Medical Image Analysis (2020)

Contents lists available at ScienceDirect

Medical Image Analysis

journal homepage: www.elsevier.com/locate/media

CS^2-Net: Curvilinear Structure Segmentation Network for Medical Images

Lei Mou[a,b], Yitian Zhao[a,*], Huazhu Fu[c], Yonghuai Liu[d], Jun Cheng[e], Yalin Zheng[f,a], Pan Su[a], Jianlong Yang[a], Li Chen[g], Alejandro F Frangi[a,j,h], Jiang Liu[b,a,i,*]

[a] *Cixi Institute of Biomedical Engineering, Ningbo Institute of Materials Technology and Engineering, Chinese Academy of Sciences, Ningbo, China*
[b] *Department of Computer Science and Engineering, Southern University of Science and Technology, Shenzhen, China*
[c] *Inception Institute of Artificial Intelligence, Abu Dhabi, United Arab Emirates*
[d] *Department of Computer Science, Edge Hill University, Ormskirk, UK*
[e] *UBTech Research, UBTech Robotics Corp Ltd, Shenzhen, China*
[f] *Department of Eye and Vision Science, University of Liverpool, Liverpool, UK*
[g] *School of Computer Science and Technology, Wuhan University of Science and Technology, Wuhan, China*
[h] *Centre for Computational Imaging and Simulation Technologies in Biomedicine (CISTIB), School of Computing and School of Medicine, University of Leeds, Leeds, UK; Leeds Institute of Cardiovascular and Metabolic Medicine, School of Medicine, University of Leeds, Leeds, UK*
[i] *Guangdong Provincial Key Laboratory of Brain-inspired Intelligent Computation, Department of Computer Science and Engineering, Southern University of Science and Technology, Shenzhen, China*
[j] *Medical Imaging Research Center (MIRC), University Hospital Gasthuisberg. Cardiovascular Sciences and Electrical Engineering Departments, KU Leuven, Leuven, Belgium*

ARTICLE INFO

Article history:

2000 MSC: 41A05, 41A10, 65D05, 65D17

Keywords: Curvilinear structure, blood vessel, nerve fiber, segmentation, attention, deep neural network

ABSTRACT

The automated detection of curvilinear structures, e.g., blood vessels or nerve fibers, in medical images is becoming a paramount step in facilitating the management of many diseases. Precise measurement of the morphological changes of these curvilinear structures provides indicative information to clinicians for understanding the mechanism, diagnosis, and treatment of many diseases. In this work, we propose a general unified convolutionary neural network for curvilinear structure segmentation of images in various 2D/3D medical imaging modalities. We introduce a novel curvilinear structure segmentation network (CS^2-Net) based on Dual Attention Network, which includes a self-attention mechanism in the encoder and decoder to learn rich hierarchical representations of curvilinear structures. Two types of attention modules - spatial attention and channel attention - are utilized to enhance the inter-class discriminating power and intra-class responsiveness, so as to further integrate local features with their global dependencies and normalization, adaptively. To further facilitate the curvilinear structure segmentation in medical images, we employ a 1×3 and a 3×1 convolutional kernel to capture more boundary feature. In addition, we extend the 2D attention to the 3D field to enhance the network's ability to aggregate depth information of different layers. The proposed curvilinear structure segmentation network is rigorously validated using both 2D and 3D images in six different imaging modalities. Experimental results on nine datasets show that the proposed method outperforms on the whole state-of-the-art algorithms in different metrics.

*Corresponding author:
e-mail: yitian.zhao@nimte.ac.cn (Yitian Zhao), liuj@sustech.edu.cn (Jiang Liu)

1. Introduction

Curvilinear structures are objects with thin, long, and tree-like shapes, and different intensities compared with their surrounding pixels (Bibiloni et al., 2016). In the biomedical field, many studies (Kim and Markoulli, 2018a; Rieber et al., 2006) suggest that geometrical and morphological changes in numerous anatomical curvilinear structures - e.g., retinal blood vessels, cerebral vasculature, or nerve fibers - are closely correlated to the presence of diseases, including diabetes, risk of stroke, hypertension and keratitis.

Acquiring images of the aforementioned anatomical curvilinear structures involves two-dimensional (2D) and three-dimensional (3D) imaging techniques, such as color fundus imaging, optical coherence tomography angiography (OCTA), fluorescence angiogram (FA), confocal microscopy (CM), magnetic resonance angiography (MRA), computed tomography angiography (CTA), etc. The top row of Fig. 1 demonstrates five examples of different medical image types, which include both 2D (Fig. 1(a-d)) and 3D (Fig. 1(e)) images.

As one type of curvilinear structure, retinal blood vessels are an important component of the retina, and the morphology change of their retinal vasculature is closely related to many systemic, metabolic, and hematologic diseases (Annunziata et al., 2016; Ding et al., 2014). Retinal blood vessels are usually observed by color fundus images (Franklin and Rajan, 2014) and OCTA images (de Carlo et al., 2015). Color fundus imaging is only able to reveal the superficial vascular network, while OCTA is a new, non-invasive imaging technique that generates volumetric angiography images, and is capable of visualizing the radial peripapillary and deep capillary networks that are not well-distinguished in color fundus images. Corneal nerve fiber properties such as branching, density, and tortuosity are linked to eye and systemic diseases such as herpes, simplex keratitis and dry eye diseases (Eladawi et al., 2017; Kim and Markoulli, 2018a). In vivo corneal confocal microscopy (CCM) is a common technique for the imaging and inspection of corneal nerve fibers. Early detection of their geometrical and topological changes often helps to reduce the incidence of vision loss and blindness. MRA is an MRI examination of the human brain vessels (cerebral vasculature), which is important for the diagnosis of many serious diseases such as strokes (Liao et al., 2012). Cerebral small vessel deformation plays an indicative role in lacunar strokes and brain hemorrhages and are a leading cause of cognitive decline and functional loss in elderly patients (Cuadrado-Godia et al., 2018).

In consequence, accurate extraction of these curvilinear structures from medical images is often an essential step in quantitative image analysis and computer-aided diagnostic pipelines. The bottom row of Fig. 1 illustrates the manual annotations of five types of medical images. However, manual annotation of these curvilinear structures is an exhaustive time-consuming task for graders, and subject to human error, and thus impractical in high-throughput analysis settings like screening programs or microscopy (Zhao et al., 2018b). In addition, the commercial software available (e.g. ImageJ[1] and TubeTK[2]) still rely heavily on manual refinement. This calls for fast, accurate, and fully automated curvilinear structure extraction methods.

Over the last two decades, we have witnessed the rapid development of curvilinear structure detection methods, especially for blood vessel segmentation, as evidenced by general reviews of 2D vessel segmentation (Fraz et al., 2012; Zhao et al., 2018b), and 3D vessel segmentation (Lesage et al., 2009). Most existing segmentation methods suffer from issues posed by the high degree of anatomical variation across populations, and the varying scales of curvilinear structures within an image. On one hand, noise, poor contrast and low resolution exacerbate these problems. As such, standard image segmentation methods are usually not able to robustly detect all the curvilinear structures of interest. On the other hand, deep learning-based methods have yet to be used to segment retinal vessels in OCTA and most of them are designed for the segmentation of vessels or fibers from one specific biomedical imaging modality. Moreover, most of them are designed specifically for 2D images, and cannot easily be extended to 3D ones. In fact, it has been proven very challenging to design a single curvilinear structure detection method that works well across a variety of medical imaging modalities.

In this paper we introduce a novel Channel and Spatial Attention Network (CS^2-Net) to extract curvilinear structures from images in different imaging modalities. Our work was inspired by Dual Attention Network (DANet) (Fu et al., 2019) that were designed for the segmentation of natural images. While medical images contain more unique features, such as simpler semantics and unitary patterns, we first construct a network backbone based on encoder-decoder framework, and then introduce a 1×3 and a 3×1 convolutional kernel to capture more boundary feature to assist the segmentation of curvilinear structures, rather than only up-sampling the attention features in the last layer of DANet. Such approach is more attractive to the researchers and practitioners, since they do not have to choose a particular method for each imaging modality and it is more applicable to various imaging modalities. It is worth noting that the proposed method is a substantial extension of our previous work (Mou et al., 2019), which focused on 2D curvilinear structure segmentation in medical images only. In this work, we have improved it so that it is applicable to segment the curvilinear structure from both 2D and 3D imaging modalities. We also expand our data pool for evaluation from three biomedical imaging modalities to six with a total of nine different datasets. Overall, this work makes the following contributions:

1) A new curvilinear structure segmentation network is proposed based on dual self-attention modules, which can deal with both 2D and 3D imaging modalities in an unified manner;
2) Two self-attention mechanisms are employed in the channel and spatial spaces to generate attention-aware expressive features. They can enhance the network to capture

[1] https://imagej.nih.gov/ij/
[2] http://tubetk.org/

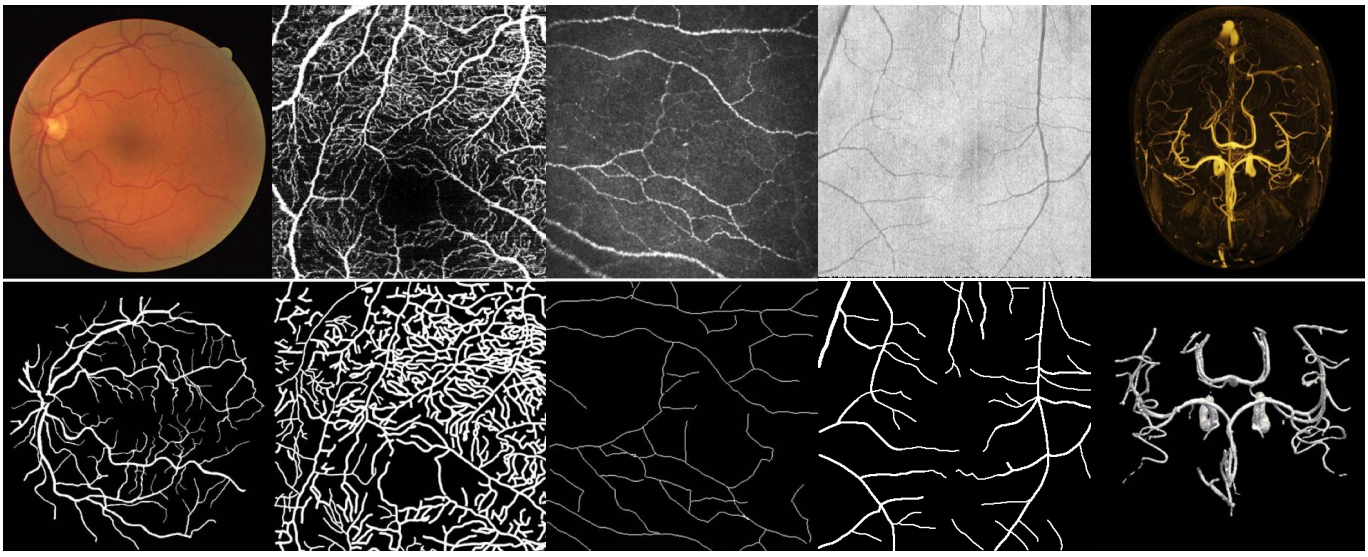

Fig. 1. Images (tow row) and their manual annotations of curvilinear structures (bottom row) in different medical imaging modalities. From the left to right column: Retinal color fundus image; Retinal optical coherence tomography angiogram (OCTA); Corneal confocal microscopy (CCM) image; Optical coherence tomography (OCT) and Brain MRA. Note that the manual annotations of OCTA and CCM are made at a centerline level, and the cerebral vasculatures are visualized in 3D by maximum intensity projection.

long-range dependencies of features at different spatial locations and make full use of the space available for the representation and normalization of the features in different channels, so as to enable the network to classify the curvilinear structure from background more effectively;

3) Experimental results on nine datasets (six 2D datasets and three 3D datasets) demonstrate that our proposed CS^2-Net achieves on the whole state-of-the-art performances in detecting curvilinear structures from different biomedical imaging modalities both quantitatively and qualitatively.

2. Related Works

2.1. 2D Curvilinear Segmentation

As vasculatures or fibers in 2D medical images are curvilinear structures distributed across different orientations and scales, various filtering methods have been proposed, including Hessian matrix-based filters (Frangi et al., 1998a), matched filters (Zhao et al., 2017b), multi-oriented filters (Soares et al., 2006; Zhang et al., 2017), symmetry filter (Zhao et al., 2018b), and tensor-based filter (Cetin and Unal, 2015), active contours-based methods (Shang et al., 2019; Al-Diri et al., 2009) and minimal geodesic paths-based approaches (Chen et al., 2019). These filtering-based methods aim to suppress non-vascular or non-fiber structures and imaging noise, and enhance the curvilinear structures, thereby benefiting the subsequent segmentation problem. For instances, (Zhao et al., 2015, 2018a) proposed infinite perimeter active contour model with hybrid region information and a weighted symmetry filter to detect vessels. (Zhang et al., 2016) designed multi-scale rotation invariant filters for retinal vessel and corneal nerve fiber segmentation based on a locally adaptive framework in the position and orientation spaces. This framework is adaptive to the local changes of curvilinear structures and is able to deal with typically difficult cases. (Soares et al., 2006) used a multi-scale Gabor transform to extract texture features of vessels for more accurate vessel detection. There are also several filter-based vessel segmentation methods, including Hessian matrix-based filters (Frangi et al., 1998a; Zhang et al., 2016), tensor-based filters (Cetin and Unal, 2015) and symmetry filters (Zhao et al., 2018b). These approaches aim to remove undesired intensity variations in the images, and suppress background structures and imaging noise, thereby facilitating the subsequent segmentation task. However, these filter-based methods rely heavily on manual parameter adjustment during implementation, and usually are designed for a specific imaging modality, which may not be effective when applied to other image types.

Recently, deep learning-based methods have made significant progress in the field of computer vision. These include classification networks, e.g., ResNet (He et al., 2016) and Inception series networks (Szegedy et al., 2015; Ioffe and Szegedy, 2015; Szegedy et al., 2016, 2017); object detection networks, e.g., Faster-RCNN (Ren et al., 2015) and R-FCN (Dai et al., 2016); segmentation networks, e.g., SegNet (Badrinarayanan et al., 2017), PSPNet (Zhao et al., 2017a); and networks designed for medical image segmentation, e.g., U-Net (Ronneberger et al., 2015) and CE-Net (Gu et al., 2019). These deep learning based methods have been modified and applied for blood vessel segmentation (Fu et al., 2016; Alom et al., 2018) and nerve fiber tracing in color fundus and CCM images (Colonna et al., 2018; Williams et al., 2020) , respectively. (Maninis et al., 2016) proposed a multi-task structure for both vessel detection and optic disc segmentation. (Liskowski and Krawiec, 2016) introduced a retinal vessel segmentation method based on a convolutional neural network (CNN), and (Fu et al., 2016) further applied the CNN along with conditional

random fields for the detection of retinal vessels. (Alom et al., 2018) embedded a recurrent neural network into the U-shaped network (R2U-Net) for the segmentation of vessels. (Wang and Chung, 2019) proposed a novel detector, named Oriented Cylinder Flux (OCF), for the detection of blood vessel structures. (Wang et al., 2019a) proposed a new curvilinear structure segmentation method using context-aware and spatio-recurrent networks. Instead of directly segmenting the entire image or densely segmenting fixed-size local patches, it uses a learning strategy to repeatedly sample the target image with different proportions. More details on recent vessel segmentation works can be found in (Shin et al., 2019; Jin et al., 2019; Wang et al., 2019c). (Colonna et al., 2018) proposed a deep neural network based on U-Net (Ronneberger et al., 2015) for corneal fiber tracing in CCM images. (Hosseinaee et al., 2019) developed a fully automated segmentation method for the segmentation of corneal nerves on a series of enface UHR-OCT images obtained from healthy human subjects. (Kim and Markoulli, 2018b) and (Oakley et al., 2019) systematically summarized the unsupervised and supervised methods for corneal nerve segmentation and analyzed the role of corneal neuromorphological features in disease diagnosis. Eladawi *et al.* proposed a joint Markov-Gibbs random field (MGRF) model to segment blood vessels based on different retinal maps from OCTA scans. (Díaz et al., 2019) developed a fully automatic system that identifies and precisely segments the foveal avascular zone (FAZ). (Heisler et al., 2019) also proposed a novel automated deep learning method to segment and quantify retinal images from prototype OCTA machines with larger fields of view. For more automated vascular segmentation and fiber tracing methods, please refer to the review by (Fraz et al., 2012). Although these methods have achieved promising segmentation results, most of these methods concentrate on the segmentation of curvilinear structure for single imaging modality. In addition, most of these methods are hard to be extended for the curvilinear structure segmentation in 3D volumes.

2.2. 3D Curvilinear Segmentation

Three-dimensional volumes contain richer features with depth information that is not available in 2D slices/images. Three-dimensional vascular segmentation is an important prior step in the characterization of cerebral aneurysms, which has proven useful for the pre-treatment planning of Guglielmi separable coils (GDC) (Wilson, 1998). With the development of imaging devices, more computer vision methods have been developed to deal with 3D data for biomedical data analysis. Moreover, most methods perform better on this volumetric data compared to 2D image counterpart, especially in medical imaging. (Zhao et al., 2018b) proposed a weighted symmetry filter for automatic 2D vessel enhancement and segmentation, and further extended it to the 3D case for vascular segmentation. (Çiçek et al., 2016) extended U-Net to 3D U-Net with a weighted cross entropy loss to perform Xenopus kidney segmentation, which has been proven an effective method for the segmentation of tubular structural organs under sparse annotations. (Gibson et al., 2018) used a dense V-Net (Milletari et al., 2016) to segment multiple 3D tubular organs. (Chung and Noble, 1999) adopted a Rician distribution to segment 3D brain vasculatures in order to extract cerebral aneurysm features. (Tetteh et al., 2018) proposed DeepVesselNet to segment vessels, detect vessel centerlines and bifurcate 3D angiographic volumes. (Liao et al., 2012) proposed to segment human brain vessels using fast matching with an anisotropic orientation being a priori. Recently, (Zhang et al., 2019) proposed a novel method for 3D retinal OCTA microvascular segmentation and surface reconstruction. Intrinsic shape analysis was performed to extract useful surface-based 3D geometric and topological biomarkers. (Wang et al., 2019b) proposed a teacher-student learning framework for fast neuron segmentation, where the segmentation inference is performed using a light-weighted student network which benefits from knowledge distillation by a teacher network with a higher capacity. (Zhao et al., 2019) proposed to perform 3D vessel segmentation by utilizing a deep feature regression (DFR) method based on a convolutional regression network (CRN) and a stable point clustering mechanism. (Poulain et al., 2019) proposed a new method by combining the information of a tree-spline with a registration algorithm to perform 3D coronary vessel tree tracking. (Sanchesa et al., 2019) proposed a Uception network based on Inception modules and the U-Net-like architecture to segment cerebrovascular in MRA images. However, many modules that rely heavily on GPU resources are used in Uception, which makes the method require huge GPU memory support in the training and inference stages.

Like the previous 2D segmentation methods, many filter-based 3D tubular structure segmentation methods rely heavily on manual tuning. Moreover, many methods based on learning strategies do not pay attention to the tubular structure by designing particular network modules, which plays an important role in their accurate segmentation.

3. Proposed Method

3.1. Network Architecture

The proposed CS^2-Net is designed for curvilinear structure segmentation of both 2D and 3D medical images. It consists of three modules: the encoder module, the channel and spatial attention module (CSAM), and the decoder module. Fig. 2 and Fig. 3 illustrate the architectures for 2D and 3D images, respectively. The encoder module is used to extract the features of input data. Then, these features are fed into two parallel attention blocks - the channel attention block (CAB) and a spatial attention block (SAB) - to generate channel-spatial attention-aware expressive features. The SAB selectively aggregates the features in each spatial location through the weighted sum of the features in all spatial locations, which allows the model to capture the long-range dependency of the features, and similar features will be related to each other regardless of their distance. Meanwhile, the CAB makes sure that the full space is used to represent and normalize and thus enhance the contrast of the features in different channels, allowing the model to be assembled with improved discrimination capabilities. Finally, the decoder module is employed to reconstruct curvilinear features and produce the segmentation result.

Instead of directly upsampling the features of the CSAM to the original image dimensions (Fu et al., 2019), we introduce a feature decoder module that restores the dimensions of the high-level semantic features layer by layer. The encoder and decoder modules include four blocks, each of which employs a residual network (ResNet) as the backbone, and then followed by a max-pooling layer to increase the receptive field for better extraction of global features. Similar to the U-shaped network (Ronneberger et al., 2015; Çiçek et al., 2016), a skip connection between each layer of the encoder and decoder is introduced to combine the features in different levels to compensate for the loss of information caused by the max-pooling operations. At the end of the CS^2-Net, we apply a 1×1 kernel (1×1×1 kernel in the 3D phase) convolutional layer and a sigmoid layer on the output of the encoder to obtain the final segmentation map.

3.2. 2D Attention Network

Several recent works have shown that the local feature representations produced by traditional fully convolutional networks (FCNs) may lead to object misclassification (Zhao et al., 2017a; Peng et al., 2017). The CS^2-Net, which consists of a 2D encoder, 2D CASM and 2D decoder, is designed to reduce this limitation and segment curvilinear structures in 2D images more effectively. The 2D version CSAM is shown in Fig. 2, which includes a 2D SAB and a 2D CAB. We use 2D convolutional, 2D batch normalization and 2D deconvolutional layers in all the modules. Their working principles are explained as follows.

3.2.1. 2D Spatial Attention Module

To model rich contextual dependencies over local feature representations, the first step is to generate a spatial attention matrix, which models spatial relationships between the features of any two pixels. In practice, tree-like structures are always distributed throughout the biomedical images. Following (Fu et al., 2019), we modify the SAB to encode a wider range of contextual information about local features, and increase their representative capability. However, unlike (Fu et al., 2019), we introduce a 3×1 and a 1×3 convolutional layer with batch normalization and ReLU layers to capture the edge information of the tree-like structures in horizontal and vertical orientations, respectively. More importantly, compared with many complex natural images, medical images contain rare and almost fixed structures. Considering this aspect, the curvilinear structure segmentation network requires skip-connection operations to fuse low-level information and compensate for the lost spatial information. Therefore, we transplant the proposed attention module into the encoder-decoder framework, rather than directly encoding the image and resampling the original one as in (Fu et al., 2019).

Specifically, we place the two types of layers (3×1 and 1×3 convolutional layer) after the input features $F \in \mathbb{R}^{C\times H\times W}$ to generate two new feature maps $Q_y \in \mathbb{R}^{C\times H\times W}$, and $K_x \in \mathbb{R}^{C\times H\times W}$, respectively, where C denotes the dimensionality of the input features, and H and W are the height and width of the input image, Q_y and K_x represent the features of the curvilinear structures captured in the vertical and horizontal directions. These two new feature maps are then reshaped to $\mathbb{R}^{C\times N}$, where $N = H\times W$ is the number of features. In consequence, the intra-class spatial association can be obtained by applying a softmax layer on the matrix multiplication of the transpose of Q and K, as:

$$\mathcal{S}_{(x,y)} = \frac{\exp\left(Q^{\mathrm{T}}{}_y \cdot K_x\right)}{\sum_{x'=1}^{N} \exp\left(Q^{\mathrm{T}}{}_y \cdot K_{x'}\right)}, \tag{1}$$

where $\mathcal{S}_{(x,y)}$ denotes the y^{th} position's impact on the x^{th} position. In other words, matrix multiplication computes and outputs the feature correlation matrix $\mathcal{S}_{(x,y)}$ between any two points, that is, the two similar spatial points promote each other and the two dissimilar spatial points suppress each other. Through this operation, the network can fully utilize and learn the curvilinear structure of different spatial locations. Then we apply softmax on the correlation matrix to obtain the attention map of the similarity between each spatial position and the others, in which the higher the similarity, the greater the response between the two points. Meanwhile, another new feature $V \in \mathbb{R}^{C\times H\times W}$ is obtained by applying a 1×1 convolutional layer with batch normalization and ReLU layers on the input features and we also reshape it to $\mathbb{R}^{C\times N}$, which is then used to perform a matrix multiplication with $\mathcal{S}_{(x,y)}$ to obtain the attention enhanced features $F' \in \mathbb{R}^{C\times N}$. Finally, we reshape it to $\mathbb{R}^{C\times H\times W}$, and perform a pixel-level summation of F and F' to construct the output of SAB. Thus, SAB gains a global contextual view and selectively aggregates context information according to the spatial attention map to achieve a more accurate segmentation performance for curvilinear structures.

3.2.2. 2D Channel Attention Module

Since each channel of a high-level feature can be regarded as a specific-class response, we further exploit the interdependencies of channel maps in this section, and propose the CAB module to improve the feature representation by using the space available. A channel-wise attention map is obtained by performing matrix multiplication between the input feature F (named as $F_x \in \mathbb{R}^{C\times H\times W}$) and its transpose (named as $F^{\mathrm{T}} \in \mathbb{R}^{H\times W\times C}$). Then, the channel affinity matrix is obtained by applying a softmax layer on the channel-wise attention map, as:

$$C_{(x,y)} = \frac{\exp\left(F_x \cdot F_y^{\mathrm{T}}\right)}{\sum_{x'=1}^{C} \exp\left(F_{x'} \cdot F_y^{\mathrm{T}}\right)}, \tag{2}$$

where $C_{(x,y)}$ denotes the similarity between the x^{th} channel and the y^{th} channel. Therefore, we can obtain the channel dependency matrix ($\mathbb{R}^{C\times C}$, where C denotes the number of channels) by performing a matrix multiplication. In this case, two similar channels will promote each other, while dissimilar channels will inhibit each other, which also enhances the discrimination between the curvilinear structure and the background. The process is similar to the spatial attention module. The difference lies in two aspects: (i) while the former operates the original features F directly, the latter operates on new derived features Q_y, K_x and V, and (ii) while the former models the attention of the features in one channel on those in another, the latter models

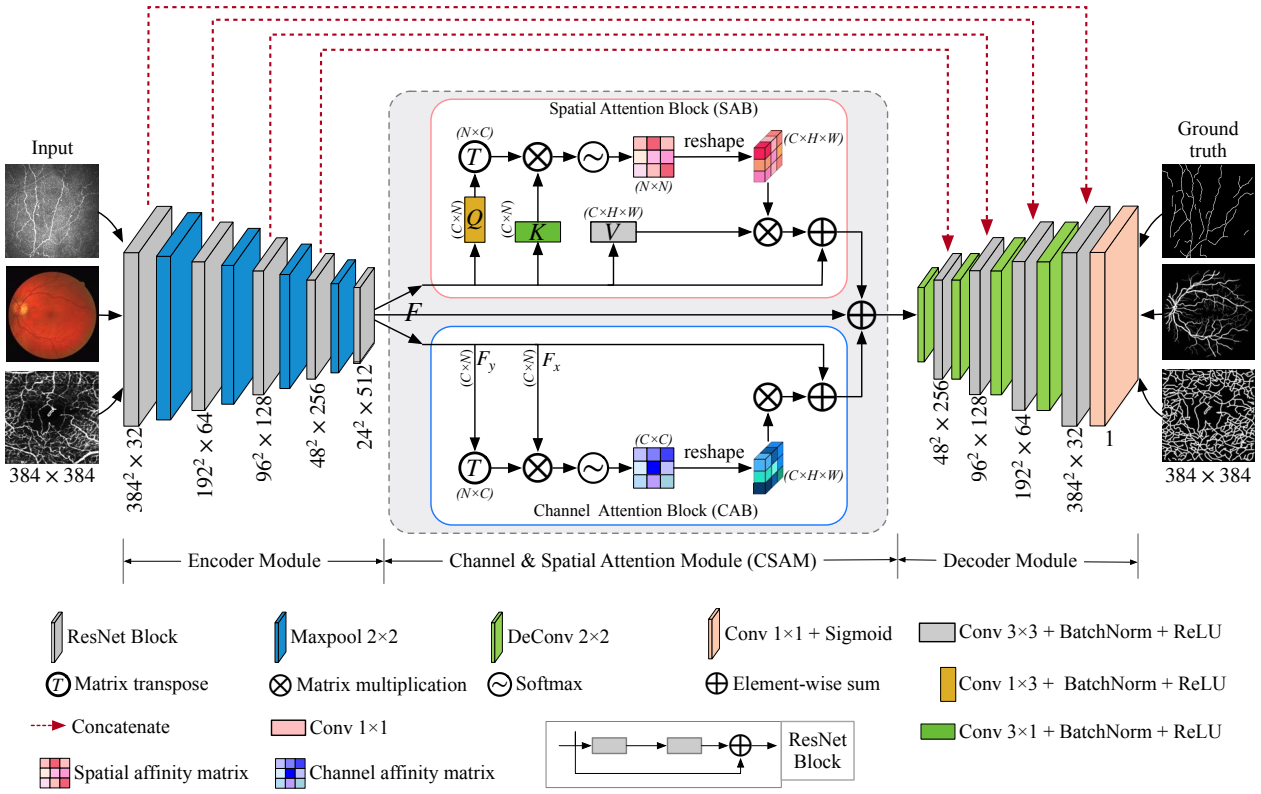

Fig. 2. The architecture of the proposed CS^2-Net over the 2D images: an encoder, a CSAM module and a decoder, which are designed to extract global features, enhance the feature expression ability, and reconstruct curvilinear features, respectively.

the attention of one pixel on that of another. After that, a softmax is applied on the channel dependency matrix to enhance the discrimination between curvilinear structures and backgrounds. Similar to SAB, we then perform a multiplication between $C_{(x,y)}$ and F to obtain the attention enhanced features F''. The final output of CAB is defined as $F + F''$ over each pixel. Such operations tend to enhance the contrast between class-dependent features and help improve their expressiveness.

3.2.3. Objective Function

It is worth noting that all datasets contain complete annotations, and curvilinear structure segmentation in a 2D image can be regarded as a pixel-level binary classification task: curvilinear structure or background. In this work, the binary cross-entropy (BCE) loss is thus adopted as the objective function for the training of the network, as it is a pixel-wise objective function that directly evaluates the distance between the ground truth and prediction. The BCE loss is defined as:

$$\mathcal{L}_{BCE} = -\frac{1}{N}\sum_{i=1}^{N} g_i \cdot \log(p_i) + (1 - g_i) \cdot \log(1 - p_i), \qquad (3)$$

where $g_i \in \{0, 1\}$ indicates the ground truth as curvilinear structure of a pixel, $p_i \in [0, 1]$ is its predicted probability, and N is the number of pixels.

3.3. 3D Attention Network

In recent years, many methods based on learning and manual design have been proposed for the detection of curvilinear structures in 2D images (Staal et al., 2004; Kim and Markoulli, 2018a; Li et al., 2015). However, there are relatively few methods, especially learning-based methods, for segmenting curvilinear structures in 3D images. Moreover, the spatial attention and the channel attention in (Fu et al., 2019) focus only on the 2D domain. Directly applying a 2D attention on 3D images lacks feature integration in the depth direction, which is crucial for improving the results of segmentation of curvilinear structures. To enable our proposed CS^2-Net to extract the 3D tree-like structures, we extend it from the 2D to the 3D, as shown in Fig. 3. For the encoder and decoder modules, we replace all their 2D operations with 3D ones. However, due to changes in the modality of the dataset, the proposed CSAM in the 3D mode is quite different from that in the 2D one. We detail the 3D CSAM in the following section.

3.3.1. 3D Spatial Attention Module

Similar to the 2D SAB, we first feed the input features $F \in \mathbb{R}^{C\times H\times W\times D}$ into a $1\times3\times1$ and $3\times1\times1$ layer with batch normalization and ReLU activations to generate two feature maps $Q_y \in \mathbb{R}^{C\times H\times W\times D}$ and $K_x \in \mathbb{R}^{C\times H\times W\times D}$ to capture the boundary features of tublar structure along y axis and x axis, where C indicates the number of input channels, and H, W and D indicate the height, width and depth of the input 3D image , respectively. However, this operation only encodes the relationship between features in the width and height directions, lacking feature integration in the depth direction. To overcome this limitation, we also feed F into a $1 \times 1 \times 3$ convolutional layer and then opti-

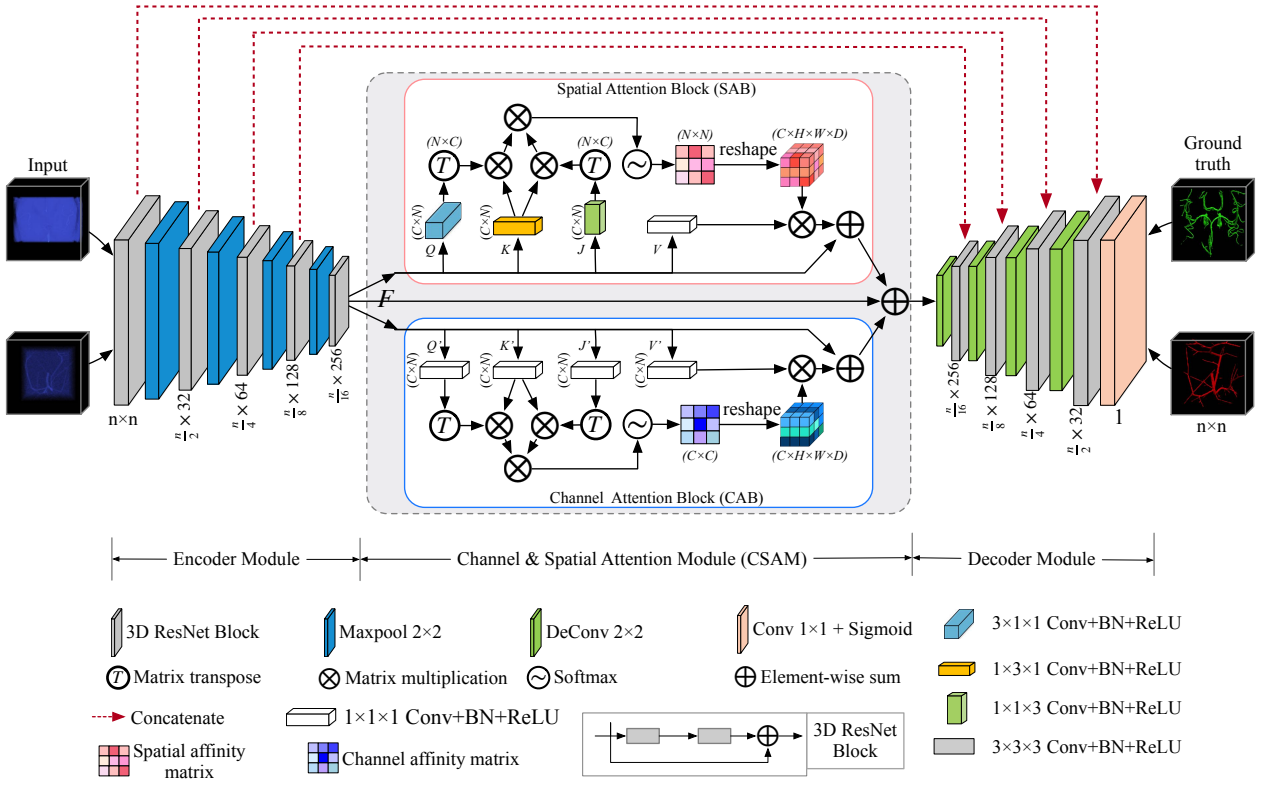

Fig. 3. The diagram of the 3D CS^2-Net. It includes an encoder, a 3D CSAM and a decoder. N is the size of cropped volumes during training.

mize and activate it with batch normalization and ReLU layers. Consequently, a new feature map $J_z \in \mathbb{R}^{C\times H\times W\times D}$ is obtained. Therefore, we use Q_y, K_x and J_z to capture the edge information of tree-like structures in the width, height and depth directions. In the next step, we reshape these three feature maps to $\mathbb{R}^{C\times N}$ to construct three activation matrices, where $N = H \times W \times D$. We first perform matrix multiplication on Q_y^{T} and K_x to encode the feature relationships in the width and height directions, and then perform the same operation on K_x and J_z^{T} to encode the feature relationships in the height and depth directions, where Q_y^{T} and J_z^{T} are the transpose of Q_y and J_z, respectively. To encode the relevance of features in the width and depth directions, we apply a matrix multiplication between the two outputs from the previous step. Finally, a softmax layer is applied to obtain the voxel-level, intra-class affinities:

$$\mathcal{S}_{(x,y,z)} = \frac{\exp\left[\left(Q_y^{\mathrm{T}} \cdot K_x\right) \cdot \left(J_z^{\mathrm{T}} \cdot K_x\right)\right]}{\sum_{x'=1}^{N} \exp\left[\left(Q_y^{\mathrm{T}} \cdot K_{x'}\right) \cdot \left(J_z^{\mathrm{T}} \cdot K_{x'}\right)\right]}, \tag{4}$$

where $\mathcal{S}_{(x,y,z)}$ denotes the mutual impacts of features at the x^{th}, y^{th} and z^{th} positions. Similarly, we gain a dimension-reduced feature map $V \in \mathbb{R}^{C\times H\times W\times D}$ by applying a $1\times1\times1$ kernel convolutional layer on the input feature map F, and we reshape it to $\mathbb{R}^{C\times N}$. Then, a matrix multiplication is performed between V and $\mathcal{S}_{(x,y,z)}$ to obtain the voxel-level enhanced features F', which is then reshaped to $\mathbb{R}^{C\times H\times W\times D}$. Finally, we perform voxel-wise addition of $F' + F$ to get the output of the 3D SAB. The schematic diagram of CSAM in Fig. 3 shows the details of this process. Our proposed 3D SAB not only performs feature mapping in the width and height directions, but also performs the mutual mapping of the 3D features in the depth direction. In the end, it is expected to increase the feature expression ability of the network.

3.3.2. 3D Channel Attention Module

Inspired by the 2D channel attention mechanism, we further extend it to the 3D domain. Similar to the 2D CAB, we apply a $1\times1\times1$ kernel convolutional layer on $F \in \mathbb{R}^{C\times H\times W\times D}$ to obtain four channel attention maps $Q'_y \in \mathbb{R}^{C\times H\times W\times D}$, $K'_x \in \mathbb{R}^{C\times H\times W\times D}$, $J'_z \in \mathbb{R}^{C\times H\times W\times D}$ and $V' \in \mathbb{R}^{C\times H\times W\times D}$, respectively. Then, we reshape Q'_y, K'_x, J'_z and V' to $\mathbb{R}^{C\times N}$. Then we perform the same matrix operations on Q'_y, K'_x and J'_z as in the 3D SAB:

$$C_{(x,y,z)} = \frac{\exp\left[\left(K'_x \cdot Q'^{\mathrm{T}}_y\right) \cdot \left(K'_x \cdot J'^{\mathrm{T}}_z\right)\right]}{\sum_{x'=1}^{C} \exp\left[\left(K'_{x'} \cdot Q'^{\mathrm{T}}_y\right) \cdot \left(K'_{x'} \cdot J'^{\mathrm{T}}_z\right)\right]}, \tag{5}$$

where $C_{(x,y,z)}$ denotes the mutual affinities between the x^{th}, y^{th} and z^{th} channels. In addition, a matrix multiplication is performed between the transpose of C and V' to obtain the voxel-level channel-wise attention output F'' and reshape it to $\mathbb{R}^{C\times H\times W\times D}$. Similarly, we perform a voxel-wise addition of F'' and F to obtain the output of the 3D CAB.

To gather the spatial and channel attention maps, a voxel-level matrix summation is applied as the output of the 3D CSAM between the outputs of the 3D SAB and the 3D CAB and the original input feature F.

3.3.3. Loss Function

In general, the labels for 3D cerebrovascular regions are sparse, and only a portion of them have high-quality annotations. Thus, we choose as our loss function the weighted cross entropy loss L_{WCE} (WCE), which is able to adjust learning bias between a vascularity and background during training. Moreover, we also introduce Dice coefficient loss L_{Dice} to ensure the micro-cerebrovascular segmentation. Finally, we define the 3D optimization loss function for the training of the proposed CS^2-Net as:

$$\mathcal{L} = \alpha L_{WCE} + (1-\alpha) L_{Dice}, \tag{6}$$

where α is the weight balance parameter between L_{MSE} and L_{Dice}, which is empirically set as $\alpha = 0.6$. For our binary segmentation task, the WCE loss and the Dice coefficient loss are defined as follows:

$$L_{WCE} = -\frac{1}{N}\sum_{i=1}^{N}\left(\omega g_i \log p_i + (1-g_i)\log(1-p_i)\right), \tag{7}$$

$$L_{Dice} = 1 - \frac{2\sum_{i=1}^{N} p_i g_i + \epsilon}{\sum_{i=1}^{N} p_i^2 + \sum_{i=1}^{K} g_i^2 + \epsilon}, \tag{8}$$

where ω is the class weight of curvilinear structure, and can be obtained by the class estimation probabilities p_i of all the voxels:

$$\omega = \frac{N - \sum_{i=1}^{N} p_i}{\sum_{i=1}^{N} p_i}.$$

Here, N denotes the number of voxels, and $p_i \in [0,1]$ and $g_i \in \{0,1\}$ denote the predicted probability and ground truth value of the i^{th} voxel as the curvilinear structure, respectively. The parameter ϵ is a Laplace smoothing factor used to avoid numerical instability problems and accelerate the convergence of the training process ($\epsilon = 1.0$ in this paper).

Table 1. Details of the datasets used to evaluate the proposed method.

Datasets	Number	Resolution	Data type	Public
DRIVE	40	565×584	Fundus	Public
STARE	20	605×700	Fundus	Public
IOSTAR	30	1024×1024	Fundus	Public
CORN-1	1698	384×384	CCM	Public
OCTA	30	1376×968	OCTA	Private
OCT RPE	36	384×379	OCT	Private

4. Experimental Results over 2D Images

In this section, the proposed segmentation network is first validated over 2D medical images for the extraction of their curvilinear structures. Many datasets are available online and aim to train and validate an automatic approach for the segmentation of vessels or nerve fibres from 2D medical images, as blood vessels or nerve fibers are closely correlated to the presence of pathology. We refer readers to (Zhao et al., 2018b) for more detailed introduction and discussions. In this work, we selected two most commonly-used (DRIVE and STARE), two newly-released (IOSTAR and CORN-1) publicly available datasets, as well as two private (OCTA and OCT RPE) datasets for evaluation of our method and the competitors.

In this work, we selected the most commonly used datasets in the research community to evaluate the proposed CS^2-Net, so that we can make a direct comparison of segmentation results with those obtained by the state-of-the-art methods. Regarding to two private datasets, OCT and OCTA are two new emerging non-invasive imaging technique, with the ability to produce high-resolution 3D images of retinal vasculature, and has been increasingly taken as a valuable imaging tool to observe retinal vascular. To our best knowledge, there is no publicly available OCTA or OCT RPE dataset with manually graded vessels for training and validation. We use these two datasets to test our model, keep growing in the size of these two datasets, and will release them online in the future.

4.1. Materials

Six 2D datasets in total are used for evaluation, whose details are provided as follows.

DRIVE[3] contains 40 colored fundus images, which were originally divided into 20 images for training and 20 images for testing. The images were acquired using a Canon CR5 non-mydriatic 3-CCD camera with a field of view (FOV) being 45°. Each image in this dataset has dimensions of 565×584. We follow the same partition of the images in our training and testing.

STARE[4] comprises 20 colored fundus images. The images were captured using a Topcon TRV-50 fundus camera with a FOV being 35° . Half of the images contain pathological indications and the other half come from healthy subjects. Each image has dimensions of 700×605. However, unlike the DRIVE dataset above, there is no fixed partition of training and testing sets. In this paper, we adopt the k-fold (k=4) cross-validation method for the training and testing phases, similar to that in (Mo and Zhang, 2017). Therefore, 15 images are used for training and the remaining 5 images are used for testing in each fold. We use the manual annotations from the first observer as the ground truth for all the images.

IOSTAR[5] includes 30 images with a resolution of 1024×1024 pixels. The images were acquired with an EasyScan camera (i-Optics Inc., the Netherlands), which is based on a SLO technique with a FOV being 45 degrees. For reasonable data division, we also adopt the k-fold (k=5) cross-validation method for training and evaluating, that is, 24 images are used for training and 6 images for testing.

CORN-1[6] is a publicly available CCM dataset, and contains a total of 1698 CCM images of corneal subbasal epithelium using a Heidelberg Retina Tomograph equipped with a Rostock Cornea Module (HRT-III) microscope. All of these images were acquired by the Peking University Third Hospital, China and University of Padova, Italy[7]. Each image has a resolution

[3] http://www.isi.uu.nl/Research/Databases/DRIVE/
[4] http://www.ces.clemson.edu/ahoover/stare/
[5] http://www.retinacheck.org/
[6] http://imed.nimte.ac.cn/
[7] http://bioimlab.dei.unipd.it/

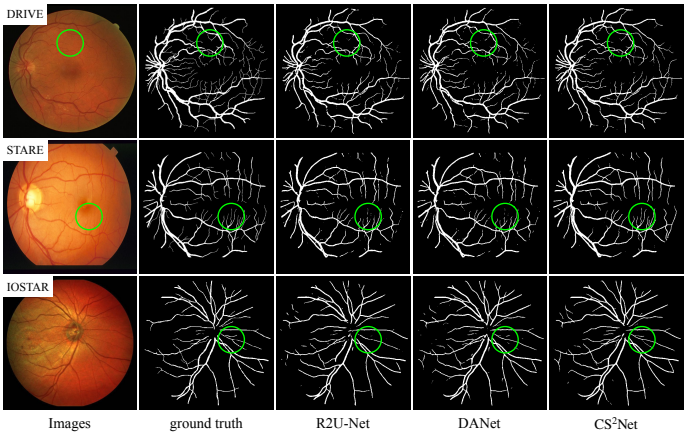

Fig. 4. Retinal vessel segmentation results of three randomly selected images from three different datasets by R2U-Net, DANet and our proposed CS^2-Net respectively.

of 384×384 pixels covering a field of view of $400 \times 400 \mu m^2$. The manual annotations of the nerve fibres in these two datasets were traced by an ophthalmologist using the open source software ImageJ.

OCTA dataset is an in-house data collection with 30 retinal OCTA scans. All these scans were acquired using a Heidelberg Spectralis device (Heidelberg, Germany) and all the vessels within the superficial vascular plexus (SVP) were manually traced by a clinical expert using an in-house program written in Matlab (Mathworks R2018, Natwick) as the ground truth. In this paper, we use a k-fold (k=5) cross-validation method to divide the training and testing datasets.

OCT RPE is also an in-house dataset which consists of 36 images of retina vessel shadows projected on the retinal pigment epithelium (RPE) layers of OCT volumes. These 3D volumes were captured using a Spectralis OCT system (Heidelberg Engineering GmbH) from 18 healthy volunteers, and have a size of 379×496×384. The manual annotations of these vessels were labeled by an image analysis expert using the open source software ImageJ. This dataset was originally designed for the task of eliminating retinal vessel shadows in *en face* choroidal OCT.

4.2. Experimental Setup

The proposed CS^2-Net was implemented in the PyTorch library with a dual NVIDIA GPU (GeForce GTX Titan Xp). We use adaptive moment estimation (Adam) as the overall optimizer. The initial learning rate is set to 0.0001 and we use a weight decay of 0.0005 with a poly learning rate policy, where the learning rate is multiplied by $\left(1 - \frac{iter}{max\_iter}\right)^{power}$ with a power of 0.9 and a maximum number of epochs of 100. Due to the limited amount of data, data augmentation is used to improve the performance, which includes random cropping (with a size of 384×384), contrast enhancement, random rotation (from -45° to 45°), random flipping, and mirror flipping about the image centre in the training phase. Note that we do not perform any augmentation on the test images. In this paper, we set the batch size to 8 for all the datasets and the proposed method is trained on each imaging modality separately.

To facilitate the observation and objective evaluation of the proposed method, the following metrics are adopted, accuracy (*ACC*), sensitivity (*SE*), specificity (*SP*), and Area under ROC curve (AUC):

$$ACC = \frac{TP + TN}{TP + FP + TN + FN}, \tag{9}$$

$$SE = \frac{TP}{TP + FN}, \quad SP = \frac{TN}{TN + FP}, \tag{10}$$

where TP, FN, TN, and FP denote true positive, false negative, true negative and false positive, respectively. Area under ROC curve (AUC) reflects the trade-offs between sensitivity and specificity, and thus evaluates the quality of our vessel segmentation results more reliably. In addition, we compute the *p*-values of all the evaluation metrics between the proposed method and the compared methods on each dataset for statistical analysis, and $p < 0.05$ is considered statistically significant.

4.3. Results

4.3.1. Vessel Segmentation in Color Fundus Image

To demonstrate the curvilinear structure segmentation performance of the proposed method, we first evaluate it on three public datasets (DRIVE, STARE and IOSTAR) that are common and highly recognized in medical imaging. Seven state-of-the-art segmentation methods were selected for comparison, which include two conventional filtering-based vessel segmentation methods (Combination of Shifted Filter Responses (COSFIRE) (Azzopardi et al., 2015) and Weighted Symmetry Filter (WSF) (Zhao et al., 2018b)), two specially designed deep learning-based vessel segmentation methods (DeepVessel (Fu et al., 2016) and Context Encoder Network (CE-Net) (Gu et al., 2019)), and three state-of-the-art segmentation networks (U-Net (Ronneberger et al., 2015), Recurrent Residual U-Net (R2U-Net) (Alom et al., 2018), and Dual Attention Network (DANet) (Fu et al., 2019)).

Table 2 shows the segmentation results of different methods on the retinal fundus datasets, where our proposed CS^2-Net outperforms all the competing methods on ACC and AUC scores. Thus, it can be confirmed that the spatial and channel attention modules are beneficial for retinal vessel detection in colored fundus images. Morevoer, Fig. 4 shows the visual comparison between the vessel segmentation results of R2U-Net (Alom et al., 2018), DANet (Fu et al., 2019) and the proposed CS^2-Net. We can observe that CS^2-Net achieves better performance than R2U-Net and DANet, extracting more vessels in a representative patch (green disc) with multiple scales of vessels in low contrast regions. To better observe the significance of the proposed method and comparison methods in segmenting retinal vessels, we compute the *p*-value for statistical analysis. The results shown that the differences between the proposed method and competing methods are significant with all *p*-values <0.05.

4.3.2. Vessel Segmentation in In-house OCTA Images

In order to justify that our proposed method can also segment the curvilinear structure on other modal medical images,

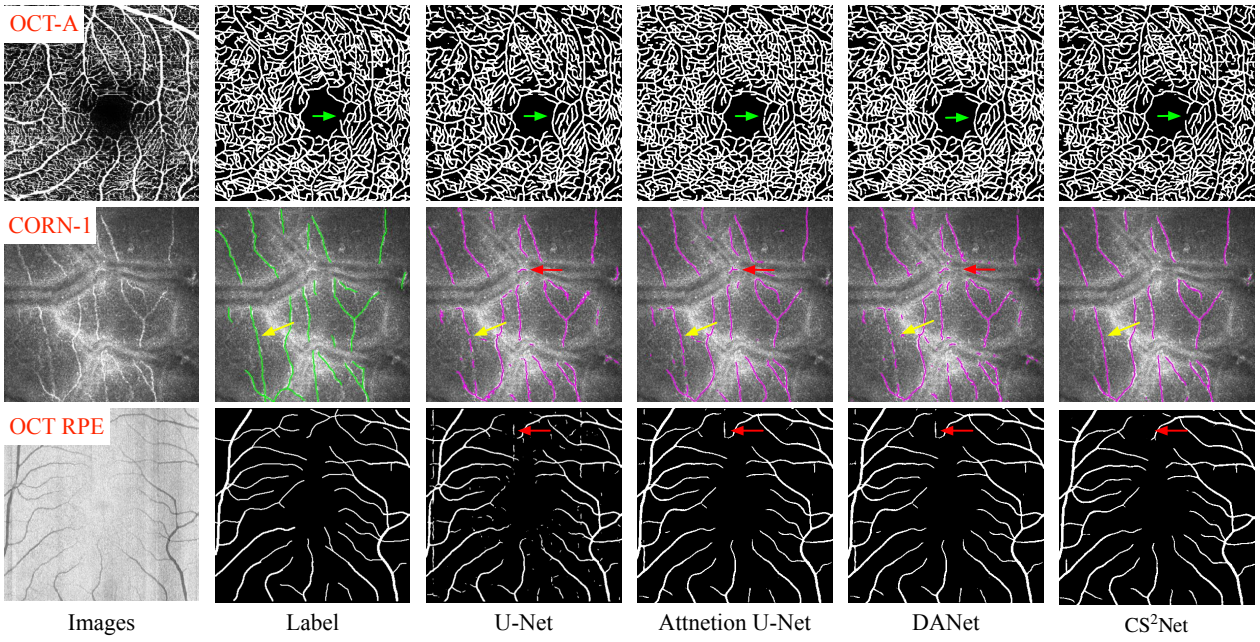

Fig. 5. Results of different methods for vessel segmentation of different images in different imaging modalities. From the left to right column: the original images, labels, and segmentation results of U-Net, Attention U-Net, DANet and the proposed CS^2-Net method, respectively. From the top to bottom row: OCTA, CORN-1 and OCT RPE Layer, respectively.

Table 2. Vessel segmentation performances in different metrics of different methods over three retinal fundus datasets.

Datasets	Methods	ACC	AUC	SE	SP
DRIVE	BCOSFIRE (Azzopardi et al., 2015)	0.9442	0.9614	0.7655	0.9704
	WSF (Zhao et al., 2018b)	0.9580	0.9750	0.7740	0.9790
	DeepVessel (Fu et al., 2016)	0.9533	0.9789	0.7603	0.9776
	U-Net (Ronneberger et al., 2015)	0.9531	0.9601	0.7537	0.9639
	R2U-Net (Alom et al., 2018)	0.9556	0.9784	0.7792	0.9813
	CE-Net (Gu et al., 2019)	0.9545	0.9779	**0.8309**	0.9747
	DANet (Fu et al., 2019)	0.9615	0.9808	0.8075	0.9841
	CS^{2}Net	**0.9632**	**0.9825**	0.8218	**0.9890**
STARE	BCOSFIRE (Azzopardi et al., 2015)	0.9497	0.9563	0.7716	0.9701
	WSF (Zhao et al., 2018b)	0.9570	0.9590	0.7880	0.9760
	DeepVessel (Fu et al., 2016)	0.9609	0.9790	0.7412	0.9701
	U-Net (Ronneberger et al., 2015)	0.9409	0.9705	0.7675	0.9631
	R2U-Net (Alom et al., 2018)	0.9712	0.9914	0.8298	**0.9862**
	CE-Net (Gu et al., 2019)	0.9583	0.9787	0.7841	0.9725
	DANet (Fu et al., 2019)	0.9679	0.9781	0.7705	0.9873
	CS^{2}Net	**0.9752**	**0.9932**	**0.8816**	0.9840
IOSTAR	BCOSFIRE (Azzopardi et al., 2015)	0.9410	0.9550	0.7610	0.9670
	WSF (Zhao et al., 2018b)	0.9480	0.9600	0.7720	0.9670
	DeepVessel (Fu et al., 2016)	-	-	-	-
	U-Net (Ronneberger et al., 2015)	0.9675	0.9464	0.8044	0.9793
	R2U-Net (Alom et al., 2018)	0.9652	0.9530	0.8042	0.9779
	CE-Net (Gu et al., 2019)	-	-	-	-
	DANet (Fu et al., 2019)	0.9720	0.9504	0.8298	**0.9832**
	CS^{2}Net	**0.9722**	**0.9758**	**0.8341**	0.9831

we perform comparative experiments on our recently released dataset: In-house OCTA. We compare the proposed network with five state-of-the-art segmentation networks: U-Net (Ronneberger et al., 2015), Deep ResUNet (Zhang et al., 2018), UNet++ (Zhou et al., 2018), Attention U-Net (Oktay et al., 2018), and DANet (Fu et al., 2019). The first column of Fig. 5 shows the visual comparison of the vessel segmentation results of different methods on a typical OCTA *en face* image. Overall, these methods perform well on segmenting major vessels. Attention U-Net (Oktay et al., 2018) is able to detect most

large structures, but it also falsely enhances background features where elongated intensity inhomogeneities are present. U-Net (Ronneberger et al., 2015) mis-detects vessels with small diameters, which leads to a relatively lower sensitivity. In contrast, the proposed CS^2-Net adaptively integrates local features with global dependencies and normalization. Hence, it shows superior performance in detecting small vessels, indicated by the green arrow in Fig. 5, and provides higher sensitivity. These findings are also confirmed by the evaluation measures reported in Table 3, where CS^2-Net achieves its highest segmentation performance in terms of all the metrics, since it employs an attention mechanism to build the powerful representation among features. In particular, the p values of the proposed method in the pairwise comparison with the the state-of-the-art methods are all less than 0.05, revealing that the proposed method achieves a significant performance improvement over them.

4.3.3. Corneal Nerve Fiber Tracing in the CORN-1 Images

We further evaluate the performance of our CS^2-Net for corneal nerve fiber tracing on the CORN-1 dataset that we have published. For validation, we compute the sensitivity and *false discovery rate* (FDR) (Guimaraes et al., 2016) between the predicted centerlines of the nerve fibres and groundtruth. FDR is defined as the fraction of the total pixels incorrectly detected as nerve segments over the total pixels of the traced nerves in the ground truth. As is customary in the evaluation process (Guimaraes et al., 2016), if any pixel on the extracted pixel-wide curves is within the three-pixel tolerance region around the manually traced nerves, it is considered as a true positive.

Similar to the evaluation on the OCTA images, we again employed U-Net (Ronneberger et al., 2015), Deep ResUNet (Zhang et al., 2018), UNet++ (Zhou et al., 2018), Attention U-Net (Oktay et al., 2018), and DANet (Fu et al., 2019) as the baselines for comparison. The second row of Fig. 5 illustrates a sample image from the CORN-1 dataset. Although all methods present visually appealing results, both U-Net, Attention U-Net and DANet falsely detect parts of the K-structures (Yokogawa et al., 2008) (indicated by the red arrows) as nerve fibers, due to the fact that they share similar morphological characteristics. In contrast, our CS^2-Net ensures continuous fiber tracing (indicated by the yellow arrows). Table 4 shows the performances of different methods for fiber tracing on the CORN-1 dataset. The challenge of corneal nerve fiber tracing is to preserve the continuity of the fibers. As a basic network, U-Net (Ronneberger et al., 2015) performs worse than the other methods. Deep ResUNet (Zhang et al., 2018) and DANet (Fu et al., 2019) obtain similar results in SE. In contrast, our method achieves the best tracing performance in terms of either the SE or FDR. In addition, the p value of the proposed method is less than 0.015, which shows that there is a significant difference in performance between the proposed method and the state-of-the-art methods.

4.3.4. Vessel Segmentation in the OCT RPE Layers

The proposed method is also validated on another different modal dataset for curviliear structure segmnetation: OCT RPE Layers. The vascular projections in the RPE layers are not true blood vessels. However, they can be considered important features to assist artifact removal on the choroid. We use the same metrics as those for the color fundus vascular segmentation to evaluate the performance of the RPE vascular projection segmentation methods. Similarity, we use U-Net (Ronneberger et al., 2015), Deep ResUNet (Zhang et al., 2018), UNet++ (Zhou et al., 2018), Attention U-Net (Oktay et al., 2018), and DANet (Fu et al., 2019) to make comparisons with the proposed CS^2-Net. Metric scores are shown in Table 5, which demonstrates the superior vascular projection performance of our model, and there are also significant differences in performance among the comparison methods, indicated by $p = 0.041$. The last row of Fig. 5 shows a randomly selected RPE image, in which the proposed method clearly demonstrates more resistance to the interference caused by the capturing device. Moreover, the proposed method extracts tiny blood vessels more effectively than U-Net (Ronneberger et al., 2015), and it does not produce over-segmentation like Attention U-Net and DANet, as indicated by the red arrows.

In Fig. 6, we show the ROC curves of our proposed CS^2-Net over different datasets for the segmentation of curvilinear structures: DRIVE, STARE, IOSTAR, CORN-1, OCT-A, and OCT RPE, compared with those of the state-of-the-art methods at particular TP and FP rates for the sake of readability. It can be seen from the local enlarged view of Fig. 6 that the proposed method outperforms on the whole state-of-the-art methods for curvilinear structure segmentation, despite the variation of structure, contrast and imaging noise from one imaging modality to another.

5. Experimental Results over 3D Volumes

5.1. Materials

To further demonstrate the broad applicability of the proposed method for the segmentation of 3D vasculatures in different modalities, we evaluate our method over 3D volumes from three public-available datasets: one brain MRA dataset (i.e., MIDAS) and two synthetic datasets (i.e., Synthetic, and VascuSynth).

MIDAS[8] is a publicly available MRA dataset. This dataset contains 50 MRA volumes which were acquired from 25 male and 25 female healthy volunteers, aged from 18 to 60+ years. Images were captured using a 3T MRI scanner under standardized protocols, with a voxel size of $0.5 \times 0.5 \times 0.8$ mm^3. These were reconstructed as a $448 \times 448 \times 128$ matrix. Manual annotations of Circles of Willis (CoW) were provided by Prof. Alejandro Frangi from the University of Leeds, where 3D vasculatures were generated by tracing the centerlines of the vessels, and the vessel surfaces were extracted using the geodesic active contour method (Bogunovic et al., 2011).

Synthetic[9] was originally generated using the method proposed in (Schneider et al., 2012), and includes 136 volumes of

[8] http://hdl.handle.net/1926/594

[9] https://github.com/giesekow/deepvesselnet/wiki/Datasets

Table 3. Vessel segmentation performances in different metrics of different methods over our private OCTA dataset.

Methods	ACC	AUC	SE	SP
U-Net (Ronneberger et al., 2015)	0.8422	0.9108	0.7867	0.8780
Deep ResUNet (Zhang et al., 2018)	0.8659	0.9175	0.8032	0.8863
UNet++ (Zhou et al., 2018)	0.8965	0.9203	0.8309	0.9101
Attention U-Net (Oktay et al., 2018)	0.9125	0.9290	0.8274	0.9007
DANet (Fu et al., 2019)	0.8869	0.9183	0.8427	0.8681
CS^{2}Net	**0.9183**	**0.9453**	**0.8631**	**0.9192**

Table 4. Nerve fibre tracing performances in different metrics of different methods over the CORN-1 dataset (mean ± standard deviation).

Methods	SE ↑	FDR ↓
U-Net (Ronneberger et al., 2015)	0.7757±0.0144	0.3961±0.0208
Deep ResUNet (Zhang et al., 2018)	0.8038±0.0140	0.2911±0.0214
UNet++ (Zhou et al., 2018)	0.8274±0.0127	0.2715±0.0118
Attention U-Net (Oktay et al., 2018)	0.8166±0.0131	0.2761±0.0120
DANet (Fu et al., 2019)	0.8012±0.0043	0.3850±0.0011
CS^{2}Net	**0.8398±0.0098**	**0.2556±0.0028**

size $325 \times 304 \times 600$ with their corresponding labels for vessel segmentation, centerlines and bifurcation detection.

VascuSynth[10] aims to provide an abundance of 3D images for the automated analysis of tree-like structures, which includes vessel segmentation and detection of bifurcation points using the VascuSynth Software (Jassi and Hamarneh, 2011). It simulates volumetric images (a size of $100 \times 100 \times 100$ voxels) of vascular trees and generates the corresponding ground truth for segmentation, bifurcation locations, branch properties, and tree hierarchy.

5.2. Experimental Setup

The proposed 3D CS^2-Net was implemented in the PyTorch framework with a dual NVIDIA GPU (Titan Xp). Adam serves as the optimizer for all comparative experiments. We adopt a poly learning strategy with an initial learning rate of 0.0001 and a weight decay of 0.0005. Due to the different sizes of 3D volumes, we have different crop sizes for different datasets, the experimental details can be found in the following subsections. In addition, we normalized the volume including training and test data and set the maximum training iteration to 200.

To better evaluate the binary segmentation performance of the proposed 3D CS^2-Net, we follow (Zhao et al., 2018a) and adopt the following metrics: true positive rate (TPR), false negative rate (FNR), and false positive rate (FPR). In order to better demonstrate that the proposed model can learn more vascular features in sparsely labeled annotations and has better discrimination ability for non-vascular patterns, we introduce two new metrics, over-segmentation rate (OR) and under-segmentation rate (UR), to evaluate the model:

$$OR = \frac{O_s}{R_s + O_s}, \quad UR = \frac{U_s}{R_s + O_s}, \tag{11}$$

where R_s denotes all the voxels inside the ground truth, O_s denotes the voxels inside the predicted volume but not inside the ground truth, and U_s indicates the voxels inside the ground truth but not in the predicted volume. According to the definition, it can be seen that $OR \in [0, 1]$ and $UR \in [0, 1]$. The lower the values of these metrics, the better the performance of the method.

5.3. Brain Vessel Segmentation in MRA Volumes

In this section, we evaluate the proposed curvilinear structure segmentation method on the cerebral MRA images. Since the manual annotations in the MIDAS dataset are sparse, i.e., many vascular voxels do not have labels. In this case, the metrics such as Dice coefficient (DC) and Intersection over Union (IoU) are not appropriate to validate its performance, since there are significantly more non-vascular voxels than the vascular ones. We perform a center crop of the raw data along the axial plane with a size of $224 \times 224 \times 64$. While the original labels are triangular polygon surfaces which cannot be directly used as input for 3D convolutions, we employ the open source medical image processing toolkit *The Visualization Toolkit* (VTK)[11] to voxelize these surfaces. These operations greatly reduce the size of the volumes, which allows us to set a larger batch size of 2 in this paper. To better justify the vessel shape and structure extraction performance of the proposed 3D CS^2-Net in the real-world scenarios, we finally evaluate it over the MRA images. Under this setting, we compare the proposed method with six state-of-the-art models: 3D Multiscale Vessel Enhancement Filtering (MVEF) (Frangi et al., 1998b), 3D Isotropic Undecimated Wavelet Filtering (IUWF) (Bankhead et al., 2012), 3D Quadrature Filters across Multiple Scales (QFMS) (Läthén et al., 2010), 3D Weighted Symmetry Filter (WSF) (Zhao et al., 2018a), V-Net (Milletari et al., 2016), 3D UNet (Çiçek et al., 2016), and Uception (Sanchesa et al., 2019). To validate the cerebrovascular segmentation performance, we compute the TPR, FNR, FPR, OR and UR between the predicted volume and ground truth. Note that we obtained the results of TPR,

[10] http://vascusynth.cs.sfu.ca/Data.html

[11] https://vtk.org/

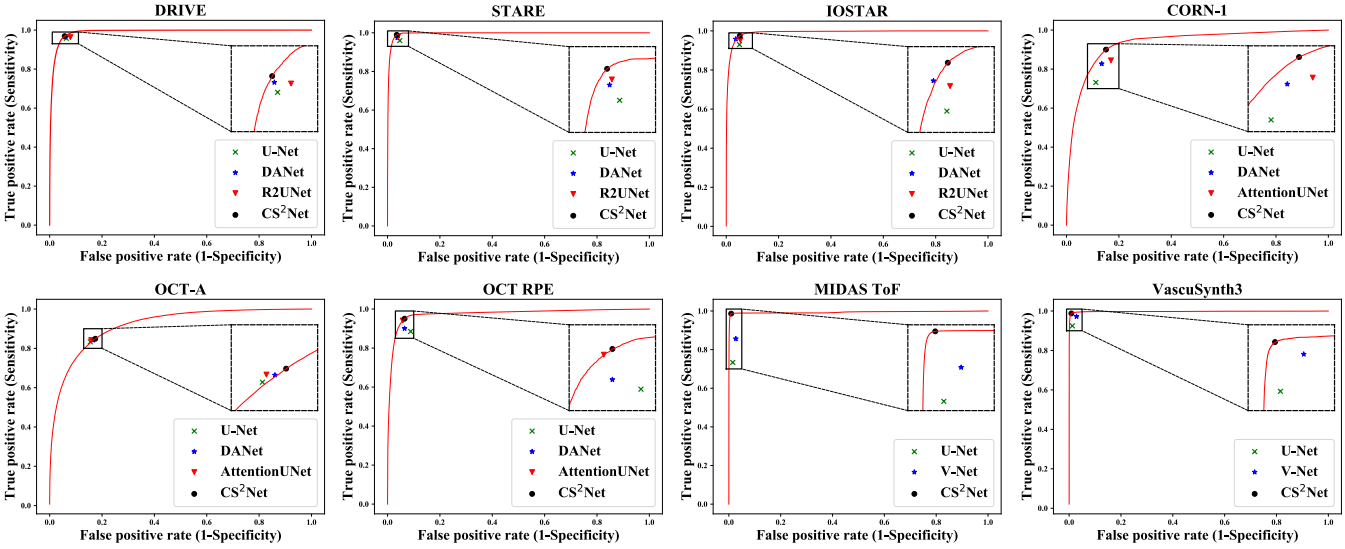

Fig. 6. ROC curves of our proposed CS^2-Net for curvilinear structure segmentation over different datasets: DRIVE, STARE, IOSTAR, CORN-1, OCT-A, OCT RPE, MIDAS and VascuSynth3 datasets, compared with those of the state-of-the-art methods at particular TP and FP rates.

Table 5. Vessel segmentation performances in different metrics of different methods over our OCT RPE Layer dataset.

Methods	ACC	AUC	SE	SP
U-Net (Ronneberger et al., 2015)	0.9550	0.9370	0.7875	0.9807
Deep ResUNet (Zhang et al., 2018)	0.9601	0.9591	0.8071	0.9838
UNet++ (Zhou et al., 2018)	0.9654	0.9578	0.8138	0.9822
Attention U-Net (Oktay et al., 2018)	0.9664	0.9584	0.8142	0.9813
DANet (Fu et al., 2019)	0.9686	0.9667	0.7849	0.9850
CS^{2}Net	**0.9693**	**0.9686**	**0.8296**	**0.9840**

FNR and FPR from (Frangi et al., 1998b; Bankhead et al., 2012) and (Zhao et al., 2018b) and put them into Table 6. We also use the OpenSource code of 3D UNet to train the model and carefully fine-tune it to reach optimality. The evaluation metrics are computed according to the predicted results. All the results are shown in Table 6. We can see that the proposed method achieves better performance in the 3D cerebrovascular segmentation task, surpassing all the other methods in terms of TPR, FNR and FPR.

In particular, the proposed method reduces the FPR of the other six methods by 0.0644, 0.0598, 0.0587, 0.0548, 0.0049 and 0.0005, respectively. This means that the proposed 3D CS^2-Net is better at distinguishing cerebrovasculatures from the complex background artifacts in MRA images. We can conclude from the UR metric in Table 6 that the predicted cerebrovasculatures of both the 3D UNet and the proposed method achieve the highest similarity with the ground truth, since the under-segmentation rates are only 0.0393 and 0.0291, respectively. However, the proposed 3D CS^2-Net achieves a lower under-segmentation rate by 0.0102. Based on the TPR metric, we can see that the proposed method is able to segment the cerebrovasculatures with the highest segmentation rate (up to 0.9706). The larger OR achieved by 3D UNet indicates that more unlabeled cerebrovascular vessels are segmented as vascular ones. Overall, the proposed method shows better reliability for the segmentation of cerebral blood vessels in terms of TPR, UR and OR, Moreover, compared with the selected methods, the proposed method gains a p value of less than 0.001, which shows that the proposed method is significantly better than other methods in segmentation performance.

Fig. 7 shows the segmentation results of 3D U-Net and the proposed 3D CS^2-Net on one of the images in the MRA dataset. Our 3D CS^2-Net presents better vascular extraction performance than 3D UNet, especially for tiny vessels, as indicated by the red arrows in Fig. 7. On the other hand, it can be observed from Fig. 7 that the cerebral vessels segmented by 3D UNet are thinner than the ground truth, while the cerebral vessels segmented by the proposed 3D CS^2-Net are more similar. Therefore, 3D UNet tends to under-segment or miss cerebral vessels, which can be verified from the OR and UR metrics in Table 6 respectively.

5.4. Vessel Segmentation in Synthetic Data

In order to further demonstrate the advantage of the proposed 3D CS^2-Net, we also report its segmentation performance over two synthetic datasets: Synthetic and VascuSynth. For both datasets, we apply k-fold ($k = 4$) cross validation to divide the training and testing datasets, i.e. 5 randomly selected volumes serve as the testing set and the remaining ones are used to train the model. In addition, Gaussian noise with three standard variances σ^2 is added to the VascuSynth dataset to mimic imaging artifacts, so as to demonstrate the superiority of the proposed

Table 6. Vessel segmentation performances in different metrics of different methods over the MRA dataset.

Methods	TPR ↑	FNR ↓	FPR ↓	OR ↓	UR ↓
MVEF (Frangi et al., 1998b)	0.9143 ± 0.0287	0.0424 ± 0.0201	0.0648 ± 0.0188	-	-
IUWF (Bankhead et al., 2012)	0.9387 ± 0.0193	0.0402 ± 0.0179	0.0602 ± 0.0178	-	-
QFMS (Läthén et al., 2010)	0.9512 ± 0.0167	0.0383 ± 0.0154	0.0591 ± 0.0121	-	-
WSF (Zhao et al., 2018a)	0.9678 ± 0.0113	0.0342 ± 0.0109	0.0562 ± 0.0128	-	-
3D UNet (Çiçek et al., 2016)	0.9521 ± 0.0001	0.0479 ± 0.0011	0.0053 ± 0.0000	0.0833 ± 0.0042	0.0393 ± 0.0001
V-Net (Milletari et al., 2016)	0.9616 ± 0.0637	0.0483 ± 0.0014	0.0009 ± 0.0000	0.1043 ± 0.0417	0.0352 ± 0.0093
Uception (Sanchesa et al., 2019)	0.9567 ± 0.0014	0.0433 ± 0.0177	0.0006 ± 0.0000	0.2005 ± 0.0012	0.0318 ± 0.0001
CS^{2}Net	**0.9683 ± 0.0014**	**0.0285 ± 0.0003**	**0.0004 ± 0.0000**	**0.0801 ± 0.0034**	**0.0291 ± 0.0003**

Table 7. Vessel segmentation results in different metrics of different methods over different 3D datasets.

Datasets	Methods	TPR ↑	FNR ↓	FPR ↓	DC ↑
Synthetic	3D UNet (Çiçek et al., 2016)	0.9965	0.0035	0.0001	0.9106
	V-Net (Milletari et al., 2016)	0.9949	0.0051	0.0001	0.9237
	Uception (Sanchesa et al., 2019)	0.9984	0.0026	0.0003	0.9785
	CS^{2}Net	**0.9986**	**0.0014**	**0.0000**	**0.9913**
VascuSynth-1	ITM (Cetin et al., 2012)	0.9423	0.0577	0.0471	0.9406
	CBS (Cheng et al., 2015)	0.9529	0.0471	0.0563	0.9489
	WSF (Zhao et al., 2018a)	0.9678	0.0342	0.0562	0.9601
	3D UNet (Çiçek et al., 2016)	0.9704	0.0096	0.0007	0.9552
	V-Net (Milletari et al., 2016)	0.9763	0.0088	0.0003	0.9594
	Uception (Sanchesa et al., 2019)	0.9800	0.0071	0.0004	0.9426
	CS^{2}Net	**0.9841**	**0.0068**	**0.0001**	**0.9637**
VascuSynth-2	ITM (Cetin et al., 2012)	0.9423	0.0577	0.0471	0.9406
	CBS (Cheng et al., 2015)	0.9529	0.0471	0.0563	0.9489
	WSF (Zhao et al., 2018a)	0.9603	**0.0451**	0.0526	0.9543
	3D UNet (Çiçek et al., 2016)	0.9602	0.0502	0.0013	0.9587
	V-Net (Milletari et al., 2016)	0.9605	0.0503	0.0011	0.9584
	Uception (Sanchesa et al., 2019)	0.9607	0.0510	0.0009	0.9468
	CS^{2}Net	**0.9611**	0.0494	**0.0004**	**0.9593**
VascuSynth-3	3D UNet (Çiçek et al., 2016)	0.9338	0.0661	0.0024	0.9112
	V-Net (Milletari et al., 2016)	0.9365	0.0598	0.0027	0.9037
	Uception (Sanchesa et al., 2019)	0.9413	0.1033	0.0066	0.9157
	CS^{2}Net	**0.9484**	**0.0416**	**0.0005**	**0.9256**

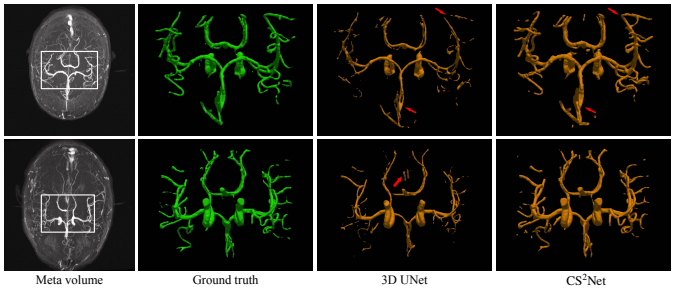

Fig. 7. 3D renderings of curvilinear structure segmentation results of an image in the MRA dataset. From the left to right column: a MIP view of a sample MRA image, the segmentation of ground truth, the 3D U-Net and the proposed CS^2-Net respectively.

method in detecting curvilinear structures in noise-corrupted data. In the remainder of this paper, we refer to these noise-corrupted versions as: VascuSynth-1 ($\sigma^2 = 20$), VascuSynth-2 ($\sigma^2 = 60$), and VascuSynth-3 ($\sigma^2 = 100$). An additional random crop operation with a size of $128 \times 128 \times 128$ is adopted to reduce the training cubes and we set the batch size to 6 in this part.

We first evaluate our proposed method and compare it with the state-of-the-art ones: Intensity-based Tensor Model (ITM) (Cetin et al., 2012), Constrained B-Snake (CBS) (Cheng et al., 2015), Weighted Symmetry Filtering (WSF) (Zhao et al., 2018a), 3D UNet (Çiçek et al., 2016), V-Net (Milletari et al., 2016) and Uception (Sanchesa et al., 2019) on the Synthetic dataset. The results are shown in Table 7. We observe that the proposed method successfully segments 3D curvilinear structures with competitive performance and outperforms the 3D UNet in terms of TPR, FNR, FPR, and particularly DC (up by 0.0807). Fig. 8 further demonstrates the segmentation performance of both methods. Compared with the 3D UNet (Çiçek et al., 2016), the proposed method shows better discrimination ability at boundaries, which can be verified from the enlarged view (green box) of the first row in Fig. 8. The synthetic vessels segmented by the 3D UNet are thicker than the results of the proposed method. This proves that the proposed method is able to achieve better edge discrimination ability through the pro-

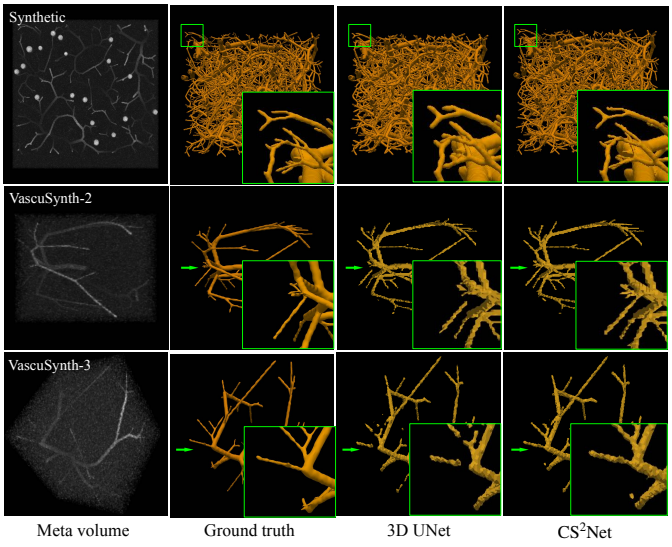

Fig. 8. 3D renderings of curvilinear structure segmentation results of different methods over Synthetic and VascuSynth. The first column shows volumes with the different levels of noise ($\sigma^2 = 20$ for Synthetic, $\sigma^2 = 60$ for VascuSynth-2 and $\sigma^2 = 100$ for VascuSynth-3). Segmentation results of different methods in the second to right column: ground truth, 3D UNet and the proposed CS^2-Net, respectively. The green box in first row shows an enlarged view of the local segmentation results.

posed CSAM. It can also be analyzed from $p = 0.027 < 0.05$ that the proposed method performs significantly better than both the 3D U-Net and the V-Net.

We further make comparisons between some of state-of-the-art methods and the proposed method. All the performances are evaluated based on TPR, FNR, FPR and DC. We follow (Zhao et al., 2018a) and add Gaussian noise with a standard variance of $\sigma^2 = 20$ to generate the VascuSynth-1 dataset. To fully verify the proposed model, we also follow (Zhao et al., 2018a) to compare the proposed method with other state-of-the-art models (Intensity based Tensor Model (ITM) (Cetin et al., 2012), Constrained B-Snake (CBS) (Cheng et al., 2015), WSF (Zhao et al., 2018a)), as well as deep learning-based model (3D UNet (Çiçek et al., 2016), V-Net (Milletari et al., 2016)), and Uception (Sanchesa et al., 2019) on different standard variances of $\sigma^2 = 20$ and $\sigma^2 = 60$ respectively. These experimental results are shown in Table 7. As can be seen, the proposed CS^2-Net outperforms other state-of-the-art methods in the segmentation of 3D curvilinear structures. Fig. 8 illustrates the 3D segmentation results of two sample 3D images by the proposed method and 3D UNet. As indicated by the green arrow and representative patches, we can observe that the 3D UNet detects discontinuous vessels and misses the small ones in the middle left of the figure. In sharp contrast, the proposed CS^2-Net detects all the vessels more completely, even though they vary in thickness, length, and local contrast with the background. For VascuSynth-1 and VascuSynth-2, the proposed method has better performance in segmenting the 3D curvilinear structure than other methods, confirmed by $p = 0.011$ and $p = 0.015$ respectively.

Since 3D UNet, V-Net, Uception and the proposed network are deep learning based methods, we apply a higher-level noise of $\sigma^2 = 100$ on the volume data to further confirm their performance for 3D curvilinear structure segmentation. Quantitative results are shown in Table 7. The results of different methods for the curvilinear structure segmentation of a randomly selected image in the VascuSynth-3 dataset are presented in the last row of Fig. 8. From the table and the detailed view in the green box of the figure, we can conclude that the proposed method has still detected more complete vessels compared with the other methods. This is because the attention model in the proposed method evaluates the expression capability of the features globally over the whole images and normalises them in the feature space and are thus more robust to the local noise and variation in size of the vessels. Compared with 3D UNet, V-Net, and Uception as the noise level increases, the performance of the proposed method increases significantly, which can be concluded from the change of p-value from $0.01 < p < 0.05$ ($\sigma^2 = 20$ and $\sigma^2 = 60$) to $p < 0.001$ ($\sigma^2 = 100$).

6. Discussions

The proposed CS^2-Net utilizes a spatial and channel attention modules to capture the structure information of the tree-like structures in horizontal and vertical directions, respectively. In this work, we carefully designed a network focusing on the extraction of the curvilinear structure in medical images. Compared with natural images, medical images contain more unique features, such as simpler semantics and unitary patterns. Therefore, we first construct a network backbone based on the encoder-decoder framework. More specifically, we introduce a 1×3 and a 3×1 convolutional kernel to capture more boundary feature to assist the segmentation of curvilinear structures. DANet (Fu et al., 2019) uses a pre-trained model to extract features, and up-samples the attention features in the last layer of the model, and this is the architectural difference between the proposed method and DANet. Secondly, we introduce batch normalization and ReLU activations after the convolutional layers in the spatial attention module to ensure that the mean and variance of the input distribution are fixed within a certain range, reducing the internal covariate shift in the network, and mitigating the gradient disappearance to a certain extent. Thirdly, since 3D volume contains rich depth information that is not included in 2D medical images, and many important lesions can be better observed through different layers of the 3D volume. In this case, we extend the 2D attention to the 3D field to enhance the network's ability to aggregate depth information of different layers. Therefore, we design a 3D volume segmentation network, and introduce $1 \times 1 \times 3$, $1 \times 3 \times 1$ and $3 \times 1 \times 1$ convolution kernels in the 3D attention module with batch normalization and ReLU activations, which makes the network more suitable for 3D medical data analysis. In addition, we evaluate the proposed method on a variety of medical datasets in different modalities, and the evaluation results also confirmed that our proposed method are effective for segmenting curvilinear structures.

To better support improved segmentation results, we further visualize some intermediate attention maps of our proposed CS^2-Net over different datasets, as shown in Fig. 9(a). By analyzing and comparing the blood vessels and nerve fibers in the

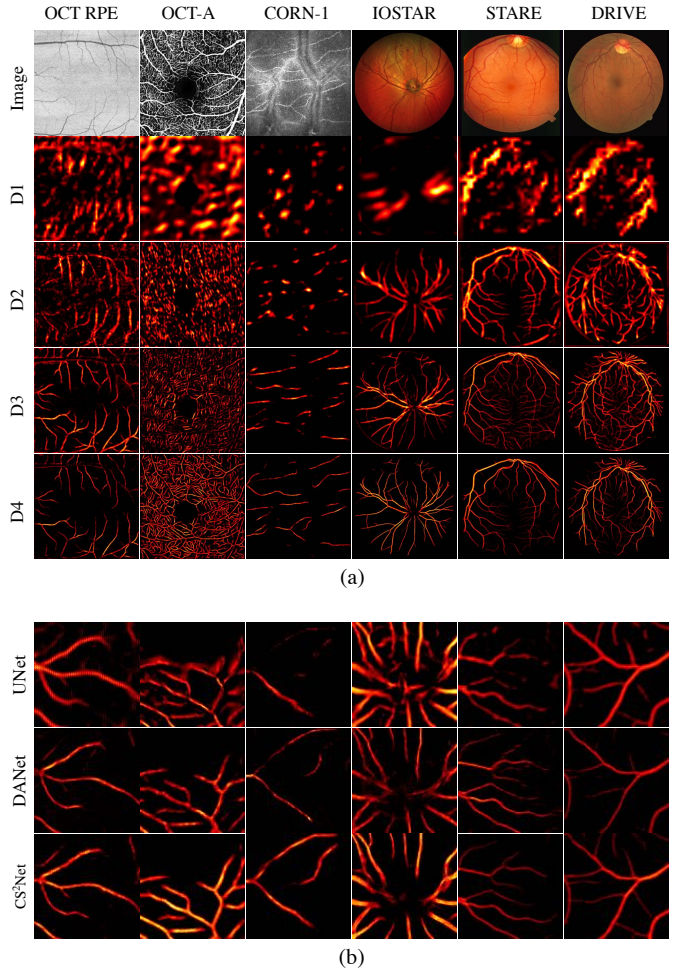

Fig. 9. Attention maps of different methods in the intermediate layers of decoding parts. (a) the attention maps of the proposed CS^2-Net in different decoding layers on different datasdets: DRIVE, STARE, IOSTAR, CORN-1, OCT-A, and OCT RPE datasets, respectively. D1 ~ D4 display the attention maps representing the incremental refinement in curvilinear structure segmentation; (b) the enlarged local intermediate attention maps of different methods: UNet, DANet, and CS^2-Net.

attention maps from $D1$ to $D4$, it can be found that the proposed model can pay more attention on the curvilinear structure during training, that is, the curvilinear structures gradually become clearer and smoother. In the low level attention maps, the highlighted areas basically are distributed at the curvilinear structure region, which reflects that the CAB module focuses more classified information on the curvilinear structure. On the other hand, the highlighted areas at different spatial locations also confirm that the SAB module can enhance the ability of the proposed network to capture long-range dependencies of curvilinear structure. In addition, we present several sets of attention maps of the proposed method and its two backbone methods (DANet and U-Net) in the same intermediate layer in Fig. 9(b) to more intuitively verify the performance of the proposed attention modules. Overall, it can be observed from the comparison of each column that the proposed CS^2-Net has stronger response than both DANet and U-Net in terms of curvilinear structure information aggregation. In this case, the proposed CS^2-Net is more responsive to vessels and nerve fibers than

Table 8. TPR, FNR, FPR, Over-segmentation Rate (OR) and Under-segmentation Rate (UR) of the proposed method with a combination of different components for the curvilinear structure segmentation of the 3D images in the MIDAS ToF MRA dataset.

Methods	TPR ↑	FNR ↓	FPR ↓	OR ↓	UR ↓
Backbone	0.9517	0.0493	0.0103	0.0877	0.0808
Backbone+CAB	0.9663	0.0310	0.0024	0.1082	0.0341
Backbone+SAB	0.9565	0.0413	0.0018	0.1147	0.0532
CS^{2}Net	**0.9706**	**0.0285**	**0.0004**	**0.1027**	**0.0296**

DANet, which can be clearly seen from the brighter highlights of curvilinear structures in Fig. 9(b). Moreover, by comparing the attention maps of CS^2-Net and U-Net, it may be seen clearly that the proposed CS^2-Net is more powerful in suppressing the background interference than U-Net.

Moreover, in order to demonstrate the effectiveness of the 3D CSAM in CS^2-Net, we carry out an ablation study over the MIDAS dataset. Firstly, we test the backbone of our network, e.g., Deep ResUNet (Zhang et al., 2018), without the CSAM. For fair comparison, we retrain the backbone network under the same hyperparameter settings as the proposed method and use the same metrics to perform its evaluation. Secondly, we perform a further ablation study by removing the CAB but retaining the SAB inside the original CS^2-Net. For the final ablation study, we remove the SAB in the CS^2-Net but retain the CAB to form the final set of the ablation study. All computed metrics are shown in Table 8. The results reveal that the proposed CSAM can effectively extract the features of curvilinear structures. For the performance of the backbone (Backbone) is slightly improved by that with the CAB only (Backbone+CAB) from 0.9517 to 0.9663 in TPR and from 0.0493 to 0.0310 in FNR. However, much better performance is achieved with the SAB only (Backbone+SAB). This is because the CAB normalises the features in the feature space for simple binary classification (vascular and non-vascular) tasks, while the SAB enhances the features over the whole image and thus increases the contrasts of different objects at different locations. Thus, the CSAM integrates the advantages of both the CAB and SAB to make the model better at producing inter-class discrimination and intra-class responses, and thus obtains the best performance with a TPR of 0.9706 and an FNR of 0.0285.

In addition to numerical verification, we also obtain and visualize the outputs before and after CSAM for different image modalities through a heat map. Note that we applied upsampling and sigmoid operations on the outputs to resize the feature maps to the corresponding image size and normalize the outputs to [0, 1], respectively. The visualization results are shown in Fig. 10. As indicated by the green arrow, the voxels with the probabilities as being curvilinear structure by our proposed CSAM are better clustered and indicative than those without it. These results clearly show that the CSAM is effective in extracting and aggregating the edge features of the curvilinear structures and enhancing the network's ability to distinguish between vascular and non-vascular patterns.

Fig. 10. The output of the proposed CSAM on a randomly selected image from the MIDAS dataset. From the left to right: the original volume, the predicted probabilities of voxles as curvilinear structure before and after the application of the proposed CSAM, respectively.

7. Conclusion and Future Works

Curvilinear structure segmentation is a fundamental step in automated diagnosis of many diseases, and it remains as a challenging medical image analysis problem despite considerable research efforts. In this paper, we have developed a new curvilinear structure segmentation network, named CS^2-Net, which is applicable to both 2D images and 3D volumes. Our CS^2-Net improves the inter-class discrimination and intra-class aggregation abilities, by applying a self-attention mechanism to high-level features in the channel and spatial dimensions (Fu et al., 2019). The experimental results over 9 datasets in 6 imaging modalities have demonstrated that the proposed method can improve the segmentation of curvilinear structures. Its superior performance confirms its great potential as a powerful tool for a wide range of healthcare applications and beyond.

Over the past few years there has been an increasing number of AI models that have been proposed and published. The lack of evaluation of their usefulness across different images in different imaging modalities appears to make them slow to be adopted in real applications. It is our intention to provide extensive evaluation not only to demonstrate the effectiveness of our model in the segmentation but also their applicability over many different images. We are expecting that this strategy will lead to fast impact on real applications.

Although these studies have revealed important findings about the potential and applicability of our proposed CS^2-Net method for curvilinear structure segmentation, the following limitations of our method will be addressed in the future. Due to the presence of pathological cells similar to the curvilinear structure in some regions of the image, the model will thus oversegment the curvilinear structure. It may be possible to improve this segmentation by applying additional information to the model such as local neighbourhood and continuity. Another limitation is that 3D volume segmentation consumes a lot of GPU resources, which makes it difficult to train the network. It is possible to simplify the architecture and thus reduce the complexity of the proposed method.

Acknowledgments

This work was supported by the Zhejiang Provincial Natural Science Foundation grant (LZ19F010001) and Ningbo Natural Science Foundation grant (2018A610055).

References

Al-Diri, B., Hunter, A., Steel, D., 2009. 3D an active contour model for segmenting and measuring retinal vessels. IEEE Transactions on Medical Imaging 28, 1488–1497.

Alom, M., et al., 2018. Recurrent residual convolutional neural network based on u-net (r2u-net) for medical image segmentation. arXiv:1802.06955 .

Annunziata, R., et al., 2016. A fully automated tortuosity quantification system with application to corneal nerve fibres in confocal microscopy images. Med. Image Anal. 32, 216–232.

Azzopardi, G., Strisciuglio, N., Vento, M., Petkov, N., 2015. Trainable cosfire filters for vessel delineation with application to retinal images. Medical Image Analysis 19, 46–57.

Badrinarayanan, V., Kendall, A., Cipolla, R., 2017. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence 39, 2481–2495.

Bankhead, P., Scholfield, C.N., McGeown, J.G., Curtis, T.M., 2012. Fast retinal vessel detection and measurement using wavelets and edge location refinement. PloS one 7, 332435.

Bibiloni, P., González-Hidalgo, M., Massanet, S., 2016. A survey on curvilinear object segmentation in multiple applications. Pattern Recognition 60, 949–970.

Bogunovic, H., Pozo, J., Frangi, A., 2011. Automated segmentation of cerebral vasculature with aneurysms in 3DRA and TOF-MRA using geodesic active regions: An evaluation study. Med. Phys. 38, 210–222.

de Carlo, T., Romano, A., Waheed, N., Duker, J., 2015. A review of optical coherence tomography angiography (octa). International Journal of Retina and Vitreous 1, 5.

Cetin, S., Demir, A., Yezzi, A., Degertekin, M., Unal, G., 2012. Vessel tractography using an intensity based tensor model with branch detection. IEEE Transactions on Medical Imaging 32, 348–363.

Cetin, S., Unal, G., 2015. A higher-order tensor vessel tractography for segmentation of vascular structures. IEEE Transactions on Medical Imaging 34, 2172–2185.

Chen, D., Zhang, J., Cohen, L.D., 2019. 3D minimal paths for tubular structure segmentation with coherence penalty and adaptive anisotropy. IEEE transactions on Image Processing 28, 1271–1284.

Cheng, Y., Hu, X., Wang, J., Wang, Y., Tamura, S., 2015. Accurate vessel segmentation with constrained b-snake. IEEE Transactions on Image Processing 24, 2440–2455.

Chung, A., Noble, J., 1999. Statistical 3d vessel segmentation using a rician distribution, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 82–89.

Çiçek, Ö., Abdulkadir, A., Lienkamp, 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.

Colonna, A., Scarpa, F., Ruggeri, A., 2018. Segmentation of corneal nerves using a u-net-based convolutional neural network, in: Computational Pathology and Ophthalmic Medical Image Analysis, Springer.

Cuadrado-Godia, E., Dwivedi, P., Sharma, S., Santiago, A., Gonzalez, J., Balcells, M., Laird, J., Turk, M., Suri, H., Nicolaides, A., 2018. Cerebral small vessel disease: a review focusing on pathophysiology, biomarkers, and machine learning strategies. Journal of stroke 20, 302.

Dai, J., Li, Y., He, K., Sun, J., 2016. R-fcn: Object detection via region-based fully convolutional networks, in: Advances in neural information processing systems, pp. 379–387.

Díaz, M., Novo, J., Cutrín, P., Gómez-Ulla, F., Penedo, M., Ortega, M., 2019. Automatic segmentation of the foveal avascular zone in ophthalmological oct-a images. PloS ONE 14, e0212363.

Ding, C., Xia, Y., Li, Y., 2014. Supervised segmentation of vasculature in retinal images using neural networks, in: 2014 International Conference on Orange Technologies, IEEE. pp. 49–52.

Eladawi, N., et al., 2017. Automatic blood vessels segmentation based on different retinal maps from octa scans. Comp. Biolog. Med. 89, 150–161.

Frangi, A., Niessen, W., Vincken, K., Viergever, M., 1998a. Multiscale vessel enhancement filtering, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 130–137.

Frangi, A.F., Niessen, W.J., Vincken, K.L., Viergever, M.A., 1998b. Multiscale vessel enhancement filtering, in: International conference on medical image computing and computer-assisted intervention, Springer. pp. 130–137.

Franklin, S.W., Rajan, S.E., 2014. Computerized screening of diabetic retinopa-

thy employing blood vessel segmentation in retinal images. Biocybernetics and Biomedical Engineering 34, 117–124.

Fraz, M., Remagnino, P., Hoppe, A., Uyyanonvara, B., Rudnicka, A., Owen, C., Barman, S., 2012. Blood vessel segmentation methodologies in retinal images - a survey. Computer Methods and Programs in Biomedicine 108, 407–433.

Fu, H., Xu, Y., Lin, S., Wong, D., Liu, J., 2016. Deepvessel: Retinal vessel segmentation via deep learning and conditional random field, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 132–139.

Fu, J., Liu, J., Tian, H., Li, Y., Bao, Y., Fang, Z., Lu, H., 2019. Dual attention network for scene segmentation, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3146–3154.

Gibson, E., Giganti, F., Hu, Y., Bonmati, E., Bandula, S., Gurusamy, K., Davidson, B., Pereira, S., Clarkson, M., Barratt, D., 2018. Automatic multi-organ segmentation on abdominal ct with dense v-networks. IEEE Transactions on Medical Imaging 37, 1822–1834.

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 transactions on medical imaging .

Guimaraes, P., Wigdahl, J., Ruggeri, A., 2016. A fast and efficient technique for the automatic tracing of corneal nerves in confocal microscopy. Translational Vision Science & Technology 5.

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.

Heisler, M., Chan, F., Mammo, Z., Balaratnasingam, C., Prentasic, P., Docherty, G., Ju, M., Rajapakse, S., Lee, S., Merkur, A., 2019. Deep learning vessel segmentation and quantification of the foveal avascular zone using commercial and prototype oct-a platforms. arXiv preprint arXiv:1909.11289 .

Hosseinaee, Z., Tan, B., Kralj, O., Han, L., Wong, A., Sorbara, L., Bizheva, K., 2019. Fully automated corneal nerve segmentation algorithm for corneal nerves analysis from in-vivo uhr-oct images, in: Ophthalmic Technologies XXIX, International Society for Optics and Photonics. p. 1085823.

Ioffe, S., Szegedy, C., 2015. Batch normalization: Accelerating deep network training by reducing internal covariate shift. arXiv preprint arXiv:1502.03167 .

Jassi, P., Hamarneh, G., 2011. Vascusynth: Vascular tree synthesis software. Insight Journal , 1–12.

Jin, Q., Meng, Z., Pham, T., Chen, Q., Wei, L., Su, R., 2019. Dunet: A deformable network for retinal vessel segmentation. Knowledge-Based Systems 178, 149–162.

Kim, J., Markoulli, M., 2018a. Automatic analysis of corneal nerves imaged using in vivo confocal microscopy. Clinical and Experimental Optometry 101 2, 147–161.

Kim, J., Markoulli, M., 2018b. Automatic analysis of corneal nerves imaged using in vivo confocal microscopy. Clinical and Experimental Optometry 101, 147–161.

Läthén, G., Jonasson, J., Borga, M., 2010. Blood vessel segmentation using multi-scale quadrature filtering. Pattern Recognition Letters 31, 762–767.

Lesage, D., Angelini, E., Bloch, I., Funka-Lea, G., 2009. Bayesian maximal paths for coronary artery segmentation from 3D CT angiograms, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 222–229.

Li, Y., Gong, H., Wu, W., Liu, G., Chen, G., 2015. An automated method using hessian matrix and random walks for retinal blood vessel segmentation, in: International Congress on Image and Signal Processing (CISP), IEEE. pp. 423–427.

Liao, W., Rohr, K., Kang, C., Cho, Z., Wörz, S., 2012. Automatic human brain vessel segmentation from 3d 7 tesla mra images using fast marching with anisotropic directional prior, in: 2012 9th IEEE International Symposium on Biomedical Imaging (ISBI), pp. 1140–1143.

Liskowski, P., Krawiec, K., 2016. Segmenting retinal blood vessels with deep neural networks. IEEE Transactions on Medical Imaging 35, 2369–2380.

Maninis, K., Pont-Tuset, J., Arbeláez, P., Gool, L.V., 2016. Deep retinal image understanding, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer. pp. 140–148.

Milletari, F., Navab, N., Ahmadi, S., 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.

Mo, J., Zhang, L., 2017. Multi-level deep supervised networks for retinal vessel segmentation. International journal of computer assisted radiology and surgery 12, 2181–2193.

Mou, L., et al., 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.

Oakley, J., Russakoff, D., Weinberg, R., McCarron, M., Izzi, J., Mankowski, J., 2019. Deep learning-based analysis of macaque corneal sub-basal nerve fibers in confocal microscopy images. bioRxiv , 758433.

Oktay, O., et al., 2018. Attention u-net: learning where to look for the pancreas. arXiv:1804.03999 .

Peng, C., Zhang, X., Yu, G., Luo, G., Sun, J., 2017. Large kernel matters–improve semantic segmentation by global convolutional network, in: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 4353–4361.

Poulain, E., Malandain, G., Vaillant, R., 2019. 3d coronary vessel tree tracking in x-ray projections, in: International Conference on Functional Imaging and Modeling of the Heart, Springer. pp. 388–396.

Ren, S., He, K., Girshick, R., Sun, J., 2015. Faster r-cnn: Towards real-time object detection with region proposal networks, in: Advances in neural information processing systems, pp. 91–99.

Rieber, J., Huber, A., Erhard, I., Mueller, S., Schweyer, M., Koenig, A., Schiele, T., Theisen, K., Siebert, U., Schoenberg, S., 2006. Cardiac magnetic resonance perfusion imaging for the functional assessment of coronary artery disease: a comparison with coronary angiography and fractional flow reserve. European Heart Journal 27, 1465–1471.

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, pp. 234–241.

Sanchesa, 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.

Schneider, M., Reichold, J., Weber, B., Székely, G., Hirsch, S., 2012. Tissue metabolism driven arterial tree generation. Medical image analysis 16, 1397–1414.

Shang, Y., Deklerck, R., Nyssen, E., Markova, A., de Mey, J., Yang, X., Sun, K., 2019. 3D vascular active contour for vessel tree segmentation. IEEE Transactions on Biomedical Engineering 58, 1023–1032.

Shin, S., Lee, S., Yun, I., Lee, K., 2019. Deep vessel segmentation by learning graphical connectivity. Medical Image Analysis 58, 101556.

Soares, J.V.B., Leandro, J.J.G., Cesar, R.M., Jelinek, H.F., Cree, M.J., 2006. Retinal vessel segmentation using the 2-d gabor wavelet and supervised classification. IEEE Transactions on Medical Imaging 25, 1214–1222.

Staal, J., Abràmoff, M., Niemeijer, M., Viergever, M., Van Ginneken, B., 2004. Ridge-based vessel segmentation in color images of the retina. IEEE Transactions on Medical Imaging 23, 501–509.

Szegedy, C., Ioffe, S., Vanhoucke, V., Alemi, A.A., 2017. Inception-v4, inception-resnet and the impact of residual connections on learning, in: Thirty-First AAAI Conference on Artificial Intelligence.

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A., 2015. Going deeper with convolutions, in: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 1–9.

Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., Wojna, Z., 2016. Rethinking the inception architecture for computer vision, in: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 2818–2826.

Tetteh, G., Efremov, V., Forkert, N., Schneider, M., Kirschke, J., Weber, B., Zimmer, C., Piraud, M., Menze, B., 2018. Deepvesselnet: Vessel segmentation, centerline prediction, and bifurcation detection in 3-d angiographic volumes. arXiv preprint arXiv:1803.09340 .

Wang, F., Gu, Y., Liu, W., Yu, Y., He, S., Pan, J., 2019a. Context-aware spatio-recurrent curvilinear structure segmentation, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 12648–12657.

Wang, H., Zhang, D., Song, Y., Liu, S., Wang, Y., Feng, D., Peng, H., Cai, W., 2019b. Segmenting neuronal structure in 3d optical microscope images via knowledge distillation with teacher-student network, in: 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019), IEEE. pp. 228–231.

Wang, J., Chung, A., 2019. High-order oriented cylindrical flux for curvilinear structure detection and vessel segmentation, in: International Conference on Information Processing in Medical Imaging, Springer. pp. 479–491.

Wang, X., Jiang, X., Ren, J., 2019c. Blood vessel segmentation from fundus image by a cascade classification framework. Pattern Recognition 88, 331–341.

Williams, B.M., Borroni, D., Liu, R., Zhao, Y., Zhang, J., Lim, J., Ma, B., Romano, V., Qi, H., Ferdousi, M., et al., 2020. An artificial intelligence-based deep learning algorithm for the diagnosis of diabetic neuropathy using corneal confocal microscopy: a development and validation study. Diabetologia 63, 419–430.

Wilson, D., 1998. An improved planning protocol for the endovascular treatment of intracranial aneurysms. Ph.D. thesis. University of Oxford.

Yokogawa, H., Kobayashi, A., Sugiyama, K., 2008. Mapping of normal corneal k-structures by in vivo laser confocal microscopy. Cornea 27, 879–883.

Zhang, J., Chen, Y., Bekkers, E., Wang, M., Dashtbozorg, B., ter Haar Romeny, B., 2017. Retinal vessel delineation using a brain-inspired wavelet transform and random forest. Pattern Recognition 69, 107–123.

Zhang, J., Dashtbozorg, B., Bekkers, E., Pluim, J., Duits, R., ter Haar Romeny, B., 2016. Robust retinal vessel segmentation via locally adaptive derivative frames in orientation scores. IEEE Transactions on Medical Imaging 35, 2631–2644.

Zhang, J., Qiao, Y., Sarabi, M.S., Khansari, M.M., Gahm, J.K., Kashani, A.H., Shi, Y., 2019. 3D shape modeling and analysis of retinal microvasculature in oct-angiography images. IEEE transactions on medical imaging .

Zhang, Z., Liu, Q., Wang, Y., 2018. Road extraction by deep residual u-net. IEEE Geoscience and Remote Sensing Letters 15, 749–753.

Zhao, H., Shi, J., Qi, X., Wang, X., Jia, J., 2017a. Pyramid scene parsing network, in: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 2281–2890.

Zhao, J., Ai, D., Yang, Y., Song, H., Huang, Y., Wang, Y., Yang, J., 2019. Deep feature regression (dfr) for 3d vessel segmentation. Physics in Medicine & Biology 64, 115006.

Zhao, Y., Rada, L., Chen, K., Harding, S.P., Zheng, Y., 2015. Automated vessel segmentation using infinite perimeter active contour model with hybrid region information with application to retinal images., in: 34 (Ed.), IEEE Transactions on Medical Imaging, pp. 1797–1807.

Zhao, Y., Zhao, J., Yang, J., Liu, Y., Zhao, Y., Zheng, Y., Xia, L., Wang, Y., 2017b. Saliency driven vasculature segmentation with infinite perimeter active contour model. Neurocomputing 259, 201–209.

Zhao, Y., Zheng, Y., Liu, Y., Zhao, Y., Luo, L., Yang, S., Na, T., Wang, Y., Liu, J., 2018a. Automatic 2-d/3-d vessel enhancement in multiple modality images using a weighted symmetry filter. IEEE Transactions on Medical Imaging 37, 438–450.

Zhao, Y., Zheng, Y., Liu, Y., Zhao, Y., Luo, L., Yang, S., Na, T., Wang, Y., Liu, J., 2018b. Automatic 2D/3D vessel enhancement in multiple modality images using a weighted symmetry filter. IEEE Transactions on Medical Imaging 37, 438–450.

Zhou, Z., Siddiquee, M., Tajbakhsh, N., Liang, J., 2018. Unet++: A nested u-net architecture for medical image segmentation, in: Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support, Springer. pp. 3–11.