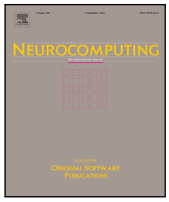
Neurocomputing 599 (2024) 128038

Contents lists available at ScienceDirect

Neurocomputing

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



ROSE: Multi-level super-resolution-oriented semantic embedding for 3D microvasculature segmentation from low-resolution images

Yifan Wang [a], Haikuan Zhu [a], Hongbo Li [a], Guoli Yan [a], Sagar Buch [b], Ying Wang [b], Ewart Mark Haacke [b], Jing Hua [a], Zichun Zhong [a,*]

[a] Department of Computer Science, Wayne State University, Detroit, 48202, MI, USA
[b] Department of Radiology, Wayne State University, Detroit, 48201, MI, USA

ARTICLE INFO

Communicated by G. Yang

Keywords:
3D microvasculature segmentation
Deep neural network
Joint multi-level embedding
Super-resolution semantics

ABSTRACT

Current state-of-the-art segmentation methods often require high-resolution input to attain the high performance, which pushes the limit of data acquisition and brings large computation budgets. Instead, we present an end-to-end deep learning-based method, ROSE, for robust and precise segmentation of high-quality 3D super-resolution (SR) microvasculatures from low-resolution (LR) images as input, which can transform data from the LR imaging domain to the SR semantic domain (cross different modalities and scales). More specifically, a multi-tasking two-stream deep learning framework is proposed to learn the high-fidelity microvasculature SR image and semantic hybrid features simultaneously. During the proposed joint learning process, the high-resolution features of microvasculatures are further enhanced by the learned fine-grained structural/textural features from the microvasculature SR stream with a multi-level embedding scheme through the oriented feature aggregation at different fusion stages. In the constructed joint multi-level hybrid embedding spaces, the instance semantic embedding and the SR imaging embedding can be connected and integrated synergistically. We have conducted extensive experiments using public and real patient micro-cerebrovascular image datasets and compare our framework with traditional 3D vessel segmentation methods and the other state-of-the-art in deep learning. This robust and precise microvascular visualization in different brain regions by our method demonstrates its potential impact in magnetic resonance (MR) angiography and venography for the diagnosis of microvascular disease.

1. Introduction

Nowadays, effectively acquiring, visualizing, processing, and analyzing complicated microstructures in the raw datasets becomes increasingly important in scientific and engineering discoveries. Current state-of-the-art semantic exploration methods often take use of high-resolution input to attain and enhance the high performance, which brings large computation budgets. The high computational cost makes them impractical to the microstructured object extraction, such as the *in-vivo* micro-level vessel extraction and visualization during the scanning. Capturing vascular abnormalities *in-vivo* at the micro-level [1] has attracted attention from more and more researchers since many neurological disorders and vascular diseases in animal and human studies are characterized by significant small blood vessel involvement, such as hypertension, arteriosclerosis, cerebral amyloid angiopathy, diabetes, ischemia, stroke [2–4]. There is an emerging trend and urgent demand for super-high resolution magnetic resonance imaging (MRI) images that are capable of pinpointing vascular abnormalities. However, the high-resolution (HR) MRI requires longer scanning time, lower signal-to-noise ratio (SNR), smaller spatial coverage, more expensive acquisition devices, and more labor-intensive processing work, which become the notorious bottlenecks (challenges) in many medical and research explorations/applications. The motivation of this work is to extract the high-fidelity microstructured objects from low-resolution image to speed up the data acquisition and processing.

In recent decades, data-driven approaches have been proposed to statistically investigate the resolutional correlations between different instances without relying on hard-coded metrics. Deep learning-based super-resolution (SR) approaches have been proposed in visualization, computer graphics, and computer vision fields, such as SRCNN [5], FSRCNN [6], ESPCN [7], SRGAN [8], ESRGAN [9], EnhancedNet [10], SPSR [11], etc. However, these methods only handle 2D images and do not work for SR-based semantics of images. In scientific visualization field, SR techniques have been applied to enhance the volumetric visualization quality, such as volume upscaling [12], tempoGAN [13], SSR-TVD [14], TSR-TVD [15], STNet [16], SSR-VFD [17], TSR-VFD [18],

* Corresponding author.
E-mail address: zichunzhong@wayne.edu (Z. Zhong).

https://doi.org/10.1016/j.neucom.2024.128038
Received 15 September 2023; Received in revised form 19 April 2024; Accepted 7 June 2024
Available online 11 June 2024

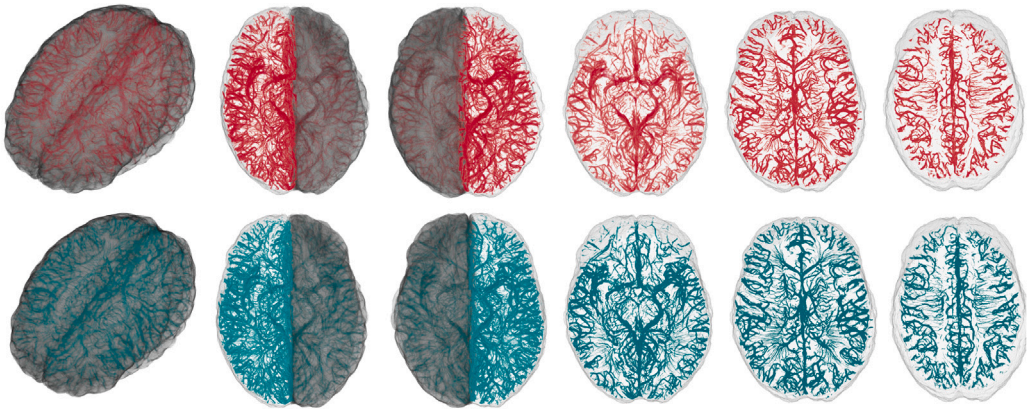


Fig. 1. Examples of 3D SR micro-cerebrovascular segmentation and visualization of whole region and subareas at the center of a human brain from an LR image input: volume rendering of our ROSE result (1st row) and the ground truth (2nd row).

DHSR [19], etc. All of the above methods do not consider the semantics generation in SR microstructure from low-resolution (LR) images/volumes. In medical image field, SR techniques have also been applied, such as CSN [20], 3D-ESPCN [21], DBAN [22], DC-SRN [23], mDCSRN-GAN [24], P-GANs [25], Multimodal-Boost [26], SR-Dict [27], etc. Although the above methods can handle 3D SR image generation, none of them explores the SR image generation from the LR grayscale images and its correlational interaction with a particular semantic analysis. For semantic extraction, there are few SR-based image segmentation methods, such as DSRL [28], SegSRGAN [29], and PFSeg [30]; and moreover, there is no method investigating how to fully and systematically leverage the joint SR-based embedding learning to improve the small-/micro-structured 3D object capture from LR images.

Recently, for the microstructure extraction applications, several deep learning based methods have been proposed to extract vessels from 2D retinal images, such as DeepVessel [31], multi-level deep supervised networks [32], deep neural network (DNN)-based method [33], unified convolutional neural network (CNN) and graph neural network (GNN) [34], etc. These methods can perform 2D vessel extraction tasks well, but are far from satisfactory for the 3D vessel scenario. There are some deep learning approaches dedicated to 3D vessel extraction, such as Uception [35], DeepVesselNet [36], VesselNet [37], VC-Net [38], JointVesselNet [39], Topo-Vascular [40], ML-Residual [41], Vascular-MIP [42], etc. The above existing methods do not consider SR techniques in the vessel extraction and are not designed for solving the challenges in 3D microvascular extraction.

The fundamental rationale of the proposed work is to present a new *SR-oriented semantic joint embedding* guided computing paradigm to enhance the intrinsic volumetric features for high-fidelity 3D microvasculature exploration from low-resolution images. In order to fill the gap in the robust and precise 3D micro-cerebrovascular extraction and visualization from the *in-vivo* MR data, in this paper, we present a deep learning-based method, *ROSE*, that can infer extra-fine vessel detail from subtle clues on low-resolution images using lightweight super-resolution (SR) technology and extract the micro-level vessel structure at the higher resolution simultaneously. The core *novelty* is to automatically leverage the SR imaging technique to enhance the 3D data exploration, especially for fine-grained 3D *in-vivo* microvasculature segmentation and visualization, at the deep neural network learning level. The key idea of our network is to integrate the trustworthy auxiliary information from the learned compound 2D SR-oriented features into the 3D volume segmentation and visualization network, instead of using more complicated and heavyweight networks empirically. Experimental results are evaluated and compared with traditional 3D vessel segmentation and state-of-the-art deep learning methods, using extensive public and real patient micro-cerebrovascular image datasets. Our key *contributions* are:

- To our knowledge, this is the first time that a deep learning framework is proposed to construct the joint multi-level hybrid (semantics and imaging) embedding spaces, where the computed joint micro vessel probabilities for the 3D volume processing (segmentation) can be synergistically augmented and enhanced from the compound 2D image SR acquisition.
- A multi-tasking two-stream convolutional neural network (CNN) framework is designed to semantically integratively and SR-orientedly learn the feature vectors of 3D volume and compound 2D SR image, and explore their inter-dependencies.
- It proposes an effective end-to-end deep learning method to segment and visualize high-fidelity 3D microvasculature with complicated geometry and tiny sizes from the raw LR volumetric images.
- The experiments on the accurate segmentation and visualization of complicated and tiny 3D microvascular structure in different brain regions by our method (such as Fig. 1) demonstrate the potential in a novel approach using MR angiography and venography in the diagnosis of microvascular disease.

2. Related work

In this section, we review related work on deep learning-based image SR and visualization, SR image segmentation, and vessel segmentation in computer vision and medical imaging.

Deep Learning for Image Super-Resolution. The main idea of an image super-resolution (SR) task is to increase the resolution of a 2D image from low to high. In computer graphics and computer vision fields, the deep learning-based SR has become the mainstream in image SR tasks during recent years. For instance, SRCNN [5] first proposes to use a three-layer CNN to solve the single 2D natural image in a pre-upsampling style, in which the low-resolution image is interpolated to a high spatial size for patch/feature extraction at the beginning so the consumption of the computational power is high accordingly. FSRCNN [6] improves SRCNN by removing the pre-upsampling step and uses deconvolution layer instead to realize post-upsampling, which also saves model parameters by reducing filter numbers and sizes. ESPCN [7] uses sub-pixel to replace the deconvolution layer for upsampling and thus solves the corresponding checkerboard issue in super-resolution results. [43–46] all apply deep residual blocks in their network for deep feature extraction. [47–49] employ recurrent neural networks to share network parameters in convolution layers and thus reduce the network footprints. Instead of aforementioned methods, which perform optimization in terms of pixel difference, GAN-based SR methods introduce discriminator networks and corresponding adversarial losses to yield much better perceptual quality in SR results. SRGAN [8] introduces a GAN-based network together with a VGG-feature based perceptual loss to greatly enhance the perceptual quality

of the restored SR images and GAN-based network plays the major role in SR problem since then. For instance, ESRGAN [9] improves SRGAN by involving a relativistic discriminator. EnhancedNet [10] employs an extra term in the loss function to capture finer texture information. SPSR [11] improves the image structure detail by involving additional image gradient supervision. In addition to these fields, SR has been widely used in many other areas, such as facial and satellite images or videos. There are some recent work [50–52] focused on inferring HR face images from the given LR ones. Deeba et al. [27] proposed a transferred wide residual single image SR remote sensing deep neural network model. Xiao et al. [53] generated satellite video SR by multiscale deformable convolution alignment and temporal grouping projection. Yi et al. [54] proposed an omniscient framework for video SR. Dharejo et al. [55] proposed a frequency domain-based spatio-temporal remote sensing single image SR technique to reconstruct the HR image combined with GANs on various frequency bands. However, none of the above methods work on 3D volume images and SR-based semantics of images.

In 3D scientific visualization, SR techniques have been applied to enhance the volumetric visualization quality. Zhou et al. [12] proposed a CNN-based volume upscaling method. Weiss et al. [56] developed a deep learning-based upscaling of a low-resolution sampling of an isosurface to a higher resolution with spatial detail and shading. tempoGAN [13] proposes a temporally coherent generative model addressing the superresolution problem for fluid flows. SSR-TVD [14] produces coherent spatial super-resolution (SSR) of time-varying data (TVD) using adversarial learning. TSR-TVD [15] generates temporal super-resolution (TSR) of TVD using adversarial learning. STNet [16] is an end-to-end generative framework that synthesizes spatiotemporal super-resolution volumes for time-varying data. SSR-VFD [17] presents a deep learning framework that produces coherent SSR of 3D vector field data (VFD). TSR-VFD [18] presents a deep learning solution that recovers TSR of 3D VFD for unsteady flow. Although the aforementioned methods can handle 3D SR generation/visualization for spatial and/or temporal data, all of them do not consider SR-based semantics generation in microstructures from LR images/volumes. In medical image field, SR techniques have also been applied, such as CSN [20], 3D-ESPCN [21], DBAN [22], DCSRN [23], mDCSRN-GAN [24], P-GANs [25], Multimodal-Boost [26], SR-Dict [27], etc. Although the above methods can handle 3D SR image generation, none of them explores the SR image generation from the LR grayscale images and its correlational interaction with a particular semantic analysis.

Deep Learning for Super-Resolution Image Segmentation. Recently, some researchers start to achieve image segmentation in terms of higher resolution from low resolution input. There are few research works along this direction. DSRL [28] proposes a dual-stream multi-task network to predict SR images and its semantic segmentation at the same time. Both tasks share the same feature encoder and the inter-stream features at the decoder stage are related by a feature affinity loss. However, their feature affinity loss involves a subsampling process, which thus lacks the feasibility when applying to the segmentation of sparse and tiny objects such as micro vessels. Furthermore, this method can only work on 2D natural images for dense semantic segmentation. SRDA-Net [57] proposes a multi-task model for super-resolution and semantic segmentation (SRS), and then the pixel-level and output-space domain classifiers are designed to guide the SRS model to learn domain-invariant features by the adversarial learning, which has been applied in remote sensing 2D images. SRSNet [58] includes three subnetworks, two of which are used for extracting HR feature image stacks, and the other one is used for segmenting the HR feature image stacks. The SR module and the segmentation module work sequentially. In this case, the performance of the final segmentation is highly dependent on the quality of the former steps (which is not co-optimized). SegSRGAN [29] applies an SRGAN-like generator to predict SR image and the corresponding segmentation together, and trains a discriminator to distinguish whether the segmentation results are from real HR images or SR ones. Inspired by [28], PFSeg [30] proposes a patch-free 3D medical image segmentation method to realize HR segmentation with LR input, but a cropped HR patch from the original image is still needed as restoration guidance during training, which limits the performance on testing when only having LR input. Another limitation is that it is really difficult to realize fully patch-free once the volume image is relatively large (such as large-scale microstructure images) due to the hardware affordability. In short, there is no method exploring how the joint SR-based semantic learning can be systematically leveraged to improve the microstructured 3D object capture from LR images.

Deep Learning for Vessel Segmentation. Since our work uses micro vessel applications as test beds, we briefly review most related work on data-driven vessel extraction and segmentation. Recently, several deep learning based methods have been proposed to extract vessels from 2D retinal images. DeepVessel [31] addresses retinal vessel segmentation as a boundary detection task by using a CNN with a side-output layer to learn discriminative representations, and a conditional random field layer that accounts for non-local pixel correlations. Li et al. [59] presented a supervised method for vessel segmentation by using the cross-modality data transformation from retinal image to vessel map. Mo and Zhang [32] developed a deep supervised fully convolutional network (FCN) by leveraging hierarchical features of the deep networks for retinal vessel segmentation. Liskowski and Krawiec [33] 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. [34] incorporated a graph neural network (GNN) into a unified CNN architecture to jointly exploit both local appearances and global vessel structures. These methods can perform well on the 2D vessel segmentation task, but are far from being satisfactory or feasible in the 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 [35] presents a network inspired by the 3D U-Net [60] and the Inception modules [61] for segmentation of the cerebrovascular network in MRA images. DeepVesselNet (DVN) [36] and VesselNet [37] 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. Recently, Wang et al. [38,39] presented end-to-end deep learning methods for robust extraction and visualization of 3D microvascular structures through embedding the image composition, generated by maximum intensity projection (MIP), into the 3D volumetric image learning process. Banerjee et al. [40] presented a topology-aware learning strategy for volumetric segmentation of intracranial cerebrovascular structures. Pal et al. [41] proposed an end-to-end multiscale residual dual attention deep neural network for resilient major brain vessel segmentation. Guo et al. [42] employed MIP to decrease the dimensionality of 3D volume to 2D image for efficient annotation on 3D vascular segmentation. However, none of the above methods considers applying SR techniques to segment and visualize high-fidelity 3D microvascular structures, which is viable to address the challenges for complicated geometry and topology as well as accounting for the small size of micro vessels.

3. ROSE

The fundamental inspiration of the proposed work is to achieve the SR-oriented semantic joint embedding guided computation. Our work presents a new computing and learning paradigm to combine direct 3D volume processing and SR fine-grained imaging clues for effective 3D microvasculature exploration. In this work, we design a novel method to support the 3D data analytics, such as semantic extraction, by using

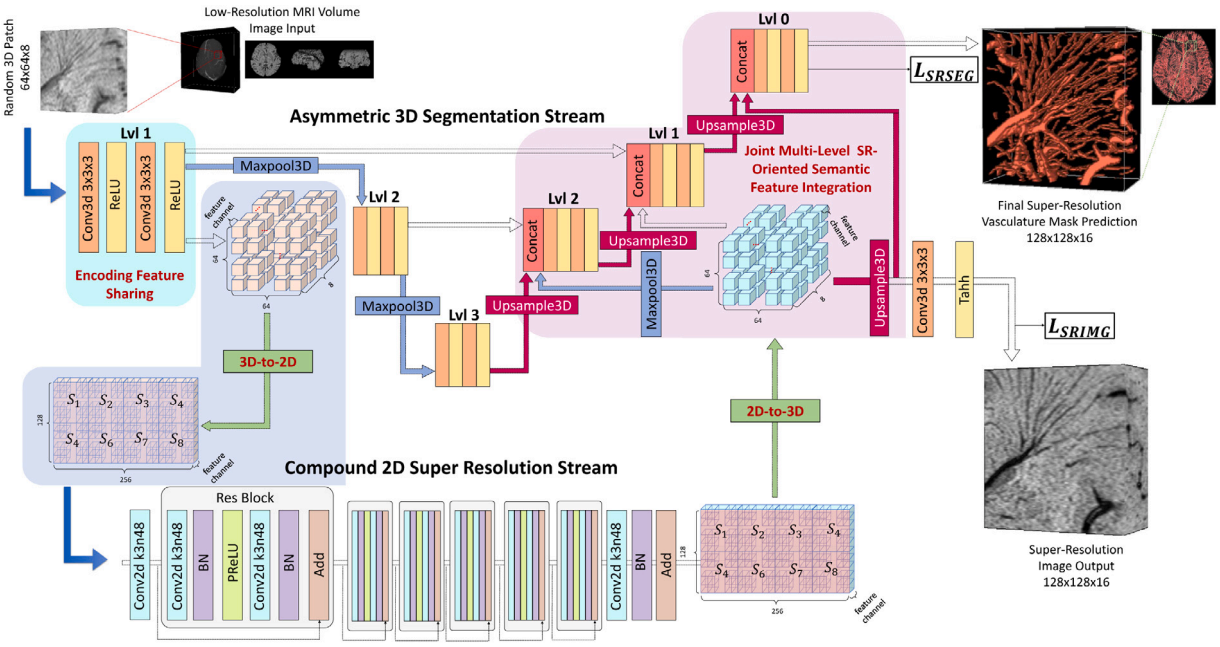


Fig. 2. Our network architecture consists of two streams: an asymmetric 3D segmentation stream and a lightweight compound 2D super resolution stream. Both streams share Lvl 1 encoder at the beginning (highlighted in a bright blue frame, explained the details in Section 3.1.2), then the feature of the input 3D patch volume F_V is converted into the feature of 2D compound plane F_S through our 3D-to −2D process (highlighted in a light indigo frame, explained the details in Fig. 3) for the 2D SR learning. The learned compound 2D SR feature F_S from the SR stream is restored to the 3D SR feature F_V through a 2D-to −3D process (inversion of the 3D-to −2D process). The dimension expansion happens in 3D and F_V is orientedly integrated and fused at the multiple decoder stages/levels in the asymmetric segmentation stream (highlighted in a lavender frame, explained the details in Fig. 4).

the SR-oriented cross-modal feature interaction and aggregation. Instead of conducting the processing sequentially as in the traditional 3D data analytics pipelines, this paradigm enables qualitative exploration of the 3D volume data through jointly learning the compound 2D SR imaging computation. Finally, the 3D data analytics can be conducted in a joint multi-level hybrid semantic-imaging embedding space. In the following, we introduce the components of the ROSE model: network architecture and loss function in detail. Due to the page limit, details of the spatial and frequency LR dataset generation are provided in Supplementary Material.

3.1. Network architecture

The motivation of our ROSE is to utilize the SR technology to orient the microvasculature signals from LR images in favor of boosting the microvasculature segmentation performance, such as on small/micro vessel detection and extraction. The proposed deep neural network is designed to simultaneously learn the major task – 3D microvasculature segmentation, and the auxiliary task – 3D volume image SR synergistically in an end-to-end manner.

Therefore, the proposed ROSE framework mainly consists of a two-stream component (i.e., an asymmetric 3D volume segmentation stream and a compound 2D SR stream), and the bi-directional joint learning between these two streams (i.e., 3D-to −2D feature sharing process and 2D-to −3D joint multi-level oriented feature integration process). The overall architecture is demonstrated in Fig. 2. Due to the limited micro-cerebrovascular data availability and their large size, our network is trained patch-wisely. We use non-cubic volume patches $V \in R^{m \times n \times s}$ (s is along the slicing axis k_z) with a larger spatial size in the axial plane (defined by two MR phase encoding axes k_x, k_y).

3.1.1. Dual-stream components for multi-tasking

Asymmetric 3D Processing Stream for Segmentation. The asymmetric segmentation stream is inspired by the 3D U-Net architecture [60]. It generally follows the encoder–decoder style. The features are extracted from the input 3D image and encoded into deeper levels through hierarchical convolutions and poolings at the encoder stage. The pooling operations further expand the reception field of the convolution and save the memory for deep feature channels. At the decoder stage, the segmentation result is gradually learned from the deepest encoding bottleneck feature through hierarchical upsampling and convolution with the complement features directed from the encoder stage through the skip-connection. However, unlike the classic U-Net, in this task, the encoder and decoder are not symmetric in terms of depth since we need more levels of upsampling to achieve 3D SR segmentation in a larger spatial size (e.g., $2^3\times$ of the input volume size in our experiments). It is noted that the network architecture supports even higher scaling factor of the spatial size. The detailed design information of the asymmetric 3D segmentation stream network is shown in Fig. 2.

2D Imaging Stream for Compound SR. For the counterpart SR stream, we apply a fully convolutional network, which mainly consists of six Res-Blocks inspired by the generator in SRGAN [8] before the dimension expanding layers. In Res-Blocks, the convolution kernel size is $k = 3$, the number of feature maps is $n = 48$, and the stride is 1, which indicate in each convolutional layer. The reason why we choose to compute the 2D image SR here (instead of 3D SR) is more in consideration of efficiency and it is theoretically justifiable. First, as an auxiliary task, a 2D SR subnetwork with Res-Blocks is much more lightweight than its 3D version to carry and train, thus it is not a big burden to the whole network and our major segmentation task. For instance, within the current GPU memory limit, we can only employ one 3D Res-Block, instead of current six 2D Res-Blocks. This will degrade the effectiveness of the SR stream. Second, according to Supplementary Material Section 1 and the MRI imaging formation, the k-space upsampling in our work is mainly along two MR phase encoding directions k_x and k_y, so the slice-wise in-plane (compound) 2D SR feature encoding/vector from the 2D Res-Blocks is adequate in practice. Once we obtain the high-dimensional 2D SR feature encoding / vector from the Res-Block series, we restore it to the SR feature of the original input 3D patch volume followed by a couple of 3D convolutional layers and deconvolutional layers to achieve the final 3D SR image volume in a larger spatial size (upsampling). Note that

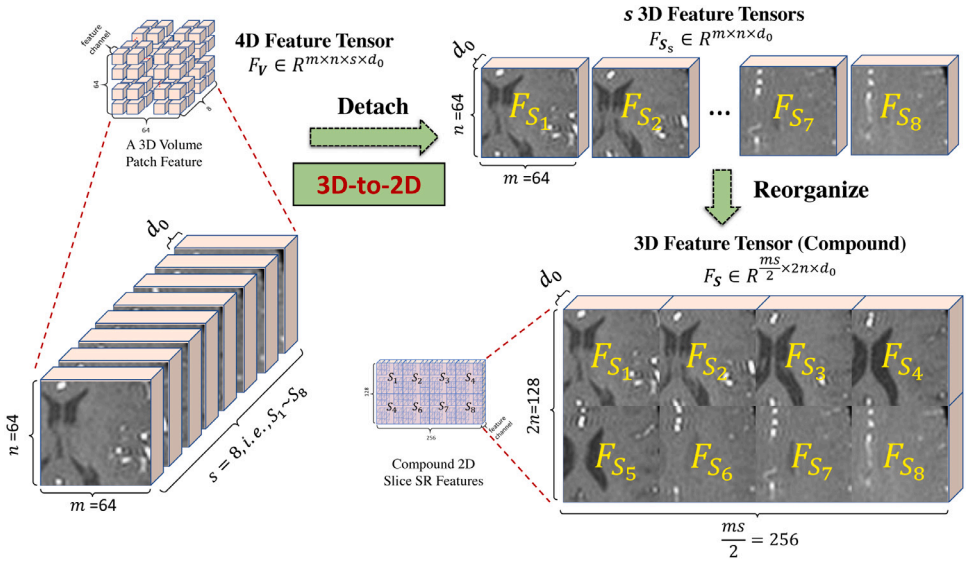


Fig. 3. 3D-to −2D operations: a 3D volume patch feature $F_{\mathbf{V}}$ (i.e., a 4D tensor) is detached into s 2D slice features $F_{\mathbf{S}_1}$, $F_{\mathbf{S}_2}$, …, $F_{\mathbf{S}_s}$ (i.e., 3D tensors) along the slicing axis, and the s 2D slice features are reorganized as one larger 2D compound slice feature $F_{\mathbf{S}}$ (i.e., a larger 3D tensor). 2D-to −3D operation in the joint multi-level oriented feature integration is the inverse process of 3D-to −2D operation.

the spatial upsampling happens in 3D so that the volumetric contextual information could be captured during this process.

3.1.2. *Joint learning for SR-oriented semantics*

In the following, we introduce how the learned hybrid (i.e., semantics and imaging) embedding features from the two streams orient and integrate between each other during different stages and levels based on the above developed 3D semantic processing and 2D SR imaging formulation. There are three major modules, i.e., encoding feature sharing for joint learning, 3D-to −2D operations, and joint multi-level oriented feature integration/fusion.

Encoding Feature Sharing for Joint Learning. Semantic segmentation and image SR are quite different tasks in terms of not only learning objective but also learning process. Our aim is to design an effective and efficient joint learning process for both tasks; as a result, the two-task streams share the initial low-level feature extractor (i.e., two 3D convolution layers and two ReLU layers) at the encoder before the first pooling in the segmentation stream (at the beginning). The advantage of this design is to let both streams capture the common low-level 3D local microvasculature features and then pass them into two streams. The learning information/feature exchange between two streams is expected to be as discriminative and sufficient as possible to acquire the specific segmentation-oriented SR features during the encoding. Consequently, such SR features in turn can better enhance the segmentation result during the decoding. In Section 4, we will demonstrate that the SR-oriented features from deep stream multi-level fusions are more functionally favorable to the SR segmentation task than the standalone/single-level SR features. However, the segmentation stream contains intensive spatial pooling operations favoring the feature expanding in the encoder which the SR stream does not have.

3D-to −2D Operations. In this work, we detach the shared encoding feature, i.e., a 4D feature tensor $F_{\mathbf{V}} \in R^{m\times n\times s\times d_0}$ (d_0 is the feature channel number) of patch volume $\mathbf{V}$ along the slicing axis k_z (vertical axis) into s 3D feature tensors ($F_{\mathbf{S}_1}, F_{\mathbf{S}_2}, \ldots, F_{\mathbf{S}_s}$) with s 2D slices of $m\times n$ resolution ($\mathbf{S}_1, \mathbf{S}_2, \ldots, \mathbf{S}_s$), and then reorganize them into the feature $F_{\mathbf{S}}$ with the larger compound 2D spatial size $\mathbf{S} \in R^{\frac{ms}{2}\times 2n}$ as shown in Fig. 3 to maintain the correct relationship between the spatial domain and the feature channels for the 2D convolution operations in the SR stream. We then apply an additional 2D convolutional layer with 3×3 kernel along with the previous two stream-shared 3D convolutional layers to reach similar effect of the single large-kernel convolution layer (e.g., the 9×9 convolution layer in SRGAN [8]) for a wide reception field capture for image SR task.

Joint Multi-Level SR-Oriented Semantic Feature Integration/Fusion. After the compound 2D slice SR feature $F_{\mathbf{S}} \in R^{\frac{ms}{2}\times 2n\times d_1}$ (d_1 is the feature channel number) is learned through a series of Res-Blocks, it is detached from the compound 2D plane and restored to the SR feature $F_{\mathbf{V}} \in R^{m\times n\times s\times d_1}$ of a 3D patch volume $\mathbf{V}$ (the inverse process of the aforementioned 3D-to −2D operation) before the spatial dimension expansions of the 4D feature tensors. After that, we construct the cascaded three-level hybrid (semantics $F_{\mathbf{V}}^{sem}$ and imaging $F_{\mathbf{V}}^{img}$) embedding spaces, where the direct restored SR feature (at Level 1) is further concatenated at one deeper-decoder level (Level 2) by maxpooling and at one upscaled-decoder level (Level 0) by upsampling to further enhance the multi-scale SR information integration for 3D segmentation. The multi-level semantics and imaging feature fusions are computed by:

$$\begin{aligned}
&\text{Level 2}: F_{\mathbf{V}}^{sem} \in R^{\frac{m}{2}\times\frac{n}{2}\times\frac{s}{2}\times d_1} \oplus F_{\mathbf{V}}^{img} \in R^{\frac{m}{2}\times\frac{n}{2}\times\frac{s}{2}\times d_1},\\
&\text{Level 1}: F_{\mathbf{V}}^{sem} \in R^{m\times n\times s\times d_1} \oplus F_{\mathbf{V}}^{img} \in R^{m\times n\times s\times d_1},\\
&\text{Level 0}: F_{\mathbf{V}}^{sem} \in R^{2\ m\times 2n\times 2s\times d_1} \oplus F_{\mathbf{V}}^{img} \in R^{2\ m\times 2n\times 2s\times d_1},
\end{aligned} \tag{1}$$

where $\oplus$ is the concatenation operation, $F_{\mathbf{V}}^{img} \in R^{\frac{m}{2}\times\frac{n}{2}\times\frac{s}{2}\times d_1} = MaxPool(F_{\mathbf{V}}^{img} \in R^{m\times n\times s\times d_1})$, $F_{\mathbf{V}}^{img} \in R^{2\ m\times 2n\times 2s\times d_1} = UpSample(F_{\mathbf{V}}^{img} \in R^{m\times n\times s\times d_1})$. With this joint multi-level SR-oriented semantic feature fusion design, we can compute the joint microvasculature probabilities with the specified task orientation from different-scale (e.g., $32\times 32\times 4$, $64\times 64\times 8$, and $128\times 128\times 16$) and different-modality (e.g., vessel semantic and SR imaging features) embeddings. Finally, the 3D SR volume processing (segmentation) can be synergistically enhanced and complemented from the compound 2D SR imaging computation via the cascaded three-level hybrid embedding feature fusion as:

$$\begin{aligned}
&F_{\mathbf{V}}^{srsem} \in R^{2\ m\times 2n\times 2s\times d_1}\\
&= UpSample\left(\ UpSample(F_{\mathbf{V}}^{sem} \in R^{\frac{m}{2}\times\frac{n}{2}\times\frac{s}{2}\times d_1} \oplus F_{\mathbf{V}}^{img} \in\right.\\
&\left. R^{\frac{m}{2}\times\frac{n}{2}\times\frac{s}{2}\times d_1}) \oplus F_{\mathbf{V}}^{img} \in R^{m\times n\times s\times d_1}\ \right) \oplus F_{\mathbf{V}}^{img} \in R^{2\ m\times 2n\times 2s\times d_1},
\end{aligned} \tag{2}$$

where $F_{\mathbf{V}}^{srsem}$ is the final joint SR semantic embedding feature. The details of the proposed scheme are shown in Fig. 4 and its effectiveness

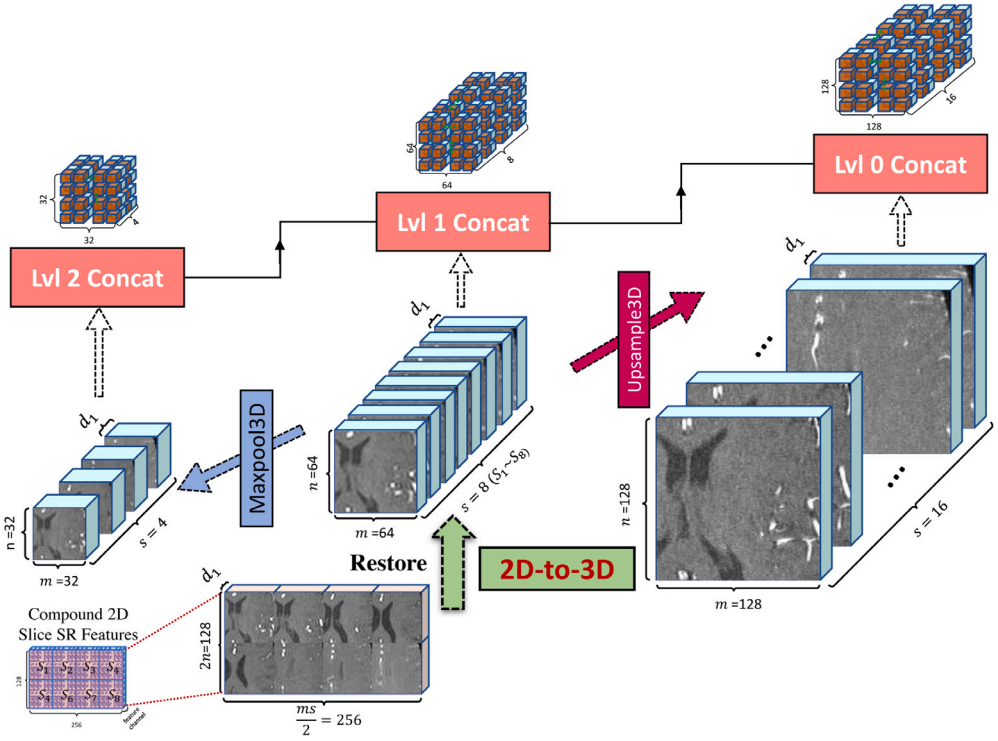


Fig. 4. Joint multi-level SR-oriented semantic feature integration: the restored 3D volume SR feature F_V is directly fused (concatenated) to decoder at Lvl 1 stage (the 3D segmentation decoder is simplified to sketch here (refer Fig. 2 for specific architecture). Meanwhile, F_V is also downsampled and upsampled (for spatial dimensions) to fuse at deeper (Lvl 2) and higher (Lvl 0) decoder stages correspondingly, so that we can construct the cascaded three-level hybrid (semantics and imaging) embedding spaces for the final SR microvasculature segmentation.

will be demonstrated in Section 4. The key goal of this step is to integrate the semantic feature embedding space with the SR imaging feature embedding space at multiple scales in order to effectively complement and enhance the joint feature representation for microvasculature extraction from the developed multi-tasking components.

3.2. Loss function

Our ROSE is a two-stream multi-task learning network aiming to use simultaneous SR enhancement to better strengthen the weak/subtle/blurring microvasculature signal in LR images to facilitate the segmentation in finer details. Consequently, the network loss function contains a major SR segmentation term L_{SRSEG} and an auxiliary SR imaging term L_{SRIMG}. L_{SRSEG} is a Dice Similarity defined in 3D SR semantic space:

$$L_{SRSEG} = 1 - \frac{2\Sigma_{x\in\mathbf{V}} p(x)g(x) + \delta}{\Sigma_{x\in\mathbf{V}} p(x) + \Sigma_{x\in\mathbf{V}} g(x) + \delta}, \tag{3}$$

where $p(x)$ and $g(x)$ are the predicted voxel-wise microvasculature probability maps and ground truth binary labels within the query volume patch $\mathbf{V}$ of size $W \times H \times D$, respectively. δ is a small smooth constant.

L_{SRIMG} is a voxel-wise L_1 loss (mean absolute error) in terms of the intensity between the predicted SR image and the ground truth HR image, which is defined in 3D SR imaging space as:

$$L_{SRIMG} = \frac{\Sigma_{x\in\mathbf{V}} \left| I^{HR}(x) - I^{SR}(x) \right|}{W \cdot H \cdot D}, \tag{4}$$

where $I^{HR}(x)$ and $I^{SR}(x)$ are the voxel intensity of the HR and SR images, respectively.

The final total loss L is defined as follows:

$$L = L_{SRSEG} + \lambda L_{SRIMG}, \tag{5}$$

where λ is a hyperparameter adjusting the training balance between the two tasks of the 3D SR semantic segmentation and the compound SR imaging. We set it as 0.1 based on our extensive experiments.

4. Experiments and results

4.1. Datasets

In this work, we evaluate our ROSE model performance on the challenging microvasculatures in brain vasculature with complicated geometric and topological structures. To our knowledge, the following real patient datasets are two largest datasets currently in public and clinical research study on microvascular segmentation and extraction.

TubeTK MRA Dataset comes from a public MRA dataset of the University of North Carolina at Chapel Hill [62], acquired on a 3T MR system. In this dataset, there are 42 patient cases with manually-labeled vessel segmentation masks. The voxel spacing of the MRA image is $0.5 \times 0.5 \times 0.8$ mm^3 with a volume size $448 \times 448 \times 128$.

MICRO-MRI Dataset represents the next generation of microvascular imaging, named as Microvascular In-vivo Contrast Revealed Origins Magnetic Resonance Imaging (MICRO-MRI) [63–65], which is acquired and collected on a 3T MR scanner by neurologists and radiologists within our collaborative group. Here we use the susceptibility weighted imaging (SWI) data in MICRO-MRI dataset, since currently it is the only imaging modality available containing micro-level vessels that can well demonstrate our model's capability of capturing tiny vessels. There are 12 patient cases in this SWI dataset with the high-quality brain region micro-level vessel labels. The voxel spacing is $0.22 \times 0.22 \times 1$ mm^3 with a volume size $1024 \times 832 \times 56$. The ground truth vessel labels are acquired by applying an existing state-of-the-art deep learning-based vessel segmentation method [38] at first and then followed by the post manual refinement from our collaborative domain experts. It is noted that vessel labels for the SWI images contain not only micro-vessels, but also major-level vessels. Unlike the vessel labels for SWI images in [38], which are manually refined based on an adaptive threshold-based region growing method (ATRG) [66], the vessel labels in the current SWI dataset are much clearer and focus more on the brain center region micro-level vasculatures.

4.2. Implementation details

For both TubeTK dataset and MICRO-MRI SWI dataset (HR and LR image examples are provided in Supplementary Material and Video), we apply the MR-based skull-stripping [67] to extract the pure brains. We then normalize the image intensity range into [0, 1] for a better numeric balance between the probability-based loss term of the segmentation stream and intensity-based loss term of the SR stream. We use these original TubeTK MRA dataset and MICRO-MRI SWI dataset as the HR images in our experiments. The ground truth vessel labels for training are generated on HR images. Due to the data availability and the large volume size in our micro-cerebrovascular image datasets, we trained and tested our ROSE patch-wisely. According to Supplementary Material Section 1, in practice we set $c_x = 3$, $c_y = 3$ and $c_z = 1$ indicating the k-space downsampling happens in MR phase encoding plane. The spatial downsampling factor r is set to 2. Consequently, the input and output patch sizes in our network are $64 \times 64 \times 8$ (LR) and $128 \times 128 \times 16$ (SR), respectively. The input patches are randomly extracted from the image volumes with overlapping, focusing mainly on the brain area. More specifically, due to the large background without any vessels in the volume image, we detect the tight boundary of the brain in the volume image and extract the patches within brain regions, so that patches are guaranteed to contain brain vessels accordingly. The patch numbers for each TubeTK MRA image and SWI image are 80 and 300, respectively. The random training/validation/testing patient case splits are 33/3/6 (42 in total) and 7/2/3 (12 in total) for the TubeTK dataset and the MICRO-MRI SWI dataset, respectively. As a result, the actual training/testing sample amounts are 2640 (80×33)/480 (80×6) and 2100 (300×7) / 900 (300×3) for these two datasets, respectively, which are adequate in terms of training and testing patch-wise sample numbers. All the numerical evaluations are reported in terms of the whole image patched with no overlapping average.

The network is trained with Adam Optimizer and the training batch size is 10 on both datasets. For the TubeTK dataset and MICRO-MRI SWI dataset, the learning rate, the learning rate decay factor, and the learning patience are 0.0005/0.0001, 0.8/0.5, 10/10, respectively. The network is implemented in TensorFlow framework and the total training time is around 4∼5 h (for the above two datasets) on one NVIDIA GeForce GTX 3090 GPU with 24 GB GDDR6X memory. The detailed training time for all methods is given in Table 3. As for the inference/testing time, the per volume inference times on our ROSE method are 9.3 s (on a TubeTK volume) and 42 s (on a MICRO-MRI volume), respectively. All other deep learning methods in comparison are quite similar and efficient on the reference time as ours. The only exception is the Vesselness algorithm, since it is a model-driven method and all the parameters are manually adjusted, which is not comparable with the end-to-end deep learning methods. Its inference time is 186.6 s on a TubeTK volume and 790 s on a MICRO-MRI volume. *Dataset and source code will be made available upon publication.*

4.3. Evaluation metrics

All the segmentation results from our model and other methods in comparison are numerically evaluated using the following metrics. Most are defined based on the confusion matrix (TP, FP, TN, FN), where TP is true positive, FP is false positive, TN is true negative, and FN is false negative. Considering the vessel segmentation result is extremely class-imbalanced, we do not include the metrics of Accuracy and Specificity, which involve TN significantly.

Dice Similarity and **Jaccard Index**, which are defined as $2TP/(2TP+FP+FN)$ and $|Pred \cap GT| / |Pred \cup GT|$, respectively, measuring the similarity between the prediction ($Pred$) and the ground truth (GT) vessel labels in different ways, where the later only counts TP once in both the numerator and denominator.

Sensitivity, $TP/(TP+FN)$, measures the model ability to capture as many vessels as possible from the large sparse-object image volume.

Precision, $TP/(TP+FP)$, measures the model ability to extract the correct vessel voxels regardless the challenging noises and the huge portion of the background.

For the side-products from our SR stream, even though they are not our major objective, we use the following standard image metric to present the corresponding results quantitatively for comparison with other methods on the SR image outputs.

Peak Signal-to-Noise Ratio (PSNR) measures the ratio between the maximum possible power of a signal and the power of corrupting noise that affects the fidelity of its representation. It is defined as:

$$PSNR = 10\log_{10}\frac{(\max(GT_i))^2}{\frac{1}{V}\sum_{i\in V}|GT_i - Pred_i|^2}, \tag{6}$$

where V is the image volume and i represents the voxel index in V.

The quantitative comparisons are provided on both TubeTK MRA and MICRO-MRI SWI datasets, and the qualitative/visualization comparisons are provided on MICRO-MRI SWI dataset in order to better demonstrate the micro vessel segmentation. We apply different visualization approaches to qualitatively evaluate our method from multiple scales and views, such as volume rendering (Fig. 1) for a large-scale/global comparison, surface rendering (Fig. 5) for a middle-scale/regional comparison, image-based visualization (Figs. 6, 7, and 8) in a fine-scale/voxel-level comparison due to the tininess of the micro vessels. There are additional results in Supplementary Material.

4.4. Comparison with SR-based segmentation methods

In this section, we compare our ROSE model with three state-of-the-art deep learning methods designed for simultaneous segmentation and image SR, i.e., DSRL [28], SegSRGAN [29], and PFSeg [30]. Meanwhile, we demonstrate that our ROSE has the unique properties for computing both SR-oriented instance segmentation and segmentation-oriented image SR. The 3D surface rendering visualization between our method and state-of-the-art SR-based segmentation methods on MICRO-MRI dataset is shown in Fig. 5 and Supplementary Video. It shows the complicated and superbly dense microvasculature extraction in both subarea and whole brain center region.

As shown in Table 3, DSRL has less model parameters since it only deals with 2D images and consequently lacks ability of expansion on the third dimension. To adapt DSRL in our experiment setting and compensate the 3D contextual downsampling, we first apply DSRL to accomplish 2× SR in MR encoding plane, use post cubic interpolation to raise the spatial size along the slicing axis, and adjust their network to suit our medical dataset. From Table 1 we can see that our ROSE outperforms DSRL greatly in terms of all segmentation metrics on two datasets. It is noted that DSRL has competitive SR performance; however, this advantage fails to bring the effective improvement in the segmentation result due to the lack of intrinsic connection and enhancement of its SR to segmentation. From the zoomed-in patches in Fig. 5 and error maps in Fig. 6, it can be seen that DSRL does not capture as many micro vessels as our method does, indicated as blue vessels (FN) in error maps, even though both methods have similar SR results. Besides the error maps, we also present the vessel segmentation visualization from our GUI, in which the semi-transparent vessel masks are mapped onto the interpolated LR SWI image as a background (since the vessel segmentation is generated from LR inputs). Through this way, one can observe the underlying vessels to get a better sense of the segmentation quality. This visualization format works as a good complement for the error map, as sometimes the slight difference on error maps does not necessarily indicate inaccurate prediction as long as the vessel objects are captured. The error maps for small object segmentation may sometimes exaggerate the discrepancy between the prediction and the ground truth.

Comparing with SegSRGAN [29], we follow their implementation on both vessel datasets and the results are reported in Table 1. Our

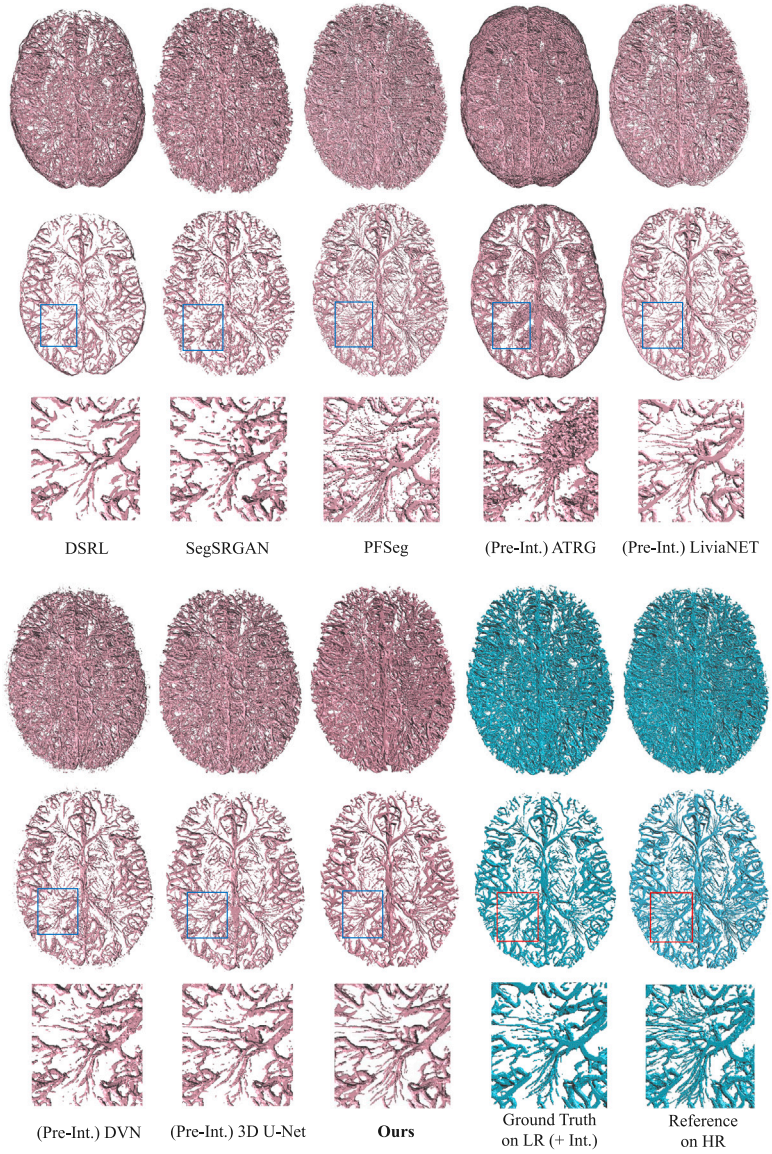


Fig. 5. The qualitative 3D surface rendering visualization comparison between our method and state-of-the-art SR-based segmentation and vessel segmentation from one testing case of MICRO-MRI dataset. 1st and 4th rows: overview of the dense whole brain center region microvasculature results; 2nd and 5th rows: the thalamus subareas of $S_{15} - S_{28}$ with very dense microvasculature distribution; 3rd and 6th rows: the zoomed-in patches for one specific region of interest on such subareas.

Table 1
The quantitative evaluation of different SR-based segmentation methods on TubeTK and MICRO-MRI datasets. Note: Sens.: Sensitivity, Prec.: Precision, Jacc.: Jaccard. Best results are in bold and gray rows are our results.

Datasets	Methods/Metrics	Dice (%) ↑	Sens. (%) ↑	Prec. (%) ↑	Jacc. (%) ↑	PSNR ↑
TubeTK	DSRL [28]	57.64	53.06	63.42	40.49	**39.74**
	SegSRGAN [29]	50.32	45.27	56.71	33.62	36.87
	PFSeg [30]	63.73	**61.04**	66.90	46.80	26.68
	Ours	**65.59**	60.70	**71.46**	**48.79**	37.80
MICRO-MRI	DSRL [28]	65.89	57.45	77.28	49.13	23.28
	SegSRGAN [29]	67.47	62.35	73.92	50.92	22.37
	PFSeg [30]	77.48	74.42	81.23	63.40	14.63
	Ours	**80.57**	**75.55**	**86.36**	**67.47**	**25.69**

method significantly outperforms SegSRGAN on all segmentation metrics. From Figs. 5 and 6, we can see that SegSRGAN lacks the accuracy of targeting at small and sparse objects with the complicated topology. Unlike our oriented-multi-stream-in-parallel method where SR serves as a sub-objective to boost the segmentation, SegSRGAN treats SR and segmentation equally, which is sub-optimal for the multi-tasking scenario. Consequently, the numeric segmentation and SR performance are compromised to each other. However, it is worth mentioning that SegSRGAN provides the best visual/qualitative effect of the SR image despite some GAN-based artifacts as circled in Fig. 6. SegSRGAN-generated SR images are sharper, more realistic to human eyes, and contain more details along with higher noise, while SR images from our method and DSRL are much smoother due to pure mean absolute error (MAE) constraint in the loss function. The results which we generate from their code conform the observation from their original paper, which may benefit from GAN-based model.

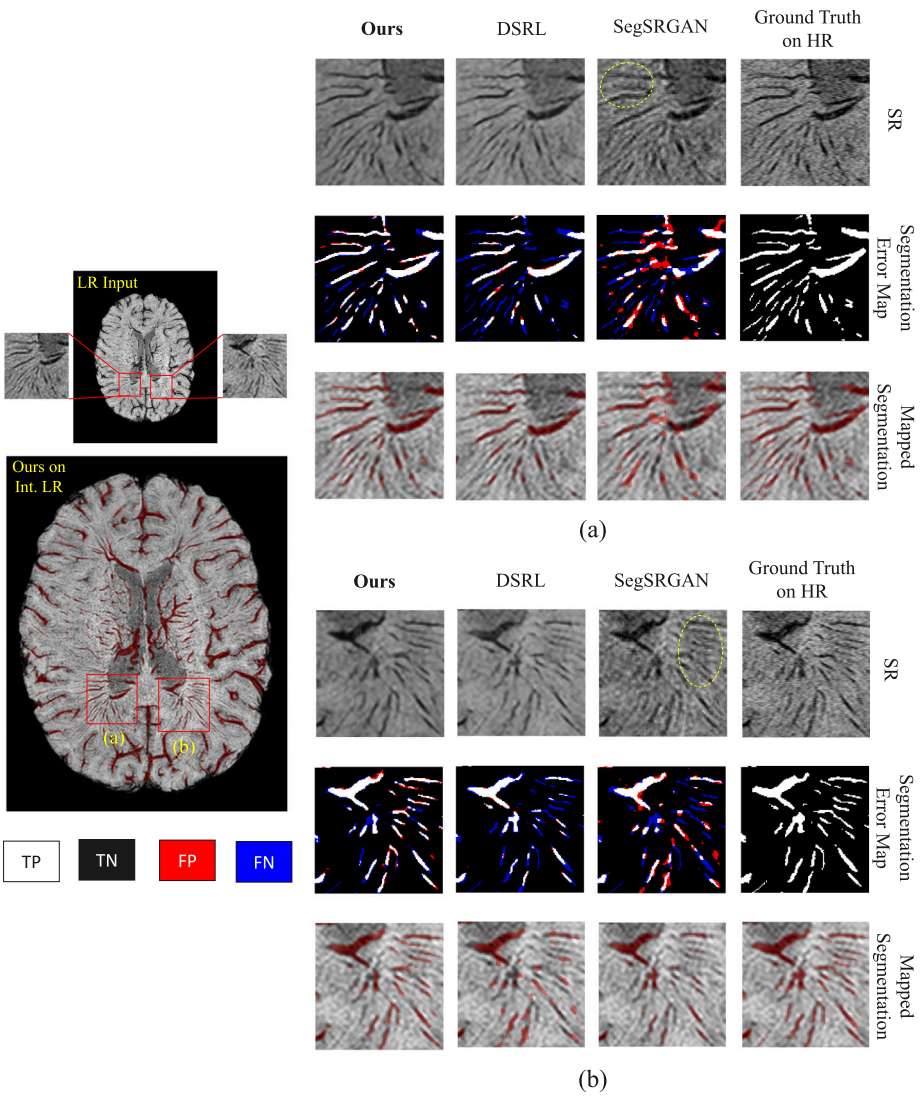


Fig. 6. The qualitative comparison between our method and state-of-the-art SR-based segmentation methods on MICRO-MRI dataset. Two patches (a) and (b) from one selected slice are zoomed in for the evaluation. In each patch, we visualize the SR results, segmentation error maps, and the semi-transparent red vessel masks mapped on the interpolated LR SWI image (as a background) from our GUI tool. In the error maps, we use white color for True Positive (TP), black color for True Negative (TN), red color for False Positive (FP), blue color for False Negative (FN).

However, we would like to emphasize that our work focuses mainly on the segmentation while SR is the side-product. From SNR perspective, higher resolution usually implies lower SNR. In both datasets, the average SNR of our generated SR images against ground truth HR ones is 59.31%/57.28% and 77.55%/75.50%, respectively. This phenomenon in turn justifies that our SR weakens the noise interference while strengthens vessel signals, resulting from the segmentation-oriented SR in our joint learning design. In addition, SegSRGAN requires pre-upsampling before sending the LR input into the network, which is more memory-consuming. The total number of parameters is over two times as ours (in Table 3) and their super-heavy discriminator (large 3D convolutional kernel size, redundant embedding feature channels) and adversarial training make the training process quite painful. Furthermore, the direct SR on an entire image (without the joint vessel feature sharing and multi-level oriented integration like ours) may also increase the noises and artifacts, so that it is clear to see that the PSNR of SegSRGAN is a bit low. This does hurt the small vessel extraction.

PFSeg [30] is an SR-based patch-free 3D medical image segmentation framework. However, our microvasculature extraction task needs much larger size of the volume image than their design (e.g., 1024 × 832 × 56 vs 192 × 192 × 128), so PFSeg still needs to apply patch-wise technique in order to meet the GPU memory limit. Furthermore, they directly use an HR patch cropped from the original image as restoration guidance for both segmentation and SR tasks, which enhance both vessel signals as well as noise in their feature learning. This is really harmful in the final SR segmentation. From Table 1, it is clear to see that PFSeg has reasonable performance in terms of segmentation metrics on both datasets, but it has lowest PSNR among the compared methods since their SR image result is very noisy. Fig. 5 also shows that PFSeg can capture many micro vessels as well as too much noise. Meanwhile, PFSeg has two heavyweight 3D neural networks for both SR and segmentation tasks, whose parameter size is about three times as ours (in Table 3).

4.5. Comparison with vessel segmentation methods

We discuss the performance of our ROSE and several state-of-the-art (model-based and deep learning-based) vessel segmentation methods on both TubeTK MRA and MICRO-MRI SWI datasets in this section. Since all vessel segmentation methods in comparison have no scheme to raise the resolution in their original techniques, we add the upsampling/interpolation procedure to the LR inputs before feeding them into

Table 2
The quantitative evaluation of different vessel segmentation methods on TubeTK and MICRO-MRI datasets. Note: Pre-Int.: Pre-Interpolation, Sens.: Sensitivity, Prec.: Precision, Jacc.: Jaccard. Best results are in bold and gray rows are our results. *Vascular-MIP [42] result is generated based on the input with high-resolution MRA images. Since the annotation of MIP images on MICRO-MRI dataset is not available, there is no result in this table. '-' means 'not applicable' due to lack of their implementations or results.

Datasets	Methods/Metrics	Dice (%) ↑	Sens. (%) ↑	Prec. (%) ↑	Jacc. (%) ↑
TubeTK	(Pre-Int.) Vesselness [68,69]	31.79	23.55	48.99	19.62
	(Pre-Int.) LiviaNET (3D FCN) [70]	57.75	49.39	71.11	40.68
	(Pre-Int.) DVN [36]	57.84	55.14	60.89	40.70
	(Pre-Int.) Full 3D U-Net [60]	63.75	56.92	**72.64**	46.80
	Vascular-MIP* [42]	64.52	-	-	-
	Ours	**65.59**	**60.70**	71.46	**48.79**
MICRO-MRI	(Pre-Int.) ATRG [66]	65.23	**83.17**	54.38	48.46
	(Pre-Int.) LiviaNET (3D FCN) [70]	72.96	65.56	82.78	57.48
	(Pre-Int.) DVN [36]	74.97	71.10	79.51	59.97
	(Pre-Int.) Full 3D U-Net [60]	77.56	72.41	83.61	61.59
	Ours	**80.57**	75.55	**86.36**	**67.47**

their methods/networks (on training, validation, and testing) in order to simulate using standard SR methods followed by segmentation networks as well as fair comparisons on our 3D SR vessel extraction. One fundamental limitation of this line of methods is that the SR methods increase both noise and vessel signals, which will affect the downstream segmentation tasks. The 3D surface rendering visualization comparison between our method and state-of-the-art vessel segmentation methods on MICRO-MRI dataset is shown in Fig. 5 and Supplementary Video.

Frangi-filter [68] based Vesselness algorithm [69] and ATRG [66] are two intensity-based model-driven methods that domain experts are currently using for MRA and SWI data modality, respectively. From Table 2, we can observe that our method outperforms Vesselness algorithm on all metrics at a great margin on TubeTK MRA dataset. While in MICRO-MRI dataset, our method has much better numeric results over all other metrics except the sensitivity. The high sensitivity may cause from the bold threshold cutting (i.e., the threshold is set to allow as many vessels to be covered as possible). Note that the SWI image has a large volume of major-level vessels, which creates redundant coverage for the ATRG method on these large vessels and thus achieves higher sensitivity. However, as shown in Fig. 7, ATRG method also induces more false positive (red on error maps) than other methods, which leads to a low precision. Moreover, its high sensitivity does not necessarily imply good micro-level vessel capture. For instance, we zoom in two specified regions (micro vessels are concentrated) from a single SWI slice where tiny vessels are challenging to be extracted, and we can see that ATRG method captures the least micro vessels among the four methods. Fig. 5 also shows the same conclusion.

LiviaNET [70] is a 3D fully convolutional neural network (3D FCN with 0.71M parameters in Table 3) and DeepVesselNet (DVN) [36] is a lightweight (0.06M parameters in Table 3) FCN. Their difference is that DVN utilizes cross-hair convolution kernel to further reduce model parameters and introduces an advanced Cross-Entropy loss to deal with the strong class imbalance. In general, their performance is quite similar. In Fig. 5, we can see that LiviaNET captures more vessels but also has more noise. From Fig. 7 one can see that DVN fails to effectively extract extra-fine micro-level vessels through the SR SWI images from the pre-interpolation of the LR inputs. Table 2 shows that our method is obviously better than LiviaNET and DVN on both datasets. 3D U-Net [60] is one of the most robust neural networks for 3D medical image segmentation. Here we use the full 3D U-Net for comparison, whose parameter number is almost four times as ours shown in Table 3. From Fig. 7 and Table 2, one can see that our method achieves better performance compared with 3D U-Net on both datasets. 3D U-Net lacks some sensitivity to detect very subtle micro-level vessel soundly. The key comparison parts between ours and 3D U-Net are circled in yellow on error maps. Although Fig. 7 only presents two zoomed-in patches from a single slice for the convenient demonstration, actually our method's out-performance over 3D U-Net on tiny vessels spreads all over the brain area.

To alleviate the reliance on 3D vascular annotation, Vascular-MIP [42] guides the segmentation of blood vessels in 3D space via 2D MIP annotations. We compare Vascular-MIP on the fully supervised model training with the annotation of 30 MIP images as they did in the original paper. Its Dice value is 64.52% and our Dice value is 65.59% on TubeTK dataset. It is noted that the input of their model is the exact high-resolution MRA images; however, if we use upsampling/interpolation procedure to the LR inputs as ours, their Dice value could be much worse.

4.6. Results on abnormal micro vessel segmentation

To our knowledge, we are the first to investigate and apply our ROSE for 3D microvasculature segmentation and visualization *in-vivo* from an LR input. Several our collaborative domain experts, such as neurologists and radiologists, have tested our ROSE and found that it exceeds the quality of the ground truth on LR images (such as shown in Fig. 5).

Furthermore, our method demonstrates a compelling ability to extract and visualize abnormal micro vessels in clinical cases we evaluated. Fig. 8 shows one real patient micro vasculature abnormality case captured by our method and another one is shown in Supplementary Material. The HR SWI data were collected on patients with metastases and relapsing-remitting multiple sclerosis (RRMS) on a 3T MRI scanner.

In Fig. 8, even with the challenging downgraded LR input SWI images, our method is still capable of generating convincing micro vessel segmentation. The SR branch effectively alleviates the vessel diffusion as well as artifacts in LR SWIs, achieving more thorough extraction of the micro vessels in the regions of interest (ROIs). There are two vascular anomalies for this metastases case as shown in Fig. 8. The global infiltration of vessels is framed in yellow, in which the vessels are irregularly dense. The circular pattern of vessels is framed in red, where the vessels are compacted at the boundary of the lesion. Our method can well reveal the abnormal microvascular morphology and distribution in finer detail. More details are shown in Supplementary Video.

4.7. Ablation study on ROSE model

In this section, we investigate the effectiveness of the proposed modules in Section 3.1.2, i.e., encoding feature sharing for joint learning and joint multi-level oriented feature integration at different levels/stages. All the ablation experiments are conducted on TubeTK MRA dataset under the same setting. The baseline here is only one stream with our asymmetric 3D segmentation stream (ROSE w/o SR), i.e., a 3D U-Net with reduced parameters followed by two additional convolutional layers (Conv3D + Deconv3D) for the spatial upsampling. This baseline does not involve any SR learning paradigm (i.e., compound 2D SR stream) and the final SR segmentation performance is shown in Table 4, which is the lowest among all other settings.

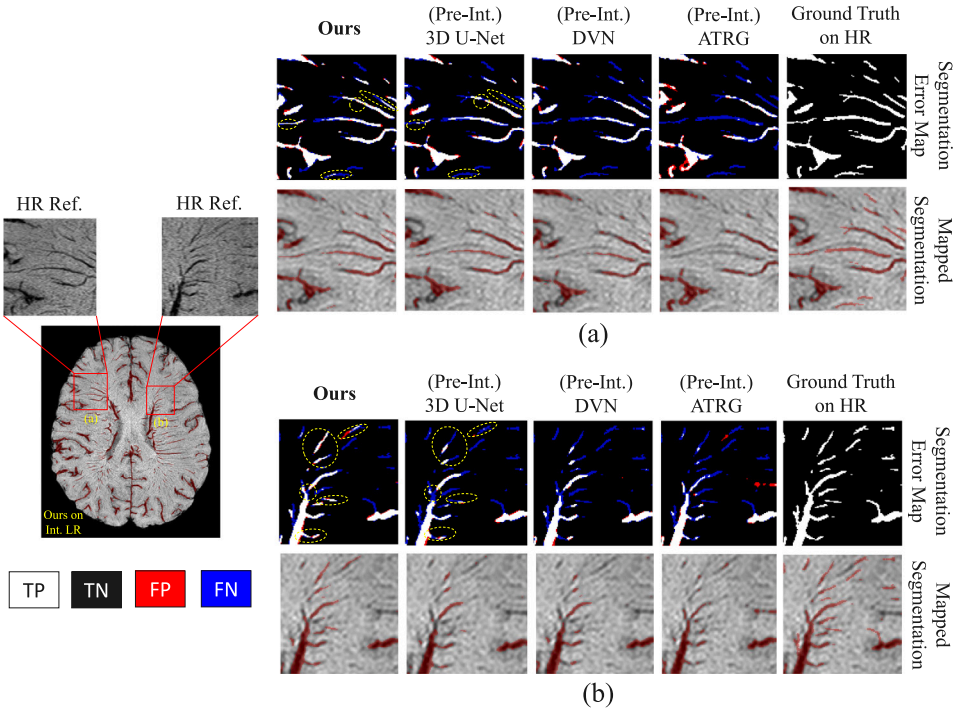


Fig. 7. The qualitative comparison between our method and state-of-the-art vessel segmentation methods on MICRO-MRI dataset. Two patches (a) and (b) are selected from two typical brain regions on a single slice, where micro-level vessels are intensively distributed. Some subtle differences are circled in yellow. Error maps, vessel masks mapped on the interpolated LR SWI image, and the HR SWI patch reference are provided.

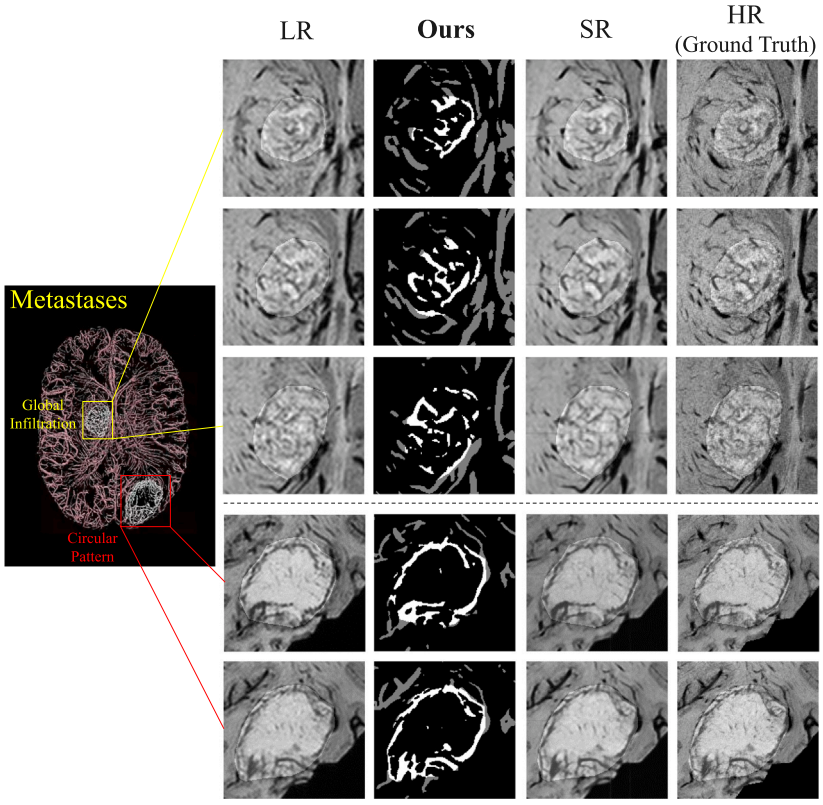


Fig. 8. The segmentation and visualization result on one vascular anomaly case (Metastases) by our method. The left one shows 3D surface visualization of our SR micro vessel extraction (highlighted tumor regions). The zoomed-in patches (with different volume slices) in different formats are listed. (Here the LR input SWI images are interpolated to the same size of the HR ones for a better examination, and the actual input size is halved along three axes).

Table 3
Comparison of model parameter size and training time on different deep learning methods. Note: M is million. Gray column is ours. The training time is captured on TubeTK dataset.

Methods	DVN	LiviaNET	3D U-Net	DSRL	SegSRGAN	PFSeg	Ours
# Conv. Dim.	2D orth.	3D	3D	2D	3D	3D	3D + 2D
# Para.	0.06 M	0.71 M	19.07 M	0.67 M	12.63 M	14.6 M	5.34 M
Training Time	0.5 h	1.5 h	16.1 h	1.2 h	10.6 h	12.3 h	4.2 h

Table 4
The quantitative ablation evaluation on our ROSE model with TubeTK dataset. Note: Sens.: Sensitivity, Prec.: Precision, Jacc.: Jaccard. Best results are in bold and gray row is our full model result.

Methods/Metrics	Dice (%) ↑	Sens. (%) ↑	Prec. (%) ↑	Jacc. (%) ↑	PSNR ↑
ROSE (w/o SR)	60.33	54.34	68.10	43.21	–
ROSE (w/ SR + EnDeTopFuse)	64.13	58.72	70.87	47.20	37.59
ROSE (w/ SR + DeMultiFuse)	64.88	60.03	70.80	48.03	**38.59**
Full: ROSE (w/ SR + EnDeMultiFuse)	**65.59**	**60.70**	**71.46**	**48.79**	37.80

Upon this baseline, we then add the encoding feature sharing and the joint single-level feature fusion between two streams at the top level (i.e., Lvl 0) of decoder in the segmentation stream, i.e., ROSE (w/ SR + EnDeTopFuse). These encoder modules induce our compound 2D SR stream to form the segmentation-driven SR features, and the SR restored vessels under this setting appear to have higher contrast against the background vicinity in our observation. With the encoding feature sharing, the segmentation performance improves a lot with only the top level of the shallow decoder fusion. This setting demonstrates the oriented feature fusion in the decoder strengthens the vessel signals intentionally for the segmentation purpose.

We then evaluate the effectiveness of the pure multi-level (i.e., Lvl 0, Lvl 1, and Lvl 2) oriented feature fusion at the decoder stage, i.e., ROSE (w/ SR + DeMultiFuse). In this setting, we disable the encoding feature sharing but keep the multi-level oriented decoding feature fusion, which is to fuse SR-oriented features into the segmentation stream at multi-level asymmetric 3D segmentation decoder stages. By doing so, the SR features can be effectively and orientedly integrated into the segmentation decoding right after the deepest bottleneck, enabling thorough SR feature supplement and enhancement to extract the vessel masks with multi-level vessel semantic and content features. Table 4 shows that this setting also outperforms the baseline at a great margin.

Our full network design, i.e., ROSE (w/ SR + EnDeMultiFuse), incorporates all the aforementioned modules, and the final 3D vessel segmentation performance is further improved on all segmentation metrics according to Table 4. One may notice that in this setting, we slightly sacrifice the SR performance for obtaining a better segmentation result. This small decrease on the SR metric (i.e., PSNR) may be the consequence from our encoding feature sharing, when compared with the previous ROSE (w/ SR + DeMultiFuse). As we discussed above, the segmentation-driven SR involves the functionally enhanced vessel signals, which may affect the overall image structure/content and thus undermine the SR numeric performance. It is necessary to emphasize that our ultimate goal is to learn the effective SR-oriented features to enhance the vessel segmentation, instead of optimizing the SR image performance measured by pure SR metrics, and our experiments already show the SR images generated solely by SR-metrics oriented optimization do not necessarily benefit a better segmentation.

5. Conclusion and discussion

In this work, we have proposed ROSE, a deep neural network to extract and visualize 3D super-resolution microvascular structure from the next-generation microvascular images. ROSE has two major components, i.e., 3D and 2D dual-domain multi-task streams and bi-directional joint learning for SR-oriented semantics. By developing the encoding feature sharing, 3D-to −2D operations, and multi-level oriented feature fusion for joint learning, the proposed framework can strengthen the 3D microvasculature representation by better capturing the small/micro vessels with complicated geometry and topology variations in the SR scale, which outperforms the state-of-the-art classical and deep learning-based methods. The SR intrinsically trained along with vessel segmentation can effectively orient and enhance the feature representation for microvasculature extraction. In medical practice, this work can be used as the key functions for *in-vivo* precise microvascular visualization in an MR angiographic and venographic diagnosis of microvascular disease, e.g., using faster (lower-resolution) scanning to generate the higher-resolution and better-quality MR imaging-semantic analytics.

Currently, this work is done in a supervised learning style. It is expensive to acquire the ground truth labels for microvasculatures. Although we have used the method in [38] to automatically generate micro vessel labels with decent coverage and accuracy, it still requires domain experts' intervention to refine the vessel labels to obtain the 'ground truth'. On the other hand, the current dataset is relatively limited compared with some large open-source medical datasets, and it is also challenging to find publicly available microvasculature datasets from other institutions in order to further evaluate the generalization of the proposed framework. Moreover, current method extracts vessels with no region- / lesion-specific focus. With more lesion-specific data given or the disease-specific methods developed, the performance will be improved by focusing more on ROIs for clinical purposes. It is necessary to mention that our ROSE network architecture supports an even larger scaling factor (e.g., 4 or 8). However, in our current microvasculature extraction experiments, it is better to use the spatial downsampling factor of 2 for each dimension in 3D to achieve real clinical values; otherwise LR images will miss too many micro vessels (diameter of a micro vessel is 1 or 2 voxels).

In the future, we will apply our method to some other microvasculature datasets besides micro-cerebrovascular images, and continue to explore research problems related to "SR-oriented semantic embedding computation" supported 3D exploration and analysis. We will extend the current joint multi-level hybrid embedding learning into other applications, such as VR/AR/MR and robotics.

CRediT authorship contribution statement

Yifan Wang: Conceptualization, Methodology, Validation, Software, Visualization, Writing – original draft, Writing – review & editing. **Haikuan Zhu:** Software, Visualization. **Hongbo Li:** Validation. **Guoli Yan:** Conceptualization. **Sagar Buch:** Data curation, Investigation, Resources, Validation, Writing – review & editing. **Ying Wang:** Conceptualization, Validation. **Ewart Mark Haacke:** Conceptualization, Data curation, Investigation, Methodology, Validation, Writing – review & editing. **Jing Hua:** Conceptualization, Funding acquisition,

Methodology, Supervision, Writing – review & editing. **Zichun Zhong:** Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing.

Declaration of competing interest

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

Data availability

Data will be made available on request.

Acknowledgments

We would like to thank the reviewers for their valuable comments and suggestions. This work was partially supported by the National Science Foundation (NSF), USA under Grant Numbers OAC-1845962, OAC-1910469, and OAC-2311245.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.neucom.2024.128038.

References

[1] M. Mott, K. Pahigiannis, W. Koroshetz, Small blood vessels: Big health problems: National Institute of Neurological Disorders and stroke update, Stroke 45 (12) (2014) e257–e258.
[2] W. Brown, C. Thore, Cerebral microvascular pathology in ageing and neurodegeneration, Neuropathol. Appl. Neurobiol. 37 (1) (2011) 56–74.
[3] A. Dorr, B. Sahota, L. Chinta, M. Brown, A. Lai, K. Ma, C.A. Hawkes, J. McLaurin, B. Stefanovic, Amyloid-β-dependent compromise of microvascular structure and function in a model of Alzheimer's disease, Brain 135 (10) (2012) 3039–3050.
[4] A. Gouw, A. Seewann, W. Van Der Flier, F. Barkhof, A. Rozemuller, P. Scheltens, J. Geurts, Heterogeneity of small vessel disease: A systematic review of MRI and histopathology correlations, J. Neurol. Neurosurg. Psychiatry 82 (2) (2011) 126–135.
[5] C. Dong, C. Loy, K. He, X. Tang, Learning a deep convolutional network for image super-resolution, in: Proceedings of the European Conference on Computer Vision, 2014, pp. 184–199.
[6] C. Dong, C. Loy, X. Tang, Accelerating the super-resolution convolutional neural network, in: Proceedings of the European Conference on Computer Vision, 2016, pp. 391–407.
[7] W. Shi, J. Caballero, F. Huszár, J. Totz, A. Aitken, R. Bishop, D. Rueckert, Z. Wang, Real-time single image and video super-resolution using an efficient subpixel convolutional neural network, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 1874–1883.
[8] C. Ledig, L. Theis, F. Huszár, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, et al., Photo-realistic single image super-resolution using a generative adversarial network, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 4681–4690.
[9] X. Wang, K. Yu, S. Wu, J. Gu, Y. Liu, C. Dong, Y. Qiao, C. Change Loy, ESRGAN: Enhanced super-resolution generative adversarial networks, in: Proceedings of the European Conference on Computer Vision Workshops, 2018.
[10] M. Sajjadi, B. Scholkopf, M. Hirsch, EnhanceNet: Single image super-resolution through automated texture synthesis, in: Proceedings of the IEEE International Conference on Computer Vision, 2017, pp. 4491–4500.
[11] C. Ma, Y. Rao, Y. Cheng, C. Chen, J. Lu, J. Zhou, Structure-preserving super resolution with gradient guidance, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 7769–7778.
[12] Z. Zhou, Y. Hou, Q. Wang, G. Chen, J. Lu, Y. Tao, H. Lin, Volume upscaling with convolutional neural networks, in: Proceedings of the Computer Graphics International Conference, 2017, pp. 1–6.
[13] Y. Xie, E. Franz, M. Chu, N. Thuerey, tempoGAN: A temporally coherent, volumetric GAN for super-resolution fluid flow, ACM Trans. Graph. 37 (4) (2018) 1–15.
[14] J. Han, C. Wang, SSR-TVD: Spatial super-resolution for time-varying data analysis and visualization, IEEE Trans. Vis. Comput. Graphics (2020).
[15] J. Han, C. Wang, TSR-TVD: Temporal super-resolution for time-varying data analysis and visualization, IEEE Trans. Vis. Comput. Graphics 26 (1) (2019) 205–215.
[16] J. Han, H. Zheng, D.Z. Chen, C. Wang, STNet: An end-to-end generative framework for synthesizing spatiotemporal super-resolution volumes, IEEE Trans. Vis. Comput. Graphics 28 (1) (2021) 270–280.
[17] L. Guo, S. Ye, J. Han, H. Zheng, H. Gao, D.Z. Chen, J.-X. Wang, C. Wang, SSR-VFD: Spatial super-resolution for vector field data analysis and visualization, in: Proceedings of IEEE Pacific Visualization Symposium, 2020.
[18] J. Han, C. Wang, TSR-VFD: Generating temporal super-resolution for unsteady vector field data, Comput. Graph. (2022).
[19] S. Wurster, H. Guo, H.-W. Shen, T. Peterka, J. Xu, Deep hierarchical super resolution for scientific data, IEEE Trans. Vis. Comput. Graphics (2022).
[20] X. Zhao, Y. Zhang, T. Zhang, X. Zou, Channel splitting network for single MR image super-resolution, IEEE Trans. Image Process. 28 (11) (2019) 5649–5662.
[21] R. Tanno, D. Worrall, A. Ghosh, E. Kaden, S. Sotiropoulos, A. Criminisi, D. Alexander, Bayesian image quality transfer with CNNs: Exploring uncertainty in dMRI super-resolution, in: Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, 2017, pp. 611–619.
[22] K. Liu, Y. Ma, H. Xiong, Z. Yan, Z. Zhou, P. Fang, C. Liu, Medical image super-resolution method based on dense blended attention network, 2019, arXiv preprint arXiv:1905.05084.
[23] Y. Chen, Y. Xie, Z. Zhou, F. Shi, A. Christodoulou, D. Li, Brain MRI super resolution using 3D deep densely connected neural networks, in: Proceedings of the IEEE International Symposium on Biomedical Imaging, 2018, pp. 739–742.
[24] Y. Chen, F. Shi, A. Christodoulou, Y. Xie, Z. Zhou, D. Li, Efficient and accurate MRI super-resolution using a generative adversarial network and 3D multi-level densely connected network, in: Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, 2018, pp. 91–99.
[25] D. Mahapatra, B. Bozorgtabar, R. Garnavi, Image super-resolution using progressive generative adversarial networks for medical image analysis, Comput. Med. Imaging Graph. 71 (2019) 30–39.
[26] F. Dharejo, M. Zawish, F. Deeba, Y. Zhou, K. Dev, S. Khowaja, N. Qureshi, Multimodal-boost: Multimodal medical image super-resolution using multi-attention network with wavelet transform, IEEE/ACM Trans. Comput. Biol. Bioinform. (2022).
[27] F. Deeba, S. Kun, F. Ali Dharejo, Y. Zhou, Sparse representation based computed tomography images reconstruction by coupled dictionary learning algorithm, IET Image Process. 14 (11) (2020) 2365–2375.
[28] L. Wang, D. Li, Y. Zhu, L. Tian, Y. Shan, Dual super-resolution learning for semantic segmentation, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 3774–3783.
[29] Q. Delannoy, C.-H. Pham, C. Cazorla, C. Tor-Díez, G. Dollé, H. Meunier, N. Bednarek, R. Fablet, N. Passat, F. Rousseau, SegSRGAN: Super-resolution and segmentation using generative adversarial networks-application to neonatal brain MRI, Comput. Biol. Med. 120 (2020) 103755.
[30] H. Wang, L. Lin, H. Hu, Q. Chen, Y. Li, Y. Iwamoto, X.-H. Han, Y.-W. Chen, R. Tong, Patch-free 3D medical image segmentation driven by super-resolution technique and self-supervised guidance, in: Proceedings of the Medical Image Computing and Computer Assisted Intervention, 2021, pp. 131–141.
[31] H. Fu, Y. Xu, S. Lin, D. Wong, 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, 2016, pp. 132–139.
[32] J. Mo, L. Zhang, Multi-level deep supervised networks for retinal vessel segmentation, Int. J. Comput. Assist. Radiol. Surg. 12 (12) (2017) 2181–2193.
[33] P. Liskowski, K. Krawiec, Segmenting retinal blood vessels with deep neural networks, IEEE Trans. Med. Imaging 35 (11) (2016) 2369–2380.
[34] S. Shin, S. Lee, I. Yun, K. Lee, Deep vessel segmentation by learning graphical connectivity, Med. Image Anal. 58 (2019) 101556.
[35] P. Sanchesa, C. Meyer, V. Vigon, B. Naegel, Cerebrovascular network segmentation of MRA images with deep learning, in: Proceedings of IEEE International Symposium on Biomedical Imaging, 2019, pp. 768–771.
[36] G. Tetteh, V. Efremov, N. Forkert, M. Schneider, J. Kirschke, B. Weber, C. Zimmer, M. Piraud, B. Menze, DeepVesselNet: Vessel segmentation, centerline prediction, and bifurcation detection in 3-D angiographic volumes, 2018, arXiv preprint arXiv:1803.09340.
[37] T. Kitrungrotsakul, X.-H. Han, Y. Iwamoto, L. Lin, A. Foruzan, W. Xiong, Y.-W. Chen, VesselNet: A deep convolutional neural network with multi pathways for robust hepatic vessel segmentation, Comput. Med. Imaging Graph. 75 (2019) 74–83.
[38] Y. Wang, G. Yan, H. Zhu, S. Buch, Y. Wang, E. Haacke, J. Hua, Z. Zhong, VC-Net: Deep volume-composition networks for segmentation and visualization of highly sparse and noisy image data, IEEE Trans. Vis. Comput. Graphics 27 (2) (2021) 1301–1311.

[39] Y. Wang, G. Yan, H. Zhu, S. Buch, Y. Wang, E. Haacke, J. Hua, Z. Zhong, JointVesselNet: Joint volume-projection convolutional embedding networks for 3D cerebrovascular segmentation, in: Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, 2020, pp. 106–116.
[40] S. Banerjee, D. Toumpanakis, A. Dhara, J. Wikstrm, R. Strand, Topology-aware learning for volumetric cerebrovascular segmentation, in: Proceedings of the IEEE International Symposium on Biomedical Imaging, 2022, pp. 1–4.
[41] S. Pal, D. Toumpanakis, J. Wikstrm, C. Ahuja, R. Strand, A. Dhara, Multi-level residual dual attention network for major cerebral arteries segmentation in MRA towards diagnosis of cerebrovascular disorders, IEEE Trans. NanoBiosci. (2023).
[42] Z. Guo, Z. Tan, J. Feng, J. Zhou, 3D vascular segmentation supervised by 2D annotation of maximum intensity projection, IEEE Trans. Med. Imaging (2024).
[43] B. Lim, S. Son, H. Kim, S. Nah, K. Mu Lee, Enhanced deep residual networks for single image super-resolution, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2017, pp. 136–144.
[44] Y. Zhang, K. Li, K. Li, L. Wang, B. Zhong, Y. Fu, Image super-resolution using very deep residual channel attention networks, in: Proceedings of the European Conference on Computer Vision, 2018, pp. 286–301.
[45] Y. Fan, H. Shi, J. Yu, D. Liu, W. Han, H. Yu, Z. Wang, X. Wang, T.S. Huang, Balanced two-stage residual networks for image super-resolution, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2017, pp. 161–168.
[46] F. Kong, M. Li, S. Liu, D. Liu, J. He, Y. Bai, et al., Residual local feature network for efficient super-resolution, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2022, pp. 766–776.
[47] J. Kim, J.K. Lee, K.M. Lee, Deeply-recursive convolutional network for image super-resolution, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 1637–1645.
[48] Y. Tai, J. Yang, X. Liu, Image super-resolution via deep recursive residual network, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 3147–3155.
[49] K. Jiang, Z. Wang, P. Yi, J. Jiang, Hierarchical dense recursive network for image super-resolution, Pattern Recognit. 107 (2020) 107475.
[50] T. Lu, Y. Wang, Y. Zhang, J. Jiang, Z. Wang, Z. Xiong, Rethinking prior-guided face super-resolution: A new paradigm with facial component prior, IEEE Trans. Neural Netw. Learn. Syst. (2022).
[51] Y. Wang, T. Lu, Y. Zhang, Z. Wang, J. Jiang, Z. Xiong, FaceFormer: Aggregating global and local representation for face hallucination, IEEE Trans. Circuits Syst. Video Technol. (2022).
[52] Y. Wang, T. Lu, Y. Yao, Y. Zhang, Z. Xiong, Learning to hallucinate face in the dark, IEEE Trans. Multimed. (2023).
[53] Y. Xiao, X. Su, Q. Yuan, D. Liu, H. Shen, L. Zhang, Satellite video super-resolution via multiscale deformable convolution alignment and temporal grouping projection, IEEE Trans. Geosci. Remote Sens. 60 (2021) 1–19.
[54] P. Yi, Z. Wang, K. Jiang, J. Jiang, T. Lu, X. Tian, J. Ma, Omniscient video super-resolution, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021, pp. 4429–4438.
[55] F.A. Dharejo, F. Deeba, Y. Zhou, B. Das, M.A. Jatoi, M. Zawish, Y. Du, X. Wang, TWIST-GAN: Towards wavelet transform and transferred GAN for spatio-temporal single image super resolution, ACM Trans. Intell. Syst. Technol. 12 (6) (2021) 1–20.
[56] S. Weiss, M. Chu, N. Thuerey, R. Westermann, Volumetric isosurface rendering with deep learning-based super-resolution, IEEE Trans. Vis. Comput. Graphics 27 (6) (2019) 3064–3078.
[57] Z. Tang, B. Pan, E. Liu, X. Xu, T. Shi, Z. Shi, SRDA-Net: Super-resolution domain adaptation networks for semantic segmentation, 2020, arXiv preprint arXiv:2005.06382.
[58] Z. Hang, Q. Tingwei, H. Qing, L. Tian, C. Tingting, Z. Shaoqun, Super-resolution segmentation network for reconstruction of packed neurites, 2020, bioRxiv.
[59] Q. Li, B. Feng, L. Xie, P. Liang, H. Zhang, T. Wang, A cross-modality learning approach for vessel segmentation in retinal images, IEEE Trans. Med. Imaging 35 (1) (2015) 109–118.
[60] Ö. Çiçek, A. Abdulkadir, S. Lienkamp, T. Brox, 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, 2016, pp. 424–432.
[61] C. Szegedy, S. Ioffe, V. Vanhoucke, A. Alemi, Inception-v4, inception-resnet and the impact of residual connections on learning, in: Proceedings of the AAAI Conference on Artificial Intelligence, 2017.
[62] TubeTK MRA, URL https://public.kitware.com/Wiki/TubeTK/Data.
[63] Y. Shen, J. Hu, K. Eteer, Y. Chen, S. Buch, H. Alhourani, K. Shah, Q. Jiang, Y. Ge, E. Haacke, Detecting sub-voxel microvasculature with USPIO-enhanced susceptibility-weighted MRI at 7 T, Magn. Reson. Imaging 67 (2020) 90–100.
[64] H. Wang, Q. Jiang, Y. Shen, L. Zhang, E. Haacke, Y. Ge, S. Qi, J. Hu, The capability of detecting small vessels beyond the conventional MRI sensitivity using iron-based contrast agent enhanced susceptibility weighted imaging, NMR Biomed. (2020).
[65] S. Buch, Y. Chen, P. Jella, Y. Ge, E. Haacke, Vascular mapping of the human hippocampus using ferumoxytol-enhanced MRI, NeuroImage 250 (2022) 118957.
[66] J. Jiang, M. Dong, 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, 2005, pp. ii–465.
[67] M. Jenkinson, M. Pechaud, S. Smith, BET2: MR-based estimation of brain, skull and scalp surfaces, in: Proceedings of Annual Meeting of the Organization for Human Brain Mapping, 2005.
[68] A. Frangi, W. Niessen, K. Vincken, M. Viergever, Multiscale vessel enhancement filtering, in: Proceedings of International Conference on Medical Image Computing and Computer Assisted Intervention, 1998, pp. 130–137.
[69] M. Descoteaux, D. Collins, K. Siddiqi, A geometric flow for segmenting vasculature in proton-density weighted MRI, Med. Image Anal. 12 (4) (2008) 497–513.
[70] J. Dolz, C. Desrosiers, I.B. Ayed, 3D fully convolutional networks for subcortical segmentation in MRI: A large-scale study, NeuroImage 170 (2018) 456–470.

Yifan Wang is a Ph.D. candidate in Computer Science at Wayne State University. She received the Master's degree (2016) in Electrical Engineering at University of Minnesota, Twin cities, USA and the Bachelor's degree (2014) in Electrical Engineering at Zhengzhou University, China. Her research interests include deep learning, scientific visualization, computer vision, computer graphics, and medical image processing.

Haikuan Zhu is a Ph.D. candidate in Computer Science at Wayne State University. He received the Master's degree (2019) in Applied Mathematics at Xiamen University, China and the Bachelor's degree (2016) in Mathematics and Applied Mathematics at Beijing University of Chemical Technology, China. His research interests include computer graphics, image processing, and 3D printing.

Hongbo Li is a Ph.D. candidate in Computer Science at Wayne State University. He received the Master's degree (2021) in Computer Science at The University of Texas at Dallas, USA and the Bachelor's degree (2018) in Software Engineering at East China Normal University, China. His research interests include 3D geometric deep learning and computer graphics.

Guoli Yan is a Ph.D. candidate in Computer Science at Wayne State University. She received the Master's degree (2017) in Computer Science and Technology at Zhejiang Gongshang University, China and the Bachelor's degree (2014) in Computer Science and Technology at Shaoxing University, China. Her research interests include video and image processing, deep learning, and pattern recognition.

Sagar Buch is an Assistant Professor (Research) in the Department of Neurology at Wayne State University, School of Medicine. He received the Ph.D. degree (2015) and Master's degree (2012) in Biomedical Engineering at McMaster University, Canada, respectively. He was an MR Research Consultant in the Department of Radiology at Wayne State University from 2019 to 2022. His research interests focus on MRI physics, development and optimization of MR-based protocols for quantitative imaging and studying the cerebral microvasculature in neurodegenerative and neurovascular diseases.

Ying Wang is a Research Assistant in the Department of Radiology at Wayne State University. She received one Master's degree (2015) in Computer Science at Wayne State University, and the other Master's degree (2005) in Automatic Control at Northeastern University, China. She was a team lead at Neusoft Medical Inc. Her research interests include image registration, image segmentation, and quantitative analysis.

Ewart Mark Haacke is a Professor in the Department of Radiology at School of Medicine, Professor and Vice-Chairman in the Department of Biomedical Engineering at Wayne State University, and Director of the Perinatal MR Imaging Research Program. He received his Ph.D. degree (1978) and Master's degree (1975) in High Energy Physics, Bachelor's degree (1973) in Math and Physics from University of Toronto, Canada, respectively. He is an original pioneer of the MR angiographic imaging, fast imaging and cardiovascular imaging, and has developed new methods for imaging veins, micro-hemorrhaging and iron called Susceptibility Weighted Imaging (SWI).

Jing Hua is a Professor of Computer Science at Wayne State University. He received his Ph.D. degree (2004) in Computer Science from the State University of New York at Stony Brook, USA. His research interests lie in the area of visual computing including computer graphics and visualization, computer vision, image analysis and informatics, etc. He received the Gaheon Award for the Best Paper of International Journal of CAD/CAM in 2009, the Best Paper Award at ACM Solid Modeling 2004, and the Best Demo Awards at GENI Engineering Conference 21 (2014) and 23 (2015), respectively.

Zichun Zhong is an Associate Professor of Computer Science at Wayne State University. He received the Ph.D. degree (2014) in Computer Science at The University of Texas at Dallas, USA. He was a Postdoctoral Fellow (2014–2015) in Department of Radiation Oncology at UT Southwestern Medical Center at Dallas, USA. His research interests focus on 3D geometric modeling and computing for computer graphics, visualization, computer vision, deep learning, medical image processing, and GPU algorithms. He received the NSF CAREER Award in 2019, NSF CRII Award in 2017.