TopNet: Topology Preserving Metric Learning for Vessel Tree Reconstruction and Labelling

Deepak Keshwani[1], Yoshiro Kitamura[1], Satoshi Ihara[1], Satoshi Iizuka[2], and Edgar Simo-Serra[3]

[1] Imaging technology center, Fujifilm corporation, Japan
[2] Center for Artificial Intelligence Research, University of Tsukuba, Japan
[3] Department of Computer science and engineering, Waseda university, Japan
`deepak.keshwani@fujifilm.com`

Abstract. Reconstructing Portal Vein and Hepatic Vein trees from contrast enhanced abdominal CT scans is a prerequisite for preoperative liver surgery simulation. Existing deep learning based methods treat vascular tree reconstruction as a semantic segmentation problem. However, vessels such as hepatic and portal vein look very similar locally and need to be traced to their source for robust label assignment. Therefore, semantic segmentation by looking at local 3D patch results in noisy misclassifications. To tackle this, we propose a novel multi-task deep learning architecture for vessel tree reconstruction. The network architecture simultaneously solves the task of detecting voxels on vascular centerlines (i.e. nodes) and estimates connectivity between center-voxels (edges) in the tree structure to be reconstructed. Further, we propose a novel connectivity metric which considers both inter-class distance and intra-class topological distance between center-voxel pairs. Vascular trees are reconstructed starting from the vessel source using the learned connectivity metric using the shortest path tree algorithm. A thorough evaluation on public IRCAD dataset shows that the proposed method considerably outperforms existing semantic segmentation based methods. To the best of our knowledge, this is the first deep learning based approach which learns multi-label tree structure connectivity from images.

Keywords: Deep Learning · Vessel Tree Reconstruction · Vessel Segmentation · Liver Vessel · Centerline Detection · Computed Tomography.

1 Introduction

Primary liver cancer is the third most common cause of cancer mortality [2]. Liver resection is currently the most prevalent treatment method with 5-year survival rates of up to 40% [11]. To perform a safe liver resection, preoperative surgical simulations have been found to be very useful [19,17]. In these simulations, a 3D model of the liver which shows the anatomical structures of portal and hepatic veins together with tumour location is reconstructed. Using vascular reconstructions, blood flow patterns around the tumour is analysed to compute
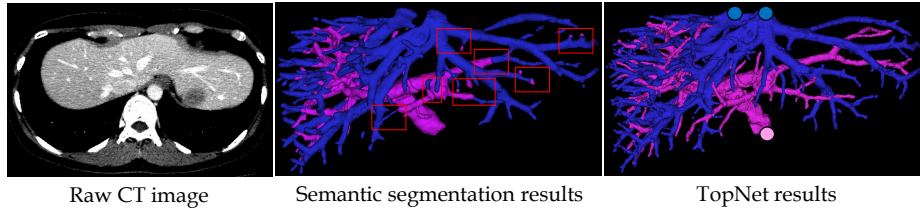
Raw CT image   Semantic segmentation results   TopNet results

Fig. 1. Portal and hepatic vein structures obtained with semantic segmentation and proposed TopNet. The red rectangles indicate vessel misclassifications.

the resection region. The analysis of blood flow requires not just segmentation of vascular structures but their representation as a tree structure.

Rule based methods for hepatic and portal vein reconstruction have been studied and a comprehensive literature survey can be found in [6]. Most popular rule based methods are based on global optimization techniques like graph cuts [20,1]. Graph cut based techniques use a handcrafted energy term between vascular nodes. A downside of such methods is that they work well only in conditions for which the rules were handcrafted. For instance when both portal and hepatic veins are high in contrast [20]. Recently, deep learning-based semantic segmentation methods have been proposed for liver vessel recognition tasks [9,13,14]. In these methods, semantic segmentation is used to classify a voxel by extracting information from a local 3D patch (also known as the networks receptive field) centered around it. However, since portal and hepatic veins look very similar locally, these methods result in noisy vessel misclassifications as shown in Figure 1. Such noisy misclassifications magnify the tree reconstruction error. Additionally, within semantic segmentation literature, simultaneous segmentation of portal and hepatic vein from CT volumes has not been tackled. Either only hepatic [13,14] or only portal vein [9] segmentation has been targeted.

In this work, we propose TopNet, a deep metric learning method for vascular tree reconstruction. The main idea is to learn the connectivity metric between vascular voxel pairs rather than to assign absolute labels. Conceptually, TopNet is a two stage network. The first stage computes center-voxels along the center-lines of all the vessels. The second stage learns the connectivity metric between center-voxel pairs in the form of topological distance between the voxels along the vascular tree. To generate a global tree, center-voxels are connected consecutively starting from the vessel source using the learned topology metric. We summarize the main contributions of this work as follows:

1. Proposed a novel multi task architecture designed for tree reconstruction, which detects nodes (voxels on vascular center lines) and estimates edges (connectivity between center-voxels) in the tree to be reconstructed.
2. Proposed a novel topology metric which learns both inter-class distance and intra-class topological distance between vascular voxel pairs in multiple trees.

3. Verified that the proposed method achieves higher accuracy than existing methods with a large margin for portal and hepatic vein reconstruction.

2 Related work

TopNet uses deep metric learning to connect detected center-voxels. In this regard, our approach is closer to deep metric learning approaches for image instance segmentation tasks [4,15,5,18]. These methods map pixels from image space to features space such that pixels belonging to the same instance are close to each other and separated by a margin if they belong to different instances. The methods differ in how they measure similarity in the feature space. Among the similarity metrics, cosine similarity has been used extensively [15,18]. Another way in which these methods differ is the way pixels are clustered in the learned feature space to generate instances. In this work, the task is not just instance segmentation (vessel segmentation) but to learn connectivity within the instance (vessel trees). For this purpose, we propose a novel deep metric learning approach.

3 TopNet

The proposed TopNet is a multi-task 3D Fully Convolutional Neural Network (3D-FCN). All the tasks share a common base encoder and task specific decoders as shown in Figure 2. The proposed architecture is similar to 3D-UNet [3], with the difference that our approach has three decoders for three different tasks instead of one. Such Multi-task architectures with shared encoder and task specific decoders have also been proposed earlier [12]. For detailed architecture, please refer to the supplementary material. The first task is extraction of all the vessels from the image. The second task is assigning a centerness score to all the voxels within the vessel mask such that the centermost voxel is assigned a minimum value. Using this centerness score, center-voxels are computed using non-maximum suppression. The final task is learning connectivity between center-voxels, which is proportional to topological distance between the center-voxels. Figure 2 shows the conceptual representation of topological distance with solid red lines. The network is trained by minimizing the sum of loss terms from each of the three tasks. The formulation of loss for each task is described in the following sections.

3.1 Vessel Extraction Decoder

The task of separating all the vessels from the background is formulated as a semantic segmentation problem. The loss function used for this task is the dice loss function which has shown to improve the accuracy in segmentation tasks with severe class imbalance [16]. We compute vessel extraction loss V_{loss} as
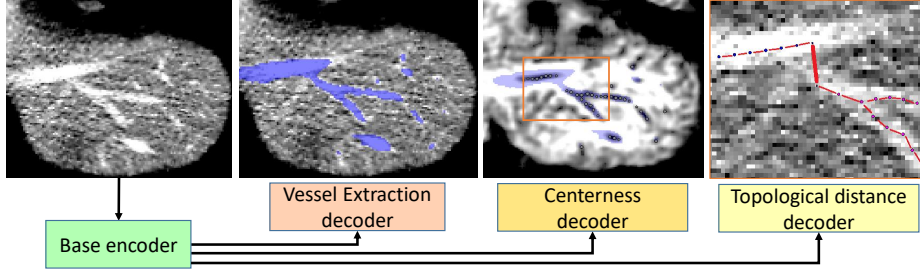
Fig. 2. TopNet methodology with base encoder and three task specific decoders.

$$V_{loss} = 1 - \frac{2\sum_{i=1}^{N} v_i v'_i}{\sum_{i=1}^{N} v_i + \sum_{i=1}^{N} v'_i}, \qquad (1)$$

where the sums run over the N voxels, of the ground truth binary vessel volume $v_i \in V$, and predicted vessel volume $v'_i \in V'$.

3.2   Centerness Decoder

The task of generating a centerness score map is formulated as a regression problem. The ground truth centerness score map is the distance transform of binary volume $m_i \in M$ representing the vessel center voxels. We consider the centerness score map only within the ground truth binary vessel volume V. The voxels outside the vessel region are masked out when computing the training loss. The training loss for the regression task is formulated as weighted $smooth_{L1}$ loss between the ground truth centerness score S_i and predicted centerness score S'_i summed over the voxels belonging to vessel mask V. We use $smooth_{L1}$ function as defined in [8] and the weight is inverse square of true centerness score i.e. S_i^2. Mathematically, the centerness loss C_{loss} is given as

$$C_{loss} = \frac{1}{\sum_{i=1}^{N} v_i} \sum_{i|v_i=1} \frac{1}{S_i^2} \left(smooth_{L1}(S_i - S'_i) \right). \qquad (2)$$

Weighting the loss function is essential because the loss would otherwise be highly biased towards the voxels away from the center. During the inference, the centerness score is masked using the predicted vessel mask. The vessel center-points are extracted by first thresholding the centerness score map followed by applying non-maximum suppression (NMS) on negative centerness score map. In this work, we use a threshold of 1.5 and NMS window of $5 \times 5 \times 5$.

3.3   Topology distance Decoder

Topological distance decoder outputs an 8-dimensional feature vector for each voxel in CT volume. The network is trained to map vessel center-voxels into the

feature space such that the L_2 norm between two center-voxels in the feature space is proportional to the topological distance between along the vascular trees. For voxel pairs belonging to different vascular trees, the L_2 norm is greater than a set margin. The pairwise loss between the two voxels i and j with associated vessel labels l_i and l_j is defined as

$$Top(x_i, x_j) = \begin{cases} smooth_{L1}\left(\parallel x_i - x_j \parallel -\alpha D_{ij}\right) & \forall i, j \mid l_i = l_j, i \neq j, \\ \gamma * \max(0, [K- \parallel x_i - x_j \parallel]) & otherwise. \end{cases} \quad (3)$$

Here, $\parallel x_i - x_j \parallel$ is the L_2 norm between the feature representations of the two voxels, D_{ij} is the topological distance between the voxels along the vascular tree. α and γ are constants whose values are explained subsequently. To compute the topological loss, all such center voxel pairs which are in the local neighborhood of each other are considered. The total loss then becomes

$$T_{loss} = \frac{1}{n} \sum_{i|m_i=1} \sum_{j \in N_i} Top\left(x_i, x_j\right). \quad (4)$$

Here, N_i is the set of voxels in the local neighborhood of voxel i and n is the total number of such voxel pairs for which the loss is computed. In this work, we use all voxel pairs which lie inside a 3D sphere of radius 15 voxels in the image space. Thus, the topological distance is normalized from 0 to 1 using the proportionality factor $\alpha = 1/15$. This roughly means that the voxels belonging to the same vascular tree will follow $0 \leq \parallel x_i - x_j \parallel \leq 1$ and if they belong to different vascular trees, then $\parallel x_i - x_j \parallel \geq 3$. To balance out the two loss terms, we use $\gamma = 1/3$.

3.4   Vascular tree reconstruction

Vascular tree is constructed by aggregating the center-voxels starting from the vessel sources using the topological distance metric learned by the TopNet. We train a separate network which outputs the location of portal and hepatic vein sources (see supplementary material). Typical portal (pink disc) and hepatic vein (blue disc) sources are shown in Figure 1. For computed vessel sources, the vascular trees are constructed using Dijkstras multi-source shortest path tree algorithm. To that end, the undirected graph is formulated as follows. The vertices of the graph are center-voxels. The weighted edges of the graph are given by the set $\left\{(w_{ij}/\alpha)^2\right\}_{w_{ij} \leq 2}$. Here, $w_{ij} = \parallel x_i - x_j \parallel$, $i \mid m_i = 1$ and $j \mid m_j = 1$ represents two center-voxels in a local neighborhood defined by a 3D sphere of radius 15. The term $(w_{ij}/\alpha)^2$ is essentially the square of topological distance between the two center-voxels i and j. All the pairs for which $w_{ij \geq 2}$ are not considered as feasible edges, since such edges most likely belong to different vascular trees.

Table 1. Characteristics of INTERNAL and public dataset IRCAD.

Dataset	Training	Test	Contrast phase	Slice Thickness
INTERNAL	115 CT volumes	20 CT volumes	Late portal	$0.5 \sim 1.0$ mm
IRCAD [10]	None	20 CT volumes	Late portal	$1.0 \sim 4.0$ mm

4 Experiments

We divide the experiments into comparison and ablation studies. The comparison of TopNet is made against single task 3D-UNet [3] with the dice loss function [16] baseline. This is a standard and competitive baseline for 3D image segmentation and an upgrade over existing 2D-FCN based existing liver vessel segmentation methods [9,13,14]. In the ablation study, we investigate the performance with different learning metrics (dice loss based classification, cosine, and topology) using the same multi-task architecture. For the Multi-task dice loss based classification approach, the last two layers of topology distance decoder are replaced to output a two-channel probability map for each hepatic and portal vein. Here, we use a 2 class dice loss summed over all the center-voxels. Multi-task cosine metric learning method uses the exact same architecture as TopNet but uses cosine metric following from similar works [15,18]. The network uses cosine metric loss between voxel pairs i and j associated with feature vectors x_i and x_j and vessel labels l_i and l_j is defined as

$$L\left(x_i, x_j\right) = \begin{cases} 1 - 0.5(1 + S_{ij}) & \forall i, j \mid l_i = l_j, i \neq j, \\ 0.5(1 + S_{ij}) & otherwise. \end{cases} \tag{5}$$

where, $S_{ij} = \frac{\|x_i^T x_j\|}{\|x_i\|\|x_j\|}$ is the cosine similarity metric. To reconstruct vascular tree using the learned metric as mentioned in section 3.4, the weighted edges of the graph are given by the set $\{E_{ij}(1 - S_{ij})\}_{S_{ij} \geq 0}$ Here, E_{ij} is the Euclidean distance between the center-voxels.

4.1 Dataset and preprocessing

We used an internal dataset for training and both the internal as well as public IRCAD dataset for evaluation [10]. The characteristics of both the datasets is shown in Table 1. For training, the CT values were normalized from 0 to 1, voxel spacing in all three dimensions was normalized to 0.7mm. We also crop the CT volume using the liver region before setting it as input to the network.

4.2 Evaluation metrics and results

We evaluate the results of portal and hepatic vein reconstruction by comparing respective ground truth and predicted centerlines. When comparing the centerlines, an exact overlap is not possible. For this, we introduce a variable tolerance
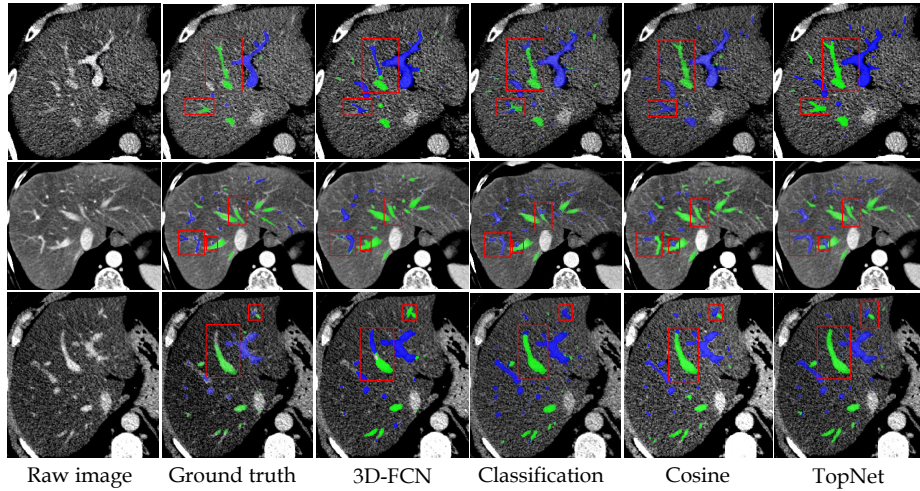
Raw image   Ground truth   3D-FCN   Classification   Cosine   TopNet

Fig. 3. Comparison between 3D-UNet and proposed TopNet for hepatic and portal vein extraction. Hepatic vein is shown in green while portal vein is shown in blue. Red boxes show regions of interest for comparing misclassifications.

term δ which is equal to the radius defined at a ground truth centerpoint following from [7]. Based on this, an illustration of True Positives (TP), False Positives (FP), False Negatives (FN), and True Negatives (TN) of portal vessel class is shown in the supplementary material. Using these metrics, we compute the dice coefficient, sensitivity and specificity for hepatic and portal centerlines. Predicted centerpoints for which there are no associated ground truth centerpoints within the tolerance are ignored while computing the metrics. Usually, these points would be counted as false positives but we ignore them because in the IRCAD dataset, small vessels are often unlabelled.

The proposed TopNet considerably outperforms the baseline i.e. single task 3D-UNet method with over 7-8% improvement in all metrics as shown in Table 2. When Multi-task dice loss classification is compared to single-task 3D-UNet which also uses the dice loss, a performance improvement of 3-4% is seen. By this, we can say that multi-stage vessel segmentation works better than single-stage method. Within the Multi-task network architecture, TopNet gives the best results. We observe that if not for topology metric learning, the problem of abrupt change in vessel labels along the vascular tree (large and medium red boxes) persists as shown in Figure 3. Such misclassifications also occur when portal and hepatic veins cross each other in close vicinity (small red boxes in second and third row). We reason that such misclassifications arise because conventional margin based techniques like semantic segmentation or even cosine metric learning classify or seperate the center-voxels based on frequently seen vessel patterns in different liver regions within the training dataset. If unique vessel patterns are observed in test data, the classification fails. With Topology metric learning on the other hand, we explicitly constrain the network to learn
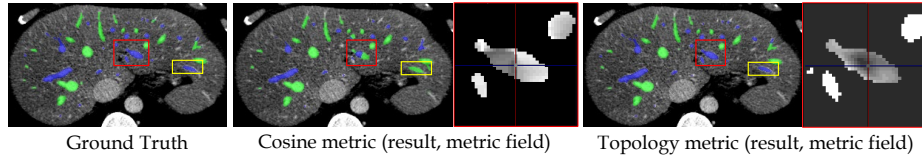
Ground Truth   Cosine metric (result, metric field)   Topology metric (result, metric field)

Fig. 4. Comparison between Topology and cosine metric learning. Hepatic vein is shown in green while portal vein is shown in blue. Boxes show regions of interest for comparing misclassifications.

Table 2. Comparison between TopNet and existing methods.

Method	Vessel	IRCAD			INTERNAL		
		Dice	Specificity	Sensitivity	Dice	Specificity	Sensitivity
Single Task							
3D-UNet	Portal	0.84	0.85	0.82	0.88	0.86	0.87
[3]	Hepatic	0.84	0.82	0.85	0.86	0.87	0.86
Multi Task							
Classification	Portal	0.87	0.85	0.88	0.92	0.93	0.90
	Hepatic	0.86	0.88	0.85	0.91	0.90	0.93
Cosine metric	Portal	0.85	0.85	0.81	0.88	0.84	0.9
	Hepatic	0.80	0.82	0.86	0.85	0.90	0.84
TopNet	Portal	**0.91**	**0.96**	**0.91**	**0.95**	**0.96**	**0.94**
	Hepatic	**0.92**	**0.91**	**0.96**	**0.94**	**0.94**	**0.96**

to trace the vessels. To verify this direction of reasoning, we show the similarity metric field from a reference center-voxel to it's neighborhood voxels in a misclassified region in Figure 3. It can be seen that the Topology metric shows a smooth transitioning of the similarity field as compared to cosine metric which shows an abrupt change. Since the cosine metric learning method does not learn absolute labels, such misclassifications propagate as shown with the yellow box in Figure 4. This explains why the cosine metric learning works poorly as compared to the multi-task classification approach as shown in Table 2.

5   Conclusion

We proposed a deep metric learning method which learns multi-label tree structure connectivity from images. We converted a hard problem of assigning absolute labels to vessels to a simpler one. Topological metric learning is simpler because it just learns local connectivity and does not require global image context. For global connectivity, we use the shortest path tree algorithm which uses the learned metric to connect vascular voxels. The results show that using topological metric learning, the issue of noisy misclassifications is resolved. We believe that the approach is general enough to be applicable to vessel segmentation tasks

in other organs. This is because the network is constrained to learn connectivity rather than other absolute labels. There are a few downsides to this approach as well. First, the method has multiple manually fine tuned parameters involved both in detecting the center-voxels and creating vascular trees using the learned topology metric. Second, the shortest path tree algorithm which is used to create global connectivity is sensitive to misdetection of vascular center voxels. Due to this, in the future it would be interesting to explore a method which can embed tree generation algorithms like the shortest path into the training process.

References

1. Bauer, C., Pock, T., Sorantin, E., Bischof, H., Beichel, R.: Segmentation of inter-woven 3d tubular tree structures utilizing shape priors and graph cuts. Medical image analysis **14**(2), 172–184 (2010)
2. Center, M.M., Jemal, A.: International trends in liver cancer incidence rates. Cancer Epidemiology and Prevention Biomarkers **20**(11), 2362–2368 (2011)
3. Çiçek, Ö., Abdulkadir, A., Lienkamp, S.S., Brox, T., Ronneberger, O.: 3d u-net: learning dense volumetric segmentation from sparse annotation. In: International conference on medical image computing and computer-assisted intervention. pp. 424–432. Springer (2016)
4. De Brabandere, B., Neven, D., Van Gool, L.: Semantic instance segmentation with a discriminative loss function. arXiv preprint arXiv:1708.02551 (2017)
5. Fathi, A., Wojna, Z., Rathod, V., Wang, P., Song, H.O., Guadarrama, S., Murphy, K.P.: Semantic instance segmentation via deep metric learning. arXiv preprint arXiv:1703.10277 (2017)
6. Fraz, M.M., Remagnino, P., Hoppe, A., Uyyanonvara, B., Rudnicka, A.R., Owen, C.G., Barman, S.A.: Blood vessel segmentation methodologies in retinal images–a survey. Computer methods and programs in biomedicine **108**(1), 407–433 (2012)
7. Gegúndez-Arias, M.E., Aquino, A., Bravo, J.M., Marín, D.: A function for quality evaluation of retinal vessel segmentations. IEEE transactions on medical imaging **31**(2), 231–239 (2011)
8. Girshick, R.: Fast r-cnn. In: Proceedings of the IEEE international conference on computer vision. pp. 1440–1448 (2015)
9. Ibragimov, B., Toesca, D., Chang, D., Koong, A., Xing, L.: Combining deep learning with anatomical analysis for segmentation of the portal vein for liver sbrt planning. Physics in Medicine & Biology **62**(23), 8943 (2017)
10. IRCAD: Ircad dataset for liver vessel segmentation. `https://www.ircad.fr/research/3d-ircadb-01/` (march 2020)
11. Kazaryan, A.M., Marangos, I.P., Røsok, B.I., Rosseland, A.R., Villanger, O., Fosse, E., Mathisen, Ø., Edwin, B.: Laparoscopic resection of colorectal liver metastases: surgical and long-term oncologic outcome. Annals of surgery **252**(6), 1005–1012 (2010)
12. Keshwani, D., Kitamura, Y., Li, Y.: Computation of total kidney volume from ct images in autosomal dominant polycystic kidney disease using multi-task 3d convolutional neural networks. In: International Workshop on Machine Learning in Medical Imaging. pp. 380–388. Springer (2018)
13. Kitrungrotsakul, T., Han, X.H., Iwamoto, Y., Foruzan, A.H., Lin, L., Chen, Y.W.: Robust hepatic vessel segmentation using multi deep convolution network. In: Med-

ical Imaging 2017: Biomedical Applications in Molecular, Structural, and Functional Imaging. vol. 10137, p. 1013711. International Society for Optics and Photonics (2017)

14. Kitrungrotsakul, T., Han, X.H., Iwamoto, Y., Lin, L., Foruzan, A.H., Xiong, W., Chen, Y.W.: Vesselnet: A deep convolutional neural network with multi pathways for robust hepatic vessel segmentation. Computerized Medical Imaging and Graphics **75**, 74–83 (2019)

15. Kong, S., Fowlkes, C.C.: Recurrent pixel embedding for instance grouping. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 9018–9028 (2018)

16. Milletari, F., Navab, N., Ahmadi, S.A.: V-net: Fully convolutional neural networks for volumetric medical image segmentation. In: 2016 Fourth International Conference on 3D Vision (3DV). pp. 565–571. IEEE (2016)

17. Mise, Y., Hasegawa, K., Satou, S., Shindoh, J., Miki, K., Akamatsu, N., Arita, J., Kaneko, J., Sakamoto, Y., Kokudo, N.: How has virtual hepatectomy changed the practice of liver surgery?: experience of 1194 virtual hepatectomy before liver resection and living donor liver transplantation. Annals of surgery **268**(1), 127–133 (2018)

18. Payer, C., Štern, D., Neff, T., Bischof, H., Urschler, M.: Instance segmentation and tracking with cosine embeddings and recurrent hourglass networks. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 3–11. Springer (2018)

19. Wakabayashi, G., Cherqui, D., Geller, D.A., Buell, J.F., Kaneko, H., Han, H.S., Asbun, H., O'Rourke, N., Tanabe, M., Koffron, A.J., et al.: Recommendations for laparoscopic liver resection: a report from the second international consensus conference held in morioka. Annals of surgery **261**(4), 619–629 (2015)

20. Zeng, Y.z., Zhao, Y.q., Tang, P., Liao, M., Liang, Y.x., Liao, S.h., Zou, B.j.: Liver vessel segmentation and identification based on oriented flux symmetry and graph cuts. Computer methods and programs in biomedicine **150**, 31–39 (2017)
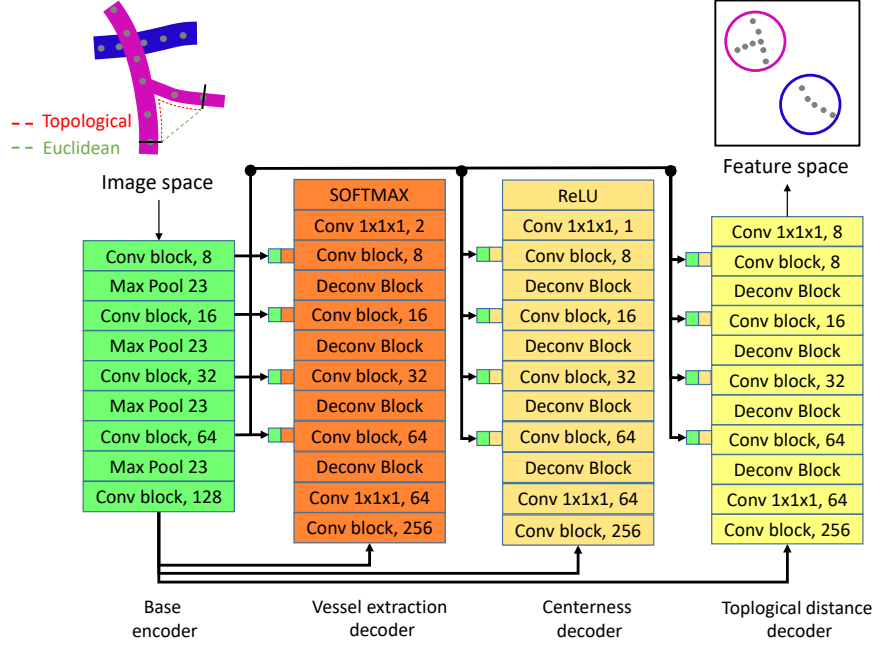
Acknowledgements

Figure 1: TopNet Architecture: The Conv block in each encoder and decoder denotes two sets of $Conv[3 \times 3 \times 3] \rightarrow BatchNorm \rightarrow ReLU$ operations. The Deconv blocks are transposed convolutional operators with filter size $4 \times 4 \times 4$. Skip connections are used between the encoder and all three task specific decoders and the features are concatenated.
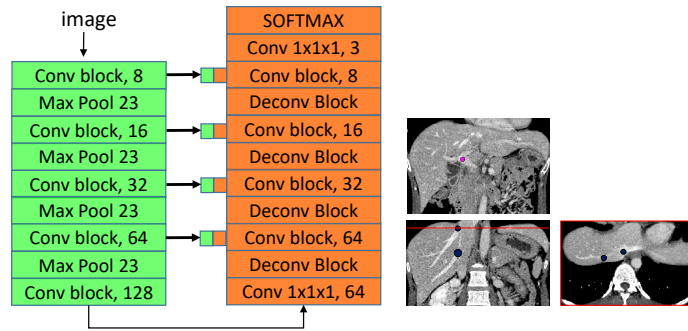

Figure 2: VesselSource detector. Left: Network architecture which outputs a 3 channel probability map for background region, portal vein source and hepatic vein source points respectively, Right: detected portal vein (pink) and hepatic vein (blue) sources.
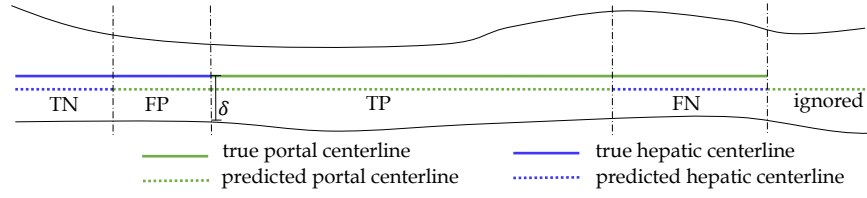
1

Figure 3: An illustrations of terms used in the evaluation metrics.
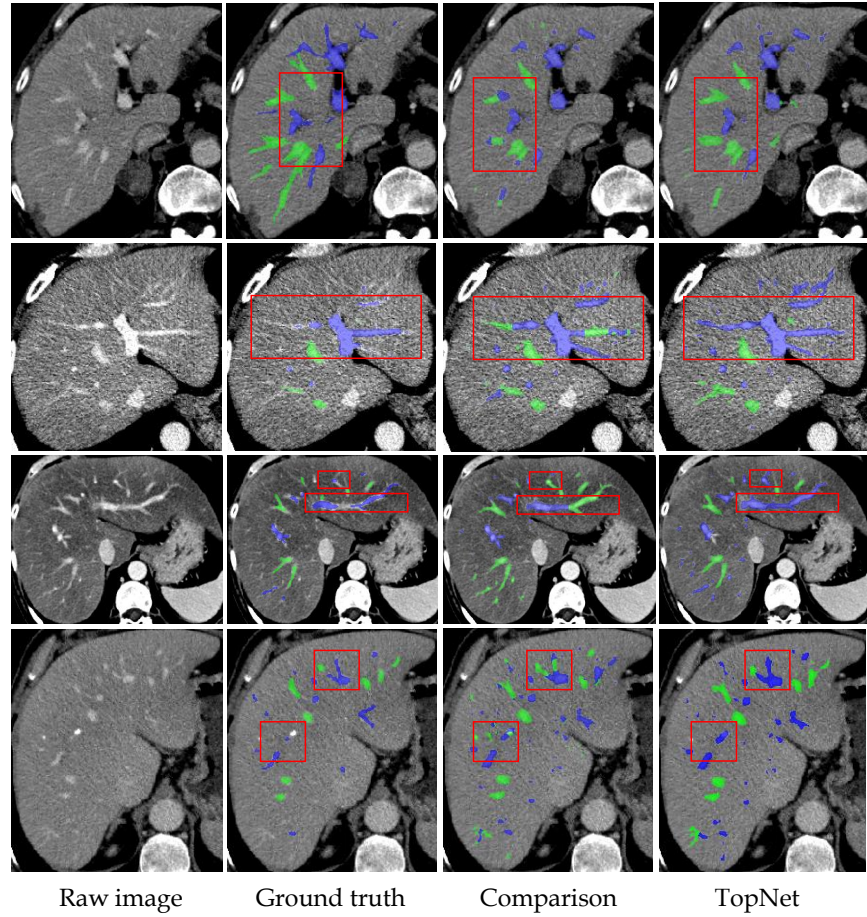

| Raw image | Ground truth | Comparison | TopNet |

Figure 4: Comparison of TopNet with different approaches on IRCAD datset. Red boxes show regions of interest. First two rows: Single-task 3D-FCN Vs TopNet. Third now: Multi-task Cosine metric Vs TopNet. Fourth row: Multi-task classification Vs TopNet.

2