STRUCTURAL PRIOR MODELS FOR 3-D DEEP VESSEL SEGMENTATION

Xuelu Li[*], *Raja Bala*[†], *Vishal Monga*[*]

[*] Pennsylvania State University, University Park, PA, [†]Amazon Research, Rochester, New York, USA

ABSTRACT

We address the problem of 3-D blood vessel segmentation with a deep learning method that incorporates domain information via priors and regularizers on vessel structure and morphology. Inspired by the observation that 3-D vessel structures project onto 2-D image slices with distinctive edges that can aid 3-D vessel segmentation, we propose a novel multi-task learning architecture comprising a shared encoder and two decoders that respectively predict vessel segmentation maps and edge profiles. 3-D features from the two branches are concatenated to facilitate edge-guidance when learning segmentation maps. We introduce new regularization terms that encourage local homogeneity of 3-D blood vessel volumes brought about by biomarkers, as well as sparsity of edge pixels. Experiments on benchmark datasets demonstrate superior performance of our method over the state-of-the-art, especially when training data is limited.

Index Terms— 3-D vessel segmentation, edge profiles, local homogeneity, deep learning, regularization

1. INTRODUCTION

Blood vessel segmentation plays a fundamental role in the accurate diagnoses of vascular-related diseases in diverse clinical applications including laryngology, oncology [1], ophthalmology [2], and neurosurgery [3].There has been recent interest to develop automatic segmentation techniques for 3-D vessel scans. 3-D analysis presents both opportunities and challenges. On the one hand, 3-D representations are appealing as they afford access to rich and intricate blood vessel geometries with 3-D context that is not explicitly available in 2-D projections. On the other hand, the dearth of datasets with 3-D voxel-level annotations, and the lack of pre-trained models make accurate 3D segmentation difficult.

Recently, deep learning has demonstrated state of the art results methods for 3-D vessel segmentation methods. Approaches can be classified into two broad categories. The first represents the 3-D volume as a series of 2-D slices and trains 2-D networks to reason on each slice either individually or sequentially [4, 5]. A major benefit of such an approach is that many architectures and pretrained models built for natural scenes can be leveraged to bootstrap feature learning.
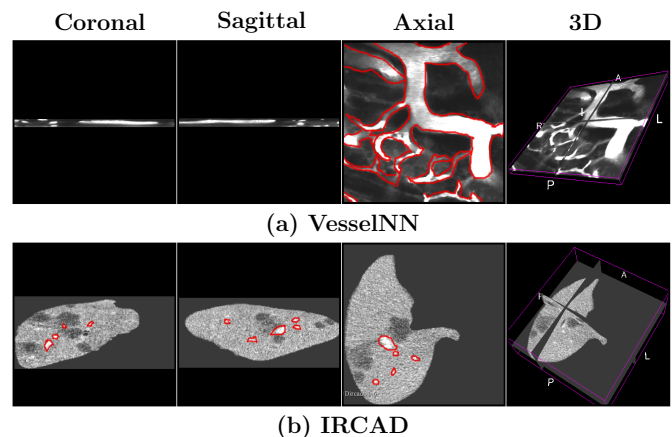
Fig. 1: Samples from VesselNN and IRCAD datasets. Top row visualizes orthogonal intersections of three selected slices. Next three rows show projections along axial, sagittal, coronal dimensions. Vessel structures are marked by red boundaries.

However, 2-D networks do not have access to 3-D context that conveys inter-slice correlations. To address this, recurrent network modules have been incorporated with promising results. For example, in [6], the authors use a recurrent neural network for exploiting the intra-slice and inter-slice contexts respectively. In [7], a novel convolutional LSTM neural network is designed to derive 3-D volumes from 4-D spatiotemporal data, and in [8], a tree-structural LSTM combined with a multi-layer perceptron encoder network is proposed to intrinsically embed the information from the upstream and downstream. The second flavor of vessel segmentation methods employ 3-D networks trained directly on voxel data. This enables utilization of full 3-D context during training and inference [9, 10]. However, the lack of 3-D annotated training datasets and pretrained models, and high computational cost for training and inference, limits deployment in practical applications [11].

Our approach belongs to the first category, processing a sequence of 2-D slices, while mining 3-D context between adjacent slices with recurrent modules. We achieve enhanced accuracy by utilizing priors on the structure and morphology of 3-D vessels. In particular, our contributions are: 1) we observe in Fig. 1 that vessel structures projected onto slices exhibit shapes with distinctive boundaries. We thus develop a learning scheme that predicts vessel regions with specific attention to edges; 2) additionally, we observe that biomark-

ers used during imaging results in highlighted blood regions and locally homogeneous vessel structures. Inspired by [12], we construct a brand new local homogeneity loss using the 3-D input volumes and the 3-D segmentation probability maps, which is also able to regularize the performance of the edge branch, and highly beneficial in practice. Our approach is dubbed Structural Prior Guided Deep 3-D Vessel Segmentation (SPG-3DVS). We conduct experiments on two well-known vessel segmentation datasets. Ablation studies and comparisons with the state-of-the-art methods are presented. The code the proposed method is made available online[1].

2. PROPOSED METHOD

2.1. Network Structure

Our network structure, shown in Fig. 2, comprises a main segmentation branch and an auxiliary edge profile branch. Both branches are constructed with 2-D feature extraction layers combined with a recurrent network that captures inter-slice context and correlations. The input to the network is an image volume $\boldsymbol{X} \in \mathbb{R}^{x \times y \times n} = \{\boldsymbol{X}_1, \cdots, \boldsymbol{X}_i, \cdots, \boldsymbol{X}_n\}$, where $\boldsymbol{X}_i \in \mathbb{R}^{x \times y}$ are 2-D slices.

Segmentation Branch: This branch predicts vessel segmentation maps from 2-D slices. The 2-D feature extraction network comprises a U-Net backbone with DenseNet blocks embedded. A given layer between each dense block in the dense encoder is skip-connected with corresponding layers between the dense blocks in the dense decoder. The parameters of the layers in the encoder and the decoder are represented respectively as Θ_E and Θ_D^{Msk}. Dense blocks are collections of densely connected layers, proven to be effective for natural image classification [13]. By embedding these blocks in the U-net structure, we inherit pretrained features, thus avoiding over-fitting when training data is limited. Each sub-slice sequence of $\boldsymbol{X}$ is passed to the encoder which extracts latent feature representations to be sent to the decoder. The output of the latter is passed to the recurrent module — a bidirectional convolutional long short term memory (BCLSTM) network [14], with parameters Θ_C^{Msk}. The BCLSTM is designed to utilize information from previous and subsequent slices in an image sequence, and has shown promising results in other medical imaging applications [7]. In our work, concatenated features of previous, current, and future hidden states are used as inputs to subsequent convolutional neural network layers in the BCLSTM, thus enabling predictions that are coherent across slices. The output of BCLSTM is a feature map corresponding to the current input slice.

Edge Profile Branch: This branch estimates the edge profiles of the vessels in each slice, and comprises a 2-D feature extraction decoder and a BCLSTM module. The edge branch shares the same encoder as the segmentation branch, thereby

both reducing the number of training parameters and coupling the learning of edges and vessel maps in a mutually reinforcing manner. The output of the dense decoder, parameterized by Θ_D^{Ed}, is sent to BCLSTM, parameterized by Θ_C^{Ed}, ensuring that the learned edge feature maps incorporate context from neighboring slices.

2.2. Concatenation of Edge and Segmentation Features

We concatenate features from edge and segmentation branches (indicated by the yellow box "C" in Fig. 2) to let the edge features guide the main segmentation task. In order to strengthen the relation between the estimated slice and adjacent slices, the hidden state features of the previous, current, and subsequent slice from both BCLSTMs are concatenated. Finally, the concatenated feature maps pass through a convolutional block to obtain the output segmentation mask. Thus effectively, we incorporate edge information as a 3-D prior for segmentation prediction, while capturing inter-slice correlations by regularizing the segmentation predictions of adjacent slices. Given an input sequence $\{\boldsymbol{X}_{t-T}, \cdots, \boldsymbol{X}_t, \cdots, \boldsymbol{X}_{t+T}\}$, the data loss term L_{seg} used to train the basic model is:

$$L_{seg} = L_{BCE}(\boldsymbol{P}_t^{Ed}, \boldsymbol{Y}_t^{Ed}) + \sum_i^{t-1,t,t+1} L_{BCE}(\boldsymbol{P}_i^{Msk}, \boldsymbol{Y}_i^{Msk}) \quad (1)$$

where t represents the index of the center slice in the input sequence, $\boldsymbol{P}_t^{Ed} \in [0,1]^{x \times y}$ is the predicted edge probability map of the center slice and $\boldsymbol{P}_i^{Msk} \in [0,1]^{x \times y}$, where $i \in \{t-1, t, t+1\}$ is the 3-D segmentation probability map obtained using the concatenated hidden state features. $\boldsymbol{Y}_t^{Ed} \in \{0,1\}^{x \times y}$ is the ground truth mask of the edge profile of tth slice and $\boldsymbol{Y}_i^{Msk} \in \{0,1\}^{x \times y}$ is the ground truth segmentation mask of the ith slice. L_{BCE} represents binary cross-entropy loss.

2.3. Structural Prior Losses

We incorporate domain priors to enhance robustness of the network to small training volumes. We exploit the fact that biomarkers that are frequently injected into blood vessels during imaging serve to enhance vessel contrast for diagnostics. This results in regions interior to the blood vessel being locally homogeneous. To exploit this characteristic, we propose a novel loss L_{spg} as shown in Eq. 2 that 1) encourages local homogeneity of 3-D vessel volumes and greater contrast between vessel and non-vessel regions 2) promotes the sparsity of the edge branch predictions:

$$L_{spg} = \underbrace{\sum_{u,v,z} \boldsymbol{P}^{Msk} \circ \left| \boldsymbol{V} - \frac{\sum_{u,v,z}(\boldsymbol{P}^{Msk} \circ \boldsymbol{V})}{\sum_{u,v,z} \boldsymbol{P}^{Msk}} \right|}_{\text{homogeneity within vessels}}$$
$$+\alpha \underbrace{\sum_{u,v,z}(1-\boldsymbol{P}^{Msk}) \circ \left| (\boldsymbol{P}^{Msk})^\sigma \circ \boldsymbol{V} - \frac{\sum_{u,v,z}\left((1-\boldsymbol{P}^{Msk}) \circ (\boldsymbol{P}^{Msk})^\sigma \circ \boldsymbol{V}\right)}{\sum_{u,v,z}(1-\boldsymbol{P}^{Msk})} \right|}_{\text{homogeneity of non-vessel background}} + \beta ||\boldsymbol{P}_t^{Ed}||_1$$
$$(2)$$
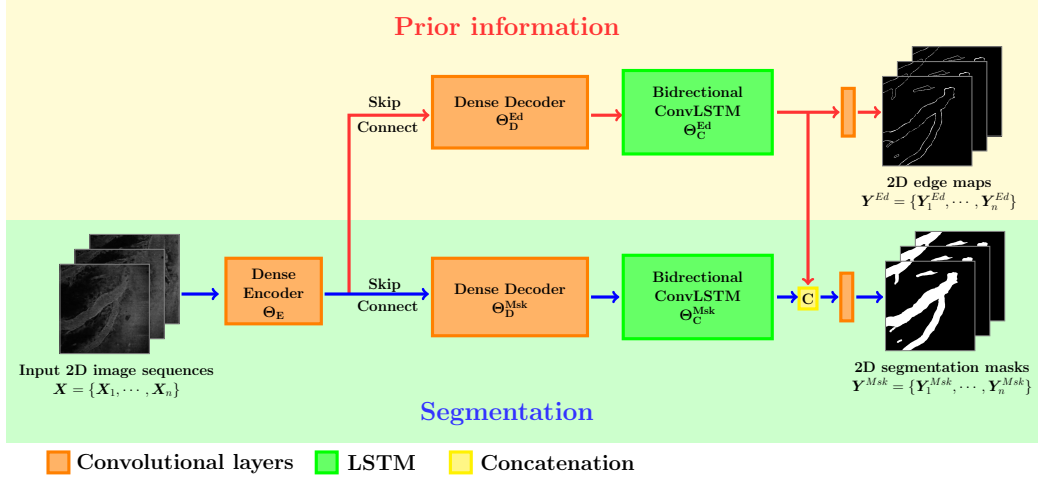
Fig. 2: SPG-3DVS Architecture: Features from a shared dense encoder are used for both segmentation and edge prediction tasks. The encoder is trained jointly with network components for both edge and segmentation tasks, so as to facilitate cross-task reinforcement.

where $\boldsymbol{V} = [\boldsymbol{X}_{t-1}, \boldsymbol{X}_t, \boldsymbol{X}_{t+1}] \in \mathbb{R}^{x \times y \times 3}$ is the input volume constructed by three adjacent slices $\boldsymbol{X}_{t-1}$, $\boldsymbol{X}_t$, and $\boldsymbol{X}_{t+1}$. u, v, z are the indexes of a feature point in three dimensions. $\boldsymbol{P}^{Msk} = [\boldsymbol{P}_{t-1}^{Msk}, \boldsymbol{P}_t^{Msk}, \boldsymbol{P}_{t+1}^{Msk}] \in \mathbb{R}^{x \times y \times 3}$ is the predicted segmentation probability map. Hyperparameter α is used to balance the effect of the regularizer within and outside of the vessels. Parameter σ is chosen based on the expected homogeneity of non-vessel regions in a particular dataset. It is set to 0 when non-vessel regions are expected to be homogeneous, and 1 when there are tissue structures outside of the vessels. Since $\boldsymbol{P}^{Msk}$ is the predicted probability mask map, it is assumed that the corresponding pixel values within the vessels in the input volume $\boldsymbol{V}$ are close to 1 while non-vessel pixel values will be close to 0. The terms $\frac{\sum_{u,v,z}(\boldsymbol{P}^{Msk} \circ \boldsymbol{V})}{\sum_{i,j,k} \boldsymbol{P}^{Msk}}$ and $\frac{\sum_{i,j,k}\left((1-\boldsymbol{P}^{Msk}) \circ (\boldsymbol{P}^{Msk})^\sigma \circ \boldsymbol{V}\right)}{\sum_{u,v,z}(1-\boldsymbol{P})}$ denote average pixel values within and outside the vessels respectively. $|\cdot|$ represents the absolute value operator on each element in the matrix. The first two loss terms achieve their lowest value if and only if vessel pixel values are equal to 1, and the non-vessel pixels are 0. The third loss term, $\boldsymbol{P}_t^{Ed} \in \mathbb{R}^{x \times y}$, represents the predicted probability of edges, wherein $||\cdot||_1$ represents the l_1 norm. This term promotes the sparsity of edges, reducing false positives introduced by random noise that manifest as fragmented edge-like artifacts. Hyperparameter β indicates the relative importance for this term. The proposed loss in Eqn 2 can be regarded as a type of prior that captures the expected structural behaviour of output segmentation and edge maps - regardless of the choice of training - thus serving to reduce overfitting on small training datasets. We use experimental cross-validation [15] to crop the input volume into $64 \times 64 \times 3$ patches to calculate local losses, and average these to form L_{spg}. The total loss function L_{total} used to train SPG-3DVS is given by:

$$L_{total} = L_{seg} + \lambda L_{spg} \qquad (3)$$

3. EXPERIMENTAL RESULTS

3.1. Experimental Setup

Datasets: We report on two datasets. The first is *VesselNN*, a computed tomography (CT) dataset [9] comprising 12 vessel image stacks acquired from the cortex of mice. The stack sizes are $512 \times 512 \times n$, where $n \in \{15, \cdots, 25\}$. Following the setup in [9], we use 10 image stacks for high-training scenario, and 2 image stacks for tests. The second dataset is *IRCAD*, a CT dataset containing blood vessel scans in and around the human liver [16]. It includes 20 contrast-enhanced CT volumes with various vessel structures, intensity distributions and contrast between liver and liver vessels. Each volume has a size of $512 \times 512 \times n$, where $n \in \{92, 256\}$. Following the protocol in [16], for the "high-training" scenario, we randomly select 10 image volumes for training and 10 image volumes for test. For the low-training scenario, we reduce the number of training volumes to five. Examples from the datasets are shown in Fig. 1.

Implementation Details: In order to obtain adequate training samples, we extract patches of size 256×256 with a stride size of $[128, 128]$ from training images, along with their corresponding ground truth. Final inference is carried out on the entire image. The regularization weight λ in Eq. 3 is chosen as 0.0001. For the VesselNN dataset, α is 1, for IR-CAD, α is set as 0.8. β is set to 1. We first train the network with the single loss in Eq. 1 for 20 epochs with a learning rate of 0.001 using the Adam optimizer [17], and then continue training using Eq. 3 for another 50 epochs, reducing the learning rate by 50% every 20 epochs. All our experiments are performed on an NVIDIA Titan X GPU (12GB) with the Pytorch package. Unless otherwise stated, the results in all tables and figures are averages from five experimental trials.

Evaluation Metrics Dice-Overlap/F1-score (DICE) and mean Intersection over Union (mIoU) are used for evaluating the results [9]. mIoU is the mean value of IoU computed

(a) VesselNN   (a) IRCAD
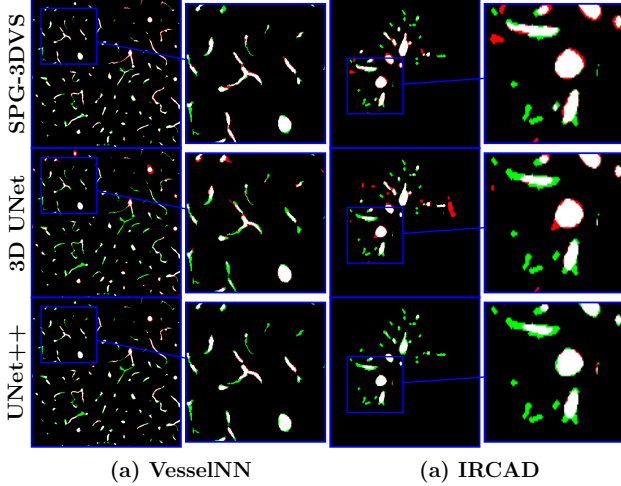
Fig. 3: Visualization of segmentation results from some sample slices in the two datasets. A small portion of each segmentation map is zoomed in and shown in the figure next to it. White: True positive pixels, black: true negative pixels, red: false positive pixels, green: false negative pixels

over a sweep of segmentation thresholds from 0.1 to 0.9 in increments of 0.1.

3.2. Ablation Study

Table 1: Ablation studies on different combinations of regularizers

Prior		VesselNN		IRCAD	
LocHomo	Sparsity	DICE	mIOU	DICE	mIOU
		0.780	0.809	0.674	0.760
✓		0.784	0.815	0.677	0.764
✓	✓	**0.786**	**0.817**	**0.679**	**0.766**

Impact of domain-specific regularizers: Table 1 shows the impact of the structural priors. We observe that each prior significantly improves performance, and their combination produces the best outcome.

3.3. Comparison with State-of-the-Art Algorithms

Table 2: Segmentation performance comparison for VesselNN

	Method	DICE	mIOU	time/voxel(1e-6s)
High training	VesselNet [18]	0.774	0.805	2.40
	U-Net++ [19]	0.772	0.808	0.82
	MFI-Net [20]	0.767	0.803	0.92
	3DUNet [21]	0.740	0.778	5.51
	DeepVess [22]	0.713	0.629	5.21
	VesselNN [23]	0.739	0.503	4.73
	SPG-3DVS	0.818	**0.817**	0.91
Low training	VesselNet [18]	0.760	0.780	—
	U-Net++ [19]	0.758	0.782	—
	MFI-Net [20]	0.736	0.776	—
	3DUNet [21]	0.679	0.668	—
	DeepVess [22]	0.680	0.690	—
	VesselNN [23]	0.685	0.697	—
	SPG-3DVS	**0.769**	**0.797**	—

We compare SPG-3DVS with recent representative state-of-the art deep methods. 2-D: **VesselNet** (CMIC2019) [18], **MFI-Net** (Plos2021) [20]. 2.5-D: **U-Net++** (MICCAI2019)

Table 3: Segmentation performance comparison for IRCAD dataset

	Method	DICE	mIOU	time/voxel(1e-6s)
High training	VesselNet [18]	0.665	0.742	2.39
	U-Net++ [19]	0.679	0.763	0.85
	MFI-Net [20]	0.658	0.750	0.96
	3DUNet [21]	0.660	0.738	5.60
	DeepVess [22]	0.630	0.715	5.19
	VesselNN [23]	0.638	0.714	4.80
	SPG-3DVS	**0.679**	**0.766**	0.95
Low training	VesselNet [18]	0.625	0.728	—
	U-Net++ [19]	0.638	0.725	—
	MFI-Net [20]	0.618	0.719	—
	3DUNet [21]	0.620	0.721	—
	DeepVess [22]	0.615	0.718	—
	VesselNN [23]	0.618	0.720	—
	SPG-3DVS	**0.644**	**0.735**	—

[19]. 3-D: **VesselNN** (Arxiv2016) [23], **3DUNet** (MICCAI2016) [21], **DeepVess** (Plos2019) [22].

For competing methods, we directly report performance from the original papers for experiments performed on the same dataset and known standard training-test split. When using a training-test split that is different than that of the original publication, we obtain results by either using author-provided implementations or implementing the methods faithfully based on network structure guidelines and parameter settings provided in the papers.

Quantitative performance and inference time: We compare the performance of our proposed method with state-of-the-art algorithms under the training setup in Section 3.1 for both "high-training" and "low-training" scenario in Table 2 and Table 3 respectively. The last column shows inference times for each method. SPG-3DVS demonstrates superior performance in the majority of scenarios, and incurs a modest inference computation time.

Qualitative Performance: We select 3DUNet and UNet++ as the representative methods for visual comparisons in Fig. 3. Note that 3D UNet is one of the best known 3-D deep networks, and UNet++ performs second best overall across the two datasets. Viewing the qualitative results in Fig. 3, SPG-3DVS exhibits noticeably fewer false positive and false negative decisions since more pixels are labeled in white.

4. CONCLUSIONS

We propose a novel, regularized deep network for 3-D vessel segmentation guided by structural priors. We predict segmentation maps and edge profiles jointly and 3-D context is mined in both the segmentation and edge prediction branches by employing LSTM modules and the edge information facilitates learning of the segmentation map. We introduce new regularization terms to encourage local homogeneity of blood vessels induced by bio-markers and edge sparsity. Experiments on benchmark datasets reveal that our proposal outperforms state-of-the-art alternatives on well-known evaluation metrics while still maintaining near real-time inference complexity.

5. REFERENCES

[1] P. Carmeliet and R. K Jain, "Angiogenesis in cancer and other diseases," *nature*, vol. 407, no. 6801, pp. 249–257, 2000.

[2] P. A Campochiaro, "Molecular pathogenesis of retinal and choroidal vascular diseases," *Progress in retinal and eye research*, vol. 49, pp. 67–81, 2015.

[3] C. Essert, et al., "Statistical study of parameters for deep brain stimulation automatic preoperative planning of electrodes trajectories," *International journal of computer assisted radiology and surgery*, vol. 10, no. 12, pp. 1973–1983, 2015.

[4] B. R. Thomson, et al., "Mr-to-us registration using multiclass segmentation of hepatic vasculature with a reduced 3d u-net," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2020, pp. 275–284.

[5] D. Hao, et al., "Sequential vessel segmentation via deep channel attention network," *Neural Networks*, 2020.

[6] J Chen, et al., "Combining fully convolutional and recurrent neural networks for 3d biomedical image segmentation," in *Advances in neural information processing systems*, 2016, pp. 3036–3044.

[7] S. van de Leemput, et al., "Stacked bidirectional convolutional lstms for deriving 3d non-contrast ct from spatiotemporal 4d ct," *IEEE transactions on medical imaging*, vol. 39, no. 4, pp. 985–996, 2019.

[8] D. Wu, et al., "Automated anatomical labeling of coronary arteries via bidirectional tree lstms," *International journal of computer assisted radiology and surgery*, vol. 14, no. 2, pp. 271–280, 2019.

[9] S Gur, et al., "Unsupervised microvascular image segmentation using an active contours mimicking neural network," in *Proceedings of the IEEE International Conference on Computer Vision*, 2019, pp. 10722–10731.

[10] A Nazir, et al., "Off-enet: An optimally fused fully end-to-end network for automatic dense volumetric 3d intracranial blood vessels segmentation," *IEEE Transactions on Image Processing*, vol. 29, pp. 7192–7202, 2020.

[11] Q Yu, et al., "Thickened 2d networks for efficient 3d medical image segmentation," *arXiv preprint arXiv:1904.01150*, 2019.

[12] T. Chan and L. A Vese, "Active contours without edges," *IEEE Transactions on image processing*, vol. 10, no. 2, pp. 266–277, 2001.

[13] G Huang, et al., "Densely connected convolutional networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 4700–4708.

[14] Q. Liu, et al., "Bidirectional-convolutional lstm based spectral-spatial feature learning for hyperspectral image classification," *Remote Sensing*, vol. 9, no. 12, pp. 1330, 2017.

[15] Vishal Monga, *Handbook of Convex Optimization Methods in Imaging Science*, vol. 1, Springer, 2017.

[16] Q Huang, et al., "Robust liver vessel extraction using 3d u-net with variant dice loss function," *Computers in biology and medicine*, vol. 101, pp. 153–162, 2018.

[17] D. P Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[18] T Kitrungrotsakul, et al., "Vesselnet: A deep convolutional neural network with multi pathways for robust hepatic vessel segmentation," *Computerized Medical Imaging and Graphics*, vol. 75, pp. 74–83, 2019.

[19] H Cui, et al., "Pulmonary vessel segmentation based on orthogonal fused u-net++ of chest ct images," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2019, pp. 293–300.

[20] Y Jiang, et al., "Mfi-net: A multi-resolution fusion input network for retinal vessel segmentation," *Plos one*, vol. 16, no. 7, pp. e0253056, 2021.

[21] Özgün Çiçek, Ahmed Abdulkadir, Soeren S Lienkamp, Thomas Brox, and Olaf Ronneberger, "3d u-net: learning dense volumetric segmentation from sparse annotation," in *International conference on medical image computing and computer-assisted intervention*. Springer, 2016, pp. 424–432.

[22] M Haft-Javaherian, et al., "Deep convolutional neural networks for segmenting 3d in vivo multiphoton images of vasculature in alzheimer disease mouse models," *PloS one*, vol. 14, no. 3, pp. e0213539, 2019.

[23] P. Teikari, et al., "Deep learning convolutional networks for multiphoton microscopy vasculature segmentation," *arXiv preprint arXiv:1606.02382*, 2016.