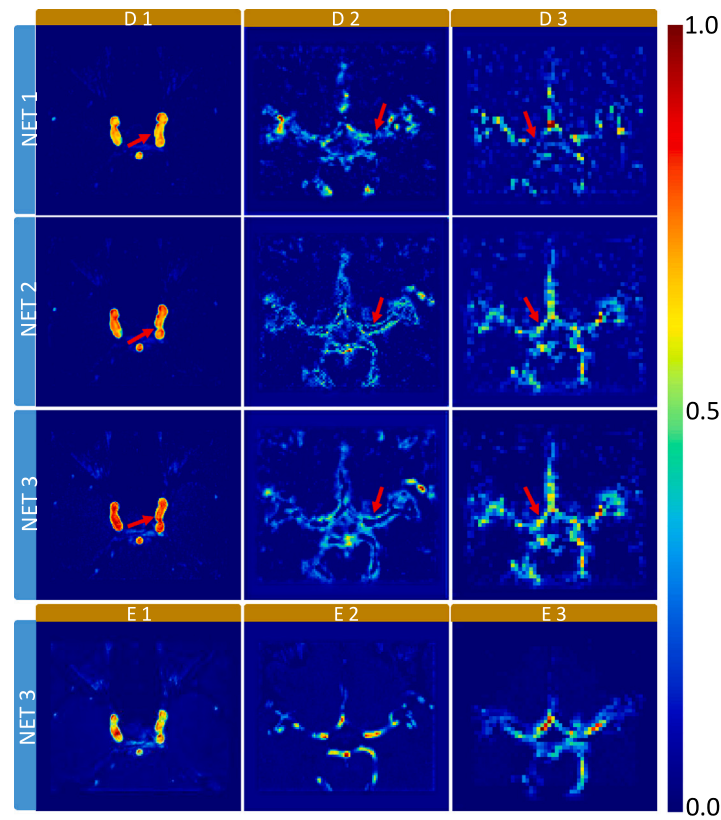


Fig. 7. Examples of attention feature map. Three encoders (E1, E2, and E3) and decoders (D1, D2, and D3) accordingly represent the output feature map in the encoding and decoding path of the network, respectively.

Fig. 6 cannot sufficiently demonstrate the advantages of the proposed network over other methods, while the results in Fig. 5 are much better. In summary, the nerve segmented by ER-Net achieves better continuity and integrity, which indicates better discrimination at the edge of the nerve structure.

It is difficult to conclusively demonstrate the superiority of ER-Net purely by the above visual inspection. Therefore, we compare various methods through a series of evaluation metrics on these nerve datasets. The quantitative results are reported in Table 2. For the OPF dataset, the proposed method achieves the highest segmentation performance in terms of all the metrics since it employs an edge-reinforced mechanism and feature selection mechanism to build optimal representations of the vessel structures. In detail, the ER-Net outperforms 3D U-Net, V-Net, 3D U-Net ++, Uception, CS²-Net, nnU-Net, BANet and RE-Net by 10.05%, 10.75%, 13.80%, 0.47%, 7.88%, 1.29%, 13.04% and 5.00% in terms of Sen, respectively, which indicates that ER-Net can obtain more neural details and can better distinguish nerves from the complex background noise in the microscope images. For the NL1 A dataset, ER-Net achieves 82.95% and 0.8532 better DSC and AHD values, which are approximately 2.37% and 0.0011 better than those of nnU-Net. It can be inferred from the above experiments that nnU-Net may not be suitable for the segmentation of 3D vessel-like structures due to its complex geometric space structure and difficult edge extraction. Likewise, BANet (the network with edge-aware attention mechanisms) performs poorly compared to our method on these four datasets, which constitutes additional proof that edge information is difficult to extract due to the slender shape in the 3D vessel-like structures. The segmentation results of the vessel-like structure are heavily affected by the extraction of the edge part. In contrast, thanks to the edge reinforced mechanisms, the proposed ER-Net can detect edges better, capture the microstructure and improve structure connectivity.

5.2. Ablation studies

In the proposed method, we have introduced REAM, FSM and ERloss. To justify the effectiveness of these modules, we conducted ablation studies on one cerebrovascular dataset and one nerve dataset. For simplicity, the MIDAS-I dataset and OPF dataset are used. First, we verify the effectiveness of the backbone. We replace the original blocks in the 3D U-Net with ResBlock and denote them as ResU-Net. The results show that ResBlock improves the segmentation results by 0.36% and 0.09% in terms of Sen and DSC on the MIDAS-I datasets, respectively, as shown in Table 3. In addition, as shown in Table 4, the Sen, DSC and AHD of 3D U-Net are approximately 1.30%, 0.60% and 0.0079 lower than those of ResU-Net on the OPF dataset, which further proves the effectiveness of the backbone.

To verify the effectiveness of the proposed modules, we conducted ablation studies based on ResU-Net. The results are illustrated in Table 5, from which we drew several conclusions.

5.2.1. Ablation study for REAM

Based on ResU-Net (Net1), we verify the effectiveness of REAM using Dice loss. The comparison results in Table 5 show that REAM (Net2) improves the segmentation results on the MIDAS-I dataset by 1.17% and 1.06% in terms of Sen and DSC. In addition, the Sen and DSC of Net1 on the OPF dataset are approximately 7.74% and 3.01% lower than those of Net2, respectively. The AHD of Net1 is greater than that of Net2, which constitutes additional proof that REAM benefits the performance of vessel-like segmentation. To further verify that REAM can extract edge features, we visualize the feature maps of different layers of Net1 (without REAM) and Net2 (with REAM) on the MIDAS-I dataset. These three encoders (E1, E2 and E3) and decoders (D1, D2 and D3) in Fig. 7 represent the output feature map in the encoding and decoding paths, respectively. They have been marked in Fig. 2. Notably, there are multiple channels in each encoder and decoder layer. We only

Table 2
Nerve segmentation results obtained by different methods on OPF and NL1A datasets.

Datasets	Methods	Sen	Spe	DSC	AHD
OPF	3D U-Net (Çiçek et al., 2016)	0.8537	0.9997	0.8293	0.1702
	V-Net (Milletari et al., 2016)	0.8467	0.9996	0.8100	0.2662
	3D U-Net ++ (Zhou et al., 2019)	0.8162	0.9997	0.8591	0.1462
	Uception (Sanches et al., 2019)	0.9495	0.9993	0.8102	0.1948
	CS ² -Net (Mou et al., 2020)	0.8754	0.9994	0.8230	0.2032
	nnU-Net (Isensee et al., 2020)	0.9413	0.9995	0.8543	0.1434
	BANet (Hatamizadeh et al., 2019)	0.8238	0.9996	0.8089	0.2832
	RE-Net (Zhang et al., 2020a)	0.9042	0.9996	0.8603	0.1573
	Ours	0.9542	0.9997	0.8843	0.1214
NL1A	3D U-Net (Çiçek et al., 2016)	0.8367	0.9974	0.7862	1.2325
	V-Net (Milletari et al., 2016)	0.7666	0.9984	0.7357	0.9544
	3D U-Net ++ (Zhou et al., 2019)	0.8285	0.9994	0.8193	1.0401
	Uception (Sanches et al., 2019)	0.8311	0.9990	0.7924	0.9855
	CS ² -Net (Mou et al., 2020)	0.8036	0.9992	0.8042	0.8559
	nnU-Net (Isensee et al., 2020)	0.8200	0.9987	0.8022	0.8521
	BANet (Hatamizadeh et al., 2019)	0.7522	0.9990	0.7188	1.4532
	RE-Net (Zhang et al., 2020a)	0.8349	0.9983	0.8193	0.8433
	Ours	0.8274	0.9995	0.8295	0.8352

Table 3
Ablation study of ResBlock on the MIDAS-I dataset.

Methods	MIDAS-I			
	Sen	Spe	DSC	AHD
U-Net	0.8913	0.9978	0.8459	0.5082
ResU-Net	0.8949	0.9978	0.8468	0.5141

Table 4
Ablation study of ResBlock on the OPF dataset.

Methods	OPF			
	Sen	Spe	DSC	AHD
U-Net	0.8537	0.9997	0.8293	0.1702
ResU-Net	0.8667	0.9997	0.8353	0.1623

show the feature map of one randomly selected channel for illustration, as shown in Fig. 7. The feature map output by the three encoding layers of Net2 in the last row, the feature map F_i output by the i th encoder layer corresponds to the foreground region due to the parameter update mechanism of the neural network. The red region denotes the high probability that it is recognized as a vessel structure. Thanks to REAM, the edge part of the vessel structure is particularly bright, as indicated by the red arrows in the second row, which implies that REAM tends to discover the edge part of the vessel, and richer context and adequate spatial information are captured. In addition, the vessel structure area with REAM is brighter, indicating that the output feature contains more representative information.

5.2.2. Ablation study for FSM

To verify the effectiveness of FSM, we compare it with Net2 (without FSM) using Dice loss, and the segmentation results are illustrated in Table 5. It is observed that the attendance of FSM can improve the segmentation results, with Sen increasing by 0.47% and 2.39% on the MIDAS-I and OPF datasets, respectively, which preliminarily explains that the FSM contributes to obtaining more details of the vessel-like structure. Moreover, AHD is significantly improved on these two datasets after FSM was added. We further visualize feature maps in the proposed Net3 on the MIDAS-I dataset, as shown in Fig. 7. On the one hand, the vessel structures of Net3 become brighter and smoother than those of Net2. This is because features significantly associated with the vessel structures are assigned larger weights owing to the existence of the FSM. On the other hand, the highlighted areas at different spatial locations confirm that the FSM can combine features from the encoder layer and the symmetric decoder layer for optimal representations of vessel structures. In summary, it is inferred from the

comparison of the attention maps of each column that the proposed attention mechanism module, namely, FSM, has a stronger response ability in terms of vascular structure feature extraction.

5.2.3. Ablation study for ERloss

To justify the benefits of ERloss, we compare it with the Dice loss based on Net3. The results in Table 5 show that our model achieves the best performance in terms of Sen, DSC and AHD in both the MIDAS-I and OPF datasets. Specifically, the use of ERloss leads to the smallest AHD of 0.4369 and 0.1214 on these two datasets, which demonstrates better edge recognition ability of the network with ERloss. To verify the effectiveness of the adaptive weight update method in \mathcal{L}_{edge} , we compared the adaptive weight with the manual selection of weights. Table 6 presents the performance of the proposed model trained on different weight combinations on the MIDAS-I and OPF datasets. Here, the AHD metric is used to evaluate the performance, as it can reflect the edge error of the segmentation results. It can be seen that the quantitative result of the proposed adaptive weight is better than those of the manually selected weights. Therefore, we conclude that the network with ERloss can effectively enhance the edge features of the vessel-like structures. All these results above suggest that the combination of REAM, FSM and ERloss contributes to the improvement of the segmentation.

6. Conclusion and discussion

3D vessel-like structure segmentation is a fundamental step in understanding disease mechanisms and diagnosing and treating diseases. Although researchers have made considerable research efforts in this field, it is still challenging to effectively detect microvascular or micro nerve structures, primarily with difficulty in extracting crisp edges. In this paper, we have proposed a new 3D vessel-like structure segmentation network based on the encoder-decoder architecture, named ER-Net, which aims to detect edges better, capture the microstructure and improve structure connectivity. There are three modules in ER-Net. (1) REAM is embedded between adjacent layers of the encoder to discover edge features by increasing the weight of voxels on the border. (2) FSM is introduced to adaptively select features from encoder and decoder at the same level, and features significantly associated with vessel-like structures are assigned larger weights. (3) A novel ERloss is introduced to constrain both the edge and nonedge voxels of the predicted vessel-like structure, thus better preserving the spatial edge information and significantly improving segmentation performance. The experimental results over four datasets have demonstrated that ER-Net can improve segmentation results in terms of Sen, DSC, Spe and AHD compared to state-of-the-art methods. The positive results imply

Table 5

Ablation studies of proposed modules on both MIDAS-I and OPF datasets.

Design	Backbone	Modules			MIDAS-I				OPF			
	ResUNet	REAM	FSM	ERloss	Sen	Spe	DSC	AHD	Sen	Spe	DSC	AHD
Net1	✓				0.8949	0.9978	0.8468	0.5141	0.8667	0.9997	0.8353	0.1623
Net2	✓	✓			0.9066	0.9979	0.8574	0.4462	0.9042	0.9996	0.8603	0.1573
Net3	✓	✓	✓		0.9113	0.9979	0.8593	0.4442	0.9281	0.9996	0.8562	0.1443
Ours	✓	✓	✓	✓	0.9212	0.9985	0.8614	0.4369	0.9542	0.9997	0.8843	0.1214

Table 6Comparison of loss weights in the edge loss (\mathcal{L}_{edge}) that were selected manually or using the proposed adaptive weight method.

Loss weights		Performance (AHD)	
κ	τ	MIDAS-I	OPF
1	1	0.4382	0.1312
1	1.5	0.4377	0.1299
1.5	1	0.4421	0.1360
Learned weights		0.4369	0.1214

that the proposed method has great potential in applying computer-aided diagnosis of medical images and automatic interpretation of biological images.

Many attention mechanisms have been proposed to make the network pay more attention to the edges to improve the segmentation performance. However, most of the existing edge-attention mechanisms are based on convolutional pyramid features and multipath aggregation to generate accurate edges, as mentioned in Liu et al. (2022), Wang et al. (2019). As proposed in Wang et al. (2019), the multiscale features of the object can be extracted through the multiscale module, and then the edge can be optimized, which is an indirect edge optimization operation. Such an operation lacks a certain interpretability. In our proposed module, edge features are directly extracted by REAM. The probability of a voxel weight corresponding to reverse is obtained by subtracting the predicted probability of the feature map from 1. The voxel containing an object with a higher original weight decreases, while that with a lower original weight increases. The intersection of the foreground and background of different layers can be regarded as the target edge feature. In addition, unlike the attention mechanism proposed in Zhang et al. (2019), Chen et al. (2020), this operation directly carries out automatic edge extraction in the network layer rather than adding additional auxiliary edge information. In addition, edge information is more difficult to extract due to the slender shape in the 3D vessel-like structure of the medical images. The extraction of edge information in such a structure is a relatively unexplored topic. To verify that REAM can effectively extract edge features, we visualized the feature map shown in Fig. 7, from which it is obvious that the vessel edge features are more visible.

It is observed that for almost all methods, the performance on MIDAS-I is much better than that on MIDAS-II, and similarly, the performance on OPF is much better than that on NL1 A. The reasons for the above phenomenon are as follows: compared with the MIDAS-I dataset, the ground truth quality of MIDAS-II is relatively poorer than that of MIDAS-I. We observed that the manual annotations of MIDAS-II do not cover entire vessel trees, and there are cases of vascular breaks (discontinuity), which lead to low segmentation metrics. The reasons why the performance on OPF is much better than that on NL1 A are similar to those described above. It is worth mentioning that the quality of the manual annotations directly determines the final segmentation performance.

Although our ER-Net has achieved promising results in segmenting 3D vessel-like structures, our model can still be improved in the following aspects. First, the threshold λ in the proposed ERloss needs to be adjusted according to different datasets, which requires subjective experience. To this end, we will develop an automated approach to achieve the adaptive setting of λ in the future. Second, ER-Net

focuses on the segmentation of vessel-like structures but it is not straightforward to adopt the proposed segmentation results in clinical practice. In a real application, the measurement of vessel/nerve density, length, tortuosity, or calibers is usually considered to guide clinical decision-making. Therefore, developing a multitask network that not only segments the vessel-like structure but also quantifies geometric biomarkers is preferred to promote the diagnosis and treatment of related diseases. Last, the segmentation algorithm proposed in this paper is based on supervised learning, which requires corresponding annotated data. However, deep learning models have high requirements for the quantity and quality of data, whereas annotated data are often difficult to obtain in the medical field. To address this issue, it is important to achieve positive results using semi-supervised learning, transfer learning, or active learning. In addition, data augmentation should also be considered to address the issue by using a generative adversarial network (GAN)-like method (Kazemini et al., 2020).

CRediT authorship contribution statement

Likun Xia: Formal analysis, Data curation, Writing – original draft, Writing – review & editing. **Hao Zhang:** Formal analysis, Data curation, Writing – original draft, Writing – review & editing. **Yufei Wu:** Formal analysis, Data curation. **Ran Song:** Conception and design of study. **Yuhui Ma:** Formal analysis. **Lei Mou:** Formal analysis, Data curation. **Jiang Liu:** Conception and design of study, review & editing. **Yixuan Xie:** Formal analysis, Data curation. **Ming Ma:** Writing – original draft, Writing – review & editing. **Yitian Zhao:** Conception and design of study, Formal analysis, Data curation, Writing – original draft, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported by the Zhejiang Provincial Natural Science Foundation, China (LR22F020008, LZ19F010001), in part by the Youth Innovation Promotion Association CAS, China (2021298), Beijing Natural Science Foundation, China (4202011), National Natural Science Foundation of China (61572076, 61772351), Ningbo major science and technology task project, China (2021Z054, 2021Z134), Ningbo Natural Science Foundation, China (202003N4039, 202003N4040) and a Key Research Grant of the Academy for Multidisciplinary Studies of CNU, China (JCKXYJY2019018). All authors have participated in the approval of the final version.

References

- Aylward, S.R., Bullitt, E., 2002. Initialization, noise, singularities, and scale in height ridge traversal for tubular object centerline extraction. *IEEE Trans. Med. Imaging* 21 (2), 61–75.
- Bogunović, H., Pozo, J.M., Villa-Urriol, M.C., Majoie, C.B., van den Berg, R., Gratama van Andel, H.A., Macho, J.M., Blasco, J., San Román, L., Frangi, A.F., 2011. Automated segmentation of cerebral vasculature with aneurysms in 3DRA and TOF-MRA using geodesic active regions: An evaluation study. *Med. Phys.* 38 (1), 210–222.

- Callara, A.L., Magliaro, C., Ahluwalia, A., Vanello, N., 2020. A smart region-growing algorithm for single-neuron segmentation from confocal and 2-photon datasets. *Front. Neuroinformatics* 14, 9.
- Cao, Y.-J., Lin, C., Li, Y.-J., 2020. Learning crisp boundaries using deep refinement network and adaptive weighting loss. *IEEE Trans. Multimed.*
- Chen, X., Lian, Y., Jiao, L., Wang, H., Gao, Y., Lingling, S., 2020. Supervised edge attention network for accurate image instance segmentation. In: *European Conference on Computer Vision*. Springer, pp. 617–631.
- Chen, S., Tan, X., Wang, B., Hu, X., 2018. Reverse attention for salient object detection. In: *Proceedings of the European Conference on Computer Vision (ECCV)*. pp. 234–250.
- Çiçek, Ö., Abdulkadir, A., Lienkamp, S.S., Brox, T., Ronneberger, O., 2016. 3D U-net: learning dense volumetric segmentation from sparse annotation. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, pp. 424–432.
- Forkert, N.D., Schmidt-Richberg, A., Fiehler, J., Illies, T., Möller, D., Säring, D., Handels, H., Ehrhardt, J., 2013. 3D cerebrovascular segmentation combining fuzzy vessel enhancement and level-sets with anisotropic energy weights. *Magn. Reson. Imaging* 31 (2), 262–271.
- Frangi, A.F., Niessen, W.J., Vincken, K.L., Viergever, M.A., 1998. Multiscale vessel enhancement filtering. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, pp. 130–137.
- Fu, F., Wei, J., Zhang, M., Yu, F., Xiao, Y., Rong, D., Shan, Y., Li, Y., Zhao, C., Liao, F., et al., 2020. Rapid vessel segmentation and reconstruction of head and neck angiograms using 3D convolutional neural network. *Nature Commun.* 11 (1), 1–12.
- Gu, Z., Cheng, J., Fu, H., Zhou, K., Hao, H., Zhao, Y., Zhang, T., Gao, S., Liu, J., 2019. Ce-net: Context encoder network for 2d medical image segmentation. *IEEE Trans. Med. Imaging* 38 (10), 2281–2292.
- Hatamizadeh, A., Terzopoulos, D., Myronenko, A., 2019. End-to-end boundary aware networks for medical image segmentation. In: *International Workshop on Machine Learning in Medical Imaging*. Springer, pp. 187–194.
- He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. pp. 770–778.
- Isensee, F., Jäger, P.F., Full, P.M., Vollmuth, P., Maier-Hein, K.H., 2020. nnU-net for brain tumor segmentation. In: *International MICCAI Brainlesion Workshop*. Springer, pp. 118–132.
- Katouzian, A., Angelini, E.D., Carlier, S.G., Suri, J.S., Navab, N., Laine, A.F., 2012. A state-of-the-art review on segmentation algorithms in intravascular ultrasound (IVUS) images. *IEEE Trans. Inf. Technol. Biomed.* 16 (5), 823–834.
- Kazemian, S., Baur, C., Kuijper, A., van Ginneken, B., Navab, N., Albarqouni, S., Mukhopadhyay, A., 2020. Gans for medical image analysis. *Artif. Intell. Med.* 101938.
- Kingma, D.P., Ba, J., 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.
- Li, P., Liu, P., Chen, C., Duan, H., Qiao, W., Ognami, O.H., 2018. The 3D reconstructions of female pelvic autonomic nerves and their related organs based on MRI: a first step towards neuronavigation during nerve-sparing radical hysterectomy. *European Radiology* 28 (11), 4561–4569.
- Li, Q., Shen, L., 2019. 3D neuron reconstruction in tangled neuronal image with deep networks. *IEEE Trans. Med. Imaging* 39 (2), 425–435.
- Liu, H., Yang, Z., Zhang, H., Wang, C., 2022. Edge detection with attention: From global view to local focus. *Pattern Recognit. Lett.*
- Livne, M., Rieger, J., Aydin, O.U., Taha, A.A., Akay, E.M., Kossen, T., Sobesky, J., Kelleher, J.D., Hildebrand, K., Frey, D., et al., 2019. A U-Net deep learning framework for high performance vessel segmentation in patients with cerebrovascular disease. *Front. Neuroscience* 13, 97.
- Ma, Y., Hao, H., Xie, J., Fu, H., Zhang, J., Yang, J., Wang, Z., Liu, J., Zheng, Y., Zhao, Y., 2020. ROSE: a retinal OCT-angiography vessel segmentation dataset and new model. *IEEE Trans. Med. Imaging* 40 (3), 928–939.
- Milletari, F., Navab, N., Ahmadi, S.-A., 2016. V-net: Fully convolutional neural networks for volumetric medical image segmentation. In: *2016 Fourth International Conference on 3D Vision (3DV)*. IEEE, pp. 565–571.
- Mou, L., Zhao, Y., Chen, L., Cheng, J., Gu, Z., Hao, H., Qi, H., Zheng, Y., Frangi, A., Liu, J., 2019. CS-Net: channel and spatial attention network for curvilinear structure segmentation. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, pp. 721–730.
- Mou, L., Zhao, Y., Fu, H., Liu, Y., Cheng, J., Zheng, Y., Su, P., Yang, J., Chen, L., Frangi, A.F., et al., 2021. CS2-Net: Deep learning segmentation of curvilinear structures in medical imaging. *Med. Image Anal.* 67, 101874.
- Mou, L., Zhao, Y., Fu, H., Liux, Y., Cheng, J., Zheng, Y., Su, P., Yang, J., Chen, L., Frangi, A.F., et al., 2020. CS2-Net: Deep learning segmentation of curvilinear structures in medical imaging. *Med. Image Anal.* 101874.
- Oktay, O., Schlemper, J., Folgoc, L.L., Lee, M., Heinrich, M., Misawa, K., Mori, K., McDonagh, S., Hammerla, N.Y., Kainz, B., et al., 2018. Attention u-net: Learning where to look for the pancreas. *arXiv preprint arXiv:1804.03999*.
- Otsu, N., 1979. A threshold selection method from gray-level histograms. *IEEE Trans. Syst. Man Cybern.* 9 (1), 62–66.
- Peng, H., Hawrylycz, M., Roskams, J., Hill, S., Spruston, N., Meijering, E., Ascoli, G.A., 2015. BigNeuron: large-scale 3D neuron reconstruction from optical microscopy images. *Neuron* 87 (2), 252–256.
- Phellan, R., Peixinho, A., Falcão, A., Forkert, N.D., 2017. Vascular segmentation in tof mra images of the brain using a deep convolutional neural network. In: *Intravascular Imaging and Computer Assisted Stenting, and Large-Scale Annotation of Biomedical Data and Expert Label Synthesis*. Springer, pp. 39–46.
- Rivest-Henault, D., Cheriet, M., 2013. 3-D curvilinear structure detection filter via structure-ball analysis. *IEEE Trans. Image Process.* 22 (7), 2849–2863.
- Ronneberger, O., Fischer, P., Brox, T., 2015. U-net: Convolutional networks for biomedical image segmentation. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, pp. 234–241.
- Salazar-Gonzalez, A., Kaba, D., Li, Y., Liu, X., 2014. Segmentation of the blood vessels and optic disk in retinal images. *IEEE J. Biomed. Health Inf.* 18 (6), 1874–1886.
- Sanches, P., Meyer, C., Vigon, V., Naegel, B., 2019. Cerebrovascular network segmentation of MRA images with deep learning. In: *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*. IEEE, pp. 768–771.
- Smith, S.M., 2002. Fast robust automated brain extraction. *Hum. Brain Mapp.* 17 (3), 143–155.
- Tetteh, G., Efremov, V., Forkert, N.D., Schneider, M., Kirschke, J., Weber, B., Zimmer, C., Piraud, M., Menze, B.H., 2018. Deepvesselnet: Vessel segmentation, centerline prediction, and bifurcation detection in 3-d angiographic volumes. *arXiv preprint arXiv:1803.09340*.
- Wang, X., Han, S., Chen, Y., Gao, D., Vasconcelos, N., 2019a. Volumetric attention for 3D medical image segmentation and detection. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, pp. 175–184.
- Wang, Y., Narayanaswamy, A., Tsai, C.-L., Roysam, B., 2011. A broadly applicable 3-D neuron tracing method based on open-curve snake. *Neuroinformatics* 9 (2), 193–217.
- Wang, Y., Zhao, S., Shen, J., Hoi, S.C., Borji, A., 2019. Salient object detection with pyramid attention and salient edges. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. pp. 1448–1457.
- Xia, S., Zhu, H., Liu, X., Gong, M., Huang, X., Xu, L., Zhang, H., Guo, J., 2019. Vessel segmentation of X-ray coronary angiographic image sequence. *IEEE Trans. Biomed. Eng.* 67 (5), 1338–1348.
- Yang, B., Chen, W., Luo, H., Tan, Y., Liu, M., Wang, Y., 2020. Neuron image segmentation via learning deep features and enhancing weak neuronal structures. *IEEE J. Biomed. Health Inf.* 25 (5), 1634–1645.
- Yang, X., Cheng, K.T., Chien, A., 2014. Geodesic active contours with adaptive configuration for cerebral vessel and aneurysm segmentation. In: *2014 22nd International Conference on Pattern Recognition*. IEEE, pp. 3209–3214.
- Yang, C., Zhou, X., Zhu, W., Xiang, D., Chen, Z., Yuan, J., Chen, X., Shi, F., 2021. Multi-discriminator adversarial convolutional network for nerve fiber segmentation in confocal corneal microscopy images. *IEEE J. Biomed. Health Inf.*
- Zhang, Z., Fu, H., Dai, H., Shen, J., Pang, Y., Shao, L., 2019. Et-net: A generic edge-attention guidance network for medical image segmentation. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, pp. 442–450.
- Zhang, H., Xia, L., Song, R., Yang, J., Hao, H., Liu, J., Zhao, Y., 2020a. Cerebrovascular segmentation in MRA via reverse edge attention network. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, pp. 66–75.
- Zhang, J., Xie, Y., Wang, Y., Xia, Y., 2020b. Inter-slice context residual learning for 3D medical image segmentation. *IEEE Trans. Med. Imaging*.
- Zhao, H., Shi, J., Qi, X., Wang, X., Jia, J., 2017. Pyramid scene parsing network. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. pp. 2881–2890.
- Zhao, Y., Zhang, J., Pereira, E., Zheng, Y., Su, P., Xie, J., Zhao, Y., Shi, Y., Qi, H., Liu, J., et al., 2020. Automated tortuosity analysis of nerve fibers in corneal confocal microscopy. *IEEE Trans. Med. Imaging* 39 (9), 2725–2737.
- Zhao, Y., Zheng, Y., Liu, Y., Zhao, Y., Luo, L., Yang, S., Na, T., Wang, Y., Liu, J., 2017b. Automatic 2-D/3-D vessel enhancement in multiple modality images using a weighted symmetry filter. *IEEE Trans. Med. Imaging* 37 (2), 438–450.
- Zhou, Z., Siddiquee, M.M.R., Tajbakhsh, N., Liang, J., 2019. Unet++: Redesigning skip connections to exploit multiscale features in image segmentation. *IEEE Trans. Med. Imaging* 39 (6), 1856–1867.