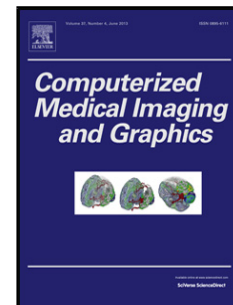


## Accepted Manuscript

Title: Vessel Net: A Deep Convolutional Neural Network with Multi Pathways for Robust Hepatic Vessel Segmentation

Author: Titinunt Kitrungrotsakul Xian-Hua Han Yutaro Iwamoto Lanfen Lin Amir Hossein Foruzan Xiong Wei Yen-Wei Chen



PII: S0895-6111(18)30409-9  
DOI: <https://doi.org/doi:10.1016/j.compmedimag.2019.05.002>  
Reference: CMIG 1634

To appear in: *Computerized Medical Imaging and Graphics*

Received date: 14 July 2018  
Revised date: 20 March 2019  
Accepted date: 13 May 2019

Please cite this article as: Titinunt Kitrungrotsakul, Xian-Hua Han, Yutaro Iwamoto, Lanfen Lin, Amir Hossein Foruzan, Xiong Wei, Yen-Wei Chen, Vessel Net: A Deep Convolutional Neural Network with Multi Pathways for Robust Hepatic Vessel Segmentation, *Computerized Medical Imaging and Graphics* (2019), <https://doi.org/10.1016/j.compmedimag.2019.05.002>

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

# Vessel Net: A Deep Convolutional Neural Network with Multi Pathways for Robust Hepatic Vessel Segmentation

Titinunt Kitrungrotsakul<sup>a</sup>, Xian-Hua Han<sup>b</sup>, Yutaro Iwamoto<sup>a</sup>, Lanfen Lin<sup>c,\*</sup>,  
Amir Hossein Foruzan<sup>d</sup>, Xiong Wei<sup>e</sup>, Yen-Wei Chen<sup>a,f,c,\*</sup>

<sup>a</sup>*Graduate School of Information Science and Engineering, Ritsumeikan University, Shiga, Japan*

<sup>b</sup>*Faculty of Science, Yamaguchi University, Yamaguchi, Japan*

<sup>c</sup>*College of Computer Science and Technology, Zhejiang University, Hangzhou, China*

<sup>d</sup>*Biomedical Engineering Group, Engineering Faculty, Shahed University, Iran*

<sup>e</sup>*Institute for Infocomm Research, A-Star, Singapore*

<sup>f</sup>*Zhejiang Lab., Hangzhou, China*

---

## Abstract

Extraction or segmentation of organ vessels is an important task for surgical planning and computer-aided diagnoses. This is a challenging task due to the extremely small size of the vessel structure, low SNR, and varying contrast in medical image data. We propose an automatic and robust vessel segmentation approach that uses a multi-pathways deep learning network. The proposed method trains a deep network for binary classification based on extracted training patches on three planes (sagittal, coronal, and transverse planes) centered on the focused voxels. Thus, it is expected to provide a more reliable recognition performance by exploring the 3D structure. Furthermore, due to the large variety of medical data device values, we transform a raw medical image into a probability map as input to the network. Then, we extract vessels based on the proposed network, which is robust and sufficiently general to handle images with different contrast obtained by various imaging systems. The proposed deep network provides a vessel probability map for voxels in the target medical data, which is used in a post-process to generate the final segmentation result. To

---

\*Corresponding authors

Email address: chen@is.ritsumei.ac.jp (Yen-Wei Chen), llf@zju.edu.cn (Lanfen Lin)

validate the effectiveness and efficiency of the proposed method, we conducted experiments with 20 data (public datasets) with different contrast levels and different device value ranges. The results demonstrate impressive performance in comparison with the state-of-the-art methods. We propose the first 3D liver vessel segmentation network using deep learning. Using a multi-pathways network, segmentation results can be improved, and the probability map as input is robust against intensity changes in clinical data.

*Keywords:* vessel segmentation, convolution neural network, 3D medical imaging

---

## 1. INTRODUCTION

The liver has a complex vascular structure. Knowledge of the morphology and topology of this vascular structure can help both anatomy education and surgical planning. Understanding the vessel structure can improve the accuracy of organ analysis, diagnostics, hepatic disease treatment, and surgery [1]. For example, understanding the branching pattern and spatial relationships among different vessels helps the development of effective surgical plans for hepatic operations, such as hepatic tumor surgery. Vessel segmentation is challenging because vessels present as a complex inhomogeneous background and significant noise directly affects segmentation results.

Typically, doctors must examine each slice manually to obtain accurate vessel segmentation. However, manual vessel segmentation is extremely tedious and depends on the experience and skill of the practitioner. Therefore, automatic vessel segmentation has attracted research attention lately.

In hepatic vessel segmentation, each voxel is assigned a label to represent whether the voxel belongs to a vessel or non-vessel (liver and other organs are in this category) region. Organ segmentation methods based on shape models perform very well and they can easily be combined with other methods to improve segmentation results. However, it is difficult to apply shape models to segment vessels with a branched tree structure due to large variations. In addition,

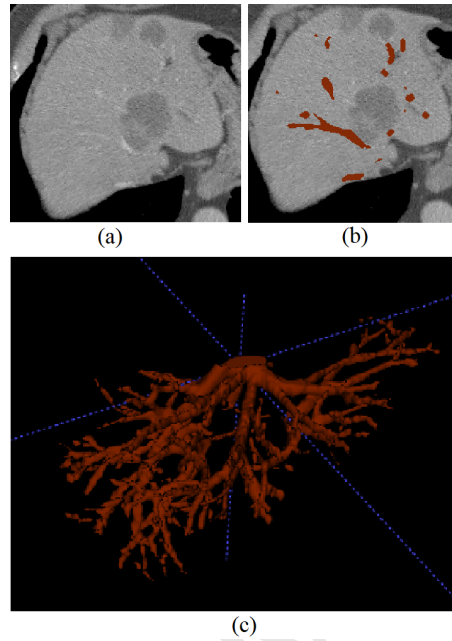


Figure 1: Liver and its vessel structure. (a) CT image of liver with tumor and vessels. (b) Ground truth of vessel segmentation in red color. (c) Visualization of vessel structure in 3D model.

noise and different intensity ranges in various data directly affect segmentation methods. As shown in Figure 1, spreading vessel branches are difficult to determine; thus, it is also difficult to achieve segmentation using a single plane slice. We have identified several problems from an analysis of liver vessel images. 1) Compared to lung vessels or tree branches in natural images, hepatic vessels are difficult to separate from vessel and non-vessel voxels. The voxels intensity value of liver has similar as vessel. 2) Some voxels are classified as noise and become non-vessel voxels when segmentation is performed using 2D methods. 3) Trained models cannot be used to segment vessel because of the different intensity of information obtained various medical data.

Although many vessel segmentation methods have been proposed, typically, the performance of existing approaches varies depending on the modality, the application domain, and other factors. Generally, automatic and semi-automatic

methods have been proposed to extract vessel structures.

35 Most automatic vessel segmentation methods use a multiscale filter to extract vesselness features [2]. These methods benefit from vesselness features by employing another method to eliminate unnecessary information, such as noise and boundary information. In addition to combining methods to eliminate unnecessary information, many methods attempt to obtain missing information  
40 from a vesselness feature, such as a branch connection [2, 3, 4]. J. Zhang et al. [5] also proposed a method based on combination of the multiscale filter with 3D patch based structured random forest (SRF) for perivascular space segmentation in brain image.

Semi-automatic segmentation methods require constraints, such as an initial  
45 point, seed point, or initial segmentation. Such methods can improve segmentation results, and by exploiting the advantage of constraints, region growing, graph cuts, and random walk methods, can yield high-performance results [6, 7, 8]. However, assigning an initial constraint, such as a seed point, is a difficult task. In addition, under- and over-segmentation may occur when initial  
50 points are positioned incorrectly.

Recently, deep learning architectures have been applied successfully to computer vision tasks by automatically learning the hierarchies of relevant features directly from the input data. For example, a deep convolutional neural network (CNN) has been applied successfully to image classification and object detection.  
55 Since 2012, CNNs have been the most successful networks in ImageNet classification [9]. Moreover, ResNet and DenseNet have been proposed in recent years with a higher classification performance [10, 11]. Peng et al[12] proposed a multi-sal ResNet for classification and quantification of emphysema, which was also a pixel-wised segmentation problem. The deep learning-based approach  
60 has superior performance over the conventional machine learning-based methods such as support vector machine, which used hand-crafted features (local binary patterns) and mid-level features (bag of visual words) [13]. Several deep learning-based methods [14, 15, 16, 17, 18, 19] have been proposed to extract retinal vessel from 2D images. Those methods can well perform on small parts,

65 however objective of vessel segmentation on liver, brain, or lung are 3D segmentation. 2D methods mostly fail on 3D image because the direction of the vessel not occur only X-Y plane but also in Z axis, that make 2D methods lose important information from Z axis. The intra-variations of the CT intensity will also significantly affect the segmentation results.

70 A fully convolutional network (FCN), based on a CNN architecture, has been proposed to perform semantic segmentation [20, 21, 22, 23]. The advantage of FCN-based methods is that the network can accept an entire image as input and performs fast and accurate segmentation.

Although these methods have achieved promising results for large organ 75 segmentation tasks, it is still a challenge to apply these methods efficiently to vessel extraction because vessels typically have detailed structures and only small parts appear in the image. C. Lian et al [24] used 3D FCN with multi-inputs types steam for perivascularspace segmentation in brain image. The method only used raw images and was also facing a problem of limited number 80 of training data compared with large amount of parameters of the 3D network, which will be shown in Sec.3.3.2

Motivated by existing conventional vessel extraction approaches and deep learning frameworks [25], we propose a specific CNN architecture with multi-pathways for liver vessel segmentation, which is called VesselNet. The proposed 85 architecture combines multiple networks, in which each network learns different features from a processed vesselness image. The main advantages and contributions of the proposed method are as follows.

1. The main problem with using semantic segmentation methods (such as FCN or U-Net) for vessel segmentation is that they cannot handle shape- 90 less structures and small objects in complete images. Thus, we use a patch-based pixel/voxel-wise classification method for accurate vessel segmentation (we will show comparison results in Sec.3.3.3).
2. The proposed CNN with multi-pathways can gather information of the vessel that occur in another axis; with these information the network can

95 has more accurate segmentation than the 2D network with one pathway  
 or branch [14, 15, 16, 17, 18, 19]. On the other hand, compared with the  
 3D network [24], the proposed network employs fewer parameters than  
 the 3D networks and shows better performance for a limited training data  
 (we will show experimental results in Sec.3.3.2). We also compare several  
 100 advanced deep networks and use DenseNet as our backbone network to  
 reduce the overfitting rate. To the best of our knowledge, this is the  
 first application of deep CNN with multi-pathways for 3D hepatic vessel  
 segmentation.

3. We propose using a vesselness probability map image as input to our  
 105 proposed network, rather than raw CT images. The proposed method is  
 more robust and insensitive to the CT intensity changes compared with  
 the existing deep learning-based vessel segmentation methods [14, 15, 24].
4. We compare our method with state-of-the-art-methods methods using  
 both clinical dataset and simulation dataset. The experiments show that  
 110 our method can achieve much higher segmentation accuracy than the  
 state-of-the-art approaches.

The remainder of this paper is organized as follows. The details of the proposed networks are described in Section 2. Section 3 presents the experimental results. The conclusions are presented in Section 4.

## 115 2. OUR VESSEL SEGMENTATION METHOD

In this section, we describe the datasets used in this study and the data preparation process, followed by a description of the proposed architecture.

### 2.1. Dataset

The datasets used for training and evaluating the proposed method are from  
 120 IRCAD [26] and VASCUSYNTH [27]. Both datasets are public; however, each  
 dataset has a different acquisition method. The IRCAD clinical dataset comprises 20 MRI scans ( $512 \times 512$  pixels per slice). Examples of the liver and its

vessels manual segmentation are provided in each dataset. Each scan has a different intensity and liver shape, and most data includes a tumor inside the liver.

125 The second dataset is VASCUSYNTH, which is a simulation dataset. It consists of 10 groups of data with  $101 \times 101 \times 101$  voxels. Each group has 12 sub-data belonging to the same structure with a different number of vessel branches. To make our network suitable for clinical data, such as the IRCAD data, and avoid a retraining process for the datasets, which have various intensity ranges, and  
130 factor for each data, we transformed intensity values to a vesselness probability map form. A multiscale method [4] can be considered a vesselness probability map method that can be used to enhance vessel information in medical images. The normalization (vesselness) value was proposed by Frangi and is expressed by Eq. 1

$$V(\sigma) = \begin{cases} 0, & \text{if } \lambda_2 > 0 \text{ or } \lambda_1 > 0 \\ (1 - \exp(\frac{R_\alpha^2}{2\alpha^2})) \exp(\frac{R_\beta^2}{2\beta^2}) (\exp(\frac{\gamma^2}{2C^2})), & \text{otherwise} \end{cases} \quad (1)$$

135 where  $\alpha, \beta, C$  are parameters that control the sensitivity of filter for balance between plate ( $R_\alpha$ ), blob ( $R_\beta$ ), and background ( $\gamma$ ), and can be formulated by Eqs 2 , 3 , 4

$$R_\alpha = \frac{|\lambda_2|}{|\lambda_1|}, \quad (2)$$

$$R_\beta = \frac{|\lambda_3|}{|\lambda_2||\lambda_1|}, \quad (3)$$

$$\gamma = \sqrt{\sum_{j \leq 3} \lambda_j^2}, \quad (4)$$

The processed data are constrained by the same data range, which is the probability of a voxel belonging to the vessel class. Figure 2 shows two typical CT data (first column) with different intensities and their probability maps  
140 (vesselness) obtained by the multiscale filter. By using the vesselness (probability map) as input, we can reduce the retraining model process. In addition,



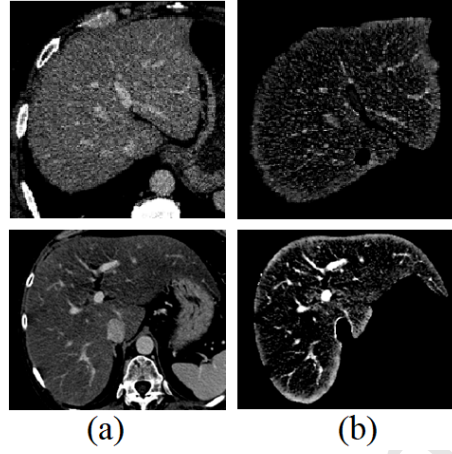


Figure 2: Data normalization from raw clinical data to probability map. First row (a) shows the raw data, while second row (b) show the probability map information results using the multiscale method.

we do not need to separate training for the scan type of each phase, which can increase the number of training datasets by mixing them together.

145 2D orthogonal (sagittal, coronal, and transverse) patches of size  $29 \times 29$  centered on a target voxel are extracted from the processed data. The class label (vessel or non-vessel) of the center voxel (target voxel) is used as the class label of the patch. The total number of patches in each data is 50,000 patches, in which 30,000 patches are for non-vessel and 20,000 patches are for vessel  
150 patches. Image patches are selected randomly from an all patch set. The leave-one-out method is used for training and testing, and the total training dataset is a set of 950,000 patches. For the VASCUSYNTH dataset, we added Gaussian noise (1-10%) to the data and then generated patches in the same manner as the IRCAD data in order to evaluate the noise robustness of the proposed method.

## 155 2.2. Proposed Architecture

Since the vessel information is a shapeless and detailed structure, we perform segmentation from patch-based classification using three 2D orthogonal patches. In most CNN-based segmentation models [28, 29] (deep convolution network

for pixel/voxel classification), a conventional CNN can be considered a single-  
 160 pathway network. Here, the input is a 2D or 3D image (usually medical images).  
 The proposed method predicts the class of the center voxel in the patches using  
 multi-networks. The proposed network architecture is shown in Figure 3. It  
 consists of three separate convolutional networks to extract the features of the  
 sagittal, coronal, and transverse planes. Three features from each network are  
 165 concatenated and used in a fully connected layer. With this concatenated of  
 three steams, the network can gather useful information from another axis and  
 the size of the proposed network is still smaller than those of 3D networks.  
 The main reason of using 2D multi-pathways is to gather information of the  
 vessel that occur in another axis; with these information the network can has  
 170 more accurate segmentation than the conventional 2D network [14, 15] with one  
 pathway of branch. Also, the proposed 2D multi-pathways network requires less  
 data than the 3D network [27] due to the small size of parameter of the network.  
 Comparison results will shown in Sec.3.3.2

We use DensNet [11] as the backbone for each pathway to learn the represent-  
 175 ed features. The DenseNet architecture is an extension of ResNet, which group-  
 s convolution layers into a block of layers. The difference between DenseNet  
 and other networks is that the network has a different connectivity pattern.  
 DenseNet is a stack of dense blocks, each of which consists of multiple layers.  
 For each dense block, the output of each layer is concatenated (instead of us-  
 180 ing addition) to form an output of the block as shown in Figure 3. It can be  
 expressed as Eq. 5:

$$y = f(x, x - 1, x - 2, \dots, x - n) \quad (5)$$

where  $x$  is the output feature from each layer and  $f$  is a set of functions,  
 namely, concatenation, batch normalization, rectified linear unit, and convolu-  
 tion. DenseNet can have narrow layers, which benefit from the implementation  
 185 of growth rate as a hyper-parameter. The feature maps of each layer are collec-  
 tive knowledge to the network. The growth rate is used to specify how much

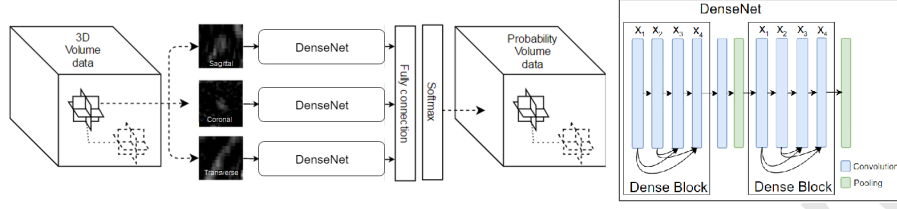


Figure 3: Architecture of our vessel segmentation, which separately trains lower feature for each type of input vector and then merges into higher level features. Each convolutional layer uses RELU as an activation function, while the final layer uses the softmax function to classify target voxel.

information each layer contributes to the collective knowledge. The number of blocks, the number of layers, and the growth rate are determined on the basis of experiments as well as the kernel size, which will be shown in Subsection 3.3 (Table 2).

As comparisons, we also use a simple four-layers CNN and a ResNet with 18 layers [10] as a backbone in our proposed deep network with multiple-pathways. The comparison results are shown in Subsection 3.3 (Table 4). Because of the multi-paths model, the output at a higher network layer, the represented features are merged into higher level features that capture complex correlations across the different input vectors, which are learned using fully connected layers. The feature map is computed as Eq. 6:

$$O_r^l = \phi\left(\sum_k W_k^{lr}\right) \cdot O_k^{(l-1)r} + b^{lr} \quad (6)$$

where  $W_k^{lr}$  is a weight matrix,  $O_k^{(l-1)r}$  is an output feature from the previous layer or input patches, and  $b^{lr}$  represents the scalar bias. While convolutional layers are used to generate a feature map, a fully connected or dense layer is used to compute a feature map for binary classification. To obtain features that are non-linear transformations of the input, an element-wise activation function is applied to the convolution result. Well-known activation functions are sigmoid, hyperbolic tangent, and rectified linear functions (ReLU)[30, 31, 30].

205 In this study, we noticed that convolutional activations strongly influence the descriptive ability of the network. In this work, we applied ReLU for convolution activation, which has been proven to speed up training processes and avoid vanishing gradients in comparison to the classic sigmoid activation function.

We apply the softmax function to the last layer to squash the output of a  
210 dense layer to a categorical probability distribution.

The neural network training methods are combined with two components, i.e., a loss function and an optimization algorithm that attempts to minimize the loss function. A loss function specifies the learning objective by mapping network weights to scalar values. We refer to the value that specifies the quality  
215 of the weights as a penalty. There are many types of loss functions, such as square loss, hinge loss, mean squared error, and cross entropy [32, 33, 34, 35, 36, 37]. Each loss type may provide a good result, improve training time and can be (or not be) easy to optimize for a given domain. Most previous deep learning networks for classification and segmentation, particularly for natural  
220 images, use cross entropy as a loss function. However, it seems to not work effectively in vessel patch classification. In this study, the networks loss function is the negative loglikelihood function, which provides better results than cross entropy. The negative loglikelihood function is defined by Eq. 7.

$$E(\theta) = -\frac{\sum_{i=1}^n \log(\sum_{j=1}^N x_j^i y_j^i)}{n} \quad (7)$$

Training is cast into the minibatch, which is performed by the stochastic gradient descent algorithm, a variant of the gradient descent algorithm commonly  
225 used to train large networks with large datasets [38]. We apply minibatch to the learning algorithm to avoid memory issues. The learning method updates the weight with a small batch of training data, rather than considering all training data. A momentum term is particularly beneficial when there is a slim margin  
230 of error. With momentum, we can accelerate the learning speed and avoid local minima [39]. The update weight function with momentum  $m$  and learning rate

at iteration  $t$  is given by Eq. 8.

$$\Delta w_i^{lr}(t) = -\alpha \frac{\partial E}{\partial w_i^{lr}} + m \Delta w_i^{lr}(t-l) \quad (8)$$

The output of the network will be in the probabilities value map for voxels in the target medical data. The output is then used in our post-processing method  
235 to generate the final segmentation result.

### 2.3. Post-processing

Normally, the output with the highest probability is selected as a label of the target voxel( $x, y, z$ ). However, the CNN-based classification will generate some noise voxels, and some voxels will not be recognized as a vessel, which  
240 will generates fragmented vessel results. To overcome these limitations, we use information of neighbor voxels to refine the label of the target voxel as Eq.10.

$$W_{x+i,y+j,z+k} = 1 - \frac{\sqrt{i^2 + j^2 + k^2}}{\sqrt{N^2 + M^2 + O^2}} \quad (9)$$

$$\hat{p}_{x+i,y+j,z+k} = \frac{\sum_{i=-N}^N \sum_{j=-M}^M \sum_{k=-O}^O (W_{x+i,y+j,z+k})(p_{x+i,y+j,z+k})}{\sum_{i=-N}^N \sum_{j=-M}^M \sum_{k=-O}^O W_{x+i,y+j,z+k}} \quad (10)$$

Here,  $p$  and  $\hat{p}$  are probabilities before and after the refinement.  $N$ ,  $M$ , and  $O$  are distance ranges. We set  $N$ ,  $M$ , and  $O$  to 1. After this post-processing, we can generate a new probability value for the target voxel ( $x,y,z$ ). This method  
245 reduces the probability of a noise image and increases the probability values of hidden vessel pixels/voxels.

## 3. EXPERIMENTS AND RESULTS

This section presents and discusses experimental results. First, we describe our implementation, evaluation methods, and experimental setup.

### 250 3.1. Implementation Detail

Our network is based on an implementation that uses the Keras and TensorFlow libraries [40, 41]. TensorFlow is an open source software library for numerical computation and deep learning. Keras is a high-level API written in Python that can run on top of TensorFlow. Both libraries support the use of a GPU, which can accelerate the execution of deep learning in the training phase.

The hyper-parameters of the network were tuned using cross-validation with a validation set [28]. The chosen hyper-parameters were those with which the model performed best with the validation set. All computations were performed using an NVIDIA GeForce GTX TITAN X GPU with 12 GB memory. To generate datasets, we generated an ROI in the liver region and applied multiscale to generate vesselness probability. The training set was randomly extracted from a voxel with a value greater than the designed threshold  $T = mean - variance$  in the vesselness probability map image without a liver boundary, while the test set used all voxels that contain a vesselness probability map value  $T > 0$  that still eliminated the liver boundary. The training set had 50k voxels from each volume. For each voxel, we extracted three 2D orthogonal patches of  $29 \times 29$  voxels from the sagittal, coronal, and transverse planes, as shown in Figure 3. The class of the extracted patch is determined by the class of the center pixel. The learning rate and momentum were set to 0.00005 and 0.9, respectively. We used a batch size of 2000 patches.

### 3.2. Evaluation Methods

The leave-one-out method was used in our experiments. We selected one CT volume as a test image, and other CT volumes were used for training. For each method, all data are performed to collect experimental results with a different training set. The overall performance of each method was evaluated using the test set. The overlap index (Dice coefficient), volumetric overlap error (VOE), Sensitivity (SEN), and Positive Predictive Value (PPV) were used as evaluation

measures, which are shown in Eqs. 11 - 14, respectively.

$$DICE = \frac{2|A \cap B|}{|A| + |B|} \quad (11)$$

$$VOE = (1 - \frac{|A \cap B|}{|A \cup B|}) \times 100\% \quad (12)$$

$$SEN = (\frac{A \cap B}{B}) \quad (13)$$

$$PPV = (\frac{A \cap B}{A}) \quad (14)$$

Here,  $A$  is the segmented vessel mask and  $B$  is the ground truth vessel mask, which was provided by the public database. The proposed method was  
 280 evaluated using two datasets. The first dataset included the IRCAD datasets, which are maintained by a French research institute [26]. Most CT images have tumors, liver abnormalities, or different intensity ranges that affect vessel information. Three CT volumes (Nos. 16, 18, and 20) from the IRCAD datasets  
 285 have different intensity ranges. Most data have an average intensity value of 90-140, while Nos. 16, 18, and 20 have an average intensity value less than 60. The other dataset was VASCUSYNTH, which includes synthetic data [42].

### 3.3. IRCAD Dataset

This section presents the experimental results obtained with the IRCAD  
 290 dataset. The first experiment focused on the effect of the loss function. The second and third experiments compare the proposed method to other deep learning networks and state-of-the-art methods.

#### 3.3.1. Tuning parameters and effect of loss function

We conducted an experiment using 20 datasets to observe the effect of the  
 295 loss functions. Table 1 compares the Dice coefficients of segmentation results per 25 epochs with different loss functions (negative loglikelihood and cross entropy) as loss functions. Compared to cross entropy, the negative loglikelihood achieved

Table 1: Comparison of dice value's of segmentation result per 25 epochs with different loss functions

Loss function	25	50	75	100
Cross-entropy	0.845041	0.858111	0.881688	0.863343
Negative log likelihood	0.8350994	0.86317	0.903459	0.862856

better performance. The segmentation accuracy (the Dice coefficient) could be improved to 0.903 as increasing the epoch to 75, while the accuracy with the cross entropy is about 0.882. So we used the negative loglikelihood as our loss function in our future experiments.

Table 2 shows the parameter turning of DenseNet as the backbone of our multi-pathways model. In this experiment, we conducted a parameter turning to find the best parameters that focus on the number of blocks, total layers, and growth rate, which have parameters that range 2-3, 10-19, and 4-16 respectively. Dice coefficient is used to evaluate the performance models. If the number of blocks is 3, total layers are 10, and the growth rate is 8, the model can generate a Dice coefficient of 0.9035, which is higher than other parameter setting. This best model was used for evaluation and comparison in other tests in this study.

### 3.3.2. Comparison of the proposed multi-pathways 2D CNN with the single-pathway 2D CNN and 3D CNN

To show the effectiveness of the proposed three-pathways deep CNN, we performed vessel segmentation experiments using six different deep learning networks, i.e., three types of single-pathway 2D CNNs (XY, XZ, and YZ patches), a three-pathways 2D CNN (proposed VesselNet), and a 3D CNNs. They are summarized in Table. 4. All patch sizes for each 2D input network were the same ( $29 \times 29$ ). Two 3D CNNs were performed with different patch sizes (different input image sizes). The first 3D CNN used a patch size of  $29 \times 29 \times 29$ , which is the same patch size as that used in 2D CNN. The second 3D CNN used a smaller patch size of  $15 \times 15 \times 15$ . Compared with number of the first



Table 2: Segmentation accuracies (Dice coefficients) with different parameters of the DenseNet backbone (for parameter tuning)

Growth rates	Data											
	2				3				4			
	Total Layers				Total Layers				Total Layers			
	10	13	16	19	10	13	16	19	10	13	16	19
4	0.87	0.89	0.88	0.89	0.88	0.88	0.87	0.88	0.89	0.89	0.89	0.88
8	0.89	0.89	0.88	0.89	<b>0.90</b>	0.89	0.88	0.88	0.89	0.88	0.89	0.87
12	0.89	0.89	0.88	0.89	0.89	0.89	0.88	0.88	0.89	0.88	0.88	0.88
16	0.88	0.88	0.85	0.89	0.88	0.88	0.87	0.88	0.89	0.88	0.88	0.88

3D CNN's parameters, the second 3D CNN has fewer parameters because of the smaller size of the input image. Since the 3D network usually increases the imbalance between vessel/background voxels, we performed an additional experiment using a 3D CNN (the third 3D CNN) with a Dice loss for optimization.

325 The Dice loss was proposed to solve the imbalance problem.

The segmentation accuracy (Dice coefficient) of each data with each network is shown in Figure 4. The average Dice coefficients are summarized in Table 3. As shown in Table 3, the segmentation accuracy of the single-pathway 2D CNN is about 0.81 and there are no significant difference among three single-pathway networks. On the other hand, the proposed three-pathways 2D CNN (VesselNet) can significantly improve the segmentation accuracy to 0.903. The segmentation accuracies of the first and second 3D CNNs are 0.77 and 0.83, respectively. Even we used the Dice loss for 3D CNN (the third 3D CNN), there are no significant improvement (the accuracy is about 0.82). Though the 3D CNN implementation could provide more comprehensive structural knowledge for segmentation, the 3D network architecture increases the number of network parameters. That means we need to use more training data for 3D CNNs to

Table 3: Comparison of the proposed multi-pathways 2D CNN with single-pathway 2D CNNs and 3D CNNs

Network	patch size	Loss	Dice
2D CNN (xy)	$29 \times 29$	Negative log likelihood	0.81916
2D CNN (yz)	$29 \times 29$	Negative log likelihood	0.78959
2D CNN (xz)	$29 \times 29$	Negative log likelihood	0.81100
3D CNN 1	$29 \times 29 \times 29$	Negative log likelihood	0.77330
3D CNN 2	$15 \times 15 \times 15$	Negative log likelihood	0.83719
3D CNN 3	$15 \times 15 \times 15$	Dice loss	0.82323
Our-VesselNet	$29 \times 29$	Negative log likelihood	0.90346

achieve the higher accuracy. As shown in Figure 4 and Table 3, we can also see that the 3D CNN 2 with a smaller input image size ( $15 \times 15 \times 15$ ) achieved higher accuracy than the 3D CNN 1 with a larger input image size ( $29 \times 29 \times 29$ ) because the 3D CNN 2 has fewer parameters than the 3D CNN 1.

This experiment also showed that the axial, coronal, and transverse planes are all important for vessel segmentation. The results obtained with the single-pathway and three-channel network show that features from each plane cannot be extracted using the same trained kernels. The Dice coefficient of the single-pathway and three-channel network was less than that of the multi-network. The proposed architecture (three-pathways CNN) demonstrated better performance than the other architectures. The 3D CNN network also generated good results; however, we found that using the 3D convolution method required more training data to obtain better results. The experimental results show that a small patch size for the 3D CNN obtained slightly better results than using only a single 2D CNN. However, the computation time was more than twice that of our multi-network and four times greater than the single 2D CNN.

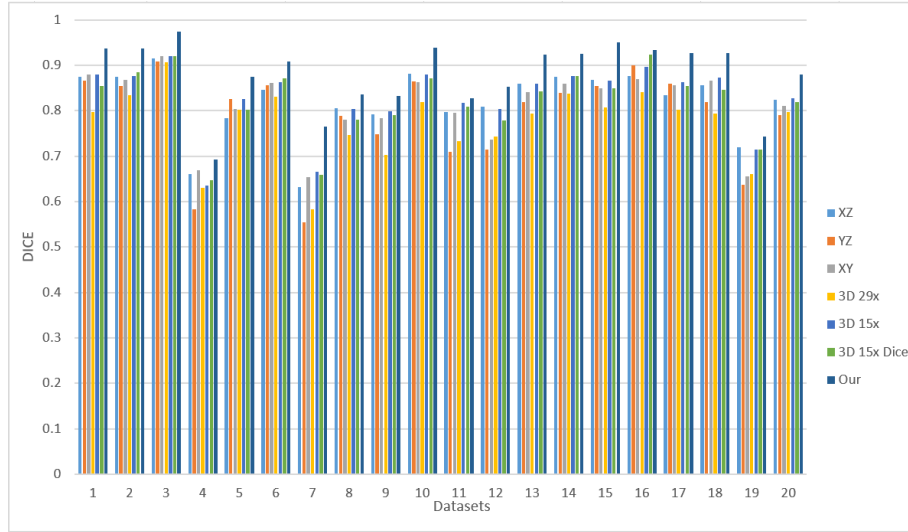


Figure 4: The segmentation accuracy of each dataset with each network.

### 3.3.3. Comparison with state-of-the-art methods

In this section, we compare the proposed CNN method with other vessel segmentation methods and a raw data input multi-network. For this experiment, we retested most methods, with the exception of the level set method, which has previously been extensively evaluated [42]. For graph cut [7] and submodular graph cut [43] methods, we performed an experiment using the same seed used to generate patches for training the proposed method, i.e., 50,000 points were provided to the graph cut methods. We also found that using raw data to generate a high-order tensor and probability data to generate a graph flow in the submodular graph cut obtained better results than using only raw data. Thus, in this experiment, we combined one for submodular graph cut and normal raw data for the conventional graph cuts. The Dice coefficient of the conventional graph cut was only 0.344 because the boundary probability term or smoothness term minimizes the cost function. This term can benefit most segmentation objects, specifically large objects, such as the liver, kidney, or lung. However, vessels have detailed information that can easily be elimi-

Table 4: Comparison of the proposed method with state-of-art method using IRCAD dataset

Method	Input	DICE	VOE(%)	SEN	PPV
Level set[42]	Processed data	0.73	-	-	-
Frangi[2]	Raw data	0.664	50.2	0.618	0.737
Graph cut[7]	Raw data	0.344	79.2	0.416	0.298
Submodular Graph cut[43]	Probability	0.750	39.3	0.776	0.727
FCN[23]	Raw data	0.107	93.9	0.171	0.078
U-NET[44]	Raw data	0.723	41.3	0.758	0.699
V-NET[45]	Raw data	0.691	42.7	0.703	0.635
Our-CNN	Probability	0.879	21.6	0.918	0.835
Our-ResNet18	Probability	0.884	20.4	0.913	0.866
Our-DenseNet	Probability	0.903	17.2	0.929	0.842

370 nated by the graph cut smoothness term. The submodular graph cut method  
 uses the submodular concept to give extra cost to a similar edge by adding a  
 high-order tensor to the cut cost. This method improved the Dice coefficient  
 from 0.344 to 0.75. Figure Table 4 shows a comparison of the proposed method  
 and other methods. Figure.5 illustrates the 3D visualization results from our  
 375 method and other methods. The Dice coefficients of the other methods includ-  
 ing deep learning-based semantic segmentation methods (FCN [23], U-Net [44]  
 and V-Net [45]) are lower than 0.8, while the proposed methods achieved higher  
 accuracies (Dice coefficients) of 0.879, 0.884 and 0.903 with a backbone network  
 of simple four-layer CNN, ResNet18 and DenseNet, respectively. Though there  
 380 are no significant difference on PPV and SEN among the three proposed meth-  
 ods, DenseNet achieved the best results in total. Therefore, we used DenseNet  
 as the backbone in our proposed multi-pathways network architecture.

### 3.3.4. Comparison of original raw CT image and vesselness probability map image as DCNN input

385 To evaluate the efficiency of the proposed probability map image as input to the DCNN, we performed experiments using the original raw CT data and their probability map images as input to our network. Figure 5 shows three typical vessel segmentation results obtained using the proposed method with two types of input. The first two rows (CT Nos. 1 and 2) demonstrate that using the original raw CT intensity and the vesselness probability map image as input  
390 can obtain good results. However, in the last row (CT No. 16), which has a lower intensity range ( $<60$ ), the network that used the original raw CT image as input could not segment the vessel well (the last row of Figure 5(e)). The network with the vesselness probability map could generate good results for all  
395 cases even for CT No.16 (the last row of Figure 5(f)). A quantitative comparison is shown in Figure 6. As can be seen, the method that used vesselness probability map is much better than the method that used the raw intensity, especially for CT volumes with different intensity ranges, such as Nos. 16, 18, and 20.

We also performed an additional experimental, which used both the raw CT image and vesselness map as inputs. In our original network using the original  
400 CT image or the vesselness map as input, we have three streams with different planes (x-y, y-z, and x-z). For the new experiments using both images, we have six streams in the network. The results are also shown in Figure 6. As shown in Fig.6, the results of the combining inputs are not so good as those using vesselness maps only. The main reason is that the result is still very sensitive to  
405 the noise. As shown in Figure 6, the accuracies for data Nos. 16, 18, 20, which have different intensities with others, are very low though they are higher than those using only original raw CT images. Another reason is that the size of training dataset is not large enough because the network with six streams has  
410 more parameters and need more dataset for training.

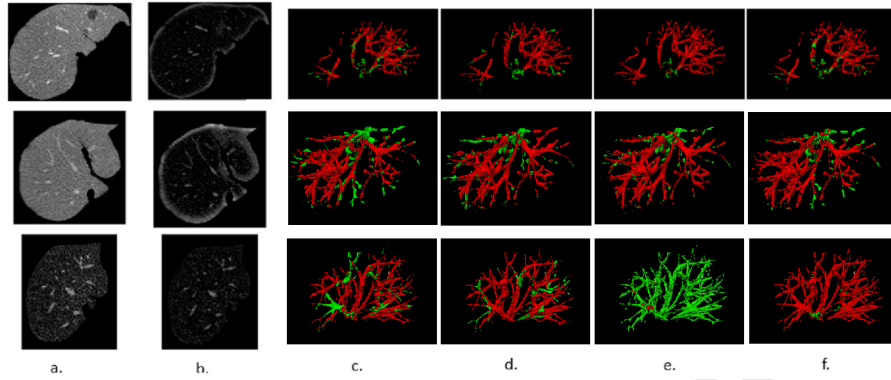


Figure 5: visualization and comparison of proposed method and state-of-the-art methods using vesselness probability as input. Row indicates different data samples. Six columns: a. CT intensity images; b. vesselness probability map images; c. visualization result of Graph Cut with submodular; d. visualization result of V-Net; e. and f. visualization results of both intensity and vesselness probability map images from our method, respectively. Red color voxels denote correct vessel results, while green color voxels are the difference between the ground-truth vessel and the segmented vessel;

### 3.3.5. VASCUSYNTH dataset

We quantitatively evaluated the proposed method using synthetic data to assess its robustness. VASCUSYNTH [27] contains synthetic volumetric data of a vascular tree that provides bifurcation location, branch properties, tree hierarchy, and ground truth segmentations. The tree generation is performed by iteratively growing a vascular structure based on a user-defined oxygen demand map. Here, the data size is  $101 \times 101 \times 101$  voxels. There are 10 groups of data, and each group contains 12 volumes, from a lower number of vessel branches (data number 1) to a higher number of vessel branches (data number 12). In this experiment, we focused on the higher number of branches. As a result, only 10 data were used in the experiments. We added Gaussian noise to the image to simulate noise data by changing the noise variance from 1% to 10%, as shown in Figure 7. We then generated five training datasets from odd noise present data and testing on even datasets.

Figure 8 shows the average Dice coefficient obtained with the VASCUSTYNTH

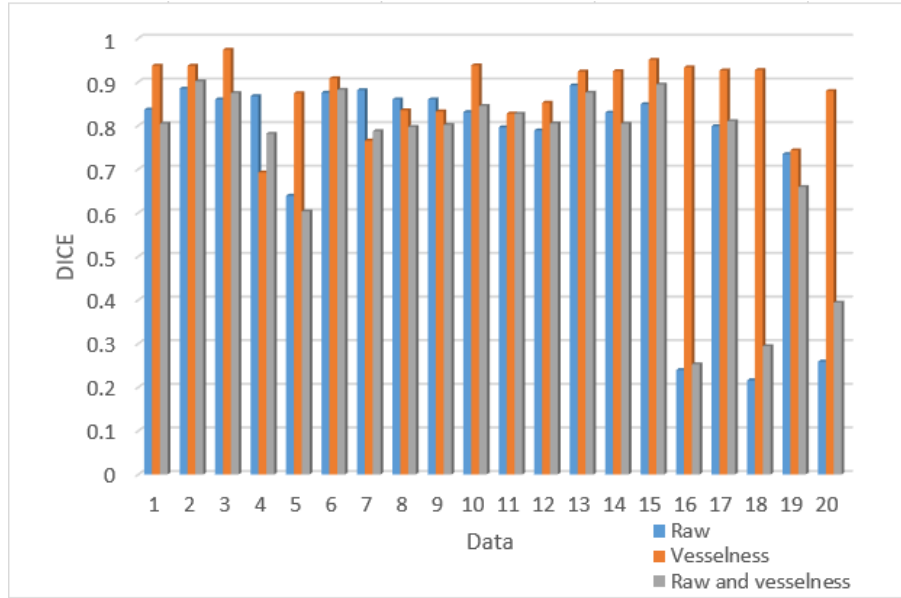


Figure 6: Comparison between the original raw CT image, the vesselness probability map, and raw & vesselness image as input of our network on IRCAD dataset.

data, which was 0.901. For 2% and 4% noise levels, we obtained high accuracy results, i.e., 0.955 and 0.942, respectively. However, the accuracy dropped to 0.899 at a 6% noise level and rapidly dropped to 0.872 and 0.837 at 8% and 10% noise levels, respectively. This rapid drop occurred because the added noise voxels were classified as vessels. It will be more beneficial to use some noise reduction techniques, such as anisotropic filter, before the vessel segmentation.

We compared our model with other approaches, including Frangi [2], Sub-modular GraphCut [43], V-Net [45], SE [46], TDF [47], and EMILOVE [48] by counting the correct segmentation, incorrect segmentation, and missed voxels. We then computed the precision and recall for all datasets on noise levels 1%, 5% , and 8%. The comparison results of the robustness against noise are listed in Table 5. When the noise level is 8, our method achieves lower precision than SE and EMILOVE. However, our method can still achieve better results on both precision and recall on all noise levels except 8% , and the average F-measure

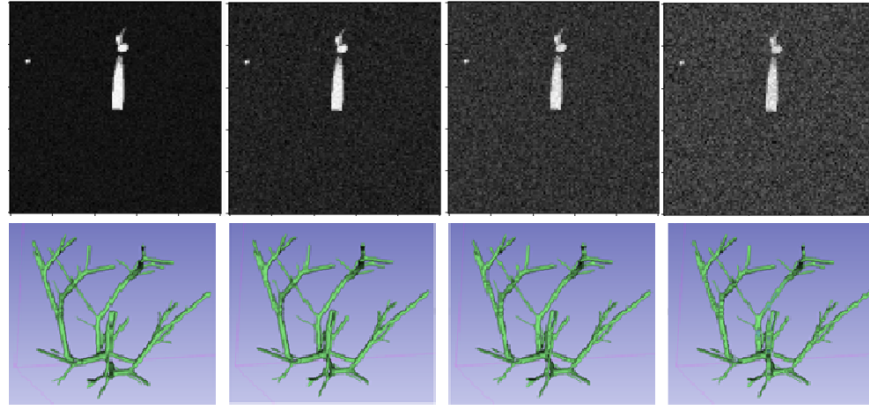


Figure 7: Segmentation results (synthetic images). Top row: Noise variance images from 2%, 4%, 6%, and 8%. Bottom: 3D visualization of the result from our model.

for all noise levels is higher than other methods. The F-measure can confirm  
 440 the robustness of our method against noise. Our method shows the best results  
 of F-measure compared with other methods in all noise levels.

#### 4. CONCLUSION

We have designed a DNN architecture to automatically segment vessels in  
 445 liver region images. The proposed networks features are learned from 2D or-  
 thogonal patches, which are first transformed from raw data to a vesselness  
 probability map using a multiscale method. This transformation can deal with  
 any intensity variation in the CT volume data, which is a problem when raw  
 intensity data are used. The results obtained with IRCAD data confirm that the  
 450 proposed model improves current state-of-the-art methods relative to accuracy.  
 However, the time required to compute the segmentation result is greater than  
 other methods because the proposed method must compute the vessel proba-  
 bility map in preprocessing and requires much more time for patch extraction  
 (approximately 7-10 min depending on the size of the liver mask). With patch  
 455 classification and segmentation, the proposed model benefits with a reduced  
 time due to a GPU implementation, requiring approximately 2-4 min depend-



Table 5: Comparison of the proposed with state-of-art method using IRCAD dataset

Method	score	1%	5%	8%
Frangi	Precision	0.69	0.66	0.52
	Recall	0.81	0.73	0.71
	F-Score	0.74	0.69	0.60
Submodular Graph cut	Precision	0.85	0.79	0.62
	Recall	0.79	0.52	0.39
	F-Score	0.81	0.62	0.48
V-Net	Precision	0.80	0.80	0.78
	Recall	0.72	0.69	0.65
	F-Score	0.76	0.75	0.72
SE	Precision	0.80	0.77	<b>0.84</b>
	Recall	0.64	0.48	0.43
	F-Score	0.71	0.59	0.57
TDF	Precision	0.77	0.69	0.65
	Recall	0.88	0.62	0.33
	F-Score	0.82	0.65	0.44
EMILOVE	Precision	0.78	0.80	0.83
	Recall	0.82	0.81	0.74
	F-Score	0.80	0.80	0.78
Our Method	Precision	<b>0.87</b>	<b>0.83</b>	0.81
	Recall	<b>0.89</b>	<b>0.85</b>	<b>0.79</b>
	F-Score	<b>0.88</b>	<b>0.84</b>	<b>0.80</b>

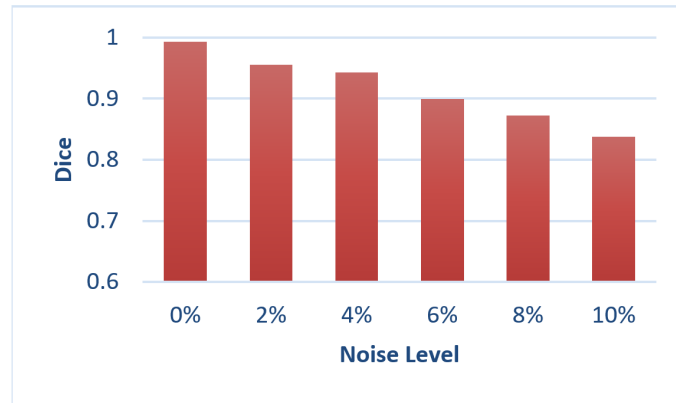


Figure 8: Evaluation of the proposed method on VASCUSYNTH data using DICE measure.

ing on the number of patches.

### Acknowledgments

This work is supported in part by Japan Society for Promotion of Science (JSPS) under Grant No. 16J09596 and KAKEN under the Grant Nos. 18H03267, 18K18078, 17H00754, 17K00420; in part by Zhejiang Lab Program under the Grant No.2018DG0ZX01; in part by the Key Science and Technology Innovation Support Program of Hangzhou under the Grant No.20172011A038; and in part by A\*STAR Research Attachment Programme.

### References

- [1] J. Brien, N. Ezquerra, Automated segmentation of coronary vessels in angiographic image sequences utilizing temporal, spatial and structural constraints, *Visualization in Biomedical Computing* 25.
- [2] A. Frangi, W. Niessen, K. Vincken, M. Viergever, Multiscale vessel enhancement filtering, *Medical Image Computing and Computer-Assisted Intervention (MICCAI)* 1496 (1998) 130–137.

- [3] Vascular active contour for vessel tree segmentation, IEEE Trans. on Biomedical Engineering 58 1023–1032, year =.
- [4] Y. Sato, S. Nakajima, H. Atsmi, T. Koller, G. Gerig, , R. Kikinis, Three-dimensional multi-scale line filter for segmentation and visualization of curvilinear structures in medical images, Medical Image Analysis 2 (1998) 143–169.
- [5] J. Zhang, Y. Gao, S. Park, X. Zong, W. Lin, D. Shen, Structured learning for 3-d perivascular space segmentation using vascular features, IEEE Transactions on Biomedical Engineering 64 (2017) 2803–2812. doi: 10.1109/TBME.2016.2638918.
- [6] Y. Masutani, K. Masamune, T. Dohi, Region-growing based feature extraction algorithm for tree-like objects, Proc. of Visualization in Biomedical Computing (1993) 161–167.
- [7] Y. Boykov, G. Funka-Lea, Graph cuts and efficient n-d image segmentation, International Journal of Computer Vision 70 (2006) 109–131.
- [8] L. Grady, Random walks for image segmentation, IEEE Transactions on pattern analysis and machine learning 28 (2006) 1768–1783.
- [9] A. Krizhevsky, I. Sutskever, G. E. Hinton, Imagenet classification with deep convolutional neural networks, In Advances in neural information processing systems (2012) 1097–1105.
- [10] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2016) 770–778.
- [11] G. Huang, Z. Liu, L. van der Maaten, K. Q. Weinberger, Densely connected convolutional networks, IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2017) 1097–1105.

- [12] L. Peng, L. Lin, H. Hu, H. Li, Q. Chen, X. Ling, D. Wang, X. Han, Y. I-wamoto, Y. Chen, Classification and quantification of emphysema using a multi-scale residual network, *IEEE J Biomed Health Inform.* 500
- [13] M. Anthimopoulos, S. Christodoulidis, L. Ebner, A. Christe, S. Mougiakakou, Lung pattern classification for interstitial lung diseases using a deep convolutional neural network, *IEEE transaction on Medical Imaging* 35 (2016) 1207–1216.
- [14] H. Fu, Y. Xu, S. Lin, D. W. K. W., J. Liu, Deepvessel: Retinal vessel segmentation via deep learning and conditional random field, *Medical Image Computing and Computer-Assisted Intervention MICCAI 2016* (2016) 132–139.
- [15] Q. Li, B. Feng, L. Xie, P. Liang, H. Zhang, T. Wang, A cross-modality learning approach for vessel segmentation in retinal images, *IEEE Transactions on Medical Imaging* 35 (2016) 109–118. doi:10.1109/TMI.2015.2457891. 510
- [16] J. Mo, L. Zhang, Multi-level deep supervised networks for retinal vessel segmentation, *International Journal of Computer Assisted Radiology and Surgery.* 515
- [17] R. Xu, G. Jiang, X. Ye, Y.-W. Chen, Retinal vessel segmentation via multiscaled deep-guidance, *19th Pacific-Rim Conference on Multimedia* (2018) 158–168.
- [18] P. Liskowski, K. Krawiec, Segmenting retinal blood vessels with deep neural networks, *IEEE Transactions on Medical Imaging* 35 (2016) 2369–2380. 520
- [19] B. Ibragimov, D. Toesca, D. Chang, A. Koong, L. Xing, Combining deep learning with anatomical analysis for segmentation of the portal vein for liver sbirt planning, *Phys Med Biol.* 62 (2017) 8943–8958.

- [20] J. Alvarez, T. Gevers, Y. LeCun, A. Lopez, Road scene segmentation from  
 525 a single image, Proceedings of the 12th European Conference on Computer  
 Vision (2012) 376–389.
- [21] D. Ciresan, A. Giusti, L. Gambardella, J. Schmidhuber, Deep neural net-  
 works segment neuronal membranes in electron microscopy images, Ad-  
 vances in neural information processing systems (2012) 2843–2851.
- 530 [22] B. Hariharan, P. Arbelaez, R. Girshick, J. Malik, Simultaneous detection  
 and segmentation, ECCV (2014) 297–312.
- [23] E. Shelhamer, J. Long, T. Darrell, Fully convolutional models for semantic  
 segmentation, PAMI.
- [24] C. Lian, J. Zhang, M. Liu, X. Zong, S.-C. Hung, W. Lin, D. Shen, Multi-  
 535 channel multi-scale fully convolutional network for 3d perivascular spaces  
 segmentation in 7t mr images, Medical Image Analysis 46. doi:10.1016/  
 j.media.2018.02.009.
- [25] T. Kitrungsakul, X. Han, Y. Iwamoto, A. Foruzan, L. Lin, Y. Chen,  
 Robust hepatic vessel segmentation using multi deep convolution network,  
 540 SPIE Medical imaging 10137.
- [26] Ircad dataset, <http://www.irca.fr/>.
- [27] G. Hamarneh, P. Jassi, Vascusynth: Simulating vascular trees for gen-  
 erating volumetric image data with ground truth segmentation and tree  
 analysis, Computerized Medical Imaging and Graphics 34 (2010) 605–616.
- 545 [28] Y. Bengio, Practical recommendation for gradient-based training of deep  
 architectures, Neural Network: tricks of the trade (2013) 437–478.
- [29] S. Tang, Y. Wang, Y.-W. Chen, Application of ica to x-ray coronary digital  
 subtraction angiography, Neurocomputing 79 (2012) 168–172.

- [30] X. Glorot, A. Bordes, , Y. Bengio, Deep sparse rectifier networks, In Proc. of 14th International Conference on Artificial Intelligence and Statistics 15 (2011) 315–323.
- [31] K. Jarrett, K. Kavukcuoglu, M. Ranzato, Y. LeCun, What is the best multi-stage architecture for object recognition?, IEEE 12th International Conference on Computer Vision (2009) 2146–2153.
- [32] L. Rosasco, E. D. Vito, A. Caponnetto, A. V. M. Piana, Are loss functions all the same?, Neural Computation (2004) 1063–1076.
- [33] Y. Shen, Loss functions for binary classification and class probability estimation, University of Pennsylvania.
- [34] A regularization tour of machine learning, MIT-9.520.
- [35] H. Masnadi-Shirazi, N. Vasconcelos, On the design of loss functions for classification: theory, robustness to outliers, and savageboost, Statistical Visual Computing Laboratory, University of California.
- [36] R. Piyush, Support vector machines, classification loss functions and regularizers, Utah CS5350/6350: Machine Learning.
- [37] R. Rifkin, R. Lippert, Notes on regularized least squares, MIT Computer Science and Artificial Intelligence Laboratory.
- [38] O. Bousquet, L. Bottou, The tradeoffs of large scale learning, In Advances in neural information processing systems 70 (2008) 161–168.
- [39] I. Sutskever, J. Martens, G. Dahl, , G. Hinton, On the importance of initialization and momentum in deep learning, In Proc. of the 30th International Conference on Machine learning (2013) 1139–1147.
- [40] F. Chollet, et al., keras, <https://github.com/fchollet/keras>.
- [41] M. Abadi, et al., Tensorflow: Large-scale machine learning on heterogeneous systems, tensorflow.org.

- 575 [42] G. Pizaine, E. D. Angelini, I. Bloch, S. Makram-Ebeid, Vessel geometry modeling and segmentation using convolution surfaces and an implicit medial axis, *IEEE International Symposium on Biomedical Imaging: From Nano to Macro* (2011) 1421–1424.
- [43] S. Jegelka, J. Bilmes, Submodularity beyond submodular energies: coupling  
580 edges in graph cuts, *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.
- [44] O. Ronneberger, P. Fischer, T. Brox, U-net: Convolutional networks for biomedical image segmentation, *Medical Image Computing and Computer-Assisted Intervention (MICCAI)* (2015) 234–241.
- 585 [45] F. Milletari, N. Navab, A. Ahmadi, V-net: Fully convolutional neural networks for volumetric medical image segmentation, *arXiv:1606.04797*.
- [46] J. A. Tyrrell, E. di Tomaso, D. Fuja, R. Tong, K. Kozak, R. K. Jain, , B. Roysam, Robust 3-d modeling of vasculature imagery using superellipsoids, *IEEE Trans. Medical Imaging*. 26 (2007) 223–237.
- 590 [47] C. Bauer, H. Bischof, A novel approach for detection of tubular objects and its application to medical image analysis, *Pattern Recognition* (2008) 163–172.
- [48] H. Skibbe, M. Reisert, S. M. et al, Efficient monte carlo image analysis for the location of vascular entity, *IEEE Trans. Medical Imaging*. 34 (2015)  
595 628–643.