

VC-Net: Deep Volume-Composition Networks for Segmentation and Visualization of Highly Sparse and Noisy Image Data

Yifan Wang, Guoli Yan, Haikuan Zhu, Sagar Buch, Ying Wang, Ewart Mark Haacke, Jing Hua, and Zichun Zhong

Abstract—The fundamental motivation of the proposed work is to present a new visualization-guided computing paradigm to combine direct 3D volume processing and volume rendered clues for effective 3D exploration. For example, extracting and visualizing microstructures *in-vivo* have been a long-standing challenging problem. However, due to the high sparseness and noisiness in cerebrovasculature data as well as highly complex geometry and topology variations of micro vessels, it is still extremely challenging to extract the complete 3D vessel structure and visualize it in 3D with high fidelity. In this paper, we present an end-to-end deep learning method, *VC-Net*, for robust extraction of 3D microvascular structure through embedding the image composition, generated by maximum intensity projection (MIP), into the 3D volumetric image learning process to enhance the overall performance. The core novelty is to automatically leverage the volume visualization technique (e.g., MIP – a volume rendering scheme for 3D volume images) to enhance the 3D data exploration at the deep learning level. The MIP embedding features can enhance the local vessel signal (through canceling out the noise) and adapt to the geometric variability and scalability of vessels, which is of great importance in microvascular tracking. A multi-stream convolutional neural network (CNN) framework is proposed to effectively learn the 3D volume and 2D MIP feature vectors, respectively, and then explore their inter-dependencies in a joint volume-composition embedding space by unprojecting the 2D feature vectors into the 3D volume embedding space. It is noted that the proposed framework can better capture the small / micro vessels and improve the vessel connectivity. To our knowledge, this is the first time that a deep learning framework is proposed to construct a joint convolutional embedding space, where the computed vessel probabilities from volume rendering based 2D projection and 3D volume can be explored and integrated synergistically. Experimental results are evaluated and compared with the traditional 3D vessel segmentation methods and the state-of-the-art in deep learning, by using extensive public and real patient (micro-)cerebrovascular image datasets. The application of this accurate segmentation and visualization of sparse and complicated 3D microvascular structure facilitated by our method demonstrates the potential in a powerful MR arteriogram and venogram diagnosis of vascular disease.

Index Terms—Deep neural network, 3D cerebrovascular segmentation and visualization, maximum intensity projection (MIP), joint embedding

1 INTRODUCTION

Nowadays, there is a pressing need for better visualizing and understanding microstructures in the raw and wild datasets. For instance, the acquisition of the *in-vivo* micro-level 3D vasculature from image data is a grand challenge. The notorious difficulties of microvascular data analytics lie in *high sparseness* of vessel data in a large-sized 3D volume, e.g., the scattered vessel fragments in the angiographic data against the otherwise encompassing white and grey matter (as well as background and noise); *high noisiness*, such as low signal-to-noise ratio (SNR), e.g., about 10:1 in cerebrovascular images; tininess of micro-level vessels, e.g., the diameter of the micro-level vessels in images is merely 1~2 voxels (e.g., 50~100 microns); and sophisticated vessel geometry and topology variations, e.g., local “crossing”, “kissing”, or “tortuous” vessel structures, etc. Currently, for such complex 3D micro-level data, it would be impossible from a timing perspective for clinicians to review all this data manually and label abnormalities slice by slice. The 3D structural / contextual information and quantitative metrics are still missing, although maximum intensity projection (MIP) [31], a widely-used approach for qualitatively visualizing and analyzing the 3D vasculature, has been employed to enhance the local vessel signal, allowing for geometric variability and scalability. The labor-intensive, time-consuming, and 3D global / contextual information-missing nature of the procedure makes it very challenging to fully take advantage of the

large number of 3D datasets (images and shapes) available for reference and comparison, and reach more informed and accurate decisions.

In recent decades, the automatic model-driven vessel extraction and segmentation approaches have been proposed, such as multiscale filtering [12], region growing techniques [27], active contours [30], geometric flow [8], level-set approach [11], nonlinear subtraction (NLS) method [47], template-based predictor-corrector algorithm [14], etc. However, these approaches are easily overwhelmed by tons of low-level handcrafted features and complicated manual parameter adjustment to overcome aforementioned difficulties and subject variations.

Recently, data-driven approaches have been proposed to robustly investigate the correlations between different objects / instances without relying on hard-coded metrics. In medical image visualization and processing, several deep learning based methods have been proposed to extract vessels from 2D retinal images, such as DeepVessel [13], multi-level deep supervised networks [29], deep neural network (DNN)-based method [26], unified convolutional neural network (CNN) and graph neural network (GNN) [37], etc. These methods can perform 2D vessel segmentation tasks well, but are far from satisfactory on 3D vessel scenario. There are still very few dedicated deep learning architectures for 3D vessel segmentation, such as Uception [34], DeepVesselNet [40], and VesselNet [22], etc. Existing methods do not consider to use the visualization techniques in the 3D vessel extraction and are not specifically designed for solving the aforementioned challenges in 3D micro-cerebrovascular segmentation.

The fundamental motivation of the proposed work is to present a new *visualization-guided computing paradigm* to combine direct 3D volume processing and volume rendered clues for effective 3D exploration. In order to fill the gap in the high-fidelity 3D micro-cerebrovascular segmentation and visualization for the medical data *in-vivo*, we present a DNN method, *VC-Net*, for robustly extracting sparse microvascular structures through embedding the 2D image slice composition by MIP into the 3D volumetric image learning process to enhance the overall performance on 3D vasculature segmentation. The core *novelty* is to automatically leverage the volume visualization technique (e.g., MIP –

• *Y. Wang, G. Yan, H. Zhu, J. Hua, and Z. Zhong are with the Department of Computer Science, Wayne State University, Detroit, MI 48202. E-mail: {yifan.wang2,guoliyan,hkzhu,jinghua,zichunzhong}@wayne.edu.*

• *S. Buch, Y. Wang, and E. M. Haacke are with the Department of Radiology, Wayne State University, Detroit, MI 48201. E-mail: {sagarbuchmri,neuying}@gmail.com, nmrimaging@aol.com.*

Manuscript received xx xxx. 201x; accepted xx xxx. 201x. Date of Publication xx xxx. 201x; date of current version xx xxx. 201x. For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org. Digital Object Identifier: xx.xxxx/TVCG.201x.xxxxxx

a volume rendering technique for 3D volume images) to enhance the qualitative 3D data exploration, especially for 3D *in-vivo* segmentation and visualization, at the deep learning level. It is noted that the proposed framework can better capture the micro vessels and improve the vessel connectivity. The key motivation of our network is to integrate the trustworthy auxiliary from learned 2D MIP features into the 3D volume segmentation and visualization network, instead of using more complicated networks empirically. Experimental results are evaluated and compared with the traditional 3D vessel segmentation methods and the state-of-the-art in deep learning, using extensive public and real patient (micro-)cerebrovascular image datasets. The key *contributions* of our work are as follows:

- It proposes an effective end-to-end deep learning method to segment and visualize high-fidelity 3D sparse microvascular structure with complicated geometry and topology variations from volumetric images with significant noise.
- A multi-stream CNN framework is designed to effectively learn the feature vectors of 3D raw volume and multislice composited 2D MIP (volume rendering), respectively, and explore inter-dependencies between 3D and 2D embedded features in a joint volume-composition embedding space by unprojecting (inverse volume rendering) the 2D features, learned from MIP, into the 3D volume embedding space.
- To our knowledge, this is the first time that a deep learning framework is proposed to construct such a joint convolutional embedding space, where the computed joint vessel probabilities from 2D projection and 3D volume can be integrated synergistically.
- The application and experiments on the accurate *in-vivo* segmentation and visualization of sparse and complicated 3D microvascular structure facilitated by our method demonstrate the potential in a novel and powerful MR arteriogram and venogram (MRAV) diagnosis of vascular disease.

2 RELATED WORK

In this section, we review most related work on 2D / 3D vessel extraction and segmentation in visualization and medical imaging domains.

2.1 Model-Driven Vessel Extraction and Segmentation

Traditionally, doctors have to manually segment each image slice to obtain accurate vessel structures, which is extremely tedious and time-consuming. Therefore, it is important to develop automatic vessel segmentation methods. For instance, Wilson and Noble [45] introduced a mixture distribution for the data, motivated by a physical model of blood flow, that is used in a two-stage segmentation algorithm with a statistical classifier and structural criteria. Chung and Noble [6] presented an extended version of the previous 3D cerebral vessel segmentation algorithm [45], and introduced a Rician distribution for background noise modeling and used a modified expectation-maximization (EM) algorithm for the parameter estimation procedure. Frangi et al. [12] developed a vessel enhancement filter by computing the multiscale second order local structure of an image (i.e., Hessian). A vesselness measure is obtained on the basis of all eigenvalues of the Hessian. Based on multiscale filtering method [12], Descoteaux et al. [8] developed a novel geometric flow for segmenting vasculature in proton-density images, which can also be applied to the cases of magnetic resonance angiography (MRA) or MRI data. Martínez-Pérez et al. [27] presented a retinal blood vessel segmentation method based on scale-space analysis of obtaining the vessel geometrical features by the first and the second derivative of the intensity in the image. Then they used a multiple pass region growing procedure which progressively segments the blood vessels. Nain et al. [30] combined image statistics and shape information to derive a region-based active contour that segments tubular structures and penalizes leakages. Liao et al. [25] introduced a fast marching approach with curvature regularization for vessel segmentation, since most vessels have a smooth path and curvature can be used to distinguish desired vessels. Florin et al. [10] proposed a particle filter based propagation approach for the segmentation of vascular structures in 3D volumes. To obtain posterior probability estimation of the vessel

location, Wang et al. [43] employed sequential Monte Carlo tracking and proposed a vessel segmentation method by fusing multiple cues extracted from CT images for enhanced segments from global path minimization. Forkert et al. [11] presented and evaluated a level-set segmentation approach with vesselness-dependent anisotropic energy weights, which focuses on the exact segmentation of malformed as well as small vessels from time-of-flight (TOF) MRA datasets. Ye et al. [47] proposed non-linear subtraction (NLS) method [47], which is employed for selective MRA enhancement utilizing the flow rephrased and dephased images. Then the vessel label can be obtained based on an enhanced angiography map. Govyadinov et al. [14] described a template-based predictor-corrector method for tracing filaments that is robust in microvascular datasets, and applied a number of glyph-based visualization techniques to represent the aggregated and biologically relevant information of the extracted microvascular network. Then, they developed a bi-modal visualization framework [15], leveraging graph-based and geometry-based techniques to achieve interactive visualization of microvascular networks. However, these approaches are exhausted by handcrafted features (e.g., gradients of the intensity, second order local structures, maximum principal curvatures) and complicated manual parameter adjustment to adapt to the subject variations. Therefore, their robustness and accuracy across subjects are limited.

2.2 Data-Driven Vessel Extraction and Segmentation

Recently, there is an emerging trend to automatically extract, segment, and reconstruct shape objects of interest from input 2D / 3D images [9, 42, 44] or 3D meshes / point clouds [21, 23, 28, 32, 46] by deep neural network (DNN) [3, 18]. Particularly for vessel structures, several deep learning based methods have been proposed to extract vessels from 2D retinal images. DeepVessel [13] addresses retinal vessel segmentation as a boundary detection task that is solved using a CNN with a side-output layer to learn discriminative representations, and a conditional random field (CRF) layer that accounts for non-local pixel correlations. Li et al. [24] presented a supervised method for vessel segmentation by using the cross-modality data transformation from retinal image to vessel map. Mo and Zhang [29] developed a deep supervised fully convolutional network by leveraging multi-level hierarchical features of the deep networks for retinal vessel segmentation. Liskowski and Krawiec [26] proposed a supervised segmentation technique that uses a DNN trained on a large number of samples preprocessed with global contrast normalization, zero-phase whitening, and augmented using geometric transformations and gamma corrections. Shin et al. [37] incorporated a graph neural network (GNN) into a unified CNN architecture to jointly exploit both local appearances and global vessel structures. Their framework has been evaluated on retinal image datasets and a coronary artery X-ray angiography dataset. These methods can perform well on the 2D vessel segmentation task, but are far from satisfaction / feasibility on 3D micro vessel scenario, since their designs either do not consider the correlation / inter-information between slices in 3D volumetric images or cannot afford the computational and memory burdens in the large 3D volume at the micro-level.

As for deep learning-based 3D vessel segmentation, for instance, Uception [34] presents a network inspired by the 3D U-Net [7] and the Inception modules [38] for segmentation of the cerebrovascular network in MRA images. DeepVesselNet [40] and VesselNet [22] propose 2D orthogonal cross-hair filters in all sagittal, coronal, and axial planes on each voxel to make use of 3D context information at a reduced computational burden and memory cost. However, the challenging problems in 3D micro-cerebrovascular segmentation are complicated vessel geometry and topology, high sparseness and noise of vessel data in a large-sized 3D volume, and the limited resource of 3D microvascular datasets. The above methods do not overcome these challenges.

3 VC-NET

The fundamental inspiration of the proposed work is to mimic the observation of human exploration in 3D aided by volume rendering. Our work presents a new paradigm to combine direct 3D volume processing and volume rendered clues for effective 3D exploration. For instance in Fig. 1, for a micro-cerebrovascular dataset, the 3D volume image can

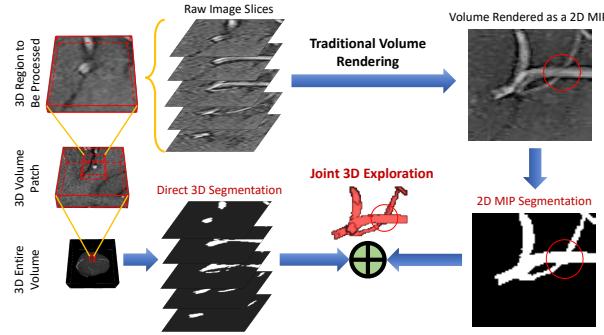


Fig. 1: A novel 3D data analytic paradigm in a joint volume-composition space, where volume rendered results are used to support the visualization-guided 3D volume processing by deep learning.

more accurately represent the 3D spatial information, but the desired task is easy to be confused by challenging SNR and sparse vesselness, as shown in the raw image slices; while the volume rendered 2D MIP image can better enhance the local vessel signal by enforcing vessel continuity and adapt to the geometric variability and scalability of vessels. However, it always lacks 3D spatial sense, e.g., two “crossing” and “kissing” vessels as circled in red (in fact, the bigger one is above the smaller one in 3D space). So, it is deficient to investigate the (sparse and noisy) 3D data from either 3D volume or volume rendered 2D MIP, respectively. In this work, we design a novel paradigm to support the 3D data analytics, such as segmentation, etc., by using the visualization-guided computing. Instead of conducting the rendering / composition at the final stage as in the traditional visualization pipelines, this paradigm qualitatively investigates the 3D volume data from 2D composited (rendered) images. Essentially, this procedure makes the visualization more important via an early and simultaneous involvement of volume rendering (composition). Finally, we explore the 3D data analytics in a joint volume-composition space. In the following, we introduce the components of the VC-Net model: network architecture and loss function, and dataset generation and preparation.

3.1 Network Architecture

The proposed VC-Net mainly consists of a dual-stream component (i.e., a 3D volume segmentation stream and a 2D composited MIP segmentation stream) and the bi-directional operations between these two streams (i.e., 3D-to-2D projection and 2D-to-3D unprojection). The overall architecture is demonstrated in Fig. 2. The two-stream segmentation component can learn vessel feature vectors in 3D volume and corresponding multiple 2D MIPs (enhanced and dense depiction of 3D relationships via a 3D-to-2D projection computation) contexts, respectively. After that, the embedded features from the 2D composited MIP are transformed from the 2D MIP domain into the 3D volume domain through a 2D-to-3D unprojection process. Then, the extracted 2D and 3D embedded features from two streams are integrated together, constructing a unified high-dimensional joint convolutional embedding space, which can strengthen the original sparse vessel features from the 3D volume. Finally, the vessel segmentation prediction can be learned at the fusion stage in this joint convolutional embedding space.

In this work, we use a 3D U-Net [7] as the 3D volume segmentation stream and a half 2D U-Net [33] (in terms of feature channel numbers) as the 2D composited MIP segmentation stream, respectively. U-Net-like networks are the most commonly-used and robust medical imaging segmentation neural networks across different data modalities for varying organ / tissue geometries, and thus it is suitable for us to justify the benefits from our 2D-to-3D unprojection and joint embedding of 3D volume and 2D composited MIP. A U-Net-like network is essentially a convolutional encoder-decoder network, which first embeds the input into a high-dimensional feature vector through hierarchical convolution and pooling at the encoder stages, and then decodes the feature vector in the hidden space through hierarchical upsampling and convolution at the decoder stages with the integration of the features directed from different encoder stages through the long-skipped feature concatenations.

In Fig. 2, the layer output feature channel numbers are denoted in the corresponding blocks and layer input spatial dimensions are shown in the horizontal levels of every block.

Due to the limited data availability and large volume size in micro-cerebrovascular image datasets, we choose to train the network patch-wisely. Specifically, from the observation that most brain MRAs have much higher resolutions in axial plane than other planes, we adaptively train our network using non-cubic patches, which have larger dimension size across axial plane, instead of resizing the data into a uniform voxel spacing through an interpolation before the network training, to avoid potential data corruption. As shown in Fig. 2, the key step in our network is the effective integration of the features from two different streams / domains. In order to fuse the 2D composited MIP stream into the 3D segmentation main stream within the network during the learning process, one may first find out that segmentation task is essentially a dense voxel (pixel)-wise classification problem, and the integration of embedded feature vectors from different learning domains and objectives (i.e., 3D volume segmentation and 2D composited MIP segmentation) must be fused voxel-wisely with the correct spatial correspondence in the volumetric domain. Accordingly, there are two main challenges that need to be overcome in this work. The first one is the effective format of the corresponding 2D composited MIPs from a randomly-extracted 3D volume patch that is adequate and suitable for delivering dense voxel correspondence in the simultaneous dual-stream learning design. The second one is the effective approach for mapping / unprojecting the feature vectors extracted from the composited MIP image plane pixels (in a dimension-reduced 2D) back to the corresponding 3D volume spacing voxels. More details are introduced in following subsections.

3.1.1 3D-to-2D Projection in Dual-Stream Design

The major motivation for projecting the 3D volume space into the 2D MIP space is to enhance the local vessel probability. Given a randomly-extracted 3D volume patch V of the size $K_1 \times K_2 \times K_3$ (e.g., we use $128 \times 128 \times 16$ in our experiments) and K_3 is along the vertical axis. We compute s -sliced (e.g., $s = 5$ in our experiments as suggested by domain experts) MIPs of V along vertical axis with overlapping coverage every t slice interval. Consequently we can get a set of m consecutive / sliding MIPs, i.e., $\mathbf{P} = \{P_1, P_2, \dots, P_k, \dots, P_{m-1}, P_m\}$, in which P_k is the MIP across the $[(k-1)t+1]^{th}$ slice to the $[(k-1)t+s]^{th}$ slice in V . It is noted that in a 2D MIP, only one voxel with the maximum intensity among the s voxels along the vertical axis in V will be recorded, which is prone to an information loss, considering the segmentation task actually needs the information of every voxel. Consequently, we set $t = 2$ as a trade-off between computation cost and information completeness / denseness. We can get m MIPs of size $K_1 \times K_2$ for V , where the MIP number m is computed as:

$$m = \left\lfloor \frac{1}{t} (K_3 - s) \right\rfloor + 1. \quad (1)$$

A MIP conveys denser vessel information and is also naturally suitable for 2D convolution. However, we now have m different MIPs and need to feed them to our network in the MIP stream in company with the 3D volume stream V as an input pair to our entire network. The information from the m MIPs is equally important, which means every pixel information should be kept during learning for later back projection. In order to avoid intuitively stacking them to a $K_1 \times K_2 \times m$ volume such that the 2D CNN (in 2D MIP stream) would essentially treat it as a 2D input of a spatial dimension $K_1 \times K_2$ with m different properties (feature channels), which is essentially deficient in terms of the spatial domain size as well as the operation motivation, we convert the m MIPs to a tiled MIP with a larger 2D spatial size, such as $0.5mK_1 \times 2K_2$. In this case, the 2D convolution is operated equally across the 2D composited MIP plane domain. The slice indices from where the MIP pixels are selected in the original V are also recorded so as to effectively restore the pixel-wise information extracted from MIP to the 3D volume space, which will be used in the 2D-to-3D unprojection in the following process. The format of the 2D composited MIP (e.g., $m = 6$ consecutive MIPs) computed from a 3D volume patch is shown in Fig. 3 (a).

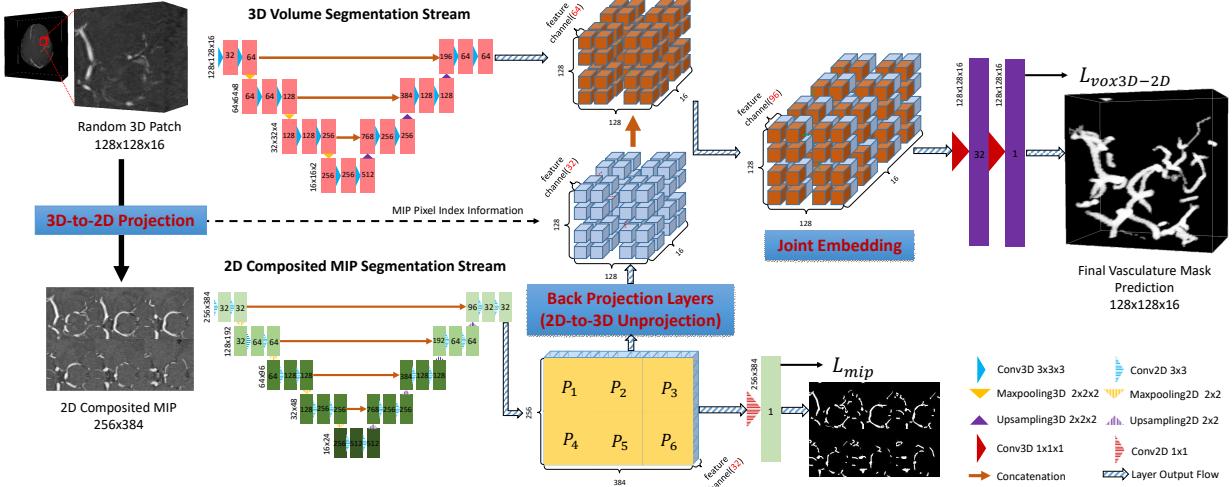


Fig. 2: The architecture of VC-Net. The major procedure includes obtaining the composited MIPs via 3D-to-2D projection, dual-stream segmentation learning for 3D volume and 2D composited MIP feature vectors, back projecting 2D composited MIP feature vectors into the 3D volume feature space via 2D-to-3D unprojection, building a joint convolutional embedding for learning the final vasculature mask.

3.1.2 2D-to-3D Unprojection for Joint Embedding

Once the 3D volume and 2D MIP streams learn their segmentation features respectively, we intend to integrate them in a unified joint hidden feature embedding space to yield the final 3D segmentation prediction. In order to achieve this, we conduct several operations within our network to unproject (i.e., back project) the pixel features extracted from the composited MIP back to their corresponding 3D voxel feature space.

The final-stage hidden feature from 2D composited MIP segmentation stream has the size $0.5mK_1 \times 2K_2$ with C_1 channels ($C_1 = 32$ as shown in Fig. 2), which is the input of the back projection layers. We first disassemble it to restore $m C_1$ -channel features for the corresponding MIPs (e.g., $P_1, P_2, \dots, P_{m-1}, P_m$, where $m = 6$ as illustrated in Fig. 3 b). Then we use the recorded index information to map the MIP pixel features back to where they are selected from V during the 2D composited MIP generation. Fig. 3 (b) shows how the feature vectors of two consecutive MIPs (e.g., P_1 and P_2) are disassembled from the composited MIP. They unproject their pixel feature space (P_{m-j} , i.e., the j -th slice among 5-sliced MIP P_m , $1 \leq j \leq 5$) back to the voxel feature space (S_n , i.e., the n -th slice in the input 3D patch, $1 \leq n \leq 16$). It is noted that the feature dimension is folded from 3D to 2D for a convenient illustration in Fig. 3 (b) (i.e., hiding the feature dimension).

For the features of overlapping slices (from the consecutive MIPs), which are covered by multiple MIPs, we take the element-wise maximum value across the overlapping restoration through the feature channels:

$$F_{S_n}[i] = \max(F_{P_{1-1}}[i], \dots, F_{P_{6-5}}[i]), 1 \leq i \leq 32, \quad (2)$$

where $F_{S_n}[i]$ represents the i -th channel in feature F at the n -th slice in the 3D patch. For example, the feature F_{S_9} is computed across the overlapping slices of $P_{3-5}, P_{4-3}, P_{5-1}$ as highlighted in pink in Fig. 3 (c). The whole process of the cross-MIP fusion in the feature channels of the 3D volume feature space is shown in Fig. 3 (c) in detail. After that, the unprojected 2D MIP features and 3D volume features from two streams are integrated together, constructing a unified high-dimensional joint convolutional embedding for predicting the final vessel segmentation.

3.1.3 Loss Function

The major learning objective of our VC-Net is to extract the sparse 3D vasculature structure from the 3D MRI volume image using a 3D segmentation network supplemented by information from multiple denser and more connected 2D MIPs. Consequently the network loss function consists of two terms:

$$L = L_{vox_{3D-2D}} + \lambda L_{mip}, \quad (3)$$

where $L_{vox3D-2D}$ is a joint 3D-2D segmentation Dice loss adopted in 3D volume stream and defined as:

$$L_{vox_{3D-2D}} = -\frac{2\Sigma_{x \in V} p(x)g(x) + \delta}{\Sigma_{x \in V} p(x) + \Sigma_{x \in V} g(x) + \delta}, \quad (4)$$

where $p(x)$ and $g(x)$ are the predicted voxel-wise vessel probability maps and ground truth binary labels within the query volume patch V , respectively. δ is a small smooth constant. L_{mip} is applied in 2D composited MIP stream and acts as a regularization term during the learning process, which is also a Dice loss function defined (similarly to $L_{vox3D-2D}$) within the 2D composited MIP plane and supervised by the ground truth MIP vessel binary labels. λ is the constant coefficient of L_{mip} , which is set to be 0.2 for our best experiment performance.

3.2 Dataset Generation and Preparation

In this work, we use two different real patient datasets to evaluate our proposed VC-Net method.

Novel MICRO-MRI Imaging and Dataset. Some researchers have recently developed a next generation of microvascular imaging, i.e., Microvascular In-vivo Contrast Revealed Origins Magnetic Resonance Imaging (MICRO-MRI) [36, 41]. Thanks to MICRO-MRI, we became the first ever to be able to acquire such brain imaging datasets and observe the complicated micro cerebral vessels. This dataset is produced by neurologists and radiologists within our collaborative group. Data was acquired with an adapted 3D gradient echo susceptibility weighted imaging (SWI) sequence [5] collected from a 3T MR scanner. The post-contrast data were acquired during a gradual increase in dose (final concentration = 4 mg / kg). Eleven healthy volunteers were scanned in brain regions with a dual echo SWI sequence at four time points: the first was acquired pre-contrast and the remaining three were acquired post-contrast during a gradual increase in dose delivered over the time frame of 20 min; with the imaging parameters: echo time (TE)1 / TE2 / repetition time (TR) = 7.5 / 22.5 / 27 ms, bandwidth = 180 Hz / ppxl, flip angle = 15° (pre-contrast and final post-contrast data) and 20° (first and second post-contrast data). The voxel spacing is $0.22 \times 0.22 \times 1 \text{ mm}^3$ with a volume size of $1024 \times 832 \times 96$ voxels.

Major-level vessel data. This protocol enables multiple image sources for producing both MR arteriogram (MRAG) and venogram (MRVG). For the MRVG, the pre-contrast quantitative susceptibility mapping (QSM) and R_2^* constitute two different representations of veins. In order to obtain the pre-contrast QSM data, the original phase data was unwrapped using the 3D best path method [2]. The sophisticated harmonic artifact reduction for phase data (SHARP) method

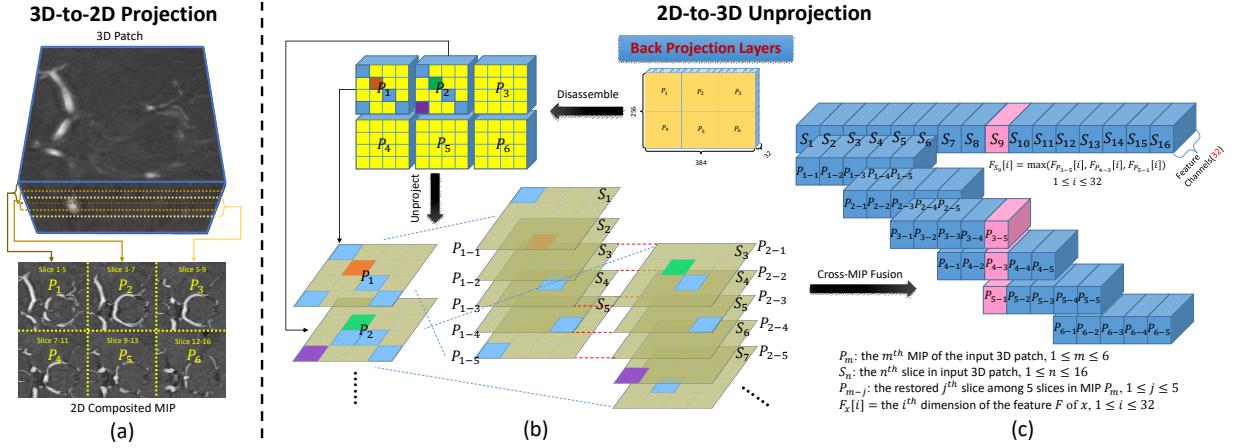


Fig. 3: (a) Illustration of the 3D-to-2D projection in the spatial domain for computing a 2D composited MIP from a 3D volume patch. (b) and (c) Illustration of the detailed computations in back projection layers for 2D-to-3D unprojection process in the embedded feature domain. As illustrated in bottom (b), the consecutive MIPs P_1 and P_2 with overlapping slice coverage of S_3 , S_4 , S_5 contribute to information completeness in 3D patch volume. A pixel location on the 5-sliced MIP 2D plane which keeps the feature information of only one voxel out of five (e.g., the middle orange pixel and the left bottom blue pixel in P_1 are back projected to S_2 and S_5) now can be supplemented by P_2 's back projection (e.g., the middle green voxel on S_3 and the left bottom purple voxel on S_6).

was used to estimate the background field and remove it from the unwrapped phase [35]. The truncated k-space inverse filter approach with an iterative geometric constraint (also known as iSWIM) was applied to the resultant phase to generate the QSM data [16, 39]. The QSM data was further refined by removing the strong phase gradients from the long TE phase data based on a quality phase mask. The resultant phase was used to obtain a QSM data QSM_{TE2} . The QSM of the short TE data QSM_{TE1} was also generated, but without using a quality map, since at a low TE the phase gradients were not that strong. Finally, the missing information on QSM_{TE2} was filled in by applying an inverted quality mask to QSM_{TE1} . To obtain the pre-contrast R_2^* , the short and long TE magnitude data $S(t)$ were fitted to the monoexponential equation: $S(t) = \rho e^{-(tR_2^*)}$, where ρ is the tissue intrinsic proton density.

Another MRVG was generated by subtracting the short TE magnitude data of pre-contrast from the short TE magnitude data of the first post-contrast. The above-mentioned subtraction provides a venous-only map V_{T_1} . The QSM, R_2^* , and V_{T_1} maps were then normalized to values between 0 and 1, and an average of these different sources produced a high-quality MRVG referred to as $MRVG_{avg}$.

An MRAG was then calculated using a nonlinear subtraction (NL-S) [47], i.e., $MRAG_{nls}$, of the long TE S' from the short TE S of the pre-contrast magnitude data as: $MRAG_{nls} = S^2 - \alpha S'^2$, where α is a constant with an empirically selected value of 1.5. Due to the T_2^* effect, this subtraction also enhances the veins, but to a much smaller extent than the arteries. Nevertheless, any venous enhancement is discarded by using a mask generated from $MRVG_{avg}$. Finally, the ultimate ground truth vessel labels are obtained by integrating the enhanced angiography (i.e., arteriogram and venogram) maps [4] from the computed $MRAG_{nls}$ and $MRVG_{avg}$, with a threshold-based method for the initial masks, followed by domain experts' post-manual labeling refinement using our developed cerebrovascular labeling and visualization tool. Supplemental Material and Video are included for demonstrating the interactive interface and basic functions in detail.

Micro-level vessel data. SWI images were generated by homodyne high-pass filtering (filter size = 96×96) the phase images to generate a phase mask, which was multiplied with the original magnitude images four times, for all time points [17]. All the original magnitude and corrected phase data were then registered to the pre-contrast data. The short TE (7.5 ms) magnitude data of pre-contrast and the first post-contrast time points were averaged. This averaged magnitude data was subtracted by the long TE (22.5 ms) SWI data from the last post-contrast time point (4 mg / kg) to enhance the vessels. The vessels were further enhanced on the resultant subtracted image by applying the vesselness algorithm [12] to obtain the micro-level vessel map. The micro-level vessels from this resultant vessel map were extracted using

an adaptive threshold-based region growing method (ATRG) [20], i.e., SWI_{ATRG} , as the initial masks, followed by domain experts' manual inspection of the extracted vessels for quality control.

Public MRA Dataset. In order to compare with the existing methods, we use a public TubeTK Toolkit MRA dataset from University of North Carolina at Chapel Hill [1], acquired by a 3T MR system. There are 42 patient cases in the whole dataset, which have the manual-labeled vessel segmentation masks. The voxel spacing of the MRA images is $0.5 \times 0.5 \times 0.8 \text{ mm}^3$ with a volume size of $448 \times 448 \times 128$ voxels.

4 RESULTS

For both MRA TubeTK and MICRO-MRI datasets (the different modalities of input image examples are provided in Supplemental Material), we first apply the MR-based skull-stripping method [19] to extract the pure brain from each image. As we mentioned in Sec. 3.1, our VC-Net network is designed for patch-wise training and the 3D training patches with the imbalanced dimensions are randomly-extracted with overlapping focusing on the brain area in the whole 3D MRA / MICRO-MRI, e.g., 80 patches for each TubeTK case and 440 patches for each MICRO-MRI major-level vessel case. The random training / validation / testing case split is 33 / 3 / 6 and 6 / 2 / 3 for the TubeTK dataset and the MICRO-MRI major-level vessel case, respectively. All the numerical evaluations are reported in terms of whole brain volume image patched with no overlapping.

Our VC-Net adopts the Adam optimizer with 0.0001 as an initial learning rate, 0.5 as the learning decay factor, and 10 epochs as the learning patience across all datasets. In our implementation, we restrict our batch size to 4 due to hardware limits. No batch normalization is adopted in either stream in VC-Net and we use ReLU (Rectified Linear Unit) activation for both 2D and 3D convolutional layers in corresponding streams and sigmoid activation for the final vessel probability output from both 2D and 3D streams. The network is implemented in TensorFlow framework and the total training time is around 10 hours on two NVIDIA GeForce GTX 1080 GPUs with 8 GB GDDR5X memory. The inference time is given in the following subsection. Data and source code of this work will be made available.

The performance of our VC-Net and all methods in comparison are numerically evaluated by the following three quantitative metrics, which are defined from the classifier confusion matrix from different aspects:

Dice Similarity (Dice), $2TP/(2TP + FP + FN)$, (the same as F-score under most of the circumstances) generally measures the intersection over union between prediction and ground truth. It involves true positive (TP), false positive (FP), and false negative (FN), so as to be the most comprehensive indicator to evaluate the sparse vessel segmentation in a large portion of background, i.e., true negative (TN).

Precision, $TP/(TP + FP)$, measures the model ability of ruling out the noise contributions and obtaining the correct vessel voxels.

False Positive Rate (FPR), $FP/(FP + TN)$, examines the model ability of distinguishing the real background and noise against vessels, which is crucial for the clinical purpose.

The best results in tables are shown in bold font. Here we do not include the metric of Accuracy, due to its extremely high value (e.g., $\geq 99\%$) for all methods. The reason is that it involves dominant portion of background (true negative) together with highly sparse target (e.g., the segmented vessels in our task) in computation and consequently loses its effectiveness for segmentation evaluation.

4.1 Comparison with the State-of-the-Art

We first compare our VC-Net performance on TubeTK dataset with four state-of-the-art deep learning based methods (i.e., 3D U-Net [7], 2D U-Net [33], DeepVesselNet [40], and Uception [34]) and one classical parametric intensity-based method (i.e., vesselness algorithm [8, 12]) in 3D vessel segmentation. All deep learning methods in comparison are trained until convergence by using the same dataset split or using the results reported from their original publication (such as Uception). For 2D U-Net, we train it with 128×128 2D patches, whose amount is over 10 times of the 3D patch amount extracted for the 3D CNN based methods in comparison with on-the-fly data augmentation for a fair data acquisition. For DeepVesselNet, we have tried different combinations of their data pre-processing process and chosen the image intensity clipping for obtaining an optimal performance on TubeTK dataset.

The quantitative performance comparison of these methods on TubeTK dataset is shown in Tab. 1. ‘–’ means ‘not applicable’ due to lack of their implementations or results. Here we also provide the per-volume inference time and the parameter number to evaluate the model efficiency besides the segmentation performance. From Tab. 1, we can see that our VC-Net has overall the best segmentation performance among all the methods on TubeTK dataset. With the 2D composited MIP feature integration, our network performs better than a pure 3D U-Net [7] over the three different metrics on segmentation results. The qualitative comparison of MIP-wise (e.g., 5-sliced) segmentation results and 3D global vessel segmentation results between our VC-Net and 3D U-Net (one of the most robust state-of-the-art deep learning based methods for biomedical image segmentation) is shown in Fig. 4 (a). With the 2D composited MIP complementary information, the final vessel segmentation shows better connectivity and better small vessel capturing as marked in red circles (3D global vessel segmentation visualization) and green circles (2D MIP vessel segmentation visualization). Besides the segmentation performance gain, the increase of time and space complexities in VC-Net is not high compared with a standalone 3D U-Net as shown in Tab. 1, since only a half 2D U-Net (i.e., 7.8 million parameters) is involved in the 2D MIP stream. On the other hand, since the computational complexity of 3D convolution operations apparently overweights that of 2D convolution operations, the 3D stream still dominates the computational complexity of the entire VC-Net. Another observation is that the 3D U-Net greatly outperforms 2D U-Net [33] even if the latter contains many more feature embedding channels, since the former method is able to capture the cross-slice continuity and that is why 3D CNN should be involved in such sparse 3D object segmentation with complex topology. Moreover, the full 2D U-Net implies much larger amount of 2D convolution operations and model parameters, which lead to the unsatisfactory model efficiency. DeepVesselNet [40] fails to yield a good performance as they reported in their own dataset, which could result from the lack of the pre-training procedure, i.e., a relatively complicated data pre-processing, and the instability of their loss function, which is severely sensitive to the training perturbation. In addition, their light-weighted network only consists of five convolutional layers for high efficiency, whereas its simplicity may undermine its cross-dataset robustness. The qualitative comparison with DeepVesselNet is given in Fig. 4 (b), from which you can see that the DeepVesselNet result is much noisier and has severer connectivity issues. Our method also outperforms the best Uception result reported in [34] on TubeTK dataset with even less data pre-processing procedure, which implies that empirical neural network

modification to increase the model complexity does not guarantee a better performance all the time. To be comprehensive, we apply the vesselness algorithm [8, 12] as a traditional benchmark method for comparison based on the available data modality in TubeTK dataset, which is a widely-used approach to segment the cylindrical vessel structures in medical field. As shown in Tab. 1, all deep learning methods greatly outperform classical vesselness method on TubeTK dataset with much higher efficiency, which is beneficial from the essence of deep learning techniques: mostly end-to-end, more adaptive non-linearity, controllable ambiguity for feature extraction and integration process than traditional complex manual parameter-driven algorithms to boost the overall robustness and accuracy for real-world image learning tasks. More qualitative comparisons with 3D U-Net and DeepVesselNet are provided in Supplemental Material.

Table 1: Quantitative performance evaluation of different methods on TubeTK dataset.

| Methods / Metrics | Dice (%) \uparrow | Precision (%) \uparrow | FPR (%) \downarrow | Time (s) \downarrow | # Para. \downarrow |
|-------------------|---------------------|--------------------------|----------------------|-----------------------|----------------------|
| Ours | 71.81 | 76.66 | 0.0821 | 10.8 | 24 M |
| 3D U-Net | 71.01 | 74.00 | 0.0958 | 7.5 | 19 M |
| Uception | 67.01 | – | – | – | – |
| DeepVesselNet | 64.12 | 63.75 | 0.1465 | 1.8 | 0.06 M |
| 2D U-Net | 65.10 | 70.05 | 0.1041 | 16.7 | 31 M |
| Vesselness | 37.71 | 47.69 | 0.1393 | 186.6 | – |

4.2 Experiments and Evaluation on MICRO-MRI Dataset

As mentioned in Sec. 2, most of the recently related work on brain vasculature segmentation tasks is limited to major arteries (in Sec. 4.1) since currently most of the available brain MRI image datasets with adequate amount and consistent quality are MRAGs. However, with assistance from the neurologists and radiologists under our collaboration, we can now extend VC-Net from general artery segmentation to the vasculature segmentation of major artery and major vein, separately. More inspiringly, we also demonstrate that our VC-Net is capable of extracting the micro vessels in the complicated real patient MICRO-MRIs. It is noted that the segmentation of brain vessels becomes more challenging in micro-level than major-level, and more challenging in veins than arteries. The following experiments show that our method has bigger improvements on more challenging cases (i.e., micro-level vessel and major-level vein segmentations) compared with other methods.

4.2.1 Major-Level Artery Segmentation and Visualization

The MRAGs in clinic MICRO-MRI datasets under our collaboration as mentioned in Sec. 3.2 focus on midbrain area from where the major-level vessels are relatively denser and more observable. Currently our collaborative domain experts apply the state-of-the-art model-driven NLS method [47] (i.e., MRAG_{nls}) followed by case-wise threshold selection to extract the clean midbrain vessels, which requires different data modalities and tedious manual parameter-tuning as stated in Sec. 3.2. However, from Fig. 5 (a) we can see that its segmentation result still fails to be free from location-dependent interference, such as superior sagittal sinus (red dotted circles in 3D visualization and green dotted circles in MIP visualization) and some random scattered voxel noise (red solid circles). Aiming to improve the segmentation performance with less manual-parameter tuning and less modality requirement yet provide a much more efficient method for major artery extraction that is well applicable for future patient case collection, we train our VC-Net with only TE1 pre-contrast SWI (a single-modal MRAG) data as input. From Tab. 2 we can see that our quantitative evaluation results outperform the MRAG_{nls} method on all metrics even if the latter one integrates and enhances artery signal from several different data modalities. Here we also include our numerical comparison with 3D U-Net (under the same experiment setting), the most competitive method on TubeTK dataset to show our network’s cross-dataset robustness and superiority. One can also observe that the numerical difference of the performance in MICRO-MRI major-level artery dataset is not as obvious as the major-level vein and micro-level vessel datasets as shown in the following two subsections, which may result from the fact that the MRAGs are relatively clearer in terms of the image dose effect and the noise type. More qualitative comparisons with the MRAG_{nls} method are provided in Supplemental Material.

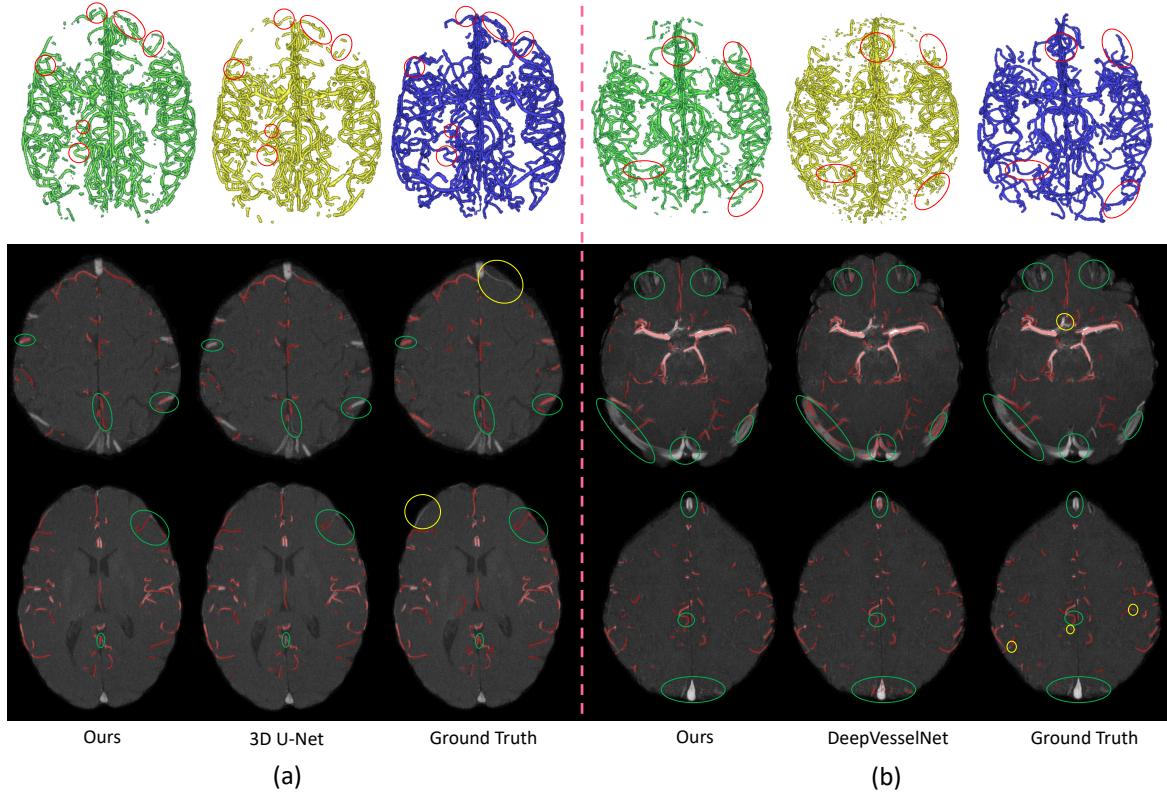


Fig. 4: Some qualitative comparison results from TubeTK dataset: The 3D global vessel segmentations are shown from superior direction. The MIP segmentations are visualized by 5-sliced MRA images, and the corresponding vessel masks in MIPs are marked in semi-transparent red. The highlighted comparison areas are marked in circles. Yellow-circled areas are some minor mistakes in the ground truth (discussed in Sec. 4.2.4).

Table 2: Quantitative performance evaluation of different methods on major-level artery segmentation.

| Methods / Metrics | Dice (%) \uparrow | Precision (%) \uparrow | FPR (%) \downarrow |
|---------------------|---------------------|--------------------------|----------------------|
| Ours | 82.98 | 83.69 | 0.0337 |
| 3D U-Net | 82.64 | 83.52 | 0.0342 |
| MRAG _{als} | 80.60 | 82.07 | 0.0354 |

Table 3: Quantitative performance evaluation of different methods on major-level vein segmentation.

| Methods / Metrics | Dice (%) \uparrow | Precision (%) \uparrow | FPR (%) \downarrow |
|---------------------|---------------------|--------------------------|----------------------|
| Ours | 76.46 | 82.20 | 0.0849 |
| 3D U-Net | 76.02 | 80.40 | 0.0946 |
| MRVG _{avg} | 73.73 | 64.99 | 0.2294 |

4.2.2 Major-Level Vein Segmentation and Visualization

Currently, the MRVGs are not readily and directly available from the scanner, so there is no raw MRI image that can produce pure veins. There are different ways that people have used to derive it such as the SWI, QSM, or R_2^* data, where the veins are highlighted. However, they all have background tissues as well as noise associated with them. In this work, the MRVGs from the MICRO-MRI dataset are ultimately acquired through the MRVG_{avg} method by enhancing vein signals from different data resources as mentioned in Sec. 3.2. Our collaborative domain experts compute the vein labels by post-manual case-wise threshold adjustment on the MRVGs. However, we can see from the major-level vein case in Fig. 5 (b) that the vein labels still fail to be free from artery artifacts as marked in red / green dotted circles. In addition, the MRVGs overall have more challenging noise type (e.g., very strong artery artifacts) than single-modal MRAGs due to modality formulation; therefore, the corresponding intensity-based vein labels tend to have more fuzzy edges. However, our VC-Net is able to effectively overcome the aforementioned difficulty as shown in Fig. 5 (b). Tab. 3 shows the numerical comparison among ours, MRVG_{avg}, and 3D U-Net under the same experiment setting. We can see that our numerical results overall outperform the other two methods. The general lower numerical performance compared to artery segmentation (in the previous subsection) may result from more challenging input data type and the sinus region (i.e., the large and thick vein-like area at the bottom in 3D global segmentation). More qualitative comparisons with the MRVG_{avg} method are provided in Supplemental Material.

4.2.3 Micro-Level Vessel Segmentation and Visualization

As mentioned in the introduction, the micro-cerebrovasculature turns out to be a good physical indicator of many neurological disorders and vascular diseases; thus it is extremely important and a breakthrough for MICRO MRAV diagnosis of vascular disease to trace small vessels and analyze their topology, morphology, density, and distribution with direct visual inspection of microvascular abnormalities *in-vivo*. Besides the major-level vessel segmentation, our VC-Net shows great capability to track the micro vessels as well. In this experiment, it is noted that we have the input data modality format which is different from those in the previous experiments as shown in the first column in Fig. 6 (c), i.e., the minimum intensity projection (MinIP) images. The different modalities of input image slice examples in the experiments are provided in Supplemental Material. All vessels, including major and micro ones, appear to be dark (very low voxel intensity) and show no contrast to the dark background, which may cause confusion to our network in the 2D composited MIP segmentation stream even if we accordingly switch to compute the MinIP instead. In order to keep the framework consistency and take advantage of the pre-trained network in the previous subsections, we inverse the voxel intensity within the brain foreground area in the whole 3D SWI image and then extract 1320 random patches from each training image (considering the vessels are much denser in a large-sized 3D volume, more patches per case can make up for limited datasets, e.g., two training cases). By fine-tuning our VC-Net pre-trained on major-level MRI images with these patches, our network is capable of capturing the continuous micro vessels clearly

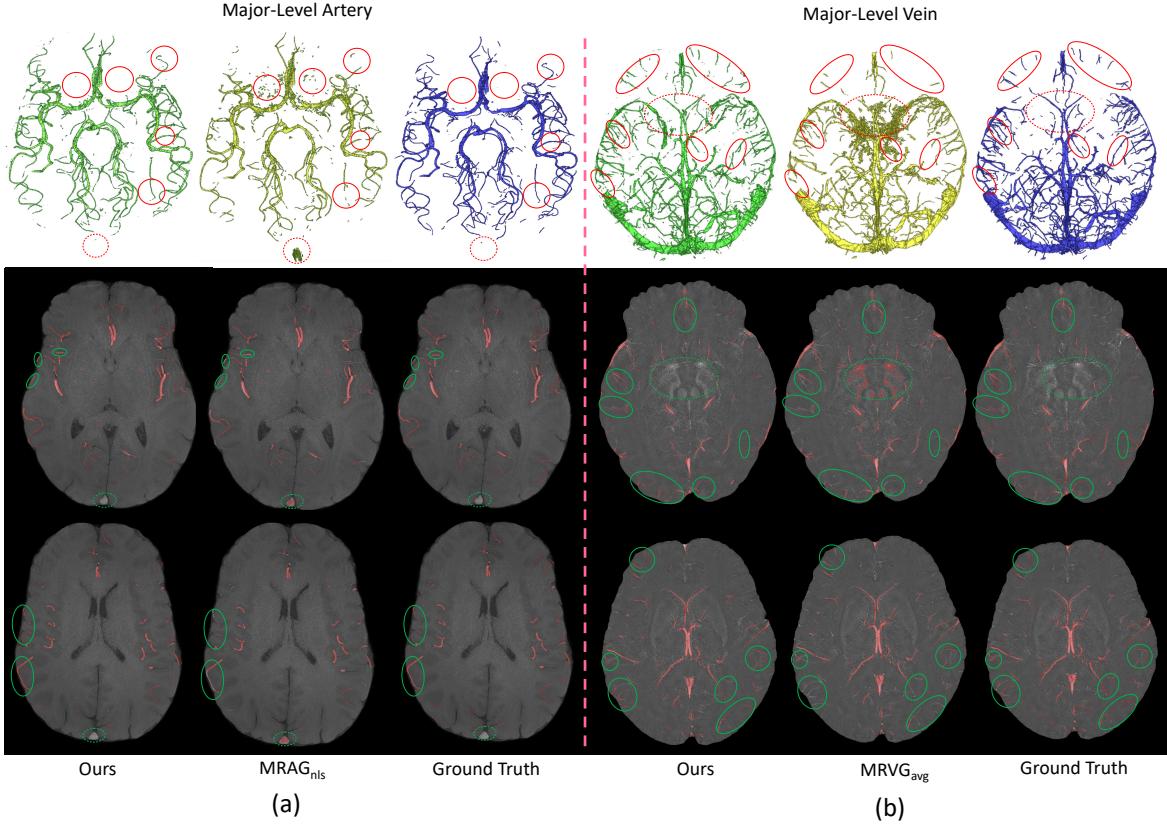


Fig. 5: Some qualitative comparison results from MICRO-MRI major-level vessel dataset: The 3D global vessel segmentations are shown from superior direction. The MIP segmentations are visualized by 5-sliced MICRO-MRI images, and the corresponding vessel masks in MIPs are marked in semi-transparent red. The highlighted comparison areas are marked in circles. The 3D MRAG / MRVG images from MICRO-MRI dataset only focus on midbrain area and thus have less vessels compared with TubeTK dataset.

as shown in Fig. 6.

Currently, SWI is the only data modality available to capture the micro-level vessels and it includes a large number of major-level vessels as well. Consequently, it is quite challenging to provide a rigorous numerical evaluation on pure micro-level vessels. Alternatively, Tab. 4 shows the numerical evaluation based on the whole SWI image for reference, in which our method quantitatively outperforms 3D U-Net (also fine-tuned from the weights pre-trained on the same major-level MRI images) and the SWI_{ATRG} method on Dice, precision, and FPR metrics. The SWI_{ATRG} method is a state-of-the-art algorithm that our collaborative domain experts are currently using as described in Sec. 3.2.

Fig. 6 (a) and (b) show our whole brain segmentation result (in gold) accompanied by non-overlapping midbrain subarea segmentation results (in red) and their corresponding ground truth (in blue). Fig. 6 (c) and (d) visualize the qualitative performance of pair-wise comparisons. From Fig. 6 (d), we can see that the result from the SWI_{ATRG} method suffers severe voxel intensity noise (circled in yellow) and unexpectedly thicker vessels (circled in blue) due to the bold intensity threshold in sacrifice to capture as many micro vessels as possible; however, it still lacks the satisfiable ability of detecting micro vessels as shown in the corresponding zoom-in green patch error maps (white: true positive, red: false positive, blue: false negative, black: true negative). In addition, the SWI_{ATRG} method requires several data modalities which are acquired from different time points as mentioned in Sec. 3.2; consequently, the corresponding computed vessel mask also has non-negligible registration errors (circled in white). However, even if 3D U-Net can alleviate most of the issues that the SWI_{ATRG} method is faced with, it is still insufficient to track the super micro vessel without the 2D MIP complementary information as shown in first two zooming-in patches (circled in yellow) and their corresponding error maps in Fig. 6 (c). Similar as the SWI_{ATRG} method, 3D U-Net also performs more boldly on covering major-level vessels (less blue on error maps) as

shown in the third zooming-in patch in Fig. 6 (c). However, 3D U-Net incurs more noises (more red on error maps) and thus it has a worse precision.

Table 4: Quantitative performance evaluation of different methods on micro-level vessel segmentation.

| Methods / Metrics | Dice (%) \uparrow | Precision (%) \uparrow | FPR (%) \downarrow |
|---------------------|---------------------|--------------------------|----------------------|
| Ours | 74.40 | 74.74 | 0.7052 |
| 3D U-Net | 74.08 | 72.70 | 0.7989 |
| SWI _{ATRG} | 70.23 | 61.84 | 1.413 |

4.2.4 Discussion on Labeling and Visualization Tool

It is interesting to note that in the TubeTK dataset even the ground truth vessel mask does not cover certain vessel continuity, which can be clearly traced on MIPs (such as some yellow circles in the ground truth MIPs in Fig. 4), since it is very difficult to label all the vessels in corresponding MRA slices in a single slice-by-slice manipulation without referring to MIP and 3D global visualization. However, based on the ground truth MIP labeling slices in the MICRO-MRI dataset from our experiments and collaborative evaluations, we can see that such issue is greatly alleviated, since our cerebrovascular labeling and visualization tool is applied to generate (refine) these ground truth vessel labels. In the future, we will further refine the ground truth vessel labels of TubeTK dataset by using our developed labeling tool (such as examples shown in Supplemental Video) under the domain experts' guidance for better public sharing and use.

Last but not least, to our knowledge, we are the first to investigate and apply our 3D brain vasculature segmentation to different vessel types and levels, especially the micro-level vessel segmentation; also, it is the first time that the whole brain vessels with different types / levels can be visualized *in-vivo*. Therefore, we have also designed a visualization tool for jointly showing different vasculature systems. Our

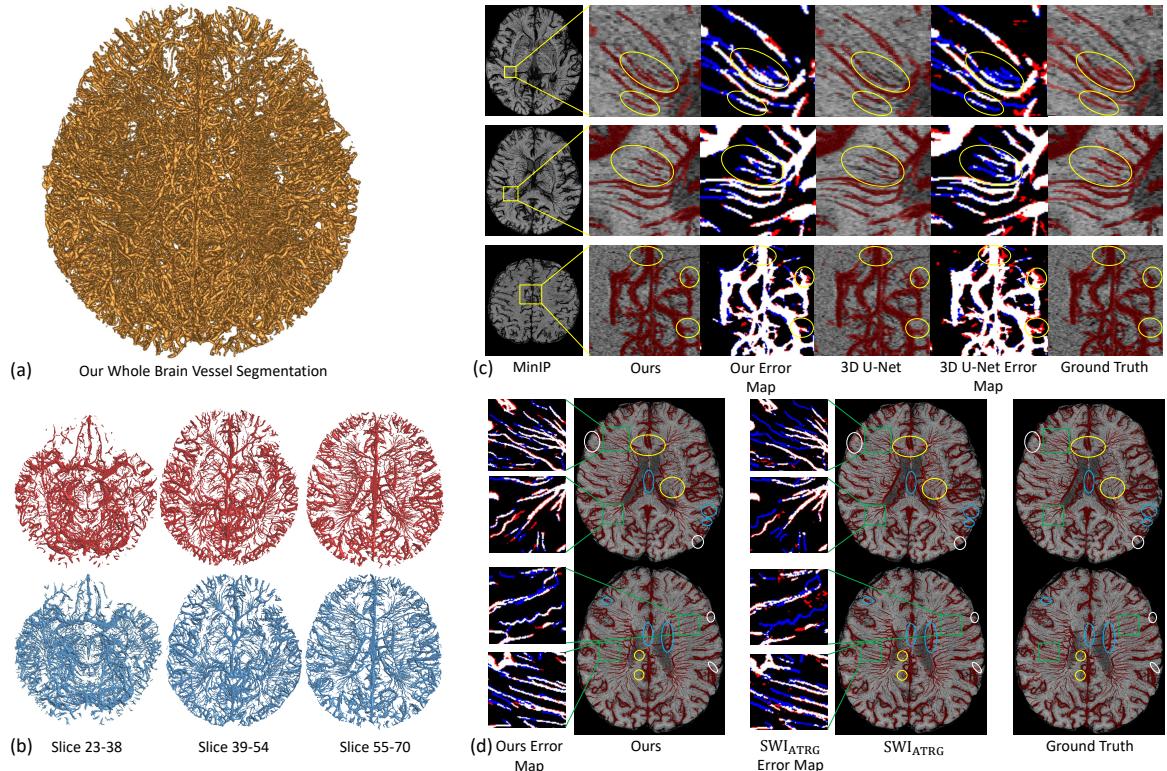


Fig. 6: (a) Whole brain micro vessel segmentation result of our method from superior direction. (b) Midbrain non-overlapping subarea detail visualization (in red) with ground truth in comparison (in blue). (c) Qualitative comparison results between our method and 3D U-Net on MICRO-MRI dataset, shown as three patch details extracted from three different 5-sliced MinIPs and their corresponding error maps on micro-level and major-level vessels. (d) Qualitative comparison results between our method and the SWI_{ATRG} method on MICRO-MRI dataset, shown as two 5-sliced MinIP segmentations and their corresponding error maps on micro-level vessels. The highlighted comparison areas are marked in circles.

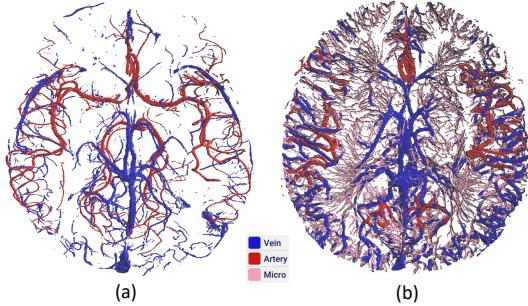


Fig. 7: Joint 3D visualization of our segmentation results on MICRO-MRI dataset in two different testing cases: (a) Whole midbrain major-level arteries (in red) and veins (in blue). (b) Major-level arteries (in red), major-level veins (in blue), and micro-level vessels (in pink) from slice No. 20 to 40 within midbrain area. Some large pink vessels are also major-level ones which are absent from major-level MRAGs and MRVGs due to the different image acquisitions.

tool enables the visualization for any combination of different vessel systems in user-defined color and lighting, and support all essential auxiliary interactions such as rotation, translation, scaling, zooming in / out, clipping, etc., for better examination. Fig. 7 (a) shows the joint visualization for our prediction results of major-level midbrain arteries and veins; and Fig. 7 (b) shows all three vasculature systems aligned together, i.e., major-level midbrain arteries and veins, and micro-level vessels, from MICRO-MRI dataset. Supplemental Video is included for demonstrating the dynamic visualization and interaction in detail.

5 CONCLUSION

In this work, we have proposed the VC-Net, a deep neural network to extract and visualize high-fidelity 3D cerebrovascular structure from

highly sparse and noisy images. VC-Net has three major components, i.e., 3D and 2D dual-domain segmentation streams, 3D-to-2D projection for two-stream design, and 2D-to-3D unprojection for joint embedding operations. By unprojecting the learned multislice composited 2D MIP feature vectors into the 3D volume embedding space, the proposed framework can strengthen the sparse 3D vascular representation by better capturing the small / micro vessels as well as improving the vessel connectivity, which outperforms the state-of-the-art classical and deep learning based methods. In medical practice, this work can be used as the key functions for real-time *in-vivo* segmentation and visualization of sparse and complicated 3D microvascular structure to improve MICRO MRAV diagnosis of vascular disease.

In the future, we will continue to explore research problems related to volume rendering supported 3D exploration and analysis to leverage both the 2D findings and the 3D knowledge and analytics by deep neural networks. We will extend current MIP-based volume rendering (i.e., a special case of volume rendering) into more general volume rendering scenarios, such as X-ray projections, full RGB composition, multi-view MIPs, and flow modeling concepts.

ACKNOWLEDGMENTS

We would like to thank the reviewers for their valuable comments. We are grateful to Yongsheng Chen from Neurology for the early discussion of this work, Pavan K. Jella from Radiology for preparing and collecting the clinical datasets, and Michelle Hua from Cranbrook Schools for pre-processing the datasets and proofreading the paper. This work was partially supported by the NSF under Grant Numbers IIS-1816511, CNS-1647200, OAC-1657364, OAC-1845962, OAC-1910469, the Wayne State University Subaward 4207299A of CNS-1821962, NIH 1R56AG060822-01A1, NIH 1R44HL145826-01A1, ZJNSF LZ16F020002, and NSFC 61972353.

REFERENCES

[1] TubeTK MRA dataset. <https://public.kitware.com/Wiki/TubeTK/Data>.

[2] H. Abdul-Rahman, M. Gdeisat, D. Burton, M. Lalor, F. Lilley, and C. Moore. Fast and robust three-dimensional best path phase unwrapping algorithm. *Applied Optics*, 46(26):6623–6635, 2007.

[3] M. Bronstein, J. Bruna, Y. LeCun, A. Szlam, and P. Vandergheynst. Geometric deep learning: going beyond Euclidean data. *IEEE Signal Processing Magazine*, 34(4):18–42, 2017.

[4] S. Buch, Y. Wang, M.-G. Park, P. Jella, J. Hu, Y. Chen, K. Shah, Y. Ge, and E. Haacke. Subvoxel vascular imaging of the midbrain using USPIO-Enhanced MRI. *NeuroImage*, p. 117106, 2020.

[5] Y. Chen, S. Liu, S. Buch, J. Hu, Y. Kang, and E. Haacke. An interleaved sequence for simultaneous magnetic resonance angiography (MRA), susceptibility weighted imaging (SWI) and quantitative susceptibility mapping (QSM). *Magnetic Resonance Imaging*, 47:1–6, 2018.

[6] A. Chung and J. Noble. Statistical 3D vessel segmentation using a Rician distribution. In *Proceedings of International Conference on Medical Image Computing and Computer Assisted Intervention*, pp. 82–89, 1999.

[7] Ö. Çiçek, A. Abdulkadir, S. Lienkamp, T. Brox, and O. Ronneberger. 3D U-Net: learning dense volumetric segmentation from sparse annotation. In *Proceedings of International Conference on Medical Image Computing and Computer Assisted Intervention*, pp. 424–432, 2016.

[8] M. Descoteaux, D. Collins, and K. Siddiqi. A geometric flow for segmenting vasculature in proton-density weighted MRI. *Medical Image Analysis*, 12(4):497–513, 2008.

[9] H. Fan, H. Su, and L. Guibas. A point set generation network for 3D object reconstruction from a single image. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 605–613, 2017.

[10] C. Florin, N. Paragios, and J. Williams. Globally optimal active contours, sequential Monte Carlo and on-line learning for vessel segmentation. In *Proceedings of European Conference on Computer Vision*, pp. 476–489, 2006.

[11] N. Forkert, A. Schmidt-Richberg, J. Fiehler, T. Illies, D. Möller, D. Säring, H. Handels, and J. Ehrhardt. 3D cerebrovascular segmentation combining fuzzy vessel enhancement and level-sets with anisotropic energy weights. *Magnetic Resonance Imaging*, 31(2):262–271, 2013.

[12] A. Frangi, W. Niessen, K. Vincken, and M. Viergever. Multiscale vessel enhancement filtering. In *Proceedings of International Conference on Medical Image Computing and Computer Assisted Intervention*, pp. 130–137, 1998.

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

[14] P. Goyadinov, T. Womack, J. Eriksen, G. Chen, and D. Mayerich. Robust tracing and visualization of heterogeneous microvascular networks. *IEEE Transactions on Visualization and Computer Graphics*, 25(4):1760–1773, 2018.

[15] P. Goyadinov, T. Womack, J. Eriksen, D. Mayerich, and G. Chen. Graph-assisted visualization of microvascular networks. In *Proceedings of IEEE Visualization Conference*, pp. 1–5, 2019.

[16] E. Haacke, J. Tang, J. Neelavalli, and Y. Cheng. Susceptibility mapping as a means to visualize veins and quantify oxygen saturation. *Journal of Magnetic Resonance Imaging*, 32(3):663–676, 2010.

[17] E. Haacke, Y. Xu, Y.-C. Cheng, and J. Reichenbach. Susceptibility weighted imaging (SWI). *Magnetic Resonance in Medicine*, 52(3):612–618, 2004.

[18] J. Hua, J. Hu, and Z. Zhong. *Spectral Geometry of Shapes: Principles and Applications*. Springer, 2019.

[19] M. Jenkinson, M. Pechaud, and S. Smith. BET2: MR-based estimation of brain, skull and scalp surfaces. 2005.

[20] J. Jiang, M. Dong, and E. Haacke. ARGDYP: an adaptive region growing and dynamic programming algorithm for stenosis detection in MRI. In *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 2, pp. ii–465, 2005.

[21] H. Jin, Y. Lian, and J. Hua. Learning facial expressions with 3D mesh convolutional neural network. *ACM Transactions on Intelligent Systems and Technology*, 10(1):1–22, 2018.

[22] T. Kitrunrotsakul, X.-H. Han, Y. Iwamoto, L. Lin, A. Foruzan, W. Xiong, and Y.-W. Chen. VesselNet: A deep convolutional neural network with multi pathways for robust hepatic vessel segmentation. *Computerized Medical Imaging and Graphics*, 75:74–83, 2019.

[23] A. Komarichev, Z. Zhong, and J. Hua. A-CNN: Annularly convolutional neural networks on point clouds. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 7421–7430, 2019.

[24] Q. Li, B. Feng, L. Xie, P. Liang, H. Zhang, and T. Wang. A cross-modality learning approach for vessel segmentation in retinal images. *IEEE Transactions on Medical Imaging*, 35(1):109–118, 2015.

[25] W. Liao, K. Rohr, and S. Wörz. Globally optimal curvature-regularized fast marching for vessel segmentation. In *Proceedings of International Conference on Medical Image Computing and Computer Assisted Intervention*, pp. 550–557, 2013.

[26] P. Liskowski and K. Krawiec. Segmenting retinal blood vessels with deep neural networks. *IEEE Transactions on Medical Imaging*, 35(11):2369–2380, 2016.

[27] M. Martínez-Pérez, A. Hughes, A. Stanton, S. Thom, A. Bharath, and K. Parker. Retinal blood vessel segmentation by means of scale-space analysis and region growing. In *Proceedings of International Conference on Medical Image Computing and Computer Assisted Intervention*, pp. 90–97, 1999.

[28] J. Masci, D. Boscaini, M. Bronstein, and P. Vandergheynst. Geodesic convolutional neural networks on Riemannian manifolds. In *Proceedings of the IEEE International Conference on Computer Vision Workshops*, pp. 37–45, 2015.

[29] J. Mo and L. Zhang. Multi-level deep supervised networks for retinal vessel segmentation. *International Journal of Computer Assisted Radiology and Surgery*, 12(12):2181–2193, 2017.

[30] D. Nain, A. Yezzi, and G. Turk. Vessel segmentation using a shape driven flow. In *Proceedings of International Conference on Medical Image Computing and Computer Assisted Intervention*, pp. 51–59, 2004.

[31] S. Napel, M. Marks, G. Rubin, M. Dake, C. McDonnell, S. Song, D. Enzmann, and R. Jeffrey Jr. CT angiography with spiral CT and maximum intensity projection. *Radiology*, 185(2):607–610, 1992.

[32] C. Qi, H. Su, K. Mo, and L. Guibas. PointNet: Deep learning on point sets for 3D classification and segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 652–660, 2017.

[33] O. Ronneberger, P. Fischer, and T. Brox. U-Net: Convolutional networks for biomedical image segmentation. In *Proceedings of International Conference on Medical Image Computing and Computer Assisted Intervention*, pp. 234–241, 2015.

[34] P. Sánchez, C. Meyer, V. Vigon, and B. Naegel. Cerebrovascular network segmentation of MRA images with deep learning. In *Proceedings of IEEE International Symposium on Biomedical Imaging*, pp. 768–771, 2019.

[35] F. Schweser, A. Deistung, B. Lehr, and J. Reichenbach. Quantitative imaging of intrinsic magnetic tissue properties using MRI signal phase: an approach to *in vivo* brain iron metabolism? *Neuroimage*, 54(4):2789–2807, 2011.

[36] Y. Shen, J. Hu, K. Eteer, Y. Chen, S. Buch, H. Alhourani, K. Shah, Q. Jiang, Y. Ge, and E. Haacke. Detecting sub-voxel microvasculature with USPIO-enhanced susceptibility-weighted MRI at 7 T. *Magnetic Resonance Imaging*, 67:90–100, 2020.

[37] S. Shin, S. Lee, I. Yun, and K. Lee. Deep vessel segmentation by learning graphical connectivity. *Medical Image Analysis*, 58:101556, 2019.

[38] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. Alemi. Inception-v4, inception-resnet and the impact of residual connections on learning. In *Proceedings of The AAAI Conference on Artificial Intelligence*, 2017.

[39] J. Tang, S. Liu, J. Neelavalli, Y. Cheng, S. Buch, and E. Haacke. Improving susceptibility mapping using a threshold-based k-space/image domain iterative reconstruction approach. *Magnetic Resonance in Medicine*, 69(5):1396–1407, 2013.

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

[41] H. Wang, Q. Jiang, Y. Shen, L. Zhang, E. Haacke, Y. Ge, S. Qi, and J. Hu. The capability of detecting small vessels beyond the conventional MRI sensitivity using iron-based contrast agent enhanced susceptibility weighted imaging. *NMR in Biomedicine*, 2020.

[42] N. Wang, Y. Zhang, Z. Li, Y. Fu, W. Liu, and Y.-G. Jiang. Pixel2Mesh: Generating 3D mesh models from single RGB images. In *Proceedings of the European Conference on Computer Vision*, pp. 52–67, 2018.

[43] S. Wang, B. Peplinski, L. Lu, W. Zhang, J. Liu, Z. Wei, and R. Summers. Sequential Monte Carlo tracking for marginal artery segmentation on CT

angiography by multiple cue fusion. In *Proceedings of International Conference on Medical Image Computing and Computer Assisted Intervention*, pp. 518–525, 2013.

[44] Y. Wang, Z. Zhong, and J. Hua. DeepOrganNet: On-the-fly reconstruction and visualization of 3D/4D lung models from single-view projections by deep deformation network. *IEEE Transactions on Visualization and Computer Graphics*, 26(1):960–970, 2019.

[45] D. Wilson and J. Noble. Segmentation of cerebral vessels and aneurysms from MR angiography data. In *Proceedings of Biennial International Conference on Information Processing in Medical Imaging*, pp. 423–428, 1997.

[46] H. Xu, M. Dong, and Z. Zhong. Directionally convolutional networks for 3D shape segmentation. In *Proceedings of the IEEE International Conference on Computer Vision*, pp. 2698–2707, 2017.

[47] Y. Ye, J. Hu, D. Wu, and E. Haacke. Noncontrast-enhanced magnetic resonance angiography and venography imaging with enhanced angiography. *Journal of Magnetic Resonance Imaging*, 38(6):1539–1548, 2013.

SUPPLEMENTAL MATERIAL

S1 SUPPLEMENTAL FIGURES

Supplemental figures are included for demonstrating additional qualitative results from TubeTK and MICRO-MRI datasets in Fig. S1, and different modalities of input image examples from TubeTK MRA and MICRO-MRI datasets in our VC-Net method in Fig. S2.

S2 ADDITIONAL QUANTITATIVE PERFORMANCE EVALUATION

In order to further demonstrate the effectiveness of our VC-Net (especially 3D-to-2D projection in dual-stream and 2D-to-3D unprojection for joint embedding in our proposed architecture), Tab. S1 shows the numerical analyses on some simple combinations of the final results from 3D U-Net and 2D U-Net through average and max operations on the probabilities.

Table S1: Quantitative performance evaluation between different combinations of 3D U-Net and 2D U-Net and our method on TubeTK dataset.

| Methods / Metrics | Dice (%) \uparrow |
|-------------------|---------------------|
| 2D U-Net | 65.10 |
| 3D U-Net | 71.01 |
| Average Fusion | 65.15 |
| Max Fusion | 69.41 |
| Ours | 71.81 |

From Tab. S1, we can see our VC-Net overall outperforms both combination methods of the final results of 3D U-Net and 2D U-Net. As shown in Sec. 4.1 of the paper, 2D U-Net performs much worse than a standalone 3D U-Net on each metric. Unlike the 2D composited MIP stream in VC-Net, 2D U-Net itself essentially does not involve any complementary or enhancement information, and the reception field of 2D U-Net is restricted to an isolated 2D slice patch every time and thus lack of the contextual information from the third dimension, which is fatal to the sparse 3D vessel segmentation. Without comprehensive 3D spacing neighborhood, 2D U-Net is more prone to strong noise perturbation (high-intensity true negative) and insensitive to weak vessel signal (low-intensity true positive), as a result, 2D U-Net performs unsatisfactorily even when equipped with more feature embedding channels. Consequently, it may not be an ideal idea to fuse the results from 3D U-Net and 2D U-Net through the simple combinations.

Here, Dice Similarity is provided since it is the most comprehensive and effective indicator / metric to justify the segmentation performance. It measures the intersection over union between the prediction and the ground truth, which comprehensively takes into account all true positive (TP), false negative (FN), as well as false positive (FP). This is also why we (as well as many other research works) select Dice Similarity as the loss function in our VC-Net.

S3 LABELING REFINEMENT AND VISUALIZATION TOOL

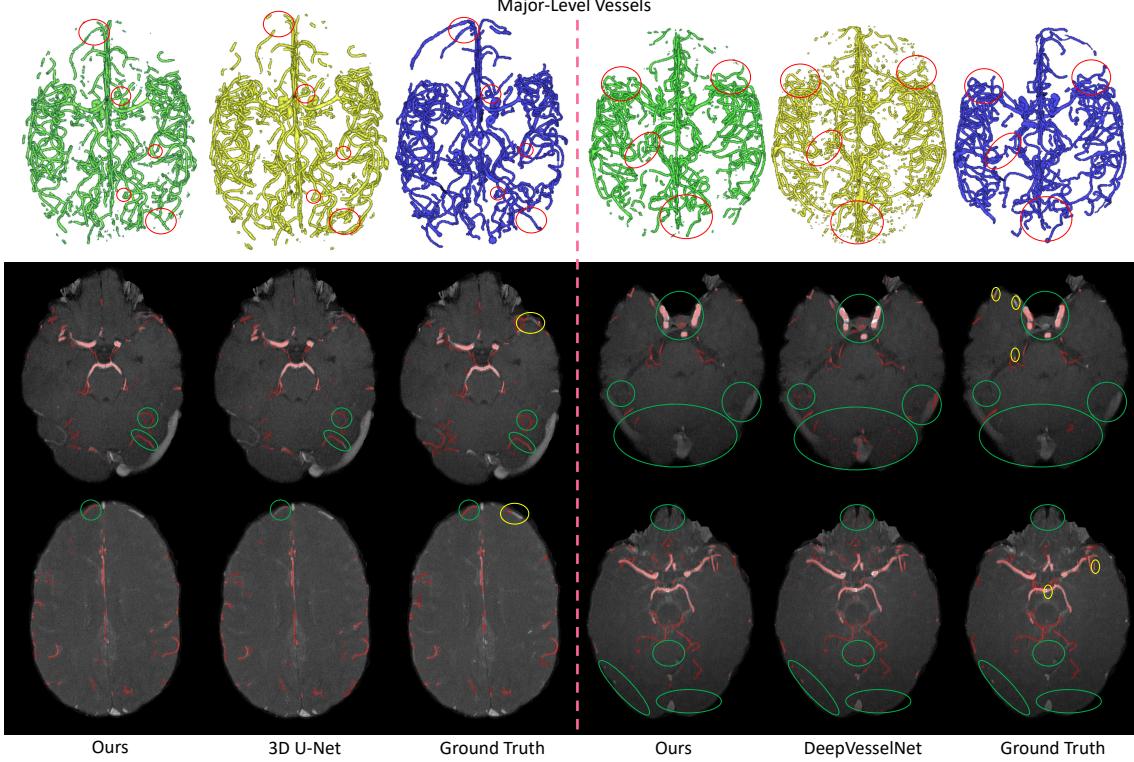
The interface and basic functions of our specifically-designed cerebrovascular labeling and visualization tool are shown in Fig. S3. Our tool enables slice-wise refinement based on the pre-computed vessel labels by MRAG_{hls} , MRVG_{avg} , and SWI_{ATRG} methods, instead of labeling from scratch manually. The interactive vessel editing is conducted in the current image slice window, e.g., manually labeling / erasing brush, automatically labeling connected components by flood-fill method as shown in Fig. S3 (a). The slice under editing is simultaneously visualized in solid red for a clearer examination in Fig. S3 (d). Unlike the operation in most of the general-purpose labeling / segmentation softwares in which the current labeling (2D) slice is usually isolated from its (3D) context and thus lacks the crucial reference, the vessel labeling in our developed tool is comprehensively assisted and guided by the following specifically-desired functions: (1) simultaneously updated 3D vasculature system from the beginning to the current slices with several interactions, such as rotation and zooming in / out, to check the cross-plane 3D vessel connectivity (Fig. S3 b); (2) synchronized brain vessel volume rendering to trace the overall segmented vasculature system (Fig. S3 c); (3) adaptive MIP labeling display (with user-defined

number of projection slices) that enables users to evaluate the contextual slices to strengthen the vessel connectivity and rule out noise (Fig. S3 e). Our tool can greatly facilitate the continuous slice-wise labeling and reduce the labeling ambiguity in some challenging areas of the micro-cerebrovascular structure, which have been extensively tested and evaluated by our collaborative domain experts. Supplemental Video is included to demonstrate the dynamic visualization and interaction in detail.

S4 SUPPLEMENTAL VIDEO

Supplemental video is included to demonstrate the joint 3D visualization of the major-level and micro-level vessels in the midbrain and the whole brain on MICRO-MRI dataset; as well as the dynamic visualization and interaction of our developed cerebrovascular labeling and visualization tool.

Additional Results on TubeTK Dataset



Additional Results on MICRO-MRI Dataset

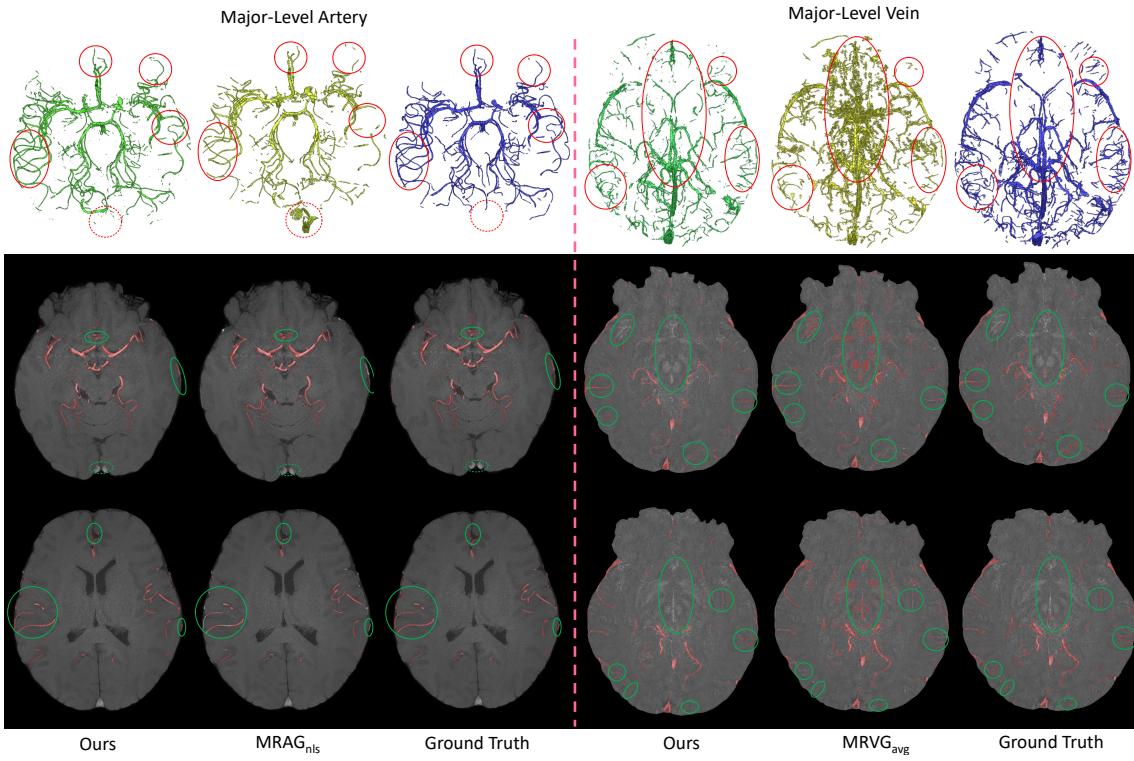


Fig. S1: Additional qualitative results from two datasets (top: TubeTK dataset, bottom: MICRO-MRI major-level vessel dataset): The 3D global vessel segmentations are shown from superior direction. The MIP segmentations are visualized by 5-sliced MRA / MICRO-MRI images, and the corresponding vessel masks in MIPs are marked in semi-transparent red. The highlighted comparison areas are marked in circles. The 3D MRAG / MRVG images from MICRO-MRI dataset only focus on midbrain area and thus have less vessels compared with TubeTK dataset. It is noted that in TubeTK dataset even the ground truth vessel label does not perfectly cover certain vessel continuity, which can be clearly traced on MIPs (such as some yellow circles in the ground truth MIPs), in the corresponding MRA slices. We will further refine the ground truth vessel labels of TubeTK dataset by using our developed labeling tool under the domain experts' guidance in our future work.

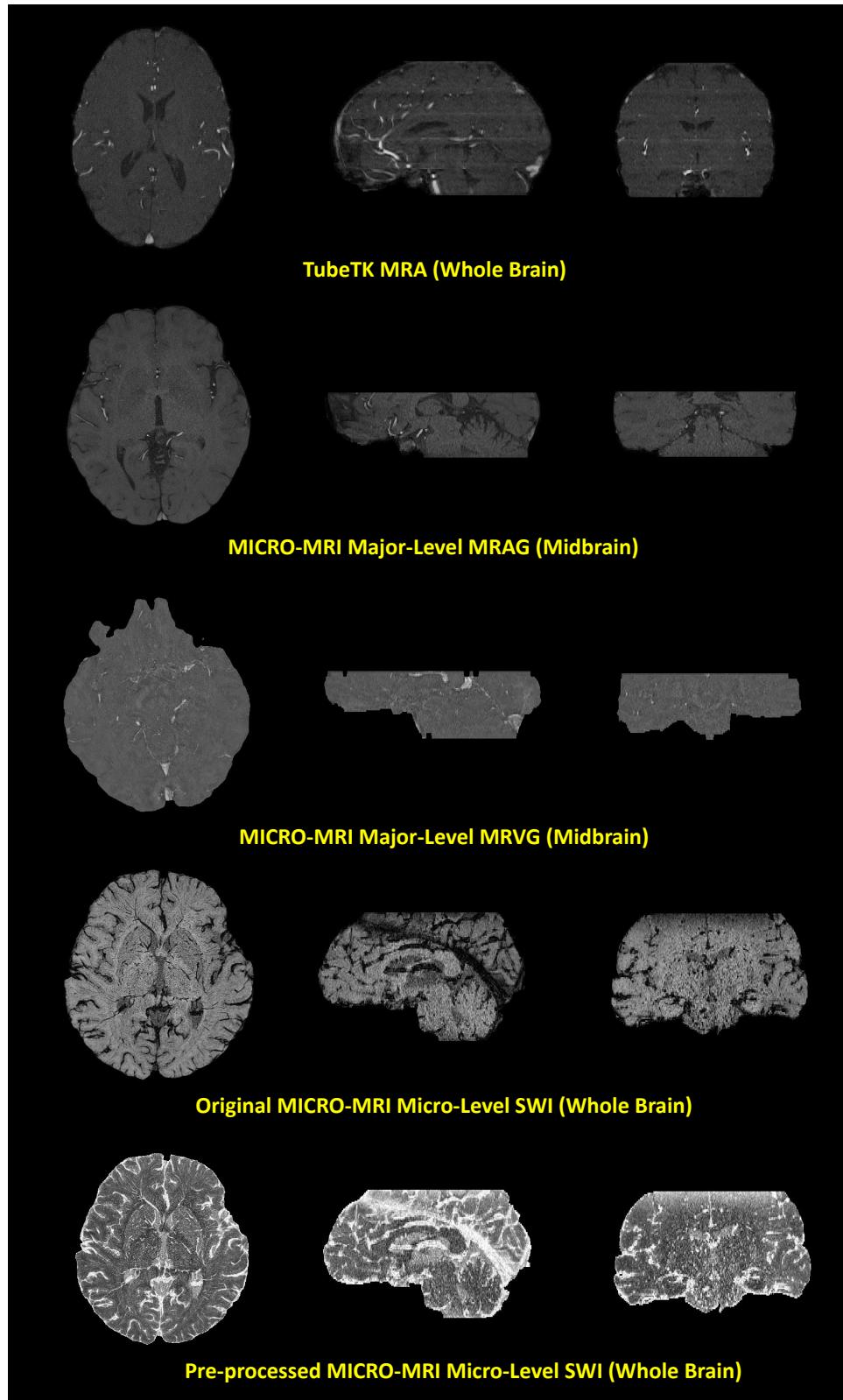


Fig. S2: The different modalities of input image examples from TubeTK MRA and MICRO-MRI datasets in our VC-Net method.

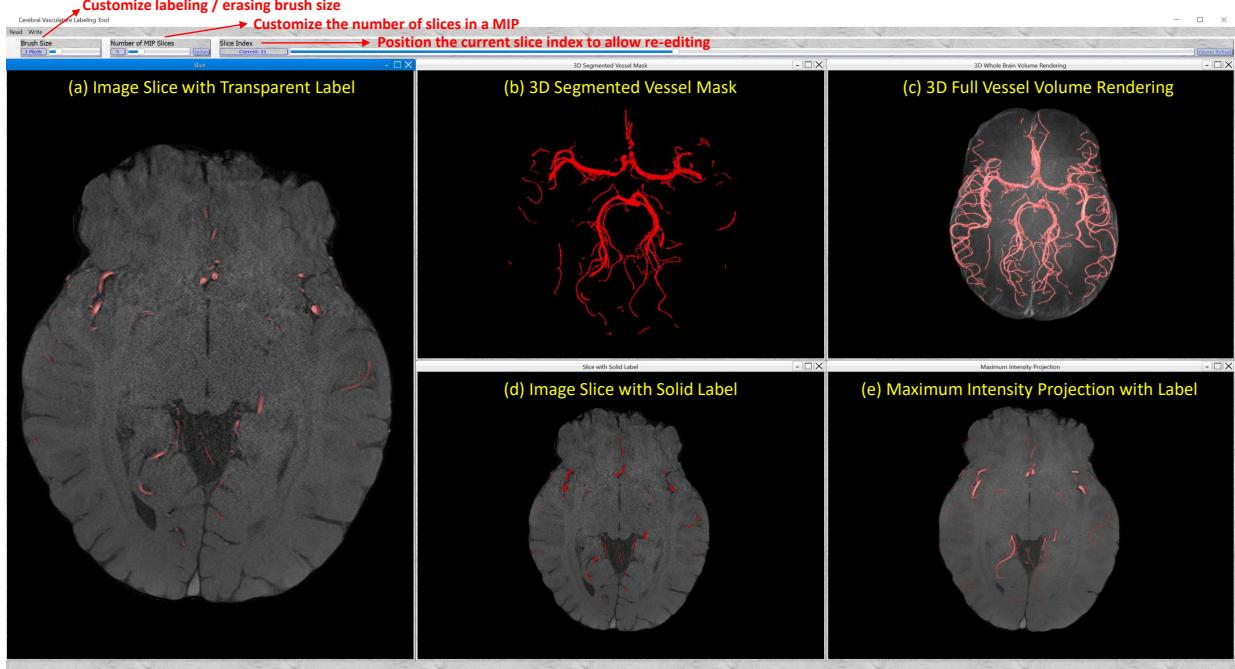


Fig. S3: Our developed cerebrovascular labeling and visualization tool (e.g., an example of major-level arterial vessels from MICRO-MRI).