Ensemble of Deep Learning Cascades for Segmentation of Blood Vessels in Confocal Microscopy Images

Yang Yang Wang¹, O. V. Glinskii²,³, Filiz Bunyak¹, Kannappan Palaniappan¹
¹Department of Electrical Engineering and Computer Science
²Department of Medical Pharmacology and Physiology
³Dalton Cardiovascular Research Center
University of Missouri-Columbia, MO, USA
yw6p8@umsystem.edu; glinskiio@health.missouri.edu; {bunyak, pal}@missouri.edu

Abstract—Detection, segmentation, and quantification of microvascular structures are the main steps towards studying microvascular remodeling. Combined with appropriate staining, confocal microscopy imaging enables exploration of the full 3D anatomical characteristics of microvascular systems. Segmentation of confocal microscopy images is a challenging task due to complexity of anatomical structures, staining and imaging issues, and lack of annotated training data. In this paper, we propose a deep learning system for robust segmentation of cranial vasculature of mice in confocal microscopy images. The proposed system is an ensemble of two deep-learning cascades consisting of two coarse-to-fine subnetworks with skip connections in between. One cascade aims to improve sensitivity, while the other aims to improve precision of the segmentation results. Our experiments on mice cranial vasculature showed promising results achieving segmentation accuracy of 92.02% and dice score of 81.45% despite being trained on very limited confocal microscopy data.

Index Terms—vessel segmentation, confocal microscopy images, deep learning, semantic segmentation

I. INTRODUCTION

Serious intracranial pathologic conditions, such as dural sinus thrombosis, dural arteriovenous fistulas, and aneurysms, involve the vessels, not of the brain itself, but its outer fibrous membrane, the dura mater. These conditions, resulting in significant neurologic morbidity and reduced cognitive abilities [1], [2], have been associated with vascular abnormalities. Meningeal vascular networks contribute to brain metabolic clearance and venous blood outflow. They constantly adjust to tissue metabolic demands through structural and functional remodeling. Defective vascular remodeling under certain pathologic conditions leads to tissue damage and limits its repair. Acute damage to meningeal vasculature caused by traumatic brain injury, resulting in disruption of meningeal vascular integrity and peripheral immune response can lead to life-threatening situations [2]. While impaired vascular integrity, capillary rarefaction, and aberrant angio-architecture can also develop under chronic conditions, for example, associated with sex hormone deprivation [3].

Fig. 1: Sample blood microvascular structures imaged using confocal microscopy. The first three columns show sample single focus slices. The last column shows fused multi-focus image.

Detection, segmentation, and quantification of microvascular structures are the main steps towards studying microvascular remodeling. Confocal microscopy [4] allows 3D image capture using optical sectioning or depth discrimination by blocking light emitted from out-of-focus planes. Each single focus image captures the details of the specimen regions that lie close to its focal plane, while the remaining regions are imaged with poor contrast. Combined with appropriate staining, confocal microscopy imaging enables exploration of the full 3D anatomical characteristics of microvascular systems.

While segmentation of vessels on traditional angiogram-based imagery or retinal imagery has advanced considerably [5]–[12], segmentation of microvasculature in confocal microscopy images remains to be a challenging task. The main challenges are due to complexities of anatomical structures such as irregular shape and varying scale of the vessels; staining issues such as non-homogeneous staining within the lumen and excessive stain in the background due to leakage; imaging issues such as low contrast or background clutter due to out of focus structures. Figure 1 that shows sample slices...
and xy-plane projections for two sample confocal microscopy stacks illustrates some of these challenges. In addition to aforementioned challenges, unlike angiogram-based or retinal vessel imagery, manually annotated training data needed for supervised machine learning approaches is severely lacking for confocal microscopy images of microvasculature. This is due to the size of data (hundreds of slices per stack), complexity of the vascular structures, and difficulty of 3D annotation.

In this paper, we propose a segmentation system for the confocal microscopy image segmentation, which is a fusion of two deep-learning cascades. The two cascades focus on improving sensitivity (recall) and precision of the segmentation results respectively. To compensate for limited confocal microscopy training data, the proposed network is first trained with an epifluorescence microscopy image dataset, then fine-tuned with a small set of fused confocal microscopy images of mice cranial microvasculature.

II. Ensemble of Deep Learning Cascades for Vessel Segmentation

A. Image preprocessing

The proposed preprocessing scheme consists of two steps: (1) multi-focus image fusion, and (2) contrast enhancement. Each confocal microscopy image stack consists of hundreds of single focus slices capturing only a small portion of the microvascular network. We use the multi-focus image fusion approach described in [13] to produce a single multi-focus image out of hundred of single-focus slices within a confocal microscopy stack. The resulting multi-focus image (as the original set of slices) typically suffers from staining issues (i.e. non-homogeneous staining within the lumen and excessive stain in the background due to leakage), imaging issues (i.e. low contrast), and background clutter due to out of focus structures. To improve image contrast, we apply adaptive histogram equalization [14] to the fused multi-focus image. Sample fused, multi-focus, confocal microscopy images before and after adaptive histogram equalization are shown in Figure 2.

B. Network Architecture

For robust and precise segmentation of microvascular structures on confocal microscopy images we have developed two deep learning network cascades: (1) deep binary attention cascade (DBAC), and (2) deep distance map attention cascade (DDMAC). The DBAC network is designed for improving sensitivity (recall) scores, while the DDMAC network is designed for improving precision scores of microvascular image segmentation results. Each cascade generates two outputs: an intermediate output from the first subnetwork acting as attention map for the second network and a final refined segmentation mask from the second subnetwork. The two final predictions from the DBAC and DDMAC networks are fused together to produce one final segmentation mask.

Fig. 2: Sample confocal microscopy fused multi-focus images before (first column) and after (second column) adaptive histogram equalization.

1) Deep binary attention cascade (DBAC): The cascade network involves two subnetworks: a classical semantic segmentation network, UNet [15], followed by a deeper U-Net++ [16] designed to exploit the convolutional features at different scales. The architecture of the proposed DBAC is shown in Figure 3. The convolutional layers have channels 32, 64, 128, 256, and 512 respectively from shallow to deeper encoder layers in both U-Net and U-Net++. The decoder layers have symmetric number of channels. Each convolutional layer is replaced with a residual block [17].

The first subnetwork takes the input image and learns to predict a soft attention map where higher values indicate likelihood of vessel presence. The second subnetwork takes the input image and is guided by the soft attention map to perform coarse-to-fine refinement of the vessel segmentation mask from the first subnetwork.

The proposed attention module connecting the two subnetworks is defined as:

\[ P_{ij} = \max(0, W_{ij}) \]  
\[ M = I \times P \]

where \( W \) is the convolutional feature map from the first subnetwork, \( P \) is the 1-channel attention map prediction from the first subnetwork trained with a binary mask, where the positive values represent foreground and the negative values represent background. \( I \) is the input image, and \( M \) is the input to the second subnetwork.
We further extend this cascade by adding two shortcut connections that feed-forward the first subnetwork at different levels of the decoder to the corresponding levels of the decoder in the second subnetwork. The forwarded feature map and target feature map are concatenated together to prevent gradient vanishing as well as to directly forward feature maps from the first network to the second network.

The proposed network is trained on 60 epifluorescence microscopy images with resolution of 1360 × 1036 pixels [7] [18], using binary cross-entropy loss [19] function and stochastic gradient descent (SGD) optimizer for the first and second subnetworks. Data augmentation is applied during the training and fine-tuning processes including random cropping (crop size is 448 × 448), rotation, flipping, scale, brightness, and contrast adjustment.

2) Deep distance map attention cascade (DDMAC): As can be observed from Figure 1, cranial vasculature’s diameter varies largely within even a small field of view. To better capture the varying size and shape of the vessels and to be able to detect thin vessels, we propose deep distance map attention cascade (DDMAC) shown in Figure 4. Dilated probability regression module (first subnetwork) aims to improve the precision score of the segmentation by applying multi-channels attention (the orange map in Figure 4) to the input image and then concatenating it with the dilated probability map \( Q \) (the green map in Figure 4).

Convolutional feature map \( Q \) is a 1-channel prediction from the first subnetwork of DDMAC and is trained with dilated probability map \( D \), which is defined as:

\[
    D = distance\_transform(1 - V) 
\]

\[
    D = \{k|D > k\} 
\]

\[
    D = 1 - \frac{D}{\max(D)} 
\]
where $V$ represents the binary mask (vessel is positive, background is negative) used for training the proposed system. $k$ is the upper bound of the pixel distance in $D$, which is set to be 20 in this study. The dilated probability regression module is defined as:

$$W = \frac{e^{W_{ijk}}}{\sum_{k=1}^{3} e^{W_{ijk}}}$$

(6)

$$W' = W \times I$$

(7)

$$M' = \text{concatenate}(W', Q)$$

(8)

where $W_{ijk}$ is the convolutional feature map with size $i \times j \times k$ representing height, width, and channels respectively. In this study, $i, j$ are the same as the height and width of the input image, and $k = 3$. The output $M'$ is the input to the second subnetwork. The second subnetwork takes $M'$ and performs coarse-to-fine refinement of the vessel segmentation mask $Q$ from the first subnetwork.

The proposed DDMAC network is trained on 60 epifluorescence microscopy images with resolution $1360 \times 1036$ [7] [18] using the same data augmentation strategies as the described DBAC network. Mean square error (MSE) [19] and binary cross-entropy loss functions are used to train the first and the second subnetworks respectively. SGD is used as training optimizer.

C. Decision fusion (late classifier fusion)

For robust vessel segmentation performance, outputs of the two proposed networks DBAC and DDMAC are fused. Two decision fusion (late classifier fusion) mechanisms, average and maximum are considered:

$$\text{Average} : S_{ij} = \frac{L_{ij} + N_{ij}}{2}$$

(9)

$$\text{Maximum} : S_{ij} = \max(L_{ij}, N_{ij})$$

(10)

where $L$ and $N$ denote the 1-channel segmentation probability maps generated by the proposed DBAC and DDMAC networks respectively, and $ij$ refers to pixel coordinates. Binary segmentation masks are produced by performing hysteresis thresholding [20] on the fused probability maps $S$. Mathematical morphology operations are applied to the predicted binary mask to fill small gaps and to remove small fragments.

III. EXPERIMENTAL RESULTS

A. Data collection

In this study, 7-8 weeks old C57BL/6J female mice were used to generate 80 3D confocal microscopy image stacks. Following sacrifice, the entire body of the mouse was perfused through the heart with Kreb’s/albunin solution containing AlexaFluor 594-conjugated soybean agglutinin (SBA) lectin to stain and identify blood microvessels [21]. Skull caps with dura mater were isolated and fixed in 10% formaldehyde. Images were acquired at 20x magnification on confocal Fluoview FV1000 inverted microscope system (Olympus). The 160-270 mm thick Z-stacks were acquired with step size 1 mm. All animal experimental procedures were approved by the University of Missouri Institutional Animal Care and Use Committee.

The single focus Z-stacks microscopy images were fused using the method presented in [13] to produce eighty $512 \times 512$ multi-focus images. Out of those multi-focus images, 62 images were selected for training and 18 images were selected for testing. Silver truth (ST) segmentation masks were generated for all the training & test images by a combination of computer-generated segmentation masks and manual annotation. More precise ground truth (GT) segmentation masks were generated for the 18 test images by further inspection and manual correction by a domain expert.

B. Evaluation Metrics

Segmentation evaluations were carried out by sensitivity (recall), precision, specificity, accuracy, and dice score measures as defined below:

$$\text{Sensitivity} = \frac{TP}{TP + FN}$$

(11)

$$\text{Precision} = \frac{TP}{TP + FP}$$

(12)

$$\text{Specificity} = \frac{TN}{TN + FP}$$

(13)

$$\text{Accuracy} = \frac{TP + TN}{FN + FP + TP + TN}$$

(14)

$$\text{Dice} = \frac{2 \times TP}{2 \times TP + FP + FN}$$

(15)

where $TP, TN, FP, FN$ represent number of true positive, true negative, false positive, and false negative pixels respectively.

C. Network inference on 2D multi-focus images

Because of limited amount of confocal microscopy training data, the proposed networks and their subnetworks were first trained with 60 epifluorescence microscopy images and their ground truth segmentation masks described in [7], [18]. Then the same networks were fine-tuned with 62 multi-focus confocal microscopy images and their associated silver truth masks described above. Both set of networks (with and without fine-tuning) were tested on the 18 multi-focus confocal microscopy test images and evaluated using corresponding ground truth segmentation masks.
TABLE II: Single network segmentation performances.

<table>
<thead>
<tr>
<th>Methods</th>
<th>Image size</th>
<th>Sensitivity</th>
<th>Precision</th>
<th>Specificity</th>
<th>Accuracy</th>
<th>Dice</th>
</tr>
</thead>
<tbody>
<tr>
<td>U-Net++ [16]</td>
<td>512 × 512</td>
<td>80.87</td>
<td>74.23</td>
<td>92.67</td>
<td>89.86</td>
<td>75.88</td>
</tr>
<tr>
<td>deeper U-Net++</td>
<td>512 × 512</td>
<td>84.31</td>
<td>67.72</td>
<td>91.05</td>
<td>89.44</td>
<td>75.91</td>
</tr>
<tr>
<td>proposed DBAC</td>
<td>1024 × 1024</td>
<td>87.73</td>
<td>66.78</td>
<td>88.39</td>
<td>88.00</td>
<td>74.22</td>
</tr>
<tr>
<td>proposed DDMAC</td>
<td>512 × 512</td>
<td>82.14</td>
<td>75.02</td>
<td>92.08</td>
<td>89.80</td>
<td>76.86</td>
</tr>
</tbody>
</table>

TABLE III: Ensemble network segmentation performances.

<table>
<thead>
<tr>
<th>DBAC with input size</th>
<th>DDMAC with input size</th>
<th>Fusion Mechanism</th>
<th>Sensitivity</th>
<th>Precision</th>
<th>Specificity</th>
<th>Accuracy</th>
<th>Dice</th>
</tr>
</thead>
<tbody>
<tr>
<td>512 × 512</td>
<td>512 × 512</td>
<td>maximum</td>
<td>86.87</td>
<td>73.61</td>
<td>91.47</td>
<td>90.44</td>
<td>79.69</td>
</tr>
<tr>
<td>512 × 512</td>
<td>512 × 512</td>
<td>average</td>
<td>81.95</td>
<td>79.87</td>
<td>94.41</td>
<td>91.67</td>
<td>80.90</td>
</tr>
<tr>
<td>1024 × 1024</td>
<td>512 × 512</td>
<td>maximum</td>
<td>86.30</td>
<td>74.63</td>
<td>92.08</td>
<td>90.75</td>
<td>80.04</td>
</tr>
<tr>
<td>1024 × 1024</td>
<td>512 × 512</td>
<td>average</td>
<td>80.60</td>
<td>82.32</td>
<td>95.26</td>
<td>92.02</td>
<td>81.45</td>
</tr>
</tbody>
</table>

1) Single network segmentation performances: First, we compare segmentation performances of the proposed DBAC and DDMAC networks with state-of-the-art segmentation network U-Net++ [16] and its deeper version with number of trainable parameters comparable to the proposed networks. The total number of trainable parameters of these networks are listed in Table I.

The 1-channel probability maps outputted from the proposed DBAC and DDMAC networks are binarized using hysteresis thresholding [20] (lower bound=0.45, higher bound=0.95). Segmentation performances of the compared networks with and without fine-tuning are listed in Table II. In order to better preserve the very thin vessels, upsampled networks with and without fine-tuning are listed in Table II. The distance based DDMAC network is more robust to scale variations.

As we can see in Table II, fine-tuning improves the dice score of all the deep learning networks by at least 1.4%. The proposed DBAC and DDMAC networks achieve the best sensitivity (recall) and precision scores respectively with and without fine-tuning. The distance based DDMAC network results in the best dice scores with and without fine-tuning.

2) Ensemble network segmentation performances: Table III summarizes segmentation performances of different configurations of the proposed ensemble network consisting of DBAC and DDMAC networks. The table explores two different input image sizes and two different decision fusion (late classifier fusion) mechanisms, average and maximum. All configurations of the proposed ensemble network outperform the best single network. Upsampling input images increases the precision scores. Ensemble network using average fusion and upsampled inputs for the proposed DBAC network improves the dice score of the best performing single network by 2.29% reaching the best dice score of 81.45%.

Figure 5 shows single network and ensemble network segmentation results for two sample multi-focus images. Red and blue pixels represent false-positive and false-negative predictions respectively. False detections (false-positives) are typically caused by adverse effects of adaptive histogram equalization. Missed detections (false-negatives) are caused by thin vessels and low contrast between microvasculature and background. Fusing the outputs of the DBAC and DDMAC networks lead to recovery of some missed vessels and removal of some spurious detections (Figure 5f).

D. Network inference on 3D image stacks

The 3D confocal microscopy image stacks in this study contain hundreds of slices. Each single slice captures the details of the specimen regions that lie close to its focal plane, while the remaining regions are imaged with poor contrast. Segmentation and visualization of the 3D image stack allow comprehensive visualization and quantification of the anatomical structure of the 3D microvasculature. 3D deep learning networks could be employed to capture and learn the full 3D morphological and anatomical characteristics of the microvasculature. However, demanding computational requirements and more importantly, the need for 3D annotation for training limit their usability. In this section, we utilize the proposed 2D image segmentation network to independently segment the single focus Z-stacks of a 3D confocal microscopy volume.

Figure 6 shows 2D multi-focus images, corresponding 2D segmentation masks, and 3D segmentation masks for two sample confocal microscopy volumes. The 2D multi-focus images were obtained using the multi-focus fusion method described in [13]. 3D segmentation masks were obtained by applying the proposed 2D segmentation network to the individual single focus images forming the confocal microscopy volume. Each input slice has been preprocessed with linear contrast enhancement. Visualization of the 3D segmentation results were generated using the Chimera software [22]. Figure
Fig. 5: Intermediate and final segmentation results for two sample multi-focus input images. (a) input images, (b-c) ground truth maps for the binary attention module in DBAC and dilated probability module in DDMAC, (d-f) predicted segmentation masks versus ground truth for DBAC, DDMAC, and ensemble networks. The red, blue, and white regions represent false-positive, false-negative, and true positive predictions respectively.

Fig. 6: Sample outputs from the proposed system. (a) multi-focus confocal microscopy image enhanced with adaptive histogram equalization; (b) predicted segmentation mask versus ground truth where the red, blue, and white regions represent false-positive, false-negative, and true positive predictions respectively. (c-d) 3D segmentation masks obtained by applying the proposed 2D segmentation network to the individual single focus images forming the confocal microscopy volume. Visualization of the 3D segmentation results were generated using the Chimera software [22]. Color fades with increasing depth.

6c and 6d show promising 3D segmentation results. In the first row of Figure 6, missed detections in 2D (Figure 6b, blue pixels within green rectangles) caused by low contrast are recovered in 3D (Figure 6c and 6d) thanks to linear contrast enhancement. In the second row of Figure 6, missed detections in 3D (Figure 6c, green rectangle) caused by lack of semantic context in each slice can be corrected by the segmentation result of the corresponding 2D multi-focus image (second row...
in Figure 6b).

IV. Conclusion

In this paper, we presented an ensemble of deep learning cascades for robust segmentation of blood vessels in confocal microscopy images. The proposed ensemble is composed of two complementary deep-learning cascades aiming to improve sensitivity and precision of the segmentation results. The proposed cascades first learn to predict two soft attention maps, one based on binary pixel classification, the other based on regression to a distance map. The attention maps guide the networks to predict an accurate vessel segmentation mask. Experiments demonstrated promising results towards segmentation of microvascular structures in both 2D and 3D datasets. Segmentation is the key first step towards objective, and quantitative analysis of microvascular systems. The proposed segmentation system will be used to study microvascular remodeling.

Acknowledgement

This work is partially supported by awards from U.S.NIH National Institute of Neurological Disorders and Stroke R01NS110915. Any opinions, findings, and conclusions or recommendations expressed in this publication are those of the authors and do not necessarily reflect the views of the U.S. Government or agency thereof.

References


